



(12) 发明专利

(10) 授权公告号 CN 108610408 B

(45) 授权公告日 2023.04.07

(21) 申请号 201810150784.4

(74) 专利代理机构 北京坤瑞律师事务所 11494

(22) 申请日 2013.09.05

专利代理人 罗天乐

(65) 同一申请的已公布的文献号

申请公布号 CN 108610408 A

(51) Int.CI.

C07K 14/47 (2006.01)

(43) 申请公布日 2018.10.02

C07K 19/00 (2006.01)

(30) 优先权数据

61/697,886 2012.09.07 US

C12N 9/22 (2006.01)

(62) 分案原申请数据

201380058037.3 2013.09.05

C12N 15/12 (2006.01)

(73) 专利权人 美国陶氏益农公司

地址 美国印第安纳州

C12N 15/62 (2006.01)

专利权人 桑格摩生物科学股份有限公司

C12N 15/10 (2006.01)

(72) 发明人 W·M·安利 S·R·韦布

C12N 15/82 (2006.01)

P·塞缪尔 D·Y·古钦

C12N 5/10 (2006.01)

J·C·米勒 L·张

(56) 对比文件

彭瑞云等.锌指核酸酶技术.《现代实验病理技术》.军事医学科学出版社,2012, (第1版), 256.

审查员 张英朔

权利要求书2页 说明书47页

序列表28页 附图9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

本申请公开了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。具体公开了在FAD2基因座中进行基因破坏、基因编辑或基因堆叠的方法和组合物,其是通过以定点的方式剪切大豆细胞FAD2基因中的一个位置,从而在FAD2基因中产生断裂(break),然后任选地将该断裂中整合入感兴趣的核酸分子。

1. 一种适用于修饰大豆细胞中的FAD2基因的锌指核酸酶,其中该锌指核酸酶在选自SEQ ID NO:14至SEQ ID NO:20的核酸靶位点处或所述核酸靶位点附近切割,其中所述FAD2基因是如SEQ ID NO:4所表示的FAD2.3基因、如SEQ ID NO:9所表示的FAD2.6基因、或前述二者,所述锌指核酸酶包含锌指核酸酶对,该对中的每个锌指核酸酶包含核酸酶结构域和位点特异性锌指蛋白,所述位点特异性锌指蛋白在SEQ ID NO:14-SEQ ID NO:20内结合,其中第一和第二锌指蛋白各自包括5个或6个锌指结构域,依次为F1至F5或F1至F6,每个锌指结构域包括识别螺旋区,其中该锌指蛋白包括在下表的单一行中排序并显示的识别螺旋区:

ZFP#	靶位点	F1	F2	F3	F4	F5	F6
37354	SEQ ID NO: 14	SEQ ID NO:24 QSSDLSR	SEQ ID NO:25 RKDALVA	SEQ ID NO:26 RSADLTR	SEQ ID NO:27 RSDDLTR	SEQ ID NO:28 RSDAMSQ	SEQ ID NO:29 RNASRTR
37355	SEQ ID NO: 14	SEQ ID NO:30 DRSNLSR	SEQ ID NO:31 HKWLRNQ	SEQ ID NO:32 DSSDRKK	SEQ ID NO:33 LRHHLTR	SEQ ID NO:34 QSGTRKT	无
37370	SEQ ID NO: 15	SEQ ID NO:35 QNAHRKT	SEQ ID NO:33 LRHHLTR	SEQ ID NO:36 QSGDLTR	SEQ ID NO:37 QTSTLSK	SEQ ID NO:38 TSGSLSR	SEQ ID NO:39 RSDHLTQ
37371	SEQ ID NO: 15	SEQ ID NO:40 RSDVLSE	SEQ ID NO:41 RSAADLSR	SEQ ID NO:24 QSSDLSR	SEQ ID NO:42 RTDALRG	SEQ ID NO:33 LRHHLTR	SEQ ID NO:43 HRSARKR
37374	SEQ ID NO: 16	SEQ ID NO:44 DRSHLTR	SEQ ID NO:45 QSGNLHV	SEQ ID NO:46 RSDHLSA	SEQ ID NO:47 RSNLLVA	SEQ ID NO:48 QSGALAR	SEQ ID NO:49 DRSLALAR
37375	SEQ ID NO: 16	SEQ ID NO:50 QSSNLAR	SEQ ID NO:51 QSSDLRR	SEQ ID NO:52 RSDTLSE	SEQ ID NO:53 QSGHLSR	SEQ ID NO:54 RSDVLST	SEQ ID NO:55 QNAHRIK
37366	SEQ ID NO: 17	SEQ ID NO:56 RSDNLSQ	SEQ ID NO:57 ASNDRKK	SEQ ID NO:58 RSDNLST	SEQ ID NO:59 MRQHLLN	SEQ ID NO:60 RSDNLAR	SEQ ID NO:61 QKKDRSY
37367	SEQ ID NO: 17	SEQ ID NO:62 RSDHLSR	SEQ ID NO:63 DRSNRKT	SEQ ID NO:64 RSDTLSA	SEQ ID NO:65 DKSTRTK	SEQ ID NO:36 QSGDLTR	SEQ ID NO:66 TSGSLTR
37384	SEQ ID NO: 18	SEQ ID NO:24 QSSDLSR	SEQ ID NO:25 RKDALVA	SEQ ID NO:26 RSAADLTR	SEQ ID NO:27 RSDDLTR	SEQ ID NO:67 RSDSLSA	SEQ ID NO:68 RSDALAR
37385	SEQ ID NO: 18	SEQ ID NO:30 DRSNLSR	SEQ ID NO:31 HKWLRNQ	SEQ ID NO:32 DSSDRKK	SEQ ID NO:33 LRHHLTR	SEQ ID NO:69 RRDILHQ	无
37398	SEQ ID NO: 19	SEQ ID NO:56 RSDNLSQ	SEQ ID NO:57 ASNDRKK	SEQ ID NO:58 RSDNLST	SEQ ID NO:59 MRQHLLN	SEQ ID NO:36 QSGDLTR	SEQ ID NO:70 QRTHLKA
37399	SEQ ID NO: 19	SEQ ID NO:62	SEQ ID NO:63	SEQ ID NO:40	SEQ ID NO:71	SEQ ID NO:36	SEQ ID NO:66

		RSDHLSR	DRSNRKT	RSDVLSE	ARSTRTN	QSGDLTR	TSGSLTR
37392	SEQ ID NO: 20	SEQ ID NO:72 QSGNLAR	SEQ ID NO:73 WRISLAA	SEQ ID NO:30 DRDSNLSR	SEQ ID NO:74 WKESLGA	SEQ ID NO:75 HRKSLSR	无
37393	SEQ ID NO: 20	SEQ ID NO:24 QSSDLSR	SEQ ID NO:76 YHWYLKK	SEQ ID NO:77 TSGHLSR	SEQ ID NO:78 TSGNLTR	SEQ ID NO:79 WWTSRAL	无

2. 如权利要求1所述的锌指核酸酶,其中所述锌指核酸酶对包括:

- (i) ZFP#37354和ZFP#37355;
- (ii) ZFP#37370和ZFP#37371;
- (iii) ZFP#37374和ZFP#37375;

- (iv) ZFP#37366和ZFP#37367;
- (v) ZFP#37384和ZFP#37385;
- (vi) ZFP#37398和ZFP#37399;或
- (vii) ZFP#37392和ZFP#37393。

3. 一种多核苷酸,其编码如权利要求1或2所述的锌指核酸酶。

4. 一种在大豆细胞中切割FAD基因的方法,该方法包括:

将如权利要求1-2中任一项所述的锌指核酸酶或如权利要求3所述的多核苷酸导入大豆细胞,从而切割所述FAD2基因。

5. 如权利要求4所述的方法,进一步包括将感兴趣的外源核酸序列整合到经切割的FAD2基因中。

6. 如权利要求5所述的方法,其中所述感兴趣的外源核酸序列是非编码序列。

7. 如权利要求5所述的方法,其中所述外源序列选自下组: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,一个或多个可选择标志物基因,及其组合。

8. 如权利要求4-6中任一项所述的方法,其中所述FAD2基因是FAD22.3、FAD2 2.6基因、或前述二者。

9. 如权利要求4-7中任一项所述的方法,其中位点特异性方式的切割对于FAD2 2.3和FAD2 2.6的一些但非全部拷贝是特异性的。

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] 本申请是2013年9月5日提交的申请号为201380058037.3 (PCT申请号为PCT/US2013/058299)、发明名称为“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”的发明专利申请的分案申请。

[0002] 相关申请交叉引用

[0003] 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。

[0004] 公开领域

[0005] 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。

[0006] 背景

[0007] 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。

[0008] 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。

[0009] 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。

[0010] 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。

[0011] 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。

[0012] 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。

[0013] 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/tracr RNA(‘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。

[0014] 然而,仍然需要有用于修饰和/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。

[0015] 公开概述

[0016] 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,本公开描述了大豆(Glycine max)(例如大豆栽培种Jack,Williams 82,X5,Westag,和Maverick)中的FAD2 2.3和FAD2 2.6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。

[0017] 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FAD2 2.3和FAD2 2.6基因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。

[0018] 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1所示。示例FAD2-结合性锌指的识别螺旋区如表2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)。

[0019] 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。在某些特定的实施方案中,双子叶植物可以是大豆(Glycine max)植物。

[0020] 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。

[0021] 具体地,本发明涉及以下各项:

[0022] 1.一种修饰大豆细胞中的FAD2基因的方法,该方法包括,该方法包括:

[0023] 以位点特异性方式剪切大豆细胞中的FAD2基因中的靶位点,从而在该FAD2基因中产生断裂;

[0024] 其中该FAD2基因在剪切之后被修饰。

[0025] 2.项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。

[0026] 3.根据项1或项2的方法,其中所述FAD2基因是FAD2 2.3、FAD2 2.6基因,或其二者。

[0027] 4.根据项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。

[0028] 5.根据项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。

[0029] 6.根据项4或项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。

[0030] 7.根据项4-6中任一项的方法,其中该融合蛋白是锌指核酸酶。

[0031] 8.根据项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2的单一行中排序并显示的识别螺旋区。

[0032] 9.根据项1-8中任一项的方法,其中位点特异性方式的剪切是对FAD2.3和FAD2.6的一些但非所有拷贝特异性的。

[0033] 10.根据项1-9中任一项的方法,其中该靶位点选自下组:SEQ ID NO:14至SEQ ID NO:20。

[0034] 11.根据项2-10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。

[0035] 12.一种大豆细胞或大豆植物,其包括根据项1-11中任一项的方法修饰的大豆细胞。

[0036] 13.根据项12的细胞或植物,其中该细胞或植物是转基因的细胞或植物,其包含整合到FAD2 2.3和/或FAD2 2.6基因的一个或多个拷贝中的感兴趣的核苷酸序列。

[0037] 14.项13的细胞或植物,其中所述核苷酸序列对所述细胞而言是异源或同源的。

[0038] 15.项14的细胞或植物,其中该同源序列包含至少一个单核苷酸多态性。

[0039] 16.根据项13-15中任一项的细胞或植物,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:14至SEQ ID NO:20。

[0040] 17. 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:14至SEQ ID NO:20。

[0041] 18. 项17的锌指核酸酶, 其中该锌指核酸酶包含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2的单一行中排序并显示的识别螺旋区。

[0042] 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。

[0043] 附图简述

[0044] 图1显示了来自Williams 82 (SEQ ID NO:4) , Westag (SEQ ID NO:5) , X5 (SEQ ID NO:6) , Jack (SEQ ID NO:7) , 和Maverick (SEQ ID NO:8) 的FAD22.3编码序列的比对。

[0045] 图2显示了来自Williams 82 (SEQ ID NO:9) , Westag (SEQ ID NO:10) , X5 (SEQ ID NO:11) , Jack (SEQ ID NO:12) , 和Maverick (SEQ ID NO:13) 的FAD2 2.6编码序列的比对。

[0046] 图3显示了FAD2 2.3与2.6基因设计的ZFN在DLSSA测定中的活性。对设计针对FAD2 2.3和2.6基因座的ZFN评估了针对克隆到哺乳动物细胞中作为报告物的FAD2 2.3与2.6序列的剪切活性。

[0047] 图4显示pDAB115620的质粒图。

[0048] 图5显示pDAB115622的质粒图。

[0049] 图6显示pDAB7221的质粒图。

[0050] 图7是描绘用于基因座破坏测定 (locus disruption assay) 的探针/引物的示意图。示出了用于FAD2 2.3和2.6基因的F2ZFN结合位点, 以及用于破坏测定的引物。

[0051] 图8显示了利用F2ZFN2锌指核酸酶在FAD2 2.3基因座中NHEJ靶定供体序列而导致的入-出PCR (In-Out PCR) 产物的序列。参考序列 (图顶部) 代表供体载体以反向取向靶定插入的构型。由Fok I 消化导致的DNA单链末端被填补 (fill in) 以制作参考序列。显示了Sanger序列。F2ZFN2ZFN结合序列加有下划线。具有与列出的序列相似的序列的质粒克隆列在右边。

[0052] 序列

[0053] 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。

[0054] 详细说明

[0055] 1. 若干实施方案的概述

[0056] 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. (2003) Nature Biotechnol. 31:686-680; Le et al. (2013) Nature Biotechnol. 31:688-691; Nekrasov et al. (2013) Nature Biotechnol. 31:

691-693,Ainley 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插入的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:对该基因座处插入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。

[0057] 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基因是靶向插入外源核酸(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。

[0058] 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,例如但不仅限于,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/tracr RNA(“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FAD3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。

[0059] 本公开举例说明了FAD2基因座作为性能基因座的用处,例如,通过描述大豆(Glycine max)中的FAD2 2.3或FAD2 2.6基因座,和可用于将外源核酸整合在FAD2 2.3和/或FAD2 2.6基因座处的相应的FAD2特异性ZFN。

[0060] 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插入方法在转基因堆叠过程中可能特别有益。

[0061] 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)。

[0062] II. 术语

[0063] 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。

[0064] 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。

[0065] 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:

[0066] 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。

[0067] 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)。

[0068] 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的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。

[0069] 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,例如但不仅限于,被破坏的或经过修饰的等位基因;转基因;PTU;和ELP。

[0070] 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。

[0071] 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,例如但不仅限于,由可遗传的基因组DNA序列编码的cDNA和/或mRNA。

[0072] 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。

[0073] 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,例如但不仅限于: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,甲基磷酸酯,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(pendant moieties),例如,肽;插层剂,例如,吖啶和补骨脂;螯合剂;烷化剂;和经过修饰的连接,例如, α -异头核酸)。

[0074] 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。

[0075] 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。

[0076] 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。

[0077] 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。实例包括但不仅限于: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)。

[0078] **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的任何方法将核酸纳入到细胞内。该术语包括核酸引入方法,包括例如但不仅限于,转染;转化;和转导。

[0079] **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(例如除草剂抗性或害虫抗性)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基因座处的感兴趣的核酸分子是转基因。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)。

[0080] 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。

[0081] **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。

[0082] **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)。

[0083] **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;和多腺昔酸化信号(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。

[0084] 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。

[0085] 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。

[0086] 美国国家生物技术信息中心(NCBI)的基本局部比对搜索工具(BLASTTM;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BLASTTM的“帮助”一节中获得。对于核酸序列的比较,可以采用BLASTTM(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。

[0087] 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。

[0088] 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。

[0089] 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。

[0090] 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。

[0091] 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,包括例如但不仅限于,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³²P标记。其他可以使用的标记物包括,例如但不仅限于,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。

[0092] 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克隆节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。

[0093] 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。这在本文中被称作“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。

[0094] 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。

[0095] 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。

[0096] 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。

[0097] 在特定的实施方案中,严格条件是在如下条件下杂交:65°C于6x盐-柠檬酸钠(SSC)缓冲液,5x Denhardt溶液,0.5% SDS和100μg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°C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。

[0098] 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。

[0099] 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。

[0100] 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。

[0101] 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。

[0102] 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。

[0103] 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。

[0104] 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。

[0105] 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。

[0106] 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。

[0107] 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”)。

[0108] 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,例如但不仅限于,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,例如但不仅限于,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。

[0109] “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。

[0110] 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 F_3 - F_5 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。

[0111] “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。

[0112] “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。

[0113] “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。

[0114] 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/组成主要来自合理的设计标准。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。

[0115] “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;US 6,200,759;WO 95/19431;WO 96/06166;WO 98/53057;WO 98/54311;WO 00/27878;WO 01/60970;WO 01/88197;WO 02/099084和美国公开号No.20110301073。

[0116] “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,包括但不仅限于,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。

[0117] “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。

[0118] “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。

[0119] 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:Mme1,大肠杆菌素-E7(CEA7_ECOLX),大肠杆菌素-E9,APFL,Endo I(END1_ECOLI),人Endo G(NUCG_HUMAN),牛Endo G(NUCG_BOVIN),R.HinP11,1-Bas1,1-Bmol,1-Hmul,1-Tev1,1-Tev11,1-Tev111,1-Two1,R.Msp1,R.Mva1,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.BspD61(R.BspD61大亚基),ss.BspD61(R.BspD61小亚基),R.PIel,Mly1,A1wl,Mva12691,Bsrl,Bsml,Nb.BtsCI,Nt.BtsCI,R1.Bts1,R2.Bts1,BbvCI亚基1,BbvCI亚基2,Bpu10Ia亚基,

Bpu10I β 亚基, Bmr1, Bfil, l-Crel, hExol (EX01JHUMAN), 酵母Exol (EX01_YEAST), 大肠杆菌Exol, 人TREX2, 小鼠TREX1, 人TREX1, 牛TREX1, 大鼠TREX1, 人DNA2, 酵母DNA2 (DNA2_YEAST)。

[0120] 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)。

[0121] 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的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设计的存在下,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Fok1(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)。

[0122] 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))。

[0123] 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。

[0124] III.FAD2性能基因座

[0125] 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。

[0126] 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。

[0127] 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:

[0128] C18:0→C18:1→C18:2→C18:3

[0129] FAD2 FAD3

[0130] FAD2基因已经在主要植物和藻类物种中被鉴定,包括但不仅限于,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。

[0131] 如Chi,X.等人((2011) Genome-wide analysis of fatty acid desaturases in soybean (Glycine max). Plant Molecular Biology Reports 29, 769-783,本文通过提述并入之)中所述的那样,FAD2在大豆中的已知的功能性基因拷贝在系统发生学上与其拟南芥中的对应者一道被归入9个亚家族:FAB2,FAD2,FAD3,FAD5,FAD6,FAD7,FAD8,SLD1,和DES1。29个去饱和酶基因被发现分布在20个大豆染色体中的至少15个上。在亚家族之中基因结构和基序组成有相当大的保守性。基于微阵列数据分析,大多数的去饱和酶基因在不同的组织和发育阶段中显示特有的时间和空间表达模式。

