



(12) 发明专利

(10) 授权公告号 CN 102282255 B

(45) 授权公告日 2014. 05. 07

(21) 申请号 200980154897. 0

(22) 申请日 2009. 11. 19

(30) 优先权数据

61/116, 109 2008. 11. 19 US

61/162, 230 2009. 03. 20 US

(85) PCT国际申请进入国家阶段日

2011. 07. 19

(86) PCT国际申请的申请数据

PCT/US2009/065048 2009. 11. 19

(87) PCT国际申请的公布数据

W02010/059763 EN 2010. 05. 27

(73) 专利权人 阿迈瑞斯公司

地址 美国加利福尼亚州

(72) 发明人 扎克·塞贝 雷蒙德·洛

杰弗里·A·尤伯萨克斯

苏尼尔·S·钱德兰

埃里克·杰迪戴亚·迪安

达伦·M·普拉特

肯尼思·俊树·竹冈

(74) 专利代理机构 北京市磐华律师事务所

11336

代理人 董巍 谢梅

(51) Int. Cl.

C12N 15/06 (2006. 01)

(56) 对比文件

W0 2008045380 A2, 2008. 04. 17,

US 20030022179 A1, 2003. 01. 30, 全文.

CN 1427079 A, 2003. 07. 02, 全文.

W0 2007038276 A2, 2007. 04. 05, 全文.

审查员 韦炜

权利要求书4页 说明书62页

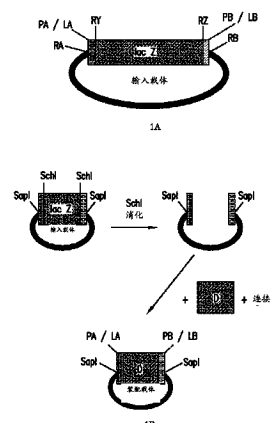
序列表68页 附图15页

(54) 发明名称

多核苷酸装配的组合物和方法

(57) 摘要

本发明提供组合物和从大量组分多核苷酸快速装配为一个或多个装配好的多核苷酸的方法。本发明的方法使用环形核酸载体,所述载体包含DNA片段D,侧邻于可退火接头序列、可退火接头序列对LA和LB、或可退火接头序列/引物结合片段对LA和PB或PA和LB。限制性内切酶消化包含用于装配的DNA片段的大量载体以产生大量DNA片段,包括元件PA-D-LB、LA-D-LB、和LA-D-PB,或D-LB、LA-D-LB、和LA-D。可退火接头序列LA和LB提供了DNA片段的互补末端,其用于宿主细胞介导的同源重组或同时在聚合酶循环装配反应中与引物结合片段PA和PB反应,从而使多种DNA片段有序装配成为一个或多个装配好的多核苷酸。



1. 一种组合物,其包含:

(a) 一个或多个第一核酸分子,其中每个第一核酸分子是环形的并以 5' 至 3' 方向包含:第一限制性酶切位点 RA_0 、选自组 D_0 的任意的 DNA 片段、可退火接头序列 LB_0 、和第二限制性酶切位点 RB_0 ;

(b) 一个或多个中间核酸分子,其中每个中间核酸分子 n 是环形的并以 5' 至 3' 的方向包含:第一限制性酶切位点 RA_n 、第一可退火接头序列 LA_n 、选自组 D_n 的任意 DNA 片段、第二可退火接头序列 LB_n 、和第二限制性酶切位点 RB_n ,其中 n 表示从 1 到中间核酸分子数目的整数;和

(c) 一个或多个最后核酸分子,其中每个最后核酸分子是环形的并以 5' 至 3' 的方向包含:第一限制性酶切位点 RA_m 、第一可退火接头序列 LA_m 、选自组 D_m 的任意 DNA 片段、和第二限制性酶切位点 RB_m ,其中 m 表示比中间核酸分子数目大 1 的整数;

其中 RA_0 至 RB_m 的所述限制性酶切位点可以被相同的 IIS 型限制性核酸内切酶切断;

其中切断 RA_0 至 RB_m 的限制性酶切位点并变性所获得的线形核酸分子之后,每个可退火接头序列 $LB_{(p-1)}$ 能够与可退火接头序列 LA_p 的互补物杂交并作为合成互补多核苷酸的起始点,其中 n 是在 1 到 $(m-1)$ 范围内的整数,其中 p 表示从 1 到 m 的整数,其中所述可退火接头序列各自独立地是至少 24 个核苷酸长,并且其中每个组 $D_0, \dots D_n, \dots D_m$ 独立地由一个或多个 DNA 片段组成。

2. 权利要求 1 的组合物,其中每个所述一个或多个第一核酸分子进一步包含一个引物结合片段 PA, PA 对于所述选自组 D_0 的 DNA 片段位于 5' 位置,并且其中每个所述一个或多个最后核酸分子进一步包含一个引物结合片段 PB, PB 对于选自组 D_m 的 DNA 片段位于 3' 位置。

3. 权利要求 1 或 2 的组合物,其中切断 RA_0 至 RB_m 的限制性酶切位点并变性所获得的线形核酸分子之后,相比于所述组合物中的其它可退火接头序列或它们的互补物,每个可退火接头序列 $LB_{(p-1)}$ 能够与可退火接头序列 LA_p 的互补物选择性杂交。

4. 权利要求 1 或 2 的组合物,其中每个可退火接头 $LB_{(p-1)}$ 与可退火接头序列 LA_p 或其互补物在序列上是相同的。

5. 权利要求 1 或 2 的组合物,其包含一个第一核酸分子和一个最后核酸分子。

6. 权利要求 1 或 2 的组合物,其中 RA_0 至 RB_m 的所述限制性酶切位点可以被限制性核酸内切酶 SapI 或 LguI 切断。

7. 权利要求 1 的组合物,其中两个或多个可退火接头序列各自具有至少 60° C 的解链温度。

8. 权利要求 1 或 2 的组合物,其中两个或多个可退火接头序列各自独立地具有至少 70% 的 G-C 含量和至少 70° C 的解链温度。

9. 权利要求 1 或 2 的组合物,其中两个或多个可退火接头序列各自独立地具有至少 30% 的 A-T 含量和至少 65° C 的解链温度,并包含基序

5' ANNNNNNNNANNNAANTANNTTANANA-3', 其中 A 代表腺嘌呤, N 代表任何核苷酸, 而 T 代表胸腺嘧啶。

10. 权利要求 1 或 2 的组合物,其中两个或多个可退火接头序列具有选自由 SEQ ID NO: 1-8 或其互补物所组成的组的序列。

11. 权利要求 1 或 2 的组合物,其中每个可退火接头序列具有选自 SEQ ID NO :1-8 或其互补物所组成的组的序列。

12. 权利要求 2 的组合物,其中每个引物结合片段具有选自 SEQ ID NO :9-10 或其互补物所组成的组的序列。

13. 权利要求 1 或 2 的组合物,其进一步包含能够切断 RA_0 至 RB_m 的限制性酶切位点的一限制性核酸内切酶。

14. 权利要求 6 的组合物,其进一步包含 SapI 或 LguI 限制性核酸内切酶。

15. 一种组合物,其包含用可以切断 RA_0 至 RB_m 的限制性酶切位点的一限制性核酸内切酶消化权利要求 1 或 2 所述的组合物后所形成的多个线形核酸分子。

16. 权利要求 15 的组合物,其中每个线形核酸分子包含粘性末端。

17. 权利要求 2 的组合物,其进一步包含与引物结合片段 PA 互补的第一引物或其互补物,和与引物结合片段 PB 互补的第二引物或其互补物。

18. 权利要求 1 或 2 的组合物,其进一步包含 DNA 聚合酶。

19. 一种组合物,其包含用 SapI 或 LguI 限制性核酸内切酶消化权利要求 6 所述组合物后所形成的多个线形核酸分子。

20. 一种核酸分子文库,其包含每种下列核酸分子的至少一种:

(a) 第一核酸分子,其中该第一核酸分子为环形并以 5' 至 3' 的方向包含:第一限制性酶切位点 RA_0 、选自组 D_0 的任何 DNA 片段、可退火接头序列 LB_0 、和第二限制性酶切位点 RB_0 ;

(b) 中间核酸分子,其中该中间核酸分子 n 为环形并以 5' 至 3' 的方向包含:第一限制性酶切位点 RA_n 、第一可退火接头序列 LA_n 、选自组 D_n 的任何 DNA 片段、第二可退火接头序列 LB_n 、和第二限制性酶切位点 RB_n ,其中 n 代表从 1 到中间核酸分子数目的整数;和

(c) 最后核酸分子,其中该最后核酸分子为环形并以 5' 至 3' 的方向包含:第一限制性酶切位点 RA_m 、第一可退火接头序列 LA_m 、选自组 D_m 的任何 DNA 片段、第二限制性酶切位点 RB_m ,其中 m 表示比中间核酸分子数目大 1 的整数;

其中 RA_0 至 RB_m 的所述限制性酶切位点可以被相同的 IIS 型限制性核酸内切酶切断;

其中切断 RA_0 至 RB_m 的限制性酶切位点并变性所获得的线形核酸分子之后,每个可退火接头序列 $LB_{(p-1)}$ 能够与可退火接头序列 LA_p 的互补物杂交并作为合成互补多核苷酸的起始点,其中 n 是在 1 到 $(m-1)$ 范围内的整数,其中 p 表示一个从 1 到 m 的整数,其中所述可退火接头序列各自独立地是至少 24 个核苷酸长,并且其中每个组 $D_0, \dots D_n, \dots D_m$ 独立地由一个或多个 DNA 片段组成。

21. 一种从多个组分多核苷酸产生装配好的多核苷酸的方法,该方法包括以下步骤:

(a) 用一限制性内切酶消化装配组合物以产生组分组合物,所述装配组合物包含:

(i) 一个或多个第一核酸分子,其中每个第一核酸分子是环形的并以 5' 至 3' 方向包含:第一限制性酶切位点 RA_0 、选自组 PA 的任意引物结合片段、选自组 D_0 的任意 DNA 片段、可退火接头序列 LB_0 、和第二限制性酶切位点 RB_0 ;

(ii) 一个或多个中间核酸分子,其中每个中间核酸分子 n 是环形的并以 5' 至 3' 的方向包含:第一限制性酶切位点 RA_n 、第一可退火接头序列 LA_n 、选自组 D_n 的任意 DNA 片段、第二可退火接头序列 LB_n 、和第二限制性酶切位点 RB_n ,其中 n 表示从 1 到中间核酸分子数目的整数;和

(iii) 一个或多个最后核酸分子,其中每个最后核酸分子是环形的并以 5' 至 3' 的方向包含:第一限制性酶切位点 RA_m 、第一可退火接头序列 LA_m 、选自组 D_m 的任意 DNA 片段、选自组 PB 的任意引物结合片段、第二限制性酶切位点 RB_m ,其中 m 表示比中间核酸分子数目大 1 的整数;

其中 RA_0 至 RB_m 的所述限制性酶切位点可以被该相同的 IIS 型限制性核酸内切酶切断;

其中通过所述限制性内切酶切断 RA_0 至 RB_m 的限制性酶切位点并变性所获得的线形核酸分子之后,每个可退火接头序列 $LB_{(p-1)}$ 能够与可退火接头序列 LA_p 的互补物杂交并作为合成互补多核苷酸的起始点,其中 n 是在 1 到 $(m-1)$ 范围内的整数,其中 p 表示从 1 到 m 的整数,并且其中每个组 $D_0, \dots D_n, \dots D_m$ 由一个或多个 DNA 片段组成;和

(b) 使所述组分组合物接触 DNA 聚合酶、三磷酸脱氧核糖核苷酸和一个或多个第一引物和一个或多个第二引物,在适合核酸分子变性的条件下,退火可退火接头序列 $LB_{(p-1)}$ 至可退火接头序列 LA_p ,其中 p 表示从 1 到 m 的整数,并从此延伸;其中每个所述第一引物能够与选自组 PA 的一个所述引物结合片段杂交,并且每个第二引物能够与选自组 PB 的一个所述引物结合片段杂交;并且将所述组分组合物进行聚合酶链式反应,

其中以 5' 到 3' 方向包含分别选自组 $D_0, \dots D_n, \dots$ 和 D_m 的一个 DNA 片段的多核苷酸被装配起来。

22. 一种制备包含多核苷酸的宿主细胞的方法,该方法包含如下步骤:

(a) 用一个或多个根据权利要求 21 的方法装配的多核苷酸转化宿主细胞;和

(b) 选择包含所述一个或多个装配好的多核苷酸的宿主细胞。

23. 一种制备包含多核苷酸的宿主细胞的方法,所述方法包含如下步骤:

(a) 用一 IIS 型限制性内切酶消化组合物,所述组合物包含:

(i) 一个或多个第一核酸分子,其中每个第一核酸分子是环形的并以 5' 至 3' 方向包含:第一限制性酶切位点 RA_0 、引物结合片段 PA、选自组 D_0 的任意 DNA 片段、可退火接头序列 LB_0 、和第二限制性酶切位点 RB_0 ;以及和

(iii) 一个或多个最后核酸分子,其中每个最后核酸分子是环形的并以 5' 至 3' 的方向包含:第一限制性酶切位点 RA_m 、可退火接头序列 LA_m 、选自组 D_m 的任意 DNA 片段、引物结合片段 PB、第二限制性酶切位点 RB_m ,其中 m 表示比中间核酸分子数目大 1 的整数;其中 RA_0 至 RB_m 的所述限制性酶切位点可以被该相同的 IIS 型限制性核酸内切酶切断;其中通过所述限制性内切酶切断 RA_0 至 RB_m 的限制性酶切位点并变性所获得的线形核酸分子之后,每个可退火接头序列 $LB_{(p-1)}$ 能够与可退火接头序列 LA_p 的互补物杂交,其中 p 表示从 1 到 m 的整数,其中 n 是在 1 到 $(m-1)$ 范围内的整数,并且其中每个组 $D_0, \dots D_n, \dots D_m$ 由一个或多个 DNA 片段组成;

和

(b) 用从步骤(a)获得的消化的组合物转化宿主细胞,其中每个可退火接头序列 $LB_{(p-1)}$ 包含与可退火接头序列 LA_p 的同源区域,其具有足够长度以起始宿主细胞介导的 $LB_{(p-1)}$ 和 LA_p 之间的同源重组,其中所述同源重组导致所述多核苷酸的装配,其中 p 表示从 1 到 m 的整数;和

(c) 选择包含所述装配好的多核苷酸的宿主细胞,其中所述装配好的多核苷酸以 5' 到 3' 方向包含选自 $D_0, \dots D_n, \dots$ 和 D_m 每一组的一个 DNA 片段。

24. 由权利要求 23 的方法制备的包含多核苷酸的宿主细胞。

25. 权利要求 23 的方法, 其中, 所述组合物进一步包含:

(ii) 一个或多个中间线形核酸分子, 其中每个中间线形核酸分子 n 以 5' 至 3' 方向包含: 第一限制性酶切位点 RA_n 、第一可退火接头序列 LA_n 、选自组 D_n 的任意 DNA 片段、第二可退火接头序列 LB_n 、和第二限制性酶切位点 RB_n , 其中 n 表示从 1 到中间核酸分子数目的整数。

26. 由权利要求 25 的方法制备的包含多核苷酸的宿主细胞。

多核苷酸装配的组合物和方法

[0001] 本申请要求根据 35U. S. C. § 119(e) 的于 2008 年 11 月 19 日提交的美国临时申请号 61/116109 和于 2009 年 3 月 23 日提交的美国临时专利申请号 61/162230 的利益,其内容在此整体引用作为参考。

1. 发明领域

[0002] 本申请一般地涉及重组 DNA 技术领域,尤其是大量的 DNA 片段有序装配为装配好的多核苷酸的改进的方法。

2. 发明背景

[0004] 多核苷酸的重组可以用本领域公知的多种方法来完成。重组核酸的传统技术使用限制性内切酶和连接酶以创造新的核酸分子。重组分子例如克隆和表达载体可以用于将感兴趣的核酸序列整合进入宿主细胞的基因组,和 / 或驱动一个或多个感兴趣的基因的表达。在所述细胞中用载体来驱动感兴趣的基因的表达,例如在酵母细胞中,需要载体包含使感兴趣的基因能够复制和表达的必要的遗传元件。这些元件可以包括,例如,感兴趣的基因,启动子序列,终止子序列,选择性标记,整合位点,等等。

[0005] 用传统的基于限制性内切酶和连接酶的方法将元件装配进单一的载体是耗时而又费力的。每个亚克隆步骤,即将新的核酸片段引入已存在的多核苷酸,需要在引入新加片段之前对所得到的克隆进行筛选和表征。通过平末端连接而产生的克隆需要确认所述片段以正确的方向引入。另一方面,粘性末端连接需要受体片段上用于产生粘性末端的限制性酶切位点也存在于供体片段上,但又不能在供体片段中打断感兴趣的序列的位点。因此,选择可使用的限制性酶切位点完全依赖于要加入的组合物片段,并且每个案例都必须小心考虑。另外,这些方法经常将外源核酸序列引入到所获得的克隆中,所述序列能够干扰需要的基因产物的结构和功能。进一步限制基限制性内切酶的克隆方法效率的因素是能够在在一个反应中连接在一起的核酸分子数量的固有限制。

[0006] 聚合酶链式反应 (PCR) 是一种强大的技术,通过其可以体外扩增特定的多核苷酸,包括基因组 DNA、cDNA 和 mRNA。PCR 典型的包括在一定条件下使目标核酸的分离的互补链接触两个寡核苷酸引物,所述条件允许在两条链上的互补引物延伸产物的形成。这些链作为需要的核酸序列合成拷贝的模板。通过在自动系统中重复分离和合成步骤,可以实现目标序列的指数复制。

[0007] PCR 的一种方法,称为“重叠区扩增拼接法”(“SOE”,参见例如美国专利号 5023171),使得不用限制性内切酶或连接酶而在精确接点装配 DNA 分子变得方便。用于重组的组分片段使用独特设计的引物在分离的聚合酶链式反应中产生,所述引物产生具有彼此互补末端的扩增子。在这些扩增子的混合和变性中,在 3' 末端具有互补序列的链彼此重叠并作为引物。这些重叠区通过 DNA 聚合酶延伸而产生核酸分子,在该分子中原始序列被“拼接”在一起。PCR 的后继循环扩增了所获得的拼接多核苷酸。

[0008] 尽管在结合大量核酸片段方面相对传统的基于连接酶的方法更加高效,SOE 却需要时间来优化引物序列和扩增条件以产生需要的产物。要拼接在一起的片段之间的每个接

点必须分别考虑,并且必须为每个片段涉及一对引物,以使得这些末端兼容。设计 PCR 引物的传统考虑,例如解链温度、G-C 含量、避免发卡和二聚化形成和错误引导位点的严格性,必须被更加仔细的考虑,因为在 SOE 反应中将被拼接的片段的数量增加了。

[0009] 因此,尽管重组 DNA 技术发展了,仍然存在提供快速而有序地装配多核苷酸的改进方法的需求。尤其需要的方法是能简化大量多核苷酸的装配,其具有最少的对中间产物的操作和和表征,而不需要引物优化步骤。这些和其它需求都可以通过本发明的组合物和方法来满足。

[0010] 3. 发明概述

[0011] 本发明提供的组合物和方法允许快速而有序地装配或“缝合”组分多核苷酸,成为装配好的多核苷酸。在一些实施方式里,本发明的方法提供使用环形核酸装配载体。在某些实施方式里,装配载体包含组分多核苷酸,其中所述组分多核苷酸包括一个 DNA 片段,其侧翼是:(i)3' 端的可退火接头;(ii)5' 端的引物结合片段和 3' 端的可退火接头;(iii)3' 端和 5' 端的可退火接头;(iv)5' 端的可退火接头和 3' 端的引物结合片段;或(v)5' 端的可退火接头。

[0012] 在一些实施方式里,多个组分多核苷酸可以通过在单一反应器里提供多个装配载体来缝合在一起。在某些实施方式里,反应器里的组分多核苷酸可以从装配载体分离。在某些实施方式里,所述组分多核苷酸然后可以被变性,可退火接头序列可以被退火为邻近的组分多核苷酸的互补链,而所述组分多核苷酸可以通过重复区延伸拼接法(SOE)缝合进入装配好的多核苷酸,然后进行 PCR。在其它实施方式里,从装配载体分离的组分多核苷酸可以在用所述组分多核苷酸转化的宿主细胞体内通过同源重组而装配进入装配好的多核苷酸。装配好的多核苷酸可以进一步在宿主细胞体内结合并介导同源重组。

[0013] 多核苷酸装配的效率可以通过提供与装配载体一起使用的可退火接头序列的标准试剂盒来促进,例如,如 SEQ ID NO:1-23 所描述的。可退火接头序列在装配反应中提供相邻组分多核苷酸之间的重叠序列。理想地,可退火接头序列在 RNA 和 DNA 水平均缺乏明显的二级结构,不以不希望的方式与另一个发生交叉反应,并具有相对较高的解链温度(T_m)。从而,多个组分多核苷酸可以被缝合在一起而不需要为每个组分多核苷酸设计独特的引物,因此节约了时间和劳动力。本发明提供的组合物和方法可以用于装配多种类型的多核苷酸,包括合成性基因,结构体,克隆载体,表达载体,染色体,基因组,肽文库等等。

[0014] 在一个方面,本发明提供了载体,即装配载体,可用于从多个组分多核苷酸装配一个或多个装配好的多核苷酸。

[0015] 在一些实施方式里,装配载体是环形多核苷酸并以 5' 至 3' 方向包含:限制性酶切位点 RA,可退火接头序列 LA, DNA 片段 D,可退火接头序列 LB,以及限制性酶切位点 RB(即 5' -RA-LA-D-LB-RB-3')。在一些实施方式里,装配载体是环形多核苷酸并以 5' 至 3' 方向包含:限制性酶切位点 RA, DNA 片段 D,可退火接头序列 LB,和限制性酶切位点 RB(即 5' -RA-D-LB-RB-3')。在一些实施方式里,装配载体是环形多核苷酸并以 5' 至 3' 方向包含:限制性酶切位点 RA,可退火接头序列 LA, DNA 片段 D,和限制性酶切位点 RB(即 5' -RA-LA-D-RB-3')。在一些实施方式里,装配载体是环形多核苷酸并以 5' 至 3' 方向包含:限制性酶切位点 RA,引物结合片段 PA, DNA 片段 D,可退火接头序列 LB,和限制性酶切位点 RB(即 5' -RA-PA-D-LB-RB-3')。在一些实施方式里,装配载体是环形多核苷酸并以 5'

至 3' 方向包含 :限制性酶切位点 RA,可退火接头序列 LA,DNA 片段 D,引物结合片段 PB,和限制性酶切位点 RB(即 5'-RA-LA-D-PB-RB-3')。在一些实施方式里,装配载体是环形多核苷酸并以 5' 至 3' 方向包含 :限制性酶切位点 RA,可退火接头序列 LA,DNA 片段 D,可退火接头序列 LB,和限制性酶切位点 RB(即 5'-RA-LA-D-LB-RB-3')。示例性的装配载体在图 1B 和图 2 中提供。

[0016] 在一些实施方式里,引物结合片段(即,PA 或 PB)可以是不与用于产生装配好的多核苷酸的可退火接头序列互补的任何核苷酸序列。在一些实施方式里,引物结合片段包括限制性内切酶识别位点和 / 或切断位点。在一些实施方式里,引物结合片段包含允许的接头序列之一的核酸序列(例如,SEQ ID NO:1-23 之一),或其互补物,不用于特定的装配反应。在一些实施方式里,引物结合片段 PA 的核酸序列选自 SEQ ID NO:24,25 和其互补物。在一些实施方式里,引物结合片段 PB 的核酸序列选自 SEQ ID NO:24,25 和其互补物。在优选实施方式里,引物结合片段 PA 和引物结合片段 PB 在序列上不完全相同。

[0017] 在一些实施方式里,两个或多个可退火接头序列为至少 24 个核苷酸长度并具有至少 60°C 的 T_m 。

[0018] 在一些实施方式里,两个或多个可退火接头序列具有至少 70% 的 G-C 含量和至少 70°C 的 T_m ,并且不形成明显的二级 DNA 结构。在一些实施方式里,可退火接头序列 LA 的核酸序列选自 SEQ ID NO:1-8 及其互补物。在一些实施方式里,可退火接头序列 LB 的核酸序列选自 SEQ ID NO:1-8 及其互补物。在一些实施方式里,可退火接头序列 LA 和可退火接头序列 LB 的核酸序列选自 SEQ ID NO:1-8 及其互补物。

[0019] 在一些实施方式里,两个或多个可退火接头序列具有至少 30% 的 A-T 含量和至少 65°C 的 T_m ,并且不形成明显的二级 DNA 或 RNA 结构。在一些实施方式里,两个或多个可退火接头序列具有低 G-C 含量和至少 65°C 的 T_m ,并包含基序 5'-ANNNNNNNNANNNAAANTANNTTNANA-3',其中 A 代表腺嘌呤,N 代表任何核苷酸,而 T 代表胸腺嘧啶。在一些实施方式里,可退火接头序列 LA 的核酸序列选自 SEQ ID NO:9-23 及其互补物。在一些实施方式里,可退火接头序列 LB 的核酸序列选自 SEQ ID NO:9-23 及其互补物。在一些实施方式里,可退火接头序列 LA 和可退火接头序列 LB 的核酸序列选自 SEQ ID NO:9-23 及其互补物。

[0020] 多个组分多核苷酸的顺序装配可以通过可退火接头序列的选择来控制,所述可退火接头序列在装配载体中侧邻于 DNA 片段。从而,在一些实施方式里,为了确保组分多核苷酸能够以有序方式被装配进去,在特定装配载体中的可退火接头序列 / 可退火接头序列对的序列是不互补的。相似的,在一些实施方式里,在特定装配载体中的引物结合片段 / 可退火接头序列对的序列是不互补的。

[0021] 在一个特别的实施方式里,限制性酶切位点 RA 和 RB 可以被同样的限制性内切酶切断,以利于从装配载体中切除组分多核苷酸。在一些实施方式里,限制性酶切位点 RA 或 RB 可以被限制性内切酶切断并留下 5' 或 3' 突出。在另一个实施方式里,限制性酶切位点 RA 或 RB 被限制性内切酶切断并留下平末端。在一些实施方式里,限制性酶切位点 RA 和 RB 可以被同样的限制性内切酶切断。在另外的实施方式里,限制性酶切位点 RA 和 RB 可以被 IIS 型限制性内切酶切断。在一些实施方式里,限制性酶切位点 RA 和 RB 可以被相同的 IIS 型限制性内切酶切断。在一个特别的实施方式里,限制性酶切位点 RA 和 RB 被 SapI 或 LguI 限制性内切酶所切断。

[0022] 在另一个方面,本发明提供输入载体,其用于制备包含 DNA 片段的装配载体。

[0023] 在一些实施方式里,所述输入载体是环形多核苷酸并以 5' 至 3' 的方向包含:限制性酶切位点 RA,限制性酶切位点 RY,DNA 片段 D,限制性酶切位点 RZ,可退火接头序列 LB,和限制性酶切位点 RB(即,5'-RA-RY-D-RZ-LB-RB-3')。在一些实施方式里,所述输入载体是环形多核苷酸并以 5' 至 3' 的方向包含:限制性酶切位点 RA,可退火接头序列 LA,限制性酶切位点 RY,DNA 片段 D,限制性酶切位点 RZ,和限制性酶切位点 RB(即,5'-RA-LA-RY-D-RZ-RB-3')。在一些实施方式里,所述输入载体是环形多核苷酸并以 5' 至 3' 的方向包含:限制性酶切位点 RA,可退火接头序列 LA,限制性酶切位点 RY,DNA 片段 D,限制性酶切位点 RZ,可退火接头序列 LB,和限制性酶切位点 RB(即,5'-RA-LA-RY-D-RZ-LB-RB-3')。在一些实施方式里,所述输入载体是环形多核苷酸并以 5' 至 3' 的方向包含:限制性酶切位点 RA,引物结合片段 PA,限制性酶切位点 RY,DNA 片段 D,限制性酶切位点 RZ,可退火接头序列 LB,和限制性酶切位点 RB(即,5'-RA-PA-RY-D-RZ-LB-RB-3')。在一些实施方式里,所述输入载体是环形多核苷酸并以 5' 至 3' 的方向包含:限制性酶切位点 RA,可退火接头序列 LA,限制性酶切位点 RY,DNA 片段 D,限制性酶切位点 RZ,引物结合片段 PB,和限制性酶切位点 RB(即,5'-RA-LA-RY-D-RZ-PB-RB-3')。图 1A 提供了示例性的输入载体。

[0024] 用能够切断 RY 和 RZ 的一种或多种限制性内切酶消化输入载体可以建立能够接受 DNA 片段的线形载体。DNA 片段通过标准克隆技术连接入 RY 和 RZ 位点以形成本发明的装配载体。在一些实施方式里,输入载体的限制性酶切位点 RY 和 RZ 可以被同样的限制性内切酶切断。在一些实施方式里,输入载体的限制性酶切位点 RY 和 RZ 可以被 IIS 型限制性内切酶切断。在一些实施方式里,输入载体的限制性酶切位点 RY 和 RZ 可以被相同的 IIS 型限制性内切酶切断。在一个特别的实施方式里,所述 IIS 型限制性内切酶是 *SchI* 或 *MlyI*。

[0025] 在一些实施方式里,输入载体的限制性酶切位点 RA 和 RB 可以被同样的限制性内切酶切断。在一些实施方式里,输入载体的限制性酶切位点 RA 和 RB 可以被 IIS 型限制性内切酶切断。在一些实施方式里,输入载体的限制性酶切位点 RA 和 RB 可以被相同的 IIS 型限制性内切酶切断。在一个特别的实施方式里,所述 IIS 型限制性内切酶是 *SapI* 或 *LguI*。

[0026] 在另一方面,本发明提供装配组合物,其包含多种装配载体,所述装配载体用于从多种组分多核苷酸装配为一种或多种装配好的多核苷酸。在一些实施方式里,所述装配组合物包含:

[0027] (a) 一个或多个第一核酸分子,其中每个第一核酸分子是环形的并以 5' 至 3' 方向包含:第一限制性酶切位点 RA_0 、选自组 D_0 的任意的 DNA 片段、可退火接头序列 LB_0 ,和第二限制性酶切位点 RB_0 ;

[0028] (b) 一个或多个中间核酸分子,其中每个中间核酸分子 n 是环形的并以 5' 至 3' 的方向包含:第一限制性酶切位点 RA_n 、第一可退火接头序列 LA_n 、选自组 D_n 的任意 DNA 片段、第二可退火接头序列 LB_n ,和第二限制性酶切位点 RB_n ,其中 n 表示从 1 到中间核酸分子数目的整数;和

[0029] (c) 一个或多个最后核酸分子,其中每个最后核酸分子是环形的并以 5' 至 3' 的方向包含:第一限制性酶切位点 RA_m 、第一可退火接头序列 LA_m 、选自组 D_m 的任意 DNA 片段、和第二限制性酶切位点 RB_m ,其中 m 表示比中间核酸分子数目大 1 的整数;

[0030] 其中切断限制性酶切位点 RA_0 至 RB_m , 变性所获得的线形核酸分子, 每个可退火接头序列 $LB_{(p-1)}$ 能够与可退火接头序列 LA_p 的互补物杂交, 其中 n 是在 1 到 $(m-1)$ 范围内的整数, 其中 p 表示一个从 1 到 m 的整数, 并且其中每个组 $D_0, \dots D_n, \dots D_m$ 独立地由一个或多个 DNA 片段组成。

[0031] 在某些实施方式里, 一个或多个第一核酸分子进一步包含一个引物结合片段 PA 位于所述选自组 D_0 的 DNA 片段的 5' 位置。在某些实施方式里, 一个或多个最后核酸分子进一步包含一个引物结合片段 PB 位于所述选自组 D_m 的 DNA 片段的 3' 位置。

[0032] 在某些实施方式里, 装配组合物包含两个或更多个中间核酸分子。在某些实施方式里, 装配组合物包含三个或更多个中间核酸分子。在某些实施方式里, 装配组合物包含四个或更多个中间核酸分子。在某些实施方式里, 装配组合物包含五个或更多个中间核酸分子。在某些实施方式里, 所述组合物包含六个或更多个中间核酸分子。在某些实施方式里, 装配组合物包含七个或更多个中间核酸分子。在某些实施方式里, 装配组合物包含八个或更多个中间核酸分子。在某些实施方式里, 装配组合物包含九个或更多个中间核酸分子。在某些实施方式里, 装配组合物包含十个或更多个中间核酸分子。在某些实施方式里, 装配组合物包含十五个或更多个中间核酸分子。在某些实施方式里, 装配组合物包含二十个或更多个中间核酸分子。

[0033] 在某些实施方式里, m 等于 1。在某些实施方式里, m 等于 2。在某些实施方式里, m 等于 3。在某些实施方式里, m 等于 4。在某些实施方式里, m 等于 5。在某些实施方式里, m 等于 6。在某些实施方式里, m 等于 7。在某些实施方式里, m 等于 8。在某些实施方式里, m 等于 9。在某些实施方式里, m 等于 10。在某些实施方式里, m 等于或大于 10。

[0034] 在一些实施方式里, 切断 RA_0 至 RB_m 的限制性酶切位点和变形所获得的线形核酸分子之后, 相比于装配组合物中的其它可退火接头序列或它们的互补物, 每个可退火接头序列 $LB_{(p-1)}$ 能够选择性地与可退火接头序列 LA_p 的互补物杂交。在一些实施方式里, 每个可退火接头序列 $L_{(p-1)}$ 在序列上与可退火接头序列 LA_p 相同。

[0035] 在一个特别的实施方式里, RA_0 至 RB_m 的限制性酶切位点可以被同样的限制性内切酶切断从而利于从装配载体中切除组分多核苷酸。在一些实施方式里, RA_0 至 RB_m 的限制性酶切位点可以被 SapI 和 / 或 LguI 限制性内切酶切断。

[0036] 在另一方面, 本发明提供包含多种线形核酸分子的组分组合物, 其中所述线形核酸分子能通过用一种或多种能够切断 RA_0 至 RB_m 的限制性酶切位点的限制性内切酶消化装配组合物而形成, 其中所述装配组合物包含:

[0037] (a) 一个或多个第一核酸分子, 其中每个第一核酸分子是环形的并以 5' 至 3' 方向包含: 第一限制性酶切位点 RA_0 , 选自组 D_0 的任意 DNA 片段, 可退火接头序列 LB_0 , 和第二限制性酶切位点 RB_0 ;

[0038] (b) 一个或多个中间核酸分子, 其中每个中间核酸分子 n 是环形的并以 5' 至 3' 的方向包含: 第一限制性酶切位点 RA_n , 第一可退火接头序列 LA_n , 选自组 D_n 的任意 DNA 片段, 第二可退火接头序列 LB_n , 和第二限制性酶切位点 RB_n , 其中 n 表示从 1 到中间核酸分子数目的整数; 和

[0039] (c) 一个或多个最后核酸分子, 其中每个最后核酸分子是环形的并以 5' 至 3' 的方向包含: 第一限制性酶切位点 RA_m , 第一可退火接头序列 LA_m , 选自组 D_m 的任意 DNA 片段, 第

二限制性酶切位点 RB_m , 其中 m 表示比中间核酸分子数目大 1 的整数;

[0040] 其中切断 RA_0 至 RB_m 的限制性酶切位点, 变性所获得的线形核酸分子之后, 每个可退火接头序列 $LB_{(p-1)}$ 能够与可退火接头序列 LA_p 的互补物杂交, 其中 n 是在 1 到 $(m-1)$ 范围内的整数, 其中 p 表示一个从 1 到 m 的整数, 并且其中每个组 $D_0, \dots D_n, \dots D_m$ 由一个或多个 DNA 片段组成。

[0041] 在某些实施方式里, 一个或多个第一核酸分子进一步包含引物结合片段 PA 位于选自组 D_0 的 DNA 片段的 5' 位置。在某些实施方式里, 一个或多个最后核酸分子进一步包含引物结合片段 PB 位于选自组 D_m 的 DNA 片段的 3' 位置。

[0042] 在某些实施方式里, 组分组合物包含两个或更多个中间核酸分子。在某些实施方式里, 组分组合物包含三个或更多个中间核酸分子。在某些实施方式里, 组分组合物包含四个或更多个中间核酸分子。在某些实施方式里, 组分组合物包含五个或更多个中间核酸分子。在某些实施方式里, 所述组合物包含六个或更多个中间核酸分子。在某些实施方式里, 组分组合物包含七个或更多个中间核酸分子。在某些实施方式里, 组分组合物包含八个或更多个中间核酸分子。在某些实施方式里, 组分组合物包含九个或更多个中间核酸分子。在某些实施方式里, 组分组合物包含十个或更多个中间核酸分子。在某些实施方式里, 组分组合物包含十五个或更多个中间核酸分子。在某些实施方式里, 组分组合物包含二十个或更多个中间核酸分子。

[0043] 在某些实施方式里, m 等于 1。在某些实施方式里, m 等于 2。在某些实施方式里, m 等于 3。在某些实施方式里, m 等于 4。在某些实施方式里, m 等于 5。在某些实施方式里, m 等于 6。在某些实施方式里, m 等于 7。在某些实施方式里, m 等于 8。在某些实施方式里, m 等于 9。在某些实施方式里, m 等于 10。在某些实施方式里, m 等于或大于 10。

[0044] 在另一方面, 本发明提供一种用于依照本发明提供的方法来装配多种多核苷酸的试剂盒。在一些实施方式里, 所述试剂盒包含: (a) 一种或多种本文描述的输入载体; (b) 一种或多种能够切断输入载体的限制性酶切位点 RA 和 RB 的限制性内切酶; 和 (c) 一种或多种能够切断输入载体的限制性酶切位点 RY 和 RZ 的限制性内切酶。

[0045] 在另一方面, 本发明提供一种核酸分子文库。在一些实施方式里, 所述文库的核酸分子包含第一限制性酶切位点 RA, DNA 片段 D, 可退火接头序列 LB, 和第二限制性酶切位点 RB。在一些实施方式里, 所述文库的核酸分子包含第一限制性酶切位点 RA, 引物结合片段 PA, DNA 片段 D, 可退火接头序列 LB, 和第二限制性酶切位点 RB。在一些实施方式里, 所述文库的核酸分子包含第一限制性酶切位点 RA, 可退火接头序列 LA, DNA 片段 D, 可退火接头序列 LB, 和第二限制性酶切位点 RB。在一些实施方式里, 所述文库的核酸分子包含第一限制性酶切位点 RA, 可退火接头序列 LA, DNA 片段 D, 和第一限制性酶切位点 RB。在一些实施方式里, 所述文库的核酸分子包含第一限制性酶切位点 RA, 可退火接头序列 LA, DNA 片段 D, 引物结合片段 PB, 和第二限制性酶切位点 RB。

[0046] 在一些实施方式里, 所述文库包含下列每种载体的至少一种:

[0047] (a) 环形多核苷酸组成的载体, 所述多核苷酸以 5' 至 3' 的方向包含: 限制性酶切位点 RA, DNA 片段 D, 可退火接头序列 LB, 和限制性酶切位点 RB;

[0048] (b) 环形多核苷酸组成的载体, 所述多核苷酸以 5' 至 3' 的方向包含: 限制性酶切位点 RA, 可退火接头序列 LA, DNA 片段 D, 可退火接头序列 LB, 和限制性酶切位点 RB; 和

[0049] (c) 环形多核苷酸组成的载体,所述多核苷酸以 5' 至 3' 的方向包含:限制性酶切位点 RA,可退火接头序列 LA, DNA 片段 D,和限制性酶切位点 RB₀。

[0050] 在一些实施方式里,所述文库包含下列每种载体的至少一种:

[0051] (a) 环形多核苷酸组成的载体,所述多核苷酸以 5' 至 3' 的方向包含:限制性酶切位点 RA,引物结合片段 PA, DNA 片段 D,可退火接头序列 LB,和限制性酶切位点 RB;

[0052] (b) 环形多核苷酸组成的载体,所述多核苷酸以 5' 至 3' 的方向包含:限制性酶切位点 RA,可退火接头序列 LA, DNA 片段 D,可退火接头序列 LB,和限制性酶切位点 RB;和

[0053] (c) 环形多核苷酸组成的载体,所述多核苷酸以 5' 至 3' 的方向包含:限制性酶切位点 RA,可退火接头序列 LA, DNA 片段 D,引物结合片段 PB,和限制性酶切位点 RB₀。

[0054] 在一些实施方式里,所述 DNA 片段 D 包含核酸序列,所述核酸序列选自选择性标记,启动子,基因组靶向序列,编码表位标签的核酸序列,和编码感兴趣的基因的核酸序列,编码终止密码子和 lacZ 的核酸序列。

[0055] 在一些实施方式里,所述文库包含下列每种核酸分子的至少一种:

[0056] (a) 第一核酸分子,其中所述第一核酸分子是环形的并以 5' 至 3' 方向包含:第一限制性酶切位点 RA₀, DNA 片段 D₀,可退火接头序列 LB₀,和第二限制性酶切位点 RB₀;

[0057] (b) 中间核酸分子,其中所述中间核酸分子 n 是环形的并以 5' 至 3' 的方向包含:第一限制性酶切位点 RA_n,第一可退火接头序列 LA_n, DNA 片段 D_n,第二可退火接头序列 LB_n,和第二限制性酶切位点 RB_n,并且其中 n 表示从 1 到中间核酸分子数目的整数;和

[0058] (c) 最后核酸分子,其中所述最后核酸分子是环形的并以 5' 至 3' 的方向包含:第一限制性酶切位点 RA_m,可退火接头序列 LA_m, DNA 片段 D_m,第二限制性酶切位点 RB_m,其中 m 表示比中间核酸分子数目大 1 的整数;

[0059] 其中切断 RA₀ 至 RB_m 的限制性酶切位点,变性所获得的线形核酸分子之后,每个可退火接头序列 LB_(p-1) 能够与可退火接头序列 LA_p 的互补物杂交,其中 n 是在 1 到 (m-1) 范围内的整数,其中 p 表示一个从 1 到 m 的整数。在一些实施方式里,第一核酸分子进一步包含引物结合片段 PA 位于选自 D₀ 的 DNA 片段的 5' 位置。在一些实施方式里,最后核酸分子进一步包含引物结合片段 PB 位于选自 D_m 的 DNA 片段的 3' 位置。

[0060] 在某些实施方式里,所述文库包含两种或更多种中间核酸分子。在某些实施方式里,所述文库包含三种或更多种中间核酸分子。在某些实施方式里,所述文库包含四种或更多种中间核酸分子。在某些实施方式里,所述文库包含五种或更多种中间核酸分子。在某些实施方式里,所述文库包含六种或更多种中间核酸分子。在某些实施方式里,所述文库包含七种或更多种中间核酸分子。在某些实施方式里,所述文库包含八种或更多种中间核酸分子。在某些实施方式里,所述文库包含九种或更多种中间核酸分子。在某些实施方式里,所述文库包含十种或更多种中间核酸分子。在某些实施方式里,所述文库包含十五种或更多种中间核酸分子。在某些实施方式里,所述文库包含二十种或更多种中间核酸分子。

[0061] 在某些实施方式里,m 等于 1。在某些实施方式里,m 等于 2。在某些实施方式里,m 等于 3。在某些实施方式里,m 等于 4。在某些实施方式里,m 等于 5。在某些实施方式里,m 等于 6。在某些实施方式里,m 等于 7。在某些实施方式里,m 等于 8。在某些实施方式里,m 等于 9。在某些实施方式里,m 等于 10。在某些实施方式里,m 等于或大于 10。

[0062] 在另一方面,本发明提供从多种组分多核苷酸装配为一种或多种装配好的多核苷

酸的方法,包含如下步骤:

[0063] (a) 用一种或多种限制性内切酶消化装配组合物以产生组分组合物,所述装配组合物包含:

[0064] (i) 一个或多个第一核酸分子,其中每个第一核酸分子是环形的并以 5' 至 3' 方向包含:第一限制性酶切位点 RA_0 ,选自组 PA 的任意引物结合片段,选自组 D_0 的任意 DNA 片段,可退火接头序列 LB_0 ,和第二限制性酶切位点 RB_0 ;

[0065] (ii) 一个或多个中间核酸分子,其中每个中间核酸分子 n 是环形的并以 5' 至 3' 的方向包含:第一限制性酶切位点 RA_n ,第一可退火接头序列 LA_n ,选自组 D_n 的任意 DNA 片段,第二可退火接头序列 LB_n ,和第二限制性酶切位点 RB_n ,其中 n 表示从 1 到中间核酸分子数目的整数;和

[0066] (iii) 一个或多个最后核酸分子,其中每个最后核酸分子是环形的并以 5' 至 3' 的方向包含:第一限制性酶切位点 RA_m ,第一可退火接头序列 LA_m ,选自组 D_m 的任意 DNA 片段,选自组 PB 的任意引物结合片段,第二限制性酶切位点 RB_m ,其中 m 表示比中间核酸分子数目大 1 的整数;其中切断 RA_0 至 RB_m 的限制性酶切位点,变性所获得的线形核酸分子之后,每个可退火接头序列 $LB_{(p-1)}$ 能够与可退火接头序列 LA_p 的互补物杂交,其中 n 是在 1 到 $(m-1)$ 范围内的整数,其中 p 表示一个从 1 到 m 的整数,并且其中每个组 $D_0, \dots D_n, \dots D_m$ 由一个或多个 DNA 片段组成;

[0067] 其中所述一种或多种限制性内切酶能够切断 RA_0 至 RB_m 的限制性酶切位点;和

[0068] (b) 使组分组合物接触 DNA 聚合酶,三磷酸脱氧核糖核苷酸和一个或多个第一引物和一个或多个第二引物,在适合核酸分子变性的条件下,退火可退火接头序列 $LB_{(p-1)}$ 至可退火接头序列 LA_p ,并从此延伸;其中每个所述第一引物能够与选自组 PA 的所述引物结合片段之一杂交,而每个第二引物能够与选自组 PB 的所述引物结合片段之一杂交;并且将所述组分组合物进行聚合酶链式反应,

[0069] 其中以 5' 到 3' 方向包含分别选自组 $D_0, \dots D_n, \dots$ 和 D_m 的一个 DNA 片段的多核苷酸被装配起来, p 表示从 1 到 m 的整数。

[0070] 在某些实施方式里,装配组合物包含两个或更多个中间核酸分子。在某些实施方式里,装配组合物包含三个或更多个中间核酸分子。在某些实施方式里,装配组合物包含四个或更多个中间核酸分子。在某些实施方式里,装配组合物包含五个或更多个中间核酸分子。在某些实施方式里,装配组合物包含六个或更多个中间核酸分子。在某些实施方式里,装配组合物包含七个或更多个中间核酸分子。在某些实施方式里,装配组合物包含八个或更多个中间核酸分子。在某些实施方式里,装配组合物包含九个或更多个中间核酸分子。在某些实施方式里,装配组合物包含十个或更多个中间核酸分子。在某些实施方式里,装配组合物包含十五个或更多个中间核酸分子。在某些实施方式里,装配组合物包含二十个或更多个中间核酸分子。

[0071] 在某些实施方式里, m 等于 1。在某些实施方式里, m 等于 2。在某些实施方式里, m 等于 3。在某些实施方式里, m 等于 4。在某些实施方式里, m 等于 5。在某些实施方式里, m 等于 6。在某些实施方式里, m 等于 7。在某些实施方式里, m 等于 8。在某些实施方式里, m 等于 9。在某些实施方式里, m 等于 10。在某些实施方式里, m 等于或大于 10。

[0072] 在一些实施方式里,装配组合物包含一个第一核酸分子和一个最后核酸分子。在

其它实施方式里,装配组合物包含多于一个第一核酸分子和多于一个最后核酸分子,这种装配方法提供了以组合的方式将多种组分多核苷酸有序装配为多个装配好的多核苷酸的有序装配方法。在某些实施方式里,装配组合物包含至少两种核酸分子,所述核酸分子包含相同的可退火接头序列 LA 和 LB,和相同的引物结合片段 PA 或 PB,或者相同的可退火接头序列对 LA 和 LB,或者相同的可退火接头序列 / 引物结合片段对 LA 和 PB,或者 LB 和 PA。

[0073] 在另一方面,本发明提供产生包含装配好的多核苷酸的宿主细胞的方法。在一些实施方式里,所述方法包含用由本文描述的多核苷酸装配方法产生的装配好的多核苷酸转化宿主细胞。在其它实施方式里,所述方法包含用由本文描述的多核苷酸装配方法产生的多种装配好的多核苷酸转化宿主细胞。在一个特别的实施方式里,所述宿主细胞通过同源重组将两种或多种装配好的多核苷酸组合进入一种或多种组合的多核苷酸。在另外的实施方式里,所述方法包含用多个组分多核苷酸转化所述宿主细胞并且允许宿主细胞通过同源重组产生一种或多种装配好的组合多核苷酸。

[0074] 在另一方面,本发明提供产生多种宿主细胞的方法,所述宿主细胞包含大量装配好的多核苷酸。在一些实施方式里,所述多种宿主细胞通过用包含多种装配好的多核苷酸的组合物转化宿主细胞而产生,其中所述装配好的多核苷酸由组分多核苷酸组合装配而产生。在其它实施方式里,所述多种宿主细胞通过用包含多种装配好的多核苷酸的组合物转化宿主细胞而产生,其中所述装配好的多核苷酸的至少两种包含选择性标记的非功能性片段,其通过宿主细胞介导的同源重组而产生功能性选择性标记,然后筛选包含组合多核苷酸的宿主细胞。在另外的实施方式里,所述多种宿主细胞通过用组分组合物转化宿主细胞的组合方法而产生,其中所述组分组合物包含多种组分多核苷酸,至少两种组分多核苷酸包含同样的可退火接头序列 LA 或 LB 或者同样的可退火序列对 LA 和 LB,然后筛选包含装配好的多核苷酸的宿主细胞。

[0075] 在另一方面,本发明提供具有选自 SEQ ID NO :1-25 的序列的多核苷酸。

[0076] 在另一方面,本发明提供包含选自 SEQ ID NO :1-25 的一个或多个序列的多核苷酸。

4. 附图说明

[0077] 图 1A 提供了用于制备本发明的装配载体的输入载体的示意图。该载体包含限制性酶切位点 RA₀,引物结合片段 PA 或可退火接头序列 LA,限制性酶切位点 RY,DNA 片段 D,限制性酶切位点 RZ,引物结合片段 PB 或可退火接头序列 LB,和限制性酶切位点 RB。

[0078] 图 1B 提供了制备输入载体的示例性方法,所述输入载体接受 DNA 片段以形成装配载体。示例性地,RY = RZ = SchI。用能够产生平末端的 IIS 型限制性内切酶 SchI 消化,导致载体与连接位点分离从而开放从而与 DNA 片段 (D) 融合。D 通过平末端连接进入输入载体可以通过传统方法实现,并施用例如 T4DNA 连接酶。

[0079] 图 2 展示了包含多个装配载体 (第一,中间,和最后) 的装配组合物的示意图,分别包含感兴趣的 DNA 片段 (D₀,D_n,D_m)。第一核酸分子包含第一限制性酶切位点 RA₀,引物结合片段 PA,DNA 片段 D₀,可退火接头序列 LB₀,和第二限制性酶切位点 RB₀。一个或多个中间核酸分子包含第一限制性酶切位点 RA_n,第一可退火接头序列 LA_n,DNA 片段 D_n,第二可退火接头序列 LB_n,和第二限制性酶切位点 RB_n,其中 n 表示从 1 到中间核酸分子数目的整数;而

最后核酸分子包含第一限制性酶切位点 RA_m ，可退火接头序列 LA_m ，DNA 片段 D_m ，引物结合片段 PB，第二限制性酶切位点 RB_m ，其中 m 表示比中间核酸分子数目大 1 的整数。

[0080] 图 3 展示了装配的示例性方法，即，从四种 (4) 组分多核苷酸“缝合”为装配好的多核苷酸。将包含用于装配的 DNA 片段的装配载体注入单独一个试管中并用 SapI 消化以从装配载体骨架中释放组分多核苷酸片段。在 SapI 加热失活之后，所述组分多核苷酸片段用于变性条件，继之以退火条件且足够互补可退火接头对的杂交。在 DNA 聚合酶和 dNTP 存在下的引物延伸之后，加入与 PA 和 PB 互补的引物，继之以传统的 PCR 扩增。作为装配反应的结果，产生了包含组分多核苷酸 D_0, D_1, D_2 ，和 D_3 以 5' 到 3' 方向装配起来的装配好的多核苷酸。

[0081] 图 4 显示了 pRYSE 载体图。

[0082] 图 5 显示了通过装配 2 到 4 组分多核苷酸 (装配表 7 中的 1 到 6) 而获得的装配好的多核苷酸，使用不同的方法以移除 SapI 限制性内切酶 (柱纯化或加热失活)，不同的装配载体 DNA 浓度 (5ng (低 DNA 浓度) 或 50ng (高 DNA 浓度) 的最小片段，和相等摩尔浓度的所有其它片段)，和 PCR 扩增的不同退火温度 (54°C 和 72°C)。

[0083] 图 6 显示了通过装配 6 或 9 组分多核苷酸 (装配表 7 中的 7，和 13 到 16) 而获得的装配好的多核苷酸，使用不同的 DNA 聚合酶 (Phusion (New Engl 和 Biolabs, Ipswich, MA) 和 PfuUltraII (Stratagene/Agilent, La Jolla, CA))。

[0084] 图 7 显示了 pMULE 载体图。该 pMULE 输入载体不同于 pRYSE 输入载体之处在于它缺少引物结合片段或可退火接头序列。

[0085] 图 8 显示了通过宿主细胞介导的同源重组来组合装配好的多核苷酸进入组合多核苷酸的示例性方法，其中组合多核苷酸被整合进入宿主细胞的染色体。装配好的多核苷酸 A 包含编码选择性标记的第一非功能性片段的 DNA 片段 D_{m1} 和编码上游基因组靶向序列的 DNA 片段 D_{01} 。装配好的多核苷酸 B 包含编码选择性标记的第二非功能性片段的 DNA 片段 D_{m2} 和编码下游基因组靶向序列的 DNA 片段 D_{02} 。宿主细胞在 DNA 片段 D_{m1} 和 D_{m2} 的同源区域联合装配好的多核苷酸 A 和装配好的多核苷酸 B，以形成包含功能性选择性标记的组合多核苷酸，并利用 DNA 片段 D_{01} 和 D_{02} 编码的基因组靶向序列来将组合多核苷酸通过同源重组插入到其染色体中。

[0086] 图 9 显示了通过宿主细胞中同源重组而产生装配好的多核苷酸和将所述装配好的多核苷酸整合进入宿主细胞的染色体的示例性方法。第一步，用限制性内切酶消化包含装配载体的装配组合物，导致组分多核苷酸从装配载体骨架中切除出来。第二步，组分多核苷酸被引入宿主细胞，在这里它们在可退火接头序列的同源区域结合起来从而形成装配好的多核苷酸，并且该装配好的多核苷酸整合进入宿主细胞的染色体。

[0087] 图 10 显示了在同一个反应中从七种 (7) 组分多核苷酸装配得到多种装配好的多核苷酸的示例性方法。将包含用于装配的 DNA 片段的装配载体注入一个单独的试管，并用 SspI 消化以从装配载体骨架中释放组分多核苷酸。在 SapI 加热失活之后，所述组分多核苷酸片段用于变性条件，继之以退火条件且足够互补可退火接头对的杂交。在 DNA 聚合酶和 dNTP 存在下的引物延伸之后，加入与 PA 和 PB 互补的引物，继之以传统的 PCR 扩增。作为装配反应的结果，产生了包含组分多核苷酸 $D_{01/02}, D_{1/2}, D_3$ ，和 $D_{41/42}$ 以 5' 到 3' 方向装配起来的装配好的多核苷酸。

[0088] 图 11 显示了产生包含组合的组合多核苷酸的多种宿主细胞的示例性方法。装配好的多核苷酸 A1 和 A2, 分别包含相同的上游基因组靶向序列和相同的选择性标记的第一非功能性部分, 和装配好的多核苷酸 B1 和 B2, 分别包含相同的下游基因组靶向序列和相同的选择性标记的第二非功能性部分, 通过宿主细胞介导的同源重组组合性结合起来一产生 4 种不同的组合多核苷酸, A1/B1, A1/B2, A2/B1, 和 A2/B2, 分别包含功能性选择性标记, 其能够被插入染色体以产生四种不同的宿主细胞。

[0089] 图 12A 显示了实施例 10 中的组分多核苷酸, 所述组分多核苷酸用于高通量产生组合的装配好的多核苷酸和包含组合的装配好的多核苷酸和组合多核苷酸的酵母细胞, 以及所期望的装配好的多核苷酸和组合多核苷酸。US = 上游基因组靶向序列, DS = 下游基因组靶向序列, P = 多种启动子序列, G = 多种蛋白质编码序列, URA = 选择性标记的 5' 片段, RA3 = 选择性标记的 3' 片段, PA = 引物结合片段 PmeI-5', PB = 引物结合片段 PmeI-3', LB₀ = 可退火接头序列 RYSE 2, LA_{n1} = 可退火接头序列 RYSE 2, LB_{n1} = 可退火接头序列 RYSE 15, LA_{n2} = 可退火接头序列 RYSE 3, LB_{n2} = 可退火接头序列 RYSE16, LA_{n3} = 可退火接头序列 RYSE 15, LB_{n3} = 可退火接头序列 RYSE 3, LA_{n4} = 可退火接头序列 RYSE 16, LB_{n4} = 可退火接头序列 RYSE 4, LA_{m1} = 可退火接头序列 RYSE 3, LA_{m2} = 可退火接头序列 RYSE 4, LA_{m3} = 可退火接头序列 RYSE 3。

[0090] 图 12B 显示了示例性的装配好的多核苷酸 (方框内), 其由实施例 10 所描述的方法产生并在 1% 琼脂糖凝胶上拆分。

[0091] 图 12C 显示了如实施例 10 所描述获得的示例性细胞克隆的酶切分析。

[0092] 图 13A 显示了实施例 11 所用的装配好的多核苷酸和组分多核苷酸, 及通过装配和宿主细胞的染色体整合而获得的期望的染色体位点。

[0093] 图 13B 显示了由实施例 11 产生的包含染色体整合了的装配好的多核苷酸的酵母细胞转化株的 cPCR 分析结果。

[0094] 图 14 显示了实施例 12 使用的组分多核苷酸, 其用于高通量产生包含染色体整合了的组合的装配好的多核苷酸和组合的组合多核苷酸的酵母细胞, 和通过装配以及宿主细胞介导的同源重组而组合后获得的组合多核苷酸。US = 上游基因组靶向序列, DS = 下游基因组靶向序列, P = 多种启动子序列, G = 多种蛋白质编码序列, URA = 选择性标记的 5' 片段, RA3 = 选择性标记的 3' 片段, PA = 引物结合片段 PmeI-5', PB = 引物结合片段 PmeI-3', LB₀ = 可退火接头序列 RYSE 2, LA_{n1} = 可退火接头序列 RYSE 2, LB_{n1} = 可退火接头序列 RYSE 15, LA_{n2} = 可退火接头序列 RYSE 3, LB_{n2} = 可退火接头序列 RYSE16, LA_{n3} = 可退火接头序列 RYSE 15, LB_{n3} = 可退火接头序列 RYSE 3, LA_{n4} = 可退火接头序列 RYSE 16, LB_{n4} = 可退火接头序列 RYSE 4, LA_{m1} = 可退火接头序列 RYSE 3, LA_{m2} = 可退火接头序列 RYSE 4, LA_{m3} = 可退火接头序列 RYSE 3。

