



(12)发明专利申请

(10)申请公布号 CN 111492059 A

(43)申请公布日 2020.08.04

(21)申请号 201880081798.3

乌-范卢恩

(22)申请日 2018.11.20

(74)专利代理机构 北京东方亿思知识产权代理
有限责任公司 11258

(30)优先权数据

17209063.1 2017.12.20 EP

代理人 赵萌

(85)PCT国际申请进入国家阶段日

2020.06.18

(51)Int.Cl.

C12N 15/81(2006.01)

C12N 15/10(2006.01)

(86)PCT国际申请的申请数据

PCT/EP2018/081942 2018.11.20

(87)PCT国际申请的公布数据

W02019/012162 EN 2019.01.17

(71)申请人 帝斯曼知识产权资产管理有限公司

地址 荷兰海尔伦

(72)发明人 瑞内·维尔瓦尔

保卢斯·佩特鲁斯·德瓦尔

弗朗辛·马鲁什卡·乔安娜·德勒

权利要求书1页 说明书22页

序列表29页 附图10页

(54)发明名称

在宿主细胞中进行基因组编辑的方法

(57)摘要

本发明涉及分子生物学和细胞生物学领域。
更具体地,本发明涉及一种基因组编辑系统。

1. 一种用于在宿主细胞中进行基因组编辑的方法,所述方法包括:

a) 使宿主细胞与以下接触:

i) 表达构建体,所述表达构建体包含可操作地连接至诱导型启动子的在表达时对所述宿主细胞的活力具有负面影响的多核苷酸,

ii) 功能性异源基因组编辑酶,或能够在所述宿主细胞中表达功能性异源基因组编辑酶的表达构建体,

(iii) 引导多核苷酸,或能够在所述宿主细胞中表达引导多核苷酸的表达构建体,以及任选地,

(iv) 外源多核苷酸,

b) 在诱导基因组编辑的条件下培养所述宿主细胞,以及

c) 在诱导对所述宿主细胞的活力具有负面影响的所述多核苷酸的表达的条件下培养所述宿主细胞;

其中至少能够在所述宿主细胞中表达功能性异源基因组编辑酶的表达构建体或能够在所述宿主细胞中表达引导多核苷酸的表达构建体位于包含在表达时对所述宿主细胞的活力具有负面影响的多核苷酸的所述表达构建体上。

2. 根据权利要求1所述的方法,其中所述宿主细胞是原核宿主细胞、真核宿主细胞、海洋真核生物、微藻或藻类宿主细胞。

3. 根据权利要求2所述的方法,其中所述宿主细胞是真核宿主细胞,以及优选地是真菌宿主,更优选地是酵母或丝状真菌宿主细胞。

4. 根据权利要求3所述的方法,其中所述酵母细胞是酵母属(*Saccharomyces*) 宿主细胞,优选地酿酒酵母(*Saccharomyces cerevisiae*) 宿主细胞。

5. 根据权利要求1至4中任一项所述的方法,其中包含在表达时对所述宿主细胞的活力具有负面影响的多核苷酸的所述表达构建体存在于附加型实体上,优选地质粒上。

6. 根据权利要求1至5中任一项所述的方法,其中所述基因组编辑酶是Cas样酶。

7. 根据权利要求1至6中任一项所述的方法,其中所述诱导型启动子是铜诱导型启动子,优选地CUP1启动子,或半乳糖诱导型启动子,优选地GAL10启动子。

8. 根据权利要求8所述的方法,其中所述CUP1启动子与SEQ ID NO:20具有至少80%的序列同一性和/或其中所述GAL10启动子与SEQ ID NO:19具有至少80%的序列同一性。

9. 根据权利要求1至8中任一项所述的方法,其中在表达时对所述宿主细胞的活力具有负面影响的所述多核苷酸与SEQ ID NO:21具有至少80%的序列同一性。

10. 一种通过根据权利要求1至9中任一项所述的方法能够获得或通过根据权利要求1至9中任一项所述的方法获得的宿主细胞。

11. 一种用于生产感兴趣的化合物的方法,所述方法包括在有利于所述感兴趣的化合物的表达的条件下培养根据权利要求10所述的宿主细胞;以及任选地分离和/或纯化所述感兴趣的化合物。

12. 一种用于生产感兴趣的化合物的方法,所述方法包括执行根据权利要求1至9中任一项所述的方法,以及随后在有利于所述感兴趣的化合物的表达的条件下培养所述宿主细胞;以及任选地分离和/或纯化所述感兴趣的化合物。

在宿主细胞中进行基因组编辑的方法

技术领域

[0001] 本发明涉及分子生物学和细胞生物学领域。更具体地,本发明涉及一种基因组编辑系统。

背景技术

[0002] 多核苷酸引导的核酸酶系统,也称为多核苷酸引导的基因组编辑系统,是已用于基因组编辑的强大工具,其中众所周知的示例是CRISPR/Cas9系统。该工具至少需要多核苷酸引导的核酸酶(多核苷酸引导的基因组编辑酶)(诸如Cas9)和引导多核苷酸(诸如引导RNA),所述引导多核苷酸使得基因组编辑酶能够靶向DNA的特定序列。此外,为了以精确方式编辑基因组,可能需要供体多核苷酸,诸如供体DNA,尤其是当依赖于用于在基因组DNA中的期望位点处进行精确基因组编辑的同源重组而不是依赖于通过随机修复过程(诸如非同源末端连接)的修复时。

[0003] 这些需要特征中的若干可以在(附加型)表达构建体上引入细胞中。在已经执行期望的基因组编辑之后,优选的是,尤其是在工业规模的发酵之前,从经编辑的细胞中去除此类(附加型)表达构建体。使用本领域中已知的方法进行反选择并不总是非常有效或便利,因为这些方法可以是耗时的并且具有不同的效率(从0%到100%),从而使得有时需要重复去除(附加型)表达构建体的耗时循环。因此,需要一种改进的系统,以在基因组编辑步骤之后从宿主细胞中去除此类(附加型)表达构建体。

附图说明

[0004] 图1描绘了用于在INT70基因座整合HXT11/2表达盒(SparTDH3p-HXT11/2N366T-EFM1t)的策略。供体DNA的一部分的5'和3'代表供体DNA与INT70基因座的同源性,d和3代表50bp的合成DNA接头序列。

[0005] 图2描绘了琼脂糖凝胶图片,以通过分析PCR片段来确认HXT11/2表达盒在INT70基因座处的整合。

[0006] 图3描绘了质粒pDB1371的丢失效率。在30°C下生长2天后,将每种条件下的20个菌落划线到YEPHd-G418和YEPHd,以对质粒丢失效率进行评分。在X轴上指示了在YEPHd或YEPHG液体培养基上的不同孵育时间。Y轴代表在每种生长条件下最初划线的20个菌落中,能够在YEPHd-G418平板上生长的菌落数(没有丢失质粒pDB1371)。

[0007] 图4描绘了质粒pDB1371(CP-71-HXT)和pCSN061(CP-61-HXT)的质粒丢失效率。在YEPHG琼脂平板上在30°C下生长2天后,将40个菌落划线到YEPHd-G418和YEPHd上,以确定含Cas9质粒的丢失效率。Y轴代表在最初划线的40个菌落中,能够在YEPHd-G418平板上生长的菌落数(没有丢失质粒pDB1371)。

[0008] 图5描绘了通过诱导pDB1372上存在的CUP1p-GIN11(M86)构建体的质粒丢失效率。在30°C下生长2天后,对每种条件下的菌落数进行计数,以对相较于无诱导条件的相对质粒丢失效率进行评分。Y轴代表能够在YEPHd-G418平板上生长的菌落的百分比(没有丢失质粒

pDB1372)。

[0009] 图6描绘了表达为了在酿酒酵母 (*S.cerevisiae*) 中表达而经密码子对优化的CAS9的单拷贝 (CEN/ARS) 载体pCSN061的载体图谱 (SEQ ID NO:18)。KanMX标记存在于载体上。

[0010] 图7描绘了pRN1120-RFP-gRNA (A) 的载体图谱, pRN1120-RFP-gRNA (A) 是基于pRS305的含natMX标记的穿梭载体, 其具有2微米起点和表达盒TDH3p-RFP-PG11t。

[0011] 图8描绘了pDB1371, 即含有pGAL10-GIN11 (M86) 多核苷酸序列的pCSN061的载体图谱。

[0012] 图9描绘了pDB1372, 即含有pCUP1-GIN11 (M86) 核苷酸序列的pCSN061的载体图谱。

[0013] 图10描绘了pDB1368的载体图谱, pDB1368是具有表达盒SparTDH3p-HXT 11/2N366T-EFM1t的克隆载体。

[0014] 序列说明

[0015] SEQ ID NO:1列出了载体pRN1120-RFP-gRNA (A) 的核苷酸序列。

[0016] SEQ ID NO:2列出了合成表达盒cFS0017 (pGAL10-GIN11 (M86)) 的核苷酸序列。

[0017] SEQ ID NO:3列出了合成表达盒cFS0018 (pCUP1-GIN11 (M86)) 的核苷酸序列。

[0018] SEQ ID NO:4列出了载体pDB1371的核苷酸序列。

[0019] SEQ ID NO:5列出了载体pDB1372的核苷酸序列。

[0020] SEQ ID NO:6列出了载体pDB1368的核苷酸序列。

[0021] SEQ ID NO:7列出了用于进行延伸PCR以将KpnI限制位点添加至cFS0017的正向引物的核苷酸序列。

[0022] SEQ ID NO:8列出了用于进行延伸PCR以将NcoMIV限制位点添加至cFS0017的反向引物的核苷酸序列。

[0023] SEQ ID NO:9列出了INT70 gRNA gBLOCK的核苷酸序列。

[0024] SEQ ID NO:10列出了用于使用pDB1368作为模板获得供体DNA PCR片段 (int70 [5']-conD-HXT11/2-con3-int70[3']) 的正向引物的核苷酸序列。

[0025] SEQ ID NO:11列出了用于使用pDB1368作为模板获得供体DNA PCR片段 (int70 [5']-conD-HXT11/2-con3-int70[3']) 的反向引物的核苷酸序列。

[0026] SEQ ID NO:12列出了用于使用pRN1120-RFP-gRNA (A) (SEQ ID NO:1) 作为模板获得gRNA受体质粒骨架的正向引物的核苷酸序列。

[0027] SEQ ID NO:13列出了用于使用pRN1120-RFP-gRNA (A) (SEQ ID NO:1) 作为模板获得gRNA受体质粒骨架的反向引物的核苷酸序列。

[0028] SEQ ID NO:14列出了用于使用INT70 gBLOCK (SEQ ID NO:9) 作为模板获得引导RNA PCR片段 (gRNA-INT70) 的正向引物的核苷酸序列。

[0029] SEQ ID NO:15列出了用于使用INT70 gBLOCK (SEQ ID NO:9) 作为模板获得引导RNA PCR片段 (gRNA-INT70) 的反向引物的核苷酸序列。

[0030] SEQ ID NO:16列出了用于确认HXT11/2表达盒在INT70基因座处的正确组装和整合的正向引物的核苷酸序列。

[0031] SEQ ID NO:17列出了用于确认HXT11/2表达盒在INT70基因座处的正确组装和整合的反向引物的核苷酸序列。

[0032] SEQ ID NO:18列出了载体pCSN061的核苷酸序列。

[0033] SEQ ID NO:19列出了pGAL10启动子的核苷酸序列。

[0034] SEQ ID NO:20列出了pCUP1启动子的核苷酸序列。

[0035] SEQ ID NO:21列出了GIN11 (M86)的核苷酸序列。

[0036] 发明详述

[0037] 发明人已经发现了使用生长抑制序列GIN11 (M86) (Akada等人, *Yeast*, 第19卷, 第393-402页, 2002) 针对含有Cas9表达盒的质粒进行主动选择的有效方法。该多核苷酸序列的过表达导致强烈的生长抑制效应。GIN11是保守的子端粒X元件的一部分, 其在染色体复制期间很重要。由于先前发现GIN11含有保守的自主复制序列 (ARS), 该保守的自主复制序列可阻碍染色体整合 (Kawahata等人, *Yeast*, 第15卷, 第1期, 第1-10页, 1999), 因此分离了丧失复制活性但在过表达时保留生长抑制效应的突变序列:GIN11 (M86)。由于多核苷酸序列不编码蛋白质, 因此丧失其生长抑制效应的突变体上的机会减少。耦合到诱导型启动子, 带有诱导型GIN11 (M86) 序列的整合构建体和附加表达的质粒显示出有效的基因丢失。该方法可以方便地用于基因组编辑方法中。

[0038] 因此, 在第一方面中, 本发明涉及一种用于在宿主细胞中进行基因组编辑的方法, 所述方法包括:

[0039] a) 使宿主细胞与以下接触:

[0040] i) 表达构建体, 所述表达构建体包含可操作地连接至诱导型启动子的在表达时对所述宿主细胞的活力具有负面影响的多核苷酸,

[0041] ii) 功能性异源基因组编辑酶, 或能够在所述宿主细胞中表达功能性异源基因组编辑酶的表达构建体,

[0042] (iii) 引导多核苷酸, 或能够在所述宿主细胞中表达引导多核苷酸的表达构建体, 以及任选地,

[0043] (iv) 外源多核苷酸,

[0044] b) 在诱导基因组编辑的条件下培养所述宿主细胞, 以及

[0045] c) 在诱导对所述宿主细胞的活力具有负面影响的所述多核苷酸的表达的条件下培养所述宿主细胞;

[0046] 其中至少能够在所述宿主细胞中表达功能性异源基因组编辑酶的表达构建体或能够在所述宿主细胞中表达所述引导多核苷酸的表达构建体位于包含在表达时对所述宿主细胞的活力具有负面影响的多核苷酸的所述表达构建体上。

[0047] 所述用于基因组编辑的方法、所述宿主细胞、所述包含在表达时对宿主细胞的活力具有负面影响的多核苷酸的表达构建体和所述诱导型启动子在本文中被称为根据本发明的用于基因组编辑的方法、根据本发明的宿主细胞、根据本发明的包含在表达时对宿主细胞的活力具有负面影响的多核苷酸的表达构建体, 和根据本发明的诱导型启动子。

[0048] 该方法的基本原理是执行基因组编辑过程, 并在基因组编辑已发生并且任选地选择已发生期望的基因组编辑的细胞之后, 在诱导对宿主细胞的活力具有负面影响的多核苷酸的表达的条件下培养一个或多个所述细胞, 所述细胞随后丢失携带对宿主细胞的活力具有负面影响的多核苷酸的表达构建体 (例如质粒)。在携带对宿主细胞的活力具有负面影响的多核苷酸的表达构建体丢失之前, 该表达构建体可以附加体形式存在或者可整合在宿主细胞的基因组中。

[0049] 在表达时对宿主细胞的活力具有负面影响的多核苷酸可以是对宿主细胞具有此类效应的任何多核苷酸。本领域的技术人员知道如何鉴定此类多核苷酸或如何调整此类多核苷酸。本领域的技术人员可以例如使用生长抑制多核苷酸GIN11 (M86) 的序列 (SEQ ID NO:21), 并调整该序列以用于除酵母外的其他生物中。取决于期望丢失的表达构建体, 在表达时对宿主细胞的活力具有负面影响的多核苷酸可以位于能够在宿主细胞中表达功能性异源基因组编辑酶的表达构建体或能够表达引导多核苷酸的表达构建体上。在表达时对宿主细胞的活力具有负面影响的多核苷酸也可以存在于能够在宿主细胞中表达功能性异源基因组编辑酶的表达构建体和能够在表达时能够表达引导多核苷酸的表达构建体两者上。可行的是, 编码功能性异源基因组编辑酶的多核苷酸和编码引导多核苷酸或其一部分的多核苷酸可以存在于单个表达构建体上。在此类情况下, 在表达时对宿主细胞的活力具有负面影响的多核苷酸将存在于该单一构建体上。

[0050] 在表达时对宿主细胞的活力具有负面影响的多核苷酸被选择性地表达, 即当在诱导对宿主细胞的活力具有负面影响的多核苷酸的表达的条件下培养宿主细胞时选择性地表达。此类选择性表达是本领域的技术人员已知的, 例如可以使用仅在选择性条件下有活性的启动子或可以使用选择性转录因子。

[0051] 任选地, 在根据本发明的方法中, 存在外源多肽。此类多肽通常是供体多核苷酸, 该供体多核苷酸将在基因组编辑步骤中引入受体多核苷酸, 诸如宿主细胞的基因组中。

[0052] 根据本发明的方法可以作为单个方法来执行, 即连续地执行步骤 (a) 至 (c), 或者可以在步骤之间有暂停的情况下单独地执行步骤。可以引入一些附加步骤, 例如选择感兴趣的细胞, 在所述感兴趣的细胞中期望的基因组编辑已在步骤 (b) 之后和步骤 (c) 之前或在步骤 (c) 之后发生。

[0053] 对活力的负面影响在本文中被解释为当对宿主细胞的活力具有负面影响的多核苷酸表达时的宿主细胞比对宿主细胞的活力具有负面影响的多核苷酸不表达的细胞活力更低。优选地, 当表达对宿主细胞的活力具有负面影响的多核苷酸时, 宿主细胞的生长受到损害。因此, 当对宿主细胞的活力具有负面影响的多核苷酸 (包括携带该多核苷酸的构建体) 从宿主细胞中丢失时, 宿主细胞具有优势。

[0054] 使根据本发明的宿主细胞与根据本发明的构建体和/或多核苷酸接触可以以本领域的技术人员已知的任何方式执行, 所述方式例如但不限于细胞或细胞的部分 (例如原生质体) 的转染或转化。本领域的技术人员将理解, 使根据本发明的宿主细胞与根据本发明的构建体和/或多核苷酸接触优选地导致并且因此优选地等同于将根据本发明的构建体和/或多核苷酸引入根据本发明的宿主细胞中。

[0055] 根据本发明的引导多核苷酸可以是本领域技术人员已知的任何引导多核苷酸。此类引导多核苷酸可以是DNA或RNA。根据本发明的引导多核苷酸至少包含能够与靶多核苷酸杂交并且能够指导异源基因组编辑系统与靶多核苷酸的序列特异性结合的引导序列。所述引导多核苷酸是根据上文列出的多核苷酸的一般定义的多核苷酸; 优选的引导多核苷酸包括核糖核苷酸, 更优选的引导多核苷酸是RNA (引导RNA)。引导RNA通常包含引导序列 (crRNA) 和引导多核苷酸结构组分 (参见例如DiCarlo等人, *Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems*. *Nucleic Acids Res.* 2013; 41 (7): 4336-4)。引导序列在本文中也称为靶序列并且本质上是靶多核苷酸的互补物, 使得该

引导多核苷酸能够与靶多核苷酸杂交,优选地在宿主细胞中在生理条件下与靶多核苷酸杂交。

[0056] 根据本发明的功能性异源基因组编辑酶可以是本领域的技术人员已知的用于本发明的所有实施方式中的任何合适的功能性基因组编辑酶,并且包括但不限于:转录激活因子样效应核酸酶(TALENs,Gaj等人,Trends in Biotechnology,2013,第31卷,第7期,397-405)、锌指核酸酶(ZFNs,Gaj等人,Trends in Biotechnology,2013,第31卷,第7期,397-405)、大范围核酸酶例如I-SceI(Cabaniols和Paques,Methods Mol Biol.2008;435:31-45)、RNA引导的内切核酸酶如CRISPR/Cas(Mali等人,Science,2013年2月15日;339(6121):823-6;Cong等人,Science.2013年2月15日;339(6121):819-23)、CRISPR/Cpf1(Zetsche等人,Cell.2015年10月22日;163(3):759-71)或Cas9直系同源物(由Mitsunobu等人综述,Trends Biotechnol.2017年10月;35(10):983-996)、具有改进性质的工程化的Cas9s例如切口酶、核酸酶死亡Cas9(dCas9)或具有改变的PAM偏好的Cas9(由Mitsunobu等人综述,Trends Biotechnol.2017年10月;35(10):983-996)、基于dCas9的转录激活物或阻遏物(由Mitsunobu等人综述,Trends Biotechnol.2017年10月;35(10):983-996)、脱氨酶介导的碱基编辑器(Komor等人,Nature.2016年5月19日;533(7603):420-424);由Hess等人综述,Mol Cell.2017年10月5日;68(1):26-43;Gaudelli等人,Nature.2017年11月23日;551(7681):464-471),或用于引入表观遗传修饰(如组蛋白乙酰化、去乙酰化或去甲基化或DNA甲基化/去甲基化)的CRISPR系统(由Montalbano等人综述,Mol Cell.2017年10月5日;68(1):44-59)。功能性基因组编辑系统是本领域的技术人员已知的,并且本领域的技术人员知道如何选择和使用合适的系统。优选的功能性基因组编辑系统是RNA或DNA引导的核酸酶系统,优选地是RNA或DNA引导的DNA核酸酶系统,更优选地是不依赖于原型间隔区相邻基序(PAM)的RNA或DNA引导的DNA核酸酶系统。

[0057] 在根据本发明的方法中,至少能够在所述宿主细胞中表达所述功能性异源基因组编辑酶的表达构建体或能够在所述宿主细胞中表达所述引导多核苷酸的表达构建体位于包含在表达时对所述宿主细胞的活力具有负面影响的所述多核苷酸的所述表达构建体上。可以原样提供引导多核苷酸或基因组编辑酶。然而,应始终通过编码引导多核苷酸或基因组编辑酶的表达构建体来提供引导多核苷酸和基因组编辑酶中的一者。本领域的技术人员将理解,多于一种表达构建体可以携带在表达时对宿主细胞的活力具有负面影响的多核苷酸;例如能够在宿主细胞中表达功能性异源基因组编辑酶的表达构建体和能够在宿主细胞中表达引导多核苷酸的表达构建体两者都可以携带在表达时对宿主细胞的活力具有负面影响的多核苷酸。其他实施方式也是可行的,例如对编码引导多核苷酸的表达构建体的文库反复地使用根据本发明的方法或使用当在上表达时对宿主细胞的活力具有负面影响的多核苷酸。本领域的技术人员将理解根据本发明的方法的多种和多重选择。

[0058] 优选地,在根据本发明的方法中,原核宿主细胞、真核宿主细胞、海洋真核生物、微藻类、原生生物或藻类宿主细胞。

[0059] 优选地,在根据本发明的方法中,宿主细胞是真核宿主细胞,并且优选地是真菌宿主细胞,更优选地是酵母或丝状真菌宿主细胞。

[0060] 优选地,在根据本发明的方法中,宿主细胞是酵母细胞,并且优选地是酵母属(*Saccharomyces*)宿主细胞,优选地是酿酒酵母(*Saccharomyces cerevisiae*)宿主细胞。

[0061] 在“一般定义”部分中列出了根据本发明的优选宿主细胞。

[0062] 优选地,在根据本发明的方法中,包含在表达时对宿主细胞的活力具有负面影响的多核苷酸的表达构建体存在于附加型实体上,该附加型实体优选地为质粒。

[0063] 优选地,在根据本发明的方法中,基因组编辑酶是Cas样酶。Cas样酶被解释为多核苷酸引导的核酸内切酶,例如但不限于RNA引导的核酸内切酶,如CRISPR/Cas (Mali等人,2013;Cong等人,2013)或CRISPR/Cpf1 (Zetsche等人,2015)。

[0064] 在根据本发明的方法中,诱导型启动子可以是本领域的技术人员已知的任何合适的诱导型启动子。此类诱导型启动子可以是营养物(例如氨、葡萄糖、半乳糖)依赖性启动子、金属依赖性启动子、pH依赖性启动子或光依赖性启动子。诱导型启动子可由激活物和/或阻遏物以顺式模式或反式模式调节。优选地,在根据本发明的方法中,所述诱导型启动子是铜诱导型启动子,优选地CUP1启动子或半乳糖诱导型启动子,优选地GAL10启动子。当启动子是CUP1启动子时,CUP1启动子优选与SEQ ID NO:20具有至少80%的序列同一性。更优选地,CUP1启动子与SEQ ID NO:20具有至少81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或至少99%的序列同一性,最优选地,CUP1启动子包含SEQ ID NO:20或由SEQ ID NO:20组成。当启动子是GAL10启动子时,GAL10启动子优选地与SEQ ID NO:19具有至少80%的序列同一性。更优选地,GAL10启动子与SEQ ID NO:19具有至少81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或至少99%的序列同一性。

[0065] 优选地,在根据本发明的方法中,在表达时对所述宿主细胞的活力具有负面影响的所述多核苷酸与SEQ ID NO:21具有至少80%的序列同一性。

[0066] 在第二方面中,本发明提供了一种能够通过根据本发明的方法获得或通过根据本发明的方法获得的宿主细胞。该第二方面的特征优选地是本发明的第一方面的特征。

[0067] 在第三方面中,本发明提供了一种生产感兴趣的化合物的方法,该方法包括在有利于所述感兴趣的化合物的表达的条件下培养根据本发明的第二方面的宿主细胞,以及任选地分离和/或纯化所述感兴趣的化合物。感兴趣的化合物可以是任何感兴趣的化合物,并且优选地是在“一般定义”部分中给出的感兴趣的化合物。

[0068] 在第四方面中,本发明提供了一种用于生产感兴趣的化合物的方法,该方法包括执行根据本发明的第一方面的方法;随后在有利于感兴趣的化合物的表达的条件下培养所述宿主细胞;以及任选地分离和/或纯化感兴趣的化合物。

[0069] 一般定义

[0070] 在整个说明书和所附权利要求书中,词语“包括”、“包含”和“具有”将被包含性地解释。也就是说,在上下文允许的情况下,这些词语旨在传达可包括未具体叙述的其他要素或整数。

[0071] 不使用数量词修饰时在本文中用于指代一个/种或多于一个/种(即,一个/种或至少一个/种)的语法对象。举例来说,“要素”可以表示一个/种要素或多于一个/种要素。

[0072] 当与数值(例如,约10)结合使用时,词语“约”或“近似”优选地意味值可以是给定值(10)加减该值的1%的范围。

[0073] 多核苷酸在本文中是指任何长度或确定的特定长度范围或长度的脱氧核糖核苷酸或核糖核苷酸或它们的混合物或类似物的核苷酸的聚合形式。多核苷酸可以具有任何三

维结构,并且可以执行任何已知或未知的功能。以下是多核苷酸的非限制性示例:基因或基因片段的编码或非编码区、根据连锁分析定义的一个或多个基因座、外显子、内含子、信使RNA (mRNA)、转移RNA (tRNA)、核糖体RNA (rRNA)、短干扰RNA (siRNA)、短发夹RNA (shRNA)、微RNA (miRNA)、核酶、cDNA、重组多核苷酸、支链多核苷酸、质粒、构建体、载体、具有任何序列的分离DNA、具有任何序列的分离RNA、核酸探针、寡核苷酸和引物。多核苷酸可包含天然和非天然核苷酸,并且可包含一个或多个修饰的核苷酸,诸如甲基化核苷酸和核苷酸类似物或核苷酸等同物,其中核苷酸类似物或等同物被定义为具有经修饰的碱基、和/或经修饰的骨架、和/或非天然的核苷间键、或这些修饰的组合物残基。根据需要,可以在组装多核苷酸之前或之后引入对核苷酸结构的修饰。多核苷酸可以在聚合后进一步修饰,诸如通过与标记化合物缀合。

