



(12) 发明专利申请

(10) 申请公布号 CN 105025702 A

(43) 申请公布日 2015. 11. 04

(21) 申请号 201380058051. 3

(22) 申请日 2013. 09. 05

(30) 优先权数据

61/697, 886 2012. 09. 07 US

(85) PCT国际申请进入国家阶段日

2015. 05. 06

(86) PCT国际申请的申请数据

PCT/US2013/058283 2013. 09. 05

(87) PCT国际申请的公布数据

W02014/039692 EN 2014. 03. 13

(71) 申请人 美国陶氏益农公司

地址 美国印第安纳州

申请人 桑格摩生物科学股份有限公司

(72) 发明人 N·科根 J·福斯特 M·海登

T·索布里奇 G·斯潘根贝格

S·R·韦布 M·古朴塔

W·M·安利 M·J·亨利

J·C·米勒 D·Y·古斯钦

(74) 专利代理机构 北京坤瑞律师事务所 11494

代理人 罗天乐

(51) Int. Cl.

A01H 5/00(2006. 01)

C07H 21/02(2006. 01)

C07H 21/04(2006. 01)

C12N 15/82(2006. 01)

C12N 5/04(2006. 01)

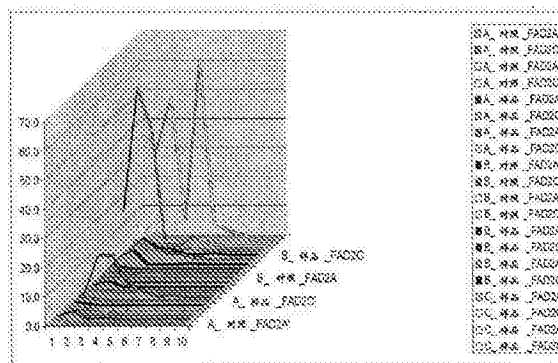
权利要求书2页 说明书68页 附图24页

(54) 发明名称

FAD2 性能基因座及相应的能够诱导靶向断裂的靶位点特异性结合蛋白

(57) 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。



1. 一种修饰细胞的基因组的方法,该方法包括:
以位点特异性方式剪切细胞中的 FAD2 基因中的靶位点,从而在该 FAD2 基因中产生断裂;
其中该 FAD2 基因在剪切之后被修饰。
2. 权利要求 1 的方法,其中该方法进一步包括向该断裂中整合感兴趣的核酸序列。
3. 根据权利要求 1 或权利要求 2 的方法,其中所述 FAD2 基因是 FAD2A, FAD2A', FAD2C 和 / 或 FAD2C' 基因。
4. 根据权利要求 1-3 中任一项的方法,其中位点特异性方式的剪切包括向细胞内引入融合蛋白或编码融合蛋白的多核苷酸,所述融合蛋白包含 DNA 结合结构域以及剪切结构域或剪切半结构域,其中该融合蛋白与所述靶位点特异性结合,并在靶位点处或其附近进行剪切从而产生断裂。
5. 根据权利要求 4 的方法,其中该 DNA 结合结构域选自下组:大范围核酸酶 DNA 结合结构域,亮氨酸拉链 DNA 结合结构域,转录激活因子样 (TAL) DNA 结合结构域, RNA 指导的 CRISPR-Cas9, 重组酶, 锌指蛋白 DNA 结合结构域, 和前述者的任何嵌合组合。
6. 根据权利要求 4 或权利要求 5 的方法,其中该剪切结构域或剪切半结构域选自下组:来自 IIS 型限制性核酸内切酶的剪切半结构域,来自 FokI 核酸内切酶的剪切半结构域,来自 StsI 核酸内切酶的剪切半结构域,和归巢核酸内切酶。
7. 根据权利要求 4-6 中任一项的方法,其中该融合蛋白是锌指核酸酶。
8. 根据权利要求 7 的方法,其中该锌指核酸酶包括 3-6 个锌指结构域,每个锌指结构域包括识别螺旋区,其中该锌指蛋白包括在表 3 的单一行中排序并显示的识别螺旋区。
9. 根据权利要求 1-8 中任一项的方法,其中位点特异性方式的剪切是对 FAD2A, FAD2A', FAD2C 和 / 或 FAD2C' 的一些拷贝但非所有拷贝特异性的。
10. 根据权利要求 1-9 中任一项的方法,其中该靶位点选自下组:SEQ ID NO:22-26, 和 SEQ ID NO:28-33 和 SEQ ID NO:35-38。
11. 根据权利要求 1-10 中任一项的方法,其中该细胞是植物、真菌、细菌或藻类细胞。
12. 根据权利要求 11 的方法,其中该植物细胞是单子叶植物细胞或双子叶植物细胞。
13. 根据权利要求 12 的方法,其中该植物细胞选自下组:芸苔属 (*Brassica* sp.), 欧洲油菜 (*Brassica napus*); 芜菁 (*Brassica rapa*); 芥菜 (*Brassica juencea*); 甘蓝 (*Brassica oleracea*); 黑芥 (*Brassica nigra*); 玉米属 (*Zea* sp.); 玉米 (*Zea mays*); 大豆属 (*Glycine* sp.); 大豆 (*Glycine max*); 小麦属 (*Triticum* sp); 普通小麦 (*Triticum aestivum*); 稻属 (*Oryza* sp.); 水稻 (*Oryza sativa*); *Triticae* sp.; *Triticae triticum*; 向日葵属 (*Heliantheae* sp.); 向日葵 (*Heliantheae helianthus*); 棉属 (*Gossypium* sp.); 陆地棉 (*Gossypium hirsutum*); 和大麦 (*Hordeum vulgare*)。
14. 根据权利要求 2-13 中任一项的方法,其中感兴趣的核酸序列选自下组:包含 DNA 结合结构域结合靶位点的序列,一个或多个杀虫剂抗性基因,一个或多个除草剂抗性基因,一个或多个氮使用效率基因,一个或多个水使用效率基因,一个或多个营养品质基因,一个或多个 DNA 结合基因,一个或多个可选择标志物基因,及其组合。
15. 一种细胞、种子或植物,其包括根据权利要求 1-14 中任一项的方法修饰的细胞。
16. 根据权利要求 15 的细胞、种子或植物,其中该细胞、种子或植物是转基因的细胞、

种子或植物,其包含整合到 FAD2A, FAD2A', FAD2C 和 / 或 FAD2C' 基因的一个或多个拷贝中的感兴趣核苷酸序列。

17. 权利要求 16 的细胞、种子或植物,其中所述核苷酸序列对所述细胞而言是异源或同源的。

18. 权利要求 14 的细胞、种子或植物,其中该同源序列包含至少一个单核苷酸多态性。

19. 根据权利要求 16-18 中任一项的细胞、种子或植物,其中该核酸序列整合到选自下组的靶位点处或其附近:SEQ ID NO:22-26, 和 SEQ ID NO:28-33 和 SEQ ID NO:35-38。

20. 一种位点特异性锌指核酸酶,其在选自下组的核酸靶位点处或其附近进行剪切:SEQ ID NO:22-26, 和 SEQ ID NO:28-33 和 SEQ ID NO:35-38。

21. 权利要求 20 的锌指核酸酶,其中该锌指核酸酶包括 3-6 个锌指结构域,每个锌指结构域包含识别螺旋区,其中该锌指蛋白包含在表 3 的单一行中排序并显示的识别螺旋区。

FAD2 性能基因座及相应的能够诱导靶向断裂的靶位点特异性结合蛋白

[0001] 相关申请交叉引用

[0002] 本申请要求获得于 2012 年 9 月 7 日提交的美国临时专利申请 No. 61/697, 886 的优先权, 本文通过提述并入其全部内容。

[0003] 公开领域

[0004] 本公开一般地涉及用于植物重组技术(例如, 用于产生转基因植物)的组合物和方法。更具体地, 本公开涉及植物细胞和植物, 它们的基因组中包含可用于位点特异性地引入任何感兴趣核酸的基因座。

[0005] 背景

[0006] 许多植物被外源核酸(例如转基因)遗传转化以引入期望的性状, 例如提高农业价值。可以通过遗传转化提高农业价值的实例包括: 提高营养品质, 提高产量, 害虫或疾病抗性, 干旱和胁迫耐受性, 提高园艺品质(例如, 改善色素沉着和/或生长), 除草剂抗性, 从植物产生工业上有用的化合物和/或材料, 和/或生产药物。将克隆的基因引入到植物细胞中并回收稳定可育的转基因植物, 可用于制造在多个世代中稳定的植物遗传修饰, 借此实现对作物的遗传工程化。

[0007] 在用于遗传转化和产生转基因植物的方法中, 通常将外源 DNA 随机引入到真核植物细胞的细胞核或质体 DNA 中, 随后分离含有被整合的外源 DNA 的细胞, 然后再生被稳定转化的植物。转基因植物通常通过土壤杆菌介导的转化技术产生。这些技术的成功刺激了其他用于将感兴趣核酸分子引入到植物基因组内的方法的发展, 例如 PEG 介导的原生质体 DNA 摄取, 基因枪(微粒轰击), 和晶须硅介导的转化。

[0008] 然而, 在所有这些植物转化方法中, 引入到植物基因组中的外源核酸被随机整合到植物细胞的基因组中, 并且拷贝数可不预测。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。例如, 转基因经常以重复序列的形式被整合, 或者是整个转基因或是其部分。这种复杂的整合模式通常会对整合核酸的表达水平具有不利的影响(例如, 由于转录后基因沉默机制破坏转录的 RNA, 或者由于诱导整合 DNA 的甲基化)。同时, 整合位点的位置通常也会影响被整合核酸的表达水平。而且, 外源 DNA 的整合可能对发生整合的基因组区域产生破坏性影响, 由此影响或干扰靶区域的正常功能, 从而产生不良的副作用。上述因素的组合导致转基因或外源 DNA 的表达水平(和整体农艺品质)在不同的转基因植物细胞和植物系之间发生很大的变异, 即便是它们是通过相同方法创建的。因为整合是随机的, 所以在从业者试图产生具有期望特征的新植物时, 这些影响不能够人为控制。

[0009] 上述考虑因素致使人们无论在任何时候研究将特定的外源核酸引入到植物中的效应时, 都必须产生和分析大量的转基因植物品系以获得显著的结果。类似地, 在产生含有特定整合核酸的转基因植物以提供具有期望表型的转基因植物时, 必须创建独立创建的转基因植物品系的大群体, 以便选择具有最佳核酸表达并对转基因植物的整体表型和性能具

有最小或者没有副作用的植物品系。在通过插入多个外源核酸（即，基因堆叠）创建转基因植物时，这些实际的考虑因素具有更大的重要性。在这样的植物中，转录后基因沉默等现象会增多。

[0010] 为了控制植物中的转基因插入，已经开发出了多种方法。见例如 Kumar and Fladung(2001)Trends Plant Sci.6:155-9。这些方法依赖基于同源重组的转基因整合，同源重组的转基因整合已经被成功地应用于原核生物和低等真核生物。Paszkowski et al. (1988)EMBO J. 7:4021-6。然而，直到最近，在植物中，转基因整合的主导机制仍然依赖于非常规重组 (illegitimate recombination)，其几乎不涉及重组 DNA 链之间的同源性。因此，这个领域中的主要挑战是检测和选择性产生罕见的同源重组事件，其会被效率高得多的通过非常规重组实现的整合事件所掩盖。而且，即使实现了对靶向的同源重组事件的选择性产生和检测，该事件也必须被靶向到宿主基因组的期望位置处方可实现这一策略的最大利益。

[0011] 例如，靶向遗传转化的一个推定的好处是，与通过随机整合获得的转化事件相比，转基因表达在事件与事件之间的变异性降低。另一个推定的好处是，显著减少在筛选引入核酸、分选转化构建体、和产生有助于在所得转基因植物中获得期望整体特征的事件中所需的事件数目。实现这些好处所需要的一个关键因素是鉴定基因组中这样的特定位置，在该位置上转基因性能是一致的，并且如果可能的话，在该位置上对宿主植物的不利影响被消除或最小化。

[0012] 最近，已经有人描述了用于靶向剪切基因组 DNA 的方法和组合物。这些靶向剪切事件能够用于，例如，诱导靶向突变，诱导细胞 DNA 序列的靶向删除，和促进在预定染色体基因座处的靶向重组和整合。见例如，Urnov et al. (2010)Nature 435(7042):646-51；美国专利公开 20030232410；20050208489；20050026157；20050064474；20060188987；20090263900；20090117617；20100047805；20110207221；20110301073；2011089775；20110239315；20110145940；和国际公开 WO 2007/014275，本文通过提述并入其全部内容用于所有目的。剪切可以通过使用特异性核酸酶，例如工程化的锌指核酸酶 (ZFN)、转录激活因子样效应物核酸酶 (TALEN)，或者使用具有引导特异性剪切的工程化 crRNA/tracrRNA (‘单向导 RNA’ (single guide RNA)) 的 CRISPR/Cas 系统而发生。美国专利公开 No. 20080182332 描述了使用非规范 (non-canonical) 的锌指核酸酶 (ZFNs) 靶向修饰植物基因组；美国专利公开 No. 20090205083 描述了植物 EPSPS 基因座的 ZFN 介导靶向修饰；美国专利公开 No. 20100199389 描述了靶向修饰植物 Zp15 基因座，美国专利公开 No. 20110167521 描述了靶向修饰参与脂肪酸生物合成的植物基因。此外，Moehle et al. (2007)Proc. Natl. Acad. Sci. USA 104(9):3055-3060 描述了使用设计的 ZFN 在特定的基因座处靶向添加基因。美国专利公开 20110041195 描述了制造纯合二倍体生物的方法。

[0013] 然而，仍然需要有助于修饰和 / 或调节植物 FAD2 基因表达的组合物和方法，包括产生在 FAD2 基因座处靶向插入期望转基因的植物。

[0014] 公开概述

[0015] 本公开描述了用于调节 FAD2 基因表达（例如在植物、藻类和真菌中）的组合物和方法，和这些基因座作为将感兴趣的核酸序列（例如外源核酸序列）靶向整合到宿主细胞内的位点的用途。在一些实施方案中，宿主细胞可能含有一个或多个基因组，所述基因组具

有一个或多个 FAD2 序列（例如，同源物和 / 或旁系同源物），其中任何一个或全部序列均可被选择性修饰和 / 或破坏。在具体的实施例中，本公开描述了欧洲油菜 (*Brassica napus*)（即欧洲油菜品系 DH12075）中的 FAD2A, FAD2A', FAD2C 和 FAD2C' 基因，以及相应的同源物或旁系同源物，和它们作为靶向整合感兴趣核酸序列的基因座的用途。如本文所述，尽管 FAD2 基因在宿主中参与脂肪酸生物合成，但是它们的修饰或破坏（例如，通过在 FAD2 的编码序列中整合外源核酸）出人意料地可能对所得的宿主生物体没有或者仅有极小的不利影响。

[0016] 本文还描述了与能够造成 FAD2 基因座内特定核酸序列的剪切和 / 或整合的多肽串联的一个或多个特定 FAD2 基因座的用途。与能够造成 FAD2 基因座内特定核酸序列的剪切和 / 或整合的多肽串联的一个或多个特定 FAD2 基因座的用途的实例包括选自下组的多肽：锌指蛋白，大范围核酸酶，TAL 结构域，RNA 引导的 CRISPR-Cas9，重组酶，亮氨酸拉链，CRISPR/Cas，和其他本领域已知的序列。特定的实例包括含有位点特异性 DNA 结合结构域多肽和剪切结构域多肽（例如核酸酶）的嵌合（“融合”）蛋白，例如含有锌指多肽和 FokI 核酸酶多肽的 ZFN 蛋白。例如，本文描述了被设计用于结合和诱导 FAD2A, FAD2A', FAD2C 和 FAD2C' 及其组合中的双链断裂、但不剪切相应的同源物或旁系同源物的特定 ZFN 在体外和体内的效力与特异性的演示。在一些实施方案中，特定的 FAD2 基因座可以和任何前述多肽一起使用，以造成感兴趣核酸的位点特异性整合，感兴趣核酸随后在宿主内表达，同时对宿主的农艺性能的不利影响极小。

[0017] 在某些方面中，本文描述包括与 FAD2 基因特异性结合的 DNA 结合结构域的多肽。在一些实施方案中，这样的多肽还可以包含核酸酶（剪切）结构域或半结构域（例如 ZFN、重组酶、转座酶或归巢核酸内切酶，包括具有经过修饰的 DNA 结合结构域、TAL 结构域、TALEN、RNA 引导的 CRISPR-Cas9 的归巢核酸内切酶），和 / 或连接酶结构域，使该多肽可以诱导靶向双链断裂，和 / 或易化感兴趣核酸在断裂位点的重组。在特定的实施方案中，靶向 FAD2 基因座的 DNA 结合结构域可以是 DNA 剪切功能结构域。前述的多肽可以在一些实施方案中用于将外源核酸引入到宿主生物体（例如植物或动物物种）基因组的一个或多个 FAD2 基因座处。在某些实施方案中，DNA 结合结构域包括具有一个或多个锌指（例如 2, 3, 4, 5, 6, 7, 8, 9 或更多个锌指）的锌指蛋白，并可以被工程化（是非天然出现的），从而结合 FAD2 基因内的任何序列。本文所述的任何锌指蛋白均可以结合靶基因编码序列内或相邻序列内（例如启动子或其他表达元件）的靶位点。在某些实施方案中，锌指蛋白结合 FAD2 基因内的靶位点，例如表 5 所示。示例 FAD2- 结合性锌指的识别螺旋区如表 4 所示。锌指蛋白的一个或多个组分锌指结合结构域可以是规范的 (C2H2) 锌指或非规范的（例如，C3H）锌指（例如，N 端和 / 或 C- 端锌指可以是非规范的指）。

[0018] 本文还描述了用于破坏或编辑 FAD2 基因的方法。此外本文还描述了通过根据本发明实施方案的方法产生的遗传修饰的宿主生物体（例如转基因植物）。在特定的实施例中，通过根据本发明实施方案的方法产生的转基因生物体可以是，但不限于，藻类、真菌、单子叶植物、双子叶植物，等。

[0019] 本文公开的 FAD2 基因可包括在任何具有一种或多种 FAD2 基因的植物、藻类、或真菌中发现的那些。

[0020] 通过参考附图对多个实施方案的详细描述，前述和其他的特征将变得更加显而易见

见。

[0021] 附图简述

[0022] 图 1, 子图 (panel) A-E, 显示了使用 AlignX® 生成的 FAD2 基因序列 (SEQ ID NO:5-8) 的序列比对。

[0023] 图 2 示意性描绘了使用 Jalview v 2.3 根据邻接距离生成的 FAD2 基因序列的系统进化树。

[0024] 图 3 显示了 pDAB104010 的质粒图, 其是一个代表性的锌指核酸酶表达盒。该构建体的布局与其他 ZFN 表达盒相似, 其中锌指结构域 24828 和 24829 与上述的备选锌指结构域相互交换。

[0025] 图 4 是一幅实例多线图, 显示了每 10,000 个序列读取中在靶 ZFN 位点具有缺失的序列的数目。图上的 X 轴指示缺失的碱基数目, Y 轴指示序列读取的数目, Z 轴指示用颜色编码的样品身份, 如图的右侧所述。所示的具体实例是 FAD2 基因家族的基因座 1 的, 其含有 3 个靶 ZFN 位点 A、B 和 C, 具有 4 个基因家族成员和两个对照转染, 作为对照样品 A 和 B 进行评价。图中从上到下列出的线 (A_对照_FADA', 位于图例顶部, 到 C_样品_FAD2C, 位于图例的底部) 是以从最接近标记的 X 轴 (A_对照_FADA') 到最远离标记的 X 轴 (C_样品_FAD2C) 显示的。

[0026] 图 5, 子图 A 和 B, 显示了 FAD2 基因的 ZFN 靶定。图 5A 是描述来自 FAD2 基因家族 ZFN 靶定基因座 4 的数据的曲线图。该基因座含有两个 ZFN 位点和两个必需的对照转染。图 5B 显示了在 ZFN 靶位点周围的具体序列背景 (SEQ ID NO:471-480), 鉴定了含有 C、T 和 G 三核苷酸重复的 FAD2A 和 C, 导致通过 FAD2A 和 C 基因座测序观察到的单碱基缺失的增加。

[0027] 图 6 显示了 pDAS000130 的质粒图。

[0028] 图 7 显示了 pDAS000031 的质粒图。

[0029] 图 8 示意性显示了用于转基因拷贝数测定的转基因靶引物和探针的结合位点。

[0030] 图 9 显示的是一个 Sequencher 文件, 显示了 FAD2A ZFN DNA 识别结构域 (bc12075_Fad2a-r272a2 和 bc12075_Fad2a-278a2), 以及 ZFN 特异性引物 (FAD2A. UnE. F1 和 FAD2A. UnE. R1) 和内源引物 (FAD2A/2C. RB. UnE. F1 和 FAD2A/2C. RB. UnE. R1) 的结合位点。

[0031] 图 10 显示的是一个示意图, 其显示了内源和转基因靶引物的结合位点, 这些引物用于检测 FAD2A 处通过完美 HDR 实现的转基因整合。

[0032] 图 11 是一个示意图, 显示了在完美编辑的 FAD2A 基因座中 Kpn1 限制性内切酶位点可能出现的位置, 和 FAD2a 5', hph 和 FAD2A 3' Southern 探针可以结合的位置。

[0033] 图 12 显示了 Kpn1 片段、FAD2A 5'、hph、FAD2A 3' 探针的位置和尺寸, 和对通过 HDR 在 FAD2A 基因座处整合了 ETIP 的植物进行 Southern 分析的预期结果。

[0034] 图 13 显示了拷贝数估计 qPCR 的代表性数据输出。左手的列代表了从已知的具有单一随机转基因插入物的 T₀转基因植物获得的数据, 并用作校准样品, 供所有其它样品对其“标准化”。右手的列是已知的具有 5 个转基因整合的 T₀转基因植物。两个植物的插入物拷贝数均使用 Southern 分析加以确定。其余的列提供了推定转基因植物的拷贝数估计。这些列标记为: 1 拷贝对照, 310420, 311819, 311821, 311822, 311823, 311824, 311827, 312524, 312525, 312526, 312527, 312529, 312530, 312532, 313810, 313811, 313905, 313941, 313942, 313944, 和 5 拷贝对照。这些列可用于确定每个转基因植物的估计拷贝数。当使用软件

估计拷贝数时,野生型植物、非转化的对照植物、和仅含质粒的对照不会产生拷贝数,因为它们并不针对 hph 和 HMG I/Y 靶标二者均具有 Cq。

[0035] 图 14 显示了 pDAS000129 的质粒图。

[0036] 图 15 显示了 pDAS000129 整合到 FAD2A 基因座内的示意图。

[0037] 图 16 显示了 pDAS000097 的质粒图。

[0038] 图 17 显示了 pDAS000389 的质粒图。

[0039] 图 18 显示了 pDAS000391 的质粒图。

[0040] 图 19 显示了 pDAS000392 的质粒图。

[0041] 图 20 显示了 pDAS000393 的质粒图。

[0042] 图 21 显示了 pDAS000394 的质粒图。

[0043] 图 22 显示了 pDAS000395 的质粒图。

[0044] 图 23 显示了 pDAS000396 的质粒图。

[0045] 图 24 显示了 pDAS000397 的质粒图。

[0046] 序列

[0047] 所示的核酸序列使用核苷酸碱基的标准字母缩写,如 37C.F.R. § 1.822 所定义的。仅显示了每个核酸序列的一条链,但是可以理解,当提到所示的链时总是包括了互补链在内。

[0048] 详细说明

[0049] 1. 若干实施方案的概述

[0050] 本发明的实施方案建立了一种将外源核酸(例如转基因)靶向整合到宿主基因组内的方法,该方法除了被整合的核酸所影响的表型之外不会严重不利地影响宿主的其它表型。一些实施方案可用于将多个核酸“堆叠”在单个宿主基因组中。这种方法利用了四种相互联系的技术的开发与部署:靶向技术,其可以用来在特定的基因组 DNA 位置处引入双链断裂(见,例如 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.* (8 月 19 日在线发表));递送技术,其可以用来递送优化的外源(供体)核酸(Bibikova et al. (2003) *Science* 300(5620):764);涉及修饰宿主基因(位于同源重组或 NHEJ 通路中)以增加靶向供体 DNA 整合的 HDR 或 NHEJ 频率的整合技术;分析工具,用以富集和表征靶向整合事件;和特定的期望宿主基因组位置(“性能基因座”),它们在遗传上良好定义并支持基因跨世代的稳定表达,而不会严重不利影响被转化的宿主生物体。另见美国专利公开 20030232410;20050208489;20050026157;20050064474;20060188987;20090263900;20090117617;20100047805;20110207221;20110301073;2011089775;20110239315;20110145940;20080182332;20090205083;20100199389;20110167521。例如,在植物中,性能基因座是这样的基因座:对该基因座处整合了转基因的植物的农艺或品质性质的不利影响可以忽略或者不存在。

[0051] 本文所述的实施方案利用了如下的出人意料的发现,即植物 FAD2 基因是靶向整合外源核酸(例如基因;非编码 DNA 序列,如工程化着陆垫 (Engineered Landing Pad) (ELPs) (美国专利申请 12/011,735) 和工程化转基因插入平台 (ETIP) (待审美国专利申请 No:61/697882);和植物转化单元)的性能基因座。FAD2 基因座在植物中普遍存在的性质,以及在芥花 (canola)、玉米、向日葵、小麦、棉花和大豆中改变或敲除 FAD2 不会造成农艺学或品质损失的证据,说明 FAD2 基因座是多种有商业意义的植物物种中的一大类性能基因座。

[0052] 一些实施方案利用 FAD2 基因座处的位点特异性双链 DNA 剪切,例如通过递送并表达靶位点特异性 DNA 识别和剪切蛋白造成的。在具体的实施方案中,这种 FAD2 特异性 DNA 识别和剪切蛋白可以是,例如但不限于,ZFN;TALEN;RNA 引导的 CRISPR-Cas9,重组酶(例如 Cre, Hin, RecA, Tre, 和 FLP 重组酶);大范围核酸酶,和源自上述任一种或其等同物的工程化蛋白。剪切还可以使用 CRISPR/Cas 系统实现,其用工程化的 crRNA/tracrRNA (“单向导 RNA”)来引导特异性剪切。在一些实施方案中,这种双链断裂可以通过供体核酸整合在 FAD2 性能基因座内的剪切位点处,例如通过同源介导的修复 (HDR) 或非同源末端连接 (NHEJ),而被修复。

[0053] 本公开举例说明了 FAD2 基因座作为性能基因座的用处,例如,通过描述芥花(欧洲油菜)中的 FAD2A, 2A', 2C 或 2C' 基因座,和可用于将外源核酸整合在 FAD2A, 2A', 2C 或 2C' 基因座处的相应的 FAD2 特异性 ZFN。

[0054] 本发明的实施方案致力于解决本领域许多尚未解决的问题。例如,本文所述的靶向整合方法的选择性可以减少或者不需要用于消除不良转基因事件所需要的重复田间试验,这些试验由于所涉及的资源和该区域中繁重的监管要求而成本高昂。而且,本文所述的靶向 DNA 整合方法在转基因堆叠过程中可能特别有益。

[0055] 尽管可以利用内源 FAD2 基因座处的固有 (native) 核苷酸序列来直接靶定感兴趣的核酸,但是在一些实施方案中,可以首先将核酸靶向到宿主的至少一个 FAD2 基因座处,使得其他感兴趣的核酸分子整合到宿主中变得容易。在其它实例中,可以使用这样的核苷酸序列:其与宿主生物体的位于 DNA 识别位点(例如锌指识别位点)之侧翼的固有序列不同源(例如,基本上随机生成的核酸序列)。

[0056] II. 术语

[0057] 如本申请,包括权利要求中使用的,除非上下文中清楚地另有说明,否则单数和单数形式的术语,例如“一”、“一个”和“该”,包括复数指代物。因此,例如,称“植物”、“该植物”或“一个植物”也指示多个植物。而且,根据上下文,使用术语“植物”还指示该植物的在遗传上相似或相同的后代。类似地,术语“核酸”可以指核酸分子的许多拷贝。类似地,术语“探针”可以指许多相似或相同的探针分子。

[0058] 数字范围包括限定该范围的数字,并明确地包括所限定的范围内的每个整数和非整数分数。除非另外指出,否则本文所用的全部技术和科学术语具有与本领域普通技术人员所普遍理解的相同的含义。

[0059] 为了方便审阅本公开所述的各种实施方案,提供了具体术语的如下解释:

[0060] 分离的:“分离的”生物组分(例如核酸或蛋白质)是与该组分天然存在的生物体细胞中的其他生物组分(即其它染色体、染色体外 DNA 和 RNA,和蛋白质)实质分离的、

与上述组分分开产生的、或者是从上述组分中纯化出来的,同时导致组分的化学或功能变化(例如,可以通过断裂将核酸与染色体的其余 DNA 连接在一起的化学键从染色体分离该核酸)。被“分离的”核酸分子和蛋白质包括通过标准纯化方法纯化的核酸分子和蛋白质。该术语还包括通过在宿主细胞中重组表达而制备的核酸和蛋白质,以及化学合成的核酸分子、蛋白质和肽。

[0061] 杂交:如本文关于植物使用的,术语“杂交”或“杂交的”是指通过授粉产生后代(例如细胞、种子和植物)实现配子融合。该术语既包括有性杂交(即,一个植物被另一个授粉)也包括自交(即,自花授粉,例如使用来自相同植物的花粉和胚珠)。

[0062] 回交:回交方法可用于将核酸序列引入到植物中。这种技术已经被广泛使用了数十年,用于向植物中引入新的性状。Jensen, N., Ed. Plant Breeding Methodology, John Wiley&Sons, Inc., 1988。在典型的回交方案中,感兴趣的原始栽培品种(轮回亲本)与携带待转染的感兴趣核酸序列的第二栽培品种(非轮回亲本)杂交。然后将此次杂交所得的后代再次与轮回亲本进行杂交,重复这个过程,直至获得如下的植物,其中被转化的植物中除了从非轮回亲本转移的核酸序列之外,还恢复了轮回植物几乎全部的期望形态和生理特征。

[0063] 基因渗入:如本文所使用的,术语“(基因)渗入”是指等位基因(或者包含外源核酸的经修饰等位基因)被传送到某个遗传背景的特定基因座处。在一些实施方案中,可以通过相同物种的两个亲本之间的有性杂交,其中至少一个亲本的基因组中具有特定的等位基因形式,来将该特定等位基因传送到至少一个后代中,从而将该特定等位基因渗入基因座。包含该特定等位基因的后代可以和具有期望遗传背景的品系反复回交。可以对回交后代选择特定等位基因形式,从而产生新的品种,其中特定的等位基因形式已被固定在该遗传背景中。在一些实施方案中,可以通过两个供体基因组之间的重组(例如,在融合的原生质体中),其中至少一个供体基因组在其基因组中具有特定的等位基因形式,来实现该特定等位基因的基因渗入。基因渗入可能涉及传送特定的等位基因形式,其可以是,例如但不限于,被破坏的或经过修饰的等位基因;转基因;PTU;和 ELP。

[0064] 种质:如本文所使用的,术语“种质”是指植物个体、植物群体(例如,植物品系、品种和家族)、和来自植物或植物群体的克隆的遗传材料。种质可以是生物体或细胞的一部分,或者它可以从生物体或细胞分出(例如,分离的)。一般地,种质提供具有特定的分子组成(makeup)的遗传材料,这是植物遗传性质的基础。如本文所用的,“种质”是指特定植物的细胞;种子;特定植物的组织(例如,可以生长出新植物的组织);和特定植物的非种子部分(例如,叶,茎,花粉,和细胞)。如本文所用的,术语“种质”与“遗传材料”同义,并可用于指代可以繁殖出植物的种子(或其他植物材料)。“种质库”可以指代不同种子或其它遗传材料(其中每种基因型被唯一地标识)的有序集合,从它们可以培养出已知的栽培种或者可以产生新的栽培种。

[0065] 基因:如本文所用的,术语“基因”(或“遗传元件”)可以指具有功能性意义的可遗传的基因组 DNA 序列。基因可以是天然核酸,或者整合到基因组中的核酸。术语“基因”也可用于指,例如但不限于,由可遗传的基因组 DNA 序列编码的 cDNA 和 / 或 mRNA。

[0066] 核酸分子:如本文所用的,术语“核酸分子”可以指核苷酸(即,核糖核苷酸,脱氧核糖核苷酸,和 / 或前述任一者的修饰形式)的聚合物形式。本文所使用的“核酸分子”与

“核酸”和“多核苷酸”是同义的。该术语包括 RNA、cDNA、基因组 DNA 的有义链和反义链，及其合成形式和混合聚合物。该术语包括任何拓扑构象，包括单链，双链，部分双链体化，三链体化，发夹化，环形和挂锁构象。核酸分子可以包括天然存在的和经过修饰的核苷酸的任一种或两种。这样的核苷酸也可以通过天然存在的和 / 或非天然存在的核苷酸键连接在一起。

[0067] 核酸分子可以被化学或生物化学修饰，或者可以含有衍生的核苷酸碱基，这是本领域技术人员容易理解的。这样的修饰包括，例如但不限于：标记；甲基化；一个或多个天然存在的核苷酸被类似物取代；和核苷酸间修饰（例如，不带电的连接，例如，甲基膦酸酯，磷酸三酯，氨基磷酸酯，和氨基甲酸酯；带电荷的连接，例如硫代磷酸酯和二硫代磷酸酯；悬垂部分 (pedant moieties)，例如，肽；插层剂，例如，吡啶和补骨脂；螯合剂；烷化剂；和经过修饰的连接，例如， α -异头核酸）。

[0068] 外源的：“外源的”分子是指这样的分子：就多核苷酸而言，其核苷酸序列和 / 或基因组位置（即基因座）（且就多肽而言，其氨基酸序列和 / 或细胞定位而言）不是规定的系统（例如种质、品种、优良品种、和 / 或植物）所固有的。在实施方案中，外源的或异源的多核苷酸或多肽可以是人为提供给生物系统（例如，植物细胞，植物基因，特定的植物物种或栽培品种，和 / 或植物染色体）的分子，并且不是该特定生物系统固有的。因此，称某个核酸为“外源的”，可以表示该核酸来自天然存在的来源之外的来源，或者可以表示该核酸具有非天然的构型、遗传位置、或者元件布置。

[0069] 相反，例如，“天然”或“内源”的核酸是这样的核酸（例如，基因），其不包含除了该核酸通常天然出现的染色体或其他遗传材料中通常存在的那些核酸元件之外的核酸元件。内源的基因转录本由处于天然染色体基因座处的核苷酸序列编码，而不是人为提供给细胞的。

[0070] 可操作地连接：当第一核酸序列与第二核酸序列处于功能关系中时，则第一核酸序列与第二核酸序列可操作地连接。例如，当启动子影响编码序列的转录或表达时，启动子与编码序列可操作地连接。当重组产生时，可操作地连接的核酸序列一般是邻接的，并且在需要连接两个蛋白编码区时，处于同一阅读框内。然而，可操作地连接的元件不必是邻接的。

[0071] 启动子：启动子是 DNA 的某个区域，其一般位于核酸的上游（朝向 5' 区域），可增强该核酸的转录。启动子允许适当地激活或抑制与之可操作地连接的核酸。启动子含有被转录因子识别的特定的序列。这些因子与启动子 DNA 序列结合，导致 RNA 聚合酶的招募，这种酶从核酸的编码区合成 RNA。转化：当载体将核酸分子转移到细胞内时，该载体“转化”或“转导”细胞。当核酸分子被细胞稳定复制时，无论是通过将该核酸分子纳入到细胞的基因组中还是通过游离型复制 (episomal replication)，细胞被核酸分子“转化”。如本文所使用的，术语“转化”包括所有能够将核酸分子引入到细胞内的技术。实例包括但不限于：病毒载体转染；质粒载体转化；电穿孔 (Fromm et al. (1986) Nature 319:791-3)；脂质体转染 (Felgner et al. (1987) Proc. Natl. Acad. Sci. USA 84:7413-7)；显微注射 (Mueller et al. (1978) Cell 15:579-85)；土壤杆菌介导的转移 (Fraley et al. (1983) Proc. Natl. Acad. Sci. USA 80:4803-7)；直接 DNA 摄取；和微粒轰击 (Klein et al. (1987) Nature 327:70)。

[0072] 引入 :如本文所使用的,术语“引入“在指外源核酸转位到细胞内时,是指使用本领域可得任何方法将核酸纳入到细胞内。该术语包括核酸引入方法,包括例如但不限于,转染 ;转化 ;和转导。

[0073] 转基因 :如本文所使用的,术语“转基因“是指感兴趣的外源核酸编码序列。例如,转基因可以编码工业上或药学上有用的化合物,或者有利于期望农艺性状(例如除草剂抗性 or 害虫抗性)的表达产物。在进一步的实例中,转基因可以是反义核酸,其中该反义核酸的表达会抑制靶核酸序列的表达。转基因可以包括与该转基因可操作地连接的调节序列(例如,启动子)。在一些实施方案中,通过位点特异性靶向引入到 FAD2 基因座处的感兴趣核酸分子是转基因。然而,在其他实施方案中,感兴趣的核酸分子可以是 PTU,ELP,ETIP,或者内源核酸序列(例如,其中内源核酸序列的额外外源基因组拷贝是期望的)。

[0074] 元件还可以包括编码结构 RNA,例如 shRNA,的 DNA。这种 RNA 能够修饰外源或内源基因,包括但不仅限于,影响功能发挥(posting)或赋予除草剂抗性。

[0075] 重组 :如本文所使用的,术语“重组”是指通过人为干预而改变的材料(例如,核酸、基因、多核苷酸、和 / 或多肽)。例如,重组分子的部分或元件的布置可能不是固有的布置,和 / 或重组分子的一级序列可能已自其固有序列被改变,例如为了优化其表达和 / 或活性。材料可以被改变,从而产生处于其天然环境或状态中或从其天然环境或状态脱离的重组材料。举一例子,如果某个核酸的开放阅读框的核苷酸序列已经脱离其天然背景且被克隆到人工的核酸分子(例如,载体)中,则该开放阅读框是重组的。用于产生重组分子(例如,重组核酸)的方案和试剂是本领域常见的,并且被常规地使用。术语“重组”在本文中还可以指包含重组材料的细胞或生物体(例如,含有重组核酸的植物和 / 或植物细胞)。在一些实例中,重组生物体是转基因生物体。

[0076] 载体 :如本文所使用的,术语“载体”是指能够将至少一个核酸节段转移到细胞内的多核苷酸或其他分子。载体可以任选地包含介导载体维持和 / 或实现其预期用途的组分 / 元件(例如,复制必需的序列,赋予药物或抗生素抗性的基因,多克隆位点,和 / 或可操作地连接的实现克隆基因表达的启动子 / 增强子元件)。载体可以源自,例如质粒、噬菌体,或植物或动物病毒。“克隆载体”、“穿梭载体”或“亚克隆载体”一般包括易化克隆或亚克隆步骤的可操作地连接的元件(例如,含有多个限制性核酸内切酶位点的多克隆位点)。

[0077] 表达载体 :如本文所使用的,术语“表达载体”是指这样的载体,其包含可以易化编码序列在特定宿主生物体中表达的可操作地连接的多核苷酸序列。例如,细菌表达载体可以易化编码序列在细菌中的表达。类似地,植物表达载体可以易化编码序列在植物细胞中的表达。易化在原核生物中表达的多核苷酸序列可以包括,例如但不仅限于 :启动子 ;操纵基因 ;和核糖体结合位点。真核表达载体(例如,植物表达载体)可包括,例如,启动子 ;增强子 ;终止信号 ;和多腺苷酸化信号(以及其它序列),它们一般与在原核表达载体中所使用的不同。

[0078] 序列同一性 :如本文在两个核酸或多肽序列内容中所使用的,术语“序列同一性”或“同一性”是指当经过比对两个序列从而在特定的比较窗口上获得最大的对应度时,两个序列中相同的残基。序列同一性的数值可以通过比较两个在比较窗口上最佳比对的序列(例如,核酸序列和氨基酸序列)加以确定,其中与参考序列(其不含有添加或缺失)相比,序列在比较窗口中的部分可以包含添加或缺失(即,缺口)以实现两个序列的最佳比对。序

列同一性可以被计算为百分比,通过确定两个序列中存在的相同核苷酸或氨基酸残基的位置的数目,用匹配位置的数目除以比较窗口中的位置的总数,并将结果乘以 100,产生序列同一性的百分比。

[0079] 用于比对比较序列的方法是本领域众所周知的。在下列文献中描述了各种程序和比对算法,例如: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。序列比对方法和同源性计算的详细描述可以在 Altschul et al. (1990)J. Mol. Biol. 215:403-10 中找到。

[0080] 美国国家生物技术信息中心 (NCBI) 的基本局部比对搜索工具 (BLAST™; Altschul et al. (1990)) 可以用于比对序列,它可以从多个来源获得,包括美国国家生物技术信息中心 (Bethesda, MD), 和互联网上,用于和多种序列分析程序联合使用。关于如何使用这一程序确定序列同一性的描述可以在互联网上在 BLAST™的“帮助”一节中获得。对于核酸序列的比较,可以采用 BLAST™(Blastn) 程序的“Blast 2sequences”功能,使用默认参数。当通过这一方法进行评估时,与参考序列具有越大相似性的核酸序列将显示越高的百分比同一性。

[0081] 如本文所使用的,术语“基本上相同的”可以指核苷酸序列超过 80% 相同。例如,基本上相同的核苷酸序列可以和参考序列至少 85%,至少 86%;至少 87%;至少 88%;至少 89%;至少 90%;至少 91%;至少 92%;至少 93%;至少 94%;至少 95%;至少 96%;至少 97%;至少 98%;至少 99%;或至少 99.5% 相同。

[0082] 基因座:如本文所使用的,术语“基因座”是指基因组上与可测量特征(例如,性状)相应的位置。在一些实施方案中,特别感兴趣的基因座是 FAD2 基因的基因组位置,在该处基因的破坏可降低或消除从野生型基因转录的 mRNA 的表达。基因座可以用在 Southern 杂交或 PCR 中与该基因座中含有的独特核苷酸序列杂交的探针来定义。

[0083] 标志物:如本文所使用的,“标志物”是指这样的基因或核苷酸序列,其能够用于鉴定可能具有特定等位基因和/或显示特定性状或表型的植物。标志物可以被描述为给定基因组座位处的变异。遗传标志物可以是短 DNA 序列,例如围绕单碱基对变化的序列(单核苷酸多态性,或“SNP”),或者是长序列,例如微卫星/简单序列重复(“SSR”)。“标志物等位基因”是指特定植物中存在的标志物版本。如本文所使用的,术语“标志物”可以指植物染色体 DNA 的克隆节段(例如,包含 FAD2 基因座,或经过修饰和/或被破坏的 FAD2 基因座的节段),还/或者可以指与植物染色体 DNA 克隆节段互补的 DNA 分子。本领域的普通技术人员会认识到,可以将获得额外的连续核苷酸序列包含在标志物中的过程重复几乎无限多次(仅受限于染色体的长度),借此沿着该染色体鉴定其他的标志物。上述标志物的任何变化均可使用在本发明的某些实施方案中。

