



(12)发明专利

(10)授权公告号 CN 106978438 B

(45)授权公告日 2020.08.28

(21)申请号 201710106331.7

A01H 6/46(2018.01)

(22)申请日 2017.02.27

审查员 李美宣

(65)同一申请的已公布的文献号

申请公布号 CN 106978438 A

(43)申请公布日 2017.07.25

(73)专利权人 北京大北农生物技术有限公司

地址 100193 北京市海淀区圆明园西路2号

院中国农业科学院原子能利用研究所

49号楼

(72)发明人 杨进孝 徐雯

(51)Int.Cl.

C12N 15/82(2006.01)

C12N 9/22(2006.01)

A01H 5/10(2018.01)

C07K 19/00(2006.01)

权利要求书1页 说明书14页

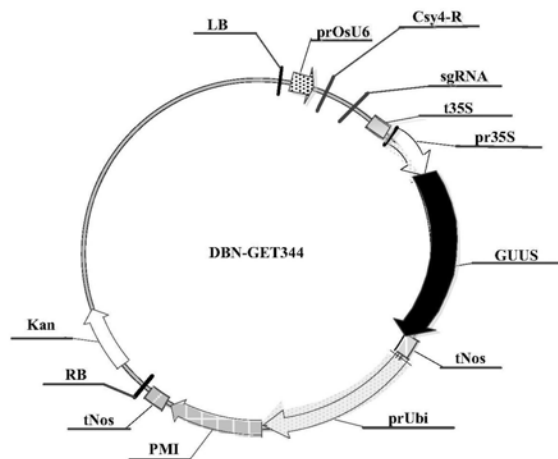
序列表16页 附图4页

(54)发明名称

提高同源重组效率的方法

(57)摘要

本发明公开了一种提高同源重组效率的方法,包括向宿主细胞中引入FokI-dCas9融合蛋白。本发明首次将FokI-dCas9融合蛋白应用于提高同源重组效率,为基因组编辑技术实现高效同源重组编辑提供了一种新的选择,提高同源重组效率的同时,也减少了转化受体的使用量。



1. 一种提高同源重组效率的方法,其特征在于,包括向宿主细胞中引入FokI-dCas9融合蛋白,所述FokI-dCas9融合蛋白具有SEQ ID NO:4和SEQ ID NO:5所示的氨基酸序列。

2. 根据权利要求1所述提高同源重组效率的方法,其特征在于,所述FokI-dCas9融合蛋白在宿主细胞中瞬时表达或稳定表达。

3. 根据权利要求1或2所述提高同源重组效率的方法,其特征在于,所述宿主细胞为植物细胞。

4. 根据权利要求3所述提高同源重组效率的方法,其特征在于,所述植物为玉米、水稻、大豆、拟南芥、棉花、油菜、高粱、小麦、大麦、粟、甘蔗或燕麦。

5. 根据权利要求4所述提高同源重组效率的方法,其特征在于,所述FokI-dCas9融合蛋白的核苷酸序列具有SEQ ID NO:1第643-5523位所示的核苷酸序列。

6. 一种基因组编辑系统,其特征在于,包含FokI-dCas9融合蛋白,所述FokI-dCas9融合蛋白具有SEQ ID NO:4和SEQ ID NO:5所示的氨基酸序列。

7. 根据权利要求6所述基因组编辑系统,其特征在于,所述FokI-dCas9融合蛋白的核苷酸序列具有SEQ ID NO:1第643-5523位所示的核苷酸序列。

8. 根据权利要求6或7所述基因组编辑系统,其特征在于,所述基因组编辑系统还包括编码序列操纵系统的多核苷酸序列。

9. 根据权利要求8所述基因组编辑系统,其特征在于,所述序列操纵系统为CRISPR/Cas系统。

10. 一种实现基因组编辑的方法,其特征在于,包括在生物体中表达权利要求6-9任一项所述基因组编辑系统。

11. 一种产生基因组编辑的植物的方法,其特征在于,包括向植物基因组中引入编码权利要求6-9任一项所述基因组编辑系统的核苷酸序列。

12. 一种产生基因组编辑植物种子的方法,其特征在于,包括将权利要求11所述方法产生的基因组编辑的植物自交,从而获得具有基因组编辑植物种子。

13. 一种培育基因组编辑植物的方法,其特征在于,包括:

种植至少一粒权利要求12所述方法产生的所述基因组编辑植物种子;

使所述种子长成植株。

14. 一种权利要求6-9任一项所述基因组编辑系统在提高同源重组效率和/或提高基因组编辑效率中的用途。

15. 一种FokI-dCas9融合蛋白在提高同源重组效率中的用途,其特征在于所述FokI-dCas9融合蛋白具有SEQ ID NO:4和SEQ ID NO:5所示的氨基酸序列。

16. 根据权利要求15所述用途,其特征在于,所述FokI-dCas9融合蛋白的核苷酸序列具有SEQ ID NO:1第643-5523位所示的核苷酸序列。

提高同源重组效率的方法

技术领域

[0001] 本发明涉及一种提高同源重组率的方法,特别涉及一种将FokI-dCas9蛋白应用于基因编辑系统中以实现提高同源重组效率的方法。

背景技术

[0002] 随着生命科学研究进入基因组时代,越来越多物种的基因组完成测序,解读与改造基因组的功能就显得非常紧迫,近几年,生物学家们巧妙地利用蛋白质结构与功能领域的研究成果,将特异识别与结合DNA的蛋白质结构与和核酸内切酶结构域融合,创造出能够按照人们意愿特异切割DNA的序列特异核酸酶(Sequence-specific nucleases,SSNs),并藉此实现了对基因组特定定位点的靶向修饰,及基因组编辑(Genome editing)。SSNs主要包括3种类型:锌指核酸酶(Zinc finger nuclease,ZFN)、类转录激活因子效应物核酸酶(Transcription activator-like effector nuclease,TALEN)和成簇的规律间隔的短回文重复序列及其相关系统(Clustered regularly interspaced short palindromic repeats/CRISPR-associated 9,CRISPR/Cas9 system)。上述几类SSNs的共同特征是能够作为核酸内切酶切割特定的DNA序列,创造DNA双链断裂(Double-strand breaks,DSBs)。

[0003] 在真核生物中,DSBs的修复机制是高度保守的,主要包括两种途径:非同源末端连接(Non-homologous end joining,NHEJ)和同源重组(Homologous recombination,HR)。通过NHEJ方式,断裂的染色体会重新连接,但往往不是精确的,断裂位置会产生少量核苷酸的插入或删除,从而产生基因敲除突变体;通过HR方式,在引入同源序列的情况下,以同源序列为模板进行合成修复,从而产生精确的定点替换或者插入突变体。在这两种途径中,NHEJ方式占绝对主导,可以发生在几乎所有类型的细胞以及不同的细胞周期中(G1、S和G2期);然而,HR发生频率很低,主要发生在S和G2期。HR根据其发生方式不同可以分为两类:单链退火(Single-strand annealing,SSA)和合成依赖式退火(Synthesis-dependent strand annealing,SDSA)。DSBs产生后,在这两种途径下DNA断裂末端都会发生5'至3'方向的DNA切除,形成3'单链末端。SSA途径类似NHEJ途径,DSBs两端各有一段同源序列,同源序列区域直接退火形成互补双链,再经过末端加工和连接修复DSBs,在基因组串联重复区域,SSA是主要的DSBs修复方式。SDSA途径是依赖DNA合成的修复过程,基因组编辑过程中同源重组通常是指这种方式。DSB经过5'至3'方向的DNA切除产生的一个3'单链末端入侵同源供体DNA模板,形成D-loop环状结构,再利用同源供体DNA的互补链作为模板进行DNA合成修复,当延伸至可以与DSB另一个单链末端互补配对的位置时,脱离D-loop结构,两个单链DSB末端退火形成双链,完成修复过程。SDSA途径最终结果就是完成从同源DNA至DSB遗传信息的转化过程。SDSA途径发生频率非常低,相同条件下只有SSA方式的10%-20%。由此可见,提高HR效率是基因组编辑研究最重要也是最迫切的任务之一。

[0004] 在CRISPR/Cas9基因编辑系统中,设计不同的sgRNA以指导Cas9内切酶完成对DNA的定点切割,通过同源重组修复机制实现目标基因中不同类型的修饰,包括基因的删除、添加和替换等。因此明确DNA修复机制特别是HR修复过程的研究将有助于人们采用适当方法

提高基因组编辑中定点插入或替换的效率。

发明内容

[0005] 本发明的目的是提供一种提高同源重组效率的方法,首次提供了FokI-dCas9融合蛋白可以提高同源重组效率。

[0006] 为实现上述目的,本发明提供了一种提高同源重组效率的方法,包括向宿主细胞中引入FokI-dCas9融合蛋白。

[0007] 进一步地,所述FokI-dCas9融合蛋白在宿主细胞中瞬时表达或稳定表达。

[0008] 更进一步地,所述宿主细胞为植物细胞。

[0009] 优选地,所述植物为玉米、水稻、大豆、拟南芥、棉花、油菜、高粱、小麦、大麦、粟、甘蔗或燕麦。

[0010] 在上述技术方案的基础上,所述FokI-dCas9融合蛋白的氨基酸序列具有SEQ ID NO:4和SEQ ID NO:5所示的氨基酸序列。

[0011] 优选地,所述FokI-dCas9融合蛋白的核苷酸序列具有SEQ ID NO:1第643-5523位所示的核苷酸序列。

[0012] 为实现上述目的,本发明还提供了一种基因组编辑系统,包含FokI-dCas9融合蛋白。

[0013] 进一步地,所述FokI-dCas9融合蛋白的氨基酸序列具有SEQ ID NO:4和SEQ ID NO:5所示的氨基酸序列。

[0014] 更进一步地,所述FokI-dCas9融合蛋白的核苷酸序列具有SEQ ID NO:1第643-5523位所示的核苷酸序列。

[0015] 可选地,所述基因组编辑系统还包括编码序列操纵系统的多核苷酸序列。

[0016] 优选地,所述序列操纵系统为CRISPR/Cas系统。

[0017] 为实现上述目的,本发明还提供了一种实现基因组编辑的方法,包括在生物体中表达所述基因组编辑系统。

[0018] 为实现上述目的,本发明还提供了一种产生基因组编辑植物的方法,包括向植物基因组中引入编码所述基因组编辑系统的核苷酸序列。

[0019] 为实现上述目的,本发明还提供了一种产生基因组编辑植物种子的方法,包括将所述方法产生的基因组编辑的植物自交,从而获得具有基因组编辑植物种子。

[0020] 为实现上述目的,本发明还提供了一种培育基因组编辑植物的方法,包括:

[0021] 种植至少一粒所述方法产生的所述基因组编辑植物种子;

[0022] 使所述种子长成植株。

[0023] 为实现上述目的,本发明还提供了一种所述基因组编辑系统在提高同源重组效率和/或提高基因组编辑效率中的用途。

[0024] 为实现上述目的,本发明还提供了一种FokI-dCas9融合蛋白在提高同源重组效率中的用途。

[0025] 进一步地,所述FokI-dCas9融合蛋白的氨基酸序列具有SEQ ID NO:4和SEQ ID NO:5所示的氨基酸序列。

[0026] 更进一步地,所述FokI-dCas9融合蛋白的核苷酸序列具有SEQ ID NO:1第643-

5523位所示的核苷酸序列。

[0027] 本发明所述的FokI是包括DNA识别结构域和催化(内切核酸酶)结构域的II型限制性内切核酸酶。本文所述的融合蛋白可包括所有FokI或仅仅催化内切核酸酶结构域,例如,基因库登录号AAA24927.1的氨基酸388-583或408-583,例如,如Li等,NucleicAcidsRes.39(1):359-372(2011);Cathomen和Joung,Mol. Ther. 16:1200-1207(2008)中描述的,或如Miller等NatBiotechnol25:778-785(2007);Szczepek等,Nat Biotechnol25:786-793(2007);或Bitinaite等,Proc.Natl.Acad.Sci.USA.95:10570-10575(1998)中描述的FokI的突变形式。

[0028] 本发明中所述的Cas9, II型CRISPR/Cas系统中一种重要的蛋白质成分,可以从生物体如链球菌属物种(*Streptococcus* sp.), 优选酿脓链球菌(*Streptococcus pyogenes*.)中分离得到。当Cas9与称为CRISPR RNA (crRNA) 和反式激活crRNA (TracrRNA) 的两个RNA复合时,形成活性核酸内切酶,从而切断入侵噬菌体或质粒中的外源遗传元件,以保护宿主细胞。crRNA从宿主基因组中的CRISPR元件转录,其中该CRISPR元件之前自外源入侵物捕获。研究表明通过融合crRNA和tracrRNA的必要部分产生的单链嵌合RNA可以取代Cas9/RNA复合体中的两个RNA以形成功能性核酸内切酶。Cas9蛋白质的变体可以是Cas9的突变体形式,其中催化性天冬氨酸残基改变为任何其它氨基酸。优选地,所述其它氨基酸可以是丙氨酸。

[0029] 本发明中所述FokI-dCas9融合蛋白,其中FokI序列任选地通过间插接头,例如2-30个氨基酸,例如4-12氨基酸的接头,例如Gly4Ser与dCas9(优选地与dCas9的氨基末端,而且还任选地与羧基末端)融合。在一些实施方案中,所述融合蛋白包括dCas9与FokI结构域之间的接头。可用于这些融合蛋白的接头(或串联结构中的融合蛋白之间)可包括不干扰融合蛋白的功能的任何序列。在优选实施方案中,所述接头是短的,例如2-20个氨基酸,并且通常是柔性的(即包含具有高度自由的氨基酸诸如甘氨酸、丙氨酸和丝氨酸)。在一些实施方案中,所述接头包含一个或多个由GGGS或GGGGS组成的单元,例如,2、3、4或更多个GGGS或GGGGS单元的重复,还可使用其它接头序列。

