



(12) 发明专利申请

(10) 申请公布号 CN 101878307 A

(43) 申请公布日 2010. 11. 03

(21) 申请号 200880118027. 3

(51) Int. Cl.

(22) 申请日 2008. 09. 25

C12N 15/82(2006. 01)

(30) 优先权数据

C12N 9/22(2006. 01)

60/995, 557 2007. 09. 27 US

C12N 15/90(2006. 01)

A01H 5/00(2006. 01)

(85) PCT申请进入国家阶段日

2010. 05. 27

(86) PCT申请的申请数据

PCT/US2008/011089 2008. 09. 25

(87) PCT申请的公布数据

W02009/042164 EN 2009. 04. 02

(71) 申请人 陶氏益农公司

地址 美国印第安纳州

申请人 桑格摩生物科学股份有限公司

(72) 发明人 曼朱·格普塔 阿莎·M·帕尔塔

斯蒂芬·诺瓦克 菲奥多·厄诺夫

桑尼塔·戈帕兰

(74) 专利代理机构 北京市柳沈律师事务所

11105

代理人 张文辉 封新琴

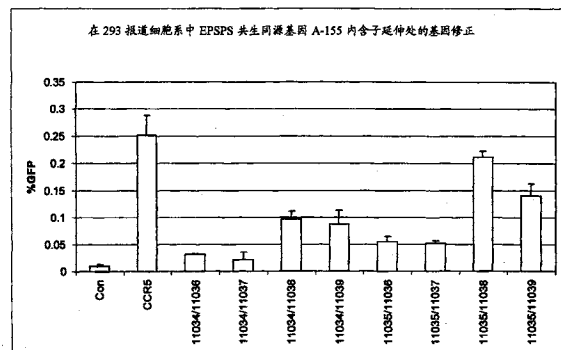
权利要求书 3 页 说明书 66 页 附图 26 页

(54) 发明名称

以 5-烯醇式丙酮酰莽草酸-3-磷酸合酶基因为靶的改造锌指蛋白

(57) 摘要

现行公开涉及在植物中以 5-烯醇式丙酮酰莽草酸-3-磷酸合酶 (EPSPS) 基因为靶的改造锌指蛋白, 以及使用上述锌指蛋白调整基因表达, 基因失活与靶向基因修饰的方法。具体而言, 本公开涉及供靶向剪切与改变 EPSPS 基因的锌指核酸酶。



1. 一种非天然存在的锌指蛋白 (ZFP), 其结合于目标 EPSPS 靶基因组区域, 所述 ZFP 包括一个或多个改造的锌指结合区域。
2. 权利要求 1 所述的 ZFP, 其中所述 ZFP 为包含一个或多个功能域的融合蛋白。
3. 权利要求 2 所述的 ZFP, 其中一个或多个功能域选自下述组成的组: 转录阻抑物、核酸内切酶、甲基转移酶、组蛋白脱乙酰酶、转录激活物和组蛋白乙酰转移酶。
4. 权利要求 3 所述的 ZFP, 其中所述 ZFP 为下述的锌指核酸酶 (ZFN), 其剪切目标 EPSPS 靶基因组区域, 所述 ZFN 包含一个或多个改造的锌指结合区域与核酸酶剪切域。
5. 权利要求 4 所述的 ZFN, 其中所述 ZFN 剪切一个或多个 EPSPS 共生同源基因或直向同源基因序列。
6. 编码权利要求 1-5 中任一所述的一个或多个锌指蛋白 (ZFP) 的多核苷酸。
7. 包含权利要求 6 所述的一个或多个多核苷酸的植物宿主细胞。
8. 权利要求 7 所述的植物宿主细胞, 其中所述细胞以一个或多个多核苷酸稳定转染。
9. 权利要求 7 所述的植物宿主细胞, 其中所述细胞以一个或多个多核苷酸暂时转染。
10. 一种在植物细胞中剪切一个或多个 EPSPS 基因的方法, 所述方法包括下述步骤: 将权利要求 6 所述的一个或多个多核苷酸导入植物细胞; 和在所述细胞中表达所述一个或多个 ZFN, 其中所述 ZFN 剪切一个或多个 EPSPS 基因。
11. 权利要求 10 所述的方法, 其中至少一个多核苷酸稳定转染入所述植物细胞中。
12. 权利要求 10 所述的方法, 其中至少一个多核苷酸暂时转染入所述植物细胞中。
13. 权利要求 10-12 中任一项所述的方法, 其中所述植物细胞中所有的 EPSPS 共生同源基因被剪切。
14. 权利要求 10-12 中任一项所述的方法, 其中所述植物细胞中一个或多个 EPSPS 共生同源基因被剪切。
15. 包含第一和第二 DNA 序列的供体载体; 其中所述第一序列和第三序列同源, 而所述第二序列与第四序列同源; 且其中所述第三与第四序列为染色体 EPSPS DNA 序列。
16. 权利要求 15 所述的供体载体, 其中第三与第四序列的近边 (near edge) 为毗连的。
17. 权利要求 15 所述的供体载体, 其中第三与第四序列的近边通过至少一个核苷酸对分隔。
18. 权利要求 15-17 中任一项所述的载体, 其中所述第三与第四序列为外源序列。
19. 权利要求 15-17 中任一项所述的载体, 其中所述第三与第四序列为内源序列。
20. 权利要求 15-19 中任一项所述的载体, 进一步包含第五序列, 其中所述第五序列:
  - (a) 插入所述第一与第二序列间; 且
  - (b) 为外源核酸序列。
21. 权利要求 20 所述的载体, 其中所述第五序列包含编码选择性标记的序列。
22. 权利要求 21 所述的载体, 其中选择性标记选自下述组成的组: 绿色荧光蛋白 (GFP)、 $\beta$ -葡萄糖醛酸糖苷酶 (GUS)、草胺膦 N-乙酰转移酶 (PAT, BAR)、新霉素磷酸转移酶、潮霉素磷酸转移酶、 $\beta$ -内酰胺酶、儿茶酚二氧合酶、 $\alpha$ -淀粉酶、酪氨酸酶、 $\beta$ -半乳糖苷酶、荧光素酶、水母发光蛋白、EPSP 合酶、胍水解酶、乙酰乳糖合酶 (ALS)、二氢叶酸还原酶 (DHFR)、茅草枯脱卤素酶及邻氨基苯甲酸合酶。

23. 权利要求 20 所述的载体,其中所述第五序列包含选自下组的序列:编码除选择性标记以外蛋白质的序列;一个或多个转录调节序列;一个或多个增强或减弱蛋白质靶向的序列;一个或多个编码蛋白质的部分的序列;小干扰 RNA;和微 RNA。

24. 权利要求 23 所述的载体,其中所述第五序列包含编码突变体 EPSPS 染色体序列的序列,其增加植物针对除草剂草甘膦的耐受性。

25. 一种将外源核酸序列导入植物细胞基因组中的方法,所述方法包括下列步骤:

在存在权利要求 15-24 中任一项所述的供体载体的情况下,在所述细胞中表达一个或多个锌指核酸酶(ZFN);

其中所述 ZFN 剪切第三或第四序列之一的 EPSPS 千碱基对;

从而使得在步骤(b)中染色体 DNA 的剪切刺激所述供体载体通过同源重组整合入基因组中。

26. 一种在植物细胞中表达外源核酸序列产物的方法,所述方法包括下述步骤:

根据权利要求 25 方法导入所述的外源序列从而使得所述外源序列的产物在植物细胞中表达。

27. 一种得自权利要求 25 或 26 所述方法的转基因植物细胞或转基因植物。

28. 一种在植物细胞基因组中刺激分子内同源重组的方法,所述方法包括下述步骤:

将权利要求 4-6 中任一项所述的 ZFN 在存在下述 DNA 片段的情况下导入所述植物细胞,其中所述片段包含 EPSPS 基因和与所述植物细胞中第二序列同源的第一序列,从而使得所述 ZFN 剪切该 DNA 片段并刺激分子内同源重组。

29. 权利要求 28 所述的方法,其中同源重组导致在所述内源 DNA 片段内的缺失。

30. 权利要求 29 所述的方法,其中所述缺失序列选自下组:编码 EPSPS 基因全部或部分的序列和编码选择性标记全部或部分的序列。

31. 权利要求 29 所述的方法,其中所述缺失 DNA 由外源序列取代,所述方法进一步包括:

将多核苷酸导入所述细胞,其中所述多核苷酸包括:

(a) 第四与第五序列,其中所述第四序列与邻近于所述第一序列的非缺失序列同源,而所述第五序列与所述第二序列同源;和

(b) 所述外源序列。

32. 权利要求 31 所述的方法,其中所述外源序列为选择性标记。

33. 权利要求 32 所述的方法,其中所述选择性标记选自下组:绿色荧光蛋白(GFP)、 $\beta$ -葡糖醛酸糖苷酶(GUS)、草胺膦 N-乙酰转移酶(PAT, BAR)、新霉素磷酸转移酶、潮霉素磷酸转移酶、 $\beta$ -内酰胺酶、儿茶酚二氧合酶、 $\alpha$ -淀粉酶、酪氨酸酶、 $\beta$ -半乳糖苷酶、荧光素酶、水母发光蛋白、EPSP 合酶、胍水解酶、乙酰乳糖合酶(ALS)、二氢叶酸还原酶(DHFR)、茅草枯脱卤素酶和邻氨基苯甲酸合酶。

34. 权利要求 31 所述的方法,其中所述外源序列为 EPSPS 基因序列。

35. 权利要求 34 所述的方法,其中所述 EPSPS 基因序列包含突变。

36. 权利要求 35 所述的方法,其中所述突变增加植物针对除草剂草甘膦的耐受性。

37. 一种从植物基因组中缺失 EPSPS 基因序列的方法,所述方法包括下述步骤:

在所述细胞中表达权利要求 4 所述的第一与第二锌指核酸酶(ZFN),其中所述第一 ZFN

在 EPSPS 基因的第一剪切位点处剪切,而所述第二 ZFN 在所述 EPSPS 基因的第二剪切位点处剪切,其中所述 EPSPS 基因序列位于所述第一剪切位点与第二剪切位点之间,其中第一与第二剪切位点的剪切导致所述 EPSPS 基因序列的缺失。

38. 权利要求 37 所述的方法,其中第一与第二剪切位点非同源末端连接。

39. 权利要求 37 或 38 所述的方法,其中所述第一与第二剪切位点通过至少 100 个核苷酸分隔。

40. 权利要求 37 所述的方法,其中所述植物细胞为包含外源 EPSPS 序列的转基因植物细胞。

41. 一种调整 EPSPS 基因表达的方法,所述方法包括:

在所述细胞中表达权利要求 1 到 3 中任一项所述的 ZFP,其中所述 ZFP 与所述 EPSPS 基因的靶位点结合,从而调整所述 EPSPS 基因的表达。

42. 权利要求 41 所述的方法,其中所述 ZFP 增加所述 EPSPS 基因的转录。

43. 权利要求 42 所述的方法,其中所述 ZFP 增加植物针对除草剂草甘膦的耐受性。

44. 权利要求 41 所述的方法,其中所述 ZFP 降低所述 EPSPS 基因的转录。

45. 权利要求 44 所述的方法,其中所述 ZFP 降低植物针对除草剂草甘膦的耐受性。

## 以 5-烯醇式丙酮酰莽草酸-3-磷酸合酶基因为靶的改造锌指蛋白

### 技术领域

[0001] 本发明主要涉及植物中基因组工程,基因打靶,靶向染色体整合及蛋白质表达等领域。特别而言,本发明涉及以 5-烯醇式丙酮酰莽草酸-3-磷酸合酶 (5-enolpyruvyl shikimate-3-phosphate synthase) 基因为靶的改造锌指蛋白 (engineered zinc finger proteins), 及在调节基因表达、基因失活及靶向基因修饰中使用该锌指蛋白的方法。更加特别而言,本发明属于催化 EPSPS 基因靶向剪切和改变的改造锌指核酸酶。

### 现有技术

[0002] 农业中一个主要让人感兴趣的方向,尤其考虑到不少植物基因组全核苷酸序列业已确定,是靶向调节基因表达和基因序列改变。特别而言,调节基因表达及修饰内源植物序列的能力可以促进许多应用,例如,农作物特性的优化,所述特性影响营养价值、产量、胁迫耐受性、病原体抗性、油质量以及对农药的抗性和 / 或改造植物用作生物工厂以生产医药化合物或工业化学品。

[0003] 改造锌指蛋白 (ZFPs) 已被方便的用于选择性的调节基因表达及靶向改变植物基因序列 (参见,如, U. S. Patents Nos. 7, 262, 054, 7, 235, 354, 7, 220, 719, 7, 001, 768 与 6, 534, 261; 以及 U. S. Serial No. 60/874, 911)。锌指蛋白是以序列特异性为特征依靠称为锌指由金属稳定化的域结合 DNA、RNA 和 / 或蛋白质的蛋白质。参见,例如, Miller 等人 (1985)EMBO J. 4 :1609-1614; Rhodes et al. (1993)Sci. Amer. 268(2) :56-65; 及 Klug (1999) J. Mol. Biol. 293 :215-218。ZFPs 通常见于转录因子,迄今为止,在数千已知或推定的转录因子中辨识出超过 10000 个锌指序列。

[0004] 对所选基因靶的调节和改变理论上可以通过设计对具有所希望生物活性,预先确定的 DNA 序列有特异性的 ZFPs 来达成。例如,将锌指结构域与调节性的域在融合蛋白中结合,从而产生改造锌指转录因子以控制基因调节 (参见,例如 U. S. Patent No. 6, 534, 261)。锌指结构域亦与核酸酶剪切域结合,产生锌指核酸酶 (ZFNs) 以特异性的靶向希望修饰 (如,缺失,突变,同源重组或插入外源序列) 的基因组双链断裂 (参见,如 U. S. Patent Application Publication Nos. 2007/0134796 与 2005/0064474)。改造 ZFPs 极大的促进了外源序列的插入或在植物中特异靶位点内源序列的修饰,并提供了比常规方法效率更高的对植物基因组的靶向改变 (参见,如 U. S. Patents Nos. 7, 262, 054, 7, 235, 354, 7, 220, 719, 7, 001, 768 与 6, 534, 261)。

[0005] 然而,基因组复制常见于植物中,而且仍有对在植物基因组中靶向改变上述共生同源基因 (paralogous genes), 以及在植物中调节共生同源基因表达的组合物与方法的需求。

### [0006] 发明概述

[0007] 本发明提供在植物细胞中调节表达及靶向改变一个或多个共生同源基因 (例如 EPSPS 基因) 的组合物与方法。植物细胞可以来自单子叶植物 (monocots) 或双子叶植物

(dicots) 物种,亦包括培养的细胞,在任何发育阶段植物中的细胞,以及从全本植物上移除的细胞和置回植物的细胞(或其后代)。植物细胞亦包括一个或多个同源或共生同源的基因序列,其中任何部分或全部可作为本发明描述方法进行修饰的靶。

[0008] 在一方面,本文描述结合作为目标 EPSPS 靶基因组区域的锌指蛋白(ZFPs),其中所述 ZFP 包括一个或更多改造的锌指结合区域。在一些实施方案中,锌指结合区域包含表 A 所示的序列。在一些实施方案中,作为 ZFP 靶的 EPSPS 基因包括选自下组的核苷酸序列: SEQ ID NOS:10-14 或与其至少有大约 80% -100% 同一性的序列,包括任何在该范围内的同一性百分比,例如与其有 81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99% 序列相同(序列同一性)。在一些实施方案中,所述 ZFP 是融合蛋白,包含一个或更多的调节区。在一个实施方案中,一个或更多的调节区选自下组:转录阻抑物、核酸内切酶、甲基转移酶、组蛋白脱乙酰酶、转录激活物以及组蛋白乙酰转移酶。在一个实施方案中,所述 ZFP 与 EPSPS 基因的靶序列结合,其中 EPSPS 表达增加或降低。在一个实施方案中,所述 ZFP 与 EPSPS 基因的转录调节序列结合。在另一个实施方案中,所述 ZFP 在 EPSPS 基因的转录起始位点上游结合。在另一个实施方案中,所述 ZFP 在 EPSPS 基因转录起始位点邻近结合。在另一个实施方案中,所述 ZFP 在 EPSPS 基因的转录起始位点下游结合。在一个实施方案中,所述 ZFP 在 EPSPS 基因转录起始位点下游的 RNA 多聚酶暂停位点(pause site) 邻近结合。

[0009] 在一个实施方案中,所述 ZFP 是剪切目标 EPSPS 靶基因组区域的锌指核酸酶(ZFN),其中所述 ZFN 包含一个或更多改造的锌指结合区域,与核酸酶剪切(cleavage) 域。在一些实施方案中,所述 ZFN 包括融合多肽,所述多肽包含对 EPSPS 基因序列有特异性的改造的锌指结合区域与剪切域,所述 ZFN 还包括一个或更多融合多肽,所述多肽包含改造的锌指结合区域与剪切半域(half domain)。在一些实施方案中,锌指结合区域包括选自自由包含表 A 所示识别域的锌指蛋白组成的组。剪切域与剪切半域可由例如各种限制性核酸内切酶和/或靶向(homing) 核酸内切酶得到。在一个实施方案中,所述剪切域与剪切半域源自 IIS 型限制性核酸内切酶(例如, Fok I)。所述 ZFN 可能特异性的剪切一种特定 EPSPS 基因序列。或者,所述 ZFN 可能剪切两种或更多同源 EPSPS 基因序列,所述同源序列可能包括 EPSPS 共生同源或正向(orthologous) 同源基因序列。

[0010] 在一些实施方案中,作为 ZFN 靶的 EPSPS 基因包含选自下组的核苷酸序列: SEQ ID NOS:10-14 或与其至少有大约 80% -100% 同一性的序列,包括任何在该范围内的同一性百分比,例如与其有 81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99% 序列同一性。

[0011] 所述 ZFN 可能在 EPSPS 基因的编码区域内或在该基因内或邻近该基因的非编码区域,例如,前导序列、尾随序列或内含子,或者在非转录区域内编码区域上游,或其下游区域结合和/或剪切 EPSPS 基因。在一些实施方案中,所述 ZFN 结合和/或剪切所述 EPSPS 基因的编码序列或调节序列。在一些实施方案中,所述 ZFN 在选自 SEQ ID NOS:10-14 的核苷酸序列组成的区域内结合并剪切 EPSPS 基因。

[0012] 在另一方面,本文描述了包括一个或更多 ZFP 的组合物,所述组合物可能包含一个或更多 ZFN。植物细胞可能包含一个独特的 EPSPS 基因或多个 EPSPS 共生同源基因。因此,组合物可能包括一个或更多 ZFP,所述 ZFP 在植物细胞中靶向一个或更多 EPSPS 基因,例

如,1、2、3、4、5或直到任意数量的EPSPS共生同源基因或植物细胞中存在的所有EPSPS共生同源基因。在一个实施方案中,所述组合物包括一个或更多在植物细胞中靶向所有EPSPS共生同源基因的ZFP。在另一个实施方案中,所述组合物包括一个在植物细胞中特异性的靶向一个特定EPSPS共生同源基因的ZFP。举例而言,所述组合物可能包括一个在植物细胞中特异性的结合并剪切一个特定EPSPS共生同源基因的ZFN,或多个在植物细胞中特异性的结合并剪切两个或更多EPSPS共生同源基因的ZFN。另外,组合物可能包含非核酸酶的ZFP,所述ZFP改变一个或更多EPSPS共生同源基因的转录调节。

[0013] 在另一方面,本文描述了编码一个或更多本文描述的ZFP的多肽。在一个实施方案中,所述多肽编码至少一个ZFN。作为示例的多肽包括编码表A所示任意锌指蛋白的核苷酸序列。

[0014] 在另一方面,本文描述了ZFP表达载体,所述载体包含编码一个或更多,可操纵的与启动子连接,本文所描述ZFP的多核苷酸。在一个实施方案中,一个或更多ZFP是ZFN。

[0015] 在另一方面,本文描述了包含一个或更多ZFP表达载体的植物宿主细胞。该植物宿主细胞可能是由一个或更多ZFP表达载体稳定的转化或暂时性的转染,或两者组合。在一个实施方案中,所述一个或更多ZFP表达载体在植物宿主细胞中表达一个或更多ZFN。

[0016] 在另一方面,本文描述了在植物细胞中剪切一个或更多共生同源基因的方法,该方法包括:(a) 将一个或更多编码一个或更多ZFN的表达载体导入植物细胞,所述ZFN与一个更多共生同源基因的靶位点在下述条件下结合:所述ZFN表达,且剪切一个或更多共生同源基因。在一些实施方案中,所述靶位点在EPSPS基因中。在一个实施方案中,在植物细胞中一个特定EPSPS共生同源基因被剪切。在另一个实施方案中,多于一个EPSPS共生同源基因被剪切,举例而言,2、3、4、5或直至任意数量的EPSPS共生同源基因或植物细胞中存在的所有EPSPS共生同源基因被剪切。

[0017] 在另一方面,本文描述了一个包括第一与第二DNA序列的供体载体,其中(i)第一序列与第三序列同源,而第二序列与第四序列同源;且(ii)第三和第四序列是染色体DNA序列。在一些实施方案中,第三和第四序列的邻近(near edge)被至少一对核苷酸分开。在一个实施方案中,第三和第四序列是内源序列。在另一个实施方案中,第三和第四序列是外源序列。在任何供体载体中,作为靶的染色体DNA序列可能为EPSPS序列。在一些实施方案中,所述染色体EPSPS DNA序列属于包括选自下组核苷酸序列的EPSPS基因:SEQ ID NOS:10-14或与其至少有大约80%-100%同一性的序列,包括任何在该范围内的同一性百分比,例如与其有81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99%序列同一性。

[0018] 在一些实施方案中,供体载体第一或第二序列中至少一个长度为100核苷酸或更短。另外,本文描述的任一载体可能进一步包括第五序列,其中所述第五序列:(a)介于第一和第二序列之间,且(b)为外源序列。在一些实施方案中,第五序列大小至少为一对碱基,但可能大到22千碱基对或更大。

[0019] 所述供体载体(比如,第五序列)可能亦包括编码蛋白质或蛋白质一部分的序列。在一些实施方案中,所述编码蛋白质的序列编码选择性标记(例如,绿色荧光蛋白(GFP)、 $\beta$ -葡糖醛酸糖苷酶(GUS)、草胺膦N-乙酰转移酶(PAT, BAR)、新霉素磷酸转移酶、潮霉素磷酸转移酶、 $\beta$ -内酰胺酶、儿茶酚二氧合酶、 $\alpha$ -淀粉酶、酪氨酸酶、 $\beta$ -半乳糖苷酶、荧光

素酶、水母发光蛋白、EPSP 合酶、腈水解酶、乙酰乳酸合酶 (ALS)、二氢叶酸还原酶 (DHFR)、茅草枯 (dalapon) 脱卤素酶及邻氨基苯甲酸合酶)。在另一些实施方案中,蛋白质编码序列(例如,第五序列)编码蛋白质或蛋白质一部分,例如与染色体序列同源的序列。

[0020] 在其它实施方案中,所述供体载体(例如,第五序列)包括一个或更多转录调节序列。举例而言,供体载体可能包括一个或更多增加或减少所述共生同源基因(例如,EPSPS)表达的转录调节序列。在一些实施方案中,供体载体包括一个或多个增强或减弱蛋白质运输的蛋白质靶向序列。

[0021] 进一步在其它实施方案中,所述供体载体(例如,第五序列)可能包括与突变体染色体序列(例如,EPSPS)对应的野生型,或作为替代,与野生型染色体序列(例如,EPSPS)对应的突变体。在一些实施方案中,所述突变体染色体序列包括一个或更多选自下组的突变:点突变、取代、缺失与插入。在一个实施方案中,所述供体载体包括突变体 EPSPS 染色体序列,所述序列增加植物对除草剂草甘膦 (glyphosate) 的耐受性。

[0022] 在本文描述的任何供体载体中,第一序列可能与第三序列有至少 35% 的同源性。类似的,在本文描述的任何供体载体中,第二序列可能与第四序列有至少 35% 的同源性。在有些实施方案中,第一序列与第三序列有至少 35% 至 50%、至少 50% 至 70%、至少 70% 至 80%、至少 80% 至 85%、至少 85% 至 90%、至少 90% 至 95%、至少 95%、96%、97%、98%、99% 或 100% 的同源性。在有些实施方案中,第二序列与第四序列有至少 35% 至 50%、至少 50% 至 70%、至少 70% 至 80%、至少 80% 至 85%、至少 85% 至 90%、至少 90% 至 95%、至少 95%、96%、97%、98%、99% 或 100% 的同源性。

[0023] 在又一方面,本文描述了一种将外源序列导入植物细胞基因组的方法。所述方法包括下列步骤:(a) 使细胞与本文描述的任何供体载体接触;(b) 在细胞中表达一个或更多锌指核酸酶,其中一个或更多锌指核酸酶在第三或第四序列任一的 0.4 和 3 千碱基对之间剪切染色体 DNA,其结果是在步骤 (b) 中染色体 DNA 的剪切促进了所述供体载体通过同源重组整合入基因组中。在一些实施方案中,一个或更多核酸酶为 IIS 型限制性核酸内切酶的剪切域与改造的锌指结合区域间的融合蛋白。

[0024] 在一些实施方案中,所述锌指核酸酶剪切包括选自下组核苷酸序列的 EPSPS 基因:SEQ ID NOS:10-14 或与其至少有大约 80% -100% 同一性的序列,包括任何在该范围内的同一性百分比,例如与其有 81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99% 序列同一性。

[0025] 在另一方面,本文描述了一种在植物中表达外源核酸序列的方法,该方法包括下列步骤:(a) 使细胞与包含外源核酸序列的供体载体接触;(b) 在细胞中表达锌指核酸酶 (ZFN),其中 ZFN 在染色体中距离第三或第四序列任一 3 千碱基对之内剪切一个或更多共生同源基因(例如,一个或更多 EPSPS 基因)。在步骤 (b) 中剪切染色体 DNA 导致供体载体通过同源重组整合入基因组及外源核酸序列产物的表达。

[0026] 在一些实施方案中,所述锌指核酸酶剪切包括选自下组核苷酸序列的 EPSPS 基因:SEQ ID NOS:10-14 或与其至少有大约 80% -100% 同一性的序列,包括任何在该范围内的同一性百分比,例如与其有 81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99% 序列同一性。

[0027] 在另一方面,本文描述了一种在植物细胞基因组中进行分子内同源重组的方法,



该方法包括下列步骤:(a) 提供包含靶基因序列并进一步包含第一序列的 DNA 片段,其中第一序列与第二序列同源;(b) 使所述 DNA 片段与本文描述的 ZFN 接触,其中 ZFN 在靶基因序列处剪切 DNA 片段,从而促进分子内同源重组。在一些实施方案中,所述 DNA 片段对所述细胞为内源的。在其它实施方案中,所述 DNA 片段对所述细胞为外源的。在一些实施方案中,所述靶基因对所述细胞为特有的(unique)。在其它实施方案中,所述靶基因为共生同源基因。在上述任何方法中,所述靶基因可能包括特有的或共生同源的 EPSPS 基因且所述 ZFN 包括表 A 所示的任何序列。在一些实施方案中,同源重组可能在染色体中发生。在一个实施方案中,第一和第二序列间的 DNA 从染色体中缺失。在一个实施方案中,从染色体中缺失的序列可能编码全部或部分靶基因。在另一个实施方案中,从染色体中缺失的序列可能编码全部或部分选择性标记,例如,绿色荧光蛋白(GFP)、 $\beta$ -葡糖醛酸糖苷酶(GUS)、phosphinothricin N-乙酰转移酶(PAT, BAR)、新霉素磷酸转移酶、潮霉素磷酸转移酶、 $\beta$ -内酰胺酶、儿茶酚二氧合酶、 $\alpha$ -淀粉酶、酪氨酸酶、 $\beta$ -半乳糖苷酶、荧光素酶、水母发光蛋白、EPSP 合酶、胍水解酶、乙酰乳糖合酶(ALS)、二氢叶酸还原酶(DHFR)、茅草枯脱卤素酶及邻氨基苯甲酸合酶。

[0028] 在一些实施方案中,所述缺失的 DNA 由外源序列取代,该方法进一步包括将多核苷酸导入细胞,其中该多核苷酸包括(i) 第四和第五序列,其中第四序列与靠近第一序列的未缺失序列同源且第五序列与靠近第二序列的未缺失序列同源;(ii) 所述外源序列。

[0029] 在一些实施方案中,缺失的 DNA 由基因序列取代,所述基因序列可能包括野生型基因序列的对应突变体。在一些实施方案中,所述突变体基因序列包括一个或更多选自下组的突变:点突变、取代、缺失及插入。在一个实施方案中,缺失的 DNA 被 EPSPS 基因序列取代,例如,包含增加植物对除草剂草甘膦耐受性的突变的 EPSPS 基因序列。

[0030] 在另一个实施方案中,所述外源序列可能是选择性标记,例如,绿色荧光蛋白(GFP)、 $\beta$ -葡糖醛酸糖苷酶(GUS)、phosphinothricin N-乙酰转移酶(PAT, BAR)、新霉素磷酸转移酶、潮霉素磷酸转移酶、 $\beta$ -内酰胺酶、儿茶酚二氧合酶、 $\alpha$ -淀粉酶、酪氨酸酶、 $\beta$ -半乳糖苷酶、荧光素酶、水母发光蛋白、EPSP 合酶、胍水解酶、乙酰乳糖合酶(ALS)、二氢叶酸还原酶(DHFR)、茅草枯脱卤素酶及邻氨基苯甲酸合酶。

[0031] 在另一个实施方案中,本文描述了从植物细胞基因组中缺失基因序列的方法,该方法包括(a) 提供包含基因序列的植物细胞;(b) 在所述细胞中表达第一和第二锌指核酸酶(ZFNs),其中第一 ZFN 在第一剪切位点剪切,而第二 ZFN 在第二剪切位点剪切,其中所述基因序列位于第一剪切位点和第二剪切位点之间,其中第一和第二剪切位点的剪切导致所述基因序列的缺失。在一些实施方案中,所述基因序列是 EPSPS 基因。缺失的基因序列大小由第一和第二剪切位点间的距离确定。相应的,可以获取在任何目标基因组区域中任何大小的缺失。可以获得 25、50、100、200、300、400、500、600、700、800、900、1000 对核苷酸,或在此范围内任意整数值对核苷酸的缺失。此外,可以使用上述方法及本文公开的组合物缺失任何大于 1000 对核苷酸的整数值对核苷酸的序列。在一个实施方案中,第一和第二剪切位点间隔至少 100 个核苷酸。在一个实施方案中,整个基因(例如,EPSPS)缺失。在另一个实施方案中,基因(例如,EPSPS)一部分缺失。在一个实施方案中,所述基因序列(例如,EPSPS 基因序列)从转基因植物细胞中缺失。所述基因序列(例如,EPSPS)可以是内源或外源序列。

[0032] 在另一方面,本文描述了调整植物基因调节的方法,所述方法包括(a)提供包含靶基因序列的植物细胞;(b)在该细胞中表达ZFP,其中所述ZFP与该靶基因调节序列结合,从而调整靶基因的调节。在一些实施方案中,所述基因序列为EPSPS基因。ZFP与调节序列的结合可能增加或减少靶基因(例如,EPSPS)的转录。在一些实施方案中,ZFP亦增加或减少植物对除草剂草甘膦的耐受性。

[0033] 在另一方面,亦提供了通过本文中描述的任何方法所得的转基因植物细胞。

[0034] 在另一方面,本文提供了包含本文描述方法所得的转基因植物细胞的植物。

[0035] 在本文描述的任何方法中,对靶植物基因序列的修饰(例如,转录调节序列或EPSPS编码序列)可用于增加或减少植物对除草剂草甘膦的耐受性。

[0036] 因此,本发明包括,但不限于,下列编号的实施方案:

[0037] 1. 一种与目标EPSPS靶基因组区域结合的锌指蛋白(ZFP),所述ZFP包括一个或更多改造的锌指结合区域。

[0038] 2. 实施方案1中所述的ZFP,其中所述靶基因组区域是在双子叶植物细胞中。

[0039] 3. 实施方案2中所述的ZFP,其中所述靶基因组区域是在菜籽(canola)细胞中。

[0040] 4. 实施方案2中所述的ZFP,其中所述靶基因组区域是在芸薹(Brassica)植物细胞中。

[0041] 5. 实施方案1中所述的ZFP,其中所述目标EPSPS靶基因组区域属于包含选自下组的核苷酸序列的EPSPS基因:SEQ ID NOS:10-14或与其至少有95%同一性的序列。

[0042] 6. 实施方案1中所述的ZFP,其中所述ZFP是包含一个或更多功能域的融合蛋白。

[0043] 7. 实施方案6中所述的ZFP,包含选自下组的一个或更多功能域:转录阻抑物、核酸内切酶、甲基转移酶、组蛋白脱乙酰酶、转录激活物以及组蛋白乙酰转移酶。

[0044] 8. 实施方案1-7中任一项所述的ZFP,其中所述ZFP与所述EPSPS基因转录调节序列结合。

[0045] 9. 实施方案1-7中任一所述的ZFP,其中所述ZFP在所述EPSPS基因转录起始序列的上游结合。

[0046] 10. 实施方案1-7中任一所述的ZFP,其中所述ZFP在所述EPSPS基因转录起始序列邻近结合。

[0047] 11. 实施方案1-10中任一所述的ZFP,其中所述ZFP增加所述EPSPS基因的转录。

[0048] 12. 实施方案1-10中任一所述的ZFP,其中所述ZFP减少所述EPSPS基因的转录。

[0049] 13. 实施方案1中所述的ZFP,其中所述ZFP为剪切EPSPS目标靶基因组区域的锌指核酸酶(ZFN),所述ZFN包含一个或更多锌指结合区域及核酸酶剪切域。

[0050] 14. 实施方案13所述的ZFN,其中所述剪切域包含两个剪切半域。

[0051] 15. 实施方案14所述的ZFN,其中所述剪切半域源自相同核酸酶。

[0052] 16. 实施方案15所述的ZFN,其中所述剪切半域源自IIS型限制性内切酶。

[0053] 17. 实施方案16所述的ZFN,其中所述IIS型限制性内切酶是Fok I。

[0054] 18. 实施方案13所述的ZFN,其中所述EPSPS目标靶基因组区域属于包含选自下组的核苷酸序列的EPSPS基因:SEQ ID NOS:10-14或与其至少有95%同一性的序列。

[0055] 19. 实施方案13所述的ZFN,其中所述ZFN与EPSPS基因的编码区域结合。

[0056] 20. 实施方案13所述的ZFN,其中所述ZFN与EPSPS基因的非编码区域结合。

- [0057] 21. 实施方案 20 所述的 ZFN,其中所述 ZFN 与 EPSPS 基因的调节序列结合。
- [0058] 22. 实施方案 13 所述的 ZFN,其中所述 ZFN 剪切一个或更多 EPSPS 共生同源或正向同源基因序列。
- [0059] 23. 实施方案 13 所述的 ZFN,其中所述 ZFN 特异性的剪切一个 EPSPS 共生同源或正向同源基因序列。
- [0060] 24. 实施方案 13 所述的 ZFN,包括含有如表 A 所示序列的锌指结合区域。
- [0061] 25. 实施方案 13 所述的 ZFN,其中所述 ZFN 在选自 SEQ ID NOS:10-14 的核苷酸序列组成的区域内结合并剪切 EPSPS 基因。
- [0062] 26. 实施方案 13 所述的 ZFN,其中所述 ZFN 包括:
- [0063] (a) 第一融合蛋白,包括第一锌指结合区域,和第一剪切半域,其中第一锌指结合区域与第一核苷酸序列结合;以及
- [0064] (b) 第二融合蛋白,包括第二锌指结合区域,和第二剪切半域,其中第二锌指结合区域与第二核苷酸序列结合。
- [0065] 27. 实施方案 26 所述的 ZFN,其中所述第二核苷酸序列位于距离第一核苷酸序列 2 和 50 个核苷酸间的位置。
- [0066] 28. 实施方案 26 所述的 ZFN,其中剪切发生在第一和第二核苷酸序列之间。
- [0067] 29. 一种包含实施方案 1-28 中任一所述的一个或更多锌指蛋白 (ZFPs) 的组合物。
- [0068] 30. 实施方案 29 所述的组合物,其中一个或更多 ZFP 为锌指核酸酶 (ZFNs)。
- [0069] 31. 实施方案 29 所述的组合物,包括一个或多个在植物细胞中靶向一个或更多 EPSPS 基因的 ZFP。
- [0070] 32. 实施方案 29 所述的组合物,包括两个或多个在植物细胞中组合靶向所有的 EPSPS 共生同源基因的 ZFP。
- [0071] 33. 实施方案 30 所述的组合物,包括一个在植物细胞中特异性的结合并剪切一个 EPSPS 共生同源基因的 ZFN
- [0072] 34. 实施方案 30 所述的组合物,包括两个或更多在植物细胞中结合并剪切两个或更多 EPSPS 共生同源基因的 ZFN。
- [0073] 35. 实施方案 30 所述的组合物,包括一个或更多在植物细胞中结合并剪切所有 EPSPS 共生同源基因的 ZFN。
- [0074] 36. 编码实施方案 1-28 中任一所述的一个或更多锌指蛋白 (ZFP) 的多核苷酸。
- [0075] 37. 实施方案 36 所述的多核苷酸,包含编码如表 A 所示锌指蛋白的核苷酸序列。
- [0076] 38. 一种 ZFP 表达载体,包含可操纵的与启动子连接的实施方案 36 或 37 所述的多核苷酸。
- [0077] 39. 一种包含一种或更多实施方案 38 所述 ZFP 表达载体的植物宿主细胞。
- [0078] 40. 实施方案 39 所述的植物宿主细胞,其中所述细胞稳定的由一种或更多 ZFP 表达载体转染。
- [0079] 41. 实施方案 39 所述的植物宿主细胞,其中所述细胞暂时性的由一种或更多 ZFP 表达载体转染。
- [0080] 42. 一种在植物细胞中剪切一个或更多 EPSPS 基因的方法,该方法包括:
- [0081] (a) 用编码一种或更多实施方案 10 所述的 ZFN 的一种或更多 ZFP 表达载体转染所

述植物细胞；

[0082] (b) 在所述细胞中表达一种或更多 ZFN, 其中所述 ZFN 剪切一个或更多 EPSPS 基因。

[0083] 43. 实施方案 42 所述的方法, 其中至少一种 ZFP 表达载体稳定转染至所述植物细胞中。

[0084] 44. 实施方案 42 所述的方法, 其中至少一种 ZFP 表达载体暂时转染至所述植物细胞中。

[0085] 45. 实施方案 42-44 任一所述的方法, 其中至少两种 ZFP 表达载体转染至所述细胞中。

[0086] 46. 实施方案 45 所述的方法, 其中至少两种 ZFP 表达载体共同转染至所述细胞中。

[0087] 47. 实施方案 45 所述的方法, 其中至少两种 ZFP 表达载体先后依次转染至所述细胞中。

[0088] 48. 实施方案 42 至 47 任一所述的方法, 其中所述植物细胞中所有的 EPSPS 共生同源基因被剪切。

[0089] 49. 实施方案 42 至 47 任一所述的方法, 其中所述植物细胞中一个 EPSPS 共生同源基因被剪切。

[0090] 50. 实施方案 42 至 47 任一所述的方法, 其中所述植物细胞中至少两个 EPSPS 共生同源基因被剪切。

[0091] 51. 一种包含第一和第二 DNA 序列的供体载体；

[0092] 其中第一序列与第三序列同源, 且第二序列与第四序列同源；且

[0093] 其中第三和第四序列是染色体 EPSPS DNA 序列

[0094] 52. 实施方案 51 所述的供体载体, 其中第三和第四序列的近边是毗连的。

[0095] 53. 实施方案 51 所述的供体载体, 其中第三和第四序列的近边被至少一对核苷酸分开。

[0096] 54. 实施方案 51 至 53 中任一所述的载体, 其中所述染色体 EPSPS DNA 序列属于包含选自下组的核苷酸序列的 EPSPS 基因: SEQ ID NOS: 10-14 或与其至少有 95% 同一性的序列。

[0097] 55. 实施方案 51 至 54 中任一所述的载体, 其中所述第三和第四序列是外源序列。

[0098] 56. 实施方案 51 至 54 中任一所述的载体, 其中所述第三和第四序列是内源序列。

[0099] 57. 实施方案 51 至 56 中任一所述的载体, 其中所述第一或第二序列至少之一长度为 100 核苷酸或更短。

[0100] 58. 实施方案 51 至 57 中任一所述的载体, 进一步包括第五序列, 其中所述第五序列

[0101] (a) 介于 (interpose) 第一和第二序列之间；且

[0102] (b) 为外源核酸序列。

[0103] 59. 实施方案 58 所述的载体, 其中所述第五序列大小至少为 1 碱基对。

[0104] 60. 实施方案 58 或 59 任一所述的载体, 其中所述第五序列包括编码选择性标记的序列。

[0105] 61. 实施方案 61 所述的载体, 其中所述选择性标记选自下组: 绿色荧光蛋白

(GFP)、 $\beta$ -葡糖醛酸糖苷酶 (GUS)、phosphinothricin N-乙酰转移酶 (PAT, BAR)、新霉素磷酸转移酶、潮霉素磷酸转移酶、 $\beta$ -内酰胺酶、儿茶酚二氧合酶、 $\alpha$ -淀粉酶、酪氨酸酶、 $\beta$ -半乳糖苷酶、荧光素酶、水母发光蛋白、EPSP 合酶、腈水解酶、乙酰乳糖合酶 (ALS)、二氢叶酸还原酶 (DHFR)、茅草枯脱卤素酶及邻氨基苯甲酸合酶。

[0106] 62. 实施方案 58 所述的载体,其中所述第五序列包括编码除选择性标记以外蛋白质的序列。

[0107] 63. 实施方案 58 至 62 所述的载体,其中所述第五序列包括一个或更多转录调节序列。

[0108] 64. 实施方案 58 至 63 所述的载体,其中所述第五序列包括一个或更多增强或减弱蛋白质靶向的序列。

[0109] 65. 实施方案 63 所述的载体,其中所述一个或更多转录调节序列增加 EPSPS 表达。

[0110] 66. 实施方案 63 所述的载体,其中所述一个或更多转录调节序列减少 EPSPS 表达。

[0111] 67. 实施方案 58 至 66 所述的载体,其中所述第五序列包括一个或更多编码小干扰 RNA 或微 RNA 或蛋白质一部分的序列

[0112] 68. 实施方案 67 所述的载体,其中所述编码所述蛋白质一部分的序列包括与 EPSPS 染色体序列同源的序列。

[0113] 69. 实施方案 58 所述的载体,其中所述第五序列包括突变体 EPSPS 染色体序列相应的野生型序列。

[0114] 70. 实施方案 58 所述的载体,其中所述第五序列包括野生型 EPSPS 染色体序列相应的突变体序列。

[0115] 71. 实施方案 70 所述的载体,其中所述突变体 EPSPS 染色体序列增加植物对除草剂草甘膦的耐受性。

[0116] 72. 实施方案 51 至 71 所述的载体,其中所述第一序列与第三序列有至少 35% 的同源性。

[0117] 73. 实施方案 51 至 72 所述的载体,其中所述第二序列与第四序列有至少 35% 的同源性。

[0118] 74. 一种将外源核酸序列导入植物细胞基因组的方法,所述方法包括下述步骤:

[0119] (a) 使所述细胞与实施方案 51 至 73 任一所述的供体载体接触;以及

[0120] (b) 在所述细胞中表达锌指核酸酶 (ZFN),其中所述 ZFN 剪切第三或第四序列之一的 3 千碱基对以内的在染色体 DNA 中的 EPSPS 基因 (anEPSPS gene in chromosomal DNA within 3 kilobase pairs of either of the third or fourth sequences)。

[0121] 从而使步骤 (b) 中染色体 DNA 的剪切促进供体载体通过同源重组整合入基因组中。

[0122] 75. 实施方案 74 所述的方法,其中所述 EPSPS 基因包含选自下组的核苷酸序列的 EPSPS 基因:SEQ ID NOS:10-14 或与其至少有 95% 同一性的序列。

[0123] 76. 一种在植物细胞中表达外源核酸序列产物的方法,所述方法包括下述步骤:

[0124] (a) 使所述细胞与实施方案 58 至 73 任一所述的供体载体接触;

[0125] (b) 在所述细胞中表达锌指核酸酶 (ZFN),其中所述 ZFN 剪切第三或第四序列之一的 3 千碱基对以内的在染色体 DNA 中的 EPSPS 基因。

[0126] 从而使步骤 (b) 中染色体 DNA 的剪切导致供体载体通过同源重组整合入基因组中以及所述外源核酸序列的表达。

[0127] 77. 实施方案 76 所述的方法, 其中所述 EPSPS 基因包含选自下组的核苷酸序列 EPSPS 基因: SEQ ID NOS: 10-14 或与其至少有 95% 同一性的序列。

[0128] 78. 根据实施方案 74 或 75 任一所述的方法获得的转基因植物细胞。

[0129] 79. 一种含有实施方案 78 所述转基因植物细胞的植物。

[0130] 80. 一种在植物细胞基因组中进行分子内同源重组的方法, 所述方法包括下述步骤:

[0131] (a) 提供包含 EPSPS 基因并进一步包含与第二序列同源的第一序列的 DNA 片段;

[0132] (b) 使所述 DNA 片段与实施方案 14 至 28 任一所述的 ZFN 接触, 其中 ZFN 在 EPSPS 基因序列处剪切所述 DNA 片段, 从而促进分子内同源重组。

[0133] 81. 实施方案 80 所述的方法, 其中所述 DNA 片段对细胞是内源的。

[0134] 82. 实施方案 80 或 81 所述的方法, 其中所述同源重组发生在染色体上。

[0135] 83. 实施方案 82 所述的方法, 其中所述第一和第二序列间的 DNA 从染色体缺失。

[0136] 84. 实施方案 80 至 83 任一所述的方法, 其中所述 EPSPS 基因在基因组中为特有的。

[0137] 85. 实施方案 80 至 83 任一所述的方法, 其中一个或更多所述 EPSPS 基因的共生同源基因存在于所述基因组中。

[0138] 86. 实施方案 80 至 85 任一所述的方法, 其中所述 ZFN 包括一对融合蛋白, 其中每个融合蛋白都是由 IIS 型限制性内切酶剪切域与改造的锌指结合区域融合而成的。

[0139] 87. 实施方案 80 至 86 任一所述的方法, 其中所述第二序列是自第一序列距离的至少 100 碱基对 (the second sequence is at least 100 base pairs from the first sequence)。

[0140] 88. 实施方案 80 至 86 任一所述的方法, 其中所述 EPSPS 基因序列是自 (from) 所述第一或第二序列的至少 100 碱基对。

[0141] 89. 实施方案 80 至 86 任一所述的方法, 其中所述 EPSPS 基因序列位于所述第一和第二序列之间。

[0142] 90. 实施方案 80 至 89 任一所述的方法, 其中所述第一或第二序列之一对生物体为外源的。

[0143] 91. 实施方案 80 至 90 任一所述的方法, 其中所述第一和第二序列两者对于生物体均为外源的。

[0144] 92. 实施方案 83 所述的方法, 其中所述从染色体缺失的序列编码 EPSPS 基因的全部或一部分。

[0145] 93. 实施方案 83 所述的方法, 其中所述从染色体缺失的序列编码选择性标记的全部或一部分。

[0146] 94. 实施方案 93 所述的方法, 其中所述选择性标记选自下组: 绿色荧光蛋白 (GFP)、 $\beta$ -葡萄糖醛酸糖苷酶 (GUS)、phosphinothricin N-乙酰转移酶 (PAT, BAR)、新霉素磷酸转移酶、潮霉素磷酸转移酶、 $\beta$ -内酰胺酶、儿茶酚二氧合酶、 $\alpha$ -淀粉酶、酪氨酸酶、 $\beta$ -半乳糖苷酶、荧光素酶、水母发光蛋白、EPSP 合酶、胍水解酶、乙酰乳糖合酶 (ALS)、二氢

叶酸还原酶 (DHFR)、茅草枯脱卤素酶及邻氨基苯甲酸合酶。

[0147] 95. 实施方案 83 所述的方法,其中所述缺失的 DNA 被外源序列取代,所述方法进一步包括:

[0148] 将多核苷酸导入所述细胞,其中所述多核苷酸包括:

[0149] (a) 第四和第五序列,其中所述第四序列与靠近第一序列的非缺失序列同源,且第五序列与靠近第二序列的非缺失序列同源;以及

[0150] (b) 所述外源序列。

[0151] 96. 实施方案 95 所述的方法,其中所述外源序列是选择性标记。

[0152] 97. 实施方案 96 所述的方法,其中所述选择性标记选自下组:绿色荧光蛋白 (GFP)、 $\beta$ -葡糖醛酸糖苷酶 (GUS)、phosphinothricin N-乙酰转移酶 (PAT, BAR)、新霉素磷酸转移酶、潮霉素磷酸转移酶、 $\beta$ -内酰胺酶、儿茶酚二氧合酶、 $\alpha$ -淀粉酶、酪氨酸酶、 $\beta$ -半乳糖苷酶、荧光素酶、水母发光蛋白、EPSP 合酶、腓水解酶、乙酰乳糖合酶 (ALS)、二氢叶酸还原酶 (DHFR)、茅草枯脱卤素酶及邻氨基苯甲酸合酶。

[0153] 98. 实施方案 95 所述的方法,其中所述外源序列为 EPSPS 基因序列。

[0154] 99. 实施方案 98 所述的方法,其中所述 EPSPS 基因序列包含突变。

[0155] 100. 实施方案 99 所述的方法,其中所述突变增加植物对除草剂草甘膦的耐受性。

[0156] 101. 一种从植物基因组中缺失 EPSPS 基因序列的方法,所述方法包括:

[0157] (a) 提供包含 EPSPS 基因序列的植物细胞;以及

[0158] (b) 在所述细胞中表达第一和第二锌指核酸酶 (ZFN),其中第一 ZFN 在第一剪切位点剪切,而第二 ZFN 在第二剪切位点剪切,其中所述 EPSPS 基因序列位于第一剪切位点和第二剪切位点之间,其中第一和第二剪切位点的剪切导致所述 EPSPS 基因序列的缺失。

[0159] 102. 实施方案 101 所述的方法,其中通过所述第一和第二剪切位点的非同源末端联接缺失所述 EPSPS 基因序列。

[0160] 103. 实施方案 101 或 102 所述的方法,其中所述第一和第二剪切位点间隔至少 100 核苷酸。

[0161] 104. 实施方案 101 或 102 所述的方法,其中所述植物细胞是转基因植物细胞。

[0162] 105. 实施方案 101 或 102 所述的方法,其中所述 EPSPS 基因是外源序列。

[0163] 106. 实施方案 101 或 102 所述的方法,其中所述 EPSPS 基因是内源序列。

[0164] 107. 一种调整 EPSPS 基因调节的方法,所述方法包括:

[0165] (a) 提供含有 EPSPS 基因序列的植物细胞;以及

[0166] (b) 在所述细胞中表达 ZFP,其中所述 ZFP 与所述 EPSPS 基因中的靶位点结合,从而调整所述 EPSPS 基因的调节。

[0167] 108. 实施方案 107 所述的方法,其中所述靶位点是所述 EPSPS 基因的调节序列。

[0168] 109. 实施方案 107 所述的方法,其中所述靶位点位于所述 EPSPS 基因转录起始位点的上游。

[0169] 110. 实施方案 107 所述的方法,其中所述靶位点邻近所述 EPSPS 基因的转录起始位点。

[0170] 111. 实施方案 107 所述的方法,其中所述靶位点位于所述 EPSPS 基因转录起始位点的下游。

- [0171] 112. 实施方案 107 所述的方法,其中所述 ZFP 增加 EPSPS 基因的转录。
- [0172] 113. 实施方案 112 所述的方法,其中所述 ZFP 增加植物对除草剂草甘膦的耐受性。
- [0173] 114. 实施方案 107 所述的方法,其中所述 ZFP 减少所述 EPSPS 基因的转录。
- [0174] 115. 实施方案 114 所述的方法,其中所述 ZFP 减少植物对除草剂草甘膦的耐受性。
- [0175] 鉴于本文的公开,本领域一般技术人员可以容易的想到本发明的这些和其它实施方案。

[0176] 附图简述

[0177] 图 1 显示 Southern Blot,提供了欧洲油菜 (*B. Napus*) 变种 Nex710、芜菁 (*B. rapa*) 与甘蓝 (*B. oleracea*) 基因组中 EPSPS 基因数的估计。(标准标记是 Promega 的分析 DNA 宽量程标记)

[0178] 图 2A-IE 显示产生 ZFN 表达构建体所用克隆策略的概图表示。使用分步克隆策略:将 ZFN 编码基因各自克隆入载体 pVAX-C2A-NLSop2-EGFP-FokMono(图 2A) 与 pVAX-N2A-NLSop2-EGFP-FokMono(图 2B) 以构建双重蛋白质盒(图 2C)。将该盒与 pDAB3731 连接以产生最终质粒(图 2D) 以表达 ZFN 异源双聚体。随即将所述 ZFN 盒使用 Gateway 技术转移入二元载体以构建土壤杆菌 (*Agrobacterium*)-介导欧洲油菜转化构建体(图 2E)。ZFN d2 = 10654-CH3-v2;ZFN rb2 = 10657-CH3-v2。

[0179] 图 3A-D 显示 EPSPS 的共生同源基因特异性扩增。图 3A-3D 分别表示共生同源基因 A、B、C 与 D 特异性 PCR 分析。1-6 泳道包含下列 DNA。泳道 1:无 DNA PCR 对照;泳道 2:欧洲油菜变种 Nex710 DNA (10ng/ $\mu$ l);泳道 3、4、5 与 6 包含共生同源基因基因 D、C、B 与 A (1570bp) 以 1ng/ $\mu$ l 的浓度作为阳性对照的扩增 DNA。PCR 产物在 2% E-GEL96 (Invitrogen, Carlsbad, CA) 上泳动,使用 GEL DPC 2000 胶文档编制系统 (Bio-Rad, Hercules, CA) 显影。捕捉图像并使用 QUANTITY ONE 软件 (Bio-Rad, Hercules, CA) 分析,并进一步使用 E-EDITOR 软件 (Invitrogen, Carlsbad, CA) 处理。片断大小以碱基对 (bp) 表示。

[0180] 图 4 显示 ZFN 对欧洲油菜 EPSPS 共生同源基因的结合与剪切位点。需要两个 ZFN 蛋白质以实施 DNA 的双链 (ds) 剪切。在剪切位点上游,以向下箭头标示,一个蛋白质 (10567 或 10658) 与下划线所示的核苷酸相联,而另一个蛋白质 (10654) 结合下划线所示序列的下游。只有当两个蛋白质均与其相应的位点结合时,剪切才发生。在 5 个共生同源基因中有一对在一个或两个 ZFN 结合位点由轻微的序列差异(下划线)提供了序列特异性,并导致在所述共生同源基因中选择性的双链剪切。

[0181] 图 5 显示在 EPSPS 共生同源基因 D 中的 ZFN 介导缺失。在 *B. napus* 中,在 EPSPS 共生同源基因 D 中存在的 ZFN 介导双链 DNA 断裂经非同源末端联接 (NHEJ) 修复导致 2bp 的缺失。所述 ZFN pDAB7151 的剪切靶为 CAGTT,这对应于所述 2bp 的 GT 缺失。底图:预期的野生型序列;顶图:使用 SEQUENCHER 软件对 26 个共生同源基因 D 克隆进行序列比对,显示所述 2bp 缺失。这些序列由 13 个克隆的正向与反向引物测序获取。

[0182] 图 6 显示在 *B. napus* EPSPS 共生同源基因 D 中 ZFN-介导双链断裂导致的 NHEJ。显示多个 NHEJ 缺失相对于野生型 DNA 中预期的剪切位点(顶)的序列比对。右手侧括号中的数字显示见于序列比对中相同的分子数。

[0183] 图 7A-7B 显示在 *B. napus* EPSPS 共生同源基因 C 和 D 中 ZFN-介导双链断裂导致的 NHEJ。显示多个 NHEJ 缺失相对于野生型与经处理的试样中预期的剪切位点(顶)的序



列比对。试样号对应表 5 所示的号码。

[0184] 图 8A-8C 显示在 *B. napus* EPSPS 共生同源基因 A 和 B 中 ZFN- 介导双链断裂导致的 NHEJ。显示多个 NHEJ 缺失相对于野生型与经处理的试样中预期的剪切位点（顶）的序列比对。试样号对应表 6 所示的号码。

[0185] 图 9 (SEQ ID NO :10) 显示 *B. napus* EPSPS 共生同源基因 A 的序列。

[0186] 图 10 (SEQ ID NO :11) 显示 *B. napus* EPSPS 共生同源基因 B 的序列。

[0187] 图 11A-11B (SEQ ID NO :12) 显示 *B. napus* EPSPS 共生同源基因 C 的序列。

[0188] 图 12A-12B (SEQ ID NO :13) 显示 *B. napus* EPSPS 共生同源基因 D 的序列。

[0189] 图 13A-13B (SEQ ID NO :14) 显示 *B. napus* EPSPS 共生同源基因 E 的序列。

[0190] 图 14, 小图 A 至 E 描述肾 293 报道细胞中示例性 EPSPS ZFN (亦参见实施例 3) 的基因修正活性。图 14A 显示使用 ZFN 对 10654 与 10658 的基因修正活性。图 14B 显示使用 ZFN 对 10654 与 10658 的基因修正活性。图 14C 显示使用 ZFN 对 9875 与 10275 的基因修正活性。图 14D 显示使用 ZFN 对 10740/10741 与 10749/10742 的基因修正活性。图 14E 显示使用每个直方 (bar) 下显示的 ZFN 对的基因修正活性。

[0191] 详细描述

[0192] 本文公开了适用于调节表达及靶向剪切与改变植物基因, 尤其是植物共生同源基因的组合物与方法。共生同源基因调节的可通过, 例如使用改造 ZF 转录因子或修饰基因调节区域的方法来调节。基因可通过, 例如靶向剪切继以染色体同源重组或靶向剪切继以外源多核苷酸 (包含一个或更多与基因核苷酸序列同源的区域) 与基因组序列同源重组的方法来改变。共生同源基因一个非限定性的例子是 EPSPS 基因。

[0193] 基因组序列包括存在于染色体、附加体 (游离基因)、细胞器基因组 (例如, 线粒体, 叶绿体)、人工染色体以及任何其它类型存在于细胞中的核酸, 例如扩增序列、双微染色体及内源或感染的细菌与病毒的基因组中的序列。基因组序列可以是正常的 (亦即野生型) 或突变体; 突变体序列可能包括, 例如插入、删除、易位、重排和 / 或点突变。基因组序列亦可包括多个不同等位基因中的一个。

[0194] 本文公开的组合物包括一个或更多 ZFP, 所述 ZFP 包括改造的锌指结合区域、编码这些多肽的多核苷酸、以及 ZFP 与编码 ZFP 的多核苷酸的组合。锌指结合区域可包括一个或更多锌指 (例如, 2、3、4、5、6、7、8、9 或更多锌指), 且可改造以与任何 EPSPS 基因组序列结合。

[0195] 本文描述的 ZFP 通过激活或阻抑基因转录可用于调节 EPSPS 基因表达。可构建包括锌指结构域连接到调节区的融合体的 ZFP 以产生激活或阻抑基因转录的嵌合转录因子。通过将锌指结构域与核酸酶剪切域 (或剪切半域) 连接以产生锌指核酸酶, ZFP 亦可用于靶向剪切目标 EPSPS 基因组区域。因此, 通过识别目标 EPSPS 基因组区域, 且在该区域期望基因调节、剪切或重组, 根据本文公开的方法, 可构建包括一个或多个融合蛋白的锌指蛋白, 所述融合蛋白包括一个或更多与经改造以在该基因组区域识别 EPSPS 基因区域的锌指结构域连接的调节区和 / 或剪切域 (或剪切半域)。细胞中所述包含融合蛋白 (或多个融合蛋白) 的 ZFP 的存在会导致所述融合蛋白结合到其结合位点, 或在所述基因组区域内或附近剪切以及调节的改变。另外, 如果 EPSPS 基因组区域被剪切且与该 EPSPS 基因组区域同源的外源多核苷酸亦存在于该细胞中, 高频度的同源重组会发生在所述 EPSPS 基因组区

域与所述外源多核苷酸之间。

[0196] 植物细胞可包含一个或更多同源或共生同源 EPSPS 基因序列,其中数个或全部可作为本文公开方法进行修饰的靶。因此,本文描述的组合物可能在植物细胞中靶向一个或更多 EPSPS 基因,例如,1、2、3、4、5 或直至任何数量或所有存在植物细胞中的 EPSPS 共生同源基因。一些 ZFP 可能在植物细胞中特异性的与一个特定的 EPSPS 共生同源基因结合。其它 ZFP 可能在植物细胞中与多个 EPSPS 共生同源基因结合。因此,一个或更多 ZFP 或编码具有不同特性 ZFP 的表达载体可以组合以在植物细胞中靶向期望的目标 EPSPS 基因。

[0197] 总述

[0198] 除非指明为其它,所述方法的实行以及本文公开的组合物的制备与用途使用分子生物学、生物化学、染色质结构与分析、计算化学、细胞培养、重组 DNA 与相关领域的常规方法,只要它们属于本领域的常规技术。这些技术在文献中完全得到阐明。例如,参见 Sambrook et al. MOLECULAR CLONING :A LABORATORY MANUAL, Second edition, Cold Spring Harbor Laboratory Press, 1989 and Third edition, 2001; Ausubel et al., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, 1987 and periodic updates; the series METHODS IN ENZYMOLOGY, Academic Press, San Diego; Wolffe, CHROMATIN STRUCTURE AND FUNCTION, Third edition, Academic Press, San Diego, 1998; METHODS IN ENZYMOLOGY, Vol. 304, "Chromatin" (P. M. Wassarman and A. P. Wolffe, eds.), Academic Press, San Diego, 1999; METHODS IN MOLECULAR BIOLOGY, Vol. 119, "Chromatin Protocols" (P. B. Becker, ed.) Humana Press, Totowa, 1999。

[0199] 定义

[0200] 术语“核酸”、“多核苷酸”与“寡聚核苷酸”可互换使用,指脱氧核糖核酸或核糖核酸多聚体,具有链状或环状构造,并为单链或双链形式。在本公开中使用时,这些术语不应解释为针对多聚体长度作出限定。该术语可包容天然核酸的已知类似物,以及在其碱基、糖和 / 或磷酸部分(例如,磷硫酰骨架)有修饰的核苷酸。一般而言,特定核苷酸的类似物具有相同的碱基配对特异性;亦即,A 的类似物将与 T 碱基配对。

[0201] 术语“多肽”、“肽”与“蛋白质”可互换使用,指氨基酸残基的多聚体。该术语亦适用于其一个或更多氨基酸是相应的天然存在氨基酸的化学类似物或修饰衍生物的氨基酸多聚体。

[0202] “结合”指序列特异的,非共价的大分子间相互作用(例如,蛋白质与核酸间)。在结合相互作用中,并不需要所有部分都是序列特异的(例如,与 DNA 骨架中磷酸残基的接触),只要作为整体该相互作用是序列特异的即可。上述相互作用通常以解离常数 ( $K_d$ ) 为  $10^{-6}M^{-1}$  或更小为特征。“亲和力”指所述结合的强度:增加结合亲和力对应  $K_d$  减少。

[0203] “结合蛋白”是具有与另一个分子非共价结合能力的蛋白质。结合蛋白可与,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)结合。对于蛋白质结合蛋白,该蛋白质可与自身(以形成同源二聚体,同源三聚体等等)和 / 或一个或更多不同的蛋白质分子结合。结合蛋白可具有多于一种类型的结合活性。举例而言,锌指蛋白具有 DNA 结合、RNA 结合与蛋白质结合活性。

[0204] “锌指 DNA 结合蛋白”(或结合区域)是通过一个或更多锌指,以序列特异性为特征与 DNA 结合的蛋白质,或更大蛋白质中的域,所述锌指是在结合区域中的氨基酸序列区域,

其结构通过锌离子的配位作用稳定化。术语锌指 DNA 结合蛋白常常缩写为锌指蛋白或 ZFP。

[0205] 锌指结合区域可经“改造”以与预先确定的核苷酸序列（例如，EPSPS 基因序列）结合。改造锌指蛋白方法的非限定性示例包括设计及选择。设计的锌指蛋白是非天然存在的蛋白质，其设计 / 组成主要来自理性的标准。设计的理性标准包括运用取代法则及计算机算法以处理存储已有的 ZFP 设计与结合数据信息的数据组中的信息。举例而言，参见 U. S. Patents 6, 140, 081 ;6, 453, 242 ;6, 534, 261 ;以及 6, 785, 613 ;亦参见 WO 98/53058 ;WO 98/53059 ;WO 98/53060 ;WO 02/016536 与 WO 03/016496 ;以及 U. S. Patents 6, 746, 838 ;6, 866, 997 与 7, 030, 215。因此，“改造”锌指蛋白或“非天然”锌指蛋白为其一个或多个锌指 DNA 结合区域组分（识别螺旋）为非天然的，且经改造以与事先选定的靶位点结合的蛋白质。

[0206] “选择的”锌指蛋白是非见于天然的蛋白质，其产生主要来自例如噬菌体展示、相互作用陷阱 (trap) 或杂交选择的经验性过程。举例而言，参见 US 5, 789, 538 ;US 5, 925, 523 ;US 6, 007, 988 ;US 6, 013, 453 ;US 6, 200, 759 ;US 6, 733, 970 ;US RE39, 229 ;以及 WO 95/19431 ;WO 96/06166 ;WO 98/53057 ;WO 98/54311 ;WO 00/27878 ;WO 01/60970WO 01/88197 与 WO 02/099084。

[0207] 术语“序列”指任何长度的核苷酸序列，可以是 DNA 或 RNA ;可以是链状、环状或分枝的，且可以是单链或双链的。术语“供体序列”指插入基因组的核苷酸序列。供体序列可以是任何长度，例如长度在 2 至 25000 核苷酸之间（或任何之间或之上的整数），优选长度在 100 至 5000 核苷酸之间（或任何之间或之上的整数），较优选长度大约在 200 至 2500 核苷酸之间。

[0208] “同源序列”指与第二序列具有部分序列同一性的第一序列，且其序列可以为与第二序列完全相同。“同源非相同序列”指与第二序列具有部分序列同一性的第一序列，但其序列与第二序列并不完全相同。举例而言，包含突变基因野生型序列的多核苷酸与该突变基因同源且不完全相同。在一些实施方案中，所述两序列间的同源程度足以使利用正常细胞机理使两者之间同源重组成为可能。两个同源非完全相同序列可以为任何长度，且其非同源程度可以小到一个核苷酸（例如，利用靶向同源重组校正基因组点突变的情况），或大到 10 或更多千碱基对（例如，在染色体中预先确定的位点插入基因的情况）。包含同源非相同序列的两个多肽无需长度相同。举例而言，可使用 20 至 10000 核苷酸或核苷酸对的外源多核苷酸（亦即，供体多核苷酸）。

[0209] 确定核酸及蛋白质序列同一性的技术在本领域是已知的。通常，上述技术包含确定基因 mRNA 的核苷酸序列和 / 或确定由其编码的氨基酸序列，并用这些序列与第二个核苷酸或氨基酸序列比较。亦可用此方式确定并比较基因组序列。一般而言，相同（同一性）指两个多核苷酸或多肽中各自准确的核苷酸 - 核苷酸或氨基酸 - 氨基酸对应。两个或更多序列（多核苷酸或氨基酸）可通过确定其同一性百分比相比较。不管是核酸还是氨基酸序列，两个序列的同一性百分比，为两个序列比对的序列间准确匹配数除以较短序列的长度再乘以 100。一个近似的核酸序列序列比对由 Smith and Waterman, Advances in Applied Mathematics 2:482-489(1981) 的局部同源算法提供。通过使用 Dayhoff, Atlas of Protein Sequences and Structure, M. O. Dayhoff ed., 5 suppl. 3:353-358, National Biomedical Research Foundation, Washington, D. C., USA 研制以及 Gribskov,

Nucl. Acids Res. 14(6) :6745-6763(1986) 正态化的评分矩阵, 该算法可用于氨基酸序列。用该算法确定序列同一性度的示例性实施由 Genetics Computer Group (Madison, WI) 提供于“BestFit”应用程序中。该方法的缺省参数描述于 Wisconsin Sequence Analysis Package Program Manual, Version 8(1995) (可自 Genetics Computer Group, Madison, WI 得到) 中。在本发明背景下确立相同率的优选方法是使用爱丁堡大学拥有版权, 由 John F. Collins 与 Shane S. Sturrok 研制, 以及 IntelliGenetics, Inc. (Mountain View, CA) 发布的 MPSRCH 程序集。在该套程序集中, 当缺省参数用于评分表时 (例如, 开放间隔罚分 (gap open penalty) 12, (gap extension penalty) 延伸间隔罚分 1, (gap) 间隔罚分 6), 可以使用 Smith-Waterman 算法。根据所得的数据, “匹配”值反映了序列同一性程度。其它适合用来计算序列间同一性百分比或相似度的程序在本领域通常是公知的, 例如, 另一个序列匹配程序是 BLAST, 使用缺省数据。例如, BLASTN 和 BLASTP 可使用下述缺省参数: 遗传密码 (geneticcode) = 标准 (standard); 过滤 (filter) = 无 (none); 链 (strand) = 双 (both); 截断值 (cutoff) = 60; 期望值 (expect) = 10; 模型 (Matrix) = BLOSUM62; 描述 (Descriptions) = 50 序列; 排列 (sort by) = 高分 (high score); 数据库 (Databases) = 非冗余 (non-redundant), Genbank+EMBL+DDBJ+PDB+GenBank CDStranslations+Swiss protein+Spupdate+PIR。这些程序的细节可以在互联网上找到。考虑本文描述的序列, 所期望的序列同一性度大约为 35% 至 100% 以及任何之间的整数值。典型的, 序列间同一性百分比为至少 35% -40%; 40% -45%; 45% -50%; 50% -60%; 60% -70%; 70-75%, 优选为 80-82%, 较优选为 85-90%, 更优选为 92%, 进一步较优选为 95%, 最优选为 98% 序列同一性。

[0210] 或者, 多核苷酸之间的序列相似性可通过下述方法确定: 在允许同源区域间形成稳定双联体的条件下进行多核苷酸杂交, 继以用单链特异性核酸酶消化, 并确定消化后片段的大小。两个核酸, 或两个多肽序列在下述情况下为基本上相互同源: 所述序列在确定长度分子上, 显示如上述方法确定的至少大约 70-75%, 优选 80% -82%, 较优选 85% -90%, 更优选 92%, 进一步更优选 95%, 最优选 98% 序列同一性。在此所用的基本上同源亦指与特定 DNA 或多肽序列显示为完全同一性的序列。亦可用 Southern 杂交实验辨识 DNA 序列基本上同源, 所述实验是在根据该系统定义的严格条件下进行的。确定合适的杂交条件属于本领域公知技术。例如, 参见 Sambrook et al., supra; Nucleic Acid Hybridization: A Practical Approach, editors B. D. Hames and S. J. Higgins, (1985) Oxford; Washington, DC; IRL Press)。

[0211] 两个核酸片段的择性杂交可以通过下述确定。两个核酸分子间的序列同一性 (identical) 度影响上述分子间杂交反应的效率和强度。部分相同的核酸序列将至少部分抑制完全相同序列对靶分子的杂交。对完全相同序列杂交的抑制可用本领域众所周知的杂交分析 (例如, Southern (DNA) blot、Northern (RNA) blot、溶液杂交, 或类似, 参见 Sambrook, et al., Molecular Cloning: A Laboratory Manual, Second Edition, (1989) Cold Spring Harbor, N. Y.) 来评估。上述分析可使用不同程度的择性进行, 例如, 使用从低到高严格程度不同的条件。如果采用低严格度条件, 可以使用甚至缺乏部分序列同一性 (例如, 与靶分子具有少于 30% 序列同一性的探针) 的第二探针评估非特异性结合不存在, 从而, 在不存在非特异性结合反应时, 第二探针将不会与靶杂交。

[0212] 当使用基于杂交的检测系统时,选择与参考核酸序列互补的核酸探针,并随即通过选择合适的条件,所述探针与所述参考序列选择性的相互杂交或结合以形成双链分子。能够与参照序列在中等严格杂交条件下选择性的与参照序列杂交的核酸分子通常在下述条件下杂交:即允许检测至少约为 10-14 核苷酸长度,具有与所选核酸探针序列有超过约 70% 的序列同一性的靶核酸序列。严格杂交条件通常允许检测至少约为 10-14 核苷酸长度,具有与所选核酸探针序列有超过约 90-95% 的序列同一性的靶核酸序列。有用于探针 / 参考序列杂交的杂交条件,其中所述探针与所述参考序列具有特定程度的序列同一性,可如本领域已知方法确定(参见,例如, Nucleic Acid Hybridization: A Practical Approach, editors B. D. Hames and S. J. Higgins, (1985) Oxford; Washington, DC; IRL Press)。

[0213] 杂交条件对本领域技术人员为众所周知的。杂交严格度指杂交条件不利于形成含有错配核苷酸杂交体的程度,严格度越高,则对错配杂交体的容忍度越低。影响杂交严格度的因素对本领域技术人员为众所周知的,且包括,但不仅限于:温度, pH, 离子强度, 有机溶剂, 例如甲酰胺与二甲亚砜的浓度。如本领域技术人员所知,杂交严格度随温度升高,离子强度降低与溶剂浓度降低而增加。

[0214] 就杂交严格条件而论,在本领域众所周知可使用很多等价条件通过,例如,改变下列因素以确立特定严格度:所述序列的长度与本性,不同序列的碱基构成,盐与其它杂交溶液组分浓度,是否在所述杂交溶液中存在封闭剂(例如,硫酸葡聚糖,与聚乙二醇),杂交反应温度与时间参数,以及,改变洗涤条件。选择特定组的杂交条件选自本领域的下列标准方法(参见,例如, Sambrook, et al., Molecular Cloning: A Laboratory Manual, Second Edition, (1989) Cold Spring Harbor, N. Y.)。

[0215] “重组”指在两个多核苷酸间交换遗传信息的过程。就本公开考虑,“同源重组(HR)”指上述交换的下述特定形式,即,例如,发生于细胞中的双链断裂修复。该过程需要核苷酸序列的同源性,使用“供体”分子作为模板修复“靶”分子(亦即,经受双链断裂的分子),该过程以“非交换基因转换(non-crossover gene conversion)”或“短路基因转换(short track gene conversion)”等不同名称为人所知,因其导致遗传信息从所述供体转移至所述靶。虽不愿拘于任何特定理论,但该转移可涉及在断裂靶与所述供体间形成的异源双链 DNA 的错配校正,和 / 或“合成依存性链退火(synthesis-dependent strand annealing)”,其中所述供体用于重新合成遗传信息,该遗传信息将成为靶的一部分,和 / 或相关过程。上述特定 HR 常导致所述靶分子序列的改变,从而使所述供体多核苷酸部分或全部序列掺入所述靶多核苷酸中。

[0216] “剪切”指 DNA 分子共价骨架的断裂。剪切可通过多种方法起始,包括但不限于,磷酸二酯键的酶或化学水解。单链剪切与双链剪切均为可能,且双链剪切可作为两个分别的单链剪切事件的结果发生。DNA 剪切可导致平端或交错末端的产生。在一些实施例中,融合多肽用于靶向双链 DNA 剪切。

[0217] “剪切域”包括一个或更多具有 DNA 剪切催化活性的多肽序列。剪切域可包含于单独的多肽链中,或剪切活性可来源于两个(或更多)多肽的联合。

[0218] “剪切半域”为下述的多肽序列,即其与第二多肽(可为相同或相异的)结合形成具有剪切活性的复合物(优选为双链剪切活性)。

[0219] “染色质”为包含细胞基因组的核蛋白结构。细胞染色质包括核酸,主要为 DNA,与蛋白质,包括组蛋白与非组蛋白染色体蛋白。真核细胞染色质绝大多数以核小体形式存在,其中核小体 (nucleosome) 核心包括大约 150 碱基对,与包含组蛋白 H2A, H2B, H3 与 H4 各两个的八聚体 (octamer) 相联系的 DNA;以及在核小体核心间延伸的接头 DNA (根据生物体不同,可为不同长度)。组蛋白 H1 分子通常与所述接头 DNA 相联系。就现行公开考虑,术语“染色质”意为包含所有类型的细胞核蛋白,原核与真核两者皆可。细胞染色质 (chromatin) 包括染色体与附加体染色质。

[0220] “染色体”为包含细胞全部或部分基因组的染色质复合物。细胞基因组经常以染色体组型为特征,所述染色体组型包含细胞基因组的所有染色体集合。细胞基因组可包括一个或更多染色体。

[0221] “附加体”为可复制的核酸,核蛋白复合物或其它包含核酸的结构,且其不为细胞染色体的染色体组型 (karyotype) 一部分。附加体的示例包括质粒与一些病毒基因组。

[0222] “可及区域 (accessible region)”为细胞染色质上的位点,所述细胞染色质中存在于该核酸中的靶位点可由识别该靶位点的外源分子结合。虽不愿拘于任何特定理论,但相信可及区域为未包装入核小体结构内的区域。可及区域的迥异结构常可通过其对化学与酶探针,例如核酸酶的敏感度进行检测。

[0223] “靶位点”或“靶序列”为限定核酸一部分的核酸序列,且若足以允许结合的条件存在,结合分子将结合其上。举例而言,序列 5'-GAATTC-3' 为 EcoRI 限制性内切酶的靶位点。

[0224] “外源”分子为非正常存在于细胞中,但可通过一种或更多遗传学,生物化学或其它方法导入细胞的分子。“正常存在于细胞”为根据细胞特定发育阶段与环境条件所确定。因此,例如,仅在肌肉胚胎发育中存在的分子就成年肌肉细胞而论为外源分子。类似的,由热激诱导的分子就非经热激的细胞而论为外源分子。外源分子可包括,例如,功能障碍性内源分子的功能性版本,或正常功能性内源分子的功能障碍性版本。

[0225] 外源分子可为,除其它外,小分子,例如,通过组合化学过程生成的,或大分子,例如蛋白质,核酸,碳水化合物,脂肪,糖蛋白,脂蛋白,多糖,任何上述分子的修饰衍生物,或任何含有一种或更多上述分子的组合物。核酸包括 DNA 与 RNA,可为单链或双链;可为线性,分支的或环状;且可为任何长度。核酸包括那些能够形成双链,以及可形成三链的核酸。参见,例如,U. S. Patent Nos. 5, 176, 996 与 5, 422, 251。蛋白质包括,但不仅限于,DNA 结合蛋白,转录因子,染色质重建因子,甲基化 DNA 结合蛋白,多聚酶,甲基酶,脱甲基酶,乙酰酶,脱乙酰酶,激酶,磷酸酶,整合酶,重组酶,连接酶,拓扑异构酶,促旋酶 (gyrase) 与解旋酶。

[0226] 外源分子可为与内源分子为相同类型的分子,例如,外源蛋白质或核酸。举例而言,外源核酸可包括感染性的病毒基因组,根癌土壤杆菌 (*Agrobacterium tumefaciens*) T- 链,导入细胞的质粒或附加体,或正常情况下不存在于细胞中的染色体。将外源分子导入细胞的方法对本领域技术人员为已知的,且包括但不限于,脂肪介导转移 (亦即,脂质体,包括中性与阳离子性脂肪),电穿孔,直接注射,细胞融合,颗粒轰击,磷酸钙共沉淀,DEAE 葡聚糖介导转移与病毒载体介导转移。外源分子包括非植物分子,例如,哺乳类 (例如,人类或人源化) 抗体。

[0227] 相反,“内源”分子为正常情况下在特定发育阶段中在特定环境条件下存在于特定细胞中的分子。举例而言,内源核酸可包括染色体,线粒体,质体或其它细胞器的基因组,或

天然存在的附加体核酸。除此之外内源分子可包括蛋白质,例如,转录因子与酶。

[0228] “融合”分子为下述分子,即其中两个或更多亚基分子相连接,优选为共价连接。所述亚基分子可为同样化学类型分子,或可为不同化学类型分子。第一种类型的融合分子的示例包括但不限于,融合蛋白(例如,包含 ZFP DNA 结合区域与剪切域融合的 ZFN)与融合核酸(例如,编码上述融合蛋白的核酸)。第二类型融合分子的示例包括但不限于,形成三链的核酸与多肽间的融合,以及小沟结合物质与核酸间的融合。

[0229] 在细胞中融合蛋白的表达可来自将所述融合蛋白递送到所述细胞中,或将编码所述融合蛋白的多核苷酸递送到细胞中,并在其中转录所述多核苷酸,翻译所得转录物以产生融合蛋白。在细胞中表达蛋白质亦可涉及反式剪接,多肽剪切与多肽连接。将多核苷酸与多肽递送至细胞的方法在本公开其它部分表示。

[0230] “基因”,对现行公开目的而言,包括编码基因产物(参见下文)的 DNA,以及所有调节基因产物产生的 DNA 区域,无论上述调节序列是否邻近编码和/或转录序列。相应的,基因包括但不限于,启动子序列,终止子,翻译调节序列例如核糖体结合位点与内核糖体进入位点,增强子,沉默基因,绝缘子(insulator),边界元件,复制起点,基质连接位点与基因座控制区。

[0231] “基因表达”指将包含于基因中的信息转换为基因产物。基因产物可为基因的直接转录产物(例如, mRNA, tRNA, rRNA, 反义 RNA, 核酶, shRNA, 微 RNA, 结构 RNA 或任何其它类型的 RNA)或 mRNA 翻译产生的蛋白质。基因产物亦包括通过例如帽化,多腺苷酸化,甲基化与编辑等过程修饰的 RNA,以及通过例如甲基化,乙酰化,磷酸化,遍在蛋白化(ubiquitination), ADP-核糖基化,十四烷基化(myristylation)以及糖基化修饰的蛋白质。

[0232] 基因表达的“调整”指基因活性的改变。表达的调整可包括但不限于,基因激活与基因阻抑。

[0233] “植物”细胞包括但不限于,单子叶植物(monocots)或双子叶植物(dicots)的细胞。单子叶植物的非限定性示例包括谷类植物例如玉米,水稻,大麦,燕麦,小麦,高粱,黑麦,甘蔗,菠萝,洋葱,香蕉与椰子。双子叶植物的非限定性示例包括烟草,番茄,向日葵,棉花,甜菜,马铃薯,生菜,甜瓜,大豆,canola(菜籽)与苜蓿。植物细胞可来自所述植物任何部分和/或植物发育的任何阶段。

[0234] “目标区域”为细胞染色质的任何区域,例如,基因或者在基因内或邻近基因的非编码序列,其中期望其与外源分子结合(例如,EPSPS 目标区域包含在 EPSPS 基因内或邻近其的区域)。结合可以靶向 DNA 剪切和/或靶向重组为目的。目标区域可存在于染色体,附加体,细胞器基因组(例如,线粒体或质体),或例如感染病毒基因组。目标区域可在基因编码区域内,可在转录的非编码区域内,例如,引导序列,尾随序列或内含子,或在非转录区域内,无论是所述编码区域的上游或下游。目标区域可小至 1 对核苷酸或直至长度为 25000 对核苷酸,或任何整数值的核苷酸对。

[0235] 术语“可操作的连接”与“可操作的连接于”可互换用于指将两个或更多组分(例如序列元件)并列,其中所述组分如此排列,使得两个组分均正常运作,并使至少其中一个组分能够介导施加于至少一个其它组分的功能。欲举例说明,则可称转录调节序列,例如启动子与编码序列为可操作的连接,如若所述转录调节序列根据是否存在一个或更多转录调

节因子控制所述编码序列的转录水平。转录调节序列通常与编码序列顺式可操纵的连接，但无需直接与其相邻。举例而言，增强子为可操作的连接于编码序列的转录调节序列，尽管其并不毗连。

[0236] 就融合多肽而论，术语“可操作的连接”可指下述事实，即其每个组分与其它组分相连接时实施同样功能，正如其未如此连接时仍将如此实施一般。举例而言，就 ZFP DNA 结合区域与剪切域融合的融合多肽而论，所述 ZFP DNA 结合区域与所述剪切域为可操作的连接，如若在此融合多肽中，所述 ZFP DNA 结合区域能够与其靶位点和 / 或其结合位点结合，而所述剪切域能够在所述靶位点附近剪切 DNA。

[0237] 蛋白质，多肽或核酸的“功能性片段”为其序列不同于全长蛋白质，多肽或核酸，但其保留与全长蛋白质，多肽或核酸相同功能的蛋白质，多肽或核酸。功能片段可拥有较之相应天然分子更多，更少或相同数量残基，和 / 或可含有一个或更多氨基酸或核苷酸取代。确定核酸功能的方法（例如，编码功能，与其它核酸杂交的能力）在本领域为众所周知的。类似的，确定蛋白质功能的方法为众所周知的。举例而言，多肽的 DNA 结合功能可通过，例如，滤膜结合，电泳迁移率变动或免疫沉淀分析进行确定。DNA 剪切可通过凝胶电泳来分析。参见 Ausubel et al., 见上。蛋白质与另一种蛋白质相互作用的能力可通过下述方法确定：例如，共同免疫沉淀，双杂交分析或者互补分析，可为遗传的亦可为生物化学的。参见，例如 Fields et al. (1989) Nature 340 :245-246 ;U. S. Patent No. 5, 585, 245 与 PCT WO 98/44350。