[0074] 通常,密码子优化是指通过以下方式来修饰核酸序列以增强在感兴趣的宿主细胞中的表达的过程:用在所述宿主细胞的基因中更频繁或最频繁使用的密码子替换天然序列的至少一个密码子(例如,多于1个、2个、3个、4个、5个、10个、15个、20个、25个、50个或更多个密码子),同时保持天然氨基酸序列。不同物种对特定氨基酸的某些密码子表现出特定偏好。密码子偏好(生物体之间密码子使用的差异)通常与信使RNA (mRNA)的翻译效率相关,而这继而据信又尤其取决于被翻译的密码子的性质和特定转移RNA (tRNA)分子的可用性。所选择的tRNA在细胞中的主导性通常反映了肽合成中最常使用的密码子。因此,可以基于密码子优化来定制基因以实现给定生物中的最佳基因表达。密码子使用表是容易例如在“密码子使用数据库(Codon Usage Database)”中获得的,并且这些表可以以多种方式进行调整。参见例如Nakamura, Y.等人,2000。用于对特定序列进行密码子优化以在特定宿主细胞中表达的计算机算法可是可得的,诸如Gene Forge (Aptagen; Jacobus, PA)也是可得的。优选地,编码Cas蛋白的序列中的一个或多个密码子(例如,1个、2个、3个、4个、5个、10个、15个、20个、25个、50个或更多个,或所有密码子)对应于特定氨基酸的最频繁使用的密码子。用于密码子优化的优选方法描述于W02006/077258和W02008/000632中)。W02008/000632涉及密码子对优化。密码子对优化是这样的一种方法,在所述方法中,编码多肽的核苷酸序列已经关于其密码子使用(特别是所使用的密码子对)被修饰,以获得编码所述多肽的核苷酸序列的改善表达和/或所编码多肽的改进产生。密码子对被定义为编码序列中的一组两个紧接三联体(密码子)。

[0075] 在具有5'-帽的RNA分子中,7-甲基鸟苷酸残基位于RNA的5'末端上(诸如通常在真核生物中的mRNA中)。RNA聚合酶II (Pol II)在真核生物中转录mRNA。信使RNA加帽通常发生如下:由RNA末端磷酸酶去除mRNA转录物的最末端5'磷酸基团,留下两个末端磷酸酯。由鸟苷酰转移酶将单磷酸鸟苷(GMP)加入到转录物的末端磷酸酯,从而在转录物末端处留下5'-5'三磷酸酯连接的鸟嘌呤。最后,用甲基转移酶甲基化该末端鸟嘌呤的7-氮。本文中术语“不具有5'-帽”用于指具有例如5'-羟基而不是5'-帽的RNA。例如,此类RNA可称为“未加帽的RNA”。未加帽的RNA可以在转录后在细胞核中更好地积累,因为5'-加帽的RNA经受核输出。

[0076] 核酶是指这样的一种或多种RNA序列,所述一种或多种RNA序列形成能够在特定位点处切割RNA的二级、三级和/或四级结构。核酶包括“自切割核酶或自加工核酶”,其能够在相对于核酶序列的c/s位点处切割RNA(即,自动催化或自切割的)。核酶溶核活性的一般性

质是本领域的技术人员已知的。用于RNA引导的核酸酶系统(诸如CRISPR/Cas)的引导RNA的生产中的自加工核酶的使用尤其由Gao等人, *Integr Plant Biol.* 2014年4月;56(4):343-9描述。

[0077] 核苷酸类似物或等同物通常包含经修饰的骨架。此类骨架的示例由吗啉代骨架、氨基甲酸酯骨架、硅氧烷骨架、硫化物骨架、亚砷骨架和砷骨架、甲酰乙酰基(formacetyl)和硫代甲酰乙酰基骨架、亚甲基甲酰乙酰基骨架、核乙酰基(riboacetyl)骨架、含烯烃的骨架、氨基磺酸酯骨架、磺酸酯骨架和磺酰胺骨架、亚甲基亚氨基骨架和亚甲基胍基骨架,以及酰胺骨架提供。进一步优选的是,骨架中残基之间的键不包含磷原子,诸如由短链烷基或环烷基核苷间键、混合的杂原子和烷基或环烷基核苷间键、或一个或多个短链杂原子或杂环核苷间键形成的键。

[0078] 优选的核苷酸类似物或等同物包含具有经修饰的聚酰胺骨架的肽核酸(PNA)(Nielsen等人,1991.*Science* 254,1497-1500)。基于PNA的分子是DNA分子在碱基对识别方面的真实模拟物。PNA的骨架由通过肽键连接的N-(2-氨基乙基)-甘氨酸单元构成,其中核碱基通过亚甲基羰基键与骨架连接。替代的骨架包括一个碳延伸的吡咯烷PNA单体(Govindaraju和Kumar,2005.*Chem. Commun.*,495-497)。因为PNA分子的骨架不含带电荷的磷酸基团,所以PNA-RNA杂合体通常分别比RNA-RNA或RNA-DNA杂合体更稳定(EghoIm等人,1993.*Nature* 365,566-568)。

[0079] 另一优选的骨架包含吗啉代核苷酸类似物或等同物,其中核糖或脱氧核糖被6元吗啉代环替换。最优选的核苷酸类似物或等同物包括磷酰二胺吗啉代寡聚物(PMO),其中核糖或脱氧核糖被6元吗啉代环替换,并且相邻吗啉代环之间的阴离子磷酸二酯键被非离子磷酰二胺键替换。

[0080] 另一优选的核苷酸类似物或等同物包括对磷酸二酯键中的至少一个非桥连氧的取代。这种修饰使碱基配对稍微不稳定,但增加了对核酸酶降解的显著抗性。优选的核苷酸类似物或等同物包括硫代磷酸酯、手性硫代磷酸酯、二硫代磷酸酯、磷酸三酯、氨基烷基磷酸三酯、H-膦酸酯、甲基膦酸酯和其他烷基膦酸酯(包括3'-亚烷基膦酸酯、5'-亚烷基膦酸酯和手性膦酸酯)、次膦酸酯、氨基磷酸酯(包括3'-氨基氨基磷酸酯和氨基烷基氨基磷酸酯)、硫代氨基磷酸酯、硫代烷基膦酸酯、硫代烷基磷酸三酯、硒代磷酸酯或碲代磷酸酯。

[0081] 另一优选的核苷酸类似物或等同物包含一个或多个在2'、3'和/或5'位置处被单取代或二取代的糖部分,诸如-OH;-F;取代或未取代的、直链或支链的低级(C1-C10)烷基、烯基、炔基、烷芳基、烯丙基、芳基或芳烷基,其可被一个或多个杂原子中断;O-烷基、S-烷基或N-烷基;O-烯基、S-烯基或N-烯基;O-炔基、S-炔基或N-炔基;O-烯丙基、S-烯丙基或N-烯丙基;O-烷基-O-烷基,-甲氧基,-氨基丙氧基;氨氧基、甲氧基乙氧基;-二甲氨基乙氧基乙氧基。糖部分可以是吡喃糖或其衍生物,或脱氧吡喃糖或其衍生物,优选核糖或其衍生物,或脱氧核糖或其衍生物。这种优选的衍生糖部分包括锁核酸(Locked Nucleic Acid,LNA),其中2'-碳原子与糖环的3'或4'碳原子连接,从而形成双环糖部分。优选的LNA包含2'-O,4'-C-亚乙基桥连核酸(Morita等人,2001.*Nucleic Acid Res* 增刊No.1:241-242)。这些取代使得核苷酸类似物或等同物对RNA酶H和核酸酶的抗性并增加了对靶标的亲和力。

[0082] 本发明上下文中氨基酸序列或核酸序列的“序列同一性”或“同一性”在本文中被

定义为两个或更多个氨基酸(肽、多肽或蛋白质)序列或两个或更多个核酸(核苷酸、寡核苷酸、多核苷酸)序列之间的关系,如通过比较所述序列确定的。在本领域中,“同一性”还意指氨基酸或核苷酸序列(视情况而定)之间的序列相关性程度,如通过具有此类序列的链之间的匹配所确定的。在本发明中,与特定序列的序列同一性优选是指在所述特定多肽或多核苷酸序列的整个长度上的序列同一性。

[0083] 两个氨基酸序列之间的“相似性”是通过比较一种肽或多肽的氨基酸序列及其保守氨基酸取代物与第二肽或多肽的序列来确定的。在一个优选的实施方式中,计算整个序列(SEQ ID NO:)的同一性或相似性,如在本文中所鉴定的。“同一性”和“相似性”可以通过已知方法容易地计算,所述已知方法包括但不限于在Computational Molecular Biology, Lesk, A.M. 编辑, Oxford University Press, New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D.W. 编辑, Academic Press, New York, 1993; Computer Analysis of Sequence Data, 第I部分, Griffin, A.M. 和 Griffin, H.G. 编辑, Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heine, G., Academic Press, 1987; 和 Sequence Analysis Primer, Gribskov, M. 和 Devereux, J. 编辑, M Stockton Press, New York, 1991; 以及 Carillo, H. 和 Lipman, D., SIAM J. Applied Math., 48:1073 (1988) 中描述的那些方法。

[0084] 用于确定同一性的优选方法被设计成给出所测试的序列之间的最大匹配。用于确定同一性和相似性的方法被编入了公开可用的计算机程序中。用于确定两个序列之间的同一性和相似性的优选计算机程序方法包括例如GCG程序包(Devereux, J. 等人, Nucleic Acids Research 12(1):387 (1984))、BestFit、BLASTP、BLASTN和FASTA(Altschul, S.F. 等人, J. Mol. Biol. 215:403-410 (1990))。BLAST X程序可从NCBI和其他来源(BLAST Manual, Altschul, S. 等人, NCBI NLM NIH Bethesda, MD 20894; Altschul, S. 等人, J. Mol. Biol. 215:403-410 (1990)) 公开获得。还可以使用众所周知的史密斯-沃特曼算法(Smith Waterman algorithm)来确定同一性。

[0085] 多肽序列比较的优选参数包括以下:算法:Needleman和Wunsch, J. Mol. Biol. 48:443-453 (1970); 比较矩阵:BLOSSUM62, 来自Hentikoff和Hentikoff, Proc. Natl. Acad. Sci. USA. 89:10915-10919 (1992); 空位罚分:12; 以及空位长度罚分:4。使用这些参数的可用程序可作为来自位于Madison, WI的Genetics Computer Group的“Ogap”程序公开获得。上述参数是用于氨基酸比较的默认参数(并且没有针对末端空位的罚分)。

[0086] 用于核酸比较的优选参数包括以下:算法:Needleman和Wunsch, J. Mol. Biol. 48:443-453 (1970); 比较矩阵:匹配=+10, 不匹配=0; 空位罚分:50; 空位长度罚分:3。可从位于Madison, Wis的Genetics Computer Group获得Gap程序。以上给出的是用于核酸比较的默认参数。

[0087] 任选地,在确定氨基酸相似性程度时,技术人员还可以考虑所谓的“保守性”氨基酸取代,如本领域技术人员所清楚的。保守氨基酸取代是指具有相似侧链的残基的可互换性。例如,一组具有脂肪族侧链的氨基酸是甘氨酸、丙氨酸、缬氨酸、亮氨酸和异亮氨酸;一组具有脂肪族-羟基侧链的氨基酸是丝氨酸和苏氨酸;一组具有含酰胺侧链的氨基酸是天冬酰胺和谷氨酰胺;一组具有芳族侧链的氨基酸是苯丙氨酸、酪氨酸和色氨酸;一组具有碱性侧链的氨基酸是赖氨酸、精氨酸和组氨酸;并且一组具有含硫侧链的氨基酸是半胱氨酸

和甲硫氨酸。优选的保守氨基酸取代组是：缬氨酸-亮氨酸-异亮氨酸、苯丙氨酸-酪氨酸、赖氨酸-精氨酸、丙氨酸-缬氨酸，以及天冬酰胺-谷氨酰胺。本文公开的氨基酸序列的取代变体是这样的变体，其中已去除了所公开序列中的至少一个残基并在其位置插入不同残基。优选地，氨基酸变化是保守性的。每种天然存在的氨基酸的优选保守取代如下：Ala至ser；Arg至lys；Asn至gln或his；Asp至glu；Cys至ser或ala；Gln至asn；Glu至asp；Gly至pro；His至asn或gln；Ile至leu或val；Leu至ile或val；Lys至arg；gln或glu；Met至leu或ile；Phe至met、leu或tyr；Ser至thr；Thr至ser；Trp至tyr；Tyr至trp或phe；以及Val至ile或leu。

[0088] 根据本发明的多核苷酸由核苷酸序列表示。根据本发明的多肽由氨基酸序列表示。根据本发明的核酸构建体被定义为从天然存在的基因分离的多核苷酸，或者经过修饰以含有多核苷酸的区段(segment)的多核苷酸，所述多核苷酸的区段以天然不存在的方式组合或并置。

[0089] 本文提供的序列信息不应狭窄地解释为需要包含错误鉴定的碱基。技术人员能够辨别出此类错误鉴定的碱基并且知道如何校正此类错误。

[0090] 表达被理解为包括参与多肽产生的任何(单一)步骤，所述步骤包括但不限于转录、转录后修饰、翻译、翻译后修饰和分泌。

[0091] 术语“表达构建体”在本文中术语“表达盒”可互换使用，并且通常包含根据本发明的多核苷酸和表达多核苷酸的必要组分，例如启动子、终止子、Kozak序列等。表达构建体可以位于载体上；此类载体可以是质粒。

[0092] 在本发明所有实施方式的上下文中，感兴趣的化合物可以是任何生物化合物。生物化合物可以是生物质或生物聚合物或代谢物。生物化合物可以由单个多核苷酸或构成生物合成或代谢途径的一系列多核苷酸编码，或者可以是单个多核苷酸的产物或一系列多核苷酸的产物的直接结果，所述多核苷酸可以是基因，所述一系列多核苷酸可以是基因簇。在本发明的所有实施方式中，编码感兴趣的生物化合物或与所述感兴趣的生物化合物相关的生物合成或代谢途径的单个多核苷酸或一系列多核苷酸是根据本发明的组合物和方法的优选靶标。生物化合物可以是对于宿主细胞天然的或对于宿主细胞异源的。

[0093] 术语“异源生物化合物”在本文中定义为对细胞不是天然生物化合物；或这样的天然生物化合物，在所述天然生物化合物中已经进行结构修饰以改变所述天然生物化合物。

[0094] 术语“生物聚合物”在本文中被定义为具有相同、相似或不相似的亚单元(单体)的链(或聚合物)。生物聚合物可以是任何生物聚合物。生物聚合物可以是例如但不限于核酸、聚胺、多元醇、多肽(或聚酰胺)，或多糖。

[0095] 生物聚合物可以是多肽。多肽可以是具有感兴趣的生物活性的任何多肽。术语“多肽”在本文中并不是指特定长度的编码产物，因此涵盖肽、寡肽和蛋白质。术语多肽是指任何长度的氨基酸聚合物。聚合物可以是直链或支链的，其可以包含经修饰的氨基酸，并且其可以被非氨基酸中断。该术语还包括已经修饰的氨基酸聚合物；例如，二硫键形成、糖基化、脂化、乙酰化、磷酸化，或任何其他操纵，诸如与标记组分缀合。如本文所用，术语“氨基酸”包括天然和/或非天然或合成的氨基酸，包括甘氨酸和D或L旋光异构体二者，以及氨基酸类似物和肽模拟物。多肽还包括上述多肽和杂合多肽的天然存在的等位基因变体和工程化变体。多肽可以是对于宿主细胞天然的或可以是对于宿主细胞异源的。多肽可以是胶原蛋白

或明胶,或它们的变体或杂合体。多肽可以是抗体或其部分、抗原、凝血因子、酶、激素或激素变体、受体或其部分、调节蛋白、结构蛋白、报告分子、或转运蛋白、参与分泌过程的蛋白质、参与折叠过程的蛋白质、分子伴侣、肽氨基酸转运蛋白、糖基化因子、转录因子、合成肽或寡肽、细胞内蛋白质。细胞内蛋白质可以是酶,诸如蛋白酶、神经酰胺酶、环氧化物水解酶、氨肽酶、酰基转移酶、醛缩酶、羟化酶、氨肽酶、脂肪酶。多肽也可以是细胞外分泌的酶。此类酶可以属于以下项的组:氧化还原酶、转移酶、水解酶、裂解酶、异构酶、连接酶、过氧化氢酶、纤维素酶、几丁质酶、角质酶、脱氧核糖核酸酶、葡聚糖酶、酯酶。所述酶可以是糖酶,例如纤维素酶(诸如内切葡聚糖酶、 β -葡聚糖酶、纤维二糖水解酶或 β -葡糖苷酶)、半纤维素酶或果胶裂解酶,诸如木聚糖酶、木糖苷酶、甘露聚糖酶、半乳聚糖酶、半乳糖苷酶、果胶甲基酯酶、果胶裂解酶、果胶酸裂解酶、内多聚半乳糖醛酸酶、外多聚半乳糖醛酸酶、鼠李半乳糖醛酸酶、阿拉伯聚糖酶、阿拉伯糖呋喃糖苷酶、阿拉伯糖基木聚糖水解酶、半乳糖醛酸酶、裂解酶或淀粉分解酶;水解酶、异构酶或连接酶、磷酸酶诸如植酸酶、酯酶诸如脂肪酶、蛋白水解酶、氧化还原酶诸如氧化酶、转移酶,或异构酶。酶可以是植酸酶。所述酶可以是氨肽酶、天冬酰胺酶、淀粉酶、产麦芽糖淀粉酶、糖酶、羧肽酶、内切蛋白酶、金属蛋白酶、丝氨酸蛋白酶过氧化氢酶、几丁质酶、角质酶、环糊精糖基转移酶、脱氧核糖核酸酶、酯酶、 α -半乳糖苷酶、 β -半乳糖苷酶、葡糖淀粉酶、 α -葡糖苷酶、 β -葡糖苷酶、卤过氧化物酶、蛋白质脱氨酶、转化酶、漆酶、脂肪酶、甘露糖苷酶、变构酶(mutanase)、氧化酶、果胶裂解酶、过氧化物酶、磷脂酶、半乳糖脂酶、叶绿素酶、多酚氧化酶、核糖核酸酶、转谷氨酰胺酶、或葡糖氧化酶、己糖氧化酶、单加氧酶。

[0096] 根据本发明,感兴趣的化合物可以是具有改进的分泌特征的多肽或酶,如W02010/102982中所述。根据本发明,感兴趣的化合物可以是融合或杂合多肽,另一种多肽在所述多肽或其片段的N末端或C末端处与其融合。通过将编码一种多肽的核酸序列(或其部分)与编码另一种多肽的核酸序列(或其部分)融合来产生融合多肽。

[0097] 用于产生融合多肽的技术是本领域已知的,并且包括连接编码多肽的编码序列,以使得它们在框内(in frame)并且融合多肽的表达在相同的启动子和终止子的控制下。杂合多肽可包含从至少两种不同多肽获得的部分或完整多肽序列的组合,其中一种或多种所述多肽可以是对宿主细胞异源的。融合多肽和信号序列融合体的示例例如如W02010/121933中所述。

[0098] 生物聚合物可以是多糖。多糖可以是任何多糖,包括但不限于粘多糖(例如,肝素和透明质酸)和含氮多糖(例如,几丁质)。在优选的选项中,多糖是透明质酸。根据本发明的编码感兴趣的化合物或编码参与感兴趣的化合物产生的化合物的多核苷酸可以编码参与初级或次级代谢产物(诸如有机酸、类胡萝卜素、(β -内酰胺)抗生素和维生素)的合成的酶。此类代谢产物可以被认为是根据本发明的生物化合物。

[0099] 术语“代谢产物”包括初级和次级代谢产物两者;代谢产物可以是任何代谢产物。优选的代谢产物是柠檬酸、葡糖酸、己二酸、富马酸、衣康酸和琥珀酸。

[0100] 代谢产物可以由一种或多种基因编码,诸如在生物合成或代谢途径中。初级代谢产物是细胞的初级代谢或一般性代谢的产物,所述产物涉及能量代谢、生长和结构。次级代谢产物是次级代谢的产物(参见例如,R.B.Herbert,The Biosynthesis of Secondary Metabolites,Chapman和Hall,New York,1981)。

[0101] 初级代谢产物可以是但不限于氨基酸、脂肪酸、核苷、核苷酸、糖、甘油三酯,或维生素。

[0102] 次级代谢产物可以是但不限于生物碱、香豆素、类黄酮、聚酮、奎宁、类固醇、肽,或萜烯。次级代谢产物可以是抗生素、拒食素、引诱剂、杀细菌剂、杀真菌剂、激素、杀虫剂,或杀鼠剂。优选的抗生素是头孢菌素和 β -内酰胺。其他优选的代谢产物是外代谢产物。外代谢产物的示例是Aurasperone B、Funalenone、Kotinin、Nigragillin、Orlandin、其他萘并- γ -吡喃酮、吡喃黑杆菌素A (Pyranonigrin A)、Tensidol B、伏马菌素B2和赭曲霉素A。

[0103] 生物化合物也可以是选择性标记的产物。选择性标记是感兴趣的多核苷酸的产物,该产物提供杀生物剂或病毒抗性、重金属抗性、营养缺陷体的原营养等。选择性标记包括但不限于amdS(乙酰胺酶)、argB(鸟氨酸氨甲酰转移酶)、bar(草胺磷乙酰转移酶)、hygB(潮霉素磷酸转移酶)、niaD(硝酸还原酶)、pyrG(乳清酸核苷-5'-磷酸脱羧酶)、sC(硫酸腺苷酰转移酶)、trpC(邻氨基苯甲酸合酶)、ble(腐草霉素抗性蛋白)、hyg(潮霉素)、NAT或NTC(诺尔丝菌素),以及它们的等同物。

[0104] 根据本发明,感兴趣的化合物优选是感兴趣的化合物列表中描述的多肽。

[0105] 根据本发明的另一个实施方式,感兴趣的化合物优选是代谢产物。

[0106] 根据本发明的细胞可以已经能够产生感兴趣的化合物。根据本发明的细胞还可以被提供编码多肽的同源或异源核酸构建体,其中所述多肽可以是感兴趣的化合物或参与感兴趣的化合物生产的多肽。本领域技术人员知道如何修饰微生物宿主细胞,以使所述微生物宿主细胞能够产生感兴趣的化合物。

[0107] 本发明的所有实施方式涉及细胞,而不是无细胞的体外系统;换句话说,根据本发明的系统是细胞系统,而不是无细胞的体外系统。

[0108] 在本发明的所有实施方式中,例如,根据本发明的细胞可以是单倍体细胞、二倍体细胞或多倍体细胞。

[0109] 根据本发明的细胞在本文中可互换地称为“细胞”、“根据本发明的细胞”、“宿主细胞”和“根据本发明的宿主细胞”;所述细胞可以是任何细胞,例如原核细胞、藻类、微藻类、海洋真核生物、网黏菌纲(Labyrinthulomycetes)或真核细胞。优选地,所述细胞不是哺乳动物细胞。

[0110] 当细胞是原核细胞时,原核宿主细胞优选地是细菌宿主细胞。术语“细菌宿主细胞”包括革兰氏阴性微生物和革兰氏阳性微生物。优选地,根据本发明的细菌宿主细胞来自自由以下项组成的组的属:埃希氏菌属(Escherichia)、鱼腥藻属(Anabaena)、柄杆菌属(Caulobacter)、葡糖杆菌属(Gluconobacter)、红杆菌属(Rhodobacter)、假单胞菌属(Pseudomonas)、副球菌属(Paracoccus)、丙酸杆菌属(Propionibacterium)、芽孢杆菌属(Bacillus)、短杆菌属(Brevibacterium)、棒状杆菌属(Corynebacterium)、根瘤菌属(Rhizobium)(中华根瘤菌属(Sinorhizobium))、黄杆菌属(Flavobacterium)、克雷伯氏菌属(Klebsiella)、肠杆菌属(Enterobacter)、乳杆菌属(Lactobacillus)、乳球菌属(Lactococcus)、甲基杆菌属(Methylobacterium)、葡萄球菌属(Staphylococcus)或链霉菌属(Streptomyces)。更优选地,细菌宿主细胞选自由以下项组成的组:枯草芽孢杆菌(B. subtilis)、解淀粉芽孢杆菌(B. amyloliquefaciens)、地衣芽孢杆菌(B. licheniformis)、潘蒂芽孢杆菌(B. pumilus)、巨大芽孢杆菌(B. megaterium)、耐盐芽孢

杆菌 (*B. halodurans*)、短小芽孢杆菌 (*B. pumilus*)、氧化葡萄糖酸杆菌 (*G. oxydans*)、新月柄杆菌 (*Caulobacter crescentus*) CB 15、扭脱甲基杆菌 (*Methylobacterium extorquens*)、亚比多球型红杆菌 (*Rhodobacter sphaeroides*)、*Pseudomonas zeaxanthinifaciens*、脱氮副球菌 (*Paracoccus denitrificans*)、大肠杆菌 (*Escherichia coli*)、谷氨酸棒杆菌 (*Corynebacterium glutamicum*)、肉葡萄球菌 (*Staphylococcus carnosus*)、变铅青链霉菌 (*Streptomyces lividans*)、苜蓿中华根瘤菌 (*Sinorhizobium melioli*) 和放射形根瘤菌 (*Rhizobium radiobacter*)、优选地,所述细胞是真菌,即酵母细胞或丝状真菌细胞。优选地,细胞缺乏NHEJ(非同源末端连接)组分。所述与NHEJ相关的组分优选是酵母Ku70、Ku80、MRE11、RAD50、RAD51、RAD52、XRS2、SIR4和/或LIG4的同源物或直系同源物。或者,在根据本发明的细胞中,可以通过使用抑制RNA连接酶IV的化合物(诸如SCR7)来使NHEJ缺陷(Vartak SV和Raghavan, FEBS J. 2015Nov; 282(22): 4289-94)。本领域技术人员知道如何调节NHEJ及其对RNA引导的核酸酶系统的影响,参见例如W02014130955A1; Chu等人, Nat Biotechnol 2015, 33, 543-548; Yu等人, Cell Stem Cell, 2015, 16, 142-147.; 所有这些文献都以引用方式并入本文。术语“缺陷”在本文其他地方定义。

[0111] 当根据本发明的细胞是酵母细胞时,优选的酵母细胞来自自由以下项组成的组的属:念珠菌属 (*Candida*)、汉逊酵母属 (*Hansenula*)、伊萨酵母属 (*Issatchenkia*)、克鲁维酵母属 (*Kluyveromyces*)、毕赤酵母属 (*Pichia*)、酵母属 (*Saccharomyces*)、裂殖酵母属 (*Schizosaccharomyces*)、耶氏酵母属 (*Yarrowia*)、或接合酵母属 (*Zygosaccharomyces*); 更优选地,酵母宿主细胞选自自由以下项组成的组:乳酸克鲁维酵母 (*Kluyveromyces lactis*)、乳酸克鲁维酵母 (*Kluyveromyces lactis* NRRL) Y-1140、马克斯克鲁维酵母 (*Kluyveromyces marxianus*)、耐热克鲁维酵母 (*Kluyveromyces thermotolerans*)、克鲁斯假丝酵母 (*Candida krusei*)、萨纳瑞西斯假丝酵母 (*Candida sonorensis*)、光滑念珠菌 (*Candida glabrata*)、酿酒酵母 (*Saccharomyces cerevisiae*)、酿酒酵母 CEN.PK113-7D、粟酒裂殖酵母 (*Schizosaccharomyces pombe*)、多形汉逊酵母 (*Hansenula polymorpha*)、东方伊萨酵母 (*Issatchenkia orientalis*)、解脂耶氏酵母 (*Yarrowia lipolytica*)、解脂耶氏酵母 CLIB122、树干毕赤酵母 (*Pichia stipidis*) 和巴斯德毕赤酵母 (*Pichia pastoris*)。优选的酵母细胞是酿酒酵母。

[0112] 根据本发明的宿主细胞是丝状真菌宿主细胞。如本文所定义的丝状真菌包括亚门真菌门 (*Eumycota*) 和卵菌门 (*Oomycota*) 的所有丝状形式(如由 Hawksworth 等人在 Ainsworth and Bisby's Dictionary of The Fungi, 第8版, 1995, CAB International, University Press, Cambridge, UK 中所定义)。

[0113] 丝状真菌宿主细胞可以是分类单位发菌科 (*Trichocomaceae*) 的任何丝状形式的细胞(如由 Houbraeken 和 Samson 在 *Studies in Mycology* 70:1-51. 2011 中所定义的)。在另一个优选实施方式中,丝状真菌宿主细胞可以是三个科曲霉科 (*Aspergillaceae*)、嗜热子囊菌科 (*Thermoascaceae*) 和发菌科中的任一科的任何丝状形式的细胞,所述三个科容纳在分类单位发菌科中。

