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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 March 2014 (13.03.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/039692 A2

(51) International Patent Classification:
C12N 15/66 (2006.01)

(21) International Application Number:
PCT/US2013/058283

(22) International Filing Date:
5 September 2013 (05.09.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/697,886 7 September 2012 (07.09.2012) US

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

(72) Inventors: **COGAN, Noel**; c/o Dow Agrosciences LLC, 9330 Zionsville Road, Indianapolis, IN 46268-1054 (US). **FORSTER, John**; c/o Dow Agrosciences LLC, 9330 Zionsville Road, Indianapolis, IN 46268-1054 (US). **HAYDEN, Matthew**; c/o Dow Agrosciences LLC, 9330 Zionsville Road, Indianapolis, IN 46268-1054 (US). **SAWBRIDGE, Tim**; c/o Dow Agrosciences LLC, 9330 Zionsville Road, Indianapolis, IN 46268-1054 (US). **SPANGENBERG, German**; c/o Dow Agrosciences LLC, 9330 Zionsville Road, Indianapolis, IN 46268-1054 (US). **WEBB, Steven, R.**; c/o Dow Agrosciences LLC, 9330 Zionsville Road, Indianapolis, IN 46268-1054 (US). **GUPTA, Manju**; c/o Dow Agrosciences LLC, 9330 Zionsville Road, Indianapolis, IN 46268-1054 (US). **AINLEY, William, Michael**; c/o Dow Agrosciences LLC, 9330 Zionsville Road, Indianapolis, IN 46268-1054 (US). **HENRY, Matthew, J.**; c/o Dow Agrosciences LLC, 9330 Zionsville Road, Indianapolis, IN 46268-1054 (US). **MILLER, Jeffrey, C.**; c/o Sangamo Biosciences, Inc., Point Richmond Tech Center, 501 Canal Blvd., Suite A100, Richmond, CA 94804 (US). **GUSCHIN, Dmitry, Y.**; c/o Sangamo Biosciences, Inc., Point Richmond Tech Center, 501 Canal Blvd., Suite A100, Richmond, CA 94804 (US).

(74) Agent: **PASTERNAK, Dahna, S.**; Pasternak Patent Law, 1900 Embarcadero Rd., Suite 211, Palo Alto, CA 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, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, 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.

[Continued on next page]

(54) Title: FAD2 PERFORMANCE LOCI AND CORRESPONDING TARGET SITE SPECIFIC BINDING PROTEINS CAPABLE OF INDUCING TARGETED BREAKS

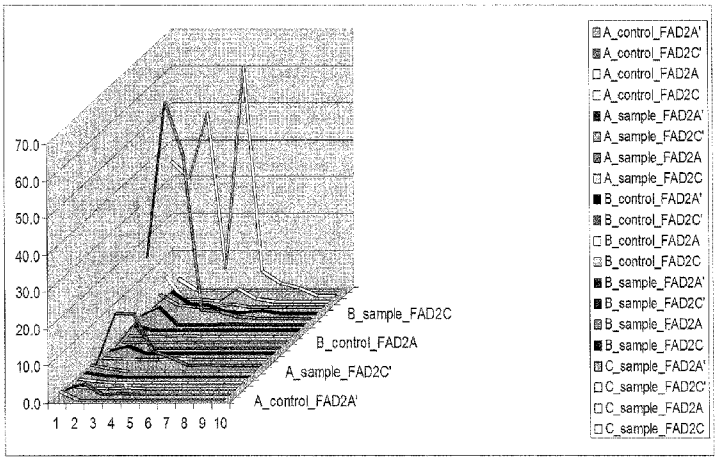


Figure 4

(57) Abstract: A method of gene editing or gene stacking within a FAD2 loci by cleaving, in a site directed manner, a location in a FAD2 gene in a cell, to generate a break in the FAD2 gene and then ligating into the break a nucleic acid molecule associated with one or more traits of interest is disclosed.

WO 2014/039692 A2

(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:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

FAD2 PERFORMANCE LOCI AND CORRESPONDING TARGET SITE SPECIFIC BINDING PROTEINS CAPABLE OF INDUCING TARGETED BREAKS

CROSS -REFERENCE TO RELATED APPLICATIONS

[0001] The present application claims priority to the benefit of U.S. Provisional Patent Application No. 61/697,886, filed September 7, 2012, the disclosure of which is hereby incorporated by reference in its entirety.

FIELD OF THE DISCLOSURE

[0002] The present disclosure relates generally to compositions and methods for use in recombinant plant technology (for example, for generating a transgenic plant). More specifically, the present disclosure relates to plant cells and plants including loci within their genomes that may be used for the site-specific introduction of any nucleic acid of interest.

BACKGROUND

[0003] Many plants are genetically transformed with exogenous nucleic acids (*e.g.*, transgenes) to introduce desirable traits, for example, to improve agricultural value. Examples of improvements in agricultural value that can be achieved through genetic transformation include: improved nutritional quality, increased yield, pest or disease resistance, drought and stress tolerance, improved horticultural quality (*e.g.*, improved pigmentation and/or growth), herbicide resistance, production of industrially useful compounds and/or materials from the plant, and/or production of pharmaceuticals. The introduction of cloned genes into plant cells and recovery of stable fertile transgenic plants can be used to make a genetic modification of a plant stable through multiple generations, and thereby allow the genetic engineering of a crop plant.

[0004] In methods for genetic transformation and transgenic plant production, exogenous DNA is typically randomly introduced into the nuclear or plastid DNA of a eukaryotic plant cell, followed by isolation of cells containing integrated exogenous DNA, and subsequent regeneration of a stably transformed plant. Transgenic plants were typically generated by *Agrobacterium*-mediated transformation technology. Successes with these techniques spurred the development of other methods to introduce a nucleic acid molecule of interest into the genome of a plant, such as PEG-mediated DNA uptake in protoplasts, microprojectile bombardment, and silicon whisker-mediated transformation.

[0005] In all of these plant transformation methods, however, the exogenous nucleic acids incorporated in the plant genome are integrated randomly in the genome of the plant cell, and in unpredictable copy number. 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. For example, the transgenes are frequently integrated in the form of sequence repeats, either of the whole transgene or of parts thereof. Such a complex integration pattern commonly adversely impacts the expression level of the integrated nucleic acid (*e.g.*, by destruction of transcribed RNA through post-transcriptional gene silencing mechanisms, or by inducing methylation of the integrated DNA). Also, the location of the integration site commonly influences the level of expression of the integrated nucleic acid. Moreover, the integration of the exogenous DNA may have a disruptive effect on the region of the genome where the integration occurs, and thereby influence or disturb the normal function of that target region to produce undesirable side-effects. The combination of factors including the foregoing results in a wide variation in the level of expression of transgene or exogenous DNA (and overall agronomic quality) between different transgenic plant cell and plant lines, even those created by the same methods. Because the integration is random, these effects are not able to be controlled by the practitioner while he or she attempts to produce a new plant with desirable characteristics.

[0006] The foregoing considerations necessitate that, whenever the effects of introducing a particular exogenous nucleic acid into a plant is investigated, a large number of transgenic plant lines must be generated and analyzed in order to obtain significant results. Likewise, in the generation of a transgenic plant containing a particular integrated nucleic acid so as to provide the transgenic plant with a desired phenotype, a large population of independently created transgenic plant lines must be created to allow the selection of a plant line with optimal expression of the nucleic acid, and with minimal or no side-effects on the overall phenotype and performance of the transgenic plant. These practical considerations take on added importance in transgenic plants created by inserting multiple exogenous nucleic acids (*i.e.*, gene stacking). In such plants, phenomena such as post-transcriptional gene silencing may be amplified.

[0007] Several methods have been developed in an effort to control transgene insertion in plants. *See, e.g.*, Kumar and Fladung (2001) Trends Plant Sci. 6:155-9. These methods rely on homologous recombination-based transgene integration, which has been successfully applied both in prokaryotes and lower eukaryotes. Paszkowski *et al.* (1988) EMBO J. 7:4021-6. However, until recently in plants, the predominant mechanism for transgene integration has been

based on illegitimate recombination, which involves little homology between recombining DNA strands. A major challenge in this area is therefore the detection and selective generation of rare homologous recombination events, which are masked by far more efficient integration events *via* illegitimate recombination. Moreover, even if the selective generation and detection of targeted homologous recombination events is achieved, the event must be targeted to a desirable location in the host genome in order to realize the maximum benefit of this strategy.

[0008] For example, an assumed benefit of targeted genetic transformation is the reduction in event-to-event variability of transgene expression, as compared to transformation events that are obtained from random integration. A further assumed benefit is a significant reduction in the number of events required to screen introduced nucleic acids, sort transformation constructs, and produce events that contribute to desirable overall characteristics in the resulting transgenic plant. A critical factor required to realize these benefits is the identification of specific locations in the genome where transgene performance is consistent, and if possible, where adverse effects on the host plant are eliminated or minimized.

[0009] 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 Publications 20030232410; 20050208489; 20050026157; 20050064474; 20060188987; 20090263900; 20090117617; 20100047805; 20110207221; 20110301073; 2011089775; 20110239315; 20110145940; and International Publication WO 2007/014275, the disclosures of which are incorporated by reference in their entireties for all purposes. 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 Publication No. 20090205083 describes ZFN-mediated targeted modification of a plant EPSPS locus; U.S. Patent Publication No. 20100199389 describes targeted modification of a plant *Zp15* locus and U.S. Patent Publication No. 20110167521 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.

[0010] However, there remains a need for compositions and methods for modifying and/or modulating expression of *FAD2* genes in plants, including generation of plants with targeted insertions of desired transgenes at the *FAD2* locus.

BRIEF SUMMARY OF THE DISCLOSURE

[0011] The present disclosure describes compositions and methods for modulating expression of *FAD2* genes (*e.g.*, in plants, algae, and fungi) and the use of these loci as sites for the targeted integration of a nucleic acid sequence of interest (e.g., an exogenous nucleic acid sequence) into a host cell. In some embodiments, a host cell may contain one or more genomes with one or more FAD2 sequences (*e.g.*, homeologues or paralogs), any or all of which may be selectively modified and/or disrupted. In specific examples, the present disclosure describes FAD2A, FAD2A', FAD2C and FAD2C' genes, as well as corresponding homeologues or paralogs, in *Brassica napus* (*i.e.*, *B. napus* line, DH12075) and their use as loci for targeted integration of a nucleic acid sequence of interest. As described herein, though FAD2 genes are involved in fatty acid biosynthesis in the host, their modification or disruption (e.g., by integration of an exogenous nucleic acid in the FAD2 coding sequence) unexpectedly may have no or minimal adverse effects on the resultant host organism.

[0012] Also described herein is the use of one or more particular FAD2 loci in tandem with a polypeptide capable of effecting cleavage and/or integration of specific nucleic acid sequences within the FAD2 loci. Examples of the use of FAD2 loci in tandem with a polypeptide capable of effecting cleavage and/or integration of the FAD2 loci include a polypeptide selected from the group consisting of zinc finger proteins, meganucleases, TAL domains, TALENs, RNA-guided CRISPR-Cas9, recombinases, leucine zippers, CRISPr/Cas and others known to those in the art. Particular examples include a chimeric ("fusion") protein comprising a site-specific DNA binding domain polypeptide and cleavage domain polypeptide (e.g., a nuclease), such as a ZFN protein comprising a zinc-finger polypeptide and a FokI nuclease polypeptide. For example, described herein is a demonstration of the *in vitro* and *in vivo* efficacy and specificity of particular ZFNs designed to bind and induce double stranded breaks in FAD2A, FAD2A', FAD2C, FAD2C', and in combinations thereof without cleaving corresponding homeologues or paralogs. In some embodiments, particular FAD2 loci may be used with any of the foregoing polypeptides to effect site-specific integration of a nucleic acid of

interest that is subsequently expressed in the host while having a minimal adverse impact on the agronomic performance of the host.

[0013] In certain aspects, described herein are polypeptides comprising a DNA-binding domain that specifically binds to a FAD2 gene. In some embodiments such a polypeptide may also comprise a nuclease (cleavage) domain or half-domain (*e.g.*, a ZFN, a recombinase, a transposase, or a homing endonuclease, including a homing endonuclease with a modified DNA-binding domain, TAL domains, TALENs, RNA-guided CRISPR-Cas9), and/or a ligase domain, such that the polypeptide may induce a targeted double-stranded break, and/or facilitate recombination of a nucleic acid of interest at the site of the break. In particular embodiments, a DNA-binding domain that targets a FAD2 locus may be a DNA-cleaving functional domain. The foregoing polypeptides may be used in some embodiments to introduce an exogenous nucleic acid into the genome of a host organism (e.g., a plant or animal species) at one or more FAD2 loci. In certain embodiments, the DNA-binding domains comprise a zinc finger protein with one or more zinc fingers (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9 or more zinc fingers), and can which is engineered (non-naturally occuring) to bind to any sequence within a *FAD2* gene. 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 *FAD2* gene, for example, as shown in Table 5. The recognition helix regions of exemplary *FAD2*-binding zinc fingers are shown in Table 4. 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).

[0014] Also described herein are methods for disrupting or editing a FAD2 gene. Additionally described herein are genetically modified host organisms (e.g., transgenic plants) produced by methods according to embodiments of the invention. In particular examples, a transgenic organism produced by a method according to an embodiment of the invention may be, without limitation, algae, a fungus, a monocotyledonous plant, a dicotyledonous plant, etc.

[0015] The FAD2 genes disclosed herein may include those found in any plant, algae, or fungi that have one or more FAD2 genes.

[0016] The foregoing and other features will become more apparent from the following detailed description of several embodiments, which proceeds with reference to the accompanying figures.

BRIEF DESCRIPTION OF THE FIGURES

[0017] **Figure 1, panels A to E**, show a sequence alignment of FAD2 gene sequences (SEQ ID NOs:5-8), generated using AlignX®.

[0018] **Figure 2** is a schematic depicting a phylogenetic tree of FAD2 gene sequences generated using Jalview v 2.3 based on neighbor joining distances.

[0019] **Figure 3** shows a plasmid map of pDAB104010 which that is a representative Zinc Finger Nuclease expression cassette. The lay-out of this construct was similar for the other ZFN expression cassettes, wherein the Zinc Finger domains, 24828 and 24829, were exchanged with alternative Zinc Finger domains that are described above.

[0020] **Figure 4** is an example multiple line graph showing number of sequence reads per 10,000 sequence reads with deletions at the target ZFN site. The X axis on the graph denotes number of bases deleted, the Y axis denotes number of sequence reads and the Z axis denotes colour-coded sample identity as described to the right of the graph. Specific example shown is for locus 1 of the FAD2 gene family that contains 3 target ZFN sites, A, B and C with the four gene family members and two control transfections assessed as control samples A and B. The lines listed from top to bottom (A-control_FADA' at the top of the legend to C_sample_FAD2C at the bottom of the legend) are shown on the graph from closest to the labeled X-axis (A_control_FADA') to farthest from the labeled X-axis (C_sample_FAD2C).

[0021] **Figure 5, panels A and B,** show ZFN targeting of FAD2 genes. Figure 5A is a graph depicting data from ZFN targeting locus 4 of the FAD2 gene family. The locus contains two ZFN sites and two requisite control transfections. Figure 5B shows specific sequence context (SEQ ID NOs:471-480) surrounding the ZFN target site, identifying FAD2A and C containing tri-nucleotide repeats of C, T and G, leading to the observed increase in single base deletions through sequencing of the FAD2A and C loci.

[0022] **Figure 6** shows a plasmid map of pDAS000130.

[0023] **Figure 7** shows a plasmid map of pDAS000031.

[0024] **Figure 8** is a schematic showing binding sites of transgene target primers and probe for transgene copy number estimation assay.

[0025] **Figure 9** shows a Sequencher file showing FAD2A ZFN DNA recognition domain (bc12075_Fad2a-r272a2 and bc12075_Fad2a-278a2), and binding sites of ZFN specific primers (FAD2A.UnE.F1 and FAD2A.UnE.R1) and endogenous primers (FAD2A/2C.RB.UnE.F1 and FAD2A/2C.RB.UnE.R1).

[0026] **Figure 10** shows a schematic showing binding sites of endogenous and transgene target primers used in the detection of transgene integration at FAD2A via perfect HDR.

[0027] **Figure 11** is a schematic showing where Kpn1 restriction endonuclease sites would occur in a perfectly edited FAD2A locus, and where FAD2a 5', *hph* and FAD2A 3' Southern probes bind.

[0028] **Figure 12** shows the location and size of Kpn1 fragments, FAD2A 5', hph, FAD2A 3' probes and expected outcomes of Southern analysis for plants that have integration of ETIP at FAD2A locus via HDR.

[0029] **Figure 13** shows representative data output from copy number estimation qPCR. The left hand column represents data obtained from a known T_0 transgenic plant with a single random transgene insert and is used as the calibrator sample to which all other samples are 'normalized' against. The right hand column is a known T_0 transgenic plant with 5 transgene integrations. The insert copy numbers for both plants was determined using Southern analysis. The remaining columns provide copy number estimates for the putative transgenic plants. The columns are labeled as; 1 copy control, 310420, 311819, 311821, 311822, 311823, 311824, 311827, 312524, 312525, 312526, 312527, 312529, 312530, 312532, 313810, 313811, 313905, 313941, 313942, 313944, and 5 copy control. The columns can be used to determine the estimated copy number for each transgenic plant. When using the software to estimate copy numbers, wildtype plants, non-transformed control plants, and plasmid only controls do not result in a copy number as they do not possess a Cq for both the *hph* and *HMG I/Y* target.

[0030] **Figure 14** shows a plasmid map of pDAS000129.

[0031] **Figure 15** shows a schematic for integration of pDAS000129 into the FAD2A locus.

[0032] **Figure 16** shows plasmid map pDAS000097.

[0033] **Figure 17** shows a plasmid map of pDAS000389.

[0034] **Figure 18** shows a plasmid map of pDAS000391.

[0035] **Figure 19** shows a plasmid map of pDAS000392.

[0036] **Figure 20** shows a plasmid map of pDAS000393.

[0037] **Figure 21** shows a plasmid map of pDAS000394.

[0038] **Figure 22** shows a plasmid map of pDAS000395.

[0039] **Figure 23** shows a plasmid map of pDAS000396.

[0040] **Figure 24** shows a plasmid map of pDAS000397.

SEQUENCES

[0041] The nucleic acid sequences are shown using standard letter abbreviations for nucleotide bases, as defined in 37 C.F.R. § 1.822. Only one strand of each nucleic acid sequence is shown, but the complementary strand is understood to be included by any reference to the displayed strand.

DETAILED DESCRIPTION

I. Overview of several embodiments

[0042] Embodiments of the invention establish an approach for targeted integration of exogenous nucleic acids (*e.g.*, transgenes) in a host genome without greatly adversely impacting other phenotypes of the host beyond those affected by the integrated nucleic acid. Some embodiments may be used for "stacking" multiple nucleic acids in a single host genome. Such an approach requires the development and deployment of four inter-connected technologies: targeting technologies allowing the introduction of double stranded breaks in specific genomic DNA locations (*see, e.g.*, Puchta *et al.* (1993) Nucleic Acids Res. 21:5034-40; Siebert and Puchta (2002) Plant Cell 14:1121-31; D'Halluin et al. (2008) Plant Biotechnol. J. 6(1):93-102; Cai *et al.* (2009) Plant Mol. Biol. 69(6):699-709; Shukla *et al.* (2009) Nature 459(7245):437-41); Shan *et al.* (2103) Nature Biotechnol. 31:686-680; Le *et al.* (2013) Nature Biotechnol 31: 688-691; Nekrasov *et al.* (2013) Nature Biotechnol. 31:691-693, Ainely *et al.* (2013) Plant Biotechnol. J. (On Line 19 Aug); delivery technologies allowing the delivery of an optimized exogenous (donor) nucleic acid (Bibikova *et al.* (2003) Science 300(5620):764); integration technologies involving modification of the host genes (located either in the homologous recombination or NHEJ pathways) so as to increase the HDR or NHEJ frequencies for targeted donor DNA integration; analytical tools to enrich and characterize targeted integration events; and specific desired host genomic locations ("performance loci") that are genetically well-defined and that support stable gene expression across generations without greatly adversely affecting the transformed host organism. *See, also,* United States Patent Publications

20030232410; 20050208489; 20050026157; 20050064474; 20060188987; 20090263900; 20090117617; 20100047805; 20110207221; 20110301073; 2011089775; 20110239315; 20110145940; 20080182332; 20090205083; 20100199389; 20110167521. For example, in plants, a performance locus is a locus where the negative impact on the agronomic or quality properties of a transgenic plant wherein a transgene has been integrated at the locus is negligible or non-existent.

[0043] Embodiments described herein take advantage of the unexpected finding that plant *FAD2* genes are performance loci for the targeted integration of exogenous nucleic acids (*e.g.*, gene(s); non-coding DNA sequences, such as an Engineered Landing Pads (ELPs) (U.S. Application 12/011,735) and Engineered Transgene Insertion Platform (ETIP) (pending U.S. Application No: 61/697882); and plant transformation unit(s)). The ubiquitous nature of FAD2 loci in plants, and evidence that alteration or knock-out of FAD2 in canola, corn, sunflower, wheat, cotton, and soybean does not carry an agronomic or quality penalty, identifies FAD2 loci as a broad class of performance loci across commercially-relevant plant species.

[0044] Some embodiments utilize site-specific double-stranded DNA cleavage at a *FAD2* locus, for example, resulting from the delivery and expression of a target-site specific DNA recognition and cleavage protein. In specific examples, such a FAD2-specific DNA recognition and cleavage protein may be, for example and without limitation, a ZFN; a TALEN; RNA-guided CRISPR-Cas9, a recombinase (*e.g.*, Cre, Hin, RecA, Tre, and FLP recombinases); a meganuclease, and an engineered protein derived from any of the foregoing or their equivalents. Cleavage may also be effected using the CRISPR/Cas system with an engineered crRNA/tracr RNA ('single guide RNA') to guide specific cleavage. In some embodiments, such a double-strand break may be repaired via integration of a donor nucleic acid at the cleavage site within the *FAD2* performance locus, for example, by Homology Directed Repair (HDR) or Non-Homologous End Joining (NHEJ).

[0045] This disclosure exemplifies the utility of *FAD2* loci as performance loci, for example, by describing the FAD2A, 2A', 2C or 2C' locus in canola (*Brassica napus*), and corresponding FAD2-specific ZFNs that may be utilized to integrate an exogenous nucleic acid at the FAD2A, 2A', 2C or 2C' locus.

[0046] Embodiments of the present invention address many unsolved problems in the art. For example, the selectivity of the targeted integration approach described herein may reduce or eliminate the necessity of repeated field trials required for elimination of unwanted transgenic events, which trials are costly due to the resources involved and the burdensome

regulatory requirements in this area. Furthermore, the targeted DNA integration approaches described herein may be particularly beneficial in the process of transgene stacking.

[0047] Although the native nucleotide sequence at an endogenous FAD2 locus may be used to directly target a nucleic acid of interest, in some embodiments, a nucleic acid may first be targeted to at least one FAD2 locus of the host, such that the integration of further nucleic acid molecules of interest into the host is facilitated. In other examples, nucleotide sequences that are not homologous to native sequences of the host organism (*e.g.*, essentially randomly generated nucleic acid sequences) that flank a DNA recognition site (*e.g.*, zinc finger recognition sites) may be utilized.

II. Terms

[0048] As used in this application, including the claims, terms in the singular and the singular forms, "a," "an," and "the," for example, include plural referents, unless the content clearly dictates otherwise. Thus, for example, a reference to "plant," "the plant," or "a plant" also refers to a plurality of plants. Furthermore, depending on the context, use of the term, "plant," may also refer to genetically-similar or identical progeny of that plant. Similarly, the term, "nucleic acid," may refer to many copies of a nucleic acid molecule. Likewise, the term, "probe," may refer to many similar or identical probe molecules.

[0049] Numeric ranges are inclusive of the numbers defining the range, and expressly include each integer and non-integer fraction within the defined range. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art.

[0050] In order to facilitate review of the various embodiments described in this disclosure, the following explanation of specific terms is provided:

[0051] Isolated: An "isolated" biological component (such as a nucleic acid or protein) has been substantially separated, produced apart from, or purified away from other biological components in the cell of the organism in which the component naturally occurs (*i.e.*, other chromosomal and extra-chromosomal DNA and RNA, and proteins), while effecting a chemical or functional change in the component (*e.g.*, a nucleic acid may be isolated from a chromosome by breaking chemical bonds connecting the nucleic acid to the remaining DNA in the chromosome). Nucleic acid molecules and proteins that have been "isolated" include nucleic acid molecules and proteins purified by standard purification methods. The term also embraces

nucleic acids and proteins prepared by recombinant expression in a host cell, as well as chemically-synthesized nucleic acid molecules, proteins, and peptides.

[0052] Cross: As used herein in regard to plants, the term "cross" or "crossed" refers to the fusion of gametes via pollination to produce progeny (*e.g.*, cells, seeds, and plants). This term encompasses both sexual crosses (*i.e.*, the pollination of one plant by another) and selfing (*i.e.*, self-pollination, for example, using pollen and ovule from the same plant).

[0053] Backcrossing: Backcrossing methods may be used to introduce a nucleic acid sequence into a plant. This technique has been widely used for decades to introduce new traits into plants. Jensen, N., Ed. Plant Breeding Methodology, John Wiley & Sons, Inc., 1988. In a typical backcross protocol, the original variety of interest (recurrent parent) is crossed to a second variety (non-recurrent parent) that carries a nucleic acid sequence of interest to be transferred. The resulting progeny from this cross are then crossed again to the recurrent parent, and the process is repeated until a plant is obtained wherein essentially all of the desired morphological and physiological characteristics of the recurrent plant are recovered in the converted plant, in addition to the transferred nucleic acid sequence from the non-recurrent parent.

[0054] Introgression: As used herein, the term "introgression" refers to the transmission of an allele (or modified allele comprising an exogenous nucleic acid) into a genetic background at a particular locus. In some embodiments, introgression of a specific allele at the locus may occur by transmitting the allele to at least one progeny via a sexual cross between two parents of the same species, where at least one of the parents has the specific allele form in its genome. Progeny comprising the specific allele may be repeatedly backcrossed to a line having a desired genetic background. Backcross progeny may be selected for the specific allele form, so as to produce a new variety wherein the specific allele form has been fixed in the genetic background. In some embodiments, introgression of a specific allele may occur by recombination between two donor genomes (*e.g.*, in a fused protoplast), where at least one of the donor genomes has the specific allele form in its genome. Introgression may involve transmission of a specific allele form that may be, for example and without limitation, a disrupted or modified allele; a transgene; a PTU; and an ELP.

[0055] Germplasm: As used herein, the term "germplasm" refers to genetic material of or from an individual plant, a group of plants (*e.g.*, a plant line, variety, and family), and a clone derived from a plant or group of plants. A germplasm may be part of an organism or cell, or it may be separate (*e.g.*, isolated) from the organism or cell. In general, germplasm provides

genetic material with a specific molecular makeup that is the basis for hereditary qualities of the plant. As used herein, "germplasm" refers to cells of a specific plant; seed; tissue of the specific plant (*e.g.*, tissue from which new plants may be grown); and non-seed parts of the specific plant (*e.g.*, leaf, stem, pollen, and cells). As used herein, the term "germplasm" is synonymous with "genetic material," and it may be used to refer to seed (or other plant material) from which a plant may be propagated. A "germplasm bank" may refer to an organized collection of different seed or other genetic material (wherein each genotype is uniquely identified) from which a known cultivar may be cultivated, and from which a new cultivar may be generated.

[0056] Gene: As used herein, the term "gene" (or "genetic element") may refer to a heritable genomic DNA sequence with functional significance. A gene may be a native nucleic acid, or a nucleic acid that has been integrated into the genome. The term "gene" may also be used to refer to, for example and without limitation, a cDNA and/or an mRNA encoded by a heritable genomic DNA sequence.

[0057] Nucleic acid molecule: As used herein, the term "nucleic acid molecule" may refer to a polymeric form of nucleotides (*i.e.*, ribonucleotides, deoxyribonucleotides, and/or a modified form of either of the foregoing). A "nucleic acid molecule" as used herein is synonymous with "nucleic acid" and "polynucleotide." The term includes both sense and anti-sense strands of RNA, cDNA, genomic DNA, and synthetic forms and mixed polymers thereof. The term includes any topological conformation, including single-stranded, double-stranded, partially duplexed, triplexed, hairpinned, circular, and padlocked conformations. A nucleic acid molecule can include either or both of naturally-occurring and modified nucleotides. Such nucleotides may be linked together by naturally-occurring and/or non-naturally-occurring nucleotide linkages.

[0058] Nucleic acid molecules may be modified chemically or biochemically, or may contain derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example and without limitation: labels; methylation; substitution of one or more of the naturally-occurring nucleotides with an analog; and inter-nucleotide modifications (*e.g.*, uncharged linkages, for example, methyl phosphonates, phosphotriesters, phosphoramidates, and carbamates; charged linkages, for example, phosphorothioates and phosphorodithioates; pendent moieties, for example, peptides; intercalators, for example, acridine and psoralen; chelators; alkylators; and modified linkages, for example, alpha anomeric nucleic acids).

[0059] Exogenous: An "exogenous" molecule is a molecule that is not native to a specified system (*e.g.*, a germplasm, variety, elite variety, and/or plant) with respect to nucleotide sequence and /or genomic location (*i.e.*, locus) for a polynucleotide (and with respect to amino acid sequence and/or cellular localization for a polypeptide). In embodiments, exogenous or heterologous polynucleotides or polypeptides may be molecules that have been artificially supplied to a biological system (*e.g.*, a plant cell, a plant gene, a particular plant species or variety, and/or a plant chromosome) and are not native to that particular biological system. Thus, the designation of a nucleic acid as "exogenous" may indicate that the nucleic acid originated from a source other than a naturally-occurring source, or it may indicate that the nucleic acid has a non-natural configuration, genetic location, or arrangement of elements.

[0060] In contrast, for example, a "native" or "endogenous" nucleic acid is a nucleic acid (*e.g.*, a gene) that does not contain a nucleic acid element other than those normally present in the chromosome or other genetic material on which the nucleic acid is normally found in nature. An endogenous gene transcript is encoded by a nucleotide sequence at its natural chromosomal locus, and is not artificially supplied to the cell.

[0061] Operably linked: A first nucleic acid sequence is operably linked with a second nucleic acid sequence when the first nucleic acid sequence is in a functional relationship with the second nucleic acid sequence. For instance, a promoter is operably linked with a coding sequence when the promoter affects the transcription or expression of the coding sequence. When recombinantly produced, operably linked nucleic acid sequences are generally contiguous and, where necessary to join two protein-coding regions, in the same reading frame. However, elements need not be contiguous to be operably linked.

[0062] Promoter: A promoter is a region of DNA that generally is located upstream (towards the 5' region) of a nucleic acid that enhances transcription of the nucleic acid. Promoters permit the proper activation or repression of the nucleic acid(s) with which they are operably linked. A promoter contains specific sequences that are recognized by transcription factors. These factors bind to the promoter DNA sequences and result in the recruitment of RNA polymerase, the enzyme that synthesizes the RNA from the coding region of the nucleic acid. Transformed: A vector "transforms" or "transduces" a cell when it transfers nucleic acid molecules into the cell. A cell is "transformed" by a nucleic acid molecule when the nucleic acid molecule becomes stably replicated by the cell, either by incorporation of the nucleic acid molecule into the cellular genome or by episomal replication. As used herein, the term "transformation" encompasses all techniques by which a nucleic acid molecule can be

introduced into a cell. Examples include, but are not limited to: transfection with viral vectors; transformation with plasmid vectors; electroporation (Fromm *et al.* (1986) Nature 319:791-3); lipofection (Felgner *et al.* (1987) Proc. Natl. Acad. Sci. USA 84:7413-7); microinjection (Mueller *et al.* (1978) Cell 15:579-85); *Agrobacterium*-mediated transfer (Fraley *et al.* (1983) Proc. Natl. Acad. Sci. USA 80:4803-7); direct DNA uptake; and microprojectile bombardment (Klein *et al.* (1987) Nature 327:70).

[0063] Introduced: As used herein, the term "introduced," when referring to translocation of an exogenous nucleic acid into a cell, refers to the incorporation of the nucleic acid into the cell using any methodology available in the art. This term encompasses nucleic acid introduction methods including, for example and without limitation, transfection; transformation; and transduction.

[0064] Transgene: As used herein, the term "transgene" refers to an exogenous nucleic acid coding sequence of interest. For example, a transgene may encode an industrially or pharmaceutically useful compound, or an expression product that contributes to a desirable agricultural trait (*e.g.*, herbicide resistance or pest resistance). In a further example, a transgene may be an antisense nucleic acid, wherein expression of the antisense nucleic acid inhibits expression of a target nucleic acid sequence. A transgene may comprise regulatory sequences operably linked to the transgene (*e.g.*, a promoter). In some embodiments, a nucleic acid molecule of interest to be introduced by site-specific targeting at a FAD2 locus is a transgene. However, in other embodiments, a nucleic acid molecule of interest may be a PTU, an ELP, an ETIP, or an endogenous nucleic acid sequence (*e.g.*, wherein additional, exogenous genomic copies of the endogenous nucleic acid sequence are desired).

[0065] Elements can also include DNA that encodes for a structural RNA, such as shRNA. Such RNA can modify exogenous or endogenous genes including, but not limited to affecting postings or conferring herbicide resistance.

[0066] Recombinant: As used herein, the term "recombinant" refers to a material (*e.g.*, nucleic acid, gene, polynucleotide, and/or polypeptide) that has been altered by human intervention. For example, the arrangement of the parts or elements of a recombinant molecule may not be a native arrangement, and/or the primary sequence of the recombinant molecule may have been changed from its native sequence, *e.g.*, to optimize its expression and/or activity. A material may be altered to produce a recombinant material within or removed from its natural environment or state. As one example, an open reading frame of a nucleic acid is recombinant if the nucleotide sequence of the open reading frame has been removed from it natural context and

cloned into an artificial nucleic acid molecule (*e.g.*, a vector). Protocols and reagents to produce recombinant molecules (*e.g.*, recombinant nucleic acids) are common in the art, and their use is routine. The term "recombinant" may also refer herein to a cell or organism that comprises recombinant material (*e.g.*, a plant and/or plant cell that comprises a recombinant nucleic acid). In some examples, a recombinant organism is a transgenic organism.

[0067] Vector: As used herein, the term "vector" refers to a polynucleotide or other molecule that is capable of transferring at least one nucleic acid segment(s) into a cell. A vector may optionally comprise components/elements that mediate vector maintenance and/or enable its intended use (*e.g.*, sequences necessary for replication, genes imparting drug or antibiotic resistance, a multiple cloning site, and/or operably linked promoter/enhancer elements that enable the expression of a cloned gene). Vectors may be derived, for example, from plasmids, bacteriophages, or plant or animal viruses. A "cloning vector," "shuttle vector," or "subcloning vector" generally comprises operably linked elements to facilitate cloning or subcloning steps (*e.g.*, a multiple cloning site containing multiple restriction endonuclease sites).

[0068] Expression Vector: The term "expression vector," as used herein, refers to a vector comprising operably linked polynucleotide sequences that may facilitate expression of a coding sequence in a particular host organism. For example, a bacterial expression vector may facilitate expression of a coding sequence in a bacterium. Likewise, a plant expression vector may facilitate expression of a coding sequence in a plant cell. Polynucleotide sequences that facilitate expression in prokaryotes may include, for example and without limitation, a promoter; an operator; and a ribosome binding site. Eukaryotic expression vectors (*e.g.*, a plant expression vector) may comprise, for example, promoters; enhancers; termination signals; and polyadenylation signals (and other sequences) that are generally different from those used in prokaryotic expression vectors.

[0069] Sequence identity: The term "sequence identity" or "identity," as used herein in the context of two nucleic acid or polypeptide sequences, refers to the residues in the two sequences that are the same when aligned for maximum correspondence over a specified comparison window. A value of sequence identity may be determined by comparing two optimally aligned sequences (*e.g.*, nucleic acid sequences and amino acid sequences) over a comparison window, wherein the portion of the sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The sequence identity is calculated as a percentage by determining the number of positions at which the

identical nucleotide or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the comparison window, and multiplying the result by 100 to yield the percentage of sequence identity.

[0070] Methods for aligning sequences for comparison are well-known in the art. Various programs and alignment algorithms are described in, for example: Smith and Waterman (1981) Adv. Appl. Math. 2:482; Needleman and Wunsch (1970) J. Mol. Biol. 48:443; Pearson and Lipman (1988) Proc. Natl. Acad. Sci. U.S.A. 85:2444; Higgins and Sharp (1988) Gene 73:237-44; Higgins and Sharp (1989) CABIOS 5:151-3; Corpet *et al.* (1988) Nucleic Acids Res. 16:10881-90; Huang *et al.* (1992) Comp. Appl. Biosci. 8:155-65; Pearson *et al.* (1994) Methods Mol. Biol. 24:307-31; Tatiana *et al.* (1999) FEMS Microbiol. Lett. 174:247-50. A detailed consideration of sequence alignment methods and homology calculations can be found in Altschul *et al.* (1990) J. Mol. Biol. 215:403-10.

[0071] The National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST™; Altschul *et al.* (1990)) may be used to align sequences, and it is available from several sources, including the National Center for Biotechnology Information (Bethesda, MD), and on the internet, for use in connection with several sequence analysis programs. A description of how to determine sequence identity using this program is available on the internet under the "help" section for BLAST™. For comparisons of nucleic acid sequences, the "Blast 2 sequences" function of the BLAST™ (Blastn) program may be employed using the default parameters. Nucleic acid sequences with greater similarity to the reference sequences will show increasing percentage identity when assessed by this method.

[0072] As used herein, the term "substantially identical" may refer to nucleotide sequences that are more than 80% identical. For example, a substantially identical nucleotide sequence may be at least 85%, at least 86%; at least 87%; at least 88%; at least 89%; at least 90%; at least 91%; at least 92%; at least 93%; at least 94%; at least 95%; at least 96%; at least 97%; at least 98%; at least 99%; or at least 99.5% identical to the reference sequence.

[0073] Locus: As used herein, the term "locus" refers to a position on a genome that corresponds to a measurable characteristic (*e.g.*, a trait). In some embodiments, a locus of particular interest is the genomic position of a FAD2 gene, where disruption of the gene reduces or eliminates expression of the mRNA transcribed from the wild-type gene. A locus may be defined by a probe that hybridizes to a unique nucleotide sequence contained within the locus either during Southern hybridization or PCR.

[0074] Marker: As used herein, a "marker" refers to a gene or nucleotide sequence that can be used to identify plants that are likely to have a particular allele and/or exhibit a particular trait or phenotype. A marker may be described as a variation at a given genomic locus. A genetic marker may be a short DNA sequence, such as a sequence surrounding a single base-pair change (single nucleotide polymorphism, or "SNP"), or a long sequence, for example, a minisatellite/simple sequence repeat ("SSR"). A "marker allele" refers to the version of the marker that is present in a particular plant. The term marker as used herein may refer to a cloned segment of plant chromosomal DNA (*e.g.*, a segment comprising a FAD2 locus, or a modified and/or disrupted FAD2 locus), and may also or alternatively refer to a DNA molecule that is complementary to a cloned segment of plant chromosomal DNA. As is recognized by those of ordinary skill in the art, the process of obtaining additional, contiguous nucleotide sequence for inclusion in a marker may be repeated nearly indefinitely (limited only by the length of the chromosome), thereby identifying additional markers along the chromosome. Any and all of the above-described varieties of markers may be used in some embodiments of the present invention.

[0075] In some embodiments, the presence of a transgene or marker (which are characterized by a "target" sequence) in a germplasm may be detected through the use of a nucleic acid probe; *e.g.*, an oligonucleotide. A probe may be a DNA molecule or an RNA molecule. An oligonucleotide probe may be prepared synthetically or by cloning. Suitable cloning vectors are well-known to those of skill in the art. RNA probes can be synthesized by means known in the art, for example, using a DNA molecule template.

[0076] An oligonucleotide probe may be labeled or unlabeled. A wide variety of techniques exist for labeling nucleic acid molecules, including, for example and without limitation, radiolabeling by nick translation; random priming; and tailing with terminal deoxytransferase, where the nucleotides employed are labeled, for example, with radioactive ^{32}P. Other labels which may be used include, for example and without limitation, fluorophores; enzymes; enzyme substrates; enzyme cofactors; and enzyme inhibitors. Alternatively, the use of a label that provides a detectable signal, by itself or in conjunction with other reactive agents, may be replaced by ligands to which receptors bind, where the receptors are labeled (for example, by the above-indicated labels) to provide detectable signals, either by themselves, or in conjunction with other reagents. *See*, *e.g.*, Leary *et al.* (1983) Proc. Natl. Acad. Sci. USA 80:4045-9.

[0077] A probe may be an exact copy of a transgene or marker to be detected. A probe may also be a nucleic acid molecule comprising, or consisting of, a nucleotide sequence that is substantially identical to a cloned segment of chromosomal DNA comprising the transgene or marker to be detected. A probe may further comprise additional nucleic acid sequences, for example, promoters; transcription signals; and/or vector sequences.

[0078] A probe may contain all or a portion of the target nucleotide sequence and additional, contiguous nucleotide sequence from the genome. This is referred to herein as a "contiguous probe." The additional, contiguous nucleotide sequence is referred to as "upstream" or "downstream" of the original target, depending on whether the contiguous nucleotide sequence from the chromosome is on the 5' or the 3' side of the original marker, as conventionally understood. A probe may also contain a nucleotide sequence that is not contiguous to that of the original target; this probe is referred to herein as a "non-contiguous probe." The sequence of the non-contiguous probe may be located sufficiently close to the sequence of the original target on the chromosome so that the non-contiguous probe is linked to the original marker or transgene.

[0079] In some embodiments, a probe is a nucleic acid molecule that is "specifically hybridizable" or "specifically complementary" to an exact copy of the target to be detected. "Specifically hybridizable" and "specifically complementary" are terms that indicate a sufficient degree of complementarity, such that stable and specific binding occurs between the nucleic acid molecule and the target. A nucleic acid molecule need not be 100% complementary to its target sequence to be specifically hybridizable. A nucleic acid molecule is specifically hybridizable when there is a sufficient degree of complementarity to avoid non-specific binding of the nucleic acid to non-target sequences under conditions where specific binding is desired, for example, under stringent hybridization conditions.

[0080] Hybridization conditions resulting in particular degrees of stringency will vary depending upon the nature of the hybridization method of choice and the composition and length of the hybridizing nucleic acid sequences. Generally, the temperature of hybridization and the ionic strength (especially the Na^+ and/or Mg^{++} concentration) of the hybridization buffer will determine the stringency of hybridization, though wash times also influence stringency. Calculations regarding hybridization conditions required for attaining particular degrees of stringency are known to those of ordinary skill in the art, and are discussed, for example, in Sambrook *et al.* (ed.) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989, chapters 9 and 11; and Hames and

Higgins (eds.) Nucleic Acid Hybridization, IRL Press, Oxford, 1985. Further detailed instruction and guidance with regard to the hybridization of nucleic acids may be found, for example, in Tijssen, "Overview of principles of hybridization and the strategy of nucleic acid probe assays," in Laboratory Techniques in Biochemistry and Molecular Biology- Hybridization with Nucleic Acid Probes, Part I, Chapter 2, Elsevier, NY, 1993; and Ausubel *et al.*, Eds., Current Protocols in Molecular Biology, Chapter 2, Greene Publishing and Wiley-Interscience, NY, 1995.

[0081] As used herein, "stringent conditions" encompass conditions under which hybridization will only occur if there is less than 25% mismatch between the hybridization molecule and the DNA target. "Stringent conditions" include further particular levels of stringency. Thus, as used herein, "moderate stringency" conditions are those under which molecules with more than 25% sequence mismatch will not hybridize; conditions of "medium stringency" are those under which molecules with more than 15% mismatch will not hybridize; and conditions of "high stringency" are those under which sequences with more than 10% mismatch will not hybridize. Conditions of "very high stringency" are those under which sequences with more than 6% mismatch will not hybridize.

[0082] In particular embodiments, stringent conditions are hybridization at 65°C in 6x saline-sodium citrate (SSC) buffer, 5x Denhardt's solution, 0.5% SDS, and 100 μg sheared salmon testes DNA, followed by 15-30 minute sequential washes at 65°C in 2x SSC buffer and 0.5% SDS, followed by 1x SSC buffer and 0.5% SDS, and finally 0.2x SSC buffer and 0.5% SDS.

