Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 2908512 C 2023/10/24

(11)(21) **2 908 512**

(12) **BREVET CANADIEN**
**CANADIAN PATENT**

(13) **C**

(86) Date de dépôt PCT/PCT Filing Date: 2014/04/02
(87) Date publication PCT/PCT Publication Date: 2014/10/09
(45) Date de délivrance/Issue Date: 2023/10/24
(85) Entrée phase nationale/National Entry: 2015/09/29
(86) N° demande PCT/PCT Application No.: US 2014/032706
(87) N° publication PCT/PCT Publication No.: 2014/165612
(30) Priorités/Priorities: 2013/04/05 (US61/809,097); 2013/05/07 (US61/820,461)

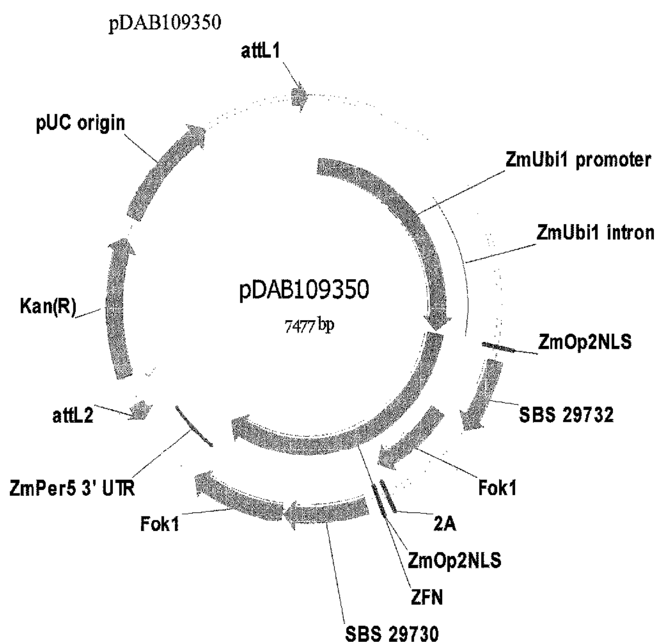
(51) Cl.Int./Int.Cl. *C12N 5/04* (2006.01), *A01H 1/00* (2006.01), *C07K 14/00* (2006.01), *C12N 15/04* (2006.01), *C12N 15/90* (2006.01)

(72) Inventeurs/Inventors:
AINLEY, W. MICHAEL, US;
GUSCHIN, DMITRY Y., US;
HAYDEN, MATTHEW, US;
ISENEGGER, DANIEL, US;
MASON, JOHN, US;
...

(73) Propriétaires/Owners:
SANGAMO THERAPEUTICS, INC., US;
CORTEVA AGRISCIENCE LLC, US

(74) Agent: SMART & BIGGAR LP

(54) Titre : PROCEDES ET COMPOSITIONS PERMETTANT D'INTEGRER UNE SEQUENCE EXOGENE AU SEIN DU GENOME DE PLANTES
(54) Title: METHODS AND COMPOSITIONS FOR INTEGRATION OF AN EXOGENOUS SEQUENCE WITHIN THE GENOME OF PLANTS




(57) Abrégé/Abstract:
Disclosed herein are methods and compositions for parallel or sequential transgene stacking in plants to produce plants with selected phenotypes. The present disclosure provides methods and compositions for precision transformation, gene targeting,

(11)(21) **2 908 512**

(13) **C**

(72) **Inventeurs(suite)/Inventors(continued)**: MILLER, JEFFREY C., US; PETOLINO, JOSEPH F., US; RAN, YIDONG, US; SAWBRIDGE, TIM, US; SPANGENBERG, GERMAN, US; WEBB, STEVEN R., US

(57) **Abrégé(suite)/Abstract(continued)**:
targeted genomic modification and protein expression in plants. In particular, the present disclosure describes a novel, transgenic marker-free strategy for integrating an exogenous sequence and to stack traits that exploit differential selection at an endogenous locus, e.g., the acetohydroxyacid synthase (AHAS) locus) in plant genomes.

CA 02908512 2015-09-29

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
9 October 2014 (09.10.2014)

(10) International Publication Number
WO 2014/165612 A3

(51) International Patent Classification: *A01H 1/00* (2006.01)

(21) International Application Number: PCT/US2014/032706

(22) International Filing Date: 2 April 2014 (02.04.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/809,097 5 April 2013 (05.04.2013) US
61/820,461 7 May 2013 (07.05.2013) US

(71) Applicants: **DOW AGROSCIENCES LLC** [US/US]; 9330 Zionsville Road, Indianapolis, Indiana 46268 (US). **SANGAMO BIOSCIENCES, INC.** [US/US]; 501 Canal Blvd., Suite A100, Richmond, California 94804 (US).

(72) Inventors: **AINLEY, W. Michael**; c/o Dow AgroSciences LLC, 9330 Zionsville Road, Indianapolis, Indiana 46268 (US). **GUSCHIN, Dmitry Y.**; c/o Sangamo BioSciences, Inc., 501 Canal Blvd., Suite A100, Richmond, California 94804 (US). **HAYDEN, Matthew**; c/o Dow AgroSciences LLC, 9330 Zionsville Road, Indianapolis, Indiana 46268 (US). **ISENEGGER, Daniel**; c/o Dow AgroSciences LLC, 9330 Zionsville Road, Indianapolis, Indiana 46268 (US). **MASON, John**; c/o Dow AgroSciences LLC, 9330 Zionsville Road, Indianapolis, Indiana 46268 (US). **MILLER, Jeffrey C.**; c/o Sangamo BioSciences, Inc., 501 Canal Blvd., Suite A100, Richmond, California 94804 (US). **PETOLINO, Joseph F.**; c/o Dow AgroSciences LLC, 9330 Zionsville Road, Indianapolis, Indiana 46268 (US). **RAN, Yidong**; c/o Dow AgroSciences LLC, 9330 Zionsville Road, Indianapolis, Indiana 46268 (US). **SAWBRIDGE, Tim**; c/o Dow AgroSciences LLC, 9330 Zionsville Road, Indianapolis, Indiana 46268 (US). **SPANGENBERG, German**; c/o Dow AgroSciences LLC, 9330 Zionsville Road, Indianapolis, Indiana 46268 (US). **WEBB, Steven R.**; c/o Dow AgroSciences LLC, 9330 Zionsville Road, Indianapolis, Indiana 46268 (US).

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

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME,

*[Continued on next page]*

**(54) Title**: METHODS AND COMPOSITIONS FOR INTEGRATION OF AN EXOGENOUS SEQUENCE WITHIN THE GENOME OF PLANTS

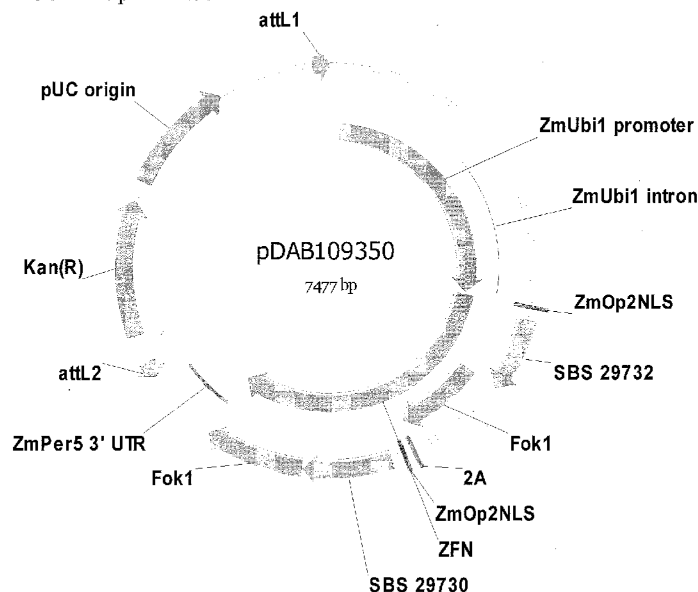



**FIGURE 1:** pDAB109350

**(57) Abstract**: Disclosed herein are methods and compositions for parallel or sequential transgene stacking in plants to produce plants with selected phenotypes. The present disclosure provides methods and compositions for precision transformation, gene targeting, targeted genomic modification and protein expression in plants. In particular, the present disclosure describes a novel, transgenic marker-free strategy for integrating an exogenous sequence and to stack traits that exploit differential selection at an endogenous locus, e.g., the acetohydroxyacid synthase (AHAS) locus) in plant genomes.

WO 2014/165612 A3

MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

**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))*

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

**(88) Date of publication of the international search report**: 15 January 2015

# METHODS AND COMPOSITIONS FOR INTEGRATION OF AN EXOGENOUS SEQUENCE WITHIN THE GENOME OF PLANTS

## CROSS-REFERENCE TO RELATED APPLICATIONS

**[0001]** The application claims the benefit of U.S. Provisional Application No. 61/809,097, filed on April 5, 2013 and U.S. Provisional Application No. 61/820,461, filed on May 7, 2013.

## STATEMENT OF RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH

**[0002]** Not applicable.

## REFERENCE TO SEQUENCE LISTING SUBMITTED ELECTRONICALLY

**[0003]** The official copy of the sequence listing is submitted electronically via EFS-Web as an ASCII formatted sequence listing concurrently with the specification.

## TECHNICAL FIELD

**[0004]** The present disclosure is in the field of genomic engineering, particularly in the integration of exogenous sequences into plants, including simultaneous genomic editing of multiple alleles over multiple genomes, including in polyploid plants.

## BACKGROUND

**[0005]** To meet the challenge of increasing global demand for food production, many effective approaches to improving agricultural productivity (*e.g.*, enhanced yield or engineered pest resistance) rely on either mutation breeding or introduction of novel genes into the genomes of crop species by transformation. Both processes are inherently non-specific and relatively inefficient. For example, conventional plant transformation methods deliver exogenous DNA that integrates into the genome at random locations. The random nature of these methods makes it

necessary to generate and screen hundreds of unique random-integration events per construct in order to identify and isolate transgenic lines with desirable attributes. Moreover, conventional transformation methods create several challenges for transgene evaluation including: (a) difficulty for predicting whether pleiotropic effects due to unintended genome disruption have occurred; and (b) difficulty for comparing the impact of different regulatory elements and transgene designs within a single transgene candidate, because such comparisons are complicated by random integration into the genome. As a result, conventional plant trait engineering is a laborious and cost intensive process with a low probability of success.

**[0006]** Precision gene modification overcomes the logistical challenges of conventional practices in plant systems, and as such has been a longstanding but elusive goal in both basic plant biology research and agricultural biotechnology. However, with the exception of "gene targeting" via positive-negative drug selection in rice or the use of pre-engineered restriction sites, targeted genome modification in all plant species, both model and crop, has until recently proven very difficult. Terada et al. (2002) *Nat Biotechnol* 20(10):1030; Terada et al. (2007) *Plant Physiol* 144(2):846; D'Halluin et al. (2008) *Plant Biotechnology J.* 6(1):93.

**[0007]** Recently, methods and compositions for targeted cleavage of genomic DNA have been described. Such targeted cleavage events can be used, for example, to induce targeted mutagenesis, induce targeted deletions of cellular DNA sequences, and facilitate targeted recombination and integration at a predetermined chromosomal locus. *See*, for example, Urnov *et al.* (2010) *Nature* 435(7042):646-51; United States Patent Nos. 8,586,526; 8,586,363; 8,409,861; 8,106,255; 7,888,121; 8,409,861and U.S. Patent Publications 20030232410; 20050026157; 20090263900; 20090117617; 20100047805; 20100257638; 20110207221; 20110239315; 20110145940. Cleavage can occur through the use of specific nucleases such as engineered zinc finger nucleases (ZFN), transcription-activator like effector nucleases (TALENs), or using the CRISPR/Cas system with an engineered crRNA/tracr RNA ('single guide RNA') to guide specific cleavage. U.S. Patent Publication No. 20080182332 describes the use of non-canonical zinc finger nucleases (ZFNs) for targeted modification of plant genomes; U.S. Patent No. 8,399,218 describes ZFN-mediated targeted modification of a plant EPSPS locus; U.S. Patent No. 8,329,986 describes targeted modification of a plant *Zp15* locus and U.S. Patent No. 8,592,645 describes

targeted modification of plant genes involved in fatty acid biosynthesis. In addition, Moehle *et al.* (2007) *Proc. Natl. Acad, Sci. USA* 104(9):3055-3060 describes using designed ZFNs for targeted gene addition at a specified locus. U.S. Patent Publication 20110041195 describes methods of making homozygous diploid organisms.

**[0008]** Transgene (or trait) stacking has great potential for production of plants, but has proven difficult. *See, e.g.,* Halpin (2005) *Plant Biotechnology Journal* 3:141–155. In addition, polyploidy, where the organism has two or more duplicated (autoploidy) or related (alloploid) paired sets of chromosomes, occurs more often in plant species than in animals. For example, wheat has lines that are diploid (two sets of chromosomes), tetraploid (four sets of chromosomes) and hexaploid (six sets of chromosomes). In addition, many agriculturally important plants of the genus *Brassica* are also allotetraploids.

**[0009]** Thus, there remains a need for compositions and methods for the identification, selection and rapid advancement of stable targeted integration into precise locations within a plant genome, including simultaneous modification of multiple alleles across different genomes of polyploid plants, for establishing stable, heritable genetic modifications in a plant and its progeny.

## SUMMARY

**[0010]** The present disclosure provides methods and compositions for precision transformation, gene targeting, targeted genomic modification and protein expression in plants. In particular, the present disclosure describes a novel, transgenic marker-free strategy for integrating an exogenous sequence and to stack traits that exploit differential selection at an endogenous locus (*e.g.,* the acetohydroxyacid synthase (AHAS) locus) in plant genomes. The strategy facilitates generation of plants that have one or more transgenes (or one or more genes of interest (GOI), wherein the transgenes do not include transgenic selectable marker genes) precisely positioned at an endogenous plant locus, for example, at one or more AHAS paralogs. The methods and compositions described herein enable both parallel and sequential transgene stacking in plant genomes at precisely the same genomic location, including simultaneous editing of multiple alleles across multiple genomes of polyploid plant species. In addition, the methods and compositions of the invention allow for exogenous transgenic selectable marker-free selection and/or genomic modification of an endogenous gene in which the genomic modification produces a mutation in the

endogenous gene such that the endogenous gene produces a product that results in an herbicide tolerant plant (*e.g.,* by virtue of exploiting known mutations in an endogenous gene such as known mutations in AHAS gene that confer tolerance to Group B herbicides, or ALS inhibitor herbicides such as imidazolinone or sulfonylurea). Also provided are cells (*e.g.,* seeds), cell lines, organisms (*e.g.,* plants), *etc.* comprising these transgene-stacked and/or simultaneously-modified alleles. The targeted genomic editing (insertions, deletions, mutations, transgene stacking) can result, for example, in increased crop yield, a protein encoding disease resistance, a protein that increases growth, a protein encoding insect resistance, a protein encoding herbicide tolerance and the like. Increased yield can include, for example, increased amount of fruit or grain yield, increased biomass of the plant (or fruit or grain of the plant), higher content of fruit flesh, larger plants, increased dry weight, increased solids context, higher total weight at harvest, enhanced intensity and/or uniformity of color of the crop, altered chemical (*e.g.,* oil, fatty acid, carbohydrate, protein) characteristics, *etc.*

**[0011]** Thus, in one aspect, disclosed herein are methods and compositions for precise, genomic modification (*e.g.,* transgene stacking) at one or more endogenous alleles of a plant gene. In certain embodiments, the transgene(s) is(are) integrated into an endogenous locus of a plant genome (*e.g.,* polyploid plant). Transgene integration includes integration of multiple transgenes, which may be in parallel (simultaneous integration of one or more transgenes into one or more alleles) or sequential. In certain embodiments, the transgene does not include a transgenic marker, but is integrated into an endogenous locus that is modified upon integration of the transgene comprising a trait, for example, integration of the transgene(s) into an endogenous acetohydroxyacid synthase (AHAS) locus (*e.g.,* the 3' untranslated region of the AHAS locus) such that the transgene is expressed and the AHAS locus is modified to alter herbicide tolerance (*e.g.,* Group B herbicides, or ALS inhibitor herbicides such as imidazolinone or sulfonylurea). The transgene(s) is(are) integrated in a targeted manner using one or more non-naturally occurring nucleases, for example zinc finger nucleases, meganucleases, TALENs and/or a CRISPR/Cas system with an engineered single guide RNA. The transgene can comprise one or more coding sequences (*e.g.,* proteins), non-coding sequences and/or may produce one or more RNA molecules (*e.g.,* mRNA, RNAi, siRNA, shRNA, *etc.*). In certain embodiments, the transgene integration is simultaneous (parallel). In other

embodiments, sequential integration of one or more transgenes (GOIs) is achieved, for example by the AHAS locus, by alternating between different herbicide (Group B, or ALS inhibitor herbicides such as imidazolinone or sulfonylurea) chemical selection agents and known AHAS mutations conferring tolerance to those specific herbicides. Furthermore, any of the plant cells described herein may further comprise one or more additional transgenes, in which the additional transgenes are integrated into the genome at a different locus (or different loci) than the target allele(s) for transgene stacking. Thus, a plurality of endogenous loci may include integrated transgenes in the cells described herein.

**[0012]** In another aspect, disclosed herein are polyploid plant cells in which multiple alleles of one or more genes across the different genomes (sub-genomes) have been simultaneously modified. The targeted modifications may enhance or reduce gene activity (*e.g.,* endogenous gene activity and/or activity of an integrated transgene) in the polyploid plant, for example mutations in AHAS that alter (*e.g.,* increase) herbicide tolerance.

**[0013]** In certain embodiments, the targeted genomic modification in the polyploid plant cell comprises a small insertion and/or deletion, also known as an indel. Any of the plant cells described herein may be within a plant or plant part (*e.g.,* seeds, flower, fruit), for example, any variety of: wheat, soy, maize, potato, alfalfa, rice, barley, sunflower, tomato, *Arabidopsis*, cotton, *Brassica* species (including but not limited to *B. napus, B. rapa, B. oleracea, B. nigra, B. juncea, B. carinata*), *Brachypodium*, timothy grass and the like.

**[0014]** In another aspect, described herein is a DNA-binding domain (*e.g.,* zinc finger protein (ZFP)) that specifically binds to a gene involved in herbicide tolerance, for example, an *AHAS* gene. The zinc finger protein can comprise one or more zinc fingers (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9 or more zinc fingers), and can be engineered to bind to any sequence within a polyploid plant genome. Any of the zinc finger proteins described herein may bind to a target site within the coding sequence of the target gene or within adjacent sequences (*e.g.,* promoter or other expression elements). In certain embodiments, the zinc finger protein binds to a target site in an AHAS gene, for example, as shown in Table 3 and Table 13. The recognition helix regions of exemplary AHAS-binding zinc fingers are shown in Table 2 and Table 12. One or more of the component zinc finger binding domains of the zinc finger protein

can be a canonical (C2H2) zinc finger or a non-canonical (*e.g.,* C3H) zinc finger (*e.g.,* the N-terminal and/or C-terminal zinc finger can be a non-canonical finger).

**[0015]** In another aspect, disclosed herein are fusion proteins, each fusion protein comprising a DNA-binding domain (*e.g.*, a zinc finger protein) that specifically binds to multiple alleles of a gene in polyploid plant genomes. In certain embodiments, the proteins are fusion proteins comprising a zinc finger protein and a functional domain, for example a transcriptional activation domain, a transcriptional repression domain and/or a cleavage domain (or cleavage half-domain). In certain embodiments, the fusion protein is a zinc finger nuclease (ZFN). Cleavage domains and cleavage half domains can be obtained, for example, from various restriction endonucleases and/or homing endonucleases. In one embodiment, the cleavage half-domains are derived from a Type IIS restriction endonuclease (*e.g.*, *Fok* I).

**[0016]** In other aspects, provided herein are polynucleotides encoding any of the DNA-binding domains and/or fusion proteins described herein. In certain embodiments, described herein is an expression vector comprising a polynucleotide, encoding one or more DNA-binding domains and/or fusion proteins described herein, operably linked to a promoter. In one embodiment, one or more of the fusion proteins are ZFNs.

**[0017]** The DNA-binding domains and fusion proteins comprising these DNA-binding domains bind to and/or cleave two or more endogenous genes in a polyploid genome (*e.g.,* an AHAS gene) within the coding region of the gene or in a non-coding sequence within or adjacent to the gene, such as, for example, a leader sequence, trailer sequence or intron, or promoter sequence, or within a non-transcribed region, either upstream or downstream of the coding region, for example the 3' untranslated region. In certain embodiments, the DNA-binding domains and/or fusion proteins bind to and/or cleave a coding sequence or a regulatory sequence of the target gene.

**[0018]** In another aspect, described herein are compositions comprising one or more proteins, fusion proteins or polynucleotides as described herein. Polyploid plant cells contain multiple genomic allelic targets. Thus, compositions described herein may comprise one or more DNA-binding proteins (and polynucleotides encoding same) that target (and simultaneously modify) multiple alleles present in multiple genomes (also referred to as sub-genomes) of a polyploid plant cell. The DNA-

binding proteins may target all genes (paralogs), one or multiple (but less than all) selected alleles.

**[0019]** In another aspect, provided herein is a method for simultaneously altering multiple alleles across the multiple genomes of a polyploid plant cell, the method comprising, expressing one or more DNA-binding domain proteins (*e.g.,* zinc finger proteins such as zinc finger nucleases) in the cell such that multiple alleles of the polyploid plant are altered. In certain embodiments, altering expression of one or more AHAS genes in a plant cell, the method comprising, expressing one or more DNA-binding domain containing proteins (*e.g.*, zinc finger proteins) in the cell such that expression of AHAS is altered. In certain embodiments, the methods comprise using a pair of zinc finger nucleases to create a small insertion and/or deletion ("indel") that disrupts endogenous gene expression. In other embodiments, the methods comprise using a pair of zinc finger nucleases to enhance gene expression, for example via targeted insertion of an exogenous sequence (*e.g.*, donor sequence, GOI, or transgene) or expression enhancing element. The altered gene expression/function can result in increased photosynthesis, increased herbicide tolerance and/or modifications in growth within plant cells.

**[0020]** In another aspect, provided herein are nucleic acids and antibodies, and methods of using the same, for detecting and/or measuring altered expression of and modifications to multiples alleles of a gene (*e.g.,* AHAS).

**[0021]** In another aspect, described herein is a method for simultaneously modifying one or more endogenous genes in a polyploid plant cell. In certain embodiments, the method comprising: (a) introducing, into the polyploid plant cell, one or more expression vectors encoding one or more nucleases (*e.g.,* ZFNs, TALENs, meganucleases and/or CRISPR/Cas systems) that bind to a target site in the one or more genes under conditions such that the nucleases cleave the one or more endogenous genes, thereby modifying the one or more endogenous (*e.g.,* AHAS) genes. In other embodiments, more than one allele of an endogenous gene is cleaved, for example in polyploid plants. In other embodiments, one or more alleles of more than one endogenous gene is cleaved. Furthermore, in any of the methods described herein, cleavage of the one or more genes may result in deletion, addition and/or substitution of nucleotides in the cleaved region, for example such that AHAS activity is altered (*e.g.,* enhanced or reduced), thereby allowing for assessment of, for example, transgene integration at or near the modified endogenous genes.

**[0022]** In yet another aspect, described herein is a method for introducing one or more exogenous sequences into the genome of a plant cell, the method comprising the steps of: (a) contacting the cell with the one or more exogenous sequences (e.g., donor vector, transgene or GOI, or combinations thereof); and (b) expressing one or more nucleases (*e.g.,* ZFNs, TALENs, meganucleases and/or CRISPR/Cas systems) as described herein in the cell, wherein the one or more nucleases cleave chromosomal DNA; such that cleavage of chromosomal DNA in step (b) stimulates incorporation of the exogenous sequence into the genome by homologous recombination. In certain embodiments, the chromsomal DNA is modified such that the chromosomal sequence (*e.g.,* endogenous gene) is mutated to_expresses a product that produces a selectable phenotype (*e.g.,* herbicide tolerance). Multiple exogenous sequences may be integrated simultaneously (in parallel) or the steps may be repeated for sequential addition of transgenes (transgene stacking). In certain embodiments, the one or more transgenes are introduced within an AHAS gene, for example the 3' untranslated region. In any of the methods described herein, the one or more nucleases may be fusions between a nuclease (cleavage) domain (*e.g.,* a cleavage domain of a Type IIs restriction endonuclease or a meganuclease) and an engineered zinc finger binding domain. In other embodiments, the nuclease comprises a TAL effector domain, a homing endonuclease and/or a Crispr/Cas single guide RNA. In any of the methods described herein, the exogenous sequence may encode a protein product and/or produce an RNA molecule. In any of the methods described herein, the exogenous sequence may be integrated such that endogenous locus into which the exogenous sequence(s) is(are) inserted is modified to produce one or more measurable phenotypes or markers (*e.g.,* herbicide tolerance by mutation of endogenous AHAS).

**[0023]** In yet another aspect, disclosed herein is a plant cell comprising a targeted genomic modification to one or more alleles of an endogenous gene in the plant cell, wherein the genomic modification follows cleavage by a site specific nuclease, and wherein the genomic modification produces a mutation in the endogenous gene such that the endogenous gene produces a product that results in an herbicide tolerant plant cell. In an embodiment, the genomic modification comprises integration of one or more exogenous sequences. In a further embodiment, the genomic modification comprises introduction of one or more indels that mutate the endogenous gene. In an additional embodiment, the endogenous gene with the

genomic modification encodes a protein that confers tolerance to sulfonylurea herbicides. In an embodiment, the endogenous gene with the genomic modification encodes a protein that confers tolerance to imidazolinone herbicides. In a further embodiment, the exogenous sequence does not encode a transgenic selectable marker. In an additional embodiment, the exogenous sequence encodes a protein selected from the group consisting of a protein that increases crop yield, a protein encoding disease resistance, a protein that increases growth, a protein encoding insect resistance, a protein encoding herbicide tolerance, and combinations thereof. In subsequent embodiments, the increased crop yield comprises an increase in fruit yield, grain yield, biomass, fruit flesh content, size, dry weight, solids content, weight, color intensity, color uniformity, altered chemical characteristics, or combinations thereof. In certain embodiments, the endogenous gene is an endogenous acetohydroxyacid synthase (AHAS) gene. In additional embodiments, the two or more exogenous sequences are integrated into the endogenous gene. In a further aspect, the plant cell is a polyploid plant cell. In an embodiment, the site specific nuclease comprises a zinc finger DNA-binding domain, and a FokI cleavage domain. In yet another embodiment, the zinc finger DNA-binding domain encodes a protein that binds to a target site selected from the group consisting of SEQ ID NOs:35-56 and 263-278. In a further embodiment, the plant is selected from the group consisting of wheat, soy, maize, potato, alfalfa, rice, barley, sunflower, tomato, *Arabidopsis*, cotton, *Brassica* species, and timothy grass.

**[0024]** In yet another aspect, disclosed herein is a plant, plant part, seed, or fruit comprising one or more plant cells comprising a targeted genomic modification to one or more alleles of an endogenous gene in the plant cell, wherein the genomic modification follows cleavage by a site specific nuclease, and wherein the genomic modification produces a mutation in the endogenous gene such that the endogenous gene produces a product that results in an herbicide tolerant plant cell.

**[0025]** In yet another aspect, disclosed herein is a method for making a plant cell as disclosed herein above, the method comprising: expressing one or more site specific nucleases in the plant cell; and, modifying one or more alleles of an endogenous gene across multiple genomes of a polyploid plant cell. In an embodiment, the endogenous gene is an acetohydroxyacid synthase (AHAS) gene. In a further embodiment, the modification disrupts expression of the endogenous gene. In yet another embodiment, the modification comprises integration of one or

more exogenous sequences into one or more alleles of the endogenous gene. Furthermore, a plant, plant part, seed, or fruit comprising one or more plant cells produced by the method are disclosed herein as an aspect of the disclosure.

**[0026]** In yet another aspect, disclosed herein is a zinc finger protein that binds to a target site selected from the group consisting of SEQ ID NOs:35-56 and 263-278. In a further embodiment, the zinc finger proteins comprise the recognition helix regions shown in a single row of Table 2 or Table 12.

**[0027]** In yet another aspect, described herein is a method of integrating one or more exogenous sequences into the genome of a plant cell, the method comprising: expressing one or more site specific nucleases in the plant cell, wherein the one or more nucleases target and cleave chromosomal DNA of one or more endogenous loci; integrating one or more exogenous sequences into the one or more endogenous loci within the genome of the plant cell, wherein the one or more endogenous loci are modified such that the endogenous gene is mutated to expresses a product that results in a selectable phenotype in the plant cell; and, selecting plant cells that express the selectable phenotype, wherein plant cells are selected which incorporate the one or more exogenous sequences. In a further embodiment, the one or more exogenous sequences are selected from the group consisting of a donor polynucleotide, a transgene, or any combination thereof. In a subsequent embodiment, the integration of the one or more exogenous sequences occurs by homologous recombination or non-homologous end joining. In an additional embodiment, the one or more exogenous sequences are incorporated simultaneously or sequentially into the one or more endogenous loci. In further embodiments, the one or more endogenous loci comprise an acetohydroxyacid synthase (AHAS) gene. In an embodiment, the AHAS gene is located on an A, B, or D genome of a polyploidy genome. In another embodiment, the one or more exogenous sequences are integrated into the AHAS gene. In yet another embodiment, the one or more exogenous sequences encode a S653N AHAS mutation. In an additional embodiment, the one or more exogenous sequences encode a P197S AHAS mutation. In a subsequent embodiment, the site specific nuclease is selected from the group consisting of a zinc finger nuclease, a TAL effector domain nuclease, a homing endonuclease, and a Crispr/Cas single guide RNA nuclease. In a further embodiment, the site specific nuclease comprises a zinc finger DNA-binding domain, and a FokI cleavage domain. In an embodiment, the one or more exogenous sequences encode a transgene or produce an RNA molecule.

In a subsequent embodiment, the transgene encodes a protein selected from the group consisting of a protein that increases crop yield, a protein encoding disease resistance, a protein that increases growth, a protein encoding insect resistance, a protein encoding herbicide tolerance, and combinations thereof. In further embodiments, the integration of the transgene further comprises introduction of one or more indels that disrupt expression of the one or more endogenous loci and produce the selectable phenotype. Subsequent embodiments of the method further comprise the steps of; culturing the selected plant cells comprising the one or more exogenous sequences; and, obtaining a whole plant comprising the one or more exogenous sequences integrated within the one or more endogenous loci of the plant genome. In an additional embodiment, a selection agent comprising an imidazolinone, or a sulfonylurea selection agent is used to select the plant cells. In other embodiments, the whole plant comprising the one or more exogenous sequences integrated within the one or more endogenous loci of the plant genome is further modified to incorporate an additional exogenous sequence within the endogenous loci of the plant genome. In further embodiments, the one or more exogenous sequences do not encode a transgenic selectable marker.

**[0028]** In a still further aspect, a plant cell obtained according to any of the methods described herein is also provided.

**[0029]** In another aspect, provided herein is a plant comprising a plant cell as described herein.

**[0030]** In another aspect, provided herein is a seed from a plant comprising the plant cell that is obtained as described herein.

**[0031]** In another aspect, provided herein is fruit obtained from a plant comprising plant cell obtained as described herein.

**[0032]** In any of the compositions (cells or plants) or methods described herein, the plant cell can comprise a monocotyledonous or dicotyledonous plant cell. In certain embodiments, the plant cell is a crop plant, for example, wheat, tomato (or other fruit crop), potato, maize, soy, alfalfa, *etc.*

**[0032a]** In one aspect of the present invention, there is provided a plant cell comprising a targeted genomic modification to one or more alleles of an

Date Reçue/Date Received 2021-01-19

endogenous acetohydroxyacid synthase (AHAS) gene of SEQ ID NO: 1. SEQ ID NO: 2 or SEQ ID NO:3 in the plant cell, wherein the genomic modification follows cleavage by a site-specific nuclease, wherein the site-specific nuclease is a zinc finger nuclease comprising a Fok I cleavage domain and a DNA binding domain that binds to a sequence comprising a target site as shown in any one of SEQ ID NOs: 35-56 or 263-278, and wherein the genomic modification produces a mutation in the endogenous AHAS gene such that the endogenous gene produces a product that results in an imidazolinone herbicide tolerant plant cell, wherein the genomic modification comprises (a) introduction of one or more indels that disrupt expression of the endogenous gene, or (b) integration of one or more exogenous sequences, wherein (i) the exogenous sequence does not encode a transgenic selectable marker, or (ii) the exogenous sequence encodes a protein selected from the group consisting of a protein that increases crop yield, a protein encoding disease resistance, a protein that increases growth, a protein encoding insect resistance, a protein encoding herbicide tolerance, and combinations thereof, or (iii) two or more exogenous sequences are integrated into the endogenous gene.

**[0032b]** In another aspect of the present invention, there is provided a method for making a plant cell as described herein, the method comprising: expressing one or more site-specific zinc finger nucleases in the plant such that one or more alleles of the endogenous AHAS gene across multiple genomes of a polyploid plant cell are modified, wherein the modification comprises integration of one or more exogenous sequences into one or more alleles of the endogenous AHAS gene.

**[0032c]** In yet another aspect of the present invention, a zinc finger protein (ZFP) that binds to a target site in an endogenous acetohydroxyacid synthase (AHAS) gene, the zinc finger protein comprising from four to six zinc finger domains ordered F1 to F4, F1 to F5 or F1 to F6, each zinc finger domain comprising a recognition helix region and wherein the zinc finger protein comprises the recognition helix regions ordered as shown in a single row of the following Table:

Date Reçue/Date Received 2021-01-19

| ZFP designation | F1 | F2 | F3 | F4 | F5 | F6 |
|---|---|---|---|---|---|---|
| 29964 | QSSHLTR SEQ ID NO:181 | RSDDLTR SEQ ID NO:182 | RSDDLTR SEQ ID NO:182 | YRWLLRS SEQ ID NO:183 | QSGDLTR SEQ ID NO:184 | QRNARTL SEQ ID NO:185 |
| 29965 | RSDNLSV SEQ ID NO:186 | QKINLQV SEQ ID NO:187 | DDWNLSQ SEQ ID NO:188 | RSANLTR SEQ ID NO:189 | QSGHLAR SEQ ID NO:190 | NDWDRRV SEQ ID NO:191 |
| 29966 | RSDDLTR SEQ ID NO:182 | YRWLLRS SEQ ID NO:183 | QSGDLTR SEQ ID NO:184 | QRNARTL SEQ ID NO:185 | RSDHLSQ SEQ ID NO:192 | DSSTRKK SEQ ID NO:193 |
| 29967 | RSDDLTR SEQ ID NO:182 | YRWLLRS SEQ ID NO:183 | QSGDLTR SEQ ID NO:184 | QRNARTL SEQ ID NO:185 | RSDVLSE SEQ ID NO:194 | DRSNRIK SEQ ID NO:195 |
| 29968 | RSDNLSN SEQ ID NO:196 | TSSSRIN SEQ ID NO:197 | DRSNLTR SEQ ID NO:198 | QSSDLSR SEQ ID NO:199 | QSAHRKN SEQ ID NO:200 | N/A |
| 29969 | DRSHLTR SEQ ID NO:201 | QSGHLSR SEQ ID NO:202 | RSDNLSV SEQ ID NO:186 | QKINLQV SEQ ID NO:187 | DDWNLSQ SEQ ID NO:188 | RSANLTR SEQ ID NO:189 |
| 29970 | QSGDLTR SEQ ID NO:184 | QRNARTL SEQ ID NO:185 | RSDVLSE SEQ ID NO:194 | DRSNRIK SEQ ID NO:195 | RSDNLSE SEQ ID NO:203 | HSNARKT SEQ ID NO:204 |
| 29971 | TSGNLTR SEQ ID NO:205 | HRTSLTD SEQ ID NO:206 | QSSDLSR SEQ ID NO:199 | HKYHLRS SEQ ID NO:207 | QSSDLSR SEQ ID NO:199 | QWSTRKR SEQ ID NO:208 |
| 29730 | DRSHLTR SEQ ID NO:201 | QSGHLSR SEQ ID NO:202 | RSDNLSN SEQ ID NO:196 | TSSSRIN SEQ ID NO:197 | DRSNLTR SEQ ID NO:198 | N/A |
| 29731 | RSDVLSE SEQ ID NO:194 | SPSSRRT SEQ ID NO:209 | RSDTLSE SEQ ID NO:210 | TARQRNR SEQ ID NO:211 | DRSHLAR SEQ ID NO:212 | N/A |
| 29732 | RSDSLSA SEQ ID NO:213 | RSDALAR SEQ ID NO:214 | RSDDLTR SEQ ID NO:182 | QKSNLSS SEQ ID NO:215 | DSSDRKK SEQ ID NO:216 | N/A |
| 30006 | TSGNLTR SEQ ID NO:205 | WWTSRAL SEQ ID NO:217 | DRSDLSR SEQ ID NO:218 | RSDHLSE SEQ ID NO:219 | YSWRLSQ SEQ ID NO:220 | N/A |

Date Reçue/Date Received 2021-01-19

| | | | | | | |
|---|---|---|---|---|---|---|
| 30008 | RSDSLSV SEQ ID NO:221 | RNQDRKN SEQ ID NO:222 | QSSDLSR SEQ ID NO:199 | HKYHLRS SEQ ID NO:207 | QSGDLTR SEQ ID NO:184 | N/A |
| 29753 | QSGNLAR SEQ ID NO:223 | DRSALAR SEQ ID NO:224 | RSDNLST SEQ ID NO:225 | AQWGRTS SEQ ID NO:226 | N/A | N/A |
| 29754 | RSADLTR SEQ ID NO:227 | TNQNRIT SEQ ID NO:228 | RSDSLLR SEQ ID NO:229 | LQHHLTD SEQ ID NO:230 | QNATRIN SEQ ID NO:231 | N/A |
| 29769 | QSGNLAR SEQ ID NO:223 | DRSALAR SEQ ID NO:224 | RSDNLST SEQ ID NO:225 | AQWGRTS SEQ ID NO:226 | N/A | N/A |
| 29770 | QSGDLTR SEQ ID NO:184 | MRNRLNR SEQ ID NO:232 | DRSNLSR SEQ ID NO:233 | WRSCRSA SEQ ID NO:234 | RSDNLSV SEQ ID NO:186 | N/A |
| 30012 | HSNARKT SEQ ID NO:204 | QSGNLAR SEQ ID NO:223 | DRSALAR SEQ ID NO:224 | RSDNLST SEQ ID NO:225 | AQWGRTS SEQ ID NO:226 | N/A |
| 30014 | HSNARKT SEQ ID NO:204 | QSGNLAR SEQ ID NO:223 | DRSALAR SEQ ID NO:224 | RSDHLSQ SEQ ID NO:192 | QWFGRKN SEQ ID NO:235 | N/A |
| 30018 | QSGDLTR SEQ ID NO:184 | MRNRLNR SEQ ID NO:232 | DRSNLSR SEQ ID NO:233 | WRSCRSA SEQ ID NO:234 | QRSNLDS SEQ ID NO:34 | N/A |
| 29988 | QSGDLTR SEQ ID NO:184 | QWGTRYR SEQ ID NO:33 | DRSNLSR SEQ ID NO:233 | HNSSLKD SEQ ID NO:32 | QSGNLAR SEQ ID NO:223 | N/A |
| 29989 | RSDVLSA SEQ ID NO:31 | RNDHRIN SEQ ID NO:30 | RSDHLSQ SEQ ID NO:192 | QSAHRTN SEQ ID NO:29 | DRSNLSR SEQ ID NO:233 | DSTNRYR SEQ ID NO:28 |
| 34456 | RSADLTR SEQ ID NO:227 | RSDDLTR SEQ ID NO:182 | RSDDLTR SEQ ID NO:182 | RSDALTQ SEQ ID NO:236 | ERGTLAR SEQ ID NO:237 | RSDDLTR SEQ ID NO:182 |
| 34457 | QSGDLTR SEQ ID NO:184 | DTGARLK SEQ ID NO:238 | RSDDLTR SEQ ID NO:182 | HRRSRDQ SEQ ID NO:239 | DRSYRNT SEQ ID NO:240 | N/A |

Date Reçue/Date Received 2021-01-19

| | | | | | | |
|---|---|---|---|---|---|---|
| 34470 | RSADLSR SEQ ID NO:241 | RSDHLSA SEQ ID NO:242 | QSSDLRR SEQ ID NO:243 | DRSNLSR SEQ ID NO:233 | RSDDRKT SEQ ID NO:244 | N/A |
| 34471 | QSGDLTR SEQ ID NO:184 | RRADRAK SEQ ID NO:245 | RSDDLTR SEQ ID NO:182 | TSSDRKK SEQ ID NO:246 | RSADLTR SEQ ID NO:227 | RNDDRKK SEQ ID NO:247 |
| 34472 | RSADLTR SEQ ID NO:227 | DRSNLTR SEQ ID NO:198 | ERGTLAR SEQ ID NO:237 | RSDDLTR SEQ ID NO:182 | DRSDLSR SEQ ID NO:218 | DSSTRRR SEQ ID NO:248 |
| 34473 | RSDHLSE SEQ ID NO:219 | HSRTRTK SEQ ID NO:249 | RSDTLSE SEQ ID NO:210 | NNRDRTK SEQ ID NO:250 | ERGTLAR SEQ ID NO:237 | DRSALAR SEQ ID NO:224 |
| 34474 | ERGTLAR SEQ ID NO:237 | RSDDLTR SEQ ID NO:182 | DRSDLSR SEQ ID NO:218 | DSSTRRR SEQ ID NO:248 | DRSNLTR SEQ ID NO:198 | N/A |
| 34475 | RSDHLSR SEQ ID NO:249 | QQWDRKQ SEQ ID NO:73 | DRSHLTR SEQ ID NO:201 | DSSDRKK SEQ ID NO:216 | DRSNLSR SEQ ID NO:233 | VSSNLTS SEQ ID NO:251 |
| 34476 | DRSDLSR SEQ ID NO:218 | DSSTRRR SEQ ID NO:248 | DRSNLSR SEQ ID NO:233 | QSGDLTR SEQ ID NO:184 | DRSNLTR SEQ ID NO:198 | N/A |
| 34477 | ERGTLAR SEQ ID NO:237 | RSDHLSR SEQ ID NO:249 | RSDALSV SEQ ID NO:252 | DSSHRTR SEQ ID NO:253 | DSSDRKK SEQ ID NO:216 | N/A |
| 34478 | RSDNLTR SEQ ID NO:254 | RSDNLAR SEQ ID NO:255 | DRSALAR SEQ ID NO:224 | DRSHLSR SEQ ID NO:256 | TSGNLTR SEQ ID NO:205 | N/A |
| 34479 | RSDALSV SEQ ID NO:252 | DSSHRTR SEQ ID NO:253 | RSDNLSE SEQ ID NO:203 | ARTGLRQ SEQ ID NO:254 | ERGTLAR SEQ ID NO:237 | DRSALAR SEQ ID NO:224 |
| 34480 | RSDNLAR SEQ ID NO:255 | DRSALAR SEQ ID NO:224 | DRSHLSR SEQ ID NO:256 | TSGNLTR SEQ ID NO:205 | RSDHLSR SEQ ID NO:249 | TSSNRKT SEQ ID NO:257 |
| 34481 | DRSALAR SEQ ID NO:224 | RSDALSV SEQ ID NO:252 | DSSHRTR SEQ ID NO:253 | RSDNLSE SEQ ID NO:203 | ARTGLRQ SEQ ID NO:254 | N/A |
| 34482 | RSDDLSK SEQ ID NO:258 | RSDNLTR SEQ ID NO:254 | RSDSLSV SEQ ID NO:221 | RSAHLSR SEQ ID NO:259 | RSDALST SEQ ID NO:260 | DRSTRTK SEQ ID NO:261 |
| 34483 | DSSDRKK SEQ ID NO:216 | RSAHLSR SEQ ID NO:259 | DRSDLSR SEQ ID NO:218 | RSDHLSE SEQ ID NO:219 | TSSDRTK SEQ ID NO:262 | N/A |

Date Reçue/Date Received 2021-01-19

**[0032d]** In still another aspect of the invention, a zinc finger nuclease comprising a pair of zinc finger proteins as described herein, wherein the pair comprises the following ZFPs: (i) a ZFP designated 29964 and a ZFP designated 29965; (ii) a ZFP designated 29966 and a ZFP designated 29968; (iii) a ZFP designated 29967 and a ZFP designated 29968; (iv) a ZFP designated 29967 and a ZFP designated 29969; (v) a ZFP designated 29970 and a ZFP designated 29971; (vi) a ZFP designated 29730 and a ZFP designated 29732; (vii) a ZFP designated 29731 and a ZFP designated 29732; (viii) a ZFP designated 30006 and a ZFP designated 30008; (ix) a ZFP designated 29753 and a ZFP designated 29754; (x) a ZFP designated 29769 and a ZFP designated 29770; (xi) a ZFP designated 30012 and a ZFP designated 30018; (xii) a ZFP designated 30014 and a ZFP designated 30018; (xiii) a ZFP designated 29988 and a ZFP designated 29989; (xiv) a ZFP designated 34456 and a ZFP designated 34457; (xv) a ZFP designated 34470 and a ZFP designated 34471; (xvi) a ZFP designated 34472 and a ZFP designated 34473; (xvii) a ZFP designated 34474 and a ZFP designated 34475; (xviii) a ZFP designated 34476 and a ZFP designated 34477; (xix) a ZFP designated 34478 and a ZFP designated 34479; (xx) a ZFP designated 34480 and a ZFP designated 34481; and (xxi) a ZFP designated 34482 and a ZFP designated 34483.