[0084] 在一些实施方案中,种质中转基因或标志物的存在(用“靶”序列表征)可以通过使用核酸探针(例如寡核苷酸)进行检测。探针可以是 DNA 分子或 RNA 分子。寡核苷酸探

针可以合成制备或者通过克隆制备。合适的克隆载体是本领域技术人员所熟知的。RNA 探针可以通过本领域已知的手段合成,例如使用 DNA 分子模板。

[0085] 寡核苷酸探针可以是带标记的或者无标记的。用于标志物核酸分子的技术有很多,包括例如但不限于,通过切口平移的放射标记;随机引发;和用末端脱氧转移酶 (terminal deoxytransferase) 加尾 (tailing),其中所用的核苷酸是带标记的,例如带放射性 ^{32}P 标记。其他可以使用的标记物包括,例如但不限于,荧光团;酶;酶底物;酶的辅助因子;和酶抑制剂。或者,作为使用自身提供或者与其它反应试剂联合提供可检测信号的标记物的替代手段,可以使用能结合受体的配体,其中受体被标记(例如,被上述的标记物所标记),从而自身提供或者与其他试剂一起提供可检测信号。参见,例如,Leary et al. (1983)Proc. Natl. Acad. Sci. USA 80:4045-9。

[0086] 探针可以是被检测的转基因或标志物的精确拷贝。探针也可以是这样的核酸分子,其包含与包含待检测的转基因或标志物的染色体 DNA 克隆片段基本上相同的核苷酸序列,或者由这样的核苷酸序列构成。探针可以进一步包括额外的核酸序列,例如启动子;转录信号;和/或载体序列。

[0087] 探针可以包含靶核苷酸序列的全部或一部分和来自基因组的额外的邻接核苷酸序列。这在本文中被称为“邻接探针”。上述额外的邻接核苷酸序列是指原始靶的“上游”或“下游”,这取决于来自染色体的邻接核苷酸序列位于原始靶的 5' 侧还是 3' 侧,如常规上所理解的。探针还可以包含不与原始靶邻接的核苷酸序列;这种探针在本文中被称作“非邻接探针”。非邻接探针可以定位于和染色体上的原始靶序列足够接近的位置,使得该非邻接探针与原始标志物或转基因连接。

[0088] 在一些实施方案中,探针是与待检测的靶的精确拷贝“能够特异性杂交”或“特异性互补”的核酸分子。术语“能够特异性杂交”和“特异性互补”表示有足够程度的互补性,使得核酸分子和靶之间发生稳定而特异的结合。核酸分子不需要与其靶序列 100% 互补才能特异性杂交。当存在足够程度的互补性从而避免在期望特异性结合的条件下,例如在严格的杂交条件下,发生核酸与非靶序列的非特异性结合时,核酸分子是能够特异性杂交的。

[0089] 导致特定严格度的杂交条件将随着所选杂交方法的性质和杂交核酸序列的组成和长度而改变。一般地,杂交温度和杂交缓冲液的离子强度(尤其是 Na^+ 和 / 或 Mg^{++} 浓度)将决定杂交的严格度,尽管清洗时间也会影响严格度。获得特定严格程度所需的杂交条件的计算是本领域普通技术人员已知的,并且在下列文献中有讨论,例如, Sambrook et al. (ed.) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY., 1989, 第 9 和 11 章;和 Hames and Higgins (eds.) Nucleic Acid Hybridization, IRL Press, Oxford, 1985 年。关于核酸杂交更详细的说明和指导可以在下列文献中找到,例如, Tijssen, “Overview of principles of hybridization and the strategy of nucleic acid probe assays,” 其汇编在 Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes, 第 I 部分, 第 2 章中, Elsevier, NY, 1993; 和 Ausubel 等人编辑, Current Protocols in Molecular Biology, 第 2 章, Greene Publishing and Wiley-Interscience, NY, 1995。

[0090] 如本文所使用的,“严格条件”包括只有当杂交分子与 DNA 靶之间的错配小于 25%

时才会发生杂交的条件。“严格条件”包括更具体水平的严格度。因此,如本文所使用的,“温和 (moderate) 严格”条件是序列错配超过 25% 的分子不会杂交的条件;“中等 (medium) 严格”条件是序列错配超过 15% 的分子不会杂交的条件;“高严格”条件是序列错配超过 10% 的分子不会杂交的条件。“极高严格”条件是序列错配超过 6% 的分子不会杂交的条件。

[0091] 在特定的实施方案中,严格条件是在如下条件下杂交:65℃于 6x 盐-柠檬酸钠 (SSC) 缓冲液,5x Denhardt 溶液,0.5% SDS 和 100 μg 剪切的鲑鱼精 DNA,接着在下列缓冲液中顺次清洗 15-30 分钟:65℃于 2x SSC 缓冲液和 0.5% SDS,随后是 1x SSC 缓冲液和 0.5% SDS,最后是 0.2x SSC 缓冲液和 0.5% SDS。

[0092] 连锁(不)平衡:如本文所使用的,术语“连锁平衡”是指这样的情况,其中标志物和第二核酸(例如,转基因,PTU,和第二标志物)独立地分离;即,标志物和第二核酸在后代中随机地组合(sort)。显示连锁平衡的核酸被认为是不连锁的(无论它们是否位于同一染色体上)。如本文所使用的,术语“连锁不平衡”是指如下情况,其中标志物和第二核酸以非随机的方式分离;即,核酸的重组频率小于 50% (因此,根据定义,在同一连锁群上的距离小于 50cM)。在一些实施例中,显示连锁不平衡的核酸被认为是连锁的。

[0093] 连锁,紧密连锁,和极紧密连锁:如本文所使用的,标志物与第二核酸(例如,转基因,PTU,和第二标志物)之间的连锁可以指这样的现象,其中染色体上的核酸显示出可测量的一起传递到下一代的个体中的概率。因此,一个标志物与第二核酸的连锁可以用重组频率来度量和/或表示。两个核酸越彼此接近,这个概率越接近“1”。因此,术语“连锁”可以指一个或多个基因或标志物与第二核酸以大于 0.5 的概率一起传递(这是根据标志物/基因在不同染色体上的位置的自由组合(independent assortment)而预期的)。当基因(例如,转基因)的存在有助于个体中的表型时,与该基因连锁的标志物可以被称作与该表型连锁。因此,术语“连锁”可以指标志物和基因或者标志物和表型之间的关系。

[0094] 相对遗传距离(通过杂交频率确定,并以厘摩(cM)度量)通常与两个连锁标志物或基因在染色体上彼此分离的物理距离(用碱基对度量)成比例。一厘摩被定义为显示 1% 重组频率(即,两个标志物在每 100 次细胞分裂中发生一次交换事件)的两个遗传标志物之间的距离。一般地,一个标志物与另一个标志物或基因越接近(无论它们之间的距离是按照遗传距离还是物理距离度量的),它们连锁得越紧密。因为染色体距离与性状间重组事件的频率大致成比例,因此存在与重组频率相关的近似物理距离。在这种关联是众所周知的,或者在主要作物和许多其他生物中可以容易地确定(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)。例如,1cM 在酵母中对应于大约 2.5-3.0kb,在拟南芥中对应于大约 140kb,在向日葵中对应于大约 400kb,在桉树中对应于约 350kb。

[0095] 术语“连锁的”在本文中可以指一个或多个显示重组频率小于 50% (即,小于 50cM) 的核酸。例如,“连锁”的核酸的重组频率可以为大约 45%或更低,大约 40%或更低,大约 35%或更低,大约 30%或更低,大约 25%或更低,大约 20%或更低,大约 15%或更低,

和大约 10% 或更低。这些核酸在同一染色体上的物理距离（位于不同染色体上的核酸预期是连锁平衡的）对应于前述的重组频率，依赖于宿主基因组，并且可以按照上文所提出的容易地进行计算。

[0096] 如本文所使用的，术语“紧密连锁”可以指显示大约 20% 或更低的重组频率（即，大约 20cM 或更小）的一个或多个核酸。例如，“紧密连锁的”核酸可以以 22% 或更低，大约 18% 或更低，大约 16% 或更低，大约 14% 或更低，大约 12% 或更低，大约 10% 或更低，大约 8% 或更低，大约 6% 或更低，大约 4% 或更低，和大约 2% 或更低的频率重组。

[0097] 如本文所使用的，术语“极紧密连锁”可以指显示大约 10% 或更低的重组频率（即，大约 10cM 或更小）的一个或多个核酸。例如，“极紧密连锁的”核酸可以以 11% 或更低，大约 9% 或更低，大约 8% 或更低，大约 7% 或更低，大约 6% 或更低，大约 5% 或更低，大约 4% 或更低，大约 3% 或更低，大约 2% 或更低，和大约 1% 或更低的频率的重组。

[0098] 特定核酸与编码贡献于特定表型的多肽的基因越接近（无论是用遗传距离还是用物理距离测量），该特定核酸与该表型越紧密连锁。鉴于上述情况可以理解，与特定基因或表型连锁的核酸包括那些与该基因或表型紧密连锁的核酸，和那些与该基因或表型极紧密连锁的核酸。在一些实施方案中，特定核酸与 FAD2 基因座（例如，经过修饰或被破坏的 FAD2 基因座）越接近，无论是用遗传距离还是用物理距离测量，该特定核酸与由整合在该 FAD2 基因座处的外源核酸所赋予的任何性状 / 表型（或者在基因座未修饰的情况下，与野生型 FAD2 的表型）的连锁越紧密。因此，与包含整合的外源核酸的 FAD2 基因座相连锁、紧密连锁、和 / 或极紧密连锁的遗传标志物可在 MAS 程序中用于鉴定包含整合核酸的生物体（例如，植物或植物品种），鉴定包含由整合核酸赋予的表型的生物体，和将这种整合核酸和 / 或由该整合核酸赋予的表型培育到其他相容的生物体中。

[0099] 标志物辅助育种：如本文所使用的，术语“标志物辅助育种”可以指直接针对一个或多个性状（例如多基因性状）培育植物的方法。在当前的实践中，植物育种者试图鉴定容易检测的性状，例如花的颜色，种皮外观，或者与农艺上期望的性状相连锁的同工酶变体。然后，植物育种者通过跟踪该容易检测的性状的分离，在分离和育种群体中跟踪农艺学性状。然而，只有非常少的这些感兴趣性状与容易检测的性状之间的连锁关系可供用于植物育种。在本发明的一些实施方案中，标志物辅助育种包括鉴定一个或多个与 FAD2 基因座连锁的遗传标志物（例如，SNP，同工酶和 / 或 SSR 标志物），其中有利于感兴趣性状的外源核酸整合在该 FAD3 基因座中，并通过跟踪该一个或多个遗传标志物的分离，在分离的育种群体中跟踪感兴趣的性状。在一些实例中，一个或多个遗传标志物的分离可以用该一个或多个遗传标志物的探针通过测定来自后代植物的遗传样品中是否存在该一个或多个遗传标志物加以确定。标志物辅助育种可以为植物栽培品种的改良提供时间上和成本上高效率的过程。

[0100] 性状或表型：术语“性状”和“表型”在本文中可互换使用。为本公开的目的，特别感兴趣的性状包括农艺上重要的性状，例如可以在例如作物植物中表达的，和转基因表达产物自靶向整合事件的产生。术语“分子表型”可以指这样的表型，其可以在（一个或多个）分子的群体的水平上被检测到。在一些实例中，该分子表型可以仅能在分子水平上被检测到。表型的可检测分子可以是核酸（例如，基因组 DNA 或 RNA）；蛋白质；和 / 或代谢物。例如，分子表型可以是一个或多个基因产物（例如，在植物发育的特定阶段、或响应于环境条

件或胁迫)的表达谱。

[0101] 数量性状基因座:由于遗传(加性、显性和上位性)和环境影响而连续变化的性状通常被称为“数量性状”。数量性状可以根据两个因素区别于“定性”或“不连续”性状:对基因表达的环境影响,其导致表型的连续分布;和由多基因遗传产生的复杂分离模式。与数量性状表达相连锁的一个或多个基因组区域的鉴定将这样的区域定义为数量性状基因座(“QTL”)。

[0102] 植物:如本文所使用的,术语“植物”是指全植物,来自植物的细胞或组织培养物,和/或任一前述的任何部分。因此,术语“植物”包括,例如但不限于,全植物;植物组分和/或器官(例如,叶、茎和根);植物组织;种子;和植物细胞。植物细胞可以是,例如但不限于,植物中和/或属于植物的细胞,从植物分离的细胞,和通过培养分离自植物的细胞而获得的细胞。

[0103] “转基因植物”是在其至少一个细胞中包含外源多核苷酸的植物。术语“转基因”在本文中用于指任何基因型由于外源核酸的存在而被改变的细胞,细胞系,愈伤组织,组织,植物部分或植物。因此,该术语包括最初被改变从而包含外源多核苷酸的转基因生物体和细胞,和通过最初的转基因生物体或细胞的杂交或无性繁殖产生的那些生物体和细胞。如本文所使用的,术语“转基因”不包括通过常规植物育种方法(例如,单纯非转基因生物体的杂交)或通过天然出现的事件(例如随机杂交授粉,非重组病毒感染,非重组细菌转化,非重组转座,和自发突变)引入的基因组(染色体或染色体外)改变。

[0104] 植物“品系”,“品种”或“株”是一群具有相同家系(parentage)的个体植物。一个品系的植物在一定程度是近交的,并且一般在大多数遗传基因座(例如,FAD2基因座)上是纯合的且均质的。“亚系”可以指来自共同祖先的后代的近交子集,它们在遗传上与来自相同祖先的其他相似的近交子集的后代不同。在一些实施方案中,“亚系”可以通过这样产生:使在F₃-F₅代选出的个体转基因植物的种子近交,直至残余的分离中的基因座(segregating locus)在大多数或全部基因座处均是纯合的。

[0105] “结合蛋白”是能够与另一个分子结合的蛋白。结合蛋白能够结合,例如,DNA分子(DNA结合蛋白),RNA分子(RNA结合蛋白)和/或蛋白质分子(蛋白质结合蛋白)。在蛋白质结合蛋白的情况下,它能够与自身结合(形成同二聚体,同三聚体,等),和/或它能够与其他不同蛋白的一个或多个分子结合。结合蛋白可以具有不止一种类型的结合活性。例如,锌指蛋白具有DNA结合,RNA结合和蛋白结合活性。

[0106] “锌指DNA结合蛋白”(或结合结构域)是通过一个或多个锌指(这是结合结构域内的氨基酸序列区,其结构通过锌离子配位加以稳定)以序列特异性的方式结合DNA的蛋白质或较大蛋白内的结构域。术语锌指DNA结合蛋白通常被缩写为锌指蛋白或ZFP。

[0107] “TALE DNA结合结构域”或“TALE”是包含一个或多个TALE重复结构域/单元的多肽。该重复结构域参与TALE与其关联靶DNA序列的结合。单一“重复单元”(也称为“重复”)通常长度为33-35个氨基酸,并与天然存在的TALE蛋白内的其他TALE重复序列显示至少一些序列同源性。

[0108] 锌指和TALE结合结构域可以被“工程化”以结合预定的核苷酸序列,例如通过将天然存在的锌指或TALE蛋白的识别螺旋区工程化(改变一个或多个氨基酸)。因此,工程化的DNA结合蛋白(锌指或TALE)是非天然存在的蛋白质。用于工程化DNA结合蛋白的方法的

非限制性实例是设计和选择。设计的DNA结合蛋白不是天然存在的蛋白,其设计/组成主要来自合理标准。用于设计的合理标准包括采用替换规则和计算机化算法来处理现有ZFP和/或TALE设计和结合数据的数据库分选信息中的信息。参见,例如,美国专利6140081;6453242;和6534261;另见WO 98/53058;WO 98/53059;WO 98/53060;WO 02/016536和WO 03/016496和美国公开No. 20110301073。

[0109] “选定的”锌指蛋白或TALE是未在自然界中发现的蛋白,其产生主要来自经验性方法,例如噬菌体展示,相互作用捕获(interaction trap)或杂交选择。参见例如US 5,789,538;US 5,925,523;US 6,007,988;US 6,013,453;US6,200,759;WO 95/19431;WO 96/06166;WO 98/53057;WO 98/54311;WO 00/27878;WO 01/60970WO 01/88197,WO 02/099084和美国公开号No. 20110301073。

[0110] “剪切”是指DNA分子共价骨架的断裂。剪切可以通过多种方法触发,包括但不限于,磷酸二酯键的酶或化学水解。单链剪切和双链剪切均是可能的,并且双链剪切可以作为两个迥异的单链剪切事件的结果而发生。DNA剪切可导致产生平末端或粘末端。在某些实施方案中,融合多肽被用于靶向的双链DNA剪切。

[0111] “剪切半结构域”是这样的多肽序列,其与第二多肽(相同的或不同的)一道形成具有剪切活性(优选地,双链剪切活性)的复合体。术语“第一和第二剪切半结构域”、“+和-剪切半结构域”和“右和左剪切半结构域”可互换使用,指示可二聚体化的剪切半结构域对。

[0112] “工程化的剪切半结构域”是指已经被修饰,使其与另一个剪切半结构域(例如另一个被工程化的剪切半结构域)形成专性(obligate)异二聚体的剪切半结构域。另见美国专利公开号2005/0064474,20070218528,2008/0131962和2011/0201055,本文引用其全部内容作为参考。

[0113] 用于产生双链DNA断裂的手段:如本文所使用的,术语“用于产生双链DNA断裂的手段”意在援引美国法典第35编第112条(35U.S.C. § 112)第六款中美国国会授权的特殊权利要求规定。具体地,“用于产生双链DNA断裂的手段”是指一种分子结构,其能够剪切双链DNA分子的两条链。这样的结构包括在许多已知的核酸酶蛋白中含有的多肽结构域,例如,FokI核酸酶结构域,选自下组蛋白的催化结构域:MmeI,大肠杆菌素-E7(CEA7_ECOLX),大肠杆菌素-E9,APFL,EndA,Endo I(END1_ECOLI),人Endo G(NUCG_HUMAN),牛Endo G(NUCG_BOVIN),R.HinPII,1-BasI,1-BmOI,1-HmuI,1-TevI,1-TevII,1-TevIII,1-TwoI,R.MspI,R.MvaI,NucA,NucM,Vvn,Vvn_CLS,葡萄球菌核酸酶(NUC_STAAU),葡萄球菌核酸酶(NUC_STAHY),微球菌核酸酶(NUC_SHIFL),核酸内切酶yncB,脱氧核糖核酸内切酶(Endodeoxyribonuclease)I(ENRN_BPT7),Metnase,Nb.BsrDI,BsrDI A,Nt.BspD6I(R.BspD6I大亚基),ss.BspD6I(R.BspD6I小亚基),R.PIeI,MlyI,AlwI,MvaI269I,BsrI,BsmI,Nb.BtsCI,Nt.BtsCI,Rl.BtsI,R2.BtsI,BbvCI亚基1,BbvCI亚基2,BpuI0Iα亚基,BpuI0Iβ亚基,BmrI,BfiI,1-CreI,hExoI(EX01JHUMAN),酵母ExoI(EX01_YEAST),大肠杆菌ExoI,人TREX2,小鼠TREX1,人TREX1,牛TREX1,大鼠TREX1,人DNA2,酵母DNA2(DNA2_YEAST)。

[0114] 用于修复双链DNA断裂的手段:如本文所使用的,术语“修复双链DNA断裂的装置”意在援引美国法典第35编第112条第六款中美国国会授权的特殊权利要求规定。具体地,

“修复双链 DNA 断裂的手段”是指一种分子结构,其能够易化 / 催化双链 DNA 分子末端的接合,例如通过接合由单条双链 DNA 分子剪切产生的末端,或者通过接合由单条双链 DNA 分子剪切产生的一个末端与一个外源双链 DNA 分子的末端。这样的结构包括许多已知的连接酶蛋白,例如 Cre 重组酶中含有的多肽结构域。在一些实例中,同一分子结构可以同时作为用于产生双链 DNA 断裂的手段,又作为用于修复双链 DNA 断裂的手段,其中同一结构既易化双链 DNA 分子的剪切,又易化双链 DNA 分子的修复(例如, Hin 重组酶)。

[0115] 基因组中位点特异性双链断裂的诱发可诱导宿主植物细胞 DNA 修复通路,其通过同源性介导的修复 (HDR) 或者非同源末端连接 (NHEJ) 修复恢复双链断裂。在植物中,科学文献报道,基因或供体 DNA 到固有基因组位置中或预先工程化的位置的精确整合涉及这样的输入 (incoming) 供体 DNA 构建体,它们包含不同量的与靶定的双链断裂的侧翼序列同源的序列。此类供体到特定的靶基因座中的整合可以认为是依赖 HDR 通路。植物中完全依赖 HDR 方法的基因打靶可能是受限的,因为已有报道,与 NHEJ 相比, HDR 修复通路不是占主导地位 DNA 修复通路。在已经出版的利用靶特异性 DNA 断裂 (ZFN, TALEN, 或工程化的大范围核酸酶,等) 的植物科学文献中, NHEJ 通路已经被报道用作向基因组中引入特定点突变 (插入或删除) 的方法。本文中我们报道,在植物中,在设计为具有 0 到 <10bp 的各种同源区域的供体 DNA 设计存在下,可以通过 NHEJ 修复通路在靶断裂处特异性插入位点特异性双链断裂 (通过 ZFN, TALEN 等诱导)。多种不同的 DNA 供体设计具有零同源性到小的 1 - 10bp 范围,从线性到环形,从单链到双链,均可以使用 NHEJ 通路被靶向到特定的位置。基于 NHEJ 的供体 DNA 的植物基因组靶向可以是基于“粘末端捕捉”,其中基因组中的由 FokI (或其它 II 型核酸内切酶结构域) 产生的靶向双链断裂和相应的粘末端在 NHEJ 供体 DNA 设计上。粘末端供体 DNA 可以作为具有预定突出端的线性供体 DNA 被直接递送给细胞。一种替代方法是通过共递送宿主靶 ZFN 和含有至少一个与靶识别位点相同的 ZFN 识别位点的环形 DNA 供体分子,在体内产生供体 DNA 粘末端。至少一个 ZFN 的表达会切割宿主基因组 DNA (固有的或预先工程化的) 和环形供体 DNA,从而产生粘末端,粘末端使用宿主的 NHEJ 修复通路解析 (resolve)。

[0116] 在供体分子上可能具有一个或多个 ZFN 切割位点 (单独一个 ZFN 切割位点用来使整个供体分子线性化, 2 个同样的 ZFN 位点用来释放较小的供体 DNA 片段, 或者 2 个不同的 ZFN 位点用来从供体释放一个片段、并从宿主基因组 DNA 释放一个对应的片段 (DNA 替换))。

[0117] 因此,供体多核苷酸可以是 DNA 或 RNA,单链和 / 或双链,并能够以线性或环形形式被引入到细胞中。参见,例如,美国专利公开 Nos. 20100047805 和 20110207221。在本发明的某些实施方案中,还可以包括线性外源 (供体) 核酸,包含这些核酸的组合物,和用于制造和使用这些线性供体分子的方法。在某些实施方案中,线性供体分子在引入线性供体分子的细胞中稳定保持。在其它实施方案中,线性供体分子被修饰成抵抗外切性核酸切割 (exonucleolytic cleavage),例如通过在供体分子末端的一个或多个碱基对之间放置一个或多个硫代磷酸酯磷酸二酯键。该线性外源核酸还可以包括单链的特异 DNA。

[0118] III. FAD2 性能基因座

[0119] 命名为 FAD2 (脂肪酸去饱和酶 2) 的基因座包含在涉及植物中脂肪酸含量复杂多基因性状遗传的 QTL 中。FAD2 编码负责将油酸 (18:1) 去饱和为亚油酸 (C18:2) 的

酶。Tanhuanpaa et al. (1998) Mol. Breed. 4:543-50 ; Schierholt et al. (2001) Crop Sci. 41:1444-9。

[0120] 在植物油生物合成通路中, 脂肪酸去饱和酶 (FAD) 在植物脂质生物合成中发挥关键作用, 并且它们的活性显著影响脂肪酸组成。FAD 在植物中丰富, 并且表达分析提示 FAD mRNA 以过量丰度产生。而且, FAD 基因在各种组织和细胞类型, 以及包括质体和内质网在内的亚细胞区室中表达。

[0121] 植物的脂肪酸组成, 以及由其产生的油在许多应用中的性能, 是由主要脂肪酸组分——油酸、亚油酸和亚麻酸 (C18:3) 的相对浓度决定的。这些脂肪酸的浓度主要受到酶 FAD2 和 FAD3 的功能的调节。在植物中, 油酸根据下述过程被转变成亚油酸和亚麻酸:

[0122]



FAD2 FAD3

[0123] FAD2 基因已经在主要植物和藻类物种中被鉴定, 包括但不限于, 玉米、大豆、棉花、拟南芥、小麦、牧草、水稻、向日葵和芸苔属 (Brassica), 并且 FAD2 表达的修饰可导致这些生物体中的脂肪酸谱的改变。而且, 包含经过修饰的 FAD2 基因的植物已经被商业化, 并且已经有人显示, 破坏 FAD2 基因能够改良由宿主植物产生的油的营养和功能性质, 但不会对宿主植物的造成农艺学上的损失。例如, 已经以商品名 Nexera[®] (Dow AgroSciences, LLC) 被商业化的芥花 (canola) 和向日葵品种的特征在于, 与野生型芥花和向日葵谱相比, 具有较高的油酸、较低的亚油酸和较低的亚麻酸 (和较低的饱和脂肪酸) 组成。

[0124] 在欧洲、北美和澳大利亚种植的主导芥花物种是欧洲油菜, 它是一种被认为是通过甘蓝 (*B. oleracea*) (具有二倍体 C 基因组) 与芜菁 (*B. rapa*) (具有二倍体 A 基因组) 杂交产生的多倍体油菜物种。细胞基因组研究显示, AA 和 CC 基因组显示一定程度的亲缘性, 彼此部分同源。A 和 C 基因组均含有高百分比的同源或旁系同源基因。因此, 有人认为, AA 和 CC 基因组来自共同的祖先基因组。Prakash and Hinata (1980) Opera Botanica 55:1-57。尽管两个祖先物种的基因组在技术上均被归类为二倍体, 但是这些基因组含有高百分比的彼此重复的区域。Song et al. (1991) Theor. Appl. Genet. 82:296-304。详细的细胞器和细胞核 RFLP 分析显示, 芜菁的 AA 基因组向油菜贡献 10 个染色体, 而甘蓝从其作为母本供体的 CC 基因组贡献了 9 个染色体。Song et al. (1992) Genome 35:992-1001。通过两个祖先基因组中基因组重复的数目, 以及 A、B 和 C 基因组之间高百分比的相似性, 已经出现了 FAD2 和 FAD3 基因的数个拷贝。实际上, 这个事实使得培育具有这些基因的修饰和 / 或破坏拷贝的芥花以产生特定的脂肪酸谱变得特别具有挑战性。

[0125] 芥花中 FAD2 的已知的功能性基因拷贝均位于 A 基因组的连锁群 N4 上。Scheffler et al. (1997) TAG 94(5):583-91; Schierholt et al. (2000) TAG 101(5-6):897-901。更近些时候, 已经将芥花的高油酸性状与位于 A 基因组上的修饰和破坏的 FAD2 基因联系起来。美国专利申请公开 No. US2006/0248611A1; 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, 2006年1月14-18日, 2006, San Diego, CA. 失活性的 FAD2 等位基因通过减少亚油酸去饱和为亚麻酸而贡献于对油酸含量的控制。这种高油酸和 FAD3 性状在欧洲油菜品种 (DMS100) 中被鉴定, 其具有大约 77% 的特征性油酸含量。见美国申请 No. 10/545, 100。此外, FAD2 基因新近被定位在 A5 染色体上, 且有人声称它是高 C18:1 含量的原因。见, Yang et al., "Brassica napus genome" Theor Appl Genet (2012) 125:715-729。进一步, 已经开发出了用于辅助将 Fad2 和高油酸性状基因渗入到芥花中的遗传标志物。

[0126] 可以在植物中修饰和 / 或破坏 FAD2 基因座而不会不利地影响植物价值, 而且对于许多目的而言, 实际上可以增加其价值, 包括改变 FAD2 表达, 改变油含量 / 比例, 和 / 或整合和表达期望的转基因。而且, 根据 FAD 基因座在植物中普遍存在的性质, 可以为了至少某些目的在许多物种中修饰和 / 或破坏 FAD2 基因座而不会造成损害, 这些物种包括, 例如但不限于: 芥花; 大豆; 玉米; 小麦; 牧草; 芸苔属植物; 水稻; 番茄; 大麦; 燕麦; 高粱; 棉花; 和向日葵, 以及真菌和藻类。本发明的实施方案包括 FAD2 基因座, 及其作为性能基因座用于整合外源核酸的用途。已经发现在 FAD2 基因用作性能基因座的背景中有数种特征是理想的, 在实例中 FAD2 基因表现出其中至少一种, 包括, 例如但不限于: 在宿主生物体整个生命周期中具有大体一致的表达水平; 以及, 令人惊讶地, 供体 DNA 整合到 FAD2 基因座处不会诱发宿主的品质或健康 (fitness) 损失。

[0127] 在本发明的一些实施方案中, 至少一个 FAD2 基因座 (例如, FAD2A, FAD2A', FAD2C, 和 / 或 FAD2C' 基因座) 被用作外源核酸 (例如, 包含编码感兴趣多肽的核苷酸序列的核酸) 的位点特异性整合的靶位点。在特定的实施方案中, 外源核酸的整合产生修饰的基因座。例如, 外源核酸的整合可以修饰基因座, 从而产生被破坏的 (即, 失活的) FAD2 基因。

[0128] 在一些实施方案中, FAD2 基因座可以包括与选自下组的核苷酸序列的互补物能够特异性杂交的核苷酸序列: SEQ ID NO: 22-26, SEQ ID NO: 28-33 和 SEQ ID NO: 35-38。例如, FAD2 基因座可以包括选自下组的核苷酸序列: SEQ ID NO: 22-26, SEQ ID NO: 28-33 和 SEQ ID NO: 35-38。在一些实施方案中, FAD3 基因座可以包括与选自下组的核苷酸序列基本上相同的核苷酸序列: SEQ ID NO: 22-26, SEQ ID NO: 28-33 和 SEQ ID NO: 35-38。例如, 在一些实施方案中, FAD2 基因座是包含与选自下组的核苷酸序列具有至少大约 85% 的同一性的核苷酸序列的 FAD3 同源物 (例如, 直系同源物或旁系同源物): SEQ ID NO: 22-26, SEQ ID NO: 28-33 和 SEQ ID NO: 35-38。FAD2 同源物可包含这样的核苷酸序列, 其与选自下组的核苷酸序列例如但不限于至少 80%; 至少 85%; 至少约 90%; 至少约 91%; 至少约 92%; 至少约 93%; 至少约 94%; 至少约 95%; 至少约 96%; 至少约 97%; 至少约 98%; 至少约 99%; 至少约 99.5%; 99.6%, 99.7%, 99.8% 和 / 或至少约 99.9% 相同: SEQ ID NO: 22-26, SEQ ID NO: 28-33 和 SEQ ID NO: 35-38。这样的 FAD2 同源物对于多种生物而言, 可以容易地识别并从本领域技术人员容易获得的任何完整或部分基因组中分离。

[0129] IV. 核酸在 FAD2 基因座处的靶向整合

[0130] 外源核酸在 FAD2 基因座处的位点特异性整合可以通过本领域技术人员已知的任何技术实现。在一些实施方案中, 在 FAD2 基因座处整合外源核酸包括使细胞 (例如, 分离的细胞或组织或生物体中的细胞) 与包含外源核酸的核酸分子接触。在实例中, 这样的核酸分子可以包括位于外源核酸侧翼的、易化该核酸分子与至少一个 FAD2 基因座之间的同

源重组的核苷酸序列。在特定实例中,位于外源核酸侧翼的易化同源重组的核苷酸序列可以和 FAD2 基因座的内源核苷酸互补。在特定的实例中,位于外源核酸侧翼的易化同源重组的核苷酸序列可以和既往整合的外源核苷酸互补。在一些实施方案中,多个外源核酸可以被整合在一个 FAD2 基因座处,如在基因堆叠中。

[0131] 在一些实施方案中,可以通过宿主细胞的内源细胞机构,例如但不限于,内源 DNA 和内源重组酶,来易化(例如催化)FAD2 基因座处的核酸整合。在一些实施方案中,FAD2 基因座处的核酸整合可以被一种或多种提供给宿主细胞的因子(例如,多肽)所易化。例如,可以提供核酸酶,重组酶和/或连接酶多肽(或是独立的,或是作为嵌合多肽的一部分),通过使该多肽与宿主细胞接触或者通过在宿主细胞内表达该多肽来提供。因此,在一些实例中,可以向宿主细胞中引入包含编码至少一种核酸酶,重组酶和/或连接酶多肽的核苷酸序列的核酸,该核酸可以与要位点特异性整合到 FAD2 基因座处的核酸同时或顺次引入,其中在宿主细胞中,该至少一种核酸酶,重组酶,和/或连接酶多肽自该核苷酸序列表达。

[0132] A. DNA-结合多肽

[0133] 在一些实施方案中,位点特异性整合可以利用能够识别并结合特定核苷酸序列(例如宿主生物体基因组中的特定核苷酸序列)的因子来实现。例如,许多蛋白质包含能够识别并以位点特异性的方式与 DNA 结合的多肽结构域。被 DNA 结合多肽识别的 DNA 序列可被称为“靶”序列。能够识别 DNA 并以位点特异性的方式与之结合的多肽结构域一般可正确折叠并独立地发挥功能,从而以位点特异性的方式结合 DNA,即使是当它表达在与最初分离出该结构域的蛋白质不同的多肽中也是如此。类似地,被 DNA 结合多肽识别和结合的靶序列一般能够被这样的多肽识别并结合,即使是存在于大 DNA 结构中(例如,染色体)也是如此,特别是当靶序列所处的位点是已知可溶性细胞蛋白质可以达到的位点(例如,基因)时。

[0134] 尽管从自然界中存在的蛋白质中鉴定的 DNA 结合多肽通常与离散的核苷酸序列或基序(例如,共有识别序列)结合,但是在本领域中存在并且知晓有方法来修饰许多这样的 DNA 结合多肽从而识别不同的核苷酸序列或基序。DNA 结合多肽包括,例如但不限于:锌指 DNA 结合结构域;亮氨酸拉链;UPA DNA 结合结构域;GAL4;TAL;LexA;Tet 抑制子;LacR;和类固醇激素受体。

[0135] 在一些实例中,DNA 结合多肽是锌指。单独的锌指基序可以被设计成靶向并特异性结合多种多样的 DNA 位点中的任何种。规范的 Cys₂His₂(以及非规范的 Cys₃His) 锌指多肽通过将 α -螺旋插入到靶 DNA 双螺旋的大沟中来结合 DNA。锌指识别 DNA 是模块性的;每个指主要与靶中的三个连续碱基对接触,并由多肽中的少数关键残基介导识别。通过在靶向性核酸内切酶中包含多个锌指 DNA 结合结构域,靶向性核酸内切酶的 DNA 结合特异性可以被进一步提高(因此,由其赋予的任何基因调节效应的特异性也被提高)。见例如 Urnov et al. (2005) *Nature* 435:646-51。因此,可以工程构建并使用一个或多个锌指 DNA 结合多肽,使得引入到宿主细胞中的靶向性核酸内切酶与宿主细胞基因组内独特的 DNA 序列相互作用。

[0136] 优选地,锌指蛋白是非天然存在的,即其是被工程构建为结合所选的靶位点的。参见,例如 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 ;美国专利 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 ;和美国专利公开 Nos. 2005/0064474 ;2007/0218528 ;2005/0267061, 本文引用其全部内容作为参考。

[0137] 与天然存在的锌指蛋白相比,工程化的锌指结合结构域可以具有新的结合特异性。工程化方法包括,但不仅限于,合理设计和各种类型的选择。合理设计包括,例如,使用包含三链体(或四链体)核苷酸序列和单个锌指氨基酸序列的数据库,其中每个三链体或四链体核苷酸序列与结合该特定三链体或四链体序列的一个或多个锌指氨基酸序列相关。参见,例如共同拥有的美国专利 6, 453, 242 和 6, 534, 261, 本文引用其全部内容作为参考。

[0138] 示例性选择方法,包括噬菌体展示和双杂交系统,在下列文献中被公开:美国专利 5, 789, 538 ;5, 925, 523 ;6, 007, 988 ;6, 013, 453 ;6, 410, 248 ;6, 140, 466 ;6, 200, 759 ;和 6, 242, 568 ;以及 WO 98/37186 ;WO 98/53057 ;WO 00/27878 ;WO 01/88197 和 GB 2, 338, 237。此外,例如,在共同拥有的 WO 02/077227 中描述了增强锌指结合结构域的结合特异性。

[0139] 此外,如这些和其他参考文献所公开的,锌指结构域和/或多指(multi-fingered)锌指蛋白可以使用任何合适的接头序列,包括例如,长度为5个或更多个氨基酸的接头,连接在一起。示例性的长度为6个或更多个氨基酸的接头序列可另见美国专利 Nos. 6, 479, 626 ;6, 903, 185 ;和 7, 153, 949。本文所述的蛋白质可以包含该蛋白的各个锌指之间的合适接头的任意组合。

[0140] 靶位点的选择;ZFP和用于设计和构建融合蛋白(和编码它们的多核苷酸)的方法是本领域技术人员已知的,并且在下列文献中有详述:美国专利 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 01/60970WO 01/88197 ;WO 02/099084 ;WO 98/53058 ;WO 98/53059 ;WO 98/53060 ;WO 02/016536 和 WO 03/016496。

[0141] 此外,如这些和其他参考文献所公开的,锌指结构域和/或多指锌指蛋白可以使用任何合适的接头序列,包括例如长度为5个或更多个氨基酸的接头,连接在一起。关于长度为6个或更多个氨基酸的示例性接头序列可另见美国专利 Nos. 6, 479, 626 ;6, 903, 185 ;和 7, 153, 949。本文所述的蛋白质可以包含该蛋白的各个锌指之间的合适接头的任意组合。

[0142] 在一些实例中,DNA结合多肽是来自 GAL4 的 DNA 结合结构域。GAL4 是酿酒酵母中的模块性反式激活因子,但它也可以在许多其他生物体中充当反式激活因子。参见,例如 Sadowski et al. (1988) Nature 335:563-4。在这种调节系统中,编码酿酒酵母半乳糖代谢通路中的酶的基因的表达受到可得的碳源的严格调节。Johnston(1987) Microbiol. Rev. 51:458-76。这些代谢酶的转录控制由正调节蛋白 GAL4 和 GAL4 特异性结合的 17bp 对称 DNA 序列(UAS)之间的相互作用介导。

[0143] 天然的 GAL4 包括 881 个氨基酸残基,分子量为 99kDa。GAL4 包括功能性的自主结构域,它们的总活性是 GAL4 的体内活性的原因。Ma and Ptashne(1987) Cell 48:847-53) ;

Brent and Ptashne(1985)Cell 43(3Pt 2):729-36。GAL4 的 N 端 65 个氨基酸包括 GAL4DNA 结合结构域。Keegan et al. (1986)Science231:699-704 ;Johnston(1987)Nature 328:353-5。序列特异性结合要求与 DNA 结合结构域中的 6 个 Cys 残基配位的二价阳离子的存在。含有配位阳离子的结构域通过与 DNA 螺旋的大沟直接接触与 17bp UAS 每端的保守 CCG 三链体相互作用并识别。Marmorstein et al. (1992)Nature 356:408-14。该蛋白的 DNA 结合功能使得 C 端转录激活结构域定位于启动子附近,从而使该激活结构域能够指导转录。

[0144] 其他在某些实施方案中可以使用的 DNA 结合多肽包括,例如但不限于,来自 AVRBS3 可诱导基因的结合序列;来自 AVRBS3 可诱导基因或从其工程化的合成结合序列的共有结合序列(例如 UPA DNA 结合结构域);TAL;LexA 蛋白(见,例如, Brent&Ptashne(1985), 同上);LacR(见,例如, Labow et al. (1990)Mol. Cell. Biol. 10:3343-56;Baim et al. (1991)Proc. Natl. Acad. Sci. USA 88(12):5072-6);类固醇激素受体(Elilliston et al. (1990)J. Biol. Chem. 265:11517-121);Tet 抑制子(美国专利 6271341)和突变的 Tet 抑制子,其在四环素(Tc)存在下可结合 tet 操纵基因序列,但在没有四环素(Tc)时则不结合;NF- κ B 的 DNA 结合结构域;和 Wang et al. (1994)Proc. Natl. Acad. Sci. USA91(17):8180-4 中描述的调节系统的组分,其利用 GAL4、激素受体以及 VP16 的融合物。

[0145] 在某些实施方案中,在本文所述的方法和组合物中使用的一种或多种核酸酶的 DNA 结合结构域包括天然存在的或工程化的(非天然存在)TAL 效应物 DNA 结合结构域。参见,例如,美国专利公开号 20110301073,本文引用其全部内容作为参考。黄单胞菌属(*Xanthomonas*)的植物病原细菌已知会在重要的作用植物中导致多种疾病。黄单胞菌的致病性取决于一个保守的 III 型分泌(T3S)系统,该系统向植物细胞内注入超过 25 种不同的效应物蛋白。这些注入的蛋白质包括转录激活因子样(TAL)效应物,其模拟植物转录激活因子并操纵植物转录组(见 Kay et al(2007)Science 318:648-651)。这些蛋白质包含 DNA 结合结构域和转录激活结构域。其中一种最良好表征的 TAL 效应物是来自野油菜黄单胞菌叶斑致病亚型(*Xanthomonas campestris* pv. *Vesicatoria*)的 AvrBs3(见 Bonas et al(1989)Mol Gen Genet 218:127-136 和 W02010079430)。TAL 效应物含有一个由串联重复构成的集中(centralized)结构域,每个重复含有大约 34 个氨基酸,其对于这些蛋白质的 DNA 结合特异性是关键。此外,它们含有核定位序列和酸性转录激活结构域(综述见 Schornack S, et al(2006)J Plant Physiol 163(3):256-272)。此外,在植物病原细菌青枯雷尔氏菌(*Ralstonia solanacearum*)中,两个被命名为 brg11 和 hpx17 的基因被发现与青枯雷尔氏菌生物型 1 菌株 GMI1000 和生物型 4 菌株 RS1000 中的黄单胞菌 AvrBs3 家族同源(见 Heuer et al(2007)Appl and Envir Micro 73(13):4379-4384)。这些基因的核苷酸序列彼此具有 98.9% 相同,差异在于在 hpx17 的重复结构域中有一个 1,575bp 的缺失。然而,这两种基因产物与黄单胞菌 AvrBs3 家族蛋白的序列同一性小于 40%。见例如美国专利 Nos., 8,420,782 和 8,440,431 和美国专利公开 No. 20110301073。