[0030] 本发明所用2A肽(T2A)是一种可“自我剪切的短小肽链”,最初在手足口病毒(foot-and-mouth disease virus,FMDV)中发现,平均长度为18-22个氨基酸,2A肽可在蛋白翻译时通过核糖体跳跃从自身最后2个氨基酸C末端断裂(de Felipe et al.,2003)。甘氨酸和脯氨酸之间的肽链结合群在2A为电视受损的,能引发核糖体跳跃而从第2个密码子开始翻译,从而使1个转录单元里2个蛋白独立表达。这中2A介导的剪切广泛存在与所有的真核动物细胞当中。利用2A较高的剪切效率及促使上下游基因平衡表达的能力,可以改进异源多聚蛋白(如细胞表面受体、细胞因子、免疫球蛋白等)的表达效率。

[0031] 本发明中所述的向导RNA或引导RNA(guide RNA,gRNA),也称为小向导RNA(small guide RNA,sgRNA),作用于动质体(kinetoplastid)体内一种称为RNA编辑(RNA editing)的后转录修饰过程中,也是一种小型非编码RNA。可与pre-mRNA配对,并在其中插入一些尿嘧啶(U),产生具有作用的mRNA。向导RNA编辑的RNA分子,长度大约是60-80个核苷酸,是由单独的基因转录的,在gRNA的5'末端有一段锚定区,以特殊的G-U配对方式与非编辑的pre-mRNA序列互补,锚定序列促进gRNA与pre-mRNA中的编辑区互补,特意结合;在gRNA分子的中间部位有一个编辑区负责在被编辑的pre-mRNA分子中插入U的位置,其与被编辑mRNA精确互补;在gRNA分子的3'末端,有一段转录后加入的由大约15个非编码的PolyU序列,功能为

把gRNA链接到pre-mRNA的编辑区的5'上游富含嘌呤碱基的核苷酸序列上。在编辑时,形成一个编辑体(editosome),以gRNA内部的序列作为模板进行转录物的校正,同时产生编辑的mRNA。

[0032] 有三种类型CRISPR/Cas系统,其中涉及Cas9蛋白质和crRNA、tracrRNA的II型CRISPR/Cas系统是代表性的。Cas9蛋白质通过人工修饰的向导RNA的导向作用可以打靶DNA序列的5'-N20-NGG-3'(N代表任何脱氧核苷酸碱基),N20是与gRNA的5'序列相同的20个碱基,NGG是PAM区(原型间隔子邻近基序,Protospacer-adjacent motif)。Cas9剪切的位点就是PAM附近的区域。相对于锌指和转录激活因子样效应物DNA结合蛋白提供了优势-因为在核苷酸结合CRISPR-Cas蛋白中的位点特异性由RNA分子调控而不是DNA结合蛋白调控。

[0033] 本发明中所述重组,当用于例如细胞、核酸、蛋白质或载体时,表示该细胞、核酸、蛋白质或载体已通过引入异源核酸或蛋白质、或改变天然核酸或蛋白质而被修饰,或该细胞源自此修饰的细胞。

[0034] 本发明中向导RNA可以以编码该向导RNA的RNA或DNA的形式转移到细胞或生物体中。向导RNA可以是分离的RNA、并入病毒载体的RNA的形式、或者在载体中编码。优选地,载体可以是病毒载体、质粒载体、或农杆菌载体。

[0035] 编码向导RNA的DNA可以是包含编码向导RNA序列的载体。例如,可以通过用分离的向导RNA或包含编码向导RNA的序列和启动子的质粒DNA转染细胞或生物体,将向导RNA转染到细胞或生物体。

[0036] 本发明中所述切割或剪切是指核苷酸分子共价骨架的断裂。向导RNA可以制备为特异于任何待切割的靶,通过向导RNA的靶特异性部分切割任何靶DNA。

[0037] 本发明中所述非同源末端连接(Non-homologous end joining, NHEJ)指完全不需要任何模版的帮助,修复蛋白可以直接将DNA断裂的末端彼此拉近,再借由DNA连接酶的帮助将断裂的末端重新接合的修复机制。

[0038] 本发明中所述同源重组(Homologous Recombination, HR)是指发生在非姐妹染色单体之间或同一染色体上含有同源序列的DNA分子之间或分子之内的重新组合。同源重组需要一系列的蛋白质催化,如原核生物细胞内的RecA、RecBCD、RecF、RecO、RecR等;以及真核生物细胞内的Rad51、Mre11-Rad50等等。同源重组反应通常根据交叉分子或Holliday结构的形成和拆分为三个阶段,即前联会体阶段、联会体形成和Holliday结构的拆分。同源重组反应依赖于DNA分子之间的同源性,100%同源性的DNA分子之间的重组常见于非姐妹染色体之间的同源重组,称为Homologous Recombination,而小于100%同源性的DNA分子之间或分子之内的重组,则被称为Hemologus Recombination。后者可被负责碱基错配对的蛋白如原核细胞内的MutS或真核生物细胞内的MSH2-3等蛋白质“编辑”。同源重组可以双向交换DNA分子,也可以单向转移DNA分子,后者又被称为基因转换(Gene Conversion)。

[0039] 本发明中,单链退火(SSA)模型是由Lin等于1984年提出的。在SSA模型中,重组在DNA双链断裂处开始,在单链特异性核酸外切酶的作用下,断裂点两侧逐渐形成DNA单链区域,这一过程持续到两个断点处出现互补DNA单链。互补DNA单链退火,非互补末端被切除,单链缺口修复连接,完成DNA重组。SSA模型没有其它模型所需的双链DNA的识别配对过程,不形成Holliday结构作为重组的中间过渡形式。因此,重组产生DNA双链交换,并丢失非退火区域单链DNA顺序。

[0040] 本发明中,所述串联是指将二个或二个以上向导RNA (sgRNA) 排成一串,每个sgRNA的首端和前一个sgRNA的尾端由Csy4切割识别序列连接。

[0041] 本发明中所述的植物、植物组织或植物细胞的基因组,是指植物、植物组织或植物细胞内的任何遗传物质,且包括细胞核和质体和线粒体基因组。

[0042] 本发明中所述的多核苷酸和/或核苷酸形成完整“基因”,在所需宿主细胞中编码蛋白质或多肽。本领域技术人员很容易认识到,可以将本发明的多核苷酸和/或核苷酸置于目的宿主中的调控序列控制下。

[0043] 如本申请,包括权利要求中使用的,除非上下文中清楚地另有说明,否则单数和单数形式的术语,例如“一”、“一个”和“该”,包括复数指代物。因此,例如“植物”、“该植物”或“一个植物”也指示多个植物。并且根据上下文,使用术语“植物”还可指示该植物在遗传上相似或相同的后代。类似地,术语“核酸”可以指核酸分子的许多拷贝。类似地,术语“探针”可以指代相同或相似的探针分子。

[0044] 数字范围包括限定该范围的数字,并明确地包括所限定的范围内的每个整数和非整数分数。除非另外指出,否则本文所用的全部技术和科学术语具有与本领域普通技术人员普遍理解的相同的含义。

[0045] 本发明中,术语“核酸”、“核苷酸”、“核苷酸序列”、“寡核苷酸”和“多核苷酸”可互换使用,它们是指具有任何长度的核苷酸的聚合形式,根据上下文含义,可指代DNA或RNA、或其类似物。其中DNA包括但不限于cDNA、基因组DNA、合成DNA (例如人工合成) 和含核酸类似物的DNA (或RNA)。多核苷酸可具有任何三维结构,并且可以执行已知或未知的任何功能。核酸可为双链或单链 (既有义链或反义单链)。多核苷酸的非限制性示例包括基因、基因片段、外显子、内含子、信使RNA (mRNA)、转移RNA、核糖体RNA、核酶、cDNA、重组多核苷酸、分支多核苷酸、质粒、载体、任何序列的分离DNA、任何序列的分离RNA、核酸探针和引物、以及核酸类似物。

[0046] 本发明中“野生型”表示生物、菌株、基因的典型形式或者当它在自然界存在时区别于突变体或变体形式的特征。

[0047] 本发明中“突变体”或“变体”是指发生突变的个体,其具有与野生型不同的序列,可能会导致其中序列的至少部分功能已丢失的序列,例如,在启动子或增强子区域中序列的变化将至少部分地影响生物体中编码序列的表达。术语“突变”是指可由诸如缺失、添加、取代或重排引起的核酸序列中序列的任何变化。突变还可影响该序列参与的一个或多个步骤。例如,DNA序列中的变化可导致有活性的、有部分活性的或无活性的改变的mRNA和/或蛋白的合成。

[0048] 本发明中“非天然存在的”表明人工的参与。当指核酸分子或多肽时,表示该核酸分子或多肽至少基本上从它们在自然界中或如发现于自然界中的与其结合的至少另一种组分游离出来。

[0049] 本发明中“表达”指感兴趣的序列转录产生对应mRNA并且该mRNA翻译产生对应产物,即肽、多肽或蛋白。调节元件控制或调整感兴趣的序列表达,所述调节元件包括5' 调节元件如启动子。

[0050] 本发明中“多肽”、“肽”和“蛋白质”可互换地使用,是指具有任何长度的氨基酸聚合物。该聚合物可以是直链或支链的,它可以包含修饰的氨基酸,并且它可以被非氨基酸中

断。这些术语还涵盖已经被修饰的氨基酸聚合物；这些修饰例如二硫键形成、糖基化、脂化、乙酰化、磷酸化、或任何其他操作，如与标记组分的结合。术语“氨基酸”包括天然的和/或非天然的或者合成的氨基酸，包括甘氨酸以及D和L旋光异构体、以及氨基酸类似物和肽模拟物。

[0051] 本发明中术语“载体”是指能够在宿主细胞内复制的DNA分子。质粒和粘粒是示例性载体。此外，术语“载体”和“媒介”可互换使用以指将DNA片段从一种细胞转移至另一种细胞的核酸分子，因此细胞不必要属于相同的生物（例如将DNA片段从农杆菌细胞转移至植物细胞）。

[0052] 本发明中术语“表达载体”是指含有所需要的编码序列和在特定宿主生物中表达有效连接的编码序列所需要的适宜核酸序列的重组DNA分子。

[0053] 本发明中术语“重组表达载体”指来自任何来源、能够整合入基因组或自主复制的任何因子如质粒、粘粒、病毒、BAC（细菌人工染色体）、自主复制型序列、噬菌体、或者线性或环状单链或双链DNA或RNA核苷酸序列，包括其中一种或多种DNA序列使用众所周知的重组DNA技术以功能性可操作方式连接的DNA分子。

[0054] 在本发明中，“定位结构域”可任选地添加为蛋白部分的部分，定位结构域可将蛋白部分或编程的蛋白部分或组装的复合物定位至活细胞中特定细胞或亚细胞定位。定位结构域可通过将蛋白部分的氨基酸序列融合到掺入包括以下结构域的氨基酸来构建：核定位信号（NLS）；线粒体前导序列（MLS）；叶绿体前导序列；和/或被设计以将蛋白运输或引导或定位至含有核酸的细胞器、细胞区室或细胞的任何细分部分的任何序列。在一些实施方案中，生物体是真核生物，且定位结构域包括允许蛋白进入细胞核和基因组DNA内的核定位结构域（NLS）。所述NLS的序列可包括带正电荷序列的任何功能NLS。在另一些实施方案中，定位结构域可包括使蛋白部分或编程的核蛋白进入细胞器的前导序列，使细胞器DNA被修饰成为可能。

[0055] 本发明中，真核生物有3类RNA聚合酶，负责转录3类不同的启动子。由RNA聚合酶I负责转录的rRNA基因，启动子（I型）比较单一，由转录起始位点附近的两部分序列构成：第一部分是核心启动子（core promoter），由-45—+20位核苷酸组成，单独存在时就足以起始转录；另一部分由-170—-107位序列组成，称为上游调控元件，能有效地增强转录效率。

[0056] 由RNA聚合酶III负责转录的是5S rRNA、tRNA和某些核内小分子RNA（snRNA），其启动子（III型）组成较复杂，又可被分为两个亚类：一类属于结构基因内启动子，一类属于结构基因外启动子。内启动子的有效启动依赖于基因内部包含的两个不连续的DNA片段，这两个DNA片段包括一些不同的连续DNA序列A、B或C区，两个区之间被隔开。根据两个区的不同组合，可以分为I类和II类两种：I类包括A和C区，目前发现其仅存在于5S rRNA的基因中；II类包括A和B区，存在于tRNA的基因、7SLRNA的基因、腺病毒VAI与VAII的RNA中。A、B或A、C内部DNA序列是转录因子TF IIIA和TF IIIC转录启动结合部位。在多数情况下，5'端还会有其它调节或关键元件，这些元件对RNA的高效转录是必需的。这些序列的存在与否之间影响着转录效率。这些序列呈现复杂的多样性，不过大多数启动子5'端-30到-20处存在着TATA盒样序列，与外启动子相似。外启动子缺乏相应的内部序列，仅在5'端有顺式作用元件，在基因的末端有一组由4个或更多胸腺嘧啶组成的终止信号，如脊椎动物U6核小RNA和7SK RNA启动子，这些启动子都高度相似或完全相同，其位置和碱基序列高度保守，其结构与pol II启动子也

有一定的相似性。它们的5'端顺式作用元件包括几个控制元件,在其上游约-30处存在一个TATA样序列,近-60处有一个snRNA PSE (snRNA近似序列) 和一个或多个被称为OCT的修饰序列5' -ATGCAAAT-3'。TATA样序列是pol III对snRNA基因进行转录特异的。TATA样元件和PSE元件共同决定了转录起始位点的选择和转录效率。TATA样元件和PSE元件之间的距离决定了RNA聚合酶转录的特异性,但似乎TATA样元件更重要,因为在PSE缺失的U6RNA和7SK RNA基因转录中,只有转录效率的下降。而且,PSE元件可能和B盒(boxB)相关,B盒在一定程度上可以替代PSE元件。这些序列对下游基因的转录具有决定性的作用,而且它们与起始位点相距较远,往往超过150bp,而对pol III内启动子来说,一般在80bp以内;与pol III外启动子相比,pol II启动子5'端各顺式作用元件的作用刚好相反,如果TATA样元件缺失,PSE则可以履行TATA样元件功能,决定转录的起始部位。在PSE上游,外启动子还有一个远端控制序列,其结构与pol II增强子OCT骨架相似,但较其复杂。而在-223处还有一个CACC序列与OCT骨架相连。这些远端控制序列的存在可大大提高U6RNA和7SK RNA的表达效率。