**[0032e]** In a further aspect of the invention, a method of integrating one or more exogenous sequences into the genome of a plant cell, the method comprising: a) expressing one or more site-specific nucleases as described herein in the plant cell; b) integrating one or more exogenous sequences into the endogenous AHAS loci within the genome of the plant cell, wherein the AHAS loci is modified such that the endogenous AHAS gene is mutated to express a product that results in a selectable phenotype in the plant cell; and c) selecting plant cells that express the selectable phenotype, wherein plant cells are selected which incorporate the one or more exogenous sequences, wherein preferably the one or more exogenous sequences are selected from the group consisting of a donor polynucleotide, a transgene, or any combination thereof.

## BRIEF DESCRIPTION OF THE DRAWINGS

**[0033] Figure 1** is a plasmid map of pDAB109350.

**[0034] Figure 2** is a plasmid map of pDAB100360.

**[0035] Figure 3** is a plasmid map of pDAS000132.

[0036] **Figure 4** is a plasmid map of pDAS000133.

[0037] **Figure 5** is a plasmid map of pDAS000134.

[0038] **Figure 6** is a plasmid map of pDAS000135.

[0039] **Figure 7** is a plasmid map of pDAS000131.

[0040] **Figure 8** is a plasmid map of pDAS000153.

[0041] **Figure 9** is a plasmid map of pDAS000150.

[0042] **Figure 10** is a plasmid map of pDAS000143.

[0043] **Figure 11** is a plasmid map of pDAS000164.

[0044] **Figure 12** is a plasmid map of pDAS000433.

[0045] **Figure 13** is a plasmid map of pDAS000434.

[0046] **Figure 14, panels A and B,** are schematics depicting 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 14A depicts a first transgene stack; Figure 14B depicts a second transgene stack.

[0047] **Figure 15, panels A and B,** are schematics depicting 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 15A depicts a first transgene stack; Figure 15B depicts a second transgene stack.

[0048] **Figure 16** is a schematic showing a linear map of pDAS000435.

[0049] **Figure 17** is a schematic showing a linear map of pDAS000436.

[0050] **Figure 18** is a plasmid map of pDAS0000004.

[0051] **Figure 19** is a plasmid map of QA_pDAS000434.

## DETAILED DESCRIPTION

[0052] The present disclosure relates to methods and compositions for exogenous sequence integration, including parallel (simultaneous) or sequential exogenous sequence integration (including transgene stacking) in a plant species, including in a polyploid plant. The methods and compositions described herein are advantageous in providing targeted integration into a selected locus without the use of an exogenous transgenic marker to assess integration. In particular, differential selection at an endogenous locus, with a transgenic marker-free donor design, has been demonstrated to bias selection for targeted transgenic events by reducing the number of illegitimate integrated events recovered (Shukla et al. (2009) *Nature*

459(7245):437-41). In addition, the disclosure relates to genomic modification (*e.g.,* mutation) of an endogenous locus, which mutation can result in production of a product that serves as a marker (phenotype). Thus, the present disclosure provides for exogenous sequence integration, including transgene stacking, into an endogenous locus, which endogenous locus can serve as a marker for integration (*e.g.,* the AHAS locus in which single mutations can impart herbicide tolerance).

**[0053]** Integration of the exogenous sequence(s) (*e.g.,* into the AHAS locus) is facilitated by targeted double-strand cleavage of endogenous sequence, for example by cleavage of a sequence located in the 3' untranslated region. Cleavage is targeted to this region through the use of fusion proteins comprising a DNA-binding domain, such as a meganuclease DNA-binding domain, a leucine zipper DNA-binding domain, a TAL DNA-binding domain, a zinc finger protein (ZFP); or through the use of a Crispr/Cas RNA or chimeric combinations of the aforementioned. Such cleavage stimulates integration of the donor nucleic acid sequence(s) at, or near the endogenous cleavage site. Integration of exogenous sequences can proceed through both homology-dependent and homology-independent mechanisms, and the selection of precisely targeted events is achieved through screening for a selectable marker (*e.g.,* tolerance to a specific Group B herbicide, or ALS inhibitor herbicides such as imidazolinone or sulfonylurea) which is only functional in correctly targeted events.

**[0054]** In certain embodiments, the nuclease(s) comprise one or more ZFNs, one or more TALENs, one or more meganucleases and/or one or more CRISPR/Cas nuclease systems. ZFNs and TALENs typically comprise a cleavage domain (or a cleavage half-domain) and a zinc finger DNA binding or TALE-effector DNA binding domain and may be introduced as proteins, as polynucleotides encoding these proteins or as combinations of polypeptides and polypeptide-encoding polynucleotides. ZFNs and TALENs can function as dimeric proteins following dimerization of the cleavage half-domains. Obligate heterodimeric nucleases, in which the nuclease monomers bind to the "left" and "right" recognition domains can associate to form an active nuclease have been described. *See, e.g.,* U.S. Patent Nos. 8,623,618; 7,914,796; 8034,598. Thus, given the appropriate target sites, a "left" monomer could form an active nuclease with any "right" monomer. This significantly increases the number of useful nuclease sites based on proven left and right domains that can be used in various combinations. For example, recombining the binding sites of four homodimeric nucleases yields an additional twelve

heterodimeric nucleases. More importantly, it enables a systematic approach to transgenic design such that every new introduced sequence becomes flanked with a unique nuclease binding site that can be used to excise the gene back out or to target additional genes next to it. Additionally, this method can simplify strategies of stacking into a single locus that is driven by nuclease-dependent double-strand breaks.

**[0055]** A zinc finger binding domain can be a canonical (C2H2) zinc finger or a non-canonical (*e.g.*, C3H) zinc finger. *See, e.g.*, U.S. Patent Publication No. 20080182332. Furthermore, the zinc finger binding domain can comprise one or more zinc fingers (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9 or more zinc fingers), and can be engineered to bind to any sequence within any endogenous gene, for example an AHAS gene. The presence of such a fusion protein (or proteins) in a cell results in binding of the fusion protein(s) to its (their) binding site(s) and cleavage within the target gene(s).

**General**

**[0056]** 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, 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, “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**

**[0057]** 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.

**[0058]** 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 corresponding naturally-occurring amino acids.

**[0059]** "Binding" refers to a sequence-specific, non-covalent interaction between macromolecules (*e.g.*, between a protein and a nucleic acid). Not all 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 $K_d$.

**[0060]** A "binding protein" is a protein that is able to bind 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.

**[0061]** 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. The term zinc finger DNA binding protein is often abbreviated as zinc finger protein or ZFP.

**[0062]** 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 “repeat”) is typically 33-35 amino acids and includes hypervariable diresidues at positions 12 and/or 13 referred to as the Repeat Variable Diresidue (RVD) involved in DNA-

binding specificity. TALE repeats exhibit at least some sequence homology with other TALE repeat sequences within a naturally occurring TALE protein. *See, e.g.,* U.S. Patent No. 8,586,526.

**[0063]** Zinc finger binding and TALE domains can be "engineered" to bind to a predetermined nucleotide sequence. Non-limiting examples of methods for engineering zinc finger proteins are design and selection. A designed zinc finger 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 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; WO 02/016536 and WO 03/016496.

**[0064]** 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. 8,586,526, U.S. 5,789,538; U.S. 5,925,523; U.S. 6,007,988; U.S. 6,013,453; U.S. 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.

**[0065]** 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" refers to a nucleotide sequence that is inserted into a genome. A donor sequence can be of any length, for example between 2 and 10,000 nucleotides in length (or any integer value therebetween or thereabove), preferably between about 100 and 1,000 nucleotides in length (or any integer therebetween), more preferably between about 200 and 500 nucleotides in length.

**[0066]** A "homologous, non-identical sequence" refers to a first sequence which shares a degree of sequence identity with a second sequence, but whose sequence is not identical to that of the second sequence. For example, a polynucleotide comprising the wild-type sequence of a mutant gene is homologous and non-identical to the sequence of the mutant gene. In certain embodiments, the degree of homology between the two sequences is sufficient to allow homologous recombination therebetween, utilizing normal cellular mechanisms. Two homologous non-identical sequences can be any length and their degree of non-homology can be

as small as a single nucleotide (*e.g.*, for correction of a genomic point mutation by targeted homologous recombination) or as large as 10 or more kilobases (*e.g.*, for insertion of a gene at a predetermined ectopic site in a chromosome). Two polynucleotides comprising the homologous non-identical sequences need not be the same length. For example, an exogenous polynucleotide (*i.e.*, donor polynucleotide) of between 20 and 10,000 nucleotides or nucleotide pairs can be used.

**[0067]** Techniques for determining nucleic acid and amino acid sequence identity are known in the art. Typically, such techniques include determining the nucleotide sequence of the mRNA for a gene and/or determining the amino acid sequence encoded thereby, and comparing these sequences to a second nucleotide or amino acid sequence. Genomic sequences can also be determined and compared in this fashion. In general, identity refers to an exact nucleotide-to-nucleotide or amino acid-to-amino acid correspondence of two polynucleotides or polypeptide sequences, respectively. Two or more sequences (polynucleotide or amino acid) can be compared by determining their percent identity. The percent identity of two sequences, whether nucleic acid or amino acid sequences, is the number of exact matches between two aligned sequences divided by the length of the shorter sequences and multiplied by 100. An approximate alignment for nucleic acid sequences is provided by the local homology algorithm of Smith and Waterman, Advances in Applied Mathematics 2:482-489 (1981). This algorithm can be applied to amino acid sequences by using the scoring matrix developed by Dayhoff, Atlas of Protein Sequences and Structure, M.O. Dayhoff ed., 5 suppl. 3:353-358, National Biomedical Research Foundation, Washington, D.C., USA, and normalized by Gribskov, Nucl. Acids Res. 14(6):6745-6763 (1986). An exemplary implementation of this algorithm to determine percent identity of a sequence is provided by the Genetics Computer Group (Madison, WI) in the "BestFit" utility application. Suitable programs for calculating the percent identity or similarity between sequences are generally known in the art, for example, another alignment program is BLAST, used with default parameters. For example, BLASTN and BLASTP can be used using the following default parameters: genetic code = standard; filter = none; strand = both; cutoff = 60; expect = 10; Matrix = BLOSUM62 (for BLASTP); Descriptions = 50 sequences; sort by = HIGH SCORE; Databases = non-redundant, GenBank + EMBL + DDBJ + PDB + GenBank CDS translations + Swiss protein + Spupdate + PIR. Details of these programs can be found on the internet. With respect to

sequences described herein, the range of desired degrees of sequence identity is approximately 80% to 100% and any integer value therebetween. Typically the percent identities between sequences are at least 70-75%, preferably 80-82%, more preferably 85-90%, even more preferably 92%, still more preferably 95%, and most preferably 98% sequence identity.

**[0068]** Alternatively, the degree of sequence similarity between polynucleotides can be determined by hybridization of polynucleotides under conditions that allow formation of stable duplexes between homologous regions, followed by digestion with single-stranded-specific nuclease(s), and size determination of the digested fragments. Two nucleic acid, or two polypeptide sequences are substantially homologous to each other when the sequences exhibit at least about 70%-75%, preferably 80%-82%, more preferably 85%-90%, even more preferably 92%, still more preferably 95%, and most preferably 98% sequence identity over a defined length of the molecules, as determined using the methods above. As used herein, substantially homologous also refers to sequences showing complete identity to a specified DNA or polypeptide sequence. DNA sequences that are substantially homologous can be identified in a Southern hybridization experiment under, for example, stringent conditions, as defined for that particular system. Defining appropriate hybridization conditions is known to those with skill of the art. See, *e.g.*, Sambrook et al., *supra*; Nucleic Acid Hybridization: A Practical Approach, editors B.D. Hames and S.J. Higgins, (1985) Oxford; Washington, DC; IRL Press).

**[0069]** Selective hybridization of two nucleic acid fragments can be determined as follows. The degree of sequence identity between two nucleic acid molecules affects the efficiency and strength of hybridization events between such molecules. A partially identical nucleic acid sequence will at least partially inhibit the hybridization of a completely identical sequence to a target molecule. Inhibition of hybridization of the completely identical sequence can be assessed using hybridization assays that are well known in the art (*e.g.*, Southern (DNA) blot, Northern (RNA) blot, solution hybridization, or the like, see Sambrook, *et al.*, *Molecular Cloning: A Laboratory Manual*, Second Edition, (1989) Cold Spring Harbor, N.Y.). Such assays can be conducted using varying degrees of selectivity, for example, using conditions varying from low to high stringency. If conditions of low stringency are employed, the absence of non-specific binding can be assessed using a secondary probe that lacks even a partial degree of sequence identity (for example, a

probe having less than about 30% sequence identity with the target molecule), such that, in the absence of non-specific binding events, the secondary probe will not hybridize to the target.

**[0070]** When utilizing a hybridization-based detection system, a nucleic acid probe is chosen that is complementary to a reference nucleic acid sequence, and then by selection of appropriate conditions the probe and the reference sequence selectively hybridize, or bind, to each other to form a duplex molecule. A nucleic acid molecule that is capable of hybridizing selectively to a reference sequence under moderately stringent hybridization conditions typically hybridizes under conditions that allow detection of a target nucleic acid sequence of at least about 10-14 nucleotides in length having at least approximately 70% sequence identity with the sequence of the selected nucleic acid probe. Stringent hybridization conditions typically allow detection of target nucleic acid sequences of at least about 10-14 nucleotides in length having a sequence identity of greater than about 90-95% with the sequence of the selected nucleic acid probe. Hybridization conditions useful for probe/reference sequence hybridization, where the probe and reference sequence have a specific degree of sequence identity, can be determined as is known in the art (see, for example, Nucleic Acid Hybridization: A Practical Approach, editors B.D. Hames and S.J. Higgins, (1985) Oxford; Washington, DC; IRL Press).

**[0071]** Conditions for hybridization are well-known to those of skill in the art. Hybridization stringency refers to the degree to which hybridization conditions disfavor the formation of hybrids containing mismatched nucleotides, with higher stringency correlated with a lower tolerance for mismatched hybrids. Factors that affect the stringency of hybridization are well-known to those of skill in the art and include, but are not limited to, temperature, pH, ionic strength, and concentration of organic solvents such as, for example, formamide and dimethylsulfoxide. As is known to those of skill in the art, hybridization stringency is increased by higher temperatures, lower ionic strength and lower solvent concentrations.

**[0072]** With respect to stringency conditions for hybridization, it is well known in the art that numerous equivalent conditions can be employed to establish a particular stringency by varying, for example, the following factors: the length and nature of the probe sequences, base composition of the various sequences, concentrations of salts and other hybridization solution components, the presence or absence of blocking agents in the hybridization solutions (*e.g.*, dextran sulfate, and

polyethylene glycol), hybridization reaction temperature and time parameters, as well as, varying wash conditions. The selection of a particular set of hybridization conditions is selected following standard methods in the art (see, for example, Sambrook, et al., Molecular Cloning: A Laboratory Manual, Second Edition, (1989) Cold Spring Harbor, N.Y.).

**[0073]** "Recombination" refers to a process of exchange of genetic information between two polynucleotides. 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. This process requires nucleotide sequence homology, that 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 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 incorporated into the target polynucleotide.

**[0074]** "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 cleavage.

**[0075]** A "cleavage domain" comprises one or more polypeptide sequences which possesses catalytic activity for DNA cleavage. A cleavage domain can be contained in a single polypeptide chain or cleavage activity can result from the association of two (or more) polypeptides.

**[0076]** 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).

**[0077]** 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 Nos. 7,914,796; 8,034,598; 8,623,618 and U.S. Patent Publication No. 2011/0201055.

**[0078]** "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 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.

**[0079]** 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.

**[0080]** 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.

**[0081]** An "accessible region" is a site in cellular chromatin in which a target site present in the nucleic acid can be bound by an exogenous molecule which recognizes the target site. Without wishing to be bound by any particular theory, it is believed that an accessible region is one that is not packaged into a nucleosomal structure. The distinct structure of an accessible region can often be detected by its sensitivity to chemical and enzymatic probes, for example, nucleases.

**[0082]** 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. For example, the sequence 5'-GAATTC-3' is a target site for the Eco RI restriction endonuclease. In addition Table 3 and 13 list the target sites for the binding of the ZFP recognition helices of Table 2 and Table 12.

**[0083]** 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 in cells only during the early stages of development of a flower is an exogenous molecule with respect to the cells of a fully developed flower. 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 coding sequence for any polypeptide or fragment thereof, a functioning version of a malfunctioning endogenous molecule or a malfunctioning version of a normally-functioning endogenous molecule. Additionally, an exogenous molecule can comprise a coding sequence from another species that is an ortholog of an endogenous gene in the host cell.

**[0084]** 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. Thus, the term includes "transgenes" or "genes of interest" which are exogenous sequences introduced into a plant cell.

**[0085]** 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 skill in the art and include, but are not limited to, protoplast transformation, silicon carbide (*e.g.*, WHISKERS™), *Agrobacterium*-mediated transformation, lipid-mediated transfer (*i.e.*, liposomes, including neutral and cationic lipids), electroporation, direct injection, cell fusion, particle bombardment (*e.g.*, using a "gene gun"), calcium phosphate co-precipitation, DEAE-dextran-mediated transfer and viral vector-mediated transfer.

**[0086]** 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-occurring episomal nucleic acid. Additional endogenous molecules can include proteins, for example, transcription factors and enzymes.

**[0087]** As used herein, the term "product of an exogenous nucleic acid" includes both polynucleotide and polypeptide products, for example, transcription products (polynucleotides such as RNA) and translation products (polypeptides).

**[0088]** A "fusion" molecule is a molecule in which two or more subunit molecules are linked, preferably covalently. The subunit molecules can be the same 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 DNA-binding domain (*e.g.*, ZFP, TALE and/or meganuclease DNA-binding domains) and a nuclease (cleavage) domain (*e.g.*, endonuclease, meganuclease, *etc.* and fusion nucleic acids (for example, a nucleic acid encoding the fusion protein described herein). Examples of the second type of fusion 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.

**[0089]** 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 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.

**[0090]** 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 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 and locus control regions.

**[0091]** "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 translation of a 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, myristoylation, and glycosylation.

**[0092]** "Modulation" of gene expression refers to a change in the activity of a gene. Modulation of expression can include, but is not limited to, gene activation and gene repression.

**[0093]** A "transgenic selectable marker" refers to an exogenous sequence comprising a marker gene operably linked to a promoter and 3'-UTR to comprise a chimeric gene expression cassette. Non-limiting examples of transgenic selectable markers include herbicide tolerance, antibiotic resistance, and visual reporter markers. The transgenic selectable marker can be integrated along with a donor sequence via targeted integration. As such, the transgenic selectable marker expresses a product that is used to assess integration of the donor. In contrast, the methods and compositions described herein allow for integration of any donor sequence without the need for co-integration of a transgenic selectable marker, for example by using a donor which mutates the endogenous gene into which it is integrated to produce a selectable marker (*i.e.*, the selectable marker as used in this instance is not transgenic) from the endogenous target locus. Non-limiting examples of selectable markers

include herbicide tolerance markers, including a mutated AHAS gene as described herein.

**[0094]** "Plant" cells include, but are not limited to, cells of monocotyledonous (monocots) or dicotyledonous (dicots) plants. Non-limiting examples of monocots include cereal plants such as maize, rice, barley, oats, wheat, sorghum, rye, sugarcane, pineapple, onion, banana, and coconut. Non-limiting examples of dicots include tobacco, tomato, sunflower, cotton, sugarbeet, potato, lettuce, melon, soy, canola (rapeseed), and alfalfa. Plant cells may be from any part of the plant and/or from any stage of plant development.

**[0095]** A "region of interest" is any region of cellular chromatin, such as, for 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.

**[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 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 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.

**[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 DNA-binding domain (ZFP, TALE) is fused to a cleavage domain (*e.g.,* endonuclease domain such as FokI, meganuclease domain, *etc.*), the DNA-binding domain and the cleavage domain are in operative linkage if, in the fusion polypeptide, the DNA-binding domain portion is able to bind its target site and/or its binding site, while the cleavage (nuclease) domain is able to cleave DNA in the vicinity of the target site. The nuclease domain may also exhibit DNA-binding capability (*e.g.*, a nuclease fused to a ZFP or TALE domain that also can bind to DNA). Similarly, with respect to a fusion polypeptide in which a DNA-binding domain is fused to an activation or repression domain, the DNA-binding domain and the activation or repression domain are in operative linkage if, in the fusion polypeptide, the DNA-binding domain portion is able to bind its target site and/or its binding site, while the activation domain is able to upregulate gene expression or the repression domain is able to downregulate gene expression. 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.

**DNA-binding domains**

**[0098]** Any DNA-binding domain can be used in the methods disclosed herein. In certain embodiments, the DNA binding domain comprises a zinc finger protein. A zinc finger binding domain comprises one or more zinc fingers. Miller *et al.* (1985) *EMBO J.* **4:**1609-1614; Rhodes (1993) *Scientific American* Feb.:56-65; U.S. Patent No. 6,453,242. The zinc finger binding domains described herein generally include 2, 3, 4, 5, 6 or even more zinc fingers.

**[0099]** Typically, a single zinc finger domain is about 30 amino acids in length. Structural studies have demonstrated that each zinc finger domain (motif) contains two beta sheets (held in a beta turn which contains the two invariant cysteine residues) and an alpha helix (containing the two invariant histidine residues), which are held in a particular conformation through coordination of a zinc atom by the two cysteines and the two histidines.

**[0100]** Zinc fingers include both canonical $C_2H_2$ zinc fingers (*i.e.*, those in which the zinc ion is coordinated by two cysteine and two histidine residues) and non-canonical zinc fingers such as, for example, $C_3H$ zinc fingers (those in which the zinc ion is coordinated by three cysteine residues and one histidine residue) and $C_4$ zinc fingers (those in which the zinc ion is coordinated by four cysteine residues). See also WO 02/057293 and also U.S. Patent Publication No. 20080182332 regarding non-canonical ZFPs for use in plants.

**[0101]** 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 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.

**[0102]** Exemplary selection methods, including phage display and two-hybrid systems, are disclosed in U.S. 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.

**[0103]** Enhancement of binding specificity for zinc finger binding domains has been described, for example, in U.S. Patent No. 6,794,136.

**[0104]** Since an individual zinc finger binds to a three-nucleotide (*i.e.*, triplet) sequence (or a four-nucleotide sequence which can overlap, by one nucleotide, with the four-nucleotide binding site of an adjacent zinc finger), the length of a sequence to which a zinc finger binding domain is engineered to bind (*e.g.*, a target sequence) will determine the number of zinc fingers in an engineered zinc finger binding domain. For example, for ZFPs in which the finger motifs do not bind to overlapping subsites, a six-nucleotide target sequence is bound by a two-finger binding domain; a nine-nucleotide target sequence is bound by a three-finger binding domain, *etc*. As noted herein, binding sites for individual zinc fingers (*i.e.*, subsites) in a target site need not be contiguous, but can be separated by one or several nucleotides, depending on the length and nature of the amino acids sequences between the zinc fingers (*i.e.*, the inter-finger linkers) in a multi-finger binding domain.

**[0105]** In a multi-finger zinc finger binding domain, adjacent zinc fingers can be separated by amino acid linker sequences of approximately 5 amino acids (so-called "canonical" inter-finger linkers) or, alternatively, by one or more non-canonical linkers. See, *e.g.*, U.S. Patent Nos. 6,453,242 and 6,534,261. For engineered zinc finger binding domains comprising more than three fingers, insertion of longer ("non-canonical") inter-finger linkers between certain of the zinc fingers may be desirable in some instances as it may increase the affinity and/or specificity of binding by the binding domain. See, for example, U.S. Patent No. 6,479,626 and WO 01/53480. Accordingly, multi-finger zinc finger binding domains can also be characterized with respect to the presence and location of non-canonical inter-finger linkers. For example, a six-finger zinc finger binding domain comprising three fingers (joined by two canonical inter-finger linkers), a long linker and three additional fingers (joined by two canonical inter-finger linkers) is denoted a 2x3 configuration. Similarly, a binding domain comprising two fingers (with a canonical linker therebetween), a long linker and two additional fingers (joined by a canonical linker) is denoted a 2x2 configuration. A protein comprising three two-finger units (in each of which the two fingers are joined by a canonical linker), and in which each two-finger unit is joined to the adjacent two finger unit by a long linker, is referred to as a 3x2 configuration.

**[0106]** The presence of a long or non-canonical inter-finger linker between two adjacent zinc fingers in a multi-finger binding domain often allows the two fingers to bind to subsites which are not immediately contiguous in the target sequence. Accordingly, there can be gaps of one or more nucleotides between

subsites in a target site; *i.e.*, a target site can contain one or more nucleotides that are not contacted by a zinc finger. For example, a 2x2 zinc finger binding domain can bind to two six-nucleotide sequences separated by one nucleotide, *i.e.*, it binds to a 13-nucleotide target site. *See* also Moore *et al.* (2001a) Proc. Natl. Acad. Sci. USA **98:**1432-1436; Moore *et al.* (2001b) Proc. Natl. Acad. Sci. USA **98:**1437-1441 and WO 01/53480.

**[0107]** As discussed previously, a target subsite is a three- or four-nucleotide sequence that is bound by a single zinc finger. For certain purposes, a two-finger unit is denoted a "binding module." A binding module can be obtained by, for example, selecting for two adjacent fingers in the context of a multi-finger protein (generally three fingers) which bind a particular six-nucleotide target sequence. Alternatively, modules can be constructed by assembly of individual zinc fingers. See also WO 98/53057 and WO 01/53480.

**[0108]** Alternatively, the DNA-binding domain may be derived from a nuclease. For example, the recognition sequences of homing endonucleases and meganucleases such as I-*Sce*I, I-*Ceu*I, PI-*Psp*I, PI-*Sce*, I-*Sce*IV, I-*Csm*I, I-*Pan*I, I-*Sce*II, I-*Ppo*I, I-*Sce*III, I-*Cre*I, I-*Tev*I, I-*Tev*II and I-*Tev*III 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; Jasin (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 DNA-binding domain (*e.g.,* zinc finger protein or TALE) or to a heterologous cleavage domain. DNA-binding domains derived from meganucleases may also exhibit DNA-binding activity.

**[0109]** In other embodiments, the DNA-binding domain comprises a naturally occurring or engineered (non-naturally occurring) TAL effector DNA binding

domain. *See, e.g.,* U.S. Patent No. 8,586,526. 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 effectors (TALE) 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 TALEs is AvrBs3 from *Xanthomonas campestgris* pv. *Vesicatoria* (see Bonas *et al* (1989) *Mol Gen Genet* 218: 127-136 and WO2010079430). TALEs 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, 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 biovar 4 strain RS1000 (See Heuer *et al* (2007) *Appland 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*.

**[0110]** Thus, in some embodiments, the DNA binding domain that binds to a target site in a target locus is an engineered domain from a TAL effector similar to those derived from the plant pathogens Xanthomonas (see Boch *et al*, (2009) *Science* 326: 1509-1512 and Moscou and Bogdanove, (2009) *Science*326: 1501) and Ralstonia (see Heuer *et al* (2007) *Applied and Environmental Microbiology* 73(13): 4379-4384); U.S. Patent Nos. 8,586,526; 8,420,782 and 8,440,431. TALENs may include C-cap and/or N-cap sequences (*e.g.,* C-terminal and/or N-terminal truncations of the TALE backbone (*e.g.,* "+17", "+63" C-caps). *See, e.g.,* U.S. Patent No. 8,586,526.

**[0111]** As another alternative, the DNA-binding domain may be derived from a leucine zipper protein. Leucine zippers are a class of proteins that are involved in protein-protein interactions in many eukaryotic regulatory proteins that are important transcriptional factors associated with gene expression. The leucine zipper refers to a

common structural motif shared in these transcriptional factors across several kingdoms including animals, plants, yeasts, *etc.* The leucine zipper is formed by two polypeptides (homodimer or heterodimer) that bind to specific DNA sequences in a manner where the leucine residues are evenly spaced through an α-helix, such that the leucine residues of the two polypeptides end up on the same face of the helix. The DNA binding specificity of leucine zippers can be utilized in the DNA-binding domains disclosed herein.

**Cleavage Domains**

**[0112]** As noted above, any DNA-binding domain may be associated with a cleavage (nuclease) domain. For example, homing endonucleases may be modified in their DNA-binding specificity while retaining nuclease function. In addition, zinc finger proteins may also be fused to a nuclease (cleavage) domain to form a zinc finger nuclease (ZFN). TALE proteins may be linked to a nuclease (cleavage) domain to form a TALEN.

**[0113]** The cleavage domain portion of the fusion proteins disclosed herein can be obtained from any endonuclease or exonuclease. Exemplary endonucleases from which a cleavage domain can be 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 al.* (eds.) *Nucleases*, Cold Spring Harbor Laboratory Press,1993). Non limiting examples of homing endonucleases and meganucleases include I-*Sce*I, I-*Ceu*I, PI-*Psp*I, PI-*Sce*, I-*Sce*IV, I-*Csm*I, I-*Pan*I, I-*Sce*II, I-*Ppo*I, I-*Sce*III, I-*Cre*I, I-*Tev*I, I-*Tev*II and I-*Tev*III 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; Jasin (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. One or more of these enzymes (or functional fragments thereof) can be used as a source of cleavage domains and cleavage half-domains.

**[0114]** 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, U.S. 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.

**[0115]** 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 FokI 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.

**[0116]** 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.

**[0117]** Exemplary Type IIS restriction enzymes are described in U.S. Publication No. 20070134796.

**[0118]** To enhance cleavage specificity, cleavage domains may also be modified. In certain embodiments, variants of the cleavage half-domain are employed these variants minimize or prevent homodimerization of the cleavage half-domains.

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. Non-limiting examples of such modified cleavage half-domains are described in detail in U.S. Patent Nos. 7,888,121; 7,914,796 and 8,034,598. See, also, Examples.

**[0119]** Additional engineered cleavage half-domains of *Fok*I that form obligate heterodimers can also be used in the ZFNs described herein. 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.

**[0120]** 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 Nos. 7,914,796 and 8,034,598. 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 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 U.S. Patent 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).

**[0121]** Engineered cleavage half-domains described herein can be prepared using any suitable method, for example, by site-directed mutagenesis of wild-type cleavage half-domains (*Fok* I) as described in U.S. Patent Nos. 7,914,796; 8,034,598 and 8,623,618; and U.S. Patent Publication No. 20110201055.

**[0122]** In other embodiments, the nuclease comprises an engineered TALE DNA-binding domain and a nuclease domain (*e.g.,* endonuclease and/or meganuclease domain), also referred to as TALENs. Methods and compositions for engineering these TALEN proteins for robust, site specific interaction with the target sequence of the user's choosing have been published (see U.S. Patent No. 8,586,526). In some embodiments, the TALEN comprises a endonuclease (*e.g.,* FokI) cleavage domain or cleavage half-domain. In other embodiments, the TALE-nuclease is a mega TAL. These mega TAL nucleases are fusion proteins comprising a TALE DNA binding domain and a meganuclease cleavage domain. The meganuclease cleavage domain is active as a monomer and does not require dimerization for activity. (See Boissel *et al.*, (2013) *Nucl Acid Res*: 1-13, doi: 10.1093/nar/gkt1224). In addition, the nuclease domain may also exhibit DNA-binding functionality.

**[0123]** In still further embodiments, the nuclease comprises a compact TALEN (cTALEN). These are single chain fusion proteins linking a TALE DNA binding domain to a TevI nuclease domain. The fusion protein can act as either a nickase localized by the TALE region, or can create a double strand break, depending upon where the TALE DNA binding domain is located with respect to the TevI nuclease domain (see Beurdeley *et al* (2013) *Nat Comm*: 1-8 DOI: 10.1038/ncomms2782). Any TALENs may be used in combination with additional TALENs (*e.g.,* one or more TALENs (cTALENs or FokI-TALENs) with one or more mega-TALs).

**[0124]** Nucleases may be assembled using standard techniques, including *in vivo* at the nucleic acid 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.

**[0125]** Nucleases can be screened for activity prior to use, for example in a yeast-based chromosomal system as described in U.S. Patent No. 8,563,314. 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 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.

**[0126]** In certain embodiments, the nuclease comprises a CRISPR/Cas system. 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.

**[0127]** 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 foreign 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 foreign 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 foreign DNA *etc.*

**[0128]** 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, 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, 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 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 produce Cas protein and is genetically engineered to produce a Cas protein. *See, e.g.,* U.S. Provisional Application No. 61/823,689.

**[0129]** Thus, the nuclease comprises a DNA-binding domain in that specifically binds to a target site in any gene in combination with a nuclease domain that cleaves DNA at or near the binding site.

**Fusion proteins**

**[0130]** Methods for design and construction of fusion proteins (and polynucleotides encoding same) are known to those of skill in the art. For example, methods for the design and construction of fusion proteins comprising DNA-binding domains (*e.g.,* zinc finger domains, TALEs) and regulatory or cleavage domains (or cleavage half-domains), and polynucleotides encoding such fusion proteins, are described in U.S. Patents 8,586,526; 8,592,645; 8,399,218; 8,329,986; 7,888,121; 6,453,242; and 6,534,261 and U.S. Patent Application Publications 2007/0134796. In certain embodiments, polynucleotides encoding the fusion proteins are constructed. These polynucleotides can be inserted into a vector and the vector can be introduced into a cell (see below for additional disclosure regarding vectors and methods for introducing polynucleotides into cells).

**[0131]** In certain embodiments of the methods described herein, a zinc finger nuclease or TALEN comprises a fusion protein comprising a zinc finger binding domain or a TALE DNA binding domain and a nuclease domain (*e.g.,* Type IIS restriction enzyme and/or meganuclease domain). In certain embodiments, the ZFN or TALEN comprise a cleavage half-domain from the *Fok*I restriction enzyme, and two such fusion proteins are expressed in a cell. Expression of two fusion proteins in a cell can result from delivery of the two proteins to the cell; delivery of one protein and one nucleic acid encoding one of the proteins to the cell; delivery of two nucleic acids, each encoding one of the proteins, to the cell; or by delivery of a single nucleic acid, encoding both proteins, to the cell. In additional embodiments, a fusion protein comprises a single polypeptide chain comprising two cleavage half domains and a zinc finger or TALE binding domain. In this case, a single fusion protein is expressed in a cell and, without wishing to be bound by theory, is believed to cleave DNA as a result of formation of an intramolecular dimer of the cleavage half-domains.

**[0132]** In certain embodiments, the components of the fusion proteins (*e.g.,* ZFP-*Fok*I fusions) are arranged such that the DNA-binding domain is nearest the amino terminus of the fusion protein, and the cleavage half-domain is nearest the carboxy-terminus. This mirrors the relative orientation of the cleavage domain in naturally-occurring dimerizing cleavage domains such as those derived from the *Fok*I enzyme, in which the DNA-binding domain is nearest the amino terminus and the cleavage half-domain is nearest the carboxy terminus. In these embodiments,

dimerization of the cleavage half-domains to form a functional nuclease is brought about by binding of the fusion proteins to sites on opposite DNA strands, with the 5' ends of the binding sites being proximal to each other.

**[0133]** In additional embodiments, the components of the fusion proteins (*e.g.*, ZFP-*Fok*I fusions) are arranged such that the cleavage half-domain is nearest the amino terminus of the fusion protein, and the zinc finger domain is nearest the carboxy-terminus. In these embodiments, dimerization of the cleavage half-domains to form a functional nuclease is brought about by binding of the fusion proteins to sites on opposite DNA strands, with the 3' ends of the binding sites being proximal to each other.

**[0134]** In yet additional embodiments, a first fusion protein contains the cleavage half-domain nearest the amino terminus of the fusion protein, and the zinc finger domain nearest the carboxy-terminus, and a second fusion protein is arranged such that the zinc finger domain is nearest the amino terminus of the fusion protein, and the cleavage half-domain is nearest the carboxy-terminus. In these embodiments, both fusion proteins bind to the same DNA strand, with the binding site of the first fusion protein containing the zinc finger domain nearest the carboxy terminus located to the 5' side of the binding site of the second fusion protein containing the zinc finger domain nearest the amino terminus.

**[0135]** In certain embodiments of the disclosed fusion proteins, the amino acid sequence between the zinc finger domain and the cleavage domain (or cleavage half-domain) is denoted the "ZC linker." The ZC linker is to be distinguished from the inter-finger linkers discussed above. *See, e.g.,* U.S. Patent No. 7,888,121for details on obtaining ZC linkers that optimize cleavage.

**[0136]** In one embodiment, the disclosure provides a ZFN comprising a zinc finger protein having one or more of the recognition helix amino acid sequences shown in Table 2 (*e.g.,* a zinc finger protein made up of component zinc finger domains with the recognition helices as shown in a single row of Table 2). In another embodiment, provided herein is a ZFP expression vector comprising a nucleotide sequence encoding a ZFP having one or more recognition helices shown in Tables 2 or 12. In another embodiment, provided herein is a ZFP that binds to a target site as shown in Tables 3 or 13 or a polynucleotide encoding a ZFP that binds to a target site shown in Tables 3 or 13.

**Target Sites**

**[0137]** The disclosed methods and compositions include fusion proteins comprising a DNA-binding domain (*e.g.,* ZFP, TALE, *etc.*) and a regulatory domain or cleavage (*e.g.,* nuclease) domain (or a cleavage half-domain), in which the DNA-binding domain, by binding to a sequence in cellular chromatin in one or more plant genes, induces cleavage and targeted integration of one or more exogenous sequences (including transgenes) into the vicinity of the target sequence.

**[0138]** As set forth elsewhere in this disclosure, a DNA-binding domain can be engineered to bind to virtually any desired sequence. Accordingly, after identifying a region of interest containing a sequence at which gene regulation, cleavage, or recombination is desired, one or more DNA-binding domains can be engineered to bind to one or more sequences in the region of interest. In certain embodiments, the DNA-binding domain comprises a zinc finger protein that binds to a target site in one or more AHAS genes as shown in Table 3 or Table 13.

**[0139]** Selection of a target site in a genomic region of interest in cellular chromatin of any gene for binding by a DNA-binding domain (*e.g.,* a target site) can be accomplished, for example, according to the methods disclosed in U.S. Patent No. 6,453,242. It will be clear to those skilled in the art that simple visual inspection of a nucleotide sequence can also be used for selection of a target site. Accordingly, any means for target site selection can be used in the claimed methods.

**[0140]** Target sites are generally composed of a plurality of adjacent target subsites. In the case of zinc finger proteins, a target subsite refers to the sequence (usually either a nucleotide triplet, or a nucleotide quadruplet that can overlap by one nucleotide with an adjacent quadruplet) bound by an individual zinc finger. *See,* for example, U.S. Patent No. 6,794,136. If the strand with which a zinc finger protein makes most contacts is designated the target strand "primary recognition strand," or "primary contact strand," some zinc finger proteins bind to a three base triplet in the target strand and a fourth base on the non-target strand. A target site generally has a length of at least 9 nucleotides and, accordingly, is bound by a zinc finger binding domain comprising at least three zinc fingers. However binding of, for example, a 4-finger binding domain to a 12-nucleotide target site, a 5-finger binding domain to a 15-nucleotide target site or a 6-finger binding domain to an 18-nucleotide target site, is also possible. As will be apparent, binding of larger binding domains (*e.g.*, 7-, 8-, 9-finger and more) to longer target sites is also possible.

**[0141]** It is not necessary for a target site to be a multiple of three nucleotides. For example, in cases in which cross-strand interactions occur (see, *e.g.*, U.S. Patent 6,453,242 and 6,794,136), one or more of the individual zinc fingers of a multi-finger binding domain can bind to overlapping quadruplet subsites. As a result, a three-finger protein can bind a 10-nucleotide sequence, wherein the tenth nucleotide is part of a quadruplet bound by a terminal finger, a four-finger protein can bind a 13-nucleotide sequence, wherein the thirteenth nucleotide is part of a quadruplet bound by a terminal finger, *etc.*

**[0142]** In certain embodiments, the target site is in an AHAS locus (including untranslated regions such as the 3' untranslated region of AHAS). Non-limiting examples of suitable AHAS target sites are shown in Table 3 and Table 13. The AHAS (also known as AHAS/ALS) genes are present in all major plant species including but not limited to maize, soybean, cotton, *Arabidopsis*, rice, sunflower, wheat, barley, sugarbeet and *Brassica*. Specific amino acid modifications to the AHAS structural gene sequence have been described that alter the resulting proteins sensitivity to various structural classes of herbicides without a negative penalty on plant performance. For example, imidazolinone-tolerant maize (*Zea mays* L.) [Currie RS, Kwon CS and Penner D, Magnitude of imazethapyr resistance of corn (*Zea mays*) hybrids with altered acetolactate synthase. *Weed Sci* **43**:578–582 (1995), Wright TR and Penner D, Corn (*Zea mays*) acetolactate synthase sensitivity to four classes of ALS-inhibiting herbicides. *Weed Sci* **46**:8–12 (1998), Siehl DL, Bengtson AS, Brockman JP, Butler JH, Kraatz GW, Lamoreaux RJ and Subramanian MV, Patterns of cross tolerance to herbicides inhibiting acetohydroxyacid synthase in commercial corn hybrids designed for tolerance to imidazolinones. *Crop Sci* **36**:274–278 (1996), and Bailey WA and Wilcut JW, Tolerance of imidazolinone-resistant corn (*Zea mays*) to diclosulam. *Weed Technol* **17**:60–64 (2003)], rice (*Oryza sativa* L.) [Webster EP and Masson JA, Acetolactate synthase-inhibiting herbicides on imidazolinone-tolerant rice. *Weed Sci* **49**:652–657 (2001) and, Gealy DR, Mitten DH and Rutger JN, Gene flow between red rice (*Oryza sativa*) and herbicide-resistant rice (*O sativa*): implications for weed management. *Weed Technol* **17**:627–645 (2003)], bread wheat (*Triticum aestivum* L.) [Newhouse K, Smith WA, Starrett MA, Schaefer TJ and Singh BK, Tolerance to imidazolinone herbicides in wheat. *Plant Physiol* **100**:882–886 (1992), and Pozniak CJ and Hucl PJ, Genetic analysis of imidazolinone resistance in mutation-derived lines of common wheat. *Crop Sci* **44**:23–30 (2004)], and oilseed

rape (*Brassica napus* and *B. juncea* L. Czern.) [Shaner DL, Bascomb NF and Smith W, Imidazolinoneresistant crops: selection, characterization and management, in *Herbicide resistant crops*, edited by Duke SO, CRC Press, Boca Raton, pp 143–157 (1996) and Swanson EB, Herrgesell MJ, Arnoldo M, Sippell DW and Wong RSC, Microspore mutagenesis and selection: canola plants with field tolerance to the imidazolinones. *Theor Appl Genet* **78**:525–530 (1989)], were developed through mutagenesis, selection, and conventional breeding technologies and have been commercialized since 1992, 2003, 2002, and 1996, respectively. Several *AHAS* genes encoding AHAS enzymes that are tolerant to imidazolinone herbicides have been discovered in plants as naturally occurring mutations and through the process of chemically-induced mutagenesis. The S653N mutation is among the five most common single-point mutations in *AHAS* genes that result in tolerance to imidazolinone herbicides in plants (Tan, S., Evans, R.R., Dahmer, M.L., Singh, B.K., and Shaner, D.L. (2005) Imidazolinone-tolerant crops: History, current status and future. *Pest Manag. Sci.* 61:246-257).