[0095] 5. 发明详述

[0096] 5.1 定义

[0097] 如本文所述, 术语“多核苷酸”是指由核苷酸单位组成的聚合物, 如本领域技术人员所理解的。优选的核苷酸单位包括但不限于包含腺嘌呤 (A), 鸟嘌呤 (G), 胞嘧啶 (C), 胸腺嘧啶 (T), 和尿嘧啶 (U)。有用的修饰的核苷酸单位包括但不限于: 4-乙酰胞苷, 5-(羧基羟基甲基) 尿苷, 2-O-甲基胞苷, 5-羧基甲基氨基甲基-2-硫代尿苷, 5-羧基甲基氨基-甲

基尿苷,二氢尿苷,2-0-甲基假尿苷,2-0-甲基乌苷,肌苷,N6-异戊基腺苷,1-甲基腺苷,1-甲基假尿苷,1-甲基乌苷,1-甲基肌苷,2,2-二甲基乌苷,2-甲基腺苷,2-甲基乌苷,3-甲基胞苷,5-甲基胞苷,N6-甲基腺苷,7-甲基乌苷,5-甲基氨基甲基尿苷,5-甲氧基氨基甲基-2-硫代尿苷,5-甲氧基尿苷,5-甲氧羰基甲基-2-硫代尿苷,5-甲氧羰基甲基尿苷,2-甲基硫代-N6-异戊基腺苷,尿苷-5-氧乙酸-甲酯,尿苷-5-氧乙酸,wybutoxosine,wybutosine,假尿苷,q核苷,2-硫代胞苷,5-甲基-2-硫代尿苷,2-硫代尿苷,4-硫代尿苷,5-甲基尿苷,2-0-甲基-5-甲基尿苷,2-0-甲基尿苷,等等。多核苷酸包括天然存在的核酸例如脱氧核糖核酸(“DNA”)和核糖核酸(“RNA”),以及核酸类似物。核酸类似物包括包含以下的物质:非天然存在的碱基,以与天然存在的二磷酸键不同的方式参与与其它核苷的连接,或包含通过二磷酸键之外的连接而连接的碱基。因此,核酸类似物包含,示例性的和非限制性的,硫代磷酸酯,二硫代磷酸酯,磷酸三酯,氨基磷酸酯,硼烷磷酸酯,甲基磷酸酯,手性-甲基磷酸酯,2-0-甲基核苷酸,肽-核酸(PNA),等等。

[0098] 如本文所述,“组分多核苷酸”是指可以用本文描述的多核苷酸装配方法装配在一起而形成“装配好的多核苷酸”的多核苷酸序列。当多种装配载体被一种或多种限制性内切酶消化而从装配载体中释放组分多核苷酸之后,获得的组分多核苷酸的群体能够包含将被装配进入装配好的多核苷酸的所有DNA片段。

[0099] 如本文所述,“装配好的多核苷酸”是指由本文所述的多核苷酸装配方法所制造的多核苷酸。所述装配好的多核苷酸能够包含两种或更多组分多核苷酸。在一些实施方式里,装配好的多核苷酸包含2,3,4,5,6,7,8,9,10,11,12,13,14,15或更多组分多核苷酸。装配好的多核苷酸长度可以在大约100到大约20000个核苷酸范围内。在一些实施方式里,装配好的多核苷酸的长度在大约200到大约10000,大约200到大约8000,大约200到大约5000,大约200到大约3000,或大约200到大约1000个核苷酸范围内。在其它实施方式里,装配好的多核苷酸的长度在大约200到大约2000,大约2000到大约5000,大约5000到大约10000,大约10000到大约20000,或大于20000个核苷酸的范围内。

[0100] 本文使用常规标注来描述多核苷酸序列:单链多核苷酸序列的左手末端是5'-末端;双链多核苷酸序列的左手方向是指向其5'-方向。

[0101] 如本文所述,术语“DNA片段”,在下文实施例中可选地称为“Bit”,指任何分离的或未分离的DNA分子。有用的实例包括但不限于蛋白质编码序列,报告基因,荧光标记编码序列,启动子,增强子,终止子,内含子,外显子,聚-A尾,多克隆位点,核酸定位信号,mRNA稳定信号,选择性标记,整合位点,表位标签编码序列,降解信号,或任何其它天然存在或人工合成的DNA分子。在一些实施方式里,DNA片段可以是天然来源的。可选地,DNA片段可以是完全的合成来源的,体外合成。进一步,DNA片段可以包含任何分离的天然存在的DNA分子的组合,或任何分离的天然存在的DNA分子和人工合成的DNA分子的组合。例如,DNA片段可以包含异源启动子可操作地连接蛋白质编码序列,蛋白质编码序列连接聚-A尾,蛋白质编码序列在框内连接表位标签编码序列,等等。

[0102] “互补”是指两个多核苷酸的相互作用的表面拓扑学上相容或匹配,如本领域技术人员所理解的。因此,两个序列如果他们能够相互杂交而形成稳定的反平行的、双链核酸结构,则他们相互“互补”。如果第一多核苷酸的核苷酸序列与第二多核苷酸的多核苷酸结合伴侣的核苷酸序列完全相同,或者第一多核苷酸能够在严格杂交条件下与第二多核苷酸杂

交,则该第一多核苷酸与第二多核苷酸互补。因此,序列为 5'-TATAC-3'的多核苷酸与序列为 5'-GTATA-3'的多核苷酸互补。

[0103] “引物”是指这样的多核苷酸序列,它能够特异性地杂交多核苷酸模板序列,例如,引物结合片段,并能够在适合的合成条件下提供合成互补多核苷酸的起始点,即,在核苷酸和能够催化合成反应的试剂(例如 DNA 聚合酶)存在的条件下。引物与多核苷酸模板序列互补,但是不需要与多核苷酸模板序列精确互补。例如,引物可以与多核苷酸模板序列的互补物具有至少 80,85,90,95,96,97,98,或 99%的相同。引物可以具有可变的长度但是通常至少 15 个碱基。在一些实施方式里,引物在 15 和 35 个碱基长度之间。在一些实施方式里,引物为大于 35 个碱基长度。在其它实施方式里,引物具有解链温度 (T_m) 至少 50°C,所述解链温度是指在该温度下 DNA 双链体的一半将分离形成单链。在其它实施方式里,引物具有 50°C 和 70°C 之间的 T_m 。在另外的实施方式里,引物不形成明显的 DNA 或 RNA 二级结构,从而不影响与多核苷酸模板序列的杂交效率。

[0104] 如本文所述,术语“引物结合片段”是这样的多核苷酸序列,它与引物结合以提供在适合合成的条件下合成互补多核苷酸的起始点。在一些实施方式里,引物结合序列是本发明的可退火接头之一。在已知的装配组合物之内的装配载体或组分多核苷酸之中,当不存在互补接头时,序列成为引物结合序列而不是可退火接头。在一些实施方式里,引物结合片段作为基因组靶向序列起作用,例如,上游或下游基因靶向序列。

[0105] 如本文所述,术语“接头序列”和“可退火接头序列”可交替地使用,指包含在本文所述的输入载体和装配载体之内的多核苷酸序列。特别是,可退火接头序列在输入载体或装配载体之内侧邻于 DNA 片段。在从装配载体切除组分多核苷酸和变性该组分多核苷酸之后,可退火接头能够在多核苷酸装配反应中特异性杂交相邻组分多核苷酸的互补可退火接头序列,如本文所述。可退火接头序列,在与互补接头链退火后,能提供合成互补多核苷酸的起始点。

[0106] 如本文所述,术语“载体”在参考文献中用于指细胞内能够复制的染色体外的核酸分子,插入序列能够可操作地连接载体以使插入序列复制。有用的实例包括但不限于环形 DNA 分子例如质粒构建体,噬菌体构建体,黏粒载体,等等,以及线形核酸构建体(例如,λ 噬菌体构建体,细菌人工染色体(BAC),酵母人工染色体(YAC),等等)。载体可以包括表达信号例如启动子和/或终止子,选择性标记例如赋予抗生素抗性的基因,和一个或多个限制性酶切位点,插入序列可以克隆进入其中。载体可以具有其它独特的特性(例如其能够适应的 DNA 插入片段的大小)。

[0107] 如本文所述,术语“输入载体”是指这样的克隆载体质粒,它能够在用于本发明提供的多核苷酸装配方法的装配载体的制备中作为亲本载体。输入载体包含两个可退火接头序列,或一个可退火接头序列和一个引物结合片段,其侧邻于限制性酶切位点,所述位点可用于引入 DNA 片段以形成装配载体。如本文所述,“装配载体”是指已经引入 DNA 片段的输入载体。装配载体在本文描述的多核苷酸装配方法中用来提供组分多核苷酸,以装配进入装配好的多核苷酸。

[0108] 如本文所述,术语“装配载体”指这样的载体,其包含一个可退火接头序列,两个可退火接头序列,或一个可退火接头序列和一个引物结合片段,以及一个 DNA 片段。

[0109] 如本文所述,术语“限制性酶”或“限制性内切酶”是指这样一类催化分子的成员,

其结合同一类 DNA 序列并在该序列中的一个精确位置切断该 DNA 分子。限制性内切酶包括 IIS 型限制性内切酶。这种酶与其它限制性内切酶的不同之处在于其识别序列与切断序列相分离。IIS 型限制性酶的一些实例包括 AlwI, BsaI, BbsI, BbuI, BsmAI, BsrI, BsmI, BspMI, EarI, Esp3I, FokI, HgaI, HphI, LguI, MboII, MnlI, PleI, SapI, SchI, SfaNI, 等等。很多这些限制性内切酶可以商业上获得并为本领域技术人员所熟知。

[0110] 如本文所述, 术语“可退火接头序列双链体”是指一个可退火接头序列链与充分互补的可退火接头序列链以反平行连接方式联合。互补不需要完美; 可退火接头序列双链体可以包含错配碱基对或不配对碱基, 尽管在特别的实施方式里, 可退火接头序列双链体包含两个具有完美互补性的可退火接头序列链。

[0111] 如本文所述, 术语“基因组靶向序列”是指在宿主细胞的基因组上一定位置的核酸序列, 在该位置本发明的多核苷酸通过宿主细胞介导的同源重组而插入其中。术语“上游基因组靶向序列”和“下游基因组靶向序列”是指在宿主细胞基因组中相对位于上游和下游的基因组靶向序列。

[0112] 如本文所述, 术语“染色体靶向序列”是指宿主细胞的染色体上一定位置的核酸序列, 在该位置本发明的多核苷酸通过宿主细胞介导的同源重组而插入其中。术语“上游染色体靶向序列”和“下游染色体靶向序列”是指在宿主细胞染色体中相对位于上游和下游的染色体靶向序列。

[0113] 5.2 多核苷酸装配方法

[0114] 在一方面, 本发明提供了将多种组分多核苷酸有序装配为一种或多种装配好的多核苷酸的快速、牢固和高通量的方法。本发明的方法使用环形核酸载体, 称为装配载体, 其分别包含 DNA 片段 D, 侧邻于可退火接头序列 (即, LA 或 LB), 一对可退火接头序列 (即, LA 和 LB), 或可退火接头序列和引物结合片段 (即, LA 和 PB 或者 LB 和 PA), 和一对限制性酶切位点, RA 和 RB (图 1B)。限制性内切酶在限制性酶切位点 RA 和 RB 消化多种装配载体产生了多种组分多核苷酸, 所述组分多核苷酸包含元件 5'-LA-D-3', 5'-D-LB-3', 5'-LA-D-LB-3', 5'-LA-D-PB-3', 或 5'-LB-D-PA-3' (图 3)。在本发明的方法里, 可退火接头序列 LA 和 LB 提供具有互补末端的组分多核苷酸, 以用于结合重叠延伸装配反应, 继之以聚合酶链式反应 (SOE/PCR) 以使组分多核苷酸以有序顺序装配进入装配好的多核苷酸。

[0115] 特别是, 所述方法能用于将多个功能性 DNA 元件装配进入单一的装配好的多核苷酸, 所述功能性 DNA 元件包括但不限于蛋白质编码序列, 报告基因, 荧光标记编码基因, 启动子, 增强子, 终止子, 内含子, 外显子, 聚-A 尾, 多克隆位点, 核定位信号, mRNA 定位信号, 选择性标记, 整合位点, 表位标签编码序列, 和降解信号。该方法可用于装配任何类型的装配好的多核苷酸, 包括但不限于合成基因, 构建体, 克隆载体, 表达载体, 染色体, 基因组整合构建体, 基因组, 和 DNA 文库。进一步, 所述方法可用于在单一反应中装配 DNA 片段, 而不需要处理和表征中间产物。

[0116] 在一些实施方式里, 所述方法也可用于将多种组分多核苷酸装配为装配好的多核苷酸, 其中所述组分多核苷酸不是来源于装配载体 (即, 通过本领域公知的标准过程获得的 DNA 片段, 例如 PCR 扩增, 化学合成, 等等, 其侧邻于一或两个可退火接头序列, LA 和 / 或 LB, 或可退火接头序列和引物结合片段 (即, LA 和 PB 或者 LB 和 PA))。不来源于装配载体的组分多核苷酸可以在 SOE/PCR 反应或者宿主细胞介导的同源重组之前的任何步骤中

加入装配反应,以装配进入装配好的多核苷酸。因此,在一些实施方式里,该装配方法可以用于装配:(1) 组分多核苷酸,其来自装配载体,该装配载体包含一个或两个可退火接头序列,或一个可退火接头序列和一个引物结合片段,通过消化所述装配载体而产生组分多核苷酸;(2) 非载体 DNA 片段,侧邻于一个或两个可退火接头序列,或一个可退火接头序列和一个引物结合片段,和 (3) 其组合。

[0117] 在一些实施方式里,本发明提供将多种组分多核苷酸装配进入一种或多种装配好的多核苷酸的方法,包含如下步骤:

[0118] (a) 用一种或多种限制性内切酶消化装配组合物以产生组分组合物,所述装配组合物包含:

[0119] (i) 一个或多个第一核酸分子,其中每个第一核酸分子是环形的并以 5' 至 3' 方向包含:第一限制性酶切位点 RA_0 ,选自组 PA 的任意引物结合片段,选自组 D_0 的任意 DNA 片段,可退火接头序列 LB_0 ,和第二限制性酶切位点 RB_0 ;

[0120] (ii) 一个或多个中间核酸分子,其中每个中间核酸分子 n 是环形的并以 5' 至 3' 的方向包含:第一限制性酶切位点 RA_n ,第一可退火接头序列 LA_n ,选自组 D_n 的任意 DNA 片段,第二可退火接头序列 LB_n ,和第二限制性酶切位点 RB_n ,其中 n 表示从 1 到中间核酸分子数目的整数;和

[0121] (iii) 一个或多个最后核酸分子,其中每个最后核酸分子是环形的并以 5' 至 3' 的方向包含:第一限制性酶切位点 RA_m ,第一可退火接头序列 LA_m ,选自组 D_m 的任意 DNA 片段,选自组 PB 的任意引物结合片段,第二限制性酶切位点 RB_m ,其中 m 表示比中间核酸分子数目大 1 的整数;其中切断 RA_0 至 RB_m 的限制性酶切位点,变性所获得的线形核酸分子之后,每个可退火接头序列 $LB_{(p-1)}$ 能够与可退火接头序列 LA_p 的互补物杂交,其中 n 是在 1 到 $(m-1)$ 范围内的整数,其中 p 表示一个从 1 到 m 的整数,并且其中每个组 $D_0, \dots D_n, \dots D_m$ 由一个或多个 DNA 片段组成;

[0122] 其中所述一种或多种限制性内切酶能够切断 RA_0 至 RB_m 的限制性酶切位点;和

[0123] (b) 使组分组合物接触 DNA 聚合酶,三磷酸脱氧核糖核苷酸和一个或多个第一引物和一个或多个第二引物,在适合核酸分子变性的条件下,退火可退火接头序列 $LB_{(p-1)}$ 至可退火接头序列 LA_p ,并从此延伸;其中每个所述第一引物能够与选自组 PA 的所述引物结合片段之一杂交,而每个第二引物能够与选自组 PB 的所述引物结合片段之一杂交;并且将所述组分组合物进行聚合酶链式反应,

[0124] 其中以 5' 到 3' 方向包含分别选自组 $D_0, \dots D_n, \dots$ 和 D_m 的一个 DNA 片段的多核苷酸被装配起来。在所述方法里, p 表示从 1 到 m 的整数。

[0125] 图 3 为示意性性目的描述了本发明的装配方法的一个实施方式。在该实例里,总共四种组分多核苷酸装配产生了装配好的多核苷酸。但是,本发明提供的装配方法可以用于将任意数量的组分多核苷酸装配进入一种或多种装配好的多核苷酸。在一些实施方式里,本发明提供的方法导致 2,3,4,5,6,7,8,9,10,11,12,13,14,15,或更多组分多核苷酸被装配进入一种或多种装配好的多核苷酸。

[0126] 在图 3 示意的实例里,装配组合物包含四个输入装配载体,表示为“第一”,“中间 1(int₁)”,“中间 2(int₂)”,和“最后”,装配好的多核苷酸由所述装配组合物产生。每个装配载体分别包含 DNA 片段侧邻于一个可退火接头序列和一个引物结合片段,或者两个可退火

接头序列。特别地, DNA 片段 D_0 侧邻于 5' 引物结合片段 PA 和 3' 可退火接头序列 LB_0 。DNA 片段 D_1 侧邻于 5' 和 3' 可退火接头序列 LA_1 和 LB_1 , 而 DNA 片段 D_2 侧邻于 5' 和 3' 可退火接头序列 LA_2 和 LB_2 。DNA 片段 D_3 侧邻于 3' 引物结合片段 PB 和 5' 可退火接头序列 LA_3 。装配载体中的元件 5' -PA-D-LB-3', 5' -LA-D-LB-3', 或 5' -LA-D-PB-3' 进一步侧邻于 SapI 限制性酶切位点。

[0127] 在图 3 展示的装配反应的第一步, 装配组合物被用 SapI 消化, 导致组分多核苷酸被从装配载体骨架中切除出来而成为组分组合物, 所述组分多核苷酸包含元件 5' -PA-D-LB-3', 5' LA-D-LB-3', 或 5' -LA-D-PB-3'。由于 SapI 是 IIS 型限制性内切酶, 其识别位点在其切断位点的远端, 而其切断发生在其识别序列之外。这一性质使 IIS 型限制性内切酶尤其适用于根据本发明提供的方法来装配多核苷酸, 因为不包含限制性酶切位点伤痕的多核苷酸可以被装配起来, 这是不同于用非 IIS 型限制性内切酶切断限制性酶切位点 RA 和 RB 的结果。参考图 2, RA_0 , RA_n , 和 RA_m 的 IIS 型识别位点分别在其对应切断位点的 5' 端, 而 RB_0 , RA_n , 和 RA_m 的 IIS 型识别位点分别在其对应切断位点的 3' 端。因此, 限制性酶切位点 RA_0 到 RB_m 是定向的从而其被一种或多种能够切断 RA_0 到 RB_m 的 IIS 型限制性内切酶所切断, 导致了 RA_0 从 D_0 分离, LB_0 从 RB_0 分离, RA_n 从 LA_n 分离, LB_n 从 RB_n 分离, RA_m 从 LA_m 分离, 和 D_m 从 RB_m 分离, 结果产生的线形核酸分子包含 D_0 , LB_0 , RA_n , LB_n , LA_m 或 D_m 而不包含 RA_0 到 RB_m 的任何一个。从而, 获得的组分多核苷酸不包括限制性内切酶的识别或切断位点的任何痕迹。因此本发明的装配多核苷酸的方法可以用于多次转化宿主细胞而不引入重复序列, 后者可能导致遗传不稳定性。

[0128] 随后, 限制性内切酶可选地失活。如果需要失活, 本领域公知的失活核酸内切酶活性的任何方法都可以使用, 包括基于柱或凝胶的纯化方法。一种方便的方法是加热失活, 例如, 65°C 20 分钟, 仅需要很少或不需要在反应试管外操作所述组分组合物。

[0129] 组分多核苷酸之间的互补末端的重叠链形成的序列双链体启动了组分多核苷酸装配进入装配好的多核苷酸。特别是, 可退火接头序列设计为可退火接头序列 LB_0 能杂交可退火接头序列 LA_1 的互补物, 可退火接头序列 LB_1 能杂交可退火接头序列 LA_2 的互补物, 而可退火接头序列 LB_2 能杂交可退火接头序列 LA_3 的互补物。从而, 在装配反应的第二步, 组分多核苷酸用于变性条件 (例如, 加热) 以产生单链组分多核苷酸, 并且伴随着或者在装配反应的变性步骤之后接触耐热 DNA 聚合酶和三磷酸脱氧核糖核苷。

[0130] 耐热 DNA 聚合酶可以是本领域技术人员认为合适的任何耐热 DNA 聚合酶。适合用于本方法的耐热 DNA 聚合酶包括但不限于 *Thermus thermophilus* (Tth) DNA 聚合酶, *Thermus aquaticus* (Taq) DNA 聚合酶, *Thermotoga neopolitana* (Tne) DNA 聚合酶, *Thermotoga maritima* (Tma) DNA 聚合酶, *Thermococcus lit* 和 *alis* (Tli 或 VENT™) DNA 聚合酶, *Pyrrococcus furiosus* (Pfu 或 DEEPVENT™) DNA 聚合酶, *Pyrrococcus woosii* (Pwo) DNA 聚合酶, *Bacillus stearothermophilus* (Bst) DNA 聚合酶, *Sulfolobus acidocaldarius* (SAC) DNA 聚合酶, *Thermoplasma acidophilum* (Tac) DNA 聚合酶, *Thermus flavus* (Tfl/Tub) DNA 聚合酶, *Thermus ruber* (Tru) DNA 聚合酶, *Thermus brockianus* (DYNAZYME™) DNA 聚合酶, *Methanobacterium thermoautotrophicum* (Mth) DNA 聚合酶, 及其突变物, 变体, 和其衍生物。优选具有高保真性 (即, 校对机制) 和低错配率的耐热 DNA 聚合酶。在某些实施方式里, DNA 聚合酶是 Phusion™ DNA 聚合酶 (New Engl 和 Biolabs, Ipswich, MA)。在其它实施

方式里,DNA 聚合酶是 PfuUltra™ II Fusion DNA 聚合酶 (Stratagene/Agilent,La Jolla, CA)。

[0131] 然后将装配反应置于允许从重叠可退火接头序列的 3' 羟基部分开始链延伸的条件下,期间耐热 DNA 聚合酶填充重叠的可退火接头序列之间的部分。装配反应经受一定数量的变性 / 退火 / 延伸的重复循环 (例如,5-15 个循环),期间形成了足量的双链的装配好的多核苷酸。在该循环期间,组分多核苷酸即作为引物也作为模板以产生装配好的多核苷酸的全长模板。在某些实施方式里,PCR 的所述退火和延伸步骤都在 72°C 下进行。

[0132] 与可退火接头序列 LA 和 LB 相比,引物结合片段 PA 和 PB 设计为相互之间或者与任何可退火接头序列或 DNA 片段之间都不重叠,但是作为用于扩增全长的装配好的多核苷酸的引物的结合位点。因此,在装配反应的 4 和 5,加入与引物结合片段 PA 和 PB 互补的引物,该组合物置于传统 PCR 扩增条件下。PCR 扩增条件可以是本领域技术人员认为合适的任何 PCR 扩增条件,包括如《PCR Technology :Principles 和 Applications f 和 DNA Amplification》,ed.HA Erlich, Stockton Press, New Y 和 k, N.Y. (1989) ;《PCR Prot ° C ols :A Guide to methods 和 Applications》,eds. Innis, Gelfl 和, Snisky, 和 White, Academic Press, San Diego, Calif(1990) ;Mattila et al. (1991)Nucleic Acids Res. 19 :4967 ;Eckert, K. A. 和 Kunkel, T. A. (1991)PCR Methods 和 Applications 1 :17 ;和 U. S. 专利号 4, 683, 202 和 4, 965, 188 中所描述的,它们分别引入本文作为参考。在某些实施方式里,装配反应的 PCR 步骤包含大约 35 个循环的变性,退火和延伸,在与引物结合片段 PA 和 PB 互补的引物存在的条件下。在某些实施方式里,PCR 的退火和延伸步骤可以都在 72°C 下进行。但是,本领域技术人员将明白成功扩增的最佳条件依赖于所用的耐热 DNA 聚合酶和可退火接头序列,这些条件从而可以调整。

[0133] 任选地,装配好的多核苷酸可以通过本领域技术人员抑制的任何技术来纯化,例如,凝胶电泳纯化方法,然后用于各种目的。例如,装配好的多核苷酸可以插入表达载体骨架以进行序列验证。

[0134] 5.3 制备包含装配好的多核苷酸的宿主细胞的方法

[0135] 在另一方面,本发明提供制备包含装配好的多核苷酸的宿主细胞的方法。在一些实施方式里,装配好的多核苷酸是至少 3kb 大小。在其它实施方式里,装配好的多核苷酸是至少 5kb 大小。在其它实施方式里,装配好的多核苷酸是至少 6,7,8,9 或 10kb 大小。在其它实施方式里,装配好的多核苷酸为大于 10kb 大小。在其它实施方式里,装配好的多核苷酸为大于 15kb 大小。在其它实施方式里,装配好的多核苷酸为大于 20kb 大小。

[0136] 在一些实施方式里,提供的方法包含用由本文描述的多核苷酸装配方法产生的装配好的多核苷酸转化宿主细胞。装配好的多核苷酸可以再转化之前环形化,也可以以线形分子转化。装配好的多核苷酸可以在宿主细胞内以染色体外多核苷酸来维持。可选地,装配好的多核苷酸可以整合进入宿主细胞的基因组,例如,通过宿主细胞介导的同源重组。为了通过同源重组将装配好的多核苷酸整合进入基因组,装配好的多核苷酸必须在一个末端包含具有上游基因组靶向序列的核酸序列,而在另一个末端包含具有下游基因组靶向序列的核酸序列。从而,要整合进入宿主细胞染色体的装配好的多核苷酸是从这样的装配组合物产生的,所述装配组合物包含包括上游染色体整合序列的第一核酸分子和包括下游染色体整合序列的最后核酸分子,每个染色体靶向序列都足够长度以起始宿主细胞介导的它与

染色体之间的同源重组。

[0137] 在其它实施方式里,所述方法包含用由本文描述的多核苷酸装配方法产生的大量装配好的多核苷酸转化宿主细胞。在一个特别的实施方式里,宿主细胞通过同源重组将两个或更多装配好的多核苷酸组合进入单一的组合多核苷酸。包含组合多核苷酸的宿主细胞转化体由于表达选择性标记而被筛选出来,所述选择性标记在组合装配好的多核苷酸的过程中产生。所述方法对通过同源重组将相对大的多核苷酸片段插入目标多核苷酸尤其有用。为了发生染色体整合,组合多核苷酸必须分别包含上游基因组靶向序列位于选择性标记的编码序列的 5' 或 3' 位置,和下游基因组靶向序列位于选择性标记的编码序列的 3' 或 5' 位置。本文所用的基因组整合包括染色体整合,即,多核苷酸整合进入宿主细胞的染色体。*Saccharomyces cerevisiae* 中合适的染色体整合位点包括但不限于 NDT80, HO, GAL2, 和 GAL1-GAL10-GAL7 位点。该方法也可以用于产生包含染色体外维持的多核苷酸的宿主细胞,例如,质粒和表达载体。当组合多核苷酸不包含相同的可退火接头序列或 DNA 片段排列为正向重复的时候,染色体整合的和染色体外保持的组合多核苷酸的稳定性都会增加,否则这种正向重复会起始额外的同源重组事件并导致组分多核苷酸片段的切除。因此,在一些实施方式里,装配好的多核苷酸包含独特的可退火接头序列和 DNA 片段。在其它实施方式里,装配好的多核苷酸包含一个或多个同样的可退火接头序列或 DNA 片段,但它们在装配好的多核苷酸组合之后在组合多核苷酸中以反向重复来排列。

[0138] 图 8 示意了示例性的组合多核苷酸的制备和通过同源重组将组合的多核苷酸整合进入宿主细胞的染色体中。能够同源重组的宿主细胞吸收了两种装配好的多核苷酸 A 和 B。每种装配好的多核苷酸分别包含 DNA 片段 D_m , 其编码选择性标记的一个片段,其中装配好的多核苷酸 A 的 DNA 片段 D_{m1} 编码选择性标记的第一片段,而装配好的多核苷酸 B 的 DNA 片段 D_{m2} 编码选择性标记的第二片段,其中 DNA 片段 D_{m1} 和 DNA 片段 D_{m2} 包含足够起始宿主细胞介导的同源重组的同源区域,并且其中 DNA 片段 D_{m1} 和 DNA 片段 D_{m2} 均不能产生功能性选择性标记,但是通过宿主细胞介导的同源重组,产生了功能性的选择性标记。每种装配好的多核苷酸分别进一步包含 DNA 片段 D_0 。其编码染色体靶向序列并具有足够长度以起始宿主细胞介导的同源重组,其中装配好的多核苷酸 A 的 DNA 片段 D_{01} 编码上游染色体靶向序列,而装配好的多核苷酸 B 的 DNA 片段 D_{02} 编码下游染色体靶向序列。一旦进入细胞,宿主细胞将在 DNA 片段 D_{m1} 和 D_{m2} 的同源区域组合装配好的多核苷酸 A 和装配好的多核苷酸 B,以形成组合的多核苷酸。此外,宿主细胞使用 DNA 片段 D_{01} 和 D_{02} 编码的染色体靶向序列而使组合的多核苷酸通过同源重组插入其染色体。包含组合的多核苷酸的宿主细胞很容易基于其产生的功能性选择性标记而鉴别出来。

[0139] 在其它实施方式里,所述方法包含用多种组分多核苷酸转化宿主细胞,并允许宿主细胞通过同源重组产生一种或多种装配好的多核苷酸。装配好的多核苷酸可以在宿主细胞内染色体外保持或者整合进入宿主细胞的染色体。图 9 示意了通过宿主细胞中同源重组而产生示例性的装配好的多核苷酸和将装配好的多核苷酸整合进入宿主细胞染色体。第一步,包含装配载体的装配组合物用 IIS 型限制性例如 SapI 或 LguI 内切酶消化导致组分多核苷酸从装配载体骨架中切除出来。在该实施方式里, D_0 和 D_3 可以是上游或下游染色体靶向序列,这种情况下第一装配载体和最后装配载体中具有引物结合片段是可选的。或者,这两个引物结合片段能够作为上游和下游基因组靶向序列起作用。