[0238] 靶位点

[0239] 所述披露的方法与组合物包括含有下述融合蛋白的 ZFP：即所述融合蛋白包括调节区或剪切域（或剪切半域）与锌指结构域，其中所述锌指结构域，通过结合于细胞染色体上的序列（例如，EPSPS 基因靶位点或结合位点），将所述调节区或剪切域（或剪切半域）的活性引导至所述序列附近，并且由此调整转录，或在所述靶序列的附近诱导剪切。如在该披露其它地方所提出的一般，锌指结构域可经改造以与几乎任何所期望的序列结合。相应的，在识别目标区域含有下述序列，即在此期望基因调节，剪切或重组，则一个或更多锌指结合区域可经改造以与一个或更多在目标区域内的序列结合。

[0240] 在细胞染色质中选择 EPSPS 目标基因组区域以供锌指结构域（例如，靶位点）结合可由下述方法达成，例如，根据公开于共有的 US Patent No. 6, 453, 242 (Sept. 17, 2002) 中的方法，其亦公开了涉及 ZFP 以与选择序列结合的方法。本领域技术人员应清楚简地对核苷酸序列的肉眼检查亦可用于选择靶位点。相应的，在所要求的方法中可使用任何选择靶位点的方法。

[0241] 靶位点通常包括多个邻近的靶亚位点。靶亚位点指由单独锌指结合的序列（通常或者为核苷酸三联体，或可与邻近的核苷酸四联体有一个核苷酸重叠的核苷酸四联体）。参见，例如，WO 02/077227。若将锌指蛋白接触最多的链定义为靶链“主要 (primary) 识别链”或“主要接触链”，则一些锌指蛋白与靶链上的三碱基三联体以及非靶链的第四个碱基结合。靶位点通常至少长为 9 个核苷酸，且相应的，由含有至少三个锌指的锌指结合区域结合。然而，例如，亦有可能 4- 指结合区域与 12- 核苷酸靶位点结合，5- 指结合区域与 15- 核苷酸靶位点结合，或 6- 指结合区域与 18- 核苷酸靶位点结合。显而易见，更大的结合区域（例如，7-，8-，9- 指或更多）与更长靶位点的结合亦为可能的。



[0242] 靶位点并不需要为三的倍数个核苷酸。例如,当发生交叉链相互作用时(参见,例如,US Patent 6,453,242与WO 02/077227),多-指结合区域的一个或更多单独锌指可与重叠的四联体亚位点结合。其结果为,三-指蛋白质可与10-核苷酸序列结合,其中第十个核苷酸为由末端的指结合的四联体的一部分,四-指蛋白质可与13-核苷酸序列结合,其中第十三个核苷酸为由末端的指结合的四联体的一部分,依此类推。

[0243] 在单独锌指之间的氨基酸接头的长度和特性亦影响针对靶序列的结合。举例而言,在多-指结合区域邻近锌指间所谓“非规范接头”,“长接头”或“结构化接头”的存在可允许这些指与非直接毗连的亚位点结合。上述接头的非限定性示例描述于,例如,US Patent No. 6,479,626与WO 01/53480中。相应的,在锌指结合区域靶位点中的一个或更多亚位点,可彼此相隔1,2,3,4,5或更多核苷酸。仅举一例,4-指结合区域可与下述13-核苷酸靶位点结合,即所述靶位点顺次包含两个毗连的3-核苷酸亚位点,插入核苷酸,与两个毗连的三联体亚位点。关于连接人工核酸酶以结合于相隔不同数量的核苷酸的靶位点的组合物与方法亦参见,U.S. Application No. 61/130,099。序列(例如,靶位点)间的距离指两个序列间,从所述序列最靠近彼此的边缘测量的插入核苷酸或核苷酸对的数量。

[0244] 在有些实施方案中,设计了具有转录因子功能的ZFP。对转录因子功能,通常所有所需要的仅为单纯的与启动子足够近的结合。相对启动子的准确位置,取向,以及距离,只要其在限定范围内,均无足轻重。该特性允许在选择构建人工转录因子的靶位点时相当的灵活性。ZFP所识别的靶位点因此可为所述靶基因上任何合适的位点,只需其允许ZFP激活或抑制基因表达,所述ZFP可操作的连接于调节区。优选的靶位点包括转录开始位点邻近,下游或上游的区域。另外还包括,位于增强子区域的靶位点,阻抑物位点, RNA多聚酶暂停位点以及特异性调节位点(例如,SP-1位点,低氧(hypoxia)反应元件,核受体识别元件,p53结合位点),cDNA编码区域或已表达序列标志(EST)编码区域上的位点。

[0245] 在其它实施方案中,设计了具有核酸酶活性的ZFP。在细胞中,包含锌指结合区域与剪切域融合蛋白的ZFN(或两个融合蛋白,每个均包含锌指结合区域与剪切半域)的表达,导致在靶序列附近的剪切。在有些实施方案中,剪切依赖两个锌指结构域/剪切半域融合分子的结合以使结合位点分离。所述两个靶位点可位于不同的DNA链上,或者,两个靶位点可位于同一个DNA链上。

[0246] 锌指结合区域

[0247] 锌指结合区域包括一个或更多锌指。Miller et al. (1985)EMBO J. 4:1609-1614; Rhodes(1993)Scientific American Feb. :56-65;US Patent No. 6,453,242。通常,单独锌指结合区域长度为约30氨基酸。结构研究说明每个锌指结合区域(基序)含有两个beta折叠(sheet)(由包含两个不变半胱氨酸残基的beta转角支持),以及alpha螺旋(包含两个不变组氨酸残基),并通过所述两个半胱氨酸与两个组氨酸对锌原子的配位将其支撑为特定的构象。

[0248] 锌指既包括规范的(canonical) $C_2H_2$ 锌指(亦即,其中锌离子由两个半胱氨酸与两个组氨酸残基配位的那些),以及非规范的锌指,例如, $C_3H$ 锌指(其中锌离子由三个半胱氨酸残基与一个组氨酸残基配位的那些)以及 $C_4$ 锌指(其中锌离子由四个半胱氨酸残基配位的那些)。亦参见WO 02/057293。

[0249] 锌指结合区域可经改造以与所选的序列结合。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。改造的锌指结合区域可较之天然存在的锌指蛋白具有新型的结合特异性。改造方法包括但不限于,理性设计与多种类型的选择。理性设计包括,例如,使用包含三联体(或四联体)核苷酸序列与单独锌指氨基酸序列的数据库,其中每个三联体或四联体核苷酸序列与一个或更多锌指氨基酸序列相联系,所述锌指与该特定三联体或四联体序列结合。参见,例如,共有的 U. S. Patents 6, 453, 242 与 6, 534, 261。其它设计方法公开于,例如, U. S. Patents 6, 746, 838 ;6, 785, 613 ;6, 866, 997 ;以及 7, 030, 215 中。

[0250] 示例性的选择方法,包括噬菌体展示与二杂交系统,公开于 US Patents 5, 789, 538 ;5, 925, 523 ;6, 007, 988 ;6, 013, 453 ;6, 410, 248 ;6, 140, 466 ;6, 200, 759 ;与 6, 242, 568, 以及 WO 98/37186 ;WO 98/53057 ;WO 00/27878 ;WO 01/88197 与 GB 2, 338, 237 中。

[0251] 增强锌指结合区域的结合特异性描述于,例如,共有的 US Patent No. 6, 794, 136 中。

[0252] 因为单独锌指与三-核苷酸(亦即,三联体)序列(或可与邻近锌指的四-核苷酸结合位点重叠以一个核苷酸的四-核苷酸序列)结合,锌指结合区域经改造以结合的序列(例如,靶序列)长度将决定在改造的锌指结合区域中锌指的数量。举例而言,对于其中指基序不结合于重叠亚位点的 ZFP,六-核苷酸靶序列由二-指结合区域结合;九-核苷酸靶序列由三-指结合区域结合,依此类推。如本文所示,在靶位点上单独锌指的结合位点(亦即,亚位点)并不需要为毗连的,而是可以由一个或数个核苷酸所分隔,取决于在多-指结合区域中所述锌指间氨基酸序列(亦即,指间接头)的长度与特性。

[0253] 在多-指锌指结合区域中,邻近锌指可由 5 个氨基酸左右的氨基酸接头(所谓“规范的”指间接头(inter-finger linkers))序列,或者,一个或更多非规范的接头所分隔。参见,例如,共有的 US Patent Nos. 6, 453, 242 与 6, 534, 261。对包含多于三个指的改造的锌指结合区域,在有些锌指间插入较长的(“非规范的”)指间接头可为优选的,这是因为其可能增加结合区域的结合亲和力和/或特异性。参见,例如,U. S. Patent No. 6, 479, 626 与 WO 01/53480。相应的,多-指锌指结合区域亦可以非规范性指间接头的存在与位置的为特征。举例而言,包括三个指(由两个规范的指间接头联结),长接头与三个附加指(由两个规范的指间接头联结)的六-指锌指结合区域表示为 2x3 构型。类似的,包括两个指(与其间的规范的指间接头),长接头与两个附加指(由规范的指间接头联结)的结合区域表示为 2x2 蛋白质。包含三个二-指单元(在其中任一所述两个指由规范的接头联结),其中每个二-指单元由长接头联结于邻近的二指单元的蛋白质,称为 3x2 蛋白质。

[0254] 在多-指结合区域两个邻近锌指间长或非规范指间接头的存在常常允许所述两个指结合于在靶位点上并非直接紧接的亚位点。相应的,在靶位点上亚位点间可有一个或更多核苷酸的间隔,亦即,靶位点可包含一个或更多不由锌指接触的核苷酸。举例而言,2x2 锌指结合区域可与两个由一个核苷酸分隔的六-核苷酸序列结合,亦即,其结合于 13-核苷酸靶序列。亦参见 Moore et al. (2001a)Proc. Natl. Acad. Sci. USA 98 :1432-1436 ;Moore et al. (2001b)Proc. Natl. Acad. Sci. USA 98 :1437-1441 与 WO 01/53480。

[0255] 如之前所提及的,靶位点为由单独锌指结合的三-或四-核苷酸序列。出于某些考虑,二-指单元称为结合组件。结合组件可通过,例如,在多-指蛋白质(通常为三指)的环境下选择两个邻近指来获取,所述多-指蛋白质与特定六-核苷酸靶序列结合。或者,组件可通过组合单独的锌指来构建。亦参见 WO 98/53057 与 WO 01/53480。

[0256] 锌指结合区域可设计为与一个或更多同源的(例如,正向同源的或共生同源的)EPSPS 靶基因组序列结合。举例而言,锌指结合区域可设计为特异性的结合于一个独特的EPSPS 靶序列。或者,锌指结合区域可设计为与多个正向同源的或共生同源的EPSPS 基因组序列结合。

[0257] 在一个实施方案中,本文描述了包括示于表 A 氨基酸序列的锌指结合区域。在另一个实施方案中,本公开提供了编码锌指结合区域的多核苷酸,其中所述锌指结合区域包括示于表 A 的氨基酸序列。

[0258] 调节区

[0259] 本文所述的 ZFP 可任选的与调节区相联系以调整基因表达。ZFP 可共价的或非共价的与一个或更多调节区,或者两个或更多调节区相联系,其中所述两个或更多域为同一域的两个拷贝,或两个不同域。所述调节区可作为融合蛋白的一部分,共价的与所述 ZFP 相联,例如,通过氨基酸接头。所述 ZFP 亦可通过非共价的二聚体化功能域,例如,亮氨酸拉链, STAT 蛋白质 N 末端域或 FK506 结合蛋白(参见,例如, O' Shea, Science 254 : 539(1991), Barahmand-Pour et al, Curr. Top. Microbiol. Immunol. 211 :121-128(1996); Klemm et al., Annu. Rev. Immunol. 16 :569-592(1998); Klemm et al., Annu. Rev. Immunol. 16 :569-592(1998); Ho et al., Nature 382 :822-826(1996); 以及 Pomeranz et al., Biochem. 37 :965(1998)) 与调节区相联系。所述调节区可与 ZFP 在任何合适的位置相联系,包括所述 ZFP 的 C 或 N 末端。

[0260] 除 ZFP 外通常的调节区包括,例如,转录因子的效应者(effector)结构域(激活物,阻抑物,共激活物,共阻抑物),沉默基因,核激素受体,癌基因转录因子(例如, myc, jun, fos, myb, max, mad, rel, ets, bcl, myb, mos 家族成员等等);DNA 修复酶及其相关因子与修饰因子;DNA 重排酶及其相关因子与修饰因子;染色质相关蛋白及其修饰因子(例如, 激酶,乙酰酶与脱乙酰酶);以及 DNA 修饰酶(例如, 甲基转移酶,拓扑异构酶,解螺旋酶,连接酶,激酶,磷酸酶,多聚酶,核酸内切酶)及其相关因子与修饰因子。

[0261] 可从其获取调节区的转录因子多肽包括涉及调节与基础转录中的那些。上述多肽包括转录因子,其效应者结构域,共激活物,沉默基因,核激素受体(对关于转录中涉及的蛋白质与核酸元件的综述参见,例如, Goodrich et al., Cell 84 :825-30(1996); 转录因子的总体综述参见 Barnes & Adcock, Clin. Exp. Allergy 25 Suppl. 2 :46-9(1995) and Roeder, Methods Enzymol. 273 :165-71(1996))。已知专用于转录因子的数据库(参见,例如, Science 269 :630(1995))。核激素受体转录因子描述于,例如, Rosen et al., J Med. Chem. 38 :4855-74(1995)。C/EBP 家族转录因子的综述见于 Wedel et al., Immunobiology 193 :171-85(1995)。介导通过核激素受体的转录调节的共激活物与共阻抑物的综述见于,例如, Meier, Eur. J Endocrinol. 134(2) :158-9(1996); Kaiser et al., Trends Biochem. Sci. 21 :342-5(1996); 以及 Utley et al., Nature 394 :498-502(1998))。GATA 转录因子,其涉及血细胞生成的调节,描述于,例如, Simon, Nat. Genet. 11 :9-11(1995); Weiss et al.,

Exp. Hematol. 23 :99-107。TATA 盒结合蛋白 (TBP) 与其相关的 TAP 多肽 (其包括 TAF30, TAF55, TAF80, TAF 10, TAFI 50 与 TAF250) 描述于 Goodrich & Tjian, Curr. Opin. Cell Biol. 6 :403-9(1994) and Hurley, Curr. Opin. Struct. Biol. 6 :69-75(1996)。STAT 家族转录因子的综述见于, 例如, Barahmand-Pour et al., Curr. Top. Microbiol. Immunol. 211 :121-8(1996)。涉及疾病的转录因子的综述见于 Aso et al., J Clin. Invest. 97 :1561-9(1996)。

[0262] 在一个实施方案中, 来自人类 KOX-1 蛋白质的 KRAB 阻抑域用作转录阻抑物 (Thiesen et al., New Biologist 2 :363-374(1990); Margolin et al., PNAS 91 :4509-4513(1994); Pengue et al., Nucl. Acids Res. 22 :2908-2914(1994); Witzgall et al., PNAS 91 :4514-4518(1994))。在另一个实施方案中, KAP-1, KRAB 共阻抑物, 与 KRAB 一同使用 (Friedman et al., Genes Dev. 10 :2067-2078(1996))。或者, KAP-1 可单独与 ZFP 使用。其它作为转录阻抑物作用的优选转录因子和转录因子域包括 MAD (see, e. g., Sommer et al., J. Biol. Chem. 273 :6632-6642(1998); Gupta et al., Oncogene 16 :1149-1159(1998); Queva et al., Oncogene 16 :967-977(1998); Larsson et al., Oncogene 15 :737-748(1997); Laherty et al., Cell 89 :349-356(1997); 以及 Cultraro et al., Mol. Cell. Biol. 17 :2353-2359(1997)); FKHR (横纹肌肉瘤叉头基因 (forkhead in rhabdomyosarcoma gene); Ginsberg et al., Cancer Res. 15 :3542-3546(1998); Epstein et al., Mol. Cell. Biol. 18 :4118-4130(1998)); EGR-1 (早期生长反应基因产物 -1 (early growth response gene product -1); Yan et al., PNAS 95 :8298-8303(1998); and Liu et al., Cancer Gene Ther. 5 :3-28(1998)); ets2 阻抑物因子阻抑物域 (the ets2 repressor factor repressor domain) (ERD; Sgouras et al., EMBO J 14 :4781-4793(1995)); 以及 MAD smSIN3 相互作用域 (SID; Ayer et al., Mol. Cell. Biol. 16 :5772-5781(1996))。

[0263] 在一个实施方案中, HSV VP16 激活域用作转录激活物 (参见, 例如, Hagmann et al., J. Virol. 71 :5952-5962(1997))。其它可提供激活域的优选转录因子包括 VP64 激活域 (Seipel et al., EMBO J. 11 :4961-4968(1996)); 核激素受体 (参见, 例如, Torchia et al., Curr. Opin. Cell. Biol. 10 :373-383(1998)); 核因子 kappa B 的 p65 亚基 (Bitko & Barik, J. Virol. 72 :5610-5618(1998) and Doyle & Hunt, Neuroreport 8 :2937-2942(1997)); 以及 EGR-1 (早期生长反应基因产物 -1; Yan et al., PNAS 95 :8298-8303(1998); 以及 Liu et al., Cancer Gene Ther. 5 :3-28(1998))。

[0264] 激酶, 磷酸酶与其它修饰涉及基因调节的多肽亦作为 ZFP 的调节区为有用的。上述修饰因子常常涉及由, 例如, 激素介导的打开或关闭转录。涉及转录调节的激素的综述见于 Davis, Mol. Reprod. Dev. 42 :459-67(1995), Jackson et al., Adv. Second Messenger Phosphoprotein Res. 28 :279-86(1993), 以及 Boulikas, Crit. Rev. Eukaryot. Gene Expr. 5 :1-77(1995), 而磷酸酶的综述见于, 例如, Schonthal & Semin, Cancer Biol. 6 :239-48(1995)。核酪氨酸激酶描述于 Wang, Trends Biochem. Sci. 19 :373-6(1994)。

[0265] 如上所述, 有用的域亦可得自癌基因 (例如, myc, jun, fos, myb, max, mad, rel, ets, bcl, myb, mos 家族成员) 的基因产物与其相关因子与修饰因子。癌基因描述于, 例如, Cooper. Oncogenes, 2nd ed., The Jones and Bartlett Series in Biology, Boston, Mass., Jones and Bartlett Publishers, 1995。ets 转录因子的综述见于 Waslylk et al.,

Eur. J. Biochem. 211 :7-18(1993) 与 Crepieux et al., Crit. Rev. Oncog. 5 :615-38(1994)。Myc 癌基因的综述见于, 例如, Ryan et al., Biochem. J. 314 :713-21(1996)。jun 与 fos 转录因子描述于, 例如, The Fos and Jun Families of Transcription Factors, Angel & Herrlich, eds. (1994)。max 癌基因的综述见于 Hurlin et al., Cold Spring Harb. Symp. Quant. Biol. 59 :109-16。myb 基因家族的综述见于 Kanei-Ishii et al., Curr. Top. Microbiol. Immunol. 211 :89-98(1996)。mos 家族的综述见于 Yew et al., Curr. Opin. Genet. Dev. 3 :19-25(1993)。

[0266] ZFP 可包括得自 DNA 修复酶与其相关因子与修饰因子的调节区。DNA 修复系统的综述见于, 例如, Vos, Curr. Opin. Cell Biol. 4 :385-95(1992) ; Sancar, Ann. Rev. Genet. 29 :69-105(1995) ; Lehmann, Genet. Eng. 17 :1-19(1995) ; 以及 Wood, Ann. Rev. Biochem. 65 :135-67(1996)。DNA 重排酶与其相关因子与修饰因子亦可用作调节区 (参见, 例如, Gangloff et al., Experientia 50 :261-9(1994) ; Sadowski, FASEB J. 7 :760-7(1993))。

[0267] 类似的, 调节区亦可源自 DNA 修饰酶 (例如, DNA 甲基转移酶, 拓扑异构酶, 解螺旋酶, 连接酶, 激酶, 磷酸酶, 多聚酶) 与其相关因子与修饰因子。解螺旋酶的综述见于 Matson et al., Bioessays, 16 :13-22(1994), 而甲基转移酶描述于 Cheng, Curr. Opin. Struct. Biol. 5 :4-10(1995)。染色质相关蛋白与其修饰因子 (例如, 激酶, 乙酰酶与脱乙酰酶), 例如组蛋白脱乙酰酶 (Wolffe, Science 272 :371-2(1996)) 亦有用于作为域以添加到所选的 ZFP。在一个优选实施方案中, 所述调节区为作为转录阻抑物作用的 DNA 甲基转移酶, (参见, 例如, Van den Wyngaert et al., FEBS Lett. 426 :283-289(1998) ; Flynn et al., J. Mol. Biol. 279 :101-116(1998) ; Okano et al., Nucleic Acids Res. 26 :2536-2540(1998) ; 以及 Zardo & Caiafa, J. Biol. Chem. 273 :16517-16520(1998))。在另一个优选实施方案中, 核酸内切酶, 如 FokI 用作转录阻抑物, 其通过基因剪切起作用 (参见, 例如, W095/09233 ; 以及 PCT/US94/01201)。

[0268] 控制染色质与 DNA 结构, 活动与定位的因子及其相关因子与修饰因子 ; 源自微生物 (例如, 原核生物, 真核生物与病毒) 的因子, 以及与其相联系或修饰其的因子亦可用于获取嵌合蛋白质。在一个实施方案中, 重组酶与整合酶用作调节区。在一个实施方案中, 组蛋白乙酰转移酶用作转录激活物 (参见, 例如, Jin & Scotto, Mol. Cell. Biol. 18 :4377-4384(1998) ; Wolffe, Science 272 :371-372(1996) ; Taunton et al., Science 272 :408-411(1996) ; 以及 Hassig et al., PNAS 95 :3519-3524(1998))。在另一个实施方案中, 组蛋白脱乙酰酶用作转录阻抑物 (参见, 例如, Jin & Scotto, Mol. Cell. Biol. 18 :4377-4384(1998) ; Syntichaki & Thireos, J. Biol. Chem. 273 :24414-24419(1998) ; Sakaguchi et al., Genes Dev. 12 :2831-2841(1998) ; 以及 Martinez et al., J. Biol. Chem. 273 :23781-23785(1998))。

[0269] 可包含多肽域间的接头域, 例如, 两个 ZFP 间或 ZFP 与调节区间的。上述接头通常为多肽序列, 例如大约 5 到 200 个氨基酸间的多聚甘氨酸序列。优选的接头通常为弹性的 (flexible) 氨基酸序列, 其作为重组融合蛋白的一部分合成。参见, 例如, U. S. Patent No. 6, 534, 261 ; Liu et al., Proc. Nat. Acad. Sci. USA, 95 :5525-5530(1997) ; Pomerantz et al., Proc. Nat. Acad. Sci. USA 92 :9752-9756(1995) ; Kim et al., Proc. Nat. Acad. Sci. USA 93 :1156-1160(1996) ; 以其全文引用的形式包含于本文中。或者, 弹性的接头可

使用下述的计算机程序,即其能够对 DNA 结合位点与肽自身两者进行建模 (Desjarlais & Berg, Proc. Nat. Acad. Sci. USA 90 :2256-2260 (1993), Proc. Nat. Acad. Sci. USA 91 :11099-11103 (1994), 或通过噬菌体展示方法理性的进行设计。

[0270] 在其他实施方案中,使用化学接头以连接合成产生或重组产生的域序列。上述弹性的接头对本领域技术人员为已知的。举例而言,聚乙二醇接头可得自 Shearwater Polymers, Inc. Huntsville, Ala. 这些接头可任选的具有酰胺键,氢硫键或异功能键 (heterofunctional linkage)。除将 ZFP 共价的连接于调节区外,可使用非共价方法以产生 ZFP 与调节区相联系的分子。

[0271] 剪切域

[0272] 本文所公开的融合蛋白的剪切域部分可得自任何核酸内切酶或核酸外切酶。剪切域可源自的核酸内切酶示例包括但不限于,限制性核酸内切酶与靶向核酸内切酶。参见,例如,2002-2003 Catalogue, New England Biolabs, Beverly, MA ;and Belfort et al. (1997) Nucleic Acids Res. 25 :3379-3388。附加的剪切 DNA 的酶为已知的 (例如, S1 核酸酶 ;绿豆核酸酶 ;胰 DNase I ;微球菌核酸酶 ;酵母 H0 核酸内切酶 ;亦参见 Linn et al. (eds.) Nucleases, Cold Spring Harbor Laboratory Press, 1993)。这些酶 (或其功能性片段) 中的一种或更多可用作剪切域与剪切半域的来源。

[0273] 类似的,剪切半域 (例如,包含锌指结合区域与剪切半域的融合蛋白) 可源自下述任何核酸酶或其部分,如上面所提出的,即其剪切活性需要二聚作用。通常,若融合蛋白包括剪切半域,则需要两个融合蛋白进行剪切。或者,可使用包括两个剪切半域的单蛋白。所述两个剪切半域可源自同一个核酸内切酶 (或其功能性的片段),或每个剪切半域可源自不同的核酸内切酶 (或其功能性的片段)。此外,所述两个融合蛋白的靶位点优选以下述的方式相互排列,即使得两个融合蛋白分别与其靶位点的结合将剪切半域置为下述两者间的空间取向,即其允许所述剪切半域形成功能性的剪切域,例如,通过二聚化。从而,在有些实施方案中,所述靶位点的近边间隔 5-8 个核苷酸或 15-18 个核苷酸。然而任何整数数量的核苷酸或核苷酸对可插入到两个靶位点之间 (例如,从 2 到 50 个核苷酸或更多)。通常,剪切点处于靶位点之间。

[0274] 限制性核酸内切酶 (限制性酶) 存在于许多物种中,且能够序列特异性的结合于 DNA (在识别位点),并在或接近结合位点处剪切 DNA。有些限制性酶 (例如, IIS 型) 在远离识别位点的位点剪切 DNA,且具有可分离的结合与剪切域。举例而言, IIS 型酶 FokI 在距离其一个链上的识别位点 9 个核苷酸,另一个链上识别位点 13 个核苷酸处催化 DNA 的双链剪切。参见,例如, US Patents 5,356,802 ;5,436,150 与 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。从而,在一个实施方案中,融合蛋白包括来自至少一个 IIS 型限制性酶的剪切域 (或剪切半域) 以及一个或更多锌指结合区域,其可能为经或未经改造。

[0275] 其剪切域可与结合区域分离的 IIS 型限制性酶示例为 FokI。该特定酶作为二聚体有活性 (Bitinaite et al. (1998) Proc. Natl. Acad. Sci. USA 95 :10,570-10,575)。相应的,就现行公开考虑,用于公开的融合蛋白中的 FokI 酶部分被视为剪切半域。从而,针对使

用锌指-FokI 融合蛋白的靶向双链剪切和 / 或靶向细胞序列取代, 两个融合蛋白, 每个均包含 FokI 剪切半域, 可用于重建具有催化活性的剪切域。或者, 亦可使用含有锌指结合区域与两个 FokI 剪切半域的单多肽。针对使用锌指-FokI 融合蛋白的靶向剪切与靶向序列变更在本公开其它地方提供。

[0276] 剪切域或剪切半域可为蛋白质的任何部分, 只要其保留剪切活性的, 或保留多聚化 (例如, 二聚化) 以形成功能性剪切域的能力。

[0277] IIS 型限制性酶的示例列于表 1。附加的限制性酶亦包含可分离的结合与剪切域, 且其为现行公开所考虑。参见, 例如 Roberts et al. (2003) *Nucleic Acids Res.* 31: 418-420。

[0278] 表 1: 一些 IIS 型限制性酶

[0279]	Aar I	BsrB I	SspD5 I
[0280]	Ace III	BsrD I	Sth132 I
[0281]	Aci I	BstF5I	Sts I
[0282]	Alo I	Btr I	TspDT I
[0283]	Bae I	Bts I	TspGW I
[0284]	Bbr7 I	Cdi I	Tth111 II
[0285]	Bbv I	CjeP I	UbaP I
[0286]	Bbv II	Drd II	Bsa I
[0287]	BbvC I	Eci I	BsmB I
[0288]	Bcc I	Eco31 I	
[0289]	Bce83 I	Eco57 I	
[0290]	BceA I	Eco57M I	
[0291]	Bcef I	Esp3 I	
[0292]	Bcg I	Fau I	
[0293]	BciV I	Fin I	
[0294]	Bfi I	Fok I	
[0295]	Bin I	Gdi II	
[0296]	Bmg I	Gsu I	
[0297]	Bpu10 I	Hga I	
[0298]	BsaX I	Hin4 II	
[0299]	Bsb I	Hph I	
[0300]	BscA I	Ksp632 I	
[0301]	BscG I	Mbo II	
[0302]	BseR I	Mly I	
[0303]	BseY I	Mme I	
[0304]	Bsi I	Mnl I	
[0305]	Bsm I	Pf11108 I	
[0306]	BsmA I	Ple I	
[0307]	BsmF I	Ppi I	

[0308] Bsp24 I Psr I

[0309] BspG I RleA I

[0310] BspM I Sap I

[0311] BspNC I SfaN I

[0312] Bsr I Sim I

[0313] 锌指融合蛋白

[0314] 设计与构建融合蛋白（与编码其的多核苷酸）的方法对本领域技术人员为已知的。举例而言，设计与构建包含锌指结构域与调节或剪切域（或剪切半域）的融合蛋白，以及编码上述融合蛋白的多核苷酸的方法，描述于共有的 Patents 6, 453, 242 与 6, 534, 261 以及 U.S. Patent Application Publications 2007/0134796 与 2005/0064474；其以全文引用的方式包含于本文中。在有些实施方案中，构建了编码所述融合蛋白的多核苷酸。该多核苷酸可插入载体中，且该载体可导入细胞中（更多关于载体与将多核苷酸导入细胞的方法的公开见下）。

[0315] 在本文所述方法的有些实施方案中，锌指核酸酶包括含有锌指结合区域与来自 FokI 限制性酶的剪切半域的融合蛋白，且在细胞中表达两个上述的融合蛋白。在细胞中两个融合蛋白的表达可来自下述方法：将两个蛋白质递送到所述细胞中；将一个蛋白质与编码一个蛋白质的核酸递送到所述细胞中；将两个核酸，每个均编码一个蛋白质递送到所述细胞中；或通过将编码两个蛋白质单一核酸递送到所述细胞中。在其它实施方案中，融合蛋白包括含有两个剪切半域与锌指结合区域的单一多肽链。在此情况下，单一融合蛋白表达于细胞中，且虽不愿拘于理论，但相信作为所述剪切半域分子内二聚体形成的结果，该融合蛋白剪切 DNA。

[0316] 在有些实施方案中，所述融合蛋白（例如，ZFP-FokI 融合蛋白）的组分经排列，使得所述锌指功能域最接近所述融合蛋白的氨基末端，而所述剪切半域最接近其羧基末端。这与在天然存在的二聚体剪切域，例如那些源自 FokI 酶的中，剪切域的相对取向一致，即其中 DNA 结合区域最接近氨基末端而剪切半域最接近羧基末端。在这些实施方案中，所述剪切半域的二聚化形成功能性的核酸酶通过将所述融合蛋白结合至相反 DNA 链上的位点来达成，其中所述结合位点的 5' 端彼此接近。

[0317] 在其它实施方案中，所述融合蛋白（例如，ZFP-FokI 融合蛋白）的组分经排列，使得所述剪切半域最接近所述融合蛋白的氨基末端，而所述锌指结构域最接近其羧基末端。在这些实施方案中，所述剪切半域的二聚化形成功能性的核酸酶通过将所述融合蛋白结合至相反 DNA 链上的位点来达成，其中所述结合位点的 3' 端彼此接近。

[0318] 在仍旧其它的实施方案中，第一个融合蛋白在其氨基末端最接近处含有所述剪切半域，而在其羧基末端最接近处含有所述锌指结构域，而第二个融合蛋白如下述排列，使得所述锌指结构域在其氨基末端最接近处，而所述剪切半域在其羧基末端最接近处。在这些实施方案中，两个融合蛋白均结合于同一个 DNA 链，其中第一个其锌指结构域最靠近其羧基末端的融合蛋白的结合位点位于第二个其锌指结构域最接近其氨基末端的融合蛋白的结合位点的 5' 侧。

[0319] 在有些实施方案中，所公开的融合蛋白其在所述锌指结构域与所述剪切域（或剪切半域）间的氨基酸序列称为“ZC 接头”。所述 ZC 接头须与上面讨论的指间接头区



别开来。关于获取优化剪切的 ZC 接头的细节,参见,例如,U.S.Patent Publications 20050064474A1 与 20030232410 以及 International Patent Publication W005/084190。

[0320] 在一个实施方案中,本公开提供了包含具有示于表 A 的识别螺旋氨基酸序列的锌指蛋白的 ZFN。在另一个实施方案中,本文提供了包含具有示于表 A 识别螺旋的 ZFP 的编码核苷酸序列的 ZFP 表达载体。

[0321] 基因表达的调节

[0322] 多种分析可用于确定 ZFP 是否调整基因表达。特定 ZFP 的活性可使用多种体外与体内分析通过测定,例如,蛋白质或 mRNA 水平,产物水平,酶活性,报道基因的转录激活或阻抑,使用,例如,免疫分析(例如,使用抗体的 ELISA 与免疫组织化学分析),杂交分析(例如,Rnase 保护,Northern,原位杂交,寡核苷酸点阵研究),色度法分析,扩增分析,酶活性分析,表型分析以及类似的来进行评价。

[0323] ZFP 通常首先针对其活性在体外使用 ELISA 分析,随即使用肾脏细胞来测试。所述 ZFP 通常首先使用具有报道基因的暂时表达系统测试,并随即所述靶内源基因的调节在细胞中与在全植物中在体内与体外均进行测试。所述 ZFP 可在细胞中重组表达,可在移植到植物的细胞中重组表达,或在转基因植物中重组表达,以及作为蛋白质使用下述的给药载体施用于植物或细胞。所述细胞可为固定化的,可在溶液中,可注射于植物,或可天然存在于转基因或非转基因植物。

[0324] 基因表达的调节使用本文所述的一种体外或体内分析进行测试。将试样或测试经 ZFP 处理,并与无测试化合物的对照试样比较以检查调整的程度。对内源基因表达的调节,所述 ZFP 通常具有 200nM 或更少的  $K_d$ ,较优选 100nM 或更少,更优选 50nM,最优选 25nM 或更少。

[0325] 所述 ZFP 的作用可通过检查任何上述参数来进行测定。任何合适的基因表达,表型或生理学变化均可用于评价 ZFP 的影响。当使用完整细胞或植物确定功能性后果时,亦可测定多种作用,例如植物生长,已知与未特征化的遗传标记的转录变化(例如,Northern blot 或寡核苷酸点阵研究),细胞代谢的变化,例如细胞生长或 pH 的变化,以及胞内第二信使,例如 cGMP 的变化。

[0326] ZFP 调节内源基因表达的优选分析可在体外实施。在一个优选的体外分析形式中,在培养细胞中 ZFP 对内源基因表达的调节通过使用 ELISA 分析检查蛋白质生成来测定。将所述测试试样与经空载体或以其它基因为目标,不相关的 ZFP 处理的对照细胞进行比较。

[0327] 在另一个实施方案中,ZFP 对内源基因表达的调节在体外通过测定靶基因 mRNA 表达的水平来确定。基因表达水平使用扩增,例如,使用 PCR,LCR 或杂交分析,例如 Northern 杂交,Rnase 保护,点印迹法,来进行测定。Rnase 保护用于一个实施方案中。蛋白质与 mRNA 的水平使用直接或间接标记的检测剂,例如,荧光或放射性标记的核酸,放射性或酶标记的抗体,以及类似物进行检测,如本文所述。

[0328] 或者,可使用可操作性的连接于报道基因,例如荧光素酶,绿色荧光蛋白,CAT 或  $\beta$ -gal 的靶基因启动子设计报道基因系统。所述报道构建体通常共转染至培养细胞中。当使用所选的 ZFP 处理后,报道基因转录,翻译或活性的量根据本领域技术人员所知的标准技术进行测定。

[0329] 转基因与非转基因植物亦用作检查内源基因体内表达调节的优选实施方案。转基

因植物可稳定的表达所选的 ZFP。或者,可使用暂时表达所选 ZFP 的植物,或对其 ZFP 通过递送载体施与的植物。内源基因表达的调节使用任何一个本文描述的分析进行测试。

#### [0330] 靶向剪切方法

[0331] 公开的方法与组合物可用于在细胞染色质上的目标区域(例如,在基因组中所期望的或预决的位点上,例如,在 EPSPS 基因内或邻近该基因,无论该基因是突变体或野生型)剪切 DNA。为进行上述靶向 DNA 剪切,锌指结合区域经改造与在预决的剪切位点上或邻近其的靶位点结合,且在细胞内表达包含所述改造的锌指结合区域与剪切域的融合蛋白。在所述融合蛋白的锌指部分与所述靶位点结合后,剪切域在所述靶位点附近剪切 DNA。剪切的准确位点可取决于 ZC 接头的长度。

[0332] 或者,两个融合蛋白,每个包含锌指结合区域与剪切半域,在细胞中表达,且与以下述方式并置的靶位点结合,即重建了功能性的剪切域,且 DNA 在所述靶位点附近被剪切。在一个实施方案中,剪切发生在两个锌指结合区域的靶位点之间。所述锌指结合区域中的一个或两者可经改造。

[0333] 为使用锌指结合区域-剪切域融合多肽进行靶向剪切,结合位点可涵盖剪切位点,或所述结合位点的近边可为距离所述剪切位点 1,2,3,4,5,6,10,25,50 或更多个核苷酸(或任何在 1 到 50 个核苷酸之间的整数值)。所述结合位点相对于所述剪切位点的准确位置将取决于特定的剪切域,以及 ZC 接头的长度。对于使用两个融合多肽,其中每个均包含锌指结合区域与剪切半域的方法的,所述结合位点通常位于所述剪切位点两侧。从而第一个结合位点的近边可为距所述剪切位点一侧 1,2,3,4,5,6,25 或更多个核苷酸(或任何在 1 到 50 个核苷酸之间的整数值),而第二个结合位点的近边可为距所述剪切位点另一侧 1,2,3,4,5,6,25 或更多个核苷酸(或任何在 1 到 50 个核苷酸之间的整数值)。在体外与体内对剪切位点进行定位的方法对本领域技术人员为已知的。

[0334] 从而,本文所述方法可使用融合于剪切域的改造的锌指结合区域。在此情况下,所述结合区域经改造,在期望进行剪切的位置或其附近与靶序列结合。将所述融合蛋白,或编码其的多核苷酸,导入植物细胞。一旦导入所述细胞或在其中表达,所述融合蛋白结合于靶序列,并在靶序列上或其附近进行剪切。剪切的准确位点取决于所述剪切域的特性和/或所述结合与剪切域之间接头序列的存在和/或特性。在使用两个融合蛋白,每个均包含剪切半域的情况下,所述结合位点近边之间的距离可为 1,2,3,4,5,6,7,8,9,10,25 或更多个核苷酸(或任何 1 到 50 个核苷酸之间的整数值)。剪切的水平亦可既取决于在所述两个融合蛋白上结合位点之间的距离(参见,例如,Smith et al. (2000) *Nucleic Acids Res.* 28 :3361-3369; Bibikova et al. (2001) *Mol. Cell. Biol.* 21 :289-297),又取决于每个融合蛋白上 ZC 接头的长度。亦参见 U.S. Patent Publication 20050064474A1 以及 International Patent Publications W005/084190, W005/014791 与 W003/080809。

[0335] 在有些实施方案中,所述剪切域包含两个剪切半域,即第一剪切半域与第二剪切半域,两者均为包含结合区域的单一多肽的一部分。所述剪切半域可具有相同或不同的氨基酸序列,只要它们具有剪切 DNA 的功能。

[0336] 剪切半域亦可能由单独分子分别提供。举例而言,两个融合多肽可能导入细胞,其中每个多肽均包含结合区域与剪切半域。所述剪切半域可具有相同或不同的氨基酸序列,只要它们具有剪切 DNA 的功能。进一步,所述结合区域与靶位点结合,所述靶位点通常以下

述方式配置,即当所述融合多肽结合后,所述两个剪切半域表现出下述的相对彼此的空间取向,即其允许剪切域的重建(例如,通过所述半域的二聚化),从而将所述半域置于相对彼此的位置以形成功能性的剪切域,导致在目标区域剪切细胞染色质。一般而言,所述重建剪切域的剪切发生在位于所述两个靶序列之间的位点。所述蛋白质中的一个或两者可经改造以与其靶位点结合。

[0337] 所述两个融合蛋白可以相同或者相反极性结合于所述目标区域,且其结合位点(亦即,靶位点)可间隔任何数量的核苷酸,例如,从0到200个核苷酸或任何其间的整数值。在有些实施方案中,两个融合蛋白,其中每个均包含锌指结合区域与剪切半域,的结合位点,自每个结合位点最接近另一个结合位点的边缘(edge)起计算可彼此间距5到18之间个核苷酸,例如,间距5-8个核苷酸,或间距15-18个核苷酸,或间距6个核苷酸,或间距16个核苷酸,且剪切发生于所述结合位点之间。

[0338] DNA剪切的位点通常位于所述两个融合蛋白的结合位点之间。DNA的双链断裂通常为两个单链断裂,亦称“切口”,彼此偏移1,2,3,4,5,6或更多个核苷酸的结果(举例而言,天然FokI剪切双链DNA为彼此偏移4个核苷酸的单链断裂的结果)。从而,剪切未必需要发生于每个DNA链上准确相对的位点。另外,所述融合蛋白的结构与所述靶位点间的距离可影响剪切是否在邻近一个单独核苷酸对处发生,或剪切是否发生于多个位点。然而,对于许多应用,包括靶向重组与靶向突变(见下),在一定核苷酸范围内的剪切通常已经足够,而不需要剪切发生于特定碱基对间。

[0339] 如上所指明,所述融合蛋白可作为多肽和/或多核苷酸导入。举例而言,两个多核苷酸,每个包含编码一个前述多肽的序列,可导入细胞,且当所述多肽表达,并每个均与其靶序列结合时,剪切发生在靶序列处或其附近。另外,亦可将包含编码两个融合多肽的单一多核苷酸导入细胞。多核苷酸可为DNA, RNA或其任何修饰形式或类似物,或DNA和/或RNA。

[0340] 为增强剪切特异性,亦可在本文所述方法中使用附加的组合物。举例而言,单一剪切半域可展示受限的双链剪切活性。在下述方法中,即其中两个融合蛋白,每个均含有三指锌指结构域与剪切半域,导入细胞,则任一蛋白质特定于约9-核苷酸的靶位点。尽管该总计(aggregate)18个核苷酸的靶序列似在哺乳类基因组中为独特的,任何给定9-核苷酸靶位点平均在人类基因组中出现约23000次。从而,由于单一半域的位点特异性结合,可能会出现非特异性剪切。相应的,本文所述方法考虑使用剪切半域,例如FokI(或编码其的核酸)的显性失活突变体,其与两个融合蛋白一起表达于细胞中。所述显性失活突变体能够二聚化,但是无法剪切,并亦阻遏与其二聚化的半域的剪切活性。通过提供摩尔数上针对所述融合蛋白为过量的显性失活突变体,仅仅下述区域,即其中两个融合蛋白均结合的,才将具有足够高局部浓度的功能性剪切半域以供二聚作用以及剪切发生。在下述位点,即仅仅两个融合蛋白之一结合的,其剪切半域与所述显性失活突变体半域形成二聚体,而不发生非期望的,非特异性的剪切。