**[0143]** The length and nature of amino acid linker sequences between individual zinc fingers in a multi-finger binding domain also affects binding to a target sequence. For example, the presence of a so-called "non-canonical linker," "long linker" or "structured linker" between adjacent zinc fingers in a multi-finger binding domain can allow those fingers to bind subsites which are not immediately adjacent. Non-limiting examples of such linkers are described, for example, in U.S. Patent Nos. 6,479,626 and 7,851,216. Accordingly, one or more subsites, in a target site for a zinc finger binding domain, can be separated from each other by 1, 2, 3, 4, 5 or more nucleotides. To provide but one example, a four-finger binding domain can bind to a 13-nucleotide target site comprising, in sequence, two contiguous 3-nucleotide subsites, an intervening nucleotide, and two contiguous triplet subsites. See, also, U.S. Patent Publication Nos. 20090305419 and 20110287512 for compositions and methods for linking artificial nucleases to bind to target sites separated by different numbers of nucleotides. Distance between sequences (*e.g.*, target sites) refers to the number of nucleotides or nucleotide pairs intervening between two sequences, as measured from the edges of the sequences nearest each other.

**[0144]** In certain embodiments, DNA-binding domains with transcription factor function are designed, for example by constructing fusion proteins comprising a

DNA-binding domain (*e.g.,* ZFP or TALE) and a transcriptional regulatory domain (*e.g.,* activation or repression domain). For transcription factor function, simple binding and sufficient proximity to the promoter are all that is generally needed. Exact positioning relative to the promoter, orientation, and within limits, distance does not matter greatly. This feature allows considerable flexibility in choosing target sites for constructing artificial transcription factors. The target site recognized by the DNA-binding domain therefore can be any suitable site in the target gene that will allow activation or repression of gene expression, optionally linked to a regulatory domain. Preferred target sites include regions adjacent to, downstream, or upstream of the transcription start site. In addition, target sites that are located in enhancer regions, repressor sites, RNA polymerase pause sites, and specific regulatory sites (*e.g.*, SP-1 sites, hypoxia response elements, nuclear receptor recognition elements, p53 binding sites), sites in the cDNA encoding region or in an expressed sequence tag (EST) coding region.

**[0145]** In other embodiments, ZFPs with nuclease activity are designed. Expression of a ZFN comprising a fusion protein comprising a zinc finger binding domain and a cleavage domain (or of two fusion proteins, each comprising a zinc finger binding domain and a cleavage half-domain), in a cell, effects cleavage in the vicinity of the target sequence. In certain embodiments, cleavage depends on the binding of two zinc finger domain/cleavage half-domain fusion molecules to separate target sites. The two target sites can be on opposite DNA strands, or alternatively, both target sites can be on the same DNA strand.

**[0146]** A variety of assays can be used to determine whether a ZFP modulates gene expression. The activity of a particular ZFP can be assessed using a variety of *in vitro* and *in vivo* assays, by measuring, *e.g.*, protein or mRNA levels, product levels, enzyme activity, transcriptional activation or repression of a reporter gene, using, *e.g.*, immunoassays (*e.g.*, ELISA and immunohistochemical assays with antibodies), hybridization assays (*e.g.*, RNase protection, northerns, *in situ* hybridization, oligonucleotide array studies), colorimetric assays, amplification assays, enzyme activity assays, phenotypic assays, and the like.

**[0147]** ZFPs are typically first tested for activity *in vitro* using ELISA assays and then using a yeast expression system. The ZFP is often first tested using a transient expression system with a reporter gene, and then regulation of the target endogenous gene is tested in cells and in whole plants, both *in vivo* and *ex vivo*. The

ZFP can be recombinantly expressed in a cell, recombinantly expressed in cells transplanted into a plant, or recombinantly expressed in a transgenic plant, as well as administered as a protein to plant or cell using delivery vehicles described below. The cells can be immobilized, be in solution, be injected into a plant, or be naturally occurring in a transgenic or non-transgenic plant.

**[0148]** Transgenic and non-transgenic plants are also used as a preferred embodiment for examining regulation of endogenous gene expression *in vivo*. Transgenic plants can stably express the ZFP of choice. Alternatively, plants that transiently express the ZFP of choice, or to which the ZFP has been administered in a delivery vehicle, can be used. Regulation of endogenous gene expression is tested using any one of the assays described herein.

**Methods for targeted cleavage**

**[0149]** The disclosed methods and compositions can be used to cleave DNA at a region of interest in cellular chromatin (*e.g.*, at a desired or predetermined site in a genome, for example, within or adjacent to an AHAS gene). For such targeted DNA cleavage, a DNA-binding domain (*e.g.,* zinc finger protein or TALE) is engineered to bind a target site at or near the predetermined cleavage site, and a fusion protein comprising the engineered zinc finger binding domain and a cleavage domain is expressed in a cell. Upon binding of the DNA-binding portion of the fusion protein to the target site, the DNA is cleaved near the target site by the cleavage domain.

**[0150]** Alternatively, two fusion proteins, each comprising a DNA- binding domain and a cleavage half-domain, are expressed in a cell, and bind to target 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 sites. In one embodiment, cleavage occurs between the target sites of the two DNA-binding domains. One or both of the zinc finger binding domains can be engineered.

**[0151]** For targeted cleavage using a zinc finger binding domain-cleavage domain fusion polypeptide, the binding site can encompass the cleavage site, or the near edge of the binding site can be 1, 2, 3, 4, 5, 6, 10, 25, 50 or more nucleotides (or any integral value between 1 and 50 nucleotides) from the cleavage site. The exact location of the binding site, with respect to the cleavage site, will depend upon the particular cleavage domain, and the length of the ZC linker. For methods in which two fusion polypeptides, each comprising a zinc finger binding domain and a

cleavage half-domain, are used, the binding sites generally straddle the cleavage site. Thus the near edge of the first binding site can be 1, 2, 3, 4, 5, 6, 10, 25 or more nucleotides (or any integral value between 1 and 50 nucleotides) on one side of the cleavage site, and the near edge of the second binding site can be 1, 2, 3, 4, 5, 6, 10, 25 or more nucleotides (or any integral value between 1 and 50 nucleotides) on the other side of the cleavage site. Methods for mapping cleavage sites *in vitro* and *in vivo* are known to those of skill in the art.

**[0152]** Thus, the methods described herein can employ an engineered zinc finger binding domain fused to a cleavage domain. In these cases, the binding domain is engineered to bind to a target sequence, at or near where cleavage is desired. The fusion protein, or a polynucleotide encoding same, is introduced into a plant cell. Once introduced into, or expressed in, the cell, the fusion protein binds to the target sequence and cleaves at or near the target sequence. The exact site of cleavage depends on the nature of the cleavage domain and/or the presence and/or nature of linker sequences between the binding and cleavage domains. In cases where two fusion proteins, each comprising a cleavage half-domain, are used, the distance between the near edges of the binding sites can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 25 or more nucleotides (or any integral value between 1 and 50 nucleotides). Optimal levels of cleavage can also depend on both the distance between the binding sites of the two fusion proteins (*see*, for example, Smith *et al.* (2000) *Nucleic Acids Res.* 28:3361-3369; Bibikova *et al.* (2001) *Mol. Cell. Biol.* 21:289-297) and the length of the ZC linker in each fusion protein. *See, also,* U.S. Patent Publication 20050064474A1 and International Patent Publications WO05/084190, WO05/014791 and WO03/080809.

**[0153]** In certain embodiments, the cleavage domain comprises two cleavage half-domains, both of which are part of a single polypeptide comprising a binding domain, a first cleavage half-domain and a second cleavage half-domain. The cleavage half-domains can have the same amino acid sequence or different amino acid sequences, so long as they function to cleave the DNA.

**[0154]** Cleavage half-domains may also be provided in separate molecules. For example, two fusion polypeptides may be introduced into a cell, wherein each polypeptide comprises a binding domain and a cleavage half-domain. The cleavage half-domains can have the same amino acid sequence or different amino acid sequences, so long as they function to cleave the DNA. Further, the binding domains

bind to target sequences which are typically disposed in such a way that, upon binding of the fusion polypeptides, the two cleavage half-domains are presented in a spatial orientation to each other that allows reconstitution of a cleavage domain (*e.g.*, by dimerization of the half-domains), thereby positioning the half-domains relative to each other to form a functional cleavage domain, resulting in cleavage of cellular chromatin in a region of interest. Generally, cleavage by the reconstituted cleavage domain occurs at a site located between the two target sequences. One or both of the proteins can be engineered to bind to its target site.

**[0155]** The two fusion proteins can bind in the region of interest in the same or opposite polarity, and their binding sites (*i.e.*, target sites) can be separated by any number of nucleotides, *e.g.*, from 0 to 200 nucleotides or any integral value therebetween. In certain embodiments, the binding sites for two fusion proteins, each comprising a zinc finger binding domain and a cleavage half-domain, can be located between 5 and 18 nucleotides apart, for example, 5-8 nucleotides apart, or 15-18 nucleotides apart, or 6 nucleotides apart, or 16 nucleotides apart, as measured from the edge of each binding site nearest the other binding site, and cleavage occurs between the binding sites.

**[0156]** The site at which the DNA is cleaved generally lies between the binding sites for the two fusion proteins. Double-strand breakage of DNA often results from two single-strand breaks, or "nicks," offset by 1, 2, 3, 4, 5, 6 or more nucleotides, (for example, cleavage of double-stranded DNA by native *Fok* I results from single-strand breaks offset by 4 nucleotides). Thus, cleavage does not necessarily occur at exactly opposite sites on each DNA strand. In addition, the structure of the fusion proteins and the distance between the target sites can influence whether cleavage occurs adjacent a single nucleotide pair, or whether cleavage occurs at several sites. However, for many applications, including targeted recombination and targeted mutagenesis (see *infra*) cleavage within a range of nucleotides is generally sufficient, and cleavage between particular base pairs is not required.

**[0157]** As noted above, the fusion protein(s) 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.

**[0158]** To enhance cleavage specificity, additional compositions may also be employed in the methods described herein. For example, single cleavage half-domains can exhibit limited double-stranded cleavage activity. In methods in which two fusion proteins, each containing a three-finger zinc finger domain and a cleavage half-domain, are introduced into the cell, either protein specifies an approximately 9-nucleotide target site. Although the aggregate target sequence of 18 nucleotides is likely to be unique in a mammalian and plant genomes, any given 9-nucleotide target site occurs, on average, approximately 23,000 times in the human genome. Thus, non-specific cleavage, due to the site-specific binding of a single half-domain, may occur. Accordingly, the methods described herein contemplate the use of a dominant-negative mutant of a nuclease (or a nucleic acid encoding same) that is expressed in a cell along with the two fusion proteins. The dominant-negative mutant is capable of dimerizing but is unable to induce double-stranded cleavage when dimerized. By providing the dominant-negative mutant in molar excess to the fusion proteins, only regions in which both fusion proteins are bound will have a high enough local concentration of functional cleavage half-domains for dimerization and double-stranded cleavage to occur.

**[0159]** In other embodiments, the nuclease domain(s) are nickases in that they induce single-stranded break. In certain embodiments, the nickase comprises two nucleases domains one of which is modified (*e.g.*, to be catalytically inactive) such that the nuclease makes only a single-stranded break. Such nickases are described for example in U.S. Patent Publication No. 20100047805. Two nickases may be used to induce a double-stranded break.

**Expression vectors**

**[0160]** A nucleic acid encoding one or more fusion proteins (*e.g.,* ZFNs, TALENs, *etc.*) as described herein can be cloned into a vector for transformation into prokaryotic or eukaryotic cells for replication and/or expression. Vectors can be prokaryotic vectors (*e.g.*, plasmids, or shuttle vectors, insect vectors) or eukaryotic vectors. A nucleic acid encoding a fusion protein can also be cloned into an expression vector, for administration to a cell.

**[0161]** To express the fusion proteins, sequences encoding the fusion proteins are typically subcloned into an expression vector that contains a promoter to direct transcription. Suitable prokaryotic and eukaryotic promoters are well known in the art and described, *e.g.*, in Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (2nd ed. 1989; 3rd ed., 2001); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); and *Current Protocols in Molecular Biology* (Ausubel *et al.*, *supra.* Bacterial expression systems for expressing the ZFP are available in, *e.g., E. coli, Bacillus sp.*, and *Salmonella* (Palva *et al.*, *Gene* 22:229-235 (1983)). Kits for such expression systems are commercially available. Eukaryotic expression systems for mammalian cells, yeast, and insect cells are well known by those of skill in the art and are also commercially available.

**[0162]** The promoter used to direct expression of a fusion protein-encoding nucleic acid depends on the particular application. For example, a strong constitutive promoter suited to the host cell is typically used for expression and purification of fusion proteins.

**[0163]** In contrast, when a fusion protein is administered *in vivo* for regulation of a plant gene (see, "Nucleic Acid Delivery to Plant Cells" section below), either a constitutive, regulated (*e.g.*, during development, by tissue or cell type, or by the environment) or an inducible promoter is used, depending on the particular use of the fusion protein. Non-limiting examples of plant promoters include promoter sequences derived from *A. thaliana* ubiquitin-3 (ubi-3) (Callis, *et al.*, 1990, *J. Biol. Chem.* 265-12486-12493); *A. tumefaciens* mannopine synthase (Δmas) (Petolino *et al.*, U.S. Patent No. 6,730,824); and/or Cassava Vein Mosaic Virus (CsVMV) (Verdaguer *et al.*, 1996, *Plant Molecular Biology* 31:1129-1139). *See, also,* Examples.

**[0164]** In addition to the promoter, the expression vector typically contains a transcription unit or expression cassette that contains all the additional elements required for the expression of the nucleic acid in host cells, either prokaryotic or eukaryotic. A typical expression cassette thus contains a promoter (comprising ribosome binding sites) operably linked, *e.g.*, to a nucleic acid sequence encoding the fusion protein, and signals required, *e.g.*, for efficient polyadenylation of the transcript, transcriptional termination, or translation termination. Additional elements of the cassette may include, *e.g.*, enhancers, heterologous splicing signals, the 2A

sequence from *Thosea asigna* virus (Mattion *et al.* (1996) *J. Virol.* 70:8124-8127), and/or a nuclear localization signal (NLS).

**[0165]** The particular expression vector used to transport the genetic information into the cell is selected with regard to the intended use of the fusion proteins, *e.g.*, expression in plants, animals, bacteria, fungus, protozoa, *etc.* (see expression vectors described below). Standard bacterial and animal expression vectors are known in the art and are described in detail, for example, U.S. Patent Publication 20050064474A1 and International Patent Publications WO 05/084190, WO 05/014791 and WO 03/080809.

**[0166]** Standard transfection methods can be used to produce bacterial, plant, mammalian, yeast or insect cell lines that express large quantities of protein, which can then be purified using standard techniques (*see, e.g.,* Colley *et al.*, *J. Biol. Chem.* 264:17619-17622 (1989); *Guide to Protein Purification, in Methods in Enzymology*, vol. 182 (Deutscher, ed., 1990)). Transformation of eukaryotic and prokaryotic cells are performed according to standard techniques (*see, e.g.,* Morrison, *J. Bact.* 132:349-351 (1977); Clark-Curtiss & Curtiss, *Methods in Enzymology* 101:347-362 (Wu *et al.*, eds., 1983).

**[0167]** Any of the well-known procedures for introducing foreign nucleotide sequences into such host cells may be used. These include the use of calcium phosphate transfection, polybrene, protoplast fusion, electroporation, ultrasonic methods (*e.g.*, sonoporation), liposomes, microinjection, naked DNA, plasmid vectors, viral vectors, *Agrobacterium*-mediated transformation, silicon carbide (*e.g.,* WHISKERS™) mediated transformation, both episomal and integrative, and any of the other well-known methods for introducing cloned genomic DNA, cDNA, synthetic DNA or other foreign genetic material into a host cell (*see, e.g.,* Sambrook *et al.*, *supra*). It is only necessary that the particular genetic engineering procedure used be capable of successfully introducing at least one gene into the host cell capable of expressing the protein of choice.

**Donors**

**[0168]** As noted above, insertion of one or more exogenous sequence (also called a "donor sequence" or "donor" or "transgene"), for example for stacking can also be completed. A donor sequence can contain a non-homologous sequence flanked by two regions of homology to allow for efficient HDR at the location of

interest. Additionally, donor sequences can comprise a vector molecule containing sequences that are not homologous to the region of interest in cellular chromatin. A donor molecule can contain several, discontinuous regions of homology to cellular chromatin. For example, for targeted insertion of sequences not normally present in a region of interest, said sequences can be present in a donor nucleic acid molecule and flanked by regions of homology to sequence in the region of interest.

**[0169]** The donor polynucleotide can be DNA or RNA, single-stranded or double-stranded and can be introduced into a cell in linear or circular form. *See, e.g.,* U.S. Patent Publication Nos. 20100047805; 20110281361; and 20110207221. If introduced in linear form, the ends of the donor sequence can be protected (*e.g.*, from exonucleolytic degradation) by methods known to those of skill in the art. For example, one or more dideoxynucleotide residues are added to the 3' terminus of a linear molecule and/or self-complementary oligonucleotides are ligated to one or both ends. *See*, for example, Chang *et al.* (1987) *Proc. Natl. Acad. Sci. USA*84:4959-4963; Nehls *et al.* (1996) *Science* 272:886-889. Additional methods for protecting exogenous polynucleotides from degradation include, but are not limited to, addition of terminal amino group(s) and the use of modified internucleotide linkages such as, for example, phosphorothioates, phosphoramidates, and O-methyl ribose or deoxyribose residues.

**[0170]** A polynucleotide can be introduced into a cell as part of a vector molecule having additional sequences such as, for example, replication origins, promoters and genes encoding antibiotic resistance. Moreover, donor polynucleotides can be introduced as naked nucleic acid, as nucleic acid complexed with an agent such as a liposome or poloxamer, or can be delivered by viruses (*e.g.*, adenovirus, AAV, herpesvirus, retrovirus, lentivirus and integrase defective lentivirus (IDLV)). *See, e.g.,* U.S, Patent Publication No. 20090117617.

**[0171]** The donor is generally inserted so that its expression is driven by the endogenous promoter at the integration site, namely the promoter that drives expression of the endogenous gene into which the donor is inserted. However, it will be apparent that the donor may comprise a promoter and/or enhancer, for example a constitutive promoter or an inducible or tissue specific promoter. Furthermore, the donor molecule may be inserted into an endogenous gene such that all, some or none of the endogenous gene is expressed.

**[0172]** Furthermore, although not required for expression, exogenous sequences 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.

**[0173]** The donor sequence is introduced into an endogenous gene (or multiple alleles of the gene) such that the function of the endogenous gene is altered to act as an endogenous marker for transgene integration, thereby resulting in a genomic modification. In certain embodiments, the endogenous locus into which the transgene(s) is (are) introduced is an AHAS locus. Several mutations in the AHAS gene are known to confer Group B, or ALS inhibitor herbicide tolerance (for example imidazolinone or sulfonylurea), including a single mutation of serine at position 653 to asparagine (S653N). *See, e.g.,* Lee *et al.* (2011) *Proc. Nat'l. Acad. Sci. USA* 108: 8909–8913, and Tan, S., Evans, R.R., Dahmer, M.L., Singh, B.K., and Shaner, D.L. (2005) Imidazolinone-tolerant crops: History, current status and future. *Pest Manag. Sci.* 61:246-257.

**[0174]** AHAS is one desirable locus because the gene is transcriptionally active at all stages of plant development, it is not prone to gene silencing (*e.g.,* by DNA, Histone Methylation, iRNA, *etc.*), where the insertion of a new gene or plant transformation unit into this locus does not have a negative impact on the agronomic or quality properties of the host plant. The ubiquitous nature of the AHAS locus and clear commercial evidence that alteration AHAS locus or loci in canola, corn, sunflower, cotton, soybean, sugar beet, wheat, and any other plant does not carry an agronomic or quality penalty means the AHAS loci represents broad class of a preferred target loci across all commercially relevant plant species.

**[0175]** Integration of the donor DNA into the wild type (herbicide susceptible) AHAS locus typically both introduces an exogenous sequence (e.g., a transgene) and a mutation to the endogenous AHAS to produce a genomic modification that confers tolerance to imidazolinones (*i.e.*, a product that results in an herbicide tolerant plant cell), thus allowing regeneration of correctly targeted plants using an endogenous imidazolinone selection system rather than a transgenic selection marker system. Stacking of a second transgene at the AHAS locus can be achieved by integration of a donor DNA that introduces one or more additional transgenes, confers susceptibility to imidazolines but tolerance to sulfonylureas (*i.e.*, a product that results in an herbicide tolerant plant cell), thus allowing regeneration of correctly targeted plants

using a sulfonylurea selection agent. Stacking of a third transgene can be achieved by integration of a donor DNA that introduces further transgene(s) and confers susceptibility to sulfonylurea and tolerance to imidazolinones, thus allowing 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 molecules that introduce mutations (*e.g.*, genomic modification) to wild-type AHAS thus allowing differential cycling between sulfonylurea and imidazolinone chemical selection agents.

**Nucleic Acid Delivery to Plant Cells**

**[0176]** As noted above, DNA constructs (*e.g.*, nuclease(s) and/or donor(s)) may be introduced into (*e.g.*, into the genome of) a desired plant host by a variety of conventional techniques. For reviews of such techniques see, for example, Weissbach & Weissbach *Methods for Plant Molecular Biology* (1988, Academic Press, N.Y.) Section VIII, pp. 421-463; and Grierson & Corey, *Plant Molecular Biology* (1988, 2d Ed.), Blackie, London, Ch. 7-9. *See, also,* U.S. Patent Publication Nos. 20090205083; 20100199389; 20110167521 and 20110189775. It will be apparent that one or more DNA constructs can be employed in the practice of the present invention, for example the nuclease(s) may be carried by the same construct or different constructs as the construct(s) carrying the donor(s).

**[0177]** The DNA construct(s) may be introduced directly into the genomic DNA of the plant cell using techniques such as electroporation and microinjection of plant cell protoplasts, or the DNA constructs can be introduced directly to plant tissue using biolistic methods, such as DNA particle bombardment (see, *e.g.*, Klein *et al.* (1987) *Nature* **327**:70-73). Alternatively, the DNA construct can be introduced into the plant cell via nanoparticle transformation (*see, e.g.,* U.S. Patent Publication No. 20090104700). Alternatively, the DNA constructs may be combined with suitable T-DNA border/flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector. *Agrobacterium tumefaciens*-mediated transformation techniques, including disarming of oncogenes and the development and use of binary vectors, are well described in the scientific literature. See, for example Horsch *et al.*

(1984) *Science* **233**:496-498, and Fraley *et al.* (1983) *Proc. Nat'l. Acad. Sci. USA* **80**:4803.

**[0178]** In addition, gene transfer may be achieved using non-Agrobacterium bacteria or viruses such as *Rhizobium sp.* NGR234, *Sinorhizoboium meliloti, Mesorhizobium loti,* potato virus X, cauliflower mosaic virus and cassava vein mosaic virus and/or tobacco mosaic virus, *See, e.g.,* Chung *et al.* (2006) *Trends Plant Sci.* **11**(1):1-4.

**[0179]** The virulence functions of the *Agrobacterium tumefaciens* host will direct the insertion of a T-strand containing the construct and adjacent marker into the plant cell DNA when the cell is infected by the bacteria using binary T-DNA vector (Bevan (1984) *Nuc. Acid Res.* **12**:8711-8721) or the co-cultivation procedure (Horsch *et al.* (1985) *Science* **227**:1229-1231). Generally, the *Agrobacterium* transformation system is used to engineer dicotyledonous plants (Bevan *et al.* (1982) *Ann. Rev. Genet* **16**:357-384; Rogers *et al.* (1986) *Methods Enzymol.* **118**:627-641). The *Agrobacterium* transformation system may also be used to transform, as well as transfer, DNA to monocotyledonous plants and plant cells. *See* U.S. Patent No. 5, 591,616; Hernalsteen *et al.* (1984) *EMBO J* **3**:3039-3041; Hooykass-Van Slogteren *et al.* (1984) *Nature* **311**:763-764; Grimsley *et al.* (1987) *Nature* **325**:1677-179; Boulton *et al.* (1989) *Plant Mol. Biol.* **12**:31-40; and Gould *et al.* (1991) *Plant Physiol.* **95**:426-434.

**[0180]** Alternative gene transfer and transformation methods include, but are not limited to, protoplast transformation through calcium-, polyethylene glycol (PEG)- or electroporation-mediated uptake of naked DNA (see Paszkowski *et al.* (1984) *EMBO J* **3**:2717-2722, Potrykus *et al.* (1985) *Molec. Gen. Genet.* **199**:169-177; Fromm *et al.* (1985) *Proc. Nat. Acad. Sci.* USA **82**:5824-5828; and Shimamoto (1989) *Nature* **338**:274-276) and electroporation of plant tissues (D'Halluin *et al.* (1992) *Plant Cell* **4**:1495-1505). Additional methods for plant cell transformation include microinjection, silicon carbide (*e.g.*, WHISKERS™) mediated DNA uptake (Kaeppler *et al.* (1990) *Plant Cell Reporter* **9**:415-418), and microprojectile bombardment (see Klein *et al.* (1988) *Proc. Nat. Acad. Sci. USA* **85**:4305-4309; and Gordon-Kamm *et al.* (1990) *Plant Cell* **2**:603-618). Finally, nanoparticles, nanocarriers and cell penetrating peptides can be utilized to deliver DNA, RNA, peptides and/or proteins into plant cells (see WO/2011/26644, WO/2009/046384, and WO/2008/148223).

**[0181]** The disclosed methods and compositions can be used to insert exogenous sequences into an AHAS gene. This is useful inasmuch as expression of an introduced transgene into a plant genome depends critically on its integration site and, as noted above, AHAS provides a suitable site for transgene integration. Accordingly, genes encoding, *e.g.*, herbicide tolerance, insect resistance, nutrients, antibiotics or therapeutic molecules can be inserted, by targeted recombination, into regions of a plant genome favorable to their expression.

**[0182]** Transformed plant cells which are produced by any of the above transformation techniques can be cultured to regenerate a whole plant which possesses the transformed genotype and thus the desired phenotype. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been introduced together with the desired nucleotide sequences. Plant regeneration from cultured protoplasts is described in Evans, *et al.*, "Protoplasts Isolation and Culture" in *Handbook of Plant Cell Culture,* pp. 124-176, Macmillan Publishing Company, New York, 1983; and Binding, *Regeneration of Plants, Plant Protoplasts*, pp. 21-73, CRC Press, Boca Raton, 1985. Regeneration can also be obtained from plant callus, explants, organs, pollens, embryos or parts thereof. Such regeneration techniques are described generally in Klee *et al.* (1987) *Ann. Rev. of Plant Phys.* **38**:467-486.

**[0183]** Nucleic acids introduced into a plant cell can be used to confer desired traits on essentially any plant. A wide variety of plants and plant cell systems may be engineered for the desired physiological and agronomic characteristics described herein using the nucleic acid constructs of the present disclosure and the various transformation methods mentioned above. In preferred embodiments, target plants and plant cells for engineering include, but are not limited to, those monocotyledonous and dicotyledonous plants, such as crops including grain crops (*e.g.*, wheat, maize, rice, millet, barley), fruit crops (*e.g.*, tomato, apple, pear, strawberry, orange), forage crops (*e.g.*, alfalfa), root vegetable crops (*e.g.*, carrot, potato, sugar beets, yam), leafy vegetable crops (*e.g.*, lettuce, spinach); flowering plants (*e.g.*, petunia, rose, chrysanthemum), conifers and pine trees (*e.g.*, pine fir, spruce); plants used in phytoremediation (*e.g.*, heavy metal accumulating plants); oil crops (*e.g.*, sunflower, rape seed) and plants used for experimental purposes (*e.g.*, *Arabidopsis*). Thus, the disclosed methods and compositions have use over a broad

range of plants, including, but not limited to, species from the genera Asparagus, Avena, Brassica, Citrus, Citrullus, Capsicum, Cucurbita, Daucus, Erigeron, Glycine, Gossypium, Hordeum, Lactuca, Lolium, Lycopersicon, Malus, Manihot, Nicotiana, Orychophragmus, Oryza, Persea, Phaseolus, Pisum, Pyrus, Prunus, Raphanus, Secale, Solanum, Sorghum, Triticum, Vitis, Vigna, and Zea.

**[0184]** The introduction of nucleic acids introduced into a plant cell can be used to confer desired traits on essentially any plant. In certain embodiments, the integrated transgene(s) in plant cells results in plants having increased amount of fruit yield, increased biomass of plant (or fruit of the plant), higher content of fruit flesh, concentrated fruit set, larger plants, increased fresh weight, increased dry weight, increased solids context, higher total weight at harvest, enhanced intensity and/or uniformity of color of the crop, altered chemical (*e.g.,* oil, fatty acid, carbohydrate, protein) characteristics, *etc.*

**[0185]** One with skill in the art will recognize that an exogenous sequence can be transiently incorporated into a plant cell. The introduction of an exogenous polynucleotide sequence can utilize the cell machinery of the plant cell in which the sequence has been introduced. The expression of an exogenous polynucleotide sequence comprising a ZFN that is transiently incorporated into a plant cell can be assayed by analyzing the genomic DNA of the target sequence to identify and determine any indels, inversions, or insertions. These types of rearrangements result from the cleavage of the target site within the genomic DNA sequence, and the subsequent DNA repair. In addition, the expression of an exogenous polynucleotide sequence can be assayed using methods which allow for the testing of marker gene expression known to those of ordinary skill in the art. Transient expression of marker genes has been reported using a variety of plants, tissues, and DNA delivery systems. Transient analyses systems include but are not limited to direct gene delivery via electroporation or particle bombardment of tissues in any transient plant assay using any plant species of interest. Such transient systems would include but are not limited to electroporation of protoplasts from a variety of tissue sources or particle bombardment of specific tissues of interest. The present disclosure encompasses the use of any transient expression system to evaluate a site specific endonuclease (*e.g.,* ZFN) and to introduce transgenes and/or mutations within a target gene (*e.g.,* AHAS) to result in a genomic modification. Examples of plant tissues envisioned to test in transients via an appropriate delivery system would include but are not limited to leaf

base tissues, callus, cotyledons, roots, endosperm, embryos, floral tissue, pollen, and epidermal tissue.

**[0186]** One of skill in the art will recognize that an exogenous polynucleotide sequence can be stably incorporated in transgenic plants. Once the exogenous polynucleotide sequence is confirmed to be operable, it can be introduced into other plants by sexual crossing. Any of a number of standard breeding techniques can be used, depending upon the species to be crossed.

**[0187]** A transformed plant cell, callus, tissue or plant may be identified and isolated by selecting or screening the engineered plant material for a phenotype encoded by the markers present on the exogenous DNA sequence. Markers may also be described and referred to as selectable markers, or reporter markers. The markers can be utilized for the identification and selection of transformed plants ("transformants"). Typically, the marker is incorporated into the genome of a plant cell as an exogenous sequence. In some examples, the exogenous marker sequence is incorporated into the plant genome at a site specific target loci as a donor sequence, wherein the donor sequence contains mutations which result in tolerance to a selection agent (*e.g.*, herbicides, *etc.*). In other examples, the exogenous marker sequence is incorporated into the plant genome as a transgene (*i.e.*, "transgenic selectable marker"), wherein the marker gene is operably linked to a promoter and 3'-UTR to comprise a chimeric gene expression cassette. The expression of the marker gene results in expression of a visual marker protein or in tolerance to a selection agent (*e.g.*, herbicide, antibiotics, *etc.*).

**[0188]** For instance, selection can be performed by growing the engineered plant material on media containing an inhibitory amount of the antibiotic or herbicide (*i.e.*, also described as a selective agent) to which the transforming gene construct confers tolerance. In an embodiment, selectable marker genes include herbicide tolerance genes.

**[0189]** Herbicide tolerance markers code for a modified target protein insensitive to the herbicide, or for an enzyme that degrades and detoxifies the herbicide in the plant before it can act. For example, a modified target protein insensitive to an herbicide would include tolerance to glyphosate. Plants tolerant to glyphosate have been obtained by using genes coding for mutant target enzyme *5-enolpyruvylshikimate-3-phosphate synthase* (EPSPS). Genes and mutants for EPSPS are well known, and include mutant *5-enolpyruvylshikimate-3-phosphate synthase*

(EPSPs), *dgt-28*, and aroA genes. Such genes provide tolerance to glyphosate via the introduction of recombinant nucleic acids and/or various forms of *in vivo* mutagenesis of the native EPSPs genes. An example of enzymes that degrade and detoxify herbicides in the plant would include tolerance to glufosinate ammonium, bromoxynil, and 2,4-dichlorophenoxyacetate (2,4-D). Tolerance to these herbicides has been obtained by expressing bacterial genes that encode *pat* or *DSM-2*, a *nitrilase*, an *aad*-1 or an *aad*-12 gene within a plant cell as a transgene. Tolerance genes for phosphono compounds include *bar* and *pat* genes from *Streptomyces* species, including *Streptomyces hygroscopicus* and *Streptomyces viridichromogenes*, and pyridinoxy or phenoxy proprionic acids and cyclohexones (ACCase inhibitor-encoding genes). Exemplary genes conferring tolerance to cyclohexanediones and/or aryloxyphenoxypropanoic acid (including Haloxyfop, Diclofop, Fenoxyprop, Fluazifop, Quizalofop) include genes of acetyl coenzyme A carboxylase (ACCase), such as; Acc1-S1, Acc1-S2 and Acc1-S3. In an embodiment, herbicides can inhibit photosynthesis, including triazine (psbA and 1s+ genes) or benzonitrile (nitrilase gene). Other herbicide tolerant gene sequences are known by those with skill in the art.

**[0190]** Antibiotic resistant markers code for an enzyme that degrades and detoxifies an antibiotic in the plant before it can act on the plant. Various types of antibiotics are known that can impede plant growth and development when used at proper concentrations, such as kanamycin, chloramphenicol, spectinomycin, and hygromycin. Exogenous sequences can be obtained (*e.g.*, bacterial genes) and expressed as a transgene to breakdown the antibiotic. For example, antibiotic resistant marker genes include exogenous sequences encoding antibiotic resistance, such as the genes encoding neomycin phosphotransferase II (NEO), chloramphenicol acetyltransferase (CAT), alkaline phosphatase, spectinomycin resistance, kanamycin resistance, and hygromycin phosphotransferase (HPT).

**[0191]** Further, transformed plants and plant cells can also be identified by screening for the activities of a reporter gene that encode a visible marker gene. Reporter genes are typically provided as recombinant nucleic acid constructs and integrated into the plant cell as a transgene. Visual observation of proteins such as reporter genes encoding β-glucuronidase (GUS), luciferase, green fluorescent protein (GFP), yellow fluorescent protein (YFP), DsRed, β-galactosidase may be used to

identify and select transformants. Such selection and screening methodologies are well known to those skilled in the art.

**[0192]** The above list of marker genes is not meant to be limiting. Any reporter or selectable marker gene is encompassed by the present disclosure. Moreover, it should be appreciated that markers (*e.g.*, herbicide tolerant markers) are primarily utilized for the identification and selection of transformed plants, as compared to a trait (*e.g.*, herbicide tolerant traits) that are utilized for providing tolerance to herbicides applied in a field environment to control weed species.

**[0193]** Physical and biochemical methods also may be used to identify plant or plant cell transformants containing stably inserted gene constructs, or plant cell containing target gene altered genomic DNA which results from the transient expression of a site-specific endonuclease (*e.g.,* ZFN). These methods include but are not limited to: 1) Southern analysis or PCR amplification for detecting and determining the structure of the recombinant DNA insert; 2) Northern blot, S1 RNase protection, primer-extension or reverse transcriptase-PCR amplification for detecting and examining RNA transcripts of the gene constructs; 3) enzymatic assays for detecting enzyme or ribozyme activity, where such gene products are encoded by the gene construct; 4) protein gel electrophoresis, Western blot techniques, immunoprecipitation, or enzyme-linked immunoassays (ELISA), where the gene construct products are proteins. Additional techniques, such as *in situ* hybridization, enzyme staining, and immunostaining, also may be used to detect the presence or expression of the recombinant construct in specific plant organs and tissues. The methods for doing all these assays are well known to those skilled in the art.

**[0194]** Effects of gene manipulation using the methods disclosed herein can be observed by, for example, northern blots of the RNA (*e.g.,* mRNA) isolated from the tissues of interest. Typically, if the mRNA is present or the amount of mRNA has increased, it can be assumed that the corresponding transgene is being expressed. Other methods of measuring gene and/or encoded polypeptide activity can be used. Different types of enzymatic assays can be used, depending on the substrate used and the method of detecting the increase or decrease of a reaction product or by-product. In addition, the levels of polypeptide expressed can be measured immunochemically, *i.e.*, ELISA, RIA, EIA and other antibody based assays well known to those of skill in the art, such as by electrophoretic detection assays (either with staining or western blotting). As one non-limiting example, the detection of the AAD-1 and PAT proteins

using an ELISA assay is described in U.S. Patent Publication No. 20090093366. A transgene may be selectively expressed in some tissues of the plant or at some developmental stages, or the transgene may be expressed in substantially all plant tissues, substantially along its entire life cycle. However, any combinatorial expression mode is also applicable.

**[0195]** The present disclosure also encompasses seeds of the transgenic plants described above wherein the seed has the transgene or gene construct. The present disclosure further encompasses the progeny, clones, cell lines or cells of the transgenic plants described above wherein said progeny, clone, cell line or cell has the transgene or gene construct.

**[0196]** Fusion proteins (*e.g.*, ZFNs) and expression vectors encoding fusion proteins can be administered directly to the plant for gene regulation, targeted cleavage, and/or recombination. In certain embodiments, the plant contains multiple paralogous target genes. For example, for AHAS, *Brassica napus* includes 5 paralogs and wheat includes 3 paralogs. Thus, one or more different fusion proteins or expression vectors encoding fusion proteins may be administered to a plant in order to target one or more of these paralogous genes in the plant.

**[0197]** Administration of effective amounts is by any of the routes normally used for introducing fusion proteins into ultimate contact with the plant cell to be treated. The ZFPs are administered in any suitable manner, preferably with acceptable carriers. Suitable methods of administering such modulators are available and well known to those of skill in the art, and, 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.

**[0198]** Carriers may also be used and are determined in part by the 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 carriers that are available.

# EXAMPLES

## Example 1: Characterization of AHAS Genomic Target Sequences

### *Identification of AHAS Sequences*

**[0199]** The transcribed regions for three homoeologous AHAS genes were identified and determined. These novel sequences are listed as SEQ ID NO:1, SEQ ID NO:2, and SEQ ID NO:3. 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, AY273827.1) were used to determine the transcribed region for the homoeologous copies of the AHAS gene (SEQ ID NOs: 1-3).

**[0200]** The novel, non-coding sequences located upstream and downstream of the transcribed region were characterized for the first time. To completely characterize theses 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:4-6).

***Sequence Analysis of AHAS Genes Isolated From Triticum aestivum* cv. *Bobwhite MPB26RH***

**[0201]** 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 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.

**[0202]** A cohort of PCR primers (Table 1) were designed for amplification of the AHAS genes. 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.

**[0203]** 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 | 7 | TCTGTAAGTTATCGCCTGAATTGCTT |
| AHAS-p_Rvs6 | 5' UTR | D | 8 | CATTGTGACATCAGCATGACACAA |
| AHAS-p_Fwd4 | 5' UTR | D | 9 | AAGCAYGGCTTGCCTACAGC |
| AHAS-p_Rvs3 | 5' UTR | D | 10 | AACCAAATRCCCCTATGTCTCTCC |
| AHAS-p_Fwd1 | 5' UTR | A, B, and D | 11 | CGTTCGCCCGTAGACCATTC |
| AHAS-p_Rvs1 | 5' UTR | A, B, and D | 12 | GGAGGGGTGATGKTTTTGTCTTT |
| AHAS_1F1_transcribed | Coding | A, B, and D | 13 | TCG CCC AAA CCC TCG CC |
| AHAS_1R1_transcribed | Coding | A, B, and D | 14 | GGG TCG TCR CTG GGG AAG TT |
| AHAS_2F2_transcribed | Coding | A, B, and D | 15 | GCC TTC TTC CTY GCR TCC TCT GG |
| AHAS_2R2_transcribed | Coding | A, B, and D | 16 | GCC CGR TTG GCC TTG TAA AAC CT |
| AHAS_3F1_transcribed | Coding | A, B, and D | 17 | AYC AGA TGT GGG CGG CTC AGT AT |
| AHAS_3R1_transcribed | Coding | A, B, and D | 18 | GGG ATA TGT AGG ACA AGA AAC TTG CAT GA |
| AHAS-6A.PS.3'.F1 | 3'UTR | A | 19 | AGGGCCATACTTGTTGGATATCAT*C |
| AHAS-6A.PS.3'.R2 | 3'UTR | A | 20 | GCCAACACCCTACACTGCCTA*T |
| AHAS-6B.PS.3'.F1 | 3'UTR | B | 21 | TGCGCAATCAGCATGATACC*T |
| AHAS-6B.PS.3'.R1 | 3'UTR | B | 22 | ACGTATCCGCAGTCGAGCAA*T |
| AHAS-6D.PS.3'.F1 | 3'UTR | D | 23 | GTAGGGATGTGCTGTCATAAGAT*G |
| AHAS-6D.PS.3'.R3 | 3'UTR | D | 24 | TTGGAGGCTCAGCCGATCA*C |

UTR = untranslated region
Coding = primers designed for the transcribed regions
asterisk (*) indicates the incorporation of a phosphorothioate sequence

**[0204]** 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 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 $MgCl_2$, 0.25 units IMMOLASE DNA POLYMERASE™ (Bioline, Taunton, MA), 0.2 μM each of forward and reverse primer, and about 50 ng genomic DNA. 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 μl 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:25, SEQ ID NO:26, and SEQ ID NO:27) from cv. Bobwhite MPB26RH. CLUSTALW™ was used to produce a multiple consensus sequence alignment from which homoeologous sequence variation distinguishing between the AHAS gene copies was confirmed.