[0132] 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,例如但不仅限于: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插入FAD2基因座不会诱发宿主的品质或健康(fitness)损失。

[0133] 在本发明的一些实施方案中,至少一个FAD2基因座(例如,FAD2 2.3基因座和/或FAD2 2.6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。

[0134] 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:14至SEQ ID NO:20。例如,FAD2基因座可以包括选自下组的核苷酸序列:SEQ ID NO:14至SEQ ID NO:20。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:SEQ ID NO:14至SEQ ID NO:20。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仅限于至少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:SEQ ID NO:14至SEQ ID NO:20。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。

[0135] IV. 核酸在FAD2基因座处的靶向整合

[0136] 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。

[0137] 在一些实施方案中,可以通过宿主细胞的内源细胞机构,例如但不仅限于,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。

[0138] A. DNA-结合多肽

[0139] 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。

[0140] 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,例如但不仅限于: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。

[0141] 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。

[0142] 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。

[0143] 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。

[0144] 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。

[0145] 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。

[0146] 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 01/60970 WO 01/88197; WO 02/099084; WO 98/53058; WO 98/53059; WO 98/53060; WO 02/016536 和 WO 03/016496。

[0147] 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。

[0148] 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。

[0149] 天然的GAL4由881个氨基酸残基组成,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) *Science* 231:699-704; 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。

[0150] 其他在某些实施方案中可以使用的DNA结合多肽包括,例如但不仅限于,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(E111iston 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. USA* 91 (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。

[0151] 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结合特异性是关键的。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。

[0152] 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。

[0153] 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。

[0154] 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, 但不仅限于, 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, 但不仅限于: 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。

[0155] 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,例如但不仅限于,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)。

[0156] B. 靶向核酸内切酶

[0157] 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,例如但不仅限于,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,例如但不仅限于: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。

[0158] 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。

[0159] 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。

[0160] 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。

[0161] 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,但不仅限于,下述多肽的组合:锌指DNA结合多肽;FokI核酸酶多肽;TALE结构域;亮氨酸拉链;转录因子的DNA结合基序;以及从例如但不仅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。

[0162] 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,但不仅限于,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H0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。

[0163] 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。

[0164] 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。

[0165] 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。

[0166] 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。

[0167] 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。

[0168] 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。

[0169] 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。

[0170] 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。

[0171] 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。

[0172] 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。

[0173] C. 锌指核酸酶

[0174] 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FokI)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)。FokI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。

[0175] 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。

[0176] 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。

[0177] V. 用于整合在FAD2基因座处的外源核酸

[0178] 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,例如但不仅限于,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。

[0179] A. 用于位点特异性整合的外源核酸分子

[0180] 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”)的插入,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,为了靶向插入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。

[0181] 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,将一个或更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. USA 84:4959-4963; Nehls et al. (1996) Science 272:886-889。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。

[0182] 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。

[0183] 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。

[0184] 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。

[0185] 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。

[0186] 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。

[0187] 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。

[0188] 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(太平洋折翅蠊(*Dipteroptera puntata*)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(phenylproanoid 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蛋白 (Tachyplesin) 的肽衍生物) ; 和PCT国际专利公开No. WO 95/18855 (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)) 。

[0189] 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, 例如但不仅限于: 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 (包括但不仅限于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, 例如来自链霉属物种, 包括吸水链霉菌和 *Streptomyces viridichromogenes* 的草铵膦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. 0 333 033 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-膦丝菌素的抗性。示例性 PAT 基因的核苷酸序列在欧洲专利申请No. 0 242 246, 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WO 2005012515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WO 2005107437和WO 2007053482中被描述。

[0190] 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, 例如但不仅限于: 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) 或苯腈 (腈

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)。腈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)。

[0191] 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,增值性状例如但不仅限于: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.Bacteol.* 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))。

[0192] 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。

[0193] 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,包括但不仅限于,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。

[0194] 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个指的组合有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个基因组(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。

[0195] B. 包含编码靶向性核酸内切酶的核苷酸序列的核酸分子

[0196] 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。

[0197] 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,例如但不仅限于,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。

[0198] 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,由任何序列组成,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。

[0199] C. 载体和表达构建体

[0200] 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。

[0201] 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,其包括例如但不仅限于,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。

[0202] 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。

[0203] 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 -薏苡辛（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）。

[0204] 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）。

[0205] 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，但不仅限于；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)。

[0206] 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.Virology* 64:1590-7);和AGRinos(GenBank登录号No.V00087;和Bevan et al. (1983),同上)。

[0207] 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。

[0208] 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。

[0209] 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,但不仅限于: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;和氨基蝶呤抗性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,包括例如但不仅限于,氨苄青霉素;博莱霉素;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。

[0210] 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) ;xy1E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 (Ikatu 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。

[0211] 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。

[0212] 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) 。

[0213] 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, 包括例如但不仅限于: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。

[0214] 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。根瘤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。

[0215] 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 0 122 791),或发根土壤杆菌的R_i质粒。额外的植物转化载体包括,例如但不仅限于,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 0 120 516,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_i质粒和合适的二元载体进行基因转移。

[0216] 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。

[0217] 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。

[0218] 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。

[0219] 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,但不仅限于,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。

[0220] 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F₀植物)的有性结合(自交)产生F₁种子而获得。四分之一的所产F₁种子就转基因而言是纯合的。令F₁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)。

[0221] 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。

[0222] 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。

[0223] 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的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的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。

[0224] 正向和反向寡核苷酸引物可以被设计成与引入的感兴趣的核酸分子特异性退火,例如,与对应于感兴趣的核酸分子编码区的序列或感兴趣的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。

[0225] 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

[0226] 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)(例如,大豆FAD2 2.3基因座合伙FAD2 2.6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。

[0227] 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一个或多个特征: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，包括例如但不仅限于，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。

[0228] 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。

[0229] 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，例如但不仅限于：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)。

[0230] 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。

[0231] 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。

[0232] 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

[0233] 提供了与大豆中的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大豆基因组连锁群中的位置。

[0234] 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

[0235] 实施例1:来自五种大豆栽培种的FAD2靶序列的测序

[0236] 测序反应

[0237] 从大豆组织分离基因组DNA。对于栽培种X5和Westag,从冻干的胚胎发生性悬浮细胞分离并纯化基因组DNA;对于栽培种Jack、Williams 82和Maverick,从幼叶分离并纯化基因组DNA。基因组DNA使用DNeasy Plant Mini KitTM(Qiagen;Carlsbad,CA)按照制造商的规程提取。

[0238] FAD2 2.3和2.6基因

[0239] FAD2 2.3和2.6基因组DNA序列使用引物MA49(SEQ ID NO:1caagggttccaaacacaaaagcc)与MA51(SEQ ID NO:2catcaatacttggcctgtacc),或MA50(SEQ ID NO:3gaagaaggcctctcaagggttc)与MA51通过PCR扩增。针对1140bp基因的大致碱基40至1140的片段获得了基因组DNA序列。PCR反应条件为初始变性98°C 1min;然后35个循环的98°C 30s,60°C 15s,72°C 3min,最终延伸为72°C 5min。

[0240] 将所得的PCR扩增子悬浮在TE缓冲液中(1μg于TE缓冲液中),并使用Covaris E220SystemTM超声仪(Covaris;Woburn,MA)以下述设定剪断成约300bp的片段:峰值入射功率(peak incident power)140,占空因数10%,每次脉冲200个循环,处理时间430秒。使用PrepX DNA library kitTM(IntegenX;Pleasanton,CA),在Apollo 324TM Automation SystemTM(IntegenX)上按照制造商推荐的规程制备IlluminaTM(San Diego,CA)配对末端测序文库。简而言之,将经剪断的带有5'或3'突出端的DNA片段转化为磷酸化的平端DNA。将单一腺嘌呤(A)延伸添加到末端配对的DNA片段的3'末端,然后连接到索引的(indexed) IlluminaTM配对末端衔接子。最后,从机器人回收衔接子连接的文库产物,并通过PCR富集,PCR使用Illumina TruSeq PCRTM产物,在如下热循环条件下进行:初始变性98°C 30s,然后10个循环的98°C 10s、60°C 30s、72°C 30s,最终延伸为72°C 5min。将经富集的文库标准化为2nM并汇集。然后将汇集的文库用氢氧化钠变性,在杂交缓冲液中稀释至6pM,加载到Miseq flow cellTM(Illumina;San Diego,CA)上。根据Illumina的推荐规程在MiseqTM上进行2x 150循环运行,其中6个索引循环(index cycle)。

[0241] 对从 Illumina MiSeqTM 仪器的序列反应得出的配对末端读取 (paired-end reads) 进行修剪用于 TruSeq adapter sequencesTM (Illumina)。在修剪之后, 使用 Burrows Wheeler Aligner (BWA) (Li H. and Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics*, 25:1754-60) 将读取定位到 Williams 82 栽培种的大豆参考脚手架上。

[0242] 每种大豆栽培种作为单独的样品对待, 因此每个栽培种的测序读取单独定位。检查了脚手架上被定位的测序读取的深度大于零的区域。由于这些是来自扩增子的测序读取, 预期只有脚手架上的特定区域才有读取定位到它们。不同的样品的测序读取均定位到大豆染色体 10 和 20。对于每个样品, 共有序列使用 Mpileup 计算机程序获得。这些结果表明, 测序读取定位于两个旁系同源的推定 FAD 基因。将所得的序列读取比对, 以鉴定从栽培种 X5, Westag, Jack, Williams 82 和 Maverick 获得的旁系同源推定基因序列之间的 SNP/ 变化。

[0243] 序列比对用 Vector NTI Advance 11.0 计算机程序 (Life Technologies, Carlsbad, CA) 中的 AlignX[®] 程序进行, 显示于图 1 和图 2 中。AlignX[®] 使用修改的 Clustal W 程序来产生蛋白质或核酸序列的多重序列比对, 以供相似性比较和注释。如图 1 和图 2 所示, 对分离的序列的分析表明各个 FAD2 2.3 和 FAD2 2.6 序列分别享有高水平的序列相似性。

[0244] FAD2 2.3 基因对应于 Williams 82 参考基因组序列的染色体 10 上的碱基 49,417,070 至 49,418,219 (SEQ ID NO:4)。来自五种栽培种 (Williams 82, Jack, Maverick, X5 和 Westag) 的基因的序列从碱基 41-1140 起是相同的 (以所用的引物获得覆盖)。FAD2 2.6 对应于 Williams 82 参考基因组序列的染色体 20 上的碱基 34,178,330 至 34,179,475 (SEQ ID NO:9)。X5 序列中在下列位置上鉴定了相对于 Williams 82 参考序列的序列差异: (C>T), 352 (A>G), 633 (C>T), 645 (T>C), 658 (T>C), 894 (A>G)。Maverick 在 352 和 894 位置上具有与 X5 相同的碱基改变。

[0245] 实施例 2: 特异针对 FAD2 基因的锌指结合结构域的设计

[0246] 针对编码 FAD2 基因座各种功能序列的 DNA 序列的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 1 (识别螺旋区设计) 和表 2 (识别螺旋区域设计) 所示。锌指核酸酶 (ZFN) 靶位点被设计成结合 FAD2A 的靶位点。FAD2A 锌指设计被合并在锌指表达载体中, 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 Fok I 的核酸酶结构域 (Wah et al., (1998) *Proc. Natl. Acad. Sci. USA* 95:10564-10569) 的氨基酸 384-579) 融合, 形成 FAD2A 锌指核酸酶 (ZFNs)。对于 ZFN 1 至 3 构建了野生型 Fok I 和 eHF-Fok I 结构域二者 (见美国专利公布号 20110201055), 而对于 ZFN 4 至 7 仅使用了 eHF-Fok I 结构域。

[0247] 在 DLSSA 测定 (见美国专利申请公布号 20110301073) 中测试了 FAD2 2.3 和 2.6 设计的 ZFN 的活性, 以鉴定活性最高的 ZFN。评估了 ZFN 对克隆到哺乳动物细胞中的相关 FAD2 2.3 和 2.6 序列的剪切 (图 3)。将活性与高活性参比 ZFN (8266:8196) 比较; 标注了基线活性。

[0248] 表 1: FAD2 锌指蛋白的靶位点

质粒名称	靶 FAD2 基因	SEQ ID NO:	ZN 结合位点的序列	用于结合ZN 结合位点的ZN 单体
[0249]	pDAB115603	2.3	14 agccatgcgcgcacactccaa acagggtccctgac	37354, 37355
	pDAB115600	2.3	14 agccatgcgcgcac tccaa acagggtccctgac	37354, 37355
	pDAB115605	2.3	15 tctaccgtgtcaaccct gaaagg gttggttggctgtatg	37370, 37371
	pDAB115601	2.3	15 tctaccgtgtcaaccct gaaagg gttggttggctgtatg	37370, 37371
	pDAB115606	2.3	16 gccttgccttacatcgat tcac agaatggactggctgaa	37374, 37375
	pDAB115604	2.3	17 ctgtgacttactctct accgt gttgcacccctgaaagg	37366, 37367
	pDAB115607	2.6	18 agccatgcgcgcaccac tccaa acgggtccctgac	37384, 37385
	pDAB115602	2.6	18 agccatgcgcgcaccac tccaa acgggtccctgac	37384, 37385
	pDAB115609	2.6	19 ctgtgacttactgtct accgt gttgcacatgaaagg	37398, 37399
	pDAB115608	2.6	20 ttcaatgtctgttgc agacc tatgtatggtttgc	37392, 37393

[0250] 表2: 用于ZN结合位点的ZN单体的FAD2锌指设计

ZFP#	F1	F2	F3	F4	F5	F6
37354	SEQ ID NO:24 QSSDLR	SEQ ID NO:25 RKDALVA	SEQ ID NO:26 RSADLTR	SEQ ID NO:27 RSDDLTR	SEQ ID NO:28 RSDAMSQ	SEQ ID NO:29 RNASRTR
37355	SEQ ID NO:30 DRSNLSR	SEQ ID NO:31 HKWLRLNQ	SEQ ID NO:32 DSSDRKK	SEQ ID NO:33 LRHHLTR	SEQ ID NO:34 QSGTRKT	NA
37370	SEQ ID NO:35 QNAHRKT	SEQ ID NO:33 LRHHLTR	SEQ ID NO:36 QSGDLTR	SEQ ID NO:37 QTSTLSK	SEQ ID NO:38 TSGSLSR	SEQ ID NO:39 RSDHLTQ
37371	SEQ ID NO:40 RSDVLSE	SEQ ID NO:41 RSADLSR	SEQ ID NO:24 QSSDLSR	SEQ ID NO:42 RTDALRG	SEQ ID NO:33 LRHHLTR	SEQ ID NO:43 HRSARKR
37374	SEQ ID NO:44 DRSHLTR	SEQ ID NO:45 QSGNLHV	SEQ ID NO:46 RSDHLSA	SEQ ID NO:47 RSNLLVA	SEQ ID NO:48 QSGALAR	SEQ ID NO:49 DRSALAR
37375	SEQ ID NO:50 QSSNLAR	SEQ ID NO:51 QSSDLRR	SEQ ID NO:52	SEQ ID NO:53	SEQ ID NO:54	SEQ ID NO:55

			RSDTLSE	QSGHLSR	RSDVLST	QNAHRIK
[0252]	37366	SEQ ID NO:56 RSDNLSQ	SEQ ID NO:57 ASNDRKK	SEQ ID NO:58 RSDNLST	SEQ ID NO:59 MRQHLLN	SEQ ID NO:60 RSDNLAR
	37367	SEQ ID NO:62 RSDHLSR	SEQ ID NO:63 DRSNRKT	SEQ ID NO:64 RSDTLSA	SEQ ID NO:65 DKSTRTK	SEQ ID NO:66 TSGSLTR
	37384	SEQ ID NO:24 QSSDLSR	SEQ ID NO:25 RKDALVA	SEQ ID NO:26 RSADLTR	SEQ ID NO:27 RSDDLTR	SEQ ID NO:68 RSDALAR
	37385	SEQ ID NO:30 DRSNLSR	SEQ ID NO:31 HKWLRNQ	SEQ ID NO:32 DSSDRKK	SEQ ID NO:33 LRHHLTR	SEQ ID NO:69 RRDILHQ
	37398	SEQ ID NO:56 RSDNLSQ	SEQ ID NO:57 ASNDRKK	SEQ ID NO:58 RSDNLST	SEQ ID NO:59 MRQHLLN	SEQ ID NO:70 QRTHLKA
	37399	SEQ ID NO:62 RSDHLSR	SEQ ID NO:63 DRSNRKT	SEQ ID NO:40 RSDVLSE	SEQ ID NO:71 ARSTRTN	SEQ ID NO:66 TSGSLTR
	37392	SEQ ID NO:72 QSGNLAR	SEQ ID NO:73 WRISLAA	SEQ ID NO:30 DRSNLSR	SEQ ID NO:74 WKESLGA	SEQ ID NO:75 HRKSLSR
	37393	SEQ ID NO:24 QSSDLSR	SEQ ID NO:76 YHWYLKK	SEQ ID NO:77 TSGHLSR	SEQ ID NO:78 TSGNLTR	SEQ ID NO:79 WWTSRAL

[0253] 实施例3:锌指核酸酶和供体构建体

[0254] ZFN构建体

[0255] 含有使用如实施例2所述的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。

[0256] 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,每个构建体由单个开放阅读框构成,开放阅读框由两个opaque-2核定位信号::锌指核酸酶融合序列构成,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ORF23 3' 非翻译区 (Atu0RF23 3' UTR)。

[0257] 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。

[0258] 在递送到大豆原生质体之前,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。

[0259] 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:pDAB115603 (包含

具有eHF-FokI的ZFN37354和ZFN37355构建体),pDAB115600(包含具有野生型FokI的ZFN37354和ZFN37355构建体),pDAB115605(包含具有eHF-FokI的ZFN37370和ZFN37371构建体),pDAB115601(包含具有野生型FokI的ZFN37370和ZFN37371构建体),pDAB115606(包含具有eHF-FokI的ZFN37374和ZFN37375构建体),pDAB115604(包含具有eHF-FokI的ZFN37366和ZFN37367构建体),pDAB115607(包含具有eHF-FokI的ZFN37384和ZFN37385构建体),pDAB115602(包含具有野生型FokI的ZFN37384和ZFN37385构建体),pDAB115609(包含具有eHF-FokI的ZFN37398和ZFN37399构建体),和pDAB115608(包含具有eHF-FokI的ZFN37392和ZFN37393构建体)。