[0140] 一旦切除出来,每个切除的组分多核苷酸分别包含可退火接头序列 LB,其同源于另一种组分多核苷酸的可退火接头序列 LA,并具有足够长度以起始宿主细胞介导的同源重组。从第一装配载体切下的组分多核苷酸进一步包含上游染色体靶向序列,而从最后装配载体切下的组分多核苷酸进一步包含下游染色体靶向序列,其中两个染色体靶向序列都具有足够长度以在宿主细胞中起始宿主细胞介导的与染色体之间的同源重组。限制性内切酶随后被失活。在所述方法的第二步,组分组合物被引入能够同源重组的宿主细胞。一旦进入细胞,宿主细胞在可退火接头序列之间的同源区域组合所述组分多核苷酸以形成装配好的多核苷酸,而装配好的多核苷酸整合进入染色体。包含装配好的多核苷酸的宿主细胞可以通过基于装配好的多核苷酸的 DNA 片段编码的选择性标记而容易地鉴别出来。

[0141] 任何宿主细胞可以用于本文描述的方法。在特定的实施方式里,合适的宿主细胞是能够基于序列互补性而组合多核苷酸的宿主细胞,所述序列互补性例如本发明提供的选择性标记片段,基因组靶向序列,和可退火接头序列。这种宿主细胞的示意性实例包括但不限于 *Saccharomyces cerevisiae*。适合这些宿主细胞摄入 DNA 的条件是本领域熟知的。

[0142] 包含装配好的多核苷酸或组合多核苷酸的宿主细胞转化体可以由于装配好的多核苷酸或者组合多核苷酸编码的选择性标记的表达而被容易地鉴别出来,所述选择性标记允许通过促进或者抑制细胞的生长来选择。选择性标记可以由装配组合物的装配载体中出现的单独 DNA 片段来编码。可选地,选择性标记的非功能性片段可以由装配组合物的多种装配载体中或者多种装配好的多核苷酸中出现的 DNA 片段来编码,而功能性选择性标记仅分别由装配好的多核苷酸或者组合多核苷酸产生。

[0143] 本领域已知多种选择性标记(例如参见 Kaufman, *Meth. Enzymol.*, 185:487(1990); Kaufman, *Meth. Enzymol.*, 185:537(1990); Srivastava 和 Schlessinger, *Gene*, 103:53(1991); Romanos et al., in *DNA Cloning 2: Expression Systems*, 2nd Edition, pages 123-167 (IRL Press 1995); Markie, *Methods Mol. Biol.*, 54:359(1996); Pfeifer et al., *Gene*, 188:183(1997); Tucker 和 Burke, *Gene*, 199:25(1997); Hashida-Okado et al., *FEBS Letters*, 425:117(1998))。在一些实施方式里,选择性标记是药物抗性标记。药物抗性标记使细胞能够对一种外源药物解毒,否则该药物将杀死细胞。药物抗性标记的示意性实例包括但不限于赋予对抗生素的抗性,例如氨苄青霉素,四环素,卡那霉素,博来霉素,链霉素,潮霉素,新霉素, ZeocinTM, 等等。在其它实施方式里,选择性标记是营养缺陷型标记。营养缺陷型标记允许细胞在缺乏某种必要组分的培养基里生长的时候能够合成必要组分(通常是氨基酸)。选择性营养缺陷型基因序列包括,例如, *hisD*, 其允许在组氨酸存在的情况下在组氨酸缺乏的培养基里生长。其它选择性标记包括博来霉素抗性基因,金属硫蛋白基因,潮霉素 B-磷酸转移酶基因, *AUR1* 基因,腺苷脱氨基酶基因,氨基糖苷磷酸转移酶基因,二氢叶酸还原酶基因,胸苷激酶基因,黄嘌呤-鸟嘌呤磷酸核糖转移酶基因,等等。

[0144] 营养缺陷型也可以用于鉴别包含染色体整合的装配好的多核苷酸或组合多核苷酸的宿主细胞转化株,当装配好的多核苷酸或组合多核苷酸的整合导致中断了宿主细胞需要用来合成细胞生长所必须的组分的基因的时候,使细胞成为营养缺陷突变体。

[0145] 包含染色体整合的装配好的多核苷酸或组合多核苷酸的宿主细胞转化株也可以通过筛选具有其它特性的宿主细胞转化株而鉴别出来,所述特性由单独 DNA 片段或者组合 DNA 片段编码,例如,表达发光的肽,或宿主细胞单克隆的分子分析,例如,通过限制性酶图

谱, PCR 扩增, 或者分离装配好的多核苷酸或染色体整合为点的序列分析。

[0146] 5.4 多核苷酸装配的组合方法和宿主细胞的制备

[0147] 在另一方面, 本发明提供将多种组分多核苷酸有序装配进入多种装配好的多核苷酸的快速, 牢固, 和高通量的方法。该方法依赖于使用包含装配载体的装配组合物, 其中所述装配载体分别包含 DNA 片段 D, 侧邻于可退火接头序列 LA 或 LB, 一对可退火接头序列 LA 和 LB, 或者一个可退火接头序列和一个引物结合片段, 即, LA 和 PB 或者 LB 和 PA, 侧邻于一对限制性酶切位点 RA 和 RB (图 1B)。但是, 为了用本发明公开的方法产生多样的装配好的多核苷酸, 选择可退火接头序列和引物结合片段以使得多于一种的组分多核苷酸的组合能在反应中被装配进入装配好的多核苷酸。因此, 在某些实施方式里, 装配组合物包含至少两种装配载体, 其具有相同可退火接头序列 LA 或 LB 或者相同引物结合片段 PA 或 PB, 但是至于 DNA 片段则不相同。在其它实施方式, 装配组合物包含至少两种装配载体, 其具有同样的可退火接头序列对 LA 和 LB, 或者同样的可退火接头序列可退火接头序列和引物结合片段对, 即, LA 和 PB 或者 LB 和 PA, 至于 DNA 片段则不相同。

[0148] 图 10 描述了在同一个反应中由七种 (7) 组分多核苷酸产生多种装配好的多核苷酸的示例性方法。包含要装配的 DNA 片段的装配载体注入一个单一的试管, 用 SapI 消化以从装配载体骨架中释放组分多核苷酸片段。加热失活 SapI 之后, 组分多核苷酸经历变性条件, 继之以退火条件使足够互补的可退火接头序列对杂交。随后在 DNA 聚合酶和 dNTP 存在下引物延伸, 加入与引物结合片段 PA 和 PB 互补的引物, 以 PCR 扩增八种 (8) 不同的全长的装配好的多核苷酸, 包括在各种可能的组合下装配的 DNA 片段 $D_{01/02}$, $D_{1/2}$, D_3 , 和 $D_{41/42}$ 。单独的装配好的多核苷酸可以从混合的装配好的多核苷酸的组合物中分离出来, 例如, 通过另一个轮 PCR 扩增并使用与 DNA 片段 D_{01} , D_{02} , D_{41} , 和 D_{42} 的区域互补的引物。可选地, 可以使用包含一组引物结合片段 PA 和 / 或 PB 的第一装配载体和最后装配载体并使用仅杂交选择的引物结合片段 PA 和 PB 的亚组的引物, 通过 PCR 扩增而使一系列装配好的多核苷酸分离出来。分离的装配好的多核苷酸可以用于, 例如, 转化宿主细胞以产生包含装配好的多核苷酸的多种宿主细胞。可选择地, 宿主细胞可以用混合的装配好的多核苷酸的组合物直接转化, 而包含每种装配好的多核苷酸的宿主细胞转化株可以被分离出来, 例如, 通过宿主细胞单克隆的分子分析, 或者通过选择包含选择性标记的宿主细胞转化株, 或者展现由单独 DNA 片段或 DNA 片段组合编码的其它特性的细胞转化株。

[0149] 在其它实施方式里, 包含多种通过组合方法装配的多核苷酸的多种宿主细胞通过如下方法产生: 用包含多种装配好的多核苷酸的组合物转化宿主细胞, 其中至少两种装配好的多核苷酸包含选择性标记的非功能性片段, 在此基础上同源重组能产生功能性选择性标记, 然后筛选包含组合的多核苷酸的宿主细胞。图 11 示意了产生多种包含组合的多核苷酸的宿主细胞的组合性尝试。在该实例里, 装配好的多核苷酸 A1 和 A2, 分别包含相同的上游染色体靶向序列和相同的选择性标记的第一部分, 而装配好的多核苷酸 B1 和 B2, 分别包含相同的下游染色体靶向序列和相同的选择性标记的第二部分, 被通过宿主细胞介导的同源重组组合性组合起来, 以产生四种不同的组合多核苷酸, A1/B1, A1/B2, A2/B1, 和 A2/B2, 它们可以被插入染色体以产生四种不同的宿主细胞。

[0150] 在另外的其它实施方式里, 包含通过组合方法装配和组合的多种多核苷酸的多种宿主细胞由以下方法产生: 用包含多种组分多核苷酸的组分组合物转化宿主细胞, 其中至

少两种组分多核苷酸包含选择性标记的非功能性片段,在此基础上同源重组能产生功能性选择性标记,然后筛选包含组合多核苷酸的宿主细胞。

[0151] 5.5 输入载体

[0152] 在另一方面,本发明提供了载体,即,输入载体,它可以用于制备装配载体。在一些实施方式里,输入载体是环形多核苷酸并包含选择性标记,复制起始位点,和中间的 DNA 片段,它侧邻两个限制性酶切位点,后者有利于在本发明提供的方法中亚克隆用于装配的不同的 DNA 片段。输入载体进一步包含一个或两个可退火接头序列,或者一个可退火接头序列和一个引物结合片段,侧邻于限制性酶切位点。输入载体进一步包含额外的一对限制性酶切位点位于 DNA 片段的外侧,例如,在所述一个或两个可退火接头序列或者所述可退火接头序列和引物结合片段的外侧。从而,在一些实施方式里,输入载体是环形多核苷酸并以 5' 至 3' 方向包含:限制性酶切位点 RA,可退火接头序列 LA,限制性酶切位点 RY, DNA 片段 D,限制性酶切位点 RZ,和限制性酶切位点 RB。在其它实施方式里,输入载体是环形多核苷酸并以 5' 至 3' 方向包含:限制性酶切位点 RA,限制性酶切位点 RY, DNA 片段 D,限制性酶切位点 RZ,可退火接头序列 LB,和限制性酶切位点 RB。在其它实施方式里,输入载体是环形多核苷酸并以 5' 至 3' 方向包含:限制性酶切位点 RA,引物结合片段 PA 或者可退火接头序列 LA,限制性酶切位点 RY, DNA 片段 D,限制性酶切位点 RZ,引物结合片段 PB 或者可退火接头序列 LB,和限制性酶切位点 RB。

[0153] 在一些实施方式里,输入载体的 DNA 片段 D 的序列是 lac Z 报告基因。lacZ 报告基因用于通过蓝/白筛选来区分于转化了包含 lacZ 之外的 DNA 片段的载体克隆,例如,在制备本文所描述的装配载体的过程中。

[0154] 在一些实施方式里,输入载体是环形多核苷酸并以 5' 至 3' 方向包含:限制性酶切位点 RA,可退火接头序列 LA,限制性酶切位点 RY, DNA 片段 D,限制性酶切位点 RZ,和限制性酶切位点 RB(即,

[0155] 5' -RA-LA-RY-D-RZ-RB-3')。在一些实施方式里,输入载体是环形多核苷酸并以 5' 至 3' 方向包含:限制性酶切位点 RA,限制性酶切位点 RY, DNA 片段 D,限制性酶切位点 RZ,可退火接头序列 LB,和限制性酶切位点 RB(即,

[0156] 5' -RA-RY-D-RZ-LB-RB-3')。在一些实施方式里,输入载体是环形多核苷酸并以 5' 至 3' 方向包含:限制性酶切位点 RA,可退火接头序列 LA,限制性酶切位点 RY, DNA 片段 D,限制性酶切位点 RZ,可退火接头序列 LB,和限制性酶切位点 RB(即, 5' -RA-LA-RY-D-RZ-LB-RB-3')。在一些实施方式里,输入载体是环形多核苷酸并以 5' 至 3' 方向包含:限制性酶切位点 RA,引物结合片段 PA,限制性酶切位点 RY, DNA 片段 D,限制性酶切位点 RZ,可退火接头序列 LB,和限制性酶切位点 RB(即,

[0157] 5' -RA-PA-RY-D-RZ-LB-RB-3')。在一些实施方式里,输入载体是环形多核苷酸并以 5' 至 3' 方向包含:限制性酶切位点 RA,可退火接头序列 LA,限制性酶切位点 RY, DNA 片段 D,限制性酶切位点 RZ,引物结合片段 PB,和限制性酶切位点 RB(即, 5' -RA-LA-RY-D-RZ-PB-RB-3')。图 1A 提供了一个示例性的输入载体。

[0158] 引物结合片段可以是与用于制造装配好的多核苷酸的任意可退火接头序列不互补的任何核苷酸序列。在一些实施方式里,两个引物结合片段包括限制性内切酶识别和切断位点。在一些实施方式里,引物结合片段仅仅是用于特定的装配反应接头序列。在一

些实施方式里,引物结合片段 PA 的核酸序列选自 SEQ ID NO:24 和 25。在一些实施方式里,引物结合片段 PB 的核酸序列选自 SEQ ID NO:24 和 25。在一些实施方式里,引物结合片段 PA 和引物结合片段 PB 的核酸序列选自 SEQ ID NO:24 和 25。在优选实施方式里,PA 和 PB 在序列上是不完全相同的。

[0159] 在一些实施方式里,可退火接头序列 LA 或 LB 的核酸序列为至少 24 个核苷酸并具有至少 60℃的 T_m 。在一些实施方式里,可退火接头序列 LA 的核酸序列选自 SEQ ID NO:1-23。在一些实施方式里,可退火接头序列 LB 的核酸序列选自 SEQ ID NO:1-23。在一些实施方式里,可退火接头序列 LA 和可退火接头序列 LB 的核酸序列选自 SEQ ID NO:1-23。

[0160] 限制性酶切位点 RY 和 RZ 可用作克隆位点以引入各种 DNA 片段,从而产生装配载体。在一些实施方式里,RY 和 RZ 在序列上不相同。在一些实施方式里,RY 和 RZ 可以被相同的限制性内切酶切断。在一些实施方式里,RY 和 RZ 在序列上相同。在一些实施方式里,限制性酶切位点 RY 和 RZ 可以被产生交错末端的限制性内切酶切断,即,末端具有 5' 或 3' 突出。在其它实施方式里,限制性酶切位点 RY 和 RZ 可以被产生平末端的限制性内切酶切断。

[0161] 尽管限制性酶切位点 RY 和 RZ 可以是本领域已知的任何限制性酶切位点,被 IIS 型限制性内切酶识别的限制性酶切位点尤其有用。IIS 型限制性内切酶具有与其切断区域不同的 DNA 结合区域。因此,它们识别一个特定的序列但是在一定距离之外切断。例如,IIS 型限制性内切酶 *SchI* (也称为 *MlyI*) 结合包含序列 GAGTC 的识别位点,而切断距离识别位点四个 (4) 碱基对处,形成平末端 DNA 分子。IIS 型限制性酶切位点在由输入载体制备装配载体时尤其有用。例如,在这样的亚克隆过程中,其中输入载体的 DNA 片段,例如 *lacZ*,用感兴趣的 DNA 片段代替,则具有 IIS 型限制性内切酶的 *lacZ* 的切除能够导致该限制性酶切位点识别序列的完全移除。从而,在连接所述感兴趣的 DNA 片段和线形化输入载体之后,可退火接头序列或者引物结合片段与新引入的 DNA 片段之间的额外序列是最小的。

[0162] 因此,在一些实施方式里,限制性酶切位点 RY 和 RZ 是本领域已知的任何 IIS 型限制性内切酶能够识别和切断的限制性酶切位点。合适的 IIS 型限制性内切酶包括但不限于下列核酸内切酶及其同裂酶,以字母顺序排列:*Alw26I* (*BsmAI*), *AlwI* (*AcIWI*, *BinI*), *AsuHPI* (*HphI*), *BbvI* (*Bst71I*), *BceFI*, *BstF5I* (*BseGI*, *FokI*), *FauI*, *HgaI*, *SapI* (*LguI*), *MboII*, *PleI*, *SapI*, *SchI* (*MlyI*), *SfaNI*, 和 *TspRI*, *AceIII*, *BbsI* (*BbvII*, *BpiI*, *BpuAI*), *Bce83I*, *BciVI*, *BfiI* (*BmrI*), *BpmI* (*GsuI*), *BsaI* (*Eco31I*), *BseRI*, *BsgI*, *BsmBI* (*Esp3I*), *BsmFI*, *BspMI*, *BsrDI* (*Bse3DI*), *Bsu6I* (*Eam1104I*, *EarI*, *Ksp632I*), *Eco57I*, *FauI*, *MmeI*, *RleAI*, *TaqII*, 和 *Tth111III*。在特别的实施方式里,限制性酶切位点 RY 和 RZ 能够被 *SchI* 限制性内切酶识别和切断。

[0163] 在一些实施方式里,RA 和 RB 在序列上不相同。在一些实施方式里,RA 和 RB 可以被相同的限制性内切酶切断。在一些实施方式里,RA 和 RB 在序列上相同。在一些实施方式里,限制性酶切位点 RA 和 RB 可以被产生交错末端的限制性内切酶切断,即,末端具有 5' 或 3' 突出。在其它实施方式里,限制性酶切位点 RA 和 RB 可以被产生平末端的限制性内切酶切断。

[0164] 尽管限制性酶切位点 RA 和 RB 可以是本领域已知的任何限制性酶切位点,在一种或多种生物的 DNA (例如, cDNA) 内相对不频繁的限制性酶切位点 (即,不常见切断点) 尤其

有用。在一些实施方式里,限制性酶切位点 RA 和 RB 被在人类 DNA 中相对不频繁的限制性酶切位点的限制性内切酶识别和切断。在一些实施方式里,限制性酶切位点 RA 和 RB 被在鼠 DNA 中相对不频繁的限制性酶切位点的限制性内切酶识别和切断。在一些实施方式里,限制性酶切位点 RA 和 RB 被在酵母 DNA 中相对不频繁的限制性酶切位点的限制性内切酶识别和切断,例如,在 *Saccharomyces cerevisiae*, *Pichia pastoris*, *Kluyveromyces lactis*, *Arxula adeninivorans*, 或者 *Hansenula polymorpha* 的 DNA 中。在一些实施方式里,限制性酶切位点 RA 和 RB 被在细菌 DNA 中几乎没有限制性酶切位点的限制性内切酶识别和切断,例如,在 *Escherichia coli* 或者 *Bacillus subtilis* 的 DNA 中。

[0165] 在一些实施方式里,限制性酶切位点 RA 和 RB 被 IIS 型限制性内切酶识别和切断,其识别位点是例如包含 PA/LA-D-PB/LB 的多核苷酸序列的末梢。在一些实施方式里,限制性酶切位点 RA 和 RB 各自独立的可以被选自下列的限制性内切酶切断:MssI, NruI (Bsp68I, MluB2I, Sbo13I, SpoI), SnaBI (BstSNI, Eco105I), SrfI, 和 SwaI (BstRZ246I, BstSWI, MspSWI, SmlI), HpaI, HincII, PshAI, OliI, AluI, Alw26I, BalI, DraI, DpnI, Ec 和 47III, Ec 和 CRI, Ec 和 V, FokI, HaeIII, HincII, MboI, MspAII, NaeI, RsaI, PvuII, ScaI, SmaI, SspI, StuI, XmnI, EcaBC3I, SciI, HincII, DraI, BsaBI, Cac8I, Hpy8I, MlyI, PshAI, SspD51, BfrBI, BsaAI, BsrBI, BtrI, CdiI, CviJI, CviRI, Eco47III, Eco78I, EcoICRI, FnuDII, FspAI, HaeI, LpnI, MlyI, MslI, MstI, NaeI, NlaIV, NruI, NspBII, OliI, PmaCI, PshAI, PsiI, SrfI, StuI, XcaI, XmnI, ZraI, 及其同裂酶。在一个特别的实施方式里,限制性酶切位点 RA 和 RB 被 SapI 或 LguI 限制性内切酶识别并切断。LguI 是 SapI 的同裂酶,具有相同的识别和切断特异性。

[0166] 在一些实施方式里,本发明提供的输入载体也包含一个或多个在载体的复制,保持,或整合(例如,复制起点)中起作用的核酸序列,以及一个或多个选择性标记。复制起点是一种独特的多核苷酸,包含多个短重复序列并可以被多亚基起始结合蛋白所识别,从而在装配 DNA 复制酶在起始位点起着关键性作用。本发明提供的输入载体和装配载体中使用的合适的复制起点包括但不限于 *E. coli* oriC, colE1 质粒 origin, 2 μ 和 ARS(均用于酵母系统), sf1, SV40EBV oriP(用于哺乳动物系统),或者在 pSC101 中发现的这些。选择性标记是载体中的有用元件,因为它们提供促进或抑制细胞生长而筛选成功转化了包含选择性标记的载体并表达该标记的细胞的方法。

[0167] 在一些实施方式里,任何载体均可用于构建本发明提供的输入载体。特别是,本领域已知的载体和商业上可获得的载体(及其变体或衍生物)可以被加工为包括限制性酶切位点 RA,可选的引物结合片段 PA 或可退火接头序列 LA,限制性酶切位点 RY, DNA 片段 D,限制性酶切位点 RZ,可选的引物结合片段 PB 或可退火接头序列 LB,和限制性酶切位点 RB,以用于本发明提供的方法。这种载体可以例如从以下源获得:Vector Laboratories Inc., InVitrogen, Promega, Novagen, NEB, Clontech, Boehringer Mannheim, Pharmacia, EpiCenter, OriGenes Technologies Inc., Stratagene, Perkin Elmer, Pharmingen, Life Technologies, Inc., 和 Research Genetics。尤其吸引人的常规载体种类包括真核和/或原核克隆载体,表达载体,融合载体,双杂交或反向双杂交载体,用于不同宿主的穿梭载体,突变载体(mutagenesis vector),转录载体,接受大的插入片段的载体,等等。感兴趣的其它载体包括病毒源载体(M13载体,噬菌体 λ 载体,腺病毒载体,和反转录病毒载体),高、低

和合适拷贝数目的载体,具有兼容性复制子以用于在单一宿主中组合的载体 (PACYC184 和 pBR322) 和真核附加体 (episomal) 复制载体 (pCDM8)。

[0168] 在一个特别的实施方式里,用于本发明提供的方法的输入载体是 pRYSE 载体,其具有 SEQ ID NO:207 到 221 的核苷酸序列。图 4 提供了 pRYSE 载体的示意图,而 pRYSE 载体的制备在下文实施例 1 中描述。

[0169] 5.6 装配载体

[0170] 在另一方面,本发明提供了载体,即,装配载体,用于将多种组分多核苷酸装配进入一种或多种装配好的多核苷酸。在一些实施方式里,装配载体是环形多核苷酸并包含选择性标记,复制起点,和 DNA 片段,该 DNA 片段侧邻于可退火接头序列、可退火接头序列对或者可退火接头序列 / 引物结合片段对,并侧邻于一对限制性酶切位点。限制性酶切位点有益于在装配反应中将组分多核苷酸从装配载体骨架上切除出来。因此,在一些实施方式里,装配载体是环形多核苷酸并以 5' 至 3' 方向包含:限制性酶切位点 RA,引物结合片段 PA 或可退火接头序列 LA, DNA 片段 D,和限制性酶切位点 RB。在一些实施方式里,装配载体是环形多核苷酸并以 5' 至 3' 方向包含:限制性酶切位点 RA,DNA 片段 D,引物结合片段 PB 或可退火接头序列 LB,和限制性酶切位点 RB。在某些实施方式里,装配载体是环形多核苷酸并以 5' 至 3' 方向包含:限制性酶切位点 RA,引物结合片段 PA 或可退火接头序列 LA, DNA 片段 D,引物结合片段 PB 或可退火接头序列 LB,和限制性酶切位点 RB。

[0171] 在一些实施方式里,装配载体是环形多核苷酸并以 5' 至 3' 方向包含:限制性酶切位点 RA,可退火接头序列 LA,DNA 片段 D,和限制性酶切位点 RB(即,5'-RA-LA-D-RB-3')。在一些实施方式里,装配载体是环形多核苷酸并以 5' 至 3' 方向包含:限制性酶切位点 RA,DNA 片段 D,可退火接头序列 LB,和限制性酶切位点 RB(即,5'-RA-D-LB-RB-3')。在一些实施方式里,装配载体是环形多核苷酸并以 5' 至 3' 方向包含:限制性酶切位点 RA,可退火接头序列 LA,DNA 片段 D,可退火接头序列 LB,和限制性酶切位点 RB(即,5'-RA-LA-D-LB-RB-3')。在一些实施方式里,装配载体是环形多核苷酸并以 5' 至 3' 方向包含:限制性酶切位点 RA,引物结合片段 PA, DNA 片段 D,可退火接头序列 LB,和限制性酶切位点 RB(即,5'-RA-PA-D-LB-RB-3')。在一些实施方式里,装配载体是环形多核苷酸并以 5' 至 3' 方向包含:限制性酶切位点 RA,可退火接头序列 LA,DNA 片段 D,引物结合片段 PB,和限制性酶切位点 RB(即,5'-RA-LA-D-PB-RB-3')。图 1B 和图 2 提供了示例性的装配载体。

[0172] 在一些实施方式里,引物结合片段 PA 的核酸序列选自 SEQ ID NO:24 和 25。在一些实施方式里,引物结合片段 PB 的核酸序列选自 SEQ ID NO:24 和 25。在一些实施方式里,引物结合片段 PA 和引物结合片段 PB 的核酸序列选自 SEQ ID NO:24 和 25。在优选实施方式里,PA 和 PB 在序列上是不完全相同的。

[0173] 在一些实施方式里,可退火接头序列 LA 或 LB 的核酸序列为至少 24 个核苷酸并具有至少 60°C 的 T_m 。在一些实施方式里,可退火接头序列 LA 的核酸序列选自 SEQ ID NO:1-23。在一些实施方式里,可退火接头序列 LB 的核酸序列选自 SEQ ID NO:1-23。在一些实施方式里,可退火接头序列 LA 和可退火接头序列 LB 的核酸序列选自 SEQ ID NO:1-23。

[0174] 在一些实施方式里,RA 和 RB 在序列上不相同。在一些实施方式里,RA 和 RB 可以被相同的限制性内切酶切断。在一些实施方式里,RA 和 RB 在序列上相同。在一些实施方式里,限制性酶切位点 RA 和 RB 可以被产生交错末端的限制性内切酶切断,即,末端具有 5'

或 3' 突出。在其它实施方式里,限制性酶切位点 RA 和 RB 可以被产生平末端的限制性内切酶切断。

[0175] 尽管限制性酶切位点 RA 和 RB 可以是本领域已知的任何限制性酶切位点,在一种或多种生物的 DNA(例如, cDNA) 内相对不频繁出现的限制性酶切位点(即,不常见切断点)尤其有用。在一些实施方式里,限制性酶切位点 RA 和 RB 被在人类 DNA 中相对不频繁出现的限制性酶切位点的限制性内切酶识别和切断。在一些实施方式里,限制性酶切位点 RA 和 RB 被在鼠 DNA 中相对不频繁出现的限制性酶切位点的限制性内切酶识别和切断。在一些实施方式里,限制性酶切位点 RA 和 RB 被在酵母 DNA 中相对不频繁出现的限制性酶切位点的限制性内切酶识别和切断,例如,在 *Saccharomyces cerevisiae*, *Pichia pastoris*, *Kluyveromyces lactis*, *Arxula adeninivorans*, 或者 *Hansenula polymorpha* 的 DNA 中。在一些实施方式里,限制性酶切位点 RA 和 RB 被在细菌 DNA 中几乎不存在的限制性酶切位点的限制性内切酶识别和切断,例如,在 *Escherichia coli* 或者 *Bacillus subtilis* 的 DNA 中。

[0176] 在一些实施方式里,限制性酶切位点 RA 和 RB 被 IIS 型限制性内切酶识别和切断。IIS 型限制性内切酶的示意性实例包括但不限于:MssI, NruI (Bsp68I, MluB2I, Sbo13I, SpoI), SnaBI (BstSNI, Eco105I), SrfI, 和 SwaI (BstRZ246I, BstSWI, MspSWI, SmiI), HpaI, HincII, PshAI, OliI, AluI, Alw26I, BalI, DraI, DpnI, Ec 和 47III, Ec 和 CRI, Ec 和 V, FokI, HaeIII, HincII, MboI, MspAII, NaeI, RsaI, PvuII, ScaI, SmaI, SspI, StuI, XmnI, EcaBC3I, SciI, HincII, DraI, BsaBI, Cac8I, Hpy8I, MlyI, PshAI, SspD5I, BfrBI, BsaAI, BsrBI, BtrI, CdiI, CviJI, CviRI, Eco47III, Eco78I, EcoICRI, FnuDII, FspAI, HaeI, LpnI, MlyI, MslI, MstI, NaeI, NlaIV, NruI, NspBII, OliI, PmaCI, PshAI, PsiI, SrfI, StuI, XcaI, XmnI, ZraI, 及其同裂酶。在一个特别的实施方式里,限制性酶切位点 RA 和 RB 被 SapI 或 LguI 限制性内切酶识别并切断。

[0177] 优选地,装配载体的 DNA 片段不包含能切断所述装配载体中 RA 和 RB 任一限制性酶切位点的限制性内切酶识别和切断的核酸序列。这使得所述 DNA 片段在装配反应的第一阶段能保持完整,期间组分多核苷酸被从装配载体骨架中切除出来。在特别的实施例里,所述 DNA 不含 SapI/LguI 位点,而 RA 和 RB 能被 SapI 或 LguI 切断。在将所述 DNA 片段连接进入输入载体之前或之后,可以通过定点突变(参见 Carter, Biotech. J. 237: 1-7(1986); Zoller 和 Smith, Methods Enzymol. 154: 329-50(1987)), 盒式诱变, 限制性选择诱变(Wells et al., Gene 34: 315-323(1985)), 寡核苷酸借到的(定点)诱变, PCR 诱变, 或其它已知技术可以用于修饰所述 DNA 片段中的任何序列。