**Example 2: Design of Zinc Finger Binding Domains Specific to AHAS Gene Sequences**

**[0205]** 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 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 in 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 residue, and a region about 400-bp downstream of the serine 653 amino acid residue.

**Table 2: AHAS zinc finger designs (N/A indicates "not applicable")**

| ZFP # | F1 | F2 | F3 | F4 | F5 | F6 |
|---|---|---|---|---|---|---|
| 29964 | QSSHLTR SEQ ID NO:181 | RSDDLTR SEQ ID NO:182 | RSDDLTR SEQ ID NO:182 | YRWLLRS SEQ ID NO:183 | QSGDLTR SEQ ID NO:184 | QRNARTL SEQ ID NO:185 |
| 29965 | RSDNLSV SEQ ID NO:186 | QKINLQV SEQ ID NO:187 | DDWNLSQ SEQ ID NO:188 | RSANLTR SEQ ID NO:189 | QSGHLAR SEQ ID NO:190 | NDWDRRV SEQ ID NO:191 |
| 29966 | RSDDLTR SEQ ID NO:182 | YRWLLRS_ SEQ ID NO:183 | QSGDLTR SEQ ID NO:184 | QRNARTL_ SEQ ID NO:185 | RSDHLSQ_ SEQ ID NO:192 | DSSTRKK SEQ ID NO:193 |
| 29967 | RSDDLTR SEQ ID NO:182 | YRWLLRS_ SEQ ID NO:183 | QSGDLTR SEQ ID NO:184 | QRNARTL SEQ ID NO:185 | RSDVLSE SEQ ID NO:194 | DRSNRIK SEQ ID NO:195 |
| 29968 | RSDNLSN SEQ ID NO:196 | TSSSRIN SEQ ID NO:197 | DRSNLTR SEQ ID NO:198 | QSSDLSR SEQ ID NO:199 | QSAHRKN SEQ ID NO:200 | N/A |
| 29969 | DRSHLTR SEQ ID NO:201 | QSGHLSR SEQ ID NO:202 | RSDNLSV SEQ ID NO:186 | QKINLQV SEQ ID NO:187 | DDWNLSQ SEQ ID NO:188 | RSANLTR SEQ ID NO:189 |
| 29970 | QSGDLTR SEQ ID NO:184 | QRNARTL SEQ ID NO:185 | RSDVLSE SEQ ID NO:194 | DRSNRIK SEQ ID NO:195 | RSDNLSE SEQ ID NO:203 | HSNARKT SEQ ID NO:204 |
| 29971 | DRSHLTR SEQ ID NO:201 | QSGHLSR SEQ ID NO:202 | RSDNLSN SEQ ID NO:196 | TSSSRIN SEQ ID NO:197 | DRSNLTR SEQ ID NO:198 | N/A |
| 29730 | TSGNLTR SEQ ID NO:205 | HRTSLTD SEQ ID NO:206 | QSSDLSR SEQ ID NO:199 | HKYHLRS SEQ ID NO:207 | QSSDLSR SEQ ID NO:199 | QWSTRKR SEQ ID NO:208 |
| 29731 | RSDVLSE SEQ ID NO:194 | SPSSRRT SEQ ID NO:209 | RSDTLSE SEQ ID NO:210 | TARQRNR SEQ ID NO:211 | DRSHLAR SEQ ID NO:212 | N/A |
| 29732 | RSDSLSA_ SEQ ID NO:213 | RSDALAR_ SEQ ID NO:214 | RSDDLTR_ SEQ ID NO:182 | QKSNLSS_ SEQ ID NO:215 | DSSDRKK_ SEQ ID NO:216 | N/A |
| 30006 | TSGNLTR_ SEQ ID NO:205 | WWTSRAL_ SEQ ID NO:217 | DRSDLSR_ SEQ ID NO:218 | RSDHLSE_ SEQ ID NO:219 | YSWRLSQ_ SEQ ID NO:220 | N/A |
| 300 | RSDSLSV_ SEQ ID | RNQDRKN_ SEQ ID | QSSDLSR SEQ ID | HKYHLRS_ SEQ ID | QSGDLTR_ SEQ ID | N/A |

| 08 | NO:221 | NO:222 | NO:199 | NO:207 | NO:184 | |
|---|---|---|---|---|---|---|
| 29753 | QSGNLAR_ SEQ ID NO:223 | DRSALAR_ SEQ ID NO:224 | RSDNLST_ SEQ ID NO:225 | AQWGRTS_ SEQ ID NO:226 | N/A | N/A |
| 29754 | RSADLTR_ SEQ ID NO:227 | TNQNRIT_ SEQ ID NO:228 | RSDSLLR_ SEQ ID NO:229 | LQHHLTD_ SEQ ID NO:230 | QNATRIN_ SEQ ID NO:231 | N/A |
| 29769 | QSGNLAR_ SEQ ID NO:223 | DRSALAR_ SEQ ID NO:224 | RSDNLST_ SEQ ID NO:225 | AQWGRTS_ SEQ ID NO:226 | N/A | N/A |
| 29770 | QSGDLTR SEQ ID NO:184 | MRNRLNR_ SEQ ID NO:232 | DRSNLSR_ SEQ ID NO:233 | WRSCRSA SEQ ID NO:234 | RSDNLSV_ SEQ ID NO:186 | N/A |
| 30012 | HSNARKT SEQ ID NO:204 | QSGNLAR SEQ ID NO:223 | DRSALAR SEQ ID NO:224 | RSDNLST SEQ ID NO:225 | AQWGRTS_ SEQ ID NO:226 | N/A |
| 30014 | HSNARKT SEQ ID NO:204 | QSGNLAR SEQ ID NO:223 | DRSALAR SEQ ID NO:224 | RSDHLSQ SEQ ID NO:192 | QWFGRKN_ SEQ ID NO:235 | N/A |
| 30018 | QSGDLTR SEQ ID NO:184 | MRNRLNR SEQ ID NO:232 | DRSNLSR SEQ ID NO:233 | WRSCRSA SEQ ID NO:234 | QRSNLDS_ SEQ ID NO:34 | N/A |
| 29988 | QSGDLTR SEQ ID NO:184 | QWGTRYR SEQ ID NO:33 | DRSNLSR SEQ ID NO:233 | HNSSLKD SEQ ID NO:32 | QSGNLAR_ SEQ ID NO:223 | N/A |
| 29989 | RSDVLSA SEQ ID NO:31 | RNDHRIN SEQ ID NO:30 | RSDHLSQ SEQ ID NO:192 | QSAHRTN SEQ ID NO:29 | DRSNLSR_ SEQ ID NO:233 | DSTNRYR_ SEQ ID NO:28 |

**Table 3: Target site of AHAS zinc fingers**

| ZFP | AHAS Region | Target Site (5'→3') | SEQ ID NO: |
|---|---|---|---|
| 29964 | 500-bp upstream of S653 | ggATAGCAtATTGCGGCGGGAtggcctc | 35 |
| 29965 | 500-bp upstream of S653 | gtACTGGAtGAGCTGaCAAAAGgggagg | 36 |
| 29966 | 500-bp upstream of S653 | gtACCTGGATAGCAtATTGCGgcgggat | 37 |
| 29967 | 500-bp upstream of S653 | agTACCTGgATAGCAtATTGCGgcggga | 38 |
| 29968 | 500-bp upstream of S653 | gaTGAGCTGACAAAAGGggaggcgatca | 39 |
| 29969 | 500-bp upstream of S653 | atGAGCTGaCAAAAGgGGAGGCgatcat | 40 |
| 29970 | 500-bp upstream of S653 | tcATCCAGTACCTGgATAGCAtattgcg | 41 |
| 29971 | 500-bp upstream of S653 | ctGACAAAAGGGGAGGCgatcattgcca | 42 |

| 29730 | Within 30-bp upstream of S653 | agGCAGCACGTGCTCCTGATgcgggact | 43 |
|---|---|---|---|
| 29731 | Within 30-bp upstream of S653 | taGGCAGCACGtgCTCCTGatgcgggac | 44 |
| 29732 | Within 30-bp upstream of S653 | gaTCCCAAGCGGTGGTGctttcaaggac | 45 |
| 30006 | Within 30-bp upstream S653 | tgATGCGGGACTATGATatccaacaagt | 46 |
| 30008 | Within 30-bp upstream S653 | gaGCACGTGCTgCCTATGatcccaagcg | 47 |
| 29753 | Within 80-bp downstream of S653N | tcTTGTAGGTCGAAatttcagtacgagg | 48 |
| 29754 | Within 80-bp downstream of S653N | ctACAAGTGTGaCATGCGcaatcagcat | 49 |
| 29769 | Within 80-bp downstream of S653N | cTTGTAGGTCGAAa | 50 |
| 29770 | Within 80-bp downstream of S653N | cAAGTGTGACaTGCGCAa | 51 |
| 30012 | Within 80-bp downstream of S653N | tcTTGTAGGTCGAAATTtcagtacgagg | 52 |
| 30014 | Within 80-bp downstream of S653N | tcTTGTAGGTCGAAATTtcagtacgagg | 53 |
| 30018 | Within 80-bp downstream of S653N | taCAAgTGTGACaTGCGCAatcagcatg | 54 |
| 29988 | 400-bp downstream of S653 | caGAACCTGACACAGCAgacatgtaaag | 55 |
| 29989 | 400-bp downstream of S653 | atAACGACCGATGGAGGGTGgtcggcag | 56 |

**[0206]** 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 *Fok*I (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.

**[0207]** 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 sites, 13 ZFNs were identified as having *in vivo* activity at high levels, and selected for further experimentation. Eleven of the ZFNs were designed to bind to the three homoeologous gene copies and two ZFNs (29989-2A-29988 and 30006-2A-30008) were designed to only bind the gene copy on chromosome 6D. The 13 ZFNs were characterized as being capable of efficiently binding and cleaving the unique AHAS genomic polynucleotide target sites *in planta*. Exemplary vectors are described below.

**Example 3: Evaluation of zinc finger nuclease cleavage of AHAS genes using transient assays**

***ZFN Construct Assembly***

**[0208]** Plasmid vectors containing ZFN gene expression constructs, which were identified using the yeast assay as described in Example 2, were designed and completed using skills and techniques commonly known in the art. 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 the zinc finger nuclease.

**[0209]** Expression of the fusion proteins was driven by the 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 polyadenylation site) from the *Zea mays* peroxidase (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.

**[0210]** 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, 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 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).

**[0211]** The resulting 13 plasmid constructs: pDAB109350 (ZFNs 29732-2A-29730), pDAB109351 (ZFNs 29732-2A-29731), pDAB109352 (ZFNs 29753-2A-29754), pDAB109353 (ZFNs 29968-2A-29967), pDAB109354 (ZFNs 29965-2A-29964), pDAB109355 (ZFNs 29968-2A-29966), pDAB109356 (ZFNs 29969-2A-29967), pDAB109357 (ZFNs 29971-2A-29970), pDAB109358 (ZFNs 29989-2A-29988), pDAB109359 (ZFNs 30006-2A-30008), pDAB109360 (ZFNs 30012-2A-30018), pDAB109361 (ZFNs 30014-2A-30018) and pDAB109385 (ZFNs 29770-2A-29769) were confirmed via restriction enzyme digestion and via DNA sequencing.

**[0212]** Representative plasmids pDAB109350 and pDAB109360 are shown in Figure 1 and Figure 2.

***Preparation of DNA from ZFN Constructs for Transfection***

**[0213]** 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.

***Isolation of Wheat Mesophyll Protoplasts***

**[0214]** Mesophyll protoplasts from the wheat line cv. Bobwhite MPB26RH were prepared for transfection using polyethylene glycol (PEG)-mediated DNA delivery as follows.

**[0215]** 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 Skooge (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.

**[0216]** About 2-3 grams of leaf material from the germinated seeds was cut into 2-3 cm lengths and placed in a pre-weighed PETRI™ 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, 1 mM $CaCl_2$, 0.1% bovine serum albumin, 0.025% v/v pluronic acid, 5 mM β-mercaptoethanol, pH 5.7) was pipetted into the PETRI™ 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 added to the PETRI™ 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 PETRI™ dish was sealed with PARAFILM® and incubated at 28°C with gentle rotational shaking for 4-5 hours.

**[0217]** 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 PETRI™ dish, swirling gently for 1 min, followed by passing of the 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 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 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***

**[0218]** 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 μl wash buffer containing 70 μg 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 $Ca(NO_3)_2$, 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.

**[0219]** 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 PETRI™ dish and incubated in the dark for 24°C for 72 h.

*Genomic DNA Isolation from Mesophyll Protoplasts*

**[0220]** Transfected protoplasts were transferred from the 3 cm PETRI™ 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 PETRI™ dish was rinsed three times with 1 mL of wash buffer. Each rinse was performed by swirling the wash buffer in the PETRI™ 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 and the protoplast cells were added directly to the lysis buffer.

*PCR Assay of Protoplast Genomic DNA for ZFN Sequence Cleavage*

**[0221]** 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 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 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.

**[0222]** A total of four sets of PCR primers were designed to amplify the ZFN 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 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.

**[0223]** 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*cactctttccctacacgacgctcttccgatctTCCTCTAGGATTCAAGACTTTTG*G | 57 |
| AHAS-500ZFN.R1 | 500-bp upstream of S653 | g*tgactggagttcagacgtgtgctcttccgatctCGTGGCCGCTTGTAAGTGTA*A | 58 |
| AHASs653ZFN.F1 | Within 30-bp upstream of S653 | a*cactctttccctacacgacgctcttccgatctGAGACCCCAGGGCCATACTT*G | 59 |
| AHASs653ZFN.R3 | Within 30-bp upstream of S653 | g*tgactggagttcagacgtgtgctcttccgatctCAAGCAAACTAGAAAACGCATG*G | 60 |
| AHASs653ZFN.F5 | Within 80-bp downstream of S653N | a*cactctttccctacacgacgctcttccgatctATGGAGGGTGATGGCAGGA*C | 61 |
| AHASs653ZFN.R1 | Within 80-bp downstream of S653N | g*tgactggagttcagacgtgtgctcttccgatctATGACAGCACATCCCTACAAAAG*A | 62 |
| AHAS+400ZFN.F1 | 400-bp downstream of S653 | a*cactctttccctacacgacgctcttccgatctAACAGTGTGCTGGTTCCTTTCT*G | 63 |
| AHAS+400ZFN.R3 | 400-bp downstream of S653 | g*tgactggagttcagacgtgtgctcttccgatctTYTYYCCTCCCAACTGTATTCAG*A | 64 |

asterisk (*) is used to indicate a phosphorothioate

[0224] 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 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 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 together and 200 μl of pooled product per ZFN was purified using a QIAQUICK MINIELUTE PCR PURIFICATION KIT™ (Qiagen) following the manufacturer's instructions.

[0225] To enable the sample to be sequenced using Illumina short read 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 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 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***

**[0226]** 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.

**[0227]** 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, 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.

**[0228]** Following sample assignment, a quality filter was passed across all sequences. The quality filter was implemented in custom developed PERL script. 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.

**[0229]** Next, the quality trimmed sequences were attributed to the wheat sub-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.

**[0230]** 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.

**[0231]** 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.

**[0232]** 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.

**[0233]** 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.

**Example 4: Evaluation of Donor Designs for ZFN-mediated AHAS Gene Editing Using Transient Assays**

**[0234]** 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 homologous recombination (HR)-directed and 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 HR-Directed DNA Repair***

**[0235]** Donor DNA designs were based on a plasmid DNA vector containing 750-bp homology arms (*i.e.* sequence identical to the endogenous AHAS gene) flanking each side of the target cleavage site for ZFNs 29732 and 29730. A plasmid DNA vector was designed for each of the three wheat sub-genomes: pDAS000132 (Figure 3), pDAS000133 (Figure 4) and pDAS000134 (Figure 5) were designed to the A-, B- and D-genome, respectively (SEQ ID NO:65, SEQ ID NO:66, and SEQ ID NO:67). Each plasmid DNA vector was designed to introduce an S653N (AGC→ATT) mutation as a genomic modification conferring tolerance to imidazolinone class herbicides at the target homoeologous copy of the endogenous AHAS gene by ZFN-mediated HR-directed DNA repair. Two additional plasmid DNA constructs were also designed to target the D-genome. The first plasmid DNA, pDAS000135 (SEQ ID NO: 68) (Figure 6), was identical to pDAS000134 except that it contained two additional (synonymous) single nucleotide point mutations, one each located at 15-bp upstream and downstream of the S653N mutation. The second plasmid DNA, pDAS000131(SEQ ID: 69) (Figure 7), did not contain the S653N mutation, but was designed to introduce an *Eco*RI restriction endonuclease recognition site (*i.e.*, GAATTC) at the double strand DNA break created by target ZFN cleavage in the D-genome copy of the endogenous AHAS gene.

***Donor Designs for NHEJ-Directed DNA Repair***

**[0236]** Two types of donor DNA designs were used for NHEJ-directed DNA repair.

**[0237]** 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:74 and SEQ ID NO:75), 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: 76 and SEQ ID NO:77) 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.

**[0238]** 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:78 and SEQ ID NO:79) (Figure 8), 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:80 and SEQ ID NO:81) (Figure 9), 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

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 and HDR-Directed DNA Repair*

**[0239]** Standard cloning methods commonly known by one skilled in the art 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.

**[0240]** 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 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.

*Isolation of Wheat Protoplasts Derived from Somatic Embryogenic Callus*

**[0241]** 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:

**[0242]** 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 $m^2$ 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 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 PETRI™ dish containing 2-4 medium without TIMENTIN™, with the uncut scutellum oriented upwards. A total of 25 scutella were plated onto each 9 cm PETRI™ 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 TIMENTIN™ and incubated for a further 3 weeks in the dark at 24°C. Sub-culturing of SEC was repeated for a total of three times before being used for protoplast preparation.

**[0243]** About one gram of SEC was chopped into 1-2 mm pieces using a sharp scalpel blade in a 10 cm PETRI™ 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 $CaCl_2$, 0.8 mM $MgSO_4$, 0.7 mM $KH_2PO_4$, 0.6 M Mannitol, pH 5.8) to prevent the callus from dehydrating. Additional callus digest mix was added to the PETRI™ dish to a volume of 10 mL per gram fresh weight of callus and subject to vacuum (20" Hg) pressure for 30 min. The PETRI™ dish was sealed with PARAFILM® and incubated at 28°C with gentle rotational shaking at 30-40 rpm for 4-5 hours.

**[0244]** 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 PETRI™ 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 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, 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 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.

***PEG-Mediated Transfection of SEC Protoplasts***

**[0245]** 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 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 $Ca(NO_3)_2$, 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.

**[0246]** 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 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 was transferred to a sterile 3 cm PETRI™ dish and incubated in the dark for 24°C for 72 h.

***Isolation of Scutella from Immature Zygotic Wheat Embryos***

**[0247]** Scutella of immature zygotic wheat embryos from the donor wheat line cv. Bobwhite MPB26RH were prepared for transfection using biolistics-mediated DNA delivery as follows.

**[0248]** 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 $m^2$ 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 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 PETRI™ dish containing osmotic MS (E3 maltose) medium, with the uncut scutellum oriented upwards. A total of 20 scutella were plated onto each 9 cm PETRI™ 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.

***Transfection of Scutella of Immature Zygotic Wheat Embryos by Biolistic-mediated DNA Delivery***

**[0249]** 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 μL aliquot of the gold particle suspension was transferred to a 1.5 mL microtube containing 5 μg of DNA resuspended in 5 μL of sterile water. Following thorough mixing by vortexing, 50 μL of 2.5 M $CaCl_2$ and 20 μL 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 centrifugation, as described above, and the supernatant discarded. The DNA-coated gold particles were resuspended in 110 μL of 100% ethanol and maintained on ice.

Following a brief vortex, 10 μL of the gold particle solution was placed centrally onto a macro-carrier membrane and allowed to air dry.

**[0250]** The PDS-1000/HE PARTICLE GUN DELIVERY SYSTEM™ (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 900 psi rupture disc. Each PETRI™ 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.

***Genomic DNA Isolation from SEC Protoplasts***

**[0251]** Genomic DNA was extracted from SEC protoplasts using the 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).

***Genomic DNA Isolation from Scutella of Immature Zygotic Embryos***

**[0252]** 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 MAXI™ KIT (Qiagen) following the manufacturer's instructions.

**[0253]** 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 MAXI™ KIT (Qiagen).

***PCR Assay of Genomic DNA for ZFN-mediated AHAS Editing***

**[0254]** To investigate ZFN-mediated genomic editing at the endogenous AHAS genes in wheat using HR- and 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: 59 and SEQ ID NO: 60) 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 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 HR-directed Editing at AHAS Gene Locus***

**[0255]** 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 HR-directed editing at the target ZFN sites.

**[0256]** To identify sequence reads with molecular evidence for HR-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) were used to identify reads that contained sequence for both the donor DNA molecule used for transfection and the endogenous AHAS locus. To ensure unequivocal discernment between sequence reads arising from ZFN-mediated HR-directed gene editing and those resulting from the carryover of (any) donor DNA used for transfection, molecular evidence for gene editing was declared only if the sequence read also contained a NHEJ deletion at the position of the double strand DNA break created by the ZFN; *i.e.*, the sequence read showed evidence for the outcome of imperfect HR-directed DNA repair. 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 HR-directed gene editing.

**[0257]** From the results of three biological replicates performed for each plasmid donor DNA design, molecular evidence was obtained for the enrichment of sequence reads showing ZFN-mediated HR-directed editing at the three homoeologous copies of the endogenous AHAS genes in wheat (Table 5 and Table 6). Strong molecular evidence was obtained for the addition of an *Eco*RI restriction endonuclease site at the position of the double strand DNA break created by ZFNs 29732 and 29730 in all three homoeologous copies of the endogenous AHAS gene in both samples of SEC protoplasts and scutella of immature zygotic embryos that were transfected with pDAB109350 and pDAS000131. The frequency of ZFN-mediated

HR-directed gene editing was highest in the D-genome to which the donor DNA molecule was targeted. Similarly, strong molecular evidence was obtained for the introduction of donor polynucleotide containing the S653N mutation in all three homoeologous copies of the endogenous AHAS genes in samples of scutella of immature zygotic embryos that were transfected with pDAB109350 and either pDAS000132, pDAS000133 or pDAS000134; strong molecular evidence was also observed for samples of SEC protoplasts transfected with pDAB109350 and pDAS000134. The frequency of ZFN-mediated HR-directed gene editing was again highest in the sub-genome for which the donor DNA was designed. Importantly, the editing frequency in samples of SEC protoplasts and scutella of immature zygotic embryos transfected with pDAB109350 and pDAS000135 was lower (about 10-fold) than that observed for samples transfected with pDAB109350 and pDAS000134. This result was expected due to the penalty imposed on the efficiency for HR-directed DNA repair by the presence of the flanking mutations in the pDAS00135 donor design.

**Table 5: Average HR-directed editing frequency in parts per million (ppm) across three biological replicates of scutella transfected with plasmid donor DNA designs. "na" indicates "not applicable."**

| Donor | Sub-genome targeted | ZFN | Donor-to-ZFN molar ratio | Editing Frequency in Wheat Sub-Genome | Editing Frequency (ppm) |
|---|---|---|---|---|---|
| pDAS000131 | D | n/a | n/a | A | 0 |
| pDAS000131 | D | 29732-2A-29730 | 5:1 | A | 251 |
| pDAS000131 | D | 29732-2A-29730 | 10:1 | A | 46 |
| pDAS000131 | D | n/a | n/a | B | 0 |
| pDAS000131 | D | 29732-2A-29730 | 5:1 | B | 106 |
| pDAS000131 | D | 29732-2A-29730 | 10:1 | B | 19 |
| pDAS000131 | D | n/a | n/a | D | 3 |
| pDAS000131 | D | 29732-2A-29730 | 5:1 | D | 2,577 |
| pDAS000131 | D | 29732-2A-29730 | 10:1 | D | 642 |
| pDAS000132 | A | n/a | n/a | A | 5 |
| pDAS000132 | A | 29732-2A-29730 | 5:1 | A | 2,353 |

| | | | | | |
|---|---|---|---|---|---|
| pDAS000132 | A | 29732-2A-29730 | 10:1 | A | 1,800 |
| pDAS000132 | A | n/a | n/a | B | 0 |
| pDAS000132 | A | 29732-2A-29730 | 5:1 | B | 42 |
| pDAS000132 | A | 29732-2A-29730 | 10:1 | B | 30 |
| pDAS000132 | A | n/a | n/a | D | 0 |
| pDAS000132 | A | 29732-2A-29730 | 5:1 | D | 110 |
| pDAS000132 | A | 29732-2A-29730 | 10:1 | D | 61 |
| pDAS000133 | B | n/a | n/a | A | 0 |
| pDAS000133 | B | 29732-2A-29730 | 5:1 | A | 230 |
| pDAS000133 | B | 29732-2A-29730 | 10:1 | A | 149 |
| pDAS000133 | B | n/a | n/a | B | 8 |
| pDAS000133 | B | 29732-2A-29730 | 5:1 | B | 5,528 |
| pDAS000133 | B | 29732-2A-29730 | 10:1 | B | 4,472 |
| pDAS000133 | B | n/a | n/a | D | 0 |
| pDAS000133 | B | 29732-2A-29730 | 5:1 | D | 0 |
| pDAS000133 | B | 29732-2A-29730 | 10:1 | D | 0 |
| pDAS000134 | D | n/a | n/a | A | 2 |
| pDAS000134 | D | 29732-2A-29730 | 5:1 | A | 316 |
| pDAS000134 | D | 29732-2A-29730 | 10:1 | A | 959 |
| pDAS000134 | D | n/a | n/a | B | 1 |
| pDAS000134 | D | 29732-2A-29730 | 5:1 | B | 110 |
| pDAS000134 | D | 29732-2A-29730 | 10:1 | B | 318 |
| pDAS000134 | D | n/a | n/a | D | 19 |
| pDAS000134 | D | 29732-2A-29730 | 5:1 | D | 4,662 |
| pDAS000134 | D | 29732-2A-29730 | 10:1 | D | 9,043 |
| pDAS000135 | D | n/a | n/a | A | 0 |
| pDAS000135 | D | 29732-2A-29730 | 5:1 | A | 38 |
| pDAS000135 | D | 29732-2A-29730 | 10:1 | A | 97 |

| | | | | | |
|---|---|---|---|---|---|
| pDAS000135 | D | n/a | n/a | B | 0 |
| pDAS000135 | D | 29732-2A-29730 | 5:1 | B | 14 |
| pDAS000135 | D | 29732-2A-29730 | 10:1 | B | 31 |
| pDAS000135 | D | n/a | n/a | D | 1 |
| pDAS000135 | D | 29732-2A-29730 | 5:1 | D | 541 |
| pDAS000135 | D | 29732-2A-29730 | 10:1 | D | 1,191 |

**Table 6: Average HR-directed editing frequency in parts per million (ppm) across three biological replicates of SEC protoplasts transfected with plasmid donor DNA designs. "na" indicates "not applicable."**

| Donor | Sub-genome targeted | ZFN | Donor-to-ZFN molar ratio | Editing Frequency in Wheat Sub-Genome | Editing Frequency (ppm) |
|---|---|---|---|---|---|
| pDAS000131 | D | n/a | n/a | A | 0 |
| pDAS000131 | D | 29732-2A-29730 | 7:1 | A | 50 |
| pDAS000131 | D | n/a | 7:1 | B | 0 |
| pDAS000131 | D | 29732-2A-29730 | 7:1 | B | 0 |
| pDAS000131 | D | n/a | 7:1 | D | 4 |
| pDAS000131 | D | 29732-2A-29730 | 7:1 | D | 212 |
| pDAS000134 | D | n/a | 7:1 | A | 0 |
| pDAS000134 | D | 29732-2A-29730 | 7:1 | A | 0 |
| pDAS000134 | D | n/a | 7:1 | B | 0 |
| pDAS000134 | D | 29732-2A-29730 | 7:1 | B | 0 |
| pDAS000134 | D | n/a | 7:1 | D | 32 |
| pDAS000134 | D | 29732-2A-29730 | 7:1 | D | 258 |
| pDAS000135 | D | n/a | 7:1 | A | 0 |
| pDAS000135 | D | 29732-2A-29730 | 7:1 | A | 0 |
| pDAS000135 | D | n/a | 7:1 | B | 0 |
| pDAS000135 | D | 29732-2A-29730 | 7:1 | B | 0 |
| pDAS000135 | D | n/a | 7:1 | D | 0 |
| pDAS000135 | D | 29732-2A-29730 | 7:1 | D | 1 |

***Data Analysis for Detecting ZFN-mediated NHEJ-directed Editing at AHAS Genes***

**[0258]** 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.

**[0259]** 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.

**[0260]** 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. 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.

**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-29730 | 5:1 | D | 4 |
| pDAS000150 | 29732-2A-29730 | 10:1 | D | 0 |
| pDAS000150 | n/a | n/a | A | 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-29730 | 5:1 | D | 0 |
| pDAS000153 | 29732-2A-29730 | 10:1 | D | 0 |
| pDAS000153 | 29732-2A-29730 | 20:1 | D | 0 |
| pDAS000153 | n/a | n/a | A | 0 |
| pDAS000153 | 29732-2A-29730 | 5:1 | A | 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 |

**[0261]** 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 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 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.

**Example 5: Development of a Transformation System for Producing AHAS Edited Plants**

**[0262]** The endogenous AHAS gene locus in wheat was selected as a model 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 , or ALS inhibitor herbicides such as imidazolinone or sulfonylurea ), knowledge of prerequisite information of sub-genome-specific gene coding sequence, and knowledge of specific mutations conferring tolerance to group B herbicides, or ALS inhibitor 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 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***

**[0263]** *Triticum aestivum cv.* Clearfield Janz, a commercially released bread 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.

**[0264]** 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: 82 and SEQ ID NO: 83) 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.

**[0265]** 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™***

**[0266]** 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.

**[0267]** 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.

**[0268]** 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.

**[0269]** 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 sub-culturing on multiplication medium. Following multiplication, plants for each genotype were transferred to 10 cm PETRI™ 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 PETRI™ dish. A total of 30 plants per genotype were tested for basal 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 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 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.

**[0270]** The basal tolerance of the donor wheat line cv. Bobwhite MPB26RH 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 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 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 PETRI™ dishes containing RM medium supplemented with 0, 100, 200, 300, 400, 500 nM IMAZAMOX®, respectively. Twenty regenerated plants were placed in each PETRI™ dish. A total of 60 regenerated plants were tested for basal tolerance 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.

**[0271]** 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 sub-culturing 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.

**[0272]** 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.

***Design and Synthesis of Donor DNA for ZFN-mediated NHEJ-directed AHAS Gene Editing***

**[0273]** Two types of donor DNA molecule were designed to promote precise ZFN-mediated NHEJ-directed gene editing at the endogenous AHAS genes in wheat. Both 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).

**[0274]** 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:84 and SEQ ID NO:85), 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 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.

**[0275]** 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:86 and SEQ ID NO:87), 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.

**[0276]** 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)***

**[0277]** Standard cloning methods were used in the construction of binary vector pDAS000143 (SEQ ID: 88) (Figure 10). 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 base-pairs 1880 and 1181 mutated from CG to AT in order to induce an amino acid change from serine (S) to aspargine (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 was comprised of the promoter, 5' untranslated region and intron from the actin 1 (*Act*1) 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 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).

**[0278]** 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 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).

***Biolistic-mediated Transformation System for Generating AHAS Edited Wheat Plants***

**[0279]** 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.

**[0280]** Biolistic-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.

**[0281]** 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 sub-cultured once onto fresh callus induction medium and kept in the same conditions for 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 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. $T_1$ seed was harvested from individual plants, following bagging of individual spikes to prevent out-crossing.

**[0282]** 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.

**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 |

**[0283]** 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 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.

**Example 6: Molecular Characterization of Edited Wheat Plants**

**[0284]** 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.

**[0285]** 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 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.

***Genomic DNA Isolation from Regenerated Wheat Plants***

**[0286]** 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 x $10^{-3}$ mBar pressure. The lyophilized 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***

**[0287]** 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: 89, SEQ ID NO: 90 and SEQ ID NO: 91 for forward and reverse primers and probe sequence, respectively) and a region of the Actin (Act1) promoter present on pDAS000143 (SEQ ID NO: 92, SEQ ID NO: 93 and SEQ ID NO: 94 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.

**[0288]** 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***

**[0289]** 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:95 and SEQ ID NO:96) 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 SEQUENCHER 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.

**[0290]** 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. 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 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:97-180 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 | | |
|---|---|---|---|---|---|---|---|---|
| | | Allele 1 | Allele 2 | Allele 1 | Allele 2 | Allele 1 | Allele 2 | SEQ ID NO: |
| Plant No.1 | Status | PE | NHEJ | IE | UE | IE | UE | 97-102 |
| | No. clones[1] | 13 | 20 | 12 | 19 | 14 | 22 | |
| Plant No.2 | Status | NHEJ | UE | UE | nd | IE | UE | 103-108 |
| | No. clones[1] | 9 | 3 | 16 | 0 | 75 | 17 | |
| Plant No.3 | Status | PE | UE | UE | nd | UE | nd | 109-114 |
| | No. clones[1] | 7 | 11 | 29 | 0 | 35 | 0 | |
| Plant No.4 | Status | PE | UE | IE | UE | PE | IE | 115-120 |
| | No. clones[1] | 6 | 11 | 44 | 30 | 6 | 11 | |
| Plant No.5 | Status | PE | UE | NHEJ | UE | UE | nd | 121-126 |
| | No. clones[1] | 10 | 9 | 15 | 26 | 21 | 0 | |
| Plant No.6 | Status | UE | nd | PE | UE | UE | nd | 127-132 |
| | No. clones[1] | 22 | 0 | 11 | 18 | 43 | 0 | |
| Plant No.7 | Status | PE | UE | UE | nd | UE | nd | 133-138 |
| | No. clones[1] | 5 | 12 | 26 | 0 | 22 | 0 | |
| Plant No.8 | Status | UE | nd | UE | nd | UE | nd | 139-144 |
| | No. clones[1] | 32 | 0 | 40 | 0 | 26 | 0 | |
| Plant No.9 | Status | PE | nd | IE | UE | UE | nd | 145-150 |
| | No. clones[1] | 24 | 0 | 13 | 21 | 33 | 0 | |
| Plant No.10 | Status | PE | UE | UE | nd | UE | nd | 151-156 |
| | No. clones[1] | 10 | 19 | 37 | 0 | 29 | 0 | |
| Plant No.11 | Status | UE | nd | UE | nd | PE | UE | 157-162 |
| | No. clones[1] | 35 | 0 | 37 | 0 | 15 | 11 | |
| Plant No.12 | Status | UE | nd | UE | nd | IE | NHEJ | 163-168 |
| | No. clones[1] | 34 | 0 | 40 | 0 | 14 | 8 | |

[1]Number 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.
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.

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

| | | A-genome | | B-genome | | D-genome | | |
|---|---|---|---|---|---|---|---|---|
| | | Allele 1 | Allele 2 | Allele 1 | Allele 2 | Allele 1 | Allele 2 | SEQ ID NO: |
| Plant No.12a | Status | IE | UE | UE | nd | IE | nd | 169-174 |
| | No. clones[1] | 5 | 14 | 53 | 0 | 1 | 24 | |
| Plant No.13a | Status | IE | UE | UE | nd | UE | nd | 175-180 |
| | No. clones[1] | 10 | 12 | 49 | 0 | 18 | 0 | |

[1]Number 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.
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.

**Example 7: Design of Zinc Finger Binding Domains Specific to Region in AHAS Genes Encoding the P197 Amino Acid Residue**

**[0291]** Zinc finger proteins directed against DNA sequence of the homoeologous copies of the AHAS genes were designed as previously described (see also Example 2). Exemplary target sequence and recognition helices are shown in Table 12 (recognition helix regions designs) and Table 13 (target sites). In Table 13, 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 upstream (from 2 to 510 nucleotides upstream) of the region in the AHAS gene encoding the proline 197 (P197) amino acid residue.

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

| ZFP# | F1 | F2 | F3 | F4 | F5 | F6 |
|---|---|---|---|---|---|---|
| 34456 | SEQ ID NO:227 RSADLTR | SEQ ID NO:182 RSDDLTR | SEQ ID NO:182 RSDDLTR | SEQ ID NO:236 RSDALTQ | SEQ ID NO:237 ERGTLAR | SEQ ID NO:182 RSDDLTR |
| 34457 | SEQ ID NO:184 QSGDLTR | SEQ ID NO:238 DTGARLK | SEQ ID NO:182 RSDDLTR | SEQ ID NO:239 HRRSRDQ | SEQ ID NO:240 DRSYRNT | N/A |
| 34470 | SEQ ID NO:241 RSADLSR | SEQ ID NO:242 RSDHLSA | SEQ ID NO:243 QSSDLRR | SEQ ID NO:233 DRSNLSR | SEQ ID NO:244 RSDDRKT | N/A |
| 34471 | SEQ ID NO:184 QSGDLTR | SEQ ID NO:245 RRADRAK | SEQ ID NO:182 RSDDLTR | SEQ ID NO:246 TSSDRKK | SEQ ID NO:227 RSADLTR | SEQ ID NO:247 RNDDRKK |
| 34472 | SEQ ID NO:227 RSADLTR | SEQ ID NO:198 DRSNLTR | SEQ ID NO:237 ERGTLAR | SEQ ID NO:182 RSDDLTR | SEQ ID NO:218 DRSDLSR | SEQ ID NO:248 DSSTRRR |
| 34473 | SEQ ID NO:219 RSDHLSE | SEQ ID NO:249 HSRTRTK | SEQ ID NO:210 RSDTLSE | SEQ ID NO:250 NNRDRTK | SEQ ID NO:237 ERGTLAR | SEQ ID NO:224 DRSALAR |
| 34474 | SEQ ID NO:237 ERGTLAR | SEQ ID NO:182 RSDDLTR | SEQ ID NO:218 DRSDLSR | SEQ ID NO:248 DSSTRRR | SEQ ID NO:198 DRSNLTR | N/A |
| 34475 | SEQ ID NO:249 RSDHLSR | SEQ ID NO:73 QQWDRKQ | SEQ ID NO:201 DRSHLTR | SEQ ID NO:216 DSSDRKK | SEQ ID NO:233 DRSNLSR | SEQ ID NO:251 VSSNLTS |
| 34476 | SEQ ID NO:218 DRSDLSR | SEQ ID NO:248 DSSTRRR | SEQ ID NO:233 DRSNLSR | SEQ ID NO:184 QSGDLTR | SEQ ID NO:198 DRSNLTR | N/A |
| 34477 | SEQ ID NO:237 ERGTLAR | SEQ ID NO:249 RSDHLSR | SEQ ID NO:252 RSDALSV | SEQ ID NO:253 DSSHRTR | SEQ ID NO:216 DSSDRKK | N/A |
| 34478 | SEQ ID NO:254 RSDNLTR | SEQ ID NO:255 RSDNLAR | SEQ ID NO:224 DRSALAR | SEQ ID NO:256 DRSHLSR | SEQ ID NO:205 TSGNLTR | N/A |
| 34479 | SEQ ID NO:252 RSDALSV | SEQ ID NO:253 DSSHRTR | SEQ ID NO:203 RSDNLSE | SEQ ID NO:254 ARTGLRQ | SEQ ID NO:237 ERGTLAR | SEQ ID NO:224 DRSALAR |
| 34480 | SEQ ID NO:255 RSDNLA | SEQ ID NO:224 DRSALAR | SEQ ID NO:256 DRSHLSR | SEQ ID NO:205 TSGNLTR | SEQ ID NO:249 RSDHLSR | SEQ ID NO:257 TSSNRKT |

| | | | | | | |
|---|---|---|---|---|---|---|
| | R | | | | | |
| 34481 | SEQ ID NO:224 DRSALAR | SEQ ID NO:252 RSDALSV | SEQ ID NO:253 DSSHRTR | SEQ ID NO:203 RSDNLSE | SEQ ID NO:254 ARTGLRQ | N/A |
| 34482 | SEQ ID NO:258 RSDDLSK | SEQ ID NO:254 RSDNLTR | SEQ ID NO:221 RSDSLSV | SEQ ID NO:259 RSAHLSR | SEQ ID NO:260 RSDALST | SEQ ID NO:261 DRSTRTK |
| 34483 | SEQ ID NO:216 DSSDRKK | SEQ ID NO:259 RSAHLSR | SEQ ID NO:218 DRSDLSR | SEQ ID NO:219 RSDHLSE | SEQ ID NO:262 TSSDRTK | N/A |

**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: cnGCGGCCATGGCGGCGGCGagggtttg | 263 |
| | | 34457: acCTCcCCCGCCGTCGCAttctcnggcg | 264 |
| pDAB111855 (34470-2A-34471) | 109-bp upstream | 34470: ggCCGGACGCGCGGGCGtanccggacgc | 265 |
| | | 34471: cgTCGGCGTCTGCGTCGCCAcctccggc | 266 |
| pDAB111856 (34472-2A-34473) | 99-bp upstream | 34472: acGCCGACGCGGCCgGACGCGcgggcgt | 267 |
| | | 34473: gcGTCGCCaCCTCCGGCCCGGgggccac | 268 |
| pDAB111857 (34474-2A-34475) | 96-bp upstream | 34474: caGACGCCGACGCGGCCggacgcgcggg | 269 |
| | | 34475: gtCGCCACcTCCGGCCCGGGGgccacca | 270 |
| pDAB111858 (34476-2A-34477) | 90-bp upstream | 34476: gcGACGCAGACGCCGACgcggccggacg | 271 |
| | | 34477: ccTCCGGCCCGGGGGCCaccaacctc | 272 |

| | | | |
|---|---|---|---|
| | | gt | |
| pDAB111859 (34478-2A-34479) | 24-bp upstream | 34478: ggGATGGAGTCGAGGAGngcgtcngcga | 273 |
| | | 34479: tgGTCGCCATCACGGGCCAGgtcccccg | 274 |
| pDAB111860 (34480-2A-34481) | 18-bp upstream | 34480: acCATGGGGATGGAGTCGAGgagngcgt | 275 |
| | | 34481: ccATCACGGGCCAGGTCccccgccgcat | 276 |
| pDAB111861 (34482-2A-34483) | 16-bp upstream | 34482: cgACCATGGGGATGGAGTCGaggagngc | 277 |
| | | 34483: caTCACGGGCCAGGTCCcccgccgcatg | 278 |

[0292] The AHAS zinc finger designs were incorporated into zinc finger expression vectors and verified for cleavage activity using a budding yeast system, as described in Example 2. Of the numerous ZFNs that were designed, produced and tested to bind to the putative AHAS genomic polynucleotide target sites, 14 ZFNs were identified as having *in vivo* activity at high levels, and selected for further experimentation. All 14 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*.

## Example 8: Evaluation of Zinc Finger Nuclease Cleavage of AHAS Genes Using Transient Assays

### *ZFN Construct Assembly*

[0293] Plasmid vectors containing ZFN expression constructs verified for cleavage activity using the yeast system (as described in Example 7) were designed and completed as previously described in Example 3. The resulting 14 plasmid constructs: pDAB111850 (ZFNs 34456-2A-34457), pDAB111851 (ZFNs 34458-2A-34459), pDAB111852 (ZFNs 34460-2A-34461), pDAB111853 (ZFNs 34462-2A-34463), pDAB111854 (ZFNs 34464-2A-34465), pDAB111855 (ZFNs 34470-2A-34471), pDAB111856 (ZFNs 34472-2A-34473), pDAB111857 (ZFNs 34474-2A-34475), pDAB111858 (ZFNs 34476-2A-34477), pDAB111859 (ZFNs 34478-2A-

34479), pDAB111860 (ZFNs 34480-2A-34481), pDAB111861 (ZFNs 34482-2A-34483), pDAB111862 (ZFNs 34484-2A-34485) and pDAB111863 (ZFNs 34486-2A-34487) were confirmed via restriction enzyme digestion and via DNA sequencing.

