



(12) 发明专利申请

(10) 申请公布号 CN 115605597 A

(43) 申请公布日 2023.01.13

(21) 申请号 202180017085.2

(22) 申请日 2021.03.01

(30) 优先权数据

20160961.7 2020.03.04 EP

(85) PCT国际申请进入国家阶段日

2022.08.26

(86) PCT国际申请的申请数据

PCT/EP2021/054993 2021.03.01

(87) PCT国际申请的公布数据

WO2021/175759 EN 2021.09.10

(71) 申请人 巴斯夫欧洲公司

地址 德国莱茵河畔路德维希港

(72) 发明人 M·F·菲勒 C·索尔 N·韦尔奇

M·阿佩尔鲍姆 T·施韦德尔

(74) 专利代理机构 北京市中咨律师事务所

11247

专利代理师 胡志君 黄革生

(51) Int.Cl.

C12N 15/63 (2006.01)

C12N 15/75 (2006.01)

权利要求书2页 说明书33页

序列表31页 附图7页

(54) 发明名称

用于产生赋予低至中等表达的组成型细菌启动子的方法

(57) 摘要

本发明属于分子生物学领域并且提供在细菌中产生低至中等表达的组成型启动子的方法和用所述方法产生的启动子。

1. 用于产生与细菌细胞中相应的起始调节核酸分子相比, 赋予减少的组成型表达的一种或多种合成性调节核酸分子的方法, 所述方法包括步骤

a. 鉴定至少一种在细菌细胞中赋予组成型表达的起始调节核酸分子, 和
b. 有效连接所述起始调节核酸分子至编码与所述起始调节核酸分子异源的蛋白质的编码区, 和

c. 向载体引入包含与编码区有效连接的所述起始调节核酸分子的构建体, 所述载体包含在细菌细胞中赋予所述载体高拷贝数的复制起点, 其中所述构建体赋予所述编码区的高表达, 其中所述编码区在细菌细胞中的高表达使所述细菌细胞承压, 从而导致生长减少或消除, 和

d. 将所述载体转化入细菌细胞中, 和

e. 培育所述转化的细菌细胞以回收单一克隆, 和

f. 分离显示与不包含所述构建体的相应细菌菌株可比较的生长速率的单一克隆, 和

g. 从所述克隆分离所述构建体; 和

h. 对包含于所述构建体中的合成性调节核酸分子测试与所述合成性调节核酸分子有效连接的基因的功能表达; 和任选地

i. 对包含于所述构建体中的相应调节核酸分子测序, 由此鉴定在细菌细胞中赋予减少的组成型表达的合成性调节核酸分子。

2. 根据权利要求1所述的方法, 其中合成性调节核酸分子在与其中产生重组核酸的细胞不同的细菌细胞中赋予减少的表达。

3. 根据权利要求1或2所述的方法, 其中合成性调节核酸分子在革兰氏阳性细菌和革兰氏阴性细菌的细胞中有活性。

4. 根据权利要求3所述的方法, 其中合成性调节核酸分子在芽孢杆菌纲 (Bacilli) 或 γ -变形菌纲 (Gammaproteobacteria) 的细胞中有活性。

5. 根据权利要求4所述的方法, 其中合成性调节核酸分子在芽孢杆菌科 (Bacillaceae) 和肠杆菌科 (Enterobacteriaceae) 的细胞中有活性。

6. 根据权利要求5所述的方法, 其中合成性调节核酸分子在芽孢杆菌属 (Bacilli) 或埃希氏菌属 (Escherichia) 的细胞中有活性。

7. 根据权利要求6所述的方法, 其中合成性调节核酸分子在芽孢杆菌属的细胞中有活性。

8. 根据权利要求7所述的方法, 其中合成性调节核酸分子在嗜碱芽孢杆菌 (*Bacillus alkalophilus*)、解淀粉芽孢杆菌 (*Bacillus amyloliquefaciens*)、短芽孢杆菌 (*Bacillus brevis*)、环状芽孢杆菌 (*Bacillus circulans*)、克劳氏芽孢杆菌 (*Bacillus clausii*)、凝固芽孢杆菌 (*Bacillus coagulans*)、坚强芽孢杆菌 (*Bacillus firmus*)、灿烂芽孢杆菌 (*Bacillus lautus*)、迟缓芽孢杆菌 (*Bacillus lentus*)、地衣芽孢杆菌 (*Bacillus licheniformis*)、巨大芽孢杆菌 (*Bacillus megaterium*)、短小芽孢杆菌 (*Bacillus pumilus*)、嗜热脂肪芽孢杆菌 (*Bacillus stearothermophilus*)、甲基营养芽孢杆菌 (*Bacillus methylotrophicus*)、蜡样芽孢杆菌 (*Bacillus cereus*)、副地衣芽孢杆菌 (*Bacillus paralicheniformis*)、枯草芽孢杆菌 (*Bacillus subtilis*) 或苏云金芽孢杆菌 (*Baccillus thuringiensis*) 的细胞中有活性。

9. 根据权利要求8所述的方法,其中合成性调节核酸分子在至少三个不同芽孢杆菌物种的细胞中有活性。

10. 根据权利要求9所述的方法,其中合成性调节核酸分子在至少二个不同芽孢杆菌物种的细胞中有活性。

11. 根据权利要求10所述的方法,其中合成性调节核酸分子在至少一个芽孢杆菌物种的细胞中有活性。

12. 根据权利要求7至11所述的方法,其中杆菌物种包含枯草芽孢杆菌、地衣芽孢杆菌或短小芽孢杆菌中的至少一者。

13. 根据权利要求12所述的方法,其中合成性调节核酸分子在地衣芽孢杆菌细胞中有活性。

14. 根据权利要求1至14所述的方法,其中在细菌细胞中赋予组成型表达的起始调节核酸分子选自

a. SEQ ID NO:28和29,

b. 核酸分子,其包含与SEQ ID NO:28或29所述的序列的20个连续碱基对相同的至少20个连续碱基对,和

c. 核酸分子,其在SEQ ID NO:28或29所述的序列的整个长度范围内具有至少90%同一性,和

d. 核酸分子,其在高严格条件下与SEQ ID NO:28或29所述的核酸分子的至少20个连续碱基对的核酸分子杂交,和

e. 如a)至d)中定义的核酸分子中任一者的互补序列。

15. 合成性调节核酸分子,其中调节核酸分子包含于以下者组成的组中

a. 具有SEQ ID NO 35、36、37、38、39、40、42、43、45、46或47的序列的核酸分子,和

b. 核酸分子,其包含与SEQ ID NO:35、36、37、38、39、40、42、43、45、46或47所述的序列的20个连续碱基对相同的至少20个连续碱基对,和

c. 核酸分子,其在SEQ ID NO:35、36、37、38、39、40、42、43、45、46或47所述的序列的整个长度范围内具有至少90%同一性,和

d. 核酸分子,其在高严格条件下与SEQ ID NO:35、36、37、38、39、40、42、43、45、46或47中任一者所述的核酸分子的至少20个连续碱基对的核酸分子杂交,和

e. 如a)至d)中定义的核酸分子中任一者的互补序列,

其中如b)至e)中所定义的序列不同于相应的起始性核酸分子。

16. 根据权利要求15所述的合成性调节核酸分子,其中核酸分子如权利要求1至14中任一项所述地产生。

17. 表达构建体,其包含权利要求15或16所述的合成性调节核酸分子。

18. 载体,其包含权利要求15或16所述的调节核酸分子或权利要求17所述的表达构建体。

19. 微生物,其包含权利要求15或16所述的调节核酸分子或权利要求17所述的表达构建体或权利要求18所述的载体。

用于产生赋予低至中等表达的组成型细菌启动子的方法

[0001] 发明概述

[0002] 本发明属于分子生物学领域并且提供在细菌中产生低至中等表达的组成型启动子的方法和用所述方法产生的启动子。

[0003] 发明介绍

[0004] 工业中如今通过利用微生物的发酵能力广泛应用微生物。特别使用微生物作为发酵生产多种物质如酶、蛋白质、化学物质、糖和聚合物的宿主。出于这些目的,微生物是基因工程的对象,旨在针对特定生产工艺的需求改造其基因表达。微生物的理性基因工程需要靶特异性基因组编辑技术如引入点突变、基因删除、基因插入、基因重复。

[0005] 已经针对若干物种开发了许多不同的基因组编辑方法。它们当中大部分要求引入双链DNA断口或两个毗邻的单链DNA断口,以通过非同源末端接合(NHEJ)在基因组中特定位点处引入随机突变或以使用要求递送供体DNA分子的同源重组修复机制(HR)引入、替换或删除DNA。所用技术例如是Zn指核酸酶、TALEN、归巢核酸内切酶等。基于CRISPR(成簇规律间隔的短回文重复序列)的系统的最新发展使基因组编辑甚至更有吸引力,原因在于其精度、效率和速度。

[0006] 起初鉴定CRISPR系统为链球菌属(*Streptococcus*)细菌的适应性防御机制(WO2007/025097)。那些细菌性CRISPR系统依赖与剪切性蛋白质复合的向导RNA(gRNA)以指引入侵性病毒DNA内部存在的互补序列降解。CRISPR/Cas系统中首个鉴定的蛋白质Cas9是一种大的单体DNA核酸酶,所述核酸酶被两种非编码性RNA(crRNA和反式激活性crRNA(tracrRNA))的复合体导向到毗邻于PAM(原间隔子毗邻基序)序列基序的DNA靶序列。后来,显示通过crRNA与tracrRNA融合所创建的合成性RNA嵌合体(单一向导RNA或sgRNA)同等地有功能(Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. 和 Charpentier, E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity (适应性细菌免疫中的可编程双重RNA引导的DNA核酸内切酶). *Science* 337 (6096), 816-821. 17-8-2012)。

[0007] 数个研究团体已经发现,CRISPR切割特性可以用来前所未有地轻松破坏几乎任何生物的基因组中的基因(Mali P等人(2013) *Science*. 339 (6121): 819-823; Cong L等人(2013) *Science* 339 (6121))。最近,已经清楚,为修复提供模板允许以几乎任何目的序列在几乎任何位点编辑基因组,从而使CRISPR转化成强有力的基因编辑工具(WO/2014/150624、WO/2014/204728)。

[0008] 驱动宿主细胞中基因表达的关键要素是启动子序列。为了基因表达发生,RNA聚合酶必须结合至基因附近的启动子序列。因此,启动子含有为RNA聚合酶及还为召集RNA聚合酶至识别序列的其他蛋白质(即,转录因子)提供结合位点的特定DNA序列。在细菌中,启动子通常由RNA聚合酶和相关 σ 因子识别,后两者因激活物蛋白质与其附近自身DNA结合位点结合被导向启动子DNA(Lee, D. J., Minchin, S. D. 和 Busby, S. J. Activating transcription in bacteria (细菌中的激活性转录). *Annu. Rev. Microbiol.* 66, 125-152. 2012)。例如驱动许多持家基因表达的组成型启动子与激活蛋白或阻遏蛋白所致的激活或去阻遏无关并且

RNA聚合酶通过识别sigA特异性DNA序列元件-35框和-10框的相关 σ 因子sigA(大肠杆菌中也称作sig70)与组成型启动子结合。已经充分研究了芽孢杆菌(Bacillus)和大肠杆菌的sigA依赖性启动子并且比较sigA启动子序列的共有基序提示,芽孢杆菌衍生和大肠杆菌衍生的sigA启动子分别被大肠杆菌RNA聚合酶和芽孢杆菌RNA聚合酶以及相应sig70因子和sigA因子交叉识别(Helmann,J.D.Compilation and analysis of Bacillus subtilis sigma A-dependent promoter sequences:evidence for extended contact between RNA polymerase and upstream promoter DNA(枯草芽孢杆菌(Bacillus subtilis) σ A-依赖性启动子序列编纂和分析:RNA聚合酶和上游启动子DNA之间扩展接触的证据).Nucleic Acids Res.23(13),2351-2360.11-7-1995)。

[0009] 在真核生物中,该过程更复杂,并且多种因子为RNA聚合酶与启动子结合所需要。受核酸序列影响,启动子可以赋予低、中等或高表达水平并且可以为组成型或诱导型。

[0010] 已经描述了芽孢杆菌的许多组成型启动子。veg基因的启动子Pveg是一个充分描述的组成型强启动子。另外,已经构建了包含启动子强度不同的芽孢杆菌组成型启动子的表达模块文库(Guiziuo,S.,等人(2016).Nucleic Acids Res.44(15),7495-7508)。

[0011] 通过添加诱导物分子至细胞,则激活诱导型启动子或使其去阻遏。因而,激活蛋白与紧邻启动子序列的序列结合并且活跃地召集RNA聚合酶和相关 σ 因子以允许转录启动。熟知的已述实例是受添加阿拉伯糖时改变启动子构象并且作为二聚体结合至操纵基因位点I₁和I₂的araC调节的大肠杆菌P_{BAD}启动子和受激活物manR调节的芽孢杆菌甘露糖诱导型启动子系统PmanP。诱导型启动子如lacUV5启动子、用于大肠杆菌中表达的T7-噬菌体启动子和芽孢杆菌Pspac-I和Ppac-I启动子受lac阻遏物(由lacI基因编码)负向调节,所述阻遏物在诱导物分子不存在情况下与启动子序列内部(例如-35sigA识别位点和-10sigA识别位点之间)或附近(即启动子序列的3'或5')的其特异性lac操纵基因位点结合以阻止转录。另一个实例是广泛用于芽孢杆菌表达系统的来自巨大芽孢杆菌(Bacillus megaterium)的Pxy1A诱导型启动子系统。Pxy1A启动子受包含转录起始位点3'处xy1R操纵基因位点的xy1R阻遏蛋白负向调节。

[0012] 诱导型启动子系统通常有利于表达载体中的克隆过程,因为这类启动子控制下的基因表达大幅度减少并且因此,例如关于剥夺细胞资源、干扰细胞代谢等的不利影响最小化,然而,需要针对其中使用该启动子的每个菌株就所添加诱导物分子的数量和诱导表达的时间点仔细分析对所需蛋白质表达的调适。相反,组成型启动子具有不依赖施加诱导物的优点,不需要特定调节物或转运蛋白,因而在广泛类型的细菌中有活性。

[0013] 质粒是在宿主细胞中自主复制,因而与宿主染色体复制无关的染色体外环状DNA。

[0014] 为了自主复制,质粒包含复制起点,所述复制起点使得载体在讨论的宿主细胞中自主复制成为可能。细菌复制起点的实例是允许在芽孢杆菌中复制的质粒pUB110、pE194、pC194、pTB19、pAM β 1、pTA1060的复制起点和允许在大肠杆菌中复制的质粒pBR322、colE1、pUC19、pSC101、pACYC177和pACYC184的复制起点(Sambrook,J.和Russell,D.W.《分子克隆实验室指南》,第3版,Cold Spring Harbor Laboratory Press,Cold Spring Harbor, NY.2001)。

[0015] 将质粒拷贝数定义为正常生长条件下每个细菌细胞或每条染色体的质粒平均数目。另外,存在不同类型在细菌宿主中导致不同拷贝数的复制起点(也称作复制子)。

[0016] 质粒复制子pBS72和质粒pTB19和衍生物pTB51、pTB52赋予低拷贝数,在芽孢杆菌细胞内部分别为6个拷贝和1至8个拷贝,而质粒pE194和pUB110赋予低-中等拷贝数,每个细胞分别14-20个拷贝,及中等拷贝数,每个细胞分别30-50个拷贝。更详细地分析了质粒pE194 (Villafane, 等人 (1987) : J. Bacteriol. 169 (10), 4822-4829) 并且描述了在芽孢杆菌内部具有范围从85个拷贝至202个拷贝的高拷贝数的数个pE194-cop突变体。另外,质粒pE194为温度敏感性,高达37°C时拷贝数稳定,然而在43°C以上复制消除。另外,它存在一种称作pE194ts的pE194变体,在复制子区域内部带2个点突变,导致更夸张的温度敏感性——高达32°C时拷贝数稳定,然而37°C时仅每个细胞1至2个拷贝。

[0017] 在大肠杆菌中,携带pMB1复制子或其近亲——大肠菌素E1 (colE1) 复制子的pBR322质粒维持低-中等拷贝数,即每个细菌细胞中15-20个拷贝。在colE1和pMB1质粒衍生物内部删除rop/rom基因略微地增加大肠杆菌细胞内部质粒拷贝数至中等拷贝数25-50个。pUC载体系列是每个大肠杆菌细胞多达200个拷贝的小型高拷贝质粒,衍生自缺少rop蛋白的突变型pBR322质粒。pUC质粒是充分建立的克隆载体,原因在于相比上文提到的pBR322衍生和ColE1衍生载体,其尺寸小和质粒制备时产率高。

[0018] 备选地,pACYC177/184质粒中存在的p15A复制子赋予低-中等拷贝数,每个细胞20个拷贝,并且pSC101复制子赋予低拷贝数,每个细胞5-10个拷贝。具有低至中等拷贝数且编码有毒或不利表达构建体的质粒通常稳定维持于细胞内部,然而,质粒制备时产率低。对于细菌细胞的后续转化,相比高拷贝质粒的质粒制备,质粒DNA的量变成限制性。当平行进行多个制备时,这尤其对中等至高通量应用有意义。

[0019] 质粒拷贝数和用于基因表达的启动子选择的组合决定蛋白质总体表达水平并且因此影响细胞的生存力和质粒稳定性。

[0020] 已经成功地应用了应用于革兰氏阳性生物(如芽孢杆菌属物种)的基于CRISPR的表达系统,所述表达系统基于单一质粒系统方法,即包含Cas9核酸内切酶、gRNA(例如sgRNA或crRNA/tracrRNA)、在一个单一大肠杆菌-芽孢杆菌穿梭载体上的修复同源性序列(供体DNA)。

[0021] Altenbuchner创建了一系列高拷贝pUC复制子,所述复制子基于CRISPR/Cas9基因组编辑用于枯草芽孢杆菌(*B. subtilis*)的大肠杆菌-芽孢杆菌穿梭质粒,与诱导型启动子PmanP、PxylA和PtetLM组合以表达Cas9核酸内切酶(Altenbuchner, (2016) : Applied and environmental microbiology 82 (17), 5421-5427)。这允许大肠杆菌克隆宿主内部高度有效的质粒DNA制备和稳定维持。同样地,产生了一种构建高拷贝pUC衍生型CRISPRi-大肠杆菌-芽孢杆菌穿梭质粒以应用于甲醇芽孢杆菌(*Bacillus methanolicus*)中的相似方法。驱动缺陷性Cas9表达的甲醇芽孢杆菌甘露醇激活基因mt1R的启动子通过在启动子3'处引入lacO位点受到修饰,因此在大肠杆菌中用完整lacI有效阻断转录活性(Schultenkämper等人(2019) : Applied microbiology and biotechnology 103 (14), 5879-5889)。

[0022] CRISPR/Cas9应用于枯草芽孢杆菌中的另一个单一质粒方案使用低至中等拷贝数复制子p15A以与使用表达Cas9的诱导物非依赖性启动子(解淀粉芽孢杆菌(*B. amyloliquefaciens*) PamyQ-淀粉酶启动子)组合,允许成功克隆并稳定维持大肠杆菌中基于CRISPR/Cas9的基因组编辑用大肠杆菌-芽孢杆菌穿梭质粒。应用了中等拷贝pBR322衍生型大肠杆菌-芽孢杆菌穿梭载体与组成型强启动子控制下的Cas9的相似组合(Zhou等人

(2019):International journal of biological macromolecules 122,329-337)。

[0023] 尽管低拷贝和中等拷贝主链减少代谢负担,但这伴随来自大肠杆菌的质粒产率降低并且阻碍按照难以转化芽孢杆菌菌株或难以在高通量应用中应用的许多转化方案所要求的规模分离质粒DNA。诱导物依赖性启动子系统并不总是适用于广泛类型的不同微生物中并且此外需要分析诱导物-分子的数量和诱导启动子的时间点。另外,相比组成型启动子,需要一个借助添加诱导物分子至细胞的额外启动子激活步骤,这拉长基因组编辑方法的总体时间范围。

[0024] 因此本领域需要提供允许使用与使用组成型启动子组合的高拷贝载体来克服这些限制的系统。这类系统的一个要素是提供在细菌中赋予减少表达的组成型启动子,所述的减少表达不干扰或仅不显著干扰细菌的生长和/或生长势。

[0025] 发明详述

[0026] 本发明的第一实施方案包括一种用于产生一种或多种与细菌细胞中相应的起始调节核酸分子相比,赋予减少的组成型表达的合成性调节核酸分子的方法,所述方法包括步骤

[0027] a. 鉴定至少一种在细菌细胞中赋予组成型表达的起始调节核酸分子,并且

[0028] b. 有效连接所述起始调节核酸分子至编码相对于所述起始调节核酸分子为异源的蛋白质的编码区,并且

[0029] c. 向载体引入包含与编码区有效连接的所述起始调节核酸分子的构建体,所述载体包含在细菌细胞内部赋予所述载体高拷贝数的复制起点,其中所述构建体赋予所述编码区的高表达,其中所述编码区在细菌细胞中的高表达使所述细菌细胞承压,从而导致生长减少或消除,并且

[0030] d. 将所述载体转化入细菌细胞中,并且

[0031] e. 培育所述转化的细菌细胞以回收单一克隆,并且

[0032] f. 分离单一克隆显示生长速率与可分离显示与不包含所述构建体的相应WT菌株可比较的生长速率的单一克隆,并且

[0033] g. 从所述克隆分离所述构建体;并且

[0034] h. 对包含于所述构建体中的合成性调节核酸分子测试与所述合成性调节核酸分子有效连接的编码区的功能表达并且任选地

[0035] i. 将合成性调节核酸赋予的表达与起始性调节核酸赋予的表达比较并且任选地

[0036] j. 对包含于所述构建体中的相应调节核酸分子测序,因而鉴定在细菌细胞中赋予减少的组成型表达的合成性调节核酸分子。

[0037] 减少的生长意指在平板上对于相应细菌适宜的条件下孵育某个时间段后,在包含如上文所述的构建体的细菌菌落和不包含所述构建体的细菌菌落之间可见相应菌落的可见尺寸差异。如相比不包含构建体的细菌菌落,包含所述构建体的细菌菌落将显示更小的菌落。例如,将在36-37°C孵育大肠杆菌细菌8-16小时,之后比较菌落尺寸的差异。

[0038] 在组成型强启动子控制下使表达编码区的细菌承压的所述编码区可以例如是编码大于150kDa的蛋白质(例如Cas9或Cas12a)的任何编码区、编码诱导DNA链断裂或突变的酶(例如Cas9、Cas12a和任何其他CRISPR Cas酶、归巢核酸内切酶、大范围核酸酶、腺苷脱氨酶或DNA糖基化酶)的编码区、编码干扰细菌代谢的酶(例如涉及产生能量等同物(ATP)或辅

因子如NADP的酶)的编码区或编码干扰底物摄取或细菌细胞脱毒的转运蛋白或跨膜蛋白的编码区。

[0039] 细菌细胞中的组成型表达意指衍生自相应启动子的表达强度在多种条件下基本上恒定。在本说明书中,组成型表达意指衍生自一个启动子的表达在以下条件下差异小于10倍、优选地小于9倍、优选地小于8倍、优选地小于7倍、优选地小于6倍、优选地小于5倍、优选地小于4倍,更优选地小于3倍、甚至更优选地小于2倍:对相应细胞为最佳的温度条件下于丰富培养基中,例如LB培养基中、替代有糖、例如蔗糖、乳糖或葡萄糖、优选地浓度0.1%至0.5%之间、优选地0.3%的葡萄糖的丰富培养基中以及极限盐培养基中,例如补充有糖、例如蔗糖、乳糖或葡萄糖、优选地浓度0.1%至0.5%之间、优选地0.3%的葡萄糖的M9培养基中指数生长阶段、过渡阶段和静止阶段。

[0040] 为了确定基因是否差异性表达,将基因表达跨这些条件至少一式三份测量并且使用本领域标准方法DESeq2软件包(Love,M.I.等人,Genome Biology 15(12):550(2014))测量这些值的差异。这种分析将评估诸条件之间的观测倍数变化以及这种差异应归因于随机机率的概率。较上文定义者更上调或和/下调并且具有因随机机率所致5%以下概率的任何基因视为差异性表达,因此,未组成型表达。

[0041] 组成型启动子与其他细胞调节因子无关并且转录启动取决于 σ 因子A(sigA)。sigA依赖性启动子包含 σ 因子A特异性识别位点'-35'-区和'-10'-区。

[0042] 优选地,组成型启动子序列选自以下者组成的组:基因表达强度不同的启动子Pveg、PlepA、PserA、PymdA、PfbA及其衍生物(Guiziou等人,(2016):Nucleic Acids Res.44(15),7495-7508)、噬菌体SP01启动子P4、P5、P15(W015118126)、来自苏云金芽孢杆菌(*Bacillus thuringiensis*)的cryIIIA启动子(W09425612)及其组合,或其活性片段或变体。

[0043] 赋予高拷贝数的复制起点(ORI)意指一种ORI,其在其中该ORI有功能的相应细菌细胞内导致相应载体的至少51个拷贝。由于拷贝数目取决于培育相应细菌的温度,故优选地该定义指在技术人员已知如对例如各种菌株所述的实验室中培育相应细菌的温度(Bronikowski等人(2001):Evolution 55(1):33-40)。

[0044] 优选地对于大肠杆菌,这意味着在36-37°C生长下检测拷贝数,对于芽孢杆菌,这意味着在36-37°C生长下检测拷贝数。

[0045] 一个ORI赋予中等拷贝数意指维持25-50个载体拷贝的ORI,一个赋予低-中等拷贝数的ORI意指维持每个细胞11-24个拷贝的ORI并且一个赋予低拷贝数的ORI意指在细菌细胞内部维持1-10个载体拷贝的ORI。