[0083] Linkage (dis)equilibrium: As used herein, the term "linkage equilibrium" refers to the situation where a marker and a second nucleic acid (*e.g.*, transgene, PTU, and second marker) independently segregate; *i.e.*, the marker and the second nucleic acid sort randomly among progeny. Nucleic acids that show linkage equilibrium are considered unlinked (whether or not they lie on the same chromosome). As used herein, the term "linkage disequilibrium" refers to the situation where a marker and a second nucleic acid segregate in a non-random manner; *i.e.*, the nucleic acids have a recombination frequency of less than 50% (and thus by definition, are separated by less than 50 cM on the same linkage group). In some examples, nucleic acids that show linkage disequilibrium are considered linked.

[0084] Linked, tightly linked, and extremely tightly linked: As used herein, linkage between a marker and a second nucleic acid (*e.g.*, transgene, PTU, and second marker) may refer to the phenomenon in which nucleic acids on a chromosome show a measurable

probability of being passed on together to individuals in the next generation. Thus, linkage of one marker to a second nucleic acid may be measured and/or expressed as a recombination frequency. The closer two nucleic acids are to each other, the closer to "1" this probability becomes. Thus, the term "linked" may refer to one or more genes or markers that are passed together with a second nucleic acid with a probability greater than 0.5 (which is expected from independent assortment where markers/genes are located on different chromosomes). When the presence of a gene (*e.g.*, a transgene) contributes to a phenotype in an individual, markers that are linked to the gene may be said to be linked to the phenotype. Thus, the term "linked" may refer to a relationship between a marker and a gene, or between a marker and a phenotype.

[0085] A relative genetic distance (determined by crossing over frequencies and measured in centimorgans (cM)) is generally proportional to the physical distance (measured in base pairs) that two linked markers or genes are separated from each other on a chromosome. One centimorgan is defined as the distance between two genetic markers that show a 1% recombination frequency (*i.e.*, a crossing-over event occurs between the two markers once in every 100 cell divisions). In general, the closer one marker is to another marker or gene (whether the distance between them is measured in terms of genetic distance or physical distance,) the more tightly they are linked. Because chromosomal distance is approximately proportional to the frequency of recombination events between traits, there is an approximate physical distance that correlates with recombination frequency. This correlation is generally known or readily determinable across the major crop plants (Helentjaris and Burr (eds.) (1989) Development and Application of Molecular Markers to Problems in Plant Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY; Gresshoff (ed.) (1994) Plant Genome Analysis. CRC Press, Boca Raton, FL; Lander *et al.* (1987) Genomics 1:174-81; Tanksley *et al.* (1988) "Molecular mapping of plant chromosomes," In Chromosome Structure and Function. Gustafson and Appels (eds.) Plenum Press, NY, pp. 157-73) and many other organisms. For example, 1 cM corresponds to about 2.5-3.0 kb in yeast, about 140 kb in *Arabidopsis*, about 400 kb in sunflower, and about 350 kb in *Eucalyptus*.

[0086] The term "linked" may refer herein to one or more nucleic acids that show a recombination frequency of less than 50% (*i.e.*, less than 50 cM). For example, "linked" nucleic acids may recombine with a frequency of about 45% or less, about 40% or less, about 35% or less, about 30% or less, about 25% or less, about 20% or less, about 15% or less, and about 10% or less. The physical distances between such nucleic acids on the same chromosome (nucleic acids on different chromosomes are expected to be in linkage equilibrium) that correspond to the

foregoing recombination frequencies depend on the host genome, and may be easily calculated as set forth, *supra*.

[0087] As used herein, the term "tightly-linked" may refer to one or more nucleic acids that show a recombination frequency of about 20% or less (*i.e.*, about 20 cM or less). For example, "tightly linked" nucleic acids may recombine with a frequency of 22% or less, about 18% or less, about 16% or less, about 14% or less, about 12% or less, about 10% or less, about 8% or less, about 6% or less, about 4% or less, and about 2% or less.

[0088] As used herein, the term "extremely tightly-linked" may refer to one or more nucleic acids that show a recombination frequency of about 10% or less (*i.e.*, about 10 cM or less). For example, "extremely tightly linked" nucleic acids may recombine with a frequency of 11% or less, about 9% or less, about 8% or less, about 7% or less, about 6% or less, about 5% or less, about 4% or less, about 3% or less, about 2% or less, and about 1% or less.

[0089] The closer a particular nucleic acid is to a gene that encodes a polypeptide that contributes to a particular phenotype (whether measured in terms of genetic or physical distance), the more tightly-linked is the particular nucleic acid to the phenotype. In view of the foregoing, it will be appreciated that nucleic acids linked to a particular gene or phenotype include those nucleic acids that are tightly linked, and those nucleic acids that are extremely tightly linked, to the gene or phenotype. In some embodiments, the closer a particular nucleic acid is to a *FAD2* locus (*e.g.*, a modified or disrupted *FAD2* locus), whether measured in terms of genetic or physical distance, the more tightly-linked is the particular nucleic acid to any trait/phenotype conferred by an exogenous nucleic acid integrated at the *FAD2* locus (or to a wild-type *FAD2* phenotype in the case of an unmodified locus). Thus, genetic markers that are linked, tightly linked, and/or extremely tightly linked to a *FAD2* locus comprising an integrated exogenous nucleic acid may be useful in an MAS program to identify organisms (*e.g.*, plants and plant varieties) comprising the integrated nucleic acid, to identify organisms comprising a phenotype conferred by the integrated nucleic acid, and to breed such an integrated nucleic acid and/or a phenotype conferred by the integrated nucleic acid into other compatible organisms.

[0090] Marker-assisted breeding: As used herein, the term "marker-assisted breeding" may refer to an approach to breeding plants directly for one or more trait(s) (*e.g.*, a polygenic trait). In current practice, plant breeders attempt to identify easily detectable traits, such as flower color, seed coat appearance, or isozyme variants that are linked to an agronomically desired trait. The plant breeders then follow the agronomic trait in the segregating, breeding populations by following the segregation of the easily detectable trait. However, there are very

few of these linkage relationships between traits of interest and easily detectable traits available for use in plant breeding. In some embodiments of the invention, marker-assisted breeding comprises identifying one or more genetic markers (*e.g.*, SNP, isozyme, and/or SSR markers) that are linked to a *FAD2* locus wherein an exogenous nucleic acid contributing to a trait of interest has been integrated, and following the trait of interest in a segregating, breeding population by following the segregation of the one or more genetic markers. In some examples, the segregation of the one or more genetic markers may be determined utilizing a probe for the one or more genetic markers by assaying a genetic sample from a progeny plant for the presence of the one or more genetic markers. Marker-assisted breeding provides a time- and cost-efficient process for improvement of plant varieties.

[0091] Trait or phenotype: The terms "trait" and "phenotype" are used interchangeably herein. For the purposes of the present disclosure, traits of particular interest include agronomically important traits, as may be expressed, for example, in a crop plant, and the production of transgene expression products from a targeted integration event. The term "molecular phenotype" may refer to a phenotype that is detectable at the level of a population of (one or more) molecules. In some examples, the molecular phenotype may only be detectable at the molecular level. The detectable molecules of the phenotype may be nucleic acids (*e.g.*, genomic DNA or RNA); proteins; and/or metabolites. For example, a molecular phenotype may be an expression profile for one or more gene products (*e.g.*, at a specific stage of plant development, or in response to an environmental condition or stress).

[0092] Quantitative Trait Locus: Traits that are continuously varying due to genetic (additive, dominant, and epistatic) and environmental influences are commonly referred to as "quantitative traits." Quantitative traits may be distinguished from "qualitative," or "discrete," traits on the basis of two factors; environmental influences on gene expression that produce a continuous distribution of phenotypes, and the complex segregation pattern produced by multigenic inheritance. The identification of one or more regions of the genome linked to the expression of a quantitative trait defines such regions as Quantitative Trait Loci ("QTL").

[0093] Plant: As used herein, the term "plant" may refer to a whole plant, a cell or tissue culture derived from a plant, and/or any part of any of the foregoing. Thus, the term "plant" encompasses, for example and without limitation, whole plants; plant components and/or organs (*e.g.*, leaves, stems, and roots); plant tissue; seed; and a plant cell. A plant cell may be, for

example and without limitation, a cell in and/or of a plant, a cell isolated from a plant, and a cell obtained through culturing of a cell isolated from a plant.

[0094] A "transgenic plant" is a plant comprising within at least one of its cells an exogenous polynucleotide. The term "transgenic" is used herein to refer to any cell, cell line, callus, tissue, plant part, or plant, the genotype of which has been altered by the presence of a exogenous nucleic acid. Thus, this term encompasses transgenic organisms and cells that have been initially altered to comprise the exogenous polynucleotide, and those organisms and cells created by crosses or asexual propagation of the initial transgenic organism or cell. The term "transgenic," as used herein, does not encompass genome (chromosomal or extra-chromosomal) alternations introduced by conventional plant breeding methods (*e.g.*, crosses of only non-transgenic organisms) or by naturally-occurring events (*e.g.*, random cross-fertilization, non-recombinant viral infection, non-recombinant bacterial transformation, non-recombinant transposition, and spontaneous mutation).

[0095] A plant "line," "variety," or "strain" is a group of individual plants having the same parentage. Plants of a line generally are inbred to some degree, and are generally homozygous and homogeneous at most genetic loci (*e.g.*, a *FAD2* locus). A "subline" may refer to an inbred subset of descendents from a common progenitor that are genetically distinct from other similarly inbred subsets descended from the same progenitor. In some embodiments, a "subline" may be produced by inbreeding seed from an individual transgenic plant selected at the F_3 to F_5 generation until the residual segregating loci are homozygous across most or all loci.

[0096] 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.

[0097] 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.

[0098] 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 in length and exhibits at least some sequence homology with other TALE repeat sequences within a naturally occurring TALE protein.

[0099] Zinc finger and TALE binding domains can be "engineered" to bind to a predetermined nucleotide sequence, for example via engineering (altering one or more amino acids) of the recognition helix region of a naturally occurring zinc finger or TALE protein. Therefore, engineered DNA binding proteins (zinc fingers or TALEs) are proteins that are non-naturally occurring. Non-limiting examples of methods for engineering DNA-binding proteins are design and selection. A designed DNA binding protein is a protein not occurring in nature whose design/composition results principally from rational criteria. Rational criteria for design include application of substitution rules and computerized algorithms for processing information in a database storing information of existing ZFP and/or TALE designs and binding data. See, for example, US 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 and U.S. Publication No. 20110301073.

[0100] 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., US 5,789,538; US 5,925,523; US 6,007,988; US 6,013,453; US 6,200,759; WO 95/19431; WO 96/06166; WO 98/53057; WO 98/54311; WO 00/27878; WO 01/60970 WO 01/88197, WO 02/099084 and U.S. Publication No. 20110301073.

[0101] "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.

[0102] A "cleavage half-domain" is a polypeptide sequence which, in conjunction with a second polypeptide (either identical or different) forms a complex having cleavage activity (preferably double-strand cleavage activity). The terms "first and second cleavage

half-domains;" "+ and – cleavage half-domains" and "right and left cleavage half-domains" are used interchangeably to refer to pairs of cleavage half-domains that dimerize.

[0103] 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 Publication Nos. 2005/0064474, 20070218528, 2008/0131962 and 2011/0201055, incorporated herein by reference in their entireties.

[0104] Means for generating a double strand DNA break: As used herein, the term "means for generating a double strand DNA break" is intended to invoke the special claiming provisions authorized by Congress in 35 U.S.C. § 112, sixth paragraph. Specifically, a "means for generating a double strand DNA break" refers to a molecular structure that is capable of cleaving both strands of a double-stranded DNA molecule. Such structures include polypeptide domains comprised within many known nuclease proteins, for example, the FokI nuclease domain, the catalytic domain is selected from the group consisting of proteins Mmel, Colicin-E7 (CEA7_ECOLX), Colicin-E9, APFL, EndA, Endo I (END1_EC0LI), Human Endo G (NUCG_HUMAN), Bovine Endo G (NUCG_BOVIN), R.HinPll, I-Basl, I-Bmol, I-Hmul, I-Tevl, I-Tevll, I-Tevlll, I-Twol, R.Mspl, R.Mval, NucA, NucM, Vvn, Vvn_CLS, Staphylococcal nuclease (NUC_STAAU), Staphylococcal nuclease (NUC_STAHY), Micrococcal nuclease (NUC_SHIFL), Endonuclease yncB, Endodeoxyribonuclease I (ENRN_BPT7), Metnase, Nb.BsrDI, BsrDI A, Nt.BspD6l (R.BspD6l large subunit), ss.BspD6l (R.BspD6l small subunit), R.PIel, Mlyl, Alwl, Mval269l, Bsrl, Bsml, Nb.BtsCI, Nt.BtsCI, Rl.Btsl, R2.Btsl, BbvCI subunit 1, BbvCI subunit 2, BpulOI alpha subunit, BpulOI beta subunit, Bmrl, Bfil, I-Crel, hExol (EX01JHUMAN), Yeast Exol (EX01_YEAST), E.coli Exol, Human TREX2, Mouse TREX1, Human TREX1, Bovine TREX1, Rat TREX1, Human DNA2, Yeast DNA2 (DNA2_YEAST).

[0105] Means for repairing a double strand DNA break: As used herein, the term "means for repairing a double strand DNA break" is also intended to invoke the special claiming provisions authorized by Congress in 35 U.S.C. § 112, sixth paragraph. Specifically, a "means for repairing a double strand DNA break" refers to a molecular structure that is capable of facilitating/catalyzing the joining of the ends of double-stranded DNA molecules, for example, by joining ends generated by cleaving a single double-stranded DNA molecule, or by joining one end generated by cleaving a single double-stranded DNA molecule with the end of an exogenous double-stranded DNA molecule. Such structures include polypeptide domains

comprised within many known ligase proteins, for example, Cre recombinase. In some examples, the same molecular structure may serve as both a means for generating a double strand DNA break and a means for repairing a double strand DNA break, where the same structure facilitates both the cleavage and repair of double-stranded DNA molecules (*e.g.*, Hin recombinase).

[0106] The induction of the site specific double stranded breaks in the genome induces the host plant cell DNA repair pathway which resolves the double stranded break through homology-directed repair (HDR) or non-homologous end-joining (NHEJ) repair. In plants, the scientific literature reports that precise gene or donor DNA integration into native genomic or at pre-engineered locations have involved incoming donor DNA construct(s) that comprise varying amounts of sequence homologous to the sequences flanking the targeted double stranded break. The integration of such donors into the specific target locus presumably has relied on the HDR pathway. Exclusively relying on the HDR approach for gene targeting in plants can have limitations due to reports that the HDR repair pathway is not the dominate DNA repair pathway when compared to NHEJ. The published plant scientific literature utilizing target specific DNA breaks (ZFN, TALeNs, or Engineered Meganucleases, etc.) the NHEJ pathway has been reported as the method to introduce specific point mutations (insertions, or deletions) into the geneome. Here we report that site specific double stranded breaks (induced by ZFN, TALeNs, etc.) in the presents of various donor DNA design with homology regions of 0 to <10 bp can be specifically inserted at targeted break via the NHEJ repair pathway in plants. A variety of different DNA donor designs with zero homology to small 1 – 10 bp of ranging from linear to circular, single stranded to double stranded can be targeted to specific locations using the NHEJ pathway. NHEJ based donor DNA plant genome targeting can be based on "sticky end capture", where the targeted double stranded break in the genome generated by *Fok*1 (or other Type II endonuclease domains) and the corresponding sticky ends are on the NHEJ donor DNA designs. The sticky ends donor DNA can be delivered directly to the cell as linear donor DNA with predefined overhangs. An alternative approach is to produce the donor DNA sticky ends *in vivo* by co-delivering the host target ZFN and a circular DNA donor molecule that contains at least one ZFN recognition site that is identical to the target recognition site. Expression of at least one ZFN cuts the host genomic DNA (native or pre-engineered) and the circular donor DNA to produce sticky ends that are resolved using the hosts NHEJ repair pathway.

[0107] It is possible to have one or more ZFN cuts sites on the donor molecule (a single ZFN cut site to linearize the entire donor molecule, 2 of the same ZFN sites to release a smaller donor DNA fragment or 2 different ZFN sites to release a fragment from the donor and a corresponding fragment from the host genomic DNA (DNA replacement).

[0108] Thus, the donor polynucleotide can be DNA or RNA, single-stranded and/or double-stranded and can be introduced into a cell in linear or circular form. *See, e.g.,* U.S. Patent Publication Nos. 20100047805 and 20110207221. In certain, embodiments of the present invention may also include linear exogenous (donor) nucleic acid(s), compositions comprising these nucleic acids and methods of making and using these linear donor molecules. In certain embodiments, the linear donor molecule stably persists in the cell into which it is introduced. In other embodiments, the linear donor molecule is modified to resist exonucleolytic cleavage, for example by placing one or more phosphorothioate phosphodiester bonds between one or more base pairs on the ends of the donor molecule. The linear exogenous nucleic acid may also include single stranded specific DNA.

III. FAD2 Performance Loci

[0109] The loci designated *FAD2* (*fatty acid desaturase 2*) are included in QTLs involved in the inheritance of the complex multigenic trait of fatty acid content in plants. *FAD2* encodes the enzyme responsible for the desaturation of oleic acid (18:1) to linoleic acid (C18:2). Tanhuanpaa *et al.* (1998) Mol. Breed. 4:543-50; Schierholt *et al.* (2001) Crop Sci. 41:1444-9.

[0110] Within the plant oil biosynthetic pathway the fatty acid desaturases (FADs) play a key role in plant lipid biosynthesis and their activity significantly influences the fatty acid composition. FADs are abundant in plants, and expression analysis suggested that FAD mRNAs are produced in over-abundance. Furthermore, FAD genes are expressed in various, tissues, and cell types, as well as subcellular compartments including the plastid and endoplasmic reticulum.

[0111] The fatty acid composition of plants, and the performance of oils produced therefrom in many applications, is determined by the relative concentrations of the major fatty acid constituents; oleic, linoleic, and linolenic (C18:3). The concentrations of these fatty acids are predominantly regulated by the function of the enzymes FAD2 and FAD3. Oleic acid is converted to linoleic acid and linolenic acid in plants according to the scheme:

C18:0 → C18:1 → C18:2 → C18:3

FAD2 FAD3

FAD2 genes have been identified in major plant and algal species including but not limited to maize, soybean, cotton, *Arabidopsis*, wheat, forage grasses, rice, sunflower and *Brassica*, and modification of FAD2 expression leads to altered fatty acid profiles in such organisms. Furthermore, plants comprising modified FAD2 genes have been commercialized, and disruption of a *FAD2* gene has been shown to be able to improve the nutritional and functional properties of oil produced by a host plant without an agronomic penalty to the host plant. For example, canola and sunflower varieties that have been commercialized under the Nexera® brand (Dow AgroSciences, LLC) are characterized by a higher oleic acid, lower linoleic aced, and lower linolenic acid (and lower saturated fatty acid) composition, when compared to wild-type canola and sunflower profiles.

[0112] The dominant canola species grown in Europe, North America, and Australia is *Brassica napus*, a polyploid *Brassica* species considered to have arisen from the hybridization of *B. oleracea* (having a diploid C genome) and *B. rapa* (having a diploid A genome). Cytogenetic investigation revealed the AA and CC genomes show a degree of relatedness as being partially homologous to one another. Both the A and C genomes contain a high percentage of homeologous or paralogous genes. Thus, it is thought that the AA and CC genomes are derived from a common ancestor genome. Prakash and Hinata (1980) Opera Botanica 55:1-57. Although the genomes of both progenitor species are technically classified as diploids, these genomes contain a high percentage of regions that are duplicative of one another. Song *et al.* (1991) Theor. Appl. Genet. 82:296-304. A detailed organelle and nuclear RFLP analysis revealed that the AA genome of *B. rapa* contributed ten chromosomes to *B. napus*, while *B. oleracea* contributed nine chromosomes from its CC genome as the maternal donor. Song *et al.* (1992) Genome 35:992-1001. Through the number of genome duplications in both ancestral genomes, as well as the high percentage of similarity between the A, B and C genomes, there have arisen several copies of FAD2 and FAD3 genes. As a practical matter, this fact makes breeding canola with modified and/or disrupted copies of these genes in order to produce a particular fatty acid profile particularly challenging.

[0113] The known functional gene copies of *FAD2* in canola are located on linkage group N4 of the A genome. Scheffler *et al.* (1997) TAG 94(5):583-91; Schierholt *et al.* (2000) TAG 101(5-6):897-901. More recently, a high oleic trait in canola has been associated with a modified and disrupted *FAD2* gene that is located on the A genome. U.S.

Patent Application Publication No. US 2006/0248611 A1; Hu *et al.* (2006) "Identification and Mapping of FAD2 and FAD3 Mutations and Development of Allele-specific Markers for High Oleic and Low Linolenic Acid Contents in Canola (*Brassica napus* L.)," Plant & Animal Genomes XIV Conference, January 14-18, 2006, San Diego, CA. An inactivating FAD2 allele contributes to the control of oleic acid content by reducing the desaturation of oleic acid to linoleic acid. This high oleic acid and *fad2* trait was identified in a *B. napus* variety (DMS100) that has a characteristic oleic acid content of about 77%. See, U.S. Application No: 10/545,100. Additionally, FAD2 genes were recently located on the A5 chromosome and were alleged to be responsible for high C18:1 content. See, Yang *et al.*, "*Brassica napus* genome" Theor Appl Genet (2012 125:715-729. Further, genetic markers have been developed to assist the introgression of the *FAD2* and high oleic acid trait into canola.

[0114] FAD2 loci may be modified and/or disrupted in a plant without detrimentally affecting the value of the plant, and for many purposes, with an actual increase in its value, including alteration of FAD2 expression, alteration of oil content/ratios and/or integration and expression of desired transgenes. Furthermore, according to the ubiquitous nature of FAD loci in plants, *FAD2* loci may be modified and/or disrupted without detriment for at least some purposes in many species, including, for example and without limitation: canola; soybean; maize; wheat; forage grasses; *brassica* sp.; rice, tomatoes, barley; oats; sorghum; cotton; and sunflower, as well as fungi and algae. Embodiments of the invention include FAD2 loci, and the use thereof as performance loci for integration of exogenous nucleic acids. In examples, a FAD2 locus exhibits at least one of several features that have been found to be desirable within the context of its use as a performance locus, including, for example and without limitation: that there is an approximately consistent level of expression during the life cycle of the host organism; and surprisingly, that integration of donor DNA at a FAD2 locus does not induce a quality or fitness penalty on the host.

[0115] In some embodiments of the present invention, at least one FAD2 locus (e.g., a FAD2A, FAD2A', FAD2C and/or FAD2C' locus) is used as a target site for the site-specific integration of an exogenous nucleic acid (e.g., a nucleic acid comprising a nucleotide sequence encoding a polypeptide of interest). In particular embodiments, integration of the exogenous nucleic acid results in a modified locus. For example, integration of the exogenous nucleic acid may modify the locus so as to produce a disrupted (i.e., inactivated) FAD2 gene.

[0116] In some embodiments, a FAD2 locus may comprise a nucleotide sequence that is specifically hybridizable to the complement of a nucleotide sequence selected from the group consisting of SEQ ID NOs: 22-26, SEQ ID NOs: 28-33 and SEQ ID NOs: 35-38. For example, a FAD2 locus may comprise a nucleotide sequence selected from the group consisting of SEQ ID NOs: 22-26, SEQ ID NOs: 28-33 and SEQ ID NOs: 35-38. In some embodiments, a FAD2 locus may comprise a nucleotide sequence that is substantially identical to a nucleotide sequence selected from the group consisting of SEQ ID NOs: 22-26, SEQ ID NOs: 28-33 and SEQ ID NOs: 35-38. For example, in some embodiments, a FAD2 locus is a FAD2 homologue (e.g., an ortholog or a paralog) that comprises a nucleotide sequence that is at least about 85% identical to a nucleotide sequence selected from the group consisting of SEQ ID NOs: 22-26, SEQ ID NOs: 28-33 and SEQ ID NOs: 35-38. A FAD2 homologue may comprise a nucleotide sequence that is, for example and without limitation: at least 80%; at least 85%; at least about 90%; at least about 91%; at least about 92%; at least about 93%; at least about 94%; at least about 95%; at least about 96%; at least about 97%; at least about 98%; at least about 99%; at least about 99.5%; 99.6%, 99.7%, 99.8% and/or at least about 99.9% identical to a nucleotide sequence selected from the group consisting of SEQ ID NOs: 22-26, SEQ ID NOs: 28-33 and SEQ ID NOs: 35-38. Such a FAD2 homologue may be readily identified and isolated from any complete or partial genome readily available to those of skill in the art for a variety of organisms.

IV. Targeted integration of a nucleic acid at a FAD2 locus

[0117] Site-specific integration of an exogenous nucleic acid at a FAD2 locus may be accomplished by any technique known to those of skill in the art. In some embodiments, integration of an exogenous nucleic acid at a FAD2 locus comprises contacting a cell (e.g., an isolated cell or a cell in a tissue or organism) with a nucleic acid molecule comprising the exogenous nucleic acid. In examples, such a nucleic acid molecule may comprise nucleotide sequences flanking the exogenous nucleic acid that facilitate homologous recombination between the nucleic acid molecule and at least one FAD2 locus. In particular examples, the nucleotide sequences flanking the exogenous nucleic acid that facilitate homologous recombination may be complementary to endogenous nucleotides of the FAD2 locus. In particular examples, the nucleotide sequences flanking the exogenous nucleic acid that facilitate homologous recombination may be complementary to previously integrated exogenous

nucleotides. In some embodiments, a plurality of exogenous nucleic acids may be integrated at one FAD2 locus, such as in gene stacking.

[0118] Integration of a nucleic acid at a FAD2 locus may be facilitated (e.g., catalyzed) in some embodiments by endogenous cellular machinery of a host cell, such as, for example and without limitation, endogenous DNA and endogenous recombinase enzymes. In some embodiments, integration of a nucleic acid at a FAD2 locus may be facilitated by one or more factors (e.g., polypeptides) that are provided to a host cell. For example, nuclease(s), recombinase(s), and/or ligase polypeptides may be provided (either independently or as part of a chimeric polypeptide) by contacting the polypeptides with the host cell, or by expressing the polypeptides within the host cell. Accordingly, in some examples, a nucleic acid comprising a nucleotide sequence encoding at least one nuclease, recombinase, and/or ligase polypeptide may be introduced into the host cell, either concurrently or sequentially with a nucleic acid to be integrated site-specifically at a FAD2 locus, wherein the at least one nuclease, recombinase, and/or ligase polypeptide is expressed from the nucleotide sequence in the host cell.

A. DNA-binding polypeptides

[0119] In some embodiments, site-specific integration may be accomplished by utilizing factors that are capable of recognizing and binding to particular nucleotide sequences, for example, in the genome of a host organism. For instance, many proteins comprise polypeptide domains that are capable of recognizing and binding to DNA in a site-specific manner. A DNA sequence that is recognized by a DNA-binding polypeptide may be referred to as a "target" sequence. Polypeptide domains that are capable of recognizing and binding to DNA in a site-specific manner generally fold correctly and function independently to bind DNA in a site-specific manner, even when expressed in a polypeptide other than the protein from which the domain was originally isolated. Similarly, target sequences for recognition and binding by DNA-binding polypeptides are generally able to be recognized and bound by such polypeptides, even when present in large DNA structures (*e.g.*, a chromosome), particularly when the site where the target sequence is located is one known to be accessible to soluble cellular proteins (*e.g.*, a gene).

[0120] While DNA-binding polypeptides identified from proteins that exist in nature typically bind to a discrete nucleotide sequence or motif (*e.g.*, a consensus recognition sequence), methods exist and are known in the art for modifying many such DNA-binding polypeptides to recognize a different nucleotide sequence or motif. DNA-binding polypeptides

include, for example and without limitation: zinc finger DNA-binding domains; leucine zippers; UPA DNA-binding domains; GAL4; TAL; LexA; a Tet repressor; LacR; and a steroid hormone receptor.

[0121] In some examples, a DNA-binding polypeptide is a zinc finger. Individual zinc finger motifs can be designed to target and bind specifically to any of a large range of DNA sites. Canonical Cys_2His_2 (as well as non-canonical Cys_3His) zinc finger polypeptides bind DNA by inserting an α-helix into the major groove of the target DNA double helix. Recognition of DNA by a zinc finger is modular; each finger contacts primarily three consecutive base pairs in the target, and a few key residues in the polypeptide mediate recognition. By including multiple zinc finger DNA-binding domains in a targeting endonuclease, the DNA-binding specificity of the targeting endonuclease may be further increased (and hence the specificity of any gene regulatory effects conferred thereby may also be increased). *See, e.g.*, Urnov *et al.* (2005) Nature 435:646-51. Thus, one or more zinc finger DNA-binding polypeptides may be engineered and utilized such that a targeting endonuclease introduced into a host cell interacts with a DNA sequence that is unique within the genome of the host cell.

[0122] Preferably, the zinc finger protein is non-naturally occurring in that it is engineered to bind to a target site of choice. See, for example, See, for example, Beerli et al. (2002) Nature Biotechnol. 20:135-141; Pabo et al. (2001) Ann. Rev. Biochem. 70:313-340; Isalan et al. (2001) Nature Biotechnol. 19:656-660; Segal et al. (2001) Curr. Opin. Biotechnol. 12:632-637; Choo et al. (2000) Curr. Opin. Struct. Biol. 10:411-416; U.S. Patent Nos. 6,453,242; 6,534,261; 6,599,692; 6,503,717; 6,689,558; 7,030,215; 6,794,136; 7,067,317; 7,262,054; 7,070,934; 7,361,635; 7,253,273; and U.S. Patent Publication Nos. 2005/0064474; 2007/0218528; 2005/0267061, all incorporated herein by reference in their entireties.

[0123] 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. See, for example, co-owned U.S. Patents 6,453,242 and 6,534,261, incorporated by reference herein in their entireties.

[0124] Exemplary selection methods, including phage display and two-hybrid systems, are disclosed in US Patents 5,789,538; 5,925,523; 6,007,988; 6,013,453; 6,410,248; 6,140,466; 6,200,759; and 6,242,568; as well as WO 98/37186; WO 98/53057; WO 00/27878; WO 01/88197 and GB 2,338,237. In addition, enhancement of binding specificity for zinc finger binding domains has been described, for example, in co-owned WO 02/077227.

[0125] In addition, as disclosed in these and other references, zinc finger domains and/or multi-fingered zinc finger proteins may be linked together using any suitable linker sequences, including for example, linkers of 5 or more amino acids in length. See, also, U.S. Patent Nos. 6,479,626; 6,903,185; and 7,153,949 for exemplary linker sequences 6 or more amino acids in length. The proteins described herein may include any combination of suitable linkers between the individual zinc fingers of the protein.

[0126] Selection of target sites; ZFPs and methods for design and construction of fusion proteins (and polynucleotides encoding same) are known to those of skill in the art and described in detail in U.S. Patent Nos. 6,140,0815; 789,538; 6,453,242; 6,534,261; 5,925,523; 6,007,988; 6,013,453; 6,200,759; WO 95/19431; WO 96/06166; WO 98/53057; WO 98/54311; WO 00/27878; WO 01/60970 WO 01/88197; WO 02/099084; WO 98/53058; WO 98/53059; WO 98/53060; WO 02/016536 and WO 03/016496.

[0127] In addition, as disclosed in these and other references, zinc finger domains and/or multi-fingered zinc finger proteins may be linked together using any suitable linker sequences, including for example, linkers of 5 or more amino acids in length. See, also, U.S. Patent Nos. 6,479,626; 6,903,185; and 7,153,949 for exemplary linker sequences 6 or more amino acids in length. The proteins described herein may include any combination of suitable linkers between the individual zinc fingers of the protein.

[0128] In some examples, a DNA-binding polypeptide is a DNA-binding domain from GAL4. GAL4 is a modular transactivator in *Saccharomyces cerevisiae*, but it also operates as a transactivator in many other organisms. *See, e.g.*, Sadowski *et al.* (1988) Nature 335:563-4. In this regulatory system, the expression of genes encoding enzymes of the galactose metabolic pathway in *S. cerevisiae* is stringently regulated by the available carbon source. Johnston (1987) Microbiol. Rev. 51:458-76. Transcriptional control of these metabolic enzymes is mediated by the interaction between the positive regulatory protein, GAL4, and a 17 bp symmetrical DNA sequence to which GAL4 specifically binds (the UAS).

[0129] Native GAL4 includes 881 amino acid residues, with a molecular weight of 99 kDa. GAL4 comprises functionally autonomous domains, the combined activities of which account for activity of GAL4 *in vivo*. Ma and Ptashne (1987) Cell 48:847-53); Brent and Ptashne (1985) Cell 43(3 Pt 2):729-36. The N-terminal 65 amino acids of GAL4 comprise the GAL4 DNA-binding domain. Keegan *et al.* (1986) Science 231:699-704; Johnston (1987) Nature 328:353-5. Sequence-specific binding requires the presence of a divalent cation coordinated by 6 Cys residues present in the DNA binding domain. The coordinated cation-containing domain interacts with and recognizes a conserved CCG triplet at each end of the 17 bp UAS *via* direct contacts with the major groove of the DNA helix. Marmorstein *et al.* (1992) Nature 356:408-14. The DNA-binding function of the protein positions C-terminal transcriptional activating domains in the vicinity of the promoter, such that the activating domains can direct transcription.

[0130] Additional DNA-binding polypeptides that may be utilized in certain embodiments include, for example and without limitation, a binding sequence from a AVRBS3-inducible gene; a consensus binding sequence from a AVRBS3-inducible gene or synthetic binding sequence engineered therefrom (*e.g.*, UPA DNA-binding domain); TAL; LexA (*see, e.g.*, Brent & Ptashne (1985), *supra*); LacR (*see, e.g.*, Labow *et al.* (1990) Mol. Cell. Biol. 10:3343-56; Baim *et al.* (1991) Proc. Natl. Acad. Sci. USA 88(12):5072-6); a steroid hormone receptor (Elliston *et al.* (1990) J. Biol. Chem. 265:11517-121); the Tet repressor (U.S. Patent 6,271,341) and a mutated Tet repressor that binds to a *tet* operator sequence in the presence, but not the absence, of tetracycline (Tc); the DNA-binding domain of NF-κB; and components of the regulatory system described in Wang *et al.* (1994) Proc. Natl. Acad. Sci. USA 91(17):8180-4, which utilizes a fusion of GAL4, a hormone receptor, and VP16.

[0131] In certain embodiments, the DNA-binding domain of one or more of the nucleases used in the methods and compositions described herein comprises a naturally occurring or engineered (non-naturally occurring) TAL effector DNA binding domain. *See, e.g.,* U.S. Patent Publication No. 20110301073, incorporated by reference in its entirety herein. The plant pathogenic bacteria of the genus Xanthomonas are known to cause many diseases in important crop plants. Pathogenicity of Xanthomonas depends on a conserved type III secretion (T3S) system which injects more than 25 different effector proteins into the plant cell. Among these injected proteins are transcription activator-like (TAL) effectors which mimic plant transcriptional activators and manipulate the plant transcriptome (see Kay et al (2007) Science 318:648-651). These proteins contain a DNA binding domain and a

transcriptional activation domain. One of the most well characterized TAL-effectors is AvrBs3 from Xanthomonas campestgris pv. Vesicatoria (see Bonas et al (1989) Mol Gen Genet 218: 127-136 and WO2010079430). TAL-effectors contain a centralized domain of tandem repeats, each repeat containing approximately 34 amino acids, which are key to the DNA binding specificity of these proteins. In addition, 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) Appl and Envir Micro 73(13): 4379-4384). These genes are 98.9% identical in nucleotide sequence to each other but differ by a deletion of 1,575 bp in the repeat domain of hpx17. However, both gene products have less than 40% sequence identity with AvrBs3 family proteins of Xanthomonas. See, e.g., U.S. Patent Nos., 8,420,782 and 8,440,431 and U.S. Patent Publication No. 20110301073.

[0132] In other 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.

[0133] 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 Wastson-Crick base-pairing between the spacer on the crRNA and the protospacer on the target DNA next to the protospacer adjacent motif (PAM), an additional requirement for target recognition. Finally, Cas9 mediates cleavage of target DNA to create a double-stranded break within the

protospacer. Activity of the CRISPR/Cas system comprises of three steps: (i) insertion of alien DNA sequences into the CRISPR array to prevent future attacks, in a process called 'adaptation', (ii) expression of the relevant proteins, as well as expression and processing of the array, followed by (iii) RNA-mediated interference with the alien nucleic acid. Thus, in the bacterial cell, several of the so-called 'Cas' proteins are involved with the natural function of the CRISPR/Cas system and serve roles in functions such as insertion of the alien DNA etc.

[0134] 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.

[0135] In particular embodiments, a DNA-binding polypeptide specifically recognizes and binds to a target nucleotide sequence comprised within a genomic nucleic acid of a host organism. Any number of discrete instances of the target nucleotide sequence may be found in the host genome in some examples. The target nucleotide sequence may be rare within the genome of the organism (*e.g.*, fewer than about 10, about 9, about 8, about 7, about 6, about 5, about 4, about 3, about 2, or about 1 copy(ies) of the target sequence may exist in the genome). For example, the target nucleotide sequence may be located at a unique site within the genome

of the organism. Target nucleotide sequences may be, for example and without limitation, randomly dispersed throughout the genome with respect to one another; located in different linkage groups in the genome; located in the same linkage group; located on different chromosomes; located on the same chromosome; located in the genome at sites that are expressed under similar conditions in the organism (*e.g.*, under the control of the same, or substantially functionally identical, regulatory factors); and located closely to one another in the genome (*e.g.*, target sequences may be comprised within nucleic acids integrated as concatemers at genomic loci).

B. Targeting endonucleases

[0136] In particular embodiments, a DNA-binding polypeptide that specifically recognizes and binds to a target nucleotide sequence may be comprised within a chimeric polypeptide, so as to confer specific binding to the target sequence upon the chimeric polypeptide. In examples, such a chimeric polypeptide may comprise, for example and without limitation, nuclease, recombinase, and/or ligase polypeptides, as these polypeptides are described above. Chimeric polypeptides comprising a DNA-binding polypeptide and a nuclease, recombinase, and/or ligase polypeptide may also comprise other functional polypeptide motifs and/or domains, such as for example and without limitation: a spacer sequence positioned between the functional polypeptides in the chimeric protein; a leader peptide; a peptide that targets the fusion protein to an organelle (*e.g.*, the nucleus); polypeptides that are cleaved by a cellular enzyme; peptide tags (*e.g.*, Myc, His, etc.); and other amino acid sequences that do not interfere with the function of the chimeric polypeptide.

[0137] Functional polypeptides (*e.g.*, DNA-binding polypeptides and nuclease polypeptides) in a chimeric polypeptide may be operatively linked. In some embodiments, functional polypeptides of a chimeric polypeptide may be operatively linked by their expression from a single polynucleotide encoding at least the functional polypeptides ligated to each other in-frame, so as to create a chimeric gene encoding a chimeric protein. In alternative embodiments, the functional polypeptides of a chimeric polypeptide may be operatively linked by other means, such as by cross-linkage of independently expressed polypeptides.

[0138] In some embodiments, a DNA-binding polypeptide that specifically recognizes and binds to a target nucleotide sequence may be comprised within a natural isolated protein (or mutant thereof), wherein the natural isolated protein or mutant thereof also comprises a nuclease polypeptide (and may also comprise a recombinase and/or ligase polypeptide). Examples of

such isolated proteins include TALENs, recombinases (*e.g.*, Cre, Hin, Tre, and FLP recombinase), RNA-guided CRISPR-Cas9, and meganucleases.

[0139] As used herein, the term "targeting endonuclease" refers to natural or engineered isolated proteins and mutants thereof that comprise a DNA-binding polypeptide and a nuclease polypeptide, as well as to chimeric polypeptides comprising a DNA-binding polypeptide and a nuclease. Any targeting endonuclease comprising a DNA-binding polypeptide that specifically recognizes and binds to a target nucleotide sequence comprised within a *FAD2* locus (*e.g.*, either because the target sequence is comprised within the native sequence at the locus, or because the target sequence has been introduced into the locus, for example, by recombination) may be utilized in certain embodiments.

[0140] Some examples of chimeric polypeptides that may be useful in particular embodiments of the invention include, without limitation, combinations of the following polypeptides: zinc finger DNA-binding polypeptides; a FokI nuclease polypeptide; TALE domains; leucine zippers; transcription factor DNA-binding motifs; and DNA recognition and/or cleavage domains isolated from, for example and without limitation, a TALEN, a recombinase (*e.g.*, Cre, Hin, RecA, Tre, and FLP recombinases), RNA-guided CRISPR-Cas9, a meganuclease; and others known to those in the art. Particular examples include a chimeric protein comprising a site-specific DNA binding polypeptide and a nuclease polypeptide. Chimeric polypeptides may be engineered by methods known to those of skill in the art to alter the recognition sequence of a DNA-binding polypeptide comprised within the chimeric polypeptide, so as to target the chimeric polypeptide to a particular nucleotide sequence of interest.

[0141] In certain embodiments, the chimeric polypeptide comprises a DNA-binding domain (*e.g.*, zinc finger, TAL-effector domain, etc.) and a nuclease (cleavage) domain. The cleavage domain may be heterologous to the DNA-binding domain, for example a zinc finger DNA-binding domain and a cleavage domain from a nuclease or a TALEN DNA-binding domain and a cleavage domain, or meganuclease DNA-binding domain and cleavage domain from a different nuclease. Heterologous cleavage domains can be obtained from any endonuclease or exonuclease. Exemplary endonucleases from which a cleavage domain can be 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). One or more of these enzymes (or functional fragments thereof) can be used as a source of cleavage domains and cleavage half-domains.

[0142] Similarly, a cleavage half-domain can be derived from any nuclease or portion thereof, as set forth above, that requires dimerization for cleavage activity. In general, two fusion proteins are required for cleavage if the fusion proteins comprise cleavage half-domains. Alternatively, a single protein comprising two cleavage half-domains can be used. The two cleavage half-domains can be derived from the same endonuclease (or functional fragments thereof), or each cleavage half-domain can be derived from a different endonuclease (or functional fragments thereof). In addition, the target sites for the two fusion proteins are preferably disposed, with respect to each other, such that binding of the two fusion proteins to their respective target sites places the cleavage half-domains in a spatial orientation to each other that allows the cleavage half-domains to form a functional cleavage domain, *e.g.*, by dimerizing. Thus, in certain embodiments, the near edges of the target sites are separated by 5-8 nucleotides or by 15-18 nucleotides. However any integral number of nucleotides or nucleotide pairs can intervene between two target sites (*e.g.*, from 2 to 50 nucleotide pairs or more). In general, the site of cleavage lies between the target sites.

[0143] 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, for example, such that one or more exogenous sequences (donors/trangsenes) are integrated at or near the binding (target) sites. 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.

[0144] An exemplary Type IIS restriction enzyme, whose cleavage domain is separable from the binding domain, is *Fok* I. This particular enzyme is active as a dimer.

Bitinaite *et al.* (1998) *Proc. Natl. Acad. Sci. USA* 95: 10,570-10,575. Accordingly, for the purposes of the present disclosure, the portion of the *Fok* I enzyme used in the disclosed fusion proteins is considered a cleavage half-domain. Thus, for targeted double-stranded cleavage and/or targeted replacement of cellular sequences using zinc finger-*Fok* I fusions, two fusion proteins, each comprising a *Fok*I cleavage half-domain, can be used to reconstitute a catalytically active cleavage domain. Alternatively, a single polypeptide molecule containing a DNA binding domain and two *Fok* I cleavage half-domains can also be used.

[0145] 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.

[0146] Exemplary Type IIS restriction enzymes are described in U.S. Patent Publication No. 20070134796, incorporated herein in its entirety. Additional restriction enzymes also contain separable binding and cleavage domains, and these are contemplated by the present disclosure. *See*, for example, Roberts *et al.* (2003) *Nucleic Acids Res.* **31**:418-420.

[0147] In certain embodiments, the cleavage domain comprises one or more engineered cleavage half-domain (also referred to as dimerization domain mutants) that minimize or prevent homodimerization, as described, for example, in U.S. Patent Publication Nos. 20050064474; 20060188987 and 20080131962, the disclosures of all of which are incorporated by reference in their entireties herein. Amino acid residues at positions 446, 447, 479, 483, 484, 486, 487, 490, 491, 496, 498, 499, 500, 531, 534, 537, and 538 of *Fok* I are all targets for influencing dimerization of the *Fok* I cleavage half-domains.

[0148] 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.

[0149] Thus, in one embodiment, a mutation at 490 replaces Glu (E) with Lys (K); the mutation at 538 replaces Iso (I) with Lys (K); the mutation at 486 replaced Gln (Q) with Glu (E); and the mutation at position 499 replaces Iso (I) with Lys (K). Specifically, the engineered cleavage half-domains described herein were prepared by mutating positions 490 (E→K) and 538 (I→K) in one cleavage half-domain to produce an engineered cleavage half-domain designated "E490K:I538K" and by mutating positions 486 (Q→E) and 499 (I→L) in

another cleavage half-domain to produce an engineered cleavage half-domain designated "Q486E:I499L". The engineered cleavage half-domains described herein are obligate heterodimer mutants in which aberrant cleavage is minimized or abolished. *See, e.g.,* U.S. Patent Publication No. 2008/0131962, the disclosure of which is incorporated by reference in its entirety for all purposes.