[0114] 丝状真菌的特征在于由几丁质、纤维素、葡聚糖、壳聚糖、甘露聚糖和其他复合多糖构成的菌丝体壁。营养生长是通过菌丝伸长进行的,并且碳分解代谢是专性需氧的。丝状真菌菌株包括但不限于以下菌株:支顶孢属 (*Acremonium*)、伞菌属 (*Agaricus*)、曲霉属

(*Aspergillus*)、短梗霉属(*Aureobasidium*)、金孢子菌属(*Chrysosporium*)、鬼伞属(*Coprinus*)、隐球菌属(*Cryptococcus*)、丝梗霉属(*Filibasidium*)、镰刀菌属(*Fusarium*)、腐质霉属(*Humicola*)、巨座壳属(*Magnaporthe*)、被孢霉属(*Mortierella*)、毛霉属(*Mucor*)、毁丝霉属(*Myceliophthora*)、新美鞭菌属(*Neocallimastix*)、脉孢菌属(*Neurospora*)、拟青霉属(*Paecilomyces*)、青霉菌属(*Penicillium*)、瘤胃壶菌属(*Piromyces*)、原毛平革菌属(*Panerochaete*)、侧耳属(*Pleurotus*)、裂褶菌属(*Schizophyllum*)、篮状菌属(*Talaromyces*)、踝节菌属(*Rasamsonia*)、热子囊菌属(*Thermoascus*)、梭孢壳属(*Thielavia*)、弯颈霉属(*Tolyposcladium*)和木霉属(*Trichoderma*)。根据本发明的优选丝状真菌宿主细胞来自选自由以下项组成的组的属:支顶孢属、曲霉属、金孢子菌属、毁丝霉属、青霉菌属、篮状菌属、踝节菌属、梭孢壳属、镰刀菌属和木霉属;更优选地来自选自由以下项组成的组的物种:黑曲霉(*Aspergillus niger*)、亚拉巴马枝顶孢(*Acremonium alabamense*)、泡盛曲霉(*Aspergillus awamori*)、臭曲霉(*Aspergillus foetidus*)、酱油曲霉(*Aspergillus sojae*)、烟曲霉菌(*Aspergillus fumigatus*)、埃默森篮状菌(*Talaromyces emersonii*)、踝节菌属埃默森蓝状菌(*Rasamsonia emersonii*)、踝节菌属埃默森蓝状菌CBS393.64、米曲霉(*Aspergillus oryzae*)、鲁克文金孢子菌(*Chrysosporium lucknowense*)、尖孢镰刀菌(*Fusarium oxysporum*)、高山被孢霉(*Mortierella alpina*)、高山被孢霉ATCC 32222、嗜热毁丝霉(*Myceliophthora thermophila*)、里氏木霉(*Trichoderma reesei*)、太瑞斯梭孢壳菌(*Thielavia terrestris*)、产黄青霉菌(*Penicillium chrysogenum*)和产黄青霉菌(*P. chrysogenum*)威斯康星54-1255(ATCC28089);甚至更优选地,根据本发明的丝状真菌宿主细胞是黑曲霉。

[0115] 当根据本发明的宿主细胞是黑曲霉宿主细胞时,所述宿主细胞优选是CBS 513.88、CBS124.903或它们的衍生物。

[0116] 在许多菌种保藏单位中若干丝状真菌菌株是公众可容易获得的,所述菌种保藏单位例如美国典型培养物保藏中心(American Type Culture Collection, ATCC)、德意志微生物保藏中心(Deutsche Sammlung von Mikroorganismen and Zellkulturen GmbH, DSM)、荷兰微生物菌种保藏中心(Centraalbureau Voor Schimmelcultures, CBS)、农业研究服务专利菌种保藏所北方地区研究中心(Agricultural Research Service Patent Culture Collection, Northern Regional Research Center, NRRL),以及俄罗斯科学院全俄微生物保藏中心(All-Russian Collection of Microorganisms of Russian Academy of Sciences, 俄文缩写-VKM, 英文缩写-RCM),俄罗斯莫斯科。作为根据本发明的宿主细胞的优选菌株是黑曲霉CBS 513.88、CBS124.903,米曲霉ATCC 20423、IFO 4177、ATCC 1011、CBS205.89、ATCC 9576、ATCC14488-14491、ATCC 11601、ATCC12892、产黄青霉菌CBS 455.95、产黄青霉菌威斯康辛54-1255(ATCC28089)、橘青霉(*Penicillium citrinum*) ATCC 38065、产黄青霉菌P2、太瑞斯梭孢壳菌NRRL8126、踝节菌属埃默森蓝状菌CBS393.64、埃默森篮状菌CBS 124.902、产黄头孢霉(*Acremonium chrysogenum*) ATCC 36225或ATCC 48272、里氏木霉ATCC 26921或ATCC 56765或ATCC 26921、酱油曲霉ATCC11906、嗜热毁丝霉C1、Garg 27K、VKM-F 3500D、鲁克文金孢子菌C1、Garg 27K、VKM-F 3500D、ATCC44006,以及它们的衍生物。

[0117] 优选地,根据本发明的宿主细胞具有这样的修饰,优选在其基因组中具有这样的

修饰,所述修饰导致如果当在相同条件下分析时与未经修饰的亲本宿主细胞相比,产生减少的或者不产生如本文所定义的不期望的化合物。

[0118] 可以通过本领域技术人员已知的任何手段来引入修饰,诸如但不限于传统菌株改良、随机诱变后进行选择。也可以通过定点诱变来引入修饰。

[0119] 可以通过引入(插入)、取代(置换)或去除(缺失)多核苷酸序列中的一个或多个核苷酸来完成修饰。可以实现编码不期望的化合物(诸如多肽)的多核苷酸的完全或部分缺失。不期望的化合物可以是本文其他地方列出的任何不期望的化合物;其也可以是合成不期望的化合物(诸如代谢物)的生物学途径中的蛋白质和/或酶。或者,编码所述不期望的化合物的多核苷酸可以部分或完全地被这样的多核苷酸序列替换,所述多核苷酸序列不编码所述不期望的化合物或编码所述不期望的化合物的部分或完全失活形式。在另一个替代方案中,可以将一个或多个核苷酸插入编码所述不期望的化合物的多核苷酸中,从而导致对所述多核苷酸的破坏以及因此使由被破坏的多核苷酸编码的所述不期望的化合物的部分或完全失活。

[0120] 在一个实施方式中,根据本发明的宿主细胞在其基因组中包含选自以下的修饰

[0121] a) 编码不期望的化合物的多核苷酸的完全或部分缺失,

[0122] b) 将编码不期望的化合物的多核苷酸用这样的多核苷酸序列完全或部分地替换,所述多核苷酸序列不编码所述不期望的化合物或编码所述不期望的化合物的部分或完全失活形式。

[0123] c) 通过多核苷酸序列中插入一个或多个核苷酸来破坏编码不期望的化合物的多核苷酸,并因此通过被破坏的多核苷酸使所述不期望的化合物部分或完全失活。

[0124] 该修饰可以例如在所述不期望的化合物的转录或翻译所需的编码序列或调节元件中。例如,可以插入或删除核苷酸以导致终止密码子的引入、起始密码子的去除,或编码序列的开放阅读框的改变或移框。对编码序列或其调节元件的修饰可以通过定点或随机诱变、DNA改组方法、DNA重组方法、基因合成(参见例如Young和Dong, (2004), *Nucleic Acids Research* 32(7)或Gupta等人, (1968), *Proc. Natl. Acad. Sci USA*, 60:1338-1344; Scarpulla等人, (1982), *Anal. Biochem.* 121:356-365; Stemmer等人, (1995), *Gene* 164:49-53)或根据本领域已知的方法进行PCR产生的诱变来完成。随机诱变程序的示例是本领域熟知的,诸如化学(例如NTG)诱变或物理(例如UV)诱变。定点诱变程序的示例是QuickChange™ 定点诱变试剂盒(Stratagene Cloning Systems, La Jolla, CA)、Altered Sites® II体外诱变系统(Promega Corporation),或通过使用如在Gene. 1989年4月15日; 77(1):51-9. (Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR "Site-directed mutagenesis by overlap extension using the polymerase chain reaction")中所述的PCR或使用如在Molecular Biology: Current Innovations and Future Trends. (A.M. Griffin和H.G. Griffin编辑, ISBN 1-898486-01-8; 1995 Horizon Scientific Press, PO Box 1, Wymondham, Norfolk, U.K.)中所述的PCR进行重叠延伸。

[0125] 优选的修饰方法为基于重组遗传操纵技术,诸如部分或完全基因替换或部分或完全基因缺失。

[0126] 例如,在替换多核苷酸、核酸构建体或表达盒的情况下,可以在待替换的靶基因座处引入合适的DNA序列。合适的DNA序列优选存在于克隆载体上。优选的整合克隆载体包含

这样的DNA片段,所述DNA片段与待替换基因座侧翼的多核苷酸具有同源性和/或与多核苷酸同源,以将克隆载体的整合靶向至该预定基因座。为了促进靶向整合,优选在转化细胞之前将克隆载体线性化。优选地,执行线性化,使得克隆载体的至少一个但优选两端侧接与待替换的DNA序列(或侧翼序列)同源的序列。该过程称为同源重组,并且该技术也可用于实现(部分)基因缺失。

[0127] 例如,与内源多核苷酸对应的多核苷酸可以被缺陷型多核苷酸替换;所述缺陷型多核苷酸是不能生产(全功能性)多肽的多核苷酸。通过同源重组,缺陷型多核苷酸替换内源多核苷酸。可期望的是,缺陷型多核苷酸还编码标记,所述标记可用于选择核酸序列已被修饰的转化体。替代地或与其他提到的技术组合,可以使用基于大肠杆菌细胞中粘粒的重组的技术,如在以下中所述:A rapid method for efficient gene replacement in the filamentous fungus *Aspergillus nidulans* (2000) Chaveronche, M-K., Ghico, J-M. 和 d'Enfert C; *Nucleic acids Research*, 第28卷, 第22期。

[0128] 或者,可以通过使用与多核苷酸的核酸序列互补的核苷酸序列进行已建立的反义技术来执行修饰,在所述修饰中所述宿主细胞产生较少或不产生蛋白质,诸如如本文所描述且由本文所述的多核苷酸编码的具有淀粉酶活性,优选 α -淀粉酶活性的多肽。更特别地,可以通过引入与多核苷酸的核酸序列互补的核苷酸序列来减少或消除宿主细胞对所述多核苷酸的表达,所述与所述多核苷酸的核酸序列互补的核苷酸序列可以在细胞中转录并且能够与所述细胞中产生的mRNA杂交。在允许互补反义核苷酸序列与mRNA杂交的条件下,翻译的蛋白质的量因此减少或被消除。表达反义RNA的示例显示于

Appl. Environ. Microbiol. 2000年2月; 66 (2) : 775-82中。(Characterization of a foldase, protein disulfide isomerase A, in the protein secretory pathway of *Aspergillus niger*. Ngiam C, Jeenes DJ, Punt PJ, Van Den Hondel CA, Archer DB中) 或 (Zrenner R, Willmitzer L, Sonnwald U. Analysis of the expression of potato uridinediphosphate-glucose pyrophosphorylase and its inhibition by antisense RNA. *Planta*. (1993) ; 190 (2) : 247-52.)。

[0129] 导致不期望的化合物的产生减少或不产生的修饰优选是由于如果与未经修饰的亲本微生物宿主细胞相比并且在相同条件下测量时,编码所述不期望的化合物的mRNA的产生减少。

[0130] 可以经由RNA干扰(RNAi)技术(Mouyna等人, 2004)来获得导致从编码不期望的化合物的多核苷酸转录的mRNA量减少的修饰。在该方法中,将表达要受影响的核苷酸序列的相同有义和反义部分彼此前后克隆,且在其间具有核苷酸间隔区,并插入到表达载体中。在转录此类分子后,小核苷酸片段的形成将导致要影响的mRNA的靶向降解。特定mRNA的消除可以是不同程度的。RNA干扰技术描述于例如W02008/053019、W02005/05672A1和W02005/026356A1中。

[0131] 可以通过不同方法来获得导致不期望的化合物的产生减少或不产生的修饰,例如通过针对此类不期望的化合物的抗体或化学抑制剂或蛋白质抑制剂或物理抑制剂(Tour O. 等人, (2003) *Nat. Biotech*: Genetically targeted chromophore-assisted light inactivation., 第21卷, 编号12: 1505-1508) 或肽抑制剂或反义分子或RNAi分子(R. S. Kamath等人, (2003) *Nature*: Systematic functional analysis of the

Caenorhabditis elegans genome using RNAi.,第421卷,231-237)。

[0132] 除了上述技术之外或作为替代方案,还可以抑制不期望的化合物的活性,或借助于替代信号序列(Ramon de Lucas,J.,Martinez O,Perez P.,Isabel Lopez,M.,Valenciano,S.和Laborda,F.The Aspergillus nidulans carnitine carrier encoded by the acuH gene is exclusively located in the mitochondria.FEMS Microbiol Lett.2001年7月24日;201(2):193-8.)或保留信号(Derkx,P.M.和Madrid,S.M.The foldase CYPB is a component of the secretory pathway of Aspergillus niger and contains the endoplasmic reticulum retention signal HEEL.Mol.Genet.Genomics.2001年12月;266(4):537-545)或通过诸如多肽的不期望的化合物靶向到能够与参与细胞分泌途径的细胞的膜结构融合而导致细胞外分泌所述多肽的过氧化物酶体(例如,如在W02006/040340中所述)来重新定位诸如蛋白质的不期望的化合物。

[0133] 替代地或与上述技术组合,还可以获得不期望的化合物的产生减少或不产生,例如通过UV或化学诱变(Mattern,I.E.,van Noort J.M.,van den Berg,P.,Archer,D.B.,Roberts,I.N.和van den Hondel,C.A.,Isolation and characterization of mutants of Aspergillus niger deficient in extracellular proteases.Mol Gen Genet.1992年8月;234(2):332-6.)或通过使用抑制如本文所述的不期望的多肽的酶促活性的抑制剂(例如,野尻霉素,其用作 β -葡糖苷酶的抑制剂(Carrel F.L.Y.和Canevascini G.Canadian Journal of Microbiology(1991)37(6):459-464;Reese E.T.,Parrish F.W.和Ettlinger M.Carbohydrate Research(1971)381-388))。

[0134] 在本发明的一个实施方式中,根据本发明的宿主细胞的基因组中的修饰是对编码不期望的化合物的多核苷酸的至少一个位置的修饰。

[0135] 在化合物(例如不期望的化合物,诸如不期望的多肽和/或酶)产生方面有缺陷的细胞在本文中定义为这样的突变微生物宿主细胞,所述突变微生物宿主细胞已经修饰,优选在其基因组中经修饰,以产生表型特征,其中所述细胞与未修饰的亲本宿主细胞相比,当在相同条件下分析时:a)产生更少的不期望的化合物或基本上不产生不期望的化合物,和/或b)产生具有降低的活性或降低的比活性的不期望的化合物或不具有活性或不具有比活性的不期望的化合物,以及这些可能性中的一种或多种的组合。

[0136] 优选地,根据本发明的经修饰的宿主细胞如果与未经修饰的亲本宿主细胞相比且在相同条件下测量,则产生少1%的不期望的化合物、至少少5%的不期望的化合物、至少少10%的不期望的化合物、至少少20%的不期望的化合物、至少少30%的不期望的化合物、至少少40%的不期望的化合物、至少少50%的不期望的化合物、至少少60%的不期望的化合物、至少少70%的不期望的化合物、至少少80%的不期望的化合物、至少少90%的不期望的化合物、至少少91%的不期望的化合物、至少少92%的不期望的化合物、至少少93%的不期望的化合物、至少少94%的不期望的化合物、至少少95%的不期望的化合物、至少少96%的不期望的化合物、至少少97%的不期望的化合物、至少少98%的不期望的化合物、至少少99%的不期望的化合物、至少少99.9%的不期望的化合物,或最优选少100%的不期望的化合物。

[0137] 本文中提及的专利文献或作为现有技术给出的其他事项不应被视为承认该文献

或事项是已知的,或者该文献或事项所包含的信息是在任何权利要求的优先权日期时的公知常识的一部分。

[0138] 本文所述的每篇参考文献的公开内容均以引用方式整体并入本文。

[0139] 通过以下实施例进一步说明本发明。

实施例

[0140] 在以下实施例中,说明了本发明的多种实施方式。根据以上描述和这些实施例中,本领域技术人员可以对本发明进行各种改变和修改以使其适应各种用途和条件。

[0141] 材料与amp;方法

[0142] 一般分子生物学技术

[0143] 除非另有说明,否则使用的方法是标准生化技术。合适的通用方法教科书的示例包括Sambrook等人,Molecular Cloning,a Laboratory Manual (1989) 和Ausubel等人,Current Protocols in Molecular Biology (1995), John Wiley&Sons, Inc.

[0144] 质粒、寡核苷酸引物和菌株

[0145] 实施例中使用的质粒列于表1中。用于进一步菌株工程化的菌株列于表2中。

[0146] 培养基

[0147] 实验中使用的培养基是补充有糖(即YEPHD, 20g/l葡萄糖;YEPHG, 20g/l半乳糖)的YEPH培养基(10g/l酵母提取物、20g/l植物蛋白胨(BD BioSciences, Temse, 比利时)和固体YNB培养基(6.7g/l酵母基础氮源、15g/l琼脂)。对于固体YEPH培养基,在灭菌之前将15g/l琼脂加入到液体培养基中。

[0148] 对于CuSO₄诱导实验,使用矿物培养基。Verduyn等人已经描述了矿质培养基的组成(Yeast, 1992, 第8卷, 第501-517页)。用2.3g/l尿素代替硫酸铵作为氮源。培养基的初始pH为4.6。

[0149] 表1. 实施例中使用的质粒的列表。

名称	特性	来源
pCSN061	CEN6.ARSH4, kanMX, Cas9.	PCT/EP2016/050136 SEQ ID NO: 18, 图 6.
pRN1120-RFP-gRNA(A)	基于 pRS305 的带有 natMX 的穿梭载体, 带有 2 微米起点和表达盒 TDH3p-RFP-PGI1t.	SEQ ID NO: 1, 图 7
[0150] pDB1371	在 KpnI/NgoMIV 位点中插入了 pGAL10-GIN11(M86)的 pCSN061。	实施例 1 SEQ ID NO: 4, 图 8
pDB1372	在 KpnI/NgoMIV 位点中插入了 pCUP1-GIN11(M86)的 pCSN061。	实施例 1 SEQ ID NO: 5, 图 9
pDB1368	具有表达盒 SparTDH3p-HXT11/2 ^{N366T} -EFM1t.的克隆载体。	实施例 2 SEQ ID NO: 6, 图 10

[0151] 表2. 在实施例中使用和产生的酿酒酵母菌株的列表。

菌株名称	相关基因型	来源
CEN.PK113-7D	<i>MATa URA3 HIS3 LEU2 TRP1 MAL2-8 SUC2</i>	Van Dijken 等, Enzyme Microb Technol. 2000 Jun 1;26(9-10):706-714.
[0152] CP-61	CEN.PK113-7D pCSN061 (6 个转化体的合并)	实施例 2
CP-71	CEN.PK113-7D pDB1371 (6 个转化体的合并)	实施例 2
CP-72	CEN.PK113-7D pDB1372 (6 个转化体的合并)	实施例 2
CP-61-HXT	CEN.PK113-7D <i>int70::[SparTDH3p-HXT11/2^{N366T}-EFM1t]</i> pCSN061	实施例 2, 4
[0153] CP-71-HXT	CEN.PK113-7D <i>int70::[SparTDH3p-HXT11/2^{N366T}-EFM1t]</i> pDB1371	实施例 2, 3, 4
CP-72-HXT	CEN.PK113-7D <i>int70::[SparTDH3p-HXT11/2^{N366T}-EFM1t]</i> pDB1372	实施例 2, 5

[0154] 实施例1:克隆带有GIN11M86的Cas9表达质粒

[0155] 为了能够控制GIN11 (M86) 的表达, 该序列被设计在酿酒酵母中众所周知的诱导型启动子的上游: CUP1 (Mascorro-Gallardo等人, Gene. 1996年6月12日; 172 (1) : 169-170) 或 GAL10的600bp上游序列 (Partow等人, Yeast. 2010年11月; 27 (11) : 955-964)。GAL10被葡萄糖抑制并被半乳糖诱导, 并且CUP1启动子被铜诱导。pGAL10-GIN11 (M86) (cFS0017; SEQ ID NO: 2) 和pCUP1-GIN11 (M86) (cFS0018; SEQ ID NO: 3) 表达盒均在ATUM (Menlo Park, CA, USA)

处合成。通过使用正向引物(SEQ ID NO:7)和反向引物(SEQ ID NO:8)的延伸PCR扩增将KpnI和NgoMIV位点添加至pGAL10-GIN11 (M86) 构建体,在此之后使用上述限制性位点将PCR产物克隆到pCSN061(带有化脓性链球菌(*S. pyogenes*) Cas9表达盒的质粒)中,从而产生质粒pDB1371(SEQ ID NO:4)。使用作为合成构建体的一部分的KpnI和NgoMIV位点将pCUP1-GIN11 (M86) 盒克隆到pCSN061中,从而产生质粒pDB1372(SEQ ID NO:5)。所用的pGAL10启动子序列如SEQ ID NO:19所示,所用的pCUP1启动子序列如SEQ ID NO:20所示。GIN11 (M86) 序列如SEQ ID NO:21所示。

[0156] 实施例2:带有GIN11M86的Cas9表达质粒有效地促进基因组修饰

[0157] 在专利申请PCT/EP2013/056623和PCT/EP2016/050136中描述了所遵循的菌株构建方法。PCT/EP2013/056623描述了这样的技术,该技术能够构建来自各种感兴趣的基因的表达盒,使得这些盒在转化该酵母后组合进通路并整合到该酵母基因组的特定基因座中。

[0158] PCT/EP2016/050136的实施例9描述了CRISPR-Cas9系统用于将表达盒整合到宿主细胞(在这种情况下是酿酒酵母)的基因组中的用途。在第一轮转化中,将pCSN061、pDB1371或pDB1372单独引入酵母中,pCSN061、pDB1371或pDB1372均为带有化脓性链球菌Cas9表达盒的G418可选择的附加型质粒。用500ng的pCSN061、pDB1371或pDB1372转化CEN.PK113-7D。在补充有200微克/毫升G418的固体琼脂YEPHD培养基(YEPHD-G418, Invivogen)上选择正确的转化体。随后,将几个转化体在YEPHD-G418(200微克/毫升)琼脂上重新划线,以获得纯菌落。合并六个菌落以继续进行下一轮转化。产生的三个转化体合并物分别被命名为:CP-61、CP-71和CP-72(参见表2)。

[0159] 在第二轮转化中,将gRNA受体质粒骨架PCR片段、与gRNA受体质粒骨架PCR片段具有同源性的引导RNA PCR片段(其允许体内重组到含有诺尔丝菌素选择性标记的环状质粒中)和供体DNA表达盒转化到预表达Cas9的细胞中,从而产生预期的修饰。为了将含有预期修饰的供体DNA表达盒引入酵母基因组,选择酵母基因组中的整合位点。通过使用引物的延伸PCR将与选定的整合位点具有约50bp同源性的DNA侧翼添加到供体DNA中,从而将侧翼序列引入到产生的PCR产物中。这些侧翼(在供体DNA表达盒的5'末端和3'末端处的大小为50bp)允许在酵母中转化后将供体DNA片段正确整合到预期的基因座。在用DNA片段转化酵母细胞后,在期望位置发生体内重组和整合到基因组中。

[0160] 整合位点:表达盒靶向于INT70基因座。INT70整合位点是位于酿酒酵母的染色体XIV上的YNL180C与YNL178W之间的非编码区。靶向INT70的引导序列是使用gRNA设计工具(<https://www.dna20.com/eCommerce/cas9/input>)设计的。

[0161] gRNA表达盒(如由DiCarlo等人,Nucleic Acids Res.2013年4月;41(7):4336-4343)是作为合成DNA盒(gBLOCK)在Integrated DNA Technologies (Leuven,比利时).INT70 gBLOCK;SEQ ID NO:9)订购的。

[0162] gRNA受体质粒骨架:随后,通过共转化来源于酵母载体pRN1120-RFP-gRNA(A)的线性PCR片段来完成gRNA表达质粒的体内组装。pRN1120-RFP-gRNA(A)是一种多拷贝酵母穿梭载体,该多拷贝酵母穿梭载体包含赋予对诺尔丝菌素(NTC)的抗性的功能性natMX标记盒(SEQ ID NO:1,图7)。该质粒的骨架基于pRS305(Sikorski和Hieter,Genetics 1989,第122卷,第19-27页),包括功能性2微米ORI序列、功能性natMX标记盒和RFP表达盒,以便能够基于荧光或通过肉眼可见的菌落的粉红色至紫色着色来追踪携带该质粒的菌落。

[0163] 供体DNA表达盒的构建:在ATUM (Menlo Park, CA, USA) 处合成开放阅读框 (ORF)、启动子序列和终止子。如Engler等人, PLoS One. 2008; 3 (11) :e3647和Engler等人, PLoS One. 2009; 4 (5) :e5553以及其中的参考文献所述, 通过使用Golden Gate技术来重组启动子、ORF和终止子序列。将表达盒克隆到标准克隆载体中。所得质粒 (也在表1中列出) 是pDB1368。pDB1368 (SEQ ID NO:6) 带有在奇异酵母 (*Saccharomyces paradoxus*) TDH3启动子 (SparTDH3p) 和酿酒酵母EFM1终止子 (EFM1t) 控制下的嵌合戊糖转运体HXT11/2的表达盒 (Shin等人, 2017, Biotechnol Bioeng. 2017年9月; 114 (9) :1937-1945. doi:10.1002/bit.26322)。如下所述, 通过延伸PCR将用于整合到INT70基因座内的侧翼添加至供体DNA表达盒的5'末端和3'末端。

[0164] 用指定的PCR片段转化CP-61、CP-71、CP-72

[0165] 对于第二轮转化, 用以下片段转化预表达Cas9的CP-61、CP-71和CP-72, 从而导致HXT11/2^{N366T}表达盒的组装和在INT70基因座处的整合 (图1):

[0166] 1) 使用pDB1368作为模板, 用正向引物 (SEQ ID NO:10) 和反向引物 (SEQ ID NO:11) 生成供体DNA表达盒PCR片段 (int70[5']-conD-HXT11/2-con3-int70[3']);

[0167] 2) 使用pRN1120-RFP-gRNA (A) (SEQ ID NO:1) 作为模板, 用正向引物 (SEQ ID NO:12) 和反向引物 (SEQ ID NO:13) 生成gRNA受体质粒骨架PCR片段 (骨架-1120-RFP-gRNA [A]);

[0168] 3) 使用INT70 gBLOCK (SEQ ID NO:9) 作为模板, 用正向引物 (SEQ ID NO:14) 和反向引物 (SEQ ID NO:15) 生成引导RNA PCR片段 (gRNA-INT70)。

[0169] 在含有200微克/毫升的G418和200微克/毫升的NTC (Werner Bioagents) 的YEPHD-琼脂平板 (YEPHD-G418-NTC) 上选择转化体。对从每次转化的一个菌落分离的基因组DNA执行使用正向引物 (SEQ ID NO:16) 和反向引物 (SEQ ID NO:17) 的诊断PCR, 以确认HXT11/2表达盒在INT70基因座处的正确组装和整合 (图2)。这些结果表明, Cas9质粒pDB1371和pDB1372是功能性的, 并且使得能够进行HXT11/2表达盒在INT70基因座处的正确的CRISPR/Cas9介导的整合。将用于pCSN061、pDB1371或pDB1372的进一步质粒丢失实验的所得菌落分别命名为CP-61-HXT、CP-71-HXT或CP-72-HXT (表2)。