[0046] 在一个优选的实施方法中,大肠杆菌ORI选自高拷贝数ORI并且芽孢杆菌ORI选自低拷贝数ORI、低-中等拷贝数ORI和中等拷贝数ORI。

[0047] 更优选地,大肠杆菌ORI选自高拷贝数ORI并且芽孢杆菌ORI选自低-中等拷贝数ORI。

[0048] 更优选地,大肠杆菌ORI选自高拷贝数ORI并且芽孢杆菌ORI选自温度敏感性低-中等拷贝数ORI例如在36-37°C赋予低-中等拷贝数并在30-33°C赋予低-中等拷贝数及高于43°C不复制的质粒pE194衍生物。

[0049] 更优选地,大肠杆菌ORI选自高拷贝数ORI,例如例如pUC ORI,并且芽孢杆菌ORI选

自温度敏感性低-中等拷贝数ORI例如在36-37°C赋予低拷贝数并在30-33°C赋予低-中等拷贝数及高于38°C不复制的质粒pE194ts衍生物。

[0050] 术语“显示与不包含所述构建体的相应WT菌株可比的生长速率的克隆”意指用如上文定义的构建体转化的克隆,其中与不包含此种构建体或未用其转化的细菌相比时,所述克隆显示生长速率为WT细菌至少50%的生长速率。优选地它们具有如WT细菌至少60%、65%、70%、75%、80%、85%的生长速率。更优选地,具有如WT细菌至少90%、95%的生长速率如或具有与WT细菌相同的生长速率。可以例如依据在液体培养物中某个孵育时间后的细胞密度或依据平板上的菌落大小测定生长速率。

[0051] 编码区的功能性表达意指这种编码区的表达至少例如通过RNA检测方法如RT-PCR、qPCR或通过使用可检出蛋白质如荧光蛋白、GUS、对相应酶特异的酶反应或编码在基因组中诱导双链断裂的酶(如CRISPR/Cas酶)的编码区的基因删除效率是可检出的。

[0052] 本发明的又一个实施方案是如上文定义的方法,其中合成性调节核酸分子在与其中产生重组核酸的细胞区别的细菌中赋予低至中高表达。例如,将起始调节核酸分子在大肠杆菌中测试并突变并且稍后在芽孢杆菌属物种(*Bacillus* sp.)中用于低至中等组成型表达。为此目的,可以将上文定义的方法中所用的构建体克隆入穿梭载体,所述穿梭载体包含用于大肠杆菌的高拷贝数ORI和用于选择芽孢杆菌属物种的另一个ORI。

[0053] 在本发明的又一个实施方案中,合成性调节核酸分子在革兰氏阳性或革兰氏阴性细菌的细胞中、优选地芽孢杆菌纲(*Bacilli*)或 γ -变形菌纲(*Gammaproteobacteria*)的细胞中;更优选地在芽孢杆菌科(*Bacillaceae*)或肠杆菌科(*Enterobacteriaceae*)的细胞中;甚至更优选地在芽孢杆菌属(*Bacilli*)或埃希氏菌属(*Escherichia*);并且甚至更优选地在芽孢杆菌属的细胞中有活性。

[0054] 优选的芽孢杆菌属细胞包含嗜碱芽孢杆菌(*Bacillus alkalophilus*)、解淀粉芽孢杆菌、短芽孢杆菌(*Bacillus brevis*)、环状芽孢杆菌(*Bacillus circulans*)、克劳氏芽孢杆菌(*Bacillus clausii*)、凝固芽孢杆菌(*Bacillus coagulans*)、坚强芽孢杆菌(*Bacillus firmus*)、灿烂芽孢杆菌(*Bacillus lautus*)、迟缓芽孢杆菌(*Bacillus lentus*)、地衣芽孢杆菌、巨大芽孢杆菌(*Bacillus megaterium*)、短小芽孢杆菌(*Bacillus pumilus*)、嗜热脂肪芽孢杆菌(*Bacillus stearothermophilus*)、甲基营养芽孢杆菌(*Bacillus methylotrophicus*)、蜡样芽孢杆菌(*Bacillus cereus*)、副地衣芽孢杆菌(*Bacillus paralicheniformis*)、枯草芽孢杆菌和苏云金芽孢杆菌(*Baccillus thuringiensis*)细胞。

[0055] 优选地,合成性调节核酸分子在至少三个不同的芽孢杆菌属物种的细胞中、至少两个不同的芽孢杆菌属物种的细胞中或至少一个芽孢杆菌属物种的细胞中有活性。

[0056] 更优选地,芽孢杆菌属物种包括以下至少一者:枯草芽孢杆菌、地衣芽孢杆菌(*Bacillus licheniformis*)或短小芽孢杆菌(*Bacillus pumilus*)。最优选地,合成性调节核酸分子在地衣芽孢杆菌细胞中有活性。

[0057] 在本发明的又一个实施方案中,可以使用任何在细菌中有活性的赋予高表达的组成型调节核酸分子。Guiziou等人(Guiziou等人,(2016):*Nucleic Acids Res.*44(15), 7495-7508)描述了适于本发明方法的各种调节核酸分子并且进一步介绍如何鉴定适合本发明方法的额外调节核酸分子的方法。优选地,在细菌细胞中赋予组成型表达的起始调节

核酸分子选自

[0058] a) SEQ ID NO:28和29,

[0059] b) 核酸分子,其包含与SEQ ID NO:28或29所述的序列的20、优选地25、更优选地50、更优选地75、更优选地100、甚至更优选地110、甚至更优选地120个连续碱基对相同的至少20、优选地25、更优选地50、更优选地75、更优选地100、甚至更优选地110、甚至更优选地120个连续碱基对,和

[0060] c) 核酸分子,其在SEQ ID NO:28或29所述的序列的整个长度范围内具有至少90%同一性、优选地至少91%、92%、93%、94%或95%、更优选地至少96%、97%、98%或99%同一性,和

[0061] d) 核酸分子,其在高严格条件下与SEQ ID NO:28或29所述的核酸分子的至少20个连续碱基对、优选地25、更优选地50、更优选地75、更优选地100、甚至更优选地110、甚至更优选地120连续碱基对的核酸分子杂交,

[0062] e) 如a)至d)中定义的核酸分子中任一者的互补序列。

[0063] 本发明的又一个实施方案是一种合成性调节核酸分子,其中调节核酸分子包含于以下者组成的组中

[0064] A) 具有SEQ ID NO 35、36、37、38、39、40、42、43、45、46或47的序列的核酸分子,和

[0065] B) 核酸分子,其包含与SEQ ID NO:35、36、37、38、39、40、42、43、45、46或47所述的序列的20、优选地25、更优选地50、更优选地75、更优选地100、甚至更优选地110、甚至更优选地120个连续碱基对相同的至少20、优选地25、更优选地50、更优选地75、更优选地100、甚至更优选地110、甚至更优选地120个连续碱基对,和

[0066] C) 核酸分子,其在整个长度范围内与依据SEQ ID NO:35、36、37、38、39、40、42、43、45、46或47所述的序列具有至少90%同一性、优选地至少91%、92%、93%、94%或95%、更优选地至少96%、97%、98%或99%同一性,和

[0067] D) 核酸分子,其在高严格条件下与SEQ ID NO:35、36、37、38、39、40、42、43、45、46或47中任一者所述的核酸分子的至少20、优选地25、更优选地50、更优选地75、更优选地100、甚至更优选地110、甚至更优选地120个连续碱基对的核酸分子杂交,和

[0068] E) 如A)至D)中定义的核酸分子中任一者的互补序列,

[0069] 其中如B)至E)中所定义的序列不同于具有SEQ ID NO 28或29的相应起始调节核酸分子,并且与相应起始性调节核酸相比,优选地包含至少一个碱基删除或插入。

[0070] 本发明的又一个实施方案是如上文所述的合成性调节核酸分子,其中应用如上文定义的方法产生该核酸分子。

[0071] 包含如上文定义合成性调节核酸分子的表达构建体也是本发明的一个实施方案。优选地,所述表达构建体包含合成性调节核酸分子并且功能性连接于编码CRISPR/Cas蛋白、大范围核酸酶蛋白或TALE/N的编码区、优选地作为Cas9或Cas12a蛋白的CRISPR/Cas蛋白。

[0072] 包含如上文定义的合成性调节核酸分子或上文定义的表达构建体的载体是本发明的又一个实施方案。

[0073] 本发明的又一个实施方案是包含如上文定义的调节核酸分子或表达构建体或载体的微生物。

[0074] 定义

[0075] 缩略语:GFP-绿色荧光蛋白,GUS- β -半乳糖苷酶,BAP-6-苄氨基嘌呤,2,4-D-2,4-二氯苯氧乙酸,MS-Murashige-Skoog培养基,NAA-1-萘乙酸,MES,2-(N-吗啉代)-乙磺酸,IAA:吲哚乙酸,Kan:硫酸卡那霉素,GA3-赤霉素,TimentinTM:替卡西林二钠/克拉维酸钾,microl:微升。

[0076] 应当理解,本发明不限于具体的方法或方案。还应当理解本文所用的术语仅仅出于描述具体实施方案的目的,并且不意图限制本发明的范围,本发明将仅受所附权利要求限制。必须指出,如本文中和所附权利要求中所用,单数形式“一个(a)”、“一种(an)”和“该(the)”包括复数指称,除非上下文另外明确地指明并非如此。因此,例如,对“一种载体”的称谓是对一种或多种载体的称谓并且包括本领域技术人员已知的其等同物等。术语“约”在本文中用来指大约、大致、左右和在……范围内。当术语“约”与一个数字范围联合使用时,它通过扩展界限值高于和低于所述数值而修饰该范围。通常而言,术语“约”在本文中用来通过20%、优选地10%之上或之下(更高或更低)变异而修饰高于和低于所述值的数值。如本文中所用,词汇“或”意指特定列出的任何一成员并且还包含该列出的成员的任意组合。在本说明书中及以下权利要求中使用,词汇“包含”、“包含着”、“包括”、“包括着”和“包括了”意图指明一个或多个所述特征、整数、组分或步骤的存在,但是它们不排除一个或多个其他特征、整数、组分、步骤或其组的存在或添加。为清晰起见,本说明书中使用的某些术语如下定义并使用。

[0077] 编码区:如本文中所用,术语“编码区”在谈及结构基因使用时,指编码在作为mRNA分子翻译结果的新生多肽中存在的氨基酸的核苷酸序列。编码区在5'侧以编码起始物甲硫氨酸的核苷酸三联体“ATG”为界并且在3'侧以指定终止密码子的三种三联体(即TAA、TAG、TGA)为界。备选地,核苷酸三联体可以是“GTG”或“TTG”并且视为起始核苷酸三联体,因为相对于所述核苷酸三联体的5',核糖体结合位点(Shine Dalgarno)距离4个核苷酸至12个核苷酸存在。基因的基因组形式也可以包含位于RNA转录物中存在的序列的5'-端及3'-端上的序列。这些序列称作“侧翼”序列或区(这些侧翼序列位于mRNA转录物上存在的非翻译序列的5'或3')。5'-侧翼区可以含有控制或影响基因转录的调节序列如启动子和增强子及控制或影响mRNA翻译的核糖体结合位点(Shine Dalgarno)。3'-侧翼区可以含有指导转录终止和转录后剪切的序列。

[0078] 互补的:“互补的”或“互补性”指包含反平行核苷酸序列的两个核苷酸序列,其中一旦在反平行核苷酸序列中的互补性碱基残基之间形成氢键,则所述反平行核苷酸序列能够相互配对(通过碱基配对原则)。例如,序列5'-AGT-3'与序列5'-ACT-3'互补。互补性可以是“部分的”或“全部的”。“部分”互补性是其中一个或多个核酸碱基根据碱基配对规则未匹配的情况。核酸分子之间的“全部”或“完全”互补性是其中每个核酸碱基按照碱基配对规则与另一个碱基匹配的情况。核酸分子链之间互补性的程度对核酸分子链之间杂交的效率和强度具有明显影响。如本文中所用的核酸序列“互补物”指这样的核苷酸序列,其核酸分子显示与该核酸序列的核酸分子的全部互补性。

[0079] 供体DNA分子:如本文所用,本文中均互换使用的术语“供体DNA分子”、“修复DNA分子”或“模板DNA分子”意指具有待引入细胞基因组的序列的DNA分子。它可以在5'和/或3'末端旁侧分布有与所述细胞的基因组的靶区域中序列同源或相同的序列。它可以包含相应细

胞中不天然存在的序列如应当向靶区域引入的ORF、非编码性RNA或调节元件或它可以包含除至少一个突变(基因编辑)外与靶区域同源的序列:可以将供体DNA分子的序列添加至基因组或它可以替代基因组中供体DNA序列长度的序列。

[0080] 双链RNA:“双链RNA分子”或“dsRNA”分子包含核苷酸序列的有义RNA片段和该核苷酸序列的反义RNA片段,二者均包含彼此互补的核苷酸序列,因而允许有义RNA片段和反义RNA片段配对并形成双链RNA分子。

[0081] 内源的:“内源的”核苷酸序列指存在于未转化的细胞的基因组中的核苷酸序列。

[0082] 表达:“表达”指基因产物的生物合成,优选地指细胞中核苷酸序列例如内源基因或异源基因的转录和/或翻译。例如,在结构基因的情况下,表达涉及结构基因转录成mRNA并且任选地随后mRNA翻译成一种或多种多肽。在其他情况下,表达可以仅指携带RNA分子的DNA的转录。

[0083] 表达构建体:如本文中所述的“表达构建体”意指能够指导特定核苷酸序列在适宜植物部分或植物细胞中表达的DNA序列,该DNA序列包含在将引入此DNA序列的所述植物部分或植物细胞中有功能的启动子,所述启动子与任选地有效连接至终止信号的目的核苷酸序列有效连接。如果需要翻译,该DNA序列一般还包含为正确翻译所述核苷酸序列所需的序列。编码区可以编码目的蛋白,但也可以按有义或反义方向编码功能性目的RNA,例如RNAa、siRNA、snoRNA、snRNA、microRNA、ta-siRNA或任何其他非编码的调节性RNA。包含目的核苷酸序列的表达构建体可以是嵌合的,这意指该表达构建体的组件中一者或多者相对于该表达构建体的其他组件中一者或多者是异源的。该表达构建体也可以是一种这样的表达构建体,它天然地存在,但已经以用于异源表达的重组形式获得。然而,一般而言,该表达构建体相对于宿主为异源,即,该表达构建体的特定DNA序列不天然存在于宿主细胞中并且必须已经通过转化事件引入宿主细胞或该宿主细胞的祖先中。该表达构建体盒中的核苷酸序列的表达可以处于组成型启动子或处于仅在宿主细胞暴露于一些特定外部刺激时才启动转录的诱导型启动子的控制下。就细胞发育而言,启动子也可以针对特定发育阶段(例如生物被膜形成、孢子形成)为特异。

[0084] 外来:术语“外来”指这样的任何核酸分子(例如,基因序列),所述核酸分子通过实验操作引入细胞的基因组中并且可以包含该细胞中存在的序列,只要引入的序列含有一些修饰(例如,点突变、存在可选择标记基因等)并且因此相对于天然存在序列而不同。

[0085] 功能性连接:将术语“功能性连接”或“功能性连接的”理解为意指例如调节性元件(例如启动子)与待表达的核酸序列并且根据需要进行与其他调节性元件以如此方式依次排列,从而每种调节性元件可以履行其目的功能以允许、修饰、促进或影响所述核酸序列的表达。作为同义词,可以使用“有效连接”或“有效连接的”。取决于核酸序列的排列,表达可以产生有义或反义RNA。为此目的,不必需要化学意义上的直接连接。遗传控制序列例如增强子序列也能够从远离的位置或甚至从其它DNA分子对靶序列产生其作用。优选的排列是这样的排列,其中待表达的核酸序列重组地位于充当启动子的序列之后,从而这两个序列相互共价地连接。该启动子序列与待重组表达的核酸序列之间的距离优选地小于200碱基对、特别地优选小于100碱基对、非常特别优选地小于50碱基对。在优选的实施方案中,待转录的核酸序列以如此方式位于启动子之后,其中转录起点与本发明嵌合RNA的合乎需要的开端相同。可以借助如(例如,在Maniatis T,Fritsch EF和Sambrook J(1989)Molecular

Cloning: A Laboratory Manual, 第2版, Cold Spring Harbor Laboratory, Cold Spring Harbor (NY); Silhavy等人(1984) Experiments with Gene Fusions, Cold Spring Harbor Laboratory, Cold Spring Harbor (NY); Ausubel等人(1987) Current Protocols in Molecular Biology, Greene Publishing Assoc. and Wiley Interscience; Gelvin等人(编著)(1990) Plant Molecular Biology Manual; Kluwer Academic Publisher, Dordrecht, The Netherlands)所述的惯用重组与克隆技术产生功能性连接和表达构建体。然而,其他序列,例如充当带限制性酶特定切割位点的接头或充当信号肽的序列,也可以位于这两个序列之间。序列的插入也可以导致融合蛋白表达。优选地,由调节性区域例如启动子和待表达核酸序列的连接组成的表达构建体可以按载体整合的形式存在并且被插入植物基因组,例如通过转化法做到。

[0086] 基因:术语“基因”指与能够以某种方式调节基因产物(例如,多肽或功能性RNA)表达的适宜调节序列有效连接的区域。基因包括DNA中位于编码区(可读框,ORF)之前(上游)和之后(下游)的非翻译调节区(例如启动子、增强子、阻抑物等),以及根据需要,在各个编码区(例如外显子)之间的间插序列(例如内含子)。如本文中所述的术语“结构基因”意图指转录成mRNA的DNA序列,其中所述的mRNA随后翻译成作为具体多肽的特征的氨基酸序列。

[0087] “基因编辑”在本文中使用时意指在细胞基因组的特定位置引入特定突变。可以应用更先进技术例如使用CRISPR Cas系统和供体DNA或与致突变活性如脱氨酶关联的CRISPR Cas系统,通过精确编辑过程引入基因编辑(WO15133554,WO17070632)。

[0088] 基因组和基因组DNA:术语“基因组”或“基因组DNA”指宿主生物的可遗传信息。在真核生物中,所述基因组DNA包括胞核的DNA(也称作染色体DNA),还包括质体(例如,叶绿体)和其他细胞器(例如,线粒体)的DNA。优选地,术语“基因组”或基因组“DNA”指胞核的染色体DNA。在原核生物中,所述基因组DNA包括细菌细胞内部的染色体DNA。

[0089] 异源的:就核酸分子或DNA而言,术语“异源的”指这样的核酸分子,所述核酸分子有效连接于或受到操作以变成有效连接于自然界中(例如,野生型(WT)植物的基因组中)不与该核酸分子有效连接或自然界中(例如,WT植物的基因组中)与该核酸分子在不同部位或位置有效连接的第二种核酸分子,例如启动子。

[0090] 优选地,就核酸分子或DNA(例如NEENA)而言,术语“异源的”指这样的核酸分子,所述核酸分子有效连接于或受到操作以变成有效连接于自然界中并不与之有效连接的第二种核酸分子,例如,启动子。

[0091] 包含核酸分子和与之连接的一个或多个调节核酸分子(如启动子或转录终止信号)的异源表达构建体例如是源自实验操作的构建体,在所述构建体中,a)所述核酸分子或b)所述调节核酸分子或c)二者(即(a)和(b))不位于其天然(原有)遗传环境中或已经因实验操作受到修饰,修饰的实例是一个或多个核苷酸残基的置换、添加、删除、倒位或插入。天然遗传环境指源生物中的天然染色体基因座或指存在于基因组文库中。在基因组文库的情况下,优选地保留、至少部分地保留该核酸分子的序列的天然遗传环境。该环境分布在该核酸序列的至少一侧并且具有至少50bp、优选地至少500bp、特别优选地至少1,000bp、非常特别优选地至少5,000bp序列长度。天然存在的表达构建体例如启动子与相应基因的天然存在组合-在通过非天然的合成性“人工”方法例如诱变法修饰时变成转基因表达构建体。已经描述了此类方法(US 5,565,350;WO 00/15815)。例如,与启动子有效连接的编码蛋白质

的核酸分子相对于该启动子视为异源,其中所述的启动子不是该核酸分子的天然启动子。优选地,异源DNA相对于导入该异源DNA的细胞并非内源或不天然与该细胞相关,但是已经从另一种细胞获得或已经合成。异源DNA还包括含有一些修饰的内源DNA序列、不天然存在的多拷贝内源DNA序列或这样的DNA序列,该DNA序列不与同物理连接于所述DNA序列的另一个DNA序列天然接合。通常地,虽然不必然地,异源DNA编码正常情况下引入所述异源DNA的细胞不编码的RNA或蛋白质。

[0092] 如本文中定义的术语“杂交”是其中基本上互补的核苷酸序列相互复性的过程。杂交过程可以完全在溶液中进行,即两种互补性核酸均处于溶液中。杂交过程也可以在互补性核酸之一固定到中等如磁珠、琼脂糖凝胶珠或任何其他树脂的情况下发生。杂交过程也可以在互补性核酸之一固定至固相支持体如硝酸纤维素膜或尼龙膜上或通过例如照相平版印刷术固定至载体,包括但不限于硅酸玻璃支持物(后者称作核酸阵列或微阵列或称作核酸芯片)上的情况下进行。为使杂交发生,通常将核酸分子热变性或化学变性以使双链解链成为两条单链和/或去除来自单链核酸的发夹或其它二级结构。

[0093] 杂交分子的这种形成或解链取决于多种参数,包括但不限于温度。温度增加有利于解链,而温度下降有利于杂交。但是,这种杂交分子形成过程不以线性方式随施加的温度变化而:杂交过程是动态的,并且已经形成的核苷酸对也支持相邻核苷酸配对。因而,在良好逼近情况下,杂交是一个是或非过程,并且存在一个基本上规定杂交与非杂交之间界限的温度。这个温度是解链温度(T_m)。 T_m 是以摄氏度计的温度,在所述温度,50%具有给定出核苷酸序列的全部分子杂交成双链,并且50%作为单链存在。

[0094] 解链温度(T_m)取决于所分析核酸序列的物理特性并且因此可以指示两个不同序列之间的关系。但是,解链温度(T_m)还受完全与序列不相关的各种其他参数影响,并且必须考虑施加的杂交实验条件。例如,增加盐(例如一价阳离子)导致 T_m 较高。

[0095] 可以通过进行物理杂交实验确定给定杂交条件的 T_m ,但是也可以计算机模拟估计给定DNA序列对的 T_m 。在这个实施方案中,Meinkoth和Wahl的等式(Anal. Biochem., 138: 267-284, 1984)用于具有50个或更多个碱基长度的片段: $T_m = 81.5^\circ\text{C} + 16.6(\log M) + 0.41(\%GC) - 0.61(\text{甲酰胺}\%) - 500/L$ 。

[0096] M 是一价阳离子的体积摩尔浓度,%GC是DNA片段中鸟苷和胞嘧啶的百分数,甲酰胺%是杂交溶液中甲酰胺的百分数,并且 L 是以碱基对计的杂交分子的长度。该等式用于0.01至0.4M的盐范围和30%至75%的GC%。

[0097] 尽管上述 T_m 是完美匹配探针的温度,对每1%错配, T_m 降低约 1°C (Bonner等人, J. Mol. Biol. 81:123-135, 1973): $T_m = [81.5^\circ\text{C} + 16.6(\log M) + 0.41(\%GC) - 0.61(\%甲酰胺) - 500/L] - \text{非同一性}\%$ 。

[0098] 这个等式可用于具有35或更多个核苷酸的探针并且在科学方法文献(例如,引用于:“Recombinant DNA Principles and Methodologies”, James Greene, 章节“Biochemistry of Nucleic acids”, Paul S. Miller, 第55页; 1998, CRC Press)、许多专利申请(例如,引用于:US 7026149)及还在商业企业的数据文件(例如来自www.genomics.agilent.com的“Equations for Calculating T_m ”)中广泛应用。

[0099] 在这个实施方案中不太优选的用于计算 T_m 的其他公式仅可能用于所示的情形:

[0100] 对于DNA-RNA杂交分子(Casey, J. 和Davidson, N. (1977) Nucleic Acids Res., 4:

1539) :

[0101] $T_m = 79.8^\circ\text{C} + 18.5 (\log M) + 0.58 (\%GC) + 11.8 (\%GC * \%GC) - 0.5 (\text{甲酰胺}\%) - 820/L$ 。

[0102] 对于RNA-RNA杂交分子 (Bodkin, D.K. 和 Knudson, D.L. (1985) *J. Virol. Methods*, 10:45) :

[0103] $T_m = 79.8^\circ\text{C} + 18.5 (\log M) + 0.58 (\%GC) + 11.8 (\%GC * \%GC) - 0.35 (\text{甲酰胺}\%) - 820/L$ 。

[0104] 对少于20个碱基的寡核苷酸探针 (Wallace, R.B. 等人 (1979) *Nucleic Acid Res.* 6:3535) : $T_m = 2x_n (A+T) + 4x_n (G+C)$, 其中n是形成杂交分子的探针中相应碱基的数目。

[0105] 对于20-35个核苷酸的寡核苷酸探针, 改良的Wallace计算可以适用: $T_m = 22 + 1.46n (A+T) + 2.92n (G+C)$, 其中n是形成杂交分子的探针中相应碱基的数目。

[0106] 对于其他寡核苷酸, 应连同适宜的热动力数据一起, 使用计算解链温度的最近邻模型:

[0107] $T_m = (\sum (\Delta H_d) + \Delta H_i) / (\sum (\Delta S_d) + \Delta S_i + \Delta S_{self} + R \times \ln (cT/b)) + 16.6 \log [Na^+] - 273.15$ (Breslauer, K.J., Frank, R., Blöcker, H., Marky, L.A. 1986 Predicting DNA duplex stability from the base sequence (从碱基序列预测DNA双链体稳定性). *Proc. Natl Acad. Sci. USA* 83:3746-3750; Alejandro Panjkovich, Francisco Melo, 2005. Comparison of different melting temperature calculation methods for short DNA sequences (比较短DNA序列的不同解链温度计算方法) *Bioinformatics*, 21 (6) :711-722)

[0108] 其中:

[0109] T_m 是以摄氏度的解链温度;