[0341] 在所述FokI剪切半域中已经识别出了三个催化性的氨基酸残基:Asp450, Asp467与Lys469。Bitinaite et al. (1998)Proc. Natl. Acad. Sci. USA 95:10,570-10,575。从而,在这些残基中的一个中一个或更多突变可用于产生显性失活突变。进一步,其它IIS型核酸内切酶的许多催化性氨基酸残基为已知和/或可以通过,例如,与FokI序列进行序列比

对和 / 或通过产生并测试突变体的催化活性来确定。

[0342] 在剪切半域中二聚体化功能域突变

[0343] 下述方法,即涉及使用 ZFP 与剪切半域间的融合物(例如,ZFP/FokI 融合物)以进行靶向剪切的,需要使用两个上述融合分子,每个均通常导向独特的靶序列。所述两个融合蛋白的靶序列可如此选取,从而将靶向剪切导向基因组中的独特位点,如上所讨论。使得剪切特异性减少的潜在来源为所述两个 ZFP/ 剪切半域融合物之一的同源二聚化的结果。该现象可能由于下列原因发生,例如,在基因组中存在所述两个 ZFP/ 剪切半域融合物之一的靶序列的反向重复序列,且所述靶序列的位置允许所述相同融合蛋白的两个拷贝互相以下述取向与间距结合,即其允许功能性二聚体的形成。

[0344] 一种减少在除预期靶位点以外的序列发生这种类型的异常剪切的概率的方法涉及产生所述剪切半域的变体,其最小化或阻止同源二聚化。优选的,改变所述半域中涉及其二聚化的区域中的一个或更多氨基酸。在所述 FokI 蛋白质二聚体的晶体结构中,有报道称所述剪切半域的结构类似于在 FokI 剪切 DNA 时所述剪切半域的排列。Wah et al. (1998) Proc. Natl. Acad. Sci. USA95 :10564-10569。该结构指明在位置 483 与 487 处的氨基酸残基在所述 FokI 剪切半域的二聚化中起关键作用。该结构亦指明在位置 446,447,479,483,484,486,487,490,491,496,498,499,500,531,534,537 与 538 处的氨基酸残基均足够靠近二聚作用的界面,从而影响到二聚作用。相应的,在前述位置中的一个或更多处发生的氨基酸序列改变似将改变所述剪切半域的二聚作用特性。上述变化可通过,例如,构建含有(或编码)在这些位置不同氨基酸残基的文库并选取具有所期望特性的变体,或通过理性设计个体的突变体来导入。除阻止同源二聚作用外,亦可能这些突变中一些可能增加上述由两个野生型剪切半域所得的剪切效力。

[0345] 相应的,在 FokI 剪切半域中任何氨基酸残基处的改变,若其影响二聚化,则其可用于阻止一对 ZFP/FokI 融合物间发生下述二聚作用,即其可导致在非期望序列处的剪切。从而,为使用一对 ZFP/FokI 融合物进行靶向剪切,所述融合蛋白中的一个或两者可包括一个或更多的改变,其抑制自身二聚作用,但允许所述两个融合蛋白间的异源二聚作用发生,从而使得剪切发生在期望的靶位点。在有些实施方案中,两个融合蛋白均存在改变,且所述改变具有附加作用;亦即,最小化或消除了任一融合物的同源二聚作用,其导致异常剪切,同时较之得自使用野生型剪切半域的,促进了所述两个融合蛋白间的异源二聚作用。

[0346] 对共生同源基因组序列的靶向改变方法与靶向重组方法

[0347] 本文亦描述了取代基因组序列的方法,例如,将一个或更多共生同源基因(例如,细胞染色质中作为 EPSPS 靶的目标基因组区域)取代为同源非等同序列(亦即,靶向重组)。之前取代特定序列的尝试涉及将细胞与包含具有与染色体区域有同源性的序列(亦即,供体 DNA)接触,继以选择其中所述供体 DNA 分子同源重组入基因组的细胞。这些方法的成功率较低,这是由于同源重组的不良效率,以及所述供体 DNA 高频度的非特异性插入基因组上除目标位点外的区域。

[0348] 现行公开提供了靶向序列改变的方法,其特征为更高效率的靶向重组与更低频率的非特异性插入事件。所述方法涉及制造与使用改造的锌指结合区域,其在共生同源基因序列(例如,EPSPS 基因序列)处或其附近结合,融合于剪切域(或剪切半域)以在细胞 DNA 中造成一个或更多靶向双链断裂。因为细胞 DNA 中双链断裂在所述剪切位点附近数千倍的

激发细胞修复机理,上述靶向剪切使得基因序列(例如,EPSPS)的改变或取代(通过同源导向修复)可以发生在基因组中几乎任何位点。

[0349] 本文所述方法可应用于来自任何生物或物种的任何共生同源基因(例如,EPSPS)序列。在有些实施方案中,所述改变的EPSPS靶基因组区域属于包含选自下组的核苷酸序列的EPSPS基因:SEQ ID NOS:10-14或预期具有至少约80-100%序列同一性度的,包含在此区间内任何百分比的相同度,例如与其具有81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99%序列同一性度。

[0350] 除本文所述的融合分子外,对选定基因组序列的靶向取代亦需要导入所述取代(或供体)序列。所述供体序列可在表达所述融合蛋白之前,同时或之后导入所述细胞。所述供体多核苷酸针对基因组序列(例如,EPSPS)含有足够的同源性以支持其同与之具有同源性的EPSPS基因组序列间的同源重组(或同源导向修复)。供体与基因组序列间大约25,50,100,200,500,750,1000,1500,2000或更多个核苷酸的序列同源性(或任何在10到2000之间整数个核苷酸,或更多)将支持其间的同源重组。供体序列可在从10到5000个核苷酸(或任何其间整数个核苷酸)或更长的长度范围内浮动。显而易见所述供体序列通常与其取代的基因组序列不完全相同。举例而言,所述供体多核苷酸的序列可较之所述基因组序列含有一个或更多单碱基变化,插入,缺失,倒位或重组,只要存在与染色体序列足够的同源性。或者,供体序列可含有两侧为两个同源区域的非同源序列。此外,供体序列可包括含有下述序列的载体分子,即其与细胞染色质中的目标区域不同源。一般而言,供体序列的同源区域将具有与期望其与之进行重组的基因组序列至少50%的序列同一性度。在有些实施方案中,存在60%,70%,80%,90%,95%,98%,99%或99.9%的序列同一性度。可存在任何在1%与100%之间值的序列同一性度,取决于所述供体多核苷酸的长度。

[0351] 供体分子可含有数个,非毗连的区域,其与细胞染色质同源。举例而言,为靶向插入通常不存在于目标区域的序列,所述序列可存在于供体核酸分子上,且其两侧为与目标区域的基因序列具有同源性的区域。

[0352] 为使确定成功插入所述供体序列的分析(例如,杂交,PCR,限制性酶消化)简单化,有些相对EPSPS基因组序列的序列差异可能存在于供体序列中。优选地,若其位于编码区域,则上述核苷酸序列差异将不改变氨基酸序列,或将造成沉默氨基酸变化(亦即,不影响所述蛋白质结构或功能的改变)。所述供体多核苷酸可任选的含有对应于在目标区域锌指结构域结合位点的序列改变,以阻止剪切已通过同源重组导入细胞染色质的供体序列。

[0353] 所述供体多核苷酸可为DNA或RNA,单链或双链,且可以线性或环状形式导入细胞。若以线性形式导入,所述供体序列的两端可通过本领域技术人员已知的方法保护(例如,防止核酸外切性降解)。举例而言,一个或更多双脱氧核苷酸残基可添加到线性分子的3'末端,和/或将自身互补的寡核苷酸连接到一个或两个末端。参见,例如,Chang et al. (1987)Proc. Natl. Acad. Sci. USA 84:4959-4963; Nehls et al. (1996)Science 272:886-889。附加的保护外源多核苷酸防止降解的方法包括但不限于,添加末端氨基基团或使用修饰的核苷酸间连接,例如,磷硫酰化合物,氨基磷酸酯,以及O-甲基核糖或脱氧核糖残基。

[0354] 多核苷酸可作为载体分子的一部分导入细胞,所述载体分子具有其它序列,例如,复制起点,启动子与编码抗生素抗性的基因。另外,供体多核苷酸可作为裸露核酸,作为与

试剂,例如脂质体或聚羟体 (poloxamer) 的复合物导入,或这可通过细菌或病毒(例如,土壤杆菌,根瘤菌属 (*Rhizobium* sp.) NGR234,苜蓿中华根瘤菌 (*Sinorhizobium meliloti*),百脉根中生根瘤菌 (*Mesorhizobium loti*),烟草花叶病病毒,马铃薯病毒 X,花椰菜花叶病病毒与木薯叶脉花叶病病毒)。参见,例如,Chung et al. (2006) Trends Plant Sci. 11(1): 1-4。

[0355] 虽不愿拘于单一理论,但看来在细胞序列中双链断裂的存在,加上与邻近或围绕所述断裂的区域具有同源性的外源 DNA 分子的存在,激活下述细胞机理,即其通过将序列信息从所述供体分子转移到细胞(例如,基因组或染色体)序列中来修复所述断裂;亦即,通过同源导向修复过程,亦称为“基因转换”。申请人的方法有利的将改造 ZFP 的强力靶向能力与剪切域(或剪切半域)组合以特异性的靶向共生同源基因,例如 EPSPS 基因,从而使靶序列的剪切在下述基因组区域产生双链断裂,即在所述区域期望插入外源序列。

[0356] 为改变染色体序列,并不需要将供体的全序列拷贝入所述染色体,只要拷贝足够多的供体序列以实现所期望的序列改变。

[0357] 通过同源重组插入供体序列的效率与细胞 DNA 中双链断裂与期望重组发生的位点间的距离负相关。换言之,当双链断裂越靠近期望重组发生的位点时,同源重组的效率就越高。当其中准确的重组位点未预决时(例如,所期望的重组事件可在一段基因组序列范围上发生时),选择所述供体序列的长度与序列,以及剪切位点,以获取所期望的重组事件。当其中所述期望的事件设计为在基因组序列中改变一个单独核苷酸对的序列时,细胞染色体在该核苷酸对任一侧 10000 个核苷酸以内受剪切。在有些实施方案中,剪切发生在距需改变其序列的核苷酸对任一侧 1,000,500,200,100,90,80,70,60,50,40,30,20,10,5 或 2 个核苷酸以内,或为任何在 2 到 1000 之间整数值的核苷酸。

[0358] 如上所详述,两个融合蛋白,每个均包括锌指结合区域与剪切半域,其结合位点,自每个结合位点靠近另一个结合位点最近的边缘测定,可为 5-8 或 15-18 个核苷酸以外,且剪切发生在所述结合位点之间。剪切是发生在所述结合位点间的单一位点处还是多个位点处并不重要,这是因为受剪切的基因组序列由供体序列所取代。从而,为有效的通过靶向重组改变一个单独核苷酸对的序列,结合位点间区域的中点为距该核苷酸对 10000 核苷酸以内,优选 1000 核苷酸以内,或 500 个核苷酸以内,或 200 个核苷酸以内,或 100 个核苷酸以内,或 50 个核苷酸以内,或 20 个核苷酸以内,或 10 个核苷酸以内,或 5 个核苷酸以内,或 2 个核苷酸以内,或 1 个核苷酸以内,或就在目标核苷酸对处。

[0359] 在有些实施方案中,同源染色体可适用作供体多核苷酸。从而,例如,在异源杂合体中突变的校正可通过改造出下述融合蛋白来达成,即所述蛋白结合于并剪切在一个染色体上的所述突变序列,但并不剪切在同源染色体上的野生型序列。在携带突变的染色体上的双链断裂激发了基于同源的“基因转换”过程,其中来自同源染色体的野生型序列被拷贝入受剪切的染色体,从而恢复了两个拷贝的野生型序列。

[0360] 亦提供了可能增强靶向重组水平的方法与组合物,包括但不限于,使用附加的 ZFP- 功能性结构域融合物以激活涉及同源重组的基因的表达,例如, RAD52 上位显性组 (epistasis group) 成员(例如, Rad50, Rad51, Rad51B, Rad51C, Rad51D, Rad52, Rad54, Rad54B, Mre11, XRCC2, XRCC3),其产物与前述基因产物(例如, BRCA1, BRCA2)相互作用的基因,和/或在 NBS1 复合物中的基因。参见,例如,Boyko et al. (2006) Plant Physiology

141:488-497 and LaFarge et al. (2003) *Nucleic Acids Res* 31(4):1148-1155。可使用类似的 ZFP- 功能性结构域融合物,与本文公开的方法与组合物组合,以阻抑涉及非同源末端连接的基因表达(例如, Ku70/80, XRCC4, poly(ADP ribose) polymerase, DNA ligase 4)。参见,例如, Riha et al. (2002) *EMBO* 21:2819-2826; Freisner et al. (2003) *Plant J.* 34:427-440; Chen et al. (1994) *European Journal of Biochemistry* 224:135-142。公开了使用锌指结合区域与功能性结构域间融合物激活与阻抑基因表达的方法,例如,于共有的 US Patents 6,534,261; 6,824,978 与 6,933,113 中。其它阻抑方法包括使用向需要阻抑的基因靶向的反义寡核苷酸和 / 或小干扰 RNA (siRNA 或 RNAi) 或 shRNA。

[0361] 作为激活涉及同源重组的基因产物表达的备选方法或附加方法,这些蛋白质(或其功能性片段)与以目标基因组区域(例如, EPSPS)为靶的锌指结合区域的融合物,可用于将这些蛋白质(重组蛋白质)招致到目标区域,从而增加其局部浓度,并进一步激发同源重组过程。或者,如上所述涉及同源重组的多肽(或其功能性片段)可为包含锌指结合区域,剪切域(或剪切半域)以及所述重组蛋白(或其功能性片段)的三重融合蛋白的一部分。其它涉及基因转换与重组相关染色质重建,且可用于前述方法与组合物的蛋白质,包括组蛋白乙酰转移酶(例如, Esa1p, Tip60),组蛋白甲基转移酶(例如, Dot1p),组蛋白激酶与组蛋白磷酸酶。参见,例如, Bhat et al. (1999) *Plant J.* 33:455-469。

[0362] 在包含锌指 / 核酸酶融合分子与供体 DNA 分子的细胞中,靶向重组效率的进一步增加,通过将细胞阻遏于细胞周期的 G<sub>2</sub> 期来达成,在该期同源性驱动的修复过程具最大限活性。上述停止可用多种方法达成。举例而言,细胞可经下述物质处理,例如,药物,影响细胞周期进行从而将细胞停止在 G<sub>2</sub> 期的化合物和 / 或小分子。这类分子的示例包括但不限于,影响微管多聚化的化合物(例如,长春花碱,诺考达唑(nocodazole),紫杉醇),与 DNA 相互作用的化合物(例如,顺-铂(II)二胺二氯化物,顺铂,阿霉素)和 / 或影响 DNA 合成的化合物(例如,胸腺嘧啶核苷,羟脲, L- 含羞草碱,依托泊苷(etoposide),5- 氟代尿嘧啶)。重组效率的额外增加通过使用组蛋白脱乙酰酶(HDAC)抑制剂(例如,丁酸钠,曲古抑菌素(trichostatin A))达成,所述抑制剂改变染色质结构以使细胞重组装置更易接近基因组 DNA。

[0363] 其它停止细胞周期的方法包括过表达抑制 CDK 细胞周期激酶活性的蛋白,例如,通过将编码该蛋白的 cDNA 导入所述细胞或者通过将激活编码该蛋白的基因表达的改造 ZFP 导入所述细胞。细胞周期停止亦可通过抑制细胞周期蛋白与 CDK 活性的方法来达成,例如,使用 RNAi 方法(例如, U. S. Patent No. 6,506,559)或通过将下述改造 ZFP 导入所述细胞,即所述 ZFP 阻抑一种或更多涉及细胞周期进行的基因,例如,细胞周期蛋白和 / 或 CDK 基因的表达。关于合成改造锌指蛋白以调节基因表达的方法参见,例如,共有的 U. S. Patent No. 6,534,261。

[0364] 或者,在有些情况下,在缺乏供体多核苷酸的情况下进行靶向剪切(优选在 S 或 G<sub>2</sub> 期),而重组发生在同源染色体间。

[0365] 表达载体

[0366] 编码一个或更多 ZFP 的核酸可克隆入载体内以供转化原核或真核细胞以进行复制和 / 或表达。载体可为原核载体,例如,质粒,或穿梭载体,昆虫载体或真核载体。编码 ZFP 的核酸亦可克隆入表达载体内,以施用于植物细胞。

[0367] 为表达 ZFP, 编码 ZFP 的序列通常亚克隆入含有启动子以引导转录的表达载体中。合适的细菌与真核启动子在本领域为众所周知的, 且描述于, 例如, Sambrook et al., *Molecular Cloning, A Laboratory Manual* (2nd ed. 1989; 3<sup>rd</sup> ed., 2001); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); and *Current Protocols in Molecular Biology* (Ausubel et al., 见上) 中。针对, 例如, 大肠杆菌, 芽孢杆菌属, 沙门氏菌属的表达所述 ZFP 的细菌表达系统是可获取的 (Palva et al., *Gene* 22: 229-235 (1983))。上述表达系统的试剂盒是可市购的。针对哺乳类细胞, 酵母与昆虫细胞的真核表达系统对本领域技术人员是众所周知的, 且亦为可市购的。

[0368] 用于引导 ZFP 编码核酸表达的启动子取决于特定的用途。举例而言, 适于宿主细胞的强组成型启动子通常用于表达与提纯 ZFP。

[0369] 相反, 当在体内施用 ZFP 以调节植物基因时 (参见下面的“对植物细胞的核酸递送”部分), 使用组成型或诱导型的启动子, 取决于所述 ZFP 的特定用途。植物启动子的非限定性示例包括源自 *A. thaliana* 泛肽 -3 (ubi-3) (Callis, et al., 1990, *J. Biol. Chem.* 265-12486-12493); *A. tumifaciens* 甘露碱合酶 ( $\Delta$ mas) (Petolino et al., U. S. Patent No. 6, 730, 824); 和 / 或木薯叶脉花叶病病毒 (CsVMV) (Verdaguer et al., 1996, *Plant Molecular Biology* 31:1129-1139) 的启动子序列。亦参见实施例。

[0370] 除启动子外, 表达载体通常包含转录单元或表达盒, 其包含需用于在宿主细胞, 无论是原核还是真核细胞中表达核酸的所有额外元件。通常的表达盒从而含有启动子, 其可操作性的连接于, 例如, 编码所述 ZFP 的核酸序列, 以及信号, 其需用于, 例如, 转录物的有效多聚腺苷化, 转录终止, 核糖体结合位点与翻译终止。其它盒元件可能包括, 例如, 增强子, 以及异源剪接信号。

[0371] 用于将遗传信息转送入细胞的特定表达载体根据 ZFP 的预期用途选取, 例如, 在植物, 动物, 细菌, 真菌, 原生生物等等中的表达 (参见下面所述的表达载体)。标准细菌与动物表达载体在本领域为已知的, 且详述于, 例如, U. S. Patent Publication 20050064474A1 以及 International Patent Publications W005/084190, W005/014791 与 W003/080809 中。

[0372] 标准转染方法可用于产生下述细菌, 哺乳类, 酵母或昆虫细胞系, 即其表达大量的蛋白质, 所述蛋白质可随即使用标准技术提纯 (参见, 例如, Colley et al., *J. Biol. Chem.* 264:17619-17622 (1989); *Guide to Protein Purification, in Methods in Enzymology*, vol. 182 (Deutscher, ed., 1990))。真核与原核细胞的转化根据标准技术实施 (参见, 例如, Morrison, *J. Bact.* 132:349-351 (1977); Clark-Curtiss & Curtiss, *Methods in Enzymology* 101:347-362 (Wu et al., eds., 1983))。

[0373] 可使用任何众所周知的将外源核苷酸序列导入上述宿主细胞中的过程。这些包括使用磷酸钙转染, polybrene, 原生质体融合, 电穿孔, 超声波方法 (例如, 声致孔 (sonoporation)), 脂质体, 显微注射, 裸露 DNA, 质粒载体, 病毒载体, 包括附加体性的和整合性的, 以及其它任何众所周知的方法将克隆的基因组 DNA, cDNA, 人工合成 DNA 或其它外源基因材料导入宿主细胞中 (参见, 例如, Sambrook et al., 见上)。仅仅需要所用的特定遗传工程过程能够成功的将至少一个基因导入能够表达所选蛋白质的宿主细胞。

[0374] 针对植物细胞的核酸递送



[0375] 如上所指明, DNA 构建体可通过多种常规技术导入(例如, 导入其基因组)所期望的植物宿主。关于上述技术的综述, 参见, 例如, Weissbach & Weissbach *Methods for Plant Molecular Biology* (1988, Academic Press, N. Y.) Section VIII, pp. 421-463; 以及 Grierson & Corey, *Plant Molecular Biology* (1988, 2d Ed.), Blackie, London, Ch. 7-9。

[0376] 举例而言, 所述 DNA 构建体可能使用例如电穿孔与对植物细胞原生质体的显微注射的方法直接导入所述植物细胞的基因组 DNA, 或者所述 DNA 构建体可使用生物弹射击 (biolistic) 方法, 例如 DNA 颗粒轰击 (参见, 例如, Klein et al (1987) *Nature* 327 :70-73) 直接导入植物组织。或者, 所述 DNA 构建体可能与合适的 T-DNA 侧翼区域合并, 并导入常规的根癌土壤杆菌宿主载体。根癌土壤杆菌介导转化技术, 包括解除武装 (disarming) 与使用二元载体, 在科学文献中得到充分描述。参见, 例如, Horsch et al (1984) *Science* 233 :496-498, and Fraley et al (1983) *Proc. Nat'l Acad. Sci. USA* 80 :4803。

[0377] 另外, 基因转移可能使用非农杆菌细菌或病毒, 例如, *Rhizobium* sp. NGR234, *Sinorhizobium meliloti*, *Mesorhizobium loti*, 马铃薯病毒 X, 花椰菜花叶病病毒与木薯叶脉花叶病病毒和 / 或烟草花叶病病毒。参见, 例如, Chunget al. (2006) *Trends Plant Sci.* 11(1) :1-4。

[0378] 当细胞使用 binary T DNA 载体 Bevan (1984) *Nuc. Acid Res.* 2 :8711-8721) 或共培养过程 (Horsch et al (1985) *Science* 227 :1229-1231) 由细菌感染时, 所述根癌土壤杆菌宿主的毒性功能将引导所述构建体与邻近的标记插入植物细胞的 DNA。一般而言, 所述土壤杆菌转化系统用于改造双子叶植物 (Bevan et al (1982) *Ann. Rev. Genet* 16 :357-384; Rogers et al (1986) *Methods Enzymol.* 118 :627-641)。所述土壤杆菌转化系统亦可用于转化, 以及转移 DNA 于单子叶植物与植物细胞。参见 U. S. Patent No. 5, 591, 616; Hernalsteen et al (1984) *EMBO J* 3 :3039-3041; Hooykass-Van Slogteren et al (1984) *Nature* 311 :763-764; Grimsley et al (1987) *Nature* 325 :1677-179; Boulton et al (1989) *Plant Mol. Biol.* 12 :31-40. ; 以及 Gould et al (1991) *Plant Physiol.* 95 :426-434。

[0379] 备选的基因转移与转化方法包括但不限于, 通过钙离子-, 聚乙二醇 (PEG)- 或电穿孔-介导裸露 DNA 的摄入进行原生质体转化 (参见 Paszkowski et al. (1984) *EMBO J* 3 :2717-2722, Potrykus et al. (1985) *Molec. Gen. Genet.* 199 :169-177; Fromm et al. (1985) *Proc. Nat. Acad. Sci. USA* 82 :5824-5828; and Shimamoto (1989) *Nature* 338 :274-276) 以及对植物组织进行电穿孔 (D' Halluin et al. (1992) *Plant Cell* 4 :1495-1505)。其它植物细胞转化的方法包括显微注射, 碳化硅介导 DNA 摄入 (Kaeppler et al. (1990) *Plant Cell Reporter* 9 :415-418), 以及微粒轰击 (参见 Klein et al. (1988) *Proc. Nat. Acad. Sci. USA* 85 :4305-4309; 以及 Gordon-Kamm et al. (1990) *Plant Cell* 2 :603-618)。

[0380] 本公开的方法与组合物可用于将外源序列插入植物细胞基因组中预决的位置。这是有用的, 考虑到导入植物基因组的转基因表达严重取决于其整合位点。相应的, 编码, 例如, 养分, 抗生素或治疗分子的基因可通过靶向重组插入到植物基因组中对其表达有利的区域。

[0381] 通过上述任何转化技术产生的转化植物细胞可经培养再生出完整植物, 其拥有转化的基因型以及由此, 所期望的表型。上述再生技术依赖于在组织培养生长培养基中操作某些植物激素, 通常依赖于抗微生物剂和 / 或除草剂标记, 其可与所期望的核苷酸序列一

同导入。自培养原生质体的植物再生描述于“*Protoplasts Isolation and Culture*” in *Handbook of Plant Cell Culture*, pp.124-176, Macmillian Publishing Company, New York, 1983; 以及 *Binding, Regeneration of Plants, Plant Protoplasts*, pp. 21-73, CRC Press, Boca Raton, 1985。再生亦可得自植物愈伤组织, 外植体, 器官, 花粉, 胚或其部分。上述再生技术一般的描述于 Klee et al (1987) *Ann. Rev. of Plant Phys.* 38 :467-486。

[0382] 导入植物细胞的核酸可用于给予基本上任何植物以所期望的性状。使用现行公开的核酸构建体与上面所述多种转化方法, 广泛种类的植物与植物细胞系统可经改造以获取本文所述的所期望的生理学和农学特征。在优选实施方案中, 需改造的目标植物与植物细胞包括但不限于, 那些单子叶与双子叶植物, 例如作物, 包括谷类作物 (例如, 小麦, 玉米, 稻, 黍, 大麦), 果实作物 (例如, 西红柿, 苹果, 梨, 草莓, 橙子), 草料作物 (例如, 苜蓿), 根菜作物 (例如, 胡萝卜, 马铃薯, 甜菜, 甘薯 (yam)), 叶菜作物 (例如, 莴苣 (lettuce), 菠菜); 有花植物 (例如, 牵牛花, 玫瑰, 雏菊), 松柏类树木 (例如, 松杉, 云杉); 用于 phytoremediation 的植物 (例如, 积聚重金属的植物); 油料作物 (例如, 向日葵, 油菜籽) 以及实验用的植物 (例如, 拟南芥 (*Arabidopsis*))。从而, 本公开的方法与组合物具有对广泛范围的植物的用途, 包括但不限于, 来自天门冬属 (*Asparagus*), 燕麦属 (*Avena*), 芸薹属 (*Brassica*), 柑橘属 (*Citrus*), 西瓜属 (*Citrullus*), 辣椒属 (*Capsicum*), 南瓜属 (*Cucurbita*), 胡萝卜属 (*Daucus*), 飞蓬属 (*Erigeron*), 大豆属 (*Glycine*), 棉属 (*Gossypium*), 大麦属 (*Hordeum*), 莴苣属 (*Lactuca*), 黑麦草属 (*Lolium*), 番茄属 (*Lycopersicon*), 苹果属 (*Malus*), 木薯属 (*Manihot*), 烟草属 (*Nicotiana*), 诸葛菜属 (*Orychophragmus*), 稻属 (*Oryza*), 鳄梨属 (*Persea*), 菜豆属 (*Phaseolus*), 豌豆属 (*Pisum*), 梨属 (*Pyrus*), 李属 (*Prunus*), 萝卜属 (*Raphanus*), 黑麦属 (*Secale*), 茄科 (*Solanum*), 高粱属 (*Sorghum*), 小麦属 (*Triticum*), 葡萄属 (*Vitis*), 豇豆属 (*Vigna*) 与玉蜀黍属 (*Zea*)。

[0383] 本领域技术人员将认定在所述表达盒稳定的掺入转基因植物并确证为可操作的之后, 其可通过有性杂交导入其它植物中。多种标准育种技术任意均可使用, 取决于需进行杂交的物种。

[0384] 转化的植物细胞, 愈伤组织, 组织或植物可根据由存在于所述转化 DNA 上的标记基因编码的性状通过选择或筛选所述改造植物材料进行识别与分离。举例而言, 可通过将改造植物材料在下述培养基中生长而实施选择, 即所述培养基含有抗生素或除草剂的量足以具有抑制性, 但针对其所述转化基因构建体提供抗性。进一步, 转化植物与植物细胞亦可能通过筛选任何可见标记基因 (例如,  $\beta$ -葡萄糖醛酸苷酶, 荧光素酶, B 或 C1 基因) 的活性来识别, 所述标记基因可能存在于重组核酸构建体上。上述选择与筛选方法对本领域技术人员是众所周知的。

[0385] 亦可使用物理与生物化学方法以识别包含插入基因构建体的植物或植物细胞转化体。这些方法包括但不限于: 1) Southern 分析或 PCR 扩增以检测与确定重组 DNA 插入的结构; 2) Northern blot, S1 RNase 保护, 引物延伸或反转录酶-PCR 扩增以检测与检查所述基因构建体的 RNA 转录物; 3) 酶分析以检测酶或核酶活性, 其中上述基因产物由所述基因构建体编码; 4) 蛋白质凝胶电泳, Western blot 技术, 免疫沉淀或酶联免疫分析, 其中所述基因构建体产物为蛋白质。其它技术, 例如原位杂交, 酶染色以及免疫染色亦可用于在特定植物器官与组织中检测所述重组构建体的存在或表达。实施所有这些分析的方法对本领

域技术人员为众所周知的。

[0386] 使用本文公开方法进行基因操作的结果可通过,例如,对自目标组织分离的 RNA 进行 Northern blot 来加以观察。通常,若 mRNA 的量增加,则可假定其相应的内源基因较之之前表达速率增加。可使用其它测定基因和 / 或 CYP74B 活性的方法。可使用不同类型的酶分析,取决于所使用的底物以及检测反应产物或副产物增加或减少的方法。此外,CYP74B 水平和 / 或表达量可用免疫化学方法,亦即,ELISA,RIA,EIA 与其它对本领域技术人员为众所周知的基于抗体的分析,例如通过电泳检测分析(可进行染色或 Westernblotting),来进行测定。转基因可选择性的表达于所述植物的某些组织或某些发展阶段,或所述转基因可基本上在所有植物组织中,基本上在其整个生活史中表达。然而,任何组合式的表达模式亦可适用。

[0387] 现行公开亦涵盖了上述转基因植物的种子,其中所述种子具有所述转基因或基因构建体。现行公开进一步涵盖了上述转基因植物的后代,克隆,细胞系或细胞,其中所述后代,克隆,细胞系或细胞具有所述转基因或基因构建体。

[0388] ZFP 与编码 ZFP 的表达载体可直接施用于植物以供基因调节,靶向剪切和 / 或重组。在有些实施方案中,所述植物包含多个共生同源靶基因。已知植物可含有多个共生同源基因,举例而言,欧洲油菜具有 5 个共生同源的 EPSPS 基因 (SEQ ID NOS :10-14),其可为一个或更多 ZFP 的靶(参见实施例)。从而,一个或更多不同 ZFP 或编码 ZFP 的表达载体可施用于植物以在该植物中靶向一个或更多 EPSPS 基因。举例而言,对 1,2,3,4,5 或直至任何数量的共生同源基因(例如,EPSPS 共生同源基因)或存在于植物中的所有共生同源基因(例如,EPSPS 共生同源基因)可进行靶向。

[0389] 在有些实施方案中,作为靶的 EPSPS 基因包含选自下组的核苷酸序列:SEQ ID NOS :10-14 或与其至少有大约 80% -100% 同一性的序列,包括任何在该范围内的同一性百分比,例如与其有 81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99% 序列同一性。

[0390] 有效量的使用可通过通常用于引导 ZFP 与所需处理的植物细胞最终接触的任何途径。所述 ZFP 以任何合适的方法施用,优选使用医药上可接受的载体。施用上述调整剂的合适方法是可用的,且对本领域技术人员为众所周知的,而且,尽管可使用多于一种途径施用特定组合物,特定途径常常可提供较之其它途径更加迅捷而更加有效的反应。

[0391] 载体亦可部分根据所施用的特定组合物,以及用于施用该组合物的特定方法使用与确定。相应的,有广泛种类的医药组合物的合适配制可供使用。

[0392] 应用

[0393] 农业上一个特定的目标领域是遗传改良植物使其具有除草剂抗性。许多除草剂通过抑制植物需用于生长的关键酶或蛋白质来其作用。举例而言,除草剂草甘膦通过抑制合成芳香族氨基酸的酶 5-烯醇式丙酮酰-3-磷酸莽草酸合酶(EPSPS)来杀灭植物。对草甘膦具有耐受性的植物已通过将 EPSPS 转基因插入植物基因组,过表达 EPSPS,并选择性的突变 EPSPS 以产生草甘膦抗性突变体来产生(参见,例如,U. S. Pat. Nos. 5,312,910 与 6,987,213;以及 Papanikou et al. (2004) *Planta* 218(4):589-598)。

[0394] 举例而言,所述公开方法与组合物可用于调整表达以及 EPSPS 基因的靶向改变。在一方面,改造 ZFP 用于在植物中上调与下调 EPSPS 的表达。ZFP 可选的与供调整基因表

达的调节区相联系,其可为共价或非共价的联系,且激活或阻抑 EPSPS 基因转录。上述 ZFP 可用于增加或减少所述 EPSPS 酶的产生,在植物中控制芳香族氨基酸的生物合成,或者增加或减少植物对除草剂草甘膦的耐受性,例如,使作物对除草剂草甘膦有抗性,增加作物产量,或消除杂草或野生植物对草甘膦的抗性。

[0395] 包含一个或更多 ZFP 的组合物,或编码其的核苷酸,可施用于植物细胞。在一个实施方案中,至少两个识别 EPSPS 基因上或者相同靶序列或者不同靶序列的 ZFP,或编码上述 ZFP 的多核苷酸,施用于细胞。第一个 ZFP 可选的与第二个 ZFP 相联系,可为共价的或非共价的。由或者相联系的或者单独的 ZFP 识别邻近靶位点可用于产生 ZFP 的协同结合,导致较之单独结合于其各自靶位点的 ZFP 亲和力更大的亲和力。

[0396] 对于阻抑基因表达,通常将所述基因的表达减少约 20% (亦即,80%的非 ZFP 调整表达),较优选约 50% (亦即,50%的非 ZFP 调整表达),更优选约 70-100% (亦即,25%到 0%的非 ZFP 调整表达)。对于激活基因表达,通常将表达激活约 1.5 倍 (亦即,150%的非 ZFP 调整表达),优选 2 倍 (亦即,200%的非 ZFP 调整表达),更优选 5-10 倍 (亦即,500-1000%的非 ZFP 调整表达),直至至少 100 倍或更多。

[0397] 改造 ZFP 激活剂与阻抑剂的表达亦可通过小分子调节系统,例如 tet- 调节系统与 RU-486 系统来控制 (参见,例如, Gossen & Bujard, PNAS 89:5547(1992); Oligino et al., Gene Ther. 5:491-496(1998); Wang et al., Gene Ther. 4:432-441(1997); Neering et al, Blood 88:1147-1155(1996); and Rendahl et al., Nat. Biotechnol. 16:757-761(1998))。这些调节系统对所述 ZFP 激活剂与阻抑剂的表达给予小分子控制,且对所述目标基因 (例如, EPSPS) 施加进一步水平的控制。

[0398] 在另一方面, ZFN 用于在 EPSPS 基因组序列中诱导突变,例如,通过在两个位点剪切并缺失其间序列,通过在一个单独位点剪切继以非同源末端连接,在一个或两个位点剪切,并在所述断裂间插入外源序列和 / 或通过在位点剪切从而移除一个或两个或几个核苷酸。靶向剪切亦可用于构建基因敲除 (例如,以供功能性基因组学或靶标确认) 与协助将序列靶向插入基因组 (亦即,基因敲入 (knock-in)); 例如,为细胞工程或蛋白质过表达的目的。插入可通过下述方法进行,即通过同源重组或靶向整合取代染色体序列,其中将两侧具有与染色体上目标区域同源的序列的新序列 (亦即,不存在于目标区域内的序列) 插入于预决的靶位点。同样的方法亦可用于用突变体 EPSPS 基因序列取代野生型 EPSPS 基因序列,或将一个等位基因变换为不同的等位基因。本文所描述的组合物与方法亦可用于产生具有可诱导的 ZFP 和 / 或稳定整合入基因组的 ZFN 的植物种系。相应的,编码所述含锌指蛋白质的稳定整合序列可在合适的诱导下表达以在所述植物中经历多个植物世代以及在植物发育的任何阶段达成所期望的效果。

[0399] 此外,感染或整合植物病原体的靶向剪切可用于在植物宿主中通过下述方法治疗病原体感染:例如,通过剪切所述病原体的基因组从而使其病原性减少或消失。此外,靶向剪切编码植物病毒受体的基因可用于阻遏上述载体的表达,从而在植物中阻止病毒感染和 / 或病毒扩散。

[0400] 植物病原体的示例包括但不限于:植物病毒,例如苜蓿花叶病毒 (Alfamoviruses), 甲型隐蔽病毒 (Alphacryptoviruses), 香蕉病毒 (Badnaviruses), 乙型隐蔽病毒 (Betacryptoviruses), 二双粒病毒 (Bigeminiviruses), 雀麦花叶病

毒 (Bromoviruses), 大麦黄花叶病毒 (Bymoviruses), 发状病毒 (Capilloviruses), 香石竹潜隐病毒 (Carlaviruses), 香石竹斑驳病毒 (Carmoviruses), 花椰菜花叶病毒 (Caulimoviruses), 黄化线状病毒 (Closteroviruses), 豇豆花叶病毒 (Comoviruses), 黄瓜花叶病毒 (Cucumoviruses), 质型弹状病毒 (Cytorhabdoviruses), 香石竹环斑病毒 (Dianthoviruses), 耳突花叶病毒 (Enamoviruses), 蚕豆病毒 (Fabaviruses), 斐济病毒 (Fijiviruses), 真菌传杆状病毒 (Furoviruses), 大麦病毒 (Hordeiviruses), 杂双粒病毒 (Hybrigeminiviruses), 悬钩子病毒 (Idaeoviruses), 等变环点病毒 (Ilarviruses), 甘薯病毒 (Ipomoviruses), 黄症病毒 (Luteoviruses), 玉米褪绿斑驳病毒 (Machlomoviruses), 柘橙病毒 (Macluraviruses), 玉蜀黍雷亚多精致病毒 (Marafiviruses), 单双粒病毒 (Monogeminiviruses), 矮缩病毒 (Nanaviruses), 坏死病毒 (Necroviruses), 线虫传多面体病毒 (Nepoviruses), 核型弹状病毒 (Nucleorhabdoviruses), 水稻病毒 (Oryzaviruses), 欧尔密病毒 (Ourmiaviruses), 植物呼肠孤病毒 (Phytoreoviruses), 马铃薯 x 病毒 (Potexviruses), 马铃薯 y 病毒 (Potyvirus), 黑麦草花叶病毒 (Rymovirus), 卫星 (satellite) RNAs, 卫星病毒 (satelliviruses), 随伴病毒 (Sequiviruses), 南方菜豆花叶病毒 (Sobemoviruses), 纤细病毒 (Tenuiviruses), 烟草花叶病毒 (Tobamovirus), 烟草脆裂病毒 (Tobravirus), 番茄丛矮病毒 (Tomusvirus), 番茄斑萎病毒 (Tospovirus), 纤毛病毒 (Trichovirus), 芜菁黄花叶病毒 (Tymovirus), 幽影病毒 (Umbravirus), 叶脉曲张病毒 (Varicosavirus) 与矮化病毒 (Waikavirus); 真菌病原体, 例如黑穗病 (例如黑粉菌 (Ustilaginales)), 锈病 (锈菌 (Uredinales)), 麦角症 (麦角菌 (Claviceps pupurea)) 与霉病; 霉菌 (卵菌 (Oomycetes)) 例如马铃薯晚疫病菌 (*Phytophthora infestans*) (马铃薯疫病); 细菌病原体例如欧文氏菌 (*Erwinia*) (例如, 草生欧文氏菌 (*E. herbicola*)), 假单胞菌 (*Pseudomonas*) (例如, 绿脓杆菌 (*P. aeruginosa*), 丁香假单胞菌 (*P. syringae*), 荧光假单胞菌 (*P. fluorescense*) 与恶臭假单胞菌 (*P. putida*)), 劳尔氏菌 (*Ralstonia*) (例如, 茄科劳尔氏菌 (*R. solanacearum*)), 土壤杆菌 (*Agrobacterium*) 与黄单胞菌 (*Xanthomonas*); 线虫 (线虫动物门 (Nematoda)); 以及植物粘液虫 (*Phytomyxa*) (多粘菌 (*Polymyxa*) 与根肿菌 (*Plasmodiophora*))。

[0401] 公开的靶向重组方法可用于用同源非相同序列取代一个或更多 EPSPS 基因组序列。举例而言, 突变体基因组序列可用野生型序列取代, 或者, 野生型基因组序列可用突变体序列取代, 用以, 例如, 使作物具有对除草剂草甘膦的抗性, 增加植物产量, 消除杂草或野生植物对草甘膦的抗性, 等等。同样的, 基因的一个等位基因可由不同等位基因使用本文所披露的靶向重组方法进行取代。

[0402] 在许多这样的情况下, 目标 EPSPS 基因组区域包含突变, 而供体多核苷酸包含相应的野生型序列。类似的, 若期望, 野生型基因组序列可由突变体序列取代。举例而言, 可用下述方法在植物中消除或减少草甘膦抗性, 即用野生型基因取代突变的或外源 EPSPS 基因, 移除外源 EPSPS 基因, 突变 EPSPS 基因至较低的草甘膦抗性, 或者用支持较低水平 EPSPS 表达的序列取代 EPSPS 基因的对照序列。或者, 在植物中可将突变导入 EPSPS 基因使其具有针对草甘膦的抗性, 或者通过突变所述 EPSPS 基因以产生草甘膦耐受性 EPSPS 酶, 或者通过用增加 EPSPS 表达水平的序列取代所述 EPSPS 基因的对照序列。增加针对除草剂草甘膦耐受性的 EPSPS 基因修饰与突变体 EPSPS 酶在本领域为已知的 (参见, 例如, U. S. Patent

Nos. 7, 238, 508, 7, 214, 535, 7, 141, 722, 7, 045, 684, 6, 803, 501, 6, 750, 377, 6, 248, 876, 6, 225, 114, 6, 040, 497, 5, 866, 775, 5, 804, 425, 5, 776, 760, 5, 633, 435, 5, 627, 061, 5, 554, 798, 5, 463, 175, 5, 312, 910, 5, 310, 667, 5, 188, 642, 5, 145, 783, 4, 971, 908 与 4, 940, 835, 以及 W000/66748 ;以引用方式包含于本文中)。