[0146] 在其他实施方案中,核酸酶包括 CRISPR/Cas 系统。CRISPR(成簇的、规律间隔的短回文重复序列(clustered regularly interspaced short palindromic repeats))基因座编码系统的 RNA 组件,和 cas(CRISPR-相关)基因座,其编码蛋白质(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), 构成 CRISPR/Cas 核酸酶系统的基因序列。微生物宿主中的 CRISPR 基因座含有 CRISPR- 相关 (Cas) 基因以及能够编程 CRISPR 介导的核酸切割特异性的非编码 RNA 分子元件的组合。

[0147] II 型 CRISPR 是最良好表征的系统之一, 并以四个顺序的步骤实施靶向 DNA 双链断裂。首先, 从 CRISPR 基因座转录两个非编码 RNA, 前 -crRNA 阵列和 tracrRNA。第二, tracrRNA 与前 -crRNA 的重复区域杂交, 并且介导将前 -crRNA 加工成含有单独间隔序列的成熟 crRNAs。第三, 成熟的 crRNA:tracrRNA 复合体指导 Cas9 靶向 DNA, 其中靶向 DNA 通过 crRNA 上的间隔子与靶 DNA 上紧邻前间隔子邻近基序 (PAM) 的原间隔子 (protospacer) 之间的 Watson-Crick 碱基配对来实现, 这是靶向识别的额外要求。最后, Cas9 介导靶 DNA 的剪切, 从而在前间隔子内产生双链断裂。CRISPR/Cas 系统的活性包括三个步骤: (i) 外源 DNA 序列插入到 CRISPR 阵列中, 以防止将来的攻击, 这个过程称作“适应”, (ii) 相关蛋白质的表达, 以及阵列的表达和加工, 随后是 (iii) 对外来核酸的 RNA 介导的干扰。这样, 在细菌细胞中, 多个所谓的“Cas”蛋白质参与 CRISPR/Cas 系统的天然功能, 并在功能中发挥作用, 例如外源 DNA 的插入等。

[0148] 在某些实施方案中, Cas 蛋白质可以是天然存在的 Cas 蛋白的“功能性衍生物”。天然序列多肽的“功能性衍生物”是指具有与天然序列多肽共同的定性生物学性质的化合物。“功能性衍生物”包括, 但不限于, 天然序列的片段, 以及天然序列多肽及其片段的衍生物, 只要它们具有与相应的天然序列多肽一样的生物学活性即可。本文构想的生物学活性是功能衍生物将 DNA 底物水解成片段的能力。术语“衍生物”涵盖多肽的氨基酸序列变异体、共价修饰、及其融合体。Cas 多肽或其片段的合适衍生物包括, 但不限于: Cas 多肽或其片段的突变体、融合体、共价修饰。Cas 蛋白, 包括 Cas 蛋白质或其片段, 以及 Cas 蛋白或其片段的衍生物, 可以从细胞获得或者化学合成或者通过这两种方法的组合获得。细胞可以是天然产生 Cas 蛋白的细胞, 或者天然产生 Cas 蛋白并被遗传工程化从而以更高的表达水平产生内源 Cas 蛋白的细胞, 或者天然产生 Cas 蛋白并被遗传工程化从而从外源引入的核酸产生 Cas 蛋白的细胞, 其中该核酸编码与内源 Cas 相同或不同的 Cas。在一些情况下, 细胞不天然产生 Cas 蛋白, 并且被遗传工程化从而产生 Cas 蛋白。

[0149] 在特定的实施方案中, DNA 结合多肽特异性识别并结合包含在宿主生物体基因组核酸内的靶核苷酸序列。在一些实例中, 在宿主基因组中可以发现靶核苷酸序列的任何数量的离散实例。靶核苷酸序列在生物体的基因组中可以是罕见的 (例如, 在基因组中可能存在靶序列的小于约 10, 约 9, 约 8, 约 7, 约 6, 约 5, 约 4, 约 3, 约 2, 或约 1 个拷贝)。例如, 靶核苷酸序列可以位于生物体基因组内的独特的位点。各靶核苷酸序列可以, 例如但不限于, 相对于彼此随机分散在整个基因组中; 位于基因组的不同连锁群中; 位于相同连锁群中; 位于不同染色体上; 位于相同染色体上; 位于基因组中在生物体的相似条件下表达 (例如, 在同一种, 或者在功能上基本上相同的调节因子的控制之下) 的位点处; 和位于在基因组中彼此紧邻的位置 (例如, 靶序列可以包含在作为串联体被整合到基因座中的核酸内)。

[0150] B. 靶向核酸内切酶

[0151] 在特定的实施方案中,特异性识别并结合靶核苷酸序列的 DNA 结合多肽可以包含在嵌合多肽中,从而赋予嵌合多肽对靶序列的特异性结合。在实例中,这样的嵌合多肽可以包括,例如但不限于,核酸酶,重组酶,和 / 或连接酶多肽,这些多肽如上文所述。包含 DNA 结合多肽与核酸酶,重组酶,和 / 或连接酶多肽的嵌合多肽还可以包括其它的功能性多肽基序和 / 或结构域,例如但不限于:位于嵌合蛋白中的功能性多肽之间的间隔子序列;前导肽;将融合蛋白导向到细胞器(例如,细胞核)的肽;被细胞酶剪切的多肽;肽标签(例如 Myc, His 等);和其它不干扰嵌合多肽功能的氨基酸序列。

[0152] 嵌合多肽中的功能多肽(例如, DNA 结合多肽和核酸酶多肽)可以是可操作地连接的。在一些实施方案中,嵌合多肽的各个功能多肽可以通过从这样的单个多核苷酸表达而可操作地连接,该多核苷酸至少编码彼此对框连接的各功能多肽,从而产生编码嵌合蛋白的嵌合基因。在替代实施方案中,嵌合多肽的各个功能多肽可以通过其它方式可操作地连接,例如通过各个独立表达的多肽的交联。

[0153] 在一些实施方案中,特异性识别并结合靶核苷酸序列的 DNA 结合多肽可以包含在天然的分离的蛋白质(或其突变体)中,其中天然的分离的蛋白质或其突变体还包括核酸酶多肽(并且还可以包括重组酶和 / 或连接酶多肽)。这样的分离的蛋白质的实例包括 TALEN,重组酶(例如,Cre,Hin,Tre,和 FLP 重组酶),RNA 指导的 CRISPR-Cas9,和大范围核酸酶。

[0154] 如本文所使用的,术语“靶向核酸内切酶”(targeting endonuclease)是指包含 DNA 结合多肽与核酸酶多肽的分离的天然或工程化蛋白质及其突变体,以及包含 DNA 结合多肽与核酸酶多肽的嵌合多肽。任何包含特异性识别并结合 FAD2 基因座内包含的靶核苷酸序列(例如,或者是因为该靶序列包含在该基因座处的天然序列中,或者是因为该靶序列被引入到该基因座中,例如,通过重组)的 DNA 结合多肽的靶向核酸内切酶均可以在某些实施方案中使用。

[0155] 可以在本发明的特定实施方案中使用的嵌合多肽的一些实例包括,但不限于,下述多肽的组合:锌指 DNA 结合多肽;Fok I 核酸酶多肽;TALE 结构域;亮氨酸拉链;转录因子的 DNA 结合基序;以及从例如但不限于下述各项分离的 DNA 识别和 / 或剪切结构域:TALEN,重组酶(例如,Cre,Hin,RecA,Tre,和 FLP 重组酶),RNA 指导的 CRISPR-Cas9,大范围核酸酶;和其他本领域技术人员已知者。具体的实例包括嵌合蛋白,其包含位点特异性 DNA 结合多肽和核酸酶多肽。嵌合多肽可以通过本领域技术人员已知的方法进行工程化,从而改变该嵌合多肽内包含的 DNA 结合多肽的识别序列,以便将该嵌合多肽靶向到感兴趣的特定核苷酸序列。

[0156] 在某些实施方案中,嵌合多肽包含 DNA 结合结构域(例如,锌指,TAL-效应物结构域,等)和核酸酶(剪切)结构域。剪切结构域可以是相对 DNA 结合结构域为异源的,例如锌指 DNA 结合结构域与来自核酸酶的剪切结构域,或者 TALEN DNA 结合结构域与剪切结构域,或者大范围核酸酶 DNA 结合结构域与来自不同核酸酶的剪切结构域。异源的剪切结构域可以从任何核酸内切酶或核酸外切酶获得。可以衍生该剪切结构域的示例性核酸内切酶包括,但不限于,限制性核酸内切酶和归巢核酸内切酶。参见,例如,2002-2003 年目录,New England Biolabs,Beverly,MA;和 Belfort et al.(1997)Nucleic Acids Res. 25:3379-3388。其他可切割 DNA 的酶是已知的(例如,S1 核酸;绿豆核酸酶;胰 DNA

酶 I ;微球菌核酸酶 ;酵母 HO 核酸内切酶 ;另见 Linn 等人 (编辑)Nucleases, Cold Spring Harbor Laboratory Press, 1993)。这些酶 (或其功能片段) 中的一种或多种可以用作剪切结构域和剪切半结构域的来源。

[0157] 类似地, 剪切半结构域可以从如上所述的任何其剪切活性需要二聚体化的核酸酶或其部分产生。一般地, 如果融合蛋白包含剪切半结构域, 则需要两个融合蛋白用于剪切。或者, 可以使用包含两个剪切半结构域的单一蛋白。两个剪切半结构域可以来自同一核酸内切酶 (或其功能片段), 或者每个剪切半结构域可以来自不同的核酸内切酶 (或其功能片段)。此外, 两个融合蛋白的靶位点优选地相对彼此如下布置, 使得两个融合蛋白与其各自靶位点的结合将剪切半结构域彼此置于一定的空间取向, 该空间取向允许剪切半结构域形成功能性的剪切结构域 (例如通过二聚体化)。因此, 在某些实施方案中, 靶位点的邻近边缘被 5-8 个核苷酸或者 15-18 个核苷酸所分开。然而, 在两个靶位点之间可以介入任意整数的核苷酸或核苷酸对 (例如, 2-50 个核苷酸对或者更多)。一般地, 剪切位点位于靶位点之间。

[0158] 限制性核酸内切酶 (限制酶) 存在于许多物种中, 并且能够序列特异性结合 DNA (在识别位点处), 并在结合位点处或其附近剪切 DNA, 从而例如使一个或多个外源序列 (供体 / 转基因) 被整合在该结合 (靶) 位点处或其附近。某些限制酶 (例如, IIS 型) 在远离识别位点的地方剪切 DNA, 并具有可分开的结合和剪切结构域。例如, IIS 型酶 Fok I 催化 DNA 的双链剪切, 剪切位点在一条链上位于距其识别位点达 9 个核苷酸处, 而在另一条链上位于距其识别位点 13 个核苷酸处。参见, 例如, 美国专利 5, 356, 802 ; 5, 436, 150 和 5, 487, 994 ; 以及 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。因此, 在一个实施方案中, 融合蛋白包括来自至少一种 IIS 型限制酶的剪切结构域 (或剪切半结构域) 和一个或多个锌指结合结构域, 其可以是或者不是被工程化的。

[0159] 具有可以分开的剪切结构域与结合结构域的示例性 IIS 型限制酶是 Fok I。这种酶作为二聚体发挥作用。Bitinaite et al. (1998) Proc. Natl. Acad. Sci. USA 95:10, 570-10, 575。因此, 为本公开的目的, 在被公开的融合蛋白中使用的 Fok I 酶的部分视为剪切半结构域。因此, 对于使用锌指 -Fok I 融合物的细胞序列的靶向双链剪切和 / 或靶向替换, 可以使用两个融合蛋白, 每一个融合蛋白都包含 Fok I 剪切半结构域, 来重建具有催化活性的剪切结构域。或者, 也可以使用含有 DNA 结合结构域和两个 Fok I 剪切半结构域的单一多肽分子。

[0160] 剪切结构域或剪切半结构域可以是蛋白质的保留剪切活性、或者保留多聚体化形成功能性剪切结构域的能力的任何部分。

[0161] IIS 型限制酶的实例在美国专利公开 No. 20070134796 中有描述, 本文引用其全部内容作为参考。额外的限制酶也还有可分开的结合和剪切结构域, 并且是本公开所构想的。参见, 例如, Roberts et al. (2003) Nucleic Acids Res. 31:418-420。

[0162] 在某些实施方案中, 剪切结构域包括一个或多个工程化的剪切半结构域 (也称作二聚体化结构域突变体), 其最小化或阻止同二聚体化 (homodimerization), 如例如美国专利公开 Nos. 20050064474 ; 20060188987 和 20080131962 所述, 本文引用其全部内容作为

参考。位于 Fok I 位置 446, 447, 479, 483, 484, 486, 487, 490, 491, 496, 498, 499, 500, 531, 534, 537, 和 538 的氨基酸残基均是供用于影响 Fok I 剪切半结构域二聚体化的靶点。

[0163] 可形成专性异二聚体的 Fok I 的示例性工程化剪切半结构域包括这样的一对, 其中第一剪切半结构域在 Fok I 的位置 490 和 538 的氨基酸残基处包括突变, 第二剪切半结构域在氨基酸残基 486 和 499 处包括突变。

[0164] 因此, 在一个实施方案中, 490 处的突变用 Lys(K) 代替 Glu(E); 538 处的突变用 Lys(K) 代替 Iso(I); 486 处的突变用 Glu(E) 代替 Gln(Q); 499 处的突变用 Lys(K) 代替 Iso(I)。具体地, 本文所述的工程化剪切半结构域是通过使一个剪切半结构域中的位置 490(E → K) 和 538(I → K) 突变产生命名为“E490K:I538K”的工程化剪切半结构域, 并将另一个剪切半结构域中的位置 486(Q → E) 和 499(I → L) 突变产生命名为“Q486E:I499L”的工程化剪切半结构域而制备的。本文所述的工程化剪切半结构域是专性异源二聚体突变体, 其中异常的剪切被最小化或消除。参见, 例如美国专利公开号 No. 2008/0131962, 本文出于所有目的引用其全部公开内容作为参考。

[0165] 在某些实施方案中, 工程化的剪切半结构域包括位置 486、499 和 496 处的突变(相对于野生型 Fok I 编号), 例如用 Glu(E) 残基代替位置 486 处的野生型 Gln(Q) 残基的突变, 用 Leu(L) 残基代替位置 499 处的野生型残基 Iso(I) 的突变, 用 Asp(D) 或 Glu(E) 残基代替位置 496 处的野生型 Asn(N) 残基的突变(亦分别称作“ELD”和“ELE”结构域)。在其他实施方案中, 工程化的剪切半结构域包括位置 490、538 和 537 处的突变(相对于野生型 Fok I 编号), 例如用 Lys(K) 残基代替位置 490 处的野生型 Glu(E) 残基的突变, 用 Lys(K) 残基代替位置 538 处的野生型 Iso(I) 残基的突变, 用 Lys(K) 或 Arg(R) 残基代替位置 537 处的野生型 His(H) 残基的突变(亦分别被称作“KKK”和“KKR”结构域)。在其他实施方案中, 工程化的剪切半结构域包括位置 490 和 537 处的突变(相对于野生型 Fok I 编号), 例如用 Lys(K) 残基代替位置 490 处的野生型 Glu(E) 残基的突变, 用 Lys(K) 或 Arg(R) 残基代替位置 537 处的野生型 His(H) 残基的突变(亦分别被称作“KIK”和“KIR”结构域)。(参见美国专利公开 No. 20110201055)。本文所述的工程化剪切半结构域可以使用任何合适的方法制备, 例如通过野生型剪切半结构域(Fok I)的定点诱变, 如美国专利公开 Nos. 20050064474; 20080131962; 和 20110201055 所述。

[0166] 或者, 可以使用所谓的“分割(split)-酶”技术(参见, 例如美国专利公开 No. 20090068164)在核酸靶位点处体内组装核酸酶。这样的分割酶的组分或者可以在各别的表达构建体上表达, 或者可以将它们连接在一个开放阅读框中, 其中各个组件被分隔开, 例如, 被自剪切性 2A 肽或 IRES 序列分隔开。组件可以是单独的锌指结合结构域, 或者是大范围核酸酶的核酸结合结构域的结构域。

[0167] C. 锌指核酸酶

[0168] 在具体的实施方案中, 嵌合多肽是定制设计的锌指核酸酶(ZFN), 其可以被设计成递送靶位点特异性的双链 DNA 断裂, 该断裂中可以整合外源核酸或供体 DNA(见共同拥有的美国专利公开 20100257638, 本文引用其内容作为参考)。ZFN 是嵌合多肽, 其包含来自限制性核酸内切酶(例如 Fok I)的非特异性剪切结构域, 和锌指 DNA 结合结构域多肽。见, 例如 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。在一些实施方案中, ZFN 包括非规范的锌指 DNA 结合结构域 (参见共同拥有的美国专利公开 20080182332, 本文引用其内容作为参考)。Fok I 限制性内切酶必须通过核酸酶结构域二聚体化, 以便切割 DNA 和引入双链断裂。因此, 含有来自这类核酸内切酶的核酸酶结构域的 ZFN 也需要核酸酶结构域的二聚体化, 以便切割靶 DNA。Mani et al. (2005) *Biochem. Biophys. Res. Commun.* 334:1191-7; Smith et al. (2000) *Nucleic Acids Res.* 28:3361-9。ZFN 的二聚体化可以被两个相邻的、方向相反的 DNA 结合位点所易化。同上。

[0169] ZFN 系统的灵活性和特异性提供了先前通过已知的重组酶介导基因编辑策略所不能实现的控制水平。作为一个实例, ZFN 可以容易地工程化, 以便, 例如, 识别特定的核酸序列。Wu et al. (2007) *Cell. Mol. Life Sci.* 64:2933-44。(见, 美国专利公开 20090205083, 20110189775, 20110167521 和 20100199389, 本文引用其全部内容作为参考)。对锌指识别残基的密码子的随机化有助于选择对任意选定的 DNA 序列具有高亲和性的新的指。而且, 锌指是天然的 DNA 结合分子, 并且已经显示工程化的锌指可在活细胞中对其设计的靶具有作用。因此, 基于锌指的核酸酶可以靶向到特定的而非任意的识别位点。

[0170] 在特定的实例中, 用于将外源核酸位点特异性地整合到宿主的至少一个 FAD2 性能基因座中的方法包括将 ZFN 引入到宿主的细胞中, 其中该 ZFN 识别并结合靶核苷酸序列, 其中该靶核苷酸序列包含在宿主至少一个 FAD2 基因座内。在某些实例中, 靶核苷酸序列并非包含在宿主基因组除所述至少一个 FAD2 基因座之外的任何其他位置处。例如, ZFN 的 DNA 结合多肽可以被工程化以识别并结合在所述至少一个 FAD2 基因座内被鉴定 (例如通过测序 FAD2 基因座) 的靶核苷酸序列。对于包括将 ZFN 引入到宿主细胞中、用于位点特异性地将外源核酸整合到宿主的至少一个 FAD2 性能基因座中的方法, 其还可以包括将外源核酸引入到细胞内, 其中 ZFN 对靶序列的位点特异性识别和结合 (和随后对包含 FAD2 基因座的核酸的剪切) 使得该外源核酸更容易重组到宿主的包含所述至少一个 FAD2 基因座的核酸之中。

[0171] V. 用于整合在 FAD2 基因座处的外源核酸

[0172] 本发明的实施方案可以包括一个或多个选自下组的核酸: 用于位点特异性整合在至少一个 FAD2 基因座处的外源核酸, 例如但不限于, PTU, ELP, ETIP 或 ORF; 包含编码靶向核酸内切酶的核苷酸序列的核酸; 和包含前述的至少一者或两者的载体。因此, 用于一些实施方案的特定核酸包括编码多肽的核苷酸序列, 结构核苷酸序列, 和 / 或 DNA 结合多肽识别与结合位点。

[0173] A. 用于位点特异性整合的外源核酸分子

[0174] 如上文指出的, 提供了外源序列 (也称作“供体序列”或“供体”或“转基因”) 的整合, 用于例如表达多肽, 修正突变基因, 或用于增加野生型基因的表达。可以显见, 供体序列通常与其被置于的基因组序列不相同。供体序列可以包含非同源序列, 其侧翼是两个具有同源性的区域, 用于在感兴趣的位置处实现高效的 HDR。此外, 供体序列可以包括载体分子, 其含有与细胞染色质中感兴趣的区域不同源的序列。供体分子可以包含多个不连续的

与细胞染色质具有同源性的区域。例如,为了靶向整合通常不存在于感兴趣区域内的序列,所述序列可以存在于供体核酸分子中,并且被与感兴趣区域内的序列具有同源性的序列所侧翼。

[0175] 供体多核苷酸可以是 DNA 或 RNA,单链或双链的,并且可以以线性或环形形式引入到细胞内。见例如,美国专利公开 Nos. 20100047805, 20110281361, 20110207221 和美国专利申请 No. 13/889, 162。如果以线性形式引入,则供体序列的末端可以通过本领域技术人员已知的方法被保护(例如,保护免于核酸外切酶降解)。例如,将一个或多个双脱氧核苷酸残基添加到线性分子的 3' 端和 / 或将自互补的寡核苷酸与一个或两个末端连接。参见,例如,Chang et al. (1987)Proc. Natl. Acad. Sci. USA84:4959-4963;Nehls et al. (1996)Science272:886-889。其它用于保护外源多核苷酸免于降解的方法包括,但不仅限于,添加末端氨基和使用经过修饰的核苷酸间连接键,例如硫代磷酸酯,氨基磷酸酯和 O- 甲基核糖或脱氧核糖残基。

[0176] 多核苷酸可以作为具有额外序列(例如复制起点、启动子和编码抗生素抗性的基因)的载体的一部分被引入到细胞内。而且,供体多核苷酸可以作为裸核酸,作为与脂质体或泊洛沙姆等作用剂复合的核酸被引入,或者可以通过病毒(例如,腺病毒,腺伴随病毒,疱疹病毒,逆转录病毒,慢病毒和整合缺陷的慢病毒(IDLV))被递送。

[0177] 供体一般被整合,使得其表达被整合位点处的内源启动子,即驱动整合了该供体的内源基因(例如 FAD3)的表达的启动子所驱动。然而,容易想到的是,供体可以包含启动子和 / 或增强子,例如组成型启动子,或可诱导的或组织特异性的启动子。

[0178] 而且,尽管不是表达必需的,但是外源序列还可以包含转录或转录调节序列,例如启动子、增强子、绝缘子(insulators)、内部核糖体进入位点、编码 2A 肽和 / 或多腺苷酸化信号的序列。

[0179] 可以在实施方案中位点特异性方式整合到至少一个 FAD2 基因座中从而修饰该 FAD2 基因座的外源核酸包括,例如但不仅限于,包含编码感兴趣多肽的核苷酸序列的核酸;包含农艺学基因的核酸;包含编码 RNAi 分子的核苷酸序列的核酸;或破坏 FAD2 基因的核酸。

[0180] 在一些实施方案中,外源核酸被整合在 FAD2 基因座处,以便修饰该 FAD2 基因座,其中该核酸包括农艺学基因或编码感兴趣多肽的核苷酸序列,使得该农艺学基因或核苷酸序列在宿主中从该 FAD2 基因座表达。在一些实例中,感兴趣的多肽(例如,外来蛋白质)以商业量从编码该感兴趣多肽的核苷酸序列表达。在这样的实例中,感兴趣的多肽可以从宿主细胞、组织或生物量提取。在一些实施方案中,宿主是植物,并且被提供用于商业生产感兴趣多肽的植物材料可以是植物,植物部分,植物组织或植物细胞。在一些实例中,该植物部分可以是植物种子。从植物生物量提取蛋白质可以通过已知的方法实现,其在例如 Heney and Orr(1981)Anal. Biochem. 114:92-6 中有讨论。

[0181] 类似地,农艺学基因可以在被转化的植物细胞、植物和 / 或其后代中表达。例如,植物可以通过特定实施方案的方法被遗传工程化,从而从至少一个 FAD2 基因座表达各种感兴趣的农艺学表型。

[0182] 在一些实施方案中,包含农艺学基因或编码感兴趣多肽的核苷酸序列的核酸可以包括,例如但不仅限于:赋予对害虫或疾病的抗性的基因(见例如, Jones et al. (1994)

Science 266:789(用于抵抗叶霉病菌的番茄 Cf-9 基因的克隆);Martin et al.(1993) Science 262:1432;Mindrinos et al.(1994)Cell 78:1089(用于丁香假单胞菌抗性的 RSP2 基因);PCT 国际专利公开 No. WO 96/30517(对大豆胞囊线虫的抗性);PCT 国际专利公开 No. WO 93/19181);编码苏云金芽孢杆菌蛋白的基因、其衍生物、或以其为模型的合成多肽(见例如 Geiser et al.(1986)Gene 48:109(Bt δ -内毒素基因的克隆和核苷酸序列);而且编码 δ -内毒素基因的 DNA 分子可以从美国典型培养物保藏中心(Manassas, VA)以例如 ATCC 登录号 Nos. 40098;67136;31995;和 31998 购买));编码凝集素的基因(见例如 Van Damme et al.(1994)Plant Molec. Biol. 24:25(多种君子兰甘露糖结合凝集素基因的核苷酸序列);编码维生素结合蛋白的基因,例如链霉亲和素(见 PCR 国际专利公开 No. US93/06487(使用链霉亲和素和链霉亲和素同系物作为针对昆虫害虫的杀幼虫剂));编码酶抑制剂,例如蛋白酶(protease, proteinase)抑制剂、或淀粉酶抑制剂,的基因(见例如 Abe et al.(1987)J. Biol. Chem. 262:16793(水稻半胱氨酸蛋白酶抑制剂的核苷酸序列);Huub et al.(1993)Plant Molec. Biol. 21:985(编码番茄蛋白酶抑制剂 I 的 cDNA 的核苷酸序列);Sumitani et al.(1993)Biosci. Biotech. Biochem. 57:1243(硝孢链霉菌(*Streptomyces nitrosporeus*) α -淀粉酶抑制剂的核苷酸序列)和美国专利 5,494,813);编码昆虫特异性激素或信息素(pheromone)的基因,例如蜕皮激素或保幼激素,其变异体,基于它的模拟物,或其拮抗剂或激动剂(见例如 Hammock et al.(1990)Nature 344:458(杆状病毒表达克隆的保幼激素酯酶,一种保幼激素去活剂);编码在表达时会破坏受影响害虫的生理的昆虫特异性肽或神经肽的基因(见例如 Regan(1994)J. Biol. Chem. 269:9(表达克隆产生编码昆虫利尿激素受体的 DNA);Pratt et al.(1989)Biochem. Biophys. Res. Comm. 163:1243(太平洋折翅蠅(*Diploptera punctata*)中的咽侧体抑制素(allostatin));和美国专利 5,266,317(编码昆虫特异性瘫痪神经毒素的基因);天然由蛇、黄蜂或其他生物产生的编码昆虫特异性毒液的基因(见例如 Pang et al.(1992)Gene 116:165(在植物中异源表达编码蝎子昆虫毒性肽的基因));编码负责单萜、倍半萜烯、类固醇、羟肟酸、苯丙衍生物(phenylpropanoid derivative)或其它具有杀虫活性的分子的超积累的酶的基因;编码参与生物活性分子(例如糖醇解酶、蛋白水解酶、脂肪分解酶、核酸酶、环化酶、转氨酶、酯酶、水解酶、磷酸酶、激酶、磷酸化酶、聚合酶、弹性蛋白酶、几丁质酶或葡聚糖酶)的修饰(包括翻译后修饰)的酶的基因,无论是天然的还是合成的(见,例如 PCT 国际专利公开 No. WO 93/02197(callase 基因的核苷酸序列);而且,含有几丁质酶编码序列的 DNA 分子可以从例如 ATCC 按照登录号 39637 和 67152 获得;Kramer et al.(1993)Insect Biochem. Molec. Biol. 23:691(编码烟草天蛾几丁质酶的 cDNA 的核苷酸序列);和 Kawalleck et al.(1993)Plant Molec. Biol. 21:673(欧芹ubi4-2多聚泛素基因的核苷酸序列));编码刺激信号转导的分子的基因(见,例如 Botella et al.(1994)Plant Molec. Biol. 24:757(绿豆钙调蛋白 cDNA 克隆的核苷酸序列);和 Griess et al.(1994)Plant Physiol. 104:1467(玉米钙调蛋白 cDNA 克隆的核苷酸序列));编码疏水性 moment 肽的基因(见例如 PCT 国际专利公开 No. WO 95/16776(抑制真菌植物病原体的螯素(Tachyplesin)的肽衍生物);和 PCT 国际专利公开 No. WO95/18855(赋予疾病抗性的合成抗微生物肽));编码膜通透酶、通道构成物(former)或通道阻断剂的基因(见例如 Jaynes et al.(1993)Plant Sci 89:43(异源表达天蚕杀菌肽(cecropin)- β 裂解肽类似物,赋予

转基因烟草植物对青枯假单胞菌的抗性));编码病毒侵入蛋白或从其衍生的复合体毒素的基因(见例如Beachy et al. (1990)Ann. rev. Phytopathol. 28:451);编码昆虫特异性抗体或从其衍生的免疫毒素的基因(见例如, Taylor et al., Abstract#497, Seventh Int'l Symposium on Molecular Plant-Microbe Interactions (Edinburgh, Scotland) (1994) (通过产生单链抗体片段在转基因烟草中的酶去活));编码病毒特异性抗体的基因(见例如Tavladoraki et al. (1993)Nature 366:469(表达重组抗体基因的转基因植物被保护免于病毒攻击));编码由病原体或寄生虫自然产生的发育阻断蛋白的基因(见例如, Lamb et al. (1992)Bio/Technology 10:1436(真菌内切 α -1,4-D-多聚半乳糖醛酸酶通过溶解植物细胞壁均聚 α -1,4-D-半乳糖醛酸而易化真菌定植(colonization)和植物营养素释放);Toubart et al. (1992)Plant J. 2:367(克隆和表征编码豆类内切多聚半乳糖醛酸酶抑制蛋白的基因);编码自然界中由植物产生的发育阻断蛋白的基因(见例如Logemann et al. (1992)Bio/Technology 10:305(表达大麦核糖体失活基因的转基因植物具有更高的针对真菌疾病的抗性))。

[0183] 在一些实施方案中,包含农艺学基因或编码感兴趣多肽的核苷酸序列的核酸还可以和/或可选择地包括,例如但不限于:赋予对除草剂(例如抑制生长点或分生组织的除草剂,例如咪唑啉酮或磺酰脲)的抗性的基因(这类示例性基因编码突变的ALS和AHAS酶,例如分别如Lee et al. (1988)EMBO J. 7:1241, 和Miki et al. (1990)Theor. Appl. Genet. 80:449所述);草甘膦抗性,如由例如突变的5-烯醇式丙酮酰莽草酸-3-磷酸合酶(EPSP)所赋予的(通过引入重组核酸和/或天然EPSP基因(包括但不限于CP4, DMMG, 和DGT-28)的各种形式的体内突变)、aroA基因、和草甘膦乙酰转移酶(GAT)基因);其它膦化合物,例如来自链霉菌物种,包括吸水链霉菌和Streptomyces viridichromogenes的草铵膦丝菌素乙酰转移酶(PAT)基因;和吡啶氧基或苯氧基丙酸和环己酮(ACC酶抑制剂编码基因)。见例如,美国专利4940835和6248876(能够赋予植物草甘膦抗性的各种形式EPSP的核苷酸序列)。编码突变体aroA基因的DNA分子能够根据ATCC登录号39256获得。另见美国专利No. 4,769,061(突变体aroA基因的核苷酸序列)。欧洲专利申请No. 0333033和美国专利No. 4,975,374公开了谷氨酰胺合成酶基因的核苷酸序列,其可以赋予对除草剂例如L-丝菌素的抗性。示例性PAT基因的核苷酸序列在欧洲专利申请No. 0242246, 和DeGreef et al. (1989)Bio/Technology 7:61(产生表达编码PAT活性的嵌合bar基因的转基因植物)中提供。赋予对苯氧基丙酸和环己酮(例如稀禾定和盖草)的抗性的示例性基因包括ACC1-S1, ACC1-S2和ACC1-S3基因,如Marshall et al. (1992)Theor. Appl. Genet. 83:435所述。能够赋予草甘膦抗性的GAT基因在例如W02005012515中被描述。赋予对2,4-D、苯氧基丙酸和吡啶氧基生长素除草剂抗性的基因在例如W0 2005107437和W0 2007053482中被描述。

[0184] 包含农艺学基因或编码感兴趣多肽的核苷酸序列的核酸还可以包括,例如但不限于:赋予对抑制光合作用的除草剂的抗性的基因,例如三嗪(psbA和gs+基因)或胍脂(胍水解酶基因)。见例如Przibila et al. (1991)Plant Cell 3:169(用编码突变体psbA基因的质粒转化衣藻)。胍水解酶基因的核苷酸序列在美国专利4,810,648中被公开,含有这些基因的DNA分子可以根据ATCC登录号Nos. 53435;67441;和67442获得。另见Hayes et al. (1992)Biochem. J. 285:173(克隆和表达编码谷胱甘肽S转移酶的DNA)。

[0185] 在一些实施方案中,包含农艺学基因或编码感兴趣多肽的核苷酸序列的核酸还可以和/或可选择地包括赋予或者有助于增值性状的基因,增值性状例如但不限于:经过修饰的脂肪酸代谢,例如,通过用硬脂酰-ACP 去饱和酶的反义基因转化植物以增加植物的硬脂酸含量(见例如 Knultzon et al. (1992) *Proc. Natl. Acad. Sci. U. S. A.* 89:2624);减少植酸盐/酯含量,例如,引入植酸酶编码基因可以提高植酸盐/酯的裂解,向被转化植物增加更多的游离磷酸盐(参见,例如, Van Hartingsveldt et al. (1993) *Gene* 127:87(黑曲霉植酸酶基因的核苷酸序列));可以被引入到玉米中以降低植酸含量的基因,例如,这可以通过克隆然后再次引入与可能负责低植酸水平的玉米突变体特征的单等位基因相关的DNA 实现(见 Raboy et al. (1990) *Maydica* 35:383));和经过修饰的碳水化合物组成,其通过例如用编码可以改变淀粉分支模式的酶的基因转化植物来实现(见例如 Shiroza et al. (1988) *J. Bacteriol.* 170:810(链球菌突变体果糖转移酶基因的核苷酸序列);Steinmetz et al. (1985) *Mol. Gen. Genet.* 20:220(果聚糖蔗糖酶基因);Pen et al. (1992) *Bio/Technology* 10:292(α -淀粉酶);Elliot et al. (1993) *Plant Molec. Biol.* 21:515(西红柿转化酶基因的核苷酸序列);Sogaard et al. (1993) *J. Biol. Chem.* 268:22480(大麦 α -淀粉酶基因);和 Fisher et al. (1993) *Plant Physiol.* 102:1045(玉米胚乳淀粉分支酶 II))。

[0186] 在一些实施方案中,外源核酸被整合在FAD2 基因座处以修饰FAD2 基因座,其中该核酸包括PTU 或ELP,使得,例如,随后的第二外源核酸易于位点特异性整合在PTU 或ELP 的位点处。另见美国专利 No. 13/889,162。

[0187] 为了通过靶向整合在靶向性核酸内切酶的介导下将感兴趣的核酸分子靶向整合到植物基因组中,需要递送靶向性核酸内切酶或者靶向性核酸内切酶编码核酸分子,随后在宿主内表达功能性的靶向性核酸内切酶蛋白。在靶向性核酸内切酶被递送到宿主细胞中或在宿主细胞中表达时,外源核酸也优选地存在于宿主细胞中,从而使该功能性靶向性核酸内切酶蛋白在至少一个FAD2 基因座中的靶位点处诱导双链断裂,这些断裂随后被修复,例如通过同源性驱动的外源核酸整合到基因座中而被修复。本领域的技术人员可以预见,功能性靶向性核酸内切酶蛋白的表达可以通过数种方法实现,包括但不限于,编码靶向性核酸内切酶的构建体的基因转移,和编码靶向性核酸内切酶的构建体的瞬时表达。在这两种情况下,均可以同时实现宿主细胞中功能性靶向性核酸内切酶的表达和外源核酸的递送,以便驱动FAD2 基因座处的靶向整合。

[0188] 在利用ZFN 作为靶向性核酸内切酶的实施方案中获得的一个特别优势在于,对嵌合锌指核酸酶剪切结构域二聚体化的需要赋予高水平的序列特异性,且因此赋予高水平的剪切特异性。因为每个3 个指的组结合9 个连续的碱基对,所以如果每个锌指结构域具有完美的特异性,则2 个嵌合核酸酶实际上要求一个18bp 的靶点。任何给定的这种长度的序列均可预期在单各基因组(推定大约 10^9 bp) 中是唯一的。Bibikova et al. (2001) *Mol. Cell. Biol.* 21(1):289-97;Wu et al. (2007), 前文。而且,额外的指可以提供更高的特异性,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,因此可以增加每个DNA 结合结构域中的锌指数目以提供更高的特异性。例如,通过使用一对可识别24bp 序列的4 指、5 指、6 指或更多指的ZFN,可以进一步提高特异性。

Urnov et al. (2005) Nature 435:646-51。因此,可以使用 ZFN,使得被引入到宿主植物基因组中的识别序列在基因组中是唯一的。

[0189] B. 包含编码靶向性核酸内切酶的核苷酸序列的核酸分子

[0190] 在一些实施方案中,编码靶向性核酸内切酶的核苷酸序列可以通过操作(例如连接)编码该靶向性核酸内切酶中包含的多肽的天然核苷酸序列而工程化。例如,可以对编码包含 DNA 结合多肽的蛋白质的基因的核苷酸序列进行检查,以鉴定对应于该 DNA 结合多肽的基因的核苷酸序列,并可以使用该核苷酸序列作为编码包含该 DNA 结合多肽的靶向性核酸内切酶的核苷酸序列的元件。或者,可以利用靶向性核酸内切酶的氨基酸序列来推导编码该靶向性核酸内切酶的核苷酸序列,例如,根据遗传密码的简并性。

[0191] 在包含编码靶向性核酸内切酶的核苷酸序列的示例性核酸分子中,编码核酸酶多肽的第一多核苷酸序列的最后一个密码子与编码 DNA 结合多肽的第二多核苷酸序列的第一个密码子可以相隔任意数量的核苷酸三联体,例如不编码内含子或“停止(STOP)”。类似地,编码 DNA 结合多肽的第一多核苷酸序列的最后一个密码子与编码核酸内切酶的第二多核苷酸序列的第一个密码子可以相隔任意数量的核苷酸三联体。在这些和进一步的实施方案中,编码核酸酶多肽的第一多核苷酸序列与编码 DNA 结合多肽的第二多核苷酸序列的最后一个(即,核酸序列的最 3') 的最后一个密码子可以和与之直接邻接、或者仅相隔短肽序列(例如由合成的核苷酸接头(例如可能已被用于实现融合的核苷酸接头)编码的短肽序列)的其他多核苷酸编码序列的第一个密码子相位对准地(phase-register)融合。其他多核苷酸序列的实例包括,例如但不限于,标签、靶向肽、和酶剪切位点。类似地,第一和第二多核苷酸序列的最 5' (在核酸序列中) 的第一个密码子可以和与之直接邻接、或者仅相隔短肽序列的其他多核苷酸编码序列的最后一个密码子相位对准地融合。

[0192] 将编码靶向性核酸内切酶中的功能性多肽(例如,DNA 结合多肽和核酸酶多肽)的各多核苷酸序列分隔开来的序列可以,例如,包括任何序列,使得所编码的氨基酸序列不太可能显著改变靶向性核酸内切酶的翻译。由于已知的核酸酶多肽和已知的 DNA 结合多肽的自主性质,在实例中,间插序列不会干扰这些结构的各自功能。

[0193] C. 载体和表达构建体

[0194] 在一些实施方案中,可以将包含至少一个编码感兴趣多肽和/或靶向性核酸内切酶的外源多核苷酸序列的至少一个核酸分子引入到细胞、组织、或生物体中,用于在其中表达。例如,可以将包含编码特异性识别至少一个 FAD2 基因座内包含的核苷酸序列的靶向性核酸内切酶的多核苷酸序列的核酸分子引入到细胞中,使得编码该感兴趣多肽的多核苷酸序列整合到至少一个 FAD2 基因座中(这可以通过例如被表达的靶向性核酸内切酶在该基因座处引入双链断裂后进行同源重组而实现),并从整合的多核苷酸序列表达该感兴趣的多肽。

[0195] 在一些实施方案中,核酸分子,例如前述的其中一个,可以是例如载体系统,其包括例如但不限于,线性质粒或闭环质粒。在特定的实例中,载体可以是表达载体。根据特定实施方案的核酸序列可以,例如,被整合到载体中,使得该核酸序列与一个或多个调节序列可操作地连接。可以获得许多载体用于这一目的,并且特定载体的选择可能取决于,例如,待插入在载体中的核酸的大小,待用该载体转化的特定宿主细胞,和/或任何编码多肽期望被表达的量。载体通常包含各种组件,它们的身份取决于载体的功能(例如,DNA 扩增

或 DNA 表达), 和与载体相容的特定宿主细胞。

[0196] 在一些实施方案中, 与一个或多个编码序列可操作地连接的调节序列可以是在宿主细胞中, 如在细菌细胞、藻类细胞、真菌细胞或植物细胞中发挥功能的启动子序列, 在该细胞中该核酸分子被扩增或表达。一些实施方案可以包括植物转化载体, 其包括含有至少一个调节序列的核苷酸序列, 该调节序列与一个或多个编码感兴趣多肽或靶向性核酸内切酶的核苷酸序列可操作地连接, 其中该一个或多个核苷酸序列可以在该调节序列的控制下在植物细胞、组织或生物体中表达, 从而产生感兴趣的多肽或靶向性核酸内切酶。

