



(12) 发明专利申请

(10) 申请公布号 CN 105209624 A

(43) 申请公布日 2015. 12. 30

(21) 申请号 201480028497. 6

C12N 15/86(2006. 01)

(22) 申请日 2014. 03. 14

A01H 5/00(2006. 01)

(30) 优先权数据

61/790, 694 2013. 03. 15 US

(85) PCT国际申请进入国家阶段日

2015. 11. 16

(86) PCT国际申请的申请数据

PCT/US2014/028445 2014. 03. 14

(87) PCT国际申请的公布数据

W02014/144155 EN 2014. 09. 18

(71) 申请人 明尼苏达大学董事会

地址 美国明尼苏达州

(72) 发明人 D·F·沃塔斯 P·阿特金斯

N·J·巴特斯

(74) 专利代理机构 上海专利商标事务所有限公

司 31100

代理人 杨昀 陶家蓉

(51) Int. Cl.

C12N 15/82(2006. 01)

权利要求书2页 说明书10页

序列表32页 附图6页

(54) 发明名称

采用 CRISPR/Cas 系统的植物基因组的工程改造

(57) 摘要

用于采用本文所述的成簇且规律间隔的短回文重复序列 /CRISPR- 相关的 (CRISPR/Cas) 系统进行基因靶向的材料和方法。

1. 一种修饰植物细胞中的基因组物质的方法,所述方法包括:
 - (a) 向所述细胞引入包含 crRNA 和 tracrRNA 或嵌合型 cr/tracrRNA 杂合体的核酸,其中,所述 crRNA 和 tracrRNA 或所述 cr/tracrRNA 杂合体靶向至对所述植物细胞为内源性的序列;和
 - (b) 向所述细胞引入 Cas9 核酸内切酶分子,该 Cas9 核酸内切酶分子诱导所述 crRNA 和 tracrRNA 序列靶向的序列处或附近的双链断裂,或诱导所述 cr/tracrRNA 杂合体靶向的序列处或附近的双链断裂。
2. 如权利要求 1 所述的方法,其特征在于,所述引入步骤包括:向所述植物细胞递送编码 Cas9 核酸内切酶的核酸和编码所述 crRNA 和 tracrRNA 或所述 cr/tracrRNA 杂合体的核酸,并且其中,所述递送通过 DNA 或 RNA 病毒进行。
3. 如权利要求 2 所述方法,其特征在于,所述 DNA 病毒是双生病毒。
4. 如权利要求 2 所述的方法,其特征在于,所述 RNA 病毒是烟草脆裂病毒。
5. 如权利要求 1 所述的方法,其特征在于,所述引入步骤包括:向所述植物细胞递送 T-DNA,所述 T-DNA 包含编码所述 Cas9 核酸内切酶的核酸序列和编码所述 crRNA 和 tracrRNA 或所述 cr/tracrRNA 杂合体的核酸序列,并且其中所述递送通过农杆菌 (*Agrobacterium*) 或剑菌 (*Ensifer*) 进行。
6. 如权利要求 5 所述的方法,其特征在于,编码所述 Cas9 核酸内切酶的核酸序列操作性连接至启动子,所述启动子是组成型、细胞特异型、诱导型或被自杀外显子的可变剪接激活的启动子。
7. 如权利要求 1 所述的方法,其特征在于,所述引入步骤包括:微粒轰击编码 Cas9 的核酸和所述 crRNA 和 tracrRNA 或所述 cr/tracrRNA 杂合体。
8. 如权利要求 1 所述的方法,其特征在于,所述植物细胞来自单子叶植物。
9. 如权利要求 8 所述的方法,其特征在于,所述单子叶植物是小麦、玉米、水稻或狗尾草。
10. 如权利要求 1 所述的方法,其特征在于,所述植物细胞来自双子叶植物。
11. 如权利要求 10 所述的方法,其特征在于,所述双子叶植物是番茄、大豆、烟草、马铃薯、木薯或拟南芥。
12. 如权利要求 1 所述的方法,其特征在于,所述方法还包括:在所述引入步骤之后筛选所述植物细胞,以确定所述 crRNA 和 tracrRNA 或所述 cr/tracrRNA 杂合体靶向的序列处或附近是否已发生双链断裂。
13. 如权利要求 1 所述的方法,其特征在于,所述方法还包括从所述植物细胞再生植物。
14. 如权利要求 13 所述的方法,其特征在于,所述方法还包括:杂交繁育所述植物以获得遗传上所需的植物株系。
15. 一种植物细胞,其包含编码与 SEQ ID NO:12 具有至少 80% 的序列相同性的多肽的核酸。
16. 一种植物细胞,其包含编码如下多肽的核酸:所述多肽包含与 SEQ ID NO:12 的氨基酸 810-872 具有至少 80% 的序列相同性的氨基酸序列。
17. 一种病毒载体,其包含编码 Cas9 多肽的核苷酸序列。

18. 如权利要求 17 所述的病毒载体,其特征在于,所述载体包含编码如下多肽的核苷酸序列:所述多肽包含与 SEQ ID NO:12 具有至少 90%相同性的氨基酸序列。

19. 如权利要求 17 所述的病毒载体,其特征在于,所述病毒是烟草脆裂病毒或双生病毒。

20. 一种 T-DNA,其包含编码如下多肽的核酸序列,所述多肽包含与 SEQ ID NO:12 的氨基酸 810-872 具有至少 80%的序列相同性的氨基酸序列。

21. 一种包含如权利要求 20 所述的 T-DNA 的农杆菌 (Agrobacterium) 菌株。

22. 一种使 Cas 蛋白在植物细胞中表达的方法,所述方法包括:提供包含 T-DNA 的农杆菌 (Agrobacterium) 或剑菌 (Ensifer) 载体,所述 T-DNA 包含编码如下多肽的核酸序列,所述多肽包含与 SEQ ID NO:12 的氨基酸 810-872 具有至少 80%的序列相同性的氨基酸序列,其中,该多肽编码序列操作性连接至启动子;

使所述农杆菌 (Agrobacterium) 或剑菌 (Ensifer) 载体与所述植物细胞接触;和使所述核酸序列在所述植物细胞中表达。

23. 如权利要求 22 所述的方法,其特征在于,所述启动子是诱导型启动子。

24. 如权利要求 23 所述的方法,其特征在于,所述诱导型启动子是雌激素诱导型启动子。

25. 如权利要求 22 所述的方法,其特征在于,所述方法还包括使所述植物细胞与编码向导 RNA 的核酸接触,所述向导 RNA 与 Cas 蛋白相关联。

26. 如权利要求 22 所述的方法,其特征在于,所述植物细胞是原生质体。

采用 CRISPR/Cas 系统的植物基因组的工程改造

[0001] 相关申请的交叉参考

[0002] 本申请要求 2013 年 3 月 15 日提交的美国临时申请序列号 61/790,694 的优先权。

[0003] 有关联邦资助的研究的声明

[0004] 本发明得到政府的支持,在国立卫生研究院 (National Institutes of Health) 授予的 GM 834720 和国家自然科学基金会 (National Science Foundation) 授予的 DBI0923827 的资助下完成。政府对本发明拥有某些权利。

技术领域

[0005] 本发明涉及用于在植物中基因靶向的材料和方法,具体而言,涉及采用 CRISPR/Cas 系统进行基因靶向的方法。

[0006] 背景

[0007] 允许在活细胞中精确修饰 DNA 序列的技术对于基础和应用研究而言均具有价值。精确基因组修饰 - 靶向诱变或基因靶向 (GT) - 依赖靶细胞的 DNA- 修复机制。对于靶向诱变,序列 - 特异性核酸酶 (SSN) 介导的 DNA 双链断裂 (DSB) 通常被易于出错的非同源末端连接 (NHEJ) 通路修复,导致断裂位点处的突变。另一方面,如果供体分子与 SSN 共同递送,后续 DSB 可能会刺激所述断裂位点附近的序列与供体分子上存在的序列之间的同源重组 (HR)。因此,供体分子携带的任何经修饰的序列将被稳定地整合进入基因组 (称为 GT)。意在在植物中进行 GT 的尝试常常受困于极低的 HR 频率。大多数情况下,供体 DNA 分子通过 NHEJ 不合理地整合。该过程的发生与同源“臂”的大小无关;使同源长度增加至约 22kb,导致 GT 的增强不显著 (ThykJaer 等,Plant Mol Biol,35:523-530,1997)。然而,伴随 SSN 引入 DSB 能够通过 HR 大幅提高 GT 频率 (Shukla 等,Nature 459:437-441,2009;和 Townsend 等,Nature 459:442-445,2009)。

发明内容

[0008] 本发明部分基于如下发现:可采用成簇且规律间隔的短回文重复序列 (Clustered Regularly Interspersed Short Palindromic Repeats)/CRISPR-相关的 (CRISPR/Cas) 系统来进行植物基因组工程改造。所述 CRISPR/Cas 系统为在基因组 DNA 中的选定位点处产生修饰提供了相对简单、有效的手段。CRISPR/Cas 系统可用于产生靶向的 DSB 或单链断裂,并且能够用于 (不限于):靶向诱变、基因靶向、基因替换、靶向缺失、靶向倒位、靶向易位、靶向插入,和通过共表达多重靶向 RNA 导向的单细胞中的多重 DSB 进行的多重基因组修饰。该技术可用于提高植物中功能性遗传学研究的速度,以及用于工程改造出具有改善的特点的植物,所述改善特点包括,营养质量增强、抗病和抗应激性提高,以及市场价值化合物的产量提高。

[0009] 一方面,本发明涉及用于对植物细胞中的基因组物质进行修饰的方法。所述方法可包括:(a) 向所述细胞引入一核酸,所述核酸包含 crRNA 和 tracrRNA 或嵌合型 cr/tracrRNA 杂合体,其中所述 crRNA 和 tracrRNA 或所述 cr/tracrRNA 杂合体靶向至对该植

物细胞为内源性的序列；和 (b) 向该细胞引入 Cas9 核酸内切酶分子，其诱导所述 crRNA 和 tracrRNA 序列靶向的序列处或附近的双链断裂，或诱导所述 cr/tracrRNA 杂合体靶向的序列处或附近的双链断裂。所述引入步骤可包括：向所述植物细胞递送编码所述 Cas9 核酸内切酶的核酸，和编码所述 crRNA 和 tracrRNA 或所述 cr/tracrRNA 杂合体的核酸，其中，所述递送通过 DNA 病毒（例如，双生病毒）或 RNA 病毒（例如，烟草脆裂病毒）。所述引入步骤可包括向所述植物细胞递送 T-DNA，所述 T-DNA 包含编码所述 Cas9 核酸内切酶的核酸序列和编码所述 crRNA 和 tracrRNA 或所述 cr/tracrRNA 杂合体的核酸序列，其中所述递送通过农杆菌 (*Agrobacterium*) 或剑菌 (*Ensifer*) 进行。编码所述 Cas9 核酸内切酶的核酸序列可操作地连接至启动子，所述启动子是组成型（例如，花椰菜花叶病毒 35S 启动子）、细胞特异型、诱导型或通过自杀外显子 (suicide exon) 的可变剪接激活的启动子。所述引入步骤可包括：微粒轰击编码 Cas9 的核酸，以及所述 crRNA 和 tracrRNA 或所述 cr/tracrRNA 杂合体。编码所述 Cas9 核酸内切酶的核酸序列能操作性地连接至启动子，所述启动子是组成型、细胞特异型、诱导型或通过自杀外显子的可变剪接激活的启动子。所述植物细胞可来自单子叶植物（例如，小麦、玉米、水稻或狗尾草）或来自双子叶植物（例如，番茄、大豆、烟草、马铃薯、木薯或拟南芥）。所述方法还可包括：在引入步骤之后筛选所述植物细胞，以确定在所述 crRNA 和 tracrRNA 或所述 cr/tracrRNA 杂合体靶向的序列处或附近是否已发生双链断裂。所述方法还可包括从所述植物细胞再生植物，并且在一些实施方式中，所述方法可包括：杂交繁育所述植物以获得遗传上所需的植物株系。

[0010] 在另一个方面，本发明涉及一种细胞，该细胞包含编码多肽的核酸，所述多肽与 SEQ ID NO:12 具有至少 80% 的序列相同性，还涉及一种植物细胞，所述植物细胞包含编码多肽的核酸，所述多肽具有与 SEQ ID NO:12 的氨基酸 810-872 具有至少 80% 的序列相同性的氨基酸序列。

[0011] 在另一个方面，本发明涉及一种病毒载体，其包含编码 Cas9 多肽的核苷酸序列。所述病毒载体能够包含编码多肽的核苷酸序列，所述多肽的氨基酸序列与 SEQ ID NO:12 具有至少 90% 相同性。所述病毒载体可来自烟草脆裂病毒或双生病毒。

[0012] 在另一个方面，本发明涉及一种 T-DNA，所述 T-DNA 包含编码多肽的核酸序列，所述多肽具有与 SEQ ID NO:12 的氨基酸 810-872 有至少 80% 的序列相同性的氨基酸序列。本发明还涉及一种农杆菌 (*Agrobacterium*) 菌株，其包含所述 T-DNA。

[0013] 另一方面，本发明涉及一种用于在植物细胞中表达 Cas 蛋白的方法。所述方法可包括：提供一种农杆菌 (*Agrobacterium*) 或剑菌 (*Ensifer*) 载体，所述载体包含 T-DNA，所述 T-DNA 包含：编码具有与 SEQ ID NO:12 的氨基酸 810-872 具有至少 80% 的序列相同性的氨基酸序列的多肽的核酸序列，其中该多肽编码序列操作性地连接至启动子；使所述农杆菌 (*Agrobacterium*) 或剑菌 (*Ensifer*) 载体与植物细胞接触；和，使所述核酸序列在所述植物细胞中表达。所述启动子可以是诱导型启动子（例如，雌激素诱导型启动子）。所述方法还可包括：使所述植物细胞与编码向导 RNA 的核酸接触，所述向导 RNA 与 Cas 蛋白相关联。所述植物细胞可以是原生质体。

[0014] 除非另外定义，本文使用的所有技术和科学术语的意义与本发明所属领域普通技术人员通常所理解的相同。虽然与本文所述类似或等同的方法和材料可用来实施本发明，但在下文描述合适的方法和材料。本文中述及的所有出版物、专利申请、专利和其他参考文献

献都通过引用全文纳入本文。在抵触的情况下,以本说明书(包括定义在内)为准。此外,材料、方法和实施例都仅是说明性的,并不意在构成限制。

[0015] 在附图和以下描述中详细说明了本发明的一种或多种实施方式。通过说明书和权利要求书,不难了解本发明的其它特征、目的和优点。

附图说明

[0016] 图 1 是 pMDC32 质粒(标准 T-DNA 表达质粒)的示意图,其包含 Cas9 编码序列和 cr/tracrRNA 杂合序列。该质粒的核苷酸序列示于 SEQ ID NO:6。

[0017] 图 2 是 pFZ19 质粒(雌激素-诱导型 T-DNA 表达载体)的示意图,其包含 Cas9 编码序列和 cr/tracrRNA 杂合 @ 序列。该质粒的核苷酸序列示于 SEQ ID NO:7。

[0018] 图 3 是 pNJB121 质粒(双生病毒-复制子 T-DNA 载体)的示意图,其包含 Cas9 编码序列和 cr/tracrRNA 杂合体。该质粒的核苷酸序列示于 SEQ ID NO:8。

[0019] 图 4A-4D 提供证据显示植物细胞中的 CRISPR/Cas 功能,其中 Cas9 编码序列和 cr/tracrRNA 杂合体通过农杆菌(Agrobacterium)或双生病毒复制子递送。图 4A 说明载有植物密码子最优化的 Cas9 序列的 T-DNA。所述 cr/tracrRNA 杂合体(称为 sgRNA)置于拟南芥 AtU6-26 启动子(PU6)的下游。“棒棒糖(lollypop)”指示对于复制酶(Rep)介导的复制而言具有重要性的长基因间区段(LIR)。灰色方框代表同样对于复制子功能具有重要性的短基因间区段(SIR)。未标记的灰色箭头是 35S 启动子,其能够在复制子环化之后驱动 Cas9 表达。Cas9 表达也可通过 LIR 驱动,其具有启动子功能。所述的整个构建体被称为 LSL T-DNA。图 4B 是包含 PCR 产物的琼脂糖凝胶照片,证明植物细胞中双生病毒复制子的环化。采用 PCR 引物(图 4A 中的小箭头)从携带所述复制子的农杆菌(Agrobacterium) T-DNA 感染的细胞扩增 DNA。仅在编码双生病毒复制酶(pRep)的存在下发生复制子的环化和和扩增。图 4C 显示,在烟草(Nicotiana tabacum) SurA/SurB 基因座处,缺失 Cas9-诱导的突变。烟草叶组织用图 4A 所示的包含 pREP 和 LSL T-DNA 的两种农杆菌(Agrobacterium)菌株进行注射器浸润;进行该处理以测试采用双生病毒复制子的 CRISPR/Cas9 介导的诱变。或者,叶组织用仅包含 LSL T-DNA 的单一农杆菌(Agrobacterium)菌株浸润;进行该处理以测试通过标准农杆菌(Agrobacterium) T-DNA 递送进行的 CRISPR/Cas9 介导的诱变。浸润后五天,基因组 DNA 经分离,并用作 PCR 反应的模板,该 PCR 反应设计以扩增 SurA/SurB 中的 Cas9 靶位点。所得的复制子用 AlwI 消化,通过凝胶电泳分离条带。图 4D 显示来自用 LSL T-DNA 和 pREP T-DNA 转化的样品中的切割抵抗性复制子的序列(SEQ ID NO:1-5)。PAM,原型间隔子邻近基序(protospacer adjacent motif)。

[0020] 图 5 是编码非功能性荧光蛋白(YFP)的报告质粒的示意图。

[0021] 图 6 是显示采用 YFP 报告质粒证明原生质体中 CRISPR/Cas 功能的荧光水平。制备烟草原生质体并用不同构建体转化,以测试 CRISPR/Cas9 的靶向切割,并通过流式细胞术检测 YFP 荧光。柱 1 显示用 YFP 报告子和表达 Cas9 的构建体以及由 AtU6-26 启动子表达的 cr/tracr RNA 转化的细胞中观察到的荧光水平。柱 2 显示用报告子、Cas9 和由 At7SL2-2 启动子表达的 cr/tracr RNA 转化的细胞中观察到的荧光水平。柱 3 显示仅用报告子转化的细胞中观察到的荧光(阴性对照);柱 4 显示用表达 YFP 的构建体转化的细胞中的荧光(阴性对照)。

[0022] 发明详述

[0023] 对植物进行高效的基因组工程改造能够通过向待修饰的 DNA 序列中导入靶向的双链断裂 (DSB) 来实现。该 DSB 激活细胞 DNA 修复通路,能够对其进行控制以在断裂位点附近获得所需的 DNA 序列修饰。可采用序列特异性核酸酶 (SSN) 来引入靶向的 DSB,所述序列特异性核酸酶是一类特殊化的蛋白质,其包括转录活化剂样 (TAL) 效应物核酸内切酶、锌指核酸酶 (ZFN),和寻靶核酸内切酶 (HE)。对于特异性 DNA 序列的识别通过与 SSN 编码的特异性氨基酸的相互作用来实现。在开发 TAL 效应物核酸内切酶之前,工程改造 SSN 的一大挑战是结合至 DNA 序列的氨基酸之间的不可预测的内容依赖性。尽管 TAL 效应物核酸内切酶大大降低了该难度,其较大尺寸(各 TAL 效应物核酸内切酶单体平均包含 2.5-3kb 的编码序列)和反复的性质阻碍了其在注重载体大小和稳定性的应用中的使用 (Voytas, *Annu Rev Plant Biol*, 64:327-350, 2013)。

[0024] 本发明部分基于如下发现:CRISPR/Cas 系统在植物基因组工程改造中可用作简单、有效的工具。CRISPR/Cas 分子是原核获得性免疫系统(其利用 RNA 碱基配对来指导 DNA 切割)的组分。指导 DNA DSB 需要如下两种组分:Cas9 蛋白,其作为核酸内切酶,和 CRISPR RNA (crRNA) 以及示踪 RNA (tracrRNA) 序列,其协助指导 Cas9/RNA 复合物靶向 DNA 序列 (Makarova 等, *Nat Rev Microbiol*, 9(6):467-477, 2011)。对于单一靶向 RNA 的修饰可足以改变 Cas 蛋白的核苷酸靶标。在一些情况中,crRNA 和 tracrRNA 可被工程改造为单一 cr/tracrRNA 杂合体,以指导 Cas9 切割活性 (Jinek 等, *Science*, 337(6096):816-821, 2012)。所述 CRISPR/Cas 系统可用于细菌、酵母、人类和斑马鱼,如他处所述(参见例如, Jiang 等, *Nat Biotechnol*, 31(3):233-239, 2013; Dicarolo 等, *Nucleic Acids Res*, doi:10.1093/nar/gkt135, 2013; Cong 等, *Science*, 339(6121):819-823, 2013; Mali 等, *Science*, 339(6121):823-826, 2013; Cho 等, *Nat Biotechnol*, 31(3):230-232, 2013 和 Hwang 等, *Nat Biotechnol*, 31(3):227-229, 2013)。

[0025] 先前并没有证明所述 CRISPR/Cas 系统在植物中的作用。该 CRISPR/Cas 系统源自具有相对小基因组的原核细胞,其中 Cas9 在显著 RNA 酶 III 活性的存在下于细胞中稳定表达。因此,当本文所述的植物细胞作用起始时,并不确定 Cas9 转基因的表达是否在植物细胞中可行,以及 Cas9 是否会与植物内容中的 RNA- 向导以及 RNA 酶 III 活性正确合作。此外,植物细胞中异源蛋白质的表达一般存在难度,这归因于不同密码子的采用。此外,预计植物中的 Cas9 表达会产生一些毒性,因为植物基因组较大,这会增加基因组 DNA 的非特异性切割的可能性,可能会诱导该细胞的基因毒性。本文报告的所述 CRISPR/Cas9 系统与包含 10-20 个核苷酸的特定识别序列发挥作用,其特异性小于大多数其它稀切核酸内切酶系统,例如 TAL 效应物核酸内切酶、大范围核酸酶,和锌指核酸酶。