[0170] 实施例3: 从CP-71-HXT的半乳糖诱导的质粒丢失表明当诱导GIN11 (M86) 序列时有更有效的质粒丢失

[0171] 为了比较在诱导和不诱导的情况下通过诱导pDB1371上存在的GAL10p-GIN11 (M86) 构建体的质粒丢失效率, 在摇瓶中在30°C和250rpm下在液体YEPHG培养基上诱导来自YEPHD-G418-NTC琼脂平板的CP-71-HXT菌落0小时、6小时、24小时和48小时。对于0小时条件, 将细胞直接铺板在YEPHG琼脂培养基上。作为对照条件, 还在不应发生GAL10p-GIN11 (M86) 的诱导的液体YEPHD培养基中培养菌落。基于OD600, 将培养物稀释, 以通过取决于先前的诱导培养基铺板在YEPHG或YEPHD琼脂培养基上来在平板上获得单个菌落。在30°C下生长2天后, 将每种条件下的20个菌落重新铺板到YEPHD-G418和YEPHD, 以对质粒丢失效率进行评分 (图3)。无法在YEPHD-G418上生长但能够在YEPHD琼脂平板上生长的菌落丢失了Cas9表达质粒 (pDB1371)。在没有诱导的情况下 (YEPHD培养基), 约75%的菌落保留了Cas9质粒。当在YEPHG上诱导时, 所有菌落都丢失了Cas9质粒。即使没有诱导并在液体YEPHG上生长, 在时间点0小时处, 20个菌落中也只有1个在G418上生长。这表明铺板在YEPHG平板上应该足以

选择丢失Cas9质粒pDB1371的菌落。

[0172] 实施例4:通过铺板细胞在YEPHG上来直接诱导GIN11 (M86)

[0173] 考虑到第一实验的结果(实施例3),执行第二实验,其中在没有事先在液体YEPHG培养基中进行诱导的情况下,将菌落转移到YEPHG平板上(图4)。为了确定GIN11 (M86) 序列的特异性效应,将含有pCSN061的CP-61-HXT细胞作为对照,该pCSN061是不携带GIN11 (M86) 序列的Cas9表达质粒。将来自YEPHD-G418-NTC转化平板的四十个转化体直接划线到YEPHG平板上。随后,在30℃下生长2天后,将40个菌落划线到YEPHD和YEPHD-G418琼脂平板上,并在30℃下再次生长2天。在用pCSN061 (CP-61-HXT) 转化的40个菌落中,有6个仍然能够在YEPHD-G418琼脂上生长,表明CP-61-HXT菌落中的一些仍然存在含Cas9的质粒。在YEPHD-G418琼脂平板上未出现用pDB1371 (CP-71-HXT) 转化的菌落,表明所有CP-71-HXT菌落均丢失了含Cas9 GIN11 (M86) 的质粒。CP-61-HXT的40个重划线和CP-71-HXT的40个重划线如预期的那样在YEPHD平板上生长。由于不需要在液体培养基中进行诱导,因此可以在YEPHG琼脂平板上生长仅2天内从CEN.PK113-7D去除pDB1371质粒。

[0174] 实施例5:铜诱导的CP-61-HXT和CP-72-HXT的质粒丢失

[0175] 为了比较在诱导和不诱导的情况下通过诱导pDB1372上存在的CUP1p-GIN11 (M86) 构建体的质粒丢失效率,不诱导(0mM CuSO₄)或在补充有20g/l葡萄糖和0.1或0.2mM CuSO₄的矿物质培养基上诱导来自YEPHD-G418-NTC琼脂平板的CP-72-HXT菌落达24小时和48小时。随后将菌落铺板在YEPHD-G418和YEPHD琼脂培养基上。在30℃下生长2天后,对每种条件下的菌落数进行计数,以对相较于接种在YEPHD上的无诱导条件(含20g/l葡萄糖且无CuSO₄的矿物培养基)的相对质粒丢失效率进行评分。对于在0.1mM或0.2mM CuSO₄上诱导的CP-62-HXT,在诱导24小时后,>90%的细胞丢失了Cas9质粒(图5)。在诱导48h后,百分比是可比的,但更长的诱导后可能会有略微更多的细胞丢失pDB1372质粒(图5)。结果表明,含有Cas9的质粒pDB1372在被CuSO₄诱导时从CP-72-HXT被有效地丢失。

[0176] 总结

[0177] 在原养型菌株中去除含有显性抗生素抗性标记(如KanMX)的质粒的经典方式是基于细胞分裂期间的自发丢失。此经典工序需要6步并且具有不同的成功率(0-100%)。为了增加对该过程的控制,将GIN11 (M86) 自杀基因添加到含Cas9的质粒中,从而使得能够针对该质粒进行主动选择。示例描述了经由两种诱导系统诱导GIN11 (M86) 序列:

[0178] 如果GIN11 (M86) 偶联至GAL10启动子,则当直接铺板在YEPHG上以诱导GIN11 (M86) 的表达时可以在2天内以100%的效率去除Cas9质粒。

[0179] 如果GIN11 (M86) 与CUP1启动子偶联,则可以以(实施例5)的效率去除Cas9质粒。在0.1mM CuSO₄中诱导24小时作用最有效。

序列表

<110> 帝斯曼知识产权资产管理有限公司

<120> 在宿主细胞中进行基因组编辑的方法

<130> 32804-W0-PCT

<150> EP17209063.1

<151> 2017-12-20

<160> 21

<170> PatentIn版本3.5

<210> 1

<211> 7794

<212> DNA

<213> 人工序列

<220>

<223> nucleotide sequence of vector pRN1120-RFP-gRNA (A)

<400> 1

```

tcatgtttga cagcttatca tcgataatcc ggagctagca tgcggccgct tctttgaaaa 60
gataatgtat gattatgctt tcaactatat ttatacagaa acttgatggt ttctttcgag 120
tatatacaag gtgattacat gtacgtttga agtacaactc tagatthttgt agtgcctctt 180
tgggctagcg gtaaagggtc gcattttttc acaccctaca atgttctggt caaaagattt 240
tgggtcaaacg ctgtagaagt gaaagttggt gcgcatgttt cggcgttcga aacttctccg 300
cagtgaaaga taaatgatct tgttccagac acgacgtcag ttttagagct agaaatagca 360
agttaaaata aggctagtcc gttatcaact tgaaaaagtg gcaccgagtc ggtggtgctt 420
tttttgTTTT ttatgtctcc cgatatcaag cttatcgacg ctttccggca tcttcagac 480
cacagtatat ccatccgcct cctgttgagg accggtttat cattatcaat actgccattt 540
caaagaatac gtaaataatt aatagtagtg attttctaa ctttatttag tcaaaaaatt 600
agccttttaa ttctgctgta acccgtagat gcccaaaata gggggcgggt tacacagaat 660
atataacatc gtaggtgtct ggggtgaacag tttattcctg gcattcacta aatataatgg 720
agccccgctt ttaagctggc atccagaaaa aaaaagaatc ccagcaccaa aatattgttt 780
tcttcaccaa ccatcagttc ataggtccat tctcttagcg caactacaga gaacaggggc 840
acaaacaggc aaaaaacggg cacaacctca atggagtgat gcaacctgcc tggagtaaat 900
gatgacacaa ggcaattgac ccacgcatgt atctatctca ttttcttaca ctttctatta 960
ccttctgctc tctctgattt ggaaaaagct gaaaaaaaag gttgaaacca gttccctgaa 1020
attattcccc tacttgacta ataagtatat aaagacggta ggtattgatt gtaattctgt 1080
aaatctatth cttaaacttc ttaaatteta cttttatagt tagtcttttt tttagtttta 1140
aaacaccaag aacttagttt cgaataaaca cacataaaga attcaaaatg gtttcaaaag 1200
gtgaagaaga taatatggct attattaaag aatttatgag atttaaagtt catatggaag 1260
gttcagttaa tggatcatgaa tttgaaattg aaggtgaagg tgaaggtaga ccatatgaag 1320
gtactcaaac tgctaaattg aaagttacta aaggtggtcc attaccattt gcttgggata 1380

```

ttttgtcacc acaatttatg tatggttcaa aagcttatgt taaacatcca gctgatattc 1440
cagattatatt aaaattgtca tttccagaag gttttaaagt ggaaagagtt atgaattttg 1500
aagatgggtg tgttgttact gttactcaag attcatcatt acaagatggg gaatttattt 1560
ataaagttaa attgagaggt actaattttc catcagatgg tccagttatg caaaaaaaaa 1620
ctatggggtg ggaagcttca tcagaaagaa tgtatccaga agatggtgct ttaaaagggtg 1680
aaattaaaca aagattgaaa ttaaaagatg gtggtcatta tgatgctgaa gttaaaacta 1740
cttataaagc taaaaaacca gttcaattac caggtgctta taatgttaat attaaattgg 1800
atattacttc acataatgaa gattatacta ttgttgaaca atatgaaaga gctgaaggta 1860
gacattcaac tgggtggtatg gatgaattat ataaataatc tagacaaatc gctcttaaat 1920
atatacctaa agaacattaa agctatatta taagcaaaga tacgtaaatt ttgcttatat 1980
tattatacac atatcatatt tctatatttt taagatttgg ttatataatg tacgtaatgc 2040
aaaggaaata aattttatac attattgaac agcgtccaag taactacatt atgtgcacta 2100
atagtttagc gtcgtgaaga ctttattgtg tcgcaaaaag taaaaatttt aaaaattaga 2160
gcacctgaa cttgcgaaaa aggttctcat caactgttta aaagatctga gctcgcagct 2220
tttgttcct ttagtgaggg ttaattccga gcttggcgta atcatggtea tagctgtttc 2280
ctgtgtgaaa ttgttatccg ctcaaatc cacacaacat aggagccgga agcataaagt 2340
gtaaagcctg ggggtgcctaa tgagtgaggt aactcacatt aattgcgttg cgctcactgc 2400
ccgctttcca gtcgggaaac ctgtcgtgcc agctgcatta atgaatcggc caacgcgcgg 2460
ggagaggcgg tttgcgtatt gggcgtctt ccgcttctc gctcactgac tcgctgcgct 2520
cggtcgttcg gctgcggcga gcggtatcag ctactcaaa ggcgtaata cggttatcca 2580
cagaatcagg ggataacgca ggaaagaaca tgtgagcaaa aggccagcaa aaggccagga 2640
accgtaaaaa ggccgcgtt ctggcgtttt tccataggct cggccccct gacgagcatc 2700
acaaaaatcg acgctcaagt cagaggtggc gaaaccgac aggactataa agataccagg 2760
cgttcccccc tggaaagctc ctctgtcgct ctctgttcc gacctgccg cttaccggat 2820
acctgtccgc ctttctccct tcgggaagcg tggcgcttcc tcaatgctca cgctgtaggt 2880
atctcagttc ggtgtaggtc gttcgtcca agctgggctg tgtgcacgaa cccccgttc 2940
agccccaccg ctgcgcctta tccgtaact atcgtcttga gtccaaccg gtaagacacg 3000
acttatcgcc actggcagca gccactggta acaggattag cagagcgagg tatgtaggcg 3060
gtgctacaga gttcttgaag tgggtggcta actacggcta cactagaagg acagtatttg 3120
gtatctgcgc tctgctgaag ccagttacct tcggaaaaag agttggtagc tcttgatccg 3180
gcaaacaaac caccgctggt agcggtggtt ttttgtttg caagcagcag attacgcgca 3240
gaaaaaaaaagg atctcaagaa gatccttga tctttctac ggggtctgac gctcagtgga 3300
acgaaaactc acgttaaggg attttggtea tgagattatc aaaaaggatc ttcacctaga 3360
tccttttaaa ttaaaaatga agttttaa atcaatctaaag tatatatgag taaacttggg 3420
ctgacagtta ccaatgctta atcagtgagg cacctatctc agcgatctgt ctatttcggt 3480
catccatagt tgctgactg cccgtcgtgt agataactac gatacgggag ggcttaccat 3540
ctggccccag tgctgcaatg ataccgcgag acccagctc accggtcca gatttatcag 3600
caataaacca gccagccgga agggccgagc gcagaagtgg tcctgcaact ttatccgctt 3660
ccatccagtc tattaattgt tgccgggaag ctagagtaag tagttcgcca gttaatagtt 3720

tgcgcaacgt tgttgccatt gctacaggca tcgtggtgtc acgctcgtcg tttggtatgg 3780
cttcattcag ctccggttcc caacgatcaa ggcgagttac atgatcccc atgttgtgaa 3840
aaaaagcggg tagctccttc ggtcctccga tcgttgtcag aagtaagttg gccgcagtgt 3900
tatcactcat gggtatggca gcaactgcata attctcttac tgtcatgcca tccgtaagat 3960
gcttttctgt gactggtgag tactcaacca agtcattctg agaatagtgt atgcbggcagc 4020
cgagttgctc ttgcccggcg tcaatacggg ataataccgc gccacatagc agaactttaa 4080
aagtgctcat cattggaaaa cgttcttcgg ggcgaaaact ctcaaggatc ttaccgctgt 4140
tgagatccag ttcgatgtaa cccactcgtg cacccaactg atcttcagca tcttttactt 4200
tcaccagcgt ttctgggtga gcaaaaacag gaaggcaaaa tgccgcaaaa aaggaataa 4260
ggcgacacg gaaatgttga atactcatac tcttctttt tcaatattat tgaagcattt 4320
atcagggtta ttgtctcatg agcggataca tatttgaatg tatttagaaa aataaaciaa 4380
taggggttcc gcgcacattt ccccgaagag tgccacctga cgtctaagaa accattatta 4440
tcatgacatt aacctataaa aataggcgta tcacgagcc ctttcgtctc gcgcgtttcg 4500
gtgatgacgg tgaaaacctc tgacacatgc agctcccga gacggtcaca gcttgtctgt 4560
aagcggatgc cgggagcaga caagcccgtc agggcgcgtc agcgggtgtt ggcgggtgtc 4620
ggggctggct taactatgcg gcatcagagc agattgtact gagagtgcac catatcgact 4680
acgtcgtaaag gccgtttctg acagagtaaa attcttgagg gaactttcac cattatggga 4740
aatggttcaa gaaggtattg acttaaactc catcaaatgg tcaggtcatt gagtgttttt 4800
tatttgttgt attttttttt ttttagagaa aatcctcaa tatcaaatta ggaatcgtag 4860
tttcatgatt ttctgttaca cctaactttt tgtgtggtgc cctctcctt gtcaatatta 4920
atgttaaagt gcaattcttt ttcttatca cgttgagcca ttagtatcaa tttgcttacc 4980
tgtattcctt tactatcctc cttttctcc ttcttgataa atgtatgtag attgcgtata 5040
tagtttcgtc taccctatga acatattcca ttttgtaatt tcgtgtcgtt tctattatga 5100
atttcattta taaagtttat gtacacctag gatccgtcga cactggatgg cggcgttagt 5160
atcgaatcga cagcagtata gcgaccagca ttacatacg attgacgat gatattactt 5220
tctgcgcact taacttcgca tctgggcaga tgatgtcag gcgaaaaaaaa atataaatca 5280
cgctaacatt tgattaaaat agaacaacta caatataaaa aactataca aatgacaagt 5340
tcttgaaaac aagaatcttt ttattgtcag tactaggggc agggcatgct catgtagagc 5400
gcctgctcgc cgtccgaggc ggtgccgtc tacagggcgg tgtccaggcc gcagagggtg 5460
aaccccatcc gccggtacgc gtggatcgc ggtgcgttga cgttggtgac ctccagccag 5520
aggtgcccgg cgcctcgtc gcgggcgaac tccgtcgcga gcccataca cgcgcgccc 5580
accccgctcc cccggtgctc cggggcgacc tcgatgtct cgcaggtcag ccggcggttc 5640
cagccggagt acgagacgac cacgaagccc gccaggtcgc cgtcgtcccc gtacgcgacg 5700
aacgtccggg agtccgggtc gccgtctcc cggcgtccg attcgtcgtc cgattcgtcg 5760
tcggggaaca ccttggtcag gggcgggtcc accggcact cccgcagggt gaagccgtcc 5820
ccggtggcgg tgacgcggaa gacggtgtcg gtggtgaagg acccatccag tgctcgtatg 5880
gcctcggcgt ccccgggac actggtgcgg taccgtaag ccgtgtcgtc aagagtggtc 5940
attttacatg gttgtttatg ttcggatgtg atgtgagaac tgtatcctag caagatttta 6000
aaaggaagta tatgaaagaa gaacctcagt ggcaaatcct aaccttttat atttctctac 6060

aggggcgcgg cgtggggaca attcaacgcg tctgtgaggg gagcgtttcc ctgctcgcag 6120
 gtctgcagcg aggagccgta atttttgctt cgcgccgtgc ggccatcaaa atgtatggat 6180
 gcaaatgatt atacatgggg atgtatgggc taaatgtacg ggcgacagtc acatcatgcc 6240
 cctgagctgc gcacgtcaag actgtcaagg agggatttct gggcctccat gtcgctggcc 6300
 gggtgacccg gcggggacga ggccttaagt tcgaacgtac gagctccggc attgcgaata 6360
 ccgctttcca caaacattgc tcaaaagtat ctctttgcta tataatctctg tgctatatcc 6420
 ctatataacc taccatcca ctttcgctc ctggaacttg catctaaact cgacctctac 6480
 attttttatg tttatctcta gtattactct ttagacaaaa aaattgtagt aagaactatt 6540
 catagagtga atcgaaaaca atacgaaaat gtaaacttt cctatacgta gtatatagag 6600
 acaaaataga agaaaccgtt cataattttc tgaccaatga agaatcatca acgctatcac 6660
 tttctgttca caaagtatgc gcaatccaca tcggtataga atataatcgg ggatgccttt 6720
 atcttgaaaa aatgcacccg cagcttcgct agtaatcagt aaacgcggga agtggagtca 6780
 ggcttttttt atggaagaga aaatagacac caaagtagcc ttcttctaac cttaacggac 6840
 ctacagtgca aaaagtattc aagagactgc attatagagc gcacaaagga gaaaaaagt 6900
 aatctaagat gctttgtag aaaaatagcg ctctcgggat gcatttttgt agaacaaaa 6960
 agaagtatag attctttgtt ggtaaaatag cgctctcgcg ttgcatttct gttctgtaaa 7020
 aatgcagctc agattctttg tttgaaaaat tagcgtctc gcgttgcatt tttgttttac 7080
 aaaaatgaag cacagattct tcgttgtaa aatagcgtt tcgcttgca tttctgttct 7140
 gtaaaaatgc agctcagatt ctttgtttga aaaattagcg ctctcgcgtt gcatttttgt 7200
 tctacaaaat gaagcacaga tgcttcgta acaaagatat gctattgaag tgcaagatgg 7260
 aaacgcagaa aatgaaccgg ggatgcgacg tgcaagatta cctatgcaat agatgcaata 7320
 gtttctccag gaaccgaaat acatacattg tcttccgtaa agcgttagac tatatattat 7380
 tatacagggt caaatatact atctgtttca gggaaaactc ccaggttcgg atgttcaaaa 7440
 ttcaatgatg ggtaacaagt acgatcgtaa atctgtaaaa cagtttgctg gatattaggc 7500
 tgtatctcct caaagcgtat tcgaatatca ttgagaagct gcagcgtcac atcggataat 7560
 aatgatggca gccattgtag aagtgccttt tgcatttcta gtctcttctt cggcttagct 7620
 agttttacta catcgcgaag atagaatctt agatcacact gcctttgctg agctggatca 7680
 atagagtaac aaaagagtgg taaggcctcg ttaaaggaca aggacctgag cggaagtgt 7740
 tcgtacagta gacggagtat actaggtata gtctatagtc cgtggaatta attc 7794

<210> 2

<211> 1387

<212> DNA

<213> 人工序列

<220>

<223> nucleotide sequence of synthetic expression cassette cFS0017

(pGAL10-GIN11 (M86))

<400> 2

ctgaaccgc ggcatgtgaa taagaagtaa tacaaccga aatgttgaa agtattagtt 60
 aaagtggta tgcagttttt gcatttata atctgttaat agatcaaaaa tcatcgttc 120

gctgattaat taccccagaa ataaggctaa aaaactaatc gcattatcat cctatggttg 180
 ttaatttgat tcggtcattt gaaggtttgt ggggccaggt tactgccaat ttttcctctt 240
 cataaccata aaagctagta ttgtagaatc tttattgttc ggagcagtgc ggcgcgaggc 300
 acatctgcgt ttcaggaacg cgaccggtga agacgaggac gcacggagga gagtcttctt 360
 tcggagggct gtcacccgct cggcggcttc taatccgtac ttcaatatag caatgagcag 420
 ttaagcgat tactgaaaagt tccaaagaga aggttttttt aggctaagat aatggggctc 480
 tttacatttc cacaacatat aagtaagatt agatatggat atgtatatgg atatgtatat 540
 ggtggtaatg ccatgtaata tgattattaa acttctttgc gtccatcaa aaaaaagta 600
 agaatttttg aaaattcaag gaatttcgac ggatcaataa cagtgtttgt ggagcatttt 660
 ctgaatacaa taaacccaaa acagaaactt cctttttgta tcaactgttct ggaaaagggg 720
 tgggcggtaa taaagctaat aggggtgtgtc cataagtaat actgaacttg gaaatgtgcg 780
 gctttgcagc attttgtctt tctataaaaa tgtgtcgttc ctttttttca ttttttggcg 840
 cgtcgcctcg gggtcgtata gaatatgctt cacttttaaa aataagattg cagatcaggg 900
 caaaacaagt agcaaatcat agcaagagac cctgattttt gtgacataaa tttttttact 960
 tctgtgttag gttaactttt tatgtaactg taaatggaat agagttgagg ggatagtgcc 1020
 cacaagtcaa tatgtttatt ttgtaaagtt gaaagataat ttttttatg ctcaggtgat 1080
 tttgggtgtg aattttctgt aatattaaca taagagtaat acattgagtg gttagtatat 1140
 ggtgtaaaag tggataaacg catgtattaa gagcagttat acaatatttg gggccgctga 1200
 atgagatata gatattaata tgtggataat catgggcttt atgggtaaat ggaacagggt 1260
 atagaccact gaggcaagtg ccgtgcataa tgatatgagt gcatctagtg gcgaacgtgg 1320
 cgagaaagga agggaagaaa gcgagtcca tctgtgcaga caaacgcac aggatactag 1380
 tccttga 1387

<210> 3

<211> 1385

<212> DNA

<213> 人工序列

<220>

<223> nucleotide sequence of synthetic expression cassette cFS0018
(pCUP1-GIN11 (M86))

<400> 3

aaccgcgggg tacctaagga gatttcagat ttttaaatgg aaagagaagt tgtccaaagg 60
 agtataatta ttgacaagga tttggaatct gataatctgg gtattactac ggcaaacttc 120
 aacgatttct atgatgcatt ttataattag taagecgatc ccattaccga catttgggcg 180
 ctatacgtgc atatgttcat gtatgtatct gtatttaaaa cacttttgta ttatttttcc 240
 tcatatatgt gtataggttt atacggatga ttttaattatt acttcaccac cttttatttc 300
 aggctgatat cttagccttg ttactagtta gaaaaagaca tttttgctgt cagtcactgt 360
 caagagattc ttttgctggc atttcttcta gaagcaaaaa gagcgatgcg tcttttccgc 420
 tgaaccgttc cagcaaaaaa gactaccaac gcaatatgga ttgtcagaat catataaaag 480
 agaagcaaat aactccttgt cttgtatcaa ttgcattata atatcttctt gttagtgcaa 540

tatcatatag aagtcacga aatagatatt aagaaaaaca aactgtacaa tcaatcaatc 600
aatcatcaca taaaaggaat ttcgacggat caataacagt gtttgtggag cattttctga 660
atacaataaa cccaaaacag aaacttcctt tttgtatcac tgttctggaa aaggggtggg 720
cggtaataaa gctaataagg tgtgtccata agtaatactg aacttgaaa tgtgcggtt 780
tgcagcattt tgtctttcta taaaaatgtg tcgttccttt ttttcatttt ttggcgcgtc 840
gcctcggggg cgtatagaat atgcgtcact tttaaaaata agattgcaga tcagggcaaa 900
acaagtagca aatcatagca agagaccctg atttttgtga cataaatatt tttacttctg 960
tgtaggtta actttttatg taactgtaa tggaatagag ttgaggggat agtgcccaca 1020
agtcaatatg tttattttgt aaagttgaaa gataattatt tttatgctca ggtgattttg 1080
gtgttgaatt ttctgtaata ttaacataag agtaatacat tgagtgggta gtatatgggtg 1140
taaaagtggg ataacgcatg tattaagagc agttatacaa ttttggggc cgctgaatga 1200
gatatagata ttaaaatgtg gataatcatg ggctttatgg gtaaattgaa cagggtatag 1260
accactgagg caagtgccgt gcataatgat atgagtgcac ctagtggcga acgtggcgag 1320
aaaggaaggg aagaaagcga gtgccatctg tgcagacaaa cgcacagga tgccggcccc 1380
cggaa 1385