[0178] 在一些实施方式里,本发明提供的装配载体也包含一个或多个在载体的复制,保持,或整合(例如,复制起点)中起作用的核酸序列,以及一个或多个选择性标记。复制起点是独特的多核苷酸,包含多个短重复序列并可以被多亚基起始结合蛋白所识别,从而在装配 DNA 复制酶在起始位点起关键性作用。本发明提供的输入载体和装配载体中使用的合适的复制起点包括但不限于 *E. coli* OriC, colE1 质粒 origin, 2 μ 和 ARS(均用于酵母系统), sf1, SV40EBV oriP(用于哺乳动物系统), 或者在 pSC101 中发现的这些。选择性标记是载体中的有用元件,因为它们提供通过促进或抑制细胞生长而筛选成功转化了包含选择性标记的载体并表达该标记的细胞的方法。

[0179] 在一些实施方式里,任何载体均可用于构建本发明提供的装配载体。特别是,本领域已知的载体和商业上可获得的载体(及其变体或衍生物)可以被加工为包含限制性酶切位点 RA,可选的引物结合片段 PA 或可退火接头序列 LA, DNA 片段 D,可选的引物结合片段 PB 或可退火接头序列 LB,和限制性酶切位点 RB,以用于本发明提供的方法。这种载体可以例如从以下来源获得:Vector Laboratories Inc., InVitrogen, Promega, Novagen, NEB, Clontech, Boehringer Mannheim, Pharmacia, EpiCenter, OriGenes Technologies Inc., Stratagene, Perkin Elmer, Pharmingen, Life Technologies, Inc., 和 Research Genetics。尤其吸引人的常规载体种类包括真核和/或原核克隆载体,表达载体,融合载体,双杂交或反向双杂交载体,用于不同宿主的穿梭载体,突变载体(mutagenesis vector),转录载体,接受大的插入片段的载体,等等。感兴趣的其它载体包括病毒源载体(M13 载体,噬菌体 λ 载体,腺病毒载体,和反转录病毒载体),高,低和合适拷贝数目的载体,具有兼容性复制子以用于在单一宿主中组合的载体(PACYC184 和 pBR322)和真核附加体(episomal)复制载体(pCDM8)。

[0180] 装配载体可以由输入载体制备。为了从输入载体制备装配载体,输入载体用一种或多种可以切断 RY 和 RZ 限制性内切酶消化,从而线形化所述载体,使其能接受 DNA 片段。所述 DNA 片段可以通过标准克隆技术连接进 RY 和 RZ 位点以形成本发明的装配载体。例如,所述 DNA 片段可以从以下途径获得:通过本领域已知的标准程序从克隆 DNA(例如, DNA “文库”),通过化学合成,通过 cDNA 克隆,或者通过基因组 DNA 克隆,或其片段,从确定的细胞纯化,或通过 PCR 扩增和克隆。参见,例如,Sambrook et al., Molecular Cloning, A Laboratory Manual, 3d. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (2001); Glover, D. M. (ed.), DNA Cloning: A Practical Approach, 2d. ed., MRL Press, Ltd., Oxford, U. K. (1995)。

[0181] 装配载体也可以从另一个载体来制备,所述载体不包含可退火接头序列,可退火接头序列对,或者侧邻于 DNA 片段插入位点的可退火接头序列/引物结合片段对。为了从这样的载体来制备装配载体,所述载体用一种或多种能在适合插入 DNA 片段的位点切断该载体的限制性内切酶消化,例如在多克隆位点,从而线形化该载体使其能接受 DNA 片段。要插入的 DNA 片段可以通过本领域已知的标准过程来获得,例如,克隆,化学合成,或 PCR 扩增。所述 DNA 片段包含侧邻于可退火接头序列、可退火接头序列对或者可退火接头序列/引物结合片段对的 DNA 片段。从而,在一些实施方式里,所述 DNA 片段以 5' 至 3' 方向包含:可退火接头序列 LA 或者引物结合片段 PA, DNA 片段 D, 和可退火接头序列 LB 或引物结合片段 PB(即, 5' -LA-D-LB-3' 或 5' -PA-D-LB-3' 或 5' -LA-D-PB-3')。在一些实施方式里,所述 DNA 片段以 5' 至 3' 方向包含:DNA 片段 D, 和可退火接头序列 LB 或引物结合片段 PB(即, 5' -D-LB-3' 或 5' -D-PB-3')。在一些实施方式里,所述 DNA 片段以 5' 至 3' 方向包含:可退火接头序列 LA 或引物结合片段 PA, 和 DNA 片段 D(即, 5' -LA-D-3' 或 5' -PA-D-3')。所述 DNA 片段可以进一步包含一对限制性酶切位点,其侧邻于可退火接头序列、可退火接头序列对或者可退火接头序列/引物结合片段对,并且限制性内切酶切断产生的末端与线形化 DNA 的片段将要插入的载体所产生的末端相匹配。可选的,所述 DNA 片段可以在产生时就包含这样的匹配末端而不需要额外的限制性内切酶消化来产生匹配末端。DNA 片段与线形化载体连接产生装配载体之后,用于产生匹配末端的限制性酶切位点被保持以作为装配

载体的限制性酶切位点 RA 和 RB。或者,所述连接可以移除原始限制性酶切位点但是另外的限制性酶切位点可以出现在线形化的载体中,并作为装配载体的限制性酶切位点 RA 和 RB。

[0182] 下文实施例 6 提供了从输入载体(即, pRYSE)或者从另一个载体(即, pMULE 载体)产生装配载体的示例性方法。

[0183] 5.7 可退火接头序列

[0184] 在另一方面,本发明提供了可退火接头序列,其侧邻于位于输入载体和装配载体之内的 DNA 片段。可退火接头序列在装配反应中提供相邻组分多核苷酸之间的序列重叠,从而启动组分多核苷酸装配进入装配好的多核苷酸。从而,在优选的实施方式里,输入载体和装配载体的可退火接头序列 LA 和 LB 被优化以在装配反应期间为互补可退火接头序列提供有效和准确的引发。

[0185] 在一些实施方式里,可退火接头序列的长度足够长以提供与其互补可退火接头序列的足够特异性,同时足够短以在装配反应的退火温度下能容易地退火。在一些实施方式里,可退火接头序列的长度是足够长以允许宿主细胞介导的与互补可退火接头序列之间的同源重组。

[0186] 在一些实施方式里,可退火接头序列为大约 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 或 80 个核苷酸长度。在一些实施方式里,可退火接头序列为至少 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 或 30 个核苷酸长度。在一些实施方式里,可退火接头序列为大于 30, 40, 50, 60, 70, 80, 90, 100, 500, 1000, 5000, 或 10,000 个核苷酸长度。在一些实施方式里,可退火接头为至少 18 个核苷酸长度并且是可以被 3 整除的数字,从而当连接在编码 DNA 片段中时,能便于通读接头的转录物。在特别的实施方式里,可退火接头是 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 或 60 个核苷酸长度。

[0187] 在一些实施方式里,可退火接头序列具有相对高的解链温度(T_m),即,在该温度下退火的可退火接头序列双链体的一半将分离以成为单链。可退火接头的 T_m 可以根据 SantaLucia, PNAS, 95 :1460-1465(1998) 用最接近邻居算法计算得出。相对高的 T_m 可以在装配反应中提供更高特异性的配对。相对高的 T_m 也可以允许 PCR 的退火和延伸步骤的组合,或者减少调整 PCR 的退火和延伸步骤之间的温度所需要的时间的量,从而使用本发明的装配方法具有更高的效率。因此,在一些实施方式里,可退火接头序列双链体具有大约 60°C -80°C 的 T_m 。在一些实施方式里,可退火接头序列双链体具有大约 65°C -75°C 的 T_m 。在一些实施方式里,可退火接头序列双链体具有大于 50°C, 55°C, 60°C, 65°C, 70°C, 75°C, 80°C, 85°C, 或 90°C 的 T_m 。

[0188] 在一些实施方式里,可退火接头序列在本文描述的方法的条件下不通过分子内(即,在相同分子内)作用形成明显的二级结构(例如,发卡,自二聚体),在 DNA 水平或 RNA 水平或 DNA 和 RNA 水平均如此。DNA 出现二级结构会导致装配反应产生很少或没有产生装配好的多核苷酸。RNA 出现二级结构会导致翻译效率降低,当退火接头序列用于将包含启动子和蛋白质编码序列的组分多核苷酸装配为装配好的多核苷酸时,可退火接头序列位于启动子和蛋白质编码序列之间,这时候 RNA 的二级结构尤其具有重要影响。从而,用于本发明的装配方法的可退火接头序列被设计为不形成 RNA 和 / 或 DNA 二级结构。可退火接头序列形成 RNA 或 DNA 二级结构的能力可以使用软件工具来确定,例如, IDT Oligo Analyzer(Integrated DNA Technologies, Coralville, IA), mFold(Zuker 2003 Nucleic

Acids Res. 31(13), 3406-15), 或 RNAfold(Hofacker & Stadler(2006)Bioinformatics 22(10):1172-6)。通常来说,这些工具计算序列从线形转变为折叠状态的吉布斯自由能(ΔG)。 ΔG 越大,则该序列越不会形成二级结构。从而,在一些实施方式中,可退火接头序列被设计为具有从线形转变为折叠状态的大的 ΔG 值。在一些实施方式里,可退火接头序列设计为从线形转变为折叠状态的 ΔG 值等于或大于 *Saccharomyces cerevisiae* 基因组中高表达基因的编码序列的直接上游的 n 个碱基由线形转变为折叠状态的 ΔG 值,其中 n 表示可退火接头序列中的碱基数目的整数。在一些实施方式里,可退火接头序列为 36 个碱基长度并具有从线形转变为折叠状态的 ΔG 值为 -1 或更大。

[0189] 在一些实施方式里,可退火接头序列被设计为避免不希望出现的分子间作用(即,在不同分子间)。从而,在一些实施方式里,可退火接头序列基本上不与包含可退火接头序列的装配载体(例如,载体骨架序列)的其它任何序列退火,和/或不与在本发明提供的多核苷酸装配方法所需要的互补可退火接头序列之外的装配组合物的其它载体退火。在一些实施方式,可退火接头序列基本上不与本发明提供的装配组合物的装配载体中的其它可退火接头序列退火。

[0190] 在一些实施方式里,可退火接头序列具有高 G-C 含量,即,可退火接头序列中鸟嘌呤和胞嘧啶的数量占可退火接头序列中碱基总数的百分比高。具有高 G-C 含量的可退火接头序列通常在本发明的方法中很有用,因为高 G-C 含量通常提供了高 T_m , 后者从而可以在装配反应中提供更高特异性的配对,并且允许 SOE/PCR 的退火和延伸步骤的组合,从而节约时间和程序。在一些实施方式里,可退火接头序列的 G-C 含量在大约 20-80% 之间。在一些实施方式里,可退火接头序列的 G-C 含量在大约 40-60% 之间。在一些实施方式里,可退火接头序列的 G-C 含量为大约 40, 45, 50, 55, 60 或 70%。在特定的实施方式里,可退火接头序列具有大于 70% 的 G-C 含量。SEQ ID NO:1-8 是具有高 G-C 含量且不形成明显二级 DNA 结构,并具有 70°C 或更高 T_m 的可退火接头序列的示意性实例。

[0191] 在一些实施方式里,可退火接头序列具有高 A-T 含量,即,可退火接头序列中腺嘌呤和胸腺嘧啶的数量占可退火接头序列中碱基总数的百分比高。高的 A-T 含量可以减小可退火接头序列形成二级结构的倾向,这在可退火接头序列用于将包含启动子和蛋白质编码序列的组分多核苷酸装配为装配好的多核苷酸的过程中尤其有用,其中可退火接头序列位于所述启动子和蛋白质编码序列之间。在一些实施方式里,可退火接头序列的 A-T 含量在大约 20-80% 之间。在一些实施方式里,可退火接头序列的 A-T 含量在大约 40-60% 之间。在一些实施方式里,可退火接头序列的 A-T 含量为大约 30, 35, 40, 45, 50, 55, 或 60%。在一些实施方式里,可退火接头序列具有大于 30% 的 A-T 含量。在一些实施方式里,可退火接头序列的最 3' 端的 26 个碱基的序列符合下列共有基序:

[0192] 5'-ANNNNNNNANNNAANTANNTTANANA-3', 其中 A 代表腺嘌呤, N 代表任意核苷酸, T 代表胸腺嘧啶。该共有基序在 *Saccharomyces cerevisiae* 基因组中高表达基因的起始密码子的上游 26 个碱基出频繁出现。SEQ ID NO:9-23 是包含该共有基序,具有相对高的 A-T 含量,不形成明显的 RNA 或 DNA 二级结构,并具有 60°C 或更高的 T_m 的可退火接头序列的示意性实例。

[0193] 在一些实施方式里,可退火接头序列包含一个或多个限制性酶切位点。可退火接头序列中含有限制性酶切位点使得从输入或装配载体中切除 DNA 片段的同时能将限制性

酶切位点 RA 和 RB 保持在输入载体或装配载体之中。可退火接头序列的限制性酶切位点有利于正向克隆 DNA 片段进入其它输入或装配载体。这一特征有利于高效构建包含同样的 DNA 片段但是具有不同可退火接头序列对或者引物结合片段 / 可退火接头序列对的装配载体,例如,以产生包含下文所述的不同可退火接头序列的装配载体的文库。这一特征也可以在构建包含所述 DNA 片段的其它装配载体时避免需要再扩增和测序该 DNA 片段。从而,在一些实施方式里,可退火接头序列包含独特的限制性酶切位点。在一些实施方式里,所述限制性酶切位点是 7 碱基对限制性酶切位点,即,它可以被识别 7 碱基对核酸序列的限制性内切酶切断。在一些实施方式里,所述限制性酶切位点是 8 碱基对限制性酶切位点。在特别的实施方式里,可退火接头序列内的限制性酶切位点可以被 MreI, FseI, SbfI, AsiSI, NotI, AscI, 或 BbvCI 识别和切断。

[0194] 在一些实施方式里,可退火接头序列包含使其在连接至编码 DNA 片段后能通读转录物的序列。在一些实施方式里,可退火接头序列允许以 5' 至 3' 和 3' 至 5' 方向通读转录物。在这些实施方式里,可退火接头序列的长度优选地为被三 (3) 整除的数目的核苷酸。

[0195] 在特别的实施方式里,可退火接头序列不包含在 *Escherichia coli* (*E. coli*) 或者 *Saccharomyces cerevisiae* (*S. cerevisiae*) 中的稀有密码子。*E. coli* 或 *S. cerevisiae* 中异源基因的有效表达可以被不常见密码子负面影响,而当稀有密码子被更常见密码子取代时,异源蛋白的表达水平通常会提高。参见,例如,Williams et al., *Nucleic Acids Res.* 16 :10453-10467, 1988 和 Hoog et al., *Gene* 43 :13-21, 1986。从而,包含读通序列的可退火接头序列优选不包含 *E. coli* 或 *S. cerevisiae* 中的稀有密码子,以便于包含可退火接头序列的装配好的多核苷酸编码的蛋白质的有效表达。

[0196] 在一些实施方式里,可退火接头序列系列是即将进入的宿主生物中没有的独特序列。在一些实施方式里,可退火接头序列系列是在 *E. coli* 中没有的独特序列。在其它实施方式里,可退火接头序列系列是在 *S. cerevisiae* 中没有的独特序列。

[0197] 在一些实施方式里,合适的可退火接头序列在测试性装配好的多核苷酸里鉴别。测试性装配好的多核苷酸包含要测试的可退火接头序列和允许可退火接头序列测试的其它元件。例如,为了测试可退火接头是否适合装配包含启动子序列的第一组分多核苷酸和包含要在装配好的多核苷酸中处于所述启动子控制之下的蛋白质编码序列的第二组分多核苷酸,可以从第一组分多核苷酸和第二组分多核苷酸来装配测试性装配好的多核苷酸,其中所述第一组分多核苷酸以 5' 至 3' 方向包含:引物结合片段或可退火接头序列,包含启动子的 DNA 片段,和要测试的可退火接头序列,而所述第二组分多核苷酸以 5' 至 3' 方向包含:要测试的可退火接头序列,编码报告基因(例如,绿色荧光蛋白(GFP))的 DNA 片段,和引物结合片段或可退火接头序列。测试性装配好的多核苷酸可以在体内或体外测试其表达报告基因的效率。可以装配相似的测试性多核苷酸以测试用于装配包含其它元件的 DNA 片段的组分多核苷酸的可退火接头序列的合适性,所述其它元件例如增强子,终止子,聚-A尾,核定位信号, mRNA 稳定化信号,选择性标记,表位标签编码序列,降解信号,等等。测试性装配好的多核苷酸可以包含额外的组分多核苷酸,后者使允许测试例如基因组靶向序列和选择性标记,它们使得可以在体内测试中将测试性装配好的多核苷酸引入宿主细胞并筛选阳性转化株。

[0198] 表 1 展示了根据 SEQ ID NO :1-23 的示例性可退火接头序列的 T_m, 限制性酶切位

点,和通读氨基酸。

[0199]

1. 表 1-可退火接头序列的序列和表征								
可退火接头序列	序列名称	长度 (碱基)	% G-C	% A-T	解链 温度 (T _m)	限制 性酶 位点	通读氨基酸	
							正向	反向
SEQ ID NO: 1	RYSE 1	24	79.2	20.8	72.4			
SEQ ID NO: 2	RYSE 2	24	75.0	25.0	71.4	MreI		
SEQ ID NO: 3	RYSE 3	24	75.0	25.0	73.7	FseI		TAGQA RGD
SEQ ID NO: 4	RYSE 4	24	70.8	29.2	71.5	SbfI	NLQA ASAD	IGARG LQV
SEQ ID NO: 5	RYSE 5	24	70.8	29.2	71.2	AsiSI	NAIAD AAD	IGGVG DRV
SEQ ID NO: 6	RYSE 6	24	70.8	29.2	70.9	NotI	KAAA GEDD	ISLASG RL
SEQ ID NO: 7	RYSE 7	24	70.8	29.2	71.5	AscI	KARH GRRD	
SEQ ID NO: 8	RYSE 8	24	75.0	25.0	70.7	BbvCI		
SEQ ID NO: 9	RYSE 9	36	50.0	50.0	67.4			
SEQ ID NO: 10	RYSE 10	36	52.8	47.2	67.7			
SEQ ID NO: 11	RYSE 11	36	58.3	41.7	69.2			
SEQ ID NO: 12	RYSE 12	36	50.0	50.0	67.4			
SEQ ID NO: 13	RYSE 13	36	58.3	41.7	69.4			
SEQ ID NO: 14	RYSE 14	36	52.8	47.2	67.4			
SEQ ID NO: 15	RYSE 15	36	52.8	47.2	67.8			
SEQ ID NO: 16	RYSE 16	36	52.8	47.2	67.8			
SEQ ID NO: 17	RYSE 17	36	52.8	47.2	68.4			
SEQ ID NO: 18	RYSE 18	36	50.0	50.0	67.8			
SEQ ID NO: 19	RYSE 19	36	52.8	47.2	68.1			
SEQ ID NO: 20	RYSE 20	36	55.6	44.4	68.3			
SEQ ID NO: 21	RYSE 21	36	55.6	44.4	67.9			
SEQ ID NO: 22	RYSE 22	36	52.8	47.2	67.4			
SEQ ID NO: 23	RYSE 23	36	55.6	44.4	68.8			

[0200] 5.8 文库

[0201] 在另一方面,本发明提供包了含多种装配载体的文库。该文库能有利于将多种组分多核苷酸高效装配进一个或多个装配好的多核苷酸并在原核或真核细胞中起作用,从而有利于产生独特的生物,例如,细菌或酵母的重组株,而不需要使用耗费时间的基于限制性内切酶和连接酶的克隆技术。本发明提供的装配方法和组合物有利于在表达构建体中的有效替换或引入功能性 DNA 单位,例如启动子、增强子、复制起点等等,从而能提供宿主生物体内的表达构建体的复制和 / 或表达的有效优化。

[0202] 所述文库可以包含在单一组合物或容器内装配的多种装配载体,例如,包含适合实现本发明提供的装配方法的组合物。可选地,所述文库可以包含不在同一组合物或容器内装配的多种装配载体。在一些实施方式里,所述文库包含至少 3、至少 6、至少 10、至少 20、至少 50、或者多于 50 种装配载体,每种分别包含 DNA 片段。

[0203] 在一些实施方式里,所述文库包含多种装配载体,其中每种装配载体分别以 5' 至 3' 方向包含:第一限制性酶切位点 RA, DNA 片段 D, 可退火接头序列 LB, 和第二限制性酶切位点 RB。在一些实施方式里,所述文库包含多种装配载体,其中每种装配载体分别以 5' 至 3' 方向包含:第一限制性酶切位点 RA, 引物结合片段 PA 或第一可退火接头序列 LA, DNA 片段 D, 和第二限制性酶切位点 RB。在一些实施方式里,所述文库包含多种装配载体,其中每种装配载体分别以 5' 至 3' 方向包含:第一限制性酶切位点 RA, 第一可退火接头序列 LA, DNA 片段 D, 可退火接头序列 LB 或引物结合片段 PB, 和第二限制性酶切位点 RB。在一些实施方式里,所述文库的每种载体中的可退火接头序列对或可退火接头序列 / 引物结合片段对不包含相同序列。在一些实施方式里,每种装配载体中的可退火接头序列 LA 和 / 或 LB 的核酸序列选自 SEQ ID NO:1-23。在一些实施方式里,每种装配载体中的引物结合片段 PA 或 PB 的核酸序列选自 SEQ ID NO:24 和 25。

[0204] 在一些实施方式里,所述文库包含下列载体的每种的至少一种:

[0205] (a) 由环形多核苷酸组成的载体,该多核苷酸以 5' 到 3' 方向包含:限制性酶切位点 RA, DNA 片段 D, 可退火接头序列 LB, 和限制性酶切位点 RB;

[0206] (b) 由环形多核苷酸组成的载体,该多核苷酸以 5' 到 3' 方向包含:限制性酶切位点 RA, 可退火接头序列 LA, DNA 片段 D, 可退火接头序列 LB, 和限制性酶切位点 RB; 和

[0207] (c) 由环形多核苷酸组成的载体,该多核苷酸以 5' 到 3' 方向包含:限制性酶切位点 RA, 可退火接头序列 LA, DNA 片段 D, 和限制性酶切位点 RB₀。

[0208] 在一些实施方式里,所述文库包含下列载体的每种的至少一种:

[0209] (a) 由环形多核苷酸组成的载体,该多核苷酸以 5' 到 3' 方向包含:限制性酶切位点 RA, 引物结合片段 PA, DNA 片段 D, 可退火接头序列 LB, 和限制性酶切位点 RB;

[0210] (b) 由环形多核苷酸组成的载体,该多核苷酸以 5' 到 3' 方向包含:限制性酶切位点 RA, 可退火接头序列 LA, DNA 片段 D, 可退火接头序列 LB, 和限制性酶切位点 RB; 和

[0211] (c) 由环形多核苷酸组成的载体,该多核苷酸以 5' 到 3' 方向包含:限制性酶切位点 RA, 可退火接头序列 LA, DNA 片段 D, 引物结合片段 PB, 和限制性酶切位点 RB₀。

[0212] 在一些实施方式里,引物结合片段 PA 的核酸序列选自 SEQ ID NO:24 和 25。在一些实施方式里,引物结合片段 PB 的核酸序列选自 SEQ ID NO:24 和 25。在一些实施方式里,引物结合片段 PA 和引物结合片段 PB 的核酸序列选自 SEQ ID NO:24 和 25。

[0213] 在一些实施方式里,所述文库的任何可退火接头序列 LA 和任何可退火接头序列 LB 的核酸序列选自 SEQ ID NO:1-23。在一些实施方式里,所述文库的至少一个可退火接头序列 LA 和至少一个可退火接头序列 LB 的核酸序列选自 SEQ ID NO:1-23。在一些实施方式里,所述文库的每个可退火接头序列 LA 和可退火接头序列 LB 的核酸序列选自 SEQ ID NO:1-23。

[0214] 在一些实施方式里, DNA 片段 D 包含选自下列的核酸序列:选择性标记,启动子,基因组靶向序列,编码表位标签的核酸序列,编码感兴趣的基因的核酸序列,编码终止密码子

的核酸序列,和 lacZ。

[0215] 在一些实施方式里,所述文库包含下列核酸分子每种的至少一种:

[0216] (a) 第一核酸分子,其中所述第一核酸分子为环形并以 5' 至 3' 的方向包含:第一限制性酶切位点 RA_0 ,选自组 D_0 的任何 DNA 片段,可退火接头序列 LB_0 ,和第二限制性酶切位点 RB_0 ;

[0217] (b) 中间核酸分子,其中所述中间核酸分子 n 为环形并以 5' 至 3' 的方向包含:第一限制性酶切位点 RA_n ,第一可退火接头序列 LA_n ,选自组 D_n 的任何 DNA 片段,第二可退火接头序列 LB_n ,和第二限制性酶切位点 RB_n ,其中 n 代表从 1 到中间核酸分子数目的整数;和

[0218] (c) 最后核酸分子,其中每个最后核酸分子为环形的并以 5' 至 3' 的方向包含:第一限制性酶切位点 RA_m 、第一可退火接头序列 LA_m 、选自组 D_m 的任何 DNA 片段、第二限制性酶切位点 RB_m ,其中 m 表示比中间核酸分子数目大 1 的整数;

[0219] 其中切断 RA_0 至 RB_m 的限制性酶切位点,变性所获得的线形核酸分子之后,每个可退火接头序列 $LB_{(p-1)}$ 能够与可退火接头序列 LA_p 的互补物杂交,其中 n 是在 1 到 $(m-1)$ 范围内的整数,其中 p 表示一个从 1 到 m 的整数,并且其中每个组 $D_0, \dots D_n, \dots D_m$ 独立地由一个或多个 DNA 片段组成。在一些实施方式里,第一核酸分子进一步包含引物结合片段 PA 位于选自 D_0 的 DNA 片段的 5' 位置。在一些实施方式里,最后核酸分子进一步包含引物结合片段 PB 位于选自 D_m 的 DNA 片段的 3' 位置。

[0220] 在一些实施方式里,切断 RA_0 至 RB_m 的限制性酶切位点,变性所获得的线形核酸分子之后,与组分组合物中的其它可退火接头序列和其互补物相比,每个可退火接头序列 $LB_{(p-1)}$ 能够选择性地与可退火接头序列 LA_p 的互补物杂交。在一些实施方式里,每个可退火接头序列 $LB_{(p-1)}$ 在序列上与可退火接头序列 LA_p 相同。

[0221] 在一个特别的实施方式里, RA_0 到 RB_m 的限制性酶切位点可以被相同的限制性内切酶切断,以利于从装配载体中切除组分多核苷酸。在一些实施方式里, RA_0 到 RB_m 的限制性酶切位点可以被 SapI 和 LguI 限制性内切酶切断。

[0222] 在一些实施方式里,引物结合片段 PA 的核酸序列选自 SEQ ID NO:24 和 25。在一些实施方式里,引物结合片段 PB 的核酸序列选自 SEQ ID NO:24 和 25。在一些实施方式里,引物结合片段 PA 和引物结合片段 PB 的核酸序列选自 SEQ ID NO:24 和 25。在优选实施方式里,引物结合片段 PA 和引物结合片段 PB 的核酸序列不相同。

[0223] 在一些实施方式里,所述文库的任何可退火接头序列 LA 和任何可退火接头序列 LB 的核酸序列选自 SEQ ID NO:1-23。在一些实施方式里,所述文库的至少一个可退火接头序列 LA 和至少一个可退火接头序列 LB 的核酸序列选自 SEQ ID NO:1-23。在一些实施方式里,所述文库的每个可退火接头序列 LA 和可退火接头序列 LB 的核酸序列选自 SEQ ID NO:1-23。在一些实施方式里,所述组合物中的每个可退火接头序列 LA 的核酸序列是彼此不同的。在一些实施方式里,所述组合物中的每个可退火接头序列 LB 的核酸序列是彼此不同的。

[0224] 在一个特别的实施方式里,所述文库包含下列核酸分子:

[0225] (a) 两种第一核酸分子,其中一种第一核酸分子以 5' 至 3' 方向包含:第一限制性酶切位点 RA_0 ,引物结合片段 PA, DNA 片段 D_{01} ,可退火接头序列 LB_0 ,和第二限制性酶切位点 RB_0 ,其中另一种第一核酸分子以 5' 至 3' 方向包含:第一限制性酶切位点 RA_0 ,引物结合片

段 PA, DNA 片段 D_{02} , 可退火接头序列 LB_0 , 和第二限制性酶切位点 RB_0 , 其中 DNA 片段 D_{01} 编码第一基因组靶向序列, 其中 DNA 片段 D_{02} 编码在靶向基因组中位于第一基因靶向序列下游的第二基因组靶向序列, 其中 DNA 片段 D_{02} 和 DNA 片段 D_{01} 位于相对引物结合片段 PA 和可退火接头序列 LB_0 相反的方向;

[0226] (b) 至少一种中间核酸分子, 以 5' 至 3' 方向包含: 第一限制性酶切位点 RA_n , 第一可退火接头序列 LA_n , DNA 片段 D_n , 第二可退火接头序列 LB_n , 和第二限制性酶切位点 RB_n , 其中 n 表示从 1 到中间核酸分子数目的整数; 和