[0403] 靶向剪切与靶向重组亦可用于改变非编码序列(例如,调节序列如启动子,增强子,起始子,终止子,剪接位点)以改变EPSPS基因产物的表达水平。上述方法可用于,例如,增加草甘膦耐受性 EPSPS 变体在作物中的表达。

[0404] EPSPS 基因的失活可通过下述方式达成,例如,通过一个单独剪切事件,通过剪切继以非同源末端连接,通过在两个位点剪切继以连接从而缺失所述两个剪切位点间的序列,通过将错义或无义密码子靶向重组入编码区域,或通过将无关序列(亦即,“填充”序列)靶向重组入所述基因或其调节区域,从而破坏该基因或其调节区域。

[0405] 染色质结构的靶向修饰,如公开于共有的 WO 01/83793 中的,可用于促进融合蛋白与细胞染色质的结合。

[0406] 在其它实施方案中,除了或替代本文所公开的锌指-剪切域融合物,可使用一个或更多锌指结合区域与重组酶(或其功能性片段)的融合物,以促进靶向重组。参见,例如,共有的 US patent No. 6, 534, 261 与 Akopian et al. (2003) Proc. Natl. Acad. Sci. USA 100 : 8688-8691。

[0407] 在其它实施方案中,本公开的方法与植物用于提供其活性需要二聚作用(无论是同源二聚作用或异源二聚作用)的 ZFP 结合区域与转录激活或阻抑域的融合物。在这些情况下,融合多肽包含锌指结合区域与功能性域单体(例如,来自二聚转录激活或阻抑域的单体)。两个上述融合多肽结合于置于合适位置的靶位点允许二聚作用,从而重建功能性转录激活与阻抑域。

[0408] 进一步,如上所公开的,本文提出的方法与组合物可用于将外源序列靶向整合入细胞基因组中的目标区域(例如,EPSPS 基因的调节或编码序列),例如,其中剪切增强通过同源依存性机理进行插入的(例如,插入包含外源序列与一个或更多与预决的基因组序列(亦即,靶位点)或者为相同的,或者为同源非同一性的序列一起的供体序列)。

[0409] 如上所指明,在有些实施方案中,通过同源依存性以及同源非依存性机理进行靶向整合涉及将外源序列插入由剪切所产生的末端之间。所述插入的外源序列可为任何长度,例如,在 1 到 50 个核苷酸长度间较短的“补片”序列(例如,2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45 或 50 个核苷酸的序列)

[0410] 当其中靶向整合为同源依存性时,供体核酸或供体序列包含外源序列与一个或更多与预决的基因组序列(亦即,靶位点)或者为相同的,或者为同源非同一性的序列一起。在有些实施方案中,在所述外源序列的两侧存在两个相同序列,或两个同源非相同序列(或两者各一)。外源序列(或外源核酸或外源多核苷酸)为含有正常情况下不存在于目标区域内的核酸序列的。

[0411] 外源序列的示例包括但不限于, cDNA, 启动子序列, 增强子序列, 表位标记, 标记基因, 剪切酶识别位点与多种类型的表达构建体。标记基因包括但不限于, 编码介导化学或抗生素抗性(例如, 氨苄青霉素抗性, 卡那霉素抗性, G418 抗性, 潮霉素 B 抗性, 嘌呤霉素

抗性,好必思 (herbiace) 抗性) 的蛋白质的序列,编码有色或荧光或发光蛋白质(例如,绿色荧光蛋白,增强绿色荧光蛋白,红色荧光蛋白,荧光素酶)的序列,以及介导细胞生长和/或基因扩增的蛋白质(例如,四氢叶酸还原酶)。表位标记包括,例如,一个或更多拷贝的 FLAG, His, myc, Tap, HA 或者任何可检出的氨基酸序列。

[0412] 蛋白质表达构建体包括但不限于, cDNA 与可操纵的连接于 cDNA 的转录控制序列。转录控制序列包括启动子,增强子与绝缘子。其它可包含在表达构建体中的转录与翻译调节序列包括,例如,内核糖体进入位点,编码 2A 肽与多腺苷酸化信号的序列。蛋白质表达构建体的示例为包含编码抗体重链的序列与编码抗体轻链的序列的抗体表达构建体,每个序列可操纵的连接于启动子(所述启动子可为相同或不同),而所述序列其一或两者可选的可操纵的连接于增强子(且当两个编码序列均连接于增强子时,所述增强子可为相同或不同)。

[0413] 剪切酶识别位点包括,例如,由限制性核酸内切酶,靶向核酸内切酶和/或大范围核酸酶识别的序列。剪切酶识别位点的靶向整合(或者通过同源依存性或者通过同源非依存性机理)对产生其基因组仅包含一个单独的特定酶剪切位点的细胞为有用的。将上述细胞与识别并在该单独位点剪切的酶相接触促进随后外源序列的靶向整合(或者通过同源依存性或者通过同源非依存性机理)和/或在该受剪切的位点处的靶向突变。

[0414] 靶向核酸内切酶的示例包括 I-CeuI, PI-PspI, PI-Sce, I-SceIV, I-CsmI, I-PanI, I-SceI, I-SceII, I-PpoI, I-SceIII, I-CreI, I-TevI, I-TevII 与 I-TevIII。其识别序列为已知。亦参见 U. S. Pat. No. 6, 833, 252, U. S. Pat. No. 5, 420, 032 ;Belfort et al. (1997) Nucleic Acids Res. 25 :3379-3388 ;Dujon et al. (1989)Gene 82 :115-118 ;Perler et al. (1994)Nucleic Acids Res. 22,1125-1127 ;Jasin(1996)Trends Genet.12 :224-228 ;Gimble et al. (1996)J. Mol. Biol. 263 :163-180 ;Argast et al. (1998)J. Mol. Biol. 280 :345-353 与 New England Biolabs 产品目录。

[0415] 尽管大部分靶向核酸内切酶的剪切特异性相对其识别位点并非绝对,但所述位点足够长,从而使得通过在含有一个单独拷贝的其识别位点的细胞中表达靶向核酸内切酶可在每个哺乳类大小的基因组中获取一个单独的剪切事件。亦有报道称可改造剪切酶以结合非天然靶位点。参见,例如, Chevalier et al. (2002)Molec. Cell 10 :895-905 ;Epinat et al. (2003)Nucleic Acids Res. 31 :2952-2962 ;Ashworth et al. (2006)Nature 441 :656-659。

[0416] 之前使用靶向核酸内切酶以获取靶向重组与整合的方法受下述问题困扰,即所述识别位点的靶向插入极端低效,需要艰巨的筛选以识别含有所述识别位点插入所期望位置的细胞。现行方法通过允许高效的靶向整合 DNA 剪切酶识别位点(或者为同源依存性或者为同源非依存性)克服了这些问题。

[0417] 在有些实施方案中,靶向整合用于插入 RNA 表达构建体,例如,负责调节微 RNA, shRNA 或 siRNA 表达的序列。启动子,增强子与其它转录调节序列,如上所述,亦可掺入 RNA 表达构建体。

[0418] 在其中靶向整合通过同源依存性机理发生的实施方案中,供体序列在外源序列两侧的区域含有足够的同源性以支持在基因组序列中进行同源导向修复,从而将所述外源序列插入基因组靶位点。因此,所述供体核酸可为任何足以支持外源序列通过同源依存性修

复机理（例如，同源重组）整合的大小。虽不愿拘于任何特定理论，但在所述外源序列两侧的同源区域被认为向断裂染色体末端提供了模板以供在所述双链断裂位点重新合成遗传信息。

[0419] 外源序列的靶向整合，如本文所公开，可用于产生细胞与细胞系以供蛋白质表达。参见，例如共有的 U.S. Patent Application Publication No. 2006/0063231（基于所有目的，其公开由此以其全文引用的方式包含于本文中）。为由整合入基因组的外源序列编码的一种或更多蛋白质的最佳表达，所述染色体整合位点应与整合序列的高水平转录相容，优选在广泛细胞类型与发育状态中均如此。然而，观察到整合序列的转录依赖于整合位点而浮动，其余因素姑且不计，这是由于所述整合位点的基因组染色质结构。相应的，期望支持高水平整合序列转录的基因组靶位点。在有些实施方案中，亦期望外源序列的整合不导致一种或更多细胞基因（例如，癌基因）的异位激活。另一方面，当整合启动子和 / 或增强子序列时，可能期望异位表达。

[0420] 对有些实施方案，期望整合位点并不存在于必需基因（例如，对细胞存活必要的基因）中，从而使得外源序列的整合并不导致所述必需基因的失活。另一方面，如果意图为关闭基因功能（亦即，构建基因“敲除”），外源序列的靶向整合以破坏内源基因为有效的方法。在这些情况下，所述外源序列可为任何序列，只要其能够阻遏所述内源基因转录，或产生非功能性的翻译产物，例如，可能可检出的一小片氨基酸残基（见上）。在有些实施方案中，所述外源序列可包括标记基因（如上所述），允许对进行了靶向整合的细胞加以选择。

[0421] 其它支持高水平转录整合序列的基因组靶位点可识别为开放染色质或“易接近的区域”，如描述于，例如，共有的 U.S. Patent Application Publications 2002/0064802（2002 年 5 月 30 日）与 2002/0081603（2002 年 6 月 27 日）。

[0422] 在基因组序列中双链断裂的存在不仅促进外源序列的同源依存性整合（亦即，同源重组），亦促进外源序列在所述双链断裂位点同源非依存性整合入基因组。相应的，本文所公开的组合物与方法可用于靶向剪切基因组序列，继以外源序列非同源依存性整合于该靶剪切位点或其附近。举例而言，细胞可与一个或更多如本文所述经改造以在基因组中目标区域剪切的 ZFP- 剪切域（或剪切半域）融合蛋白（或一种或更多编码上述融合蛋白的多核苷酸）以及包含缺乏针对目标区域同源性的外源序列的多核苷酸相接触，以获取下述细胞，即其中所述外源序列的全部或部分整合入目标区域。

[0423] 本文所公开的靶向整合（亦即，将外源序列插入基因组）方法，无论是同源依存性的还是同源非依存性的，均可用于多种目的。这些包括但不限于，将基因或 cDNA 序列插入细胞基因组以使细胞能够表达所述基因或 cDNA 的转录和 / 或翻译产物。在其中疾病或病状可来自多种突变中一种的情况下（例如，多个点突变散布于所述具有的序列中），野生型基因 cDNA 拷贝的靶向整合（或者为同源依存性的或者为同源非依存性的）特别有效。举例而言，将上述野生型 cDNA 插入未翻译的先导序列或位于所有已知突变上游的基因第一个外显子。在有些整合体中，若其保持翻译阅读框，则其结果为所述野生型 cDNA 表达且其表达由合适的内源转录调节序列调节。在其它实施方案中，上述整合的 cDNA 序列可包括置于所述野生型 cDNA 下游，突变体内源基因上游的转录（和 / 或翻译）终止信号。以此方法，使致病基因的野生型拷贝表达，且所述变异体内源基因不表达。在其它的实施方案中，将野生型 cDNA 的一部分插入基因的合适区域中（例如，其中致病突变成簇聚集的基因）。



**[0424] 实施例**

[0425] 下列为实施现行公开的特定实施方式的实施例。所述实施例仅为说明目的而提供,且并无以任何方式限定现行公开范围的意图。

[0426] 经努力保证所用数值(例如,量,温度,等等)的准确度,但理所当然的,应容许一些实验误差与偏差。

[0427] 实施例 1:在欧洲油菜中识别靶序列

**[0428] A. 序列鉴定**

[0429] 将已知功能的天然菜籽油基因 DNA 序列选为使用改造锌指核酸酶进行基因组剪辑的靶。这些基因,称作 5-烯醇式丙酮酰莽草酸-3-磷酸合酶(5-enolpyruvyl shikimate-3-phosphate synthase)(EPSPS) 基因的序列源自欧洲油菜 L, Nex710。该酶 EPSPS 为莽草酸途径的第六个酶,所述途径对在植物中合成芳香族氨基酸与许多芳香族代谢物至关重要(Bentley(1990)Crit. Rev. Biochem. Mol. Biol. 25 :307-384)。其催化磷酸烯醇式丙酮酸(PEP)的烯醇式丙酮酸半基转移至莽草酸-3-磷酸(S3P)的 5-羟基上。因为欧洲油菜为合并芜菁( $2n = 20$ , AA)与甘蓝( $2n = 18$ , CC)染色体组所得的双二倍体(Morinaga, 1934 ;U, 1935),期望在该物种中应具有多于一个基因的 EPSPS。

**[0430] B. DNA 分离**

[0431] 将欧洲油菜变种 Nex710 种子植于温室中。在栽植 13 日后收获试样,将其在液氮中闪冻(flash frozen),并储藏于  $-80^{\circ}\text{C}$  直至使用。

[0432] 基因组 DNA 通过使用或者溴化十六烷基三甲基铵沉淀,或者为分离植物 DNA 的 Plant DNeasy 抽提试剂盒(Qiagen, Valencia, CA)分离。对使用 CTAB 的过程,将 1g 叶组织(六株植物的组)在液氮中磨碎。如 Permingeat et al. (PlantMol. Biol. Repr. (1998) 16 :1-6 ;以引用方式包含于本文中)所描述分离 DNA,但是修饰了抽提缓冲液。所述修饰的抽提缓冲液含有 100mM Tris-HCl(pH8.0), 2M NaCl, 25mM EDTA, 2.5% CTAB(Sigma Catalog#H-5882) 与 1.5% 聚乙烯吡咯烷酮-40(PVP-40)。总基因组 DNA 使用 PLANT DNEASY extractionkit(Qiagen, Valencia, CA), 根据生产商的推荐分离,除了一处修饰。将 PVP-40 添加入 AP1 缓冲液(Qiagen) 至终浓度为 1%。若 DNA 需经限制性酶消化,其将进一步通过两次下述的聚乙二醇(PEG, MW8000) 沉淀步骤提纯。将等体积的 1.2M NaCl/13% PEG 添加到 DNA 中,并在冰上培育 2 小时。随即将试样以  $5000\times g$  旋转 10 分钟,弃去上清,用 70% 乙醇洗涤沉淀。用冷冻干燥完全去除乙醇,并将所述 DNA 沉淀重新悬浮于 EB 缓冲液中(Qiagen)。

[0433] 随即使用 PicoGreen 荧光核酸染色根据贩卖商的指示(Molecular Probes, Eugene, OR) 通过在 260 与 280nm 处的吸光度读数测定 DNA。DNA 的质量通过在 0.8% 琼脂糖胶上使用 Tris-乙酸-EDTA(TAE) 缓冲液将 DNA 试样进行跑胶来检查(Sambrook et al. (1989) Gel electrophoresis of DNA, MolecularCloning. Cold Spring Harbor Laboratory Press, p. 6. 7)。

[0434]

**C. 通过 Southern 分析估计 EPSPS 基因拷贝数**

[0435] 在欧洲油菜基因组 DNA 基因扩增,克隆与测序之前通过 Southern 分析实施对 EPSPS 基因拷贝数的估计。消化基因组 DNA 的限制性酶通过下述方法选择,即其在所述基因中切割 DNA 一次(Gasser and Klee(1990)NucleicAcid Research 18 :2821),并在其两

侧的基因组序列中切割第二次,从而使得对每个 EPSPS 基因在与 EPSPS 基因探针杂交后得到独特大小的基因组 DNA 片段。多数所选的限制性酶 (Pvu II, Nde I, Bsr BI, Bsa I, Bcl I, Bsm I, Afl II) 或者靠近 5' 端或者在基因中间切割,只有 BclI 例外,其在所述基因的 3' 端,即探针杂交处切割 (见下)。

[0436] DNA 试样 (对 Nex710 每个 5  $\mu$  g, 对欧洲油菜每个 4  $\mu$  g, 对甘蓝每个 3  $\mu$  g) 用 30 单位的每种限制性酶 Pvu II, Nde I, Bsr BI, Bsa I, Bcl I, Bsm I 与 Afl II 分别在 eppendorf 试管中根据生产商的指示 (New England BioLabs) 消化过夜。对消化的 DNA 试样随即进行乙醇沉淀,并冷冻干燥所得沉淀。

[0437] 干燥过的沉淀在 2x 上样缓冲液中重新悬浮,装载于 0.85% 琼脂糖胶,并对其在 0.4xTris-乙酸缓冲液在 pH8.0 下进行电泳 (Sambrook et al. (1989) Gelelectrophoresis of DNA, Molecular Cloning, Cold Spring Harbor Laboratory Press, p. 6.7)。随即将所得胶用溴乙锭染色,并用 UV 将 DNA 条带显影。随即将 DNA 在 25mM 焦磷酸钠 buffer (Murray et al. (1992) Plant Mol. Biol. Repr. 10 :173-177) 转移到 GENESCREEN PLUS 杂交转移膜上 (DuPont NEN, Boston, MA, USA)。预杂交在 65°C 下于 SIGMA PERFECT HYB PLUS 杂交缓冲液 (Sigma, St. Louis, MO) 中至少进行 2 小时。杂交在该缓冲液中在添加放射性探针 (见下) 后过夜进行。杂交与预杂交两个步骤均使用杂交烤炉 (RobbinsScientific Corp, Sunnyvale, CA, USA) 进行。在 20 倍稀释的包含 200mM 磷酸钠 pH7.8, 50mM 焦磷酸钠, 10mM EDTA 与 2% SDS (Murray et al., 见上) 的洗涤缓冲液中洗涤所得膜。使用起初 5 分钟的漂洗继以两次每次 15 分钟的洗涤。所得印迹随即在室温下暴露于磷光成像屏 12 小时,然后在 BIORADPERSONAL FX 磷光成像仪 (Bio-Rad, Hercules, CA) 中扫描。

[0438] 所述供 Southern blot 杂交的 EPSPS 探针通过 PCR 使用欧洲油菜变种 Nex710 作为基因组 DNA 模板产生。引物从基于公开的欧洲油菜基因组 DNA 序列 (Gasser and Klee, 见上) 的外显子 8 序列使用 VECTORNTI 软件 (Invitrogen, Carlsbad, CA) 设计,并由 MWG BIOTECH, Inc. (High Point, NC, USA) 定制合成。正向与反向引物的序列分别为 TTGGAGCTACAGTGAAGAAGGTT (SEQ ID NO :1) 与 CGATTGCATCTCACTCAGTTCATTA (SEQ ID NO :2)。PCR 反应包含 5  $\mu$  l 10X HOT START PCR 缓冲液 (Qiagen, Valencia, CA, USA), 2  $\mu$  l 25mM MgCl<sub>2</sub>, 4  $\mu$  l 10mM 核苷酸混合液, 1  $\mu$  l 每种 (20  $\mu$  M), 1.5 单位 HOT STARTTaq DNA 多聚酶 (Qiagen, Valencia, CA), 5  $\mu$  l Nex710 模板 DNA, 以及灭菌水使总体积为 50  $\mu$  l。扩增在 ICYCLER IQ 实时 PCR 仪上 (Bio-Rad, Hercules, CA) 进行,使用下列参数:起初在 95°C 下变性 15 分钟,继以在 95°C 下 35 周期 30 秒,在 55.5°C 与 52.9°C 下退火 30 秒,以及在 72°C 下 30 秒。350 碱基对的 PCR 产物用 QIAQUICK 核苷酸移除试剂盒 (Qiagen, Valencia, CA) 纯化。DNA 大小与完整性通过在 2.0% E-GEL 琼脂糖胶 (Invitrogen, Carlsbad, CA) 上进行电泳来进行验证。使用 PICOGREEN DNA 量化试剂 (Invitrogen, Carlsbad, CA) 来确定片段的数量。使用 READY-TO-GO DNA 标记珠 (-dCTP) (Amersham Biosciences, Piscataway, NJ) 标记 DNA 探针。

[0439] Southern blot 分析显示多个欧洲油菜 EPSPS 特异性条带 (四个或更多),以及可能有同样多的基因存在 (图 1)。芜菁与甘蓝 DNA 与较少条带杂交,且其数量与位置合计不足以得到欧洲油菜的模式,表明在双二倍体化之后,亲本与欧洲油菜基因组中发生了序列分歧。次要条带可能来自与其它具有少量序列同源性的欧洲油菜基因交叉杂交。

**[0440] D. 基因扩增与序列分析**

[0441] 在现行研究中,芜菁 EPSPS cDNA 序列 (GenBank Accession No. AY512663, SEQ ID NO :3) 用于使用 BLAST 算法查询 TIGR 欧洲油菜 EST 数据组 (可在互联网上 [tigrblast.tigr.org/tgi/](http://tigrblast.tigr.org/tgi/) 获取)。识别出了两个序列, TC1307 (部分且未标注) 与 TC1308 (全长 EPSPS)。所述 TC1307 序列为 EPSPS 基因序列。使用 AY512663 与 TC1307 序列涉及多种短寡聚核苷酸用作 PCR 引物,包括下列正向寡聚核苷酸:

[0442] 5' -ATGGCGCAAGCTAGCAGAATCTGCC-3' (SEQ ID NO :4)

[0443] 5' -ATGGCGCAAGCTAGCAGAATC-3' (SEQ ID NO :5)

[0444] 5' -CCAGCAGCAGCGTGGAGCTTATCAGATA-3' (SEQ ID NO :6), 以及下列反向寡聚核苷酸:

[0445] 5' -GGCCCAAAACTGATTCAACGATTGC-3' (SEQ ID NO :7)

[0446] 5' -CGTTGCCACCAGCAGCAGTA-3' (SEQ ID NO :8)

[0447] 5' -GATGGTCCAGTCACAGTCACACTGTTCTCTGT-3' (SEQ ID NO :9)。

[0448] 所有的寡聚核苷酸引物由 Integrated DNA Technologies (IDT, Coralville, IA) 合成并购自。

[0449] 对基于 PCR 的分析, DNA 扩增在含有 2.5  $\mu$ l 的 10X LA PCR bufferII ( $Mg^{2+}$  plus) (Takara Bio Inc., Otsu, Shiga, Japan), 0.7  $\mu$ l 的 25mM  $MgCl_2$ , 4  $\mu$ l 的 10mM nucleotide mix, 0.5  $\mu$ l 的每种引物 (20  $\mu$ M), 1.25 单位 TAKARALA Taq 多聚酶 (Takara Bio Inc.), 1  $\mu$ l 的欧洲油菜变种 Nex710 模板 DNA (3-10ng DNA), 以及灭菌水至 25  $\mu$ l 的总体积的 PCR 反应混合物中进行。扩增或者在 MJthermocycler (Bio-Rad, Hercules, CA) 或者在 GENEAMP PCR system 9700 (Applied Biosystems, Foster City, CA) 中, 使用下列参数进行: 起初在 94°C 下变性 1 分钟, 继以 30 周期在 94°C 下 20 秒, 在 59°C 下 30 秒, 以及在 72°C 下 2 分钟。PCR 产物的大小与完整度通过电泳验证。

[0450] 当用 SR130 与 SR131 引物 (分别为 SEQ ID NOS 4 与 7) 实施 PCR 时, 扩增出了 2.6kb-3kb 大小的 DNA 片段。将这些片段使用来自 Invitrogen (Carlsbad, CA) 的 TA cloning kit 根据生产商的推荐方法直接克隆入载体 pCR2.1 (Invitrogen, Carlsbad, CA)。克隆片段在 DAS 使用 CEQ dye terminator cycle sequencing kit (Beckman Coulter, Fullerton, CA) 根据生产商的推荐进行测序或将测序服务包给 Cogenics 之前为 Lark Technologies, Inc. Houston, TX)。多个克隆的序列分析揭示了三个相异的基因片段。这些基因片段称为 EPSPS 共生同源基因 C, D 与 E (SEQ ID NOS :12-14)。

[0451] 为了识别所述基因其它可能存在于欧洲油菜基因组中的变体, 在如上所述同样的 PCR 条件下以温度梯度运行 PCR。扩增在 ICYCLER IQ 实时 PCR 仪 (Bio-Rad, Hercules, CA) 中使用下述参数进行: 起初在 94°C 下变性 1 分钟, 继以 30 周期在 94°C 下 20 秒, 在 40°C 到 60°C 温度梯度下 30 秒, 以及 72°C 下 4 分钟。进行 72°C 下 30 分钟的最后延伸, 继以 4°C 下无限期暂停。在这些条件下, 一个特定的对应约 2.5kb 扩增 DNA 的条带在 52.5°C 下产生。将扩增的 DNA 克隆入载体 pCR2.1, 并对其如前所述进行测序。多个克隆的序列分析清楚的指明该 PCR 产物代表不同的基因, 其识别为 EPSPS 共生同源基因 B (SEQ ID NO :11)。

[0452] 对应于片段 TC1307 的引物 1307F (SEQ ID NO :5) 与 1307R (SEQ ID NO :8) 用于下列组成的 PCR 反应液: 5  $\mu$ l 的 10X Hot Start PCR buffer (Qiagen, Valencia, CA), 3  $\mu$ l 的

25mM MgCl<sub>2</sub>, 4 μl 的 10mM 核苷酸混合物, 1 μl 的每种引物 (20 μM), 1.5 单位的 HOT START Taq DNA 多聚酶 (Qiagen, Valencia, CA), 5 μl (20ng) 的 Nex710 模板 DNA, 以及灭菌水至总体积 50 μl。扩增在 ICYCLER IQ PCR 仪 (Bio-Rad, Hercules, CA) 中使用下列参数进行: 起初在 95°C 下变性 15 分钟, 继以 35 周期的 95°C 下 30 秒, 在 40°C 到 60°C 的温度梯度下 30 秒, 以及在 72°C 下 1 分钟。在 72°C 下进行 10 分钟的最后延伸继以在 4°C 下无限期暂停。在这些条件下, 在 41.4°C 产生了 700bp 的扩增 DNA 条带。将该片段克隆于 TOPO PCR2.1 载体中并如前所述进行测序。多个克隆序列比对后得到 669bp 的序列, 其识别为 EPSPS 共生同源基因 A (SEQ ID NO:10)。

[0453] 其它 PCR 反应使用引物 4<sup>th</sup>\_Gene\_F2 (SEQ ID NO:6) 与 EPSP\_cDNA\_R9 (SEQ ID NO:9) 进行以扩增共生同源基因 A 的更长序列。PCR 反应混合物为下列组成: 10.0 μl 的 ACCUPRIME SUPERMIX II PCR 扩增试剂 (Invitrogen, Carlsbad, CA), 0.5 μl 的每种引物 (20 μM), 3 μl 的 Nex710 模板 DNA, 以及灭菌水至总体积 20 μl。扩增使用下列参数进行: 95°C 下 3 分钟, 继以 10 周期的 95°C 下 30 秒, 73°C (-0.5°C / 周期) 下 30 秒, 以及 68°C 下 3 分钟, 继以 30 周期 95°C 下 30 秒, 68°C 下 30 秒与 68°C 下 3 分钟。再继以 68°C 下最终延伸 30 分钟。将约 2kb 的扩增片段克隆入 TOPO PCR2.1 载体并随即如前所述进行测序。多个克隆序列比对后得到 1571bp 的共生同源基因 A 序列。

[0454] 芜菁 cDNA (SEQ ID NO:3) 与欧洲油菜基因组 DNA (SEQ ID NO:14) 的比较显示在 EPSPS 基因中存在 8 个外显子与 7 个内含子。基于该比较, 从欧洲油菜变种 Nex710 DNA 分离所得的所有 5 个共生同源基因的序列比对表明, 在所述基因之间在编码区域有小的差异, 例如单核苷酸多态性 (SNPs), 而内含子序列在核苷酸水平上的具有显著更大的差异。总体而言, 在 5 个 EPSPS 共生同源基因间有 84% 或更多的序列同源性 (参见表 2)。

[0455] 表 2. EPSPS 共生同源基因 A-E 间的序列同源性 (%)

[0456] (SEQ ID NOS:10-14)

[0457]

EPSPS 共生同源基因	共生同源 基因 E	共生同源 基因 A	共生同源 基因 B	共生同源 基因 C
共生同源基因 D	98	88	84	97
共生同源基因 E		88	84	95
共生同源基因 A			92	87
共生同源基因 B				84

[0458] 注意到了所述 5 个共生同源基因间的差异, 这是因为其凸显了可由序列依存性 DNA 结合蛋白, 例如锌指蛋白辨别的序列区域。期望设计与尽管非常相似的序列中一个基因序列结合但不与另一个结合的锌指 DNA 结合区域。对所述共生同源基因中的四个, B, C, D 与 E (SEQ ID NOS:11-14), 其近乎全长的基因序列, 以及对 1575kb 的共生同源基因 A, 其部分基因序列受选择作为设计锌指核酸酶的靶, 如下面所述。

[0459] 实施例 2: EPSPS 锌指 DNA 结合区域的设计

[0460] 使用自欧洲油菜 EPSPS 共生同源基因 A-E (实施例 1, 图 9-13) 中识别的靶位点, 为

EPSPS 锌指选择识别螺旋。代表性 EPSPS 锌指涉及的识别螺旋如下示于表 A

[0461] 表 A :EPSPS 锌指设计

[0462]

ZFN 名称	F1	F2	F3	F4	F5	F6
10654 EPSPS	QSGDLTR (SEQ ID NO :91)	RSDTLST (SEQ ID NO :92)	RNDNRIT (SEQ ID NO :93)	QSSDLR (SEQ ID NO :94)	QSSDLTR (SEQ ID NO :95)	无
10658* EPSPS	DRSNLSR (SEQ ID NO :96)	RSDALAR (SEQ ID NO :97)	QNAHRKT (SEQ ID NO :98)	RSDHLSE (SEQ ID NO :99)	NSRNRKT (SEQ ID NO :100)	无
9875 EPSPS	QSSDLR (SEQ ID NO :94)	RSDHLR (SEQ ID NO :101)	QSSDLRR (SEQ ID NO :102)	QSGNLR (SEQ ID NO :103)	QSGNLR (SEQ ID NO :103)	无
10275 EPSPS	RSDVLSQ (SEQ ID NO :104)	RNANRKK (SEQ ID NO :105)	RSDNLST (SEQ ID NO :106)	RNDARIT (SEQ ID NO :107)	RSDNLST (SEQ ID NO :106)	DNSSRIT (SEQ ID NO :108)
10740 EPSPS	RSDVLSE (SEQ ID NO :109)	TSGHLR (SEQ ID NO :110)	RSDDLK (SEQ ID NO :111)	DSSARKK (SEQ ID NO :112)	无	无
10741 EPSPS	TSGNLR (SEQ ID NO :113)	TSGSLR (SEQ ID NO :114)	RSDHLST (SEQ ID NO :115)	QSANRKT (SEQ ID NO :116)	无	无
10742 EPSPS	TSGNLR (SEQ ID NO :113)	TSGSLR (SEQ ID NO :114)	RSDHLSQ (SEQ ID NO :117)	TSSNRIT (SEQ ID NO :118)	无	无
9876 EPSPS	QSSDLRR (SEQ ID NO :102)	RSDHLR (SEQ ID NO :101)	QSSDLRR (SEQ ID NO :102)	DRSALS (SEQ ID NO :119)	QSGNLR (SEQ ID NO :103)	无
9882 EPSPS	QSSDLRR (SEQ ID NO :102)	RSDHLST (SEQ ID NO :115)	HSDTRKK (SEQ ID NO :120)	QSGNLR (SEQ ID NO :103)	QSGNLR (SEQ ID NO :103)	无

ZFN 名称	F1	F2	F3	F4	F5	F6
11038 EPSPS	QSGNLAR (SEQ ID NO :103)	TSGSLTR (SEQ ID NO :114)	RSDHLST (SEQ ID NO :115)	QSANRTK (SEQ ID NO :116)	无	无
11039 EPSPS	QSGNLAR (SEQ ID NO :103)	TSGSLTR (SEQ ID NO :114)	RSDHLSQ (SEQ ID NO :117)	TSSNRIT (SEQ ID NO :118)	无	无
10744 EPSPS	RSDDLSE (SEQ ID NO :121)	TNSNRKR (SEQ ID NO :122)	RSDLSA (SEQ ID NO :123)	TSANLSR (SEQ ID NO :124)	无	无
10743 EPSPS	RREDLIT (SEQ ID NO :125)	TSSNLSR (SEQ ID NO :126)	RSDLSE (SEQ ID NO :127)	QANRKT (SEQ ID NO :128)	无	无
10745 EPSPS	RSDLSE (SEQ ID NO :127)	TSANLSR (SEQ ID NO :124)	RSDLSA (SEQ ID NO :123)	TSANLSR (SEQ ID NO :124)	无	无
9892 EPSPS	RSDNSA (SEQ ID NO :129)	QNRDRKN (SEQ ID NO :130)	QSGDLTR (SEQ ID NO :91)	RSDALAR SEQ ID NO :97)	RSDNLRE (SEQ ID NO :131)	无
9895 EPSPS	RSAALAR (SEQ ID NO :132)	RSDDLTR (SEQ ID NO :133)	QSGDLTR (SEQ ID NO :91)	RSDTLSQ (SEQ ID NO :134)	QSGSLTR (SEQ ID NO :135)	无
9896 EPSPS	RSDALAR (SEQ ID NO :97)	RSDDLTR (SEQ ID NO :133)	QSGDLTR (SEQ ID NO :91)	RSDTLSQ (SEQ ID NO :134)	QSGSLTR (SEQ ID NO :135)	无
10657* EPSPS	DRSNLSR (SEQ ID NO :96)	RSDALAR (SEQ ID NO :97)	QNAHRKT (SEQ ID NO :98)	RSDHLSE (SEQ ID NO :99)	NSRNRKT (SEQ IN NO :100)	无
12385 EPSPS	QSGDLTR SEQ ID NO :91)	RSDLTST (SEQ ID NO :92)	RNDNRIT (SEQ ID NO :93)	QSSDLR (SEQ ID NO :94)	LLTTLKA (SEQ ID NO :136)	无

ZFN 名称	F1	F2	F3	F4	F5	F6
12202* EPSPS	DRSNLSR (SEQ ID NO :96)	RSDALAR (SEQ ID NO :97)	QNAHRKT (SEQ ID NO :98)	RSDHLSE (SEQ ID NO :99)	KNFNLHQ (SEQ ID NO :137)	无
14318* EPSPS	DRSNLSR (SEQ ID NO :96)	RSDALAR (SEQ ID NO :97)	QNAHRKT (SEQ ID NO :98)	RSDHLSE (SEQ ID NO :99)	KNFNLHQ (SEQ ID NO :137)	无
14320* EPSPS	DRSNLSR (SEQ ID NO :96)	RSDALAR (SEQ ID NO :97)	QNAHRKT (SEQ ID NO :98)	RSDHLSE (SEQ ID NO :99)	KNFNLHQ (SEQ ID NO :137)	无
13969 EPSPS	DRSNLSR (SEQ ID NO :96)	RSDALAR (SEQ ID NO :97)	TSTGLLI (SEQ ID NO :138)	RSDHLSE (SEQ ID NO :99)	KNFNLHQ (SEQ ID NO :137)	无
12540 EPSPS	DRSNLSR (SEQ ID NO :96)	RSDALAR (SEQ ID NO :97)	VSHTRLD (SEQ ID NO :139)	RSDHLSE (SEQ ID NO :99)	NSRNRKT (SEQ ID NO :100)	无
12352 EPSPS	QSGDLTR (SEQ ID NO :91)	RSDLTST (SEQ ID NO :92)	TRYKLMS (SEQ ID NO :140)	QSSDLR (SEQ ID NO :94)	QSSDLTR (SEQ ID NO :95)	无
11034 EPSPS	RSDVLSE (SEQ ID NO :109)	TSGHLSR (SEQ ID NO :110)	RSDDLK (SEQ ID NO :111)	DSSARKK (SEQ ID NO :112)	无	无
11036 EPSPS	TSGNLTR (SEQ ID NO :113)	TSGSLTR (SEQ ID NO :114)	RSDHLST (SEQ ID NO :115)	QSANRTK (SEQ ID NO :116)	无	无
11037 EPSPS	TSGNLTR (SEQ ID NO :113)	TSGSLTR (SEQ ID NO :114)	RSDHLSQ (SEQ ID NO :117)	TSSNRIT (SEQ ID NO :118)	无	无

[0463]

[0464]

[0465] (\*备注 -10657 与 10658, 12202, 14318 与 14320, 10740 与 11034, 10741 与 11036,

以及 10742 与 11037 ZFN 彼此的区别在于并非位于识别螺旋上的突变)

[0466] 所述锌指设计的靶位点如下所示于表 B。ZFP10654 与 10658 为共生同源基因 C 与 D 上的靶位点设计 ;ZFP9875 与 10275 为共生同源基因 D 上的靶位点设计 ;而 ZFP10740, 10741 与 10742 为结合于共生同源基因 A 与 B 上的靶位点而设计。

[0467] 表 B :EPSPS 锌指的靶位点

[0468]

ZFN 名称	靶位点 (5' to 3' )
10654 EPSPS	ttACTGCTgCAGGTGGCAac (SEQ ID NO :141)
10658 EPSPS	ggCAGCGGTGAGTGGACgc (SEQ ID NO :142)
9875 EPSPS	ttGAAGAAGCTGGGGCTta (SEQ ID NO :143)
10275 EPSPS	gcATCAAGcATGTAGTTGATGtc (SEQ ID NO :144)
10740 EPSPS	aaATCTCGGGTCTGat (SEQ ID NO :145)
10741 EPSPS	tcTAATGGGTTGAAgc (SEQ ID NO :146)
10742 EPSPS	tcTAATGGGTTGAAgc (SEQ ID NO :146)
9876 EPSPS	ttGAAGAAGCTGGGGCTta (SEQ ID NO :143)
9882 EPSPS	ttGAAGAAGCTGGGGCTta (SEQ ID NO :143)
11038 EPSPS	tcTAATGGGTTGAAgc (SEQ ID NO :146)
11039 EPSPS	tcTAATGGGTTGAAgca (SEQ ID NO :146)
10744 EPSPS	gaGATTTGGATCCGgg (SEQ ID NO :147)
10743 EPSPS	tcCAACCGGATTCTtc (SEQ ID NO :148)
10745 EPSPS	gaGATTTGGATCCGgg (SEQ ID NO :147)
9892 EPSPS	tgCAGGTGGCAaCGCAAgat (SEQ ID NO :149)
9895 EPSPS	caGTAACGGCAGCGGTGag (SEQ ID NO :150)
9896 EPSPS	caGTAACGGCAGCGGTGag (SEQ ID NO :150)



ZFN 名称	靶位点 (5' to 3')
10657 EPSPS	ggCAGCGGTGAGTGGACgc (SEQ ID NO :142)
12385 EPSPS	ttACTGCTgCAGGTGGCAac (SEQ ID NO :141)
12202 EPSPS	ggCAGCGGTGAGTGGACgc (SEQ ID NO :142)
14318 EPSPS	ggCAGCGGTGAGTGGACgc (SEQ ID NO :142)
14320 EPSPS	ggCAGCGGTGAGTGGACgc (SEQ ID NO :142)
13969 EPSPS	GCAGCGGTGAGTGGACG (SEQ ID NO :151)
12540 EPSPS	tgCAGCTGTAAGTGGACgc (SEQ ID NO :152)
12352 EPSPS	ttACTGCTgCTGGTGGCAac (SEQ ID NO :153)
11034 EPSPS	aaATCTCGGGTCTGat (SED ID NO :145)
11036 EPSPS	tcTAATGGGTTGAAgc (SEQ ID NO :146)
11037 EPSPS	tcTAATGGGTTGAAgc (SEQ ID NO :146)

[0469]

[0470] 将所述 EPSPS 设计并入下述锌指表达载体, 即其编码具有至少一个有 CCHC 结构的指的蛋白质。参见, U. S. Patent Application Serial No. 60/874, 911。尤其是, 每个蛋白的最后指具有 CCHC 骨架。随即将锌指编码序列融合于 IIS 型限制性酶 FokI 的核酸酶结构域 (通过四氨基酸的 ZC 接头与 Wah et al. (1998) Proc. Natl. Acad. Sci. USA 95 :10564-10569 序列的 384-579 位氨基酸) 以形成 EPSPS ZFN。对多种 ZFN 分析其生物学活性和 / 或毒性, 如 U. S. Patent Application No. 60/995, 566 所述。

[0471] 实施例 3 : 在 HEK293 细胞中 EPSPS 特异性 ZFN 的功能性验证

[0472] 本文所述的 EPSPS ZFN 促进同源重组的能力在描述于 Urnov (2005) Nature 435 (7042) :646-51 与 U. S. Patent Publication No. 20050064474 (例如, 实施例 6-11) 中的 GFP 系统中得到了测试。简言之, 携带所述 EPSP 基因目标区域的 HEK293 报道细胞系如下所述产生, 通过 PCR 扩增所述 EPSP 基因的目标区域, 并随后将其克隆入 pcDNA4TO-GIF。使用上述质粒转染 HEK293, 并随后在转染后 48 小时在 400  $\mu$ g/ml 博莱霉素 (Zeocin) 存在下进行选择。

[0473] 所得稳定克隆的池随即使用下述 ZFN 进行测试, 即所述 ZFN 导向在上述所产生的报道细胞系中 EPSP 基因的特定目标区域, 测试方法如下所述。将报道细胞系以 350000 细胞 / 孔接种于 12 孔板中的 1mL DMEM, 10% FBS 培养基 (无 PSG), 并用 50 或 100ng 每种 ZFN 进

行转染,并且将 500ng 无启动子的 GFP 供体 (Urnov (2005) Nature) 转染入 500000 报道细胞中,使用每试样 2uL Lipofectamine 2000 (Invitrogen),根据 Invitrogen Lipofectamine 2000 实验方案进行。每对 ZFN 的转染重复三次。在转染后翌日,对每个孔添加 1mL DMEM 培养基,其中含有 1.5uL 长春花碱使终浓度为 0.2  $\mu$ M 到 1mM,并在转染 72 小时后将其移除。在转染 5 天后通过在 Guava 桌上 FACS 分析仪上通过每次转染测量 40000 细胞来分析所得细胞的 GFP 表达。示例性的结果示于图 14,小图 A 到 E。

[0474] 实施例 4:在欧洲油菜变种 Nex710 中一个 ZFN 可剪切两个或更多 EPSPS 共生同源基因

[0475] 为了评价设计的锌指核酸酶在植物细胞中的功能性,利用了在生活植物细胞中表达上述蛋白质的方法。可在 DNA 没有掺入植物细胞基因组的条件下将编码锌指核酸酶蛋白质的 DNA 递送入植物细胞。从而,所述 DNA 暂时维持在植物细胞中,并充当基因表达的模板。或者,可在允许 DNA 掺入植物细胞基因组的条件下将编码锌指核酸酶蛋白质的 DNA 递送入植物细胞,导致锌指核酸酶编码基因的转基因,从而使得所得 DNA 分子稳定维持在植物细胞中,并充当基因表达的模板。本领域技术人员可利用锌指核酸酶编码 DNA 的暂时或转基因表达之一以评价这些蛋白在生活植物细胞中的功能性。