[0260] 供体构建体

[0261] 通过将从头合成的线性片段与高拷贝质粒载体组合来构建FAD2供体载体。使用pDAB115620(图4)和pDAB115622(图5)在大豆基因组中的FAD2基因座内整合供体。两种供体载体均被合成为含有锌指核酸酶结合结构域。pDAB115620(“供体1”)包含37354:37355ZFN结合结构域、37366:37367ZFN结合结构域、37370:37371ZFN结合结构域、和37374:37375ZFN结合结构域。pDAB115622(“Donor 2”)包含37384:37385ZFN结合结构域、37392:37393ZFN结合结构域、和37398:37399ZFN结合结构域。ZFN结合结构域被相应的表达的锌指核酸酶所识别,并且在与供体载体和锌指核酸酶载体的共转化过程中在植物细胞中被剪切。

[0262] 实施例4:大豆原生质体的转化

[0263] 开发了一种基于大豆(例如大豆栽培种Maverick)原生质体的转化方法。原生质体自Maverick悬浮培养物分离,该Maverick悬浮培养物源自从叶外植体产生的愈伤组织。下述的技术描述了该方法。

[0264] 培养物维持

[0265] 大豆细胞悬液每7日通过用新鲜LS培养基(Linsmaier与Skoog 1965)1:5稀释来亚培养,LS培养基含有3% (w/v) 蔗糖、0.5mg/L 2,4-D,和7g细菌琼脂,pH 5.7。所有实验均在亚培养后7日开始实施,基于下述的规程。

[0266] 原生质体分离

[0267] 将30毫升亚培养后7日的Maverick悬浮培养物转移到50ml锥形离心管中,200g离心3分钟,每管产生大约10ml沉降细胞体积(settled cell volume,SCV)。在不扰动细胞沉淀的同时去除上清液。每4个SCV的悬浮细胞加入20毫升酶溶液(0.3% 果胶裂解酶(320952; MP Biomedicals),3% 纤维素酶(“Onozuka”R10TM; Yakult Pharmaceuticals, Japan),溶于MMG溶液(4mMMES, 0.6M甘露醇, 15mM MgCl₂, pH 6.0),用ParafilmTM包住管。将管放置在平台式振荡器上过夜(约16-18小时)。次日早晨,取一等份消化细胞显微镜检以确保细胞壁消化充分。

[0268] 原生质体纯化

[0269] 将细胞/酶溶液缓慢过滤通过100μM细胞滤网。用10ml W5+培养基(1.82mM MES, 192mM NaCl, 154mM CaCl₂, 4.7mM KC1, pH 6.0)漂洗细胞滤网。过滤步骤用70μM筛重复。通过加入10ml W5+培养基将最终体积定为40ml。通过颠倒管混合细胞。将8ml缓冲垫溶液(500mM蔗糖, 1mM CaCl₂, 5mM MES-KOH, pH 6.0)加到含有细胞的50ml锥形离心管的底部来使原生质体缓慢地在缓冲垫溶液上成层。将离心管在翻斗式转头中350g离心15分钟。使用5ml移液器吸头缓慢地移出原生质体条带(约7-8ml)。然后将原生质体转移到50ml锥形管

中,加入25ml的W5+洗涤液。将管缓慢颠倒,200g离心10分钟。移出上清液,加入10ml MMG溶液,并将管缓慢颠倒以重悬原生质体。使用血细胞计数仪或流式细胞仪确定原生质体密度。典型地,4PCV的细胞悬液产生大约2百万个原生质体。

[0270] 使用PEG转化原生质体

[0271] 用MMG将原生质体浓度调整至1.6百万个/ml。将300μl的原生质体等份(大约500,000个原生质体)转移2ml无菌管中。在将原生质体转移到管的过程中定期混合原生质体悬液。按照实验设计将质粒DNA加入原生质体等份中。将容纳原生质体管的架子缓慢颠倒3次,每次1分钟,以混合DNA和原生质体。将原生质体在室温温育5分钟。向原生质体中加入300毫升聚乙二醇(PEG4000)溶液(40%乙二醇(81240-Sigma Aldrich),0.3M甘露醇,0.4M CaCl₂),将架上的管混合1分钟,温育1分钟,在温育过程中轻柔颠倒两次。向管中缓慢加入1毫升W5+,将架上的管颠倒15-20次。然后将管350g离心5分钟,在不扰动沉淀的同时移出上清液。向每个管中加入1毫升WI培养基(4mM MES 0.6M甘露醇,20mM KC1,pH 6.0),将架子轻柔颠倒以重悬沉淀。用铝箔覆盖架子并使之侧卧,23℃温育过夜。

[0272] 测量转化效率并收获原生质体

[0273] 使用Quanta Flow CytometerTM(Beckman-Coulter Inc)实施原生质体定量和转化效率的测量。在转化后大约16-18小时,从每个重复取样100μl,置于96孔平板中,用WI溶液1:1稀释。将各重复重悬3次,使用流式细胞术对100μl定量。将样品送去分析之前,将样品200g离心5min,去除上清液,并将样品在液氮中快速冷冻。然后将样品置于-80℃冰柜中直至对其处理用于分子分析。

[0274] 实施例5:锌指核酸酶剪切和供体整合

[0275] 利用上文所述的转化方法将设计的ZFN转化到大豆原生质体中。通过如美国临时专利申请号61/736,856中描述的一种基因座破坏测定法,测试了各种ZFN对FAD2基因座的剪切效率。此外,通过一种入-出PCR测定法评估了该FAD2基因座内供体序列的锌指核酸酶介导的整合,并对所得的PCR扩增子进行测序,以表征大豆基因组内的供体整合。

[0276] 实验包括:仅包括供体载体的处理组、仅包括ZFN载体的处理组,或包括ZFN载体和供体载体二者的处理组(表3)。此外,实验还包括未转化细胞或用对照载体pDAB7221(图6)转化的细胞的阴性对照处理组,其中pDAB7221包含在高拷贝数质粒内由CsVMV启动子驱动、被AtuORF243'-UTR侧翼的GFP表达盒。在转染后大约18-24小时收集转化的样品。实验数据显示了ZFN质粒pDAB115601的高活性,将这种ZFN质粒用作所有后续实验的阳性对照。

[0277] 如表3详细说明的,转化实验包含总共80μg DNA,视需要加入pDAB7221质粒以使DNA的总浓度达到80μg。供体载体对ZFN表达质粒的比大约为10:1。每个实验或处理由6个实验重复组成,各重复独立地处理和分析。在其中两组实验中进行了评估ZFN的实验,在所有最终实验中均用ZFN质粒pDAB115601作为阳性对照。

[0278] 表3:实验设计。在两组(F2ZFN1-3和F2ZFN 4-7)中评估了ZFN质粒。适合ZFN质粒的供体质粒被测试并用于靶定实验。每个实验进行了6个重复。

[0279]

样品 ID	供体质粒	供体质粒的量(μg)	ZFN 质粒(μg)	ZFN 质粒的量(μg)	pDAB7221 的量(GFP)
未处理	--	--	--	--	--
GFP 对照	--	--	--	--	80
仅供体 1	pDAB115620	36	--	--	44
仅供体 2	pDAB115622	36	--	--	44
仅 F2 ZFN1_WT	--	--	pDAB115600	4	76
仅 F2 ZFN2_WT	--	--	pDAB115601	4	76
仅 F2 ZFN3_WT	--	--	pDAB115602	4	76
仅 F2 ZFN1_HF	--	--	pDAB115603	4	76
仅 F2 ZFN2_HF	--	--	pDAB115605	4	76
仅 F2 ZFN3_HF	--	--	pDAB115607	--	--
供体 1+F2 ZFN1_WT	pDAB115620	36	pDAB115600	4	40
供体 1+F2 ZFN2_WT	pDAB115620	36	pDAB115601	4	40
供体 2+F2 ZFN3_WT	pDAB115622	36	pDAB115602	4	40
供体 1+F2 ZFN1_HF	pDAB115620	36	pDAB115603	4	40
供体 1+F2 ZFN2_HF	pDAB115620	36	pDAB115605	4	40
供体 2+F2 ZFN3_HF	pDAB115622	36	pDAB115607	4	40

[0280]

样品 IDs	供体质粒	供体质粒的量(μg)	ZFN 质粒(μg)	ZFN 质粒的量(μg)	pDAB7221 (GFP) 的量
未处理	--	--	--	--	--
GFP 对照	--	--	--	--	80
仅供体 1	pDAB115620	36	--	--	44
仅供体 2	pDAB115622	36	--	--	44
仅 F2 ZFN2_WT	--	--	pDAB115601	4	76
仅 F2 ZFN4_HF	--	--	pDAB115609	4	76
仅 F2 ZFN5_HF	--	--	pDAB115608	4	76
仅 F2 ZFN6_HF	--	--	pDAB115606	4	76
仅 F2 ZFN7_HF	--	--	pDAB115604	4	76
供体 1+F2 ZFN2_WT	pDAB115620	36	pDAB115601	4	40
供体 2+F2 ZFN4_HF	pDAB115622	36	pDAB115609	4	40
供体 2+F2 ZFN5_HF	pDAB115622	36	pDAB115608	4	40
供体 1+F2 ZFN6_HF	pDAB115620	36	pDAB115606	4	40
供体 1+F2 ZFN7_HF	pDAB115620	36	pDAB115604	4	40

[0281] 靶定的分析

[0282] 对于来自靶定实验的DNA样品, 使用基因座破坏测定法进行分析以检测FAD2ZFN剪切位点的修饰或评估NHEJ靶定。涉及了一种qPCR测定来测量FAD2靶中的完整ZFN结合位点。ZFN介导的供体插入或剪切和随后的NHFJ修复导致ZFN结合位点丧失, 随后可检测的qPCR信号减少。具有显著的剪切活性的ZFN导致产生与单独供体处理相比信号减少的扩增子。基因座破坏测定中使用的引物和探针在表4中给出, 它们在FAD2基因座上的相对位置在图7中给

出。

[0283] 将结果与从未处理的大豆细胞中的完整FAD2基因座获得的信号比较。在合适的供体载体存在下用FAD2 2.3ZFN2_WT ZFN(两个实验) 和FAD2 2.6ZFNs ZFN4_HF(一个实验) 和F2ZFN5_HF(两个实验) 处理原生质体导致与从完整序列(单独的供体)获得的信号相比统计学上显著更低的信号。

[0284] 表4用于破坏PCR的引物和探针

[0285]	引物名称	序列	探针(荧光团/淬灭物)	靶ZFN

[0286]	GMS116 SOY F	SEQ ID NO:21 GTAATATGGGCTCAGAGGAATGGT	--	--
	GMS116 SOY R	SEQ ID NO:22 ATGGAGAAGAACATTGGAATTGC	--	--
	GMS116 SOY	SEQ ID NO:23 CCATGGCCCGGTACCATCTGGTC	HEX	--
	MAS723	SEQ ID NO:80 CACGAGTGTGGTCACCATGCCTT	--	ZFN1
	MAS724	SEQ ID NO:81 TGAGTGTGACGAGAAGAGAACAGC C	--	ZFN1
	MAS725_FAM	SEQ ID NO:82 AGCAAGTACCAATGGGTTGATGATG TTGTG	FAM	ZFN1
	MAS727	SEQ ID NO:83 TGCAAGCCACTACCACCCTTATGC	--	ZFN2/ZFN7
	MAS728	SEQ ID NO:84 GGCAAAGTGTGTGCTGCAAATAT G	--	ZFN2/ZFN7
	MAS729_FAM	SEQ ID NO:85 CTAACCGTGAGAGGGCTTCTGATCTAT GTCTCTGA	FAM	ZFN2/ZFN7
	MAS731	SEQ ID NO:86 TGAGTGTGATGAGAAGAGAAGCAGC C	--	ZFN3
	MAS732_FAM	SEQ ID NO:87 AGCAAGTACCCATGGGTTGATGATG TTATG	FAM	ZFN3
	MAS723	SEQ ID NO:80 CACGAGTGTGGTCACCATGCCTT	--	ZFN3
	MAS812	SEQ ID NO:88	--	ZFN6

	TTGGTTGGCTGCTATGTGTTATGG		
MAS813	SEQ ID NO:89 TGTGGCATTGTAGAGAAGAGATGGT GAG	--	ZFN6
MAS814_FAM	SEQ ID NO:90 AGGGAGCTTGGCAACTATGGACAG AGATTAT	FAM	ZFN6
[0287] MAS824	SEQ ID NO:91 AGCCTTCAATGTCTCTGGCAGACCCT	--	ZFN4/ZFN5
MAS818	SEQ ID NO:92 GGCATAAGTGTGTGCTGCAGATAT G	--	ZFN4/ZFN5
MAS817_FAM	SEQ ID NO:93 CAAATCGTGAGAGGGCTTTGATCTAT GTCTCTGA	FAM	ZFN4/ZFN5

[0288] 基因座特异性入-出PCR

[0289] 为了确证靶向供体插入,对来自所有处理的DNA进行了基因座特异性入-出PCR测定。这些实验中的供体载体被设计为含有所有被测试FAD2基因座内靶向整合的ZFN的结合位点。向大豆细胞内共投送ZFN和供体导致目标处和供体载体中的ZFN结合位点被剪切,以及随后供体通过非同源末端连接机制整合到被剪切的FAD2基因座之内。在整合到FAD2基因座内之前,ZFF剪切所产生的FAD2染色体位点及线性化供体载体的末端接受加工,可能导致不完美的末端连接产物。目标处的靶向整合的确认基于“入-出”PCR策略来进行,其中“出”引物识别天然基因座基因座处的序列,“入”引物结合供体DNA之内的序列。入-出PCR测定在插入节点的5' 和3' 末端都进行。

[0290] 如供体和ZFN样品中的PCR产物确定的,所有被测试的ZFN均在至少一个实验中显示了供体片段靶定和整合到FAD2大豆基因座中的某些证据。利用下述ZFN:F2ZFN2_WT、F2ZFN2_HF和F2ZFN4_HF获得的供体整合靶定的结果是可重复的,因为5' 和3' 末端的6个实验重复中至少都有2个产生的PCR产物(表5)。

[0291] 表5:大豆原生质体中FAD2基因座处的NHEJ靶定的概览。示出了各实验或处理的在独立的靶定实验中入-出PCR阳性的重复的数目。

[0292]	ZFN ID	F2ZFN1-3A	F2ZFN1-3B	F2ZFN4-7A	F2ZFN4-7B
	ZFN 1WT	1/6	0/6	--	--
	ZFN 1HF	1/6	4/6	--	--
	ZFN 2WT	3/6	5/6	5/6	5/6
	ZFN 2HF	4/6	3/6	--	--

ZFN 3WT	0/6	0/6	--	--
ZFN 3HF	0/6	0/6	--	--
ZFN 4HF	--	--	2/6	2/6
ZFN 5HF	--	--	0/6	0/6
ZFN 6HF	--	--	0/6	0/6
ZFN 7HF	--	--	4/6	0/6

[0293] 入-出PCR产物的测序

[0294] 将来自用pDAB1115620与F2ZFN2_WT、或pDAB1115620与F2ZFN2_HF完成的入-出PCR靶向实验的两个扩增子(具有预期的大小)克隆到质粒中。用Sanger测序法对所得的质粒测序。将序列与参考序列比对,在参考序列中,预测由FokI剪切导致的单链4bp末端被重复,以代表这些末端的所有可能的组合。从获得的23个克隆的序列中找到了10种独特的序列模式(图8)。所有序列模式均在ZFN结合位点(GAAATTTC)之间保留了FAD2基因组参考序列的一部分,但序列模式相对于FAD2基因组参考序列也具有缺失。序列4WT1和4WT4含有延伸到GAAATTTC序列的3'端的ZFN结合位点中的缺失。两条序列,1HF4和6HF4,具有单碱基插入。观察到的DNA序列模式表明发生了供体DNA到大豆FAD2基因座中的靶定。

[0295] 本文中描述了某些示例性的实施方案,但本领域普通技术人员会意识到并且理解,可以对这些示例性实施方案进行许多添加、删除和修改而不背离后述的权利要求的范围。此外,来自一个实施方案的特征可以与来自另一个实施方案的特征组合。

- [0001] 序列表
[0002] <110> DOW AGROSCIENCES LLC
[0003] <120> 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
[0004] <130> 8326-4012.40
[0005] <140> PCT/US2013/058299
[0006] <141> 2013-09-05
[0007] <150> 61/697,886
[0008] <151> 2012-09-07
[0009] <160> 130
[0010] <170> PatentIn version 3.5
[0011] <210> 1
[0012] <211> 22
[0013] <212> DNA
[0014] <213> 人工序列
[0015] <220>
[0016] <223> 人工序列描述:合成引物
[0017] <400> 1
[0018] caagggttcc aaacacaaaag cc 22
[0019] <210> 2
[0020] <211> 22
[0021] <212> DNA
[0022] <213> 人工序列
[0023] <220>
[0024] <223> 人工序列描述:合成引物
[0025] <400> 2
[0026] catcaatact tggccctgtt cc 22
[0027] <210> 3
[0028] <211> 23
[0029] <212> DNA
[0030] <213> 人工序列
[0031] <220>
[0032] <223> 人工序列描述:合成引物
[0033] <400> 3
[0034] gaagaaggct ctctcaaggg ttc 23
[0035] <210> 4
[0036] <211> 1100
[0037] <212> DNA
[0038] <213> 大豆
[0039] <400> 4
[0040] ggaagaagcc tctctcaagg gttccaaaca caaagccacc attcactgtt ggccaactca 60
[0041] agaaaagcaat tccaccacac tgctttcagc gctccctcct cacttcattc tcctatgtt 120