*Preparation of DNA from ZFN Constructs for Transfection*

**[0294]** 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.

*Isolation and Transfection of Wheat Mesophyll Protoplasts*

**[0295]** Mesophyll protoplasts from the donor wheat line cv. Bobwhite MPB26RH were prepared and transfected using polyethylene glycol (PEG)-mediated DNA delivery as previously described in Example 3.

*PCR Assay of Protoplast Genomic DNA for ZFN Sequence Cleavage*

**[0296]** 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 in Example 3. 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 in Example 3.

**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*cactctttccctacacgacgctcttccgatctTCCCCAATTCCAACCCTCT*C | 279 |
| AHAS-P197ZFN.R1 | Set 1 | g*tgactggagttcagacgtgtgctcttccgatctCGTCAGCGCCTGGTGGATC*T | 280 |
| AHAS-P197ZFN.F5 | Set 2 | a*cactctttccctacacgacgctcttccgatctGCCCGTCCGAGCCCCGCA*A | 281 |
| AHAS-P197ZFN.R1 | Set 2 | g*tgactggagttcagacgtgtgctcttccgatctCGTCAGCGCCTGGTGGATC*T | 282 |
| AHAS- | Set 3 | a*cactctttccctacacgacgctcttccgatctGC | 283 |

| P197ZFN.F7 | | GCTCGCCCGTCATCA*C | |
|---|---|---|---|
| AHAS-P197ZFN.R5 | Set 3 | g*tgactggagttcagacgtgtgctcttccgatctATGGGGATGGAGTCGAGGA*G | 284 |
| AHAS-P197ZFN.F9 | Set 4 | a*cactctttccctacacgacgctcttccgatctCTTCCGCCACGAGCAGG*G | 285 |
| AHAS-P197ZFN.R5 | Set 4 | g*tgactggagttcagacgtgtgctcttccgatctATGGGGATGGAGTCGAGGA*G | 286 |
| AHAS-P197ZFN.F11 | Set 5 | a*cactctttccctacacgacgctcttccgatctTCGTCTCCGCGCTCGCTG*A | 287 |
| AHAS-P197ZFN.R6 | Set 5 | g*tgactggagttcagacgtgtgctcttccgatctTCCACTATGGGCGTCTCCT*G | 288 |

***Data Analysis for Detecting NHEJ at Target ZFN Sites***

**[0297]** Following generation of Illumina short read sequence data for sample libraries prepared for transfected mesophyll protoplasts, bioinformatics analysis (as previously described in Example 3) 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.

**[0298]** 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 |
|---|---|---|---|
| pDAB111850 (34456-2A-34457) | 12,567 | 1,716 | 10,399 |
| pDAB111851 (34458-2A-34459) | 2,088 | 995 | 874 |
| pDAB111852 (34460-2A-34461) | 2 | 2 | 3 |
| pDAB111853 (34462-2A-34463) | 3 | 0 | 3 |
| pDAB111854 (34464-2A-34465) | 47 | 92 | 308 |

| | | | |
|---|---|---|---|
| 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 |
| pDAB111858 (34476-2A-34477) | 48,339 | 56,001 | 44,459 |
| pDAB111859 (34478-2A-34479) | 3,069 | 2,731 | 3,069 |
| pDAB111860 (34480-2A-34481) | 11,790 | 11,946 | 11,790 |
| pDAB111861 (34482-2A-34483) | 28,719 | 33,888 | 28,719 |
| pDAB111862 (34484-2A-34485) | 216 | 111 | 216 |
| pDAB111863 (34486-2A-34487) | 54 | 28 | 54 |

**[0299]** 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.

**Example 9: Artificial Crossing and Molecular Analysis to Recover Plants with Specific Combinations of Precise Genome Modifications**

**[0300]** Wheat events that 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 of the target endogenous locus. As shown previously in Example 6, ZFN-mediated genome modification effectuates 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 in Example 5 that have precisely modified AHAS genes with the S653N mutation can be used to produce wheat plants that have the S653N mutation in either a specific sub-genome, in any combination of multiple sub-genomes, or in all three sub-genomes.

**[0301]** Similarly, self-pollination of transformation events having genome modifications at multiple copies of the target endogenous locus can be subsequently used to produce wheat events that have the S653N mutation at only a specific sub-

genome. Subsequent self-pollination of transformation events is especially useful for removing undesirable genome modifications from an event, such as imperfect editing at one or more copies of the target endogenous locus.

**[0302]** Molecular and phenotypic assays, such as those previously described, can be used to track the inheritance of specific genome modifications in the progeny derived from artificial crossing and self-pollination of transformed events.

***Inheritance and Expression of Precision Genome Modifications in Wheat***

**[0303]** To verify stable expression and inheritance of the AHAS herbicide tolerance phenotype conferred by the S653N mutation carried by the wheat transformation events generated in Example 5, T1 seed from three wheat events were subjected to molecular and phenotypic analysis. The three independent wheat events each carried the integrated S653N mutation in the AHAS gene located within the A-genome.

**[0304]** T1 seed were derived from self-pollination of each T0 event. The seeds were surface sterilized and germinated *in vitro* by sub-culturing the sterilized seeds on multiplication medium, as described previously. After 10 days of growth at 24°C under a 16/8 (light/dark) hour photoperiod, the roots of the germinated seedlings were removed and the seedlings were transferred onto rooting medium containing 200 nM IMAZAMOX® (imidazolinone). The seedlings were incubated for 2-3 weeks under the same conditions and the presence or absence of root re-growth was recorded. Leaf tissue harvested from each seedling was used for DNA extraction, and a PCR assay to test for the presence of the modified AHAS gene using primers AHAS_3F1 and AHAS_3R1 (SEQ ID NO:95 and SEQ ID NO:96) was completed, as described previously. Electrophoretic separation of the resulting PCR products on agarose gel was used to detect the presence of the modified AHAS gene. The amplification of only a 750-bp fragment PCR product indicated the absence of the modified AHAS gene. Comparatively, the amplification of only a 850-bp fragment indicated the presence of the modified AHAS gene in the homozygous state. Furthermore, the amplification of both a 750-bp and 850-bp fragment indicated the presence of the modified AHAS gene in the hemizygous state.

**[0305]** Next, a chi-square test was used to confirm the inheritance of the modified AHAS gene as a single genetic unit. Expected Mendelian inheritance was observed in the T1 generation for each of the three wheat transformation events. The

modified AHAS gene segregated at the 3:1 ratio expected for a PCR test producing a dominant marker (Table 16) in the T1 seedlings. Similarly, IMAZAMOX® tolerance showed 3:1 segregation, as expected for the dominant AHAS herbicide tolerance phenotype conferred by the S653N mutation (Table 17) in the T1 seedlings.

**Table 16: Segregation of modified AHAS gene in T1 seedlings derived from self-pollination of transformed wheat plants from Example 5.**

| Event | No. of T1 plants | No. of T1 plants with exogenous sequence | No. of T1 plants without exogenous sequence | Segregation ratio tested | P-value |
|---|---|---|---|---|---|
| mb1k-7783-1-1 | 25 | 19 | 6 | 3:1 | $p < 0.05$ |
| yr00-7794-1-1 | 54 | 44 | 10 | 3:1 | $p < 0.05$ |
| yt02-7786-1-1 | 33 | 27 | 6 | 3:1 | $p < 0.05$ |

**Table 17: Segregation of IMAZAMOX® tolerance phenotype in T1 seedlings derived from self-pollination of transformed wheat plants from Example 5.**

| Event | No. of T1 plants | No. of T1 plants IMI tolerance | No. of T1 plants without IMI tolerance | Segregation ratio tested | P-value |
|---|---|---|---|---|---|
| mb1k-7783-1-1 | 25 | 19 | 6 | 3:1 | $p < 0.05$ |
| yr00-7794-1-1 | 54 | 44 | 10 | 3:1 | $p < 0.05$ |
| yt02-7786-1-1 | 33 | 27 | 6 | 3:1 | $p < 0.05$ |

**[0306]** The stability of expression of the modified AHAS gene was verified by its correspondence with the AHAS herbicide tolerance phenotype. Complete concordance was observed between the presence of one or more copies of the modified AHAS gene and IMAZAMOX® tolerance.

***Self-pollination and artificial crossing to recover plants with specific combinations of precise genome modifications***

**[0307]** Artificial crossing between wheat transformation events produced in Example 5 can be used to generate wheat plants that have the S653N mutation on a specific sub-genome, on multiple sub-genomes, or on all three sub-genomes.

[0308] To generate homozygous wheat plants having the S653N mutation on a specific sub-genome, three wheat events from Example 5 were allowed to self-pollinate and produce T1 seed. The three events; mb1k-7783-1-1, yw06-7762-2-1 and yw06-7834-1-1 were selected to have hemizygous AHAS genome modifications on the A-genome, B-genome and D-genome, respectively. About 15 T1 seed from each event were germinated and grown under glasshouse containment conditions to produce T2 seed. Leaf material harvested from each T1 plant was used for DNA extraction and PCR assays were completed to determine the zygosity of the modified AHAS gene. This PCR zygosity test was designed to amplify a fragment from each of the three homoeologous copies of the endogenous AHAS gene within a region containing the binding site for ZFNs 29732 and 29730 (encoded on plasmid pDAB190350), and to include genomic nucleotide sequence variation. Enough genomic nucleotide sequence variation was included to differentiate between the AHAS homoeologs, such that the resulting amplicons could be unequivocally attributed (at the sequence level) to the wheat sub-genome from which they were derived. The primer pairs were synthesized with the Illumina™ SP1 and SP2 sequences at the 5' end to provide compatibility with Illumina™ sequencing-by-synthesis chemistry. The synthesized primers also contained a phosphorothioate linkage at the penultimate 5' and 3' nucleotides. The 5' phosphorothioate linkage afforded protection against exonuclease degradation of the Illumina™ SP1 and SP2 sequences. Likewise, the 3' phosphorothioate linkage improved PCR specificity for amplification of the target AHAS sequences using on-off PCR (Yang *et al.*, (2005) *Biochem. Biophys. Res. Commun.*, Mar. 4:328(1):265-72). The sequences of the primer pairs are provided in Table 18.

**Table 18: Primer sequences used to assess the zygosity of the modified AHAS gene in transgenic wheat events from Example 5.**

| Primer Name | Primer Sequence (5'→3') | SEQ ID NO: |
|---|---|---|
| AHASs653ZFN.F2 | a*cactctttccctacacgacgctcttccgatctGCAATCAAGAAGATGCTTGAGAC*C | 297 |
| AHASs653ZFN.R1 | g*tgactggagttcagacgtgtgctcttccgatctTCTTTTGTAGGGATGTGCTGTCA*T | 298 |

The asterisk(*) indicates a phosphorothioate; lowercase font indicates SP1 and SP2 sequences, and uppercase font indicates the genomic DNA sequence.

**[0309]** The resulting PCR amplicons were prepared for deep sequencing as described previously, and sequenced on an Illumina MiSEQ™ instrument to generate 250-bp paired-end sequence reads, according to the manufacturer's instructions. The resultant sequence reads were computationally processed, as described previously, to assign each read to sample (based on the barcode index) and the sub-genome from which they were derived (based on nucleotide variation that distinguished between homoeologous copies of the AHAS gene), and to perform quality filtering to ensure that only high quality sequences were used for subsequent analyses. Custom developed PERL scripts and manual data manipulation in MICROSOFT EXCEL 2010™ (Microsoft Corporation) were used to process the data and determine the zygosity of the modified AHAS gene in each T1 wheat event.

**[0310]** As the integration of pDAS000267 into the endogenous AHAS locus resulted in only a 95-bp size difference between the wild-type (unmodified) and resulting transgenic (modified) allele, the PCR zygosity assay was expected to amplify both the wild-type and modified AHAS gene. Consequently, T1 plants, homozygous for the target genome modification, were expected to produce only sequence reads that originate from the amplification of the transgenic allele at the modified AHAS locus. These alleles were distinguishable at the sequence level by the six mutations deliberately introduced into the AHAS exon in pDAS000267 (*e.g.*, the two mutations encoding the S653N mutation, and the four codon-optimized, synonymous mutations positioned across the binding site of ZFN 29732 prevented re-cleavage of the integrated donor). The T1 plants hemizygous for the target genome modification were expected to produce sequence reads originating from both the wild-type and transgenic allele at the modified AHAS locus. Whereas, T1 plants without the modified AHAS gene were expected to only produce sequence reads originating from the wild-type allele at the modified AHAS locus. Based on the PCR zygosity test, T1 plants homozygous for the S653N mutation in only the A-genome, B-genome, or D-genome were identified (Table 19).

**Table 19: PCR zygosity assay results for T1 plants derived from self-pollination of transgenic wheat events from Example 5.**

| Event | T1 plant | A-genome | | B-genome | | D-genome | | Genotype[3] |
|---|---|---|---|---|---|---|---|---|
| | | No. of WT reads[1] | No. of ED reads[2] | No. of WT reads | No. of ED reads | No. of WT reads | No. of ED reads | |
| mb1k-7783-1 | mb1k-7783-1-29 | 39,305 | 46,481 | 92,167 | 2,011 | 85,048 | 2,222 | AaBBDD |
| mb1k-7783-1 | mb1k-7783-1-31 | 95,696 | 61,451 | 203,228 | 3,913 | 200,232 | 4,087 | AaBBDD |
| mb1k-7783-1 | mb1k-7783-1-33 | 32,608 | 27,270 | 67,551 | 1,440 | 70,588 | 1,632 | AaBBDD |
| mb1k-7783-1 | mb1k-7783-1-39 | 37,172 | 56,416 | 76,005 | 1,693 | 77,899 | 1,787 | AaBBDD |
| mb1k-7783-1 | mb1k-7783-1-41 | 31,782 | 37,945 | 74,540 | 1,478 | 76,916 | 1,892 | AaBBDD |
| mb1k-7783-1 | mb1k-7783-1-43 | 3,784 | 93,125 | 189,570 | 4,164 | 160,769 | 3,931 | aaBBDD |
| mb1k-7783-1 | mb1k-7783-1-46 | 208,627 | 4,902 | 241,948 | 4,567 | 247,912 | 5,094 | AABBDD |
| mb1k-7783-1 | mb1k-7783-1-47 | 66,472 | 39,215 | 134,076 | 2,464 | 126,823 | 2,613 | AaBBDD |
| mb1k-7783-1 | mb1k-7783-1-49 | 83,048 | 1,906 | 85,267 | 1,586 | 87,773 | 1,794 | AABBDD |
| mb1k-7783-1 | mb1k-7783-1-53 | 41,810 | 34,455 | 81,446 | 1,603 | 82,871 | 1,776 | AaBBDD |
| mb1k-7783-1 | mb1k-7783-1-55 | 73,129 | 48,692 | 164,791 | 3,233 | 155,375 | 3,205 | AaBBDD |
| mb1k-7783-1 | mb1k-7783-1-57 | 2,971 | 119,900 | 97,509 | 2,161 | 96,476 | 2,563 | aaBBDD |
| mb1k-7783-1 | mb1k-7783-1-58 | 2,076 | 60,517 | 62,638 | 1,444 | 59,721 | 1,827 | aaBBDD |
| mb1k-7783-1 | mb1k-7783-1-59 | 1,777 | 78,101 | 56,566 | 1,239 | 55,302 | 1,326 | aaBBDD |
| mb1k-7783-1 | mb1k-7783-1-61 | 64,093 | 57,599 | 135,703 | 2,713 | 132,205 | 2,863 | AaBBDD |
| yw06-7762-2 | yw06-7762-2-23 | 13,123 | 374 | 21,286 | 532 | 21,471 | 560 | AABBDD |
| yw06-7762-2 | yw06-7762-2-24 | 56,120 | 1,382 | 87,745 | 1,635 | 82,753 | 2,170 | AABBDD |
| yw06-7762-2 | yw06-7762-2-25 | 39,091 | 1,053 | 1,525 | 38,594 | 61,284 | 1,578 | AAbbDD |
| yw06-7762-2 | yw06-7762-2-27 | 24,551 | 804 | 1,428 | 19,364 | 37,500 | 1,184 | AAbbDD |
| yw06-7762-2 | yw06-7762-2-28 | 44,494 | 1,234 | 32,935 | 18,811 | 64,736 | 1,733 | AABbDD |
| yw06-7762-2 | yw06-7762-2-29 | 33,554 | 964 | 22,898 | 11,718 | 45,887 | 1,221 | AABbDD |
| yw06-7762-2 | yw06-7762-2-30 | 33,410 | 1,011 | 1,481 | 26,659 | 46,214 | 1,430 | AAbbDD |
| yw06-7762-2 | yw06-7762-2-31 | 56,639 | 1,516 | 44,649 | 17,155 | 85,830 | 2,116 | AABbDD |
| yw06- | yw06- | 45,753 | 1,223 | 35,723 | 13,649 | 69,858 | 1,781 | AABbDD |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 7762-2 | 7762-2-32 | | | | | | | |
| yw06-7762-2 | yw06-7762-2-33 | 12,239 | 306 | 17,611 | 333 | 18,324 | 498 | AABBDD |
| yw06-7762-2 | yw06-7762-2-34 | 38,709 | 1,001 | 32,109 | 14,549 | 61,150 | 1,620 | AABbDD |
| yw06-7762-2 | yw06-7762-2-35 | 48,185 | 1,329 | 40,719 | 16,138 | 75,876 | 1,953 | AABBDD |
| yw06-7762-2 | yw06-7762-2-36 | 44,420 | 1,096 | 71,463 | 1,374 | 72,604 | 1,721 | AABBDD |
| yw06-7762-2 | yw06-7762-2-37 | 23,752 | 685 | 37,126 | 796 | 36,283 | 941 | AABBDD |
| yw06-7834-1 | yw06-7834-1-28 | 43,467 | 1,092 | 68,043 | 1,317 | 65,748 | 1,677 | AABBDD |
| yw06-7834-1 | yw06-7834-1-29 | 47,463 | 1,177 | 72,531 | 1,390 | 38,007 | 14,387 | AABBDd |
| yw06-7834-1 | yw06-7834-1-31 | 51,138 | 1,484 | 77,266 | 1,797 | 1,770 | 27,955 | AABBdd |
| yw06-7834-1 | yw06-7834-1-32 | 42,666 | 1,336 | 70,422 | 1,578 | 38,234 | 17,932 | AABBDd |
| yw06-7834-1 | yw06-7834-1-33 | 33,075 | 907 | 55,545 | 1,331 | 28,610 | 10,916 | AABBDd |
| yw06-7834-1 | yw06-7834-1-34 | 47,971 | 1,277 | 78,765 | 1,671 | 1,536 | 29,627 | AABBdd |
| yw06-7834-1 | yw06-7834-1-35 | 44,355 | 1,043 | 74,365 | 1,347 | 68,161 | 1,634 | AABBDD |
| yw06-7834-1 | yw06-7834-1-36 | 67,661 | 1,788 | 93,068 | 2,329 | 2,214 | 31,935 | AABBdd |
| yw06-7834-1 | yw06-7834-1-37 | 33,663 | 826 | 49,051 | 973 | 52,989 | 1,274 | AABBDD |
| yw06-7834-1 | yw06-7834-1-38 | 45,974 | 1,080 | 67,706 | 1,258 | 67,774 | 1,619 | AABBDD |
| yw06-7834-1 | yw06-7834-1-39 | 2,687 | 27,436 | 88,976 | 2,084 | 92,612 | 2,892 | AABBDD |
| yw06-7834-1 | yw06-7834-1-40 | 62,142 | 1,713 | 93,532 | 2,233 | 49,886 | 21,129 | AABBDd |
| yw06-7834-1 | yw06-7834-1-41 | 50,781 | 1,381 | 77,168 | 1,696 | 37,412 | 14,167 | AABBDd |
| yw06-7834-1 | yw06-7834-1-42 | 44,020 | 1,233 | 61,262 | 1,517 | 1,374 | 27,505 | AABBdd |
| yw06-7834-1 | yw06-7834-1-43 | 68,958 | 1,456 | 48,972 | 1,009 | 91,624 | 2,062 | AABBDD |

[1]Number of sequence reads originating from the specified sub-genome and having the sequence haplotype corresponding to the wild-type (unmodified) AHAS locus. The usage of "WT" indicates wild-type.

[2]Number of sequence reads originating from the specified sub-genome and having the sequence haplotype corresponding to the transgenic (modified) AHAS locus. The usage of "ED" indicates edited.

[3]Genotype for the T1 plant, where uppercase and lowercase letters indicate the presence of the wild-type and transgenic AHAS loci on the specified sub-genome, respectively. For example, AaBBDD indicates the T1 plant has a hemizygous AHAS genome modification on the A-genome and homozygous wild-type AHAS loci on the B- and D-genomes. The zygosity at each of the three endogenous AHAS loci is determined from the frequency of the sequence reads corresponding to the wild-type

and modified alleles originating from each sub-genome. Hemizygous genotypes have a similar frequency of wild-type and modified alleles originating from an endogenous AHAS locus, where homozygous genotypes reveal predominantly wild-type or modified alleles. The low frequency of alternate alleles originating from homozygous AHAS loci is due to PCR chimerism between reads originating from different sub-genomes.

**[0311]** One skilled in the art can deploy subsequent rounds of artificial crossing between different wheat transformation events, in combination with the described PCR zygosity test, to produce homozygous wheat plants having the S653N mutation on any combination of multiple sub-genomes (*e.g.*, the A-genome and B-genome, the A-genome and D-genome, or the B-genome and D-genome), or on all three sub-genomes. For example, artificial crossing of T1 plant *mb1k-7783-1-43* (*i.e.*, aaBBDD genotype) with T1 plant *yw06-7762-2-25* (*i.e.*, AAbbDD genotype) would produce T2 seed that are hemizygous for modified AHAS genes in the A-genome and B-genomes; *i.e.*, with the AaBbDD genotype. Subsequent, growth and self-pollination of T2 plants would produce T3 seed segregating for homozygous genotypes for the modified AHAS genes on the A- and B-genomes (*i.e.*, aabbDD genotype), which can be identified using the described PCR zygosity assay.

**Example 10: Development of a Transformation System for Sequential, Exogenous Transgene Stacking at the Endogenous AHAS Loci in Wheat**

**[0312]** The endogenous AHAS gene locus in wheat was selected as a model locus to develop a ZFN-mediated, exogenous 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 of one or more transgenes) at precisely the same genomic location. In addition, the transformation system includes simultaneous parallel or sequential stacking at multiple alleles across multiple sub-genomes. The strategies exploit incorporating mutations in the AHAS gene that confer tolerance to Group B herbicides (*e.g.*, ALS inhibitors such as imidazolinone or sulfonylurea).

**[0313]** ZFN-mediated integration of a donor DNA into the wild-type (herbicide susceptible) AHAS locus was used to introduce transgene(s) and a mutation to the endogenous AHAS gene that conferred tolerance to imidazolinones, thus allowing the

regeneration of correctly targeted plants that possess tolerance to an imidazolinone selection agent.

**[0314]** Stacking of a second transgene(s) at the AHAS locus is achieved by integration of a 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.

**[0315]** Stacking of a third transgene is achieved by integration of a donor molecule that introduces further transgene(s) and confers susceptibility to sulfonylurea and tolerance to imidazolinones, thus allowing the regeneration of correctly targeted plants using an imidazolinone selection agent.

**[0316]** 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 gene for differential cycling between imidazolinone and sulfonylurea selection agents. The transgenes can be integrated within the AHAS gene and stacked via an NHEJ and/or HDR pathway. The desired repair and recombination pathway can be determined by the design of the donor transgene. In an embodiment, exogenous sequences that are integrated and stacked within the AHAS gene would be designed to contain a 5' and 3' region of homology to the genomic integration site; *i.e.* the AHAS gene. The 5' and 3' region of homology would flank the payload (*e.g.*, AHAS mutation and gene of interest). Accordingly, such a design would utilize an HDR pathway for the integration and stacking of the donor polynucleotide within the chromosome. In a subsequent 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 would utilize an NHEJ pathway for the integration and stacking of the donor polynucleotide within the chromosome.

***Design and Production of Donor DNA for First Sequential Transgene Stacking at an Endogenous AHAS Locus Using NHEJ-Directed DNA Repair***

**[0317]** The donor DNA for the first round of transgene stacking was designed to promote precise donor integration at an endogenous AHAS locus via ZFN-mediated, NHEJ-directed repair. The design was 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, Figure 1).

**[0318]** The donor molecule backbone of pDAS000433 (SEQ ID NO:71; Figure 12) comprised several polynucleotide sequence features. The 5' end contained sequence that was nearly identical to the endogenous AHAS gene encoded in the D-genome. This sequence was made up of a fragment that spanned from the target ZFN cleavage site and finished at the AHAS stop codon. In addition, seven deliberate mutations were introduced into the sequence: the two mutations that encoded the S653N mutation and the five codon-optimized, synonymous mutations positioned across the binding site of ZFN 29732. The five codon-optimised, synonymous mutations were included to prevent re-cleavage of the integrated donor. Next, the stop codon was followed by 316-bp of non-coding sequence corresponding to the conserved 3'untranslated region (3'UTR) across the AHAS homoeologs. In addition, the 3'UTR sequence was 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 second round of transgene stacking. The self-excision Zinc Finger binding sites were followed by two additional Zinc Finger binding sites, which were flanked by 100-bp of random sequence. These two additional Zinc-Finger binding sites were immediately followed by a pair of unique restriction endonuclease cleavage sites that were used to insert the transgene expression cassette (*i.e.*, the PAT expression cassette, as described below). Following the two unique restriction endonuclease sites were two more Zinc Finger binding sites, which were again flanked by 100-bp of random sequence. The inclusion of the four additional Zinc Finger binding sites enable future excision of transgenes integrated at an AHAS locus by sequential marker-free transgene stacking, or continued sequential transgene stacking at the same genomic location using an alternate stacking method.

**[0319]** The donor backbone cassette was synthesized by a commercial gene service vendor (GeneArt, Life Technologies) 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 were compatible with the ligation overhangs generated by ZFNs 29732 and 29730 (encoded on plasmid pDAB109350) upon cleavage of an endogenous AHAS locus.

**[0320]** The PAT expression cassette was inserted, using standard methods known to a person skilled in the art, into the donor backbone cassette of pDAS000433 between the two unique restriction endonuclease sites to produce the donor molecule cassette "QA_pDAS000434" (SEQ ID NO:314; Figure 19). The PAT selection cassette was comprised of the promoter, 5' untranslated region, and intron from the Actin (*Act*1) 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 comprising phosphinothricin, glufosinate, and bialaphos (Wohlleben *et al.,* (1988) *Gene*, 70(1): 25-37). This cassette was terminated with the 3' UTR comprising the transcriptional terminator and polyadenylation sites from the 35s gene of cauliflower mosaic virus (CaMV) (Chenault *et al.,* (1993) *Plant Physiology* 101 (4): 1395-1396). Plasmid DNA for "QA_pDAS000434" was prepared using the PURE YIELD PLASMID MAXIPREP SYSTEM™ (Promega Corporation, WI) following the manufacturer's instructions.

**[0321]** PCR amplification of "QA_pDAS000434" followed by digestion with restriction endonuclease *Bbs*I was used to produce linear double-stranded DNA donor molecules with protruding 5' and 3' ends that were compatible with the ligation overhangs generated by ZFNs 29732 and 29730 (encoded on plasmid pDAB109350) upon cleavage of an endogenous AHAS locus. PCR amplification was performed with primers AHAS_TSdnr1_F1 and AHAS_TSdnr1_R1 (SEQ ID NO: 297 and 298, respectively), which were designed to the short stretch of additional sequence added to the 5' and 3' ends of the donor backbone cassette "QA_pDAS000434". The resulting amplicons were purified using the Agencourt AMPure™ XP-PCR purification kit (Beckman Coulter) and digested with *Bbs*I (New England Biolabs). The amplicons were purified a second time using the Agencourt AMPure™ XP-PCR purification kit (Beckman Coulter), followed by ethanol precipitation and resuspension in sterile water at a DNA concentration appropriate for wheat transformation. Standard methods known to a person skilled in the art were used to prepare the linear double-stranded DNA donor molecule.

***Production of Control Binary Vector Encoding AHAS (S653N)***

**[0322]** A binary vector pDAS000143 (SEQ ID NO:88, Figure 10) containing AHAS(S653N) expression and PAT selection cassettes was designed and assembled using skills and techniques commonly known in the art as previously described. Plasmid DNA for the binary was prepared using the PURE YIELD PLASMID MAXIPREP SYSTEM™ (Promega Corporation, WI) following the manufacturer's instructions. The binary vector pDAS000143 was transformed into wheat cells as a control.

***Biolistics-Mediated Transformation for Generating Wheat Events with First Sequential Transgene Stack at an Endogenous AHAS Locus Using NHEJ-Directed DNA Repair***

**[0323]** A total of 55,468 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 with the formulations as described above. For transfections performed using the linear double-stranded donor DNA derived from "QA_pDAS000434" or pDAS000433, the donor DNA was mixed at a 5:1 molar ratio with plasmid DNA for pDAB109350 (encoding ZFNs 29732 and 29730). Transfections performed using pDAS000143 were performed using gold particles that were coated only with plasmid DNA for pDAS000143.

**[0324]** Biolistic-mediated transfections were performed as described previously. Following bombardment, the transfected scutella were incubated at 26°C in the dark for 16 h before being transferred onto medium for callus induction.

**[0325]** Two different chemical selection strategies were used to enrich for regenerated wheat plants with an integrated linear double-stranded donor molecule. The first strategy based on IMAZAMOX® was used to recover wheat events that had the donor molecule precisely integrated into one or more homoeologous copies of the endogenous AHAS gene by ZFN-mediated NHEJ-directed gene editing. Such events are expected to have the AHAS herbicide tolerance phenotype conferred by the S653N mutation. The second strategy based on BASTA® (DL-Phosphinothricin) was used to recover events that had the donor molecule integrated at either a random (non-targeted) position in the wheat genome, or imperfectly integrated into one or more homoeologous copies of the endogenous AHAS gene by ZFN-mediated NHEJ-

directed gene editing. These events are expected to exhibit the BASTA® herbicide tolerance phenotype conferred by the PAT gene, but not necessarily the AHAS herbicide tolerance phenotype conferred by the S653N mutation. The purpose of the second chemical selection strategy was to allow the frequency of precise (on-target) versus random (off-target) donor integration to be quantified, as well as the frequency of perfect and imperfect integration at the endogenous AHAS loci. The two chemical selection strategies are described in Table 20.

**Table 20: Chemical selection strategies used to regenerate wheat plants that had an integrated donor molecule ("IMI" indicates IMAZAMOX® and "PPT" indicates BASTA® selection).**

| Plant Regeneration Stage | IMI Selection | PPT Selection |
|---|---|---|
| Callus Induction (CIM) | 150 nM | None |
| Plant Regeneration (DRM) | 150 nM | 5 mg/ml PPT |
| Rooting (RM) | 200 nM | 5 mg/ml PPT |

**[0326]** A total of 34,546 and 23,550 transfected scutella were subject to IMAZAMOX® and BASTA® selection, respectively. For each strategy, scutella were cultured in the dark on callus induction medium at 24°C for 2 weeks. The resultant calli were sub-cultured once onto fresh callus induction medium and kept in the same conditions for a further two weeks. Somatic embryogenic callus (SEC) were 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. For IMAZAMOX® selection, the regenerated plants were sub-cultured for a total of three times on rooting media. At the end of each round, the roots of regenerated plants were removed and the plants were again sub-cultured on rooting media under the same conditions. Plantlets with roots were transferred to soil and grown under glasshouse containment conditions. $T_1$ seed was harvested from individual plants, following bagging of individual spikes to prevent out-crossing.

**[0327]** The scutella explants bombarded with gold particles coated with pDAS000143 were used to monitor the selection stringency across both the IMAZAMOX® and BASTA® chemical selection strategies. Plants transformed with pDAS000143 were regenerated using the process described above.

**[0328]** A total of 36 wheat plants were recovered from each chemical selection strategy for scutella explants transfected with pDAS000143. Molecular testing of these events using the hydrolysis probe assay described in Example 6 confirmed that all of the recovered wheat plants carried at least one randomly integrated copy of the pDAS000143 insert. These results indicated that the IMAZAMOX® and BASTA® selection conditions were sufficiently stringent to ensure a low escape rate (*i.e.*, recovery of wheat plants that were not transformed), whilst allowing the recovery of events carrying one or more integrated copies of the AHAS (S653N) and PAT donor polynucleotides, respectively.

**[0329]** No wheat plants having the AHAS herbicide tolerance phenotype conferred by the S653N mutation were recovered from IMAZAMOX® selection under the specific selection conditions described above. As IMAZAMOX® selection is expected only to recover wheat plants that have precise integration of the donor molecule into one or more copies of the homoeologous AHAS gene, these results suggest that the chemical selection regime was sub-optimal, and that the conditions should be modified for precise ZFN-mediated NHEJ-directed integration of pDAS000433 donor at an endogenous AHAS locus, or that the scale of transformation was not appropriate for the chemical selection conditions used in the current work. In contrast, 1,652 wheat plants were recovered from BASTA® selection. As BASTA® is expected to recover wheat plants that have both targeted and non-targeted (random) donor integration, molecular characterization of these events can distinguish between targeted and non-targeted donor integration, which can provide guidance for refining IMAZAMOX® selection conditions.

***Molecular Characterisation of BASTA®-Selected Wheat Plants for Evidence of First Transgene Stacking at an Endogenous AHAS Locus***

**[0330]** A total of 1,162 wheat plants recovered from BASTA®-selection were molecularly characterized to assess the frequency of targeted and off-target (random) donor integration, as well as the frequency of targeted perfect and imperfect donor integration at the endogenous AHAS loci.

**[0331]** Three molecular assays were performed for each wheat plant using genomic DNA extracted with the DNEASY® PLANT DNA EXTRACTION MINI KIT™ (Qiagen) from freeze-dried leaf tissue, as described previously.

**[0332]** The first molecular test was used to confirm that the regenerated wheat plants carried at least one integrated copy of the linear double-strand DNA derived from "QA_pDAS000434". This test involved a PCR assay to amplify a region of the Actin (Act1) promoter present in "QA_pDAS000434" (SEQ ID NOs: 92 and 93 for forward and reverse primers, respectively), followed by electrophoretic separation of the resulting amplicon on an agarose gel. The presence of a PCR fragment of expected size (218-bp) indicated integration of at least one copy of the donor molecule. Of the 1,162 wheat events, 1,065 (92%) produced a PCR fragment of the expected size.

**[0333]** The second molecular test was used to identify wheat plants having the donor molecule putatively integrated into one or more copies of the endogenous AHAS locus. This test comprised an on-off PCR assay using a primer designed to hybridize to a region upstream of the binding site for ZFNs 29732 and 29730 (encoded on plasmid pDAB190350) in each of the homoeologous copies of the endogenous AHAS gene, and a primer designed to hybridize to a region within the 100-bp of random sequence flanking the binding site for ZFNs 34480 and 34481 (encoded on plasmid pDAB111860) in "QA_pDAS000434" (SEQ ID NO: 299 and 300 for forward and reverse primers, respectively). Each primer was designed with a phosphorothioate linkage positioned at the penultimate base to maximize specificity for primer extension during PCR amplification. Amplification of a PCR fragment with size greater than 300-bp when separated by electrophoresis on agarose gel was considered as suggestive evidence for targeted integration (of least a portion) of the donor molecule into one or more copies of the endogenous AHAS gene. Of the 1,065 wheat events tested, 543 (51%) amplified a PCR fragment of greater than 300-bp in size.

**[0334]** The third molecular assay was used to further characterize wheat plants showing suggestive evidence for targeted integration of the donor molecule in one or more copies of the endogenous AHAS gene. This test involved a PCR assay using a pair of primers designed to amplify a 256-bp region from the three homoeologous copies of the endogenous AHAS gene. This region contained the binding site for ZFNs 29732 and 29730 (encoded on plasmid pDAB190350), and to include genomic nucleotide sequence variation. Enough genomic nucleotide sequence variation was included to differentiate between the AHAS homoeologs, such that the resulting amplicons could be unequivocally attributed (at the sequence level) to the wheat sub-

genome from which they were derived. The primer pairs were synthesized with the Illumina™ SP1 and SP2 sequences at the 5' end, respectively, to provide compatibility with Illumina™ sequencing-by-synthesis chemistry. The synthesized primers also contained a phosphorothioate linkage at the penultimate 5' and 3' nucleotides. The 5' phosphorothioate linkage 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. These sequences of the primer pair are given in Table 21.

**Table 21: Primer sequences used to further characterize wheat plants having suggestive evidence for targeted integration of the donor molecule in one or more copies of the endogenous AHAS gene.**

| Primer Name | Primer Sequence (5'→3') | SEQ ID NO: |
|---|---|---|
| AHASs653ZFN.F2 | a*cactctttccctacacgacgctcttccgatctGCAATCA AGAAGATGCTTGAGAC*C | 301 |
| AHASs653ZFN.R3 | g*tgactggagttcagacgtgtgctcttccgatctCAAGCA AACTAGAAAACGCATG*G | 302 |

The asterisk(*) indicates a phosphorothioate; lowercase font indicates SP1 and SP2 sequences, and upper case font indicates the genomic DNA sequence.

**[0335]** PCR amplicons produced by the third molecular assay were prepared for deep sequencing by performing an additional round of PCR 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. Following amplification, the generated products were sequenced on an Illumina MiSEQ™ instrument to generate 250-bp paired-end sequence reads, according to the manufacturer's instructions.

**[0336]** The resultant paired-end 250-bp sequence reads were computationally processed, as described previously, to assign each read to sample (based on the barcode index) and the sub-genome from which they were derived (based on nucleotide variation that distinguished between homoeologous copies of the AHAS gene), and to perform quality filtering to ensure that only high quality sequences were used for subsequent analyses. Custom developed PERL scripts and manual data

manipulation in MICROSOFT EXCEL 2010™ (Microsoft Corporation) were used, as described below, to identify reads that contained evidence for targeted integration of the donor into one or more copies of the endogenous AHAS gene.

**[0337]** As the hybridization site for primer AHASs653ZFN.R3 (Table 21) was also present in the AHAS 3' untranslated region (UTR) in "QA_pDAS000434", the third molecular assay allowed for differentiation between targeted and random donor integration, as well as between perfect and imperfect donor integration at one or more copies of the endogenous AHAS locus. Wheat plants having perfect hemizygous on-target editing are expected to produce sequence reads that originate from amplification of both the wild-type (unedited) and edited alleles at each modified AHAS locus. These alleles are distinguishable at the sequence level by the seven deliberate mutations introduced into the AHAS exon in "QA_pDAS000434" (*i.e.*, the two mutations encoding the S653N mutation and the five codon-optimized, synonymous mutations positioned across the binding site of ZFN 29732, which were incorporated to prevent re-cleavage of the integrated donor). Theoretically, the frequency of reads corresponding to the wild-type and edited alleles should occur at a ratio of 1:1 for each endogenous AHAS locus with perfect hemizygous editing. In contrast, wheat plants having perfect homozygous on-target editing are expected to only generate sequence reads that originate from the pair of edited alleles at each modified endogenous AHAS locus. As the primer pair used in the third molecular assay were designed to amplify all three homoeologous copies of the AHAS gene, the expected generation of reads originating from all three wheat sub-genomes can also be used to detect on-target imperfect donor integration (*e.g.*, integration of a partial donor fragment, or integration of the donor fragment in the wrong orientation). Imperfect on-target donor integration is expected to result in amplification of only the wild-type (unedited) allele from each modified endogenous AHAS locus due to PCR competition favoring the amplification of the shorter wild-type fragment. Consequently, hemizygous on-target imperfect donor integration is expected to generate about half as many reads originating from the sub-genome into which integration occurred, compared to unedited sub-genomes. For homozygous on-target imperfect donor integration, no reads are expected to originate from the sub-genome into which integration occurred. Conversely, off-target (random) donor integration is expected to generate an equal proportion of sequence reads originating from all three homoeologous copies of the AHAS gene.

**[0338]** Sequence analysis of the 543 wheat plants tested revealed 38 events with molecular evidence for on-target donor integration in one or more copies of the endogenous AHAS gene. Event di01-9632-1-1 had perfect hemizygous donor integration in the AHAS locus situated in the B-genome. These results were indicated by the presence of both wild-type and perfectly edited reads originating from the B-genome, and only wild-type alleles originating from the A- and D-genome (Table 22). Two events had imperfect hemizygous donor integration in the AHAS loci on the A- and D-genomes, respectively. Event yl02-9453-1-2 had both wild-type and imperfectly edited reads originating from the D-genome, and only wild-type alleles originating from the A- and B- genomes. Comparatively, event yl02-9552-21-1 had both wild-type and imperfectly edited reads originating from the A-genome, and only wild-type alleles originating from the other sub-genomes.

**[0339]** The remaining 35 events showed molecular evidence for imperfect donor integration into at least one copy of the endogenous AHAS gene, where the donor molecule was likely to be truncated or integrated in the wrong orientation (Table 22). These events were characterized by a lower than expected frequency of reads originating from one or more of the wheat sub-genomes. For example, event yl02-9552-7-1 had a statistically significant lower frequency of wild-type AHAS reads originating from the B-genome than expected for an unedited locus. The remaining 453 events showed only evidence for random integration of the donor elsewhere in the wheat genome, indicating that the amplified product from the second molecular assay most likely arose from PCR chimerism. The consensus sequences for the edited alleles present in the B, D and A sub-genome of wheat events di01-9632-1-1, yl02-9453-1-2 and yl02-9552-21-1 are provided as SEQ ID NOs:303, 304 and 305, respectively.