[0197] 根据一些实施方案, 适用于核酸分子的启动子包括可诱导的、组织特异性的、病毒、合成或组成型启动子, 它们都是本领域众所周知的。可用于本发明实施方案的启动子的非限制性实例在下列文献中提供: 美国专利 Nos. 6, 437, 217(玉米 RS81 启动子); 5, 641, 876(水稻肌动蛋白启动子); 6, 426, 446(玉米 RS324 启动子); 6, 429, 362(玉米 PR-1 启动子); 6, 232, 526(玉米 A3 启动子); 6, 177, 611(组成型玉米启动子); 5, 322, 938, 5, 352, 605, 5, 359, 142, 和 5, 530, 196(35S 启动子); 6, 433, 252(玉米 L3 油质蛋白启动子); 6, 429, 357(水稻肌动蛋白 2 启动子, 和水稻激动蛋白 2 内含子); 6, 294, 714(光诱导的启动子); 6, 140, 078(盐诱导的启动子); 6, 252, 138(病原体诱导的启动子); 6, 175, 060(磷缺乏诱导的启动子); 6, 388, 170(双向启动子); 6, 635, 806(γ -蕈苣辛(coixin)启动子); 5, 447, 858(大豆热激蛋白启动子); 和美国专利申请系列 No. 09/757, 089(玉米叶绿体醛缩酶启动子)。

[0198] 额外的示例性启动子包括胭脂碱合成酶(NOS)启动子(Ebert et al. (1987) Proc. Natl. Acad. Sci. USA 84(16):5745-9); 章鱼碱合成酶(OCS)启动子(其由根癌土壤杆菌的肿瘤诱导质粒携带); 花椰菜花叶病毒属启动子, 例如花椰菜花叶病毒(CaMV) 19S 启动子(Lawton et al. (1987) Plant Mol. Biol. 9:315-24); CaMV 35S 启动子(Odell et al. (1985) Nature 313:810-2); 玄参花叶病毒 35S 启动子(Walker et al. (1987) Proc. Natl. Acad. Sci. USA 84(19):6624-8); 蔗糖合酶启动子(Yang and Russell(1990) Proc. Natl. Acad. Sci. USA 87:4144-8); R 基因复合物启动子(Chandler et al. (1989) Plant Cell 1:1175-83); 叶绿素 a/b 结合蛋白基因启动子; CaMV35S(美国专利 Nos. 5, 322, 938, 5, 352, 605, 5, 359, 142, 和 5, 530, 196); FMV35S(美国专利 Nos. 6, 051, 753, 和 5, 378, 619); PC1SV 启动子(美国专利 No. 5, 850, 019); SCP1 启动子(美国专利 No. 6, 677, 503); 和 AGRtu.nos 启动子(GenBank 登录号 No. V00087; Depicker et al. (1982) J. Mol. Appl. Genet. 1:561-73; Bevan et al. (1983) Nature 304:184-7)。

[0199] 在特定的实施方案中, 核酸分子可以包括组织特异性启动子。组织特异性启动子是指导可操作地连接的核苷酸序列在该启动子特异的组织中以相对于生物体的其它组织更高的水平进行转录的核苷酸序列。组织特异性启动子的实例包括, 但不限于: 绒毡层特异性启动子; 花药特异性启动子; 花粉特异性启动子(见例如美国专利 No. 7, 141, 424 和国际 PCT 公开 No. WO 99/042587); 胚珠特异性启动子;(见例如, 美国专利申请 No. 2001/047525A1); 果实特异性启动子(见例如, 美国专利 Nos. 4, 943, 674, 和 5, 753, 475); 和种子特异性启动子(见例如, 美国专利 Nos. 5, 420, 034, 和 5, 608, 152)。在一些实施方案中, 可以使用发育阶段特异性启动子(例如, 在发育晚期有活性的启动子)。

[0200] 其他在一些实施方案中可以和核酸分子可操作地连接的调节序列包括位于启动

子序列和编码序列之间的 5' UTR,其作为翻译前导序列。该翻译前导序列存在于完全加工好的 mRNA 中,并且可以影响初级转录本的加工,和 / 或 RNA 稳定性。翻译前导序列的实例包括玉米和矮牵牛热激蛋白前导 (美国专利 No. 5, 362, 865),植物病毒外被蛋白前导,植物核酮糖-1, 5-二磷酸羧化酶 / 加氧酶前导,和其他。参见,例如,Turner and Foster(1995) *Molecular Biotech.* 3(3):225-36。下面提供了 5' UTR 的非限制性实例:GmHsp(美国专利 NO. 5, 659, 122);PhDnaK(美国专利 NO. 5, 362, 865);AtAnt1;TEV(Carrington and Freed(1990) *J. Virol.* 64:1590-7);和 AGRinos(GenBank 登录号 NO. V00087;和 Bevan et al. (1983),同上)。

[0201] 其他在一些实施方案中可以和核酸分子可操作地连接的调节序列还包括 3' 非翻译序列,3' 转录终止区,或多腺苷酸化区。这些都是位于核苷酸序列下游的遗传元件,并且包括这样的多核苷酸,其提供多腺苷酸化信号,和 / 或其他能够影响转录或 mRNA 加工的调节信号。植物中多腺苷酸化信号的功能是向 mRNA 前体的 3' 末端添加聚腺苷酸核苷酸。多腺苷酸化序列可以来自多种植物基因或来自 T-DNA 的基因。3' 转录终止区的非限制性实例是胭脂碱合成酶 3' 区(nos 3';Fraley et al. (1983) *Proc. Natl. Acad. Sci. USA* 80:4803-7)。在 Ingelbrecht et al. (1989) *Plant Cell* 1:671-80 中提供了使用不同 3' 非翻译区的实例。多腺苷酸化信号的非限制性实例包括来自豌豆 RbcS2 基因 (Ps. RbcS2-E9;Coruzzi et al. (1984) *EMBO J.* 3:1671-9) 和 AGRtu.nos(GenBank 登录号 E01312) 的多腺苷酸化信号。

[0202] 关于可能用于特定实施方案的调节序列的额外信息在,例如 Goeddel (1990)“Gene Expression Technology,”*Methods Enzymol.* 185, Academic Press, San Diego, CA 中有描述。

[0203] 重组核酸分子或载体可以包括可选择标志物,其给被转化的细胞,例如植物细胞,赋予可选择的表型。可选择标志物还可用于选择包含含有可选择标志物的核酸分子的细胞或生物体。标志物可以编码杀生物剂抗性,抗生素抗性(例如,卡那霉素,遗传霉素(G418),博来霉素,和潮霉素),或除草剂抗性(例如,草甘膦)。可选择标志物的实例包括,但不限于:neo 基因,其赋予卡那霉素抗性并可以使用例如卡那霉素和 G418 进行选择;bar 基因,其赋予双丙氨膦抗性;突变的 EPSP 合酶基因,其赋予草甘膦抗性;腈水解酶基因,其赋予对溴苯腈的抗性;突变的乙酰乳酸合酶基因(ALS),其赋予咪唑啉酮或磺酰脲抗性;和氨甲蝶呤抗性 DHFR 基因。可以获得多种可选择标志物,其赋予对化学剂的抗性,包括例如但不限于,氨苄青霉素;博来霉素;氯霉素;庆大霉素;潮霉素;卡那霉素;林可霉素;甲氨蝶呤;膦丝菌素;嘌呤霉素;壮观霉素;利福平;链霉素;和四环素。这样的可选择标志物的实例在例如,美国专利 5, 550, 318;5, 633, 435;5, 780, 708 和 6, 118, 04 中被举例。

[0204] 核酸分子或载体还可以或者可选择地包括可筛选标志物。可筛选标志物可用于监测表达。示例性的可筛选标志物包括 β -葡萄糖醛酸酶或 uidA 基因(GUS),其编码一种酶,该酶的各种生色底物已知的(Jefferson et al. (1987) *Plant Mol. Biol. Rep.* 5:387-405);R-基因座基因,其编码可调节植物组织中花青素色素(红色)生产的产物(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(263-82 页); β -内酰胺酶基因(Sutcliffe et al. (1978) *Proc. Natl. Acad. Sci. USA* 75:3737-41);编码各种生色底物已知的酶的基因(例如,

PADAC, 发色头孢菌素); 荧光素酶基因 (Ow et al. (1986) Science 234:856-9); xyle 基因, 其编码可以转化显色儿茶酚的儿茶酚双加氧酶 (Zukowski et al. (1983) Gene 46(2-3):247-55); 淀粉酶基因 (Ikata et al. (1990) Bio/Technol. 8:241-2); 酪氨酸酶基因, 其编码能够氧化酪氨酸为 DOPA 和多巴醌的酶, 其进一步可缩合成黑色素 (Katz et al. (1983) J. Gen. Microbiol. 129:2703-14); 和 α -半乳糖苷酶。

[0205] 编码例如感兴趣的特定多肽或特定靶向性核酸内切酶的所有核苷酸序列, 可以被本领域的技术人员立即识别。遗传密码子的简并性为特定的氨基酸序列提供了有限数目的编码序列。根据本发明的实施方案选择编码多肽的特定序列在从业者的判断力之内。在不同的应用中可能期望不同的编码序列。

[0206] 在一些实施方案中, 可能期望修饰核酸的核苷酸, 例如提高该核酸中包含的多核苷酸序列在特定宿主中的表达。遗传密码子是冗余的, 具有 64 种可能的密码子, 但是大多数生物体优先使用这些密码子的子集。在物种中最经常使用的密码子被称作最佳密码子, 不是非常经常使用的密码子被归类为稀有或低使用率密码子。Zhang et al. (1991) Gene 105:61-72。密码子可以被取代, 以反映特定宿主的优选密码子使用, 这个过程有时被称作“密码子优化”。含有特定原始生物或真核生物宿主优选的密码子的优化编码序列可以被制备, 以提高翻译速度或产生具有期望性质的重组 RNA 转录本 (例如, 与从非优化序列产生的转录本相比, 具有更长的半衰期)。

[0207] 在本发明的实施方案中, 核酸可以用本领域技术人员已知的任何方法引入到宿主细胞中, 包括例如但不限于: 原生质体转化 (见例如美国专利 5, 508, 184); 干燥 / 抑制介导的 DNA 摄取 (见例如, Potrykus et al. (1985) Mol. Gen. Genet. 199:183-8); 电穿孔 (见例如, 美国专利 5, 384, 253); 用碳化硅纤维搅拌 (见例如, 美国专利 5, 302, 523 和 5, 464, 765); 土壤杆菌介导的转化 (见例如, 美国专利 5, 563, 055, 5, 591, 616, 5, 693, 512, 5, 824, 877, 5, 981, 840, 和 6, 384, 301); 并通过 DNA 涂覆的颗粒加速 (见例如, 美国专利 5, 015, 580, 5, 550, 318, 5, 538, 880, 6, 160, 208, 6, 399, 861, 和 6, 403, 865)。通过诸如这些技术的应用, 几乎任何种类的细胞均可以稳定地转化。在一些实施方案中, 转化 DNA 被整合到宿主细胞的基因组中。在多细胞物种的情况下, 转基因细胞可以再生为转基因生物。任何这些技术均可用于产生转基因植物, 例如, 在该转基因植物的基因组中包含一个或多个本发明的核酸序列。

[0208] 用于将表达载体导入到植物中的最广泛使用的方法是基于土壤杆菌的天然转化系统。根癌土壤杆菌和发根土壤杆菌是植物致病性土壤细菌, 其可以遗传转化植物细胞。根癌土壤杆菌和发根土壤杆菌的 Ti 和 Ri 质粒分别携带负责植物遗传转化的基因。Ti (肿瘤诱导) 质粒含有一个大片段, 被称为 T-DNA, 其可被转移到被转化的植物。Ti 质粒的另一个片段, vir 区域, 负责 T-DNA 转移。T-DNA 区被左手和右手边界包围, 其每一个由末端重复核苷酸序列构成。在一些经过修饰的二元载体中, 肿瘤诱导的基因已被删除, 并使用 vir 区域的功能转移被 T-DNA 边界序列包围的外源 DNA。T 区还可以含有, 例如, 用于高效回收转基因植物和细胞的可选择标志物, 用于插入转移序列的多克隆位点, 例如编码本发明融合蛋白的核酸。

[0209] 因此, 在一些实施方案中, 植物转化载体来自根癌土壤杆菌的 T_i 质粒 (见例如, 美国专利 Nos. 4, 536, 475, 4, 693, 977, 4, 886, 937, 和 5, 501, 967; 和欧洲专利 EP 0122791), 或

发根土壤杆菌的 R_1 质粒。额外的植物转化载体包括,例如但不限于,下列文献中描述的那些:Herrera-Estrella et al. (1983) *Nature* 303:209-13; Bevan et al. (1983), 上文; Klee et al. (1985) *Bio/Technol.* 3:637-42; 和欧洲专利 EP 0120516, 和从上述任意一种衍生的那些。其它细菌,例如中华根瘤菌,根瘤菌,和中慢生根瘤菌,其与植物自然相互作用,可以被修饰从而介导向大量不同植物转移基因。这些植物相关的共生细菌可以被制成感受态,通过获取卸甲 (disarmed) 的 T_1 质粒和合适的二元载体进行基因转移。

[0210] 向受体细胞提供外源 DNA 以后,一般对被转化细胞进行鉴定,用于进一步的培养和植株再生。为了提高鉴定转化细胞的能力,可能期望在用于产生转化体的载体中采用可选择或可筛选的标志物基因,如前所述。在使用可选择标志物的情况下,通过将细胞暴露于一种或多种选择剂,在潜在的被转化细胞中鉴定出被转化的细胞。在使用可筛选标志物的情况下,可以根据期望的标志物基因性状对细胞进行筛选。

[0211] 暴露于选择剂后存活的细胞或者在筛选试验中被评定为阳性的细胞,可以在支持植物再生的培养基中进行培养。在一些实施方案中,可以通过包含其它物质,例如生长调节剂,对任何合适的植物组织培养基(例如 MS 和 N6 培养基)进行修饰。组织可以保持在具有生长调节剂的基本培养基中,直到获得足够的组织来开始再生植物的工作,或者进行反复多轮的手动选择,直到组织的形态适合于再生(例如,至少 2 周),然后转移到有利于芽(shoot)形成的培养基中。培养物被周期性地转移,直到产生足够的芽苗。一旦芽苗形成,便将它们转移到有利于根形成的培养基中。一旦形成足够的根,便将植物转移到土壤中,用于进一步的生长和成熟。

[0212] 为了确认在再生植物中存在感兴趣的核酸分子(例如,编码含有至少一个本发明融合蛋白的多肽的核苷酸序列),可以执行各种测定。这样的测定包括,例如:分子生物学测定,例如 Southern 和 Northern 印迹,PCR 和核酸测序;生物化学测定,例如检测蛋白质产物的存在,例如,通过免疫学手段(ELISA 和/或 Western 印迹)或通过酶促功能;植物部分测定,例如叶或根测定;和整个再生植物的表型分析。

[0213] 整合事件可以通过,例如,PCR 扩增进行分析,使用例如特异针对感兴趣核苷酸序列的寡核苷酸引物。PCR 基因分型被理解为包括,但不限于,对来自预测含有被整合到基因组内的感兴趣核酸分子的分离宿主植物组织的基因组 DNA 进行聚合酶链反应(PCR),随后对 PCR 扩增产物进行标准克隆和序列分析。PCR 基因分型的方法已经得到了很好的描述(参见,例如,Rios, G. et al. (2002) *Plant J.* 32:243-53),并且可应用于来自任何植物物种或组织类型(包括细胞培养物)的基因组 DNA。

[0214] 使用土壤杆菌依赖的转化方法形成的转基因植物通常含有重组 DNA 的单个到多个拷贝。单个重组 DNA 序列被称作一个“转基因事件”或“整合事件”。这类转基因植物就插入的 DNA 序列而言是杂合的。在一些实施方案中,转基因的转基因植物纯合子可以通过含有单个外源基因序列的独立分离的转基因植物与自身(例如 F_0 植物)的有性结合(自交)产生 F_1 种子而获得。四分之一的所产 F_1 种子就转基因而言是纯合的。令 F_1 种子萌发可产生能够用于测试杂合性的植物,这通常使用 SNP 测定或热扩增测定,其能够区分杂合子和纯合子(即,接合性测定)。

[0215] 除了在一些实施方案中用核酸分子直接转化植物或植物细胞之外,在特定的实施方案中,可以通过使具有至少一个转基因事件的第一植物与缺少这种事件的第二植物杂

交,来制备转基因植物。例如,可以向适于转化的第一植物品系中引入包含至少一个经过修饰的 FAD2 基因座的核酸,其中外源核酸已经以位点特异性的方式被整合,从而产生转基因植物,该转基因植物可以和第二植物品系杂交,从而将至少一个经过修饰的 FAD2 基因座(以及因此导致的外源核酸)基因渗入到第二植物品系中。

[0216] 为了确认在再生植物中存在感兴趣的核酸分子,可以实施多种测定。这样的测定包括,例如:分子生物学测定,例如 Southern 和 Northern 印迹和 PCR;生物化学测定,例如检测蛋白质产物的存在,例如,通过免疫学手段(ELISA 和/或 Western 印迹)或通过酶促功能;植物部分测定,例如叶或根测定;和整个再生植物的表型分析。

[0217] 靶向整合事件可以通过例如 PCR 扩增进行筛选,使用例如特异针对感兴趣核酸分子的寡核苷酸引物。PCR 基因分型被理解为包括,但不仅限于,对来自预测含有被整合到基因组内的感兴趣核酸的分离宿主植物愈伤组织的基因组 DNA 进行聚合酶链反应(PCR)扩增,随后对 PCR 扩增产物进行标准克隆和序列分析。PCR 基因分型的方法已经有很多描述(例如, Rios, G. et al. (2002) Plant J. 32:243-53),并且可以应用于来自任何植物物种或组织类型(包括细胞培养物)的基因组 DNA。结合靶序列和被引入的序列二者的寡核苷酸引物的组合可以顺次使用或复用于 PCR 扩增反应。设计成退火至靶位点、引入的核酸序列、和/或两者的组合的寡核苷酸引物是可行的。因此,PCR 基因分型策略可以包括(但不仅限于)扩增植物基因组中的特定序列,扩增植物基因组中的多个特定序列,扩增植物基因组中的非特异性的序列,或其组合。本领域的技术人员可以设计出额外的引物和扩增反应的组合来查询基因组。例如,一组正向和反向寡核苷酸引物可以被设计为与被引入的核酸序列的边界之外的靶标所特有的核酸序列退火。

[0218] 正向和反向寡核苷酸引物可以被设计成与引入的感兴趣核酸分子特异性退火,例如,与对应于感兴趣核酸分子编码区的序列或感兴趣核酸分子其它部分的序列退火。这些引物可以和上述的引物一起使用。寡核苷酸引物可以根据期望的序列来合成,并且可以商购获得(例如,从 Integrated DNA Technologies, Inc., Coralville, IA)。扩增之后可以进行克隆和测序,或者对扩增产物进行直接序列分析。本领域技术人员可以想到替代方法,用于对在 PCR 基因分型期间产生的扩增产物进行分析。在一个实施方案中,在 PCR 扩增中使用特异针对靶基因的寡核苷酸引物。

[0219] VI. 包含整合在 FAD2 性能基因座处的核酸的转基因植物和植物材料

[0220] 在一些实施方案中,提供了转基因植物,其中该植物包括含有至少一个经过修饰的 FAD2 基因座(例如,被破坏和/或靶向整合了外源序列的 FAD2 基因座)。在特定的实施方案中,这样的植物可以通过转化植物组织或植物细胞,然后再生整株植物来产生。在进一步的实施方案中,这样的植物可以通过以位点特异性的方式在至少一个 FAD2 基因座处引入外源核酸,或者通过将经过修饰的 FAD2 基因座基因渗入到种质中而获得。还提供了包含这样的植物细胞的植物材料。这样的植物材料可从包含该植物细胞的植物获得。

[0221] 在一些实施方案中,包含含有至少一个经过修饰的 FAD2 基因座的植物细胞的转基因植物或植物材料可以表现出这样的或多个特征:在植物细胞内表达靶向性核酸内切酶;在植物细胞内(或者在其中的质体内)表达感兴趣的多肽;在植物细胞的细胞核内表达靶向性核酸内切酶;靶向性核酸内切酶定位在植物细胞内;整合在植物细胞基因组中的 FAD2 基因座处;编码感兴趣多肽的核苷酸序列或农艺学基因整合在植物细胞基因组中

的 FAD2 基因座处 ;和 / 或存在与整合在植物细胞基因组 FAD2 基因座处的编码序列对应的 RNA 转录本。这样的植物可以额外具有一个或多个期望的性状,包括例如但不限于,通过内源或转基因核苷酸序列表达产生的性状,表达受到感兴趣多肽或整合在植物细胞基因组 FAD2 基因座处的农艺学基因调节的性状 ;抗昆虫、其它害虫和致病剂 ;抗除草剂 ;增加稳定性、产量或保质期 ;环境耐受 ;药物生产 ;工业产品生产 ;和营养增强。

[0222] 根据本发明的转基因植物可以是任何能够根据本文所述的方法被核酸转化、该核酸随后整合到至少一个 FAD2 基因座中的植物。因此,植物可以是双子叶植物或单子叶植物。可用于本方法的双子叶植物的非限制性实例包括拟南芥,紫花苜蓿,豆类,西兰花 (broccoli), 卷心菜, 芥花, 胡萝卜, 花椰菜 (cauliflower), 芹菜, 大白菜, 棉花, 黄瓜, 茄子, 莴苣, 甜瓜, 豌豆, 胡椒, 花生, 马铃薯, 南瓜 (pumpkin), 萝卜, 油菜, 菠菜, 大豆, 倭瓜 (squash), 甜菜, 向日葵, 烟草, 番茄, 和西瓜。可用于本方法的单子叶植物的非限制性实例包括玉米, 大麦, 洋葱, 水稻, 高粱, 小麦, 黑麦, 粟, 甘蔗, 燕麦, 黑小麦, 稷, 和草坪草。根据本发明的转基因植物可以用任何方式使用或培养。

[0223] 一些实施方案还提供了由本发明的转基因植物生产的商业产品。商业产品包括,例如但不限于 :含有一个或多个整合在至少一个 FAD2 基因座处的核苷酸序列的植物的食品、谷物粗粉 (meal)、油,或压碎的或全谷粒或种子。在一种或多种商品或商业产品中检测到一个或多个这样的核苷酸序列事实上证明该商品或商品产物的至少一部分是由根据本发明的实施方案产生的转基因植物产生的。在一些实施方案中,包含在基因组中含有至少一个经过修饰的 FAD2 基因座的植物细胞的转基因植物或种子,可能在其基因组中含有至少一个其它的转基因事件,包括但不限于 :RNAi 分子被转录的转基因事件 ;编码杀虫蛋白 (例如,苏云金芽孢杆菌杀虫蛋白) 的基因 ;除草剂抗性基因 (例如,提供对草甘膦抗性的基因) ;和有助于转基因植物的期望表型的基因 (例如,增加产量,改变脂肪酸代谢,或恢复细胞质雄性不育)。

[0224] 包含含有至少一个经过修饰的 FAD2 基因座的植物细胞的转基因植物可以具有一种或多种期望的性状。这样的性状可以包括,例如 :抗昆虫、其它害虫和致病剂 ;抗除草剂 ;增加稳定性、产量或保质期 ;环境耐受 ;药物生产 ;工业产品生产 ;和营养增强。期望的性状可以由一个或多个通过靶向重组被整合在 FAD2 基因座处的核酸分子赋予,这些核酸分子在显示该期望表型的植物中表达。因此,在一些实施方案中,期望的性状可能是由于植物中存在转基因,该转基因被引入在植物基因组中至少一个经过修饰的 FAD2 基因座位置处。在另外的实施方案中,期望的性状可以通过常规育种获得,该性状可以由一个或多个通过靶向重组被整合在至少一个经过修饰的 FAD2 基因座处的核酸分子赋予。

[0225] 根据本发明的转基因植物可以按照任何方式加以使用和培养,其中至少一个经过修饰的 FAD2 基因座的存在是期望的。因此,植物可以如下所述工程化而 (除其他事项外) 具有一种或多种期望的性状 :用核酸分子进行转化,该核酸分子随后根据本发明以位点特异性的方式被整合在至少一个 FAD2 基因座处,并按照本领域技术人员已知的任何方法进行种植 (crop) 和培养。

[0226] VII. 包含整合在 FAD2 性能基因座处的核酸的转基因植物的标志物辅助育种

[0227] 提供了与芸苔属植物中的 *fad2* 连锁 (例如,紧密连锁) 的分子标志物。例如,鉴定了含有涉及 H0 性状 (*fad2*) 的序列的 DNA 片段。这些片段位于和基因连锁群中突变等位

基因连锁（例如，紧密连锁）的标志物的周围和所述标志物之间。因此，还提供了包含具有失活突变的突变体 FAD2 基因的核酸分子。所鉴定的片段及其标志物被包含在本主题中，这部分是由于它们在欧洲油菜基因组连锁群中的位置。例如，FAD2 和与之连锁的分子标志物可以位于连锁群 N5 和 N1 中。

[0228] 本文引用在这里引用的全部参考文献，包括出版物、专利和专利申请，的内容作为参考，它们与本公开的具体细节没有冲突，因此在这里引用它们的内容就如同每一篇参考文献被单独且具体地引用并且在本文中阐述了其全部内容一样。提供本文所讨论的参考文献仅仅是由于它们在本申请的申请日期之前被公开。本文的任何内容均不应被理解为由于先前发明而承认本发明人没有资格先于这些公开内容。提供了这样的实施例以举例说明某些特定的特征和 / 或实施方案。这些实施例不应被理解为将本公开限制于所举例的这些特定的特征或实施方案。

实施例

[0229] 实施例 1: 从细菌人工染色体文库鉴定 FAD2 靶序列

[0230] BAC 文库构建

[0231] 细菌人工染色体 (BAC) 文库来自商业供应商 (Amplicon Express, Pullman, WA)。BAC 文库包含 110,592 个 BAC 克隆，其含有从欧洲油菜 L. var. DH10275 分离的高分子量基因组 DNA (gDNA) 片段。gDNA 用 BamHI 或 HindIII 限制酶消化。将大约 135Kbp 的分离 gDNA 片段连接到 pCC1BAC 载体 (Epicentre, Madison, WI) 中，并转化进入大肠杆菌 DH10B 株 (Invitrogen)。BAC 文库由相同数目的用两种不同限制酶构建的 BAC 克隆构成。这样，HindIII 构建的 BAC 文库包含在 144 个单独的 384 孔板中。类似地，BamHI 构建的 BAC 文库包含在 144 个单独的 384 孔板中。总共分离了 110,592 个 BAC 克隆，并排列在 288 个单独的 384 孔板中。288 个单独的 384 孔板的每一个均由供应商作为用于快速 PCR 筛选的单个 DNA 提取体系提供。最终的 BAC 文库覆盖大约 15Gbp 的 gDNA，其对应于欧洲油菜 L. var. DH10275 基因组的 12 倍基因组覆盖（欧洲油菜 L. 的基因组估计为大约 1.132Gbp，如 Johnston et al. (2005) Annals of Botany 95:229-235 所述）。

[0232] 从 BAC 文库分离的 FAD2 编码序列的序列分析

[0233] 使用构建的 BAC 文库分离 FAD2 基因编码序列。进行测序实验从欧洲油菜 L. var. DH10275 中鉴定 4 个 FAD2 基因同源和旁系同源物的特定基因序列。

[0234] FAD2 基因序列最初是在模式物种拟南芥中被鉴定的。该基因序列收录在 Genbank 中，基因座标签为 At3g12120。模式植物物种拟南芥与二倍体芜菁（其是四倍体欧洲油菜的一个祖先）之间的比较基因组关系先前已有描述 (Schrantz et al. (2006) Trends in Plant Science 11(11):535-542)。具体就 FAD2 基因而言，比较分析预测在二倍体欧洲油菜基因组中可能存在该基因的 3-4 个拷贝。Scheffler et al. (1997) Theoretical and Applied Genetics 94:583-591 完成了进一步的基因定位研究。这些基因定位研究的结果表明，欧洲油菜中存在 FAD3 基因的 4 个拷贝。

[0235] 从欧洲油菜 L. var. DH10275 构建的 BAC 文库的测序分析结果分离了 4 个 BAC 序列 (SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, 和 SEQ ID NO:4)，从中确定了 FAD2A (SEQ ID NO:5)，FAD2-1 (SEQ ID NO:6)，FAD2-2 (SEQ ID NO:7)，和 FAD2-3 (SEQ ID NO:8) 基因的编码

序列。FAD2A, FAD2-1, FAD2-2, 和 FAD2-3 的基因序列被鉴定和基因定位。4 个 FAD2 基因的序列分析使用序列比对程序和邻接树利用同一性百分比加以实施。序列比对使用 Vector NTI Advance 11.0 计算机程序 (Life Technologies, Carlsbad, CA) 的 AlignX® 程序实施, 如图 1 所示。AlignX® 使用经过修改的 Clustal W 算法生成蛋白质或核酸序列的多个序列比对结果, 用于相似性比较和注释。。

[0236] 邻接树用 Jalview v2.3® 软件产生, 并如图 2 所示。(Waterhouse et al. (2009) Bioinformatics 25(9)1189-1191)。如图 2 所示, 分离序列的分析表明, FAD2A 和 FAD2-3 序列具有高水平的序列相似性, 并且, 类似地, FAD2-1 和 FAD2-2 具有高水平的序列相似性。这四个序列可以分成 2 个进化枝 (clade), 其中 FAD2A 和 FAD2-3 构成第一枝, FAD2-1 和 FAD2-2 构成第二枝。

[0237] 接着, 使用从欧洲油菜新分离的 FAD2 序列对从芜菁基因组 BAC 文库和甘蓝乌枪基因组序列读取分离的基因组文库进行 BLAST。芜菁和甘蓝均是欧洲油菜的二倍体祖先, 欧洲油菜是双二倍体物种 (AC 基因组, $n = 19$)。欧洲油菜源自芜菁 (A 亚基因组, $n = 10$) 和甘蓝 (C 亚基因组, $n = 9$) 之间的一次新近杂交事件。使用 BLASTn 分析将二倍体祖先序列与从欧洲油菜分离的 4 个不同 FAD2 编码序列进行比较。这种序列分析从芜菁和甘蓝中鉴定了特异性的、经过注释的基因序列, 它们与新近发现的欧洲油菜 FAD2 序列具有最高的序列相似性。表 1 列举了新近鉴定的 FAD2 编码序列和相应的祖先参考序列登录号和源生物体。

[0238] 表 1: 来自欧洲油菜的 FAD2 序列和相应的祖先生物和 FAD 序列登录号

[0239]

分离基因序列	祖先生物和序列登录号	
FAD2A	芜菁	Genbank 登录号 No: KBrB063G23 (A05)
FAD2-3	甘蓝	Genbank 登录号 No: GSS23580801
FAD2-1	芜菁	Genbank 登录号 No: KBrB130I19
FAD2-2	甘蓝	Genbank 登录号 No: GSS 17735412

[0240] FAD2 基因存在于欧洲油菜基因组中, 每个基因在每个亚基因组中具有 2 个拷贝。每个基因的一个拷贝位于 A 亚基因组, 类似地, 每个基因的一个拷贝位于 C 亚基因组。描述了新的命名规则来指示每个基因位于哪个亚基因组上。从欧洲油菜 BAC 基因组 DNA 文库分离的 4 个不同 FAD2 编码序列与祖先序列数据之间的高水平的序列相似性提示 FAD2-3 是来自 C 亚基因组的 FAD2 序列的复制品, 并可以重新标记为 FAD2C; FAD2-1 是来自 A 亚基因组的 FAD2 序列的复制品, 因此可以标记为 FAD2A'; 最后, FAD2-2 是第二个从 C 亚基因组的 FAD2 序列复制的拷贝, 并可以标记为 FAD2C'。

[0241] 基于 PCR 的筛选

[0242] 设计了一系列 PCR 引物用于筛选前述的 BAC 文库。引物或者被设计成通用引物, 其可扩增该基因家族的所有成员, 或者被设计成基因特异性引物, 用于靶向等位基因扩增。PCR 引物被设计成 20bp 长 (+/-1bp), 并且 G/C 含量为 50% (+/-8%)。表 2 列举了设计并合成的引物。汇集 BAC 文库的克隆, 并通过聚合酶链式反应 (PCR) 进行筛选。

[0243] 表 2: 设计用于 BAC 文库筛选以进行 FAD2 基因鉴定的 PCR 引物序列

[0244]

引物名称	SEQ ID NO:	序列
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

[0245] 使用两组不同的条件进行聚合酶链式反应 (PCR)。第一组 PCR 反应条件包括: 1X PCR 缓冲液 (含有 dNTPs); 1.5mM $MgCl_2$; 200 μM 的 0.25U **Immolase®** DNA 聚合酶 (Bioline, London, UK); 250nM 每种引物; 和大约 5–10ng 模板 DNA。第二组 PCR 反应条件被开发用于扩增基因组 DNA, 并且包括: 5–10ng 基因组 DNA, 1X PCR 缓冲液, 2mM dNTPs, 0.4 μM 正向和反向引物, 和 0.25U **Immolase®** DNA 聚合酶 (Bioline, London, UK)。将扩增物汇集至终体积为 13 μL , 并使用 MJ **PTC200®** 热循环仪 (BioRad, Hercules, CA) 或 ABI 9700Gene Amp **System®** (Life Technologies, Carlsbad, CA) 进行扩增。对特定平板的基于 PCR 的筛选使用 4 维筛选方法实施, 其是基于 Bryan 等人 (Scottish Crops Research Institute annual report: 2001–2002) 所述的系统并使用上述 PCR 条件。在对汇集的 BAC 文库进行基于 PCR 的筛选之后, 使用直接 Sanger 测序方法对扩增的 PCR 产物进行测序。扩增的产物用乙醇、乙酸钠和 EDTA 按照 **BigDye®** v3.1 方案 (Applied Biosystems) 进行纯化, 并在 **ABI3730xl®** 自动毛细管电泳平台上进行电泳。

[0246] 在基于 PCR 的筛选和构象 Sanger 测序之后, 鉴定含有各种不同 FAD2 基因家族成员的平板集合。总共鉴定了 4 个独特的 FAD2 旁系同源基因序列 (表 2)。每个 FAD2 旁系同源物基因序列的总共 2 个平板被选择进行平板筛选, 以鉴定平板内含有 FAD2 基因的特定孔和克隆 (表 3)。为两个平板均鉴定特定的孔, 并为 FAD2 基因家族的每个成员选择一个单独的克隆。

[0247] 表 3: 鉴定与精细 (detailed) PCR 引物组合具有阳性反应的 BAC 克隆平板和继续用于在平板内进行克隆鉴定的两个平板的身份

[0248]

基因名称	引物组	阳性平板池	被选平板	孔 Id
FAD2A	F4+R1, F1+R1,	8, 27, 30, 83, 109, 147, 180,	平板 199	L23
	F1+R4, F3+R3	199, 209, 251, 288	平板 27	D20
FAD2-1	F1+R4, F4+R1,	12, 89, 123, 148, 269	平板 123	N17
	F1+R1, F2+R2		平板 148	B15
FAD2-2	F4+R1, F1+R1,	24, 44, 46, 47, 80, 91, 104,	平板 44	H03
	F1+R4, F2+R2	110, 119, 121, 124, 248	平板 121	A17
FAD2-3	F1+R4, F4+R1,	8, 62, 113, 205, 276	平板 62	I16
	F1+R1, F3+R3		平板 205	K11

[0249] 通过测序对每个已鉴定的 FAD 基因家族成员的单个 BAC 克隆进行进一步的分析。使用 Large Construct 试剂盒® (Qiagen, Valencia, CA) 按照制造商的使用说明为 BAC 克隆分离 DNA 并准备用于测序。使用 GS-FLX Titanium Technology® (Roche, Indianapolis, IN) 按照制造商的使用说明准备用于测序的提取 BAC DNA。测序反应使用物理分区的 (sectored) GS-FLX TI Pico- 滴定板® 利用成对收集的 BAC 实施, 用于优化数据输出。成对合并 BAC, 其中 FAD3 基因与 FAD3 基因配对。产生的全部序列数据用 Newbler v2.0.01.14® (454Life Sciences, Branford, CT) 组装。组装的重叠群 (contig) 使用 Sequencher v3.7® (GeneCodes, Ann Arbor, MI) 手动分析相应 FAD 基因重叠群的存在。

[0250] 在鉴定并完全表征了所有 4 个 FAD2 基因的全基因组序列之后, 设计锌指核酸酶用于结合每个特定基因家族成员的序列。

[0251] 实施例 2: 特异针对 FAD2 基因的锌指结合结构域的设计

[0252] 针对编码 FAD2 基因座各种功能序列的 DNA 序列的新颖锌指蛋白的设计基本上如前所述。参见例如 Urnov et al. (2005) Nature 435:646-651。示例性靶序列和识别螺旋如表 4 (识别螺旋区设计) 和表 5 (靶位点) 所示。在表 5 中, 与 ZFP 识别螺旋接触的靶位点的核苷酸用大写字母指示; 非接触的核苷酸用小写字母指示。

[0253] 锌指核酸酶 (ZFN) 靶位点被设计成结合 FAD2A 的 5 个靶位点。FAD2A 锌指设计被合并到锌指表达载体中, 其编码具有至少一个具有 CCHC 结构的锌指的蛋白质。参见, 美国专利公开 No. 2008/0182332。特别地, 每个蛋白的最后一个指具有用于识别螺旋的 CCHC 骨架。非规范的锌指编码序列通过一个四氨基酸 ZC 接头和来自玉米的 opaque-2 核定位信号与 IIS 型限制酶 FokI 的核酸酶结构域 (Wah et al., (1998) Proc. Natl. Acad. Sci. USA 95:10564-10569 的序列的氨基酸 384-579) 融合, 形成 FAD2A 锌指核酸酶 (ZFNs)。融合蛋白的表达受到相对较强的组成型启动子驱动, 例如来自木薯脉花叶病毒 (CsVMV) 启动子的启动子, 并且侧翼是根癌土壤杆菌 ORF233' 非翻译区 (AtuORF233' UTR v1)。在克隆到构建体中的两个锌指核酸酶融合蛋白之间添加来自明脉扁刺蛾 (Thauma octocaryon) 病毒的自水解 2A 编码核苷酸序列 (Szymczak et al., 2004)。示例性载体或质粒如下文表 5 所述。

[0254] 使用先前显示用于鉴定活性核酸酶的基于出芽酵母的系统验证最优的 FAD2 锌指核酸酶的剪切活性。参见例如美国专利公开 No. 20090111119 ;Doyon et al. (2008) Nat Biotechnol. 26:702 - 708 ;Geurts et al. (2009) Science 325:433。用于多种功能性结构域的锌指被选择用于体内使用。在被设计、产生和测试与推定 FAD 基因组多核苷酸靶位点结合的大量 ZFN 中,有 11 个 ZFN 被鉴定具有高水平的体内活性,并被选择用于进一步的实验。这些 ZFN 被表征为能够高效结合和剪切植物内的独特 FAD2 基因组多核苷酸靶位点。

[0255] 表 4:FAD2 锌指设计

[0256]

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

[0257]

	(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 NO:94	HSHARIK SEQ ID NO:95	TSGHLSR SEQ ID NO:136	RSDNLSV SEQ ID NO:137	IRSTLRD SEQ ID NO:138	N/A
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	DRSDLSR SEQ ID	RSDHLTQ SEQ ID	QSGALAR SEQ ID	QSGNLAR SEQ ID	N/A

[0258]

	NO:124	NO:125	NO:108	NO:110	NO:104	
24797	RS AVLSE 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	QSSDLR 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 NO:110	QSGNLAR SEQ ID NO:104	DRSHLAR SEQ ID NO:169	DRSDLSR SEQ ID NO:125	RSDNLTR SEQ ID NO:179	N/A
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	RSSDLR SEQ ID NO:193	QSSDLR SEQ ID NO:161	RSDHLTQ SEQ ID NO:108
24832	RSDSLR SEQ ID NO:114	RKDARIT SEQ ID NO:115	DRSHLSR SEQ ID NO:198	QSGNLAR SEQ ID NO:104	QSSDLR SEQ ID NO:161	DRSALAR SEQ ID NO:201

[0259]

24833	RSDDL\$K	RSDTRKT	DRSNLSR	DRSNLWR	RSDSLSR	NNDHRKT
	SEQ ID	SEQ ID	SEQ ID	SEQ ID	SEQ ID	SEQ ID
	NO:202	NO:203	NO:99	NO:121	NO:114	NO:207

[0260] 表 5:FAD2 锌指的靶位点

[0261]

ZFP	质粒 No.	靶位点(5'至 3')	SEQ ID NO.中存在的 ZFP 靶/结合位点
24800	pDAB104001	ccCAAAGGGTTGTTGAGgtacttgccgt	SEQ ID NO:17
24801	pDAB104001	cgCACCGTGATGTAAACggttcagtca	SEQ ID NO:18
24794	pDAB104002	taAGGGACGAGGAGGAAGgagtgaaga	SEQ ID NO:19
24795	pDAB104002	ttCTCCTGGAAGTACAGtcacgcacgcc	SEQ ID NO:20
24810	pDAB104003	gtCGCTGAAGGcGTGGTGgcccgcactcg	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	tcCCAAAGGGTtGTTGAGgtaettgccg	SEQ ID NO:25
24819	pDAB104005	acCGTGATGTAAACGGTtcagttcactc	SEQ ID NO:26
24796	pDAB104006	gaGAAGTAAGGGACGAGgaggaaggagt	SEQ ID NO:23
24797	pDAB104006	tgGAAGTAcAGTCATCGAcgccaccatt	SEQ ID NO:28
24836	pDAB104007	gtAGAGACcGTAGCAGACgcccagagatg	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	gaGAGGCCaGGCGAAGTAaggagagagg	SEQ ID NO:33
24821	pDAB104009	acTGGGCCTGCCAGGGCtgcgtcctaac	SEQ ID NO:32
24828	pDAB104010	agGCCCAGtAGAGAGGCCaggcgaagta	SEQ ID NO:35
24829	pDAB104010	ccAGGGCTGCGTCCTAACCgcgctctgg	SEQ ID NO:36
24832	pDAB104011	taGTCGCTGAAGGCGTGGTGgcccgcact	SEQ ID NO:37
24833	pDAB104011	agTGGCTGGACGACaCCGTCCgcctcat	SEQ ID NO:38

[0262] 实施例 3: 锌指核酸酶剪切 FAD3 基因的评估

[0263] 构建体组装

[0264] 含有使用如实施例 2 所述的酵母测定法而鉴定的示例性锌指核酸酶的 ZFN 表达构建体的质粒载体使用本领域公知的技术设计并制成。每个锌指编码序列与位于锌指核酸酶的上游的编码 opaque-2 核定位信号 (Maddaloni et al. (1989) Nuc. Acids Res. 17(18):7532) 的序列融合。

[0265] 接着,将 opaque-2 核定位信号::锌指核酸酶融合序列与互补的 opaque-2 核定位信号::锌指核酸酶融合序列配对。这样,每个构建体含有由两个 opaque-2 核定位信号::锌指核酸酶融合序列构成的单个开放阅读框,两个融合序列被来自明脉扁刺蛾 (Thosea asigna) 病毒的 2A 序列 (Mattion et al. (1996) J. Virol. 70:8124-8127) 分隔。融合蛋白的表达受到相对较强的组成型启动子驱动,例如来自木薯叶脉花叶病毒 (CsVMV) 启动子的启动子,并且侧翼是根癌土壤杆菌 ORF233' 非翻译区 (AtuORF233' UTR)。