[0057] 由RNA聚合酶II负责转录的II型基因包括所有蛋白质编码基因和部分snRNA基因,后者的启动子结构与III型基因启动子中的第三亚类相似,编码蛋白质的II型基因启动子在结构上有共同的保守序列。转录起始位点没有广泛的序列同源性,但第一个碱基为腺嘌呤,而两侧是嘧啶碱基。这个区域被称为起始子(initiator, Inr),序列可表示为Py2CAPy5。Inr元件位于-3—+5。仅由Inr元件组成的启动子是具有可被RNA聚合酶II识别的最简单启动子形式。多数II型启动子有一个被称为TATA盒的共有序列,通常处于-30区,相对于转录起始位点的位置比较固定。TATA盒存在于所有真核生物中,TATA盒是一个保守的七碱基对,也有一些II型启动子不含有TATA盒,这样的启动子称为无TATA盒启动子。

[0058] 本发明中所述“有效连接”或“可操作地连接”表示核酸序列的联结,所述联结使得一条序列可提供对相连序列来说需要的功能。在本发明中所述“有效连接”可以为将启动子与感兴趣的序列相连,使得该感兴趣的序列的转录受到该启动子控制和调控。当感兴趣的序列编码蛋白并且想要获得该蛋白的表达时“有效连接”表示:启动子与所述序列相连,相连的方式使得得到的转录物高效翻译。如果启动子与编码序列的连接是转录物融合并且想要实现编码的蛋白的表达时,制造这样的连接,使得得到的转录物中第一翻译起始密码子是编码序列的起始密码子。备选地,如果启动子与编码序列的连接是翻译融合并且想要实现编码的蛋白的表达时,制造这样的连接,使得5'非翻译序列中含有的第一翻译起始密码子与启动子相连接,并且连接方式使得得到的翻译产物与编码想要的蛋白的翻译开放读码框的关系是符合读码框的。可以“有效连接”的核酸序列包括但不限于:提供基因表达功能的序列(即基因表达元件,例如启动子、5'非翻译区域、内含子、蛋白编码区域、3'非翻译区域、聚腺苷化位点和/或转录终止子)、提供DNA转移和/或整合功能的序列(即T-DNA边界序列、位点特异性重组酶识别位点、整合酶识别位点)、提供选择性功能的序列(即抗生素抗性标记物、生物合成基因)、提供可计分标记物功能的序列、体外或体内协助序列操作的序列(即多接头序列、位点特异性重组序列)和提供复制功能的序列(即细菌的复制起点、自主复制序列、着丝粒序列)。

[0059] 本发明中,调节元件可操作地连接至CRISPR系统的一个或多个元件,从而驱动该CRISPR系统的所述一个或多个元件的表达。一般而言,CRISPR(规律间隔成簇短回文重复),也称为SPIDR(Spacer间隔开的同向重复),构成通常对于特定细菌物种而言特异性的DNA基

因座的家族。该CRISPR座位包含在大肠杆菌中被识别的间隔开的短序列重复 (SSR) 的一个不同类、以及相关基因。类似的间隔开的SSR已经鉴定于地中海富盐菌、化脓链球菌、鱼腥草属和结核分枝杆菌中。这些CRISPR座位典型地不同于其它SSR的重复结构,这些重复已被称为规律间隔的短重复 (SRSR)。一般而言,这些重复是以簇存在的短元件,其被具有基本上恒定长度的独特间插序列规律地间隔开。虽然重复序列在菌株之间是高度保守的,许多间隔开的重复和这些间隔区的序列一般在菌株与菌株之间不同,已经在40种以上的原核生物中鉴定出CRISPR座位。

[0060] 本发明中,“靶序列”或“靶点序列”或“靶位点序列”或“靶多核苷酸”是待被作用的任何期望的预定的核酸序列,包括但不限于编码或非编码序列、基因、外显子或内含子、调节序列、基因间序列、合成序列和细胞内寄生物序列。在一些实施方案中,靶序列存在于靶细胞、组织、器官或生物体内。

[0061] 术语“引物”是一段分离的核酸分子,其通过核酸杂交,退火结合到互补的目标DNA链上,在引物和目标DNA链之间形成杂合体,然后在聚合酶(例如DNA聚合酶)的作用下,沿目标DNA链延伸。本发明的引物对涉及其在目标核酸序列扩增中的应用,例如,通过聚合酶链式反应(PCR)或其他常规的核酸扩增方法。

[0062] 引物的长度一般是11个多核苷酸或更多,优选的是18个多核苷酸或更多,更优选的是24个多核苷酸或更多,最优选的是30个多核苷酸或更多。这种引物在高度严格杂交条件下与目标序列特异性地杂交。尽管不同于目标DNA序列且对目标DNA序列保持杂交能力的引物是可以通过常规方法设计出来的,但是,优选的,本发明中的引物与目标序列的连续核酸具有完全的DNA序列同一性。

[0063] 本发明的引物在严格条件下与目标DNA序列杂交。核酸分子或其片段在一定情况下能够与其他核酸分子进行特异性杂交。如本发明使用的,如果两个核酸分子能形成反平行的双链核酸结构,就可以说这两个核酸分子彼此间能够进行特异性杂交。如果两个核酸分子显示出完全的互补性,则称其中一个核酸分子是另一个核酸分子的“互补物”。如本发明使用的,当一个核酸分子的每一个核苷酸都与另一个核酸分子的对应核苷酸互补时,则称这两个核酸分子显示出“完全互补性”。如果两个核酸分子能够以足够的稳定性相互杂交从而使它们在至少常规的“低度严格”条件下退火且彼此结合,则称这两个核酸分子为“最低程度互补”。类似地,如果两个核酸分子能够以足够的稳定性相互杂交从而使它们在常规的“高度严格”条件下退火且彼此结合,则称这两个核酸分子具有“互补性”。从完全互补性中偏离是可以允许的,只要这种偏离不完全阻止两个分子形成双链结构。为了使一个核酸分子能够作为引物或探针,仅需保证其在序列上具有充分的互补性,以使得在所采用的特定溶剂和盐浓度下能形成稳定的双链结构。

[0064] 术语“特异性结合(目标序列)”是指在严格杂交条件下引物仅与包含目标序列的样品中的目标序列发生杂交。

[0065] 本发明中,“试剂盒”可包括本发明描述的基因组修饰系统与以下的任一种或全部:测定试剂、缓冲液、探针和/或引物、和无菌盐水或另一种药学上可接受的乳剂和悬液基底。此外,试剂盒可包括含有用于实践本发明描述的方法的用法说明(例如,操作方案)的说明性材料。

[0066] 转化方案以及将核苷酸序列导入植物的方案依定向转化的植物或植物细胞类型

而异,即单子叶植物或双子叶植物。将核苷酸序列导入植物细胞并随后插入植物基因组中的适合方法包括但不限于,农杆菌介导的转化、微量发射轰击、直接将DNA摄入原生质体、电穿孔或晶须硅介导的DNA导入。已经转化的细胞可按照常规的方式生长成植物。这些植物被培育,用相同的转化株或不同的转化株授粉,得到的杂交体表达所需的被鉴定的表现型特征。可培育二代或多代以保证稳定地保持和遗传所需表现型特征的表达,然后收获可保证得到所需表现型特征表达的种子。

[0067] 本发明提供了一种提高同源重组效率的方法,具有以下优点:

[0068] 1、本发明首次将FokI-dCas9融合蛋白应用于提高同源重组效率,利用FokI二聚体的内切酶活性,使得同源重组靶位点被切开的概率提高,以提高同源重组效率。

[0069] 2、本发明为基因组编辑技术实现高效同源重组编辑提供了一种新的选择,提高同源重组效率的同时,也减少了转化受体的使用量。

[0070] 下面通过附图和实施例,对本发明的技术方案做进一步的详细描述。

附图说明

[0071] 图1为本发明提高同源重组效率方法中重组克隆剪刀载体DBN01-T构建流程图;

[0072] 图2为本发明提高同源重组效率方法中重组表达载体DBN-GET326构建流程图;

[0073] 图3为本发明提高同源重组效率方法中重组表达载体DBN-GET344载体结构示意图;

[0074] 图4为本发明提高同源重组效率方法中重组表达载体DBN-GET345载体结构示意图;

[0075] 图5为本发明提高同源重组效率方法中水稻抗性愈伤GUS染色标准图。

具体实施方式

[0076] 下面通过具体实施例进一步说明本发明提高同源重组效率方法的技术方案。

[0077] 第一实施例、剪刀载体构建

[0078] 1. 构建基础载体和重组克隆剪刀载体

[0079] 将pCAMBIA2300 (CAMBIA机构可以提供) 载体进行改造,利用常规的酶切方法构建载体是本领域技术人员所熟知的,通过点突变去掉pCAMBIA2300载体上的BsaI位点,同时将卡那霉素表达盒去掉,得到pDBN骨架载体。向所述pDBN骨架载体引入PAT表达盒,得到表达载体DBN-PAT用于下述载体构建。

[0080] 将合成的Csy4-T2A-FokI-dCas9核苷酸序列连入克隆载体pGEM-T (Promega, Madison, USA, CAT:A3600) 上,操作步骤按Promega公司产品pGEM-T载体说明书进行,得到重组克隆剪刀载体DBN01-T,其构建流程如图1所示(其中,Amp表示氨苄青霉素抗性基因;f1表示噬菌体f1的复制起点;LacZ为LacZ起始密码子;SP6为SP6RNA聚合酶启动子;T7为T7RNA聚合酶启动子;Csy4-T2A-FokI-dCas9为Csy4-T2A-FokI-dCas9核苷酸序列(Csy4-T2A-FokI-dCas9核苷酸序列如SEQ ID NO:1所示;Csy4氨基酸序列如SEQ ID NO:2所示,T2A多肽序列如SEQ ID NO:3所示;FokI的氨基酸序列如SEQ ID NO:4所示;dCas9氨基酸序列如SEQ ID NO:5所示);MCS为多克隆位点)。

[0081] 然后将重组克隆剪刀载体DBN01-T用热激方法转化大肠杆菌T1感受态细胞

(Transgen, Beijing, China, CAT:CD501), 其热激条件为:50 μ L大肠杆菌T1感受态细胞、10 μ L质粒DNA(重组克隆剪刀载体DBN01-T), 42 $^{\circ}$ C水浴90秒;37 $^{\circ}$ C振荡培养1小时(100rpm转速下摇床摇动), 在表面涂有IPTG(异丙基硫代- β -D-半乳糖苷)和X-gal(5-溴-4-氯-3-吡啶- β -D-半乳糖苷)的氨苄青霉素(100mg/L)的LB固体培养基(胰蛋白胨10g/L, 酵母提取物5g/L, NaCl 10g/L, 琼脂15g/L, 用NaOH调pH至7.5)上生长过夜。挑取白色菌落, 在LB液体培养基(胰蛋白胨10g/L, 酵母提取物5g/L, NaCl 10g/L, 氨苄青霉素100mg/L, 用NaOH调pH至7.5)中于温度37 $^{\circ}$ C条件下培养过夜。碱法提取其质粒:将菌液在12000rpm转速下离心1min, 去上清液, 沉淀菌体用100 μ L冰预冷的溶液I(25mM Tris-HCl, 10mM EDTA(乙二胺四乙酸), 50mM葡萄糖, pH8.0)悬浮;加入200 μ L新配制的溶液II(0.2M NaOH, 1%SDS(十二烷基硫酸钠)), 将管子颠倒4次, 混合, 置冰上3-5min;加入150 μ L冰冷的溶液III(3M醋酸钾, 5M醋酸), 立即充分混匀, 冰上放置5-10min;于温度4 $^{\circ}$ C、转速12000rpm条件下离心5min, 在上清液中加入2倍体积无水乙醇, 混匀后室温放置5min;于温度4 $^{\circ}$ C、转速12000rpm条件下离心5min, 弃上清液, 沉淀用浓度(V/V)为70%的乙醇洗涤后晾干;加入30 μ L含RNase(20 μ g/mL)的TE(10mM Tris-HCl, 1mM EDTA, pH8.0)溶解沉淀;于温度37 $^{\circ}$ C下水浴30min, 消化RNA;于温度-20 $^{\circ}$ C保存备用。

[0082] 提取的质粒经SnaBI和SpeI酶切鉴定后, 对阳性克隆进行测序验证, 结果表明重组克隆剪刀载体DBN01-T中插入的所述FokI-dCas9核苷酸序列为序列表中SEQ ID NO:1所示的核苷酸序列, 即Csy4-T2A-FokI-dCas9核苷酸序列正确插入。

[0083] 2. 构建重组表达剪刀载体

[0084] 用限制性内切酶SnaBI和SpeI分别酶切重组克隆剪刀载体DBN01-T和表达载体DBN-PAT, 将切下的Csy4-T2A-FokI-dCas9核苷酸序列插入表达载体DBN-PAT, 利用常规的酶切方法构建载体是本领域技术人员所熟知的, 构建成重组表达载体DBN-GET326, 其构建流程如图2所示(RB:右边界;pr35S:花椰菜花叶病毒35S启动子(SEQ ID NO:6);Csy4-T2A-FokI-dCas9:Csy4-T2A-FokI-dCas9核苷酸序列(Csy4-T2A-FokI-dCas9核苷酸序列如SEQ ID NO:1所示;Csy4氨基酸序列如SEQ ID NO:2所示;T2A多肽序列如SEQ ID NO:3所示;FokI的氨基酸序列如SEQ ID NO:4所示;dCas9氨基酸序列如SEQ ID NO:5所示);t35S:花椰菜花叶病毒35S终止子(SEQ ID NO:7);PAT:草丁膦乙酰转移酶基因(SEQ ID NO:8);LB:左边界)。