[0026] 如本文所述,CRISPR/Cas 系统可用于产生靶向的 DSB 或单链断裂,并且能够用于(不限于):靶向诱变、基因靶向、基因替换、靶向缺失、靶向倒位、靶向易位、靶向插入,和通过共表达多重靶向 RNA 导向的单细胞中的多重 DSB 进行的多重基因组修饰。该技术可用于提高植物中功能性遗传学研究的速率,以及用于工程改造出具有改善的特点的植物,所述改善特点包括,营养质量增强、抗病和抗应激性提高,以及市场价值化合物的产量提高。可在植物叶组织中通过向报告基因和内源性基因座靶向 DSB 来进行概念验证实验。然后,可将该技术适用于原生质体和完整植物,以及基于病毒的递送系统。最终,通过向相同基因组

中的多个位点靶向 DSB 可显示多重基因组工程改造。

[0027] 一般而言,本文所述的系统和方法包括至少两种组分:RNA(crRNA 和 tracrRNA 或单一的 cr/tracrRNA 杂合体),其与植物细胞中的特定序列(例如,植物基因组中或染色体外质粒中,例如报告子)互补(并因而靶向至)该特定序列,和 Cas9 核酸内切酶,其能够在靶序列切割植物 DNA。代表性的 Cas9 编码序列示于 SEQ ID NO:6 的核苷酸 9771-14045(以及 SEQ ID NO:7 的核苷酸 4331-8605,和 SEQ ID NO:8 的核苷酸 9487-13761)。在一些情况中,系统还可包括含有靶向至植物序列的供体序列的核酸。所述核酸内切酶可造成所需基因座(或多个基因座)位置处的靶标 DNA 双链断裂,而所述植物细胞能够利用供体 DNA 序列来修复该双链断裂,由此将修饰稳定地纳入所述植物基因组。

[0028] Cas9 蛋白包含两个不同活性位点 - RuvC- 样核酸酶结构域和 HNH- 样核酸酶结构域,在 DNA 链的相对位点上产生位点特异性切口(Gasiunas 等,Proc Natl Acad Sci USA 109(39):E2579-E2586,2012)。RuvC- 样结构域处于 Cas9 蛋白的氨基末端附近,并且认为会切割与 crRNA 不互补的靶 DNA,而 HNH- 样结构域位于所述蛋白质中间,并且认为会切割与 crRNA 互补的靶 DNA。来自嗜热链球菌(*Streptococcus thermophilus*)的代表性的 Cas9 序列示于 SEQ ID NO:11(也可参见 UniProtKB 编号 Q03JI6),而来自化脓链球菌(*S. pyogenes*)的代表性的 Cas9 序列示于 SEQ ID NO:12(也可参见 UniProtKB 编号 Q99ZW2)。因此,本文所述的方法可采用编码 Cas9 多肽的核苷酸序列进行,所述 Cas9 多肽具有 SEQ ID NO:11 或 SEQ ID NO:12 所示序列。然而,在一些实施方式中,本文所述的方法能够采用编码 Cas9 功能性变体的核苷酸序列进行,所示 Cas9 功能性变体与 SEQ ID NO:11 或 SEQ ID NO:12 具有至少 80%(例如,至少 85%,至少 90%,至少 95%或至少 98%)的序列相同性。此外,Cas9 可分为两个部分,一部分包含 HNH 结构域,另一部分包含 RuvC 结构域。所述 HNH 结构域本身,与 RNA- 向导相关联,可具有一些切割活性,因此,本发明还设想采用包含 HNH 结构域的 Cas9 多肽,所述 HNH 结构域与 SEQ ID NO:11 中的 HNH 结构域(例如,SEQ ID NO:11 的氨基酸 828-879)或 SEQ ID NO:12 中的 HNH 结构域(例如,SEQ ID NO:12 的氨基酸 810-872)具有至少 80%(例如,至少 85%,至少 90%,至少 95%或至少 98%)的序列相同性。

[0029] 如下所述测定特定核酸或氨基酸序列与特定序列识别号所指示的序列之间的序列相同性百分比。首先,使用来自含有 BLASTN 版本 2.0.14 和 BLASTP 版本 2.0.14 的 BLASTZ 独立版本的 BLAST 2 序列(B12seq)程序比较核酸或氨基酸序列与特定序列识别号所指示的序列。该 BLASTZ 独立版本可在 fr.com/blast 或 ncbi.nlm.nih.gov 在线获取。解释如何使用 B12seq 程序的说明书可参见 BLASTZ 附带的自述文件。B12seq 使用 BLASTN 或 BLASTP 算法在两种序列之间进行比较。BLASTN 用于比较核酸序列,而 BLASTP 用于比较氨基酸序列。为比较两种核酸序列,如下所述设置选项:-i 设为含有待比较的第一核酸序列的文件(如 C:\seq1.txt);-j 设为含有待比较的第二核酸序列的文件(如 C:\seq2.txt);-p 设为 blastn;-o 设为任何需要的文件名(如 C:\output.txt);-q 设为 -1;-r 设为 2;且所有其他选项都设为其默认设置。例如,以下命令可用于生成含有两种序列之间比较的输出文件:C:\B12seq-i c:\seq1.txt-j c:\seq2.txt-p blastn-o c:\output.txt-q 1-r 2。为比较两种氨基酸序列,如下所述设置 B12seq 的选项:-i 设为含有待比较的第一氨基酸序列的文件(如 C:\seq1.txt);-j 设为含有待比较的第二氨基酸序列的文件(如 C:\seq2.txt);-p 设为 blastp;-o 设为任何需要的文件名(如 C:\output.txt);且所有其他选项都设为其默认

设置。例如,以下命令可用于生成含有两种氨基酸序列之间比较的输出文件 :C:\B12seq-ic:\seq1.txt-j c:\seq2.txt-p blastp-o c:\output.txt。如果两种比较的序列具有同源性,则指定的输出文件会以经比对序列的形式显示那些同源区域。如果比较的两种序列不具有同源性,则指定的输出文件不会显示经比对的序列。

[0030] 一旦进行比对,即通过对两种序列中存在相同核苷酸或氨基酸的位置的数目进行计数来测定匹配的数目。将匹配的数目除以鉴定的序列(如 SEQ ID NO:11)中所列序列长度或连接的长度(如来自鉴定的序列中所列序列的 100 个连续核苷酸或氨基酸残基),随后将结果值乘以 100,从而测定序列相同性百分比。例如,与 SEQ ID NO:11 中所列序列比对时具有 1300 个匹配的氨基酸序列与 SEQ ID NO:11 中所列序列具有 93.7% 相同性(即 $1300 \div 1388 \times 100 = 93.7$)。应注意序列相同性百分比值四舍五入至小数点后第一位。例如,75.11、75.12、75.13 和 75.14 向下四舍五入至 75.1,而 75.15、75.16、75.17、75.18 和 75.19 向上四舍五入至 75.2。还应注意,长度值始终是整数。

[0031] 本文中术语“功能性变体”意指蛋白质或蛋白质结构域的催化活性突变体。所述突变体可与亲本蛋白质或蛋白质结构域具有相同的活性水平,或具有高于或低于亲本蛋白质或蛋白质结构域的活性水平。

[0032] 可采用,例如,生物弹射轰击,将包含所示 crRNA、tracrRNA、cr/tracrRNA 杂合体、核酸内切酶编码序列,以及供体序列(可行时)的构建体递送至植物、植物部分或植物细胞。或者,可采用农杆菌(Agrobacterium)介导的转化递送该系统组分。在一些实施方式中,所示系统组分可在如下载体中递送:病毒载体(例如,来自 DNA 病毒的载体,例如但不限于,双生病毒(例如,卷心菜卷叶病毒、菜豆黄矮病毒、小麦矮病毒、番茄卷叶病毒、玉米条纹病毒、烟草卷叶病毒或番茄金花叶病毒)或纳米病毒(例如,蚕豆坏死黄化病毒);或者来自 RNA 病毒的载体,例如但不限于,烟草脆裂病毒(例如,烟草响叶病毒,烟草花叶病毒)、马铃薯 X 病毒组(例如,马铃薯 X 病毒)或大麦病毒(例如,大麦条纹花叶病毒)。

[0033] 在植物、植物部分或植物细胞被核酸内切酶编码序列和 crRNA 以及 tracrRNA 或 cr/tracrRNA 杂合体(在一些情况中,供体序列)感染或转染后,可采用任何合适方法来测定在靶位点是否已发生 GT 或靶向诱变。在一些实施方式中,表型变化可指示供体序列已被纳入靶位点。还可采用基于 PCR 的方法来确定基因组靶位点是否包含靶向突变或供体序列,和/或在所述供体的 5' 和 3' 端是否已发生精确重组。检测靶向突变的一种方法,本文中称作“PCR 消化”,描述于 Zhang 等(Proc Natl Acad Sci USA 107 ;12028:-120332010)。用于检测精确重组的方法包括,采用与供体序列具有同源性的探针的 southern 印迹。

[0034] 在一些实施方式中,本文提供的方法可包括:向植物、植物部分或植物细胞引入核酸,所述核酸包括 crRNA 和 tracrRNA 或嵌合型 cr/tracrRNA 杂合体,其中所述 crRNA 和 tracrRNA 或所述 cr/tracrRNA 杂合体靶向至对所述植物细胞为内源性的核苷酸序列,和还向所述植物、植物部分或植物细胞中引入 Cas9 核酸内切酶分子(例如,Cas9 多肽或其部分,例如包含 HNH 结构域的 Cas9 多肽的部分,或编码 Cas9 多肽或其部分的核酸),其中所述 Cas9 核酸内切酶分子在所述 crRNA 和 tracrRNA 序列(或所述 cr/tracrRNA 杂合体)靶向的序列处或附近包含双链断裂。

[0035] 用于本文提供的方法的植物、植物部分和植物细胞可来自任何植物物种。例如,在一些实施方式中,本文提供的方法可采用单子叶植物、其部分或其细胞。示例性的单子叶

植物包括但不限于,小麦、玉米、水稻、兰科植物、洋葱、芦荟、真百合 (true lily)、杂草 (例如、狗尾草)、木质灌木和树 (例如、棕榈和竹子),以及食用植物,例如菠萝和甘蔗。示例性的双子叶植物包括但不限于,番茄、木薯、大豆、烟草、马铃薯、拟南芥、玫瑰、三色堇、向日葵、葡萄、草莓、南瓜、菜豆、豌豆,和花生。

[0036] 在一些实施方式中,本文所述的方法可包括筛选所述植物、植物部分或植物细胞,以确定在所述 crRNA 和 tracrRNA 或所述 cr/tracrRNA 杂合体靶向的序列处或附近是否已经出现 DSB。例如, Zhang 等. (同上) 所述的 PCR- 消化试验可用于确定 DSB 是否已经出现。其它有用的方法包括但不限于, T7 试验、Surveyor 试验, 和 Southern 印迹 (限制酶结合序列是否存在于预定的切割位点处或附近)。

[0037] 此外,在其中采用植物部分或植物细胞的一些实施方式中,本文提供的方法可包括,从所述植物部分或植物细胞再生植物。所述方法还可包括繁育所述植物 (例如,其中引入核酸的植物或将植物部分或植物细胞用作起始材料再生后获得的在植物), 以获得遗传上所需的植物株系。用于再生和繁育植物的方法已在本领域中成熟建立。

[0038] 本文中还提供植物、植物部分和植物细胞,其包含编码 Cas9 多肽的核酸,所述 Cas9 多肽具有与 SEQ ID NO:11 或 SEQ ID NO:12 所示的氨基酸序列具有至少 80% (例如,至少 85%,至少 90%,至少 95%或至少 98%) 的相同性的氨基酸序列,或者编码 Cas9 多肽的核酸,所述 Cas9 包含与 SEQ ID NO:11 的氨基酸 828-879 或 SEQ ID NO:12 的氨基酸 810-872 具有至少 80% (例如,至少 85%,至少 90%,至少 95%或至少 98%) 的相同性的氨基酸序列。

[0039] 本发明还提供包含编码 Cas9 多肽的核苷酸序列的病毒载体。例如,病毒载体可包括编码多肽的核苷酸序列,所述多肽具有与 SEQ ID NO:11 或 SEQ ID NO:12 中所示氨基酸序列具有至少 80% (例如,至少 85%,至少 90%,至少 95%或至少 98%) 相同性的氨基酸序列。在一些实施方式中,病毒载体可具有编码 Cas9 多肽的核苷酸序列,所述 Cas9 多肽包含与 SEQ ID NO:11 的氨基酸 828-879 或 SEQ ID NO:12 的氨基酸 810-872 具有至少 80% (例如,至少 85%,至少 90%,至少 95%或至少 98%) 序列相同性的氨基酸序列。所述载体可来自任何合适类型的病毒,例如烟草脆裂病毒或双生病毒。

[0040] 本文中还提供包含编码 Cas9 多肽的核酸序列的 T-DNA 分子,所述 Cas9 多肽具有与 SEQ ID NO:11 或 SEQ ID NO:12 所示氨基酸序列具有至少 80% (例如,至少 85%,至少 90%,至少 95%或至少 98%) 相同性的氨基酸序列。在一些实施方式中, T-DNA 可包括编码 Cas9 多肽的核苷酸序列,所述 Cas9 多肽包含与 SEQ ID NO:11 的氨基酸 828-879 或 SEQ ID NO:12 的氨基酸 810-872 具有至少 80% (例如,至少 85%,至少 90%,至少 95%或至少 98%) 序列相同性的氨基酸序列。

[0041] 本发明还提供包含本文所述的 T-DNA 的农杆菌 (Agrobacterium) 菌株。

[0042] 此外,本发明提供用于在植物、植物部分或植物细胞中表达 Cas 蛋白的方法。所述方法可包括,例如, (a) 提供包含 T-DNA 的农杆菌 (Agrobacterium) 或剑菌 (Ensifer) 载体,所述 T-DNA 包含编码 Cas9 多肽的核酸序列,所述 Cas9 多肽具有与 SEQ ID NO:11 或 SEQ ID NO:12 具有至少 80% (例如,至少 85%,至少 90%,至少 95%或至少 98%) 序列相同性的氨基酸序列,其中该 Cas9 编码序列操作性地连接至启动子, (b) 使所述农杆菌 (Agrobacterium) 或剑菌 (Ensifer) 载体与植物、植物部分或植物细胞接触,和 (c) 在所述

植物、植物部分或植物细胞中表达所述核酸序列。所述启动子可以是,例如,组成型启动子(例如,CaMV 35S启动子),诱导型启动子(例如,雌二醇-诱导的XVE启动子;Zuo等,Plant J 24:265-273,2000),细胞特异型启动子或通过自杀外显子的可变剪接激活的启动子。在一些实施方式中,所述方法还可包括使所述植物、植物部分或植物细胞接触编码向导RNA的核酸,所述向导RNA与Cas蛋白相关联,以及,表达所述向导RNA。

[0043] 以下实施例进一步描述了本发明,这些实施例并不限制权利要求书所述的本发明范围。

实施例

[0044] 实施例 1 - 用于表达 CRISPR/Cas 组分的质粒

[0045] 为了证明所述 CRISPR/Cas 系统在植物中的基因组编辑功能性,构建质粒以编码 Cas9、crRNA 和 tracrRNA、cr/tracrRNA 杂合体,和用于表达 crRNA、tracrRNA 或 cr/tracrRNA 杂合体的 RNA 聚合酶 III 启动子(例如,AtU6-26 或 At7SL-2)。植物密码子最优化的 Cas9 编码序列经合成并克隆进入多位点 Gateway 进入型质粒。此外,由 RNA 聚合酶 III(PolIII) 启动子 AtU6-26 和 At7SL2-2 驱动的 crRNA 和 tracrRNA 或 cr/tracrRNA 杂合体经合成并克隆进入第二多位点 Gateway 进入型质粒。未能够有效重建所述 crRNA 序列(用于再导向 CRISPR/Cas 介导的 DSB),向 crRNA 核苷酸序列中插入反向 IIS 型限制酶位点(例如,BsaI 和 Esp3I)。通过采用合适的 IIS 型限制酶进行消化,可采用寡核苷酸将靶序列高效克隆进入 crRNA 序列。将用于 Cas9 和用于从 RNA 聚合酶 III 启动子(AtU6-26 或 At7SL2-2) 表达所述 crRNA 和 tracrRNA 或所述 cr/tracrRNA 杂合体的进入型质粒重组进入 pMDC32(标准 T-DNA 表达质粒,具有 2x35S 启动子;图 1 和 SEQ ID NO:6)、pFZ19(雌激素-诱导型 T-DNA 表达载体;图 2 和 SEQ ID NO:7;Zuo 等,Plant J. 24(2):265-273,2000),和 pNJB121(双生病毒-复制子 T-DNA 载体;图 3 和 SEQ ID NO:8)。

[0046] 实施例 2 - 体细胞植物组织中的 CRISPR/Cas 活性

[0047] 为了证明 CRISPR/Cas 系统作为 SSN 发挥作用的能力,修饰双生病毒-复制子 T-DNA 载体,pNJB121,以编码 Cas9 和 cr/tracrRNA 杂合体序列(图 4A)。靶向 RNA 序列(由 crRNA 中的核苷酸序列编码;负责指导 Cas9 切割)设计为与内源性 SuRA 和 SuRB 基因中的序列同源。crRNA 与 SuR 基因座匹配的靶向部分的序列是 5'-GUGGGAGGAUCGGUUCUAUA(SEQ ID NO:9;5' G 不与 SuR 基因座匹配,但是 RNA 聚合酶 III 转录所需的)。尽管 pNJB121 是双生病毒-复制子,在不存在复制酶(Rep)的情况下,无扩增发生。因此,在 Rep 不存在的情况下,pNJB121 是标准 T-DNA 载体,且无复制子形成。该修饰的 pNJB121 质粒通过根癌农杆菌(Agrobacterium tumefaciens)递送至烟草(Nicotiana tabacum)叶组织。浸润后五天,采用 PCR-消化评估 SuRA/SuRB 序列的 Cas9 介导的突变(图 4C)。对应的靶序列处突变的存在指示 CRISPR/Cas 系统在植物叶细胞中的功能性。

[0048] 实施例 3 - 原生质体中的 CRISPR/Cas 活性

[0049] 为了进一步证明 CRISPR/Cas 系统在植物中的活性,评估拟南芥(thaliana)和烟草(Nicotiana tabacum)原生质体中的 DNA 序列的靶向诱变。靶向 crRNA 序列重新设计为与内源性 ADH1 或 TT4 基因(拟南芥)中存在的序列同源,或整合 gus:nptII 报告基因或 SuRA/SuRB(烟草)。原生质体从拟南芥和烟草叶组织分离,并分别用编码 Cas9 和 ADH1-或

TT4-靶向 crRNA 的质粒,和编码 Cas9 和 gus:nptII- 或 SuRA/SuRB- 靶向 crRNA 的质粒转染。转染后 5-7 天提取基因组 DNA,并评估对应的靶序列处的突变。在 ADH1、TT4、gus:nptII 或 SuRA/SuRB 基因中检测到突变指示 CRISPR/Cas 系统在植物原生质体中靶向内源性基因的功能性。

[0050] 在初始研究中,评估了所述 CRISPR/Cas 系统切割染色体外报告质粒的能力,采用的是与 Zhang 等所述类似的方法 (Plant Physiol 161:20-27, 2013)。所述报告质粒编码非功能性荧光蛋白 (YFP;图 5 和 SEQ ID NO:10)。YFP 表达被侧接 Cas9/crRNA 复合物的靶序列的内部编码序列的直接重复所干扰。靶向 DSB 在 Cas9/crRNA 靶序列处的生成导致直接重复序列的重组,由此恢复了 YFP 基因功能。来自烟草 SuRA/SuRB 基因座的序列被克隆进入直接重复之间的 YFP 报告子。然后产生靶向该位点的 cr/tracrRNA 杂合体构建体。靶向 SuR 基因座的 crRNA 的部分的序列是 5'-GUGGGAGGAUCGGUUCUAUA (SEQ ID NO:9;同样地所述 5'G 不匹配 SuR 基因座,但是 RNA 聚合酶 III 转录所需的)。烟草 (Nicotiana tabacum) 原生质体用编码 Cas9、cr/tracrRNA 杂合体,和 YFP 报告子的质粒转化,并且通过流式细胞术监测到了因 CRISPR/Cas 核酸酶活性所致的 YFP 表达的恢复。采用编码 YFP 的阳性对照质粒,94.7%的细胞被转化并表达 YFP (图 6,柱 4)。用单独报告子转化的细胞给出仅稍高于背景的活性水平 (图 6,柱 3)。当细胞用表达 Cas9 和 cr/tracrRNA 的构建体转化时,观察到显著的活性,指示 Cas9/crRNA 复合物切割了靶标。对于从 AtU6-26 启动子表达的 cr/tracrRNA,18.8%的细胞显荧光 (图 6,柱 1)。当从 At7SL2-2 启动子表达 cr/tracrRNA 时,20.7%的细胞呈 YFP 阳性 (图 6,柱 2)。YFP- 表达型细胞的检测指示 CRISPR/Cas 系统在植物原生质体中的功能性。

[0051] 实施例 4 - 采用 CRISPR/Cas 系统在原生质体中的多重基因组工程改造

[0052] CRISPR/Cas 系统在不同 DNA 序列产生多重 DSB 的能力采用植物原生质体评估。为了指导对于 TT4、ADH1 和染色体外 YFP 报告质粒 (相同拟南芥原生质体中) 的 Cas9 核酸酶活性,crRNA 和 tracrRNA 或 cr/tracrRNA 杂体质粒经修饰以表达多重 crRNA 靶向序列。这些序列经设计与 TT4、ADH1 和 YFP 报告质粒中存在的序列同源。在用 Cas9、crRNA、tracrRNA 或所述 cr/tracrRNA 杂合体,以及 YFP 报告质粒转染进入拟南芥原生质体之后,对 YFP- 表达型细胞定量并分离,然后提取基因组 DNA。观察到 YFP- 表达型细胞的 ADH1 和 TT4 基因中的突变,说明 CRISPR/Cas 能够促进拟南芥细胞中的多重基因组工程改造。

[0053] 为了证明在烟草原生质体中的多重基因组工程改造,修饰包含 crRNA 的质粒以编码与整合的 gus:nptII 报告基因、SuRA/SuRB 和 YFP 报告质粒同源的序列。与拟南芥原生质体中所述的方法类似,烟草原生质体用 Cas9、crRNA、tracrRNA 或所述 cr/tracrRNA 杂合体以及 YFP 报告质粒转染。YFP- 表达型细胞经定量并分离,然后提取基因组 DNA。观察到 YFP- 表达型细胞中整合的 gus:nptII 报告基因和 SuRA/SuRB 中的突变,说明 CRISPR/Cas 能够促进烟草细胞中的多重基因组工程改造。