[0110] $\sum (\Delta H_d)$ 和 $\sum (\Delta S_d)$ (相应地) 是对全部内部最近邻二重峰计算的焓总和和熵总和;

[0111] ΔS_{self} 是自身互补序列的熵罚分;

[0112] ΔH_i 和 ΔS_i 分别是起始焓和起始熵的总和;

[0113] R是气体常数 (在1,987 cal/K · mol 固定);

[0114] cT是以摩尔单位计的总链浓度;

[0115] 对非自身互补序列, 常数b采用值4, 或对自身互补链的双链体或对链之一明显过量时的双链体, 其等于1。

[0116] 热动力计算假设: 复性在缓冲溶液中接近7.0的pH发生并且出现双态转变。

[0117] 用于该计算的热动力值可以获自 (Alejandro Panjkovich, Francisco Melo, 2005. Comparison of different melting temperature calculation methods for short DNA sequences (比较短DNA序列的不同解链温度计算方法). *Bioinformatics*, 21 (6) :711-722) 的表1, 或获自原始研究论文 (Breslauer, K.J., Frank, R., Blöcker, H., Marky, L.A. 1986 Predicting DNA duplex stability from the base sequence (从碱基序列预测DNA双链体稳定性). *Proc. Natl Acad. Sci. USA* 83:3746-3750; Santa Lucia, J., Jr, Allawi, H.T., Seneviratne, P.A. 1996 Improved nearest-neighbor parameters for predicting DNA duplex stability (用于预测DNA双链体稳定性的改良最近邻居参数). *Biochemistry*

353555-3562; Sugimoto, N., Nakano, S., Yoneyama, M., Honda, K. 1996 Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes (预测DNA双链体稳定性的改良热力学参数和螺旋起始因子). *Nucleic Acids Res.* 24:4501-4505)。

[0118] 为了计算机模拟估计根据这个实施方案的 T_m , 首先在两个序列之间生成一个生物信息学序列比对结果集合。可以通过本领域技术人员已知的各种工具产生这类比对结果, 如产生局部比对结果的程序“Blast” (NCBI)、“Water” (EMBOSS) 或“Matcher” (EMBOSS) 或产生全局比对结果的程序“Needle” (EMBOSS)。这些工具应当配合其默认参数集应用, 还配合某些参数变动应用。例如, 程序“MATCHER”可以随空位开放/空位延伸的各种参数(如14/4; 14/2; 14/5; 14/8; 14/10; 20/2; 20/5; 20/8; 20/10; 30/2; 30/5; 30/8; 30/10; 40/2; 40/5; 40/8; 40/10; 10/2; 10/5; 10/8; 10/10; 8/2; 8/5; 8/8; 8/10; 6/2; 6/5; 6/8; 6/10) 应用并且程序“WATER”可以随空位开放/空位延伸的各种参数(如10/0.5; 10/1; 10/2; 10/3; 10/4; 10/6; 15/1; 15/2; 15/3; 15/4; 15/6; 20/1; 20/2; 20/3; 20/4; 20/6; 30/1; 30/2; 30/3; 30/4; 30/6; 45/1; 45/2; 45/3; 45/4; 45/6; 60/1; 60/2; 60/3; 60/4; 60/6) 应用, 并且这些程序还应当通过使用核苷酸序列作为给出, 还随处于其反互补形式的序列之一而应用。例如, BlastN (NCBI) 可以配合增加的 e -值截值(例如 $e+1$ 或甚至 $e+10$) 应用, 旨在还鉴定非常短的比对结果, 特别地在小规格数据库中如此。

[0119] 重要是考虑局部比对, 因为杂交可能并不必然地在两个序列的整个长度范围内发生, 但可能在不同区域最佳发生, 这些区域则决定实际解链温度。因此, 从产生的全部比对结果中, 必须确定比对长度、比对结果的GC含量% (以更精确方式, 比对结果内部匹配性碱基的%GC含量) 及比对同一性。随后, 必须对每个比对结果计算预测的解链温度(T_m)。最高计算 T_m 用来预测实际解链温度。

[0120] 如本文定义的术语“在本发明的完整序列范围内杂交”意指对于本发明的序列片段化成为约300至500碱基长度的小片时长度超过300个碱基的序列, 每个片段必须杂交。例如, 可以通过使用一种限制性酶或使用限制性酶组合, 使DNA片段化成小片。通过如上文所述的相同程序进行 T_m 的生物信息学计算机模拟计算, 仅对每个片段进行。可以通过标准DNA印迹分析或本领域技术人员已知的可比方法, 分析各个片段的实际杂交。

[0121] 如本文定义的术语“严格性”描述了可能籍此在两个核苷酸序列之间形成杂交分子的难易度。“较高严格性”的条件需要一个序列的更多碱基与另一个序列配对(“较高严格性”的条件下解链温度 T_m 降低), “较低严格性”的条件允许某些更多碱基不配对。因此, 两个序列之间关系的程度可以通过它们仍能够形成杂交分子的实际严格性条件来估计。可以通过保持实验杂交温度恒定并降低盐浓度, 或通过保持实验性杂交温度恒定并增加实验性杂交温度或这些参数的组合, 实现增加严格性。另外, 增加甲酰胺将增加严格性。技术人员知晓可以在杂交期间加以改变和将要维持或变更严格性条件的额外参数(Sambrook等人(2001) *Molecular Cloning: a laboratory manual*, 第3版, Cold Spring Harbor Laboratory Press, CSH, New York 或 *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989和年度更新版))。

[0122] 通过初始杂交步骤, 后续一个至几个洗涤步骤, 进行常见杂交实验。用于这些步骤的溶液可以含有防止分析序列降解和/或防止探针的非特异性背景结合的额外组分, 如

EDTA、SDS、片段化精子DNA或相似试剂,这些组分是本领域技术人员已知(Sambrook等人(2001)Molecular Cloning:a laboratory manual,第3版,Cold Spring Harbor Laboratory Press,CSH,New York或Current Protocols in Molecular Biology,John Wiley&Sons,N.Y.(1989和年度更新版))。

[0123] 通过随机引发标记方法产生杂交实验的常见探针,所述方法最初由Feinberg和Vogelstein开发(Anal.Biochem.,132(1),6-13(1983);Anal.Biochem.,137(1),266-7(1984)并基于全部可能六核苷酸的混合物与待标记DNA的杂交。标记的探针产物实际上将是一组具有可变长度、一般长度在100-1000个核苷酸大小范围内的片段,最高片段浓度一般约为200至400bp。最终作为杂交实验的探针使用的探针片段的实际大小范围也可以受以下影响:所用标记方法的参数、已产生的探针的后续纯化(例如琼脂糖凝胶)和用于标记的所用模板DNA的大小(标记前,可以例如使用4bp切割酶(例如HaeIII)限制性消化大模板)。

[0124] 对于本发明,通过杂交实验分析本文所述的序列,其中探针从另一个序列产生,并且通过标准随机引发标记方法产生这种探针。对于本发明,探针由一组具有约200-400个核苷酸大小的标记的寡核苷酸组成。在本发明的序列和另一个序列之间杂交意指探针的杂交在本发明的完整序列范围内出现,如上文定义。通过实现依据最终洗涤步骤的严格性的最高严格性,进行杂交实验。最终洗涤步骤具有与以下洗涤条件的严格性可比的严格性条件:至少洗涤条件1:在50℃时1.06x SSC、0.1% SDS、0%甲酰胺;在另一个实施方案中至少洗涤条件2:在55℃时1.06x SSC、0.1% SDS、0%甲酰胺;在另一个实施方案中至少洗涤条件3:在60℃时1.06x SSC、0.1% SDS、0%甲酰胺;在另一个实施方案中至少洗涤条件4:在65℃时1.06x SSC、0.1% SDS、0%甲酰胺;在另一个实施方案中至少洗涤条件5:在65℃时0.52x SSC、0.1% SDS、0%甲酰胺;在另一个实施方案中至少洗涤条件6:在65℃时0.25x SSC、0.1% SDS、0%甲酰胺;在另一个实施方案中至少洗涤条件7:在65℃时0.12x SSC、0.1% SDS、0%甲酰胺;在另一个实施方案中至少洗涤条件8:在65℃时0.07x SSC、0.1% SDS、0%甲酰胺。

[0125] “低严格洗涤”具有这样的严格性条件,其与至少洗涤条件1的严格性条件可比,但是没有比洗涤条件3更严格,其中洗涤条件如上文所述。

[0126] “高严格洗涤”具有这样的严格性条件,其与至少洗涤条件4、在另一个实施方案中至少洗涤条件5、在另一个实施方案中至少洗涤条件6、在另一个实施方案中至少洗涤条件7、在另一个实施方案中至少洗涤条件8可比,其中洗涤条件如上文所述。

[0127] “同一性”:在比较两个或多个核酸或氨基酸分子使用时,“同一性”意指所述分子的序列共有某种程度的序列相似性,即所述序列是部分相同的。

[0128] 酶变体可以由它们与亲本酶相比时的序列同一性定义。序列同一性通常作为“序列同一性%”或“同一性%”提供。为了确定两个氨基酸序列之间的同一性百分数,在第一步骤中,在这两个序列之间生成配对序列比对结果,其中将两个序列在其整个长度范围内比对(即,配对的全局比对)。用实施Needlem和Wunsch算法(J.Mol.Biol.(1979)48,第443-453页)的程序,优选地通过使用程序“NEEDLE”(欧洲分子生物学开放软件套件(European Molecular Biology Open Software Suite)(EMBOSS)),采用程序默认参数(空位开口=10.0、空位延伸=0.5和矩阵=EDNAFULL),生成比对结果。

[0129] 以下实例意在说明两个核苷酸序列,但是相同计算适用于蛋白质序列:

[0130] Seq A:AAGATACTG长度:9个碱基

[0131] Seq B:GATCTGA长度:7个碱基

[0132] 因此,较短的序列是序列B。

[0133] 产生在其整个长度范围内显示两个序列的两两全局比对产生了

Seq A: AAGATACTG-

[0134] ||| |||

Seq B: --GAT-CTGA

[0135] 比对结果中的“|”符号表示相同残基(其意指DNA的碱基或蛋白质的氨基酸)。相同残基的数目是6。

[0136] 比对结果中的“-”符号表示空位。Seq B内部因比对引入的空位数是1。通过比对引入的空位的编号在Seq B的边界处是2,并且在Seq A的边界处为1。

[0137] 在其整个长度范围内显示比对序列的比对长度是10。

[0138] 根据本发明产生在其整个长度范围内显示较短序列的两两比对因此产生:

Seq A: GATACTG-

[0139] ||| |||

Seq B: GAT-CTGA

[0140] 根据本发明产生在其整个长度范围内显示序列A的两两比对因此产生:

Seq A: AAGATACTG

[0141] ||| |||

Seq B: --GAT-CTG

[0142] 根据本发明产生在其整个长度范围内显示序列B的两两比对因此产生:

Seq A: GATACTG-

[0143] ||| |||

Seq B: GAT-CTGA

[0144] 在其整个长度范围内显示较短序列的比对长度是8(存在一个在较短序列的比对长度中考虑的缺口)。

[0145] 因此,在其整个长度范围内显示Seq A的比对长度将是9(意指Seq A是本发明的序列)。

[0146] 因此,在其整个长度范围内显示Seq B的比对长度将是8(意指Seq B是本发明的序列)。

[0147] 在比对两个序列后,在第二步骤中,从产生的比对结果测定同一性值。为了本说明书的目的,通过以下方式计算同一性百分数:同一性%=(相同残基/在其整个长度范围内显示本发明相应序列的比对区域的长度)*100。因此,根据这个实施方案,通过相同残基的数目除以其整个长度范围内显示本发明相应序列的比对区域的长度,计算与比较两个氨基酸序列有关的序列同一性。这个值乘以100以得出“同一性%”。根据上文提供的实例,同一性%:对于作为本发明序列的Seq A $(6/9) * 100 = 66.7\%$;对于作为本发明序列的Seq B、 $(6/8) * 100 = 75\%$ 。

[0148] Indel(插入/删除)是针对生物的基因组中随机插入或删除与NHEJ修复DSB相关的

碱基的术语。将它在小遗传性变异当中分类,计量1至10 000个碱基对长度。如本文所用,它指在靶位点中或紧邻靶位点(例如其上游和/或下游小于1000bp、900bp、800bp、700bp、600bp、500bp、400bp、300bp、250bp、200bp、150bp、100bp、50bp、40bp、30bp、25bp、20bp、15bp、10bp或5bp)随机插入或删除碱基。

[0149] 就靶DNA的靶位点中引入供体DNA分子而言,术语“引入”、“导入”等意指例如通过将供体DNA分子或其部分物理整合入靶区域而将供体DNA分子的序列向靶区域中的任何引入或将供体DNA分子或其部分的序列引入其中使用供体DNA作为聚合酶模板的靶区域中。

[0150] 同基因的:除可以因存在或不存在异源DNA序列而不同之外,在遗传上相同的生物(例如植物)。

[0151] 分离的:如本文中所述的术语“分离”意指材料已经通过人工取出并且离开其原来的天然环境存在并且因此不是自然界的产物。分离的材料或分子(如DNA分子或酶)可以按纯化的形式存在或可以存在于非天然环境中如例如存在于转基因宿主细胞中。例如,活植物中存在的天然存在多核苷酸或多肽不是分离的,然而与该天然系统中一些或全部共存物质分开的相同多核苷酸或多肽是分离的。此类多核苷酸可以是载体的一部分和/或此类多核苷酸或多肽可以是组合物的一部分,和是分离的,因为这种载体或组合物不是其原始环境的一部分。优选地,术语“分离的”在与核酸分子相关地使用时,如在“分离的核酸序列”中使用,是指被鉴定并且与该核酸序列天然来源中通常与其接合的至少一种杂质性核酸分子分开的核酸序列。分离的核酸分子是这样的核酸分子,其在与自然界中找到该核酸分子的形式或环境不同的形式或环境下存在。相反,未分离的核酸分子是其在自然界中存在的状态所找到的核酸分子如DNA和RNA。例如,在宿主细胞染色体上相邻基因附近找到给定的DNA序列(例如,基因);作为与编码多种蛋白质的众多其他mRNA的混合物,在细胞中找到RNA序列,如编码特定蛋白质的特定mRNA序列。但是,包含例如SEQ ID NO:12的分离的核酸序列以举例方式包括细胞中通常含有SEQ ID NO:12的此类核酸序列,其中核酸序列处在不同于天然细胞的染色体位置或染色体外位置或否则旁侧分布有与自然界中存在者不同的核酸序列。分离的核酸序列可以以单链或双链形式存在。当分离的核酸序列用来表达蛋白质时,该核酸序列将最少含有有义链或编码链的至少一部分(即,该核酸序列可以为单链)。备选地,它可以含有有义链和反义链(即,该核酸序列可以为双链)。

[0152] 非编码的:术语“非编码”指核酸分子中不编码所表达蛋白质的部分或全部的序列。非编码序列包括,但不限于内含子、增强子、启动子区、3'非翻译区和5'非翻译区。

[0153] 核酸和核苷酸:术语“核酸”和“核苷酸”指天然存在的或合成的或人工核酸或核苷酸。术语“核酸”和“核苷酸”包括处于单链或双链、有义或反义形式的脱氧核糖核苷酸或核糖核苷酸或其任何核苷酸类似物和聚合物或杂合体。除非另外说明,否则具体核酸序列也内在包括其保守方式修饰的变体(例如简并密码子置换)和互补序列,以及明确指出的序列。术语“核酸”在本文中“基因”、“cDNA”、“mRNA”、“寡核苷酸”和“多核苷酸”互换地使用。核苷酸类似物包括在碱基、糖和/或磷酸酯的化学结构中具有修饰的核苷酸,包括但不限于5-位置嘧啶修饰、8-位置嘌呤修饰、胞嘧啶环外环胺处的修饰、5-溴-尿嘧啶置换等;和2'-位置糖修饰,包括但不限于糖修饰的核糖核苷酸,其中2'-OH由选自H、OR、卤素、SH、SR、NH₂、NHR、NR₂或CN的基团替换。短发夹RNA(shRNA)也可以包含非天然元件如非天然碱基,例如,肌昔和黄嘌呤,非天然糖,例如,2'-甲氧基核糖,或非天然的磷酸二酯键,例如甲基磷酸酯、

硫代磷酸酯和肽。

[0154] 核酸序列:短语“核酸序列”指从5'-端至3'-端读取的脱氧核糖核苷酸或核糖核苷酸的单链或双链聚合物。它包括染色体DNA、自我复制型质粒、DNA或RNA的感染性聚合物和主要发挥结构性作用的DNA或RNA。“核酸序列”也指代表核苷酸的缩写、字母、字符或字的连续串。在一个实施方案中,核酸可以是“探针”,所述探针是相对短的核酸,通常长度小于100个核苷酸。经常地,核酸探针具有约50个核苷酸长度至约10个核苷酸长度。核酸的“靶区域”是核酸中被鉴定为有目标的部分。核酸的“编码区”是核酸的部分,其中置于适宜的调节性序列控制下时,所述部分以序列特异性方式转录和翻译以产生特定的多肽或蛋白质。称该编码区编码这种多肽或蛋白质。

[0155] 寡核苷酸:术语“寡核苷酸”指核糖核酸(RNA)或脱氧核糖核酸(DNA)或其模拟物的低聚物或聚合物,以及具有类似发挥作用的非天然存在部分的寡核苷酸。此类修饰或取代的寡核苷酸因合乎需要的特性,例如增强的细胞摄取、增强的核酸靶亲和力和在核酸酶存在下增加的稳定性而经常胜过其天然形式为优选。寡核苷酸优选地包括通过键(例如,磷酸二酯键)或取代键(substitute linkage)相互共价偶联的两个或多个核苷酸单体(nucleomonomer)。

[0156] 突出端:“突出端”是双链寡核苷酸分子的5'-或3'-羟基端上相对短的单链核苷酸序列(也称作“延伸”、“伸出端”或“粘末端”)。

[0157] 多肽:术语“多肽”、“肽”、“寡肽”、“多肽”、“基因产物”、“表达产物”和“蛋白质”在文中使用互换地适用,用来指连续氨基酸残基的聚合物或低聚物。

[0158] 前蛋白:正常情况下靶向细胞器如叶绿体并且仍包含其转运肽的蛋白质。

[0159] 就靶区域中引入供体DNA分子而言,“精确”意指如与未包含于供体DNA分子序列中的靶区域的未改变DNA序列相比,将供体DNA分子的序列在无任何插入/删除、重复或其他突变的情况下引入靶区域。

[0160] 初级转录物:如本文中所述的术语“初级转录物”指基因的不成熟RNA转录物。“初级转录物”例如仍包含内含子和/或仍不包含聚腺苷酸尾或帽结构和/或是缺少对其作为转录物正常发挥作用必需的其他修饰,例如修剪或编辑。

[0161] “启动子”或“启动子序列”或“调节核酸”是位于某基因上游与该基因相同链上的能够实现该基因转录的核苷酸序列。启动子后接基因的转录起始位点。启动子被RNA聚合酶(连同任何需要的转录因子一起)识别,这启动转录。启动子的功能性片段或功能性变体是由RNA聚合酶可识别并且能够启动转录的核苷酸序列。

[0162] 纯化的:如本文中所述,术语“纯化的”指从其天然环境取出、分离或分开的分子,即核酸序列或氨基酸序列。“基本上纯化的”分子至少60%没有、优选地至少75%没有和更优选地至少90%没有与它们天然结合在一起的其他组分。纯化的核酸序列可以是分离的核酸序列。

[0163] 重组:就核酸分子而言,术语“重组”指通过重组DNA技术产生的核酸分子。重组核酸分子也可以包括本身不存在于自然界中但是被人修饰、改变、突变或操纵的分子。优选地,“重组核酸分子”是在序列方面与天然存在的核酸分子有至少一个核酸不同的非天然存在的核酸分子。“重组核酸分子”也可以包括“重组构建体”,其包含不天然地以这种顺序存在的一系列核酸分子,所述核酸分子优选地有效地连接。用于产生所述重组核酸分子的优

选方法可以包括克隆技术、定向或非定向诱变法、合成或重组技术。

[0164] 减少的表达:本文中同等使用“减少”或“降低”核酸分子在细胞中的表达,并且这意指该核酸分子在应用本发明方法之后在细胞中的表达水平比其在该方法之前在细胞中的表达更低,或与缺少本发明重组核酸分子的参考细胞相比更低。例如,参考细胞包含相同的构建体,所述构建体包含本发明的起始调节核酸分子且不包含本发明的合成性调节核酸分子。如本文中所述的术语“减少”或“降低”是同义的,并且在本文中意指待表达核酸分子的表达减少、优选显著减少。如本文中所述,“减少”某物质(如蛋白质、mRNA或RNA)的水平意指相对于基本上相同条件下培育的、缺少本发明重组核酸分子(例如包含本发明起始调节核酸分子且不包含本发明合成性调节核酸分子)的基本上相同细胞,该水平降低。如本文所用,某物质(如例如由靶基因表达的preRNA、mRNA、rRNA、tRNA、snoRNA、snRNA)的水平和/或由靶基因编码的蛋白质产物的水平“降低”意指相对于缺少本发明重组核酸分子(例如包含本发明起始调节核酸分子且不包含本发明合成性调节核酸分子)的细胞,该水平降低10%或更多,例如20%或更多、30%或更多、40%或更多、优选地50%或更多、例如60%或更多、70%或更多、80%或更多、90%或更多。可以通过技术人员熟悉的方法测定降低情况。因而,可以例如通过蛋白质的免疫学检测法确定核酸或蛋白数量的降低。另外,可以使用技术如蛋白质测定法、荧光法、RNA杂交法、核酸酶保护测定法、逆转录法(定量RT-PCR)、ELISA(酶联免疫吸附测定法)、蛋白质印迹法、放射免疫测定法(RIA)或其他免疫测定法和荧光激活的细胞分析(FACS)来测量细胞中的特定蛋白质或RNA。取决于已减少蛋白质产物的类型,也可以确定其活性或对生物或细胞表型的影响。技术人员已知用于确定蛋白质数量的方法。可以提到的实例是:微量Biuret法(Goa J(1953) Scand J Clin Lab Invest 5:218-222)、Folin-Ciocalteu法(Lowry OH等人(1951) J Biol Chem 193:265-275)或测量CBB G-250的吸光度(Bradford MM(1976) Analyt Biochem 72:248-254)。

[0165] 有义:术语“有义”理解为意指核酸分子,其具有与靶序列互补或相同的序列,例如与蛋白质转录因子结合和参与给定基因表达的序列。根据一个优选的实施方案,该核酸分子包含目的基因和允许表达该目的基因的元件。

[0166] 显著的增加或下降:大于测量技术中固有误差幅度的增加或减少,例如在酶活性或在基因表达方面,优选地是对照酶的活性或对照细胞中的表达增加或减少约2倍或更多,更优选地增加或减少约5倍或更多,并且最优选地增加或减少约10倍或更多。

[0167] 小核酸分子:“小核酸分子”理解为由核酸或其衍生物如RNA或DNA组成的分子。它们可以是双链或单链的,并且其长度在约15和约30bp之间,例如在15和30bp之间,更优选在约19和约26bp之间,例如在19和26bp之间,甚至更优选在约20和约25bp之间,例如在20和25bp之间。在一个特别优选的实施方案中,寡核苷酸的长度在约21和约24bp之间,例如在21和24bp之间。在最优选的实施方案中,小核酸分子的长度是约21bp和约24bp,例如21bp和24bp。

[0168] 基本上互补的:在最广意义上,本文中就核苷酸序列相对于参考或目标核苷酸序列而言使用时,术语“基本上互补的”意指这样的核苷酸序列,其具有在基本上互补的核苷酸序列和所述参考或目标核苷酸序列的完全互补序列之间至少60%、更希望地至少70%、更希望地至少80%或85%、优选地至少90%、更优选地至少93%、仍更优选地至少95%或96%、仍旧更优选地至少97%或98%、依旧更优选地至少99%或最优选地100%的同一性百

分数(后者等同于本上下文中的术语“相同”)。优选地,在至少19个核苷酸长度、优选地至少50个核苷酸长度、更优选核酸序列的全部长度范围内针对所述参比核苷酸序列评估同一性(如果不是,则另外在下文说明)。使用威斯康辛大学GCG的默认GAP分析,GAP的SEQWEB应用,基于Needleman和Wunsch算法(Needleman和Wunsch(1970) J Mol. Biol. 48:443-453;如上文定义)实施序列比较。与参比核苷酸序列“基本上互补的”核苷酸序列与该参比核苷酸序列在低严格性条件、优选地中等严格性条件、最优选地高严格性条件(如上文定义)下杂交。

[0169] 如本文所用的“靶区域”意指这样的区域,其距离靶位点几乎例如10个碱基、20个碱基、30个碱基、40个碱基、50个碱基、60个碱基、70个碱基、80个碱基、90个碱基、100个碱基、125个碱基、150个碱基、200个碱基或500个碱基或更多个碱基或包含其中将供体DNA分子序列向细胞基因组引入的靶位点。

[0170] 如本文所用的“靶位点”意指基因组中使用重组技术如Zn指、TALEN、限制性酶、归巢核酸内切酶、RNA导向核酸酶、RNA-导向切刻酶如CRISPR/Cas核酸酶或切刻酶等诱导双链断裂或一个或一对单链断裂(切刻)的位置。

[0171] 转基因:如本文中所述的术语“转基因”指通过实验操作导入细胞的基因组中的任何核酸序列。转基因可以是“内源DNA序列”或“异源DNA序列”(即,“外来DNA”)。术语“内源DNA序列”指这样的核苷酸序列,其天然存在于引入该核苷酸序列的细胞中,只要它相对于天然存在序列不含有一些修饰(例如,例如,点突变、存在可选择标记基因等)。

[0172] 转基因的:当提及生物时,术语“转基因的”意指用重组DNA分子转化生物,优选地稳定转化该生物,其中所述重组DNA分子优选地包含与目的DNA序列有效连接的合适启动子。