[0085] 将重组表达载体DBN-GET326用热激方法转化大肠杆菌T1感受态细胞, 其热激条件为:50 μ L大肠杆菌T1感受态细胞、10 μ L质粒DNA(重组表达载体DBN-GET326), 42 $^{\circ}$ C水浴90秒;37 $^{\circ}$ C振荡培养1小时(100rpm转速下摇床摇动);然后在含50mg/L卡那霉素(Kanamycin)的LB固体培养基(胰蛋白胨10g/L, 酵母提取物5g/L, NaCl 10g/L, 琼脂15g/L, 用NaOH调pH至7.5)上于温度37 $^{\circ}$ C条件下培养12小时, 挑取白色菌落, 在LB液体培养基(胰蛋白胨10g/L, 酵母提取物5g/L, NaCl 10g/L, 卡那霉素50mg/L, 用NaOH调pH至7.5)中于温度37 $^{\circ}$ C条件下培养过夜。碱法提取其质粒。将提取的质粒用限制性内切酶SnaBI和SpeI酶切后鉴定, 并将阳性克隆进行测序鉴定, 结果表明重组表达载体DBN-GET326在SnaBI和SpeI位点间的核苷酸序列为序列表中SEQ ID NO:1所示核苷酸序列, 即Csy4-T2A-FokI-dCas9核苷酸序列。

[0086] 第二实施例、水稻GUUS验证载体的构建

[0087] 1、GUUS靶点的选择

[0088] 将GUUS之间的靶序列信息输入到ZiFiT网站

[0089] (<http://zifit.partners.org/ZiFiT/ChoiceMenu.aspx>) 中,选择一对可用的靶点,即靶点1序列(如SEQ ID NO:9所示)和靶点2序列(如SEQ ID NO:10所示)。

[0090] 2、水稻无靶点载体的构建

[0091] 本实施例中,无靶点载体设计为pr0sU6+sgRNA+t35S结构。向第一实施例中所述pDBN骨架载体引入PMI表达盒、GUUS表达盒,利用常规的酶切方法构建载体是本领域技术人员所熟知的,构建成水稻无靶点载体DBN-GET344,其载体结构示意图如图3所示(LB:左边界;pr0sU6:水稻U6启动子(SEQ ID NO:14);Csy4-R:Csy4切割识别序列(如SEQ ID NO:11所示);sgRNA:sgRNA序列(如SEQ ID NO:15所示);t35S:花椰菜花叶病毒35S终止子(SEQ ID NO:7);pr35S:花椰菜花叶病毒35S启动子(SEQ ID NO:6);GUUS:含有靶点1序列和靶点2序列的GUS基因(如SEQ ID NO:16所示);tNos:胭脂碱合成酶基因的终止子(SEQ ID NO:17);prUbi:玉米泛素(Ubiquitin)基因启动子(SEQ ID NO:18);PMI:磷酸甘露糖异构酶基因(SEQ ID NO:19);RB:右边界)。

[0092] 按照第一实施例2中的方法,将无靶点载体DBN-GET344用热激方法转化大肠杆菌;碱法提取的质粒经AscI和AvrII酶切鉴定后,对阳性克隆进行测序验证,结果表明无靶点载体DBN-GET344构建正确。

[0093] 3、水稻靶点载体的构建

[0094] 本实施例中,靶点载体设计为pr0sU6+靶点+sgRNA+t35S结构。两个靶点+sgRNA之间连接有Csy4切割识别序列。通过限制性内切酶BsaI将各个片段无缝连接在一起。Csy4切割识别序列如SEQ ID NO:11所示。本实施例中使用的2个靶点为:

[0095] 靶点1序列,如SEQ ID NO:9所示;

[0096] 靶点2序列,如SEQ ID NO:10所示。

[0097] 引入靶点1和靶点2的引物如下:

[0098] 正向引物:**acatcaggtctc**caaaccgaggcattggtgcttcttggttttagagctagaata,如SEQ ID NO:12所示;

[0099] 反向引物:taggatggtctc**gaaa**acgtcgaggatgcctgggttgccctgcctatacggcagtgaaccac,如SEQ ID NO:13所示;

[0100] 其中,上述引物5'端的粗体小写字母为保护碱基,斜体小写字母为酶切位点BsaI,带下划线的小写字母为酶切位点BsaI的粘性末端。

[0101] 以合成的sgRNA+cys4识别序列为模板(扩增体系中<250ng),将靶点1序列和靶点2序列通过上述正向引物和反向引物带入模板,用Pfu酶(NEB)进行PCR扩增,PCR体系如下:

	Phusion® DNA 聚合酶	0.5 μL
	5X Phusion® HF or GC Buffer	10 μL
	10 μM 正向引物	2.5 μL
[0102]	10 μM 反向引物	2.5 μL
	10 mM dNTPs	1 μL
	二甲基亚砷(DMSO)	1.5 μL
	补充水(ddH ₂ O)至	50 μL

[0103] PCR反应条件为:98℃预变性30s,然后进入下列循环:98℃变性10s,56-60℃退火30s,72℃延伸30s/kb,共30-32个循环,最后72℃延伸5-10min;于4℃保存。

[0104] PCR扩增后得到含有靶位点序列+sgRNA+Csy4切割识别序列的产物,使用过柱纯化试剂盒(购自北京全式金生物技术有限公司)对上述PCR产物过柱纯化,具体方法参考其产品说明书;BsaI酶切PCR产物和表达载体DBN-GET344,切胶回收对应酶切产物后,将酶切后的表达载体DBN-GET344产物和PCR产物按照1:10的比例用T4连接酶于温度16℃下连接30min,利用常规的酶切方法构建载体是本领域技术人员所熟知的,构建成水稻靶点载体DBN-GET345,其载体结构示意图如图4所示(LB:左边界;pr0sU6:水稻U6启动子(SEQ ID NO:14);Csy4-R:Csy4切割识别序列(如SEQ ID NO:11所示);靶点1:靶1序列(SEQ ID NO:9);靶点2:靶点2序列(SEQ ID NO:10);sgRNA:sgRNA序列(如SEQ ID NO:15所示);t35S:花椰菜花叶病毒35S终止子(SEQ ID NO:7);pr35S:花椰菜花叶病毒35S启动子(SEQ ID NO:6);GUUS:含有靶点1序列和靶点2序列的GUS基因(如SEQ ID NO:16所示);tNos:胭脂碱合成酶基因的终止子(SEQ ID NO:17);prUbi:玉米泛素(Ubiquitin)基因启动子(SEQ ID NO:18);PMI:磷酸甘露糖异构酶基因(SEQ ID NO:19);RB:右边界)。

[0105] 按照第一实施例2中的方法,将靶点载体DBN-GET345用热激方法转化大肠杆菌;碱法提取的质粒经KpnI和AscI酶切鉴定后,对阳性克隆进行测序验证,结果表明靶点载体DBN-GET345中2个靶点(靶点1和靶点2)正确插入。

[0106] 第三实施例、剪刀载体和GUUS验证载体转化农杆菌

[0107] 将已经构建正确的重组表达载体DBN-GET326、DBN-GET344和DBN-GET345用液氮法转化到农杆菌LBA4404中,其转化条件为:100μL农杆菌LBA4404、3μL质粒DNA(重组表达载体);置于液氮中5分钟,37℃温水浴5分钟;将转化后的农杆菌LBA4404接种于LB试管中于温度28℃、转速为200rpm条件下培养2小时,涂于含50mg/L的利福平(Rifampicin)和50mg/L的卡那霉素的LB固体培养基上直至长出阳性单克隆,挑取单克隆培养并提取其质粒,用限制性内切酶进行酶切验证,结果表明重组表达载体DBN-GET326、DBN-GET344和DBN-GET345结构完全正确。

[0108] 按照以下组合将菌液等体积混合:DBN-GET326和DBN-GET345菌液(靶点处理)、DBN-GET326和DBN-GET344菌液(无靶点处理)、DBN-GET344菌液(对照处理),室温静置3h,获得对应处理的农杆菌悬浮液。

[0109] 第四实施例、获得稳定转化的水稻愈伤

[0110] 对于农杆菌介导的水稻转化,简要地,把水稻种子(日本晴,中国农业大学提供)接种在诱导培养基(N6盐3.1g/L、N6维他命、干酪素300mg/L、麦芽糖30g/L、2,4-二氯苯氧乙酸(2,4-D)2mg/L、植物凝胶3g/L,pH5.8)上,从水稻成熟胚诱导出愈伤组织(步骤1:愈伤诱导步骤),之后,优选愈伤组织,用上述3种处理的农杆菌悬浮液接触愈伤组织,其中农杆菌能够将目的构建体传递至愈伤组织上的至少一个细胞(步骤2:侵染步骤)。在此步骤中,愈伤组织优选地浸入农杆菌悬浮液(OD₆₆₀=0.3,侵染培养基(N6盐3.1g/L、N6维他命、干酪素300mg/L、蔗糖30g/L、葡萄糖10g/L、乙酰丁香酮(AS)40mg/L、2,4-二氯苯氧乙酸(2,4-D)2mg/L、pH5.4))中以启动接种。愈伤组织与农杆菌共培养一段时期(3天)(步骤3:共培养步骤)。优选地,愈伤组织在侵染步骤后在固体培养基(N6盐3.1g/L、N6维他命、干酪素300mg/L、蔗糖30g/L、葡萄糖10g/L、乙酰丁香酮(AS)40mg/L、2,4-二氯苯氧乙酸(2,4-D)2mg/L、植

物凝胶3g/L, pH5.8) 上培养。在此共培养阶段后, 有一个“恢复”步骤。在“恢复”步骤中, 恢复培养基(N6盐3.1g/L、N6维他命、干酪素300mg/L、蔗糖30g/L、2,4-二氯苯氧乙酸(2,4-D) 2mg/L、植物凝胶3g/L, pH5.8) 中至少存在一种已知抑制农杆菌生长的抗生素(头孢霉素150-250mg/L), 不添加植物转化体的选择剂(步骤4: 恢复步骤)。优选地, 愈伤组织在有抗生素但没有选择剂的固体培养基上培养, 以消除农杆菌并为侵染细胞提供恢复期。接着, 接种的愈伤组织在含选择剂(甘露糖和/或草铵膦) 的培养基上培养并选择生长着的转化愈伤组织(步骤5: 选择步骤)。优选地, 所述靶点处理的愈伤组织和所述无靶点处理的愈伤组织在有甘露糖和草铵膦的筛选固体培养基(N6盐3.1g/L、N6维他命、干酪素300mg/L、蔗糖5g/L、甘露糖12.5g/L、草铵膦4mg/L、2,4-二氯苯氧乙酸(2,4-D) 2mg/L、植物凝胶3g/L, pH5.8) 上培养, 所述对照处理的愈伤组织在有甘露糖的筛选固体培养基(N6盐3.1g/L、N6维他命、干酪素300mg/L、蔗糖5g/L、甘露糖12.5g/L、2,4-二氯苯氧乙酸(2,4-D) 2mg/L、植物凝胶3g/L, pH5.8) 上培养, 导致转化的细胞选择性生长。将上述筛选得到的抗性愈伤组织进行GUS染色分析。

[0111] 第五实施例、水稻愈伤的GUS染色测定

[0112] 分别取稳定转化获得的靶点处理的抗性愈伤组织、无靶点处理的抗性愈伤组织和对照处理的抗性愈伤组织作为样品, 参照Jefferson等(Jefferson R.A., Burgess S.M., Hirsh D. Beta-glucuronidase from *Escherichia coli* as a gene fusion marker. *Proc. Natl. Acad. Sci.*, 1986, 83:8447-8454) 的方法, 通过在GUS染色液中37°C密封染色1-2天, 从组织化学上检验GUS的表达方式, 即GUS突变为GUS酶后会将X-gluc原位分解产生蓝色沉淀, 从而可以说明FokI-dCas9有助于使GUS恢复GUS染色。每个处理设3次重复, 每个重复做10个抗性愈伤组织, 取平均值。具体方法如下:

[0113] 步骤1、5-溴-4-氯-3-吡啶-β-D-葡萄糖苷酸(X-gluc) 按照40mg/mL的浓度溶解到二甲亚砜(DMSO) 中, 用锡箔纸密封后置于-80°C冰箱中保存;

[0114] 步骤2、配制GUS染色液: 100mM NaH₂PO₄、10mM Na₂EDTA、0.5mM K₄[Fe(CN)₆] • 3H₂O、0.5mM K₃[Fe(CN)₆]、体积比为1%的聚乙二醇辛基苯基醚(Triton X-100), 用pH计调pH至7.0, 加水定容至1L;

[0115] 步骤3、在步骤2配制好的GUS染色液中加入步骤1密封保存的X-Gluc, 使X-Gluc终浓度为0.5mg/mL, 用于GUS染色;

[0116] 步骤4、分别取靶点处理的抗性愈伤组织、无靶点处理的抗性愈伤组织和对照处理的抗性愈伤组织各30个, 每3粒抗性愈伤组织放入1支2mL离心管中, 加入步骤3获得的GUS染色液并没过样品, 于37°C恒温箱中放置24-48h后, 目测染色情况。

[0117] GUS染色结果如表1所示, 在上述GUS染色验证同源重组效率的实验中, 将GUS染色的程度分为四个等级, 即+++、++、+、-, 依次表示大部分细胞深蓝色、仅不到一半细胞蓝色、少数细胞蓝色、无蓝色), GUS染色标准如图5所示, GUS染色实验结果如表1所示, 约有24%所述靶点处理的抗性愈伤组织发生了GUS回复突变(染色程度为+的占14.00%), 说明靶点与FokI-dCas9融合蛋白共转化时可以促进同源重组的发生; 仅有3.00%的所述对照处理的抗性愈伤组织的少数细胞被GUS染成蓝色(染色程度为+); 值得注意的是, 在没有靶点存在的情况下, 约有17.20%所述无靶点处理的抗性愈伤组织发生了GUS回复突变(染色程度为+), 比所述对照处理的抗性愈伤组织的同源重组效率提高4.7倍, 说明当不存在切割(无靶点)

时,FokI-dCas9融合蛋白能单独促进同源重组的产生,并且通过过量表达能显著提高同源重组效率。

表 1.不同处理的抗性愈伤组织 GUS 染色结果处理	GUS 染色程度及对应比例			
	+++	++	+	-
靶点处理	0.80%	8.80%	14.00%	76.40%
无靶点处理	/	/	17.20%	82.80%
对照处理	/	/	3.00%	97.00%