[0266] 载体用 In-Fusion™ Advantage Technology (Clontech, Mountain View, CA) 组装。限制性核酸内切酶从 New England BioLabs (NEB; Ipswich, MA) 获得,并使用 T4DNA 连接酶 (Invitrogen) 进行 DNA 连接。质粒制备使用 NucleoSpin® 质粒试剂盒 (Macherey-Nagel Inc., Bethlehem, PA) 或 Plasmid Midi 试剂盒 (Qiagen) 按照制造商的使用说明进行。DNA 片段在琼脂糖 Tris- 乙酸凝胶电泳之后使用 QIAquick 凝胶提取试剂盒™ (Qiagen) 进行分离。所有已组装质粒的菌落通过微量制备的 DNA 的限制性消化进行最初筛选。对选出的克隆的质粒 DNA 通过商业测序载体 (Eurofins MWG Operon, Huntsville, AL) 进行测序。序列数据使用 Sequencher™ 软件 (Gene Codes Corp., Ann Arbor, MI) 进行组装和分析。在递送到欧洲油菜原生质体之前,使用 Pure Yield Plasmid Maxiprep System® (Promega Corporation, Madison, WI) 或 Plasmid Maxi Kit® (Qiagen, Valencia, CA) 按照供应商的使用说明从大肠杆菌培养物制备质粒 DNA。

[0267] 通过限制酶消化和通过 DNA 测序确认了所得的 11 个质粒构建体:pDAB104008 (含有 ZFN24845 和 ZFN24844 构建体),pDAB104009 (含有 ZFN24820 和 ZFN24821 构建体),pDAB104010 (含有 ZFN24828 和 ZFN24829 构建体) (图 3),pDAB104003 (含有 ZFN24810 和 ZFN24811 构建体),pDAB104011 (含有 ZFN24832 和 ZFN24833 构建体),pDAB104002 (含有 ZFN24794 和 ZFN24795 构建体),pDAB104006 (含有 ZFN24796 和 ZFN24797 构建体),pDAB104004 (含有 ZFN24814 和 ZFN24815 构建体),pDAB104001 (含有 ZFN24800 和 ZFN24801 构建体),pDAB104005 (含有 ZFN24818 和 ZFN24819 构建体),和 pDAB104007 (含有 ZFN24836 和 ZFN24837 构建体)。表 6 列举了不同的构建体和每个 ZFN 被设计用于剪切和结合的特定 FAD2A 序列。

[0268] 表 6: 列出锌指蛋白结合基序和相应的构建体编号。每个锌指被设计成结合并剪切表中所述的 FAD2A

[0269]

ZFN 设计	构建体 No.	基因座 ID.	FAD2A 序列中的靶切割位点
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-24815	pDAB104004	FAD2_ZFN_Locus3_F2H	408-410
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

[0270] 用于转染的 DNA 的制备

[0271] 上述载体的质粒 DNA 通过沉淀,在 100% (v/v) 乙醇中清洗,并在层流净化罩内干燥进行灭菌。将 DNA 沉淀 (pellet) 重悬于 30 μ L 无菌双蒸水中,终浓度为 0.7 μ g/ μ L,用于如下所述地转染进入原生质体细胞中。进行质粒 DNA 制备,产生用于瞬时转染的超螺旋质粒 DNA 和用于稳定转染的线性化质粒 DNA。对于原生质体细胞的瞬时转染,不需要向转化质粒中添加载体 DNA (例如,鱼精 DNA)。对于瞬时研究,每次转染中,每 10⁶ 个原生质体使用大约 30 μ g 质粒 DNA。

[0272] 转染

[0273] 欧洲油菜 L. var. DH10275 的转染如 Spangenberg et al., (1986) Plant Physiology 66:1-8 所述实施,培养基配方如 pangenberg G. and Protrykus I. (1995) Polyethylene Glycol-Mediated Direct Gene Transfer in Tobacco Protoplasts, 其集录于: Gene Transfer to Plants. (Protrykus I. and Spangenberg G. Eds.) Springer-Verlag, Berlin 所述。欧洲油菜的种子用 70% 乙醇表面灭菌。将种子浸没在 12mL 70% 乙醇溶液中,并通过轻柔摇动该混合物 (cocktail) 10 分钟进行混合。通过倾倒溶液除去 70% 乙醇溶液,并用由 1% w/v 次氯酸钙和 0.1% v/v 吐温-20 组成的种子灭菌溶液进行交换。将种子浸没在种子灭菌溶液中,并通过轻柔摇动该混合物 (cocktail) 25 分钟进行混合。倾倒种子灭菌溶液,并将经过灭菌的种子在 50mL 无菌水中漂洗 3 次。最后,将种子转移到无菌 80mm Whatman 滤纸盘® (Fisher-Scientific, St. Louis, MO) 上,其置于培养皿 (Petri dish) 中,并用无菌水将种子稍微饱和。培养皿用 Parafilm® (Fisher-Scientific, St. Louis, MO) 密封,并将平板在 25°C 下完全黑暗温育 1-2 天。在可以从种子观察到幼苗萌发迹象之后,将幼苗转移到含有凝固的 GEM 培养基的培养皿中,以易化进一步的种子萌发。幼苗在 GEM 培养基上 25°C 温育 4-5 天。

[0274] 向无菌培养皿中倾倒一定体积的液体 PS 培养基（大约 10mL）。使用无菌镊子和手术刀除去并丢弃处于生长和发育 4 叶期的 4-5 日龄幼苗的地上部分。确定长度为 20-40mm 的下胚轴节段产生的富含细胞质的小原生质体的群体最高。将下胚轴节段无菌切下并转移到液体 PS 培养基中。将切下的下胚轴节段合在一起，横切成 5-10mm 节段。接着，将下胚轴节段转移到新鲜 PS 培养基中，并在室温下温育 1 小时。将质壁分离的下胚轴转移到含有酶溶液的培养皿中。将全部下胚轴节段小心地浸没在溶液中。用 Parafilm® 密封培养皿，并在轻柔摇动下 20-22℃ 过夜温育 16-18 小时。

[0275] 原生质体细胞从下胚轴节段释放。轻轻搅拌下胚轴的过夜消化物，使原生质体释放到酶溶液中。将培养皿略微倾斜，以帮助转移酶溶液和植物碎片的消化悬液。使用 10mL 移液管将消化悬液转移到无菌的原生质体过滤单元（100 微米网筛滤器）中，进一步分离原生质体和植物碎片。轻扣过滤单元，以释放被筛网捕获的多余液体。轻柔混合大约 8-9mL 的原生质体悬液，并分配到 14mL 无菌塑料圆底离心管中。每个悬液用 1.5mL W5 溶液覆盖。将 W5 溶液小心置于原生质体悬液上，以一定的角度并逐滴添加，以便使扰动最小化。向原生质体悬液添加 W5 溶液导致产生富含原生质体的界面。这个界面用移液管进行收集。接着，将收集的原生质体转移到一个新的 14mL 离心管中，并轻轻混合。将所产生或获得的原生质体用血细胞计数器进行检测，以确定每毫升中原生质体的数量。重复本方法，其中消化叶片组织以产生叶肉原生质体。

[0276] 接着，添加 W5 溶液到 10mL 的体积，并在 70g 离心原生质体，然后除去 W5 溶液。通过轻轻摇晃将剩余的原生质体悬浮液重新悬浮。含有原生质体悬浮液的每个管中填充 5mL W5 溶液，并在室温下温育 1-4 小时。在 70g 离心原生质体悬液，并除去全部的 W5 溶液。接着，向每个含有分离的原生质体的沉淀原生质体悬液中添加 300 μ L 转化缓冲液。向每个管中添加 10 μ g 质粒 DNA 到原生质体悬液中。质粒 DNA 含有如上所述的锌指核酸酶构建体（例如 pDAB104010）。接着，向原生质体悬液中添加 300 μ L 预热的 PEG4000 溶液，并轻轻叩击管。将原生质体悬液和转化混合物在室温下温育 15 分钟，不进行任何搅拌。以每份 1mL, 1mL, 1mL, 2mL, 2mL 和 3mL 的量向每个管中顺序添加额外的 10mL W5 溶液，并在每次添加 W5 溶液之间将管轻轻颠倒。通过 70g 离心沉淀原生质体。除去全部的 W5 溶液，留下纯的原生质体悬液。

[0277] 接着，向离心沉淀的原生质体细胞添加 0.5mL K3 培养基，并将细胞重悬浮。将重悬的原生质体细胞置于培养皿的中心，和 1:1 浓度的 5mL K3 和 0.6mL Sea Plaque™ 琼脂糖（Cambrex, East Rutherford, NJ）。单次轻柔涡旋运动摇动培养皿，并在室温下温育 20-30 分钟。培养皿用 Parafilm® 密封，并将原生质体在完全黑暗下培养 24 小时。在黑暗中温育后，将培养皿在昏暗光（Osram L36W/21Lumilux 日光灯（white tube），5 μ Mol m⁻²s⁻¹）下培养 6 天。在培养步骤后，使用无菌刮刀将含有原生质体的琼脂糖切分成四份。将分离的四等分物置于含有 20mL A 培养基的 250 毫升塑料培养容器中，并在旋转摇床上以 80rpm 和 1.25cm 冲程（throw）在 24℃ 下在连续昏暗的灯光下温育 14 天，然后分析确定每个锌指核酸酶构建体的活性水平。

[0278] 从芥花原生质体分离基因组 DNA

[0279] 将经转染的原生质体置于在单独的 1.5 或 2.0mL 微离心管中。将细胞离心沉淀在含有缓冲溶液的管的底部。DNA 提取如下实施：将细胞在液氮中速冻，然后在 Labconco

Freezone 4.5® (Labconco, Kansas City, MO) 中在 -40°C 和大约 $133 \times 10^{-3} \text{ mBar}$ 压强下冷冻干燥细胞大约 48 小时。使用 DNeasy® (QIAGEN, Carlsbad, CA) 植物试剂盒对冻干的细胞进行 DNA 提取, 按照制造商的使用说明进行, 只是不需要组织破坏, 而是将原生质体细胞直接添加到裂解缓冲液中。

[0280] 用于在芥花原生质体中进行基因组 DNA 序列剪切的 FAD2A ZFN 的测试

[0281] 将 FAD2A 基因座中的 ZFN 靶位点的设计进行聚类, 从而设计出覆盖靶位点的多对 ZFN。ZFN 靶位点的聚类使得人们能够设计 PCR 引物, 使之能够从所有 FAD2A 基因家族成员扩增 100bp 窗口内的周围基因组序列, 从而涵盖全部的覆盖 ZFN 靶位点。这样, Illumina 短读取序列技术可用于评估被转染的原生质体中靶 ZFN 位点的完整性。此外, 设计的 PCR 引物需要包括这样的特定核苷酸碱基, 其可将序列读取归结于 FAD2A 家族的特定基因成员。因此, 需要全部的 PCR 引物在距离任何 ZFN 靶切割位点 5-10 个核苷酸处结合, 因为已知非同源末端连接 (NHEJ) 活性可导致小的缺失, 这些小的缺失能够除去引发位点, 抑制扩增, 从而干扰 NHEJ 活性的评估。

[0282] 引物被设计成结合 FAD2A 基因家族全部的 ZFN 靶基因座 (表 7), 并通过对 PCR 扩增产物进行基于 Sanger 的测序, 经验性地测试它们对全部基因家族成员的扩增。在若干情况下, 无法建立能够区分全部基因家族成员的引物 (表 8), 然而, 在所有情况下, FAD2A 的靶基因序列可以被区分。在 PCR 引物设计之后, 将定制的 DNA 条码 (barcode) 序列纳入到 PCR 引物中, 这些条码序列可用于区分不同的 ZFN 靶基因座, 并鉴定对转染和 ZFN 的特定序列读取 (表 7 和 8)。

[0283] 表 7: 设计用于评估 FAD2ZFN 活性的引物序列。引物包括定制条码 (custom barcode), 连同两个必要的用于构建 Illumina 文库的 Illumina 衔接子序列用于测序 - 合成分析。所购引物是所给出的全部三个列的总和。

[0284]

基因座 ID	SEQ ID NO:	Illumina 接头引物序列条码基因座引物
FAD2_ZFN_Locus 1_F2A	SEQ ID NO:39	ACACTCTTTCCTACACGACGCTCTTCCGATCTACGTACCCCTCTCYCTACYTCGCC

[0285]

FAD2_ZFN_Locus 1_F2B	SEQ ID NO:39	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGTACCCCTCTCYCY TACYTCGCC
FAD2_ZFN_Locus 1_F2C	SEQ ID NO:39	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTACGCCCTCTCYCY TACYTCGCC
FAD2_ZFN_Locus 2_F1D	SEQ ID NO:39	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTACGTGTCATAGCCC ACGAGTGCGGC
FAD2_ZFN_Locus 2_F1E	SEQ ID NO:39	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTGACGTCATAGCCC ACGAGTGCGGC
FAD2_ZFN_Locus 3_F2F	SEQ ID NO:39	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGACTGTCGGCCTC ATCTTCCACTCC
FAD2_ZFN_Locus 3_F2G	SEQ ID NO:39	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGACTGGTCGGCCTC ATCTTCCACTCC
FAD2_ZFN_Locus 3_F2H	SEQ ID NO:39	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACTGAGTCGGCCTC ATCTTCCACTCC
FAD2_ZFN_Locus 4_F1J	SEQ ID NO:39	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCTAGCAGACATCAA GTGGTACGGC
FAD2_ZFN_Locus 4_F1K	SEQ ID NO:39	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTAGCCAGACATCAA GTGGTACGGC
FAD2_ZFN_Locus 5_F1L	SEQ ID NO:39	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAGCTATCTCCGACG CTGGCATCCTC
FAD2_ZFN_Locus 1_R1A	SEQ ID NO:40	CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTACGTA CTGGT AGTCGCTGAAGGCGT
FAD2_ZFN_Locus 1_R1B	SEQ ID NO:40	CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTCGTACCTGGT AGTCGCTGAAGGCGT
FAD2_ZFN_Locus 1_R1C	SEQ ID NO:40	CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTGTACGCTGGT AGTCGCTGAAGGCGT
FAD2_ZFN_Locus 2_R1D	SEQ ID NO:40	CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTTACGTGGACG AGGAGGAAGGAGTGGA
FAD2_ZFN_Locus 2_R1E	SEQ ID NO:40	CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTCTGACGGACG AGGAGGAAGGAGTGGA
FAD2_ZFN_Locus 3_R1F	SEQ ID NO:40	CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTTGACTAGTGT GGAATGGTGGCGTCG
FAD2_ZFN_Locus 3_R1G	SEQ ID NO:40	CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTGACTGAGTGT TGGAATGGTGGCGTCG
FAD2_ZFN_Locus 3_R1H	SEQ ID NO:40	CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTACTGAAGTGT TGGAATGGTGGCGTCG
FAD2_ZFN_Locus 4_R1J	SEQ ID NO:40	CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTGCTAGCCCGA GACGTTGAAGGCTAAG
FAD2_ZFN_Locus 4_R1K	SEQ ID NO:40	CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTCTAGCCCGA GACGTTGAAGGCTAAG
FAD2_ZFN_Locus 5_R1L	SEQ ID NO:40	CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTTAGCTGAAGG ATGCGTGTGCTGCAAG

[0286] 表 8 显示了所设计的 PCR 引物对 FAD2 基因家族的扩增性能。“X”表示基因拷贝检测特异性，灰色阴影和“+”表示在讨论的特定基因座由两个引物设计的序列读取无法区分。

[0287] 表 8:FAD2A 和 FAD2C 基因座处的剪切结果

[0288]

ZFN 基因座	FAD 基因拷贝			
	FAD2A	FAD2C	FAD2A'	FAD2C'
Locus 1	X	X	X	X
Locus 2	X	X	X	X

[0289]

Locus 3	+	+	X	X
Locus 4	X	X	X	X
Locus 5	X	X	X	X

[0290] 在对被 ZFN 转染的芥花原生质体进行 DNA 提取之后,对靶 ZFN 基因座进行 PCR 扩增,以便以正确的格式产生必需的基因座特异性 DNA 分子,用于通过合成技术的基于 Illumina 的测序。每个测定经过优化以便在 25ng 起始 DNA(欧洲油菜基因组的大约 12,500 个细胞的当量)上工作。为每个样品进行多个反应,以提供以合适的水平评估 NHEJ 效率和特异性所需的覆盖率,大约 16 个 PCR 反应相当于从单个原生质体获得的欧洲油菜基因组的 200,000 个拷贝。为所有使用相同的测定的待测试的样品制备 PCR 扩增主预混物(master-mixes);用曾被用于确定要对靶组织实施的最佳循环数的定量 PCR 方法测定一个 3 个重复的反应,以确保 PCR 扩增未受到试剂限制,并且仍然处于指数扩增阶段。实验以及必要的阴性对照反应以 96 孔格式使用 MX3000P thermocycler® (Stratagene, LaJolla, CA) 实施。

[0291] 根据从定量 PCR 平台收集的输出对荧光的逐循环相对增加描点作图,并为每个测定确定能够输送足够的扩增同时又不会使反应受到试剂限制的循环数,以试图减少过度循环和共有转录本或分子的扩增。将未使用的主预混物保持在冰上直至定量 PCR 分析完成且循环数得以确定,然后分装到期望数目的反应管中(每个 ZFN 测定大约 16 个),并进行 PCR 反应。

[0292] 扩增后,将单个 ZFN 基因座的样品汇集在一起,并使用 MinElute PCR 纯化试剂盒® (Qiagen) 按照制造商的使用说明对每个 ZFN 的 200 μL 汇集产物进行清洁。为了使样品能够用 Illumina 短读取技术进行测序,需要通过扩增将额外的配对末端引物附接到已产生的片段上。这通过 PCR 扩增实现,PCR 扩增使用的引物与在第一轮扩增中添加的序列部分互补,但还含有所需的配对末端序列。使用经过了如前所述的定量 PCR 循环分析的样品,再次确定要实施的 PCR 循环的最佳数目,PCR 循环的最佳数目能够向模板添加配对末端引物但不会过度扩增共有片段。

[0293] 在 PCR 扩增之后,使用 MinElute 柱® (Qiagen) 按照制造商的使用说明对所产生的产物进行清洁,并在 2.5% 琼脂糖凝胶上进行分离。DNA 片段用 Syber® Safe (Life Technologies, Carlsbad, CA) 可视化,对具有正确大小的条带进行凝胶提取,除去任何残余的 PCR 产生的引物二聚体和其它假片段,使用 MinElute 凝胶提取试剂盒® (Qiagen) 按照制造商的使用说明从凝胶片提取 DNA。在完成凝胶提取后,使用 AMPure 磁珠® (Beckman-Coulter, Brea, CA) 对 DNA 进行额外的清洁,DNA 与珠子比例为 1:1.7。然后,使用基于定量 PCR 的文库定量试剂盒估测 DNA 浓度,以 1/40,000 和 1/80,000 稀释用于 Illumina 测序(KAPA),反应实施三个重复。根据定量 PCR 结果,将 DNA 稀释成 2nM 的标准浓度,并合并所有的文库用于 DNA 测序。使用 cBot cluster generation kit® (Illumina, San Diego, CA) 制备用于测序的样品,并在 Illumina GA2x® 上按照制造商的使用说明进行测

序,实施 100bp 配对末端测序读取。

[0294] 用于检测靶锌指位点处非同源末端连接的数据分析方法

[0295] 在完成测序反应并使用用于碱基识别 (base calling) 的 Illumina 生物信息管道 (pipeline) 进行原始数据读出之后,实施完全分析来鉴定每种情况下靶 ZFN 位点处缺失的碱基。设计定制的 PERL 脚本 (script) 从 DNA 序列中根据一系列输入序列用计算机提取和分选条码。条码必须以大于 30 的 Phred 得分与参考序列相匹配,以减少错误指认 (misattributing) 的序列读取。在序列读取被分选到所用的不同条码组中之后,对所有序列进行质量过滤。质量过滤是第二种定制开发的 PERL 脚本。如果有三个碱基被读出为“N”,或者如果中位数 Phred 得分小于 20,或者有三个连续碱基的 Phred 得分小于 20,或者序列读取的长度小于 40bp,则排除该序列读取。使用 NextGENe® (SoftGenetics, State College, PA) 软件包合并剩余的序列,其中两个配对序列读取均可获得。然后,使用第三种定制的 PERL 脚本将剩余的合并序列读取减小为一组唯一的序列读取,将被鉴定的冗余序列的数目计数记录在剩余序列识别器的末端。然后,使用 NextGENe® 软件将唯一的序列读取与 FAD2 参考序列进行比对,产生有缺口的 FASTA 比对文件。

[0296] 使用该带缺口的 FASTA 文件,使用第四种定制的 PERL 脚本将有缺口的碱基位置编号转换成输入参考。这能够在组装的数据中鉴定这样的碱基,其可以区分不同的基因家族成员 (不同基因家族成员之间的同源或旁系同源序列变异)。一旦实施了碱基编号的转换,便有可能为每个唯一的序列读取产生单倍型报告,并将该读取指派到特定的基因家族成员。一旦读取被基因分组,便可以在 ZFN 靶位点周围鉴定并评估一个 10bp 的窗口。记录每个基因具有缺失的序列数目和错配碱基的数目。

[0297] 然后,将数据图形显示为多线图,图上有每 10,000 个序列读取中在靶 ZFN 位点处具有 1-10 个碱基缺失的序列的数目。为所有 ZFN 转染以及对照转染实施这种分析。在几种情况下,天然 DNA 序列中的重复导致靶 ZFN 位点中的测序错误增加,这种错误可以一般性地视为在所有样品中 (包括被 ZFN 转染的和被对照转染的) 被报告的单碱基缺失的普遍增加 (图 5)。

[0298] 从这些结果可知,在 FAD2 靶位点处的最高水平的 ZFN 活性,如根据更大的 NHEJ 活性确定的,在位点 E 被鉴定。选择质粒 pDAB104010 (即 ZFN24828 和 24829) 上编码的 ZFN 用于在植物内靶向工程化的转基因整合平台 (ETIP),这是考虑到它具有显著的基因组 DNA 剪切活性和最小的非靶活性等特征。

[0299] 实施例 4: 用于工程化的转基因整合平台 (ETIP) 芥花植物品系的 DNA 构建体

[0300] 下面描述的质粒载体是使用本领域技术人员公知的方法和技术构建的。本段所描述的具体试剂和技术的应用是本领域技术人员容易知道的,并且容易和其它试剂和技术交换,以实现构建质粒载体构建体的理想目的。限制性核酸内切酶从 New England BioLabs (NEB; Ipswich, MA) 获得。连接用 T4DNA 连接酶 (Invitrogen, Carlsbad, CA) 完成。Gateway 反应使用 Gateway® LR Clonase® 酶预混物 (Invitrogen) 实施,用于将一个进入载体 (entry vector) 组装到单个目的载体中。In-Fusion™ 反应使用 In-Fusion™ Advantage Technology (Clontech, Mountain View, CA) 实施,用于将一个进入载体

组装到单个目的载体中。质粒制备使用 NucleoSpin[®] 质粒试剂盒 (Macherey-Nagel Inc., Bethlehem, PA) 或 Plasmid Midi Kit[®] (Qiagen) 按照供应商的使用说明实施。DNA 片段在琼脂糖 Tris- 乙酸凝胶电泳之后使用 QIAquick 凝胶提取试剂盒[™] (Qiagen) 提取。所选克隆的质粒 DNA 通过商业测序供应商 (Eurofins MWG Operon, Huntsville, AL) 进行测序。序列数据用 SEQUENCHER[™] 软件 (Gene Codes Corp., Ann Arbor, MI) 组装和分析。

[0301] 用于精确整合 ETIP 到芥花 FAD2 基因座中的直接投送载体

[0302] 使用标准克隆方法构建含有 ETIP 的载体 pDAS000130 (图 6, T 链插入物为 SEQ ID NO:61), 用于特异整合到欧洲油菜的 FAD2A 基因中。这种构建体设计为被投送到具有锌指核酸酶构建体 pDAB104010 的芥花原生质体中。锌指核酸酶构建体将切割 FAD2A 基因座, 然后 pDAS000130 构建体将通过同源性介导的或非同源末端连接修复机制整合到芥花基因组中。ETIP 包括 4 个表达盒 (两个是不完整的), 它们被额外的 ZFN 识别序列分隔, 和一个工程化的登录垫 (Engineered Landing Pad) (ELP), 其含有另一个 ZFN 识别序列。这些额外的 ZFN 识别序列是独特的, 被设计作为靶标用于将多核苷酸序列引入到 ETIP 和 ELP 转基因插入物中。类似地, ZFN 识别序列可以用于切出多核苷酸序列。第一表达盒是一个不完整的 dsRED 表达盒, 并且包含来自拟南芥多聚泛素 (polyubiquitin) 10 (AtUbi 启动子) 基因的启动子, 5' 非翻译区和内含子 (Callis, et al., (1990) J. Biol. Chem., 265:12486-12493), 随后是 210bp 的来自珊瑚礁珊瑚 (reef coral *Discosoma* sp.) (Clontech, Mountain View, CA) 的 dsRed 基因, 其被密码子优化用于在双子叶植物中表达 (dsRED (双子叶植物优化的) 外显子 1), 随后是来自拟南芥硫还原酶样基因 (thioreductase-like gene) 的内含子 (来自拟南芥硫还原酶的内含子 1: 登录号 No: NC_00374) 和包含玉米 Viviparous-1 (Vp1) 基因转录终止子和聚腺苷酸化位点的 3' 非翻译区 (Zmlip 终止子: Paek et al., (1998) Molecules and Cells, 8(3):336-342)。第二个表达盒含有 19S 启动子, 其包括来自花椰菜花叶病毒的 5' UTR (CaMV 19S: Cook and Penon (1990) Plant Molecular Biology 14(3):391-405), 随后是来自大肠杆菌的 hph 基因, 其被密码子优化用于在双子叶植物中表达 (hph (HygR): Kaster et al., (1983) Nucleic Acids Research 11(19):6895-6911), 和包含根癌土壤杆菌 pTi15955 开放阅读框 1 的转录终止子和聚腺苷酸化位点的 3' UTR (At-ORF1 终止子: Barker et al., (1983) Plant Molecular Biology 2(6):335-50)。第三个表达盒是完整的 PAT 表达盒, 包含来自拟南芥 4- 香豆酰-CoA 合酶的第一内含子 (内含子 #24- 香豆酰-CoA 合酶 v: 登录号 At3g21320/NC003074), 随后是从绿色产色链霉菌 (*Streptomyces viridochromogenes*) 分离的膦丝菌素乙酰转移酶基因的一个合成、植物优化版本的最后 256bp, 其编码赋予对谷氨酰胺合成酶抑制剂, 包括膦丝菌素、草铵膦和双丙氨膦的抗性的蛋白质 (PAT(v6) 3' 端: Wohlleben et al., (1988) Gene 70(1):25-37)。这个盒终止于 3' UTR, 其包括根癌土壤杆菌 pTi15955 开放阅读框 23 的转录终止子和聚腺苷酸化位点 (AtuORF23 终止子: Barker et al., (1983) Plant Molecular Biology 2(6):335-50)。第四个表达盒是 ipt 基因盒, 包含来自拟南芥 DNA 结合蛋白 MYB32 基因 (U26933) 的启动子和 5' UTR 的 588bp 截短形式 (AtMYB32(T) 启动子: Li et al., (1999) Plant Physiology 121:313), 随后是来自根癌土壤杆菌的异戊二烯基转移酶 (ipt) 基因和 35s 终止子, 其包含来自花椰菜花叶病毒的转录终止子和聚腺苷酸化位点

(CaMV 35S 终止子:Chenault et al., (1993)Plant Physiology 101(4):1395-1396)。为了投送到FAD2A,ETIP序列的每个末端侧翼有1kb的FAD2A基因组序列,该基因组序列来自通过向欧洲油菜的FAD2A基因中投送pDAB104010中编码的ZFN而诱导的双链断裂的任一侧。

[0303] ETIP序列由商业基因合成供应商(GeneArt,Life Technologies)合成。使用Qiagen DNeasy plant mini kit® (Qiagen,Hilden),按照制造商提供的使用说明,从自欧洲油菜DH12075的叶片组织纯化的基因组DNA中扩增FAD2A基因组序列的1kb节段。使用T4连接酶(NEB,Ipswich,MA)将这些1kb FAD2A序列连接到ETIP中。所有组装质粒的菌落通过微量制备DNA的限制性消化进行初始筛选。限制性核酸内切酶从New England BioLabs(NEB,Ipswich,MA)和Promega(Promega Corporation,WI)获得。质粒制备用QIAprep Spin Miniprep Kit® (Qiagen)或Pure Yield Plasmid Maxiprep System® (Promega Corporation,WI)按照供应商的使用说明进行。所选克隆的质粒DNA用ABI Sanger测序和Big Dye Terminator v3.1cycle sequencing protocol® (Applied Biosystems,Life Technologies)进行测序。使用SEQUENCHER™软件评估和分析序列数据(Gene Codes Corp.,Ann Arbor,MI)。

[0304] 对照载体

[0305] 使用对照载体开发基于荧光激活细胞分选(FACS)的细胞分选方法。使用标准克隆方法构建包含两个基因表达盒的对照质粒pDAS000031(图7:T-链插入物为SEQ ID NO:62)。第一个基因表达盒包含花椰菜花叶病毒19S启动子(CaMV 19S启动子;Shillito,et al., (1985)Bio/Technology 3:1099-1103)::潮霉素抗性基因(hph(HygR);美国专利No. 4,727,028)::和根癌土壤杆菌开放阅读框1的3'非翻译区(AtORF1终止子;Huang et al., (1990)J.Bacteriol. 1990172:1814-1822)。第二个基因表达盒包含拟南芥泛素10启动子(AtUbi10启动子;Callis,et al., (1990)J.Biol.Chem., 265:12486-12493)::dsRED(dsRED(D);美国专利No. 6,852,849)和来自拟南芥的内含子(内含子#1;GenBank:AB025639.1)::根癌土壤杆菌开放阅读框23的3'非翻译区(AtORF23终止子;美国专利No. 5,428,147),以反方向(例如,头对头方向)对框融合。质粒载体的组装使用In-Fusion™ Advantage技术(Clontech,Mountain View,CA)。

[0306] 实施例5:ETIP芥花植物系的产生

[0307] 欧洲油菜的转化

[0308] FAD2A位点特异性整合利用ETIP构建体(pDAS000130),连同锌指核酸酶(pDAB104010),和如实施例4所述的DS-Red对照构建体(pDAS000031)。将二元载体转化到根癌土壤杆菌菌株GV3101:PM90中。欧洲油菜原生质体细胞的转化使用实施例3所述的转染规程完成,并具有一些修改。

[0309] 该方案的修改包括使用海藻酸钠代替Sea Plaque™琼脂糖。将锌指核酸酶构建体pDAB104010和ETIP构建体pDAS000130共投送到欧洲油菜原生质体细胞内的转染实验,用摩尔比为5:1和12:1的DNA浓度完成,其中pDAS000130的浓度为27.8 μg质粒DNA,pDAB104010的浓度为2.2 μg质粒DNA。对照质粒构建体用30 μg质粒DNA的浓度转化。

[0310] 该方案的其他修改包括在含有1.5mg/mL潮霉素的培养基中从被转化的原生质体

细胞繁殖整株植物。整株植物的繁殖要求每2周更换一次A培养基,并对原生质体来源的集落的生长进行监视。在原生质体来源的集落生长至直径大约2-3mm后,将集落转移到含有凝固MS morpho培养基的12孔Costar®平板(Fisher Scientific, St. Louis, MO)的各个孔中。平板在连续的昏暗光下24℃温育1-2周,直至愈伤组织的尺寸扩增至直径为8-10mm。在原生质体细胞达到直径1-2cm后,将原生质体细胞转移到含有MS morpho培养基的单个250mL培养容器中。该容器在16小时光照($20 \mu\text{Mol m}^{-2}\text{s}^{-1}$, Osram L36W/21Lumilux 白光灯管)和8小时黑暗的条件下24℃温育。在1-2周内,可以见到多个芽。在芽长度达到3-4cm后,将芽转移到含有MS培养基的250mL培养容器中。将250mL培养容器在16小时光照($20 \mu\text{Mol m}^{-2}\text{s}^{-1}$, Osram L36W/21Lumilux 白光灯管)和8小时黑暗的条件下24℃温育。芽保持在该培养容器中,直至它们发育成小植株,此时将它们转移到温室中生长至成熟。

[0311] 实施例6: 含有ETIP的T-DNA整合在芥花中的分子确认

[0312] 使用DNeasy Plant Mini试剂盒Kit™(Qiagen)按照制造商的使用说明从所有推定的转基因植物叶片组织中提取基因组DNA,所不同的是组织用80 μL AE缓冲液洗脱。将30毫克来自再生植物的嫩叶组织在液氮中速冻,之后研磨成粉。

[0313] FAD2A基因座的分子表征使用三种独立的测定法执行。这些测定法使用如下的对照设计和优化:已表征的包含单个随机整合转基因的转基因事件,具有5个随机整合转基因的已表征转基因事件,野生型芥花栽培种DH12075植株和无模板对照反应。将下列三个分子分析的结果在一起进行考虑,以提供ETIP通过HDR整合在FAD2A处的证据。

[0314] 通过实时聚合酶链式反应鉴定转基因整合

[0315] 使用下列引物对每个植物的四个重复样品进行分析,其特异针对hph(也称为hpt)靶基因(SEQ ID NO:63, hpt F7915' CTTACATGCTTAGGAT CGGACTTG 3'; SEQ ID NO:64, hpt R9095' AGTTCCAGCACCAGATC TAACG 3'; SEQ ID NO:65, hpt Taqman 8725' CCCTGAGCCCAAGCAG CATCATCG 3' FAM)(图8)和编码高迁移率族蛋白I/Y的基因(HMG I/Y)(SEQ ID NO:66, F 5' CGGAGAGGGCGTGGAAGG 3'; SEQ ID NO:67, R 5' TTCGATTTGCTACAGCGTCAAC 3'; SEQ ID NO:68, 探针 5' AG GCACCATCGCAGGCTTCGCT 3' HEX)。使用如下的条件扩增反应:95℃ 10分钟,随后是40个循环的95℃ 30秒,60℃ 1分钟,扩增数据在每个退火步骤结束时捕获。拷贝数的计算使用 ΔC_q 方法,其中 $\Delta C_q = C_q(\text{靶基因}) - C_q(\text{参考基因})$ 。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。具有hph和HMG I/Y扩增并且拷贝数为0.5或更高的植株被看作是转基因的,并且拷贝数 ≥ 0.5 和 ≤ 1.2 的植株被评定为推定单拷贝。扩增在具有FastStart Universal Probe Master(R0X), (Roche, Basel, Switzerland)的BioRad CFX96Touch™ Real-Time PCR检测系统上进行。

[0316] 被破坏的FAD2A ZFN位点的检测

[0317] 在破坏基因座测试(disrupted locus test)中,对每个植物分析内源靶标的扩增存在与否。破坏基因座测试是一种显性测定(dominant assay)。该测定是一种SYBR® Green I qPCR测定,并且是单重式(singleplex),但是每个反应在同一PCR平板上同时运行,靶定内源基因座(FAD2A/2C. RB. UnE. F1, SEQ ID NO:69, 5' CTCCACTCCTCCTCCTCGT* C 3' 和 FAD2A/2C. RB. UnE. R1, 5' SEQ ID NO:70, GCGTCCCAAAGGGTTGTTGA*G 3') 和 ZFN

基因座 (在该基因座处 ZFN pDAB104010 结合并切割基因组) (FAD2A. UnE. F1, SEQ ID NO:71, 5' TCTCTACTGGGCCTGCCAGGG*C 3' 和 FAD2A. UnE. R1, SEQ ID NO:72, 5' CCCCAGACGT TGAAGGCTAAGTACAA*A 3') (图 9)。两个引物对均使用如下的条件进行扩增: 98℃ 30 秒, 随后是 35 个循环 (98℃ 10 秒, 65℃ 20 秒, 72℃ 90 秒), 然后是 95℃ 10 秒, 随后进行 50℃ -95℃ 的熔解分析, 增量为每 0.05 秒 0.5℃, 并在每次增量时进行平板读取。反应条件如表 9 所列。

[0318] 表 9: 用于 PCR 扩增的单一反应试剂组分和浓度

[0319]

试剂组分	体积(μl)
10 mM dNTP	0.40
5X Phusion HF 缓冲液	4.00
Phusion 热启动 II 高保真 DNA 聚合酶(2U/μl) (Thermo Scientific)	0.25
正向引物 10μM	0.40
反向引物 10μM	0.40
1:10000 稀释的 SYBR Green I 染料(Invitrogen)	1.00
分子生物学级 H ₂ O	11.55
基因组 DNA 模板(~20ng/μl)	2.00
总体积	20.00

[0320] 具有内源靶标扩增但没有 ZFN 靶标扩增的植株评定为破坏基因座测试阳性, 并被考虑为具有被破坏的 ZFN 基因座。当 FAD2A 基因座处两个等位基因上的 ZFN 结合位点均被破坏时, 这一测定被认为是阳性的。

[0321] 对 FAD2A 处通过同源性介导的修复的转基因整合的 PCR 检测

[0322] 利用终点法 (endpoint) 对每个推定的植物转化子进行分析, 其使用被设计成扩增如下转基因靶的 PCR 引物: hph(hph_ExoDigPC_F1, SEQ ID NO:73, 5' TTGCGCTGACG GATTCTACAAGGA 3' 和 hph_ExoDigPC_R1, SEQ ID NO:74, 5' TCCATCAGTCCAAACAGCAGCAG A 3'), FAD2A 内源基因座 (FAD2A. Out. F1, SEQ ID NO:75, 5' CATAGCAGTCTCACGTCCTGGT* C 3' 和 FAD2A. Out. Rvs3, SEQ ID NO:76, 5' GGAAGCTAAGCCATTACACTGTTCA*G 3'), 跨越任何通过 HDR 被整合到 FAD2A 基因座中的转基因的 5' 端的区域, 位于进入 FAD2A 基因座内的转基因的上游 (FAD2A. Out. F1, SEQ ID NO:77, 5' CATAGCAGTCTCACGTCCTGGT*C 3' 和 QA520, SEQ ID NO:78, 5' CCTGATCCGTTGACCTGCAG 3') 和跨越任何通过 HDR 整合到 FAD2A 基因座中的转基因的 3' 端、延伸到转基因的下游的 FAD2A 基因座内的区域 (QA558, SEQ ID NO:79, 5' GTGTGAGGTGGCTAGGCATC 3' 和 FAD2A. Out. Rvs3, SEQ ID NO:80, 5' GGAAGCTAAGCCA TTACACTGTTCA*G 3') (图 3)。所有引物对用如下的条件进行扩增: 98℃ 30 秒, 随后是 35 个循环 (98℃ 10 秒, 65℃ 20 秒, 72℃ 90 秒)。反应试剂条件如表 10 所列。

[0323] 表 10: 用于 PCR 扩增的单一反应试剂组分和浓度

[0324]

试剂组分	体积 (μl)
5x Phusion HF 缓冲液	6.00
10mM dNTPs	0.60
正向引物 10μM	0.60
反向引物 10μM	0.60
Phusion 热启动 II 高保真 DNA 聚合酶(2U/μl) (Thermo Scientific)	0.25
分子生物学级 H ₂ O	19.95
基因组 DNA 模板(~20ng/μl)	2.00
总体积	30.0

[0325] 靶侧翼的 5' 转基因 - 基因组的扩增和 / 或靶侧翼的 3' 转基因 - 基因组的扩增指示推定的插入事件。必须注意, 由于 pDAS000130 盒中大约 1,000bp 的 FAD2A 同源臂 (包含与 ZFN 切割位点直接上游和下游的 FAD2A 区域具有 100% 序列同一性的多核苷酸序列), PCR 反应可能产生假阳性 PCR 产物扩增, 这是由于脱靶 ETIP 整合事件的扩增产生的 PCR 嵌合。hph 靶的扩增确认发生了转基因整合。FAD2A 靶的扩增提示, FAD2A 基因座是完整的, 或者仅含有部分插入。由于 ETIP 的尺寸 (ETIP 盒为 11,462bp, 或者包含 FAD2A 同源臂和 ETIP 盒为 13,472bp), 可以预期, 当完整的 ETIP 被整合到 FAD2A 基因组中时, FAD2A 引物不会扩增产物。

[0326] FAD2A 编辑的 Southern 检测

[0327] 将扩增出了 5' 基因组 - 转基因侧翼靶产物和 / 或扩增出了的 3' 转基因 - 基因组侧翼靶产物、或者没有扩增出 ZFN 基因座靶标、或者两者的植株进行 Southern 分析, 检测转基因在 FAD2A 基因座处的整合。使用经过修改的 CTAB 法从 5g 叶片组织中纯化基因组 DNA (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)。接着, 用 KpnI-HF (New England BioLabs) 消化 12 μg 基因组 DNA, 并在 0.8% 琼脂糖凝胶上通过电泳分离消化片段, 随后用标准 Southern 印迹方案转移到膜上。使用针对如下基因的引物产生探针: FAD2A 5' 靶区域 (F, SEQ ID NO:81, 5' AGAGAGGAGACAGAGAGAGAGT 3' 和 R, SEQ ID NO:82, 5' AGACAGCATCAAGATTTACACA 3'), FAD2A 3' 靶区域 (F, SEQ ID NO:83, 5' CAACGGCGAGCGTAATCTTAG 3' 和 R, SEQ ID NO:84, 5' GTTCCTGGAATTGCTGATAGG 3') 和 hph (F, SEQ ID NO:85, 5' TGTTGGTGAAGAGGATACG 3' 和 R, SEQ ID NO:86, 5' ATCAGCAGCAGCGATAGC 3'), 使用 DIG Easy Hyb System® (Roche, South San Francisco, CA) 按照制造商的使用说明检测 FAD2A 基因座内 ETIP 的存在 (图 11)。执行杂交的温度为: 42°C 用于 FAD2A 5' 区、45°C 用于 FAD2A 3' 区, 42°C 用于