[0054] 实施例 5 - 植物中的 CRISPR/Cas 活性

[0055] 为证明 CRISPR/Cas 在植物中的活性,修饰 pFZ19T-DNA 以编码 Cas9 和所述 crRNA 和 tracrRNA 或所述 cr/tracrRNA 杂合体序列。靶 DNA 序列存在于内源性 ADH1 或 TT4 基因组中。通过浸花法,采用农杆菌 (Agrobacterium),将所得的 T-DNA 整合进入拟南芥基因组。通过直接接触雌激素,在原生转基因植物中诱导 Cas9 表达。提取来自体细胞叶组织的基因

组 DNA, 并通过 PCR- 消化评估其在对应的基因组基因座位置处的突变。观察到 ADH1 或 TT4 基因中的突变, 证明 CRISPR/Cas 在植物中的活性。或者, CRISPR/Cas 活性可通过对 T2 种子 (产自诱导的 T1 亲本) 筛选对应的基因组基因座的杂合或纯合突变来评估。此外, CRISPR/Cas 实现多重基因组工程改造的能力通过用与 ADH1 和 TT4 质粒同源的序列修饰包含多重 crRNA 的质粒来评估。将所得的 T-DNA 质粒整合进入拟南芥基因组, 在原生转基因植物中诱导 Cas9 表达, 并通过评价 T1 和 T2 植物中的 ADH1 和 TT4 基因来评估 CRISPR/Cas 活性。在 ADH1 和 TT4 基因中均观察到突变, 说明 CRISPR/Cas 能够促进拟南芥植物中的多重基因组工程改造。

[0056] 实施例 6 - CRISPR/Cas 组分的病毒递送

[0057] 植物病毒可以是用于递送异源性核酸序列的有效载体, 例如用于 RNAi 试剂或用于表达异源蛋白质。有用的植物病毒包括 RNA 病毒 (例如, 烟草花叶病毒、烟草响叶病毒、马铃薯 X 病毒, 和大麦条纹花叶病毒) 以及 DNA 病毒 (例如, 卷心菜卷叶病毒、菜豆黄矮病毒、小麦矮病毒、番茄卷叶病毒、玉米条纹病毒、烟草卷叶病毒、番茄金花叶病毒, 和蚕豆坏死黄化病毒; Rybicki 等, Curr Top Microbiol Immunol, 2011; 和 Gleba 等, Curr Opin Biotechnol 2007, 134-141)。所述植物病毒可被修饰以递送 CRISPR/Cas9 组分。概念验证实验在烟草 (*Nicotiana tabacum*) 叶细胞中进行, 采用 DNA 病毒 (双生病毒复制子; Baltes 等, Plant Cell 26:151-163, 2014)。为此, 修饰 crRNA 序列以包含对于内源性 SuRA/SuRB 基因座的同源性。将所得的质粒连同 Cas9 克隆进入 pNJB121 (双生病毒复制所需的具有顺式作用元件的 T-DNA 目的地载体 (LSL T-DNA)) (图 4A)。农杆菌 (*Agrobacterium*) 的 LSL T-DNA 与编码复制酶蛋白 T-DNA (Rep; REP T-DNA) 的共表达导致双生病毒复制子的复制性释放 (图 4B)。T-DNA 通过用农杆菌 (*Agrobacterium*) 注射器浸润被递送至烟草叶组织。浸润后五 - 七天, 采用 PCR- 消化评估 SuRA/SuRB 序列的 Cas9 介导的突变 (图 4C)。抵抗消化的 PCR 复制子经克隆并序列。对应的靶序列存在突变, 指示植物病毒是递送 CRISPR/Cas 组分的有效载体 (图 4D)。

[0058] 其它实施方式

[0059] 应理解虽然本发明已结合其详述进行描述, 但以上描述意在说明而不是限制本发明的范围, 该范围由所附权利要求的范围限定。其他方面、优点和修改在以下权利要求的范围内。

[0001]

序列表

- <110> 明尼苏达大学董事会 (Regents of the University of Minnesota)
- <120> 采用 CRISPR/Cas 系统的植物基因组的工程改造
- <130> 09531-0335W01
- <150> 61/790, 694
- <151> 2013-03-15
- <160> 12
- <170> FastSEQ Windows 4.0 版
- <210> 1
- <211> 37
- <212> DNA
- <213> 烟草 (*Nicotiana tabacum*)
- <400> 1
ggttcaatgg gaggatcggg tctataagge taacaga 37
- <210> 2
- <211> 33
- <212> DNA
- <213> 人工序列
- <220>
- <223> 突变的序列
- <400> 2
ggttcaatgg gaggatcggg taaggctaac aga 33
- <210> 3
- <211> 30
- <212> DNA
- <213> 人工序列
- <220>
- <223> 突变的序列

[0002]

<400> 3	
ggttcaatgg gaggatataa ggctaacaga	30
<210> 4	
<211> 17	
<212> DNA	
<213> 人工序列	
<220>	
<223> 突变的序列	
<400> 4	
ggtataaggc taacaga	17
<210> 5	
<211> 16	
<212> DNA	
<213> 人工序列	
<220>	
<223> 突变的序列	
<400> 5	
ggttcaatgg gaggat	16
<210> 6	
<211> 15198	
<212> DNA	
<213> 人工序列	
<220>	
<223> 重组质粒	
<400> 6	
cgtaatcatg gtcatagetg tticcctgtgt gaaattgtta tccgetcaca attccacaca 60	
acatacgagc cggaagcata aagtgtaaag cctgggggtgc ctaatgagtg agctaaactca 120	
cattaattgc gttgcgetca ctgeccgctt tccagteggg aaacctgtcg tgcagctgc 180	
attaatgaat cggccaacgc gcggggagag gcggtttgcg tattggctag agcagcttgc 240	
caacatggtg gagcagcaca ctctcgtcta ctccaagaat atcaaagata cagtctcaga 300	
agaccaaagg gctattgaga cttttcaaca aagggttaata teggaaacc tcctcggatt 360	
ccattgccc a gctatctgtc acttcatcaa aaggacagta gaaaaggaag gtggcaccta 420	
caaatgcat cattgcgata aaggaaaggc tatcgttcaa gatgcctctg ccgacagtgg 480	
tcceaaagat ggacccccac ccacgaggag catcgtggaa aaagaagacg ttccaaccac 540	
gtcttcaag caagtggatt gatgtgataa catgggtggag caegaeacte tcgtctactc 600	

[0003]

caagaatatc aaagatacag tctcagaaga ccaaagggtc attgagactt ttcaacaaag 660
ggtaaatatcg ggaaacctcc teggattcca ttgccagct atctgtcact tcatcaaaag 720
gacagtagaa aaggaaggtg gcacctaaa atgccatcat tgcgataaag gaaaggctat 780
cgttcaagat gectctgceg acagtggccc caaagatgga cccccacca cgaggagcat 840
cgtggaaaaa gaagacgttc caaccacgic ticaaagcaa gtggattgat gfgatattc 900
cactgacgta agggatgacg cacaatccca ctatccttcg caagaccitc ctctatataa 960
ggaagtfcatt ttcatttggg gaggacacgc tgaateace agtctctctc tacaaateta 1020
tctctctcga gcttttcgag atccccgggg gcaatgagat atgaaaaage ctgaaactac 1080
cgcgacgtct gtcgagaagt ttctgatcga aaagtctgac agcgtctccg acctgatgca 1140
gctctcggag ggcgaagaat ctctgtcttt cagcttcgat gtaggagggc gtggatattg 1200
cctgcccggta aatagctgcg ccgatggttt ctacaaagat cgttatgttt atcggcactt 1260
tgcacggtcc gcctcccgga ttccggaagt gcttgacatt ggggagttta gcgagagcct 1320
gacctattgc atctcccgcc gtgcacaggc tgtcaegtig caagacctgc ctgaaaccga 1380
actgcccgct gttctacaac cggctcggga ggetatggat gcgatcctg cgcccgatct 1440
fagccagacg agcgggttcg gcccaatcgg accgcaagga atcgggtcaal aactacatg 1500
gcgtgatttc atatgcgcga ttgtgatcc ccatgtgtat cactggcaaa ctgtgatgga 1560
cgacaccgtc agtgcgtccg tcgctcagcc tctcgtatgag ctgatgcttt gggccgagga 1620
ctgccccgaa gtcggcacc tcgtgcacc ggatttcggc tccaacaatg tctgacgga 1680
caatggccgc ataacagcgg tcattgactg gagcgaggcg atgttcgggg atccccata 1740
cgaggctgcc aacatcttct tctggagccc gtggttggct tgtatggage agcagacgcg 1800
ctacttcgag cggagccatc cggagcttgc aggatcggca cgactccggg cgtatatgct 1860
ccgcattggt cttgaccaac tctatcagag cttggttgac ggcaatttcg atgatgcagc 1920
ttgggcgcag ggtcgtatgc acgcaatcgt ccgatccgga gccgggactg tcgggcgtac 1980
acaaatcgcc cgcagaagcg cgccctctg gaccgatggc tgtgtagaag tactcggcca 2040
lagtggaaac cgacgcccc gactctgccc gagggcaaaag aaalagagta gatgccgacc 2100
ggatctgtcg atcgacaage tcgagtttct ccataataat gtgtgagtag ttcccagata 2160
agggaaattag gtttccatata gggtttcgct catgtgttga gcatataaga aacccttagt 2220
atgtatttgt atttgtaaaa tacttctate aataaaattt ctaattceta aaaccaaatt 2280
ccagtactaa aatccagatc ccccgaatta attcggcgtt aattcagtac attaaaaacg 2340
tcgcgaatgt gttattaagt tgtetaagcg tcaatttgtt tacaccaca tafatctgce 2400
caccagccag ccaacagctc cccgaccggc agctcggcac aaaatcacca ctcgatacag 2460
gcagcccatc agtccgggac ggcgtcagcg ggagagccgt tgaaggcgg cagactttgc 2520
tcatgttacc gatctattc ggaagaacgg caactaagct gccgggtttg aaacacggat 2580
gatctcggcg aggtagcat gttgattgta acgatgacag agegttctg cctgtgatca 2640
ccgctgtttc aaaatcggct ccgtcgatac latgltatc gccaaatttg aaaacaactt 2700
tgaaaaagct gttttctggt atttaagggt ttagaatgca aggaacagtg aattggagtt 2760
cgtcttgtta taattagctt cttggggat ctttaaatat ttagaaaaag aggaaggaaa 2820
faataaatgg ctaaaatgag aatacaccg gaattgaaa aactgatcga aaaataccgc 2880
tgcgtaaaag atacggaagg aatgtctct gctaaggat ataagctggt gggagaaaat 2940
gaaaacctat atttaaaaat gacggacagc cgttataaag ggaccaceta tgatgtgga 3000
cgggaaaagg acatgatgct atggtcggaa ggaaagctgc ctgttccaaa ggtcctgcac 3060
ttgaacggc atgatggctg gagcaatctg ctcatgagt aggccgatgg cgtcctttgc 3120
tcggaagagt atgaagatga acaaagccct gaaaagatta tegagctgta tgcggagtgc 3180
atcaggctct ttaactecat cgacatateg gattgtccct atacgaatag cttagacagc 3240

[0004]

cgcttagcgg aattggatta cttactgaat aacgatctgg ccgatgtgga ttgcgaaaac 3300
 tgggaagaag acactccatt taaagatccg cgcgagctgt atgatttttt aaagacggaa 3360
 aagcccgaag aggaacttgt cttttccac ggcgacctgg gagacagcaa catctttgtg 3420
 aaagatggca aagtaagtgg ctttattgat cttgggagaa ggcgcaggge ggacaagtgg 3480
 tatgacattg ccttctcggt ccggtcgate agggaggata tegggaaga acagtatgtc 3540
 gagctatfff ttgacttact ggggatcaag cctgattggg agaaaataaa atattatatt 3600
 ttaetggatg aattgtttta gtacctagaa tgcattgacca aatccctta acgtgagttt 3660
 tctttccact gagegtcaga ccccgtagaa aagatcaaag gatcttcttg agatcctttt 3720
 tttctgcgcg taatctgctg cttgcaaaca aaaaaccac cgctaccage ggtggtttgt 3780
 ttgcccgate aagagctacc aactcctttt ccgaaggtaa ctggcttcag cagagcgcag 3840
 ataccaata ctgtccttct agttagccg tagttaggcc accacttcaa gaactctgta 3900
 gcaccgccta catacctcgc tctgctaate ctgttaccag tggctcctgc cagtggcgat 3960
 aagtcgtgtc ttaccgggtt ggaactcaaga cgatagttae eggataagge gcagcggtcg 4020
 ggctgaacgg ggggttcgtg cacacagccc agcttggage gaacgacctt caccgaactg 4080
 agatacctac agcgtgagct atgagaaage gccacgcttc ccgaaggag aaaggcggac 4140
 aggtatccgg taagcggcag ggtcggaaaca ggagagcgcg egagggagct tccaggggga 4200
 aacgcctggg atctttatag tctgtcggg tttgccacc tctgacttga gcgtcgattt 4260
 ttgtgatgct cgtcaggggg gcggagccta tggaaaaacg ccagcaacgc ggctttttta 4320
 cggttcctgg ctttttgctg gccttttct cacaatgtct ttcctgcgtt atcccctgat 4380
 tctgtggata accgtattac cgcctttgag tgagctgata ccgctcgcg cagecgaacg 4440
 accgagcgcg gcgagtcagt gagcggaggaa gcggaagagc gectgatgcg gtattttctc 4500
 cttacgcate tgtcgggat ttcacaccgc atatggtgca ctctcagtae aatctgetct 4560
 gatgccgat agttaagcca gtatacact cgetatcgtt acgtgactgg gtcattgctg 4620
 cgccccgaca cccccaaca cccgtgacg cccctgacg ggttgtctg ctcccggcat 4680
 ccgcttacag acaagctgtg accgtctccg ggagctgcat gtgtcagagg ttttaccgt 4740
 catcaccgaa acgcgcgagg cagggtgctt tgatgtgggc gccggcggtc gactggcgac 4800
 ggcgcggctt gtcgcgccc tggtagattg cctggccgta ggccagccat ttttgagcgg 4860
 ccagcggcgc cgataggccg acgcgaagc gcggggcgta gggagcgcag cgaccgaagg 4920
 gtaggcgctt tttgcagctc ttgggtgtg cgtggtccag acagttatgc acagccagg 4980
 cgggttttaa gagtttaat aagttttaa gagttttagg eggaaaaac gcctttttt 5040
 tcttttatat cagtcactta catgtgtgac cgtttccaa tgtacggtt tgggttccca 5100
 atgtacgggt tccggttccc aatgtacggc tttgggttcc caatgtact gctatccaca 5160
 ggaaagagac cttttcgacc tttttccct gctagggeaa tttgccctag catctgetcc 5220
 gtacattagg aaccggcgga tgettcgcc tgcattcagg tgcggtagcg catgactagg 5280
 atcgggccag cctgccccgc ctctccttc aatcgtact ccggcaggtc atttgaccgc 5340
 atcagcttgc gcacggtgaa acagaacttc ttgaactctc cggcgcgtgc actgcgttcg 5400
 tagatcgtct tgaacaacca tctggttctt gccttgcctg cggcgcggcg tgcaggcgg 5460
 tagagaaaac ggccgatgcc gggatcgate aaaaagtaat cggggtgaa cgtcagcacg 5520
 tccgggttct tgccttctgt gatctcggg tacatccaat cagctagctc gatctcgatg 5580
 tactccggcc gcccggttcc gctctttacg atctttagc ggctaatcaa ggttccacc 5640
 tcggataccg tcaccaggcg gccgttcttg gccttcttcg tacgttgcg ggcaacgtgc 5700
 gtggtgttta accgaatgca gttttctacc aggtcgtctt tctgcttcc gccatcggct 5760
 cgccggcaga acttgagtac gtccgcaacg tgtggaegga acacgcggcc gggcttgtct 5820
 ccttccctt cccggtatcg gttcatggat tgggttagat gggaaacegc catcagtae 5880

[0005]

```

aggtcgtaat cccacacact ggccatgccg gccggccctg cggaaacctc tacgtgcccg 5940
tctggaagct cgtagcggat cacctcgcca gctcgtcggc cacgcttcga cagacggaaa 6000
acggccacgt ccatgatgct gcgactatcg cgggtgcccc egtcatagag catcggaacg 6060
aaaaaatctg gttgctcgtc gcccttgggc ggttctctaa tcgacggcgc accggetgce 6120
ggcggttgcc gggattcttt gcgatttega tcageggccg ctigccaega ttcaccgggg 6180
cgtgcttctg cctcgatgcg ttgcegetgg gggcctgceg eggcctteaa cttctccacc 6240
aggteatcac ccagcgcgcg gccgatttgt accgggcccg atggfittgag accgtcacgc 6300
cgattcctcg gcttggggg ttccagtgcc attgcagggc cggcagacaa cccagecgtc 6360
tacgcctggc caaccgccg ttctccaca catggggcat tccacggcgt cgggtgcctgg 6420
ttgttcttga ttttccatgc cgcctcttt agccctaaa attcatctac tcatttattc 6480
atttgetcat ttactctggt agctgcgca tgtattcaga tagcagctcg gtaatggctc 6540
tgcccttggc taccgcgtac atcttcagct tgggtgtgat ctcgccggc aactgaaagt 6600
tgaccgcctt catggctggc gtgtctgcca ggetggccaa cgttgcagec ttgctgetgc 6660
gtggetcggg accggccgca cttagecgtg ttgtctttt gctcatttct tctttacctc 6720
attaactcaa atgagttttg atttaatttc agcggccagc gcctggacct cgcgggcagc 6780
gtgcacctcg ggttctgatt caagaacggg tgtgccggcg gcggcagtcg ctgggtaget 6840
cacgcctgct gtgatacggg actcaagaat gggcagctcg taccggcca gcgcctcggc 6900
aacctcaccg ccgatgcgcg tgccttctgat cgcgccgac acgacaaagg ccgctttag 6960
ccttccatcc gtgacctcaa tgcctgctt aaccagctcc accaggctcg cgggtggcca 7020
tatgtogtaa ggcttggct gcaccggaat cagcacgaag teggetgcct tgatcgcgga 7080
cacagecaag tccgccctt ggggcctcc gtcgatact acgaagtcgc gccggccgat 7140
ggccttcacg tcgggtcaa tcgtcggcg gtcgatgccg acaacggta gcggttgatc 7200
ttccgcacg gcccccatt cgcggcact gccctgggga teggaatcga ctaacagAAC 7260
atcggccccg gcgagttgca gggcggggc tagatgggtt gcgatggctg tcttgcctga 7320
cccgccttct tggtaagta cagcgataac cttcatgctg tccccttgcg tatttgttta 7380
ttactcacc gcacatata cgcagcagc gcacgacga agctgtttta ctcaaataca 7440
catcaccttt ttagacggcg gcgctcgggt tcttcagcgg ccaagetggc cggccaggcc 7500
gccagcttgg catcagacaa accggccagg attteatgca gccgcacggt tgagacgtgc 7560
gcgggcggct cgaacacgta cccggccgcg atcatctccg cctcgatctc ttcggtaatg 7620
aaaaacggtt cgtcctggcc gtcctgggct ggttctatgc ttgttctct tggcgtteat 7680
tctcggcggc cgcaggggcg tcggcctcgg tcaatgcgct ctcacggaag gcaccgcgcc 7740
gccfggcctc ggtgggcgct acttctctgc tgcctcaag tgcggctac agggctcagc 7800
gatgcacgcc aagcagtgca gccgcctctt tcaaggctcg gccctctcgg tcgatcagct 7860
cgcgggcgtg cgcgatctgt gccggggtga gggtagggcg ggggccaaac ttcacgcctc 7920
ggccttggc ggctcgcgc ccgctcggg tgcggtcgat gattagggaa cgctcgaact 7980
cggcaatgcc ggcgaacacg gteaacacca tgcggccggc cggcgtggtg gtgtcggccc 8040
accgctctgc caggetaccg aggccgcgc cggcctctg gatgcctcg gcaatgtcca 8100
gtaggtcgcg ggtgctcgg gccaggcgt ctagcctggt caetgtcaca acgtcggcag 8160
ggcgtagggt gtaaacatc ctggccagct cggggcggtc gcgcctggtg ccggtgatct 8220
tctcggaaaa cagcttgggt cagccggccg cgtgcagttc ggcccgttgg ttggtcaagt 8280
cctggctcgt ggtgctgacg cgggcatagc ccagcagcc agcggcgggc ctcttgttca 8340
tggcgtaatg tctccggttc tagtcgaag tattctactt tatgcgacta aaacacgcga 8400
caagaaaacg ccaggaaaag ggcaggcgcg cagcctgctg cgtaaacttag gacttgtgag 8460
acatgtcgtt ttcagaagac ggtgcaactg aacgtcagaa gccgactgca ctatagcagc 8520