[0118] 综上所述,本发明了首次提供了FokI-dCas9融合蛋白能够促进同源重组的发生,并且显著提高细胞内同源重组效率,同时也大大的减少了对转化受体的需求,为高效同源重组编辑提供了一种新的选择。

[0119] 最后所应说明的是,以上实施例仅用以说明本发明的技术方案而非限制,尽管参照较佳实施例对本发明进行了详细说明,本领域的普通技术人员应当理解,可以对本发明的技术方案进行修改或者等同替换,而不脱离本发明技术方案的精神和范围。

[0001]	SEQUENCE LISTING	
[0002]	<110> 北京大北农生物技术有限公司	
[0003]	<120> 提高同源重组效率的方法	
[0004]	<130> DBNBC120	
[0005]	<160> 19	
[0006]	<170> PatentIn version 3.3	
[0007]	<210> 1	
[0008]	<211> 5523	
[0009]	<212> DNA	
[0010]	<213> Artificial sequence	
[0011]	<220>	
[0012]	<223> Csy4-T2A-FokI-dCas9核苷酸序列	
[0013]	<400> 1	
[0014]	atgggcgacc actacctgga catcaggctg aggccggacc cggagttccc gccggcccag	60
[0015]	ctgatgagcg tgctgttcgg caagctgcac caggcactgg tggcccaggg cggcgacagg	120
[0016]	atcggcgtga gcttcccgga cctggacgag agcaggagca ggctgggcga gaggctgaga	180
[0017]	atccacgcca gcgccgacga cctgagggca ctgctggcca ggccgtggct ggagggcctg	240
[0018]	agggaccacc tgcaattcgg cgagccggcc gtggtgccgc acccgacccc gtacaggcag	300
[0019]	gtgagcaggg tgcaggccaa gagcaaccgg gagaggctga ggaggaggct gatgaggagg	360
[0020]	cacgacctga gcgaggagga ggccaggaag agaatcccgg acaccgtggc aagggcctg	420
[0021]	gacctgccgt tcgtgacctt gaggagccag agcaccggcc agcacttcag gctgttcata	480
[0022]	aggcacggcc cgctacaggt gaccgccgag gaggcggtt tcacctgcta cggcctgagc	540
[0023]	aaggcggtt tcgtgccgtg gttcgagggc agggcgagcc tgctgacctg cggcgacgtg	600
[0024]	gaggagaacc cgggcccgat gccgaagaag aagaggaagg tgcctccca gctcgtgaag	660
[0025]	tccgagctcg aggagaagaa gtccgagctc cgccacaagc tcaagtacgt gccgcacgag	720
[0026]	tacatcgagc tcatcgagat cgcccgaac tccaccagg accgcatcct cgagatgaag	780
[0027]	gtgatggagt tcttcatgaa ggtgtacggc taccgcgca agcacctcgg cggctcccgc	840
[0028]	aagccggacg gcgccatcta caccgtgggc tccccgatcg actacggcgt gatcgtggac	900
[0029]	accaaggcct actccggcgg ctacaacctc ccgatcggcc aggccgacga gatgcagcgc	960
[0030]	tacgtggagg agaaccagac ccgcaacaag cacatcaacc cgaacgagtg gtggaagggtg	1020
[0031]	taccgtcctt ccgtgaccga gttcaagttc ctcttcgtgt ccggccactt caagggaac	1080
[0032]	tacaaggccc agctcaccgg cctcaaccac ataccaact gcaacggcgc cgtgctctcc	1140
[0033]	gtggaggagc tcctcatcgg cggcgagatg atcaaggcgg gcaccctcac cctcgaggag	1200
[0034]	gtgcgccgca agttcaaca cggcgagatc aacttcggcg gcggcgagcag catggactac	1260
[0035]	aaggaccag acggggatta caaagaccac gacatagact acaaggatga cgatgacaaa	1320
[0036]	atggcaccga agaaaaaaaa gaaggtcgga atccatggcg ttccagctgc cgataagaaa	1380
[0037]	tattccatcg gactcgccat tggcacgaat agcgtcggat gggctgttat tactgatgag	1440
[0038]	tacaaagttc cgtctaagaa gttcaaggtg ctgggcaaca cagaccgcca cagcataaag	1500

[0039] aaaaatctca tcggtgcaact ccttttcgat agtggggaga ctgcagaagc gacaagattg 1560
[0040] aaaaggactg cgagaaggcg ctatacacgg cgtaagaata gaatctgcta ccttcaggag 1620
[0041] attttctcta acgaaatggc taaggtcgat gacagtttct ttcatagact tgaggaatcg 1680
[0042] ttcttggttg aggaggataa gaaacatgag aggcacccga tatttgaaa catcgtgat 1740
[0043] gaggtcgcat atcatgaaaa gtaccccaca atctaccacc tgagaaagaa actcgttgat 1800
[0044] tccaccgaca aagcggattt gagactcatc tacctcgctc ttgcccataat gataaagttc 1860
[0045] cgcggaactt ttctgatcga gggcgacctc aacctgata atagcgacgt cgataagctc 1920
[0046] ttcattccagt tggttcaaac ctacaatcag ctctttgagg aaaacccaat taatgctagt 1980
[0047] ggagtggatg caaaaagcga actgtcggcc agactctcca agagcagaag gttggagaac 2040
[0048] ctgatcgctc aacttcctgg agaaaagaaa aacggtcttt ttgggaattt gattgccttg 2100
[0049] tctctgggcc tcacacaaa cttcaagtca aattttgacc tcgctgagga tgccaaactt 2160
[0050] cagttgtcta aggataccta tgatgacgat cttgacaatt tgctggcaca aattggcgac 2220
[0051] cagtacgagg atctgttctt cgcagcgaag aatctgagtg atgctattct cctttcggac 2280
[0052] atactcaggg ttaacactga gatcacaaaa gcaccttga gtgcgtcgat gattaagcgc 2340
[0053] tatgatgaac atcaccaaga cctcactttg ctgaaggccc ttgtcggca gcaattgcca 2400
[0054] gagaagtaca aagaaatctt ctttgaccaa tctaagaac gatacctgg ctatattgat 2460
[0055] ggaggagctt ctcaggagga attctataag tttatcaaac ctatacttga gaagatggat 2520
[0056] ggtacagagg aactccttgt taaattgaac agagaagatt tgctgcgcaa gcaacggacc 2580
[0057] tttgacaac gatcaattcc gcatcagata cacctcggcg agcttcatgc catccttcgc 2640
[0058] cggcaggaag atttctacce ctttttgaag gacaaccgcg agaagataga aaaaatcctt 2700
[0059] acgttccgga ttccttacta tgtgggtcca ttggcaaggg ggaattcccg ctttgcgtgg 2760
[0060] atgactcgga aaagcgagga aactatcaca ccgtggaact tcgaggaagt tgtggacaag 2820
[0061] ggagcttctg cccaatcatt cattgagagg atgactaact tcgataagaa cctgccgaac 2880
[0062] gagaaagttc tcccaagca ctccctcctt tacgagtatt tcaccgtgta taacgaactt 2940
[0063] acgaaggta aatacgtgac tgagggtatg aggaagccag cattcttgag cggggaacaa 3000
[0064] aagaaagcga ttgttgattt gctgtttaaa actaatcgca aggtgacagt caagcagctc 3060
[0065] aaagaggatt atttcaagaa aattgaatgt ttcgactctg tggagatatac aggagtcgaa 3120
[0066] gataggttta acgcttccct tggcacatac catgacctc ttaagatcat taaggacaaa 3180
[0067] gatttcttgg ataacgagga aatgaggac atcctcgaag atattgttct taccttgacg 3240
[0068] ctgtttgagg atcgcgaaat gatcgaggaa cggttaaga cgtatgctca cttgttcgac 3300
[0069] gataaggta tgaagcagct caagcgtaga aggtacactg gatggggccg tctgtctaga 3360
[0070] aagctcatca acggaatac tgataaacia agtggcaaga caattttgga ttttctgaag 3420
[0071] tcggacggat tcgccaacag aaattttatg cagctgattc atgacgatag tctcacctc 3480
[0072] aaagaggaca tacagaaggc tcaagtgagt ggtcaagggg attcgtgca tgaacacatc 3540
[0073] gcaaacctcg cgggttcacc ggccataaag aaaggaatcc ttcaaactgt taaggtcgtt 3600
[0074] gatgagtttg ttaaagtgat gggtaggcac aagccgaaa acatagtgat cgagatggct 3660
[0075] cgcgaaaatc agactacaca aaaagggcag aagaactctc gcgagcggat gaaaaggatt 3720
[0076] gaggaaggaa tcaaggaact gggctcacag attctcaag agcatccagt cgaaaacaca 3780
[0077] cagctgcaaa atgagaagct ctatctttac tatctcaaaa atggccggga catgtatggt 3840

[0078] gatcaggagc ttgacatcaa ccgtttgtcc gactatgatg tggacgccat tgtcccgcaa 3900
 [0079] tctttcetta aggacgattc aatcgataat aagggttga cccggagcga taaaaaccgt 3960
 [0080] ggaaagtctg acaatgtccc ttcagaggaa gtggttaaga agatgaagaa ctactggaga 4020
 [0081] caattgctga atgcaaaact gatcacacag agaaagttcg acaacctcac caaagcagag 4080
 [0082] agaggtgggc tcagtgaact tgataaagcg ggcttcatta agcgtcagct cgttgagact 4140
 [0083] agacagatca cgaagcatgt cgcgcagatt ttggattcgc ggatgaacac gaagtacgac 4200
 [0084] gagaatgata aactgatacg tgaagtcaag gttatcactc ttaagtccaa attggtgagc 4260
 [0085] gatttcagaa aggacttcca attctataag gtcagggaga tcaacaatta tcatcacgct 4320
 [0086] cacgatgctt accttaatgc tgttgtgggg accgccetta ttaagaaata ccctaaattg 4380
 [0087] gagtctgaat tcgtttacgg ggattataag gtctacgacg ttaggaaaat gatagctaag 4440
 [0088] agtgagcagg agatcggtaa agcaactgcg aagtatttct tttactcgaa catcatgaat 4500
 [0089] ttctttaaga ccgagataac gctggcaaat ggcgaaatta gaaagaggcc tctcatagag 4560
 [0090] actaacggtg agacagggga aatcgtctgg gataagggtg gggactttgc gacagtgcgc 4620
 [0091] aaggctctct ctatgccgca agttaatatt gtgaagaaaa ccgaggtgca gacgggaggc 4680
 [0092] ttctccaagg aaagcatact tcccaaaccg aactctgata agttgatcgc tcgtaagaaa 4740
 [0093] gattgggacc ctaagaaata tgggtgggttc gattcccaa ctggttctta cagcgtgctg 4800
 [0094] gtcgttgcca aggtcgagaa gggtaaatcc aagaaactca aaagcgtaa ggaactcctt 4860
 [0095] gggattacta tcatggagag atcttcattc gaaaagaatc ctatcgactt tcttgaggcc 4920
 [0096] aaaggatata aggaagttaa gaaagatctg ataatacaac tcccaaagta ctctattgtt 4980
 [0097] gagctggaaa acggcaggaa gcgcatgctt gcttccgccg gagagttgca gaaagggaac 5040
 [0098] gagttggctc tgccttctaa gtatgttaac ttctctatc ttgcctctca ttacgagaag 5100
 [0099] ctcaaaggct caccagagga caacgaacag aaacaacttt ttgtcgagca acataagcac 5160
 [0100] tatttgatg agattataga acagatcagt gaattctcga aaagggttat ccttgcagat 5220
 [0101] gcgaatcttg acaaggtggt gtctgcatac acaaacata gagataagcc gatcaggag 5280
 [0102] caagcggaaa atatcattca cctcttact cttacaaact tgggtgctcc cgctgccttc 5340
 [0103] aagtattttg ataccacgat tgaccggaaa cgttacacct caacgaagga ggtgctggat 5400
 [0104] gccacctca tccaccaatc tattaccgga ctctacgaga ctagaatcga tctctcacag 5460
 [0105] ctggcgggg ataaaagacc agcagcgacg aaaaaggcag gacaggctaa gaagaagaaa 5520
 [0106] tag 5523
 [0107] <210> 2
 [0108] <211> 188
 [0109] <212> PRT
 [0110] <213> *Pseudomonas aeruginosa*
 [0111] <400> 2
 [0112] Met Gly Asp His Tyr Leu Asp Ile Arg Leu Arg Pro Asp Pro Glu Phe
 [0113] 1 5 10 15
 [0114] Pro Pro Ala Gln Leu Met Ser Val Leu Phe Gly Lys Leu His Gln Ala
 [0115] 20 25 30
 [0116] Leu Val Ala Gln Gly Gly Asp Arg Ile Gly Val Ser Phe Pro Asp Leu

[0117] 35 40 45
 [0118] Asp Glu Ser Arg Ser Arg Leu Gly Glu Arg Leu Arg Ile His Ala Ser
 [0119] 50 55 60
 [0120] Ala Asp Asp Leu Arg Ala Leu Leu Ala Arg Pro Trp Leu Glu Gly Leu
 [0121] 65 70 75 80
 [0122] Arg Asp His Leu Gln Phe Gly Glu Pro Ala Val Val Pro His Pro Thr
 [0123] 85 90 95
 [0124] Pro Tyr Arg Gln Val Ser Arg Val Gln Ala Lys Ser Asn Pro Glu Arg
 [0125] 100 105 110
 [0126] Leu Arg Arg Arg Leu Met Arg Arg His Asp Leu Ser Glu Glu Glu Ala
 [0127] 115 120 125
 [0128] Arg Lys Arg Ile Pro Asp Thr Val Ala Arg Ala Leu Asp Leu Pro Phe
 [0129] 130 135 140
 [0130] Val Thr Leu Arg Ser Gln Ser Thr Gly Gln His Phe Arg Leu Phe Ile
 [0131] 145 150 155 160
 [0132] Arg His Gly Pro Leu Gln Val Thr Ala Glu Glu Gly Gly Phe Thr Cys
 [0133] 165 170 175
 [0134] Tyr Gly Leu Ser Lys Gly Gly Phe Val Pro Trp Phe
 [0135] 180 185
 [0136] <210> 3
 [0137] <211> 18
 [0138] <212> PRT
 [0139] <213> Foot-and-mouth disease virus
 [0140] <400> 3
 [0141] Glu Gly Arg Gly Ser Leu Leu Thr Cys Gly Asp Val Glu Glu Asn Pro
 [0142] 1 5 10 15
 [0143] Gly Pro
 [0144] <210> 4
 [0145] <211> 198
 [0146] <212> PRT
 [0147] <213> Flavobacterium okeanokoites
 [0148] <400> 4
 [0149] Ser Ser Gln Leu Val Lys Ser Glu Leu Glu Glu Lys Lys Ser Glu Leu
 [0150] 1 5 10 15
 [0151] Arg His Lys Leu Lys Tyr Val Pro His Glu Tyr Ile Glu Leu Ile Glu
 [0152] 20 25 30
 [0153] Ile Ala Arg Asn Ser Thr Gln Asp Arg Ile Leu Glu Met Lys Val Met
 [0154] 35 40 45
 [0155] Glu Phe Phe Met Lys Val Tyr Gly Tyr Arg Gly Lys His Leu Gly Gly