[0042] ttatgacct ttcatttgcc ttcatttct acattgccac cacctacttc caccccttc 180
 [0043] ctcaccctt ttccctcatt gcatggccaa tctattgggt tctccaaggt tgccttctca 240
 [0044] ctgggtgtg ggtgattgtc cacgagtgtg gtcaccatgc cttcagcaag taccaatggg 300
 [0045] ttgatgatgt tgggggttg acccttact caacacttt agtcccttat ttctcatgga 360
 [0046] aaataagcca tcgcccatt cactccaaca caggtccct tgaccgtat gaagtgtttg 420
 [0047] tccaaaacc aaaatccaaa gttcatgtt ttccaagta cttaacaac cctctaggaa 480
 [0048] gggctgtttc tcttcgttc acactcacaa tagggtgcc tatgtattta gccttcaatg 540
 [0049] tctctggtag accctatgtat agtttgcaa gccactacca cccttatgtc cccatatatt 600
 [0050] ctaaccgtga gaggcttctg atctatgtct ctgatgttc tttttttct gtgacttact 660
 [0051] ctctctaccg tggcaacc ctgaaagggt tggttgggt gctatgtgtt tatgggtgc 720
 [0052] cttgctcat tggcaacgggt ttcttgcata ctatcacata tttgcagcac acacacttt 780
 [0053] cttgcctca ttacgattca tcagaatggg actggctgaa gggagcttg gcaactatgg 840
 [0054] acagagatta tgggattctg aacaagggtt ttcatcacat aactgataact catgtggctc 900
 [0055] accatctt ctctacaatg ccacattacc atgcaatggg ggcaaccaat gcaatcaagc 960
 [0056] caatattggg tggactac caatttgatg acacaccatt ttacaaggca ctgtggagag 1020
 [0057] aagcgagaga gtgccttat gtggagccag atgaaggaac atccgagaag ggcgtgtatt 1080
 [0058] ggtacaggaa caagtattga 1100
 [0059] <210> 5
 [0060] <211> 1100
 [0061] <212> DNA
 [0062] <213> 大豆
 [0063] <400> 5
 [0064] ggaagaagcc tctctcaagg gttccaaaca caaagccacc attactgtt gccaaactca 60
 [0065] agaaagcaat tccaccacac tgcttcagc gctccctcct cacttcattc tcctatgtt 120
 [0066] ttatgacct ttcatttgcc ttcatttct acattgccac cacctacttc caccccttc 180
 [0067] ctcaccctt ttccctcatt gcatggccaa tctattgggt tctccaaggt tgccttctca 240
 [0068] ctgggtgtg ggtgattgtc cacgagtgtg gtcaccatgc cttcagcaag taccaatggg 300
 [0069] ttgatgatgt tgggggttg acccttact caacacttt agtcccttat ttctcatgga 360
 [0070] aaataagcca tcgcccatt cactccaaca caggtccct tgaccgtat gaagtgtttg 420
 [0071] tccaaaacc aaaatccaaa gttcatgtt ttccaagta cttaacaac cctctaggaa 480
 [0072] gggctgtttc tcttcgttc acactcacaa tagggtgcc tatgtattta gccttcaatg 540
 [0073] tctctggtag accctatgtat agtttgcaa gccactacca cccttatgtc cccatatatt 600
 [0074] ctaaccgtga gaggcttctg atctatgtct ctgatgttc tttttttct gtgacttact 660
 [0075] ctctctaccg tggcaacgggt tggttgggt gctatgtgtt tatgggtgc 720
 [0076] cttgctcat tggcaacgggt ttcttgcata ctatcacata tttgcagcac acacacttt 780
 [0077] cttgcctca ttacgattca tcagaatggg actggctgaa gggagcttg gcaactatgg 840
 [0078] acagagatta tgggattctg aacaagggtt ttcatcacat aactgataact catgtggctc 900
 [0079] accatctt ctctacaatg ccacattacc atgcaatggg ggcaaccaat gcaatcaagc 960
 [0080] caatattggg tggactac caatttgatg acacaccatt ttacaaggca ctgtggagag 1020
 [0081] aagcgagaga gtgccttat gtggagccag atgaaggaac atccgagaag ggcgtgtatt 1080
 [0082] ggtacaggaa caagtattga 1100
 [0083] <210> 6

[0084] <211> 1100
 [0085] <212> DNA
 [0086] <213> 大豆
 [0087] <400> 6
 [0088] ggaagaagcc tctctcaagg gttccaaaca caaagccacc attcaactgtt gcccaactca 60
 [0089] agaaaagcaat tccaccacac tgcttcagc gctccctcct cacttcattc tcctatgtt 120
 [0090] ttatgacct ttcatttgcc ttcatttct acattgccac cacctacttc caccccttc 180
 [0091] ctcaaccctt ttccctcatt gcatggccaa tctattgggt tctccaaggt tgcctctca 240
 [0092] ctgggtgtgt ggtgattgtt cacgagtgtg gtcaccatgc cttcagcaag taccaatggg 300
 [0093] ttgatgatgt tgggggtttg acccttact caacactttt agtcccttat ttctcatgga 360
 [0094] aaataagcca tcgcccacat cactccaaca caggttcct tgaccgtat gaagtgtttt 420
 [0095] tcccaaaacc aaaatccaaa gttcatggt ttccaagta cttaacaac cctctaggaa 480
 [0096] gggctgttgc tcttcgttc acactcacaa tagggtgcc tatgtattta gccttcaatg 540
 [0097] tctctggtag accctatgtat agtttgcaa gccactacca cccttatgtt cccatatatt 600
 [0098] ctaaccgtga gaggcttctg atctatgtct ctgatgttgc tttgtttct gtgacttact 660
 [0099] ctctctaccg tggcaacc ctgaaagggt tgggtggct gctatgtgtt tatgggggtgc 720
 [0100] cttgctcat tggcaacggt ttcttgta ctatcacata tttcagcac acacactttt 780
 [0101] cttgcctca ttacgattca tcagaatggg actggctgaa gggagcttt gcaactatgg 840
 [0102] acagagatta tgggattctg aacaagggtt ttcatcacat aactgataact catgtggctc 900
 [0103] accatctt ctctacaatg ccacattacc atgcaatggg ggcaaccaat gcaatcaagc 960
 [0104] caatattggg tggactac caatttgatg acacaccatt ttacaaggca ctgtggagag 1020
 [0105] aagcgagaga gtgccttat gtggagccag atgaaggaac atccgagaag ggcgtgtatt 1080
 [0106] ggtacaggaa caagtattga 1100
 [0107] <210> 7
 [0108] <211> 1100
 [0109] <212> DNA
 [0110] <213> 大豆
 [0111] <400> 7
 [0112] ggaagaagcc tctctcaagg gttccaaaca caaagccacc attcaactgtt gcccaactca 60
 [0113] agaaaagcaat tccaccacac tgcttcagc gctccctcct cacttcattc tcctatgtt 120
 [0114] ttatgacct ttcatttgcc ttcatttct acattgccac cacctacttc caccccttc 180
 [0115] ctcaaccctt ttccctcatt gcatggccaa tctattgggt tctccaaggt tgcctctca 240
 [0116] ctgggtgtgt ggtgattgtt cacgagtgtg gtcaccatgc cttcagcaag taccaatggg 300
 [0117] ttgatgatgt tgggggtttg acccttact caacactttt agtcccttat ttctcatgga 360
 [0118] aaataagcca tcgcccacat cactccaaca caggttcct tgaccgtat gaagtgtttt 420
 [0119] tcccaaaacc aaaatccaaa gttcatggt ttccaagta cttaacaac cctctaggaa 480
 [0120] gggctgttgc tcttcgttc acactcacaa tagggtgcc tatgtattta gccttcaatg 540
 [0121] tctctggtag accctatgtat agtttgcaa gccactacca cccttatgtt cccatatatt 600
 [0122] ctaaccgtga gaggcttctg atctatgtct ctgatgttgc tttgtttct gtgacttact 660
 [0123] ctctctaccg tggcaacc ctgaaagggt tgggtggct gctatgtgtt tatgggggtgc 720
 [0124] cttgctcat tggcaacggt ttcttgta ctatcacata tttcagcac acacactttt 780
 [0125] cttgcctca ttacgattca tcagaatggg actggctgaa gggagcttt gcaactatgg 840

[0126] acagagatta tgggattctg aacaagggtgt ttcatacacat aactgatact catgtggctc 900
 [0127] accatctt ctctacaatg ccacattacc atgcaatgga ggcaaccaat gcaatcaagc 960
 [0128] caatattggg tgagtactac caatggatg acacaccatt ttacaaggca ctgtggagag 1020
 [0129] aagcgagaga gtgccttat gtggagccag atgaaggaac atccgagaag ggcgtgtatt 1080
 [0130] ggtacagggaa caagtattga 1100
 [0131] <210> 8
 [0132] <211> 1100
 [0133] <212> DNA
 [0134] <213> 大豆
 [0135] <400> 8
 [0136] ggaagaagcc tctctcaagg gttccaaaca caaagccacc attcaactgtt gccaaactca 60
 [0137] agaaagcaat tccaccacac tgcttcage gctccctct cacttcattc tcctatgtt 120
 [0138] ttatgacct ttcatttgcc ttcatggct acattgccac cacctacttc caccccttc 180
 [0139] ctcaaccctt ttccctcatt gcatggccaa tctattgggt tctccaaggt tgcctctca 240
 [0140] ctgggtgtg ggtgattgtc cacgagtgtg gtcaccatgc cttcagcaag taccaatggg 300
 [0141] ttgatgatgt tgggggtttg acccttactt caacactttt agtcccttat ttctcatgga 360
 [0142] aaataagcca tcgcccacat cactccaaca caggtccct tgaccgtat gaagtgttt 420
 [0143] tccaaaacc aaaatccaaa gttcatggt tttccaagta cttaacaac cctctaggaa 480
 [0144] gggctgttcc tcttcgtc acactcacaa tagggtgcc tatgtattta gccttcaatg 540
 [0145] tctctggtag accctatgt agtttgcaaa gccactacca cccttatgtc cccatatatt 600
 [0146] ctaaccgtga gaggcttctg atctatgtct ctgatgttgc tttttttct gtgacttact 660
 [0147] ctcttaccg tggcaacc ctgaaagggt tgggggttgc gctatgtt tatgggtgc 720
 [0148] ctttgctcat tggcaacggt ttcttgcata ctatcacata tttgcagcac acacacttt 780
 [0149] cttgcctca ttacgattca tcagaatggg actggctgaa gggagcttg gcaactatgg 840
 [0150] acagagatta tgggattctg aacaagggtgt ttcatacacat aactgatact catgtggctc 900
 [0151] accatctt ctctacaatg ccacattacc atgcaatgga ggcaaccaat gcaatcaagc 960
 [0152] caatattggg tgagtactac caatggatg acacaccatt ttacaaggca ctgtggagag 1020
 [0153] aagcgagaga gtgccttat gtggagccag atgaaggaac atccgagaag ggcgtgtatt 1080
 [0154] ggtacagggaa caagtattga 1100
 [0155] <210> 9
 [0156] <211> 1101
 [0157] <212> DNA
 [0158] <213> 大豆
 [0159] <400> 9
 [0160] cagaagaagc ctctctcaag ggtccaaac acaaagccac cattcactgt tggccaaactc 60
 [0161] aagaaagcca ttccaccgca ctgcttcag cgttccctcc tcacttcatt gtcctatgtt 120
 [0162] gtttatgacc ttcatggc ttcatggc tacattgca ccacccactt ccacccctc 180
 [0163] cttcacccct ttccctcat tgcattggca atctattggg ttctccaagg ttgcatttt 240
 [0164] actggcgtgt gggtgattgc tcacgagtgt ggtcaccatg cttcagcaa gtacccatgg 300
 [0165] gttgatgatgt ttatgggtt gaccgttac tcagcacttt tagtccctta ttctcatgg 360
 [0166] aaaataagcc atcgccgcca ccactccaac acgggttccc ttgaccgtga tgaagtgttt 420
 [0167] gttccaaaacc caaaaatccaa agttgcatgg tacaccaagt acctgaacaa ccctcttagga 480

[0168] agggctgctt ctcttctcat cacactcaca atagggtggc ctttgttattt agccttaat 540
 [0169] gtctctggca gaccctatga tgggtttgct agccactacc acccttatgc tcccatatat 600
 [0170] tcaaatcgta agaggctttt gatctatgtc tctgatgtt cttgttttc tgtgacttac 660
 [0171] ttgctctacc gtgttgcac tatgaaaggg ttgggttggc tgctatgtgt ttatgggtg 720
 [0172] ccattgctca ttgtgaacgg tttcttgtg accatcacat atctgcagca cacacactat 780
 [0173] gccttcctc actatgattc atcagaatgg gattggctga ggggtgcctt ggcaactatg 840
 [0174] gacagagatt atggaattct gaacaagggtg tttcaccaca taactgatac tcatgtggct 900
 [0175] caccatctt tctctacaat gccacattac catgcaacgg aggcaaccaa tgcaatgaag 960
 [0176] ccaatattgg gtgagtacta ccgatttgc gacacaccat tttacaaggc actgtggaga 1020
 [0177] gaagcaagag agtgcctcta tgtggagcca gatgaaggaa catccgagaa gggcgtgtat 1080
 [0178] tggtacagga acaagtattt a 1101
 [0179] <210> 10
 [0180] <211> 1101
 [0181] <212> DNA
 [0182] <213> 大豆
 [0183] <400> 10
 [0184] cagaagaagc ctctctcaag ggttccaaac acaaagccac cattcactgt tggccaactc 60
 [0185] aagaaagcca ttccaccgca ctgcttcag cggtccctcc tcacttcatt gtcctatgtt 120
 [0186] gtttatgacc ttccattggc ttccattttc tacattgcca ccacccactt ccacccctc 180
 [0187] cctcaccctt ttccctcat tgcattggc atctattggg ttcccaagg ttgcattctt 240
 [0188] actggcgtgt gggtgattgc tcacgagtgt ggtcaccatg cttcagcaaa gtacccatgg 300
 [0189] gttgatgatg ttatgggtt gacggttac tcagcactt tagtccctt tttctcatgg 360
 [0190] aaaataagcc atgcggccca ccactccaac acgggttccc ttgaccgtga tgaagtgttt 420
 [0191] gtcccaaaac caaaatccaa agttgcatgg tacaccaagt acctgaacaa ccctctagga 480
 [0192] agggctgctt ctcttctcat cacactcaca atagggtggc ctttgttattt agccttaat 540
 [0193] gtctctggca gaccctatga tgggtttgct agccactacc acccttatgc tcccatatat 600
 [0194] tcaaatcgta agaggctttt gatctatgtc tctgatgtt cttgttttc tgtgacttac 660
 [0195] ttgctctacc gtgttgcac tatgaaaggg ttgggttggc tgctatgtgt ttatgggtg 720
 [0196] ccattgctca ttgtgaacgg tttcttgtg accatcacat atctgcagca cacacactat 780
 [0197] gccttcctc actatgattc atcagaatgg gattggctga ggggtgcctt ggcaactatg 840
 [0198] gacagagatt atggaattct gaacaagggtg tttcaccaca taactgatac tcatgtggct 900
 [0199] caccatctt tctctacaat gccacattac catgcaacgg aggcaaccaa tgcaatgaag 960
 [0200] ccaatattgg gtgagtacta ccgatttgc gacacaccat tttacaaggc actgtggaga 1020
 [0201] gaagcaagag agtgcctcta tgtggagcca gatgaaggaa catccgagaa gggcgtgtat 1080
 [0202] tggtacagga acaagtattt a 1101
 [0203] <210> 11
 [0204] <211> 1101
 [0205] <212> DNA
 [0206] <213> 大豆
 [0207] <400> 11
 [0208] cagaagaagc ctctctcaag ggttccaaac acaaagccac cattcactgt tggccaactc 60
 [0209] aagaaagcca ttccaccgca ctgcttcag cggtccctcc tcacttcatt gtcctatgtt 120

[0210] gtttatgacc tttcattggc tttcatttc tacattgcca ccacctaatt ccacccctc 180
 [0211] cctcaccctt tttccat tgcattggca atctattggg ttctccaagg ttgcatttt 240
 [0212] actggcgtgt gggtgattgc tcacgagtgt ggtcaccatg cttcagcaa gtacccatgg 300
 [0213] gttgatgatg ttatgggaaa gaccgttac tcagcacttt tagtccctta ttctcatgg 360
 [0214] aaaataagcc atcgccgcca ccactccaac acgggtccc ttgaccgtga tgaagtgttt 420
 [0215] gtcccaaaac caaaatccaa agttgcatgg tacaccaagt acctaaca ccctctagga 480
 [0216] agggctgctt ctcttctcat cacactcaca atagggtggc ctttgtattt agccttcaat 540
 [0217] gtctctggca gaccctatga tgggtttgct agccactacc acccttatgc tccttatata 600
 [0218] tcaaaccgtg agaggcttct gatctatgtc tctgatgtt cttgttttgc tgtgacttac 660
 [0219] ttgctctacc gtgttgcac tatgaaagggtt ttgggttggc tgctatgtgt ttatgggggt 720
 [0220] ccattgctca ttgtgaacgg tttcttgc accatcacat atctgcagca cacacactat 780
 [0221] gccttcctc actatgatc atcagaatgg gattggctga ggggtgctt ggcaactatg 840
 [0222] gacagagatt atgggattct gaacaagggtt tttcaccaca taactgatac tcatgtggct 900
 [0223] caccatctt tctctacaat gccacattac catgcaacgg aggcaaccaa tgcaatgaag 960
 [0224] ccaatattgg gtgagttacta ccgatttgat gacacaccat tttacaaggc actgtggaga 1020
 [0225] gaagcaagag agtgcctcta tgtggagccaa gatgaaggaa catccgagaa gggcgtgtat 1080
 [0226] tggtacagga acaagtattt a 1101
 [0227] <210> 12
 [0228] <211> 1101
 [0229] <212> DNA
 [0230] <213> 大豆
 [0231] <400> 12
 [0232] cagaagaagc ctctctcaag gttccaaac acaaagccac cattcactgt tggccaaactc 60
 [0233] aagaaagcca ttccaccgca ctgtttcag cgttccctcc tcacttcatt gtcctatgtt 120
 [0234] gtttatgacc tttcattggc tttcatttc tacattgcca ccacctaatt ccacccctc 180
 [0235] cctcaccctt tttccat tgcattggca atctattggg ttctccaagg ttgcatttt 240
 [0236] actggcgtgt gggtgattgc tcacgagtgt ggtcaccatg cttcagcaa gtacccatgg 300
 [0237] gttgatgatg ttatgggaaa gaccgttac tcagcacttt tagtccctta ttctcatgg 360
 [0238] aaaataagcc atcgccgcca ccactccaac acgggtccc ttgaccgtga tgaagtgttt 420
 [0239] gtcccaaaac caaaatccaa agttgcatgg tacaccaagt acctaaca ccctctagga 480
 [0240] agggctgctt ctcttctcat cacactcaca atagggtggc ctttgtattt agccttcaat 540
 [0241] gtctctggca gaccctatga tgggtttgct agccactacc acccttatgc tccttatata 600
 [0242] tcaaaccgtg agaggcttct gatctatgtc tctgatgtt cttgttttgc tgtgacttac 660
 [0243] ttgctctacc gtgttgcac tatgaaagggtt ttgggttggc tgctatgtgt ttatgggggt 720
 [0244] ccattgctca ttgtgaacgg tttcttgc accatcacat atctgcagca cacacactat 780
 [0245] gccttcctc actatgatc atcagaatgg gattggctga ggggtgctt ggcaactatg 840
 [0246] gacagagatt atgggattct gaacaagggtt tttcaccaca taactgatac tcatgtggct 900
 [0247] caccatctt tctctacaat gccacattac catgcaacgg aggcaaccaa tgcaatgaag 960
 [0248] ccaatattgg gtgagttacta ccgatttgat gacacaccat tttacaaggc actgtggaga 1020
 [0249] gaagcaagag agtgcctcta tgtggagccaa gatgaaggaa catccgagaa gggcgtgtat 1080
 [0250] tggtacagga acaagtattt a 1101
 [0251] <210> 13