```

[0006]

ggaggggttg gatcaaagla etttgatccc gaggggaacc ctgtggttgg catgcacata 8580
 caaatggacg aacggataaa ccttttcacg cccittitaa tatccgttat tctaataaac 8640
 gctcttttct cttaggttta cccgccataa taccctgtca aacactgata gtttaaactg 8700
 aaggcgggaa acgacaatct gatccaagct caagctgctc tagcattcgc cattieagget 8760
 gcgcaactgt tgggaaggge gatcgggtgc ggcctcttcg ctattacgcc agctggcgaa 8820
 agggggatgt getgcaagge gattaagttg ggtaacgcca gggtttccc agtcacgacg 8880
 ttgtaaaaag acggccagtg ccaagcttgg cgtgcctgca ggtaacatg gtggagcaag 8940
 acacacttgt ctactccaaa aatatcaaag atacagctct agaagaccaa agggcaattg 9000
 agacttttca acaaagggtg atatccggaa acctcctcgg attcattgc ccagctatct 9060
 gtcactttat tgtgaagata gtggaaaagg aagggtggctc ctacaaatgc catcattgcg 9120
 ataaaggaaa ggccatcgtt gaagatgcct ctgccgacag tggccccaaa gatggacccc 9180
 caccacgag gagcatcgtg gaaaaagaag acgttccaac cacgtcttca aagcaagtgg 9240
 attgatgtga taacatgggt gagcacgaca cacttgtctc ctccaaaaat atcaaagata 9300
 cagctcaga agaccaaagg gcaattgaga ctttcaaca aagggttaata tccggaaacc 9360
 tectcggatt ccattgcccc getatctgtc actttattgt gaagatagtg gaaaaggaag 9420
 gtggctccta caaatgcca catctgcgata aaggaaaggc catcgtigaa gatgctctg 9480
 ccgacagtgg tcccaaagat ggacccccac ccacgaggag catcgtggaa aaagaagacg 9540
 ttccaaccac gtcttcaaag caagtggatt gatgtgatat ctccactgac gtaagggatg 9600
 acgcacaatc cactatcct tcgcaagacc ctctctctat ataaggaagt tcatttcatt 9660
 tggagaggac ctgactctc gaggatcccc gggtaaccgg cccccctcg aggcgcgcca 9720
 agctacaagt ttgtacaaaa aagcagctg ctccgaattc gcccttacc atggctagt 9780
 cagattacaa agatcacgat ggagattaca aagatcacga tattgattac aaagatgatg 9840
 atgataaaat ggcaccaaag aaaaagagaa aagttggaat tcattggtgt cctgtgcag 9900
 ataagaagta cagtatagga ctgatctcgt gtaaccaatc tgttggatgg gctgtgatta 9960
 ctgatgaata taaggttcca tctaaaaagt tcaaaagtgt gggtaacact gatagacact 10020
 ctattaagaa gaatcttata ggagctcttt tgtttgatag tggtaaaaca gctgaggcaa 10080
 ccagacttaa gaggacagca agaaggagat acaccaggag aaaaaacaga atctgttatt 10140
 tgcaagaaat cttttctaag gagatggcta aggttgatga ttcatcttcc cataggttgg 10200
 aagagagttt tctcgtggaa gagataaaaa agcatgaaag acacctata ttcggaaata 10260
 ttgttgatga agtggcttac catgagaagt acccaacaat ctatcactta agaaaaaagc 10320
 ttgttgatc taccgataag gctgatctta ggttaactca ccttgccttg gcaatataga 10380
 tcaagtttag aggtcacttc ttaatagaag gagatttgaa tctgataac tcagatgtg 10440
 ataagttgtt tattcaactc gtgcagacat ataaccaact tttcgaagag aatccaatca 10500
 acgcttcagg agttgatgct aaggcaatcc tcagtccaag acttagtaaa tctaggagac 10560
 tcgaaaatct tafagctcag ttgcctggag aaaagaagaa tggtttgttc ggtaacctca 10620
 tgcactctc tctcggctct acaccaaact tcaagtcaaa ctctgatctt gctgaagatg 10680
 caaagttgca actctctaa gatacttacg atgatgatt ggataatctc ttagectcaga 10740
 ttggagatca atatgctgat cttttcttg ctgcaagaa ccttcagat gctatccttt 10800
 tgagtgatat cttgagagtt aacactgaaa taacaaagge tctctttca gcaagtatga 10860
 tcaaaagata cgatgagcat caccaagatt tgaactctct aaaagcttgg gttagacaac 10920
 agctcccaga aaagtataag gaaatcttct ttgatcagtc taagaatggg tatgctgggt 10980
 acatcgaagg aggtgcatca caagaagagt tctacaaatt cattaagcct atcttgaaa 11040
 agatggatgg aaccgaagag cttttgggta agcttaacag agaggatctc ttaagaaaac 11100
 agaggacttt cgataatggt tctatccctc atcaaatata ccttggtgaa ttgatgcta 11160

[0007]

```

ttctcaggag acaagaggat ttctacccat tccttaagga taacagagaa aagatcgaga 11220
agatthttgac ttttagaatt ccttattacg ttggaccact tgctaggggt aacagtaggt 11280
tcgcatggat gactaggaag tetgaagaga caattacccc ttggaattht gaagagttg 11340
tggataaagg tgcctctgca cagctcttth ttgaaagaat gacaaacttc gataagaatc 11400
ttcctaacga gaaggctctt ccaaaagcatt ctctthttgta cgaatacttc actgltgaca 11460
atgagcttac aaaggthtaag tatgtgaccg aaggaatgag gaagccagct thtctthtcag 11520
gagagcagaa aaaggcaatt gttgatctct tattcaagac taacagaaag gttacagtga 11580
aacaactthaa ggaagattac thcaaaaaga tctgagtctt cgattctgtt gaaatactag 11640
gagtggagga tagattcaat gcttctthtg gtacttatca tgatctthtg aaaattatca 11700
aagataagga thtctthtgat aatgaggaga acgaagatac cctcagggat atcgthttga 11760
ctctcacatt attcgaagat agagagatga thgaagagag gthgaagaca tacctcacc 11820
ctctcagatga thaaagthtat aagcaactta agaggagaag gtataccgga tggggtagat 11880
tatctaggaa acttateaac ggaataaggg ataagcagtc aggaaagact attctcgatt 11940
ctthaaagag tgatggatc gctaacagaa acttcatgca actcatccat gatgattctc 12000
thactththaa ggaagatact caaaaagctc aggtthtcagg acagggagat agthlacalg 12060
agcacattgc thaatctthgca ggatcacctg ctattaagaa ggtatththt caaaactgta 12120
aggthtgtaga tgaactthgtt aaggthtatgg gaagacataa gccagagAAC atcgtgatag 12180
aaatggctag ggagaaccaa actacacaga agggthcaaaa gaattcaaga gaaaggatga 12240
agagaatcga agagggcaatt aaggagctcg thtcacaaat ththaaaggaa catccagthg 12300
agaatacaca attgcagAAC gaaaagctth actthgacta cctccagAAC ggaagagata 12360
tgtatgtgga thcaagagthg gatathtaata ggtctcagthg thacgatgth gatcacatcg 12420
tgcctcagtc thtctthtaag gatgattcaa tctgataaaa agthctthact agatcagata 12480
agaatagggg thaatctgat aacgtgccat cagaagaggt thtgaaaaag atgaaaaact 12540
actggagaca gctctctaac gctaaagthg thactcaaaag aaaathcgat aatctcaca 12600
aggctgaaag gggaggctct thctgagthgg athaaaggcagg ththththaaag agacagthgg 12660
thgaaactag gcaaaattaca aagcatgtgg ctcaaatctc thgattcaagg atgaatacta 12720
agtacgatga aaacgataag thgatcagag aggtthaaagt gattacattg aagagthaac 12780
tcgthctctga thtctcagaaag gatthctcagth thtacaaaagt gagggagatt aataactacc 12840
atcacgctca cgatgcatac ththaatgctg thgtgggAAC agcatthgat aaaaagthac 12900
ctaagctcga atctgagtht gththacggag athtacaagt thacgatgth agaaagatga 12960
tcgctaagtc agaacaagag athggaaagg ctactgcaaa gthactththt thatagthaca 13020
thtatgaacth thtcaagacc gaaatcacth thgctaacgg agagatcaga aaaaggccac 13080
thtatagaAAC caatggthgaa actggagaga thgtthtgga thaaaggthg gaththcga 13140
cagthtaggaa ggtgctthct atgctcaggt thaacatctg gaaaaagaca gaagthcaaa 13200
ccggaggtht tagthaggag ththctctct caaaagagaaa ctcagataag ctthctgctc 13260
ggaaaaagga thgggacct aagaagthac gaggtththg thgtccaaca gthgctthatt 13320
ctgtgttggt thgtggcaaaa gthgaaaagg gaaagthctaa gaagctthaa thctgtgaaag 13380
agctththgg thactactatt atggaaagat cactactcga gaagaacct athgattct 13440
thggaagctaa gggatacaaa gaggtthaaag aggatcttht ththaaagct ccaaagthact 13500
cactctctega attagagaat ggtagaaaga ggaatgtggc thgtgcagga gaactcaaaa 13560
agggthaacga gctctgctct cctthctaaat atgtthaaath thctthactth gcatctcatt 13620
acgaaaagct thaaaggthca ccagaagata atgagcaaaa gcagthattc gthgaaacagc 13680
ataaacacta cctcagatgaa atcatagagc aaatactcaga gththagthag agagthattt 13740
thgctgatgc aaactthgat aaagthctct ctgctthaca thagcatgag gataagccta 13800

```

[0008]

```

taaggaaca agcagagaac attatccacc tttttacett gactaatctc ggagctccag 13860
ctgcttttaa gtacttcgat accactattg atagaaagag gtatacaagt accaaggaag 13920
ttctcgatgc aactcttata catcagtcta tcaccggtct ttatgagact agaatagatt 13980
tgtcacagtt aggaggagat tcaagagcag atccaaagaa gaaaagaaaa gttagatcta 14040
gctagtcgac gtccgatcgt tcaaacattt ggcaataaag tttettaaga ttgaatectg 14100
ttgccggtct tgcgatgatt atcatataat ttctgttgaa ttacgttaag catgtaataa 14160
ftaacatgta atgcgatgac ttatttatga gatgggtttt tatgattaga gteccgcaat 14220
tatacattta atacgcgata gaaaacaaaa tatagecgcg aaactaggat aaattatcgc 14280
gcgcggtgtc atctatgtta ctagatcggg aattgatccc ccctcgacag cttccgaaa 14340
gggccaattc caacttttgt atacaaagtt gccgagctcg cggccgatg ttgttgttac 14400
cagaaagtaa ataaatgttc aatctctgat gttctcaagt aagtgagttt tattgggaat 14460
aatattaact tatgttcttc ttgcatttga tttctttgcc gctctcttct tctatcttaa 14520
atctgtgtat actatttcac tattgggett tttattagtc tataatggga ctcaaaaataa 14580
ggctttggcc cacatcaaaa agataagtca caaatcaaaa ctaaattcag agtcttttct 14640
cccacatcgg tcaactgtact ctttttgtgt ttgtttatat attacacgaa ccgatctttg 14700
gtacggagac ggagtcgatt cgtctcgttt tagagctaga aatagcaagt taaaataagg 14760
ctagtccgtt atcaacttga aaaagtggca ccgagtcggt gcttttttct tagaccacgc 14820
tttcttgtag aaagtggca ttactgcagg tcgacggacc cagctttctt gtacaaagtg 14880
tcgataattc ctttaattaac tagttctaga gcggccgccc accgcggtgg agctcgaatt 14940
tccccgatcg ttcaaacatt tggcaataaa gttcttaag attgaatect gttgccggtc 15000
ttgcgatgat tatcatataa tttctgttga attacgttaa gcatgtaata attaacatgt 15060
aatgcatgac gttatttatg agatgggttt ttatgattag agtcccgcaa ttatacattt 15120
aatacgcgat agaaaacaaa atatagecgc caaactagga taaattatcg cgcgcggtgt 15180
catctatgtt
actgaatt

```

15198<210> 7

<211> 16818

<212> DNA

<213> 人工序列

<220>

<223> 重组质粒

<400> 7

```

gggccaaca ctgatagttt aaactgaagg cgggaaacga caatctgatc caagctcaag 60
ctaagcttgc atgccgtcag gatatcgtgg atccaagctt gccacgtgcc gccacgtgcc 120
gccacgtgcc gccacgtgcc tctagaggat ccattctccac tgacgtaagg gatgacgcac 180
aatcccacta tcttccgcaa gaccttctct ctatataagg aagtctcattt catttggaga 240
ggacacgctg ggatccccc aaatcggcgg aatgaaagcg ttaacggcca ggcaacaaga 300
ggtgtttgat ctcatccgtg atcaccatcag ccagacaggt atgccgccga cgcgtgcgga 360
aatcgcgcag cgtttggggt tccgttcccc aaacgcggtt gaagaacatc tgaaggcgc 420
ggcacgcaaa ggcgttattg aaattgttct cggcgcatac cgcgggatcc gtctgttgca 480
ggaagaggaa gaagggttgc cgctggtagg tcgtgtggct gccggtgaac cgtcgcgcgc 540
cccccgacc gatgtcagcc tgggggacga gctccactta gacggcgagg acgtggcgat 600
ggcgcgatgc gacgcgctag acgatttcca tctggacatg ttgggggacg gggattcccc 660

```

[0009]

```

gggtccggga tttaccccc acgactccgc cccctacggc gctctggata tggccgactt 720
cgagtttgag cagatgttta ccgatgccct tggaaatgac gagtacggtg gggatccgtc 780
tgctggagac atgagagctg ccaaccttgg gccaaagccg ctcgatgatc aacgetctaa 840
gaagaacage ctggccttgt ccttgacggc cgaccagatg gtcagtgeet tgttgatgc 900
lgagcccccc alaactelall ccgagialga tccclaccaga cccllcaglg aagcttccgat 960
gatgggetta ctgaccaacc tggcagacag ggagctgggt cacatgatca actgggcgaa 1020
gagggtgcca ggtcttgggg atttgacct ccatgatcag gtccacetic tagaatgtgc 1080
ctggctagag atcctgatga ttggctctgt ctggcgctcc atggagcacc cagggaagct 1140
actgtttgct cctaacttgc tcttggacag gaaccagga aatgtgtag agggcatggt 1200
ggagatcttc gacatgetgc tggctacate atctcggttc cgcgatgafa atctgcaggg 1260
agaggagttt gtgtgcctca aatctattat ttgtctaat tetggagtgt acacatttct 1320
gtccagcacc ctgaagtctc tggaaagaaa ggaccatac caccgagtc tggacaagat 1380
cacagacact ttgatccacc tgatggccaa ggcaggcctg accctgcagc agcagcacca 1440
gcggctggcc cagctcctcc teatcctctc ccacatcagg cacatgagta acaaaggcat 1500
ggagcactcg tacageatga aglgaagaaa cglgglgccc ctctalgacc tgcctgctga 1560
gatgctggac gccaccgcc tacatgcgcc cactagecgt ggaggggcat ccgtggagga 1620
gacggaccaa agccacttgg ccaactcggg ctctactica tgcattcct tgcaaaagta 1680
ttacatcacg ggggaggcag agggtttccc tggcacagtc tgagagctcc ctggcggaat 1740
tcccagagat gttagctgaa atcatcacta atcagatacc aaaatattca aatggaaata 1800
tcaaaaagct tetgtttcat caaaaatgac tegacctaac tgagtaagct agcttgttcg 1860
agtattatgg cattgggaaa actgttttct ttgtaccatt tgttgtgctt gtaatttact 1920
gtgtttttta ttcggttttc gctatcgaae tgtgaaatgg aatggatgg agaagagtta 1980
atgaatgata tggctctttt gttcattctc aaattaatat tatttgtttt ttctcttatt 2040
tgttgttgtt tgaatttgaa attataagag atatgcaaac attttgtttt gagtaaaaat 2100
glgtcaaatc glggcctcta atgaccgaag ttaatatgag gagliaaaaaca ctagatcccc 2160
aaacaagctt gaaactgaag gcgggaaacg acaatctgat catgagcgga gaattaaggg 2220
agtcacgtta tgacccccgc cgatgacgcg ggacaagccg ttttacgitt ggaactgaca 2280
gaaccgcaac gttgaaggag ccactcagcc gcgggtttct ggagttaat gagctaagca 2340
catacgtcag aaaccattat tgcgcgttca aaagtcgctt aaggtcacta tcagctagca 2400
aatatttctt gtcaaaaatg ctccactgac gtccataaa tteccctcgg taiccaatta 2460
gagtctcata ttcactctca atccaaataa tetgaccgg atccgctaga ggatctcgac 2520
ctgcaagatc ccggggggca atgagatag aaaaagcctg aactcaccgc gacgtctgtc 2580
gagaagtttc tgatcgaaaa gttcgacagc gtctccgacc tgatgcagct ctccggaggc 2640
gaagaatctc gtgtttcag ctccgatgta ggagggcgtg gatatgtct cggggtaaata 2700
agctgcgcgc atggtttcta caaagatcgt latgtllalc ggcactttgc atcggccgcg 2760
ctcccagatc cggaaagtgt tgacattggg gaattcagc agagcctgac ctattgcate 2820
tcccgcgctg cacagggtgt cacgttgcga gaectgctg aaaccgaact gcccgtgtt 2880
ctgcagccgg tccggaggc catggatgcg atcgtcgcgg ccgatcttag ccagacgagc 2940
gggttcggcc cattcggacc gcaaggaate ggtcaataca ctacatggcg tgatttcata 3000
tgcgcgattg ctgateccca tgtgtateac tggcaaaactg tgatggaaga caccgtcagt 3060
gcgtccgtcg cgcaggctct cgatgagctg atgctttggg ccgaggactg ccccgaagtc 3120
cggcacctcg tgcacgcgga tttcggctcc aacaatgtcc tgacggacaa tggccgcata 3180
acagcggtea ttgaactggag cgaggcagat ttcggggatt cccaatacga ggtcgcacac 3240
atcttcttct ggaggcctgt gttgcttgt atggagcagc agacgcgcta ctteggagcg 3300

```

[0010]

```

aggcatccgg agcttgcagg ategccgagg ctccgggagt atatgctcgg cattggcttt 3360
gaccaactct atcagagctt ggttgacggc aatttcgatg atgcagcttg ggcgcagggt 3420
cgatgacgac caatcgtccg atccggagcc gggactgtcg ggcgtacaca aatcggccgc 3480
agaagcgagg ccgctggac cgaigcctgt gtagaagtac tcgccgatag tggaaaecga 3540
cgccccagea ctctccggg atcttgagg tgatgtaaca tgateacaag ctgatcccc 3600
gaatttcccc gatcgttcaa acatttggca ataaagttc ttaagattga atcctgttgc 3660
eggtcttggc atgattatca tataatttct gttgaattac gtttaageatg taataattaa 3720
catgtaatgc atgacgttat ttatgagatg ggtttttatg attagagtec cgcaattata 3780
catttaatac gcgatagaaa acaaaatata gcgcgcaaac taggataaat tatcgcgcgc 3840
ggtgtcatct atgttactag atcgggaatt gateccccct cgacagcttg catgccgctt 3900
gggctgcagg tcgaggctaa aaaactaatc gcattatcat cccctcgacg tactgtacat 3960
ataaccactg gttttatata cagcagtact gtacatataa ccactggttt tatatacagc 4020
agtgcagcta ctgtacatat aaccactggt tttatataca gcagtactgt acatataacc 4080
actggtttta tatacagcag tcgaggtaa gattagatat gatatgtata tggatatgta 4140
tatggtggta atgccatgta atatgctega ctctaggatc ttcgcaagac ccttctctta 4200
tataaggaag ttcatttcat ttggagagga cacgctgaag ctagtgaact ctagcctega 4260
ggcgcgccaa getatcgcgc gccaaactac ccagcttict tgtacaaagt ggcegaattc 4320
gcccttcacc atggctagtt cagattacaa agatcacgat ggagattaca aagatcacga 4380
tattgattac aaagatgatg atgataaaat ggcaccaaa gaaaaagaga aagttggaat 4440
tcattggtgt cctgctgcag ataagaagta cagtatagga cttgatatcg gtaccaattc 4500
tgtttggatg gctgtgatta ctgatgaata taagttcca tctaaaaagt tcaaagtgtt 4560
gggtaacact gatagacact ctattaagaa gaatcttata ggagctcttt tgtttgatag 4620
tggtgaaaca gctgaggcaa ccagacttaa gaggacagca agaaggagat acaccaggag 4680
aaaaaacaga atctgttatt tgaagaaat ctttctaat gagatggeta aggtttgatga 4740
ttcatttttc cataggttgg aagagagttt tctcgtggaa gaggataaaa agcatgaaag 4800
acacctata ttccgaaata ttgttgatga agtggettac catgagaagt acccaacaat 4860
ctatcactta agaaaaaagc ttgttgattc taccgataag gctgatctta ggttaatcta 4920
ccttgctttg gcacatatga tcaagtttag aggtcacttc ttaatagaag gagatttgaa 4980
tcctgataac tcagatgttg ataagttgtt tatteaactc gtgcagacat ataaccaact 5040
ttfegaagag aatccaatca acgctcagg agttgatgct aaggeaatcc tcagtgcaag 5100
acttagtaaa tctaggagac tcgaaaatct tatagctcag ttgccctggag aaaagaagaa 5160
tggtttgttc ggtaacctca tcgcactctc tctcgttctt acaccaaact tcaagtcaaa 5220
cttcgatctt gctgaagatg caaagttgca actctctaag gatacttaag atgatgattt 5280
ggataatctc tttagctcaga ttggagatca atatgctgat ctttttcttg ctgcaaagaa 5340
cctttcagat gctatccttt tgagtgatat cttgagagtt aacactgaaa taacaaaggc 5400
tcctctttca gcaagtatga tcaaaagata cgatgagcat caccaagatt tgaactctct 5460
aaaagctttg gttagacaac agctcccaga aaagtataag gaaattttct ttgatcagtc 5520
taagaatggt tatgctggtt acatcgatgg aggtgcaatc caagaagagt tctacaaatt 5580
cattaagcct atcttggaaa agatggatgg aaccgaagag cttttggta agcttaacag 5640
agaggatctc ttaagaaaac agaggacttt cgataatggt tctatccctc atcaaatata 5700
ccttgggtgaa ttgcattgta ttctcaggag acaagaggat ttctacccat tccttaagga 5760
taacagagaa aagatcgaga agattttgac ttttagaatt ccttattacg ttggaccact 5820
tgctaggggt aacagtaggt tcgcatggat gactaggaag tctgaagaga caattacccc 5880
ttggaatttt gaagagggtg tggataaagg tgcttctgca cagtctttta ttgaaagaat 5940