<210> 4

<211> 12783

<212> DNA

<213> 人工序列

<220>

<223> nucleotide sequence of vector pDB1371

<400> 4

ccatttgaat aagaagtaat acaaaccgaa aatgttgaaa gtattagtta aagtggttat 60
gcagtttttg catttatata tctgttaata gatcaaaaat catcgcttcg ctgattaatt 120
accccagaaa taaggctaaa aaactaatcg cattatcatc ctatggttgt taatttgatt 180
cgttcatttg aaggtttgtg gggccagggt actgccaatt tttcctcttc ataaccataa 240
aagctagtat ttagaatctt ttattgttcg gagcagtgcg gcgagaggca catctgcgtt 300
tcaggaacgc gaccggtgaa gacgaggacg cacggaggag agtcttcctt cggagggctg 360
tcaccgctc ggcggttctt aatccgtact tcaatatagc aatgagcagt taagcgtatt 420
actgaaagt ccaaagagaa ggttttttta ggctaagata atggggctct ttacatttcc 480
acaacatata agtaagatta gatatggata tgtatatgga tatgtatatg gtggtaatgc 540
catgtaatat gattattaaa cttctttgcg tccatccaaa aaaaagtaa gaatttttga 600
aaattcaagg aatttcgacg gatcaataac agtgtttgtg gagcattttc tgaatacaat 660
aaacccaaaa cagaaacttc cttttgtat cactgttctg gaaaaggggt gggcggtaat 720
aaagctaata ggggtgtgcc ataagtaata ctgaacttgg aatgtgcgg ctttgcagca 780
ttttgtcttt ctataaaaat gtgtcgttcc ttttttcat tttttggcgc gtcgcctcgg 840
ggtcgtatag aatatgcgtc acttttaaaa ataagattgc agatcagggc aaaacaagta 900
gcaaatcata gcaagagacc ctgatttttg tgacataaat atttttactt ctgtgttagg 960
ttaacttttt atgtaactgt aaatggaata gagttgaggg gatagtgcc acaagtcaat 1020

atgtttatTT tgtaaagttg aaagataatt atTTTTatgc tcaggtgatt ttggtgttga 1080
atTTTctgta atattaacat aagagtaata cattgagTgg ttagtatatg gtgTaaaagt 1140
ggTataacgc atgtattaag agcagttata caatattTgg ggccgctgaa tgagatatag 1200
atattaaaaT gtggataatc atgggctTTa tgggTaaatg gaacagggtA tagaccactg 1260
aggcaagTgc cgtgcataat gatatgagTg catctagTgg cgaacgtggc gagaaaggaa 1320
gggaagaaaag cgagTgccat ctgtgcagac aaacgcatac ggatgccggc tttccccgTc 1380
aagctctaaa TcgggggTc ctttagggT tccgatttag Tgctttacgg cacctcgacc 1440
ccaaaaaaact Tgattagggt gatggTtTc gtagTgggC atcgccctga tagacggTtt 1500
ttcgcccttt gacgtTggag TccacgtTct ttaatagTgg actctTgtTc caaactggaa 1560
caacactcaa ccctatctcg gtctattctt ttgatttata agggattTtTg ccgattTcgg 1620
cctattggTt aaaaaatgag ctgattTaaC aaaaattTaa cgcgaattTt aacaaaatat 1680
taacgtTtTc aattTcctga TgcggTattt TctctTtag catctgtgcg gtattTcaca 1740
ccgcataggc aagTgcacaa acaatactta aataaatact actcagTaat aacctattTc 1800
ttagcattTt Tgacgaaatt TgctattTtTg ttagagTctt ttacaccatt Tgtctccaca 1860
cctccgTtTc catcaacacc aataacgcca tTtaatctaa gcgcatacacc aacattTtct 1920
ggcgtcagTc caccagctaa cataaaaTgt aagctTtTcg ggctctctTg cttTccaacc 1980
cagTcagaaa TcgagTtcca atccaaaagT Tcacctgtcc cacctgctTc Tgaatcaaac 2040
aagggaaTaa acgaaTgagg tTtctgtgaa gctgcactga gtagtatgtt gcagTctTtt 2100
ggaaatacga gtctTtTaat aactggcaaa ccgaggaact cttggTattc ttgccacgac 2160
Tcatctccat gcagTtggac gatatcaatg ccgTaatcat Tgaccagagc caaaacatcc 2220
Tccttaggtt gattacgaaa cacgccaacc aagTattTcg gagtgcctga actattTtTtA 2280
tatgtctTtTc caagactTga aattTtctt gcaataaccg ggtcaattgt TctctTtTcTc 2340
ttgggcacac atataatacc cagcaagTcA gcatcggaat ctagagcaca ttctgcggcc 2400
Tctgtgctct gcaagccgca aactTtTcacc aatggaccag aactacctgt gaaattaata 2460
acagacatac Tccaagctgc cttTgtgtgc tTaatcacgt atactcacgt gctcaatagt 2520
caccaatgcc ctccctctTg gccctctct tTtctTtTtTt cgaccgaatt aattctTaat 2580
cggcaaaaaa agaaaagTc cggatcaaga ttgtacgTaa ggtgacaagc tattTtTtcaa 2640
taaagaatat ctTccactac Tgccatctgg cgtcataact gcaaagtaca catatattac 2700
gatgctgtct attaaatgct Tcctatatta tatatatagt aatgtcgTtt atggTgcact 2760
ctcagTacaa Tctgctctga Tgccgcatag tTaaagccagc cccgacacc gccaacacc 2820
gctgacgcgc cctgacggc ttgtctgtc ccggcatccg cttacagaca agctgtgacc 2880
gtctccggga gctgcatgtg TcagaggTtt Tcacctcat caccgaaacg cgcgagacga 2940
aagggcctcg Tgatacgctt atTTTtatag gTtaatgtcA Tgataataat ggtTtcttag 3000
gacggatcgT ttgcctgTaa cttacacgcg cctcgtatct tTtaatgatg gaataattTg 3060
ggaattTact ctgtgtTtTt tTattTtTtTt gTttTgtatt TggattTtag aaagTaaata 3120
aagaaggtag aagagTtTcG gaatgaagaa aaaaaataa acaaaggtTt aaaaaattTc 3180
aacaaaaagc gtactTtTaca tatatattTtA ttagacaaga aaagcagatt aaatagatat 3240
acattcgatt aacgataagT aaaaTgTaaa atcacaggat tTtctgtgtgT ggtctTctac 3300
acagacaaga Tgaaacaatt cggcattaat acctgagagc aggaagagca agataaaagg 3360

tagtatttgt tggcgatccc cctagagtct tttacatctt cggaaaacaa aaactatttt 3420
ttctttaatt tcttttttta ctttctattt ttaatttata tatttatatt aaaaaattta 3480
aattataatt atttttatag cacgtgatga aaaggacca ggtggcactt ttcggggaaa 3540
tgtgcgcgga acccctattt gtttattttt ctaaatacat tcaaataatgt atccgctcat 3600
gagacaataa ccctgataaa tgcttcaata atattgaaaa aggaagagta tgagtattca 3660
acatttccgt gtcgccctta ttcccttttt tgcggcattt tgccttctctg tttttgctca 3720
cccagaaacg ctggtgaaag taaaagatgc tgaagatcag ttgggtgcac gagtgggtta 3780
catcgaactg gatctcaaca gcggtaaagat ccttgagagt tttcgccccg aagaacgttt 3840
tccaatgatg agcactttta aagttctgct atgtggcgcg gtattatccc gtattgacgc 3900
cgggcaagag caactcggtc gccgcataca ctattctcag aatgacttgg ttgagtactc 3960
accagtcaca gaaaagcatc ttacggatgg catgacagta agagaattat gcagtgtctgc 4020
cataaccatg agtgataaca ctgcggccaa cttacttctg acaacgatcg gaggaccgaa 4080
ggagctaacg gcttttttgc acaacatggg ggatcatgta actcgccttg atcgttggga 4140
accggagctg aatgaagcca taccaaacga cgagcgtgac accacgatgc ctgtagcaat 4200
ggcaacaacg ttgcgcaaac tattaactgg cgaactactt actctagctt cccggcaaca 4260
attaatagac tggatggagg cggataaagt tgcaggacca cttctgcgct cggcccttcc 4320
ggctggctgg tttattgctg ataaatctgg agccggtgag cgtgggtctc gcggtatcat 4380
tgcagcactg gggccagatg gtaagccctc ccgtatcgta gttatctaca cgacggggag 4440
tcaggcaact atggatgaac gaaatagaca gatcgtgag ataggtgcct cactgattaa 4500
gcattggtaa ctgtcagacc aagtttactc atatatactt tagattgatt taaaacttca 4560
tttttaattt aaaaggatct aggtgaagat cttttttgat aatctcatga ccaaaatccc 4620
ttaacgtgag ttttcgttcc actgagcgtc agaccccgta gaaaagatca aaggatcttc 4680
ttgagatcct ttttttctgc gcgtaatctg ctgcttgcaa acaaaaaaac caccgctacc 4740
agcggtggtt tgtttgccgg atcaagagct accaactctt tttccgaagg taactggctt 4800
cagcagagcg cagataccaa atactgtcct tctagtgtag ccgtagttag gccaccactt 4860
caagaactct gtagcaccgc ctacatacct cgctctgcta atcctgttac cagtggctgc 4920
tgccagtggc gataagtcgt gtcttaccgg gttggactca agacgatagt taccggataa 4980
ggcgcagcgg tcgggctgaa cgggggggtc gtgcacacag cccagcttgg agcgaacgac 5040
ctacaccgaa ctgagatacc tacagcgtga gctatgagaa agcggccacgc ttcccgaagg 5100
gagaaaaggcg gacaggtatc cggtaaagcgg cagggtcgga acaggagagc gcacgagga 5160
gcttccaggg gaaaacgcct ggtatcttta tagtctgtc gggtttcgcc acctctgact 5220
tgagcgtcga tttttgtgat gctcgtcagg ggggaggagc ctatggaaaa acgccagcaa 5280
cgcgcccttt ttacggttcc tggccttttg ctggcctttt gctcacatgt tctttctctgc 5340
gttatcccct gattctgtgg ataaccgtat taccgctttt gactgagctg ataccgctcg 5400
ccgcagccga acgaccgagc gcagcgagtc agtgagcgag gaagcgaag agcgcceaat 5460
acgcaaaccg cctctccccg cgcgttggcc gattcattaa tgcagctggc acgacaggtt 5520
tcccactgg aaagcgggca gtgagcgcaa cgcaattaat gtgagttacc tctctatta 5580
ggcaccaccag gctttacact ttatgcttcc ggctcctatg ttgtgtggaa ttgtgagcgg 5640
ataacaattt cacacaggaa acagctatga ccatgattac gccaaagcgc caattaacc 5700

tcactaaagg gaacaaaagc tggagctcca ccgcggtggc ggccgcatag gccactagtg 5760
gatctgattc gaattctacc gttcgtatag catacattat acgaagtat gagctcgttt 5820
tcgacactgg atggcggcgt tagtatcgaa tcgacagcag tatagcgacc agcattcaca 5880
tacgattgac gcatgatatt actttctgcg cacttaactt cgcactctggg cagatgatgt 5940
cgaggcgaaa aaaaatataa atcacgctaa catttgatta aatagaaca actacaatat 6000
aaaaaaaaacta tacaaatgac aagttcttga aaacaagaat ctttttattg tcagtactga 6060
ttagaaaaaac tcatcgagca tcaaatgaaa ctgcaattta ttcatatcag gattatcaat 6120
accatatttt tgaaaaagcc gtttctgtaa tgaaggagaa aactcaccca ggcagttcca 6180
taggatggca agatcctggt atcggctctgc gattccgact cgtccaacat caatacaacc 6240
tattaatttc cctcgtcaa aaataagggt atcaagttag aatcacat gagtgacgac 6300
tgaatccggt gagaatggca aaagcttatg catttcttct cagacttggt caacaggcca 6360
gccattacgc tcgtcatcaa aatcaactgc atcaacaaa cegtattca ttcgtgattg 6420
cgctgagcg agacgaaata cgcgatcgt gttaaaagga caattacaaa caggaatcga 6480
atgcaaccgg cgcaggaaca ctgccagcgc atcaacaata ttttcacctg aatcaggata 6540
ttcttcta atctggaatg ctgttttgcc ggggatcgca gtggtgagta accatgcatc 6600
atcaggagta cggataaaat gcttgatggt cggaagaggc ataaattccg tcagccagtt 6660
tagtctgacc atctcatctg taacatcatt ggcaacgcta cctttgccat gtttcagaaa 6720
caactctggc gcatcgggct tcccatacaa tcgatagatt gtcgcacctg attgcccgcac 6780
attatcgca gccatttat accatataa atcagcatcc atgttggaat ttaatcgcg 6840
cctcgaaacg tgagtctttt ccttaccat gttgtttat gttcggatgt gatgtgagaa 6900
ctgtatccta gcaagatttt aaaaggaagt atatgaaaga agaacctcag tggcaaatcc 6960
taacctttta tatttctcta caggggcgcg gcgtggggac aattcaacgc gtctgtgagg 7020
ggagcgtttc cctgctcgca ggtctgcagc gaggagccgt aatttttgct tcgcgccgtg 7080
cggccatcaa aatgtatgga tgcaaatgat tatacatggg gatgtatggg ctaaattgac 7140
ggcgacagc cacatcatgc cctgagctg cgcacgta gactgtcaag gagggtattc 7200
tgggcctcca tgtcgtggc cgggtgacc ggcggggac aggcaagcta aacagatcta 7260
taacttcgta tagcatacat tatacgaacg gtagaattcg tcgacctgca gcgtacgaag 7320
cttcagctgg cggccgcaac aatagcgtc gaaaggcgg caataggtct agaaacttgt 7380
tcaagttctt agagactata tggcgtacag aaggtcttgc ggccctctac acgggcctgg 7440
cagccagagt aattaagata gcgccaagtt gcgccatcat gatattagt tatgagatct 7500
ccaaaaaagt atttgaaac aaattgcatc agtgaataaa ggcttgtaa tatagatata 7560
tagtaaacga aaagaagcat atacgtataa ttatttggg gaacggctct agaaaagaaa 7620
actttgcctt taactcctt acaccttct cttcttctt ggatcagctc tggaatcacc 7680
accaattga gacaagtcaa ttctggttct gtacaaaccg gtaatagatt ggtgaatcaa 7740
agtagcatct aaaacttctt tgggtggaagt gtatctctt ctgtcaatgg tagtgcgaa 7800
gtatttgaaa gcagctggag cacccaagtt ggttaaagt acaaatgaa tgatgttttc 7860
agcttgctt ctgattggtt tgtctctgtg cttgttgtaa gcgataaaa cttatctaa 7920
gttagcatca gcaaaaatga ctctcttggga gaattcggag atttgttcaa taatttcac 7980
caagtagtgc ttgtgtgtt caacaaaca ttgctttgt tcgttctt ctggggaacc 8040

tttcaacttt tcgtagtggg aagccaagta caagaagtta acgtacttag atggtaaagc 8100
caattcgta cctttttgta gttcaccagc ggaggccaac attctctttc taccgttttc 8160
caattcgaac aaggagtatt ttggtaactt aatgatcaaa tccttcttga cttccttgta 8220
acccttagct tccaaaaagt cgattgggtt cttttcgaag gaggatcttt ccatgatggt 8280
gatacccaac aattccttaa cagacttcaa cttcttagac ttaccctttt caaccttagc 8340
aacgaccaaaa acggagtaag caacagttgg agagtcgaaa ccaccgtact tcttaggatc 8400
ccagtccttc tttctagcaa tcaactgtc agagtttctc tttggcaaaa tggattcctt 8460
agagaaacca ccagtttgaa cttcagtctt cttgacgatg ttaacttggtg gcatagacaa 8520
aacctttctg acggtagcga aatctctacc cttgtcccag acaatttcac cagtttcacc 8580
attggtttca atcaatggac gctttctaatt ttcaccgtta gctaaagtga tttcagtctt 8640
gaaaaagttc atgatgtag agtagaagaa gtacttagca gtggccttac caatttcttg 8700
ttcagacttg gcgatcatct ttctaacatc gtaaacctg tagtcaccgt aaacgaattc 8760
agattccaac tttgggtact ttttgattaa ggcagtaccg acaacagcgt tcaagtaggc 8820
atcgtgagcg tgatgtagt tgttgatttc tctgacctg taaaattgga agtcctttct 8880
gaagtcagaa accaacttag acttcaaagt gatgacctt acttctctaa ttagtttgtc 8940
gttttcatcg tacttagtgt tcattctgga atccaagatt tgagcaacat gcttggtgat 9000
ttgtctagtt tcgactaatt gtctcttgat gaaaccggct ttgtccaatt cggacaaacc 9060
acctctttca gccttggta agttgtcga cttcttttga gtgatcaact tagcattcaa 9120
caattgtctc cagtagttct tcacttctt aacaacttct tcagatggaa cgttatcaga 9180
cttacctctg ttcttgtcag atctagtcaa aactttgttg tcaatggaat cgtccttcaa 9240
gaacgattgt gggacgatat gatcgacatc gtagtcagac aatctgttga tatccaattc 9300
ttggtcgacg tacatgtcac gaccgtttt caagtagtac aagtatagct tttcgttttg 9360
taattgagtg ttttcgactg ggtgttcttt caaaatttga gaaccaact ccttgatacc 9420
ttcttcaatt ctctcatac gttctctaga gttcttttga cccttttgag tagtttggtt 9480
ttctctagcc atttcgatga caatattttc tggcttgtgt ctaccatga ctttgaccaa 9540
ttcatcaacg acctgacgg tttgtaagat accttctta atagctggag aaccagccaa 9600
gttagcgatg tgttcgtgca aagaatcacc ttgaccagag acttgggctt tttgaatgtc 9660
ttccttgaaa gtcaaagaat cgtcgtgaat caattgcata aagtttctgt tagcgaaacc 9720
atcggatttc aaaaagtcta aaatagtctt accggattgc ttgtctctga taccgttaat 9780
caactttctg gacaatctac cccaaccagt gtatcttctt ctcttttagt gcttcataac 9840
tttatcgctg aacaagtgag cgtaggtctt caatctctt tcaatcattt ctctgtcctc 9900
gaaaagagtc aaggtaaaaa cgatatctt caagatgtct tegtcttctt cgttatctaa 9960
aaagtccttg tccttgatga tctttaacaa atcgtggtag gtgcccaaag aagcattgaa 10020
acggtcttca acaccggaga tttcgacgga atcgaaacat tcaatcttct tgaagtagtc 10080
ttccttcaat tgcttaacag tgaccttctt gttggtctta aacaacaagt caacaatagc 10140
cttcttttgt tcaccggata ggaaagctgg ctttctcata ccttcagtaa cgtatttaac 10200
cttggttaat tcgltgtaga cggatgaagta ttcgtacaac aaagagtgtt ttggcaagac 10260
cttctcgttt ggcaagttct tatcaaagtt ggtcattctt tcgatgaaag attgggcaga 10320
agcacccttg tcgacgactt cttcgaagtt ccatggggtg atggttctt cagactttct 10380

ggatcatcaa gcgaatctag aattacctct ggccaatgga cgcacgtagt atgggattct 10440
gaaagttaag atcttttcga tcttttctct gttgtccttt aggaatggat agaaatcttc 10500
ctgtcttctc aaaatggcgt gcaattcacc caagtggatt tgggtgtggga tagaacctgt 10560
atcgaaggta cgttgctttc tcaataagtc ttctctgttc aacttaacca ataattcttc 10620
agtaccatcc atcttttcca aaattggctt gatgaacttg tagaattctt cctgagaagc 10680
accaccgtca atgtaaccgg cgtaaccatt tttggattgg tcgaagaaga tttccttgta 10740
cttttctggc aattgttgtc taaccaaagc cttcaacaaa gtcagatctt ggtgggtgttc 10800
gtcgtatctt ttgatcatag aagcagacaa tggagccttg gtaatttcag tgttaactct 10860
caagatgtca gacaacaaga tagcgtcaga taagtttttg gcagccaaga acaagtcggc 10920
gtattggca ccgatttgag ccaacaagtt gtctaagtcg tcgtcgtagg tgtccttgga 10980
caattgcaac ttggcatctt cagccaagtc gaagttggac ttgaagtttg gggtaaaacc 11040
caaggacaaa gcgatcaagt taccgaacaa accgttttct tttcaccag gcaattgagc 11100
aatcaagttt tccaaacgct tagacttgga caaacggca gataagatgg ccttagcatc 11160
aacaccagaa gcgttaattg ggttttctc gaataattgg ttgtaggttt ggaccaattg 11220
gatgaacaat ttgtcgacgt cagagttgtc tgggttcaag tcaccttcaa tcaagaagtg 11280
acctctgaac ttgatcatgt gagccaaggc caaatagatc aatctcaaat cagccttgctc 11340
gggtggaatcg accaacttct ttctcaaatg gtagatggta ggggtattttt cgtggtaagc 11400
aacttcgtca acgatgttac cgaagattgg atgtctttcg tgcttcttgt cttcttcaac 11460
caagaaagat tcttccaatc tgtggaagaa agagtcgtca accttgcca tttcgttaga 11520
aaagatttct tgcaagtaac aaatacggtt tttacgtcta gtgtatctac gacgagcggc 11580
acgcttcaat ctggtagctt cagcggtttc accggagtcg aacaacaaag caccgatcaa 11640
attcttcttg atagagtgtc tgtcagtggt acccaagacc ttgaatttct tggatggaac 11700
cttgatttcg tcggtgatga cagccaacc gacggagttg gtcccgatat ccaaaccaat 11760
agagtatttc ttgtccattt ttgataagta tttaaagcag tgactgaaga ataatattct 11820
atgaggtttt aagctaaaaa tgaatatagt aaaattatta taataatagt gtagaagaag 11880
agaaaagaat atattaaaat ctgaatggta gctgctgtat atatactctt ttttttgcct 11940
cctccttggg taagtttctc tatttcagta gataaaaaaa aaacaataaa gcagtaatat 12000
tttttatcat atctcctcat aacaggaaaa atcagaaact aaggtttcta aatttgtcat 12060
ttttttggcc acgccccagt ggccgcagta caaatcgca gatttgtttg tttttgtca 12120
tgaaaagaaaa aatgaaaaa taagaaacta aacaaaaaa aaaaaaacgc caacagtgag 12180
aacgggcttt acaacggact tttacagtgg gttctaaaag aggtaaaaac tagcccgtag 12240
ccagcagact ggtattttga ttacctgtct tagcgttgac tggttgctta ttctactaga 12300
ccgcgtggta gcagacaata atgggagaat acgactcgtt tttgcccctt tcgtagtcaa 12360
tcctctcagt tcctcttctc cgaaagtaga aacgcagca ctttctcatt ggcgagtatt 12420
ttggtttttg ttttttgtt tcggtttcca cgtataatga gtggagtttt cggtttgttg 12480
aaccgtgttt gctttgctag tagaaactta ccaattttg aaaaaaattt gacaggctaa 12540
ggctagctta cctgaggcta gctagggtaa gcaagatttt gctaggattt ccacctatgg 12600
cggatttctc gttctactcc gtagtgctta catttctgag tcgttaacag cgcttgcctt 12660
caaatctaaa aaatctagaa aaagagttcc tagactaatt tggataaata gtaggattcg 12720

ttgctctacc tactcctctg cctccccgcc acatgggggt gaccgcaaaa aaagaaaagg 12780
tac 12783
<210> 5
<211> 12777
<212> DNA
<213> 人工序列
<220>
<223> nucleotide sequence of vector pDB1372
<400> 5
tcgcgcgttt cggatgatgac ggtgaaaacc tctgacacat gcagctcccg gagacgggtca 60
cagcttgtct gtaagcggat gccgggagca gacaagcccc tcagggcgcg tcagcgggtg 120
ttggcgggtg tcggggctgg cttactatg cggcatcaga gcagattgta ctgagagtgc 180
accataaacg acattactat atatataata taggaagcat ttaatagaca gcatcgtaat 240
atatgtgtac tttgcagtta tgacgccaga tggcagtagt ggaagatatt ctttattgaa 300
aaatagcttg tcaccttacg tacaatcttg atccggagct tttctttttt tgccgattaa 360
gaattaattc ggtcgaaaaa agaaaaggag agggccaaga gggaggcat tggtgactat 420
tgagcacgtg agtatacgtg attaagcaca caaaggcagc ttggagtatg tctgttatta 480
atctcacagg tagttctggt ccattggtga aagtttgcgg cttgcagagc acagaggccg 540
cagaatgtgc tctagattcc gatgctgact tgctgggtat tatatgtgtg cccaatagaa 600
agagaacaat tgacctggtt attgcaagga aaatttcaag tcttgtaaaa gcatataaaa 660
atagttcagg cactccgaaa tacttggttg gcgtgtttcg taatcaacct aaggaggatg 720
ttttggctct ggtcaatgat tacggcattg atatcgtcca actgcatgga gatgagtcgt 780
ggcaagaata ccaagagttc ctcggtttgc cagttattaa aagactcgta tttccaaaag 840
actgcaacat actactcagt gcagcttcac agaaacctca ttcgtttatt cccttgtttg 900
attcagaagc aggtgggaca ggtgaacttt tggattggaa ctcgatttct gactgggttg 960
gaaggcaaga gagccccgaa agcttacatt ttatgttagc tgggtgactg acgccagaaa 1020
atgttggtga tgcgcttaga ttaaattggc ttattggtgt tgatgtaagc ggaggtgtgg 1080
agacaaatgg tgtaaaagac tctaacaaaa tagcaaatc cgtcaaaaat gctaagaaat 1140
aggttattac tgagtagtat ttatttaagt attgtttgtg cacttgccca tgcggtgtga 1200
aataccgcac agatgcgtaa ggagaaaata ccgcatcagg aaattgtaaa cgtaaatatt 1260
ttgttaaaaat tcgctttaa tttttgttaa atcagctcat tttttaacca ataggccgaa 1320
atcggcaaaa tcccttataa atcaaaaagaa tagaccgaga tagggttgag tgttgttcca 1380
gtttggaaca agagtccact attaaagaac gtggactcca acgtcaaagg gcgaaaaacc 1440
gtctatcagg gcgatggccc actacgtgaa ccatcacctt aatcaagttt tttggggtcg 1500
aggtgccgta aagcactaaa tcggaacctt aaaggagacc cccgatttag agcttgacgg 1560
ggaaagccgg catcctgatg cgtttgtctg cacagatggc actcgtttc ttcccttctt 1620
ttctcggcac gttcggcaact agatgcactc atatcattat gcacggcact tgctcagtg 1680
gtctataacc tgttccattt acccataaag cccatgatta tccacatttt aatatctata 1740
tctcattcag cggccccaaa tattgtataa ctgctcttaa tacatgcgtt ataccacttt 1800

tacaccatat actaaccact caatgtatta ctcttatggt aatattacag aaaattcaac 1860
 accaaaaatca cctgagcata aaaataatta tctttcaact ttacaaaata aacatattga 1920
 cttgtgggca ctatcccctc aactctattc catttacagt tacataaaaa gttaacctaa 1980
 cacagaagta aaaatattta tgtcacaaaa atcagggtct cttgctatga tttgctactt 2040
 gttttgccct gatctgcaat cttattttta aaagtgcgc atattctata cgaccccgag 2100
 gcgacgcgcc aaaaaatgaa aaaaaggaac gacacatttt tatagaaaga caaaatgctg 2160
 caaagccgca catttccaag ttcagtatta cttatggaca caccctatta gctttattac 2220
 cgcccccccc ttttccagaa cagtataca aaaggggaagt ttctgttttg ggtttattgt 2280
 attcagaaaa tgctccacaa aactgttat tgatccgtcg aaattccttt tatgtgatga 2340
 ttgattgatt gattgtacag tttgtttttc ttaatatcta tttcgatgac ttctatatga 2400
 tattgcacta acaagaagat attataatgc aattgataca agacaaggag ttatttgctt 2460
 ctcttttata tgattctgac aatccatatt gcgttggtag tcttttttgc tggaacggtt 2520
 cagcggaaaa gacgcatcgc tctttttgct tctagaagaa atgccagcaa aagaatctct 2580
 tgacagtgac tgacagcaaa aatgtctttt tctaactagt aacaaggcta agatatcagc 2640
 ctgaaataaa ggggtggtgaa gtaataatta aatcatccgt ataaacctat acacatatat 2700
 gaggaanaat aatacaaaag tgttttaaat acagatacat acatgaacat atgcacgtat 2760
 agcgcceaaa tgtcggtaat gggatcggct tactaattat aaaatgcatc atagaaatcg 2820
 ttgaagtttg ccgtagtaat acccagatta tcagattcca aatccttgtc aataattata 2880
 ctcttttga caacttctct ttccattaaa aaatctgaaa tctccttagg taccttttct 2940
 ttttttgcgg tcacccccat gtggcgggga ggcagaggag taggtagagc aacgaatcct 3000
 actatttatc caaattagtc taggaactct ttttctagat tttttagatt tgagggcaag 3060
 cgctgttaac gactcagaaa tgtaagcact acggagtaga acgagaaatc cgccataggt 3120
 ggaaatccta gcaaaatcct gcttacccta gctagcctca ggtaagctag ccttagcctg 3180
 tcaaattttt ttcaaaattt ggtaagtffc tactagcaaa gcaaacacgg ttcaacaac 3240
 cgaaaactcc actcattata cgtggaaacc gaacaaaaa acaaaaacc aaaatactcg 3300
 ccaatgagaa agttgctgcg tttctacttt cgaggaagag gaactgagag gattgactac 3360
 gaaaggggca aaaacgagtc gtattctccc attattgtct gctaccacgc ggtctagtag 3420
 aataagcaac cagtcaacgc taagacaggt aatcaaaata ccagtctgct ggctacgggc 3480
 tagtttttac ctcttttaga acccactgta aaagtccggt gtaaagcccg ttctactgt 3540
 tggcgttttt ttttttttgg ttttagttct tatttttcat ttttttctt catgaccaa 3600
 aacaaacaaa tctcgcgatt tgtactgcgg ccaactgggc gtggccaaaa aaatgacaaa 3660
 tttagaaacc ttagtttctg atttttctg ttatgaggag atatgataaa aaatattact 3720
 gctttattgt tttttttta tctactgaaa tagagaaact tacccaagga ggaggcaaaa 3780
 aaaagagtat atatacagca gctaccatc agattttaat atattctttt ctcttcttct 3840
 acactattat tataataatt ttactatatt catttttagc ttaaacctc atagaatatt 3900
 attcttcagt cactcgctta aatacttatc aaaaatggac aagaataact ctattggttt 3960
 ggatatcggg accaactccg tcggttgggc tgtcatcacc gacgaataca aggttccatc 4020
 caagaaatc aaggtcttgg gtaacactga cagacactct atcaagaaga atttgatcgg 4080
 tgctttgttg ttcgactccg gtgaaaccgc tgaagctacc agattgaagc gtaccgctcg 4140