[0150] 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 US Patent Publication No. 20110201055). 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 Publication Nos. 20050064474; 20080131962; and 20110201055.

[0151] Alternatively, nucleases may be assembled *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.

C. Zinc finger nucleases

[0152] In specific embodiments, a chimeric polypeptide is a custom-designed zinc finger nuclease (ZFN) that may be designed to deliver a targeted site-specific double-strand DNA break into which an exogenous nucleic acid, or donor DNA, may be integrated (*See* co-owned US Patent publication 20100257638, incorporated by reference herein). ZFNs are chimeric polypeptides containing a non-specific cleavage domain from a restriction endonuclease (for example, *FokI*) and a zinc finger DNA-binding domain polypeptide. *See, e.g.*, Huang *et al.* (1996) J. Protein Chem. 15:481-9; Kim *et al.* (1997a) Proc. Natl. Acad. Sci. USA 94:3616-20; Kim *et al.* (1996) Proc. Natl. Acad. Sci. USA 93:1156-60; Kim *et al.* (1994) Proc Natl. Acad. Sci. USA 91:883-7; Kim *et al.* (1997b) Proc. Natl. Acad. Sci. USA 94:12875-9; Kim *et al.* (1997c) Gene 203:43-9; Kim *et al.* (1998) Biol. Chem. 379:489-95; Nahon and Raveh (1998) Nucleic Acids Res. 26:1233-9; Smith *et al.* (1999) Nucleic Acids Res. 27:674-81. In some embodiments, the ZFNs comprise non-canonical zinc finger DNA binding domains (see co-owned US Patent publication 20080182332, incorporated by reference herein). The *FokI* restriction endonuclease must dimerize *via* the nuclease domain in order to cleave DNA and introduce a double-strand break. Consequently, ZFNs containing a nuclease domain from such an endonuclease also require dimerization of the nuclease domain in order to cleave target DNA. Mani *et al.* (2005) Biochem. Biophys. Res. Commun. 334:1191-7; Smith *et al.* (2000) Nucleic Acids Res. 28:3361-9. Dimerization of the ZFN can be facilitated by two adjacent, oppositely oriented DNA-binding sites. *Id.*

[0153] The flexibility and specificity of the ZFN system provides a level of control previously unachievable by known recombinase-mediated gene editing strategies. As one example, ZFNs can be easily engineered, for example, to recognize specific nucleic acid sequences. Wu *et al.* (2007) Cell. Mol. Life Sci. 64:2933-44 (*See*, US Patent Publications 20090205083, 20110189775, 20110167521 and 20100199389, incorporated by reference in their entireties herein). Randomization of the codons for zinc finger recognition residues allows the selection of new fingers that have high affinity for arbitrarily chosen DNA sequences. Furthermore, zinc fingers are natural DNA-binding molecules, and engineered zinc fingers have been shown to act on their designed targets in living cells. Thus, nucleases based on zinc fingers are targetable to specific but arbitrary recognition sites.

[0154] In particular examples, a method for the site-specific integration of an exogenous nucleic acid into at least one *FAD2* performance locus of a host comprises introducing into a cell of the host a ZFN, wherein the ZFN recognizes and binds to a target nucleotide sequence,

wherein the target nucleotide sequence is comprised within at least one *FAD2* locus of the host. In certain examples, the target nucleotide sequence is not comprised within the genome of the host at any other position than the at least one *FAD2* locus. For example, a DNA-binding polypeptide of the ZFN may be engineered to recognize and bind to a target nucleotide sequence identified within the at least one *FAD2* locus (*e.g.*, by sequencing the *FAD2* locus). A method for the site-specific integration of an exogenous nucleic acid into at least one *FAD2* performance locus of a host that comprises introducing into a cell of the host a ZFN may also comprise introducing into the cell an exogenous nucleic acid, wherein recombination of the exogenous nucleic acid into a nucleic acid of the host comprising the at least one *FAD2* locus is facilitated by site-specific recognition and binding of the ZFN to the target sequence (and subsequent cleavage of the nucleic acid comprising the *FAD2* locus).

V. Exogenous nucleic acids for integration at a FAD2 locus

[0155] Embodiments of the invention may include one or more nucleic acids selected from the group consisting of: an exogenous nucleic acid for site-specific integration in at least one *FAD2* locus, for example and without limitation, a PTU, ELP, ETIP or an ORF; a nucleic acid comprising a nucleotide sequence encoding a targeting endonuclease; and a vector comprising at least one of either or both of the foregoing. Thus, particular nucleic acids for use in some embodiments include nucleotide sequences encoding a polypeptide, structural nucleotide sequences, and/or DNA-binding polypeptide recognition and binding sites.

A. Exogenous nucleic acid molecules for site-specific integration

[0156] As noted above, integration of an exogenous sequence (also called a "donor sequence" or "donor" or "transgene") is provided, for example for expression of a polypeptide, correction of a mutant gene or for increased expression of a wild-type gene. It will be readily apparent that the donor sequence is typically not identical to the genomic sequence where it is placed. 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 integration 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.

[0157] 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, 20110207221 and U.S. Application No. 13/889,162. 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.

[0158] 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)).

[0159] The donor is generally integrated 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 integrated (*e.g., FAD2*). 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.

[0160] 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.

[0161] Exogenous nucleic acids that may be integrated in a site-specific manner into at least one *FAD2* locus, so as to modify the *FAD2* locus, in embodiments include, for example and without limitation, nucleic acids comprising a nucleotide sequence encoding a polypeptide

of interest; nucleic acids comprising an agronomic gene; nucleic acids comprising a nucleotide sequence encoding an RNAi molecule; or nucleic acids that disrupt the *FAD2* gene.

[0162] In some embodiments, an exogenous nucleic acid is integrated at a *FAD2* locus, so as to modify the *FAD2* locus, wherein the nucleic acid comprises an agronomic gene or nucleotide sequence encoding a polypeptide of interest, such that the agronomic gene or nucleotide sequence is expressed in the host from the *FAD2* locus. In some examples, the polypeptide of interest (*e.g.*, a foreign protein) is expressed from a nucleotide sequence encoding the polypeptide of interest in commercial quantities. In such examples, the polypeptide of interest may be extracted from the host cell, tissue, or biomass. In some embodiments, the host is a plant, and plant material provided for commercial production of a polypeptide of interest may be a plant, plant part, plant tissue, or plant cell. In some examples, the plant part may be plant seed. Protein extraction from a plant biomass may be accomplished by known methods which are discussed, for example, in Heney and Orr (1981) Anal. Biochem. 114:92-6.

[0163] Likewise, agronomic genes may be expressed in transformed plant cells, plants, and/or their progeny. For example, a plant may be genetically engineered *via* methods of particular embodiments to express various phenotypes of agronomic interest from at least one *FAD2* locus.

[0164] In some embodiments, nucleic acids comprising an agronomic gene or nucleotide sequence encoding a polypeptide of interest may include, for example and without limitation: a gene that confers resistance to a pests or disease (*See, e.g.*, Jones *et al.* (1994) Science 266:789 (cloning of the tomato Cf-9 gene for resistance to *Cladosporium fulvum*); Martin *et al.* (1993) Science 262:1432; Mindrinos *et al.* (1994) Cell 78:1089 (*RSP2* gene for resistance to *Pseudomonas syringae*); PCT International Patent Publication No. WO 96/30517 (resistance to soybean cyst nematode); PCT International Patent Publication No. WO 93/19181); a gene that encodes a *Bacillus thuringiensis* protein, a derivative thereof, or a synthetic polypeptide modeled thereon (*See, e.g.*, Geiser *et al.* (1986) Gene 48:109 (cloning and nucleotide sequence of a Bt δ-endotoxin gene; moreover, DNA molecules encoding δ-endotoxin genes can be purchased from American Type Culture Collection (Manassas, VA), for example, under ATCC Accession Nos. 40098; 67136; 31995; and 31998)); a gene that encodes a lectin (*See, e.g.*, Van Damme *et al.* (1994) Plant Molec. Biol. 24:25 (nucleotide sequences of several *Clivia miniata* mannose-binding lectin genes)); a gene that encodes a vitamin-binding protein, *e.g.*, avidin (*See* PCT International Patent Publication No. US93/06487 (use of avidin and avidin homologues as larvicides against insect pests)); a gene that encodes an enzyme inhibitor, *e.g.*, a

protease, proteinase inhibitor, or amylase inhibitor (*See, e.g.*, Abe *et al.* (1987) J. Biol. Chem. 262:16793 (nucleotide sequence of rice cysteine proteinase inhibitor); Huub *et al.* (1993) Plant Molec. Biol. 21:985 (nucleotide sequence of cDNA encoding tobacco proteinase inhibitor I); Sumitani *et al.* (1993) Biosci. Biotech. Biochem. 57:1243 (nucleotide sequence of *Streptomyces nitrosporeus* alpha-amylase inhibitor) and U.S. Patent 5,494,813); a gene encoding an insect-specific hormone or pheromone, *e.g.*, an ecdysteroid or juvenile hormone, a variant thereof, a mimetic based thereon, or an antagonist or agonist thereof (*See, e.g.*, Hammock *et al.* (1990) Nature 344:458 (baculovirus expression of cloned juvenile hormone esterase, an inactivator of juvenile hormone)); a gene encoding an insect-specific peptide or neuropeptide that, upon expression, disrupts the physiology of the affected pest (*See, e.g.*, Regan (1994) J. Biol. Chem. 269:9 (expression cloning yields DNA coding for insect diuretic hormone receptor); Pratt *et al.* (1989) Biochem. Biophys. Res. Comm. 163:1243 (an allostatin in *Diploptera puntata*); and U.S. Patent 5,266,317 (genes encoding insect-specific, paralytic neurotoxins)); a gene encoding an insect-specific venom produced in nature by a snake, a wasp, or other organism (*See, e.g.*, Pang *et al.* (1992) Gene 116:165 (heterologous expression in plants of a gene coding for a scorpion insectotoxic peptide)); a gene encoding an enzyme responsible for a hyperaccumulation of a monoterpene, a sesquiterpene, a steroid, hydroxamic acid, a phenylpropanoid derivative or other molecule with insecticidal activity; a gene encoding an enzyme involved in the modification, including the post-translational modification, of a biologically active molecule, *e.g.*, a glycolytic enzyme, a proteolytic enzyme, a lipolytic enzyme, a nuclease, a cyclase, a transaminase, an esterase, a hydrolase, a phosphatase, a kinase, a phosphorylase, a polymerase, an elastase, a chitinase, or a glucanase, whether natural or synthetic (*See, e.g.*, PCT International Patent Publication No. WO 93/02197 (nucleotide sequence of a callase gene); moreover, DNA molecules containing chitinase-encoding sequences can be obtained, for example, from the ATCC, under Accession Nos. 39637 and 67152; Kramer *et al.* (1993) Insect Biochem. Molec. Biol. 23:691 (nucleotide sequence of a cDNA encoding tobacco hornworm chitinase); and Kawalleck *et al.* (1993) Plant Molec. Biol. 21:673 (nucleotide sequence of the parsley ubi4-2 polyubiquitin gene)); a gene encoding a molecule that stimulates signal transduction (*See, e.g.*, Botella *et al.* (1994) Plant Molec. Biol. 24:757 (nucleotide sequences for mung bean calmodulin cDNA clones); and Griess *et al.* (1994) Plant Physiol. 104:1467 (nucleotide sequence of a maize calmodulin cDNA clone)); a gene that encodes a hydrophobic moment peptide (*See, e.g.*, PCT International Patent Publication No. WO 95/16776 (peptide derivatives of Tachyplesin which inhibit fungal plant pathogens); and PCT International Patent Publication No. WO 95/18855

(synthetic antimicrobial peptides that confer disease resistance)); a gene that encodes a membrane permease, a channel former, or a channel blocker (*See, e.g.*, Jaynes *et al.* (1993) Plant Sci 89:43 (heterologous expression of a cecropin-β lytic peptide analog to render transgenic tobacco plants resistant to *Pseudomonas solanacearum*)); a gene that encodes a viral-invasive protein or complex toxin derived therefrom (*See, e.g.*, Beachy *et al.* (1990) Ann. rev. Phytopathol. 28:451); a gene that encodes an insect-specific antibody or immunotoxin derived therefrom (*See, e.g.*, Taylor *et al.*, Abstract #497, Seventh Int'l Symposium on Molecular Plant-Microbe Interactions (Edinburgh, Scotland) (1994) (enzymatic inactivation in transgenic tobacco via production of single-chain antibody fragments)); a gene encoding a virus-specific antibody (*See, e.g.*, Tavladoraki *et al.* (1993) Nature 366:469 (transgenic plants expressing recombinant antibody genes are protected from virus attack)); a gene encoding a developmental-arrestive protein produced in nature by a pathogen or a parasite (*See, e.g.*, Lamb *et al.* (1992) Bio/Technology 10:1436 (fungal endo α-1,4-D-polygalacturonases facilitate fungal colonization and plant nutrient release by solubilizing plant cell wall homo-α-1,4-D-galacturonase); Toubart *et al.* (1992) Plant J. 2:367 (cloning and characterization of a gene which encodes a bean endopolygalacturonase-inhibiting protein)); a gene encoding a developmental-arrestive protein produced in nature by a plant (*See, e.g.*, Logemann *et al.* (1992) Bio/Technology 10:305 (transgenic plants expressing the barley ribosome-inactivating gene have an increased resistance to fungal disease)).

[0165] In some embodiments, nucleic acids comprising an agronomic gene or nucleotide sequence encoding a polypeptide of interest may also and/or alternatively include, for example and without limitation: genes that confer resistance to an herbicide, such as an herbicide that inhibits the growing point or meristem, for example, an imidazolinone or a sulfonylurea (exemplary genes in this category encode mutant ALS and AHAS enzymes, as described, for example, by Lee *et al.* (1988) EMBO J. 7:1241, and Miki *et al.* (1990) Theor. Appl. Genet. 80:449, respectively); glyphosate resistance as conferred by, *e.g.*, mutant 5-enolpyruvylshikimate-3-phosphate synthase (EPSPs) genes (*via* the introduction of recombinant nucleic acids and/or various forms of *in vivo* mutagenesis of native EPSPs genes (including but not limited to CP4, DMMG, and DGT-28); *aroA* genes and glyphosate acetyl transferase (GAT) genes, respectively); other phosphono compounds, such as glufosinate phosphinothricin acetyl transferase (PAT) genes from *Streptomyces* species, including *Streptomyces hygroscopicus* and *Streptomyces viridichromogenes*); and pyridinoxy or phenoxy proprionic acids and cyclohexones (ACCase inhibitor-encoding genes). *See, e.g.*, U.S. Patents 4,940,835 and

6,248,876 (nucleotide sequences of forms of EPSPs which can confer glyphosate resistance to a plant). A DNA molecule encoding a mutant aroA gene can be obtained under ATCC accession number 39256. *See also* U.S. Pat. No. 4,769,061 (nucleotide sequence of a mutant aroA gene). European patent application No. 0 333 033 and U.S. Pat. No. 4,975,374 disclose nucleotide sequences of glutamine synthetase genes, which may confer resistance to herbicides such as L-phosphinothricin. Nucleotide sequences of exemplary PAT genes are provided in European application No. 0 242 246, and DeGreef *et al.* (1989) Bio/Technology 7:61 (production of transgenic plants that express chimeric bar genes coding for PAT activity). Exemplary of genes conferring resistance to phenoxy proprionic acids and cyclohexones, such as sethoxydim and haloxyfop, include the Acc1-S1, Acc1-S2 and Acc1-S3 genes described by Marshall *et al.* (1992) Theor. Appl. Genet. 83:435. GAT genes capable of conferring glyphosate resistance are described, for example, in WO 2005012515. Genes conferring resistance to 2,4-D, phenoxyproprionic acid and pyridyloxy auxin herbicides are described, for example, in WO 2005107437 and WO 2007053482.

[0166] Nucleic acids comprising an agronomic gene or nucleotide sequence encoding a polypeptide of interest may also include, for example and without limitation: a gene conferring resistance to an herbicide that inhibits photosynthesis, such as a triazine (psbA and gs+ genes) or a benzonitrile (nitrilase gene). *See, e.g.*, Przibila *et al.* (1991) Plant Cell 3:169 (transformation of *Chlamydomonas* with plasmids encoding mutant psbA genes). Nucleotide sequences for nitrilase genes are disclosed in U.S. Patent 4,810,648, and DNA molecules containing these genes are available under ATCC Accession Nos. 53435; 67441; and 67442. *See also* Hayes *et al.* (1992) Biochem. J. 285:173 (cloning and expression of DNA coding for a glutathione S-transferase).

[0167] In some embodiments, nucleic acids comprising an agronomic gene or nucleotide sequence encoding a polypeptide of interest may also and/or alternatively include, genes that confer or contribute to a value-added trait, for example and without limitation: modified fatty acid metabolism, *e.g.*, by transforming a plant with an antisense gene of stearyl-ACP desaturase to increase stearic acid content of the plant (*See, e.g.*, Knultzon *et al.* (1992) Proc. Natl. Acad. Sci. U.S.A. 89:2624); decreased phytate content, *e.g.*, introduction of a phytase-encoding gene may enhance breakdown of phytate, adding more free phosphate to the transformed plant (*See, e.g.*, Van Hartingsveldt *et al.* (1993) Gene 127:87 (nucleotide sequence of an *Aspergillus niger* phytase gene); a gene may be introduced to reduce phytate content- in maize, for example, this may be accomplished by cloning and then reintroducing DNA

associated with the single allele which may be responsible for maize mutants characterized by low levels of phytic acid (*See* Raboy *et al.* (1990) Maydica 35:383)); and modified carbohydrate composition effected, *e.g.*, by transforming plants with a gene encoding an enzyme that alters the branching pattern of starch (*See*, *e.g.*, Shiroza *et al.* (1988) J. Bacteol. 170:810 (nucleotide sequence of *Streptococcus* mutant fructosyltransferase gene); Steinmetz *et al.* (1985) Mol. Gen. Genet. 20:220 (levansucrase gene); Pen *et al.* (1992) Bio/Technology 10:292 (α-amylase); Elliot *et al.* (1993) Plant Molec. Biol. 21:515 (nucleotide sequences of tomato invertase genes); Sogaard *et al.* (1993) J. Biol. Chem. 268:22480 (barley α-amylase gene); and Fisher *et al.* (1993) Plant Physiol. 102:1045 (maize endosperm starch branching enzyme II)).

[0168] In some embodiments, an exogenous nucleic acid is integrated at a *FAD2* locus, so as to modify the *FAD2* locus, wherein the nucleic acid comprises a PTU or ELP, such that, for example, the subsequent site-specific integration of a second exogenous nucleic acid at the site of the PTU or ELP is facilitated. *See, also,* U.S. Application No. 13/889,162.

[0169] Targeting endonuclease-mediated integration of a nucleic acid molecule of interest into a plant genome *via* targeted integration requires delivery of targeting endonucleases or targeting endonuclease-encoding nucleic acid molecules, followed by expression of a functional targeting endonuclease protein in the host. An exogenous nucleic acid is preferably also be present in the host cell at the same time as the targeting endonuclease is delivered or expressed therein, such that functional targeting endonuclease protein induces double-stranded breaks at the target site(s) in the at least one *FAD2* locus, which are then repaired, for example *via* homology-driven integration of the exogenous nucleic acid into the locus. One skilled in the art may envision that expression of a functional targeting endonuclease protein may be achieved by several methods, including, but not limited to, transgenesis of a targeting endonuclease-encoding construct, and transient expression of a targeting endonuclease-encoding construct. In both these cases, expression of a functional targeting endonuclease protein and delivery of an exogenous nucleic acid in the host cell may be simultaneously achieved in order to drive targeted integration at a *FAD2* locus.

[0170] A particular advantage obtained in embodiments utilizing ZFNs as targeting endonucleases, is that the requirement for dimerization of cleavage domains of chimeric zinc finger nucleases imparts a high level of sequence, and hence cleavage, specificity. Since each set of three fingers binds nine consecutive base pairs, two chimeric nucleases effectively demand an 18 bp target if each zinc finger domain has perfect specificity. Any given sequence of this length is predicted to be unique within a single genome (assuming approximately 10^9 bp).

Bibikova *et al.* (2001) Mol. Cell. Biol. 21(1):289-97; Wu *et al.* (2007), *supra.* Furthermore, additional fingers can provide enhanced specificity, Beerli *et al.* (1998) Proc. Natl. Acad. Sci. USA 95:14628-33; Kim and Pabo (1998) Proc. Natl. Acad. Sci. USA 95:2812-7; Liu *et al.* (1997) Proc. Natl. Acad. Sci. USA 94:5525-30, so the number of zinc fingers in each DNA-binding domain may be increased to provide even further specificity. For example, specificity may be further increased by using a pair of 4-, 5-, 6- or more finger ZFNs that recognize a 24 bp sequence. Urnov *et al.* (2005) Nature 435:646-51. Thus, ZFNs may be used such that a recognition sequence is introduced into the host plant genome is unique within the genome.

B. Nucleic acid molecules comprising a nucleotide sequence encoding a targeting endonuclease

[0171] In some embodiments, a nucleotide sequence encoding a targeting endonuclease may be engineered by manipulation (*e.g.*, ligation) of native nucleotide sequences encoding polypeptides comprised within the targeting endonuclease. For example, the nucleotide sequence of a gene encoding a protein comprising a DNA-binding polypeptide may be inspected to identify the nucleotide sequence of the gene that corresponds to the DNA-binding polypeptide, and that nucleotide sequence may be used as an element of a nucleotide sequence encoding a targeting endonuclease comprising the DNA-binding polypeptide. Alternatively, the amino acid sequence of a targeting endonuclease may be used to deduce a nucleotide sequence encoding the targeting endonuclease, for example, according to the degeneracy of the genetic code.

[0172] In exemplary nucleic acid molecules comprising a nucleotide sequence encoding a targeting endonuclease, the last codon of a first polynucleotide sequence encoding a nuclease polypeptide, and the first codon of a second polynucleotide sequence encoding a DNA-binding polypeptide, may be separated by any number of nucleotide triplets, *e.g.*, without coding for an intron or a "STOP." Likewise, the last codon of a nucleotide sequence encoding a first polynucleotide sequence encoding a DNA-binding polypeptide, and the first codon of a second polynucleotide sequence encoding a nuclease polypeptide, may be separated by any number of nucleotide triplets. In these and further embodiments, the last codon of the last (*i.e.*, most 3' in the nucleic acid sequence) of a first polynucleotide sequence encoding a nuclease polypeptide, and a second polynucleotide sequence encoding a DNA-binding polypeptide, may be fused in phase-register with the first codon of a further polynucleotide coding sequence directly contiguous thereto, or separated therefrom by no more than a short peptide sequence, such as

that encoded by a synthetic nucleotide linker (*e.g.*, a nucleotide linker that may have been used to achieve the fusion). Examples of such further polynucleotide sequences include, for example and without limitation, tags, targeting peptides, and enzymatic cleavage sites. Likewise, the first codon of the most 5' (in the nucleic acid sequence) of the first and second polynucleotide sequences may be fused in phase-register with the last codon of a further polynucleotide coding sequence directly contiguous thereto, or separated therefrom by no more than a short peptide sequence.

[0173] A sequence separating polynucleotide sequences encoding functional polypeptides in a targeting endonuclease (*e.g.*, a DNA-binding polypeptide and a nuclease polypeptide) may, for example, include any sequence, such that the amino acid sequence encoded is not likely to significantly alter the translation of the targeting endonuclease. Due to the autonomous nature of known nuclease polypeptides and known DNA-binding polypeptides, intervening sequences will not in examples interfere with the respective functions of these structures.

C. Vectors and expression constructs

[0174] In some embodiments, at least one nucleic acid molecule(s) comprising at least one exogenous polynucleotide sequence encoding a polypeptide of interest, and/or a targeting endonuclease, may be introduced into a cell, tissue, or organism for expression therein. For example, a nucleic acid molecule comprising a polynucleotide sequence encoding a targeting endonuclease that specifically recognizes a nucleotide sequence comprised within at least one FAD2 locus may be introduced into a cell for expression of the targeting endonuclease, and a nucleic acid molecule comprising a polynucleotide sequence encoding a polypeptide of interest may be introduced into the cell, such that the polynucleotide sequence encoding the polypeptide of interest is integrated into the at least one *FAD2* locus, *e.g.*, by homologous recombination following introduction of a double strand break at the locus by the expressed targeting endonuclease, and the polypeptide of interest is expressed from the integrated polynucleotide sequence.

[0175] In some embodiments, a nucleic acid molecule such as one of the foregoing may, for example, be a vector system including, for example and without limitation, a linear plasmid, or a closed circular plasmid. In particular examples, the vector may be an expression vector. Nucleic acid sequences according to particular embodiments may, for example, be integrated into a vector, such that the nucleic acid sequence is operably linked to one or more

regulatory sequences. Many vectors are available for this purpose, and selection of the particular vector may depend, for example, on the size of the nucleic acid to be inserted into the vector, the particular host cell to be transformed with the vector, and/or the amount of any encoded polypeptide that is desired to be expressed. A vector typically contains various components, the identity of which depend on a function of the vector (*e.g.*, amplification of DNA or expression of DNA), and the particular host cell(s) with which the vector is compatible.

[0176] In some embodiments, a regulatory sequence operably linked to one or more coding sequence(s) may be a promoter sequence that functions in a host cell, such as a bacterial cell, algal cell, fungal cell, or plant cell, wherein the nucleic acid molecule is to be amplified or expressed. Some embodiments may include a plant transformation vector that comprises a nucleotide sequence comprising at least one regulatory sequence operably linked to one or more nucleotide sequence(s) encoding a polypeptide of interest or a targeting endonuclease, wherein the one or more nucleotide sequence(s) may be expressed, under the control of the regulatory sequence(s), in a plant cell, tissue, or organism to produce the polypeptide of interest or the targeting endonuclease.

[0177] Promoters suitable for use in nucleic acid molecules according to some embodiments include those that are inducible, tissue-specific, viral, synthetic, or constitutive, all of which are well known in the art. Non-limiting examples of promoters that may be useful in embodiments of the invention are provided by: U.S. Patent Nos. 6,437,217 (maize RS81 promoter); 5,641,876 (rice actin promoter); 6,426,446 (maize RS324 promoter); 6,429,362 (maize PR-1 promoter); 6,232,526 (maize A3 promoter); 6,177,611 (constitutive maize promoters); 5,322,938, 5,352,605, 5,359,142, and 5,530,196 (35S promoter); 6,433,252 (maize L3 oleosin promoter); 6,429,357 (rice actin 2 promoter, and rice actin 2 intron); 6,294,714 (light-inducible promoters); 6,140,078 (salt-inducible promoters); 6,252,138 (pathogen-inducible promoters); 6,175,060 (phosphorous deficiency-inducible promoters); 6,388,170 (bidirectional promoters); 6,635,806 (gamma-coixin promoter); 5,447,858 (soybean heat shock promoter); and U.S. Patent Application Serial No. 09/757,089 (maize chloroplast aldolase promoter).

[0178] Additional exemplary promoters include the nopaline synthase (NOS) promoter (Ebert *et al.* (1987) Proc. Natl. Acad. Sci. USA 84(16):5745-9); the octopine synthase (OCS) promoter (which is carried on tumor-inducing plasmids of *Agrobacterium tumefaciens*); the caulimovirus promoters such as the cauliflower mosaic virus (CaMV) 19S promoter (Lawton *et al.* (1987) Plant Mol. Biol. 9:315-24); the CaMV 35S promoter (Odell *et al.* (1985) Nature 313:810-2; the figwort mosaic virus 35S-promoter (Walker *et al.* (1987) Proc. Natl. Acad. Sci.

USA 84(19):6624-8); the sucrose synthase promoter (Yang and Russell (1990) Proc. Natl. Acad. Sci. USA 87:4144-8); the R gene complex promoter (Chandler *et al.* (1989) Plant Cell 1:1175-83); the chlorophyll a/b binding protein gene promoter; CaMV35S (U.S. Patent Nos. 5,322,938, 5,352,605, 5,359,142, and 5,530,196); FMV35S (U.S. Patent Nos. 6,051,753, and 5,378,619); a PC1SV promoter (U.S. Patent No. 5,850,019); the SCP1 promoter (U.S. Patent No. 6,677,503); and AGRtu.nos promoters (GenBank Accession No. V00087; Depicker *et al.* (1982) J. Mol. Appl. Genet. 1:561-73; Bevan *et al.* (1983) Nature 304:184-7).

[0179] In particular embodiments, nucleic acid molecules may comprise a tissue-specific promoter. A tissue-specific promoter is a nucleotide sequence that directs a higher level of transcription of an operably linked nucleotide sequence in the tissue for which the promoter is specific, relative to the other tissues of the organism. Examples of tissue-specific promoters include, without limitation: tapetum-specific promoters; anther-specific promoters; pollen-specific promoters (*See, e.g.*, U.S. Patent No. 7,141,424, and International PCT Publication No. WO 99/042587); ovule-specific promoters; (See, e.g., U.S. Patent Application No. 2001/047525 A1); fruit-specific promoters (*See, e.g.*, U.S. Patent Nos. 4,943,674, and 5,753,475); and seed-specific promoters (*See, e.g.*, U.S. Patent Nos. 5,420,034, and 5,608,152). In some embodiments, a developmental stage-specific promoter (*e.g.*, a promoter active at a later stage in development) may be used.

[0180] Additional regulatory sequences that may in some embodiments be operably linked to a nucleic acid molecule include 5' UTRs located between a promoter sequence and a coding sequence that function as a translation leader sequence. The translation leader sequence is present in the fully-processed mRNA, and it may affect processing of the primary transcript, and/or RNA stability. Examples of translation leader sequences include maize and petunia heat shock protein leaders (U.S. Patent No. 5,362,865), plant virus coat protein leaders, plant rubisco leaders, and others. *See, e.g.*, Turner and Foster (1995) Molecular Biotech. 3(3):225-36. Non-limiting examples of 5' UTRs are provided by: GmHsp (U.S. Patent No. 5,659,122); PhDnaK (U.S. Patent No. 5,362,865); AtAnt1; TEV (Carrington and Freed (1990) J. Virol. 64:1590-7); and AGRtunos (GenBank Accession No. V00087; and Bevan *et al.* (1983), *supra*).

[0181] Additional regulatory sequences that may in some embodiments be operably linked to a nucleic acid molecule also include 3' non-translated sequences, 3' transcription termination regions, or poly-adenylation regions. These are genetic elements located downstream of a nucleotide sequence, and include polynucleotides that provide polyadenylation signal, and/or other regulatory signals capable of affecting transcription or mRNA processing.

The polyadenylation signal functions in plants to cause the addition of polyadenylate nucleotides to the 3' end of the mRNA precursor. The polyadenylation sequence can be derived from a variety of plant genes, or from T-DNA genes. A non-limiting example of a 3' transcription termination region is the nopaline synthase 3' region (nos 3'; Fraley *et al.* (1983) Proc. Natl. Acad. Sci. USA 80:4803-7). An example of the use of different 3' nontranslated regions is provided in Ingelbrecht *et al.* (1989) Plant Cell 1:671-80. Non-limiting examples of polyadenylation signals include one from a *Pisum sativum* RbcS2 gene (Ps.RbcS2-E9; Coruzzi *et al.* (1984) EMBO J. 3:1671-9) and AGRtu.nos (GenBank Accession No. E01312).

[0182] Additional information regarding regulatory sequences that may be useful in particular embodiments is described, for example, in Goeddel (1990) "Gene Expression Technology," Methods Enzymol. 185, Academic Press, San Diego, CA.

[0183] A recombinant nucleic acid molecule or vector may comprise a selectable marker that confers a selectable phenotype on a transformed cell, such as a plant cell. Selectable markers may also be used to select for cells or organisms that comprise a nucleic acid molecule comprising the selectable marker. A marker may encode biocide resistance, antibiotic resistance (*e.g.*, kanamycin, Geneticin (G418), bleomycin, and hygromycin), or herbicide resistance (*e.g.*, glyphosate). Examples of selectable markers include, but are not limited to: a *neo* gene that confers kanamycin resistance and can be selected for using, *e.g.*, kanamycin and G418; a *bar* gene that confers bialaphos resistance; a mutant EPSP synthase gene that confers glyphosate resistance; a nitrilase gene that confers resistance to bromoxynil; a mutant acetolactate synthase gene (ALS) that confers imidazolinone or sulfonylurea resistance; and a methotrexate-resistant DHFR gene. Multiple selectable markers are available that confer resistance to chemical agents including, for example and without limitation, ampicillin; bleomycin; chloramphenicol; gentamycin; hygromycin; kanamycin; lincomycin; methotrexate; phosphinothricin; puromycin; spectinomycin; rifampicin; streptomycin; and tetracycline. Examples of such selectable markers are illustrated in, *e.g.*, U.S. Patents 5,550,318; 5,633,435; 5,780,708 and 6,118,047.

[0184] A nucleic acid molecule or vector may also or alternatively include a screenable marker. Screenable markers may be used to monitor expression. Exemplary screenable markers include a β-glucuronidase or uidA gene (*GUS*) which encodes an enzyme for which various chromogenic substrates are known (Jefferson *et al.* (1987) Plant Mol. Biol. Rep. 5:387-405); an R-locus gene, which encodes a product that regulates the production of anthocyanin pigments (red color) in plant tissues (Dellaporta *et al.* (1988) "Molecular cloning of the maize R-nj allele by transposon tagging with Ac." In 18th Stadler Genetics Symposium, P. Gustafson and R.

Appels, eds., Plenum, NY (pp. 263-82); a β-lactamase gene (Sutcliffe *et al.* (1978) Proc. Natl. Acad. Sci. USA 75:3737-41); a gene which encodes an enzyme for which various chromogenic substrates are known (*e.g.*, PADAC, a chromogenic cephalosporin); a luciferase gene (Ow *et al.* (1986) Science 234:856-9); a *xylE* gene that encodes a catechol dioxygenase that converts chromogenic catechols (Zukowski *et al.* (1983) Gene 46(2-3):247-55); an amylase gene (Ikatu *et al.* (1990) Bio/Technol. 8:241-2); a tyrosinase gene which encodes an enzyme capable of oxidizing tyrosine to DOPA and dopaquinone, which in turn condenses to melanin (Katz *et al.* (1983) J. Gen. Microbiol. 129:2703-14); and an α-galactosidase.

[0185] All of the nucleotide sequences that encode, for example, a particular polypeptide of interest or a particular targeting endonuclease, will be immediately recognizable by those of skill in the art. The degeneracy of the genetic code provides a finite number of coding sequences for a particular amino acid sequence. The selection of a particular sequence to encode a polypeptide according to embodiments of the invention is within the discretion of the practitioner. Different coding sequences may be desirable in different applications.

[0186] In some embodiments, it may be desirable to modify the nucleotides of a nucleic acid, for example, to enhance expression of a polynucleotide sequence comprised within the nucleic acid in a particular host. The genetic code is redundant with 64 possible codons, but most organisms preferentially use a subset of these codons. The codons that are utilized most often in a species are called optimal codons, and those not utilized very often are classified as rare or low-usage codons. Zhang *et al.* (1991) Gene 105:61-72. Codons may be substituted to reflect the preferred codon usage of a particular host in a process sometimes referred to as "codon optimization." Optimized coding sequences containing codons preferred by a particular prokaryotic or eukaryotic host may be prepared, for example, to increase the rate of translation or to produce recombinant RNA transcripts having desirable properties (*e.g.*, a longer half-life, as compared with transcripts produced from a non-optimized sequence).

[0187] Nucleic acids may be introduced into a host cell in embodiments of the invention by any method known to those of skill in the art, including, for example and without limitation: by transformation of protoplasts (*See, e.g.*, U.S. Patent 5,508,184); by desiccation/inhibition-mediated DNA uptake (*See, e.g.*, Potrykus *et al.* (1985) Mol. Gen. Genet. 199:183-8); by electroporation (*See, e.g.*, U.S. Patent 5,384,253); by agitation with silicon carbide fibers (*See, e.g.*, U.S. Patents 5,302,523 and 5,464,765); by *Agrobacterium*-mediated transformation (*See, e.g.*, U.S. Patents 5,563,055, 5,591,616, 5,693,512, 5,824,877, 5,981,840, and 6,384,301); and by acceleration of DNA-coated particles (*See, e.g.*, U.S. Patents 5,015,580, 5,550,318,

5,538,880, 6,160,208, 6,399,861, and 6,403,865). Through the application of techniques such as these, the cells of virtually any species may be stably transformed. In some embodiments, transforming DNA is integrated into the genome of the host cell. In the case of multicellular species, transgenic cells may be regenerated into a transgenic organism. Any of these techniques may be used to produce a transgenic plant, for example, comprising one or more nucleic acid sequences of the invention in the genome of the transgenic plant.

[0188] The most widely-utilized method for introducing an expression vector into plants is based on the natural transformation system of *Agrobacterium*. *A. tumefaciens* and *A. rhizogenes* are plant pathogenic soil bacteria that genetically transform plant cells. The T_i and R_i plasmids of *A. tumefaciens* and *A. rhizogenes*, respectively, carry genes responsible for genetic transformation of the plant. The T_i (tumor-inducing)-plasmids contain a large segment, known as T-DNA, which is transferred to transformed plants. Another segment of the T_i plasmid, the vir region, is responsible for T-DNA transfer. The T-DNA region is bordered by left-hand and right-hand borders that are each composed of terminal repeated nucleotide sequences. In some modified binary vectors, the tumor-inducing genes have been deleted, and the functions of the vir region are utilized to transfer foreign DNA bordered by the T-DNA border sequences. The T-region may also contain, for example, a selectable marker for efficient recovery of transgenic plants and cells, and a multiple cloning site for inserting sequences for transfer such as a nucleic acid encoding a fusion protein of the invention.

[0189] Thus, in some embodiments, a plant transformation vector is derived from a T_i plasmid of *A. tumefaciens* (*See, e.g.*, U.S. Patent Nos. 4,536,475, 4,693,977, 4,886,937, and 5,501,967; and European Patent EP 0 122 791) or a R_i plasmid of *A. rhizogenes*. Additional plant transformation vectors include, for example and without limitation, those described by Herrera-Estrella *et al.* (1983) Nature 303:209-13; Bevan *et al.* (1983), *supra*; Klee *et al.* (1985) Bio/Technol. 3:637-42; and in European Patent EP 0 120 516, and those derived from any of the foregoing. Other bacteria, such as *Sinorhizobium*, *Rhizobium*, and *Mesorhizobium*, that naturally interact with plants can be modified to mediate gene transfer to a number of diverse plants. These plant-associated symbiotic bacteria can be made competent for gene transfer by acquisition of both a disarmed T_i plasmid and a suitable binary vector.

[0190] After providing exogenous DNA to recipient cells, transformed cells are generally identified for further culturing and plant regeneration. In order to improve the ability to identify transformed cells, one may desire to employ a selectable or screenable marker gene, as previously set forth, with the vector used to generate the transformant. In the case where a

selectable marker is used, transformed cells are identified within the potentially transformed cell population by exposing the cells to a selective agent or agents. In the case where a screenable marker is used, cells may be screened for the desired marker gene trait.

[0191] Cells that survive the exposure to the selective agent, or cells that have been scored positive in a screening assay, may be cultured in media that supports regeneration of plants. In some embodiments, any suitable plant tissue culture media (*e.g.*, MS and N6 media) may be modified by including further substances, such as growth regulators. Tissue may be maintained on a basic media with growth regulators until sufficient tissue is available to begin plant regeneration efforts, or following repeated rounds of manual selection, until the morphology of the tissue is suitable for regeneration (*e.g.*, at least 2 weeks), then transferred to media conducive to shoot formation. Cultures are transferred periodically until sufficient shoot formation has occurred. Once shoots are formed, they are transferred to media conducive to root formation. Once sufficient roots are formed, plants can be transferred to soil for further growth and maturity.

[0192] To confirm the presence of a nucleic acid molecule of interest (for example, a nucleotide sequence encoding a polypeptide comprising at least one fusion protein of the invention) in a regenerating plant, a variety of assays may be performed. Such assays include, for example: molecular biological assays, such as Southern and Northern blotting, PCR, and nucleic acid sequencing; biochemical assays, such as detecting the presence of a protein product, *e.g.*, by immunological means (ELISA and/or Western blots) or by enzymatic function; plant part assays, such as leaf or root assays; and analysis of the phenotype of the whole regenerated plant.

[0193] Integration events may be analyzed, for example, by PCR amplification using, *e.g.*, oligonucleotide primers that are specific for a nucleotide sequence of interest. PCR genotyping is understood to include, but not be limited to, polymerase-chain reaction (PCR) amplification of genomic DNA derived from isolated host plant tissue predicted to contain a nucleic acid molecule of interest integrated into the genome, followed by standard cloning and sequence analysis of PCR amplification products. Methods of PCR genotyping have been well described (*see, e.g.*, Rios, G. *et al.* (2002) Plant J. 32:243-53), and may be applied to genomic DNA derived from any plant species or tissue type, including cell cultures.

[0194] A transgenic plant formed using *Agrobacterium*-dependent transformation methods typically contains a single to multiple copies of recombinant DNA. The single recombinant DNA sequence is referred to as a "transgenic event" or "integration event." Such

transgenic plants are heterozygous for the inserted DNA sequence. In some embodiments, a transgenic plant homozygous with respect to a transgene may be obtained by sexually mating (selfing) an independent segregant transgenic plant that contains a single exogenous gene sequence to itself, for example, an F_0 plant, to produce F_1 seed. One fourth of the F_1 seed produced will be homozygous with respect to the transgene. Germinating F_1 seed results in plants that can be tested for heterozygosity, typically using a SNP assay or a thermal amplification assay that allows for the distinction between heterozygotes and homozygotes (*i.e.*, a zygosity assay).

[0195] In addition to direct transformation of a plant or plant cell with a nucleic acid molecule in some embodiments, transgenic plants may be prepared in particular embodiments by crossing a first plant having at least one transgenic event with a second plant lacking such an event. For example, a nucleic acid comprising at least one modified *FAD2* locus, wherein an exogenous nucleic acid has been integrated in a site-specific manner, may be introduced into a first plant line that is amenable to transformation, to produce a transgenic plant, which transgenic plant may be crossed with a second plant line to introgress the at least one modified *FAD2* locus (and therefore the exogenous nucleic acid) into the second plant line.

[0196] To confirm the presence of a nucleic acid molecule of interest in regenerating plants, a variety of assays may be performed. Such assays include, for example: molecular biological assays, such as Southern and Northern blotting and PCR; biochemical assays, such as detecting the presence of a protein product, *e.g.*, by immunological means (ELISA and/or Western blots) or by enzymatic function; plant part assays, such as leaf or root assays; and analysis of the phenotype of the whole regenerated plant.

[0197] Targeted integration events may be screened, for example, by PCR amplification using, *e.g.*, oligonucleotide primers specific for nucleic acid molecules of interest. PCR genotyping is understood to include, but not be limited to, polymerase-chain reaction (PCR) amplification of genomic DNA derived from isolated host plant callus tissue predicted to contain a nucleic acid molecule of interest integrated into the genome, followed by standard cloning and sequence analysis of PCR amplification products. Methods of PCR genotyping have been well described (for example, Rios, G. *et al.* (2002) Plant J. 32:243-53) and may be applied to genomic DNA derived from any plant species or tissue type, including cell cultures. Combinations of oligonucleotide primers that bind to both target sequence and introduced sequence may be used sequentially or multiplexed in PCR amplification reactions. Oligonucleotide primers designed to anneal to the target site, introduced nucleic acid sequences,

and/or combinations of the two are feasible. Thus, PCR genotyping strategies may include (but are not limited to) amplification of specific sequences in the plant genome, amplification of multiple specific sequences in the plant genome, amplification of non-specific sequences in the plant genome, or combinations thereof. One skilled in the art may devise additional combinations of primers and amplification reactions to interrogate the genome. For example, a set of forward and reverse oligonucleotide primers may be designed to anneal to nucleic acid sequence(s) specific for the target outside the boundaries of the introduced nucleic acid sequence.

[0198] Forward and reverse oligonucleotide primers may be designed to anneal specifically to an introduced nucleic acid molecule of interest, for example, at a sequence corresponding to a coding region within the nucleic acid molecule of interest, or other parts of the nucleic acid molecule of interest. These primers may be used in conjunction with the primers described above. Oligonucleotide primers may be synthesized according to a desired sequence, and are commercially available (*e.g.*, from Integrated DNA Technologies, Inc., Coralville, IA). Amplification may be followed by cloning and sequencing, or by direct sequence analysis of amplification products. One skilled in the art might envision alternative methods for analysis of amplification products generated during PCR genotyping. In one embodiment, oligonucleotide primers specific for the gene target are employed in PCR amplifications.