[0476] A. 载体设计

[0477] 在欧洲油菜细胞中构建了质粒载体以供表达 ZFN 蛋白。为了优化表达与两个需用以形成功能性的锌指核酸酶异源二聚体的不同蛋白质的相对化学量,采用了下述的表达策略,即其导致两个 ZFN 单体的开放阅读框均插入一个单独的载体,由一个单独的启动子驱动。该策略利用了源自 *Thesoa assigna* 病毒的 2A 序列 (Mattion, et al. (1996) J. Virol. 70,8124-8127), 或者为 SMV 病毒核定位信号 (NLS) (PKKKRKV (SEQ ID NO: 15); Kalderon et al. (1984a) Nature 311:33-38; Kalderon et al. (1984b) Cell 39:499-509), 或者为来自 opaque-2 基因 (op-2; Maddaloni et al. (1989) Nucleic Acids Research 17:7532; Van Eenennaam et al. (2004) Metabolic Engineering 6:101-108) 的玉米 NLS, 的功能性,以及源自木薯叶脉花叶病毒启动子或 CsVMV 的启动子 (参见表 3)。

[0478] 表 3. 存在于用于欧洲油菜转化的多种构建体中的 ZFN 对与表达元件的描述

[0479]

S. N.	ZFN 对	构建体号	构建体类型	基因卡盒
1	10654-v2/10657-v2	pDAB7147	二元的	CsVMV/ZFN 对 /AtuORF23// AtUbi10/Pat/AtuORF1
2	10654-v3/10658-v3	pDAB7150	二元的	CsVMV/ZFN 对 /AtuORF23// AtUbi10/Pat/AtuORF1
3	10654-v2/10657-v2	pDAB7151	非二元的	CsVMV/ZFN 对 /AtuORF23
4	10654-v3/10658-v3	pDAB7154	非二元的	CsVMV/ZFN 对 /AtuORF23

S. N.	ZFN 对	构建体号	构建体类型	基因卡盒
5	10740-v2/10741-v2	pDAB7185	二元的	RB7 MAR//CsVMV//ZFN 对 //AtuORF23// 中断的 ipt(onc 4) 基因 Orf//4OCS delta mas 2' //PAT//AtuORF1
6	10740-v2/10742-v2	pDAB7186	二元的	RB7 MAR//CsVMV//ZFN 对 //AtuORF23// 中断的 ipt(onc 4) 基因 Orf//4OCS delta mas 2' //PAT//AtuORF1

[0480] CsVMV = 木薯叶脉花叶病病毒启动子与 517bp 的先导序列 (Verdaguer et al. (1998) Plant Mol. Biol. 37 :1055-1067) ;AtuORF23 = 根癌土壤杆菌 ORF23 3' UTR ; AtUbi10 = 拟南芥泛肽基因 10 启动子 ;Pat = 来自绿色产色链霉菌 (*Streptomyces viridochromogenes*) 草胺膦乙酰转移酶基因。其为报道于 USpatent 5633434 中的重建基因 ;AtuORF1 = *Agrobacterium tumefaciens* ORF1 3' UTR (Genebank accession number X00493, NC\_002377) ;RB7MAR = 烟草基质附着区 ;4OCS delta mas 2' = 修饰的甘露碱合酶启动子, 其包含 4X OCS 元件以增强表达 ;中断的 ipt(onc 4) 基因 Orf = 根癌土壤杆菌破坏的 ipt 基因 (使用 genebank 序列 ID ATMRPTI 以供设计)。

[0481] 设计了分步的模块式克隆方案以开发这些针对任何给定对的 ZFN 编码基因的表达载体, 所述 ZFN 编码基因为选自文库档案或从头合成。首先, pVAX 载体 (参见, 例如 U. S. Patent Publication 2005-0267061, 所述公开以引用方式包含于此) 经修饰以涵盖示于图 2A-2E 的 N- 末端表达域。该修饰质粒 (pVAX-N2A-NLSop2-EGFP-FokMono) (Figure 2A) 的特征包括重新设计与合成的 NLS 编码区段, 与重新设计的与合成的编码 FokI 核酸酶结构域, 使用双子叶植物偏好的密码子。另外, 在独特的 XhoI 位点下游的一个单独核苷酸插入 (C) 形成了额外的 SacI 位点以便克隆。

[0482] 其次, pVAX 载体 (参见, 例如 U. S. Patent Publication 2005-0267061) 亦经修饰以涵盖 C- 末端表达域。该修饰质粒 (pVAX-C2A-NLSop2-EGFP-FokMono) (Figure 2B) 的特征包括重新设计的与合成的 NLS 编码区段, 与重新设计的与合成的编码 FokI 核酸酶结构域, 使用双子叶植物偏好的密码子。另外, 来自 *Thosea asigna* 病毒的 2A 序列 (EGRGSLTTCGDVEENPGP, SEQ ID NO :16) 导入 ZFN ORF 的 N- 末端以达随后连接所述两个蛋白质编码域的目的。

[0483] 使用限制性酶 KpnI 与 BamHI 产生相容末端通过连接将编码个别锌指蛋白 ORF 的基因盒克隆入 N2A 或 C2A 载体之一。其次, 将来自 C2A 载体的 BglIII/XhoI 片段使用相同限制性位点插入 N2A 载体, 得到含有下述盒的中间构建体, 即所述盒包括夹在 NcoI 与 SacI 限制性位点之间的两个 ZFN 编码域 (图 2C)。

[0484] 最终, 来自该中间构建体的 NcoI/SacI 盒 (图 2C), 包含两个 ZFN 基因, 使用这些限制性酶切出, 并连接入质粒骨架 pDAB3731。所得的质粒, 例如 pDAB7151 (图 2D), 包含所述 ZFN 基因加上相关的启动子与终止子序列, 加上选择性标记以供质粒维持 (表 2)。所述

序列通过限制性酶消化与测序来确证。在该构建体中,所述 ZFN 表达盒(包括启动子与终止子元件)两侧为 attL 位点以供使用来自 Invitrogen(Carlsbad,CA) 的 GATEWAY system 进行方便的操作。将每个使用该克隆方案产生的 ZFN 构建体转化入大肠杆菌 DH5  $\alpha$  细胞(Invitrogen, Carlsbad, CA),并随后在合适的选择下维持。

[0485] 对农杆菌介导植物转化,将所述 ZFN 盒使用 GATEWAY LR CLONASE 反应(Invitrogen, cat#11791-019) 克隆入二元构建体。所得的二元构建体(图 2E)通过限制性酶消化确证,并随即转化入根癌土壤杆菌菌株 Z707s。包含该克隆的菌落通过限制性酶消化与测序反应确证。

#### [0486] B. 暂时与稳定表达系统

[0487] ZFN 表达构建体的质粒制备,例如 pDAB7151,如图 2 所描述,自 2L 生长于含有抗生素的培养的大肠杆菌使用无核酸内切酶的 GIGAPREP kit(Qiagen, Valencia, CA) 根据生产商的推荐方法产生。使用多种方法将质粒 DNA 直接递送于欧洲油菜的下胚轴细胞。

[0488] 在暂时 ZFN 递送的一个实施例中,对菜籽下胚轴节片(部分)通过触须-介导的下胚轴节片暂时转化进行 DNA 递送。欧洲油菜变种 Nex710 的种子经用 10% (v/v) CLOROX(5.25%次氯酸钠)表面灭菌 10 分钟并用灭菌蒸馏水漂洗三次。随后,在含有 1/2 浓度 MS 培养基(1/2MS 基础盐分与维生素,1%蔗糖,0.8%琼脂糖,pH5.8)的 Phytatray 上,以每 Phytatray 25 种子使种子发芽。将所述种子置于培养室以使其在 23°C 下发芽 5 天,其光周期为 16 小时光照,8 小时黑暗。于日 5,将 3mm 长的下胚轴节片在无菌情况下切出,并置于无菌水中以阻止当在切其它节片时其变干。弃去茎与根节片。将所得节片水平置于一张灭菌滤纸之上,静止在 MSK1D1 培养基(MS 基础盐分与维生素,1mg/L 激动素,1mg/L 2,4-二氯苯氧基乙酸 [2,4-D],30g/L 蔗糖,7g/L TC 琼脂糖,pH5.8)表面上。将节片在 23°C 与 16 小时光照下培养 3 天。

[0489] 在触须处理日,将 300 根部分愈伤组织的下胚轴节片与 30ml “高渗培养基”(MS 盐与 B5 维生素,4.42mg/L 2,4-D 以及 12%蔗糖)一起置于 Sorvall 瓶中在室温下进行一小时预处理。该预处理是使组织部分胞质分离的手段以试着在随后的触须处理中当破坏细胞壁时缓解对细胞的损害。随后,将 8.1ml 5% Silar S0-9 碳化硅触须(Advanced Composite Materials, LLC Greer, SC) 溶液与 170  $\mu$ g 如上述制备的非二元 ZFN 质粒 DNA(表 2)添加至 Sorvall 瓶。所述瓶随即在颜料混合器(Red Devil Equipment Co., Minneapolis, MN) 上剧烈振荡 30 秒,其中所述颜料罐钳组件经改造翻新以握持 Sorvall 瓶。在振荡后,将 100ml “高渗培养基”立刻添加到瓶中,并允许其在室温下恢复 20 分钟。随即通过将所述瓶的内容物倾倒通过合适大小的灭菌线网的方法回收以将节片自触须与瓶的液体内容物分离。最后,将节片置回到有滤纸的新鲜的 MSK1D1 培养基板上。在触须处理后日 1,2,3 与 7 取约 100mg 的试样以供暂时表达分析。

[0490] 在另一个实施例中,所述暂时递送系统使用聚乙二醇(PEG)介导转化下胚轴原生质体。原生质体使用由 Sun et al. (Can. J. Bot. (1998) 76 :530-541) 描述的方法经修饰从欧洲油菜变种 Nex710 苗的下胚轴组织制备。种子经表面灭菌并在 MS 菜籽培养基上如上所述发芽 7 天。为每次处理收集一克下胚轴。将下胚轴切成  $\leq$  1mm 大小的薄片,并置于含有 MS9m 培养基(9%甘露醇,5mM MES,10mM 精氨酸,0.3%聚乙烯吡咯烷酮-40(PVP-40)) 的 100mm 培养板中。在将所有下胚轴节片都置于培养皿中后,用吸量管移除液体培养基,并

用 6ml 酶溶液 (包含 0.1% Macerozyme-R10 (Yakult Honsha Co. Ltd, Tokyo, Japan), 1% Cellulase-R10 (Yakult Honsha Co. Ltd, Tokyo, Japan), 1% Pectinase (Sigma Chemical Co.) 的 MS9m) 取代。

[0491] 将所得组织在 25°C 下置于暗处在旋转振荡器上以 40rpm 轻微振荡 16 小时以消化细胞壁。在培育后, 在无菌条件下将酶-原生质体溶液通过置于 50ml 可丢弃离心管之上的 100  $\mu$ m 细胞滤过器 (Sigma Chemical Co.) 过滤。将溶液在 50xg 下离心 5 分钟。在弃去上清后, 将原生质体沉淀重新悬浮于 4ml MS9m 培养基中。在 15ml 离心管中将原生质体悬浮液轻柔的层叠于含 4.5ml 含 0.5M 蔗糖溶液的 MS 之上并在 50xg 下离心。用微量吸管将原生质体从位于相之间的厚条带中取出, 并通过在 50xg 下离心 5 分钟用 5ml MS9m 培养液洗涤。

[0492] 为进行 DNA 处理, 将所述原生质体沉淀重新悬浮于 200  $\mu$ l 甘露醇镁溶液中以达到最终浓度为  $1 \times 10^5$  原生质体/ml。将 50  $\mu$ l 非二元质粒 DNA, 例如 pDAB51 (表 2) 的试样添加到置于 15ml 可丢弃灭菌离心管中的 200  $\mu$ l 原生质体溶液中并混匀。将等体积的 40% PEG-3350 溶液 (Sigma Chemical Co) 添加到所述原生质体溶液中, 并在室温下培育 20 分钟。随后, 添加 0.8ml W5 培养基 (125mM  $\text{CaCl}_2 \cdot \text{H}_2\text{O}$ , 154mM NaCl, 5mM KCl 与 5mM 葡萄糖), 并另外培育 10 分钟继以在 180xg 下离心 3 分钟。用吸量管移除 PEG 溶液, 并将原生质体重新悬浮于 1ml WI 溶液。随即将试管在暗处培育约 18 小时, 继以在 180xg 下离心 3 分钟。移除上清, 并将 100  $\mu$ l 原生质体悬浮液转移至 2ml Eppendorf 管中。在 DNA 处理后 0, 1, 2, 3 日收集原生质体 ( $10^5$ /ml), 并储藏于 -80°C 直至分析。将 10  $\mu$ l 原生质体悬浮液试样与二乙酸荧光素染料混合, 并在血细胞计数器中计量存活的原生质体。

[0493] 在另一个实施例中, 该暂时递送系统使用农杆菌介导转化下胚轴节片。将节片在灭菌滤纸上于愈伤组织诱导培养基 MSK1D1 上培养, 并如上所述触须实验方案给予 3 日预处理。在农杆菌处理前日, 将二元质粒, 例如 pDAB7147 (表 2) 的细菌培养 (1 环) 接种到含有包括合适抗生素的 35ml YEP 培养基的烧瓶中。允许所述细菌培养在暗处以 200rpm 恒定振荡在 28°C 下生长过夜 ~ 16 小时。翌日, 将农杆菌溶液在液体 M 培养基中制备到浓度为 Klett 50。将下胚轴节片从滤纸转移到 100x25mm 含 40ml 农杆菌悬浮液的培养皿并在室温下每 10 分钟周期性回荡培育 30 分钟。在处理期间最后, 移除农杆菌溶液并将下胚轴节片转移回原来含有 MSK1D1 培养基与滤纸的板中。在 Percival 中, 或在培养室中通过将板覆以铝箔减少光强度将节片共培养 3 日。

[0494] 在三日共培养后, 将节片转移至愈伤组织诱导培养基 MSK1D1TC (MS, 1mg/l 激动素, 1mg/l 2,4-D, 0.5mg/l MES, 5mg/l  $\text{AgNO}_3$ , 300mg/l 特泯菌 (Timentin), 200mg/l 羧苄青霉素 (carbenicillin), 3% 蔗糖, 0.7% 植物琼脂 (Phytagar))。在农杆菌处理后 0, 2, 3, 4 与 7 日取约 100mg 下胚轴组织并在 -80°C 下储藏直至分析。

[0495] 在一个使用稳定转基因表达系统的 ZFN 递送的实施例中, 将欧洲油菜变种 Nex710 的种子表面灭菌, 发芽 5 日, 制备  $\leq 1$ mm 的下胚轴节片, 并如触须处理中所述预处理 3 日。在 3 日后, 用任何二元农杆菌菌株 (表 2) 处理下胚轴节片, 并如上述农杆菌介导暂时表达系统所描述的共培养 3 日。

[0496] 在共培养 3 日后, 将 300 个下胚轴节片转移到愈伤组织诱导培养基 MSK1D1H1 (MS, 1mg/l 激动素, 1mg/l 2,4-D, 0.5gm/l MES, 5mg/l  $\text{AgNO}_3$ , 300mg/l 特泯菌, 200mg/l 羧苄青霉素, 1mg/l 好必思, 3% 蔗糖, 0.7% 植物琼脂) 中用低水平除草剂选择 7 日。随即将下胚轴节

片转移至含有高水平选择的 MSK1D1H3 (MS, 1mg/1 激动素, 1mg/1 2,4-D, 0.5gm/1 MES, 5mg/1 AgNO<sub>3</sub>, 300mg/1 特泯菌, 200mg/1 羧苄青霉素, 3mg/1 好必思, 3%蔗糖, 0.7%植物琼脂) 中 2 周, 并随即转移至 MSK1D1H5 培养基 (MS, 1mg/1 激动素, 1mg/12,4-D, 0.5gm/1 MES, 5mg/1 AgNO<sub>3</sub>, 300mg/1 特泯菌, 200mg/1 羧苄青霉素, 5mg/1 好必思, 3%蔗糖, 0.7%植物琼脂) 再两周。对每个二元构建体得到总共 203 与 227 个愈伤组织系 (表 2), 其愈伤组织频率分别为 67.6% 与 75.5%。在农杆菌处理 5-7 周后随即对五十个随机愈伤组织系进行 DNA 分析。

#### [0497] C: 靶向 ZFN 介导双链剪切的 EPSPS 共生同源基因分析

[0498] 在本实施例中 ZFN 的功能性应理解为包括 (但不仅限于) ZFN 在作物物种细胞中表达, 以及该 ZFN 在该作物的内源基因组中通过识别, 结合与剪切其所期望的靶以介导双链 (ds) 断裂的能力。亦应理解, 在本实施例中, 所述 ZFN 的靶为在内源基因座的基因, 以及其在所述作物基因组内的构象。为了评价改造 ZFN 是否具有针对所期望的靶基因在基因组环境中的功能性, 采用了基于 DNA 序列的分析。预计在 DNA 中 ZFN 诱导 ds 断裂诱导修复机理, 例如非同源末端连接 (NHEJ) (由 Cahill et al. (2006) Front Biosci. 1:1958-1976 综述)。NHEJ 的一个结果为部分断裂 DNA 链会以非完美的形式修复, 导致在剪切位点的少量缺失, 插入或取代。本领域技术人员可通过多种方法检测 DNA 序列中的这些变化。

[0499] 为在 EPSPS 共生同源基因中识别 NHEJ, 使用基于 PCR 的方法开发了基因特异性分析。在五个 EPSPS 共生同源基因中四个, 即 A, B, C 与 D 中的序列差异足以允许开发共生同源基因特异性的分析。在靶基因座, 共生同源基因 D 与 E 的序列无法足够分辨开来, 导致仅开发了一种分析来代表这两个共生同源基因。PCR 扩增使用对靶基因具有特异性, 并侧接所述 ZFN 的预期剪切位点的寡聚核苷酸引物进行。共生同源基因特异性 PCR 引物如下所述:

[0500] 共生同源基因 A:

[0501] 正向引物: 5' -TCCCAGCTTCTTTAGATTCTAAGG-3' (SEQ ID NO:17)

[0502] 反向引物: 5' -CTGCAACTTTTCACATAGCAA-3' (SEQ ID NO:18)

[0503] 共生同源基因 B:

[0504] 正向引物:

[0505] 5' -CAAGAGTGATATCGAGTTGTACCTTGGGAATGCT-3' (SEQ IDNO:19)

[0506] 反向引物: 5' -AGGCCATCATATCGAGCAAACGCAGT-3' (SEQ ID NO:20)

[0507] 共生同源基因 C:

[0508] 正向引物: 5' -GGGTAAACAACCGTGCTGTA-3' (SEQ ID NO:21)

[0509] 反向引物: 5' -AAAGACTGCTGCAAACAAGATC-3' (SEQ ID NO:22)

[0510] 共生同源基因 D/E:

[0511] 正向引物: 5' -GGTTGTTGAAGGATGCGGT-3' (SEQ ID NO:23)

[0512] 反向引物: 5' -GCAAACAATTCATAGAGTAAATGTG-3' (SEQ ID NO:24)

[0513] 所有的正向与反向 PCR 引物均对给定直向同源基因组合使用以扩增或者纯化的基因组 DNA 或者含有每个共生同源基因的阳性对照质粒 DNA, 反应条件如下: 25  $\mu$  l 反应体积包含 2.5  $\mu$  l DNA 模板 (10ng/ $\mu$  l) 或质粒 DNA 阳性对照 (1ng/ $\mu$  l), 0.625  $\mu$  l 每种引物 (每种 10  $\mu$  M), 2.5  $\mu$  l 的 10x ACCUPRIME PCR bufferII, 以及 0.15  $\mu$  l (0.75 单位) 在酶生产商提供的缓冲液中的 ACCUPRIME TaqDNA 多聚酶 (Invitrogen, Carlsbad, CA)。扩增在 ICYCLER IQ (Bio-Rad, Hercules, CA) 使用下述参数进行: 94°C 下 2 分钟, 35 周期 (94°C 下 30

秒,退火(参见下述梯度)30秒,68°C下1分钟),68°C下10分钟,4°C下无限期暂停。

[0514] 运行梯度以确定最佳反应条件。梯度温度为65.0°C与50.0°C之间。共生同源基因A,B,C与D/E分别在62.1°C,65.0°C,65.0°C与59.3°C的退火温度下显示最佳扩增(图3A-3D),且在后续研究中使用这些温度。将所有4个共生同源基因的PCR产物克隆入TOPO pCR2.1载体(Invitrogen, Carlsbad, CA),并对其通过对共生同源基因特异性扩增产物的测序加以确证。

[0515] 比较了上述将ZFN递送到下胚轴细胞的三种暂时方法与一种稳定方法以鉴定评价ZFN效力(由存在最多数量的NHEJ确定)的最有效方法。在该研究中使用了在肾293细胞中证明为有效的共生同源基因D特异性ZFN。预测这些ZFN蛋白质将与两个共生同源基因D的EPSPS基因特异性短序列结合构成剪切双链DNA的异源二聚核酸酶(图4)。这些ZFN基因存在于四个构建体中;两个二元构建体,pDAB7147与pDAB7150,其对农杆菌介导转化具有特异性,而剩下两个构建体,pDAB7151与pDAB7154则供暂时转化(表2)。稳定转化的愈伤组织进一步归类为“绿色”与“棕色”试样,其中在所述“棕色”组织中具有较高ZFN表达可能性,从而具有高频率的NHEJ的机会。后者可能导致细胞毒性而使组织变“棕”。收集了所有的试样(参见4B部分),包括未处理的对照,将其冻结并冷冻干燥,但原生质体试样除外,其直接用于DNA分离。基因组DNA通过上述Qiagen方法分离。3μg的所有基因组DNA用5单位MaeIII限制性内切酶(Roche Applied Science, Indianapolis, IN)依照生产商的推荐消化过夜。

[0516] 所述DNA随即通过下述乙醇沉淀方法提纯,即添加0.1体积3M乙酸钠,pH5.2与2体积100%乙醇继以在微离心管中以10000g离心5分钟。所述DNA随即用70%乙醇洗涤,且在SPEEDVAC蒸发器(Savant)中将沉淀干燥并将其重悬浮于水中。随即对所述DNA进行第二次MaeIII消化过夜,并如前所指示用乙醇沉淀。所述限制性酶位点位于两个成对的单体ZFN结合位置间(图4),且靠近FokI域二聚化处,并在基因组DNA中诱导双链断裂。因此,所述限制性酶消化富集发生了NHEJ从而导致其丧失限制性酶识别位点的片段。

[0517] 随即使用对共生同源基因D具有特异性且包夹所述ZFN预测剪切位点的寡聚核苷酸引物进行了PCR扩增。将对作为靶的EPSPS共生同源基因D具有特异性的正向PCR引物(5'-GGTTGTTGAAGGATGCGGT-3')(SEQ ID NO:23)与反向PCR引物(5'-GCAAACAATTCATAGAGTAAATGTG-3')(SEQ ID NO:24)组合使用以在下述条件下扩增纯化的基因组DNA:55μl反应体积包含10μl经MaeIII消化的gDNA(26.4ng)模板,1.25μl每种引物(每种10μM),5μl的10x ACCUPRIME PCR buffer II,5μl的10%PVP-40与1μl(5单位)ACCUPRIME TaqDNA多聚酶(Invitrogen, Carlsbad, CA)。自下述扩增循环得到了预期大小的扩增产物:94°C下2分钟,40周期(94°C下30秒,59.3°C下30秒,68°C下1分钟),68°C下10分钟,并在4°C下无限期暂停。使用来自Invitrogen(Carlsbad, CA)的TA克隆试剂盒将所得扩增片段直接克隆入载体pCR2.1。

[0518] 在每个时点每个处理下对大约90个单独克隆片段使用存在于pCR2.1上的M13正向与M13反向启动位点进行测序。在给定处理中包括未处理的对照。大约3000个克隆通过此法进行了测序。

[0519] 对包括两个相异ZFN处理所得的所有测序结果的分析揭示了13个克隆(经正向与反向测序引物两者确证)正好在pDAB7151中存在的ZFN预期剪切位点处含有少量缺失,

表明所述 NHEJ 机理在该位点介导对所述 DNA 序列的非完美修复 (图 5)。这些特定克隆得自 ZFN 转化 3 日后的原生质体 DNA 试样。这些结果说明了所述改造 ZFN 以特定方式在作物物种中内源基因基因座处诱导靶向双链断裂的能力。在任何其它 ZFN 处理方法中使用该类型的测序未发现 NHEJ。

[0520] D. 大规模并行测序分析 (Massively Parallel Sequencing Analysis)

[0521] 在另一个实施例中, PCR 与大规模并行焦磷酸测序 (massively parallel pyrosequencing) 的组合用于在上述所得试样中调查共生同源基因 D。相同组的正向与反向共生同源基因 D 特异性引物 (SEQ ID NO:23 与 SEQ ID NO:24) 用于扩增代表将 ZFN 转化入菜籽下胚轴细胞的 3 种暂时与一种稳定方法的所有试样 DNA。扩增条件如上所述。

[0522] 该初步扩增产物随即使用 MINELUTE PCR purification kit (Qiagen, Valencia, CA) 纯化, 并洗脱出 10  $\mu$ l DNA。涉及第二组嵌套引物以扩增适于大规模并行测序的大约 100bp 的片段。合成了所述正向引物 PCR 的六个变体 (5' -XXX AGTTGTACCTTGGGAATG-3') (SEQ ID NO:25) 其中 XXX = GGC, CGC, GGC, CGG, CCG 或 GCG, 以及所述反向 PCR 引物的六个变体 (5' -XXXATCAATTTCTTGACAATAACA-3') (SEQ ID NO:26) 其中 XXX = GGC, CGC, GGC, CGG, CCG 或 GCG, 并将其用 HPLC 纯化 (IDT, Coralville, IA)。每个引物 5' 末端的 3bp 标签用作识别密钥, 并表明扩增子源自哪个细胞试样。具有匹配鉴定标签 (密钥) (key) 的引物对组合用于在下述条件下扩增源自上述试样的纯化初步 PCR 扩增子: 50  $\mu$ l 反应体积包含 10  $\mu$ l 稀释 1:10 的纯化 PCR 扩增子, 1.25  $\mu$ l 每种引物 (每种 10  $\mu$ M), 5  $\mu$ l 10x ACCUPRIME PCR 缓冲液 I 与 0.31  $\mu$ l (1.5 单位) 在酶生产商提供的缓冲液中的 ACCUPRIME TaqDNA 高保真度多聚酶 (Invitrogen, Carlsbad, CA)。预期大小的扩增产物得自下述扩增循环: 94 $^{\circ}$ C 下 2 分钟, 30 周期 (94 $^{\circ}$ C 下 30 秒, 62 $^{\circ}$ C 下 30 秒, 68 $^{\circ}$ C 下 30 秒), 68 $^{\circ}$ C 下 5 分钟, 以及 4 $^{\circ}$ C 下无限期暂停, 并使用 MINELUTE PCR purification kit (Qiagen, Valencia, CA) 根据生产商的推荐方法对其进行纯化。

[0523] 大规模并行焦磷酸测序反应 (亦称为 454 测序) 如 (Margulies et al. (2005) Nature 437:376-380) 所述直接对 PCR 产物实施。454 测序结构的分析通过鉴定出在所述 DNA 分子中含有预期大小与位置的缺失的序列表示来进行。这些分析的结果表明多个 9-12bp 的小缺失存在于这些 ZFN 的预期剪切位点, 如图 6 所示。四十八个缺失中的四十六个见于得自经 ZFN 构建体 pDAB8147 稳定转化的绿色愈伤组织的序列表示 (表 3)。获得了使用相同 ZFN 对的两个其它缺失, 一个来自经暂时转化处理的原生质体 DNA (pDAB7151) 而另一个来自经农杆菌 (pDAB7147) 暂时转化处理的下胚轴组织。这些缺失恰好位于 ZFN 靶位点处, 且表明产生了 ZFN 诱导的 ds 断裂, 其随后由 NHEJ 机理所修复。

[0524] 因为通过用于本实施例的 PCR 分析无法分辨共生同源基因序列 D 与 E, 所述共生同源基因两者其一或均由所述 ZFN 剪切都是可能的。这些结果进一步说明了所述改造 ZFN 以特定方式在作物物种内的内源基因基因座处诱导靶向双链断裂的能力。其进一步证明在现行实验条件下, ZFN 稳定转化的方法为供筛选 NHEJ 最有效的方法。存在于构建体 pDAB7150 与 pDAB7154 中的 ZFN 在经不同转化方法处理的多种方法中并不显示任何缺失 (参见表 4)。

[0525] 表 4. 大规模并行测序的结果, 显示在得自用 ZFN pDAB7147 与 pDAB7151 暂时与稳定转化的欧洲油菜下胚轴节片 (片段) 的 EPSPS 共生同源基因 D 目标序列中的 NHEJ。所述



对照试样由未经 ZFN 处理的组织组成。

[0526]

S. N.	试样	ZFN 构建体 #	试样 #- 引物	分析的序列	#of NHEJ
1	对照	pDAB7147+ pDAB7151	1- 正向	18, 215	0
2	对照	pDAB7147+ pDAB7151	1- 反向	18, 922	0
3	原生质体	pDAB7151	2- 正向	45, 896	0
4	原生质体	pDAB7151	2- 反向	48, 606	1
5	触须	pDAB7151	3- 正向	19, 601	0
6	触须	pDAB7151	3- 反向	19, 628	0
7	暂时农杆菌	pDAB7147	4- 正向	31, 281	1
8	暂时农杆菌	pDAB7147	4- 反向	31, 595	0
9	稳定农杆菌 -G	pDAB7147	5- 正向	13, 795	29
10	稳定农杆菌 -G	pDAB7147	5- 反向	13, 221	17
11	稳定农杆菌 -B	pDAB7147	6- 正向	8, 167	0
12	稳定农杆菌 -13	pDAB7147	6- 反向	7, 549	0
	总计			276, 476	48

[0527]

[0528] 为努力在其它 EPSPS 共生同源基因中分析 ZFN 诱导双链断裂, 实施 PCR 与大规模并行焦磷酸测序的组合以调查剩余 EPSPS 共生同源基因 DNA 中的 ZFN 诱导双链断裂。使用了来自如部分 4D-E 所述的经含 pDAB7147 与 pDAB7150 的农杆菌菌株稳定转化的同一“绿色”愈伤组织的经 MaeIII 消化的基因组 DNA。PCR 扩增随即使用对 EPSPS 共生同源基因 A, B, C 与 D 有特异性, 锚于基因组 DNA 上所述 ZFN 预期剪切位点两侧的寡聚核苷酸引物进行。将针对共生同源基因 A 的正向 PCR 引物 (5'-TCCCAGCTTCTTTAGATTCTAAGG-3') (SEQ ID

NO:17) 与反向 PCR 引物 (5' -CTGCAACTTTTCACATAGCAA-3') (SEQ ID NO:18), 针对共生同源基因 B 的正向 PCR 引物 (5' -CAAGAGTGATATCGAGTTGTACCTTGGGAATGCT-3') (SEQ ID NO:19) 与反向 PCR 引物 (5' -AGGCCATCATATCGAGCAAACGCAGT-3') (SEQ ID NO:20), 针对共生同源基因 C 的正向 PCR 引物 (5' -GGGTAAACAACCGTGCTGTA-3') (SEQ IDNO:21) 与反向 PCR 引物 (5' -AAAGACTGCTGCAAACAAGATC-3') (SEQ IDNO:22) 以及如部分 4D-E 所述同样组的正向与反向引物 (SEQ ID NO:23 与 SEQID NO:24), 组合用于针对每个共生同源基因分别扩增基因组 DNA:50  $\mu$ l 反应体积含有 200ng 经 Mae III 消化的 gDNA 模板 (10  $\mu$ l), 1.25  $\mu$ l 每种引物 (每种 10  $\mu$ M), 5  $\mu$ l 的 10x Accuprime PCR buffer II, 5  $\mu$ l 的 10% PVP-40 以及 0.3  $\mu$ l (1.5 单位) 在酶生产商提供的缓冲液中的 ACCUPRIME TaqDNA 高保真度多聚酶 (Invitrogen, Carlsbad, CA)。扩增产物由下述扩增循环产生:94 $^{\circ}$ C 下 2 分钟, 25 周期 (94 $^{\circ}$ C 下 30 秒, 退火 30 秒, 68 $^{\circ}$ C 下 1 分钟), 68 $^{\circ}$ C 下 5 分钟, 在 4 $^{\circ}$ C 下无限期暂停。下列共生同源基因的退火温度如下所述:A = 62.1 $^{\circ}$ C B = 65 $^{\circ}$ C, C = 65 $^{\circ}$ C 而 D = 59.3 $^{\circ}$ C。

[0529] 该初步扩增产物随即使用 Qiagen MinElute PCR Purification Kit (Qiagen, Valencia, CA) 纯化并以 10  $\mu$ l EB 缓冲液洗脱。合成并用 HPLC (IDT, Coralville, IA) 纯化了共生同源基因 A 正向 PCR 引物的三个变体 (5' -XXXATCGAGTTGTACCTTGGGAATG-3') (SEQ ID NO:27), 其中 XXX = GGC, CGG 或 GCC 以及其反向 PCR 引物的三个变体 (5' -XXX AATAAGTCCTTAACCTTACCTT-3') (SEQ ID NO:28), 其中 XXX = GGC, CGG 或 GCC。合成并用 HPLC (IDT, Coralville, IA) 纯化了共生同源基因 B 正向 PCR 引物的三个变体 (5' -XXX AGAGTGATATCGAGTTGTACCTTG-3') (SEQ IDNO:29), 其中 XXX = CGG, CGC 或 GCC 以及其反向 PCR 引物的三个变体 (5' -XXX AACTCCTTAACCTTACCTT-3') (SEQ ID NO:30), 其中 XXX = CGG, CGC 或 GCC。合成并用 HPLC (IDT, Coralville, IA) 纯化了共生同源基因 C 正向 PCR 引物的三个变体 (5' -XXXAGAGTGATATTGAGTTGTACCTTG-3') (SEQ ID NO:31), 其中 XXX = CGG, GGC 或 GCC 以及其反向 PCR 引物的三个变体 (5' -XXXAAAGCTCCTTAACCTTTACCT-3') (SEQ ID NO:32), 其中 XXX = CGG, GGC 或 GCC。对共生同源基因 D 进行二次 PCR 扩增的引物 (SEQ ID NO:25 与 SEQID NO:26) 描述于部分 4D。每个引物 5' 末端的 3bp 标签用作识别密钥并表明该扩增子来自哪个欧洲油菜试样。将具有匹配识别标签 (密钥) 的引物对组合用于扩增纯化的下述初步 PCR 扩增子, 即所述扩增子以下述条件源自上述试样:50  $\mu$ l 反应体积含有 10  $\mu$ l 以 1:10 稀释的纯化 PCR 扩增子, 1.25  $\mu$ l 每种引物 (每种 10  $\mu$ M), 5  $\mu$ l 10x Accuprime PCR 缓冲液 I 以及 0.3  $\mu$ l (1.5 单位) 在酶生产商提供的缓冲液中 Accuprime TaqDNA 高保真度聚合酶 (Invitrogen, Carlsbad, CA)。预期大小的扩增产物得自下述扩增循环:94 $^{\circ}$ C 下 2 分钟, 33 周期 (94 $^{\circ}$ C 下 30 秒, 62 $^{\circ}$ C 下 30 秒, 68 $^{\circ}$ C 下 30 秒), 68 $^{\circ}$ C 下 5 分钟并在 4 $^{\circ}$ C 下无限期暂停, 并使用 Qiagen (Valencia, CA) 的 MinElute PCR purification kit 根据生产商的推荐方法纯化。

[0530] 直接对 PCR 产物如部分 4D 所述进行大规模并行焦磷酸测序反应。所述测序结果的分析通过鉴定出在所述 DNA 分子中含有预期大小与位置的缺失的序列表示来进行。

[0531] 这些分析的结果表明在共生同源基因 C 与 D 中在预期的 ZFN 剪切位点处存在多个小尺度缺失 (图 7)。pDAB7147 ZFN 再一次对剪切两个共生同源基因, C 与 D (以及 E), 为有效的。这些 5-32bp 的剪切精确位于 ZFN 的靶位点, 并说明了 pDAB7147 ZFN 剪切 2 个或更多 EPSPS 共生同源基因。这些结果进一步说明了这些改造 ZFN 以特定方式在作物物种中在

同源基因基因座处诱导靶向双链断裂的能力。

[0532] 在经 pDAB7147 与 pDAB7150 处理的试样 5,6 与 9 的共生同源基因 A 与 B 中亦各发现了一个 NHEJ。这些 NHEJ 见于预期的位置。然而,因为一些对照试样亦含有一个 NHEJ(表 5 中的试样 2,7 与 13),所述 ZFN 无法视为有效剪切这些共生同源基因。

[0533] 表 5. 大规模并行测序的结果,显示下述的 DNA 分子,即其发生了 ZFN 介导双链断裂,继以在欧洲油菜的四个 EPSPS 共生同源基因中的靶序列上发生了 NHEJ 修复。所述对照试样代表未经 ZFN 处理的转基因愈伤组织。

[0534]

试样号	靶共生同源基因	ZFN 构建体	引物	比对的序列数	总 NHEJ
1	A	对照	正向	18,451	0
2	A	对照	反向	19,005	1
3	A	pDAB7150	正向	14,711	0
4	A	pDAB7150	反向	15,616	0
5	A	pDAB7147	正向	21,478	1
6	A	pDAB7147	反向	22,459	1
7	B	对照	正向	12,682	1
8	B	对照	反向	18,033	0
9	B	pDAB7150	正向	11,361	1
10	B	pDAB7150	反向	13,017	0
11	B	pDAB7147	正向	16,565	0
12	B	pDAB7147	反向	17,499	0
13	C	对照	正向	20,516	1
14	C	对照	反向	14,790	0
15	C	pDAB7150	正向	9,073	0
16	C	pDAB7150	反向	9,096	1

试样号	靶共生同源基因	ZFN 构建体	引物	比对的序列数	总 NHEJ
17	C	pDAB7147	正向	12, 696	3
18	C	pDAB7147	反向	14, 719	19
23	D	pDAB7147	正向	7, 028	5
24	D	pDAB7147	反向	6, 832	9

[0535]

[0536] 实施例 5:第二个 ZFN 在欧洲油菜中剪切所述五个共生同源基因中剩下的两个

[0537] 接下来,着眼于在剩下两个共生同源基因 A 与 B 中诱导双链断裂,尝试了 ZFN 介导双链断裂。使用了两个新的改造 ZFN,其以位于距第一个 ZFN 结合位置 5' 约 350bp 的不同序列为靶(图 4)。这些特定 ZFN 构建体, pDAB7185 与 pDAB7186(表 3),用于欧洲油菜下胚轴节片的稳定转化,如实施例 4,部分 B 所述。根据之前实施例 4,部分 C 所述将用 ZFN 稳定转化的愈伤组织冻结,冷冻干燥并提取 DNA。随即使用或者 BsoB1(New England Biolabs, Ipswich, MA) 或者 Lwe I(Fermentas, Inc., Hanover, MD) 消化 DNA 以富集发生了 NHEJ 的片段(图 4)。所述消化根据生产商的指示实施过夜并如前所述通过乙醇沉淀纯化。随即使用对靶基因具有特异性,且包夹所述 ZFN 预期剪切位点的引物进行 PCR 扩增。共生同源基因 A 的正向 PCR 引物(5' -CAGCGTGGAGCTTATCAGA-3')(SEQ ID NO:33)与反向 PCR 引物(5' -AAACGCAACACTAAGCAAAC-3')(SEQ ID NO:35),共生同源基因 B 的正向 PCR 引物(5' -GAAGAGTAACAACGGCTCTGTG-3')(SEQ ID NO:34)与反向 PCR 引物(5' -GAAAGAAAGAAGCAAACCGAC-3')(SEQ ID NO:90),其对相应靶共生同源基因具有特异性,组合用于在下述条件下扩增纯化基因组 DNA。

[0538] 对共生同源基因 A,50  $\mu$ l 反应体积包含 420-700ng BsoBI 消化的基因组 DNA 模板(10  $\mu$ l),1.25ul 每种引物(每种 10  $\mu$ M),5  $\mu$ l 的 10x Accuprime PCR 缓冲液 II,5  $\mu$ l 的 10% PVP-40 与 0.3  $\mu$ l(1.5 单位)在酶生产商提供的缓冲液中的 Accuprime TaqDNA 高保真度多聚酶(Invitrogen, Carlsbad, CA)。扩增产物由下述扩增循环产生:94 $^{\circ}$ C 下 2 分钟,28 周期(94 $^{\circ}$ C 下 30 秒,60 $^{\circ}$ C 下 30 秒,68 $^{\circ}$ C 下 1 分钟),68 $^{\circ}$ C 下 5 分钟,4 $^{\circ}$ C 下无限期暂停。对共生同源基因 B,50  $\mu$ l 反应体积包含 420-700ng BsoBI 消化的基因组 DNA 模板(10  $\mu$ l),1.25ul 每种引物(每种 10  $\mu$ M),5  $\mu$ l 的 10x Accuprime PCR 缓冲液 II,5  $\mu$ l 的 10% PVP-40 与 0.3  $\mu$ l(1.5 单位)在酶生产商提供的缓冲液中的 Accuprime TaqDNA 高保真度多聚酶(Invitrogen, Carlsbad, CA)。扩增产物由下述扩增循环产生:94 $^{\circ}$ C 下 2 分钟,28 周期(94 $^{\circ}$ C 下 30 秒,58 $^{\circ}$ C 下 30 秒,68 $^{\circ}$ C 下 1 分钟),68 $^{\circ}$ C 下 5 分钟,4 $^{\circ}$ C 下无限期暂停。该初步扩增产物随即使用 MINELUTE PCR purification kit(Qiagen, Valencia, CA) 分离并以 10  $\mu$ l EB 缓冲液洗脱。合成并用 HPLC(IDT, Coralville, IA) 纯化了共生同源基因 A 正向 PCR 引物的三个变体(5' -XXXTCTGTTTCCACGGCGGAG-3')(SEQ ID NO:36),其中 XXX = CCG, GCG 或 CGC 及其反向 PCR 引物的三个变体((5' -XXXAAGCGGCAAGAAGAAGAATC-3')(SEQ ID NO:37),其中 XXX = CCG, GCG 或 CGC。合成并用 HPLC(IDT, Coralville, IA) 纯化了共生同源基因

B 正向 PCR 引物的三个变体 (5' -XXX TCTGTTTCCACGGCTGAG-3') (SEQ ID NO :38), 其中 XXX = GGC, GCC 或 CGG 以及其反向 PCR 引物的三个变体 (5' -XXXATTGGACAGAGATTTGGGTC-3') (SEQ ID NO :39), 其中 XXX = GGC, GCC 或 CGG。每个引物 5' 末端的 3bp 标签用作识别密钥并表明该扩增子来自哪个欧洲油菜试样。将具有匹配识别标签 (密钥) 的引物对组合用于扩增纯化的下述初步 PCR 扩增子, 即所述扩增子以下述条件源自上述试样: 50  $\mu$  l 反应体积含有 10  $\mu$  l 以 1 : 10 稀释的纯化 PCR 扩增子, 1.25  $\mu$  l 每种引物 (每种 10  $\mu$  M), 5  $\mu$  l 10x ACCUPRIME PCR 缓冲液 I 以及 0.3  $\mu$  l (1.5 单位) 在酶生产商提供的缓冲液中 ACCUPRIME TaqDNA 高保真度聚合酶 (Invitrogen, Carlsbad, CA)。

