



(12)发明专利申请

(10)申请公布号 CN 108138156 A

(43)申请公布日 2018.06.08

(21)申请号 201680060783.X

(74)专利代理机构 中国专利代理(香港)有限公司 72001

(22)申请日 2016.10.17

代理人 翟建伟 黄希贵

(30)优先权数据

62/243719 2015.10.20 US

62/309033 2016.03.16 US

62/359254 2016.07.07 US

(51)Int.Cl.

C12N 9/22(2006.01)

C12N 15/01(2006.01)

C12N 15/82(2006.01)

C12N 15/90(2006.01)

(85)PCT国际申请进入国家阶段日

2018.04.17

(86)PCT国际申请的申请数据

PCT/US2016/057272 2016.10.17

(87)PCT国际申请的公布数据

W02017/070029 EN 2017.04.27

(71)申请人 先锋国际良种公司

地址 美国依阿华州

权利要求书2页 说明书69页

序列表76页 附图12页

(72)发明人 A.M.西甘 S.斯维塔舍夫

(54)发明名称

用于无标记基因组修饰的方法和组合物

(57)摘要

提供了用于在不使用可选择标记的情况下修饰植物细胞的基因组中的核苷酸序列的组合物和方法。所述方法和组合物使用指导多核苷酸/Cas核酸内切酶系统在位于核苷酸序列中的靶位点中制造双链断裂并且在不使用可选择标记的情况下获得植物细胞,并且提供用于修饰植物、植物细胞或种子的基因组内的靶位点的有效系统。还提供了用于在不使用可选择标记的情况下产生在其基因组中具有修饰的核苷酸序列的植物细胞、愈伤组织或植物的组合物和方法。

核苷酸	PAM	评分	SPCIS IS1
CGGTTAAATATACGGGTACGGGTACGGT	↓		42
CGGTTAAATATACGGGTACGGGTACGGT		1.6	49
CGGTTAAATATACGGGTACGGGTACGGT		3.2	50
CGGTTAAATATACGGGTACGGGTACGGT		4.8	51
CGGTTAAATATACGGGTACGGGTACGGT		6.4	52
CGGTTAAATATACGGGTACGGGTACGGT		8.0	53
CGGTTAAATATACGGGTACGGGTACGGT		9.6	54
CGGTTAAATATACGGGTACGGGTACGGT		11.2	55
CGGTTAAATATACGGGTACGGGTACGGT		12.8	56
CGGTTAAATATACGGGTACGGGTACGGT		14.4	57
CGGTTAAATATACGGGTACGGGTACGGT		16.0	58
CGGTTAAATATACGGGTACGGGTACGGT		17.6	59
CGGTTAAATATACGGGTACGGGTACGGT		19.2	60
CGGTTAAATATACGGGTACGGGTACGGT		20.8	61
CGGTTAAATATACGGGTACGGGTACGGT		22.4	62
CGGTTAAATATACGGGTACGGGTACGGT		24.0	63
CGGTTAAATATACGGGTACGGGTACGGT		25.6	64
CGGTTAAATATACGGGTACGGGTACGGT		27.2	65
CGGTTAAATATACGGGTACGGGTACGGT		28.8	66
CGGTTAAATATACGGGTACGGGTACGGT		30.4	67
CGGTTAAATATACGGGTACGGGTACGGT		32.0	68
CGGTTAAATATACGGGTACGGGTACGGT		33.6	69
CGGTTAAATATACGGGTACGGGTACGGT		35.2	70
CGGTTAAATATACGGGTACGGGTACGGT		36.8	71
CGGTTAAATATACGGGTACGGGTACGGT		38.4	72
CGGTTAAATATACGGGTACGGGTACGGT		40.0	73
CGGTTAAATATACGGGTACGGGTACGGT		41.6	74
CGGTTAAATATACGGGTACGGGTACGGT		43.2	75
CGGTTAAATATACGGGTACGGGTACGGT		44.8	76
CGGTTAAATATACGGGTACGGGTACGGT		46.4	77
CGGTTAAATATACGGGTACGGGTACGGT		48.0	78
CGGTTAAATATACGGGTACGGGTACGGT		49.6	79
CGGTTAAATATACGGGTACGGGTACGGT		51.2	80
CGGTTAAATATACGGGTACGGGTACGGT		52.8	81
CGGTTAAATATACGGGTACGGGTACGGT		54.4	82
CGGTTAAATATACGGGTACGGGTACGGT		56.0	83
CGGTTAAATATACGGGTACGGGTACGGT		57.6	84
CGGTTAAATATACGGGTACGGGTACGGT		59.2	85
CGGTTAAATATACGGGTACGGGTACGGT		60.8	86
CGGTTAAATATACGGGTACGGGTACGGT		62.4	87
CGGTTAAATATACGGGTACGGGTACGGT		64.0	88
CGGTTAAATATACGGGTACGGGTACGGT		65.6	89
CGGTTAAATATACGGGTACGGGTACGGT		67.2	90
CGGTTAAATATACGGGTACGGGTACGGT		68.8	91
CGGTTAAATATACGGGTACGGGTACGGT		70.4	92
CGGTTAAATATACGGGTACGGGTACGGT		72.0	93
CGGTTAAATATACGGGTACGGGTACGGT		73.6	94
CGGTTAAATATACGGGTACGGGTACGGT		75.2	95
CGGTTAAATATACGGGTACGGGTACGGT		76.8	96
CGGTTAAATATACGGGTACGGGTACGGT		78.4	97
CGGTTAAATATACGGGTACGGGTACGGT		80.0	98
CGGTTAAATATACGGGTACGGGTACGGT		81.6	99
CGGTTAAATATACGGGTACGGGTACGGT		83.2	100

1. 一种用于在不使用可选择标记的情况下修饰植物细胞的基因组中的核苷酸序列的方法,所述方法包括:

将能够在位于所述核苷酸序列中的靶位点中制造双链断裂的指导RNA/Cas核酸内切酶复合物引入至少一个植物细胞中;并且,

选择在所述核苷酸序列中具有修饰的植物细胞,其中所述选择在不使用可选择标记的情况下发生。

2. 一种用于在不使用可选择标记的情况下产生在其基因组中具有修饰的核苷酸序列的植物的方法,所述方法包括:

将能够在位于核苷酸序列中的靶位点中制造双链断裂的指导RNA/Cas核酸内切酶复合物引入至少一个植物细胞中;

从所述植物细胞获得植物;并且,

选择在所述核苷酸序列中具有修饰的植物,其中所述选择在不使用可选择标记的情况下发生。

3. 一种用于在不使用可选择标记的情况下产生在其基因组中具有修饰的核苷酸序列的植物愈伤组织的方法,所述方法包括:

将能够在位于核苷酸序列中的靶位点中制造双链断裂的指导RNA/Cas核酸内切酶复合物引入至少一个植物细胞中;

从所述植物细胞获得愈伤组织;并且,

选择在所述核苷酸序列中具有修饰的愈伤组织,其中所述选择在不使用可选择标记的情况下发生。

4. 如权利要求1所述的方法,其中所述修饰选自下组,该组由以下各项组成:所述靶位点中至少一个核苷酸的插入、至少一个核苷酸的缺失、或至少一个核苷酸的取代。

5. 如权利要求1所述的方法,所述方法进一步包括将多核苷酸修饰模板引入所述植物细胞中,其中所述多核苷酸修饰模板包含所述核苷酸序列的至少一个核苷酸修饰。

6. 如权利要求3所述的方法,其中所述多核苷酸修饰模板的至少一个核苷酸修饰选自下组,该组由以下各项组成:(i)至少一个核苷酸的替代、(ii)至少一个核苷酸的缺失、(iii)至少一个核苷酸的插入、和(iv) (i)-(iii)的任何组合。

7. 如权利要求1所述的方法,所述方法进一步包括将供体DNA引入(a)的植物细胞中,其中所述供体DNA包含至少一个待插入所述靶位点中的目的多核苷酸。

8. 如权利要求1所述的方法,其中所述引入不包括将可选择标记引入所述细胞中。

9. 如权利要求1所述的方法,其中所述引入不包括将破坏的可选择标记基因恢复成编码功能性可选择标记蛋白的非破坏的可选择标记基因。

10. 如权利要求1所述的方法,其中所述引入不导致在所述细胞内产生可选择标记。

11. 如权利要求1所述的方法,其中所述选择不包括可选择标记的鉴定或使用。

12. 如权利要求1所述的方法,其中所述选择经由对所述植物的DNA进行测序来发生。

13. 如权利要求1所述的方法,其中将所述指导RNA/Cas核酸内切酶复合物作为核糖核苷酸-蛋白质引入。

14. 如权利要求1所述的方法,其中将所述指导RNA/Cas核酸内切酶复合物的组分作为指导RNA和Cas核酸内切酶蛋白引入,所述指导RNA和所述Cas核酸内切酶蛋白能够形成所述

指导RNA/Cas核酸内切酶复合物。

15. 如权利要求1所述的方法,其中将所述指导RNA/Cas核酸内切酶复合物的组分作为编码所述Cas核酸内切酶的mRNA和作为包含指导RNA的RNA引入。

16. 如权利要求1所述的方法,其中将所述指导RNA/Cas核酸内切酶复合物的组分作为编码指导RNA和Cas核酸内切酶蛋白的重组DNA分子引入。

17. 如权利要求1所述的方法,其中所述指导RNA/Cas核酸内切酶复合物在所述细胞内部组装。

18. 如权利要求13所述的方法,其中将所述核糖核苷酸-蛋白质包被到颗粒递送基质上或与颗粒递送基质组合以形成核糖核苷酸-蛋白质-基质复合物,其中将所述核糖核苷酸-蛋白质-基质复合物引入所述细胞中。

19. 如权利要求1所述的方法,其中所述植物细胞是体细胞胚胎细胞。

20. 如权利要求1所述的方法,其中所述植物细胞不是原生质体。

21. 如权利要求1所述的方法,其中所述植物细胞选自下组,该组由以下各项组成:单子叶植物和双子叶植物细胞。

22. 如权利要求1所述的方法,其中所述植物细胞选自下组,该组由以下各项组成:玉蜀黍、稻、高粱、黑麦、大麦、小麦、粟、燕麦、甘蔗、草坪草、或柳枝稷、大豆、低芥酸菜籽、苜蓿、向日葵、棉花、烟草、花生、马铃薯、番茄、烟草、拟南芥和红花细胞。

23. 如权利要求1所述的方法,所述方法进一步包括从所述植物细胞再生植物。

24. 一种通过如权利要求23所述的方法产生的植物。

25. 一种如权利要求24所述的植物的后代植物,其中所述后代植物没有选自下组的任何组分,该组由以下各项组成:指导RNA、Cas核酸内切酶、多核苷酸修饰模板和供体DNA。

## 用于无标记基因组修饰的方法和组合物

[0001] 本申请要求于2015年10月20日提交的美国临时申请号62/243719、于2016年3月16日提交的美国临时申请号62/309033以及于2016年7月7日提交的美国临时申请号62/359254的权益,这些申请通过引用以其全文结合在此。

### 技术领域

[0002] 本披露涉及分子生物学领域,具体涉及用于改变细胞基因组的方法。

### [0003] 以电子方式提交的序列表的引用

[0004] 该序列表的官方副本经由EFS-Web作为ASCII格式的序列表以电子方式提交,文件名为20161011\_7158PCT\_SeqLs.txt,创建于2016年10月11日,且具有185千字节大小,并与本说明书同时提交。包含在该ASCII格式的文件中的序列表是说明书的一部分并通过引用以其全文结合在此。

### 背景技术

[0005] 重组DNA技术使得可以在靶标基因组位置处插入DNA序列和/或修饰(编辑)特异性内源染色体序列,从而改变生物体的表型。已经使用了采用位点特异性重组系统的位点特异性整合技术以及其他类型的重组技术来在各种生物体中产生目的基因的靶向插入。基因组编辑技术例如设计师的锌指核酸酶(ZFN)或转录激活子样效应子核酸酶(TALEN)或归巢大范围核酸酶可以用于产生靶向基因组干扰,但这些系统倾向于具有低特异性并且使用需要对每个靶位点进行重新设计的经设计的核酸酶,这使得它们的制备成本高昂且耗时。

[0006] 尽管已经开发了若干种方法来靶向在生物体基因组中的特定位点用于修饰,但仍然需要可负担的、易于建立的、可扩展的、并且适于靶向生物体基因组中的多个位置的新的基因组工程技术。

### 发明内容

[0007] 提供了用于在不使用可选择标记的情况下修饰植物细胞的基因组中的核苷酸序列的组合物和方法。所述方法和组合物使用指导多核苷酸/Cas核酸内切酶系统在位于核苷酸序列中的靶位点中制造双链断裂并且在不使用可选择标记的情况下获得植物细胞,并且提供用于修饰植物、植物细胞或种子的基因组内的靶位点的有效系统。还提供了用于在不使用可选择标记的情况下产生在其基因组中具有修饰的核苷酸序列的植物细胞、愈伤组织或植物的组合物和方法。

[0008] 在本披露的一个实施例中,该方法包括用于在不使用可选择标记的情况下修饰植物细胞的基因组中的核苷酸序列的方法,所述方法包括将能够在位于所述核苷酸序列中的靶位点中制造双链断裂的指导RNA/Cas核酸内切酶复合物引入至少一个植物细胞中,并且选择在所述核苷酸序列中具有修饰的植物细胞,其中所述选择在不使用可选择标记的情况下发生。

[0009] 在本披露的一个实施例中,该方法包括用于在不使用可选择标记的情况下产生在

其基因组中具有修饰的核苷酸序列的植物的方法,所述方法包括:将能够在位于核苷酸序列中的靶位点中制造双链断裂的指导多核苷酸/Cas核酸内切酶复合物引入至少一个植物细胞中;从所述植物细胞获得植物;并且,选择在所述核苷酸序列中具有修饰的植物,其中所述选择在不使用可选择标记的情况下发生。

[0010] 在本披露的一个实施例中,该方法包括用于在不使用可选择标记的情况下产生在其基因组中具有修饰的核苷酸序列的植物愈伤组织的方法,所述方法包括:将能够在位于核苷酸序列中的靶位点中制造双链断裂的指导多核苷酸/Cas核酸内切酶复合物引入至少一个植物细胞中;从所述植物细胞获得愈伤组织;选择在所述核苷酸序列中具有修饰的愈伤组织,其中所述选择在不使用可选择标记的情况下发生。该指导多核苷酸/Cas核酸内切酶复合物可以是指导RNA/Cas核酸内切酶复合物。

[0011] 本披露的方法可以进一步包括引入多核苷酸模板,其中所述多核苷酸修饰模板包含在所述细胞的基因组中含有靶位点的核苷酸序列的至少一个核苷酸修饰,其中所述多核苷酸修饰模板的所述至少一个核苷酸修饰选自下组,该组由以下各项组成:(i)至少一个核苷酸的替代、(ii)至少一个核苷酸的缺失、(iii)至少一个核苷酸的插入、以及(iv) (i)-(iii)的任何组合。该方法还可以进一步包括引入供体DNA,其中所述供体DNA包含至少一个目的多核苷酸。可以经由本领域已知的任何方法将该指导多核苷酸/Cas核酸内切酶复合物引入细胞中,所述方法例如但不限于选自下组的递送系统,该组由以下各项组成:颗粒介导的递送、晶须介导的递送、细胞穿透肽介导的递送、电穿孔、PEP介导的转染和纳米颗粒介导的递送。

[0012] 可以在不使用表型标记或可选择标记并且不应用选择试剂的情况下使用本披露的方法。这些方法包括在也不将可选择标记引入植物细胞中的情况下将指导多核苷酸/Cas核酸内切酶复合物引入所述植物细胞中,或其中引入所述指导多核苷酸/Cas核酸内切酶复合物不伴随着将破坏的可选择标记基因恢复成编码功能性可选择标记蛋白的非破坏的可选择标记基因,或者其中引入所述指导多核苷酸/Cas核酸内切酶复合物不导致在所述细胞内产生可选择标记。本披露的方法可以进一步包括选择在其基因组中包含修饰的核苷酸序列的植物细胞、愈伤组织或植物,其中所述选择在不使用可选择标记的情况下发生。

[0013] 还提供了通过本文所述的方法产生的核酸构建体、细胞、植物、后代植物、微生物、外植体、种子和谷粒。在本文中示出了本披露的这些方法和组合物的另外的实施例。

[0014] 附图和序列表的说明

[0015] 根据下列的详细描述和附图以及序列表,可以更全面地理解本披露,所述详细描述和附图以及序列表形成本申请的一部分。这些序列描述以及所附序列表遵守如37C.F.R. §§1.821-1.825所列出的管理专利申请中核苷酸和氨基酸序列披露内容的规则。这些序列描述包含如在37C.F.R. §§1.821-1.825中所定义的用于氨基酸的三字母代码,将其通过引用结合在此。

[0016] 附图

[0017] 图1描绘了由本文所述的玉蜀黍优化的指导RNA/Cas核酸内切酶系统诱导的最常见的前10种的NHEJ突变的比对和计数。通过深度测序来鉴定这些突变。参考序列(SEQ ID NO:48)表示未修饰的基因座,每个靶位点以粗体显示。还指出了PAM序列(灰色)和预期切割位点(箭头)。由于不完美的NHEJ导致的缺失或插入分别以“-”或斜体加下划线的核苷酸显

示。靶位点的参考和突变1-10分别对应于SEQ ID NO:49-58。在玉蜀黍中,对于大多数靶位点,由Cas9-gRNA系统产生的突变的最普遍类型是单个核苷酸插入( $\geq 60\%$ ) (计数显示为16,861)。

[0018] 图2描绘了ALS2基因的部分核苷酸序列(SEQ ID NO:59、61、63)和部分氨基酸序列(SEQ ID NO:60、62、64)以及两个编辑修复模板(Oligo1和Oligo2);修饰的核苷酸被加下划线,并且靶向用于基因编辑的密码子序列(Pro至Ser)被加框。

[0019] 图3描绘了具有经编辑的用于抗氯磺隆的ALS2等位基因的玉蜀黍植物(左)和野生型植物(右)。向四周龄的植物喷洒氯磺隆(100mg/L)。显示了处理后三周的植物。

[0020] 图4A-4C显示选择用于修饰的ALS2基因片段(SEQ ID NO:65、67、69)和使用ALS2作为可选择标记的示意图。编码的氨基酸序列显示在每个核苷酸序列的下方。(SEQ ID NO:66、68)。图4A:可以去除在位置165处的单个核苷酸(G)(粗体并加下划线),以便产生氯磺隆抗性ALS2基因的经编辑的特定敲除版本。图4B描绘了具有导致翻译移码和ALS2基因敲除的单个核苷酸缺失(G被去除)的新核苷酸序列。图4C:在经由NHEJ途径进行DSB修复的过程中通过插入单个核苷酸(N,粗体并加下划线)恢复ALS2基因功能和氯磺隆抗性。

[0021] 图5A-5B.作为可选择标记的失活的ALS2P165S的重新激活。图5A(SEQ ID NO:70)。ALS 2P165S基因的设计,其含有位于PAM的5'的3个核苷酸处的上游框外翻译起始密码子。在第一个AUG处的翻译的起始(通过箭头下面序列所描绘)编码一种4个氨基酸的多肽,该多肽阻止ALS2的翻译起始密码子(灰色字母)的起始。图5B(SEQ ID NO:71)。导致上游AUG缺失的单个核苷酸插入(C、A或T)或缺失(或任何组合)允许在ALS2的起始密码子处开始翻译(通过箭头下面序列所描绘),从而恢复全长ALS2P165S除草剂抗性基因的翻译。

[0022] 图6A-6C显示包含选择用于修饰的内源靶位点的目的多核苷酸片段(SEQ ID NO:72)的示意图。编码的氨基酸序列显示在每个核苷酸序列的下方。(SEQ ID NO:73、75、77)。图6A描绘了可以通过NHEJ去除位于核酸内切酶切割位点(由箭头所示)旁边的单个核苷酸(在该实例中为C,以粗体并加下划线显示)。图6B描绘了具有导致产生新的切割位点(由箭头表示)和翻译移码的单个碱基缺失的所得的目的多核苷酸(SEQ ID NO:74)。图6C:可以在不使用多核苷酸修饰(修复)模板的情况下通过NHEJ插入位于核酸内切酶切割位点旁边的单个核苷酸(在该实例中为T,以粗体并加下划线显示),导致目的多核苷酸的单个核苷酸编辑(SEQ ID NO:76)。PAM序列以灰色突出显示。

[0023] 图7:顶图:用于将UBI:Cas9稳定整合到玉蜀黍基因组中的农杆菌(Agrobacterium)载体。底图:用于将MDH:Cas9稳定整合到玉蜀黍基因组中的农杆菌载体。MDH是温度调节型启动子,调节Cas9的表达。这些载体还含有可视标记基因(END2:AmCYAN),其被用于选择稳定转化的愈伤组织部分。红色荧光蛋白(DsRED)的序列包含由343-bp的间隔子分开的同向重复的369bp片段,其包含用于通过两种gRNA和LIG3:4大范围核酸酶进行识别和靶向的序列。H2B是指组蛋白H2B基因启动子。

[0024] 序列

[0025] 表1.核酸和蛋白质SEQ ID号的汇总

[0026]

描述	核酸 SEQ ID NO.	蛋白质 SEQ ID NO.
Cas9 编码序列	1	
马铃薯 ST-LS1 内含子	2	
SV40 NLS		3
VirD2 NLS	4	
玉蜀黍优化的 Cas9 表达盒	5	
Lig-CR3 指导 RNA 表达载体	6	
玉蜀黍基因组靶位点 MS26Cas-1 加 PAM 序列	7	
玉蜀黍基因组靶位点 MS26Cas-2 加 PAM 序列	8	
玉蜀黍基因组靶位点 MS26Cas-3 加 PAM 序列	9	
玉蜀黍基因组靶位点 LIGCas-1 加 PAM 序列	10	
玉蜀黍基因组靶位点 LIGCas-2 加 PAM 序列	11	
玉蜀黍基因组靶位点 LIGCas-3 加 PAM 序列	12	
玉蜀黍基因组靶位点 MS45Cas-1 加 PAM 序列	13	
玉蜀黍基因组靶位点 MS45Cas-2 加 PAM 序列	14	
玉蜀黍基因组靶位点 MS45Cas-3 加 PAM 序列	15	
玉蜀黍基因组靶位点 ALSCas-1 加 PAM 序列	16	
玉蜀黍基因组靶位点 ALSCas-2 加 PAM 序列	17	
玉蜀黍基因组靶位点 ALSCas-3 加 PAM 序列	18	
引物序列	19-38	
ALS1-DNA 序列	39	
ALS2-DNA 序列	40	
全长 Zm-ALS2 蛋白质		41
玉蜀黍基因组靶位点 ALSCas-4 加 PAM 序列	42	
794 bp 多核苷酸修饰模板	43	
127 bp 多核苷酸修饰模板, oligo1	44	
127 bp 多核苷酸修饰模板, oligo2	45	
含有玉蜀黍密码子优化的 Cas9 和玉蜀黍 UBI 启动子的农杆菌	46	

[0027]

载体		
含有玉蜀黍密码子优化的 Cas9 和玉蜀黍 MDH 启动子的农杆菌载体	47	
示于图 1 中的序列	48-58	
示于图 2 中的序列	59、61、63	60、62、64
示于图 4A-4C 中的序列	65、67、69	66、68
示于图 5A-5B 中的序列	70-71	
示于图 6A-6C 中的序列	72、74、76	73、75、77
IN2 启动子	78	
ALSCas7 靶位点	79	
ALSCas7-1 靶位点，其是经修饰的 ALSCas7 靶位点	80	
玉蜀黍脱靶位点	81	

### 具体实施方式

[0028] 提供了用于在不使用可选择标记的情况下修饰植物细胞的基因组中的核苷酸序列的组合物和方法。所述方法和组合物使用指导多核苷酸/Cas核酸内切酶系统在位于待修饰的核苷酸序列中的靶位点中制造双链断裂并且在不使用可选择标记的情况下获得植物细胞。还提供了用于在不使用可选择标记的情况下产生在其基因组中具有修饰的核苷酸序列的植物细胞、愈伤组织或植物的组合物和方法。本文所述的方法可以进一步包括将能够在位于核苷酸序列中的靶位点中制造双链断裂的指导多核苷酸/Cas核酸内切酶复合物引入至少一个植物细胞中，并且选择在所述核苷酸序列中具有修饰的植物细胞，其中所述选择在不使用可选择标记的情况下发生。该方法可以进一步包括从植物细胞获得愈伤组织或植物并且选择在所述核苷酸序列中具有修饰的愈伤组织或植物，其中所述选择在不使用可选择标记的情况下发生。

[0029] 本领域技术人员可以理解，常规的基因组修饰方法和植物转化方法主要依靠将可选择标记基因引入待修饰的细胞中以允许选择方案，在这些选择方案中，例如使用抗生素或除草剂(选择试剂)来抑制或杀死不包含该可选择标记基因的细胞或组织，并且由于该可选择标记(抗性)基因的表达，包含该可选择标记基因的细胞或组织继续生长。相比之下，可以在不使用可选择标记并且不应用选择试剂的情况下使用本披露的方法。

[0030] CRISPR基因座(规律间隔成簇短回文重复序列)(又称SPIDR--间隔区散在同向重复序列)构成DNA基因座的家族。CRISPR基因座由部分回文的短而高度保守的DNA重复序列(通常为24至40bp,重复从1至140次,也称为CRISPR重复序列)组成。重复序列(通常具有物种特异性)由恒定长度的可变序列(通常为20至58,依赖于CRISPR基因座(WO 2007/025097, 2007年3月1日公开))间隔。细菌和古细菌已经发展了适应性免疫防御,称为使用短RNA直接降解外来核酸的规律间隔成簇短回文重复序列(CRISPR)/CRISPR相关(Cas)系统(WO2007/025097, 2007年3月1日公开)。已经描述了包括具有多亚基效应子复合物的1类系统和具有单一蛋白质效应子(例如但不限于Cas9、Cpf1、C2c1、C2c2、C2c3)的2类系统的多重CRISPR-Cas系统。(Zetsche等人, 2015, Cell [细胞] 163, 1-13; Shmakov等人, 2015, Molecular Cell



[分子细胞]60,1-13;Makarova等人,2015,Nature Reviews Microbiology[自然评论-微生物]第13卷:1-15;于2013年11月23日公开的WO 2013/176772A1,并将其通过引用以其全文结合在此)。

[0031] 来自细菌的II型CRISPR/Cas系统使用crRNA (CRISPR RNA) 和tracrRNA (反式激活CRISPR RNA) 来将Cas核酸内切酶指导到其DNA靶标上。该crRNA包含与双链DNA靶标的一条链互补的间隔子区域和与tracrRNA (反式激活CRISPR RNA) 碱基配对的区域,该tracrRNA形成指导Cas核酸内切酶切割DNA靶标的RNA双链体。通过未完全理解的涉及Cas1和Cas2蛋白的过程获得间隔子。除cas9基因之外,所有的II型CRISPR-Cas基因座包含cas1和cas2基因 (Makarova等人2015,Nature Reviews Microbiology[自然评论-微生物]第13卷:1-15)。Cas基因包括通常与侧翼CRISPR基因座偶合、相关或相近或相邻的基因。术语“Cas基因”、“CRISPR相关的 (Cas) 基因”在本文中可互换地使用。对Cas蛋白家族的全面综述呈现于以下文献中:Haft等人 (2005) Computational Biology[计算生物学],PLoS Comput Biol[科学公共图书馆计算生物学]1 (6) :e60.doi:10.1371/journal.pcbi.0010060。如其中所述,除了四个先前已知的基因家族之外,还描述了41个CRISPR相关的 (Cas) 基因家族。它表明CRISPR系统属于不同类别,具有不同的重复模式、基因的组和种类范围。在给定CRISPR基因座处的Cas基因的数目可以在物种之间变化 (Han等人,2005,Computational Biology[计算生物学],PLoS Comput Biol[科学公共图书馆计算生物学]1 (6) :e60.doi:10.1371/journal.pcbi.0010060;Makarova等人,2015,Nature Reviews Microbiology[自然评论-微生物]第13卷:1-15;2013年11月23日公开的WO 2013/176772 A1,并将其通过引用以其全文结合在此)。

[0032] 本文中术语“Cas核酸内切酶”是指由Cas (CRISPR-相关的) 基因编码的蛋白质。当与适合的多核苷酸组分复合时,Cas核酸内切酶能够识别、结合特定DNA靶序列的全部或部分、并任选地使特定DNA靶序列的全部或部分产生切口或切割特定DNA靶序列的全部或部分。本文描述的Cas核酸内切酶包含一个或多个核酸酶结构域。本披露的Cas核酸内切酶包括具有HNH或HNH样核酸酶结构域和/或RuvC或RuvC样核酸酶结构域的那些 (Makarova等人,2015,Nature Reviews Microbiology[自然评论-微生物]第13卷:1-15)。Cas包括Cas9蛋白、Cpf1蛋白、C2c1蛋白、C2c2蛋白、C2c3蛋白、Cas3、Cas3-HD、Cas5、Cas7、Cas8、Cas10或这些的复合物。

[0033] 如本文所用,术语“指导多核苷酸/Cas核酸内切酶复合物”、“指导多核苷酸/Cas核酸内切酶系统”、“指导多核苷酸/Cas复合物”、“指导多核苷酸/Cas系统”、“指导Cas系统”、“PGEN”在本文中可互换地使用,并且是指能够形成多核苷酸-蛋白质复合物的至少一种指导多核苷酸和至少一种Cas核酸内切酶蛋白,其中所述指导多核苷酸/Cas核酸内切酶复合物可以将Cas核酸内切酶引导至DNA靶位点,使Cas核酸内切酶能够识别、结合到DNA靶位点、并任选地使DNA靶位点产生切口或切割 (引入单链或双链断裂) DNA靶位点。本文中指导多核苷酸/Cas核酸内切酶复合物可以包含四种已知的CRISPR系统 (Horvath和Barrangou, Science[科学]327:167-170) (例如I型、II型或III型CRISPR系统) 中任一种的一种或多种Cas蛋白和一种或多种合适的多核苷酸组分。Cas核酸内切酶在靶序列处解开DNA双链体并任选地切割至少一条DNA链,如通过由与Cas蛋白复合的多核苷酸 (例如但不限于crRNA或指导RNA) 识别靶序列所介导的。如果正确的前间区序列邻近基序 (PAM) 位于或相邻于DNA靶序

列的3'末端,则通过Cas核酸内切酶对靶序列进行的此类识别和切割通常会发生。可替代地,本文中的Cas蛋白可能缺乏DNA切割或切口活性,但是当与合适的RNA组分复合时,仍然可以特异性结合DNA靶序列。(还参见于2015年3月19日公开的美国专利申请US 2015-0082478 A1和于2015年2月26日公开的US 2015-0059010 A1,两者均通过引用以其全文特此结合)。

[0034] 指导多核苷酸/Cas核酸内切酶复合物可以切割DNA靶序列的一条或两条链。可以切割DNA靶序列的两条链的指导多核苷酸/Cas核酸内切酶复合物通常包含具有处于功能状态的所有其核酸内切酶结构域的Cas蛋白(例如野生型核酸内切酶结构域或其变体在每个核酸内切酶结构域中保留一些或全部活性)。因此,在Cas蛋白的每个核酸内切酶结构域中保留一些或全部活性的野生型Cas蛋白(例如,本文披露的Cas9蛋白)或其变体是可以切割DNA靶序列的两条链的Cas核酸内切酶的合适实例。包含功能性RuvC和HNH核酸酶结构域的Cas9蛋白是可以切割DNA靶序列的两条链的Cas蛋白的实例。可以切割DNA靶序列的一条链的指导多核苷酸/Cas核酸内切酶复合物可以在本文中表征为具有切口酶活性(例如,部分切割能力)。Cas切口酶通常包含一个功能性核酸内切酶结构域,该结构域允许Cas仅切割DNA靶序列的一条链(即,形成切口)。例如,Cas9切口酶可以包含(i)突变的、功能失调的RuvC结构域和(ii)功能性HNH结构域(例如野生型HNH结构域)。作为另一个实例,Cas9切口酶可以包含(i)功能性RuvC结构域(例如野生型RuvC结构域)和(ii)突变的、功能失调的HNH结构域。适用于本文使用的Cas9切口酶的非限制性实例披露于以下文献中:Gasiunas等人(Proc.Natl.Acad.Sci.U.S.A.[美国科学院院报]109:E2579-E2586)、Jinek等人(Science[科学]337:816-821)、Sapranauskas等人(Nucleic Acids Res.[核酸研究]39:9275-9282)和美国专利申请公开号2014/0189896,将其通过引用结合在此。

[0035] 可以使用一对Cas9切口酶来增加DNA靶向的特异性。一般来说,这可以通过引入两种Cas9切口酶来进行,这两种Cas9切口酶通过与具有不同引导序列的RNA组分缔合,在希望靶向的区域的相反链上在DNA序列附近进行靶向和产生切口。每个DNA链的这样的附近切割产生双链断裂(即,具有单链突出端的DSB),其然后被识别为非同源末端连接(NHEJ)(倾向于产生导致突变的不完美修复)或同源重组(HR)的底物。在这些实施例中的每个切口可以,例如,彼此分离至少约5、10、15、20、30、40、50、60、70、80、90、或100(或在5与100之间的任何整数)个碱基。本文中的一种或两种Cas9切口酶蛋白可以用于Cas9切口酶对。例如,可以使用具有突变的RuvC结构域但具有功能性HNH结构域的Cas9切口酶(即,Cas9HNH+/RuvC-) (例如,化脓性链球菌(*Streptococcus pyogenes*) Cas9HNH+/RuvC-)。通过使用本文中的合适的RNA组分(具有将每个切口酶靶向每个特异性DNA位点的指导RNA序列),将每个Cas9切口酶(例如,Cas9HNH+/RuvC-)引导到彼此邻近(分离多达100个碱基对)的特定的DNA位点。

[0036] Cas蛋白可以是包含一个或多个异源蛋白质结构域(例如除Cas蛋白之外的1、2、3或更多个结构域)的融合蛋白的一部分。这样的融合蛋白可以包含任何另外的蛋白质序列,以及任选地在任何两个结构域之间(例如在Cas和第一异源结构域之间)的连接体序列。可以与本文中的Cas蛋白融合的蛋白质结构域的实例包括但不限于表位标签(例如,组氨酸[His]、V5、FLAG、流感血球凝集素[HA]、myc、VSV-G、硫氧还蛋白[Trx]);报道子(例如谷胱甘肽-5-转移酶[GST]、辣根过氧化物酶[HRP]、氯霉素乙酰转移酶[CAT]、 $\beta$ -半乳糖苷酶、 $\beta$ -葡萄糖醛酸酶[GUS]、荧光素酶、绿色荧光蛋白[GFP]、HcRed、DsRed、青色荧光蛋白[CFP]、黄色

荧光蛋白[YFP]、蓝色荧光蛋白[BFP])；以及具有一个或多个以下活性的结构域：甲基化酶活性、脱甲基酶活性、转录激活活性(例如,VP16或VP64)、转录抑制活性、转录释放因子活性、组蛋白修饰活性、RNA切割活性和核酸结合活性。Cas蛋白还可以与结合DNA分子或其他分子的蛋白质融合,例如麦芽糖结合蛋白(MBP)、S-标签、Lex A DNA结合结构域(DBD)、GAL4ADNA结合结构域和单纯疱疹病毒(HSV)VP16。

[0037] 本文中的Cas蛋白可以来自以下任何属：气火菌属(Aeropyrum)、火棒菌属(Pyrobaculum)、硫化叶菌属(Sulfolobus)、古球状菌属(Archaeoglobus)、嗜盐小盒菌属(Haloarcula)、甲烷杆菌属(Methanobacterium)、产甲烷球菌属(Methanococcus)、甲烷八叠球菌属(Methanosarcina)、甲烷火菌属(Methanopyrus)、火球菌属(Pyrococcus)、嗜酸菌属(Picrophilus)、Thermioplasmia、棒状杆菌属(Corynebacterium)、分支杆菌属(Mycobacterium)、链霉菌属(Streptomyces)、产液菌属(Aquifrix)、卟啉单胞菌属(Porphromonas)、绿菌属(Chlorobium)、Thermus、芽孢杆菌属(Bacillus)、利斯特菌属(Listeria)、葡萄球菌属(Staphylococcus)、梭菌属(Clostridium)、高温厌氧杆菌属(Thermoanaerobacter)、支原菌属(Mycoplasma)、梭形杆菌属(Fusobacterium)、Azarcus、色素杆菌属(Chromobacterium)、奈瑟氏菌属(Neisseria)、亚硝化单胞菌属(Nitrosomonas)、脱硫弧菌属(Desulfovibrio)、土杆菌属(Geobacter)、Myrococcus、弯曲杆菌属(Campylobacter)、沃林氏菌属(Wolinella)、不动细菌属(Acinetobacter)、欧文氏菌属(Erwinia)、埃希氏菌属(Escherichia)、军团杆菌属(Legionella)、甲基球菌属(Methylococcus)、巴斯德氏菌属(Pasteurella)、发光杆菌属(Photobacterium)、沙门氏菌属(Salmonella)、黄单胞杆菌属(Xanthomonas)、耶尔森氏菌属(Yersinia)、链球菌属(Streptococcus)、密螺旋体属(Treponema)、弗朗西斯氏菌属(Francisella)、或栖热袍菌属(Thermotoga)。可替代地,本文中的Cas蛋白可以例如由SEQ ID NO:462-465、467-472、474-477、479-487、489-492、494-497、499-503、505-508、510-516或517-521编码,如披露于美国申请公开号2010/0093617中,其通过引用结合在此。

[0038] 在某些实施例中指导多核苷酸/Cas核酸内切酶复合物可以结合DNA靶位点序列,但不切割在靶位点序列处的任何链。这样的复合物可以包含其中所有核酸酶结构域都是突变的、功能失调的Cas蛋白。例如,可以结合到DNA靶位点序列但在靶位点序列处不切割任何链的本文的Cas9蛋白可以包含突变的、功能失调的RuvC结构域和突变的、功能失调的HNH结构域。结合但不切割靶DNA序列的本文中的Cas蛋白可以用于调节基因表达,例如,在该情况下,Cas蛋白可以与转录因子(或其部分)融合(例如抑制子或激活子,例如本文披露的那些中的任一种)。

[0039] 该Cas核酸内切酶基因可以是II型Cas9核酸内切酶基因,例如但不限于,2007年3月1日公开的并通过引用结合在此的WO2007/025097的SEQ ID NO:462、474、489、494、499、505和518中列出的Cas9基因。在另一个实施例中,Cas核酸内切酶基因是植物、玉蜀黍或大豆优化的Cas9核酸内切酶基因。本文中的Cas核酸内切酶基因可以是植物或微生物密码子优化的Cas9核酸内切酶基因。Cas核酸内切酶基因可以有效地连接至Cas密码子区域上游的SV40核靶向信号和Cas密码子区域下游的二分型VirD2核定位信号(Tinland等人,(1992) Proc.Natl.Acad.Sci.USA[美国科学院院报]89:7442-6)。

[0040] 本文中的“Cas9”(以前称为Cas5、Csn1、或Csx12)是指与cr核苷酸和tracr核苷酸

或与单指导多核苷酸形成复合物的II型CRISPR系统的Cas核酸内切酶,其用于特异性识别和切割DNA靶序列的全部或部分。Cas9蛋白包含RuvC核酸酶结构域和HNH(H-N-H)核酸酶结构域,它们各自可以在靶序列处切割单个DNA链(两个结构域的协同作用导致DNA双链切割,而一个结构域的活性导致一个缺口)。通常,RuvC结构域包含亚结构域I、II和III,其中结构域I位于Cas9的N-末端附近,并且亚结构域II和III位于蛋白质的中间,即位于HNH结构域的侧翼(Hsu等人,Cell[细胞],157:1262-1278)。II型CRISPR系统包括利用与至少一种多核苷酸组分复合的Cas9核酸内切酶的DNA切割系统。例如,Cas9可以与CRISPR RNA(crRNA)和反式激活CRISPR RNA(tracrRNA)复合。在另一个实例中,Cas9可以与单一指导RNA复合。

[0041] 本文所述的Cas9蛋白的氨基酸序列,以及本文中某些其他Cas蛋白,可能源自于,例如链球菌属(例如化脓性链球菌(*S. pyogenes*)、肺炎链球菌(*S. pneumoniae*)、嗜热链球菌(*S. thermophilus*)、无乳链球菌(*S. agalactiae*)、副血链球菌(*S. parasanguinis*)、口腔链球菌(*S. oralis*)、唾液链球菌(*S. salivarius*)、猕猴链球菌(*S. macacae*)、停乳链球菌(*S. dysgalactiae*)、咽峡炎链球菌(*S. anginosus*)、星座链球菌(*S. constellatus*)、假豕链球菌(*S. pseudoporcinus*)、变异链球菌(*S. mutans*)、李斯特菌属(*Listeria*) (例如,无害利斯特菌(*L. innocua*))、螺原体属(*Spiroplasma*) (例如,蜂螺旋原体(*S. apis*)、*S. syrphidicola*)、消化链球菌科(*Peptostreptococcaceae*)、奇异菌属(*Atopobium*)、卟啉单胞菌属(*Porphyromonas*) (例如,*P. catoniae*)、普雷沃菌属(*Prevotella*) (例如,中间普雷沃菌(*P. intermedia*))、韦荣球菌属(*Veillonella*)、密螺旋体属(*Treponema*) (例如,索氏密螺旋体(*T. socranskii*)、齿垢密螺旋体(*T. denticola*))、二氧化碳噬细胞菌属(*Capnocytophaga*)、芬戈尔德菌属(*Finegoldia*) (例如,大芬戈尔德菌(*F. magna*))、红蝽菌科(*Coriobacteriaceae*) (例如,*C. bacterium*)、*Olsenella* (例如,*O. Profusa*)、嗜血杆菌属(*Haemophilus*) (例如,*H. sputorum*、*H. pittmaniae*)、巴斯德氏菌属(*Pasteurella*) (例如,贝氏巴斯德菌(*P. bettyae*))、*Olivibacter* (例如,*O. sitiensis*)、*Epilithonimonas* (例如,*E. tenax*)、*Mesonia* (例如,*M. mobilis*)、乳杆菌属(*Lactobacillus*) (例如,植物乳杆菌(*L. plantarum*))、芽孢杆菌属(*Bacillus*) (例如,蜡样芽孢杆菌(*B. Cereus*))、*Aquimarina* (例如,*A. Muelleri*)、金黄杆菌属(*Chryseobacterium*) (例如,*C. Palustre*)、多形杆状菌属(*Bacteroides*) (例如,*B. grammisolvens*)、奈瑟氏菌属(*Neisseria*) (例如,脑膜炎奈瑟氏菌(*N. Meningitidis*))、弗朗西斯氏菌属(*Francisella*) (例如,新凶手弗朗西斯菌(*F. Novicida*))、或黄杆菌属(*Flavobacterium*) (例如,寒冷黄杆菌(*F. Frigidarium*)、*F. soli*)物种。作为另一个实例,Cas9蛋白可以是披露于Chylinski等人(RNA Biology[RNA生物学]10:726-737以及2015年5月15日提交的美国专利申请62/162377,其通过引用结合在此)中的任何Cas9蛋白。

[0042] 因此,本文中Cas9蛋白的序列可以包括,例如,在以下各项中披露的任何Cas9氨基酸序列:基因库登录号G3ECR1(嗜热链球菌(*S. thermophilus*))、WP\_026709422、WP\_027202655、WP\_027318179、WP\_027347504、WP\_027376815、WP\_027414302、WP\_027821588、WP\_027886314、WP\_027963583、WP\_028123848、WP\_028298935、Q03JI6(嗜热链球菌)、EGP66723、EGS38969、EGV05092、EHI65578(假豕链球菌)、EIC75614(口腔链球菌)、EID22027(星座链球菌)、EIJ69711、EJP22331(口腔链球菌)、EJP26004(咽峡炎链球菌)、EJP30321、EPZ44001(化脓性链球菌)、EPZ46028(化脓性链球菌)、EQL78043(化脓性链球菌)、EQL78548

(化脓性链球菌)、ERL10511、ERL12345、ERL19088 (化脓性链球菌)、ESA57807 (化脓性链球菌)、ESA59254 (化脓性链球菌)、ESU85303 (化脓性链球菌)、ETS96804、UC75522、EGR87316 (停乳链球菌)、EGS33732、EGV01468 (口腔链球菌)、EHJ52063 (猕猴链球菌)、EID26207 (口腔链球菌)、EID33364、EIG27013 (副血链球菌)、EJF37476、EJ019166 (链球菌属物种BS35b)、EJU16049、EJU32481、YP\_006298249、ERF61304、ERK04546、ETJ95568 (无乳链球菌)、TS89875、ETS90967 (链球菌属物种SR4)、ETS92439、EUB27844 (链球菌属物种BS21)、AFJ08616、EUC82735 (链球菌属物种CM6)、EWC92088、EWC94390、EJP25691、YP\_008027038、YP\_008868573、AGM26527、AHK22391、AHB36273、Q927P4、G3ECR1、或Q99ZW2 (化脓性链球菌)，其通过引用结合。可以使用任何这些Cas9蛋白序列的变体，但是当与本文中的RNA组分缔合时，它们应当具有针对DNA的特异性结合活性和任选地内切核苷酸活性。这样的变体可以包含与参比Cas9的氨基酸序列具有至少约80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、或99%同一性的氨基酸序列。

[0043] 可替代地，例如，本文中的Cas9蛋白可以由如披露于美国申请公开号2010/0093617 (通过引用结合在此)中的SEQ ID NO:462 (嗜热链球菌)、474 (嗜热链球菌)、489 (无乳链球菌)、494 (无乳链球菌)、499 (变异链球菌)、505 (化脓性链球菌)、或518 (化脓性链球菌)中的任何一个编码。仍可替代地，Cas9蛋白可以包含例如，与任何前述氨基酸序列具有至少约80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%同一性的氨基酸序列。当与本文中的RNA组分缔合时，这样的变体Cas9蛋白应当具有针对DNA的特异性结合活性，以及任选地切割或切口活性。

[0044] 本文中的Cas蛋白 (如Cas9) 可以包括异源核定位序列 (NLS)。例如，本文中的异源NLS氨基酸序列可能具有足够的强度来驱动在本文的酵母细胞细胞核中可检测的量的Cas蛋白的积累。NLS可以包含碱性、带正电荷的残基 (例如赖氨酸和/或精氨酸) 的一个 (单分型) 或多个 (例如，二分型) 短序列 (例如，2至20个残基)，并且可以位于Cas氨基酸序列中的任何地方，但使得其暴露于蛋白质表面上。例如，NLS可以有效地连接到本文中的Cas蛋白的N-末端或C-末端。两个或更多个NLS序列可以连接到Cas蛋白，例如在Cas蛋白的N-末端和C-末端两者。本文中合适的NLS序列的非限制性实例包括在美国专利号6660830和7309576 (例如，其中的表1) 中披露的那些，其都通过引用结合在此。

[0045] 该Cas核酸内切酶可以包括Cas9多肽的修饰形式。该Cas9多肽的修饰形式可以包括降低Cas9蛋白的天然存在的核酸酶活性的氨基酸变化 (例如，缺失、插入、或取代)。例如，在一些情况下，该Cas9蛋白的修饰形式具有低于50%、低于40%、低于30%、低于20%、低于10%、低于5%、或低于1%的相应的野生型Cas9多肽 (2014年3月6日公开的美国专利申请US20140068797A1) 的核酸酶活性。在一些情况下，该Cas9多肽的修饰形式不具有实质性的核酸酶活性并且被称作催化“失活的Cas9”或“失活的cas9 (dCas9)”。催化失活的Cas9变体包括在HNH和RuvC核酸酶结构域中包含突变的Cas9变体。这些催化失活的Cas9变体能够与sgRNA相互作用，并且在体内结合至靶位点但不能切割靶DNA的任一链。

[0046] 催化失活的Cas9可以融合到异源序列 (于2014年3月6日公开的美国专利申请US20140068797 A1)。合适的融合配偶体包括，但不限于提供活性的多肽，该活性通过直接作用于靶DNA上或与该靶DNA相关的多肽 (例如，组蛋白或其他DNA-结合蛋白) 上间接地增加

转录。另外的合适的融合配偶体包括,但不限于提供甲基转移酶活性、脱甲基酶活性、乙酰基转移酶活性、脱乙酰基酶活性、激酶活性、磷酸酶活性、泛素连接酶活性、去泛素化酶活性、腺苷酸化活性、去腺苷酸化活性、苏素化活性、去苏素化活性、核糖基化活性、去核糖基化活性、豆蔻酰化活性,或去豆蔻酰化活性的多肽。此外合适的融合配偶体包括,但不限于直接提供靶核酸的增加的转录的多肽(例如,募集转录激活因子、小分子/药物-应答性转录调节因子等的转录激活因子或其片段,蛋白质或其片段)。还可以将催化失活的Cas9融合到FokI核酸酶从而产生双链断裂(Guilinger等人Nature biotechnology[自然生物技术],第32卷,第6期,2014年6月)。

[0047] 术语Cas核酸内切酶的“功能性片段”、“功能上等效的片段”和“功能等效片段”在本文中可互换地使用,并且意指Cas核酸内切酶序列的一部分或子序列,其中保留了识别、结合靶位点、并任选地使靶位点产生切口或切割(引入单链或双链断裂)靶位点的能力。

[0048] 术语Cas核酸内切酶的“功能性变体”、“功能上等效的变体”和“功能等效变体”在本文中可互换地使用,并且意指Cas核酸内切酶的变体,其中保留了识别、结合靶位点、并任选地使靶位点产生切口或切割(引入单链或双链断裂)靶位点的能力。这些片段和变体可以通过如定点诱变和合成构建等方法来获得。

[0049] 该Cas核酸内切酶基因包括可以识别N(12-30)NGG(其原则上可以被靶向)的任何基因组序列的植物密码子优化的化脓性链球菌Cas9基因或源自以下生物体的Cas9核酸内切酶,所述生物体选自下组,该组由以下各项组成:侧孢短芽孢杆菌、罗伊氏乳杆菌M1c3、红色乳杆菌DSM 15814、戊糖片球菌SL4、Lactobacillus nodensis JCM 14932、硫磺单胞菌属物种SCADC、嗜热双歧杆菌DSM 20210、Loktanellela vestfoldensis、三赞鞘氨醇单胞菌NX02、Epilithonimonas tenax DSM 16811、粘球生孢噬纤维菌、和扭曲冷弯曲菌ATCC 700755,其中所述Cas9核酸内切酶可以形成指导RNA/Cas核酸内切酶复合物,该复合物能够识别、结合DNA靶序列的全部或部分、并任选地使DNA靶序列的全部或部分产生切口或切割DNA靶序列的全部或部分。其他Cas核酸内切酶系统已经在2015年5月15日提交的美国专利申请62/162,377和2015年5月15日提交的美国专利申请62/162,353中描述,这两个申请通过引用结合在此。

[0050] Cas9核酸内切酶可以用于靶向的基因组编辑(经由单个和多个双链断裂和缺口)和靶向的基因组调控(经由将表观遗传效应子结构域系链到Cas9或sgRNA)。Cas9还可以被工程化作为RNA指导的重组酶起作用,并且经由RNA系链可以充当用于组装多蛋白和核酸复合物的支架(Mali等人2013Nature Methods[自然方法]第10卷:957-963)。

[0051] 如本文所用,术语“指导多核苷酸”涉及可以与Cas核酸内切酶形成复合物的多核苷酸序列,并且使得Cas核酸内切酶能够识别、结合并任选地切割DNA靶位点。指导多核苷酸可以是单分子或双分子。指导多核苷酸序列可以是RNA序列(称为指导RNA、gRNA)、DNA序列或其组合(RNA-DNA组合序列)。任选地,指导多核苷酸可以包含至少一种核苷酸、磷酸二酯键或连接修饰,例如但不限于锁核酸(LNA)、5-甲基dC、2,6-二氨基嘌呤、2'-氟代A、2'-氟代U、2'-O-甲基RNA、硫代磷酸酯键、与胆固醇分子的连接、与聚乙二醇分子的连接、与间隔子18(六乙二醇链)分子的连接、或导致环化的5'至3'共价连接。仅仅包含核糖核酸的指导多核苷酸也被称为“指导RNA”或“gRNA”(还参见于2015年3月19日公开的美国专利申请US 2015-0082478 A1和2015年2月26日公开的US 2015-0059010 A1,两者均通过引用以其全文

特此结合)。

[0052] 指导多核苷酸可以是包含cr核苷酸序列和tracr核苷酸序列的双分子(也称为双链体指导多核苷酸)。cr核苷酸包括可以与靶DNA中的核苷酸序列杂交的第一核苷酸序列结构域(称为可变靶向结构域或VT结构域)和作为Cas核酸内切酶识别(CER)结构域的一部分的第二核苷酸序列(也称为tracr配对序列)。tracr配对序列可以沿互补区域与tracr核苷酸杂交,并一起形成Cas核酸内切酶识别结构域或CER结构域。CER结构域能够与Cas核酸内切酶多肽相互作用。双链体指导多核苷酸的cr核苷酸和tracr核苷酸可以是RNA、DNA和/或RNA-DNA组合序列。在一些实施例中,双链体指导多核苷酸的cr核苷酸分子被称为“crDNA”(当由DNA核苷酸的连续延伸构成时)或“crRNA”(当由RNA核苷酸的连续延伸构成时)或“crDNA-RNA”(当由DNA和RNA核苷酸的组合构成时)。cr核苷酸可以包括在细菌和古细菌中天然存在的crRNA的片段。可以存在于本文披露的cr核苷酸中的、细菌和古细菌中天然存在的crRNA片段的大小可以是但不限于2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20或更多个核苷酸。在一些实施例中,tracr核苷酸被称为“tracrRNA”(当由RNA核苷酸的连续延伸构成时)或“tracrDNA”(当由DNA核苷酸的连续延伸构成时)或“tracrDNA-RNA”(当由DNA和RNA核苷酸的组合构成时)。在一个实施例中,指导RNA/Cas9核酸内切酶复合物的RNA是包含双链体crRNA-tracrRNA的双链体化的RNA。在5'-至-3'方向上,tracrRNA(反式激活CRISPR RNA)包含(i)与CRISPR II型crRNA的重复区退火的序列和(ii)含茎环的部分(Del tcheva等人,Nature[自然]471:602-607)。双链体指导多核苷酸可以与Cas核酸内切酶形成复合物,其中所述指导多核苷酸/Cas核酸内切酶复合物(也被称为指导多核苷酸/Cas核酸内切酶系统)可以将Cas核酸内切酶引导至基因组靶位点,使得Cas核酸内切酶可以识别、结合该靶位点、并任选地使该靶位点产生切口或切割(引入单链或双链断裂)该靶位点。(还参见于2015年3月19日公开的美国专利申请US 2015-0082478 A1和于2015年2月26日公开的US 2015-0059010 A1,两者均通过引用以其全文特此结合)。

[0053] 指导多核苷酸也可以是包含连接至tracr核苷酸序列的cr核苷酸序列的单分子(也称为单指导多核苷酸)。单指导多核苷酸包含可以与靶DNA中的核苷酸序列杂交的第一核苷酸序列结构域(称为可变靶向结构域或VT结构域)和与Cas核酸内切酶多肽相互作用的Cas核酸内切酶识别结构域(CER结构域)。“结构域”意指可以为RNA、DNA和/或RNA-DNA组合序列的核苷酸的连续延伸。单指导多核苷酸的VT结构域和/或CER结构域可以包含RNA序列、DNA序列或RNA-DNA组合序列。由来自cr核苷酸和tracr核苷酸的序列构成的单指导多核苷酸可以被称为“单指导RNA”(当由RNA核苷酸的连续延伸构成时)或“单指导DNA”(当由DNA核苷酸的连续延伸构成时)或“单指导RNA-DNA”(当由RNA和DNA核苷酸的组合构成时)。单指导多核苷酸可以形成具有Cas核酸内切酶的复合物,其中所述指导多核苷酸/Cas核酸内切酶复合物(还称作指导多核苷酸/Cas核酸内切酶系统)可以将Cas核酸内切酶引导至基因组靶位点,使该Cas核酸内切酶能够识别、结合靶位点、并任选地使靶位点产生切口或切割(引入单链或双链断裂)靶位点。(还参见于2015年3月19日公开的美国专利申请US 2015-0082478 A1和于2015年2月26日公开的US 2015-0059010 A1,两者均通过引用以其全文特此结合)。

[0054] 术语“可变靶向结构域”或“VT结构域”在本文中可互换使用,并且包括能与双链DNA靶位点的一条链(核苷酸序列)杂交(互补)的核苷酸序列。第一核苷酸序列结构域(VT结构域)与靶序列之间的互补%可以为至少50%、51%、52%、53%、54%、55%、56%、57%、

58%、59%、60%、61%、62%、63%、63%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%。可变靶向结构域可以是至少12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29或30个核苷酸长度。在一些实施例中,可变靶向结构域包含12至30个核苷酸的连续延伸。可变靶向结构域可以由DNA序列、RNA序列、修饰的DNA序列、修饰的RNA序列或其任何组合构成。

[0055] 术语(指导多核苷酸的)“Cas核酸内切酶识别结构域”或“CER结构域”在本文中可互换地使用,并且包括与Cas核酸内切酶多肽相互作用的核苷酸序列。CER结构域包含tracr核苷酸配对序列,随后是tracr核苷酸序列。CER结构域可以由DNA序列、RNA序列、修饰的DNA序列、修饰的RNA序列(参见例如2015年2月26日公开的US 2015-0059010 A1,其通过引用以其全文结合在此)或其任何组合构成。

[0056] 连接单指导多核苷酸的cr核苷酸和tracr核苷酸的核苷酸序列可以包含RNA序列、DNA序列或RNA-DNA组合序列。在一个实施例中,连接单指导多核苷酸的cr核苷酸和tracr核苷酸的核苷酸序列可以是至少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、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99或100个核苷酸的长度。在另一个实施例中,连接单指导多核苷酸的cr核苷酸和tracr核苷酸的核苷酸序列可以包含四核苷酸环序列,例如但不限于GAAA四核苷酸环序列。

[0057] 指导多核苷酸、VT结构域和/或CER结构域的核苷酸序列修饰可以选自但不限于由以下各项组成的组:5'帽、3'聚腺苷酸尾、核糖开关序列、稳定性控制序列、形成dsRNA双链体的序列、将指导多核苷酸靶向亚细胞位置的修饰或序列、提供跟踪的修饰或序列、提供蛋白质结合位点的修饰或序列、锁核酸(LNA)、5-甲基dC核苷酸、2,6-二氨基嘌呤核苷酸、2'-氟A核苷酸、2'-氟U核苷酸、2'-O-甲基RNA核苷酸、硫代磷酸酯键、与胆固醇分子的连接、与聚乙二醇分子的连接、与间隔物18分子的连接、5'至3'共价连接、或其任何组合。这些修饰可以产生至少一个另外的有益特征,其中该另外的有益特征选自由以下组成的组:修改的或调节的稳定性、亚细胞靶向、跟踪、荧光标记、用于蛋白质或蛋白质复合物的结合位点、对互补靶序列的修改的结合亲和力、修改的细胞降解抗性和增加的细胞通透性。

[0058] 术语指导RNA、crRNA或tracrRNA的“功能片段”、“功能上等效的片段”和“功能等效片段”在本文中可互换地使用,并且分别意指指导RNA、crRNA或tracrRNA的一部分或子序列,其中分别保留用作指导RNA、crRNA或tracrRNA的能力。

[0059] 术语指导RNA、crRNA或tracrRNA(分别地)的“功能变体”、“功能上等效的变体”和“功能等效变体”在本文中可互换地使用,并且分别意指指导RNA、crRNA或tracrRNA的变体,其中分别保留用作指导RNA、crRNA或tracrRNA的能力。

[0060] 术语“单指导RNA”、“gRNA”和“sgRNA”在本文中可互换使用,并涉及两个RNA分子的合成融合,其中包含可变靶向结构域(与tracrRNA杂交的tracr配对序列连接)的crRNA(CRISPR RNA)与tracrRNA(反式激活CRISPR RNA)融合。单指导RNA可以包含可与II型Cas核酸内切酶形成复合物的II型CRISPR/Cas系统的crRNA或crRNA片段和tracrRNA或tracrRNA



片段,其中所述指导RNA/Cas核酸内切酶复合物可以将Cas核酸内切酶引导至DNA靶位点,使得Cas核酸内切酶能够识别、结合DNA靶位点、并任选地使DNA靶位点产生切口或切割(引入单链或双链断裂)DNA靶位点。

[0061] 术语“指导RNA/Cas核酸内切酶复合物”、“指导RNA/Cas核酸内切酶系统”、“指导RNA/Cas复合物”、“指导RNA/Cas系统”、“gRNA/Cas复合物”、“gRNA/Cas系统”、“RNA-指导的核酸内切酶”、“RGEN”在本文中可互换地使用,并且意指能够形成复合物的至少一种RNA组分和至少一种Cas核酸内切酶蛋白,其中所述指导RNA/Cas核酸内切酶复合物可以将Cas核酸内切酶引导至DNA靶位点,使Cas核酸内切酶能够识别、结合DNA靶位点、并任选地使DNA靶位点产生切口或切割(引入单链或双链断裂)DNA靶位点。本文中的指导RNA/Cas核酸内切酶复合物可以包含已知的CRISPR系统(Zetsche等人,2015,Cell[细胞]163,1-13;Shmakov等人,2015,Molecular\_Cell[分子细胞]60,1-13;Makarova等人2015,Nature Reviews Microbiology[自然评论-微生物]第13卷:1-15;Horvath和Barrangou,Science[科学]327:167-170)(例如I型、II型或III型CRISPR系统)中任一种的一种或多种Cas蛋白和一种或多种合适的RNA组分。指导RNA/Cas核酸内切酶复合物可以包括II型Cas9核酸内切酶和至少一种RNA组分(例如,crRNA和tracrRNA、或gRNA)。(还参见于2015年3月19日公开的美国专利申请US 2015-0082478 A1和于2015年2月26日公开的US 2015-0059010 A1,两者均通过引用以其全文特此结合)。

[0062] Cas核酸内切酶可以通过本领域已知的任何方法引入细胞中,这些方法例如但不限于瞬时引入法、转染、显微注射、和/或局部施用、或间接经由重组构建体。植物细胞与人类和动物细胞的不同之处在于,植物细胞含有植物细胞壁,其可以作为将Cas9核酸内切酶直接递送到植物细胞中的屏障。编码Cas9核酸内切酶的重组DNA构建体已经成功地导入植物细胞中(Svitashev等人,Plant Physiology[植物生理学],2015,第169卷,第931-945页)以允许在靶位点进行基因组编辑。在植物细胞中稳定引入重组DNA构建体的一个可能的缺点是Cas9核酸内切酶的持续存在可能增加脱靶效应。

[0063] 如本文所述,可以通过颗粒介导的递送将Cas核酸内切酶直接递送到植物细胞中。基于本文所述的实验,技术人员现在可以预想任何其他直接递送方法(例如但不限于聚乙二醇(PEG)介导的对原生质体的转染、晶须介导的递送、电穿孔、粒子轰击、细胞穿透肽或介孔二氧化硅纳米颗粒(MSN)介导的直接蛋白质递送)都可以成功用于将Cas9核酸内切酶递送到植物细胞中。

[0064] 可以通过使用本领域已知的任何方法将Cas蛋白、编码Cas核酸内切酶的mRNA、和/或RNA指导的核酸内切酶核糖核苷酸-蛋白质复合物(RGEN)本身(作为核糖核苷酸-蛋白质复合物)引入细胞中来实现Cas核酸内切酶的直接递送(也称为Cas核酸内切酶的无DNA递送)。经由编码Cas核酸内切酶的mRNA或经由多肽分子直接递送Cas核酸内切酶在本文中也称为Cas核酸内切酶的无DNA递送,因为没有DNA分子参与Cas核酸内切酶蛋白的产生。类似地,作为RNA分子的指导RNA的直接递送在本文中也称为指导RNA的无DNA递送。类似地,作为核糖核苷酸-蛋白质复合物的指导RNA/核酸内切酶复合物本身(RGEN)的直接递送在本文中也称为RGEN的无DNA递送。

[0065] 将Cas核酸内切酶作为蛋白质或作为mRNA分子与gRNA一起或作为RGEN核糖核苷酸-蛋白质本身直接引入,允许在靶位点处进行基因组编辑,然后迅速降解RGEN复合物,并

且仅允许细胞中暂时存在该复合物,这导致脱靶效应降低(如实例12中所述)。

[0066] 这些组分的直接递送可以伴随着可以促进接受RGEN组分的细胞的富集和/或可视化的其他mRNA的直接递送(共递送)。例如,也可以使用编码可筛选的可视标记如荧光蛋白(例如但不限于红色、绿色、黄色、蓝色或其组合)的mRNA的递送来代替经修复的破坏的、非功能性基因产物的直接选择或与其相偶联。

[0067] 本文描述了通过恢复破坏的基因的核苷酸序列来恢复非功能性基因产物的功能,这样使得恢复的核苷酸序列编码功能性基因产物的方法。

[0068] 破坏的基因是指如下基因:该基因已经被修饰(破坏)这样使得其基因产物失去其功能(被称为非功能性基因产物)、或当与不具有破坏的相应基因(也被称为未被破坏的基因)的产物相比时具有降低的功能。例如,编码功能性多肽或蛋白质的基因可以被破坏(修饰),这样使得破坏的基因的翻译产物产生失去其功能或具有降低的功能的多肽。

[0069] 功能性基因产物包括具有生物学或非生物学功能的功能性蛋白质或多肽。

[0070] 非功能性基因产物包括提及破坏的基因的基因产物。非功能性基因产物包括失去其功能(功能缺失)或当与相应的未破坏的基因的基因产物相比时具有降低的功能的多肽。

[0071] 与通过NHEJ恢复基因功能(通过例如将RGEN组分或RGEN复合物本身递送至细胞)一致,可以通过同时添加其他指导多核苷酸来完成对其他靶标的修饰。这样的其他靶标(除了通过NHEJ恢复基因功能的靶标之外)可以是基因组中的任何靶标,包括转基因基因座。当一种gRNA通过NHEJ靶向并激活可选择标记(例如但不限于赋予除草剂耐受性)并且另外一种或多种gRNA在一个或多个靶位点处促进一个或多个不同于可选择标记的DSB(或其他破坏的基因设计)并且可以促进靶向诱变、缺失、基因编辑或位点特异性性状基因插入时,同时递送两种或更多种gRNA的方法可以允许完全瞬时靶向的基因组修饰,因为所有其他必需组分(Cas9、gRNA)能以蛋白质和/或体外转录的RNA分子的形式递送。

[0072] 破坏的基因包括提及如下标记基因(例如但不限于表型标记基因和可选择标记基因),该标记基因已被修饰(破坏)这样使得其基因产物失去其功能(例如,在除草剂破坏的可选择标记基因的情况下,破坏的基因不再赋予除草剂抗性)。

[0073] 可选择标记和可筛选标记在本文中可互换使用,并且包括提及允许人们通常在特定条件下鉴定、或选择含有它的分子或细胞或对其进行选择的DNA区段(如可选择标记基因)。这些标记可以编码活性,例如但不限于RNA、肽或蛋白质的产生,或可以提供RNA、肽、蛋白质、无机和有机化合物或组合物等的结合位点。可选择标记进一步包括当修饰或敲除时在细胞中产生以下特性的基因,所述特性允许人们鉴定或选择含有所述特性的细胞(或对其进行选择)。

[0074] 在一方面,可选择标记允许通过应用选择方案来选择细胞,其中例如使用选择试剂(例如但不限于抗生素或除草剂)来抑制或杀死不包含该可选择标记的细胞或组织,并且包含该可选择标记的细胞或组织由于该可选择标记基因的表达而继续生长。

[0075] 在一方面,可选择标记允许通过应用选择方案对细胞进行视觉选择,其中例如使用可视化标记(如荧光分子)来选择包含该可视化标记的细胞。

[0076] 可选择标记基因包括但不限于氯磺隆抗性基因、磷酸甘露糖异构酶基因(PMI)、双丙氨磷抗性基因(BAR)、草丁膦乙酰转移酶(PAT)基因、潮霉素抗性基因(NPTII)、草甘膦抗性基因、包含限制性内切酶位点的DNA区段;编码对另外的毒性化合物提供抗性的产物的

DNA区段,所述毒性化合物包括抗生素,例如壮观霉素、氨基青霉素、卡那霉素、四环素、Basta、新霉素磷酸转移酶II (NEO) 和潮霉素磷酸转移酶 (HPT);编码在受体细胞中本身缺乏的产物的DNA区段(例如,tRNA基因、营养缺陷型标记);编码易于鉴定的产物的DNA区段(称为可视化标记基因)。可视化标记基因包括提及荧光标记基因(如红色荧光标记基因、蓝色荧光标记基因、绿色荧光标记基因、黄色荧光标记基因);编码DsRED (RFP,即红色荧光蛋白)、CFP、GFP (绿色荧光蛋白)的基因;以及编码表型标记(如 $\beta$ -半乳糖苷酶、GUS)、荧光蛋白(如绿色荧光蛋白 (GFP)、青色荧光蛋白 (CFP)、黄色荧光蛋白 (YFP)、红色荧光蛋白 (RFP))和细胞表面蛋白的基因。可选择标记基因进一步包括产生用于PCR的新引物位点(例如,以前未并列的两个DNA序列的并列),包含通过限制性核酸内切酶或其他DNA修饰酶、化学品等不起作用或起作用的DNA序列;并且包含允许其鉴定的特异性修饰(例如,甲基化)所需的DNA序列。

[0077] 另外的可选择标记包括赋予除草剂化合物(例如草胺磷、溴草腈、咪唑啉酮和2,4-二氯苯氧基乙酸酯(2,4-D))抗性的基因。参见例如,Yarranton,(1992) *Curr. Opin. Biotech.* [生物技术新见] 3:506-11; Christopherson等人,(1992) *Proc. Natl. Acad. Sci. USA* [美国科学院院报] 89:6314-8; Yao等人,(1992) *Cell* [细胞] 71:63-72; Reznikoff,(1992) *Mol. Microbiol* [分子微生物学] 6:2419-22; Hu等人,(1987) *Cell* [细胞] 48:555-66; Brown等人,(1987) *Cell* [细胞] 49:603-12; Figge等人,(1988) *Cell* [细胞] 52:713-22; Deuschle等人,(1989) *Proc. Natl. Acad. Sci. USA* [美国科学院院报] 86:5400-4; Fuerst等人,(1989) *Proc. Natl. Acad. Sci. USA* [美国科学院院报] 86:2549-53; Deuschle等人,(1990) *Science* [科学] 248:480-3; Gossen,(1993) 博士学位论文, University of Heidelberg [德国海德堡大学]; Reines等人,(1993) *Proc. Natl. Acad. Sci. USA* [美国科学院院报] 90:1917-21; Labow等人,(1990) *Mol. Cell. Biol.* [分子细胞生物学] 10:3343-56; Zambretti等人,(1992) *Proc. Natl. Acad. Sci. USA* [美国科学院院报] 89:3952-6; Baim等人,(1991) *Proc. Natl. Acad. Sci. USA* [美国科学院院报] 88:5072-6; Wyborski等人,(1991) *Nucleic Acids Res.* [核酸研究] 19:4647-53; Hillen和Wissman,(1989) *Topics Mol. Struct. Biol.* [分子与结构生物学专题] 10:143-62; Degenkolb等人,(1991) *Antimicrob. Agents Chemother.* [抗微生物剂化学疗法] 35:1591-5; Kleinschmidt等人,(1988) *Biochemistry* [生物化学] 27:1094-104; Bonin,(1993) 博士学位论文, University of Heidelberg [德国海德堡大学]; Gossen等人,(1992) *Proc. Natl. Acad. Sci. USA* [美国科学院院报] 89:5547-51; Oliva等人,(1992) *Antimicrob. Agents Chemother.* [抗微生物剂化学疗法] 36:913-9; Hlavka等人,(1985) *Handbook of Experimental Pharmacology* [实验药理学手册], 第78卷 (Springer-Verlag, Berlin [柏林施普林格出版社]); Gill等人,(1988) *Nature* [自然] 334:721-4. *J. Bacteriol.* [细菌学杂志] 170:5837-5847) 促进聚羟基链烷酸酯 (PHA) 的表达。