检测 hph。

[0328] 膜结合的基因组 DNA 用特定的顺序进行探测：首先，探测 FAD2A 5' 序列，然后探测 FAD2A 3' 序列，最后探测 hph 序列（图 12）。这样做的理由如下。第一探针（FAD2A 5'）是诊断探针，如果 ETIP 通过完美的 HDR 被整合到 FAD2A 中，则在膜上可以见到一个 5,321bp 片段。所得条带大小容易在电泳过程中被区分，并且会位于 DIG 标记的 Roche DNA 分子量 Marker III®（Roche, Indianapolis, IN）中的 5,148bp 片段附近。膜的第二探针是 FAD2A 3' 探针，被编辑的植株会具有一个 22,433bp 片段，而未被编辑的植株会有一个 16,468bp 片段。用 FAD2A 3' 探针鉴定的相同 22,433bp 片段也应当可以被 hph 探针结合和鉴定。这些片段难以在凝胶上区分，因为它们极其大，并且可能难以确定在 DIG 标记的 Roche DNA 分子量 Marker III® 中最大的 21,226bp 片段之上或下面显现的片段之间的任何差异。因此，这些探针通过使用 FAD2A 5' 探针显现一条 5kb 片段，提供可能强化 ETIP 通过同源性介导的修复（HDR）整合到 FAD2A 中的鉴定的证据。限制酶 KpnI 是唯一适用于本测定的限制性核酸内切酶，因为 KpnI 位点在 ETIP 盒中以单基因座出现，以单基因座切割 ETIP 盒，并且在 FAD2A ZFN 基因座中的两个位点存在。一个位点位于 FAD2A 同源臂的上游，第二个位点位于其下游。此外，KpnI 是甲基化不敏感的，可以作为具有更高保真度的重组酶供使用（New England Biolabs）。

[0329] 分子和 Southern 分析的结果

[0330] 在转染、培养和选择之后，将转基因植株转移到土壤中。该过程的结果有 139 株植物存活，对它们进行组织采样用于 gDNA 提取和分析。对全部 139 株植物进行分析，估计拷贝数。这 139 株植物中，56 株对 ETIP 阳性，56 株阳性植物中的 11 株具有推定的单拷贝整合（图 13）（表 11）。在 56 株对 ETIP 整合呈阳性的植物中，有 7 株植物发生了 FAD2A 5' - 基因组 - 转基因侧翼序列的扩增。在这 56 株对 ETIP 整合呈阳性的植株中均没有发生 FAD2A 3' - 转基因 - 基因组侧翼序列的扩增。此外，在 56 株对 ETIP 整合呈阳性的植株中，有 11 株植物对破坏的基因座 qPCR 测试呈阳性。对于 14 株对 FAD2A 5' - 基因组 - 转基因侧翼序列扩增呈阳性和 / 或对破坏的基因座 qPCR 测试呈阳性的植株，使用上述的 3 个探针进行了 Southern 分析。对于这 14 株进行了 Southern 分析的植株，所有的植株均显示在 FAD2A 基因座内具有部分整合，但是当用 FAD2A 5' 探针、FAD2A 3' 和 hph 探针进行探测时，这些植株均没有显示出完整的全长 ETIP 通过 HDR 整合在 FAD2A 基因座处的证据。没有出现如下的条带：i) 大于 WT，和 ii) 与这些样品用 FAD2A 3' 探针进行探测时所观察到的条带相同（表 11）。

[0331] 表 11:ETIP 整合分析结果概览

[0332]

土壤中存活的植物数目	139
采样植物数目	139
完成 qPCR 拷贝数分析的植物数目	139
ETIP 整合阳性的植物数目	56

包含推定单拷贝插入物的植物数目	11
ETIP/FAD2 入 - 出 5' 反应的数目	7(自 56 个)
ETIP/FAD2 入 - 出 3' 反应的数目	0(自 56 个)
基因座破坏 qPCR 测试的数目	9(自 56 个)
ETIP 中靶 (on-target) (Southern)	0(自 14 个)

[0333] 用 PDAS000130 和 PDAB104010 转化的 ETIP 转基因芥花的结果

[0334] 通过 pDAS000130 和 pDAB104010 转化产生的欧洲油菜转基因事件导致来自 pDAS000130 的 ETIP 多核苷酸序列的单拷贝、全长 T 链插入物被整合在 FAD2A 基因座内。对 3-4 个事件进行了完全表征,并确认含有被整合的 ETIP。确认是使用入 - 出 PCR 扩增方法,并进一步通过 Southern 印迹进行验证。使选出的 T₀事件生长至 T₁发育阶段。对 T₁植物进行再次筛选,以确定所整合 T- 链的接合性。将所筛选的事件分类为纯合子、半合子或空。

[0335] 通过前述的方法使用纯合子事件产生原生质体。随后用至少一个锌指核酸酶和供体质粒共转染原生质体,该锌指核酸酶被设计成靶向纳入在 ETIP 序列内的结合位点,该供体质粒与 ETIP 的特定区域具有同源性,其中该供体通过 HDR 机制被整合在 ETIP 内。类似地,随后用至少一个锌指核酸酶和供体质粒共转染原生质体,该锌指核酸酶被设计成靶向纳入在 ETIP 序列内的结合位点,该供体质粒与 ETIP 的特定区域不具有同源性,其中该供体通过非同源末端连接机制被整合在 ETIP 内。ZFN 剪切 ETIP 基因座,并且供体质粒通过同源性介导的修复或非同源末端连接被整合在欧洲油菜细胞的基因组内。

[0336] 作为该供体质粒整合的结果,部分 DS-red 转基因被修复成长全 DS-red 转基因。利用此时功能齐全 DS-red 转基因的表达来在 FACS 方法中分选原生质体细胞。用实施例 7 中所述的 FACS 方法分选推定的转基因植株,并将分离的原生质体再生成成熟植株。使用分子确认方法,确认供体质粒被整合在 ETIP 靶向的植株中。这样,ETIP 基因座发挥位点特异性基因座的作用,用于供体多核苷酸序列的基因靶向整合。

[0337] 实施例 7: 基于 FACS 的原生质体细胞分选

[0338] 对于用 DS-Red 对照构建体 pDAS000031 转染的欧洲油菜原生质体,使用 BD Biosciences Influx-Cell sorter™(San Jose, CA),通过 FACS 介导的细胞分选法进行分选。原生质体细胞的分离和转染如实施例 3 所述。在细胞用 pDAS000031 转染之后,使用 FACS 分选器利用表 7 所述的条件对细胞进行分选。

[0339] 表 7: 用于分选被 pDAS000031 转染的原生质体细胞的调节

[0340]

参数	
滴落频率 (Drop frequency)	6.1 KHz
喷嘴直径	200 μ m
鞘压力	4 psi
回收培养基	W5 培养基
培养调节	珠子型培养, 使用 sea-plaque 琼脂糖和海藻酸钠
分选标准	根据叶绿体自发荧光、报告基因表达 (Ds-Red) 进行分选
分选回收率 (%)	50-75

[0341]

分选后存活能力(%)	>95
------------	-----

[0342] 分选和分离表达 DS-red 转基因的原生质体。FACS 分离的原生质体用分选器进行计数。在 FACS 分离后的第一天, 将大约 1×10^5 – 1.8×10^5 个细胞置于 24 孔微滴定板的孔内。将细胞转移至珠子培养历时 5–20 天。测试相似的条件, 其中在 FACS 分离后的第二天, 将大约 1×10^4 个细胞置于 24 孔微滴定板的孔内。测试的各种条件导致回收了具有一定活力的细胞, 或者 95–98% 的总分离原生质体细胞。将 FACS 分选的原生质体细胞转移到珠子培养历时 3–20 天。在含有 1.5mg/mL 潮霉素的培养基上使用上述的实验方案将 FACS 分选的原生质体细胞再生成植株。推定的转基因植株通过分子确认方案被确认含有来自 pDAS000031 的完整 T 链插入物。

[0343] FACS 分选方法可直接用于筛选任何荧光转基因序列, 并用于分离一部分被用荧光转基因打靶的欧洲油菜原生质体细胞, 所述荧光转基因通过同源性介导的修复被靶向在基因组基因座中 ETIP 区的特定定位点内。

[0344] 实施例 8: 欧洲油菜 ω -3 脂肪酸去饱和酶 (FAD2) 通过同源性介导的修复的靶向整合和破坏

[0345] 特异针对 FAD2 的锌指结合结构域的选择

[0346] 鉴定并表征同源 FAD2 基因的被转录区域, 设计结合并剪切这些位点的锌指核酸酶用于供体序列的 NHEJ 介导的打靶。如上所述地设计针对来自 FAD2 序列同源物的 DNA 序列的锌指蛋白 (ZFN) 并进行测试。从显示中靶 (on-target) 活性的 ZFN 中, 选择一个可高

效切割 FAD2 靶标的锌指蛋白 :ZFP24828-2A-24829 识别 SEQ ID NO:355'-agGCCCAGtAGAGA GGCCaggcgaagta-3'和 SEQ ID NO:365'-ccAGGGCTGCGTCCTAACCGgcgtcttg-3'。这个 ZFN 显示可特异性结合并剪切 FAD2A 基因座。

[0347] 用于 HDR 指导的 DNA 修复的“供体”载体的设计和构建

[0348] 为了通过 HDR 整合供体序列,构建了单个载体。该载体编码潮霉素 (hph 或 hpt) 抗性基因表达盒。该潮霉素抗性基因表达盒包括含有来自花椰菜花叶病毒 (CaMV) 5' UTR 的 19S 启动子 (Cook and Penon Plant Molecular Biology 1990 14(3), 391-405), 随后是潮霉素磷酸转移酶 (hph) 基因 (Kaster et al Nucleic Acids Research 1983 11(19), 6895-6911)。hph 基因被密码子优化,用于在双子叶植物中表达,并且侧翼是含有根癌土壤杆菌 pTi15955 开放阅读框 1 (ORF1) 的转录终止子和聚腺苷酸化位点的 3' UTR (Barker et al, Plant Molecular Biology 1983, 2(6), 335-50)。表达盒由商业基因合成供应商 (GeneArt, Life Technologies, Regensburg, Germany) 合成。侧翼 FAD2A 序列添加在该基因表达盒的上游和下游。将潮霉素抗性盒克隆在每个载体的特定限制酶位点中,产生“供体”载体 :SEQ ID NO:87 (图 14)。

[0349] 通过对从大肠杆菌过夜培养物纯化的 DNA 的限制性核酸内切酶消化,对组装好的质粒的菌落进行初始筛选。限制性核酸内切酶从 NEW ENGLAND BIOLABS™ (NEB, Ipswich, MA) 和 PROMEGA™ (Promega Corporation, WI) 获得。质粒制备使用 QIAPREP SPIN MINIPREP KIT™ (Qiagen, Hilden, Germany) 或 PURE YIELD PLASMID MAXIPREP SYSTEM™ (Promega Corporation, WI) 按照供应商的使用说明进行。在通过对所得片段的琼脂糖凝胶电泳确认了限制性片段之后,使用 ABI Sanger 测序和 BIG DYE TERMINATOR V3.1™ 循环测序方案 (Applied Biosystems, Life Technologies) 对所选克隆的质粒 DNA 进行测序。序列数据的组装和分析使用 SEQUENCHER™ 软件 (Gene Codes, Ann Arbor, MI)。

[0350] 欧洲油菜的转化

[0351] PEG-介导的原生质体转染和靶向到 FAD2a 和 FAD2a ZFN 的 ETIP (精确事件) 的再生

[0352] 叶肉来源的原生质体从欧洲油菜 (DH10275) 3 周龄无菌芽培养物中分离。使相应的种子萌发。种子用 70% 乙醇表面灭菌 1 分钟,温柔震荡,随后在无菌双蒸水中漂洗 3-4 次,随后用 20% 漂白剂和 10 μl 吐温 20™ 进行灭菌;种子在摇床 (台式旋转摇床大约 100RPM) 上用漂白剂处理 15 分钟,然后在无菌双蒸水中漂洗 3-4 次。将种子小心转移到无菌滤纸上,以除去多余的水分,并铺在种子萌发培养基上 (1/2 强度的 MS/B5 维生素 +1% 蔗糖 +0.8% 琼脂; pH 5.8, 每个 Petri™ 培养皿 (15X 100mm) 倾倒 50-60ML 培养基,用支撑物将平板以微小的角度放置)。每个平板上放置大约 50 粒种子。将平板在 16h/d 光照 (20 μmol m⁻²s⁻¹) 下的 22℃ 直立温育 6 天。从 6 天龄幼苗切下 0.5cm 大小的下胚轴段,并培养在芽诱导培养基上 (MS/B5 维生素 +3% 蔗糖 +500mg/L MES+BAP (13 μm) + 玉米素 (5 μm) + 硝酸银 (5mg/L) +0.8% 琼脂 (pH 5.8), 倾倒到 100x 20mm 无菌 PETRI™ 培养皿内), 在每个平板上放置大约 20 个外植体。将 3-4 周后出现的芽分生组织转移到芽伸长培养基上 (MS/B5 维生素 +2% 蔗糖 +500mg/L MES+BAP (2 μm) +GA-3 (0.1 μm) +0.8% 琼脂 (pH 5.8), 并倾倒在 250mL 培养容器中), 在一轮亚培养期间,将培养物保持在这一培养基中达 4 周。然后,将 2-3cm 高的芽转移到发根培养基上 (1/2 强度 MS/B5 维生素 +1% 蔗糖 +500mg/

L MES+IBA (2.5 μm)+0.6%琼脂 (pH 5.8),并倾倒在700mL培养容器中),用于根系发育。将发根的芽亚培养在新鲜的发根培养基中,以3-4周间隔进行枝条扦插,经过2-3轮后使用。在整个期间,将培养物保持在22°C,16h/d光照(30 $\mu\text{mol m}^{-2}\text{s}^{-1}$)。

[0353] 原生质体的分离和纯化

[0354] 体外生长的DH12075欧洲油菜植株用作分离叶肉原生质体的外植体源。使用锋利的手术刀将3-4周龄小植株的上面第3-第4片完全伸展的叶片切成小条(0.5-1mm)。酶消化如下实施:用25mL消化缓冲液(1.2% (w/v) 纤维素酶"Onozuka" R10®和0.2% (w/v) Macerozyme® R10溶于K4培养基(Spangenberg et al. 1998)中)处理250-500mg叶片材料。用Parafilm™密封含有叶片材料和消化缓冲液的PETRI™培养皿,并在室温下黑暗温育12-15h。过夜温育后,通过BD®细胞过滤网(网眼尺寸为70 μm)过滤消化物。将原生质体悬液(5-6mL)收集在14mL圆底管中,并用1mL清洗液W5缓冲液(154mM NaCl, 125mM CaCl₂, 5mM KCl和5mM蔗糖;pH 5.8Menzel et al. 1981)覆于上层,以400RPM离心10min。离心后,取出浮在中间相中的原生质体,并通过400RPM离心10min用10mL W5缓冲液进行清洗。最后一次清洗后,将分离的原生质体重悬,密度为每毫升W5缓冲液1X 10⁶个原生质体,并在转染之前温育1小时。

[0355] 原生质体产量和活力的评估

[0356] 原生质体产量用血细胞计数仪按照Sambrook and Russel, 2006所述进行评估,活力用埃文斯蓝染色(400mg/L埃文斯蓝染料溶解于0.5M甘露醇中)按照Huang et al. 1996所述方案进行测试,进行了很少的次要修改。

[0357] PEG 4000介导的稳定DNA投送

[0358] 使用Pure Yield Plasmid Maxiprep System™(Promega Corporation, WI)或Plasmid Maxi Kit™(Qiagen, Hilden)按照供应商的使用说明从大肠杆菌培养物中分离含有ETIP的载体pDAS000129和靶向FAD2A基因座的ZFN载体(pDAB104010)的质粒DNA。将质粒DNA以0.7 $\mu\text{g}/\mu\text{L}$ 的密度溶解在无菌双蒸水中。向悬浮在100 μL 转化缓冲液(15mM MgCl₂, 0.1% (w/v) 吗啉代乙磺酸(MES)和0.5M甘露醇;pH 5.8)中的100万个原生质体(活力 ≥ 95)施加总共30微克(30 μg)质粒DNA(pDAS000129和pDAB104010摩尔比为5:1),随后是150 μL PEG溶液(40% (w/v) PEG 4000溶于0.4M甘露醇和0.1M Ca(NO₃)₂ (pH 6-7), Spangenberg and Potrykus(1995))。对照转化包括总共30微克(30 μg) pDAS000129或pDAB104010的质粒DNA。在室温下温育10-15分钟后,以逐滴方式加入5mL W5缓冲液,轻轻混合原生质体,并将另外的5mL W5缓冲液缓慢流加到原生质体混悬液中。轻轻混合原生质体,并以400RPM离心10分钟,小心除去W5上清液,留下沉淀饼形式的原生质体。然后将被转染的原生质体在1mL W5缓冲液中在室温下进行温育,直至它们将包埋在珠子型培养物中。在转染后即刻,用构建体pDAS000129和pDAB104010共转染的原生质体样品显示了60-80%范围的细胞活力。按照seaplaque琼脂糖或海藻酸钠法包埋被转染的原生质体。

[0359] 培养叶肉来源的原生质体以回收活的微愈伤组织(microcalli)

[0360] 在包埋前,将被转染的原生质体以400RPM离心10min,并小心除去W5缓冲液。然后,将原生质体重悬在0.5mL K3培养基(Spangenberg et al. 1998)中。将精确的0.5mL被转染原生质体混悬液(约5x 10⁵个原生质体)置于6cm培养皿中,向其中添加

4. 5ml 预热的 (在微波炉中熔化,并在 40–45℃ 水浴中温育) K3:H 培养基 (Spangenberg et al. 1998) 1:1 混合物,其含有 0.6% Sea Plaque™ 琼脂糖。轻轻混合琼脂糖和原生质体悬液,并使其凝固。凝固后 (20–30min 后),用 Parafilm® 密封培养皿,并将原生质体在 24℃ 黑暗培养 24h,随后在连续的昏暗灯光下 ($5\text{--}10\ \mu\text{mol m}^{-2}\text{s}^{-1}$) 培养 6 天,其中发生第一次和多次细胞分裂。6 天后,将包埋在琼脂糖内的原生质体切成四等份,并置于 700ml 培养容器内的 100ml A 培养基中 (Spangenberg et al. 1998)。液体 A 培养基补充了 1.5mg/l 潮霉素。培养物在连续昏暗灯光下在旋转摇床上以 80–100RPM 在 24℃ 下进行温育。在 Sea-plaque 琼脂糖和藻酸钠法的情况下,抗性菌落分别于原生质体涂板后 5–6 周和 3–4 周后出现。通过轻柔破碎琼脂糖珠子,将大小为 2–3mm 直径的微愈伤组织转移到 B1 培养基上 (MS/MS 维生素 +3.5% 蔗糖 +500mg/L MES+BAP ($5\ \mu\text{m}$) +NAA ($5\ \mu\text{m}$) +2,4-D ($5\ \mu\text{m}$) +1.5mg/L 潮霉素 +0.7% I 型琼脂糖 (pH 6.0),并倾倒在 100x 20mm 无菌培养皿中)。将这样获得的微愈伤组织重悬在足够量的液体 A 中 (50ml 液体 A 用于 1 毫升沉降细胞体积 (SCV: 这是在将全部释放的愈伤组织转移到 50 或 15ml falcon 管中并使之沉降 5min 之后测量的)。在将微愈伤组织均匀混合后,将 0.5ml 悬浮在液体 A 中的微愈伤组织转移到 B1 平板上,并使用 1–2ml 额外的液体 A 培养基将微愈伤组织均匀分布在 B 培养基中,并从每个平板小心除去多余的液体 A。平板用微孔带密封,其可增强胚成熟。

[0361] 海藻酸钠法

[0362] 在包埋前,将被转染的原生质体以 400RPM 离心 10min,并小心除去 W5 缓冲液。然后,将原生质体重悬在 1.0ml 0.5M 甘露醇中,并在冰上温育。向其添加等体积的 1.0% 海藻酸钠,并轻轻混合。将原生质体悬液在冰上温育,直至其被包埋。用血清吸管将成珠溶液 (0.4M 甘露醇 +50mM CaCl_2 (pH 5.8)) 转移到无菌的 6 孔板中 (每孔 3–4mL)。使用 1mL 移液管将精确的 1.0mL 原生质体悬液逐滴添加到成珠溶液中,并且每个孔包埋每一个被转染样品 (大约 5×10^5 个原生质体)。原生质体悬液在室温下温育 1–2 小时以形成海藻酸钠珠子。温育期之后,小心除去成珠溶液,并更换为补充有 1.5mg/L 潮霉素的 4–5mL K3+H:A 培养基 (Spangenberg et al. 1998) 的 1:2 混合物。将原生质体在摇床上 (50RPM) 22℃ 黑暗培养 3–4 周。3–4 周后,用解聚缓冲液 (0.3M 甘露醇 +20mM 柠檬酸钠 (pH 5.8)) 处理来释放抗性的微愈伤组织 (0.5–1.0mm)。除去液体培养基之后,向含有珠子型培养物的每个孔添加 3–4mL 解聚缓冲液,并在室温下温育 2 小时。使用无菌镊子轻轻混合珠子,以提高微愈伤组织的高效释放。使用无菌的 1.0mL 移液管轻轻混合并除去在解聚缓冲液中释放的胶凝剂 (gelling agent)。微愈伤组织用 5mL 液体 A 培养基清洗 2 次,并将微愈伤组织重悬浮在足量的液体 A 中 (50mL 液体 A 用于 1 毫升沉降细胞体积 (SCV: 这是在将全部释放的愈伤组织转移到 50 或 15mL falcon 管中并使之沉降 5min 之后测量的)。在将微愈伤组织均匀混合后,将 0.5mL 悬浮在液体 A 培养基中的微愈伤组织转移到 B1 培养基 (MS/MS 维生素 +3.5% 蔗糖 +500mg/L MES+BAP ($5\ \mu\text{m}$) +NAA ($5\ \mu\text{m}$) +2,4-D ($5\ \mu\text{m}$) +1.5mg/L 潮霉素 +0.7% I 型琼脂糖 (pH 6.0),并倾倒在 100x 20mm 无菌 Petri 培养皿中),并使用 1–2mL 额外的液体 A 培养基将微愈伤组织均匀分布在 B1 培养基中,并从每个平板小心除去多余的液体 A 培养基。平板用微孔胶带密封,其可增强胚成熟。将培养物保持在 22℃,16h/d 光照 ($30\ \mu\text{mol m}^{-2}\text{s}^{-1}$)。

[0363] 从叶肉原生质体分离基因组 DNA

[0364] 将被转染的原生质体从 3cm PETRI™ 皿转移到 2mL 微离心管中。通过 70g 离心沉

淀细胞,并除去上清液。为了最大限度回收被转染的原生质体, PETRI™皿用 1mL 清洗缓冲液漂洗 3 次。每次漂洗将 PETRI™皿中的清洗缓冲液涡旋 1 分钟,随后将液体转移到相同的 2mL 微离心管中。在每次漂洗结束时,通过 70g 离心沉淀细胞,并除去上清液。将沉淀的原生质体在液氮中速冻,随后在 Labconco Freezone 4.5® (Labconco, Kansas City, MO) 中在 -40℃和 133×10^3 mBar 压力下进行冷冻干燥 24h。使用 DNeasy®植物 DNA 微量提取试剂盒 (Qiagen) 按照制造商的说明对冻干细胞进行 DNA 提取,只是不要求组织破坏而是将原生质体细胞直接加入到裂解缓冲液中。

[0365] 从愈伤组织分离基因组 DNA

[0366] 将单个愈伤组织在液氮中速冻,随后在 Labconco Freezone 4.5® (Labconco, Kansas City, MO) 中在 -40℃和 133×10^3 mBar 压力下进行冷冻干燥 24h。使用 DNeasy®植物 DNA 微量提取试剂盒 (Qiagen, Hilden, Germany) 按照制造商的说明对冻干的愈伤组织进行 DNA 提取。

[0367] 在繁殖的愈伤组织中检测通过同源性介导的修复添加到 FAD2A 的基因

[0368] 从原生质体池 (每个池含有 100 万个原生质体) 提取基因组 DNA,其中在 24 小时前已经向原生质体池投送了编码功能性 HGH 报告盒 (pDAS000129) 的供体 DNA, ZFN DNA (pDAB104010), 或供体和 ZFN DNA 的混合物。用于转化的 DNA 投送量如上文所述。将 PCR 产物克隆到质粒载体中。在每个细胞中独立发生基因组编辑,产生若干不同的插入事件,通过克隆到质粒载体中,可以对每个基因组编辑进行无歧义的测序。在 ABI3730XL® 自动毛细管电泳平台上对几个克隆进行了测序。基因序列分析使用 SEQUENCHER SOFTWARE V5.0™ (GeneCodes, Ann Arbor, MI) 完成。

[0369] 使用表 13 所述的引物从提取自原生质体的基因组 DNA 中同时扩增出 5' 和 3' FAD2A 盒和接点提供了通过同源性介导的修复添加基因到 FAD2A 基因座中的证据。从单独投送了 ZFN 质粒或供体质粒的原生质体中没有观察到扩增。所有接点序列均可指示 hgh 盒通过 HDR 介导的修复途径插入到 Fad2A 基因座处。观察到了基因组和盒子中任一个或两个的各种长度的缺失,以及来自载体骨架 (或是来自供体,或是来自 ZFN) 的额外序列被插入在基因组和盒子之间 (图 15)。

[0370] 表 13: 用于 FAD2A 基因座内供体整合的构建体的列表

[0371]

处理	构建体	Hph 测定	入-出 PCR (LB)	入-出 PCR (RB)
ZFN	pDAB104010 (无 Hyg)	NO	NO	NO
ZFN	pDAB104010 (1.5mg/L Hyg)	NO	NO	NO
DNR	pDAS000129 (无 Hyg)	YES	NO	NO
DNR	pDAS000129 (1.5mg/L Hyg)	YES	NO	NO
DNR + ZFN2c	pDAS000129 + pDAB104010 (无 Hyg)	YES	NO	NO
DNR + ZFN2c	pDAS000129 + pDAB104010 (1.5mg/L Hyg)	YES	YES	YES
DsRed Ctrl	pDAS00097 (2mg/L PPT)	NO	NO	NO
阴性对照	未转染的 DH12075	NO	NO	NO
阴性对照	水	NO	v	NO

[0372] 实施例 9: 通过非同源末端连接实现欧洲油菜 Ω -3 脂肪酸去饱和酶 (FAD2) 的靶向整合和破坏

[0373] 特异针对 FAD2A 的锌指结合结构域的选择

[0374] 鉴定并表征了同源 FAD2 基因的被转录区域, 设计结合并剪切这些位点的锌指核酸酶, 用于供体序列的 NHEJ 介导的靶向。如上所述地设计导向针对来自 FAD2 序列同源物的 DNA 序列的锌指蛋白 (ZFN), 并进行测试。从显示中靶 (on-target) 活性的 ZFN 中, 选择一个可高效切割 FAD2 靶标的锌指蛋白: ZFP 24828-2A-24829 识别 SEQ ID NO:355'-agCCCC AGtAGAGAGGCCcaggcgaagta-3' 和 SEQ ID NO:365'-ccAGGGCTGCGTCCTAACCGgcgtcttg-3'。这个 ZFN 显示可特异性结合并剪切 FAD2A 基因组基因座。构建了在前文中有描述的质粒构建体 pDAB104010, 并用于转化实验。

[0375] 设计和构建编码锌指核酸酶的表达载体和特异针对 FAD2A 的供体载体

[0376] 为了通过 NHEJ 整合供体序列, 构建了单个供体载体。该载体编码 yige dsRED 报告基因表达盒。该 dsRED 报告基因表达盒包括拟南芥泛素 10 启动子 (Callis, et al., 1990, J. Biol. Chem., 265:12486-12493), 随后是 dsRED 基因 (Dietrich et al. (2002) Biotechniques 2(2):286-293)。dsRED 基因被密码子优化用于在双子叶植物中表达, 并且侧翼是含有根癌土壤杆菌 pTi15955 开放阅读框 23 (ORF23) 的转录终止子和聚腺苷酸化位点的 3' UTR (Barker et al, Plant Molecular Biology 1983, 2(6), 335-50)。可选择标记盒包括与 pat 转基因融合的 CsVMV 启动子。pat 转基因终止于根癌土壤杆菌 pTi15955 开放阅读框 1 (ORF1) (Barker et al, Plant Molecular Biology 1983, 2(6), 335-50)。dsRED 抗性盒被克隆到每个载体的特异限制酶位点, 产生“供体”载体: pDAS000097 (SEQ ID NO:88, 图 16)。pDAS00097 供体被设计为作为线性 DNA 或环状 DNA 投送到植物细胞中, 并在 ZFN

pDAB104010 剪切 FAD2A 基因座时整合在 FAD2A 基因座内。线性 DNA 介导的整合是线性化 pDAS000097 质粒在转化过程中整合到植物细胞中的结果。该质粒可以通过独特限制位点处的剪切被线性化。环状 DNA 介导的整合是环化 PDA000097 质粒在转化过程中整合到植物细胞中的结果。pDAS000097 被修饰成含有能够被 ZFP24828-2A-24829 锌指核酸酶切割的锌指结合位点。环状质粒 pDAS000097 在植物细胞中被 pDAB104010 编码的锌指核酸酶切割,并且 dsRED 基因盒被整合到 FAD2A 基因组基因座中。

[0377] 通过对从大肠杆菌过夜培养物纯化的 DNA 的限制性核酸内切酶消化,对组装质粒的菌落进行初始筛选。限制性核酸内切酶从 NEW ENGLAND BIOLABS™(NEB, Ipswich, MA) 和 PROMEGA™(Promega Corporation, WI) 获得。质粒制备使用 QIAPREP SPIN MINIPREP KIT™(Qiagen, Hilden, Germany) 或 PURE YIELD PLASMID MAXIPREP SYSTEM™(Promega Corporation, WI) 按照供应商的使用说明实施。在通过对所得片段的琼脂糖凝胶电泳确认了限制性片段之后,使用 ABI Sanger 测序和 BIG DYE TERMINATOR V3.1™周期测序方案 (Applied Biosystems, Life Technologies) 对所选克隆的质粒 DNA 进行测序。序列数据的组装和分析使用 SEQUENCHER™软件 (Gene Codes, Ann Arbor, MI)。

[0378] 欧洲油菜的转化

[0379] 如上所述地从欧洲油菜 (DH10275) 植株分离和制备叶肉来源的原生质体。原生质体用纯合的质粒 DNA 转染。等份的供体和 ZFN 质粒 DNA 制备成三个摩尔比:1:1 (30 μg 每种质粒), 5:1 (供体质粒与 ZFN 质粒, 总共 30 μg 质粒 DNA) 和 10:1 (供体质粒与 ZFN 质粒, 总共 30 μg 质粒 DNA)。此外,制备仅有供体和仅有 ZFN 的等份 (30 μg) 作为对照。表 14 总结了通过 PEG4000 介导的转化被投送到欧洲油菜原生质体中的 DNA 的量。被转化的原生质体细胞被如前所述地培养,其中选择培养基是草铵膦选择培养基,对推定的转化子通过 qPCR 分析测定转基因插入。

[0380] 表 14: 投送到原生质体的 ZFN 和供体 DNA 的量

[0381]

	质粒 DNA 的摩尔比	投送到 100 万个原生质体中的 DNA 总量(μg)
剪接	仅供体质粒	30
	仅 ZFN 质粒(pDAB104010)	30
	1:1 供体:ZFN	60
	5:1 供体:ZFN	30
	10:1 供体:ZFN	30
编辑	仅供体质粒	30
	1:1: ZFN 质粒(pDAB104010)	30
	1:1:1 供体:ZFN:ZFN	90

[0382]

	5:1:1 供体:ZFN:ZFN	30
	10:1:1 供体:ZFN:ZFN	30

[0383] 在原生质体中检测通过非同源末端连接添加到 FAD2A 的基因

[0384] 从原生质体池（每个池含有 100 万个原生质体）提取基因组 DNA，其中在 24 小时前已经向原生质体池投送了编码功能性 dsRFP 报告盒（pDAS000097）的供体 DNA，ZFN DNA（pDAB104010），或供体与 ZFN DNA 的混合物。用于转化的 DNA 投送量如上文所述。将 PCR 产物克隆到质粒载体中。在每个细胞中独立发生基因组编辑，产生若干不同的插入事件，通过克隆到质粒载体中，可以对每个基因组编码进行无歧义测序。就几个克隆在 ABI3730XL® 自动毛细管电泳平台上被测序。基因序列分析使用 SEQUENCHER SOFTWARE V5.0™（GeneCodes, Ann Arbor, MI）完成。

[0385] 从提取自原生质体的基因组 DNA 中扩增出 5' 和 3' FAD2A- 盒接点二者，提供了基因通过编辑或剪切被添加到 FAD2A 基因座中的证据。从单独投送了 ZFN 质粒或供体质粒的原生质体中没有观察到扩增。所有接点序列均可指示 dsRED 盒通过 NHEJ 介导的修复途径插入到 FAD2A 基因座处。观察到基因组和盒子中任一个或两个的各种长度的缺失，以及来自载体骨架（来自供体或 ZFN）的序列被添加插入到基因组和盒子之间。

[0386] 在从原生质体再生的愈伤组织中检测通过非同源末端连接添加到 FAD2A 的基因

[0387] 从自接受选择的原生质体再生的愈伤组织获得了 FAD2A 基因座被剪切和编辑的进一步的证据，其中向原生质体中投送了编码 dsRED 盒的供体 DNA（pDAS000097），仅 ZFN DNA（pDAB104010），或供体和 ZFN DNA。对于每个比例，从大约 80 个愈伤组织中提取了 DNA。

[0388] 通过 TAQMAN™ qPCR 使用特异针对供体插入物和基因组侧翼序列的引物和探针确认了 dsRED 盒被整合到欧洲油菜基因组中。相对定量的计算是根据 $2^{-\Delta\Delta C_t}$ 方法（Livak and Schmittgen, 2001），其估算被插入到基因组中的 dsRED 盒的拷贝数。通过用一个特异针对 FAD2A 的引物和特异针对 dsRED 盒的启动子或终止子的第二引物进行 PCR 测定，获得了 FAD2A 被 NHEJ 介导的剪切和编辑的证据。PCR 产物用 QIAQUICK MINIELUTE PCR PURIFICATION KIT™（Qiagen）进行凝胶纯化，并用直接 Sanger 测序方法进行测序。测序产物用乙醇、乙酸钠和 EDTA 按照 BIGDYE® v3.1 方案（Applied Biosystems）进行纯化，并如上所述地进行测序和分析。

[0389] 确定了每个实验中含有供体盒的愈伤组织的数目。通过跨越 ZFN 切割位点和 5' 和 3' FAD2A-dsRED 盒接点的 PCR 扩增，提供了供体基因通过编辑和 / 或剪切添加到 FAD2A 基因座的证据。对从仅用 dsRED 质粒（pDAS000097）或仅用 ZFN 质粒（pDAB104010）转化的对照原生质体回收的愈伤组织中分离的基因组 DNA 进行 PCR 扩增，没有产生 PCR 扩增产物。

[0390] 从琼脂糖凝胶纯化从 5' 和 3' FAD2A-dsRED 盒接点扩增产生的 PCR 扩增子，并测序以确认 FAD2A 基因组基因座内的整合特异性。PCR 产物测序分析结果表明，从各个被转化原生质体产生的每个分离愈伤组织仅产生了单一 PCR 扩增产物，不含有具有混合基因型的细胞。

[0391] 植物中通过非同源末端连接添加到 FAD2A 的基因的检测

[0392] 从自原生质体再生并转移到盆栽培养基（potting medium）中的植株提取 DNA。估

计所回收的大部分植株仅含有供体 DNA 编码的 dsRED 盒的 1-2 个拷贝。使用与愈伤组织所述相同序列的测定对植株进行分析,并通过测定确定盒子是否已经插入在 FAD2A 基因座中。

[0393] 使用上述的 PCR 测定确定中靶剪接(其中 dsRED 盒被插入在 FAD2A 基因座中)的频率。对所得扩增子条带进行测序,以确定侧翼序列。此外,筛选植株的脱靶插入,以确定 dsRED 盒整合在 FAD2A 之外的位点处的频率。

[0394] 实施例 10: 欧洲油菜 Ω -3 脂肪酸去饱和酶 (FAD2) 和农艺学重要基因的靶向整合

[0395] 设计并构建含有赋予对除草剂草甘膦的抗性的 DGT-28 转基因的构建体(国际专利申请 No. WO/2013/116700, 本文引用其内容作为参考)用于整合在欧洲油菜的 FAD2A 基因组基因座中。示例性供体构建体包括 pDAS000389(图 17, SEQ ID NO:89),用于 NHEJ 整合在 FAD2A 基因座内, pDAS000391(图 18, SEQ ID NO:90),用于 NHEJ 整合在 FAD2A 基因座内, pDAS000392(图 19, SEQ ID NO:91),用于 NHEJ 整合在 FAD2A 基因座内, pDAS000393(图 20, SEQ ID NO:92),用于 NHEJ 整合在 FAD2A 基因座内, pDAS000394(图 21, SEQ ID NO:93),用于 HDR 整合在 FAD2A 基因座的 ETIP 位点内, pDAS000395(图 22, SEQ ID NO:208),用于 HDR 整合在 FAD2A 基因座内, pDAS000396(图 23, SEQ ID NO:209),用于 HDR 整合在 FAD2A 基因座内,和 pDAS000397(图 24, SEQ ID NO:210),用于 HDR 整合在 FAD2A 基因座内。如上所述地,将构建体和相关锌指核酸酶构建体(例如, pDAB104010)转化到欧洲油菜细胞中。通过如前文所述的分子确认测定对转化子进行鉴定和确认。分离包含被整合 dgt-28 转基因的 FAD2A 染色体整合子。dgt-28 转基因整合在 FAD2A 基因座内的实例是通过 NHEJ 介导的整合和 HDR 介导的整合。整合在 FAD2A 基因座内可以被导向到 FAD2A 内源序列中,或者导向到先前所述的、被稳定整合在 FAD2A 基因座内的 ETIP(pDAS000130)中。通过 NHEJ 介导的机制整合在 FAD2A 基因座内,能够使用线性化供体或环状供体 DNA 设计实现。获得被转化的 DGT-28 欧洲油菜事件,并测试 DGT-28 的强表达和因此导致的对除草剂草甘膦的抗性。

[0396] 尽管本文描述了某些示例性实施方案,但是本领域普通技术人员将认识并意识到,可以在不背离下述权利要求的范围下对示例性实施方案进行许多添加、缺失和修改。此外,来自一个实施方案的特征可以和另一个实施方案的特征相结合。

	1	40
FAD2-3 (SEQ ID NO:8)	(1) ATGGGCGCAGGTGGAAGAATGCAAGTCTCTCCTCCCTCCA	
FAD2A (SEQ ID NO:5)	(1) ATGGGCGCAGGTGGAAGAATGCAAGTCTCTCCTCCCTCCA	
FAD2-2 (SEQ ID NO:7)	(1) ATGGGCGCAGGTGGAAGAATGCAAGTCTCTCCTCCCTCCA	
FAD2-1 (SEQ ID NO:6)	(1) ATGGGCGCAGGTGGAAGAATGCAAGTCTCTCCTCCCTCCA	
	41	80
FAD2-3 (SEQ ID NO:8)	(41) AGAAAGGGAACCAACATCAACGGGTACCCCTGCGA	
FAD2A (SEQ ID NO:5)	(41) AGAAAGGGAACCAACATCAACGGGTACCCCTGCGA	
FAD2-2 (SEQ ID NO:7)	(41) AGAAAGGGAACCAACATCAACGGGTACCCCTGCGA	
FAD2-1 (SEQ ID NO:6)	(41) AGAAAGGGAACCAACATCAACGGGTACCCCTGCGA	
	81	120
FAD2-3 (SEQ ID NO:8)	(81) GACACCCCTTCACTCTCGGAGAACTCAAGAAAGCAATC	
FAD2A (SEQ ID NO:5)	(81) GACACCCCTTCACTCTCGGAGAACTCAAGAAAGCAATC	
FAD2-2 (SEQ ID NO:7)	(81) GACACCCCTTCACTCTCGGAGAACTCAAGAAAGCAATC	
FAD2-1 (SEQ ID NO:6)	(81) GACACCCCTTCACTCTCGGAGAACTCAAGAAAGCAATC	
	121	160
FAD2-3 (SEQ ID NO:8)	(121) CCACCGCACTGTTCAAACGCTCATCCCGCCTCTTCT	
FAD2A (SEQ ID NO:5)	(121) CCACCGCACTGTTCAAACGCTCATCCCGCCTCTTCT	
FAD2-2 (SEQ ID NO:7)	(121) CCACCGCACTGTTCAAACGCTCATCCCGCCTCTTCT	
FAD2-1 (SEQ ID NO:6)	(121) CCACCGCACTGTTCAAACGCTCATCCCGCCTCTTCT	
	161	200
FAD2-3 (SEQ ID NO:8)	(161) CCTACCTCTCGGAGCAT--CATCTAGCTCCTCTCTC	
FAD2A (SEQ ID NO:5)	(161) CCTACCTCTCGGAGCAT--CATCTAGCTCCTCTCTC	
FAD2-2 (SEQ ID NO:7)	(161) CCTACCTCTCGGAGCAT--CATCTAGCTCCTCTCTC	
FAD2-1 (SEQ ID NO:6)	(161) CCTACCTCTCGGAGCAT--CATCTAGCTCCTCTCTC	
	201	240
FAD2-3 (SEQ ID NO:8)	(199) TACCACTCTCGGAGCAT--CATCTAGCTCCTCTCTC	
FAD2A (SEQ ID NO:5)	(199) TACCACTCTCGGAGCAT--CATCTAGCTCCTCTCTC	
FAD2-2 (SEQ ID NO:7)	(199) TACCACTCTCGGAGCAT--CATCTAGCTCCTCTCTC	
FAD2-1 (SEQ ID NO:6)	(200) TACCACTCTCGGAGCAT--CATCTAGCTCCTCTCTC	
	241	280

图 1A

FAD2-3 (SEQ ID NO:8) (239) GGGGCTCTACGCGGCTGGCCCTCTACTGGGCTGCCA
FAD2A (SEQ ID NO:5) (239) GGGGCTCTACGCGGCTGGCCCTCTACTGGGCTGCCA