[0156]	50	55	60
[0157]	Ser Arg Lys Pro Asp Gly Ala Ile Tyr Thr Val Gly Ser Pro Ile Asp		
[0158]	65	70	75 80
[0159]	Tyr Gly Val Ile Val Asp Thr Lys Ala Tyr Ser Gly Gly Tyr Asn Leu		
[0160]	85	90	95
[0161]	Pro Ile Gly Gln Ala Asp Glu Met Gln Arg Tyr Val Glu Glu Asn Gln		
[0162]	100	105	110
[0163]	Thr Arg Asn Lys His Ile Asn Pro Asn Glu Trp Trp Lys Val Tyr Pro		
[0164]	115	120	125
[0165]	Ser Ser Val Thr Glu Phe Lys Phe Leu Phe Val Ser Gly His Phe Lys		
[0166]	130	135	140
[0167]	Gly Asn Tyr Lys Ala Gln Leu Thr Arg Leu Asn His Ile Thr Asn Cys		
[0168]	145	150	155 160
[0169]	Asn Gly Ala Val Leu Ser Val Glu Glu Leu Leu Ile Gly Gly Glu Met		
[0170]	165	170	175
[0171]	Ile Lys Ala Gly Thr Leu Thr Leu Glu Glu Val Arg Arg Lys Phe Asn		
[0172]	180	185	190
[0173]	Asn Gly Glu Ile Asn Phe		
[0174]	195		
[0175]	<210> 5		
[0176]	<211> 1423		
[0177]	<212> PRT		
[0178]	<213> Artificial Sequence		
[0179]	<220>		
[0180]	<223> dCas9氨基酸序列		
[0181]	<400> 5		
[0182]	Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp		
[0183]	1 5 10 15		
[0184]	Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg Lys Val		
[0185]	20 25 30		
[0186]	Gly Ile His Gly Val Pro Ala Ala Asp Lys Lys Tyr Ser Ile Gly Leu		
[0187]	35 40 45		
[0188]	Ala Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr		
[0189]	50 55 60		
[0190]	Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His		
[0191]	65 70 75 80		
[0192]	Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu		
[0193]	85 90 95		
[0194]	Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr		

[0195]		100		105		110												
[0196]	Arg	Arg	Lys	Asn	Arg	Ile	Cys	Tyr	Leu	Gln	Glu	Ile	Phe	Ser	Asn	Glu		
[0197]			115						120						125			
[0198]	Met	Ala	Lys	Val	Asp	Asp	Ser	Phe	Phe	His	Arg	Leu	Glu	Glu	Ser	Phe		
[0199]			130						135						140			
[0200]	Leu	Val	Glu	Glu	Asp	Lys	Lys	His	Glu	Arg	His	Pro	Ile	Phe	Gly	Asn		
[0201]			145						150						155			160
[0202]	Ile	Val	Asp	Glu	Val	Ala	Tyr	His	Glu	Lys	Tyr	Pro	Thr	Ile	Tyr	His		
[0203]									165						170			175
[0204]	Leu	Arg	Lys	Lys	Leu	Val	Asp	Ser	Thr	Asp	Lys	Ala	Asp	Leu	Arg	Leu		
[0205]									180						185			190
[0206]	Ile	Tyr	Leu	Ala	Leu	Ala	His	Met	Ile	Lys	Phe	Arg	Gly	His	Phe	Leu		
[0207]									195						200			205
[0208]	Ile	Glu	Gly	Asp	Leu	Asn	Pro	Asp	Asn	Ser	Asp	Val	Asp	Lys	Leu	Phe		
[0209]									210						215			220
[0210]	Ile	Gln	Leu	Val	Gln	Thr	Tyr	Asn	Gln	Leu	Phe	Glu	Glu	Asn	Pro	Ile		
[0211]									225						230			235
[0212]	Asn	Ala	Ser	Gly	Val	Asp	Ala	Lys	Ala	Ile	Leu	Ser	Ala	Arg	Leu	Ser		
[0213]									245						250			255
[0214]	Lys	Ser	Arg	Arg	Leu	Glu	Asn	Leu	Ile	Ala	Gln	Leu	Pro	Gly	Glu	Lys		
[0215]									260						265			270
[0216]	Lys	Asn	Gly	Leu	Phe	Gly	Asn	Leu	Ile	Ala	Leu	Ser	Leu	Gly	Leu	Thr		
[0217]									275						280			285
[0218]	Pro	Asn	Phe	Lys	Ser	Asn	Phe	Asp	Leu	Ala	Glu	Asp	Ala	Lys	Leu	Gln		
[0219]									290						295			300
[0220]	Leu	Ser	Lys	Asp	Thr	Tyr	Asp	Asp	Asp	Leu	Asp	Asn	Leu	Leu	Ala	Gln		
[0221]									305						310			315
[0222]	Ile	Gly	Asp	Gln	Tyr	Ala	Asp	Leu	Phe	Leu	Ala	Ala	Lys	Asn	Leu	Ser		
[0223]									325						330			335
[0224]	Asp	Ala	Ile	Leu	Leu	Ser	Asp	Ile	Leu	Arg	Val	Asn	Thr	Glu	Ile	Thr		
[0225]									340						345			350
[0226]	Lys	Ala	Pro	Leu	Ser	Ala	Ser	Met	Ile	Lys	Arg	Tyr	Asp	Glu	His	His		
[0227]									355						360			365
[0228]	Gln	Asp	Leu	Thr	Leu	Leu	Lys	Ala	Leu	Val	Arg	Gln	Gln	Leu	Pro	Glu		
[0229]									370						375			380
[0230]	Lys	Tyr	Lys	Glu	Ile	Phe	Phe	Asp	Gln	Ser	Lys	Asn	Gly	Tyr	Ala	Gly		
[0231]									385						390			395
[0232]	Tyr	Ile	Asp	Gly	Gly	Ala	Ser	Gln	Glu	Glu	Phe	Tyr	Lys	Phe	Ile	Lys		
[0233]									405						410			415

[0234]	Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu
[0235]	420 425 430
[0236]	Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser
[0237]	435 440 445
[0238]	Ile Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg
[0239]	450 455 460
[0240]	Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu
[0241]	465 470 475 480
[0242]	Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg
[0243]	485 490 495
[0244]	Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile
[0245]	500 505 510
[0246]	Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln
[0247]	515 520 525
[0248]	Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu
[0249]	530 535 540
[0250]	Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr
[0251]	545 550 555 560
[0252]	Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro
[0253]	565 570 575
[0254]	Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe
[0255]	580 585 590
[0256]	Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe
[0257]	595 600 605
[0258]	Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp
[0259]	610 615 620
[0260]	Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile
[0261]	625 630 635 640
[0262]	Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu
[0263]	645 650 655
[0264]	Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu
[0265]	660 665 670
[0266]	Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys
[0267]	675 680 685
[0268]	Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys
[0269]	690 695 700
[0270]	Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp
[0271]	705 710 715 720
[0272]	Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile

[0273]		725		730		735
[0274]	His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val					
[0275]		740		745		750
[0276]	Ser Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly					
[0277]		755		760		765
[0278]	Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp					
[0279]		770		775		780
[0280]	Glu Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile					
[0281]	785		790		795	800
[0282]	Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser					
[0283]		805		810		815
[0284]	Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser					
[0285]		820		825		830
[0286]	Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu					
[0287]		835		840		845
[0288]	Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp					
[0289]		850		855		860
[0290]	Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp Ala Ile					
[0291]	865		870		875	880
[0292]	Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu					
[0293]		885		890		895
[0294]	Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu					
[0295]		900		905		910
[0296]	Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala					
[0297]		915		920		925
[0298]	Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg					
[0299]		930		935		940
[0300]	Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu					
[0301]	945		950		955	960
[0302]	Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser					
[0303]		965		970		975
[0304]	Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val					
[0305]		980		985		990
[0306]	Lys Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp					
[0307]		995		1000		1005
[0308]	Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala					
[0309]		1010		1015		1020
[0310]	His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys					
[0311]		1025		1030		1035

[0312]	Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys		
[0313]	1040	1045	1050
[0314]	Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile		
[0315]	1055	1060	1065
[0316]	Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn		
[0317]	1070	1075	1080
[0318]	Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys		
[0319]	1085	1090	1095
[0320]	Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp		
[0321]	1100	1105	1110
[0322]	Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met		
[0323]	1115	1120	1125
[0324]	Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly		
[0325]	1130	1135	1140
[0326]	Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu		
[0327]	1145	1150	1155
[0328]	Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe		
[0329]	1160	1165	1170
[0330]	Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val		
[0331]	1175	1180	1185
[0332]	Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu		
[0333]	1190	1195	1200
[0334]	Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile		
[0335]	1205	1210	1215
[0336]	Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu		
[0337]	1220	1225	1230
[0338]	Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly		
[0339]	1235	1240	1245
[0340]	Arg Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn		
[0341]	1250	1255	1260
[0342]	Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala		
[0343]	1265	1270	1275
[0344]	Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln		
[0345]	1280	1285	1290
[0346]	Lys Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile		
[0347]	1295	1300	1305
[0348]	Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp		
[0349]	1310	1315	1320
[0350]	Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp		

[0351]	1325	1330	1335
[0352]	Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr		
[0353]	1340	1345	1350
[0354]	Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr		
[0355]	1355	1360	1365
[0356]	Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp		
[0357]	1370	1375	1380
[0358]	Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg		
[0359]	1385	1390	1395
[0360]	Ile Asp Leu Ser Gln Leu Gly Gly Asp Lys Arg Pro Ala Ala Thr		
[0361]	1400	1405	1410
[0362]	Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys		
[0363]	1415	1420	
[0364]	<210> 6		
[0365]	<211> 328		
[0366]	<212> DNA		
[0367]	<213> Cauliflower mosaic virus		
[0368]	<400> 6		
[0369]	ccattgccca gctatctgtc actttattgt gaagatagtg gaaaaggaag gtggctccta 60		
[0370]	caaatgcat cattgcgata aaggaaaggc catcgttgaa gatgcctctg cgcacagtgg 120		
[0371]	tcccaaagat ggacccccac ccacgaggag catcgtggaa aaagaagacg ttccaaccac 180		
[0372]	gtcttcaaag caagtggatt gatgtgatat ctccactgac gtaagggatg acgcacaatc 240		
[0373]	ccactatcct tcgcaagacc ctctctctat ataaggaagt tcatttcatt tggagaggac 300		
[0374]	acgctgacaa gctgactcta gcagatct 328		
[0375]	<210> 7		
[0376]	<211> 195		
[0377]	<212> DNA		
[0378]	<213> Cauliflower mosaic virus		
[0379]	<400> 7		
[0380]	ctgaaatcac cagtctctct ctacaaatct atctctctct ataataatgt gtgagtagtt 60		
[0381]	cccagataag ggaattaggg ttcttatagg gtttcgctca tgtgttgagc atataagaaa 120		
[0382]	cccttagtat gtatttgtat ttgtaaaata ctctatcaa taaaatttct aattcctaaa 180		
[0383]	accaaaatcc agtgg 195		
[0384]	<210> 8		
[0385]	<211> 552		
[0386]	<212> DNA		
[0387]	<213> Streptomyces viridochromogenes		
[0388]	<400> 8		
[0389]	atgagccctg aaagacggcc tgtggagatt agaccagcga cggcagcgga catggcggcg 60		

[0390] gtgtgcgaca tcgtgaacca ttacatcgaa acttcaacgg tgaacttccg cacagagccc 120
[0391] caaacaccac aggagtggat cgacgatctg gagagacttc aagacagata cccgtggctt 180
[0392] gttgcagagg tcgagggcgt ggtcgcgggg atcgcgtatg cggccccgtg gaaggcgagg 240
[0393] aacgcctacg attggacagt ggaatccacc gtgtatgtca gccatcgcca ccagaggctg 300
[0394] ggcctcggca gcactctcta cacccatctc ctgaagagca tggaggcgca gggcttcaag 360
[0395] tccgtggtcg cagtgattgg cctgcctaac gatccatccg tgagactcca tgaggccctc 420
[0396] ggctacactg cgcgcggcac tctgcgcgcc gcgggtata agcacggcgg gtggcatgac 480
[0397] gtgggcttct ggcagagaga ctttgaactt cccgctcccc caagacctgt cagacccggt 540
[0398] acgcagatct aa 552
[0399] <210> 9
[0400] <211> 20
[0401] <212> DNA
[0402] <213> *Oryza sativa*
[0403] <400> 9
[0404] ggaggcattg gtgcttcttg 20
[0405] <210> 10
[0406] <211> 20
[0407] <212> DNA
[0408] <213> *Oryza sativa*
[0409] <400> 10
[0410] gcaaccagg catcctcgac 20
[0411] <210> 11
[0412] <211> 20
[0413] <212> DNA
[0414] <213> *Pseudomonas aeruginosa*
[0415] <400> 11
[0416] gttcactgcc gtataggcag 20
[0417] <210> 12
[0418] <211> 55
[0419] <212> DNA
[0420] <213> Artificial Sequence
[0421] <220>
[0422] <223> 正向引物
[0423] <400> 12
[0424] acatcaggtc tccaaacgga ggcattggtg cttcttggtt ttagagctag aaata 55
[0425] <210> 13
[0426] <211> 62
[0427] <212> DNA
[0428] <213> Artificial Sequence