- [0252] <211> 1101
[0253] <212> DNA
[0254] <213> 大豆
[0255] <400> 13
[0256] cagaagaagc ctctctcaag ggttccaaac acaaagccac cattcactgt tggccaactc 60
[0257] aagaaagcca ttccaccgca ctgcttcag cgttccctcc tcacttcatt gtcctatgtt 120
[0258] gtttatgacc tttcattggc tttcattttc tacattgcca ccacctactt ccacccctc 180
[0259] ctcacccct tttccctcat tgcattggca atctattggg ttctccaagg ttgcattctt 240
[0260] actggcgtgt gggtgattgc tcacgagtgt ggtcaccatg cttcagcaa gtacccatgg 300
[0261] gttgatgatg ttgtgggtt gaccgttac tcagcactt tagtccctta tttctcatgg 360
[0262] aaaataagcc atgcggcca ccactccaac acgggttccc ttgaccgtga tgaagtgttt 420
[0263] gtcacccaaac caaaatccaa agtgcatgg tacaccaagt acctaaca ccctctagga 480
[0264] agggctgtt ctcttctcat cacactcaca atagggtggc ctttgttattt agccttcaat 540
[0265] gtctctggca gaccctatga tgggtttgct agccactacc acccttatgc tcccatatat 600
[0266] tcaaattcgta agaggctttt gatctatgtc tctgatgtt ctgtttttc tgtgacttac 660
[0267] ttgctctacc gtgttgcac tatgaaaggg ttgggttggc tgctatgtgt ttatgggtt 720
[0268] ccattgctca ttgtgaacgg tttcttgtg accatcacat atctgcagca cacacactat 780
[0269] gccttcctc actatgattc atcagaatgg gattggctga ggggtgcctt ggcaactatg 840
[0270] gacagagatt atgggattct gaacaagggtg tttcaccaca taactgatac tcatgtggct 900
[0271] caccatctt tctctacaat gcccattac catgcacccgg aggcaaccaa tgcaatgaag 960
[0272] ccaatattgg gtgagtacta ccgatttgat gacacaccat tttacaaggc actgtggaga 1020
[0273] gaagcaagag agtgcctcta tgtggagcca gatgaaggaa catccgagaa gggcgtgtat 1080
[0274] tggtacagga acaagtattt a 1101
[0275] <210> 14
[0276] <211> 39
[0277] <212> DNA
[0278] <213> 人工序列
[0279] <220>
[0280] <223> 人工序列描述:合成寡核苷酸
[0281] <400> 14
[0282] agccatcgcc gccatcaactc caacacaggt tcccttgac 39
[0283] <210> 15
[0284] <211> 43
[0285] <212> DNA
[0286] <213> 人工序列
[0287] <220>
[0288] <223> 人工序列描述:合成寡核苷酸
[0289] <400> 15
[0290] tctaccgtgt tgcaaccctg aaagggttgg tttggctgct atg 43
[0291] <210> 16
[0292] <211> 41
[0293] <212> DNA

- [0294] <213> 人工序列
- [0295] <220>
- [0296] <223> 人工序列描述:合成寡核苷酸
- [0297] <400> 16
- [0298] gccttcctc attacgattc atcagaatgg gactggctga a 41
- [0299] <210> 17
- [0300] <211> 41
- [0301] <212> DNA
- [0302] <213> 人工序列
- [0303] <220>
- [0304] <223> 人工序列描述:合成寡核苷酸
- [0305] <400> 17
- [0306] ctgtgactta ctctctcac cgtttgcaa ccctgaaagg g 41
- [0307] <210> 18
- [0308] <211> 39
- [0309] <212> DNA
- [0310] <213> 人工序列
- [0311] <220>
- [0312] <223> 人工序列描述:合成寡核苷酸
- [0313] <400> 18
- [0314] agccatcgcc gccaccactc caacacgggt tcccttgac 39
- [0315] <210> 19
- [0316] <211> 41
- [0317] <212> DNA
- [0318] <213> 人工序列
- [0319] <220>
- [0320] <223> 人工序列描述:合成寡核苷酸
- [0321] <400> 19
- [0322] ctgtgactta ctgtctcac cgtttgcaa ctatgaaagg g 41
- [0323] <210> 20
- [0324] <211> 36
- [0325] <212> DNA
- [0326] <213> 人工序列
- [0327] <220>
- [0328] <223> 人工序列描述:合成寡核苷酸
- [0329] <400> 20
- [0330] ttcaatgtct ctggcagacc ctatgatgg tttgct 36
- [0331] <210> 21
- [0332] <211> 24
- [0333] <212> DNA
- [0334] <213> 人工序列
- [0335] <220>

- [0336] <223> 人工序列描述:合成引物
- [0337] <400> 21
- [0338] gtaatatggg ctcagaggaa tggt 24
- [0339] <210> 22
- [0340] <211> 23
- [0341] <212> DNA
- [0342] <213> 人工序列
- [0343] <220>
- [0344] <223> 人工序列描述:合成引物
- [0345] <400> 22
- [0346] atggagaaga acattggaat tgc 23
- [0347] <210> 23
- [0348] <211> 23
- [0349] <212> DNA
- [0350] <213> 人工序列
- [0351] <220>
- [0352] <223> 人工序列描述:合成探针
- [0353] <400> 23
- [0354] ccatggcccg gtaccatctg gtc 23
- [0355] <210> 24
- [0356] <211> 7
- [0357] <212> PRT
- [0358] <213> 人工序列
- [0359] <220>
- [0360] <223> 人工序列描述:合成肽
- [0361] <400> 24
- [0362] Gln Ser Ser Asp Leu Ser Arg
- [0363] 1 5
- [0364] <210> 25
- [0365] <211> 7
- [0366] <212> PRT
- [0367] <213> 人工序列
- [0368] <220>
- [0369] <223> 人工序列描述:合成肽
- [0370] <400> 25
- [0371] Arg Lys Asp Ala Leu Val Ala
- [0372] 1 5
- [0373] <210> 26
- [0374] <211> 7
- [0375] <212> PRT
- [0376] <213> 人工序列
- [0377] <220>

- [0378] <223> 人工序列描述:合成肽
[0379] <400> 26
[0380] Arg Ser Ala Asp Leu Thr Arg
[0381] 1 5
[0382] <210> 27
[0383] <211> 7
[0384] <212> PRT
[0385] <213> 人工序列
[0386] <220>
[0387] <223> 人工序列描述:合成肽
[0388] <400> 27
[0389] Arg Ser Asp Asp Leu Thr Arg
[0390] 1 5
[0391] <210> 28
[0392] <211> 7
[0393] <212> PRT
[0394] <213> 人工序列
[0395] <220>
[0396] <223> 人工序列描述:合成肽
[0397] <400> 28
[0398] Arg Ser Asp Ala Met Ser Gln
[0399] 1 5
[0400] <210> 29
[0401] <211> 7
[0402] <212> PRT
[0403] <213> 人工序列
[0404] <220>
[0405] <223> 人工序列描述:合成肽
[0406] <400> 29
[0407] Arg Asn Ala Ser Arg Thr Arg
[0408] 1 5
[0409] <210> 30
[0410] <211> 7
[0411] <212> PRT
[0412] <213> 人工序列
[0413] <220>
[0414] <223> 人工序列描述:合成肽
[0415] <400> 30
[0416] Asp Arg Ser Asn Leu Ser Arg
[0417] 1 5
[0418] <210> 31
[0419] <211> 7

- [0420] <212> PRT
[0421] <213> 人工序列
[0422] <220>
[0423] <223> 人工序列描述:合成肽
[0424] <400> 31
[0425] His Lys Trp Leu Arg Asn Gln
[0426] 1 5
[0427] <210> 32
[0428] <211> 7
[0429] <212> PRT
[0430] <213> 人工序列
[0431] <220>
[0432] <223> 人工序列描述:合成肽
[0433] <400> 32
[0434] Asp Ser Ser Asp Arg Lys Lys
[0435] 1 5
[0436] <210> 33
[0437] <211> 7
[0438] <212> PRT
[0439] <213> 人工序列
[0440] <220>
[0441] <223> 人工序列描述:合成肽
[0442] <400> 33
[0443] Leu Arg His His Leu Thr Arg
[0444] 1 5
[0445] <210> 34
[0446] <211> 7
[0447] <212> PRT
[0448] <213> 人工序列
[0449] <220>
[0450] <223> 人工序列描述:合成肽
[0451] <400> 34
[0452] Gln Ser Gly Thr Arg Lys Thr
[0453] 1 5
[0454] <210> 35
[0455] <211> 7
[0456] <212> PRT
[0457] <213> 人工序列
[0458] <220>
[0459] <223> 人工序列描述:合成肽
[0460] <400> 35
[0461] Gln Asn Ala His Arg Lys Thr

- [0462] 1 5
[0463] <210> 36
[0464] <211> 7
[0465] <212> PRT
[0466] <213> 人工序列
[0467] <220>
[0468] <223> 人工序列描述:合成肽
[0469] <400> 36
[0470] Gln Ser Gly Asp Leu Thr Arg
[0471] 1 5
[0472] <210> 37
[0473] <211> 7
[0474] <212> PRT
[0475] <213> 人工序列
[0476] <220>
[0477] <223> 人工序列描述:合成肽
[0478] <400> 37
[0479] Gln Thr Ser Thr Leu Ser Lys
[0480] 1 5
[0481] <210> 38
[0482] <211> 7
[0483] <212> PRT
[0484] <213> 人工序列
[0485] <220>
[0486] <223> 人工序列描述:合成肽
[0487] <400> 38
[0488] Thr Ser Gly Ser Leu Ser Arg
[0489] 1 5
[0490] <210> 39
[0491] <211> 7
[0492] <212> PRT
[0493] <213> 人工序列
[0494] <220>
[0495] <223> 人工序列描述:合成肽
[0496] <400> 39
[0497] Arg Ser Asp His Leu Thr Gln
[0498] 1 5
[0499] <210> 40
[0500] <211> 7
[0501] <212> PRT
[0502] <213> 人工序列
[0503] <220>

- [0504] <223> 人工序列描述:合成肽
[0505] <400> 40
[0506] Arg Ser Asp Val Leu Ser Glu
[0507] 1 5
[0508] <210> 41
[0509] <211> 7
[0510] <212> PRT
[0511] <213> 人工序列
[0512] <220>
[0513] <223> 人工序列描述:合成肽
[0514] <400> 41
[0515] Arg Ser Ala Asp Leu Ser Arg
[0516] 1 5
[0517] <210> 42
[0518] <211> 7
[0519] <212> PRT
[0520] <213> 人工序列
[0521] <220>
[0522] <223> 人工序列描述:合成肽
[0523] <400> 42
[0524] Arg Thr Asp Ala Leu Arg Gly
[0525] 1 5
[0526] <210> 43
[0527] <211> 7
[0528] <212> PRT
[0529] <213> 人工序列
[0530] <220>
[0531] <223> 人工序列描述:合成肽
[0532] <400> 43
[0533] His Arg Ser Ala Arg Lys Arg
[0534] 1 5
[0535] <210> 44
[0536] <211> 7
[0537] <212> PRT
[0538] <213> 人工序列
[0539] <220>
[0540] <223> 人工序列描述:合成肽
[0541] <400> 44
[0542] Asp Arg Ser His Leu Thr Arg
[0543] 1 5
[0544] <210> 45
[0545] <211> 7

- [0546] <212> PRT
[0547] <213> 人工序列
[0548] <220>
[0549] <223> 人工序列描述:合成肽
[0550] <400> 45
[0551] Gln Ser Gly Asn Leu His Val
[0552] 1 5
[0553] <210> 46
[0554] <211> 7
[0555] <212> PRT
[0556] <213> 人工序列
[0557] <220>
[0558] <223> 人工序列描述:合成肽
[0559] <400> 46
[0560] Arg Ser Asp His Leu Ser Ala
[0561] 1 5
[0562] <210> 47
[0563] <211> 7
[0564] <212> PRT
[0565] <213> 人工序列
[0566] <220>
[0567] <223> 人工序列描述:合成肽
[0568] <400> 47
[0569] Arg Ser Asn Leu Leu Val Ala
[0570] 1 5
[0571] <210> 48
[0572] <211> 7
[0573] <212> PRT
[0574] <213> 人工序列
[0575] <220>
[0576] <223> 人工序列描述:合成肽
[0577] <400> 48
[0578] Gln Ser Gly Ala Leu Ala Arg
[0579] 1 5
[0580] <210> 49
[0581] <211> 7
[0582] <212> PRT
[0583] <213> 人工序列
[0584] <220>
[0585] <223> 人工序列描述:合成肽
[0586] <400> 49
[0587] Asp Arg Ser Ala Leu Ala Arg

- [0588] 1 5
[0589] <210> 50
[0590] <211> 7
[0591] <212> PRT
[0592] <213> 人工序列
[0593] <220>
[0594] <223> 人工序列描述:合成肽
[0595] <400> 50
[0596] Gln Ser Ser Asn Leu Ala Arg
[0597] 1 5
[0598] <210> 51
[0599] <211> 7
[0600] <212> PRT
[0601] <213> 人工序列
[0602] <220>
[0603] <223> 人工序列描述:合成肽
[0604] <400> 51
[0605] Gln Ser Ser Asp Leu Arg Arg
[0606] 1 5
[0607] <210> 52
[0608] <211> 7
[0609] <212> PRT
[0610] <213> 人工序列
[0611] <220>
[0612] <223> 人工序列描述:合成肽
[0613] <400> 52
[0614] Arg Ser Asp Thr Leu Ser Glu
[0615] 1 5
[0616] <210> 53
[0617] <211> 7
[0618] <212> PRT
[0619] <213> 人工序列
[0620] <220>
[0621] <223> 人工序列描述:合成肽
[0622] <400> 53
[0623] Gln Ser Gly His Leu Ser Arg
[0624] 1 5
[0625] <210> 54
[0626] <211> 7
[0627] <212> PRT
[0628] <213> 人工序列
[0629] <220>

- [0630] <223> 人工序列描述:合成肽
[0631] <400> 54
[0632] Arg Ser Asp Val Leu Ser Thr
[0633] 1 5
[0634] <210> 55
[0635] <211> 7
[0636] <212> PRT
[0637] <213> 人工序列
[0638] <220>
[0639] <223> 人工序列描述:合成肽
[0640] <400> 55
[0641] Gln Asn Ala His Arg Ile Lys
[0642] 1 5
[0643] <210> 56
[0644] <211> 7
[0645] <212> PRT
[0646] <213> 人工序列
[0647] <220>
[0648] <223> 人工序列描述:合成肽
[0649] <400> 56
[0650] Arg Ser Asp Asn Leu Ser Gln
[0651] 1 5
[0652] <210> 57
[0653] <211> 7
[0654] <212> PRT
[0655] <213> 人工序列
[0656] <220>
[0657] <223> 人工序列描述:合成肽
[0658] <400> 57
[0659] Ala Ser Asn Asp Arg Lys Lys
[0660] 1 5
[0661] <210> 58
[0662] <211> 7
[0663] <212> PRT
[0664] <213> 人工序列
[0665] <220>
[0666] <223> 人工序列描述:合成肽
[0667] <400> 58
[0668] Arg Ser Asp Asn Leu Ser Thr
[0669] 1 5
[0670] <210> 59
[0671] <211> 7

- [0672] <212> PRT
[0673] <213> 人工序列
[0674] <220>
[0675] <223> 人工序列描述:合成肽
[0676] <400> 59
[0677] Met Arg Gln His Leu Leu Asn
[0678] 1 5
[0679] <210> 60
[0680] <211> 7
[0681] <212> PRT
[0682] <213> 人工序列
[0683] <220>
[0684] <223> 人工序列描述:合成肽
[0685] <400> 60
[0686] Arg Ser Asp Asn Leu Ala Arg
[0687] 1 5
[0688] <210> 61
[0689] <211> 7
[0690] <212> PRT
[0691] <213> 人工序列
[0692] <220>
[0693] <223> 人工序列描述:合成肽
[0694] <400> 61
[0695] Gln Lys Lys Asp Arg Ser Tyr
[0696] 1 5
[0697] <210> 62
[0698] <211> 7
[0699] <212> PRT
[0700] <213> 人工序列
[0701] <220>
[0702] <223> 人工序列描述:合成肽
[0703] <400> 62
[0704] Arg Ser Asp His Leu Ser Arg
[0705] 1 5
[0706] <210> 63
[0707] <211> 7
[0708] <212> PRT
[0709] <213> 人工序列
[0710] <220>
[0711] <223> 人工序列描述:合成肽
[0712] <400> 63
[0713] Asp Arg Ser Asn Arg Lys Thr

- [0714] 1 5
[0715] <210> 64
[0716] <211> 7
[0717] <212> PRT
[0718] <213> 人工序列
[0719] <220>
[0720] <223> 人工序列描述:合成肽
[0721] <400> 64
[0722] Arg Ser Asp Thr Leu Ser Ala
[0723] 1 5
[0724] <210> 65
[0725] <211> 7
[0726] <212> PRT
[0727] <213> 人工序列
[0728] <220>
[0729] <223> 人工序列描述:合成肽
[0730] <400> 65
[0731] Asp Lys Ser Thr Arg Thr Lys
[0732] 1 5
[0733] <210> 66
[0734] <211> 7
[0735] <212> PRT
[0736] <213> 人工序列
[0737] <220>
[0738] <223> 人工序列描述:合成肽
[0739] <400> 66
[0740] Thr Ser Gly Ser Leu Thr Arg
[0741] 1 5
[0742] <210> 67
[0743] <211> 7
[0744] <212> PRT
[0745] <213> 人工序列
[0746] <220>
[0747] <223> 人工序列描述:合成肽
[0748] <400> 67
[0749] Arg Ser Asp Ser Leu Ser Ala
[0750] 1 5
[0751] <210> 68
[0752] <211> 7
[0753] <212> PRT
[0754] <213> 人工序列
[0755] <220>

