(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 111094570 A
(43)申请公布日 2020.05.01

(21)申请号 201880055333.0

(22)申请日 2018.06.29

(30)优先权数据
62/527,442 2017.06.30 US
62/584,270 2017.11.10 US

(85)PCT国际申请进入国家阶段日
2020.02.26

(86)PCT国际申请的申请数据
PCT/US2018/040312 2018.06.29

(87)PCT国际申请的公布数据
WO2019/006312 EN 2019.01.03

(71)申请人 PTT全球化学股份有限公司
地址 泰国曼谷

(72)发明人 罗杰·R·约卡姆 T·格拉巴
泰龙·赫尔曼
克里斯托弗·约瑟夫·马丁
莱恩·西勒斯 余晓辉 周小美

(74)专利代理机构 北京德琦知识产权代理有限公司 11018
代理人 潘怀仁 王珍仙

(51)Int.Cl.
C12N 15/52(2006.01)
C12N 15/90(2006.01)

权利要求书2页 说明书23页
序列表18页 附图4页

(54)发明名称

具有稳定的功能DNA序列的拷贝数的微生物及相关方法

(57)摘要

本发明提供了用于鉴定和追踪可发生在经典菌株的开发期间或通过特定的遗传操纵最初构建用于生产生化物质的微生物菌株的代谢进化期间的基因组重复的方法、使用适当的可选择的标记使期望的基因组重复稳定化的拷贝数的方法,和具有稳定的功能DNA序列的拷贝数的非天然存在的微生物。

CN 111094570 A

1.一种具有稳定的功能DNA序列的拷贝数的非天然存在的微生物,所述非天然存在的微生物包括功能DNA序列的n个拷贝和在所述功能DNA序列的所述n个拷贝的至少一个相邻的对之间的至少一个可选择的标记,其中所述非天然存在的微生物是祖先微生物的后代,其中所述祖先微生物包括不多于所述功能DNA序列的n-1个拷贝,并且其中n为至少2。

2.如权利要求1所述的非天然存在的微生物,其中所述微生物是细菌、酵母、丝状真菌或古细菌。

3.如权利要求1所述的非天然存在的微生物,其中所述微生物选自由下述组成的组中:埃希杆菌属、克雷伯杆菌属、酵母菌属、青霉属、芽孢杆菌属、伊萨酵母属、毕赤酵母属、假丝酵母属、棒状杆菌属、链霉菌属、放线菌属、梭菌属、曲霉属、木霉属、根霉菌属、毛霉菌属、乳酸杆菌属、连接酵母属和克鲁维酵母菌属。

4.如权利要求1所述的非天然存在的微生物,其中所述微生物是大肠埃希杆菌或马克斯克鲁维酵母菌。

5.如权利要求4所述的非天然存在的微生物,其中所述功能DNA序列是在B重复中重复的序列,PDR12基因或PDR12基因的同源物。

6.如权利要求1所述的非天然存在的微生物,其中所述可选择的标记是编码磷酸丙糖异构酶的基因或用于代谢蔗糖的盒。

7.如权利要求1所述的非天然存在的微生物,其中在相似的培养条件下,所述微生物产生的期望的发酵产物的滴度比所述祖先微生物产生的所述期望的发酵产物的所述滴度大至少10%。

8.一种用于创建具有稳定的功能DNA序列的拷贝数的非天然存在的微生物的方法,所述方法包括下述步骤:

(a)提供包括功能DNA序列的n个拷贝的非天然存在的微生物,其中所述非天然存在的微生物是祖先微生物的后代,其中所述祖先微生物包括不多于所述功能DNA序列的n-1个拷贝,其中n为至少2;和

(b)在所述功能DNA序列的所述n个拷贝中的至少一个相邻的对之间,将至少一个可选择的标记插入到所述非天然存在的微生物中,以使所述功能DNA序列的所述拷贝数稳定化。

9.如权利要求8所述的方法,其中所述微生物是细菌、酵母、丝状真菌或古细菌。

10.如权利要求9所述的方法,其中所述微生物选自由下述组成的组中:埃希杆菌属、克雷伯杆菌属、酵母菌属、青霉属、芽孢杆菌属、伊萨酵母属、毕赤酵母属、假丝酵母属、棒状杆菌属、链霉菌属、放线菌属、梭菌属、曲霉属、木霉属、根霉菌属、毛霉菌属、乳酸杆菌属、连接酵母属和克鲁维酵母菌属。

11.如权利要求8所述的方法,其中所述微生物是大肠埃希杆菌或马克斯克鲁维酵母菌。

12.如权利要求11所述的方法,其中所述功能DNA序列是在B重复中重复的序列,PDR12基因或PDR12基因的同源物。

13.一种用于鉴定后代微生物的方法,所述后代微生物具有由祖先微生物的功能DNA序列的重复导致的至少一个改进的发酵参数,所述方法包括下述步骤:

(a)使所述祖先微生物在产生至少一种后代微生物的第一组发酵条件下生长;

(b)使所述至少一种后代微生物在第二组发酵条件下生长;

(c)鉴定与所述祖先微生物相比具有至少一个改进的发酵参数的至少一种后代微生物;

(d)确定具有至少一个改进的发酵参数的所述至少一种后代微生物的DNA序列和所述祖先微生物的所述DNA序列;和

(e)将具有至少一个改进的发酵参数的所述至少一种后代微生物的所述DNA序列与所述祖先微生物的所述DNA序列进行比较,以鉴定在所述后代微生物中所述功能DNA序列的重复。

14.如权利要求13所述的方法,其中所述第一组发酵条件是厌氧生长或微需氧生长,并且所述第二组发酵条件是需氧生长。

15.如权利要求13所述的方法,其中所述第一组发酵条件是需氧生长,并且所述第二组发酵条件是厌氧生长或微需氧生长。

16.如权利要求13所述的方法,其中所述祖先微生物是细菌、酵母、丝状真菌或古细菌。

17.如权利要求13所述的方法,其中所述祖先微生物选自由下述组成的组中:埃希杆菌属、克雷伯杆菌属、酵母菌属、青霉属、芽孢杆菌属、伊萨酵母属、毕赤酵母属、假丝酵母属、棒状杆菌属、链霉菌属、放线菌属、梭菌属、曲霉属、木霉属、根霉菌属、毛霉菌属、乳酸杆菌属、连接酵母属和克鲁维酵母菌属。

18.如权利要求13所述的方法,其中所述微生物是大肠埃希杆菌或马克斯克鲁维酵母。

19.如权利要求13所述的方法,其中所述功能DNA序列是在B重复中重复的序列,PDR12基因或PDR12基因的同源物。

具有稳定的功能DNA序列的拷贝数的微生物及相关方法

[0001] 相关申请的交叉引用

[0002] 本申请要求2017年6月30日提交的美国临时专利申请第62/527,442号和2017年11月10日提交的美国临时专利申请第62/584,270号的优先权。

[0003] 关于联邦政府赞助的研究或开发的声明

[0004] 不适用。

[0005] 联合研究协议

[0006] 不适用。

[0007] 参考序列表

[0008] 与本申请相关的序列表通过引用以其整体并入本文。含有序列表的文本文件通过USPTO电子归档系统(EFS-Web)电子提交。

技术领域

[0009] 本发明涉及通过微生物的发酵产生化学品的领域。更具体地,本发明涉及通过增加期望的化学品(也称为“发酵产物”)的滴度来改善通过发酵产生化学品的方法的经济性。甚至更具体地,可以通过创建、发现和遗传稳定化增加期望的化学品的滴度的染色体DNA序列重复来增加滴度。一旦鉴定,有益的复制可与其他有益的遗传性状结合。

[0010] 可对微生物(比如真细菌、酵母、丝状真菌和古细菌)进行基因工程改造,使其进化以产生有用的化学品,例如有机酸、醇、聚合物、氨基酸、胺、类胡萝卜素、脂肪酸、酯和蛋白质。在许多情况下,可通过基因工程和/或经典遗传学来构建生产有机体,使得期望的化学品的生产与生长耦合,从而细胞能够生长的唯一方式是将供给的碳源代谢为期望的化学品(例如见Jantama等008a;Jantama等2008b;(Zhu,Tan等2014);美国专利第8,691,539号;美国专利第8,871,489号;WO专利申请WO2011063055A2)。在构建了这样的菌株之后,可使其经受如下所述的称为“代谢进化”的过程。对于代谢进化的方法,使菌株生长在适当的条件下,比如在上述涉及琥珀酸的生产的示例中,在伴随pH控制的微需氧条件下,在最小的葡萄糖培养基中生长许多代,通过允许从小接种体(例如,以约0.01至0.5的低的起始OD600)生长足够的时间,以便发生实质的生长(例如,达到约1至30的OD600),然后再次以约0.01至0.5的低的OD600将该培养物接种至新鲜的培养基中,然后重复整个过程多次,直到培养物中进化出更快生长的菌株。在以上示例中,再接种的每个步骤也称为“转移”。每次转移时的发酵可为分批的或补料的。

[0011] 作为可选方案,不是重复地转移到新鲜的培养基中,而是可在连续的培养物(也称为“恒化器”、“auxostat”或“pHstat”)中实现代谢进化,其中将新鲜的培养基以可控的方式泵入或虹吸到充分混合的发酵罐中,并且以相似的速率去除发酵液,使得培养物的工作体积保持恒定,并且细胞生长的速率能够通过进入的新鲜培养基跟上稀释的速率。在恒化器中,调节进料速率以维持特定的细胞密度。在auxostat中,通过测量营养物或产物的浓度来确定进料速率,目的是维持所述营养物或产物的一定浓度。在pHstat中,调节进料速率以维持特定的期望的pH。随着生长速率因进化而增加,进料速率也增加,以允许连续的选择更快

生长的变体。

[0012] 在任一类型的代谢进化(即转移或连续培养)期间,在个体细胞中出现自发突变。如果偶然的情况下,自发突变赋予了生长优势,那么突变细胞的后代将缓慢地或快速地(取决于生长优势的大小)超越其他细胞生长并占领种群。在代谢进化过程期间的任何时候,都可以通过在含有等价培养基加上2%琼脂的陪替氏平皿上划线来从液体培养物中分离出个体细胞,并且与起始菌株和中间分离物相比,可以挑选和测试个体菌落。

[0013] 用于代谢进化的优选起始菌株是已通过完全缺失相关基因而消除了不希望的竞争途径的菌株,从而任何不希望的途径都不太可能通过回复或抑制而重新出现。在任何给定步骤中,可以将培养物暴露于诱变剂,例如亚硝基胍(NTG)、甲烷磺酸乙酯(EMS)、过氧化氢或紫外线辐射,以便增加突变的频率。

[0014] 在已通过代谢进化获得了经济上有吸引力的菌株后,可以通过本领域公知的任何方法对其基因组进行测序,例如鸟枪克隆和测序、Illumina平台和PacBio平台。然后可将所得的DNA序列与起始菌株或祖先菌株的DNA序列比较,从而可以鉴定在代谢进化期间发生的每个突变。可选地,可以通过下述而以单独的或组合的方式研究每个突变:将野生型等位基因(多个野生型等位基因)重新安装到进化的菌株中以研究哪些突变有助于观察到的产物形成的改善,或者“逆向工程”,其中将来自基因组序列的个体突变或起组合引入天然或野生型菌株。

[0015] 已经对被工程改造以产生琥珀酸、乙醇和D-乳酸以及其他产物的大肠埃希杆菌(大肠埃希杆菌)菌株进行了上述过程(例如,WO专利申请WO2011063055A2)。通常,设计用于处理原始基因组序列数据并将原始数据组装成完整的基因组序列的各种软件包,例如Lasergene Genomic Suite,产生当与参考(亲本)菌株比较时进化的菌株中发现的突变表。然而,这样的计算机生成的表可能是不完整的或难以精确地说明,尤其是对于重复的DNA序列。例如,大肠埃希杆菌菌株通常含有各种插入(IS)元件的多个拷贝。例如,大肠埃希杆菌Crooks(ATCC 8739),Genbank登录号NC_010468,含有许多IS4的拷贝。由于IS4为约1400个碱基对,而Illumina平台序列读数的长度仅为50-250个碱基,而双端读数通常来自长度约为500个碱基对的片段,因此测序软件无法准确地放置不同的IS4的变体,因为来自IS4拷贝的中间的序列读数将不与周围的DNA重叠。此外,即使使用PacBio平台可能具有更长的读数,在含有许多基因的大型重复中间的多样化的等位基因(比如本文公开的复制的类型)也将不易处理。

[0016] 众所周知,一个或多个基因的复制在菌株开发期间发生(Elliott,Cuff等2013)。例如,在产生青霉素的菌株的开发中,这主要通过诱变和筛选完成(与基因工程和代谢进化相对),青霉素生物合成基因簇显示出自发扩增直至5或6个串联拷贝(Fierro,Barredo等1995)。枯草芽孢杆菌(Bacillus subtilis)中基因盒的有意扩增是公知的方法,其中进入的盒含有抗生素抗性基因,例如四环素抗性基因,其提供有限的抗性,然后通过使菌株在增加抗生素的浓度时生长而使得到的转化体进化以获得更高的抗性(EP1214420)。然后确认所得菌株在串联重复中具有约2至7个盒的拷贝。然而,如果菌株在没有抗生素的情况下生长,则这样的串联重复可通过同源重组塌缩成单个拷贝。另一方面,在高浓度的抗生素的存在下生长是不切实际和不期望的。

[0017] 在大肠埃希杆菌中使用相似的方法扩增整合盒的拷贝数(Tyo,Ajikumar等2009)。

然而，再一次使用抗生素抗性基因以实现扩增，并且缺失了recA基因以保留扩增的拷贝数。众所周知，recA细胞的群体含有高百分比的死细胞，因此这种方案并不理想。在动物细胞中也已实现了基因拷贝数的扩增，但是当去除特定的选择压力时，拷贝数不稳定(Tyo，Ajikumar等2009)。

[0018] 还已知在酵母菌属(Saccharomyces)酵母中的以串联阵列的基因盒的扩增(US 7,527,927，(Lopes，de Wijs等1996))。待扩增的盒可以含有与重复的核糖体DNA基因同源的序列和可选择的标记，比如抗生素G418抗性基因(US 7,52,927)或营养缺陷型互补标记，比如dLEU2基因(Lopes，de Wijs等1996)。然而，再次，在没有强选择的情况下，扩增的拷贝数不稳定并且丢失(Lopes，de Wijs等1996)。

[0019] 在以上给出的青霉素示例中，没有可选择的标记来稳定化和维持扩增的拷贝数。尽管没有讨论扩增盒的稳定性，但是理论上，通过同源重组可能容易丢失拷贝数。维持扩增的拷贝数的唯一方式将是精心策划和频繁测试品种，以维持原始生产率。在其他示例中，待扩增拷贝数的盒开始含有可选择的标记，但是该标记需要特殊条件以便维持扩增的拷贝数，高浓度或昂贵浓度的抗生素，或者化学限定培养基。因此，在发明人已知的所有现有技术的情况下，没有用于在期望的培养条件下(即廉价、实用且非常适合产生期望的产物的培养条件下)使复制的DNA序列的拷贝数稳定化的方法。也没有已知的用于在有或没有有意选择压力的情况下使自发产生的复制的DNA序列稳定化的方法，其中没有与复制的DNA序列相关联的容易可用的可选择的标记。因此，仍然需要稳定化和维持被工程改造用于经济上有吸引力的商业化生产化学品的微生物中串联重复的DNA序列的有用的拷贝数的方法，无论该化学品是比如琥珀酸的商品，还是较高价值的化学品(比如特定的蛋白质)。本发明提供了这样的方法和衍生自所述方法的菌株。

[0020] 另一个令人惊讶的发现是，即使在大量的代谢进化和基因组测序之后，微生物菌株仍可以在常规处理和存储期间进一步进化。这样的进一步的进化可以由在不同于代谢进化期间使用的发酵条件的发酵条件下培养菌株而导致。

[0021] 对原油的增加的需求导致全球努力从可再生资源中产生可替换的燃料和化学品，以替代目前的燃料和石油衍生的化学品。2004年，美国能源部制定了来自生物质的前十二种潜在化学品的清单。这些化学品之一是琥珀酸。

[0022] 琥珀酸可以化学转化为各种各样的在工业中公知的目标化合物，包括1,4-丁二醇(BDO)、四氢呋喃(THF)、γ-丁内酯(GBL)和N-甲基吡咯烷酮。琥珀酸还可用于制造若干大宗商业产品，包括动物饲料、增塑剂、聚结剂、冷冻剂(congealers)、纤维、塑料和聚合物(比如PBS(聚丁二酸丁二醇酯))。PBS是可生物降解的聚合物，其能够代替石油衍生且不可生物降解的现有聚合物。

[0023] 琥珀酸($C_4H_6O_4$)，也称为1,4-丁二酸，是一种二羧酸，其容易采取琥珀酸阴离子的形式，并且在活生物中具有多种作用。琥珀酸盐是三羧酸(TCA)循环(所有需氧生物共有的产生能量的循环)的中间体，并且是许多细菌产生的发酵产物之一。琥珀酸盐是由葡萄糖($C_6H_{12}O_6$)(一种己糖)作为原料，通过一系列酶催化反应产生的，具有下述总体化学计量：$7C_6H_{12}O_6+6CO_2>12C_4H_6O_4+6H_2O$。当在厌氧条件下进行还原和氧化途径的氧化还原平衡的组合时，该反应的最大理论产率为每摩尔葡萄糖1.71摩尔琥珀酸或每克葡萄糖1.12克琥珀酸。

[0024] 已开发出微生物生物催化剂，以用于使用若干碳源的琥珀酸的商业规模的发酵生

产。大肠埃希杆菌KJ122菌株借助于下述从大肠埃希杆菌Crooks菌株衍生:在各种代谢途径(ΔldhA、ΔadhE、ΔackA、ΔfocA-pflB、ΔmgsA、ΔpoxB、ΔtdcDE、ΔcitF、ΔaspC、ΔsfcA)的操作中涉及的若干基因中引入突变,并在基因工程的各个阶段期间使基因工程改造的菌株经受代谢进化的过程((Jantama,Haupt等2008);(Jantama,Zhang等2008);Zhu等2014;以及美国专利第8,691,539号)。

[0025] 在最近几年中,大肠埃希杆菌KJ122已进一步改进以使用除葡萄糖以外的若干碳源。KJ122在由纤维素水解衍生的C5糖和C6糖的存在下经受代谢进化,以开发出可以在琥珀酸生产中使用C5糖和C6糖两者的菌株(美国专利第8,871,489号)。KJ122还已经受基因工程改造,以产生可以使用蔗糖(WO2012/082720)或甘油(WO2011373671)作为用于生产琥珀酸的碳源的菌株。还已经努力改善了糖输入的效率作为增加KJ122细菌菌株中琥珀酸产量的方式(WO2015/013334)。