[0078] 表型标记基因包括编码可筛选或可选择标记的基因,所述标记包括可视化标记,不管它是阳性还是阴性可选择标记。可以使用任何表型标记。

[0079] 如本文所述,表型和可选择标记基因可以被修饰以作为编码非功能性基因产物的破坏基因引入植物细胞中,并且被诱导双链断裂的核酸内切酶用作通过指导RNA引入和DNA修复来恢复回编码功能性基因产物的非破坏的基因的靶标。

[0080] 待破坏的表型或可选择标记基因可以是先前引入细胞中并稳定整合到细胞基因组中的标记基因。此类预整合的可选择标记基因还可以用其他基因进行补充,例如细胞发育增强基因(ZmODP2和ZmWUS,参见例如2016年8月26日提交的PCT/US16/49144和2016年8月26日提交的PCT/US16/49128,通过引用结合在此)。

[0081] 本领域技术人员可以理解,常规的基因组修饰方法主要依靠将可选择标记基因引入待修饰的细胞中以允许选择方案,在这些选择方案中,例如使用抗生素或除草剂(选择试剂)来抑制或杀死不包含该可选择标记基因的细胞或组织,并且由于该可选择标记(抗性)基因的表达,包含该可选择标记基因的细胞或组织继续生长。相比之下,可以在不使用可选择标记并且不应用选择试剂的情况下使用本披露的方法。

[0082] 在该披露的一个实施例中,该方法包括用于将能够在位于核苷酸序列中的靶位点中制造双链断裂的指导多核苷酸/Cas核酸内切酶复合物引入至少一个植物细胞中的方法;并且,选择在所述核苷酸序列中具有修饰的植物细胞,其中所述选择在不使用可选择标记的情况下发生。该指导多核苷酸/Cas核酸内切酶复合物可以是指导RNA/Cas核酸内切酶复合物。

[0083] 在本披露的一个实施例中,该方法包括用于在不使用可选择标记的情况下产生在其基因组中具有修饰的核苷酸序列的植物的方法,所述方法包括:将能够在位于核苷酸序列中的靶位点中制造双链断裂的指导多核苷酸/Cas核酸内切酶复合物引入至少一个植物细胞中;从所述植物细胞获得植物;并且,选择在所述核苷酸序列中具有修饰的植物,其中所述选择在不使用可选择标记的情况下发生。该指导多核苷酸/Cas核酸内切酶复合物可以是指导RNA/Cas核酸内切酶复合物。

[0084] 在本披露的一个实施例中,该方法包括用于在不使用可选择标记的情况下产生在其基因组中具有修饰的核苷酸序列的植物愈伤组织的方法,所述方法包括:将能够在位于核苷酸序列中的靶位点中制造双链断裂的指导多核苷酸/Cas核酸内切酶复合物引入至少一个植物细胞中;从所述植物细胞获得愈伤组织;选择在所述核苷酸序列中具有修饰的愈伤组织,其中所述选择在不使用可选择标记的情况下发生。该指导多核苷酸/Cas核酸内切酶复合物可以是指导RNA/Cas核酸内切酶复合物。

[0085] 本文所述的方法中的选择步骤可以包括本领域技术人员已知的不基于使用可选择标记的任何选择(鉴定、获得),例如像通过基因型手段(例如但不限于DNA测序)或表型手段(例如植物形态学特征或与其基因组中所希望的修饰有关的可选择表型)选择在其基因组中包含所希望的修饰的植物细胞、愈伤组织或植物。

[0086] 在一方面,该方法包括用于在不使用可选择标记的情况下修饰植物细胞的基因组中的核苷酸序列的方法,所述方法包括:将能够在位于所述核苷酸序列中的靶位点中制造双链断裂的指导RNA/Cas核酸内切酶复合物引入至少一个植物细胞中,其中在不使用重组DNA构建体的情况下将指导RNA/Cas核酸内切酶复合物引入细胞中;从(a)的植物细胞获得植物;并且,选择在所述核苷酸序列中具有修饰的植物,其中所述选择经由对所述植物的DNA测序而发生。

[0087] 本文所述的方法可以进一步包括引入多核苷酸模板,其中所述多核苷酸修饰模板包含所述细胞的基因组中的核苷酸序列的至少一个核苷酸修饰,其中所述多核苷酸修饰模板的所述至少一个核苷酸修饰选自下组,该组由以下各项组成:(i)至少一个核苷酸的替

代、(ii)至少一个核苷酸的缺失、(iii)至少一个核苷酸的插入、以及(iv) (i)-(iii)的任何组合。该方法还可以进一步包括引入供体DNA,其中所述供体DNA包含至少一个目的多核苷酸。可以经由选自下组的递送系统将指导多核苷酸/Cas核酸内切酶复合物引入细胞中,该组由以下各项组成:颗粒介导的递送、晶须介导的递送、细胞穿透肽介导的递送、电穿孔、PEP介导的转染和纳米颗粒介导的递送。

[0088] 可以在不使用表型标记或可选择标记并且不应用选择试剂的情况下使用本披露的方法。这些方法包括在也不将可选择标记引入植物细胞中的情况下将指导RNA/Cas核酸内切酶复合物引入所述植物细胞中,或其中引入所述指导RNA/Cas核酸内切酶复合物不伴随着将破坏的可选择标记基因恢复成编码功能性可选择标记蛋白的非破坏的可选择标记基因,或者其中引入所述指导RNA/Cas核酸内切酶复合物不导致在所述细胞内产生可选择标记。所述方法包括选择在其基因组中包含修饰的核苷酸序列的植物细胞、愈伤组织或植物,其中所述选择在不使用可选择标记的情况下发生。

[0089] 任何指导的核酸内切酶可以用于本文披露的方法中。此类核酸内切酶包括但不限于Cas和Cpf1核酸内切酶。迄今为止已经描述了许多核酸内切酶,其可以识别特定的PAM序列(参见例如-于2015年5月15日提交的美国专利申请62/162377和于2015年5月15日提交的美国专利申请62/162353以及Zetsche B等人2015,Ce11[细胞]163,1013)并且在特定位置处切割靶DNA。应当理解的是,基于本文所述的使用指导的Cas系统的方法和实施例,现在人们可以定制这些方法这样使得它们可以利用任何指导的核酸内切酶系统。例如,可以设想将本文所述的用于在不使用可选择标记的情况下修饰植物细胞的基因组中的核苷酸序列的方法改编为包括引入指导的Cpf1核酸内切酶复合物而不是引入指导的Cas核酸内切酶复合物的方法。可用于本文所披露的方法中的其他指导的核酸内切酶和核苷酸-蛋白质复合物包括WO2013/088446中描述的那些。

[0090] 核酸内切酶是切割多核苷酸链内的磷酸二酯键的酶,并且包括在特定位点切割DNA而不损害碱基的限制性核酸内切酶。限制性核酸内切酶包括I型、II型、III型、和IV型核酸内切酶,这些限制性核酸内切酶进一步包括亚型。在I型和III型系统中,甲基化酶和限制性酶活性二者均包含在单复合物中。核酸内切酶还包括大范围核酸酶,也称为归巢核酸内切酶(HE酶),该酶类似于限制性核酸内切酶,在特定识别位点处结合并且切割,然而对于大范围核酸酶,这些识别位点通常更长,约18bp或更长(于2012年3月22日提交的专利申请PCT/US12/30061)。基于保守的序列基序将大范围核酸酶分类为四个家族,这些家族是LAGLIDADG、GIY-YIG、H-N-H、和His-Cys box家族。这些基序参与金属离子的配位和磷酸二酯键的水解。HE酶的显著之处在于它们的长识别位点,并且还在于耐受其DNA底物中的一些序列多态性。对于大范围核酸酶的命名约定类似于对其他限制性核酸内切酶的约定。大范围核酸酶还分别特征在于针对由独立的ORF、内含子、和内含肽编码的酶的前缀F-、I-、或PI-。在重组过程中的一个步骤涉及在识别位点处或在识别位点附近的多核苷酸切割。该切割活性可用于产生双链断裂。对于位点特异性重组酶和它们的识别位点的综述,参见,Sauer (1994) *Curr Op Biotechnol* [生物技术新见] 5:521-7;以及Sadowski (1993) *FASEB* [美国实验生物学学会联合会杂志] 7:760-7。在一些实例中,重组酶来自Integrase或Resolvase家族。

[0091] TAL效应子核酸酶是新一类的序列-特异性核酸酶,该酶可以用于在植物或其他生

物体的基因组中的特定靶序列处造成双链断裂 (Miller等人 (2011) Nature Biotechnology [自然生物技术] 29:143-148)。锌指核酸酶 (ZFN) 是由锌指DNA结合结构域和双链-断裂-诱导剂结构域组成的工程化双链断裂诱导剂。识别位点特异性由锌指结构域赋予, 该锌指结构域通常包含两个、三个、或四个锌指, 例如具有C2H2结构, 然而其他锌指结构是已知的并且已经被工程化。锌指结构域适于设计特异性结合所选择的多核苷酸识别序列的多肽。ZFN包括连接至非特异性核酸内切酶结构域 (例如来自II型核酸内切酶, 例如FokI的核酸酶结构域) 的工程化DNA结合锌指结构域。另外的功能性可以融合到锌指结合结构域中, 这些另外的功能性包括转录激活子结构域、转录抑制子结构域、和甲基化酶。在一些实例中, 核酸酶结构域的二聚化是切割活性所需的。每个锌指在靶DNA中识别三个连续的碱基对。例如, 3指结构域识别9个连续核苷酸的序列, 由于该核酸酶的二聚化需要, 因此两组锌指三联体用于结合18个核苷酸的识别序列。

[0092] DNA双链断裂 (DSB) 技术 (ZFN、TALEN和CRISPR-Cas) 在学术研究、基因治疗和动植物育种计划中有广泛的应用。这些技术已成功用于在多种植物物种 (包括主要作物, 如玉蜀黍、小麦、大豆和稻) 中引入基因组修饰。植物基因组编辑受到当前转化和基因修饰方法、DNA递送效率和植物再生频率低的限制。与人类和动物系统相比, 在每个植物细胞周围的厚壁的存在从根本上影响植物转化和植物基因修饰方案。该细胞壁使得不可能使用转染或电穿孔, 其广泛用于哺乳动物基因组编辑实验中的核酸和/或蛋白质递送。出于这个原因, 植物转化和植物基因组修饰主要依赖于DNA载体上的指导RNA/Cas核酸内切酶试剂的农杆菌介导的递送和基因枪递送 (弹道传递)。结果, gRNA和Cas9表达盒经常被整合到基因组中并可能导致基因破坏、植物镶嵌和潜在的脱位点切割。虽然这些不希望的二次变化可以通过与野生型亲本植物回交几轮而分离开来, 但是该过程可能是耗时的, 尤其是对于具有复杂多倍体基因组和长繁育周期的作物, 例如但不限于大豆和小麦。如本文所述, 将Cas核酸内切酶和gRNA以RGEN复合物的形式递送到植物细胞中可以减轻许多这些副作用 (实例11-12)。在实例10中描述了通过将RGEN复合物递送到植物中而促进的出乎意料的高频率的NHEJ介导的诱变。考虑到使用RGEN复合物进行的这种高频率的诱变, 无DNA和无可选择标记的基因修饰可能成为产生基因敲除的实用方法。由于植物体细胞中HDR途径的频率低, 这种无DNA和无可选择标记的方法对于基因编辑和基因插入 (当与通过NHEJ进行的基因突变相比) 应用可能不太实用。此外, DNA分子通常整合到靶向的DSB位点, 这降低了基因编辑并且尤其是基因插入的效率。已经证明, 编码递送到植物细胞中的基因 (例如, Cas9、gRNA、可选择标记基因和性状基因) 的DNA载体具有共同整合到同一DSB位点中的倾向, 从而随着位点特异性性状基因插入而显著降低可用事件的发生率。将递送的DNA分子限制到供体DNA (例如, 具有同源臂的性状基因) 可以增加具有期望的基因型的事件的可能性。因此, 本文所述的可以在RGEN递送后被激活的破坏的 (失活的) 内源或预整合的可选择标记基因的概念可以使得用于基因编辑和基因插入的无DNA和无可选择标记的方法变得非常实用。

[0093] 使用本领域已知的任何方法 (例如, 但不限于, 粒子轰击、晶须介导的转化、农杆菌转化或局部施用), 可以将作为单链多核苷酸或双链多核苷酸的指导多核苷酸直接引入细胞中。指导RNA还可以通过引入 (经由例如但不限于粒子轰击或农杆菌转化等方法) 包含编码指导RNA的异源核酸片段的重组DNA分子被间接引入细胞中, 该重组DNA分子与能够在所述细胞中转录该指导RNA的特异性启动子有效地连接。特异性启动子可以是但不限于RNA聚

合酶III启动子,其允许具有精确定义的未修饰的5'-和3'-端的RNA转录(DiCarlo等人, *Nucleic Acids Res.* [核酸研究]41:4336-4343;Ma等人, *Mol. Ther. Nucleic Acids* [分子治疗-核酸]3:e161)。如本文所述,可以通过颗粒介导的递送将sgRNA直接递送到植物细胞中。基于本文所述的实验,技术人员现在可以预想任何其他直接递送方法(例如但不限于聚乙二醇(PEG)介导的对原生质体的转染、晶须介导的转化、电穿孔、粒子轰击、细胞穿透肽或介孔二氧化硅纳米颗粒(MSN)介导的直接蛋白质递送)都可以成功用于将gRNA递送到植物细胞中。

[0094] 该指导多核苷酸可以通过本领域已知的任何方法产生,包括化学合成指导多核苷酸(例如但不限于Hendel等人2015, *Nature Biotechnology* [自然生物技术]33,985-989)、体外产生的指导多核苷酸、和/或自剪接指导RNA(例如但不限于Xie等人2015, *PNAS* [美国国家科学院院刊]112:3570-3575)。

[0095] 术语“靶位点”、“靶序列”、“靶位点序列”、“靶DNA”、“靶基因座”、“基因组靶位点”、“基因组靶序列”、“基因组靶基因座”和“前间区”在本文中可互换地使用,并且是指多核苷酸序列,例如,但不限于,在细胞的染色体、附加体、转基因基因座或基因组中的任何其他DNA分子(包括染色体DNA、叶绿体DNA、线粒体DNA、质粒DNA)上的核苷酸序列,在这些序列处指导多核苷酸/Cas核酸内切酶复合物可以进行识别、结合并任选地产生切口或进行切割。靶位点可以是细胞基因组中的内源性位点,或者可替代地,靶位点可以与细胞异源并且从而不是天然存在于细胞的基因组中,或者与在自然界发生的位置相比,可以在异质基因组位置中找到靶位点。如本文所用,术语“内源性靶序列”和“天然靶序列”在本文中可互换使用,指的是对细胞基因组来说是内源的或天然的、并且位于细胞基因组中该靶序列的内源或天然位置处的靶序列。细胞包括但不限于人类、非人类、动物、细菌、真菌、昆虫、酵母、非常规酵母和植物细胞,以及通过本文所述的方法产生的植物和种子。“人工靶位点”或“人工靶序列”在本文中可互换地使用,并且意指已经引入细胞基因组中的靶序列。这样的人工靶序列可以在序列上与细胞基因组中的内源性或天然靶序列相同,但是位于细胞基因组中的不同位置(即,非内源性的或非天然的位置)处。

[0096] “改变的靶位点”、“改变的靶序列”、“修饰的靶位点”、“修饰的靶序列”在本文中可互换地使用,并且意指如本文披露的靶序列,当与非改变的靶序列相比时,该靶序列包括至少一个改变。例如,此类“改变”包括:(i)至少一个核苷酸的替代、(ii)至少一个核苷酸的缺失、(iii)至少一个核苷酸的插入、或(iv) (i)-(iii)的任何组合。

[0097] 靶DNA序列(靶位点)的长度可以变化,并且包括例如为至少12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30或更多个核苷酸长度的靶位点。还有可能靶位点可以是回文的,即,一条链上的序列与在互补链上以相反方向的读取相同。切口/切割位点可以在靶序列内,或者切口/切割位点可以在靶序列之外。在另一种变异中,切割可以发生在彼此正好相对的核苷酸位置处,以产生平端切割,或者在其他情况下,切口可以交错以产生单链突出端,也称为“粘性末端”,其可以是5'突出端抑或或3'突出端。还可以使用基因组靶位点的活性变体。这样的活性变体可以包含与给定靶位点至少65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高的序列同一性,其中该活性变体保留生物活性,因此能够被Cas核酸内切酶识别和切割。测量由核酸内切酶引起的靶位点的单链或双链断裂的测定是本领域已知的,并且通常测量试剂在包含识别位点的

DNA底物上的总体活性和特异性。

[0098] 本文中的“前间区邻近基序”(PAM)是指与由指导多核苷酸/Cas核酸内切酶系统识别的靶序列(前间区)邻近的短核苷酸序列。如果靶DNA序列不在PAM序列后面,则Cas核酸内切酶可能无法成功识别该靶DNA序列。本文中的PAM的序列和长度可以取决于所使用的Cas蛋白或Cas蛋白复合物而不同。该PAM序列可以是任何长度,但典型地是1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸长度。

[0099] 术语“靶向”、“基因靶向”和“DNA靶向”在本文中可互换地使用。本文中的DNA靶向可能是在特定的DNA序列(例如细胞的染色体或质粒)中特异性引入敲除、编辑、或敲入。通常,DNA靶向可以通过在具有与合适的多核苷酸组分缔合的Cas蛋白的细胞中的特异性DNA序列处切割一条或两条链来进行。这样的DNA切割,如果是双链断裂(DSB),可以促进NHEJ或HDR过程,这可能导致靶位点处的修饰。

[0100] 术语“敲除”、“基因敲除”和“遗传敲除”在本文中可互换地使用。敲除代表通过用Cas蛋白靶向而部分或完全不起作用的细胞的DNA序列;敲除之前这样的DNA序列例如可以已经编码氨基酸序列或者可以具有调节功能(例如,启动子)。可以通过插入缺失(通过NHEJ在靶DNA序列中插入或缺失核苷酸碱基),或通过特异性去除在靶向位点处或其附近处降低或完全破坏序列功能的序列来产生敲除。

[0101] 指导多核苷酸/Cas核酸内切酶系统可以与共同递送的多核苷酸修饰模板组合使用以允许编辑(修饰)目的基因组核苷酸序列。(也参见2015年3月19日公开的美国专利申请US 2015-0082478 A1和2015年2月26日公开的W02015/026886 A1,两者均通过引用以其全文特此结合)。

[0102] “修饰的核苷酸”或“编辑的核苷酸”是指当与其非修饰的核苷酸序列相比时,包含至少一个改变的目的核苷酸序列。例如,此类“改变”包括:(i)至少一个核苷酸的替代、(ii)至少一个核苷酸的缺失、(iii)至少一个核苷酸的插入、或(iv) (i)-(iii)的任何组合。

[0103] 术语“多核苷酸修饰模板”包括,当与待编辑的核苷酸序列相比时,包含至少一个核苷酸修饰的多核苷酸。核苷酸修饰可以是至少一个核苷酸取代、添加或缺失。任选地,多核苷酸修饰模板可以进一步包含位于至少一个核苷酸修饰侧翼的同源核苷酸序列,其中侧翼同源核苷酸序列为待编辑的希望的核苷酸序列提供了充足同源性。

[0104] 可以通过本领域已知的任何方法将多核苷酸修饰模板引入细胞中,所述方法例如但不限于瞬时引入方法、转染、电穿孔、显微注射、颗粒介导的递送、局部施用、晶须介导的递送、经由细胞穿透肽的递送或介孔二氧化硅纳米颗粒(MSN)介导的直接递送。

[0105] 可以将多核苷酸修饰模板作为单链多核苷酸分子、双链多核苷酸分子或作为环状DNA(载体DNA)的一部分引入细胞中。该多核苷酸修饰模板还可以与指导RNA和/或Cas核酸内切酶进行系链。系链的DNA可以允许共定位靶标和模板DNA,可用于基因组编辑和靶向的基因组调控,并且还可以用于靶向有丝分裂后期细胞,在这些细胞中内源性HR机制的功能预计会大大降低(Mali等人2013Nature Methods[自然方法]第10卷:957-963。)该多核苷酸修饰模板可以瞬时地存在于细胞中,或可以经由病毒复制子引入。

[0106] 待编辑的核苷酸可以位于由Cas核酸内切酶识别和切割的靶位点的内部或外部。在一个实施例中,该至少一个核苷酸修饰不是由Cas核酸内切酶识别和切割的靶位点上的修饰。在另一个实施例中,该待编辑的至少一个核苷酸和基因组靶位点之间有至少1、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、30、40、50、100、200、300、400、500、600、700、900或1000个核苷酸。

[0107] 可以使用任何可用的基因编辑方法来完成基因组编辑。例如,可以通过向宿主细胞中引入含有宿主细胞基因组内的基因的靶向修饰的多核苷酸修饰模板(有时也称为基因修复寡核苷酸)来完成基因编辑。用于这类方法的多核苷酸修饰模板可以是单链或双链。这类方法的实例通常描述于例如美国公开号2013/0019349中。

[0108] 基于本文所述的实验,技术人员现在可以预想任何其他直接递送方法(例如但不限于聚乙二醇(PEG)介导的对原生质体的转染、晶须介导的转化、电穿孔、粒子轰击、细胞穿透肽或介孔二氧化硅纳米颗粒(MSN)介导的直接蛋白质递送)都可以成功用于将多核苷酸修饰模板递送到植物细胞中。

[0109] 在一些实施例中,可以通过在所期望的改变附近的基因组中在限定位置诱导双链断裂(DSB)来促进基因编辑。可以使用可用的任何DSB诱导剂诱导DSB,该诱导剂包括但不限于,TALEN、大范围核酸酶、锌指核酸酶、核酸指导的核酸内切酶系统(例如,Cas9-gRNA系统(基于细菌性CRISPR-Cas系统))等。在一些实施例中,可以将DSB的引入与多核苷酸修饰模板的引入组合。

[0110] 编辑组合有DSB和修饰模板的基因组序列的过程通常包括:向宿主细胞引入DSB诱导剂或编码DSB诱导剂的核酸(识别染色体序列中的靶序列并且能够诱导基因组序列中的DSB),和与待编辑的核苷酸序列相比时包含至少一个核苷酸改变的至少一个多核苷酸修饰模板。多核苷酸修饰模板还可以包含侧翼于该至少一个核苷酸改变的核苷酸序列,其中侧翼序列与侧翼于DSB的染色体区域基本同源。例如,以下申请中已经描述了使用DSB-诱导剂(如Cas9-gRNA复合物)进行基因组编辑:2015年3月19日公开的美国专利申请US 2015-0082478 A1、2015年2月26日公开的W02015/026886 A1、2014年7月7日公开的美国申请62/023246,和2014年8月13日公开的美国申请62/036,652,将其全部通过引用结合在此。

[0111] 术语“敲入”、“基因敲入”、“基因插入”和“遗传敲入”在本文中可互换地使用。敲入表示通过用Cas蛋白靶向在细胞中的特异性DNA序列处进行的DNA序列的替换或插入(通过HR,其中还使用合适的供体DNA多核苷酸)。敲入的实例是异源氨基酸编码序列在基因的编码区中的特异性插入,或转录调控元件在遗传基因座中的特异性插入。

[0112] 可以采用不同方法和组合物来获得细胞或生物体,该细胞或生物体具有插入针对Cas核酸内切酶的靶位点中的目的多核苷酸。此类方法可以采用同源重组以提供目的多核苷酸在靶位点处的整合。在提供的一个方法中,将在供体DNA构建体中的目的多核苷酸提供至生物体细胞。

[0113] 如本文所用,“供体DNA”包括提及DNA构建体,其包括待插入到Cas核酸内切酶的靶位点的目的多核苷酸。供体DNA构建体可以进一步包含位于目的多核苷酸侧翼的同源的第一区域和第二区域。供体DNA的同源的第一区域和第二区域分别与存在于细胞或生物体基因组的靶位点中或位于该靶位点侧翼的第一和第二基因组区域共享同源性。供体DNA可以与指导多核苷酸和/或Cas核酸内切酶进行系链。系链的供体DNA可以允许共定位靶标和供体DNA,可用于基因组编辑和靶向的基因组调控,并且还可以用于靶向有丝分裂后期细胞,在这些细胞中内源性HR机制的功能预计会大大降低(Mali等人2013Nature Methods[自然方法]第10卷:957-963。)

[0114] “同源性”意指相似的DNA序列。例如，在供体DNA上发现的“与基因组区域同源的区域”是与细胞或生物体基因组中给定的“基因组序列”具有相似序列的DNA的区域。同源的区域可以具有足以促进在切割的靶位点处的同源重组的任何长度。例如，同源的区域可以包括至少5-10、5-15、5-20、5-25、5-30、5-35、5-40、5-45、5-50、5-55、5-60、5-65、5-70、5-75、5-80、5-85、5-90、5-95、5-100、5-200、5-300、5-400、5-500、5-600、5-700、5-800、5-900、5-1000、5-1100、5-1200、5-1300、5-1400、5-1500、5-1600、5-1700、5-1800、5-1900、5-2000、5-2100、5-2200、5-2300、5-2400、5-2500、5-2600、5-2700、5-2800、5-2900、5-3000、5-3100或更多个碱基，这样使得同源的区域具有足够的同源性以与相应的基因组区域进行同源重组。“足够的同源性”表示两个多核苷酸序列具有足够的结构相似性以充当同源重组反应的底物。结构相似性包括每个多核苷酸片段的总长度以及多核苷酸的序列相似性。序列相似性可以通过在序列的整个长度上的百分比序列同一性和/或通过包含局部相似性(例如具有100%序列同一性的连续核苷酸)的保守区域以及在序列长度的一部分上的百分比序列同一性来描述。

[0115] 由靶标和供体多核苷酸共享的同源性或序列同一性的量可以变化，并且包括总长度和/或在约1-20bp、20-50bp、50-100bp、75-150bp、100-250bp、150-300bp、200-400bp、250-500bp、300-600bp、350-750bp、400-800bp、450-900bp、500-1000bp、600-1250bp、700-1500bp、800-1750bp、900-2000bp、1-2.5kb、1.5-3kb、2-4kb、2.5-5kb、3-6kb、3.5-7kb、4-8kb、5-10kb、或多达并包括靶位点的总长度的范围内具有单位整数值的区域。这些范围包括该范围内的每个整数，例如1-20bp的范围包括1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19和20bp。同源性的量也可以通过在两个多核苷酸的完整比对长度上的百分比序列同一性来描述，其包括约至少50%、55%、60%、65%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%的百分比序列同一性。足够的同源性包括多核苷酸长度、总体百分比序列同一性，和任选地连续核苷酸的保守区域或局部百分比序列同一性的任何组合，例如，足够的同源性可以被描述为与靶标基因座的区域具有至少80%序列同一性的75-150bp的区域。还可以通过用来在高严格条件下特异性杂交的两个多核苷酸的预测能力来描述足够的同源性，参见例如Sambrook等人，(1989) *Molecular Cloning: A Laboratory Manual* [分子克隆：实验室手册] (Cold Spring Harbor Laboratory Press, NY [纽约冷泉港实验室出版社])；*Current Protocols in Molecular Biology* [分子生物学现代方案]，Ausubel等人，编辑(1994) *Current Protocols* [实验室指南] (Greene Publishing Associates, Inc. [格林出版合伙公司]和John Wiley&Sons, Inc. [约翰威利父子公司])；以及Tijssen(1993) *laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes* [生物化学和分子生物学中的实验室技术--与核酸探针杂交] (Elsevier [爱思唯尔出版社]，纽约)。

[0116] 如本文所用，“基因组区域”是存在于靶位点任一侧上的细胞的基因组中的染色体的区段，或者可替代地，还包含靶位点的一部分。基因组区域可以包含至少5-10、5-15、5-20、5-25、5-30、5-35、5-40、5-45、5-50、5-55、5-60、5-65、5-70、5-75、5-80、5-85、5-90、5-95、5-100、5-200、5-300、5-400、5-500、5-600、5-700、5-800、5-900、5-1000、5-1100、5-1200、5-1300、5-1400、5-1500、5-1600、5-1700、5-1800、5-1900、5-2000、5-2100、5-2200、5-

2300、5-2400、5-2500、5-2600、5-2700、5-2800、5-2900、5-3000、5-3100或更多个碱基,使得基因组区域具有足够的同源性以与相应的同源的区域进行同源重组。

[0117] 目的多核苷酸和/或性状可以在复合物性状基因座中堆叠在一起,如在2013年10月3日公开的US-2013-0263324-A1和2013年1月24日公开的PCT/US13/22891中所述,将这两个申请通过引用特此结合。本文所述的指导多核苷酸/Cas9核酸内切酶系统提供了用来产生双链断裂并允许将性状在复杂性状基因座中堆叠的有效系统。

[0118] 指导多核苷酸/Cas核酸内切酶系统可以通过将一种或多种指导多核苷酸、一种Cas核酸内切酶以及任选地一种或多种供体DNA引入植物细胞中而将一种或多种目的多核苷酸或一种或多种目的形状引入到一个或多个靶位点中。(如在2015年3月19日公开的美国专利申请US-2015-0082478-A1中所述,通过引用结合在此)。可育的植物可以从在所述一个或多个靶位点处包含改变的植物细胞产生,其中该改变选自下组,该组由以下各项组成:(i)至少一个核苷酸的替代、(ii)至少一个核苷酸的缺失、(iii)至少一个核苷酸的插入、或(iv) (i)-(iii)的任何组合。包含这些改变的靶位点的植物可以与包含相同复杂性状基因座中的至少一个目的基因或目的性状的植物进行杂交,从而进一步堆叠所述复杂性状基因座中的性状(还参见2013年10月3日公开的US-2013-0263324-A1以及2013年1月24日公开的PCT/US13/22891,通过引用结合在此)。

[0119] 在给定的基因组区域和在供体DNA上发现的相应的同源的区域之间的结构相似性可以是允许同源重组发生的任何程度的序列同一性。例如,由供体DNA的“同源的区域”和生物体基因组的“基因组区域”共享的同源性或序列同一性的量可以是至少50%、55%、60%、65%、70%、75%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%序列同一性,这样使得序列进行同源重组。

[0120] 供体DNA上的同源的区域可以与靶位点侧翼的任何序列具有同源性。虽然在一些实施例中,同源的区域与紧邻靶位点侧翼的基因组序列共享显著的序列同源性,但是应当认识到同源的区域可以被设计为与可能更靠近靶位点的5'或3'的区域具有足够的同源性。在又其他实施例中,同源的区域还可以与靶位点的片段以及下游基因组区域具有同源性。在一个实施例中,第一同源的区域进一步包含靶位点中的第一片段,并且第二同源的区域包含靶位点中的第二片段,其中第一片段和第二片段不同。

[0121] 一旦在DNA中诱导双链断裂,则细胞的DNA修复机制被激活来修复断裂。非同源末端连接(NHEJ)途径是用来将断裂的末端结合在一起的最常见的修复机制(Bleuyard等人,(2006)DNA Repair[DNA修复]5:1-12)。染色体的结构完整性通常通过修复来保持,但是缺失、插入、或其他重排是可能的。一个双链断裂的两个末端是NHEJ最普遍的底物(Kirik等人,(2000)EMBO J[欧洲分子生物学学会杂志]19:5562-6),然而如果发生两种不同的双链断裂,则来自不同断裂的游离端可以被连接,并且导致染色体缺失(Siebert和Puchta,(2002)Plant Cell[植物细胞]14:1121-31),或在不同的染色体之间的染色体易位(Pacher等人,(2007)Genetics[遗传学]175:21-9)。易错DNA修复机制可以在双链断裂位点处产生突变。非同源末端连接(NHEJ)途径是用来将断裂的末端结合在一起的最常见的修复机制(Bleuyard等人,(2006)DNA Repair[DNA修复]5:1-12)。

[0122] 可替代地,该双链断裂可以通过同源DNA序列之间的同源重组(HR)来修复。一旦双

链断裂周围的序列被改变,例如通过涉及双链断裂的成熟的外切核酸酶活性,则基因转换途径可以恢复原始结构(如果有同源序列的话),例如非分裂的体细胞中的同源染色体,或DNA复制后的姊妹染色单体(Molinier等人,(2004) *Plant Cell* [植物细胞]16:342-52)。异位的和/或表观遗传的DNA序列还可以充当用于同源重组的DNA修复模板(Puchta,(1999) *Genetics* [遗传学]152:1173-81)。还可以将附加体DNA分子连接至双链断裂中,例如,将T-DNA整合至染色体双链断裂中(Chilton和Que,(2003) *Plant Physiol* [植物生理学]133:956-65;Salomon和Puchta,(1998) *EMBO J* [欧洲分子生物学学会杂志]17:6086-95)。

[0123] 如本文所用,“同源重组(HR)”包括在同源的位点处的两个DNA分子之间的DNA片段的交换。同源重组的频率受多个因素影响。不同的生物体相对于同源重组的量和同源与非同源重组的相对比例而变化。通常,同源的区域长度影响同源重组事件的频率:同源的区域越长,频率越高。为观察同源重组而需要的同源区域的长度也是随物种而异的。在许多情况下,已经利用了至少5kb的同源性,但已经观察到具有仅25-50bp的同源性的同源重组。参见,例如,Singer等人,(1982) *Cell* [细胞]31:25-33;Shen和Huang,(1986) *Genetics* [遗传学]112:441-57;Watt等人,(1985) *Proc. Natl. Acad. Sci. USA* [美国科学院院报]82:4768-72;Sugawara和Haber,(1992) *Mol Cell Biol* [分子细胞生物学]12:563-75;Rubnitz和Subramani,(1984) *Mol Cell Biol* [分子细胞生物学]4:2253-8;Ayares等人,(1986) *Proc. Natl. Acad. Sci. USA* [美国科学院院报]83:5199-203;Liskay等人,(1987) *Genetics* [遗传学]115:161-7。

[0124] 同源-定向修复(HDR)是在细胞中用来修复双链DNA和单链DNA断裂的机制。同源-定向修复包括同源重组(HR)和单链退火(SSA)(Lieber.2010 *Annu. Rev. Biochem* [生物化学年鉴].79:181-211)。HDR的最常见形式称为同源重组(HR),其在供体和受体DNA之间具有最长的序列同源性要求。HDR的其他形式包括单链退火(SSA)和断裂诱导的复制,并且这些需要相对于HR更短的序列同源性。缺口(单链断裂)处的同源-定向修复可以经由与在双链断裂处的HDR不同的机制发生(Davis和Maizels. *PNAS* [美国国家科学院院刊] (0027-8424), 111 (10), 第E924-E932页)。

[0125] 例如,通过同源性同源定向修复(HDR)来改变植物细胞的基因组对于遗传工程而言是有力的工具。尽管在高等植物中的低频率的同源重组,但是存在植物内源基因的成功同源重组的少数实例。用于植物中同源重组的参数主要已经通过挽救引入的截短的可选择标记基因进行了研究。在这些实验中,同源DNA片段通常在0.3kb至2kb之间。针对同源重组观察到的频率是在 $10^{-4}$ 至 $10^{-5}$ 数量级。参见,例如Halfter等人,(1992) *Mol Gen. Genet.* [分子和普通遗传学]231:186-93;Offringa等人,(1990) *EMBO J* [欧洲分子生物学学会杂志]9:3077-84;Offringa等人,(1993) *Proc. Natl. Acad. Sci. USA* [美国科学院院报]90:7346-50;Paszkowski等人,(1988) *EMBO J* [欧洲分子生物学学会杂志]7:4021-6;Hourda和Paszkowski,(1994) *Mol Gen Genet* [分子和普通遗传学]243:106-11;以及Risseuw等人,(1995) *Plant J* [植物杂志]7:109-19。

[0126] DNA双链断裂似乎是刺激同源重组途径的有效因子(Puchta等人,(1995) *Plant Mol Biol* [植物分子生物学]28:281-92;Tzfira和White,(2005) *Trends Biotechnol* [生物技术趋势]23:567-9;Puchta,(2005) *J Exp Bot* [实验植物学杂志]56:1-14)。使用DNA断裂剂,在植物中的人工构建的同源DNA重复序列之间观察到同源重组的两倍至九倍的增加

(Puchta等人, (1995) Plant Mol Biol [植物分子生物学]28:281-92)。在玉蜀黍原生质体中,用线状DNA进行的实验证实了在质粒之间增强的同源重组(Lyznik等人, (1991) Mol Gen Genet [分子和普通遗传学]230:209-18)。

[0127] 可以通过本领域已知的任何手段引入供体DNA。例如,提供了具有靶位点的植物。可以通过本领域已知的任何递送方法(包括例如农杆菌介导的转化、晶须介导的转化或生物射弹粒子轰击)来提供供体DNA。该供体DNA可以瞬时地存在于细胞中,或可以经由病毒复制子引入。在Cas核酸内切酶和靶位点的存在下,供体DNA被插入植物基因组中。

[0128] 如本文所述,可以通过颗粒介导的递送将供体DNA直接递送到植物细胞中。基于本文所述的实验,技术人员现在可以预想任何其他直接递送方法(例如但不限于聚乙二醇(PEG)介导的对原生质体的转染、电穿孔、粒子轰击、晶须介导的递送、细胞穿透肽或介孔二氧化硅纳米颗粒(MSN)介导的直接蛋白质递送)都可以成功用于将供体DNA递送到植物细胞中。

[0129] 指导RNA/Cas核酸内切酶系统的另外的用途已进行了描述(参见2015年3月19日公开的美国专利申请US 2015-0082478A1、2015年2月26日公开的W02015/026886 A1、2015年2月26日公开的US 2015-0059010 A1、2014年7月7日提交的美国申请62/023246,和2014年8月13日提交的美国申请62/036,652,将其全部通过引用结合在此),并包括,但不限于修饰或取代目的核苷酸序列(如调节元件)、目的多核苷酸插入、基因敲除、基因敲入、剪接位点的修饰和/或引入交替剪接位点、编码目的蛋白的核苷酸序列的修饰、氨基酸和/或蛋白质融合物、以及通过在目的基因中表达反向重复序列引起的基因沉默。

[0130] 在本文中进一步描述了目的多核苷酸,并且包括反映涉及作物发育的那些的商业市场和利益的多核苷酸。感兴趣的作物和市场发生变化,以及随着发展中国家打开国际市场,新作物和技术也将出现。此外,随着我们对农学性状和特征(例如产量和杂种优势增加)的理解逐渐深入,对用于遗传工程的基因的选择将会相应变化。

[0131] 进一步提供了用于鉴定至少一个植物细胞的方法,该植物细胞在其基因组中包含在靶位点处整合的目的多核苷酸。可以使用多种方法来鉴定在靶位点处或靶位点附近插入到基因组中的那些植物细胞,而不使用可筛选的标记表型。此类方法可被认为是直接分析靶序列以检测靶序列中的任何变化,包括但不限于PCR方法、测序方法、核酸酶消化、DNA印迹法、及其任何组合。参见,例如,美国专利申请12/147,834,将该申请通过引用结合在此至本文所述方法所必需的程度。该方法还包括从植物细胞重新获得包含整合至其基因组中的目的多核苷酸的植物。该植物可以是不育的或可育的。应当认识到,可以提供任何目的多核苷酸,将该多核苷酸在靶位点处整合到植物的基因组中,并在植物中表达。

[0132] 目的多核苷酸/多肽包括但不限于,除草剂-抗性编码序列、杀昆虫编码序列、杀线虫编码序列、抗微生物编码序列、抗真菌编码序列、抗病毒编码序列、非生物和生物胁迫耐受性编码序列、或修饰植物性状(例如产量、谷粒质量、营养成分、淀粉质量和数量、固氮和/或氮利用、脂肪酸、以及含油量和/或油组成)的序列。更具体的目的多核苷酸包括但不限于:改进作物产量的基因、改进作物合意性的多肽、编码赋予对非生物胁迫(例如干旱、氮、温度、盐度、毒性金属、或痕量元素)的抗性的蛋白质,或赋予对毒素(例如杀有害生物剂和除草剂)、或对生物胁迫(例如真菌、病毒、细菌、昆虫和线虫的攻击以及与这些生物体相关的疾病的发展)的抗性的那些蛋白质的基因。目的基因的一般类别包括,例如涉及信息的那

些基因(例如锌指),涉及通讯的那些基因(例如激酶),以及涉及管家的那些基因(例如热休克蛋白)。转基因的更具体类别包括,例如,编码对农学、昆虫抗性、疾病抗性、除草剂抗性、可育性或不育性、谷粒特征、和商业产品而言重要的性状的基因。目的基因通常包括涉及油、淀粉、碳水化合物、或营养代谢的那些,以及可以与如本文所述的其他性状(例如但不限于除草剂抗性)堆叠或组合使用的影响籽粒大小、蔗糖负载量等的那些基因。

[0133] 除了使用传统的育种方法之外,还可通过遗传方式改变农学上重要的性状(例如油、淀粉、和蛋白质含量)。修饰包括增加油酸、饱和及不饱和油的含量、增加赖氨酸和硫的水平、引入必需氨基酸、以及还有对淀粉的修饰。在美国专利号5,703,049、5,885,801、5,885,802、和5,990,389中描述了Hordomionin蛋白质修饰,将这些专利通过引用结合在此。

[0134] 目的多核苷酸序列可以编码涉及引入疾病或有害生物抗性的蛋白质。“疾病抗性”或“有害生物抗性”意在是植物避免为植物-病原体相互作用后果的有害症状的发生。有害生物抗性基因可以编码对严重影响产量的有害生物的抗性,这些有害生物例如根虫、切根虫、欧洲玉米螟等。疾病抗性基因和昆虫抗性基因,例如用于抗细菌保护的溶菌酶或天蚕杀菌肽,或用于抗真菌保护的蛋白质,例如防御素、葡聚糖酶、或几丁质酶,或用于控制线虫或昆虫的苏云金芽孢杆菌内毒素、蛋白酶抑制剂、胶原酶、凝集素、或糖苷酶,均是有用的基因产物的实例。编码疾病抗性性状的基因包括解毒基因,例如抗伏马毒素(美国专利号5,792,931);无毒力(avr)和疾病抗性(R)基因(Jones等人(1994) Science[科学]266:789;Martin等人(1993) Science[科学]262:1432;和Mindrinis等人(1994) Cell[细胞]78:1089);等。昆虫抗性基因可以编码对严重影响产量的有害生物的抗性,这些有害生物例如根虫、切根虫、欧洲玉米螟等。此类基因包括,例如,苏云金芽孢杆菌毒性蛋白基因(美国专利号5,366,892;5,747,450;5,736,514;5,723,756;5,593,881;和Geiser等人(1986) Gene[基因]48:109);等。

[0135] “除草剂抗性蛋白”或由“除草剂抗性编码核酸分子”表达生成的蛋白质包括这样的蛋白质,其赋予细胞与未表达该蛋白质的细胞相比耐受更高浓度除草剂的能力,或赋予细胞与未表达该蛋白质的细胞相比对某种浓度的除草剂耐受更长时段的能力。除草剂抗性性状可通过如下基因引入进植物中:编码对起到抑制乙酰乳酸合酶(ALS,也称为AHAS)的作用的除草剂(特别是磺酰脲(sulfonylurea)(UK:磺酰脲(sulphonylurea))类除草剂)的抗性的基因、编码对起到抑制谷氨酰胺合酶的作用的除草剂(例如草丁膦或basta)的抗性的基因(例如bar基因)、编码对草甘膦的抗性的基因(例如EPSP合酶基因和GAT基因)、编码对HPPD抑制剂的抗性的基因(例如HPPD基因)或本领域已知的其他此类基因。参见,例如,美国专利号7,626,077、5,310,667、5,866,775、6,225,114、6,248,876、7,169,970、6,867,293,以及美国临时申请号61/401,456,其各自通过引用结合在此。bar基因编码对除草剂basta的抗性,nptII基因编码对抗生素卡那霉素和遗传霉素的抗性,以及ALS-基因突变体编码对除草剂氯磺隆的抗性。

[0136] 如本文所用,“磺酰脲耐受性多肽”包含当在植物中表达时对至少一种磺酰脲赋予耐受性的任何多肽。磺酰脲类除草剂通过阻断乙酰乳酸合成酶(ALS)(也称为乙酰羟酸合酶(AHAS))来抑制高等植物的生长。在ALS中含有特定突变(例如S4和/或HRA突变)的植物耐受磺酰脲类除草剂。磺酰脲耐受性植物的生产更充分地描述于美国专利号5,605,011;5,013,659;5,141,870;5,767,361;5,731,180;5,304,732;4,761,373;5,331,107;5,928,937;和

5,378,824;以及国际公开WO 96/33270(出于全部目的,将以上各项通过引用以其全文结合在此)中,以及在Tan等人2005.Imidazolinone-tolerant crops:history,current status and future.[耐咪唑啉酮的作物:历史、现状和未来]Pest Manag Sci[有害生物管理科学]61:246-257中。磺酰脲耐受性多肽可以由例如ALS的SuRA或SuRB基因座编码。在具体的实施例中,该ALS抑制剂耐受性多肽包含C3ALS突变体、HRA ALS突变体、S4突变体或S4/HRA突变体或其任何组合。已知ALS中的不同突变对不同除草剂和除草剂的组(和/或亚组)赋予耐受性;参见,例如,Tranel和Wright(2002)Weed Science[杂草科学]50:700-712。还参见美国专利号5,605,011、5,378,824、5,141,870和5,013,659,其各自通过引用以其全文结合在此。在一个实施例中发现ALS中的HRA突变特别有用。该突变导致相比于野生型蛋白对至少一种磺酰脲化合物具有抗性的乙酰乳酸合酶多肽的产生。

[0137] 编码磺酰脲耐受性多肽的基因被称为磺酰基耐受性基因或磺酰基抗性基因。术语磺酰基耐受性基因或磺酰基抗性基因在本文中可互换使用。

[0138] 破坏的磺酰脲抗性(ALS)基因是指被破坏的基因,其相应的未破坏的基因编码磺酰脲耐受性多肽,将该多肽修饰这样使得其基因产物不再编码功能性磺酰脲耐受性多肽。

[0139] 也可以将磺酰脲响应性阻遏物系统的组分(如2012年9月4日发布的US 8,257,956中所述)引入植物基因组中,以在所述植物中产生阻遏物/操纵子/诱导物系统,在所述植物中多肽可以特异性结合操作子,其中该特异性结合由磺酰脲类化合物调节。

[0140] 不育性基因还可以在表达盒中编码,并为物理去雄提供替代方案。以此类方式使用的基因的实例包括雄性可育性基因,例如MS26(参见例如美国专利7,098,388、7,517,975、7,612,251)、MS45(参见例如美国专利5,478,369、6,265,640)、或MSCA1(参见例如美国专利7,919,676)。玉蜀黍植物(玉米)可以通过自花授粉和异花传粉技术二者来进行育种。玉蜀黍在同一植物上具有位于雄穗上的雄花和位于雌穗上的雌花。玉蜀黍可自花授粉(“自交”)或异花授粉。当风将花粉从雄穗吹到从初期(incipient)雌穗顶端伸出的须上时,玉蜀黍中就发生了自然授粉。授粉可以通过本领域技术人员已知的技术容易地控制。对玉蜀黍杂交体的开发需要开发纯合纯合近交系,将这些纯合近交系杂交,以及对这些杂交的评价。谱系育种和轮回选择是用于从种群开发近交系的两种育种方法。育种程序将来自两个或更多个近交系或不同基础广泛来源的所希望的性状组合进育种库,从该育种库中通过自交和对所希望表型的选择从而开发出新的近交系。杂交玉蜀黍品种是两种此类近交系的杂交,每种都可能具有由另一种所缺少的一个或多个所希望的特征,或补足另一种。新的近交系与其他近交系杂交,并对来自这些杂交的杂交体进行评价,以确定哪些具有商业潜力。第一代的杂交后代被指定为F1。F1杂种比其近交的亲本更有活力。这种杂交活力或杂种优势可以按许多方式表现,包括增加的营养生长和增加的产量。

[0141] 杂交玉蜀黍种子可以通过结合人工去雄在内的雄性不育系统来生产。为了产生杂交种子,将雄花穗从生长的雌性近交亲本中除去,这些雌性近交亲本可以与雄性近交亲本以不同交替行排列模式种植。因此引入,如果与外来玉蜀黍花粉的来源有足够的隔离,则雌性近交系的穗将只能来自雄性近交系的花粉进行受精。因此,所得的种子是杂种(F1),并将形成杂种植物。

[0142] 影响植物发育的田间变化可以导致在对雌性亲本的人工去雄完成后的植物抽穗。或,在去雄过程中雌性近交植物雄穗可能没有完全去除。无论如何,结果是雌性植物将成功

地散播花粉,并且一些雌性植物将会自花授粉。这将导致雌性近交系的种子与正常产生的杂交种子一起被收获。雌性近交系种子不显示杂种优势,并且因此不如F1种子多产。此外,雌性近交系种子的存在可以代表生产杂交体的公司的种质安全风险。

[0143] 可替代地,雌性近交系可以通过机器进行机械地去雄。机械去雄和手工去雄的可靠性大致相同,但更快并且成本更低。然而,相比手工去雄,大多数去雄机器会对植物造成更多伤害。因此,目前没有令人完全满意的去雄形式,并且仍存在对替代方案的需求,这些替代方案进一步降低生产成本,并且消除在杂交种子的生产中的雌性亲本的自花授粉。

[0144] 在植物中引起雄性不育的突变具有可用于作物植物(例如玉蜀黍)的杂交种子生产的方法的潜力,并且可以通过消除对耗费大量劳动力从用作杂交亲本的母本植物上去除雄花(也称为去雄)的需要来降低生产成本。在玉蜀黍中引起雄性不育的突变已通过多种方法产生,这些方法例如X射线或紫外线照射、化学处理、或转座元件插入(ms23、ms25、ms26、ms32)(Chaubal等人(2000) *Am J Bot*[美国植物学杂志]87:1193-1201)。通过可育性/不育性“分子转换”对可育性基因进行条件调节可以增强用于设计新的雄性不育系统来进行作物改进的选项(Unger等人(2002) *Transgenic Res*[转基因研究]11:455-465)。

[0145] 此外,认识到目的多核苷酸还可以包括与针对目的所靶向的基因序列的信使RNA(mRNA)的至少一部分互补的反义序列。构建反义核苷酸以与相应的mRNA杂交。可以对该反义序列作出修饰,只要该序列与相应的mRNA杂交并干扰相应的mRNA的表达。在该方式中,可以使用与相应的反义序列具有70%、80%、或85%序列同一性的反义构建体。此外,反义核苷酸的部分可以用来破坏该靶基因的表达。通常,可以使用至少50个核苷酸、100个核苷酸、200个核苷酸、或更多个核苷酸的序列。

[0146] 此外,目的多核苷酸还可以按有义取向来使用从而抑制植物中内源基因的表达。以有义取向使用多核苷酸用于抑制植物中基因表达的方法是本领域已知的。这些方法通常涉及用包含启动子的DNA构建体的转化植物,该启动子有效地连接到至少一部分的对应于该内源基因的转录物的核苷酸序列上,驱动在植物中的表达。通常,此类核苷酸序列与内源基因的转录物的序列具有实质性的序列同一性,通常大于约65%序列同一性、约85%序列同一性、或大于约95%序列同一性。参见美国专利号5,283,184和5,034,323;通过引用结合在此。

[0147] 目的多核苷酸还可以是表型标记。

[0148] 重组DNA分子、目的DNA序列、和目的多核苷酸可以包括用于基因沉默的一个或多个DNA序列。用于涉及在植物中DNA序列表达的基因沉默的方法是本领域已知的,包括但不限于,共抑制、反义抑制、双链RNA(dsRNA)干扰、发夹RNA(hpRNA)干扰、包含内含子的发夹RNA(ihpRNA)干扰、转录基因沉默、以及微小RNA(miRNA)干扰。

[0149] 如本文所用,“核酸”意指多核苷酸,并且包括脱氧核糖核苷酸或核糖核苷酸碱基的单链或双链聚合物。核酸还可以包括片段和修饰的核苷酸。因此,术语“多核苷酸”、“核酸序列”、“核苷酸序列”和“核酸片段”可互换使用以表示单链或双链的RNA和/或DNA的聚合物,任选地包含合成的、非天然的或改变的核苷酸碱基。核苷酸(通常发现处于其5'-单磷酸形式)可以通过单字母名称表示如下:“A”用于腺苷或脱氧腺苷(分别针对RNA或DNA),“C”用于胞嘧啶或脱氧胞嘧啶,“G”用于鸟苷或脱氧鸟苷,“U”用于尿苷,“T”用于脱氧胸苷,“R”用于嘌呤(A或G),“Y”用于嘧啶(C或T),“K”用于G或T,“H”用于A或C或T,“I”用于肌苷,并且“N”



用于任何核苷酸。

[0150] “可读框”缩写为ORF。

[0151] 术语“在功能上等价的亚片段”和“功能等价亚片段”在本文中可互换地使用。这些术语意指分离的核酸片段的一部分或子序列,其中不管该片段或亚片段是否编码活性酶,改变基因表达或产生某种表型的能力被保留。例如,片段或亚片段可用于设计基因以在转化的植物中产生所希望的表型。可以将基因设计为用于在抑制中使用,无论该基因是否编码活性酶,通过以相对于植物启动子序列的有义或反义取向连接其核酸片段或其亚片段。

[0152] 术语“保守结构域”或“基序”是指沿进化相关蛋白的比对序列在特定位置处保守的一组氨基酸。虽然同源蛋白质之间在其他位置处的氨基酸可以发生变化,但在特定位置处高度保守的氨基酸表明对蛋白质的结构、稳定性或活性来说是必需的氨基酸。因为它们通过蛋白质同系物家族的比对序列中的高度保守性而被鉴定,所以它们可以用作标识符或“特征”,以确定具有新确定的序列的蛋白质是否属于先前鉴定的蛋白质家族。

[0153] 多核苷酸和多肽序列、其变体、以及这些序列的结构关系,可用术语“同源性”、“同源的”、“基本上同一的”、“基本上相似的”、以及“基本上相应”来描述,这些术语在本文中可互换地使用。这些意指多肽或核酸片段,其中在一个或多个氨基酸或核苷酸碱基上的变化不影响分子的功能,例如介导基因表达或产生某种表型的能力。这些术语还意指相对于初始未修饰的片段,基本上不改变所得核酸片段的功能特性的核酸片段的一个或多个修饰。这些修饰包括在核酸片段中一个或多个核苷酸的缺失、取代、和/或插入。

[0154] 所涵盖的基本上相似的核酸序列可以通过这些核酸序列与本文所示例的序列杂交,或与本文所披露的并且与任何本文所披露的核酸序列在功能上等价的核苷酸序列的任何部分杂交(在中严格条件下,例如0.5X SSC,0.1%SDS,60°C)的能力来定义。可以调节严格条件以筛选中度相似的片段(例如来自远缘生物体的同源序列),到筛选高度相似的片段(例如从近缘生物体复制功能性酶的基因)。杂交后的洗涤决定了严格条件。

[0155] 术语“选择性杂交”包括参考在严格的杂交条件下将核酸序列杂交到特定的核酸靶序列上,相比其杂交到非靶核酸序列和基本上排除非靶核酸,该杂交达到可检测地更大程度(例如,至少为背景值的2倍)。选择性杂交序列通常彼此具有约至少80%序列同一性、或90%序列同一性、高达并且包括100%序列同一性(即,完全互补)。

[0156] 术语“严格条件”或“严格杂交条件”包括提及在体外杂交测定中探针将与其靶序列选择性杂交的条件。严格条件是序列依赖性的,并且在不同情况下将有所不同。通过控制杂交条件和/或洗涤条件的严格性,可以鉴定与探针100%互补的靶序列(同源探测)。可替代地,也能够调节严格条件以允许序列中的某些错配,以便检测到更低程度的相似性(异源探测)。通常,探针长度为小于约1000个核苷酸,任选地是长度小于500个核苷酸。

[0157] 通常,严格条件将是以下条件:在pH 7.0至8.3下盐浓度为小于约1.5M Na离子、通常约0.01至1.0M Na离子浓度(或其他一种或多种盐),并且对于短探针(例如,10至50个核苷酸)为至少约30°C,并且对于长探针(例如,超过50个核苷酸)为至少约60°C。添加去稳定剂例如甲酰胺也可以实现严格条件。示例性低严格条件包括在37°C下用30%至35%甲酰胺、1M NaCl、1%SDS(十二烷基硫酸钠)的缓冲溶液杂交,并且在50°C至55°C下在1X至2X SSC(20X SSC=3.0M NaCl/0.3M柠檬酸三钠)中洗涤。示例性中严格条件包括在37°C下在40%至45%甲酰胺、1M NaCl、1%SDS中杂交,并且在55°C至60°C下在0.5X至1X SSC中洗涤。

示例性高严格条件包括在37℃下在50%甲酰胺、1M NaCl、1%SDS中杂交,并且在60℃至65℃下在0.1X SSC中洗涤。

[0158] 在核酸的或多肽的序列的背景下,“序列同一性”或“同一性”是指在两个序列中的核酸碱基或氨基酸残基当在指定的比较窗口上比对最大对应度时是相同的。

[0159] “序列同一性的百分比”是指通过在比较窗口上比较两个最佳比对的序列所确定的值,其中与参考序列(其不包含添加或缺失)比较两个序列的最佳比对时,该多核苷酸或多肽序列在比较窗口中的部分可以包含添加或缺失(即空位)。通过以下方式计算该百分比:确定在两个序列中出现相同核酸碱基或氨基酸残基的位置的数目以产生匹配位置的数目,将匹配位置的数目除以比较窗口中的位置的总数目,然后将该结果乘以100以产生序列同一性的百分比。百分比序列同一性的有用实例包括但不限于50%、55%、60%、65%、70%、75%、80%、85%、90%或95%,或从50%至100%的任何整数百分比。可以使用在此描述的任何程序确定这些同一性。

[0160] 序列比对和百分比同一性或相似性计算可以使用设计用于检测同源序列的多种比较方法来确定,这些方法包括但不限于LASERGENE生物信息计算包(DNASTAR公司(DNASTAR Inc.),麦迪逊(Madison),威斯康星州)的MegAlign™程序。在此申请的背景内,应当理解的是,在使用序列分析软件来分析的情况下,分析的结果将基于参考的程序的“默认值”,除非另外指明。如本文所用,“默认值”将意指当第一次初始化时,最初加载该软件的任何组的值或参数。

[0161] “比对的Clustal V方法”对应于标记为Clustal V的比对方法(由以下描述:Higgins和Sharp,(1989)CABIOS 5:151-153;Higgins等人,(1992)Comput Appl Biosci[生物学中的计算机应用]8:189-191),并且发现于LASERGENE生物信息计算包(DNASTAR公司(DNASTAR Inc.),麦迪逊(Madison),威斯康星州)的MegAlign™程序中。对于多重比对,默认值对应于空位罚分(GAP PENALTY)=10和空位长度罚分(GAP LENGTH PENALTY)=10。使用Clustal方法进行逐对比对和蛋白质序列的百分比同一性计算的默认参数为KTUPLE=1、空位罚分=3、窗口(WINDOW)=5、以及存储的对角线(DIAGONALS SAVED)=5。对于核酸,这些参数是KTUPLE=2、空位罚分=5、窗口=4、以及存储的对角线=4。使用Clustal V程序比对序列后,可能通过查看同一程序中的“序列距离”表来获得“百分比同一性”。

[0162] “Clustal W比对方法”对应于标记为Clustal W的比对方法(由以下描述:Higgins和Sharp,(1989)CABIOS 5:151-153;Higgins等人,(1992)Comput Appl Biosci[生物学中的计算机应用]8:189-191),并且发现于LASERGENE生物信息计算包(DNASTAR公司(DNASTAR Inc.),麦迪逊(Madison),威斯康星州)的MegAlign™ v6.1程序中。用于多重比对的默认参数(空位罚分=10、空位长度罚分=0.2、延迟发散序列(Delay Divergen Seqs)(%)=30、DNA转换权重=0.5、蛋白质权矩阵=Gonnet系列、DNA权矩阵=IUB)。使用Clustal W程序比对序列后,可能通过查看同一程序中的“序列距离”表来获得“百分比同一性”。

[0163] 除非另行说明,否则在此提供的序列同一性/相似性值是指使用采用以下参数的GAP版本10((GCG,Accelrys公司,圣地亚哥,加利福尼亚州)获得的值:核苷酸序列的%同一性和%相似性采用50的空位产生罚分权重和3的空位长度延伸罚分权重以及nwsgapdna.cmp评分矩阵;氨基酸序列的%同一性和%相似性采用8的空位产生罚分权重和

2的空位长度延伸罚分权重以及BLOSUM62评分矩阵(Henikoff和Henikoff,(1989) Proc.Natl.Acad.Sci.USA[美国科学院院报]89:10915)。GAP使用Needleman和Wunsch(1970) J Mol Biol[分子生物学杂志]48:443-53的算法来找到使匹配数目最大化并且使空位数目最小化的两个完整序列的比对。GAP考虑所有可能的比对和空位位置,并且创建具有最大数目的配对碱基和最小空位的比对,使用以配对碱基为单位的空位创建罚分和空位延伸罚分。

[0164] “BLAST”是美国国家生物技术信息中心(National Center for Biotechnology Information,NCBI)提供的用于寻找生物序列之间的相似性的区域的搜索算法。该程序将核苷酸或者蛋白质序列与序列数据库比较,并计算匹配的统计显著性以鉴定出与查询序列具有足够的相似性的序列,这样使得相似性不会被预测为已经随机发生。BLAST报告鉴定的序列和它们与查询序列的局部比对。

[0165] 本领域技术人员很清楚地理解,许多水平的序列同一性在鉴定来自其他物种的多肽或修饰的天然或合成的多肽中是有用的,其中这样的多肽具有相同或相似的功能或活性。百分比同一性的有用实例包括但不限于50%、55%、60%、65%、70%、75%、80%、85%、90%或95%,或从50%至100%的任何整数百分比。实际上,在描述本披露中,从50%至100%的任何整数氨基酸同一性会是有用的,例如51%、52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%。

[0166] “基因”包括表达功能分子(例如,但不限于特定蛋白质)的核酸片段,包括编码序列之前的调节序列(5'非编码序列)和编码序列之后的调节序列(3'非编码序列)。“天然基因”是指自然界中发现的具有其自身调节序列的基因。

[0167] “突变基因”是通过人为干预已经改变的基因。这样的“突变基因”具有通过至少一个核苷酸添加、缺失或取代而与相应的非突变基因的序列不同的序列。在本披露的某些实施例中,该突变的基因包含由如本文披露的指导多核苷酸/Cas核酸内切酶系统引起的改变。突变的植物是包含突变基因的植物。

[0168] 如本文所用,“靶向突变”是通过使用涉及双链断裂诱导剂的方法,改变天然基因内的靶序列,而造成的天然基因中的突变,该双链断裂诱导剂能够在如在此披露的或本领域已知的靶序列的DNA中诱导双链断裂。

[0169] 指导RNA/Cas核酸内切酶诱导的靶向突变可以发生在位于由Cas核酸内切酶识别和切割的基因组靶位点内部或外部的核苷酸序列中。

[0170] 术语“基因组”当应用于植物细胞时不仅涵盖在细胞核内发现的染色体DNA,还涵盖在细胞的亚细胞组分(例如线粒体、或质体)内发现的细胞器DNA。

[0171] “密码子修饰的基因”或“密码子偏好的基因”或“密码子优化的基因”是其密码子使用的频率被设计为模拟宿主细胞的偏好的密码子使用的频率的基因。

[0172] “等位基因”是占据染色体上给定基因座的基因的若干种替代形式中的一种。当染色体上在给定基因座处存在的所有等位基因都相同时,该植物在该基因座处是纯合的。如果染色体上在给定基因座处存在的等位基因不同,则该植物在该基因座处是杂合的。

[0173] “编码序列”是指编码特定氨基酸序列的多核苷酸序列。“调节序列”是指位于编码

序列的上游(5'非编码序列)、内部或下游(3'非编码序列)的核苷酸序列,并且其影响相关的编码序列的转录、RNA加工或稳定性、或翻译。调节序列可以包括但不限于:启动子、翻译前导序列、5'非翻译序列、3'非翻译序列、内含子、聚腺苷酸化靶序列、RNA加工位点、效应子结合位点、和茎环结构。

[0174] “植物优化的核苷酸序列”是为了增加在植物中表达而优化的核苷酸序列。例如,可以使用一个或多个植物偏好的密码子来提高表达,通过修饰编码蛋白质(例如,如本文披露的双链断裂诱导剂(例如,核酸内切酶))的核苷酸序列,来合成植物优化的核苷酸序列。有关宿主偏好的密码子使用的讨论,参见例如Campbell和Gowri(1990) *Plant Physiol.* [植物生理学]92:1-11。

[0175] 本领域中可获得用于合成植物偏好基因的方法。参见,例如,美国专利号5,380,831和5,436,391,以及Murray等人(1989) *Nucleic Acids Res.* [核酸研究]17:477-498,通过引用结合在此。已知另外的序列修饰以增强在植物宿主中的基因表达。这些包括例如消除以下项:编码假的聚腺苷酸化信号、一个或多个外显子-内含子剪接位点信号的一个或多个序列、一个或多个转座子样重复序列、以及可能对基因表达有害的其他此类充分表征的序列。可以将序列的G-C含量调节至通过参考宿主植物细胞中表达的已知基因而计算出的给定植物宿主的平均水平。当可能时,修饰序列以避免出现一个或多个预测的发夹二级mRNA结构。因此,本披露的“植物-优化的核苷酸序列”包括一个或多个此类序列修饰。

[0176] 启动子是参与RNA聚合酶和其他蛋白质的识别和结合以开始转录的DNA区域。启动子序列由近端和更远端的上游元件组成,后一元件通常被称为增强子。“增强子”是可以刺激启动子活性的DNA序列,并且可以是该启动子的固有元件或被插入以增强启动子的水平或组织特异性的异源元件。启动子可以全部来源于天然基因,或者由来源于在自然界存在的不同启动子的不同元件构成,和/或包含合成的DNA区段。本领域技术人员应当理解,不同的启动子可能引导基因在不同组织或细胞类型中、或在不同发育阶段、或者响应于不同环境条件的表达。还认识到,由于在大多数情况下调节序列的确切界限仍未完全限定进一步认识到,由于在大多数情况下调节序列的确切边界尚未完全限定,一些变异的DNA片段可能具有相同的启动子活性。在多数情况下引起基因在大多数细胞型中表达的启动子通常称为“组成型启动子”。

[0177] 已示出某些启动子能够以相比其他启动子更高的速率引导RNA合成。这些被称为“强启动子”。已经显示某些其他启动子仅以较高的水平在特定类型的细胞或组织中指导RNA合成,并且如果所述启动子优选在某些组织中而且还以降低的水平在其他组织中指导RNA合成则通常将其称为“组织特异性启动子”或“组织偏好性启动子”。由于引入到植物中的嵌合基因(或基因)的表达模式是使用启动子来控制的,因此在分离能够在特定组织类型中或在特定的植物发育阶段以某种水平控制嵌合基因或(基因)表达的新颖启动子中是具有持续兴趣的。

[0178] 植物启动子可以包括能够在植物细胞中启动转录的启动子;关于植物启动子的综述,参见Potenza等人,(2004) *In Vitro Cell Dev Biol* [体外细胞与发育生物学]40:1-22。组成型启动子包括,例如Rsyn7启动子的核心启动子和其他在WO 99/43838和美国专利号6,072,050中披露的组成型启动子;核心CaMV 35S启动子(Ode11等人,(1985) *Nature* [自然]313:810-2);稻肌动蛋白(McElroy等人,(1990) *Plant Cell* [植物细胞]2:163-71);泛素

(Christensen等人, (1989) *Plant Mol Biol* [植物分子生物学] 12:619-32; Christensen等人, (1992) *Plant Mol. Biol* [植物分子生物学] 18:675-89); pEMU (Last等人, (1991) *Theor. Appl. Genet* [理论与应用遗传学] 81:581-8); MAS (Velten等人, (1984) *EMBO J* [欧洲分子生物学学会杂志] 3:2723-30); ALS启动子 (美国专利号5,659,026) 等。其他组成型启动子被描述于例如美国专利号5,608,149; 5,608,144; 5,604,121; 5,569,597; 5,466,785; 5,399,680; 5,268,463; 5,608,142和6,177,611中。在一些实例中,可以使用诱导型启动子。在被病原体感染后诱导的病原体诱导型启动子包括但不限于调节PR蛋白、SAR蛋白、 $\beta$ -1,3-葡聚糖酶、几丁质酶等的表达的启动子。

[0179] 可以使用化学调节型启动子通过应用外源化学调节剂来调节植物中的基因表达。在应用化学品诱导基因表达的情况下启动子可以是化学诱导型启动子,或者在应用化学品阻抑基因表达的情况下启动子可以是化学阻抑型启动子。化学品诱导型启动子包括但不限于:由苯磺酰胺除草剂安全剂激活的玉米In2-2启动子 (De Veylder等人, (1997) *Plant Cell Physiol* [植物细胞生理学] 38:568-77)、由用作出苗前除草剂的疏水性亲电子化合物激活的玉米GST启动子 (GST-11-27, W093/01294)、以及由水杨酸激活的烟草PR-1启动子 (Ono等人, (2004) *Biosci Biotechnol Biochem* [生物科学生物技术生物化学] 68:803-7)。其他化学调节的启动子包括类固醇反应启动子 (参见,例如,糖皮质激素诱导型启动子 (Scheda等人, (1991) *Proc. Natl. Acad. Sci. USA* [美国科学院院报] 88:10421-5; McNellis等人, (1998) *Plant J* [植物杂志] 14:247-257); 四环素诱导型启动子和四环素阻抑型启动子 (Gatz等人, (1991) *Mol Gen Genet* [分子和普通遗传学] 227:229-37; 美国专利号5,814,618和5,789,156)。