[0227] (c) 两种最后核酸分子, 其中一种最后核酸分子以 5' 至 3' 方向包含: 第一限制性酶切位点 RA_m , 可退火接头序列 LA_m , DNA 片段 D_{m1} , 引物结合片段 PB, 和第二限制性酶切位点 RB_m , 其中另一种最后核酸分子以 5' 至 3' 方向包含: 第一限制性酶切位点 RA_m , 可退火接头序列 LA_m , DNA 片段 D_{m2} , 引物结合片段 PB, 和第二限制性酶切位点 RB_m , 其中 m 表示比中间核酸分子数目大 1 的整数, 其中 DNA 片段 D_{m1} 编码选择性标记的第一片段, 其中 DNA 片段 D_{m2} 编码选择性标记的第二片段, 其中 DNA 片段 D_{m2} DNA 片段 D_{m1} 位于相对于可退火接头序列 LA_m 和引物结合片段 PB 相反的方向, 其中 DNA 片段 D_{m1} 和 DNA 片段 D_{m2} 均产生非功能性选择性标记, 但是在经过同源重组之后 DNA 片段 D_{m1} 和 D_{m2} 产生功能性标记;

[0228] 其中每个可退火接头序列 $LB_{(p-1)}$ 分别与可退火接头序列 LA_p 相同, 其中 p 表示从 1 到 m 的整数。

[0229] 在一些实施方式里, 所述文库包含多种装配载体, 其中每种装配载体包含相同的可退火接头序列, 可退火接头序列对或可退火接头序列 / 引物结合片段对, 但是在序列上与其对应的 DNA 片段 D 不同。

[0230] 在其它实施方式里, 所述文库包含多种装配载体, 其中每种装配载体包含相同的 DNA 片段 D, 其侧邻于独特的可退火接头序列, 可退火接头序列对或可退火接头序列 / 引物结合片段对。这种文库可以有利于相对于装配进入装配好的多核苷酸的其它 DNA 片段, 将 DNA 片段 D 快速装配进入特定的位置或方向。

[0231] 在一些实施方式里, 所述文库的成员包含具有共有结构或功能性特征的 DNA 片段。例如, 文库可以包含含有同样的功能性 DNA 单位的多种装配载体。示例性功能 DNA 单位包括但不限于蛋白质编码序列, 报告基因, 荧光标记, 启动子, 增强子, 终止子, 内含子, 外显子, 聚 -A 尾, 多克隆位点, 核定位信号, 核输出信号, mRNA 稳定化信号, 选择性标记, 整合位点, 表位标签, 和降解信号。在一些实施方式里, 所述文库包含多种装配载体, 其中每种装配载体包含相同的启动子。所述装配载体可以包含本领域已知的任何原核或真核启动子序列。示例性的真核启动子包括但不限于金属硫蛋白启动子, 组成型腺病毒主要晚期启动子 (constitutive adenovirus major late promoter), 地塞米松诱导的 MMTV 启动子, SV40 启动子, MRPp01III 启动子, 组成型 MPSV 启动子, RSV 启动子, 四环素诱导的 CMV 启动子 (例如人即刻早期 CMV 启动子), 和组成型 CMV 启动子。在特别的实施方式里, 装配载体包含酵母启动子序列。示例性的酵母启动子包括但不限于 PGAL3, PGAL7, PCTR3, PMET3, PPGK1, PTDH1, PTDH3, PFBA1, PTEF1, PEN01, PEN02, PCYC1, PTDH2, PCUP1, PGAL80, PGAL2, PBNA6, PTMA29, PSBP1, PPUP3, PACS2, PTP01, PRPT1, PAAT2, PAHP1, PSSE1, PTEF2, PNPL3, PPET9, PTUB2, POLE1, PCPR1, PIPPP1, 和 PSOD1。

[0232] 在一些实施方式里, 所述文库包含多种装配载体, 其中每种装配载体包含相同的

终止子序列。所述装配载体可以包含本领域已知的任何原核或真核终止子序列。在特别的实施方式里,装配载体包含酵母终止子序列。示例性的酵母终止子包括但不限于 TADH1, TEN01, TEN02, TCYC1, TNDT80, TTDH3, TTDH1, 和 TPGK1。

[0233] 在一些实施方式里,所述文库包含多种装配载体,其中每种装配载体包含相同的选择性标记序列。所述装配载体可以包含本领域已知的任何原核或真核选择性标记。示例性的选择性标记包括但不限于抗生素抗性标记(例如,卡那霉素、氨苄青霉素、氯霉素、庆大霉素或三甲氧苄二氢嘧啶的抗性编码基因)和代谢标记(例如,氨基酸合成基因或转移 RNA 基因)。

[0234] 5.9 试剂盒

[0235] 在另一方面,本发明提供了用于装配多核苷酸的试剂盒,所述试剂盒包含下列中的两个或更多:(a) 一种或多种本文所述的输入载体;(b) 能切断所述一种或多种输入载体的限制性酶切位点 RA 和 RB 的一种或多种限制性内切酶;(c) 能切断所述一种或多种输入载体的限制性酶切位点 RY 和 RZ 的一种或多种限制性内切酶;和(d) 能与所述一种或多种输入载体的引物结合片段退火的寡核苷酸引物。

[0236] 在一些实施方式里,所述试剂盒的每种输入载体的限制性酶切位点 RA 和 RB 被 SapI 限制性内切酶识别和切断,而所述试剂盒包含 SapI 限制性内切酶。在一些实施方式里,所述试剂盒的每种输入载体的限制性酶切位点 RY 和 RZ 被 SchI(或 MlyI)限制性内切酶识别和切断,而所述试剂盒包含 SchI(或 MlyI)限制性内切酶。

[0237] 在一些实施方式里,所述试剂盒中的一种或多种输入载体的引物结合片段 PA 的核酸序列选自 SEQ ID NO:24 和 25。在一些实施方式里,所述试剂盒中的一种或多种输入载体的引物结合片段 PB 的核酸序列选自 SEQ ID NO:24 和 25。在优选实施方式里,引物结合片段 PA 和引物结合片段 PB 的核酸序列不相同。

[0238] 在一些实施方式里,所述试剂盒中的一种或多种输入载体的可退火接头序列 LA 的核酸序列选自 SEQ ID NO:1-23。在一些实施方式里,所述试剂盒中的一种或多种输入载体的可退火接头序列 LB 的核酸序列选自 SEQ ID NO:1-23。在一些实施方式里,所述试剂盒中的所有输入载体的可退火接头序列 LA 和可退火接头序列 LB 的核酸序列选自 SEQ ID NO:1-23。

[0239] 在一些实施方式里,所述试剂盒包含 pRYSE 载体 #1,其序列在本文中提供为 SEQ ID NO:221。在一些实施方式里,所述试剂盒包含 pRYSE 载体 #2,其序列在本文中提供为 SEQ ID NO:207。在一些实施方式里,所述试剂盒包含 pRYSE 载体 #3,其序列在本文中提供为 SEQ ID NO:208。在一些实施方式里,所述试剂盒包含 pRYSE 载体 #4,其序列在本文中提供为 SEQ ID NO:209。在一些实施方式里,所述试剂盒包含 pRYSE 载体 #5,其序列在本文中提供为 SEQ ID NO:210。在一些实施方式里,所述试剂盒包含 pRYSE 载体 #6,其序列在本文中提供为 SEQ ID NO:211。在一些实施方式里,所述试剂盒包含 pRYSE 载体 #7,其序列在本文中提供为 SEQ ID NO:212。在一些实施方式里,所述试剂盒包含 pRYSE 载体 #8,其序列在本文中提供为 SEQ ID NO:213。在一些实施方式里,所述试剂盒包含 pRYSE 载体 #9,其序列在本文中提供为 SEQ ID NO:214。在一些实施方式里,所述试剂盒包含 pRYSE 载体 #10,其序列在本文中提供为 SEQ ID NO:215。在一些实施方式里,所述试剂盒包含 pRYSE 载体 #11,其序列在本文中提供为 SEQ ID NO:216。在一些实施方式里,所述试剂盒包含 pRYSE 载体 #12,

其序列在本文中提供为 SEQ ID NO:217。在一些实施方式里,所述试剂盒包含 pRYSE 载体 #13,其序列在本文中提供为 SEQ ID NO:218。在一些实施方式里,所述试剂盒包含 pRYSE 载体 #14,其序列在本文中提供为 SEQ ID NO:219。在一些实施方式里,所述试剂盒包含 pRYSE 载体 #15,其序列在本文中提供为 SEQ ID NO:220。

[0240] 在一些实施方式里,所述试剂盒进一步包含使用本文所描述的多核苷酸装配方法的说明书。在一些实施方式里,多核苷酸聚合酶,例如耐热 DNA 聚合酶(例如,Pfu DNA 聚合酶),和三磷酸脱氧核糖核苷(dNTPs)也包含在所述试剂盒里。在一些实施方式里,分别包含要装配进入装配好的多核苷酸的组分多核苷酸的两种或更多装配载体可以由所述试剂盒提供。例如,可以提供装配载体可以包含用于标定和/或作为阳性对照使用的组分多核苷酸,以鉴别所述试剂盒的正确操作。其它实例包括但不限于包含如下组分多核苷酸的装配载体:蛋白质编码序列,报告基因,荧光标记编码序列,启动子,增强子,终止子,外显子,聚-A 尾,多克隆位点,核定位信号,mRNA 稳定化信号,选择性标记,整合位点,表位标签编码序列,和降解信号。

6. 实施例

[0241] 本发明由下列实施例来说明,但并不意味着以任何方式限定。实施例中描述的 *Saccharomyces cerevisiae* 构建体源于 *Saccharomyces cerevisiae* 株 CEN.PK2。与 *Saccharomyces cerevisiae* 株 S288c 不同,菌株 CEN.PK2 的基因组序列是不能公开获得的。所描述的有些构建体是序列证实的,因此所提供的序列是真实的源自 CEN.PK2 的构建体。对于未序列证实的构建体,所提供的序列是基于菌株 S288c 公开的基因组序列,因此可能包含与真实的源自 CEN.PK2 的构建体的序列的多态性差异。

[0242] 实施例 1

[0243] 本实施例描述制备 pRYSE 载体的方法。pRYSE 载体以 5'至 3'方向包含:第一 SapI 限制性内切酶识别位点,第一可退火接头序列或引物结合片段,第一 SchI 限制性内切酶识别位点,绿色荧光蛋白(GFP)或 lacZ 标记基因,第二 SchI 限制性内切酶识别位点,第二可退火接头序列或引物结合片段,和第二 SapI 限制性内切酶识别位点。

[0244] 用 PCR 引物 JCB158-17A(SEQ ID NO:227)和 JCB158-17B(SEQ ID NO:228)以定点诱变的方式将 pUC19 的 bla 基因中的 SchI 限制性内切酶识别位点从 pUC19 移除之后,从 pUC19 载体(GenBank accession L09137)以引物 JCB158-17C(SEQ ID NO:229)和 JCB158-17D(SEQ ID NO:230)PCR 扩增编码 β -乳糖酶的 DNA 片段。PCR 产物通过凝胶纯化,然后连接至 TOPO 载体(Invitrogen,Carlsbad,CA),通过限制性内切酶 SphI 和 MfeI 完全消化该构建体之后所述 PCR 产物被重新释放出来,产生“bla DNA 片段”。

[0245] DNA 片段 1040(SEQ ID NO:224)、1041(SEQ ID NO:225)和 1042(SEQ ID NO:226)通过人工合成(Biosearch Technologies,Novato,CA)。DNA 片段 1040 和 1041 用限制性内切酶 BstXI 完全消化,每个消化的片段分别连接至 2.65kb 的载体骨架,所述载体骨架通过用限制性内切酶 SacI 和 KpnI 完全切断 pAM1466(SEQ ID NO:223;由 Biosearch Technologies,Novato,CA 人工合成)而产生。1040_pAM1466 DNA 构建体用限制性内切酶 BsmBI 和 BstXI 完全消化,反应混合物用凝胶电泳拆分,含有 1040DNA 片段的大约 3.5kb 的 DNA 片段通过凝胶纯化。1041_pAM1466 DNA 构建体用限制性内切酶 BsaI 和 BstXI 完全消化,

反应混合物用凝胶电泳拆分,含有 1041DNA 片段的大约 0.9kb 的 DNA 片段通过凝胶纯化。连接纯化得到的 DNA 片段,产生 DNA 构建体 1040_1041_pAM1466。使用引物 J036(SEQ ID NO:69) 和 J037(SEQ ID NO:70) 来产生 1040_1041 DNA 片段,使用引物 J038(SEQ ID NO:71) 和 J039(SEQ ID NO:72) 来产生具有与所述 1040_1041DNA 片段的末端序列相重叠的末端片段的 1042DNA 片段,使用引物 J039(含有 SphI 限制性内切酶识别位点)(SEQ ID NO:72) 和 J036(包含 MfeI 限制性内切酶识别位点)(SEQ ID NO:69) 来连接这两个 PCR 产物,通过 PCR “缝合”反应而将 DNA 片段 1042 连接到 DNA 构建体 1040_1041 上。1040_1041_1042PCR 产物用 SphI 和 MfeI 限制性内切酶完全消化,反应混合物用凝胶电泳拆分,大约 2.4kb 的 1040_1041_1042 DNA 片段用凝胶纯化,纯化的 DNA 片段连接凝胶纯化的 bla 片段,产生“1040_1041_1402_bla”DNA 构建体。

[0246] 编码 GFP 基因的 1040_1041_1042_bla DNA 构建体的片段用 PCR 引物 1 和 2(见表 2)PCR 扩增。对于扩增的 GFP 片段,通过使用在第一轮 PCR 反应中产生的凝胶提取的 GFP 片段作为模板、使用 PCR 引物 3 和 4(见表 2)PCR 扩增从而将末端 SacI 和 XhoI 限制性内切酶识别位点加入所述 GFP 片段。扩增的 PCR 产物用凝胶提取,用 XhoI 和 SacI 限制性内切酶完全消化,所述限制性内切酶加热至 65°C 20 分钟使其失活,消化后的 PCR 产物用柱纯化,然后与 1040_1041_1042_bla DNA 片段经 XhoI 和 SacI 限制性内切酶完全消化后通过凝胶纯化的获得的大约 2.2kb 的 DNA 片段相连接。获得的载体用 PCR 引物 5 和 6(见表 3)PCR 扩增,反应混合物用凝胶电泳拆分,用凝胶纯化大约 2.2kb 的“pRYSE 载体骨架”。

[0247]

表 2— 用于制备 GFP 插入片段的 PCR 引物, 所述 GFP 片段侧邻于可退火接头对或可退火接头/引物结合片段对和 SacI 和 XhoI 限制性酶切位点。

GFP 片段	可退火接头 或引物结合 片段 1	可退火接头 或引物结合 片段 2	引物 1	引物 2	引物 3	引物 4
1	Pme1-5'	RYSE 1	J018 (SEQ ID NO: 73)	J073 (SEQ ID NO: 106)	J055 (SEQ ID NO: 88)	J064 (SEQ ID NO: 97)
2	RYSE 1	RYSE 2	J019 (SEQ ID NO: 74)	J074 (SEQ ID NO: 107)	J056 (SEQ ID NO: 89)	J065 (SEQ ID NO: 98)
3	RYSE 2	RYSE 3	J020 (SEQ ID NO: 75)	J029 (SEQ ID NO: 82)	J057 (SEQ ID NO: 90)	J066 (SEQ ID NO: 99)
4	RYSE 3	RYSE 4	J021 (SEQ ID NO: 76)	J030 (SEQ ID NO: 83)	J058 (SEQ ID NO: 91)	J067 (SEQ ID NO: 100)
5	RYSE 4	RYSE 5	J022 (SEQ ID NO: 77)	J031 (SEQ ID NO: 84)	J059 (SEQ ID NO: 92)	J068 (SEQ ID NO: 101)
6	RYSE 5	RYSE 6	J023 (SEQ ID NO: 78)	J032 (SEQ ID NO: 85)	J060 (SEQ ID NO: 93)	J069 (SEQ ID NO: 102)
7	RYSE 6	RYSE 7	J024 (SEQ ID NO: 79)	J033 (SEQ ID NO: 86)	J061 (SEQ ID NO: 94)	J070 (SEQ ID NO: 103)
8	RYSE 7	RYSE 8	J025 (SEQ ID NO: 80)	J034 (SEQ ID NO: 87)	J062 (SEQ ID NO: 95)	J071 (SEQ ID NO: 104)
9	RYSE 2	Pme1-3'	J020 (SEQ ID NO: 75)	J075 (SEQ ID NO: 108)	J057 (SEQ ID NO: 90)	J072 (SEQ ID NO: 105)
10	RYSE 3	Pme1-3'	J021 (SEQ ID NO: 76)	J075 (SEQ ID NO: 108)	J058 (SEQ ID NO: 91)	J072 (SEQ ID NO: 105)
11	RYSE 4	Pme1-3'	J022 (SEQ ID NO: 77)	J075 (SEQ ID NO: 108)	J059 (SEQ ID NO: 92)	J072 (SEQ ID NO: 105)
12	RYSE 5	Pme1-3'	J023 (SEQ ID NO: 78)	J075 (SEQ ID NO: 108)	J060 (SEQ ID NO: 93)	J072 (SEQ ID NO: 105)
13	RYSE 6	Pme1-3'	J024 (SEQ ID NO: 79)	J075 (SEQ ID NO: 108)	J061 (SEQ ID NO: 94)	J072 (SEQ ID NO: 105)

[0248]

14	RYSE 7	Pme1-3'	J025 (SEQ ID NO: 80)	J075 (SEQ ID NO: 108)	J062 (SEQ ID NO: 95)	J072 (SEQ ID NO: 105)
15	RYSE 8	Pme1-3'	J026 (SEQ ID NO: 81)	J075 (SEQ ID NO: 108)	J063 (SEQ ID NO: 96)	J072 (SEQ ID NO: 105)

[0249]

表 3—pRYSE 载体中出现的可退火接头序列对或可退火接头序列/引物结合片段对, 和用于产生 pRYSE 载体骨架的 PCR 引物。

pRYSE 载体	可退火接头或引物结合片段 1 (见表 1)	可退火接头或引物结合片段 2 (见表 1)	引物 5	引物 6
1	Pme1-5'	RYSE 1	S001(SEQ ID NO: 46)	S002(SEQ ID NO: 47)
2	RYSE 1	RYSE 2	S003(SEQ ID NO: 48)	S004(SEQ ID NO: 49)
3	RYSE 2	RYSE 3	S005(SEQ ID NO: 50)	S006(SEQ ID NO: 51)
4	RYSE 3	RYSE 4	S007(SEQ ID NO: 52)	S008(SEQ ID NO: 53)
5	RYSE 4	RYSE 5	S009(SEQ ID NO: 54)	S010(SEQ ID NO: 55)
6	RYSE 5	RYSE 6	S011(SEQ ID NO: 56)	S012(SEQ ID NO: 57)
7	RYSE 6	RYSE 7	S013(SEQ ID NO: 58)	S014(SEQ ID NO: 59)
8	RYSE 7	RYSE 8	S015(SEQ ID NO: 60)	S016(SEQ ID NO: 61)
9	RYSE 2	Pme1-3'	S005(SEQ ID NO: 50)	S018(SEQ ID NO: 63)
10	RYSE 3	Pme1-3'	S007(SEQ ID NO: 52)	S018(SEQ ID NO: 63)
11	RYSE 4	Pme1-3'	S009(SEQ ID NO: 54)	S018(SEQ ID NO: 63)
12	RYSE 5	Pme1-3'	S011(SEQ ID NO: 56)	S018(SEQ ID NO: 63)
13	RYSE 6	Pme1-3'	S013(SEQ ID NO: 58)	S018(SEQ ID NO: 63)
14	RYSE 7	Pme1-3'	S015(SEQ ID NO: 60)	S018(SEQ ID NO: 63)
15	RYSE 8	Pme1-3'	S017(SEQ ID NO: 62)	S018(SEQ ID NO: 63)

[0250] lacZ 基因从 pUC19 载体使用引物 S012(SEQ ID NO: 65) 和 S028(SEQ ID NO: 66) PCR 扩增, 所述引物分别包含 SchI 限制性内切酶识别位点。反应混合物用凝胶电泳拆分, 大约 0.5kb 的 PCR 产物用凝胶纯化, 纯化的 PCR 产物与每个 pRYSE 载体骨架连接。在获得的载体上用 PCR 引物 L012(SEQ ID NO: 231) 和 L013(SEQ ID NO: 232) 实行定点诱变以从复制的起始处移除 SchI 限制性内切酶识别位点。最后, 用 PCR 引物 S036(SEQ ID NO: 67) 和 S037(SEQ ID NO: 68) 实行第二次定点诱变以从 lacZ 片段中移除 SchI 限制性内切酶识别位点, 从而产生 pRYSE 载体 1 到 15 (见图 4 的 pRYSE 载体的质粒图, 和 SEQ ID NO: 207-221 的 pRYSE 1-15 的核苷酸序列)。

[0251] 实施例 2

[0252] 本实施例描述制备 pRYSE 载体的可选方法。

[0253] 用 SEQ ID NO: 207-221 作为模板, 可以人工合成制备 pRYSE 载体 1 到 15 (例如, 通过 Biosearch Technologies, Novato, CA)。用 SEQ ID NO: 221 作为模板, 其中的 Pme1-5'

引物结合片段和 / 或 RYSE 1 可退火接头序列变为另一个合适的可退火接头序列或引物结合片段 (见表 1), 可以人工合成制备包含不同的可退火接头序列的其它 pRYSE。

[0254] 实施例 3

[0255] 本实施例描述制备 pMULE 载体的方法, 所述载体以 5' 至 3' 的方向包含: 第一 SapI 限制性内切酶识别位点, 第一 SchI 限制性内切酶识别位点, lacZ 标记基因, 第二 SchI 限制性内切酶识别位点, 和第二 SapI 限制性内切酶识别位点。所述 pMULE 载体用于克隆 Mule。

[0256] 用引物 K162 (SEQ ID NO:109) 和 K163 (SEQ ID NO:110) PCR 扩增 pRYSE 载体 8 的骨架。反应混合物用凝胶电泳解析, 大约 2.2kb 的载体骨架用凝胶纯化。用 SchI 限制性内切酶完全消化 pRYSE 载体而产生包含 lacZ 基因的 DNA 片段, 加热至 65°C 20 分钟使酶失活, 反应混合物用凝胶电泳解析, 凝胶纯化大约 0.5kb 的 DNA 片段。包含 lacZ 基因的纯化 DNA 片段与纯化的载体骨架连接, 产生 pMULE 载体 (见图 7 的质粒图)。

[0257] 实施例 4

[0258] 本实施例描述制备 “Bit” 的方法。“Bit” 是可插入 pRYSE 载体以产生装配载体的 DNA 片段, 所述装配载体包含可以通过本发明公开的方法装配进入装配好的多核苷酸的多核苷酸组分。Bit 可以编码感兴趣的基因或遗传元件 (例如, 启动子, 终止子, 选择性标记, 多克隆位点)。Bit 用表 4 所描述的模板和引物 PCR 扩增。

[0259]

表 4- 扩增的 Bit

Bit	类型 *	引物	大小 (bp)	模板
atoB	Gs	L229(SEQ ID NO: 40) L230(SEQ ID NO: 41)	1185	质粒 DNA, 包含来自大肠杆菌的 <i>atoB</i> 基因(GenBank accession number NC_000913 REGION: 2324131..2325315)
mvaS	Gs	L235(SEQ ID NO: 42) L236(SEQ ID NO: 43)	1152	合成的 DNA 片段, 包含来自 <i>Enterococcus faecalis</i> 的 <i>mvaS</i> 基因 (GenBank accession number AF290092 REGION: 142..1293), 密码子优化以在 <i>Saccharomyces cerevisiae</i> 中表达, 并包含 110 位的丙氨酸变为甘氨酸的修饰以增强酶活(见 Steussy <i>et al.</i> (2006) <i>Biotechnology</i> 45(48):14407-14414)
ERG13-1	GsT	L109(SEQ ID NO: 235) L110(SEQ ID NO: 26)	1726	<i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
3' NDT80	D	L221(SEQ ID NO: 34) L222(SEQ ID NO: 35)	516	<i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
5' NDT80	U	L219(SEQ ID NO: 32) L220(SEQ ID NO: 33)	495	<i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
tP _{FBA1}	P	L225(SEQ ID NO: 37) L057(SEQ ID NO: 234)	526	<i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
tP _{TDH3}	P	L224(SEQ ID NO: 36) L054(SEQ ID NO: 233)	559	<i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
ERG10-1	Gs	L226(SEQ ID NO: 38) L227(SEQ ID NO: 39)	1182	编码 <i>Ralstonia eutropha</i> 的乙酰辅酶 A 乙酰转移酶(GenBank accession NC_008313 REGION: 183291..184469) 的合成片段, 密码子优化以在 <i>Saccharomyces cerevisiae</i> 中表达, 随后是额外的终止密码子。
tENO1	T	L248(SEQ ID NO: 44) L176(SEQ ID NO: 27)	265	<i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
tTDH3	T	L185(SEQ ID NO: 28) L186(SEQ ID NO: 29)	260	<i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
HphA	M	TRIX_L_193(SEQ ID NO: 184) TRIX_L_194(SEQ ID NO: 185)	1912	质粒 DNA, 包含 <i>Kluyveromyces lactis</i> 的 TEF1 启动子和终止子(分别是 GenBank accession CR382122 REGIONS:788874..789380 和 787141..787496)和 <i>Klebsiella pneumoniae</i> 的 hph 基因

[0260]

tHMG1	GsT	TRIX_L_232(SEQ ID NO: 186) TRIX_L_233(SEQ ID NO: 187)	1742	<i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
tP _{GAL1,10}	P	TRIX_L_266(SEQ ID NO: 190) TRIX_L_267(SEQ ID NO: 191)	620	<i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
ERG10-2	GsT	TRIX_L_106(SEQ ID NO: 170) TRIX_L_107(SEQ ID NO: 171)	1467	<i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
ERG13-2	GsT	TRIX_L_109(SEQ ID NO: 172) TRIX_L_110(SEQ ID NO: 173)	1726	<i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
GAL80US	U	JU-218-168-130-GAL80US- F(SEQ ID NO: 134) JU-219-168-130-GAL80US- R(SEQ ID NO: 135)	500	<i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
GAL80DS	D	JU-220-168-130-GAL80DS- F(SEQ ID NO: 136) JU-221-168-130-GAL80DS- R(SEQ ID NO: 137)	500	<i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
P _{TDH3}	P	L224(SEQ ID NO: 36) TRIX_L_053(SEQ ID NO: 169)	583	<i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 基因组 DNA

[0261]

表 4- 扩增的 Bit(续)

Bit	类型 *	引物	大小 (bp)	模板
NatA	M	TRIX_L_193(SEQ ID NO: 184) TRIX_L_194(SEQ ID NO: 185)	1456	质粒 DNA, 包含 <i>Kluyveromyces lactis</i> 的 TEF1 启动子和终止子(分别是 GenBank accession CR382122 REGIONS:788874..789380 和 787141..787496)和 <i>S. noursei</i> 的 nat1 基因
ERG12	GsT	TRIX_L_112(SEQ ID NO: 174) TRIX_L_113(SEQ ID NO: 175)	1582	<i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
ERG8	GsT	TRIX_L_118(SEQ ID NO: 178) TRIX_L_119(SEQ ID NO: 179)	1616	<i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
P _{GAL4oc}	P	TRIX_K_131 (SEQ ID NO: 165) PW-91-093-CPK422-G (SEQ ID NO: 162)	270	质粒 DNA, 包含 <i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 的 GAL4 基因的启动子的“操作的构成性”形式 (Griggs & Johnston(1991) <i>PNAS</i> 88(19):8597-8601)
GAL4-1	G	JU-286-275-31-GAL4-F (SEQ ID NO: 140) JU-285-275-31-GAL4-FIX-R2 (SEQ ID NO: 139)	526	<i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
GAL4-2	G	JU-284-275-31-GAL4-FIX-F2(SEQ ID NO: 138) JU-287-275-31-GAL4-R (SEQ ID NO: 141)	2414	<i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
KanA	M	TRIX_L_193(SEQ ID NO: 184) TRIX_L_194(SEQ ID NO: 185)	1696	质粒 DNA, 包含 <i>Kluyveromyces lactis</i> 的 TEF1 启动子和终止子(分别是 GenBank accession CR382122 REGIONS:788874..789380 和 787141..787496)和 Tn903 转座子的 kanR 基因
ERG19	GsT	TRIX_L_115(SEQ ID NO: 176) TRIX_L_116(SEQ ID NO: 177)	1441	<i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
ERG20	GsT	TRIX_L_124(SEQ ID NO: 182)	1319	<i>Saccharomyces cerevisiae</i> 菌株

[0262]

		TRIX_L_125(SEQ ID NO: 183)		CEN.PK2 基因组 DNA
P _{GAL7}	P	TRIX_L_34(SEQ ID NO: 166) TRIX_L_35(SEQ ID NO: 167)	500	<i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
tP _{GAL7}	P	TRIX_L_34(SEQ ID NO: 166) TRIX_L_36(SEQ ID NO: 168)	476	<i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
IDI1	GsT	TRIX_L_121(SEQ ID NO: 180) TRIX_L_122(SEQ ID NO: 181)	1127	<i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
tP _{CTR3}	P	TRIX_K_0142(SEQ ID NO: 163) TRIX_K_0143(SEQ ID NO: 164)	710	质粒 DNA, 包含 <i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 的 CTR3 基因的启动子
LEU2US	U	JU-164-168-110-LEU2 US-f (SEQ ID NO: 129) JU-165-168-110-LEU2 US-r (SEQ ID NO: 130)	500	<i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
LEU2DS	D	JU-162-168-110-LEU2 DS-f(SEQ ID NO: 127) JU-163-168-110-LEU2 DS-r (SEQ ID NO: 128)	500	<i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
ERG9US	U	JU-108-168-110-ERG9 US-f (SEQ ID NO: 126) JU-172-168-110-ERG9 US-r1 (SEQ ID NO: 133)	499	<i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
ERG9CDS	G	JU-106-168-110-ERG9 CDS-f (SEQ ID NO: 124) JU-107-168-110-ERG9 CDS-r (SEQ ID NO: 125)	501	<i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
STE5US	U	TRIX_RN017(SEQ ID NO: 192) TRIX_RN018(SEQ ID NO: 193)	600	<i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
STE5DS	D	TRIX_RN019(SEQ ID NO: 194) TRIX_RN020(SEQ ID NO: 195)	600	<i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 基因组 DNA

[0263]

表 4- 扩增的 Bit(续)