tcgtagatac actagacgta aaaaccgtat ttgttacttg caagaaatct tttctaacga 4200
aatggccaag gttgacgact ctttcttcca cagattggaa gaatctttct tggttgaaga 4260
agacaagaag cacgaaagac atccaatctt cggtaacatc gttgacgaag ttgcttacca 4320
cgaaaaatac cctaccatct accatttgag aaagaagttg gtcgattcca ccgacaaggc 4380
tgatttgaga ttgatctatt tggccttggc tcacatgac aagttcagag gtcacttctt 4440
gattgaaggt gacttgaacc cagacaactc tgacgtcgac aaattgttca tccaattggt 4500
ccaaacctac aaccaattat tcgaggaaaa cccaattaac gcttctggtg ttgatgctaa 4560
ggccatctta tctgcccgtt tgtccaagtc tagacgtttg gaaaacttga ttgctcaatt 4620
gcctggtgaa aagaaaaacg gtttgttcgg taacttgac gctttgtcct tgggtttgac 4680
cccaaacttc aagtccaact tcgacttggc tgaagatgcc aagttgcaat tgtccaagga 4740
cacctacgac gacgacttag acaacttggt ggctcaaate ggtgaccaat acgccgactt 4800
gttcttggtt gccaaaaact tatctgacgc tatcttgttg tctgacatct tgagagttaa 4860
cactgaaatt accaaggttc cattgtctgc ttctatgac aaaagatacg acgaacacca 4920
ccaagatctg actttgttga aggttttggg tagacaacaa ttgccagaaa agtacaagga 4980
aatcttcttc gaccaatcca aaaatggtta cgccggttac attgacggtg gtgcttctca 5040
ggaagaattc tacaagttca tcaagccaat tttggaaaag atggatggta ctgaagaatt 5100
attggttaag ttgaacagag aagacttatt gagaaagcaa cgtaccttcg ataacggttc 5160
tatcccacac caaatccact tgggtgaatt gcacgccatt ttgagaagac aggaagattt 5220
ctatccattc ctaaaggaca acagagaaaa gatcgaaaag atcttaactt tcagaatccc 5280
atactacgtc ggtccattgg ccagaggtaa ttctagattc gcttggatga ccagaaagtc 5340
tgaagaaacc atcaccccat ggaacttcca agaagtcgtc gacaagggtg cttctgcccc 5400
atctttcatc gaaagaatga ccaacttga taagaacttg ccaaacgaga aggtcttgcc 5460
aaagcactct ttgtgttac aatacttcac cgtctacaac gaattaacca aggttaaata 5520
cgttactgaa ggtatgagaa agccagcttt cctatccggt gaacaaaaga aggtattgt 5580
tgacttgttg ttaagacca acagaaaggt cactgttaag caattgaagg aagactactt 5640
caagaagatt gaatgtttcg attccgtcga aatctccggt gttgaagacc gtttcaatgc 5700
ttctttgggc acctaccagc atttgttaaa gatcatcaag gacaaggact ttttagataa 5760
cgaagaaaac gaagacatct tggaagatat cgttttgacc ttgactcttt tcgaggacag 5820
agaaatgatt gaagagagat tgaagaccta cgctcacttg ttcgacgata aagttatgaa 5880
gcaactaaag agaagaagat aacttggttg gggtagattg tccagaaagt tgattaacgg 5940
tatcagagac aagcaatccg gtaagactat ttagacttt ttgaaatccg atggtttcgc 6000
taacagaaac tttatgcaat tgattcacga cgattctttg actttcaagg aagacattca 6060
aaaagcccaa gtctctggtc aaggtgattc tttgcacgaa cacatcgcta acttggtg 6120
ttctccagct attaagaagg gtatcttaca aaccgtcaag gtcgttgatg aattggtcaa 6180
agtcatgggt agacacaagc cagaaaatat tgtcatcgaa atggctagag aaaaccaaac 6240
tactcaaaag ggtcaaaaga actctagaga acgtatgaag agaattgaag aaggtatcaa 6300
ggagttgggt tctcaaattt tgaaagaaca cccagtcgaa aacctcaat taaaaacga 6360
aaagctatac ttgtactact tgcaaacggc tcgtgacatg tacgtcgacc aagaattgga 6420
tatcaacaga ttgtctgact acgatgtcga tcatatcgtc ccacaatcgt tcttgaagga 6480

cgattccatt gacaacaaag ttttgactag atctgacaag aacagaggta agtctgataa 6540
cgttccatct gaagaagttg ttaagaagat gaagaactac tggagacaat tgttgaatgc 6600
taagttgatc actcaaagaa agttcgacaa cttgaccaag gctgaaagag gtggtttgtc 6660
cgaattggac aaagccggtt tcatcaagag acaattagtc gaaactagac aatcaccaa 6720
gcatgttgct caaatcttgg attccagaat gaacactaag tacgatgaaa acgacaaaact 6780
aattagagaa gttaaggtca tcactttgaa gtctaagttg gtttctgact tcagaaagga 6840
cttccaattt tacaaggtca gagaaatcaa caactacat cacgctcacg atgcctactt 6900
gaacgctggt gtcggtactg ccttaatcaa aaagtaccca aagttggaat ctgaattcgt 6960
ttacggtgac tacaaggtt acgatgtag aaagatgac gccaaagtctg aacaagaaat 7020
tggtaaaggc actgctaagt acttcttcta ctctaacatc atgaactttt tcaagactga 7080
aatcacttta gctaacggtg aaattagaaa gcgtccattg attgaaacca atggtgaaac 7140
tggtgaaatt gtctgggaca agggtagaga tttcgctacc gtcagaaagg ttttgtctat 7200
gccacaagtt aacatcgtca agaagactga agttcaaact ggtggtttct ctaaggaatc 7260
cattttgcca aagagaaaact ctgacaagtt gattgctaga aagaaggact gggatcctaa 7320
gaagtacggt ggtttcgact ctccaactgt tgcttactcc gttttggtcg ttgctaaggt 7380
tgaaaagggt aagtctaaga agttgaagtc tgtaaggaa ttggtgggta tcaccatcat 7440
ggaaagatcc tccttcgaaa agaaccatcg acttttttg gaagctaagg gttacaagga 7500
agtcaagaag gatttgatca ttaagttacc aaaatactcc ttgttcgaat tggaaaacgg 7560
tagaaagaga atgttggcct ccgctggtga actacaaaa ggtaacgaat tggctttacc 7620
atctaagtac gttaacttct tgtacttggc ttcccactac gaaaagttga aaggttcccc 7680
agaagacaac gaacaaaagc aattgtttgt tgaacaacac aagcactact tggatgaaat 7740
tattgaacaa atctccgaat tctccaagag agtcattttg gctgatgcta acttagataa 7800
ggttttatcc gcttacaaca agcacagaga caaccaatc agagaacaag ctgaaaacat 7860
cattcatttg ttcactttaa ccaacttggg tgctccagct gctttcaat acttcgacac 7920
taccattgac agaaagagat acaactccac caagaagtt ttagatgcta ctttgattca 7980
ccaatctatt accggtttgt acgaaaccag aattgacttg tctcaattgg gtggtgattc 8040
cagagctgat ccaaagaaga agagaaaggt gtaaaggagt taaaggcaa gttttctttt 8100
ctagagccgt tcccacaaat aattatacgt atatgcttct tttcgtttac tatatatcta 8160
tatttacaag cttttattca ctgatgcaat ttgtttcaa atactttttt ggagatctca 8220
taactagata tcatgatgac gcaacttggc gctatcttaa ttactctggc tgccaggccc 8280
gtgtagaggc ccgcaagacc ttctgtacgc catatagtct ctaagaactt gaacaagttt 8340
ctagacctat tgccgcttt cggatcgtca ttgttgccgc cgccagctga agcttcgtac 8400
gctgcaggtc gacgaattct accgttcgta taatgtatgc tatacgaagt tatagatctg 8460
tttagcttgc ctctgccccg ccgggtcacc cggccagcga catggaggcc cagaataccc 8520
tccttgacag tcttgacgtg cgcagctcag gggcatgatg tgactgtcgc ccgtacattt 8580
agcccataca tccccatgta taatcatttg catccataca ttttgatggc cgcacggcgc 8640
gaagcaaaaa ttacggetcc tcgctgcaga cctgcgagca gggaaacgct ccctcacag 8700
acgcgttgaa ttgtccccac gccgcgcccc tgtagagaaa tataaaaggt taggatttgc 8760
cactgagggt cttctttcat atacttcctt ttaaaatctt gctaggatac agttctcaca 8820

tcacatccga acataaacia ccatgggtaa ggaaaagact cacgtttcga ggccgcgatt 8880
aaattccaac atggatgctg atttatatgg gtataaatgg gctcgcgata atgtcgggca 8940
atcagggtcg acaatctatc gattgtatgg gaagcccgat gcgccagagt tgtttctgaa 9000
acatggcaaa ggtagcgtt ccaatgatgt tacagatgag atggtcagac taaactggct 9060
gacggaattt atgcctcttc cgaccatcaa gcattttatc cgtactcctg atgatgcatg 9120
gttactcacc actgcgatcc ccggcaaac agcattccag gtattagaag aatcctctga 9180
ttcagggtgaa aatattgttg atgcgctggc agtgttcctg cgcgggttg attcgattcc 9240
tgtttgtaat tgcctttta acagcgatcg cgtatttcgt ctcgctcagg cgcaatcacg 9300
aatgaataac ggtttggttg atgcgagtga tttgatgac gagcgtaatg gctggcctgt 9360
tgaacaagtc tggaaagaaa tgcataagct tttgccattc tcaccggatt cagtcgtcac 9420
tcatggatgat ttctcacttg ataaccttat tttgacgag gggaaattaa taggttgat 9480
tgatgttga cgagtcggaa tgcgagaccg ataccaggat cttgccatcc tatggaactg 9540
cctcggtagg ttttctcctt cattacagaa acggcttttt caaaaatatg gtattgataa 9600
tctgatatg aataaattgc agtttcattt gatgctcgat gagtttttct aatcagtact 9660
gacaataaaa agattcttgt tttcaagaac ttgtcatttg tatagttttt ttatattgta 9720
gttgttctat tttaatcaaa tgtagcgtg atttatattt tttttcgcct cgacatcatc 9780
tgcccagatg cgaagttaag tgcgagaaa gtaatatcat gcgtcaatcg tatgtgaatg 9840
ctggtcgcta tactgctgct gattcgatac taacgccgcc atccagtgtc gaaaacgagc 9900
tcataacttc gtataatgta tgctatacga acggtagaat tcgaatcaga tccactagtg 9960
gcctatgcgg ccgccaccgc ggtggagctc cagcttttgt tccctttagt gagggttaat 10020
tgccgccttg gcgtaatcat ggtcatagct gtttctctgt tgaattgtt atccgctcac 10080
aattccacac aacataggag ccggaagcat aaagtgtaaa gcctgggggtg cctaatgagt 10140
gaggtaactc acattaattg cgttgcgctc actgcccgct ttccagtcgg gaaacctgtc 10200
gtgccagctg cattaatgaa tcggccaacg cgcggggaga ggccggttg gtattgggag 10260
ctcttccgct tctcgcctca ctgactcgtc gcgctcggtc gttcggctgc ggcgagcgg 10320
atcagctcac tcaaaggcgg taatacggtt atccacagaa tcaggggata acgcaggaaa 10380
gaacatgtga gcaaaaaggc agcaaaaaggc caggaaccgt aaaaaggccg cgttgctggc 10440
gtttttccat aggctccgcc cccctgacga gcatcaciaa aatcgacgct caagtccagag 10500
gtggcgaaac ccgacaggac tataaagata ccaggcgttt ccccctggaa gctccctcgt 10560
gcgctctcct gtccgaccc tgcgcttac cggatacctg tccgcctttc tcccttcggg 10620
aagcgtggcg ctttctcata gctcacgctg taggtatctc agttcgggtg aggtcgttccg 10680
ctccaagctg ggctgtgtgc acgaaccccc cgttcagccc gaccgctgcg ctttatccgg 10740
taactatcgt cttgagtcca acccggttaag acacgactta tcgccactgg cagcagccac 10800
tggtaacagg attagcagag cgaggtatgt aggcgggtgct acagagttct tgaagtgggtg 10860
gcctaactac ggctacacta gaaggacagt atttggtatc tgcgctctgc tgaagccagt 10920
taccttcgga aaaagagttg gtagctcttg atccggcaaa caaacaccg ctggtagcgg 10980
tggttttttt gtttgcaagc agcagattac gcgcagaaaa aaaggatctc aagaagatcc 11040
tttgatcttt tctacgggtg ctgacgctca gtggaacgaa aactcacgtt aagggatttt 11100
ggatcatgaga ttatcaaaaa ggatcttcac ctagatcctt ttaaattaa aatgaagttt 11160

taaatcaatc taaagtatat atgagtaaac ttggtctgac agttaccaat gcttaatcag 11220
 tgaggcacct atctcagcga tctgtctatt tcgttcatcc atagttgcct gactccccgt 11280
 cgtgtagata actacgatac gggagggtct accatctggc cccagtgtg caatgatacc 11340
 gcgagaccca cgctcaccgg ctccagattt atcagcaata aaccagccag ccggaagggc 11400
 cgagcgcaga agtggctcctg caactttatc cgctccatc cagtctatta attggtgccg 11460
 ggaagctaga gtaagtagtt cgccagttaa tagtttgcgc aacgttggtg ccattgctac 11520
 aggcatcgtg gtgtcacgct cgtcgtttgg tatggcttca ttcagctccg gttcccaacg 11580
 atcaaggcga gttacatgat ccccatggtt gtgcaaaaaa gcggttagct ccttcgggtc 11640
 tccgatcgtt gtcagaagta agttggccgc agtgttatca ctcatggtta tggcagcact 11700
 gcataattct cttactgtca tgccatccgt aagatgcttt tctgtgactg gtgagtactc 11760
 aaccaagtca ttctgagaat agtgtatgcg gcgaccgagt tgctcttgcc cggcgtcaat 11820
 acgggataat accgcgccac atagcagaac tttaaaagtg ctcatcattg gaaaacgttc 11880
 ttcggggcga aaactctcaa ggatcttacc gctgttgaga tccagttcga tgtaaccac 11940
 tcgtgcacc cactgatctt cagcatcttt tactttcacc agcgtttctg ggtgagcaaa 12000
 aacaggaagg caaaatgccg caaaaaaggg aataaggcg acacggaat gttgaatact 12060
 catactcttc ctttttcaat attattgaag catttatcag ggttattgtc tcatgagcgg 12120
 atacatattt gaatgtattt agaaaaataa acaaataggg gttccgcgca catttccccg 12180
 aaaagtgcc cctgggtcct tttcatcacg tgctataaaa ataattataa tttaaatttt 12240
 ttaatataaa tatataaatt aaaaaataga agtaaaaaaa gaaattaaag aaaaaatagt 12300
 ttttgttttc cgaagatgta aaagactcta ggggatcgc caacaatac taccttttat 12360
 cttgctcttc ctgctctcag gtattaatgc cgaattgttt catcttgtct gtgtagaaga 12420
 ccacacacga aaatcctgtg attttacatt ttacttatcg ttaatcgaat gtatatctat 12480
 ttaatctgct tttctgtct aataaatata tatgtaaagt acgctttttg ttgaaatttt 12540
 ttaaaccctt gttattttt tttcttcat tccgtaacte ttctacctc tttatttact 12600
 ttctaaaatc caaatacaaa acataaaaat aaataaacac agagtaaatt cccaaattat 12660
 tccatcatta aaagatacga ggcgcgtgta agttacaggc aagcgatccg tcctaagaaa 12720
 ccattattat catgacatta acctataaaa ataggcgtat cagaggccc tttcgtc 12777

<210> 6

<211> 5042

<212> DNA

<213> 人工序列

<220>

<223> nucleotide sequence of vector pDB1368

<400> 6

tagaaaaact catcgagcat caaatgaaac tgcaatttat tcatatcagg attatcaata 60
 ccatattttt gaaaaagccg tttctgtaat gaaggagaaa actcaccgag gcagttccat 120
 aggatggcaa gatcctggta tcggtctgcg attccgactc gtccaacatc aatacaacct 180
 attaatttcc cctcgtcaaa aataaggtta tcaagtgaga aatcacatg agtgacgact 240
 gaatccggtg agaatggcaa aagtttatgc atttctttcc agacttggtc aacaggccag 300

ccattacgct cgtcatcaaa atcactcgca tcaaccaaac cgttattcat tcgtgattgc 360
 gcctgagcga ggcgaaatac gcgatcgctg ttaaaaggac aattacaaac aggaatcgag 420
 tgcaaccggc gcaggaacac tgccagcgca tcaacaatat tttcacctga atcaggatat 480
 tcttctaata cctggaacgc tgtttttccg gggatcgcag tggtagtaa ccatgcatca 540
 tcaggagtac ggataaaatg cttgatggtc ggaagtggca taaattccgt cagccagttt 600
 agtctgacca tctcatctgt aacatcattg gcaacgctac ctttgccatg tttcagaaac 660
 aactctggcg catcgggctt cccatacaag cgatagattg tcgcacctga ttgcccgaca 720
 ttatcgcgag cccatttata cccatataaa tcagcatcca tgttggaatt taatcgcggc 780
 ctcgacgttt cccgttgaat atggctcata ttcttcttt ttcaatatta ttgaagcatt 840
 tatcagggtt attgtctcat gagcggatac atatttgaat gtatttagaa aaataaaca 900
 atagggtca gtgttacaac caattaacca attctgaaca ttatcgcgag cccatttata 960
 cctgaatatg gctcataaca ccccttgttt gcttggcggc agtagcgcgg tggteccacc 1020
 tgaccccatg ccgaaactcag aagtgaaacg ccgtagcgcg gatggtagtg tggggactcc 1080
 ccatgcgaga gtagggaact gccaggcatc aaataaacg aaaggctcag tcgaaagact 1140
 gggcctttcg cccgggctaa ttaggggtg tcgcccttat tcgactctat agtgaagttc 1200
 ctattctcta gaaagtatag gaacttctga agtggggaac gttgtccagg tttgtatcca 1260
 cgtgtgtccg ttccgccaat attccgcgtg cgttttattt ctgctgcat ccgtaaagtc 1320
 caggatttga gcgggttaca caatatact catattttcg gtgtctgggt cactacttta 1380
 ctcttggcat ccactaaata tattggatcc tgctttttaa actggcttcc agaaaaaat 1440
 caatggagtg atgcaaactg cctggagtaa aagatgacac aaggcgattg acctacgcat 1500
 gtatctatct cattttctta caccttctat ttatttctaa ctctttgatt tggaaaacac 1560
 ctaagaaaaa aaaggtgaa atcagttccc tgaaattgtc cccctacttg actaataaat 1620
 atataaagac ggtaggtatt gactgtaatt cgtaaacta tacttcttaa acttcttcaa 1680
 atttactttt ttggatagtc ttatttttgg tttcaatacc ccaagaactt agtttcaaat 1740
 aaatacacat acaaacaaaa tgtcaggtgt taataataca tccgcaatg agttatctac 1800
 tacatgtct aactctaact cagcagtagg cgctccctct gttagactg aacacggtga 1860
 ctctaaaaat tcccttaacc tagatgcaa tgagccacct attgacttac ctcaaaaacc 1920
 cctcgccgca tattggactg ttatctgttt atgtctaatt attgcattg gtgggtttgt 1980
 ctttggttgg gatactggtg ccatctctgg tttgttaat caaacgatt tcaaaagaag 2040
 atttgggtcaa atgaaatctg atggtacctt ttatctttcg gacgtccgga ctggtttgat 2100
 cgttgggtatc ttcaatattg gttgtgctt tgggtgggtta accttaggac gtctgggtga 2160
 tatgtatgga cgtagaattg gtttcatgtg cgtcgttctg gtatacatcg ttggtattgt 2220
 gattcaaat gcttctagt acaaatggta ccaatattc attggtagaa ttatctctgg 2280
 tatgggtgtc ggtggtattg ctgtctatc tccaacttg atttccgaa cagcaccaaa 2340
 acacattaga ggtacctgt tttctttcta tcagttaatg atcactctag gtattttctt 2400
 aggttactgt accaactatg gtactaaaga ctactccaat tcagttcaat ggagagtgcc 2460
 tttgggtttg aactttgct tcgctatfff catgatcgct ggtatgctaa tggttccaga 2520
 atctccaaga ttcttagtcg aaaaaggcag atacgaagac gctaaacgtt ctttggcaaa 2580
 atctaacaaa gtcaccattg aagatccaag tattgttgcg gaaatggata caattatggc 2640

caacgttgaa actgaaagat tagccggtaa cgcttcttgg ggtgagttat tctccaacaa 2700
agggtgctatt ttacctcgtg tgattatggg tattatgatt caatccttac aacaattaac 2760
tggtacaacat tacttcttct attatggtac tactatnttc aacgccgtcg gtatgaaaga 2820
ttctttccaa acttccatcg ttttaggtat agtcacgttc gcatccactt tcgtggcctt 2880
atacactggt gataaatttg gtcgtcgtaa gtgtctattg ggtggttctg cttccatggc 2940
catttgtttt gttatcttct ctactgtcgg tgcacaagc ttatatccaa atggtaaaga 3000
tcaacctatct tccaaggctg ccggtaacgt catgattgtc tttacctgtt tattcatttt 3060
cttcttcgct attagttggg ccccaattgc ctacgttatt gttgccgaat cctatccttt 3120
gcgtgtcaaa aatcgtgcta tggctattgc tgttggtgcc aactggattt ggggtttctt 3180
gattggtttc ttactccct tcattacaag tgcaattgga tttcatacag ggtatgtctt 3240
catgggctgt ttggtatntt cattcttcta cgtgtnttct tttgtctgtg aaaccaaggg 3300
cttaacatta gaggaagtta atgaaatgta tgttgaagggt gtcaaacat ggaaatctgg 3360
tagctggatc tcaaaaagaaa aaagagtttc cgaggaataa atttgatctg tagcctaagt 3420
ataaaattct acgtatgtat atatttacet gcaatnttct cttnttccaa ttcatgcctc 3480
agaaagcctg tatgcgaagc cacaatcctt tccaacagac catactaagt aaaatgaagt 3540
gaagttccta tactttctag agaataggaa cttctatagt gagtccaata agggcgacac 3600
aaaatttatt ctaaatgcat aataaatact gataacatct tatagtttgt attatatttt 3660
gtattatcgt tgacatgtat aattttgata tcaaaaactg atnttccctt tattatnttc 3720
gagatttatt ttcttaattc tctttaacaa actagaaata ttgtatatac aaaaaatcat 3780
aaataataga tgaatagttt aattataggt gttcatcaat cgaaaaagca acgtatctta 3840
tttaaagtgc gttgctnttct tctcatttat aaggttaaat aattctcata tatcaagcaa 3900
agtgacaggc gccctaaat attctgacaa atgctcttct cctaaactcc ccccataaaa 3960
aaacccgccg aagcgggttt ttacgttatt tgcggattaa cgattactcg ttatcagaac 4020
cgcccagggg gcccgagctt aagactggcc gtcgtnttct aacacagaaa gagttttag 4080
aaacgcaaaa aggccatccg tcaggggctt tctgcttagt ttgatgcctg gcagttccct 4140
actctcgcct tccgcttctt cgtcactga ctcgctgcgc tcggtcgttc ggctgcggcg 4200
agcggtatca gctcactcaa aggcggtaat acggttatcc acagaatcag gggataacgc 4260
aggaaagaac atgtgagcaa aaggccagca aaaggccagg aaccgtaaaa aggccgcgtt 4320
gctggcgttt ttccatagcc tccgcccccc tgacgagcat cacaaaaatc gacgctcaag 4380
tcagaggtag cgaaaccgca caggactata aagataccag gcgtnttccc ctggaagctc 4440
cctcgtgcgc tctcctgttc cgacctgct gcttaccgga tacctgtccg cttntctccc 4500
ttcggaagc gtggcgcttt ctcatagctc acgctgtagg tatctcagtt cgggttaggt 4560
cgttcgtctc aagctgggct gtgtgcacga accccccgtt cagccccacc gctgcgcctt 4620
atccggtaac tatcgtcttg agtccaacct ggtaagacac gacttatcgc cactggcagc 4680
agccactggt aacaggatta gcagagcgag gtatgtaggc ggtgctacag agttcttgaa 4740
gtgggtgggct aactacggct aactagaag aacagtatntt ggtatctgcg ctctgctgaa 4800
gccagttacc ttcggaaaaa gagttgtag ctcttgatcc ggcaaaaaa ccaccgctgg 4860
tagcggtagt tttnttgttt gcaagcagca gattacgcgc agaaaaaaag gatctcaaga 4920
agatcctttg atctnttcta cggggtctga cgtcagtagg aacgacgcgc gcgtaactca 4980

cgtaaagggga ttttgggtcat gagcttgccg cgtcccgtca agtcagcgtata atgctctgct 5040
 tt 5042
 <210> 7
 <211> 35
 <212> DNA
 <213> 人工序列
 <220>
 <223> nucleotide sequence of Forward primer for extension PCR to add
 KpnI restriction site to cFS0017
 <400> 7
 aaccgcggggg tacccatttg aataagaagt aatac 35
 <210> 8
 <211> 33
 <212> DNA
 <213> 人工序列
 <220>
 <223> nucleotide sequence of Reverse primer for extension PCR to add
 NgoMIV restriction site to cFS0017
 <400> 8
 ttccgcggggc cggcatcctg atgcgtttgt ctg 33
 <210> 9
 <211> 516
 <212> DNA
 <213> 人工序列
 <220>
 <223> nucleotide sequence of the INT70 gRNA gBLOCK
 <400> 9
 gctatacgaa cggtagaatt cgatatacaga tccactagtgc gcctatgcgg ccgccaccgc 60
 ggtcttttgaa aagataatgt atgattatgc tttcactcat atttatacag aaacttgatg 120
 ttttcttttcg agtatataca aggtgattac atgtacgttt gaagtacaac tctagatttt 180
 gtagtgcctt cttgggctag cggtaaaggt gcgcattttt tcacacccta caatgttctg 240
 ttcaaaaagat tttgggtcaaa cgctgtagaa gtgaaagttg gtgcgcatgt ttcggcgttc 300
 gaaacttctc cgcagtgaaa gataaatgat cggagagaaa ggcccggggc tgtttttagag 360
 ctgaaaatag caagttaaaa taaggctagt ccgttatcaa cttgaaaaag tggcaccgag 420
 tcggtgggtgc tttttttggt ttttatgtct ccgcggtgga gctccagctt ttgttccctt 480
 tagtgagggt taattgcgcg cttggcgtaa tcatgg 516
 <210> 10
 <211> 82
 <212> DNA

<213> 人工序列

<220>

<223> nucleotide sequence of the forward primer to obtain donor DNA PCR fragment (int70[5']-conD-HXT11/2-con3-int70[3']) using pDB1368 as template

<400> 10

gaccggtcta agctcttaga ggttctcgca tacccaagta aaagctaaga ccgaagcaaa 60
aacgttgtcc aggtttgtat cc 82

<210> 11

<211> 84

<212> DNA

<213> 人工序列

<220>

<223> nucleotide sequence of the reverse primer to obtain donor DNA PCR fragment (int70[5']-conD-HXT11/2-con3-int70[3']) using pDB1368 as template

<400> 11

tttgtttcctt tattgttttt atttttacga cattttcccc tcgaagaata tttatccgaa 60
acttagtatg gtctgttga aagg 84

<210> 12

<211> 70

<212> DNA

<213> 人工序列

<220>

<223> nucleotide sequence of the forward primer to obtain a gRNA-recipient plasmid backbone using pRN1120-RFP-gRNA(A) (SEQ ID NO: 1) as template

<400> 12

acgctttccg gcatcttcca gaccacagta tatccatccg cctcctgttg aggaccggtt 60
tatcattatc 70

<210> 13

<211> 24

<212> DNA

<213> 人工序列

<220>

<223> set out the nucleotide sequence of the reverse primer to obtain a gRNA-recipient plasmid backbone using pRN1120-RFP-gRNA(A) (SEQ ID NO: 1) as template