[0180] 组织偏好性启动子可以用于靶向特定植物组织内的增强的表达。组织偏好性启动子包括,例如, Kawamata等人, (1997) *Plant Cell Physiol* [植物细胞生理学] 38:792-803; Hansen等人, (1997) *Mol Gen Genet* [分子和普通遗传学] 254:337-43; Russell等人, (1997) *Transgenic Res* [转基因研究] 6:157-68; Rinehart等人, (1996) *Plant Physiol* [植物生理学] 112:1331-41; Van Camp等人, (1996) *Plant Physiol*. [植物生理学] 112:525-35; Canevascini等人, (1996) *Plant Physiol*. [植物生理学] 112:513-524; Lam, (1994) *Results Probl Cell Differ* [细胞分化中的结果与问题] 20:181-96; 以及Guevara-Garcia等人, (1993) *Plant J*. [植物杂志] 4:495-505。叶偏好性启动子包括,例如, Yamamoto等人, (1997) *Plant J* [植物杂志] 12:255-65; Kwon等人, (1994) *Plant Physiol* [植物生理学] 105:357-67; Yamamoto等人, (1994) *Plant Cell Physiol* [植物细胞生理学] 35:773-8; Gotor等人, (1993) *Plant J* [植物杂志] 3:509-18; Orozco等人, (1993) *Plant Mol Biol* [植物分子生物学] 23:1129-38; Matsuoka等人, (1993) *Proc. Natl. Acad. Sci. USA* [美国科学院院报] 90:9586-90; Simpson等人, (1958) *EMBO J* [欧洲分子生物学学会杂志] 4:2723-9; Timko等人, (1988) *Nature* [自然] 318:57-8。根偏好性启动子包括,例如, Hire等人, (1992) *Plant Mol Biol* [植物分子生物学] 20:207-18 (大豆根特异性谷氨酰胺合酶基因); Miao等人, (1991) *Plant Cell* [植物细胞] 3:11-22 (胞质谷氨酰胺合酶 (GS)); Keller和Baumgartner, (1991) *Plant Cell* [植物细胞] 3:1051-61 (法国菜豆的GRP 1.8基因中的根特异性控制元件); Sanger等人, (1990) *Plant Mol Biol* [植物分子生物学] 14:433-43 (根癌农杆菌 (*A. tumefaciens*) 的甘露氨酸合酶 (MAS) 的根特异性启动子); Boguszc等人, (1990) *Plant*

Cell [植物细胞] 2:633-41 (从榆科糙叶山黄麻 (*Parasponia andersonii*) 和山黄麻 (*Trema tomentosa*) 分离的根特异性启动子); Leach和Aoyagi, (1991) *Plant Sci* [植物科学] 79:69-76 (发根农杆菌 (*A. rhizogenes*) *rolC*和 $rolD$ 根诱导型基因); Teeri等人, (1989) *EMBO J* [欧洲分子生物学学会杂志] 8:343-50 (农杆菌伤口诱导的TR1' 和TR2' 基因);  $\nu$ FENOD-GRP3基因启动子 (Kuster等人, (1995) *Plant Mol Biol* [植物分子生物学] 29:759-72); 以及 $rolB$ 启动子 (Capana等人, (1994) *Plant Mol Biol* [植物分子生物学] 25:681-91); 菜豆球蛋白基因 (Murai等人, (1983) *Science* [科学] 23:476-82; Sengopta-Gopalen等人, (1988) *Proc. Natl. Acad. Sci. USA* [美国科学院院报] 82:3320-4)。还参见美国专利号5,837,876; 5,750,386; 5,633,363; 5,459,252; 5,401,836; 5,110,732和5,023,179。

[0181] 种子偏好性启动子包括在种子发育期间有活性的种子特异性启动子以及在种子发芽期间有活性的种子发芽性启动子两者。参见Thompson等人, (1989) *BioEssays* [生物学分析] 10:108。种子偏好性启动子包括但不限于Cim1 (细胞分裂素诱导的信号); cZ19B1 (玉蜀黍19kDa玉米蛋白); 和milps (肌醇-1-磷酸盐合酶); (W000/11177; 以及美国专利6,225,529)。对于双子叶植物, 种子偏好性启动子包括但不限于: 菜豆 $\beta$ -菜豆素、油菜籽蛋白、 $\beta$ -伴大豆球蛋白、大豆凝集素、十字花科蛋白等。对于单子叶植物, 种子偏好性启动子包括但不限于玉蜀黍15kDa玉米蛋白、22kDa玉米蛋白、27kDa  $\gamma$ 玉米蛋白、蜡质、收缩素1、收缩素2、球蛋白1、油质蛋白和nucl。还参见WO 00/12733, 其中披露了来自END1和END2基因的种子偏好性启动子。

[0182] 术语“诱导型启动子”意指对内源或外源刺激的存在, 例如通过化学化合物 (化学诱导剂) 响应, 或对环境、激素、化学品、和/或发育信号响应, 选择性表达编码序列或功能RNA的启动子。诱导型或调节型启动子包括例如通过光、热、胁迫、水淹或干旱、盐胁迫、渗透胁迫、植物激素、伤口或化学品 (例如乙醇、脱落酸 (ABA)、茉莉酮酸酯、水杨酸或安全剂) 诱导或调节的启动子。

[0183] 胁迫诱导型启动子的实例是RD29A启动子 (Kasuga等人 (1999) *Nature Biotechnol* [自然生物技术]. 17:287-91)。本领域技术人员熟悉模拟干旱条件并评价植物耐旱性的规程, 所述植物已经遭受了模拟的或天然存在的干旱条件。技术人员可以通过给予植物比正常需要更少的水或在一个时段内不提供水来模拟干旱条件, 并且技术人员可通过寻找在生理和/或物理条件上的差异来评价耐旱性, 包括 (但不限于) 活力、生长、大小、或根长、或具体地讲叶片颜色或叶片面积大小。用于评价耐旱性的其他技术包括测量叶绿素荧光、光合作用速率和换气速率。此外, 本领域技术人员熟悉模拟胁迫条件 (例如渗透胁迫、盐胁迫、和温度胁迫) 并评价植物的胁迫耐受性的规程, 所述植物已经遭受了模拟的或天然存在的胁迫条件。

[0184] 在植物细胞中有用的诱导型启动子的另一个实例已经在于2013年11月21日公开的美国专利申请US 2013-0312137A1中进行了描述, 将该申请通过引用结合在此。美国专利申请US 2013-0312137A1描述了来自CBSU-Anther\_Subtraction文库 (CAS1) 基因的ZmCAS1启动子及其功能片段, 该基因编码来自玉蜀黍的甘露醇脱氢酶。该ZmCAS1启动子 (还称为“CAS1启动子”、“甘露醇脱氢酶启动子”, “mdh启动子”) 可以被如上所述的化学或胁迫处理诱导。该化学品可以是安全剂, 例如, 但不限于, N-(氨基羰基)-2-氯苯磺酰胺 (2-CBSU)。该胁迫处理可以是热处理, 例如, 但不限于, 热休克处理 (还参见, 2015年2月25日提交的美国

临时专利申请62/120421,并且通过引用结合在此)。

[0185] 不断发现在植物细胞中有用的不同类型的新启动子;许多实例可以在Okamuro和Goldberg, (1989) *The Biochemistry of Plants* [植物生物化学], 115卷, Stumpf和Conn, 编辑(纽约, 纽约州: 学术出版社) 1-82页的汇编中发现。

[0186] “翻译前导序列”意指位于基因的启动子序列和编码序列之间的多核苷酸序列。翻译前导序列存在于翻译起始序列的mRNA上游。翻译前导序列可以影响初级转录物对mRNA的加工、mRNA稳定性、或翻译效率。已经描述了翻译前导序列的实例(例如, Turner和Foster, (1995) *Mol Biotech* 1 [分子生物技术] 3: 225-236)。

[0187] “3' 非编码序列”、“转录终止子”、或“终止序列”意指位于编码序列的下游的DNA序列, 并且包括聚腺苷酸化识别序列和编码能够影响mRNA加工或基因表达的调节信号的其他序列。聚腺苷酸化信号通常特征在于影响聚腺苷酸片添加到mRNA前体的3' 末端。由Ingelbrecht等人, (1989) *Plant Cell* [植物细胞] 1: 671-680示例了不同的3' 非编码序列的用途。

[0188] “RNA转录物”是指由DNA序列的RNA聚合酶催化的转录产生的产物。当RNA转录物是DNA序列的完全互补拷贝时, 该RNA转录物被称为初级转录物或前mRNA。当RNA转录物是源自初级转录物前mRNA的转录后加工的RNA序列时, RNA转录物被称为成熟RNA或mRNA。“信使RNA”或“mRNA”是指不含内含子并且可以被细胞翻译成蛋白质的RNA。“crDNA”是指与mRNA模板互补并且使用逆转录酶从mRNA模板合成的DNA。cDNA可以是单链的或者可以使用DNA聚合酶I的Klenow片段转化成双链形式。“正义”RNA是指包含mRNA并且可以在细胞内或体外翻译成蛋白质的RNA转录物。“反义RNA”是指与靶初级转录物或mRNA的全部或部分互补、并且阻断靶基因的表达的RNA转录物(参见, 例如美国专利号5, 107, 065)。反义RNA的互补性可以是与特定基因转录物的任何部分, 即在5' 非编码序列、3' 非编码序列、内含子或编码序列处。“功能性RNA”是指反义RNA、核酶RNA、或不能被翻译但对细胞加工有影响的其他RNA。术语“互补序列”和“反向互补序列”在本文中关于mRNA转录物可互换使用, 并且意在限定信使的反义RNA。

[0189] 术语有效地连接是指单个核酸片段上的核酸序列的关联, 这样使得其中一个核酸序列的功能被另一个核酸序列调节。例如, 当启动子能够调节编码序列的表达(即, 该编码序列在启动子的转录控制下)时, 启动子与该编码序列有效地连接。编码序列可以在正义或反义取向上有效地连接到调节序列。在另一个实例中, 互补的RNA区域可以直接或间接有效地连接至靶mRNA的5'、或靶mRNA的3'、或靶mRNA内、或第一个互补区是5'且其互补序列是靶mRNA的3'。

[0190] 本文使用的标准重组DNA和分子克隆技术是在本领域熟知的, 并且更全面地描述于Sambrook等人, *Molecular Cloning: A Laboratory Manual* [分子克隆: 实验室手册], Cold Spring Harbor Laboratory: Cold Spring Harbor, NY [冷泉港实验室: 冷泉港, 纽约州] (1989) 中。转化方法是本领域技术人员熟知的并且在下文中进行了描述。

[0191] “PCR”或“聚合酶链反应”是用于合成特定DNA片段的技术, 并且“PCR”或“聚合酶链式反应”是用于特定DNA区段的合成的技术, 由一系列重复变性、退火和延伸循环组成。通常, 将双链DNA进行热变性, 并将两条与靶区段的3' 边界互补的引物与该DNA在低温下退火, 并且然后在中等温度下延伸。将这三个连续步骤的一组称为一个“循环”。

[0192] 术语“重组”是指例如通过化学合成或者通过基因工程技术操纵分离的核酸区段来将两个原本分离的序列区段进行人工组合。

[0193] 术语“质粒”、“载体”和“盒”意指染色体外元件,其通常携带非细胞中心代谢的一部分的基因,并且通常处于双链DNA的形式。这样的元件可以是来源于任何来源的、单链或双链DNA或RNA的、处于直链或环状形式的自主复制序列、基因组整合序列、噬菌体、或核苷酸序列,其中许多核苷酸序列已经被连接或重组成能够将目的多核苷酸引入细胞中的独特构造。“转化盒”意指包含基因并具有促进特定宿主细胞转化的基因之外的元件的特定载体。“表达盒”是指包含基因并具有允许在宿主中表达该基因的基因之外的元件的特定载体。

[0194] 术语“重组DNA分子”、“重组构建体”、“表达构建体”、“构建体”、“构建体”、和“重组DNA构建体”在此可互换地使用。重组构建体包含核酸片段,例如在自然界中未全部一起发现的调节序列和编码序列的人工组合。例如,构建体可以包含源自于不同来源的调节序列和编码序列,或者包括源自于相同来源但以不同于天然存在的方式排列的调节序列和编码序列。这样一个构建体可以单独使用或可以与载体结合使用。如果使用载体,则载体的选择取决于如本领域技术人员熟知的将用于转化宿主细胞的方法。例如,可以使用质粒载体。技术人员充分了解必须存在于载体上以便成功转化,选择和繁殖宿主细胞的遗传元件。技术人员还将认识到,不同的独立转化事件可以导致不同水平和模式的表达(Jones等人,(1985)EMBOJ[欧洲分子生物学学会杂志]4:2411-2418;De Almeida等人,(1989)Mol Gen Genetics[分子和普通遗传学]218:78-86),并且因此通常筛选多个事件,从而获得显示希望的表达水平和模式的品系。此类筛选可以是完成的标准分子生物学测定、生物化学测定以及其他测定,这些测定包括DNA的印迹分析、mRNA表达的Northern分析、PCR、实时定量PCR(qPCR)、逆转录PCR(RT-PCR)、蛋白表达的免疫印迹分析、酶测定或活性测定、和/或表型分析。

[0195] 如本文所用,术语“表达”是指处于前体抑或成熟形式的功能性终产物(例如,mRNA、指导RNA或蛋白质)的产生。

[0196] 术语“引入”包括提及在细胞中引入、提供、接触化合物(例如但不限于核酸(例如表达构建体)或肽、多肽或蛋白质)。引入包括多核苷酸(例如RNA、DNA、RNA-DNA杂合体、单链或双链寡核苷酸、线性或环状多核苷酸)的直接递送并且/或包括蛋白质(多肽)的直接递送。引入包括提及将核酸或多肽掺入真核细胞或原核细胞中,其中核酸可以被掺入细胞基因组中,并且包括提及将核酸或蛋白质被瞬时引入细胞中。引入包括提及稳定或瞬时转化方法、转染、转导、显微注射、电穿孔、病毒方法、农杆菌介导的转化、弹道粒子加速、晶须介导的转化以及有性杂交。因此,在将核酸片段(例如,重组DNA构建体/表达构建体、指导RNA、指导DNA、模板DNA、供体DNA)插入细胞中的上下文中,“引入”包括“转染”、“转化”或“转导”,并且包括提及将核酸片段掺入真核或原核细胞中,其中可以将该核酸片段掺入细胞的基因组(例如,染色体、质粒、质体或线粒体DNA)中,转化成自主复制子或进行瞬时表达(例如,经转染的mRNA)。

[0197] 用于在生物体中引入、接触和/或提供组合物的各种方法是已知的,包括稳定转化方法、瞬时转化方法、病毒介导的方法、有性杂交和有性繁殖。稳定转化表明将引入的多聚核苷酸整合到生物体的基因组中,并且能够由其后代进行遗传。瞬时转化表明所引入的组



合物仅在生物体中暂时表达或存在。

[0198] 用于在细胞或生物体中接触、提供、引入多核苷酸和多肽的方案是已知的并且包括显微注射(Crossway等人,(1986)Biotechniques[生物技术]4:320-34和美国专利号6,300,543);分生组织转化(美国专利号5,736,369);电穿孔(Riggs等人,(1986)Proc.Natl.Acad.Sci.USA[美国科学院院报]83:5602-6);农杆菌介导的转化(美国专利号5,563,055和5,981,840);晶须介导的转化(Ainley等人2013,Plant Biotechnology Journal[植物生物技术杂志]11:1126-1134;Shaheen A.和M.Arshad 2011Properties and Applications of Silicon Carbide[碳化硅的特性和应用](2011),345-358,编辑:Gerhardt,Rosario.,出版商:印天科技公司(InTech),里耶卡(Rijeka),克罗地亚(Croatia),代码:69PQBP;ISBN:978-953-307-201-2);直接基因转移(Paszkowski等人,(1984)EMBO J[欧洲分子生物学学会杂志]3:2717-22);以及弹道粒子加速(美国专利号4,945,050;5,879,918;5,886,244;5,932,782;Tomes等人,(1995)“Direct DNA Transfer into Intact Plant Cells via Microprojectile Bombardment”[经由微粒轰击将DNA直接转移到完整植物细胞中]在Plant Cell, Tissue, and Organ Culture: Fundamental Methods[植物细胞、组织和器官培养:基本方法],编辑Gamborg和Phillips(Springer-Verlag,Berlin[柏林施普林格出版社];McCabe等人(1988)Biotechnology[生物技术]6:923-6;Weissinger等人,(1988)Ann Rev Genet[遗传学年鉴]22:421-77;Sanford等人,(1987)Particulate Science and Technology[微粒科学与技术]5:27-37(洋葱);Christou等人,(1988)Plant Physiol[植物生理学]87:671-4(大豆);Finer和McMullen,(1991)In Vitro Cell Dev Biol[体外细胞与发育生物学]27P:175-82(大豆);Singh等人,(1998)Theor Appl Genet[理论与应用遗传学]96:319-24(大豆);Datta等人,(1990)Biotechnology[生物技术]8:736-40(稻);Klein等人,(1988)Proc.Natl.Acad.Sci.USA[美国科学院院报]85:4305-9(玉蜀黍);Klein等人,(1988)Biotechnology[生物技术]6:559-63(玉蜀黍);美国专利号5,240,855;5,322,783和5,324,646;Klein等人,(1988)Plant Physiol[植物生理学]91:440-4(玉蜀黍);Fromm等人,(1990)Biotechnology[生物技术]8:833-9(玉蜀黍);Hooykaas-Van Slogteren等人,(1984)Nature[自然]311:763-4;美国专利号5,736,369(谷类);Bytebier等人,(1987)Proc.Natl.Acad.Sci.USA[美国科学院院报]84:5345-9(百合科(Liliaceae));De Wet等人,(1985)在The Experimental Manipulation of Ovule Tissues[胚珠组织的实验操作]中,编辑Chapman等人,(Longman,New York[纽约朗文出版社]),第197-209页(花粉);Kaeppler等人,(1990)Plant Cell Rep[植物细胞报告]9:415-8)以及Kaeppler等人,(1992)Theor Appl Genet[理论与应用遗传学]84:560-6(晶须介导的转化);D’Halluin等人,(1992)Plant Cell[植物细胞]4:1495-505(电穿孔);Li等人,(1993)Plant Cell Rep[植物细胞报告]12:250-5;Christou和Ford(1995)Annals Botany[植物学年鉴]75:407-13(稻)以及Osjoda等人,(1996)Nat Biotechnol自然生物技术]14:745-50(经由根癌农杆菌转化的玉蜀黍)。

[0199] 可替代地,可以通过使细胞或生物体与病毒或病毒核酸接触来将多核苷酸引入细胞或生物体中。通常,此类方法涉及将多核苷酸掺入病毒DNA或RNA分子内。在一些实例中,可以最初将目的多肽作为病毒多聚蛋白的一部分合成,然后将合成的多肽在体内或在体外通过蛋白水解加工从而产生所希望的重组蛋白。用于将多核苷酸引入植物,并且表达在其

中编码的蛋白质(涉及病毒DNA或RNA分子)的方法是已知的,参见例如,美国专利号5,889,191、5,889,190、5,866,785、5,589,367、以及5,316,931。瞬时转化方法包括但不限于将多肽(例如双链断裂诱导剂)直接引入生物体,将多核苷酸(例如DNA和/或RNA多核苷酸)引入,以及将RNA转录物(例如编码双链断裂诱导剂的mRNA)引入生物体中。此类方法包括例如显微注射或粒子轰击。参见,例如Crossway等人,(1986)Mol Gen.Genet[分子和普通遗传学]202:179-85;Nomura等人,(1986)Plant Sci.[植物科学]44:53-8;Hepler等人,(1994)Proc.Natl.Acad. Sci.USA[美国科学院院报],91:2176-80;以及,Hush等人,(1994)J Cell Sci[细胞科学杂志]107:775-84。

[0200] 可以通过任何方法将核酸和蛋白质提供给细胞,所述方法包括使用分子来促进指导的Cas系统(蛋白质和/或核酸)的任何或所有组分(例如细胞穿透肽和纳米载体)的摄取的方法。还参见US20110035836Nanocarrier based plant transfection and transduction[基于植物转染和转导的纳米载体],和EP 2821486 A1 Method of introducing nucleic acid into plant cells[将核酸引入植物细胞中的方法],通过引用结合在此。

[0201] 将指导RNA/Cas核酸内切酶复合物引入细胞中包括将所述复合物的各组分单独地或组合地引入细胞中,并且直接地(作为指导的RNA和Cas核酸内切酶的蛋白质直接递送)或经由表达这些组分(指导RNA、Cas核酸内切酶)的重组构建体引入。将指导RNA/Cas核酸内切酶复合物引入细胞中包括将该指导RNA/Cas核酸内切酶复合物作为核糖核苷酸-蛋白质引入细胞中。可以将该核糖核苷酸-蛋白质在引入如本文所述的细胞中之前进行组装。

[0202] 植物细胞与人类和动物细胞的不同之处在于,植物细胞含有植物细胞壁,其可以作为RGEN核糖核蛋白的直接递送和/或RGEN组分的直接递送的屏障。

[0203] 如本文所述,可以通过颗粒介导的递送(粒子轰击)实现将RGEN核糖核蛋白直接递送到植物细胞中。基于本文所述的实验,技术人员现在可以预想任何其他直接递送方法(例如但不限于聚乙二醇(PEG)介导的对原生质体的转染、电穿孔、细胞穿透肽或介孔二氧化硅纳米颗粒(MSN)介导的直接蛋白质递送)都可以成功用于将RGEN核糖核蛋白递送到植物细胞中。

[0204] 如本文所述,RGEN核糖核蛋白的直接递送允许在细胞基因组中的靶位点进行基因组编辑,其后可以迅速降解复合物,并且仅允许细胞中短暂存在该复合物。RGEN复合物的这种短暂存在可能导致脱靶效应降低。相比之下,经由质粒DNA序列递送RGEN组分(指导RNA、Cas9核酸内切酶)可以导致RGEN从这些质粒的的恒定表达,该恒定表达可以加强脱靶效应(Cradick,T.J.等人(2013)Nucleic Acids Res[核酸研究]41:9584-9592;Fu,Y等人(2014)Nat.Biotechnol.[自然生物技术]31:822-826)。

[0205] 直接递送可以通过将RNA指导的核酸内切酶的任何一种组分(指导RNA、Cas蛋白、编码gRNA或Cas核酸内切酶的mRNA)或RGEN复合物本身与包含微粒(例如但不限于金颗粒、钨颗粒和碳化硅晶须颗粒)的颗粒递送基质进行组合来实现。本文所述的用于将微粒与质粒DNA和目的DNA结合的组合方法的实例还可以用于将指导RNA分子、mRNA分子、Cas蛋白和RGEN复合物包被到微粒上。

[0206] 可以通过本领域已知的任何直接方法(例如实例8中描述的粒子轰击方法)将这些包被的微粒引入细胞中。微粒和RGEN组分或RGEN复合物可以在任何物质中组合(混合)以允

许将RGEN组分包被到微粒上。例如,使用任何合适的缓冲液,例如但不限于水溶性阳离子脂质(例如但不限于TransIT-2020转染试剂(目录号MIR 5404,Mirus公司,美国)),可以将RGEN组分沉淀到直径范围为至少0.1 $\mu$ m、0.2 $\mu$ m、0.3 $\mu$ m、0.4 $\mu$ m、0.5 $\mu$ m、0.6 $\mu$ m、0.7 $\mu$ m、0.8 $\mu$ m、0.9 $\mu$ m或1.0 $\mu$ m直径的金球粒上。使用至少0.1 $\mu$ g、0.2 $\mu$ g、0.3 $\mu$ g、0.4 $\mu$ g、0.5 $\mu$ g、0.6 $\mu$ g、0.7 $\mu$ g、0.8 $\mu$ g、0.9 $\mu$ g、1.0 $\mu$ g、2.0 $\mu$ g、3.0 $\mu$ g、4.0 $\mu$ g、5.0 $\mu$ g、6.0 $\mu$ g、7.0 $\mu$ g、8.0 $\mu$ g、9.0 $\mu$ g或10 $\mu$ g的RNA(指导的RNA或mRNA)或Cas核酸内切酶蛋白,可以在冰上(或在适合于使微粒结合的任何温度下)制备RGEN组分溶液。为了得到RGEN复合物的预混合RGEN组分,可以添加至少1 $\mu$ l至20 $\mu$ l的制备的微粒并小心混合。

[0207] 将指导RNA/Cas核酸内切酶复合物引入细胞中的方法包括以下方法:所述方法包括将至少一种指导RNA分子和至少一种Cas核酸内切酶蛋白进行组合以形成核糖核苷酸-蛋白质,并将所述核糖核苷酸-蛋白质与颗粒递送基质进行组合以允许所述核糖核苷酸-蛋白质和基质结合并形成核糖核苷酸-蛋白质-基质复合物;并且,将所述核糖核苷酸-蛋白质-基质复合物引入所述细胞中。该颗粒递送基质可以包含与阳离子脂质结合的微粒。

[0208] 术语“阳离子脂质”包括提及水溶性阳离子脂质(例如但不限于TransIT-2020)或阳离子脂质溶液(例如但不限于包含N,N,N',N'-四甲基-N,N'-双(2-羟乙基)-2,3-二(油酰基氧基)-1,4-丁烷二铵碘化物和L-二油酰基磷脂酰乙醇胺(DOPE)的阳离子脂质溶液)(还参见2007年8月2日公开的US2007/0178593,通过引用结合在此)。

[0209] 5. 要求保护的方法。

[0210] 该颗粒递送基质可以包含选自下组的微粒,该组由以下各项组成:金颗粒、钨颗粒和碳化硅晶须颗粒。

[0211] 该颗粒递送基质可以进一步包含选自下组的化合物,该组由以下各项组成:Tfx-10<sup>TM</sup>、Tfx-20<sup>TM</sup>、Tfx-50<sup>TM</sup>、Lipofectin<sup>TM</sup>、Lipofectamine<sup>TM</sup>、Cellfectin<sup>TM</sup>、Effectene<sup>TM</sup>、Cytfectin GSV<sup>TM</sup>、Perfect Lipids<sup>TM</sup>、DOTAp<sup>TM</sup>、DMRIE-C<sup>TM</sup>、FuGENE-6<sup>TM</sup>、Superfect<sup>TM</sup>、Polyfect<sup>TM</sup>、聚乙烯亚胺、壳聚糖、鱼精蛋白C1、组蛋白H1、组蛋白CENH3、聚-L赖氨酸、和DMSA(2007年8月2日公开的US2007/0178593,通过引用结合在此)。

[0212] 还可以在包被在微粒上之前,通过将至少0.1 $\mu$ g、0.2 $\mu$ g、0.3 $\mu$ g、0.4 $\mu$ g、0.5 $\mu$ g、0.6 $\mu$ g、0.7 $\mu$ g、0.8 $\mu$ g、0.9 $\mu$ g、1.0 $\mu$ g、2.0 $\mu$ g、3.0 $\mu$ g、4.0 $\mu$ g、5.0 $\mu$ g、6.0 $\mu$ g、7.0 $\mu$ g、8.0 $\mu$ g、9.0 $\mu$ g或10 $\mu$ g的指导RNA与至少0.1 $\mu$ g、0.2 $\mu$ g、0.3 $\mu$ g、0.4 $\mu$ g、0.5 $\mu$ g、0.6 $\mu$ g、0.7 $\mu$ g、0.8 $\mu$ g、0.9 $\mu$ g、1.0 $\mu$ g、2.0 $\mu$ g、3.0 $\mu$ g、4.0 $\mu$ g、5.0 $\mu$ g、6.0 $\mu$ g、7.0 $\mu$ g、8.0 $\mu$ g、9.0 $\mu$ g或10 $\mu$ g的Cas核酸内切酶在适合于允许复合物形成的溶液(例如但不限于Cas9缓冲液(NEB))中,在允许复合物形成的任何温度(例如温度范围为1 $^{\circ}$ C、2 $^{\circ}$ C、3 $^{\circ}$ C、4 $^{\circ}$ C、5 $^{\circ}$ C、6 $^{\circ}$ C、7 $^{\circ}$ C、8 $^{\circ}$ C、9 $^{\circ}$ C、10 $^{\circ}$ C、11 $^{\circ}$ C、12 $^{\circ}$ C、13 $^{\circ}$ C、14 $^{\circ}$ C、15 $^{\circ}$ C、16 $^{\circ}$ C、17 $^{\circ}$ C、18 $^{\circ}$ C、19 $^{\circ}$ C、20 $^{\circ}$ C、21 $^{\circ}$ C、22 $^{\circ}$ C、23.0 $^{\circ}$ C、24 $^{\circ}$ C、25 $^{\circ}$ C、26 $^{\circ}$ C、27 $^{\circ}$ C、28 $^{\circ}$ C、29 $^{\circ}$ C、30 $^{\circ}$ C、31 $^{\circ}$ C、32 $^{\circ}$ C、33.0 $^{\circ}$ C、34 $^{\circ}$ C、35 $^{\circ}$ C、36 $^{\circ}$ C、37 $^{\circ}$ C、38 $^{\circ}$ C、39 $^{\circ}$ C和40 $^{\circ}$ C)下将RGEN组分进行组合。

[0213] 将指导RNA/Cas核酸内切酶复合物引入细胞中的方法包括以下方法:将至少一种指导RNA分子和至少一种Cas核酸内切酶蛋白引入细胞中,并且在合适的条件下使所述细胞生长以允许所述指导RNA和所述Cas核酸内切酶蛋白在所述细胞内部形成复合物。

[0214] 将指导RNA/Cas核酸内切酶复合物引入细胞中的方法包括以下方法:将至少一种指导RNA分子和至少一种编码Cas核酸内切酶蛋白的mRNA引入细胞中,并且在合适的条件下

使所述细胞生长以允许所述mRNA翻译所述Cas核酸内切酶蛋白并与所述指导RNA形成复合物。

[0215] 将指导RNA/Cas核酸内切酶复合物引入细胞中的方法包括以下方法：将至少一种指导RNA分子和至少一种Cas核酸内切酶蛋白进行组合以形成核糖核苷酸-蛋白质，并将所述核糖核苷酸-蛋白质与颗粒递送基质进行组合以允许所述核糖核苷酸-蛋白质和基质结合并形成核糖核苷酸-蛋白质-基质复合物；并且，将所述核糖核苷酸-蛋白质-基质复合物与至少一种多核苷酸模板一起引入所述细胞中，其中所述多核苷酸修饰模板包含所述细胞的基因组中的核苷酸序列的至少一个核苷酸修饰，其中所述多核苷酸修饰模板的所述至少一个核苷酸修饰选自下组，该组由以下各项组成：(i) 至少一个核苷酸的替代、(ii) 至少一个核苷酸的缺失、(iii) 至少一个核苷酸的插入、以及(iv) (i)-(iii)的任何组合。

[0216] 将指导RNA/Cas核酸内切酶复合物引入细胞中的方法包括以下方法：将至少一种指导RNA分子和至少一种Cas核酸内切酶蛋白进行组合以形成核糖核苷酸-蛋白质，并将所述核糖核苷酸-蛋白质与颗粒递送基质进行组合以允许所述核糖核苷酸-蛋白质和基质结合并形成核糖核苷酸-蛋白质-基质复合物；并且，将所述核糖核苷酸-蛋白质-基质复合物与供体DNA一起引入所述细胞中，其中所述供体DNA包含至少一个目的多核苷酸。

[0217] 用于使细胞生长的合适条件是本领域熟知的，并且技术人员可以使用基于细胞类型的任何生长条件(例如适合于植物细胞的条件)。如实例8所述，植物胚胎或细胞可以在本领域已知的任何植物维持培养基(例如但不限于560P,实例8)中在范围从26°C至37°C的温度下孵育12至48小时，并且然后放置在26°C。在5至7天之后，将胚胎/细胞转移至本领域已知的任何选择培养基(例如但不限于560R,实例8)，其后进行传代培养。

[0218] 可以在包被在微粒上(与微粒结合)之前，通过将至少0.1μg、0.2μg、0.3μg、0.4μg、0.5μg、0.6μg、0.7μg、0.8μg、0.9μg、1.0μg、2.0μg、3.0μg、4.0μg、5.0μg、6.0μg、7.0μg、8.0μg、9.0μg或10μg的指导RNA与至少0.1μg、0.2μg、0.3μg、0.4μg、0.5μg、0.6μg、0.7μg、0.8μg、0.9μg、1.0μg、2.0μg、3.0μg、4.0μg、5.0μg、6.0μg、7.0μg、8.0μg、9.0μg或10μg的Cas核酸内切酶在适合于允许复合物形成的溶液(例如但不限于Cas9缓冲液(NEB))中，在允许复合物形成的任何温度(例如温度范围为1°C、2°C、3°C、4°C、5°C、6°C、7°C、8°C、9°C、10°C、11°C、12°C、13°C、14°C、15°C、16°C、17°C、18°C、19°C、20°C、21°C、22°C、23°C、24°C、25°C、26°C、27°C、28°C、29°C、30°C、31°C、32°C、33°C、34°C、35°C、36°C、37°C、38°C、39°C和40°C)下将RGEN组分(包括指导RNA、Cas核酸内切酶蛋白)组合以形成核糖核苷酸-蛋白质复合物(RNP)。

[0219] “成熟”蛋白质是指翻译后加工的多肽(即，从其中已经除去存在于初级翻译产物中的任何前肽(pre-peptide)或原肽(propeptide)的一种多肽)。“前体”蛋白质是指mRNA的翻译的初级产物(即，仍存在前肽或原肽)。前肽或原肽可以是但不限于细胞内定位信号。

[0220] “稳定转化”是指将核酸片段转移至宿主生物体的基因组中，包括细胞核的和细胞器的基因组，导致遗传稳定的遗传。相比之下，“瞬时转化”是指将核酸片段转移至宿主生物体的细胞核中或其他包含DNA的细胞器中，导致基因表达而无整合或稳定的遗传。包含转化的核酸片段的宿主生物体被称为“转基因”生物体。

[0221] 基因改良种质的商业开发也已经进展至将多个性状引入作物植物中的阶段，其通常被称为基因堆叠法(gene stacking approach)。在该方法中，可以将赋予目的不同特征的多种基因引入植物中。基因堆叠可通过许多手段实现，包括但不限于共转化、再转化以及

具有目的不同基因的品系的杂交。

[0222] 细胞包括但不限于人类、非人类、动物、细菌、真菌、昆虫、酵母和植物细胞,以及通过本文所述的方法产生的植物和种子。植物细胞包括选自下组的细胞,该组由以下各项组成:玉蜀黍、稻、高粱、黑麦、大麦、小麦、粟、燕麦、甘蔗、草坪草、或柳枝稷、大豆、低芥酸菜籽、苜蓿、向日葵、棉花、烟草、花生、马铃薯、番茄、烟草、拟南芥、和红花细胞。

[0223] 术语“植物”包括提及整株植物、植物器官、植物组织、种子、和植物细胞及其后代。植物细胞包括但不限于来自以下物质的细胞:种子、悬浮培养物、胚胎、分生区、愈伤组织、叶、根、芽、配子体、孢子体、花粉和小孢子。植物部分包括分化和未分化的组织,包括但不限于根、茎、芽、叶、花粉、种子、肿瘤组织和各种形式的细胞和培养物(例如单细胞、原生质体、胚胎和愈伤组织)。植物组织可以是在植物中或在植物器官、组织或细胞培养物中的。术语“植物器官”意指植物组织或构成植物的形态上和功能上不同部分的一组组织。术语“基因组”意指存在于生物体或病毒或细胞器的每个细胞中的遗传物质的全部互补序列(基因和非编码序列);和/或从一个亲本遗传为(单倍体)单位的完整染色体组。“后代”包括植物的任何后续世代。

[0224] 转基因植物包括例如在其基因组中包含通过转化步骤引入的异源多核苷酸的植物。异源多核苷酸可以稳定地整合到基因组内,这样使得多核苷酸被传递给连续世代。异源多核苷酸可以单独地或作为重组DNA构建体的部分整合进基因组中。转基因植物还可以在其基因组内包含多于一个异源多核苷酸。各异源多核苷酸均可对所述转基因植物产生不同的性状。异源多核苷酸可以包括源自外来物种的序列,或者如果源自相同物种,可以从其天然形式上进行实质修饰的序列。转基因可以包括其基因型已经通过异源核酸的存在改变的任何细胞、细胞系、愈伤组织、组织、植物部分或植物,这些异源核酸包括最初如此改变的那些转基因以及通过有性杂交或无性繁殖从初始转基因产生的那些。通过常规植物育种方法,通过在此所述的不导致外源多核苷酸的插入的基因组编辑程序,或通过天然存在的事件例如随机异花受精、非重组病毒感染、非重组细菌转化、非重组转座或自发突变的基因组(染色体或染色体外)的改变并不旨在被视为转基因。

[0225] 在本披露的某些实施例中,可育植物是产生活雄配子和雌配子并且是自身可育的植物。这样的自体受精的植物可以产生后代植物,而没有来自任何其他植物的配子及其中所含的遗传物质的贡献。本披露的其他实施例可以涉及使用非自身可育的植物,因为该植物不产生有活力的或在其他情况下能够受精的雄配子或雌配子或二者。如本文所用,“雄性不育植物”是不产生有活力的或在其他情况下能够受精的雄配子的植物。如本文所用,“雌性不育植物”是不产生有活力的或在其他情况下能够受精的雌配子的植物。应当认识到雄性不育植物和雌性不育植物可以分别是雌性可育的和雄性可育的。应当进一步认识到,雄性可育(但雌性不育)植物当与雌性可育植物杂交时可以产生有活力的后代,并且雌性可育(但雄性不育)植物当与雄性可育植物杂交时可以产生有活力的后代。

[0226] 本文中的非常规酵母是指不是酵母属(*Saccharomyces*) (例如,酿酒酵母(*S.cerevisiae*))或裂殖酵母属(*Schizosaccharomyces*) 酵母物种的任何酵母。非常规酵母描述于Non-Conventional Yeasts in Genetics, Biochemistry and Biotechnology: Practical Protocols [遗传学、生物化学和生物技术中的非常规酵母菌:实践方案] (K.Wolf、K.D.Breunig、G.Barth编辑, Springer-Verlag, Berlin, Germany [德国柏林施普林

格出版社],2003),将其通过引用结合在此。在某些实施例中,非常规酵母可能另外地(或可替代地)是偏爱非同源末端连接(NHEJ)DNA修复过程超过由同源重组(HR)介导的修复过程的酵母。按照这些原则,非常规酵母的定义-偏好NHEJ超过HR-被Chen等人(PLoS ONE 8:e57952,其通过引用结合在此)进一步披露。本文中优选的非常规酵母是耶氏酵母属的那些(例如解脂耶氏酵母)。本文中的术语“酵母”是指主要以单细胞形式存在的真菌物种。本文中酵母可以替代地称为“酵母细胞”(还参见2014年8月13日提交的美国临时申请62/036,652,将其通过引用结合在此)。

[0227] “厘摩”(cM)或“图距单位”是两个连锁的基因、标记、靶位点、基因座或它们的任何配对之间的距离,其中1%的减数分裂的产物是重组的。因此,一厘摩与等于两个连锁的基因、标记、靶位点、基因座或它们的任何配对之间的1%平均重组频率的距离相当。

[0228] 本披露可用于繁育包含一个或多个引入性状的植物。最常见的是,作为基于农杆菌、生物射弹、或其他常用程序的转化系统的结果,转基因性状被随机地插入整个植物的基因组中。最近,已经开发了使能够实现定向转基因插入的基因靶向方案。一个重要的技术,位点特异性整合(SSI)使转基因能够靶向至与先前插入的转基因相同的染色体位置。专门设计的大范围核酸酶和专门设计的锌指大范围核酸酶允许研究者设计核酸酶以靶向特定的染色体位置,并且这些试剂允许将转基因靶向到由这些核酸酶切割的染色体位点处。

[0229] 用于真核生物基因组(例如植物基因组)的精准遗传工程的目前使用的系统,依赖归巢核酸内切酶、大范围核酸酶、锌指核酸酶、和转录激活子样效应子核酸酶(TALEN),这些需要用于每个新靶基因座的从头蛋白质工程。本文描述的高度特异性的、RNA指导的DNA核酸酶、指导RNA/Cas9核酸内切酶系统更容易可定制,并且因此当对许多不同靶序列的修饰是目的时更有用。

[0230] 本文所述的指导RNA/Cas系统在核酸酶离靶切割会对靶细胞有毒性的情况下,对于基因组工程(尤其是植物的基因组工程)是尤其有用的。在本文所述的指导RNA/Cas系统的一个实施例中,将表达优化的Cas9基因稳定整合到靶基因组(例如植物基因组)中。Cas9基因的表达是在启动子(例如植物启动子)的控制下进行的,该启动子可以是组成型启动子、组织特异性启动子、或诱导型启动子,例如,温度诱导型、胁迫诱导型、发育阶段诱导型、或化学品诱导型启动子。在不存在指导RNA或crRNA的情况下,Cas9蛋白不能切割DNA,并且因此其在植物细胞中的存在应该具有很少的结果或没有结果。因此,本文所描述的指导RNA/Cas系统的关键优点是产生并且维持能够有效表达对细胞活力具有很少的结果或没有结果的Cas9蛋白的细胞系或转基因生物体的能力。为了诱导在所希望的基因组位点处切割以实现靶向的遗传修饰,可以通过多种方法将指导RNA或crRNA引入包含稳定整合的且表达的cas9基因的细胞中。例如,指导RNA或crRNA可以化学合成或酶合成,并且经由直接的递送方法(例如粒子轰击或电穿孔)引入到Cas9表达细胞中。可替代地,可以经化学合成、酶合成、在生物学系统中合成能够在靶细胞中有效地表达指导RNA或crRNA的基因,并且可以将这些基因经由直接递送方法,例如粒子轰击、电穿孔、或生物学递送方法(例如农杆菌介导的DNA递送)引入Cas9表达细胞中。

[0231] 介导基因靶向的指导RNA/Cas系统可以在用于按如在W02013/0198888(于2013年8月1日公开)中披露的相似的方式指导转基因插入和/或用于产生包含多个转基因的复合转基因性状基因座的方法中使用,其中使用如本文披露的指导RNA/Cas系统来代替使用双链

断裂诱导剂,以引入目的基因。复杂性状基因座包括具有彼此遗传连锁的多个转基因的基因组基因座。通过将独立的转基因插入在彼此的0.1、0.2、0.3、0.4、0.5、1.0、2、或甚至5厘摩(cM)内,这些转基因可以作为单个遗传基因座进行育种(参见,例如,美国专利申请13/427,138或PCT申请PCT/US2012/030061)。在选择包含转基因的植物后,可以将包含(至少)一个转基因的植物进行杂交从而形成包含全部两个转基因的F1。在来自这些F1(F2或BC1)的后代中,1/500的后代将具有重组在相同的染色体上的两个不同的转基因。然后,可以将复合物基因座繁育为具有全部两个转基因性状的单遗传基因座。可以重复该过程以堆叠尽可能多的性状。

[0232] 可以鉴定与目的表型或性状相关的染色体区间。本领域熟知的多种方法可用于鉴定染色体区间。此类染色体区间的边界扩展到涵盖将与控制目的性状的基因连锁的标记。换句话说,扩展染色体区间,这样使得位于区间内的任何标记(包括限定区间的边界的末端标记)可以用作北方叶枯病抗性的标记。在一个实施例中,染色体区间包含至少一个QTL,并且此外,确实可以包含多于一个QTL。相同区间中非常接近的多个QTL可以搅乱特定标记与特定QTL的关联,因为一个标记可显示与多于一个QTL连锁。相反地,例如如果非常接近的两个标记显示与期望表型性状共分离,则有时分不清楚是否那些标记中的每一个鉴定相同QTL或两个不同的QTL。术语“数量性状基因座”或“QTL”意指在至少一种遗传背景下(例如在至少一个育种种群中),与数量表型性状的差异表达关联的DNA区域。QTL的区域涵盖或紧密地连锁于影响所考虑的性状的一个或多个基因。“QTL的等位基因”可以包含在连续的基因组区域或连锁群中的多个基因或其他遗传因子,例如单倍型。QTL的等位基因可以表示在指定窗口内的单倍型,其中所述窗口是可以由一组的一个或多个多态性标记定义和追踪的连续的基因组区域。单倍型可以指定被窗口内的每一标记的等位基因的独特指纹定义。

[0233] 可以使用多种方法来鉴定在靶位点处或靶位点附近具有改变的基因组的那些细胞,而不使用可筛选标记表型。此类方法可被认为是直接分析靶序列以检测靶序列中的任何变化,包括但不限于PCR方法、测序方法、核酸酶消化、DNA印迹法、及其任何组合。

[0234] 可以按不同方式改变蛋白质,这些方式包括氨基酸取代、缺失、截短、和插入。用于此类操作的方法通常是已知的。例如,可以通过在DNA中的突变制备一种或多种蛋白质的氨基酸序列变体。用于诱变和核苷酸序列改变的方法包括,例如,Kunkel,(1985) Proc.Natl.Acad.Sci.USA[美国科学院院报]82:488-92;Kunkel等人,(1987) Meth Enzymol [酶学方法]154:367-82;美国专利号4,873,192;Walker和Gaastra编辑(1983) Techniques in Molecular Biology[分子生物学技术](MacMillan Publishing Company,New York[麦克米伦出版公司,纽约]),以及其中所引用的文献。发现关于不太可能影响蛋白质生物学活性的氨基酸取代的引导,例如,在Dayhoff等人,(1978) Atlas of Protein Sequence and Structure[蛋白质序列和结构图谱集](Natl Biomed Res Found,Washington,D.C.[国家生物医学研究基金会,美国华盛顿哥伦比亚特区])的模型中。保守取代,例如将一个氨基酸与具有相似特性的另一个氨基酸交换,会是优选的。未预期保守缺失、插入、和氨基酸取代会产生在蛋白质特征中的根本变化,并且可以通过常规筛选测定来评价任何取代、缺失、插入、或其组合的作用。对双链-断裂-诱导活性的测定是已知的,并且通常测量试剂对包含靶位点的DNA底物的总体活性和特异性。

[0235] 术语“双子叶植物”是指被子植物的亚类,也被称为“双子叶植物类”,并且包括提

及整株植物、植物器官(例如叶、茎、根、等)、种子、植物细胞、及其后代。如本文所用,植物细胞包括但不限于种子、悬浮培养物、胚胎、分生区、愈伤组织、叶、根、芽、配子体、孢子体、花粉和小孢子。

[0236] 在本披露的上下文中,术语“杂交的”或“杂交”(cross或crossing)是指经由授粉将配子融合从而产生后代(即,细胞、种子、或植物)。该术语涵盖有性杂交(一株植物被另一株植物授粉)和自交(自体授粉,即当花粉和胚珠(或小孢子和大孢子)是来自同一植物或基因相同的植物时)。

[0237] 术语“渗入”指基因座的期望等位基因从一种遗传背景传递到另一种遗传背景的现象。例如,可以经由两个亲本植物之间的有性杂交将指定基因座处的所希望的等位基因的渗入传递给至少一个后代植物,其中至少一个亲本植物在其基因组内具有所希望的等位基因。可替代地,例如等位基因的传递可以通过两个供体基因组之间的重组而发生,例如在融合原生质体中,其中至少其中一个供体原生质体在其基因组中具有所希望的等位基因。所希望的等位基因可以是,例如转基因、修饰的(突变的或编辑的)天然等位基因、或标记或QTL的选择的等位基因。

[0238] 标准的DNA分离、纯化、分子克隆、载体构建、和验证/表征方法是完善确立的,参见,例如Sambrook等人,(1989)Molecular Cloning:A Laboratory Manual[分子克隆:实验室手册]Cold Spring Harbor Laboratory Press,NY[纽约冷泉港实验室出版社]。载体和构建体包括环状质粒和包含目的多核苷酸的线状多核苷酸,以及任选地包括接头、衔接子、用于调节或其他分析的其他组分。在一些实例中,识别位点和/或靶位点可以包含在内含子、编码序列、5' UTR、3' UTR、和/或调节区内。

[0239] 本披露进一步提供了用于在植物、植物细胞、或植物部分中表达指导RNA/Cas系统的表达构建体,该指导RNA/Cas系统能够结合靶位点并且在靶位点处产生双链断裂。在一个实施例中,本披露的表达构建体包括有效地连接至编码Cas基因的核苷酸序列的启动子,以及有效地连接至本披露的指导RNA的启动子。该启动子能够驱动在植物细胞中有效地连接的核苷酸序列的表达。

[0240] 可使用任何植物,包括单子叶植物和双子叶植物。可以使用的单子叶植物的实例包括但不限于,玉米(玉米(*Zea mays*))、稻(稻(*Oryza sativa*))、黑麦(黑麦(*Secale cereale*))、高粱(双色高粱(*Sorghumbicolor*)、高粱(*Sorghum vulgare*))、粟(例如,珍珠粟、御谷(*Pennisetum glaucum*))、黍稷(粟米(*Panicum miliaceum*))、谷子(谷子(*Setaria italica*))、糜子(龙爪稷(*Eleusine coracana*))、小麦(小麦(*Triticum aestivum*))、甘蔗(甘蔗属物种(*Saccharum spp.*))、燕麦(燕麦属(*Avena*))、大麦(大麦属(*Hordeum*))、柳枝稷(柳枝黍(*Panicum virgatum*))、菠萝(菠萝(*Ananas comosus*))、香蕉(香蕉属物种(*Musa spp.*))、棕榈、观赏植物、草坪草、以及其他草。可以使用的双子叶植物的实例包括大豆(大豆(*Glycine max*))、低芥酸菜籽(欧洲油菜(*Brassica napus*)和白菜型油菜(*B.campestris*))、苜蓿(紫花苜蓿(*Medicago sativa*))、烟草(烟草(*Nicotiana tabacum*))、拟南芥(拟南芥(*A.thaliana*))、向日葵(向日葵(*Helianthus annuus*))、棉花(木本棉(*Gossypmm arboreum*))、和花生(花生(*Arachis hypogaea*))、番茄(番茄(*Solanum lycopersicum*))、马铃薯(马铃薯(*Solanum tuberosum*))等。

[0241] 缩写的含义如下:“sec”意指秒、“min”意指分钟、“h”意指小时、“d”意指天、“ $\mu$ L”意



指微升、“mL”意指毫升、“L”意指升、“ $\mu$ M”意指微摩尔、“mM”意指毫摩尔、“M”意指摩尔、“mmol”意指毫摩尔、“ $\mu$ mole”微摩尔、“g”意指克、“ $\mu$ g”意指微克、“ng”意指纳克、“U”意指单位、“bp”意指碱基对、并且“kb”意指千碱基。

[0242] 本文披露的组合物和方法的非限制性实例如下：

[0243] 1. 一种用于在不使用可选择标记的情况下修饰植物细胞的基因组中的核苷酸序列的方法，该方法包括：将能够在位于核苷酸序列中的靶位点中制造双链断裂的指导多核苷酸/Cas核酸内切酶复合物引入至少一个植物细胞中；并且，

[0244] 选择在所述核苷酸序列中具有修饰的植物细胞，其中该选择在不使用可选择标记的情况下发生。

[0245] 1b. 一种用于在不使用可选择标记的情况下修饰植物细胞的基因组中的核苷酸序列的方法，该方法包括：将能够在位于所述核苷酸序列中的靶位点中制造双链断裂的指导RNA/Cas核酸内切酶复合物引入至少一个植物细胞中；并且，选择在所述核苷酸序列中具有修饰的植物细胞，其中所述选择在不使用可选择标记的情况下发生。

[0246] 2. 一种用于在不使用可选择标记的情况下产生在其基因组中具有修饰的核苷酸序列的植物的方法，该方法包括：将能够在位于核苷酸序列中的靶位点中制造双链断裂的指导多核苷酸/Cas核酸内切酶复合物引入至少一个植物细胞中；从所述植物细胞获得植物；并且，选择在所述核苷酸序列中具有修饰的植物，其中所述选择在不使用可选择标记的情况下发生。

[0247] 2b. 一种用于在不使用可选择标记的情况下产生在其基因组中具有修饰的核苷酸序列的植物的方法，该方法包括：将能够在位于核苷酸序列中的靶位点中制造双链断裂的指导RNA/Cas核酸内切酶复合物引入至少一个植物细胞中；从所述植物细胞获得植物；并且，选择在所述核苷酸序列中具有修饰的植物，其中所述选择在不使用可选择标记的情况下发生。

[0248] 3. 一种用于在不使用可选择标记的情况下产生在其基因组中具有修饰的核苷酸序列的植物愈伤组织的方法，该方法包括：将能够在位于核苷酸序列中的靶位点中制造双链断裂的指导多核苷酸/Cas核酸内切酶复合物引入至少一个植物细胞中；从所述植物细胞获得愈伤组织；并且，选择在所述核苷酸序列中具有修饰的愈伤组织，其中所述选择在不使用可选择标记的情况下发生。

[0249] 3b. 一种用于在不使用可选择标记的情况下产生在其基因组中具有修饰的核苷酸序列的植物愈伤组织的方法，该方法包括：将能够在位于核苷酸序列中的靶位点中制造双链断裂的指导RNA/Cas核酸内切酶复合物引入至少一个植物细胞中；从所述植物细胞获得愈伤组织；并且，选择在所述核苷酸序列中具有修饰的愈伤组织，其中所述选择在不使用可选择标记的情况下发生。

[0250] 4. 如实施例1-3b所述的方法，其中所述修饰选自下组，该组由以下各项组成：所述靶位点中至少一个核苷酸的插入、至少一个核苷酸的缺失、或至少一个核苷酸的取代。

[0251] 5. 如实施例1-3b所述的方法，所述方法进一步包括将多核苷酸修饰模板引入所述植物细胞中，其中所述多核苷酸修饰模板包含所述核苷酸序列的至少一个核苷酸修饰。

[0252] 6. 如实施例5所述的方法，其中所述多核苷酸修饰模板的所述至少一个核苷酸修饰选自下组，该组由以下各项组成：(i) 至少一个核苷酸的替代、(i i) 至少一个核苷酸的缺

失、(iii)至少一个核苷酸的插入、和(iv)(i)-(iii)的任何组合。

[0253] 7.如实施例1-3b所述的方法,所述方法进一步包括将供体DNA引入(a)的植物细胞中,其中所述供体DNA包含至少一个待插入所述靶位点中的目的多核苷酸。

[0254] 8.如实施例1-3b所述的方法,其中所述引入不包括将可选择标记引入所述细胞中。

[0255] 9.如实施例1-3b所述的方法,其中所述引入不包括将破坏的可选择标记基因恢复成编码功能性可选择标记蛋白的非破坏的可选择标记基因。

[0256] 10.如实施例1-3b所述的方法,其中所述引入不导致可选择标记在所述细胞内的产生。

[0257] 11.如实施例1-3b所述的方法,其中所述选择不包括可选择标记的鉴定或使用。

[0258] 12.如实施例1-3b所述的方法,其中所述选择经由对所述植物的分离的DNA进行测序来发生。

[0259] 13.如实施例1-3b所述的方法,其中将所述指导RNA/Cas核酸内切酶复合物在体外组装并作为核糖核苷酸-蛋白质复合物引入。

[0260] 13b.如实施例1-3b所述的方法,其中将所述指导RNA/Cas核酸内切酶复合物作为核糖核苷酸-蛋白质复合物引入。

[0261] 13c.如实施例1-3b所述的方法,其中在不使用重组DNA构建体的情况下将所述指导RNA/Cas核酸内切酶复合物引入所述细胞中。

[0262] 14.如实施例1-3b所述的方法,其中将所述指导RNA/Cas核酸内切酶复合物的组分作为指导RNA和Cas核酸内切酶蛋白引入,所述指导RNA和所述Cas核酸内切酶蛋白能够形成所述指导RNA/Cas核酸内切酶复合物。

[0263] 15.如实施例1-3b所述的方法,其中将所述指导RNA/Cas核酸内切酶复合物的组分作为编码所述Cas核酸内切酶的mRNA和作为包含指导RNA的RNA引入。

[0264] 16.如实施例1-3b所述的方法,其中将所述指导RNA/Cas核酸内切酶复合物的组分作为编码指导RNA和Cas核酸内切酶蛋白的重组DNA分子引入。

[0265] 17.如实施例1-3b所述的方法,其中所述指导RNA/Cas核酸内切酶复合物在所述细胞内部组装。

[0266] 18.如实施例13所述的方法,其中将所述核糖核苷酸-蛋白质复合物包被到颗粒递送基质上或与颗粒递送基质组合以形成核糖核苷酸-蛋白质-基质复合物,其中将所述核糖核苷酸-蛋白质-基质复合物引入所述细胞中。

[0267] 18b.如实施例18所述的方法,其中所述颗粒递送基质包含至少一种微粒。

[0268] 18c.如实施例18所述的方法,其中该颗粒递送基质包含与阳离子脂质组合的至少一种微粒。

[0269] 18d.如实施例18c-18c所述的方法,其中所述微粒选自下组,该组由以下各项组成:金颗粒、钨颗粒和碳化硅晶须颗粒。

[0270] 18e.如实施例1-3b所述的方法,其中所述植物细胞是体细胞胚胎细胞。

[0271] 19.如实施例1-3b所述的方法,其中所述植物细胞不是原生质体。

[0272] 20.如实施例1-3b所述的方法,其中所述植物细胞选自下组,该组由以下各项组成:单子叶植物和双子叶植物细胞。

[0273] 21.如实施例21所述的方法,其中所述植物细胞选自下组,该组由以下各项组成:玉蜀黍、稻、高粱、黑麦、大麦、小麦、粟、燕麦、甘蔗、草坪草、或柳枝稷、大豆、低芥酸菜籽、苜蓿、向日葵、棉花、烟草、花生、马铃薯、番茄、烟草、拟南芥和红花细胞。

[0274] 22.如实施例1-3b所述的方法,所述方法进一步包括从所述植物细胞再生植物。

[0275] 23.一种通过如实施例22所述的方法产生的植物。

[0276] 24.一种如实施例23所述的植物的后代植物,其中所述后代植物没有选自下组的任何组分,该组由以下各项组成:指导RNA、Cas核酸内切酶、多核苷酸修饰模板和供体DNA。

[0277] 实例

[0278] 在以下实例中,除非另有说明,份数和百分比以重量计,并且度数为摄氏度。应当理解的是,尽管这些实例说明了本披露的实施例,但仅是通过说明的方式给出的。从上述讨论和这些实例,本领域技术人员可以对本披露进行各种改变和修改以使其适应各种用途和条件。这样的修改也旨在落入所附权利要求的范围内。

[0279] 实例1

[0280] 通过递送Cas9核酸内切酶和指导RNA表达盒来修饰植物细胞基因组中的靶DNA序列

[0281] 来自化脓性链球菌M1GAS (SF370)的Cas9基因 (SEQ ID NO:1) 是使用本领域已知的标准技术进行玉蜀黍密码子优化的,并且引入马铃薯ST-LS1内含子 (SEQ ID NO:2) 以便于消除该基因在大肠杆菌和农杆菌中的表达。为了促进Cas9蛋白质在玉蜀黍细胞中的核定位,将猿猴病毒40 (SV40) 单分型氨基端核定位信号 (MAPKKRKV, SEQ ID NO:3) 和根癌农杆菌二分型VirD2T-DNA边界核酸内切酶羧基端核定位信号 (KRPRDRHDGELGGRKRAR, SEQ ID NO:4) 分别并入Cas9可读框的氨基和羧基端处。将玉蜀黍优化的Cas9基因通过标准分子生物学技术与玉蜀黍组成型启动子 (泛素) 有效地连接。通过添加来自马铃薯蛋白酶抑制剂II基因 (PinII) 的3' 序列来终止转录以产生UBI:Cas9:PinII载体。泛素驱动的玉蜀黍优化的Cas9表达盒的序列示于SEQ ID NO:5中。

[0282] 使用由Mali等人,2013 (Science [科学] 339:823-26) 描述的方法来设计单指导RNA (gRNA)。将玉蜀黍U6聚合酶III启动子和终止子分离并将其分别用于引导gRNA的起始和终止。以反向串联取向引入两个BbsI限制性核酸内切酶位点,其中切割以向外方向取向 (如Cong等人,2013 (Science [科学] 339:819-23) 中所述) 以促进将玉蜀黍基因组DNA靶序列快速引入gRNA表达构建体中。仅使用以G核苷酸开始的靶序列来促进gRNA的有利的聚合酶III表达。将gRNA表达盒亚克隆到Bluescript SK载体 (SEQ ID NO:6) 中。

[0283] 为了测试玉蜀黍优化的Cas9-gRNA复合物是否可以通过非同源末端连接 (NHEJ) 修复途径来识别、切割和促进玉蜀黍染色体DNA中的靶向突变,将5个玉蜀黍基因座 (每个基因座中三个不同的基因组序列) 靶向用于切割 (参见表2) 并通过扩增子深度测序检查突变的存在。

[0284] 表2. 由Cas9-gRNA系统靶向的玉蜀黍基因组位点

[0285]

基因座	位置	靶位点名称	玉蜀黍基因组 靶位点序列	PAM 序 列	SEQ ID NO.
MS26	染色体 1: 51.81cM	MS26Cas-1	GTACTCCATCCGCCCATCGAGTA	GGG	7
		MS26Cas-2	GCACGTACGTCACCATCCCGC	CGG	8
		MS26Cas-3	GACGTACGTGCCCTACTCGAT	GGG	9
LIG	染色体 2: 28.45cM	LIGCas-1	GTACCGTACGTGCCCGGCGG	AGG	10
		LIGCas-2	GGAATTGTACCGTACGTGCCC	CGG	11
		LIGCas-3	GCGTACGCGTACGTGTG	AGG	12
MS45	染色体 9: 119.15cM	MS45Cas-1	GCTGGCCGAGGTCGACTAC	CGG	13
		MS45Cas-2	GGCCGAGGTCGACTACCGGC	CGG	14
		MS45Cas-3	GGCGCGAGCTCGTGCTTCAC	CGG	15
ALS1	1 - 染色体	ALSCas-1	GGTGCCAATCATGCGTCG	CGG	16
		ALSCas-2	GGTCGCCATCACGGGAC	AGG	17

[0286]

和 ALS2	4: 107.73cM 2 - 染色体 5: 115.49cM	ALSCas-3	GTCGCGGCACCTGTCCCGTGA	TGG	18
-----------	---	----------	-----------------------	-----	----

[0287] MS26 = 雄性不育基因26, LIG = Liguleless-1基因启动子, MS45 = 雄性不育基因45, ALS1 = 乙酰乳酸合酶基因1 (染色体4), ALS2 = 乙酰乳酸合酶基因2 (染色体5)。

[0288] 通过用可选择和可视的标记 (UBI:MoPAT:DsRED融合物) (参见实例8) 以及发育基因ZmODP-2 (BBM) 和ZmWUS2 (WUS) (参见实例9) 进行粒子轰击, 将含有特定玉蜀黍可变靶向结构域的玉蜀黍优化的Cas9核酸内切酶和gRNA表达盒共同递送到60-90个Hi-II未成熟的玉蜀黍胚胎中。将仅用Cas9或gRNA表达盒转化的Hi-II玉蜀黍胚胎作为阴性对照。7天后, 合并来自每种处理的20-30个最均匀转化的胚胎 (基于DsRED荧光蛋白的瞬时表达) 并提取总基因组DNA。用Phusion®高保真PCR预混合物 (新英格兰生物实验室 (New England Biolabs), M0531L), 添加对于扩增子-特异性条形码必要的序列, 对预期靶位点周围的区域进行PCR扩增, 并且使用“加尾的”引物通过两轮PCR进行Illumina测序。初级PCR反应中使用的引物示于表3中。

[0289] 表3. PCR引物序列

[0290]

靶位点	引物	初级 PCR 引物序列	SEQ ID NO:
MS26Cas-1	正向	CTACACTCTTTCCCTACACGACGCTCTTCCGATCT AGGACCGGAAGCTCGCCGCGT	19
MS26Cas-1	反向	CAAGCAGAAGACGGCATAACGAGCTCTTCCGATC TTCCTGGAGGACGACGTGCTG	20
MS26Cas-2	正向	CTACACTCTTTCCCTACACGACGCTCTTCCGATCT AAGGTCCTGGAGGACGACGTGCTG	21
MS26Cas-2	反向	CAAGCAGAAGACGGCATAACGAGCTCTTCCGATC TCCGGAAGCTCGCCGCGT	22
MS26Cas-3	正向	CTACACTCTTTCCCTACACGACGCTCTTCCGATC TTCCTCCGGAAGCTCGCCGCGT	23
MS26Cas-3	反向	CAAGCAGAAGACGGCATAACGAGCTCTTCCGATC TTCCTGGAGGACGACGTGCTG	20
LIGCas-1	正向	CTACACTCTTTCCCTACACGACGCTCTTCCGATCT AGGACTGTAACGATTACGCACCTGCTG	24
LIGCas-1	反向	CAAGCAGAAGACGGCATAACGAGCTCTTCCGATC TGCAAATGAGTAGCAGCGCACGTAT	25
LIGCas-2	正向	CTACACTCTTTCCCTACACGACGCTCTTCCGATC TTCCTCTGTAACGATTACGCACCTGCTG	26
LIGCas-2	反向	CAAGCAGAAGACGGCATAACGAGCTCTTCCGATC TGCAAATGAGTAGCAGCGCACGTAT	25
LIGCas-3	正向	CTACACTCTTTCCCTACACGACGCTCTTCCGATCT AAGGCGCAAATGAGTAGCAGCGCAC	27
LIGCas-3	反向	CAAGCAGAAGACGGCATAACGAGCTCTTCCGATC TCACCTGCTGGGAATTGTACCGTA	28
MS45Cas-1	正向	CTACACTCTTTCCCTACACGACGCTCTTCCGATCT	29

[0291]

		AGGAGGACCCGTTTCGGCCTCAGT	
MS45Cas-1	反向	CAAGCAGAAGACGGCATAACGAGCTCTTCCGATC TGCCGGCTGGCATTGTCTCTG	30
MS45Cas-2	正向	CTACACTCTTTCCCTACACGACGCTCTTCCGATC TTCCTGGACCCGTTTCGGCCTCAGT	31
MS45Cas-2	反向	CAAGCAGAAGACGGCATAACGAGCTCTTCCGATC TGCCGGCTGGCATTGTCTCTG	30
MS45Cas-3	正向	CTACACTCTTTCCCTACACGACGCTCTTCCGATC TGAAGGGACCCGTTTCGGCCTCAGT	32
MS45Cas-3	反向	CAAGCAGAAGACGGCATAACGAGCTCTTCCGATC TGCCGGCTGGCATTGTCTCTG	30
ALSCas-1	正向	CTACACTCTTTCCCTACACGACGCTCTTCCGATCT AAGGCGACGATGGGCGTCTCCTG	33
ALSCas-1	反向	CAAGCAGAAGACGGCATAACGAGCTCTTCCGATC TGCGTCTGCATCGCCACCTC	34
ALSCas-2	正向	CTACACTCTTTCCCTACACGACGCTCTTCCGATC TTTCCCACGATGGGCGTCTCCTG	35
ALSCas-2	反向	CAAGCAGAAGACGGCATAACGAGCTCTTCCGATC TGCGTCTGCATCGCCACCTC	34
ALSCas-3	正向	CTACACTCTTTCCCTACACGACGCTCTTCCGATC TGGAACGACGATGGGCGTCTCCTG	36
ALSCas-3	反向	CAAGCAGAAGACGGCATAACGAGCTCTTCCGATC TGCGTCTGCATCGCCACCTC	34

[0292] 在次级PCR反应中使用的引物为AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACG (正向, SEQ ID NO:37) 和CAAGCAGAAGACGGCATA (反向, SEQ ID NO:38)。

[0293] 用凯杰公司 (Qiagen) PCR纯化旋转柱对所得PCR扩增物进行纯化, 用基于Hoechst染料的荧光测定测量浓度, 按等摩尔比组合, 并且在Illumina的MiSeq个人测序仪上用PhiX对照v3 (Illumina, FC-110-3001) 的30%-40% (v/v) 加标以抵消序列偏差来进行单读数100个核苷酸长度的深度测序。将只在集中在预期切割位点上的10个核苷酸窗口内产生的具有 $\geq 1$ 个核苷酸插入缺失并且未在阴性对照中以相似的水平发现的那些读数分类为突变。对具有相同突变的突变体读数进行计数并使其压缩成单个组, 并且目视确认在预期切割位点内产生的前10个最常见的突变。然后基于含有与条形码和正向引物完全匹配的合适长度的读数的总数, 使用突变的总数来计算突变体读数的百分比。

[0294] 通过扩增子深度测序揭示的针对靶向全部15个位点的Cas9-gRNA系统的突变频率示于表4中。

[0295] 表4.5个靶基因座 (15个靶位点) 处的突变体读数的百分比

[0296]

靶标	DSB 试剂	读数的总数	突变体靶基因读数的数目	靶基因中突变体读数的百分比
LIG (染色体 2)	LIG-CR1 gRNA+Cas9	716,854	33,050	4.61%

[0297]

	LIG-CR2 gRNA+Cas9	711,047	16,675	2.35%
	LIG-CR3 gRNA+Cas9	713,183	27,959	3.92%
MS26 (染色体 1)	MS26-CR1 gRNA+Cas9	575,671	10,073	1.75%
	MS26-CR2 gRNA+Cas9	543,856	16,930	3.11%
	MS26-CR3 gRNA+Cas9	538,141	13,879	2.58%
MS45 (染色体 9)	MS45-CR1 gRNA+Cas9	812,644	3,795	0.47%
	MS45-CR2 gRNA+Cas9	785,183	14,704	1.87%
	MS45-CR3 gRNA+Cas9	728,023	9,203	1.26%
ALS1 (染色体 4) 和 ALS2 (染色体 5)	ALS-CR1 gRNA+Cas9	434,452	9,669	2.23%
	ALS-CR2 gRNA+Cas9	472,351	6,352	1.35%
	ALS-CR3 gRNA+Cas9	497,786	8,535	1.72%
对照	仅 Cas9	640,063	1	0.00%
	仅 LIG-CR1 gRNA	646,774	1	0.00%

[0298] 进一步分析证明,由Cas9-gRNA系统促进的最常见类型的突变是单核苷酸插入(例如,参见图1,SEQ ID NO:49-58)。对大多数测试的gRNA观察到类似的结果(表5)。

[0299] 表5.由Cas9-gRNA系统促进的15个靶位点中单核苷酸插入和缺失的频率

[0300]

靶标 基因座	DSB 试剂	突变体读数总数中单核 核苷酸插入%	突变体读数总数中单核 核苷酸缺失%
LIG	LIG-CR1 gRNA+Cas9	86%	5%
	LIG-CR2 gRNA+Cas9	49%	25%
	LIG-CR3 gRNA+Cas9	62%	20%
MS26	MS26-CR1 gRNA+Cas9	46%	16%
	MS26-CR2 gRNA+Cas9	78%	8%
	MS26-CR3 gRNA+Cas9	45%	18%
MS45	MS45-CR1 gRNA+Cas9	45%	17%
	MS45-CR2 gRNA+Cas9	41%	23%
	MS45-CR3 gRNA+Cas9	20%	24%
ALS1 (染色体 4) 和 ALS2 (染色体 5)	ALS-CR1 gRNA+Cas9	22%	76%
	ALS-CR2 gRNA+Cas9	60%	27%
	ALS-CR3 gRNA+Cas9	84%	12%

[0301] 这个实例证明RNA指导的Cas9产生双链断裂,导致高频率的突变。对多个靶位点的突变分析表明,虽然观察到各种大小的缺失和/或插入,但是对于玉蜀黍中测试的大多数靶位点而言,单个核苷酸插入和单个核苷酸缺失是由Cas9-gRNA技术产生的最常见类型的突变。