[0026] 已使用全基因组测序鉴定在代谢进化的过程期间KJ122菌株中发生的各种突变。之后进行逆向遗传分析以建立通过全基因组测序鉴定的那些突变在琥珀酸生产中的重要性。已出乎意料地发现,若干KJ122品种在7L发酵罐中的性能有所不同,并且这些KJ122品种之间的性能的差异在琥珀酸盐滴度方面最大减少30%或增加50%(取决于使用哪种品种作为参考品种)。对这些不同的KJ122的品种进行全基因组测序,然后使用来自DNAStar(Madison,WI,USA)的Lasergene Genomic Suite软件包对这些KJ122品种的基因组序列进行比较分析,鉴定了在一些KJ122品种中的某些基因组区域中的若干功能DNA序列重复,并且发现这些基因组重复中的至少一个与琥珀酸生产的增加的滴度相关。然而,期望的基因组重复是不稳定的,因此需要一种用于使有用的重复稳定化的方法。本发明提供了一种在生产菌株中使期望的基因组重复稳定化的方法。

[0027] 由于前述原因,在本领域中对于下述仍存在未满足的需求:(1)创建增强期望的产物的生产的大型基因重复,(2)鉴定生产菌株中的大型基因重复,(3)确定大型基因重复的精确的结构,和(4)使针对拷贝数的有益的大型基因重复稳定化。

[0028] 本文公开了克服现有技术中存在的问题和局限性的方法和微生物菌株。

发明内容

[0029] 本发明涉及用于鉴定和追踪可以发生在经典菌株的开发期间或通过特定的遗传操纵最初构建用于生产生化物质的微生物菌株的代谢进化期间的基因组重复的方法、使用适当的可选择的标记使期望的基因组重复的拷贝数稳定化的方法,和具有稳定的功能DNA序列的拷贝数的非天然存在的微生物。

[0030] 在一个实施方式中,本发明提供了一种涉及全基因组DNA测序以鉴定功能DNA序列重复的方法,该功能DNA序列重复发生在通过有意的遗传操纵最初被工程改造用于生产生化物质的微生物菌株的代谢进化期间。在该实施方式的一方面中,本发明涉及比较基因组分析,其涉及产生琥珀酸的大肠埃希杆菌菌株KJ122的几种分离物和衍生物,以及鉴定包括基因组的大型多基因部分的功能DNA序列的串联重复(本文称为"B重复",其与高滴度的琥珀酸生产相关联)。在该实施方式的另一方面中,本发明已经确认并解决了在具有B重复的KJ122菌株中引入进一步的遗传修饰的挑战。

[0031] 在另一个实施方式中,本发明提供了用于使在代谢进化过程期间发生的期望的基

因组重复稳定化的方法。在该实施方式的一个方面中，本发明提供了一种用于通过在重复的基因的两个相邻拷贝之间插入可选择的标记使代谢进化期间发生的期望的基因组重复稳定化的方法。适用于该目的的可选择的标记的组包括但不限于抗生素抗性基因、编码涉及看家功能的一种或多种蛋白质的基因、必需基因、有条件必需基因、营养缺陷型互补基因以及任何编码具有可选择的表型(比如利用蔗糖作为生长的唯一碳源的能力)的蛋白质的外源基因。

[0032] 在又另一个实施方式中，本发明提供了一种用于构建和稳定化具有B重复的基因工程改造的菌株的方法，以用于发酵生产期望的生化物质。在该实施方式的一个方面中，本发明描述了构建具有针对琥珀酸生产的高滴度以及降低水平的作为副产物的乙酸的微生物菌株。

[0033] 除非另有定义，否则本文使用的所有术语具有与本发明所属领域的普通技术人员通常所理解的相同含义。尽管下面描述了用于实践的合适方法和材料的示例，但是本领域技术人员将知道，基于所公开的示例，与本文所述相似或等同的方法和材料可用于本发明的实践或测试中。本文提及的所有出版物、专利申请、专利和其他参考文献通过引用以起整体并入本文。在发生冲突的情况下，以说明书(包括术语和定义)为准。本文包括的材料、方法、示例和附图仅是说明性的，而不期望是限制性的。

附图说明

[0034] 图1是绘制读取频率(覆盖深度)相对于来自Illumina(San Diego,California,USA)平台基因组测序的基因组碱基对位置的图，和来自DNAStar(Madison,Wisconsin,USA)的LaserGene Genomic Suite测序软件，其显示了读取频率突然增加两倍导致的B重复。

[0035] 图2是描绘在三个步骤中用于在KJ122中形成B重复的逻辑假设机制的图。

[0036] 图3是描绘用于确定B重复的存在的诊断PCR的图。

[0037] 图4是描绘用于使用可选择的标记使B重复稳定化的机制的图。

[0038] 图5是描绘具有glf、glk和B重复的细菌菌株的构建的图。

[0039] 图6是描绘具有rrsG::cscBAK和稳定的B重复的细菌菌株的构建的图。

[0040] 术语

[0041] 为了促进对本发明的理解，下面提供术语的描述。

[0042] 关于术语，细菌基因或编码区通常用斜体小写字母命名，例如来自大肠埃希杆菌的“tpiA”或简称为“tpi”是编码磷酸丙糖异构酶的基因的名称，而该基因编码的酶或蛋白质可用相同的字母命名，但首字母大写且无斜体，例如“TpiA”或“Tpi”。酵母基因或编码区通常用斜体大写字母命名，例如“PDC1”，其编码丙酮酸脱羧酶，而由该基因编码的酶或蛋白质可以用相同的字母命名，但首字母大写且无斜体，例如“Pdc1”或“Pdc1p”，其后者是酵母中用于指定酶或蛋白质的约定的示例。“p”是蛋白质的缩写，由指定的基因编码。酶或蛋白质也可用更具描述性的名称来指代，例如磷酸丙糖异构酶或丙酮酸脱羧酶，分别指以上两个示例。编码具有特定催化活性的酶的一个示例的基因或编码区可以具有几个不同的名称，这是由于历史上不同的起源、功能上冗余的基因、不同地调控的基因，或者因为该基因来自不同的物种。例如，可以将编码甘油-3-磷酸脱氢酶的基因命名为GPD1、GDP2或DAR1，以及其他名称。

[0043] 定义

[0044] 为了促进对本发明的理解，下面定义了若干术语，并且其他术语在说明书的其他地方找到。

[0045] “微生物”意为细菌、酵母、丝状真菌或古细菌的任何细胞或菌株。通过使用一种或多种公知的方法，例如通过诱变、交配、育种、基因工程、进化、选择和筛选，微生物可以被有意地遗传改变，或允许其自发地遗传变化。在这样的过程中，遗传改变起始菌株(本文称为“祖先微生物”或“祖先菌株”)，以创建新的菌株，该新的菌株可以通过获得一个或多个遗传变化后来自所述祖先菌株的一个或多个细胞分裂而繁殖。在遗传上与祖先菌株不同但通过遗传改变和随后的细胞分裂从所述祖先菌株衍生而来的菌株在本文中称为“后代微生物”或“后代菌株”。相对于其祖先菌株，后代微生物可以具有一个或多于一个遗传变化。后代微生物可以是来自其祖先微生物的任何有限数量的世代(细胞分裂)的结果。一种类型的后代微生物是“衍生微生物”，其意为通过稳定化添加、去除或改变祖先微生物中的DNA序列而创建的微生物。后代微生物可以通过DNA测序或通过任何其他可测量的表型，比如改进的发酵参数，而与其相关的祖先微生物区分开。

[0046] “盒”、“表达盒”或“基因盒”意为，当安装在宿主生物中时，能够引起或增加，或可替换地消除或降低，一种或多种期望的蛋白质、酶或代谢物的生产的脱氧核糖核酸(DNA)序列。用于产生蛋白质或酶的盒通常包括至少一个启动子、至少一个蛋白质编码序列和可选地至少一个终止子序列。如果要表达的基因是异源或外源的，则启动子和终止子通常衍生自两个不同的基因或衍生自异源基因，以便防止与该启动子或终止子从其衍生的天然基因双重重组。盒可以可选地并且优选地在一端或两端上含有与祖先(或“宿主”或“亲本”)生物中的DNA序列同源的一个或两个侧翼序列，使得对于染色体或质粒，该盒可以经历与宿主生物的基因组的同源重组，导致在与所述侧翼序列同源的位点将盒整合到所述染色体或质粒中。如果只有盒的一端含有侧翼同源物，则可以通过在侧翼序列处的单个重组整合环状形式的盒。如果盒的两端均含有侧翼同源物，则以线性或环状形式提供的盒可以通过与周围的侧翼双重重组而整合，或者环状形式可以通过一次交换事件以其整体整合。盒可以通过基因工程(其中例如，从非天然启动子表达编码序列)构建，或者盒可以使用天然相关联的启动子。盒可以构造成质粒，该质粒可以是环状的，或者其可以是通过限制性酶切、聚合酶链反应(PCR)、引物延伸PCR或通过体内或体外同源重组创建的线性DNA。盒可以被设计成包括可选择的标记。盒可以使用本领域公知的方法在一个或多个步骤中构建，例如通过连接反应连接限制性酶产生的片段、使用NEBuilder试剂盒(New England BioLabs，Ipswitch，Massachusetts，USA)的“Gibson法”以及体内组装。

[0047] “可选择的标记”意为这样的任何基因、盒或其他形式的DNA序列：其在亲本微生物菌株或祖先微生物菌株中功能上不存在，但可以安装到所述祖先微生物菌株中，可以在安装在所述祖先微生物中后起作用以产生一种后代菌株，并且在至少一组生长条件下对于所述后代菌株的生长是必需的。这样，可选择的标记单独地或当与一个或多个其他有用的DNA序列一起包含在盒中时，可在转化、转导、转染、繁殖或交配到先前不含有所述可选择的标记的菌株时，用于选择或筛选具有或不具有额外附接的DNA序列的所述可选择的标记的成功安装。在一些情况下，可选择的标记可以在菌株中突变(优选从菌株中缺失)，于是然后可以将未突变的形式用作所得突变的或缺失的菌株中的可选择的标记。在大多数情况下，没

有可选择的标记的祖先菌株能够在一组特定条件下(例如,丰富的培养基、营养补充的培养基或缺乏抗生素的培养基)生长,但是在将可选择的标记安装到亲本菌株中后,后代或后代菌株能够在亲本菌株不能生长的一组条件下(例如,最低限度培养基、缺乏营养的培养基或含有抗生素的培养基)生长。有用的可选择的标记基因包括但不限于功能抗生素抗性基因或盒、赋予在特定碳源(比如蔗糖或木糖)上的生长能力的基因或盒,以及在某些生长条件下可以缺失的生物合成途径基因(例如大肠埃希杆菌中的tpiA,其在最低限度的葡萄糖或最低限度的蔗糖培养基中是必需的,但在比如Luria Broth的丰富的培养基中则不是必需的)。为了将生物合成途径基因(例如pyrF或URA3)用作可选择的标记,亲本菌株当然必须在相应的基因中含有突变,优选无效突变(导致功能有效丧失的突变)。对于抗生素抗性基因,抗性基因通常需要启动子,该启动子在宿主菌株中具有足够好的功能以能够进行选择。尽管可以以盒的形式将期望表达的可选择的标记安装在宿主菌株中,但是可以将DNA序列(例如从起始密码子到终止密码子的编码序列)整合到宿主染色体或质粒中,而无需启动子或终止子,使得进入的编码序列精确地或近似地替换宿主菌株原有基因的编码序列,使得整合后,进入的编码区从已与被进入的编码序列代替的宿主编码序列相关联的其余启动子表达。

[0048] “转化体”意为由将自主复制或不自主复制的线性或环状的期望的DNA序列安装到先前不含有所述DNA序列的宿主或亲本菌株或祖先菌株中而得到的细胞或菌株。“转化”意为用于获得转化体的任何过程。

[0049] “滴度”意为化合物在发酵液中的浓度,通常以克/升(g/l)或重量百分比/体积(%w/v)表示。滴度通过任何合适的分析方法确定,比如定量分析色谱法,例如高压液相色谱法(HPLC)或气相色谱法(GC),使用由外标制成的标准曲线,以及可选地使用内标。

[0050] “异源”意为不是在生物中自然或天然发现的,而是可以通过基因工程(比如通过转化、交配、转染或转导)引入生物中的基因或蛋白质。异源基因可以整合(即,插入或安装)到染色体中,或包含在质粒上。“外源”意为为了相对于亲本菌株或宿主菌株的活性增加、减少或消除活性的目的而通过基因工程,比如通过转化、交配、转导或诱变,引入生物中或在生物中改变的基因或蛋白质。外源基因或蛋白质可以是异源的,或者其可以是宿主生物原有的但已通过一种或多种方法改变(例如,突变、缺失、启动子的变化、终止子的变化、在染色体或质粒中复制或插入一个或多个额外的拷贝)的基因或蛋白质。因此,例如,如果将DNA序列的第二拷贝插入染色体中与天然位点不同的位点,则第二拷贝将是外源的。

[0051] “质粒”意为基本上小于染色体的环状或线性DNA分子,与微生物的一个或多个染色体分开,并且与一个或多个染色体分开复制。质粒可以以每个细胞约一个拷贝或每个细胞多于一个拷贝存在。在微生物细胞中维持质粒通常需要在选择质粒存在的培养基中生长,例如使用抗生素抗性基因或染色体营养缺陷型的互补。然而,一些质粒不需要选择压力用于稳定的维持,例如在许多酵母菌属菌株中的2微米圆质粒。

[0052] “染色体”或“染色体DNA”意为线性或环状DNA分子,其实质上大于质粒,并且通常不需要任何抗生素或营养选择用于维持。在本发明中,酵母人工染色体(YAC)可以用作安装异源和/或外源基因的载体,但是其将通常需要选择压力用于维持。

[0053] “过表达”意为使得由基因或编码区编码的酶或蛋白质在亲本微生物或宿主微生物中产生的水平高于在相同或相似的生长条件下在亲本微生物或宿主微生物的野生型形

式中发现的水平。例如,这可以通过一种或多种下述方法来实现:安装更强的启动子、安装更强的核糖体结合位点、安装终止子或更强的终止子、改善编码区中一个或多个位点处的密码子的选择、改善mRNA的稳定性,或通过在染色体中引入多个拷贝或将盒置于多拷贝质粒上来增加基因的拷贝数。由过度表达的基因产生的酶或蛋白质被称为是“过度产生的”。过表达的基因或过度产生的蛋白质可以是宿主微生物原有的基因,或者其可以是通过基因工程方法从不同的生物移植到宿主微生物中的基因,在这种情况下酶或蛋白质以及编码该酶或蛋白质的基因或编码区称为“外来的”或“异源的”。根据定义,外来的或异源的基因和蛋白质是过表达的和过度产生的,因为它们不存在于未经工程改造的宿主生物中。

[0054] “同源物”意为第一基因或DNA序列与第二DNA序列相比具有至少50%的序列同一性,或当所述第一DNA序列被翻译成第一蛋白质序列并且将所述第一蛋白质序列与通过翻译所述第二DNA序列而衍生的第二蛋白质序列进行比较时具有至少25%的氨基酸序列同一性(如由用于序列比较的Basic Local Alignment Search Tool(BLAST)计算机程序确定的(Altschul,Gish等1990,Altschul,Madden等1997),并允许缺失和插入)。如果发现第一DNA或蛋白质序列是第二DNA或蛋白质序列的同源物,则将这两个序列称为“同源物”或是“同源的”。当意图将盒整合到基因组的特定位点时,优选侧翼同源DNA序列与被靶向的DNA序列具有100%同一性或几乎100%同一性。

[0055] “相似物”意为基因、DNA序列或蛋白质与另一种基因、DNA序列或蛋白质实施相似的生物学功能,但是与所述另一种基因、DNA序列或蛋白质有小于25%的同一性(当比较蛋白质序列或比较由基因序列衍生的蛋白质序列时)(由用于序列比较的BLAST计算机程序确定(Altschul等1990;Altschul等1997),并允许缺失和插入)。酿酒酵母(Saccharomyces cerevisiae)Gpd1蛋白的相似物的一个示例是酿酒酵母Gut2蛋白,因为这两种蛋白质都是催化相同反应的酶,但是两种酶或其各自的基因之间没有显著的序列同源性。本领域普通技术人员将知道,可以在许多不同的生物中找到许多作为同源物或相似物的具有特定生物学功能的酶和蛋白质(在刚刚上面的示例中是甘油-3-磷酸脱氢酶),并且由于这样的酶或蛋白质家族的成员共有相同的功能,尽管它们的结构可能略有不同或实质上不同。在许多情况下,可以使用当前的基因工程方法,使用同一家族的不同成员来实施相同的生物学功能。因此,例如,可以从许多不同生物中的任何一种获得编码磷酸丙糖异构酶的基因。

[0056] “突变”意为来自天然或亲本或祖先DNA序列的任何变化,例如,倒位、复制、一个或多个碱基对的插入、一个或多个碱基对的缺失、导致创建提前终止密码子的碱基变化的点突变,或改变该位置处编码的氨基酸的错义突变。突变甚至可以意为DNA序列中的单个或多个碱基对变化,其不导致由所述DNA序列编码的预测氨基酸序列的变化。“无效突变”意为有效消除基因的功能的突变。编码区的完全缺失将是无效突变,但是单个碱基的变化也可导致无效突变。“突变体”、“突变菌株”、“突变的菌株”或“已突变的”菌株意为与天然、野生型、亲本或祖先菌株相比包括一个或多个突变的菌株。

[0057] “消除或降低其功能的突变”意为降低基因、蛋白质或酶的任何可测定参数或输出(比如菌株的mRNA的水平、蛋白质浓度、代谢产物生产或特定酶活性)(当测量所述可测定参数或输出并将其与在相似条件下生长的未突变的亲本菌株的可测定参数或输出进行比较时)的突变。这样的突变优选是缺失突变,但是它可以是实现期望的功能消除或降低的任何类型的突变。

[0058] “强组成型启动子”意为通常位于被RNA聚合酶转录的DNA序列或基因上游(当以常规5’至3’方向描绘时,基因的5’侧)的DNA序列,并且其导致所述DNA序列或基因以容易通过任何适当的测定程序直接或间接检测到的水平通过被RNA聚合酶转录来表达。合适的测定程序的示例包括定量逆转录酶加PCR、编码的酶的酶测定、考马斯蓝染色的蛋白质凝胶或由于所述转录而间接产生的代谢物的可测量的生产,并且这样的可测量的转录发生无论存在或不存在特异性调节转录水平的蛋白质、代谢物或诱导化学品。通过使用公知的方法,可以使用强组成型启动子替代天然启动子(另外在DNA序列或基因上游的天然存在的启动子),得到可置于质粒或染色体中的表达盒,其以高于天然启动子的水平提供期望的DNA序列或基因的表达水平。强组成型启动子可以对一物种或属具有特异性,但是来自酵母的强组成型启动子通常可以在远缘相关的酵母中发挥良好的功能。例如,来自棉阿舒囊霉(Ashbya gossypii)的TEF1(翻译延伸因子1)启动子在许多其他酵母属(包括酿酒酵母)中都发挥良好的作用。