<400> 13

agcggccgca tgctagctcc ggat 24
<210> 14
<211> 80
<212> DNA
<213> 人工序列
<220>
<223> nucleotide sequence of the forward primer to obtain a guide RNA
PCR fragment (gRNA-INT70) using INT70 gBLOCK (SEQ ID NO: 9) as
template
<400> 14
tcatgtttga cagcttatca tcgataatcc ggagctagca tgcggccgct gttccgcggt 60
ctttgaaaag ataatgtatg 80
<210> 15
<211> 74
<212> DNA
<213> 人工序列
<220>
<223> nucleotide sequence of the reverse primer to obtain a guide RNA
PCR fragment (gRNA-INT70) using INT70 gBLOCK (SEQ ID NO: 9) as
template
<400> 15
caacaggagg cggatggata tactgtggtc tggaagatgc cggaaagcgc catttgatgg 60
agttccgcgg agac 74
<210> 16
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> nucleotide sequence of the forward primer to confirm to confirm
the correct assembly and integration of the HXT11/2 expression
cassettes at the INT70 locus
<400> 16
gtctgcatag gagccttctg 20
<210> 17
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> nucleotide sequence of the reverse primer to confirm the correct

assembly and integration of the HXT11/2 expression cassettes at the INT70 locus

<400> 17

aatttaccac tgcccatggg 20

<210> 18

<211> 11742

<212> DNA

<213> 人工序列

<220>

<223> nucleotide sequence of vector pCSN061

<400> 18

```

tcgcgcgttt cggatgatgac ggtgaaaacc tctgacacat gcagctcccg gagacgggtca 60
cagcttgtct gtaagcggat gccgggagca gacaagcccc tcagggcgcg tcagcgggtg 120
ttggcgggtg tcggggctgg cttactatg cggcatcaga gcagattgta ctgagagtgc 180
accataaacg acattactat atatataata taggaagcat ttaatagaca gcatcgtaat 240
atatgtgtac tttgcagtta tgacgccaga tggcagtagt ggaagatatt ctttattgaa 300
aaatagcttg tcaccttacg tacaatcttg atccggagct tttctttttt tgccgattaa 360
gaattaattc ggtcgaaaaa agaaaaggag agggccaaga gggaggcat tggtgactat 420
tgagcacgtg agtatacgtg attaagcaca caaaggcagc ttggagtatg tctgttatta 480
atttcacagg tagttctggt ccattggtga aagtttgcgg cttgcagagc acagaggccg 540
cagaatgtgc tctagattcc gatgctgact tgctgggtat tatatgtgtg cccaatagaa 600
agagaacaat tgaccgggtt attgcaagga aaatttcaag tcttgtaaaa gcatataaaa 660
atagttcagg cactccgaaa tacttggttg gcgtgtttcg taatcaacct aaggaggatg 720
ttttggctct ggtcaatgat tacggcattg atatcgtcca actgcatgga gatgagtcgt 780
ggcaagaata ccaagagttc ctcggtttgc cagttattaa aagactcgta tttccaaaag 840
actgcaacat actactcagt gcagcttcac agaaacctca ttcgtttatt cccttgtttg 900
atcagaagc aggtgggaca ggtgaacttt tggattggaa ctcgatttct gactgggttg 960
gaaggcaaga gagccccgaa agcttacatt ttatgttagc tgggtgactg acgccagaaa 1020
atgttgggtga tgcgcttaga ttaaattggcg ttattggtgt tgatgtaagc ggaggtgtgg 1080
agacaaatgg tgtaaaagac tctaacaaaa tagcaaatc cgtcaaaaat gctaagaaat 1140
aggttattac tgagtagtat ttatttaagt attgtttgtg cacttgcccta tgcggtgtga 1200
aataccgcac agatgcgtaa ggagaaaata ccgcatcagg aaattgtaaa cgттаатatt 1260
ttgttaaaaat tcgcgttaaa tttttgttaa atcagctcat tttttaacca ataggccgaa 1320
atcggcaaaa tcccttataa atcaaaaagaa tagaccgaga tagggttgag tgttgttcca 1380
gtttggaaca agagtccact attaaagaac gtggactcca acgtcaaagg gcgaaaaacc 1440
gtctatcagg gcgatggccc actacgtgaa ccataccctt aatcaagttt tttggggtcg 1500
aggtgccgta aagcactaaa tcggaacctt aaaggagacc cccgatttag agcttgacgg 1560
ggaaagccgg cgaacgtggc gagaaaggaa gggaagaaag cgaaaggagc gggcgctagg 1620
gcgctggcaa gtgtagcggc cacgctgcgc gtaaccacca caccgcccgc gcttaatgcg 1680

```

ccgctacagg gcgcgtcgcg ccattcgcca ttcaggctgc gcaactgttg ggaagggcga 1740
tcgggtgcggg cctcttcgct attacgccag ctggcgaaag ggggatgtgc tgcaaggcga 1800
ttaagttggg taacgccagg gttttcccag tcacgacggt gtaaacgac ggccagttag 1860
cgcgcgtaaat acgactcact atagggcgaa ttgggtacct tttctttttt tgcggtcacc 1920
cccatgtggc ggggaggcag aggagtaggt agagcaacga atcctactat ttatccaaat 1980
tagtctagga actctttttc tagatTTTTT agatttgagg gcaagcgctg ttaacgactc 2040
agaaaatgtaa gcactacgga gtagaacgag aaatccgcca taggtggaaa tcctagcaaa 2100
atcttgctta ccctagctag cctcaggtaa gctagcctta gcctgtcaaa tttttttcaa 2160
aatttggtaa gtttctacta gcaaagcaaa cacggttcaa caaacgaaa actccactca 2220
ttatacgtgg aaaccgaaac aaaaaaaca aaacccaaat actcgccaat gagaaagttg 2280
ctgcgtttct actttcgagg aagaggaact gagaggattg actacgaaag gggcaaaaac 2340
gagtcgtatt ctcccattat tgtctgctac cacgcggtct agtagaataa gcaaccagtc 2400
aacgctaaga caggtaatca aaataccagt ctgctggcta cgggctagtt tttacctctt 2460
ttagaaccca ctgtaaaaagt ccgttgtaaa gcccgttctc actgttggcg tttttttttt 2520
tttggtttag tttcttattt ttcatttttt tctttcatga ccaaaaaca acaaatctcg 2580
cgatttgtag tgcggccact ggggcgtggc caaaaaaatg acaaattag aaaccttagt 2640
ttctgatttt tcctgttatg aggagatatg ataaaaata ttactgcttt attgtttttt 2700
ttttatctac tgaaatagag aaacttacc aaggaggagg caaaaaaag agtatatata 2760
cagcagctac cattcagatt ttaatatatt ctttctctt cttctacact attattataa 2820
taattttact atattcattt ttagcttaaa acctcataga atattattct tcagtcactc 2880
gcttaaatac ttatcaaaaa tggacaagaa atactctatt ggtttggata tcgggacca 2940
ctccgctcgg tgggctgtca tcaccgacga atacaaggtt ccatccaaga aattcaaggt 3000
cttgggtaac actgacagac actctatcaa gaagaattg atcgggtgctt tgttgttcga 3060
ctccggtgaa accgctgaag ctaccagatt gaagcgtacc gctcgtcgta gatacactag 3120
acgtaaaaac cgtatttggt acttgcaaga aatctttct aacgaaatg ccaaggttga 3180
cgactctttc ttccacagat tggaagaatc tttcttggtt gaagaagaca agaagcacga 3240
aagacatcca atcttcggtg acatcgttga cgaagttgct taccacgaaa aataccctac 3300
catctaccat ttgagaaaaga agttggtcga ttccaccgac aaggctgatt tgagattgat 3360
ctatttggcc ttggctcaca tgatcaagtt cagaggtcac ttcttgattg aaggtgactt 3420
gaaccagac aactctgacg tcgacaaatt gttcatcaa ttggtccaaa cctacaacca 3480
attattcgag gaaaacccaa ttaacgctt tggtgttgat gctaaggcca tcttatctgc 3540
ccgtttgtcc aagtctagac gtttggaaaa ctgattgct caattgctg gtgaaaagaa 3600
aaacggtttg ttcggttaact tgatcgctt gtcttgggt ttgacccca acttcaagtc 3660
caacttcgac ttggctgaag atgccaagtt gcaattgtcc aaggacacct acgacgacga 3720
cttagacaac ttgttggtc aaatcggtga ccaatacgc gacttgttct tggctgcca 3780
aaacttatct gacgtatct tgttgctga catcttgaga gtttaactg aaattacca 3840
ggctccattg tctgcttcta tgatcaaaag atacgacgaa caccaccaag atctgacttt 3900
gttgaaggct ttggttagac aacaattgcc agaaaagtac aaggaaatct tcttcgacca 3960
atccaaaaat ggttacgcc gttacattga cgggtggtct tctcaggaag aattctaca 4020

gttcatcaag ccaatthttgg aaaagatgga tggactgaa gaattattgg ttaagttgaa 4080
cagagaagac ttattgagaa agcaacgtac cttegataac ggttctatcc cacaccaa 4140
ccacttgggt gaattgcacg ccattttgag aagacaggaa gatttctatc cattcctaaa 4200
ggacaacaga gaaaagatcg aaaagatctt aactttcaga atcccatact acgtcgggcc 4260
attggccaga ggtaattcta gattcgcttg gatgaccaga aagtctgaag aaaccatcac 4320
cccatggaac ttcgaagaag tcgtcgacaa ggggtgcttct gcccaatctt tcatcgaaag 4380
aatgaccaac tttgataaga acttgccaaa cgagaaggct ttgccaaagc actctttggt 4440
gtacgaatac ttcaccgtct acaacgaatt aaccaagggt aaatacgtta ctgaaggat 4500
gagaaagcca gctttcctat ccggtgaaca aaagaaggct attggtgact tgttgtttaa 4560
gaccaacaga aaggctactg ttaagcaatt gaaggaagac tacttcaaga agattgaatg 4620
tttcgattcc gtcgaaatct ccggtgttga agaccgtttc aatgcttctt tgggcaccta 4680
ccacgatttg ttaaagatca tcaaggacaa ggacttttta gataacgaag aaaacgaaga 4740
catcttgga gatatcgtht tgacctgac tcttttcgag gacagagaaa tgattgaaga 4800
gagattgaag acctacgctc acttgttcga cgataaagtt atgaagcaac taaagagaag 4860
aagatacact ggttggggta gattgtccag aaagttgatt aacggtatca gagacaagca 4920
atccggtaag actatthtag actthttgaa atccgatggt ttcgctaaca gaaacttht 4980
gcaattgatt cacgacgatt ctttgactth caaggaagac attcaaaaag cccaagtctc 5040
tggtaagggt gattctthgc acgaacacat cgctaacttg gctggttctc cagctattaa 5100
gaagggtatc ttacaaaccg tcaaggtcgt tgatgaattg gtcaaagtca tgggtagaca 5160
caagccagaa aatattgtca tcgaaatggc tagagaaaac caactactc aaaagggtca 5220
aaagaactct agagaacgta tgaagagaat tgaagaagg atcaaggagt tgggttctca 5280
aattthgaaa gaacaccag tcgaaaacac tcaattacaa aacgaaaagc tatacttgta 5340
ctacttgcaa aacggtcgtg acatgtactg cgaccaagaa ttggatatca acagattgtc 5400
tgactacgat gtcgatcata tcgtcccaca atcgttcttg aaggacgatt ccattgacaa 5460
caaagthttg actagatctg acaagaacag aggtaagtct gataacgttc catctgaaga 5520
agttgttaag aagatgaaga actactggag acaattgttg aatgctaagt tgatcactca 5580
aagaaagttc gacaactga ccaaggctga aagaggtggt ttgtccgaat tggacaaagc 5640
cgtttcatc aagagacaat tagtcgaaac tagacaaatc accaagcatg ttgctcaaat 5700
cttgattcc agaatgaaca ctaagtacga tgaaaacgac aaactaatta gagaagttaa 5760
ggtcatcact ttgaagtcta agttggtttc tgacttcaga aaggacttcc aattthtaca 5820
ggtcagagaa atcaacaact accatcacgc tcacgatgcc tacttgaacg ctgttgctcg 5880
tactgcctta atcaaaaagt acccaaagtt ggaatctgaa ttcgthtacg gtgactacaa 5940
ggtthtacgat gttagaaaga tgatcgccaa gtctgaacaa gaaattggta aggccactgc 6000
taagtacttc thtacteta acatcatgaa cththtcaag actgaaatca cthtagctaa 6060
cgtgaaatt agaaagcgtc cattgattga aaccaatggt gaaactggtg aaattgtctg 6120
ggacaagggt agagatttcg ctaccgtcag aaaggtthttg tctatgccac aagttaacat 6180
cgtcaagaag actgaagttc aaactggtgg thtctetaag gaatccattt tgccaaagag 6240
aaactctgac aagttgattg ctagaaagaa ggactgggat cctaagaagt acggtggtth 6300
cgactctca actgttgctt actccgthtt ggtcgttgct aaggttgaaa agggttaagtc 6360

taagaagttg aagtctgtta aggaattggt gggatcacc atcatggaaa gatcctcctt 6420
 cgaaaagaac ccaatcgact ttttgaagc taagggttac aaggaagtca agaaggattt 6480
 gatcattaag ttacccaaat actccttggt cgaattggaa aacggtagaa agagaatggt 6540
 ggccctccgct ggtgaactac aaaaaggtaa cgaattggct ttaccatcta agtacgttaa 6600
 cttcttgtag ttggcttccc actacgaaaa gttgaaaggt tccccagaag acaacgaaca 6660
 aaagcaattg tttgttgaac aacacaagca ctacttggat gaaattattg aacaaatctc 6720
 cgaattctcc aagagagtca ttttggctga tgctaactta gataaggttt tatccgctta 6780
 caacaagcac agagacaaac caatcagaga acaagctgaa aacatcattc atttgttcac 6840
 tttaaccaac ttgggtgctc cagctgcttt caaatacttc gacactacca ttgacagaaa 6900
 gagatacact tccaccaaag aagttttaga tgctactttg attaccaat ctattaccgg 6960
 tttgtacgaa accagaattg acttgtctca attgggtggt gattccagag ctgatccaaa 7020
 gaagaagaga aaggtgtaaa ggagttaaag gcaaagtttt cttttctaga gccgttccca 7080
 caaataatta tacgtatatg cttcttttctg tttactatat atctatattt acaagccttt 7140
 attcactgat gcaatttgtt tccaaatact tttttggaga tctcataact agatatcatg 7200
 atggcgcaac ttggcgctat ctttaattact ctggctgcca ggcccgtgta gagggccgca 7260
 agaccttctg tacgccatat agtctctaag aacttgaaca agtttctaga cctattgccg 7320
 cttttcggat cgctattggt gcggccgcca gctgaagctt cgtacgctgc aggtcgacga 7380
 attctaccgt tcgtataatg tatgtatac gaagttatag atctgtttag cttgcctcgt 7440
 ccccgccggg tcaccggcc agcgacatgg aggccagaa taccctcctt gacagtcttg 7500
 acgtgcgcag ctcaggggca tgatgtgact gtcgccgta catttagccc atacatcccc 7560
 atgtataatc atttgcattc atacattttg atggccgca ggcgcaagc aaaaattacg 7620
 gctcctcgct gcagacctgc gagcaggaa acgctccct cacagacgcg ttgaattgtc 7680
 cccacgccgc gccctgtag agaaataaa aaggttagga tttgccactg aggttcttct 7740
 ttcatatact tccttttaa atcttgctag gatacagttc tcacatcaca tccgaacata 7800
 aacaaccatg ggtaaggaaa agactcacgt ttcgaggccg cgattaaatt ccaacatgga 7860
 tgctgattta tatgggtata aatgggctcg cgataatgtc gggcaatcag gtgcgacaat 7920
 ctatcgattg tatgggaagc ccgatgcgcc agagttgttt ctgaaacatg gcaaaggtag 7980
 cgttgccaat gatgttacag atgagatggt cagactaac tggctgacgg aatttatgcc 8040
 tcttccgacc atcaagcatt ttatccgtac tctgatgat gcatggttac tcaccactgc 8100
 gatccccggc aaaacagcat tccaggtatt agaagaatat cctgattcag gtgaaaatat 8160
 tgttgatgcg ctggcagtgt tctgcgccg gttgcattcg attcctgttt gtaattgtcc 8220
 ttttaacagc gatcgcgtat ttcgtctcgc tcaggcgca tcacgaatga ataacggttt 8280
 ggttgatgcg agtgattttg atgacgagcg taatggctgg cctgttgaac aagtctggaa 8340
 agaaatgcat aagcttttgc cattctcacc ggattcagtc gtcactcatg gtgatttctc 8400
 acttgataac cttatttttg acgaggggaa attaataggt tgtattgatg ttggacgagt 8460
 cggaatcgca gaccgatacc aggatcttgc cactctatgg aactgcctcg gtgagtttct 8520
 tccttcatta cagaaacggc tttttcaaaa atatggtatt gataatctg atatgaataa 8580
 attgcagttt catttgatgc tcgatgagtt tttctaata gtactgacaa taaaagatt 8640
 cttgttttca agaactgtc atttgtatag ttttttata ttgtagttgt tctattttaa 8700

tcaaatgtta gcgtgattta tatttttttt cgctcgaca tcactgccc agatgcgaag 8760
ttaagtgcgc agaaagtaat atcatgcgtc aatcgatgt gaatgctggt cgctatactg 8820
ctgtcgattc gatactaacg ccgcatcca gtgtcgaaaa cgagctcata acttcgtata 8880
atgtatgcta tacgaacggt agaattcgaa tcagatccac tagtggccta tgcggccgcc 8940
accgcggtgg agctccagct tttgttcct ttagtgagg ttaattgctg gcttggcgta 9000
atcatggtca tagctgtttc ctgtgtgaaa ttgttatccg ctcaaatc cacacaacat 9060
aggagccgga agcataaagt gtaaagcctg ggtgcctaa tgagtgaggt aactcacatt 9120
aattgcggtg cgctcactgc ccgctttcca gtcgggaaac ctgtcgtgcc agctgcatta 9180
atgaatcggc caacgcgcgg ggagaggcgg tttgctatt gggcgtctt ccgcttctc 9240
gctcactgac tcgctgcgct cggctgttcg gctgcggcga gcggtatcag ctactcaaa 9300
ggcgtaata cggttatcca cagaatcagg ggataacga ggaaagaaca tgtgagcaa 9360
aggccagcaa aaggccagga accgtaaaaa ggccgcgtt ctggcgttt tccataggct 9420
ccgccccct gacgagcctc acaaaaatcg acgctcaagt cagaggtggc gaaaccgcac 9480
aggactataa agataccagg cgtttcccc tggaagctc ctctgctgct ctctgttcc 9540
gacctgccg cttaccgat acctgtccg ctttctcct tcgggaagcg tggcgctttc 9600
tcatagctca cgctgtaggt atctcagttc ggtgtagtc gttcgtcca agctgggctg 9660
tgtgcacgaa cccccgttc agcccgacc ctgcgcctta tccgtaact atcgtcttga 9720
gtccaaccg gtaagacacg acttatgcc actggcagca gccactggtg acaggattag 9780
cagagcgagg tatgtagcg gtgctacaga gttcttgaag tggtagccta actacggcta 9840
cactagaagg acagtattg gtatctgctc tctgctgaag ccagttacct tcgaaaaag 9900
agttggtagc tcttgatccg gcaaaaaac caccgctggt agcgggtggt ttttgtttg 9960
caagcagcag attacgcga gaaaaaagg atctcaagaa gatccttga tctttctac 10020
ggggtctgac gctcagtga acgaaaactc acgttaaggg attttggta tgagattatc 10080
aaaaaggatc ttcacctaga tccttttaa ttaaaatga agttttaat caatctaaag 10140
tatatatgag taaacttgg ctgacagta ccaatgctta atcagtgagg cacctatctc 10200
agcgatctgt ctatttcgt catccatagt tgctgactc ccgctcgtg agataactac 10260
gatacgggag ggcttaccat ctggccccag tgctgcaatg ataccgcgag acccacgctc 10320
accgctcca gattatcag caataacca gccagccgga agggccgagc gcagaagtgg 10380
tcctgcaact ttatccgct ccatccagtc tattaattg tggcgggag ctagagtaag 10440
tagttcgcca gttaatagtt tgcgcaacgt gtttgcatt gctacaggca tcgtggtgct 10500
acgctcgtcg tttggtatgg cttcattcag ctccggttc caacgatcaa ggcgagttac 10560
atgatcccc atgttgtgca aaaaagcgg tagctcttc ggtctcctga tcgttgtcag 10620
aagtaagttg gccgcagtgt tatcactcat ggttatggca gactgcata attctcttac 10680
tgtcatgcca tccgtaagat gttttctgt gactggtgag tactcaacca agtcattctg 10740
agaatagtg atgcggcgac cgagttgctc ttgcccggcg tcaatacggg ataataccgc 10800
gccacatagc agaactttaa aagtgetcat cattggaaaa cgttcttcgg ggcgaaaact 10860
ctcaaggatc ttaccgctgt tgagatccag ttcgatgtaa cccactcgtg cacccaactg 10920
atcttcagca tcttttactt tcaccagcgt ttctgggtga gcaaaaacag gaaggcaaaa 10980
tgccgcaaaa aagggaataa gggcgacacg gaaatgttga atactcatac tcttctttt 11040

tcaatattat tgaagcattt atcagggtta ttgtctcatg agcggataca tatttgaatg 11100
 tatttagaaa aataaaciaa taggggttcc gcgcacattt ccccgaaaag tgccacctgg 11160
 gtccttttca tcacgtgcta taaaaataat tataatntaa attttttaat ataaatatat 11220
 aaattaaaaa tagaaagtaa aaaaagaaat taaagaaaaa atagtntttg ttttccgaag 11280
 atgtaaaaaga ctctaggggg atcgccaaca aatactacct tttatcttgc tcttcctgct 11340
 ctcaggtatt aatgccgaat tgtttcatct tgtctgtgta gaagaccaca cacgaaaatc 11400
 ctgtgatttt acattttact tatcgntaat cgaatgtata tctatntaat ctgctntttct 11460
 tgtctaataa atatatatgt aaagtacgct tttgttgaa atntntntaa cttttgttta 11520
 tntntntttc ttcatntcgt aactctteta cttctnttat ttactntteta aaatccaaat 11580
 acaaaacata aaaataaata aacacagagt aaatteccaa attatntcat cattaaaaga 11640
 tacgaggcgc gtgtaagnta caggcaagcg atcctgcteta agaaaccatt attatcatga 11700
 cattaaccta taaaaatagg cgtatcacga ggccctnttcg tc 11742

<210> 19

<211> 605

<212> DNA

<213> 人工序列

<220>

<223> nucleotide sequence of the pGAL10 promoter

<400> 19

catttgaata agaagtaata caaacgaaa atgttgaaag tattagttaa agtggttatg 60
 cagntnttgc atttatatat ctgntaatag atcaaaaatc atcgtntcgc tgattaatta 120
 ccccgaaaat aaggctaaaa aactaatcgc attatcatcc tatggntgnt aatttgattc 180
 gntcatntga aggnttgntg ggccagntta ctgccaatnt ttctctntca taaccataaa 240
 agctagtatt gtagaatctt tattgntcgg agcagtgcgg cgcgaggcac atctgcgntt 300
 caggaacgcg accggntgaag acgaggacgc acggaggaga gtctntcctc ggagggctgt 360
 cacccgctcg gcggctnteta atccgtactt caatatagca atgagcagnt aagcgtatta 420
 ctgaaagntc caaagagaag gntntntntag gctaagataa tggggctctt tacatnttcca 480
 caacatataa gtaagantag atatggatat gtatatggat atgtatatgg tgntaatgcc 540
 atgtaatatg attatntaac ttctntgcgt ccatccaaaa aaaaagntag aatntntntgaa 600
 aatntc 605

<210> 20

<211> 600

<212> DNA

<213> 人工序列

<220>

<223> nucleotide sequence of the pCUP1 promoter

<400> 20

ttatgtgatg attgattgat tgattgtaca gntntgntntt ctntaatatct atntcगतga 60
 cttctatatg atatntgact aacaagaaga tattataatg caatntgatac aagacaagga 120

gttatttgct tctcttttat atgattctga caatccatat tgcgttgta gtcttttttg 180
 ctggaacggg tcagcggaaa agacgcatcg ctctttttgc ttctagaaga aatgccagca 240
 aaagaatctc ttgacagtga ctgacagcaa aaatgtcttt ttctaactag taacaaggct 300
 aagatatcag cctgaaataa aggggtggta agtaataatt aatcatccg tataaaccta 360
 tacacatata tgaggaaaaa taatacaaaa gtgttttaaa tacagataca tacatgaaca 420
 tatgcacgta tagcgcccaa atgtcggtaa tgggatcggc ttactaatta taaaatgcat 480
 catagaaatc gttgaagttt gccgtagtaa taccagatt atcagattcc aaatccttgt 540
 caataattat actcctttgg acaacttctc tttcattaa aaaatctgaa atctccttag 600

<210> 21

<211> 758

<212> DNA

<213> 人工序列

<220>

<223> GIN11 (M86)

<400> 21

aaggaatttc gacggatcaa taacagtgtt tgtggagcat tttctgaata caataaaccc 60
 aaaacagaaa ctccctttt gtatcactgt tctggaaaag ggggtggcgg taataaagct 120
 aatagggtgt gtccataagt aatactgaac ttggaaatgt gcggctttgc agcattttgt 180
 ctttctataa aaatgtgtcg ttcctttttt tcattttttg gcgcgtcgcc tcggggtcgt 240
 atagaatatg cgtcactttt aaaaataaga ttgcagatca gggcaaaaca agtagcaaat 300
 catagcaaga gaccctgatt tttgtgacat aaatattttt acttctgtgt taggttaact 360
 ttttatgtaa ctgtaaagtg aatagagttg aggggatagt gccacaagt caatatgttt 420
 attttgtaaa gttgaaagat aattattttt atgctcaggt gattttggtg ttgaattttc 480
 tgtaatatta acataagagt aatacattga gtggttagta tatggtgtaa aagtggata 540
 acgcatgtat taagagcagt tatacaatat ttggggccgc tgaatgagat atagatatta 600
 aaatgtggat aatcatgggc tttatgggta aatggaacag ggtatagacc actgaggcaa 660
 gtgccgtgca taatgatatg agtgcattca gtggcgaac tggcgagaaa ggaaggaag 720
 aaagcgagtg ccatctgtgc agacaaacgc atcaggat 758