**Table 22: Molecular evidence for integration of QA_pDAS000434 into one or more homoeologous copies of the endogenous AHAS locus.**

| | A-genome | | | | |
|---|---|---|---|---|---|
| Event | No. of reads | % Reads | % WT | % PE | % IE |
| di01-9632-1-1 | 17,312 | 9 | 99 | 0 | 1 |
| yl02-9453-1-2 | 8,548 | 20 | 97 | 3 | 0 |
| yl02-9552-21-1 | 3,049 | 10 | 47 | 0 | 53 |
| yl02-9552-7-1 | 43,845 | 66 | 100 | 0 | 0 |
| gt19-9595-10-1 | 48,681 | 62 | 100 | 0 | 0 |
| yr00-9553-3-1 | 16,212 | 16 | 98 | 1 | 0 |
| yr00-9580-9-1 | 69,153 | 35 | 97 | 2 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| yl02-9532-1-1 | 85,431 | 43 | 100 | 0 | 0 |
| yl02-9532-16-1 | 14,318 | 29 | 100 | 0 | 0 |
| di01-9603-10-1 | 825 | 1* | 100 | 0 | 0 |
| yl02-9578-1-1 | 1,662 | 1* | 100 | 0 | 0 |
| di01-9603-2-1 | 833 | 5* | 100 | 0 | 0 |
| yc06-9547-1-1 | 831 | 1* | 100 | 0 | 0 |
| yl02-9532-9-1 | 2,168 | 1* | 100 | 0 | 0 |
| yc06-9522-1-1 | 4,233 | 2* | 100 | 0 | 0 |
| mb1k-9539-31-1 | 2,355 | 2* | 100 | 0 | 0 |
| yl02-9503-1-1 | 1,381 | 1* | 100 | 0 | 0 |
| mb1k-9546-4-1 | 1,971 | 2* | 100 | 0 | 0 |
| di01-9603-18-1 | 1,436 | 1* | 100 | 0 | 0 |
| di01-9603-25-1 | 819 | 1* | 100 | 0 | 0 |
| yl02-9503-2-1 | 1,241 | 1* | 100 | 0 | 0 |
| di01-9550-14-1 | 2,846 | 2* | 100 | 0 | 0 |
| yr00-9580-28-1 | 708 | 0* | 100 | 0 | 0 |
| yl02-9552-19-1 | 4,127 | 2* | 100 | 0 | 0 |
| hw12-9569-5-1 | 1,959 | 1* | 100 | 0 | 0 |
| gt19-9582-2-1 | 244 | 0* | 99 | 0 | 1 |
| gt19-9593-6-1 | 9,426 | 7* | 100 | 0 | 0 |
| mb1k-9539-25-1 | 982 | 1* | 100 | 0 | 0 |
| yl02-9457-7-1 | 467 | 0* | 100 | 0 | 0 |
| yr00-9553-16-1 | 433 | 0* | 100 | 0 | 0 |
| yw06-9345-15-1 | 146 | 4* | 100 | 0 | 0 |
| mb1k-9546-2-1 | 93,058 | 97 | 99 | 0 | 1 |
| yr00-9541-5-1 | 131,675 | 93 | 100 | 0 | 0 |
| yl02-9552-47-1 | 180,989 | 97 | 100 | 0 | 0 |
| gt19-9551-4-1 | 144,978 | 99 | 100 | 0 | 0 |
| yc06-9340-5-1 | 96,105 | 98 | 100 | 0 | 0 |
| yc06-9584-2-1 | 98,385 | 98 | 100 | 0 | 0 |
| yr00-9541-1-1 | 115,671 | 98 | 100 | 0 | 0 |
| | B-genome | | | | |
| Event | No. of reads | % Reads | % WT | % PE | % IE |
| di01-9632-1-1 | 9,498 | 5 | 70 | 29 | 1 |
| yl02-9453-1-2 | 13,374 | 32 | 97 | 3 | 0 |
| yl02-9552-21-1 | 16,817 | 55 | 100 | 0 | 0 |
| yl02-9552-7-1 | 6,254 | 9* | 100 | 0 | 0 |
| gt19-9595-10-1 | 5,146 | 7* | 100 | 0 | 0 |
| yr00-9553-3-1 | 8,683 | 8* | 100 | 0 | 0 |
| yr00-9580-9-1 | 1,768 | 1* | 98 | 1 | 1 |
| yl02-9532-1-1 | 6,644 | 3* | 100 | 0 | 0 |
| yl02-9532-16-1 | 34,310 | 70 | 100 | 0 | 0 |
| di01-9603-10-1 | 3,228 | 4* | 100 | 0 | 0 |
| yl02-9578-1-1 | 2,176 | 1* | 100 | 0 | 0 |

| | | | | | |
|---|---|---|---|---|---|
| di01-9603-2-1 | 1,225 | 7* | 100 | 0 | 0 |
| yc06-9547-1-1 | 723 | 1* | 100 | 0 | 0 |
| yl02-9532-9-1 | 1,012 | 0* | 100 | 0 | 0 |
| yc06-9522-1-1 | 3,979 | 2* | 100 | 0 | 0 |
| mb1k-9539-31-1 | 2,359 | 2* | 100 | 0 | 0 |
| yl02-9503-1-1 | 601 | 0* | 100 | 0 | 0 |
| mb1k-9546-4-1 | 364 | 0* | 100 | 0 | 0 |
| di01-9603-18-1 | 106,322 | 96 | 100 | 0 | 0 |
| di01-9603-25-1 | 101,834 | 98 | 100 | 0 | 0 |
| yl02-9503-2-1 | 221,040 | 99 | 100 | 0 | 0 |
| di01-9550-14-1 | 130,434 | 96 | 100 | 0 | 0 |
| yr00-9580-28-1 | 174,074 | 99 | 100 | 0 | 0 |
| yl02-9552-19-1 | 174,186 | 95 | 100 | 0 | 0 |
| hw12-9569-5-1 | 260,971 | 98 | 100 | 0 | 0 |
| gt19-9582-2-1 | 67,764 | 99 | 100 | 0 | 0 |
| gt19-9593-6-1 | 110,669 | 84 | 100 | 0 | 0 |
| mb1k-9539-25-1 | 75,915 | 96 | 100 | 0 | 0 |
| yl02-9457-7-1 | 125,465 | 99 | 100 | 0 | 0 |
| yr00-9553-16-1 | 111,825 | 99 | 100 | 0 | 0 |
| yw06-9345-15-1 | 3,655 | 93 | 100 | 0 | 0 |
| mb1k-9546-2-1 | 1,448 | 2* | 100 | 0 | 0 |
| yr00-9541-5-1 | 4,403 | 3* | 100 | 0 | 0 |
| yl02-9552-47-1 | 2,236 | 1* | 100 | 0 | 0 |
| gt19-9551-4-1 | 740 | 1* | 100 | 0 | 0 |
| yc06-9340-5-1 | 620 | 1* | 100 | 0 | 0 |
| yc06-9584-2-1 | 617 | 1* | 100 | 0 | 0 |
| yr00-9541-1-1 | 781 | 1* | 100 | 0 | 0 |
| | D-genome | | | | |
| Event | No. of reads | % Reads | % WT | % PE | % IE |
| di01-9632-1-1 | 170,321 | 86 | 99 | 0 | 1 |
| yl02-9453-1-2 | 19,841 | 48 | 68 | 32 | 0 |
| yl02-9552-21-1 | 10,665 | 35 | 100 | 0 | 0 |
| yl02-9552-7-1 | 15,936 | 24 | 100 | 0 | 0 |
| gt19-9595-10-1 | 24,091 | 31 | 100 | 0 | 0 |
| yr00-9553-3-1 | 79,529 | 76 | 98 | 1 | 0 |
| yr00-9580-9-1 | 128,317 | 64 | 97 | 2 | 1 |
| yl02-9532-1-1 | 105,821 | 53 | 100 | 0 | 0 |
| yl02-9532-16-1 | 434 | 1* | 99 | 1 | 0 |
| di01-9603-10-1 | 84,718 | 95 | 100 | 0 | 0 |
| yl02-9578-1-1 | 152,767 | 98 | 100 | 0 | 0 |
| di01-9603-2-1 | 14,671 | 88 | 100 | 0 | 0 |
| yc06-9547-1-1 | 71,423 | 98 | 100 | 0 | 0 |
| yl02-9532-9-1 | 230,632 | 99 | 100 | 0 | 0 |
| yc06-9522-1-1 | 167,492 | 95 | 100 | 0 | 0 |

| | | | | | |
|---|---|---|---|---|---|
| mb1k-9539-31-1 | 142,061 | 97 | 100 | 0 | 0 |
| yl02-9503-1-1 | 199,717 | 99 | 100 | 0 | 0 |
| mb1k-9546-4-1 | 89,309 | 97 | 100 | 0 | 0 |
| di01-9603-18-1 | 2,921 | 3* | 100 | 0 | 0 |
| di01-9603-25-1 | 1,715 | 2* | 96 | 0 | 4 |
| yl02-9503-2-1 | 1,741 | 1* | 100 | 0 | 0 |
| di01-9550-14-1 | 3,140 | 2* | 100 | 0 | 0 |
| yr00-9580-28-1 | 1,012 | 1* | 100 | 0 | 0 |
| yl02-9552-19-1 | 5,470 | 3* | 100 | 0 | 0 |
| hw12-9569-5-1 | 2,479 | 1* | 100 | 0 | 0 |
| gt19-9582-2-1 | 496 | 1* | 99 | 0 | 1 |
| gt19-9593-6-1 | 11,821 | 9* | 100 | 0 | 0 |
| mb1k-9539-25-1 | 1,898 | 2* | 100 | 0 | 0 |
| yl02-9457-7-1 | 555 | 0* | 100 | 0 | 0 |
| yr00-9553-16-1 | 604 | 1* | 100 | 0 | 0 |
| yw06-9345-15-1 | 150 | 4* | 100 | 0 | 0 |
| mb1k-9546-2-1 | 1,191 | 1* | 100 | 0 | 0 |
| yr00-9541-5-1 | 4,766 | 3* | 100 | 0 | 0 |
| yl02-9552-47-1 | 3,537 | 2* | 100 | 0 | 0 |
| gt19-9551-4-1 | 1,171 | 1* | 99 | 0 | 0 |
| yc06-9340-5-1 | 1,186 | 1* | 100 | 0 | 0 |
| yc06-9584-2-1 | 1,234 | 1* | 100 | 0 | 0 |
| yr00-9541-1-1 | 1,566 | 1* | 100 | 0 | 0 |

"No. of reads" indicates the number of sequence reads assigned to the wheat sub-genome; "% Reads" indicates the percentage of sequence reads assigned to the wheat sub-genome as a proportion of all assigned reads; "% WT" indicates the percentage of sequence reads identified as wild type (unedited) alleles; "% PE" indicates the
5 percentage of sequence reads indicating precise donor integration into the wheat sub-genome; "% IE" indicates the percentage of sequence reads indicating imperfect donor integration into the wheat sub-genome; Asterisks(*) indicate occurrence of statistically significant fewer sequence reads than expected for an unedited endogenous AHAS locus
10

**[0340]** Overall, 3% (38/1,162) of the BASTA®-selected wheat events showed molecular evidence for targeted donor integration into one or more of the homoeologous copies of the endogenous AHAS gene.

15 **Example 11: Development of a Transformation System for Sequential, Exogenous Marker-Free Transgene Stacking at the Endogenous AHAS Loci in Wheat**

**[0341]** Wheat plants containing a donor integrated polynucleotide within the AHAS locus to introduce the S653N mutation are produced via the previously

described methods. For example, the regeneration of event di01-9632-1-1 (Table 23) showing molecular evidence of perfect hemizygous integration of "QA_pDAS000434" in the B-genome of wheat indicates that donor DNA and zinc finger nuclease constructs can be utilized for the integration of donor molecule sequences at one or more copies of the target endogenous AHAS locus within wheat. Producing such an event, that is free of any additional transgenic selectable markers, is the initiating act for sequential, exogenous transgenic selectable marker-free stacking of a donor polynucleotide at an endogenous AHAS locus in the genome of wheat. The edited plant events are obtained via alternative selection conditions as previously described in Example 10.

**[0342]** The previously described selection conditions can be modified by a number of methodologies. Other approaches can be implemented to enhance the recovery of wheat plants with precise integration of the S653N mutation (as encoded on "QA_pDAS000434" or pDAS000433) into one or more copies of the endogenous AHAS locus, without using a transgenic selectable marker.

**[0343]** Two additional approaches can be implemented to enhance the recovery of wheat plants with precise integration of the S653N mutation into one or more copies of the endogenous AHAS locus, without the usage of a transgenic selectable marker.

**[0344]** For example, IMAZAMOX® selection conditions are modified, to include selection at differing stages of culturing and/or lower concentrations of the herbicide. Accordingly, selection at the plant regeneration stage is reduced by lowering the concentration of IMAZAMOX® added to the plant regeneration media or as another alternative the usage of herbicide at this plant regeneration stage is completely eliminated. As such, stronger growth of regenerated plantlets is observed, thereby ensuring larger plantlets that are less susceptible to tissue damage when sub-cultured to rooting media. Furthermore, the plantlets may be required to be dissected from the embryogenic callus from which they originate. Smaller plantlets are more susceptible to tissue damage, which can result in tissue necrosis and potential loss of transformed plantlets during sub-culturing. Maintenance of IMAZAMOX® selection at the callus induction stage helps to restrict embryogenesis from untransformed cells, while its maintenance at the rooting stage would provide strong selection for plantlets with precise integration of pDAS000433 at one or more copies of the endogenous AHAS locus, which is required to produce the AHAS herbicide tolerance phenotypes

conferred by the S653N mutation. The success of such IMAZAMOX® selection strategies for generating precisely edited wheat plants was demonstrated in Example 5.

**[0345]** In another example, a different transformation system is used to generate wheat plants with precisely integrated donor DNA. For example, protoplast-based transformation could be used to produce individual calli, where each callus is derived from a single cell. Protoplast-derived calli provide several advantages over callus derived from biolistic-bombarded scutella of immature zygotic embryos. Unlike callus derived from biolistics-bombardment, and which is chimeric for both transformed and untransformed cells, protoplast-derived callus is clonal. Hence, cell survival in callus derived from a transformed protoplast in which precise pDAS000433 integration has occurred cannot be compromised by the presence of neighboring untransformed cells when subject to IMAZAMOX® selection. In the case of callus derived from biolistics-bombardment, the chimeric composition of the callus means that the survival of a precisely transformed can be compromised by the death of surrounding untransformed cells when subjected to IMAZAMOX® selection. Protoplast-based transformation systems also provide the advantage of scalability, compared to biolistics bombardment, since many more cells can be transformed for an given amount of effort, thereby providing for higher probability for recovering wheat plants with precise integration of pDAS000433 in one or more copies of the endogenous AHAS gene. Several protoplast-based transformation systems for wheat have been described in published scientific literature (Qiao *et al.* (1992) *Plant Cell Reports* 11:262-265; Ahmed and Sagi (1993) *Plant Cell Reports* 12:175-179; Pauk *et al.*(1994) *Plant Cell, Tissue and Organ Culture* 38: 1-10; He *et al.* (1994) *Plant Cell Reports* 14: 92-196; Gu and Lang (1997) *Plant Cell, Tissue and Organ Culture* 50: 139–145; and Li *et al.* (1999) *Plant Cell, Tissue and Organ Culture* 58: 119–125).

**[0346]** A series of experiments are performed to determine optimal selection conditions for regenerating wheat plants expressing the AHAS(S653N) mutation conferring tolerance to IMAZAMOX® from a protoplast-based transformation system such as those described above.

**[0347]** IMAZAMOX® selection conditions are optimized using protoplasts derived from somatic embryogenic callus (SEC)-derived cell suspension culture of the wheat line cv. Bobwhite MPB26RH. While protoplasts derived from Bobwhite

MPB26RH are non-totipotent (*i.e.*, cannot be used to regenerate entire plants), the selection conditions established for enriching for events expressing the AHAS(653N) mutation are expected to be transferrable to any protoplast-based transformation system based on a totipotent wheat genotype, which those in the art would recognize. The experiments conducted establish the basal tolerance of the wild-type donor wheat line cv. Bobwhite MPB26RH (S653/S653 genotype, which confers susceptibility to imidazolinones) to IMAZAMOX®. The use of IMAZAMOX® selection conditions stronger than basal tolerance will strongly enrich for transformed cells expressing the AHAS(S653N) mutation.

**[0348]** Further transformation methods are applicable. For example a cell suspension culture for wheat line cv. Bobwhite MPB26RH can be established. Somatic embryogenic callus (SEC) is induced from immature zygotic embryos of wheat line cv. Bobwhite MPB26RH as described previously. A fast growing callus line is selected after six cycles of sub-culturing on callus induction media. For each cycle of sub-culturing, the fast-growing calli are transferred onto new callus induction media and cultured in the dark at 26°C for 14 d.

**[0349]** A cell suspension culture is initiated by transferring 1gram calli of the fast-growing callus line to a flask containing 20 ml liquid growth medium and culturing at 25°C in the dark on a gyratory shaker at 90 rpm. Every seven days the cell suspension culture is sub-cultured by passing the culture through a fine gauze to remove cell clumps greater than 2 mm in diameter, and replacing two thirds of the culture media with fresh medium. After 3 months of repeated filtration and sub-culturing a fast-growing SEC-derived cell suspension culture is established.

**[0350]** Next protoplasts are isolated from the SEC-derived cell suspension culture. About 4 grams fresh weight of cell clumps are obtained by passing 7 day old SEC-derived cell suspension culture through a fine-mesh. The cell clumps are digested in wheat callus digest mix, as described previously, to release the protoplasts. The yield of SEC-derived cell suspension culture protoplasts is estimated using a Neubauer™ haemocytometer. Evans Blue stain is used to determine the proportion of live cells recovered.

**[0351]** The protoplast culture selection conditions with the herbicide IMAZAMOX® are selected. An agarose bead-type culture system is used for protoplast culture. About 1 x $10^6$ protoplasts are precipitated by gentle centrifugation and the supernatant is removed. The protoplasts are resuspended by gentle agitation in

1 ml of melted 1.2% Sea-Plaque™ agarose cooled to 40°C and transferred to a 3.5 cm petri dish. Following agarose solidification, 1 ml culture medium is added to the petri dish and the plate is incubated at 25°C in the dark for 1 week. The agarose plug is transferred into a 20 cm petri dish containing 10 ml culture medium and incubated at 25°C in the dark on a gyratory shaker at 90 rpm. Every 14 days the culture medium is replaced with fresh media. Protoplast cell division is typically observed 3 days after embedding in agarose, with clumps of multiple cells visible after 7 days.

**[0352]** The basal tolerance of wheat line cv. Bobwhite MPB26RH to IMAZAMOX® is determined by incubating the agarose bead-type cultures in media supplemented with 0, 50, 100, 200, 400 and 600 nM IMAZAMOX® and assessing the rate of calli growth after 2 weeks. IMAZAMOX® concentrations higher than 200 nM impede calli development, indicating that concentrations of 200 nM and higher are optimal for enriching and selecting wheat cells having the AHAS (S653N) mutation.

**[0353]** Establishment of tissue culture selection conditions for obtaining transgenic plants with a donor integrated fragment resulting in the S653N mutation within the AHAS locus are obtained. The edited plant events are used to generate explant material (*e.g.*, protoplasts or scutella of immature zygotic embryos) for a second round of transfection. As described in the next example, the explant material is subsequently co-transfected with a donor DNA molecule and a plasmid encoding a ZFN 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.

**Example 12: Alternate Transformation Systems for Sequential, Exogenous Marker-Free Transgene Stacking at the Endogenous AHAS Loci in Wheat**

**[0354]** Molecular evidence provided in Example 10 for the regenerated wheat plant event di01-9632-1-1 demonstrates the technical feasibility for sequential, exogenous marker-free transgene stacking at the endogenous AHAS loci in wheat. Refinement of IMAZAMOX® selection conditions or use of a different transformation system permit the production of wheat plants with sequentially stacked transgenes at an endogenous AHAS locus. This example describes approaches for achieving exogenous marker-free sequential transgene stacking at an endogenous AHAS locus by alternating between different selective agents (*e.g.*, imidiazolinone

and sulfonylurea) and corresponding AHAS mutations (*e.g.*, S653N and P197S). First, the selection conditions for sulfonylurea were determined.

***Optimization of Chemical Selection Conditions; Generation of Low-Copy, Randomly Integrated T-DNA Wheat Plants with AHAS(P197S) Expression Constructs***

**[0355]** A binary vector pDAS000164 (SEQ ID NO:289, Figure 11) containing 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 Physiology*, 100: 1503-07) followed by the coding sequence (1,935bp) 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 the transcriptional terminator and polyadenylation site 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 (*Act*1) 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 comprising phosphinothricin, glufosinate, and bialaphos (Wohlleben *et al.,* (1988) *Gene*, 70(1): 25-37). This cassette was terminated with the 3' UTR comprising the transcriptional terminator and polyadenylation sites from the 35s gene of the cauliflower mosaic virus (CaMV) (Chenault *et al.,* (1993) *Plant Physiology,* 101 (4): 1395-1396).

**[0356]** 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 (*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 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 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 SEQUENCHER™ software (Gene Codes Corporation, Ann Arbor, MI).

**[0357]** 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 $T_0$ transgenic events expressing the AHAS (P197) expression constructs were selected for phosphinothricin (PPT) tolerance, the phenotype conferred by the PAT transgenic selectable marker, and transferred to soil. The $T_0$ plants were grown under glasshouse containment conditions and $T_1$ seed was produced.

**[0358]** Genomic DNA from each $T_0$ plant was extracted from leaf tissue, using the protocols as previously described in Example 6, and tested for the presence or absence of carryover *Agrobacterium tumefaciens* strain and for the number of integrated copies of the T-DNA encoding AHAS(P197S). The presence or absence of the *A. tumefaciens* strain was performed using a duplex hydrolysis probe qPCR assay (analogous to TAQMAN™) to amplify the endogenous ubiquitin gene (SEQ ID NO:290, SEQ ID NO:291, and SEQ ID NO:292 for forward and reverse primers and probe sequence, respectively) from the wheat genome, and virC from pTiBo542 (SEQ ID NO: 293, SEQ ID NO:294, and SEQ ID NO:70 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 in Example 6, based on the *puroindoline-b* gene (Pinb) from the D genome of hexaploid wheat and a region of the Actin (Act1) promoter present on pDAS000164. Overall, 35

independent $T_0$ events with fewer than three randomly integrated copies of T-DNA were generated.

***Optimization of Chemical Selection Conditions; Conditions for Regenerating Wheat Plants on Sulfometuron Methyl***

**[0359]** 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/P197genotype, 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 (P197S) mutation, which confers tolerance to sulfonylurea selection agents.

**[0360]** 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 in Example 4,and placed in 10 cm petri dishes containing CIM medium supplemented with 0, 100, 500, 1000, 1500 and 2000 nM sulfometuron methyl, respectively. Twenty scutella were placed in each petri 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 formation for cv. Bobwhite MPB26RH was reduced by about 70% at 100 nM sulfometuron methyl, compared to untreated samples.

**[0361]** 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 petri dishes containing CIM medium. Then SEC 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, 500, 1000, 1500, 2000, 2500 and 3000 nM sulfometuron methyl, respectively. Twenty CIM were placed in each petri 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.

**[0362]** 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 petri dishes containing CIM medium. Then SEC 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 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 petri 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 petri 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.

**[0363]** The basal tolerance of transgenic wheat events with randomly integrated, low-copy (≤ 3) T-DNA expressing the AHAS (P197S) mutation to sulfometuron methyl from pDAS000164 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 were transferred to 10 cm petri 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 petri dish. A total of 3 plants per event were tested for basal tolerance 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 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.

***Design and Synthesis of Donor DNA for First Sequential Transgene Stacking at an Endogenous AHAS Locus Using NHEJ-Directed DNA Repair***

**[0364]** The donor DNA of the pDAS000433 construct (Figure 12) for the first round of transgene stacking is designed and synthesized as described in Examples 10 and 11 to promote precise donor integration (containing the S653N mutation) at an endogenous AHAS locus via ZFN-mediated, NHEJ-directed repair. Whole plants that are resistant to IMAZAMOX® are obtained and prepared for a second round of targeting to introduce the

***Design and Synthesis of Donor DNA for Second Sequential Transgene Stack at an Endogenous AHAS Locus Using NHEJ-Directed DNA Repair***

**[0365]** The donor DNA (pDAS000434; Figure 13; SEQ ID NO:72) containing a P197S mutation 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 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 pDAS000434 donor molecule comprises several portions of polynucleotide sequences. The 5' end contains sequence nearly 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 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 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 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 unique restriction endonuclease cleavage

sites, and which enable insertion of a transgene expression cassette (*e.g.*, the DGT-28 expression cassette, as described in U.S. Pat. Pub. No. 20130205440). The additional Zinc Finger binding sites enable future excision of transgenes that can be integrated at an AHAS locus by sequential marker-free 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 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.

**[0366]** The donor molecule with protruding 5' and 3' ends is generated by digesting plasmid DNA containing the donor molecule, or following PCR amplification as described for "QA_pDAS000434" and/or pDAS000433, with the restriction endonuclease *BbsI* using standard methods known to one in the art.

***Transformation System for Exogenous Marker-free, Sequential Transgene Stacking at an Endogenous AHAS Locus in Wheat Using NHEJ-Directed DNA Repair***

**[0367]** 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 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 correctly targeted plants using IMAZAMOX® as a selection agent, as previously described in Example 5. Figure 14a 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 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 susceptibility to imidazolinones (Figure 14b). 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. 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.

**[0368]** The transformation system used to regenerate wheat plants with sequentially stacked transgenes at an endogenous AHAS locus is based on the previously described approach for biolistics-mediated DNA delivery to scutella of immature zygotic wheat embryos, or direct DNA delivery to wheat protoplasts using approaches known to one skilled in the art; for example, using the method of He *et al.* (1994) *Plant Cell Reports* 14: 92-196, or any of the methods described in Example 11.

***Design and Synthesis of Donor DNA for First Sequential Transgene Stacking at an Endogenous AHAS Locus Using HDR-Directed DNA Repair***

**[0369]** 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, HDR-directed homology 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; Figure 16; SEQ ID NO:295) is identical in sequence to pDAS000433 (Figure 12).

**[0370]** The donor cassette is synthesized by a commercial gene service vendor (*e.g.*, GeneArt, Life Sciences, *etc.*) with 750-bp homology arms at each end. The homology arms at the 5' and 3' ends of the donor correspond to endogenous AHAS sequence immediately upstream and downstream of the double strand DNA break created by ZFNs 29732 and 29730 (encoded on plasmid pDAB109350).

***Design and Synthesis of Donor DNA for Second Sequential Transgene Stacking at an Endogenous AHAS Locus Using HDR-Directed DNA Repair***

**[0371]** 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, HDR-directed homology repair. The design is based on the 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 (pDAS000436; Figure 17; SEQ ID NO:296) is identical in sequence to pDAS000434 (Figure 13).

**[0372]** The donor cassette is synthesized by a commercial gene service vendor (*e.g.*, GeneArt, Life Sciences, *etc.*) with 750-bp homology arms at each end. The homology arm at the 5' end of the donor corresponds to endogenous AHAS sequence immediately upstream of the double strand DNA break created by ZFNs 34480 and 34481 (encoded on plasmid pDAB111860). The homology arm at the 3' end of the donor corresponds to GOI-1 sequence adjacent to the double stand DNA break created by ZFNs 34480 and 34481 in the donor DNA integrated in the first transgene stack.

***Transformation System for Exogenous Marker-Free, Sequential Transgene Stacking at an Endogenous AHAS Locus in Wheat using HDR-Directed DNA Repair***

**[0373]** Transgenic wheat events with multiple transgenes stacked at the same endogenous AHAS locus are produced by exogenous transgenic marker-free, sequential stacking of transgenes encoding traits (without use of a transgenic marker) 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 in Example 5. Figure 15a 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

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. Subsequently, the integration of the second transgene allows 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 susceptibility to imidazolinones (Figure 15b). 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 imidazoliones, thus allowing the regeneration of correctly targeted plants using IMAZAMOX® as a selection agent. 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 imidiazolinone and sulfonylurea selection agents.

**[0374]** The transformation system used to regenerate wheat plants with sequentially stacked transgenes at an endogenous AHAS locus is based on the previously described approach for biolistics-mediated DNA delivery to scutella of immature zygotic wheat embryos, or direct DNA delivery to wheat protoplasts using approaches known to one skilled in the art; for example, using the of He *et al.* (1994) Plant Cell Reports 14: 92-196, or any of the methods described in Example 11.

## Example 13: Development of a Transformation System for Exogenous Marker-Free Genome Editing at a Non-Selectable Trait Locus in Wheat

**[0375]** Precision genome modification of endogenous loci provides an effectual approach to modify trait expression. The generation of exogenous marker-free transformation events with precise genome modifications at one or more non-selectable endogenous trait loci provides opportunities to create new and novel high-value alleles for crop improvement. Here, we describe the development of a transformation system for ZFN-mediated, exogenous marker-free, precision genome editing at non-selectable trait loci in wheat that can be adapted for both integrative and non-integrative trait modification.

**[0376]** The transformation system is based on a two-step process. In the first step, ZFN-mediated precision genome modification is used to simultaneously modify

two independent loci in the plant genome; one locus is modified to confer tolerance to a selectable marker, the other is modified to alter expression for a non-selectable trait of interest. Transformation T0 events co-edited at both loci are generated by selecting for the introduced exogenous selectable marker. In the second step, marker-free events with only the modified trait locus are recovered by PCR screening of segregating T1 plants. The approach can be adapted for non-integrative precision genome modification that results in either the ablation of the non-selectable endogenous gene, or re-writing (editing) of the nucleotide sequence of the non-selectable endogenous gene. Alternatively, the approach can be adapted for integrative precision genome modification in which the function of the non-selectable endogenous gene is altered. More broadly, the approach could be adapted for non-integrative precision genome modification in which previously integrated exogenous DNA, for example a transgene, is excised.

**[0377]** The endogenous AHAS gene in wheat was selected as a model locus to establish and validate the transformation system for exogenous marker-free precision genome editing at a non-selectable trait locus in wheat.

***Preparation of Donor DNA for ZFN-mediated NHEJ-directed AHAS Gene Editing***

**[0378]** The donor DNA molecule, pDAS000267 (SEQ ID NO:84 and SEQ ID NO:85) was designed and synthesized as described in Example 6. Briefly, the donor DNA consisted a 95-bp double stranded molecule that was designed to integrate 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 pDAS000267 construct consisted of two parts. The 5' end contained sequence nearly 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 synonymous mutations (in which a silent mutation was incorporated into the donor sequence). The 3' end of the donor molecule contained a unique sequence that 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.

***Preparation of ZFN Construct DNA***

**[0379]** Plasmid DNA for pDAB109350 (Figure 1) encoding ZFNs 29732 and 29730 was prepared from cultures of *E. coli* using the PURE YIELD PLASMID MAXIPREP SYSTEM® (Promega Corporation, Madison, WI) following the manufacturer's instructions.

***Design and Production of Binary Vector Encoding PAT Selection Cassette***

**[0380]** Standard cloning methods were used to construct the binary vector pDAS000004 (SEQ ID:303; Figure 18). The PAT selection cassette consisted of the promoter, 5' untranslated region and intron from the Actin (*Act*1) 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 comprising phosphinothricin, glufosinate, and bialaphos (Wohlleben *et al.,* (1988) *Gene*, 70(1): 25-37). This cassette was terminated with the 3' UTR comprising the transcriptional terminator and polyadenylation sites from the 35s gene of cauliflower mosaic virus (CaMV) (Chenault *et al.,* (1993) *Plant Physiology* 101 (4): 1395-1396).

**[0381]** The selection cassette was synthesized by a commercial gene synthesis vendor (GeneArt, Life Technologies, *etc.*) and cloned into Gateway-enabled binary vector. Colonies of the assembled plasmid were screened by restriction digestion of miniprep DNA using restriction endonucleases obtained from New England BioLabs and Promega. Plasmid preparations were performed using the QIAPREP *SPIN MINIPREP* KIT™ 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). Plasmid DNA used for transfection was prepared from cultures of *E. coli* using the PURE YIELD PLASMID MAXIPREP SYSTEM® (Promega Corporation, Madison, WI) following the manufacturer's instructions.

***Biolistic-mediated Transformation System for Generating Exogenous Marker-Free Wheat Plants with Precise Genome Modifications at Non-Selectable Endogenous Trait Loci***

**[0382]** A total of 2,320 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 using a DNA mixture comprising 2.5 μg of donor pDAS000267 and plasmid pDAB109350 (at a molar ratio of 7:1, respectively) and 2.5 μg of plasmid pDAS000004.

**[0383]** Following bombardment, the transfected scutella were incubated at 26°C in the dark for 16 h before being transferred onto medium for callus induction. The scutella were cultured in the dark on callus induction medium at 24°C for 2 weeks. The resultant calli were sub-cultured once onto fresh callus induction medium, and kept in the same conditions for a further two weeks. The SEC was transferred onto plant regeneration medium containing 5 mg/ml BASTA® 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 containing 5 mg/ml BASTA® and cultured under the same conditions for 2-3 weeks. Regenerated plantlets producing roots were expected have one or more copies of the PAT selection cassette randomly inserted into the plant genome. The roots of these plantlets were removed and the plants were again sub-cultured on rooting media containing 200 nM IMAZAMOX® under the same conditions for 2-3 weeks. Plants with regrown roots were expected to have the S653N mutation (resulting from precise integration of pDAS000267) in one or more copies of endogenous AHAS gene.

**[0384]** A total of 170 wheat plants producing strong root growth on rooting medium containing BASTA® were obtained from the transfection of the 2,320 scutella of immature zygotic embryos from the donor wheat line cv. Bobwhite MPB26RH. Of these, two wheat plants produced roots when transferred to rooting medium containing IMAZAMOX®. These plants were transferred to soil and grown under glasshouse containment conditions to produce T1 seed.

***Optimization of BASTA® Chemical Selection for Enrichment of Transformed Events in a Wheat Protoplast-Based Transformation System***

**[0385]** A series of experiments are performed to determine optimal selection conditions for regenerating wheat plants expressing the PAT gene conferring tolerance to BASTA® from a protoplast-based transformation system such as those described by Qiao *et al.* (1992) Plant Cell Reports 11:262-265; Ahmed and Sagi (1993) Plant Cell Reports 12:175-179; Pauk *et al.*(1994) Plant Cell, Tissue and Organ Culture 38: 1-10; He *et al.* (1994) Plant Cell Reports 14: 92-196; Gu and Lang (1997) Plant Cell, Tissue and Organ Culture 50: 139–145; and Li *et al.* (1999) Plant Cell, Tissue and Organ Culture 58: 119–125.

**[0386]** BASTA® selection conditions are optimized using protoplasts derived from somatic embyrogenic callus (SEC)-derived cell suspension culture of the wheat line cv. Bobwhite MPB26RH. While protoplasts derived from Bobwhite MPB26RH are non-totipotent (*i.e.*, cannot be used to regenerate entire plants), the selection conditions established for enriching the events that express the PAT gene are expected to be transferrable to any protoplast-based transformation system based on a totipotent wheat genotype. The experiments are conducted, and the basal tolerance of the wild-type donor wheat line cv. Bobwhite MPB26RH to BASTA® is established. The use of BASTA® selection conditions stronger than basal tolerance are identified and used to select for transformed cells expressing the PAT gene.

***Establishment of Agarose Bead-Type Cultures and BASTA® Selection Conditions***

**[0387]** Protoplasts are isolated from an established SEC-derived cell suspension culture and used to establish agarose bead-types cultures, as described previously. The basal tolerance of wheat line cv. Bobwhite MPB26RH to BASTA® is determined by incubating the agarose bead-type cultures in media supplemented with 0, 0.5, 2.5, 5, 7.5, 10, 20, 30, 40 and 50 mg/L BASTA® and assessing the rate of calli growth after 2 weeks. The BASTA® concentrations (*e.g.*, higher than 20 mg/L) that severely impeded calli development are optimal for enriching and selecting wheat cells having the PAT gene.

***Molecular Characterization of the Transformed Wheat Plants with BASTA® and IMAZAMOX® Tolerant Phenotypes***

**[0388]** The two wheat plants having both the BASTA® and IMAZAMOX® herbicide tolerant phenotypes were molecularly characterized to identify the endogenous AHAS gene that contained the S653N mutation resulting from integration of pDAS000267 donor at a genomic double cleavage site created by ZFNs 29732 and 29730 encoded on pDAB109350.

**[0389]** Two molecular assays were performed for each wheat plant using genomic DNA extracted with the DNEASY® PLANT DNA EXTRACTION MINI KIT™ (Qiagen) from freeze-dried leaf tissue, as described previously.

**[0390]** The first molecular test was used to confirm that the regenerated wheat plants had at least one randomly integrated copy of the PAT gene. A duplex hydrolysis probe qPCR assay (analogous to TAQMAN®) was used to amplify the endogenous single copy gene, *puroindoline-b* (Pinb) gene, from the D genome of hexaploid wheat (Gautier *et al*., (2000) *Plant Science* 153, 81–91; SEQ ID NO: 89, SEQ ID NO: 90 and SEQ ID NO: 91 for forward primer, reverse primer, and probe sequence, respectively) and a region of the Actin (Act1) promoter present on pDAS000004 (SEQ ID NO: 92, SEQ ID NO: 93 and SEQ ID NO: 94 for forward primer, reverse primer, and probe sequence, respectively). Assessment for the presence, and estimated copy number of pDAS000004 was performed according to the method described in Livak and Schmittgen (2001) *Methods* 25(4):402-8. From the results, evidence was obtained for the integration of the PAT polynucleotide sequence into the genome of wheat plant events *yc06-9110-1* and *yr00-9311-1*, respectively.

**[0391]** The second molecular test was used to characterize the sub-genomic location and outcome for ZFN-mediated NHEJ-directed donor integration at the endogenous AHAS genes. PCR with primers AHASs653ZFN.F2 and AHASs653ZFN.R1 (SEQ ID NO: 301 and 302; Table 18) was used to amplify the DNA fragment from each of the three homoeologous copies of the endogenous AHAS gene. The amplified fragment contained a region containing the binding site for ZFNs 29732 and 29730 (encoded on plasmid pDAB190350), and to include genomic nucleotide sequence variation. Enough genomic nucleotide sequence variation was included to differentiate between the AHAS homoeologs, such that the resulting amplicons could be unequivocally attributed (at the sequence level) to the wheat sub-genome from which they were derived. The resulting amplicons were prepared for

deep sequencing as described in Example 12 and sequenced on an Illumina MiSEQ™ instrument to generate 250-bp paired-end sequence reads, according to the manufacturer's instructions. The resultant sequence reads were computationally processed, as described previously, to assign each read to sample (based on the barcode index) and the sub-genome from which they were derived (based on nucleotide variation that distinguished between homoeologous copies of the AHAS gene). As described in Example 9, the integration of pDAS000267 into an endogenous AHAS locus results in a 95-bp size difference between the wild-type (unmodified) and resulting transgenic (modified) allele. Hence, PCR amplification of both the wild-type and modified AHAS gene loci is expected. Custom developed PERL scripts and manual data manipulation in MICROSOFT EXCEL 2010™ (Microsoft Corporation) were used to characterize the sub-genomic location and outcome for donor integration into the endogenous AHAS genes.

**[0392]** From the results of the second molecular assay, conclusive evidence for precise ZFN-mediated NHEJ-directed gene editing at an endogenous AHAS locus was demonstrated for both wheat plants. Event *yc06-9110-1* had perfect hemizygous donor integration in the B-genome (Table 24). Event *yr00-9311-1* had simultaneous donor integration into multiple sub-genomes. In the A-genome, independent editing of both endogenous AHAS loci was observed. One allele had partial donor integration that resulted in the expected integration of the S653N mutation for expression of the AHAS herbicide tolerance phenotype. However, a fragment spanning 24-bp nucleotides were deleted from the 3' end of the donor molecule. The other allele had integration of a 51-bp polynucleotide sequence of unknown origin. No sequence reads originating from the B-genome were obtained, suggesting independent integration of a large polynucleotide sequence into each of the endogenous AHAS loci (Table 24). Consensus sequences for the alleles present in each sub-genome for two regenerated wheat plants are provided as SEQ ID NOs: 304-313. The absence of evidence of sequence originating from pDAS0000004 in both wheat plant events indicates that the PAT gene conferring tolerance to BASTA® was randomly integrated into a different locus in the plant genome.

**Table 24: ZFN-mediated NHEJ-directed AHAS editing outcomes for wheat plants *yc06-9110-1* and *yr00-9311-1***

| | | A-genome | | B-genome | | D-genome | | |
|---|---|---|---|---|---|---|---|---|
| | | Allele 1 | Allele 2 | Allele 1 | Allele 2 | Allele 1 | Allele 2 | SEQ ID NO: |
| *yc06-9110-1* | Status | UE | UE | PE | UE | UE | UE | 304-309 |
| | No. Reads[1] | 143,159 | | 76,903 | 110,846 | 219,858 | | |
| *yr00-9311-1* | Status | IE | IE | nd | nd | UE | UE | 310-313 |
| | No. Reads[1] | 164,038 | 138,539 | 0 | | 556,123 | | |

[1]Number of sequence reads originating from the specified sub-genome and having the sequence haplotype corresponding to wild-type (unmodified) or transgenic (modified) AHAS loci.
"PE" indicates perfect edit; *i.e.*, ZFN-mediated NHEJ-directed genome editing produced a predicted outcome.
"IE" indicates imperfect edit; *i.e.*, ZFN-mediated NHEJ-directed genome editing produced an unpredicted outcome.
"UE" indicates unedited allele; *i.e.*, allele had wild-type sequence.
"nd" indicates not detected.

**[0393]** These results disclose for the first time a transformation method which can be utilized to generate exogenous marker-free wheat plants having precise genome modifications at one or more non-selectable trait loci. Wheat plants comprising an integrated AHAS donor polynucleotide encoding a S653N mutation conferring tolerance to imidazolinone class herbicides are exemplified. As will be appreciated by one skilled in the art, wheat plants without the exogenous transgenic selectable marker (*e.g.*, PAT) can be recovered by screening T1 plants derived from these events using PCR assays specific for either the PAT or the modified AHAS genes.

**[0394]**

**[0395]** Although disclosure has been provided in some detail by way of 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:

1. A plant cell comprising a targeted genomic modification to one or more alleles of an endogenous acetohydroxyacid synthase (AHAS) gene of SEQ ID NO: 1, SEQ ID NO: 2 or SEQ ID NO:3 in the plant cell, wherein the genomic modification follows cleavage by a site-specific nuclease, wherein the site-specific nuclease is a zinc finger nuclease comprising a Fok I cleavage domain and a DNA binding domain that binds to a sequence comprising a target site as shown in any one of SEQ ID NOs: 35-56 or 263-278, and wherein the genomic modification produces a mutation in the endogenous AHAS gene such that the endogenous gene produces a product that results in an imidazolinone herbicide tolerant plant cell, wherein the genomic modification comprises

(a) introduction of one or more indels that disrupt expression of the endogenous gene, or

(b) integration of one or more exogenous sequences, wherein

(i) the exogenous sequence does not encode a transgenic selectable marker, or

(ii) the exogenous sequence encodes a protein selected from the group consisting of a protein that increases crop yield, a protein encoding disease resistance, a protein that increases growth, a protein encoding insect resistance, a protein encoding herbicide tolerance, and combinations thereof, or

(iii) two or more exogenous sequences are integrated into the endogenous gene.

2. The plant cell of claim 1, wherein the endogenous AHAS gene with the genomic modification encodes a protein that confers tolerance to sulfonylurea herbicides, or to imidazolinone herbicides.

3. The plant cell of claim 1 or claim 2, wherein the plant cell is a polyploid plant cell.

4. The plant cell of any one of claims 1 to 3, wherein the plant is selected from the group consisting of wheat, soy, maize, potato, alfalfa, rice, barley, sunflower, tomato, *Arabidopsis*, cotton, *Brassica* species, and timothy grass.

Date Reçue/Date Received 2022-12-05

5. The plant cell of any one of claims 1 to 4, further comprising one or more transgenes integrated into the genome of the plant cell at one or more loci different from the endogenous AHAS gene.

6. A method for making a plant cell according to any one of claims 1 to 5, the method comprising:

expressing one or more site-specific zinc finger nucleases in the plant such that one or more alleles of the endogenous AHAS gene across multiple genomes of a polyploid plant cell are modified, wherein the modification comprises integration of one or more exogenous sequences into one or more alleles of the endogenous AHAS gene.

7. The method of claim 6, wherein the modification disrupts expression of the endogenous AHAS gene.

8. The method of claim 6 or claim 7, wherein at least one of the integrated exogenous sequences expresses a product that results in a selectable phenotype in the plant cell; and the method further comprises the step of

selecting plant cells that express the selectable phenotype, wherein the plant cells are selected which incorporate the one or more exogenous sequences.

9. The method of claim 8, wherein integrating the one or more exogenous sequences occurs by homologous recombination or non-homologous end joining.

10. The method of claim 8 or claim 9, wherein the one or more exogenous sequences are incorporated simultaneously or sequentially into the one or more endogenous AHAS loci.

11. The method of any one of claims 8 to 10, wherein the AHAS gene is located on an A, B, or D genome of a polyploidy genome.

Date Reçue/Date Received 2022-12-05

12. The method of any one of claims 8 to 11, wherein the one or more exogenous sequences encode a S653N AHAS mutation and/or a P197S AHAS mutation.

13. The method of any one of claims 8 to 12, the method further comprising the steps of;

culturing the selected plant cells comprising the one or more exogenous sequences; and

obtaining a whole plant comprising the one or more exogenous sequences integrated within the endogenous AHAS gene of the plant genome.