[0059] “微需氧”或“微需氧发酵条件”意为向发酵罐有意供应的空气少于每体积液体肉汤每分钟0.1体积的空气(vvm)。“厌氧”或“厌氧发酵条件”意为没有有意向发酵罐供应空气。“有氧”或“有氧发酵条件”意为每分钟有意向发酵罐供应每体积液体肉汤每分钟0.1体积或更多体积的空气(vvm)。典型地,“发酵”是指微生物的厌氧或微需氧培养。然而,为了简化本说明书,术语“发酵”或“发酵条件”意为微生物的任何类型的生长或培养,包括厌氧、微需氧或需氧,并且包括在液体培养基中或固体培养基上的生长,例如在琼脂培养皿上的生长。“发酵器”意为在其中进行发酵或可以进行发酵的任何容器。在摇瓶培养(也称为振荡瓶培养)中,如果不精确控制通气条件,则发酵条件可以是厌氧、微需氧或需氧的,并且在培养过程期间,条件可以变化。例如,在摇瓶培养开始时,接种物水平低(例如,起始OD600为0.5或更小),并且剧烈摇动(例如200rpm或更大),松动的装配盖或多孔盖,发酵条件可能是有氧的。然而,随着培养物生长到更高的密度(例如OD600为10或更大),相对于氧气进入烧瓶的速率,微生物消耗的氧气可能会很大,导致厌氧或微需氧条件。摇瓶中的发酵条件可以通过使用空气阱(例如,允许二氧化碳逸出的鼓泡器)或空气不可渗透的盖或闭合器强制为厌氧或微需氧的。应当理解,除非使用严格的条件(例如,用氮气、二氧化碳或氩气充气),否则不能达到严格的厌氧条件,从而在厌氧和微需氧发酵条件之间存在连续性。因此,术语厌氧和微需氧在本文中通常一起使用。

[0060] “重复”意为功能DNA序列在相关祖先微生物中每个单倍体基因组中存在n-1个拷贝之后,在后代微生物中每个单倍体基因组中存在n个拷贝,其中n是大于或等于2的整数。术语“重复”是指所述功能DNA的所有n个或更多个拷贝。“重复的DNA”、“被重复的DNA”或“重复的DNA序列”意为所述功能DNA的任何一个单拷贝。在重复中,重复的DNA拷贝的末端可能在各个拷贝之间有所不同。因此,对于本文公开的B重复的示例,功能DNA序列的一个拷贝在menC编码序列的中间以IS4插入元件结尾,而第二个拷贝以menC基因的完整拷贝结尾。重复的DNA的两个或更多个拷贝在它们的序列中可以含有微小差异。两个或更多个拷贝彼此相邻或几乎彼此相邻的重复称为“串联重复”。为清楚起见,术语“重复”和“重复的DNA”不指代所述功能DNA的额外的拷贝,其通常在微生物的基因组的复制期间创建,作为细胞分裂或细胞出芽的正常前体。

[0061] “相似的培养条件”或“相似的发酵条件”意为设计用来比较两种不同微生物的性

能的条件，其中，实验者试图将两种培养物中的所有条件设置和控制为尽可能相同。本文使用术语“相似”是因为众所周知，实际上不可能使两个独立的微生物培养物中的条件绝对相同。

[0062] “发酵参数”意为发酵或培养的几个可测量的方面之一，例如，完成时间、温度、pH、以克/升(g/l)为单位的产物的滴度、以克产物/克输入营养物为单位的产率，或比生产率(每小时克产物/克细胞质量)。当两种微生物都在相似的培养条件下培养时，如果一个或多个发酵参数在经济上对后代微生物更有利，则当与相关祖先微生物相比，后代微生物的发酵参数据称“得到改善”。改善的发酵参数的示例为增加的滴度、产率或比生产率，或缩短的完成时间。

[0063] “化学限定培养基”、“最低限度培养基”或“矿物质培养基”意为任何这样的生长培养基：其由纯化或部分纯化的化学品构成，比如矿物盐(例如，钠、钾、铵、镁、钙、磷酸盐、硫酸盐和氯化物)，其提供必要的元素，比如氮、硫、镁、磷(有时钙和氯化物)、维生素(必要时或刺激微生物生长时)、一种或多种纯化的碳源(比如糖、甘油、乙醇、甲烷、微量金属(必要时或刺激微生物生长时)(比如铁、锰、铜、锌、钼、镍、硼和钴))，以及可选地渗透保护剂(比如甘氨酸甜菜碱，也称为甜菜碱)。除了可选的渗透保护剂和维生素外，这样的培养基不含有大量的对正在发酵的微生物的生长不必需的任何营养物或多于一种营养物的混合物。如果微生物是营养缺陷型，例如氨基酸或核苷酸营养缺陷型，则可以支持所述营养缺陷型的生长的最低限度培养基将必须含有所需的营养，但是对于最低限度培养基，添加的营养物将是基本上纯的形式。最低限度培养基不含有任何大量的丰富或复杂的营养物混合物，比如酵母提取物、蛋白胨、蛋白质水解物、糖蜜、肉汤、植物提取物、动物提取物、微生物提取物、乳清和菊芋粉。为了通过发酵产生商品化学品(其中在通过简单蒸馏纯化期望的化学品在经济上不是具有吸引力的选择)，最低限度培养基优选于丰富培养基。

[0064] “发酵生产培养基”意为在其中使微生物生长以产生期望的产物(例如琥珀酸)的过程中，在包括一个或多个罐、容器或发酵器的系列中，在最后的罐、容器或发酵器中使用的培养基。为了通过发酵生产商品化学品(比如琥珀酸)(其中必须或期望进行大量纯化)，作为最低限度培养基的发酵生产培养基优选于丰富培养基，因为最低限度培养基通常较便宜，并且发酵结束时的发酵液通常含有较低浓度的需要从期望的化学品中纯化出来的不希望的污染化学品。尽管一般优选在这样的发酵中最小化丰富营养物的浓度，但是在一些情况下，对于整个过程，使接种培养物生长在不同于发酵生产培养基的培养基中是有利的，例如使相对小体积(通常为发酵生产培养基体积的10％或更少)的接种物生长在含有一种或多种丰富成分的培养基中。如果接种物培养物相对于生产培养物小，则可以将接种物培养物的丰富组分稀释到发酵生产培养基中，达到它们基本上不干扰期望的产物的纯化的程度。发酵生产培养基必须含有碳源，其通常是糖、甘油、脂肪、脂肪酸、二氧化碳、甲烷、酒精或有机酸。在一些地理位置，例如在美国中西部，D-葡萄糖(右旋糖)相对便宜，因此可用作碳源。关于酵母生产乳酸的大多数现有技术出版物使用右旋糖作为碳源。然而，在一些地理位置，比如巴西和东南亚大部分地区，蔗糖比右旋糖便宜，因此蔗糖是这些地区的优选碳源。

[0065] “不等交换”意为使DNA序列变成复制的或复制的DNA序列变成不被复制的机制。当复制的DNA序列有两个拷贝彼此紧邻时，通常在DNA复制期间发生不等交换(Reams和Roth，

2015)。

[0066] “功能DNA序列”是当存在于生物的基因组中时,单独地或当与另一DNA序列串联附接时,产生可测量的表型或输出的任何DNA序列。功能DNA序列的示例是菌株KJ122-RY的染色体的段,其包括111个基因,该111个基因在菌株KJ122-RY中以一个拷贝存在,并且在菌株KJ122-F475中以两个拷贝存在(存在于B重复中)(参考表1和表2)。当存在第二个拷贝时,可测量的表型或输出为在典型的可比较的发酵中琥珀酸盐滴度从约57g/L增加至约89g/L。功能DNA序列的另一个示例是在青霉素生产菌株中扩增的青霉素生物合成基因簇(Fierro,Barredo等1995)。在这种情况下,表型或输出为增加的青霉素的产量。

具体实施方式

[0067] 在一个实施方式中,本发明提供了一种用于增加期望的发酵产物的滴度的方法,该方法包括在第一组条件(例如厌氧或微需氧条件)下,使菌株代谢进化,然后使所得的进化的菌株在第二组条件(例如需氧条件)下生长,以允许进一步进化,并在从第二条件分离的菌株中筛选改进了生产期望的发酵产物的菌株。

[0068] 在另一个实施方式中,本发明提供了用于鉴定和追踪基因组重复的方法,该基因组重复导致期望的发酵产物的改进的生产。

[0069] 在另一个实施方式中,本发明提供了用于使有益基因组重复稳定化的方法。

[0070] 在另一个实施方式中,本发明提供了具有稳定的功能DNA序列的拷贝数的非天然存在的微生物。

[0071] 在通过生物发酵生产有价值的生物化学物质的微生物菌株的产生中,可以遵循两步过程。在第一步骤中,将合理设计的遗传修饰引入微生物细胞中。在第二步骤中,使经遗传修饰的微生物细胞经受代谢进化,以获得具有期望的表型的微生物催化剂。例如,在开发用于产生琥珀酸的细菌菌株的情况下,引入了若干遗传修饰以将微生物细胞内的碳的路径朝琥珀酸生产引导。有意的遗传修饰是基于我们对微生物细胞中代谢途径的了解而设计的。期望的遗传修饰是分阶段进行的。在基因修饰的阶段之间以及所有相关基因修饰的最后,可使微生物菌株经受代谢进化,以允许微生物细胞群达到期望的表型,即期望的生物化学物质的生长、滴度和生产率的增加。换句话说,当生长与特定化学品的生产耦合时,选择更快的生长(代谢进化)导致该化学品的更快生产。

[0072] 预期代谢进化的过程产生有利于期望的表型的特定突变,即针对期望的产物的更有利的生产参数。因此,在代谢进化的最后,微生物菌株将具有获得的某些特定的遗传修饰。在现有技术示例中,这些遗传修饰已显示出包括简单的基因组的重要部分的核苷酸变化、插入和缺失。基因组测序成本的急剧下降使得有可能对代谢进化的菌株的整个基因组进行测序,以确认仍然保留了最初引入的特定突变,并且鉴定在代谢进化的过程期间获得的突变。

[0073] 一旦鉴定了代谢进化期间的微生物菌株中发生的突变,就可能通过逆向遗传分析来确认所鉴定的突变的功能重要性。当突变在单个基因中时,逆向遗传分析容易进行。

[0074] 当突变是在代谢进化期间或在随后的第二组条件下培养已发生的主要基因组重复时,逆向遗传分析将是困难的或不可能的。然而,如果通过密切相关的微生物菌株之间的可比较的基因组和表型分析发现主要基因组重复与期望的表型密切相关联,则符合期望的

是维持该基因重复在菌株的随后的培养和大规模使用中不丢失。建立稳定地维持DNA复制的方法的第一步是了解复制的精确结构以及导致复制的机制。一旦发明人了解了导致即时主要基因组重复的结构和分子机制以及所得到的结构，就有可能进一步对菌株进行基因工程改造以稳定地维持复制。开发基于聚合酶链反应(PCR)的诊断方法以检测微生物菌株中存在或不存在基因重复也是有用的，以便容易地表明已实现了稳定化。

[0075] 一种稳定地维持原本未标记的基因重复(缺乏方便或实用的可选择的标记的复制)的新型方法是在相邻的一对复制的DNA序列之间插入可选择的标记(可以在至少一种培养条件下选择的DNA序列，比如基因、基因的组或操纵子)，而不干扰复制的序列内的任何基因的表达。可以容易地引入基因组重复的区域的一种类型的可选择的标记是编码抗生素抗性的基因，也称为抗生素抗性基因或抗生素抗性标记。若干容易有用的抗生素抗性基因是本领域公知的。在大肠埃希杆菌中，存在公知的基因，其赋予针对例如但不限于青霉素(例如氨苄青霉素)、四环素、卡那霉素、氯霉素、链霉素、壮观霉素或红霉素的抗性。然而，如上所述，一般不期望将抗生素抗性基因用于大规模发酵。适用于本目的的优选的可选择的标记是编码必需或条件必需的蛋白质的内源性(天然)或外源性基因。在这种情况下，在将基因插入在复制中的适当位置(例如在复制的基因的两个拷贝之间的连接处)之前或之后，野生型基因从其天然位点突变或缺失。例如，可以通过基因工程或经典遗传学方法使编码磷酸丙糖异构酶的天然tpiA基因缺失或基本上失活。具有失活的内源性tpiA基因的微生物细胞不能在含有葡萄糖或另一种糖作为碳源的最低限度培养基中生长；然而，tpiA突变体可以在比如LB(Luria Broth)的丰富培养基中繁殖。一旦将外源性tpiA基因盒(被设计为整合在非天然位点处的盒)插入缺少功能tpiA基因的菌株的基因重复中，后代菌株将恢复在包括葡萄糖或其他糖作为碳源的最低限度培养基中生长的能力，并且结果，基因重复将得以稳定地维持。用于稳定地维持基因重复的又另一种方式涉及使用编码蛋白质或操纵子的外源性基因或编码赋予可选择的表型的一组蛋白质的另一组基因。例如，如果已获得基因重复的微生物菌株由于没有编码一个或多个参与蔗糖代谢的蛋白质的一个或多个基因而缺乏利用蔗糖作为碳源的能力，则可以将编码蔗糖利用所需的一个或多个蛋白质的基因盒插入到基因重复中，然后通过在含有蔗糖作为唯一碳源的培养基中使所得菌株生长来稳定地维持复制。

[0076] 使用特定的可选择的标记来稳定地维持基因重复的决定取决于总体环境。例如，当葡萄糖是期望的碳源时，使用蔗糖利用基因作为可选择的标记将是不合适的。尽管抗生素抗性基因可以很好地起到使基因重复稳定化的作用，但是一般优选使用必需或条件必需基因，比如tpiA作为可选择的标记，从而避免需要使用昂贵的抗生素，并且避免大规模生产潜在可传播的抗生素抗性基因。

[0077] 代谢进化通常在特定组条件下进行，比如在最低限度培养基中的厌氧或微需氧发酵(例如，见Jantama等2008a;Jantama等2008b;(Zhu,Tan等2014);美国专利第8,691,539号;美国专利第8,871,489号;WO专利申请WO2011063055A2)。然而，在本发明中，发现使进化的菌株经受第二组条件，例如有氧培养，可以出乎意料地导致进一步的有益进化，该有益进化在没有在第二组条件下生长的情况下预期不发生。在下面给出的示例中，在祖先菌株KJ122-RY的培养物中第二组发酵条件期间，发生了111个基因的大型重复(“B重复”)，并且如果该菌株没有在第二组条件下生长，该重复不可能发生。可能导致B重复的唯一逻辑系列

事件如下。首先，可转座的IS4元件将其自身插入menC基因的中间。menC基因是menFDHBCE操纵子中的第五个基因，其编码生物合成甲萘醌所必需的酶。甲萘醌是大肠埃希杆菌和其他微生物在厌氧生长期间使用的电子载体。menC的突变导致在最低限度的葡萄糖培养基上厌氧生长不良(Guest 1977)。由于KJ122的代谢进化在最低限度葡萄糖培养基中微需氧条件下进行，因此由于在menC无效突变体将处于生长劣势，所以不可能在进化期间发生IS4插入menC中。在导致B重复的第二步中，在大肠埃希杆菌Crooks基因组的Genbank形式(登录号NC_010468)中标注为EcolC_1276的IS4的拷贝与menC基因中的IS4的拷贝之间的111个基因区域为精确地重复，大概是因为上述IS4的两个拷贝之间的不等交换。该得到的中间菌株仍将缺少功能menC基因，所以同样地，该事件将不可能发生在微需氧进化期间。与此相符的事实是，从研究型大学实验室获得的KJ122的原始分离株(此处命名为KJ122-RY)没有B重复。在第三步中，精确切除B重复的第二个拷贝的末端的menC中的IS4元件，以重建功能menC基因，允许所得菌株KJ122-F475在微需氧条件下良好生长(这些步骤的图解形式见图4)。发明人的标准实践是在最低限度葡萄糖培养基中在陪替氏培养皿上和摇瓶中使菌株(比如KJ122-RY和KJ122-F475)需氧生长，以制备-80℃冷冻品种，并制备用于7升发酵的接种物。因此，我们认为导致B重复的前两个步骤一定是在那些需氧生长期期间发生的。由于B重复至少在三种情况下独立出现，因此发明人推断必须存在选择压力以发生重复，例如在第二组发酵条件下具有在重复的DNA序列中含有的一个或多个基因的两个拷贝的选择优势。第三步可能发生在摇瓶培养期间，在溶解氧浓度相对低，并且普遍存在微需氧条件，导致重新产生menC基因的功能拷贝的压力的时候。虽然导致B重复的形成的系列事件由下述得到：在微需氧条件下培养微生物，然后在需氧条件下培养微生物，最后一步导致新型菌株KJ-122-F475(见下文)，即可能在菌株再次经受微需氧条件时发生的从menC基因中精确切除IS4元件。因此，有益的遗传事件，比如最终形式的B重复的形成和建立，可能由下述得到：首先在微需氧条件下，然后随后在需氧条件下培养微生物，或者首先在需氧条件下，然后随后在厌氧或微需氧条件下培养微生物。

[0078] 一旦产生了具有稳定的与期望的表型相关联的基因重复的微生物菌株，该菌株就可以用作构建用于商业生产增值化学品的改进的微生物菌株的起点。在另一种方式中，也可能例如通过交配将与期望的表型相关联的基因重复引入已经进行基因工程改造以产生增值生化物质的微生物菌株中，或将另外的期望的性状引入已经含有重复的菌株(比如KJ122-F475)中。

[0079] 实施例

[0080] 实施例1

[0081] 各种KJ122品种和噬菌体抗性衍生物的基因组结构

[0082] 表1提供了通过源于原始大肠埃希杆菌KJ122菌株的各种不同品种(包括几种噬菌体抗性衍生物)获得的基因组结构和琥珀酸盐滴度的总结。使用来自Illumina，Inc.(San Diego，California，USA)的技术对表1中列出的所有7个菌株进行全基因组序列，并使用Lasergene Genomic Suite软件包(DNAStar，Madison，WI)分析基因组数据。从DNA序列数据分析中，明显的是，当与亲本菌株大肠埃希杆菌Crooks相比时，名为“KJ122-F475”(有时称为“KJ122-F”)的特别品种出乎意料地获得了两个多基因重复(图1)。这两个多基因重复在本文中称为“A重复”和“B重复”。本文中我们也将包括B重复的菌株称为“B+”。

[0083] A重复包括66个基因，并且B重复包括111个基因。插入元件IS4出现在B重复中重复的序列的连接处(图2)。IS4能够拷贝其自身并将该拷贝插入染色体中的随机位置。表1中所示的琥珀酸盐滴度的模式以及A重复和/或B重复的存在或不存在强烈表明，B重复是由一些菌株，例如由KJ122-F475产生的较高琥珀酸盐滴度的唯一原因。此外，当依靠用于葡萄糖输入的促进的扩散并产生显著较少的乙酸盐副产物的菌株MH141(见WO2015/013334)经受代谢进化而产生新的菌株FES33时，基因组DNA测序揭示的唯一变化是精确相同的B重复的获得。由于FES33在补料分批发酵中始终产生较高的琥珀酸盐滴度，因此该观察结果进一步支持了B重复促进较高的琥珀酸盐滴度的说法。最后，菌株KR122-RY的噬菌体抗性衍生物MYR585-4E独立获得了B重复和琥珀酸盐滴度的增加，又进一步支持了B重复是琥珀酸盐滴度增加的原因的假设。