图 1B

FAD2-2 (SEQ ID NO:7)	(239)	CACGCGCCCTACCTCGCCTGCCCCCTCTACTGGGGCTGCCA	
FAD2-1 (SEQ ID NO:6)	(231)	-----CTACCTCGCCTGACCCCTCTACTGGGGCTGCCA	
		281	320
FAD2-3 (SEQ ID NO:8)	(279)	AGGGTGGCTCCTAACCGGCCTCTGGGTCATAGCCCAAGAG	
FAD2A (SEQ ID NO:5)	(279)	GGCTTGGCTCCTAACCGGCCTCTGGGTCATAGCCCAAGAG	
FAD2-2 (SEQ ID NO:7)	(279)	AGGGTGGCTCCTAACCGGCCTCTGGGTCATAGCCCAAGAA	
FAD2-1 (SEQ ID NO:6)	(265)	AGGGTGGCTCCTAACCGGCCTCTGGGTCATAGCCCAAGAG	
		321	360
FAD2-3 (SEQ ID NO:8)	(319)	TGCGGCCACCAAGCCTTCAGCGACCAACAGTGGCTTGACG	
FAD2A (SEQ ID NO:5)	(319)	TGCGGCCACCAAGCCTTCAGCGACCAACAGTGGCTTGACG	
FAD2-2 (SEQ ID NO:7)	(319)	TGCGGCCACCAAGCCTTCAGCGACCAACAGTGGCTTGACG	
FAD2-1 (SEQ ID NO:6)	(305)	TGCGGCCACCAAGCCTTCAGCGACCAACAGTGGCTTGACG	
		361	400
FAD2-3 (SEQ ID NO:8)	(359)	ACACCGCTGGCTCTCATCTTCCACTCCTTCCTCCTCGTCCC	
FAD2A (SEQ ID NO:5)	(359)	ACACCGCTGGCTCTCATCTTCCACTCCTTCCTCCTCGTCCC	
FAD2-2 (SEQ ID NO:7)	(359)	ACACCGCTGGCTCTCATCTTCCACTCCTTCCTCCTCGTCCC	
FAD2-1 (SEQ ID NO:6)	(345)	ACACCGCTGGCTCTCATCTTCCACTCCTTCCTCCTCGTCCC	
		401	440
FAD2-3 (SEQ ID NO:8)	(399)	GTACTTCTCCTGGAAGTACAGTCATGACGCGCACCATTC	
FAD2A (SEQ ID NO:5)	(399)	GTACTTCTCCTGGAAGTACAGTCATGACGCGCACCATTC	
FAD2-2 (SEQ ID NO:7)	(399)	GTACTTCTCCTGGAAGTACAGTCATGACGCGCACCATTC	
FAD2-1 (SEQ ID NO:6)	(385)	GTACTTCTCCTGGAAGTACATCAT-GACGCGCACCATTC	
		441	480
FAD2-3 (SEQ ID NO:8)	(439)	AACACCGGCTCCCTCGAGAGAGAGGAAGTGTTCCTCCCA	
FAD2A (SEQ ID NO:5)	(439)	AACACCGGCTCCCTCGAGAGAGAGGAAGTGTTCCTCCCA	
FAD2-2 (SEQ ID NO:7)	(439)	AACACCGGATCCCTCGAGAGAGATGAAGTGTTCCTCCCA	
FAD2-1 (SEQ ID NO:6)	(424)	AACACCGGATCCCTCGATAGAGAGAGTGTTCCTCCCA	
		481	520

图 1B(续)

FAD2-3 (SEQ ID NO:8)	(479)	AGAAGAAATCAGACATCAAGTGGTACGGGAAGTACCTCAA
FAD2A (SEQ ID NO:5)	(479)	AGAAGAAATCAGACATCAAGTGGTACGGGAAGTACCTCAA
FAD2-2 (SEQ ID NO:7)	(479)	AGAAGAAATCAGACATCAAGTGGTACGGGAAGTACCTCAA
FAD2-1 (SEQ ID NO:6)	(464)	AGAAGAAATCAGACATCAAGTGGTACGGGAAGTACCTCAA
	521	560

图 1C

FAD2-3 (SEQ ID NO:8)	(519)	CAACCCCTGGACGACCGTGATGTAACGTTCAGTTTC
FAD2A (SEQ ID NO:5)	(519)	CAACCCCTGGACGACCGTGATGTAACGTTCAGTTTC
FAD2-2 (SEQ ID NO:7)	(519)	CAACCCCTAGGACGACCGTGATGCTAACCGTCCAGTTTC
FAD2-1 (SEQ ID NO:6)	(504)	CAACCCCTAGGACGACCGTGATGCTAACCGTCCAGTTTC
	561	600
FAD2-3 (SEQ ID NO:8)	(559)	AAGCTCGGCTGGCCCTTGTTACTTAGCCTTCAACGTCTCG
FAD2A (SEQ ID NO:5)	(559)	AAGCTCGGCTGGCCCTTGTTACTTAGCCTTCAACGTCTCG
FAD2-2 (SEQ ID NO:7)	(559)	AAGCTCGGCTGGCCCTTGTTACTTAGCCTTCAACGTCTCTG
FAD2-1 (SEQ ID NO:6)	(544)	AAGCTCGGCTGGCCCTTGTTACTTAGCCTTCAACGTCTCG
	601	640
FAD2-3 (SEQ ID NO:8)	(599)	GGAGACCTTACACCGCGCTTCGCTTCCCATTTCCACCC
FAD2A (SEQ ID NO:5)	(599)	GGAGACCTTACACCGCGCTTCGCTTCCCATTTCCACCC
FAD2-2 (SEQ ID NO:7)	(599)	GGAGACCTTACACCGCGCTTCGCTTCCCATTTCCACCC
FAD2-1 (SEQ ID NO:6)	(584)	GGAGACCTTACACCGCGCTTCGCTTCCCATTTCCACCC
	641	680
FAD2-3 (SEQ ID NO:8)	(639)	CAACGCTCCCATCTACAACGACCGGAGCGTCTCCAGATA
FAD2A (SEQ ID NO:5)	(639)	CAACGCTCCCATCTACAACGACCGGAGCGTCTCCAGATA
FAD2-2 (SEQ ID NO:7)	(639)	CAACGCTCCCATCTACAACGACCGGAGCGTCTCCAGATA
FAD2-1 (SEQ ID NO:6)	(624)	CAACGCTCCCATCTACAACGACCGGAGCGTCTCCAGATA
	681	720
FAD2-3 (SEQ ID NO:8)	(679)	TACATCTCCGACCGCTGGCATCCTCGCGTTCGCTACGGTC
FAD2A (SEQ ID NO:5)	(679)	TACATCTCCGACCGCTGGCATCCTCGCGTTCGCTACGGTC
FAD2-2 (SEQ ID NO:7)	(679)	TACATCTCCGACCGCTGGCATCCTCGCGTTCGCTACGGTC
FAD2-1 (SEQ ID NO:6)	(664)	TACATCTCCGACCGCTGGCATCCTCGCGTTCGCTACGGTC
	721	760

图 1C(续)

FAD2-3 (SEQ ID NO:8)	(719)	TCTTCCGTTACGCCGTCGTCAGGGAGTGGCTCTGTTGGT	
FAD2A (SEQ ID NO:5)	(719)	TCTACCGCTACGCTGGTGTCCAGGGAGTTGCCCTCTGTTGGT	
FAD2-2 (SEQ ID NO:7)	(719)	TCTACCGCTACGCTGGTGTCCAGGGAGTGGCTCTGTTGGT	
FAD2-1 (SEQ ID NO:6)	(704)	TCTACCGTTACGCTGGTGTCCAGGGAGTACCTCTGTTGGT	
		761	800
FAD2-3 (SEQ ID NO:8)	(759)	CTGCTCTACGGAGTCCCTCTTGTGATTGTCAATGGTTTC	
FAD2A (SEQ ID NO:5)	(759)	CTGCTCTACGGAGTCCCTCTTGTGATTGTCAAAGGGTTTC	
FAD2-2 (SEQ ID NO:7)	(759)	CTGCTCTACGGAGTCCCTCTTGTGATTGTCAAAGGGTTTC	

图 1D

FAD2-1 (SEQ ID NO:6)	(744)	CTGCTCTACGGAGTCCCTCTTGTGATTGTCAAAGGGTTTC	
		801	840
FAD2-3 (SEQ ID NO:8)	(799)	CTGCTGTGATCACTTACTTGCAGCACACGCCACCTTCCC	
FAD2A (SEQ ID NO:5)	(799)	TTAGTTTTGATCACTTACTTGCAGCACACGCCACCTTCCC	
FAD2-2 (SEQ ID NO:7)	(799)	CTGCTGTGATCACTTACTTGCAGCACACGCCACCTTCCC	
FAD2-1 (SEQ ID NO:6)	(784)	CTGCTGTGATCACTTACTTGCAGCACACGCCACCTTCCC	
		841	880
FAD2-3 (SEQ ID NO:8)	(839)	TGCCTCACTACGATCGTCTGAGTGGGATTGGTTGAGAGG	
FAD2A (SEQ ID NO:5)	(839)	TGCCTCACTACGATCGTCTGAGTGGGATTGGTTGAGAGG	
FAD2-2 (SEQ ID NO:7)	(839)	TGCCTCACTACGATCGTCTGAGTGGGATTGGTTGAGAGG	
FAD2-1 (SEQ ID NO:6)	(824)	TGCCTCACTACGATCGTCTGAGTGGGATTGGTTGAGAGG	
		881	920
FAD2-3 (SEQ ID NO:8)	(879)	AGCTTTGGCTACCTTGACAGAGACTACGGAATCTTGAAC	
FAD2A (SEQ ID NO:5)	(879)	AGCTTTGGCTACCTTGACAGAGACTACGGAATCTTGAAC	
FAD2-2 (SEQ ID NO:7)	(879)	AGCTTTGGCTACCTTGACAGAGACTACGGAATCTTGAAC	
FAD2-1 (SEQ ID NO:6)	(864)	AGCTTTGGCTACCTTGACAGAGACTACGGAATCTTGAAC	
		921	960
FAD2-3 (SEQ ID NO:8)	(919)	AAGGTCTTGCACAAATTACCGACACGCACGTGGCGCATC	
FAD2A (SEQ ID NO:5)	(919)	AAGGTCTTGCACAAATTACCGACACGCACGTGGCGCATC	
FAD2-2 (SEQ ID NO:7)	(919)	AAGGTCTTGCACAAATTACCGACACGCACGTGGCGCATC	
FAD2-1 (SEQ ID NO:6)	(904)	AAGGTCTTGCACAAATTACCGACACGCACGTGGCGCATC	
FIG1E		961	1000

图 1D(续)

FAD2-3 (SEQ ID NO:8)	(959)	A CT GTTCCTCCAC AT GCCCGCATTAT AC GCGATGGAAGC	
FAD2A (SEQ ID NO:5)	(959)	ACCTGTTCTC AC CCATGCCCGCATTAT AC GCGATGGAAGC	
FAD2-2 (SEQ ID NO:7)	(959)	A CT GTTCCTC AC ATGCCCGCATTATAACGCGATGGAAGC	
FAD2-1 (SEQ ID NO:6)	(944)	A CT GTTCCTC AC ATGCCCGCATTATAACGCGATGGAAGC	
		1001	1040
FAD2-3 (SEQ ID NO:8)	(999)	AC CGAAGGCGATAAAGCCGATACTG-GGAGAGTATTAC CA	
FAD2A (SEQ ID NO:5)	(999)	AC CGAAGGCGATAAAGCCGATACTG-GGAGAGTATTAC CA	
FAD2-2 (SEQ ID NO:7)	(999)	GAC CA AAGGCGATAAAGCCGATACT CG -GGAGAGTATTACCA	
FAD2-1 (SEQ ID NO:6)	(984)	GAC CA AAGGCGATAAAGCCGATACT CG TGGAGAGTATTACCA	
		1041	1080
FAD2-3 (SEQ ID NO:8)	(1038)	GTTGATGG AC GCCCG CG GGTTAAGGCGATGTGGAGGGAG	

图 1E

FAD2A (SEQ ID NO:5)	(1038)	GTTGATGG AC GCCCG CG GGTTAAGGCGATGTGGAGGGAG	
FAD2-2 (SEQ ID NO:7)	(1038)	GTT CG ATGGAAACGCCCG CG GGTTAAGGCGATGTGGAGGGAG	
FAD2-1 (SEQ ID NO:6)	(1024)	GTT CG ATGGAAACGCCCG CG GGTTAAGGCGATGTGGAGGGAG	
		1081	1120
FAD2-3 (SEQ ID NO:8)	(1078)	GCGAAGGAGTGTATCTATGTGGAACCGGACAGGCCAAGGTG	
FAD2A (SEQ ID NO:5)	(1078)	GCGAAGGAGTGTATCTATGTGGAACCGGACAGGCCAAGGTG	
FAD2-2 (SEQ ID NO:7)	(1078)	GCGAAGGAGTGTATCTATGT CG AACCGGA AG GCCAAGGTG	
FAD2-1 (SEQ ID NO:6)	(1064)	GCGAAGGAGTGTATCTATGT CG AACCGGA AG GCCAAGGTG	
		1121	1160
FAD2-3 (SEQ ID NO:8)	(1118)	AGAAGAAAGGTGTGTCTGG-----	
FAD2A (SEQ ID NO:5)	(1118)	AGAAGAAAGGTGTGTCTGG TACAGCAATAGTTA CTTG	
FAD2-2 (SEQ ID NO:7)	(1118)	AGAAGAAAGGTGTGTCTGG TACAGCAATAGTTA AGAG	
FAD2-1 (SEQ ID NO:6)	(1104)	AGAAGAAAGGTGTGTCTGG TACAGCAATAGTTA -----	
		1161	
FAD2-3 (SEQ ID NO:8)	(1138)	-----	
FAD2A (SEQ ID NO:5)	(1158)	CTAA	
FAD2-2 (SEQ ID NO:7)	(1158)	ATCA	
FAD2-1 (SEQ ID NO:6)	(1135)	-----	

图 1E(续)