Bit	类型 *	引物	大小 (bp)	模板
URA3	M	JU-169-168-110-URA3-f (SEQ ID NO: 131) JU-170-168-110-URA3-r (SEQ ID NO: 132)	1554	<i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
* G=基因; s=终止密码子; T=终止子; M=标记; D=下游整合区域; U=上游整合区域; P=启动子。				

[0264] PCR 扩增使用 Phusion DNA 聚合酶 (New England Biolabs, Ipswich, MA) 根据生产商建议的流程来进行。PCR 结果用凝胶电泳拆分, Bit 用凝胶纯化, 纯化的 Bit 用 T4 多核苷酸激酶 (PNK) (New England Biolabs, Ipswich, MA) 根据生产商建议的流程来处理。PNK 加热至 65℃ 20 分钟失活, 样品储存在 -20℃。

[0265] 实施例 5

[0266] 本实施例描述制备“MULE”的方法。MULE 是可以插入 pMULE 载体从而产生装配的载体的 DNA 片段, 所述装配的载体包含可以用本文公开的方法装配进入装配好的多核苷酸的多核苷酸组分。MULE 可以编码感兴趣的基因或遗传元件 (例如, 启动子、终止子、选择性标记、整合位点、表位标签、定位信号、降解信号、荧光标记、多克隆位点), 并侧邻于可退火接头序列对或可退火接头序列 / 引物结合片段对。MULE 从模板用引物 PCR 扩增, 所述引物的 3' 端与目标序列退火, 而 5' 端包含可退火接头序列或引物结合片段 (合适的可退火接头序列见表 1), 如表 5 所描述。

[0267]

表 5- 扩增的 MULE

MULE	类型 *	引物	大小 (bp)	模板
tHMG1-a	G	KMH8-276-1-linker4.tHMG1.fwd (SEQ ID NO: 157) KMh9-276-1-linker9.tHMG1.rev (SEQ ID NO: 160)	1794	RABit 254 质粒 DNA
ERG12	G	KMH46-276-43-ERG12linker4.fwd (SEQ ID NO: 151) KMh14-276-4-linker9.ERG12.rev (SEQ ID NO: 145)	1634	RABit 250 质粒 DNA
ERG19	G	KMH47-276-43-ERG19linker4.fwd (SEQ ID NO: 152) KMh15-276-4-linker9.ERG19.rev (SEQ ID NO: 146)	1493	RABit 241 质粒 DNA
P _{TDH3} -a	P	KMH81-276-116-TDH3.rev.tHMG1 (SEQ ID NO: 155) S004(SEQ ID NO: 49)	626	RABit 54 质粒 DNA
P _{TDH3} -b	P	KMH91-276-116-TDH3.rev.FS (SEQ ID NO: 158) S004(SEQ ID NO: 49)	546	RABit 54 质粒 DNA
tHMG1-b	G	KMH82-276-116-tHMG1.fwd.TDH3 (SEQ ID NO: 156) S009(SEQ ID NO: 54)	1801	RABit 20 质粒 DNA
IME1US	U	KB454-266-53(SEQ ID NO: 142) KB455-266-53(SEQ ID NO: 143)	578	<i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
IME1DS	D	KMH93-276-130-3'IME.linker4.fwd (SEQ ID NO: 161) KB457-266-53(SEQ ID NO: 144)	554	<i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
LEU2	M	VH296-235-55-Leu2 12-1 F (SEQ ID NO: 30) VH296-235-55-Leu2 12-1 R (SEQ ID NO: 31)	1795	质粒 DNA, 包含 <i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 的 LEU2 位点(Sikorski RS, Hieter(1989)Genetics 122(1):19-27)

[0268]

FS-a	G	KMH5-276-1-linker3.FS(Kozak).fwd (SEQ ID NO: 153) KMH7-276-1-linker4.TCYC1.rev (SEQ ID NO: 154)	1981	质粒 DNA, 包含 <i>Artemisia annua</i> 的法尼烯合成酶的编码序列 (GenBank accession number AY835398) 密码子优化以在 <i>Saccharomyces cerevisiae</i> 中表达, 和 <i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 的 CYC1 基因的终止子
FS-b	G	KMH92-276-116-FS.fwd.TDH3 (SEQ ID NO: 159) KMH7-276-1-linker4.TCYC1.rev (SEQ ID NO: 154)	1976	质粒 DNA, 包含 <i>Artemisia annua</i> 的法尼烯合成酶的编码序列 (GenBank accession number AY835398) 密码子优化以在 <i>Saccharomyces cerevisiae</i> 中表达, 和 <i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 的 CYC1 基因的终止子

[0269]

表 5 – 扩增的 MULEs(续)				
MULE	类型 *	引物	大小(bp)	模板
URA3blaster	M	VH228-235-7-URA3LOF3RYSE12-1F (SEQ ID NO: 204) VH229-235-7-URA3LOF3RYSE12-1R (SEQ ID NO: 205)	1565	URA-3 blaster 模板 **
* G=基因; s=终止密码子; T=终止子; M=标记; D=下游整合区域; U=上游整合区域; P=启动子。 ** The URA-3 blaster 模板通过下述方法来制备: 首先制备 DNA 片段, 所述 DNA 片段侧邻于序列 A (由包含 SEQ ID NO:206 的人工合成 DNA 片段用 PCR 引物 TRI_X_Z025(SEQ ID NO: 196)和 TRI_X_Z026(SEQ ID NO: 197)制备), 侧邻于序列 B(由包含 SEQ ID NO: 206 的人工合成 DNA 片段用 PCR 引物 TRI_X_Z027(SEQ ID NO: 198)和 TRI_X_Z028(SEQ ID NO: 199)制备), URA3-c(由 <i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 基因组 DNA 用 PCR 引物 TRI_X_Z033(SEQ ID NO: 200)和 TRI_X_Z036(SEQ ID NO: 203)制备), 和 URA3-d(由 <i>Saccharomyces cerevisiae</i> 菌株 CEN.PK2 基因组 DNA 用 PCR 引物 TRI_X_Z034(SEQ ID NO: 201)和 TRI_X_Z035(SEQ ID NO: 202)制备)。侧邻于序列 A、URA3-c 和 URA-3-d 的 DNA 片段随后通过 PCR 引物 TRI_X_Z025 和 TRI_X_Z034 缝合在一起形成 DNA 片段 A, 而 DNA 片段 URA3-c、URA3-d 和侧邻序列 B 用 PCR 引物 TRI_X_Z028 和 TRI_X_Z033 缝合在一起以形成片段 B。最后, DNA 片段 A 和 B 用 PCR 引物 TRI_X_Z025 和 TRI_X_Z028 缝合在一起, 形成 URA-3 blaster 模板。				

[0270] PCR 扩增使用 Phusion DNA 聚合酶 (New England Biolabs, Ipswich, MA) 根据生产商建议的流程来进行。PCR 结果用凝胶电泳来分析, MULE 用凝胶纯化, 纯化的 Bit 用 T4 多核苷酸激酶 (PNK) (New England Biolabs, Ipswich, MA) 根据生产商建议的流程来处理。PNK 加热至 65°C 20 分钟失活, 样品储存在 -20°C。

[0271] 实施例 6

[0272] 本实施例描述将 Bit 插入 pRYSE 载体或将 MULE 插入 pMULE 载体以形成装配载体的方法。

[0273] pRYSE 载体 1 到 8 和 pRYSE 载体 15 用 SchI 限制性内切酶完全消化,消化的 DNA 片段用南极磷酸酶 (Antarctic Phosphatase) (New England Biolabs,Ipswich,MA) 处理。磷酸酶加热至 65°C 20 分钟失活,反应混合物用凝胶电泳解析,大约 2.2kb 的 pRYSE 载体骨架 (缺少 lacZ) 用凝胶纯化。纯化的 pRYSE 载体骨架与表 6 所描述的 Bit 连接,以形成装配载体。

[0274] pMULE 载体用 SchI 限制性内切酶完全消化,反应混合物用凝胶电泳解析,大约 2.2kb 的 pMULE 载体骨架 (缺少 lacZ) 用凝胶纯化。用磷酸酶 (例如南极磷酸酶 (New England Biolabs, Ipswich, MA), CIAP (New England Biolabs, Ipswich, MA), SAP (New England Biolabs, Ipswich, MA ;Fermentas, Glen Burnie, MD), 或 FastAP (Fermentas, Glen Burnie, MD)) 处理纯化的 pMULE 载体骨架,磷酸酶加热失活 (例如 65°C 20 分钟),pMULE 载体骨架与 MULE 连接,从而形成装配载体。

[0275]

表 6- 制备的装配载体

Bit(见表 4)	pRYSE 载体(见表 3)	装配载体
atoB	4	2
mvaS	7	5
ERG13-1	7	12
3' NDT80	15	29
	10	24
5' NDT80	1	30
	1	97
tP _{FBA1}	6	35
tP _{TDH3}	3	53
ERG10-1	4	60
tENO1	8	62
tTDH3	5	64
GAL80US	1	270
HphA	2	22
tHMG1	3	254
tP _{GAL1,10}	4	229
ERG10-2	5	244
ERG13-2	6	253
tP _{GAL1,10}	7	228
tHMG1	8	255
GAL80DS	15	271
LEU2US	1	187
NatA	2	262
ERG12	3	250
ERG8	5	252
P _{GAL4oc}	6	268
GAL4 *	7	265
LEU2DS	14	263
ERG9US	1	186
KanA	2	261
ERG19	3	241
ERG20	5	251
tP _{GAL7}	6	249
IDI1	7	237
tP _{CTR3}	8	269
ERG9CDS	15	185
P _{GAL7}	3	44
STE5US	1	567
URA3	2	556(方向 1)
		555(方向 2)
P _{TDH3}	3	54
tHMG1	4	20
STE5DS	11	563

[0276]

通过使用 50ng 载体骨架, 3 摩尔过量的 Bit, 和连接酶(例如快速连接酶(Quick Ligase)(New England Biolabs, Ipswich, MA), T4 DNA 连接酶(常规和高浓度; vendor, Fermentas, Glen Burnie, MD), 快速连接酶(Fermentas, Glen Burnie, MD))根据生产商建议的流程来进行连接。

*通过用引物 JU-286-275-31-GAL4-F(SEQ ID NO: 140)和 JU-287-275-31-GAL4-R(SEQ ID NO: 141)将 BitGAL4-1 和 GAL4-2(见表 4)缝合在一起而 BitGAL4。

[0277] 装配载体转化化学感受态 TOP10 大肠杆菌亲本细胞 (Invitrogen, Carlsbad, CA)。宿主细胞转化体用含 100ug/ml 羧苄青霉素和 40ug/ml X-gal 的 Luria Bertoni (LB) 琼脂进行筛选。单独的白色克隆从 LB 琼脂转移至含 5mL 的 LB 液体培养基和羧苄青霉素的培养试管中, 培养液在 37°C 摇床以 250rpm 培养过夜。提取质粒 DNA, 测序以确认克隆含有正确的序列并且方向正确。细胞保存在含有由 400uL 无菌 50% 甘油和 600uL 液体培养基组成的 1mL 储存液的 -80°C 冷冻管里。

[0278] 实施例 7

[0279] 本实施例描述用装配载体和 / 或 MULE 将组分多核苷酸装配到装配的多核苷酸中的方法。

[0280] 装配载体 (见表 7) 置于同一个试管中 (333fmole 的每个 RABit) 并用 LguI 限制性内切酶 (Fermentas, Glen Burnie, MD) 消化。限制性内切酶通过柱浓缩移除或加热至 65°C 20 分钟失活。对涉及 MULE 或装配的多核苷酸的装配反应, 333fmole 的 MULE 或组装的多核苷酸 (见表 7) 置于一个试管中或加入消化的装配载体。样品分为 3 个 30uL 反应体系; 向每个反应混合物加入水, 缓冲液, dNTP, 和 DNA 聚合酶, 起始第一个轮 PCR 扩增。样品置于冰中, 向反应混合物中加入 0.5uM 的每种末端引物 (表 7), 并进行第二轮 PCR 扩增。3 个 PCR 反应混合物合并在一个试管, 反应混合物用凝胶电泳解析, PCR 产物用凝胶纯化。

[0281]

表 7- 用于装配的多核苷酸的末端引物

装配	装配载体(见表 6)或 MULE(见表 5)待续*	装配的多核苷酸大小 (kb) (序列)	末端引物 1	末端引物 2
1	30_22_53_60	4.3	S000 (SEQ ID NO: 45)	S009 (SEQ ID NO: 54)
2	30_22_53	3.1	S000 (SEQ ID NO: 45)	S007 (SEQ ID NO: 52)
3	22_53_60	3.7	S002 (SEQ ID NO: 47)	S009 (SEQ ID NO: 54)
4	30_22	2.5	S000 (SEQ ID NO: 45)	S005 (SEQ ID NO: 50)
5	22_53	2.5	S002 (SEQ ID NO: 47)	S007 (SEQ ID NO: 52)
6	53_60	1.8	S004 (SEQ ID NO: 49)	S009 (SEQ ID NO: 54)
7	30_22_53_60_64_35_12_62_29	7.7 (SEQ ID NO: 222)	S000 (SEQ ID NO: 45)	S019 (SEQ ID NO: 64)
8	30_22_53_60_64_35_5_62_29	7.1	S000 (SEQ ID NO: 45)	S019 (SEQ ID NO: 64)
9	30_22_53_2_64_35_5_62_29	7.1	S000 (SEQ ID NO: 45)	S019 (SEQ ID NO: 64)
10	60_64_35_5_62_29	4.1	S006 (SEQ ID NO: 51)	S019 (SEQ ID NO: 64)
11	2_64_35_5_62_29	4.1	S006 (SEQ ID NO: 51)	S019 (SEQ ID NO: 64)
Phase I-A	270_22_254_229_24_4_253	8.1 (SEQ ID NO: 111)	S000 (SEQ ID NO: 45)	S013 (SEQ ID NO: 58)
Phase I-B	228_255_271	3.0 (SEQ ID NO: 112)	S013 (SEQ ID NO: 58)	S019 (SEQ ID NO: 64)
Phase II 完全	187_262_250_229_2_52_268_265_263	9.7 (SEQ ID NO: 113)	S000 (SEQ ID NO: 45)	S019 (SEQ ID NO: 64)
Phase III-A	186_261_241_229	4.4 (SEQ ID NO: 114)	S000 (SEQ ID NO: 45)	S008 (SEQ ID NO: 53)
Phase III-B	251_249_237_269_1_85	4.3 (SEQ ID NO: 115)	S009 (SEQ ID NO: 54)	S018 (SEQ ID NO: 63)
Phase I 标记再循环	270_URA3blaster_44_FS-a_tHMG1-a	6.3 (SEQ ID NO: 116)	S000 (SEQ ID NO: 45)	S019 (SEQ ID NO: 64)
Phase II 标记再循环	187_URA3blaster_44_FS-a_ERG12	6.2 (SEQ ID NO: 117)	S000 (SEQ ID NO: 45)	S019 (SEQ ID NO: 64)
Phase III 标记再循环	186_URA3blaster_44_FS-a_ERG19	6.0 (SEQ ID NO: 118)	S000 (SEQ ID NO: 45)	S019 (SEQ ID NO: 64)

[0282]

STE5 敲除	567_556_P _{TDH3} -a_tH MG1-b_563	5.2 (SEQ ID NO: 119)	S000 (SEQ ID NO: 45)	S019 (SEQ ID NO: 64)
IME1 敲除	IME1US_LEU2_P _{TD} H3-b_FS-b_IME1DS	5.4 (SEQ ID NO: 120)	S000 (SEQ ID NO: 45)	S019 (SEQ ID NO: 64)
第一轮 PCR 扩增如下进行: 1 个循环的 98℃ 变性 2 分钟; 5 个循环的 98℃ 变性 30 秒和 72℃ 的退火/延伸 30 秒每千碱基 PCR 产物。第二轮 PCR 扩增如下进行: 一个循环的 98℃ 变性 2 分钟; 35 个循环的 98℃ 变性 12 秒和 72℃ 退火/延伸 20-25 秒每千碱基 PCR 产物; 一个循环的 72℃ 最终延伸 7 分钟; 和最终保持 4℃。当退火温度不是 72℃ (例如, 当它是 54℃ 或者 65℃), 在 PCR 扩增的第一轮 1 分钟退火步骤之后接着在 72℃ 下 30 秒每千碱基 PCR 产物的延伸步骤, 而在 PCR 扩增的第二轮 15 秒的退火步骤之后接着 72 分钟退火步骤下 20 秒每千碱基 PCR 产物延伸步骤。 * 装配载体用数字标注, 而 MULE 用名称标注。				

[0283] 如图 5 和 6 所示, 组分多核苷酸 2 到 9 被正确的装配进入 7.7kb 长的装配的多核苷酸。

[0284] 实施例 8

[0285] 本实施例描述由本文公开的方法使用装配的多核苷酸来制备遗传改变了的宿主微生物的方法。

[0286] 将 Phase I-A 和 Phase I-B 装配的多核苷酸 (见表 7) 克隆进入 TOP0 Zero Blunt II 克隆载体 (Invitrogen, Carlsbad, CA), 分别产生质粒 TOP0-Phase I-A 和 TOP0-Phase I-B。所述构建体在含 50ug/ml 卡那霉素的 LB 琼脂上生长的 TOP10 细胞 (Invitrogen, Carlsbad, CA) 中增殖。每个质粒用 NotI 限制性内切酶完全消化, Phase I-A 和 Phase I-B 插入片段用凝胶纯化试剂盒 (Qiagen, Valencia, CA) 凝胶提取, 等摩尔量的纯化的 DNA 片段用 T4DNA 连接酶 (New England Biolabs, Ipswich, MA) 连接, 产生质粒 TOP0-Phase I。所述构建体在含 50ug/ml 卡那霉素的 LB 琼脂上生长的 TOP10 细胞 (Invitrogen, Carlsbad, CA) 中增殖。

[0287] 将 Phase II 完整装配的多核苷酸 (见表 7) 克隆进入 TOP0 Zero Blunt II 克隆载体 (Invitrogen, Carlsbad, CA), 产生质粒 TOP0-Phase II。所述构建体在含 50ug/ml 卡那霉素的 LB 琼脂上生长的 TOP10 细胞 (Invitrogen, Carlsbad, CA) 中增殖。

[0288] 将 Phase III-A 和 Phase III-B 装配的多核苷酸 (见表 7) 克隆进入 TOP0 ZeroBlunt II 克隆载体 (Invitrogen, Carlsbad, CA), 分别产生质粒 TOP0-Phase III-A 和 TOP0-Phase III-B。所述构建体在含 50ug/ml 卡那霉素的 LB 琼脂上生长的 TOP10 细胞 (Invitrogen, Carlsbad, CA) 中增殖。每个质粒用 BamHI 和 SbfI 限制性内切酶完全消化, Phase III-A 和 Phase III-B 插入片段用凝胶纯化试剂盒 (Qiagen, Valencia, CA) 凝胶提取, 等摩尔量的纯化的 DNA 片段用 T4DNA 连接酶 (New England Biolabs, Ipswich, MA) 连接, 产生 Phase III 完整装配的多核苷酸。将 Phase III 完整装配的多核苷酸 (见表 7) 克隆进入 TOP0Zero Blunt II 克隆载体 (Invitrogen, Carlsbad, CA), 产生质粒 TOP0-Phase III。所述构建体在含 50ug/ml 卡那霉素的 LB 琼脂上生长的 TOP10 细胞 (Invitrogen, Carlsbad, CA) 中增殖。

[0289] 对于酵母细胞转化, 25ml 的酵母浸出粉胨葡萄糖 (YPD) 培养基用于培养单个克隆

的起始宿主菌株。培养物在 30℃ 和 200rpm 旋转振荡下生长过夜。测量培养液的 OD600, 随后用培养液孵育 50ml 的 YPD 培养基至 OD600 达到 0.15。新孵育的培养液在 30℃ 和 200rpm 旋转振荡下生长至 OD600 达到 0.7-0.9, 这时候用 1 μg 的 DNA 转化细胞。然后允许细胞在 YPD 培养基中恢复 4 小时, 之后平铺在含有鉴别宿主细胞转化体的选择试剂的琼脂上。

[0290] 通过将活性干 PE-2 酵母 (于 1994 年在 Santelisa Vale 由 Sertãozinho, Brazil 所分离) 悬浮于 5mL 的包含 100ug/mL 羧苄青霉素和 20ug/mL 卡那霉素的 YPD 培养基中而获得起始宿主菌株 Y1198。培养物在 30℃ 和 200rpm 旋转振荡下孵育过夜。然后将 10uL 培养物平铺于 YPD 平板上并使其干燥。细胞连续分散为单个克隆, 在 30℃ 下孵育 2 天。挑选 12 个单个克隆, 置于一个新的 YPD 平板上, 30℃ 生长过夜。克隆菌株的身份通过分析其染色体大小来区分, 通过 Bio-Rad CHEF DR II 系统 (Bio-Rad, Hercules, CA) 使用 Bio-Rad CHEF 基因组 DNA 插头试剂盒 (Bio-Rad, Hercules, CA) 根据生产商的说明来进行。挑选了一个克隆并作为菌株 Y1198 保存。

[0291] 菌株 Y1661, Y1662, Y1663 和 Y1664 由菌株 Y1198 通过呈递单倍体菌株而产生。菌株 1198 在一个滚轮中的玻璃试管中的 5mL 的 YPD 培养基中在 30℃ 下生长过夜。测量其 OD600, 细胞稀释在含 2% 醋酸钾的 5mL YP 培养基中, 使其 OD600 为 0.2。培养物在滚轮中的玻璃试管中 30℃ 下生长过夜。测量其 OD600, 然后通过 5000g 离心 2 分钟来收集 40D600*ml 的细胞。细胞块用无菌水冲洗一次, 然后重悬于 3mL 的含 0.02% 棉子糖的 2% 醋酸钾中。细胞在滚轮中的玻璃试管中在 30℃ 下生长 3 天。通过显微镜确认孢子形成。33uL 培养物转移至 1.5mL 的微量离心管中, 以 14000rpm 离心 2 分钟。细胞块重悬于含 2uL 的 10mg/mL 酵母裂解酶 100T (MPBiomedicals, Solon, OH) 的 50uL 无菌水中, 细胞在 30℃ 水浴中孵育 10 分钟。试管转移至冰上, 加入 150uL 冰冷的水。将 10uL 混合物加入 12mL 的 YPD 平板中, 用 Singer MSM 300 解剖显微镜 (Singer, Somerset, UK) 分开四分体。YPD 平板在 30℃ 下孵育 3 天, 之后孢子转移至新鲜的 YPD 平板中, 并 30℃ 生长过夜。来自 8 个四孢子四分体的每个孢子的交配类型用克隆 PCR 分析。挑选具有 2 个 MATA 和 2 个 MATalpha 孢子的单独的 4 孢子四分体, 保存为菌株 Y1661 (MATA), Y1662 (MATA), Y1663 (MATalpha) 和 Y1664 (MATalpha)。

[0292] 宿主菌株 1515 通过用 PmeI 限制性内切酶完全消化的质粒 TOP0-Phase I 转化菌株 Y1664 而产生。宿主细胞转化株在含 300ug/mL 潮霉素 B 的 YPD 培养基上筛选。

[0293] 通过用 PmeI 限制性内切酶完全消化的质粒 TOP0-Phase II 转化菌株 Y1515 而产生宿主菌 1762。宿主细胞转化株在含 100ug/mL 诺尔丝菌素的 YPD 培养基上选择。

[0294] 经过表达质粒 pAM404 和 PmeI 限制性内切酶完全消化的质粒 TOP0-Phase III 两步转化通过菌株 Y1762 而获得宿主菌 1770。表达质粒 pAM404 源自质粒 pAM353, 后者通过向 pRS425-Gall 载体 (Mumberg et. al. (1994) Nucl. Acids. Res. 22 (25) :5767-5768) 插入编码 β-法尼烯合成酶的核苷酸序列而产生。所述核苷酸插入序列可以人工合成, 以 *Artemisia annua* (GenBank accession number AY835398) 的 β-法尼烯合成酶的编码序列为模板, 并通过 *Saccharomyces cerevisiae* 中表达的密码子优化 (SEQ ID NO:121)。所述人工合成的核苷酸序列侧邻 5' BamHI 和 3' XhoI 限制性酶切位点, 并且因此可以克隆进克隆载体的互补限制性酶切位点例如标准的票 UC 或 pACYC 源载体。所述人工合成的核苷酸序列通过 DNA 合成构建体经 BamHI 和 XhoI 限制性内切酶完全消化而分离出来。反应混合物通过凝胶电泳拆分, 凝胶回收大约 1.7kb 的含 β-法尼烯合成酶编码序列的 DNA 片段, 分离的 DNA

片段连接至 pRS425-Gal1 载体的 BamHI XhoI 限制性酶切位点,产生表达质粒 pAM353。编码 β -法尼烯合成酶的核苷酸序列从 pAM353 用引物 GW-52-84pAM326 BamHI (SEQ ID NO:188) 和 GW-52-84pAM326NheI (SEQ ID NO:189) 经 PCR 扩增。得到的 PCR 产物用 BamHI 和 NheI 限制性内切酶完全消化,凝胶电泳回收反应混合物,凝胶提取大约 1.7kb 的含 β -法尼烯合成酶编码序列的 DNA 片段,分离的 DNA 片段连接进入载体 pAM178 (SEQ ID NO:122) 的 BamHI NheI 限制性酶切位点,产生表达载体 pAM404。pAM404 的宿主细胞转化株用缺少蛋氨酸和亮氨酸的完全合成培养基 (Complete Synthetic Medium, CSM) 筛选。pAM404 和 Phase III 完全装配的核苷酸的宿主细胞转化株用缺少蛋氨酸和亮氨酸并含 200ug/mL 的 G418 的 CSM 筛选。

[0295] 宿主菌 1793 通过 URA3 敲除构建体 (SEQ ID NO:123) 转化菌株 Y1770 而产生。所述敲除构建体用以下方法产生:首先产生 DNA 片段 URA3US (从 *Saccharomyces cerevisiae* 菌株 CEN.PK2 基因组 DNA 用 PCR 引物 KMH33-276-21-URA3 5'.fwd (SEQ ID NO:147) 和 KMH34-276-21-URA35'.rev (SEQ ID NO:148) 产生) 和 URA3DS (从 *Saccharomyces cerevisiae* 菌株 CEN.PK2 基因组 DNA 用 PCR 引物 KMH35-276-21-URA3 3'.fwd (SEQ ID NO:149) 和 KMH36-276-21-URA3 3'.rev (SEQ ID NO:150) 产生), 随后用 PCR 引物 KMH33-276-21-URA3 5'.fwd 和 KMH36-276-21-URA3 3'.rev 将这 2 个 DNA 片段缝合在一起。宿主细胞转化株用含 5-FOA 的 YPD 培养基筛选。

[0296] 宿主菌株 YAAA 由 Phase I 标记循环装配的多核苷酸 (见表 7) 转化菌株 Y1793 而产生。宿主细胞转化株在缺乏蛋氨酸和尿嘧啶的 CSM 上筛选。URA3 标记通过以下步骤来去除:YPD 培养基上 30°C 并 200rpm 旋转振荡培养细胞过夜,然后将细胞平铺于含 5-FOA 的琼脂上。标记去除通过克隆 PCR 来确认。

[0297] 宿主菌株 YBBB 由 Phase II 标记循环装配的多核苷酸 (见表 7) 转化菌株 YAAA 而产生。宿主细胞转化株在缺乏蛋氨酸和尿嘧啶的 CSM 上筛选。URA3 标记通过以下步骤来去除:YPD 培养基上 30°C 并 200rpm 旋转振荡培养细胞过夜,然后将细胞平铺于含 5-FOA 的琼脂上。标记去除通过克隆 PCR 来确认。

[0298] 宿主菌株 Y1912 由 Phase III 标记循环装配的多核苷酸 (见表 7) 转化菌株 YBBB 而产生。宿主细胞转化株在缺乏蛋氨酸和尿嘧啶的 CSM 上筛选。URA3 标记通过以下步骤来去除:YPD 培养基中在 30°C 并 200rpm 旋转振荡培养细胞过夜,然后将细胞平铺于含 5-FOA 的琼脂上。标记去除通过克隆 PCR 来确认。

[0299] 宿主菌株 Y1913 由 STE5 敲除的装配的多核苷酸 (见表 7) 转化菌株 Y1912 而产生。宿主细胞转化株在缺乏蛋氨酸和尿嘧啶的 CSM 上筛选。

[0300] 宿主菌株 Y1915 从菌株 Y1913 产生,从 pAM404 中分离菌株 Y1913 并用 IME1 敲除的装配的多核苷酸 (见表 7) 转化获得的菌株。菌株 Y1913 在非选择性 YPD 培养基中在 30°C 以 200rpm 旋转振荡增殖。大约 100 个细胞平铺在 YPD 固体培养基上并 30°C 生长 3 天,然后重新平铺在缺少蛋氨酸和亮氨酸的 CSM 平板中,30°C 再生长 3 天。分离的细胞通过其在含亮氨酸的矿物培养基上生长的能力和其在缺乏亮氨酸的培养基上不能生长而被鉴定。一个单独的这种克隆被挑选出来并用 IME1 敲除的装配的多核苷酸转化。宿主细胞转化体在缺乏蛋氨酸和尿嘧啶的 CSM 上筛选。

[0301] 实施例 9

[0302] 本实施例描述筛选可退火接头序列的方法,所述可退火接头序列用于通过本文描述的发明方法将编码启动子的组分多核苷酸和蛋白编码序列装配进入装配好的多核苷酸。

[0303] 如表 8 所描述,通过 PCR 扩增编码启动子并紧随两个不同的候选可退火接头序列(可退火接头序列 RYSE 15(R15;SEQ ID NO:15)和可退火接头序列 RYSE 7(R7;SEQ ID NO:7))的 MULE 以及编码 GFP 并且之前有 2 个可退火接头序列的 MULE。

[0304]