[0084] 表2列出了与本发明有关的各种菌株。由于期望将B重复与在作为唯一碳源的蔗糖上生长的能力结合，因此尝试了结合菌株SD14的rrsG::cscBAK特征(KJ122-RY rrsG::cscBAK；见WO2012/082720)。结合两个特征的第一种尝试通过下述进行：在SD14上使P1vir转导噬菌体生长，并且在最低限度蔗糖平皿上选择，转导至KJ122-F475。获得了许多蔗糖+转导子，但是都已经丢失了B重复(通过使用引物BY296和BY297的诊断性PCR显示)，表明B重复不稳定且可能丢失，大概是由于重复的两个拷贝的简单的同源重组(环出)或重复的两个拷贝之间的不等交换。第二种尝试是成功的，其使用公知的重组DNA转化方法将来自SD14的rrsG::cscBAK等位基因转移到KJ122-F475，即在质粒pKD46上安装噬菌体λred重组系统(见表3)，然后用含有与整合靶位点侧翼同源的线性DNA转化(Jantama等2008a；Jantama等2008b)。这成功表明rrsG::cscBAK等位基因与B重复不是根本不相容的。

[0085] 考虑到B重复的潜在不稳定性和丢失，期望发明一种用于使B重复稳定化以防丢失的方法。如本文所公开的，可以通过在B重复的两个拷贝之间的连接处插入可选择的标记，比如必需基因盒或条件必需基因盒，来实现期望的稳定化，使得B重复塌缩回到一个拷贝将导致插入的可选择的标记的丢失。条件可选择的标记的具体示例是(1)cscBAK操纵子，其在菌株背景中不存在蔗糖利用基因的情况下，对于在最低限度蔗糖培养基上的生长是必需的(见WO2012/082720)，以及(2)编码磷酸丙糖异构酶的基因，比如大肠埃希杆菌Crooks的tpiA基因，该基因对于在最低限度葡萄糖培养基上的生长是必需的，但对于在丰富培养基，比如LB(Luria Broth)上的生长不是必需的。本领域技术人员将知道，可以以与本文用于使串联基因重复或串联多基因重复稳定化所述的相似的方式使用各种各样的可选择的标记和/或基因盒和/或操纵子。唯一的要求是可选择的标记或基因盒对于在至少一种生长条件下的生长是必需的(即，其是必需的或条件必需的)。其他示例是看家基因、抗生素抗性基因、补充营养缺陷的基因以及赋予在宿主菌株无法生长的营养源，例如蔗糖、木糖、尿素、乙酰胺或硫酸盐上生长的能力的基因。本领域技术人员也将知道，可选择的标记无需是亲本生物原有的。例如，编码可在亲本生物或宿主生物中起作用的来自异源生物的磷酸丙糖异构酶的基因或DNA序列可用作可选择的标记。

[0086] 实施例2

[0087] 琥珀酸生产菌株中B重复的PCR诊断

[0088] 设计了两种引物(BY296和BY297)用于B重复的PCR(聚合酶链反应)诊断。如图3中阐释，这两个引物正好分别在B重复的连接位点的上游和下游杂交。BY296从大肠埃希杆菌

C_1386基因(B重复3'端附近的基因)的3'端(从终止密码子上游35bp)引发。BY297位于大肠埃希杆菌C_1277基因(B重复的第二个基因)的3'端。

[0089] 根据NEB制造商推荐,PCR在下述中进行:50μl的总工作体积,具有1微升的DNA模板(来自单个菌落或液体细胞培养物)、两种引物(BY296和BY297)、Quick-Load®Taq 2X Master Mix(New England Biolabs,Ipswitch,Massachusetts,USA),以及PCR级水。PCR程序包括在94℃下进行3分钟的初始变性步骤,然后进行30秒94℃、30秒55℃和2分钟68℃的35个循环,然后在68℃下进行10分钟的最终延伸。通过琼脂糖凝胶电泳分析PCR产物。从B重复阳性菌株产生了1788bp片段,阴性对照菌株或未添加模板的对照PCR不产生片段。1788bp PCR产物显示B重复是串联重复,其中两个拷贝是相邻的。PCR产物和基因组DNA测序表明,IS4插入元件存在于B重复的两个拷贝之间的连接处,这又提示了可能出现重复的机制(见图2)。可以使用对照PCR反应,以显示PCR方法和条件恰当地运行,该对照PCR反应使用从重复中未包括的序列产生相似的但可测量的不同大小的片段的适当引物。

[0090] 实施例3

[0091] 关于各种细菌菌株的细节

[0092] 表2中提供了关于本发明中构建的各种细菌菌株的细节。表3中提供了本发明中使用的质粒的列表。表4和表5中提供了关于引物和基因的序列信息。

[0093] 实施例4

[0094] 包括稳定的B重复的菌株的构建

[0095] 如上文实施例1中所描述的其中B重复已被稳定的各种菌株的构建显示在图5和图6中。所有稳定的菌株均通过使用已知的方法构建,该方法为用由λred重组酶系统辅助的线性DNA的整合转化,然后选择同源整合,然后用于正确整合的诊断PCR,和/或代谢进化(Jantama等2008a;Jantama等2008b)。

[0096] 实施例5

[0097] B重复的稳定化

[0098] 使菌株XZ174(其含有B重复和安装在B重复的连接处的可选择的标记cscBAK)的18个个体菌落,以及菌株XZ132(其含有B重复但不含有安装在B重复的连接处的稳定化可选择的标记)的13个个体菌落在含有蔗糖作为碳源的液体培养物中需氧生长约50代。然后如上所述,通过使用引物BY296和BY297的诊断PCR测试每种培养物的B重复的存在,只是将68℃下的延伸时间从两分钟增加到六分钟,以在启动子位点之间容纳包括cscBAK操纵子的额外4千个碱基的DNA。来自稳定的菌株XZ174的所有18个培养物均保留了B重复和可选择的标记,产生5.8千碱基的PCR产物,而对于其中B重复尚未稳定的XZ132菌株,13个培养物中有2个丢失了B重复。因此,在B重复的两个拷贝之间的连接处的可选择的标记确实使B重复的拷贝数稳定化。

[0099] 实施例6

[0100] 不同大肠埃希杆菌菌株的蔗糖发酵

[0101] 将本专利申请中描述的三种不同的大肠埃希杆菌菌株在7升的发酵容器中生长,使用蔗糖作为最低限度培养基中的唯一碳源,并监测它们的发酵性能。这些发酵以及表1中给出的那些发酵的方法的细节先前描述在美国专利9,845,513中。该发酵研究的结果提供在表5中。显示的数字是两次独立发酵的平均值。就滴度、产率和发酵时间而言,含有B重复

的两个菌株(XZ132和XZ174)都比不含B重复的对照菌株SD14表现更好。

[0102] 实施例7

[0103] 导致多于两个拷贝的重复

[0104] 在一些情况下,DNA序列的重复可以导致串联的DNA序列的多于两个拷贝(Fierro,Barredo等1995)。可以通过若干适当的方法中的任何一种来确定拷贝数,例如检查原始Illumina平台数据中的读取频率、直接从原始PacBio数据中观察(如果重复不是大型的)、定量PCR、限制性酶切和琼脂糖凝胶电泳或整个染色体的电泳(对于相对大型的重复)。当有重复的n个拷贝时,如上面实施例1中所述,可以通过在重复的每个相邻对之间安装可选择的标记来实现所有n个拷贝的稳定化。优选将不同的可选择的标记(即,可以独立于先前使用的任何其他标记选择的标记)用于复制的每个相邻对,使得重复的n个拷贝可以优选地用n-1个可选择的标记稳定化。由于可以使许多不同的必需基因成为有条件地需要,因此有许多不同的可能的可选择的标记可用,因此可以使复制的许多拷贝稳定化。例如,可以通过来自大肠埃希杆菌菌株的突变(优选缺失)使pyrF基因失活,该大肠埃希杆菌菌株含有期望被稳定的重复,从而使菌株依赖于添加的尿嘧啶来生长。在第二步骤中,当菌株在不存在尿嘧啶的情况下生长时,在重复的序列的相邻对的连接处安装功能pyrF基因(天然同源物或异源同源物)将使该对重复稳定化。因此,例如,如果重复在大肠埃希杆菌菌株中含有重复的DNA的三个拷贝,则可将tipA基因用作第一对的可选择的标记,并且可将pyrF基因用于第二对。考虑到大多数微生物中大量的生物合成途径基因,可通过延伸和重复该概念来使大量的重复的拷贝稳定化。类似地,TPI、URA3和其他生物合成酵母基因也可以以在酵母中相同的方式使用。理论上,可以发现或构建简单营养缺陷型的任何基因都可以通过这种方式利用。

[0105] 实施例8

[0106] B重复的进一步增强

[0107] 解释由B重复导致的琥珀酸盐滴度增加的一种可能机制是B重复导致甲萘醌生物合成操纵子menFDHB的前四个基因的第二个拷贝,以及所导致的这些基因的表达的增加使甲萘醌的浓度增加,其又增加了细胞在厌氧或微需氧条件下生长的能力。然而,由于插入在menC基因中间的IS4元件,在B重复中没有重复men操纵子menCE的最后两个基因。因此,将menCE的完整的拷贝插入如图4所描绘的B重复的第一拷贝的3'端,以给出men操纵子的完整的第二拷贝,将进一步增加细胞合成甲萘醌的能力。可通过本领域公知的用于通过同源重组将线性DNA片段整合到染色体中的方法来容易地实现这种插入,例如,在第一步骤中,将cat、sacB盒整合在menC与在B重复连接处的IS4之间的边界处,选择对氯霉素的抗性,然后在第二步骤中,通过在含蔗糖的平皿上反向选择sacB基因来整合重构的menCE基因盒(Jantama等,2008a;Jantama等,2008b;(Zhu,Tan等2014);以及美国专利第8,691,539号)。重构的menCE盒将含有适当的侧翼同源物,例如在menC基因的5'端上游约500个碱基对,以及IS4的5'端约500个碱基对。

[0108] 实施例9

[0109] 酵母中的重复的稳定化

[0110] 从马克斯克鲁维酵母的基因组序列可以看出,注释为PDR12的基因作为重复存在。从酿酒酵母的研究,已知Pdr12p是弱酸诱导的多药转运体,是弱有机酸抗性所需要的。酿酒

酵母菌株中PDR12的缺失导致对乳酸抑制的增加的敏感性(Nygard,Mojzita等2014)。因此，可以推断在马克斯克鲁维酵母中PDR12同源物的重复对于其在低pH下对有机酸的特殊抗性可能是重要的(专利申请WO 2014/043591)。为了防止PDR12重复的丢失,因此在经工程改造以通过酵母发酵产生有机酸的菌株中稳定化PDR12重复是合乎逻辑的。类似于以上给出的实施例,这可以通过在PsiI限制性位点(即,PDR12基因的第一拷贝的终止密码子下游约380个碱基对)插入可选择的标记来完成。在所有步骤中,通过诊断PCR确认正确的结构。对于第一步骤,通过用线性缺失盒(SEQ ID NO11)转化,从含有PDR12重复的祖先马克斯克鲁维酵母菌株(比如DMKU3菌株)中缺失URA3编码序列,并如上所述选择对5-氟乳清酸的抗性。在第二步骤中,如下将马克斯克鲁维酵母TPI基因的拷贝插入PRD12重复的重复连接处。第二盒(例如,SEQ ID NO 12)按所列顺序由下述DNA序列组装而成:(1)上游侧翼同源物,其包括从PDR12基因的5'拷贝的终止密码子到PsiI限制性位点的380个碱基对序列的拷贝;(2)TPI基因的拷贝,包括其启动子和终止子;(3)编码不重要功能,并且不含有与宿主菌株("序列X")中的任何序列具有同源性的任何300个碱基对的DNA序列的拷贝;(4)酿酒酵母URA3基因的拷贝,包括其启动子和终止子;(5)序列X的第二平行拷贝,以及(6)如上所述恰好从PRD12基因的5'拷贝下游的PsiI限制性位点的3'的400个碱基对序列的拷贝。通过在缺乏尿嘧啶的最低限度葡萄糖培养基上选择,转化并整合该盒。在第三步骤中,将得到的菌株以每个平皿约10^8个细胞(含有1g/L 5-氟乳清酸和每升24mg尿嘧啶)进行平板接种,以反向选择URA3基因,使得其通过序列X的两个平行的拷贝之间的重组而划掉。在第四步骤中,如下缺失其天然位点处的TPI基因。构建盒,包括按所列顺序的下述DNA序列(SEQ ID NO 13):(1)天然地位于TPI1启动子的恰好上游500个碱基对的拷贝,与在上述第二步骤中使用的携带TPI1基因的启动子没有任何重叠;(2)酿酒酵母URA3基因的拷贝,包括其启动子和终止子;(3)在以上步骤2中使用的天然地位于TPI1终止子恰好下游的500个碱基对的拷贝。然后通过在含有缺乏尿嘧啶的最低限度葡萄糖培养基的平皿上选择,将该盒整合到来自第三步骤4的菌株中。所得菌株将含有TPI1基因,包括其启动子和终止子,从其天然基因座移植到PDR12基因的两个拷贝之间的连接处。TPI1基因的天然的拷贝将被酿酒酵母URA3基因替换。本领域技术人员将知道,以上构造仅仅是如何稳定化PDR12重复的一个实施例,并且许多其他功能上等效的方式也是可能的。

[0111] 实施例10

[0112] 本发明的概括

[0113] 尽管本文公开的实施例使用大肠埃希杆菌或马克斯克鲁维酵母作为宿主生物和琥珀酸作为期望的产物,但是本领域技术人员将知道,同源重组和不等交换在几乎所有经检查的微生物中是公知的现象。因此,本文公开的方法和原理可以应用于许多微生物和许多DNA序列,唯一的限制是这里有用于例如通过转化、转导、转染或交配将外源DNA引入微生物的方法。例如,拜耳连接酵母(Zygosaccharomyces bailii),一种公知的其在低pH下对有机酸具有耐受性的酵母,含有与PDR12高度同源的基因的三个串联的拷贝。这三个串联重复的基因在GenBank登录号HG316458中命名为BN860_04456g、BN860_04478g和BN860_04500g,其公开了拜耳连接酵母菌株CLIB213的基因组序列的支架5。如本文所述,可通过在每对之间安装可选择的标记来稳定化这三个拷贝。

[0114] 以上给出的具体实施例涉及已被工程改造以产生琥珀酸的大肠埃希杆菌的菌株,

并且大肠埃希杆菌的菌株通过称为B重复的功能DNA序列的重复而得到改进。然而，本文公开的原理可以应用于任何其中功能DNA序列的串联的重复可以通过筛选、选择或代谢进化产生的微生物，并通过诊断PCR、DNA测序、凝胶电泳或功能测定(例如，编码基因的酶测定或来自发酵的产物的滴度)发现。一旦发现，可以通过在重复的序列的相邻拷贝之间安装合适的可选择的标记来稳定化这样的重复以如本文公开的维持其扩增的拷贝数。这可在这样的任何微生物中实现：其中可安装包括可选择的标记和与重复的序列之间连接处周围的序列同源(优选与之相同)的侧翼序列的DNA序列，例如通过转化以及所述侧翼序列与重复连接周围的序列之间的同源重组(其导致选择性标记整合在重复的序列的相邻对之间的连接处)。在其中进入的DNA的非同源末端连接比用可选择的标记转化后的同源定向整合更为频繁的微生物中，首先突变(优选缺失)一个或多个负责所述非同源末端连接的基因，例如马克斯克鲁维酵母中的NEJ1、KU70、KU80或LIG4基因中的一个或多个是有用的。在其中用可选择的标记转化后的染色体整合是相对罕见的事件的一些微生物(比如大肠埃希杆菌)中，首先提供增加同源重组的频率的功能(例如来自噬菌体λ的red基因，如pKD46或pCP225上所含有的)是有用的(见表3，Jantama等，2008a和Jantama等，2008b)。应用本发明的优选的微生物是已知可用于商业目的的微生物，比如选自下述属的微生物：埃希杆菌属(Escherichia)、克雷伯杆菌属(Klebsiella)、酵母属(Saccharomyces)、青霉属(Penicillium)、芽孢杆菌属(Bacillus)、伊萨酵母属(Issatchenkia)、毕赤酵母属(Pichia)、假丝酵母属(Candida)、棒状杆菌属(Corynebacterium)、链霉菌属(Streptomyces)、放线菌属(Actinomyces)、梭菌属(Clostridium)、曲霉属(Aspergillus)、木霉属(Trichoderma)、根霉菌属(Rhizopus)、毛霉属(Mucor)、乳酸杆菌属(Lactobacillus)、连接酵母属(Zygosaccharomyces)或克鲁维酵母菌属(Kluyveromyes)。

[0115]

表 1.各种 KJ122 品种和噬菌体抗性衍生物的基因组结构

菌株	A 重复	B 重复	琥珀酸盐滴度(g/L)
KJ122-F475 品种	+	+	89
MYR592 MLPI，LA1 噬菌体^{R}	+	--	68
MYR593 MLPI，LA1 噬菌体^{R}	+	--	49
MYR585 MLPI，LA1 噬菌体^{R}	--	--	58
LU16 Leuna 和 LA1 噬菌体^{R}	--	+	88
MYR585-4E 噬菌体^{R}进化的	--	+	87
KJ122－RY 品种	--	--	57

表 2.本发明中使用的细菌

细菌菌株	基因型/描述
KJ122-F475	大肠埃希杆菌 Crooks，Δ*ldhA*、Δ*adhE*、Δ*ackA*、Δ*focA-pflB*、Δ*mgsA*、Δ*poxB*、Δ*tdcDE*、Δ*citF*、Δ*aspC*、Δ*sfcA*、B 重复$^{+}$(*B*+)
KJ122-RY	大肠埃希杆菌 Crooks，Δ*ldhA*、Δ*adhE*、Δ*ackA*、Δ*focA-pflB*、Δ*mgsA*、Δ*poxB*、Δ*tdcDE*、Δ*citF*、Δ*aspC*、Δ*sfcA*
AC15	KJ122-RY，Δ*ptsHI*、Δ*galP*、*glf*$^{+}$-*glk*$^{+}$(来自运动发酵单胞菌(*Zymomonasmobilis*))
YSS41	进化的 AC15 (两个 *glf*$^{+}$-*glk*$^{+}$mRNA 突变)
MH141	YSS41 *pflD::crr*
FES33	进化的 MH141，*B*+
XZ132	KJ122-F475 *rrsG::cscBAK*
XZ156	FES33 *rrsG::cscBAK*，*B*+
XZ157	
XZ158	
XZ159	

[0116]

XZ162	FES33 *B*+，pCP225
XZ167	FES33 Δ*tpiA*，*B*+
XZ171	FES33 *B*+::*cscBAK*
XZ172	
XZ173	KJ122-F475 *B*+::*cscBAK*
XZ174	
XZ175	XZ167 pKD46，Amp^{R}
XZ192	FES33 Δ*tpiA,B*+::*tpiA*
TG500	XZ132 Δ*tpiA*
TG501	
TG502	
TG503	
TG504	
TG505	TG500 Δ*tpiA*，*B*+::*tpiA*
TG506	
TG507	
TG508	

[0117]

表 3.本发明中使用的质粒

质粒	基因型/描述
pKD46	P_{BAD} *redα redβ redγ*，*amp*^{R}
pCP225	pKD46 + cas9，*amp*^{R}*kan*^{R}
pXZ016	携带用于 *tpiA* 缺失的 gRNA 的质粒
pXZ017	pCL1921 *cscBAK*，*spc*^{R}，一种用于在 B 重复连接位点整合蔗糖利用基因盒 *cscBAK* 而构建的质粒，具有两个分别约 450 个碱基对的侧翼同源物。
pXZ020	pBluescript II KS-*tpiA*，*amp*^{R}，一种用于在 B 重复连接位点整合 *tpiA* 基因而构建的质粒，具有 457bp 的上游侧翼同源物和 447bp 的下游侧翼同源物。