14. The method of any one of claims 8 to 13, wherein a selection agent comprising an imidazolinone, or a sulfonylurea selection agent is used to select the plant cells.

15. The method of claim 13, wherein the whole plant comprising the one or more exogenous sequences integrated within the endogenous AHAS gene of the plant genome is further modified to incorporate an additional exogenous sequence within the endogenous AHAS gene of the plant genome or

wherein the one or more exogenous sequence does not encode a transgenic selectable marker.

16. The plant cell of claim 1, wherein the increased crop yield comprises an increase in fruit yield, grain yield, biomass, fruit flesh content, size, dry weight, solids content, weight, color intensity, color uniformity, altered chemical characteristics, or combinations thereof.

17. A zinc finger protein (ZFP) that binds to a target site in an endogenous acetohydroxyacid synthase (AHAS) gene as shown in any one of SEQ ID NOs: 35-56 or 263-278, the zinc finger protein comprising from four to six zinc finger domains ordered F1 to F4, F1 to F5 or F1 to F6, each zinc finger domain comprising a recognition helix region and wherein the zinc finger protein comprises the recognition helix regions ordered as shown in a single row of the following Table:

| ZFP designation | F1 | F2 | F3 | F4 | F5 | F6 |
|---|---|---|---|---|---|---|
| 29964 | QSSHLTR SEQ ID NO:181 | RSDDLTR SEQ ID NO:182 | RSDDLTR SEQ ID NO:182 | YRWLLRS SEQ ID NO:183 | QSGDLTR SEQ ID NO:184 | QRNARTL SEQ ID NO:185 |
| 29965 | RSDNLSV SEQ ID NO:186 | QKINLQV SEQ ID NO:187 | DDWNLSQ SEQ ID NO:188 | RSANLTR SEQ ID NO:189 | QSGHLAR SEQ ID NO:190 | NDWDRRV SEQ ID NO:191 |
| 29966 | RSDDLTR SEQ ID NO:182 | YRWLLRS SEQ ID NO:183 | QSGDLTR SEQ ID NO:184 | QRNARTL SEQ ID NO:185 | RSDHLSQ SEQ ID NO:192 | DSSTRKK SEQ ID NO:193 |
| 29967 | RSDDLTR SEQ ID NO:182 | YRWLLRS SEQ ID NO:183 | QSGDLTR SEQ ID NO:184 | QRNARTL SEQ ID NO:185 | RSDVLSE SEQ ID NO:194 | DRSNRIK SEQ ID NO:195 |
| 29968 | RSDNLSN SEQ ID NO:196 | TSSSRIN SEQ ID NO:197 | DRSNLTR SEQ ID NO:198 | QSSDLSR SEQ ID NO:199 | QSAHRKN SEQ ID NO:200 | N/A |
| 29969 | DRSHLTR SEQ ID NO:201 | QSGHLSR SEQ ID NO:202 | RSDNLSV SEQ ID NO:186 | QKINLQV SEQ ID NO:187 | DDWNLSQ SEQ ID NO:188 | RSANLTR SEQ ID NO:189 |
| 29970 | QSGDLTR SEQ ID NO:184 | QRNARTL SEQ ID NO:185 | RSDVLSE SEQ ID NO:194 | DRSNRIK SEQ ID NO:195 | RSDNLSE SEQ ID NO:203 | HSNARKT SEQ ID NO:204 |
| 29730 | TSGNLTR SEQ ID NO:205 | HRTSLTD SEQ ID NO:206 | QSSDLSR SEQ ID NO:199 | HKYHLRS SEQ ID NO:207 | QSSDLSR SEQ ID NO:199 | QWSTRKR SEQ ID NO:208 |
| 29971 | DRSHLTR SEQ ID NO:201 | QSGHLSR SEQ ID NO:202 | RSDNLSN SEQ ID NO:196 | TSSSRIN SEQ ID NO:197 | DRSNLTR SEQ ID NO:198 | N/A |
| 29731 | RSDVLSE SEQ ID NO:194 | SPSSRRT SEQ ID NO:209 | RSDTLSE SEQ ID NO:210 | TARQRNR SEQ ID NO:211 | DRSHLAR SEQ ID NO:212 | N/A |
| 29732 | RSDSLSA SEQ ID NO:213 | RSDALAR SEQ ID NO:214 | RSDDLTR SEQ ID NO:182 | QKSNLSS SEQ ID NO:215 | DSSDRKK SEQ ID NO:216 | N/A |

| | | | | | | |
|---|---|---|---|---|---|---|
| 30006 | TSGNLTR SEQ ID NO:205 | WWTSRAL SEQ ID NO:217 | DRSDLSR SEQ ID NO:218 | RSDHLSE SEQ ID NO:219 | YSWRLSQ SEQ ID NO:220 | N/A |
| 30008 | RSDSLSV SEQ ID NO:221 | RNQDRKN SEQ ID NO:222 | QSSDLSR SEQ ID NO:199 | HKYHLRS SEQ ID NO:207 | QSGDLTR SEQ ID NO:184 | N/A |
| 29753 | QSGNLAR SEQ ID NO:223 | DRSALAR SEQ ID NO:224 | RSDNLST SEQ ID NO:225 | AQWGRTS SEQ ID NO:226 | N/A | N/A |
| 29754 | RSADLTR SEQ ID NO:227 | TNQNRIT SEQ ID NO:228 | RSDSLLR SEQ ID NO:229 | LQHHLTD SEQ ID NO:230 | QNATRIN SEQ ID NO:231 | N/A |
| 29769 | QSGNLAR SEQ ID NO:223 | DRSALAR SEQ ID NO:224 | RSDNLST SEQ ID NO:225 | AQWGRTS SEQ ID NO:226 | N/A | N/A |
| 29770 | QSGDLTR SEQ ID NO:184 | MRNRLNR SEQ ID NO:232 | DRSNLSR SEQ ID NO:233 | WRSCRSA SEQ ID NO:234 | RSDNLSV SEQ ID NO:186 | N/A |
| 30012 | HSNARKT SEQ ID NO:204 | QSGNLAR SEQ ID NO:223 | DRSALAR SEQ ID NO:224 | RSDNLST SEQ ID NO:225 | AQWGRTS SEQ ID NO:226 | N/A |
| 30014 | HSNARKT SEQ ID NO:204 | QSGNLAR SEQ ID NO:223 | DRSALAR SEQ ID NO:224 | RSDHLSQ SEQ ID NO:192 | QWFGRKN SEQ ID NO:235 | N/A |
| 30018 | QSGDLTR SEQ ID NO:184 | MRNRLNR SEQ ID NO:232 | DRSNLSR SEQ ID NO:233 | WRSCRSA SEQ ID NO:234 | QRSNLDS SEQ ID NO:34 | N/A |
| 29988 | QSGDLTR SEQ ID NO:184 | QWGTRYR SEQ ID NO:33 | DRSNLSR SEQ ID NO:233 | HNSSLKD SEQ ID NO:32 | QSGNLAR SEQ ID NO:223 | N/A |
| 29989 | RSDVLSA SEQ ID NO:31 | RNDHRIN SEQ ID NO:30 | RSDHLSQ SEQ ID NO:192 | QSAHRTN SEQ ID NO:29 | DRSNLSR SEQ ID NO:233 | DSTNRYR SEQ ID NO:28 |
| 34456 | RSADLTR SEQ ID NO:227 | RSDDLTR SEQ ID NO:182 | RSDDLTR SEQ ID NO:182 | RSDALTQ SEQ ID NO:236 | ERGTLAR SEQ ID NO:237 | RSDDLTR SEQ ID NO:182 |
| 34457 | QSGDLTR SEQ ID NO:184 | DTGARLK SEQ ID NO:238 | RSDDLTR SEQ ID NO:182 | HRRSRDQ SEQ ID NO:239 | DRSYRNT SEQ ID NO:240 | N/A |
| 34470 | RSADLSR SEQ ID | RSDHLSA SEQ ID | QSSDLRR SEQ ID | DRSNLSR SEQ ID | RSDDRKT SEQ ID | N/A |

Date Reçue/Date Received 2022-12-05

| | NO:241 | NO:242 | NO:243 | NO:233 | NO:244 | |
|---|---|---|---|---|---|---|
| 34471 | QSGDLTR SEQ ID NO:184 | RRADRAK SEQ ID NO:245 | RSDDLTR SEQ ID NO:182 | TSSDRKK SEQ ID NO:246 | RSADLTR SEQ ID NO:227 | RNDDRKK SEQ ID NO:247 |
| 34472 | RSADLTR SEQ ID NO:227 | DRSNLTR SEQ ID NO:198 | ERGTLAR SEQ ID NO:237 | RSDDLTR SEQ ID NO:182 | DRSDLSR SEQ ID NO:218 | DSSTRRR SEQ ID NO:248 |
| 34473 | RSDHLSE SEQ ID NO:219 | HSRTRTK SEQ ID NO:249 | RSDTLSE SEQ ID NO:210 | NNRDRTK SEQ ID NO:250 | ERGTLAR SEQ ID NO:237 | DRSALAR SEQ ID NO:224 |
| 34474 | ERGTLAR SEQ ID NO:237 | RSDDLTR SEQ ID NO:182 | DRSDLSR SEQ ID NO:218 | DSSTRRR SEQ ID NO:248 | DRSNLTR SEQ ID NO:198 | N/A |
| 34475 | RSDHLSR SEQ ID NO:249 | QQWDRKQ SEQ ID NO:73 | DRSHLTR SEQ ID NO:201 | DSSDRKK SEQ ID NO:216 | DRSNLSR SEQ ID NO:233 | VSSNLTS SEQ ID NO:251 |
| 34476 | DRSDLSR SEQ ID NO:218 | DSSTRRR SEQ ID NO:248 | DRSNLSR SEQ ID NO:233 | QSGDLTR SEQ ID NO:184 | DRSNLTR SEQ ID NO:198 | N/A |
| 34477 | ERGTLAR SEQ ID NO:237 | RSDHLSR SEQ ID NO:249 | RSDALSV SEQ ID NO:252 | DSSHRTR SEQ ID NO:253 | DSSDRKK SEQ ID NO:216 | N/A |
| 34478 | RSDNLTR SEQ ID NO:254 | RSDNLAR SEQ ID NO:255 | DRSALAR SEQ ID NO:224 | DRSHLSR SEQ ID NO:256 | TSGNLTR SEQ ID NO:205 | N/A |
| 34479 | RSDALSV SEQ ID NO:252 | DSSHRTR SEQ ID NO:253 | RSDNLSE SEQ ID NO:203 | ARTGLRQ SEQ ID NO:254 | ERGTLAR SEQ ID NO:237 | DRSALAR SEQ ID NO:224 |
| 34480 | RSDNLAR SEQ ID NO:255 | DRSALAR SEQ ID NO:224 | DRSHLSR SEQ ID NO:256 | TSGNLTR SEQ ID NO:205 | RSDHLSR SEQ ID NO:249 | TSSNRKT SEQ ID NO:257 |
| 34481 | DRSALAR SEQ ID NO:224 | RSDALSV SEQ ID NO:252 | DSSHRTR SEQ ID NO:253 | RSDNLSE SEQ ID NO:203 | ARTGLRQ SEQ ID NO:254 | N/A |
| 34482 | RSDDLSK SEQ ID NO:258 | RSDNLTR SEQ ID NO:254 | RSDSLSV SEQ ID NO:221 | RSAHLSR SEQ ID NO:259 | RSDALST SEQ ID NO:260 | DRSTRTK SEQ ID NO:261 |
| 34483 | DSSDRKK SEQ ID NO:216 | RSAHLSR SEQ ID NO:259 | DRSDLSR SEQ ID NO:218 | RSDHLSE SEQ ID NO:219 | TSSDRTK SEQ ID NO:262 | N/A |

18. A zinc finger nuclease comprising a pair of zinc finger proteins of claim 17, wherein the pair comprises the following ZFPs:

(i) a ZFP designated 29964 and a ZFP designated 29965;

(ii) a ZFP designated 29966 and a ZFP designated 29968;

(iii) a ZFP designated 29967 and a ZFP designated 29968;

(iv) a ZFP designated 29967 and a ZFP designated 29969;

(v) a ZFP designated 29970 and a ZFP designated 29971;

(vi) a ZFP designated 29730 and a ZFP designated 29732;

(vii) a ZFP designated 29731 and a ZFP designated 29732;

(viii) a ZFP designated 30006 and a ZFP designated 30008;

(ix) a ZFP designated 29753 and a ZFP designated 29754;

(x) a ZFP designated 29769 and a ZFP designated 29770;

(xi) a ZFP designated 30012 and a ZFP designated 30018;

(xii) a ZFP designated 30014 and a ZFP designated 30018;

(xiii) a ZFP designated 29988 and a ZFP designated 29989;

(xiv) a ZFP designated 34456 and a ZFP designated 34457;

(xv) a ZFP designated 34470 and a ZFP designated 34471;

(xvi) a ZFP designated 34472 and a ZFP designated 34473;

(xvii) a ZFP designated 34474 and a ZFP designated 34475;

(xviii) a ZFP designated 34476 and a ZFP designated 34477;

(xix) a ZFP designated 34478 and a ZFP designated 34479;

(xx) a ZFP designated 34480 and a ZFP designated 34481; and

(xxi) a ZFP designated 34482 and a ZFP designated 34483.

19. A method of integrating one or more exogenous sequences into the genome of a plant cell, the method comprising:

a) expressing one or more site-specific nucleases according to claim 18 in the plant cell;

b) integrating one or more exogenous sequences into the endogenous AHAS loci within the genome of the plant cell, wherein the AHAS loci is modified such that the endogenous AHAS gene is mutated to express a product that results in a selectable phenotype in the plant cell; and

c) selecting plant cells that express the selectable phenotype, wherein plant cells are selected which incorporate the one or more exogenous sequences.

20. The method of claim 19, wherein the one or more exogenous sequences are selected from the group consisting of a donor polynucleotide, a transgene, or any combination thereof.

21. The method of claim 19 or 20, wherein integrating the one or more exogenous sequences occurs by homologous recombination or non-homologous end joining.

22. The method of any one of claims 19 to 21, wherein the one or more exogenous sequences are incorporated simultaneously or sequentially into the endogenous AHAS loci.

23. The method of any one of claims 19 to 22, wherein the AHAS gene is located on an A, B, or D genome of a polyploidy genome, and the one or more exogenous sequences encode a S653N AHAS mutation, or a P197S AHAS mutation.

24. The method of any one of claims 19 to 23, wherein the one or more exogenous sequences encode a transgene or produce an RNA molecule, wherein the transgene encodes a protein selected from the group consisting of a protein that increases crop yield, a protein encoding disease resistance, a protein that increases growth, a protein encoding insect resistance, a protein encoding herbicide tolerance, and combinations thereof and/or wherein the integration of the transgene further comprises introduction of one or more indels that disrupt expression of the one or more endogenous loci and produce the selectable phenotype.

25. The method of any of claims 19 to 24, the method further comprising the steps of;

d) culturing the selected plant cells comprising the one or more exogenous sequences; and

e) obtaining a whole plant comprising the one or more exogenous sequences integrated within the one or more endogenous loci of the plant genome.

26. The method of any one of claims 19 to 25, wherein a selection agent comprising an imidazolinone, or a sulfonylurea selection agent is used to select the plant cells.

27. The method of claim 26, wherein the whole plant comprising the one or more exogenous sequences integrated within the one or more endogenous loci of the plant genome is further modified to incorporate an additional exogenous sequence within the endogenous loci of the plant genome or wherein the one or more exogenous sequence does not encode a transgenic selectable marker.

Date Reçue/Date Received 2022-12-05

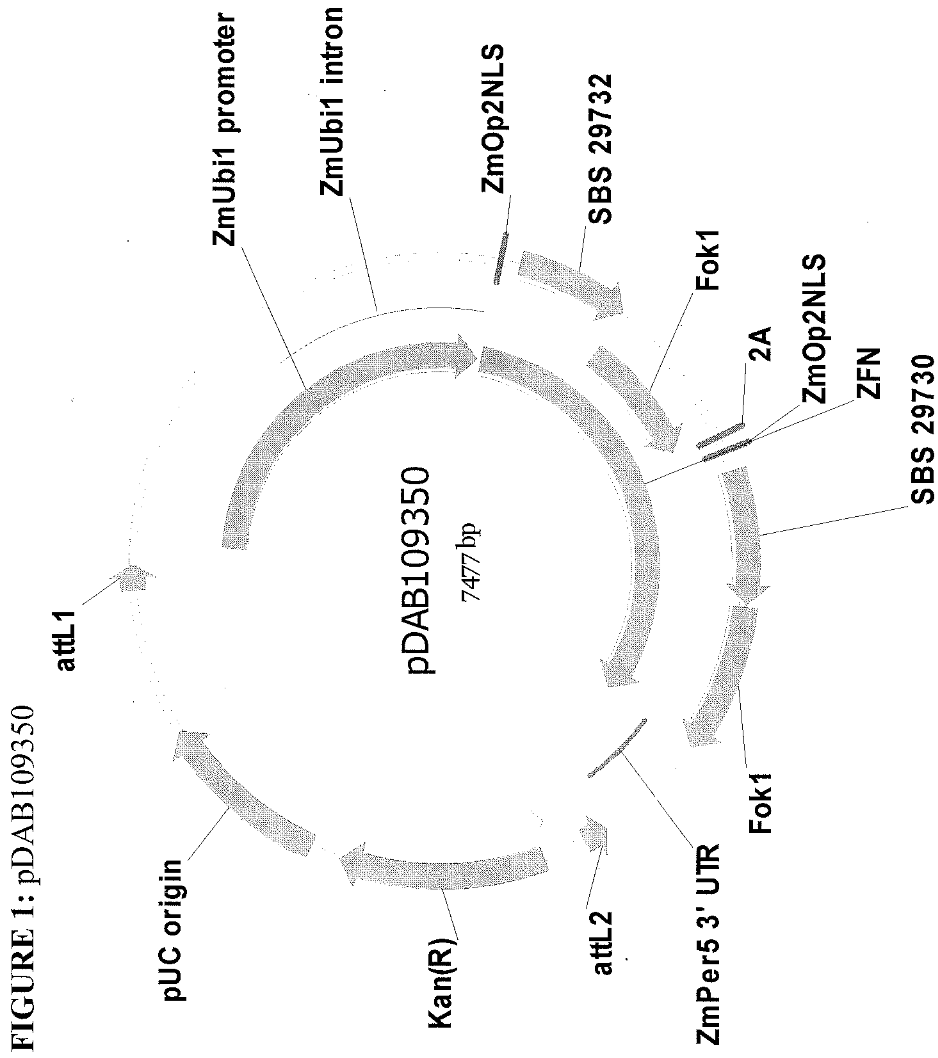



**FIGURE 1:** pDAB109350

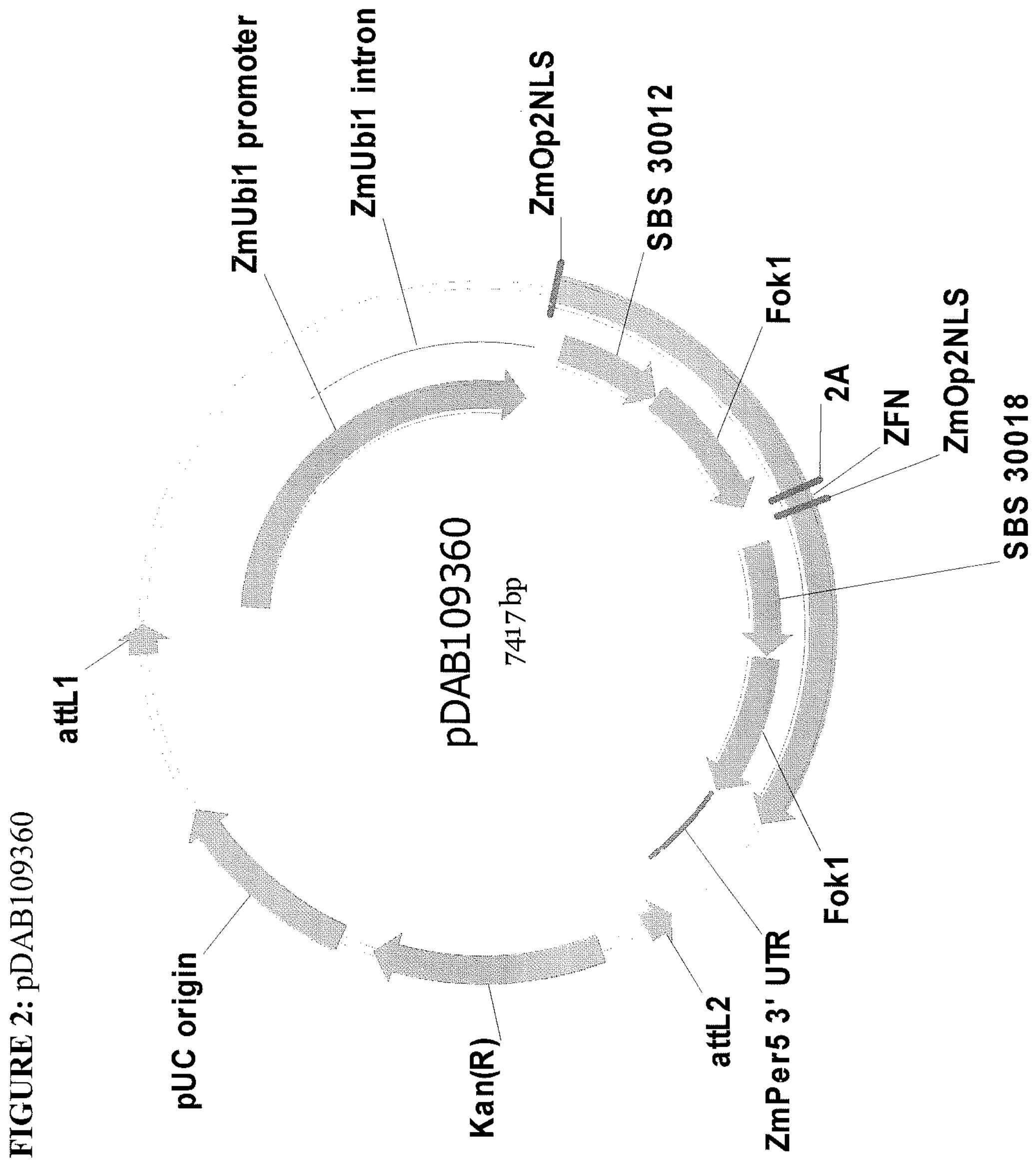



FIGURE 2: pDAB109360

FIGURE 3: pDAS000132

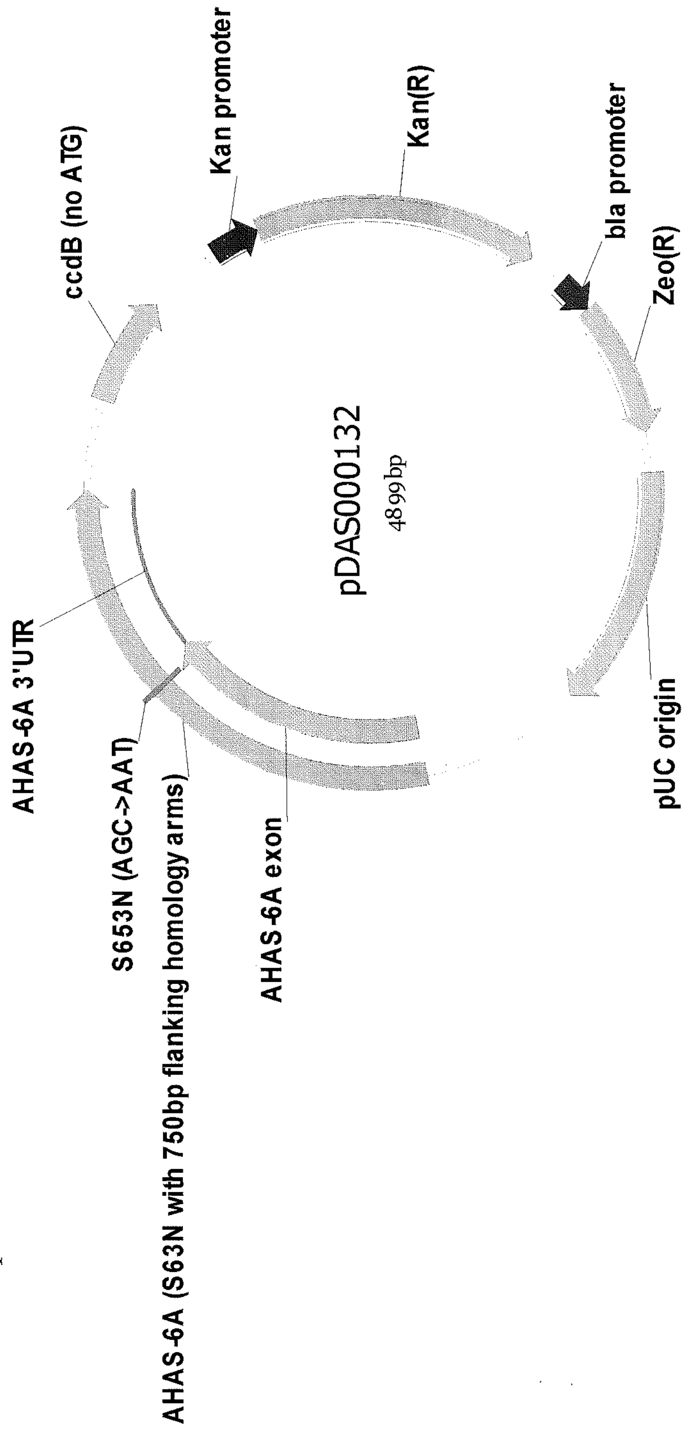
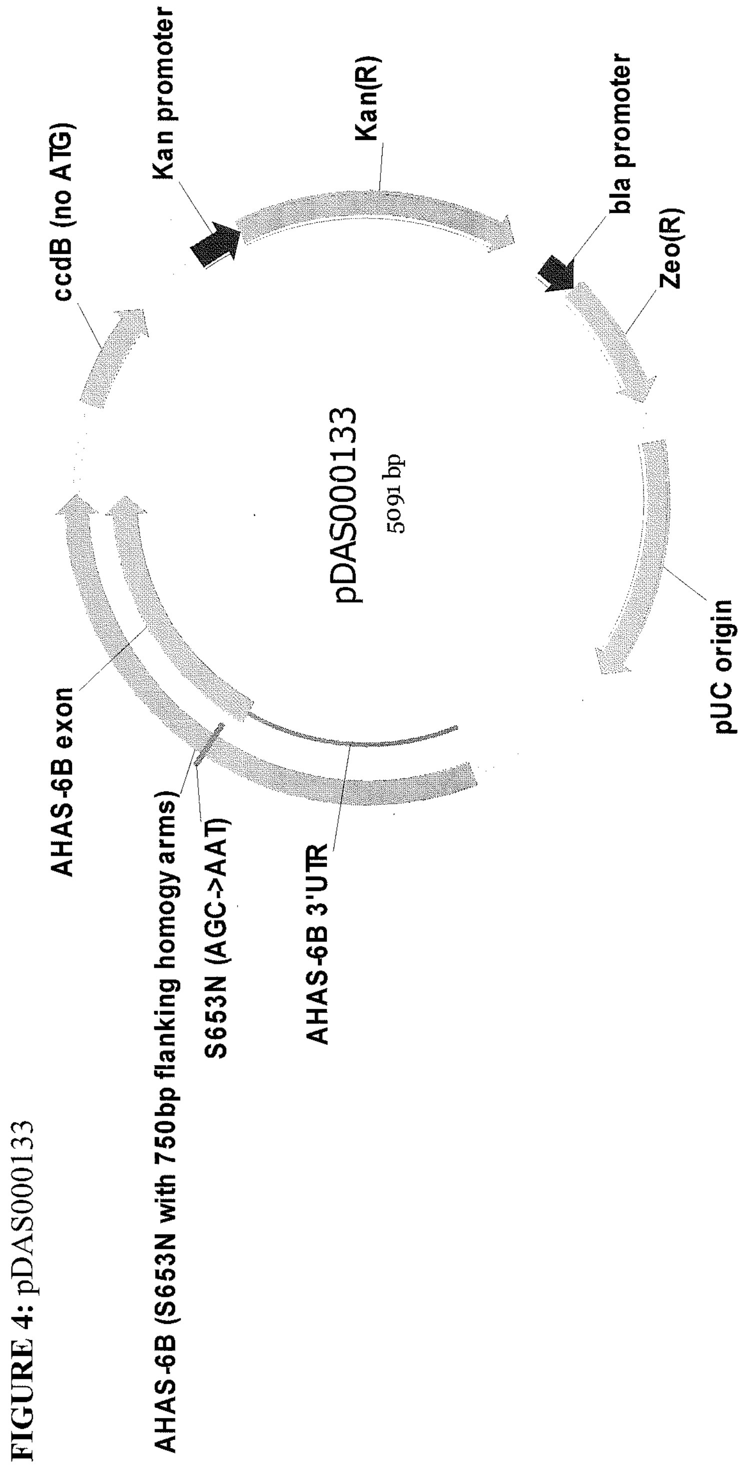


**FIGURE 4:** pDAS000133



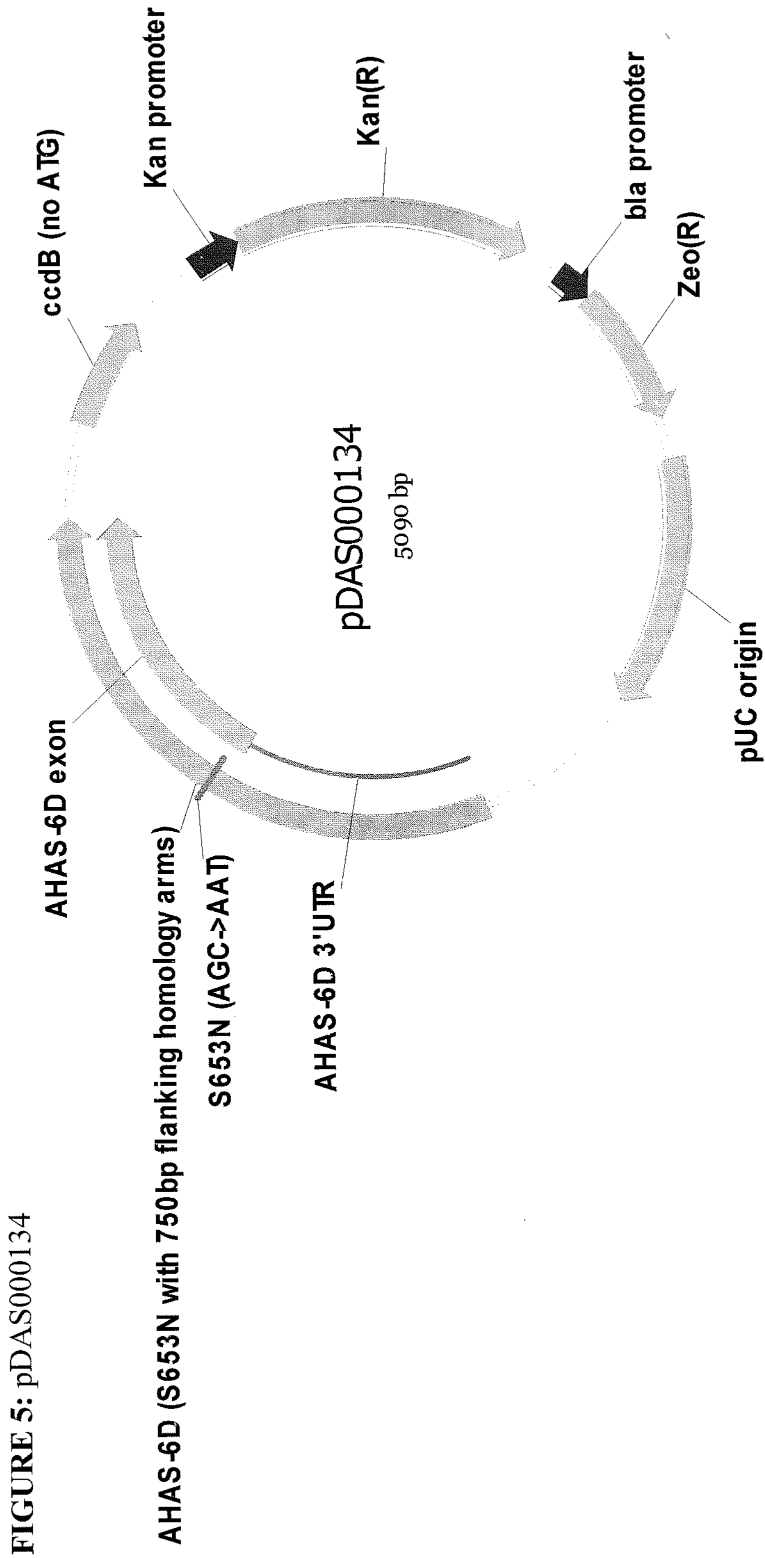



**FIGURE 5:** pDAS000134

**FIGURE 6:** pDAS000135

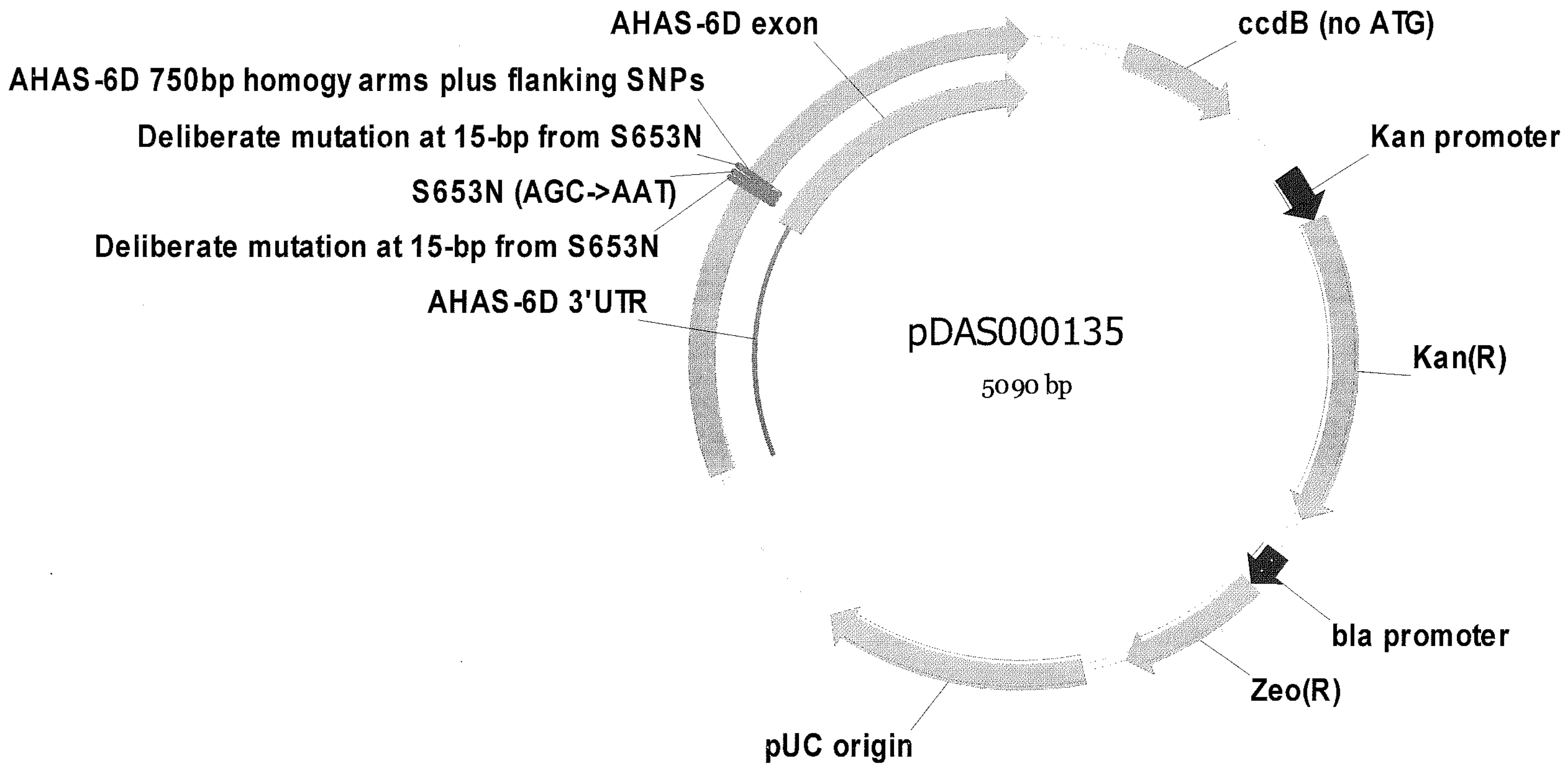


**FIGURE 7:** pDAS000131

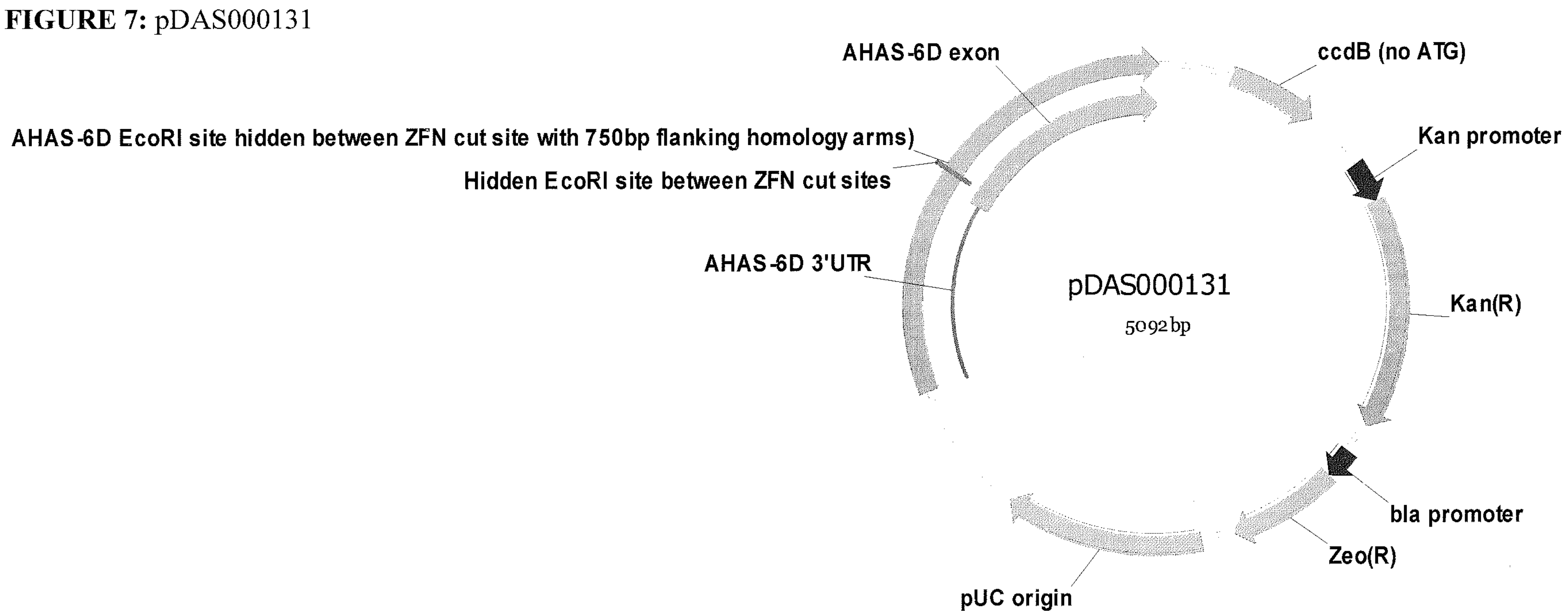
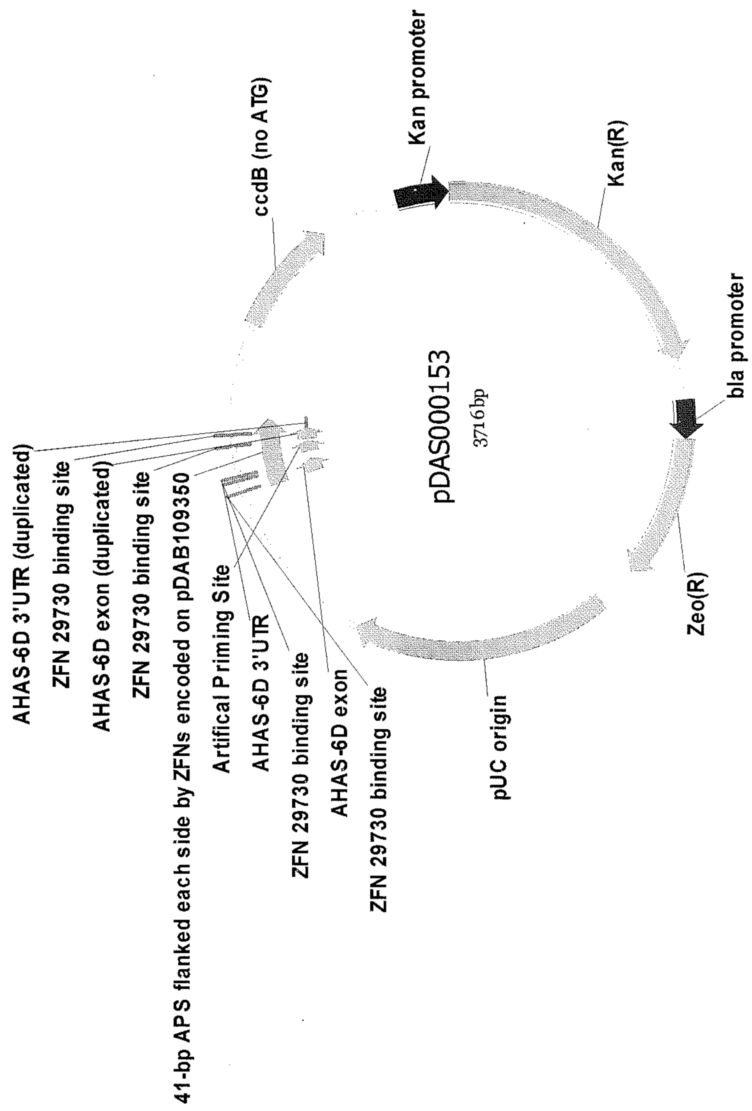