[0302] 实例2

[0303] 编辑的乙酰乳酸合酶基因赋予氯磺隆抗性

[0304] 该实例表明引入天然玉蜀黍乙酰乳酸合酶 (ALS) 基因的核苷酸序列中的一个或多个特定改变导致对磺酰脲类除草剂 (特别是氯磺隆) 的抗性。

[0305] 在玉蜀黍中存在两个ALS基因,分别位于染色体4和5上的ALS1 (SEQ ID NO:39) 和ALS2 (SEQ ID NO:40),在DNA水平上具有94%的序列同一性。

[0306] 该ALS蛋白含有N-末端转运,并且在转运到叶绿体中并随后切割转运肽之后形成成熟蛋白。该成熟蛋白从残基S41开始,导致598个氨基酸的成熟蛋白,预测分子量为65kDa (SEQ ID NO:41)。

[0307] 与内源性玉蜀黍乙酰乳酸合酶蛋白相比,导致单个氨基酸残基 (P165A或P165S,灰



色加框)改变的ALS1或ALS2的核苷酸序列的改变在玉蜀黍中提供了对除草剂的抗性。

[0308] 由于乙酰乳酸合酶是植物细胞功能的关键酶,所以ALS1和ALS2基因的同时双等位基因敲除将不期望会幸存。因此,基于ALS1和ALS2核苷酸序列之间的多态性,鉴定和测试ALS2-特异性ALSCas-4靶位点。将靶向ALS1和ALS2基因二者的ALSCas-1指导RNA表达构建体用作对照。表6呈现了有关ALSCas-1和ALSCas-4靶位点的信息。

[0309] 表6. ALSCas-1和ALSCas-4 (ALS2特异性) 靶位点

[0310]

基因座	位置	靶位点名称	玉蜀黍基因组靶位点序列	PAM 序列	SEQ ID NO:
ALS1 和 ALS2	染色体 4: 107.73cM 和染色体 5: 115.49cM	ALSCas-1	GGTGCCAATCATGCGTC G	CGG	16
		ALSCas-4	GCTGCTCGAITCCGTCC CCA	TGG	42

[0311] ALSCas-4靶位点和PAM中加下划线的核苷酸在ALS1基因中是不同的。

[0312] 通过如实例1所述的扩增子深度测序确定了ALSCas-1和ALSCas-4处的突变频率,并示于表7中。

[0313] 表7. 通过扩增子深度测序重新获得的ALSCas-1和ALSCas-4靶位点处的突变频率。

[0314]

靶位点	总读数	突变体读数 (ALS1)	突变体读数 (ALS2)
ALSCas-1	204,230	2704 (1.3%)	5072 (2.5%)
ALSCas-4	120,766	40 (0.03%)	3294 (2.7%)

[0315] 这些结果表明,ALSCas-4gRNA/Cas9系统以比ALS1基因高约90倍的效率使ALS2基因突变。因此,选择ALSCas-4靶位点和相应的ALS-CR4gRNA,用于ALS基因编辑实验。

[0316] 为了产生ALS2编辑的等位基因,将包含同源性片段的794bp多核苷酸修饰模板 (SEQ ID NO:43) 克隆到质粒载体中,并将两个127nt单链多核苷酸修饰模板 (也称为DNA寡核苷酸,0ligo1,SEQ ID NO:44,和0ligo2,SEQ ID NO:45) 作为多核苷酸修饰模板进行测试 (图2)。该794bp片段具有与0ligo1相同的序列修饰。与天然序列相比,这些多核苷酸修饰模板 (修复模板) 含有若干个核苷酸改变。单链0ligo1和794bp修复模板包括单个核苷酸改变,其将指导对应于氨基酸位置165处的脯氨酸的DNA序列编辑成丝氨酸 (P165S) 以及ALS-CR4靶位点和PAM序列内的三个另外的改变。修复模板内PAM序列的修饰将甲硫氨酸密码子 (AUG) 改变为天然存在于ALS1基因中的异亮氨酸 (AUU)。还测试了第二个127nt单链寡核苷酸修复模板 (0ligo2),其保留序列中的位置157处的甲硫氨酸,但包含将影响与ALS-CR4gRNA碱基配对的三个另外的单个核苷酸改变 (图2)。

[0317] 在DNA表达盒中用两种寡核苷酸或单一质粒修复模板 (Cas9、ALS-CR4gRNA和MoPAT-DsRED) 轰击每次处理的大约1,000个未成熟的胚胎,并置于培养基上以选择由PAT赋予的双丙氨酸抗性。转化后五周,将在选择培养基上生长的二百个 (每次处理) 随机选择的独立幼龄愈伤组织部分从胚胎中分离出来并转移到新鲜的双丙氨酸平板上。将具有发育愈伤组织事件的剩余胚胎 (每次处理 > 800个) 转移至含有100ppm氯磺隆 (作为编辑的ALS2基因的直接选择) 的平板上。一个月后,通过PCR扩增和测序来分析总共384个随机挑选的在双

丙氨酸上生长的愈伤组织部分(对于每个修复模板约130个事件)和7个在具有氯磺隆的培养基上继续生长的愈伤组织部分。在九个愈伤组织部分检测到编辑的ALS2等位基因:两个来源于生长在双丙氨酸上的愈伤组织部分并使用794bp修复DNA模板产生,并且其余7个来源于使用127nt单链寡核苷酸编辑(三个由Oligo1编辑和四个由Oligo2编辑)的氯磺隆抗性愈伤组织部分。作为NHEJ修复的结果,这些愈伤组织部分中的第二个ALS2等位基因发生了突变。对ALS1基因的分析仅揭示了证实ALS-CR4gRNA的高特异性的野生型序列。

[0318] 从含有编辑的ALS2等位基因的9个愈伤组织部分中的7个再生植物用于另外的分子分析和后代测试。对ALS2等位基因的DNA序列分析证实存在P165S修饰(ALS2-P165S)以及与各自修复模板相关的其他核苷酸改变。分析由不同愈伤组织事件(794bp修复DNA和Oligo2)产生的两种T0植物的T1和T2后代以评估编辑的ALS2等位基因的遗传。通过测序分析了来源于使用来自野生型Hi-II植物的花粉进行杂交的后代植物,并证明了在亲本植物中观察到的编辑的等位基因的有性传播预期是1:1分离比率(分别为57:56和47:49)。为了测试编辑的ALS序列是否赋予除草剂抗性,将选择的具有编辑的和野生型ALS2等位基因的四周龄的分开T1植物用四种不同浓度的氯磺隆(50、100(1x)、200和400mg/升)进行喷雾。处理后三周,具有编辑的等位基因的植物显示正常表型(图3-左),而仅具有野生型等位基因的植物显示出强烈的衰老迹象(图3-右)。

[0319] 除了对磺酰脲类除草剂(特别是氯磺隆)具有抗性以外,ALS基因可以被修饰以赋予对其他类别的AHAS抑制剂(包括三唑并嘧啶、嘧啶基硫代苯甲酸酯和咪唑啉酮除草剂)的抗性(Tan S, Evans RR, Dahmer ML, Singh BK, Shaner DL (2005) Imidazolinone-tolerant crops: history, current status and future [耐咪唑啉酮的作物: 历史、现状和未来] Pest Management Science [有害生物管理科学] 61: 246-257)。因此,对ALS基因的修饰不应该局限于本文描述的变化和赋予氯磺隆抗性。

[0320] 这些实验表明,Cas9-gRNA可以刺激玉蜀黍中HDR依赖性的靶向序列修饰,这导致具有编辑的内源基因的植物,所述编辑的内源基因适当地传递至后代。数据还表明,在内源性启动子下的单个编辑的ALS2等位基因在玉蜀黍中提供除草剂抗性。

### [0321] 实例3

#### [0322] ALS2作为内源性可选择标记基因

[0323] 本实例展示了特异性编辑的ALS2基因可以如何用于在细胞中产生可选择标记,从而替换目前用于植物转化中的外源性标记基因的递送。

[0324] 由于植物转化(转基因事件恢复)频率相对较低,所以对各种除草剂提供抗性的可选择标记基因通常与性状基因共同递送。为了赋予抗性,这些可选择标记基因需要稳定整合到植物基因组中,并且必须在连续的世代中被切除或人工繁殖消除。一些天然植物基因可以被特异性修饰(编辑)以赋予对除草剂的抗性。如实例2中所述,具有单个氨基酸改变的ALS2基因提供对氯磺隆的抗性。因此,可以预期的是,可以在没有外源提供的标记基因的情况下使用基因诱变、基因编辑或性状基因的共同递送以及一致的ALS2基因编辑。鉴于由Cas9-gRNA系统(实例1)产生的DSB的NHEJ修复导致的高突变频率,该方法对于基因诱变可能是有用的。在这种情况下,突变事件的频率预计将依赖于HDR介导的ALS基因编辑。关于基因编辑,很可能植物细胞中两种低频HDR依赖性基因组编辑过程(一种用于ALS基因修复的选择,并且另一种用于内源性基因编辑或性状基因整合)将使得使用一致的ALS2基因编辑

的方法比较不切实际。

[0325] 以下实例描述了一种方法,该方法允许克服这种低效率(和不切实际)并且改善选择对选择试剂有抗性的植物细胞的可能性。该方法不依赖于HDR依赖性基因编辑,而是依赖于通过NHEJ DNA修复的靶向诱变来恢复基因功能,这在植物体细胞中(比HDR)更常见。如实例2中所述,在玉蜀黍中存在两种ALS基因,分别位于染色体4和染色体5上的ALS1和ALS2。对两种ALS基因中任一种的特异性编辑将赋予除草剂抗性。这些基因在植物新陈代谢中起着至关重要的作用,因此,靶向它们两者并使其突变导致细胞死亡。因此,在这个实例中,通过使用不靶向ALS1基因的ALS2特异性gRNA,修饰只涉及ALS2基因,因此ALS1保持野生型。特别地,将两种修饰都引入ALS2基因;首先,特定的一个或多个核苷酸改变,例如在核苷酸位置493处的C到T(图2,oligo1)或分别在核苷酸位置493和495处的C到T以及C到G(图2,oligo2)将在氨基酸位置165处的脯氨酸转化成丝氨酸(命名为ALS2-P165S),赋予对氯磺隆的抗性(详见实例2)。其次,去除单个核苷酸(例如核苷酸位置165处的G(图4A-4B)),这导致翻译移码(图4B),并因此丧失ALS2介导的氯磺隆抗性(命名为ALS2-P165S-CC $\Delta$ )。尽管预期许多设计,但是为了ALS2基因修复的最高频率,单个核苷酸位置(示于图4A-4C中的实例)优选地应为:i)来自gRNA/Cas核酸内切酶靶位点中的PAM序列的第4个核苷酸上游(5'),和ii)密码子中的第3个核苷酸(图4A)。该第3位置对于大多数氨基酸是灵活的,并且在20种氨基酸中的8种中可以被四个核苷酸中的任何一个占据(表8)。考虑到这种灵活性,预计与第1或第2位置相比时,在第3位置正确修复的频率更高。

[0326] 表8.遗传密码。

[0327]

氨基酸	密码子	缩写的密码子
丙氨酸/Ala	GCU、GCC、GCA、GCG	GCN
精氨酸/Arg	CGU、CGC、CGA、CGG、AGA、AGG	CGN、MGR
甘氨酸/Gly	GGU、GGC、GGA、GGG	GGN
亮氨酸/Leu	CUU、CUC、CUA、CUG、UUA、UUG	CUN、YUR
脯氨酸/Pro	CCU、CCC、CCA、CCG	CCN
丝氨酸/Ser	UCU、UCC、UCA、UCG、AGU、AGC	UCN、AGY
苏氨酸/Thr	ACU、ACC、ACA、ACG	ACN

[0328]

缬氨酸/Val	GUU、GUC、GUA、GUG	GUN
异亮氨酸/Ile	AUU、AUC、AUA	AUH
天冬酰胺/Asn	AAU、AAC	AAY
天冬氨酸/Asp	GAU、GAC	GAY
半胱氨酸/Cys	UGU、UGC	UGY
谷氨酰胺/Gln	CAA、CAG	CAR
谷氨酸/Glu	GAA、GAG	GAR
组氨酸/His	CAU、CAC	CAY
赖氨酸/Lys	AAA、AAG	AAR
苯丙氨酸/Phe	UUU、UUC	UUY
酪氨酸/Tyr	UAU、UAC	UAY
甲硫氨酸/Met	AUG	-
色氨酸/Trp	UGG	-

[0329] 选择满足以上标准的四个不同靶位点和相应的gRNA。除了上述优选的单核苷酸位置之外,相应的gRNA应促进高频率的突变,其中高百分比的突变表示在切割位点的单个核苷酸插入。鉴定到,被测试的四个靶位点中只有一个满足了所有描述的偏好,并因此适合于这个实验(被称为ALSCas-7;表9)。

[0330] 表9. ALS2特异性ALSCas-7靶位点和通过扩增子深度测序进行的ALS-CR7-gRNA评估。

[0331]

靶标	靶位点序列	突变体 读数%	具有1 bp 插 入的突变体 读数%	具有1 bp 缺 失的突变体 读数%	SEQ ID NO:
ALSCas-4 (对照)	GCTGCTCGA <u>IT</u> CCGTC <u>CCCA</u>	2.73%	-	-	42
ALSCas-7	GCT <u>CCCCCGGCCAC</u> CCCGCTC	2.99%	2.23% (75%)	0.14% (5%)	79

[0332] 然后,进一步修饰赋予对氯磺隆有抗性的ALS2-P165S基因(产生破坏的基因):通过去除摆动位置处的G核苷酸(密码子中的第3个核苷酸位置)来改变ALSCas-7靶位点中由CCG编码的脯氨酸密码子(表9中加下划线的),这导致翻译移码和破坏的基因(被称为ALS2-P165S-CC $\Delta$ ;图4A-4B)。如实例1所示,对由玉蜀黍中的Cas9-gRNA系统产生的DSB的修复,并且通过NHEJ进行修复,通常导致在切割位点处的单个核苷酸插入。因此,通过在修饰的ALSCas-7位点处(被称为ALSCas-7-1 (GCTCCCCCGGCCACCCCTC; SEQ ID NO: 80))产生双链断裂(DSB)以及通过NHEJ对其的修复(参见图4B-4C),可以恢复ALS2-P165S-CC $\Delta$ 基因的功能和相继地细胞对氯磺隆有抗性。

[0333] 基于本披露,可以设想同时递送两种或更多种gRNA,这时一种gRNA通过NHEJ靶向并激活破坏的ALS2-P165S-CC $\Delta$ 基因(由此赋予除草剂抗性)并且其他一种或多种gRNA促进

不同于ALS2的一个或多个位点处的一个或多个DSB并促进希望的基因组修饰(例如靶向诱变、缺失、基因编辑或位点特异性性状基因插入)。这种方法可以允许完全瞬时的靶向基因组修饰,因为所有其他必需组分(Cas9、gRNA)能以蛋白质和/或体外转录的RNA分子的形式递送。

[0334] 为了测试这种方法,产生了具有上述特异性修饰的ALS2基因的玉蜀黍植物(Hi-II基因型)。首先,ALS2序列在氨基酸位置165处被修饰以赋予对氯磺隆的抗性,如实例2所述。然后用Cas9和靶向ALSCas-7靶位点的ALS-CR7gRNA、可选择标记(UBI:MoPAT-DSRED融合)和细胞发育增强基因的轰击来自对于编码而言纯合的植物的未成熟胚胎(详见实例8和9)。通过测序来分析来自双丙氨酸抗性愈伤组织部分的再生物。鉴定了几个具有单核苷酸缺失(在核苷酸位置165处的G)的T0植物(图4A-4B)。这种缺失导致翻译移码(图4B),并因此导致ALS2-P165S介导的氯磺隆抗性的丧失。再生对于这两种编辑(ALS2-P165S-CC $\Delta$ )而言纯合的植物,对其通过测序证实,并通过喷洒氯磺隆来测试除草剂抗性的丧失。

[0335] 将来自具有特异性修饰的内源性ALS2基因(ALS2-P165S-CC $\Delta$ )的纯合植物的胚胎用于概念验证实验。为了证明编辑的破坏的ALS2-P165S-CC $\Delta$ 基因可以如上所述被修复(以便其编码功能性蛋白质)并用作可选择标记,将编码Cas9的DNA载体、靶向ALSCas-7-1位点的gRNA(称为ALS-CR7-1)、细胞发育增强基因(ZmODP2和ZmWUS)和MS45-CR2gRNA共同递送到玉蜀黍(Hi-II)未成熟的胚胎细胞中。轰击后一周,将胚胎转移到具有100ppm氯磺隆的培养基中进行选择。大约30%的胚胎(290个中有84个)发育成除草剂抗性愈伤组织事件,将其通过测序进行分析。绝大多数事件(79个事件)在预期的ALSCas-7-1DSB位点处显示出单核苷酸插入(图4C)和ALS2-P165S基因的完全恢复。四个事件没有插入,但2bp或5bp缺失将该基因重新置于框架中,并且一个事件似乎为逃逸(escape)。83个事件中的五十个(60%)也在MS45Cas-2靶位点处显示出突变。

[0336] 本实例通过使用指导的Cas核酸内切酶系统证明了特定修饰的、无活性ALS2作为内源性可选择标记基因的概念验证和实用性。基于本文所述的结果,本领域技术人员可以使用所描述的方法并将其扩展到任何类似修饰的内源性或预整合的一种或多种外源基因,替换目前用于植物基因组编辑实验中的可选择标记基因的共同递送。

[0337] 实例4

[0338] 使由破坏的基因编码的非功能性蛋白质恢复功能的替代设计。

[0339] 在前面的实例中,将序列改变掺入到编码区或ALS2-P165S中。预计产生破坏的基因(不编码功能性蛋白质)的其他序列变化也可被设计为用作再激活序列。本实例描述了再激活序列的产生,该再激活序列不依赖于编码序列内密码子的恢复,而是依赖于消除ALS2-P165S的初级翻译起始密码子的上游和框外的起始密码子。

[0340] 根据真核翻译起始的扫描模型,相对于mRNA的5'帽的第一个AUG密码子被用于启动蛋白质合成(Kozak M.1989.The scanning model for translation:an update.[翻译扫描模型:新版]The Journal of Cell Biology[细胞生物学杂志]108:229-241)。因此,如果15RNA转录物的非编码前导子内的AUG密码子位于初级起始密码子的上游和框外,则消除由mRNA编码的多肽的蛋白质合成。为了利用该规则并应用于基因表达或功能的再激活策略,可以产生内源性ALS2-P165S等位基因以包含上游的框外翻译起始密码子(图5A-5B)。该等位基因含有Cas9PAM识别位点和ALS-CRX靶向间隔子。在本实例中位于PAM位点的5'的核

核苷酸3和4之间的Cas9切割可以促进可导致ATG密码子丢失的一个或多个核苷酸缺失或添加。通过由于NHEJ修复引起的缺失或添加的任何组合而丢失该上游框外ATG可以导致在赋予除草剂抗性的初级ALS2-P165S起始密码子处的翻译起始。

[0341] 预计使用上游框外ATG的再激活策略不限于图5A-5B中的设计。PAM和靶向间隔子还可以放置在相对于上游框外ATG的不同位置,只要由Cas9进行的靶向切割导致该起始密码子的丢失即可。例如,该PAM可以存在于反义链上,即起始密码子的5'。可以考虑其他设计;框外ATG起始密码子还可以置于5'前导序列内的不同位置。该PAM序列可以被像化脓性链球菌的其他Cas9蛋白识别,这些Cas9蛋白识别nGG PAM或非nGG PAM序列,例如嗜热链球菌CR1 (PAM序列识别nnAGAAAn) 和具有PAM序列的其他序列。其他Cas9蛋白质的用途将满足本实例以及上述实例3的再激活设计。

[0342] 用于基因激活的其他设计是预期的。如先前所述,除氯磺隆抗性外,赋予对其他除草剂有抗性的对ALS基因的修饰可以用于再激活。此外,磷酸甘露糖异构酶基因(PMI)、双丙氨酸抗性基因(BAR)、草丁膦乙酰转移酶(PAT)、潮霉素抗性基因(NPTII)、可选择标记基因、荧光标记基因(例如但不限于RFP(红色荧光蛋白)、CFP、GFP(绿色荧光蛋白))和草甘膦抗性基因可以被修饰以作为失活形式引入植物细胞中,并且用作如实例3所述的通过指导RNA引入进行再激活和通过NHEJ进行修复的靶标。还可以预计,具有多个失活基因可以作为靶标。例如,已经证明指导RNA复用在单个实验中同时修饰多个基因,因此氯磺隆和双丙氨酸抗性(但不限于这些基因)的靶向再激活可以另外被设计用于该方法。如上所述,与通过NHEJ恢复基因功能相一致,其他靶标的修饰将通过添加其他指导多核苷酸同时完成。

[0343] 另外,可以预计类似的方法可以应用于任何天然或引入的基因序列,并用作有效的基因转换机制。

[0344] 实例5

[0345] 在不引入多核苷酸修饰模板用于同源定向修复的情况下的特定基因编辑。

[0346] 实例2描述了使用特异性设计的多核苷酸修饰模板(修复模板)在ALS2基因编码区内的序列改变(P165S)。下面的实例描述了产生编辑的目的基因的不同方法,该方法不依赖于HDR机制而是依赖于NHEJ。

[0347] 如实例1所述,通过由Cas9核酸酶产生的DSB的NHEJ修复促进的两种最常见类型的突变是1bp插入和1bp缺失(表4)。基于本文所述的这些观察,开发了用于基因编辑的方法,该方法可以在两个连续的步骤中或在单个步骤中完成如下所述。

[0348] 第一步包括使用本文所述的RNA指导的Cas核酸酶系统靶向包含被Cas核酸内切酶识别的靶位点的目的基因或目的多核苷酸,导致由于NHEJ修复切割的DNA而产生具有特定核苷酸缺失的细胞或生物体(图6A中所示)。第二步需要重新靶向突变位点并选择插入所希望的核苷酸的事件(不使用多核苷酸修饰模板(修复模板)),因此特异性改变相应的氨基酸和基因功能。总的来说,这个想法如图6A-6C所示。该方法还可以用于编辑非编码DNA片段。

[0349] 可替代地,这两个步骤可以合并成单个步骤。可以使用两种不同的gRNA,一种识别原始靶位点,并且第二种识别同一位点但具有1bp缺失。在这种情况下,可以设想连续切割并修复内源性位点,这导致1bp缺失,然后切割改变的位点并对其进行具有1bp插入的修复。然后,可以选择插入所希望的核苷酸的事件。在编辑编码DNA序列的情况下,该方法实现了两个目标-恢复了翻译阅读框和替换了目的氨基酸。可以预期,不同核酸内切酶的组合可以

用于这个两步骤或一步骤系统中。例如,导致目的多核苷酸中单个碱基缺失的第一双链断裂的引入可以通过第一核酸内切酶来完成,而导致单个碱基插入的第二双链断裂的引入(以及目的多核苷酸的编辑)可以通过不同于第一核酸内切酶的第二核酸内切酶来完成。核酸内切酶之间的差异可以包括但不限于,不同的PAM识别序列、不同的靶识别序列、不同的切割活性(平端、5'或3'突出、单链、双链)、不同的DNA或氨基酸序列、起源于不同的生物体、或其任何一种组合。

[0350] 在目的基因组中编辑特定核苷酸的能力可以取决于所选择的核酸内切酶系统及其识别并切割特定靶位点的能力。新颖的引导的核酸内切酶的发现(参见例如2015年5月15日提交的美国专利申请62/162377)和/或具有各种PAM序列的引导的核酸内切酶的修饰将进一步增加可以被这些核酸内切酶识别和/或切割的靶位点的密度,最终导致使用本文所述的方法靶向基因组中任何给定核苷酸位置的能力。

#### [0351] 实例6

##### [0352] 具有稳定整合的Cas9核酸内切酶的玉蜀黍品系

[0353] 这个实例描述了具有稳定整合的Cas9表达盒的玉蜀黍品系的生成和验证。

[0354] 将含有在组成型(玉蜀黍UBI, SEQ ID NO:46)或温度调节型(玉蜀黍MDH, SEQ ID NO:47)启动子的转录控制下的,玉蜀黍密码子优化的Cas9的两个农杆菌载体(图7)引入Hi-II胚胎细胞中以建立含有Cas9核酸内切酶的预整合的基因组拷贝的品系。这些载体还含有调节蓝色荧光基因(AmCYAN)(作为可视标记)的表达的胚胎优选的END2启动子和由玉蜀黍组蛋白2B启动子转录调控的DsRED基因的中断拷贝。部分DsRED序列以正向复制(369bp片段),并由DsRED(RF-FP)基因的两个片段组成,这两个片段被可由gRNA靶向的347bp间隔子隔开。间隔子区域内的DSB促进了分子内重组恢复破坏的DsRed基因的功能,从而导致了红色荧光细胞。具有含有UBI:Cas9或MDH:Cas9的单拷贝T-DNA插入片段的玉蜀黍植物被用作未成熟胚胎的来源。将含有预整合的Cas9的发蓝色荧光的胚胎切除并在28°C(UBI:Cas9)或在37°C(MDH:Cas9)孵育24小时。轰击后,将具有MDH:Cas9的胚胎在37°C孵育24小时,并且然后转移至28°C。与对照(无gRNA)相比,用靶向347bp间隔子内的序列的两种DNA表达的gRNA轰击包含UBI:Cas9和MDH:Cas9的胚胎,容易地产生了发红色荧光的焦点。

[0355] 这些结果表明,上述玉蜀黍品系形成了单拷贝的功能性Cas9核酸内切酶。

#### [0356] 实例7

##### [0357] 在玉蜀黍中用预整合的Cas9瞬时递送gRNA到胚胎细胞中产生突变

[0358] 这个实例证明,将体外转录的RNA分子形式的gRNA递送到具有预整合的Cas9的玉蜀黍未成熟的胚胎细胞中,在靶位点处产生了突变。

[0359] 将在实例6中描述的含有UBI:Cas9或MDH:Cas9的玉蜀黍植物用作未成熟胚胎的来源用于gRNA的递送,该gRNA是作为体外转录的RNA或是作为对照的DNA表达盒。为了测量在LIG和MS26内源性靶位点处的突变频率,用实例6所述的温度处理将作为RNA分子(100ng/注射)或作为DNA载体(25ng/注射)的LIG-CR3和MS26-CR2gRNA递送到含有UBI:Cas9和MDH:Cas9的胚胎细胞中。在这些实验中,轰击后两天收获胚胎并通过扩增子深度测序进行分析。对于作为DNA载体和作为RNA分子递送的gRNA,检测到类似的频率,尤其是在由泛素启动子调控的Cas9的情况下(表10)。

[0360] 表10. 通过瞬时递送gRNA到具有在组成型(UBI)或调节型(MDH)启动子下的预整合

的Cas9的胚胎中而在玉蜀黍LIG和MS26靶位点处产生的突变体读数的百分比

靶位点	胚胎	转化	突变体读数百分比 (轰击后2天)
[0361] LIG	UBI:Cas9	gRNA (DNA)	1.22%
		gRNA (RNA)	1.86%
	MDH:Cas9 事件 1	gRNA (DNA)	0.25%
		gRNA (RNA)	0.12%
	MDH:Cas9 事件 2	gRNA (DNA)	0.57%
		gRNA (RNA)	0.26%
	MDH:Cas9 事件 3	gRNA (DNA)	0.46%
		gRNA (RNA)	0.35%
MS26	MDH:Cas9 事件 2	gRNA (DNA)	0.58%
		gRNA (RNA)	0.17%

[0362] 这些数据一起表明将处于RNA形式的gRNA直接递送到包含预整合的Cas9的玉蜀黍细胞中是对用于在植物细胞中产生突变的DNA递送的可行替代方案。

[0363] 实例8

[0364] 玉蜀黍未成熟胚胎的转化

[0365] 转化可以通过已知在植物中有效的各种方法来完成,包括颗粒介导的递送、农杆菌介导的转化、PEG介导的递送和电穿孔。

[0366] a. 颗粒介导的递送

[0367] 如下进行使用颗粒递送转化玉蜀黍未成熟胚胎。培养基配方如下。

[0368] 将穗剥皮并在30%Clorox漂白剂加上0.5%微量洗涤剂中表面消毒20分钟,并用无菌水冲洗两次。将未成熟胚胎分离,并以每个平板25个胚胎将胚胎轴侧向下(盾片侧向上)放置于560Y培养基上持续4小时,并且然后排列在2.5-cm靶区内准备进行轰击。可替代地,将分离的胚胎置于560L(起始培养基)上并在范围从26°C至37°C的温度下在黑暗中放置8至24小时,之后在26°C下放置于560Y中4小时,之后如上所述进行轰击。

[0369] 使用标准分子生物学技术构建含有双链断裂诱导剂和模板或供体DNA的质粒,并用含有发育基因ODP2(AP2结构域转录因子ODP2(胚珠发育蛋白2);US20090328252A1)和WUSCHEL(US2011/0167516)的质粒进行共同轰击。

[0370] 如下使用水溶性阳离子脂质TransIT-2020转染试剂(目录号MIR 5404,Mirus公司,美国),将质粒和目的DNA沉淀到0.6 $\mu$ m(平均直径)金颗粒上。使用总共1 $\mu$ g的DNA和/或RNA构建体(10次注射)在冰上制备DNA或DNA和RNA溶液。向预混的DNA中添加20 $\mu$ l制备的金颗粒(15mg/ml)和1 $\mu$ l TransIT-2020并小心混合。将金颗粒在微型离心机中以10,000rpm沉淀1分钟并去除上清液。添加105 $\mu$ l的100%EtOH,并通过简短的超声处理将颗粒进行重悬。然后,将10 $\mu$ l点在每个巨载体的中心上,并在轰击前允许其干燥。使用Biorad氮气枪(货架#3)以425PSI轰击样品板。

[0371] 轰击后,将胚胎在26°C至37°C的温度范围下在560P(维持培养基)上孵育12至48小时,并且然后置于26°C。在5至7天后,将胚胎转移至含有3mg/升双丙氨磷的560R选择培养基



上,并在26℃下每2周继代培养。在约10周的选择之后,将选择抗性愈伤组织克隆转移到288J培养基中以开始植物再生。在体细胞胚胎成熟(2-4周)后,将发育良好的体细胞胚胎转移到培养基上进行萌芽,并且转移到有光照的培养室中。在约7-10天后,将发育的小植物转移到试管中的272V不含激素的培养基中7-10天,直到小植物良好地生长。然后将植物转移到含有盆栽土壤的平托花盆(inserts in flats)(相当于2.5"盆)中,并在生长室中生长1周,随后在温室中再生长1-2周,然后转移到经典的600盆(1.6加仑)中并生长至成熟。对植物进行监测并对转化效率和/或再生能力的改变进行评分。

[0372] 起始培养基(560L)包含4.0g/l N6基础盐(西格玛公司(SIGMA)C-1416)、1.0ml/l Eriksson's维生素混合液(1000X西格玛公司(SIGMA)-1511)、0.5mg/l 硫胺素HCl、20.0g/l 蔗糖、1.0mg/l 2,4-D、以及2.88g/l L-脯氨酸(用D-I H2O定容,之后用KOH调节至pH 5.8);2.0g/l 结冷胶(在用D-I H2O定容之后添加);以及8.5mg/l 硝酸银(在对培养基进行灭菌并冷却至室温之后添加)。

[0373] 维持培养基(560P)包含4.0g/l N6基础盐(西格玛公司(SIGMA)C-1416)、1.0ml/l Eriksson's维生素混合液(1000X西格玛公司(SIGMA)-1511)、0.5mg/l 硫胺素HCl、30.0g/l 蔗糖、2.0mg/l 2,4-D、以及0.69g/l L-脯氨酸(用D-I H2O定容,之后用KOH调节至pH 5.8);3.0g/l 结冷胶(在用D-I H2O定容之后添加);以及0.85mg/l 硝酸银(在对培养基进行灭菌并冷却至室温之后添加)。

[0374] 轰击培养基(560Y)包含4.0g/l N6基础盐(西格玛公司(SIGMA)C-1416)、1.0ml/l Eriksson's维生素混合液(1000X西格玛公司(SIGMA)-1511)、0.5mg/l 硫胺素HCl、120.0g/l 蔗糖、1.0mg/l 2,4-D、以及2.88g/l L-脯氨酸(用D-I H2O定容,之后用KOH调节至pH 5.8);2.0g/l 结冷胶(在用D-I H2O定容之后添加);以及8.5mg/l 硝酸银(在对培养基进行灭菌并冷却至室温之后添加)。

[0375] 选择培养基(560R)包含4.0g/l N6基础盐(西格玛公司(SIGMA)C-1416)、1.0ml/l Eriksson's维生素混合液(1000X西格玛公司(SIGMA)-1511)、0.5mg/l 硫胺素HCl、30.0g/l 蔗糖、以及2.0mg/l 2,4-D(用D-I H2O定容,之后用KOH调节至pH 5.8);3.0g/l 结冷胶(在用D-I H2O定容之后添加);以及0.85mg/l 硝酸银和3.0mg/l 双丙氨膦(在对培养基进行灭菌并冷却至室温之后添加这两者)。

[0376] 植物再生培养基(288J)包含4.3g/l MS盐(GIBCO 11117-074)、5.0ml/l MS维生素储液(0.100g烟酸、0.02g/l 硫胺素HCL、0.10g/l 吡哆醇HCL、和0.40g/l 甘氨酸,用精制的D-I H2O定容)(Murashige和Skoog(1962)Physiol.Plant.[植物生理学]15:473)、100mg/l 肌醇、0.5mg/l 玉米素、60g/l 蔗糖、以及1.0ml/l的0.1mM脱落酸(用精制的D-I H2O定容,之后调节至pH 5.6);3.0g/l 结冷胶(在用D-I H2O定容之后添加);以及1.0mg/l 吲哚乙酸和3.0mg/l 双丙氨膦(在对培养基进行灭菌并冷却至60℃之后添加)。

[0377] 无激素培养基(272V)包含4.3g/l MS盐(GIBCO 11117-074)、5.0ml/l MS维生素储液(0.100g/l 烟酸、0.02g/l 硫胺素HCL、0.10g/l 吡哆醇HCL、和0.40g/l 甘氨酸,用精制的D-I H2O定容)、0.1g/l 肌醇、以及40.0g/l 蔗糖(用精制的D-I H2O定容,之后调节pH至5.6);以及6g/l 细菌用琼脂(在用精制的D-I H2O定容之后添加),灭菌并冷却至60℃。

[0378] b. 农杆菌介手的转化

[0379] 基本上如在Djukanovic等人(2006)Plant Biotech J[植物生物技术杂志]4:345-

57中所描述地进行农杆菌介导的转化。简言之,将10-12日龄的未成熟胚胎(尺寸为0.8-2.5mm)从灭菌的仁切下并放置于液体培养基(4.0g/L N6基础盐(西格玛公司(Sigma) C-1416)、1.0ml/L Eriksson's 维生素混合液(西格玛公司(Sigma) E-1511)、1.0mg/L 硫胺素HCl、1.5mg/L 2,4-D、0.690g/L L-脯氨酸、68.5g/L 蔗糖、36.0g/L 葡萄糖,pH 5.2)中。收集胚胎后,用1ml浓度为0.35-0.45 OD<sub>550</sub>的农杆菌代替培养基。将玉蜀黍胚胎与农杆菌在室温下一起孵育5分钟,然后将混合物倾倒在培养基平板上,该培养基平板包含4.0g/L N6基础盐(西格玛公司(Sigma) C-1416)、1.0ml/L Eriksson's 维生素混合液(西格玛公司(Sigma) E-1511)、1.0mg/L 硫胺素HCl、1.5mg/L 2,4-D、0.690g/L L-脯氨酸、30.0g/L 蔗糖、0.85mg/L 硝酸银、0.1nM 乙酰丁香酮、以及3.0g/L 结冷胶,pH 5.8。将胚胎轴侧向下在20℃在黑暗中孵育3天,然后在28℃在黑暗中孵育4天,然后转移到新的培养基平板上,该培养基平板包含4.0g/L N6基础盐(西格玛公司(Sigma) C-1416)、1.0ml/L Eriksson's 维生素混合液(西格玛公司(Sigma) E-1511)、1.0mg/L 硫胺素HCl、1.5mg/L 2,4-D、0.69g/L L-脯氨酸、30.0g/L 蔗糖、0.5g/L MES 缓冲液、0.85mg/L 硝酸银、3.0mg/L 双丙氨磷、100mg/L 羧苄青霉素、以及6.0g/L 琼脂,pH 5.8。将胚胎每三周进行继代培养,直到鉴定到转基因事件。通过将少量组织转移到再生培养基(4.3g/L MS盐(Gibco 11117)、5.0ml/L MS 维生素储液、100mg/L 肌醇、0.1μM ABA、1mg/L IAA、0.5mg/L 玉米素、60.0g/L 蔗糖、1.5mg/L 双丙氨磷、100mg/L 羧苄青霉素、3.0g/L 结冷胶,pH 5.6)上来诱导体细胞胚胎发生,并在28℃下在黑暗中孵育两周。将所有具有可视芽和根的物质都转移到以下培养基上,该培养基包含4.3g/L MS盐(Gibco 11117)、5.0ml/L MS 维生素储液、100mg/L 肌醇、40.0g/L 蔗糖、1.5g/L 结冷胶(pH 5.6),并在28℃下在人造光下孵育。一周后,将小植物移入含有相同培养基的玻璃管中并生长直到它们被取样和/或移植到土壤中。

#### [0380] 实例9

#### [0381] ZmODP-2和ZmWUS的瞬时表达增强了转化

[0382] 可以修改转化方案的参数以确保BBM活性是瞬时的。一种这样的方法涉及例如通过使用化学品PEI,以允许转录和表达、但是排除随后的DNA释放的方式沉淀含有BBM的质粒。

[0383] 在一个实例中,用PEI将BBM质粒沉淀到金颗粒上,同时使用标准的氯化钙方法将待整合的转基因表达盒(UBI:MoPAT-GFPm:PinII;MoPAT是玉蜀黍优化的PAT基因)沉淀到金颗粒上。

[0384] 简言之,如下用PEI包被金颗粒。首先,洗涤金颗粒。在微量离心管中称出35mg的平均直径为1.0的金颗粒(A.S.I.#162-0010),并添加1.2ml纯EtOH并涡旋一分钟。将该管在室温下孵育15分钟,并且然后在4℃下使用微型离心机高速离心15分钟。弃去上清液,并添加新鲜的1.2ml等分试样的乙醇(EtOH),涡旋一分钟,离心一分钟,并再次弃去上清液(重复两次)。添加新鲜的1.2ml等分试样的EtOH,并将该悬浮液(在EtOH中的金颗粒)在-20℃储存数周。为了用聚乙烯亚胺(PEI;西格玛公司(Sigma) #P3143)包被颗粒,将250μl的洗涤过的金颗粒/EtOH混合物进行离心并弃去EtOH。将颗粒在100μl ddH<sub>2</sub>O中洗涤一次以去除残留的乙醇,添加250μl的0.25mM PEI,然后进行脉冲超声处理以悬浮颗粒,并且然后将该管插入干冰/EtOH浴中以快速冷冻该悬浮液,然后将其冷冻干燥过夜。此时,干燥的包被的颗粒可以在-80℃储存至少3周。在使用之前,用250μl等分试样的2.5mM HEPES缓冲液(pH7.1)冲洗颗

粒3次,伴随1x脉冲超声处理,并且然后在每次离心之前快速涡旋。然后将颗粒悬浮在最终体积为250 $\mu$ l的HEPES缓冲液中。在附着DNA之前,将25 $\mu$ l等分试样的颗粒添加到新鲜的管中。为了附着未包被的DNA,将颗粒进行脉冲超声处理,然后添加1 $\mu$ g的DNA(在5 $\mu$ l水中),随后通过用Pipettman上下移液几次而混合并孵育10分钟。短暂(即10秒)旋转颗粒,弃去上清液,并添加60 $\mu$ l EtOH。将具有PEI沉淀的DNA-1的颗粒在60 $\mu$ l的EtOH中洗涤两次。将颗粒离心,弃去上清液,并将颗粒重悬于45 $\mu$ l水中。为了附着第二DNA(DNA-2),使用TransIT-2020进行沉淀。短暂超声处理45 $\mu$ l的颗粒/DNA-1悬浮液,并且然后添加5 $\mu$ l的100ng/ $\mu$ l的DNA-2和1 $\mu$ l的TransIT-2020。将溶液置于旋转式摇床上10分钟,以10,000g离心1分钟。去除上清液,并将颗粒重悬于60 $\mu$ l的EtOH中。将该溶液点在巨载体上,并使用针对PDS-1000的标准方案将其上依次附着有DNA-1和DNA-2的金颗粒递送到10个DAP Hi-II未成熟胚胎的盾片细胞中。对于该实验,该DNA-1质粒含有UBI:RFP:PinII表达盒,并且DNA-2含有UBI:CFP:PinII表达盒。轰击后两天,观察到CFP和RFP荧光标记两者的瞬时表达,因为未成熟胚胎表面上有许多红色和蓝色细胞。然后将胚胎置于非选择性培养基上并使其生长3周,然后对稳定的菌落进行评分。在这3周后,与仅有一个红色菌落相比,观察到10个多细胞的稳定表达的蓝色菌落。这表明,PEI沉淀可用于有效地引入DNA用于瞬时表达,同时显著地减少PEI引入的DNA的整合,并因此减少表达RFP的转基因事件的回收率。以此方式,PEI沉淀可以用于递送BBM和/或WUS2的瞬时表达。

[0385] 例如,首先使用PEI将颗粒用UBI:BBM:PinII包被,然后使用TransIT-2020将颗粒用UBI:MoPAT-YFP包被,并且然后轰击到未成熟胚胎的表面上的盾片细胞中。PEI介导的沉淀导致在未成熟胚胎表面上瞬时表达细胞的频率较高,并且稳定转化体的回收率极低(相对于TransIT-2020方法)。因此,预期PEI沉淀的BBM盒瞬时表达并刺激在组织的轰击表面(即盾片表面)上胚胎生长的爆发,但该质粒将不会整合。预期从Ca<sup>++</sup>/金颗粒释放的MoPAT-GFP质粒以导致转基因事件回收率大幅度提高的频率整合并表达可选择标记。作为对照处理,将含有UBI:GUS:PinII(而不是BBM)的PEI沉淀的颗粒与MoPAT-GFP/Ca<sup>++</sup>颗粒混合。将来自这两种处理的未成熟胚胎移至含有3mg/l双丙氨磷的培养基上。在6-8周后,预期在PEI/BBM处理中将以相对于对照处理(PEI/GUS)高得多的频率观察到GFP+、双丙氨磷抗性愈伤组织。

[0386] 作为一种替代方法,将BBM质粒用PEI沉淀到金颗粒上,并且然后引入未成熟胚胎表面上的盾片细胞中,并随后BBM基因的瞬时表达引起胚胎生长的快速增殖。在这个诱导生长期间,使用针对玉蜀黍的标准方法(参见实例1)用农杆菌处理外植体,其中T-DNA递送到引入转基因表达盒(如UBI:MoPAT-GFPm:PinII)的细胞中。在共培养后,允许外植体在正常培养基上回收,并且然后移至含有3mg/l双丙氨磷的培养基上。在6-8周后,预期在PEI/BBM处理中将以相对于对照处理(PEI/GUS)高得多的频率观察到GFP+、双丙氨磷抗性愈伤组织。

[0387] 通过瞬时表达BBM和/或WUS2多核苷酸产物来“启动(kick start)”愈伤组织生长可能是期望的。这可以通过递送BBM和WUS2 5' 加帽的聚腺苷酸化的RNA、含有BBM和WUS2DNA的表达盒、或BBM和/或WUS2蛋白来完成。所有这些分子都可以使用生物射弹基因枪来递送。例如,使用阿姆比恩公司(Ambion)的mMessage mMachine试剂盒可以容易地在体外制备5' 加帽的聚腺苷酸化的BBM和/或WUS2RNA。将RNA与含有目的多核苷酸和用于选择/筛选的标记的DNA(如UBI:MoPAT-GFPm:PinII)一起共同递送。预期接受该RNA的细胞将立即开始更加

快速地分裂,并且这些细胞中的大部分将整合上农艺学基因。这些事件可以进一步被验证为转基因克隆集落,因为它们也将表达该PAT~GFP融合蛋白(并因此将在适当照明下显示绿色荧光)。然后可以筛选从这些胚胎再生的植物中目的多核苷酸的存在。

#### [0388] 实例10

[0389] 在玉蜀黍中将gRNA和Cas9作为指导RNA/Cas核酸内切酶核糖核苷酸-蛋白质复合物(RGEN)直接递送到胚胎细胞中产生了突变

[0390] 这个实例表明,将处于蛋白质形式的Cas9和处于体外转录或化学合成的RNA分子形式的gRNA直接递送到玉蜀黍未成熟胚胎细胞中在相应的靶向位点处产生突变。

[0391] 为了产生处于RNA分子形式的gRNA,使用5'寡核苷酸引物通过PCR来扩增玉蜀黍优化的U6聚合酶III gRNA表达盒,所述5'寡核苷酸引物也含有就在基因的间隔子的5'处的T7聚合酶启动子和转录起始信号的序列。根据制造商的建议使用AmpliScribe T7-Flash试剂盒(Epicentre公司)进行T7体外转录,并使用NucAway Spin Columns(英杰公司(Invitrogen);生命技术公司(Life Technologies Inc))纯化产物,然后进行乙醇沉淀。

[0392] 为了产生指导RNA/Cas9核酸内切酶蛋白复合物(RGEN)(也称为指导RNA/Cas9核酸内切酶核糖核苷酸-蛋白质),将7 $\mu$ g的Cas9(化脓性链球菌Cas9)蛋白和3 $\mu$ g的gRNA分子(1:2摩尔比)在1 $\times$ Cas9缓冲液(NEB)中混合,总体积为20 $\mu$ l,并在室温下孵育15分钟。与RGEN一起,将含有泛素启动子调控的可选择的和可视的标记(MoPAT-DsRed融合物)、泛素启动子调节的玉蜀黍胚珠发育蛋白2(ZmODP2)(参见2009年12月31日公开的US20090328252)和玉蜀黍IN2启动子(Hershey等人1991,Plant Mol.Biol.[植物分子生物学117:679-690]调节的WUSCHEL、ZmWUS(参见2011年7月7日公开的US20110167516)的质粒与包含可商购的金颗粒(0.6 $\mu$ m,伯乐公司(Bio-Rad))和水溶性阳离子脂质TransIT-2020(Mirus公司,美国)的颗粒递送基质进行混合。使用颗粒介导的递送(参见实例8中描述的颗粒介导的递送)并进行一些修改,将包含指导RNA/Cas核酸内切酶核糖核苷酸-蛋白质复合物的颗粒递送基质递送到玉蜀黍胚胎细胞中。特别地,将金颗粒在微型离心机中以10,000rpm沉淀1分钟并去除上清液后,将颗粒重悬于105 $\mu$ l的无菌水中而不是100%乙醇中。然后,将10 $\mu$ l点在每个巨载体的中心上,并在轰击前允许其干燥。

[0393] 将野生型玉蜀黍植物用作未成熟胚胎的来源,用于将处于指导RNA/Cas核酸内切酶核糖核苷酸-蛋白质复合物(RGEN)形式的Cas9和gRNA与可选择的和可视的标记(UBI:MoPAT-DsRED)以及发育基因(UBI:ZmODP2和IN2:ZmWUS)一起共同递送。为了测量LIGCas-3、MS26Cas-2、MS45Cas-2和ALSCas-4内源性靶位点处的突变频率,在轰击后两天收获胚胎并通过扩增子深度测序进行分析。将未处理的胚胎和仅用Cas9蛋白轰击的胚胎用作阴性对照,而将用表达Cas9和gRNA的DNA载体轰击的胚胎用作阳性对照。对于作为DNA载体和作为指导RNA/Cas核酸内切酶核糖核苷酸-蛋白质复合物递送的Cas9-gRNA组分,检测到类似的频率(表11)。

[0394] 表11. 通过将RGEN复合物直接递送到玉蜀黍胚胎细胞中在LIG、MS26、MS45、和ALS靶位点处产生的突变体读数的百分比。

[0395]

靶位点	递送的分子	读数的总数	突变体读数的总数	突变体读数的百分比
LIGCas-3 (染色体 2)	未处理的胚胎 (对照 1)	915,198	38	0.004%
	仅 Cas9 蛋白 (对照 2)	408,348	17	0.004%
	Cas9-Lig-CR3 RGEN	439,827	2,510	0.57%
	Cas9 (DNA) +Lig-CR3 (DNA)	369,443	2,058	0.56%
MS26Cas-2 (染色体 1)	未处理的胚胎 (对照 1)	245,476	8	0.003%
	仅 Cas9 蛋白 (对照 2)	429,388	20	0.004%
	Cas9-MS26-CR2 RGEN	252,519	533	0.21%
	Cas9 (DNA) +MS26-CR2 (DNA)	186,857	812	0.43%
MS45Cas-2(染色体 9)	未处理的胚胎 (对照 1)	255,877	12	0.005%
	仅 Cas9 蛋白 (对照 2)	487,876	12	0.002%
	Cas9-MS45-CR2 RGEN	241,287	2,075	0.86%
	Cas9 (DNA) +MS45-CR2 (DNA)	304,622	1591	0.52%
ALS2Cas-4(染色体 5)	仅 Cas9 蛋白 (对照 2)	807,014	125	0.02%
	Cas9-ALS-CR4 RGEN	791,084	3,613	0.45%
	Cas9 (DNA) +ALS-CR4 (DNA)	833,130	4251	0.51%

[0396] 为了测量植物水平的突变频率,将用Cas9-MS45-gRNA复合物、ZmODP2、ZmWUS和MOPAT-DSRED共同轰击的60个胚胎置于含有双丙氨磷作为选择试剂的培养基上。从36个除草剂抗性愈伤组织部分中的每一个再生出多个植物并筛选突变。在36个事件中,17个(47%)含有突变体等位基因(10个单和7个双等位基因),而19个(53%)仅显示野生型MS45等位基因。在具有突变的植物中,每个等位基因的测序读数的数目相似,这表明植物不是嵌合体。

[0397] 为了证明直接的RGEN递送也足以在植物中的内源性基因中产生特定编辑,如实例2所述靶向玉蜀黍ALS2基因(ALS2特异性ALSCas-4靶位点)。以与上述类似的方式,将127nt单链DNA01igo2(SEQ ID NO:45)作为修复模板与Cas9/ALS-CR4RGEN复合物共同递送。轰击后两天收获胚胎并通过扩增子深度测序进行分析(表12)。

[0398] 表12. 通过将RGEN复合物和供体DNA模板递送到玉蜀黍胚胎细胞中在ALS靶位点处产生的的突变体读数和具有编辑的读数的百分比。

[0399]

靶位点	递送的分子	读数的总数	突变体读数的总数	突变体读数%	具有编辑的读数的数目	具有编辑的读数%
ALS2	仅 Cas9 蛋白	807,014	105	0.01%	-	-

[0400]

Cas9-ALS-CR4 RGEN + ss Oligo2	791,084	3,613	0.45%	209	0.02%
-------------------------------------	---------	-------	-------	-----	-------

[0401] 另外,在两个独立实验中,将40至50个轰击的胚胎转移到含有100ppm氯磺隆(作为编辑的ALS2基因的直接选择)的平板上。六周后,通过测序来分析在具有氯磺隆的培养基上继续生长的两个愈伤组织部分(每个实验一个)。在这两个事件中,一个ALS2等位基因被特异性编辑,而第二个等位基因保持野生型。从这些愈伤组织部分再生的植物含有编辑的ALS2等位基因,并且在用除草剂喷洒时对氯磺隆具有抗性。

[0402] 这些数据表明,将处于指导RNA/Cas核酸内切酶核糖核苷酸-蛋白质复合物形式的Cas9和gRNA(在使用或不使用多核苷酸修饰模板DNA的情况下)直接递送到玉蜀黍未成熟胚胎细胞中是用于植物中的靶向诱变和基因编辑的DNA递送(例如重组DNA、质粒DNA)的可行替代方案。

[0403] 实例11

[0404] 在玉蜀黍中将处于mRNA和gRNA形式的Cas9直接递送到胚胎细胞中产生突变

[0405] 这个实例表明,将处于mRNA分子形式的Cas9和处于体外转录或化学合成的RNA分子形式的gRNA直接递送到玉蜀黍未成熟胚胎细胞中在相应的靶向位点处产生突变。

[0406] 在我们先前的实验(Svitashev等人,Plant Physiology[植物生理学],2015,第169卷,第931-945页)中,处于体外合成的RNA分子形式的gRNA与表达Cas9的DNA载体的共同递送产生了比在Cas9和gRNA两者都作为DNA载体递送的实验中大约低100倍的突变频率。这种差异的一种可能的解释可以是当gRNA作为RNA递送并且Cas9作为DNA载体递送时,不满足对Cas9和gRNA的一致功能的需求。为了克服这个问题,Cas9可以作为mRNA分子递送,这将缩短从递送的时刻到功能性Cas9蛋白表达的时间。在该实验中使用可商购的Cas9mRNA(曲林克生物技术公司(TriLink Biotechnologies))。

[0407] 为了测试这个想法,每次注射用Cas9mRNA(200ng)、处于体外合成的RNA分子形式的gRNA(100ng)、含有泛素调节的MoPAT-DsRED融合物的DNA载体(25ng)以及发育基因(泛素启动子调节的ODP2和IN2启动子调节的WUS(每个12ng))共同轰击玉蜀黍胚胎细胞。在该实验中使用可商购的Cas9mRNA(曲林克生物技术公司(TriLink Biotechnologies))和如上所述体外合成的RNA分子。通过扩增子深度测序对转化后2天收集的胚胎进行突变频率分析(表13)。

[0408] 表13.通过将作为mRNA的Cas9和作为RNA分子的gRNA瞬时递送到玉蜀黍胚胎细胞中在MS45靶位点处产生的的突变体读数的百分比。

[0409]

靶位点	递送的分子	读数的总数	突变体读数的总数	突变体读数百分比
MS45	仅 Cas9 mRNA	1,097,279	799	0.07%
	Cas9 mRNA+Ms45 CR2	1,260,332	2,304	0.18%

[0410]

gRNA				
Cas9 (DNA) +Ms45 CR2 (DNA)		1,106,125	3,490	0.31%

[0411] 这些数据表明,处于RNA分子形式的Cas9和gRNA两者的递送提高了靶向突变的频率,并且与作为RGEN复合物的Cas9-gRNA递送一起,是用于植物中靶向诱变和基因编辑的DNA递送的可行替代方案。

[0412] 实例12

[0413] 在玉蜀黍中在不使用可选择标记的情况下将Cas9和gRNA作为指导RNA/Cas核酸内切酶核糖核苷酸-蛋白质复合物 (RGEN) 直接递送到胚胎细胞中产生突变

[0414] 这个实例表明,将处于蛋白质形式的Cas9和处于体外转录或化学合成的RNA分子形式的gRNA递送到玉蜀黍未成熟胚胎细胞中(没有共同递送一个或多个可选择标记基因)足以在相应靶位点处以实用频率再生具有突变的植物。

[0415] 为转化或修饰的细胞提供生长优势的可选择标记的必要性一直是植物转化和基因组修饰方案中的长期范例。因此,在所有突变、基因编辑和基因整合实验中,可选择标记都用于选择基因组编辑事件。考虑到实例10中描述的实验中RGEN复合物的出乎意料的高活性(突变频率),尝试了不使用可选择标记的情况下的完全无DNA(无载体)的基因组编辑。用靶向三种不同基因liguleless1 (LIG)、MS26和MS45的指导RNA/Cas核酸内切酶核糖核苷酸-蛋白质 (RGEN) 复合物轰击玉蜀黍胚胎细胞。在作为对照的平行实验中递送在DNA载体上的Cas9核酸内切酶和MS45-gRNA。再生植物并通过测序来分析这些植物的靶向突变。在所有实验中,以2.4%至9.7%范围的出人意料的高频率回收突变体植物(表14)。

[0416] 表14. 在将处于DNA载体形式的Cas9和gRNA递送到玉蜀黍未成熟胚胎细胞中和将RGEN复合物直接递送到玉蜀黍未成熟胚胎细胞中时,在LIG、MS26和MS45靶位点处的突变频率。对未经选择(不使用可选择标记的情况下)再生的T0植物进行分析。

[0417]

靶位点	Cas9 和 gRNA 递送方法	分析的植物	具有突变的等位基因的植物	SEQ ID NO:
LIGCas-3	RGEN	756	73 (9.7%)	12
MS26Cas-2	RGEN	756	18 (2.4%)	14
MS45 Cas-2	RGEN	1,880	70 (3.7%)	8
MS45 Cas-2	载体 DNA	940	38 (4.0%)	8

[0418] 此外,将再生的T0植物与野生型Hi-II植物杂交,并将后代植物用于分离分析。在所分析的所有后代植物中证实了在预期的孟德尔分离(1:1)下的突变的ms45等位基因的有性传播。

[0419] 为了评估RGEN递送降低玉蜀黍中脱靶切割的可能,使用DNA载体和RGEN递送来评估MS45脱位点处的突变频率。通过使用Bowtie序列比对仪(允许多达两个与中靶序列的错配)(Langmead,B.、Trapnell,C.、Pop,M.和Salzberg,S.L.Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.[短DNA序列与人类基因组的超快速和存储高效比对]Genome Biol.[基因组生物学]10:R25,2009)将

MS45Cas-2靶位点的前间区(与指导RNA间隔子碱基配对的靶位点的区域)与玉蜀黍B73参考基因组(B73RefGen\_v3,玉蜀黍遗传学和基因组学数据库)进行比对来进行与中靶位点具有密切同源性的位点的搜索。然后检查可能的脱靶位点中紧邻所鉴定的前间区脱靶的3'处的NGG前间区邻接基序(PAM)序列的存在。使用这些搜索标准只鉴定了一个单一的脱靶位点(5'-CGCCGAGGGCGACTACCGGC-3',SEQ ID NO:81)。它含有与MS45前间区靶标和AGG PAM的2bp错配(表15)。为了证实该位点在体内被切割,通过深度测序分析了在用表达Cas9和MS45-CR2gRNA的DNA载体转化的玉蜀黍胚胎中突变的存在。如表15所示,与针对中靶位点所观察到的4%频率相比,脱靶位点的突变活性为2%的频率。如表15所示,相对于在DNA载体上的Cas9和gRNA递送,RCEN脱靶活性大大降低(从2%至0%)。

[0420] 表15. 在将处于DNA载体和RCEN复合物形式的Cas9和gRNA递送到玉蜀黍未成熟胚胎细胞中时在MS45脱靶位点处的突变频率。对不经选择再生的T0植物进行分析。

[0421]

靶位点	Cas9 和 gRNA 递送方法	具有 PAM*的靶位点序列	分析的植物	具有突变的植物 (数目)	具有突变的植物 (%)	SEQ ID NO:
MS45Cas-2		GGCCGAGGTCGACTACCG GCCGG	904	38	4%	14
MS45 脱位点	DNA	<u>CGCCGAGGG</u> CGACTACCG GCAGG	940	19	2.0%	82
MS45 脱位点	RCEN	<u>CGCCGAGGG</u> CGACTACCG GCAGG	1,880	0	0.0%	82

[0422] \*PAM-前间区邻近基序是紧邻靶位点3'处的3nt序列。

[0423] 将MS45脱靶位点与预期位点相比不同的两个核苷酸以粗体和下划线显示。

[0424] 这个实例表明,使用RNA指导的核酸内切酶产生具有靶向突变的植物,不需要可选择的或可筛选的标记基因的共同递送,因此增加了引入的修饰的特异性和精确性。再生的植物仅包含靶向突变或靶向基因编辑(如果包括修复模板以修饰DNA序列),而没有DNA载体的随机整合。该递送方法提供了在主要作物物种(包括但不限于玉蜀黍、大豆、小麦、稻、粟、高粱和低芥酸菜籽)的植物细胞中进行基因诱变的完全无DNA的方法。



## 序列表

<110>	先锋良种国际有限公司 Cigan, Andrew Mark Svitashev, Sergei	
<120>	用于无标记基因组修饰的方法和组合物	
<130>	7158 PCT	
<150>	62/243719	
<151>	2015-10-20	
<150>	62/309033	
<151>	2016-03-16	
<150>	62/359254	
<151>	2016-07-07	
<160>	82	
<170>	PatentIn版本3.5	
<210>	1	
<211>	4107	
<212>	DNA	
<213>	化脓性链球菌M1 GAS(SF370)	
<400>	1	
	atggataaga aataactcaat aggettagat atcggcacia atagcgtcgg atgggcggtg	60
	atcactgatg aatataaggt tccgtctaaa aagttcaagg ttctgggaaa tacagaccgc	120
	cacagtatca aaaaaaatct tataggggct cttttatttg acagtggaga gacagcggaa	180
	gcgactcgtc tcaaacggac agctcgtaga aggtatacac gtcggaagaa tcgtatttgt	240
	tatctacagg agattttttc aatgagatg gcgaaagtag atgatagttt ctttcatcga	300
	cttgaagagt cttttttggt ggaagaagac aagaagcatg aacgtcatcc tatttttggg	360
	aatatagtag atgaagttgc ttatcatgag aaatatccaa ctatctatca tctgcgaaaa	420
	aaattggtag attctactga taaagcggat ttgcgcttaa tctatttggc cttagcgc	480
	atgattaagt ttcgtgggtca ttttttgatt gagggagatt taaatcctga taatagtgat	540
	gtggacaaac tatttatcca gttggtacaa acctacaatc aattatttga agaaaacct	600
	attaacgcaa gtggagtaga tgctaaagcg attctttctg cacgattgag taaatcaaga	660
	cgattagaaa atctcattgc tcagctcccc ggtgagaaga aaaatggctt atttgggaat	720
	ctcattgctt tgcattggg tttgaccct aattttaaat caaattttga tttggcagaa	780
	gatgctaaat tacagcttcc aaaagatact tacgatgatg atttagataa tttattggcg	840
	caaattggag atcaatatgc tgatttgttt ttggcagcta agaatttate agatgctatt	900
	ttactttcag atatcctaag agtaaatact gaaataacta aggetcccc atcagcttca	960
	atgattaaac gctacgatga acatcatcaa gacttgacte ttttaaagc tttagttcga	1020
	caacaacttc cagaaaagta taaagaaatc ttttttgatc aatcaaaaaa cggatatgca	1080
	ggttatattg atggggggagc tagccaagaa gaattttata aatttatcaa accaatttta	1140

gaaaaaatgg	atggtactga	ggaattattg	gtgaaactaa	atcgtgaaga	tttgctgcgc	1200
aagcaacgga	cctttgacaa	cggtcttatt	ccccatcaaa	ttcacttggg	tgagctgcat	1260
gctattttga	gaagacaaga	agacttttat	ccatttttaa	aagacaatcg	tgagaagatt	1320
gaaaaaatct	tgacttttcg	aattccttat	tatgttggtc	cattggcgcg	tgccaatagt	1380
cgttttgcat	ggatgactcg	gaagtctgaa	gaaacaatta	ccccatggaa	ttttgaagaa	1440
gttgctgata	aaggtgcttc	agctcaatca	tttattgaac	gcatgacaaa	ctttgataaa	1500
aatcttccaa	atgaaaaagt	actaccaaaa	catagtttgc	tttatgagta	ttttacggtt	1560
tataacgaat	tgacaaaagg	caaatatgtt	actgaaggaa	tcgaaaacc	agcatttctt	1620
tcaggngaac	agaagaaaagc	cattgttgat	ttactcttca	aaacaaatcg	aaaagtaacc	1680
gttaagcaat	taaaagaaga	ttatttcaaa	aaaatagaat	gttttgatag	tgttgaaatt	1740
tcaggagtgtg	aagatagatt	taatgcttca	ttaggtaacct	accatgattt	gctaaaaatt	1800
attaagata	aagatttttt	ggataatgaa	gaaaatgaag	atatcttaga	ggatattggt	1860
ttaacattga	ccttatttga	agatagggag	atgattgagg	aaagacttaa	aacatatgct	1920
cacctctttg	atgataaggt	gatgaaacag	cttaaactgc	gccgttatac	tggttgggga	1980
cgtttgtctc	gaaaattgat	taatggtatt	agggataagc	aatctggcaa	aacaatatta	2040
gattttttga	aatcagatgg	ttttgccaat	cgcaatttta	tgcagctgat	ccatgatgat	2100
agtttgacat	ttaaagaaga	cattcaaaaa	gcacaagtgt	ctggacaagg	cgatagttta	2160
catgaacata	ttgcaaat	agctggtagc	cctgctatta	aaaaaggtat	tttacagact	2220
gtaaaagtgtg	ttgatgaatt	ggtcaaagta	atggggcggc	ataagccaga	aaatatcggt	2280
attgaaatgg	cacgtgaaaa	tcagacaact	caaaagggcc	agaaaaatc	gcgagagcgt	2340
atgaaacgaa	tcgaagaagg	tatcaaagaa	ttaggaagtc	agattcttaa	agagcatcct	2400
gttgaaaata	ctcaattgca	aaatgaaaag	ctctatctct	attatctcca	aaatggaaga	2460
gacatgtatg	tggaccaaga	attagatatt	aatcgtttaa	gtgattatga	tgctgatcac	2520
attgttccac	aaagtttctt	taaagacgat	tcaatagaca	ataaggtctt	aacgcgttct	2580
gataaaaatc	gtggtaaact	ggataacggt	ccaagtgaag	aagtagtcaa	aaagatgaaa	2640
aactattgga	gacaacttct	aaacgccaag	ttaatcactc	aacgtaagtt	tgataattta	2700
acgaaagctg	aacgtggagg	tttgagtga	cttgataaag	ctggttttat	caaacgcca	2760
ttggttgaaa	ctcgccaaat	cactaagcat	gtggcacaaa	ttttggatag	tcgcatgaat	2820
actaaatacg	atgaaaatga	taaacttatt	cgagaggtta	aagtgattac	cttaaaatct	2880
aaattagttt	ctgacttccg	aaaagatttc	caattctata	aagtacgtga	gattaacaat	2940
taccatcatg	cccatgatgc	gtatctaaat	gccgtcgttg	gaactgcttt	gattaagaaa	3000
tatccaaaac	ttgaatcgga	gtttgtctat	ggtgattata	aagtttatga	tgttcgtaaa	3060
atgattgcta	agtctgagca	agaaaatagc	aaagcaaccg	caaaatattt	cttttactct	3120
aatatcatga	acttcttcaa	aacagaaatt	acaattgcaa	atggagagat	tcgcaaaccg	3180
cctctaatacg	aaactaatgg	ggaaaactgga	gaaattgtct	gggataaagg	gcgagatttt	3240
gccacagtgc	gcaaagtatt	gtccatgccc	caagtcata	ttgtcaagaa	aacagaagta	3300
cagacaggcg	gattctccaa	ggagtcaatt	ttaccaaaaa	gaaattcgga	caagcttatt	3360
gctcgtaaaa	aagactggga	tccaaaaaaa	tatggtggtt	ttgatagtcc	aacggtagct	3420
tattcagtcc	tagtgggtgc	taaggtggaa	aaagggaaat	cgaagaagtt	aaaatccggt	3480



&lt;220&gt;