[0118]

[0119]

表 4.序列信息

No.	名称	描述
1	SEQ ID NO 1	运动发酵单胞菌 *glf* 开放阅读框(orf)核苷酸序列
2	SEQ ID NO 2	运动发酵单胞菌 *glk* 开放阅读框(orf)核苷酸序列
3	SEQ ID NO 3	编码蔗糖通透酶-质子同向转运体的大肠埃希杆菌 *W* (ATCC 9637) *cscB* 开放阅读框(orf)，核苷酸序列
4	SEQ ID NO 4	编码转化酶的大肠埃希杆菌 *W* (ATCC 9637) *cscA* 开放阅读框(orf)核苷酸序列
5	SEQ ID NO 5	编码果糖-6 激酶的大肠埃希杆菌 *W* (ATCC 9637) *cscK* 开放阅读框(orf)，核苷酸序列

[0120]

6	SEQ ID NO 6	编码磷酸丙糖异构酶的大肠埃希杆菌 Crooks *tpiA* 开放阅读框(orf)，核苷酸序列
7	SEQ ID NO 7	大肠埃希杆菌 Crooks *tpiA* 缺失序列
8	SEQ ID NO8	B 重复连接位点核苷酸序列；复制起始于大肠埃希杆菌基因组的 EcolC_1387 基因的 139 个碱基。
9	SEQ ID NO 9	大肠埃希杆菌 Crooks TG505 *tpiA* 插入到 B 重复的连接位点。
10	SEQ ID NO 10	大肠埃希杆菌 Crooks XZ13 或 XZ174 *cscBAK* 插入到 B 重复的连接位点
11	SEQ ID NO 11	马克斯克鲁维酵母 *URA3* 缺失盒
12	SEQ ID NO 12	用于稳定化 *PDR12* 复制的马克斯克鲁维酵母
13	SEQ ID NO13	用酿酒酵母 *URA3* 基因替换马克斯克鲁维酵母中的 *TPI1* 基因的盒
14	SEQ ID NO 14	来自 SD14 的 *rrsG* 位点处的 *cscBAK* 盒的引物
15	SEQ ID NO 15	来自 SD14 的 *rrsG* 位点处的 *cscBAK* 盒的引物
16	SEQ ID NO 16	用于制备创建位于 *tpiA* 基因的 ORF 中间的 *tpiA* 缺失的 gRNA 的引物
17	SEQ ID NO 17	用于在染色体上构造 *tpiA* 缺失的合成的 gBlock 片段
18	SEQ ID NO 18	*tpiA* 缺失的诊断引物
19	SEQ ID NO 19	*tpiA* 缺失的诊断引物
20	SEQ ID NO 20	来自 SD14 的 *cscBAK* 盒(*B*:*cscAKB*,4163 bp)的引物
21	SEQ ID NO 21	来自 SD14 的 *cscBAK* 盒(*B*:*cscAKB*,4163 bp)的引物
22	SEQ ID NO 22	来自 SD14 的 *cscBAK* 盒(*B*:*cscAKB*,4123 bp)的引物
23	SEQ ID NO 23	来自 SD14 的 *cscBAK* 盒(*B*:*cscAKB*,4123 bp)的引物
24	SEQ ID NO 24	质粒 pCL1921 的骨架的引物
25	SEQ ID NO 25	质粒 pCL1921 的骨架的引物
26	SEQ ID NO 26	用于在 B 重复连接位点克隆 *cscBAK* 盒的用作上游同源物的合成的 gBlock 片段
27	SEQ ID NO 27	用于在 B 重复连接位点克隆 *cscBAK* 盒的用作下游同源物的合成的 gBlock 片段
28	SEQ ID NO 28	克隆来自 pXZ017 的 B 重复位点的 *cscBAK* 盒的引物
29	SEQ ID NO 29	克隆来自 pXZ017 的 B 重复位点的 *cscBAK* 盒的引物
30	SEQ ID NO 30	克隆来自 KJ122 gDNA 的 B 重复位点的 tpiA 片段(1161 bp)的引物
31	SEQ ID NO 31	克隆来自 KJ122 gDNA 的 B 重复位点的 tpiA 片段(1161 bp)的引物
32	SEQ ID NO 32	构造 pXZ020 的 pBluescript II KS(-)质粒的骨架的引物
33	SEQ ID NO 33	构造 pXZ020 的 pBluescript II KS(-)质粒的骨架的引物
34	SEQ ID NO 34	用于在B重复连接位点克隆 *tpiA* 片段的用作上游同源物的合成的 gBlock 片段
35	SEQ ID NO 35	用于在B重复连接位点克隆 *tpiA* 片段的用作下游同源物的合成的 gBlock 片段

[0121]

36	SEQ ID NO 36	来自质粒 pXZ020 的 *B :: tpiA* 片段的引物
37	SEQ ID NO 37	用 BY297 用于 B 重复的诊断引物
38	SEQ ID NO 38	用 BY296 用于 B 重复的诊断引物

[0122]

表 5.通过各种菌株的蔗糖到琥珀酸的发酵

菌株	存在B重复	由其稳定化B重复:	琥珀酸平均值(g/L)	乙酸平均值(g/L)	琥珀酸产率平均值(g/g)	经过的发酵时间(小时)
XZ132	+	-	80	6.7	0.89	36
XZ174	+	*cscBAK*	80	5.0	0.88	36
SD14	-	-	74	5.1	0.83	45

[0123] 参考文献

[0124] 欧洲专利第1214420号

[0125] 美国专利第7,527,927号

[0126] 美国专利第8,691,539号

[0127] 美国专利第8,871,489号

[0128] 国际专利申请公开号WO2011/063055

[0129] 国际专利申请公开号WO 2011/373671

[0130] 国际专利申请公开号WO 2012/082720

[0131] 国际专利申请公开号WO 2015/013334

[0132] Altschul,S.F.,W.Gish,W.Miller,E.W.Myers and D.J.Lipman(1990)."Basic local alignment search tool."J Mol Biol 215(3):403-410.

[0133] Altschul,S.F.,T.L.Madden,A.A.Schaffer,J.Zhang,Z.Zhang,W.Miller and D.J.Lipman(1997)."Gapped BLAST and PSI-BLAST:a new generation of protein database search programs."Nucleic Acids Res 25(17):3389-3402.

[0134] Elliott,K.T.,L.E.Cuff and E.L.Neidle(2013)."Copy number change: evolving views on gene amplification."Future Microbiol 8(7):887-899.

[0135] Fierro,F.,J.L.Barredo,B.Diez,S.Gutierrez,F.J.Fernandez and J.F.Martin (1995)."The penicillin gene cluster is amplified in tandem repeats linked by conserved hexanucleotide sequences."Proc Natl Acad Sci USA92(13):6200-6204.

[0136] Guest,J.R.(1977)."Menaquinone biosynthesis:mutants of Escherichia coli K-12requiring 2-succinylbenzoate."J Bacteriol 130(3):1038-1046.

[0137] Jantama,K.,M.J.Haupt,S.A.Svoronos,X.Zhang,J.C.Moore,K.T.Shanmugam and L.O.Ingram(2008a)."Combining metabolic engineering and metabolic evolution to develop nonrecombinant strains of Escherichia coli C that produce succinate and malate."Biotechnol Bioeng 99(5):1140-1153.

[0138] Jantama,K.,X.Zhang,J.C.Moore,K.T.Shanmugam,S.A.Svoronos and L.O.Ingram(2008b)."Eliminating side products and increasing succinate yields in engineered strains of Escherichia coli C."Biotechnol Bioeng 101(5):881-893.

[0139] Lopes,T.S.,I.J.de Wijs,S.I.Steenhauer,J.Verbakel and R.J.Planta (1996)."Factors affecting the mitotic stability of high-copy-number integration into the ribosomal DNAof Saccharomyces cerevisiae."Yeast 12(5):

467-477.

[0140] Nygard,Y.,D.Mojzita,M.Toivari,M.Penttila,M.G.Wiebe and L.Ruohonen (2014)."The diverse role of Pdr12 in resistance to weak organic acids."Yeast 31(6):219-232.

[0141] Reams,A.B.and J.R.Roth(2015)."Mechanisms of gene duplication and amplification."Cold Spring Harb Perspect Biol 7(2):a016592.

[0142] Tyo,K.E.,P.K.Ajikumar and G.Stephanopoulos(2009)."Stabilized gene duplication enables long-term selection-free heterologous pathway expression."Nat Biotechnol 27(8):760-765.

[0143] Zhu,X.,Z.Tan,H.Xu,J.Chen,J.Tang and X.Zhang(2014)."Metabolic evolution of two reducing equivalent-conserving pathways for high-yield succinate production in Escherichia coli."Metab Eng 24:87-96.

序列表

<110> PTT全球化学股份有限公司

<120> 具有稳定的功能DNA序列的拷贝数的微生物及相关方法

<130> IP40-191262

<160> 34

<170> PatentIn version 3.5

<210> 1

<211> 1422

<212> DNA

<213> 发酵单胞菌

<400> 1

atgagttctg aaagtagtca gggtctagtc acgcgactag ccctaatcgc tgctataggc 60
ggcttgcttt tcggttacga ttcagcggtt atcgctgcaa tcggtacacc ggttgatatc 120
cattttattg cccctcgtca cctgtctgct acggctgcgg cttccctttc tgggatggtc 180
gttgttgctg ttttggtcgg ttgtgttacc ggttctttgc tgtctggctg gattggtatt 240
cgcttcggtc gtcgcggcgg attgttgatg agttccattt gtttcgtcgc cgccggtttt 300
ggtgctgcgt taaccgaaaa attatttgga accggtggtt cggctttaca aatttttgc 360
tttttccggt ttcttgccgg tttaggtatc ggtgtcgttt caaccttgac cccaacctat 420
attgctgaaa ttgctccgcc agacaaacgt ggtcagatgg tttctggtca gcagatggcc 480
attgtgacgg gtgctttaac cggttatatc tttacctggt tactggctca tttcggttct 540
atcgattggg ttaatgccag tggttggtgc tggtctccgg cttcagaagg cctgatcggt 600
attgccttct tattgctgct gttaaccgca ccggatacgc cgcattggtt ggtgatgaag 660
ggacgtcatt ccgaggctag caaaatcctt gctcgtctgg aaccgcaagc cgatcctaat 720
ctgacgattc aaaagattaa agctggcttt gataaagcca tggacaaaag cagcgcaggt 780
ttgtttgctt ttggtatcac cgttgttttt gccggtgtat ccgttgctgc cttccagcag 840
ttagtcggta ttaacgccgt gctgtattat gcaccgcaga tgttccagaa tttaggtttt 900
ggagctgata cggcattatt gcagaccatc tctatcggtg ttgtgaactt catcttcacc 960
atgattgctt cccgtgttgt tgaccgcttc ggccgtaaac ctctgcttat ttggggtgct 1020
ctcggtatgg ctgcaatgat ggctgtttta ggctgctgtt tctggttcaa agtcggtggt 1080
gttttgcctt tggcttctgt gcttctttat attgcagtct ttggtatgtc atggggccct 1140
gtctgctggg ttgttctgtc agaaatgttc ccgagttcca tcaagggcgc agctatgcct 1200
atcgctgtta ccggacaatg gttagctaat atcttggtta acttcctgtt taaggttgcc 1260
gatggttctc cagcattgaa tcagactttc aaccacggtt tctcctatct cgttttcgca 1320
gcattaagta tcttaggtgg cttgattgtt gctcgcttcg tgccggaaac caaaggtcgg 1380
agcctggatg aaatcgagga gatgtggcgc tcccagaagt ag 1422

<210> 2

<211> 975

<212> DNA

<213> 发酵单胞菌

<400> 2

atggaaattg ttgcgattga catcggtgga acgcatgcgc gtttctctat tgcggaagta 60
agcaatggtc gggttctttc tcttggagaa gaaacaactt ttaaaacggc agaacatgct 120
agcttgcagt tagcttggga acgtttcggt gaaaaactgg gtcgtcctct gccacgtgcc 180
gcagctattg catgggctgg cccggttcat ggtgaagttt taaaacttac caataaccct 240
tgggtattaa gaccagctac tctgaatgaa aagctggaca tcgatacgca tgttctgatc 300
aatgacttcg gcgcggttgc ccacgcggtt gcgcatatgg attcttctta tctggatcat 360
atttgtggtc ctgatgaagc gcttcctagc gatggtgtta tcactattct tggtccggga 420
acgggcttgg gtgttgccca tctgttgcgg actgaaggcc gttatttcgt catcgaaact 480
gaaggcggtc atatcgactt tgctccgctt gacagacttg aagacaaaat tctggcacgt 540
ttacgtgaac gtttccgccg cgtttctatc gaacgcatta tttctggccc gggtcttggt 600
aatatctacg aagcactggc tgccattgaa ggcgttccgt tcagcttgct ggatgatatt 660
aaattatggc agatggcttt ggaaggtaaa gacaaccttg ctgaagccgc tttggatcgc 720
ttctgcttga gccttggcgc tatcgctggt gatcttgctt tggcacaggg tgcaaccagt 780
gttgttattg gcggtggtgt cggtcttcgt atcgcttccc atttgccaga atctggtttc 840
cgtcagcgct ttgtttcaaa aggacgcttt gaacgcgtca tgtccaagat tccggttaag 900
ttgattactt atccgcagcc tggactgttg ggtgcggcag ctgcctatgc caacaaatat 960
tctgaagttg aataa 975

<210> 3

<211> 1248

<212> DNA

<213> 大肠埃希杆菌

<400> 3

atggcactga atattccatt cagaaatgcg tactatcgtt ttgcatccag ttactcattt 60
ctcttttta tttcctggtc gctgtggtgg tcgttatacg ctatttggct gaaaggacat 120
ctagggttga cagggacgga attaggtaca ctttattcgg tcaaccagtt taccagcatt 180
ctatttatga tgttctacgg catcgttcag gataaactcg gtctgaagaa accgctcatc 240
tggtgtatga gtttcatcct ggtcttgacc ggaccgttta tgatttacgt ttatgaaccg 300
ttactgcaaa gcaatttttc tgtaggtcta attctggggg cgctattttt tggcttgggg 360
tatctggcgg gatgcggttt gcttgatagc ttcaccgaaa aaatggcgcg aaattttcat 420
ttcgaatatg gaacagcgcg cgcctgggga tcttttggct atgctattgg cgcgttcttt 480
gccggcatat tttttagtat cagtccccat atcaacttct ggttggtctc gctatttggc 540
gctgtattta tgatgatcaa catgcgtttt aaagataagg atcaccagtg cgtagcggca 600
gatgcgggag gggtaaaaaa agaggatttt atcgcagttt tcaaggatcg aaacttctgg 660
gttttcgtca tatttattgt ggggacgtgg tctttctata acatttttga tcaacaactt 720
tttcctgtct tttattcagg tttattcgaa tcacacgatg taggaacgcg cctgtatggt 780
tatctcaact cattccaggt ggtactcgaa gcgctgtgca tggcgattat tcctttcttt 840
gtgaatcggg tagggccaaa aaatgcatta cttatcggag ttgtgattat ggcgttgcgt 900

atcctttcct gcgcgctgtt cgttaacccc tggattattt cattagtgaa gttgttacat 960
gccattgagg ttccactttg tgtcatatcc gtcttcaaat acagcgtggc aaactttgat 1020
aagcgcctgt cgtcgacgat ctttctgatt ggttttcaaa ttgccagttc gcttgggatt 1080
gtgctgcttt caacgccgac tgggatactc tttgaccacg caggctacca gacagttttc 1140
ttcgcaattt cgggtattgt ctgcctgatg ttgctatttg gcattttctt cttgagtaaa 1200
aaacgcgagc aaatagttat ggaaacgcct gtaccttcag caatatag 1248

<210> 4
<211> 1437
<212> DNA
<213> 大肠埃希杆菌
<400> 4

atgttaaccc agtagccaga gtgctccatg ttgcagcaca gccactccgt gggaggcata 60
aagcgacagt tcccgttctt ctggctgcgg atagattcga ctactcatca ccgcttcccc 120
gtcgttaata aatacttcca cggatgatgt atcgataaat atccttaggg cgagcgtgtc 180
acgctgcggg aggggaatac tacggtagcc gtctaaattc tcgtgtgggt aataccgcca 240
caaaacaagt cgctcagatt ggttatcaat atacagccgc attccagtgc cgagctgtaa 300
tccgtaatgt tcggcatcac tgttcttcag cgcccactgc aactgaatct caactgcttg 360
cgcgttttcc tgcaaaacat atttattgct gattgtgcgg ggagagacag attgatgctg 420
ctggcgtaac gactcagctt cgtgtaccgg gcgttgtaga agtttgccat tgctctctga 480
tagctcgcgc gccagcgtca tgcagcctgc ccatccttca cgttttgagg gcattggcga 540
ttcccacata tccatccagc cgataacaat acgccgacca tccttcgcta aaaagctttg 600
tggtgcataa aagtcatgcc cgttatcaag ttcagtaaaa tgcccggatt gtgcaaaaag 660
tcgtcctggc gaccacattc cgggtattac gccactttga aagcgatttc ggtaactgta 720
tccctcggca ttcattccct gcggggaaaa catcagataa tgctgatcgc caaggctgaa 780
aaagtccgga cattcccaca tatagctttc acccgcatca gcgtgggcca gtacgcgatc 840
gaaggtccat tcacgcaacg aactgccgcg ataaagcagg atctgccccg tgttgcctgg 900
atctttcgcc ccgactacca tccaccatgt gtcggcttca cgccacactt taggatcgcg 960
gaagtgcatg attccttctg gtggagtgag gatcacaccc tgtttctcga aatgaatacc 1020
atcccgactg gtagccagac attgtacttc gcgaattgca tcgtcattac ctgcaccatc 1080
gagccagacg tgtccggtgt agataagtga gaggacacca ttgtcatcga cagcactacc 1140
tgaaaaacac ccgtctttgt cattatcgtc tcctggcgct agcgcaatag gctcatgctg 1200
ccagtggatc atatcgtcgc tggtggcatg tccccagtgc attggccccc agtgttcgct 1260
catcggatga tgttgataaa acgcgtgata acgatcgtta aaccagatca ggccgtttgg 1320
atcgttcatc cacccggcag gaggcgcgag gtgaaaatgg ggatagaaag tgttaccccg 1380
gtgctcatga agttttgcta gggcgttttg cgccgcatgc aatcgagatt gcgtcat 1437