[0173] 载体:如本文中所述,术语“载体”指能够运输已经与之连接的另一个核酸分子的核酸分子。一种类型的载体是基因组整合型载体,或“整合型载体”,所述载体可以整合至宿主细胞的染色体DNA中。另一个类型的载体是游离型载体,即,能够进行染色体外复制的核酸分子。本文中将能够指导与它们有效连接的基因表达的载体称作“表达载体”。在本说明书中,“质粒”和“载体”互换地使用,除非从上下文并非如此。设计旨在体外或体内产生如本文所述RNA的表达载体可以含有被任何RNA聚合酶识别的序列,所述的RNA聚合酶包括线粒体RNA聚合酶、RNA pol I、RNA pol II和RNA pol III。这些载体可以以用来在根据本发明的细胞中转录想要的RNA分子。

[0174] 附图简述

[0175] 图1

[0176] 描述了单一CRISPR/Cas9质粒pCC009的质粒图。质粒pCC009是质粒pJOE8999.1的衍生物,其携带地衣芽孢杆菌amyB基因的间隔子和分别在amyB基因5'和3'的DNA供体序列HomA和HomB。PmanP:枯草芽孢杆菌manP基因的启动子;pUC ORI:大肠杆菌高拷贝复制起点;在芽孢杆菌和大肠杆菌中均有功能的卡那霉素抗性基因;rep pE194:在芽孢杆菌中赋予温度敏感性质粒复制的质粒pE194片段;PvanP:驱动间隔区-sgRNA(crRNA重复+ 'gRNA)表达的启动子;来自λ的T0终止子;来自大肠杆菌rrnB基因的t1t2终止子;HomA和HomB:融合在一起以便进行基因删除的amyB基因5'和3'序列;Cas9:来自化脓性链球菌(S. pyogenes)的Cas9核酸内切酶。

[0177] 图2:

[0178] 显示已突变启动子序列的选定区域的序列比对-针对启动子序列PV4 (SEQ ID 028) 和PV8 (SEQ ID 029) 的nt 15至nt.128参照。在PV4 (SEQ ID 028) 和PV8 (SEQ ID 029) 启动子的参考启动子序列内部, -35区和-10区、转录起始位点 (TSS) 和Shine Dalgarno序列 (SD) 以斜体字母描述并且加灰色阴影。核苷酸删除、插入和突变以粗体描述。

[0179] 图3

[0180] 通过菌落PCR以位于基因删除用同源性区域外部的寡核苷酸SEQ ID 009和SEQ ID 010分析单菌落的地衣芽孢杆菌amyB基因删除。将地衣芽孢杆菌淀粉酶amyB基因的基因删除效率(作为具有已失活淀粉酶基因的克隆相对于总计20个针对每个基因删除构建体已分析的克隆的百分数)针对如所示的每个基因删除构建体作图。A. 描述衍生自PV4启动子变体的删除质粒的相对删除效率。B. 描述衍生自PV8启动子变体的删除质粒的相对删除效率。

[0181] 图4

[0182] A. 将地衣芽孢杆菌hag基因的基因删除效率(作为具有已失活hag基因的克隆相对于总计20个已分析克隆的百分数)针对分别如所示的两个删除用构建体和启动子变体作图。显示三个独立实验的平均数连同标准差。通过菌落PCR用位于基因删除用同源性区域外部的寡核苷酸SEQ ID 087和SEQ ID 088分析hag基因的基因删除。B. 描述两个删除用构建体和启动子变体分别对于地衣芽孢杆菌degU基因内部引入点突变的相对突变效率, 作为具有已突变degU基因的克隆相对于总计20个已分析克隆的百分数。显示三个独立实验的平均数连同标准差。通过菌落PCR用位于用来引入基因突变的同源性区域外部的寡核苷酸SEQ ID 089和SEQ ID 090分析degU基因的基因突变, 随后进行PCR片段的PstI限制性消化以区分天然型和突变型degU基因座。

[0183] 图5

[0184] A. 将枯草芽孢杆菌淀粉酶amyE基因的基因删除效率(作为具有已失活amyE基因的克隆相对于总计20个已分析克隆的百分数)针对分别如所示的两个删除用构建体和启动子变体作图。显示三个独立实验的平均数连同标准差。通过菌落PCR用位于基因删除用同源性区域外部的寡核苷酸SEQ ID 091和SEQ ID 092分析amyE基因的基因删除。B. 描述两个删除构建体和启动子变体分别对于删除枯草芽孢杆菌的枯草杆菌蛋白酶aprE基因的相对删除效率, 作为具有已失活aprE基因的克隆相对于总计20个已分析克隆的百分数。显示三个独立实验的平均数连同标准差。通过菌落PCR用位于基因删除用同源性区域外部的寡核苷酸SEQ ID 093和SEQ ID 094分析aprE基因的基因删除。

[0185] 图6

[0186] A. 将地衣芽孢杆菌vpr基因的基因删除效率(作为具有已失活vpr基因的克隆相对于总计20个已分析克隆的百分数)针对分别如所示的三个删除用构建体和间隔区变体作图。通过菌落PCR用位于基因删除用同源性区域外部的寡核苷酸SEQ ID 095和SEQ ID 096分析vpr基因的基因删除。B. 描述三个删除构建体和间隔区变体分别对于删除地衣芽孢杆菌epr基因的相对删除效率, 作为具有已失活epr基因的克隆相对于总计20个已分析克隆的百分数。通过菌落PCR用位于基因删除用同源性区域外部的寡核苷酸SEQ ID 097和SEQ ID 098分析epr基因的基因删除。

[0187] 图7

[0188] 将PaprE-GFPmut2表达盒替换地衣芽孢杆菌amyB基因的基因整合效率(作为具有

已整合PaprE-GFPmut2表达盒的克隆相对于总计20个已分析克隆的百分数)对分别如所示的两个不同地衣芽孢杆菌菌株Bli#005和P308作图。显示二个独立实验的平均数连同标准差。通过菌落PCR用位于基因整合用同源性区域外部的寡核苷酸SEQ ID 009和SEQ ID 010分析整合。

[0189] 图8

[0190] 将短小芽孢杆菌的孢子形成基因sigE、sigF和spoIIE的基因删除效率(作为具有已失活孢子形成基因的克隆相对于每个已分析孢子形成基因的总计20个克隆的百分数)如所示那样作图。通过菌落PCR分别以位于基因删除用同源性区域外部的寡核苷酸SEQ ID 099和SEQ ID 100、SEQ ID 101和SEQ ID 102以及SEQ ID 103和SEQ ID 104分析sigE、sigF和spoIIE基因的基因删除。

实施例

[0191] 材料和方法

[0192] 以下实施例仅起到说明本发明的作用。本领域技术人员显而易见的众多可能变化也落入本发明的范围内。

[0193] 除非另外声明,否则已经通过应用如基因工程和通过培养微生物来发酵生产化学化合物中使用的标准设备、方法、化学物质和生物化学物质,进行以下实验。还参见 Sambrook等人(Sambrook,J.和Russell,D.W.Molecular cloning.A laboratory manual,第3版,Cold Spring Harbor Laboratory Press,Cold Spring Harbor,NY.2001)和Chmiel等人(Bioprocessstechnik 1.Einführung in die Bioverfahrenstechnik,Gustav Fischer Verlag,Stuttgart,1991)。

[0194] 电感受态地衣芽孢杆菌细胞和电穿孔

[0195] 通过电穿孔法将DNA转化入地衣芽孢杆菌菌株DSM641和ATCC53926。电感受态地衣芽孢杆菌细胞的制备和DNA转化基本上如Brigidi等人(Brigidi,P.,Mateuzzi,D.(1991).Biotechnol.Techniques 5,5)描述那样进行,但具有以下修改:一旦转化DNA,则在1ml LBSPG缓冲液中使细胞恢复并且在37°C温育60分钟(Vehmaanperä J.,1989,FEMS Microbio.Lett.,61:165-170),此后涂布在选择性LB琼脂平板上。

[0196] 为了克服地衣芽孢杆菌菌株DSM641和ATCC53926的地衣芽孢杆菌特有的限制修饰系统,如下文描述那样从Ec#098细胞分离质粒DNA。为了转移入地衣芽孢杆菌限制酶敲除菌株中,从大肠杆菌INV110细胞(Life technologies)分离质粒DNA。

[0197] 电感受态短小芽孢杆菌细胞和电穿孔

[0198] 通过电穿孔法将DNA转化入短小芽孢杆菌DSM14395。如对地衣芽孢杆菌细胞所述那样进行电感受态短小芽孢杆菌DSM14395细胞的制备和DNA转化。

[0199] 为了克服短小芽孢杆菌特有限制性修饰系统,从大肠杆菌DH10B细胞分离质粒DNA并且根据如专利DE4005025中对地衣芽孢杆菌所述那样的方法,用来自短小芽孢杆菌DSM14395的完整细胞提取物使质粒DNA体外甲基化。

[0200] 电感受态枯草芽孢杆菌细胞和电穿孔

[0201] 分别如对地衣芽孢杆菌和短小芽孢杆菌所述那样,通过电穿孔法将DNA转化入枯草芽孢杆菌ATCC6051a。从大肠杆菌DH10B细胞分离的质粒DNA可以轻易地用于转移入枯草

芽孢杆菌中。

[0202] 质粒分离

[0203] 通过 (Sambrook, J. 和 Russell, D.W.《分子克隆实验室指南》,第3版,Cold Spring Harbor Laboratory Press,Cold Spring Harbor,NY.2001) 描述的标准分子生物学方法中或碱性裂解方法 (Birnboim, H.C., Doly, J. (1979). *Nucleic Acids Res* 7 (6) :1513-1523), 从芽孢杆菌细胞和大肠杆菌细胞分离质粒DNA。与大肠杆菌相比,将芽孢杆菌细胞用10mg/ml溶菌酶在37℃处理30分钟,之后裂解细胞。

[0204] 寡核苷酸复性以形成寡核苷酸-双链体。

[0205] 在水中调节寡核苷酸至100μM浓度。添加5μl正向寡核苷酸和5μl相应的反向寡核苷酸至90μl 30mM HEPES-缓冲液 (pH 7.8)。将反应混合物加热到95℃持续5分钟,随后按0.1℃/秒降低温度进行95℃至4℃温变来复性 (Cobb, R.E., Wang, Y. 和 Zhao, H. (2015). *High-Efficiency Multiplex Genome Editing of Streptomyces Species Using an Engineered CRISPR/Cas System* (使用工程化CRISPR/Cas系统的链霉菌属 (*Streptomyces*) 物种高效率多重基因组编辑). *ACS Synthetic Biology*, 4 (6) , 723-728)。

[0206] 分子生物学方法和技术

[0207] 分子生物学中不限于培养芽孢杆菌和大肠杆菌微生物、DNA电穿孔、分离基因组和质粒DNA、PCR反应,克隆技术的标准方法基本上如Sambrook和Russell. (Sambrook, J. and Russell, D.W.《分子克隆实验室指南》,第3版,Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.2001.) 描述那样进行。

[0208] 菌株

[0209] 大肠杆菌菌株Ec#098

[0210] 大肠杆菌菌株Ec#098是一个携带编码DNA甲基转移酶的表达质粒pMDS003的大肠杆菌INV110菌株 (Life technologies) (W02019016051)。

[0211] 生成地衣芽孢杆菌基因k.o菌株

[0212] 为了在地衣芽孢杆菌菌株DSM641和ATCC53926 (US5352604) 及其衍生物中进行基因删除,将删除质粒转化入根据Chung方法 (Chung, C.T., Niemela, S.L. 和 Miller, R.H. (1989). *One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution* (一步法制备感受态大肠杆菌: 相同溶液中转化和储存细菌细胞). *Proc. Natl. Acad. Sci. U.S.A* 86, 2172-2175) 成为感受态的大肠杆菌菌株Ec#098中,随后在37℃于含有100μg/ml氨苄青霉素和30μg/ml氯霉素的LB琼脂平板上选择。将质粒DNA从独立的克隆分离并且用于后续转化入地衣芽孢杆菌菌株中。分离的质粒DNA分别携带地衣芽孢杆菌菌株DSM641和ATCC53926的DNA甲基化模式并且转移入地衣芽孢杆菌时受到保护免于降解。

[0213] 地衣芽孢杆菌P304:删除限制性核酸内切酶

[0214] 将电感受态地衣芽孢杆菌DSM641细胞 (US5352604) 如上述制备并且用1μg从大肠杆菌Ec#098分离的pDe1006限制酶基因删除质粒,随后在30℃铺种于含有5μg/ml红霉素的LB琼脂平板上。

[0215] 基因删除过程如以下所述进行:

[0216] 将携带质粒的地衣芽孢杆菌细胞在45℃于含有5μg/ml红霉素的LB琼脂平板上培

育,从而借助Campbell重组过程驱动该删除质粒整合入染色体,同时pDe1006的同源性区域之一与aprE基因5'或3'的序列同源。挑出克隆并且在45℃于LB培养基中无选择压力情况下培育6小时,接着在30℃铺种在含有5μg/ml红霉素的LB琼脂平板上。挑出独立克隆并且通过菌落PCR分析以寡核苷酸SEQ ID 014和SEQ ID 015筛选限制酶基因的基因组成功删除者。挑出推定性删除阳性独立克隆并且在45℃于无抗生素的LB培养基中经历连续两次过夜孵育,以消除质粒,并且在37℃铺种于LB琼脂平板上孵育过夜。通过群体PCR对单一克隆分析限制酶基因的基因组成功删除。分离出正确删除限制酶基因的红霉素敏感性单一克隆并且命名为地衣芽孢杆菌P304。

[0217] 地衣芽孢杆菌P308:删除聚-γ谷氨酸合成基因

[0218] 将电感受态地衣芽孢杆菌P304细胞如上述制备并且用1μg从大肠杆菌INV110细胞(Life technologies)分离的pDe1007 pga基因删除质粒转化,随后在30℃铺种于含有5μg/ml红霉素的LB琼脂平板上。

[0219] 基因删除过程如删除限制酶基因所述那样进行。

[0220] 通过PCR以寡核苷酸SEQ ID 017和SEQ ID 018分析pga基因的删除。pga合成基因经删除的所得地衣芽孢杆菌菌株命名为地衣芽孢杆菌P308。

[0221] 地衣芽孢杆菌Bli#002:删除aprE基因

[0222] 将电感受态地衣芽孢杆菌ATCC53926细胞如上述制备并且用1μg从大肠杆菌Ec#098分离的pDe1003 aprE基因删除质粒,随后在30℃铺种于含有5μg/ml红霉素的LB琼脂平板上。

[0223] 基因删除过程如删除限制酶基因所述那样进行。通过PCR以寡核苷酸SEQ ID 020和SEQ ID 021分析aprE基因的删除。aprE基因经删除的所得地衣芽孢杆菌菌株命名为Bli#002。

[0224] 地衣芽孢杆菌Bli#005:删除聚-γ谷氨酸合成基因

[0225] 在地衣芽孢杆菌Bli#002中如地对地衣芽孢杆菌P304中删除pga基因所述那样删除聚γ-谷氨酸合成基因,区别在于pDe1007质粒分离自大肠杆菌Ec#098细胞。所得的菌株命名为Bli#005。

[0226] 质粒

[0227] pEC194RS-芽孢杆菌温度敏感性删除质粒。

[0228] 将质粒pE194用带旁侧PvuII位点的寡核苷酸SEQ ID 001和SEQ ID 002进行PCR扩增,用限制性核酸内切酶PvuII消化并且连接入限制性酶SmaI消化的载体pCE1中。pCE1是其中已经通过沉默突变移除氨苄青霉素耐药基因内部BsaI位点的pUC18衍生物。将连接混合物转化入大肠杆菌DH10B细胞(Life technologies)。将转化体铺展并在37℃于含有100μg/ml氨苄青霉素的LB琼脂平板上孵育过夜。从独立的克隆分离质粒DNA并且通过限制性消化分析正确度。所得的质粒命名为pEC194S。

[0229] 用寡核苷酸SEQ ID 003和SEQ ID 004从质粒pBSd141R(登记号:KY995200)(Radeck,J.,Meyer,D.,Lautenschlager,N.和Mascher,T.2017.Bacillus SEVA siblings: A Golden Gate-based toolbox to create personalized integrative vectors for Bacillus subtilis(芽孢杆菌SEVA同胞:基于Golden Gate的工具套件产生用于枯草芽孢杆菌的个性化整合性载体).Sci.Rep.7:14134)中PCR扩增出II型组装mRFP盒,其包含额外

的核苷酸限制位点BamHI。将PCR片段和pEC194S用限制性酶BamHI限制处理,随后连接并转化入大肠杆菌DH10B细胞(Life technologies)。将转化体铺展并在37°C于含有100μg/ml氨苄青霉素的LB琼脂平板上孵育过夜。从独立的克隆分离质粒DNA并且通过限制性消化分析正确度。所得的质粒pEC194RS携带可读框与红霉素耐药基因的可读框对头存在的mRFP盒。

[0230] pDe1003-aprE基因删除质粒

[0231] 将地衣芽孢杆菌aprE基因的基因删除质粒用质粒pEC194RS和包含aprE基因5'和3'的基因组区域的基因合成构建体SEQ ID 019构建,所述基因组区域旁侧有相容于pEC194RS的BsaI位点。采用限制性核酸内切酶BsaI的II型组装如所述那样进行(Radeck, J., Meyer, D., Lautenschlager, N. 和 Mascher, T. 2017. Bacillus SEVA siblings: A Golden Gate-based toolbox to create personalized integrative vectors for Bacillus subtilis (芽孢杆菌SEVA同胞: 基于Golden Gate的工具套件产生用于枯草芽孢杆菌的个性化整合性载体). Sci. Rep. 7: 14134)并且将反应混合物随后转化入大肠杆菌DH10B细胞(Life technologies)。将转化体铺展并在37°C于含有100μg/ml氨苄青霉素的LB琼脂平板上孵育过夜。从独立的克隆分离质粒DNA并且通过限制性消化分析正确度。所得的aprE删除质粒命名为pDe1003。

[0232] pDe1006-限制酶基因删除质粒

[0233] 将用于地衣芽孢杆菌DSM641限制修饰系统(SEQ ID 011)中限制酶基因(SEQ ID 012)的基因删除质粒用质粒pEC194RS和包含限制酶基因5'和3'的基因组区域的基因合成构建体SEQ ID 013构建,所述基因组区域旁侧有相容于pEC194RS的BsaI位点。采用限制性核酸内切酶BsaI的II型组装如上文所述那样进行并且随后将反应混合物转化入大肠杆菌DH10B细胞(Life technologies)。将转化体铺展并在37°C于含有100μg/ml氨苄青霉素的LB琼脂平板上孵育过夜。从独立的克隆分离质粒DNA并且通过限制性消化分析正确度。所得的限制酶删除质粒命名为pDe1006。

[0234] pDe1007-聚-γ-谷氨酸合成基因删除质粒

[0235] 对如pDe1006所述产生用于删除涉及地衣芽孢杆菌聚γ-谷氨酸盐(pga)产生的基因即ywsC (pgsB)、ywtA (pgsC)、ywtB (pgsA)、ywtC (pgsE)的删除质粒,然而使用基因合成构建体SEQ ID 016,其包含在ywsC、ywtA (pgsC)、ywtB (pgsA)、ywtC (pgsE)基因5'和3'侧置的基因组区域,所述基因组区域旁侧有相容于pEC194RS的BsaI位点。所得的pga删除质粒命名为pDe1007。

[0236] 质粒p689-T2A-lac

[0237] 质粒p689-T2A-lac包含旁侧分布有BpiI限制位点的lacZ-α基因,所述限制位点又在5'旁侧分布有大肠杆菌rrnB基因的T1终止子及在3'旁侧分布有T0λ终止子,并且作为基因合成构建体(SEQ ID 073)订购。

[0238] 质粒p890 PaprE-GFPmut2

[0239] 将质粒pCB56C的地衣芽孢杆菌aprE基因的启动子(US5352604)用寡核苷酸SEQ ID074和SEQ ID 075进行PCR扩增。带旁侧BpiI限制位点的GFPmut2基因变体(登录号AF302837)(SEQ ID 076)作为基因合成片段订购(Geneart雷根斯堡)。将融合于GFPmut2变体的包含地衣芽孢杆菌源PaprE启动子的基因表达构建体通过采用限制性核酸内切酶BpiI的II型组装如所述那样(Radeck, J., Meyer, D., Lautenschlager, N. 和 Mascher,

T.2017. Bacillus SEVA siblings: A Golden Gate-based toolbox to create personalized integrative vectors for Bacillus subtilis (芽孢杆菌SEVA同胞: 基于Golden Gate的工具套件产生用于枯草芽孢杆菌的个性化整合性载体). Sci. Rep. 7:14134) 克隆入质粒p689-T2A-lac并且随后将反应混合物转化入电感受态大肠杆菌DH10B细胞。将转化体铺展并在37°C于含有100µg/ml氨苄青霉素的LB琼脂平板上孵育过夜。从独立的克隆分离质粒DNA并且通过限制性消化分析正确度并且测序。所得的质粒命名为p890 PaprE-GFPmut2。

[0240] 质粒pJOE8999.1:

[0241] Altenbuchner J.2016. Editing of the Bacillus subtilis genome by the CRISPR-Cas9 system (通过CRISPR-Cas9系统编辑枯草芽孢杆菌基因组). Appl Environ Microbiol 82:5421-5.

[0242] 质粒pJOE-T2A

[0243] 为了允许基于II型组装(T2A)一步克隆用于DSB修复的sgRNA和同源性区域,如下修饰CRISPR/Cas9质粒pJOE8889.1。来自质粒pBSd141R(登记号:KY995200)(Radeck, J., Meyer, D., Lautenschlager, N.和Mascher, T.2017. Bacillus SEVA siblings: A Golden Gate-based toolbox to create personalized integrative vectors for Bacillus subtilis (芽孢杆菌SEVA同胞: 基于Golden Gate的工具套件产生用于枯草芽孢杆菌的个性化整合性载体). Sci. Rep. 7:14134)的II型组装mRFP盒经如此修饰,从而移除多个限制位点和BpiI限制位点,并且作为带旁侧SfiI限制位点的基因合成片段(SEQ ID 005)订购。质粒命名为p#732。将质粒p#732和质粒pJOE8999.1用SfiI(New England Biolabs, NEB)消化并且将p#732的mRFP盒连接入SfiI消化的pJOE8999.1,随后转化入感受态大肠杆菌DH10B细胞。在含有IPTG/X-Gal和卡那霉素(20µg/ml)的LB琼脂平板上筛选阳性克隆的紫色菌落(蓝-白色筛选及mRFP1表达)。所得的已验证序列的质粒命名为pJOE-T2A。

[0244] 质粒pBW732

[0245] 与地衣芽孢杆菌DSM641的淀粉酶amyB基因毗邻的5'同源性区域(也称作HomA)和3'同源性区域(也称作HomB)作为带旁侧XmaI限制位点的合成性基因合成片段(SEQ ID006)订购。将质粒pJOE8999.1和合成性amyB-HomAB片段用限制性核酸内切酶XmaI切割,随后用T4-DNA连接酶(NEB)连接并转化入电感受态大肠杆菌DH10B细胞。回收正确的质粒并命名为pBW732。

[0246] 质粒pBW742

[0247] 使用Geneious 11.1.5(<https://www.geneious.com>)设计针对sgRNA的amyB基因20bp靶序列。使5'磷酸化的所得寡核苷酸SEQ ID 007和SEQ ID 008复性以形成寡核苷酸双链体。通过采用限制性核酸内切酶BsaI的II型组装如所述那样(Radeck, J., Meyer, D., Lautenschlager, N.和Mascher, T.2017. Bacillus SEVA siblings: A Golden Gate-based toolbox to create personalized integrative vectors for Bacillus subtilis (芽孢杆菌SEVA同胞: 基于Golden Gate的工具套件产生用于枯草芽孢杆菌的个性化整合性载体). Sci. Rep. 7:14134)用以下组件:pBW732和寡核苷酸双链体(SEQ ID 007、SEQ ID 008)构建基于CRISPR/Cas9的针对地衣芽孢杆菌amyB基因的基因删除质粒。将反应混合物转化入大肠杆菌DH10B细胞(Life technologies)。将转化体铺展在含有20µg/ml卡那霉素的LB

琼脂平板上并在37℃温育过夜。从独立的克隆分离质粒DNA并且通过限制性消化分析正确度并且测序。所得的amyB删除质粒命名为pBW742。

[0248] T2A CRISPR目的载体pCC027和pCC028

[0249] 将质粒pCC014和pCC025如此修饰,从而覆盖间隔区-sgRNA的区域和侧置于同源区的amyB基因为来自质粒pJOE-T2A的T2A盒所替换。将pCC014和pCC025的主链用寡核苷酸SEQ ID 050和SEQ ID 051进行PCR扩增并且用寡核苷酸SEQ ID 048和SEQ ID 049从pJOE-T2A中PCR扩增出T2A组装机,接着使用High Pure PCR纯化试剂盒进行PCR纯化,用DpnI消化并进行凝胶纯化。相应的主链PCR片段和T2A盒PCR片段在10μl Gibson反应中复性,随后转化入大肠杆菌DH10B细胞(Life technologies)。将转化体铺展在含有20μg/ml卡那霉素的LB琼脂平板上并在37℃温育过夜。从独立的克隆分离质粒DNA并且通过限制性消化分析正确度并且测序。所得的pCC014衍生和pCC025衍生的T2A质粒衍生物分别命名为pCC027和pCC028。