&lt;223&gt; 玉蜀黍优化的Cas9表达盒

&lt;400&gt; 5

gtgcagcgtg acccggtcgt gcccctctct agagataatg agcattgcat gtctaagtta	60
taaaaaatta ccacataattt tttttgtcac acttgtttga agtgcagttt atctatcttt	120
atacatatat ttaaacttta ctctacgaat aatataatct atagtactac aataatatca	180
gtgtttttaga gaatcatata aatgaacagt tagacatggt ctaaaggaca attgagtatt	240
ttgacaacag gactctacag ttttatcttt ttagtgtgca tgtgttctcc ttttttttg	300
caaatagctt cacctatata atacttcate cattttatta gtacatccat ttagggttta	360
gggttaatgg tttttataga ctaatTTTT tagtacatct attttattct attttagcct	420
ctaaattaag aaaactaaaa ctctatTTTta gTTTTTTat ttaataattt agatataaaa	480
tagaataaaa taaagtgact aaaaattaaa caaatacctt ttaagaaatt aaaaaacta	540
aggaaacatt tttcttgttt cgagtagata atgccagcct gttaaacgcc gtcgacgagt	600
ctaacggaca ccaaccagcg aaccagcagc gtcgctcggg gccaagcgaa gcagacggca	660
cggcatctct gtcgctgect ctggaccctt ctcgagagtt ccgctccacc gttggacttg	720
ctccgctgtc ggcatccaga aattgcgtgg cggagcggca gacgtgagcc ggcacggcag	780
gcggcctcct cctcctctca cggcaccggc agctacgggg gattccttcc ccaccgctcc	840
ttcgctttcc ctctctgcc cgccgtaata aatagacacc cctccacac cctctttccc	900
caacctcgtg ttgttcggag cgcacacaca cacaaccaga tctccccaa atccaccctg	960
cggcacctcc gcttcaaggt acgccgctcg tctcccccc cccccctctc taccttctct	1020
agatcggcgt tccgggccat gcatggtag ggccccgtag ttctacttct gttcatgttt	1080
gtgttagatc cgtgtttgtg ttagatccgt gctgctagcg ttcgtacacg gatgcgacct	1140
gtacgtcaga cacgttctga ttgctaactt gccagtgttt ctctttgggg aatcctggga	1200
tggtcttagc cgttccgcag acgggatcga tttcatgatt tttttgttt cgttgcatag	1260
ggtttggttt gcccttttcc tttatttcaa tatatgccgt gcacttgttt gtcgggcat	1320
cttttcatgc tttttttgt cttggttgtg atgatgtggt ctggttgggc ggtcgttcta	1380
gatcggagta gaattctggt tcaactacc tgggtgattt attaatTTTt gatctgtatg	1440
tgtgtgccat acatattcat agttacgaat tgaagatgat ggatggaaat atcgatctag	1500
gataggtata catgttgatg cgggttttac tgatgcatat acagagatgc tttttgttcg	1560
cttggttgtg atgatgtggt gtggttgggc ggtcgttcat tcgttctaga tcggagtaga	1620
atactgtttc aaactacctg gtgtatttat taattttgga actgtatgtg tgtgtcatac	1680
atcttcatag ttacgagttt aagatggatg gaaatatcga tctaggatag gtatacatgt	1740
tgatgtgggt tttactgatg catatacatg atggcatatg cagcatctat tcatatgctc	1800
taaccttgag tacctatcta ttataataaa caagtatggt ttataattat tttgatcttg	1860
atatacttgg atgatggcat atgcagcagc tatatgtgga ttttttttagc cctgccttca	1920
tacgtatatt atttgettgg tactgtttct tttgtcgatg ctcacctgtg tgtttggtgt	1980
tacttctgca ggtcgactct agaggatcca tggcaccgaa gaagaagcgc aaggtgatgg	2040
acaagaagta cagcatcggc ctcgacatcg gcaccaacte ggtgggctgg gccgtcatca	2100
cggacgaata taaggtcccg tcgaagaagt tcaaggtcct cggcaataca gaccgccaca	2160

gcatcaagaa	aaacttgatc	ggcgcctcc	tgttgatag	cgcgagacc	gcgaggcga	2220
ccaggctcaa	gaggaccgcc	aggagacggt	acactaggcg	caagaacagg	atctgctacc	2280
tgcaggagat	cttcagcaac	gagatggcga	aggtggacga	ctccttcttc	caccgcctgg	2340
aggaatcatt	cctggtggag	gaggacaaga	agcatgagcg	gcaccaatc	ttcggcaaca	2400
tcgtcgacga	ggtaagtffc	tgcttctacc	tttgatata	atataataat	tatcattaat	2460
tagtagtaat	ataatatttc	aaatattttt	ttcaaaataa	aagaatgtag	tatatagcaa	2520
ttgcttttct	gtagtttata	agtgtgtata	ttttaattta	taacttttct	aatatatgac	2580
caaaacatgg	tgatgtgcag	gtggcctacc	acgagaagta	cccgacaatc	taccacctcc	2640
ggaagaaact	ggtggacagc	acagacaagg	cggacctccg	gctcatctac	cttgccctcg	2700
cgcatatgat	caagttccgc	ggcacttcc	tcategaggg	cgacctgaac	ccggacaact	2760
ccgacgtgga	caagctgttc	atccagctcg	tgcagacgta	caatcaactg	ttcgaggaga	2820
acccataaaa	cgctagcggc	gtggacgcca	aggccatctc	ctcggccagg	ctctcgaaat	2880
caagaaggct	ggagaacctt	atcgcgcagt	tgccaggcga	aaagaagaac	ggcctcttcg	2940
gcaaccttat	tgcgctcagc	ctcggcctga	cgccgaactt	caaatcaaac	ttcgacctcg	3000
cggaggacgc	caagctccag	ctctcaaagg	acacctacga	cgacgacctc	gacaacctcc	3060
tggcccagat	aggagaccag	tacgcggacc	tcttctctgc	cgccaagaac	ctctccgacg	3120
ctatcctgct	cagcgacatc	cttcgggtca	acaccgaaat	taccaaggca	ccgctgtccg	3180
ccagcatgat	taaacgctac	gacgagcacc	atcaggacct	cacgctgctc	aaggcactcg	3240
tccgccagca	gctccccgag	aagtacaagg	agatcttctt	cgaccaatca	aaaaacggct	3300
acgcgggata	tatcgacggc	ggtgccagcc	aggaagagtt	ctacaagttc	atcaaaccaa	3360
tcttgagaaa	gatggacggc	accgaggagt	tgctggtcaa	gctcaacagg	gaggacctcc	3420
tcaggaagca	gaggaccttc	gacaacggct	ccatcccgca	tcagatccac	ctgggcgaac	3480
tgcattccat	cctgcggcgc	caggaggact	tctaccegtt	cctgaaggat	aaccgggaga	3540
agatcgagaa	gatcttgacg	ttccgcatcc	catactacgt	gggcccgctg	gctcgcggca	3600
actcccgggt	cgcttgatg	accggaagt	cggaggagac	catcacacc	tggaactttg	3660
aggaggtggt	cgataagggc	gctagcgtc	agagcttcat	cgagcgcgatg	accaacttcg	3720
ataaaaacct	gccaatgaa	aaagtctctc	ccaagcactc	gctgctctac	gagtacttca	3780
ccgtgtacaa	cgagctcacc	aaggtcaaat	acgtcaccga	gggcatgcgg	aagccggcgt	3840
tcctgagcgg	cgagcagaag	aaggcgatag	tggacctctc	cttcaagacc	aacaggaagg	3900
tgaccgtgaa	gcaattaaaa	gaggactact	tcaagaaaat	agagtgtctc	gactccgtgg	3960
agatctcggg	cgtggaggat	cggttcaacg	cctcactcgg	cacgtateac	gacctctca	4020
agatcattaa	agacaaggac	ttcctcgaca	acgaggagaa	cgaggacatc	ctcgaggaca	4080
tcgtcctcac	cctgaccctg	ttcgaggacc	gcgaaatgat	cgaggagagg	ctgaagacct	4140
acgcgcacct	gttcgacgac	aaggtcatga	aacagctcaa	gaggcgcgc	tacactggtt	4200
ggggaaggct	gtcccgaag	ctcattaatg	gcatcaggga	caagcagagc	ggcaagacca	4260
tcttggaact	cctcaagtc	gacgggttcg	ccaaccgeaa	cttcatgcag	ctcattcacg	4320
acgactcgct	cacgttcaag	gaagacatcc	agaaggcaca	ggtgagcggg	cagggtgact	4380
ccctccacga	acacatgcc	aacctggccg	gctcgccggc	cattaanaag	ggcatcctgc	4440
agacggtcaa	ggtcgtcgac	gagctcgtga	aggtgatggg	ccggcacaag	cccgaanaata	4500

tcgtcataga	gatggccagg	gagaaccaga	ccacccaaaa	agggcagaag	aactcgcgcg	4560
agcggatgaa	acggatcgag	gagggcatta	aagagctcgg	gtcccagatc	ctgaaggagc	4620
accccgtgga	aaatacccag	ctccagaatg	aaaagctcta	cctctactac	ctgcagaacg	4680
gccgcgacat	gtacgtggac	caggagctgg	acattaatcg	gctatcggac	tacgacgtcg	4740
accacatcgt	gccgcagtcg	ttcctcaagg	acgatagcat	cgacaacaag	gtgctcaccc	4800
ggtcggataa	aaatcggggc	aagagcgaca	acgtgccccg	cgaggaggtc	gtgaagaaga	4860
tgaaaaacta	ctggcgccag	ctcctcaacg	cgaaactgat	caccagcgc	aagtctgaca	4920
acctgacgaa	ggcggaacgc	ggtggcttga	gcgaactcga	taaggcgggc	ttcataaaaa	4980
ggcagctggg	cgagacgcgc	cagatcacga	agcatgtcgc	ccagatcctg	gacagccgca	5040
tgaatactaa	gtacgatgaa	aacgacaagc	tgatccggga	ggtgaagggtg	atcacgctga	5100
agtccaagct	cgtgtcggac	ttccgcaagg	acttccagtt	ctacaaggtc	cgcgagatca	5160
acaactacca	ccacgcccac	gacgcctacc	tgaatgcggg	ggtcggggacc	gccctgatca	5220
agaagtacc	gaagctggag	tcggagtctg	tgtacggcga	ctacaaggtc	tacgacgtgc	5280
gcaaaatgat	cgccaagtcc	gagcaggaga	tcggcaaggc	cacggcaaaa	tactttctct	5340
actcgaacat	catgaacttc	ttcaagaccg	agateaccct	cgcgaacggc	gagatccgca	5400
agcggcccgt	catcgaaacc	aacggcgaga	cgggcgagat	cgtctgggat	aaggccggg	5460
atctcgcgac	ggtccgcaag	gtgctctcca	tgccgcaagt	caatatcgtg	aaaagacgg	5520
aggtccagac	gggcccgggtc	agcaaggagt	ccatcctccc	gaagcgaac	tccgacaagc	5580
tcatcgcgag	gaagaaggat	tgggaccgga	aaaaatatgg	cggttcgac	agccccaccg	5640
tcgcatacag	cgctctcgtc	gtggcgaagg	tggagaaggg	caagtcaaag	aagctcaagt	5700
ccgtgaagga	gctgctcggg	atcacgatta	tggagcggtc	ctccttcgag	aagaaccgga	5760
tcgacttctt	agaggccaag	ggatataagg	aggtcaagaa	ggacctgatt	attaaactgc	5820
cgaagtactc	gctcttcgag	ctggaaaacg	gccgcaagag	gatgctcgcc	tccgcaggcg	5880
agttgcagaa	gggcaacgag	ctcgccctcc	cgagcaaata	cgtcaatttc	ctgtacctcg	5940
ctagccacta	tgaaaagctc	aagggcagcc	cggaggacaa	cgagcagaag	cagctcttcg	6000
tggagcagca	caagcattac	ctggacgaga	tcatcgagca	gatcagcgag	ttctcgaagc	6060
gggtgatcct	cgccgacgcg	aacctggaca	aggtgctgtc	ggcatataac	aagcaccgcg	6120
acaaaccaat	acgcgagcag	gccgaaaata	tcateccact	cttaccctc	accaacctcg	6180
gcgctccggc	agccttcaag	tacttcgaca	ccacgattga	ccggaagcgg	tacacgagca	6240
cgaaggagggt	gctcgatcgc	acgctgatec	accagagcat	cacagggtc	tatgaaacac	6300
gcatcgacct	gagccagctg	ggcggagaca	agagaccacg	ggaccgccac	gatggcgagc	6360
tgggaggccg	caagcgggca	aggtaggtac	cgttaaccta	gacttgtcca	tcttctggat	6420
tggccaactt	aattaatgta	tgaaataaaa	ggatgcacac	atagtgacat	gctaatact	6480
ataatgtggg	catcaaagtt	gtgtgttatg	tgtaattact	agttatctga	ataaaagaga	6540
aagagatcat	ccatatttct	tatcetaaat	gaatgtcacg	tgtctttata	attctttgat	6600
gaaccagatg	catttcatta	accaaatacca	tatacatata	aatattaate	atatataatt	6660
aatatcaatt	gggttagcaa	aacaaatcta	gtctaggtgt	gttttgcgaa	tgcggcc	6717

<210> 6

<211> 4437

<212>	DNA					
<213>	人工序列					
<220>						
<223>	Lig-CR3指导RNA表达载体					
<400>	6					
aaattgtaag	cgttaatatt	ttgttaaaat	tcgcgttaaa	tttttgtaa	atcagctcat	60
tttttaacca	atagccgaa	atcggcaaaa	tccctataa	atcaaaaga	tagaccgaga	120
tagggttgag	tgttgttcca	gtttgaaca	agagtccact	attaaagaac	gtggactcca	180
acgtcaaagg	gcgaaaaacc	gtctatcagg	gcgatggccc	actacgtgaa	ccatcacct	240
aatcaagttt	tttggggtcg	aggtgccgta	aagcaactaa	tcggaacct	aaaggagcc	300
cccgatntag	agcttgacgg	ggaaagccgg	cgaacgtggc	gagaaaggaa	gggaagaaag	360
cgaaaggagc	gggcgctagg	gcgctggcaa	gtgtagcggg	cacgctgcgc	gtaaccacca	420
cacccgccgc	gcttaatgcg	ccgctacagg	gcgctgccc	ttcgccattc	aggctgcgca	480
actgttggga	agggcgatcg	gtgcgggcct	cttcgctatt	acgccagctg	gcgaaagggg	540
gatgtgctgc	aaggcgatta	agttgggtaa	cgccagggtt	ttcccagtea	cgacgttgta	600
aaacgacggc	cagtgaattg	taatacgaact	cactataggg	cgaattgggt	accgggcccc	660
ccctcgaggt	cgacggatc	gataagcttt	gagagtacaa	tgatgaacct	agattaatca	720
atgccaaagt	ctgaaaaatg	caccctcagt	ctatgatcca	gaaaatcaag	attgcttgag	780
gccctgttcg	gttgttccgg	attagagccc	cggattaatt	cctagccgga	ttacttctct	840
aatttatata	gattttgatg	agctggaatg	aatcctggct	tattccggtg	caaccgaaca	900
ggccctgaag	gataccagta	atcgctgagc	taaattggca	tgctgtcaga	gtgtcagtat	960
tgcagcaagg	tagtgagata	accggcatca	tggtgccagt	ttgatggcac	cattagggtt	1020
agagatgggtg	gccatgggcg	catgtcctgg	ccaactttgt	atgatatatg	gcagggtgaa	1080
taggaaagta	aaattgtatt	gtaaaaaggg	atttcttctg	tttgtagcgc	catgtacaag	1140
gaatgcaagt	tttgagcgag	ggggcatcaa	agatctggct	gtgtttccag	ctgtttttgt	1200
tagccccatc	gaatccttga	cataatgatc	ccgcttaaat	aagcaacctc	gcttgatag	1260
ttccttgtgc	tctaacacac	gatgatgata	agtcgtaaaa	tagtgggtgc	caaagaattt	1320
ccaggcccag	ttgtaaaaagc	taaaatgcta	ttcgaatttc	tactagcagt	aagtcgtgtt	1380
tagaaattat	ttttttatat	accttttttc	cttctatgta	cagtaggaca	cagtgtcagc	1440
gccgcgttga	cggagaatat	ttgcaaaaaa	gtaaaagaga	aagtcatagc	ggcgtatgtg	1500
ccaaaaactt	cgtcacagag	agggccataa	gaaacatggc	ccacggccca	atacgaagca	1560
ccgcgacgaa	gccccaaacag	cagtccgtag	gtggagcaaa	gcgctgggta	atacgaaac	1620
gttttgtccc	accttgacta	atcacaagag	tggagcgtac	cttataaac	gagccgcaag	1680
caccgaattg	cgtagcgcgta	cgtgtggttt	tagagctaga	aatagcaagt	taaaataagg	1740
ctagtccgtt	atcaacttga	aaaagtggca	ccgagtcggg	gctttttttt	tgcggccgcg	1800
aattcctgca	gggcccctctt	gtcggaccag	ttgcccacca	cgttgggtgag	ctcgggtgagg	1860
cccttcattg	agaggaagga	ggtcatgagg	tgctaccga	tgtgggactt	ggggccgttc	1920
ttgatggcga	agatggagta	gggggcgttc	ttcttgaggg	ccttgttgta	ggacctcacg	1980
aggttgtcct	tgaggagctg	gtactcctgc	ttgttgaggg	aggagttgcc	ggtcctgttc	2040

accctcttga	gcacgggctc	tgagttcctg	aggaactcgt	cgaggtacac	gagggggtcg	2100
atcctgccgc	gagcggagaa	gaagtagatg	tgctggaca	cggaggtctt	ggtctcggtc	2160
acgaggcact	ggatgatcac	gccgaggtac	ttgttctgac	tagttctaga	gcggccgcca	2220
ccgcggtgga	gctccagctt	ttgttccctt	tagtgagggt	taatttcgag	cttggcgtaa	2280
tcatggatcat	agctgtttcc	tgtgtgaaat	tgttatccgc	tcacaattcc	acacaacata	2340
cgagccggaa	gcataaagtg	taaagcctgg	ggtgcctaata	gagtgagcta	actcacatta	2400
attgcgttgc	gctcactgcc	cgctttccag	tcgggaaacc	tgctcgtcca	gctgcattaa	2460
tgaatcggcc	aacgcgcggg	gagagcggtt	ttgcgtattg	ggcgcctctc	cgcttccctc	2520
ctcactgact	cgctgcgctc	ggtcgttcgg	ctgcggcgag	cggatcagc	tactcaaag	2580
gcggtaaatac	ggttatccac	agaatcaggg	gataacgcag	gaaagaacat	gtgagcaaaa	2640
ggccagcaaaa	agccagggaa	ccgtaaaaag	gccgcgttgc	tggcgttttt	ccataggctc	2700
cgccccctg	acgagcatca	caaaaatcga	cgctcaagtc	agaggtggcg	aaaccgcaca	2760
ggactataaaa	gataaccagc	gtttccccct	ggaagctccc	tcgtgcgctc	tctgttccg	2820
accctgccgc	ttaccggata	cctgtccgcc	tttctccctt	cgggaagcgt	ggcgtttctt	2880
catagctcac	gctgtaggta	tctcagttcg	gtgtaggtcg	ttcgtccca	gctgggctgt	2940
gtgcacgaac	ccccgttca	gcccagccgc	tgcccttat	ccgtaacta	tcgtcttgag	3000
tccaaccggg	taagacacga	cttatcgcca	ctggcagcag	ccactggtaa	caggattagc	3060
agagcgagg	atgtaggcgg	tgctacagag	ttcttgaagt	ggtggcctaa	ctacggctac	3120
actagaagga	cagtatttgg	tatctgcgct	ctgctgaagc	cagttacctt	cggaaaaaga	3180
gttggtagct	cttgatccgg	caaacaaacc	accgctggta	gcggtggttt	ttttgtttgc	3240
aagcagcaga	ttacgcgcag	aaaaaaagga	tctcaagaag	atcctttgat	cttttctacg	3300
gggtctgacg	ctcagtggaa	cgaaaactca	cgtaagggga	ttttggatcat	gagattatca	3360
aaaaggatct	tcacctagat	ccttttaaat	taaaaatgaa	gttttaaatc	aatctaaagt	3420
atatatgagt	aaacttggtc	tgacagttac	caatgcttaa	tcagtgaggc	acctatctca	3480
gcgatctgtc	tatttcgttc	atccatagtt	gcctgactcc	ccgtcgtgta	gataactacg	3540
atacgggagg	gcttaccatc	tgccccagct	gctgcaatga	taccgcgaga	cccacgctca	3600
ccggtccag	atttatcagc	aataaaccag	ccagccggaa	gggccgagcg	cagaagtgg	3660
cctgcaactt	tatccgcctc	catccagctt	attaattggt	gccgggaagc	tagagtaagt	3720
agttcggcag	ttaatagttt	gcgcaacggt	ggtgccattg	ctacaggcat	cgtgggtgca	3780
cgctcgtcgt	ttggatggc	ttcattcagc	tcgggttccc	aacgatcaag	gcgagttaca	3840
tgatccccc	tgttgtgcaa	aaaagcgggt	agctccttcg	gtcctccgat	cgttgtcaga	3900
agtaagtgg	ccgcagtgtt	atcaactcatg	ggtatggcag	caactgataa	ttctcttact	3960
gtcatgccat	ccgtaagatg	cttttctgtg	actggtgagt	actcaaccaa	gtcattctga	4020
gaatagtgt	tgccggcacc	gagttgetct	tgccccgcgt	caatacggga	taataccgcg	4080
ccacatagca	gaactttaa	agtgetcatc	attggaaaac	gttcttcggg	gcgaaaactc	4140
tcaaggatct	taccgctgtt	gagatccagt	tcgatgtaac	ccactcgtgc	acccaactga	4200
tcttcagcat	cttttacttt	caccagcgtt	tctgggtgag	caaaaacagg	aaggcaaaat	4260
gccgcaaaaa	agggaaataag	ggcgacacgg	aaatggtgaa	tactcatact	cttctttttt	4320
caatattatt	gaagcattta	tcagggttat	tgtctcatga	gcggatacat	atttgaatgt	4380



atttagaaaa ataaacaaat aggggttccg cgcacatttc cccgaaaagt gccacct	4437
<210> 7	
<211> 27	
<212> DNA	
<213> 玉蜀黍	
<400> 7	
gtactccatc cgccccatcg agtaggg	27
<210> 8	
<211> 24	
<212> DNA	
<213> 玉蜀黍	
<400> 8	
gcacgtacgt caccatcccg ccgg	24
<210> 9	
<211> 24	
<212> DNA	
<213> 玉蜀黍	
<400> 9	
gacgtacgtg ccctactcga tggg	24
<210> 10	
<211> 24	
<212> DNA	
<213> 玉蜀黍	
<400> 10	
gtaccgtacg tgccccggcg gagg	24
<210> 11	
<211> 24	
<212> DNA	
<213> 玉蜀黍	
<400> 11	
ggaattgtac cgtacgtgcc ccgg	24
<210> 12	
<211> 20	
<212> DNA	
<213> 玉蜀黍	
<400> 12	
gcgtacgcgt acgtgtgagg	20
<210> 13	
<211> 22	

<212> DNA	
<213> 玉蜀黍	
<400> 13	
gctggccgag gtcgactacc gg	22
<210> 14	
<211> 23	
<212> DNA	
<213> 玉蜀黍	
<400> 14	
ggccgaggtc gactaccggc cgg	23
<210> 15	
<211> 23	
<212> DNA	
<213> 玉蜀黍	
<400> 15	
ggcgcgagct cgtgcttcac cgg	23
<210> 16	
<211> 21	
<212> DNA	
<213> 玉蜀黍	
<400> 16	
ggtgccaatc atgcgtcgcg g	21
<210> 17	
<211> 20	
<212> DNA	
<213> 玉蜀黍	
<400> 17	
ggtcgccatc acgggacagg	20
<210> 18	
<211> 24	
<212> DNA	
<213> 玉蜀黍	
<400> 18	
gtcgcggcac ctgtcccgtg atgg	24
<210> 19	
<211> 56	
<212> DNA	
<213> 人工序列	
<220>	

<223> MS26Cas-1正向引物	
<400> 19	
ctacactctt tcctacacg acgtctctcc gatctaggac cggaagctcg ccgcgt	56
<210> 20	
<211> 54	
<212> DNA	
<213> 人工序列	
<220>	
<223> MS26Cas-1和MS26Cas-3反向引物	
<400> 20	
caagcagaag acggcatacg agctcttccg atcttctctgg aggacgacgt gctg	54
<210> 21	
<211> 59	
<212> DNA	
<213> 人工序列	
<220>	
<223> MS26Cas-2正向引物	
<400> 21	
ctacactctt tcctacacg acgtctctcc gatctaaggt cctggaggac gacgtgctg	59
<210> 22	
<211> 51	
<212> DNA	
<213> 人工序列	
<220>	
<223> MS26Cas-2反向引物	
<400> 22	
caagcagaag acggcatacg agctcttccg atctccggaa gctcgccgcg t	51
<210> 23	
<211> 56	
<212> DNA	
<213> 人工序列	
<220>	
<223> MS26Cas-3正向引物	
<400> 23	
ctacactctt tcctacacg acgtctctcc gatcttctc cggaagctcg ccgcgt	56
<210> 24	
<211> 63	
<212> DNA	
<213> 人工序列	

<220>		
<223>	LIGCas-1正向引物	
<400>	24	
	ctacactctt tccctacacg acgtctttcc gatctaggac tgtaacgatt tacgcacctg	60
	ctg	63
<210>	25	
<211>	58	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	LIGCas-1和LIGCas-2反向引物	
<400>	25	
	caagcagaag acggcatacg agcttctccg atctgcaaat gagtagcagc gcacgtat	58
<210>	26	
<211>	63	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	LIGCas-2正向引物	
<400>	26	
	ctacactctt tccctacacg acgtctttcc gatcttcttc tgtaacgatt tacgcacctg	60
	ctg	63
<210>	27	
<211>	60	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	LIGCas-3正向引物	
<400>	27	
	ctacactctt tccctacacg acgtctttcc gatctaaggc gcaaatgagt agcagcgcac	60
<210>	28	
<211>	57	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	LIGCas-3反向引物	
<400>	28	
	caagcagaag acggcatacg agcttctccg atctcacctg ctgggaattg taccgta	57
<210>	29	

<211>	58	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	MS45Cas-1正向引物	
<400>	29	
	ctacactctt tccctacacg acgetcttcc gatctaggag gaccggttcg gcctcagt	58
<210>	30	
<211>	54	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	S45Cas-1、MS45Cas-2和MS45Cas-3反向引物	
<400>	30	
	caagcagaag acggcatacg agetcttccg atctgccgce tggcattgtc tctg	54
<210>	31	
<211>	58	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	MS45Cas-2正向引物	
<400>	31	
	ctacactctt tccctacacg acgetcttcc gatcttcttg gaccggttcg gcctcagt	58
<210>	32	
<211>	58	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	MS45Cas-3正向引物	
<400>	32	
	ctacactctt tccctacacg acgetcttcc gatctgaagg gaccggttcg gcctcagt	58
<210>	33	
<211>	58	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	ALSCas-1正向引物	
<400>	33	
	ctacactctt tccctacacg acgetcttcc gatctaagge gacgatgggc gtctcttg	58

<210>	34	
<211>	53	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	ALSCas-1、ALSCas-2和ALSCas-3反向引物	
<400>	34	
	caagcagaag acggcatacg agctcttccg atctgegtct gcatgccac etc	53
<210>	35	
<211>	58	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	ALSCas-2正向引物	
<400>	35	
	ctacactctt tccctacacg acgctcttcc gatctttccc gacgatgggc gtctcctg	58
<210>	36	
<211>	58	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	ALSCas-3正向引物	
<400>	36	
	ctacactctt tccctacacg acgctcttcc gatctggaac gacgatgggc gtctcctg	58
<210>	37	
<211>	43	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	用于次级PCR的正向引物	
<400>	37	
	aatgatacgg cgaccaccga gatctacaact ctttccctac acg	43
<210>	38	
<211>	18	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	用于次级PCR的反向引物；	
<400>	38	

caagcagaag acggcata	18
<210> 39	
<211> 1910	
<212> DNA	
<213> 玉蜀黍	
<220>	
<221> 尚未归类的特征	
<222> (1) .. (1910)	
<223> ALS1-DNA序列	
<400> 39	
atggccaccg ccgccaccgc ggccgccgcg ctaccggcg ccaactaccgc tacgcccag	60
tcgaggcgcc gagcccacca cttggccacc cggecgcgccc tcgcccgccc catcaggtgc	120
tcagcgttgt cacgcgccac gccgacgget cccccggcca ctccgctacg tccgtggggc	180
cccaacgagc cccgcaaggg ctccgacatc ctctcgagg ctctcgagcg ctgtggcgtc	240
cgtgacgtct tcgectacc cggecgcgca tccatggaga tccaccagge actaccgcgc	300
tccccgtca tcgccaacca cctcttccgc cacgaacaag gggaggcctt cgccgcctcc	360
ggctacgcgc gctcctcggg ccgcgttggc gtctgcatcg ccacctccgg ccccgcgccc	420
accaacctag tctctgcgct cgcagacgcg ttgctcgact ccgtccccat tgtegccatc	480
acgggacagg tgccgcgacg catgattggc accgacgcct ttcaggagac gcccatcgtc	540
gaggtcacc gctccatcac caagcacaac tacctggtec tcgacgtega cgacatcccc	600
cgcgctcgtgc aggaggcctt cttcctcgca tcctctggte gcccggggcc ggtgcttggt	660
gacatcccca aggacatcca gcagcagatg gcggtgccgg cctgggacac gcccatgagt	720
ctgcctgggt acatcgcgcg ccttcccaag cctcccgcga ctgaatttct tgagcaggtg	780
ctgcgtcttg ttggtgaatc acggcgcctt gttctttatg ttggcgggtg ctgtgcagca	840
tcaggtgagg agttgtgccg ctttgtggag ttgactggaa tcccagtcac aactactctt	900
atgggccttg gcaacttccc cagcgacgac ccaactgtcac tgcgcatgct tggatgcat	960
ggcacagtgt atgcaaatta tgcagtggat aaggccgatc tgttgcttgc atttggtgtg	1020
cggtttgatg atcgtgtgac agggaaaatt gaggttttg caggcagagc taagattgtg	1080
cacattgata ttgatcctgc tgagattggc aagaacaagc agccacatgt gtccatctgt	1140
gcagatgtta agcttgcttt gcagggcatg aatactcttc tggaaggaag cacatcaaag	1200
aagagctttg acttcggctc atggcatgat gaattggatc agcaaaagcg ggagtttccc	1260
cttgggtata aaatcttcaa tgaggaaatc cageccacaat atgctattca ggttcttgat	1320
gagttgacga aggggaagge catcattgcc acaggtgttg ggcagacca gatgtgggcg	1380
gcacagtatt acaattacaa gcggccaagg cagtggctgt cttcagctgg tcttggggct	1440
atgggatttg gtttgccggc tgetgctggt getgctgtgg ccaaccagc tgteactgtt	1500
gttgacatcg acggagatgg tagcttctc atgaacatc aggagctagc tatgatccgt	1560
attgagaacc tcccagtcaa ggtctttgtg ctaaacaacc agcacctcgg gatggtggtg	1620
cagtgggagg acaggttcta taaggccaat agagcacaca cattcttggg aaaccagag	1680
aacgaaagtg agatatatcc agattttgtg gcaattgcc aagggttcaa cattccagca	1740

gtccgtgtga caaagaagag cgaagtccat gcagcaatca agaagatgct tgaggctcca	1800
gggccgtacc tcttgatat aatcgccccg caccaggagc atgtgttgcc tatgatccct	1860
agtgggtgggg ctttcaagga tatgatcctg gatggtgatg gcaggactgt	1910
<210> 40	
<211> 1910	
<212> DNA	
<213> 玉蜀黍	
<220>	
<221> 尚未归类的特征	
<222> (1) .. (1910)	
<223> ALS2-DNA序列	
<400> 40	
atggccaccg ccgccgccgc gtctaccgcg ctcaactggcg ccactaccgc tgcgcccag	60
gcgaggcgcc gggcgcacct cctggccacc cgcccgcccc tcgcccgccc catcaggtgc	120
tcagcggcgt cacccgccat gccgatgget cccccggcca ccccgtccg gccgtggggc	180
cccaccgatc cccgcaaggg cgccgacatc ctctctgagt ccctcgagcg ctgcggcgtc	240
cgcgacgtct tcgectacc cgccggcgcg tccatggaga tccaccaggc actcaccgc	300
tccccgtca tcgccaacca cctcttccgc cacgagcaag gggaggcctt tgcggcctcc	360
ggctacgcgc gctcctcggg ccgcgtcggc gtctgcatcg ccacctccgg ccccggcgcc	420
accaaccttg tctccgcgct cgccgacgcg ctgctcgatt ccgtccccat ggtcgccatc	480
acgggacagg tgcccgacg catgattggc accgacgct tccaggagac gcccatcgtc	540
gaggtcacc gctccatcac caagcacaac tacctggtec tcgacgtega cgacatcccc	600
cgcgtcgtgc aggaggcttt ctctctcgcc tctctggte gaccggggcc ggtgcttgtc	660
gacatcccca aggacatcca gcagcagatg gcggtgcctg tctgggacaa gcccatgagt	720
ctgcctgggt acattgcgcg ccttcccaag ccccctgcga ctgagttgct tgagcaggtg	780
ctgcgtcttg ttggtgaatc ccggcgccct gttctttatg ttggcgggtg ctgcgcagca	840
tctggtgagg agttgcgacg ctttgtggag ctgactggaa tcccggtcac aactactctt	900
atgggcctcg gcaacttccc cagcgacgac ccactgtctc tgcgcatgct aggtatgcat	960
ggcacggtgt atgcaaatta tgcagtggat aaggccgac tgttgcttgc acttggtgtg	1020
cggtttgatg atcgtgtgac agggaagatt gagcttttg caagcagggc taagattgtg	1080
cacgttgata ttgatccgce tgagattggc aagaacaage agccacatgt gtccatctgt	1140
gcagatgtta agcttgcttt gcagggcatg aatgctcttc ttgaaggaag cacatcaaag	1200
aagagctttg actttggctc atggaacgat gagttggatc agcagaagag ggaattcccc	1260
cttgggtata aaacatctaa tgaggagate cageccacaat atgctattca ggttcttgat	1320
gagctgacga aaggcgagge catcatcggc acaggtgttg ggcagcacca gatgtgggcg	1380
gcacagtact acacttacia gcggccaagg cagtggttgt cttcagctgg tcttggggct	1440
atgggatttg gtttgccgce tgetgctggt gcttctgtgg ccaaccagg tgttactgtt	1500
gttgacatcg atggagatgg tagctttctc atgaacgttc aggagctagc tatgatccga	1560
attgagaacc tcccgtgaa ggtctttgtg ctaaacaacc agcacctggg gatggtggtg	1620





195	200	205
Leu Ala Ser Ser Gly Arg Pro Gly Pro Val Leu Val Asp Ile Pro Lys		
210	215	220
Asp Ile Gln Gln Gln Met Ala Val Pro Val Trp Asp Lys Pro Met Ser		
225	230	235
Leu Pro Gly Tyr Ile Ala Arg Leu Pro Lys Pro Pro Ala Thr Glu Leu		
245	250	255
Leu Glu Gln Val Leu Arg Leu Val Gly Glu Ser Arg Arg Pro Val Leu		
260	265	270
Tyr Val Gly Gly Gly Cys Ala Ala Ser Gly Glu Glu Leu Arg Arg Phe		
275	280	285
Val Glu Leu Thr Gly Ile Pro Val Thr Thr Thr Leu Met Gly Leu Gly		
290	295	300
Asn Phe Pro Ser Asp Asp Pro Leu Ser Leu Arg Met Leu Gly Met His		
305	310	315
Gly Thr Val Tyr Ala Asn Tyr Ala Val Asp Lys Ala Asp Leu Leu Leu		
325	330	335
Ala Leu Gly Val Arg Phe Asp Asp Arg Val Thr Gly Lys Ile Glu Ala		
340	345	350
Phe Ala Ser Arg Ala Lys Ile Val His Val Asp Ile Asp Pro Ala Glu		
355	360	365
Ile Gly Lys Asn Lys Gln Pro His Val Ser Ile Cys Ala Asp Val Lys		
370	375	380
Leu Ala Leu Gln Gly Met Asn Ala Leu Leu Glu Gly Ser Thr Ser Lys		
385	390	395
Lys Ser Phe Asp Phe Gly Ser Trp Asn Asp Glu Leu Asp Gln Gln Lys		
405	410	415
Arg Glu Phe Pro Leu Gly Tyr Lys Thr Ser Asn Glu Glu Ile Gln Pro		
420	425	430
Gln Tyr Ala Ile Gln Val Leu Asp Glu Leu Thr Lys Gly Glu Ala Ile		
435	440	445
Ile Gly Thr Gly Val Gly Gln His Gln Met Trp Ala Ala Gln Tyr Tyr		
450	455	460
Thr Tyr Lys Arg Pro Arg Gln Trp Leu Ser Ser Ala Gly Leu Gly Ala		
465	470	475
Met Gly Phe Gly Leu Pro Ala Ala Ala Gly Ala Ser Val Ala Asn Pro		
485	490	495
Gly Val Thr Val Val Asp Ile Asp Gly Asp Gly Ser Phe Leu Met Asn		
500	505	510

Val Gln Glu Leu Ala Met Ile Arg Ile Glu Asn Leu Pro Val Lys Val  
 515 520 525  
 Phe Val Leu Asn Asn Gln His Leu Gly Met Val Val Gln Trp Glu Asp  
 530 535 540  
 Arg Phe Tyr Lys Ala Asn Arg Ala His Thr Tyr Leu Gly Asn Pro Glu  
 545 550 555 560  
 Asn Glu Ser Glu Ile Tyr Pro Asp Phe Val Thr Ile Ala Lys Gly Phe  
 565 570 575  
 Asn Ile Pro Ala Val Arg Val Thr Lys Lys Asn Glu Val Arg Ala Ala  
 580 585 590  
 Ile Lys Lys Met Leu Glu Thr Pro Gly Pro Tyr Leu Leu Asp Ile Ile  
 595 600 605  
 Val Pro His Gln Glu His Val Leu Pro Met Ile Pro Ser Gly Gly Ala  
 610 615 620  
 Phe Lys Asp Met Ile Leu Asp Gly Asp Gly Arg Thr Val Tyr  
 625 630 635

<210> 42

<211> 23

<212> DNA

<213> 玉蜀黍

<400> 42

gctgctcgat tccgtcccca tgg

23

<210> 43

<211> 794

<212> DNA

<213> 人工序列

<220>

<223> 794 bp多核苷酸修饰模板

<400> 43

ttctgctcaa gcaactcagt cgcagggggc ttgggaaggc gcgcaatgta cccaggcaga 60  
 ctcatgggct tgtcccagac aggcaccgcc atctgctgct ggatgtcctt ggggatgtcg 120  
 acaagcaccg gccctggtcg accagaggag gcgaggaaga aagcctctg cacgacgegg 180  
 gggatgtcgt cgacgtcgag gaccaggtag ttgtgcttgg tgatggagcg ggtgacctcg 240  
 acgatgggcg tctcctggaa ggcgtcggtg ccaatcatgc gtcgcgacac ttggccggta 300  
 atcgccacca tggggacgga atcgagcagc gcgtcggcga gcgaggagac aaggttggtg 360  
 gcgccggggc cggagggtgc gatgcagacg ccgacgcggc ccgaggagcg cgcgtagccc 420  
 gaggccgcaa aggcctcccc ttgctcgtgg cggaagaggt ggttgccgat gacgggggag 480  
 cgggtgagtg cctgggtggat ctccatggac gcgccgccgg ggtaggcgaa gacgtcgcgg 540  
 acgccgcagc gctcgaggga ctcgacgagg atgtcggcgc ccttgccggg atcgggtggg 600

ccccacggcc ggagcggggt ggccggggga gccatcggca tggcgggtga cgccgctgag	660
cacctgatgg gcgcggcgag ggcgcggcgg gtggccagga ggtgcgcccg gcgcctcgcc	720
ttgggcgcag cggtagtggc gccagtgagc gcggtagacg cggcggcggc gttggccatg	780
gttgcggcgg ctgt	794
<210> 44	
<211> 127	
<212> DNA	
<213> 人工序列	
<220>	
<223> 127 bp多核苷酸修饰模板oligo-1	
<400> 44	
aaccttgctt ccgcgctcgc cgacgcgctt ctgactcgc tccccattgt cgccatcacg	60
ggacaggtgt cgcgacgcat gattggcacc gacgccttcc aggagacgcc catcgctcag	120
gtcaccc	127
<210> 45	
<211> 127	
<212> DNA	
<213> 人工序列	
<220>	
<223> 127 bp多核苷酸修饰模板oligo-2	
<400> 45	
aaccttgctt ccgcgctcgc cgacgcgctt ctggactcgc tgccgatggt cgccatcacg	60
ggacaggtgt cccgacgcat gattggcacc gacgccttcc aggagacgcc catcgctcag	120
gtcaccc	127
<210> 46	
<211> 56493	
<212> DNA	
<213> 人工序列	
<220>	
<223> 含有玉蜀黍密码子优化的Cas9和玉蜀黍UBI启动子的农杆菌载体	
<400> 46	
gtcggatcac cggaaaggac ccgtaaagtg ataatgatta tcactacat atcacaacgt	60
gcgtggaggc catcaaacca cgtcaaataa tcaattatga cgcaggtatc gtattaattg	120
atctgcatca acttaacgta aaaacaactt cagacaatac aatcagcga cactgaatac	180
ggggcaacct catgtcccc ccccccccc cctgcagge atcgttgtgt cacgctcgtc	240
gtttggtatg gcttcattca getccggtt ccaacgatec aggcgagtta catgateccc	300
catgttggtc aaaaaagcgg ttagctcett cggctctccg atcgttggtc gaagtaagtt	360
ggccgcagtg ttatcactca tggttatggc agcactgcat aattctctta ctgtcatgcc	420
atccgtaaga tgcttttctg tgactggtga gtactcaacc aagtcattct gagaatagtg	480

tatgcggcga	ccgagttgct	cttgccccgc	gtcaacacgg	gataataccg	cgccacatag	540
cagaacttta	aaagtgtca	tcattggaaa	acgtttctcg	gggcgaaaac	tctcaaggat	600
cttaccgctg	ttgagatcca	gttcgatgta	accactctgt	gcaccaact	gatcttcagc	660
atcttttact	ttcaccagcg	tttctgggtg	agcaaaaaca	ggaaggcaa	atgccgcaa	720
aaaggaata	agggcgacac	ggaaatgtt	aatactcata	ctcttccttt	ttcaatatta	780
ttgaagcatt	tatcagggtt	attgtctcat	gagcggatac	atatttgaat	gtatttagaa	840
aaataaaca	ataggggttc	cgcgcacatt	tccccgaaa	gtgccacctg	acgtctaaga	900
aaccattatt	atcatgacat	taacctataa	aaatagcgct	atcacgaggc	cctttcgtct	960
tcaagaattc	ggagcttttg	ccattctcac	cggattcagt	cgteactcat	ggtgatttct	1020
cacttgataa	ccttattttt	gacgagggga	aattaatagg	ttgtattgat	gttggacgag	1080
tcggaatcgc	agaccgatac	caggatcttg	ccatectatg	gaactgcctc	ggtgagtttt	1140
ctccttcatt	acagaaacgg	ctttttcaaa	aatatggtat	tgataatcct	gatatgaata	1200
aattgcagtt	tcatttgatg	ctcgatgagt	ttttctaate	agaattgggt	aattggttgt	1260
aacactggca	gagcattacg	ctgacttgac	gggacggcgg	ctttgttgaa	taaategaac	1320
ttttgctgag	ttgaaggatc	agatcacgea	tcttcccgc	aacgcagacc	gttccgtggc	1380
aaagcaaaag	ttcaaaatca	ccaactggtc	cacctacaac	aaagctctca	tcaaccgtgg	1440
ctccctcaact	ttctggctgg	atgatggggc	gattcaggcc	tggtatgagt	cagcaacacc	1500
ttcttcacga	ggcagacctc	agcgccagaa	ggccgcccaga	gaggccgagc	gcggccgtga	1560
ggcttgagac	ctagggcagg	gcatgaaaaa	gcccgtagcg	ggctgctacg	ggcgtctgac	1620
gcggttgaaa	gggggagggg	atgttgtcta	catggctctg	ctgtagtgag	tgggttgcgc	1680
tccggcagcg	gtctgatca	atcgtcacc	tttctcggtc	cttcaacgtt	cctgacaacg	1740
agcctccttt	tcgccaatcc	atcgacaatc	accgcgagtc	cctgctcgaa	cgctgcgtcc	1800
ggaccggctt	cgtcgaaggc	gtctatcgcg	gcccgaaca	gcggcgagag	cggagcctgt	1860
tcaacggtgc	cgccgcgctc	gccggcatcg	ctgtcgccgg	cctgctctc	aagcacggcc	1920
ccaacagtga	agtagctgat	tgtcatcagc	gcattgacgg	cgtecccggc	cgaaaaaccc	1980
gcctcgcaga	ggaagcgaag	ctgcgcgtcg	gccgtttcca	tctgcggtgc	gcccgtctgc	2040
gtgccggcat	ggatgcgcgc	gccatcgcg	taggcgagca	gcgcctgcct	gaagctgcgg	2100
gcattcccga	tcagaaatga	gcgccagtcg	tcgtcggtc	tcggcaccga	atgcgtatga	2160
ttctccgcca	gcatggcttc	ggccagtgcg	tcgagcagcg	cccgttgtt	cctgaagtgc	2220
cagtaaagcg	ccggctgctg	aacccccaac	cgttccgcca	gtttgcgtgt	cgtcagaccg	2280
tctacgccga	cctcgttcaa	caggtccagg	gcggcacgga	tactgtatt	cggctgcaac	2340
tttgtcatgc	ttgacacttt	atcaactgata	aacataatat	gtccaccaac	ttatcagtga	2400
taaagaatcc	gcgcgttcaa	tcggaccagc	ggaggtggt	ccggaggcca	gacgtgaaac	2460
ccaacatacc	cctgatcgta	attctgagca	ctgtcgctgt	cgacgtgtc	ggcatcgccc	2520
tgattatgcc	ggtgctgccg	ggcctcctgc	gcgatctggt	tactcgaac	gacgtaccg	2580
cccactatgg	cattctgctg	gcgctgtatg	cgttggtgca	atttgcctgc	gcacctgtgc	2640
tgggcgcgct	gtcggatcgt	ttcgggcggc	ggccaatctt	gctcgtctcg	ctggccggcg	2700
ccactgtcga	ctacgccatc	atggcgacag	cgctttctct	ttgggttctc	tatatcgggc	2760
ggatcgtggc	cggcatcacc	ggggcgactg	gggcggtagc	cggccttat	attgccgata	2820

tactgatgg	cgatgagcgc	gcgcggcact	tcggettcac	gagcgccctgt	ttcgggttcg	2880
ggatggctgc	gggacctgtg	ctcgggtggc	tgatgggcgg	tttctcccc	cacgctccgt	2940
tcttcgccgc	ggcagccttg	aacggcctca	atttcctgac	gggctgtttc	cttttgccgg	3000
agtcgcacaa	aggcgaacgc	cggccgttac	gccgggaggc	tctcaaccgc	ctcgccttcgt	3060
tccgggtggc	ccggggcatg	accgtcgtcg	ccgccctgat	ggcggctctc	ttcatcatgc	3120
aacttgctcg	acaggtgccg	gccgcgcttt	gggtcatttt	cggcgaggat	cgctttcact	3180
gggacgcgac	cacgatcggc	atttcgcttg	ccgcatttgg	cattctgcat	tactcgcgcc	3240
aggcaatgat	caccggccct	gtagccgcc	ggctcggcga	aaggcgggca	ctcatgctcg	3300
gaatgattgc	cgacggcaca	ggctacatcc	tgcttgcttc	cgcgacacgg	ggatggatgg	3360
cgttcccgat	catggtcctg	cttgcttcgg	gtggcatcgg	aatgccggcg	ctgcaagcaa	3420
tgttgtccag	gcaggtggat	gaggaacgtc	aggggcagct	gcaaggctca	ctggcggcgc	3480
tcaccagcct	gacctcgatc	gtcggacccc	tctcttccac	ggcgatctat	gcggcttcta	3540
taacaacgtg	gaacgggtgg	gcatggattg	caggecgtgc	cctctacttg	ctctgcctgc	3600
cggcgtcgcg	tcgcgggctt	tggagcggcg	cagggcaacg	agccgatcgc	tgatcgtgga	3660
aacgataggc	ctatgccatg	cgggtcaagg	cgacttccgg	caagctatac	gcgccctagg	3720
agtgcggttg	gaacgttggc	ccagccagat	actcccgatc	acgagcagga	cgccgatgat	3780
ttgaagcgc	ctcagcgtct	gatccaagaa	caaccatcct	agcaaacacg	cggtccccgg	3840
gctgagaaa	cccagtaagg	aaacaactgt	aggttcgagt	cgcgagatcc	cccgaacca	3900
aaggaagtag	gttaaaccgc	ctccgatcag	gccgagccac	gccaggccga	gaacattggt	3960
tctgttaggc	atcgggattg	gcggatcaaa	cactaaagct	actggaacga	gcagaagtcc	4020
tccggccgcc	agttgccagg	cggtaaaggt	gagcagaggc	acgggaggtt	gccacttgcc	4080
ggtcagcacg	gttccgaacg	ccatggaaac	cgccccgcc	aggcccgcctg	cgacgccgac	4140
aggatctagc	gctgcgtttg	gtgtcaaac	caacagcgcc	acgcccgcag	ttccgcaaat	4200
agcccccagg	accgccatca	atcgtatcgg	gctacctagc	agagcggcag	agatgaacac	4260
gaccatcagc	ggctgcacag	cgcctaccgt	cgccgcgacc	ccgcccggca	ggcggtagac	4320
cgaaataaac	aacaagctcc	agaatagcga	aatattaagt	gcgcccagga	tgaagatgcg	4380
catccaccag	attcccgttg	gaatctgtcg	gacgatcacc	acgagcaata	aaccggccgg	4440
caacgcccgc	agcagcatac	cggcgacccc	tcggcctcgc	tgctcgggct	ccacgaaaac	4500
gccggacaga	tcgccttgtg	gagcgtcctt	ggggccgtcc	tctgttttga	agaccgacag	4560
cccaatgatc	tcgcccgcga	tgtaggcgcc	gaatgccacg	gcattctcga	accgttcagc	4620
gaacgcctcc	atgggctttt	tctctctgtg	ctcgtaaacg	gaccgcaaca	tctctggagc	4680
tttcttcagg	gccgacaate	ggatctcgcg	gaaatctcgc	acgtcggccg	ctccaagccg	4740
tcgaatctga	gccttaatca	caattgtcaa	ttttaatcct	ctgtttatcg	gcagttcgta	4800
gagcgcgccg	tcgctcccga	gcgatactga	gcgaagcaag	tcgctcgagc	agtccccgct	4860
tgttcctgaa	atgccagtaa	agcgetgget	getgaacccc	cagccggaac	tgaccccaca	4920
aggccctagc	gtttgcaatg	caccaggtca	tcattgaccc	aggcgtgttc	caccaggccg	4980
ctgcctcga	actcttcgca	ggettcgccg	acctgctcgc	gccacttctt	cacgcgggtg	5040
gaatccgatc	cgcacatgag	gcggaaggtt	tccagcttga	gcgggtacgg	ctcccggctg	5100
gagctgaaat	agtcgaacat	ccgtcgggcc	gtcggcgaca	gcttgcggta	cttctcccat	5160

atgaatttcg	tgtagtggtc	gccagcaaac	agcacgacga	tttctctgtc	gatcaggacc	5220
tggcaacggg	acgttttctt	gccacggtcc	aggacgcgga	agcgggtgcag	cagcgacacc	5280
gattccagggt	gccaacgcg	gtcggacgtg	aagcccatcg	ccgtcgccctg	taggcgcgac	5340
aggcattcct	cggccttcgt	gtaataccgg	ccattgatcg	accagcccag	gtcctggcaa	5400
agctcgtaga	acgtgaaggt	gatcggctcg	ccgatagggg	tgcgcttcgc	gtactccaac	5460
acctgctgcc	acaccagttc	gtcatcgtcg	gcccgcagct	cgacgccggt	gtaggtgatc	5520
ttcacgtcct	tgttgacgtg	gaaaatgacc	ttgttttgca	gcgccctcgcg	cgggattttc	5580
ttgttgccgcg	tgggaacag	ggcagagcgg	gccgtgtcgt	ttggcatcgc	tcgcatcgtg	5640
tccggccacg	gcgcaatata	gaacaaggaa	agctgcattt	ccttgatctg	ctgcttcgtg	5700
tgtttcagca	acgcggcctg	cttggcctcg	ctgacctgtt	ttgccaggtc	ctcggcgcg	5760
gtttttcgtc	tcttggtcgt	catagtctct	cgcgtgtcga	tggteatcga	cttcgcaaaa	5820
cctgccgcct	cctgttcgag	acgacgcgaa	cgtccacgg	cggccgatgg	cgcgggcagg	5880
gcaggggggag	ccagttgcac	gctgtcgcgc	tcgatcttgg	ccgtagcttg	ctggaccatc	5940
gagccgacgg	actggaaggt	ttcgcggggc	gcacgcata	cgggtgcggt	tgcgatggtt	6000
tcggcatcct	cggcggaaaa	ccccgcgtcg	atcagttctt	gectgtatgc	cttcgggtca	6060
aacgtccgat	tcattcacc	tccttgccgg	attgccccga	ctcacgccg	ggcaatgtgc	6120
ccttattcct	gatttgacc	gcctgggtgc	ttgggtgtcca	gataatccac	cttatcgga	6180
atgaagtcgg	tcccgtagac	cgtctggccg	tccttctcgt	acttgggtatt	ccgaatcttg	6240
ccctgcacga	ataccagcga	ccccttgccc	aaatacttgc	cgtgggcctc	ggcctgagag	6300
ccaaaacact	tgatgcggaa	gaagtcggtg	cgtcctcgt	tgtcgcggc	atcgttgccg	6360
cactcttcat	taaccgctat	atcgaaaatt	gcttgccggt	tgttagaatt	gccatgacgt	6420
acctcgggtg	cacgggtaag	attaccgata	aactggaact	gattatggct	catatcgaaa	6480
gtctccttga	gaaaggagac	tctagttag	ctaaacattg	gttccgctgt	caagaacttt	6540
agcggctaaa	atthtgcggg	ccgcgaccaa	aggtgcgagg	ggcggcttcc	gctgtgtaca	6600
accagatatt	ttcaccaac	atccttcgtc	tgctcgatga	gcggggcatg	acgaaacatg	6660
agctgtcggg	gagggcaggg	gtttcaattt	cgthtttatc	agacttaacc	aacggttaag	6720
ccaacccctc	gttgaaggtg	atggaggcca	ttgccgacgc	cctggaaact	cccctacctc	6780
ttctcctgga	gtccaccgac	cttgaccgcg	aggcaactcg	ggagattgcg	ggtcatcctt	6840
tcaagagcag	cgtgccgccc	ggatacgaac	gcatacagtg	ggttttgccg	tcacataagg	6900
cgtttatcgt	aaagaaatgg	ggcgcgcgca	cccgaaaaa	gctgcgtgga	aggctctgac	6960
gccaagggtt	agggttgcg	cttccttctt	tagecgctaa	aacggcccct	tctctgcggg	7020
ccgtcggctc	gcgcatcata	tcgacatcct	caacgggaagc	cgtgccgcga	atggcatcgg	7080
gcgggtgcgc	tttgacagtt	gtttctctatc	agaaccctca	cgtcgtgcgg	ttcgattagc	7140
tgtttgtctt	gcaggctaaa	cactttcgggt	atatcgtttg	cctgtgcgat	aatgttgcta	7200
atgatttgtt	gcgtaggggt	tactgaaaag	tgagcgggaa	agaagagttt	cagaccatca	7260
aggagcgggc	caagcgaag	ctggaacgcg	acatgggtgc	ggacctgttg	gccgcgctca	7320
acgaccgcaa	aaccgttgaa	gtcatgctca	acgcggacgg	caaggtgtgg	cacgaacgcc	7380
ttggcgagcc	gatgcggtac	atctgcgaca	tgcggcccag	ccagtcgcag	gcgattatag	7440
aaacgggtggc	cggattccac	ggcaaaagag	tcacgcggca	ttcgcctcctc	ctggaaggcg	7500

agttcccctt	ggatggcagc	cgctttgccg	gccaatgccc	gccggtcgtg	gccgcgcca	7560
cctttgcgat	ccgcaagcgc	gcggtcgcca	tcttcacgct	ggaacagtac	gtcgaggcgg	7620
gcatcatgac	ccgcgagcaa	tacgaggcca	ttaaaagcgc	cgtcgcggcg	catcgaacaa	7680
tcctcgtcat	tggcggctact	ggctcgggca	agaccacgct	cgtaacgcg	atcatcaatg	7740
aaatggctgc	cttcaaccgg	tctgagcgcg	tcgtcatcat	cgaggacacc	ggcgaatcc	7800
agtgcgccgc	agagaacgcc	gtccaatacc	acaccagcat	cgacgtctcg	atgacgctgc	7860
tgctcaagac	aacgctgcgt	atgcgccccg	accgcatcct	ggtcggtgag	gtacgtggcc	7920
ccgaagccct	tgatctgttg	atggcctgga	acaccgggca	tgaaggaggt	gccgccacc	7980
tgacgcgcaa	caacccccaa	gcgggectga	gccggctcgc	catgcttacc	agcatgcacc	8040
cggattcacc	gaaaccatt	gagccgctga	ttggcgaggc	ggttcatgtg	gtcgtccata	8100
tcgccaggac	ccctagcggc	cgctcagctg	aagaaattct	cgaagttctt	ggttacgaga	8160
acggccagta	catcaccaaa	accctgtaag	gagtatttcc	aatgacaacg	gctgttccgt	8220
tccgtctgac	catgaatcgc	ggcattttgt	tctaccttgc	cggtttcttc	gttctcgtc	8280
tcgcgttacc	cgccatccg	gcgatggcct	cggaaggcac	cggcggcagc	ttgccatag	8340
agagctggct	gacgaacctg	cgcaactccg	taaccggccc	ggtggccttc	gcgctgtcca	8400
tcatcggcat	cgctcgcgcc	ggcggcgtgc	tgatcttcgg	cggcgaactc	aacgccttct	8460
tccgaaccct	gatcttctctg	gttctgggtga	tgccgctgct	ggtcggcgcg	cagaacgtga	8520
tgagcacctt	cttcggctctg	ggtgccgaaa	tcgcggccct	cggcaacggg	gcgctgcacc	8580
aggtgcaagt	cgccggcgcg	gatgccgtgc	gtccggttagc	ggctggacgg	ctcgcctaata	8640
catggctctg	cgccatccg	ccatccgctg	cgccagcaac	cgagaaaacc	tggtcatggg	8700
tggtgatcgt	gaactgggtga	tgctctcggg	cctgatggcg	tttgcgctga	ttttcagcgc	8760
ccaagagctg	cgggccaccg	tggtcggctt	gatcctgtgg	ttcggggcgc	tctatgcgtt	8820
ccgaatcatg	gcgaaggccc	atccgaagat	gcggttctgt	tacctgcgtc	accgccggta	8880
caagccgtat	taccggccc	gctcgacccc	gttccgcgag	aacaccaata	gccaaaggaa	8940
gcaataccga	tgatccaagc	aattgcgatt	gcaatcgcgg	gcctcggcgc	gcttctgttg	9000
ttcatcctct	ttgcccgcat	ccgcgcggtc	gatgccgaac	tgaaactgaa	aaagcatcgt	9060
tccaaggacg	ccggcctggc	cgatctgctc	aactacgccg	ctgtcgtcga	tgacggcgta	9120
atcgtgggca	agaacggcag	ctttatggct	gcttggctgt	acaaggcgca	tgacaacgca	9180
agcagcaccg	accagcagcg	cgaagtagtg	tcgccccgca	tcaaccaggc	cctcgcgggc	9240
ctgggaagtg	ggtggatgat	ccatgtggac	gccgtcggcg	gtcctgctcc	gaactacgcg	9300
gagcggggcc	gtcggcgctt	cctgaccgt	ctgacggcag	cgattgaaga	agagcgtctg	9360
gtcttgcctt	gctcgtcggg	gatgtaactc	accagctccg	cgaagtcgct	cttcttgatg	9420
gagcgcattg	ggacgtgctt	ggcaatcacg	cgccccccc	ggcgtttta	gcggctaaaa	9480
aagtcattgg	tctgccctcg	ggcggaccac	gcccattcat	accttgccaa	gctcgtcctg	9540
cttctcttctg	atcttccgca	gcagggcgag	gatcgtggca	tcaccgaacc	gcgccgtgcg	9600
cgggtcgtcg	gtgagccaga	gtttcagcag	gccgccaggg	cggccccagg	cgccattgat	9660
gcgggcccagc	tcgcccagct	gctcatagtc	cacgacgccc	gtgattttgt	agccctggcc	9720
gacggcccagc	aggtaggccc	acaggetcat	gccggcccgc	gccgcctttt	cctcaatcgc	9780
tcttcgttctg	tctggaaggc	agtacacctt	gataggtggg	ctgcccttcc	tggttgctt	9840



ggtttcatca	gccatccgct	tgccctcadc	tgttacgccc	gcggtagccc	gccagcctcg	9900
cagagcagga	ttcccgttga	gcaccgccag	gtgcgaataa	gggacagtga	agaaggaaca	9960
cccgcctcgcg	ggtgggccta	cttcacctat	cctgccccgc	tgacgccgtt	ggatacacca	10020
aggaaagtct	acacgaacct	tttgcaaaa	tcctgtatat	cgtgcgaaaa	aggatggata	10080
taccgaaaaa	atcgctataa	tgacccccgaa	gcagggttat	gcagcggaaa	agcgcctgctt	10140
ccctgctggt	ttgtggaata	tctaccgact	ggaaacaggc	aatgcagga	aattactgaa	10200
ctgaggggac	aggcgagaga	cgatgccaaa	gagctacacc	gacgagctgg	ccgagtgggt	10260
tgaatcccgc	gcggccaaaga	agcgcggcg	tgatgaggct	gcggttgcgt	tcctggcggt	10320
gagggcggat	gtcgaggcgg	cgttagcgtc	cggtatgcg	ctcgtcacca	tttgggagca	10380
catgcgggaa	acggggaagg	tcaagttctc	ctacgagacg	ttccgctcgc	acgccaggcg	10440
gcacatcaag	gccaagcccc	ccgatgtgcc	cgaccgcag	gccaaggctg	cggaacccgc	10500
gccggcaccc	aagacgccgg	agccacggcg	gccgaagcag	gggggcaagg	ctgaaaagcc	10560
ggcccccgct	gcgggccccga	ccggettca	cttcaacca	acaccggaca	aaaaggatct	10620
actgtaatgg	cgaaaattca	catggttttg	cagggcaagg	gcggggtcgg	caagtcggcc	10680
atcgccgcga	tcattgcgca	gtacaagatg	gacaaggggc	agacaccctt	gtgcatcgac	10740
accgaccggg	tgaacgcgac	gttcgagggc	tacaaggccc	tgaacgtccg	ccggtgaac	10800
atcatggccg	gcgacgaaat	taactcgcgc	aacttcgaca	ccctggtcga	gctgattgcg	10860
ccgaccaagg	atgacgtggt	gatcgacaac	ggtgccagct	cgttcgtgcc	tctgtcgcac	10920
tacctcatca	gcaaccaggt	gccggtctcg	ctgcaagaaa	tggggcatga	gctggtcadc	10980
cataccgtcg	tcaccggcgg	ccaggtcttc	ctggacacgg	tgagcggctt	cgcccagctc	11040
gccagccagt	ttccggccga	agcgtttttc	gtggtctggc	tgaaccctga	ttgggggctt	11100
atcgagcatg	agggcaagag	ctttgagcag	atgaaggcgt	acacggccaa	caaggccccg	11160
gtgtcgtcca	tcattccagat	tccggccttc	aaggaagaaa	cctacggccc	cgatttcagc	11220
gacatgctgc	aagagcggct	gacgttcgac	caggcgttgg	ccgatgaatc	gctcacgatc	11280
atgacgcggc	aacgcctcaa	gatcgtgcgg	cgcgccctgt	ttgaacagct	cgacgcggcg	11340
gccgtgctat	gagcgaccag	attgaagagc	tgatccggga	gattgcggcc	aagcacggca	11400
tcgccgtcgg	ccgcgacgac	ccggtgctga	tctgcatac	catcaacgcc	cggtctatgg	11460
ccgacagtgc	ggccaagcaa	gaggaaatcc	ttgcccgctt	caaggaagag	ctggaaggga	11520
tcgcccacgc	ttggggcgag	gacgccaagg	caaagcgga	gcggtatgctg	aacgcggccc	11580
tggcggccag	caaggacgca	atggcgaagg	taatgaagga	cagcgcggcg	caggcggccc	11640
aagcgatccg	cagggaaatc	gacgacggcc	ttggccgcca	gctcgcggcc	aaggtcgcgg	11700
acgcgcggcg	cgtagcgatg	atgaacatga	tcgccggcgg	catggtgttg	ttcgcggccc	11760
ccctggtggt	gtgggcctcg	ttatgaatcg	cagaggcgca	gatgaaaaag	cccggcgttg	11820
ccgggctttg	tttttgcgtt	agctgggett	gtttgacagg	cccaagctct	gactgcgccc	11880
gcgctcgcgc	tcctgggcct	gtttctcttc	ctgctctgc	ttgcgcatca	gggcctggtg	11940
ccgtcgggct	gcttcacgca	tcgaatccca	gtcgcggccc	agctcgggat	gctccgcgcg	12000
catcttgcgc	gtcgcagtt	cctcgatctt	gggcgcgtga	atgccatgc	cttcttgat	12060
ttcgcgcacc	atgtccagcc	gcgtgtgcag	ggtctgcaag	cgggcttgct	gttgggcctg	12120
ctgctgctgc	caggcggcct	ttgtacgcgg	cagggacagc	aagccggggg	cattggactg	12180

tagctgctgc	aaacgcgcct	gctgacggtc	tacgagctgt	tctaggecgt	cctcgatgcg	12240
ctccacctgg	tcatgctttg	cctgcacgta	gagcgcaagg	gtctgctggt	aggtctgctc	12300
gatgggcgcg	gattctaaga	gggcctgctg	ttccgtctcg	gcctcctggg	ccgcctgtag	12360
caaatcctcg	ccgctgttgc	cgctggactg	ctttactgcc	ggggactgct	gttgccctgc	12420
tcgcgccgtc	gtcgcagttc	ggcttgcccc	cactcgattg	actgcttcat	ttcgagccgc	12480
agcgatgcga	tctcggattg	cgtaacgga	cggggcagcg	cggaggtgtc	cggcttctcc	12540
ttgggtgagt	cggctgatgc	catagccaaa	ggtttccttc	caaaatgcgt	ccattgctgg	12600
accgtgtttc	tattgatgc	ccgcaagcat	cttcggcttg	accgccaggt	caagcgcgcc	12660
ttcatgggcg	gtcatgacgg	acgccgccat	gaccttgccg	ccgttgttct	cgatgtagcc	12720
gcgtaatgag	gcaatggtgc	cgcccatcgt	cagcgtgtca	tcgacaacga	tgtacttctg	12780
gccggggatc	acctccccct	cgaaaagtcgg	gttgaacgcc	aggcgatgat	ctgaaccggc	12840
tccggttcgg	gcgaccttct	cccgetgeac	aatgtccgtt	tcgacctcaa	ggccaaggcg	12900
gtcggccaga	acgaccgcca	tcatggccgg	aatcttgttg	ttccccgccc	cctcgacggc	12960
gaggactgga	acgatgcggg	gcttgctgct	gccgatcagc	gtcttgagct	gggcaacagt	13020
gtcgtccgaa	atcaggecgt	cgaccaaatt	aagcgcgcgt	tccgcgtcgc	cctgcttcgc	13080
agcctggtat	tcaggctcgt	tggtaaaaga	accaaggtcg	ccgttgcgaa	ccaccttcgg	13140
gaagtctccc	cacggtgcgc	gctcggetct	gctgtagctg	ctcaagacgc	ctcccttttt	13200
agccgctaaa	actctaacga	gtgcgcccgc	gactcaactt	gacgctttcg	gcacttacct	13260
gtgccttgcc	acttgctgca	taggtgatgc	ttttcgcact	cccgatttca	ggtactttat	13320
cgaaatctga	ccgggcgtgc	attacaaagt	tcttccccac	ctgttggtaa	atgctgccgc	13380
tatctgcgtg	gacgatgctg	ccgtcgtggc	gctgcgactt	atcggccttt	tgggccatat	13440
agatgttgta	aatgccaggt	ttcagggcc	cgctttatc	taccttctgg	ttcgtccatg	13500
cgccttgggt	ctcggctctg	acaattcttt	gccattcat	gaccaggagg	cgggtgttca	13560
ttgggtgact	cctgacgggt	gcctctgggt	ttaaactgtg	cctggctcgt	tgccggctaa	13620
aaaaaagccg	acctcggcag	ttcagggcc	gctttcccta	gagccgggcg	cgtaaggtt	13680
gttccatcta	ttttagtgaa	ctgcgttcga	ttatcagtt	actttcctcc	cgctttgtgt	13740
ttcctcccac	tcgtttccgc	gtctagccga	ccctcaaca	tagcggcctc	ttcttgggct	13800
gcctttgcct	cttgccgcgc	ttcgtcacgc	tcggettga	ccgtcgtaaa	gcgctcggcc	13860
tgcttgcccg	cctcttgcgc	cgccaacttc	ctttgctcct	ggtgggcctc	ggcgtcggcc	13920
tgcccttcg	ctttaccgc	tgccaactcc	gtgcgcaaac	tctccgcttc	gcgcctgggtg	13980
gcgtcgcgct	cgccgcgaag	cgctgcatt	tcttggttg	ccgcgtccag	ggtcttgcgg	14040
ctctcttctt	tgaatgcgcg	ggcgtcctgg	tgagcgtagt	ccagctcggc	gcgcagctcc	14100
tgcgctcgac	gctccacctc	gtcggccccg	tgcgtcgcca	gcgcggcccc	ctgctcggct	14160
cctgccaggg	cgggtgcgtc	ttcggccagg	gcttgccgct	ggcgtcgggc	cagctcggcc	14220
gcctcggcgg	cctgctgctc	tagcaatgta	acgcgcgcct	gggcttcttc	cagctcgcgg	14280
gcctgcgcct	cgaaggcgtc	ggccagctcc	ccgcgcacgg	cttccaactc	gttgcgctca	14340
cgatcccagc	cggcttgcgc	tgctgcaac	gattcattgg	caaggcctg	ggcggcttgc	14400
cagagggcgg	ccacggcctg	gttgccggcc	tgtgcaccg	cgtccggcac	ctggactgcc	14460
agcggggcgg	cctgcgccgt	gcgctggcgt	cgccattcgc	gcatgccggc	gctggcgtcg	14520

ttcatgttga	cgcgggcggc	cttacgcact	gcatccacgg	tcgggaagtt	ctcccggtcg	14580
ccttgctcga	acagctcgtc	cgcagccgca	aaaatgcggt	cgcgcgtctc	ttgtttcagt	14640
tccatgttgg	ctccggtaat	tgtaagaat	aataatactc	ttacctacct	tatcagcgca	14700
agagtttagc	tgaacagttc	tcgacttaac	ggcaggtttt	ttagcggctg	aagggcaggc	14760
aaaaaaagcc	ccgcacggtc	ggcgggggca	aagggtcagc	gggaagggga	ttagcgggcg	14820
tcgggcttct	tcatgcgtcg	gggccgcgct	tcttgggatg	gagcacgacg	aagcgcgcac	14880
gcgcatcgtc	ctcggccccta	tcggcccgcg	tcgcggtcag	gaacttgtcg	cgcgctaggt	14940
cctccctggt	gggcaccagg	ggcatgaaact	cggcctgctc	gatgtaggtc	cactccatga	15000
ccgcatcgca	gtcgagcccg	cgttccttea	ccgtctcttg	caggtcgcgg	tacgcccgct	15060
cgttgagcgg	ctggtaacgg	gccaattggt	cgtaaatggc	tgtcggccat	gagcggcctt	15120
tcctgttgag	ccagcagccg	acgacgaagc	cggcaatgca	ggcccctggc	acaaccaggc	15180
cgacgccggg	ggcaggggat	ggcagcagct	cgccaaccag	gaaccccgcc	gcgatgatgc	15240
cgatgccggt	caaccagccc	ttgaaactat	ccggccccga	aacaccctcg	cgcattgcct	15300
ggatgctgcg	ccggatagct	tgcaacatca	ggagccgttt	cttttgtteg	tcagtcatgg	15360
tccgccctca	ccagttgttc	gtatcgggtg	cggacgaact	gaaatcgcaa	gagctgccgg	15420
tatcggcca	gccgctgtcc	gtgtcgtgct	tgccgaagca	cggcgagggg	tccgcgaacg	15480
ccgcagacgg	cgtatccggc	cgcagcgcac	cgcccagcat	ggccccggtc	agcgagccgc	15540
cggccaggta	gcccagcatg	gtgctgttgg	tcgccccggc	caccagggcc	gacgtgacga	15600
aatcgccgtc	attcccctctg	gattgttctg	tgctcggcgg	ggcagtgcgc	cgcgccggcg	15660
gcgtcgtgga	tggctcgggt	tggctggcct	gcgacggccc	gcgaaaggtg	cgcagcagct	15720
cgttatcgac	cggctcgggc	gtcggggccg	ccgccttgcg	ctgcggctgg	tgttctctct	15780
tcggctcgcg	cagcttgaac	agcatgatcg	cggaaaccag	cagcaacgcc	gcgcctacgc	15840
ctcccgcgat	gtagaacagc	atcgattca	ttcttcggtc	ctcctttag	cggaaccggt	15900
gtctgtgcgg	cgcgggtggc	ccgcgccgct	gtctttgggg	atcagccctc	gatgagcgcg	15960
accagtttca	cgtcggcaag	gttcgcctcg	aactcctggc	cgctgtctctc	gtacttcaac	16020
caggcatagc	cttccgccgg	cggccgacgg	ttgaggataa	ggcgggcagg	gcgctcgtcg	16080
tgctcgacct	ggacgatggc	cttttccagc	ttgtccgggt	ccggctcctt	cgcgcccttt	16140
tccttgccgt	ccttaccgct	ctggtcgccc	tctcgcctgt	cctggccgctc	gccggcctcc	16200
gcgtcacgct	cggcatcagt	ctggccgttg	aaggcatcga	cgggtgttggg	atcgcggccc	16260
ttctcgtcca	ggaactcgcg	cagcagcttg	accgtgccgc	gcgtgatttc	ctgggtgtcg	16320
tcgtcaagcc	acgcctcgac	ttcctccggg	cgttcttga	aggccgacac	cagctcgttc	16380
accacggtca	cgtcgcgcac	gcggccggtg	ttgaacgcat	cggcgatctt	ctccggcagg	16440
tccagcagcg	tgacgtgctg	ggtgatgaac	gccggcgact	tgccgatttc	cttggcgata	16500
tcgcctttct	tcttgccttt	cgccagctcg	cggccaatga	agtcggcaat	ttcgcgcggg	16560
gtcagctcgt	tgcgttgacg	gttctcgata	acctggctcg	cttcgttga	gtcgttgcg	16620
atgaacgccg	ggatggactt	cttgcggccc	cacttcgagc	cacggtagcg	gcgggcgccg	16680
tgattgatga	tatagcggcc	cggtctctcc	tggttctcgc	gcaccgaaat	gggtgacttc	16740
accccgcgct	ctttgatcgt	ggcaccgatt	tccgcgatgc	tctccgggga	aaagccgggg	16800
ttgtcggccg	tccgcggctg	atgcggatct	tcgtcgatca	ggtccaggtc	cagctcgata	16860