<210> 5
<211> 915
<212> DNA
<213> 大肠埃希杆菌

<400> 5

atgtcagcca aagtatgggt tttaggggat gcggtcgtag atctcttgcc agaatcagac 60
gggcgcctac tgccttgtcc tggcggcgcg ccagctaacg ttgcggtggg aatcgccaga 120
ttaggcggaa caagtgggtt tataggtcgg gtgggggatg atccttttgg tgcgttaatg 180
caaagaacgc tgctaactga gggagtcgat atcacgtatc tgaagcaaga tgaatggcac 240
cggacatcca cggtgcttgt cgatctgaac gatcaagggg aacgttcatt tacgtttatg 300
gtccgcccca gtgccgatct ttttttagag acgacagact tgccctgctg gcgacatggc 360
gaatggttac atctctgttc aattgcgttg tctgccgagc cttcgcgtac cagcgcattt 420
actgcgatga cggcgatccg gcatgccgga ggttttgtca gcttcgatcc taatattcgt 480
gaagatctat ggcaagacga gcatttgctc cgcttgtgtt tgcggcaggc gctacaactg 540
gcggatgtcg tcaagctctc ggaagaagaa tggcgactta tcagtggaaa aacacagaac 600
gatcaggata tatgcgccct ggcaaaagag tatgagatcg ccatgctgtt ggtgactaaa 660
ggtgcagaag gggtggtggt ctgttatcga ggacaagttc accattttgc tggaatgtct 720
gtgaattgtg tcgatagcac gggggcggga gatgcgttcg ttgccgggtt actcacaggt 780
ctgtcctcta cgggattatc tacagatgag agagaaatgc gacgaattat cgatctcgct 840
caacgttgcg gagcgcttgc agtaacggcg aaaggggcaa tgacagcgct gccatgtcga 900
caagaactgg aatag 915

<210> 6
<211> 768
<212> DNA
<213> 大肠埃希杆菌

<400> 6

atgcgacatc ctttagtgat gggtaactgg aaactgaacg gcagccgcca catggttcac 60
gagctggttt ctaacctgcg taaagagctg gcaggtgttg ctggctgtgc ggttgcaatc 120
gcaccaccgg aaatgtacat cgatatggcg aagcgcgaag ctgaaggcag ccacatcatg 180
ctgggtgcgc aaaacgtgga cctgaacctg tccggcgcat tcaccggtga aacttctgcc 240
gctatgctga aagacatcgg cgctcagtac atcatcatcg gccactctga acgtcgtact 300
taccacaaag agtctgacga actgatcgcg aaaaaattcg cggtgctgaa agagcagggc 360
ctgactccgg ttctgtgcat cggtgaaacc gaagctgaaa acgaagcggg caaaactgaa 420
gaagtttgcg cacgtcagat cgacgcggta ctgaaaactc agggtgctgc ggcattcgaa 480
ggtgcggtta tcgcttacga acctgtatgg gcaatcggta ctggcaaatc tgcaactccg 540
gctcaggcac aggctgttca caaattcatc cgtgaccaca tcgctaaagt tgacgctaac 600
atcgctgaac aagtgatcat tcagtacggc ggctctgtaa acgcgtctaa cgctgcagaa 660
ctgtttgctc agccagatat cgacggcgcg ctggttggcg gtgcttctct gaaagctgac 720
gctttcgcag tgatcgttaa agctgcagaa gcggctaaac aggcttaa 768

<210> 7
<211> 500
<212> DNA
<213> 大肠埃希杆菌

<400> 7

cggggaaaaa caaacgttat tacaccgaga cagaaggtgc actgcgttat gttgtcgcgg 60
acaacggcga aaaggggctg accttcgctg ttgaaccgat taagctggcg ctatctgaat 120
cgcttgaagg tttgaataaa tgacaaaaag caaagccttt gtgccgatga atctctatac 180
tgtttcacag acctgctgcc ctgcggggcg gccatcttcc tttattcgct tataagcgtg 240
gagaattaaa gtctgacagg tgccggattt catatccggc acttactttc cttaactctt 300
cgccttaacg caaaatctca cactgatgat cctgaatttc ctcggctgaa gcacggttaa 360
gcgtcagtag atttcgttgt gtcgccagca atacaaatga gttatcactc tgccgtacca 420
tcgccagccc gtagcttccc atatgttccc gcgcctcagg tacttcttct gccagcatta 480
taaatgggct gcgttgtacc 500

<210> 8

<211> 3568

<212> DNA

<213> 大肠埃希杆菌

<400> 8

atgatttatc ctgatgaagc aatgctttac gcaccggttg aatggcacga ctgctccgaa 60
ggtttcgagg acattcgtta tgaaaaatcc accgacggta tcgcaaaaat caccattaat 120
cgtccgcagg tgcgcaatgc cttccgtcct ctgacggtaa aagagatgat ccaggcgctg 180
gcagatgcgc gttatgacga caacatcggc gtgatcattc tgactggtgc aggcgataaa 240
gcgttctgct ccggtggtga ccagaaagtg cgtggtgatt acggcggcta taaagatgat 300
tccggcgtac atcacctgaa tgtgctggac ttccagcgtc agatccgtac ctgtccgaaa 360
ccggttgtcg cgatggtggc tggctactcc atcggcggcg gtcacgttct gcacatgatg 420
tgcgacctga ctatcgcggc agataatgcc atcttcggtc agactggccc gaaagtcggt 480
tccttcgacg gcggctgggg cgcttcctac atggctcgca tcgtcgggca gaaaaaagcg 540
cgtgaaatct ggttcctgtg ccgtcagtac gacgcaaaac aggcgctgga tatgggcctt 600
gtgaacaccg tggtaccgct ggcggatctg gaaaaagaaa ccgtccgttg gtgccgcgaa 660
atgctgcaaa acagcccgat ggcgctgcgc tgcctgaaag ctgcactgaa cgccgactgt 720
gacgggcagg cggggctgca ggagctggcg ggcaacgcca ccatgctgtt ctacatgacg 780
gaagaaggtc aggaaggtcg caacgccttc aaccagaaac gtcagcctga cttcagcaaa 840
ttcaaacgga atccgtaatg cgtagcgcgc aggtataccg ctggcagatc cccatggacg 900
cgggggtggt tctgcgcgac aggcggttaa aaacccgcga cgggctgtat gtttgcctgc 960
gtgaaggcga gcgcgaaggg tgggggggaga tctcccataa gcgctaactt aagggttgtg 1020
gtattacgcc tgatatgatt taacgtgccg atgaattact ctcacgataa ctggtcagca 1080
attctggccc atattggtaa gcccgaagaa ctggatactt cggcacgtaa tgccggggct 1140
ctaacccgcc gccgcgaaat tcgtgatgct gcaactctgc tacgtctggg gctggcttac 1200
ggcccccgggg ggatgtcatt acgtgaagtc actgcatggg ctcagctcca tgacgttgca 1260
acattatctg acgtggctct cctgaagcgg ctgcggaatg ccgccgactg gtttggcata 1320
cttgccgcac aaacacttgc tgtacgcgcc gcagttacgg gttgtacaag cggaaagaga 1380
ttgcgtcttg tcgatggaac agcaatcagt gcgcccgggg gcggcagcgc tgaatggcga 1440

ctacatatgg gatatgatcc tcatacctgt cagttcactg attttgagct aaccgacagc 1500
agagacgctg aacggctgga ccgatttgcg caaacggcag acgagatacg cattgctgac 1560
cggggattcg gttcgcgtcc cgaatgtatc cgctcacttg cttttggaga agctgattat 1620
atcgtccggg ttcactggcg aggattgcgc tggttaactg cagaaggaat gcgctttgac 1680
atgatgggtt ttctgcgcgg gctggattgc ggtaagaacg gtgaaaccac tgtaatgata 1740
ggcaattcag gtaataaaaa agccggagct ccctttccgg cacgtctcat tgccgtatca 1800
cttcctcccg aaaaagcatt aatcagtaaa acccgactgc tcagcgagaa tcgtcgaaaa 1860
ggacgagtag ttcaggcgga aacgctggaa gcagcgggcc atgtgctatt gctaacatca 1920
ttaccggaag atgaatattc agcagagcaa gtggctgatt gttaccgtct gcgatggcaa 1980
attgaactgg cttttaagcg gctcaaaagt ttgctgcacc tggatgcttt gcgtgcaaag 2040
gaacctgaac tcgcgaaagc gtggatattt gctaatctac tcgccgcatt tttaattgac 2100
gacataatcc agccatcgct ggatttcccc cccagaagtg ccggatccga aaagaagaac 2160
taactcgttg tggagaataa caaaaatggt catctggagc ttacaggtgg ccattcgtgg 2220
gacagtatcc ctgacagcct acaaaacgca attgaagaac gcgaggcatc gtcttaacga 2280
ggcaccgagg cgtcgcattc ttcagatggt tcaaccctta agttagcgct tatgggaatt 2340
atccccggct tttttatgta tggtcttaca gcaccagtgc tgcgattgac gcagacagca 2400
cactcaccag ggtagagccg taaaccagct tcagaccgaa gcgagaaacc acgttacctt 2460
gctcttcatt caggccttta actgcacctg cgataatccc gattgaagag aagttagcga 2520
aggaaaccag gaacacagag atgatgcctt cagcacgcgg agagagcgtg gaagcaattt 2580
tctgcagatc catcatcgca acgaactcgt tggaaaccag tttggtcgcc atgatactgc 2640
ccacttgcag tgcttcactg gaaggaacac ccatcaccca tgcaatcgga tagaagatgt 2700
agcccaggat gccctggaag gagatgctgt agccaaacca gccagtaacg gtggcaaaca 2760
gtgcgttcag cgcggcgatc agggcgataa agccaatcag catcgcggca acgataatgg 2820
caactttgaa acctgccaga atgtattcac ccagcatttc gaagaagctc tgaccttcgt 2880
gcaggttgga catctggatg ttttcttcac tggcatcaac acggtaagga ttgatcagcg 2940
acagcacgat aaaggtgctg aacatgttca gtaccagcgc agcaacgacg tatttcggtt 3000
ccagcatggt catgtatgca ccaacgatgg acatcgacac ggtggacatt gccgtggcag 3060
ccatggtgta catacgatta cgggagattt tgccgaggat atctttatag gcaataaagt 3120
tttcagactg acccagaatc agggagctga cggcgttaaa ggattccagt ttgcccatgc 3180
cgttgacttt ggagagcagg aaaccaattg cgcggatgat caccggcaac acgcgaatgt 3240
gctggagaat accgatcagt gcagagataa agacgattgg gcacagcact ttcaggaaga 3300
agaatgccag gccttgatca ttcatgctac caaagacgaa gttagtccct tcgttggcaa 3360
atccgagcag tttttcgaac atttcggaga agcctttcac gaagcctaaa ccaacgtcgg 3420
agttcaggaa gaaccacgcc agtaacactt cgataacaag cagttgaata acataacgga 3480
tacgaatttt tttgcggtcg ctgcttacca gcagtgcgag aatcgcaaca acggcaagtg 3540
ccagtacaaa atgaaggacg cggtccat 3568

<210> 9

<211> 1600

<212> DNA

<213> 大肠埃希杆菌

<400> 9

caacgccttc aaccagaaac gtcagcctga cttcagcaaa ttcaaacgga atccgtaatg 60
cgtagcgcgc aggtataccg ctggcagatc cccatggacg cgggggtggt tctgcgcgac 120
aggcggttaa aaacccgcga cgggctgtat gtttgcctgc gtgaaggcga gcgcgaaggg 180
tgggggggaga tctcccataa gcgctaactt aagggttgtg gaaaaggggc tgaccttcgc 240
tgttgaaccg attaagctgg cgctatctga atcgcttgaa ggtttgaata aatgacaaaa 300
agcaaagcct ttgtgccgat gaatctctat actgtttcac agacctgctg ccctgcgggg 360
cggccatctt cctttattcg cttataagcg tggagaatta aaatgcgaca tcctttagtg 420
atgggtaact ggaaactgaa cggcagccgc cacatggttc acgagctggt ttctaacctg 480
cgtaaagagc tggcaggtgt tgctggctgt gcggttgcaa tcgcaccacc ggaaatgtac 540
atcgatatgg cgaagcgcga agctgaaggc agccacatca tgctgggtgc gcaaaacgtg 600
gacctgaacc tgtccggcgc attcaccggt gaaacttctg ccgctatgct gaaagacatc 660
ggcgctcagt acatcatcat cggccactct gaacgtcgta cttaccacaa agagtctgac 720
gaactgatcg cgaaaaaatt cgcggtgctg aaagagcagg gcctgactcc ggttctgtgc 780
atcggtgaaa ccgaagctga aaacgaagcg ggcaaaactg aagaagtttg cgcacgtcag 840
atcgacgcgg tactgaaaac tcagggtgct gcggcattcg aaggtgcggt tatcgcttac 900
gaacctgtat gggcaatcgg tactggcaaa tctgcaactc cggctcaggc acaggctgtt 960
cacaaattca tccgtgacca catcgctaaa gttgacgcta acatcgctga acaagtgatc 1020
attcagtacg gcggctctgt aaacgcgtct aacgctgcag aactgtttgc tcagccagat 1080
atcgacggcg cgctggttgg cggtgcttct ctgaaagctg acgctttcgc agtgatcgtt 1140
aaagctgcag aagcggctaa acaggcttaa gtctgacagg tgccggattt catatccggc 1200
acttactttc cttaactctt cgccttaacg caaaatctca cactgatgat cctgaatttc 1260
ctcggctgaa gcacggttaa gcgtcagtag atttcgttgt gtcgccagca atacaaatga 1320
gttatcactc tgccgtacca tcgccagccc gtagcttccc attattacgc ctgatatgat 1380
ttaacgtgcc gatgaattac tctcacgata actggtcagc aattctggcc catattggta 1440
agcccgaaga actggatact tcggcacgta atgccggggc tctaacccgc cgccgcgaaa 1500
ttcgtgatgc tgcaactctg ctacgtctgg ggctggctta cggccccggg gggatgtcat 1560
tacgtgaagt cactgcatgg gctcagctcc atgacgttgc 1600

<210> 10

<211> 4500

<212> DNA

<213> 大肠埃希杆菌

<400> 10

caacgccttc aaccagaaac gtcagcctga cttcagcaaa ttcaaacgga atccgtaatg 60
cgtagcgcgc aggtataccg ctggcagatc cccatggacg cgggggtggt tctgcgcgac 120
aggcggttaa aaacccgcga cgggctgtat gtttgcctgc gtgaaggcga gcgcgaaggg 180
tgggggggaga tctcccataa gcgctaactt aagggttgtg gaataactcc ctataatgcg 240
ccaccactga tccgttgttc cacctgatat tatgttaacc cagtagccag agtgctccat 300

gttgcagcac agccactccg tgggaggcat aaagcgacag ttcccgttct tctggctgcg 360
gatagattcg actactcatc accgcttccc cgtcgttaat aaatacttcc acggatgatg 420
tatcgataaa tatccttagg gcgagcgtgt cacgctgcgg gaggggaata ctacggtagc 480
cgtctaaatt ctcgtgtggg taataccgcc acaaaacaag tcgctcagat tggttatcaa 540
tatacagccg cattccagtg ccgagctgta atccgtaatg ttcggcatca ctgttcttca 600
gcgcccactg caactgaatc tcaactgctt gcgcgttttc ctgcaaaaca tatttattgc 660
tgattgtgcg gggagagaca gattgatgct gctggcgtaa cgactcagct tcgtgtaccg 720
ggcgttgtag aagtttgcca ttgctctctg atagctcgcg cgccagcgtc atgcagcctg 780
cccatccttc acgttttgag ggcattggcg attcccacat atccatccag ccgataacaa 840
tacgccgacc atccttcgct aaaaagcttt gtggtgcata aaagtcatgc ccgttatcaa 900
gttcagtaaa atgcccggat tgtgcaaaaa gtcgtcctgg cgaccacatt ccgggtatta 960
cgccactttg aaagcgattt cggtaactgt atccctcggc attcattccc tgcggggaaa 1020
acatcagata atgctgatcg ccaaggctga aaaagtccgg acattcccac atatagcttt 1080
cacccgcatc agcgtgggcc agtacgcgat cgaaggtcca ttcacgcaac gaactgccgc 1140
gataaagcag gatctgcccc gtgttgcctg gatctttcgc cccgactacc atccaccatg 1200
tgtcggcttc acgccacact ttaggatcgc ggaagtgcat gattccttct ggtggagtga 1260
ggatcacacc ctgtttctcg aaatgaatac catcccgact ggtagccaga cattgtactt 1320
cgcgaattgc atcgtcatta cctgcaccat cgagccagac gtgtccggtg tagataagtg 1380
agaggacacc attgtcatcg acagcactac ctgaaaaaca cccgtctttg tcattatcgt 1440
ctcctggcgc tagcgcaata ggctcatgct gccagtggat catatcgtcg ctggtggcat 1500
gtccccagtg cattggcccc cagtgttcgc tcatcggatg atgttgataa aacgcgtgat 1560
aacgatcgtt aaaccagatc aggccgtttg gatcgttcat ccacccggca ggaggcgcga 1620
ggtgaaaatg gggatagaaa gtgttacccc ggtgctcatg aagttttgct agggcgtttt 1680
gcgccgcatg caatcgagat tgcgtcattt taatcatcct ggttaagcaa atttggtgaa 1740
ttgttaacgt taacttttat aaaaataaag tcccttactt tcataaatgc gatgaatatc 1800
acaaatgtta acgttaacta tgacgttttg tgatcgaata tgcatgtttt agtaaatcca 1860
tgacgatttt gcgaaaaaga ggtttatcac tatgcgtaac tcagatgaat ttaagggaaa 1920
aaaatgtcag ccaaagtatg ggttttaggg gatgcggtcg tagatctctt gccagaatca 1980
gacgggcgcc tactgccttg tcctggcggc gcgccagcta acgttgcggt gggaatcgcc 2040
agattaggcg gaacaagtgg gtttataggt cgggtggggg atgatccttt tggtgcgtta 2100
atgcaaagaa cgctgctaac tgagggagtc gatatcacgt atctgaagca agatgaatgg 2160
caccggacat ccacggtgct tgtcgatctg aacgatcaag gggaacgttc atttacgttt 2220
atggtccgcc ccagtgccga tctttttta gagacgacag acttgccctg ctggcgacat 2280
ggcgaatggt tacatctctg ttcaattgcg ttgtctgccg agccttcgcg taccagcgca 2340
tttactgcga tgacggcgat ccggcatgcc ggaggttttg tcagcttcga tcctaatatt 2400
cgtgaagatc tatggcaaga cgagcatttg ctccgcttgt gtttgcggca ggcgctacaa 2460
ctggcggatg tcgtcaagct ctcggaagaa gaatggcgac ttatcagtgg aaaaacacag 2520
aacgatcagg atatatgcgc cctggcaaaa gagtatgaga tcgccatgct gttggtgact 2580
aaaggtgcag aaggggtggt ggtctgttat cgaggacaag ttcaccattt tgctggaatg 2640

tctgtgaatt gtgtcgatag cacgggggcg ggagatgcgt tcgttgccgg gttactcaca 2700
ggtctgtcct ctacgggatt atctacagat gagagagaaa tgcgacgaat tatcgatctc 2760
gctcaacgtt gcggagcgct tgcagtaacg gcgaaagggg caatgacagc gctgccatgt 2820
cgacaagaac tggaatagtg agaagtaaac ggcgaagtcg ctcttatctc taaataggac 2880
gtgaattttt taacgacagg caggtaatta tggcactgaa tattccattc agaaatgcgt 2940
actatcgttt tgcatccagt tactcatttc tctttttat ttcctggtcg ctgtggtggt 3000
cgttatacgc tatttggctg aaaggacatc tagggttgac agggacggaa ttaggtacac 3060
tttattcggt caaccagttt accagcattc tatttatgat gttctacggc atcgttcagg 3120
ataaactcgg tctgaagaaa ccgctcatct ggtgtatgag tttcatcctg gtcttgaccg 3180
gaccgtttat gatttacgtt tatgaaccgt tactgcaaag caatttttct gtaggtctaa 3240
ttctgggggc gctatttttt ggcttggggt atctggcggg atgcggtttg cttgatagct 3300
tcaccgaaaa aatggcgcga aattttcatt tcgaatatgg aacagcgcgc gcctggggat 3360
cttttggcta tgctattggc gcgttctttg ccggcatatt ttttagtatc agtccccata 3420
tcaacttctg gttggtctcg ctatttggcg ctgtatttat gatgatcaac atgcgtttta 3480
aagataagga tcaccagtgc gtagcggcag atgcgggagg ggtaaaaaaa gaggatttta 3540
tcgcagtttt caaggatcga aacttctggg ttttcgtcat atttattgtg gggacgtggt 3600
ctttctataa cattttgat caacaacttt ttcctgtctt ttattcaggt ttattcgaat 3660
cacacgatgt aggaacgcgc ctgtatggtt atctcaactc attccaggtg gtactcgaag 3720
cgctgtgcat ggcgattatt cctttctttg tgaatcgggt agggccaaaa aatgcattac 3780
ttatcggagt tgtgattatg gcgttgcgta tcctttcctg cgcgctgttc gttaacccct 3840
ggattatttc attagtgaag ttgttacatg ccattgaggt tccactttgt gtcatatccg 3900
tcttcaaata cagcgtggca aactttgata agcgcctgtc gtcgacgatc tttctgattg 3960
gttttcaaat tgccagttcg cttgggattg tgctgctttc aacgccgact gggatactct 4020
ttgaccacgc aggctaccag acagttttct tcgcaatttc gggtattgtc tgcctgatgt 4080
tgctatttgg cattttcttc ttgagtaaaa aacgcgagca aatagttatg gaaacgcctg 4140
taccttcagc aatatagacg taaacttttt ccggttgttg tcgatagctc tatatccctc 4200
aaccggaaaa taataatagt aaaatgctta gccctgctaa taatcgccta atccaaacgc 4260
ctgacacgga acaacggcaa acactattac gcctgatatg atttaacgtg ccgatgaatt 4320
actctcacga taactggtca gcaattctgg cccatattgg taagcccgaa gaactggata 4380
cttcggcacg taatgccggg gctctaaccc gccgccgcga aattcgtgat gctgcaactc 4440
tgctacgtct ggggctggct tacggccccg gggggatgtc attacgtgaa gtcactgcat 4500