```

[0011]

gacaaacttc gataagaatc ttccctaacga gaaggttctt ccaaagcatt ctcttttgta 6000
cgaatacttc actgtgtaca atgagcttac aaaggtaaag tatgtgaccg aaggaatgag 6060
gaagccagct tttctttcag gagagcagaa aaaggcaatt gttgatctct tattcaagac 6120
taacagaaaag gttacagtga aacaacttaa ggaagattac ttcaaaaaga tcgagtgcct 6180
cgallctggt gaaatalcag gagtggagga tagattcaal gcttctttgg glacttalca 6240
tgatcttttg aaaattatca aagataagga tttctttgat aatgaggaga acgaagatat 6300
cctcgaggat atcgttttga ctctcacatt attcgaagat agagagatga ttgaagagag 6360
gttgaagaca tacgctcacc tcttcgatga taaagttaig aagcaactta agaggagaag 6420
gtataccgga tgggtagat tatctaggaa acttatcaac ggaataaggg ataagcagtc 6480
aggaaagact attctcgatt tcttaaagag tgatggattc gctaacagaa acttcatgca 6540
actcatccat gatgattctc ttacttttaa ggaagatate caaaaagctc aggtttcagg 6600
acagggagat agtttcatg agcacattgc taatcttgea ggatcacctg ctattaagaa 6660
gggtatttta caaactgtta aggttgtgga tgaacttgtt aaggttatgg gaagacataa 6720
gccagagaac atcgtgatag aatggetag ggagaaccaa actacacaga agggtcacaaa 6780
gaattcaaga gaaaggatga agagaatcga agagggaatl aaggagctcg gttcacaaal 6840
tcttaaggaa catccagttg agaatacaca attgcagaac gaaaagcttt acttgtacta 6900
cctccagaac ggaagagata tgtatgtgga tcaagagttg gatattaata ggctcagtga 6960
ttacgatggt gatcacatcg tgcctcagtc tttctttaag gatgattcaa tcgataacaa 7020
agttcttact agatcagata agaatagggg taaatctgat aacgtgceat cagaagaggt 7080
tgtaaaaag atgaaaaact actggagaca gctcctcaac gctaagttga tcaactcaaag 7140
aaaattcgat aatctcacia aggetgaaag gggaggtctt tctgagttgg ataaggcagg 7200
ttttattaag agacagttgg ttgaaactag gcaaattaca aageatgtgg cteaaatcct 7260
tgattcaagg atgaatacta agtacgatga aaacgataag ttgatcagag aggtttaaagt 7320
gattacattg aagagtaagc tegtcttga ttccagaaag gattttcagt tctacaaagt 7380
gagggagatt aalaactacc atcacctca cgalgcatal cttaalgctg ttgtgggaac 7440
agcattgate aaaaagtate ctaagctcga atctgagtc gtttacggag attacaaagt 7500
ttacgatgtg agaaaatga tcgctaagtc agaacaagag atttgaaagg ctactgcaaa 7560
gtactttttc tatagtaaca ttatgaactt ttcaagacc gaaatcaett tggctaacgg 7620
agagatcaga aaaaggccac ttatagaaac caatggtgaa actggagaga ttgtttggga 7680
taagggtaga gatttegeaa cagttaggaa ggtgctttct atgcctcagg ttaacatcgt 7740
gaaaaagaca gaagtcaaa ccggaggitt tagtaaggag tctatctctc caaagagaaa 7800
ctcagataag cttategcta ggaaaaagga ttgggaccct aagaagtacg gaggttttga 7860
tagtccaaca gttgcttatt ctgtgttggg tgtggcaaaa gttgaaaagg gaaagtctaa 7920
gaagcttaag tetgtgaaag agettttggg tatcactatt atggaaagat catcattcga 7980
gaagaaccct attgatttct tggaaactaa gggatacaaa gagglttaaga aggatctlat 8040
tattaagctt ccaaagtact cactcttcca attagagaat ggtagaaga ggatgttggc 8100
tagtgcagga gaactccaaa agggtaacga getcctctc ccttctaaat atgttaattt 8160
tctttacttg gcatectatt acgaaaagct taagggttca ccagaagata atgagcaaaa 8220
gcagttatcc gttgaacagc ataaacacta cctcgatgaa atcatagagc aaatatcaga 8280
gtttagtaag agagttattt tggetgatgc aaacttggat aaagtgcctc ctgettacaa 8340
taagcataga gataagccta taagggaaca agcagagaac attatccacc tttttactct 8400
gactaatctc ggagctccag ctgcttttaa gtacttcgat accactattg atagaaagag 8460
gtatacaagt accaaggaag ttctcgatgc aactcttata cateagctca tcaccggtct 8520
ttatgagact agaatagatt tgtcacagtt agggaggagat tcaagagcag atccaaagaa 8580

[0012]

gaaaagaaaa gttagatcta gctagtcgac gtccgatcgt tcaaacattt ggcaataaag 8640
 tttcttaaga ttgaatcctg ttgccggtct tgcgatgatt atcatataat ttctgttgaa 8700
 ttacgttaag catgtaataa ttaacatgta atgcatgacg ttatttatga gatgggtttt 8760
 tatgattaga gteccgcaat tatacattta atacgcgata gaaaacaaaa tatagcgcgc 8820
 aaactaggat aaattategc gcgcggtgtc atctatgtta ctagatcggg aattgatecc 8880
 ecctegacag cttccgaaa gggcgaatc caacttttgt atacaaagti gccgagctcg 8940
 eggccgcctc gttgaacaac ggaaactcga cttgccttcc gcacaataca tcatttcttc 9000
 ttagcttttt ttcttcttct tcttctctac agtttttttt tgtttatcag cttacatttt 9060
 cttgaaccgt agcttttctt ttcttctttt taactttcca ttccggagttt ttgtatcttg 9120
 tttcatagtt tgtcccgagg ttagaatgat taggcacgca accttcaaga atttgattga 9180
 ataaaacatc ttcattctta agatatgaag ataactctca aaaggccctt gggaatctga 9240
 aagaagagaa gcaggcccat ttatatggga aagaacaata gtatttctta tataggcca 9300
 tftaagttag aaacaatctt caaaagtccc acatcgctta gataagaaaa cgaagctgag 9360
 tttatataca gctagagtcg aagtagtgat tgagacggag tegtctctc tegttttaga 9420
 gctagaaata gcaagttaaa ataaggctag tccgttatca acttgaaaaa gtggcaccga 9480
 gtccgtgctt tttttctaga cccagcttct ttgtacaaag ttgtcgataa ttcttaatta 9540
 actagtcgat ccaggcctcc cagctttctg ccgtatcctc ggtttcgaca acgttctgca 9600
 agttcaatgc atcagtttca ttgcccacac accagaalcc tactaagttt gagtattatg 9660
 gcattggaaa agctgttttc ttctatcatt tgttctgctt gtaatttact gtgttcttct 9720
 agtttttgtt ttcggacatc aaaatgcaaa tggatggata agagttaata aatgatatgg 9780
 tccttttgtt cattctcaaa ttattattat ctgttgtttt tactttaatg ggttgaattt 9840
 aagtaagaaa ggaactaaca gtgtgatatt aagtgcaat gttagacata taaaacagtc 9900
 tttaccctct cttttgttat gtcttgaatt ggtttgttct tteacttate tgtgtaatea 9960
 agtttactat gagtctatga tcaagtaatt atgcaatcaa gtttaagtaca gtatagctt 10020
 tttgtgtcga ggggttacct ttctccataa taatgtgtga gtagttccca gataaggaa 10080
 ttagggttcc tatagggttt cgcctcatgt ttgagcatat aagaaacct tagtatgtat 10140
 ttgtatttgt aaaatacttc tatcaataaa atttctaatt cctaaaacca aatccagta 10200
 ctaaaalcca gatccccga attaatctcg cgliaattca glacatlaaa aacglccgca 10260
 atgtgttatt aagttgteta agcgtcaatt tgtttacacc acaatatate ctgccaccag 10320
 ccagccaaca gctccccgac eggcagctcg gcacaaaatc accactgat acaggcagcc 10380
 catcagtcgg ggacggcgtc agcgggagag ccgttgttaag gcggcagact ttgctcatgt 10440
 taccgatgct attcgggaaga acggcaacta agctgccggg ttgaaacac ggatgatctc 10500
 ggggagggta gcatgttgat tghtaacgat acagagcgtt getgctctgt atcaecggcg 10560
 tttcaaaatc ggctccgtcg atactatggt atacgccaac ttgaaaaca actttgaaaa 10620
 agctgttttc tggatattta ggtttttaga tgcagggaac agtgaattgg agttctctt 10680
 gttataatta gettcttggg gtatctttaa atactgtaga aaagaggaag gaaataataa 10740
 atggetaaaa tgagaatate accggaattg aaaaaactga tcaaaaaata ccgtctgcta 10800
 aaagalacgg aaggaatgic tcttgcctag gtatataagc tggggggaga aaalgaaaac 10860
 ctatatttaa aatgacgga cagccggtat aaagggacca cctatgatgt ggaecgggaa 10920
 aaggacatga tgctatggct ggaaggaaag ctgctcttcc caaaggtcct gaacttgaa 10980
 eggcatgatg getggageaa tctgctcatg agtgaggecg atggcgtcct ttgctcgaa 11040
 gagtatgaag atgaacaaag ccttgaaaag attatcgagc tgtatgogga gtgcatcagg 11100
 ctctttcaet ccategacat atcggattgt ccctatacga atagcttaga cagccgctta 11160
 gccgaattgg aftacttact gaataacgat ctggccgatg tggattgcca aaactgggaa 11220

[0013]

gaagacactc catttaaaga tccgcgcgag ctgtatgatt ttttaaagac ggaaaagccc 11280
 gaagaggaac ttgtcttttc ccacggcgac ctgggagaca gcaacatctt tgtgaaagat 11340
 ggcaaagtaa gtggctttat tgatcttggg agaageggca gggeggacaa gtggtatgac 11400
 attgccttct gcgtccggte gatcagggag gatacggggg aagaacagta tgteagacta 11460
 ttttttgact tactggggat caagccgat lgggagaaaa laaaatatta tattttactg 11520
 gatgaattgt tttagtaect agaatgcatg accaaaaatcc cttaacgtga gttttctgtc 11580
 cactgagcgt cagaccccgat agaaaagatc aaaggatctt cttagatcc tttttttctg 11640
 cgcgtaatct gctgcttgea aacaaaaaaa ccaccgctac cagcgggtgtt ttgtttgccc 11700
 gatcaagagc taccaactct ttttccgaag gtaactggct tcagcagagc gcagatacca 11760
 aatactgtcc ttctagtgtt gccgtagtta ggccaccact tcaagaacte tgtagaccgc 11820
 cctacatacc tcgctctget aatcctgtta ccagtggtct ctgccagtgg cgataagtcg 11880
 tgtcttaccg ggttggactc aagacgatag ttaccggata aggcgcagcg gtcgggetga 11940
 acgggggggt cgtgcacaca gccagccttg gagegaacga cctacacega actgagatac 12000
 ctacagcgtg agctatgaga aagcgcacg ctccccgaag ggagaaaggc ggacaggtat 12060
 ccggtlaagc gcagggcggg aacaggagag cgcacgaggg agcttccagg gggaaacgcc 12120
 tggtatcttt atagtctgtt cgggttttgc caectctgac ttgagcgtcg atttttgtga 12180
 tgctcgtcag gggggcggag cctatggaaa aacgccagca acgcggcctt tttacggttc 12240
 ctggcctttt gctggccttt tgctcacatg ttctttctg ctgtatcccc tgattctgtg 12300
 gataaccgta ttaccgcctt tgagttagct gataccgctc gccgcagccg aacgaccgag 12360
 cgcagcaggt cagtgagega ggaagcggaa gagegcctga tgcggtattt tctcttaccg 12420
 catctgtcgc gtatttcaca ccgcatatgg tgcactctca gtacaatctg ctctgatgcc 12480
 gcatagttaa gccagtatac actccgctat cgtactgtga ctgggtcatt gctgcgcccc 12540
 gacaccgcc aacaccgcct gacgcgcctt gacgggcttg tctgctccc gcctccgctt 12600
 acagacaagc tgtgaccgte tccgggagct geatgtgtca gaggttttca ccgtcaccac 12660
 cgaaacgcgc gaggcagggt gcccttgalgt gggcgccggc ggtcagatgg cgacggcgcg 12720
 gcttgtccgc gccctggtag attgcctgge cgtaggccag ccatttttga gcggccagcg 12780
 gccgcgatag gccgacgcga agcggcgggg cgtagggagc gcagcagccg aagggttagc 12840
 gctttttgca gctcttcggc tgtgcgctgg ccagacagtt atgcacaggc caggcgggtt 12900
 ttaagagttt taataagttt taaagagttt taggcggaaa aatcgccttt tttctctttt 12960
 atateagtea cttacatgtg tgaccggttc ccaatgtacg gctttgggtt cccaatgtac 13020
 gggttccggg tcccaatgta eggcttttgg ttcccaatgt acgtgctatc cacaggaaag 13080
 agaccttttc gacctttttc cctgtctagg gcaatttgc cttagcatct ctccgtacat 13140
 taggaaccgg cggatgcttc gccctcgatc aggttgcggt agcgcacgac taggatcggg 13200
 ccagcctgcc ccgctctctc ctccaatcg tactccggca ggtcatttga cccgatcagc 13260
 ttgcgcacgg tgaaacagaa ctctttgaac tctccggcgc tgccaclgcg ttctlagatc 13320
 gtettgaaca accatctgge ttctgccctg cctgcggcgc ggcgtgccag gcggtagaga 13380
 aaacggccga tgcggggate gatcaaaaag taatcggggt gaaccgtcag caegtccggg 13440
 ttcttgcctt ctgtgatctc gcggtaacac caatcagcta gctcgatctc gatgtactcc 13500
 ggccgccccg tttctgctctt tacgatcttg tagcggctaa tcaaggcttc accctcggat 13560
 accgtcacca ggccgcccgt cttggccttc ttctgacgct geatggcaac gtgcgtggtg 13620
 tttaacggaa tgcaggtttc taccaggctc tctttctgct ttcgcccata ggctcggcgg 13680
 cagaacttga gtacgtccgc aacgtgtgga cggaacacgc ggccgggett gtctcccttc 13740
 ccttcccggg atcggttcat ggattcgggt agatgggaaa ccgccatcag taccaggtcg 13800
 taatcccaca cactggccat gccggccggc cctgcggaaa cctctacgtg cccgtctgga 13860

[0014]

```

agctcgtagc ggatcacctc gccagctcgt cggtcacgct tcgacagacg gaaaacggcc 13920
acgtccatga tgctgcgact atcgcgggtg cccacgtcat agagcatcgg aacgaaaaaa 13980
tctggttget cgtcgcctt gggeggctt ctaategacg ggcaccggc tgccggcgt 14040
tgccgggatt ctttgcggat tcgacagcgg gcccttgcg acgattcacc ggggcgtget 14100
tetgeetega tgcgttgcg ctgggeggcc lgcgcggcct lcaacttctc caccaggtea 14160
tcacccagcg ccgcgcgat ttgtaccggg ccggatggtt tgcgaccget caccgcgat 14220
ectcgggctt gggggttcca gtgccattgc agggccggca ggcacccag ccgettaccg 14280
ctgccaacc gcccgttcct ccacacatgg ggcatccac ggcgtcggtg cctggttgt 14340
cttgatttcc catgccgct cctttagccg ctaaaattca tctactcatt tattcattg 14400
ctcatttact ctggtagctg cgcgatgat tcagatagca gctcggtaat ggttctgct 14460
tggcgtaccg cgtacatctt cagcttgggt tgcctctcc ccggcaactg aaagtgacc 14520
cgcttcatgg ctggcgtgtc tgcaggetg gccaacgttg cagccttget gctgcgtgag 14580
ctcggacggc cggcacttag cgtgtttgtg cttttgctca tttctcttt acctcattaa 14640
ctcaaatgag ttttgattta atttcagcgg ccagcgcctg gacctcggg gcagcgtgc 14700
cctcgggttc tgallcaaga acggttgcg cggcggcgcc aglccctggg tagctcaccg 14760
gctgcgtgat acgggactca agaattggca gctcgtacce ggccagcgc tcggcaacct 14820
caccgccgat gcgcgtgct ttgatcgcg gcgacacgac aaaggccgct tgtagcctc 14880
catccgtgac ctcaatgcgc tgcctaacca gctccaccag gtcggcggtg gccatattg 14940
cgtaagggct tggctgcacc ggaatcagca cgaagtggc tgccttgatc gcggacacag 15000
ccaagtccgc ccctggggg gctcctcga tcaactacgaa gtcgcgcgg ccgatggct 15060
tcacgtcgcg gtcaatcgtc gggcggctga tcccgacaac ggttagcgg tgatectcc 15120
gcacggccgc ccaatcgcgg gcactgcctt ggggatcggg ategactaac agaacatcg 15180
ccccggcag ttgcagggcg cgggctagat ggttgcgat ggtcgtctt cctgaccgc 15240
ctttctggtt aagtacagcg ataaccttca tgccttccc ttgcgtattt gttatttacc 15300
tcctgcctc atalacgcag ccaccgatg acgcaagctg ltttactcaa atacacatca 15360
cctttttaga cggcggcgt cggtttctt agcggccaag ctggccggcc aggccgccag 15420
cttggcatca gacaaaccgg ccaggattt atgcagcgc acggttgaga cgtgcgcgg 15480
cggctcgaac acgtaccgg ccgcgatcat ctccgctcg atctctcgg taatgaaaa 15540
cggttcgtcc tggccgtcct ggtgcggtt catgcttgtt cctcttggc ttattctcg 15600
gggcgcgcca ggcgctggc ctccgtcaat gctcctcac ggaaggcacc gcgcgcctg 15660
gcctcgggtg gcgtcactt ctcgctgcg tcaagtgcg ggtacagggt cagcgtatg 15720
acgccaagca gtgcagccg ctctttcac gtcggcctt cctggtgat cagctcggg 15780
gcgtgcgca tctgtcggg ggtgagggt gggcgggggc caaacttca gcctcgggcc 15840
ttggcggcct ccgcccgt cgggtgcg tcatgatta gggaacctc gaactcgca 15900
atgccggcga acacggctca caccatcgg ccggccggcg tggltggltc ggcccacgg 15960
tetgccagge tacgcaggc ccgcgcggc tcttgatgc gctcgcaat gtccagtag 16020
tcgggggtgc tggggccag gcggtctagc ctggtcactg tcacaacgtc gccaggcgt 16080
aggtggtcaa gcactctgc cagctcggg cgtcgcgcc tgggtccgt gatctctcg 16140
gaaaacagct tgggtcagcc ggccgcgtc agttcggccc gttggttgg caagtctgg 16200
tcgtcgggtc tgacgcggg atagcccagc agccagcgg ccgctctctt gttcatggc 16260
taatgtctcc ggttctagt cgaagtaic tactttatgc gaactaaaaca ccgacaaga 16320
aaacgccagg aaaaggcag ggcggcagc tgcgcgtaa cttagactt gtgcgacatg 16380
tcgttttcag aagacgctg cactgaacgt cagaagccga ctgcactata gcagcggag 16440
ggttggatca aagtacttt atcccaggg gaacctgtg gttgcatgc acatacaaat 16500

```

[0015]

ggacgaacgg ataaaccttt tcacgccctt ttaaataatcc gattattcta ataaacgctc 16560
 ttttctetta ggtttaccocg ccaatataatc ctgtcaaaca ctgatagttt aaactgaagg 16620
 cgggaaacga caatctgate caagctcaag ctgctctagc attcgccatt caggctgcgc 16680
 aactgttggg aaggcgatc ggtgcgggcc tcttcgctat tacgccaget ggcgaaaggg 16740
 ggatgtgctg caaggcgatt aagttgggta acgccagggt tttcccagtc acgacgttgt 16800
 aaaacgacgg ecagtgcc

16818<210> 8

<211> 16508

<212> DNA

<213> 人工序列

<220>

<223> 重组质粒

<400> 8

catgcaagct tggcaactggc cgtcgtttta caacgtcgtg actgggaaaa cectggcggt 60
 acccaactta atcgcccttc agcaatccc ctttcgcca getggcgtaa tagcgaagag 120
 gcccgcaccg atcgcccttc ccaacagttg cgcagcctga atggcgaatg ctagagcagc 180
 ttgagcttgg atcagattgt cgtttcccgc cttcagttta aactatcagt gtttgacagg 240
 atatatggc gggtaaacct aagagaaaag agcgtttatt agaataacgg atatttataa 300
 gggcgtgaaa aggtttatcc gttcgtccat ttgtatgtgc atgccaaaca cagggttccc 360
 ctcgggatca aagtactttg atccaacccc tccgetgcta tagtgeagtc ggettctgac 420
 gttcagtgea gcgctcttct gaaaacgaca tgtegcacaa gtcctaagtt aegegacagg 480
 ctgccgccct gcccttttcc tggcgttttc ttgtcgcgtg ttttagtgcg ataaagttaga 540
 atacttgcca ctagaaccgg agacattacg ccatgaacaa gagcgcgccg gctggcctgc 600
 tgggctatgc ccgcgtcagc accgacgacc aggacttgac caaccaacgg gccgaactgc 660
 acgcggccgg ctgcaccaag ctgttttccg agaagatcac cggcaccagg cgegaccgcc 720
 cggagctggc caggatgctt gaccacctac gccctggcga cgttgtgaca gtgaccaggc 780
 tagaccgect ggcccgcagc acccgcgacc tactggacat tgcgagcgc atccaggagg 840
 ccggcgcggg cctgcgtagc ctggcagagc cgtgggcccga caccaccaag ccggccgccc 900
 gcatggtggt gaccgtgttc gccggcattg ccgagttcga gcgttcccta atcatcgacc 960
 gcaccgggag cgggcgcgag gccgccaagg cccgaggcgt gaagtttggc ccccgcctta 1020
 ccctcaaccc ggcaacagatc gcgcacgccc gcgagctgat cgaccaggaa ggccgcaccg 1080
 taaaagaggc ggtgcacatg cttggcgtgc atcgetcagc cctgtaccgc gcacttgagc 1140
 gcagcgagga agtgacgccc accgaggcca ggccgcgccc tgccttccgt gaggacgcat 1200
 tgaccgagc cagcgcctg gcggccgccc agaatgaacg ccaagaggaa caagcatgaa 1260
 accgcaccag gacggccagg acgaaccgtt ttctattacc gaagagatcg aggeggagat 1320
 gatcgcggcc ggttacgtgt tcgagccccc cgcgcacgtc tcaaccglgc ggttgcattg 1380
 aatcctggcc ggtttgtctg atgccaagct ggccgctgg ccggccaget tgcccgcctga 1440
 agaaaccgag cgcgcgctc taaaagggtg atgtgtattt gagtaaaaca gcttgcgtca 1500
 tgcggtcgtt gegtatatga tgcgatgagt aaataaaca atacgcaagg ggaacgcatg 1560
 aaggttatcg ctgtacttaa ccagaaaggc gggtcaggca agacgaccat cgcaaccat 1620
 ctagcccgg ccctgcaact cgcggggccc gatgttctgt tagtcgatte cgatccccag 1680
 ggcagtcccc gcgattgggc ggccgtgcgg gaagatcaac cgctaaccgt tgtcggcatt 1740