gggccggaac	cgccctgaga	cgccgcagga	gcgtccagga	ggctcgacag	gtcgccgatg	16920
ctatccaacc	ccaggccgga	cggctgcgcc	gcgcctgcgg	cttctgagc	ggccgcagcg	16980
gtgtttttct	tgggtgtctt	ggcttgagcc	gcagtcattg	ggaaatctcc	atcttcgtga	17040
acacgtaatc	agccagggcg	cgaacctctt	tcgatgcctt	gcgcgcggcc	gttttcttga	17100
tcttccagac	cggcacaccg	gatgcgaggg	catcggcgat	gctgctgcgc	aggccaacgg	17160
tggccggaat	catcatcttg	gggtacgcgg	ccagcagctc	ggcttgggtg	cgcgcggtggc	17220
gcggattccg	cgcatcgacc	ttgctgggca	ccatgcccaag	gaattgcagc	ttggcgttct	17280
tctggcgcac	gttcgcaatg	gtcgtgacca	tcttcttgat	gccctggatg	ctgtacgcct	17340
caagctcgat	gggggacagc	acatagtcgg	ccgcgaagag	ggcggccgcc	aggccgacgc	17400
caagggtcgg	ggccgtgtcg	atcaggcaca	cgtcgaagcc	ttggttcgcc	agggccttga	17460
tgttcgcccc	gaacagctcg	cgggcgtcgt	ccagcgacag	ccgttcggcg	ttcgccagta	17520
ccgggttggg	ctcgatgagg	gcgaggcgcg	cggcctggcc	gtcgccggct	gcgggtgcgg	17580
tttcggtcca	gccgccggca	gggacagcgc	cgaacagctt	gcttgcatgc	aggccggtag	17640
caaagtcctt	gagcgtgtag	gacgcattgc	cctgggggtc	caggtcgatc	acggcaacce	17700
gcaagcccg	ctcgaaaaag	tcgaaggcaa	gatgcacaag	ggtcgaagtc	ttgccgacgc	17760
cgctttctg	gttgcccg	accaaagttt	tcategttg	gtttctgtt	ttttcttggc	17820
gtccgcttcc	cacttccgga	cgatgtacgc	ctgatgttcc	ggcagaaccg	ccgttaccgg	17880
cgcgtacccc	tcgggcaagt	tcttgcctc	gaacgcggcc	cacacgcgat	gcaccgcttg	17940
cgacactgcg	cccctggtea	gtcccagcga	cgttgcgaa	gtcgctgtg	gcttcccatc	18000
gactaagacg	ccccgcgcta	tctcgatggt	ctgctgcccc	acttccagcc	cctggatcgc	18060
ctcctggaac	tggctttcgg	taagccgttt	cttcatggat	aacaccata	atttgcctcg	18120
cgctttgggt	gaacatagcg	gtgacagccg	ccagcacatg	agagaagttt	agctaaacat	18180
ttctcgcacg	tcaaacctt	tagccgctaa	aactcgtcct	tggcgtaaca	aaacaaaagc	18240
ccggaaaccg	ggctttcgtc	tcttgcgct	tatggctctg	caccgcgctc	catcaccaac	18300
aggctcgcga	cgcgcttcc	tcggttgcgg	atcgacactg	ccagcccaac	aaagccggtt	18360
gccgccgccg	ccaggatcgc	gccgatgatg	ccggccacac	cggccatcgc	ccaccaggtc	18420
gccgccttcc	ggttccattc	ctgctggtac	tgcttcgcaa	tgctggacct	cggctcacca	18480
taggctgacc	gctcgatggc	gtatgccgct	tctccccttg	gcgtaaaacc	cagcgcgcga	18540
ggcggcattg	ccatgctgcc	cgccgcttcc	ccgaccacga	cgcgcgacc	aggcttgcgg	18600
tccagacctt	cggccacggc	gagctgcgca	aggacataat	cagccgccga	cttggctcca	18660
cgcgctcga	tcagctcttg	cactcgcgcg	aaatccttgg	cctccacggc	cgccatgaat	18720
cgcgcacgcg	gcgaaggctc	cgcagggccg	gcgtcgtgat	cgccgccgag	aatgccttc	18780
accaagtctg	acgacacgaa	aatcatgctg	acggctatca	ccatcatgca	gacggatcgc	18840
acgaaccgcg	tgaattgaac	acgagcacgg	caccgcgcac	cactatgcca	agaatgcccc	18900
aggtaaaaaat	tgccggcccc	gccatgaagt	ccgtgaatgc	cccgcggcc	gaagtgaagg	18960
gcaggccgcc	accagggccg	ccgccctcac	tgccgggcac	ctggctcctg	aatgtcgatg	19020
ccagcacctg	cggcacgtca	atgettccgg	gcgtcgcgct	cgggctgac	gcccattccc	19080
ttactgcccc	gatcccggca	atggcaagga	ctgccagcgc	tgccattttt	ggggtgaggg	19140
cgttcgcggc	cgagggggcg	agcccctggg	gggatgggag	gcccgcgtta	gcgggcccgg	19200

agggttcgag	aagggggggc	accccccttc	ggcgtgcegg	gtcacgcgca	cagggcgcag	19260
ccctggttaa	aaacaagggt	tataaatatt	ggtttaaaag	caggttaaaa	gacaggttag	19320
cggtgccga	aaaacgggcg	gaaacccttg	caaagtctgg	atcttctgcc	tgtggacagc	19380
ccctcaaatg	tcaatagggt	cgccccctcat	ctgtcagcac	tctgccccctc	aagtgtcaag	19440
gatcgcgccc	ctcatctgtc	agtagtcgcg	cccctcaagt	gtcaataccg	cagggcactt	19500
atccccaggc	ttgtccacat	catctgtggg	aaactcgcgt	aaaatcaggc	gttttcgccc	19560
atctgcgagg	ctggccagct	ccacgtcgcc	ggccgaaatc	gagcctgccc	ctcatctgtc	19620
aacgccgcg	cgggtgagtc	ggccccctcaa	gtgtcaacgt	ccgccccctca	tctgtcagtg	19680
agggccaaagt	tttccgcgag	gtatccacaa	cgccggcgcc	cgcggtgtct	cgcacacggc	19740
ttcgacggcg	tttctggcgc	gtttgacagg	ccatagacgg	ccgccagccc	agcggcgagg	19800
gcaaccagcc	cggtagcgt	cggaaaggcg	ctggaagccc	cgtagcgacg	cggagagggg	19860
cgagacaagc	caagggcgca	ggetcgatgc	gcagcacgac	atagccggtt	ctcgcaagga	19920
cgagaatttc	cctgcgggtg	ccctcaagtg	tcaatgaaag	tttccaacgc	gagccattcg	19980
cgagagcctt	gagtccacgc	tagatgagag	ctttgttgta	ggtggaccag	ttggtgattt	20040
tgaacttttg	ctttgccacg	gaacggtctg	cgttgtcggg	aagatgcgtg	atctgatcct	20100
tcaactcagc	aaaagtctga	tttattcaac	aaagccacgt	tgtgtctcaa	aatctctgat	20160
gttacattgc	acaagataaa	aatatatcat	catgaacaat	aaaactgtct	gcttacataa	20220
acagtaatac	aaggggtggt	atgagccata	ttcaacggga	aacgtcttgc	tcgactctag	20280
agctcgttcc	tcgaggcctc	gaggcctcga	ggaacggtac	ctgcggggaa	gcttacaata	20340
atgtgtgttg	ttaagtcttg	ttgcctgtca	tcgtctgact	gactttcgtc	ataaatcccg	20400
gcctccgtaa	cccagctttg	ggcaagctca	cggatttgat	ccggcggaac	gggaatatcg	20460
agatgccggg	ctgaacgctg	cagttccagc	tttcccttcc	gggacaggta	ctccagctga	20520
ttgattatct	gctgaagggt	cttggttcca	cctcctggca	caatgcgaat	gattacttga	20580
gcgcgatcgg	gcatccaatt	ttctcccgtc	agggtcgtgg	tcaagtgcta	caaggcacct	20640
ttcagtaacg	agcgaccgtc	gatccgtcgc	cgggatacgg	acaaaatgga	gcgcagtagt	20700
ccatcgaggg	cggcgaaagc	ctcgccaaaa	gcaatacgtt	catctcgcac	agcctccaga	20760
tccgatcgag	ggtcttcggc	gtaggcagat	agaagcatgg	atacattgct	tgagagtatt	20820
ccgatggact	gaagtatggc	ttccatcttt	tctcgtgtgt	ctgcatctat	ttcgagaaag	20880
cccccgatgc	ggcgcaccgc	aacgcgaatt	gccatactat	ccgaaagtcc	cagcaggcgc	20940
gcttgatagg	aaaaggtttc	atactcggcc	gatcgcagac	gggcactcac	gaccttgaac	21000
ccttcaactt	tcagggatcg	atgctggttg	atggtagtct	cactcgacgt	ggctctgggtg	21060
tgttttgaca	tagcttctct	caaagaaagc	ggaaggctctg	gatactccag	cacgaaatgt	21120
gcccgggtag	acggatggaa	gtctagccct	gtcaatatg	aatcaacag	tacatttaca	21180
gtcaatactg	aatatacttg	ctacatttgc	aattgtctta	taacgaatgt	gaaataaaaa	21240
tagtgtaaca	acgcttttac	tcatcgataa	tcacaaaaac	atctatacga	acaaaaatac	21300
aaatgcactc	cggtttcaca	ggatagcgcg	gatcagaata	tgcaactttt	gacgttttgt	21360
tctttcaaag	ggggtgctgg	caaaaccacc	gcactcatgg	gcctttgcgc	tgctttggca	21420
aatgacggta	aacgagtggc	cctctttgat	gccgacgaaa	accggcctct	gacgcgatgg	21480
agagaaaacg	ccttacaag	cagtactggg	atcctcgtctg	tgaagtctat	tccgccgacg	21540

aaatgcccct	tcttgaagca	gcctatgaaa	atgccgagct	cgaaggattt	gattatgcgt	21600
tggccgatac	gcgtggcggc	tcgagcgagc	tcaacaacac	aatcatcgct	agctcaaacc	21660
tgettctgat	cccaccatg	ctaacgccgc	tcgacatcga	tgaggcacta	tctacctacc	21720
gctacgtcat	cgagctgctg	ttgagtgaaa	atttggcaat	tcctacagct	gttttgcgcc	21780
aacgcgtccc	ggtcggccga	ttgacaacat	cgcaacgcag	gatgtcagag	acgctagaga	21840
gccttccagt	tgtaccgtct	cccatgcatg	aaagagatgc	atttggccg	atgaaagaac	21900
gcggcatggt	gcatcttaca	ttactaaaca	cgggaaactga	tccgacgatg	cgctcatag	21960
agaggaatct	tcggattgcg	atggaggaag	tcgtgggtcat	ttcgaaactg	atcagcaaaa	22020
tcttggaggc	ttgaagatgg	caattcgcaa	gcccgcattg	tcggtcggcg	aagcacggcg	22080
gcttgcgtgg	gctcgaccgg	agatccacca	tcccaaccgg	acacttgttc	cccagaagct	22140
ggacctccag	cacttgccctg	aaaaagccga	cgagaaagac	cagcaacgtg	agcctctcgt	22200
cgccgatcac	atttacagtc	ccgatcgaca	acttaageta	actgtggatg	cccttagtec	22260
acctccgtcc	ccgaaaaagc	tccaggtttt	tctttcagcg	cgaccgccc	cgctcaagt	22320
gtcgaaaaca	tatgacaacc	tcgttcggca	atacagtecc	tcgaagtcgc	tacaaatgat	22380
tttaaggcgc	gcgttggacg	atttcgaaag	catgctggca	gatggatcat	ttcgcgtggc	22440
cccgaaaagt	tatccgatec	cttcaactac	agaaaaatcc	gttctcgttc	agacctcacg	22500
catgttccc	gttgcgttgc	tcgaggtcgc	tcgaagtcat	tttgatccgt	tggggttggga	22560
gaccgctcga	gctttcggcc	acaagctggc	taccgccgcg	ctcgcgtcat	tctttgctgg	22620
agagaagcca	tcgagcaatt	ggtgaagagg	gacctatcgg	aaccctcac	caaatattga	22680
gtgtaggttt	gaggccgctg	gccgcgtcct	cagtcacett	ttgagccaga	taattaagag	22740
ccaaatgcaa	ttggctcagg	ctgccatcgt	ccccccgtgc	gaaacctgca	cgteccgcgtc	22800
aaagaaataa	ccggcacctc	ttgctgtttt	tatcagttga	gggcttgacg	gatccgcctc	22860
aagtttgcgg	cgagccgca	aaatgagaac	atctatactc	ctgtcgtaaa	cctcctcgtc	22920
gcgtactcga	ctggcaatga	gaagttgctc	gcgcgataga	acgtcgcggg	gtttctctaa	22980
aaacgcgagg	agaagattga	actcacctgc	cgtaagtttc	acctaccgc	cagcttcgga	23040
catcaagcga	cgttgcctga	gattaagtgt	ccagtcagta	aaacaaaaag	accgtcggtc	23100
tttggagcgg	acaacgttgg	ggcgcacgcg	caaggcaacc	cgaatgcgtg	caagaaactc	23160
tctcgtacta	aacggcttag	cgataaaatc	acttgetcct	agctcgagtg	caacaacttt	23220
atccgtctcc	tcaaggcgg	cgccactgat	aattatgatt	ggaatatacag	actttgccgc	23280
cagatttcga	acgatctcaa	gcccattctc	acgacctaaa	tttagatcaa	caaccacgac	23340
atcgaccgtc	gcggaagaga	gtactctagt	gaactgggtg	ctgtcggeta	ccgcggtcac	23400
tttgaaggcg	tggatcgtaa	ggtattcgat	aataagatgc	cgcatagcga	catcgtcate	23460
gataagaaga	acgtgtttca	acggetcacc	tttcaateta	aaatctgaac	cettgttccac	23520
agcgttgag	aaattttcc	gtgaaggatg	tacaatcate	tccagctaaa	tgggcagttc	23580
gtcagaattg	cggctgaccg	cggatgacga	aaatgcgaac	caagtatttc	aattttatga	23640
caaaagtct	caatcgttgt	tacaagtgaa	acgettcgag	gttacageta	ctattgatta	23700
aggagatcgc	ctatggtctc	gccccggcgt	cgtgcgtccg	ccgcgagcca	gatctcgcct	23760
acttcataaa	cgctctcata	ggcacggaat	ggaatgatga	catcgatcgc	cgtagagagc	23820
atgtcaatca	gtgtgcgatc	ttccaageta	gcaccttggg	cgctactttt	gacaagggaa	23880

aacagtttct	tgaatccttg	gattggattc	gcgccgtgta	ttgttgaaat	cgatcccgga	23940
tgtcccgaga	cgacttcact	cagataagcc	catgctgcat	cgtcgcgcat	ctcgccaagc	24000
aatatccggg	ccggccgcat	acgcagactt	gcttggagca	agtgctcggc	gctcacagca	24060
cccagcccag	caccgttctt	ggagtagagt	agtctaacat	gattatcgtg	tggaatgacg	24120
agttcgagcg	tatcttctat	ggtgattagc	ctttcctggg	gggggatggc	gctgatcaag	24180
gtcttgcetca	ttgttgcctt	gccgcttccg	gtagggccac	atagcaacat	cgtcagtcgg	24240
ctgacgacgc	atgctgacag	aaacgcttcc	aaatccccgt	tgtcaaaatg	ctgaaggata	24300
gcttcatcat	cctgattttg	gcgtttcctt	cgtgtctgcc	actggttcca	cctcgaagca	24360
tcataacggg	aggagacttc	ttaaagacca	gaaacacgcg	agcttggccg	tcgaatggtc	24420
aagctgacgg	tgcccgaggg	aacggtcggc	ggcagacaga	ttttagtcg	ttcaccacca	24480
ggaagttcag	tggcgcagag	ggggttacgt	ggtccgacat	cctgctttct	cagcgcgccc	24540
gctaaaaatag	cgatatcttc	aagatcatca	taagagacgg	gcaaaggcat	cttggtaaaa	24600
atgccggctt	ggcgcacaaaa	tgctcttcca	ggtcgattga	tcgcaatttc	ttcagtcttc	24660
gggtcatcga	gccattccaa	aatcggettcc	agaagaaagc	gtagttgcgg	atccacttcc	24720
atttacaatg	tatcctatct	ctaagcggaa	atltgaaatc	attaagagcg	gcggttcttc	24780
ccccgcgtgg	cgccgccagt	caggcggagc	tggtaaacac	caaagaaatc	gaggtcccgt	24840
gctacgaaaa	tgaaaacggt	gtcaccttga	ttcttcttca	gggttggcgg	tatgttgatg	24900
gttgccttaa	gggctgtctc	agttgtctgc	tcaccgttat	tttgaaagct	gttgaagctc	24960
atcccgccac	ccgagctgcc	ggcgtaggtg	ctagctgcct	ggaaggcgcc	ttgaacaaca	25020
ctcaagagca	tagctccgct	aaaacgctgc	cagaagtggc	tgtcgaccga	gcccggcaat	25080
cctgagcgac	cgagttcgtc	cgcgcttggc	gatgttaacg	agatcatcgc	atggtcaggt	25140
gtctcggcgc	gatcccacaa	cacaaaaacg	cgcccatctc	cctgttgcaa	gccacgctgt	25200
atltcgccaa	caacggtggt	gccacgatca	agaagcacga	tattgttctg	tgttccacga	25260
atatacctgag	gcaagacaca	ctttacatag	cctgccaaat	ttgtgtcgat	tgcggtttgc	25320
aagatgcacg	gaattattgt	cccttgcggt	accataaaat	cgggggtgcgg	caagagcgtg	25380
gcgctgctgg	gctgcagctc	ggtgggtttc	atacgtatcg	acaaatcggt	ctcgccggac	25440
acttcgccat	tcggcaagga	gttgtcgtca	cgcttgcctt	cttgtcttcg	gcccgtgtcg	25500
ccctgaatgg	cgcgtttgc	gacccttga	tcgccgctgc	tatatgcaaa	aatcggtggt	25560
tcttccggcc	gtggctcatg	ccgctccggt	tcgccctcgc	gcggtagagg	agcagcaggc	25620
tgaacagcct	cttgaaccgc	tggaggatcc	ggcggcacct	caatcggagc	tggaatgaaat	25680
ggcttgggtg	ttgttgcgat	caaagttgac	ggcgatgcgt	tctcattcac	cttcttttgg	25740
cgcccaccta	gccaaatgag	gcttaatgat	aacgcgagaa	cgacacctcc	gacgatcaat	25800
ttctgagacc	ccgaaaagacg	ccggcgatgt	ttgtcggaga	ccagggatcc	agatgcatca	25860
acctcatgtg	ccgcttgcgt	actatcgтта	ttcattcctt	cgcccccttc	aggacgcggt	25920
tcacatcggg	cctcaccgtg	cccgtttgcg	gcctttggcc	aacgggatcg	taagcggtgt	25980
tccagataca	tagtaactgtg	tggccatccc	tcagacgcca	acctcgggaa	accgaagaaa	26040
tctcgacatc	gctcccctta	actgaatagt	tggaacacgc	ttccttgcca	tcaggattga	26100
tggtgtagat	ggagggtatg	cgtacattgc	ccgaaaagtg	gaataccgtc	gtaaatccat	26160
tgctgaagac	ttcgagtggc	aacagcgaac	gatcgccttg	ggcgacgtag	tgccaattac	26220

tgtccgccgc	accaagggct	gtgacaggct	gatccaataa	attctcagct	ttccgttgat	26280
attgtgcttc	cgcgtgtagt	ctgtccacaa	cagcettctg	ttgtgcctcc	cttcgccgag	26340
ccgccgcate	gtcggcgggg	taggcgaatt	ggacgctgta	atagagatcg	ggctgctctt	26400
tatcgagggtg	ggacagagtc	ttggaactta	tactgaaaac	ataacggcgc	atccccgagt	26460
cgcttgccgt	tagcacgatt	actggctgag	gcgtgaggac	ctggcttgcc	ttgaaaaata	26520
gataatttcc	ccgcggtagg	gctgctagat	ctttgctatt	tgaaacggca	accgctgtca	26580
ccgtttcggt	cgtaggcgaat	gttacgacca	aagtagctcc	aaccgccgtc	gagaggcgca	26640
ccacttgatc	gggattgtaa	gccaataaac	gcattgcgcg	atctagcttg	cccgccattg	26700
gagtgctctc	agcctccgca	ccagtcgcag	cggcaataaa	acatgctaaa	atgaaaagtg	26760
cttttctgat	catggttcgc	tgtggcctac	gtttgaaacg	gtatcttccg	atgtctgata	26820
ggaggtgaca	accagacctg	ccgggtttgg	tagtctcaat	ctgccgggca	agctggctac	26880
cttttcgtag	cgaactgtcg	cggctccacgt	actcaccaca	ggcattttgc	cgtaacgac	26940
gagggtcctt	ttatagcgaa	tttgetgctg	gcttgagatt	acatcatttg	aagcgatgtg	27000
ctcgacctcc	accctgccgc	gtttgccaag	aatgacttga	ggcgaactgg	gattgggata	27060
gttgaagaat	tgctggtaat	cctggcgcac	tgttggggca	ctgaagtctg	ataccaggtc	27120
gtaggcgtac	tgagcgggtg	cggcatcata	actctcgcgc	aggcgaactg	actcccacaa	27180
tgaggcgta	acgacggcct	cctcttgagt	tgcaggcaat	cgcgagacag	acacctcgt	27240
gtcaacgggtg	ccgtccggcc	gtatccatag	atatacgggc	acaagcctgc	tcaacggcac	27300
cattgtggct	atagcgaacg	cttgagcaac	atttcccaaa	atcgcgatag	ctgcgacagc	27360
tgcaatgagt	ttggagagac	gtcgcgccga	tttcgctcgc	gcggtttgaa	aggcttctac	27420
ttccttatag	tgctcggcaa	ggctttcgcg	cgccactagc	atggcatatt	caggccccgt	27480
catagcgtcc	accgcaattg	ccgagctgaa	gatctgacgg	agtaggctgc	catcgcccc	27540
cattcagcgg	gaagatcggg	cctttgcagc	tcgtaatgt	gtcgtttgtc	tggcagccgc	27600
tcaaagcgac	aactaggcac	agcaggcaat	acttcataga	attctccatt	gaggcgaatt	27660
tttgcgcgac	ctagcctcgc	tcaacctgag	cgaagcgacg	gtacaagctg	ctggcagatt	27720
gggttgccgc	gctccagtaa	ctgcctccaa	tgttgccggc	gatcgcggc	aaagcgacaa	27780
tgagcgcate	ccctgtcaga	aaaacatat	cgagttcgt	aagaccaatg	atcttggccc	27840
cggtcgtacc	ggcgaaggtg	attacaccaa	gcataagggt	gagcgcagtc	gcttcggtta	27900
ggatgacgat	cgttgccacg	aggtttaaga	ggagaagcaa	gagaccgtag	gtgataagtt	27960
gccccgatcca	cttagctcgc	atgtcccgcg	tgcgatcaaa	aatatatccg	acgaggatca	28020
gaggccccgat	cgcgagaagc	actttcgtga	gaattccaac	ggcgtcgtaa	actccgaagg	28080
cagaccagag	cgtagccgtaa	aggaccact	gtgccccttg	gaaagcaagg	atgtcctggt	28140
cgttcatcgg	accgatttcg	gatgcgattt	tctgaaaaac	ggcctgggtc	acggcgaaca	28200
ttgtatccaa	ctgtgccgga	acagtctgca	gaggcaagcc	ggttacacta	aactgctgaa	28260
caaagtttgg	gaccgtcttt	tcgaagatgg	aaaccacata	gtcttggtag	ttagcctgcc	28320
caacaattag	agcaacaacg	atggtgaccg	tgatcaccgg	agtgataccg	ctacgggtat	28380
cgacttcgcc	gcgatgact	aaaataccct	gaacaataat	ccaaagagtg	acacaggcga	28440
tcaatggcgc	actcaccgcc	tcctggatag	tctcaagcat	cgagtccaag	cctgtcgtga	28500
aggctacatc	gaagatcgta	tgaatggccg	taaacggcgc	cggaatcgtg	aaattcatcg	28560



attggacctg	aacttgactg	gtttgtcgca	taatgttgga	taaaatgagc	tcgcattcgg	28620
cgaggatgcg	ggcggatgaa	caaatcgccc	agccttaggg	gagggcacca	aagatgacag	28680
cggctctttg	atgctccttg	cgttgagcgg	ccgcctcttc	cgcctcgtga	aggccggcct	28740
gcgcggtagt	catcgtaaat	aggcttgtcg	cctgtacatt	ttgaatcatt	gcgtcatgga	28800
tctgcttgag	aagcaaacca	ttggtcacgg	ttgcctgcat	gatattgca	gatcgggaaa	28860
gctgagcaga	cgtatcagca	ttcgccgtca	agcgtttgc	catcgtttcc	agattgtcag	28920
ccgcaatgcc	agcgcgtgtt	gcggaaccgg	tgatctgca	tcgcaacagg	tccgcttcag	28980
catcactacc	cacgactgca	cgatctgtat	cgctgggtgat	cgcacgtgcc	gtggtcgaca	29040
ttggcattcg	cggcgaaaac	atttcattgt	ctaggctcct	cgtcgaagga	tactgatttt	29100
tctggttgag	cgaagtcagt	agtccagtaa	cgccgtaggc	cgacgtcaac	atcgtaacca	29160
tcgctatagt	ctgagtgaga	ttctccgcag	tcgcgagcgc	agtcgcgagc	gtctcagcct	29220
ccgttgccgg	gtcgctaaca	acaaaactgcg	cccgcgcggg	ctgaatata	agaaagctgc	29280
aggtaaaaac	tgttgcaata	agttgcgtcg	tcttcacgt	ttctacctt	atcaatcttc	29340
tgccctgtgg	tgacgggcca	tgaatteget	gagccagcca	gatgagttgc	cttcttgtgc	29400
ctcgcgtagt	cgagttgcaa	agcgcaccgt	gttggeacgc	cccgaagca	cggcgacata	29460
ttcacgcata	tcccgcagat	caaattcgca	gatgacgctt	ccactttctc	gtttaagaag	29520
aaacttacgg	ctgccgaccg	tcatgtcttc	acggatcgcc	tgaattctct	tttcggtaca	29580
tttcagtcca	tcgacataag	ccgatcgatc	tgcggttgg	gatggataga	aaatcttcgt	29640
catacattgc	gcaaccaage	tggetcctag	cggcgattcc	agaacatget	ctggttgctg	29700
cgttgccagt	attagcatcc	cgttgttttt	tcgaacggtc	aggaggaatt	tgtcgacgac	29760
agtcgaaaat	ttagggttta	acaaataggc	gcgaaactca	tcgcagctca	tcacaaaacg	29820
gcggccgctg	atcatggctc	caatccgatg	caggagatat	gctgcagcgg	gagcgcatac	29880
ttcctcgtat	tcgagaagat	gcgtcatgtc	gaagccggta	atcgacggat	ctaactttac	29940
ttcgtcaact	tcgccgtcaa	atgcccagcc	aagcgcattg	ccccggcacc	agcgttgag	30000
ccgcgctcct	gcgccttcgg	cgggccccatg	caacaaaaat	tcacgtaacc	ccgcgattga	30060
acgcatttgt	ggatcaaacg	agagctgacg	atggatacca	cggaccagac	ggcggttctc	30120
ttccggagaa	atcccacccc	gaccatcact	ctcgatgaga	gccacgatcc	attcgcgcag	30180
aaaatcgtgt	gaggctgctg	tgttttctag	gccacgcaac	ggcgccaacc	cgctgggtgt	30240
gcctctgtga	agtgccaaat	atgttctctc	tgtggcgcga	accagcaatt	cgccaccccg	30300
gtccttgtca	aagaacacga	ccgtacctgc	acggctgacc	atgctctgtt	cgagcatggc	30360
tagaacaac	atcatgagcg	tcgtcttacc	cctcccagata	ggcccgaata	ttgccgtcat	30420
gccaacatcg	tgctcatgcg	ggatatagtc	gaaaggcgtt	ccgccattgg	tacgaaatcg	30480
ggcaatcgcg	ttgccccagt	ggcctgagct	ggcgccctct	ggaaagtfff	cgaaagagac	30540
aaaccctgcg	aaattgctgt	aagtgattgc	gccagggcgt	gtgcgccact	taaaattccc	30600
cggcaattgg	gaccaatagg	ccgettccat	accaatacct	tcttgacaa	ccacggcacc	30660
tgcatccgcc	attcgtgtcc	gagcccgcgc	gcccctgtcc	ccaagactat	tgagatcgtc	30720
tgcatagacg	caaaggctca	aatgatgtga	gcccataacg	aattcgttgc	tcgcaagtgc	30780
gtcctcagcc	tcggataaatt	tgccgatttg	agtcacggct	ttatcgccgg	aactcagcat	30840
ctggctcgat	ttgaggctaa	gtttcgcgtg	cgcttgcggg	cgagtcagga	acgaaaaact	30900

ctgcgtgaga	acaagtggaa	aatcgaggga	tagcagcgcg	ttgagcatgc	ccggccgtgt	30960
ttttgcaggg	tattcgcgaa	acgaatagat	ggatccaacg	taactgtctt	ttggcgttct	31020
gatctcgagt	cctcgcttgc	cgcaaatgac	tctgtcggta	taaatcgaag	cgccgagtga	31080
gccgctgacg	accggaaccg	gtgtgaaccg	accagtcatg	atcaaccgta	gcgcttcgcc	31140
aatttcggtg	aagagcacac	cctgtcttct	gcggatgcca	agacgatgca	ggccatacgc	31200
tttaagagag	ccagcgacaa	catgccaaag	atcttccatg	ttcctgatct	ggccccgtgag	31260
atcgttttcc	ctttttccgc	ttagcttggg	gaacctctct	tttaccttcc	ctaaagccgc	31320
ctgtgggtag	acaatcaacg	taaggaagtg	ttcattgcgg	aggagttagc	cggagagcac	31380
gcgctgttca	aaagcttcgt	tcaggctagc	ggcgaaaaca	ctacggaagt	gtcgcggcgc	31440
cgatgatggc	acgtcggcat	gacgtacgag	gtgagcatat	attgacacat	gatcatcagc	31500
gatattgcgc	aacagcgtgt	tgaacgcacg	acaacgcgca	ttgcgcattt	cagtttctct	31560
aagctcgaat	gcaacgccat	caattctcgc	aatggtcatg	atcgatccgt	cttcaagaag	31620
gacgatatgg	tcgctgaggt	ggccaatata	agggagatag	atctcaccgg	atctttcggg	31680
cgttccactc	gcgccgagca	tcacaccatt	cctctcctct	gtgggggaac	cctaattgga	31740
tttgggctaa	cagtagcgcc	ccccaaact	geactataca	tgcttcttcc	cgcggctccgc	31800
aaaaatagca	ggacgacgct	cgccgcattg	tagtctcgct	ccacgatgag	ccgggctgca	31860
aaccataacg	gcacgagaac	gacttcgtag	agcgggttct	gaacgataac	gatgacaaag	31920
ccggcgaaca	tcatgaataa	ccctgccaat	gtcagtggca	ccccaaagaa	caatgcgggc	31980
cgtgtggctg	cgaggtaaag	ggtcgattct	tccaaacgat	cagccatcaa	ctaccgccag	32040
tgagcgtttg	gccgaggaag	ctcgcccaa	acatgataac	aatgccgccg	acgacgccgg	32100
caaccagccc	aagcgaagcc	cgcccgaaca	tccaggagat	cccgatagcg	acaatgccga	32160
gaacagcgag	tgactggccg	aacggaccaa	ggataaacgt	gcataatattg	ttaaccattg	32220
tggcggggtc	agtgccgcca	cccgcagatt	gcgctgcggc	gggtccggat	gaggaaatgc	32280
tccatgcaat	tgcaccgcac	aagcttgggg	cgcagctcga	tatcacgcgc	atcatcgcat	32340
tcgagagcga	gaggcgattt	agatgtaaac	ggtatctctc	aaagcatcgc	atcaatgcgc	32400
acctccttag	tataagtcga	ataagacttg	attgtcgtct	gcggatttgc	cgttgtcctg	32460
gtgtggcggt	ggcggagcga	ttaaaccgcc	agcgcctcct	tcttgcgagc	ggcgtgata	32520
tgacccccaa	acatcccacg	tctcttcgga	ttttagcgcc	tcgtgatcgt	cttttgagg	32580
ctcgattaac	gcgggcacca	gcgattgagc	agctgtttca	acttttcgca	cgtagccggt	32640
tgcaaaaaccg	ccgatgaaat	taccggtggt	gtaageggag	atgccccgac	gaagcgcaaa	32700
ttgcttctcg	tcaatcgttt	cgccgectgc	ataacgaact	ttcagcatgt	ttgcagcggc	32760
agataatgat	gtgcacgcct	ggagcgcacc	gtcaggtgtc	agaccgagca	tagaaaaatt	32820
tcgagagttt	atgtgcatga	ggccaacatc	cagcgaatgc	cgtgcatcga	gacggtgcct	32880
gacgacttgg	gttgcttggc	tgtgatcttg	ccagtgaage	gtttcggcgg	tcgtgttgct	32940
atgaatcgct	aaaggatcaa	agcgaactct	caccttagct	atgcccgcaa	gcgtagatgt	33000
cgcaactgat	ggggcacact	tgcgagcaac	atggteaaac	tcagcagatg	agagtggcgt	33060
ggcaaggctc	gacgaacaga	aggagaccat	caaggcaaga	gaaagcgacc	ccgatctctt	33120
aagcatacct	tatctcctta	gctcgcgaact	aacaccgctt	ctcccgttgg	aagaagtgcg	33180
ttgttttatg	ttgaagatta	tcgggagggt	cggttactcg	aaaattttca	attgcttctt	33240

tatgatttca	attgaagcga	gaaacctcgc	ccggcgtett	ggaacgcaac	atggaccgag	33300
aaccgcgcat	ccatgactaa	gcaaccggat	cgacctatc	aggccgcagt	tggtcaggtc	33360
aggctcagaa	cgaaaatgct	cggcgaggtt	acgctgtctg	taaaccatt	cgatgaacgg	33420
gaagcttccct	tccgattgct	cttggcagga	atattggccc	atgectgctt	gcgctttgca	33480
aatgctctta	tcgcgttgg	atcatatgcc	ttgtccgcca	gcagaaacgc	actctaagcg	33540
attatttgta	aaaatgtttc	ggatcatgcg	cggatcatggg	cttgaccgc	tgctcagcgca	33600
agacggatcg	gtcaaccgtc	ggcatcgaca	acagcgtgaa	tcttggtggt	caaaccgcca	33660
cgggaacgtc	ccatacagcc	atcgtcttga	tcccgtggtt	tcccgtcgcc	gcatgttgg	33720
ggacgcggac	acaggaactg	tcaatcatga	cgacattcta	tcgaaagcct	tggaaatcac	33780
actcagaata	tgatcccaga	cgtctgctc	acgccatcgt	acaaagcgt	tgtagcaggt	33840
tgtacaggaa	ccgtatcgat	caggaacgtc	tgcccagggc	gggcccgtcc	ggaagcgcca	33900
caagatgaca	ttgatcacc	gcgtcaacgc	gcggcacgcg	acgcggctta	tttgggaaca	33960
aaggactgaa	caacagtcca	ttcgaaatcg	gtgacatcaa	agcggggacg	ggttatcagt	34020
ggcctccaag	tcaagcctca	atgaatcaaa	atcagaccga	tttgcaaacc	tgatttatga	34080
gtgtgcggcc	taaatgatga	aatcgtcctt	ctagatcgcc	tccgtggtgt	agcaacacct	34140
cgcagtatcg	ccgtgctgac	cttggccagg	gaattgactg	gcaagggtgc	ttccatga	34200
ccgctctttt	ggccgcgata	gatgatttcg	ttgctgcttt	gggcacgtag	aaggagagaa	34260
gtcatatcgg	agaaattcct	cctggcgcga	gagcctgctc	tatcgcgacg	gcateccact	34320
gtcgggaaca	gaccggatca	ttcacgaggc	gaaagtcgctc	aacacatgcg	ttataggcat	34380
cttcccttga	aggatgatct	tgttctgctc	aatctggagg	tgcggcagcc	gcaggcagat	34440
gcgatctcag	cgcaacttgc	ggcaaaacat	ctcactcacc	tgaaaaccac	tagcgagtct	34500
cgcgatcaga	cgaaggcctt	ttacttaacg	acacaatate	cgatgtctgc	atcacaggcg	34560
tcgctatccc	agtcaatact	aaagcgggtgc	aggaactaaa	gattactgat	gacttaggcg	34620
tgccacgagg	cctgagacga	cgcgcgtaga	cagttttttg	aatcattat	caaagtgatg	34680
gcctccgctg	aagcctatca	cctctgcgcc	ggtctgtcgg	agagatgggc	aagcattatt	34740
acggctcttcg	cgcccgtaca	tgcatggac	gattgcaggg	tcaatggatc	tgagatcatc	34800
cagaggattg	ccgcccttac	cttccgtttc	gagttggagc	cagcccctaa	atgagacgac	34860
atagtcgact	tgatgtgaca	atgccaagag	agagatttgc	ttaaccgat	ttttttgctc	34920
aagcgtaagc	ctattgaagc	ttgccggcat	gacgtccgcg	ccgaaagaat	atcctacaag	34980
taaaacattc	tgcacaccga	aatgcttgg	gtagacatcg	attatgtgac	caagatcctt	35040
agcagtttcg	cttggggacc	gctccgacca	gaaataccga	agtgaactga	cgccaatgac	35100
aggaatccct	tccgtctgca	gataggtacc	atcgatagat	ctgctgctc	gcgcttttcg	35160
gtgatgacgg	tgaaaacctc	tgacacatgc	agctcccgga	gacggtcaca	gcttgtctgt	35220
aagcggatgc	cgggagcaga	caagcccgtc	agggcgcgctc	agcgggtggt	ggcgggtgctc	35280
ggggcgcagc	catgaccag	tcacgtagcg	atagcggagt	gtatactggc	ttaactatgc	35340
ggcatcagag	cagattgtac	tgagagtgea	ccatatgcgg	tgtgaaatac	cgcacagatg	35400
cgtaaggaga	aaataccgca	tcaggcgtc	ttccgcttc	tcgctcactg	actcgtcgcg	35460
ctcggtcgtt	cggctgcggc	gagcggtatc	agctcactca	aaggcggtaa	tacggttate	35520
cacagaatca	ggggataacg	caggaaagaa	catgtgagca	aaaggccagc	aaaaggccag	35580

gaaccgtaaa	aaggccgcgt	tgctggcggt	tttccatagg	ctccgcccc	ctgacgagca	35640
tcacaaaaat	cgacgctcaa	gtcagagggt	gcgaaacccg	acaggactat	aaagatacca	35700
ggcgtttccc	cctggaagct	ccctcgtgcg	ctctcctggt	ccgaccctgc	cgcttaccgg	35760
atacctgtcc	gcctttctcc	cttcgggaag	cgtggcgctt	tctcatagct	cacgctgtag	35820
gtatctcagt	tcgggttagg	tcgttcgctc	caagctgggc	tgtgtgcacg	aacccccgt	35880
tcagccccgac	cgctgcgcct	tatccggtaa	ctatcgtctt	gagtccaacc	cggtaaagaca	35940
cgacttatcg	ccactggcag	cagccactgg	taacaggatt	agcagagcga	ggtatgtagg	36000
cggtgctaca	gagttcttga	agtgggtggc	taactacggc	tacactagaa	ggacagtatt	36060
tggtatctgc	gctctgctga	agccagttac	cttcggaaaa	agagttggta	gctcttgatc	36120
cggcaaaaa	accaccgctg	gtagcgggtg	tttttttgtt	tgcaagcagc	agattacgcg	36180
cagaaaaaaaa	ggatctcaag	aagatccttt	gatcttttct	acggggctctg	acgctcagtg	36240
gaacgaaaa	tcacgttaag	ggattttgg	catgagatta	tcaaaaagga	tcttcaccta	36300
gatcctttta	aattaaat	gaagtttta	atcaatctaa	agtatatatg	agtaaacttg	36360
gtctgacagt	taccaatgct	taatcagtga	ggcacctatc	tcagcgatct	gtctatttctg	36420
ttcatccata	gttgectgac	tccccgtcgt	gtagataact	acgatacggg	agggcttacc	36480
atctggcccc	agtgctgcaa	tgataccgcg	agaccacgc	tcaccggctc	cagatttatac	36540
agcaataaac	cagccagccg	gaagggccga	gcgcagaagt	ggtcctgcaa	ctttatccgc	36600
ctccatccag	tctattaatt	gttgccggga	agctagagta	agtagttcgc	cagttaatag	36660
tttgcgcaac	gttgttgcca	ttgctgcagg	gggggggggg	gggggggttcc	attgttcatt	36720
ccacggacaa	aaacagagaa	aggaaacgac	agaggccaaa	aagctcgctt	tcagcacctg	36780
tcgtttcctt	tcttttcaga	gggtatttta	aataaaaaca	ttaagttatg	acgaagaaga	36840
acggaaacgc	cttaaaccgg	aaaattttca	taaatacgca	aaaccgcga	ggtcgccgcc	36900
ccgtaacctg	tcggatcacc	ggaaaggacc	cgtaaagtga	taatgattat	catctacata	36960
tcacaacgtg	cgtggaggcc	atcaaaccac	gtcaaataat	caattatgac	gcaggtatcg	37020
tattaattga	tctgcatcaa	cttaacgtaa	aaacaacttc	agacaataca	aatcagcgac	37080
actgaatacg	gggcaacctc	atgtcccc	cccccccc	cctgcaggca	tcgtggtgtc	37140
acgctcgtcg	tttggtatgg	cttcattcag	ctccggttcc	caacgatcaa	ggcgagttac	37200
atgatcccc	atgttgtgca	aaaaagcgg	tagctcttc	ggtcctccga	tcgttgtcag	37260
aagtaagttg	gccgcagtgt	tatactcat	ggttatggca	gactgcata	attctcttac	37320
tgtcatgcca	tccgtaagat	gctttctgt	gactggtgag	tactcaacca	agtcattctg	37380
agaatagttg	atgcggcgac	cgagttgctc	ttgcccggcg	tcaacacggg	ataataccgc	37440
gccacatagc	agaacttta	aagtgetcat	cattggaaaa	cgttcttcgg	ggcgaaaact	37500
ctcaaggatc	ttaccgctgt	tgagatccag	ttcgatgtaa	cccactcgtg	cacccaactg	37560
atcttcagca	tcttttactt	tcaccagcgt	ttctgggtga	gcaaaaacag	gaaggcaaaa	37620
tgccgcaaaa	aagggaaata	gggcgacacg	gaaatgttga	atactcatac	tcttctttt	37680
tcaatattat	tgaagcattt	atcagggtta	ttgtctcatg	agcggataca	tatttgaatg	37740
tatttagaaa	aataaaciaa	taggggttcc	gcgcacattt	ccccgaaaag	tgccacctga	37800
cgtctaagaa	accattatta	tcatgacatt	aacctataaa	aataggcgta	tcacgaggcc	37860
ctttcgtctt	caagaattgg	tcgacgatct	tgctgcgttc	ggatatttcc	gtggagttcc	37920

cgccacagac	cggattgaa	ggcgagatcc	agcaactcgc	gccagatcat	cctgtgacgg	37980
aactttggcg	cgtgatgact	ggccaggacg	tcggccgaaa	gagcgacaag	cagatcacgc	38040
ttttcgacag	cgtcggattt	gcgatcgagg	atttttcggc	gctgcgctac	gtccgcgacc	38100
gcgttgaggg	atcaagccac	agcagcccac	tcgaccttct	agccgacceca	gacgagccaa	38160
gggatctttt	tggaatgctg	ctccgtcgtc	aggctttccg	acgtttgggt	ggttgaacag	38220
aagtcaattat	cgtacggaat	gccaagcact	cccgagggga	accctgtggt	tggcatgcac	38280
atacaaatgg	acgaacggat	aaaccttttc	acgccctttt	aaatatccgt	tattctaata	38340
aacgctcttt	tctcttaggt	ttaccgcca	atatactctg	tcaaactctg	atagtttaaa	38400
ctgaaggcgg	gaaacgacaa	tctgatcatg	agcggagaat	taagggagtc	acgttatgac	38460
ccccgccgat	gacgcgggac	aagccgtttt	acgtttggaa	ctgacagAAC	cgcaacgttg	38520
aaggagccac	tcagcccAag	ctggtacgat	tgtAatacga	ctactatag	ggcgaattga	38580
gcgctgttta	aacgctcttc	aaCTggaaga	gcggttacca	gaggccagaa	tggccatctc	38640
ggaccgatat	cgctatcaac	tttgtataga	aaagttgggc	cgaattcgag	ctcggtagcg	38700
ccagaatggc	ccgaccggg	ttaccgaatt	cgagctcggT	accctgggat	ccgatatcga	38760
tgggccctgg	ccgaagcttg	catgcctgea	gtgcagcgtg	accgggtcgt	gcccctctct	38820
agagataatg	agcattgcat	gtctAagtta	tAAAAatta	ccacatattt	ttttgtcac	38880
acttgtttga	agtgcagttt	atctatcttt	atacatatat	ttaaacttta	ctctacgaat	38940
aatataatct	atagtactac	aataatatca	gtgtttttaga	gaatcatata	aatgaacagt	39000
tagacatggt	ctaaaggaca	attgagtatt	ttgacaacag	gactctacag	ttttatcttt	39060
ttagtgtgca	tgtgttctcc	tttttttttg	caaatagctt	cacctatata	atacttcate	39120
cattttatta	gtacatccat	ttaggttita	gggttaatgg	tttttataga	ctaatttttt	39180
tagtacatct	attttattct	attttagcct	ctaaattaag	aaaactaaa	ctctatttta	39240
gtttttttat	ttaataatTT	agataataaaa	tagaataaaa	taaagtgact	aaaaattaaa	39300
caaataccct	ttaagaaatt	aaaaaaacta	aggaaacatt	ttctttgttt	cgagtagata	39360
atgccagcct	gttaaacgcc	gtcgacgagt	ctaacggaca	ccaaccagcg	aaccagcagc	39420
gtcgcgtcgg	gccaagcgaA	gcagacggca	cggcatctct	gtcgcTgcct	ctggaccctt	39480
ctcgagagtt	ccgctccacc	gttggacttg	ctccgctgtc	ggcatccaga	aattgcgtgg	39540
cggagcggca	gacgtgagcc	ggcacggcag	gcggcctcct	cctcctctca	cggcaccggc	39600
agctacgggg	gattcctttc	ccaccgtctc	ttegettctc	cttctctgcc	cgccgtaata	39660
aatagacacc	ccctccacac	cctctttccc	caacctcgtg	ttgttcggag	cgcacacaca	39720
cacaaccaga	tctcccccaa	atccaccctg	cggcaectcc	gcttcaaggt	acgccgctcg	39780
tctccccccc	ccccctctc	tacctctct	agatcggcgt	tccggTccat	gcattggttag	39840
ggcccggtag	ttctacttct	gttcatgttt	gtgttagatc	cgtgtttgtg	ttagatccgt	39900
gctgctagcg	ttcgtacacg	gatgcgacct	gtacgtcaga	cacgttctga	ttgctaactt	39960
gccagtgttt	ctctttgggg	aatcctggga	tggetctage	cgttccgcag	acgggatcga	40020
tttcatgatt	ttttttgttt	cgttgcatag	ggtttggttt	gcccttttcc	tttatttcaa	40080
tatatgccgt	gcacttgttt	gtcgggtcat	cttttcatgc	ttttttttgt	cttggttgtg	40140
atgatgtggt	ctggttgggc	ggtcgttcta	gatcggagta	gaattctgtt	tcaaactacc	40200
tggTggattt	attaattttg	gatctgtatg	tgtgtgceat	acatattcat	agttacgaat	40260

tgaagatgat	ggatggaaat	atcgatctag	gataggtata	catgttgatg	cgggttttac	40320
tgatgcatat	acagagatgc	ttttgttcg	cttggttgtg	atgatgtggt	gtggttgggc	40380
ggtcgttcat	tcgttctaga	tcggagtaga	atactgtttc	aaactacctg	gtgtatttat	40440
taattttggg	actgtatgtg	tgtgtcatac	atcttcatag	ttacgagttt	aagatggatg	40500
gaaatatcga	tctaggatag	gtatacatgt	tgatgtgggt	tttactgatg	catatacatg	40560
atggcatatg	cagcatctat	tcatatgctc	taacctgag	tacctatcta	ttataataaa	40620
caagtatggt	ttataattat	tttgatcttg	atatacttgg	atgatggcat	atgcagcagc	40680
tatatgtgga	tttttttagc	cctgccttca	tacgtatatt	atttgcttgg	tactgtttct	40740
tttgtcgatg	ctcaccctgt	tgtttgggtg	tacttctgca	ggtcgactct	agaggatcca	40800
tggcaccgaa	gaagaagcgc	aaggtgatgg	acaagaagta	cagcatcggc	ctcgacatcg	40860
gcaccaactc	ggtagggctg	gccgtcatca	cggacgaata	taaggteccg	tcgaagaagt	40920
tcaaggtcct	cggcaataca	gaccgcccaca	gcatacaaga	aaacttgatc	ggcgccctcc	40980
tgttcgatag	cggcgagacc	gcggaggcga	ccaggetcaa	gaggaccgcc	aggagacggt	41040
acactaggcg	caagaacagg	atctgctacc	tgcaggagat	cttcagcaac	gagatggcga	41100
aggtggacga	ctccttcttc	caccgcttgg	aggaatcatt	cctggtggag	gaggacaaga	41160
agcatgagcg	gcacccaatc	ttcggcaaca	tcgtcgacga	ggtaagtttc	tgcttctacc	41220
tttgatata	atataataat	tatcattaat	tagtagtaat	ataatatttc	aaatattttt	41280
ttcaaaataa	aagaatgtag	tatatagcaa	ttgcttttct	gtagtttata	agtgtgtata	41340
ttttaattta	taacttttct	aatataatgac	caaaacatgg	tgatgtgcag	gtggcctacc	41400
acgagaagta	cccgacaatc	taccacctcc	ggaagaaact	ggtggacagc	acagacaagg	41460
cggacctccg	gctcatctac	cttgcctctg	cgcatatgat	caagtteccg	ggccacttcc	41520
tcatcgaggg	cgacctgaac	ccggacaact	ccgacgtgga	caagctgttc	atccagctcg	41580
tgcagacgta	caatcaactg	ttcgaggaga	acccataaa	cgctagcggc	gtggacgcca	41640
aggccatcct	ctcggccagg	ctctcgaaat	caagaaggct	ggagaacctt	atcgcgcagt	41700
tgccaggcga	aaagaagaac	ggcctcttctg	gcaaccttat	tgcgctcagc	ctcggcctga	41760
cgccgaactt	caaatcaaac	ttcgacctcg	cggaggacgc	caagctccag	ctctcaaagg	41820
acacctacga	cgacgacctc	gacaacctcc	tggcccagat	aggagaccag	tacgcgacc	41880
tcttctctgc	cgccaagaac	ctctccgacg	ctatctgct	cagcgacatc	cttcgggtca	41940
acaccgaaat	taccaaggca	ccgctgtccg	ccagcatgat	taaacgctac	gacgagcacc	42000
atcaggacct	cacgctgctc	aaggcactcg	tccgccagca	gctccccgag	aagtacaagg	42060
agatcttctt	cgaccaatca	aaaaacggct	acgcgggata	tatcgacggc	ggtgccagcc	42120
aggaagagtt	ctacaagtte	atcaaaccac	tcttgagaa	gatggacggc	accgaggagt	42180
tgctggtaaa	gctcaacagg	gaggacctcc	tcaggaagca	gaggaccttc	gacaacggct	42240
ccatcccgca	tcagatccac	ctgggcgaac	tgcatgceat	cctgcggcgc	caggaggact	42300
tctaccggtt	cctgaaggat	aaccgggaga	agatcgagaa	gatcttgacg	ttccgcatcc	42360
catactacgt	gggcccctg	gctcgcggea	actcccgggt	cgcctggatg	accgggaagt	42420
cggaggagac	catcacacc	tggaaactttg	aggaggtggt	cgataagggc	gctagcgtc	42480
agagcttcat	cgagcgcag	accaacttcg	ataaaaacct	gcccatagaa	aaagtcctcc	42540
ccaagcactc	gctgctctac	gagtacttca	ccgtgtacaa	cgagctcacc	aaggtcaaat	42600

acgtcaccga	gggcatgcgg	aagccggcgt	tcctgagcgg	cgagcagaag	aaggcgatag	42660
tggacctcct	cttcaagacc	aacaggaagg	tgacctgaa	gcaattaaaa	gaggactact	42720
tcaagaaaat	agagtgcctc	gactccgtgg	agatctcggg	cgtggaggat	cggttcaacg	42780
cctcactcgg	cacgtatcac	gacctcctca	agatcattaa	agacaaggac	ttcctcgaca	42840
acgaggagaa	cgaggacatc	ctcgaggaca	tcgtcctcac	cctgaccctg	ttcgaggacc	42900
gcgaaatgat	cgaggagagg	ctgaagacct	acgcgcacct	gttcgacgac	aaggctatga	42960
aacagctcaa	gaggcgccgc	tacactgggt	ggggaaggct	gtcccgaag	ctcattaatg	43020
gcatcagggg	caagcagagc	ggcaagacca	tctggactt	cctcaagtcc	gacgggttcg	43080
ccaaccgcaa	cttcatgcag	ctcattcacg	acgactcgtc	cacgttcaag	gaagacatcc	43140
agaaggcaca	ggtgagcggg	cagggtgact	cctccacga	acacatgcc	aacctggccg	43200
gctcgccggc	cattaaaaag	ggcatcctgc	agacgggtaa	ggtcgtcgac	gagctcgtga	43260
aggtgatggg	ccggcacaag	cccghaaaata	tcgtcataga	gatggccagg	gagaaccaga	43320
ccacccaaaa	agggcagaag	aactcgcgcg	agcggatgaa	acggatcgag	gagggcatta	43380
aagagctcgg	gtcccagatc	ctgaaggagc	accccgtgga	aaatacccag	ctccagaatg	43440
aaaagctcta	cctctactac	ctgcagaacg	gccgcgacat	gtacgtggac	caggagctgg	43500
acattaatcg	gctatcggac	tacgacgtcg	accacatcgt	gccgcagtcg	ttcctcaagg	43560
acgatagcat	cgacaacaag	gtgctcacc	ggtcggataa	aaatcggggc	aagagcgaca	43620
acgtgcccag	cgaggaggtc	gtgaagaaga	tgaaaaacta	ctggcgccag	ctcctcaacg	43680
cgaaactgat	caccagcgc	aagttcgaca	acctgacgaa	ggcggaacgc	ggtggcttga	43740
gcgaactcga	taaggcgggc	ttcataaaaa	ggcagctggt	cgagacgcgc	cagatcacga	43800
agcatgtcgc	ccagatcctg	gacagccgca	tgaataacta	gtacgatgaa	aacgacaagc	43860
tgatccggga	ggtgaagggt	atcacgtgta	agtccaagct	cgtgtcggac	ttccgcaagg	43920
acttccagtt	ctacaaggtc	cgcgagatca	acaactacca	ccacgccccac	gacgcctacc	43980
tgaatgcggg	ggtcgggacc	gccctgatca	agaagtacc	gaagctggag	tcggagttcg	44040
tgtacggcga	ctacaaggtc	tacgacgtgc	gcaaaatgat	cgccaagtcc	gagcaggaga	44100
tcggcaaggc	cacggcaaaa	tacttcttct	actcgaacat	catgaacttc	ttcaagaccg	44160
agatcaccct	cgcaaacggc	gagatccgca	agcgcgccgt	catcgaaacc	aacggcgaga	44220
cgggcgagat	cgtctgggat	aagggccggg	atttcgcgac	ggtccgcaag	gtgctctcca	44280
tgccgcaagt	caatatcgtg	aaaaagacgg	aggtccagac	ggcggggttc	agcaaggagt	44340
ccatcctccc	gaagcgcaac	tccgacaagc	tcatecgcgag	gaagaaggat	tgggaccgga	44400
aaaaatatgg	cggtctcgac	agccccaccg	tcgcatacag	cgtcctcgtc	gtggcgaagg	44460
tggagaaggg	caagtcaaaag	aagctcaagt	ccgtgaagga	gctgctcggg	atcacgatta	44520
tggagcggtc	ctccttcgag	aagaaccgca	tcgacttctc	agaggccaag	ggatataagg	44580
aggtcaagaa	ggacctgatt	attaaactgc	cgaagtactc	gctcttcgag	ctggaaaacg	44640
gccgcaagag	gatgctcgcc	tccgcagggc	agttgcagaa	gggcaacgag	ctcgcctcc	44700
cgagcaaata	cgtaatttc	ctgtacctcg	ctagccacta	tgaaaagctc	aagggcagcc	44760
cggaggacaa	cgagcagaag	cagctcttcg	tggagcagca	caagcattac	ctggacgaga	44820
tcatecgagca	gatcagcgag	ttctcgaagc	gggtgatcct	cgccgacgcg	aacctggaca	44880
aggtgctgtc	ggcatataac	aagcaccgcg	acaaaccaat	acgcgagcag	gccghaaaata	44940

tcateccact	cttcaccctc	accaacctcg	gcgctccggc	agcettcaag	tacttcgaca	45000
ccacgattga	ccggaagcgg	tacacgagca	cgaaggaggt	gctcgatgcg	acgctgatcc	45060
accagagcat	cacagggctc	tatgaaacac	gcatcgacct	gagccagctg	ggcggagaca	45120
agagaccacg	ggaccgccac	gatggcgagc	tgggaggccg	caagcgggca	aggtaggtac	45180
cgttaaccta	gacttgtcca	tcttctggat	tggccaactt	aattaatgta	tgaataaaaa	45240
ggatgcacac	atagtgacat	gctaactact	ataatgtggg	catcaaagtt	gtgtgttatg	45300
tgtaattact	agttatctga	ataaaagaga	aagagatcat	ccatatttct	tatcctaaat	45360
gaatgtcacg	tgtctttata	attctttgat	gaaccagatg	catttcatta	accaaatcca	45420
tatacatata	aatattaate	atatataatt	aatatcaatt	gggttagcaa	aacaaatcta	45480
gtctaggtgt	gttttgcgaa	tgcggccgcc	accgcggtgg	agctcgaatt	cgagctcggg	45540
accctgggat	ccagcttcgc	ttagttttta	gtttttggca	gaaaaaatga	tcaatgtttc	45600
acaaacaaa	tatttttata	acttttgatg	aaagaagatc	accacggtea	tatctagggg	45660
tggtaacaaa	ttgcgatcta	aatgttttct	cataaaaaat	aaggcttctt	aataaatttt	45720
agttcaaaa	aaatacgaat	aaagtctgat	tetaatctga	ttcgatcctt	aaattttata	45780
atgcaaaa	tagagctcat	taccacctct	agtcatatgt	ctagtctgag	gtatatccaa	45840
aaagcccttt	ctctaaatc	cacacceaac	tcagatgttt	gcaaataaat	actccgactc	45900
caaaatgtag	gtgaagtgca	actttctcca	ttttatatca	acatttgttt	ttttttgttt	45960
aacatttcac	actcaaaa	aattaataaa	atacgtggtt	gttgaacgtg	cgcacatgtc	46020
tccttacat	tatgtttttt	tatttatgta	ttattgtttg	tttctctcca	acaacttgct	46080
aacatatcat	cattggctct	taatatttat	gaatatggaa	gcctagtatt	ttacacttgg	46140
ctacacacta	gtttagtatt	tgccacttgt	ctaactgca	actctagtag	ttttgccact	46200
tgcttgccat	gcaactctag	tattgacact	tgtatagcat	ataatgcaa	tacgacacct	46260
gccttacatg	aaacattatt	tttgacactt	gtataccatg	caacattacc	attgacattt	46320
gtccatacac	attatatcaa	atatattgag	cgcatgtcac	aaactcgata	caaagctgga	46380
tgaccctccc	tcaccacatc	tataaaaacc	cgagcgctac	tgtaaatac	tcacaacaca	46440
acacatatct	tttagtaacc	tttcaatagg	cgccccca	gaactagtaa	ccatggccct	46500
gtccaacaag	ttcatcggcg	acgacatgaa	gatgacctac	cacatggacg	gctgcgtgaa	46560
cggccactac	ttcaccgtga	agggcgaggg	cagcggcaag	ccctacgagg	gcaccagac	46620
ctccaccttc	aaggtgacca	tggccaacgg	cgccccctg	gccttctct	tcgacatcct	46680
gtccaccgtg	ttcatgtacg	gcaaccgctg	cttaccgcc	taccceacca	gcatgcccgga	46740
ctacttcaag	caggccttcc	ccgacggcat	gtctacgag	agaaccttea	cctacgagga	46800
cggcggcgtg	gccaccgcca	gctgggagat	cagcctgaag	ggcaactgct	tcgagcacia	46860
gtccaccttc	cacggcgtga	acttccccgc	cgacggcccc	gtgatggcca	agaagaccac	46920
cggctgggac	ccctccttcg	agaagatgac	cgtgtgcgac	ggcatcttga	agggcgacgt	46980
gaccgccttc	ctgatgctgc	agggcgggcg	caactacaga	tgccagttec	acacctcta	47040
caagaccaag	aagcccgtga	ccatgcccc	caaccacgtg	gtggagcacc	gcatgcccg	47100
aaccgacctg	gacaagggcg	gcaacagcgt	gcagctgacc	gagcacgccg	tggccccat	47160
cacctccgtg	gtgcccttct	gaagcggccc	atggatatte	gaacgcgtag	gtaccacatg	47220
gttaacctag	acttgtccat	cttctggatt	ggccaactta	attaatgtat	gaaataaaag	47280



gatgcacaca	tagtgacatg	ctaataccta	taatgtgggc	atcaaagttg	tgtgttatgt	47340
gtaattacta	gttatctgaa	taaaagagaa	agagatcatc	catatttctt	atcctaaatg	47400
aatgtcacgt	gtctttataa	ttctttgatg	aaccagatgc	atttcattaa	ccaaatccat	47460
atacatataa	atattaatca	tatataatta	atatcaattg	ggttagcaaa	acaaatctag	47520
tctaggtgtg	ttttgcgaat	tcccatggac	ctcgaggggg	ggccccggca	cccagctttc	47580
ttgtacaaag	tggccgtaa	cggatcggcc	agaatggccc	ggaccgggtt	accgaattcg	47640
agctcggtac	cctgggatcg	gccgctctag	aactagtgga	tccccgggc	tgcaggaatt	47700
cccatggagt	caaagattca	aatagaggac	ctaacagaac	tcgccgtaa	gactggcgaa	47760
cagttcatac	agagtctctt	acgactcaat	gacaagaaga	aatcttctg	caacatgggtg	47820
gagcacgaca	cgcttgtcta	ctccaaaaat	atcaaagata	cagtctcaga	agaccaaagg	47880
gcaattgaga	cttttcaaca	aagggttaata	tccggaaacc	tectcggatt	ccattgceca	47940
gctatctgtc	actttattgt	gaagatagtg	gaaaaggaag	gtggctccta	caaatgceat	48000
cattgcgata	aaggaaaggc	catcgttgaa	gatgcctctg	ccgacagtgg	tcccaaagat	48060
ggacccccac	ccacgaggag	catcgtggaa	aaagaagacg	ttccaaccac	gtcttcaag	48120
caagtggatt	gatgtgatat	ctccactgac	gtaagggatg	acgcacaatc	ccactaagct	48180
tcggccgggg	cccatcgatc	tggcgaaag	gggatgtgct	gcaaggcgat	taagttgggt	48240
aacgccaggg	ttttcccagt	cacgacgtt	taaaacgacg	gccagtgcca	agctcagatc	48300
agcttggggc	tggatcgat	aaatgtttcc	acatagat	tgcatatcat	aatgatgttt	48360
gtcgttccgt	atctatgttt	catacaaaat	ttttacgcat	atcgcaacac	atgggcat	48420
acctagtac	tgtataactc	tgcattgatg	agtgtatgac	tatatgatgt	agtaactaat	48480
aagaagggta	gacatttgag	tgattctttt	attcctggac	ttgtaagact	tgacatttct	48540
gccttgagt	cgatacatca	tatggacagg	ggttatgcat	acactgcttg	tttgttgttt	48600
atgttctaag	agcatctcca	acaacgtgac	atatgaaaat	gccctacaat	ttaaaaatgg	48660
ttatatttta	taaaatttag	ggcataaata	aaacatccc	ctccaacatt	aaagccttaa	48720
atctattata	gggaagccca	ctatgatata	gtatatttga	ggcacttttag	agggtgcctt	48780
ataatttttt	gaccattttt	ttatgaaatg	agacactatt	ggagtatttt	ttttccgtag	48840
agcaccatat	ttcaatttga	gacaccaatt	taaggcattg	ttggagatgt	tctaaatgtt	48900
ggtttatttt	gtctgtatcg	ttgtggtttt	gatagtgggt	cctttgcaat	gtacatctta	48960
cattgacaat	aataataggt	aaaactctac	aaatttttta	tctaatggac	tcttgtatga	49020
aacattgtac	ttgcacacat	ctgatgtaaa	cactgcatac	ttttaacagt	gacaagattc	49080
tgtttcattt	tagggctagt	ttgggaacca	aattttatta	gggtttttat	tttctaagaa	49140
aaagtaattt	attttacctt	gagaaaatat	aaattacttg	agaaaataga	gttccaaact	49200
agctcttata	ttgtcgaat	cctcctctat	tcaaatgtga	catttctggc	acgtgacaac	49260
tggatgatgt	gtagactgtg	ttaagtaata	cgtgtcatta	ttactaaatg	ccatttttagt	49320
aaatgttgag	tatgtactct	actacagtaa	gtattattgg	tgtatttaca	ctagacagtt	49380
ggcggcctgg	cgggttaaagt	tatcctgtag	aaagtggggc	caggccaaaa	ccaaccgcca	49440
aaggaaaggc	cttccggccc	gcccaccttt	gcgcgccgaa	ggtcagttec	ttcagtctcc	49500
tcccgttca	gactctgacc	acgtcgacaa	tccgggccga	aacacatctg	caccgtccac	49560
ttgcgacaga	ttgaacacac	cacttctatc	cacgtcagcg	atccgtggca	ctagcccttc	49620

caccaatcag	cccaagttgc	ccctttcett	taaattegcc	gcaccattg	cttttctcac	49680
ggccatagaa	atcgaccgag	cgaatccctc	gcatcgcatt	cgcagccttt	gctgcatcac	49740
accaccgcga	aaccccagca	gccgcatctg	caggtcgact	ctagaggatc	catggcctcc	49800
tccgaggacg	tcatcaagga	gttcatgcgc	ttcaaggtgc	gcatggaggg	ctccgtgaac	49860
ggccacgagt	tcgagatcga	gggcgagggc	gagggccgcc	cctacgaggg	cacccagacc	49920
gccaaactga	aggtgaccaa	gggcggcccc	ctgcccttcg	cctgggacat	cctgtccccc	49980
cagttccagt	acggctccaa	ggtgtactgt	aagcaccccc	ccgacatccc	cgactacaag	50040
aagctgtcct	tccccgaggg	cttcaagtgg	gagcgcgtga	tgaacttcga	ggacggcggc	50100
gtgggtgacag	tgaccagga	ctcctccctg	caggacggct	ccttcattct	caaggtgaag	50160
ttcatcggcg	tgaacttccc	ctccgacggc	ccgtaatgc	agaagaagac	tatgggctgg	50220
gaggcctcca	ccgagcgctt	gtacccccgc	gacggcgtgc	tgaagggcga	gatccacaag	50280
gccctgaagc	tgaaggacgg	cggccacgct	agcccatcca	cccactact	cactcatact	50340
tgtgtgttac	gtacgagaat	ttctcgacca	accgtcgtga	gacctgcca	ccggagatcg	50400
gacgcaagag	ggtttaggca	agaatgtcgt	gcgacaggg	gagcgcgtac	tagtatacgt	50460
gagagacctt	gagatatacc	tcacacgtac	gcgtacttta	catgacgtag	gacattacga	50520
ctcaaacaga	ttcacgtcag	atttcggagt	ttctcacgcg	tgagagcctt	ggagggcgg	50580
atgtatgtca	tactatatgt	tgggatggag	ggagtgagt	agtgatatgt	ggctagcaag	50640
ggcggccccc	tgcccttcgc	ctgggacatc	ctgtccccc	agttccagta	cggctccaag	50700
gtgtactgta	agcacccccg	cgacatcccc	gactacaaga	agctgtcctt	ccccgagggc	50760
ttcaagtggg	agcgcgtgat	gaacttcgag	gacggcggcg	tggtagacgt	gaccagggac	50820
tctccttcgc	aggacggctc	cttcatctac	aaggtgaagt	tcateggcgt	gaacttcccc	50880
tccgacggcc	ccgtaatgca	gaagaagact	atgggctggg	aggcctccac	cgagcgcctg	50940
tacccccgcg	acggcgtgct	gaagggcgag	atccacaagg	ccctgaagct	gaaggacggc	51000
ggccactacc	tggtaggagtt	caagtccatc	tacatggcca	agaagcccgt	gcagctgccc	51060
ggctactact	acgtggactc	caagctggac	atcacctccc	acaacgagga	ctacaccatc	51120
gtggagcagt	acgagcgcgc	cgagggccgc	caccacctgt	tcctgtagtc	aggatctgag	51180
tcgaaacct	gacttgtcca	tcttctggat	tggccaactt	aattaatgta	tgaataaaaa	51240
ggatgcacac	atagtgacat	gctaatcact	ataatgtggg	catcaaagtt	gtgtgttatg	51300
tgttaattact	agttatctga	ataaaaagaga	aagagatcat	ccatatttct	tatcctaaat	51360
gaatgtcacg	tgtctttata	attctttgat	gaaccagatg	catttcatta	accaaataca	51420
tatacatata	aatattaate	atatataatt	aatatacaatt	gggttagcaa	aacaaatcta	51480
gtctaggtgt	gttttgcgaa	tgcggccgcc	accgcggtgg	agctcgaatt	ccggctccggg	51540
tcaccgggtc	cgggcctaga	aggccagctt	gcggccgccc	cgggcaactt	tattatacaa	51600
agttgataga	tatcgggtccg	agcggcctag	aaggcctttg	gtcacctttg	tcaccaaga	51660
tggaaactgcg	gccgctcatt	aattaagtca	ggcgcgcctc	tagttgaaga	cacgttcatg	51720
tcttcatcgt	aagaagacac	tcagtagtet	tcggccagaa	tggcctaact	caaggccatc	51780
gtggcctctt	gctcttcagg	atgaagagct	atgtttaaac	gtgcaagcgc	tactagacaa	51840
ttcagtacat	taaaaacgtc	cgcaatgtgt	tattaagttg	tctaagcgtc	aatttgttta	51900
caccacaata	tatcctgcca	ccagccagcc	aacagctccc	cgaccggcag	ctcggcacia	51960

aatcaccact	cgatacaggc	agcccatcag	tccgggacgg	cgtcagcggg	agagccgttg	52020
taaggcggca	gactttgctc	atgttaccga	tgctattcgg	aagaacggca	actaagctgc	52080
cgggtttgaa	acacggatga	tctcgcggag	ggtagcatgt	tgattgtaac	gatgacagag	52140
cgttgctgcc	tgtgatcaaa	tatcatctcc	ctcgcagaga	tccgaattat	cagccttctt	52200
attcatttct	cgcttaaccg	tgacaggctg	tcgatcttga	gaactatgcc	gacataatag	52260
gaaatcgctg	gataaagccg	ctgaggaagc	tgagtggcgc	tatttcttta	gaagtgaacg	52320
ttgacgatcg	tcgaccgtac	cccgatgaat	taattcggac	gtacgttctg	aacacagctg	52380
gatacttact	tgggcgattg	tcatacatga	catcaacaat	gtaccggtt	gtgtaaccgt	52440
ctcttgagg	ttcgtatgac	actagtggtt	cccctcagct	tcgactaga	tgttgaggcc	52500
taacatttta	ttagagagca	ggctagtgc	ttagatacat	gatcttcagg	ccgttatctg	52560
tcagggcaag	cgaaaattgg	ccatttatga	cgaccaatgc	cccgcagaag	ctcccatctt	52620
tgccgccata	gacgccgcgc	ccccctttg	gggtgtagaa	catcctttg	ccagatgtgg	52680
aaaagaagt	cgttgtccca	ttgttgcaa	tgacgtagta	gccggcgaaa	gtgcgagacc	52740
catttgcgct	atatataagc	ctacgatttc	cgttgcgact	attgtcgtaa	ttggatgaac	52800
tattatcgta	gttgctctca	gagttgtcgt	aatttgatgg	actattgtcg	taattgctta	52860
tggagtgtc	gtagttgctt	ggagaaatgt	cgtagttgga	tggggagtag	tcatagggaa	52920
gacgagcttc	atccactaaa	acaattggca	ggtcagcaag	tgccctgcccc	gatgccatcg	52980
caagtacgag	gcttagaacc	acctcaaca	gatcgcgcat	agtcttcccc	agctctctaa	53040
cgcttgagtt	aagccgcgcc	gcgaagcggc	gtcggcttga	acgaattggt	agacattatt	53100
tgccgactac	cttggtgatc	tcgcctttca	cgtagtgaac	aaattcttcc	aactgatctg	53160
cgcgcgaggc	caagcgatct	tcttgtccaa	gataagcctg	cctagcttca	agtatgacgg	53220
gctgatactg	ggccggcagg	cgctccattg	cccagtcggc	agcgacatcc	ttcggcgcga	53280
ttttgccggt	tactgcgctg	taccaaatgc	gggacaacgt	aagcactaca	tttcgctcat	53340
cgccagccca	gtcgggcggc	gagttccata	gcgttaaggt	ttcatttagc	gcctcaaata	53400
gatcctgttc	aggaaccgga	tcaaagagtt	cctccgcgcg	tggacctacc	aaggcaacgc	53460
tatgttctct	tgcttttgtc	agcaagatag	ccagatcaat	gtcgatcgtg	gctggctcga	53520
agatacctgc	aagaatgtca	ttgcgctgcc	attctccaaa	ttgcagtctg	cgcttagctg	53580
gataacgcca	cggaatgatg	tcgtcgtgca	caacaatggt	gacttctaca	gcgcggagaa	53640
tctcgccttc	tccaggggaa	gccgaagttt	ccaaaaggtc	gttgatcaaa	gctcgcgcgcg	53700
ttgtttcatc	aagccttaca	gtcaccgtaa	ccagcaaatc	aatatactg	tgtggcttca	53760
ggccgccatc	cactgcggag	ccgtacaaat	gtacggccag	caacgtcggg	tcgagatggc	53820
gctcgatgac	gccaactacc	tctgatagtt	gagtcgatac	ttcggcgatc	accgcttccc	53880
tcatgatgtt	taactcctga	attaagccgc	gccgcgaagc	ggtgtcggct	tgaatgaatt	53940
gtaggcgctc	atcctgtgct	cccgagaacc	agtaccagta	catcgtgttt	tcgttcgaga	54000
cttgaggctc	agttttatac	gtgaacaggt	caatgccgcc	gagagtaaag	ccacattttg	54060
cgtaaaaatt	gcaggcaggt	acattgttcg	tttgtgtctc	taatcgtatg	ccaaggagct	54120
gtctgcttag	tgcccacttt	ttcgaaaatt	cgatgagact	gtgcgcgact	cctttgcctc	54180
ggtgcgtgtg	cgacacaaca	atgtgttcga	tagaggctag	atcgttccat	gtaggttga	54240
gttcaatctt	cccgacaagc	tcttggtcga	tgaatgcgcc	atagcaagca	gagtcttcat	54300

cagagtcate	atccgagatg	taatccttcc	ggtaggggct	cacacttctg	gtagatagtt	54360
caaagccttg	gtcggatagg	tgacatcga	acatttcacg	aacaatgaaa	tggttctcag	54420
catccaatgt	ttccgccacc	tgctcaggga	tcaccgaaat	cttcatatga	cgcctaacgc	54480
ctggcacagc	ggatcgcaaa	cctggcgcgg	cttttggcac	aaaaggcgtg	acaggtttgc	54540
gaatccgttg	ctgccacttg	ttaacccttt	tgccagattt	gtaactata	atztatgtta	54600
gaggcgaagt	cttgggtaaa	aactggccta	aaattgctgg	ggatttcagg	aaagtaaaca	54660
tcaccttccg	gctcgatgtc	tattgtagat	atatgtagtg	tatctacttg	atcgggggat	54720
ctgctgcctc	gcgcgtttcg	gtgatgacgg	tgaaaacctc	tgacacatgc	agctcccgga	54780
gacggtcaca	gcttgtctgt	aagcggatgc	cgggagcaga	caagcccgtc	agggcgcgtc	54840
agcgggtgtt	ggcgggtgtc	ggggcgcagc	catgaccag	tcacgtagcg	atagcggagt	54900
gtatactggc	ttaactatgc	ggcatcagag	cagattgtac	tgagagtgea	ccatatgcgg	54960
tgtgaaatac	cgcacagatg	cgtaaggaga	aaataccgea	tcaggcgtc	ttccgcttcc	55020
tcgctcactg	actcgtcgcg	ctcggtcggt	cggtcgcggc	gagcggtatc	agctcactca	55080
aaggcggtaa	tacggttate	cacagaaatca	ggggataacg	caggaaagaa	catgtgagca	55140
aaaggccagc	aaaaggccag	gaaccgtaaa	aaggccgcgt	tgctggcggt	tttccatagg	55200
ctccgcccc	ctgacgagca	tcacaaaaat	cgacgctcaa	gtcagaggtg	gcgaaaccgc	55260
acaggactat	aaagatacca	ggcgtttccc	cctggaaget	ccctcgtgcg	ctctctgttt	55320
ccgaccctgc	cgcttaccgg	atacctgtcc	gcctttctcc	cttcgggaag	cgtggcgctt	55380
tctcatagct	cacgctgtag	gtatctcagt	tcgggtgtagg	tcgttcgctc	caagctgggc	55440
tgtgtgcacg	aacccccctg	tcagcccagc	cgtcgcgctt	tatccggtaa	ctatcgtctt	55500
gagtccaacc	cggtaagaca	cgacttatcg	ccactggcag	cagccactgg	taacaggatt	55560
agcagagcga	ggatgtagg	cgggtctaca	gagttcttga	agtgggtggc	taactacggc	55620
tacactagaa	ggacagtatt	tggtatctgc	gctctgctga	agccagttac	cttcggaaaa	55680
agagttggta	gctcttgatc	cggcaaacia	accaccgctg	gtagcgggtg	ttttttgttt	55740
tgcaagcagc	agattacgcg	cagaaaaaaaa	ggatctcaag	aagatccttt	gatcttttct	55800
acggggctctg	acgctcagtg	gaacgaaaac	tcacgttaag	ggattttggg	catgagatta	55860
tcaaaaagga	tcttcaccta	gatcctttta	aattaaat	gaagttttaa	atcaatctaa	55920
agtatatatg	agtaaacttg	gtctgacagt	taccaatget	taatcagtga	ggcacctatc	55980
tcagcgatct	gtctatctcg	ttcatccata	gttgectgac	tccccgtcgt	gtagataact	56040
acgatacggg	agggcttacc	atctggcccc	agtgtgcaaa	tgataccgcg	agaccacgc	56100
tcaccggctc	cagatttate	agcaataaac	cagccagccg	gaaggccgca	gcgcagaagt	56160
ggctctgcaa	ctttateccg	ctccatccag	tctattaatt	gttgccggga	agctagagta	56220
agtagttcgc	cagttaatag	tttgcgcaac	gttgttgcca	ttgctgcagg	gggggggggg	56280
gggggggact	tccattgttc	attccacgga	caaaaacaga	gaaaggaaac	gacagaggcc	56340
aaaaagcctc	gctttcagca	cctgtcgttt	cctttctttt	cagagggtat	tttaataaaa	56400
aacattaagt	tatgacgaag	aagaacggaa	acgccttaaa	ccggaaaatt	ttcataaata	56460
gcgaaaacc	gcgaggtcgc	cgccccgtaa	cct			56493