VI. Transgenic plants and plant materials comprising a nucleic acid integrated at a FAD2 performance locus

[0199] In some embodiments, a transgenic plant is provided, wherein the plant comprises a plant cell comprising at least one modified (*e.g.*, disrupted and/or targeted integration of an exogenous sequence) *FAD2* locus. In particular embodiments, such a plant may be produced by transformation of a plant tissue or plant cell, and regeneration of a whole plant. In further embodiments, such a plant may be obtained through introduction of an exogenous nucleic acid at the at least one FAD2 locus in a site-specific manner, or through introgression of the modified FAD2 locus into a germplasm. Plant materials comprising such a plant cell are also provided. Such a plant material may be obtained from a plant comprising the plant cell.

[0200] A transgenic plant or plant material comprising a plant cell comprising at least one modified *FAD2* locus may in some embodiments exhibit one or more of the following

characteristics: expression of a targeting endonuclease in a cell of the plant; expression of a polypeptide of interest in a cell of the plant (or in a plastid therein); expression of a targeting endonuclease in the nucleus of a cell of the plant; localization of a targeting endonuclease in a cell of the plant; integration at a *FAD2* locus in the genome of a cell of the plant; integration of a nucleotide sequence encoding a polypeptide of interest or an agronomic gene at a *FAD2* locus in the genome of a cell of the plant; and/or the presence of an RNA transcript corresponding to a coding sequence integrated at a *FAD2* locus in the genome of a cell of the plant. Such a plant may additionally have one or more desirable traits, including, for example and without limitation, those resulting from the expression of an endogenous or transgenic nucleotide sequence, the expression of which is regulated by a polypeptide of interest or an agronomic gene integrated at a *FAD2* locus in the genome of a cell of the plant; resistance to insects, other pests, and disease-causing agents; tolerances to herbicides; enhanced stability, yield, or shelf-life; environmental tolerances; pharmaceutical production; industrial product production; and nutritional enhancements.

[0201] A transgenic plant according to the invention may be any plant capable of being transformed with a nucleic acid that is subsequently integrated in at least one *FAD2* locus according to methods described herein. Accordingly, the plant may be a dicot or monocot. Non-limiting examples of dicotyledonous plants usable in the present methods include *Arabidopsis*, alfalfa, beans, broccoli, cabbage, canola, carrot, cauliflower, celery, Chinese cabbage, cotton, cucumber, eggplant, lettuce, melon, pea, pepper, peanut, potato, pumpkin, radish, rapeseed, spinach, soybean, squash, sugarbeet, sunflower, tobacco, tomato, and watermelon. Non-limiting examples of monocotyledonous plants usable in the present methods include corn, barley, onion, rice, sorghum, wheat, rye, millet, sugarcane, oat, triticale, switchgrass, and turfgrass. Transgenic plants according to the invention may be used or cultivated in any manner.

[0202] Some embodiments also provide commodity products produced from transgenic plants of the invention. Commodity products include, for example and without limitation: food products, meals, oils, or crushed or whole grains or seeds of a plant comprising one or more nucleotide sequences integrated in at least one FAD2 locus. The detection of one or more such nucleotide sequences in one or more commodity or commodity products is de facto evidence that the commodity or commodity product was at least in part produced from a transgenic plant produced according to an embodiment of the invention. In some embodiments, a transgenic plant or seed comprising a plant cell comprising at least one modified FAD2 locus may

comprise at least one other transgenic event in its genome, including without limitation: a transgenic event from which is transcribed an RNAi molecule; a gene encoding an insecticidal protein (*e.g.*, a *Bacillus thuringiensis* insecticidal protein); an herbicide tolerance gene (*e.g.*, a gene providing tolerance to glyphosate); and a gene contributing to a desirable phenotype in the transgenic plant (*e.g.*, increased yield, altered fatty acid metabolism, or restoration of cytoplasmic male sterility).

[0203] A transgenic plant comprising a plant cell comprising at least one modified *FAD2* locus may have one or more desirable traits. Such traits can include, for example: resistance to insects, other pests, and disease-causing agents; tolerances to herbicides; enhanced stability, yield, or shelf-life; environmental tolerances; pharmaceutical production; industrial product production; and nutritional enhancements. The desirable traits may be conferred by one or more nucleic acid molecules integrated by targeted recombination at the FAD2 locus that are expressed in the plant exhibiting the desirable traits. Thus, in some embodiments, the desired trait can be due to the presence of a transgene(s) in the plant, which is introduced into the genome of the plant at the site of at least one modified FAD2 locus. In an additional embodiment, the desirable trait can be obtained through conventional breeding, which trait may be conferred by one or more nucleic acid molecules integrated by targeted recombination at the at least one modified FAD2 locus.

[0204] Transgenic plants according to the invention may be used or cultivated in any manner, wherein presence of at least one modified FAD2 locus is desirable. Accordingly, a plant may be engineered to, *inter alia*, have one or more desired traits, by being transformed with nucleic acid molecules that are subsequently integrated in a site-specific manner in at least one FAD2 locus according to the invention, and cropped and cultivated by any method known to those of skill in the art.

VII. Marker-assisted breeding of transgenic plants comprising a nucleic acid integrated at a FAD2 performance locus

[0205] Molecular markers that are linked (*e.g.*, tightly-linked) to *fad2* in *Brasicca* spp. are provided. For example, DNA segments containing sequences involved in the HO trait (*fad2*) are identified. These segments are located around and between markers that are linked (*e.g.*, tightly-linked) to the mutant alleles in a genomic linkage group. Thus, nucleic acid molecules comprising a mutant FAD2 gene having an inactivating mutation are also provided. The segments identified, and the markers thereof, are included in the present subject matter, in part,

by their position in linkage groups in the *B. napus* genome. For example, FAD2 and molecular markers linked thereto may be located in linkage groups N5 and N1.

[0206] All references, including publications, patents, and patent applications, cited herein are hereby incorporated by reference to the extent they are not inconsistent with the explicit details of this disclosure, and are so incorporated to the same extent as if each reference were individually and specifically indicated to be incorporated by reference and were set forth in its entirety herein. The references discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the inventors are not entitled to antedate such disclosure by virtue of prior invention. The following examples are provided to illustrate certain particular features and/or embodiments. The examples should not be construed to limit the disclosure to the particular features or embodiments exemplified.

EXAMPLES

EXAMPLE 1: IDENTIFICATION OF FAD2 TARGET SEQUENCES FROM A BACTERIAL ARTIFICIAL CHROMOSOME LIBRARY

BAC LIBRARY CONSTRUCTION

[0207] A Bacterial Artificial Chromosome (BAC) library was sourced from a commercial vendor (Amplicon Express, Pullman, WA). The BAC library included 110,592 BAC clones containing high molecular weight genomic DNA (gDNA) fragments isolated from *Brassica napus* L. var. DH10275. The gDNA was digested with either the *BamHI* or *HindIII* restriction enzyme. Isolated gDNA fragments of about 135 Kbp were ligated into the pCC1BAC vector (Epicentre, Madison, WI) and transformed into *Escherichia coli* str. DH10B (Invitrogen). The BAC library was made up of an even number of BAC clones that were constructed using the two different restriction enzymes. As such, the *Hind III* constructed BAC library was contained in 144 individual 384-well plates. Likewise, the *BamHI* constructed BAC library was contained in 144 individual 384-well plates. A total of 110,592 BAC clones were isolated and arrayed into 288 individual 384-well plates. Each of the 288 individual 384 well plates were provided by the vendor as a single DNA extraction for rapid PCR based screening. The resulting BAC library covers approximately 15 Gbp of gDNA, which corresponds to a 12-fold genome coverage of *Brassica napus* L. var. DH10275genome (estimate of the *Brassica napus* L. genome is ca. 1.132 Gbp as described in Johnston *et al.* (2005) Annals of Botany 95:229-235).

SEQUENCE ANALYSIS OF FAD2 CODING SEQUENCES ISOLATED FROM THE BAC LIBRARY

[0208] The constructed BAC library was used to isolate FAD2 gene coding sequences. Sequencing experiments were conducted to identify the specific gene sequences of four FAD2 gene paralogs from *Brassica napus* L. var. DH10275.

[0209] The FAD2 gene sequence was initially identified within the model species *Arabidopsis thaliana*. The gene sequence is listed in Genbank as Locus Tag: At3g12120. Comparative genomic relationships between the model plant species *Arabidopsis thaliana* and the diploid *Brassica rapa,* one of the progenitors of the tetraploid *Brassica napus,* have been previously described. (Schranz *et al.* (2006) Trends in Plant Science 11(11):535-542). With specific relation to the FAD2 gene the comparative analysis predicted that 3-4 copies of the gene may occur within the diploid *Brassica* genome. Additional genetic mapping studies were completed by Scheffler *et al.* (1997) Theoretical and Applied Genetics 94; 583-591. The results of these genetic mapping studies indicated that four copies of the FAD2 gene were present in *Brassica napus*.

[0210] Sequencing analysis of the BAC library which was constructed from *B. napus* L. var. DH12075 resulted in the isolation of four BAC sequences (SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, and SEQ ID NO:4) from which the coding sequences for the FAD2A (SEQ ID NO:5), FAD2-1 (SEQ ID NO:6), FAD2-2 (SEQ ID NO:7), and FAD2-3(SEQ ID NO:8) genes were determined. The FAD2A, FAD2-1, FAD2-2, and FAD2-3 gene sequences were identified and genetically mapped. Sequence analysis of the four FAD2 genes was conducted using a sequence alignment program and a neighbor-joining tree using percentage of identity. The sequence alignment was made via the AlignX® program from the Vector NTI Advance 11.0 computer program (Life Technologies, Carlsbad, CA) and is shown in Figure 1. AlignX® uses a modified Clustal W algorithm to generate multiple sequence alignments of either protein or nucleic acid sequences for similarity comparisons and for annotation.

[0211] The neighbour-joining tree was created with Jalview v2.3® software and is shown in Figure 2. (Waterhouse *et al.* (2009) Bioinformatics 25 (9) 1189-1191). As shown in Figure 2, the analysis of the isolated sequences indicated that the FAD2A and FAD2-3 sequences shared high levels of sequence similarity and that, likewise, FAD2-1 and FAD2-2 shared high levels of sequence similarity. The four sequences can be categorized in two

clades, wherein FAD2A and FAD2-3 comprise a first clade, and FAD2-1 and FAD2-2 comprise a second clade.

[0212] Next, the newly isolated FAD2 sequences from *Brassica napus* were used to BLAST genomic libraries isolated from a *Brassica rapa* genomic BAC library and *Brassica oleracea* shotgun genomic sequence reads. Both, *Brassica rapa* and *Brassica oleracea* are diploid progenitors of *Brassica napus* which is an amphidiploid species (AC genome, n = 19). *Brassica napus* derived from a recent hybridization event between *Brassica rapa* (A sub-genome, n = 10) and *Brassica oleracea* (C sub-genome, n = 9). The diploid progenitor sequences were compared to the four different FAD2 coding sequences isolated from *Brassica napus* using a BLASTn analysis. This sequence analysis identified specific, annotated gene sequences from *Brassica rapa* and *Brassica oleracea* which shared the highest sequence similarity to the newly discovered *Brassica napus* FAD2 sequences. Table 1 lists the newly identified FAD2 coding sequence and the corresponding progenitor reference sequence accession number and source organism.

Table 1: FAD2 sequences from *Brassica napus* and the corresponding progenitor organism and related FAD sequence accession number.

Isolated gene sequence	**Progenitor organism and sequence accession number**	
FAD2A	*B.rapa*	Genbank Accession No: KBrB063G23 (A05)
FAD2-3	*B.oleracea*	Genbank Accession No: GSS23580801
FAD2-1	*B.rapa*	Genbank Accession No: KBrB130I19
FAD2-2	*B.oleracea*	Genbank Accession No: GSS 17735412

[0213] The FAD2 genes exist in the *Brassica napus* genome as two copies of each gene per sub-genome. One copy of each gene is located on the A sub-genome, and likewise one copy of each gene is located on the C sub-genome. New naming conventions are described to indicate which sub-genome that each gene is located on. The high levels of sequence similarity between the four different FAD2 coding sequences isolated from the *Brassica napus* BAC genomic DNA library and the progenitor sequence data suggest that

FAD2-3 is a duplicate of the FAD2 sequence from the C sub-genome and could be relabeled as FAD2C; FAD2-1 is a duplicate of the FAD2 sequence from the A sub-genome and could therefore be labeled as FAD2A'; and finally, FAD2-2 is a second copy that was duplicated from the FAD2 sequence of the C sub-genome and could be labeled as FAD2C'.

PCR BASED SCREENING

[0214] A cohort of PCR primers were design to screen the aforementioned BAC library. The primers were designed as either universal primers, which would amplify all members of the gene family, or as gene specific primers for targeted allele amplification. The PCR primers were designed to be 20bp long (+/- 1bp) and contain a G/C content of 50% (+/- 8%). Table 2 lists the primers which were designed and synthesized. The clones of the BAC library were pooled and screened via the Polymerase Chain Reaction (PCR).

Table 2: PCR primer sequences designed for BAC library screening for FAD2 gene identification.

Primer Name	**SEQ ID NO:**	**Sequence**
D_UnivF2_F1	SEQ ID NO:9	ATGGGTGCAGGTGGAAGAATG
D_UnivF2_F2	SEQ ID NO:10	AGCGTCTCCAGATATACATC
D_UnivF2_R1	SEQ ID NO:11	ATGTATATCTGGAGACGCTC
D_UnivF2_R2	SEQ ID NO:12	TAGATACACTCCTTCGCCTC
D_SpecificF2_F3	SEQ ID NO:13	TCTTTCTCCTACCTCATCTG
D_SpecificF2_R3	SEQ ID NO:14	TTCGTAGCTTCCATCGCGTG
D_UnivF2_F4	SEQ ID NO:15	GACGCCACCATTCCAACAC
D_UnivF2_R4	SEQ ID NO:16	ACTTGCCGTACCACTTGATG

[0215] Two different sets of conditions were used for the polymerase chain reactions (PCR). The first series of PCR reactions contained: 1X PCR buffer (containing dNTPs); 1.5 mM $MgCl_2$; 200 μM of 0.25 U Immolase® DNA polymerase (Bioline, London, UK); 250 nM of each primer; and, about 5-10 ng template DNA. A second series of PCR reactions were developed for the amplification of genomic DNA and contained: 5-10 ng of genomic DNA, 1X PCR buffer, 2 mM dNTPs, 0.4 μM forward and reverse primer, and 0.25 U Immolase® DNA polymerase (Bioline, London, UK). Amplifications were pooled into a final volume of 13 μL and amplified using an MJ PTC200® thermocycler (BioRad, Hercules, CA) or an ABI 9700 Gene Amp System® (Life Technologies, Carlsbad, CA). PCR based screening of specific plates was conducted using a 4 dimension screening approach based on the screening system described by Bryan *et al* (Scottish Crops Research Institute annual report: 2001-2002) with the above described PCR conditions. Following PCR based screening of pooled BAC libraries; the amplified PCR product was sequenced using a direct Sanger sequencing method. The amplified 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.

[0216] Following PCR based screening and conformational Sanger sequencing, a collection of plates were identified that contained the various different FAD2 gene family members. A total of four unique FAD2 paralogous gene sequences were identified (Table 2). A total of two plates per each FAD2 paralogous gene sequence were chosen to undergo plate screening to identify the specific well and clone within the plate that contained the FAD2 gene (Table 3). The specific wells were identified for both of the plates and an individual clone was selected for each of the FAD2 gene family members.

Table 3: Identification of the BAC clone plates that provided positive reaction with the detailed PCR primer combinations, along with two plate identities that were taken forward for clone identification within the plate

Gene Name	Primer Sets	Positive Plate Pools	Chosen Plates	Well Id
FAD2A	F4+R1, F1+R1, F1+R4, F3+R3	8, 27, 30, 83, 109, 147, 180, 199, 209, 251, 288	Plate 199 Plate 27	L23 D20

FAD2-1	F1+R4, F4+R1, F1+R1, F2+R2	12, 89, 123, 148, 269	Plate 123 Plate 148	N17 B15
FAD2-2	F4+R1, F1+R1, F1+R4, F2+R2	24, 44, 46, 47, 80, 91, 104, 110, 119, 121, 124, 248	Plate 44 Plate 121	H03 A17
FAD2-3	F1+R4, F4+R1, F1+R1, F3+R3	8, 62, 113, 205, 276	Plate 62 Plate 205	I16 K11

[0217] The single BAC clone, for each identified FAD gene family member, was further analysed via sequencing. The DNA was isolated for the BAC clone and was prepared for sequencing using a Large Construct kit® (Qiagen, Valencia, CA) following the manufacturer's instructions. The extracted BAC DNA was prepared for sequencing using GS-FLX Titanium Technology® (Roche, Indianapolis, IN) following manufacturer's instructions. Sequencing reactions were performed using a physically sectored GS-FLX TI Pico-titer plate® with the BACs pooled in pairs for optimal data output. The BACs were combined in pairs where the FAD2 gene was paired with a FAD3 gene. All generated sequence data was assembled by Newbler v2.0.01.14® (454 Life Sciences, Branford, CT). The assembled contigs were manually assessed for the presence of the corresponding FAD gene using Sequencher v3.7® (GeneCodes, Ann Arbor, MI).

[0218] After the full genomic sequence of all four FAD2 genes had been identified and fully characterized, zinc finger nucleases were designed to bind to the sequences for each specific gene family member.

EXAMPLE 2: DESIGN OF ZINC FINGER BINDING DOMAINS SPECIFIC TO FAD2 GENES

[0219] Novel zinc finger proteins directed against DNA sequences encoding various functional sequences of the FAD2 gene locus were designed essentially as previously described. *See, e.g.*, Urnov *et al.* (2005) Nature 435:646-651. Exemplary target sequence and recognition helices are shown in Table 4 (recognition helix regions designs) and Table 5 (target sites). In Table 5, nucleotides in the target site that are contacted by the ZFP

recognition helices are indicated in uppercase letters; non-contacted nucleotides indicated in lowercase.

[0220] Zinc Finger Nuclease (ZFN) target sites were designed to bind five target sites of FAD2A. The FAD2A zinc finger designs were incorporated into zinc finger expression vectors encoding a protein having at least one finger with a CCHC structure. *See*, U.S. Patent Publication No. 2008/0182332. In particular, the last finger in each protein had a CCHC backbone for the recognition helix. The non-canonical zinc finger-encoding sequences were fused to the nuclease domain of the type IIS restriction enzyme FokI (amino acids 384-579 of the sequence of Wah *et al.*, (1998) Proc. Natl. Acad. Sci. USA 95:10564-10569) via a four amino acid ZC linker and an *opaque-2* nuclear localization signal derived from *Zea mays* to form FAD2A zinc-finger nucleases (ZFNs). Expression of the fusion proteins was driven by a relatively strong constitutive promoter such as a promoter derived from the Cassava Vein Mosaic Virus (CsVMV) promoter and flanked by the *Agrobacterium tumefaciens* ORF23 3'UnTranslated Region (AtuORF23 3'UTR v1). The self-hydrolyzing 2A encoding nucleotide sequence from *Thosea asigna* virus (Szymczak *et al.*, 2004) was added between the two Zinc Finger Nuclease fusion proteins that were cloned into the construct. Exemplary vectors or plasmids are described in Table 5, below.

[0221] The optimal FAD2 zinc finger nucleases were verified for cleavage activity using a budding yeast based system previously shown to identify active nucleases. *See, e.g.*, U.S. Patent Publication No. 20090111119; Doyon *et al.* (2008) *Nat Biotechnol.* 26:702–708; Geurts *et al.* (2009) *Science* 325:433. Zinc fingers for the various functional domains were selected for *in-vivo* use. Of the numerous ZFNs that were designed, produced and tested to bind to the putative FAD genomic polynucleotide target sites, eleven ZFNs were identified as having *in vivo* activity at high levels, and selected for further experimentation. These ZFNs were characterized as being capable of efficiently binding and cleaving the unique FAD2 genomic polynucleotide target sites *in planta*.

Table 4: FAD2 Zinc Finger Designs

ZFP	F1	F2	F3	F4	F5	F6
24800	RSDNLST	HSHARIK	HRSSLRR	RSDHLSE	QNANRIT	N/A

	(SEQ ID NO:94)	SEQ ID NO:95	SEQ ID NO:96	SEQ ID NO:97	SEQ ID NO:98	
24801	DRSNLSR SEQ ID NO:99	HRSSLRR SEQ ID NO:96	TSGNLTR SEQ ID NO:101	MSHHLRD SEQ ID NO:102	DQSNLRA SEQ ID NO:103	N/A
24794	QSGNLAR SEQ ID NO:104	RSDNLSR SEQ ID NO:105	DNNARIN SEQ ID NO:106	DRSNLSR SEQ ID NO:99	RSDHLTQ SEQ ID NO:108	N/A
24795	RSDNLRE SEQ ID NO:109	QSGALAR SEQ ID NO:110	QSGNLAR SEQ ID NO:104	RSDVLSE SEQ ID NO:112	SPSSRRT SEQ ID NO:113	N/A
24810	RSDSLSR SEQ ID NO:114	RKDARIT SEQ ID NO:115	RSDHLSA SEQ ID NO:116	WSSSLYY SEQ ID NO:117	NSRNLRN SEQ ID NO:118	N/A
24811	DQSTLRN SEQ ID NO:119	DRSNLSR SEQ ID NO:99	DRSNLWR SEQ ID NO:121	DRSALSR SEQ ID NO:122	RSDALAR SEQ ID NO:123	N/A
24814	RSDALSR SEQ ID NO:124	DRSDLSR SEQ ID NO:125	RSDHLTQ SEQ ID NO:108	QSGALAR SEQ ID NO:110	QSGNLAR SEQ ID NO:104	N/A
24815	DRSNLSR SEQ ID NO:99	DSSARNT SEQ ID NO:130	DRSSRKR SEQ ID NO:131	QSGDLTR SEQ ID NO:132	LAHHLVQ SEQ ID NO:133	N/A
24818	RSDNLST SEQ ID	HSHARIK SEQ ID	TSGHLSR SEQ ID	RSDNLSV SEQ ID	IRSTLRD SEQ ID	N/A

	NO:94	NO:95	NO:136	NO:137	NO:138	
24819	TSGHLSR SEQ ID NO:136	DRSNLSR SEQ ID NO:99	HRSSLRR SEQ ID NO:96	TSGNLTR SEQ ID NO:101	MSHHLRD SEQ ID NO:102	N/A
24796	RSDALSR SEQ ID NO:124	DRSDLSR SEQ ID NO:125	RSDHLTQ SEQ ID NO:108	QSGALAR SEQ ID NO:110	QSGNLAR SEQ ID NO:104	N/A
24797	RSAVLSE SEQ ID NO:149	TNSNRIT SEQ ID NO:150	LKQHLNE SEQ ID NO:151	QSGALAR SEQ ID NO:110	QSGNLAR SEQ ID NO:104	N/A
24836	DRSNLSR SEQ ID NO:99	QSGDLTR SEQ ID NO:132	QSGALAR SEQ ID NO:110	DRSNLSR SEQ ID NO:99	QRTHLTQ SEQ ID NO:158	N/A
24837	RSDNLSN SEQ ID NO:159	TNSNRIK SEQ ID NO:160	QSSDLSR SEQ ID NO:161	QSSDLRR SEQ ID NO:162	DRSNRIK SEQ ID NO:163	N/A
24844	RSANLAR SEQ ID NO:164	RSDNLTT SEQ ID NO:165	QSGELIN SEQ ID NO:166	RSADLSR SEQ ID NO:167	RSDNLSE SEQ ID NO:168	DRSHLAR SEQ ID NO:169
24845	DRSHLAR SEQ ID NO:169	RSDNLSE SEQ ID NO:168	SKQYLIK SEQ ID NO:172	ERGTLAR SEQ ID NO:173	RSDHLTT SEQ ID NO:174	N/A
24820	QSGALAR SEQ ID	QSGNLAR SEQ ID	DRSHLAR SEQ ID	DRSDLSR SEQ ID	RSDNLTR SEQ ID	N/A

	NO:110	NO:104	NO:169	NO:125	NO:179	
24821	DRSHLAR SEQ ID NO:169	RSDNLSE SEQ ID NO:168	SKQYLIK SEQ ID NO:172	ERGTLAR SEQ ID NO:173	RSDHLTT SEQ ID NO:174	N/A
24828	DRSDLSR SEQ ID NO:125	RSDNLTR SEQ ID NO:179	QRTHLTQ SEQ ID NO:158	RSDNLSE SEQ ID NO:168	ASKTRKN SEQ ID NO:189	N/A
24829	RSDTLSE SEQ ID NO:190	QSHNRTK SEQ ID NO:191	QSDHLTQ SEQ ID NO:192	RSSDLSR SEQ ID NO:193	QSSDLSR SEQ ID NO:161	RSDHLTQ SEQ ID NO:108
24832	RSDSLSR SEQ ID NO:114	RKDARIT SEQ ID NO:115	DRSHLSR SEQ ID NO:198	QSGNLAR SEQ ID NO:104	QSSDLSR SEQ ID NO:161	DRSALAR SEQ ID NO:201
24833	RSDDLSK SEQ ID NO:202	RSDTRKT SEQ ID NO:203	DRSNLSR SEQ ID NO:99	DRSNLWR SEQ ID NO:121	RSDSLSR SEQ ID NO:114	NNDHRKT SEQ ID NO:207

Table 5: Target Sites of FAD2 Zinc Fingers

ZFP	**Plasmid No.**	**Target Site (5' to 3')**	**ZFP target/binding site present in SEQ ID Nos.**
24800	pDAB104001	ccCAAAGGGTTGTTGAGgtacttgccgt	SEQ ID NO:17
24801	pDAB104001	cgCACCGTGATGTTAACggttcagttca	SEQ ID NO:18
24794	pDAB104002	taAGGGACGAGGAGGAAggagtggaaga	SEQ ID NO:19

24795	pDAB104002	ttCTCCTGGAAGTACAGtcatcgacgcc	SEQ ID NO:20
24810	pDAB104003	gtCGCTGAAGGcGTGGTGgccgcactcg	SEQ ID NO:21
24811	pDAB104003	caGTGGCTgGACGACACCgtcggcctca	SEQ ID NO:22
24814	pDAB104004	gaGAAGTAAGGGACGAGgaggaaggagt	SEQ ID NO:23
24815	pDAB104004	gaAGTACAGTCATCGACgccaccattcc	SEQ ID NO:24
24818	pDAB104005	tcCCAAAGGGTtGTTGAGgtacttgccg	SEQ ID NO:25
24819	pDAB104005	acCGTGATGTTAACGGTtcagttcactc	SEQ ID NO:26
24796	pDAB104006	gaGAAGTAAGGGACGAGgaggaaggagt	SEQ ID NO:23
24797	pDAB104006	tgGAAGTAcAGTCATCGAcgccaccatt	SEQ ID NO:28
24836	pDAB104007	gtAGAGACcGTAGCAGACggcgaggatg	SEQ ID NO:29
24837	pDAB104007	gcTACGCTGCTgTCCAAGgagttgcctc	SEQ ID NO:30
24844	pDAB104008	gaGGCCAGGCGAAGTAGGAGagagggtg	SEQ ID NO:31
24845	pDAB104008	acTGGGCCTGCCAGGGCtgcgtcctaac	SEQ ID NO:32
24820	pDAB104009	gaGAGGCCaGGCGAAGTAggagagaggg	SEQ ID NO:33
24821	pDAB104009	acTGGGCCTGCCAGGGCtgcgtcctaac	SEQ ID NO:32
24828	pDAB104010	agGCCCAGtAGAGAGGCCaggcgaagta	SEQ ID NO:35
24829	pDAB104010	ccAGGGCTGCGTCCTAACCGgcgtctgg	SEQ ID NO:36
24832	pDAB104011	taGTCGCTGAAGGCGTGGTGgccgcact	SEQ ID NO:37
24833	pDAB104011	agTGGCTGGACGACaCCGTCGgcctcat	SEQ ID NO:38

EXAMPLE 3: EVALUATION OF ZINC FINGER NUCLEASE CLEAVAGE OF FAD2 GENES

CONSTRUCT ASSEMBLY

[0222] Plasmid vectors containing ZFN expression constructs of the exemplary zinc finger nucleases, 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 zinc finger-encoding sequence was fused to a sequence encoding an opaque-2 nuclear localization signal (Maddaloni *et al.* (1989) *Nuc. Acids Res.* 17(18):7532), that was positioned upstream of the zinc finger nuclease.

[0223] Next, the opaque-2 nuclear localization signal::zinc finger nuclease fusion sequence was paired with the complementary opaque-2 nuclear localization signal::zinc finger nuclease fusion sequence. As such, each construct included a single open reading frame comprised of two opaque-2 nuclear localization signal::zinc finger nuclease fusion sequences separated by the 2A sequence from *Thosea asigna* virus (Mattion *et al.* (1996) *J. Virol.* 70:8124-8127). Expression of the fusion proteins was driven by a relatively strong constitutive promoter such as a promoter derived from the Cassava Vein Mosaic Virus (CsVMV) promoter and flanked by the *Agrobacterium tumefaciens* ORF23 3'UnTranslated Region (AtuORF23 3'UTR).

[0224] The vectors were assembled using the IN-FUSION™ Advantage Technology (Clontech, Mountain View, CA). Restriction endonucleases were obtained from New England BioLabs (NEB; Ipswich, MA) and T4 DNA Ligase (Invitrogen) was used for DNA ligation. Plasmid preparations were performed using NUCLEOSPIN® Plasmid Kit (Macherey-Nagel Inc., Bethlehem, PA) or the Plasmid Midi Kit (Qiagen) following the instructions of the suppliers. DNA fragments were isolated using QIAquick Gel Extraction Kit™ (Qiagen) after agarose Tris-acetate gel electrophoresis. Colonies of all assembled plasmids were initially screened by restriction digestion of miniprep DNA. Plasmid DNA of selected clones was sequenced by a commercial sequencing vendor (Eurofins MWG Operon, Huntsville, AL). Sequence data were assembled and analyzed using the SEQUENCHER™ software (Gene Codes Corp., Ann Arbor, MI). Before delivery to *B. napus* protoplasts, Plasmid DNA 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.

[0225] The resulting eleven plasmid constructs; pDAB104008 (containing the ZFN24845 and ZFN24844 construct), pDAB104009 (containing the ZFN24820 and ZFN24821 construct), pDAB104010 (containing the ZFN24828 and ZFN24829 construct) (Figure 3), pDAB104003 (containing the ZFN24810 and ZFN24811 construct), pDAB104011 (containing the ZFN24832 and ZFN24833 construct), pDAB104002 (containing the ZFN24794 and ZFN24795 construct), pDAB104006 (containing the ZFN24796 and ZFN24797 construct), pDAB104004 (containing the ZFN24814 and ZFN24815 construct), pDAB104001 (containing the ZFN24800 and ZFN24801 construct), pDAB104005 (containing the ZFN24818 and ZFN24819 construct), and pDAB104007 (containing the ZFN24836 and ZFN24837 construct) were confirmed via restriction enzyme digestion and via DNA sequencing. Table 6 lists the different constructs and the specific FAD2A sequence which each ZFN was designed to cleave and bind.

Table 6: lists the Zinc Finger protein binding motif and the corresponding construct number. Each Zinc Finger was designed to bind and cleave the FAD2A which is described in the table

ZFN Design	**Construct No.**	**Locus ID.**	**Target Cut Site in FAD2A Sequence**
24844-2A-24845	pDAB104008	FAD2_ZFN_Locus1_F2A	263-265
24820-2A-24821	pDAB104009	FAD2_ZFN_Locus1_F2B	265
24828-2A-24829	pDAB104010	FAD2_ZFN_Locus1_F2C	275
24810-2A-24811	pDAB104003	FAD2_ZFN_Locus2_F1D	343-345
24832-2A-24833	pDAB104011	FAD2_ZFN_Locus2_F1E	345-346
24794-2A-24795	pDAB104002	FAD2_ZFN_Locus3_F2F	402
24796-2A-24797	pDAB104006	FAD2_ZFN_Locus3_F2G	408
24814-2A-	pDAB104004	FAD2_ZFN_Locus3_F2H	408-410

24815			
24800-2A-24801	pDAB104001	FAD2_ZFN_Locus4_F1J	531
24818-2A-24819	pDAB104005	FAD2_ZFN_Locus4_F1K	532-534
24836-2A-24837	pDAB104007	FAD2_ZFN_Locus5_F1L	724

PREPARATION OF DNA FOR TRANSFECTION

[0226] Plasmid DNA of the above described vectors was sterilized by precipitation and washing in 100% (v/v) ethanol and dried in a laminar flow hood. The DNA pellet was suspended in 30 µl of sterile double-distilled water at a final concentration of 0.7 µg/µl for transfection into protoplast cells as described below. The preparation of the plasmid DNA was undertaken to result in supercoiled plasmid DNA for transient transfection and linearized plasmid DNA for stable transfection. The addition of carrier DNA (e.g. fish-sperm DNA) to the transforming plasmid was not required for the transient transfection of protoplast cells. For transient studies about 30 µg of plasmid DNA per 10^6 protoplasts was used per transformation.

TRANSFECTION

[0227] Transfection of *Brassica napus* L. var. DH10275 was completed as described in Spangenberg *et al.*, (1986) Plant Physiology 66: 1-8, the media formulations are described in Spangenberg G. and Protrykus I. (1995) Polyethylene Glycol-Mediated Direct Gene Transfer in Tobacco Protoplasts. In: *Gene Transfer to Plants*. (Protrykus I. and Spangenberg G. Eds.) Springer-Verlag, Berlin. *Brassica napus* seeds were surface sterilized in 70% ethanol. The seeds were immersed in 12 mL of the 70% ethanol solution and mixed by gently rocking the cocktail for 10 minutes. The 70% ethanol solution was removed by decanting the solution and exchanged with a seed sterilization solution consisting of 1% w/v calcium hypochlorite and 0.1% v/v Tween-20. The seeds were immersed in the seed sterilization solution and mixed by gently rocking the cocktail for 25 minutes. The seed sterilization solution was decanted and the sterilized seeds were rinsed three times in 50 mL of sterile water. Finally, the seeds were transferred to a sterile 80 mm Whatman filter paper disc® (Fisher-Scientific, St. Louis, MO) that had been laid within a Petri dish and the seeds

were lightly saturated with sterile water. The Petri dish was sealed with Parafilm® (Fisher-Scientific, St. Louis, MO) and the plates were incubated at 25°C under complete darkness for one to two days. After signs of seedling emergence were observed from the seeds, the seedlings were transferred to Petri dish containing solidified GEM medium to encourage further seed germination. The seedlings were incubated on the GEM medium at 25°C for four to five days.

[0228] A volume of liquid PS medium (about 10 mL) was decanted into a sterile Petri dish. Using sterile forceps and a scalpel, an aerial portion of the four to five day old seedling in the 4-leaf stage of growth and development, was removed and discarded. Hypocotyl segments in lengths of 20-40 mm were determined to produce the highest population of small, cytoplasmic-rich protoplasts. The hypocotyl segments were aseptically excised and transferred to liquid PS medium. The excised hypocotyl segments were grouped together and cut transversely into 5-10 mm segments. Next, the hypocotyl segments were transferred to fresh PS medium and incubated at room temperature for 1 hour. The plasmolysed hypocotyls were transferred to a Petri dish containing enzyme solution. Care was taken to immerse all of the hypocotyl segments into the solution. The Petri dishes were sealed with Parafilm® and incubated overnight for sixteen to eighteen hours at 20 - 22°C with gentle rocking.

[0229] Protoplast cells were released from the hypocotyl segments. The overnight hypocotyl digests were gently agitated to release protoplasts into the enzyme solution. The Petri dish was angled slightly to aid the transfer of the digesting suspension of enzyme solution and plant debris. Using a 10 mL pipette the digesting suspension was transferred to a sterilized protoplast filtration (a filter of 100 micron mesh) unit to further separate the protoplasts from the plant debris. The filtration unit was tapped gently to release the excess liquid that had been caught in the sieve. The protoplast suspension, about 8 to 9 mL, was gently mixed and distributed into 14 mL sterile plastic round-bottomed centrifuge tubes. Each suspension was overlaid with 1.5 mL of W5 solution. The W5 solution was carefully dispensed over the protoplast suspension at an angle and dispensed drop-by-drop with minimal agitation. The addition of the W5 solution to the protoplast suspension resulted in the production of a protoplast rich interface. This interface was collected using a pipette. Next, the collected protoplasts were transferred into a new 14 mL centrifuge tube, and gently mixed. The yield or obtained protoplasts were determined using a haemocytometer to determine the number of protoplasts per milliliter. The method was repeated, wherein leaf tissue was digested to produce mesophyll protoplasts.

[0230] Next, W5 solution was added to a volume of 10 mL and the protoplasts were pelleted at 70g, before removing the W5 solution. The remaining protoplast suspension was resuspended by gentle shaking. Each tube containing the protoplast suspension was filled with 5 mL of W5 solution and incubated at room temperature from one to four hours. The protoplast suspensions were pelleted at 70g, and all of the W5 solution was removed. Next, 300 µL of transformation buffer was added to each of the pelleted protoplast suspensions which contained the isolated protoplasts. To each of the tubes, 10 µg of plasmid DNA was added to the protoplast suspensions. The plasmid DNA included the zinc finger nuclease constructs described above (*e.g.*, pDAB104010). Next, 300 µL of pre-warmed PEG 4000 solution was added to the protoplast suspension and the tubes were gently tapped. The protoplast suspensions and transformation mixture was allowed to incubate at room temperature for fifteen minutes without any agitation. An additional 10 mL of W5 solution was added to each tube in sequential aliquots of 1 mL, 1 mL, 1 mL, 2 mL, 2 mL, and 3 mL with gentle inversion of the tubes between each addition of W5 solution. The protoplasts were pelleted by spinning in a centrifuge at 70g. All of the W5 solution was removed leaving a pure protoplast suspension.

[0231] Next, 0.5 mL of K3 medium was added to the pelleted protoplast cells and the cells were resuspended. The resuspended protoplast cells were placed in the center of a Petri dish and 5 mL of K3 and 0.6 mL Sea Plaque™ agarose (Cambrex, East Rutherford, NJ) in a 1:1 concentration. The Petri dishes were shaken in a single gentle swirling motion and left to incubate for 20-30 minutes at room temperature. The Petri dishes were sealed with Parafilm® and the protoplasts were cultured for twenty-four hours in complete darkness. After the incubation in darkness, the Petri dishes were cultured for six days in dim light (5 µMol m^{-2} s^{-1} of Osram L36 W/21 Lumilux white tubes). After the culture step, a sterile spatula was used to divide the agarose containing the protoplasts into quadrants. The separated quadrants were placed into a 250 mL plastic culture vessel containing 20 mL of A medium and incubated on a rotary shaker at 80 rpm and 1.25 cm throw at 24 °C in continuous dim light for 14 days and then analyzed to determine the level of activity of each Zinc Finger Nuclease construct.

GENOMIC DNA ISOLATION FROM CANOLA PROTOPLASTS

[0232] Transfected protoplasts were supplied in individual 1.5 or 2.0 mL microfuge tubes. The cells were pelleted at the base of the tube in a buffer solution. DNA extraction was carried out by snap freezing the cells in liquid nitrogen followed by freeze drying the

cells, for about 48 hours in a Labconco Freezone 4.5® (Labconco, Kansas City, MO) at -40 °C and about 133 x 10^{-3} mBar pressure. The lyophilized cells were subjected to DNA extraction using the DNeasy® (QIAGEN, Carlsbad, CA) plant kit following manufactures instructions, with the exception that tissue disruption was not required and the protoplast cells were added directly to the lysis buffer.

TESTING OF FAD2A ZFN's FOR GENOMIC DNA SEQUENCE CLEAVAGE IN CANOLA PROTOPLASTS

[0233] The design of the ZFN target sites in the FAD2A gene locus were clustered, so that multiple pairs of ZFN were designed to overlapping target sites. The clustering of ZFN target sites enabled PCR primers to be designed that would amplify the surrounding genomic sequence from all FAD2A gene family members within a 100 bp window as to encompass all of the overlapping ZFN target sites. As such, the Illumina short read sequence technology could be used to assess the integrity of the target ZFN site of the transfected protoplasts. In addition, the PCR primers designs are needed to include specific nucleotide bases that would attribute sequence reads to the specific gene member of the FAD2A family. Therefore, all of the PCR primers would be required to bind 5-10 nucleotides away from any ZFN target cut site as non-homologous end joining (NHEJ) activity is known to cause small deletions that could remove a priming site, inhibit amplification and therefore distort the assessment of NHEJ activity.

[0234] Primers were designed to bind to all of the ZFN target loci for the FAD2A gene families (Table 7) and were empirically tested for amplification of all gene family members through Sanger based sequencing of PCR amplification products. In several instances primers could not be developed that would distinguish all gene family members (Table 8), however in all instances the target gene sequences of FAD2A, could be distinguished. Following PCR primer design custom DNA barcode sequences were incorporated into the PCR primers that were used to distinguish the different ZFN target loci and identify specific sequence reads to a transfection and ZFN (Tables 7 and 8).

Table 7: Primer sequences designed for FAD2 ZFN activity assessment of activity. Primers include custom barcodes, along with both requisite Illumina adaptor sequences for construction of Illumina library for sequencing-by-synthesis analysis. Purchased primer was the sum of all three columns presented.

Locus ID	**SEQ ID NO:**	**Illumina Adaptor Primer Sequence Barcode *Locus Primer***
FAD2_ZFN_Locus1_F2A	SEQ ID NO:39	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACGTA*CCCTCTCYCYT ACYTCGCC*
FAD2_ZFN_Locus1_F2B	SEQ ID NO:39	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGTAC*CCCTCTCYCYT ACYTCGCC*
FAD2_ZFN_Locus1_F2C	SEQ ID NO:39	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTACG*CCCTCTCYCYT ACYTCGCC*
FAD2_ZFN_Locus2_F1D	SEQ ID NO:39	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTACGT*GTCATAGCCCA CGAGTGCGGC*
FAD2_ZFN_Locus2_F1E	SEQ ID NO:39	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTGAC*GTCATAGCCCA CGAGTGCGGC*
FAD2_ZFN_Locus3_F2F	SEQ ID NO:39	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGACT*GTCGGCCTCAT CTTCCACTCC*
FAD2_ZFN_Locus3_F2G	SEQ ID NO:39	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGACTG*GTCGGCCTCA TCTTCCACTCC*
FAD2_ZFN_Locus3_F2H	SEQ ID NO:39	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACTGA*GTCGGCCTCA TCTTCCACTCC*
FAD2_ZFN_Locus4_F1J	SEQ ID NO:39	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCTAG*CAGACATCAAG TGGTACGGC*
FAD2_ZFN_Locus4_F1K	SEQ ID NO:39	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTAGC*CAGACATCAAG TGGTACGGC*
FAD2_ZFN_Locus5_F1L	SEQ ID	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAGCT*ATCTCCGACGC TGGCATCCTC*

	NO:39	
FAD2_ZFN_Locus1_R1A	SEQ ID NO:40	CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTACGTA*CTGGTA* *GTCGCTGAAGGCGT*
FAD2_ZFN_Locus1_R1B	SEQ ID NO:40	CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTCGTAC*CTGGTA* *GTCGCTGAAGGCGT*
FAD2_ZFN_Locus1_R1C	SEQ ID NO:40	CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTGTACG*CTGGTA* *GTCGCTGAAGGCGT*
FAD2_ZFN_Locus2_R1D	SEQ ID NO:40	CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTTACGT*GGACGA* *GGAGGAAGGAGTGGA*
FAD2_ZFN_Locus2_R1E	SEQ ID NO:40	CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTCTGAC*GGACGA* *GGAGGAAGGAGTGGA*
FAD2_ZFN_Locus3_R1F	SEQ ID NO:40	CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTTGACT*AGTGTT* *GGAATGGTGGCGTCG*
FAD2_ZFN_Locus3_R1G	SEQ ID NO:40	CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTGACTG*AGTGTT* *GGAATGGTGGCGTCG*
FAD2_ZFN_Locus3_R1H	SEQ ID NO:40	CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTACTGA*AGTGTT* *GGAATGGTGGCGTCG*
FAD2_ZFN_Locus4_R1J	SEQ ID NO:40	CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTGCTAG*CCCGAG* *ACGTTGAAGGCTAAG*
FAD2_ZFN_Locus4_R1K	SEQ ID NO:40	CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTCTAG*CCCGAG* *ACGTTGAAGGCTAAG*
FAD2_ZFN_Locus5_R1L	SEQ ID NO:40	CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTTAGCT*GAAGGA* *TGCGTGTGCTGCAAG*

[0235] Amplification performance of the designed PCR primers on the FAD2 gene families are shown in Table 8. An "X" indicates gene copy detection specificity, grey shading and an "+" indicates that at the specific locus in question the sequence reads designed by the two primers were unable to be distinguished.