[0016]

gaccgcccga cgattgaccg cgacgtgaag gccatcggcc ggcgagactt cgtagtgate 1800
gacggagcgc cccaggcggc ggacttggct gigtccgcga tcaaggcagc cgacttcgtg 1860
ctgattccgg tgcagccaag cccttacgac atatgggcca ccgccgacct ggtggagctg 1920
gtaagcagc gcattgaggt caeggatgga aggctacaag cggecttltg cgtgtgcegg 1980
gcgatcaaag gcacgcgcac cggggtgag gttgcegagg cgctggcegg gtacgagctg 2040
cccattcttg agtcccgtat cacgcagcgc gtgagctacc caggcactgc cgcgcceggc 2100
acaacegttc ttgaatcaga accegagggc gacgctgccc gcgaggtcca ggcgctggcc 2160
gctgaaatta aatcaaaact catttgagtt aatgaggtaa agagaaaatg agcaaaagca 2220
caaacacgct aagtcccggc cgtccgagcg cacgcagcag caaggctgca acgttggcca 2280
gcctggcaga cacgccagcc atgaagcggg tcaactttca gttgccggcg gaggatcaca 2340
ccaagctgaa gatgtacgcg gtacccaag gcaagaccat taccgagctg ctatctgaat 2400
acatcgcgca gctaccagag taaatgagca aatgaataaa tgagttagatg aattttagcg 2460
gctaaaggag gcggcatgga aatcaagaa caaccaggca ccgacgccgt ggaatgcccc 2520
atgtgtggag gaacgggccc ttggccaggg gtaageggct gggttgtctg ccggccctgc 2580
aatggcactg gaacccccaa gcccgaggaa tcggcgtgac ggtcgcgcaac cctccggccc 2640
ggtacaaate ggcgcggcgc tgggtgatga cctgggtggag aagtigaagg ccgcgcaggg 2700
cgcccagcgg caacgcacgc aggcagaagc acgccccggt gaatcgtggc aageggccgc 2760
tgatcgaatc cgcaagaat cccggcaacc gccggcagcc ggtgcgccgt cgattaggaa 2820
gccgcccag ggcgacgagc aaccagattt tttcgttccg atgctctatg acgtgggcac 2880
ccgcgatagt cgcagcatca tggacgtggc cgttttccgt ctgtcgaagc gtgaccgacg 2940
agctggcgag gtgateccgt acgagcttcc agacgggcac gtagaggttt ccgcagggcc 3000
ggccggcatg gccagtggtt gggattacga cctggtaact atggcggttt cccatctaac 3060
cgaatccatg aaccgatacc gggaaaggaa gggagacaag cccggcccgc tgttccgtec 3120
acaagtgcg gacgtactca agttctgccc gcgagccgat ggcggaaage agaaagacga 3180
cctggtagaa acctgcattc ggttaaacac cacgcacgtt gccatgcagc gtacgaagaa 3240
ggccaagaac ggcgcctgg tgacggtatc cgagggtgaa gccttgatta gccgctacaa 3300
gatcgtaaag agcgaaaccg ggcggccgga gtacatcgag atcgagctag ctgattggat 3360
gtaccgcgag atcacagaag gcaagaacce ggacgtgctg acggttcacc ccgattactt 3420
tttgatcgat cccggcatcg gccgttttct ctaccgectg gcacgccgag ccgcaggeaa 3480
ggcagaagcc agatggttgt tcaagacgat ctacgaacgc agtggcagcg ccggagagtt 3540
caagaagttc tgtttcaccg tgcgcaaget gatcgggtca aatgacctgc cggagtagca 3600
ttgaaggag gaggcggggc aggcctggccc gatcctagtc atgcgctacc gcaacctgat 3660
cgagggcgaa gcattccgag gttctctaatg tacggagcag atgctagggc aaattgccct 3720
agcaggggaa aaaggtcgaa aaggtctctt tctgtggat agcacgtaca ttgggaacce 3780
aaagccgtac attgggaacc ggaaccgta cattgggaac ccaaagccgt acattgggaa 3840
ccggtcacac atgtaagtga ctgatataaa agagaaaaaa ggcgattttt ccgcctaaaa 3900
ctctttaaaa ctattaaaa ctcttaaaac ccgctggccc tgtgcataac tgtctggcca 3960
gcgcacagcc gaagagctgc aaaaagcgc tacccttggc tcgctgcgct ccctacgccc 4020
cgccgcttcc cgtcggccta tcgcggccc gcgcccctca aaaatggctg gcctacggcc 4080
aggcaateta ccagggcgag gacaagcgc gccgtgcca ctcgaccgcc ggcgcccaca 4140
tcaaggcacc ctgcctcgcg cgtttcggtg atgacgggta aaacctctga cacatgcagc 4200
tcccggagac ggtcacagct tgtctgtaag cggatgccgg gagcagacaa gcccgtcagg 4260
gcgcgtcagc ggggtgtggc ggggtgtcgg gcgcagccat gaccagctca cgtagcgata 4320
gcggagtgtg tactggctta actatgcggc atcagagcag attgtactga gagtgcacca 4380

[0017]

```

tatgcggtgt gaaataccgc acagatgcgt aaggagaaaa taccgeatca ggcgctcttc 4440
cgcttcctcg ctcaactgact cgctgcgctc ggctggttcgg ctgcggcgag cggatcagc 4500
tcaactcaag gcggaataac ggttatccac agaatacagg gataacgcag gaaagaacat 4560
gtgagcaaaa ggcacgcaaa aggccaggaa ccgtaaaaag gccgcgttgc tggcgTTTTT 4620
ccataggctc cgccccctg acgagcatca caaaaatcga cgctcaagtc agagggtggcg 4680
aaacecgaca ggactataaa gataccaggc gtttccccct ggaagctccc tegtgcgctc 4740
tctgttccg acctgcgcgc ttaccggata cctgtccgcc tttctccctt cgggaagcgt 4800
ggcgcttlet catagctcac gctgtaggia tctcagttcg gtgtaggleg ttgcctccaa 4860
gctgggctgt gtgcacgaac cccccgtca gcccgaccgc tgcgccttat ccggtaacta 4920
tcgtcttgag tccaaccggg taagacacga cttatcgeca ctggcagcag ccactggtaa 4980
caggattagc agagcgaggt atgtaggcgg tgctacagag ttcttgaagt ggiggcctaa 5040
ctacggctac actagaagga cagtattgg tatctgcgct ctgctgaagc cagttacctt 5100
cggaaaaaga gttggtagct cttgatccgg caaacaacc accgctggta gcggtggttt 5160
ttttgtttgc aagcagcaga ttacgcgcag aaaaaaagga tetcaagaag atcctttgat 5220
ctttctacg gggctcgacg ctcaagtggaa cgaaaactca cgttaaggga ttttggctcat 5280
gcattctagg tactaaaaca attcaccag taaaatataa tattttattt tctcccaatc 5340
aggcttgatc ccagtaagt caaaaaatag ctgcacatac tgttcttccc cgatatctc 5400
cctgatcgac cggacgcaga aggcaatgc alaccacttg lccgccctgc cgttctccc 5460
aagatcaata aagccactta ctttgccatc tttcacaag atgttgctgt ctcccaggtc 5520
gccgtgggaa aagacaagtt cctctcggg ctttccgic tttaaaaaat catacagctc 5580
gcgcggatct ttaaatggag tgtctcttc ccagtttctg caatccacat cggccagatc 5640
gttattcagt aagtaatcca attcggctaa gcgctgtct aagctattcg tatagggaca 5700
atccgatatg tcgatggagt gaaagacct gatgcactcc geatacagct cgataatctt 5760
tfcagggett tgttcatctt catactcttc cgagcaagg acgccatcgg cctcactcat 5820
gagcagattg ctccagccat catgccttc aaagtgcagg acctttggaa caggcagctt 5880
tcttccagc catagcatca tgtcttttc cegtccaca tcataggtgg tccctttata 5940
ccggctgtcc gtcattttta aatataggtt ttcattttct cccaccagct tatatacctt 6000
agcaggagac attccttccg tatcttttac gcagegglat ttttcgatca gttttttcaa 6060
ttccggatgat attctcattt tagccattta ttatttctt cctcttttct acagtattta 6120
aagatacccc aagaagctaa ttataacaag acgaactcea attcactgtt ccttgcattc 6180
taaaacctta aataccagaa aacagctttt tcaaagtigt tttcaaagtt ggcgtataac 6240
atagtatcga cggagccgat tttgaaaccg cggatgatcac aggcagcaac gctctgtcat 6300
cgttacaate aacatgctac cctcgcgcgag atcctccttg tttcaaaacc ggcagcttag 6360
ttgccgttct tccgaatagc atcggtaaca tgagcaagt ctgcgcctt acaacggctc 6420
tcccgtgac gccgtcccgg actgatgggc tgcctgtatc gagtggatgat ttigtgccga 6480
gctgccggtc ggggagctgt tggctggctg gtggcaggat atattgtggt gtaaacaat 6540
tgacgcttag acaacttaat aacacattgc ggacgtttt aatgtactga attaacgccg 6600
aatlaattcg ggggatctgg attttagtac tggatttgg ttttaggaat tagaaaTTTT 6660
attgatagaa gtattttaca aatacaata catactaagg gtttcttata tgcctaacac 6720
atgagcgaaa cctatagga accctaatc ccttatctgg gaactactca cacattatta 6780
tggagaaact cgagcttctc gatcgacaga tccggteggc atctactcta tttctttgcc 6840
ctcggacgag tcttggggcg tccgtttcca ctatcggcga gtacttctac acagccatcg 6900
gtccagacgg ccgcgcttct gcgggcgatt tgtgtacgcc cgacagctcc ggtccggat 6960
cggacgatfg cgtcgcctgc acctcgcgcc caagctgcat catcgaatt gccgtcaacc 7020

```

[0018]

aagctctgat agagttggtc aagaccaatg cggagcatat acgcccggag tegtggegat 7080
cctgcaagct cgggatgcct ccgctcgaag tagcgcgtct gctgctccat acaagccaac 7140
cacggcctcc agaagaagat gttggcgacc tcgtattggg aatccccgaa catcgcctcg 7200
ctccagtcga fgaccgctgt tatgcggcca ttgtccgtca ggacattggt ggagecgaaa 7260
tccgegtgea cgaggtgccg gacttcgggg cagtccctcg cccaaageat cagctcatcg 7320
agagcctgag egacggacgc actgacggtg tegtccatca eagtttgcca gtgatacaca 7380
tggggatcag caatcgegca tatgaaatca cgccatgtag tgtattgacc gattccttgc 7440
gggccgaatg ggccgaaccc gctcgtctgg ctaagatcgg ccgcagcgat cgcattccata 7500
gctcccgca ccggtttag aacagcgggc agttcggttt caggcaggte ttgcaacgtg 7560
acaccctgtg cacggcggga gatgcaatag gtcaggetct cgctaaactc cccaatgtca 7620
agcacttccg gaatcgggag cgcggccgat gcaaagtgcc gataaacata acgatctttg 7680
tagaaacat ccgctcagct atttaccgc aggcataatc cagccctcc tacatcgaag 7740
ctgaaagcac gagattcttc gccctccgag agctgcatca ggtcggagac gctgtcgaac 7800
ttttagatca gaaacttctc gacagacgtc gcggtgagtt caggcttttt catatctcat 7860
tgccccccc gatctgcaaa agctcagagag agatagattt gtagagagag actgggtgatt 7920
tcagcgtgct cctcccaaat gaaatgaact tccttatata gaggaaggte ttgcaagga 7980
tagtgggatt gtgcgtcacc ccttaccgca gtggagatat cacatcaatc cacltgcttt 8040
gaagacgtgg ttggaacgtc ttcttttcc acgatgtccc tcgtgggtgg gggctccatct 8100
ttgggaccac tgcctgcaga ggcatttga acgatagcct ttcctttatc gcaatgatgg 8160
cattttagg tgccaccttc ctttctact gtccttttga tgaagtgaca gatagctggg 8220
caatggaatc cgaggaggtt tcccgatatt acectttgtt gaaaagtctc aatagccctt 8280
tggctctctg agaactgata tttagatatt ttggagtaga cgagagtgtc gtgctccacc 8340
atgttatcac atcaatccac ttgctttgaa gacgtggttg gaacgtcttc tttttccacg 8400
atgctcctcg tgggtggggg tccatcttgg ggaccactgt ccgagagagc atcttgaacg 8460
atagcctttc ctttatcgca atgatggcat ttgtaggtgc caccttccct ttctactgtc 8520
cttttgatga agtgacagat agctgggcaa tggaaatccga ggaggtttcc cgatattacc 8580
ctttgttgaa aagctcfaat agccctttgg tcttctgaga ctgtatcttt gatattcttg 8640
gagtagacga gagtgcctg ctcaccatg ttggcaagct gctctagcca ataccgaaac 8700
cgcctctccc ccgctgttgg ccgattcatt aatgcagctg gcacgacagg tttcccact 8760
ggaaagcggg cagtgagegc aacgcaatta atgtgagtt gctcactcat taggcacccc 8820
aggetttaca ctttatgett ccgctcgtat tgttgtgtgg aattgtgagc ggataacaat 8880
tccacacagg aaacagctat gaccatgatt acgctcgcag gtagcagaag gcaatgtgtt 8940
gtgactccga ggggttgcct caaactctat cttataaccg gcgtggagge atggaggcag 9000
gggtattttg gtcattttaa tagatagttg aaaatgacgt ggaatttact taaagacgaa 9060
gtctttgcga caaggggggg cccacgccga atttaatat accggcgtgg cccccctta 9120
tcgagagtgc tttagcacga gcggtccaga tttaaagtag aaaatttccc gccacttagg 9180
gttaaagggtg ttcacactat aaaagcatat acgatgtgat ggtatttgat ggagcgtata 9240
ttgtatcagg tattccggtt ggatacgaat tattcgtacg accctcccta agattcttga 9300
ttgtttataa aaccaaactc cattgtcttt gttgtgtatt gtttgcagga cgtcagagat 9360
tctcaacaca acatatacaa aacaaacgaa tctcaagcaa tcaagcattc tacttctatt 9420
gcagcaattt aatcatttcc tcaagttgt acaaaaaagc aggetgctcc gaattcgecc 9480
ttaccatgg ctagttcaga ttacaaagat cacgatggag attacaaaga tcacgatatt 9540
gattacaaag atgatgatga taaaatggca ccaaagaaaa agagaaaagt tggaaattcat 9600
gggtgtcctg ctgcagataa gaagtacagt ataggacttg atatcggtac caattctgtt 9660

[0019]

ggatgggctg tgattactga tgaatataag gttccatcta aaaagticaa agtgttgggt 9720
 aacactgata gacactctat taagaagaat cttataggag ctctttigtg tgatagtggg 9780
 gaaacagctg aggcaaccag acttaagagg acagcaagaa ggagatacac caggagaaaa 9840
 aacagaatct gttatttgea agaaatcttt tetaatgaga tggetaaggt tgatgattca 9900
 tttttccata ggttgggaaga gagttttctc glggaagagg alaaaaagca tgaagacac 9960
 cctatatctg gaaatattgt tgaatgaagt gettaccatg agaagtaccc aacaatctat 10020
 cacttaagaa aaaagcttgt tgattctacc gataaggctg atcttaggtt aatctacett 10080
 gctttggcac atatgatcaa gtttagaggt cacttcttaa tagaaggaga tttgaaacct 10140
 gataactcag atgttgataa gttgtttatt caactcgtgc agacatataa ccaacttttc 10200
 gaagagaatc caatcaacgc ttcaggaggt gatgctaagg caatcctcag tgcaagactt 10260
 agtaaacta ggagactega aaatcttata gctcagttgc ctggagaaaa gaagaatggg 10320
 ttgttcggta acctcctcgc actctctctc ggtcttacac caaacttcaa gtcaaacctc 10380
 gatcttgetg aagatgcaaa gttgcaacte tetaaggata ctacgatga tgatttggat 10440
 aatctcttag ctcagattgg agatcaatat gctgatcttt ttcttctgc aaagaacctt 10500
 tcagatgeta tctttttgag tgatattctg agagtttaaca ctgaaataac aaaggctcct 10560
 ctttcagcaa gtatgatcaa aagatacgat gacatcacc aagatttgac tctcttaaaa 10620
 gctttggta gacaacagct cccagaaaag tataaggaaa tttctttga tcagtctaag 10680
 aatggttatg ctggttacat cgatggaggt gcatcacaag aagagttcta caaatcatt 10740
 aagcctatct tggaaaagat ggatggaacc gaagagcttt tggtaagct taacagagag 10800
 gatctcttaa gaaaacagag gactttcgat aatggttcta tccctcatca aatacacctt 10860
 ggtgaattgc atgctattct caggagacaa gaggatttct accttctct taaggataac 10920
 agagaaaaaga tcgagaagat tttgactttt agaattcctt attacgttgg accacttgc 10980
 aggggtaaca gtaggttctc atggatgact aggaagtctg aagagacaat tacccttgg 11040
 aattttgaag aggttgtgga taaaggtgct tctgcacagt cttttattga aagaatgaca 11100
 aacttcgata agaattcttc taacgagaag gttcttccaa agcattctct tttgtacgaa 11160
 tacttcaactg tgtacaatga gcttacaag gtttaagtatg tgaccgaagg aatgaggag 11220
 ccagcttttc tttcaggaga gcagaaaaag gcaattgttg atctcttatt caagaciaac 11280
 agaaaggta cagtgaacaa acttaaggaa gattacttca aaaagatcga gtgettctgat 11340
 tctgttgaat tatcaggagt ggaggataga ttcaatgctt ctttgggtac ttatcatgat 11400
 cttttgaaaa ttatcaaaga taaggatttt ctgataatg aggagaacga agatacctc 11460
 gaggatctg ttttgactct cacattatct gaagatagag agatgattga agagaggttg 11520
 aagacatac ctcacctctt cgatgataaa gttatgaagc aacttaagag gagaaggtat 11580
 accggatggg gtagattatc taggaaactt atcaacggaa taagggataa gcagtcagga 11640
 aagaactatc tcgatttctt aaagagtgat ggattcgtca acagaaactt catgcaacte 11700
 atccatgatg attctcttac ttttaaggaa gataaccaaa aagctcaggt ttcaggacag 11760
 ggagatagtt tacatgagca cattgctaatt ctgacagat cactgctat taagaagggt 11820
 attttacaaa ctgttaaggt tgtggatgaa ctgttaagg ttatgggaag acataagcca 11880
 gagaacatcg tgatagaaat ggctagggag aaccaacta cacagaaggg tcaaaagaat 11940
 tcaagagaaa ggatgaagag aatcgaagag ggaaltaagg agctcggctt acaaatctt 12000
 aaggaacatc cagttgagaa tacacaattg cagaacgaaa agctttactt gtactacctc 12060
 cagaacggaa gagatatgta tgtggatcaa gatttgata ttaataggct cagtgattac 12120
 gatgttgatc acatcgtgcc tcagtctttt cttaaggatg attcaatcga taacaaagtt 12180
 ctacttagat cagataagaa taggggtaaa tctgataacg tgccatcaga agaggttgtg 12240
 aaaaagatga aaaactactg gagacagctc ctcaacgcta agttgateac tcaagaaaa 12300

[0020]


```

ttcgataatc tcacaaaggc tgaaggggga ggtctttctg agttggataa ggcaggtttt 12360
attaagagac agttggttga aactaggcaa attacaaagc atgtggctca aatccttgat 12420
tcaaggatga atactaagta cgatgaaaac gataagttga tcagagaggt taaagtgatt 12480
acattgaaga gtaagctcgt ttctgatttc agaaaaggatt ttcagttcta caaagtgagg 12540
gagattaata aclaccatec cgcicacgat gcalaictta atgetgttgi gggaacagca 12600
ttgatcaaaa agtatectaa gctcgaatct gagttcgttt acggagatta caaagtttac 12660
gatgtgagaa agatgategc taagtcagaa caagagattg gaaaggctac tgcaaagtac 12720
tttttctata gtaacattat gaactttttc aagaccgaaa tcactttggc taacggagag 12780
atcagaaaaa ggccacttat agaaaccaat ggtgaaactg gagagattgt ttgggataag 12840
ggtagagatt tcgcaacagt taggaagggt ctttctatgc ctcaggttaa catcgtgaaa 12900
aagacagaag ttcaaacegg aggttttagt aaggagtcta tccfcccana gagaaactca 12960
gataagetta tcgctaggaa aaaggattgg gaccctaaga agtacggagg ttttgatagt 13020
ccaacagttg cttattctgt gttggttgtg gcaaaagttg aaaagggaaa gtctaagaag 13080
cttaagtctg tgaagaget tttgggtatc actattatgg aaagatcacc attcagagaag 13140
aacccatattg atllctllgga agcLaaggga lacaaagagg tLaagaagga tcilattatt 13200
aagcttccaa agtactcact cttcgaatta gagaatggta gaaagaggat gttggctagt 13260
gcaggagaac tccaaaaggg taacgagctc gctctccctt ctaaatafgt taattttctt 13320
tacttggcat ctcattacga aaagcttaag ggttcaccag aagataatga gcaaaagcag 13380
ttattcgttg aacagcataa acactacctc gatgaaatca tagagcaaat atcagagttt 13440
agtaagagag ttattttggc tgatgcaaac ttggataaag tgcctctctgc ttacaataag 13500
catagagata agcctataag ggaacaagca gagaacatta tccacctttt taccttgact 13560
aatctcggag ctccagctgc ttttaagtae ttcgatacca ctattgatag aaagaggtat 13620
acaagtacca aggaagttct cgatgcaact ctatatacct agtctatcac cggctcttat 13680
gagactagaa tagatttgtc acagtttaga ggagattcaa gacgagatcc aaagaagaaa 13740
agaaaagtta gatctagcla gtcgacgtcc gatcgttcaa acattlggca ataaaagttc 13800
ttaagattga atcctgttgc cggctcttgc atgattatca tataatttct gttgaattac 13860
gtaagcatg taataattaa catgtaatgc atgacgttat ttatgagatg ggtttttatg 13920
attagagtc cccaattata catttaatac gcgatagaaa acaaaatata gcgcgcaaac 13980
taggataaat tatcgcgcgc ggtgtcatct atgttactag atcgggaatt gateccccct 14040
cgacagcttc cggaaagggc gaattccaac tttgtatac aaagtgcgc agctcgcgc 14100
cgcttctgtg aacaacggaa actcgcactg ccttccgcac aatacatcat ttctcttag 14160
cttttttctc tcttcttctg tcatacagtt tttttttgtt taicagctta cattttcttg 14220
aacgtagct ttcgttttct tctttttaac tttccattcg gagtttttgt atcttgtttc 14280
atagtttgc ccaggattag aatgattagc catcgaacct tcaagaattt gattgaataa 14340
aacatcttca ttcttaagat atgaagalaa tcttcaaaag gccccggga atctgaaaga 14400
agagaagcag gcccaattat atgggaaaga acaatagtat ttcttatata ggcccaatta 14460
agttgaaaac aatcttcaaa agtcccacat cgttagata agaaaacgaa getgagttta 14520
tatacagcta gagtcgaagt agtgattgag acggagtcga ttcgtctctg tttagagcta 14580
gaaatagcaa gttaaaataa ggctagtcgc ttatcaactt gaaaaagtgg caccgagtcg 14640
gtgctttttt tctagacceca gctttcttgt acaaagttgc tgcaggtega cgggagtgtta 14700
cttcaagtca gtgggaaate aataaaaiga ttattttatg aatataattc attgtgcaag 14760
tagatagaaa ttacatatgt tacataacac acgaaataaa caaaaaaga caatccaaa 14820
acaaacaccc caaaaaaaat aatcacttta gataaacctg tatgaggaga ggcaegtcca 14880
gtgactcgac gattcccagc caaaaaaagt ctcccctgca cacatgtagt ggggtgacga 14940