<210> 47

<211> 55518

<212> DNA

<213> 人工序列

<220>

<223> 含有玉蜀黍密码子优化的Cas9和玉蜀黍MDH启动子的农杆菌载体

<400> 47

gtcggatcac cggaaaggac ccgtaaagtg ataatgatta tcatctacat atcacaacgt	60
gcgtggaggc catcaaacca cgtcaaataa tcaattatga cgcaggtatc gtattaattg	120
atctgcatca acttaacgta aaaacaactt cagacaatac aatcagcga cactgaatac	180
ggggcaacct catgtceccc ccccccccc cctgcaggc atcgtggtgt cacgctcgtc	240
gtttggtatg gcttcattca gctccggtt ccaacgatca aggcgagtta catgatcccc	300
catgttgtgc aaaaaagcgg ttagctcett cggctectcg atcgttgtea gaagtaagtt	360
ggccgcagtg ttatcaactca tggttatggc agcaactgat aattctetta ctgtcatgcc	420
atccgtaaga tgctttctg tgactggtga gtaactcaacc aagtcattct gagaatagtg	480
tatgcggcga ccgagttgct cttgccccgc gteaacacgg gataataccg cgccacatag	540
cagaacttta aaagtgetca tcattggaaa acgttcttcg gggcgaaaac tctcaaggat	600
cttaccgctg ttgagatcca gttcgatgta acccaactgt gcacccaact gatcttcagc	660
atcttttact ttcaccagcg tttctgggtg agcaaaaaca ggaaggcaaa atgccgcaaa	720
aaaggggaata agggcgacac ggaaatgtt aataactata ctcttccttt ttcaatatta	780
ttgaagcatt tatcagggtt attgtctcat gacgggatac atatttgaat gtatttagaa	840
aaataaaca ataggggtt cgcgcacatt tccccgaaa gtgccacctg acgtctaaga	900
aaccattatt atcatgacat taacctataa aaatagcgt atcacaggc ctttcgtct	960
tcaagaattg gtcgacgatc ttgctgcgtt cggatatttt cgtggagttc ccgccacaga	1020
cccggattga aggcgagatc cagcaactcg cgccagatca tcctgtgacg gaactttggc	1080
gcgtgatgac tggccaggac gtcggccgaa agagcgacaa gcagatcacg cttttcgaca	1140
gcgtcggatt tgcgatcgag gatttttcgg cgctgcgcta cgtccgcgac cgcgttgagg	1200
gatcaagcca cagcagccca ctcgacctt tagccgacct agacgagcca aggatcttt	1260
ttggaatgct gctccgtcgt caggctttcc gacgtttggg tggttgaaca gaagtcatta	1320
tcgtacggaa tgccaagcac tcccgagggg aacctgtgg ttggcatgca catacaaatg	1380
gacgaacgga taaacctttt cacgacctt taaatccg ttattctaataaacgctctt	1440
ttctcttagg tttaccgccc aatatactct gtaaacact gatagtttaa actgaaggcg	1500
ggaaacgaca atctgatcat gagcggagaa ttaaggaggt cacgttatga ccccccca	1560
tgacgcggga caagccgttt tacgtttgga actgacagaa ccgcaacgtt gaaggagcca	1620
ctcagcccaa gctggtacga ttgtaatacg actcaactata gggcgaattg agcgtgttt	1680
aaacgctctt caactggaag agcggttacc agaggccaga atggccatct cggaccgata	1740
tcgctatcaa ctttgatag aaaagttggg ccgaattcga gctcgttacg gccagaatgg	1800
cccggaccgg gttaccgaat tcgagctcgg taccctggga tccgatatcg atgggacctg	1860
gccgaagctt ttggaagge taaggagagg aagccggcga gaaggagggg gcgttttacg	1920
tgctactgtc ctgtcgtgtt ggtgttgac acgaatcatt tcttccgcgc gtgggaagaa	1980
gaagatgcac attagcggcc tgaagtagag atgtcaatgg ggaattcccc agcggggatt	2040

aactccccag	accctaccc	atgaacatag	accggcccc	atccccgaac	ccgaaccgga	2100
cctcgggtac	gaaaatcctc	ccatacccat	tcccgaccgg	gtactaaata	cccatgggta	2160
tccatacccg	acccgattat	tcaaaaatta	atgggctttt	tatttgtaa	ccggcggacg	2220
caatgcttgg	gactctaggt	tttttactt	tgttgaccgg	ctggcggctg	ggctttttcc	2280
tacaggccca	aagtggctc	gcagccacta	ggccacacgt	cacaggcagc	ccacaagtaa	2340
atgtcgttgg	attgctggat	ggtggaataa	aaatcctaga	tgctagattg	ttctggttcc	2400
gggtattttt	ctccatggct	aatcgggtt	gggtttagcc	ctcccaaacc	cgaaccgcc	2460
atacccgatg	ggtaagggat	ttattccaaa	tctataccca	tggggatttg	ttttaaccga	2520
taccttaacc	ctaatagagg	aattccccac	gggtaatcgg	gtttcggggc	ccattgacat	2580
ctctagactg	aaggcgtcca	actcaaatca	ttaaaaagtg	ttgacgcacg	cgctgatgcg	2640
ccggccgcac	agcacagget	gcacagcccc	tttaatcagc	gatggagccc	cgcccgctag	2700
ccagccagg	ccggcgtccg	ggtctgcgcc	ctgcggcgte	actgctgtcg	ccaccgtctc	2760
cgatgggtccc	acatccatcc	agcggggccg	gcgtggtaca	aaaggctctt	cctcgccgtc	2820
aggtgcagct	gccccaaac	cagacacaga	ctccaccacc	ccgcttcgat	cttctgttgc	2880
agctgaaatc	gtcagatc	tgcagttcat	tctcatggc	accgaagaag	aagcgaagg	2940
tgatggacaa	gaagtacagc	atcggcctcg	acatcggcac	caactcggtg	ggctgggccc	3000
tcatcacgga	cgaatataag	gtcccgtcga	agaagttcaa	ggtcctcggc	aatacagacc	3060
gccacagcat	caagaaaaac	ttgatcggcg	ccctctgtt	cgatagcggc	gagaccgcgg	3120
aggcgaccag	gctcaagagg	accgccagga	gacggtacac	taggcgcaag	aacaggatct	3180
gctacctgca	ggagatcttc	agcaacgaga	tggcgaaggt	ggacgactcc	ttcttcacc	3240
gcctggagga	atcattcctg	gtggaggagg	acaagaagca	tgagcggcac	ccaatcttcg	3300
gcaacatcgt	cgacgaggta	agtttctgct	tctaccttg	atatatatat	aataattatc	3360
attaattagt	agtaataata	tatttcaaat	atTTTTTTca	aaataaaaga	atgtagtata	3420
tagcaattgc	tttctgtag	tttataagtg	tgtatatttt	aatttataac	tttctaata	3480
tatgacacaa	acatggtgat	gtgcaggtgg	ctaccacga	gaagtaccgg	acaatctacc	3540
acctccggaa	gaaactgggt	gacagcacag	acaaggcgga	cctccggctc	atctaccttg	3600
ccctcgcgca	tatgatcaag	ttccgcggcc	acttctcat	cgagggcgac	ctgaaccggg	3660
acaactccga	cgtggacaag	ctgttcatcc	agctcgtgca	gacgtacaat	caactgttcg	3720
aggagaaccc	cataaacgct	agcggcgtgg	acgccaagge	catcctctcg	gccaggctct	3780
cgaaatcaag	aaggctggag	aaccttatcg	cgcagttgcc	aggcgaaaag	aagaacggcc	3840
tcttcggcaa	ccttattgcg	ctcagcctcg	gctgacgcc	gaacttcaaa	tcaaacttcg	3900
acctcgcgga	ggacgccaag	ctccagctct	caaaggacac	ctacgacgac	gacctcgaca	3960
acctcctggc	ccagatagga	gaccagtacg	cggacctctt	cctcgccgcc	aagaacctct	4020
ccgacgtat	cctgctcagc	gacatccttc	gggtcaacac	cgaattacc	aaggcaccgc	4080
tgccgccag	catgattaaa	cgtacgacg	agcaccatca	ggacctcagc	ctgctcaagg	4140
cactcgtccg	ccagcagctc	cccgagaagt	acaaggagat	cttcttcgac	caatcaaaaa	4200
acggctacgc	gggatatac	gacggcgggtg	ccagccagga	agagttctac	aagttcatca	4260
aaccaatcct	ggagaagatg	gacggcaccg	aggagttgct	ggtcaagctc	aacagggagg	4320
acctcctcag	gaagcagagg	accttcgaca	acggetccat	cccgcacag	atccacctgg	4380

gcgaactgca	tgccatcctg	cggcgccagg	aggactteta	cccgttctg	aaggataacc	4440
gggagaagat	cgagaagatc	ttgacgttcc	gcatcccata	ctacgtgggc	ccgctggctc	4500
gcggaactc	ccggttcgcc	tggatgacce	ggaagtcgga	ggagaccatc	acaccctgga	4560
actttgagga	ggtggtcgat	aagggcgcta	gcgctcagag	cttcatcgag	cgcatgacca	4620
acttcgataa	aaacctgccc	aatgaaaaag	tcttcccaa	gcactcgctg	ctctacgagt	4680
acttcaccgt	gtacaacgag	ctcaccaagg	tcaaatacgt	caccgagggc	atgcggaagc	4740
cggcgttcct	gagcggcgag	cagaagaagg	cgatagtgga	cctcctcttc	aagaccaaca	4800
ggaaggtgac	cgtgaagcaa	ttaaaagagg	actacttcaa	gaaaatagag	tgcttcgact	4860
ccgtggagat	ctcgggctg	gaggatcgg	tcaacgctc	actcggcacg	tatcacgacc	4920
tctcaagat	cattaaagac	aaggacttcc	tcgacaacga	ggagaacgag	gacatcctcg	4980
aggacatcgt	cctcaccctg	accctgttcg	aggaccgca	aatgatcgag	gagaggctga	5040
agacctacgc	gcacctgttc	gacgacaagg	tcatgaaaca	gctcaagagg	cgccgctaca	5100
ctggttgggg	aaggtgttcc	cgcaagetca	ttaatggcat	cagggacaag	cagagcggca	5160
agaccatcct	ggacttcctc	aagtccgacg	ggttcgcaa	ccgcaacttc	atgcagctca	5220
ttcacgacga	ctcgtcagc	ttcaaggaag	acatccagaa	ggcacagggtg	agcgggcagg	5280
gtgactccct	ccacgaacac	atcgccaacc	tggccggctc	gccggccatt	aaaaagggca	5340
tctgcagac	ggtcaaggtc	gtcgacgagc	tcgtgaaggt	gatgggccgg	cacaagcccc	5400
aaaatatcgt	catagagatg	gccagggaga	accagaccac	ccaaaaaggg	cagaagaact	5460
cgcgcgagcg	gatgaaacgg	atcgaggagg	gcattaaaga	gctcgggtcc	cagatcctga	5520
aggagcacc	cgtggaaaat	accagctcc	agaatgaaaa	gctctacctc	tactacctgc	5580
agaacggccg	cgacatgtac	gtggaccagg	agctggacat	taatcggcta	tcggactacg	5640
acgtcgacca	catcgtgccg	cagtcgttcc	tcaaggacga	tagcatcgac	aacaaggtgc	5700
tcaccgggtc	ggataaaaaat	cggggcaaga	gcgacaacgt	gccagcgag	gaggtcgtga	5760
agaagatgaa	aaactactgg	cgccagctcc	tcaacgcgaa	actgatcacc	cagcgcaagt	5820
tcgacaacct	gacgaaggcg	gaacgcggtg	gcttgagcga	actcgataag	gcgggcttca	5880
taaaaaggca	gctggctcag	acgcgccaga	tcacgaagca	tgtcggccag	atcctggaca	5940
gccgatgaa	tactaagtac	gatgaaaacg	acaagctgat	ccgggagggtg	aaggtgatca	6000
cgctgaagtc	caagctcgtg	tcggacttcc	gcaaggactt	ccagttctac	aaggtccgcg	6060
agatcaacaa	ctaccaccac	gcccacgacg	cctacctgaa	tgcggtggtc	gggaccgccc	6120
tgatcaagaa	gtaccggaag	ctggagtcgg	agttcgtgta	cggcgactac	aaggtctacg	6180
acgtgcgcaa	aatgategcc	aagtccgagc	aggagatcgg	caaggccacg	gcaaaatact	6240
tcttctactc	gaacatcatg	aacttcttca	agaccgagat	caccctcgcg	aacggcgaga	6300
tccgcaagcg	cccgtcctc	gaaaccaacg	gcgagacggg	cgagatcgtc	tgggataagg	6360
gccgggattt	cgcgacggtc	cgcaaggtgc	tctccatgcc	gcaagtcaat	atcgtgaaaa	6420
agacggagg	ccagacgggc	gggttcagca	aggagtccat	cctcccgaag	cgcaactccg	6480
acaagctcat	cgcgaggaag	aaggattggg	accgaaaaa	atatggcggc	ttcgacagcc	6540
cgaccgtcgc	atacagcgtc	ctcgtcgtgg	cgaaggtgga	gaaggcaag	tcaaagaagc	6600
tcaagtcctg	gaaggagctg	ctcgggatca	cgattatgga	gcggtcctcc	ttcgagaaga	6660
accgatcga	cttcttagag	gccaagggat	ataaggaggt	caagaaggac	ctgattatta	6720

aactgccgaa	gtactcgctc	ttcgagctgg	aaaacggccg	caagaggatg	ctcgcctccg	6780
caggcgagtt	gcagaagggc	aacgagctcg	ccctcccag	caaatacgtc	aatttcctgt	6840
acctcgctag	ccactatgaa	aagctcaagg	gcagcccgga	ggacaacgag	cagaagcagc	6900
tcttcgtgga	gcagcacaag	cattacctgg	acgagatcat	cgagcagatc	agcgagttct	6960
cgaagcgggt	gatcctcgcc	gacgcgaacc	tggacaaggt	gctgtcggca	tataacaagc	7020
accgcgacaa	accaatacgc	gagcaggccg	aaaatatcat	ccacctcttc	accctcacca	7080
acctcggcgc	tccggcagcc	ttcaagtact	tcgacaccac	gattgaccgg	aagcgggtaca	7140
cgagcacgaa	ggaggtgctc	gatgcgacgc	tgatccacca	gagcatcaca	gggctctatg	7200
aaacacgcat	cgacctgagc	cagctgggcg	gagacaagag	accacgggac	cgccacgatg	7260
gcgagctggg	aggccgcaag	cgggcaaggt	aggtaccgtt	aacctagact	tgtccatctt	7320
ctggattggc	caacttaatt	aatgtatgaa	ataaaaggat	gcacacatag	tgacatgcta	7380
atcactataa	tgtgggcate	aaagttgtgt	gttatgtgta	attactagtt	atctgaataa	7440
aagagaaaga	gatcatccat	atttettate	ctaaatgaat	gtcacgtgtc	tttataattc	7500
tttgatgaac	cagatgcatt	tcattaacca	aatccatata	catataaata	ttaatcatat	7560
ataattaata	tcaattgggt	tagcaaaaca	aatctagtct	aggtgtgttt	tgcgatgcg	7620
gccgccaccg	cggtaggagct	cgaattcgag	ctcggtagcc	tgggatccag	cttcgcttag	7680
tttttagttt	ttggcagaaa	aaatgatcaa	tgtttcacaa	accaaataat	tttataactt	7740
ttgatgaaag	aagatcacca	cggtcataatc	taggggtggt	aacaaattgc	gatctaaatg	7800
tttcttcata	aaaaataaag	cttcttaata	aatttttagtt	caaaataaat	acgaataaag	7860
tctgattcta	atctgattcg	atccttaaat	tttataatgc	aaaatttaga	gctcattacc	7920
acctctagtc	atatgtctag	tctgaggtat	atccaaaaag	ccctttctct	aaattccaca	7980
cccaactcag	atgtttgcaa	ataaatactc	cgactccaaa	atgtaggtga	agtgcactt	8040
tctccatttt	atatcaacat	ttgttatttt	ttgtttaaca	tttcacactc	aaaactaatt	8100
aataaaatac	gtggttgttg	aacgtgcgca	catgtctccc	ttacattatg	tttttttatt	8160
tatgtattat	tgttgttttc	ctccgaacaa	cttgcaaca	tatcatcatt	ggtctttaat	8220
atztatgaat	atggaagcct	agttatttac	acttggtac	acactagttg	tagttttgcc	8280
acttgtctaa	catgcaactc	tagtagtttt	gccacttgcc	tggcatgcaa	ctctagtatt	8340
gacacttgta	tagcatataa	tgccaatacg	acacctgect	tacatgaaac	attatttttg	8400
acacttgat	accatgcaac	attaccattg	acatttgtec	atacacatta	tatcaaatat	8460
attgagcgca	tgtcacaac	tcgatacaaa	gctggatgac	cctccctcac	cacatctata	8520
aaaaccgag	cgctactgta	aatcaactcac	aacacaacac	atatctttta	gtaacctttc	8580
aatagcgctc	ccccagaac	tagtaacct	ggcctgttc	aacaagttea	tcggcgacga	8640
catgaagatg	acctaccaca	tggacggetg	cgtgaacggc	cactacttea	ccgtgaaggg	8700
cgagggcagc	ggcaagccct	acgagggcac	ccagacctec	accttcaagg	tgacctggc	8760
caacggcggc	cccctggcct	tctccttcga	catcctgtcc	accgtgttea	tgtacggcaa	8820
ccgctgcttc	accgectacc	ccaccageat	gcccgaactac	ttcaagcagg	ccttccccga	8880
cggcatgtcc	tacgagagaa	ccttcaceta	cgaggacggc	ggcgtggcca	ccgccagctg	8940
ggagatcagc	ctgaagggca	actgcttcga	gcacaagtcc	accttccacg	gcgtgaactt	9000
ccccgccgac	ggccccgtga	tggccaagaa	gaccaccggc	tgggaccct	ccttcgagaa	9060



gatgaccgtg	tgcgacggca	tcttgaaggg	cgacgtgacc	gccttcctga	tgctgcaggg	9120
cggcggcaac	tacagatgcc	agttccacac	ctcctacaag	accaagaagc	ccgtgaccat	9180
gcccccaac	cacgtggtgg	agcaccgcat	cgccagaacc	gacctggaca	aggcggcaa	9240
cagcgtgcag	ctgaccgagc	acgccgtggc	ccacatcacc	tccgtggtgc	ccttctgaag	9300
cggcccatgg	atattcgaac	gcgtaggtac	cacatggta	acctagactt	gtccatcttc	9360
tggattggcc	aacttaatta	atgtatgaaa	taaaaggatg	cacacatagt	gacatgctaa	9420
tcactataat	gtgggcatca	aagttgtgtg	ttatgtgtaa	ttactagtta	tctgaataaa	9480
agagaaagag	atcatccata	tttcttatec	taaatgaatg	tcacgtgtct	ttataattct	9540
ttgatgaacc	agatgcattt	cattaaccaa	atccatatac	atataaatat	taatcatata	9600
taattaatat	caattggggt	agcaaaacaa	atctagteta	ggtgtgtttt	gcaattccc	9660
atggacctcg	agggggggcc	cgggcaccca	gctttcttgt	acaaagtggc	cgtaaacgga	9720
tcggccagaa	tggcccggac	cgggttaccg	aattcgagct	cggttaccctg	ggatcggccg	9780
ctctagaact	agtggatccc	ccgggtgca	ggaattceca	tggagtcaaa	gattcaaata	9840
gaggacctaa	cagaactcgc	cgtaaagact	ggcgaacagt	tcatacagag	tctcttacga	9900
ctcaatgaca	agaagaaaa	cttcgtaaac	atggtggagc	acgacacgct	tgtctactcc	9960
aaaaatatca	aagatacagt	ctcagaagac	caaagggcaa	ttgagacttt	tcaacaaagg	10020
gtaatatccg	gaaacctcct	cggattccat	tgcccageta	tctgtcactt	tattgtgaag	10080
atagtggaaa	aggaaggtgg	ctcctacaaa	tgccatcatt	gcgataaagg	aaaggccatc	10140
gttgaagatg	cctctgccga	cagtgggtccc	aaagatggac	ccccaccac	gaggagcatc	10200
gtggaaaaag	aagacgttcc	aaccacgtct	tcaaagcaag	tggattgatg	tgatatctcc	10260
actgacgtaa	gggatgacgc	acaatcccac	taagettegg	ccggggceca	tcgatctggc	10320
gaaaggggga	tgtgctgcaa	ggcgattaag	ttgggtaacg	ccagggtttt	cccagtcacg	10380
acgttgtaaa	acgacggcca	gtgccaagct	cagatcagct	tggggctggt	atcgataaat	10440
gtttccacat	agattttgca	tatcataatg	atgtttgtcg	ttccgtatct	atgtttcata	10500
caaaattttt	acgcataatc	caacacatgg	gcacatacct	agtgactgta	taactctgca	10560
tgtatgagtg	tatgactata	tgatgtagta	actaataaga	agggtagaca	tttgagtgat	10620
tcttttattc	ctggacttgt	aagacttgac	atttctgect	tgagtgcgat	acatcatatg	10680
gacaggggtt	atgcatacac	tgettgtttg	ttgtttatgt	tctaagagca	tctccaacaa	10740
cgtgacatat	gaaaatgccc	tacaatttaa	aatgggttat	atthttataaa	atthtagggca	10800
taaataaaaac	atcccgcctc	aacattaaag	ccttaaatct	attatagggga	agcccactat	10860
gatatagtat	atthtagggca	ctthtagagg	tgcctataa	thttthtgacc	atthththtat	10920
gaaatgagac	actatthggag	tathththth	ccgtagagca	ccatathhca	atthtagaca	10980
ccaathhaag	gcathhthgg	agathhthct	aatgthhgg	taththhgtct	gtathhthgt	11040
ggtthhgata	gtgthhgcct	tgcaathhgt	atthhacatt	gacaathhata	atagthhaaa	11100
ctctacaaa	thththtct	atgthhctct	gtathhaaca	thhthhctgc	acacathhctga	11160
tgthhaacact	gcathhctth	aacathhgaca	agathhctgt	tcathhthtag	gctathhthgg	11220
gaaccaaath	thththtagg	ththththct	thhaaaaaag	ththththth	thctthhgaga	11280
aaathhaaat	thctthhgaga	thththtagct	thhaacthct	ctththctth	thgathhctct	11340
ctctathhca	atgthhctct	ctthhctct	gathhaactg	gathhctct	actgthhthaa	11400

gtaatacgtg	tcattattac	taa atgccat	tttagtaaat	gttgagtatg	tactctacta	11460
cagtaagtat	tattggtgta	tttacctag	acagttggcg	gcctggcggg	taaagttatc	11520
ctgtagaaag	ttgggccagg	ccaaaaccaa	ccgccaaagg	aaaggccttc	cggccccgcc	11580
acctttgcgc	gccgaaggtc	agttccttca	gtctcctccc	gcttcagact	ctgaccacgt	11640
cgacaatccg	ggccgaaaca	catctgcacc	gtccacttgc	gacagattga	acacaccact	11700
tctatccacg	tcagcgatcc	gtggcactag	cccttcacc	aatcagccca	agttgcccct	11760
ttcctttaaa	ttcgccgcac	ccattgctct	tctcacggcc	atagaaatcg	accgagcgaa	11820
tccctcgcat	cgcatctgca	gcctttgctg	catcacacca	ccgcgaaacc	ccagcagccg	11880
catctgcagg	tcgactctag	aggatccatg	gcctcctccg	aggacgtcat	caaggagttc	11940
atgcgcttca	aggtgcgcat	ggagggctcc	gtgaacggcc	acgagttcga	gatcgagggc	12000
gagggcgagg	gccgccccta	cgagggcacc	cagaccgcca	agctgaaggt	gaccaagggc	12060
ggccccctgc	ccttcgcctg	ggacatcctg	tccccccagt	tccagtacgg	ctccaaggtg	12120
tacgtgaagc	accccgccga	catccccgac	tacaagaagc	tgctcttccc	cgagggtctc	12180
aagtgggagc	gcgtgatgaa	cttcgaggac	ggcggcgtgg	tgacagtgac	ccaggactcc	12240
tccctgcagg	acggctcctt	catctacaag	gtgaagtcca	tcggcgtgaa	cttccccctc	12300
gacggccccg	taatgcagaa	gaagactatg	ggctgggagg	cctccaccga	gcgcctgtac	12360
ccccgcgacg	gcgtgctgaa	gggcgagatc	cacaaggccc	tgaagctgaa	ggacggcggc	12420
cacgctagcc	catccacca	ctcactcact	catactctgt	ctgtacgtac	gagaatttct	12480
cgaccaaccg	tcgtgagacc	tgcccaccgg	agatcggacg	caagagggtt	taggcaagaa	12540
tgtcgtgcga	cagggtgagc	gctgactagt	atacgtgaga	gaccttgaga	tatacctcac	12600
acgtacgcgt	actttacatg	acgtaggaca	ttacgactca	aacagattca	cgtcagattt	12660
cggagtttct	cacgcgtgag	agccttggag	ggcggtatgt	atgtcact	atatgttggg	12720
atggagggag	tgagtgagtg	atatgtggct	agcaagggcg	gccccctgcc	cttcgcctgg	12780
gacatcctgt	ccccccagtt	ccagtacggc	tccaaggtgt	acgtgaagca	ccccgccgac	12840
atccccgact	acaagaagct	gtccttcccc	gagggcttca	agtgggagcg	cgatgatgaac	12900
ttcgaggacg	gcggcgtggt	gacagtgacc	caggactcct	ccctgcagga	cggctccttc	12960
atctacaagg	tgaagttcat	cggcgtgaac	ttccccctcc	acggccccgt	aatgcagaag	13020
aagactatgg	gctgggaggc	ctccaccgag	cgctgtacc	cccgcgacgg	cgatgctgaag	13080
ggcgagatcc	acaaggccct	gaagctgaag	gacggcggcc	actacctggt	ggagttcaag	13140
tccatctaca	tgccaagaa	gcccgtgcag	ctgcccggct	actactacgt	ggactccaag	13200
ctggacatca	cctcccacaa	cgaggactac	accatcgtgg	agcagtacga	gcgcgccgag	13260
ggccgccacc	acctgttctt	gtagtcagga	tctgagtcga	aacctagact	tgtccatctt	13320
ctggattggc	caacttaatt	aatgtatgaa	ataaaaggat	gcacacatag	tgacatgcta	13380
atcaactata	tgtgggcate	aaagttgtgt	gttatgtgta	attactagtt	atctgaataa	13440
aagagaaaga	gatcatccat	atttettatc	ctaaatgaat	gtcacgtgtc	tttataatc	13500
tttgatgaac	cagatgcatt	tcattaacca	aatccatata	catataaata	ttaatcatat	13560
ataattaata	tcaattgggt	tagcaaaaca	aatctagtct	aggtgtgttt	tgcaatgcg	13620
gccgccaccg	cgggtggagct	cgaattccgg	tccgggtcac	ccggtccggg	cctagaaggc	13680
cagcttgcgg	ccgccccggg	caactttatt	atacaaagtt	gatagatata	ggtccgagcg	13740

gcctagaagg	cctttggtca	cctttgtcca	ccaagatgga	actgcggccg	ctcattaatt	13800
aagtcaggcg	cgctctagt	tgaagacacg	ttcatgtett	catcgtaaga	agacactcag	13860
tagtcttcgg	ccagaatggc	ctaactcaag	gccatcgtgg	cctcttgctc	ttcaggatga	13920
agagctatgt	ttaaactgtc	aagcgctact	agacaattca	gtacattaaa	aacgtccgca	13980
atgtgttatt	aagttgtcta	agcgtcaatt	tgtttacacc	acaatatatc	ctgccaccag	14040
ccagccaaca	gctccccgac	cggcagctcg	gcacaaaatc	accactcgat	acaggcagcc	14100
catcagtccg	ggacggcgtc	agcgggagag	ccgttgtaag	gcggcagact	ttgctcatgt	14160
taccgatgct	attcggaaaga	acggcaacta	agctgccggg	tttgaaacac	ggatgatctc	14220
gcggagggta	gcatgttgat	tgtaacgatg	acagageggt	gctgcctgtg	atcaaatatc	14280
atctccctcg	cagagatccg	aattatcagc	cttcttatte	atctctcgct	taaccgtgac	14340
aggctgtcga	tcttgagaac	tatgccgaca	taataggaaa	tcgctggata	aagccgctga	14400
ggaagctgag	tggectatt	tctttagaag	tgaacgttga	cgatcgtcga	ccgtaccccc	14460
atgaattaat	tcggacgtac	gttctgaaca	cagctggata	cttacttggg	cgattgtcat	14520
acatgacatc	aacaatgtac	ccgtttgtgt	aaccgtctct	tggaggttcg	tatgacacta	14580
gtggttcccc	tcagcttgcg	actagatggt	gaggcctaac	atctttattag	agagcaggct	14640
agttgcttag	atacatgac	ttcaggccgt	tatctgtcag	ggcaagcgaa	aattggccat	14700
ttatgacgac	caatgccccg	cagaagetcc	catctttgcc	gccatagacg	ccgcgcccc	14760
cttttgggg	gtagaacatc	cttttgccag	atgtggaaaa	gaagtctcgt	gtcccattgt	14820
tggcaatgac	gtagtagccg	gcgaaagtgc	gagaccatt	tgcgctatat	ataagcctac	14880
gatttccgtt	gcgactattg	tcgtaattgg	atgaactatt	atcgtagtgt	ctctcagagt	14940
tgctgtaatt	tgatggacta	ttgtcgtaat	tgcttatgga	gttgtcgtag	ttgcttgagg	15000
aaatgtcgta	gttgatggg	gagtagtcat	aggaagacg	agcttcatcc	actaaaacaa	15060
ttggcaggtc	agcaagtgcc	tgccccgatg	ccatcgcaag	tacgaggctt	agaaccacct	15120
tcaacagatc	gcgcatagtc	ttccccagct	ctctaacgct	tgagttaagc	cgcgccgcca	15180
agcggcgtcg	gcttgaacga	attgttagac	attatttgcc	gactaccttg	gtgatctcgc	15240
ctttcacgta	gtgaacaaat	tcttccaact	gatctgcgcg	cgaggccaag	cgatcttctt	15300
gtccaagata	agcctgccta	gcttcaagta	tgacgggctg	atactgggcc	ggcaggcgct	15360
ccattgcccc	gtcggcagcg	acatccttcg	gcgcgatttt	gccggttact	gcgctgtacc	15420
aaatgcggga	caacgtaagc	actacatttc	gctcatcgcc	agcccagtcg	ggcggcgagt	15480
tccatagcgt	taaggtttca	tttagcgctt	caaatagatc	ctgttcagga	accggatcaa	15540
agagttcctc	cgccgctgga	cctaccaagg	caacgctatg	ttctcttgct	tttgtcagca	15600
agatagccag	atcaatgtcg	atcgtggctg	gctcgaagat	acctgcaaga	atgtcattgc	15660
gctgccattc	tccaaattgc	agttcgcgct	tagctggata	acgccacgga	atgatgtcgt	15720
cgtgcacaac	aatggtgact	tctacagcgc	ggagaatctc	gctctctcca	ggggaagccg	15780
aagtttccaa	aaggtcgttg	atcaaagctc	gccgcgttgt	ttcatcaagc	cttacagtca	15840
ccgtaaccag	caaatcaata	tcactgtgtg	gcttcaggec	gccatccact	gcggagccgt	15900
acaaatgtac	ggccagcaac	gtcggttcga	gatggcgctc	gatgacgcca	actacctctg	15960
atagttgagt	cgatacttcg	gcgatcaccg	cttccctcat	gatgtttaac	tctgtaatta	16020
agccgcgccg	cgaagcggtg	tcggcttgaa	tgaattgtta	ggcgtcatcc	tgtgctcccc	16080

agaaccagta	ccagtacatc	gctgtttcgt	tcgagacttg	aggtctagtt	ttatacgtga	16140
acaggtcaat	gccgccgaga	gtaaagccac	atthttgcgta	caaattgcag	gcaggtacat	16200
tgttcgtttg	tgtctctaata	cgtatgccaa	ggagctgtct	gcttagtgcc	cactttttcg	16260
caaattcgat	gagactgtgc	gcgactcctt	tgccctcgggtg	cgtgtgcgac	acaacaatgt	16320
gttcgataga	ggctagatcg	ttccatgttg	agttgagttc	aatcttcccc	acaagctctt	16380
ggtcgatgaa	tgcgccatag	caagcagagt	cttcatcaga	gtcatcatcc	gagatgtaat	16440
ccttccggta	ggggctcaca	cttctggtag	atagttcaaa	gccttggtcg	gataggtgca	16500
catcgaacac	ttcacgaaca	atgaaatggt	tctcagcatc	caatgtttcc	gccacctgct	16560
cagggatcac	cgaaatcttc	atatgacgcc	taacgcctgg	cacagcggat	cgcaaacctg	16620
gcgcggcttt	tggcacaaaa	ggcgtgacag	gtttgcgaat	ccgttgctgc	cacttgttaa	16680
cccttttgcc	agatttggtg	actataatth	atgttagagg	cgaagtcttg	ggtaaaaact	16740
ggcctaaaaat	tgctggggat	ttcaggaaaag	taaacatcac	cttccggctc	gatgtctatt	16800
gtagatata	gtagtgtatc	tacttgatcg	ggggatctgc	tgccctcgcg	gtttcgggtg	16860
tgacggtgaa	aacctctgac	acatgcagct	cccggagacg	gtcacagctt	gtctgtaagc	16920
ggatgccggg	agcagacaag	cccgtcaggg	cgcgtcagcg	ggtgttgccg	ggtgtcgggg	16980
cgcagccatg	accagtcac	gtagcgatag	cggagtgtat	actggcttaa	ctatgcggca	17040
tcagagcaga	ttgtactgag	agtgcacat	atgcggtgtg	aaataccgca	cagatgcgta	17100
aggagaaaaat	accgcatcag	gcgctcttcc	gcttctcgc	tactgactc	gctgcgctcg	17160
gtcgttcggc	tgcggcgagc	ggtatcagct	cactcaaagg	cggtaatacg	gttatccaca	17220
gaatcagggg	ataacgcagg	aaagaacatg	tgagcaaaag	gccagcaaaa	ggccaggaac	17280
cgtaaaaagg	ccgcgttgct	ggcgtttttc	cataggetcc	gccccctga	cgagcatcac	17340
aaaaatcgac	gctcaagtca	gaggtggcga	aaccgcagag	gactataaag	ataccaggcg	17400
tttccccctg	gaagctccct	cgtgcgctct	cctgttccga	ccctgccgct	taccggatac	17460
ctgtccgcct	ttctcccttc	gggaagcgtg	gcgctttctc	atagctcacg	ctgtaggtat	17520
ctcagttcgg	tgtaggtcgt	tcgctccaag	ctgggctgtg	tgcacgaacc	ccccgttcag	17580
cccgaccgct	gcgccttata	cggtaactat	cgtcttgagt	ccaaccgggt	aagacacgac	17640
ttatcgccac	tggcagcagc	cactggtaac	aggattagca	gagcgaggta	tgtaggcgggt	17700
gctacagagt	tcttgaagtg	gtggcctaac	tacggctaca	ctagaaggac	agtatttgggt	17760
atctgcgctc	tgctgaagcc	agttaccttc	ggaaaaagag	ttggtagctc	ttgatccggc	17820
aaacaaacca	ccgctggtag	cggtggtttt	tttgtttgca	agcagcagat	tacgcgcaga	17880
aaaaaaggat	ctcaagaaga	tcctttgate	ttttctacgg	ggtctgacgc	tcagtggaac	17940
gaaaactcac	gttaagggat	tttggctcatg	agattatcaa	aaaggatctt	cacctagatc	18000
cttttaaat	aaaaatgaag	ttttaaatca	atctaaagta	tatatgagta	aacttggctc	18060
gacagttacc	aatgcttaat	cagtgaggea	ctatctcag	cgatctgtct	atthcgttca	18120
tccatagttg	cctgactccc	cgctcgttag	ataactacga	tacgggaggg	cttaccatct	18180
ggccccagtg	ctgcaatgat	accgcgagac	ccacgctcac	cggctccaga	tttatcagca	18240
ataaaccagc	cagccggaag	ggccgagcgc	agaagtggtc	ctgcaacttt	atccgcctcc	18300
atccagtcta	ttaattgttg	ccgggaagct	agagtaagta	gttcgccagt	taatagtttg	18360
cgcaacgttg	ttgccattgc	tgcagggggg	gggggggggg	gggacttcca	ttgttcattc	18420

cacggacaaa	aacagagaaa	ggaaacgaca	gaggccaaaa	agcctcgtt	tcagcacctg	18480
tcgtttcctt	tcttttcaga	gggtatttta	aataaaaaca	ttaagttatg	acgaagaaga	18540
acggaaacgc	cttaaaccgg	aaaattttca	taaatagcga	aaaccgcga	ggtcgccgcc	18600
ccgtaacctg	tcggatcacc	ggaaaggacc	cgtaaagtga	taatgattat	catctacata	18660
tcacaacgtg	cgtggaggcc	atcaaaccac	gtcaaataat	caattatgac	gcaggtatcg	18720
tattaattga	tctgcatcaa	cttaacgtaa	aaacaacttc	agacaataca	aatcagcgac	18780
actgaatacg	gggcaacctc	atgtccccc	cccccccc	cctgcaggca	tcgtggtgtc	18840
acgctcgtcg	tttggtatgg	cttcattcag	ctccggttc	caacgatcaa	ggcgagttac	18900
atgatcccc	atgttgtgca	aaaaagcgg	tagctcttc	ggtcctccga	tcgttgtcag	18960
aagtaagttg	gccgcagtgt	tatcactcat	ggttatggca	gactgcata	attctcttac	19020
tgtcatgcca	tccgtaagat	gcttttctgt	gactggtgag	tactcaacca	agtcattctg	19080
agaatagtgt	atgcggcgac	cgagttgetc	ttgcccggcg	tcaacacggg	ataataccgc	19140
gccacatagc	agaactttaa	aagtgetcat	cattggaaaa	cgttcttcgg	ggcgaaaact	19200
ctcaaggatc	ttaccgctgt	tgagatccag	ttcgatgtaa	cccactcgtg	cacccaactg	19260
atcttcagca	tcttttactt	tcaccagcgt	ttctgggtga	gcaaaaacag	gaaggcaaaa	19320
tgccgcaaaa	aagggaaataa	gggcgacacg	gaaatgttga	atactcatac	tcttcctttt	19380
tcaatattat	tgaagcattt	atcagggtta	ttgtctcatg	agcggataca	tatttgaatg	19440
tatttagaaa	aataaacaaa	taggggttcc	gcgcacattt	ccccgaaaag	tgccacctga	19500
cgtctaagaa	accattatta	tcatgacatt	aacctataaa	aataggcgta	tcacgaggcc	19560
ctttcgtctt	caagaattcg	gagcttttgc	cattctcacc	ggattcagtc	gtcactcatg	19620
gtgatttctc	acttgataac	cttatttttg	acgaggggaa	attaataggt	tgtattgatg	19680
ttggacgagt	cggaatcgca	gaccgatacc	aggatcttgc	catcctatgg	aactgcctcg	19740
gtgagttttc	tccttcatta	cagaaacggc	ttttcaaaa	atatggtatt	gataatcctg	19800
atatgaataa	attgcagttt	catttgatgc	tcgatgagtt	tttctaataca	gaattggtta	19860
attggttgta	acactggcag	agcattacgc	tgacttgacg	ggacggcggc	ttgttgtaat	19920
aaatcgaact	tttgctgagt	tgaaggatca	gatcacgcat	cttcccagca	acgcagaccg	19980
ttccgtggca	aagcaaaaagt	tcaaaatcac	caactggtcc	acctacaaca	aagctctcat	20040
caaccgtggc	tccctcactt	tctggctgga	tgatggggcg	attcaggcct	ggtatgagtc	20100
agcaacacct	tcttcacgag	gcagacctca	gcgccagaag	gccgccagag	aggccgagcg	20160
cggccgtgag	gcttgacgc	tagggcaggg	catgaaaaag	cccgtagcgg	gctgctacgg	20220
gcgtctgacg	cgggtgaaaag	ggggagggga	tgttgtctac	atggctctgc	tgtagtgagt	20280
gggttgcgct	ccggcagcgg	tctgatcaa	tcgtaccct	ttctcggtec	ttcaacgttc	20340
ctgacaacga	gcctcctttt	cgccaatcca	tcgacaatca	ccgcgagtec	ctgctcgaac	20400
gctgcgtccg	gaccggttc	gtcgaaggcg	tctatcgcgg	cccgaacag	cggcgagagc	20460
ggagcctgtt	caacggtgcc	gccgcgctcg	ccggcatcgc	tgtcgcggc	ctgctcctca	20520
agcacggccc	caacagtga	gtagctgatt	gtcactcagc	cattgacggc	gtccccggcc	20580
gaaaaacccg	cctcgcagag	gaagcgaagc	tgcgcgtcgg	ccgtttccat	ctgcggtgcg	20640
ccggctcgcg	tgccggcatg	gatgcgcgcg	ccatcgcggt	aggcgagcag	cgctgcctg	20700
aagctgcggg	cattcccgat	cagaaatgag	cgccagtcgt	cgtcggctct	cggcaccgaa	20760

tgcgatgat	tctccgccag	catggettcg	gccagtgcgt	cgagcagcgc	ccgcttgttc	20820
ctgaagtgcc	agtaaagcgc	cggctgctga	acccccaacc	gttccgccag	tttgcgtgtc	20880
gtcagaccgt	ctacgccgac	ctcgttcaac	aggtccaggg	cggcacggat	cactgtattc	20940
ggctgcaact	ttgtcatgct	tgacacttta	tcactgataa	acataatatg	tccaccaact	21000
tatcagtgat	aaagaatccg	cgcgttcaat	cggaccagcg	gaggctggtc	cggaggccag	21060
acgtgaaacc	caacataccc	ctgatcgtaa	ttctgagcac	tgtcgcgctc	gacgctgtcg	21120
gcatcggcct	gattatgccg	gtgctgccgg	gcctcctgcg	cgatctggtt	cactcgaacg	21180
acgtcaccgc	ccactatggc	attctgctgg	cgctgtatgc	gttggtgcaa	tttgcctgcg	21240
cacctgtgct	gggcgcgctg	tcggatcggt	tcgggcggcg	gccaatcttg	ctcgtctcgc	21300
tggccggcgc	cactgtcgac	tacgccatca	tggcgacagc	gcctttcctt	tgggttctct	21360
atatcgggcg	gatcgtggcc	ggcatcaccg	gggcgactgg	ggcggtagcc	ggcgcttata	21420
ttgccgatat	cactgatggc	gatgagcgcg	cgcggcaact	cggttcatg	agcgctgttt	21480
tcgggttcgg	gatggctcgcg	ggacctgtgc	tcgggtgggt	gatgggcggg	ttctccccc	21540
acgtccggtt	cttcgccgcg	gcagccttga	acggcctcaa	tttctgacg	ggctgtttcc	21600
ttttgccgga	gtcgcacaaa	ggcgaacgcc	ggccgttacg	ccgggaggct	ctcaaccgcg	21660
tcgcttcggt	ccggtgggccc	cggggcatga	ccgtcgtcgc	cgccctgatg	gcggtcttct	21720
tcatcatgca	acttgtcgga	caggtgccgg	ccgcgctttg	ggtcattttc	ggcgaggatc	21780
gctttcactg	ggacgcgacc	acgatcggca	tttcgcttgc	cgcatttgge	attctgcatt	21840
cactcgccea	ggcaatgatc	accggccctg	tagccgcccc	gctcggcgaa	aggcgggcac	21900
tcatgctcgg	aatgattgcc	gacggcacag	gctacatcct	gcttgcttcc	gcgacacggg	21960
gatggatggc	gttcccgatc	atggtcctgc	ttgcttcggg	tggcatcgga	atgccggcgc	22020
tgcaagcaat	gttgtccagg	caggtggatg	aggaacgtca	ggggcagctg	caaggctcac	22080
tggcggcgct	caccagcctg	acctgatcgc	tcggaccctt	cctcttcacg	gcgatctatg	22140
cggcttctat	aacaacgtgg	aacgggtggg	catggattgc	aggcgtgcc	ctctacttgc	22200
tctgcctgcc	ggcgtcgcgt	cgcgggcttt	ggagcggcgc	agggcaacga	gccgatcgct	22260
gatcgtggaa	acgataggcc	tatgccatgc	gggtcaaggc	gacttccggc	aagctatacg	22320
cgccctagga	gtgcggttgg	aacgttggcc	cagccagata	ctcccgatca	cgagcaggac	22380
gccgatgatt	tgaagcgcac	tcagcgtctg	atccaagaac	aaccatccta	gcaacacggc	22440
ggtccccggg	ctgagaaaagc	ccagtaagga	aacaactgta	ggttcgagtc	gcgagatccc	22500
ccggaaccaa	aggaagtagg	ttaaaccgcg	tccgatcagg	ccgagccacg	ccaggccgag	22560
aacattgggt	cctgtaggca	tcgggattgg	cggatcaaac	actaaagcta	ctggaacgag	22620
cagaagtcc	ccggccgcca	gttgccaggc	ggtaaagggtg	agcagaggca	cgggagggtg	22680
ccacttgccg	gtcagcacgg	ttccgaacgc	catggaaacc	gccccgcca	ggcccgtgc	22740
gacgccgaca	ggatctagcg	ctcgttttgg	tgtcaacacc	aacagcgcga	cgcccgcagt	22800
tccgcaaata	gccccagga	ccgccatcaa	tcgtatcggg	ctacctagca	gagcggcaga	22860
gatgaacacg	accatcagcg	getgcacagc	gcctaccgtc	gccgcgacc	cgcccggcag	22920
gcggtagacc	gaaataaaca	acaagetcca	gaatagcgaa	atattaagtg	cgccgaggat	22980
gaagatgcgc	atccaccaga	ttcccgttgg	aatctgtcgg	acgatcatca	cgagcaataa	23040
accgcccggc	aacgcccgca	gcagcatacc	ggcgaccctt	cggcctcgtc	gttcgggctc	23100

cacgaaaacg	ccggacagat	gcgcttgtg	agcgtccttg	gggcegtcct	cctgtttgaa	23160
gaccgacagc	ccaatgatct	cgccgtcgat	gtaggcgccg	aatgccacgg	catctcgcaa	23220
ccgttcagcg	aacgcctcca	tgggcttttt	ctcctcgtgc	tcgtaaacgg	acccgaacat	23280
ctctggagct	ttcttcaggg	ccgacaatcg	gatctcgccg	aaatcctgca	cgtcggccgc	23340
tccaagccgt	cgaatctgag	ccttaatcac	aattgtcaat	tttaatcctc	tgtttatcgg	23400
cagttcgtag	agcgcgccgt	gcgctcccag	cgatactgag	cgaagcaagt	gcgtcgagca	23460
gtgcccgcct	gttcctgaaa	tgccagtaaa	gcgctggctg	ctgaaccccc	agccggaact	23520
gacccccaaa	ggccctagcg	tttgcaatgc	accaggtcat	cattgacca	ggcgtgttcc	23580
accaggccgc	tgctcgcgaa	ctcttcgcag	gcttcgccga	cctgctcgcg	ccacttcttc	23640
acgcgggtgg	aatccgatec	gcacatgagg	cggaaggttt	ccagcttgag	cgggtacggc	23700
tcccgggtgcg	agctgaaata	gtcgaacatc	cgtcggggccg	tcggcgacag	cttgccgtac	23760
ttctcccata	tgaatttcgt	gtagtggtcg	ccagcaaaca	gcacgacgat	ttcctcgtcg	23820
atcaggacct	ggcaacggga	cgttttcttg	ccacggtcca	ggacgcggaa	gcggtgcagc	23880
agcgacaccg	attccaggtg	cccaacgcgg	tcggacgtga	agcccatcgc	cgtegcctgt	23940
aggcgcgaca	ggcattctc	ggccttcgtg	taataccggc	cattgatcga	ccagcccagg	24000
tcttgcaaaa	gctcgtagaa	cgtgaaggtg	atcggctcgc	cgataggggt	gcgcttcgcg	24060
tactccaaca	cctgctgcca	caccagttcg	tcctcgtcgg	cccgcagctc	gacgccggtg	24120
taggtgatct	tcacgtcctt	gttgacgtgg	aaaatgacct	tgttttgcag	cgctcgcgc	24180
gggattttct	tgttgcgcgt	ggtgaacagg	gcagagcggg	ccgtgtcgtt	tggcatcgtc	24240
cgcatcgtgt	ccggccacgg	cgcaatatcg	aacaaggaaa	gctgcatttc	cttgatctgc	24300
tgcttcgtgt	gtttcagcaa	cgcgccctgc	ttggcctcgc	tgacctgttt	tgccaggtec	24360
tcgccggcgg	tttttcgctt	cttggtcgtc	atagttctc	gcgtgtcgat	ggtcatcgac	24420
ttcgccaaac	ctgccgcctc	ctgttcgaga	cgacgcgaac	gctccacggc	ggccgatggc	24480
gcgggcaggg	cagggggagc	cagttgcacg	ctgtcgcgct	cgatcttggc	cgtagcttgc	24540
tggaccatcg	agccgacgga	ctggaagggt	tcgcggggcg	cacgcatgac	ggtgcggcct	24600
gcgatggttt	cgcatcctc	ggcggaaaac	cccgcgtcga	tcagttcttg	cctgtatgcc	24660
ttccgggtcaa	acgtccgatt	cattcacctc	ccttgccgga	ttgccccgac	tcacgccggg	24720
gcaatgtgcc	cttattcctg	attgacccg	cctggtgect	tgggtgccag	ataatccacc	24780
ttatcggcaa	tgaagtcggt	cccgtagacc	gtctggccgt	ccttctcgta	cttggtattc	24840
cgaatcttgc	cctgcacgaa	taccagcgac	cccttgecca	aatacttgcc	gtgggcctcg	24900
gcctgagagc	caaaacactt	gatgcggaag	aagtcggtgc	gctcctgctt	gtcgccggca	24960
tcgttgccgc	actcttcatt	aaccgctata	tcgaaaattg	cttgccgctt	gttagaattg	25020
ccatgacgta	cctcgggtgc	acgggtaaga	ttaccgataa	actggaactg	attatggctc	25080
atatcgaaaag	tctccttgag	aaaggagact	ctagttttagc	taaacattgg	ttccgctgtc	25140
aagaacttta	gcggctaaaa	ttttgcgggc	cgcgacaaa	ggtgcgaggg	gcggcttccg	25200
ctgtgtacaa	ccagatattt	ttcacciaaca	tccttcgtct	gctcgatgag	cggggcatga	25260
cgaaacatga	gctgtcggag	agggcagggg	tttcaatttc	gtttttatca	gacttaacca	25320
acggtaaggc	caaccctcgc	ttgaaggtga	tggaggccat	tgccgacgcc	ctggaaactc	25380
ccctacctct	tctcctggag	tccaccgacc	ttgaccgcga	ggcactcgcg	gagattgcgg	25440

gtcatccttt	caagagcagc	gtgccgcccc	gatacgaacg	catcagtgtg	gttttgccgt	25500
cacataaggc	gtttatcgtg	aagaaatggg	gcgacgacac	ccgaaaaaag	ctgcgtggaa	25560
ggctctgacg	ccaaggggta	gggcttgcac	ttccttcttt	agccgctaaa	acggcccctt	25620
ctctgcgggc	cgtcggctcg	cgcatcatat	cgacatcctc	aacggaagcc	gtgccgcgaa	25680
tggcatcggg	cgggtgcgct	ttgacagttg	ttttctatca	gaacccttac	gtcgtgcggt	25740
tcgattagct	gtttgtcttg	caggctaaac	actttcggtg	tatcgtttgc	ctgtgcgata	25800
atgttgctaa	tgatttggtg	cgtagggggt	actgaaaagt	gagcgggaaa	gaagagtttc	25860
agaccatcaa	ggagcgggcc	aagcgcaagc	tggaacgcga	catgggtgcg	gacctgttgg	25920
ccgcgctcaa	cgaccgaaaa	accgttgaag	tcgatgctcaa	cgcggacggc	aaggtgtggc	25980
acgaacgcct	tggcgagccg	atgcggtaca	tctgcgacat	gcgcccagc	cagtcgcagg	26040
cgattataga	aacggtggcc	ggattccacg	gcaaagaggt	cacgcggcat	tcgcccatec	26100
tggaaggcga	gttccccttg	gatggcagcc	gctttgcccg	ccaattgccg	ccggctcgtg	26160
ccgcgccaac	ctttgcgac	cgcaagcgcg	cggtcgccat	cttcacgctg	gaacagtacg	26220
tcgaggcggg	catcatgacc	cgcgagcaat	acgaggtcat	taaaagcgc	gtcgcggcgc	26280
atcgaaacat	cctcgtcatt	ggcggtaactg	gtcggggcaa	gaccacgctc	gtcaacgcga	26340
tcatcaatga	aatggtcgcc	ttcaaccctg	ctgagcgcgt	cgatcatc	gaggacaccg	26400
gcgaaatcca	gtgcgccgca	gagaacgccg	tccaatacca	caccagcacc	gacgtctcga	26460
tgacgctgct	gctcaagaca	acgctgcgta	tgcgccccga	ccgatcctg	gtcggtgagg	26520
tacgtggccc	cgaagccctt	gatctgttga	tggcctggaa	caccgggcat	gaaggaggtg	26580
ccgccaccct	gcacgcaaac	aaccccaaa	cgggcctgag	ccggctcgcc	atgcttatca	26640
gcatgcacc	ggattcaccg	aaaccattg	agccgctgat	tggcgaggcg	gttcatgtgg	26700
tcgtccatat	cgccaggacc	cctagcggcc	gtcagtgca	agaaattctc	gaagttcttg	26760
gttacgagaa	cggccagtac	atcaccaaaa	ccctgtaagg	agtatttcca	atgacaacgg	26820
ctgttccgtt	ccgtctgacc	atgaatcgcg	gcattttggt	ctaccttgcc	gtgttcttcg	26880
ttctcgtct	cgcgttatcc	gcgcatccgg	cgatggcctc	ggaaggcacc	ggcggcagct	26940
tgccatatga	gagctggctg	acgaacctgc	gcaactccgt	aaccggcccc	gtggccttcg	27000
cgctgtccat	catcggcacc	gtcgtcgccg	gcggcgtgct	gatcttcggc	ggcgaactca	27060
acgccttctt	ccgaaccctg	atcttctg	ttctggtgat	ggcgtgctg	gtcggcgcgc	27120
agaacgtgat	gagcaccttc	ttcggtcgtg	gtgccgaaat	cgcggccctc	ggcaacgggg	27180
cgctgcacca	ggtgcaagtc	gcggcggcgg	atgccgtgcg	tcgggtagcg	gctggacggc	27240
tcgcctaate	atggctctgc	gcacgatecc	catecgtcgc	gcaggcaacc	gagaaaacct	27300
gttcatgggt	ggtgatcgtg	aactggtgat	gttctcgggc	ctgatggcgt	ttgcgctgat	27360
tttcagcgcc	caagagctgc	gggccaccgt	ggtcggctctg	atcctgtggt	tcggggcgct	27420
ctatgcgttc	cgaatcatgg	cgaaggccga	tccgaagatg	cggttcgtgt	acctgcgtca	27480
ccgccggtac	aagccgtatt	accggccccg	ctcgaccccc	ttccgcgaga	acaccaatag	27540
ccaaggggaa	caataccgat	gatccaagea	attgcgattg	caatcgcggg	cctcggcgcg	27600
cttctgttgt	tcatectctt	tgcccgcacc	cgcgcggtcg	atgccgaact	gaaactgaaa	27660
aagcatcggt	ccaaggacgc	cggcctggcc	gatctgctca	actacgccgc	tgctcgtgat	27720
gacggcgtaa	tcgtgggcaa	gaacggcagc	tttatggctg	cctggctgta	caaggcgcat	27780



gacaacgcaa	gcagcaccga	ccagcagcgc	gaagtagtgt	ccgcccgc	caaccaggcc	27840
ctcgcgggcc	tgggaagtgg	gtggatgac	catgtggacg	ccgtgcggcg	tctgtctccg	27900
aactacgcgg	agcggggcct	gtcggcgttc	cctgaccgtc	tgacggcagc	gattgaagaa	27960
gagcgctcgg	tcttgccctg	ctcgtcggtg	atgtacttca	ccagctccgc	gaagtcgctc	28020
ttcttgatgg	agcgcattgg	gacgtgcttg	gcaatcacgc	gcaccccccg	gccgttttag	28080
cggctaaaaa	agtcatggct	ctgccctcgg	gcggaccacg	cccatcatga	ccttgccaag	28140
ctcgtcctgc	ttctcttcga	tcttcgccag	cagggcgagg	atcgtggcat	caccgaaccg	28200
cgccgtgcgc	gggtcgtcgg	tgagccagag	ttcagcagg	ccgcccaggc	ggcccaggtc	28260
gccattgatg	cgggccagct	cgcggacgtg	ctcatagtec	acgacgcccc	tgattttgta	28320
gccctggccg	acggccagca	ggtaggccga	caggetcatg	ccggccgccc	ccgccttttc	28380
ctcaatcgct	cttcgttcgt	ctggaaggca	gtacaccttg	ataggtgggc	tgcccttctc	28440
ggttggttg	gtttcatcag	ccatccgett	gcctcatct	gttacgccgg	cggtagccgg	28500
ccagcctcgc	agagcaggat	tcccgttgag	caccgccagg	tgcaataag	ggacagtga	28560
gaaggaacac	ccgctcgcgg	gtgggcctac	ttacctate	ctgcccggct	gacgccgttg	28620
gatacaccaa	ggaaagtcta	cacgaaccct	ttggcaaaat	cctgtatata	gtgcgaaaaa	28680
ggatggatat	accgaaaaaa	tcgtataat	gaccccgaag	cagggttatg	cagcggaaaa	28740
gcgctgcttc	cctgctgttt	tgtggaatat	ctaccgactg	gaaacaggca	aatgcaggaa	28800
attactgaac	tgaggggaca	ggcgagagac	gatgccaaag	agctacaccg	acgagctggc	28860
cgagtgggtt	gaatcccgcg	cggccaagaa	gcgcccggct	gatgaggctg	cggttgcgtt	28920
cctggcggtg	agggcggatg	tcgaggcggc	gttagcgtcc	ggctatgcgc	tcgtcacct	28980
ttgggagcac	atgcgggaaa	cggggaaggt	caagttctcc	tacgagacgt	tccgctcgca	29040
cgccaggcgg	cacatcaagg	ccaagcccgc	cgatgtgccc	gcaccgcagg	ccaaggctgc	29100
ggaaccccgc	ccggcaccca	agacgccgga	gccacggcgg	ccgaagcagg	ggggcaaggc	29160
tgaaaagccg	gccccgctg	cggccccgac	cggttcacc	ttcaaccaa	caccggacaa	29220
aaaggatcta	ctgtaattgg	gaaaattcac	atggttttgc	agggcaaggg	cggggtcggc	29280
aagtcggcca	tcgcccgcat	cattgcgcag	tacaagatgg	acaaggggca	gacacccttg	29340
tgatcgaca	ccgaccgggt	gaacgcgacg	ttcaggggt	acaaggccct	gaacgtccgc	29400
cggctgaaca	tcatggccgg	cgacgaaatt	aactcgcgca	acttcgacac	cctggtcgag	29460
ctgattgcgc	cgaccaagga	tgacgtggtg	atcgacaacg	gtgccagctc	gttcgtgcct	29520
ctgtcgcatt	acctcatcag	caaccagggt	ccggtctcgc	tgcaagaaat	ggggcatgag	29580
ctggctcatc	ataccgtcgt	caccggcggc	caggtctctc	tggacacggg	gagcggcttc	29640
gcccagctcg	ccagccagtt	cccggccgaa	gcgttttctg	tggctctggct	gaaccctgat	29700
tgggggccta	tcgagcatga	gggcaagagc	tttgagcaga	tgaaggcgta	cacggccaac	29760
aaggcccgcg	tgctgtccat	catccagatt	ccggeectca	aggaagaaac	ctacggccgc	29820
gatttcagcg	acatgctgca	agagcggctg	acgttcgacc	aggcgtggc	cgatgaaatg	29880
ctcacgatca	tgacgcggca	acgcctcaag	atcgtgcggc	gcggcctgtt	tgaacagctc	29940
gacgcggcgg	ccgtgctatg	agcgaccaga	ttgaagagct	gatccgggag	attgcggcca	30000
agcacggcat	cgccgtcggc	cgcgacgacc	cggtgctgat	cctgcatacc	atcaacgccc	30060
ggctcatggc	cgacagtgcg	gccaagcaag	aggaatctct	tgccgcgttc	aaggaagagc	30120