<210> 11
<211> 807
<212> DNA
<213> 马克斯克鲁维酵母菌
<400> 11

tatacctcaa tcaaaactga aattaggtgc ctgtcacggc tcttttttta ctgtacctgt 60
gacttccttt cttatttcca aggatgctca tcacaatacg cttctagatc tattatgcat 120
tataattaat agttgtagct acaaaaggta aaagaaagtc cggggcaggc aacaatagaa 180

atcggcaaaa aaaactacag aaatactaag agcttcttcc ccattcagtc atcgcatttc 240
gaaacaagag gggaatggct ctggctaggg aactaaccac catcgcctga ctctatgcac 300
taaccacgtg actacatata tgtgatcgtt tttaacattt ttcaaaggct gtgtgtctgg 360
ctgtttccat taattttcac tgattaagca gtcatattga atctgagctc atcaccaaca 420
agaaatacta ccgtaaaagt gtaaaagttc gtttaaatca tttgtaaact ggaacagcaa 480
gaggaagtat catcagctag ccccataaac taatcaaagg aggatgtaag agttctccga 540
gaacaagcag aggttcgagt gtactcggat cagaagttac aagttgatcg tttatatata 600
aactatacag agatgttaga gtgtaatggc attgcgcaca ttgtatacgc tacaagttta 660
gtcacgtgct agaagctgtt ttttgcaccg aaaatttttt ttttttttt ttttgtttt 720
ttggtgaagt acattatgtg aaatttcaca accaaagaaa aagagtttaa tacaagtgcg 780
aagaaccaaa ccttgcttct tagtcca 807

<210> 12
<211> 4614
<212> DNA
<213> 马克斯克鲁维酵母菌
<400> 12

tagtttttgt tcttccttct tttctgccaa tgaagcacta gtcatccttt atactgaaat 60
tcaattcagt ttttgtagaa tatgtcatac caatatccta tgactcttta caatatgttt 120
atacatccat taatttaacc ttaattgtta ttatcaattt tttttttca tttgacctaa 180
tctgtctttt tggtacctta ataccctatg gtgccttatt cctgatatct gacatcaatc 240
tgaaaatttc aaaatcaagc attgctaaga agaacaaaaa aaaagtcaca agtaccgttc 300
acagggtttg gcaaaagaca agatagtatg ccaatgccat gtaccgttcg gtaagaacaa 360
agtttgaact aagaaatgct cgaattataa tcaggccgcc tccccccatt cgcttctctc 420
tgccatatgt cctttccgtc ggccacgccc tgggcccacg ctccagcttc cacttttttgg 480
tgaacatttt tttttttccc gggatccgaa actgcagctt cccaaaaaac cctgcagcac 540
agcagcctcc agcctccact accaggcccc acccagcttc ccctatcccg ccccccaaac 600
aaactggcca aaaaaaaaat aaaaaagcgt ataaaacaca aaaaaaggcg tgatctcagc 660
ttccactatc attactggat gccccgctgc ctacggctac ggctacaccc attggccatt 720
ccatccatca agccctcggg tataccatac cacacaacct gattcagttc aaccgatcca 780
cactgtagta acacacacta tagactagcc tataccatac cataccatac catccatata 840
cacacacata cacacacata cacacataca tacaccatac catacgacgc ccatattcct 900
ctctctgcaa ccggaggcag tttcgagttt cgagttgcga gtttatatgc cttgactgac 960
ttcaagcagt tgttggtttg actttctgaa atttttgttg ctgcgctgtg ctttcatgcg 1020
tttctttgat tctgtgtata taagagtgga cattgtaggt atatacgaat gaaacagagt 1080
gttgtttgat gaaaggtttt ccgatttcgt tttagtatgg attggattat ttcataaccc 1140
aaggacttac aaaggactta agatacacat acacacatat ataacctatc agacatggct 1200
agaacattct tcgttggtgg taacttcaag atgaacggta ccaaggcttc cattaaggaa 1260
atcgttgaga gattgaactc tgcttcgatt ccatccaacg tggaagttgt gattgctcct 1320
ccagctgcct acttggacca cgctgtctct ttgaacaaga aggctcaagt cagtgttgct 1380

gcccaaaacg catacttgaa ggcttccggt gctttcactg gtgaaaactc tgtggagcaa 1440
atcaaggatg ttggtgccga atgggttatc ttgggtcact ccgaaagaag aacgtacttc 1500
cacgaaaccg atgaattcat tgctgacaag accaagttcg ctttggacag cggtgtcaag 1560
gttatcctat gtattggtga aaccttggaa gaaaagcaaa agggtatcac tttggaagtt 1620
gtccaaagac aattgcaagc tgttttggac aaggtccaag actggaccaa cgttgttgtt 1680
gcctacgaac cagtctgggc tattggtacc ggtttggctg ctacctctga cgatgctcaa 1740
gacatccacc actccatcag agaattcttg gccaagaagt tgaacaagga caccgctgaa 1800
aagatcagaa tcctatacgg tggttccgct aacggtaaga acgctgtcac cttcaaggac 1860
aaggccgacg ttgacggttt cttggttggt ggtgcttctt tgaagccaga attcgttgac 1920
atcatcaact ccagagtcta aattaaatta ggttctagtc caaatacgaa atattaaagg 1980
aaaaaaaaac aataaataaa taaagcctat aaagctacga tgaaatagag agtgcttttg 2040
ttttggaaaa tttttgaaat gaatttaacg gctgtatgag cacgcgcgat aatgtagtgt 2100
tgttactata tgatattgta tacttatatg tagcagtaag aacccgctta tcccaataac 2160
gaaataaaaa cgaataacaa taatttcaaa tgtttatttg cattatttga aactagggaa 2220
gacaagcaac gaaacgtttt tgaaaatttt gagtattttc aataaatttg tagaggactc 2280
agatattgaa aaaaagctac agcaattaat acttgataag aagagtattg agaagggcaa 2340
cggttcatca tctcatggat ctgcacatga acaaacacca gagtcaaacg acgttgaaat 2400
tgaggctact gcgccaattg atgacaatac agacgatgat aacaaaccga agttatctga 2460
tgtagaaaag gattaaagat gctaagagat agtgatgata tttcataaat aatgtaattc 2520
tatatatgtt aattaccttt tttgcgaggc atatttatgg tgaaggataa gttttgacca 2580
tcaaagaagg ttaatgtggc tgtggtttca gggtccataa agcttttcaa ttcatctttt 2640
tttttttgt tctttttttt gattccggtt tctttgaaat ttttttgatt cggtaatctc 2700
cgagcagaag gaagaacgaa ggaaggagca cagacttaga ttggtatata tacgcatatg 2760
tggtgttgaa gaaacatgaa attgcccagt attcttaacc caactgcaca gaacaaaaac 2820
ctgcaggaaa cgaagataaa tcatgtcgaa agctacatat aaggaacgtg ctgctactca 2880
tcctagtcct gttgctgcca agctatttaa tatcatgcac gaaaagcaaa caaacttgtg 2940
tgcttcattg gatgttcgta ccaccaagga attactggag ttagttgaag cattaggtcc 3000
caaaatttgt ttactaaaaa cacatgtgga tatcttgact gatttttcca tggagggcac 3060
agttaagccg ctaaaggcat tatccgccaa gtacaatttt ttactcttcg aagacagaaa 3120
atttgctgac attggtaata cagtcaaatt gcagtactct gcgggtgtat acagaatagc 3180
agaatgggca gacattacga atgcacacgg tgtggtgggc ccaggtattg ttagcggttt 3240
gaagcaggcg gcggaagaag taacaaagga acctagaggc cttttgatgt tagcagaatt 3300
gtcatgcaag ggctccctag ctactggaga atatactaag ggtactgttg acattgcgaa 3360
gagcgacaaa gattttgtta tcggctttat tgctcaaaga gacatgggtg gaagagatga 3420
aggttacgat tggttgatta tgacacccgg tgtgggttta gatgacaagg gagacgcatt 3480
gggtcaacag tatagaaccg tggatgatgt ggtctctaca ggatctgaca ttattattgt 3540
tggaagagga ctatttgcaa agggaaggga tgctaaggta gagggtgaac gttacagaaa 3600
agcaggctgg gaagcatatt tgagaagatg cggccagcaa aactaaaaaa ctgtattata 3660
agtaaatgca tgtatactaa actcacaaat tagagcttca atttaattat atcagttatt 3720

acccgggaat ctcggtcgta atgatttcta taatgacgaa aaaaaaaaaa ttggaaagaa 3780
aaagcttcat ggcctttata aaaaggaact atccaatacc tcgccagaac caagtaacag 3840
tattttacgg ggcacaaatc aagaacaata agacaggact gtaaagtaac aataatttca 3900
aatgtttatt tgcattattt gaaactaggg aagacaagca acgaaacgtt tttgaaaatt 3960
ttgagtattt tcaataaatt tgtagaggac tcagatattg aaaaaaagct acagcaatta 4020
atacttgata agaagagtat tgagaagggc aacggttcat catctcatgg atctgcacat 4080
gaacaaacac cagagtcaaa cgacgttgaa attgaggcta ctgcgccaat tgatgacaat 4140
acagacgatg ataacaaacc gaagttatct gatgtagaaa aggattaatt ataaaagttt 4200
tttttattga aacttaaaac ttaacgctac tggtgtcaaa tcatttttca tcatttttt 4260
ctcggttact aagttagtct atcaaatcag tatcaataag accgcatctg gatcaaacaa 4320
ggattatgtc aagaggaaat tcaagtataa tttacagctt gagttacggg atccggtaaa 4380
gtccattggt gcgatgagaa cgttgtattc actggtgctc cggttattgt agactgaagt 4440
gaagcactac cgcatgagta atgtctcttg tacagaaacg ggttttaagg acgtatcgac 4500
catagcaata aaagcaatta tttgtttaca ttctatcaaa atgtgatttg ttccagtcac 4560
tacagatcct taaatcaaca aaacacatgg cagcgccaac gcgacatatc taga 4614

<210> 13
<211> 2422
<212> DNA
<213> 马克斯克鲁维酵母菌
<400> 13

atctgccgct gatgtttacg accggtgttc cggtttccaa gagtccctat tccgccattt 60
cgagagaagc ggcggcaaga agagagatgt ggtcgcaatt gtcacccacg ggatattctt 120
gcgtgtgttc ttgatgaaat ggtttagatg gacttacgag gagtttgagt cgctcatgaa 180
cgtgccgaac ggaagcgtca tcatcatgga gctggacgaa acgctcgacc gctacgtgtt 240
gcgcaccaag ctccccaaat ggaccacaac ggggtgcgat gatgccgcgg cgggggcgtc 300
aacctcgggc tcgaactctt cttgcgcgca gaaacactgt caatgcgagc ccaccgcgat 360
cgtctgatat cacgtgaccc atacgtcatt accatatgta gcgatttaaa caaaaaaaaa 420
agttgaaaac cgacagcagc cgggtaacaa ttcacagcaa ctctctggct acagctctct 480
ggccacggct ctctggcctc aggcctagat gctaagagat agtgatgata tttcataaat 540
aatgtaattc tatatatgtt aattaccttt tttgcgaggc atatttatgg tgaaggataa 600
gttttgacca tcaaagaagg ttaatgtggc tgtggtttca gggtccataa agcttttcaa 660
ttcatctttt ttttttttgt tctttttttt gattccggtt tctttgaaat ttttttgatt 720
cggtaatctc cgagcagaag gaagaacgaa ggaaggagca cagacttaga ttggtatata 780
tacgcatatg tggtgttgaa gaaacatgaa attgcccagt attcttaacc caactgcaca 840
gaacaaaaac ctgcaggaaa cgaagataaa tcatgtcgaa agctacatat aaggaacgtg 900
ctgctactca tcctagtcct gttgctgcca agctatttaa tatcatgcac gaaaagcaaa 960
caaacttgtg tgcttcattg gatgttcgta ccaccaagga attactggag ttagttgaag 1020
cattaggtcc caaaatttgt ttactaaaaa cacatgtgga tatcttgact gatttttcca 1080
tggagggcac agttaagccg ctaaaggcat tatccgccaa gtacaatttt ttactcttcg 1140

aagacagaaa atttgctgac attggtaata cagtcaaatt gcagtactct gcgggtgtat 1200
acagaatagc agaatgggca gacattacga atgcacacgg tgtggtgggc ccaggtattg 1260
ttagcggttt gaagcaggcg gcggaagaag taacaaagga acctagaggc cttttgatgt 1320
tagcagaatt gtcatgcaag ggctccctag ctactggaga atatactaag ggtactgttg 1380
acattgcgaa gagcgacaaa gattttgtta tcggctttat tgctcaaaga gacatgggtg 1440
gaagagatga aggttacgat tggttgatta tgacacccgg tgtgggttta gatgacaagg 1500
gagacgcatt gggtcaacag tatagaaccg tggatgatgt ggtctctaca ggatctgaca 1560
ttattattgt tggaagagga ctatttgcaa agggaaggga tgctaaggta gagggtgaac 1620
gttacagaaa agcaggctgg gaagcatatt tgagaagatg cggccagcaa aactaaaaaa 1680
ctgtattata agtaaatgca tgtatactaa actcacaaat tagagcttca atttaattat 1740
atcagttatt acccgggaat ctcggtcgta atgatttcta taatgacgaa aaaaaaaaaa 1800
ttggaaagaa aaagcttcat ggcctttata aaaaggaact atccaatacc tcgccagaac 1860
caagtaacag tattttacgg ggcacaaatc aagaacaata agacaggact gtaaagaggc 1920
ctgaacgagc aacattattg tttgattccc gacacaagat agcaatagcg aagatcatta 1980
gttatttgtt gaaacatgta tacgtatata tacataccat ccatcatcac acgagaatgc 2040
cgtaatactt ggcccttgga tacgctggga gtagatccgg ccattgcacc acaaactcgt 2100
ctgccaatgg gtacagaatc ttcccagtta gactcaacct gttggtccgc gccaacgact 2160
ctacatacac aatcttggaa ggctgccata agaacaaatg gtacaatttc agccacacag 2220
taattataca acaggtcccg ggcccattga gcaaagtgag gttcggttta cccctcatag 2280
acttcttgat attatacgta atagtcatcg aagttatcag cgtctcgatg atacttatca 2340
cggaccctgc aaccgaactc ccaacatcac gcgccttctt gaaatgatgg tactctattt 2400
tcacctgtag ctgtttagaa ga 2422

<210> 14
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> 引物序列
<400> 14
aggaacgctt cggtgaactg g 21

<210> 15
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> 引物序列
<400> 15
tggtcttgcg acgttatgcg g 21

<210> 16

<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 引物序列
<400> 16
gactggcgct acaatcttcc ctgactccgg ttctgtgcat gttttagagc tagaaatagc 60
<210> 17
<211> 500
<212> DNA
<213> 人工序列
<220>
<223> 片段序列
<400> 17
cggggaaaaa caaacgttat tacaccgaga cagaaggtgc actgcgttat gttgtcgcgg 60
acaacggcga aaaggggctg accttcgctg ttgaaccgat taagctggcg ctatctgaat 120
cgcttgaagg tttgaataaa tgacaaaaag caaagccttt gtgccgatga atctctatac 180
tgtttcacag acctgctgcc ctgcggggcg gccatcttcc tttattcgct tataagcgtg 240
gagaattaaa gtctgacagg tgccggattt catatccggc acttactttc cttaactctt 300
cgccttaacg caaaatctca cactgatgat cctgaatttc ctcggctgaa gcacggttaa 360
gcgtcagtag atttcgttgt gtcgccagca atacaaatga gttatcactc tgccgtacca 420
tcgccagccc gtagcttccc atatgttccc gcgcctcagg tacttcttct gccagcatta 480
taaatgggct gcgttgtacc 500
<210> 18
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> 引物序列
<400> 18
acactcaccc cattaatgac c 21
<210> 19
<211> 22
<212> DNA
<213> 人工序列
<220>
<223> 引物序列
<400> 19
atctcttgta ttcgtcctga tg 22

<210> 20
<211> 72
<212> DNA
<213> 人工序列
<220>
<223> 引物序列
<400> 20
cgcgaagggt ggggggagat ctcccataag cgctaactta agggttgtgg aataactccc 60
tataatgcgc ca 72
<210> 21
<211> 70
<212> DNA
<213> 人工序列
<220>
<223> 引物序列
<400> 21
gttatcgtga gagtaattca tcggcacgtt aaatcatatc aggcgtaata gtgtttgccg 60
ttgttccgtg 70
<210> 22
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 引物序列
<400> 22
agggttgtgg aataactccc tataatgcgc ca 32
<210> 23
<211> 30
<212> DNA
<213> 人工序列
<220>
<223> 引物序列
<400> 23
aggcgtaata gtgtttgccg ttgttccgtg 30
<210> 24
<211> 30
<212> DNA
<213> 人工序列
<220>