[0539] 预期大小的扩增产物得自下述扩增循环: 94°C 下 2 分钟, 25 周期 (94°C 下 30 秒, 退火 30 秒, 68°C 下 30 秒), 68°C 下 5 分钟并在 4°C 下无限期暂停, 并使用 Qiagen (Valencia, CA) 的 MINELUTE PCR purification kit 根据生产商的推荐方法纯化。上面所列第二次 PCR 反应的退火温度如下所述: 对共生同源基因 A, 66°C, 而对共生同源基因 B, 64°C。

[0540] 直接对 PCR 产物进行大规模并行焦磷酸测序反应。所述测序结果的分析通过鉴定出在所述 DNA 分子中含有预期大小与位置的缺失的序列表示来进行。

[0541] 这些分析的结果表明在预期剪切位点处存在多个小尺度缺失 (图 8, 表 6)。同样, 这些缺失精确位于所述 ZFN 靶位点, 并表明由 ZFN 诱导的双链断裂产生于基因组中, 并随后由 NHEJ 修复。克隆于 pDAB7185 的 ZFN 总体上较之 ZFN pDAB7186 对引起双链断裂为更加有效。

[0542] 这些结果进一步说明了这些改造 ZFN 以特定方式在作物物种中同源基因基因座处诱导靶向双链断裂的能力。该结果亦说明了同样的 ZFN 剪切两个 EPSPS 共生同源基因, A 与 B 的能力。

[0543] 表 6. 大规模并行测序的结果, 显示下述的 DNA 分子, 即其发生了 ZFN 介导双链断裂, 继以在欧洲油菜的两个 EPSPS 共生同源基因, A 与 B 中的靶序列上发生了 NHEJ 修复。所述对照试样代表未经 ZFN 处理的转基因愈伤组织。

[0544]

S. N.	靶共生同源基因	ZFN 构建体	引物	比对的序列数	总 NHEJ
1	A	对照	正向	12903	0
2	A	对照	反向	12027	0
3	A	pDAB7185	正向	10432	9
4	A	pDAB7185	反向	9883	3
5	A	pDAB7186	正向	20,496	2
6	A	pDAB7186	反向	18697	3
7	B	对照	正向	20655	0

S. N.	靶共生同源基因	ZFN 构建体	引物	比对的序列数	总 NHEJ
8	B	对照	反向	22733	0
9	B	pDAB7185	正向	15663	32
10	B	pDAB7185	反向	15864	26
11	B	pDAB7186	正向	21333	4
12	B	pDAB7186	反向	24003	12

[0545] 实施例 6 :可剪切所有 EPSPS 共生同源基因的两个 ZFN

[0546] 在另一个实施例中,用两个 ZFN 共转化欧洲油菜下胚轴节片,并建立了含有所述 ZFN 两者的稳定转基因事件,其在所有 EPSPS 共生同源基因中表明了 NHEJ。在该实验中用于共转化 ZFN 构建体为 pDAB7147 与 pDAB7185(表 3)。如实施例 4 与 5 所述产生了稳定转基因愈伤组织事件,分离并分析了 DNA。同源,由于序列相似性,难以分辨共生同源基因 D 与 E 中的 NHEJ。所有的 NHEJ 位于所述多种共生同源基因的预期靶位点上。

[0547] 这些结构确证了下述两点:1. 可针对多基因家族的相异序列设计 ZFN 以特异性的剪切 1-2 个基因 / 共生同源基因。2. 若需要,可使用多个 ZFN 以同时剪切所有的共生同源基因。

[0548] 其它与靶向剪切,靶向重组与靶向整合的信息可参见 United States Patent Application publications US-2003-0232410 ;US-2005-0026157 ;US-2005-0064474 ;US-2005-0208489 与 US-2007-0134796,其公开以全文引用的方式包含于本文中以供所有目的。

[0549] 所有本文提及的专利,专利申请与出版物均以全文引用的方式包含于本文中,以供所有目的。

[0550] 尽管本公开以部分详述的方式通过说明与举例的方法提供以供清楚理解,对本领域技术人员显而易见可对其实施多种变化与修饰而不偏离本公开的精神与范围。相应的,前述描述与实施例不应理解为构成限定。

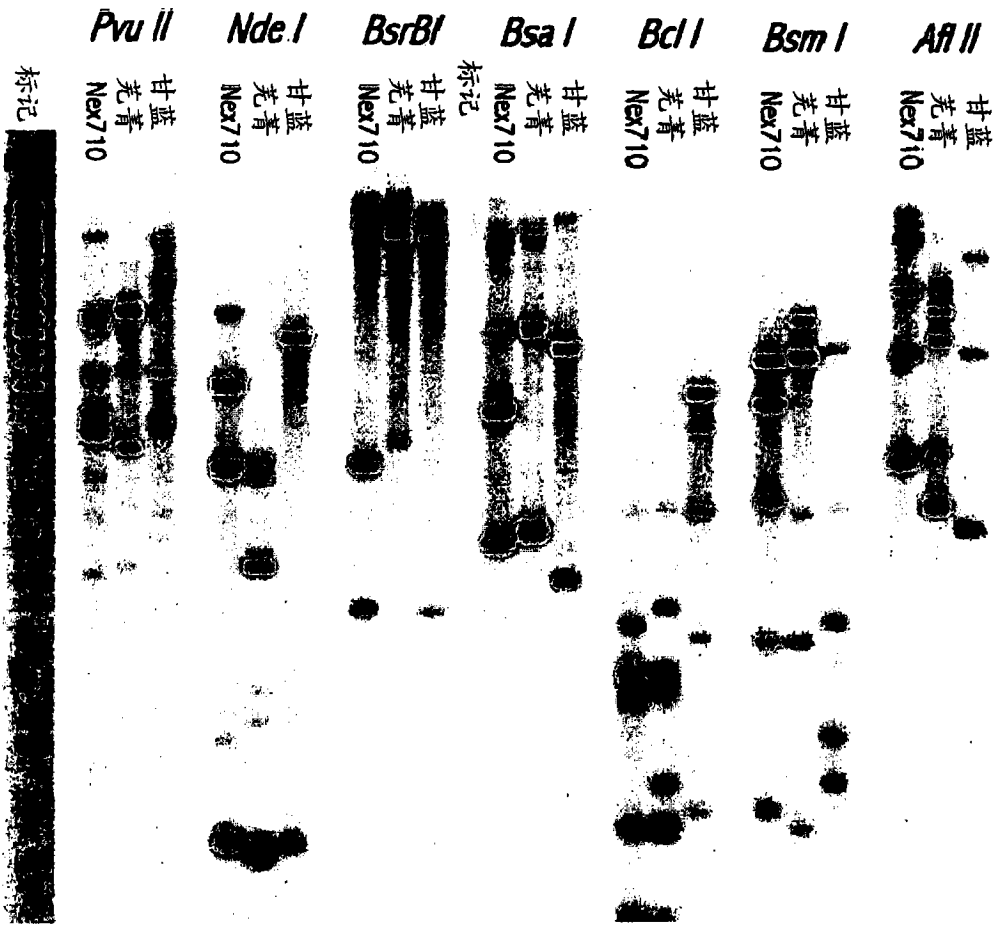


图 1

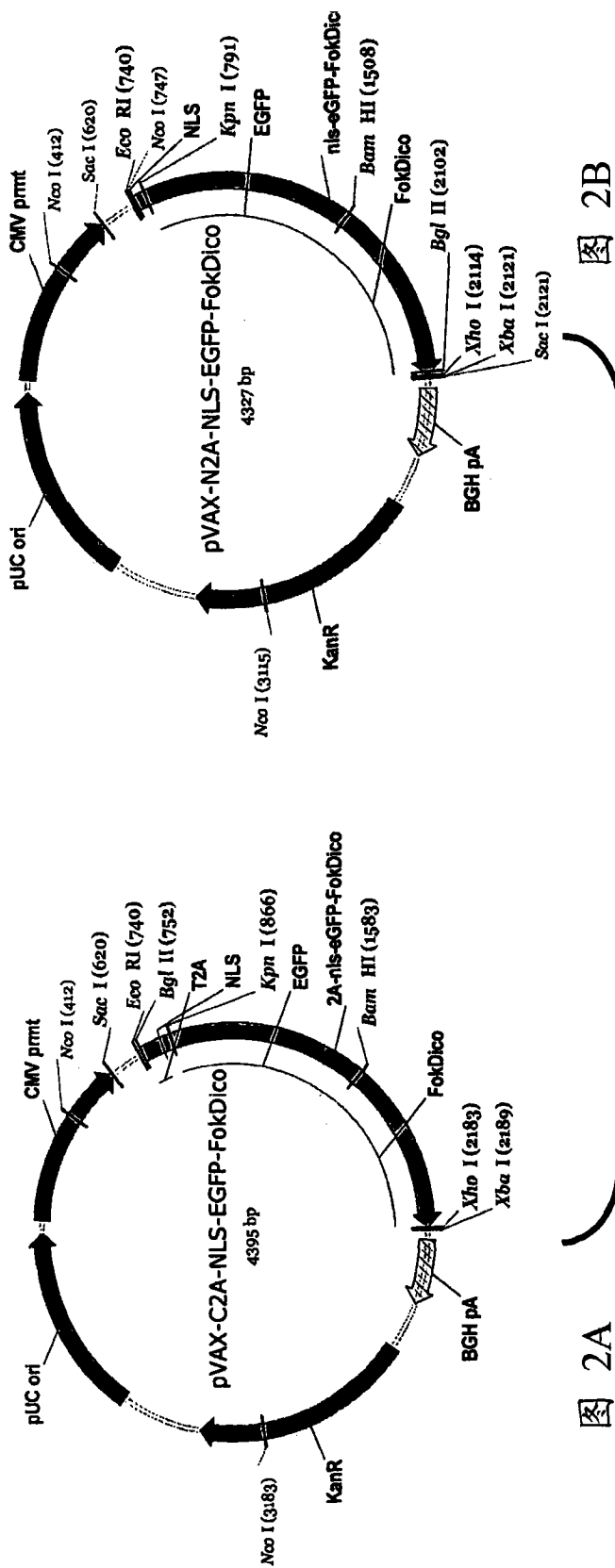


图 2B

图 2A



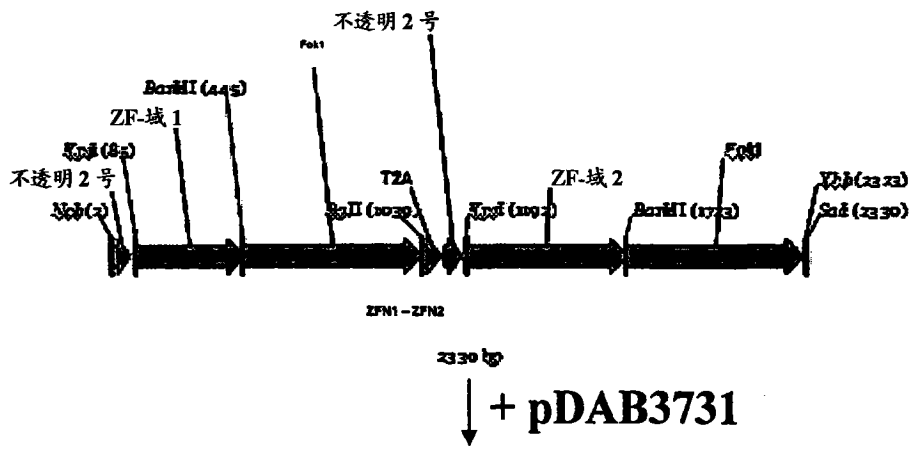
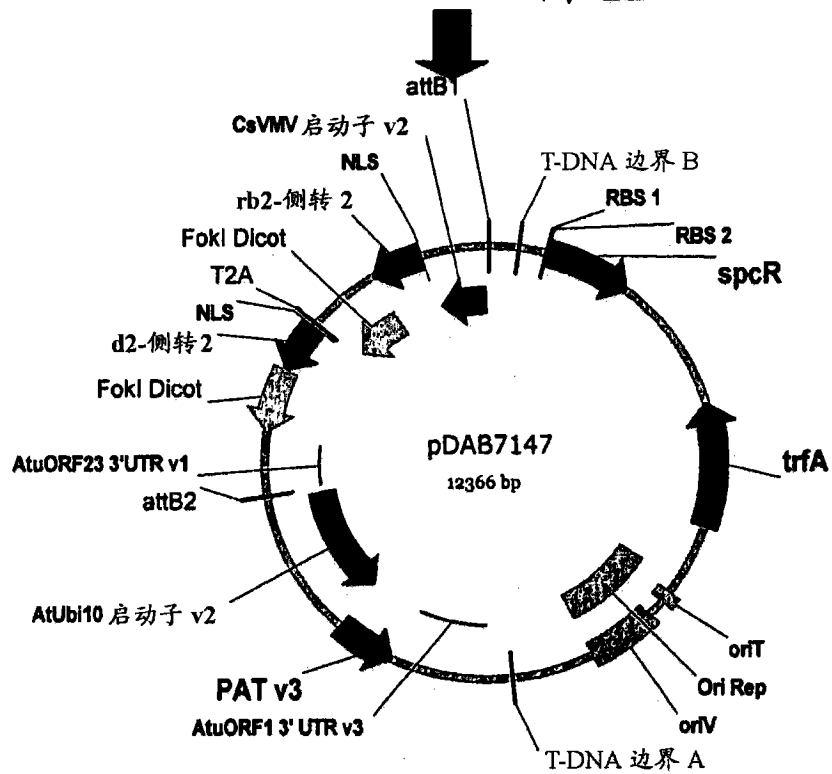
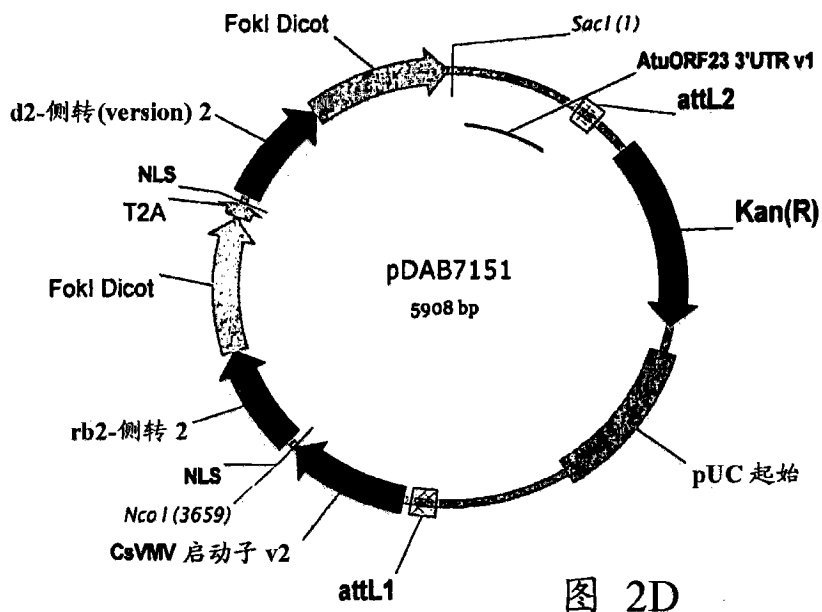


图 2C



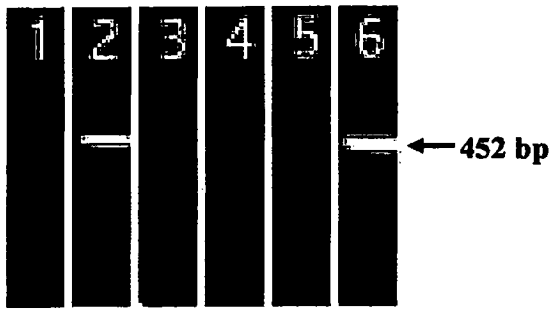


图 3A

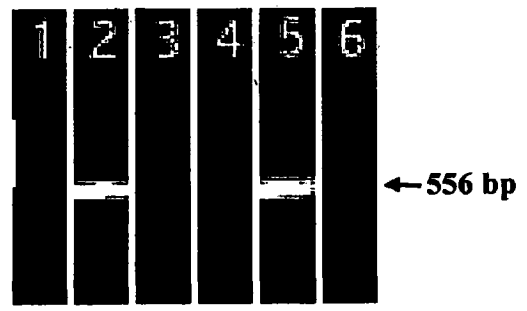


图 3B

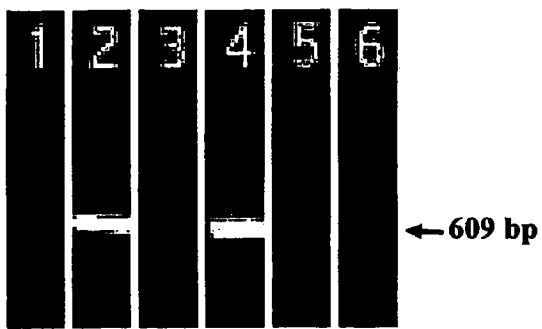


图 3C

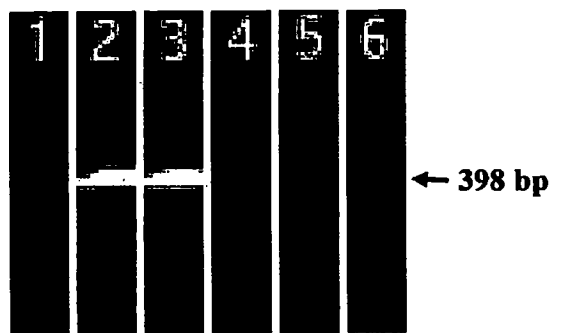
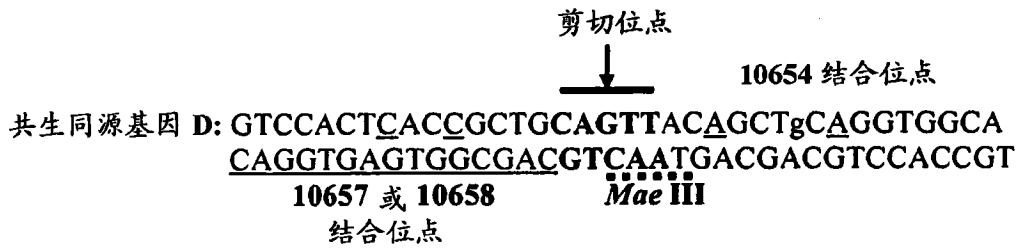
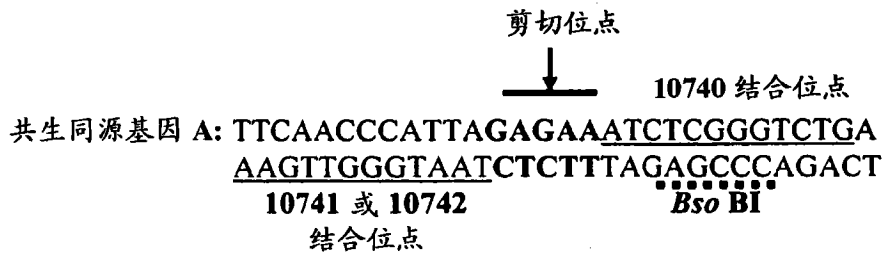


图 3D



在 EPSPS 共生同源基因中 ZFN 10654/10657 结合位点的序列比较, 小写字母的碱基在 ZFN 设计中跳过(skip).

共生同源基因 A: GTCCACTTACAGCTGCAGTTACTGCTGCTIGGTGGCA  
 共生同源基因 B: GTCCACTTACCGCTGCAGTTACTIGCTgCAGGTGGCA  
 共生同源基因 C: GTCCACTCACCGCTGCAGTTACTGCTgCAGGTGGCA  
 共生同源基因 D: GTCCACTCACCGCTGCAGTTACAGCTgCAGGTGGCA  
 共生同源基因 E: GTCCACTCACCGCTGCAGTTACAGCTgCAGGTGGCA



在 EPSPS 共生同源基因中 ZFN 10740/10741 结合位点的序列比较

共生同源基因 A: TTCAACCCATTAGAGAAATCTCGGGTCTG  
 共生同源基因 B: TTCAGCCCCATTAGAGAAATCTCGGGTCTG  
 共生同源基因 C: TTCAGCCCCATTAGAGAAATCTCGGGTCTG  
 共生同源基因 D: TTCAACCAATCAGAGAAATCTCGGGTCTC  
 共生同源基因 E: TTCAACCAATCAGAGAAATCTCGGGTCTC

图 4

Sequencher - [Contig[0001]]

File Edit Select Contig Sequence View Window Help

Overview Summary Cut Map Find Show Chromatograms ReAligner

MGU0000535_DAS-MG_F1_A02_M13R	GGGAATGCAGGAAACAGCCATGCGTCCACTCACCGCTGCA::TACAGCTGCAGGTGGCAACCGAGGTTAAGGTTAACCGAGTTT	70	80	90	100	110	120	130	140	15
MGU0000532_DAS-MG_F1_A02_M13F	GGGAATGCAGGAAACAGCCATGCGTCCACTCACCGCTGCA::TACAGCTGCAGGTGGCAACCGAGGTTAAGGTTAACCGAGTTT									
MGU0000535_DAS-MG_F1_A00_M13R	GGGAATGCAGGAAACAGCCATGCGTCCACTCACCGCTGCA::TACAGCTGCAGGTGGCAACCGAGGTTAAGGTTAACCGAGTTT									
MGU0000535_DAS-MG_F1_A00_M13F	GGGAATGCAGGAAACAGCCATGCGTCCACTCACCGCTGCA::TACAGCTGCAGGTGGCAACCGAGGTTAAGGTTAACCGAGTTT									
MGU0000538_DAS-MG_F1_A11_M13F	GGGAATGCAGGAAACAGCCATGCGTCCACTCACCGCTGCA::TACAGCTGCAGGTGGCAACCGAGGTTAAGGTTAACCGAGTTT									
MGU0000538_DAS-MG_F1_A11_M13R	GGGAATGCAGGAAACAGCCATGCGTCCACTCACCGCTGCA::TACAGCTGCAGGTGGCAACCGAGGTTAAGGTTAACCGAGTTT									
MGU0000535_DAS-MG_F1_B04_M13R	GGGAATGCAGGAAACAGCCATGCGTCCACTCACCGCTGCA::TACAGCTGCAGGTGGCAACCGAGGTTAAGGTTAACCGAGTTT									
MGU0000535_DAS-MG_F1_B04_M13F	GGGAATGCAGGAAACAGCCATGCGTCCACTCACCGCTGCA::TACAGCTGCAGGTGGCAACCGAGGTTAAGGTTAACCGAGTTT									
MGU0000535_DAS-MG_F1_E05_M13R	GGGAATGCAGGAAACAGCCATGCGTCCACTCACCGCTGCA::TACAGCTGCAGGTGGCAACCGAGGTTAAGGTTAACCGAGTTT									
MGU0000535_DAS-MG_F1_E05_M13F	GGGAATGCAGGAAACAGCCATGCGTCCACTCACCGCTGCA::TACAGCTGCAGGTGGCAACCGAGGTTAAGGTTAACCGAGTTT									
MGU0000535_DAS-MG_F1_C04_M13R	GGGAATGCAGGAAACAGCCATGCGTCCACTCACCGCTGCA::TACAGCTGCAGGTGGCAACCGAGGTTAAGGTTAACCGAGTTT									
MGU0000535_DAS-MG_F1_C04_M13F	GGGAATGCAGGAAACAGCCATGCGTCCACTCACCGCTGCA::TACAGCTGCAGGTGGCAACCGAGGTTAAGGTTAACCGAGTTT									
MGU0000538_DAS-MG_F1_D01_M13R	GGGAATGCAGGAAACAGCCATGCGTCCACTCACCGCTGCA::TACAGCTGCAGGTGGCAACCGAGGTTAAGGTTAACCGAGTTT									
MGU0000538_DAS-MG_F1_D01_M13F	GGGAATGCAGGAAACAGCCATGCGTCCACTCACCGCTGCA::TACAGCTGCAGGTGGCAACCGAGGTTAAGGTTAACCGAGTTT									
MGU0000538_DAS-MG_F1_E05_M13R	GGGAATGCAGGAAACAGCCATGCGTCCACTCACCGCTGCA::TACAGCTGCAGGTGGCAACCGAGGTTAAGGTTAACCGAGTTT									
MGU0000538_DAS-MG_F1_E05_M13F	GGGAATGCAGGAAACAGCCATGCGTCCACTCACCGCTGCA::TACAGCTGCAGGTGGCAACCGAGGTTAAGGTTAACCGAGTTT									
MGU0000538_DAS-MG_F1_F12_M13R	GGGAATGCAGGAAACAGCCATGCGTCCACTCACCGCTGCA::TACAGCTGCAGGTGGCAACCGAGGTTAAGGTTAACCGAGTTT									
MGU0000538_DAS-MG_F1_F12_M13F	GGGAATGCAGGAAACAGCCATGCGTCCACTCACCGCTGCA::TACAGCTGCAGGTGGCAACCGAGGTTAAGGTTAACCGAGTTT									
MGU0000535_DAS-MG_F1_G05_M13R	GGGAATGCAGGAAACAGCCATGCGTCCACTCACCGCTGCA::TACAGCTGCAGGTGGCAACCGAGGTTAAGGTTAACCGAGTTT									
MGU0000535_DAS-MG_F1_G05_M13F	GGGAATGCAGGAAACAGCCATGCGTCCACTCACCGCTGCA::TACAGCTGCAGGTGGCAACCGAGGTTAAGGTTAACCGAGTTT									
MGU0000535_DAS-MG_F1_H10_M13R	GGGAATGCAGGAAACAGCCATGCGTCCACTCACCGCTGCA::TACAGCTGCAGGTGGCAACCGAGGTTAAGGTTAACCGAGTTT									
MGU0000535_DAS-MG_F1_H10_M13F	GGGAATGCAGGAAACAGCCATGCGTCCACTCACCGCTGCA::TACAGCTGCAGGTGGCAACCGAGGTTAAGGTTAACCGAGTTT									
MGU0000535_DAS-MG_F1_H11_M13R	GGGAATGCAGGAAACAGCCATGCGTCCACTCACCGCTGCA::TACAGCTGCAGGTGGCAACCGAGGTTAAGGTTAACCGAGTTT									
MGU0000535_DAS-MG_F1_H11_M13F	GGGAATGCAGGAAACAGCCATGCGTCCACTCACCGCTGCA::TACAGCTGCAGGTGGCAACCGAGGTTAAGGTTAACCGAGTTT									
W8 Type EPSFS ParSeq DYE										

Select Next Ambiguous Base \*  
Spacer

图 5

野生型  
**AACAGCCATGCCGTCCACTCACCGCTGCAGTTACAGCTGCAGGTGGCAACGCCGAGGTAAGGTTAACGAGTTTTTTGTTA**

试样 2  
**AACAGCCATGCCGTCCACTCACCGCTGCAGT:::::: TGGCAACGCCGAGGTAAGGTTAACGAGTTTTTTGTTA**

试样 4  
**AACAGCCATGCCGTCCACTCACCGCTGC:::::: TGCAGGTGGCAACGCCGAGGTAAGGTTAACGAGTTTTTTGTTA**

试样 5 (35)  
**AACAGCCATGCCGTCCACTCACCGCTGC:::::: TGCAGGTGGCAACGCCGAGGTAAGGTTAACGAGTTTTTTGTTA**

试样 5  
**AACAGCCATGCCGTCCACTCACCGCTGC:::::: TGCAGGTGGCAACGCCGAGGTAAGGTTAACGAGTTTTTT: GTTA**

试样 5  
**AACAGCCATGCCGTCCACTCACCGCTGC:::::: TGCAGGTGGCAACGCCGAGGTAAGGTTAACGTTAGTTTTTTGTTA**

试样 5 (8)  
**AACAGCCATGCCGTCCAC:::::: TTACAGCTGCAGGTGGCAACGCCGAGGTAAGGTTAACGAGTTTTTTGTTA**

试样 5  
**AACAGCCATGCCGTCCAC:::::: TTACAGCTGCAGGTGGCAACGCCGAGGTAAGGTTAACGAGTTTTTT: GTTA**

图 6

共生同源基因 C-野生型反向序列

AAAGCTCCTTAACCTTACCTTGGCTGGCCACCTGCAGCAGTAACGGCAGCGGTGAGTGGACCGCATGGCTGTTCTCTGCATTTCCCAAGGTA

共生同源基因 C-试样 16 NHEJ

AAAGCTCCTTAACCTTACCTTGGCG:::.....:TGGACGCATGGCTGTTCTCTGCATTTCCCAAGGTA

共生同源基因 C-野生型正向序列

AGAGTGATATTGAGTTGTACCTTGGGAATGCAGGAACAGCCATGCGTCCACTCACCCGCTGCCGTTACTGCTGCAGGTGGCAACGCCAAGGT

共生同源基因 C-试样 17 NHEJs

AGAGTGATATTGAGTTGTACCTT:GGAAATGCAGGAACAGCCATGCGTCCACTCACCCGCTG:::ACTGCTGCAGGTGGCAACGCCAAGGT  
AGAGTGATATTGAGTTGTACCTTGGGAATGCAGGAACAGCCATGCGTCCACTCACCCGCTG:::ACTGCTGCAGGTGGCAACGCCAAGGT  
AGAGTGATATTGAGTTGTACCTTGGGAATGCAGGAACAGCCATGCGTCCACTCACCCGCTG:::ACTGCTGCAGGTGGCAACGCCAAGGT

共生同源基因 D-野生型序列

AGTTGTACCTTGGGAATGCAGGAACAGCCATGCGTCCACTCACCCGCTGCAGTTACAGCTGCAGGTGCCAACGCCGAGGTAAGGTTAACGAG

共生同源基因 D-试样正向序列

AGTTGTACCTTGGGAATGCAGGAACAGCCATGC:::.....:AGGTGGCAACGCCGAGGTAAGGTTAACGAG  
AGTTGTACCTTGGGAATGCAGGAACAGCCATGC:::.....:AGGTGGCAACGCCGAGGTAAGGTTAACGAG  
AGTTGTACCTTGGGAATGCAGGAACAGCCATGC:::.....:AGGTGGTAAACGCCGAGGTAAGGTTAACGAG  
AGTTGTACCTTGGGAATGCAGGAACAGCCATGCGTCCACTCACCCGCTG:::.....:CTGCAGGTGGCAACGCCGAGGTAAGGTTAACGAG  
AGTTGTACCTTGGGAATGCAGGAACAGCCATGCGTCCACTCACCCGCTG:::.....:CTGCAGGTGGCAACGCCGAGGTAAGGTTAACGAG

图 7A

共生同源基因 D-野生型反向序列

ATCAATTTCTTGACAATAACAAAAAATCGTTAACCTTACCTCGCGTTGCCACCTGCAGCTGTAACTGCAGCGGTGAGTGGACGCATGGC

共生同源基因 D-试样 24 NHEJs

ATCAATTTCTTGACAATAACAAAAAATACGTTAACCTTACCTCGCGTTGCCACCTGCAG::: CAGCGGTGAGTGGACGCATGGC  
 ATCAATTTCTTGACAATAACAAAAAATACGTTAACCTTACCTCGCGTTGCCACCTGCAG::: CAGCGGTGAGTGGACGCATGGC  
 ATCAATTTCTTGACAATAACAAAAAATACGTTAACCTTACCTCGCGTTGCCACCTGCAG::: CAGCGGTGAGTGGACGCATGGC  
 ATCAATTTCTTGACAATAACAAAAAATACGTTAACCTTACCTCGCGTTGCCACCTGCAG::: CAGCGGTGAGTGGACGCATGGC  
 ATCAATTTCTTGACAATAACAAAAAATACGTTAACCTTACCTCGCGTTGCCACCTGCAG::: CAGCGGTGAGTGGACGCATGGC  
 ATCAATTTCTTGACAATAACAAAAAATACGTTAACCTTACCTCGCGTTGCCACCTGCAG::: CAGCGGTGAGTGGACGCATGGC  
 ATCAATTTCTTGACAATAACAAAAAATACGTTAACCTTACCTCGCGTTGCCACCTGCAG::: CAGCGGTGAGTGGACGCATGGC  
 ATCAATTTCTTGACAATAACAAAAAATACGTTAACCTTACCTCGCGTTGCCACCTGCAG::: CAGCGGTGAGTGGACGCATGGC  
 ATCAATTTCTTGACAATAACAAAAAATACGTTAACCTTACCTCGCGTTGCCACCTGCAG::: CAGCGGTGAGTGGACGCATGGC  
 ATCAATTTCTTGACAATAACAAAAAATACGTTAACCTTACCTCGCGTTGCCACCTGCAG::: CAGCGGTGAGTGGACGCATGGC

图 7B



共生同源基因 A-野生型正向序列

AGCTTCGGAGATTGTGCTTCAACCCATTAGAGAAATCTCGGGTCTGATCAAGCTACCCGGATCCRAAATCTCTGTCCAACCG

共生同源基因 A-试样 3 NHEJs

AGCTTCGGAGATTGTGCTTCAACCC:.....:GGATCCAAATCTCTGTCCAACCG

AGCTTCGGAGATTGTGCTTCAACCCA:.....:GGTCTGATCAAGCTACCCGGATCCAAATCTCTGTCCAACCG

AGCTTCGGAGATTGTGCTTCAACCCA:.....:TCGGGTCTGATCAAGCTACCCGGATCCAAATCTCTGTCCAACCG

AGCTTCGGAGATTGTGCTTCAACCCA:.....:TCGGGTCTGATCAAGCTACCCGGATCCAAATCTCTGTCCAACCG

AGCTTCGGAGATTGTGCTTCAACCCA:.....:TCGGGTCTGATCAAGCTACCCGGATCCAAATCTCTGTCCAACCG

AGCTTCGGAGATTGTGCTTCAACCCATT:.....:AGCTACCCGGATCCAAATCTCTGTCCAACCG

AGCTTCGGAGATTGTGCTTCAACCCATTA:.....:GGGTCTGATCAAGCTACCCGGATCCAAATCTCTGTCCAACCG

AGCTTCGGAGATTGTGCTTCAACCCATT:..:AAATCTCGGGTCTGATCAAGCTACCCGGATCCAAATCTCTGTCCAACCG

AGCTTCGGAGATTGTGCTTCAACCCATTAG:.....:GGGTCTGATCAAGCTACCCGGATCCAAATCTCTGTCCAACCG

共生同源基因 B-野生型正向序列

CGGCTGAGAAATCTTCGGAGATTGTGCTTCAGCCCAATTAGAGAAATCTCGGGTCTGATCAAGCTACCCGGACCCAAATC

图 8A



CCGCTGAGAAATCTTCGGAGATTGT:.....:CTGATCAAGCTACCCGGACCCCAAATC  
 CCGCTGAGAAATCTTCGGAGATTGT:.....:CTGATCAAGCTACCCGGACCCCAAATC  
 CCGCTGAGAAATCTTCGGAGATTGTGCTTCAGCCCA:.....:CTGATCAAGCTACCCGGACCCCAAATC  
 CCGCTGAGAAATCTTCGGAGATTGTGCTTCAGCCCA:.....:GGGTCTGATCAAGCTACCCGGACCCCAAATC  
 CCGCTGAGAAATCTTCGGAGATTGTGCTTCAGCCCA:.....:GGGTCTGATCAAGCTACCCGGACCCCAAATC  
 CCGCTGAGAAATCTTCGGAGATTGTGCTTCAGCCCA:.....:AATCTCGGGTCTGATCAAGCTACCCGGACCCCAAATC  
 CCGCTGAGAAATCTTCGGAGATTGTGCTTCAGCCCA:.....:AATCTCGGGTCTGATCAAGCTACCCGGACCCCAAATC  
 CCGCTGAGAAATCTTCGGAGATTGTGCTTCAGCCCAATT:.....:GTCTGATCAAGCTACCCGGACCCCAAATC  
 CCGCTGAGAAATCTTCGGAGATTGTGCTTCAGCCCAATT:.....:CGGGTCTGATCAAGCTACCCGGACCCCAAATC  
 CCGCTGAGAAATCTTCGGAGATTGTGCTTCAGCCCAATT:.....:CGGGTCTGATCAAGCTACCCGGACCCCAAATC  
 CCGCTGAGAAATCTTCGGAGATTGTGCTTCAGCCCAATT:.....:CGGGTCTGATCAAGCTACCCGGACCCCAAATC  
 CCGCTGAGAAATCTTCGGAGATTGTGCTTCAGCCCAATT:.....:TCCTGATCAAGCTACCCGGACCCCAAATC  
 CCGCTGAGAAATCTTCGGAGATTGTGCTTCAGCCCAATT:.....:AGAAATCTCGGGTCTGATCAAGCTACCCGGACCCCAAATC

图 8C

(SEQ ID NO:10)

```

1   GCAGCAGCGT GGAGCTTATC AGATATCTTC GCGGGGGTTG AAGAAGAGCG
51  CGATGGTGCT AAACCGTTCT GTAACCTCGTC CGGTTAAGGT TATGGCCTCT
101 GTTTCCACGG CGGAGAAAGC TTCGGAGATT GTGCTTCAAC CCATTAGAGA
151 AATCTCGGGT CTGATCAAGC TACCCGGATC CAAATCTCTG TCCAACCGGA
201 TTCTTCTTCT TGCCGCTTTA TCCGAGGTTT GCTTCTTTCT TTGTTTGCTT
251 AGTGTTGCGT TTTTAACGGC GTGAGGATGA AGAAAGGTTT TGACTTTGTT
301 GTGGTTTTAT AGGGAACACT TGTAGTTGAC AACTTGTTGA ACAGTGATGA
351 CATTAACTAC ATGCTTGATG CGTTGAACAA GTTGGGGCTT AATGTGGAAC
401 GTGACAGTGA GAACAACCGT GCGGTTGTTG AAGGATGTGG CGGGATATTC
451 CCAGCTTCTT TAGATTCTAA GGGTGATATC GAGTTGTACC TTGGGAATGC
501 AGGAACAGCC ATGCGTCCAC TTACAGCTGC AGTTACTGCT GCTGGTGCCA
551 ACGCAAGGTA AGGTAAAGGA CTTATTCTGT TAGTTAGTTT TGATTATTTT
601 AAGAATCGGT CTTGTACTGA TGCTTTTTAG TTGGGTTTGT TTACCAGTTA
651 TGTGCTTGAT GGGGTGCCTA GAATGAGGGA AAGACCTATA GGAGATTTGG
701 TTGTTGGTCT TAAGCAGCTT GGTGCTGATG TTGAATGTAC TCTTGGTACT
751 AACTGTCCTC CTGTTTCGTGT CAATGCTAAT GGTGGCCTGC CCGGTGGAAA
801 GGTGAGTTTG TAATTTTCAGC ATTTGCTATG TGAAAAGTTG CAGCAATCTT
851 TGTTTCATCAC ACTGCGTTAG CTTGACATGA TTTTAGCTTT TGTATGGTTT
901 CTTGATTGAC ACATTAGACA TGTTTTTGCA TTTTTCAGGT GAAGCTTTCT
951 GGATCAATCA GTAGTCAATA CTTGACTGCA CTGCTCATGG CAGCTCCCTT
1001 AGCTCTTGGA GACGTTGAGA TTGAGATCAT TGATAAATTG ATTTCTGTTC
1051 CATATGTTGA AATGACATTG AAGTTGATGG AACGTTTTGG TGTTAGTGCC
1101 GAGCATAGTG ACAGTTGGGA TCGTTTCTTT GTCAAGGGCG GTCAGAAATA
1151 CAAGTAAGAG TTGTTTCTAA AATCACTGAA CTTATAATTA GATTGACAGA
1201 AGAGTGACTA ACCAAATGGT AAAATTTGAT TCAGGTCGCC TGGAATGCT
1251 TACGTAGAAG GTGATGCTTC TAGTGCTAGT TATTTCTTGG CTGGTGCTGC
1301 CATTACTGGT GAAACCGTTA CTGTTGAAGG TTGTGGAACA ACCAGCCTGC
1351 AGGTAACACT AAGTTTATAA TAAAATTTGC TTAGTTCAAT TTTTTTTTGT
1401 CTTTCTAAGG CTTGGCTAGT TGTGTCACCT GTGTGTAACA TATGAAGAAT
1451 CTAAGTTTAG TTTTTTTTGG TGATGAATCT CAAAGGGAGA TGTGAAGTTC
1501 GCTGAGGTTT TTGAGAAAAT GGGATGTAAA GTGTCATGGA CAGAGAACAG
1551 TGTGACTGTG ACTGGACCAT C