Table 8: Results of cleavage at the FAD2A and FAD2C locus

ZFN Locus	**FAD Gene Copy**			
	FAD2A	FAD2C	FAD2A'	FAD2C'
Locus 1	X	X	X	X
Locus 2	X	X	X	X
Locus 3			X	X
Locus 4	X	X	X	X
Locus 5	X	X	X	X

[0236] Following DNA extraction of canola protoplasts transfected with the ZFN(s), PCR amplification of the target ZFN loci was performed to generate the requisite loci specific DNA molecules in the correct format for Illumina based sequencing by synthesis technology. Each assay was optimised to work on 25 ng starting DNA (about 12,500 cell equivalents of the *Brassica napus* genome). Multiple reactions were performed, per sample to provide the coverage required to assess NHEJ efficiency and specificity at the appropriate level, about sixteen PCR reactions equivalent to 200,000 copies of the *Brassica napus* genome taken from individual protoplasts. PCR amplification master-mixes were made for all samples to be tested with the same assay and one reaction, performed in triplicate, was assayed using a quantitative PCR method that was used to determine the optimal number of cycles to perform on the target tissue, to ensure that PCR amplification had not become reagent limited and was still in an exponential amplification stage. The experimentation with the necessary negative control reactions, was performed in 96 well format using a MX3000P thermocycler® (Stratagene, LaJolla, CA). From the output gathered from the quantitative PCR platform, 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 the amplification of common transcripts or molecules. The unused master mix, remained on ice until the quantitative PCR analysis was concluded and the cycle number determined and was then aliquoted into the desired number of reaction tubes (about 16 per ZFN assay) and the PCR reaction was performed.

[0237] Following amplification, samples for a single ZFN locus were pooled together and 200 μL of pooled product per ZFN was cleaned using the MinElute PCR purification kit® (Qiagen) following manufacturer's instructions. To enable the sample to be sequenced using the Illumina short read technology additional paired end primers were required to be attached by amplification onto the generated fragments. This was achieved by PCR amplification using primers that would be, in part complementary to the sequence added in the first round of amplification, but also contain the paired end sequence required. The optimal number of PCR cycles to perform, that would add the paired end sequences without over amplifying common fragments to the template was again determined using a sample pass through a quantitative PCR cycle analysis, as described previously.

[0238] Following PCR amplification, the generated product was cleaned using a MinElute column® (Qiagen) following manufacturer's instructions and was resolved on a 2.5% agarose gel. DNA fragments visualised using Syber® Safe (Life Technologies, Carlsbad, CA) as bands of the correct size were gel extracted to remove any residual PCR generated primer-dimer or other spurious fragments, the DNA was extracted from the gel slice using a MinElute gel extraction kit® (Qiagen) following manufacturer's instructions. After completion of the gel extraction an additional clean up of the DNA was performed using AMPure magnetic beads® (Beckman-Coulter, Brea, CA) with a DNA to bead ratio of 1:1.7. The DNA was then assessed for concentration using a quantitative PCR based library quantification kit for Illumina sequencing (KAPA) with a 1/40,000 and a 1/80,000 dilution and with the reaction being performed in triplicate. Based on the quantitative PCR results the DNA was diluted to a standard concentration of 2 nM and all libraries were combined for DNA sequencing. The samples were prepared for sequencing using a cBot cluster generation kit® (Illumina, San Diego, CA) and were sequenced on an Illumina GA2x® with 100 bp paired-end sequencing reads following manufacturer's instructions.

METHOD OF DATA ANALYSIS FOR DETECTION OF NON-HOMOLOGOUS END JOINING AT TARGET ZINC FINGER SITES

[0239] Following completion of the sequencing reaction and primary data calling performed using the Illumina bioinformatic pipeline for base calling, full analysis was performed to identify deleted bases at the target ZFN site in each instance. A custom PERL script was designed to extract and sort barcodes from DNA sequences computationally following a list of input sequences. The barcode had to match the reference sequence at a Phred score of greater than 30 to be accepted, to reduce misattributing sequence reads. After

the sequence reads had been binned into the different barcode groups that had been used, a quality filter was passed across all sequences. The quality filter was a second custom developed PERL script. Sequence reads were excluded if there were more than three bases called as "N", or if the median Phred score was less than 20, or if there were 3 consecutive bases with a Phred score of less than 20, or if the sequence read was shorter than 40 bp in length. The remaining sequences were merged where both of the paired sequence reads were available using the NextGENe® (SoftGenetics, State College, PA) package. The remaining merged sequence reads were then reduced to a collection of unique sequence reads using a third custom PERL script with a count of the number of redundant sequences that had been identified recorded on the end of the remaining sequence identifier. The unique sequence reads were then aligned to the FAD2 reference sequence using the NextGENe® software that created a gapped FASTA aligned file.

[0240] Using the gapped FASTA file a conversion of the gapped base position number to the input reference was performed using a fourth custom PERL script. This enabled bases that discriminate the different gene family members (either homoeologous or paralogous sequence variation between the different gene family members) to be identified in the assembled data. Once the conversion of base numbering had been performed it was possible to generate haplotype reports for each unique sequence reads and assign the reads to specific gene family members. Once the reads had been grouped by gene a 10 bp window was identified and assessed that surrounded the ZFN target site. The number of sequences with deletions was recorded per gene along with the number of missing bases.

[0241] The data was then graphically displayed as a multiple line graph, with the number of sequences with 1 through 10 bases deleted at the target ZFN site per 10,000 sequence reads (Figure 4). This analysis was performed for all ZFN transfections along with control transfections. In several instances, repeats in the native DNA sequence lead to an increase in sequencing error in the target ZFN site, such an error can be commonly seen as an increase in the prevalence of single base deletions that were reported in all samples, both transfected with ZFN or controls (Figure 5).

[0242] From these results highest level of ZFN activity at a FAD2 target site, as determined by the greater activity of NHEJ, was identified at locus E. The ZFNs which were encoded on plasmid pDAB104010 (*i.e.*, ZFN24828 and 24829) were selected for *in planta* targeting of an Engineered Transgene Integration Platform (ETIP) given its characteristics of significant genomic DNA cleavage activity and minimal non-target activity.

EXAMPLE 4: DNA CONSTRUCTS FOR ENGINEERED TRANSGENE INTEGRATION PLATFORM (ETIP) CANOLA PLANT LINES

[0243] The plasmid vector constructs described below were built using methods and techniques commonly known by one with skill in the art. The application of specific reagents and techniques described within this paragraph are readily known by those with skill in the art, and could be readily interchanged with other reagents and techniques to achieve the desired purpose of building plasmid vector constructs. The restriction endonucleases were obtained from New England BioLabs (NEB; Ipswich, MA). Ligations were completed with T4 DNA Ligase (Invitrogen, Carlsbad, CA). Gateway reactions were performed using GATEWAY® LR CLONASE® enzyme mix (Invitrogen) for assembling one entry vector into a single destination vector. IN-FUSION™ reactions were performed using IN-FUSION™ Advantage Technology (Clontech, Mountain View, CA) for assembling one entry vector into a single destination vector 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 all assembled plasmids were initially screened by restriction digestion of miniprep DNA. Plasmid DNA of selected clones was sequenced by a commercial sequencing vendor (Eurofins MWG Operon, Huntsville, AL). Sequence data were assembled and analyzed using the SEQUENCHER™ software (Gene Codes Corp., Ann Arbor, MI).

DIRECT-DELIVERY VECTORS FOR PRECISION INTEGRATION OF ETIP IN THE FAD2A LOCUS OF CANOLA

[0244] Standard cloning methods were used in the construction of the ETIP-containing vectors pDAS000130 (Figure 6, T-strand insert as SEQ ID NO:61), for specific integration into the FAD2A gene of *B. napus*. This construct has been designed to be delivered into canola protoplasts with the Zinc Finger Nuclease construct pDAB104010. The Zinc Finger Nuclease Construct will cleave the FAD2A locus and then the pDAS000130 construct will integrate within the canola genome via a homology directed or non homologous end joining repair mechanism. The ETIP includes four expression cassettes (two incomplete) separated by additional ZFN recognition sequences and an Engineered Landing Pad (ELP) containing another ZFN recognition sequences. The additional ZFN recognition sequences are unique and have been designed to be targeted for the introduction

of polynucleotide sequences within the ETIP and ELP transgene insertions. Similarly, the ZFN recognition sequences can be utilized for excision of polynucleotide sequences. The first gene expression cassette was an incomplete dsRED expression cassette and contained the promoter, 5' untranslated region and intron from the *Arabidopsis thaliana* Polyubiquitin 10 (AtUbi promoter) gene (Callis, *et al.*, (1990) *J. Biol. Chem.*, 265: 12486-12493) followed by 210 bp of a dsRed gene from the reef coral *Discosoma* sp. (Clontech, Mountain View, CA) codon-optimised for expression in dicot plants (ds RED (dicot optimized)exon 1) followed by an intron from the *Arabidopsis thaliana* thioreductase-like gene (Intron 1 from At thioreductase: Accession No: NC_00374) and the 3' untranslated region comprising the transcriptional terminator and polyadenylation site of the Zea mays Viviparous-1 (Vp1) gene (Zmlip terminator: Paek *et al.*, (1998) Molecules and Cells , 8(3): 336-342). The second expression cassette contained the 19S promoter including 5' UTR from cauliflower mosaic virus (CaMV 19S: Cook and Penon (1990) *Plant Molecular Biology* 14(3): 391-405) followed by the *hph* gene from *E. coli*, codon-optimised for expression in dicots (hph(HygR): Kaster *et al.*, (1983) Nucleic Acids Research 11(19): 6895-6911) and the 3'UTR comprising the transcriptional terminator and polyadenylation site of open reading frame 1 of *A. tumefaciens* pTi15955 (At-ORF1 terminator: Barker *et al.*, (1983) Plant Molecular Biology 2(6): 335-50). The third expression cassette was an incomplete PAT expression cassette and contained the first intron from Arabidopsis 4-coumaryl-CoA synthase (intron#2 4-coumaryl-CoA synthase v: Accession No: At3g21320/NC003074) followed by the last 256 bp of a synthetic, plant-optimized version of phosphinothricin acetyl transferase gene, isolated from *Streptomyces viridochromogenes*, which encodes a protein that confers resistance to inhibitors of glutamine synthetase comprising phosphinothricin, glufosinate, and bialaphos (PAT(v6) 3'end: Wohlleben *et al.*, (1988) Gene 70(1): 25-37). This cassette was terminated with the 3' UTR comprising the transcriptional terminator and polyadenylation sites of open reading frame 23 of *A. tumefaciens* pTi15955 (AtuORF23 terminator: Barker *et al.*, (1983) Plant Molecular Biology 2(6): 335-50). The fourth Expression Cassette was the *ipt* gene cassette and contained a 588 bp truncated version of the promoter and 5' UTR from the Arabidopsis DNA-binding protein MYB32 gene (U26933) (AtMYB32(T) promoter: Li *et al.*, (1999) Plant Physiology 121: 313) followed by the isopentyl transferase (*ipt*) gene from *A. tumefaciens* and the 35s terminator comprising the transcriptional terminator and polyadenylation sites from cauliflower mosaic virus (CaMV 35S terminator: Chenault *et al.*, (1993) Plant Physiology 101 (4): 1395-1396). For delivery to FAD2A, each end of the ETIP

sequence was flanked by 1kb of FAD2A genomic sequence from either side of the location of the double-stranded break induced by delivery of the ZFN encoded in pDAB104010 to the FAD2A gene of *B. napus*.

[0245] The ETIP sequence was synthesized by a commercial gene synthesis vendor (GeneArt, Life Technologies). The 1kb segments of FAD2A genome sequence were amplified from genomic DNA purified from leaf tissue of *B. napus* DH12075 using a Qiagen DNeasy plant mini kit® (Qiagen, Hilden) following instructions supplied by the manufacturer. The 1 kb FAD2A sequences were ligated into the ETIP vector using T4 ligase (NEB, Ipswich, MA). 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) 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 Corp., Ann Arbor, MI).

CONTROL VECTORS

[0246] A control vector was used to develop a Fluorescence Activated Cell Sorting (FACS) cell based sorting method. Standard cloning methods were used in the construction of a control vector, pDAS000031 (Figure 7: T-strand insert as SEQ ID NO:62) including two gene expression cassettes. The first gene expression cassette contained the Cauliflower mosaic virus 19s promoter (CaMV 19S promoter; Shillito, *et al.,* (1985) *Bio/Technology* 3; 1099-1103):: hygromycin resistance gene (hph(HygR);US Patent No. 4,727,028) :: and the *Agrobacterium tumefaciens* Open Reading Frame 1 3'UnTranslated Region (AtORF1 terminator; Huang *et al.*, (1990) *J. Bacteriol.* 1990 *172*:1814-1822). The second gene expression cassette contained the *Arabidopsis thaliana* Ubiquitin 10 promoter (AtUbi10 promoter; Callis, *et al.,* (1990) *J. Biol. Chem.*, 265: 12486-12493):: dsRED (dsRED(D); US Patent No. 6,852,849) and an intron from Arabidopsis (intron #1; GenBank: AB025639.1) :: *Agrobacterium tumefaciens* Open Reading Frame 23 3'UnTranslated Region (AtORF23 terminator; US Patent No. 5,428,147) as an in-frame fusion with a *trans* orientation (*e.g.,* head to head orientation). The plasmid vector was assembled using the IN-FUSION™ Advantage Technology (Clontech, Mountain View, CA).

EXAMPLE 5: GENERATION OF ETIP CANOLA PLANT LINES

TRANSFORMATION OF *BRASSICA NAPUS*

[0247] The FAD2A site specific integration deploys the use of the ETIP construct (pDAS000130), accompanying Zinc Finger Nuclease (pDAB104010), and the DS-Red control construct (pDAS000031) described in Example 4. The binary vectors were transformed into *Agrobacterium tumefaciens* strain GV3101: PM90. Transformation of *Brassica napus* protoplast cells was completed using the transfection protocol described in Example 3 with some modification.

[0248] The modifications to the protocol included the use of sodium alginate instead of Sea Plaque™ agarose. The transfection experiments in which both the Zinc Finger Nuclease construct, pDAB104010, and the ETIP construct, pDAS000130, were co-delivered into *Brassica napus* protoplast cells were completed at DNA concentrations comprising a 5:1 and a 12:1 molar ratio, wherein pDAS000130 had a concentration of 27.8 μg of plasmid DNA and pDAB104010 had a concentration of 2.2 μg of plasmid DNA. The control plasmid constructs were transformed at concentrations of 30 μg of plasmid DNA.

[0249] Additional modifications to the protocol included the propagation of whole plants from the transformed protoplast cells in medium containing 1.5 mg/mL of hygromycin. The propagation of whole plants required that the A medium was replaced every two weeks and the growth of the protoplast-derived colonies was monitored. After the protoplast-derived colonies had grown to approximately 2-3 mm in diameter, the colonies were transferred into individual wells of a 12-well Costar® plate (Fisher Scientific, St. Louis, MO) containing solidified MS morpho medium. The plates were incubated for one to two weeks at 24 °C under continuous dim light until the calli had proliferated to a size of 8-10 mm in diameter.

[0250] After the protoplast cells had reached a diameter of 1-2 cm, the protoplast cells were transferred to individual 250 mL culture vessels containing MS morpho medium. The vessels were incubated at 24 °C under 16 h light (20 μMol m^{-2} s^{-1} of Osram L36 W/21 Lumilux white tubes) and 8 h dark conditions. Within one to two weeks, multiple shoots were visible. The shoots were transferred into 250 mL culture vessels containing MS medium after they reached a length of 3-4 cm. The 250 mL culture vessels were incubated at 24°C under 16 h light (20 μMol m^{-2} s^{-1} of Osram L36 W/21 Lumilux white tubes) and 8h dark

conditions. The shoots were maintained in the culture vessels until they developed into plantlets at which time they were transferred to a greenhouse to grow to maturity.

EXAMPLE 6: MOLECULAR CONFIRMATION OF INTEGRATION OF T-DNAS CONTAINING ETIPS IN CANOLA

[0251] Genomic DNA was extracted from leaf tissue of all putative transgenic plants using a DNeasy Plant Mini Kit™ (Qiagen) following the manufacturer's instructions, with the exception that tissue was eluted in 80µl of AE buffer. Thirty milligrams of young leaf tissue from regenerated plants was snap frozen in liquid nitrogen before being ground to a powder.

[0252] Molecular characterization of the FAD2A locus was performed using three independent assays. Assays were designed and optimized using the following controls; characterized transgenic events comprising a single randomly integrated transgene, characterized transgenic event with five randomly integrated transgenes, wildtype canola c.v. DH12075 plants and non-template control reactions. The results from the three following molecular analyses are considered together in order to provide evidence for integration of the ETIP at FAD2A via HDR.

IDENTIFYING TRANSGENE INTEGRATION BY REAL-TIME POLYMERASE CHAIN REACTION

[0253] Four replicates of each plant were analyzed using primers specific to the *hph* (also described as *hpt*) target gene (SEQ ID NO:63, hpt F791 5' CTTACATGCTTAGGATCGGACTTG 3'; SEQ ID NO:64, hpt R909 5' AGTTCCAGCACCAGATCTAACG 3'; SEQ ID NO:65, hpt Taqman 872 5' CCCTGAGCCCAAGCAGCATCATCG 3' FAM) (**Figure 8**) and reference gene encoding High Mobility Group protein I/Y *(HMG I/Y)* (SEQ ID NO:66, F 5' CGGAGAGGGCGTGGAAGG 3'; SEQ ID NO:67, R 5' TTCGATTTGCTACAGCGTCAAC 3'; SEQ ID NO:68, Probe 5' AGGCACCATCGCAGGCTTCGCT 3' HEX). The reactions were amplified using the following conditions: 95°C for 10 minutes followed by 40 cycles of 95°C for 30 seconds, 60°C for 1 minute, with amplification data being captured at the end of each annealing step. Copy number was calculated using the ΔCq method, where ΔCq = Cq(target gene) – Cq(reference gene). Livak, K.J. and T.D. Schmittgen, *Analysis of relative gene expression*

data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods, 2001. 25(4): p. 402-8. Plants with amplification of *hph* and *HMG I/Y* and a copy number of 0.5 or more were considered transgenic, while plants with a copy number of ≥ 0.5 and ≤ 1.2 were scored as putatively single copy. Amplification was performed on a BioRad CFX96 Touch™ Real-Time PCR Detection System with FastStart Universal Probe Master (ROX), (Roche, Basel, Switzerland).

DETECTION OF DISRUPTED FAD2A ZFN SITE

[0254] Each plant was analysed for presence or absence of amplification of endogenous target in the disrupted locus test, which is a dominant assay. The assay is a SYBR® Green I qPCR assay and in singleplex, but with each reaction run simultaneously on the same PCR plate, targets an endogenous locus (FAD2A/2C.RB.UnE.F1, SEQ ID NO:69, 5' CTTCCACTCCTTCCTCCTCGT*C 3' and FAD2A/2C.RB.UnE.R1, 5' SEQ ID NO:70, GCGTCCCAAAGGGTTGTTGA*G 3') and the ZFN locus (locus at which the ZFN pDAB104010 binds and cuts the genome) (FAD2A.UnE.F1, SEQ ID NO:71, 5' TCTCTACTGGGCCTGCCAGGG*C 3' and FAD2A.UnE.R1, SEQ ID NO:72, 5' CCCCGAGACGTTGAAGGCTAAGTACAA*A 3') (**Figure 9**). Both primer pairs were amplified using the following conditions: 98°C for 30 seconds followed by 35 cycles of (98°C for 10 seconds, 65°C for 20 seconds, 72°C for 90 seconds) then followed by 95°C for 10 seconds then a melt analysis from 50°C to 95°C with 0.5°C increments for 0.05seconds and a plate read at each increment. The reaction conditions are listed in Table 9.

Table 9: Single reaction reagent components and concentrations for PCR amplification

Reaction Components	**Volume (µl)**
10 mM dNTP	0.40
5X Phusion HF Buffer	4.00
Phusion Hot Start II High-Fidelity DNA Polymerase (2U/µl) (Thermo Scientific)	0.25
Forward Primer 10µM	0.40
Reverse Primer 10µM	0.40
1:10000 dilution of SYBR Green I dye (Invitrogen)	1.00
Molecular Biology Grade H_2O	11.55
Genomic DNA template (~20ng/µl)	2.00
Total Volume	20.00

[0255] Plants that had amplification of the endogenous target but no amplification of the ZFN target, were scored as positive for the disrupted locus test and were considered to

have a disrupted ZFN locus. This assay was considered to be positive when the ZFN binding site on both alleles at the FAD2A locus have been disrupted.

PCR DETECTION OF TRANSGENE INTEGRATION AT FAD2A VIA HOMOLOGY DIRECTED REPAIR

[0256] Each putative plant transformant was analysed using endpoint with PCR primers designed to amplify the transgene target *hph (hph*_ExoDigPC_F1, SEQ ID NO:73, 5' TTGCGCTGACGGATTCTACAAGGA 3' and *hph*_ExoDigPC_R1, SEQ ID NO:74, 5' TCCATCAGTCCAAACAGCAGCAGA 3'), the FAD2A endogenous locus (FAD2A.Out.F1, SEQ ID NO:75, 5' CATAGCAGTCTCACGTCCTGGT*C 3' and FAD2A.Out.Rvs3, SEQ ID NO:76, 5' GGAAGCTAAGCCATTACACTGTTCA*G 3'), the region spanning the 5' end of any transgene integrated into the FAD2A locus via HDR, upstream of the transgene into the FAD2 A locus (FAD2A.Out.F1, SEQ ID NO:77, 5' CATAGCAGTCTCACGTCCTGGT*C 3' and QA520, SEQ ID NO:78, 5' CCTGATCCGTTGACCTGCAG 3') and the region spanning the 3' end of any transgene integrated into the FAD2A locus via HDR, downstream of the transgene into the FAD2 A locus (QA558, SEQ ID NO:79, 5' GTGTGAGGTGGCTAGGCATC 3' and FAD2A.Out.Rvs3, SEQ ID NO:80, 5' GGAAGCTAAGCCATTACACTGTTCA*G 3') **(Figure 3)**. All primer pairs were amplified using the following conditions 98°C for 30 seconds followed by 35 cycles of (98°C for 10 seconds, 65°C for 20 seconds, 72°C for 90 seconds). Reaction reagent conditions are as described in **Table 10.**

Table 10: Single reaction reagent components and concentrations for PCR amplification

Reaction Components	**Volume (µl)**
5x Phusion HF Buffer	6.00
10mM dNTPs	0.60
Forward Primer 10µM	0.60
Reverse Primer 10µM	0.60
Phusion Hot Start II High-Fidelity DNA Polymerase (2U/µl) (Thermo Scientific)	0.25
Molecular Biology Grade H_2O	19.95
Genomic DNA template (~20ng/µl)	2.00
Total Volume	30.0

[0257] Amplification of the 5' transgene-genome flanking target and/or amplification of the 3' transgene-genome flanking target indicated a putative insertion event. It must be noted that due to the approximately 1,000bp FAD2A homology arms in the pDAS000130

cassette (comprising polynucleotide sequences with 100% sequence identity to the FAD2A regions immediately upstream and downstream of the ZFN cut site), the PCR reactions were subject to false positive PCR product amplification due to PCR chimerism arising from amplification of off-target ETIP integration events. Amplification of the *hph* target confirmed transgene integration had occurred. Amplification of the FAD2A target suggests that the FAD2A locus is intact or contains only a partial insertion. Due to the size of the ETIP (11,462 bp for the ETIP cassettes or 13,472 bp including the FAD2A homologous arms and the ETIP cassettes) it is expected that the FAD2A primers would not amplify a product when an intact ETIP is integrated into the FAD2A locus.

SOUTHERN DETECTION OF FAD2A EDITING

[0258] Plants that had amplification of either a 5' genome-transgene flanking target product and/or amplification of a 3' transgene-genome flanking target, or no amplification of the ZFN locus target, or both, were subject to Southern analysis for detection of transgene integration at the FAD2A locus. Genomic DNA was purified from 5g of leaf tissue using a modified CTAB method (Maguire, T.L., G.G. Collins, and M. Sedgley *A modified CTAB DNA extraction procedure for plants belonging to the family proteaceae.* Plant Molecular Biology Reporter, 1994. 12(2): p. 106-109). Next, 12 µg of genomic DNA was digested with *Kpn1*-HF (New England BioLabs) and digestion fragments were separated by electrophoresis on a 0.8% agarose gel before transfer to membrane using a standard Southern blotting protocol. Primers to *FAD2A* 5' target region (F, SEQ ID NO:81, 5' AGAGAGGAGACAGAGAGAGAGT 3' and R, SEQ ID NO:82, 5' AGACAGCATCAAGATTTCACACA 3'), *FAD2A* 3' target region (F, SEQ ID NO:83, 5' CAACGGCGAGCGTAATCTTAG 3' and R, SEQ ID NO:84, 5' GTTCCCTGGAATTGCTGATAGG 3') and *hph* (F, SEQ ID NO:85, 5' TGTTGGTGGAAGAGGATACG 3' and R, SEQ ID NO:86, 5' ATCAGCAGCAGCGATAGC 3') were used to generate probes to detect the presence of the ETIP within the FAD2A locus using the DIG Easy Hyb System® (Roche, South San Francisco, CA) following the manufacturer's instructions (**Figure 11**). Hybridization was performed at 42° C for *FAD2A* 5' region, 45°C for *FAD2A* 3' region and 42° C for detection of *hph*.

[0259] Membrane-bound genomic DNA was probed in a specific order; firstly *FAD2A* 5'sequences were probed, then the *FAD2A* 3' sequences were probe, and finally the

hph sequences were probed (**Figure 12**). The rational for this is as follows. The first probe (FAD2A 5') is the diagnostic probe, and if the ETIP has integrated into FAD2A via perfect HDR, a 5,321 bp fragment will be visible on the membrane. The resulting band size is easily differentiated during electroporation and will sit close to the 5,148 bp fragments in the DIG labeled Roche DNA Molecular Weight Marker III® (Roche, Indianapolis, IN). The second probe of the membrane is with the FAD2A 3' probe and an edited plant will have a 22,433 bp fragment whereas an unedited plant will have a 16,468 bp fragment. The same 22,433 bp fragment identified with the FAD2A 3' probe should also be bound by and identified with the *hph* probe. These fragments are difficult to differentiate on a gel as they are extremely large and it may be difficult to determine any difference between a fragment occurring above or below the largest, 21,226 bp fragment in the DIG labeled Roche DNA Molecular Weight Marker III®. As such, these probes provide evidence that may strengthen the identification of ETIP integration into FAD2A via homology directed repair (HDR), by visualization of a 5 kb fragment using the FAD2A 5' probe. The restriction enzyme, KpnI was the only suitable restriction endonuclease for use in this assay, as *KpnI* sites occurred in a single locus of the cut the ETIP cassette in a single locus, and was present in two sites of the FAD2A ZFN locus. One site was located upstream and the second site located downstream of the FAD2A homology arms. In addition, *KpnI* is not methylation sensitive, and is available as a recombinant enzyme with increased fidelity (New England Biolabs).

Results of Molecular and Southern analysis

[0260] Following transfection, culturing, and selection the transgenic plants were transferred to soil. From this process, 139 plants survived and had tissue sampled for gDNA extraction and analysis. All 139 plants were analyzed for copy number estimation. Of these 139 plants, 56 were positive for the ETIP and 11 of the 56 positive plants had a putative single copy integration (**Figure 13**) (**Table 11**). Of the 56 plants that were positive for ETIP integration, amplification of the FAD2A 5'-genome-transgene flanking sequence occurred in 7 plants. Amplification of the FAD2A 3'-transgene-genome flanking sequence did not occur in any of the 56 plants that were positive for ETIP integration. Additionally, of the 56 plants that were positive for transgene integration, 11 plants were positive for the disrupted locus qPCR test. Fourteen plants that were positive for amplification of the FAD2A 5' genome-transgene flanking sequence and/or positive for the disrupted locus qPCR test were subject to Southern analysis, with the 3 probes described above. Of the 14 plants advanced for Southern analysis, all of the plants showed partial integration within the FAD2A locus, but

none of these plants showed evidence of a complete full-length integration of the ETIP at the FAD2A locus via HDR when probed with the FAD2A 5' probe, FAD2A 3' and *hph* probes. No bands that appeared to be i) larger than WT and ii) identical to bands observed for those samples when probed with FAD2A 3' probe (**Table 11**).

Table 11: Overview of outcomes from analysis of ETIP integration

No. of plants surviving in soil	139
No. of plants sampled	139
No. of plants for which qPCR copy number analysis was completed	139
No. of plants positive for ETIP integration	56
No. of plants comprising a putative single copy insert	11
No. of ETIP/ *FAD2* in-out 5' reactions	7 (from 56)
No. of ETIP/ *FAD2* in-out 3' reactions	0 (from 56)
No. of locus disrupted qPCR tests	9 (from 56)
ETIP on-target (Southern)	0 (from 14)

RESULTS OF ETIP TRANSGENIC CANOLA TRANSFORMED WITH PDAS000130 AND PDAB104010.

[0261] The transgenic *Brassica napus* events which are produced via transformation of pDAS000130 and pDAB104010 result in the integration of a single copy, full length T-strand insertion of the ETIP polynucleotide sequence from pDAS000130 within the FAD2A locus. Three to four events are fully characterized and confirmed to contain the integrated ETIP. The confirmation is completed using an in-out PCR amplification method, and further validated via Southern blot. The selected T_0 events are grown to the T_1 stage of development. The T_1 plants are re-screened to determine the zygosity of the integrated T-strand. Screened events are categorized as homozygous, hemizygous, or null.

[0262] The homozygous events are used to produce protoplasts via the previously described method. The protoplasts are subsequently co-transformed with at least one zinc finger nuclease that is designed to target a binding site which is incorporated within the ETIP sequence and a donor plasmid which shares homology with specific regions of the ETIP wherein the donor is integrated within the ETIP via an HDR mechanism. Likewise, the protoplasts are subsequently co-transformed with at least one zinc finger nuclease that is designed to target a binding site which is incorporated within the ETIP sequence and a donor plasmid which does not share homology with specific regions of the ETIP, wherein the donor is integrated within the ETIP via an non-homologous end joining mechanism. The ZFN(s) cleave(s) the ETIP locus and the donor plasmid is integrated within the genome of *Brassica napus* cells via homology directed repair or non-homologous end joining.

[0263] As a result of the integration of the donor plasmid, the partial *DS-red* transgene is repaired to a full length *DS-red* transgene. The expression of the now fully operational *DS-red* transgene is used to sort protoplast cells with a FACS method. Putative transgenic plants are sorted using the FACS method described in Example 7 and the isolated protoplasts are regenerated into mature plants. The integration of the donor plasmid is confirmed within the ETIP-targeted plants using molecular confirmation methods. As such, the ETIP locus serves as a site-specific locus for gene targeted integration of a donor polynucleotide sequence.

EXAMPLE 7: FACS BASED SORTING OF PROTOPLAST CELLS

[0264] *Brassica napus* protoplasts that were transfected with the DS-Red control construct, pDAS000031, were sorted via FACS-mediated cell sorting using a BD Biosciences Influx-Cell sorter™ (San Jose, CA). The protoplast cells were isolated and transfected as described in Example 3. After the cells had been transfected with pDAS000031, the cells were sorted using the FACS sorter with the conditions described in Table 12.

Table 12: Conditions used for sorting protoplast cells transfected with pDAS000031.

Parameters	
Drop frequency	6.1 KHz
Nozzle diameter	200µm
Sheath pressure	4 psi

Recovery media	W5 media
Culture conditions	Bead type culture using sea-plaque agarose and sodium alginate
Sort criteria	Sorting based on chlorophyll autofluorescence, reporter gene expression (Ds-Red)
Sort recovery (%)	50-75
Viability post sorting (%)	>95

[0265] The protoplasts which expressed the *DS-red* transgene were sorted and isolated. The FACS isolated protoplasts were counted using the sorter. About $1x10^5$ to $1.8x10^5$ of cells were placed in a well of a 24-well micro titer plate on the first day after the FACS isolation. The cells were transferred to a bead culture for 5 to 20 days. Similar conditions were tested, wherein about $1x10^4$ of cells were placed in a well of a 2 or 4-well micro titer plate on the second day after the FACS isolation. The various conditions that were tested resulted in the recovery of cells at a viability or 95 – 98% of the total isolated protoplast cells. The FACS sorted protoplast cells were transferred to a bead culture for 3 – 20 days. The FACS sorted protoplast cells were regenerated into plants on media which contained 1.5 mg/mL of hygromycin using the above described protocol. The putative transgenic plants were confirmed to contain an intact T-strand insert from pDAS000031 via molecular conformation protocols.

[0266] The FACS sorting method is directly applicable to screen any fluorescent transgene sequence and is used to isolate a proportion of *Brassica napus* protoplast cells that are targeted with a fluorescent transgene via homology mediated repair within a specific site in the ETIP region within a genomic locus.

EXAMPLE 8: TARGETED INTEGRATION AND DISRUPTION OF *Brassica napus* OMEGA-3 FATTY ACID DESATURASE (FAD2) VIA HOMOLOGY DIRECTED REPAIR

SELECTION OF ZINC FINGER BINDING DOMAINS SPECIFIC TO FAD2A

[0267] The transcribed regions for homoeologous FAD2 genes were identified and characterized, zinc finger nucleases that were designed to bind and cleave these sites for

NHEJ-mediated targeting of a donor sequence. Zinc finger proteins (ZFPs) directed against DNA sequences from homeologues of FAD2 sequences were designed and tested as described above. From the ZFNs showing on-target activity, one zinc finger proteins were selected that cut the FAD2 target at high efficiency: ZFP 24828-2A-24829 recognizes SEQ ID NO:35 5'- agGCCCAGtAGAGAGGCCaggcgaagta -3' and SEQ ID NO:36 5' - ccAGGGCTGCGTCCTAACCGgcgtctgg -3'. This ZFN was shown to specifically bind and cleave the FAD2A genomic locus.

DESIGN AND CONSTRUCTION OF "DONOR" VECTORS FOR HDR-DIRECTED DNA REPAIR

[0268] For integration of a donor sequence via HDR, a single vector was constructed. The vector encoded a *hygromycin* (*hph* or *hpt*) resistance gene expression cassette. The hygromycin resistance gene expression cassette included the 19S promoter including a 5' UTR from cauliflower mosaic virus (CaMV) (Cook and Penon Plant Molecular Biology 1990 14(3), 391-405) followed by the *hygromycin phosphotransferase (hph)* gene (Kaster et al Nucleic Acids Research 1983 11 (19), 6895-6911). The *hph* gene was codon-optimised for expression in dicotyledonous plants and was flanked by a 3'UTR comprising the transcriptional terminator and polyadenylation site of Open Reading Frame 1 (ORF1) of *A. tumefaciens* pTi15955 (Barker *et al,* Plant Molecular Biology 1983, 2(6), 335-50). The cassettes were synthesized by a commercial gene synthesis vendor (GeneArt, Life Technologies, Regensberg, Germany). Flanking FAD2A sequences were added upstream and downstream of the gene expression cassette. The hygromycin resistance cassette was cloned into specific restriction enzyme sites of each vector resulting in a "donor" vectors: pDAS000129 (hygromycin-resistant gene-splicing donor: SEQ ID NO:87 Figure 14).

[0269] Colonies of the assembled plasmids were initially screened by restriction endonuclease digestion of DNA purified from overnight cultures of *E.coli*. Restriction endonucleases were obtained from NEW ENGLAND BIOLABS™ (NEB, Ipswich, MA) and PROMEGA™ (Promega Corporation, WI). Plasmid preparations were performed using the QIAPREP SPIN MINIPREP KIT™ (Qiagen, Hilden, Germany) or the PURE YIELD PLASMID MAXIPREP SYSTEM™ (Promega Corporation, WI) following the instructions of the suppliers. After the restriction fragments were confirmed by agarose gel electrophoresis of resulting fragments, plasmid DNA of selected clones were sequenced using ABI Sanger Sequencing and BIG DYE TERMINATOR V3.1™ cycle sequencing

protocol (Applied Biosystems, Life Technologies). Sequence data were assembled and analyzed using the SEQUENCHER™ software (Gene Codes, Ann Arbor, MI).

TRANSFORMATION OF *BRASSICA NAPUS*

PEG-mediated protoplast transfection and regeneration of ETIP targeted to FAD2a and FAD2a ZFN (precision events)

[0270] Mesophyll derived protoplasts were isolated from three weeks old sterile shoot cultures of *Brassica napus* (DH10275). The corresponding seeds were germinated. The seeds were surface-sterilized using 70% ethanol for 1 minute by gentle shaking followed by 3-4 rinses in sterile double-distilled water and subsequently sterilized using 20% bleach and 10 µl of Tween 20™; the seeds were treated with the bleach on a shaker (table top rotary shaker approximately 100 RPM) for 15 minutes followed by 3-4 rinses in sterile double-distilled water, seeds were carefully transferred to a sterile filter paper to remove the excess moisture and plated on seed germination medium (½ strength MS/B5 Vitamins + 1% sucrose + 0.8% Agar; pH 5.8 and 50-60 ml of the media was poured per Petri dish (15 X 100 mm) that was placed with a slight angle using a support); approximately 50 seeds were placed in each plate. The plates were incubated upright at 22°C in 16h/d light (20 µmol m^{-2} s^{-1}) for 6 days. Hypocotyl segments of 0.5 cm size were dissected from the six day old seedlings and cultured on shoot induction medium (MS/B5 Vitamins + 3% sucrose + 500mg/L MES + BAP (13µm) + Zeatin (5µm) + Silver Nitrate (5mg/L) + 0.8% Agar (pH 5.8) and poured in 100 x 20 mm sterile Petri dish) approximately 20 explants were placed on each plate . Shoot meristems that appeared after 3-4 weeks were transferred to shoot elongation medium (MS/B5 Vitamins + 2% sucrose + 500mg/L MES + BAP (2µm) + GA-3 (0.1µm) + 0.8% Agar (pH 5.8) and poured in 250 ml culture vessels) and the cultures were maintained in this medium for 4 weeks with one round of sub-culturing in between. Shoots of 2-3 cm height were then transferred to root initiation media (1/2 strength MS/B5 Vitamins + 1% sucrose + 500mg/L MES + IBA (2.5µm) + 0.6% Agar (pH 5.8) and poured in 700 ml culture vessels) for root development. Rooted shoots were sub-cultured in fresh root initiation media at 3-4 weeks intervals as stem cuttings for two-three rounds before use. The cultures were maintained throughout at 22°C in 16h/d light (30 µmol m^{-2} s^{-1}).

Protoplast isolation and purification

[0271] *In vitro* grown DH12075 *Brassica napus* plants were used as the explant source for isolating mesophyll protoplasts. The 3rd to 4^{th} upper fully expanded leaves from 3

to 4 weeks old plantlets were cut into small strips (0.5 to 1 mm) with a sharp scalpel for protoplast isolation. Enzymatic digestion was carried out by treating 250-500 mg of leaf material with 25ml of digestion buffer (1.2% (w/v) Cellulase "Onozuka" R10® and 0.2% (w/v) Macerozyme® R10 dissolved in K4 media (Spangenberg et al. 1998)). The Petri dish containing the leaf material and digestion buffer was sealed with Parafilm™ and incubated at room temperature for 12 to 15 h in darkness. After overnight incubation the digests were filtered through a BD® cell strainer (mesh size 70μm). Protoplasts suspension (5-6 ml) were collected in a 14 ml round bottomed tube that was over layered with 1 ml of washing solution W5 buffer (154 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl and 5 mM glucose; pH 5.8 Mcnczel *et al.* 1981) and centrifuged at 400 RPM for 10 min. After centrifugation, protoplasts that floated in the interphase were withdrawn and washed by centrifugation using 10 ml of W5 buffer at 400 RPM for 10 min. After the final wash, isolated protoplasts were resuspended at a density of 1X 10^6 protoplasts per mL of W5 buffer and incubated for 1 hour before transfections.

Assessment of protoplasts yield and viability

[0272] Protoplasts yield was assessed using a haemocytometer following Sambrook and Russel, 2006 and the viability was tested using Evans blue stain (400mg/L dissolved in 0.5 M of Mannitol) following Huang et al. 1996 with few modifications.

PEG 4000 mediated stable DNA delivery

[0273] Plasmid DNA of the ETIP-containing vector pDAS000129 and the ZFN vector (pDAB104010) targeting the FAD2 A locus was isolated from cultures of *E. coli* using the Pure Yield Plasmid Maxiprep System™ (Promega Corporation, WI) or Plasmid Maxi Kit™ (Qiagen, Hilden) following the instructions of the suppliers. The plasmid DNA was dissolved at a density of 0.7μg per μL of sterile double-distilled water. A total of thirty micrograms (30 μg) of the plasmid DNA (5:1 molar of pDAS000129 and pDAB104010) was applied to one million protoplasts (viability ≥95) suspended in 100 μl of transformation buffer (15mM $MgCl_2$, 0.1% (w/v) morpholinoethanesulphonic acid (MES) and 0.5 M mannitol; pH 5.8) followed by 150 μl of PEG solution (40% (w/v) PEG 4000 in 0.4 M Mannitol and 0.1 M Ca $(NO_3)_2$ (pH 6-7) Spangenberg and Potrykus (1995). Control transformations included a total of thirty micrograms (30 μg) of plasmid DNA of either pDAS000129 or pDAB104010. After 10-15 minutes of incubation at room temperature, 5 ml of W5 buffer was added in a drop wise manner and the protoplasts were gently mixed and

another 5 ml of W5 buffer was added as slow stream to the protoplasts suspension. Protoplasts were mixed gently and centrifuged at 400 RPM for 10 min and the W5 supernatant was removed carefully leaving behind the protoplasts in the form of a pellet. Transfected protoplasts were then incubated in 1 ml of W5 buffer at room temperature until they were embedded in bead type cultures. Protoplasts samples co-transfected with the construct pDAS000129 and pDAB104010 exhibited a cell viability ranging between 60-80% immediately after transfection. The transfected protoplasts were embedded following either sea plaque agarose or sodium alginate method.

Culturing of mesophyll derived protoplasts to recover viable microcalli

[0274] Before embedding the transfected protoplasts were centrifuged at 400 RPM for 10 min and the W5 buffer was carefully removed. The protoplasts were then resuspended in 0.5ml of K3 media (Spangenberg et al. 1998). Exactly 0.5 ml of the transfected protoplast suspension (ca. 5 x 10^5 protoplasts) was placed in a 6 cm Petri dish and to this 4.5 ml of pre-warmed (melted in a microwave oven and incubated in a water bath at 40-45°C) 1:1 mix of K3:H medium (Spangenberg et al. 1998) containing 0.6% Sea Plaque™ agarose was added. The agarose and the protoplasts suspension was mixed gently and allowed to set. After solidification (after 20-30 min.), seal the dishes were sealed with Parafilm® and the protoplasts were cultured for 24 h in darkness at 24°C, followed by 6 days in continuous dim light (5-10 μmol m^{-2} s^{-1}), where first and multiple cell divisions occur. After 6 days the protoplasts embedded in agarose was cut into four quadrants and placed in 100 ml of A medium (Spangenberg et al. 1998) in a 700 ml culture vessel. The liquid A medium, was supplemented with 1.5 mg/l hygromycin. The cultures were incubated on a rotary shaker with 80-100 RPM at 24°C in continuous dim light. Resistant colonies appear after 5-6 weeks and 3-4 weeks post protoplast plating in the case of sea-plaque agarose and sodium alginate method respectively. Microcalli of size between 2-3 mm diameter were transferred onto B1 medium (MS/MS Vitamins + 3.5 % Sucrose + 500mg/L MES + BAP (5μm) + NAA (5μm) + 2, 4-D (5μm) + 1.5 mg/L Hygromycin + 0.7 % Agarose Type I (pH 6.0)and poured in 100 x 20 mm sterile Petri dish) by gently breaking the agarose beads. The microcalli thus obtained was resuspended in sufficient quantity of liquid A (50 ml of liquid A was used for one ml of the settled cell volume (SCV: This was measured after transferring all the released microcalli to a sterile 50 or 15 ml falcon tube and allowed to settle down for 5 min)). After mixing the microcalli uniformly, 0.5ml of the microcalli suspended in the liquid A media was transferred

to B1 plates and using 1-2 ml of additional liquid A media the microcalli was distributed uniformly in the B1 media and the excess liquid A media was carefully removed from each plate. The plates were sealed using a micropore tape which enhanced the embryo maturation.