[0250] pCC029-hag基因删除质粒

[0251] 使用如先前所述的Geneious 11.1.5设计针对sgRNA的hag基因20bp靶序列。使5'磷酸化的所得寡核苷酸SEQ ID 056和SEQ ID 057如复性以形成上文所述的寡核苷酸双链体。用寡核苷酸SEQ ID 054和Seq ID 053及SEQ ID 052和Seq ID 55在地衣芽孢杆菌DSM641的基因组DNA上PCR扩增hag基因5'和3'的基因组区域,随后通过重叠延伸PCR与侧翼分布的寡核苷酸SEQ ID 053和SEQ ID 054融合。将所得的PCR产物做柱纯化(Qiagen PCR纯化试剂盒)。通过采用限制性核酸内切酶BsaI的II型组装如前述那样用以下组件:质粒pCC027(PV4-5启动子变体)、带旁侧BsaI限制位点的hag基因已融合同源性区域和寡核苷酸双链体(SEQ ID 056、SEQ ID 057)构建基于CRISPR/Cas9的针对地衣芽孢杆菌hag基因的基因删除质粒。将反应混合物转化入大肠杆菌DH10B细胞(Life technologies)。将转化体铺展在含有20μg/ml卡那霉素的LB琼脂平板上并在37℃温育过夜。从独立的克隆分离质粒DNA并且通过限制性消化分析正确度并且测序。所得的hag基因删除质粒命名为pCC029。

[0252] pCC030-hag基因删除质粒

[0253] 将hag基因删除构建体如pCC029那样构建,然而使用质粒pCC028(PV8-7启动子变体)。

[0254] pCC031-degU32基因编辑质粒

[0255] 如对pCC029那样构建degU32基因组编辑构建体以引入degU H12L突变,附带如下修改。

[0256] 针对degU H12L突变引入突变以及引入沉默性点突变以移除PAM位点的degU32同源性区域作为带旁侧BsaI位点的基因合成构建体(SEQ ID 058)订购(Genart,雷根斯堡)。如先前所述设计针对sgRNA的degU基因20bp靶序列并且使5'磷酸化的所得寡核苷酸SEQ ID 059和SEQ ID 060复性以形成寡核苷酸双链体。

[0257] pCC032-degU32基因编辑质粒

[0258] 如对pCC031所述那样产生degU32基因组编辑构建体,然而使用质粒pCC028(PV8-7启动子变体)。

[0259] pCC033-amyE基因删除质粒

[0260] 将包含amyE间隔区-sgRNA及枯草芽孢杆菌amyE基因5'区和3'区的同源性区域的片段从质粒pCC004(W017186550)用带旁侧BsaI限制位点的寡核苷酸SEQ ID 061和SEQ ID

062进行PCR扩增。随后通过采用限制性核酸内切酶BsaI的II型组装如上文所述那样以质粒pCC027 (PV4-5启动子变体) 和PCR扩增片段构建基于CRISPR/Cas9的针对淀粉酶amyE基因的基因删除质粒。将反应混合物转化入大肠杆菌DH10B细胞 (Life technologies)。将转化体铺展在含有20 μ g/ml卡那霉素的LB琼脂平板上并在37 $^{\circ}$ C温育过夜。从独立的克隆分离质粒DNA并且通过限制性消化分析正确度并且测序。所得的amyE基因删除质粒命名为pCC033。

[0261] pCC034-amyE基因删除质粒

[0262] 将amyE基因删除构建体如pCC033那样构建, 然而使用质粒pCC028 (PV8-7启动子变体)。

[0263] pCC035-aprE基因删除质粒

[0264] 将包含aprE间隔区 (SEQ ID 064) -sgRNA及枯草芽孢杆菌aprE基因5'区和3'区的同源性区域的片段作为带旁侧BsaI限制位点的合成性基因片段 (SEQ ID 063) 订购。随后通过采用限制性核酸内切酶BsaI的II型组装如上文所述那样以质粒pCC027 (PV4-5启动子变体) 和基因合成构建体构建基于CRISPR/Cas9的针对蛋白酶aprE基因的基因删除质粒。将反应混合物转化入大肠杆菌DH10B细胞 (Life technologies)。将转化体铺展在含有20 μ g/ml卡那霉素的LB琼脂平板上并在37 $^{\circ}$ C温育过夜。从独立的克隆分离质粒DNA并且通过限制性消化分析正确度并且测序。所得的aprE基因删除质粒命名为pCC035。

[0265] pCC036-aprE基因删除质粒

[0266] 将aprE基因删除构建体如pCC035那样构建, 然而使用质粒pCC028 (PV8-7启动子变体)。

[0267] pCC037-pCC039-vpr基因删除质粒

[0268] 将地衣芽孢杆菌蛋白酶vpr基因的CRISPR/Cas9基因删除构建体pCC037、pCC038和pCC039如对pCC035所述那样构建, 然而用包含vpr间隔区-sgRNA以及vpr基因5'区和3'区的同源性区域的合成性基因片段 (SEQ ID 065) 构建。所得的质粒pCC037、pCC038和pCC039在SEQ ID 065内部的vpr间隔序列 (SEQ ID 066、SEQ ID 067、SEQ ID 068) 方面有差异。

[0269] pCC040-pCC042-epr基因删除质粒

[0270] 将地衣芽孢杆菌蛋白酶epr基因的CRISPR/Cas9基因删除构建体pCC040、pCC041和pCC042如对pCC035所述那样构建, 然而用包含epr间隔区-sgRNA以及epr基因5'区和3'区的同源性区域的合成性基因片段 (SEQ ID 069) 构建。所得的质粒pCC040、pCC041和pCC042在SEQ ID 069内部的vpr间隔序列 (SEQ ID 070、SEQ ID 071、SEQ ID 072) 方面有差异。

[0271] pCC043-GFP基因整合质粒

[0272] 将针对sgRNA的amyB基因20bp靶序列作为作为5'磷酸化的寡核苷酸SEQ ID 007和Seq ID 008订购, 随后进行复性以形成寡核苷酸双链体。将地衣芽孢杆菌amyB基因的5'区和3'区域分别用寡核苷酸SEQ ID 077和SEQ ID 078及SEQ ID 079和SEQ ID 080进行PCR扩增。

[0273] 通过采用限制性核酸内切酶BsaI的II型组装如上述那样用以下组件: pCC027、寡核苷酸双链体 (SEQ ID 007、SEQ ID 008)、amyB基因的5'同源性区域的PCR片段、p890-PaprE-GFPmut2和amyB基因的3'同源性区域的PCR片段构建替换地衣芽孢杆菌amyB基因的基于CRISPR/Cas9的基因整合质粒构。将反应混合物转化入大肠杆菌DH10B细胞 (Life technologies)。将转化体铺展在含有20 μ g/ml卡那霉素的LB琼脂平板上并在37 $^{\circ}$ C温育过

夜。从独立的克隆分离质粒DNA并且通过限制性消化分析正确度并且测序。所得的基于CRISPR/Cas9的基因整合质粒命名为pCC043。

[0274] pCC044-短小芽孢杆菌sigE基因删除质粒

[0275] 将短小芽孢杆菌DSM14395的sigE基因的CRISPR/Cas9基因删除构建体pCC044如对pCC035所述那样构建,然而采用包含sigE间隔区(SEQ ID 081)-sgRNA和sigE基因5'区和3'区的同源性区域的合成性基因片段(SEQ ID 082)构建。

[0276] pCC045-短小芽孢杆菌sigF基因删除质粒

[0277] 将短小芽孢杆菌DSM14395的sigF基因的CRISPR/Cas9基因删除构建体pCC045如对pCC035所述那样构建,然而采用包含sigF间隔区(SEQ ID 083)-sgRNA和sigF基因5'区和3'区的同源性区域的合成性基因片段(SEQ ID 084)构建。

[0278] pCC046-短小芽孢杆菌spoIIE基因删除质粒

[0279] 将短小芽孢杆菌DSM14395的spoIIE基因的CRISPR/Cas9基因删除构建体pCC046如对pCC035所述那样构建,然而采用包含spoIIE间隔区(SEQ ID 085)-sgRNA和spoIIE基因5'区和3'区的同源性区域的合成性基因片段(SEQ ID 086)构建。

[0280] 实施例1:构建携带组成型启动子的CRISPR/Cas9基因组编辑质粒为了引入在质粒pBW742中驱动Cas9酶表达的组成型启动子,采用一个两步骤方法。

[0281] 首先,将t1t2t0终止子(衍生自pMUTIN)引入pBW742的启动子PmanP的5',以防止来自卡那霉素选择标记的潜在通读。

[0282] 通过Gibson装配(NEBuilder®HiFi DNA装配克隆试剂盒,New England Biolabs),使终止子序列t1t2t0在甘露糖启动子上游整合入pBW742。为此目的,使用pMutin2(登记号AF072806)作为模板,以寡核苷酸SEQ ID 024和SEQ ID 025通过PCR扩增终止子片段(0.44kb)。以寡核苷酸SEQ ID 022和SEQ ID 023扩增pBW742的相应载体主链。使用PCR产物纯化试剂盒(Roche)纯化pBW742扩增子。后续用DpnI(New England Biolabs)消化pBW742 PCR产物后,将两种PCR片段均使用Qia快速凝胶提取试剂盒(Qiagen,希尔登,德国)进行凝胶纯化并且按1:2比率在50°C复性1小时。将大肠杆菌菌株DH10B用装配反应物转化,接着铺种在含有20µg/ml卡那霉素的LB琼脂平板上。从独立的克隆分离质粒DNA并且通过限制性消化分析正确度并且测序。

[0283] 存在来自自己公布pMutin2参考序列的偏离。SEQ ID 026覆盖pMutin2序列的一部分,SEQ ID 027覆盖在所得质粒pCC009中存在的pMutin2的相应区域内存在的序列偏离。

[0284] 其次,甘露糖诱导型启动子PmanP为枯草芽孢杆菌组成型启动子Pveg的两个启动子变体-即衍生自Guiziou等人的PV4和PV8(Guiziou, S., V. Sauveplane, H. J. Chang, C. Clerte, N. Declerck, M. Jules和J. Bonnet. 2016. A part toolbox to tune genetic expression in *Bacillus subtilis* (微调枯草芽孢杆菌中遗传表达的部分工具套件). *Nucleic Acids Res.* 44:7495-7508)所交换。这些启动子变体包含衍生自己改编Pveg启动子文库的Pveg启动子、标准化TSS(转录起始位点)区和标准化核糖体结合位点区域R0,其中相对于所述启动子变体的已改变表达水平就枯草芽孢杆菌中单拷贝水平筛选所述启动子文库。启动子序列PV4和PV8分别作为SEQ ID 028和SEQ ID 029列出。

[0285] 两个启动子变体的整合均通过Gibson装配来实施。逐步扩增PV4片段和PV8片段。对于启动子片段,使用pCC009作为模板,将寡核苷酸SEQ ID 024和SEQ ID 030用于第一PCR

(Phusion高保真DNA聚合酶-NEB)并且所得产物充当第二PCR的模板,寡核苷酸SEQ ID 024和SEQ ID 031针对PV4并且SEQ ID 024和SEQ ID 033针对PV8。

[0286] 使用寡核苷酸SEQ ID 022和SEQ ID 032,对pCC009的载体主链进行PCR扩增。以PCR纯化试剂盒 (Roche) 纯化载体扩增子后,用DpnI消化PCR产物,以从PCR反应移除剩余的环状质粒DNA。随后,使用Qiaquick凝胶提取试剂盒 (Qiagen, 希尔登, 德国) 纯化已消化的载体和两个启动子片段。随后使载体扩增子pCC009分别与启动子片段PV4和PV8复性,因而用Pveg启动子的PV4变体和PV8变体替换甘露糖启动子PmanP。

[0287] 随后将复性反应物转化入大肠杆菌DH10B细胞 (Life technologies)。将转化体铺展在含有20 μ g/ml卡那霉素的LB琼脂平板上并在37 $^{\circ}$ C温育过夜。质粒DNA从PV4启动子的9个体克隆分离及从源自启动子变体PV8的8独立克隆分离并通过测序分析正确度。

[0288] 表1总结了各种启动子变体的测序结果:

[0289] 分析来自PV4-克隆反应的克隆揭示,仅PV4区内部有点突变、核苷酸插入或删除的序列可以恢复。

[0290] 分析来自PV8-克隆反应的克隆揭示,仅PV8区内部有点突变、核苷酸插入或删除的序列可以恢复。所得的质粒总结于表1中。

[0291] 表1

[0292]

质粒	启动子变体	SEQ ID
pCC010	Pv4-1	034
pCC011	Pv4-2	035
pCC012	Pv4-3	036
pCC013	Pv4-4	036
pCC014	Pv4-5	037
pCC015	Pv4-6	038
pCC016	Pv4-7	039
pCC017	Pv4-8	040
pCC018	Pv4-9	039

[0293]	pCC019	Pv8-1	041
	pCC020	Pv8-2	042
	pCC021	Pv8-3	043
	pCC022	Pv8-4	044
	pCC023	Pv8-5	045
	pCC024	Pv8-6	041
	pCC025	Pv8-7	046
	pCC026	Pv8-8	047

[0294] 基于CRISPR/Cas9的删除质粒的基因删除效率

[0295] 将电感受态地衣芽孢杆菌P308细胞如上述制备并且用1 μ g从大肠杆菌INV110细胞(Life technologies)分离的amyB删除质粒pCC010-012、pCC014-017、pCC019-026(带有如表1中所述的不同启动子变体)转化,随后铺种于含有20 μ g/ml卡那霉素的LB琼脂平板上并且在37 $^{\circ}$ C孵育过夜。

[0296] 次日,使每个转化反应的20个克隆经历采用寡核苷酸SEQ ID 009和SEQ ID 010的菌落PCR,以分析基于CRISPR/Cas9的amyB基因成功删除,并且进一步转移到无抗生素的新鲜LB琼脂平板上,随后在48 $^{\circ}$ C孵育过夜以进行质粒消除。

[0297] 基于相比野生型amyB基因座的更大特异性PCR扩增子出现预期的更小特异性PCR扩增子,将每种基于CRISPR/Cas9的删除质粒的amyB基因删除效率计算为相对于已分析克隆的总数以百分数计的基因成功删除比率。

[0298] 如图3中所示,基于CRISPR/Cas9的amyB基因删除质粒pCC010、pCC019和pCC022在地衣芽孢杆菌中不像携带野生型amyB基因座的全部已分析细胞那样有功能。

[0299] 其他启动子变体在地衣芽孢杆菌中有功能,驱动Cas9表达。尤其,分别具有启动子变体PV4-5、PV4-7和PV8-7的基因删除质粒pCC014、pCC016、pCC025显示最高的基因删除效率,大于60%。

[0300] 将正确的单一克隆划线到无抗生素的新鲜LB琼脂平板上,随后在48 $^{\circ}$ C第二次孵育过夜以进行质粒消除。再次通过菌落PCR对最终克隆分析amyB基因成功删除并且通过铺种在含有20 μ g/ml卡那霉素的LB琼脂平板上分析质粒丢失。将消除了删除质粒(对卡那霉素敏感)并删除了amyB基因的所得地衣芽孢杆菌菌株命名为地衣芽孢杆菌P310。

[0301] 实施例2:地衣芽孢杆菌中采用启动子PV4-5和PV8-7的基因删除和基因突变

[0302] 将电感受态地衣芽孢杆菌P308细胞如上述制备并且用1 μ g从大肠杆菌INV110细胞(Life technologies)分离的分别具有启动子PV4-5(SEQ ID 037)和PV8-7(SEQ ID 046)的hag删除质粒pCC029和pCC030每一者转化,随后铺种于含有20 μ g/ml卡那霉素的LB琼脂平板上并且在37 $^{\circ}$ C孵育过夜。

[0303] 次日,使每个转化反应的20个克隆经历采用寡核苷酸SEQ ID 087和SEQ ID 088的菌落PCR,以分析基于CRISPR/Cas9的hag基因成功删除,并且进一步转移到无抗生素的新鲜

LB琼脂平板上,随后在48℃孵育过夜以进行质粒消除。

[0304] 基于相比野生型hag基因座的更大特异性PCR扩增子出现预期的更小特异性PCR扩增子,将每种基于CRISPR/Cas9的删除质粒的hag基因删除效率计算为相对于已分析克隆的总数以百分数计的基因成功删除比率。对每种hag基因删除质粒进行三次实验。如图4A中所示,质粒pCC029和pCC030的基于CRISPR/Cas9的基因删除效率分别是95%和100%。

[0305] 为了分析点突变引入效率,将地衣芽孢杆菌P308细胞用两种degU突变质粒pCC031和pCC032如对删除hag基因所述那样转化,所述突变质粒再次在驱动Cas9组成型表达的启动子PV4-5 (SEQ ID 037) 和PV8-7 (SEQ ID 046) 方面差异。将转化的地衣芽孢杆菌细胞铺种在含有20μg/ml卡那霉素的LB琼脂平板上,接着在30℃孵育过夜。将引入H12L degU突变的突变效率计算为相对于总数20个已分析克隆,基于采用寡核苷酸SEQ ID 089和SEQ ID 090时出现与野生型degU基因座的天然degU特异性PCR-扩增子相比可以用限制性核酸内切酶PstI切割的degU特异性PCR-扩增子,以百分数计成功突变的degU基因的比率。对每种degU基因删除质粒进行三次实验。如图4B中所示,质粒pCC031和pCC032的基于CRISPR/Cas9的突变效率分别是19%和24%。

[0306] 实施例3:枯草芽孢杆菌中采用启动子PV4-5和PV8-7的基因删除

[0307] 将电感受态枯草芽孢杆菌ATCC6051a细胞如上述制备并且用1μg从大肠杆菌DH10B细胞分离的分别具有启动子PV4-5 (SEQ ID 037) 和PV8-7 (SEQ ID 046) 的amyE删除质粒pCC033和pCC034每一者转化,随后铺种于含有20μg/ml卡那霉素的LB琼脂平板上并且在37℃孵育过夜。

[0308] 次日,使每个转化反应的20个克隆经历采用寡核苷酸SEQ ID 091和SEQ ID 092的菌落PCR,以分析基于CRISPR/Cas9的amyE基因成功删除,并且进一步转移到无抗生素的新鲜LB琼脂平板上,随后在48℃孵育过夜以进行质粒消除。

[0309] 基于相比野生型amyE基因座的更大特异性PCR扩增子出现预期的更小特异性PCR扩增子,将每种基于CRISPR/Cas9的删除质粒的amyE基因删除效率计算为相对于已分析克隆的总数以百分数计的基因成功删除比率。对每种hag基因删除质粒进行三次实验。如图5A中所示,质粒pCC033和pCC034在枯草芽孢杆菌内部的基于CRISPR/Cas9的amyE基因删除效率分别是97%和100%。

[0310] 类似于对删除amyE基因所描述的方法,分析质粒pCC035和pCC036依赖启动子PV4-5 (SEQ ID 037) 和PV8-7 (SEQ ID 046) 删除枯草芽孢杆菌aprE基因的基因删除效率,然而细胞在含有20μg/ml卡那霉素的LB琼脂平板上孵育,随后在30℃转化过夜。用寡核苷酸SED ID 093和SEQ ID 094通过菌落PCR再次分析基因删除并且如上文对三个独立转化反应所述那样计算基因删除效率。如图5B中所示,质粒pCC035和pCC036在枯草芽孢杆菌内部的基于CRISPR/Cas9的aprE基因删除效率分别是32%和47%。

[0311] 实施例4:地衣芽孢杆菌中采用启动子PV4-5和PV8-7和不同间隔区的基因删除

[0312] 将电感受态地衣芽孢杆菌Bli#005细胞如上述制备并且用1μg从大肠杆菌Ec#098细胞分离的分别具有启动子PV4-5 (SEQ ID 037) 和不同vpr特有间隔序列 (SEQ ID 066-068) 的vpr删除质粒pCC037、pCC038和pCC039每一者转化,随后铺种于含有20μg/ml卡那霉素的LB琼脂平板上并且在37℃孵育过夜。

[0313] 次日,使每个转化反应的20个克隆经历采用寡核苷酸SEQ ID 095和SEQ ID 096的

菌落PCR,以分析基于CRISPR/Cas9的vpr基因成功删除,并且进一步转移到无抗生素的新鲜LB琼脂平板上,随后在48℃孵育过夜以进行质粒消除。

[0314] 基于相比野生型vpr基因座的更大特异性PCR扩增子出现预期的更小特异性PCR扩增子,将每种基于CRISPR/Cas9的删除质粒的vpr基因删除效率计算为相对于已分析克隆的总数以百分数计的基因成功删除比率。如图6A中所示,质粒pCC037、pCC038和pCC039的基于CRISPR/Cas9的vpr基因删除效率分别是100%、100%和84%。

[0315] 如对vpr基因所述那样确定具有启动子PV4-5 (SEQ ID 037) 和不同epr特有间隔序列 (SEQ ID 070-072) 的质粒pCC040、pCC041和pCC042删除地衣芽孢杆菌epr基因的基因删除效率,然而,寡核苷酸SEQ ID 097和SEQ ID 098用于基于菌落PCR的基因删除分析。如图6B中所示,质粒pCC040、pCC041和pCC042的基于CRISPR/Cas9的epr基因删除效率分别是87.5%、100%和100%。

[0316] 实施例5:地衣芽孢杆菌中采用启动子PV4-5和PV8-7的基因整合

[0317] 将电感受态地衣芽孢杆菌Bli#005细胞如上述制备并且用1μg从大肠杆菌Ec#098细胞分离的具有启动子PV4-5 (SEQ ID 037) 的基因整合质粒pCC043转化,随后铺种于含有20μg/ml卡那霉素的LB琼脂平板上并且在37℃孵育过夜。

[0318] 次日,使转化反应的20个克隆经历采用寡核苷酸SEQ ID 009和SEQ ID 010的菌落PCR,以分析基于CRISPR/Cas9的PaprE-GFPmut2表达盒成功整合来替换地衣芽孢杆菌amyB基因,并且进一步转移到无抗生素的新鲜LB琼脂平板上,随后在48℃孵育过夜以进行质粒消除。

[0319] 基于相比野生型amyB基因座的更大特异性PCR扩增子出现预期的特异性PCR扩增子,将基于CRISPR/Cas9的基因整合质粒pCC043的基因整合效率计算为相对于已分析克隆的总数以百分数计的基因成功整合比率。实验进行两次。如图7中所示,质粒pCC043基于CRISPR/Cas9进入Bli#005的基因整合效率是67%。

[0320] 按照与地衣芽孢杆菌P308菌株类似的方式确定采用质粒pCC043时PaprE-GFPmut2表达盒的基因整合效率,显示在两个独立的转化反应中72%的平均基因整合效率,如图7中所述。

[0321] 实施例6:短小芽孢杆菌中采用启动子PV4-5的基因删除

[0322] 将电感受态短小芽孢杆菌DSM14395细胞如上述制备并且用1μg具有驱动Cas9核酸内切酶表达的启动子PV4-5 (SEQ ID 037) 的孢子形成基因删除质粒pCC044 (sigE)、pCC045 (sigF) 和pCC046 (spoIIE) 转化。将质粒DNA从大肠杆菌DH10B细胞分离并且如上文所述那样体外甲基化,之后转化。将转化的短小芽孢杆菌细胞铺种在含有20μg/ml卡那霉素的LB琼脂平板上并在37℃孵育过夜。

[0323] 次日,使得每个转化反应的20个克隆经历菌落PCR,以寡核苷酸SEQ ID 099和SEQ ID 100用于分析sigE删除,以寡核苷酸SEQ ID 101and SEQ ID 102用于sigF删除并且以寡核苷酸SEQ ID 103and SEQ ID 104用于分析spoIIE删除。将单菌落进一步转移到无抗生素的新鲜LB琼脂平板上,随后在48℃孵育过夜以进行质粒消除。

[0324] 基于相比野生型基因座的更大特异性PCR扩增子出现预期的更小特异性PCR扩增子,将质粒pCC044、pCC045和pCC046在短小芽孢杆菌中的基因删除效率计算为相对于已分析克隆的总数以百分数计的基因成功删除比率。如图8中所示,质粒pCC044、pCC045和

pCC046在短小芽孢杆菌内部的基于CRISPR/Cas9的基因删除效率分别是43%、56%和50%。

序列表

<110> 巴斯夫欧洲公司	
<120> 用于产生赋予低至中等表达的组成型细菌启动子的方法	
<130> 191606W001	
<160> 104	
<170> 根据Wipo Std 25	
<210> 1	
<211> 32	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增pE194	
<400> 1	
tatatacagc tggattcaca aaaaataggc ac	32
<210> 2	
<211> 33	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增pE194	
<400> 2	
tatatacagc tggattatgt cttttgcgca gtc	33
<210> 3	
<211> 30	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增功能性mRFP盒	
<400> 3	
tatatggatc cgtaatcagg gtatcgaggc	30
<210> 4	
<211> 31	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增功能性mRFP盒	
<400> 4	
tatatggatc cctcattagg cgggctacta a	31