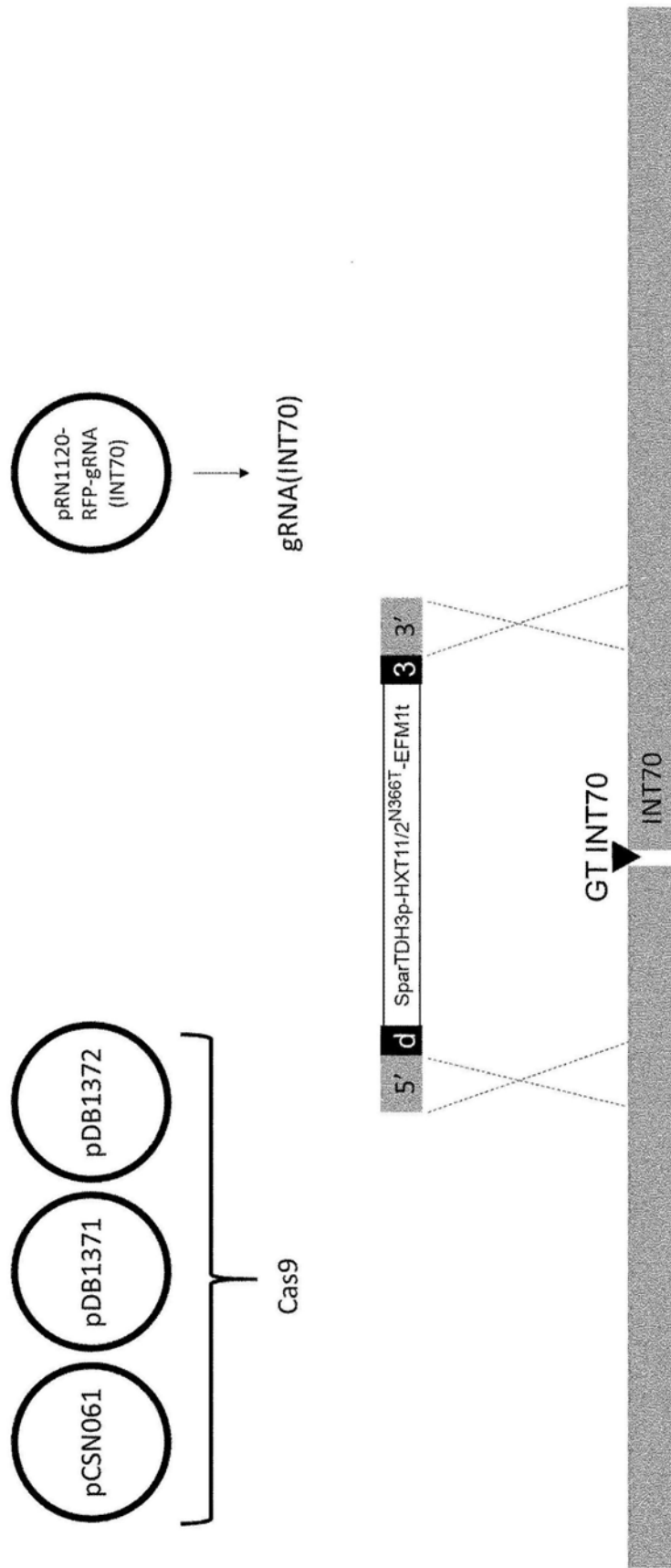


图1

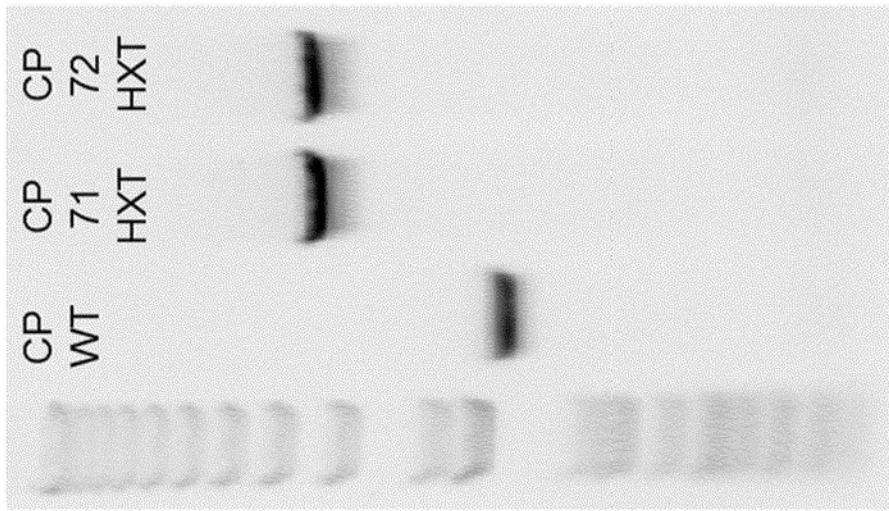


图2

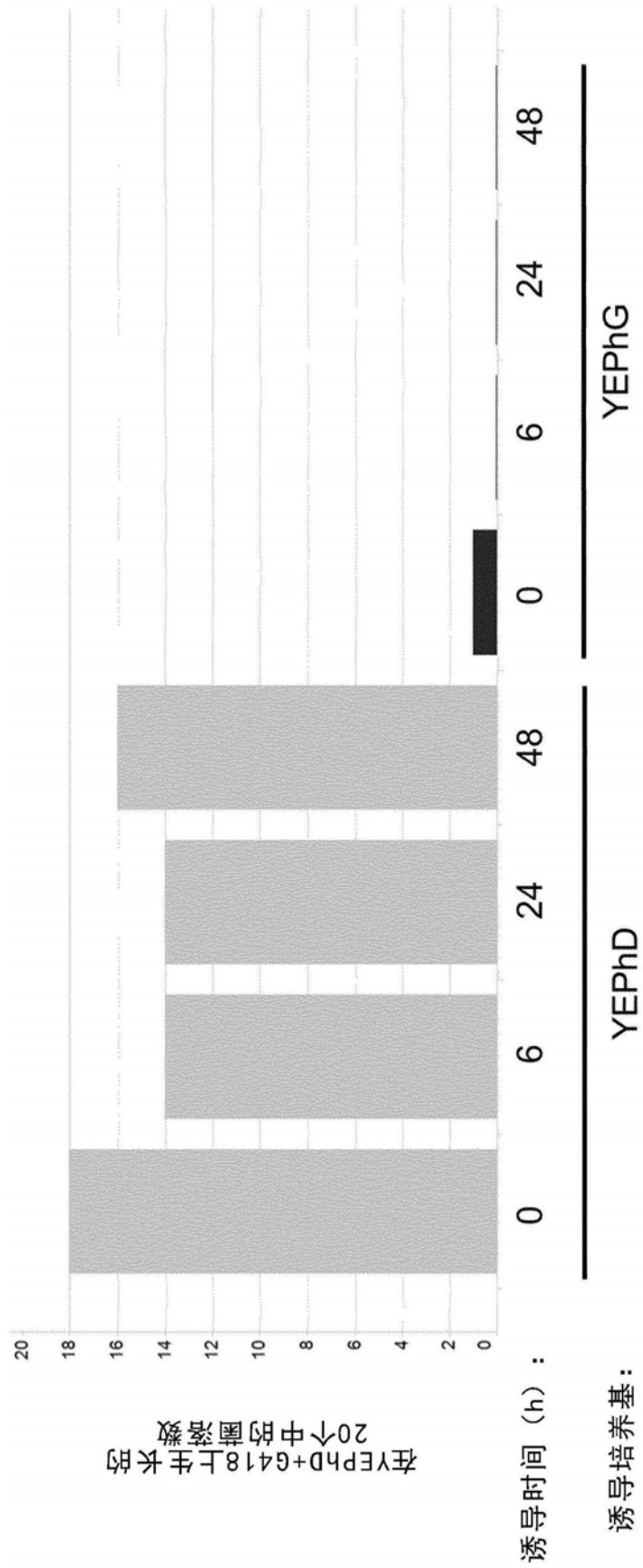


图3

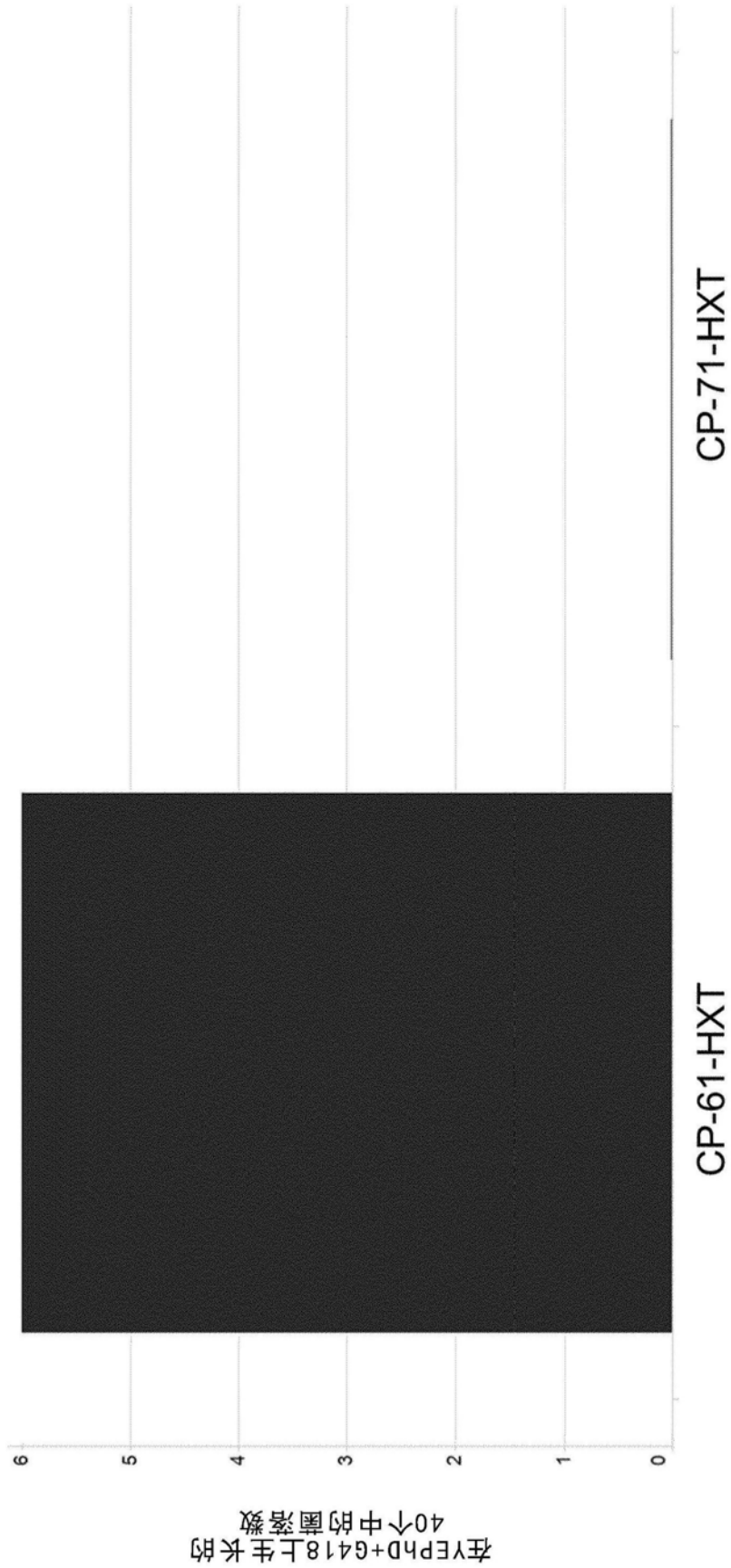


图4

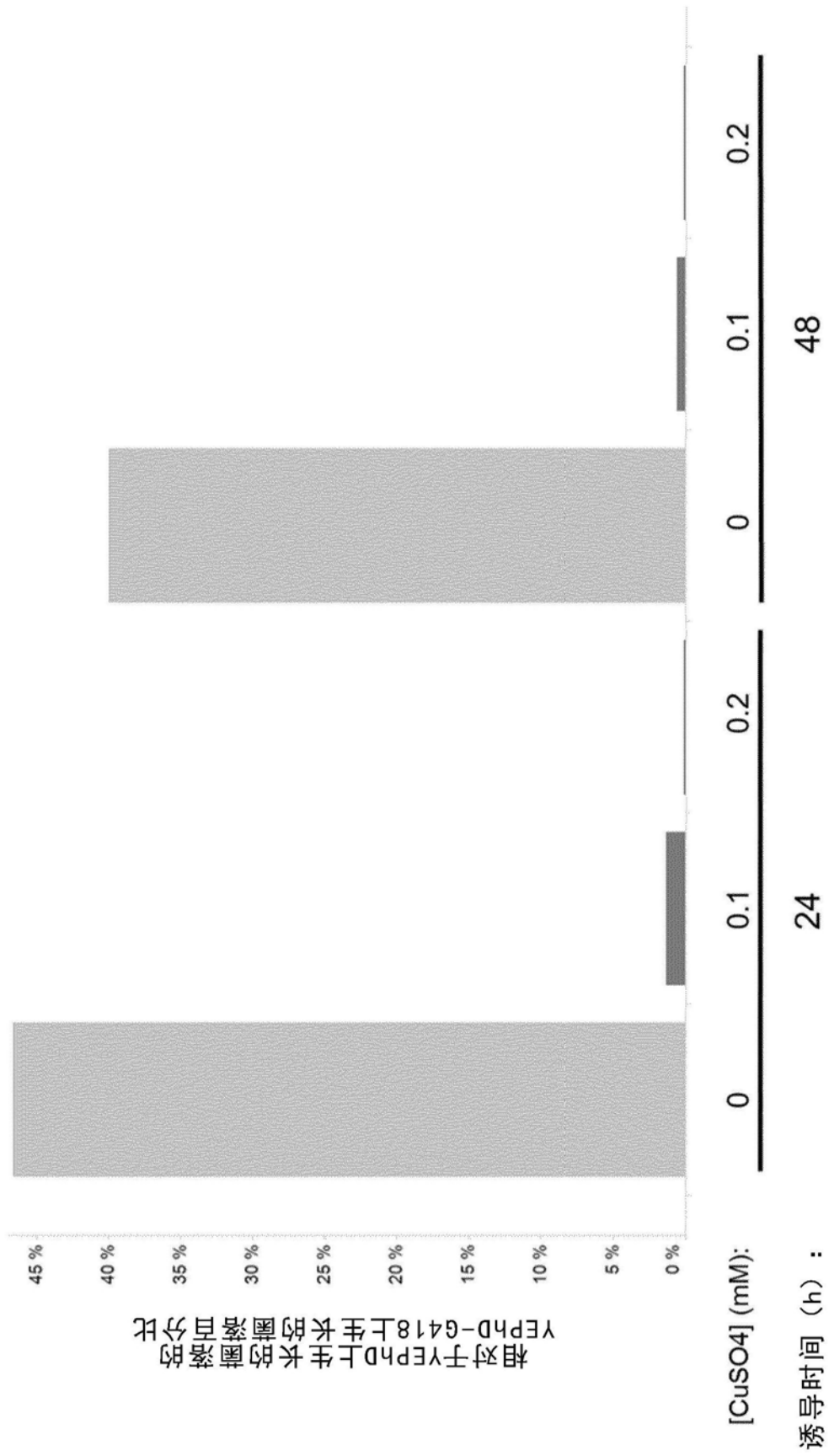


图5

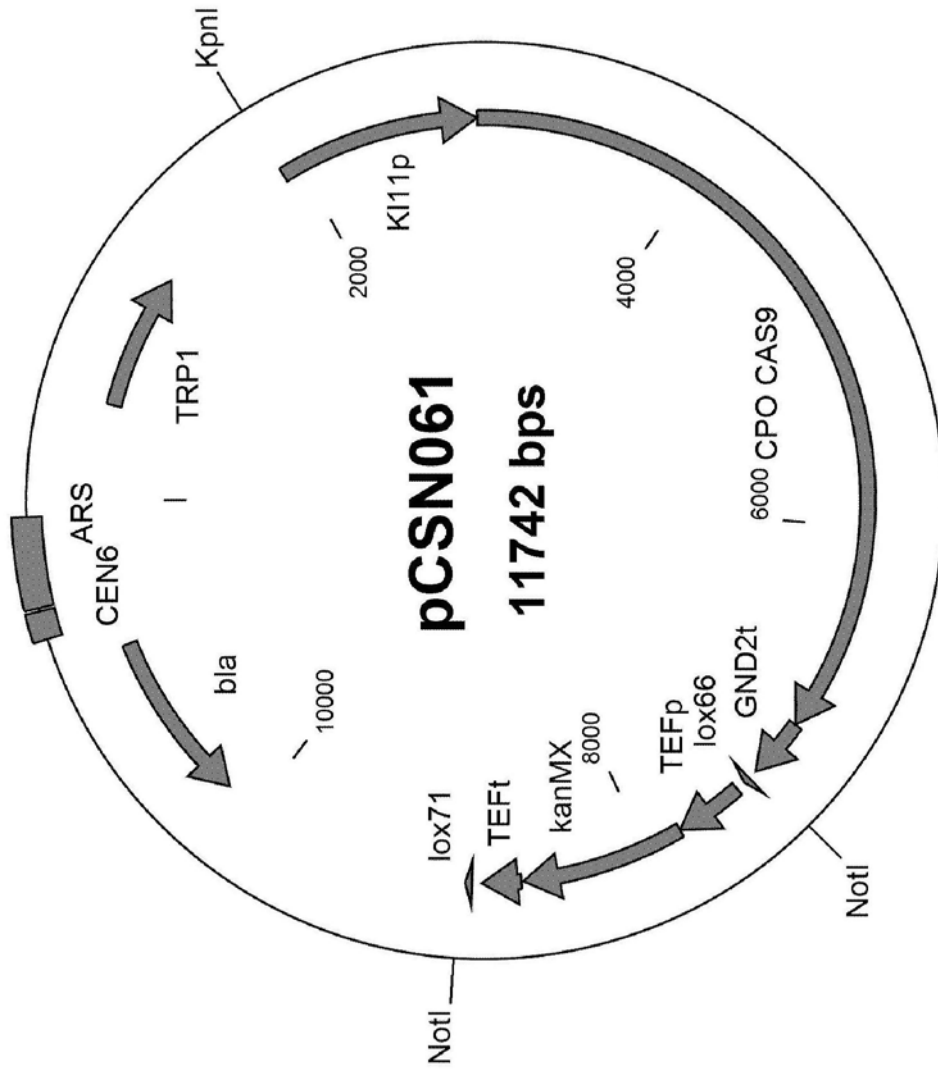


图6

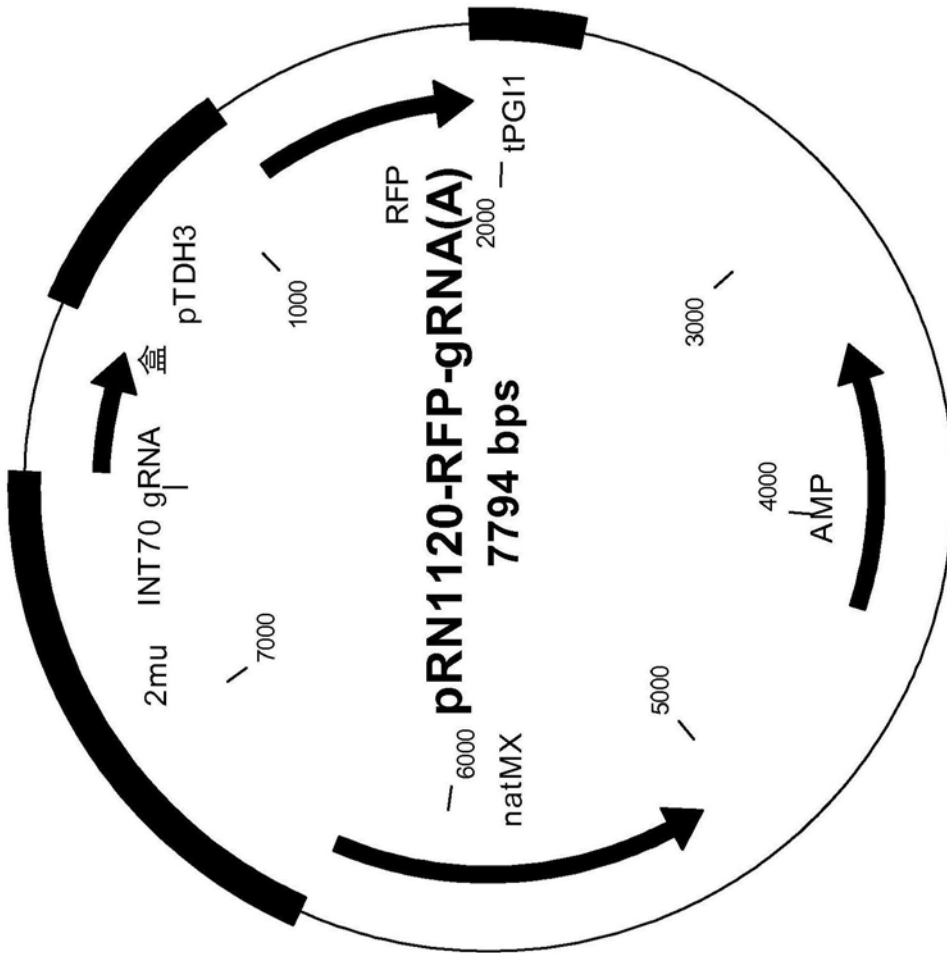


图7

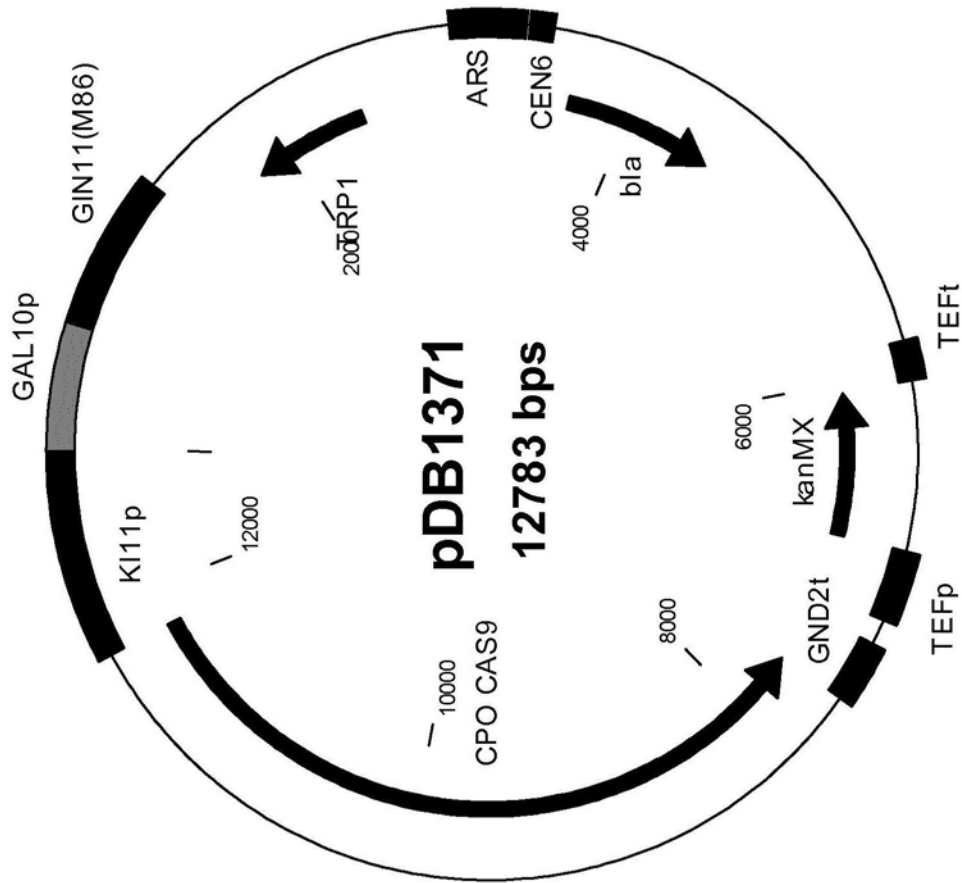


图8

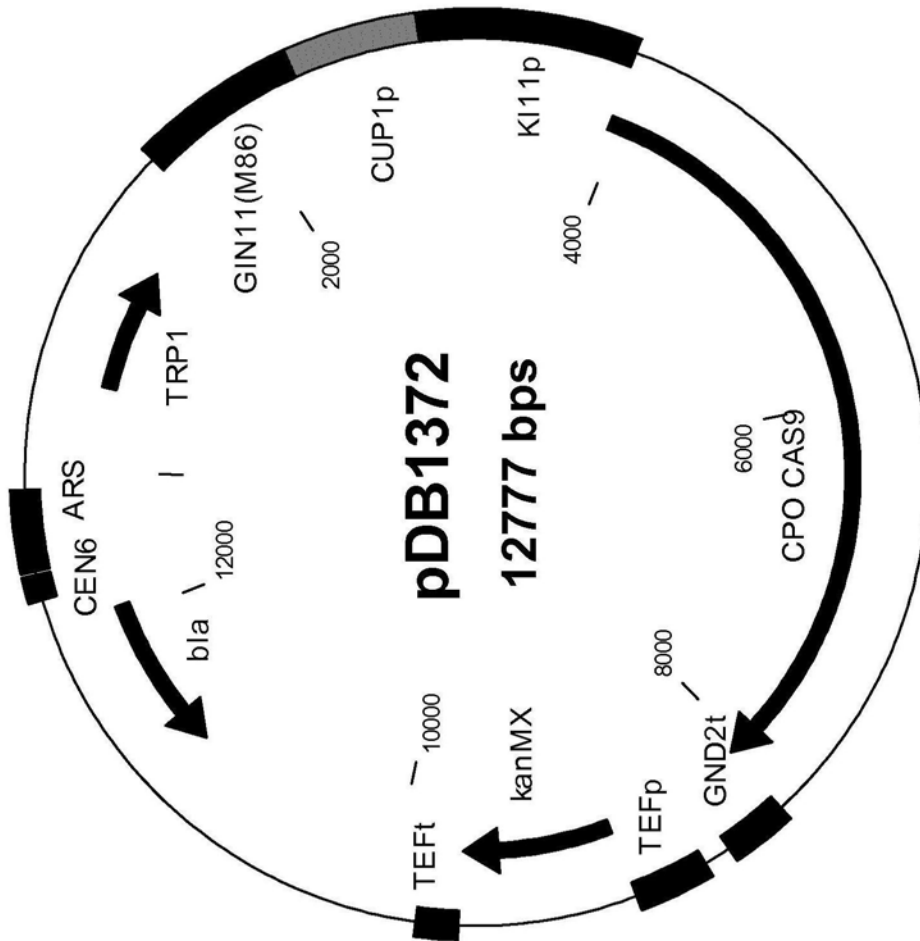


图9

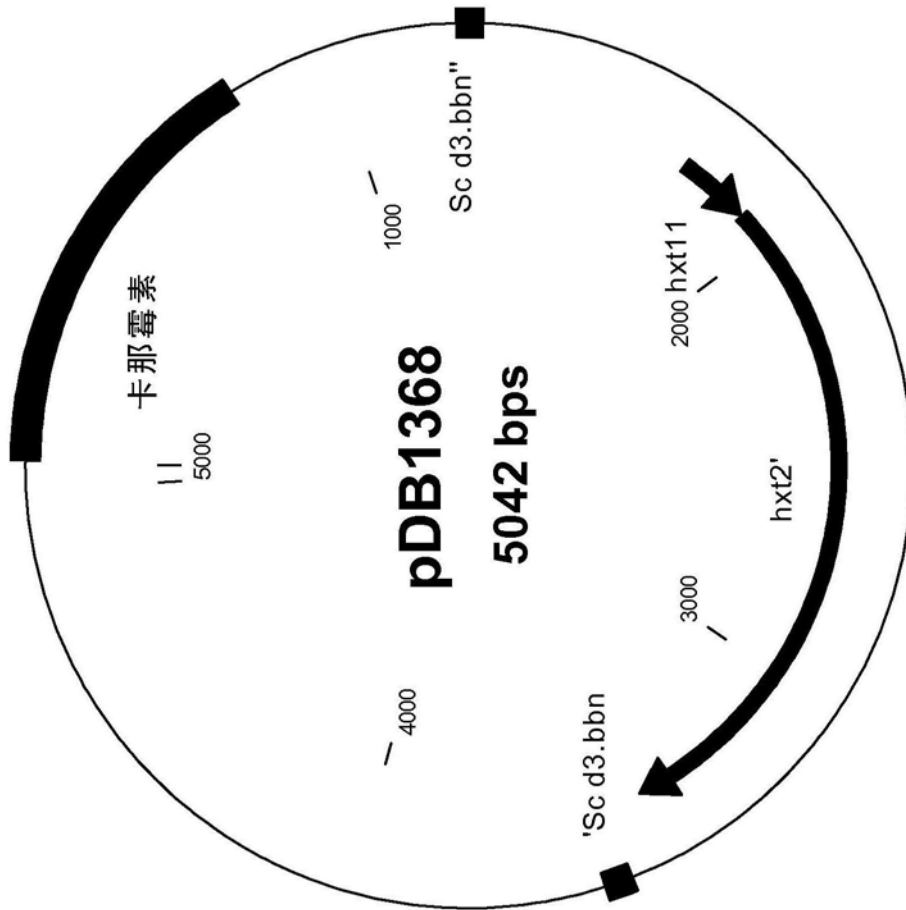


图10