Sodium-Alginate method

[0275] Before embedding the transfected protoplasts were centrifuged at 400 RPM for 10 min and the W5 buffer was carefully removed. The protoplasts were then resuspended in 1.0 ml of 0.5 M Mannitol and incubated in ice. To this equal volume of 1.0 % sodium alginate was added and mixed gently. The protoplasts suspension was incubated in ice until it was embedded. Bead forming solution (0.4 M Mannitol + 50mM $CaCl_2$ (pH 5.8)) was transferred to a sterile six well plate (3-4 ml per well) using a serological pipette. Exactly 1.0ml of the protoplasts suspension was added in a drop wise manner using a 1 ml pipette into the bead forming solution and each transfected sample (ca. 5 x 10^5 protoplasts) was embedded per well. The protoplasts suspension was incubated for 1-2 hours at room temperature to form sodium alginate beads. After the incubation period the bead forming solution was carefully removed and replaced with 4-5ml of 1: 2 mixture of K3+H:A media (Spangenberg et al. 1998) supplemented with 1.5 mg/L of Hygromycin. The protoplasts were cultured for 3-4 weeks in darkness at 22°C in a shaker (50 RPM). After 3-4 weeks the resistant microcalli (0.5-1.0mm) were released by treating with depolymerisation buffer (0.3 M Mannitol + 20mM Sodium Citrate (pH 5.8)). After removing the liquid media 3-4 ml of depolymerisation buffer (was added to each well containing the bead-type cultures and incubated at room temperature for 2 hours. Using a sterile forceps the beads were gently mixed and to enhance the efficient release of the microcalli. Using a sterile 1.0 ml pipette gently mix gelling agent released in the depolymerisation buffer was removed. The microcalli was washed twice using 5ml of liquid A media and the microcalli was resuspended in sufficient quantity of liquid A (50 ml of liquid A was used for one ml of the settled cell volume (SCV: This was measured after transferring all the released microcalli to a sterile 50 or 15 ml falcon tube and allowed to settle down for 5 min)). After mixing the microcalli uniformly, 0.5ml of the microcalli suspended in the liquid A media was transferred to B1 media (MS/MS Vitamins + 3.5 % Sucrose + 500mg/L MES + BAP (5µm) + NAA (5µm) + 2, 4-D (5µm) + 1.5 mg/L Hygromycin + 0.7 % Agarose Type I (pH 6.0)and poured in 100 x 20 mm sterile Petri dish) and using 1-2 ml of additional liquid A media the microcalli was distributed uniformly in the B1 media and the excess liquid A media was carefully removed

from each plate. The plates were sealed using a micropore tape which enhanced the embryo maturation. The cultures were maintained at 22°C in 16h/d light (30 μmol m^{-2} s^{-1}).

ISOLATION OF GENOMIC DNA FROM MESOPHYLL PROTOPLASTS

[0276] 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 minute, followed by transfer of the liquid to the same 2 mL microfuge tube. At the end of each rinse, the cells were pelleted by centrifugation at 70 g and the supernatant was removed. The pelleted protoplasts were snap frozen in liquid nitrogen before freeze drying for 24 h in a LABCONCO FREEZONE 4.5® (Labconco, Kansas City, MO) at -40°C and 133 x 10^{-3} mBar pressure. The lyophilized cells were subjected to DNA extraction using the DNEASY® PLANT DNA EXTRACTION MINI KIT (Qiagen) following the manufacturer's instructions, with the exception that tissue disruption was not required and the protoplast cells were added directly to the lysis buffer.

ISOLATION OF GENOMIC DNA FROM CALLUS TISSUE

[0277] Individual calli was snap frozen in liquid nitrogen before freeze drying for 24 h in a LABCONCO FREEZONE 4.5® (Labconco, Kansas City, MO) at -40°C and 133 x 10^{-3} mBar pressure. The lyophilized calli was subjected to DNA extraction using the DNEASY® PLANT DNA EXTRACTION MAXI kit (Qiagen, Hilden, Germany) following the manufacturer's instructions.

DETECTION OF GENE ADDITION TO *FAD2A* BY HOMOLOGY DIRECTED REPAIR IN PROLIFERATED CALLUS

[0278] Genomic DNA was extracted from protoplast pools (one million protoplast per pool) to which donor DNA encoding a functional HGH reporter cassette (pDAS000129), ZFN DNA (pDAB104010) or a mixture of donor and ZFN DNA had been delivered twenty-four hours earlier. Quantities of DNA delivered for transformation are described above. PCR products were cloned into plasmid vectors. The genomic editing occurs independently in each cell giving rise to a variety of different insertion events, by cloning into a plasmid vector, each genomic edit can be sequenced without ambiguity. Several clones were sequenced on an ABI3730XL® automated capillary electrophoresis platform. Analysis of

gene sequences was done using SEQUENCHER SOFTWARE V5.0™ (GeneCodes, Ann Arbor, MI).

[0279] Evidence of gene addition to the FAD2A locus by homologous directed repair was provided by amplification of both the 5' and 3' FAD2A cassette and junctions from genomic DNA extracted from protoplasts using the primers described in **Table 13.** No amplification was observed from protoplasts to which ZFN plasmid or donor plasmid alone had been delivered. All junction sequences were indicative of insertion of the hgh cassette at the Fad2A locus via an HDR-mediated repair pathway. Deletions of varying lengths from either or both the genome and the cassette were observed as well as the addition of sequences derived from the vector backbones (either from the donor or ZFN) being inserted between the genome and the cassette (**Figure 15**).

Table 13: List of constructs used for donor integration within the FAD2A locus

Treatments	**Constructs**	**Hph Assay**	**In-out PCR (LB)**	**In-out PCR (RB)**
ZFN	pDAB104010 (No Hyg)	**NO**	**NO**	**NO**
ZFN	pDAB104010 (1.5mg/L Hyg)	**NO**	**NO**	**NO**
DNR	pDAS000129 (No Hyg)	**YES**	**NO**	**NO**
DNR	pDAS000129 (1.5mg/L Hyg)	**YES**	**NO**	**NO**
DNR + ZFN2c	pDAS000129 + pDAB104010 (No Hyg)	**YES**	**NO**	**NO**
DNR + ZFN2c	pDAS000129 + pDAB104010 (1.5mg/L Hyg)	**YES**	**YES**	**YES**
DsRed Ctrl	pDAS00097 (2mg/L PPT)	**NO**	**NO**	**NO**
negative control	untrasfected DH12075	**NO**	**NO**	**NO**
negative control	water	**NO**	**v**	**NO**

EXAMPLE 9: TARGETED INTEGRATION AND DISRUPTION OF *Brassica napus* OMEGA-3 FATTY ACID DESATURASE (FAD2) VIA NON HOMOLOGOUS END JOINING

SELECTION OF ZINC FINGER BINDING DOMAINS SPECIFIC TO FAD2A

[0280] The transcribed regions for homoeologous FAD2 genes were identified and characterized, zinc finger nucleases that were designed to bind and cleave these sites for NHEJ-mediated targeting of a donor sequence. Zinc finger proteins (ZFPs) directed against DNA sequences from homeologues of FAD2 sequences were designed and tested as described above. From the ZFNs showing on-target activity, one zinc finger proteins were selected that cut the FAD2A target at high efficiency: ZFP 24828-2A-24829 recognizes SEQ ID NO:35 5'- agGCCCAGtAGAGAGGCCaggcgaagta -3' and SEQ ID NO:36 5' - ccAGGGCTGCGTCCTAACCGgcgtctgg -3'. This ZFN was shown to specifically bind and cleave the FAD2A genomic locus. The plasmid construct, pDAB104010 that is previously describec above was constructed and is used for transformation experiments.

DESIGN AND CONSTRUCTION OF EXPRESSION VECTORS ENCODING ZINC FINGER NUCLEASES AND DONOR VECTORS SPECIFIC TO FAD2A

[0281] For integration of a donor sequence via NHEJ, a single donor vector was constructed. The vector encoded a *dsRED* reporter gene expression cassette. The *dsRED* reporter gene expression cassette included the *Arabidopsis thalinana* Ubiquitin 10 promoter (Callis, *et al.*, 1990, *J. Biol. Chem.*, 265:12486-12493) followed by the *dsRED* gene (Dietrich et al. (2002) Biotechniques 2(2):286-293). The *dsRED* gene was codon-optimised for expression in dicotyledonous plants and was flanked by a 3'UTR comprising the transcriptional terminator and polyadenylation site of Open Reading Frame 23 (ORF23) of *A. tumefaciens* pTi15955 (Barker *et al,* Plant Molecular Biology 1983, 2(6), 335-50). The selectable marker cassette includedthe CsVMV promoter fused to the *pat* transgene. The *pat* transgene was terminated with Open Reading Frame 1 (ORF1) of *A. tumefaciens* pTi15955 (Barker *et al,* Plant Molecular Biology 1983, 2(6), 335-50). The *dsRED* resistance cassette was cloned into specific restriction enzyme sites of each vector resulting in "donor" vector: pDAS000097 (SEQ ID NO:88, Figure 16). The pDAS00097 donor is designed to be delivered as linear DNA or circular DNA into the plant cell and integrated within the FAD2A locus upon cleavage of the FAD2A genomic locus by the ZFN pDAB104010. The linear DNA mediated integration is the result of integrating a linearized pDAS000097 plasmid into the plant cell during transformation. The plasmid can be linearized by cleavage at a unique

restriction enzyme site. The circular DNA mediated integration is the result of integrating a circularized pDAS000097 plasmid into the plant cell during the transformation. pDAS000097 is modified to contain a zinc finger binding site that can be cleaved by the ZFP 24828-2A-24829 zinc finger nuclease. The circular plasmid, pDAS000097, is cleaved in the plant cell by the pDAB104010 encoded zinc finger nuclease, and the dsRED gene cassette is integrated into the FAD2A genomic locus.

[0282] Colonies of the assembled plasmids were initially screened by restriction endonuclease digestion of DNA purified from overnight cultures of *E.coli*. Restriction endonucleases were obtained from NEW ENGLAND BIOLABS™ (NEB, Ipswich, MA) and PROMEGA™ (Promega Corporation, WI). Plasmid preparations were performed using the QIAPREP SPIN MINIPREP KIT™ (Qiagen, Hilden, Germany) or the PURE YIELD PLASMID MAXIPREP SYSTEM™ (Promega Corporation, WI) following the instructions of the suppliers. After the restriction fragments were confirmed by agarose gel electrophoresis of resulting fragments, plasmid DNA of selected clones were sequenced using ABI Sanger Sequencing and BIG DYE TERMINATOR V3.1™ cycle sequencing protocol (Applied Biosystems, Life Technologies). Sequence data were assembled and analyzed using the SEQUENCHER™ software (Gene Codes, Ann Arbor, MI).

TRANSFORMATION OF *BRASSICA NAPUS*

[0283] Mesophyll derived protoplasts are isolated and prepared from *Brassica napus* (DH10275) plants as described above. The protoplasts are transformed with purified plasmid DNA. Aliquots of donor and ZFN plasmid DNA are prepared in three molar ratios: 1:1 (30 µg of each plasmid), 5:1 (donor plasmid to ZFN plasmid to a total of 30 µg of plasmid DNA) and 10:1 (donor plasmid to ZFN plasmid to a total of 30 µg of plasmid DNA). Additionally, donor-only and ZFN-only aliquots (30 µg) are prepared as controls. The amounts of DNA delivered to the *B. napus* protoplasts via a PEG4000 mediated transformation are summarized in Table 14. The transformed protoplast cells are cultured as previously described, wherein the selection medium is glufosinate selection medium, and putative transformants are assayed via qPCR analysis for transgene insertions.

Table 14: Quantities of ZFN and donor DNA delivered to protoplasts

	Molar Ratio of plasmid DNA	*Total quantity of DNA (µg) delivered to 1 million protoplasts*
Splicing	Donor plasmid only	30
	ZFN plasmid only (pDAB104010)	30
	1:1 Donor:ZFN	60
	5:1 Donor:ZFN	30
	10:1 Donor:ZFN	30
Editing	Donor plasmid only	30
	1:1: ZFN plasmids (pDAB104010)	30
	1:1:1 Donor:ZFN:ZFN	90
	5:1:1 Donor:ZFN:ZFN	30
	10:1:1 Donor:ZFN:ZFN	30

DETECTION OF GENE ADDITION TO *FAD2A* BY NON-HOMOLOGOUS END JOINING IN PROTOPLASTS

[0284] Genomic DNA is extracted from protoplast pools (one million protoplast per pool) to which donor DNA encoding a functional dsRFP reporter cassette (pDAS000097), ZFN DNA (pDAB104010) or a mixture of donor and ZFN DNA are delivered twenty-four hours earlier. Quantities of DNA delivered for transformation are described above. PCR products are cloned into plasmid vectors. The genomic editing occurs independently in each cell giving rise to a variety of different insertion events, by cloning into a plasmid vector, each genomic edit can be sequenced without ambiguity. Several clones are sequenced on an ABI3730XL® automated capillary electrophoresis platform. Analysis of gene sequences is done using SEQUENCHER SOFTWARE V5.0™ (GeneCodes, Ann Arbor, MI).

[0285] Evidence of gene addition to the FAD2A locus by editing or splicing is provided by amplification of both the 5' and 3' FAD2A-cassette junctions from genomic DNA extracted from protoplasts. No amplification is observed from protoplasts to which

ZFN plasmid or donor plasmid alone had been delivered. All junction sequences are indicative of insertion of the dsRED cassette at the FAD2A locus via an NHEJ-mediated repair pathway. Deletions of varying lengths from either or both the genome and the cassette are observed as well as the addition of sequences derived from the vector backbones (either from the donor or ZFN) being inserted between the genome and the cassette.

DETECTION OF GENE ADDITION TO *FAD2A* BY NON-HOMOLOGOUS END JOINING IN CALLUS TISSUE REGENERATED FROM PROTOPLASTS

[0286] Further evidence of splicing and editing of the FAD2A locus was obtained from callus tissue regenerated from protoplasts on selection to which donor DNA encoding a *dsRED* cassette (pDAS000097), ZFN DNA only (pDAB104010) or donor and ZFN DNA are delivered. DNA is extracted from approximately 80 calli for each ratio.

[0287] Integration of the *dsRED* cassette into the *B. napus* genome is confirmed by TAQMAN™ qPCR using primer and probes specific to the donor insert and the genomic flanking sequences. Relative quantification is calculated according to the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001), which provided an estimation of the number of copies of *dsRED* cassette inserted into the genome. Evidence of NHEJ-mediated splicing and editing of FAD2A is obtained by conducting PCR assays with one primer specific to FAD2A and a second primer specific to either the promoter or terminator of the *dsRED* cassette. PCR products are gel-purified using QIAQUICK MINIELUTE PCR PURIFICATION KIT™ (Qiagen) and sequenced using a direct Sanger sequencing method. The sequencing products are purified with ethanol, sodium acetate and EDTA following the BIGDYE® v3.1 protocol (Applied Biosystems) and sequenced and analysed as above.

[0288] The numbers of calli containing the donor cassette in each experiment are determined. Evidence of donor gene addition to the FAD2A locus by editing and/or splicing is provided by PCR amplification across the ZFN cut sites and both the 5' and 3' FAD2A-*dsRED* cassette junctions. PCR amplification of the genomic DNA isolated from callus tissue recovered from control protoplasts which are transformed with only the *dsRED* plasmid (pDAS000097) or only the ZFN plasmid (pDAB104010) do not result in the production of PCR amplification products.

[0289] The PCR amplicons produced from the amplification of the 5' and 3' FAD2A-*dsRED* cassette junctions are purified from the agarose gel and sequenced to confirm

specificity of the integration within the FAD2A genomic locus. The results of the sequencing analysis of the PCR products indicate that each isolated callus which is generated from an individually transformed protoplast only produce a single PCR amplification product and do not contain cells of mixed genotypes.

Detection of gene addition to *FAD2A* by non-homologous end joining in plants

[0290] DNA is extracted from plants that are regenerated from protoplasts and transferred to potting medium. The majority of plants recovered are estimated to contain only 1-2 copies of the *dsRED* cassette encoded in the donor DNA. Plants are analyzed with the same suite of assays described for callus tissue as well as with assays to determine if the cassette had inserted in the FAD2A locus.

[0291] The frequency of on-target splicing, where the *dsRED* cassette is inserted into FAD2A locus is determined using the PCR assays described above. The amplicon bands obtained are sequenced to determine the flanking sequences. Additionally, plants are screened for off-target insertions to determine the frequency of integration of *dsRED* at sites other than FAD2A.

EXAMPLE 10: TARGETED INTEGRATION OF *Brassica napus* OMEGA-3 FATTY ACID DESATURASE (FAD2) WITH AN AGRONOMICALLY IMPORTANT GENE

[0292] Constructs containing the DGT-28 transgene (International Patent Application No. WO/2013/116700, herein incorporated by reference) that confers resistance to the herbicide glyphosate are designed and built for integration within the FAD2A genomic loci of *Brassica napus*. Exemplary donor constructs include pDAS000389 (Figure 17, SEQ ID NO:89) for NHEJ integration within FAD2A locus, pDAS000391 (Figure 18, SEQ ID NO:90) for NHEJ integration within FAD2A locus, pDAS000392 (Figure 19, SEQ ID NO:91) for NHEJ integration within FAD2A locus, pDAS000393 (Figure 20, SEQ ID NO:92) for NHEJ integration within FAD2A locus, pDAS000394 (Figure 21, SEQ ID NO:93) for HDR integration within the ETIP site of the FAD2A locus, pDAS000395 (Figure 22, SEQ ID NO:208) for HDR integration within the ETIP site of the FAD2A locus, pDAS000396 (Figure 23, SEQ ID NO:209) for HDR integration within FAD2A locus, and pDAS000397 (Figure 24, SEQ ID NO:210) for HDR integration within FAD2A locus. The constructs and associated zinc finger nuclease constructs (*e.g.*, pDAB104010) are transformed into *Brassica*

napus cells as previously described above. Transformants are identified and confirmed via molecular confirmation assays as previously described. The FAD2A chromosomal integrants, comprising an integrated *dgt-28* transgene are isolated. The integration of the *dgt-28* transgene within the FAD2A locus is exemplified via NHEJ mediated integration and HDR mediated integration. The integration within the FAD2A locus can be directed into the FAD2A endogenous sequence or into the previouslt described ETIP (pDAS000130) that is stably integrated within the FAD2A locus. The integration within the FAD2A locus via an NHEJ mediated mechanism can be made using linearized donor or circular donor DNA designs. Transformed DGT-28 *Brassica napus* events are obtained and tested for robust expression of the DGT-28 and the subsequent resistance to the herbicide glyphosate.

[0293] While certain exemplary embodiments have been described herein, those of ordinary skill in the art will recognize and appreciate that many additions, deletions, and modifications to the exemplary embodiments may be made without departing from the scope of the following claims. In addition, features from one embodiment may be combined with features of another embodiment.

CLAIMS

What may be claimed is:

1. A method for modifying the genome of a cell, the method comprising:

cleaving, in a site specific matter, a target site in a FAD2 gene in a cell, to thereby generate a break in the FAD2 gene;

wherein the FAD2 gene is modified following cleavage.

2. The method of claim 1, wherein the method further comprises integrating into the break a nucleic acid sequence of interest.

3. The method according to claim 1 or claim 2, wherein the FAD2 gene is a FAD2A, FAD2A', FAD2C and/or a FAD2C' gene.

4. The method according to any of claims 1 to 3, wherein the cleaving in a site specific manner comprises introducing a fusion protein comprising a DNA-binding domain and a cleavage domain or cleavage half-domain into the cell or a polynucleotide encoding the fusion protein, wherein the fusion protein binds with specificity to the target site and cleaves at or near the target site to thereby generate the break.

5. The method according to claim 4, wherein the DNA-binding domain is selected from the group consisting of a meganuclease DNA-binding domain, a leucine zipper DNA-binding domain, a transcription activator-like (TAL) DNA-binding domain, a RNA-guided CRISPR-Cas9, a recombinase, a zinc finger protein DNA-binding domain, and chimeric combinations of any of the foregoing.

6. The method according to claim 4 or claim 5, wherein the cleavage domain or cleavage half-domain is selected from the group consisting of a cleavage half-domain from a type IIS restriction endonuclease, a cleavage half-domain from FokI endonuclease, a cleavage half-domain from StsI endonuclease, and a homing endonuclease.

7. The method according to any of claims 4 to 6, wherein the fusion protein is a zinc finger nuclease.

8. The method according to claim 7, wherein the zinc finger nuclease comprises from three to six zinc finger domains, each zinc finger domain comprising a recognition helix region, wherein the zinc finger protein comprises the recognition helix regions ordered and shown in a single row of Table 3.

9. The method according to any of claims 1 to 8, wherein the cleaving in a site specific manner is specific for some but not all copies of FAD2A, FAD2A', FAD2C and/or a FAD2C'.

10. The method according to any of claims 1 to 9, wherein the target site is selected from the group consisting of SEQ ID NOs: 22-26, and SEQ ID NOs: 28-33 and SEQ ID NOs:35-38.

11. The method according to any of claims 1 to 10, wherein the cell is a plant, fungi, bacteria or algae cell.

12. The method according to claim 11, wherein the plant cell is a monocot plant cell or a dicot plant cell.

13. The method according to claim 12, wherein the plant cell is selected from the group consisting of *Brassica sp.*, *Brassica napus*; *Brassica rapa*; *Brassica juencea*; *Brassica oleracea; Brassica nigra; Zea sp.*; *Zea mays*; *Glycine sp.*; *Glycine max*; *Triticum sp; Triticum aestivum; Oryza sp; Oryza sativa; Triticae sp.*; *Triticae triticum*; *Heliantheae sp.*; *Heliantheae helianthus*; *Gossypium sp.*; *Gossypium hirsutum; and Hordeum vulgar*.

14. The method according to any of claims 2 to 13, wherein the nucleic acid sequence of interest is selected from the group consisting of a sequence comprising a DNA-binding domain binding target site, one or more insecticidal resistance genes, one or more herbicide tolerance genes, one or more nitrogen use efficiency genes, one or more water use efficiency genes, one or more nutritional quality genes, one or more DNA binding genes, one or more selectable marker genes and combinations thereof.

15. A cell, seed or plant comprising a cell modified according to the method of any of claims 1 to 14.

16. The cell, seed or plant according to claim 15, wherein the cell, seed or plant is transgenic cell, seed or plant comprising a nucleotide sequence of interest integrated into one or more copies of a FAD2A, FAD2A', FAD2C and/or a FAD2C' gene.

17. The cell, seed or plant of claim 16, wherein the nucleotide sequence is heterologous or homologous to the cell.

18. The cell, seed or plant of claim 14, wherein the homologous sequence comprises at least one single nucleotide polymorphism.

19. The cell, seed or plant of any of claims 16 to 18, wherein the nucleic acid sequence is integrated at or near a target site selected from the group consisting of SEQ ID NOs: 22-26, and SEQ ID NOs: 28-33 and SEQ ID NOs:35-38.

20. A site specific zinc finger nuclease that cleaves at or near a nucleic acid target site selected from the group consisting of SEQ ID NOs: 22-26, and SEQ ID NOs: 28-33 and SEQ ID NOs:35-38.

21. The zinc finger nuclease of claim 20, wherein the zinc finger nuclease comprises from three to six zinc finger domains, each zinc finger domain comprising a recognition helix region, wherein the zinc finger protein comprises the recognition helix regions ordered and shown in a single row of Table 3.

FIG1A

```
                                        1                                  40
FAD2-3 (SEQ ID NO:8)      (1) ATGGGTGCAGGTGGAAGAATGCAAGTGTCTCCTCCCTCCA
 FAD2A (SEQ ID NO:5)      (1) ATGGGTGCAGGTGGAAGAATGCAAGTGTCTCCTCCCTCCA
FAD2-2 (SEQ ID NO:7)      (1) ATGGGCGCAGGTGGAAGAATGCAAGTCTCTCCTCCCTCCA
FAD2-1 (SEQ ID NO:6)      (1) ATGGGTGCAGGTGGAAGAATGCAAGTCTCTCCTCCCTCCA
                                       41                                  80
FAD2-3 (SEQ ID NO:8)     (41) AGAAGTCTGAAACCGACACCATCAAGCGCGTACCCTGCGA
 FAD2A (SEQ ID NO:5)     (41) AAAAGTCTGAAACCGACAACATCAAGCGCGTACCCTGCGA
FAD2-2 (SEQ ID NO:7)     (41) GCTCCCCCGAAACCAAAACCCTCAAACGCGTCCCCTGCGA
FAD2-1 (SEQ ID NO:6)     (41) GCTCCCCCGGAACCAACACCCTCAAACGCGTCCCCTGCGA
                                       81                                 120
FAD2-3 (SEQ ID NO:8)     (81) GACACCGCCCTTCACTGTCGGAGAACTCAAGAAAGCAATC
 FAD2A (SEQ ID NO:5)     (81) GACACCGCCCTTCACTGTCGGAGAACTCAAGAAAGCAATC
FAD2-2 (SEQ ID NO:7)     (81) GACACCACCCTTCACTCTCGGAGACCTCAAGAAAGCAATC
FAD2-1 (SEQ ID NO:6)     (81) GACACCACCATTCACTCTCGGAGACCTCAAGAAAGCAATC
                                      121                                 160
FAD2-3 (SEQ ID NO:8)    (121) CCACCGCACTGTTTCAAACGCTCGATCCCTCGCTCTTTCT
 FAD2A (SEQ ID NO:5)    (121) CCACCGCACTGTTTCAAACGCTCGATCCCTCGCTCTTTCT
FAD2-2 (SEQ ID NO:7)    (121) CCACCTCACTGCTTCAAACGCTCCATCCCTCGCTCCTTCT
FAD2-1 (SEQ ID NO:6)    (121) CCACCTCACTGCTTCAAACGCTCCATCCCACGCTCCTTCT
                                      161                                 200
FAD2-3 (SEQ ID NO:8)    (161) CCTACCTCATCTGGGACAT--CATCATAGCCTCCTGCTTC
 FAD2A (SEQ ID NO:5)    (161) CCTACCTCATCTGGGACAT--CATCATAGCCTCCTGCTTC
FAD2-2 (SEQ ID NO:7)    (161) CCTACCTCCTCTTCGACAT--CCTCGTCTCCTCCTCCCTC
FAD2-1 (SEQ ID NO:6)    (161) CCT-CTTCGACATCATCATCTCCTCCTCGGCTCCTCCCTC
                                      201                                 240
FAD2-3 (SEQ ID NO:8)    (199) TACTACGTCGCCACCACTTACTTCCCTCTCCTCCCTCACC
 FAD2A (SEQ ID NO:5)    (199) TACTACGTCGCCACCACTTACTTCCCTCTCCTCCCTCACC
FAD2-2 (SEQ ID NO:7)    (199) TACCACCTCTCCACAGCCTACTTCCCTCTCCTCCCCCACC
FAD2-1 (SEQ ID NO:6)    (200) TACCACCTCTCCACAGCCTACTTCCCTCTCC---------
```

FIG 1B

```
                                      241                                 280
FAD2-3 (SEQ ID NO:8)    (239) CTCTCTCCTACTTCGCCTGGCCTCTCTACTGGGCCTGCCA
 FAD2A (SEQ ID NO:5)    (239) CTCTCTCCTACTTCGCCTGGCCTCTCTACTGGGCCTGCCA
```

```
FAD2-2 (SEQ ID NO:7)    (239) CTCTCCCTTACCTCGCCTGGCCCCTCTACTGGGCCTGCCA
FAD2-1 (SEQ ID NO:6)    (231) ------CTTACCTCGCCTGACCCCTCTACTGGGCCTGCCA
                              281                                   320
FAD2-3 (SEQ ID NO:8)    (279) AGGGTGCGTCCTAACCGGCGTCTGGGTCATAGCCCACGAG
 FAD2A (SEQ ID NO:5)    (279) GGGCTGCGTCCTAACCGGCGTCTGGGTCATAGCCCACGAG
FAD2-2 (SEQ ID NO:7)    (279) AGGCTGCGTCCTAACGGGCCTCTGGGTCATCGCCCACGAA
FAD2-1 (SEQ ID NO:6)    (265) AGGCTGCGTCCTAACGGGCCTCTGGGTCATAGCCCACGAG
                              321                                   360
FAD2-3 (SEQ ID NO:8)    (319) TGCGGCCACCACGCCTTCAGCGACTACCAGTGGCTTGACG
 FAD2A (SEQ ID NO:5)    (319) TGCGGCCACCACGCCTTCAGCGACTACCAGTGGCTGGACG
FAD2-2 (SEQ ID NO:7)    (319) TGCGGCCACCACGCCTTCAGCGACCACCAGTGGCTGGACG
FAD2-1 (SEQ ID NO:6)    (305) TGCGGCCACCACGCCTTCAGCGACCACCAGTGGCTGGACG
                              361                                   400
FAD2-3 (SEQ ID NO:8)    (359) ACACCGTCGGTCTCATCTTCCACTCCTTCCTCCTCGTCCC
 FAD2A (SEQ ID NO:5)    (359) ACACCGTCGGCCTCATCTTCCACTCCTTCCTCCTCGTCCC
FAD2-2 (SEQ ID NO:7)    (359) ACGCCGTGGGCCTCGTCTTCCACTCCTTCCTCCTCGTCCC
FAD2-1 (SEQ ID NO:6)    (345) ACGCCGCCGGCCTCGTCTTCCACTCCTTCCTCCTCGTCCC
                              401                                   440
FAD2-3 (SEQ ID NO:8)    (399) TTACTTCTCCTGGAAGTACAGTCATCGACGCCACCATTCC
 FAD2A (SEQ ID NO:5)    (399) TTACTTCTCCTGGAAGTACAGTCATCGACGCCACCATTCC
FAD2-2 (SEQ ID NO:7)    (399) TTACTTCTCCTGGAAGTACAGCCATCGACGCCACCATTCC
FAD2-1 (SEQ ID NO:6)    (385) GTACTTCTCCTGGAAGTACATCCAT-GACGCCACCATTCC
                              441                                   480
FAD2-3 (SEQ ID NO:8)    (439) AACACTGGCTCCCTCGAGAGAGACGAAGTGTTTGTCCCCA
 FAD2A (SEQ ID NO:5)    (439) AACACTGGCTCCCTCGAGAGAGACGAAGTGTTTGTCCCCA
FAD2-2 (SEQ ID NO:7)    (439) AACACCGGATCCCTCGAGAGGGATGAAGTGTTCGTCCCCA
FAD2-1 (SEQ ID NO:6)    (424) AACACCGGATCCCTCGATAGGGACGAAGTGTTCGTCCCCA
FIG1C                         481                                   520
FAD2-3 (SEQ ID NO:8)    (479) AGAAGAAGTCAGACATCAAGTGGTACGGCAAGTACCTCAA
 FAD2A (SEQ ID NO:5)    (479) AGAAGAAGTCAGACATCAAGTGGTACGGCAAGTACCTCAA
FAD2-2 (SEQ ID NO:7)    (479) AGAAGAAATCCGACATCAAGTGGTACGGAAAGTACCTCAA
FAD2-1 (SEQ ID NO:6)    (464) AGAAGAAATCCGACATCAAGTGGTACGGCAAGTACCTCAA
                              521                                   560
```

```
FAD2-3 (SEQ ID NO:8)   (519) CAACCCTTTGGGACGCACCGTGATGTTAACGGTTCAGTTC
 FAD2A (SEQ ID NO:5)   (519) CAACCCTTTGGGACGCACCGTGATGTTAACGGTTCAGTTC
FAD2-2 (SEQ ID NO:7)   (519) CAACCCGCTAGGACGCACGGTGATGCTAACCGTCCAGTTC
FAD2-1 (SEQ ID NO:6)   (504) CAACCCGCTAGGACGCACGGTGATGCTAACCGTCCAGTTC
                              561                                   600
FAD2-3 (SEQ ID NO:8)   (559) ACTCTCGGCTGGCCGTTGTACTTAGCCTTCAACGTCTCGG
 FAD2A (SEQ ID NO:5)   (559) ACTCTCGGCTGGCCTTTGTACTTAGCCTTCAACGTCTCGG
FAD2-2 (SEQ ID NO:7)   (559) ACGCTCGGCTGGCCGTTGTACTTAGCCTTCAACGTCTCTG
FAD2-1 (SEQ ID NO:6)   (544) AAGCTCGGCTGGCCGTTGTACTTAGCCTTCAACGTCTCGG
                              601                                   640
FAD2-3 (SEQ ID NO:8)   (599) GAAGACCTTACGACGGCGGCTTCGCTTGCCATTTCCACCC
 FAD2A (SEQ ID NO:5)   (599) GGAGACCTTACGACGGCGGCTTCGCTTGCCATTTCCACCC
FAD2-2 (SEQ ID NO:7)   (599) GAAGACCTTACAGCGACGGTTTCGCTTGCCATTTCCACCC
FAD2-1 (SEQ ID NO:6)   (584) GAAGACCTTACAGCGACGGTTTCGCTTGCCATTTCCACCC
                              641                                   680
FAD2-3 (SEQ ID NO:8)   (639) CAACGCTCCCATCTACAACGACCGCGAGCGTCTCCAGATA
 FAD2A (SEQ ID NO:5)   (639) CAACGCTCCCATCTACAACGACCGTGAGCGTCTCCAGATA
FAD2-2 (SEQ ID NO:7)   (639) GAACGCTCCCATCTACAACGACCGCGAGCGTCTCCAGATA
FAD2-1 (SEQ ID NO:6)   (624) GAACGCTCCCATCTACAACGACCGCGAGCGTCTCCAGATA
                              681                                   720
FAD2-3 (SEQ ID NO:8)   (679) TACATCTCCGACGCTGGCATCCTCGCCGTCTGCTACGGTC
 FAD2A (SEQ ID NO:5)   (679) TACATCTCCGACGCTGGCATCCTCGCCGTCTGCTACGGTC
FAD2-2 (SEQ ID NO:7)   (679) TACATCTCTGACGCTGGCGTCCTCTCCGTATGTTACGGTC
FAD2-1 (SEQ ID NO:6)   (664) TACATCTCTGACGCTGGCGTCCTCTCCGTATGTTACGGTC
FIG1D                         721                                   760
FAD2-3 (SEQ ID NO:8)   (719) TCTTCCGTTACGCCGCCGCGCAGGGAGTGGCCTCGATGGT
 FAD2A (SEQ ID NO:5)   (719) TCTACCGCTACGCTGCTGTCCAAGGAGTTGCCTCGATGGT
FAD2-2 (SEQ ID NO:7)   (719) TCTACCGCTACGCTGGTTCGCGAGGAGTGGCCTCGATGGT
FAD2-1 (SEQ ID NO:6)   (704) TCTACCGTTACGCTGCTTCGCGAGGAGTAGCCTCTGTGGT
                              761                                   800
FAD2-3 (SEQ ID NO:8)   (759) CTGCTTCTACGGAGTCCCGCTTCTGATTGTCAATGGTTTC
 FAD2A (SEQ ID NO:5)   (759) CTGCTTCTACGGAGTTCCTCTTCTGATTGTCAACGGGTTC
FAD2-2 (SEQ ID NO:7)   (759) CTGTGTCTACGGAGTTCCGCTTATGATTGTCAACTGTTTC
```

```
FAD2-1 (SEQ ID NO:6)    (744) CTGTGTCTACGGAGTTCCGCTTCTAATTGTCAACTGTTTC
                              801                                   840
FAD2-3 (SEQ ID NO:8)    (799) CTCGTGTTGATCACTTACTTGCAGCACACGCATCCTTCCC
 FAD2A (SEQ ID NO:5)    (799) TTAGTTTTGATCACTTACTTGCAGCACACGCATCCTTCCC
FAD2-2 (SEQ ID NO:7)    (799) CTCGTCTTGATCACTTACTTGCAGCACACGCACCCTTCGC
FAD2-1 (SEQ ID NO:6)    (784) CTCGTCTTGATCACTTACTTGCAGCACACGCACCCTTCGC
                              841                                   880
FAD2-3 (SEQ ID NO:8)    (839) TGCCTCACTACGATTCGTCCGAGTGGGATTGGTTGAGGGG
 FAD2A (SEQ ID NO:5)    (839) TGCCTCACTATGACTCGTCTGAGTGGGATTGGTTGAGGGG
FAD2-2 (SEQ ID NO:7)    (839) TGCCTCACTATGATTCTTCGGAGTGGGATTGGTTGAGAGG
FAD2-1 (SEQ ID NO:6)    (824) TGCCTCACTATGATTCTTCCGAGTGGGATTGGTTGAGAGG
                              881                                   920
FAD2-3 (SEQ ID NO:8)    (879) AGCTTTGGCTACCGTTGACAGAGACTACGGAATCTTGAAC
 FAD2A (SEQ ID NO:5)    (879) AGCTTTGGCCACCGTTGACAGAGACTACGGAATCTTGAAC
FAD2-2 (SEQ ID NO:7)    (879) AGCTTTGGCTACTGTGGATAGAGACTATGGAATCTTGAAC
FAD2-1 (SEQ ID NO:6)    (864) AGCTTTGGCTACTGTGGATAGAGACTATGGAATCTTGAAC
                              921                                   960
FAD2-3 (SEQ ID NO:8)    (919) AAGGTCTTCCACAATATTACCGACACGCACGTGGCGCATC
 FAD2A (SEQ ID NO:5)    (919) AAGGTCTTCCACAATATCACGGACACGCACGTGGCGCATC
FAD2-2 (SEQ ID NO:7)    (919) AAGGTGTTTCATAACATCACGGACACGCACGTGGCGCATC
FAD2-1 (SEQ ID NO:6)    (904) AAGGTGTTCCATAACATCACGGACACGCACGTGGCGCATC
FIG1E                         961                                  1000
FAD2-3 (SEQ ID NO:8)    (959) ATCTGTTCTCCACGATGCCGCATTATCACGCGATGGAAGC
 FAD2A (SEQ ID NO:5)    (959) ACCTGTTCTCGACCATGCCGCATTATCACGCGATGGAAGC
FAD2-2 (SEQ ID NO:7)    (959) ATCTGTTCTCGACGATGCCGCATTATAACGCGATGGAAGC
FAD2-1 (SEQ ID NO:6)    (944) ATCTGTTCTCGACGATGCCGCATTATAACGCGATGGAAGC
                             1001                                  1040
FAD2-3 (SEQ ID NO:8)    (999) TACCAAGGCGATAAAGCCGATACTG-GGAGAGTATTATCA
 FAD2A (SEQ ID NO:5)    (999) TACGAAGGCGATAAAGCCGATACTG-GGAGAGTATTATCA
FAD2-2 (SEQ ID NO:7)    (999) GACCAAGGCGATAAAGCCGATACTT-GGAGAGTATTACCA
FAD2-1 (SEQ ID NO:6)    (984) GACCAAGGCGATAAAGCCGATACTTTGGAGAGTATTACCA
                             1041                                  1080
FAD2-3 (SEQ ID NO:8)   (1038) GTTCGATGGGACGCCGGTGGTTAAGGCGATGTGGAGGGAG
```

```
 FAD2A (SEQ ID NO:5)  (1038) GTTCGATGGGACGCCGGTGGTTAAGGCGATGTGGAGGGAG
FAD2-2 (SEQ ID NO:7)  (1038) GTTTGATGGAACGCCGGTGGTTAAGGCGATGTGGAGGGAG
FAD2-1 (SEQ ID NO:6)  (1024) GTTTGATGGAACGCCGGCGGTTAAGGCGATGTGGAGGGAG
                             1081                                1120
FAD2-3 (SEQ ID NO:8)  (1078) GCGAAGGAGTGTATCTATGTGGAACCGGACAGGCAAGGTG
 FAD2A (SEQ ID NO:5)  (1078) GCGAAGGAGTGTATCTATGTGGAACCGGACAGGCAAGGTG
FAD2-2 (SEQ ID NO:7)  (1078) GCGAAGGAGTGTATCTATGTTGAACCGGATAGGCAAGGTG
FAD2-1 (SEQ ID NO:6)  (1064) GCGAAGGAGTGTATCTATGTTGAACCGGATAGGCAAGGTG
                             1121                                1160
FAD2-3 (SEQ ID NO:8)  (1118) AGAAGAAAGGTGTGTTCTGG--------------------
 FAD2A (SEQ ID NO:5)  (1118) AGAAGAAAGGTGTGTTCTGGTACAACAATAAGTTATCTTG
FAD2-2 (SEQ ID NO:7)  (1118) AGAAGAAAGGTGTGTTCTGGTACAACAATAAGTTATGAGG
FAD2-1 (SEQ ID NO:6)  (1104) AGAAGAAAGGTGTGTTCTGGTACAACAATAA---------
                             1161
FAD2-3 (SEQ ID NO:8)  (1138) ----
 FAD2A (SEQ ID NO:5)  (1158) CTAA
FAD2-2 (SEQ ID NO:7)  (1158) ATGA
FAD2-1 (SEQ ID NO:6)  (1135) ----
```

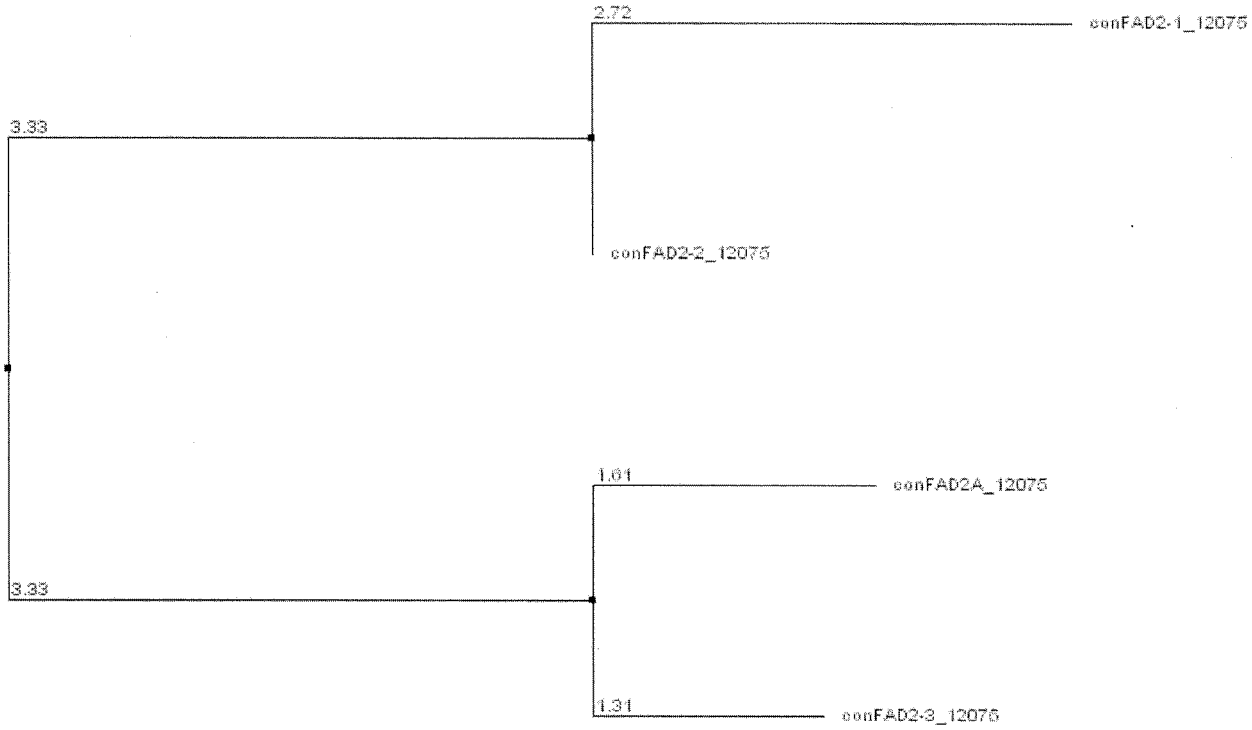

2.72
conFAD2-1_12075
3.33
conFAD2-2_12075
1.01
conFAD2A_12075
3.33
1.31
conFAD2-3_12075