- [0429] <220>
- [0430] <223> 反向引物
- [0431] <400> 13
- [0432] taggatggtc tcgaaaacgt cgaggatgcc tgggttgcc gcctatacgg cagtgaacgc 60
- [0433] ac 62
- [0434] <210> 14
- [0435] <211> 245
- [0436] <212> DNA
- [0437] <213> *Oryza sativa*
- [0438] <400> 14
- [0439] ggatcatgaa ccaacggcct ggctgtatatt ggtggttgtag tagggagatg gggagaagaa 60
- [0440] aagcccgatt ctcttcgctg tgatgggctg gatgcatgcg ggggagcggg aggcccaagt 120
- [0441] acgtgcacgg tgagcggccc acagggcgag tgtgagcgcg agaggcggga ggaacagttt 180
- [0442] agtaccacat tgcccagcta actcgaacgc gaccaactta taaaccgcg cgctgtcgcct 240
- [0443] tgtgt 245
- [0444] <210> 15
- [0445] <211> 76
- [0446] <212> DNA
- [0447] <213> Artificial Sequence
- [0448] <220>
- [0449] <223> sgRNA序列
- [0450] <400> 15
- [0451] gttttagagc tagaaatagc aagttaaaat aaggctagtc cgttatcaac ttgaaaaagt 60
- [0452] ggcaccgagt cgggtgc 76
- [0453] <210> 16
- [0454] <211> 3536
- [0455] <212> DNA
- [0456] <213> Artificial Sequence
- [0457] <220>
- [0458] <223> 含有靶点1序列和靶点2序列的GUS基因
- [0459] <400> 16
- [0460] atggtagatc tgaggtaaa tttctagttt ttctcttca tttcttggg taggaccctt 60
- [0461] ttctcttttt atttttttga gctttgatct ttctttaac tgatctatit ttttaattgat 120
- [0462] tggttatggt gtaaatatta catagcttta actgataatc tgattactit atttctgtg 180
- [0463] tctatgatga tgatgatagt tacagaaccg acgaacttct ctgtaccga tcaacaccga 240
- [0464] aaccgtggc gtcttcgacc tcaatggcgt ctggaactc aagctggact acgggaaagg 300
- [0465] actggaagag aagtggtagc aaagcaagct gaccgacact attagtagtg cgtcccaag 360
- [0466] cagttacaat gacattggcg tgaccaagga aatccgcaac catatcggat atgtctggt 420
- [0467] cgaacgtgag ttcacggtgc cggcctatct gaaggatcag cgtatcgtgc tccgcttcgg 480

[0468]	ctctgcaact	cacaaagcaa	ttgtctatgt	caatggtgag	ctggctcgtgg	agcacaaggg	540
[0469]	cggattcctg	ccattcgaag	cggaaatcaa	caactcgtctg	cgtgatggca	tgaatcgcgt	600
[0470]	caccgtcgcc	gtggacaaca	tcctcgacga	tagcacctc	ccggtggggc	tgtacagcga	660
[0471]	gcgccacgaa	gagggcctcg	gaaaagtc	tcgtaacaag	ccgaacttcg	acttcttcaa	720
[0472]	ctatgcaggc	ctgcaccgtc	cggtgaaaat	ctacacgacc	ccgtttacgt	acgtcgagga	780
[0473]	catctcggtt	gtgaccgact	tcaatggccc	aaccgggact	gtgacctata	cggtggtactt	840
[0474]	tcaaggcaaa	gccgaaaacc	tgaactgaac	tgaactgaag	gttatgacat	tccaagcgga	900
[0475]	tggaagatcc	tgccggtggt	agccgcggtg	catctggact	cgtccctgta	cgaggacccc	960
[0476]	cagcgttca	atccctggag	atggaaggtc	agtcgcaata	ggattatcag	tgtctcaagg	1020
[0477]	cgccattcag	ttccccgtgt	tccacaagaa	gcaccaatgc	ctccgccc	ggtctgtccg	1080
[0478]	tgcaaccag	gcatacctga	ccggagcattc	aggagcagga	aaaggaggag	gattgaacaa	1140
[0479]	tctacaggaa	gaggtctaaa	aagctgctctg	tgcggtggt	ggcttctgc	actgcatgca	1200
[0480]	ggtcgtatct	tgcgacgggc	gacggcgcgc	gtcgaggcgt	tggcggcatg	cgcggtcattc	1260
[0481]	gctcacgcgt	ccgcggggat	ggtggcctgc	ggtgaccgcg	gagcttgtaa	ggataatgag	1320
[0482]	gtactggctg	gaaggcccaa	gagcgggcga	ggtagaggtg	ttcgcgaacc	tgccgggctt	1380
[0483]	ccccgacaac	gtgcgtcca	acggcagggg	ccagttctgg	gtggcgatcg	actgctgccg	1440
[0484]	gacgccggcg	caggagggtg	tcgccaagag	gccgtggctc	cggacctat	acttcaagtt	1500
[0485]	cccgtgtcg	ctcaagggtc	tacttgga	ggccgccagg	aggatgcaca	cgggtgctcgc	1560
[0486]	gctcctcgac	ggcgaagggc	gcgtcgtgga	ggtgctcgag	gaccggggcc	acgaggtgat	1620
[0487]	gaagctggtg	agtgaggtgc	gggaggtggg	cagcaagctg	tggatcgga	ccgtggcgca	1680
[0488]	caaccacatc	gccaccatcc	cctaccctt	agaggactaa	ttttaccctg	ggcgtcttcg	1740
[0489]	acctcaatgg	cgtctggaac	ttcaagctgg	actacgggaa	aggactggaa	gagaagtgg	1800
[0490]	acgaaagcaa	gctgaccgac	actattagta	tggccgtccc	aagcagttac	aatgacattg	1860
[0491]	gcgtgaccaa	ggaaatccgc	aaccatattc	gatattgtctg	gtacgaacgt	gagttcacgg	1920
[0492]	tgccggccta	tctgaaggat	cagcgtatcg	tgctccgctt	cggctctgca	actcacaag	1980
[0493]	caattgtcta	tgtcaatggt	gagctggtcg	tggagcaca	ggcggattc	ctgccattcg	2040
[0494]	aagcggaaat	caacaactcg	ctgcgtgatg	gcatgaatcg	cgtcacccgtc	gccgtggaca	2100
[0495]	acatacctga	cgatagcacc	ctcccgttg	ggctgtacag	cgagcggcc	gaagagggcc	2160
[0496]	tcgaaaagt	cattcgtaac	aagccgaact	tcgacttctt	caactatgca	ggcctgcacc	2220
[0497]	gtccggtgaa	aatctacacg	accccgttta	cgtagctcga	ggacatctcg	gttgtgaccg	2280
[0498]	acttcaatgg	cccaaccggg	actgtgacct	atacgggtgga	ctttcaaggc	aaagccgaaa	2340
[0499]	ccgtgaaagt	gtcggctcgtg	gatgaggaag	gcaaagtgg	cgcaagcacc	gagggcctga	2400
[0500]	gcggtaacgt	ggagattccg	aatgtcatcc	tctgggaacc	actgaacacg	tatctctacc	2460
[0501]	agatcaaagt	ggaactggtg	aacgacggac	tgaccatcga	tgtctatgaa	gagccgttcg	2520
[0502]	gcgtgcggac	cgtggaagtc	aacgacggca	agttctctcat	caacaacaaa	ccgttctact	2580
[0503]	tcaagggtt	tggcaaacat	gaggacactc	ctataacgg	ccgtggcttt	aacgaagcga	2640
[0504]	gcaatgtgat	ggatttcaat	atcctcaaat	ggatcgggtgc	caacagcttc	cggaccgcac	2700
[0505]	actatccgta	ctctgaagag	ttgatgcgtc	ttgcggatcg	cgagggtctg	gtcgtgatcg	2760
[0506]	acgagactcc	ggcagttggc	gtgcacctca	acttcatggc	caccacggga	ctcggcgaag	2820

[0507] gcagcgagcg cgtcagtacc tgggagaaga ttcggacgtt tgagcaccat caagacgttc 2880
 [0508] tccgtgaact ggtgtctcgt gacaagaacc atccaagcgt cgtgatgtgg agcatcgcca 2940
 [0509] acgaggcggc gactgaggaa gagggcgcgt acgagtactt caagccgttg gtggagctga 3000
 [0510] ccaaggaact cgacccacag aagcgtccgg tcacgatcgt gctgtttgtg atggctaccc 3060
 [0511] cggagacgga caaagtcgcc gaactgattg acgtcatcgc gctcaatcgc tataacggat 3120
 [0512] ggtacttcca tggcggatgat ctccaagcgg ccaaagtcca tctccgccag gaatttcacg 3180
 [0513] cgtggaacaa gcgttgccca ggaaagccga tcatgatcac tgagtacggc gcagacaccg 3240
 [0514] ttgcgggctt tcacgacatt gatccagtga tgttcaccga ggaatatcaa gtcgagtact 3300
 [0515] accaggcgaa ccacgtcgtg ttcgatgagt ttgagaactt cgtgggtgag caagcgtgga 3360
 [0516] acttcgcgga cttcgcgacc tctcagggcg tgatgcgcgt ccaaggaaac aagaagggcg 3420
 [0517] tgttactcgc tgaccgaag ccgaagctcg ccgcgcacgt ctttcgcgag cgctggacca 3480
 [0518] acattccaga tttcggctac aagaacgcta gccatcacca tcaccatcac gtgtga 3536
 [0519] <210> 17
 [0520] <211> 253
 [0521] <212> DNA
 [0522] <213> *Agrobacterium tumefaciens*
 [0523] <400> 17
 [0524] gatcgttcaa acatttggca ataaagtttc ttaagattga atcctgttgc cggctcttgcg 60
 [0525] atgattatca tataatttct gttgaattac gttaagcatg taataattaa catgtaatgc 120
 [0526] atgacgttat ttatgagatg ggtttttatg attagagtcc cgcaattata catttaatac 180
 [0527] gcgatagaaa acaaaatata gcgcgcaaac taggataaat tatcgcgcg cgtgtcatct 240
 [0528] atgttactag atc 253
 [0529] <210> 18
 [0530] <211> 1992
 [0531] <212> DNA
 [0532] <213> *Zea Mays*
 [0533] <400> 18
 [0534] ctgcagtgca gcgtgaccgc gtcgtgcccc tctctagaga taatgagcat tgcattgtcta 60
 [0535] agttataaaa aattaccaca tttttttttt gtcacacttg tttgaagtgc agttttatcta 120
 [0536] tctttataca tatatttaaa ctttactcta cgaataatat aatctatagt actacaataa 180
 [0537] tatcagtgtt ttagagaatc atataaatga acagttagac atggtctaaa ggacaattga 240
 [0538] gtattttgac aacaggactc tacagtttta tctttttagt gtgcatgtgt tctccttttt 300
 [0539] ttttgcaaat agcttcacct atataatact tcatccattt tattagtaca tccatttagg 360
 [0540] gtttaggggtt aatggttttt atagactaat ttttttagta catctatttt attctatttt 420
 [0541] agcctctaaa ttaagaaaac taaaactcta ttttagtttt tttatttaaat aatttagata 480
 [0542] taaaatagaa taaaataaag tgactaaaaa ttaaacaat accctttaag aaattaaaaa 540
 [0543] aactaaggaa acatttttct tgtttcgagt agataatgcc agcctgttaa acgccgtcga 600
 [0544] cgagtctaac ggacaccaac cagcgaacca gcagcgtcgc gtcgggccaac gcgaagcaga 660
 [0545] cggcacggca tctctgtcgc tgcctctgga ccctctcga gagttccgct ccaccgttgg 720

[0546]	acttgctccg ctgtcggcat ccagaaattg cgtggcggag cggcagacgt gagccggcac	780
[0547]	ggcaggcggc ctccctctcc tctcacggca cggcagctac gggggattcc tttcccaccg	840
[0548]	ctccttcgct ttcccttctt cgcccggcgt aataaataga caccctctcc acaccctctt	900
[0549]	tccccaacct cgtgttggtc ggagcgcaca cacacacaac cagatctccc ccaaatccac	960
[0550]	ccgtcggcac ctccgcttca aggtacgccg ctctctctcc ccccccccc ctctctacct	1020
[0551]	tctctagatc ggcgttccgg tccatggtta gggcccggta gttctacttc tgttcatggt	1080
[0552]	tgtgttagat ccgtgtttgt gttagatccg tgctgctagc gttcgtacac ggatgcgacc	1140
[0553]	tgtacgtcag acacgttctg attgctaact tgccagtgtt tctctttggg gaatcctggg	1200
[0554]	atggctctag ccgttccgca gacgggatcg atttcatgat tttttttggt tcggttgcata	1260
[0555]	gggtttgggt tgcctttttc ctttatttca atatatgccg tgcacttggt tgctcgggtca	1320
[0556]	tcttttcatg cttttttttg tcttggttgt gatgatgtgg tctggttggg cggctggttct	1380
[0557]	agatcggagt agaattctgt ttcaaacctac ctggtggatt tattaatitt ggatctgtat	1440
[0558]	gtgtgtgcca tacatattca tagttacgaa ttgaagatga tggatggaaa tatcgatcta	1500
[0559]	ggataggtat acatgttgat gcgggtttta ctgatgcata tacagagatg ctttttggtc	1560
[0560]	gcttggttgt gatgatgtgg tgtggttggg cggctgttca ttcgttctag atcggagtag	1620
[0561]	aatactgttt caaacctacct ggtgtattta ttaatitttg aactgtatgt gtgtgtcata	1680
[0562]	catcttcata gttacgagtt taagatggat ggaaatatcg atctaggata ggtatacatg	1740
[0563]	ttgatgtggg ttttactgat gcatatacat gatggcatat gcagcatcta ttcatatgct	1800
[0564]	ctaacttga gtacctatct attataataa acaagtatgt tttataatta ttttgatctt	1860
[0565]	gatatacttg gatgatggca tatgcagcag ctatatgtgg attttttttag ccctgccttc	1920
[0566]	atacgtatt tatttgcttg gtactgtttc tttgtcgtat gctcacctg ttgtttggtg	1980
[0567]	ttacttctgc ag 1992	
[0568]	<210> 19	
[0569]	<211> 1176	
[0570]	<212> DNA	
[0571]	<213> Escherichia coli	
[0572]	<400> 19	
[0573]	atgcaaaaac tcattaactc agtgcaaaaac tatgcctggg gcagcaaaaac ggcgttgact	60
[0574]	gaactttatg gtatggaaaa tccgtccagc cagccgatgg ccgagctgtg gatgggcgca	120
[0575]	catccgaaaa gcagttcacg agtgcagaat gccgccggag atatcgtttc actgcgtgat	180
[0576]	gtgattgaga gtgataaatc gactctgctc ggagaggccg ttgccaaaac ctttggcgaa	240
[0577]	ctgcctttcc tgttcaaagt attatgcgca gcacagccac tctccattca ggttcatcca	300
[0578]	aacaaacaca attctgaaat cggttttgcc aaagaaaatg ccgcaggtat cccgatggat	360
[0579]	gccgccgagc gtaactataa agatcctaac cacaagccgg agctggtttt tgcgctgacg	420
[0580]	cctttccttg cgatgaacgc gtttctgtgaa ttttccgaga ttgtctccct actccagccg	480
[0581]	gtcgcaggtg cacatccggc gattgctcac tttttacaac agcctgatgc cgaacgttta	540
[0582]	agcgaactgt tcgccagcct gttgaatatg cagggtgaag aaaaatcccg cgcgctggcg	600
[0583]	attttaaaat cgccctcga tagccagcag ggtgaaccgt ggcaaacgat tcgtttaatt	660
[0584]	tctgaatitt acccggaaga cagcggctctg ttctccccgc tattgctgaa tgtggtgaaa	720