- [0756] <223> 人工序列描述:合成肽
[0757] <400> 68
[0758] Arg Ser Asp Ala Leu Ala Arg
[0759] 1 5
[0760] <210> 69
[0761] <211> 7
[0762] <212> PRT
[0763] <213> 人工序列
[0764] <220>
[0765] <223> 人工序列描述:合成肽
[0766] <400> 69
[0767] Arg Arg Asp Ile Leu His Gln
[0768] 1 5
[0769] <210> 70
[0770] <211> 7
[0771] <212> PRT
[0772] <213> 人工序列
[0773] <220>
[0774] <223> 人工序列描述:合成肽
[0775] <400> 70
[0776] Gln Arg Thr His Leu Lys Ala
[0777] 1 5
[0778] <210> 71
[0779] <211> 7
[0780] <212> PRT
[0781] <213> 人工序列
[0782] <220>
[0783] <223> 人工序列描述:合成肽
[0784] <400> 71
[0785] Ala Arg Ser Thr Arg Thr Asn
[0786] 1 5
[0787] <210> 72
[0788] <211> 7
[0789] <212> PRT
[0790] <213> 人工序列
[0791] <220>
[0792] <223> 人工序列描述:合成肽
[0793] <400> 72
[0794] Gln Ser Gly Asn Leu Ala Arg
[0795] 1 5
[0796] <210> 73
[0797] <211> 7

- [0798] <212> PRT
[0799] <213> 人工序列
[0800] <220>
[0801] <223> 人工序列描述:合成肽
[0802] <400> 73
[0803] Trp Arg Ile Ser Leu Ala Ala
[0804] 1 5
[0805] <210> 74
[0806] <211> 7
[0807] <212> PRT
[0808] <213> 人工序列
[0809] <220>
[0810] <223> 人工序列描述:合成肽
[0811] <400> 74
[0812] Trp Lys Glu Ser Leu Gly Ala
[0813] 1 5
[0814] <210> 75
[0815] <211> 7
[0816] <212> PRT
[0817] <213> 人工序列
[0818] <220>
[0819] <223> 人工序列描述:合成肽
[0820] <400> 75
[0821] His Arg Lys Ser Leu Ser Arg
[0822] 1 5
[0823] <210> 76
[0824] <211> 7
[0825] <212> PRT
[0826] <213> 人工序列
[0827] <220>
[0828] <223> 人工序列描述:合成肽
[0829] <400> 76
[0830] Tyr His Trp Tyr Leu Lys Lys
[0831] 1 5
[0832] <210> 77
[0833] <211> 7
[0834] <212> PRT
[0835] <213> 人工序列
[0836] <220>
[0837] <223> 人工序列描述:合成肽
[0838] <400> 77
[0839] Thr Ser Gly His Leu Ser Arg

- [0840] 1 5
[0841] <210> 78
[0842] <211> 7
[0843] <212> PRT
[0844] <213> 人工序列
[0845] <220>
[0846] <223> 人工序列描述:合成肽
[0847] <400> 78
[0848] Thr Ser Gly Asn Leu Thr Arg
[0849] 1 5
[0850] <210> 79
[0851] <211> 7
[0852] <212> PRT
[0853] <213> 人工序列
[0854] <220>
[0855] <223> 人工序列描述:合成肽
[0856] <400> 79
[0857] Trp Trp Thr Ser Arg Ala Leu
[0858] 1 5
[0859] <210> 80
[0860] <211> 23
[0861] <212> DNA
[0862] <213> 人工序列
[0863] <220>
[0864] <223> 人工序列描述:合成引物
[0865] <400> 80
[0866] cacgagtgtg gtcaccatgc ctt 23
[0867] <210> 81
[0868] <211> 26
[0869] <212> DNA
[0870] <213> 人工序列
[0871] <220>
[0872] <223> 人工序列描述:合成引物
[0873] <400> 81
[0874] tgagtgtgac gagaagagaa acagcc 26
[0875] <210> 82
[0876] <211> 30
[0877] <212> DNA
[0878] <213> 人工序列
[0879] <220>
[0880] <223> 人工序列描述:合成探针
[0881] <400> 82

- [0882] agcaagtacc aatgggttga tgatgttg 30
[0883] <210> 83
[0884] <211> 24
[0885] <212> DNA
[0886] <213> 人工序列
[0887] <220>
[0888] <223> 人工序列描述:合成引物
[0889] <400> 83
[0890] tgcaagccac taccaccctt atgc 24
[0891] <210> 84
[0892] <211> 26
[0893] <212> DNA
[0894] <213> 人工序列
[0895] <220>
[0896] <223> 人工序列描述:合成引物
[0897] <400> 84
[0898] ggcaaagtgt gtgtgctgca aatatg 26
[0899] <210> 85
[0900] <211> 34
[0901] <212> DNA
[0902] <213> 人工序列
[0903] <220>
[0904] <223> 人工序列描述:合成探针
[0905] <400> 85
[0906] ctaaccgtga gaggcttctg atctatgtct ctga 34
[0907] <210> 86
[0908] <211> 26
[0909] <212> DNA
[0910] <213> 人工序列
[0911] <220>
[0912] <223> 人工序列描述:合成引物
[0913] <400> 86
[0914] tgagtgtgat gagaagagaa gcagcc 26
[0915] <210> 87
[0916] <211> 30
[0917] <212> DNA
[0918] <213> 人工序列
[0919] <220>
[0920] <223> 人工序列描述:合成探针
[0921] <400> 87
[0922] agcaagtacc catgggttga tgatgtttag 30
[0923] <210> 88

- [0924] <211> 26
- [0925] <212> DNA
- [0926] <213> 人工序列
- [0927] <220>
- [0928] <223> 人工序列描述:合成引物
- [0929] <400> 88
- [0930] ttggtttggc tgctatgtgt ttatgg 26
- [0931] <210> 89
- [0932] <211> 28
- [0933] <212> DNA
- [0934] <213> 人工序列
- [0935] <220>
- [0936] <223> 人工序列描述:合成引物
- [0937] <400> 89
- [0938] tgtggcattt tagagaagag atggtag 28
- [0939] <210> 90
- [0940] <211> 32
- [0941] <212> DNA
- [0942] <213> 人工序列
- [0943] <220>
- [0944] <223> 人工序列描述:合成探针
- [0945] <400> 90
- [0946] agggagcttt ggcaactatg gacagagatt at 32
- [0947] <210> 91
- [0948] <211> 26
- [0949] <212> DNA
- [0950] <213> 人工序列
- [0951] <220>
- [0952] <223> 人工序列描述:合成引物
- [0953] <400> 91
- [0954] agccttcaat gtctctggca gaccct 26
- [0955] <210> 92
- [0956] <211> 26
- [0957] <212> DNA
- [0958] <213> 人工序列
- [0959] <220>
- [0960] <223> 人工序列描述:合成引物
- [0961] <400> 92
- [0962] ggcatagtgt gtgtgctgca gatatg 26
- [0963] <210> 93
- [0964] <211> 34
- [0965] <212> DNA

- [0966] <213> 人工序列
- [0967] <220>
- [0968] <223> 人工序列描述:合成探针
- [0969] <400> 93
- [0970] caaatcgtga gaggctttg atctatgtct ctga 34
- [0971] <210> 94
- [0972] <400> 94
- [0973] 000
- [0974] <210> 95
- [0975] <400> 95
- [0976] 000
- [0977] <210> 96
- [0978] <400> 96
- [0979] 000
- [0980] <210> 97
- [0981] <400> 97
- [0982] 000
- [0983] <210> 98
- [0984] <400> 98
- [0985] 000
- [0986] <210> 99
- [0987] <400> 99
- [0988] 000
- [0989] <210> 100
- [0990] <400> 100
- [0991] 000
- [0992] <210> 101
- [0993] <400> 101
- [0994] 000
- [0995] <210> 102
- [0996] <400> 102
- [0997] 000
- [0998] <210> 103
- [0999] <400> 103
- [1000] 000
- [1001] <210> 104
- [1002] <400> 104
- [1003] 000
- [1004] <210> 105
- [1005] <400> 105
- [1006] 000
- [1007] <210> 106

- [1008] <400> 106
- [1009] 000
- [1010] <210> 107
- [1011] <400> 107
- [1012] 000
- [1013] <210> 108
- [1014] <400> 108
- [1015] 000
- [1016] <210> 109
- [1017] <400> 109
- [1018] 000
- [1019] <210> 110
- [1020] <400> 110
- [1021] 000
- [1022] <210> 111
- [1023] <400> 111
- [1024] 000
- [1025] <210> 112
- [1026] <400> 112
- [1027] 000
- [1028] <210> 113
- [1029] <400> 113
- [1030] 000
- [1031] <210> 114
- [1032] <400> 114
- [1033] 000
- [1034] <210> 115
- [1035] <400> 115
- [1036] 000
- [1037] <210> 116
- [1038] <400> 116
- [1039] 000
- [1040] <210> 117
- [1041] <400> 117
- [1042] 000
- [1043] <210> 118
- [1044] <400> 118
- [1045] 000
- [1046] <210> 119
- [1047] <211> 62
- [1048] <212> DNA
- [1049] <213> 人工序列

- [1050] <220>
- [1051] <223> 人工序列描述:合成寡核苷酸
- [1052] <400> 119
- [1053] ttactctctc taccgtgttgc aaccctgaa atttcagggt tgcaacacgg tagagagagt 60
- [1054] aa 62
- [1055] <210> 120
- [1056] <211> 58
- [1057] <212> DNA
- [1058] <213> 人工序列
- [1059] <220>
- [1060] <223> 人工序列描述:合成寡核苷酸
- [1061] <400> 120
- [1062] ttactctctc taccgtgttgc aaccctgat cagggttgca acacggtaga gagagtaa 58
- [1063] <210> 121
- [1064] <211> 57
- [1065] <212> DNA
- [1066] <213> 人工序列
- [1067] <220>
- [1068] <223> 人工序列描述:合成寡核苷酸
- [1069] <400> 121
- [1070] ttactctctc taccgtgttgc aaccctgaa agggttgcaa cacggtagag agagtaa 57
- [1071] <210> 122
- [1072] <211> 54
- [1073] <212> DNA
- [1074] <213> 人工序列
- [1075] <220>
- [1076] <223> 人工序列描述:合成寡核苷酸
- [1077] <400> 122
- [1078] ttactctctc taccgtgttgc aaccctgag gttgcaacac ggttagagaga gtaa 54
- [1079] <210> 123
- [1080] <211> 30
- [1081] <212> DNA
- [1082] <213> 人工序列
- [1083] <220>
- [1084] <223> 人工序列描述:合成寡核苷酸
- [1085] <400> 123
- [1086] ttactctctc taccgtgttgc aaccctgaa 30
- [1087] <210> 124
- [1088] <211> 60
- [1089] <212> DNA
- [1090] <213> 人工序列
- [1091] <220>

- [1092] <223> 人工序列描述:合成寡核苷酸
- [1093] <400> 124
- [1094] ttactctctc taccgtgttgc caaccctgat ttcagggttg caacacggta gagagagtaa 60
- [1095] <210> 125
- [1096] <211> 59
- [1097] <212> DNA
- [1098] <213> 人工序列
- [1099] <220>
- [1100] <223> 人工序列描述:合成寡核苷酸
- [1101] <400> 125
- [1102] ttactctctc taccgtgttgc caaccctgat tcagggttgc aacacggtag agagagtaa 59
- [1103] <210> 126
- [1104] <211> 58
- [1105] <212> DNA
- [1106] <213> 人工序列
- [1107] <220>
- [1108] <223> 人工序列描述:合成寡核苷酸
- [1109] <400> 126
- [1110] ttactctctc taccgtgttgc caaccctatt cagggttgca acacggtaga gagagtaa 58
- [1111] <210> 127
- [1112] <211> 59
- [1113] <212> DNA
- [1114] <213> 人工序列
- [1115] <220>
- [1116] <223> 人工序列描述:合成寡核苷酸
- [1117] <400> 127
- [1118] ttactctctc taccgtgttgc caaccctgat tcagggttgc aacacggtag agagagtaa 59
- [1119] <210> 128
- [1120] <211> 59
- [1121] <212> DNA
- [1122] <213> 人工序列
- [1123] <220>
- [1124] <223> 人工序列描述:合成寡核苷酸
- [1125] <400> 128
- [1126] ttactctctc taccgtgttgc caaccctgaa tcagggttgc aacacggtag agagagtaa 59
- [1127] <210> 129
- [1128] <211> 57
- [1129] <212> DNA
- [1130] <213> 人工序列
- [1131] <220>
- [1132] <223> 人工序列描述:合成寡核苷酸
- [1133] <400> 129

- [1134] ttactctc taccgtttg caaccctgaa agggttgcaa cacggtagag agagtaa 57
- [1135] <210> 130
- [1136] <211> 59
- [1137] <212> DNA
- [1138] <213> 人工序列
- [1139] <220>
- [1140] <223> 人工序列描述:合成寡核苷酸
- [1141] <400> 130
- [1142] ttactctc taccgtttg caacccttt tcagggttgc aacacggtag agagagtaa 59
- [1143] 18

		1	50
2.3 Jack	(1)	GGAAGAAGCCTCTCTCAAGGGTCCAAACACAAAGCCACCATTCACTGTT	
2.3 Mav	(1)	GGAAGAAGCCTCTCTCAAGGGTCCAAACACAAAGCCACCATTCACTGTT	
2.3 W82	(1)	GGAAGAAGCCTCTCTCAAGGGTCCAAACACAAAGCCACCATTCACTGTT	
2.3 Wes	(1)	GGAAGAAGCCTCTCTCAAGGGTCCAAACACAAAGCCACCATTCACTGTT	
2.3 X5	(1)	GGAAGAAGCCTCTCTCAAGGGTCCAAACACAAAGCCACCATTCACTGTT	
	51		100
2.3 Jack	(51)	GGCCAACCTCAAGAAAGCAATTCCACCACACTGCTTCAGCGCTCCCTCCT	
2.3 Mav	(51)	GGCCAACCTCAAGAAAGCAATTCCACCACACTGCTTCAGCGCTCCCTCCT	
2.3 W82	(51)	GGCCAACCTCAAGAAAGCAATTCCACCACACTGCTTCAGCGCTCCCTCCT	
2.3 Wes	(51)	GGCCAACCTCAAGAAAGCAATTCCACCACACTGCTTCAGCGCTCCCTCCT	
2.3 X5	(51)	GGCCAACCTCAAGAAAGCAATTCCACCACACTGCTTCAGCGCTCCCTCCT	
	101		150
2.3 Jack	(101)	CACTTCATTCTCCTATGTTGTTATGACCTTCATTCGCCTTCATTTCT	
2.3 Mav	(101)	CACTTCATTCTCCTATGTTGTTATGACCTTCATTCGCCTTCATTTCT	
2.3 W82	(101)	CACTTCATTCTCCTATGTTGTTATGACCTTCATTCGCCTTCATTTCT	
2.3 Wes	(101)	CACTTCATTCTCCTATGTTGTTATGACCTTCATTCGCCTTCATTTCT	
2.3 X5	(101)	CACTTCATTCTCCTATGTTGTTATGACCTTCATTCGCCTTCATTTCT	
	151		200
2.3 Jack	(151)	ACATTGCCACCACCTACTTCCACCTCCTCCTCAACCCCTTTCCCTCATT	
2.3 Mav	(151)	ACATTGCCACCACCTACTTCCACCTCCTCCTCAACCCCTTTCCCTCATT	
2.3 W82	(151)	ACATTGCCACCACCTACTTCCACCTCCTCCTCAACCCCTTTCCCTCATT	
2.3 Wes	(151)	ACATTGCCACCACCTACTTCCACCTCCTCCTCAACCCCTTTCCCTCATT	
2.3 X5	(151)	ACATTGCCACCACCTACTTCCACCTCCTCCTCAACCCCTTTCCCTCATT	
	201		250
2.3 Jack	(201)	GCATGGCCAATCTATTGGGTTCTCAAGGGTGCCTTCACTGGTGTGTG	
2.3 Mav	(201)	GCATGGCCAATCTATTGGGTTCTCAAGGGTGCCTTCACTGGTGTGTG	
2.3 W82	(201)	GCATGGCCAATCTATTGGGTTCTCAAGGGTGCCTTCACTGGTGTGTG	
2.3 Wes	(201)	GCATGGCCAATCTATTGGGTTCTCAAGGGTGCCTTCACTGGTGTGTG	
2.3 X5	(201)	GCATGGCCAATCTATTGGGTTCTCAAGGGTGCCTTCACTGGTGTGTG	
	251		300
2.3 Jack	(251)	GGTGATTGCTCACGAGTGTGGTCACCATGCCTTCAGCAAGTACCAATGGG	
2.3 Mav	(251)	GGTGATTGCTCACGAGTGTGGTCACCATGCCTTCAGCAAGTACCAATGGG	
2.3 W82	(251)	GGTGATTGCTCACGAGTGTGGTCACCATGCCTTCAGCAAGTACCAATGGG	
2.3 Wes	(251)	GGTGATTGCTCACGAGTGTGGTCACCATGCCTTCAGCAAGTACCAATGGG	
2.3 X5	(251)	GGTGATTGCTCACGAGTGTGGTCACCATGCCTTCAGCAAGTACCAATGGG	
	301		350
2.3 Jack	(301)	TTGATGATGTTGTGGGTTGACCCCTCACTCAACACTTTAGTCCCTTAT	
2.3 Mav	(301)	TTGATGATGTTGTGGGTTGACCCCTCACTCAACACTTTAGTCCCTTAT	
2.3 W82	(301)	TTGATGATGTTGTGGGTTGACCCCTCACTCAACACTTTAGTCCCTTAT	
2.3 Wes	(301)	TTGATGATGTTGTGGGTTGACCCCTCACTCAACACTTTAGTCCCTTAT	
2.3 X5	(301)	TTGATGATGTTGTGGGTTGACCCCTCACTCAACACTTTAGTCCCTTAT	
	351		400
2.3 Jack	(351)	TTCTCATGGAAAATAAGCCATGCCGCCATCACTCCAACACAGGTTCCCT	
2.3 Mav	(351)	TTCTCATGGAAAATAAGCCATGCCGCCATCACTCCAACACAGGTTCCCT	
2.3 W82	(351)	TTCTCATGGAAAATAAGCCATGCCGCCATCACTCCAACACAGGTTCCCT	
2.3 Wes	(351)	TTCTCATGGAAAATAAGCCATGCCGCCATCACTCCAACACAGGTTCCCT	
2.3 X5	(351)	TTCTCATGGAAAATAAGCCATGCCGCCATCACTCCAACACAGGTTCCCT	
	401		450
2.3 Jack	(401)	TGACCGTGATGAAGTGTGTCACCATGCCAAACACAGGTTCCCT	
2.3 Mav	(401)	TGACCGTGATGAAGTGTGTCACCATGCCAAACACAGGTTCCCT	
2.3 W82	(401)	TGACCGTGATGAAGTGTGTCACCATGCCAAACACAGGTTCCCT	