Figure 2

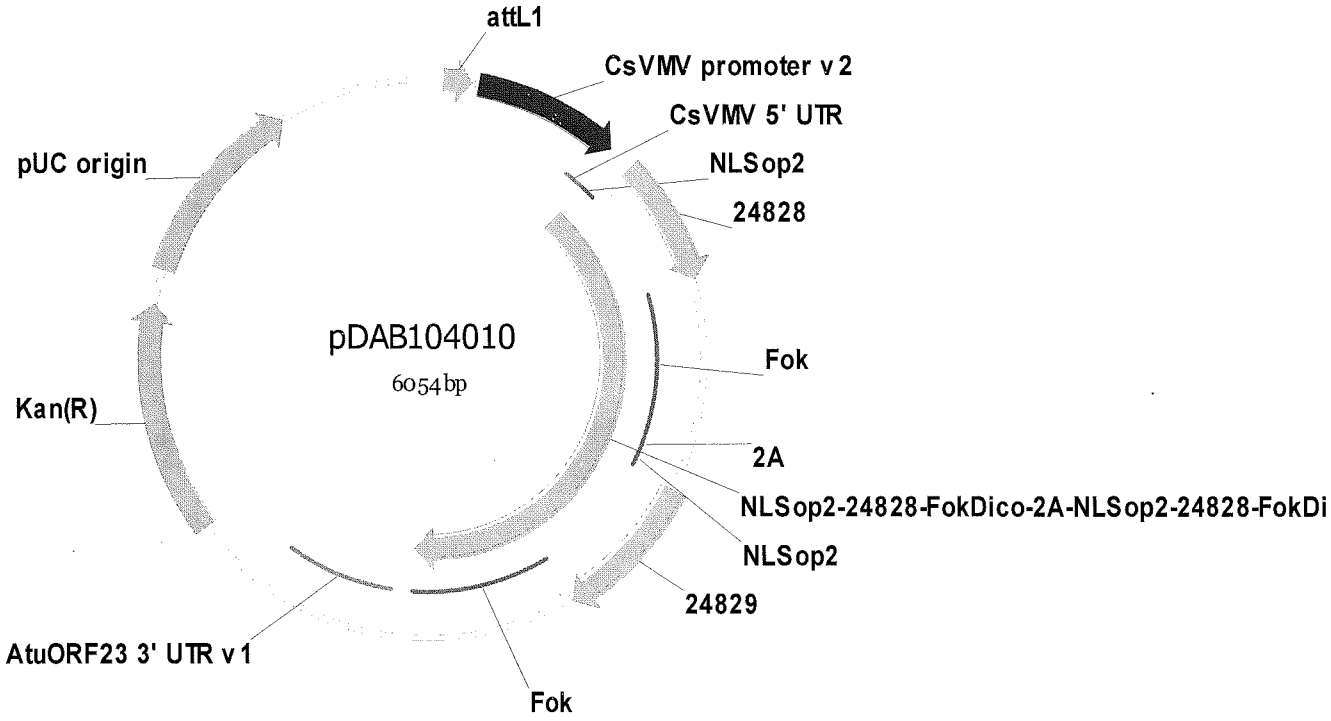

attL1
CsVMV promoter v2
CsVMV 5' UTR
NLSop2
24828
Fok
2A
NLSop2-24828-FokDico-2A-NLSop2-24828-FokDi
NLSop2
24829
Fok
AtuORF23 3' UTR v1
Kan(R)
pUC origin
pDAB104010
6054bp

Figure 3

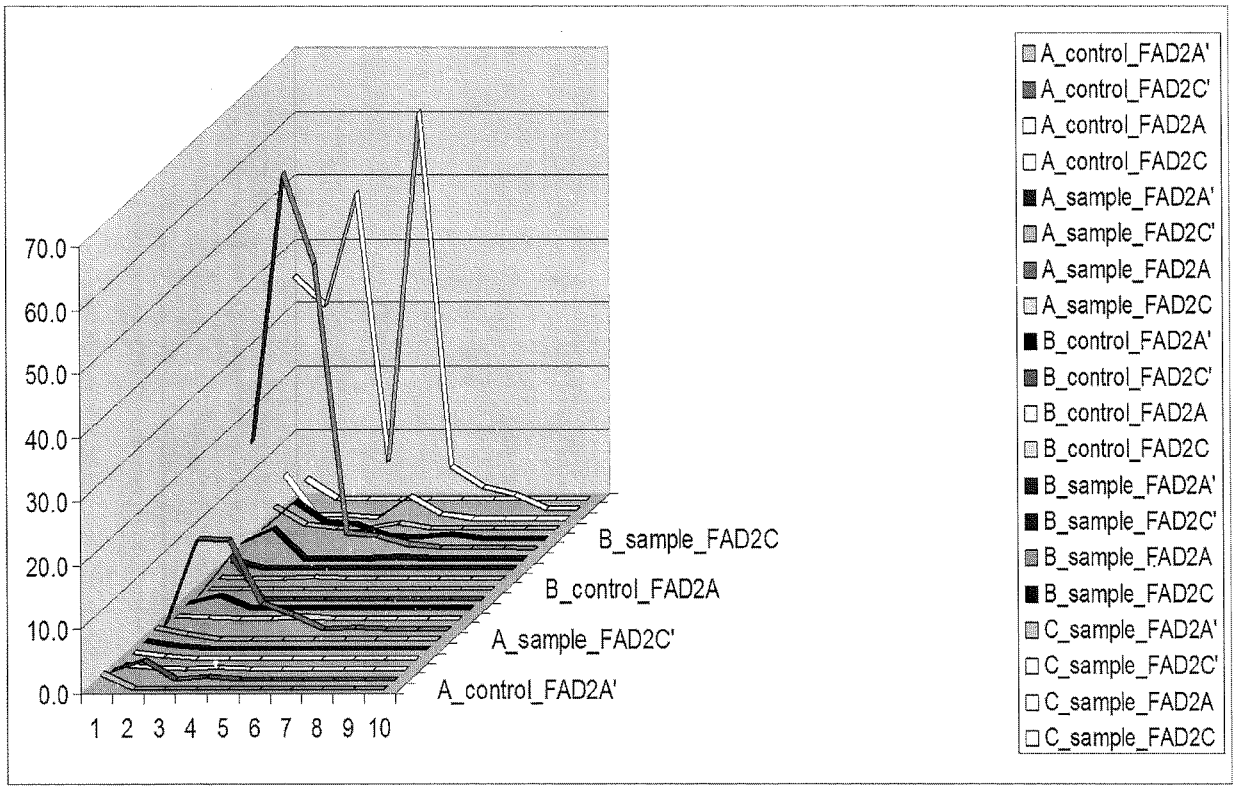

70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
1 2 3 4 5 6 7 8 9 10
B_sample_FAD2C
B_control_FAD2A
A_sample_FAD2C'
A_control_FAD2A'
A_control_FAD2A'
A_control_FAD2C'
A_control_FAD2A
A_control_FAD2C
A_sample_FAD2A'
A_sample_FAD2C'
A_sample_FAD2A
A_sample_FAD2C
B_control_FAD2A'
B_control_FAD2C'
B_control_FAD2A
B_control_FAD2C
B_sample_FAD2A'
B_sample_FAD2C'
B_sample_FAD2A
B_sample_FAD2C
C_sample_FAD2A'
C_sample_FAD2C'
C_sample_FAD2A
C_sample_FAD2C

Figure 4

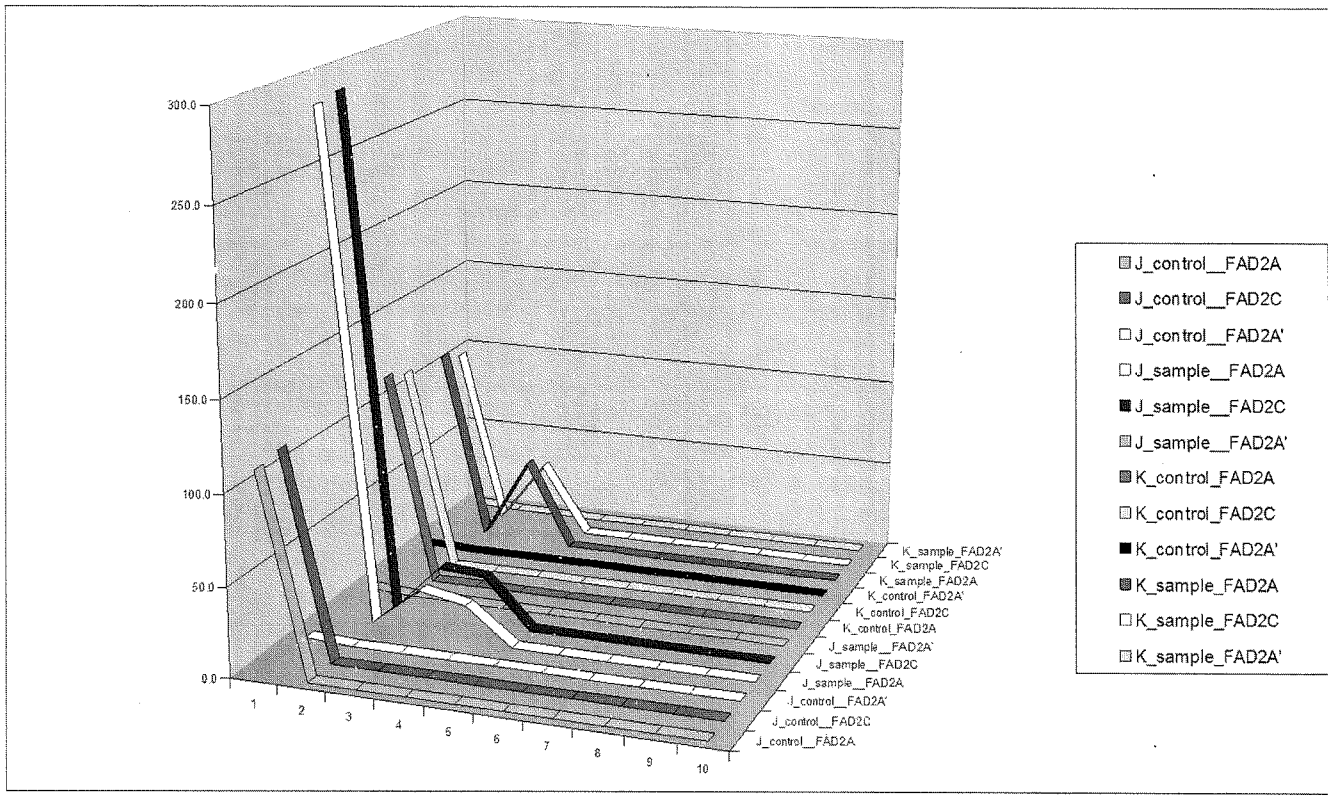


Figure 5A

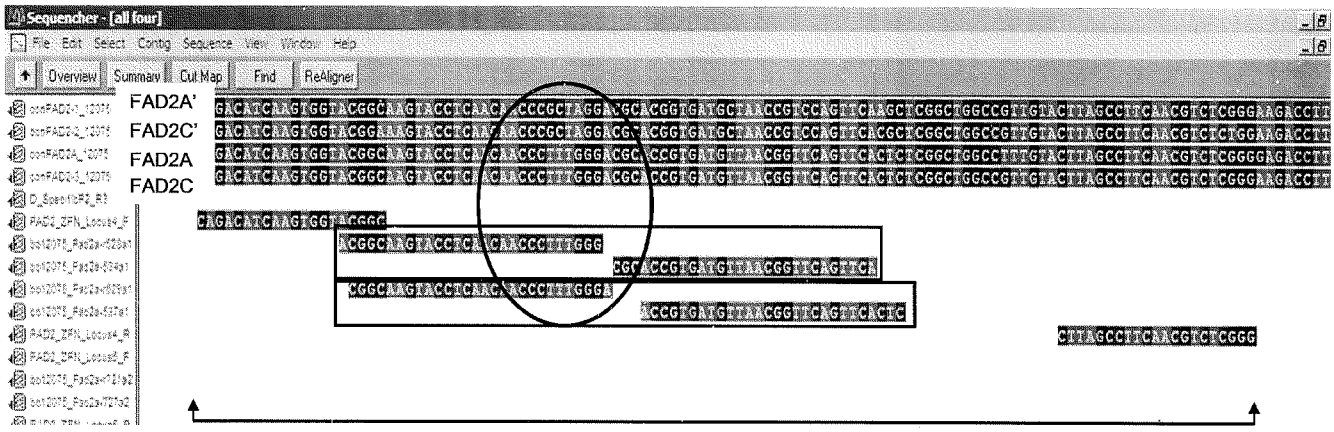


Figure 5B

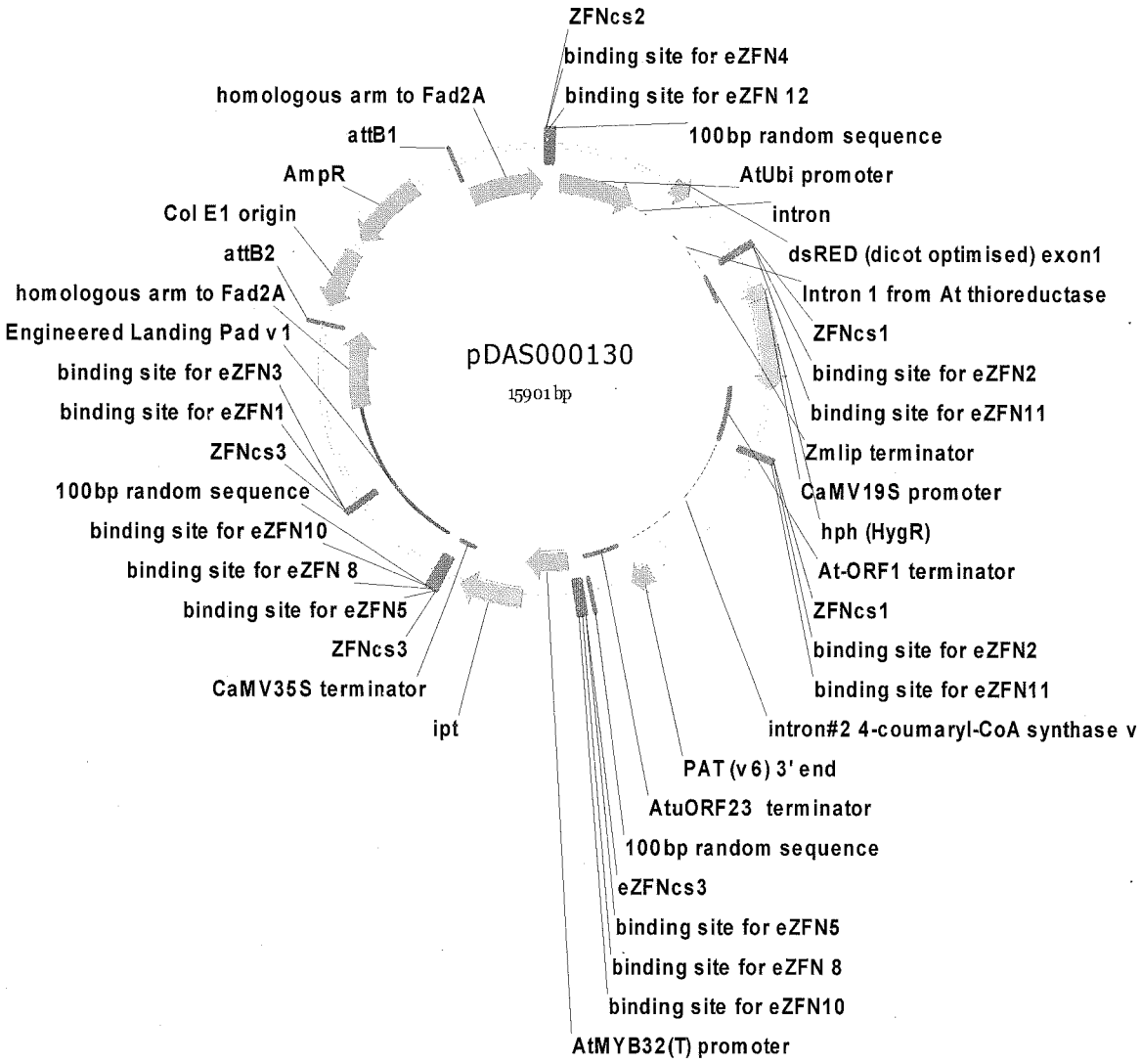

pDAS000130
15901 bp
ZFNcs2
binding site for eZFN4
binding site for eZFN 12
100bp random sequence
AtUbi promoter
intron
dsRED (dicot optimised) exon1
Intron 1 from At thioreductase
ZFNcs1
binding site for eZFN2
binding site for eZFN11
Zmlip terminator
CaMV19S promoter
hph (HygR)
At-ORF1 terminator
ZFNcs1
binding site for eZFN2
binding site for eZFN11
intron#2 4-coumaryl-CoA synthase v
PAT (v6) 3' end
AtuORF23 terminator
100bp random sequence
eZFNcs3
binding site for eZFN5
binding site for eZFN 8
binding site for eZFN10
AtMYB32(T) promoter
ipt
CaMV35S terminator
ZFNcs3
binding site for eZFN5
binding site for eZFN 8
100bp random sequence
binding site for eZFN10
ZFNcs3
binding site for eZFN1
binding site for eZFN3
Engineered Landing Pad v1
homologous arm to Fad2A
attB2
Col E1 origin
AmpR
attB1
homologous arm to Fad2A

Figure 6

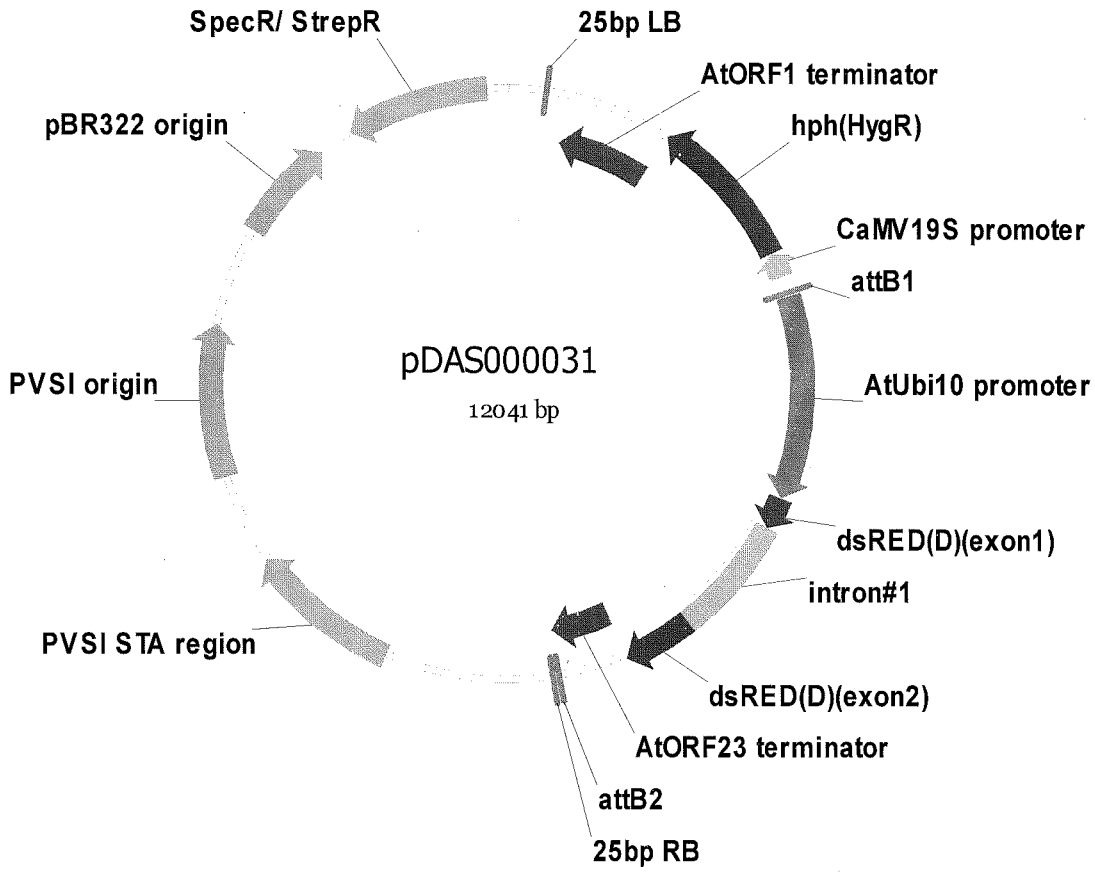
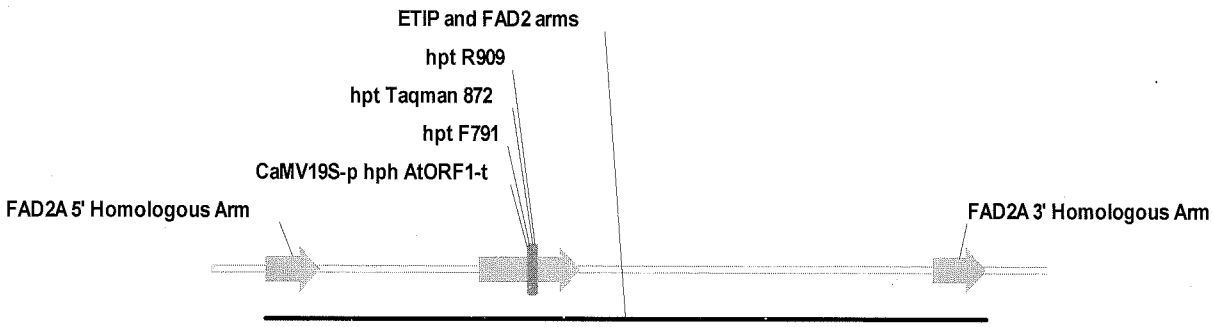

SpecR/ StrepR
25bp LB
AtORF1 terminator
pBR322 origin
hph(HygR)
CaMV19S promoter
attB1
pDAS000031
12041 bp
PVSI origin
AtUbi10 promoter
dsRED(D)(exon1)
intron#1
PVSI STA region
dsRED(D)(exon2)
AtORF23 terminator
attB2
25bp RB

Figure 7



FAD2A locus with perfect ETIP integration -hph qPCR primer binding sites

15612 bp

Figure 8

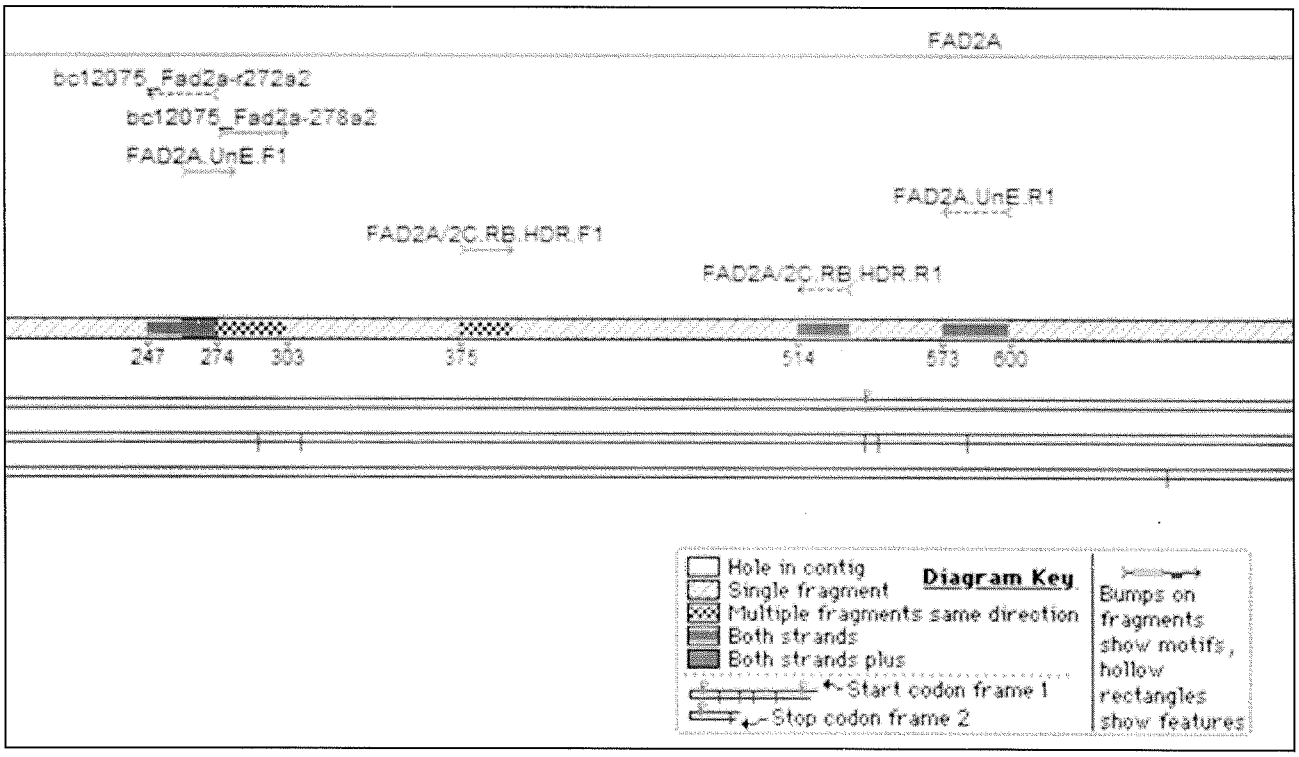


Figure 9

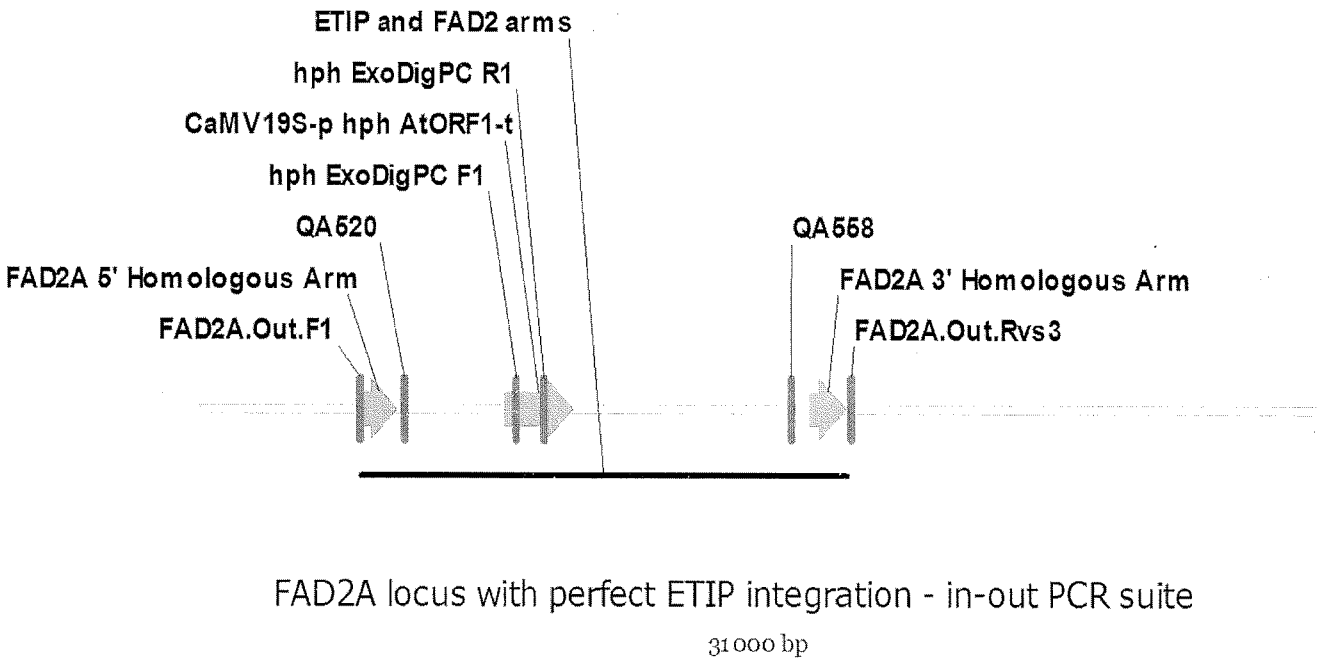


Figure 10

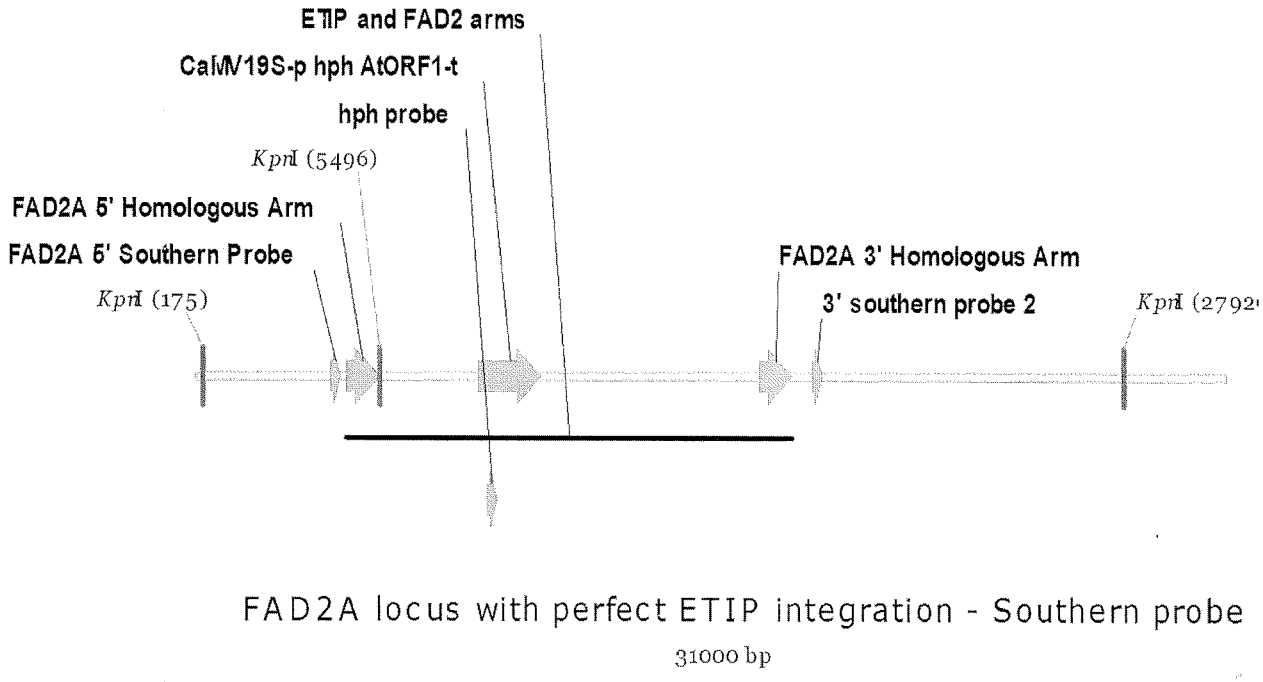


Figure 11

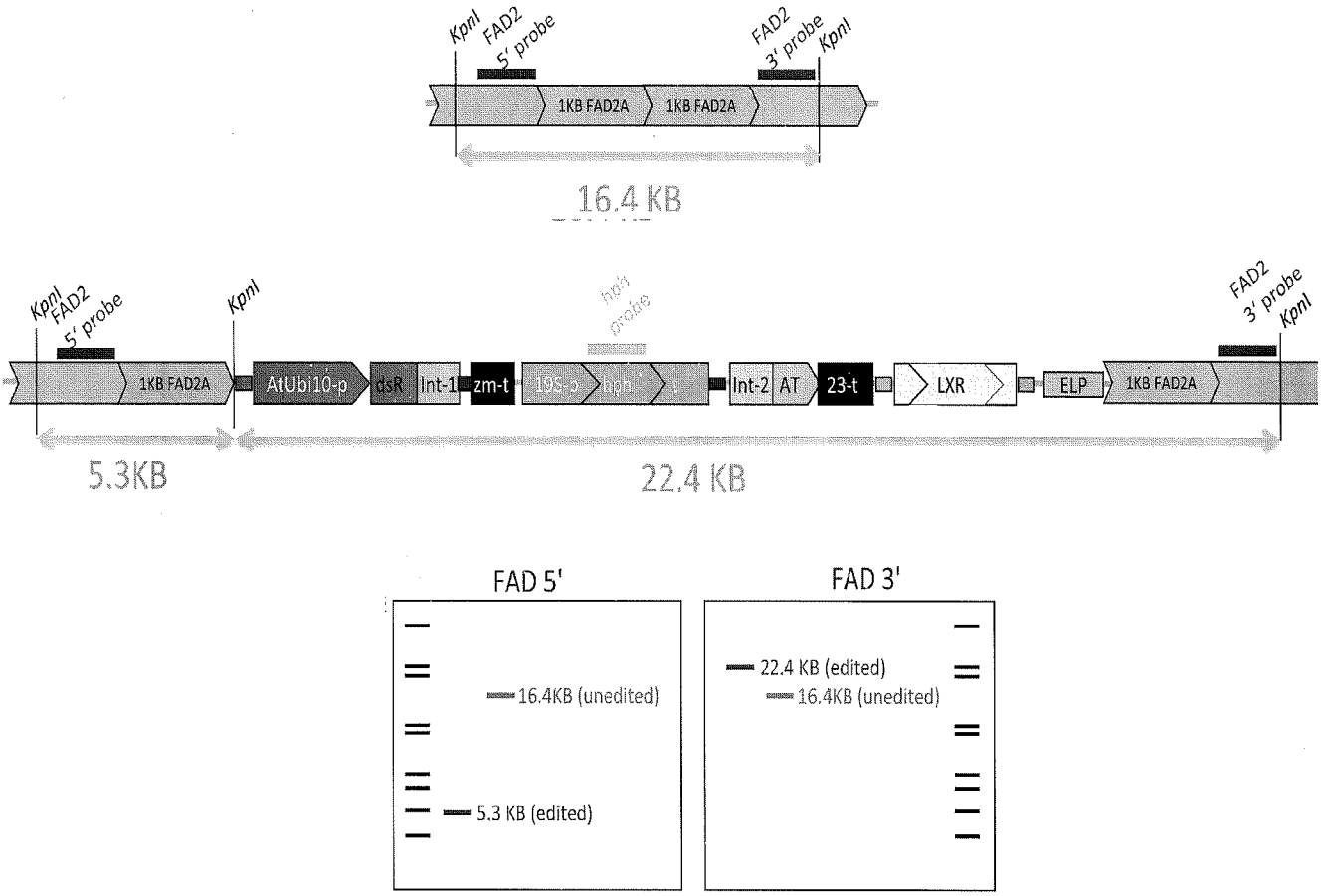

KpnI
FAD2 5' probe
FAD2 3' probe
KpnI
1KB FAD2A
1KB FAD2A
16.4 KB
KpnI
FAD2 5' probe
KpnI
hph probe
FAD2 3' probe
KpnI
1KB FAD2A
AtUbi10-p
dsR
Int-1
zm-t
I9S-p
hph
Int-2
AT
23-t
LXR
ELP
1KB FAD2A
5.3KB
22.4 KB
FAD 5'
16.4KB (unedited)
5.3 KB (edited)
FAD 3'
22.4 KB (edited)
16.4KB (unedited)

Figure 12

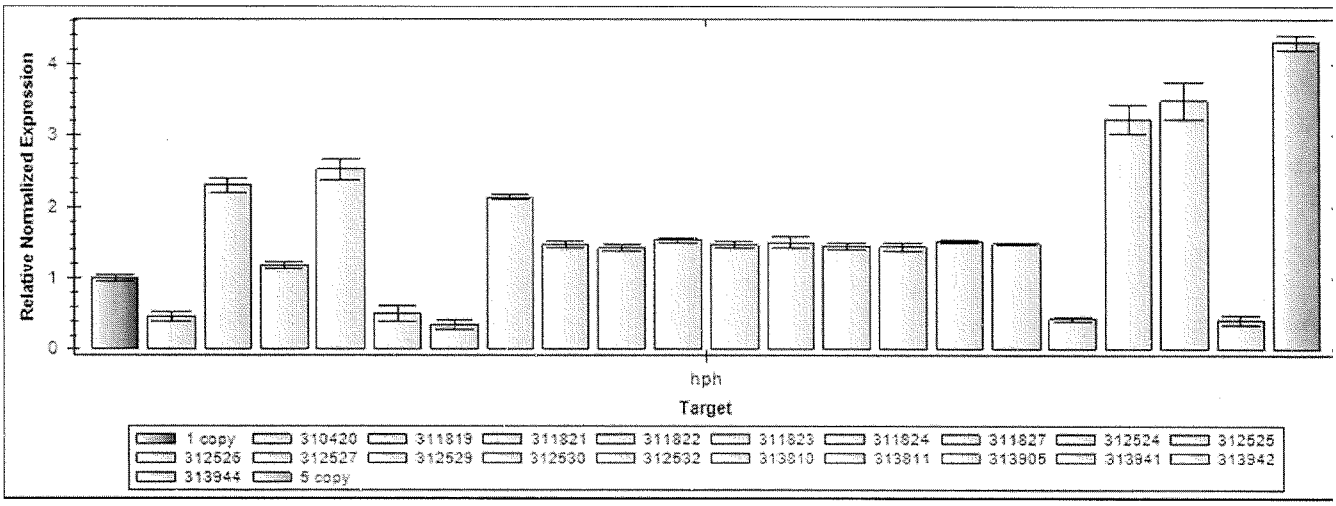


Figure 13

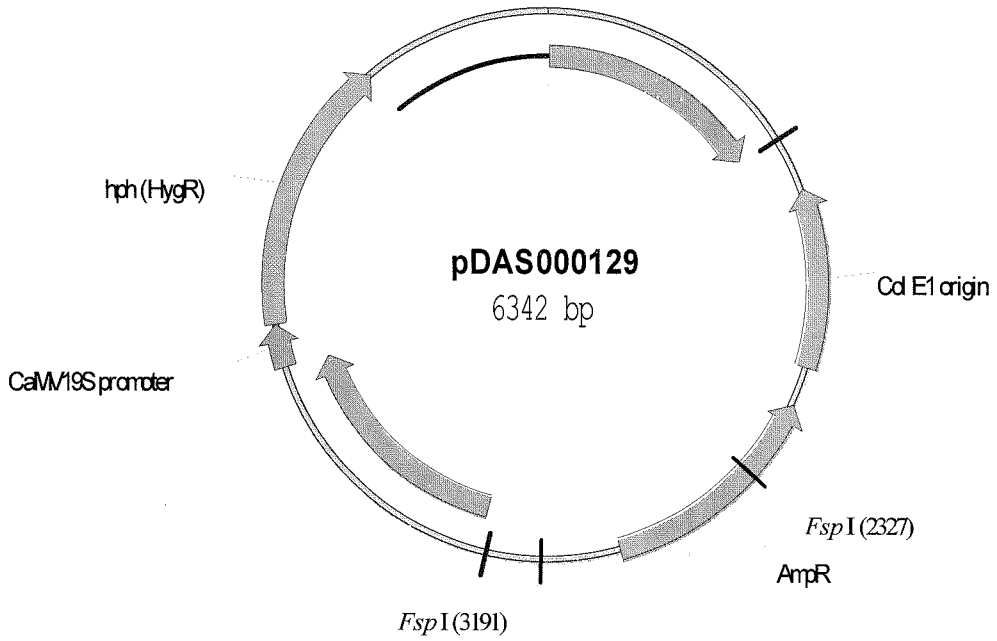


Figure 14

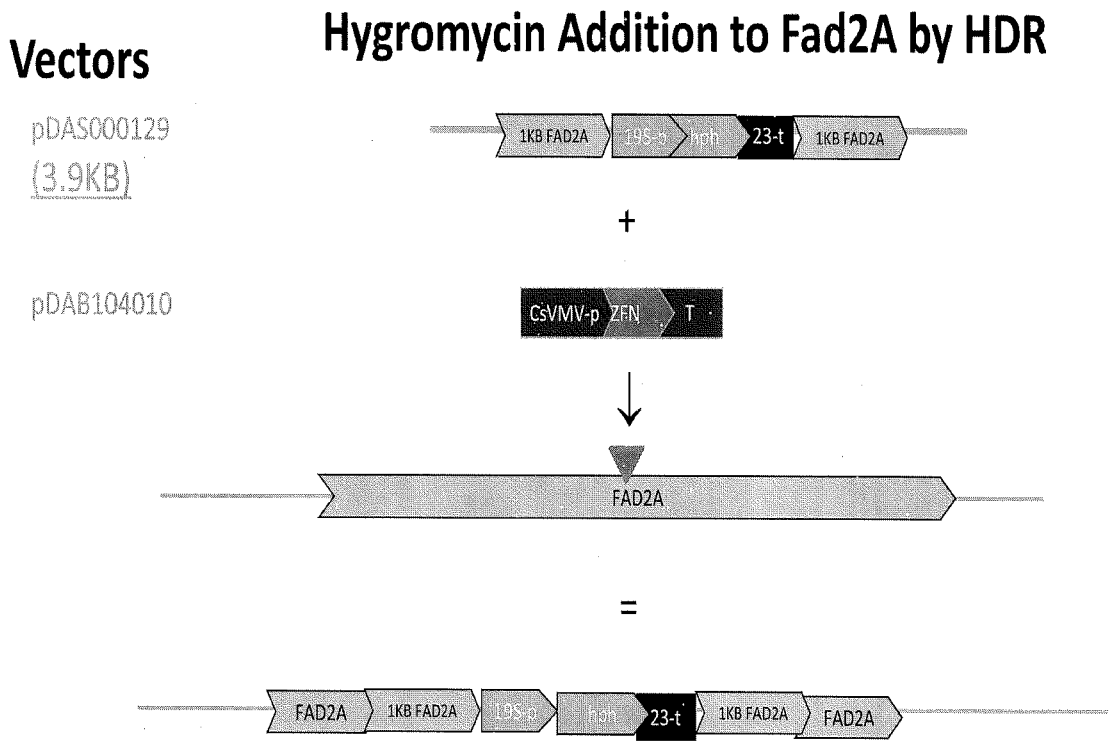

Vectors
Hygromycin Addition to Fad2A by HDR
pDAS000129
(3.9KB)
1KB FAD2A
23-t
1KB FAD2A
+
pDAB104010
CsVMV-p
ZFN
T
FAD2A
=
FAD2A
1KB FAD2A
23-t
1KB FAD2A
FAD2A

Figure 15

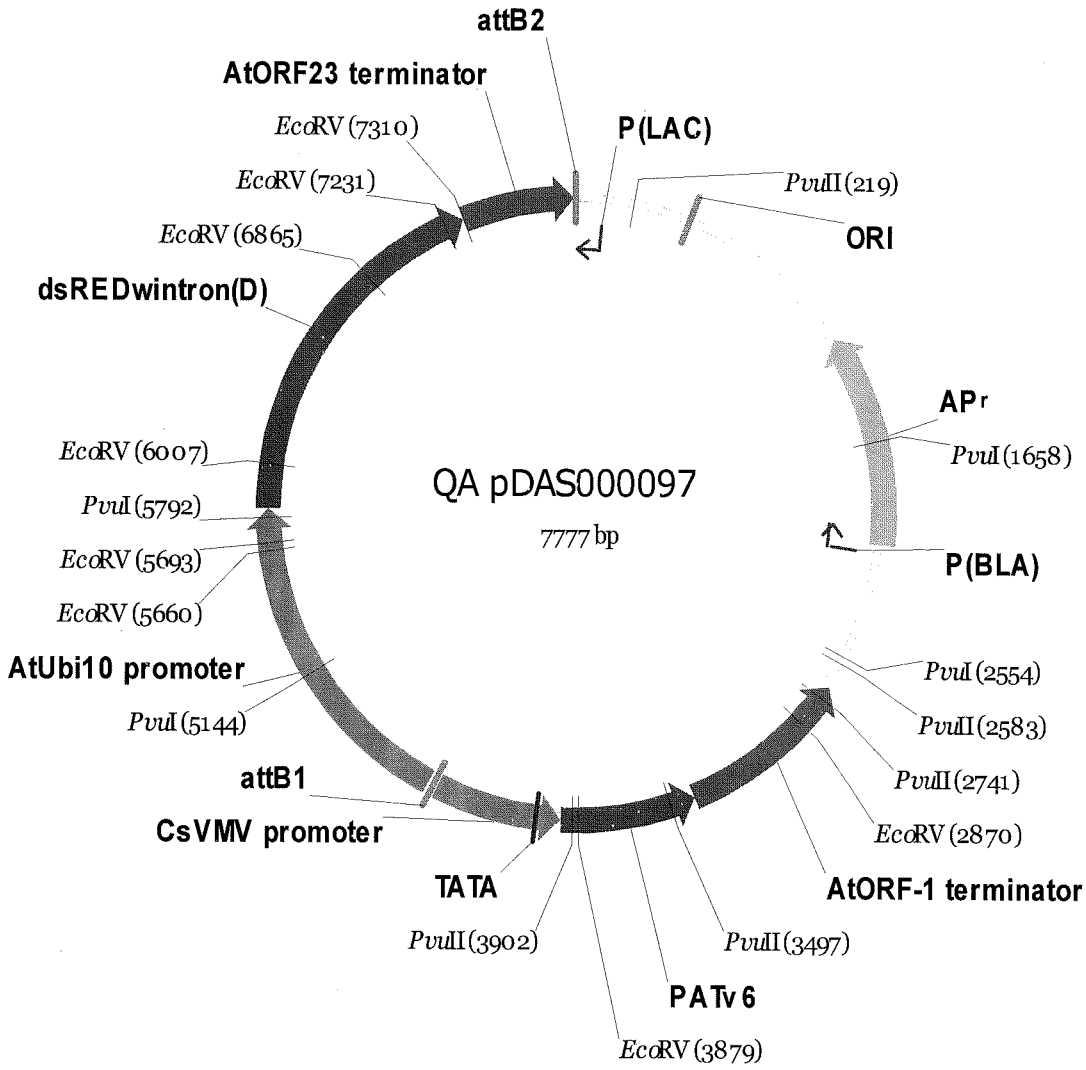

attB2
AtORF23 terminator
EcoRV (7310)
P(LAC)
EcoRV (7231)
PvuII (219)
EcoRV (6865)
ORI
dsREDwintron(D)
APr
EcoRV (6007)
PvuI (1658)
QA pDAS000097
PvuI (5792)
7777 bp
EcoRV (5693)
P(BLA)
EcoRV (5660)
AtUbi10 promoter
PvuI (2554)
PvuI (5144)
PvuII (2583)
attB1
PvuII (2741)
CsVMV promoter
EcoRV (2870)
TATA
AtORF-1 terminator
PvuII (3902)
PvuII (3497)
PATv6
EcoRV (3879)