[0585] ttgaaccctg gcaagcgat gttcctgttc gctgaaacac cgcacgctta cctgcaaggc 780
[0586] gtggcgctgg aagtgatggc aaactccgat aacgtgctgc gtgcgggtct gacgcctaaa 840
[0587] tacattgata ttccggaact ggttgccaat gtgaaattcg aagccaaacc ggctaaccag 900
[0588] ttgttgacc agccggtgaa acaaggtgca gaactggact tcccgattcc agtggatgat 960
[0589] tttgccttct cgctgcatga ccttagtgat aaagaaacca ccattagcca gcagagtgcc 1020
[0590] gccattttgt tctgcgtcga aggcgatgca acgttggtga aaggttctca gcagttacag 1080
[0591] cttaaaccgg gtgaatcagc gtttattgcc gccaacgaat caccggtgac tgtcaaaggc 1140
[0592] cacggccggt tagcgcgtgt ttacaacaag ctgtaa 1176

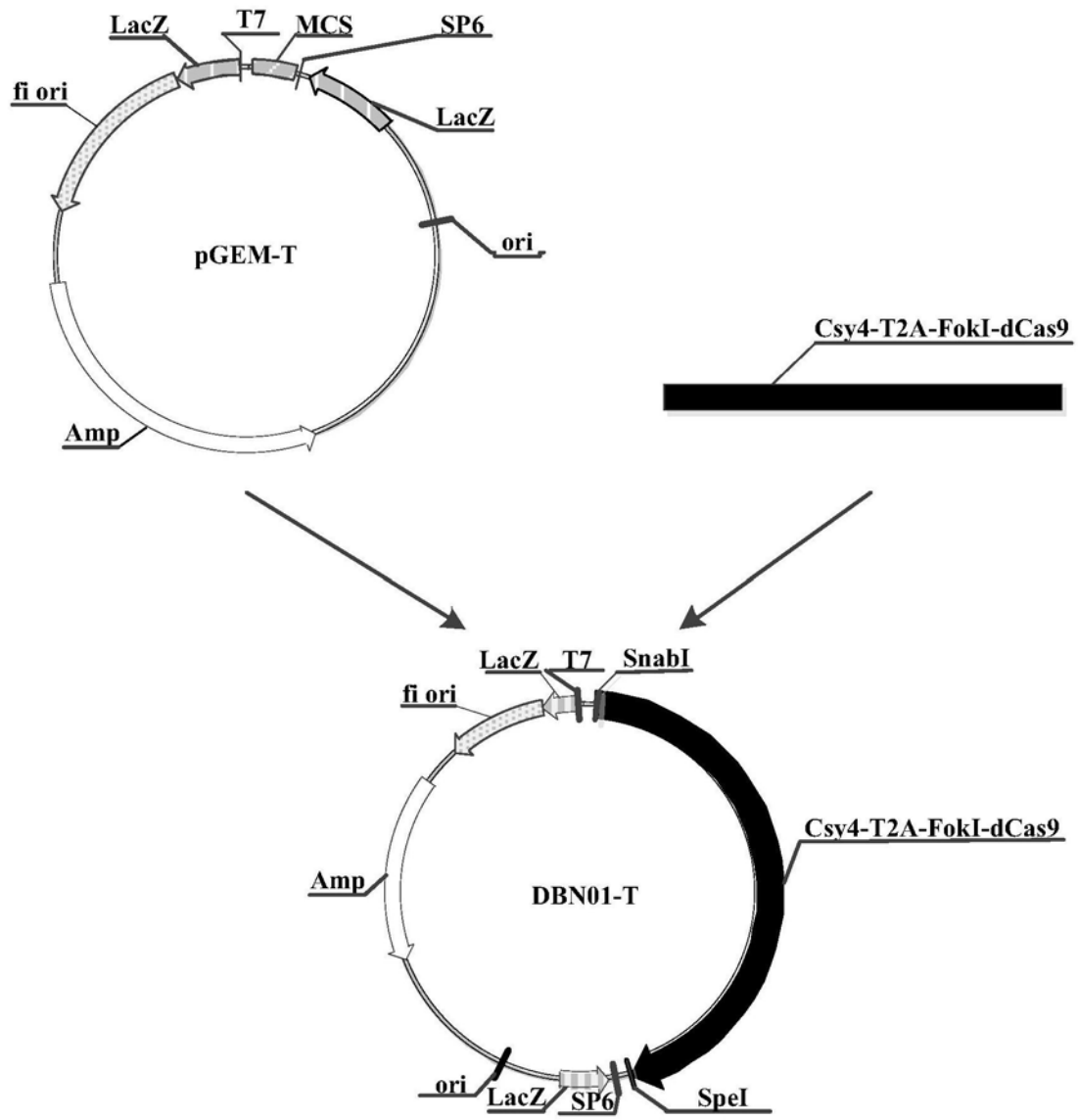


图1

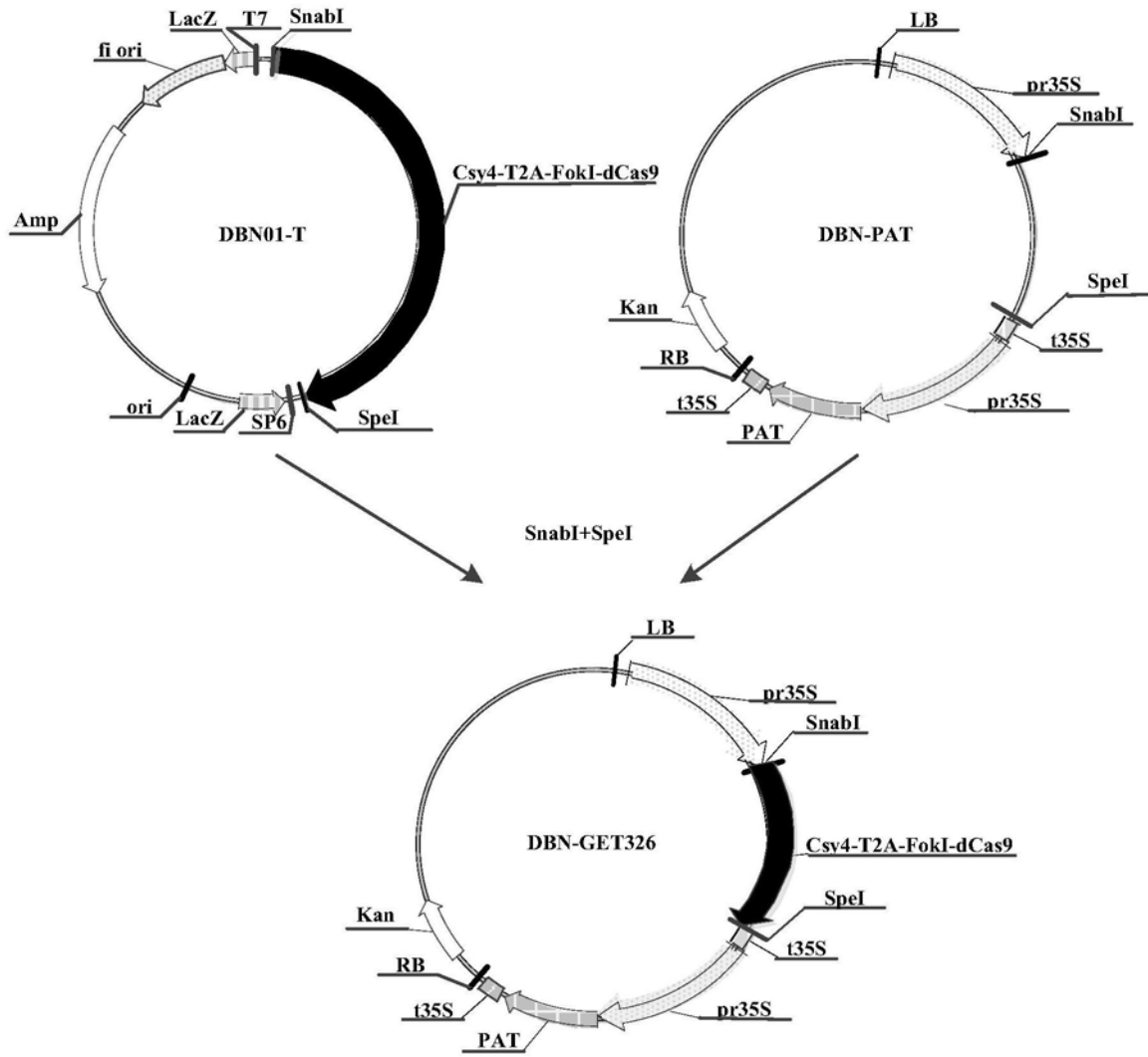


图2

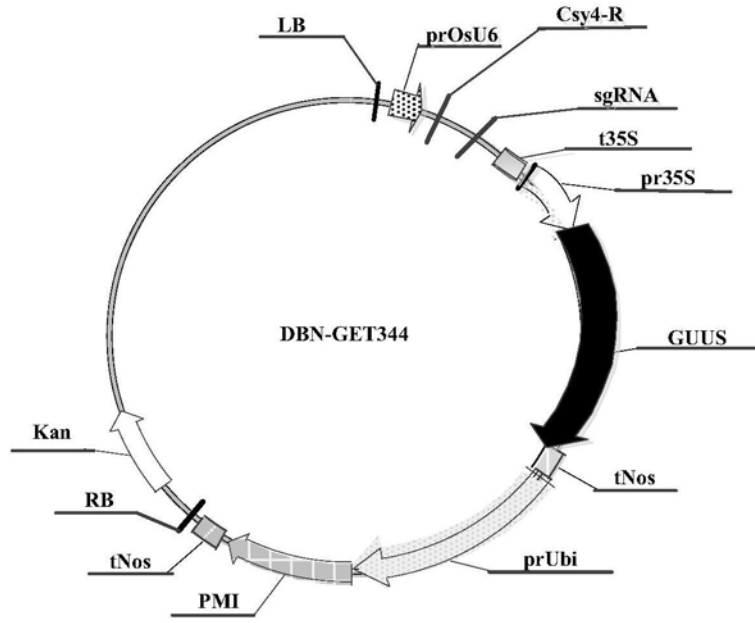


图3

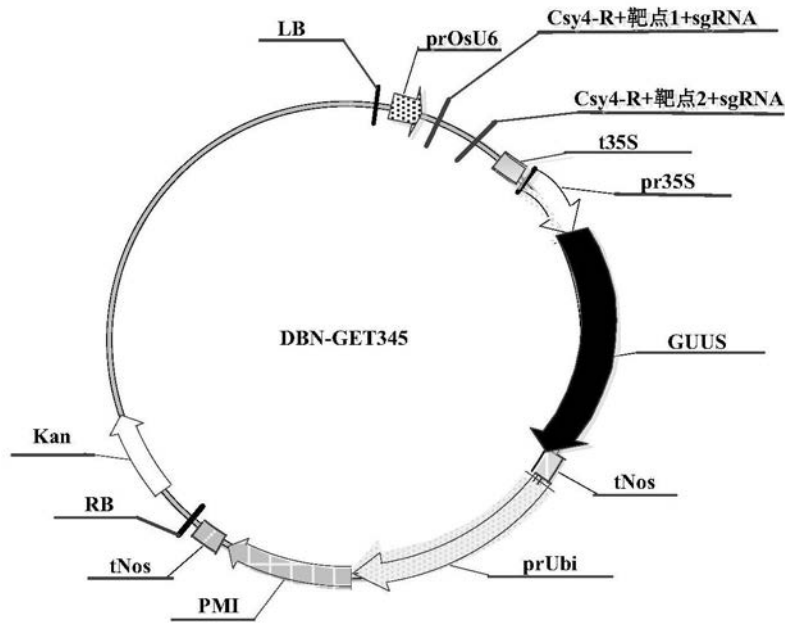
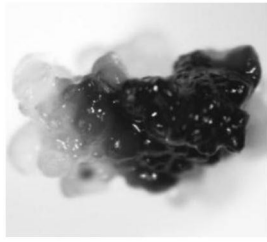
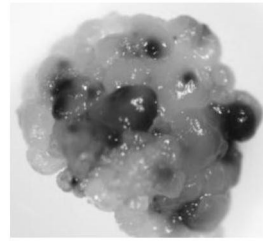


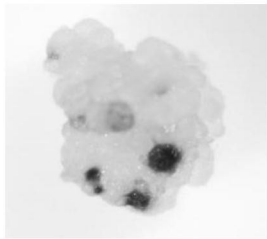
图4



+++



++



+



-

图5