<210> 5	
<211> 1345	
<212> DNA	
<213> 人工序列	
<220>	
<223> 功能性片段:T2A mRFP盒	
<400> 5	
ggaatagatc tggccaacga ggcctcgagg atccgatatc atgcatggcg cgccaccctg	60
agacgaccct gatgcaggtg accctgagac ctctgcgccc agctgtctag ggcggcggtat	120
ttgtcctact caggagagcg ttcaccgaca aacaacagat aaaacgaaag gcccgagtctt	180
tcgactgagc ctttcgtttt atttgatgcc ttaattaag taccttgctg gataaagctg	240
tgttatatta tgtcttggtg ttaaatacac acgcttaacg atttatgcag agggtgctgc	300
aggcggcagt tctgtacaaa aatgacctaa gcggaggaaa aaaaccatta tattaggagg	360
aaataacatg gcctcttcag aggatgttat taaagaattt atgcggttta aggtgaggat	420
ggaaggctcg gtgaacggac atgagttcga aattgagga gaaggtgaag gccgccctta	480
tgaaggact cagacagcga aattgaaagt cacgaaaggc ggaccgctgc cgtttgcttg	540
ggacattctc tcacctcaat ttcaatatgg ctcaaaagcc tacgtaaac acccgctga	600
catccctgat tacttaaagc tatccttccc ggagggttt aatgggaac gagttatgaa	660
ttttgaggac ggcggcgtcg ttactgtcac acaggattct tcccttcagg atggcgaatt	720
tatttacaaa gtaaaacttc gtggaactaa cttccaagt gatggtcccg tgatgcaaaa	780
aaaaacaatg ggatgggaag catctacgga acgtatgtat ccggaggatg gaccttaaa	840
gggtgaaatc aaaatgcgcc tgaaacttaa agatggcgga cactatgacg cggaagttaa	900
aacaacatat atggctaaaa aaccagtcca actgccggga gcatataaga cggatataaa	960
gttggacatt accagccata atgaagatta cacgattgtg gaacagtatg agagagcaga	1020
gggcagacat agcacaggcg cgtaagaatt aatgaaaaat aagcggcagc ctgcttttcc	1080
atgcgggctg ccgcttatcg ggttattgtc gtgactggga aaacctggc gactagtctt	1140
ggactcctgt tgatagatcc agtaatgacc tcagaactcc atctggattt gttcagaacg	1200
ctcggttgcc gccggcggtt ttttattggt gagaatccag ggtccccaa taattacgat	1260
ttggtctcac tcacacctgc tcgtctcact caatttaaat ggcggccgcg gatcctcgac	1320
ggccaataa ggccagatct ggatt	1345
<210> 6	
<211> 1012	
<212> DNA	
<213> 人工序列	
<220>	
<223> 同源性区域: amyB基因的5引物区和3引物区与侧翼XmaI位点融合	
<400> 6	
cccgggataa tgccgtcgca ctggccgata ttgagagatt tccttgtgac aagctgcaaa	60
gcataatgat gacggtccag ctgcggtctg attcccgtta acagattcat ataataaggt	120

tctgttgat ccatttcttc cagtatgagc agcttgacga cctgtgttct gttttgaacg	180
agcgcctcttg ctgcatagtt cggatatataa ttgagctcct tcattgcgga atgaacaagc	240
tttttcaatt catccgtcac agtctcagga tgattgatca cccgcgatac cgtcattttc	300
gacacatttg ctttctttgc tacatcagat aacgttgcca tttcatcccc gccttaccta	360
tgcgattcaa actgtcagca agtccttctt gagggctgat gacactttgt taaaattaat	420
tataaaatgt aatcaaagaa atttataaga cgggcaaaat aaaaaaacgg atttccttca	480
ggaaatccgt cctctctgct cttctagatt ctctctccct ttcaatgtga aacatatgat	540
attgtataaa tattccgaat ttttaacaaa taccattttc cctatatttt cttccaaaag	600
aaaagcgccg atatggcgct ttctactcat ttattcaata gcctctctgc ttcttcaact	660
cttcaagctg agatacagtt accaattgat agccttttgc tttcagcttt ttaataatct	720
cttcagcagc atctgcggac gttgcataaa tategtgcat taagacgatt tttcctctc	780
ccgcatggct catgacatga ttgacaatct tttgcttatt tttgtacttc caatcttccg	840
gatcaacatc ccacaatgaa accttcagat tggaaagcga gcggacggaa tcattgatcc	900
cgccgtatgg aggacgcaag tgtacaggca ggtgtccgct gattttttcg atcattttctt	960
gcgtgtcgtt aatctcctga tacgcttttt cgtttgacag ccttgtcccc gg	1012
<210> 7	
<211> 24	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:针对的原间隔区靶序列amyB基因 with 4nt 5-prime extension	
<400> 7	
tacgtcgcag cagaaattaa gaga	24
<210> 8	
<211> 24	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:带4nt 5-引物延伸部分的针对amyB基因的原间隔区靶序列	
<400> 8	
aaactctctt aatttctgct gcga	24
<210> 9	
<211> 20	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增amyB基因组区域	
<400> 9	

ttgcccgaat acaacgacag	20
<210> 10	
<211> 21	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增amyB基因组区域	
<400> 10	
caacaacagg ctgtctgacg g	21
<210> 11	
<211> 2143	
<212> DNA	
<213> Bacillus licheniforms	
<220>	
<223> P300 DSM641 RMS:RMS区和侧翼区	
<400> 11	
aacttctata aatgtaacga tacgatttat tgtttcatta aagtcttctt ttatttcatg	60
ttcccatatt cttttaatgt tccaatcctt ttctctgtaa tatttattaa cttccttatac	120
tcttttttta tttctttcga gttttttctc ccaatattcc gtattacttt ttggtatatt	180
cccgtgtttt tcacacgcat gccagaaaca agaatcaatg aatatgacta ttttatattt	240
ctgtattact atatctggac taccgataa tttcttaaca ttttttcgga atcttattcc	300
acggtgccat agttcttttag taaccttatac ttctaatttt gaacgagatt tgattgcctg	360
catgtttttt cttctttggt cttttgaaac cgtgtcagtc atagaagagt cctccaaagc	420
cacaataatt gtattctata aacgaggaag caagccctca agcttaccct ctcttagttc	480
cttttttgcc tacttattta ttgttttca tttcaaatt catcataaga acccatttca	540
caaaaatcaa tggagtaatg ttgttcgctg tgtttataaa caattactga gtcaatattt	600
agtctttcca gcatttcata ggttagaagt tcttttctt ctaagttcat aacttgtctt	660
agtagccatt ctccaagagc actatttggga ttagacataa gtgctttact attgtcttgg	720
cataccttgg ctgataaaag cgatttgtct ggcaagcgtactgaaaagg tttatcacga	780
gctgggaaaa atgttgggaa tacattatga atccattttg gaattggtat ataaatctcg	840
ttagggtttc gtggtcgacc taaagcattc cattggttta gaccgctttt ttctgggtaca	900
tgacgctttg agccacggtc tgaaaagagt ggaagaataa cgtgctcaag gttttcaaaa	960
ggattgacag ttggtgctgg aatttttgga atttcaaagc caaatagttt agccaattca	1020
tgataaggat tttctaagat ttcaacatta atttcttcaa taggtttatc agtgataaaa	1080
cgcttataaa ggggtgctctt agtgacatta aagctgtatt cgtgtagacc gtcttcaaag	1140
gtgattgtat ttctgttgtt acttactttc acatttgtaa ttgaggagat ttcaaccaag	1200
tccattggct cttcaaaaat aagaattttc cctggctttc ttgttacaca gtggtatatac	1260
attgaatcaa taccatattg tcttttagta aattcaattc tctcgttacg gagagaagca	1320
accgtgttta ttagctcttt tggagatttc ccacgataca agtctgagtc tttattgaat	1380

tcagctactt tttgaagagt atgaccatta ccatgaagaa aagtcttaat accaattccg	1440
acacgattta atgaagcgtc agcagaacag tctgacctc ccaagttttc agctccaaat	1500
gcttcacaaa aagcattttc cacattcctt gagaccaa ataggcgagtc actttcagag	1560
aacaaattgg atagcgaacc agttgagcgg agcatttggt tgtatgtagt gcagttgatg	1620
gctggttgat tagtatagaa cattatTTTT cctcctcttt tatgcttgtc atttcttctt	1680
tcagacccaa aaggtagtca gctgatacgt tcaatgtttc agctattctt ttgaaagtgt	1740
ccaatgatgg agttctatTT tcactttcat atagtacca agtgcttcta gtgaccccg	1800
ctttttcagc gatttggtg ggtaataacc tacgagcttc tcttgcatTT tgaatacga	1860
ttccaaggaa aggtatcatt tttgcacctc caagatttgg tgttttcaga gtatcaccag	1920
aacccccgaa aatagtccaa agttagctaa cagcaacaa ataaaaataa ataagttggt	1980
tactcttagc aaacttgTTa ctaaaatttg ataaagttat tcatttaate cagctcttat	2040
gctaaaaattg cattagcggg caagcttaat gtttgcaagg aggtataatt ttgacttate	2100
gagtaggtag tatgtttgct gggataggtg gaacttgTTt agg	2143
<210> 12	
<211> 1146	
<212> DNA	
<213> 地衣芽孢杆菌	
<220>	
<223> 编码区:限制酶P300 DSM641	
<400> 12	
ttatTTgttt tcattttcaa attcatcata agaaccatt tcacaaaaat caatggagta	60
atgttgttcg ctgtgtttat aaacaattac tgagtcaata tttagtcttt ccagcatttc	120
ataggTTaga agttctTTTT cctctaagtt cataacttgt cttagtagcc attctccaag	180
agcactatTT ggattagaca taagtgcttt actattgtct tggcatacct tggctgataa	240
aagcgatttg tctggcaagc gtaactgaaa aggtttatca cgagctggga aaaatgTTgg	300
gaatacatta tgaatccatt ttggaattgg tatataaatc tcgttagggt ttcgtggTcg	360
acctaaagca ttccattggt ttagaccgct ttttctggt acatgacgct ttgagccagc	420
gtctgaaaag agtggaaGaa taacgtgctc aaggttttca aaaggattga cagttggtgc	480
tggaatTTTT ggaatttcaa agccaaatag ttagccaat tcatgataag gattttctaa	540
gatttcaaca ttaatttctt caataggTTt atcagtgata aaacgcttat aaagggtgct	600
cttagtgaca ttaaagctgt attcgtgtag accgtcttca aaggtgattg tatttctgTT	660
gttacttact ttcacatttg taattgagga gatttcaacc aagtccattg gctcttcaaa	720
aataagaatt ttccctgget ttcttgTTac acagtggTat atcattgaat caataccata	780
tgttctTTTT gtaaattcaa ttctctcgTT acggagagaa gcaaccgtgt ttattagctc	840
ttttggagat ttcccacgat acaagtctga gtctttattg aattcagcta ctttttgaag	900
agtatgacca ttaccatgaa gaaaagtctt aataccaatt ccgacacgat ttaatgaagc	960
gtcagcagaa cagtctgacc tccccaagtt ttcagctcca aatgcttTca aaaaagcatt	1020
ttccacattc cttgagacca aataaggcga gtcacttTca gagaacaaat tggatagcga	1080
accagttgag cggagcattt gtttgtatgt agtgcagttg atggctggtt gattagtata	1140

gaacat	1146
<210> 13	
<211> 1022	
<212> DNA	
<213> 人工序列	
<220>	
<223> 同源性区域:限制酶基因的5引物区和3引物区与侧翼BsaI位点融合	
<400> 13	
ggtctcgacc caacttctat aaatgtaacg atacgattta ttgtttcatt aaagtcttcc	60
tttattttcat gttcccatat tcttttaatg ttccaatcct tttcctcgta atatttatta	120
acttccttat ctcttttttt atttctttcg agttttttct cccaatattc cgtattactt	180
tttggtatat tcccgtgttt ttcacacgca tgccagaaac aagaatcaat gaatatgact	240
atttttatatt tctgtattac tatactctgga ctaccgtata atttcttaac attttttcgg	300
aatcttattc cacgggtgcca tagttcttta gtaaccttat cttctaattt tgaacgagat	360
ttgattgcct gcatgttttt tcttctttgt tcttttgaaa ccggtgctcagt catagaagag	420
tcctccaaag ccacaataat tgtattctat aaacgaggaa gcaagccctc aagcttacc	480
cctcttagtt ccttttttgc ctacttattt atatttttcc tectctttta tgcttgtcat	540
ttcttctttc agacccaaaa ggtagtcagc tgatacgttc aatgtttcag ctattctttt	600
gaaagtgtcc aatgatggag ttctattttc actttcatat agtgaccaag tgcttctagt	660
gaccccgact ttttcagcga ttggctggg taataaccta cgagcttctc ttgcattttg	720
aatacgattt ccaaggaaag gtatcatttt tgcacctcca agatttgttg ttttcagagt	780
atcaccagaa cccccgaaaa tagtccaaag ttagctaaca gcaaacaat aaaataaat	840
aagttgttta ctcttagcaa acttgttact aaaattgat aaagttattc atttaacca	900
gctcttatgc taaaattgca ttagcggaca agcttaatgt ttgcaaggag gtataatttt	960
gacttatcga gtaggtagta tgtttgctgg gataggtgga acttgtttag gctcaggaga	1020
cc	1022
<210> 14	
<211> 21	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增限制酶基因组区域	
<400> 14	
gacaatcccc ttttactgac c	21
<210> 15	
<211> 24	
<212> DNA	
<213> 人工序列	
<220>	

<223> 寡核苷酸:PCR扩增限制酶基因组区域	
<400> 15	
ctatttcatt tgcccagaca atcc	24
<210> 16	
<211> 1363	
<212> DNA	
<213> 人工序列	
<220>	
<223> 同源性区域:pga基因的5引物区和3引物区与侧翼BsaI位点融合	
<400> 16	
ggtctcgacc cgaacactga aattttagac cgggctggga tatcgagaca tatatagggg	60
cgtttaatgg cgaatacagt aacaatgaga atagtaagaa aaattaaaaa tgttaaagtt	120
tgatgaatta tcattgaaaa aaattaatgg ctttttaaat cctaggattt taacctaaaa	180
tctgaagaaa taaggtggat cgaacgactc acaaaatatt tggatttgtc aatgaatccc	240
gctttatgct aaaagagatt ttcatttttt gatagatggt ctgattgtca taggacggat	300
ttgttttgaa gagggaacat tggtgacttt ttaacctgtt cgaaaagagc gaaaatacta	360
aaagaaaaga gacatcccgg ctgacagccc atttaaaggg gattgcggcc gggggaaaaa	420
agagatcctg aatccatcct tcaacctttc atctgaaata gggagaaaag taaaaaatc	480
ataatgtcga attttgaaag cgcatactta aaacgctgac aaaaatctga taggaattaa	540
gaactttcga tttccaaaaa tatcaataaa aagataggca ttaatgactc gggcgaggtg	600
atctttgtca cggaaaatth cgtcgtcttc tgttacataa tgccgattgt gatttcatag	660
tgaaccctga tcccgttat aaaagacctg tgaaaagcgg ccggtttgaa agggaaacac	720
gacaattttc ttaaccggtc agtgtataaa gttttataga aatcaggag gatataata	780
tggttttggg gttcatgttt attgtattct tttgaaggga ataaaaactg acaaatctcg	840
actgaagcaa aatttgaaaa tgcatacct taccaattcg ggatgggaac cgcacctcat	900
gttcatgacc tctttagaat atttccctc atctttttta tccgcgctta ggtgaaaaag	960
ctgatcatgc tgtgctgagc gtttcttctc gctatgacgc tgctgtacat gcaaaaaaag	1020
tcctttaaat atcccagttg aatgacgatg aaagaggaaa gaagaggagg aacagatcaa	1080
ttgataaaaa aagcggcaaa caaaaagttg gttttgtttt gtggaattgc ggtgctttgg	1140
atgtctttat ttttaacgaa tcataatgat gtacgcgccg atacgatcgg cgagaaaata	1200
gcggaaactg ccagacagct tgagggtgcg aaatacagct acggcggaga gaagccgaaa	1260
acggggtttg actcgtcagg ctttgtgcaa tatgtgttct aatcgtcga tattacgctt	1320
ccgagaacgg taaaggaaca atcgactctt ggctcaggag acc	1363
<210> 17	
<211> 20	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增pga基因组区域	

<400> 17	
aaagccttct cctctctatt	20
<210> 18	
<211> 20	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增pga基因组区域	
<400> 18	
ttcttgaaaa agacaaggtc	20
<210> 19	
<211> 1027	
<212> DNA	
<213> 人工序列	
<220>	
<223> 同源性区域:aprE 基因的5引物区和3引物区与侧翼BsaI位点融合	
<400> 19	
ggtctcgacc cgaagttctt ttttaacata taggtaaac aatacgtaaa aaggcgccaa	60
gtattgaaga attgcagcag ccgcggcatt tcccttttcg attgaagcaa aaaacgtata	120
ttgaacagta agcattccaa aaatggaaaa tactaaaatc gaacaaatat ctgttttttt	180
cttccatatac tgacacacat gttgaaaacc gtttttcatt gaaacatata acaagagaat	240
gactcccgat gccagaagcc tgacagagac aagcgagccg gcttcaaccg ctccccttcc	300
aaatatgtac tgtgcagcgc ttcccagata tcccacaat gaagcccctg caagcaccat	360
caatacgcct ttcacatgag ctgatttcat atctttcacc cgtttctgta tgcgatatat	420
tgcataatttt aatagatgat cgacaaggcc gcaacctcct tcggcaaaaa atgatctcat	480
aaaataaatg aatagtattt tcataaaatg agctcaataa catattctaa caaatagcat	540
atagaaaaag ctagtgtttt tagcactagc tttttcttca ttctgatgaa ggttgttcaa	600
tattttgaat ccgttccatg atcgtcggat ggccgtattt aaaaatcttg acgagaaacg	660
gcggttttgc ctcgctcagc ccggcttttg agagctcttg aaacgtcgaa accgctgcat	720
cgctgttttg cgtcagttca atcgcatact ggtcagcagc tttttcctga tgcctcgaaa	780
ctgcgttcgt aaatggagac gacgcgaaag agatgacccc catcagcatc agaagaagcg	840
gaagtgcggc tagatcggat tttctgcaa tatgaagget tcttccatag cggccgatga	900
tccgcttgta cagcttgtcg atcacataaa agacagcaag ggataaaagc agataccgcg	960
caagtccat gtaaacaatgc ttcacacat agtgccecat ttcgtgcgcc atgatgctca	1020
ggagacc	1027
<210> 20	
<211> 21	
<212> DNA	
<213> 人工序列	

<220>	
<223> 寡核苷酸:PCR扩增aprE基因组区域	
<400> 20	
ccggttgatca ttgatccttt a	21
<210> 21	
<211> 21	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增aprE基因组区域	
<400> 21	
atcctcctgc aaaaaccgta t	21
<210> 22	
<211> 41	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增pBW742	
<400> 22	
caaatttaca aaagcgactc gtgagttttc gttccactga g	41
<210> 23	
<211> 36	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增pBW742	
<400> 23	
gaacgttgct ctagagttaa gggattttgg tcatgg	36
<210> 24	
<211> 38	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增pMutin2的终止子区	
<400> 24	
gtggaacgaa aactcagag tcgcttttgg aaatttgg	38
<210> 25	
<211> 40	
<212> DNA	

<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增pMutin2的终止子区	
<400> 25	
catgaccaaaa atcccttaac tctagagcaa cgttcttgcc	40
<210> 26	
<211> 22	
<212> DNA	
<213> 人工序列	
<220>	
<223> 功能性片段: pMutin2的终止子区	
<400> 26	
ggggatctct gcagtgagat ct	22
<210> 27	
<211> 24	
<212> DNA	
<213> 人工序列	
<220>	
<223> pCC009中终止子区的序列	
<400> 27	
ggggatctct gcagtcggga agat	24
<210> 28	
<211> 128	
<212> DNA	
<213> 人工序列	
<220>	
<223> 启动子变体	
<400> 28	
aatthttgtca aaataatthtt attgacaacg tcttattaac gttgataccg gttaaatthtt	60
atthtgacaaa aatgggctcg tgttgagaa taaatgtgga gaaagattaa ctaataagga	120
ggacaaaac	128
<210> 29	
<211> 128	
<212> DNA	
<213> 人工序列	
<220>	
<223> 启动子变体	
<400> 29	
aatthttgtca aaataatthtt attgacaacg tcttattaac gttgataccg gttaaatthtt	60

atttgacaaa aatgggctcg tgttgaagaa taaatgtgga gaaagattaa ctaataagga	120
ggacaaac	128
<210> 30	
<211> 101	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸	
<400> 30	
aacacgagcc catttttgtc aaataaaatt taaccggtat caacgttaat aagacgttgt	60
caataaaatt attttgacaa aattttaata atccaaatga g	101
<210> 31	
<211> 87	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸	
<400> 31	
ctattgagta tttcttatcc atgtttgtcc tccttattag ttaatctttc tccacattta	60
ttctccaaca cgagccatt tttgtca	87
<210> 32	
<211> 48	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸	
<400> 32	
gattaactaa taaggaggac aaacatggat aagaaatact caataggc	48
<210> 33	
<211> 87	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸	
<400> 33	
ctattgagta tttcttatcc atgtttgtcc tccttattag ttaatctttc tccacattta	60
ttcttcaaca cgagccatt tttgtca	87
<210> 34	
<211> 127	

<212> DNA	
<213> 人工序列	
<220>	
<223> 启动子变体	
<400> 34	
aatTTTgtca aaataatTTT attgacaacg tcttattaac gttgataccg gttaaatttt	60
atTTgacaaa atgggctcgt gttggagaat aaatgtggag aaagattaac taataaggag	120
gacaaaac	127
<210> 35	
<211> 126	
<212> DNA	
<213> 人工序列	
<220>	
<223> 启动子变体	
<400> 35	
aatTTTgtca aaataatTTT attgacaacg tcttattaac gttgataccg gttaaatttt	60
atTTgacaaa aatgggctcg tgttggagaa aatgtggaga aagattaact aataaggagg	120
acaaaac	126
<210> 36	
<211> 127	
<212> DNA	
<213> 人工序列	
<220>	
<223> 启动子变体	
<400> 36	
aatTTTgtca aaataatTTT attgacaacg tcttattaac gttgataccg gttaaatttt	60
atTTgacaaa aatgggctcg tgttgggaat aaatgtggag aaagattaac taataaggag	120
gacaaaac	127
<210> 37	
<211> 129	
<212> DNA	
<213> 人工序列	
<220>	
<223> 启动子变体	
<400> 37	
aatTTTgtca aaataatTTT attgacaacg tcttattaac gttgataccg gttaaatttt	60
atTTgaacaa aaatgggctc gtgttggaga ataaatgtgg agaaagatta actaataagg	120
aggacaaaac	129
<210> 38	

<211> 127	
<212> DNA	
<213> 人工序列	
<220>	
<223> 启动子变体	
<400> 38	
aattttgtca aaataat tttt attgacaacg tcttattaac gttgataccg gttaaatttt	60
at ttgacaaa aatgggctcg t gttgagaat aatgtggag aaagattaac taataaggag	120
gacaaaac	127
<210> 39	
<211> 127	
<212> DNA	
<213> 人工序列	
<220>	
<223> 启动子变体	
<400> 39	
aattttgtca aaataat tttt attgacaacg tcttattaac gttgataccg gttaaatttt	60
at ttgacaaa aatgggctcg t gttgaaaat aatgtggag aaagattaac taataaggag	120
gacaaaac	127
<210> 40	
<211> 127	
<212> DNA	
<213> 人工序列	
<220>	
<223> 启动子变体	
<400> 40	
aattttgtca aaataat tttt attgacaacg tcttattaac gttgataccg gttaaatttt	60
at ttgacaaa aagg gctcgt gttggagaat aatgtggag aaagattaac taataaggag	120
gacaaaac	127
<210> 41	
<211> 127	
<212> DNA	
<213> 人工序列	
<220>	
<223> 启动子变体	
<400> 41	
aattttgtca aaataat tttt attgacaacg tcttattaac gttgataccg gttaaatttt	60
at ttgacaaa aatgggctcg t gttgaagaa aatgtggag aaagattaac taataaggag	120
gacaaaac	127

<210> 42	
<211> 130	
<212> DNA	
<213> 人工序列	
<220>	
<223> 启动子变体	
<400> 42	
aat t t t g t c a a a t a a t t t t a t t g a c a a c g t c t t a t t a a c c g t t g a t a c c g g t t a a a t t t	60
t a t t t g a c a a a a t g g g c t c g t g t t g a a g a a t a a t g t g g a g a a g a t t a a c t a a t a a g g	120
g a g g a c a a a c	130
<210> 43	
<211> 127	
<212> DNA	
<213> 人工序列	
<220>	
<223> 启动子变体	
<400> 43	
a a t t t t g t c a a a t a a t t t t a t t g a c a a c g t c t t a t t a a c g t t g a t a c c g g t t a a t t t t	60
a t t t g a c a a a a t g g g c t c g t g t t g a a g a a t a a t g t g g a a a g a t t a a c t a a t a g g a g	120
g a c a a a c	127
<210> 44	
<211> 128	
<212> DNA	
<213> 人工序列	
<220>	
<223> 启动子变体	
<400> 44	
a a t t t t g t c a a a t a a t t t t a t t g a c a a c g t c t t a t t a a c g t t g a t a c c c g g t t a a a t t t	60
t a t t t g a c a a a a t g g g c t c g t g t g a a g a a t a a t g t g g a g a a g a t t a a c t a a t a a g g a	120
g g a c a a a c	128
<210> 45	
<211> 127	
<212> DNA	
<213> 人工序列	
<220>	
<223> 启动子变体	
<400> 45	
a a t t t t g t c a a a t a a t t t t a t t g a c a a c g t c t t a t t a a c g t t g a t a c c g g t t a a t t t t	60
a t t t g a c a a a a t g g g c t c g t g t g a g a a t a a t g t g g a g a a g a t t a a c t a a t a g g a g	120

gacaaac	127
<210> 46	
<211> 127	
<212> DNA	
<213> 人工序列	
<220>	
<223> 启动子变体	
<400> 46	
aatTTTgtca aaataatTTT attgacaacg tcttattaac gttgataccg gttaaattTTT	60
atTTtgacaaa atgggctcgt gttgaagaat aaatgtggag aaagattaac taataaggag	120
gacaaac	127
<210> 47	
<211> 129	
<212> DNA	
<213> 人工序列	
<220>	
<223> 启动子变体	
<400> 47	
aatTTTgtca aaataatTTT attgacaacg tcttattaac gttgataccg gttaaattTTT	60
atTTTgacaa aaatgggctc gtgTTgaaga ataaatgtgg agaaagatta actaataagg	120
aggacaaaac	129
<210> 48	
<211> 40	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸	
<400> 48	
gaagTTTTtag atgccactct tatccatcaa tccatcactg	40
<210> 49	
<211> 41	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸	
<400> 49	
catctcaaat ttgcattta ttccaatttc cTTTTgcgt g	41
<210> 50	
<211> 40	

<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸	
<400> 50	
cacacgcaaa aaggaaattg gaataaatgc gaaatttgag	40
<210> 51	
<211> 40	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸	
<400> 51	
gaccagtgat ggattgatgg ataagagtgg catctaaaac	40
<210> 52	
<211> 20	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸	
<400> 52	
caagctatgc ttgctcaagc	20
<210> 53	
<211> 50	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸	
<400> 53	
ctgttggttc gcttgagcaa gcatagcttg gacggttcag cgtgttaagc	50
<210> 54	
<211> 37	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸	
<400> 54	
ggtggtggtc tctaccccga tcgaagagcc attcgag	37
<210> 55	