```

图 9

(SEQ ID NO:11)

```

1   ATGGCGCAAG CTAGCAGAAT CTGCCAGAAC CCATGTGTGA TCTCCAATCT
51  CTCCAAATCA AACCAACGCA AATCGCCCTT GTCTGTCTCG ATGAAGACGC
101 ACCAGATATC TTCGTGGGGG TTGAAGAAGA GTAACAACGG CTCTGTGATT
151 CGTCCGGTTC GGGTAACGGC GTCTGTTTCC ACGGCTGAGA AATCTTCGGA
201 GATTGTGCTT CAGCCCATTA GAGAAATCTC GGGTCTGATC AAGCTACCCG
251 GACCCAAATC TCTGTCCAAT CGAATCCTTC TTCTAGCCGC TCTATCCGAG
301 GTCGGTGTGC TTCTTTCTTT CTTTGTTAGC TTAGTGTTGC GTTTTTAACG
351 GCGTGAGATT GAAGAAAGGT TCACACTTTG TTGTGGTGTG ATAGGGAACC
401 ACTGTAGTTG ACAACTTGTT GAACAGTGAT GACATCAATT ACATGCTTGA
451 TGCGTTGAAG AAATTGGGGC TTAATGTGGA ACGTGACAGT GAGAATAACC
501 GTGCGGTTGT TGAAGGATGT GGCGGGATAT TCCAGCTTC TTTAGATTCC
551 AAGAGTGATA TCGAGTTGTA CCTTGGGAAT GCTGGAACAG CCATGCGTCC
601 ACTTACCGCT GCAGTTACTG CTGCAGGTGG CAACGCAAGG TAAGGTTAAG
651 GAGTGTGATT TTGTTAGTTA GTTTTGTGTT ATGTCAAGAA CCGATCTTGT
701 CCTCATGCTT TTAGTTCGGT TTATTTTCCA GTTATATTCT TGGTGGGGTG
751 CCTAGAATGA GGGAAAGGCC TATTGGAGAT TTGGTTGTTG GTCTTAAGCA
801 GCTTGGTGCT GATGTTGAAT GTACTCTTGG AACTAACTGC CCTCCTGTTC
851 GCGTCAATGC TAATGGTGGC CTTCCCGGTG GAAAGGTGAG TTTGTAATCT
901 CAGCATCTAC TATGTGGAAA GTTGCAGGAA TTTTTGTTCA TCACACTGCG
951 TTTGCTCGAT ATGATGGCCT TTGTATGGTT TCTTGATTGA CATATTAGAT
1001 ATGATTTGCA TTTTTCAGGT GAAGCTATCT GGTTNAATCA GTAGTCAATA
1051 CTTGACTGCT CTGCTCATGG CAGCTCCTTT AGCTCTTGGG GACGTTGAGA
1101 TTGAGATCGT TGATAAACTG ATCTCTGTTC CGTATGTTGA AATGACATTG
1151 AAGTTGATGG AACGTTTTGG TGTTAGTGCC GAGCATAGTG ACAGTTGGGA
1201 TCGTTTCTTT GTCAAGGGCG GTCAGAAAATA CAAGTAAGCG TTGTTTCTGA
1251 AATCACTGAA CTTATAGTTA GATTGACAGA AGAGTACTA ACCAAATGGT
1301 AAAATTTGAT TCAGGTCGCC TGGAATGCT TACGTAGAAG GTGATGCTTC
1351 TAGTGCTAGT TATTTCTTGG CTGGTGCCGC CATTACTGGT GAGACTGTTA
1401 CTGTTGAAGG TTGTGGAACA ACCAGCCTGC AGGTAACACT AAGTTTATAA
1451 TGAAATTTGC TTAGTTCAAT TTGTTTTTTT GTCTTTCTAA GGCTTTGGCT
1501 AGTTATGTGT AACATATGTT AGAATCTAAG CTCATTTTTG TTGTTGTGAT
1551 GAATCTCAA AAGGAGATGT AAGTTCGCTG AGGTTCTTGA GAAAATGGGA
1601 TGTAAGTGT CATGGACAGA GAACAGTGTG ACTGTGACTG GACCATCTAG
1651 AGATGCTTTT GGAATGAGAC ACTTGCGCGC TGTTGATGTC AACATGAACA
1701 AAATGCCTGA TGTAGCCATG ACTCTTGCCG TTGTTGCTCT CTTTGCAGAT
1751 GGTCCAACCA CCATTAGAGA TGGAAGTAC TCCCTCTAAC CATCTAATTG
1801 AGGTTTTTAA GATTCATAGT CACTTAGTTC TCCTCTCATC CAATCGTTTT
1851 ATCATATATA GTGGCTAGCT GGAGAGTAAA GGAGACAGAA AGGATGATTG
1901 CCATTTGCAC AGAGCTTAGG AAGGTA AACATTTCTTT CTGTCCCGCT
1951 CTCACTCTCT TGTTTTATG TGCTCAGTCT AGGTTAAGTT CTGCATAACT
2001 TTTGCGTGCA GCTTGGAGCT ACAGTGGAAG AGGGTTCAGA TTATTGTGTG
2051 ATAACTCCAC CAGCAAAGCT GAAACCGGCG GAGATTGACA CATATGATGA
2101 TCATAGAATG GCAATGGCAT TCTCCCTTGC AGCTTGTGCT GATGTTCCAG
2151 TAACCATCAA AGATCCTGGT TGCACCAGGA AACTTTCC TACTACTTC
2201 CAGGTCTTG AAAGTATCAC AAAGCACTAA ACAGACCTTA AAGCCCATTT
2251 GTCTTTCTT TTTGATCCAA TTGGGATCAG TTTCTCTGT TATCACTGTA
2301 AGATTACGAA AAACAAAGAG TATTAAGATT GTTTGCTTGT ACCTTAAACT
2351 GTTTGATGCA ATCGTTGAAT CAGTTTTGGG CCAAGGGC

```

图 10

## (SEQ ID NO:12)

```
1   ATGGCGCAAG CTAGCAGAAT CTGCCATGGC GTGCAGAACC CATGTGTTAT
51  CATCTCCAAT CTCTCCAAT  CAAACCAAAA CAAATCACCT TTCTCCGTCT
101 CGCTGAAGAC GCAGCAGTCT CGAGCTTCTT CGTGGGGACT AAAGAAGAGT
151 GGAACGATGC TAAACGGTTC TGTAATTCGC CCGGTTAAGG TAACAGCTTC
201 CGTTTCCACG GCCGAGAAAG CTTCAGAGAT TGTGCTTCAA CCAATTAGAG
251 AAATCTCGGG TCTCATTAAG CTACCCGGAT CCAAATCTCT CTCCAATCGG
301 ATCCTCCTTC TTGCTGCTCT ATCTGAGGTA CATATACTTG ATTAGTGTTA
351 GGCCTTTGCT GTGAGATTTT GGGAACTATA GACAATTTAG TAAGAATTTA
401 TATATTATTT TTAAAAAATT AAAAGCCTAT ATATATATAT ATTTAAAATT
451 TTCAAAAAAT TATGGAGGTT TGAGACTGAA GAAAGTTTTT TTTTAATTAT
501 TATTATAGGG AACTACTGTA GTGGACAACT TGTGTAACAG TGATGACATC
551 AACTACATGC TTGATGCGTT GAAGAAGCTG GGGCTTAACG TGGAACGTGA
601 CAGGGTAAAC AACCGTGCTG TAGTTGAAGG ATGTGGTGGG ATATCCCAG
651 CTCCTTAGA TTCCAAGAGT GATATTGAGT TGTACCTTGG GAATGCAGGA
701 ACAGCCATGC GTCCACTCAC CGCTGCCGTT ACTGCTGCAG GTGGCAACGC
751 AAGGTAAAGG TTAAGGAGCT TTTTGTATT GTCAAGAAAT TGATTTTGTG
801 TTTGATGCTT TTAGTTTGGT TTGTTTTCTA GTTATGTGCT TGATGGGGTG
851 CCTAGAATGA GGGAGAGACC TATAGGAGAT TTGGTTGTTG GTCTTAAGCA
901 GCTTGGTGCT GATGTTGAAT GACTCTCGG CACTAACTGT CCTCCTGTTT
951 GTGTCAATGC TAATGGTGGC CTTCCCGGTG GAAAGGTGAT CTTGTTTGCA
1001 GCAGTCTTTG TTCATCACAG CCTTTGCTTC ACATTATTAC ATCTTTTAGT
1051 TTGTTGTTGT GACTTGATGG ATCTTAAAAA AAGGAATTGG GAACTGGTGT
1101 GAAAGTGATT AGCAATCTTT CTCGATTCCT TGCAGGGCCG TGGGCATTAC
1151 TAAGTGAAAC ATTAGCCTAT TAACCCCAA ATATTTTGAA AAAAATTTAG
1201 TATATGGCCC CAAAATAGTT TTTAAGAAAT TAGAAAAACT TTTAATAAAT
1251 CGTCTACGGT CCCCATTTTA GAGCCGACCC TGCTTGTATG GTTTCTTGAG
1301 TGAGATATTT TACATGTTTT GCATTTTCAG GTGAAGCTTT CTGGATCAAT
1351 CAGTAGTCAA TACTTGACTG CCTTGCTCAT GGCAGCTCCT TTAGCTCTTG
1401 GAGACGTGGA GATTGAGATC ATTGATAAAC TGATTTCTGT TCCATATGTT
1451 GAAATGACAT TGAAGTTGAT GGAACGTTTT GGTGTTAGTG CCGAGCATAG
1501 TGATAGCTGG GATCGTTTCT TTGTCAAGGG CGGTCAGAAG TACAAGTAAG
```

图 11A

```
1551 AATTCTTTAA ATTAAGAAT TAGATTGAAG AAAATGACTG ATTAACCAAA
1601 TGGCAAACCT GATTCAGGTC GCCTGGTAAT GCTTATGTAG AAGGTGATGC
1651 TTCTAGTGCT AGCTACTTCT TGGCTGGTGC TGCTATTACC GGTGAAACCG
1701 TCACTGTTGA AGGTTGTGGA ACAACTAGCC TCCAGGTAGT TTCTCCACTC
1751 TGAATCATCA AATATTATAC TCCCTCCGTT TTGTATTAAG TGTCATTTTA
1801 GCTTTTAAAT TTTGTCTCAT TAAAAGTGTC ATTTTACATT TTCAATGTAT
1851 ATATTAAATA AATTTTCCAG TTTTACTAA TTCATTATAT TAAATAATAT
1901 AAAACAGAAA ATTTAACAAAT TATCGTAATT CGTGTGCAAA GTTGATTAGT
1951 TCAAAGTTGT GTGTAACATG TTTTGAAGAA TCTAAGCTCA TTCTCTTTTT
2001 ATTTTTTTTG TGATGAATCC CAAAGGGAGA TGTGAAATTC GCAGAGGTAC
2051 TTGAGAAAAT GGGATGTAAA GTGTCATGGA CAGAGAACAG TGTGACTGTG
2101 ACTGGACCAT CTAGAGATGC TTTTGGAAATG AGACACTTGC GTGCTGTTGA
2151 TGTC AACATG AATAAAATGC CCGATGTAGC CATGACTCTT GCCGTTGTTG
2201 CTCTCTTTGC CGATGGTCCA ACCACCATCA GAGATGGTAA AGCAAAACCC
2251 TCTCTTTGAA TCAGCGTCTT TTAAGAGATT CATGGTTGCT TTAACTCTAT
2301 TTGGTCAATG TAGTGGCTAG CTGGAGAGTT AAGGAGACAG AAAGGATGAT
2351 AGCCATCTGC ACAGAGCTTC GAAAGGTAAG TTTCTTTTTC TCTCATGCTC
2401 TCATTCTAAG TTAATCGTTG CATAACTTTT TGGGGTTTTT TTTTTCGCTT
2451 CAGCTTGGAG CTACAGTGGA AGAAGGTTCA GATTATTGTG TGATAACTCC
2501 ACCAGCGAAG GTGAAACCGG CGGAGATTGA TACGTATGAT GATCATAGAA
2551 TGGCGATGGC GTTCTCGCTT GCAGCATGTG CTGATGTTCC AGTCACCATC
2601 AAGGATCCTG GCTGCACCAG AAAGACTTTC CCTGACTACT TTCAAGTCCT
2651 TGAAAGTATC ACAAAGCACT AAAAAGATCA TTTCTTTTGA ATCCAAATGT
2701 GAGAATGTGT TTCTTCCTCT CTCTGTGACC ACTGTAACAT TTATTAGAAG
2751 AACAAAGTGT GTGTGTTTAA GAGTGTGTTT GCTTGTAATG AACTGAGTGA
2801 GATGCAATCG TTGAATCAGT TTTGGGCCAA GGGC
```

图 11B

## (SEQ ID NO:13)

```

1   ATGGCGCAAG CTAGCAGAAT CTGCCATGGC GTGCAGAACC CATGTGTTAT
51  CATCTCCAAT CTCTCCAAAT CCAACCAAAA CAAATCACCT CTCTCCGTCT
101 CCTTGAAGAC GCATCAGCCT CGAGCTTCTT CGTGGGGATT GAAGAAGAGT
151 GGAACGACGC TAAACGGTTC TGTAATTCGC CCGGTTAAGG TAACAGCTTC
201 TGTTCACACG TCCGAGAAAG CTTACAGAGAT TGTGCTTCAA CCAATCAGAG
251 AAATCTCGGG TCTCATTAAAG CTACCCGGAT CCAAATCTCT CTCCAATCGG
301 ATCCTCCTTC TTGCCGCTCT ATCTGAGGTA CATATACTTG CTTAGTGTTA
351 GGCCTTTGCT GTGAGATTTT GGGAACTATA GACAATTTAG TAAGAATTTA
401 TATATAATTT TTTTAAAAAA AATCAGAAGC CTATATATAT TTAAATTTTT
451 CCAAATTTT TGGAGGTTAT AGGCTTGTGT TACACCATTC TAGTCTGCAT
501 CTTTCGGTTT GAGACTGAAG AATTTTATTT TTTAAAAAAT TATTATAGGG
551 AACTACTGTA GTGGACAAC TGTGTAACAG TGATGACATC AACTACATGC
601 TTGATGCGTT GAAGAAGCTG GGGCTTAACG TGGAACGTGA CAGTGTA AAC
651 AACCGTGC GG TTGTTGAAGG ATGCGGTGGA ATATTCCCAG CTTCTTAGA
701 TTCCAAGAGT GATATTGAGT TGTACCTTGG GAATGCAGGA ACAGCCATGC
751 GTCCACTCAC CGCTGCAGTT ACAGCTGCAG GTGGCAACGC GAGGTAAGGT
801 TAACGAGTTT TTTGTTATTG TCAAGAAATT GATCTTGTGT TTGATGCTTT
851 TAGTTTGTTT TGTTTTCTAG TTATGTACTT GATGGGGTGC CTAGAATGAG
901 GGAAAGACCT ATAGGAGATT TGTTTGTGG TCTTAAGCAG CTTGGTGCTG
951 ATGTTGAGTG TACTCTTGGC ACTAACTGTC CTCCTGTTCC TGTCATGCT
1001 AATGGTGGCC TTCCC GTGG AAAGGTGATC TTCACATTTA CTCTATGAAT
1051 TGTTTGCAGC AGTCTTTGTT CATCACAGCC TTTGCTTCAC ATTATTT CAT
1101 CTTTTAGTTT GTTGTATAT TACTTGATGG ATCTTTAAA AGGAATTGGG
1151 TCTGGTGTGA AAGTGATTAG CAATCTTTCT CGATTCCTTG CAGGGCCGTG
1201 GGCATTACTA AGTGAAACAT TAGCCTATTA ACCCCCAAAA TTTTGA AAA
1251 AAATTTAGTA TATGGCCCA AAATAGTTT TAAAAAATA GAAAACTTT
1301 TAATAAATCG TCTACAGTCC CAAAAATCTT AGAGCCGGCC CTGCTTGTAT
1351 GGTTCCTCGA TTGATATATT AGACTATGTT TTGAATTTTC AGGTGAAGCT
1401 TTCTGGATCG ATCAGTAGTC AGTACTTGAC TGCCCTCCTC ATGGCAGCTC
1451 CTTTAGCTCT TGGAGACGTG GAGATTGAGA TCATTGATAA ACTGATATCT
1501 GTCCATATG TTGAAATGAC ATTGAAGTTG ATGGAACGTC TTGGTGTTAG
1551 TGCCGAGCAT AGTGATAGCT GGGATCGTTT CTTTGTCAAG GGCGGTCAGA
1601 AGTACAAGTA AGAATCTTT AAATTAAGA ATTAGATTGA AGAAAATGAC
1651 TGATTAACCA AATGGCAAAA CTGATTCAGG TCGCCTGGTA ATGCTTATGT
1701 AGAAGGTGAT GCTTCTAGTG CTAGCTACTT CTTGGCTGGT GCTGCTATTA
1751 CCGGTGAAAC CGTCACTGTT GAAGGTTGTG GAACAAC TAG CCTCCAGGTA

```

图 12A



```
1801 GTTTCTCCAC TCTGAATCAT CAAATATTAT ACTCCCTCCG TTTTGTATTA
1851 AGTGTCAATTT TAGCTTTTAA ATTTTGTCTC ATTAAAAGTG TCATTTTACA
1901 TTTTCAATGT ATATATTAAA TAAATTTTCC AGTTTTTACT AATTCATTAT
1951 ATTAAATAAT ATAAAACAGA AAATTTAACA ATTATCGTAA TTCGTGTGCA
2001 AAGTTGATTA GTTCAAAGTT GTGTGTAACA TGTTTTGAAG AATCTAAGCT
2051 CATTCTCTTC TTATTTTTTT TGTGATGAAT CCCAAAGGGA GATGTGAAAT
2101 TCGCAGAGGT ACTTGAGAAA ATGGGATGTA AAGTGTCATG GACAGAGAAC
2151 AGTGTGACTG TGACTGGACC ATCTAGAGAT GCTTTTGGAA TGAGACACTT
2201 GCGTGCTGTT GATGTCAACA TGAATAAAAT GCCCGATGTA GCCATGACTC
2251 TTGCCGTTGT TGCTCTCTTT GCCGATGGTC CAACCACCAT CAGAGATGGT
2301 AAAGCAAAAC CCTCTCTTTG AATCAGCGTC TTTTAAAAGA TTCATGGTTG
2351 CTTTAACTCT ATTTGGTCAA TGTAGTGGCT AGCTGGAGAG TTAAGGAGAC
2401 AGAAAGGATG ATAGCCATCT GCACAGAGCT TCGAAAGGTA AGTTTCCTTT
2451 TCTCTCATGC TCTCATTCTA AGTTAATCGT TGCATAACTT TTTGGGGTTT
2501 TTTTTTGCCT TCAGCTTGGA GCTACAGTGG AAGAAGGTTT AGATTATTGT
2551 GTGATAACTC CACCAGCGAA GGTGAAACCG GCGGAGATTG ATACGTATGA
2601 TGATCATAGA ATGGCGATGG CGTTCTCGCT TGCAGCATGT GCTGATGTTT
2651 CAGTCACCAT CAAGGATCCT GGCTGCACCA GAAAGACTTT CCCTGACTAC
2701 TTTCAAGTCC TTGAAAGTAT CACAAAGCAC TAAAAAGATC ATTTCTTTT
2751 AATCCAAATG TGAGAATGTG TTTCTTCCTC TCTCTGTTGC CACTGTAACA
2801 TTTATTAGAA GAACAAAGTG TGTGTGTTTA AGAGTGTGTT TGCTTGTAAT
2851 GAACTGAGTG AGATGCAATC GTTGAATCAG TTTTGGGCC
```

图 12B

(SEQ ID NO:14)

```

1   ATGGCGCAAG CTAGCAGAAT CTGCCATGGC GTGCAGAACC CATGTGTTAT
51  CATCTCCAAT CTCTCCAAAT CCAACCAAAA CAAATCACCT TTCTCCGTCT
101 CCTTGAAGAC GCATCAGCCT CGAGCTTCTT CGTGGGGATT GAAGAAGAGT
151 GGAACGATGC TAAACGGTTC TGTAATTCGC CCGGTTAAGG TAACAGCTTC
201 TGTTTCCACG TCCGAGAAAG CTTCAGAGAT TGTGCTTCAA CCAATCAGAG
251 AAATCTCGGG TCTCATTAAAG CTACCCGGAT CCAAATCTCT CTCCAATCGG
301 ATCCTCCTTC TTGCCGCTCT ATCTGAGGTA CATATACTTG CTTAGTGTTA
351 GGCTTTTGCT GTGAGATTTT GGGAECTATA GACAATTTAG TAAGAATTTA
401 TATATAATTT TTTTAAAAAA AATCAGAAGC CTATATATAT TTAAATTTTT
451 CCAAATTTT TGGAGGTTAT AGGCTTATGT TACACCATTC TAGTCTGCAT
501 CTTTCGGTTT GAGACTGAAG AATTTTATTT TTTAAAAAAT TATTATAGGG
551 AACTACTGTA GTGGACAAC TGTGTAACAG TGATGACATC AACTACATGC
601 TTGATGCGTT GAAGAAGCTG GGGCTTAACG TGGAACGTGA CAGTGTAAC
651 AACCGTGCGG TTGTTGAAGG ATGCGGTGGA ATATTCCCAG CTTCTTAGA
701 TTCCAAGAGT GATATTGAGT TGTACCTTGG GAATGCAGGA ACAGCCATGC
751 GTCCACTCAC CGCTGCAGTT ACAGCTGCAG GTGGCAACGC GAGGTAAGGT
801 TAACGAGTTT TTTGTTATTG TCAAGAAATT GATCTTGTGT TTGATGCTTT
851 TAGTTTGGTT TGTTTTCTAG TGATGTACTT GATGGGGTGC CTAGAATGAG
901 GGAAAGACCT ATAGGAGATT TGGTTGTTGG TCTTAAGCAG CTTGGTGCTG
951 ATGTTGAGTG TACTCTTGGC ACTAACTGTC CTCCTGTTCC TGTC AATGCT
1001 AATGGTGGCC TTCCCGGTGG GAAGGTGATC TTCACATTTA CTCTATGAAT
1051 TGTTTGCAGC AGTCTTTGTT CATCACAGCC TTTGCTTCAC ATTATTTTAT
1101 CTTTTAGTTT GTTGTATAT TACTTGATGG ATCTTTAAAA AGGAATTGGG
1151 TCTGGTGTGA AAGTGATTAG CAATCTTTCT CGATTCCTTG CAGGGCCGTG
1201 GGCATTAATA AGTGAAACAT TAGCCTATTA ACCCCCAAAA TTTTGGAAAA
1251 AAATTTAGTA TATGGCCCCA AAATAGTTTT TAAAAAATTA GAAAACTTT
1301 TAATAAATCG TCTACAGTCC CAAAATCTT AGAGCCGGCC CTGCTTGTAT
1351 GGTTTCTCGA TTGATATATT AGACTATGTT TTGAATTTTC AGGTGAAGCT
1401 TCCTGGATCG ATCAGTAGTC AGTACTTGAC TGCCCTCCTC ATGGCAGCTC
1451 CTTTAGCTCT TGGAGACGTG GAGATTGAGA TCATTGATAA ACTGATATCT
1501 GTTCCATATG TTGAAATGAC ATTGAAGTTG ATGGAGCGTT TTGGTGTTAG
1551 TGCCGAGCAT AGTGATAGCT GGGATCGTTT CTTTGTC AAG GCGGTCAGA
1601 AATACAAGTA ATGAGTTCTT TTAAGTTGAG AGTTAGATTG AAGAATGAAT

```

图 13A

```

1651 GACTGATTAA CCAAATGGCA AACTGATTC AGGTCGCCTG GTAATGCTTA
1701 TGTAGAAGGT GATGCTTCTA GTGCTAGCTA CTTCTTGGCT GGTGCTGCCA
1751 TTAGTGGTGA AACTGTACT GTCGAAGGTT GTGGAACAAC TAGCCTCCAG
1801 GTAGTTTATC CACTCTGAAT CATCAAATAT TATACTCCCT CCGTTTTATG
1851 TTAAGTGTCA TTAGCTTTTA AATTTTGTTT CATTAAAAGT GTCATTTTAC
1901 ATTTTCAATG CATATATTAA ATAAATTTTC CAGTTTTTAC TAATTCATTA
1951 ATTAGCAAAA TCAAACAAAA ATTATATTAA ATAATGTAAA ATTCGTAATT
2001 TGTGTGCAAA TACCTTAAAC CTTATGAAAC GGAAACCTTA TGAAACAGAG
2051 GGAGTACTAA TTTTATAATA AAATTTGATT AGTTCAAAGT TGTGTATAAC
2101 ATGTTCTGTA AGAATCTAAG CTCATTCTCT TTTTATTTTT TGTGATGAAT
2151 CCCAAAGGGA GATGTGAAAT TCGCAGAGGT TCTTGAGAAA ATGGGATGTA
2201 AGGTGTCATG GACAGAGAAC AGTGTGACTG TGACTGGACC ATCAAGAGAT
2251 GCTTTTGGAA TGAGGCACTT GCGTGCTGTT GATGTCAACA TGAACAAAAT
2301 GCCTGATGTA GCCATGACTC TAGCCGTTGT TGCTCTCTTT GCCGATGGTC
2351 CAACCACCAT CAGAGATGGT AAAGCAAAAC CCTCTCTTTG AATCAGCGTG
2401 TTTTAAAAGA TTCATGGTTG CTTAAACTCT ATTTGGTCAA TGTAGTGGCT
2451 AGCTGGAGAG TTAAGGAGAC AGAGAGGATG ATTGCCATTT GCACAGAGCT
2501 TAGAAAGGTA AGTTTCCTTT TCTCTCATGC TCTCTCATTC GAAGTTAATC
2551 GTTGCATAAC TTTTTCGGT TTTTTTTTTT GCGTTCAGCT TGGAGCTACA
2601 GTGGAAGAAG GTTCAGATTA TTGTGTGATA ACTCCACCAG CAAAGGTGAA
2651 ACCGGCGGAG ATTGATACGT ATGATGATCA TAGAATGGCG ATGGCGTTCT
2701 CGCTTGCAAG TTGTGCTGAT GTTCCAGTCA CCATCAAGGA TCCTGGCTGC
2751 ACCAGGAAGA CTTTCCCTGA CTACTTCCAA GTCCTTGAAA GTATCACAAA
2801 GCATTAAAAG ACCCTTTCCT CTGATCCAAA TGTGAGAATC TGTGCTTTC
2851 TCTTTGTTGC CACCGTAACA TTTATTAGAA GAACAAAGTG TGTGTGTTAA
2901 GAGTGTGTTT GCTTGTAATG AACTGAGTGA GATGCAATCG TTGAATCAGT
2951 TTTGGGCC

```

图 13B

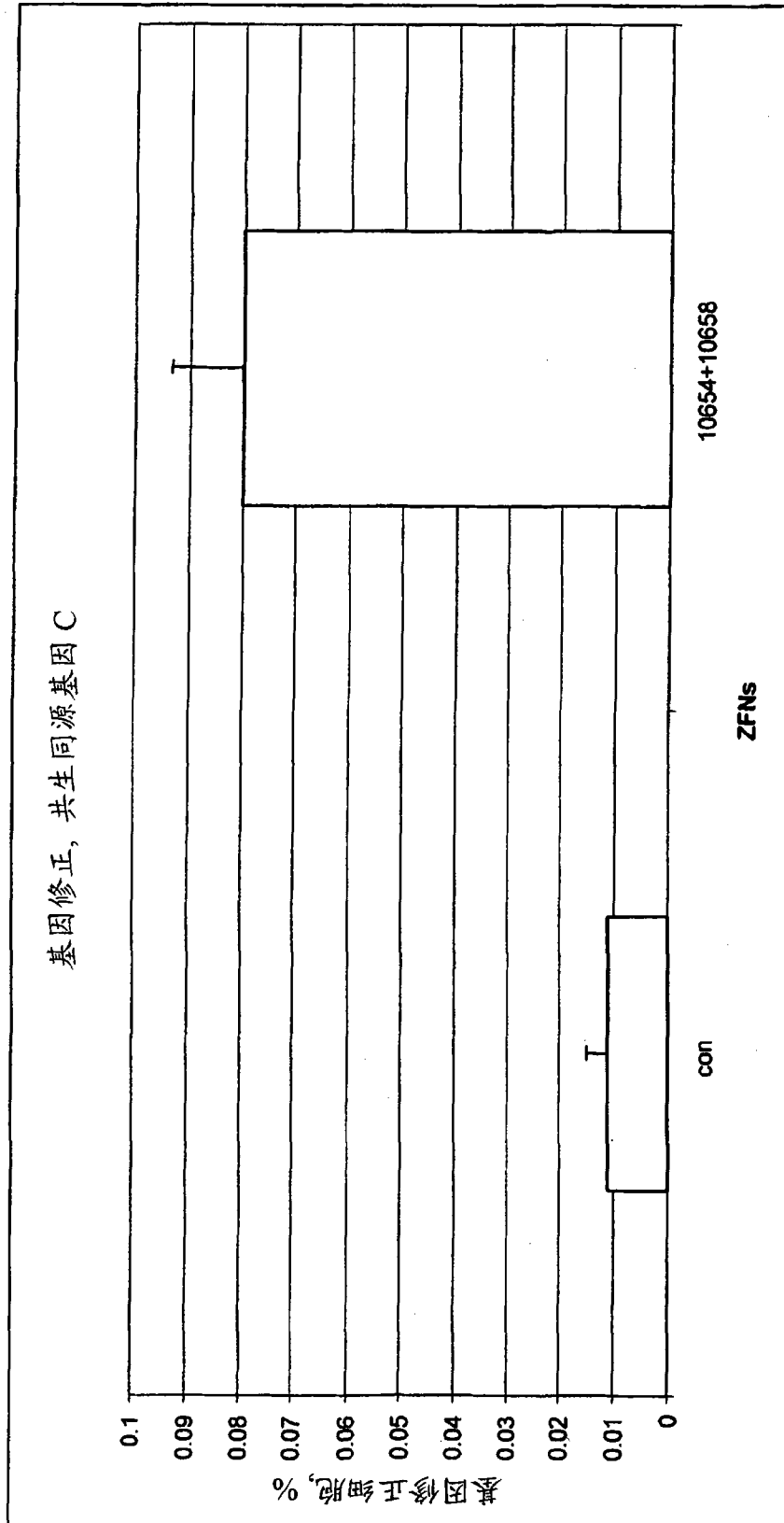


图 14A

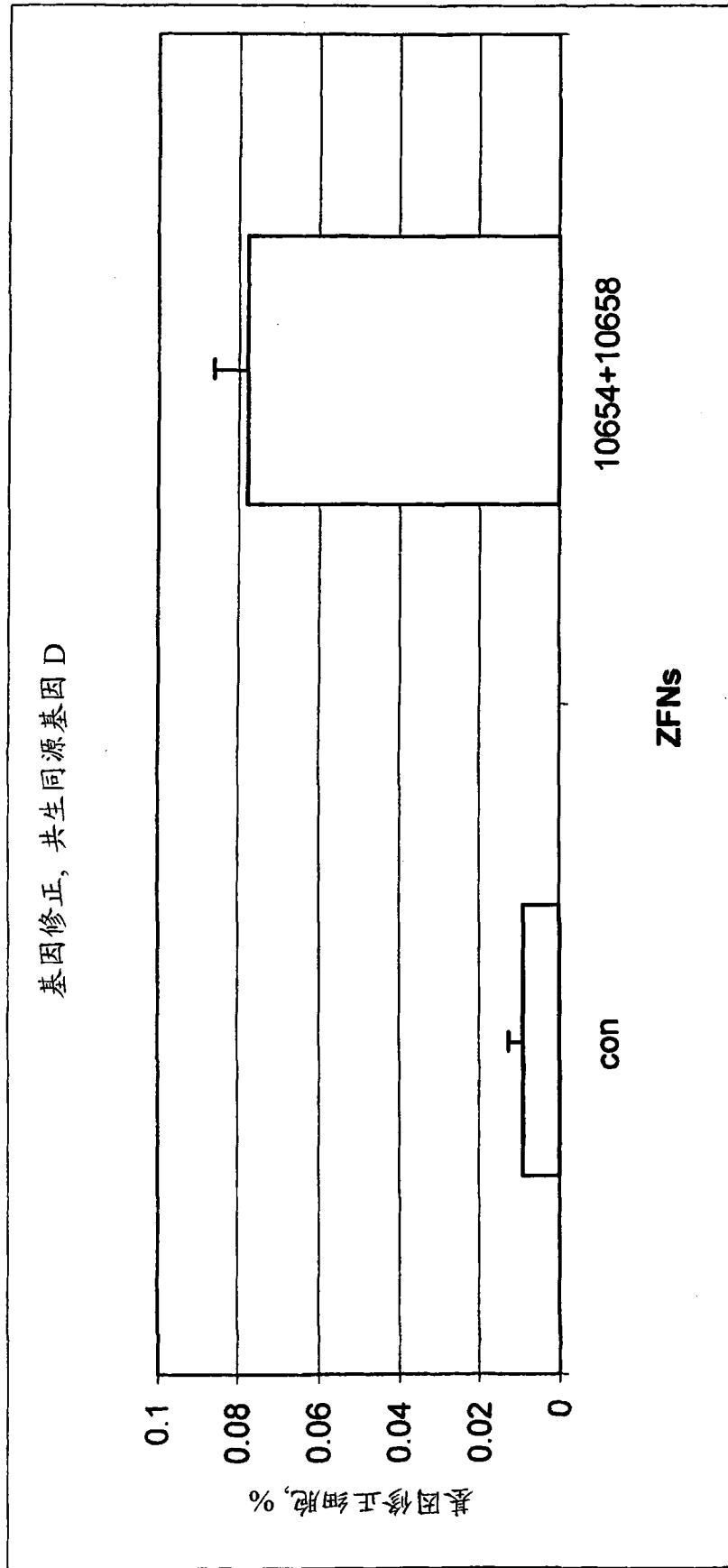


图 14B

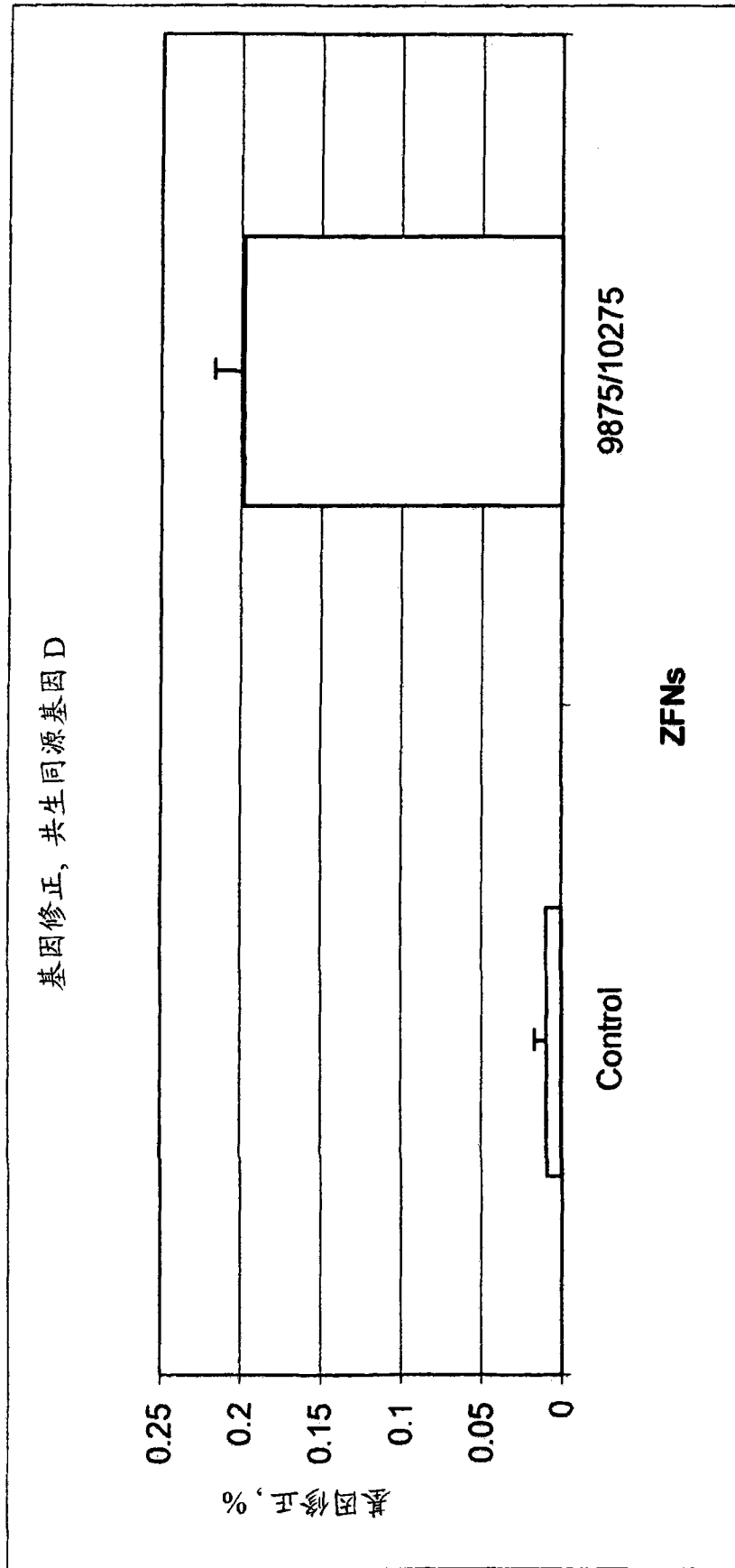
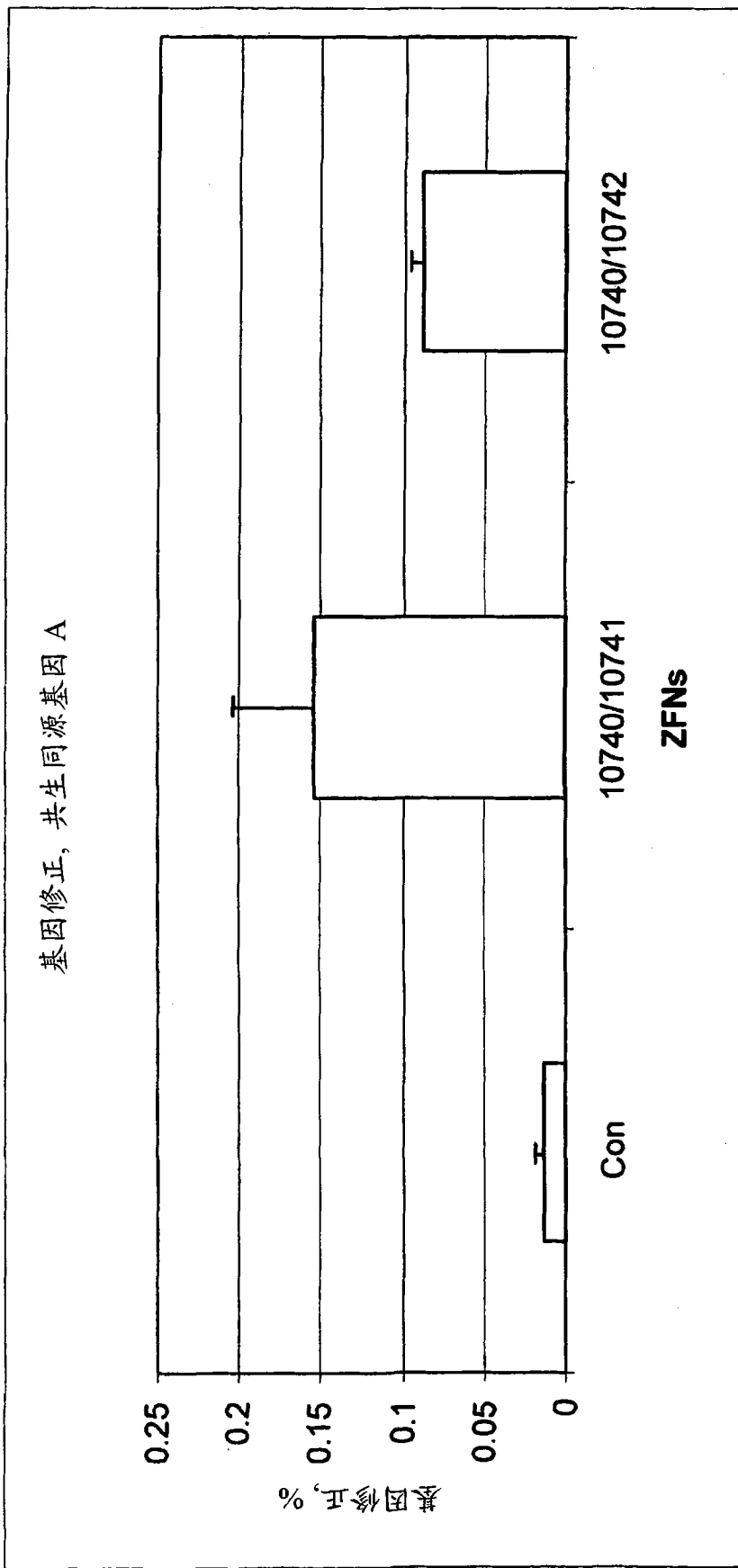


图 14C



#备注: 这些蛋白质未经 EPSPS 共生同源基因 B 报道细胞系测试

图 14D

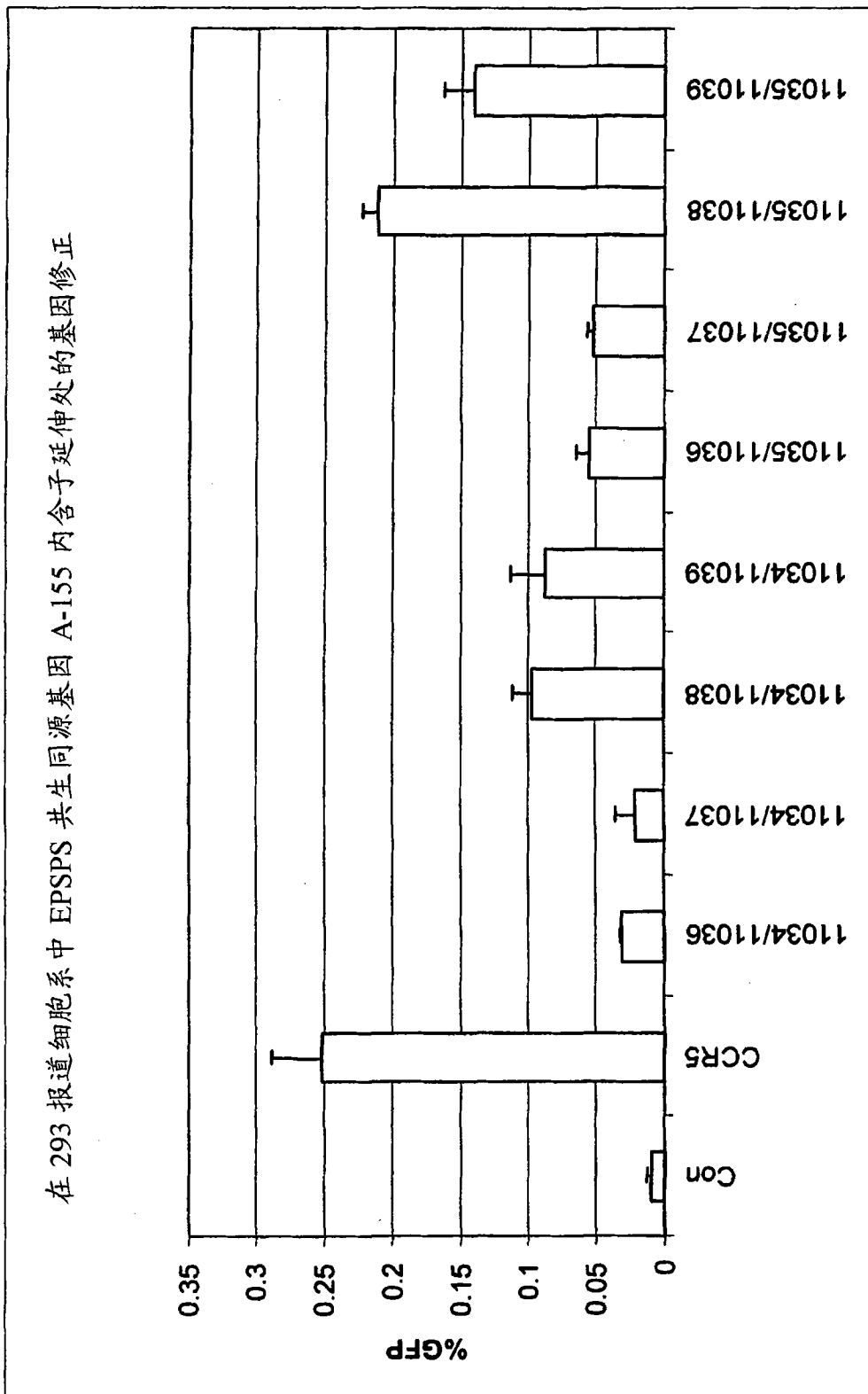


图 14E