```

[0021]

```

attatcttta aagtaatect tctgttgact tgcattgat aacatccagt cttcgtcagg 15000
attgcaaaaga attatagaag ggatcccact cgagggtcaa catggtggag cacgacacac 15060
ttgtctactc caaaaatatac aaagatacag tctcagaaga ccaaaggga attgagactt 15120
ttcaacaaaag ggtaataatcc ggaaacctcc teggattcca ttgccagct atctgtcact 15180
ttattgtgaa gatagtggaa aaggaaggtg gctcctacaa atgceatcat tgcgataaag 15240
gaaaggccat cgttgaagat gccctctgccc acagtggtcc caaagatgga cccceaceca 15300
cgaggagcat cgtggaaaaa gaagacgttc caaccacgtc ttcaaagcaa gtggattgat 15360
gtgataacat ggtggagcac gacacacttg tctactceaa aaatatcaaa gatacagtct 15420
cagaagacca aagggcaatt gagacttttc aacaaaggtt aatatccgga aacctctctg 15480
gattccattg cccagctatc tgcacttta ttgtgaagat agtggaaaag gaaggtggct 15540
cctacaaatg ccatcattgc gataaaggaa aggccatcgt tgaagatgcc tctgccgaca 15600
gtggtcccaa agatggacce ccacccaega ggagcatcgt ggaaaaagaa gacgttccaa 15660
ccacgtcttc aaagcaagtg gattgatgtg atatctccac tgacgtaagg gatgacgcac 15720
aatcccacta tccttegeaa gaccttctct ctatataagg aagttcattt catttggaga 15780
ggacctgac tctagaggat ccccttctct ctatataagg aagttcattt catttggaga 15840
ggtaagtttc actteacaca ttattactgt cttctaatac aaggtttttt atcaagetgg 15900
agaagagcat gatagtgggt agtgccatct tcatgaagct cagaagcaac accaagggaag 15960
aaaataagaa aaggtgtgag tttctcccag agaaactgga ataaatcacc tctttgagat 16020
gagcacttgg gataggttaag gaaaacatat ttagattgga gtctgaagtt cttactagca 16080
gaaggeatgt tgttgtgact ccgaggggtt gectcaaact ctatcttata accggegtgg 16140
aggcatggag gcaggggtat tttggtcatt ttaatagata gtggaaaatg acgtggaatt 16200
tacttaaaga egaagtcttt gcgacaaggg ggggcccacg ccgaatttaa tattaccggc 16260
gtggcccccc cttatcgcga gtgcttttagc acgagcggtc cagatttaaa gtagaaaatt 16320
tcccgccac tagggttaaa ggtgttcaca ctataaaage atatacgatg tgatggtatt 16380
tgatggagcg tatattgtat caggtatttc cgttggatac gaattatctg tacgacctc 16440
atagtttaaa ctgaaggcgg gaaacgacaa tctgatccaa gctcaagcta agcttgcattg 16500

```

cctgcagg

16508<210> 9

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 合成的寡核苷酸

<400> 9

gtgggaggau cgguucuaua

20

<210> 10

<211> 4940

<212> DNA

<213> 人工序列

<220>

[0022]

<223> 重组质粒

<400> 10

```

atggctccta agaagaagag aaaggttata acaatgggta gcaagggcga ggagctgttc 60
accgggggtgg tgcccaatcct ggtcgagctg gacggcgacg taaacggcea eaagttcage 120
gtgtceggcg agggcgaggg cgatgccacc tacggcaage tgacctgaa gttcatctgc 180
accaccggca agctgcccgf gccctggccc accctegtga ccaccttgg ctaccggcctg 240
cagtgtcttg cccgctaccc cgaccacatg aagcagcacg actttttcaa gtccgccaatg 300
cccgaagget acgtccagga gcgaccatc ttcttcaagg acgacggcaa ctacaagaca 360
gatcatatgt gggaggatcg gttctataag gctcgagcta gtgcagtgtc tcgcccgcga 420
ccccgaccac atgaagcagc acgacttctt caagtccgcc atgcccgaag gctacgtcca 480
ggagcgcacc atcttcttca aggacgacgg caactacaag acccgcgccg aggtgaagtt 540
cgagggcgac acactgggta accgcatcga gctgaagggc atcgacttca aggaggacgg 600
caacatcctg gggcacaage tggagtacaa ctacaacage cacaacgtct atatcatggt 660
cgacaagcag aagaacggca tcaaggtgaa ctcaagate cgccacaaca tcgaggacgg 720
cagcgtgcag ctgcccgcacc actaccagca gaacacccc atcgcgacg gccccgtgat 780
gctgcccgcac aacctactc lgagctacca gtccgccctg agcaaagacc ccaacgagaa 840
gcgcgatcac atggctctgc tggagtctgt gaccgccc gccgatcactc tcggcatgga 900
agagctgtac aagccgcggt tcccgggaga ctttagagc tcgaatttcc ccgatcgttc 960
aaacatttgg caataaagtt tettaagatt gaatctgtt gccggtcttg cgatgattat 1020
catataattt ctgttgaatt acgttaagca tgtaataatt aacatgtaat gcattgactt 1080
atftatgaga tgggttttta tgattagagt cccgcaatta tacatttaat accgataga 1140
aaacaaaata tagcgcgcaa actaggataa attatcgcgc gcggtgtcat ctatgttaet 1200
agatcgtggg cccagctcca aattcgcctt atagtgagtc gtattacggc cgtctactgg 1260
ccgtcgtttt acaacgtcgt gactgggaaa accctggcgt tacccaactt aategccttg 1320
cagcacatcc cctttcgcg agctggcgta atagegaaga ggcccgcacc gatcgccttt 1380
cccacagttt gcgcagcctg aatggcgaat ggaaattgta agcgttaata ttttgttaa 1440
atcgcgtta aattttgtt aaatcagctc attttttaac caatagccg aaatcggcaa 1500
aatcccttat aaatcaaaag aatagaccga gatagggttg agtgttgttc cagtttgtaa 1560
caagagtcca ctattaaaga acgtggactc caacgtcaaa gggcgaaaaa ccgtctatca 1620
ggcgatggc ccactacgtg aaccatcacc ctatcaagt tttttgggt cgaggtgccc 1680
taaagcacta aatcgaacc cttaaaggag cccccgattt agagcttgac ggggaaagcc 1740
ggcgaacgtg gcgagaaagg aagggaaaga agegaaagga gcgggcgcta ggcgctggc 1800
aagtgtagcg gtcacgctgc gcgtaaccac cacaccgcc gccttaatg ccccgtaca 1860
ggcgcgctca ggtggcactt ttcgggaaa tgtgcgcgga accctattt gtttattttt 1920
ctaaatacat tcaaatatgt atccgctcat gagacaataa cctgataaaa tgcctcaata 1980
atattgaaaa aggaagagta tgagtattca acatttccgt gtcgccccta ttccttttt 2040
tgcggcattt tgccttctg tttttgctca cccagaaacg ctggtgaaag taaaagatgc 2100
tgaagatcag ttgggtgcac gagtgggta catcgaactg gatctcaaea gcggttaagat 2160
ccttgagagt tttgccccg aagaacgttt tccaatgatg agcactttta aagttctgct 2220
atgtggegcg gtattatccc gtattgacgc cgggcaagag caactcggtc gccgcataca 2280
ctattctcag aatgacttgg ttgagtactc accagtcaca gaaaagcatc ttacggatgg 2340
catgacagta agagaattat gcagtctgc cataaccatg agtgataaca ctgcccga 2400
cttacttctg acaacgateg gaggaccgaa ggagctaacc gcttttttgc acaacatggg 2460

```

[0023]

```

ggatcatgta actcgccttg atcgttggga accggagctg aatgaagcca taccaaacga 2520
cgagcgtgac accacgatgc ctgtagcaat ggcaacaacg ttgcgcaaac tattaactgg 2580
cgaactactt actetagettt cccggcaaca atfaatagac tggatggagg cggataaagt 2640
tgcaggacca cttctgctct eggeccttcc ggctggctgg tttattgetg ataaatctgg 2700
agccggtgag cgtgggtctc gcggtatcat tgcagcactg gggccagatg gtaagccctc 2760
ccgtatcgta gttatctaca cgacggggag tcaggcaact atggatgaac gaaatagaca 2820
gatcgtgag ataggtgctt cactgattaa gcattggtaa ctgtcagacc aagtttactc 2880
atatatactt tagattgatt taaaacttca tttttaattt aaaaggatct aggtgaagat 2940
cctttttgat aatctcatga ccaaaatccc ttaacgtgag ttttcgttcc actgagcgtc 3000
agaccccgta gaaaagatca aaggatcttc ttgagatcct tttttctgc gcgtaatctg 3060
ctgcttgcaa acaaaaaaac caccgctacc agcggtggtt tgtttgccg atcaagagct 3120
accaactctt tttccgaagg taactggctt cagcagagcg cagataccaa atactgtcct 3180
tctagtgtag ccgtagttag gccaccactt caagaactct gtageaccgc ctacatactt 3240
cgctctgcta atcctgttac cagtggctgc tgccagtggc gataagtcgt gtcttaccgg 3300
gttggactca agacgatagt taccggataa ggcgcagcgg tcgggctgaa cgggggggtc 3360
gtgcacacag cccagcttgg agcgaacgac ctacaccgaa ctgagatacc tacagcgtga 3420
gctatgagaa agcggccacgc ttcccgaagg gagaaaggcg gacaggtatc cggtaagcgg 3480
cagggtcggg acaggagagc gcacgagggg gcttccaggg ggaaacgcct ggtatcttta 3540
tagtcctgtc gggtttcgcc acctctgact tgagcgtcga tttttgtgat gctcgtcagg 3600
ggggcggagc ctatggaaaa acgccagcaa cggggccttt ttacggttcc tggccttttg 3660
ctggcctttt gctcacatgt tctttctgc gttatccctt gattctgtgg ataaccgtat 3720
taccgccttt gactgagctg ataccgctcg ccgcagccga acgaccgagc gcagcagctc 3780
agtgagcggg gaagcgggag agcgcctaat acgcaaaccg cctctccccg cgcgttggcc 3840
gattcattaa tgcagctggc acgacaggtt tcccactgag aaagcgggca gtgagcgcga 3900
cgcaattaat gtgagttagc tcactcatta ggcaccccag gctttactt ttatgcttcc 3960
ggctcgtatg ttgtgtggaa ttgtgagcgg ataacaattt cacacaggaa acagctatga 4020
ccatgattac gccaaagcgc caattaacc tcaactaaagg gaacaaaagc tgggtactcg 4080
tacggctccc agatttgcct tttcaatttc agaaagaatg ctaaccaca gatggttaga 4140
gaggcttacg cagcaggtct catcaagacg atctaccga gcaataatct ccaggaaatc 4200
aaataccttc ccaagaaggt taaagatgca gtcaaaagat tcaggactaa ctgcatcaag 4260
aacacagaga aagatatatt tctcaagatc agaagtacta ttccagtatg gacgattcaa 4320
ggcttgcttc acaaccaag gcaagtaata gagattggag tctctaaaaa ggtagttccc 4380
actgaatcaa aggccatgga gcaaaagatt caaatagagg acctaacaga actcgcctga 4440
aagactggcg aacagttcat acagagtctc ttacgactca atgacaagaa gaaaatcttc 4500
gtcaacatgg tggagcagca cacacttgtc tactccaaa atatcaaaga tacagtctca 4560
gaagaccaa gggcaattga gacttttcaa caaagggtta tatccgaaa cctcctcgga 4620
ttccattgcc cagctatctg tcactttatt gtgaagatag tggaaaagga aggtggetcc 4680
tacaatgcc atcattgcca taaaggaaag gccatcgttg aagatgcctc tgccgacagt 4740
ggtcccaaag atggaccccc acccagcagg agcatcgttg aaaaagaaga cgttccaacc 4800
acgtcttcaa aagcaagtgg attgatgtga tatctccact gacgtaaggg gatgacgcac 4860
aatcccacta tctttcga agacccttcc tctatataag gaagttcatt tcatttggag 4920
agaacacggg ggactctaga 4940

```

[0024]

<210> 11
 <211> 1388
 <212> PRT
 <213> 嗜热链球菌(*Streptococcus thermophilus*)

<400> 11
 Met Thr Lys Pro Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
 1 5 10 15
 Gly Trp Ala Val Thr Thr Asp Asn Tyr Lys Val Pro Ser Lys Lys Met
 20 25 30
 Lys Val Leu Gly Asn Thr Ser Lys Lys Tyr Ile Lys Lys Asn Leu Leu
 35 40 45
 Gly Val Leu Leu Phe Asp Ser Gly Ile Thr Ala Glu Gly Arg Arg Leu
 50 55 60
 Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Arg Asn Arg Ile Leu
 65 70 75 80
 Tyr Leu Gln Glu Ile Phe Ser Thr Glu Met Ala Thr Leu Asp Asp Ala
 85 90 95
 Phe Phe Gln Arg Leu Asp Asp Ser Phe Leu Val Pro Asp Asp Lys Arg
 100 105 110
 Asp Ser Lys Tyr Pro Ile Phe Gly Asn Leu Val Glu Glu Lys Ala Tyr
 115 120 125
 His Asp Glu Phe Pro Thr Ile Tyr His Leu Arg Lys Tyr Leu Ala Asp
 130 135 140
 Ser Thr Lys Lys Ala Asp Leu Arg Leu Val Tyr Leu Ala Leu Ala His
 145 150 155 160
 Met Ile Lys Tyr Arg Gly His Phe Leu Ile Glu Gly Glu Phe Asn Ser
 165 170 175
 Lys Asn Asn Asp Ile Gln Lys Asn Phe Gln Asp Phe Leu Asp Thr Tyr
 180 185 190
 Asn Ala Ile Phe Glu Ser Asp Leu Ser Leu Glu Asn Ser Lys Gln Leu
 195 200 205
 Glu Glu Ile Val Lys Asp Lys Ile Ser Lys Leu Glu Lys Lys Asp Arg
 210 215 220
 Ile Leu Lys Leu Phe Pro Gly Glu Lys Asn Ser Gly Ile Phe Ser Glu
 225 230 235 240
 Phe Leu Lys Leu Ile Val Gly Asn Gln Ala Asp Phe Arg Lys Cys Phe
 245 250 255
 Asn Leu Asp Glu Lys Ala Ser Leu His Phe Ser Lys Glu Ser Tyr Asp
 260 265 270
 Glu Asp Leu Glu Thr Leu Leu Gly Tyr Ile Gly Asp Asp Tyr Ser Asp
 275 280 285
 Val Phe Leu Lys Ala Lys Lys Leu Tyr Asp Ala Ile Leu Leu Ser Gly
 290 295 300

[0025]

Phe Leu Thr Val Thr Asp Asn Glu Thr Glu Ala Pro Leu Ser Ser Ala
 305 310 315 320
 Met Ile Lys Arg Tyr Asn Glu His Lys Glu Asp Leu Ala Leu Leu Lys
 325 330 335
 Glu Tyr Ile Arg Asn Ile Ser Leu Lys Thr Tyr Asn Glu Val Phe Lys
 340 345 350
 Asp Asp Thr Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Lys Thr Asn
 355 360 365
 Gln Glu Asp Phe Tyr Val Tyr Leu Lys Lys Leu Leu Ala Glu Phe Glu
 370 375 380
 Gly Ala Asp Tyr Phe Leu Glu Lys Ile Asp Arg Glu Asp Phe Leu Arg
 385 390 395 400
 Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro Tyr Gln Ile His Leu
 405 410 415
 Gln Glu Met Arg Ala Ile Leu Asp Lys Gln Ala Lys Phe Tyr Pro Phe
 420 425 430
 Leu Ala Lys Asn Lys Glu Arg Ile Glu Lys Ile Leu Thr Phe Arg Ile
 435 440 445
 Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Asp Phe Ala Trp
 450 455 460
 Ser Ile Arg Lys Arg Asn Glu Lys Ile Thr Pro Trp Asn Phe Glu Asp
 465 470 475 480
 Val Ile Asp Lys Glu Ser Ser Ala Glu Ala Phe Ile Asn Arg Met Thr
 485 490 495
 Ser Phe Asp Leu Tyr Leu Pro Glu Glu Lys Val Leu Pro Lys His Ser
 500 505 510
 Leu Leu Tyr Glu Thr Phe Asn Val Tyr Asn Glu Leu Thr Lys Val Arg
 515 520 525
 Phe Ile Ala Glu Ser Met Arg Asp Tyr Gln Phe Leu Asp Ser Lys Gln
 530 535 540
 Lys Lys Asp Ile Val Arg Leu Tyr Phe Lys Asp Lys Arg Lys Val Thr
 545 550 555 560
 Asp Lys Asp Ile Ile Glu Tyr Leu His Ala Ile Tyr Gly Tyr Asp Gly
 565 570 575
 Ile Glu Leu Lys Gly Ile Glu Lys Gln Phe Asn Ser Ser Leu Ser Thr
 580 585 590
 Tyr His Asp Leu Leu Asn Ile Ile Asn Asp Lys Glu Phe Leu Asp Asp
 595 600 605
 Ser Ser Asn Glu Ala Ile Ile Glu Glu Ile Ile His Thr Leu Thr Ile
 610 615 620
 Phe Glu Asp Arg Glu Met Ile Lys Gln Arg Leu Ser Lys Phe Glu Asn
 625 630 635 640
 Ile Phe Asp Lys Ser Val Leu Lys Lys Leu Ser Arg Arg His Tyr Thr
 645 650 655

[0026]

Gly Trp Gly Lys Leu Ser Ala Lys Leu Ile Asn Gly Ile Arg Asp Glu
 660 665 670
 Lys Ser Gly Asn Thr Ile Leu Asp Tyr Leu Ile Asp Asp Gly Ile Ser
 675 680 685
 Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ala Leu Ser Phe Lys
 690 695 700
 Lys Lys Ile Gln Lys Ala Gln Ile Ile Gly Asp Glu Asp Lys Gly Asn
 705 710 715 720
 Ile Lys Glu Val Val Lys Ser Leu Pro Gly Ser Pro Ala Ile Lys Lys
 725 730 735
 Gly Ile Leu Gln Ser Ile Lys Ile Val Asp Glu Leu Val Lys Val Met
 740 745 750
 Gly Gly Arg Lys Pro Glu Ser Ile Val Val Glu Met Ala Arg Glu Asn
 755 760 765
 Gln Tyr Thr Asn Gln Gly Lys Ser Asn Ser Gln Gln Arg Leu Lys Arg
 770 775 780
 Leu Glu Lys Ser Leu Lys Glu Leu Gly Ser Lys Ile Leu Lys Glu Asn
 785 790 795 800
 Ile Pro Ala Lys Leu Ser Lys Ile Asp Asn Asn Ala Leu Gln Asn Asp
 805 810 815
 Arg Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Lys Asp Met Tyr Thr Gly
 820 825 830
 Asp Asp Leu Asp Ile Asp Arg Leu Ser Asn Tyr Asp Ile Asp His Ile
 835 840 845
 Ile Pro Gln Ala Phe Leu Lys Asp Asn Ser Ile Asp Asn Lys Val Leu
 850 855 860
 Val Ser Ser Ala Ser Asn Arg Gly Lys Ser Asp Asp Val Pro Ser Leu
 865 870 875 880
 Glu Val Val Lys Lys Arg Lys Thr Phe Trp Tyr Gln Leu Leu Lys Ser
 885 890 895
 Lys Leu Ile Ser Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg
 900 905 910
 Gly Gly Leu Ser Pro Glu Asp Lys Ala Gly Phe Ile Gln Arg Gln Leu
 915 920 925
 Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Arg Leu Leu Asp Glu
 930 935 940
 Lys Phe Asn Asn Lys Lys Asp Glu Asn Asn Arg Ala Val Arg Thr Val
 945 950 955 960
 Lys Ile Ile Thr Leu Lys Ser Thr Leu Val Ser Gln Phe Arg Lys Asp
 965 970 975
 Phe Glu Leu Tyr Lys Val Arg Glu Ile Asn Asp Phe His His Ala His
 980 985 990
 Asp Ala Tyr Leu Asn Ala Val Val Ala Ser Ala Leu Leu Lys Lys Tyr
 995 1000 1005

[0027]