tggaagggat	cgcccatcgt	tggggcgagg	acgccaagge	caaagcggag	cggatgctga	30180
acgcggccct	ggcggccagc	aaggacgcaa	tggcgaaggt	aatgaaggac	agcgcgcgc	30240
aggcggccga	agcgatccgc	agggaaatcg	acgacggcct	tggccgccag	ctcgcggcca	30300
aggtcgcgga	cgcgcggcgc	gtggcgaatga	tgaacatgat	cgccggcggc	atggtgtgt	30360
tcgcggccgc	cctggtgggt	tgggcctcgt	tatgaatcgc	agaggcgcag	atgaaaaagc	30420
ccggcgttgc	cgggctttgt	ttttgcgta	gctgggcttg	tttgacaggc	ccaagctctg	30480
actgcgcccg	cgctcgcgct	cctgggcctg	tttctctc	tgctcctgct	tgcgcatcag	30540
ggcctgggtg	cgctcgggctg	cttcacgeat	cgaatcccag	tcgccggcca	gctcgggatg	30600
ctccgcgcgc	atcttgcgcg	tcgccagttc	ctcgatcttg	ggcgcgtgaa	tgcccatgcc	30660
ttccttgatt	tcgcgcacca	tgtccagccg	cgtgtgcagg	gtctgcaagc	gggcttgctg	30720
ttgggcctgc	tgctgctgcc	aggcggcctt	tgtacgcggc	agggacagca	agccgggggc	30780
attggactgt	agctgctgca	aacgcgcctg	ctgacggtct	acgagctggt	ctaggcggtc	30840
ctcgatgcgc	tccacctggt	catgetttgc	ctgcacgtag	agcgaagg	tctgctggt	30900
ggtctgctcg	atgggcgcgg	attctaagag	ggcctgctgt	tccgtctcgg	cctcctgggc	30960
cgctgtagc	aaatcctcgc	cgtgtttgcc	gctggactgc	tttactgccc	gggactgctg	31020
ttgccctgct	cgcgccgctg	tcgcagttcg	gcttgcccc	actcgattga	ctgcttcatt	31080
tcgagccgca	gcgatgcgat	ctcggattgc	gtcaacggac	ggggcagcgc	ggaggtgtcc	31140
ggcttctcct	tgggtgagtc	ggtcgaatgc	atagccaaag	gtttccttcc	aaaatgcgtc	31200
cattgctgga	ccgtgtttct	cattgatgcc	cgcaagcacc	tccggttga	ccgccaggtc	31260
aagcgcgcct	tcatgggcgg	tcatgacgga	cgccgccatg	acettgccc	cgttgttctc	31320
gatgtagccg	cgtaatgagg	caatggtgcc	gcccaccgctc	agcgtgcat	cgacaacgat	31380
gtacttctgg	ccggggatca	cctccccctc	gaaagtcggg	ttgaacgcca	ggcgatgac	31440
tgaaccggct	ccggttcggg	cgacctctc	ccgctgcaca	atgtccgttt	cgacctcaag	31500
gccaaaggcgg	tcggccagaa	cgaccgccat	catggccgga	atcttgttgt	tccccgccgc	31560
ctcgacggcg	aggactggaa	cgatgcgggg	cttgcctcgc	ccgatcagcg	tcttgagctg	31620
ggcaacagtg	tcgtccgaaa	tcaggcgtc	gaccaaatta	agcggcgtt	ccgctcgc	31680
ctgcttcgca	gcctggtatt	caggctcgtt	ggtcaaagaa	ccaaggctgc	cgttgcgaa	31740
caccttcggg	aagtctcccc	acggtgcgcg	ctcggtctg	ctgtagctgc	tcaagacgcc	31800
tcccttttta	gccgctaaaa	ctctaacgag	tgcgcccgcg	actcaacttg	acgcttccgg	31860
cacttacctg	tgcttgcga	cttgcgtcat	aggtgatgct	tttgcactc	ccgatttcag	31920
gtactttatc	gaaatctgac	cgggcgtgca	ttacaaagtt	cttccccacc	tgttggtaaa	31980
tgctgccgct	atctgcgtgg	acgatgctgc	cgctcgtggc	ctgcgactta	tcggcctttt	32040
gggcatata	gatgttgtaa	atgccaggtt	tcagggeccc	ggctttatct	accttctggt	32100
tcgtccatgc	gccttgggtc	tcggtctgga	caattctttg	cccattcatg	accaggagge	32160
ggtgtttcat	tgggtgactc	ctgacggttg	cctctggtgt	taaactgtc	ctggctgctt	32220
gccggctaaa	aaaaagccga	cctcggcagt	tcgaggeccg	ctttccctag	agccgggcgc	32280
gtcaagggtg	ttccatctat	tttagtgaa	tgcgttcgat	ttatcagtta	ctttctccc	32340
gctttgtgtt	tctcccact	cgtttccgcg	tctagccgac	ccctcaacat	agcggcctct	32400
tcttgggctg	cctttgcctc	ttgccgcgct	tcgtcacgct	cggttgcac	cgctgtaaag	32460

cgctcggcct	gcctggccgc	ctcttgccgc	gccaaacttc	tttgctcctg	gtgggcctcg	32520
gcgtcggcct	gcgccttcgc	tttaccgct	gccaaactcc	tgcgaaact	ctccgcttcg	32580
cgcttggtgg	cgctcgcctc	gccgcgaagc	gcctgcattt	cctggttggc	cgctccagg	32640
gtcttgccgc	tctcttcttt	gaatgcgcgg	gcgtcctggt	gagcgtagtc	cagctcggcg	32700
cgagctcct	gcgtcgcagc	ctccacctcg	tcggcccgt	gcgtcgccag	cgcgccccgc	32760
tgctcggctc	ctgccagggc	ggtgcgtgct	tcggccaggg	cttgccgctg	gcgtcgggcc	32820
agctcggccg	cctcggcggc	ctgctgctct	agcaatgtaa	cgcgccctg	ggcttcttcc	32880
agctcgcggg	cctgcgcctc	gaaggcgtcg	gccagctccc	cgcgcacggc	ttccaactcg	32940
ttgcgctcac	gatcccagcc	ggcttgccgt	gctgcaacg	attcattggc	aagggcctgg	33000
gcggcttgcc	agagggcggc	cacggcctgg	ttgccggcct	gctgcaccgc	gtccggcacc	33060
tggactgcca	gcggggcggc	ctgcgccgtg	cgctggcgct	gccattcgcg	catgccggcg	33120
ctggcgtcgt	tcatgttgac	gcgggcggcc	ttacgcaactg	catccacggg	cggaagtctc	33180
tcccggctgc	cttgcctgaa	cagctcgtcc	gcagccgcaa	aatgcggtc	gcgcgtctct	33240
ttgttcagtt	ccatgttggc	tccgtaatt	ggtaagaata	ataaactct	tacctacctt	33300
atcagcga	gagtttagct	gaacagttct	cgacttaacg	gcaggttttt	tagcggctga	33360
agggcaggca	aaaaaagccc	cgcacggctc	gcgggggcaa	agggtcagcg	ggaaggggat	33420
tagcggggct	cgggcttctt	catgcctcgg	ggccgcgctt	cttgggatgg	agcacgacga	33480
agcgcgcacg	cgcatcgtcc	tcggccctat	cgccccgct	cgcggtcagg	aacttgctgc	33540
gcgctaggtc	ctccctgggt	ggcaccaggg	gcataaactc	ggcctgctcg	atgtaggtcc	33600
actccatgac	cgcatcgcag	tcgaggccgc	gttccttca	cgtctcttgc	aggtcgcggt	33660
acgcccgtc	gltgagcggc	tggtaacggg	ccaattggct	gtaaattggct	gtcggccatg	33720
agcggccttt	cctgttgagc	cagcagccga	cgacgaagcc	ggcaatgcag	gccccctggca	33780
caaccaggcc	gacgccgggg	gcaggggatg	gcagcagctc	gccaaccagg	aacccccgcc	33840
cgatgatgcc	gatgccggtc	aaccagccct	tgaactatc	cgccccgaa	acacccctgc	33900
gcattgcctg	gatgctgcgc	cggatagctt	gcaacatcag	gagccgtttc	ttttgttctg	33960
cagtcattgg	ccgccctcac	cagttgttcg	tatcgggtgc	ggacgaactg	aaatcgcaag	34020
agctgccggg	atcgggtccag	ccgctgtccg	tgctcgtgct	gccgaagcac	ggcgaggggt	34080
ccgcgaacgc	cgagcagggc	gtatccggcc	gcagcgcctc	gcccagcatg	gccccggctca	34140
gcgagccgcc	ggccaggtag	cccagcatgg	tgctgttggg	cgccccggcc	accagggccg	34200
acgtgacgaa	atcggcgtca	ttccctctgg	attgttctct	gctcggcggg	gcagtgccgc	34260
gcgcccggcg	cgctcgtgat	ggctcggggt	ggctggcctg	cgacggccgg	cgaaaggctc	34320
gcagcagctc	gttategacc	ggctcggcgc	tcggggccgc	cgcttgcgc	tgcggtcggt	34380
gttccttctt	cggtcgcgc	agcttgaaca	gcattgatgc	ggaaaccagc	agcaacgccg	34440
cgctacgcc	tcccgcgatg	tagaacagca	tcggattcat	tcttcggctc	tcttctgtagc	34500
ggaaccgttg	tctgtcggc	gcgggtggcc	cgccccgctg	tctttgggga	tcagccctcg	34560
atgagcgcga	ccagtttca	gtcggcaagg	ttcgctcga	actcctggcc	gtcgtcctcg	34620
tacttcaacc	agcatagcc	ttccgccggc	ggccgacggg	tgaggataag	gcgggcaggg	34680
cgctcgtcgt	gctcagactg	gacgatggcc	tttttcagct	tgtccgggtc	cggtctcttc	34740
gcgccctttt	ccttggcgtc	cttaccgtcc	tggtcgcctg	cctcgcctgc	ctggccgtcg	34800

ccggcctccg	cgtcacgctc	ggcatcagtc	tggccgttga	aggcatcgac	ggtgttggga	34860
tcgcggccct	tctcgtccag	gaactcgcgc	agcagettga	ccgtgccgcg	cgtagatttc	34920
tgggtgtcgt	cgtaagcca	cgcctcgact	tcctccgggc	gcttcttgaa	ggccgtcacc	34980
agctcgttca	ccacggtcac	gtcgcgcacg	cggccggtgt	tgaacgcate	ggcgatcttc	35040
tccggcaggt	ccagcagcgt	gacgtgctgg	gtgatgaacg	ccggcgactt	gccgatttcc	35100
ttggcgatat	cgcttttctt	cttgcccttc	gccagctcgc	ggccaatgaa	gtcggcaatt	35160
tcgcgcgggg	tcagctcgtt	gcgttgcaag	ttctcgataa	cctggtcggc	ttcgtttag	35220
tcgttgctga	tgaacgcccg	gatggacttc	ttgccggccc	acttcgagcc	acggtagcgg	35280
cgggcgccgt	gattgatgat	atagcggccc	ggctgctcct	ggttctcgcg	caccgaaatg	35340
ggtgacttca	ccccgcgctc	tttgatcgtg	gcaccgattt	ccgcgatgct	ctccggggaa	35400
aagccggggg	tgteggccgt	ccgcggctga	tgcggatctt	cgtegateag	gtccagggtcc	35460
agctcgatag	ggccggaacc	gccttgagac	gccgcaggag	cgteccaggag	gctcgcacagg	35520
tcgccgatgc	tatccaaccc	caggccggac	ggctgcgcgc	cgcttcgggc	ttcctgagcg	35580
gccgcagcgg	tgTTTTtctt	ggtggtcttg	gcttgagccg	cagtcattgg	gaaatctcca	35640
tcttcgtgaa	cacgtaatca	gccagggcgc	gaacctcttt	cgatgccttg	cgcgcggccc	35700
TTTTcttgat	cttcagacc	ggcacaccgg	atgcgagggc	atcggcgatg	ctgctgcgca	35760
ggccaacggg	ggccggaatc	atcatcttgg	ggtacgcggc	cagcagctcg	gcttggtggc	35820
gcgcgtggcg	cggattccgc	gcacgcacct	tgctgggcac	catgccaagg	aattgcagct	35880
tggcgttctt	ctggcgcacg	ttcgcaatgg	tcgtgacat	cttcttgatg	ccctggatgc	35940
tgtacgcctc	aagctcgatg	ggggacagca	catagtcggc	cgcaagagg	gcggccgcca	36000
ggccgacgcc	aagggctcggg	gccgtgtcga	tcaggcacac	gtcgaagcct	tggttcgcca	36060
gggccttgat	gttcgccccg	aacagctcgc	ggcgctcgtc	cagcgcacgc	cgttcggcgt	36120
tcgccagtac	cgggttgac	tcgatgaggg	cgagcgcgcg	ggcctggccg	tcgccggctg	36180
cgggtgcggg	ttcgggtccag	ccgccggcag	ggacagcgcg	gaacagcttg	cttgcatgca	36240
ggccggtagc	aaagtccttg	agcgtgtagg	acgcattgcc	ctgggggtcc	aggtcgatca	36300
cggcaacccg	caagccgcgc	tcgaaaaagt	cgaaggcaag	atgcacaagg	gtcgaagtct	36360
tgccgacgcc	gcctttctgg	ttggccgtga	ccaaagtttt	catcgttttg	tttctgttt	36420
tttcttgccg	tccgcttccc	acttcgggac	gatgtacgcc	tgatgttccg	gcagaaccgc	36480
cgttaccgcc	gcgtaccctt	cgggcaagtt	cttgcctcgc	aacgcggccc	acacgcgatg	36540
caccgcttgc	gacactgcgc	ccctggtcag	tcccagcgcg	gttgcgaacg	tcgctgtgg	36600
cttcccatcg	actaagacgc	cccgcgtat	ctcgatggte	tgctgcecca	cttcagccc	36660
ctggatcgcc	tcctggaact	ggctttcggg	aagccgtttc	ttcatggata	acaccataa	36720
tttgctccgc	gccttggttg	aacatagcgg	tgacagccgc	cagcacatga	gagaagttta	36780
gctaaacatt	tctcgcacgt	caacaccttt	agccgctaaa	actcgtcctt	ggcgtaaaca	36840
aacaaaagcc	cggaaaccgg	gctttcgtct	cttgccgctt	atggctctgc	accgcgctcc	36900
atcaccaaca	ggtcgcgcac	gccttcaact	cggttgcgga	tcgacactgc	cagcccaaca	36960
aagccggttg	ccgccgccgc	caggatcgcg	ccgatgatgc	cggccacacc	ggccatcgcc	37020
caccaggtcg	ccgccttccg	gttccattcc	tgctggtaact	gcttcgcaat	gctggacctc	37080
ggctcacat	aggctgaccg	ctcgatggcg	tatgccgctt	ctccccttgg	cgtaaaacce	37140

agcgccgcag	gcggcattgc	catgctgccc	gccgetttcc	cgaccacgac	gcgcgcacca	37200
ggcttgccgt	ccagaccttc	ggccacggcg	agctgcgcaa	ggacataatc	agccgccgac	37260
ttggctccac	gcgcctcgat	cagctcttgc	actcgcgcga	aatccttggc	ctccacggcc	37320
gccatgaatc	gcgcacgcgg	cgaaggctcc	gcagggccgg	cgctcgtgatc	gccgccgaga	37380
atgcccttca	ccaagttcga	cgacacgaaa	atcatgctga	cggctatcac	catcatgcag	37440
acggatcgca	cgaaccgcgt	gaattgaaca	cgagcacggc	acccgcgacc	actatgccaa	37500
gaatgcccaa	ggtaaaaatt	gccggccccg	ccatgaagtc	cgtgaatgcc	ccgacggccg	37560
aagtgaaggg	caggccgcca	cccaggccgc	cgccctcact	gcccggcacc	tggtcgcgtga	37620
atgtcgatgc	cagcacctgc	ggcacgtcaa	tgttccggg	cgctcgcctc	gggctgatcg	37680
cccatcccgt	tactgccccg	atcccggcaa	tggcaaggac	tgccagcgt	gccatttttg	37740
gggtgaggcc	gttcgcgccc	gaggggcgca	gcccctgggg	ggatgggagg	cccgcgttag	37800
cgggccggga	gggttcgaga	agggggggca	cccccttcg	gctgcgcgg	tcacgcgcac	37860
agggcgcagc	cctggttaaa	aacaagttt	ataaatattg	gtttaaaagc	aggttaaaag	37920
acaggttagc	ggtggccgaa	aaacgggcgg	aaaccttgc	aatgctgga	ttttctgct	37980
gtggacagcc	cctcaaatgt	caataggtgc	gcccctcact	tgctcagcact	ctgcccctca	38040
agtgtcaagg	atcgcgcccc	tcctctgtca	gtagtcgcgc	ccctcaagtg	tcaataccgc	38100
agggcactta	tcccagget	tgtccacatc	atctgtggga	aactcgcgta	aaatcaggcg	38160
ttttcgccga	tttgcgagge	tggccagctc	cacgtcgccg	gccgaaatcg	agcctgcccc	38220
tcctctgtca	acgccgcgcc	gggtgagtcg	gcccctcaag	tgtaacgctc	cgcccctcat	38280
ctgtcagtga	gggccaagtt	ttccgcgagg	tatccacaac	gccggcggcc	gcggtgtctc	38340
gcacacggct	tcgacggcgt	ttctggcgcg	tttgcagggc	catagacggc	cgccagccca	38400
gcggcgaggg	caaccagccc	ggtgagcgtc	ggaaaggcgc	tggaagcccc	gtagcgacgc	38460
ggagaggggg	gagacaagcc	aagggcgag	gctcgatgcg	cagcacgaca	tagccggttc	38520
tcgcaaggac	gagaatttcc	ctgcggtgcc	cctcaagtgt	caatgaaagt	ttccaacgcg	38580
agccattcgc	gagagccttg	agtccacgct	agatgagagc	tttgtttag	gtggaccagt	38640
tggtgatttt	gaacttttgc	tttgccacgg	aacggtctgc	gttgtcggga	agatgcgtga	38700
tctgatcctt	caactcagca	aaagtctgat	ttattcaaca	aagccacgtt	gtgtctcaaa	38760
atctctgatg	ttacattgca	caagataaaa	atataatc	atgaacaata	aaactgtctg	38820
cttacataaa	cagtaataca	aggggtgta	tgagccatat	tcaacgggaa	acgtcttgct	38880
cgactctaga	gctcgttctc	cgaggcctcg	aggcctcgag	gaacggtacc	tgccgggaag	38940
cttacaataa	tgtgtgttgt	taagtcttgt	tgctgtctat	cgtctgactg	actttcgtca	39000
taaatcccgg	cctccgtaac	ccagctttgg	gcaagctcac	ggatttgatc	cgccgggaacg	39060
ggaatatcga	gatgccgggc	tgaacgctgc	agttccagct	ttcccttctg	ggacaggtac	39120
tccagctgat	tgattatctg	ctgaagggtc	ttggttccac	ctcctggcac	aatgcgaatg	39180
attacttgag	cgcgatcggg	catccaattt	tctcccgtca	ggtgcgtggt	caagtgtctac	39240
aaggcacctt	tcagtaacga	gcgaccgtcg	atccgtcgcc	gggatacggg	caaaatggag	39300
cgcagtagtc	catcgagggc	ggcgaaaagc	tcgccccaaag	caatacgttc	atctcgcaca	39360
gcctccagat	ccgatcgagg	gtcttcggcg	taggcagata	gaagcatgga	tacattgctt	39420
gagagtattc	cgatggactg	aagtatggct	tccatctttt	ctcgtgtgctc	tgcatctatt	39480

tcgagaaagc	ccccgatgcg	gcgaccgca	acgcgaattg	ccatactatc	cgaaagtccc	39540
agcaggcgcg	cttgatagga	aaaggtttca	tactcggccg	atcgagacg	ggcactcacg	39600
accttgaacc	cttcaacttt	cagggatcga	tgctggttga	tggtagtctc	actcgacgtg	39660
gctctgggtg	gttttgacat	agcttcctcc	aaagaaagcg	gaaggtctgg	atactccagc	39720
acgaaatgtg	cccgggtaga	cggatggaag	tctagccctg	ctcaatatga	aatcaacagt	39780
acatttacag	tcaatactga	atatacttgc	tacatttgcg	attgtcttat	aacgaatgtg	39840
aaataaaaaat	agtgtaacaa	cgcttttact	catcgataat	cacaaaaaca	tttatacgaa	39900
caaaaataca	aatgcactcc	ggtttcacag	gataggcggg	atcagaatat	gcaacttttg	39960
acgttttgtt	ctttcaaagg	gggtgctggc	aaaaccaccg	cactcatggg	cctttgcgct	40020
gctttggcaa	atgacggtaa	acgagtggcc	ctctttgatg	ccgacgaaaa	ccggcctctg	40080
acgcgatgga	gagaaaaagc	cttacaagc	agtactggga	tectcgtctg	gaagtctatt	40140
ccgccgacga	aatgccccct	cttgaagcag	cctatgaaaa	tgccgagctc	gaaggatttg	40200
attatgcgtt	ggccgatacg	cgtggcgget	cgagcgagct	caacaacaca	atcatcgcta	40260
gctcaaacct	gcttctgac	cccaccatgc	taacgccgct	cgacatcgat	gaggcactat	40320
ctacctaccg	ctacgtcacc	gagctgctgt	tgagtgaaaa	tttgcaatt	cctacagctg	40380
ttttgcgcca	acgcgtcccg	gtcggccgat	tgacaacatc	gcaacgcagg	atgtcagaga	40440
cgctagagag	ccttccagtt	gtaccgtctc	ccatgcatga	aagagatgca	tttgccgca	40500
tgaaagaacg	cggcatgttg	catcttacct	tactaaacac	gggaactgat	ccgacgatgc	40560
gcctcataga	gaggaatctt	cggattgca	tggaggaagt	cgtggtcatt	tcgaaactga	40620
tcagcaaaaat	cttggaggct	tgaagatggc	aattcgcaag	cccgcattgt	cggtcggcga	40680
agcacggcgg	cttgctgggt	ctcgaccgca	gatccaccat	cccaaccgca	cacttgttcc	40740
ccagaagctg	gacctccagc	acttgcctga	aaaagccgac	gagaaagacc	agcaacgtga	40800
gcctctcgtc	gccgatcaca	tttacagtcc	cgatcgacaa	cttaagctaa	ctgtggatgc	40860
ccttagtcca	cctccgtccc	cgaaaaagct	ccaggttttt	ctttcagcgc	gaccgcccgc	40920
gcctcaagtg	tcgaaaacat	atgacaacct	cgttcggcaa	tacagtcctt	cgaagtgcgt	40980
acaaatgatt	ttaaggcgcg	cgttggacga	tttcgaaagc	atgctggcag	atggatcatt	41040
tcgcgtggcc	ccgaaaagtt	atccgatccc	ttcaactaca	gaaaaatccg	ttctcgttca	41100
gacctcacgc	atgttcccgg	ttgcgttget	cgaggteget	cgaagtcatt	ttgatccggt	41160
ggggttggag	accgctcgag	ctttcggcca	caagctggct	accgccgcgc	tcgcgtcatt	41220
ctttgctgga	gagaagccat	cgagcaattg	gtgaagaggg	acctatcgga	accctcacc	41280
aaatattgag	tgtaggtttg	aggccgctgg	ccgcgtctct	agtcaccttt	tgagccagat	41340
aattaagagc	caaatgcaat	tggetcaggc	tgccatcgtc	ccccgtgcg	aaacctgcac	41400
gtccgcgtca	aagaaataac	cggcaccctt	tgetgttttt	atcagttgag	ggcttgacgg	41460
atccgcctca	agtttgcggc	gcagccgcaa	aatgagaaca	tctatactcc	tgctgtaaac	41520
ctcctcgtcg	cgtaactcgac	tggcaatgag	aagttgctcg	cgcgatagaa	cgctcggggg	41580
tttctctaaa	aacgcgagga	gaagattgaa	ctcaccctgc	gtaagtttca	cctcaccgcc	41640
agcttcggac	atcaagcgac	gttgcctgag	attaagtgtc	cagtcagtaa	aacaaaaaga	41700
ccgtcggctt	ttggagcgga	caacgttggg	gcgcacgcgc	aaggcaacc	gaatgcgtgc	41760
aagaaactct	ctcgtactaa	acggcttagc	gataaaatca	cttgetccta	gctcagatgc	41820

aacaacttta	tccgtctcct	caaggcggtc	gccactgata	attatgattg	gaatatcaga	41880
ctttgccgcc	agatttcgaa	cgatctcaag	cccatettca	cgacctaaat	ttagatcaac	41940
aaccacgaca	tcgaccgtcg	cggaagagag	tactctagtg	aactgggtgc	tgctcggtac	42000
cgcggtcact	ttgaaggcgt	ggatcgtaa	gtattcgata	ataagatgcc	gcatagcgac	42060
atcgtcatcg	ataagaagaa	cgtgtttcaa	cggtcacct	ttcaatctaa	aatctgaacc	42120
cttgttcaca	gcgcttgaga	aattttcacg	tgaaggatgt	acaatcatct	ccagctaaat	42180
gggcagttcg	tcagaattgc	ggctgaccgc	ggatgacgaa	aatgcgaacc	aagtatttca	42240
atthttatgac	aaaagtctc	aatcgttggt	acaagtga	cgcttcgagg	ttacagctac	42300
tattgattaa	ggagatcgcc	tatggtctcg	ccccggcgtc	gtgcgtccgc	cgcgagccag	42360
atctcgccta	cttcataaac	gtcctcatag	gcacggaatg	gaatgatgac	atcgatcgcc	42420
gtagagagca	tgtaaatcag	tgtgcgatct	tccaagctag	caccttgggc	gctacttttg	42480
acaagggaaa	acagtttctt	gaatccttgg	attggatteg	cgccgtgtat	tgttgaaatc	42540
gatcccggat	gtcccagac	gaettcaetc	agataagccc	atgctgcatc	gtcgcgcatc	42600
tcgccaagca	atatccggtc	cgcccgcata	cgcagacttg	cttgagcaa	gtgctcggcg	42660
ctcacagcac	ccagcccage	accgttcttg	gagtagagta	gtctaacatg	attatcgtgt	42720
ggaatgacga	gttcgagcgt	atcttctatg	gtgattagcc	tttctgggg	ggggatggcg	42780
ctgatcaagg	tcttgetcat	tgttgtcttg	ccgttccgg	tagggccaca	tagcaacatc	42840
gtcagtcggc	tgacgacgca	tgcgtgcaga	aacgcttcca	aatccccgtt	gtcaaatgc	42900
tgaaggatag	cttcatcadc	ctgattttgg	cgtttcttc	gtgtctgcca	ctggttccac	42960
ctcgaagcat	cataacggga	ggagacttct	ttaagaccag	aaacacgca	gcttggccgt	43020
cgaatggtca	agctgacggt	gcccagggga	acggtcggcg	gcagacagat	ttgtagtcgt	43080
tcaccaccag	gaagttcagt	ggcgcagagg	gggttacgtg	gtccgacatc	ctgctttctc	43140
agcgcgcccc	ctaaaatagc	gatatcttca	agatcatcat	aagagacggg	caaaggcatc	43200
ttggtaaaaa	tgccggcttg	gcgcaaaaat	gcctctccag	gtcgattgat	cgcaatttct	43260
tcagtcttcg	ggatcatcgag	ccattccaaa	atcggcttca	gaagaaagcg	tagttgcgga	43320
tccacttcca	tttacaatgt	atcctatctc	taagcggaaa	tttgaattca	ttagagcgg	43380
cggttctctc	cccgcgtggc	gccgccagtc	aggcggagct	ggtaaacacc	aaagaaatcg	43440
aggctcccgtg	ctacgaaaat	ggaacggtg	tcacctgat	tcttcttcag	ggttggcgg	43500
atgttgatgg	ttgccttaag	ggctgtctca	gttgtctgct	caccgttatt	ttgaaagctg	43560
ttgaagctca	tcccgcacc	cgagctccg	gcgtaggtgc	tagctgcctg	gaaggcgcct	43620
tgaacaacac	tcaagagcat	agctccgcta	aaacgctgcc	agaagtggct	gtcgaccgag	43680
cccggcaatc	ctgagcgacc	gagttcgtcc	gcgcttggcg	atgttaacga	gatcatcgca	43740
tggtcaggtg	tctcggcgcg	atcccacaac	acaaaaacgc	gcccctctcc	ctgttgcaag	43800
ccacgctgta	tttcgccaac	aacggtgggtg	ccacgatcaa	gaagcacgat	attgttctgt	43860
gttccacgaa	tatcctgagg	caagacacac	tttacaatagc	ctgccaatt	tgtgtcgatt	43920
gcggtttgca	agatgcacgg	aattattgtc	cettgcttca	ccataaaatc	ggggtgcggc	43980
aagagcgtgg	cgctgctggg	ctgcagctcg	gtgggtttca	tacgtatcga	caaatcgttc	44040
tcgccggaca	cttcgccatt	cggcaaggag	ttgtcgtcac	gcttgccttc	ttgtcttcgg	44100
cccgtgtcgc	cctgaatggc	gcgtttgctg	acccttcat	cgccgctgct	atatgcaaaa	44160

atcgggtgtt	cttccggccg	tggctcatgc	cgctccggtt	cgccccctcg	cggttagagga	44220
gcagcaggct	gaacagcctc	ttgaaccgct	ggaggatccg	gcggcacctc	aatcggagct	44280
ggatgaaatg	gcttgggtgt	tgttgcgac	aaagttgacg	gcgatgcggt	ctcattcacc	44340
ttcttttggc	gcccacctag	ccaaatgagg	cttaatgata	acgcgagaac	gacacctccg	44400
acgatcaatt	tctgagacc	cgaaagacgc	cggcgatggt	tgtcggagac	cagggatcca	44460
gatgcatcaa	cctcatgtgc	cgcttgetga	ctatcgttat	tcatcccttc	gcccccttca	44520
ggacgcggtt	cacatcgggc	ctcaccgtgc	ccgtttgcgg	cctttggcca	acgggatcgt	44580
aagcgggtgt	ccagatacat	agtactgtgt	ggccatccct	cagacgcca	cctcgggaaa	44640
ccgaagaaat	ctcgacatcg	ctccccttaa	ctgaatagtt	ggcaacagct	tccttgccat	44700
caggattgat	ggtgtagatg	gagggtatgc	gtacattgcc	cggaaagtgg	aataccgtcg	44760
taaatccatt	gtcgaagact	tcgagtggca	acagcgaacg	atcgcttgg	gcgacgtagt	44820
gccaattact	gtccgcccga	ccaagggetg	tgacaggetg	atccaataaa	ttctcagctt	44880
tccgttgata	ttgtgcttcc	gcgtgtagtc	tgccacaac	agccttctgt	tgtgctctcc	44940
ttcggcgagc	cgccgcacg	tcggcggggt	aggcgaattg	gacgctgtaa	tagagatcgg	45000
gctgctcttt	atcgaggtgg	gacagagtct	tggaacttat	actgaaaaca	taacggcgca	45060
tcccggagtc	gcttgcggtt	agcacgatta	ctggctgagg	cgtgaggacc	tggcttgctt	45120
tgaaaaatag	ataatttccc	cgcggtaggg	ctgctagatc	tttgcatttt	gaaacggcaa	45180
ccgctgtcac	cgtttcgttc	gtggcgaatg	ttacgaccaa	agtagctcca	accgccgtcg	45240
agaggcgcac	cacttgatcg	ggattgtaag	caaataaacg	catgcgcgga	tctagcttgc	45300
ccgccattgg	agtgtcttca	gcctccgcac	cagtcgcagc	ggcaataaaa	catgctaaaa	45360
tgaaaagtgc	ttttctgac	atggttcgct	gtggcctacg	tttgaaacgg	tatcttccga	45420
tgtctgatag	gaggtgacaa	ccagacctgc	cgggttggtt	agtctcaatc	tgccgggcaa	45480
gctggtcacc	ttttctgac	gaactgtcgc	ggtccacgta	ctcaccacag	gcattttgcc	45540
gtcaacgacg	agggtccttt	tatagcgaat	ttgctgcgtg	cctggagtta	catcatttga	45600
agcgatgtgc	tcgacctcca	cctgcccgcg	tttgccaaga	atgacttgag	gcgaactggg	45660
attgggatag	ttgaagaatt	gctggtaatc	ctggcgcact	gttggggcac	tgaagtctga	45720
taccaggctcg	taggcgtact	gagcgggtgc	ggcatcataa	ctctcgcgca	ggcgaacgta	45780
ctcccacaat	gaggcgtaa	cgacggcctc	ctcttgagtt	gcaggcaatc	gcgagacaga	45840
cacctcgctg	tcaacgggtc	cgctccggccg	tatccataga	tatacgggca	caagcctgct	45900
caacggcacc	attgtggcta	tagcgaacgc	ttagcaaca	tttcccaaaa	tcgcatagc	45960
tgcgacagct	gcaatgagtt	tggagagacg	tcgcgccgat	ttegtctcgcg	cggtttgaaa	46020
ggcttctact	tccttatagt	gctcggcaag	gctttcgcgc	gccactagca	tggcatatct	46080
aggccccgtc	atagcgtcca	cccgaattgc	cgagctgaag	atctgacgga	gtaggctgcc	46140
atgccccac	attcagcggg	aagatcgggc	ccttgcaget	cgctaatgtg	tcgtttgtct	46200
ggcagccgct	caaagcgaca	actaggcaca	gcaggcaata	cttcatagaa	ttctccattg	46260
aggcgaattt	ttgcgcgacc	tagcctcget	caacctgagc	gaagcgacgg	tacaagctgc	46320
tggcagattg	ggttgcgccg	ctccagtaac	tgctccaat	gttgccggcg	atgccggca	46380
aagcgacaat	gagcgcaccc	cctgtcagaa	aaaacatctc	gagttcgtaa	agaccaatga	46440
tcttggccgc	ggtcgtaccg	gcgaaggtga	ttacaccaag	cataagggtg	agcgcagctc	46500



cttcggttag	gatgacgac	gttgccacga	ggtttaagag	gagaagcaag	agaccgtagg	46560
tgataagttg	cccgatccac	ttagctgcca	tgccccgcgt	gcgatcaaaa	atatatccga	46620
cgaggatcag	aggccccgat	gcgagaagca	ctttcgtgag	aattccaacg	gcgtcgtaaa	46680
ctccgaaggc	agaccagagc	gtgccgtaaa	ggaccactg	tgcccccttg	aaagcaagga	46740
tgtcctggtc	gttcatcgga	ccgatttcgg	atgctgattt	ctgaaaaacg	gcctgggtca	46800
cggcgaacat	tgtatccaac	tgtgccggaa	cagtctgcag	aggcaagccg	gttactactaa	46860
actgctgaac	aaagtttggg	accgtctttt	cgaagatgga	aaccacatag	tcttggtagt	46920
tagcctgccc	aacaattaga	gcaacaacga	tggtgaccgt	gatcacccga	gtgataccgc	46980
tacgggtatc	gacttcgccc	cgtatgacta	aaataacctg	aacaataatc	caaagagtga	47040
cacaggcgat	caatggcgca	ctcaccgect	cctggatagt	ctcaagcadc	gagtccaagc	47100
ctgtcgtgaa	ggctacatcg	aagatcgtat	gaatggccgt	aaacggcgcc	ggaatcgtga	47160
aattcatcga	ttggacctga	acttgactgg	tttgctgcat	aatggttgat	aaaatgagct	47220
cgcatcggc	gaggatgcgg	gcggatgaac	aaatcgccca	gccttagggg	agggcaccaa	47280
agatgacagc	ggctctttga	tgtctcttgc	gttgagcggc	cgctcttcc	gcctcgtgaa	47340
ggccggcctg	cgcggtagtc	atcgtaata	ggcttgctgc	ctgtacattt	tgaatcattg	47400
cgatcatggat	ctgcttgaga	agcaaaccat	tggtcacggt	tgcttgcctg	atattgcgag	47460
atcgggaaag	ctgagcagac	gtatcagcat	tcgccgctca	gcgtttgtcc	atcgtttcca	47520
gattgtcagc	cgcaatgcca	gcgctgtttg	cggaaccggt	gatctgcgat	cgcaacaggt	47580
ccgcttcagc	atcactacc	acgactgcac	gatctgtatc	gctggtgatc	gcacgtgccg	47640
tggtcgacat	tgccattcgc	ggcgaaaaca	tttcattgtc	taggtccttc	gtcgaaggat	47700
actgatTTTT	ctggttgagc	gaagtcagta	gtccagtaac	gccgtaggcc	gacgtcaaca	47760
tcgtaaccat	cgctatagtc	tgagtgagat	tctccgcagt	cgcgagcgca	gtcgcgagcg	47820
tctcagcctc	cgttgccggg	tcgtaacaa	caaactgcgc	ccgcgcgggc	tgaatatata	47880
gaaagctgca	ggtaaaaact	gttgcaataa	gttgctgctg	cttcatcggt	tcctacctta	47940
tcaatcttct	gcctcgtggt	gacgggccat	gaattcgtctg	agccagccag	atgagttgcc	48000
ttcttgtgcc	tcgcgtagtc	gagttgcaa	gcgcaccgtg	ttggcacgcc	ccgaaagcac	48060
ggcgacatat	tcacgcata	cccgcagatc	aaattcgcag	atgacgcttc	cactttctcg	48120
tttaagaaga	aacttacggc	tgccgaccgt	catgtcttca	cggatcgctt	gaaattcctt	48180
ttcggtagat	ttcagttccat	cgacataagc	cgatcgatct	gcggttggtg	atggatagaa	48240
aatcttcgct	atacattgcg	caaccaagct	ggctcctagc	ggcgattcca	gaacatgctc	48300
tggttgctgc	gttgccagta	ttagcatccc	gttgTTTTTT	cgaacggtea	ggaggaattt	48360
gtcgcagaca	gtcgaaaatt	tagggtttaa	caaataggcg	cgaactcat	cgagctcat	48420
cacaaaacgg	cgcccgctga	tcatggtctc	aatccgatgc	aggagatatg	ctgcagcggg	48480
agcgcatact	tcctcgtatt	cgagaagatg	cgatcatgctg	aagccggtaa	tcgacggatc	48540
taactttact	tcgtcaactt	cgccgtcaaa	tgcccagcca	agcgcattggc	cccggcacca	48600
gcgttgaggc	cgctctctg	cgcttcggc	gggcccctgc	aacaaaaatt	cacgtaacct	48660
cgcgattgaa	cgcatTTTg	gatcaaacga	gagctgacga	tggataccac	ggaccagacg	48720
gcggttctct	tccggagaaa	tcccacccc	accatcactc	tcgatgagag	ccacgatcca	48780
ttcgcgcaga	aaatcgtgtg	aggctgctgt	gttttctagg	ccacgcaacg	gcgccaacct	48840

gctgggtgtg	cctctgtgaa	gtgcaaata	tgttctctct	gtggcgcgaa	ccagcaatc	48900
gccaccccg	tccttgtcaa	agaacacgac	cgtacctgca	cggtcgacca	tgctctgttc	48960
gagcatggct	agaacaaaca	tcatgagcgt	cgtcttacc	ctcccgatag	gcccgaatat	49020
tgccgtcatg	ccaacatcgt	gctcatgcgg	gatatagtcg	aaaggcggtc	cgccattggt	49080
acgaaatcgg	gcaatcgcgt	tgccccagt	gcctgagctg	gcgccctctg	gaaagttttc	49140
gaaagagaca	aaccctgcga	aattgcgtga	agtgattgcg	ccagggcggt	tgcgccactt	49200
aaaattcccc	ggcaattggg	accaatagc	cgcttccata	ccaatacctt	cttgacaac	49260
cacggcacct	gcatccgcca	ttcgtgtccg	agccccgcgc	cccctgtccc	caagactatt	49320
gagatcgtct	gcatagacgc	aaaggctcaa	atgatgtgag	cccataacga	attcgttgc	49380
cgcaagtgcg	tcctcagcct	cggataat	gccgatttga	gtcacggctt	tatcgccgga	49440
actcagcatc	tggtctgatt	tgaggctaag	tttcgcgtgc	gcttgcgggc	gagtcaggaa	49500
cgaaaaactc	tgctgagaa	caagtggaaa	atcgagggat	agcagcgctg	tgagcatgcc	49560
cggccgtgtt	tttgacgggt	attcgcgaaa	cgaatagatg	gatccaacgt	aactgtcttt	49620
tggcgttctg	atctcgagtc	ctcgettgcc	gcaaatgact	ctgtcggtat	aaatcgaagc	49680
gccgagttag	ccgtgacga	ccggaaccgg	tgtgaaccga	ccagtcatga	tcaaccgtag	49740
cgcttcgcca	atttcggtga	agagcacacc	ctgcttctcg	cggatgcca	gacgatgcag	49800
gccatacgtc	ttaagagagc	cagcgacaac	atgccaaaga	tcttccatgt	tctgatctg	49860
gcccgtgaga	tcgttttccc	ttttccgct	tagcttgggt	aacctctct	ttaccttccc	49920
taaagccgcc	tgtgggtaga	caatcaacgt	aaggaagtgt	tcattgcgga	ggagttggcc	49980
ggagagcacg	cgctgttcaa	aagcttcgtt	caggctagcg	gcgaaaacac	tacggaagtg	50040
tcgcggcgcc	gatgatggca	cgctggcatg	acgtacgagg	tgagcatata	ttgacacatg	50100
atcatcagcg	atattgcgca	acagcgtggt	gaacgcacga	caacgcgat	tgcgatttc	50160
agtttctca	agctcgaatg	caacgccatc	aattctcgca	atggtcatga	tcgatccgtc	50220
ttcaagaagg	acgatattgt	cgctgagggt	gccaatataa	gggagataga	tctcaccgga	50280
tctttcggtc	gttccactcg	cgccgagcat	cacaccattc	ctctccctcg	tgggggaacc	50340
ctaattggat	ttgggctaac	agtagcgc	ccccaaactg	cactatcaat	gcttcttccc	50400
gcggtccgca	aaaaatagcag	gacgacgctc	gccgcattgt	agtctcgctc	cacgatgagc	50460
cgggctgcaa	accataacgg	cacgagaacg	acttcgtaga	gcgggttctg	aacgataacg	50520
atgacaaaagc	cggcgaacat	catgaataac	ctgccaatg	tcagtggcac	ccaagaaac	50580
aatgcgggcc	gtgtggctgc	gaggtaaagg	gtcgattctt	ccaaacgatc	agccatcaac	50640
taccgccagt	gagcgtttgg	ccgaggaagc	tcgccccaaa	catgataaca	atgccgccga	50700
cgacgccggc	aaccagccca	agcgaagccc	gcccgaacat	ccaggagatc	ccgatagcga	50760
caatgccgag	aacagcgagt	gactggccga	acggaccaag	gataaacgtg	catatattgt	50820
taaccattgt	ggcggggtea	gtgccgccac	ccgcagattg	cgctgcggcg	ggtccggatg	50880
aggaaatgct	ccatgcaatt	gcaccgcaca	agcttggggc	gcagctcgat	atcacgcgca	50940
tcatcgcatt	cgagagcgag	aggcgattta	gatgtaaacg	gtatctctca	aagcatcgca	51000
tcaatgcgca	cctccttagt	ataagtcgaa	taagacttga	ttgtcgtctg	cggatttgc	51060
gttgtcctgg	tgtggcggtg	gcggagcgat	taaaccgcca	gcgccatcct	cctgcgagcg	51120
gcgctgatat	gacccccaaa	catcccacgt	ctcttcggat	tttagcgct	cgtgatcgtc	51180

ttttggaggc	tcgattaacg	cgggcaccag	cgattgagca	gctgtttcaa	cttttcgcac	51240
gtagccgttt	gcaaaaccgc	cgatgaaatt	accggtgttg	taagcggaga	tcgcccgcac	51300
aagcgcaaat	tgcttctcgt	caatcgtttc	gccgcctgca	taacgacttt	tcagcatggt	51360
tgcagcggca	gataatgatg	tgcacgcctg	gagcgcaccg	tcaggtgtca	gaccgagcat	51420
agaaaaattt	cgagagtta	tttgcattgag	gccaacatcc	agcgaatgcc	gtgcatcgag	51480
acggtgcctg	acgacttggg	ttgcttggct	gtgatcttgc	cagtgaagcg	tttcgccggt	51540
cgtgttgca	tgaatcgcta	aaggatcaaa	gcgactctcc	accttagcta	tcgccgcaag	51600
cgtagatgtc	gcaactgatg	gggcacactt	gcgagcaaca	tggtaaact	cagcagatga	51660
gagtgccgtg	gcaaggctcg	acgaacagaa	ggagaccatc	aaggcaagag	aaagcgacc	51720
cgatctctta	agcatacctt	atctccttag	ctcgaacta	acaccgctc	tcccgttgga	51780
agaagtgcgt	tgttttatgt	tgaagattat	cgggagggtc	ggttactcga	aaattttcaa	51840
ttgcttcttt	atgatttcaa	ttgaagcgag	aaacctcgcc	cggcgtcttg	gaacgcaaca	51900
tggaccgaga	accgcgcate	catgactaag	caaccggatc	gacctattca	ggccgcagtt	51960
ggtcaggtea	ggctcagaac	gaaaatgctc	ggcgaggtta	cgtgtctgt	aaaccatc	52020
gatgaacggg	aagcttcctt	ccgattgctc	ttggcaggaa	tattggcca	tgcctgcttg	52080
cgctttgcaa	atgctcttat	cgcgttggtg	tcatatgctt	tgtccgccag	cagaaacgca	52140
ctctaagcga	ttatttgtaa	aaatgtttcg	gtcatgcggc	ggtcatgggc	ttgaccgct	52200
gtcagcgcaa	gacggatcgg	tcaaccgtcg	gcatcgacaa	cagcgtgaat	cttggtggtc	52260
aaaccgccac	gggaacgtcc	catacagcca	tcgtcttgat	cccgtgttt	cccgtcgccg	52320
catgttggtg	gacgcggaca	caggaactgt	caatcatgac	gacattctat	cgaaagcctt	52380
ggaaatcaca	ctcagaatat	gatcccagac	gtctgcctca	cgccatcgta	caaagcgatt	52440
gtagcagggt	gtacaggaac	cgtatcgatc	aggaacgtct	gcccaggcg	ggcccgtccg	52500
gaagcggcac	aagatgacat	tgatcacccg	cgtaacgcg	cggcacgcga	cgcggttat	52560
ttgggaacaa	aggactgaac	aacagtccat	tcgaaatcgg	tgacatcaaa	gcggggacgg	52620
gttatcagtg	gcctccaagt	caagcctcaa	tgaatcaaaa	tcagaccgat	ttgcaaacct	52680
gatttatgag	tgtgcggcct	aatgatgaa	atcgtccttc	tagatgcct	ccgtggtgta	52740
gcaacacctc	gcagtatcgc	cgtgctgacc	ttggccaggg	aattgactgg	caagggtgct	52800
ttcacatgac	cgctcttttg	gccgcgatag	atgatttcgt	tgtgctttg	ggcacgtaga	52860
aggagagaag	tcataatcga	gaaattcctc	ctggcgcgag	agcctgctct	atcgcgacgg	52920
catcccactg	tcgggaacag	accgatcat	tcacgaggcg	aaagtcgtca	acacatgcgt	52980
tataggcatc	ttcccttgaa	ggatgatctt	gttctgcca	atctggaggt	gcggcagccg	53040
caggcagatg	cgatctcagc	gcaacttgcg	gcaaaacatc	tcactcacct	gaaaaccact	53100
agcgagtctc	gcgatcagac	gaaggccttt	tacttaacga	cacaatatcc	gatgtctgca	53160
tcacaggcgt	cgctatccca	gtcaatacta	aagcgggtgca	ggaactaaag	attactgatg	53220
acttaggcgt	gccacgagge	ctgagacgac	gcgcgtagac	agttttttga	aatcattatc	53280
aaagtgatgg	cctccgctga	agcctatac	ctctgcgccg	gtctgtcgga	gagatgggca	53340
agcattatta	cggctctcgc	gcccgtacat	gcattggacg	attgcagggt	caatggatct	53400
gagatcatcc	agaggattgc	cgcccttacc	ttccgtttcg	agttggagcc	agcccctaaa	53460
tgagacgaca	tagtgcactt	gatgtgacaa	tgccaagaga	gagatttgct	taaccggatt	53520

tttttgc	ca agc	gtaagcc	tattgaagct	tgccggcatg	acgtccgcgc	cgaaagaata	53580
tcctaca	agt aaa	acattct	gcacaccgaa	atgcttgggtg	tagacatcga	ttatgtgacc	53640
aagatc	cctta gc	agtttcgc	ttggggaccg	ctccgaccag	aaataccgaa	gtgaactgac	53700
gccaat	gaca gga	atccctt	ccgtctgcag	ataggtacca	tcgatagatc	tgctgcctcg	53760
cgcg	tttcgg	tgatgacggt	gaaaacctct	gacacatgca	gctcccggag	acggtcacag	53820
cttgt	ctgta ag	cggatgcc	gggagcagac	aagcccgtca	gggcgcgtca	gcgggtgttg	53880
gcgg	gtcgcg	ggcagcc	atgaccaggt	cacgtagcga	tagcggagtg	tatactggct	53940
taact	atgcg	gcatcagagc	agattgtact	gagagtgcac	catatgcggt	gtgaaatacc	54000
gcacag	atgc	gtaaggagaa	aataccgcat	caggectct	tccgcttct	cgctcactga	54060
ctcgc	tgcgc	tcggtcgttc	ggctgcggcg	agcggatca	gctcactcaa	aggcggtaat	54120
acggt	tatcc	acagaatcag	gggataacgc	aggaaagaac	atgtgagcaa	aaggccagca	54180
aaagg	ccagg	aaccgtaaaa	aggccgcgtt	gctggcgttt	ttccatagge	tccgcccc	54240
tgacg	agcat	cacaaaaate	gacgetcaag	tcagaggtgg	cgaaaccga	caggactata	54300
aagata	accag	gcgtttcccc	ctggaagetc	cctcgtgcgc	tctctgttc	cgacctgce	54360
gcttacc	gga	tacctgtccg	cettctccc	ttcgggaage	gtggcgcttt	ctcatagctc	54420
acgct	gtagg	tatctcagtt	cgggttaggt	cgttcgtcc	aagctgggct	gtgtgcacga	54480
accccc	ggt	cagcccagcc	gctgcgcctt	atccggtaac	tatcgtcttg	agtccaacce	54540
ggtaag	acac	gacttatcgc	cactggcagc	agccactggt	aacaggatta	gcagagcgag	54600
gtatg	taggc	ggtgctacag	agttcttgaa	gtggtggcct	aactacggt	acactagaag	54660
gacagt	at	ggtatctgcg	ctctgctgaa	gccagttacc	ttcggaaaa	gagttggtag	54720
ctctt	gatcc	ggcaaaaa	ccaccgctgg	tagcgggtgt	ttttttgtt	gcaagcagca	54780
gattac	gcgc	agaaaaaag	gatctcaaga	agatccttg	atcttttcta	cggggtctga	54840
cgctc	agtg	aacgaaaact	cacgttaagg	gattttggtc	atgagattat	caaaaaggat	54900
cttcac	ctag	atccttttaa	attaaaaatg	aagttttaa	tcaatctaaa	gtatatatga	54960
gtaaac	ttgg	tctgacagtt	accaatgctt	aatcagtgag	gcacctatct	cagcgatctg	55020
tctatt	tcgt	tcatccatag	ttgcctgact	ccccgtcgtg	tagataacta	cgatacggga	55080
gggctt	acca	tctggcccc	gtgctgcaat	gataccgcga	gaccacgct	caccggctcc	55140
agatt	tatca	gcaataaacc	agccagccgg	aagggccgag	cgcagaagtg	gtcctgcaac	55200
tttat	ccgcc	tccatccagt	ctattaattg	ttgccgggaa	gctagagtaa	gtagttcgcc	55260
agtta	atagt	ttgcgcaacg	ttgttgccat	tgctgcaggg	ggggggggg	gggggttcca	55320
ttgtt	cattc	cacggacaaa	aacagagaaa	ggaaacgaca	gaggccaaa	agctcgcttt	55380
cagcac	ctgt	cgtttcttt	cttttcagag	ggtattttaa	ataaaaacat	taagttatga	55440
cgaa	agaaa	cggaaaacgc	ttaaaccgga	aaatctcat	aatagcgaa	aaccgcgag	55500
gtcgc	gccc	cgtaacct					55518
<210>		48					
<211>		61					
<212>		DNA					
<213>		人工序列					
<220>							

<223> 参考序列	
<400> 48	
cacgtatata tacgcgtacg cgtacgtgtg aggtatata atcctccgcc ggggcacgta	60
c	61
<210> 49	
<211> 62	
<212> DNA	
<213> 人工序列	
<220>	
<223> 图1中的序列	
<400> 49	
cacgtatata tacgcgtacg cgtacgttgt gaggtatata tatectcegc eggggcacgt	60
ac	62
<210> 50	
<211> 60	
<212> DNA	
<213> 人工序列	
<220>	
<223> 图1中的序列	
<400> 50	
cacgtatata tacgcgtacg cgtacggtga ggtatatata tectccgccg gggcacgtac	60
<210> 51	
<211> 60	
<212> DNA	
<213> 人工序列	
<220>	
<223> 图1中的序列	
<400> 51	
cacgtatata tacgcgtacg cgtactgtga ggtatatata tectccgccg gggcacgtac	60
<210> 52	
<211> 59	
<212> DNA	
<213> 人工序列	
<220>	
<223> 图1中的序列	
<400> 52	
cacgtatata tacgcgtacg cgtacgtgag gtatatatat cctccgccgg ggcacgtac	59
<210> 53	
<211> 32	

<212> DNA	
<213> 人工序列	
<220>	
<223> 图1中的序列	
<400> 53	
cacgtatata tatcctccgc cggggcacgt ac	32
<210> 54	
<211> 19	
<212> DNA	
<213> 人工序列	
<220>	
<223> 图1中的序列	
<400> 54	
cacgtatata tacgcgtac	19
<210> 55	
<211> 57	
<212> DNA	
<213> 人工序列	
<220>	
<223> 图1中的序列	
<400> 55	
cacgtatata tacgcgtacg cgtgtgaggt atatataatcc tccgccgggg cacgtac	57
<210> 56	
<211> 33	
<212> DNA	
<213> 人工序列	
<220>	
<223> 图1中的序列	
<400> 56	
cacgtatata tacgcgtacg ccggggcacg tac	33
<210> 57	
<211> 30	
<212> DNA	
<213> 人工序列	
<220>	
<223> 图1中的序列	
<400> 57	
cacgtatata tcctccgccg gggcacgtac	30
<210> 58	

- <211> 59  
 <212> DNA  
 <213> 人工序列  
 <220>  
 <223> 图1中的序列  
 <400> 58  
 cacgtatata tacgcgtacg cgtatgtgag gtatatatat cctccgccgg ggcacgtac 59  
 <210> 59  
 <211> 54  
 <212> DNA  
 <213> 人工序列  
 <220>  
 <223> 图2中的序列  
 <400> 59  
 gcgctgctcg attccgtccc catggtcgcc atcacgggac aggtgccgcg acgc 54  
 <210> 60  
 <211> 18  
 <212> PRT  
 <213> 人工序列  
 <220>  
 <223> 图2中的序列  
 <400> 60  
 Ala Leu Leu Asp Ser Val Pro Met Val Ala Ile Thr Gly Gln Val Pro  
 1                    5                    10                    15  
 Arg Arg  
 <210> 61  
 <211> 54  
 <212> DNA  
 <213> 人工序列  
 <220>  
 <223> 图2中的序列  
 <400> 61  
 gcgctgctcg actccgtccc cattgctgcc atcacgggac aggtgctcgcg acgc 54  
 <210> 62  
 <211> 18  
 <212> PRT  
 <213> 人工序列  
 <220>  
 <223> 图2中的序列

<400> 62

Ala Leu Leu Asp Ser Val Pro Ile Val Ala Ile Thr Gly Gln Val Ser

1                    5                    10                    15

Arg Arg

<210> 63

<211> 54

<212> DNA

<213> 人工序列

<220>

<223> 图2中的序列

<400> 63

gcgttgctgg actccgtgcc gatggtegcc atcaegggac aggtgtcccg acgc

54

<210> 64

<211> 18

<212> PRT

<213> 人工序列

<220>

<223> 图2中的序列

<400> 64

Ala Leu Leu Asp Ser Val Pro Met Val Ala Ile Thr Gly Gln Val Ser

1                    5                    10                    15

Arg Arg

<210> 65

<211> 31

<212> DNA

<213> 人工序列

<220>

<223> 图4中的序列

<400> 65

atggctcccc cggccacccc gctccggccg t

31

<210> 66

<211> 10

<212> PRT

<213> 人工序列

<220>

<223> 图4中的序列

<400> 66

Met Ala Pro Pro Ala Thr Pro Leu Arg Pro

1                    5                    10



- <210> 67  
 <211> 30  
 <212> DNA  
 <213> 人工序列  
 <220>  
 <223> 图4中的序列  
 <400> 67  
 atggctcccc cggccacccc ctccggecgt 30
- <210> 68  
 <211> 10  
 <212> PRT  
 <213> 人工序列  
 <220>  
 <223> 图4中的序列  
 <400> 68  
 Met Ala Pro Pro Ala Thr Pro Ser Gly Arg  
 1 5 10
- <210> 69  
 <211> 31  
 <212> DNA  
 <213> 人工序列  
 <220>  
 <223> 图4中的序列  
 <220>  
 <221> 尚未归类的特征  
 <222> (21) .. (21)  
 <223> n是a、c、g、或t  
 <400> 69  
 atggctcccc cggccacccc nctccggecg t 31
- <210> 70  
 <211> 40  
 <212> DNA  
 <213> 人工序列  
 <220>  
 <223> 图5中的序列  
 <400> 70  
 tcgactcget caccatgtcc ggeccatgac caccgcccgc 40
- <210> 71  
 <211> 41



<220>

<223> 图6中的序列

<400> 75

Ile Pro Pro Ala Thr Arg Arg

1                    5

<210> 76

<211> 24

<212> DNA

<213> 人工序列

<220>

<223> 图6中的序列

<400> 76

attccccgg ccacctcgtc ggcc

24

<210> 77

<211> 8

<212> PRT

<213> 人工序列

<220>

<223> 图6中的序列

<400> 77

Ile Pro Pro Ala Thr Ser Ser Ala

1                    5

<210> 78

<211> 441

<212> DNA

<213> 人工序列

<220>

<223> IN2启动子

<400> 78

atccctggcc accaaacate cctaatcate cccaaatfff ataggaacta ctaatttctc 60

taacttaaaa aaaatctaaa atagtatact ttageagect cteaactga tttgttcccc 120

aaatttgaat cctggcttcg ctctgtcacc tgtttgactc tacatgggtgc gcagggggag 180

agcctaactt ttcacgaact tgtttgtaac tgttagccag accggcgtat ttgtcaatgt 240

ataaacacgt aataaaatff acgtaccata tagtaagact ttgtatataa gacgtcacct 300

cttacgtgca tggttatatg cgacatgtgc agtgacgtta tcagatatag ctcaccctat 360

atatatagct ctgtccgggt tcagtagcaa tcaccattea tcagcacecc ggcaggtcga 420

ccccgagctc cctgcacctg c 441

<210> 79

<211> 21

<212> DNA	
<213> 玉蜀黍	
<400> 79	
gctcccccg g ccaccccgct c	21
<210> 80	
<211> 20	
<212> DNA	
<213> 玉蜀黍	
<400> 80	
gctcccccg g ccacccccte	20
<210> 81	
<211> 20	
<212> DNA	
<213> 玉蜀黍	
<400> 81	
cgccgagggc gactaccggc	20
<210> 82	
<211> 23	
<212> DNA	
<213> 玉蜀黍	
<400> 82	
cgccgagggc gactaccggc agg	23



		ALS-CR4 TS	PAM		SEQ ID NO:
ALS1 (基因组)	GCGCTGCTCGATTCCGTC	CCCA	GGTGGCCATCACGGGACAGGTG	CCCGSACGG	59
Oligo1	A L L D S V P M V A I T G Q V			P R R	60
	GCGTTGCTGGACTCCGTC	CCCA	TGTGGCCATCACGGGACAGGTG	CCCGSACGG	61
Oligo2	A L L D S V P I V A I T G Q V			S R R	62
	GCGTTGCTGGACTCCGTC	CCGATGGTGGCCATCACGGGACAGGTG	CCCGSACGG		63
	A L L D S V P M V A I T G Q V			S R R	64

图2



野生型植物

具有编辑的ALS2等位基因的植物

图3

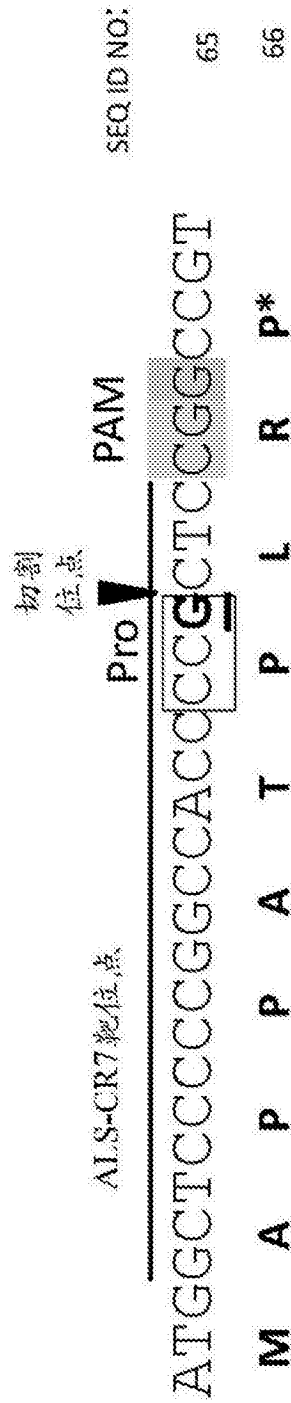
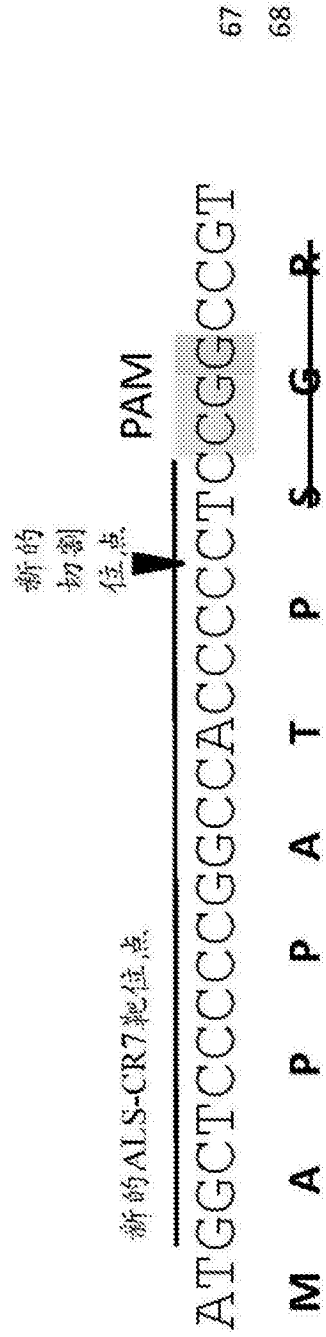


图4A





67  
68

图4B

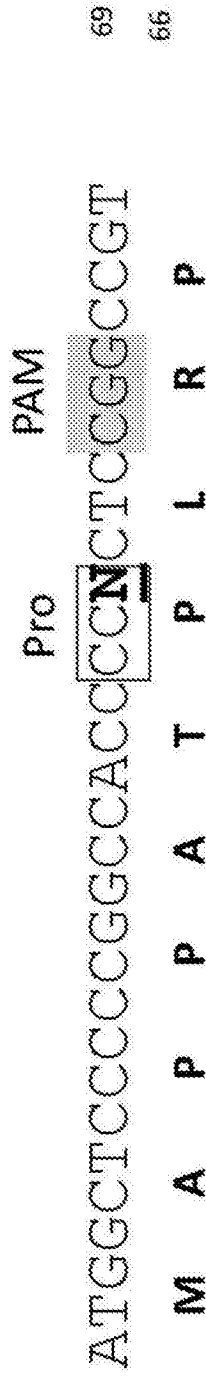


图4C

(SEQ ID NO: 70)

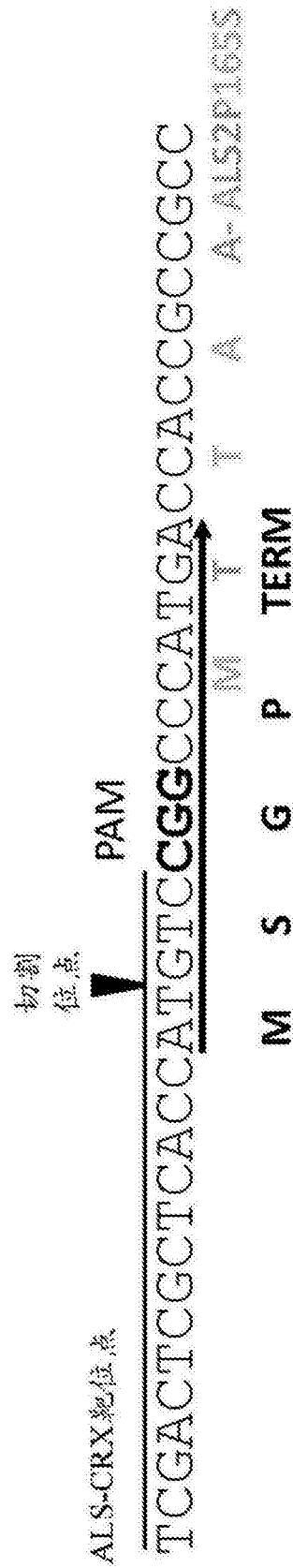


图5A



图5B

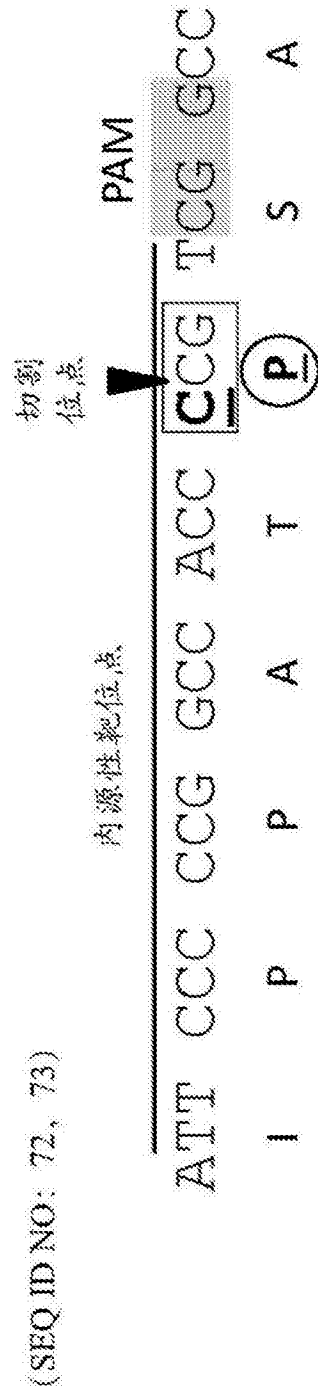


图6A

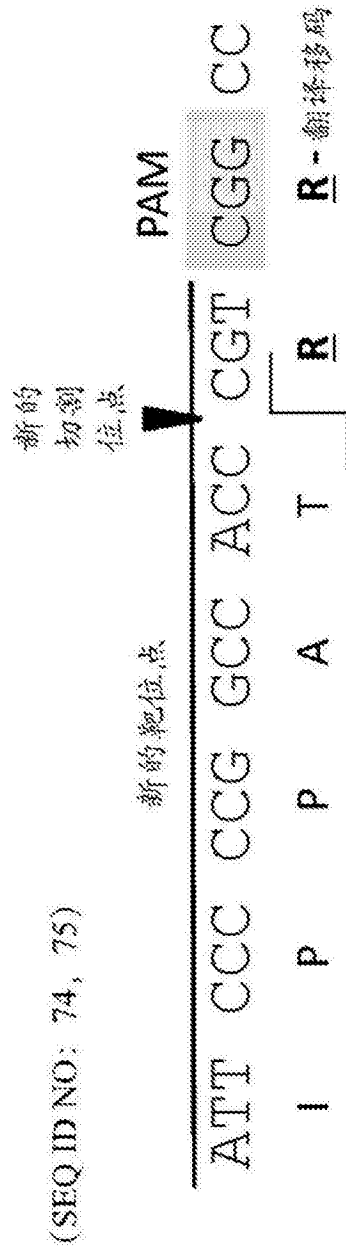


图6B

(SEQ ID NO: 75, 76)

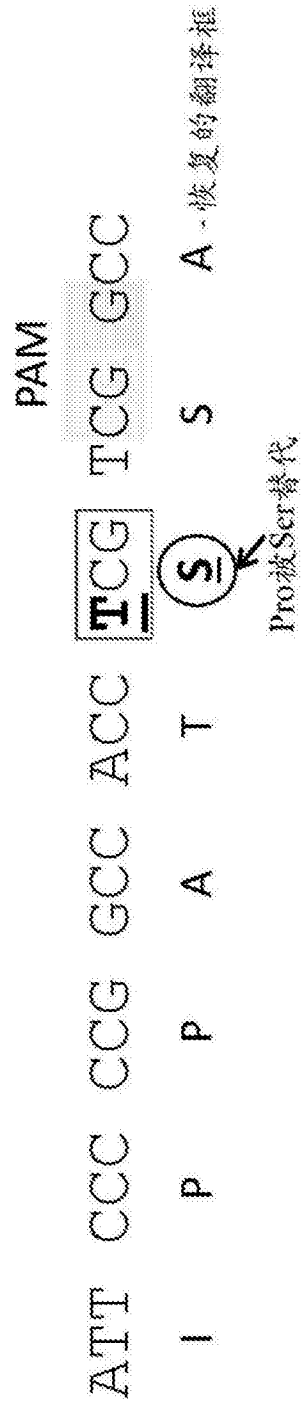


图6C

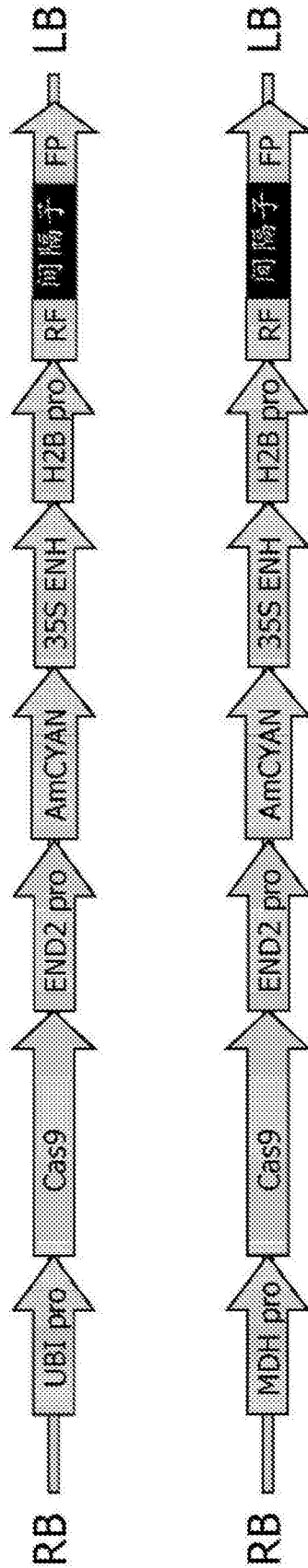


图7