图1

2.3 Wes	(401)	TGACCGTGTGAAGTGTGGCTCCAAAACCAAAATCCAAAGTGCATGGT	
2.3 X5	(401)	TGACCGTGTGAAGTGTGGCTCCAAAACCAAAATCCAAAGTGCATGGT	
	451		500
2.3 Jack	(451)	TTTCCAAGTACTTAAACAACCCCTCTAGGAAGGGCTGTTCTCTCGTC	
2.3 Mav	(451)	TTTCCAAGTACTTAAACAACCCCTCTAGGAAGGGCTGTTCTCTCGTC	
2.3 W82	(451)	TTTCCAAGTACTTAAACAACCCCTCTAGGAAGGGCTGTTCTCTCGTC	
2.3 Wes	(451)	TTTCCAAGTACTTAAACAACCCCTCTAGGAAGGGCTGTTCTCTCGTC	
2.3 X5	(451)	TTTCCAAGTACTTAAACAACCCCTCTAGGAAGGGCTGTTCTCTCGTC	
	501		550
2.3 Jack	(501)	ACACTCACAAATAGGGTGGCCTATGTATTTAGCCTTCAATGTCTCTGGTAG	
2.3 Mav	(501)	ACACTCACAAATAGGGTGGCCTATGTATTTAGCCTTCAATGTCTCTGGTAG	
2.3 W82	(501)	ACACTCACAAATAGGGTGGCCTATGTATTTAGCCTTCAATGTCTCTGGTAG	
2.3 Wes	(501)	ACACTCACAAATAGGGTGGCCTATGTATTTAGCCTTCAATGTCTCTGGTAG	
2.3 X5	(501)	ACACTCACAAATAGGGTGGCCTATGTATTTAGCCTTCAATGTCTCTGGTAG	
	551		600
2.3 Jack	(551)	ACCTATGATAGTTGCAAGCCACTACCACCCCTATGCTCCATATATT	
2.3 Mav	(551)	ACCTATGATAGTTGCAAGCCACTACCACCCCTATGCTCCATATATT	
2.3 W82	(551)	ACCTATGATAGTTGCAAGCCACTACCACCCCTATGCTCCATATATT	
2.3 Wes	(551)	ACCTATGATAGTTGCAAGCCACTACCACCCCTATGCTCCATATATT	
2.3 X5	(551)	ACCTATGATAGTTGCAAGCCACTACCACCCCTATGCTCCATATATT	
	601		650
2.3 Jack	(601)	CTAACCGTGAGAGGCTCTGATCTATGTCTGTGATGTTGCTTGTGTTCT	
2.3 Mav	(601)	CTAACCGTGAGAGGCTCTGATCTATGTCTGTGATGTTGCTTGTGTTCT	
2.3 W82	(601)	CTAACCGTGAGAGGCTCTGATCTATGTCTGTGATGTTGCTTGTGTTCT	
2.3 Wes	(601)	CTAACCGTGAGAGGCTCTGATCTATGTCTGTGATGTTGCTTGTGTTCT	
2.3 X5	(601)	CTAACCGTGAGAGGCTCTGATCTATGTCTGTGATGTTGCTTGTGTTCT	
	651		700
2.3 Jack	(651)	GTGACTTACTCTCTACCGTGTGCAACCCCTGAAAGGGTTGGTTGGCT	
2.3 Mav	(651)	GTGACTTACTCTCTACCGTGTGCAACCCCTGAAAGGGTTGGTTGGCT	
2.3 W82	(651)	GTGACTTACTCTCTACCGTGTGCAACCCCTGAAAGGGTTGGTTGGCT	
2.3 Wes	(651)	GTGACTTACTCTCTACCGTGTGCAACCCCTGAAAGGGTTGGTTGGCT	
2.3 X5	(651)	GTGACTTACTCTCTACCGTGTGCAACCCCTGAAAGGGTTGGTTGGCT	
	701		750
2.3 Jack	(701)	GCTATGTGTTATGGGGTGCCTTGCTCATTGTGAACGGTTTCTTGTGA	
2.3 Mav	(701)	GCTATGTGTTATGGGGTGCCTTGCTCATTGTGAACGGTTTCTTGTGA	
2.3 W82	(701)	GCTATGTGTTATGGGGTGCCTTGCTCATTGTGAACGGTTTCTTGTGA	
2.3 Wes	(701)	GCTATGTGTTATGGGGTGCCTTGCTCATTGTGAACGGTTTCTTGTGA	
2.3 X5	(701)	GCTATGTGTTATGGGGTGCCTTGCTCATTGTGAACGGTTTCTTGTGA	
	751		800
2.3 Jack	(751)	CTATCACATATTCGACAGCACACACTTGCCTGCCTCATTACGATTCA	
2.3 Mav	(751)	CTATCACATATTCGACAGCACACACTTGCCTGCCTCATTACGATTCA	
2.3 W82	(751)	CTATCACATATTCGACAGCACACACTTGCCTGCCTCATTACGATTCA	
2.3 Wes	(751)	CTATCACATATTCGACAGCACACACTTGCCTGCCTCATTACGATTCA	
2.3 X5	(751)	CTATCACATATTCGACAGCACACACTTGCCTGCCTCATTACGATTCA	
	801		850
2.3 Jack	(801)	TCAGAATGGACTGGCTGAAGGGAGCTTGGCAACTATGGACAGAGATTA	
2.3 Mav	(801)	TCAGAATGGACTGGCTGAAGGGAGCTTGGCAACTATGGACAGAGATTA	
2.3 W82	(801)	TCAGAATGGACTGGCTGAAGGGAGCTTGGCAACTATGGACAGAGATTA	
2.3 Wes	(801)	TCAGAATGGACTGGCTGAAGGGAGCTTGGCAACTATGGACAGAGATTA	
2.3 X5	(801)	TCAGAATGGACTGGCTGAAGGGAGCTTGGCAACTATGGACAGAGATTA	
	851		900
2.3 Jack	(851)	TGGGATTCTGAACAAGGTGTTCATCACATAACTGATACTCATGTGGCTC	
2.3 Mav	(851)	TGGGATTCTGAACAAGGTGTTCATCACATAACTGATACTCATGTGGCTC	
2.3 W82	(851)	TGGGATTCTGAACAAGGTGTTCATCACATAACTGATACTCATGTGGCTC	
2.3 Wes	(851)	TGGGATTCTGAACAAGGTGTTCATCACATAACTGATACTCATGTGGCTC	
2.3 X5	(851)	TGGGATTCTGAACAAGGTGTTCATCACATAACTGATACTCATGTGGCTC	
	901		950

图1(续)

2.3 Jack	(901)	ACCATCTCTCTACAATGCCACATTACCATGCAATGGAGGCAACCAAT	
2.3 Mav	(901)	ACCATCTCTCTACAATGCCACATTACCATGCAATGGAGGCAACCAAT	
2.3 W82	(901)	ACCATCTCTCTACAATGCCACATTACCATGCAATGGAGGCAACCAAT	
2.3 Wes	(901)	ACCATCTCTCTACAATGCCACATTACCATGCAATGGAGGCAACCAAT	
2.3 X5	(901)	ACCATCTCTCTACAATGCCACATTACCATGCAATGGAGGCAACCAAT	951
			1000
2.3 Jack	(951)	GCAATCAAGCCAATATTGGGTGAGTACTACCAATTGATGACACACCATT	
2.3 Mav	(951)	GCAATCAAGCCAATATTGGGTGAGTACTACCAATTGATGACACACCATT	
2.3 W82	(951)	GCAATCAAGCCAATATTGGGTGAGTACTACCAATTGATGACACACCATT	
2.3 Wes	(951)	GCAATCAAGCCAATATTGGGTGAGTACTACCAATTGATGACACACCATT	
2.3 X5	(951)	GCAATCAAGCCAATATTGGGTGAGTACTACCAATTGATGACACACCATT	1001
			1050
2.3 Jack	(1001)	TTACAAGGCACTGTGGAGAGAACGAGAGAGTCCTCTATGTGGAGCCAG	
2.3 Mav	(1001)	TTACAAGGCACTGTGGAGAGAACGAGAGAGTCCTCTATGTGGAGCCAG	
2.3 W82	(1001)	TTACAAGGCACTGTGGAGAGAACGAGAGAGTCCTCTATGTGGAGCCAG	
2.3 Wes	(1001)	TTACAAGGCACTGTGGAGAGAACGAGAGAGTCCTCTATGTGGAGCCAG	
2.3 X5	(1001)	TTACAAGGCACTGTGGAGAGAACGAGAGAGTCCTCTATGTGGAGCCAG	1051
			1100
2.3 Jack	(1051)	ATGAAGGAACATCCGAGAACGGCGTGTATTGGTACAGGAACAAGTATTGA	
2.3 Mav	(1051)	ATGAAGGAACATCCGAGAACGGCGTGTATTGGTACAGGAACAAGTATTGA	
2.3 W82	(1051)	ATGAAGGAACATCCGAGAACGGCGTGTATTGGTACAGGAACAAGTATTGA	
2.3 Wes	(1051)	ATGAAGGAACATCCGAGAACGGCGTGTATTGGTACAGGAACAAGTATTGA	
2.3 X5	(1051)	ATGAAGGAACATCCGAGAACGGCGTGTATTGGTACAGGAACAAGTATTGA	

图1 (续)

		1	50
2.6 Jack	(1)	CAGAAGAACGCCTCTCAAGGGTCAAACACAAAGCCACCATTCACTGT	
2.6 Mav	(1)	CAGAAGAACGCCTCTCAAGGGTCAAACACAAAGCCACCATTCACTGT	
2.6 W82	(1)	CAGAAGAACGCCTCTCAAGGGTCAAACACAAAGCCACCATTCACTGT	
2.6 Wes	(1)	CAGAAGAACGCCTCTCAAGGGTCAAACACAAAGCCACCATTCACTGT	
2.6 X5	(1)	CAGAAGAACGCCTCTCAAGGGTCAAACACAAAGCCACCATTCACTGT	
	51		100
2.6 Jack	(51)	TGGCCAACCTCAAGAAAGCCATTCCACCGCACTGCTTCAGCGTTCCTCC	
2.6 Mav	(51)	TGGCCAACCTCAAGAAAGCCATTCCACCGCACTGCTTCAGCGTTCCTCC	
2.6 W82	(51)	TGGCCAACCTCAAGAAAGCCATTCCACCGCACTGCTTCAGCGTTCCTCC	
2.6 Wes	(51)	TGGCCAACCTCAAGAAAGCCATTCCACCGCACTGCTTCAGCGTTCCTCC	
2.6 X5	(51)	TGGCCAACCTCAAGAAAGCCATTCCACCGCACTGCTTCAGCGTTCCTCC	
	101		150
2.6 Jack	(101)	TCACTTCATTGCTCTATGTTGTTATGACCTTTATTGGCTTTCATTTTC	
2.6 Mav	(101)	TCACTTCATTGCTCTATGTTGTTATGACCTTTATTGGCTTTCATTTTC	
2.6 W82	(101)	TCACTTCATTGCTCTATGTTGTTATGACCTTTATTGGCTTTCATTTTC	
2.6 Wes	(101)	TCACTTCATTGCTCTATGTTGTTATGACCTTTATTGGCTTTCATTTTC	
2.6 X5	(101)	TCACTTCATTGCTCTATGTTGTTATGACCTTTATTGGCTTTCATTTTC	
	151		200
2.6 Jack	(151)	TACATTGCCACCACCTACTTCCACCTCCCTCACCCTTTCCCTCAT	
2.6 Mav	(151)	TACATTGCCACCACCTACTTCCACCTCCCTCACCCTTTCCCTCAT	
2.6 W82	(151)	TACATTGCCACCACCTACTTCCACCTCCCTCACCCTTTCCCTCAT	
2.6 Wes	(151)	TACATTGCCACCACCTACTTCCACCTCCCTCACCCTTTCCCTCAT	
2.6 X5	(151)	TACATTGCCACCACCTACTTCCACCTCCCTCACCCTTTCCCTCAT	
	201		250
2.6 Jack	(201)	TGCATGGCCAATCTATTGGGTTCTCCAAGGGTGCATTCTACTGGCGTGT	
2.6 Mav	(201)	TGCATGGCCAATCTATTGGGTTCTCCAAGGGTGCATTCTACTGGCGTGT	
2.6 W82	(201)	TGCATGGCCAATCTATTGGGTTCTCCAAGGGTGCATTCTACTGGCGTGT	
2.6 Wes	(201)	TGCATGGCCAATCTATTGGGTTCTCCAAGGGTGCATTCTACTGGCGTGT	
2.6 X5	(201)	TGCATGGCCAATCTATTGGGTTCTCCAAGGGTGCATTCTACTGGCGTGT	
	251		300
2.6 Jack	(251)	GGGTGATTGCTCACGAGTGTGGTCACCATGCCTTCAGCAAGTACCCATGG	
2.6 Mav	(251)	GGGTGATTGCTCACGAGTGTGGTCACCATGCCTTCAGCAAGTACCCATGG	
2.6 W82	(251)	GGGTGATTGCTCACGAGTGTGGTCACCATGCCTTCAGCAAGTACCCATGG	
2.6 Wes	(251)	GGGTGATTGCTCACGAGTGTGGTCACCATGCCTTCAGCAAGTACCCATGG	
2.6 X5	(251)	GGGTGATTGCTCACGAGTGTGGTCACCATGCCTTCAGCAAGTACCCATGG	
	301		350
2.6 Jack	(301)	GTTGATGATGTTATGGGTTGACCGTTCACTCAGCACTTTAGTCCCTTA	
2.6 Mav	(301)	GTTGATGATGTTGTGGGTTGACCGTTCACTCAGCACTTTAGTCCCTTA	
2.6 W82	(301)	GTTGATGATGTTATGGGTTGACCGTTCACTCAGCACTTTAGTCCCTTA	
2.6 Wes	(301)	GTTGATGATGTTATGGGTTGACCGTTCACTCAGCACTTTAGTCCCTTA	
2.6 X5	(301)	GTTGATGATGTTGTGGGTTGACCGTTCACTCAGCACTTTAGTCCCTTA	
	351		400
2.6 Jack	(351)	TTTCTCATGGAAAATAAGCCATCGCGCCACCACTCCAACACGGGTTCCC	
2.6 Mav	(351)	TTTCTCATGGAAAATAAGCCATCGCGCCACCACTCCAACACGGGTTCCC	
2.6 W82	(351)	TTTCTCATGGAAAATAAGCCATCGCGCCACCACTCCAACACGGGTTCCC	
2.6 Wes	(351)	TTTCTCATGGAAAATAAGCCATCGCGCCACCACTCCAACACGGGTTCCC	
2.6 X5	(351)	TTTCTCATGGAAAATAAGCCATCGCGCCACCACTCCAACACGGGTTCCC	
	401		450
2.6 Jack	(401)	TTGACCGTGATGAAGTGTGTCACCAACACGGGTTCCC	
2.6 Mav	(401)	TTGACCGTGATGAAGTGTGTCACCAACACGGGTTCCC	
2.6 W82	(401)	TTGACCGTGATGAAGTGTGTCACCAACACGGGTTCCC	

图2

2.6 Wes	(401)	TTGACCGTGATGAAGTGTGTCACAAACCAAAATCCAAAGTGCATGG	
2.6 X5	(401)	TTGACCGTGATGAAGTGTGTCACAAACCAAAATCCAAAGTGCATGG	
	451		500
2.6 Jack	(451)	TACACCAAGTACCTGAACAACCCCTCTAGGAAGGGCTGCTCTCTCAT	
2.6 Mav	(451)	TACACCAAGTACCTGAACAACCCCTCTAGGAAGGGCTGCTCTCTCAT	
2.6 W82	(451)	TACACCAAGTACCTGAACAACCCCTCTAGGAAGGGCTGCTCTCTCAT	
2.6 Wes	(451)	TACACCAAGTACCTGAACAACCCCTCTAGGAAGGGCTGCTCTCTCAT	
2.6 X5	(451)	TACACCAAGTACCTGAACAACCCCTCTAGGAAGGGCTGCTCTCTCAT	
	501		550
2.6 Jack	(501)	CACACTCACAATAGGGTGGCCTTGTATTCAGCCTCAATGTCTCTGGCA	
2.6 Mav	(501)	CACACTCACAATAGGGTGGCCTTGTATTCAGCCTCAATGTCTCTGGCA	
2.6 W82	(501)	CACACTCACAATAGGGTGGCCTTGTATTCAGCCTCAATGTCTCTGGCA	
2.6 Wes	(501)	CACACTCACAATAGGGTGGCCTTGTATTCAGCCTCAATGTCTCTGGCA	
2.6 X5	(501)	CACACTCACAATAGGGTGGCCTTGTATTCAGCCTCAATGTCTCTGGCA	
	551		600
2.6 Jack	(551)	GACCCTATGATGGTTTGCTAGCCACTACCACCCCTATGCTCCCATATAT	
2.6 Mav	(551)	GACCCTATGATGGTTTGCTAGCCACTACCACCCCTATGCTCCCATATAT	
2.6 W82	(551)	GACCCTATGATGGTTTGCTAGCCACTACCACCCCTATGCTCCCATATAT	
2.6 Wes	(551)	GACCCTATGATGGTTTGCTAGCCACTACCACCCCTATGCTCCCATATAT	
2.6 X5	(551)	GACCCTATGATGGTTTGCTAGCCACTACCACCCCTATGCTCCCATATAT	
	601		650
2.6 Jack	(601)	TCAAATCGTGAGAGGCCTTGTATCTATGTCTGATGTTGCTTGTGTTTC	
2.6 Mav	(601)	TCAAATCGTGAGAGGCCTTGTATCTATGTCTGATGTTGCTTGTGTTTC	
2.6 W82	(601)	TCAAATCGTGAGAGGCCTTGTATCTATGTCTGATGTTGCTTGTGTTTC	
2.6 Wes	(601)	TCAAATCGTGAGAGGCCTTGTATCTATGTCTGATGTTGCTTGTGTTTC	
2.6 X5	(601)	TCAAACCGTGAGAGGCCTCTGATCTATGTCTGATGTTGCTTGTGTTTC	
	651		700
2.6 Jack	(651)	TGTGACTTACTTGCTCTACCGTGTTGCAACTATGAAAGGGTTGGTTGGC	
2.6 Mav	(651)	TGTGACTTACTTGCTCTACCGTGTTGCAACTATGAAAGGGTTGGTTGGC	
2.6 W82	(651)	TGTGACTTACTTGCTCTACCGTGTTGCAACTATGAAAGGGTTGGTTGGC	
2.6 Wes	(651)	TGTGACTTACTTGCTCTACCGTGTTGCAACTATGAAAGGGTTGGTTGGC	
2.6 X5	(651)	TGTGACTTACTTGCTCTACCGTGTTGCAACTATGAAAGGGTTGGTTGGC	
	701		750
2.6 Jack	(701)	TGCTATGTGTTATGGGTGCCATTGCTCATTGTGAACGGTTTCTTGTG	
2.6 Mav	(701)	TGCTATGTGTTATGGGTGCCATTGCTCATTGTGAACGGTTTCTTGTG	
2.6 W82	(701)	TGCTATGTGTTATGGGTGCCATTGCTCATTGTGAACGGTTTCTTGTG	
2.6 Wes	(701)	TGCTATGTGTTATGGGTGCCATTGCTCATTGTGAACGGTTTCTTGTG	
2.6 X5	(701)	TGCTATGTGTTATGGGTGCCATTGCTCATTGTGAACGGTTTCTTGTG	
	751		800
2.6 Jack	(751)	ACCATCACATATCTGCAGCACACACACTATGCCTTGCCCTCACTATGATT	
2.6 Mav	(751)	ACCATCACATATCTGCAGCACACACACTATGCCTTGCCCTCACTATGATT	
2.6 W82	(751)	ACCATCACATATCTGCAGCACACACACTATGCCTTGCCCTCACTATGATT	
2.6 Wes	(751)	ACCATCACATATCTGCAGCACACACACTATGCCTTGCCCTCACTATGATT	
2.6 X5	(751)	ACCATCACATATCTGCAGCACACACACTATGCCTTGCCCTCACTATGATT	
	801		850
2.6 Jack	(801)	ATCAGAATGGGATTGGCTGAGGGGTGCTTGGCAACTATGGACAGAGATT	
2.6 Mav	(801)	ATCAGAATGGGATTGGCTGAGGGGTGCTTGGCAACTATGGACAGAGATT	
2.6 W82	(801)	ATCAGAATGGGATTGGCTGAGGGGTGCTTGGCAACTATGGACAGAGATT	
2.6 Wes	(801)	ATCAGAATGGGATTGGCTGAGGGGTGCTTGGCAACTATGGACAGAGATT	
2.6 X5	(801)	ATCAGAATGGGATTGGCTGAGGGGTGCTTGGCAACTATGGACAGAGATT	
	851		900
2.6 Jack	(851)	ATGGAAATTCTGAACAAGGTGTTCACCACTACTGATACTCATGTGGCT	
2.6 Mav	(851)	ATGGGATTCTGAACAAGGTGTTCACCACTACTGATACTCATGTGGCT	
2.6 W82	(851)	ATGGAAATTCTGAACAAGGTGTTCACCACTACTGATACTCATGTGGCT	
2.6 Wes	(851)	ATGGAAATTCTGAACAAGGTGTTCACCACTACTGATACTCATGTGGCT	
2.6 X5	(851)	ATGGGATTCTGAACAAGGTGTTCACCACTACTGATACTCATGTGGCT	
	901		950