<211> 37	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸	
<400> 55	
ggtgggtggtc tcttgagctt cctctgtccg attgtcc	37
<210> 56	
<211> 24	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸	
<400> 56	
tacggcaatc tctgaaaaaa tgag	24
<210> 57	
<211> 24	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸	
<400> 57	
aaacctcatt ttttcagaga ttgc	24
<210> 58	
<211> 814	
<212> DNA	
<213> 人工序列	
<220>	
<223> 同源性区域:带侧翼BsaI位点的突变型degU基因同源性区域	
<400> 58	
ggtctcgacc cgatttgga ttgataccga cgctcagaaa atacttgaac acgatcgaag	60
attatcatgg aaaagcaaag attcatttcc aatgcatcgg agaatccgaa gaaagaagaa	120
tagcaccgcg gtttgagggt gcactattcc ggcttgaca ggaagcgggtg acaaacgcct	180
taaaacactc cgaatcaact gaaattcatg ttaaagtaga agtgacaaaa gatthttgtga	240
cgctgattat caaagacaat ggaaacggct ttgacttaaa agaagtaaaa ggcaagaaga	300
acaaatcttt cggctctgcta ggtatgaaag aaagagtcga tttgctcgaa ggctcaatga	360
caatcgattc gaaaaataggt cttgggacat ttatattgat taaagttcca ctgtctttgt	420
aaagataatt gtaaaaataga gacaaaagac atattgacca taaaagcgggt gtgtttaaca	480
atgagaatgg ggaggcgtag cttgtgacta aagtaaatat tgtaattatt gacgatctgc	540

agttattccg tgaaggtgtc aaacggattt tggatttcga gcctaccttt gaagtagtgg	600
ccgaaggaga cgacggagat gaagcggctc gcattgtcga gcactaccat cctgatgttg	660
ttatcatgga tattaatatg ccgaatgtga acggagtaga agcgacaaaa caactggtcg	720
acttgatcc ggaatcaaag gttattattt tatccatcca tgatgacgaa aactatgtta	780
cacatgcatt aaaaacagga gccctcagga gacc	814
<210> 59	
<211> 24	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸	
<400> 59	
tacgggattt cgaacctacc tttg	24
<210> 60	
<211> 24	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸	
<400> 60	
aaaccaaagg taggttcgaa atcc	24
<210> 61	
<211> 39	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸	
<400> 61	
ggttgccggtc tcatacgtga agatcaggct atcactggt	39
<210> 62	
<211> 36	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸	
<400> 62	
caacgggtct cttgagttgg caggccgctg aatttc	36
<210> 63	
<211> 1435	

<212> DNA

<213> 人工序列

<220>

<223> sgRNA和同源性区域:原间隔区aprE-sgRNA,
带侧翼BsaI位点的aprE 基因同源性区域

<400> 63

```

ggttgcggtc tcatacggaa acaaaccat accaggagtt ttagagctag aaatagcaag      60
ttaaaataag gctagtccgt tatcaacttg aaaaagtggc accgagtcgg tgctttttac      120
tccatctgga tttgttcaga acgctcggtt gccgccgggc gttttttatc taaagcttag      180
gcccagtcga aagactgggc ctttttaata cgactcacta tagggtcgac ggccaacgag      240
gcccattttc ttctgctatc aaaataacag actcgtgatt ttccaaacga gctttcaaaa      300
aagcctctgc cccttgcaaa tcggatgcct gtctataaaa ttcccgatat tggttaaaca      360
gcggcgcaat ggcgccgca tctgatgtct ttgcttggeg aatgttcate ttattttctc      420
ctccctctca ataatttttt cattctatcc cttttctgta aagtttattt ttcagaatac      480
ttttatcatc atgctttgaa aaaatatcac gataatatcc attgttctca cggaagcaca      540
cgcaggatcat ttgaacgaat tttttcgaca ggaatttgcc gggactcagg agcatttaac      600
ctaaaaaagc atgacatttc agcataatga acatttactc atgtctatth tcgttctttt      660
ctgtatgaaa atagtatttt cgagtctcta cggaaatagc gagagatgat atacctaat      720
agagataaaa tcatctcaaa aaaatgggtc tactaaaata ttattccatc tattacaata      780
aattcacaga atagtctttt aagtaagtct actctgaatt tttttaaag gagagggtaa      840
agataatagt aaaaagaagc aggttcctcc atacctgctt ctttttattt gtcagcatcc      900
tgatgttccg gcgcattctc ttctttctcc gcatgttgaa tccgttccat gatcgacgga      960
tggctgcctc tgaaaatctt cacaagcacc ggaggatcaa cctggctcag ccccgtcacg      1020
gccaaatcct gaaacgtttt aacagcggct tctctgttct ctgtcaactc gatcccatac      1080
tggtcagcct tattctcctg ataacgcgag acagcattag aaaaaggcgt aaccgcaaag      1140
ctcaaaacag aaaacaaaag caataacagc ggaagtgccg caagatcatg ccgcccttct      1200
aaatgaaaca tgctgcgggt taggcgaacc gtccgcttgt aaagcttatc aatgacataa      1260
aatccggcga gcgacacgag caaatagcca gccagaccga tgtaaactg cttcatgaca      1320
taatggccca tttcgtggcc cataataaac agaatttctg aatcgtcaag tttgttcagc      1380
gtcgtatccc acaatacaat ccgtttattg gcccattc tcaagagacc cgttg          1435

```

<210> 64

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 原间隔区:针对aprE基因的原间隔区靶序列

<400> 64

```

gaaacaaacc cataccagga          20

```

<210> 65

<211> 1241	
<212> DNA	
<213> 人工序列	
<220>	
<223> sgRNA和同源性区域:原间隔区vpr-sgRNA, 带侧翼BsaI位点的vpr 基因同源性区域	
<400> 65	
ggttgcggtc tcatacgatc aggaacggaa cgagtcggtt ttagagctag aaatagcaag	60
ttaaaataag gctagtcctg tatcaacttg aaaaagtggc accgagtcgg tgctttttac	120
tccatctgga tttgttcaga acgctcgggt gccgccgggc gttttttatc taaactagta	180
cagtcgatac ggatccaacta gtctatctct tctctttttt cegaaaagcc gcctgcctga	240
taagcatcgg cagcctgata tgtaccgccc gcgaacgccc tcaccttcat ctgattttct	300
ctgacaggca tcaccgcatt cagcagcagc gctgtccata attttaaaa tgatatcaaa	360
cctttcatac cgatccctcc agtttcggtt tgataaaact agcaactcta ttaaactttc	420
ttgctctatc ttatcccagc aaaatgaaaa tgtttgtcac aatgtgtgtg caaatgatt	480
ctagttttta gaagttttgt tgaaaactga aggaatcgca tgattcagcg gatacaaacc	540
atgaatgtaa ctactcaca gcttatccta aggataaaca catattacc acaggatata	600
tccacatata cacatactta ttcaatattt agtataagaa cgtatattcc ctacaatata	660
tatacacaag tttatttact tatacacagt aaattgtgca taaatctaat gacaagcctt	720
gttgagaacc actcaacaag gcttttttat gttaaatac ggataatgag ttcaggagaa	780
gctccccttc tcttcaaac gtgaaaaaag caatcggagg acatcgtgta tatgctttct	840
tttatcgtat tattcggctt atccttcatt attgtctgct ttatattttt cagcacttg	900
tacttcgccg tcaacctgca gaagcgcgag cccaagcctt ttcaaaaagc tgcggagcaa	960
accgctgata ccatcatcct cattccgctc agctggctgt ttaccgcttt atacatatgc	1020
attctgttta ttcttttccc aatccgcat tttctcgatt tttttcagca aaaacgctaa	1080
attgactgat gaaacgcttc ggccagcagc cggtatgaat ccaatctgtc ttgaaaatcg	1140
tgggtgatcg tcaccgcat gatttcgtcc gttccgtaag cgccggccag ttcaagcagc	1200
tgttcctagc tcgagccatg gctcactcaa gagaccggtt g	1241
<210> 66	
<211> 20	
<212> DNA	
<213> 人工序列	
<220>	
<223> 原间隔区:针对vpr基因的原间隔区靶序列	
<400> 66	
atcaggaacg gaacgagtcg	20
<210> 67	
<211> 20	
<212> DNA	

<213> 人工序列	
<220>	
<223> 原间隔区:针对vpr基因的原间隔区靶序列	
<400> 67	
gcttccgtat aatgagtatt	20
<210> 68	
<211> 20	
<212> DNA	
<213> 人工序列	
<220>	
<223> 原间隔区:针对vpr基因的原间隔区靶序列	
<400> 68	
cacgatccga aaaacccgta	20
<210> 69	
<211> 1338	
<212> DNA	
<213> 人工序列	
<220>	
<223> sgRNA和同源性区域:原间隔区epr-sgRNA, 带侧翼BsaI位点的epr基因同源性区域	
<400> 69	
ggttgcggtc tcatacgatt ccggtccagc gcttttagtt ttagagctag aaatagcaag	60
ttaaaataag gctagtccgt tatcaacttg aaaaagtggc accgagtcgg tgctttttac	120
tccatctgga tttgttcaga acgctcggtt gccgccgggc gttttttatc taaactagta	180
ggcccagtcg atacgtaaag actgggcctt ttaatacga ctactatag ggctcgacggc	240
caacgaggcc tcgaggatcc gatatcatgc atggcgcgcc accctgagac gaccctgatg	300
caggtgacct ggatccacta gtccgatttg acatcgtgct gtttaaagga cctgacaaag	360
atatattcat taaaagggtg atcgggcttc cgggcgaaac cctcaggtat gaagatgatc	420
agctgtatat caacgaagaa aagatcaaag agccttatct ggacgactta aaggccgtca	480
ccgccggagg ggacttgaca ggggatttta cactgcagga agtgaccgga gaggagaagg	540
tgccctgaaaa cgagtacttc gtccctcgggg acaaccggat ccacagcttt gacagccgcc	600
atttcggctt tgtttcagaa cgggacatcg tcgggattgt gacggaaaga attgataaga	660
agtgattgga gagtacgggg gagagtaagc ggccgaccaa ggaatacgat tacgcaaatg	720
acgagcccga aatgtcaatt agtacaacag catcaataat gacgcatttg ctaaatatga	780
aaattaaaag gccccgatga ttccgggctt ttttccgtac taagcggcgt tcgctatata	840
tatcggagga tttttgaatt ttcaaagaga aagaaagctt atcttaaggt cgcttgtcat	900
gcaccttag tttttaaacc gttaaaaaca ctgttatatc aacatttgtg aagcttcctg	960
tttattcggg aagttaaatt gggtactcca agttagtttt aaaaaagagt cataaggcca	1020
gcttctatcg atgaatcatt ttaagcgac gccttttgtc taaaatgtat aatgttactt	1080

ttgtttttgt tgaagtgaa caatgttatt gactggctta caacctaca ttttaattaaa	1140
taaaaaatag attaaaagaa gggagcggtc tcataccgtg gaaaaaaca ttaaacatga	1200
cccccaattat tacaaaaaga taattattgc attgtgttta ggggtgggtcg ctatttggat	1260
ttatcgtaca atacttacgc caatatatcc gcagattcaa gaatcattag ggaatattag	1320
tgctcaagag acccgttg	1338
<210> 70	
<211> 20	
<212> DNA	
<213> 人工序列	
<220>	
<223> 原间隔区:针对epr基因的原间隔区靶序列	
<400> 70	
attccggtcc agcgctttta	20
<210> 71	
<211> 20	
<212> DNA	
<213> 人工序列	
<220>	
<223> 原间隔区:针对epr基因的原间隔区靶序列	
<400> 71	
cgctttttca gctttggcaa	20
<210> 72	
<211> 20	
<212> DNA	
<213> 人工序列	
<220>	
<223> 原间隔区:针对epr基因的原间隔区靶序列	
<400> 72	
gcaaacatgc cggtagcgtg	20
<210> 73	
<211> 865	
<212> DNA	
<213> 人工序列	
<220>	
<223> 功能性片段:T2A lacZ盒	
<400> 73	
gcgatgttaa ggaccgtctc atctcacctg caaggtctca tctcttttac tccatctgga	60
tttgttcaga acgctcggtt gccgccgggc gttttttatc taaaactagt gtcgagggtc	120
ttcggtagcg cgatttacat atgctggcac gacaggttcc cgcactggaa agcgggcagt	180

gagcgcaacg caattaatgt gagttagctc actcattagg caccccaggc tttacacttt	240
atgcttccgg ctcgtatgtt gtgtggaatt gtgagcggat aacaatttca cacaggaac	300
agctatgacc atgattacgc caagcttgca tgctgcagg tcgactctag aggatccccg	360
ggtaccgagc tcgaattcac tggccgtcgt ttacaacgt cgtgactggg aaaaccctgg	420
cgttacccaa cttaatcgcc ttgcagcaca tcccccttc gccagctggc gtaatagcga	480
agaggcccg accgatcgcc cttccaaca gttgcgcagc ctgaatggcg aatggcgcct	540
gatgcggtat tttctcctta cgcactctgt cggtatttca caccgatat ggtgactct	600
cagtacaatc tgctctgatg ccgcatagtt aagccagccc cgacaccgc caacaccgc	660
tgccgcgttt ataatgaaga ccctagcagg catcaataa aacgaaaggc tcagtcgaaa	720
gactgggcct ttcgttttat ctgttgttg tcggtgaacg ctctctgag taggacaaat	780
ccgccccct agacagctgt cgtgagacg atcgtgaga cctcgcaagt tctcgccatc	840
gcaggtgaaa tctagatgta ttcgc	865
<210> 74	
<211> 37	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增启动子 PaprE	
<400> 74	
tatatgaaga ccttcgactc gggacctctt tcctcg	37
<210> 75	
<211> 41	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增启动子 PaprE	
<400> 75	
tatagaagac ttatatctca ctctcctct cttattcag a	41
<210> 76	
<211> 747	
<212> DNA	
<213> 人工序列	
<220>	
<223> 功能性片段:带侧翼BpiI限制性位点的GFPmut2 AF302837	
<400> 76	
attagaagac ctatatgagt aaaggagaag aacttttcac tggagttgtc ccaattcttg	60
ttgaattaga tggcgatgtt aatgggcaaa aattctctgt cagtggagag ggtgaaggtg	120
atgcaacata cggaaaactt acccttaaatt ttatttgac tactgggaag ctacctgttc	180
catggccaac acttgtcact actttcgcgt atggtcttca atgctttgcg agataccag	240

atcatatgaa acagcatgac tttttcaaga gtgccatgcc cgaaggttat gtacaggaaa	300
gaactatatt ttacaaagat gacgggaact acaagacacg tgctgaagtc aagtttgaag	360
gtgataccct tgtaataga atcgagttaa aaggattga ttttaaagaa gatggaaaca	420
ttcttggaca caaaatggaa tacaactata actcacataa tgtatacatc atggcagaca	480
aaccaaagaa tggaatcaaa gttaacttca aaattagaca caacattaaa gatggaagcg	540
ttcaattagc agaccattat caacaaaata ctccaattgg cgatggccct gtccttttac	600
cagacaacca ttacctgtcc acacaatctg ccttttcaa agatcccaac gaaaagagag	660
atcacatgat ccttcttgag tttgtaacag ctgctgggat tacacatggc atggatgaac	720
tatacaaata atagcttgtc ttcatta	747
<210> 77	
<211> 33	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增amyB基因组区域	
<400> 77	
tataggtctc aaccataat gccgtcgcac tgg	33
<210> 78	
<211> 35	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增amyB基因组区域	
<400> 78	
ccttggctctc gtcgctagaa gagcagagag gacgg	35
<210> 79	
<211> 39	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增amyB基因组区域	
<400> 79	
ttgagaggtc tcagagacat tttccctata ttttcttc	39
<210> 80	
<211> 35	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增amyB基因组区域	

<400> 80	
gtgaggctctc atgagacat ccgttattga caagg	35
<210> 81	
<211> 20	
<212> DNA	
<213> 人工序列	
<220>	
<223> 原间隔区:针对sigE基因的原间隔区靶序列	
<400> 81	
cggtgaggaa aaaacccaaa	20
<210> 82	
<211> 1338	
<212> DNA	
<213> 人工序列	
<220>	
<223> sgRNA和同源性区域:原间隔区epr-sgRNA, 带侧翼BsaI位点的sigE 基因同源性区域	
<400> 82	
ggttgcggtc tcatacgcgg tgaggaaaa acccaaagtt ttagagctag aaatagcaag	60
ttaaaataag gctagtccgt tatcaacttg aaaaagtggc accgagtcgg tgctttttac	120
tccatctgga tttgttcaga acgctcgggt gccgccgggc gttttttatc taaactagta	180
ggcccagtcg atacgtaaag actgggcctt ttaatacga ctactatag ggtcgacggc	240
caacgaggcc tcgaggatcc gatatcatgc atggcgcgcc accctgagac gaccctgatg	300
caggtgacct ggatccacta gtaagacagt atgatgaaca agtgcttgtg gaactacaca	360
ttcacggaga gacaattcgt ttaaaggac tcgtcgatc tggcaaccag ctgtatgatc	420
ctatgaccaa aacaccggtc atgatcgtcc aggccacca tctaaccgcc atttgcggag	480
aatcgtttat agaccttatg aaacagtctc atcctgttga agtcatgcaa aagatcgatg	540
atcaatttcc tcttcttgat cgattaagac ttgttccata tcgagcagtc ggtcatgatc	600
acggttttct actatgccta aaaccagata cagttgtcat ttattcaaag acgcatatga	660
ttcagccagc taagtgtttt gtaggattga gtctgagcgg cttatcggca gatcaggaat	720
ttcaatccat cattcatcca gatatgttag acgggaaat catccagggg gtgtcgtagt	780
ttttggtgat gtcttttate ttacgaggtc aacttgacat ttgaaaaat ttttttgaaa	840
agctctgtcc ctgtactgtc aaaggaaaca accttttttc tcatgattct cgtcatcgct	900
cgtgcatatt tttccaacce aaggagatac tgaactttgt acaacagctc ctgtagggag	960
ggaaaaaagt gtccagaaat aaagtggaaa tctgcggagt cgacacctcc aagctgcctg	1020
ttctgaaaaa cgacgagatg agaaaattgt tcagacagct gcaagatgag ggtgacgata	1080
cagcaagaga aaagctagtc aatggcaatt tacggttggg ttttaagtgtg attcagcgtt	1140
tcaacaacag aggcgagtat gtcgatgatc tctttcaagt aggctgtatc ggattaatga	1200
aatcaattga taattttgat ttaagccaca atgtcagatt ttcaacttat gcggttccta	1260

tgatcatagg agaaatccgt cgatacttgc gtgataataa tccgattcgg gtgtctcgct	1320
cactcaagag acccgttg	1338
<210> 83	
<211> 20	
<212> DNA	
<213> 人工序列	
<220>	
<223> 原间隔区:针对sigF基因的原间隔区靶序列	
<400> 83	
tgctgcctc gctcaagaag	20
<210> 84	
<211> 1338	
<212> DNA	
<213> 人工序列	
<220>	
<223> sgRNA和同源性区域:原间隔区epr-sgRNA, 带侧翼BsaI位点的sigF 基因同源性区域	
<400> 84	
ggttgcggtc tcatacgtgt cgtcctcgct caagaagtt ttagagctag aaatagcaag	60
ttaaaataag gctagtccgt tatcaacttg aaaaagtggc accgagtcgg tgctttttac	120
tccatctgga tttgttcaga acgctcggtt gccgccgggc gttttttatc taaactagta	180
ggcccagtcg atacgtaaag actgggcctt ttaatacga ctactatag ggtcgacggc	240
caacgaggcc tcgaggatcc gatatcatgc atggcgcgcc accctgagac gaccctgatg	300
caggtgacct ggatccacta gtaaattatc cgtatggagc cttcagagca aacggcattg	360
caaacattgg ggggtggcatc atgaggaatg aatgaacct gaccttctct gccttaagtc	420
aaaatgaatc ctttgcgagg gtgacagtcg cggcgtttat cgcccagctt gatccgacat	480
tagatgaatt aactgaaatc aaaaccgttg tgcagaggc ggtgacaaac tccattatcc	540
atggctatga tgggaatcca gatggcaagg tgcataatga agtcacactt gatgatcatg	600
ttgtgtacct gaccatccgt gacgaaggaa tgggtattac agatcttgag gaagcaagac	660
agccgctttt cacgacaaaa ccagacttag aacgctctgg catgggcttt accattatgg	720
agaattttat ggatgatgtc atgatagact catctccaga aatgggcaca accatccgtt	780
taacaaaagca tctatcaaaa agcaaagcgc tttgtaatta aatagccaat tcggctggct	840
ttttttgtgt ggtaattacc ggtaaataa gttctctcgg tatgagaacc atttttcacc	900
acatactatt tttaaccat cgtataggaa gtgacttgga tggacggaca aatctttctt	960
cggtgcgcc atcgcatata gaccgtaat gatcagctca tttatttaga agatategcc	1020
caaatcactg gtgatgagtt ggetgtgcaa aagcttagca agatgccgat atatcatgtc	1080
agtaaaaagg atcgtcacat tgccgttctt gatatcatgc atgttgtcaa aacgatcaaa	1140
aaaacatggc caaccatcga cattcaaac gtcggaggcg ctgaagccat tgttgaaatt	1200
gatacaggca aacgccagct ttctcccgtt ttatttgtgt tcgtgtggct tttattat	1260

gtcggagcgg cgcttgccat tatgaatttc cacgaggatg tcagtatgcg gctcgtccat	1320
atctcaagag acccgttg	1338
<210> 85	
<211> 20	
<212> DNA	
<213> 人工序列	
<220>	
<223> 原间隔区:针对spoIIE基因的原间隔区靶序列	
<400> 85	
cttgtagctg aacagctgat	20
<210> 86	
<211> 1338	
<212> DNA	
<213> 人工序列	
<220>	
<223> sgRNA和同源性区域:原间隔区epr-sgRNA, 带侧翼BsaI位点的spoIIE 基因同源性区域	
<400> 86	
ggttgcggtc tcatacgctt gtagctgaac agctgatggt ttagagctag aaatagcaag	60
ttaaaataag gctagtccgt tatcaacttg aaaaagtggc accgagtcgg tgctttttac	120
tccatctgga tttgttcaga acgctcgggt gccgccgggc gttttttatc taaactagta	180
ggcccagtcg atacgtaaag actgggcctt ttaatacga ctactatag ggtcgacggc	240
caacgaggcc tcgaggatcc gatatcatgc atggcgcgcc accctgagac gaccctgatg	300
caggtgaccc ggatccacta gtccctgctgt ccgcaggtgc tttttttctt gaccacacg	360
acatTTTTTg agatttcgtc atttaattta aaacttecta ttgacggaca agcgattcct	420
ttgtattata gatcctgtgc ttcttagcgc atttttatta tggcggtgta gctcagctgg	480
ctagagcgta cggttcatac ccgtgaggtc gggggttcga tcccctccgc cgctatcctt	540
ttgattagaa cataaaagca aggcccgtt gtcaagcgggt taagacaccg ccctttcacg	600
gcggtaacac gggttcgaat cccgtacggg tcattctcga aacagcttt ctttaggaaa	660
gctgtTTTTt tgtgtcttca taaaattctg atgaaagaca tcgactttca agaaagtatg	720
cctctttgac gaataaagcg tcgaacgttt tatggaaacg acaacttctt ttgacaaaat	780
ttctttttca ccttcgctat aatgacaagc aacgaatc agtgaaatat cgtataatat	840
gaatttcttc tggcgtatgat ggggatataa agcattcagt acgatcccag gaggaatgaa	900
gatgcgaaaa ggtcacgtaa accaaatctt attgattaca gatggctgct caaatcacgg	960
ggaagatcca cttgcgattg cctcattggc aaaggaaca gggattacag tcaatgttat	1020
tggcattatg gaggaaaaca gacacgacca tgaagcaatg aaagaagttg aagggattgc	1080
tctcgcaggt ggaggcatcc atcaagttgt ctacgtccag cagttatctc aaaccgtaca	1140
aatggttaca aaaaaagcga tgacacaaac cttgcaaggt gttgtgaata agaattgca	1200
gcaaatactt ggcaaggaca ctgaaattga agagctgcc cctgataaac gcggggaagt	1260

gatggaagta gtcgatgagt taggagagac gttcatctt caagtgcttg tgcttgttga	1320
tactcaagag acccgttg	1338
<210> 87	
<211> 20	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增hag基因组区域	
<400> 87	
tgatcttgat gaaacgacgg	20
<210> 88	
<211> 22	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增hag基因组区域	
<400> 88	
taatcctgat attctgatcg cc	22
<210> 89	
<211> 19	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增degU基因组区域	
<400> 89	
atgatttaag gccgatggc	19
<210> 90	
<211> 21	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增degU基因组区域	
<400> 90	
atccgccttc agctactact t	21
<210> 91	
<211> 26	
<212> DNA	
<213> 人工序列	
<220>	

<223> 寡核苷酸:PCR扩增amyE基因组区域	
<400> 91	
ggatcatgaat aatctgctga atagac	26
<210> 92	
<211> 26	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增amyE基因组区域	
<400> 92	
gcgtgtacgt tttgagggcg tgcgcc	26
<210> 93	
<211> 26	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增枯草芽孢杆菌aprE基因组区域	
<400> 93	
gagctggcag atgaagccaa tattcc	26
<210> 94	
<211> 27	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增枯草芽孢杆菌aprE基因组区域	
<400> 94	
gtacgcgcat gaggaacgac aaataag	27
<210> 95	
<211> 24	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增vpr基因组区域	
<400> 95	
ctgtcaccca cttcccatta tgag	24
<210> 96	
<211> 24	
<212> DNA	
<213> 人工序列	