Pro Lys Leu Glu Pro Glu Phe Val Tyr Gly Asp Tyr Pro Lys Tyr Asn 1010	1015	1020
Ser Phe Arg Glu Arg Lys Ser Ala Thr Glu Lys Val Tyr Phe Tyr Ser 1025	1030	1035 1040
Asn Ile Met Asn Ile Phe Lys Lys Ser Ile Ser Leu Ala Asp Gly Arg 1045	1050	1055
Val Ile Glu Arg Pro Leu Ile Glu Val Asn Glu Glu Thr Gly Glu Ser 1060	1065	1070
Val Trp Asn Lys Glu Ser Asp Leu Ala Thr Val Arg Arg Val Leu Ser 1075	1080	1085
Tyr Pro Gln Val Asn Val Val Lys Lys Val Glu Glu Gln Asn His Gly 1090	1095	1100
Leu Asp Arg Gly Lys Pro Lys Gly Leu Phe Asn Ala Asn Leu Ser Ser 1105	1110	1115 1120
Lys Pro Lys Pro Asn Ser Asn Glu Asn Leu Val Gly Ala Lys Glu Tyr 1125	1130	1135
Leu Asp Pro Lys Lys Tyr Gly Gly Tyr Ala Gly Ile Ser Asn Ser Phe 1140	1145	1150
Thr Val Leu Val Lys Gly Thr Ile Glu Lys Gly Ala Lys Lys Lys Ile 1155	1160	1165
Thr Asn Val Leu Glu Phe Gln Gly Ile Ser Ile Leu Asp Arg Ile Asn 1170	1175	1180
Tyr Arg Lys Asp Lys Leu Asn Phe Leu Leu Glu Lys Gly Tyr Lys Asp 1185	1190	1195 1200
Ile Glu Leu Ile Ile Glu Leu Pro Lys Tyr Ser Leu Phe Glu Leu Ser 1205	1210	1215
Asp Gly Ser Arg Arg Met Leu Ala Ser Ile Leu Ser Thr Asn Asn Lys 1220	1225	1230
Arg Gly Glu Ile His Lys Gly Asn Gln Ile Phe Leu Ser Gln Lys Phe 1235	1240	1245
Val Lys Leu Leu Tyr His Ala Lys Arg Ile Ser Asn Thr Ile Asn Glu 1250	1255	1260
Asn His Arg Lys Tyr Val Glu Asn His Lys Lys Glu Phe Glu Glu Leu 1265	1270	1275 1280
Phe Tyr Tyr Ile Leu Glu Phe Asn Glu Asn Tyr Val Gly Ala Lys Lys 1285	1290	1295
Asn Gly Lys Leu Leu Asn Ser Ala Phe Gln Ser Trp Gln Asn His Ser 1300	1305	1310
Ile Asp Glu Leu Cys Ser Ser Phe Ile Gly Pro Thr Gly Ser Glu Arg 1315	1320	1325
Lys Gly Leu Phe Glu Leu Thr Ser Arg Gly Ser Ala Ala Asp Phe Glu 1330	1335	1340
Phe Leu Gly Val Lys Ile Pro Arg Tyr Arg Asp Tyr Thr Pro Ser Ser 1345	1350	1355 1360

[0028]

Leu Leu Lys Asp Ala Thr Leu Ile His Gln Ser Val Thr Gly Leu Tyr
 1365 1370 1375
 Glu Thr Arg Ile Asp Leu Ala Lys Leu Gly Glu Gly
 1380 1385

<210> 12

<211> 1368

<212> PRT

<213> 化脓链球菌(Streptococcus pyogenese)

<400> 12

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
 1 5 10 15
 Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
 20 25 30
 Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
 35 40 45
 Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
 50 55 60
 Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
 65 70 75 80
 Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
 85 90 95
 Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
 100 105 110
 His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
 115 120 125
 His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
 130 135 140
 Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
 145 150 155 160
 Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
 165 170 175
 Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
 180 185 190
 Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
 195 200 205
 Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
 210 215 220
 Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
 225 230 235 240
 Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
 245 250 255

[0029]

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
 260 265 270
 Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
 275 280 285
 Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
 290 295 300
 Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
 305 310 315 320
 Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
 325 330 335
 Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
 340 345 350
 Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
 355 360 365
 Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
 370 375 380
 Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
 385 390 395 400
 Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
 405 410 415
 Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
 420 425 430
 Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
 435 440 445
 Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
 450 455 460
 Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
 465 470 475 480
 Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
 485 490 495
 Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
 500 505 510
 Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
 515 520 525
 Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
 530 535 540
 Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
 545 550 555 560
 Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
 565 570 575
 Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
 580 585 590
 Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
 595 600 605

[0030]

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr 610	615	620	
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala 625	630	635	640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr 645	650	655	
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp 660	665	670	
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe 675	680	685	
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe 690	695	700	
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu 705	710	715	720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly 725	730	735	
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly 740	745	750	
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln 755	760	765	
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile 770	775	780	
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro 785	790	795	800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu 805	810	815	
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg 820	825	830	
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys 835	840	845	
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg 850	855	860	
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys 865	870	875	880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys 885	890	895	
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp 900	905	910	
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr 915	920	925	
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp 930	935	940	
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser 945	950	955	960

[0031]

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
 965 970 975
 Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
 980 985 990
 Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
 995 1000 1005
 Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
 1010 1015 1020
 Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser
 1025 1030 1035 1040
 Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu
 1045 1050 1055
 Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile
 1060 1065 1070
 Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser
 1075 1080 1085
 Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly
 1090 1095 1100
 Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile
 1105 1110 1115 1120
 Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser
 1125 1130 1135
 Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly
 1140 1145 1150
 Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile
 1155 1160 1165
 Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala
 1170 1175 1180
 Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys
 1185 1190 1195 1200
 Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser
 1205 1210 1215
 Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr
 1220 1225 1230
 Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
 1235 1240 1245
 Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
 1250 1255 1260
 Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val
 1265 1270 1275 1280
 Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys
 1285 1290 1295
 His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu
 1300 1305 1310

[0032]

Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp
1315 1320 1325

Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp
1330 1335 1340

Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile
1345 1350 1355 1360

Asp Leu Ser Gln Leu Gly Gly Asp
1365

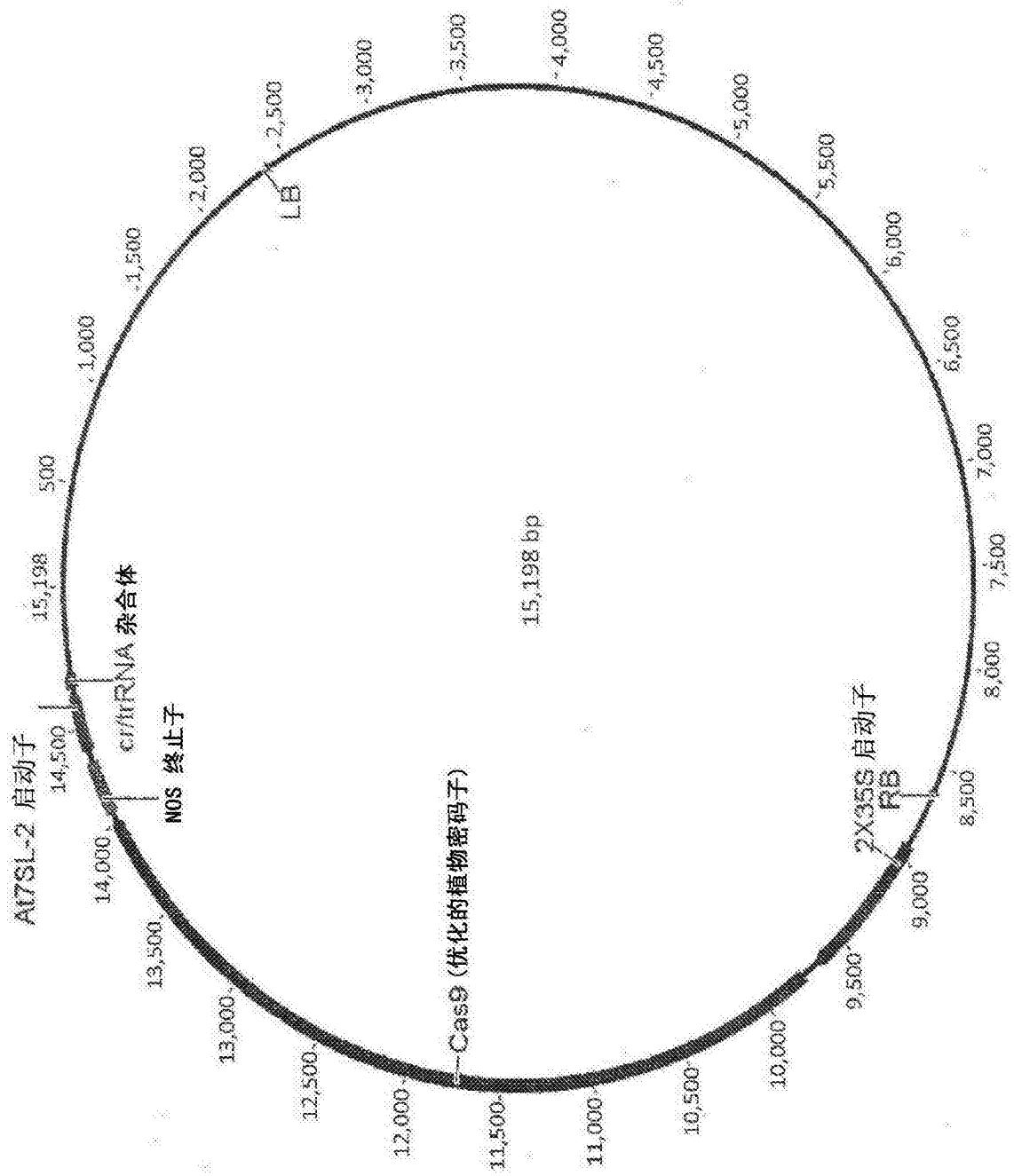


图 1

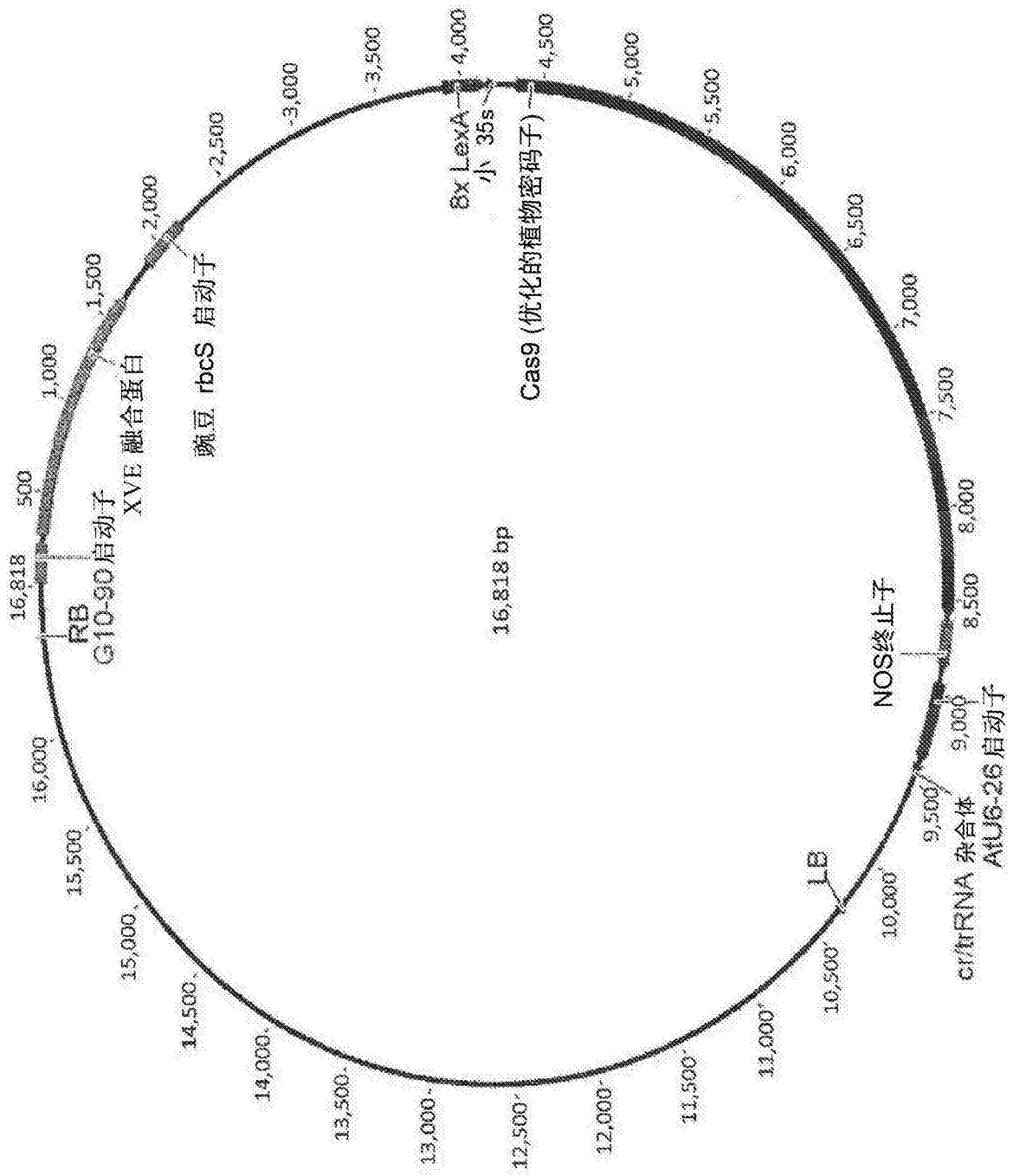


图 2

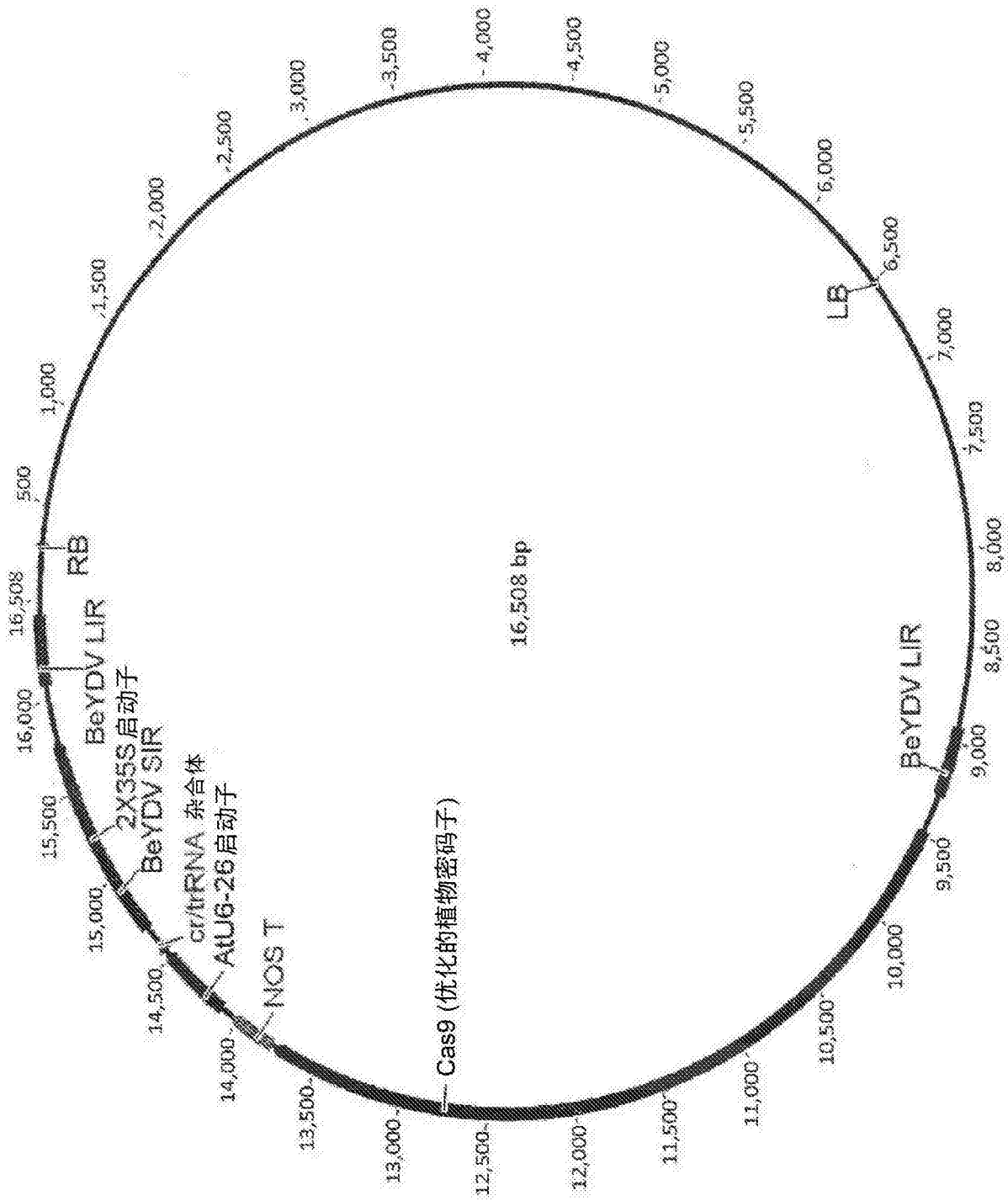


图 3

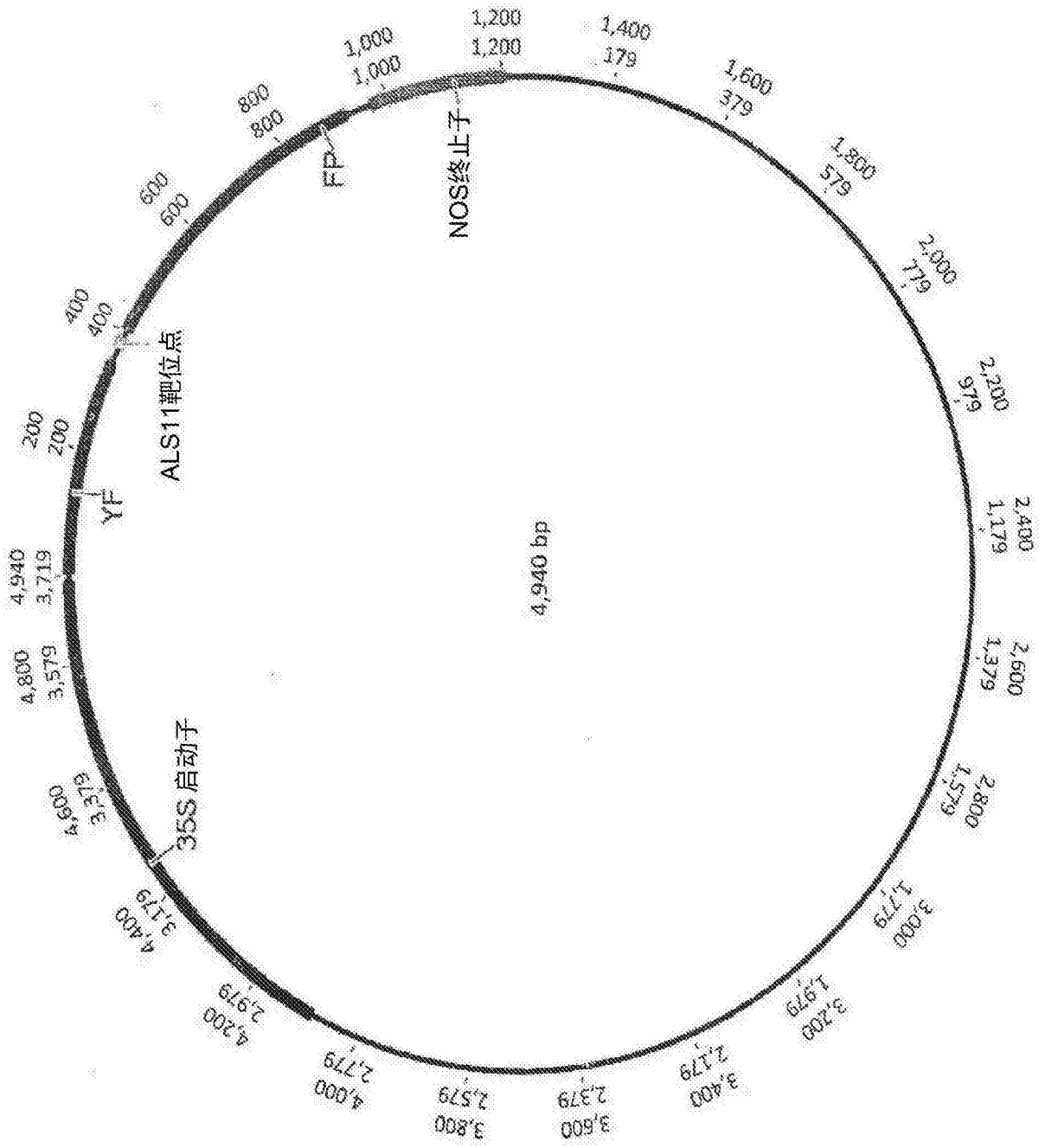


图 5

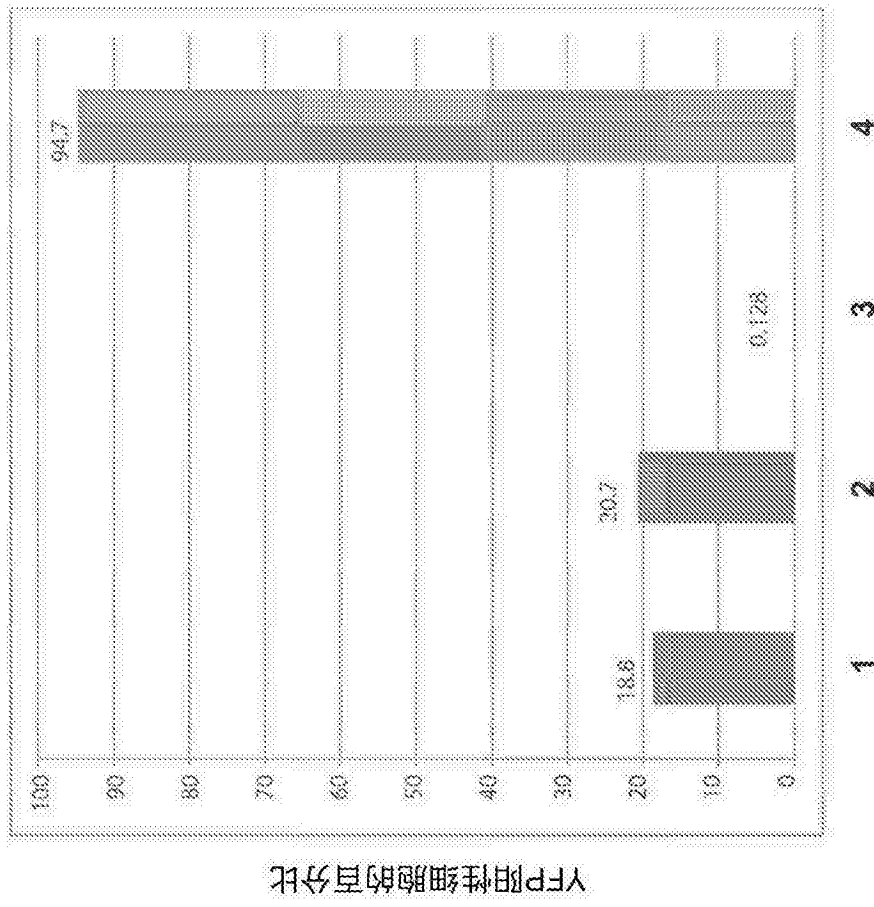


图 6