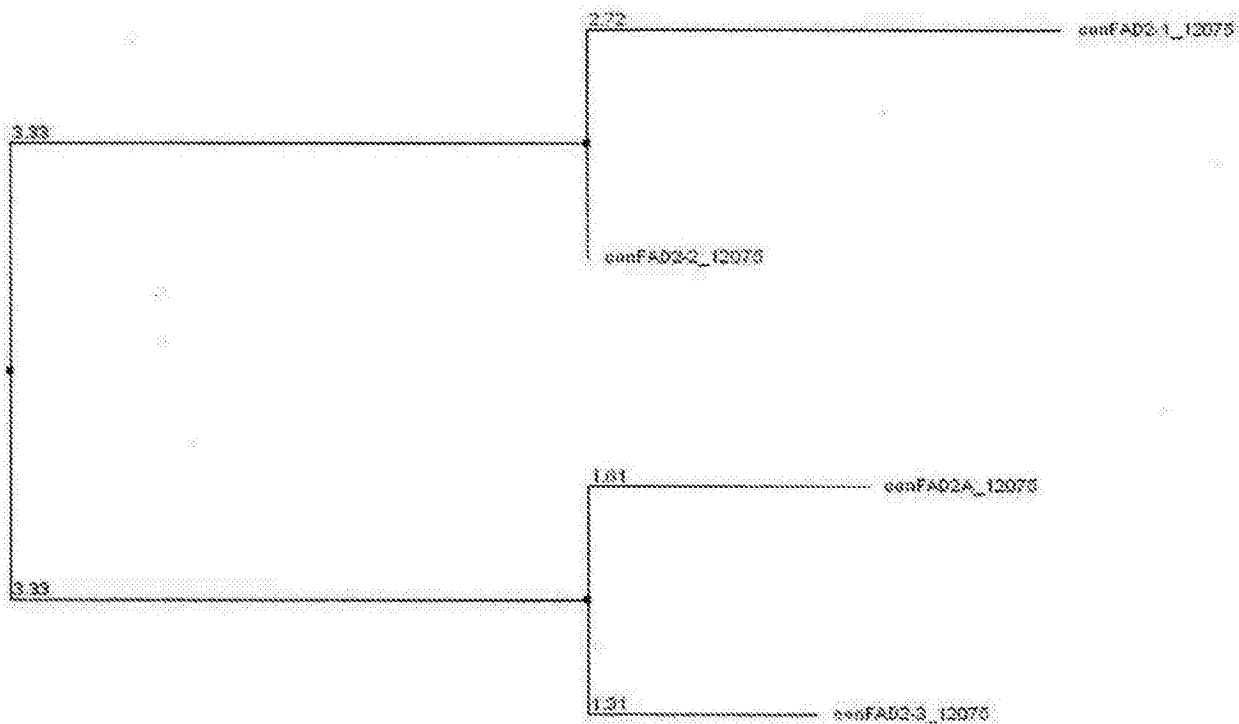


图 2

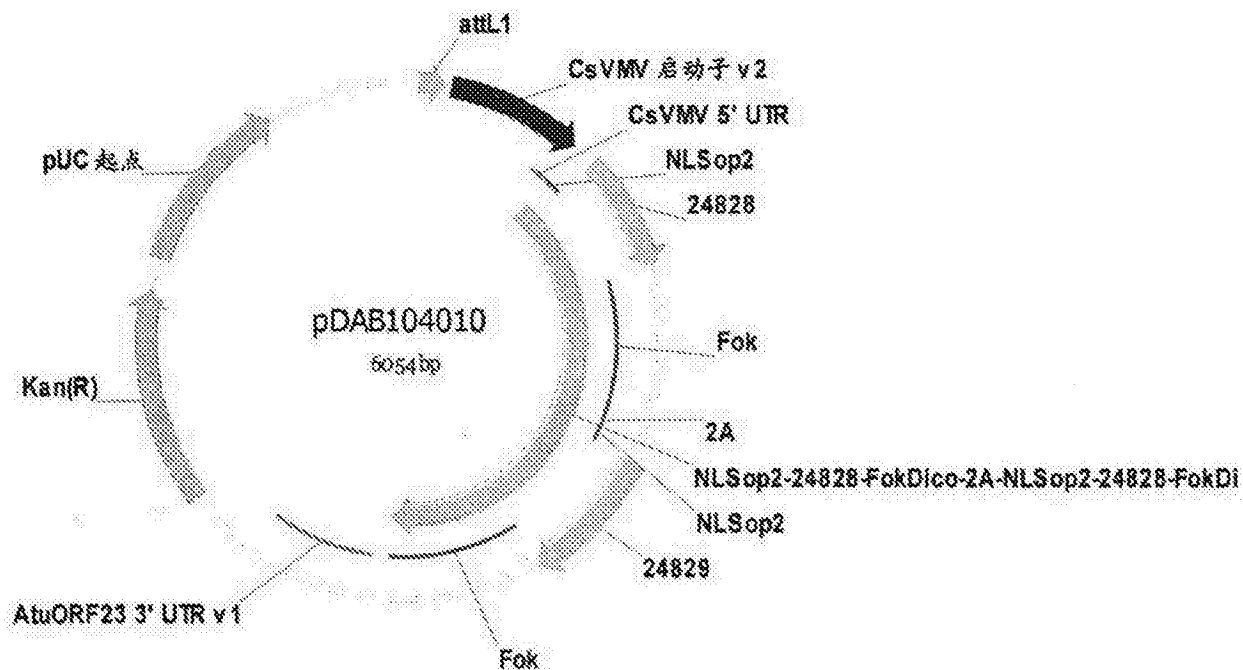


图 3

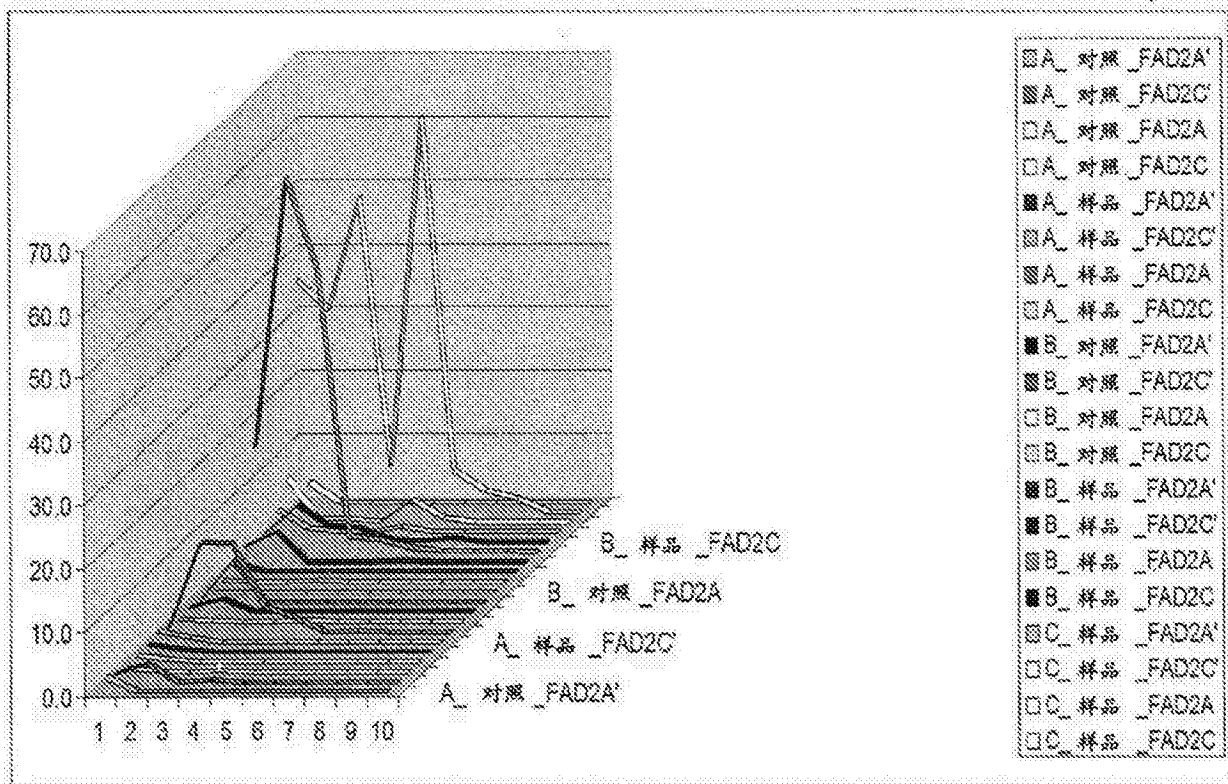


图 4

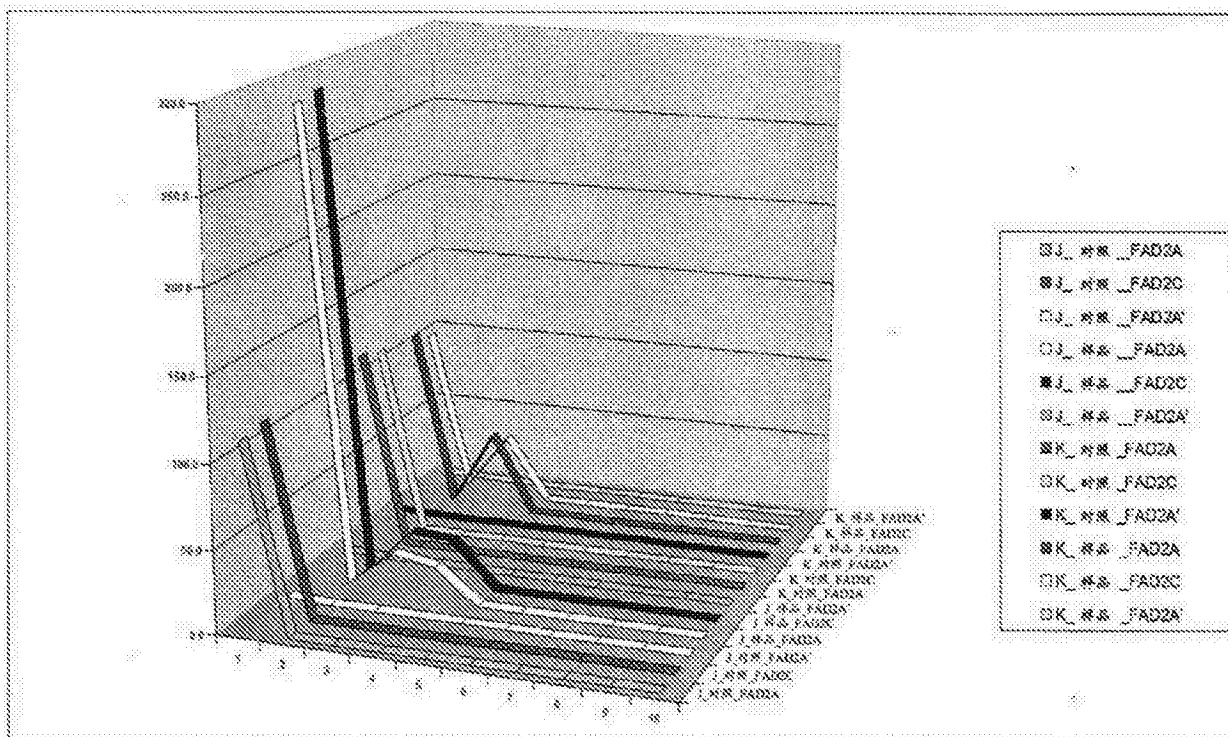


图 5A

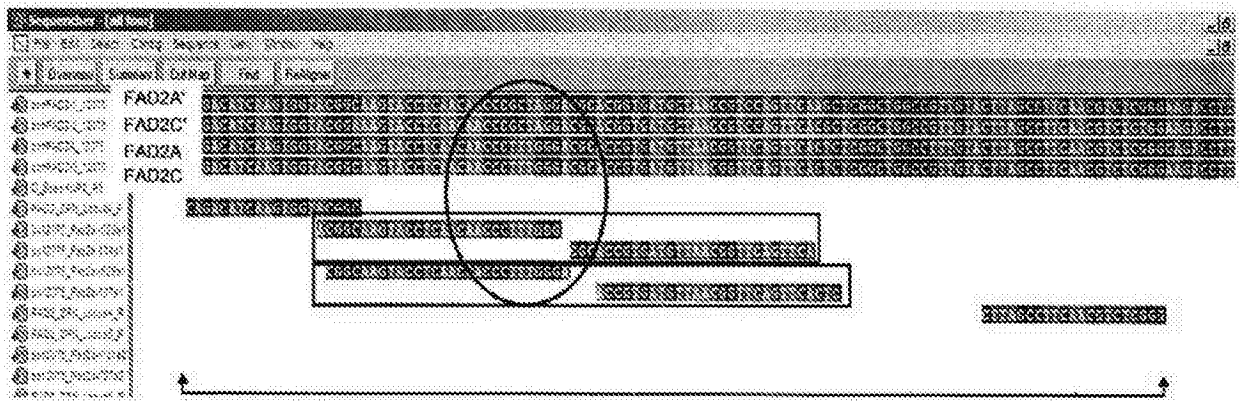


图 5B

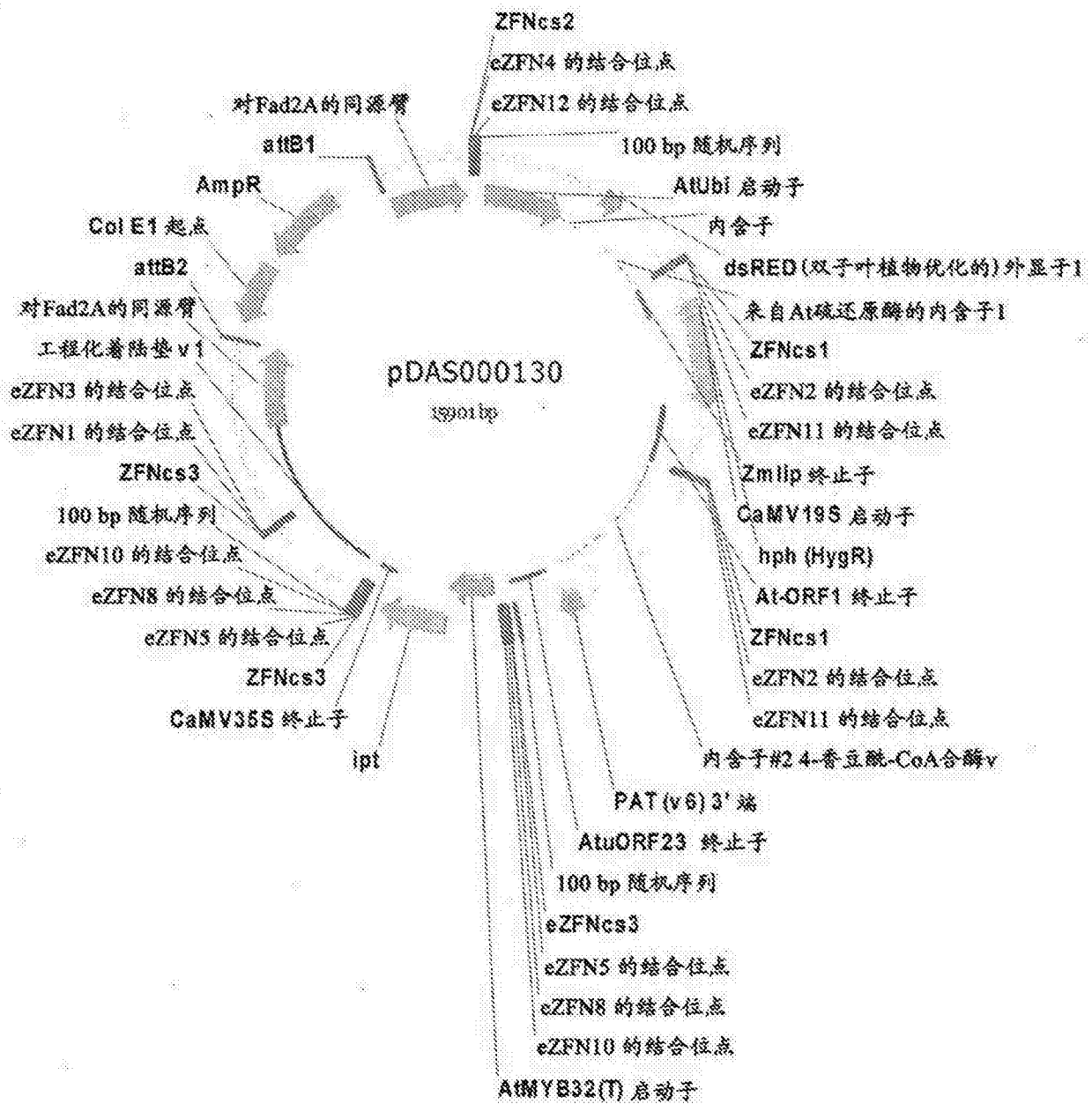


图 6

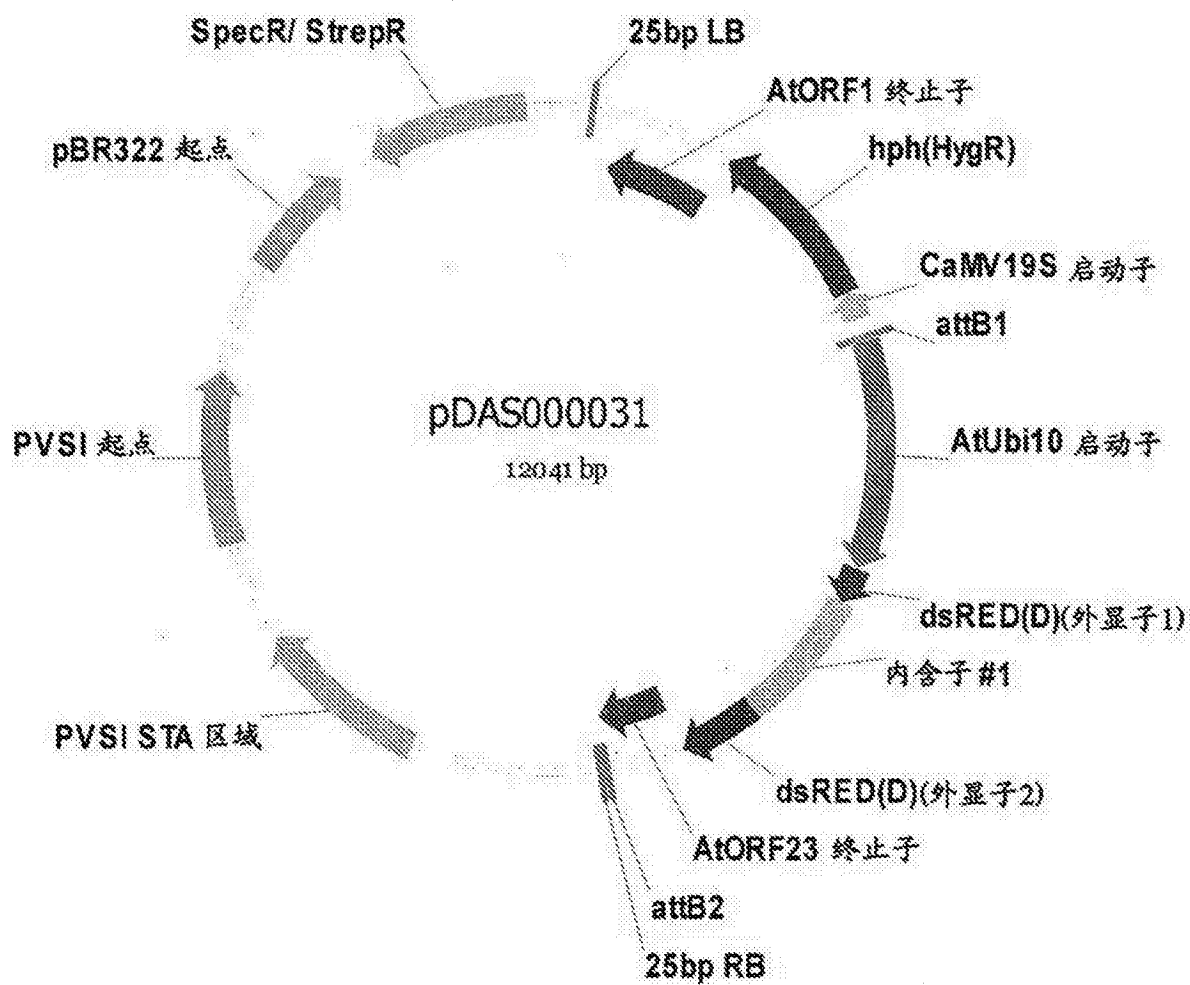


图 7

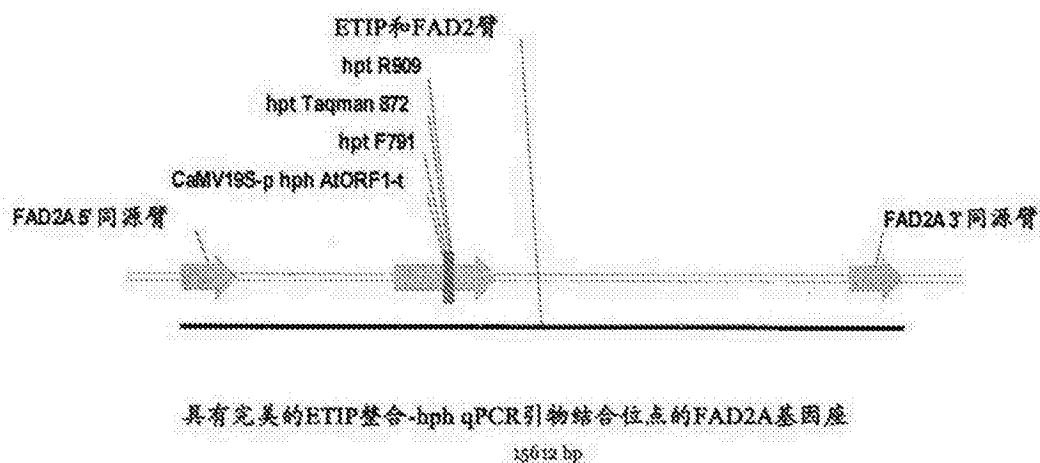


图 8

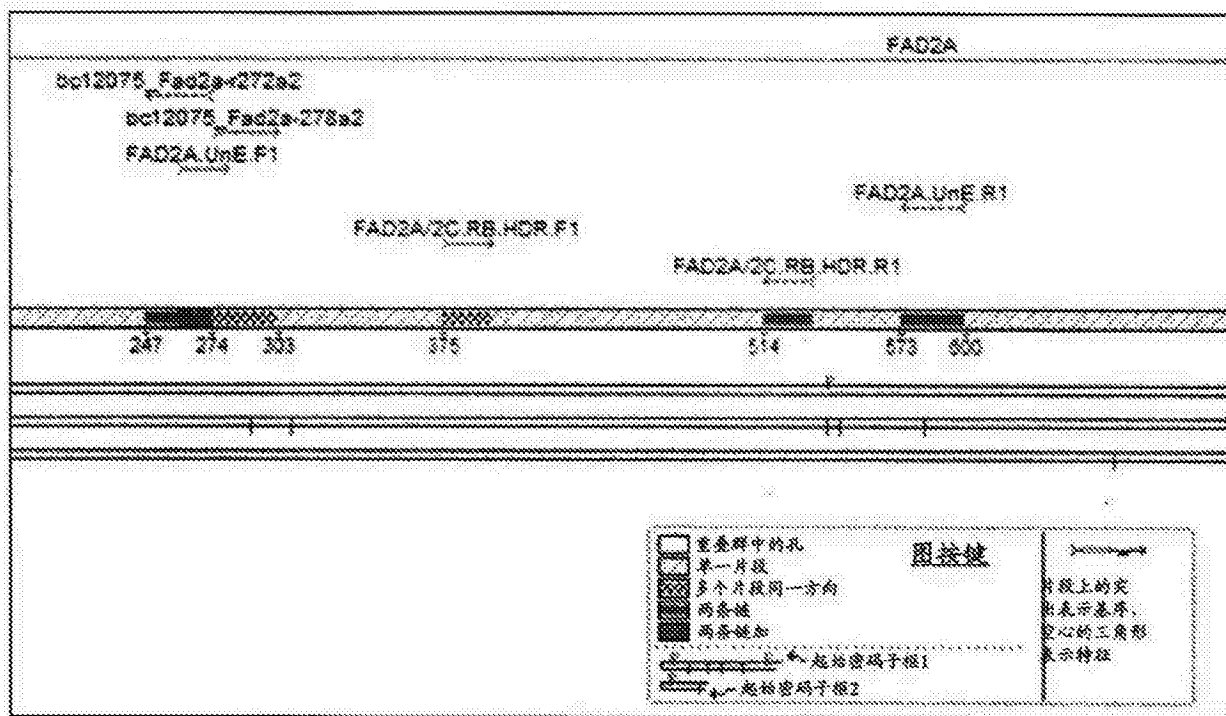


图 9

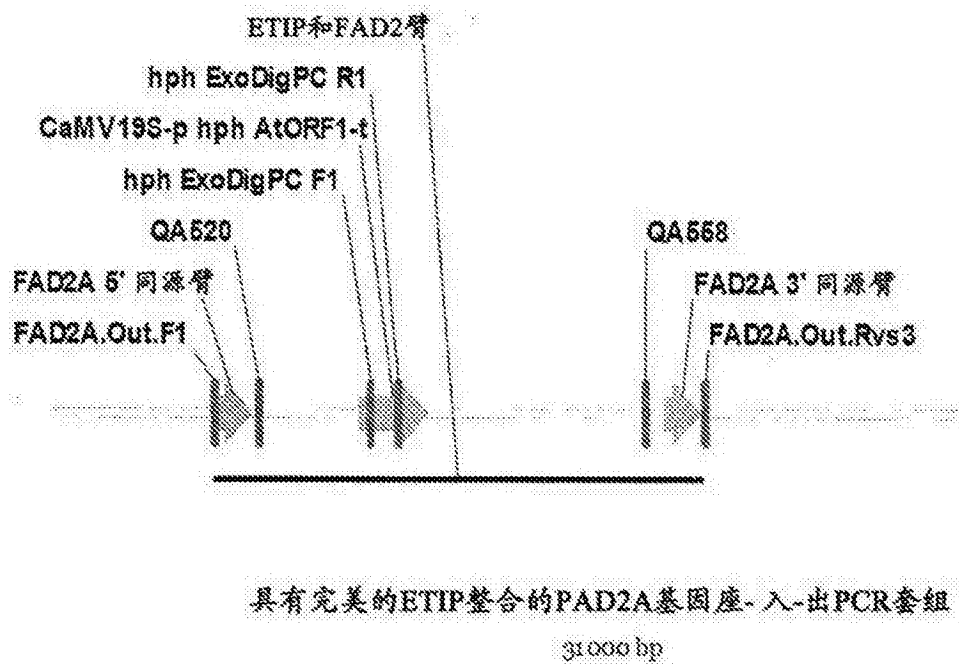


图 10

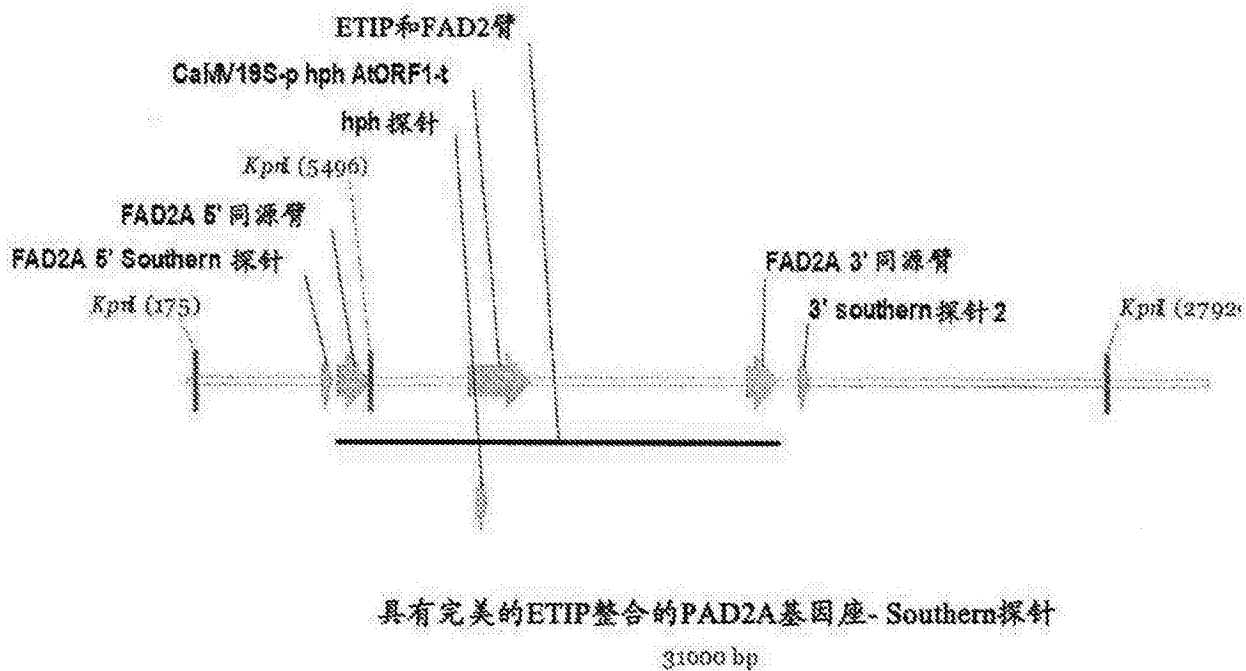


图 11

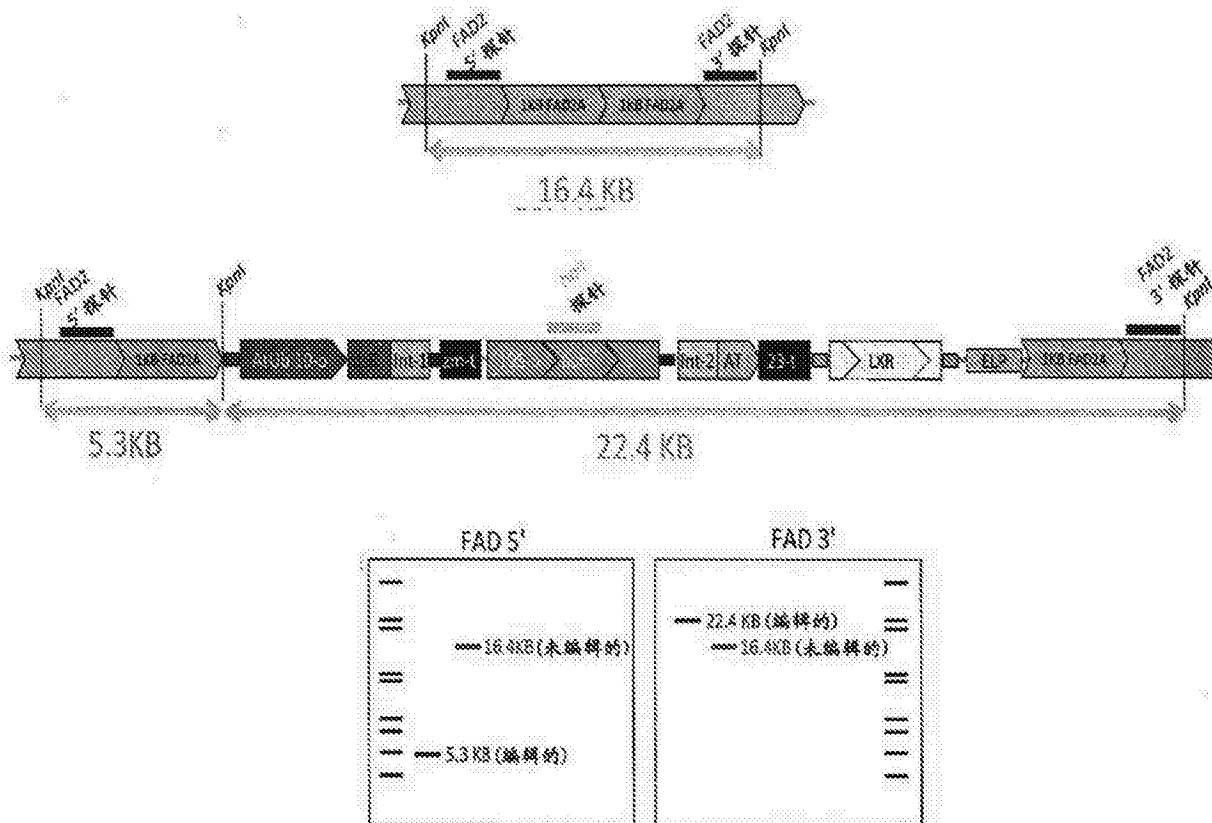


图 12

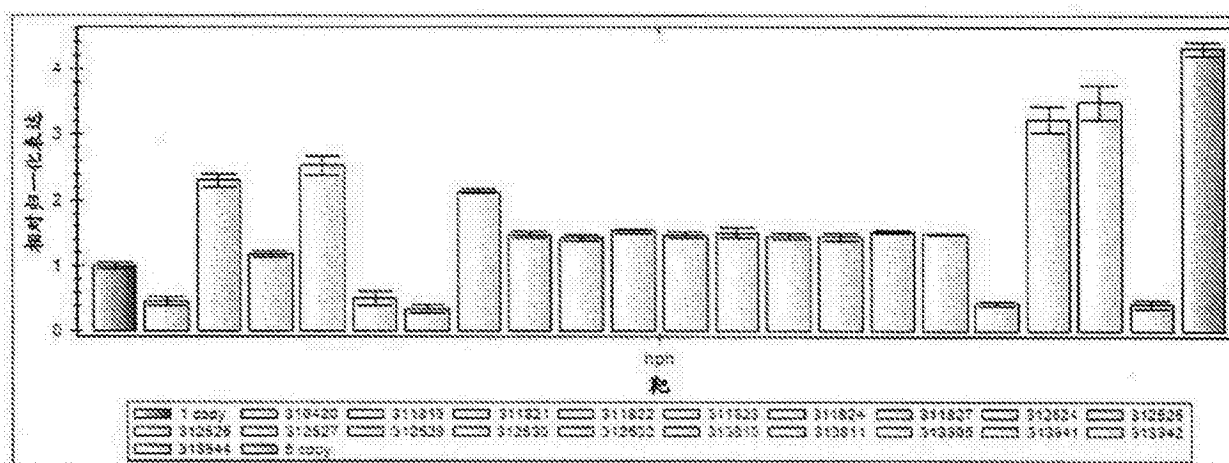


图 13

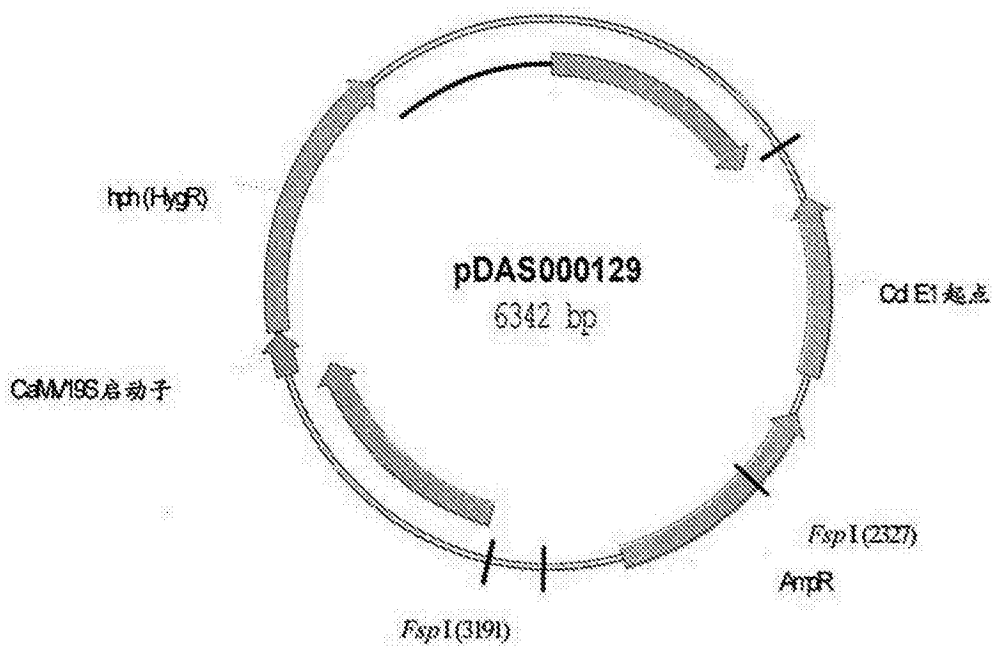


图 14

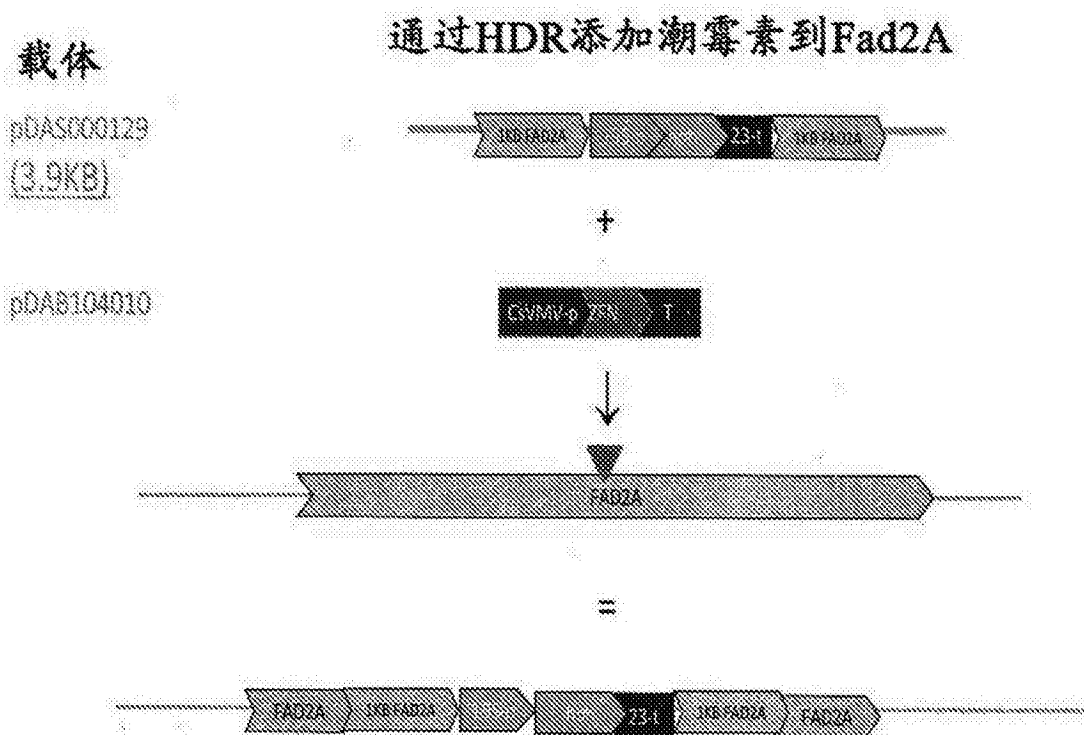


图 15

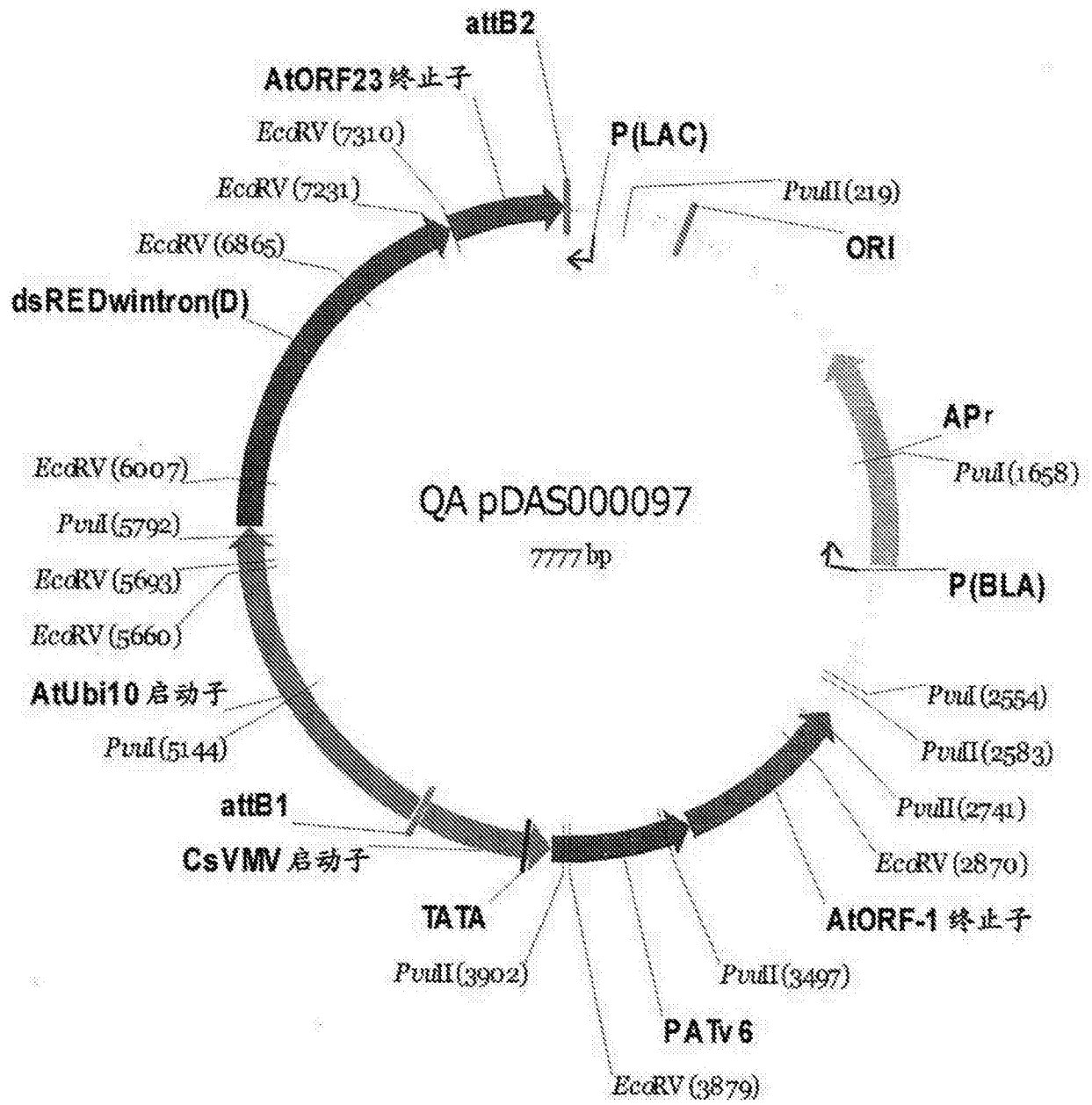


图 16

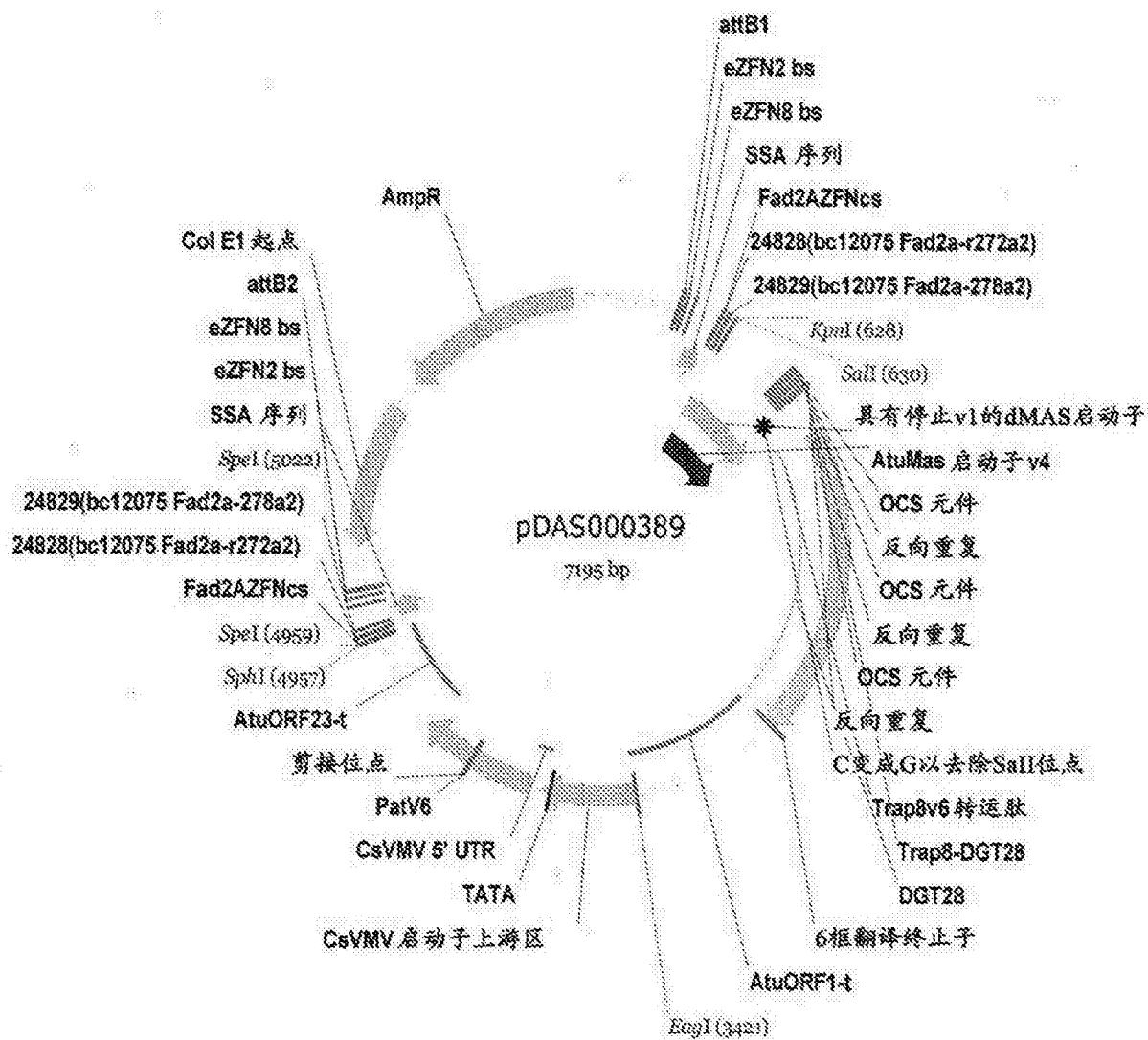


图 17

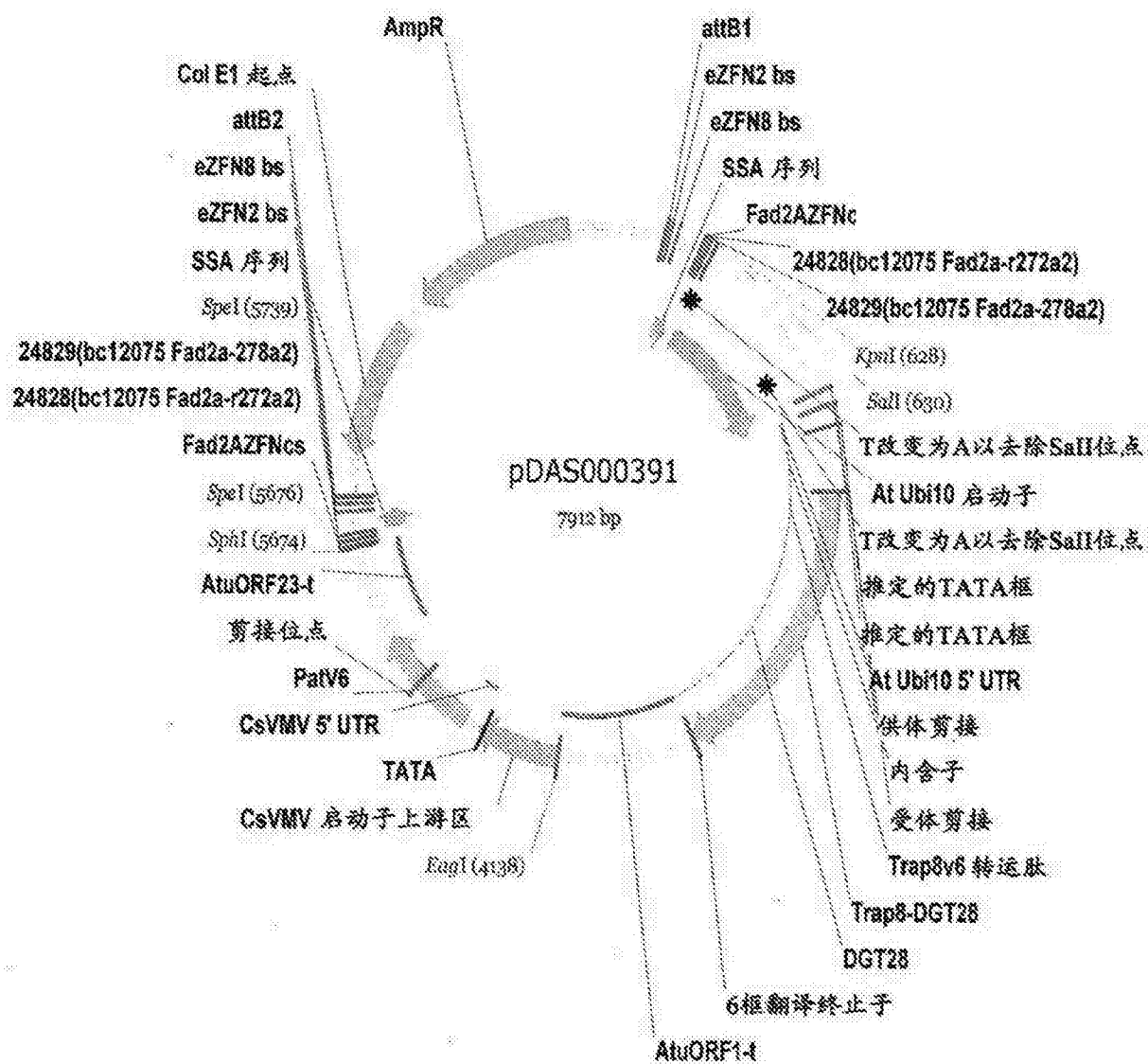


图 18

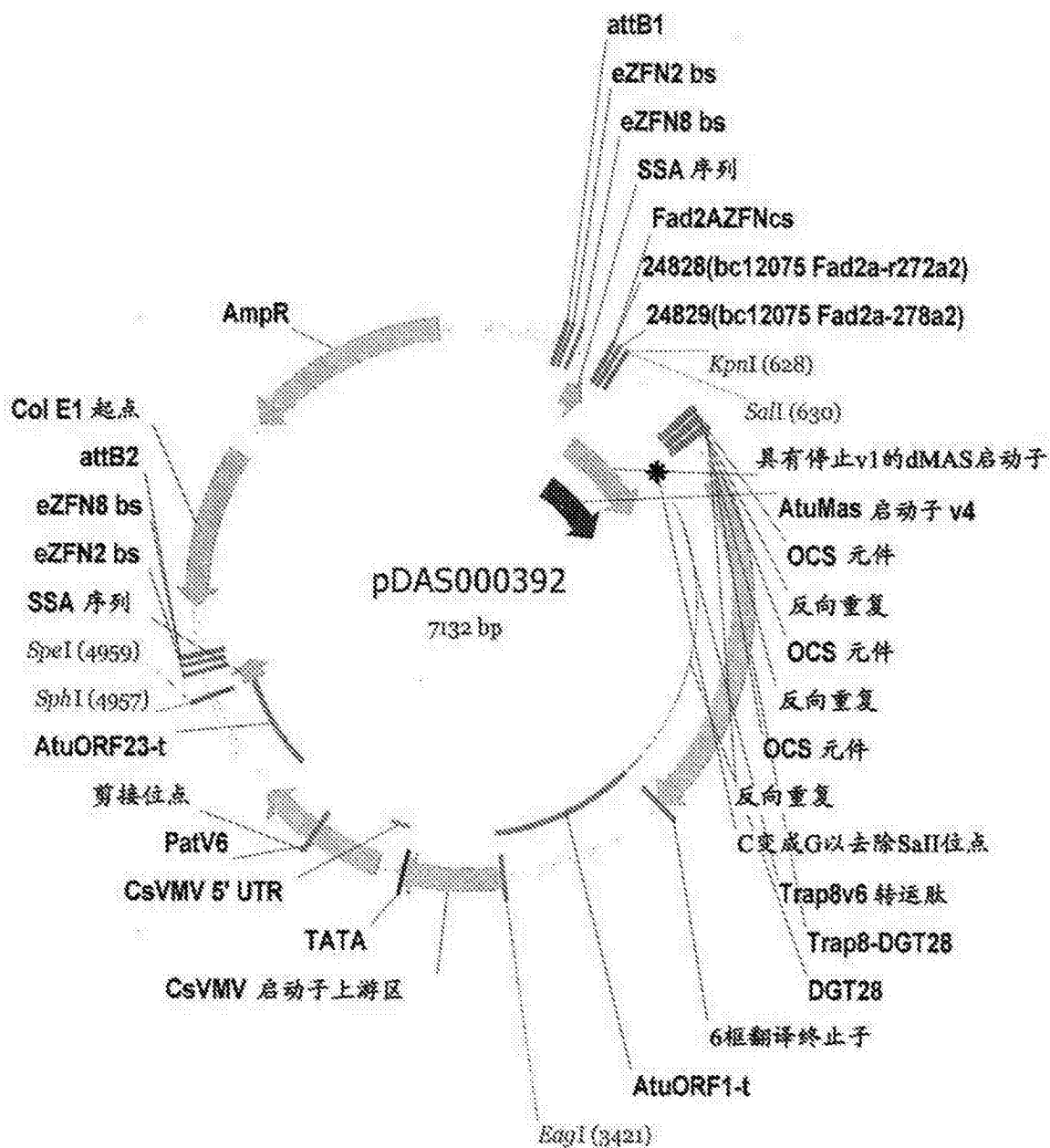


图 19

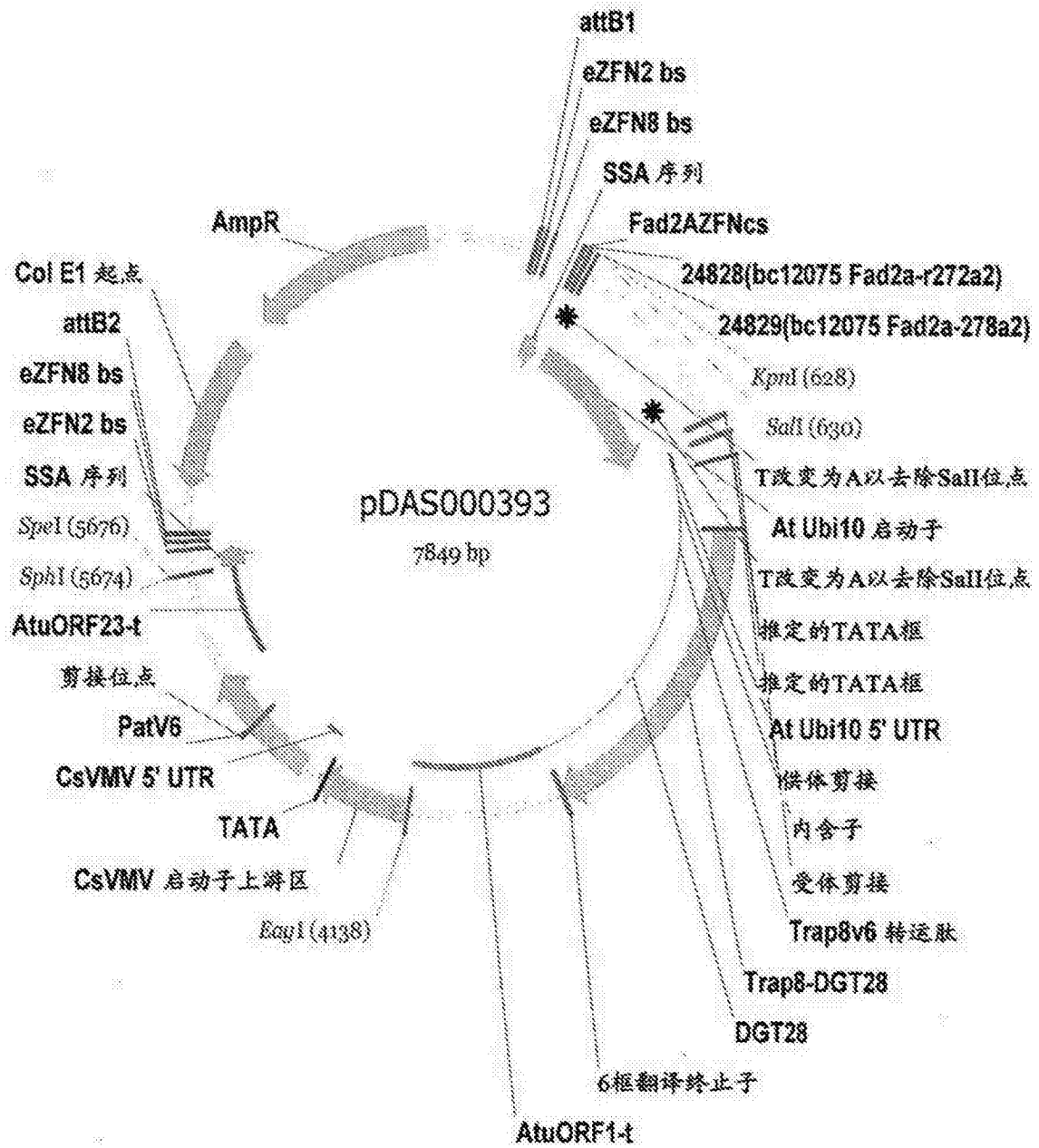


图 20

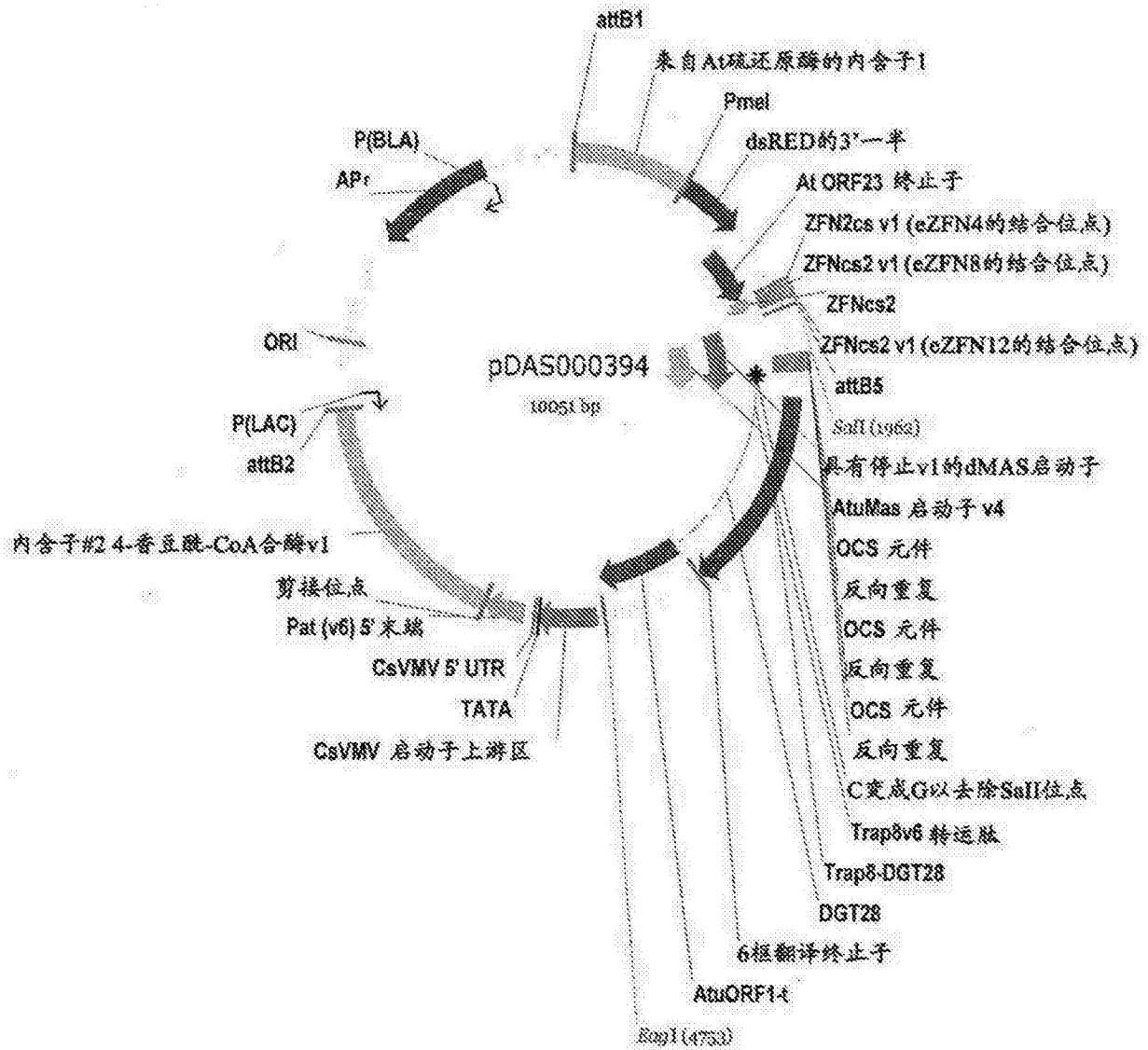


图 21

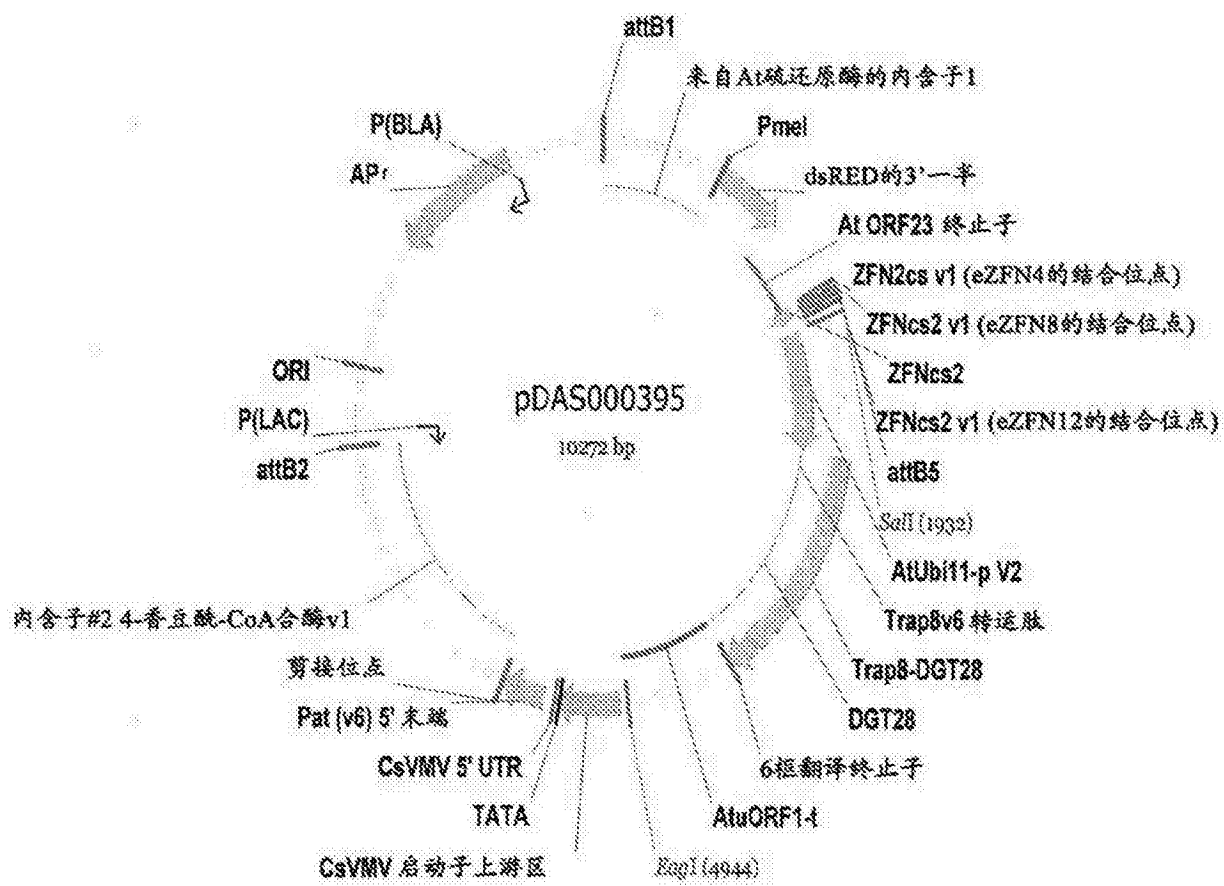


图 22

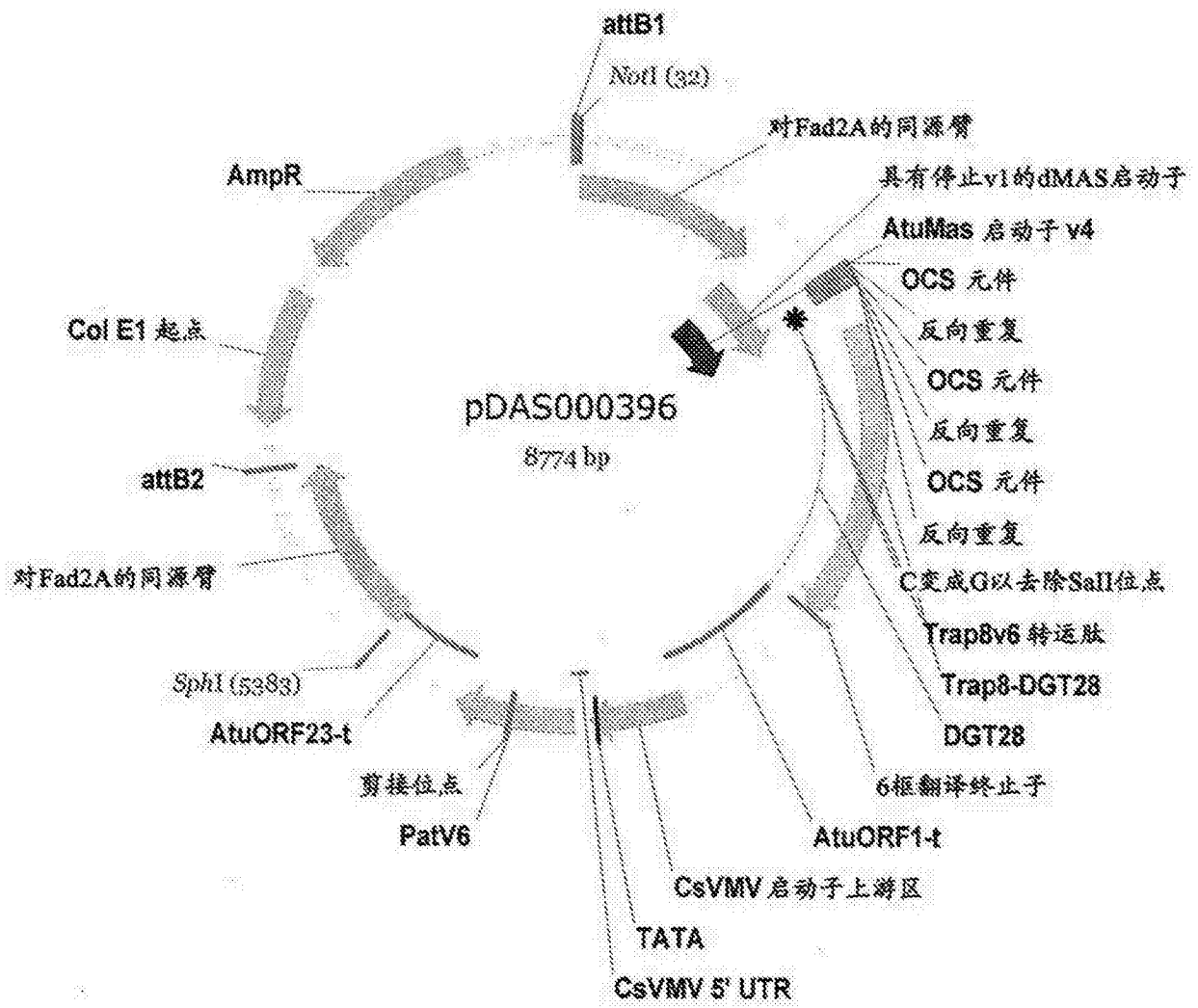


图 23

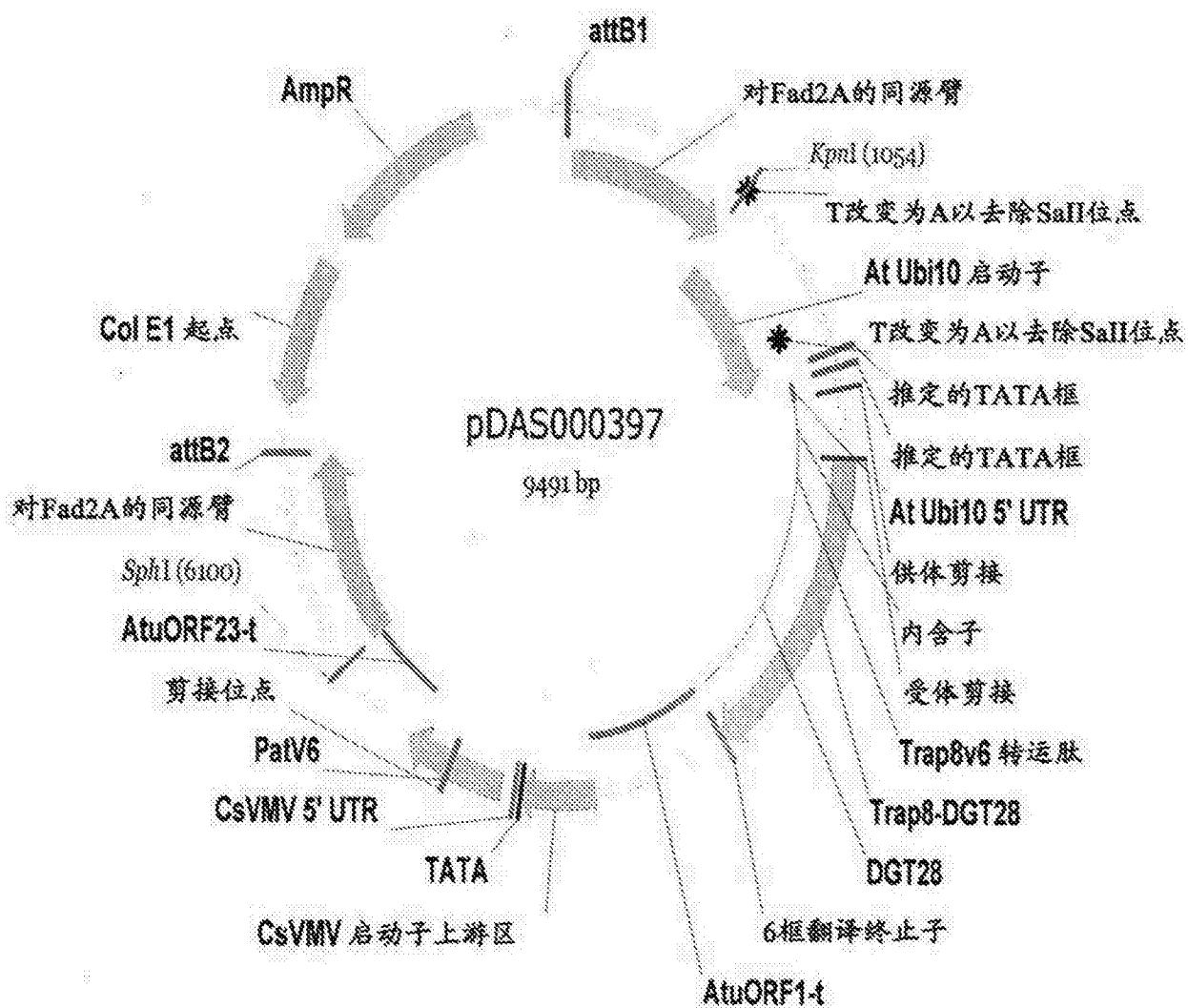


图 24