<220>	
<223> 寡核苷酸:PCR扩增vpr基因组区域	
<400> 96	
gtgaccgaag gctttccatc attg	24
<210> 97	
<211> 24	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增epr基因组区域	
<400> 97	
cttgtcatcg tcgtcgggat tcag	24
<210> 98	
<211> 24	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增epr基因组区域	
<400> 98	
gtgccaatca caaatgtagc cagc	24
<210> 99	
<211> 25	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增sigE基因组区域	
<400> 99	
gaggaggcat gatcggagtt cattc	25
<210> 100	
<211> 25	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增sigE基因组区域	
<400> 100	
ctcctccgtc attatagatc ggttc	25
<210> 101	
<211> 25	
<212> DNA	

<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增sigF基因组区域	
<400> 101	
gtgacgaatt atttgaaac agagg	25
<210> 102	
<211> 25	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增sigF基因组区域	
<400> 102	
cgacataatg atcgagatcg agctg	25
<210> 103	
<211> 25	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增spoIIE基因组区域	
<400> 103	
ctcaacaaca acaatcaaga ccgag	25
<210> 104	
<211> 25	
<212> DNA	
<213> 人工序列	
<220>	
<223> 寡核苷酸:PCR扩增spoIIE基因组区域	
<400> 104	
gatagtgaat cgaatcgagg cgtcc	25

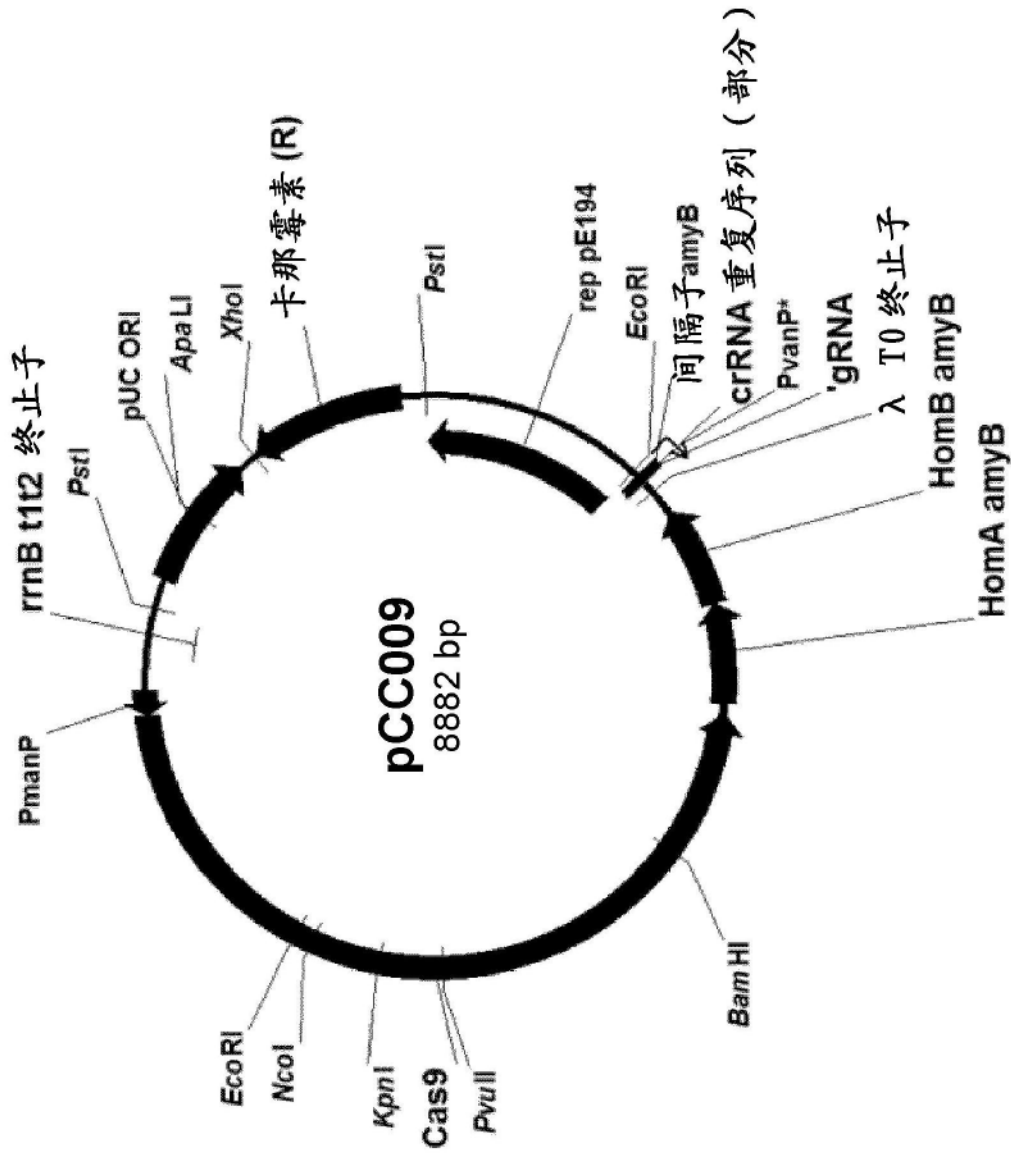


图1

```

-35          -10          TSS          SD
>SEQ_ID_028 AATTTTTTAA-CGTTGATA-CCGGTTAAATTTTAT-TTGACA-AAAATGGGCTCGTGTGGAGAAATAAATGTGGAGAAAGATTAACTAATAAG-GAGGACAAAAC
>SEQ_ID_034 AATTTTTTAA-CGTTGATA-CCGGTTAAATTTTAT-TTGACA--AAATGGGCTCGTGTGGAGAAATAAATGTGGAGAAAGATTAACTAATAAG-GAGGACAAAAC
>SEQ_ID_035 AATTTTTTAA-CGTTGATA-CCGGTTAAATTTTAT-TTGACA-AAAATGGGCTCGTGTGGAGAA--AAATGTGGAGAAAGATTAACTAATAAG-GAGGACAAAAC
>SEQ_ID_036 AATTTTTTAA-CGTTGATA-CCGGTTAAATTTTAT-TTGACA-AAAATGGGCTCGTGTGGAGAAATAAATGTGGAGAAAGATTAACTAATAAG-GAGGACAAAAC
>SEQ_ID_037 AATTTTTTAA-CGTTGATA-CCGGTTAAATTTTAT-TTGAACAAAAATGGGCTCGTGTGGAGAAATAAATGTGGAGAAAGATTAACTAATAAG-GAGGACAAAAC
>SEQ_ID_038 AATTTTTTAA-CGTTGATA-CCGGTTAAATTTTAT-TTGACA-AAAATGGGCTCGTGTGGAGAAATAAATGTGGAGAAAGATTAACTAATAAG-GAGGACAAAAC
>SEQ_ID_039 AATTTTTTAA-CGTTGATA-CCGGTTAAATTTTAT-TTGACA-AAAATGGGCTCGTGTGGAGAAATAAATGTGGAGAAAGATTAACTAATAAG-GAGGACAAAAC
>SEQ_ID_040 AATTTTTTAA-CGTTGATA-CCGGTTAAATTTTAT-TTGACA-AAAA-GGGCTCGTGTGGAGAAATAAATGTGGAGAAAGATTAACTAATAAG-GAGGACAAAAC
>SEQ_ID_029 AATTTTTTAA-CGTTGATA-CCGGTTAAATTTTAT-TTGACA-AAAATGGGCTCGTGTGGAGAAATAAATGTGGAGAAAGATTAACTAATAAG-GAGGACAAAAC
>SEQ_ID_041 AATTTTTTAA-CGTTGATA-CCGGTTAAATTTTAT-TTGACA-AAAATGGGCTCGTGTGGAGAA-AAAATGTGGAGAAAGATTAACTAATAAG-GAGGACAAAAC
>SEQ_ID_042 AATTTTTTAA-CGTTGATA-CCGGTTAAATTTTAT-TTGACA-AAAATGGGCTCGTGTGGAGAAATAAATGTGGAGAAAGATTAACTAATAAGGGGAGGACAAAAC
>SEQ_ID_043 AATTTTTTAA-CGTTGATA-CCGGTTAAATTTTAT-TTGACA-AAAATGGGCTCGTGTGGAGAAATAAATGTGGAGAAATAAATGTGGAGAAATAAATGTGGAGAAATAA
>SEQ_ID_044 AATTTTTTAA-CGTTGATA-CCGGTTAAATTTTAT-TTGACA-AAAATGGGCTCGTGTGGAGAAATAAATGTGGAGAAAGATTAACTAATAAG-GAGGACAAAAC
>SEQ_ID_045 AATTTTTTAA-CGTTGATA-CCGGTTAAATTTTAT-TTGACA-AAAATGGGCTCGTGTGGAGAAATAAATGTGGAGAAAGATTAACTAATAAG-GAGGACAAAAC
>SEQ_ID_046 AATTTTTTAA-CGTTGATA-CCGGTTAAATTTTAT-TTGACA--AAATGGGCTCGTGTGGAGAAATAAATGTGGAGAAAGATTAACTAATAAG-GAGGACAAAAC
>SEQ_ID_047 AATTTTTTAA-CGTTGATA-CCGGTTAAATTTTAT-TTGACA-AAAATGGGCTCGTGTGGAGAAAGATTAACTAATAAG-GAGGACAAAAC

```

图2

地衣芽孢杆菌中的基因删除

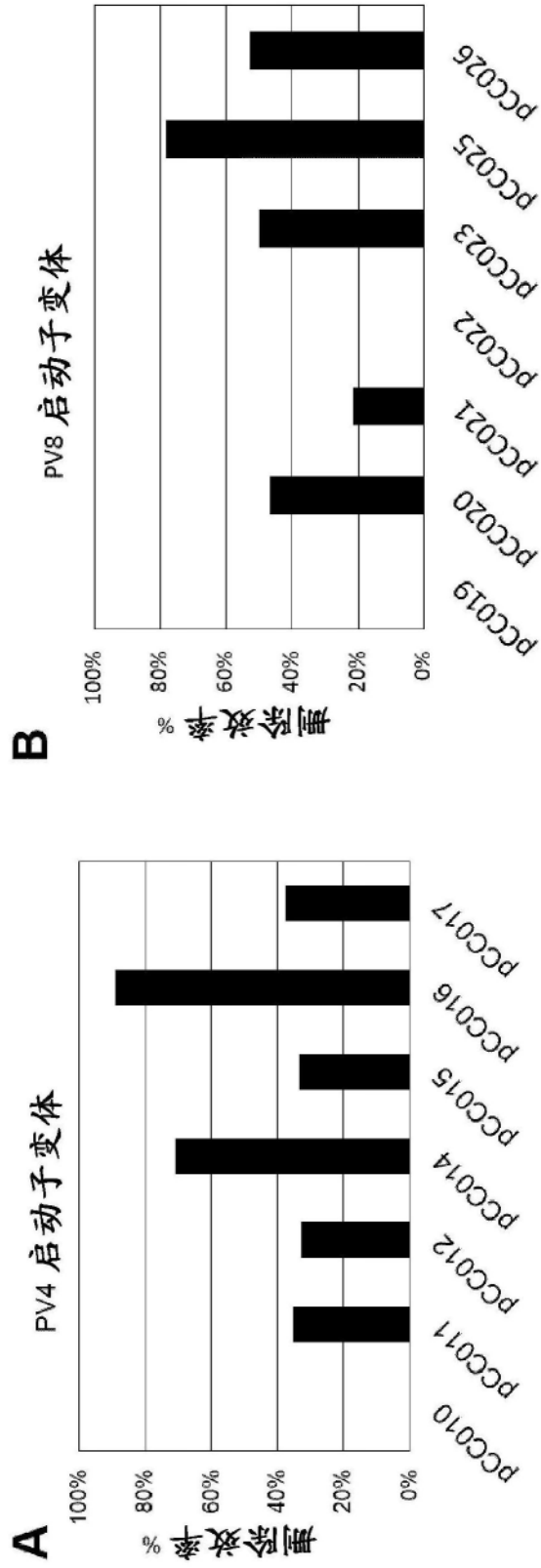


图3

地衣芽孢杆菌

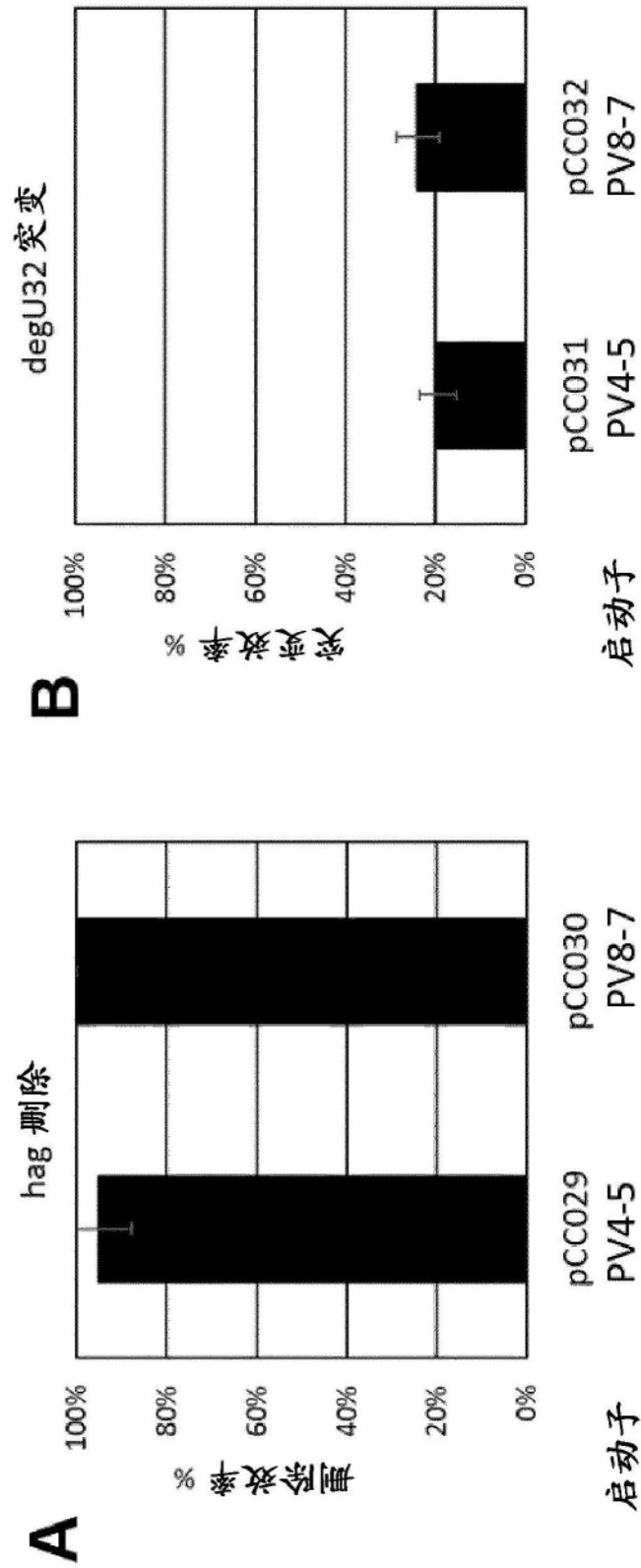


图4

枯草芽孢杆菌

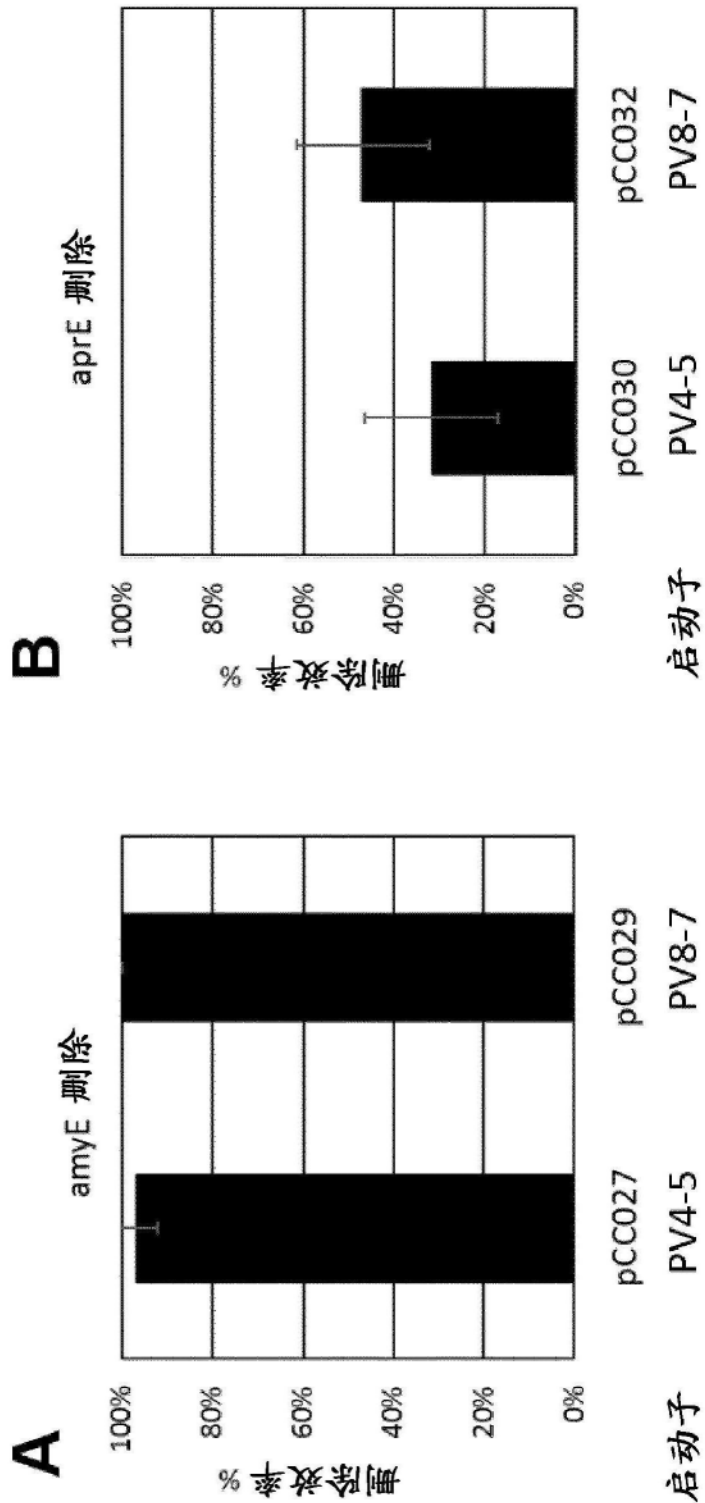


图5

地衣芽孢杆菌

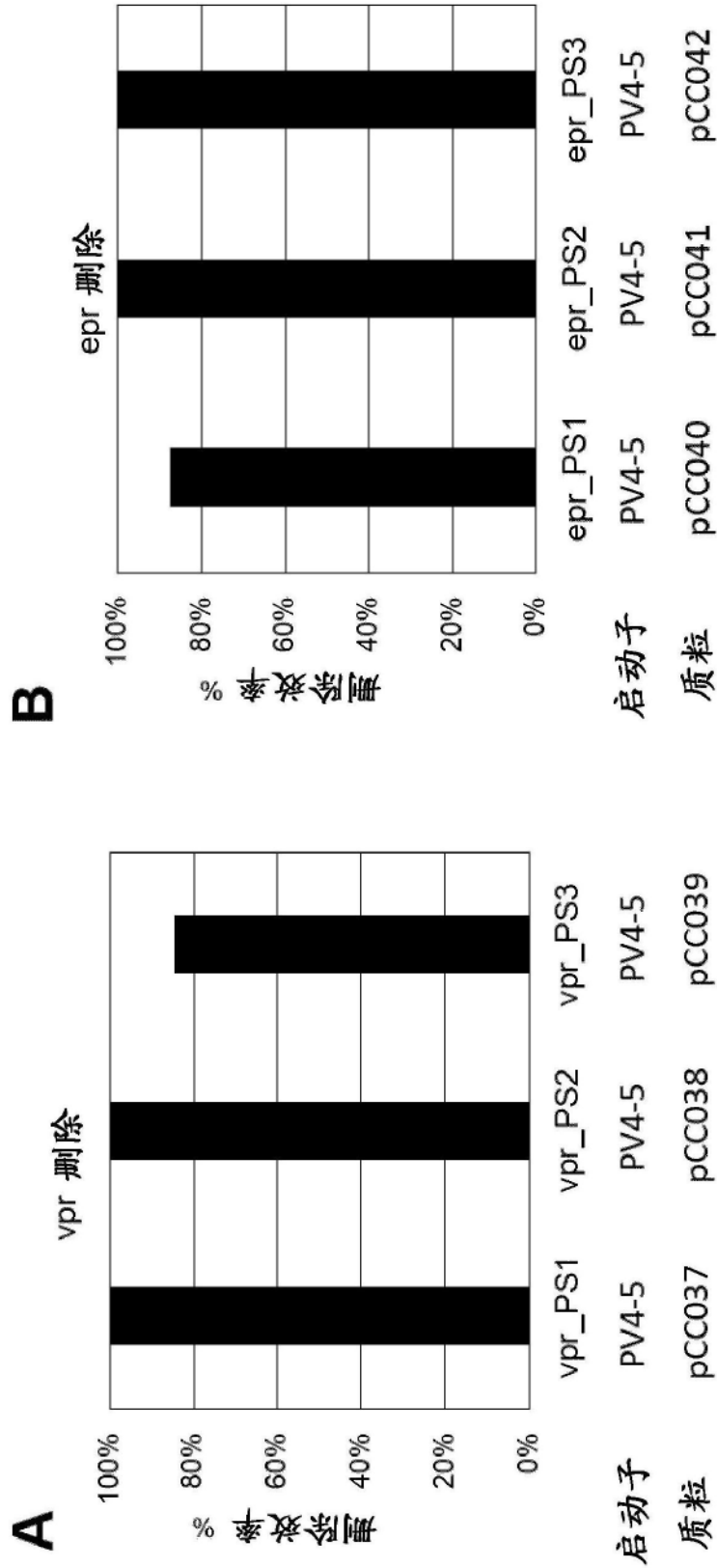


图6

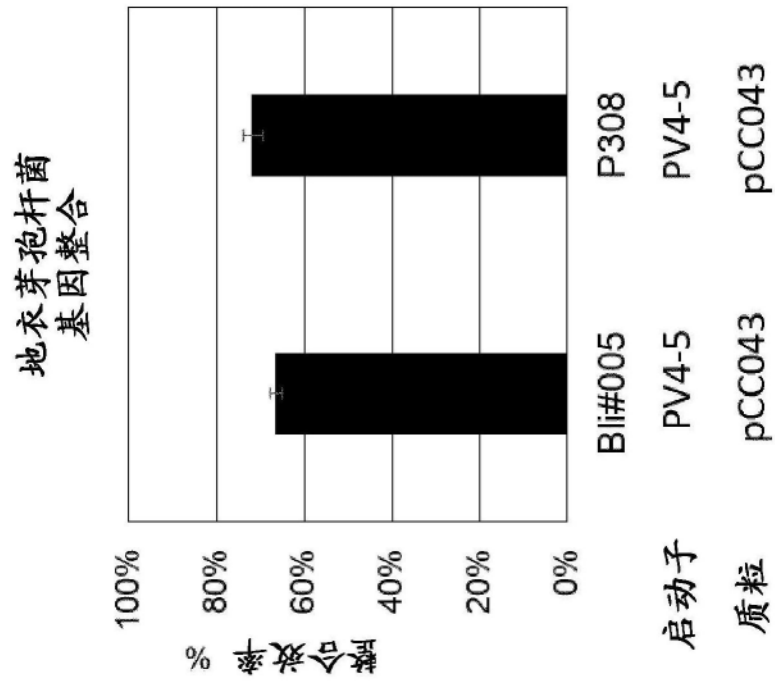


图7

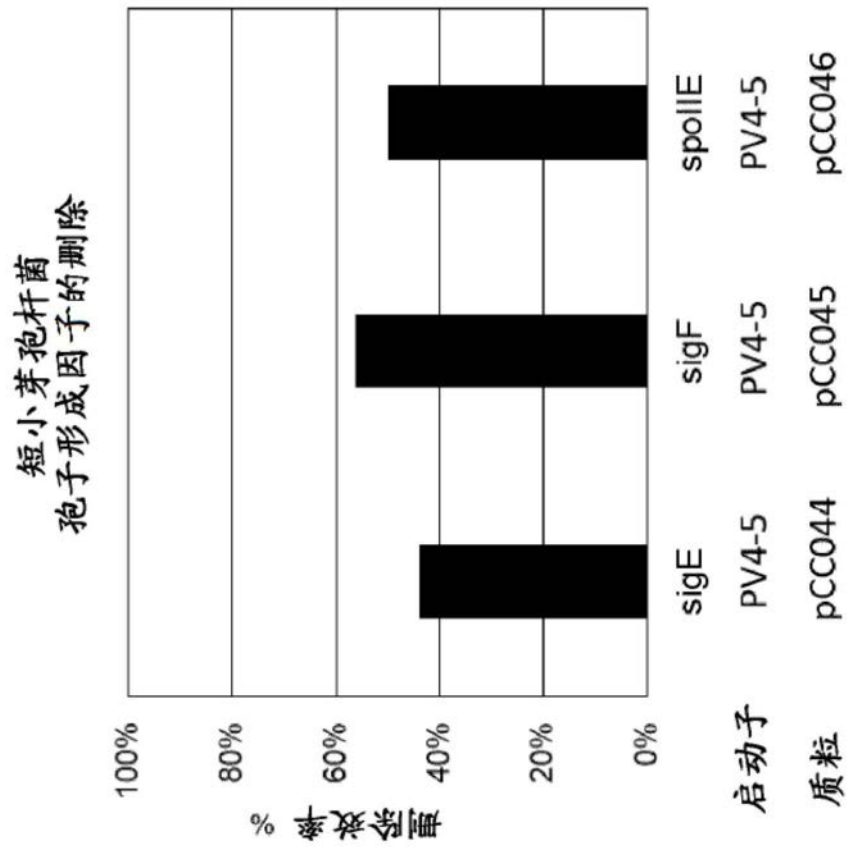


图8