**FIGURE 8:** pDAS000153



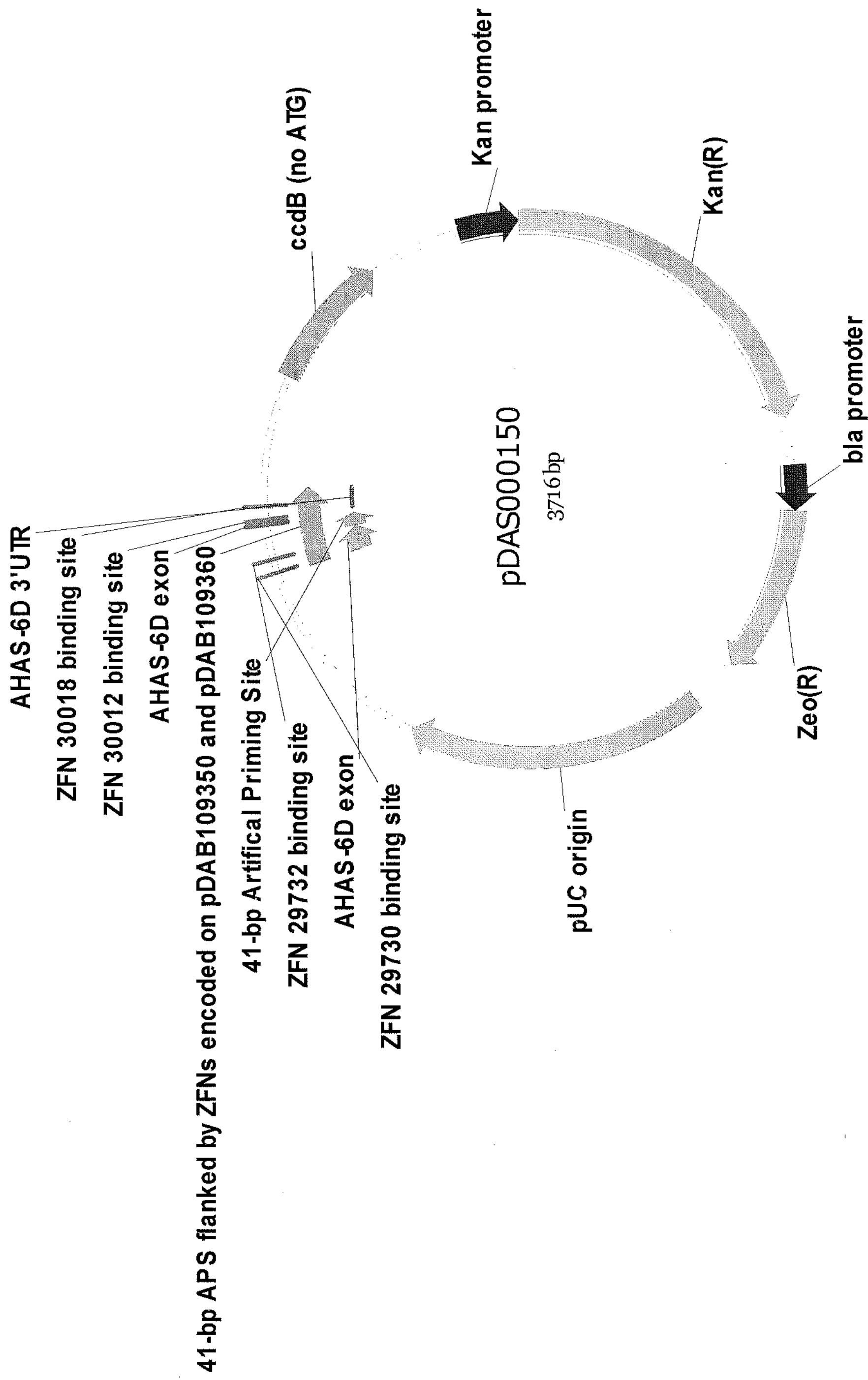



**FIGURE 9:** pDAS000150

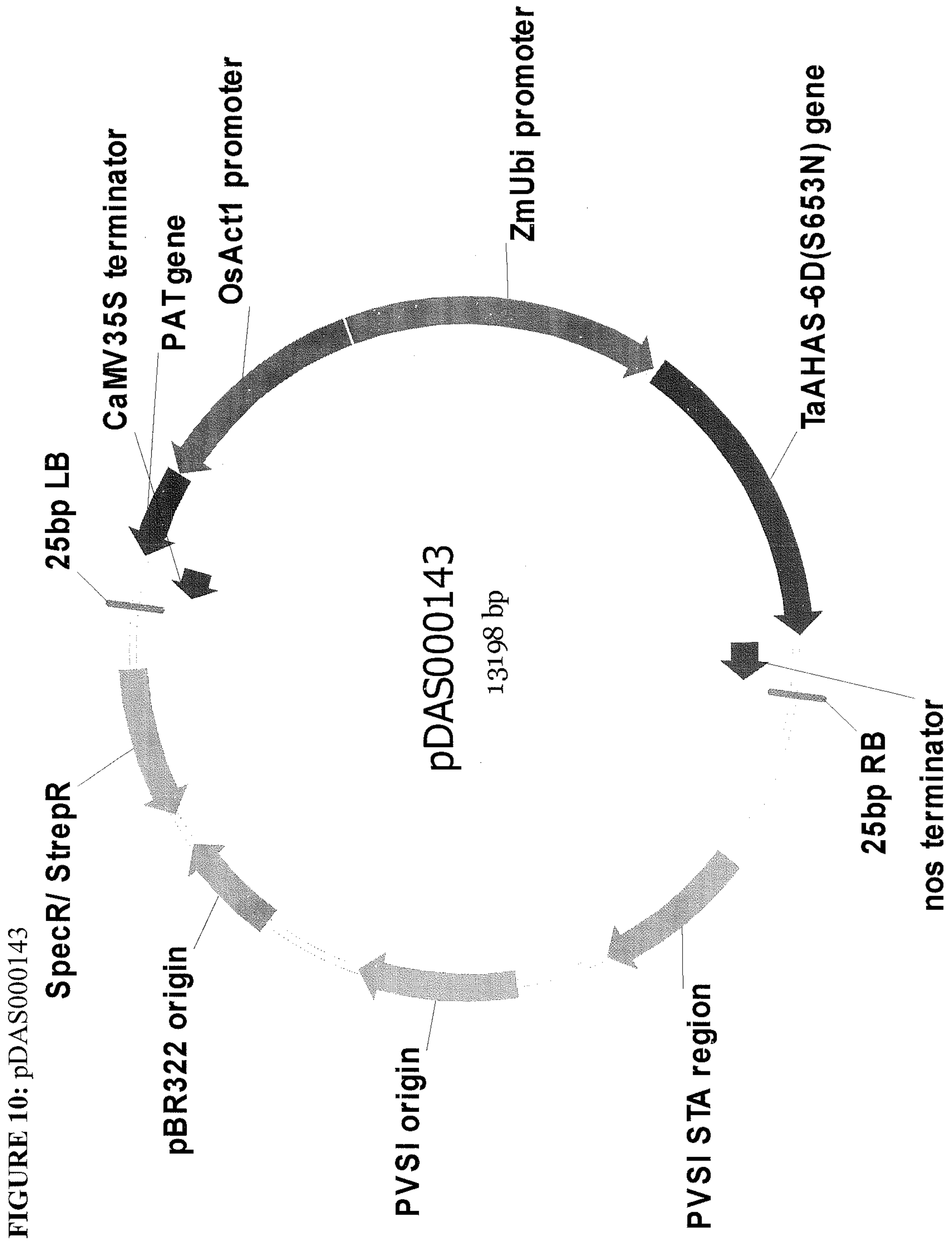



FIGURE 10: pDAS000143

FIGURE 11: pDAS000164

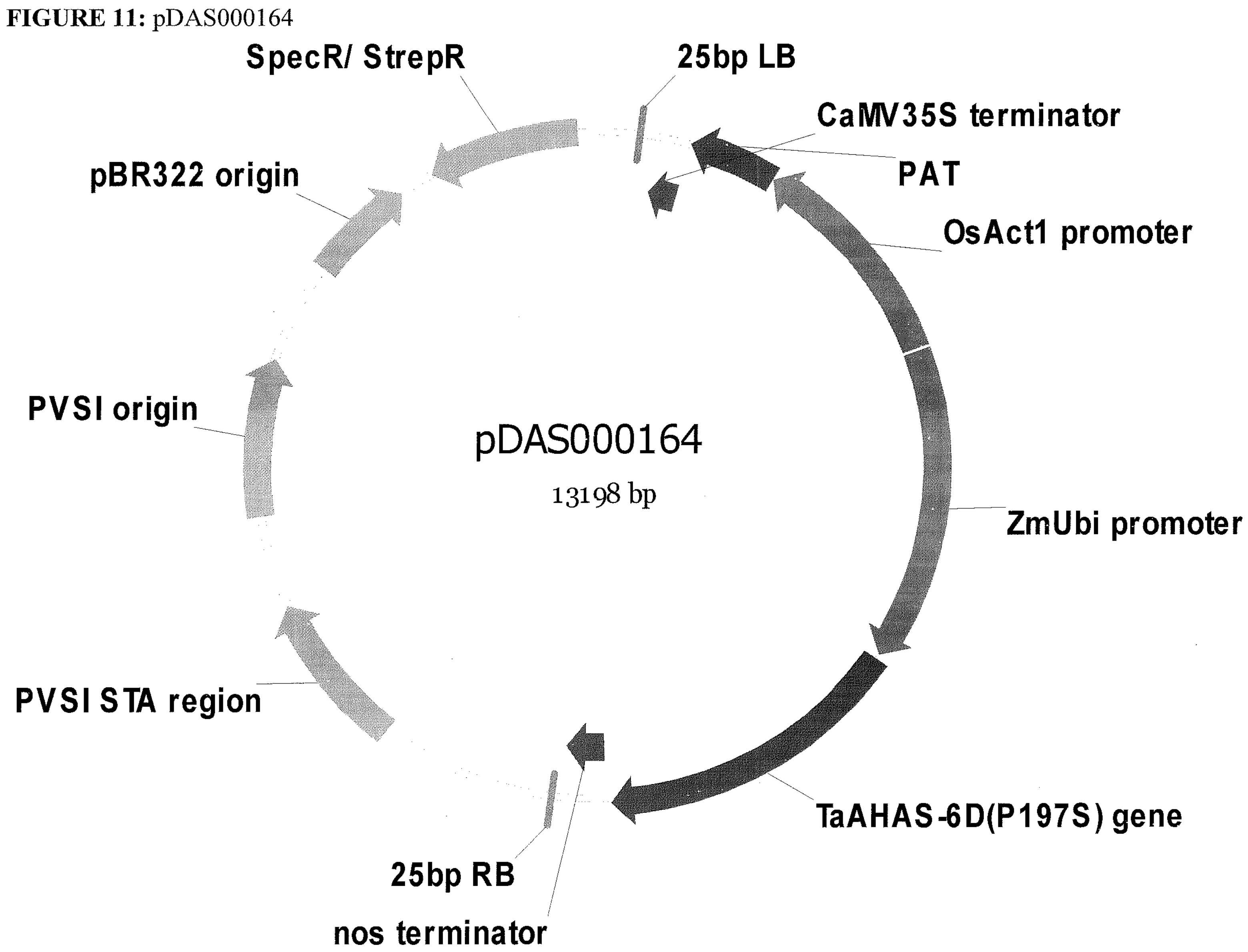
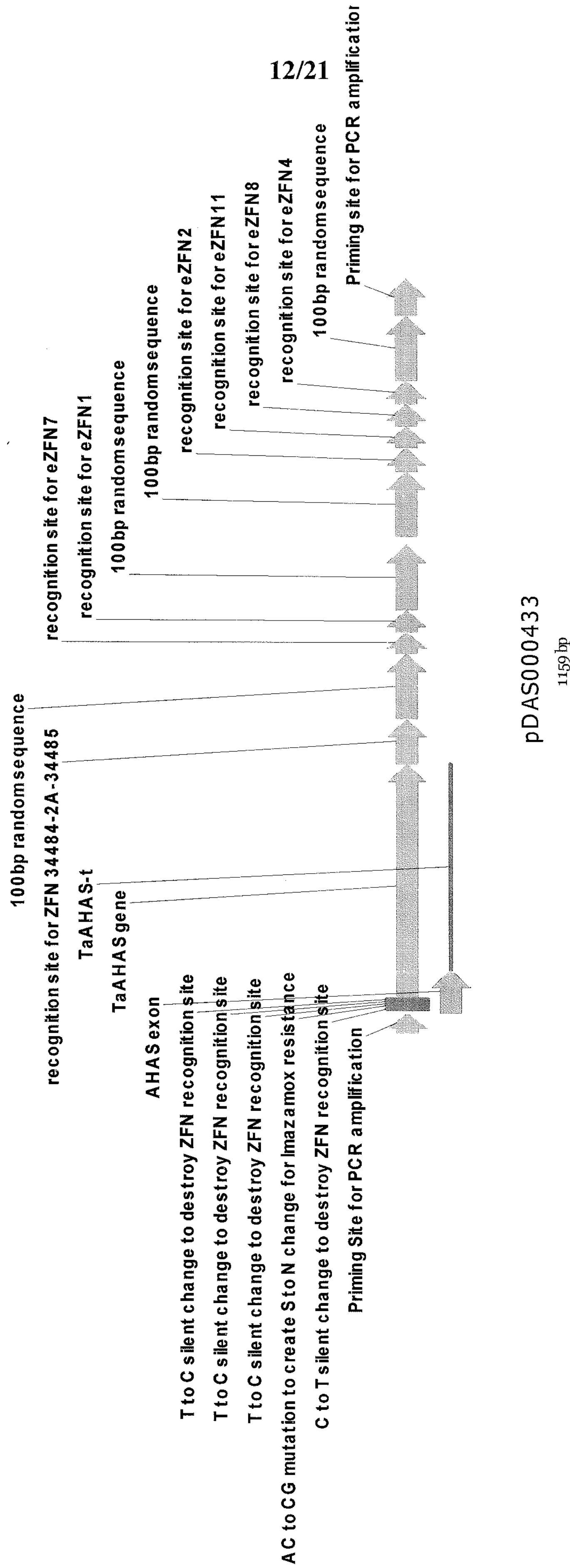


**FIGURE 12:** pDAS000433



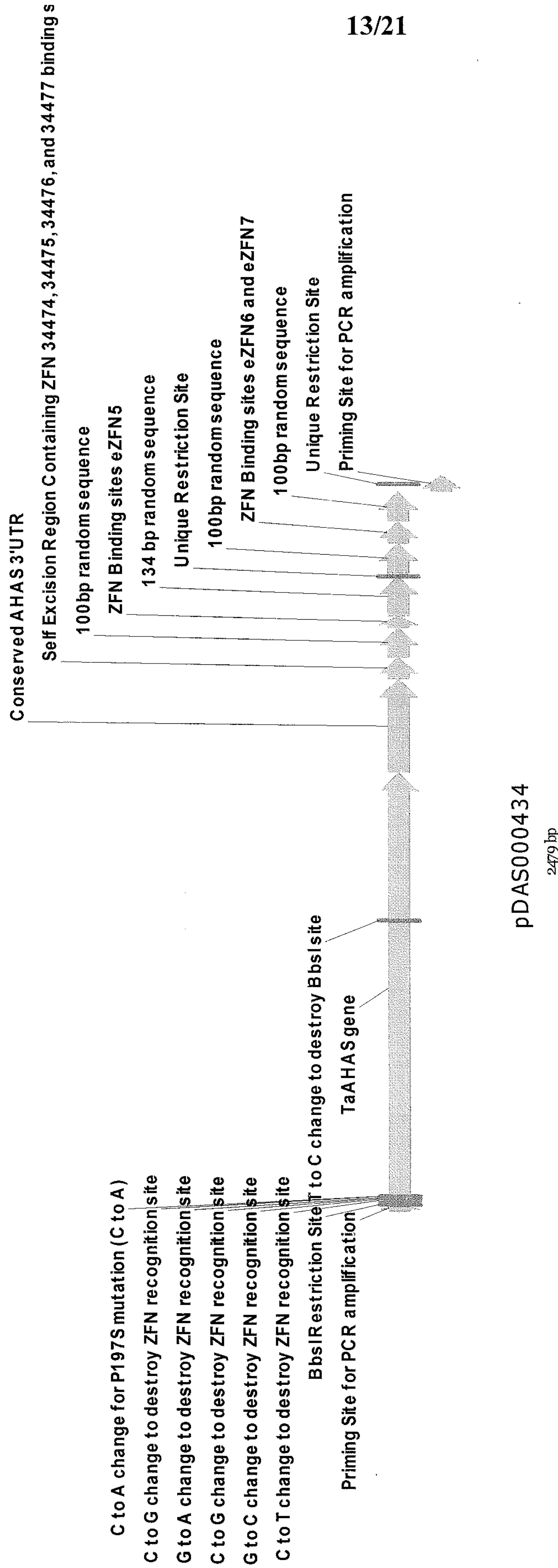
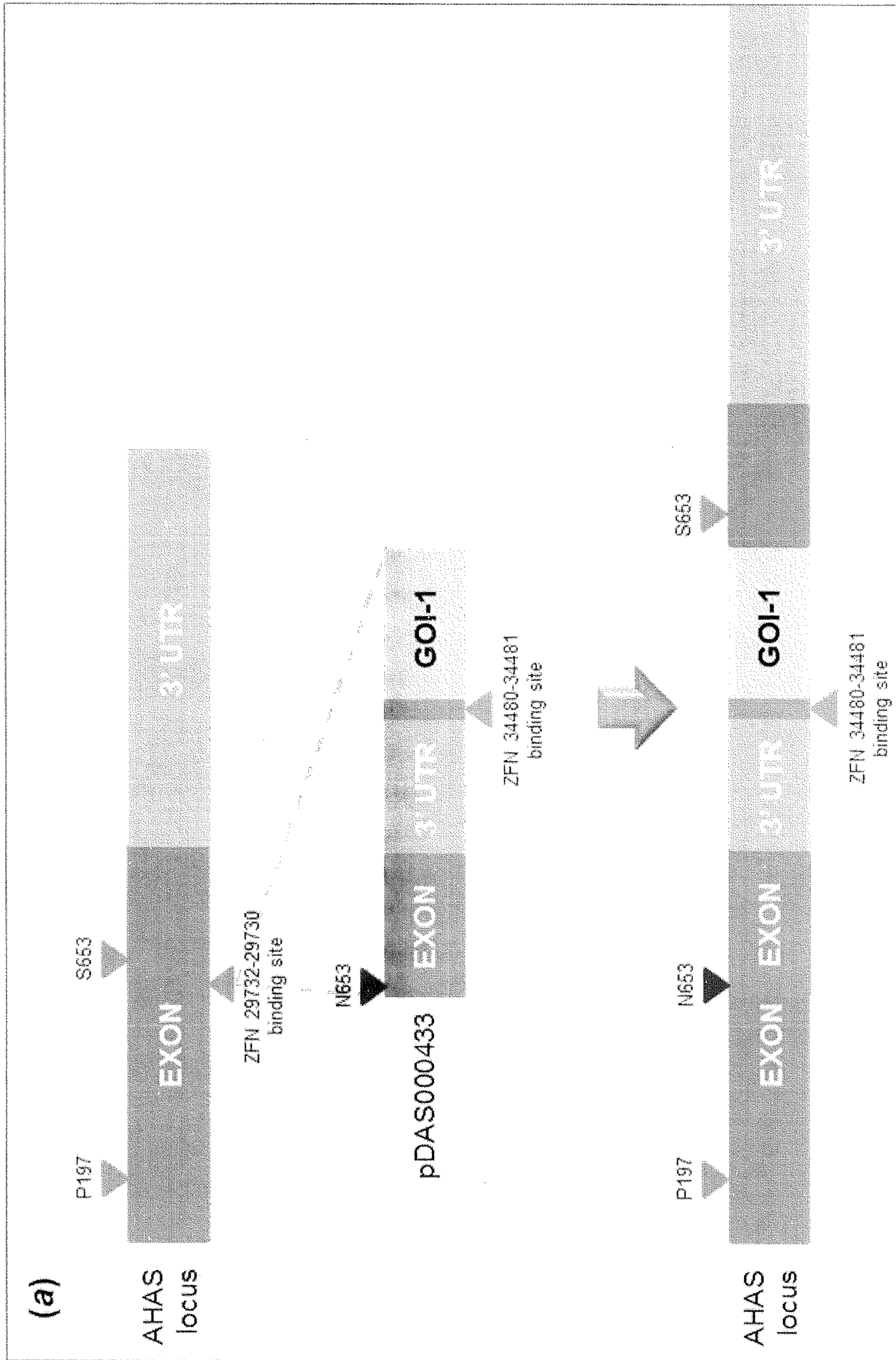



**FIGURE 13:** pDAS000434

FIGURE 14A



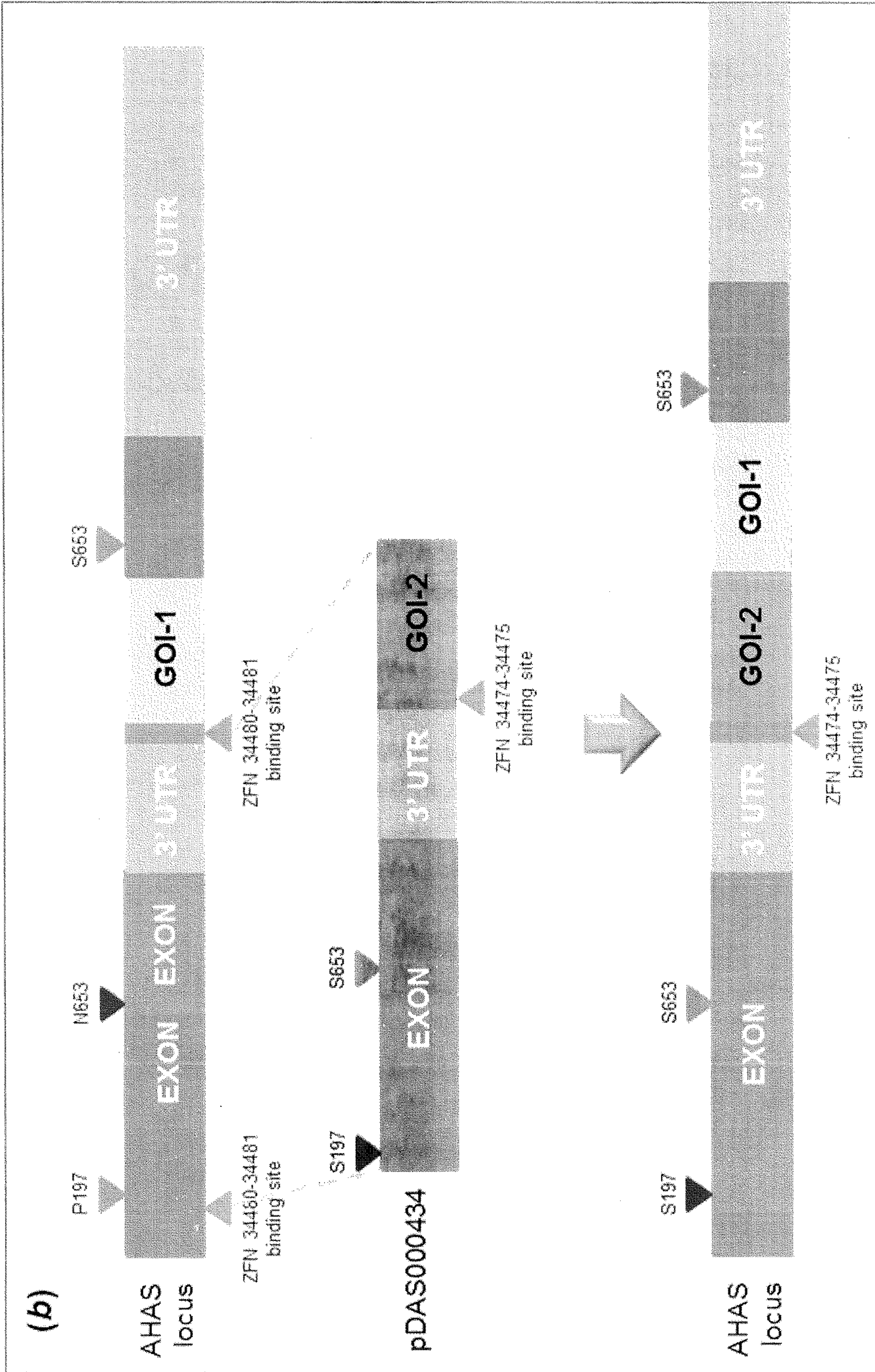
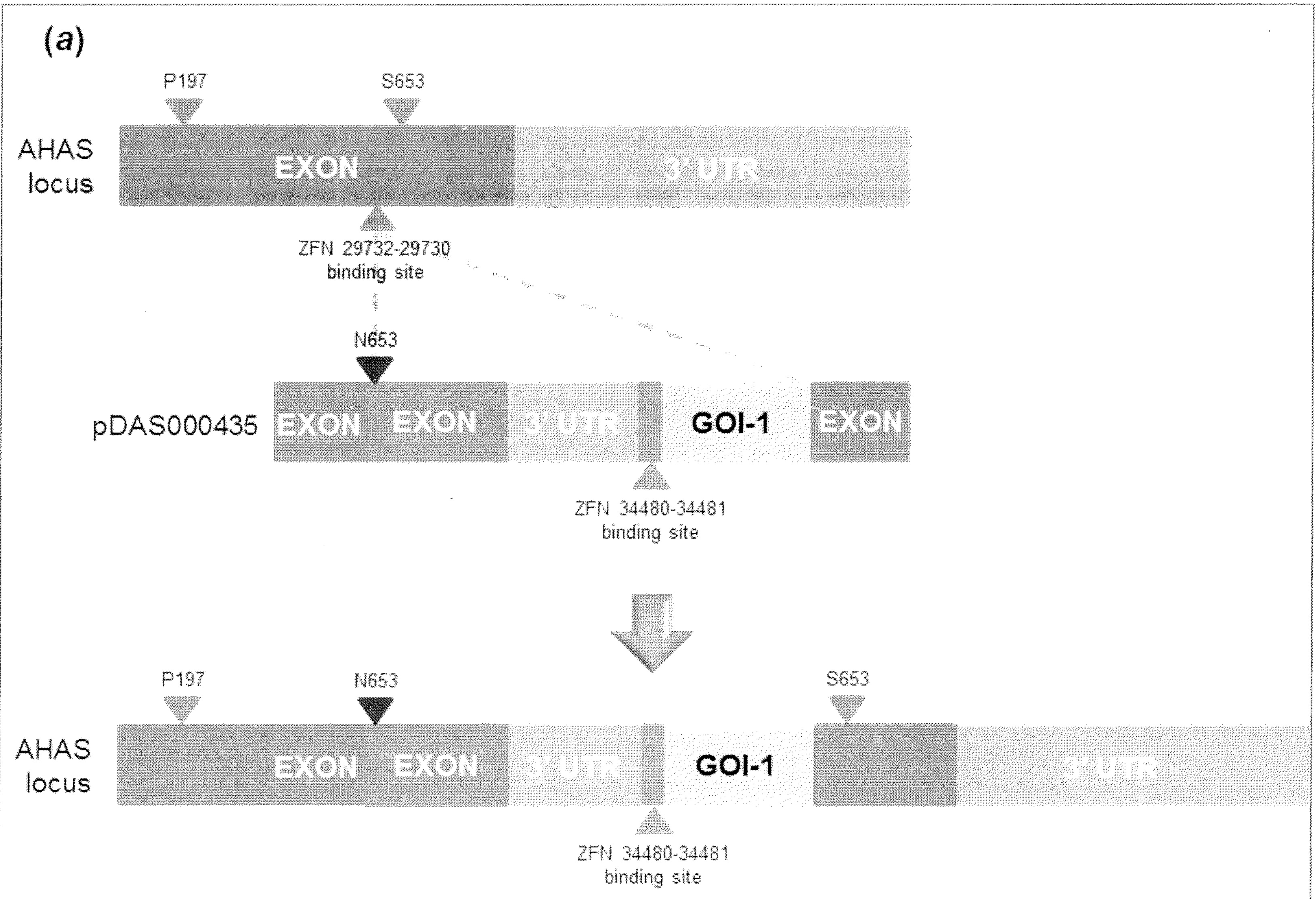



FIGURE 14B

FIGURE 15A



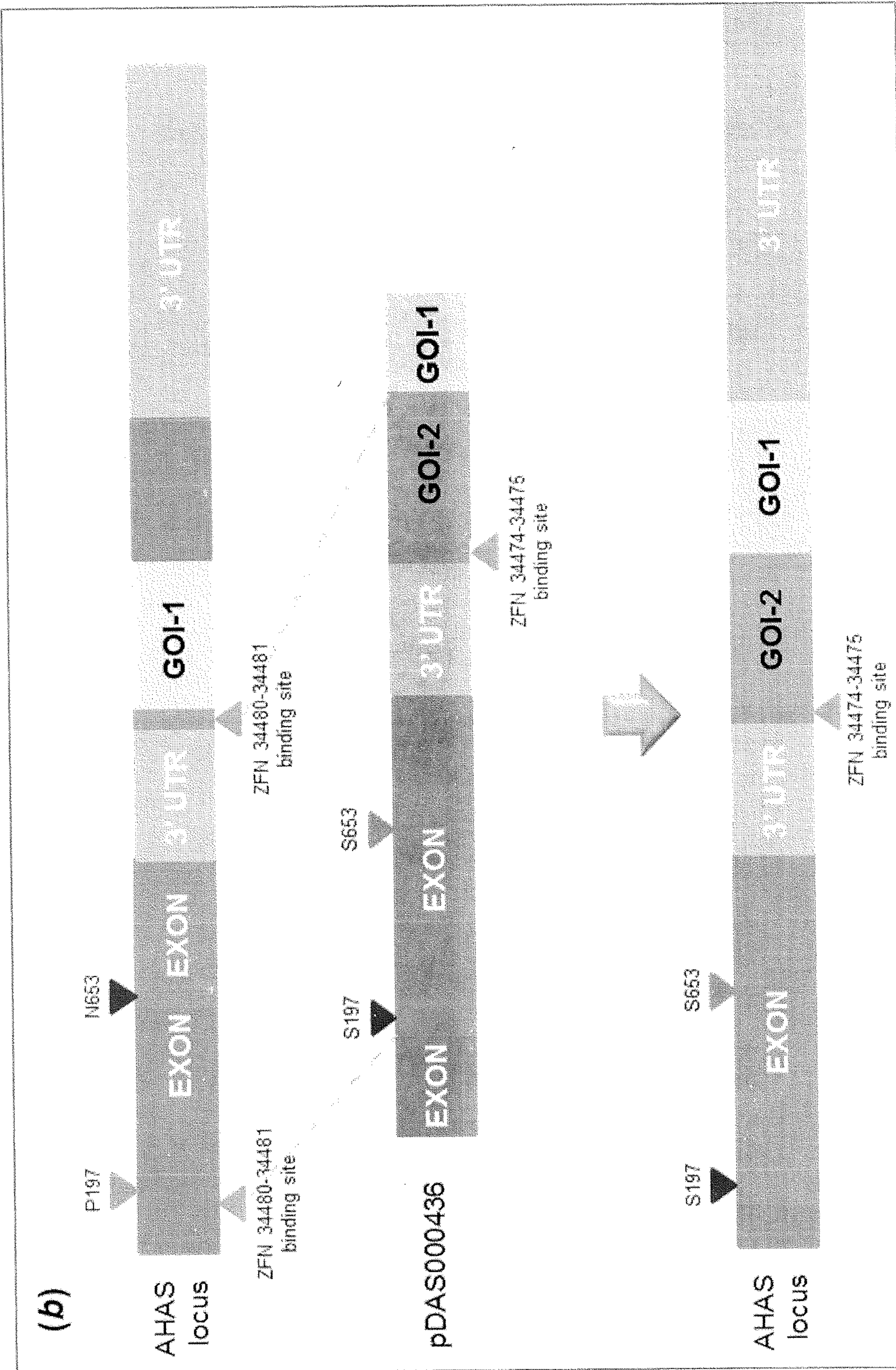



FIGURE 15B

FIGURE 16:

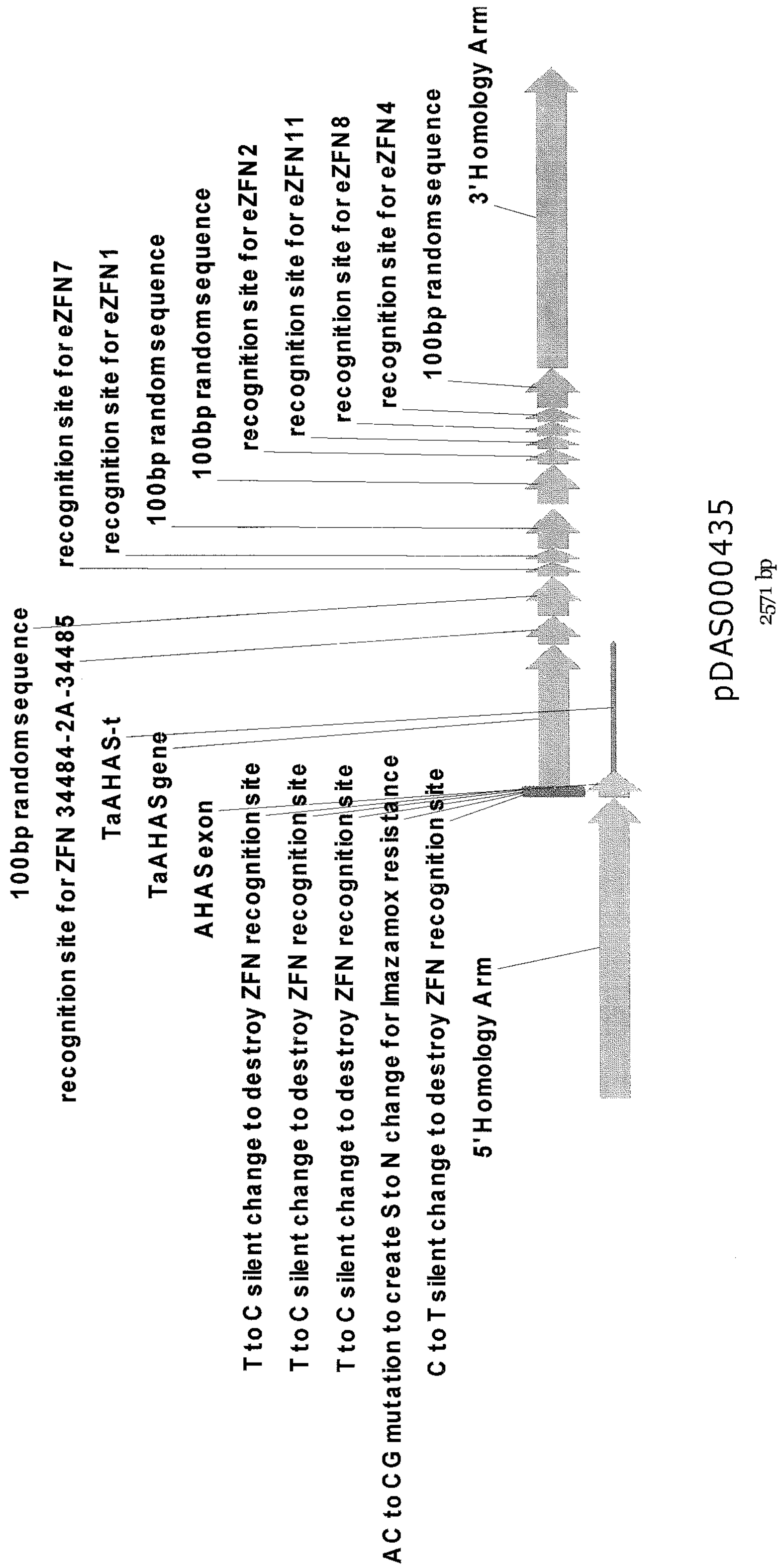


FIGURE 17:

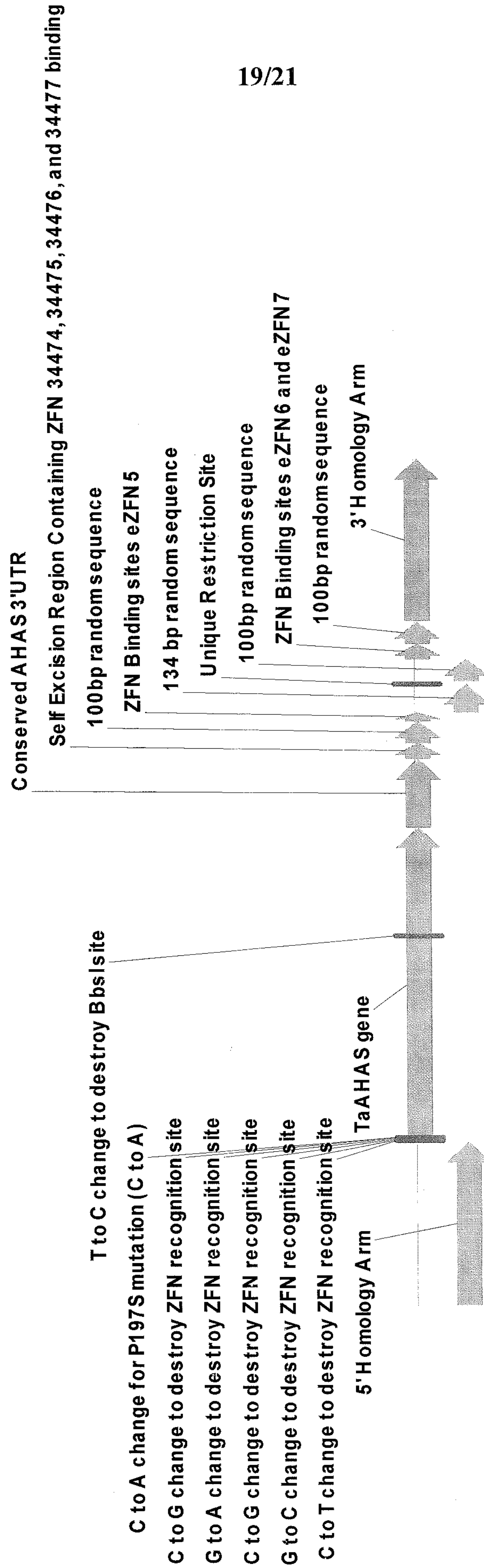


FIGURE 18: pDAS000004

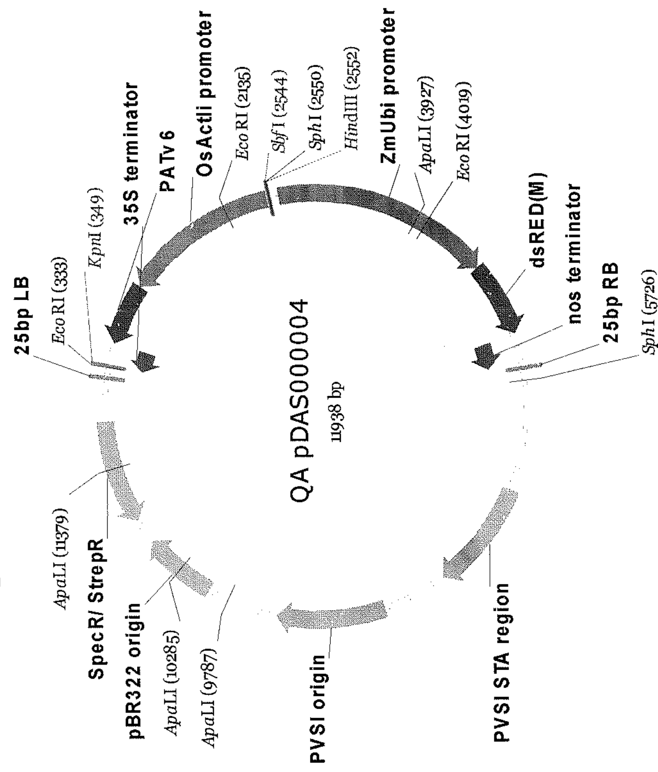
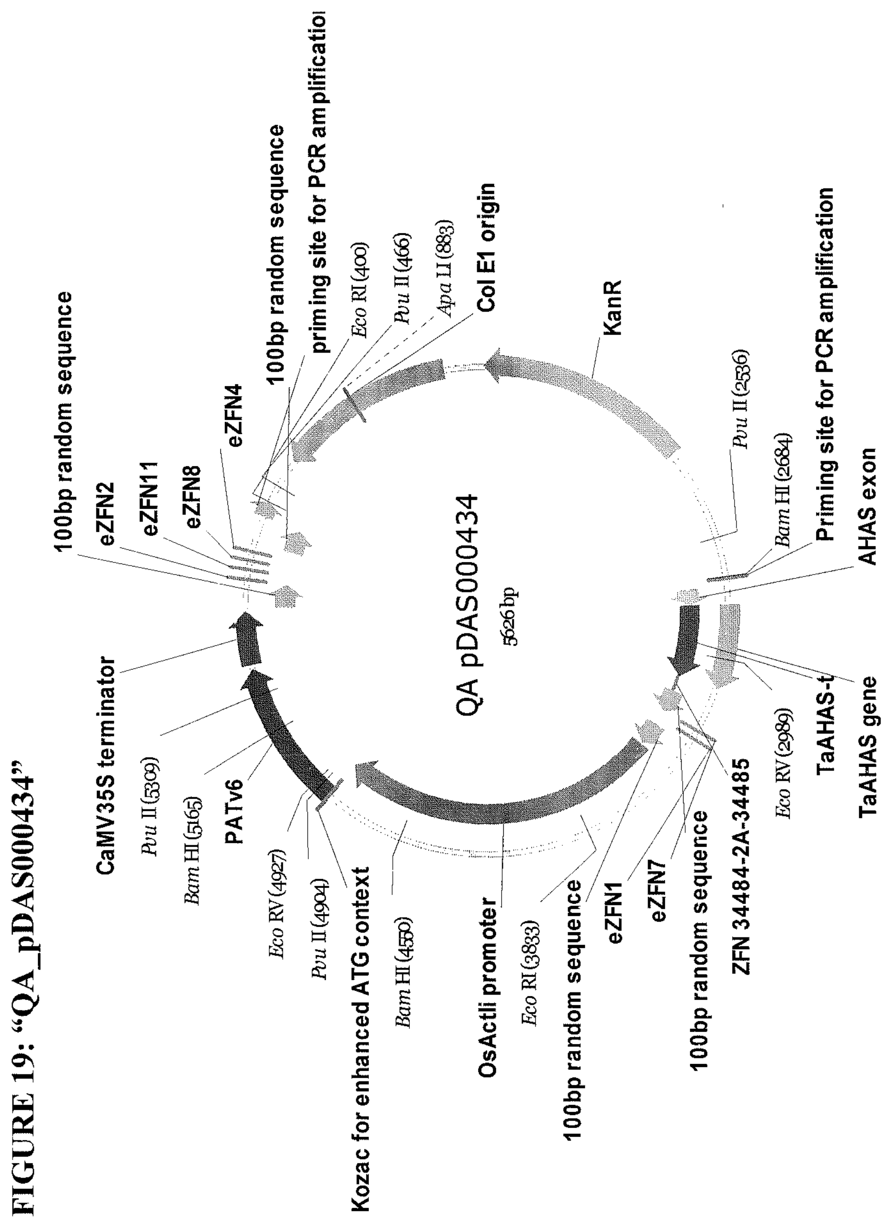
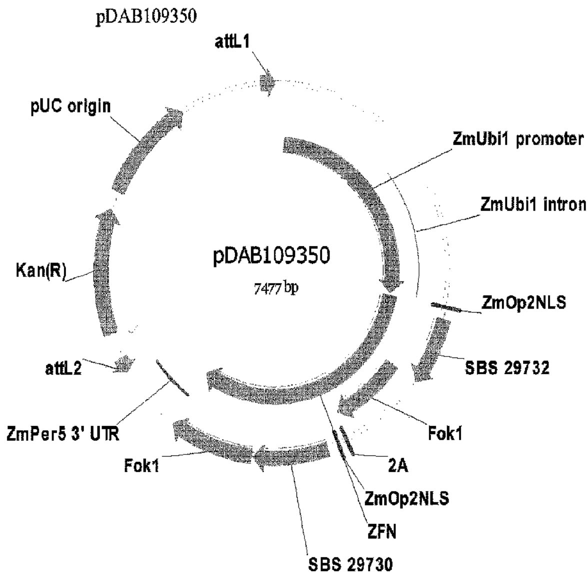


FIGURE 19: “QA_pDAS000434”




pDAB109350
attL1
pUC origin
ZmUbi1 promoter
ZmUbi1 intron
pDAB109350
7477 bp
Kan(R)
ZmOp2NLS
SBS 29732
attL2
Fok1
ZmPer5 3' UTR
2A
Fok1
ZmOp2NLS
ZFN
SBS 29730