图2(续)

2.6 Jack	(901)	CACCATCTTCTACAATGCCACATTACCATGCAACGGAGGCAACCAA
2.6 Mav	(901)	CACCATCTTCTACAATGCCACATTACCATGCAACGGAGGCAACCAA
2.6 W82	(901)	CACCATCTTCTACAATGCCACATTACCATGCAACGGAGGCAACCAA
2.6 Wes	(901)	CACCATCTTCTACAATGCCACATTACCATGCAACGGAGGCAACCAA
2.6 X5	(901)	CACCATCTTCTACAATGCCACATTACCATGCAACGGAGGCAACCAA 951
2.6 Jack	(951)	TGCAATGAAGCCAATATTGGGTGAGTACTACCGATTGATGACACACCAT
2.6 Mav	(951)	TGCAATGAAGCCAATATTGGGTGAGTACTACCGATTGATGACACACCAT
2.6 W82	(951)	TGCAATGAAGCCAATATTGGGTGAGTACTACCGATTGATGACACACCAT
2.6 Wes	(951)	TGCAATGAAGCCAATATTGGGTGAGTACTACCGATTGATGACACACCAT
2.6 X5	(951)	TGCAATGAAGCCAATATTGGGTGAGTACTACCGATTGATGACACACCAT 1001
2.6 Jack	(1001)	TTTACAAGGCACTGTGGAGAGAACAGCAAGAGAGTGCCTCTATGTGGAGCCA
2.6 Mav	(1001)	TTTACAAGGCACTGTGGAGAGAACAGCAAGAGAGTGCCTCTATGTGGAGCCA
2.6 W82	(1001)	TTTACAAGGCACTGTGGAGAGAACAGCAAGAGAGTGCCTCTATGTGGAGCCA
2.6 Wes	(1001)	TTTACAAGGCACTGTGGAGAGAACAGCAAGAGAGTGCCTCTATGTGGAGCCA
2.6 X5	(1001)	TTTACAAGGCACTGTGGAGAGAACAGCAAGAGAGTGCCTCTATGTGGAGCCA 1051
2.6 Jack	(1051)	GATGAAGGAACATCCGAGAAGGGCGTGTATTGGTACAGGAACAAGTATTG
2.6 Mav	(1051)	GATGAAGGAACATCCGAGAAGGGCGTGTATTGGTACAGGAACAAGTATTG
2.6 W82	(1051)	GATGAAGGAACATCCGAGAAGGGCGTGTATTGGTACAGGAACAAGTATTG
2.6 Wes	(1051)	GATGAAGGAACATCCGAGAAGGGCGTGTATTGGTACAGGAACAAGTATTG
2.6 X5	(1051)	GATGAAGGAACATCCGAGAAGGGCGTGTATTGGTACAGGAACAAGTATTG 1101
2.6 Jack	(1101)	A
2.6 Mav	(1101)	A
2.6 W82	(1101)	A
2.6 Wes	(1101)	A
2.6 X5	(1101)	A

图2(续)

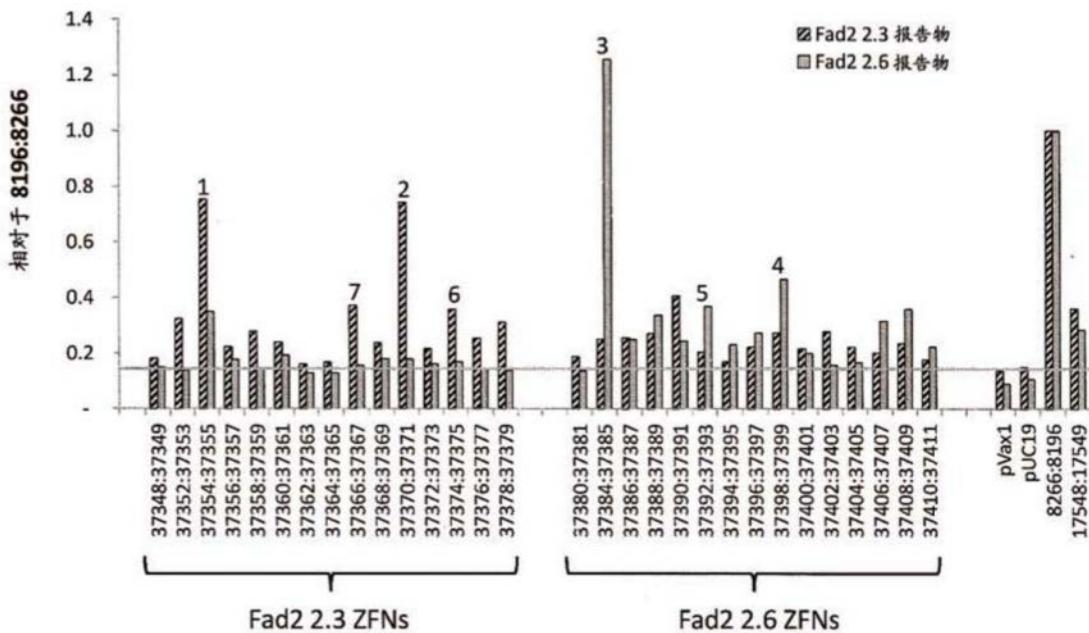


图3

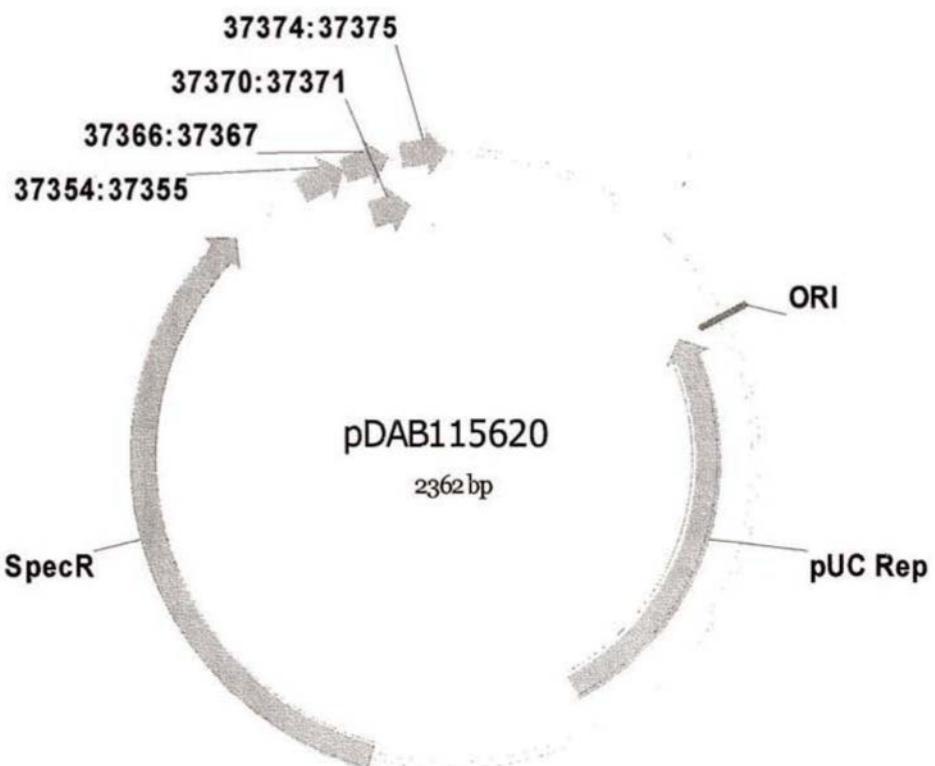


图4

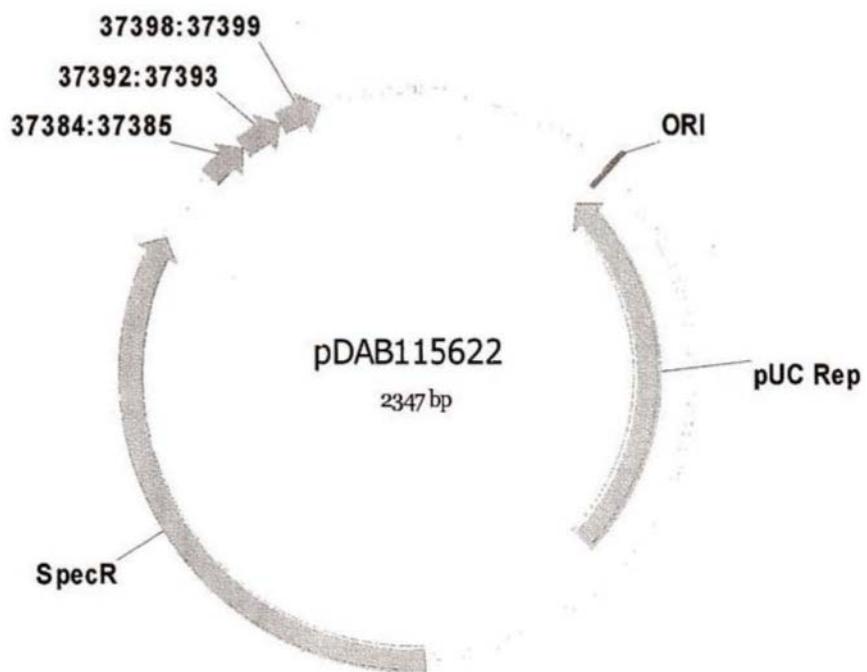


图5

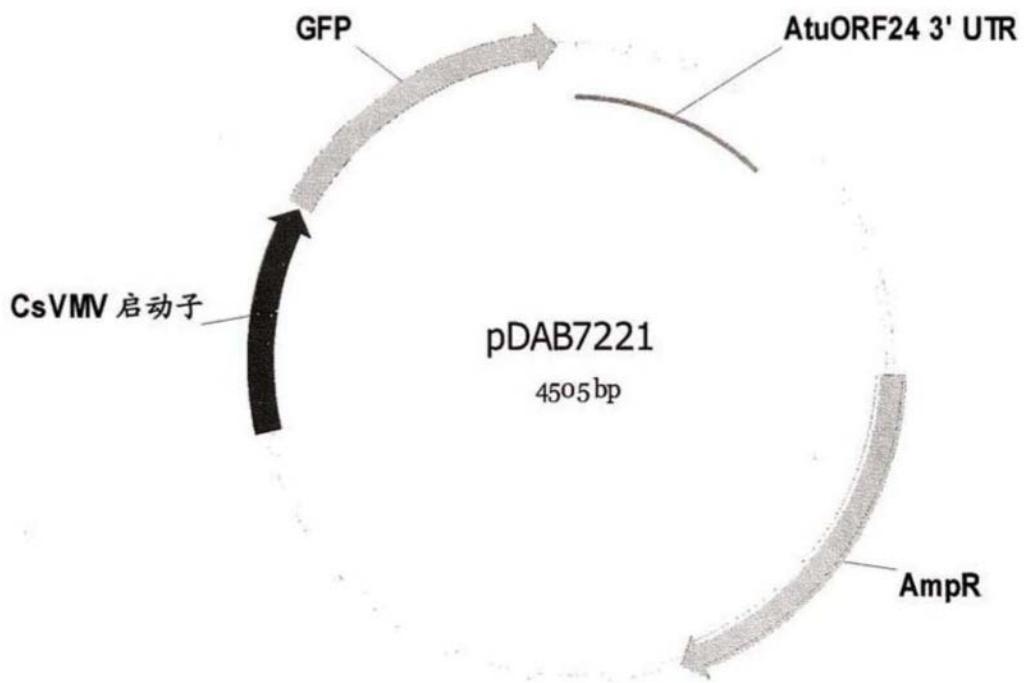


图6

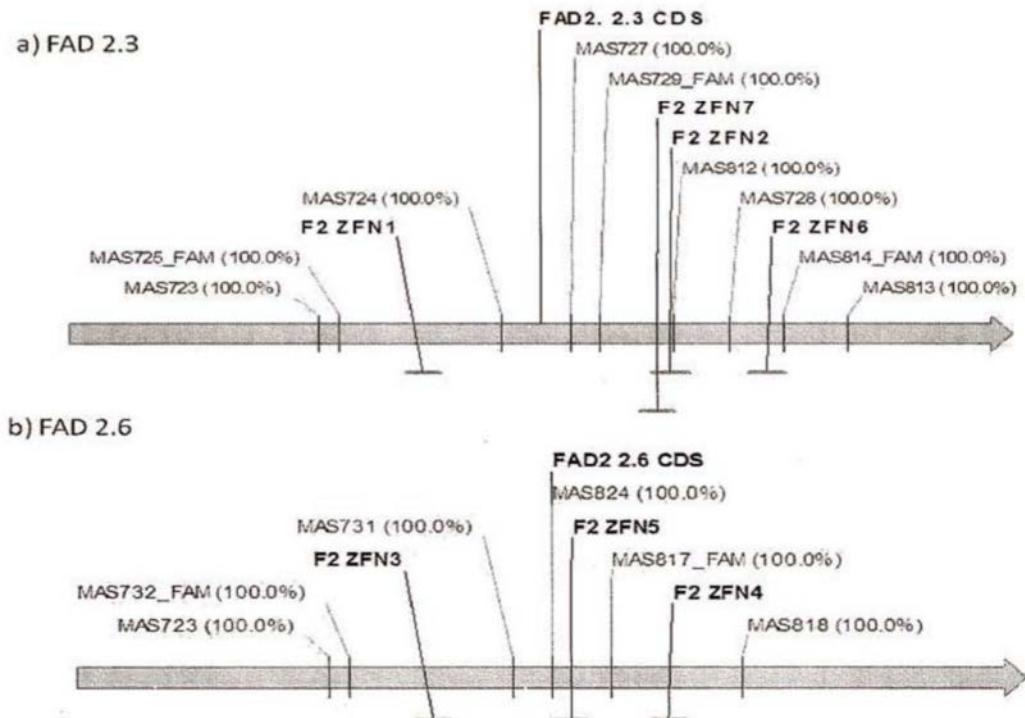


图7

名称	SEQ ID NO:	FAD2基因组序列	反向供体插入物	具有类似序列的质粒克隆
Reference	119	TTACTCTCTACCGTGGCAACCCCTGAAATTCAAGGGTTGCAACACGGT <u>AGAGAGAGTAA</u>	--	
1WT3	120	TTACTCTCTACCGTGGCAACCCCTGA <u>GAGAGAGTAA</u>	TCAGGGTTGCAACACGGTA	(1WT4,5)
1WT6	121	TTACTCTCTACCGTGGCAACCCCTGAA <u>GAGAGAGTAA</u>	AGGGTTGCAACACGGTA	--
4WT1	122	TTACTCTCTACCGTGGCAACCCCTGA <u>AGAGAGTAA</u>	GGTTGCAACACGGTAG	(4WT2,3,4,5)
4WT4	123	TTACTCTCTACCGTGGCAACCCCTGAA <u>缺失</u>	60 bp	(4WT6)
1HF1	124	TTACTCTCTACCGTGGCAACCCCTGA <u>AGAGAGAGTAA</u>	TTTCAGGGTTGCAACACGGT	(1HF3,5,6)
1HF2	125	TTACTCTCTACCGTGGCAACCCCTGA <u>GAGAGAGTAA</u>	TTCAGGGTTGCAACACGGTA	--
1HF4	126	TTACTCTCTACCGTGGCAACCCCTa <u>GAGAGAGTAA</u>	TTCAGGGTTGCAACACGGTA	--
6HF1	127	TTACTCTCTACCGTGGCAACCCCTGA <u>GAGAGAGTAA</u>	TTCAGGGTTGCAACACGGTA	--
6HF2	128	TTACTCTCTACCGTGGCAACCCCTGAA <u>AGAGAGAGTAA</u>	TCAGGGTTGCAACACGGT	--
6HF3	129	TTACTCTCTACCGTGGCAACCCCTGAA <u>GAGAGAGTAA</u>	AGGGTTGCAACACGGTA	(6HF5,6)
6HF4	130	TTACTCTCTACCGTGGCAACCCCTc <u>GAGAGAGTAA</u>	TTTCAGGGTTGCAACACGGTA	--

图8