Figure 16

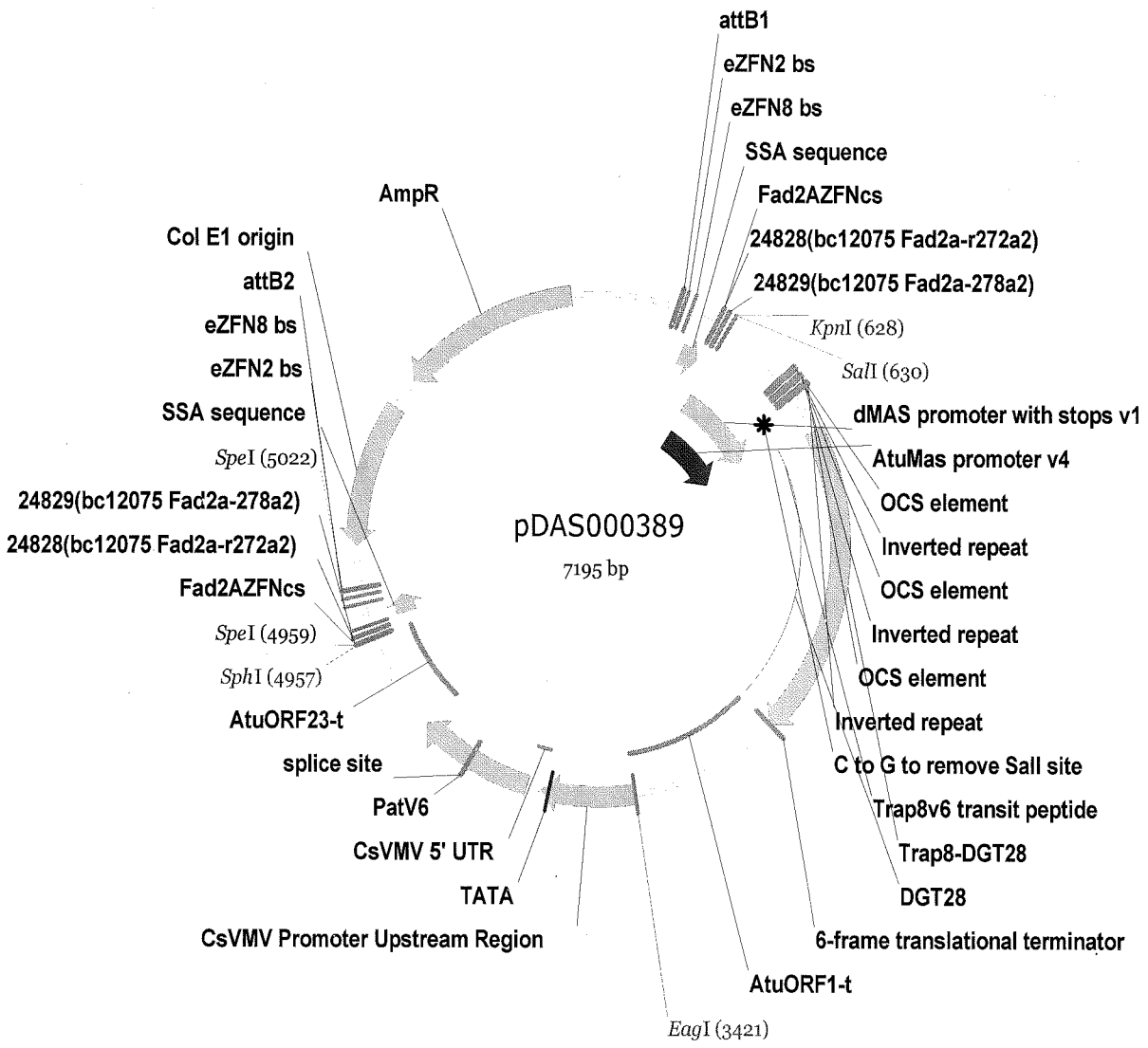

attB1
eZFN2 bs
eZFN8 bs
SSA sequence
Fad2AZFNcs
24828(bc12075 Fad2a-r272a2)
24829(bc12075 Fad2a-278a2)
KpnI (628)
SalI (630)
dMAS promoter with stops v1
AtuMas promoter v4
OCS element
Inverted repeat
OCS element
Inverted repeat
OCS element
Inverted repeat
C to G to remove SalI site
Trap8v6 transit peptide
Trap8-DGT28
DGT28
6-frame translational terminator
AtuORF1-t
EagI (3421)
CsVMV Promoter Upstream Region
TATA
CsVMV 5' UTR
PatV6
splice site
AtuORF23-t
SphI (4957)
SpeI (4959)
Fad2AZFNcs
24828(bc12075 Fad2a-r272a2)
24829(bc12075 Fad2a-278a2)
SpeI (5022)
SSA sequence
eZFN2 bs
eZFN8 bs
attB2
Col E1 origin
AmpR
pDAS000389
7195 bp

Figure 17

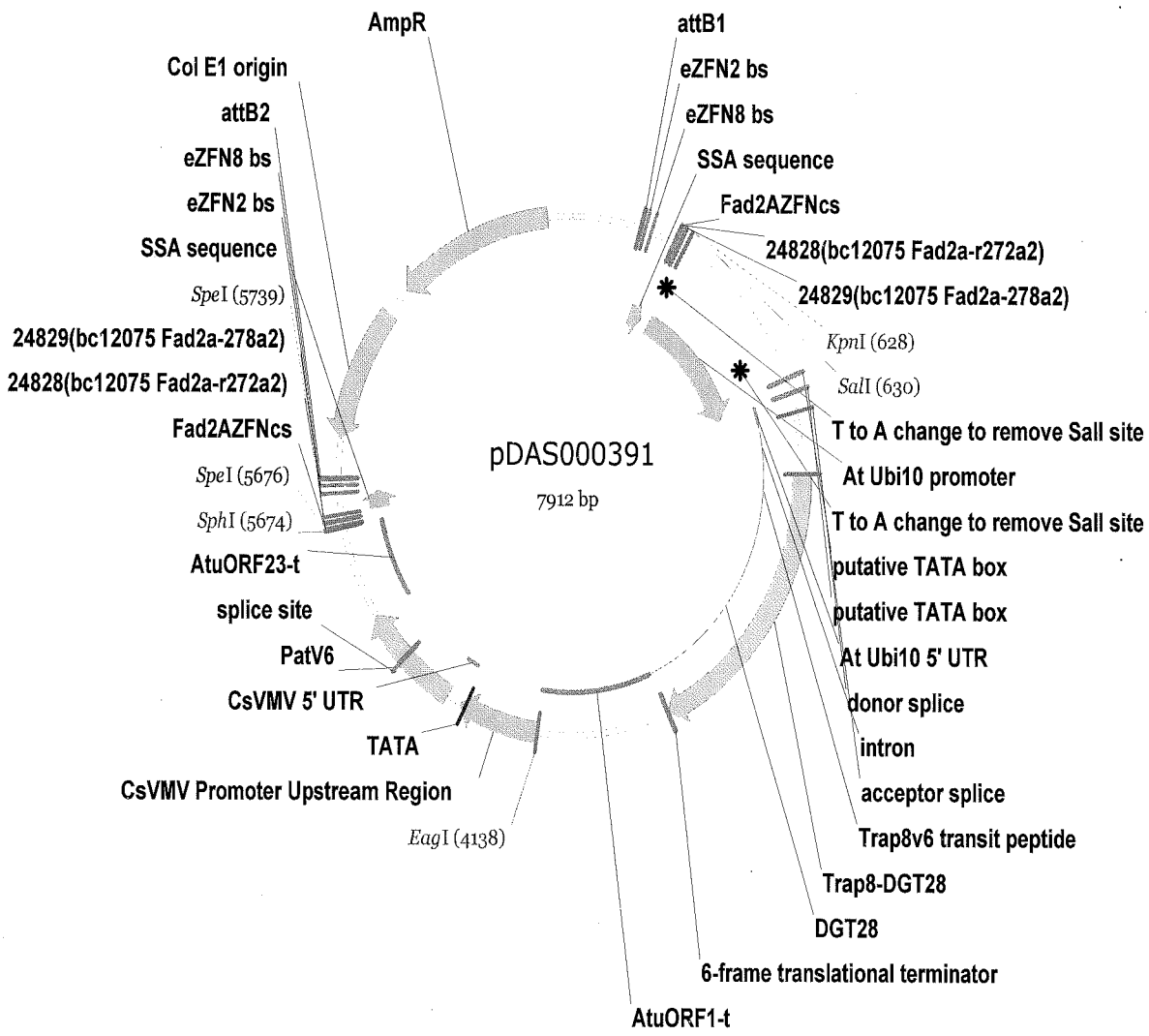

AmpR
Col E1 origin
attB2
eZFN8 bs
eZFN2 bs
SSA sequence
SpeI (5739)
24829(bc12075 Fad2a-278a2)
24828(bc12075 Fad2a-r272a2)
Fad2AZFNcs
SpeI (5676)
SphI (5674)
AtuORF23-t
splice site
PatV6
CsVMV 5' UTR
TATA
CsVMV Promoter Upstream Region
EagI (4138)
pDAS000391
7912 bp
attB1
eZFN2 bs
eZFN8 bs
SSA sequence
Fad2AZFNcs
24828(bc12075 Fad2a-r272a2)
24829(bc12075 Fad2a-278a2)
KpnI (628)
SalI (630)
T to A change to remove SalI site
At Ubi10 promoter
T to A change to remove SalI site
putative TATA box
putative TATA box
At Ubi10 5' UTR
donor splice
intron
acceptor splice
Trap8v6 transit peptide
Trap8-DGT28
DGT28
6-frame translational terminator
AtuORF1-t

Figure 18

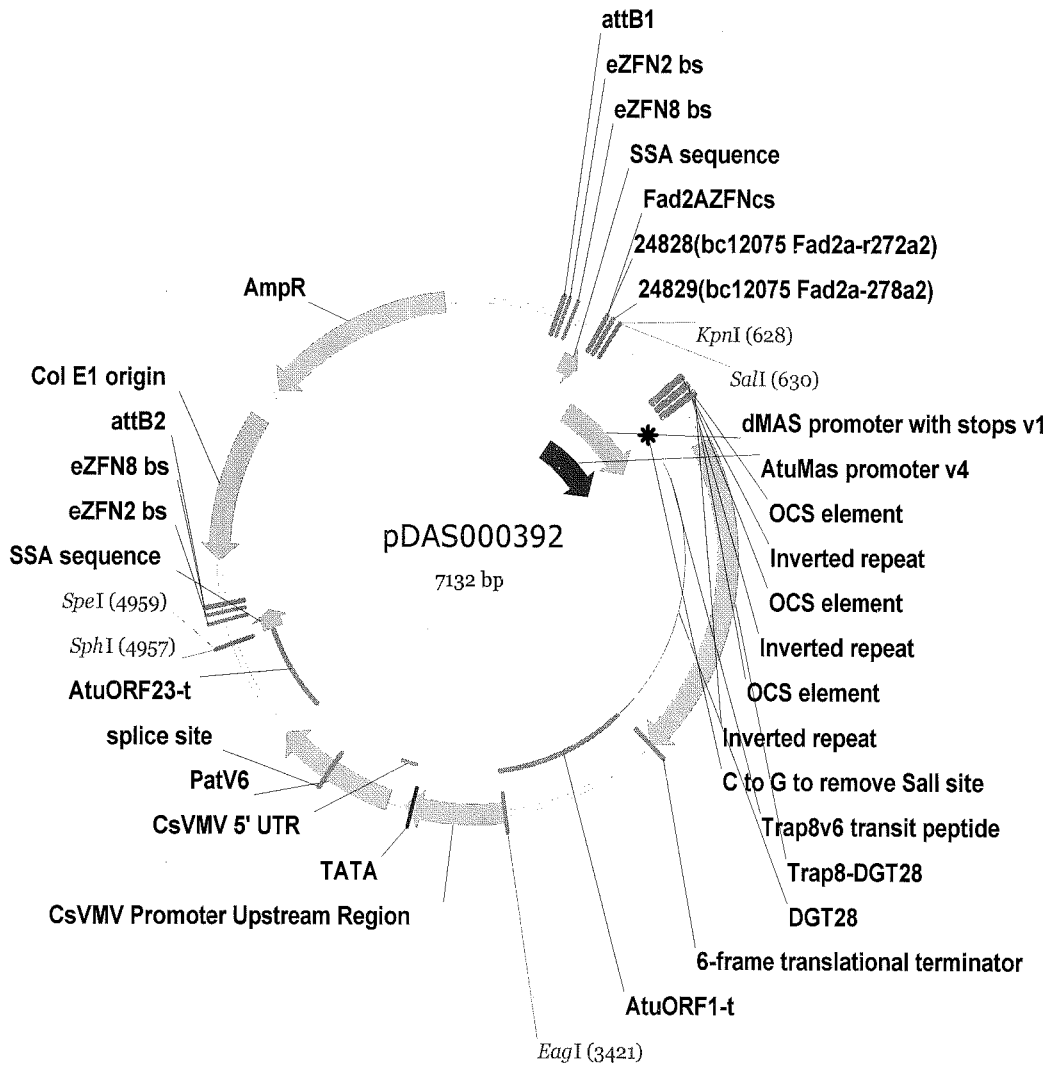

attB1
eZFN2 bs
eZFN8 bs
SSA sequence
Fad2AZFNcs
24828(bc12075 Fad2a-r272a2)
24829(bc12075 Fad2a-278a2)
KpnI (628)
SalI (630)
dMAS promoter with stops v1
AtuMas promoter v4
OCS element
Inverted repeat
OCS element
Inverted repeat
OCS element
Inverted repeat
C to G to remove SalI site
Trap8v6 transit peptide
Trap8-DGT28
DGT28
6-frame translational terminator
AtuORF1-t
EagI (3421)
CsVMV Promoter Upstream Region
TATA
CsVMV 5' UTR
PatV6
splice site
AtuORF23-t
SphI (4957)
SpeI (4959)
SSA sequence
eZFN2 bs
eZFN8 bs
attB2
Col E1 origin
AmpR
pDAS000392
7132 bp

Figure 19

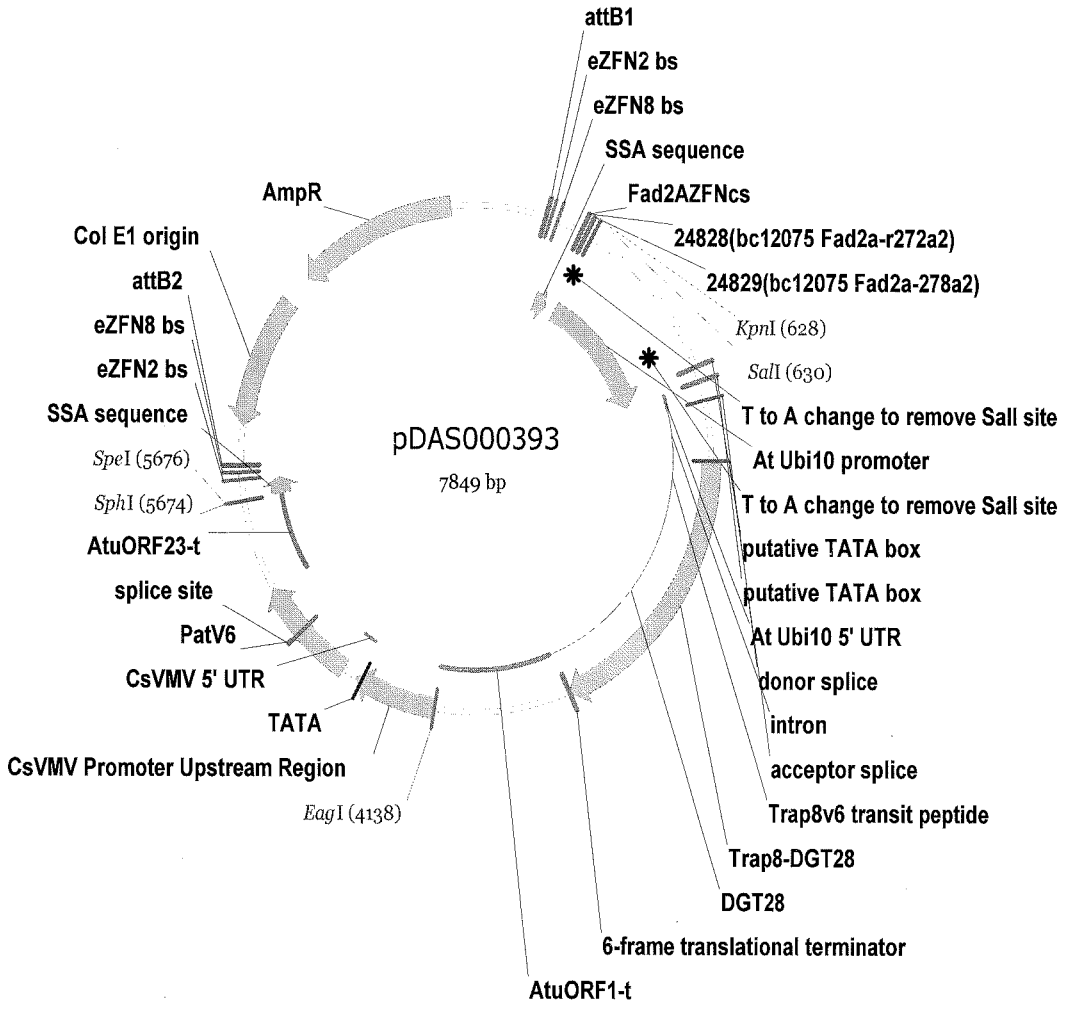

attB1
eZFN2 bs
eZFN8 bs
SSA sequence
Fad2AZFNcs
24828(bc12075 Fad2a-r272a2)
24829(bc12075 Fad2a-278a2)
KpnI (628)
SalI (630)
T to A change to remove SalI site
At Ubi10 promoter
T to A change to remove SalI site
putative TATA box
putative TATA box
At Ubi10 5' UTR
donor splice
intron
acceptor splice
Trap8v6 transit peptide
Trap8-DGT28
DGT28
6-frame translational terminator
AtuORF1-t
EagI (4138)
CsVMV Promoter Upstream Region
TATA
CsVMV 5' UTR
PatV6
splice site
AtuORF23-t
SphI (5674)
SpeI (5676)
SSA sequence
eZFN2 bs
eZFN8 bs
attB2
Col E1 origin
AmpR
pDAS000393
7849 bp

Figure 20

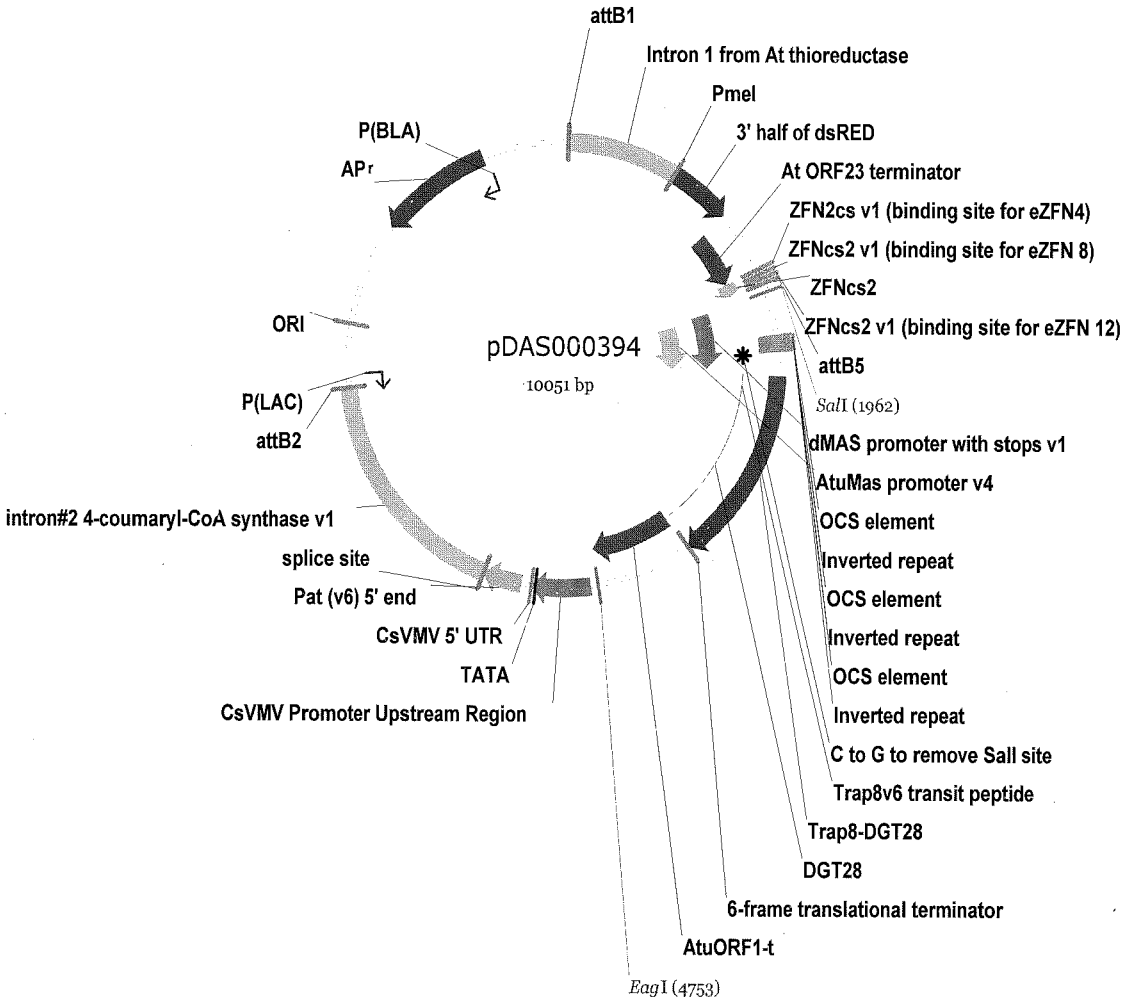

attB1
Intron 1 from At thioreductase
PmeI
3' half of dsRED
At ORF23 terminator
ZFN2cs v1 (binding site for eZFN4)
ZFNcs2 v1 (binding site for eZFN 8)
ZFNcs2
ZFNcs2 v1 (binding site for eZFN 12)
attB5
SalI (1962)
dMAS promoter with stops v1
AtuMas promoter v4
OCS element
Inverted repeat
OCS element
Inverted repeat
OCS element
Inverted repeat
C to G to remove SalI site
Trap8v6 transit peptide
Trap8-DGT28
DGT28
6-frame translational terminator
AtuORF1-t
EagI (4753)
CsVMV Promoter Upstream Region
TATA
CsVMV 5' UTR
Pat (v6) 5' end
splice site
intron#2 4-coumaryl-CoA synthase v1
attB2
P(LAC)
ORI
APr
P(BLA)
pDAS000394
10051 bp

Figure 21

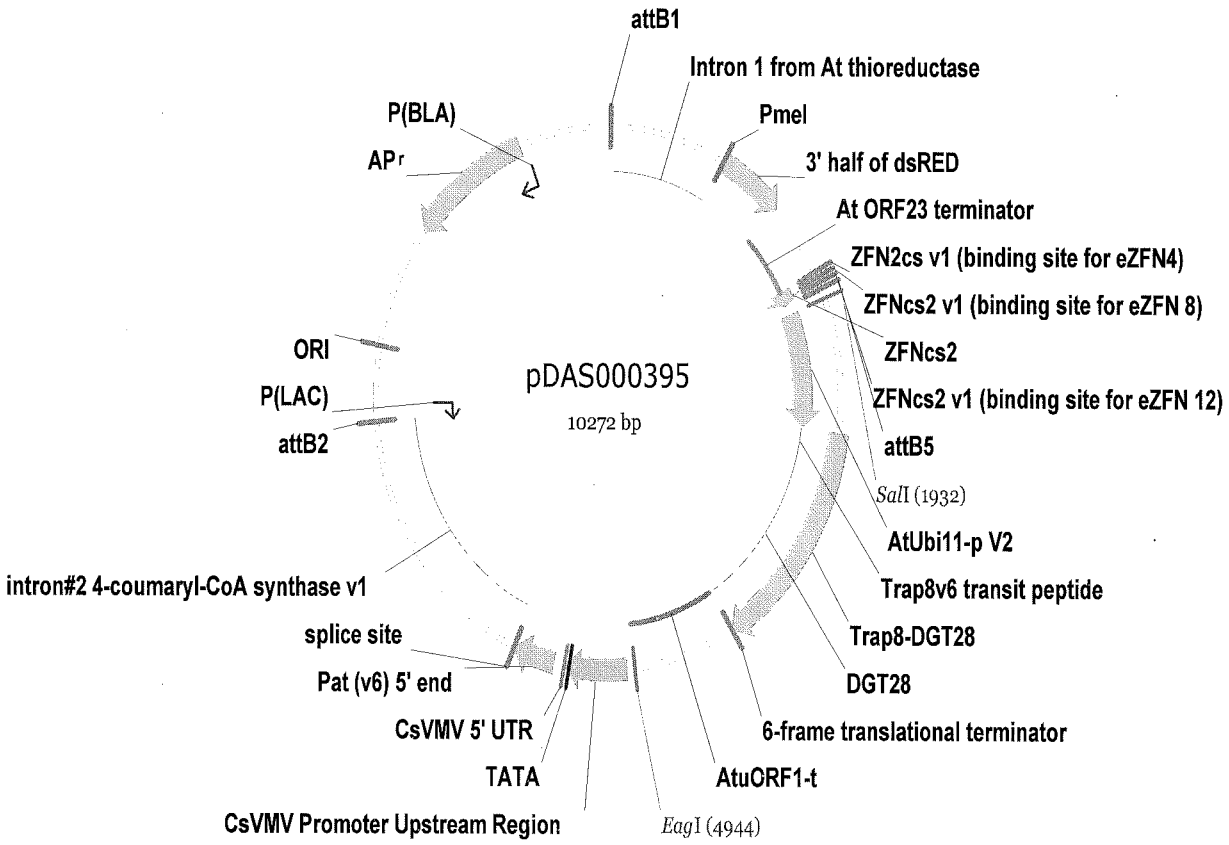

attB1
Intron 1 from At thioreductase
P(BLA)
PmeI
APr
3' half of dsRED
At ORF23 terminator
ZFN2cs v1 (binding site for eZFN4)
ZFNcs2 v1 (binding site for eZFN 8)
ZFNcs2
ORI
pDAS000395
10272 bp
P(LAC)
ZFNcs2 v1 (binding site for eZFN 12)
attB2
attB5
SalI (1932)
AtUbi11-p V2
Trap8v6 transit peptide
intron#2 4-coumaryl-CoA synthase v1
Trap8-DGT28
splice site
DGT28
Pat (v6) 5' end
6-frame translational terminator
CsVMV 5' UTR
TATA
AtuORF1-t
CsVMV Promoter Upstream Region
EagI (4944)

Figure 22

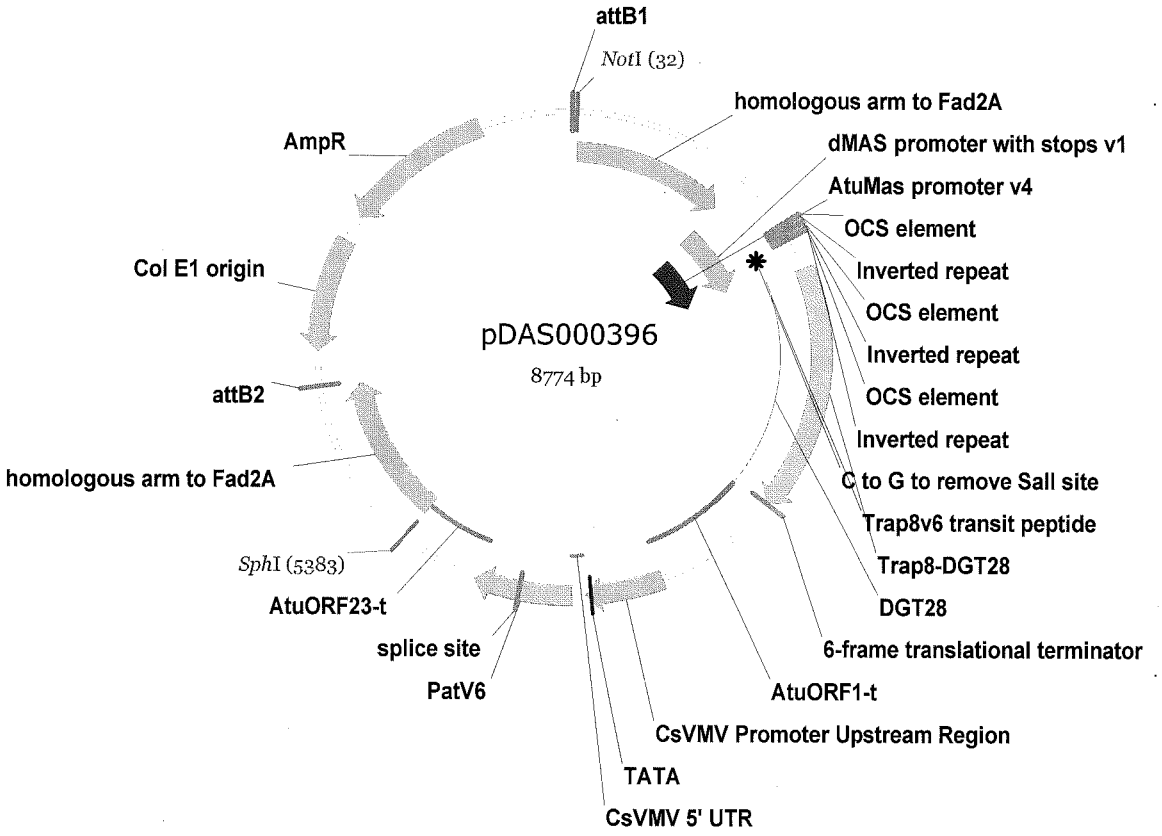

attB1
NotI (32)
homologous arm to Fad2A
dMAS promoter with stops v1
AtuMas promoter v4
OCS element
Inverted repeat
OCS element
Inverted repeat
OCS element
Inverted repeat
C to G to remove SalI site
Trap8v6 transit peptide
Trap8-DGT28
DGT28
6-frame translational terminator
AtuORF1-t
CsVMV Promoter Upstream Region
TATA
CsVMV 5' UTR
PatV6
splice site
AtuORF23-t
SphI (5383)
homologous arm to Fad2A
attB2
Col E1 origin
AmpR
pDAS000396
8774 bp

Figure 23

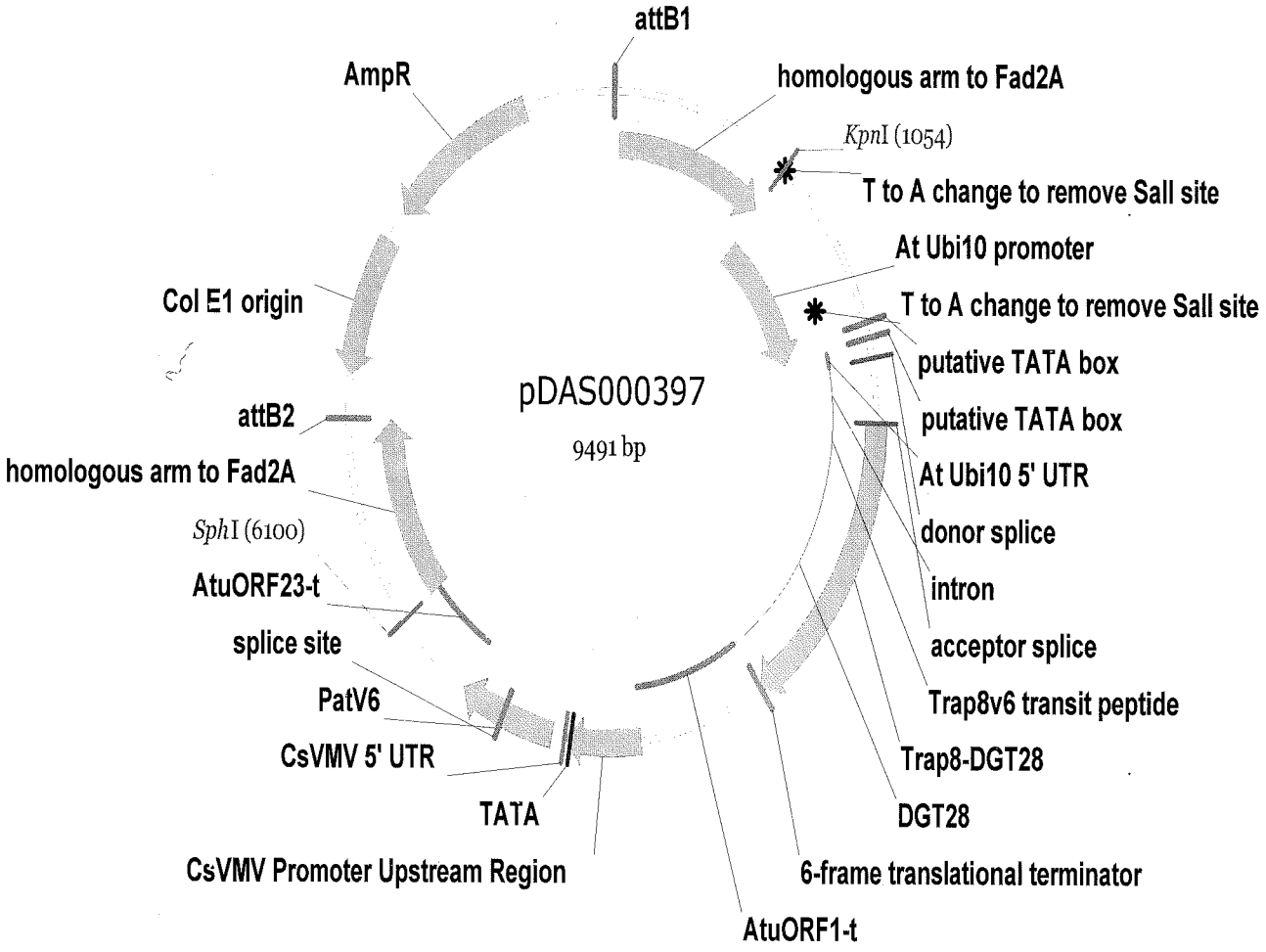

attB1
AmpR
homologous arm to Fad2A
KpnI (1054)
T to A change to remove SalI site
At Ubi10 promoter
Col E1 origin
T to A change to remove SalI site
putative TATA box
pDAS000397
9491 bp
putative TATA box
attB2
At Ubi10 5' UTR
homologous arm to Fad2A
donor splice
SphI (6100)
intron
AtuORF23-t
acceptor splice
splice site
Trap8v6 transit peptide
PatV6
Trap8-DGT28
CsVMV 5' UTR
DGT28
TATA
6-frame translational terminator
CsVMV Promoter Upstream Region
AtuORF1-t

Figure 24

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 March 2014 (13.03.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/039692 A3

(51) International Patent Classification:
A01H 5/00 (2006.01) *C12N 15/82* (2006.01)
C07H 21/02 (2006.01) *C12N 5/04* (2006.01)
C07H 21/04 (2006.01)

(21) International Application Number:
PCT/US2013/058283

(22) International Filing Date:
5 September 2013 (05.09.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/697,886 7 September 2012 (07.09.2012) US

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

(72) Inventors: **COGAN, Noel**; c/o Dow Agrosciences LLC, 9330 Zionsville Road, Indianapolis, IN 46268-1054 (US). **FORSTER, John**; c/o Dow Agrosciences LLC, 9330 Zionsville Road, Indianapolis, IN 46268-1054 (US). **HAYDEN, Matthew**; c/o Dow Agrosciences LLC, 9330 Zionsville Road, Indianapolis, IN 46268-1054 (US). **SAWBRIDGE, Tim**; c/o Dow Agrosciences LLC, 9330 Zionsville Road, Indianapolis, IN 46268-1054 (US). **SPANGENBERG, German**; c/o Dow Agrosciences LLC, 9330 Zionsville Road, Indianapolis, IN 46268-1054 (US). **WEBB, Steven, R.**; c/o Dow Agrosciences LLC, 9330 Zionsville Road, Indianapolis, IN 46268-1054 (US). **GUPTA, Manju**; c/o Dow Agrosciences LLC, 9330 Zionsville Road, Indianapolis, IN 46268-1054 (US). **AINLEY, William, Michael**; c/o Dow Agrosciences LLC, 9330 Zionsville Road, Indianapolis, IN 46268-1054 (US). **HENRY, Matthew, J.**; c/o Dow Agrosciences LLC, 9330 Zionsville Road, Indianapolis, IN 46268-1054 (US). **MILLER, Jeffrey, C.**; c/o Sangamo Biosciences, Inc., Point Richmond Tech Center, 501 Canal Blvd., Suite A100, Richmond, CA 94804 (US). **GUSCHIN, Dmitry, Y.**; c/o Sangamo Biosciences, Inc., Point Richmond Tech Center, 501 Canal Blvd., Suite A100, Richmond, CA 94804 (US).

(74) Agent: **PASTERNAK, Dahna, S.**; Pasternak Patent Law, 1900 Embarcadero Rd., Suite 211, Palo Alto, CA 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, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM,

[Continued on next page]

(54) Title: FAD2 PERFORMANCE LOCI AND CORRESPONDING TARGET SITE SPECIFIC BINDING PROTEINS CAPABLE OF INDUCING TARGETED BREAKS

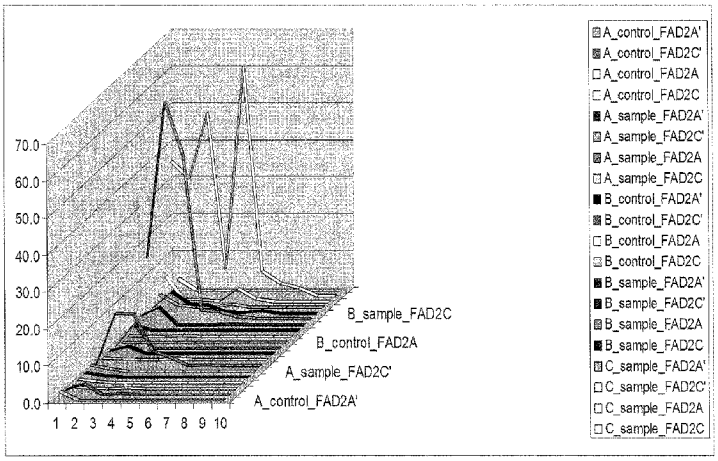


Figure 4

(57) Abstract: A method of gene editing or gene stacking within a FAD2 loci by cleaving, in a site directed manner, a location in a FAD2 gene in a cell, to generate a break in the FAD2 gene and then ligating into the break a nucleic acid molecule associated with one or more traits of interest is disclosed.

WO 2014/039692 A3

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))*

(88) **Date of publication of the international search report**:
16 July 2015

INTERNATIONAL SEARCH REPORT

PCT/US2013/058283 28.04.2014

International application No.
PCT/US 13/58283

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - A01H 5/00, C07H 21/02, C07H 21/04, C12N 15/82, C12N 5/04 (2014.01)
USPC - 800/312
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
USPC: 800/312

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC: 536/23.1, 435/468, 435/419 (keyword limited; terms below)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
PatBase, Google Patents/Scholar
Search Terms Used: FAD2, FAD2A, FAD2C, zinc finger nuclease, meganuclease, TALE DNA-binding, double strand break

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2011/0167521 A1 (DeKelver et al.) 07 July 2011 (07.07.2011) para [0011], [0018], [0083]	1-2
X -- Y	US 2011/0301073 A1 (Gregroy et al.) 08 December 2011 (08.12.2011) para [0033], [0237]	1-2 -- 3
Y	US 2011/0293785 A1 (Franklin et al.) 01 December 2011 (01.12.2011) para [0005], [0012], [0677], [0680]	3
X	US 2007/0059795 A1 (Moore et al.) 15 March 2007 (15.03.2007) para [0254]	1-3
A,P	US 2013/0326645 A1 (Cost et al.) 05 December 2013 (05.12.2013) para [0347]	1-3
A	US 2011/0123509 A1 (Jantz et al.) 26 May 2011 (26.05.20.11) para [0575]	1-3
A	Curtin et al. 'Targeted Mutagenesis of Duplicated Genes in Soybean with Zinc-Finger Nucleases1' Plant Physiology, June 2011, Vol. 156, pp. 466?473, whole doc.	1-3

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

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
07 March 2014 (07.03.2014)	28 APR 2014
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT PCT/US2013/058283 28.04.2014

International application No.
PCT/US 13/58283

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 4-19
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Group I: claims 1-3, drawn to a method for modifying the genome of a cell, the method comprising:
cleaving, in a site specific matter, a target site in a FAD2 gene in a cell, to thereby generate a break in the F AD2 gene; wherein the F AD2 gene is modified following cleavage.

Group II+: claims 20-21, drawn to nucleic acid sequences comprising zinc finger nuclease target sites and amino acid sequences comprising recognition helix regions of zinc finger proteins. The first invention is restricted to the first named zinc finger nuclease target site, SEQ ID NO: 22; and the first named zinc-finger protein, ZFP 24811, that binds to SEQ ID NO: 22, comprising recognition helix domains of SEQ ID NOs: 119, 99, 121-123. Group II+ will be searched to the extent that it reads on the above limitation. It is believed that claims 20-21 read on this first named invention. Applicants must indicate, if applicable, the claims which read on the first named invention if different than what was indicated above for this group. Failure to clearly identify how any paid additional invention fees are to be applied to the "+" group(s) will result in only the first claimed invention to be searched/examined. An exemplary election would be: target site SEQ ID NO: 23 and ZFP 24814 comprising SEQ ID NOs: 124, 125, 108, 110, 104.
- Please see extra sheet for continuation -

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-3

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

International application No.
PCT/US 13/58283

Box NO III. Observations where unity of invention is lacking

The inventions listed as Groups I and II+ do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Special Technical Features
The inventions of Group I do not include the special technical feature of nucleic acid sequences comprising zinc finger nuclease target sites and amino acid sequences comprising recognition helix regions of zinc finger proteins, as required by Group II+.

The inventions of Group II+ do not include the special technical feature of a method for generating a break in the FAD2 gene, as required by Group I.

Each invention of Group II+ includes the special technical feature of a unique amino acid and nucleic acid having a SEQ ID NO, which is not required by any other invention.

Common Technical Features
The inventions of Groups I and II+ do not share any common technical feature. The inventions of Group II+ share the technical feature of zinc finger nuclease target sites and a zinc finger nuclease comprises from three to six zinc finger domains comprising a recognition helix regions. However, this shared technical feature does not represent a contribution over prior art as being anticipated by the article entitled 'Targeted Mutagenesis of Duplicated Genes in Soybean with Zinc-Finger Nucleases' by Curtin et al. (Plant Physiology, June 2011, Vol. 156, pp. 466-473) (hereinafter 'Curtin'). Curtin teaches nucleic acid sequences of zinc finger nuclease target sites and a zinc finger nuclease comprises from three to six zinc finger domains comprising amino acid sequences of recognition helix regions (Table1, Fig. 1, 6 zinc finger domains) that are useful in generating heritable transmission of targeted mutagenesis in soybean (abstract). As said technical feature was known in the art at the time of the invention, this cannot be considered a special technical feature that would otherwise unify the groups.

Groups I and II+ therefore lack unity under PCT Rule 13 because they do not share a same or corresponding special technical feature.

摘要

公开了一种在FAD2基因座中进行基因编辑或基因堆叠的方法，其是通过以定点的方式剪切细胞FAD2基因中的一个位置，从而在FAD2基因中产生断裂(break)，然后在该断裂中接入与一种或多种感兴趣性状相关的核酸分子。