<223> 引物序列
<400> 24
gggatatgat gtgagttata cacagggctg 30
<210> 25
<211> 30
<212> DNA
<213> 人工序列
<220>
<223> 引物序列
<400> 25
catatccagc tcagcttagt aaagccctcg 30
<210> 26
<211> 457
<212> DNA
<213> 人工序列
<220>
<223> 片段序列
<400> 26
actaagctga gctggatatg ggccttgtga acaccgtggt accgctggcg gatctggaaa 60
aagaaaccgt ccgttggtgc cgcgaaatgc tgcaaaacag cccgatggcg ctgcgctgcc 120
tgaaagctgc actgaacgcc gactgtgacg ggcaggcggg gctgcaggag ctggcgggca 180
acgccaccat gctgttctac atgacggaag aaggtcagga aggtcgcaac gccttcaacc 240
agaaacgtca gcctgacttc agcaaattca aacggaatcc gtaatgcgta gcgcgcaggt 300
ataccgctgg cagatcccca tggacgcggg ggtggttctg cgcgacaggc ggttaaaaac 360
ccgcgacggg ctgtatgttt gcctgcgtga aggcgagcgc gaagggtggg gggagatctc 420
ccataagcgc taacttaagg gttgtggaat aactccc 457
<210> 27
<211> 457
<212> DNA
<213> 人工序列
<220>
<223> 片段序列
<400> 27
cggcaaacac tattacgcct gatatgattt aacgtgccga tgaattactc tcacgataac 60
tggtcagcaa ttctggccca tattggtaag cccgaagaac tggatacttc ggcacgtaat 120
gccggggctc taacccgccg ccgcgaaatt cgtgatgctg caactctgct acgtctgggg 180
ctggcttacg gccccggggg gatgtcatta cgtgaagtca ctgcatgggc tcagctccat 240
gacgttgcaa cattatctga cgtggctctc ctgaagcggc tgcggaatgc cgccgactgg 300
tttggcatac ttgccgcaca aacacttgct gtacgcgccg cagttacggg ttgtacaagc 360

ggaaagagat tgcgtcttgt cgatggaaca gcaatcagtg cgcccggggg cggcagcgct 420
gaatggcgac tacatatggg atatgatgtg agttata 457
<210> 28
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 引物序列
<400> 28
gctggatatg ggccttgtga 20
<210> 29
<211> 22
<212> DNA
<213> 人工序列
<220>
<223> 引物序列
<400> 29
atcatatccc atatgtagtc gc 22
<210> 30
<211> 31
<212> DNA
<213> 人工序列
<220>
<223> 引物序列
<400> 30
agggttgtgg aaaaggggct gaccttcgct g 31
<210> 31
<211> 30
<212> DNA
<213> 人工序列
<220>
<223> 引物序列
<400> 31
aggcgtaata atgggaagct acgggctggc 30
<210> 32
<211> 30
<212> DNA
<213> 人工序列
<220>

<223> 引物序列
<400> 32
gactacatat ggtacccagc ttttgttccc 30
<210> 33
<211> 30
<212> DNA
<213> 人工序列
<220>
<223> 引物序列
<400> 33
catatccagc ggagctccaa ttcgccctat 30
<210> 34
<211> 457
<212> DNA
<213> 人工序列
<220>
<223> 片段序列
<400> 34
ttggagctcc gctggatatg ggccttgtga acaccgtggt accgctggcg gatctggaaa 60
aagaaaccgt ccgttggtgc cgcgaaatgc tgcaaaacag cccgatggcg ctgcgctgcc 120
tgaaagctgc actgaacgcc gactgtgacg ggcaggcggg gctgcaggag ctggcgggca 180
acgccaccat gctgttctac atgacggaag aaggtcagga aggtcgcaac gccttcaacc 240
agaaacgtca gcctgacttc agcaaattca aacggaatcc gtaatgcgta gcgcgcaggt 300
ataccgctgg cagatcccca tggacgcggg ggtggttctg cgcgacaggc ggttaaaaac 360
ccgcgacggg ctgtatgttt gcctgcgtga aggcgagcgc gaagggtggg gggagatctc 420
ccataagcgc taacttaagg gttgtggaaa aggggct 457

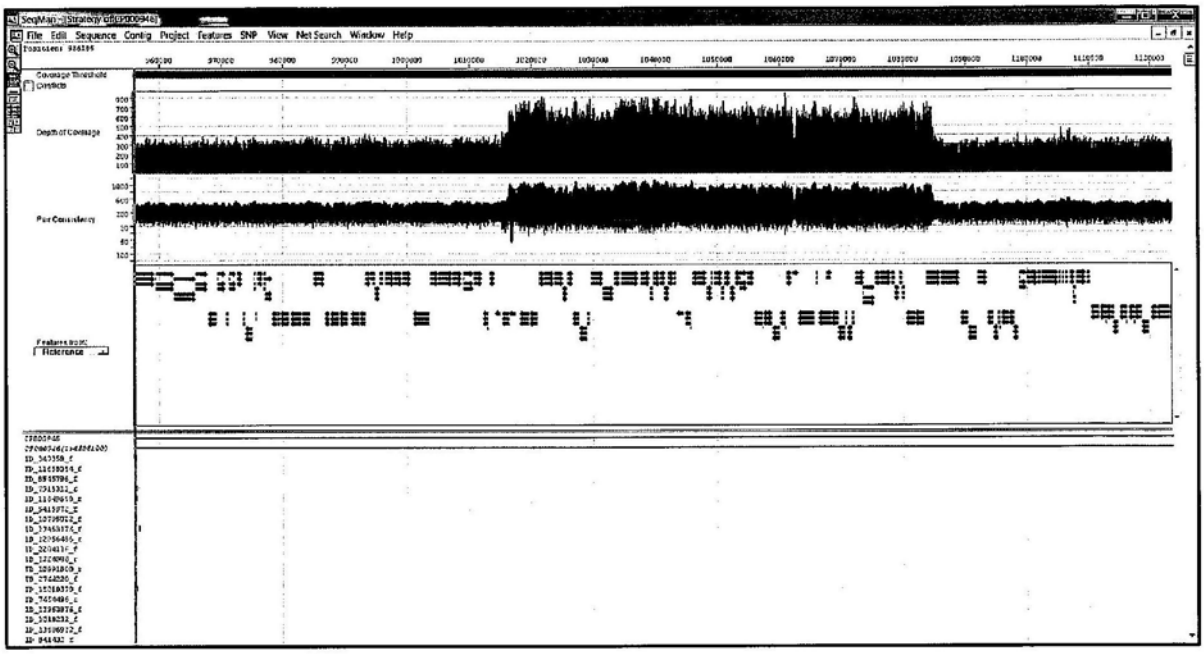


图1

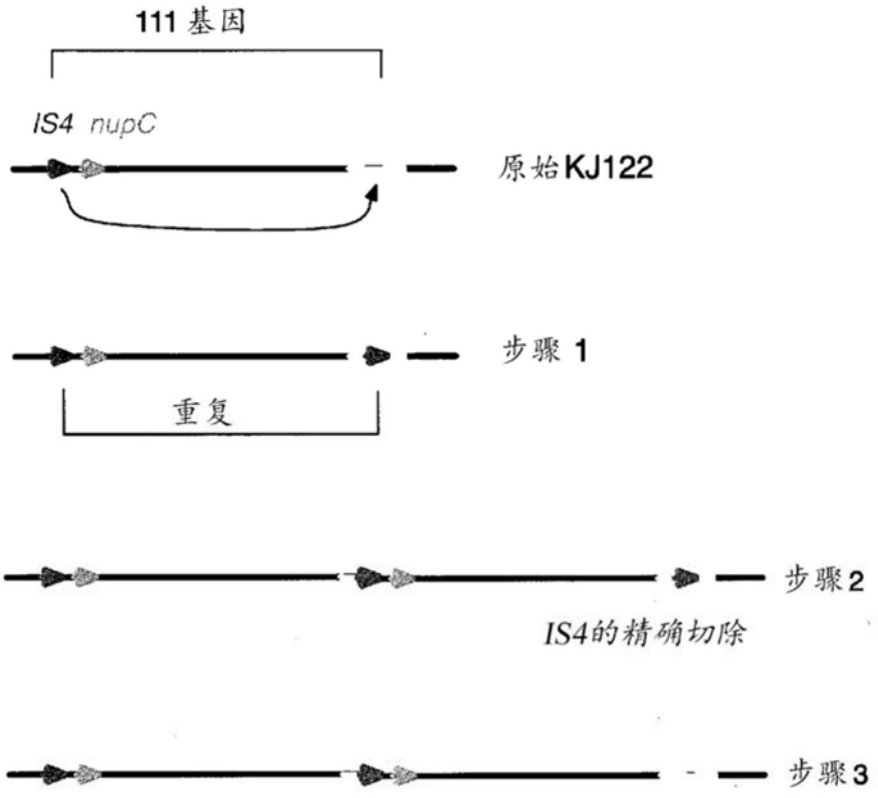


图2

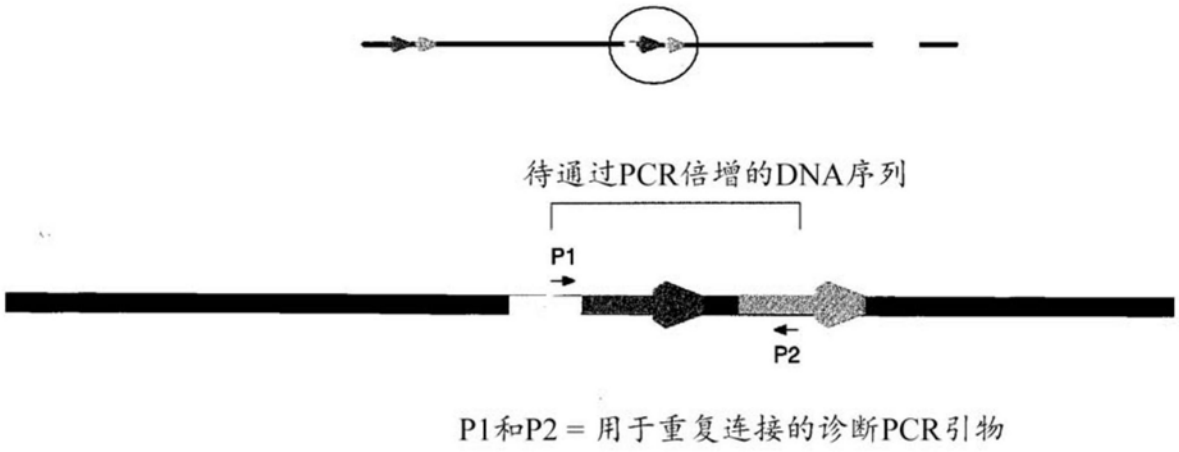

待通过PCR倍增的DNA序列
P1
P2
P1和P2＝用于重复连接的诊断PCR引物

图3

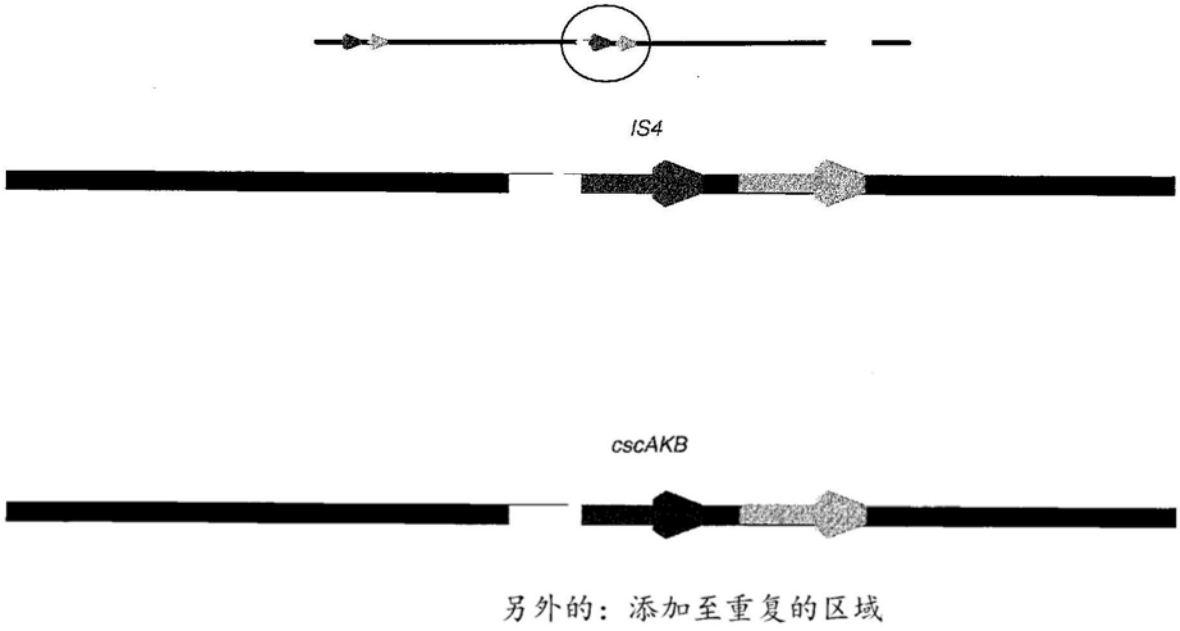

IS4
cscAKB
另外的：添加至重复的区域

图4

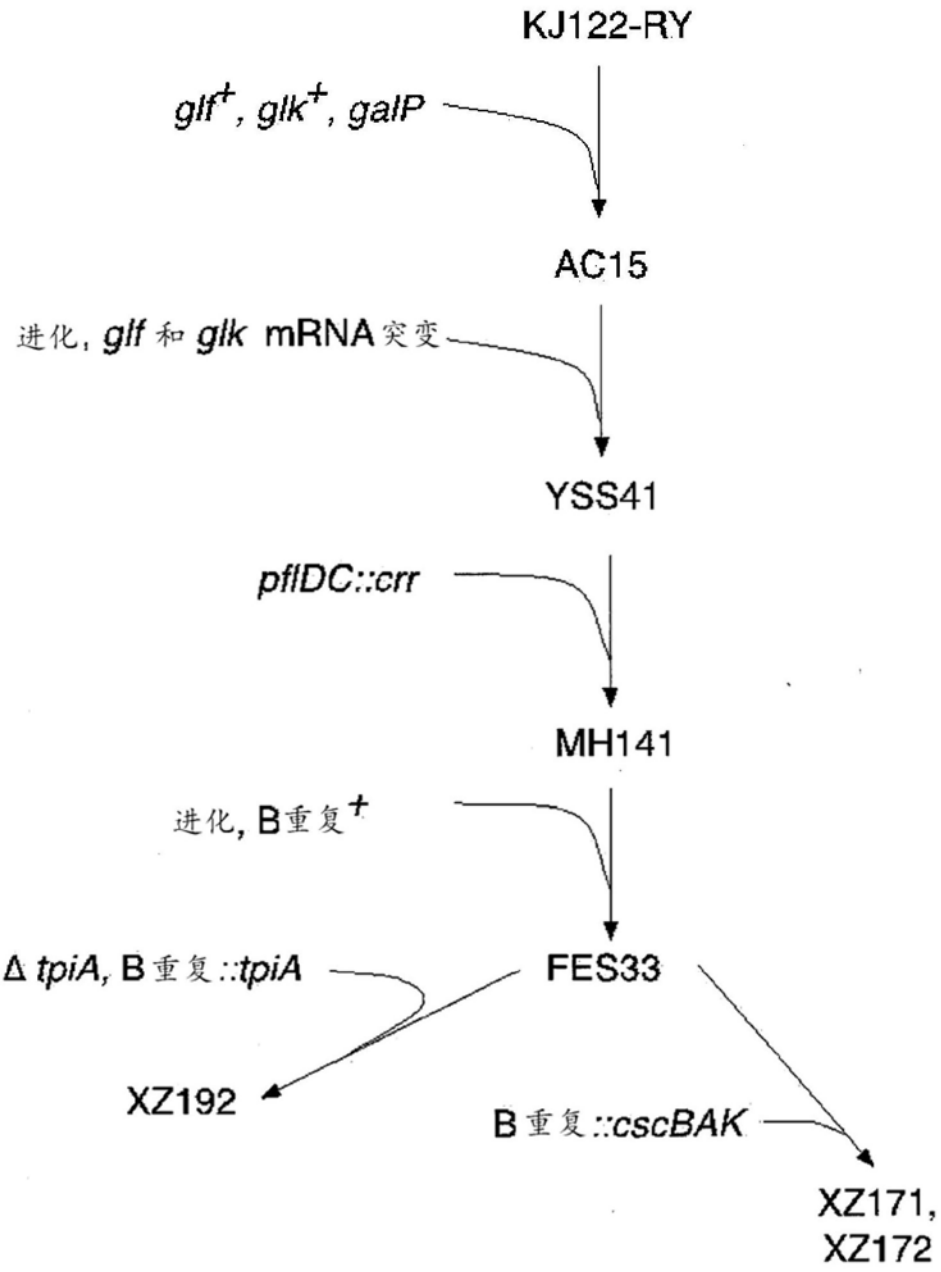

KJ122-RY
glf+, glk+, galP
AC15
进化, glf 和 glk mRNA 突变
YSS41
pflDC::crr
MH141
进化, B重复+
Δ tpiA, B重复::tpiA
FES33
XZ192
B重复::cscBAK
XZ171,
XZ172

图5

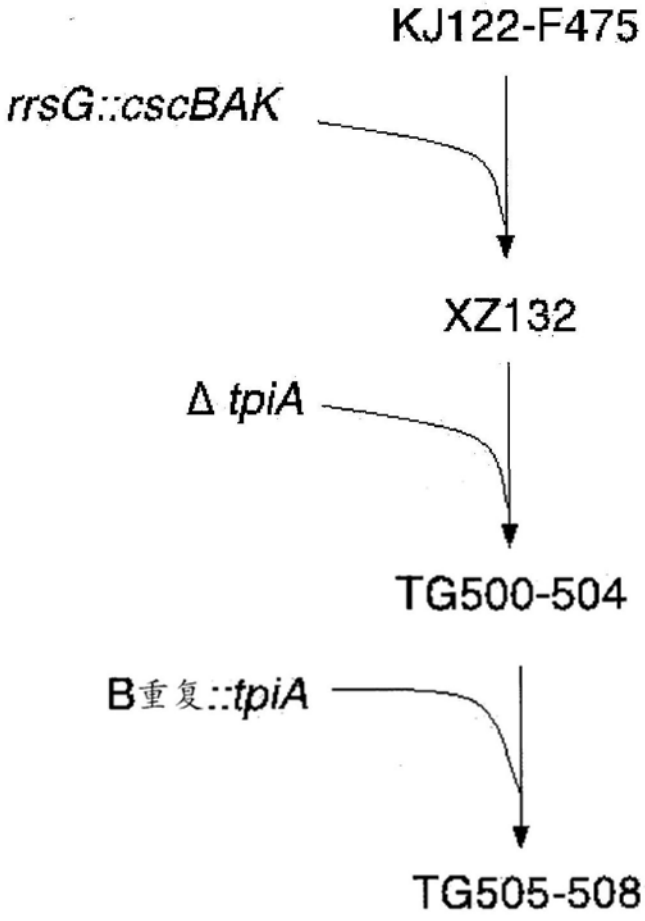

KJ122-F475
rrsG::cscBAK
XZ132
Δ tpiA
TG500-504
B重复::tpiA
TG505-508

图6