表 8 - 扩增的编码启动子和 GFP 并具有可退火接头序列 RYSE 15(R15)或可退火接头序列 RYSE 7(R7)的 MULE

MULE	类型 *	引物	大小 (bp)	模板
pGAL1-R15	P	Plan X19(SEQ ID NO: 236) Plan X20(SEQ ID NO: 237)	698	<i>S. cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
pTDH3-R15	P	Plan X47(SEQ ID NO: 238) Plan X48(SEQ ID NO: 239)	613	<i>S. cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
pCYC1-R15	P	Plan X11(SEQ ID NO: 240) Plan X12(SEQ ID NO: 241)	645	<i>S. cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
pGAL1-R7	P	Plan X19(SEQ ID NO: 236) Plan X64(SEQ ID NO: 242)	692	<i>S. cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
pTDH3-R7	P	Plan X47(SEQ ID NO: 238) Plan X71(SEQ ID NO: 243)	607	<i>S. cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
pCYC1-R7	P	Plan X11(SEQ ID NO: 240) Plan X78(SEQ ID NO: 244)	639	<i>S. cerevisiae</i> 菌株 CEN.PK2 基因组 DNA
R7-GFP	GsT	Plan X96(SEQ ID NO: 247) Plan X88(SEQ ID NO: 245)	1378	RABit 634 质粒 DNA **
A-GFP	GsT	Plan X89(SEQ ID NO: 246) Plan X88(SEQ ID NO: 245)	1385	RABit 634 质粒 DNA **

PCR 反应包括: 67 uL ddH₂O, 20 uL 5×HF 缓冲液, 2 uL 的每种引物(10uM), 1 uL dNTP 混合物(200 uM), 1 uL Phusion DNA 聚合酶(New England Biolabs, Ipswich, MA), 和 9 uL Y002 基因组 DNA 或 RABit 634 质粒 DNA。

PCR 扩增如下进行: 1 循环的 98℃ 变性 2 分钟; 9 循环的 98℃ 变性 15 秒, 61℃ 退火 30 秒且每循环减 1℃, 72℃ 延伸 1 分钟; 26 循环的 98℃ 变性 15 秒, 52℃ 退火 30 秒, 72℃ 延伸 1 分钟; 1 循环的 72℃ 最终延伸 7 分钟; 最后保持在 4℃。

* G=基因; s=终止密码子; T=终止子; P=启动子。

**RABit 634 包含绿色荧光蛋白(GFP)编码序列继之以 *Saccharomyces cerevisiae* 的 ADH1 基因的终止子。

[0305] PCR 反应通过凝胶电泳拆分,凝胶纯化 MULE,纯化的 MULE 用于装配测试性的装配好的多核苷酸。在其之后,用于装配的(见表 9)MULE 和装配载体(见表 6)置于同一个试

管中（每种装配载体 333fmole, 每种 MULE667fmole），并用 LguI 限制性内切酶（Fermentas, Glen Burnie, MD）消化。限制性内切酶加热至 65℃ 20 分钟使其失活。样品分为 3 个 30uL 反应体系；水、缓冲液、dNTP 和 DNA 聚合酶加入每个反应混合物中，起始 PCR 扩增的第一轮。末端引物加入反应混合物，进行 PCR 扩增的第二轮（见表 9）。3 个 PCR 反应混合物合并在一个试管里，反应混合物用凝胶电泳解析，凝胶纯化 PCR 产物。

[0306]

表 9 – 用于装配测试性的装配好的多核苷酸的末端引物				
装配	要组合的 MULE(见表 8)和装配载体(见表 6)*	装配好的多核苷酸大小(kb)	末端引物 1	末端引物 2
1	97 555 pGAL1-A A-GFP 24	4.7	S000 (SEQ ID NO: 45)	S019 (SEQ ID NO: 64)
2	97 555 pTDH3-A A-GFP 24	4.6		
3	97 555 pCYC1-A A-GFP 24	4.7		
7	97_555_ pGAL1-R7_R7-GFP 24	4.7		
8	97 555 pTDH3-R7 R7-GFP 24	4.6		
9	97 555 pCYC1-R7 R7-GFP 24	4.6		
PCR 反应物包括: 41 uL ddH ₂ O, 20 uL 5×HF 缓冲液, 5 uL 的每种末端引物(1 uM), 2 uL dNTP 混合物(200 uM), 1.8 uL Phusion DNA 聚合酶, 和 30 uL MULE 或 LguI 消化的装配载体。				
PCR 扩增的第一轮如下进行: 1 循环的 98℃ 变性 2 分钟; 5 循环的 98℃ 变性 30 秒, 60℃ 退火 30 秒, 和 72℃ 延伸 2.5 分钟; 随后保持在 4℃ 以加入两种末端引物。PCR 扩增的第二轮如下进行: 1 循环的 98℃ 变性 2 分钟; 35 循环的 98℃ 变性 12 秒, 60℃ 退火 30 秒, 和 72℃ 延伸 2.5 分钟; 1 循环的 72℃ 最终延伸 7 分钟; 和最后保持在 4℃。				
* 装配载体以数字标明, 而 MULE 用名称标明。				

[0307] 测试性的装配好的多核苷酸用于转化具有 URA3 缺陷和 GAl1ery80 位点删除的 *Saccharomyces cerevisiae* 宿主菌株。宿主细胞转化体用缺少尿嘧啶的 CSM 筛选，装配好的多核苷酸正确整合进入基因组通过克隆 PCR 来确认。来自每次转化的两个经鉴定的克隆被挑选进 360uL 含 2% 蔗糖的鸟食种子培养基 (Bird Seed Medium, BSM)，培养物在 30℃ 和 999rpm 旋转振荡下孵育 48 小时。从每个孔取 14.4uL 并转移进 96 孔包被板上含 4% 蔗糖的 1.1mL BSM 中，30℃ 和 999rpm 旋转振荡下培养另外 6 小时，然后 100uL 的每种培养物转移到干净底部的 96 孔板的孔中以检测 GFP 表达。每个孔的 GFP 表达在 M5Plate Reader 分光光度计 (Molecular Devices, Sunnyvale, CA) 上通过测量 485nm 激发后的 515nm 辐射来检测。通过测量的 GFP 浓度除以每份培养液的 OD600 来使 GFP 浓度相对于细胞生长标准化。

[0308] 如表 10 所示，相比于可退火接头序列 RYSE7，可退火接头序列 RYSE15 使增加的 GAL1、TDH3 和 CYC1 启动子驱动测试性的装配好的多核苷酸中 GFP 报告基因的表达。

[0309]

表 10—在启动子和 GFP 报告子之间包含可退火接头序列 RYSE 15(R15)或者可退火接头序列 RYSE 7(R7)的测试性的装配好的多核苷酸的宿主细胞的 GFP 表达

在测试性的装配好的多核苷酸中的启动子和 GFP 报告基因之间的可退火接头序列	平均%GFP 表达(相比于具有 3 个无缝对照构建体*之一的宿主细胞的平均%GFP 表达, 2 个独立宿主细胞分离株的平均值)			
	GAL1 启动子	TDH3 启动子	CYC1 启动子	所有 3 个启动子的平均值
R15	79.34	91.42	81.92	84.22
R7	27.43	54.68	46.31	42.81

* 所述无缝对照构建体具有与测试性的装配好的多核苷酸相同的结构, 除了其启动子序列无缝连接至 GFP 报告基因(即, 没有介入性可退火接头序列)。

[0310] 实施例 10

[0311] 本实施例描述多核苷酸高通量组合装配的方法, 和高通量制备包含组合的组装的多核苷酸的宿主细胞的方法。

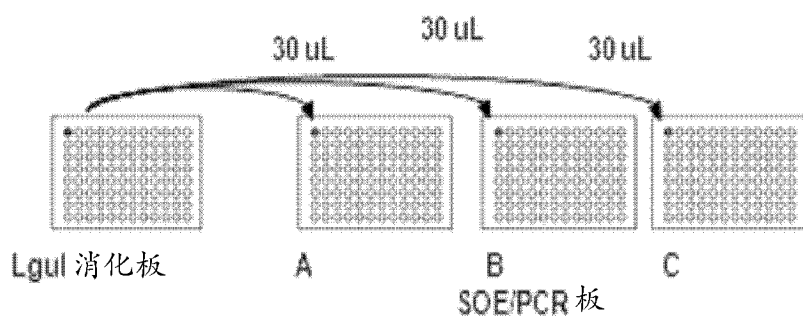
[0312] 本实施例所用的组分多核苷酸, 以及从这些组分多核苷酸产生的期望的装配好的和组合的多核苷酸, 在图 12A 中示意性展示。组分多核苷酸包含编码上游和下游染色体靶向序列的 DNA 片段 (US 和 DS), 6 个不同启动子 (P), 35 个不同的蛋白 (G), 和 URA3 选择标签的 5' 和 3' 片段 (分别是 URA 和 RA3), 侧邻于可退火接头序列对或引物结合片段 / 可退火接头序列对。

[0313] 组分多核苷酸通过用 LguI 限制性内切酶消化 RABit 或 MULE 而从装配载体中释放出来。在这之后, 设置如下表所示的 96 孔板 (“LguI 消化板”), 这些板在 37°C 下孵育 75 分钟, 之后 LguI 限制性内切酶在 PCR 仪器中加热至 65°C 20 分钟使其失活。

LguI 消化板	
组分(每孔)	体积(uL)
667 fMoleRABit 或 MULE	变量
10×Tango 缓冲液(Fermentas, Glen Burnie, MD)	10
LguI(Fermentas, Glen Burnie, MD)	2.5
ddH ₂ O	至 100

[0315] 组分多核苷酸通过 SOE 来装配。对每个 LguI 消化板, 设置三份 96 孔板 (“SOE/PCR 板”) 并在 PCR 仪器中按下表所示进行热循环:

[0316]

SOE/PCR 板			
组分(每孔)		体积(uL)	
ddH ₂ O		41	
5×Phusion HF 缓冲液 (New England Biolabs, Ipswich, MA)		20	
10 mM dNTP 混合物		2	
Phusion DNA 聚合酶(New England Biolabs, Ipswich , MA)		1.8	
LguI 消化的用于装配的 RABit 或 MULE		30	
			
		合计: 95	
热循环条件			
起始变性		98°C	2 分钟
7 个循环	变性	98°C	30 秒
	退火	67°C	30 秒
	延伸	72°C	5 分钟
保持		4°C	∞

[0317] 装配好的多核苷酸用 PCR 扩增。每个 SOE/PCR 板接受添加的试剂并在 PCR 仪器中热循环,如下表所示。SOE/PCR 板相应的孔汇集在 96 深孔板块上,装配好的多核苷酸用 Omega Biotek E-Z 96®循环-纯化试剂盒 (Omega Bio-Tek Inc., Norcross, GA) 按照生产商建议的步骤纯化 (大约终体积为 45uL)。

[0318]

SOE/PCR 板			
添加组分(每孔)		体积(uL)	
10 mM 的末端引物 S000(SEQ ID NO: 45)和 S019(SEQ ID NO: 64)		10	
热循环条件			
起始变性		98°C	2 分钟
35 个循环	变性	98°C	12 秒
	退火	67°C	30 秒
	延伸	72°C	4.5 分钟
最终延伸		72°C	7 分钟
保持		4°C	∞

[0319] 图 12B 展示了示例性装配好的多核苷酸(方框中)溶于 1% 琼脂糖凝胶。

[0320] 纯化的装配好的多核苷酸用 LguI 限制性内切酶消化以产生用于克隆的粘性末端。之后,设置 96 孔板(“LguI 装配好的多核苷酸消化板”)如下表所示,这些板在 37°C 下孵育 60 分钟,之后在 PCR 仪器中 LguI 限制性内切酶加热至 65°C 20 分钟使其失活。LguI 消化的装配好的多核苷酸用 ZR-96Zymoclean™ 凝胶 DNA 回收试剂盒(Zymo Research Corporation, Orange, CA) 按照生产推荐步骤纯化。

LguI 装配好的多核苷酸 LguI 消化板		
	组分(每孔)	体积(uL)
[0321]	纯化的组装好的多核苷酸	43
	10x Tango 缓冲液	5
	LguI	2

[0322] 装配好的多核苷酸连接至基于 pUC-19 的载体骨架。当插入片段没有连接至该载体时,pTRC 启动子(即,Saccharomyces cerevisiae 的 TRC 基因的启动子)驱动 GluRS 的表达并杀死宿主细胞。设置 96 孔板(“连接板”)如下表所示,这些板在 24°C 下孵育 15 分钟,然后 16°C 过夜。连接产物用 ZR-96DNAClean & Concentrator™-5(Zymo Research Corporation, Orange, CA) 按照生产商建议的步骤纯化。

[0323]

连接板	
组分(每孔)	体积(uL)
ddH ₂ O	5
10×T4 DNA 连接酶缓冲液	2
载体骨架	2
纯化的装配好的多核苷酸	10
T4 DNA 连接酶(NEB, Ipswich, MA)	1

[0324] 连接产物电穿孔进入大肠杆菌感受态细胞。设置预冷的 96 孔电穿孔板,并如下表所示进行电穿孔。

[0325]

电穿孔板		
组分(每孔)		体积(uL)
纯化的连接产物		10
Lucigen 10G 感受态细胞(Lucigen Corporation, Middleton, WI)		25
电穿孔设置		
2400V	750 Ω	25 μ F

[0326] 设置包含 250uL 预温的 SOC 的 1.1mL 的 96 孔培养板 (“培养板”), 从每个孔取 100uL 的 SOC 并在电穿孔之后立即加入电穿孔细胞。混合所述 SOC 和细胞, 每种混合物的 100uL 转移回培养板。培养板在 37°C 下在 Multitron II 培养摇床 (ATR Biotech, Laurel, MD) 中孵育 1 小时。两份细胞稀释物 (3uL 和 240uL) 平铺在包含 50ug/mL 卡那霉素的 LB 琼脂上, 37°C 下孵育过夜。挑选克隆并在每孔包含 1mL LB 培养基和卡那霉素的 96 深孔板中生长, 提取 DNA 用于 LguI 限制性内切酶进行限制性酶切位点分析。24 个包含大约 8kb 的组合多核苷酸的示例性克隆的中 22 个的这种限制性酶切位点分析结果显示在图 12C 中。

[0327] 包含染色体整合的组合多核苷酸的酵母细胞通过宿主细胞介导的末端染色体靶向序列和装配好的多核苷酸的选择性标记片段之间的同源重组而产生。之后, 设置 96 孔 PCR 板 (“酵母转化板”) 并在 PCR 仪器中进行热休克转化, 如下表所示。

[0328]	酵母转化板	
	组分(每孔)	体积(uL)
	Miniprep DNA(20 ng/uL)	10
	感受态酵母细胞*	40
	PEG/SS/LiAc 预混 **	200
	热休克	
	30°C	30 分钟
	42°C	45 分钟
	24°C(可选)	30 分钟
	*通过如下方法制备: 细胞在 100mL 的 YPD 中培养过夜, 稀释培养液并生长过夜至 OD600 为大约 0.8, 培养液在 3000g 旋转 5 分钟, 用 1L ddH ₂ O 洗细胞块, 用 1L 100mM 醋酸锂 (LiAc) 洗细胞块, 将细胞块重悬于 100mM LiAc 中得到总共 18mL 的体积。	
	** 足够 4 个 PCR 板的预混物包含 100 mL 50% PEG, 4 mL 煮沸的(95°C 至少 10 min)单链 DNA, 15 mL 1M LiAc。	

[0329] 酵母转化板以 2000g 旋转 2 分钟, 弃去上清液, 细胞块用 200uL ddH₂O 洗 3 次。细胞块用 100uL 冷的鸟食种子培养基 (BSM) 重悬, 所述 BSM 来自事先预冷的每孔含 360uLBSM 的冷 96 孔培养板 (“种子板”)。悬浮细胞转移到种子版, Multitron II 培养摇床中 30°C 下

生长过夜。种子板 3000g 旋转 5 分钟,移除所有液体只保留 60uL,盖好的种子板以 1000rpm 振荡以重悬细胞块。

[0330] 实施例 11

[0331] 本实施例描述通过宿主细胞介导的同源重组来产生包含装配好的多核苷酸的酵母细胞的方法。

[0332] 本实施例所用的装配好的多核苷酸和组分多核苷酸,及期望通过装配和染色体整合获得的染色体位点,在图 13A 中示例性地表示。

[0333] 酵母细胞转化用下表所描述的方法进行。热休克之后,细胞被离心,弃去上清,细胞重悬于 400uL ddH₂O,宿主细胞转化株通过将 100-200uL 的细胞悬浮液平铺在缺乏尿嘧啶的琼脂上来筛选。

[0334]

酵母转化	
组分	体积(uL)
组分和装配好的多核苷酸(分别 300-500 ng)	20
感受态酵母细胞*	细胞块*
50% PEG 溶液	240
1 M LiAc pH 8.4-8.9	36
煮沸的(95°C 5 分钟)单链 DNA(10 mg/mL)(Invitrogen, Carlsbad, CA)	10
ddH ₂ O	54
热休克	
42°C	40 min
*用以下方法制备: 来自一个克隆的细胞在 25mL YPD 中 30°C 下生长过夜至 OD600 为 0.7-0.9,离心细胞,用 5-10mL ddH ₂ O 洗涤细胞块,用 1mL ddH ₂ O 洗细胞块,用 1mL 100mM 醋酸锂(LiAc)洗细胞块,在微量离心管中离心 30 秒使细胞成块,弃去上清。	

[0335] 装配好的多核苷酸的成功整合通过 cPCR 使用 cPCR 引物 A、B、E 和 F(染色体整合位点的 5' 连接)或 cPCR 引物 C、D、G 和 H(染色体整合位点的 3' 连接)来确认(图 13A)。如图 13B 所示,分析的所有 8 个克隆都产生了 600bp 的 PCR 条带,这表明发生了期望的装配好的多核苷酸的阳性染色体整合事件,而没有出现自身位点会产生的 950bp 的 PCR 条带。

[0336] 实施例 12

[0337] 本实施例描述高通量制备通过宿主细胞介导的同源重组而产生的包含组合装配的和组合的组合多核苷酸的酵母细胞的方法。

[0338] 本实施例所用的组分多核苷酸,和通过宿主细胞介导的同源重组来装配和组合而获得的期望的组合多核苷酸,在图 14A 中示例性的表示。所述组分多核苷酸包含编码下列的 DNA 片段:上游和下游染色体靶向序列(US 和 DS),6 个不同的启动子(P),35 个不同的蛋白(G),和 RUA3 选择性标记的 5' 和 3' 片段(分别是 URA 和 RA3),侧邻于可退火接头序列对或引物结合片段/可退火接头序列对。

[0339] 组分多核苷酸通过用 LguI 限制性内切酶消化 RABit 或 MULE 而从装配载体中释放

出来。之后,设置 96 孔板 (“LguI 消化板”)如下表所示,这些板在 37℃ 下孵育 75 分钟,之后 LguI 限制性内切酶在 PCR 仪器中加热至 65℃ 20 分钟使其失活。

[0340]	LguI 消化板	
	组分(每孔)	体积(uL)
	667 fMoleRABit 或 MULE	变量
	10×Tango 缓冲液(Fermentas, Glen Burnie, MD)	5
	LguI(Fermentas, Glen Burnie, MD)	2.5
	ddH ₂ O	至 50

[0341] 为了制备包含染色体整合的组分的组装好的多核苷酸和组合的组合多核苷酸的酵母细胞,设置 96 孔 PCR 板 (“酵母转化板”),在 PCR 仪器中进行热休克转化,如下表所示。

[0342]	酵母转化板	
	组分(每孔)	体积(uL)
	组分多核苷酸	10
	感受态酵母细胞*	40
	PEG/SS/LiAc 预混 **	200
	热休克	
	30°C	30 分钟
	42°C	45 分钟
	24°C(可选)	30 分钟
*通过如下方法制备: 细胞在 100mL 的 YPD 中培养过夜, 稀释培养液并生长过夜至 OD600 为大约 0.8, 培养液在 3000g 旋转 5 分钟, 用 1L ddH ₂ O 洗涤细胞块, 用 1L 100mM 醋酸锂 (LiAc)洗涤细胞块, 将细胞块重悬于 100mM LiAc 中得到总共 18mL 的体积。		
** 足够 4 个 PCR 板的预混物包含 100 mL 50% PEG, 4 mL 煮沸的(95°C 至少 10 分钟)单链 DNA, 15 mL 1M LiAc.		

[0343] 酵母转化板以 2000g 旋转 2 分钟,弃去上清,细胞块用 200uL ddH₂O 洗三次。细胞块用 100uL 冷的乌食种子培养基 (BSM) 重悬,所述 BSM 来自事先预冷的每孔含 360uLBSM 的冷 96 孔培养板 (“种子板”)。悬浮细胞转移到种子版, Multitron II 培养摇床中 30°C 下生长过夜。种子板 3000g 旋转 5 分钟,移除所有液体只保留 60uL,盖好的种子板以 1000rpm 振荡以重悬细胞块。挑选并分析具有功能性 URA3 选择性标记的酵母细胞转化体的克隆。

[0344] 本说明书中所引用的所有出版物、专利和专利申请都纳入本文作为参考,如同每个单独的出版物或专利申请都专门地和独立地指明纳入作为参考。尽管为了清楚和易于理解的目的,上述发明以示例性或实施例的形式描述了一些细节,但是很显然具有本领域常

规技术的人员根据本发明的教导可以做出另外的改变和修饰,而不脱离所附加的权利要求的精神和范围。上文所描述的本发明的实施方式仅仅是实例性的,本领域技术人员可以认识到,或者能够使用仅仅常规实验而确认,本文所描述的特定方法的大量等同替代。所有这些等同替代被认为是在本发明的范围之内并以权利要求来覆盖。另外,如说明书和权利要求书中所用到的,单数形式的“一”或“这”包含复数形式,除非其内容做出其他明确的指示。

[0001]

序列表

<110> 阿迈瑞斯公司
扎克·塞贝
雷蒙德·洛
杰弗里·A·尤伯萨克斯
苏尼尔·S·钱德兰
埃里克·杰迪戴亚·迪安
达伦·M·普拉特
肯尼思·俊树·竹冈

<120> 多核苷酸装配的组合物和方法

<130> 11836-047-228

<150> 61/116,109

<151> 2008-11-19

<150> 61/162,230

<151> 2009-03-20

<160> 220

<170> FastSEQ for windows Version 4.0

<210> 1

<211> 24

<212> DNA

<213> 人工序列

<220>

<223> 接头 Gold (Au)

<400> 1

gctcacacgc ggccaggggg agcc

24

<210> 2

<211> 24

<212> DNA

<213> 人工序列

<220>

<223> 接头 Lapis (La)

<400> 2

cgctcgtcca acgccggcgg acct

24

<210> 3

<211> 24

<212> DNA

<213> 人工序列

<220>

<223> 接头 Copper (Cu)

<400> 3

atccccgcgt gcttgccgg ccgt

24

<210> 4

<211> 24

<212> DNA

<213> 人工序列

<220>

[0002]

<223> 接头 Quartz (Qz)	
<400> 4 aacctgcagg ccgcgagcgc cgat	24
<210> 5	
<211> 24	
<212> DNA	
<213> 人工序列	
<220>	
<223> 接头 Iron (Fe)	
<400> 5 aacgcgatcg ccgacgccgc cgat	24
<210> 6	
<211> 24	
<212> DNA	
<213> 人工序列	
<220>	
<223> 接头 obsidian (ob)	
<400> 6 aaggcggccg ctggcgaggg agat	24
<210> 7	
<211> 24	
<212> DNA	
<213> 人工序列	
<220>	
<223> 接头 Indigo (In)	
<400> 7 aaggcgcgcc acggtcgtgc ggat	24
<210> 8	
<211> 24	
<212> DNA	
<213> 人工序列	
<220>	
<223> 接头 Silver (Ag)	
<400> 8 agccccctcag cccccctagc gtcg	24
<210> 9	
<211> 28	
<212> DNA	
<213> 人工序列	
<220>	
<223> 接头 Pme1-5prime	
<400> 9 gacggcacgg ccacgcgttt aaaccgcc	28
<210> 10	
<211> 28	
<212> DNA	
<213> 人工序列	
<220>	
<223> 接头 Pme1-3prime	

[0003]

<400> 10
cggtgttttaa accccagcgc ctggcggg

28

<210> 11
<211> 2737
<212> DNA
<213> 人工序列

<220>
<223> pRYSE 输入载体 1

<400> 11
gtaaaacgac ggccagtatt aaccctcact aaagggaaact cgaggctctt cagacggcac 60
ggccacgcgt ttaaaccgcc tggcagactc catatgctat gcggcatcag agcagattgt 120
actgagagtg caccatatgc ggtgtgaaat accgcacaga tgcgtaagga gaaaataccg 180
catcaggcgc cattcgccat tcaggctgcg caactgttgg gaagggcgat cgggtgcggg 240
ctcttcgcta ttacgccagc tggcgaaagg gggatgtgct gcaaggcgat taagttgggt 300
aacgccaggg ttttcccagt caccagcgtt taaaacgacg gccagtgaat tcgagctcgg 360
taccggggga tcctctagcg tcgacctgca ggcattgcaag cttggcgtaa tcatggtcat 420
agctgtttcc tgtgtgaaat tgttatccgc tcacaattcc acacaacata cgagccggaa 480
gcataaagtg taaagcctgg ggtgcctaata gattgagcta actcacatta attgcgttgc 540
gcgctagcga gtcattccagc tcacacgcgg ccagggggag cctgaagagc gagctcccgc 600
tgagcaataa ctatagtcac agctgtttcc tgggtcgttc ggctgcggcg agcggtatca 660
gtcactcaaa agcggttaac acggttatcc acagaatcag gggataacgc aggaaagaac 720
atgtgagcaa aaggccagca aaaggccagg aaccgtaaaa aggccgcgtt gctggcggtt 780
ttccataggc tccgcccccc tgacgagcat cacaaaaatc gacgctcaag tcagaggtgg 840
cgaaacccga caggactata aagataccag gcgtttcccc ctggaagctc cctcgtgcgc 900
tctcctgttc cgaccctgcc gcttaccgca tacctgtccg cctttctccc ttcgggaagc 960
gtggcgcgtt ctcatagctc acgctgtagg tatctcagtt cggtgtagggt cgttcgctcc 1020
aagctgggct gtgtgcacga acccccgtt cagcccgacc gctgcgcctt atccggtaac 1080
tatcgtcttg attccaaccc ggtaagacac gacttatcgc cactggcagc agccactggg 1140
aacaggatta gcagagcgag gtatgtaggc ggtgctacag agttcttgaa gtggtggcct 1200
aactacggct acactagaag aacagtattt ggtatctgcg ctctgctgaa gccagttacc 1260
ttcggaaaaa gagttggtag ctcttgatcc ggcaaaacaaa ccaccgctgg tagcgggtgg 1320
ttttttgttt gcaagcagca gattacgcgc agaaaaaaag gatctcaaga agatcctttg 1380
atcttttcta cggggtctga cgctcagtg aacgaaaact cacgttaagg gatttttggtc 1440
atgagattat caaaaaggat cttcacctag atccttttaa attaaaaatg aagtttttaa 1500
tcaatctaaa gtatatatga gtaacttggt cgcatgctta ccaatgctta atcagtgagg 1560
cacctatctc agcgatctgt ctatttcgtt catccatagt tgctgactg cccgtcgtgt 1620
agataactac gatacgggag ggcttaccat ctggccccag tgctgcaatg ataccgcgag 1680
acccagctc accggctcca gatttatcag caataaacca gccagccgga agggccgagc 1740
gcagaagtgg tcctgcaact ttatccgcct ccattccagtc tattaattgt tgccgggaag 1800
ctagagtaag tagttcgcca gttaatatgt tgcgcaacgt tggtgccatt gctacaggca 1860
tcgtggtgtc acgctcgtcg tttggtatgg cttcattcag ctccggttcc caacgatcaa 1920
ggcgagttac atgatcccc atgttgtgca aaaaagcggg tagctccttc ggtcctccga 1980
tcgttgctag aagtaagtgg gccgcagtg tatcactcat gggtatggca gactgcata 2040
attctcttac tgtcatgcca tccgtaagat gcttttctgt gactggtgag tactcaacca 2100
agtcattctg agaatagtgt atgcggcgac cgagttgctc ttgcccggcg tcaatacggg 2160
ataataccgc gccacatagc agaactttaa aagtgtcat cattggaaaa cgttcttcgg 2220
ggcgaaaact ctcaaggatc ttaccgctgt tgagatccag ttcgatgtaa cccactcgtg 2280
cacccaactg atcttcagca tcttttactt tcaccagcgt ttctgggtga gcaaaaaacag 2340
gaaggcaaaa tgccgcaaaa aagggaataa gggcgacacg gaaatgttga atactcatca 2400
attgcctttt tcaatattat tgaagcattt atcagggtta ttgtctcatg agcggttaca 2460
tatttgaaatg tatttagaaa aataaaciaa taggggttcc gcgcacattt ccccgaaaag 2520
tgccacctga cgtctaagaa accattatta tcatgacatt aacctataaa aataggcgta 2580
tcacgaggcc ctttcatctc gcgcgtttcg gtgatgacgg tgaaaacctc tgacacatgc 2640
agctcccgga gacagtcaca gcttgctgtg aagcggatgc cgggagcaga caagcccgtc 2700
agggcgcgtc agcgggtgtt ggcggtgtc ggggctg 2737

<210> 12
<211> 7692
<212> DNA
<213> 人工序列

<220>
<223> 缝合7产生的DNA片段

[0004]

<400> 12

gacggcacgg	ccacgcgttt	aaaccgcctt	ccaagctgac	ataaatcgca	ctttgtatct	60
actttttttt	attcgaaaac	aaggcacaac	aatgaatcta	tcgccctgtg	agattttcaa	120
tctcaagttt	gtgtaataga	tagcggtata	ttatagaact	ataaagggtcc	ttgaatatac	180
atagtgtttc	attcctatta	ctgtatatgt	gactttacat	tgttacttcc	gcggtctattt	240
gacgtttttc	gcttcagggtg	cggcttgagg	ggcaaagtgt	cagaaaatcg	gccaggccgt	300
atgacacaaa	agagtagaaa	acgagatctc	aaatatctcg	aggcctgtcc	tctatacaac	360
cgcccagctc	tctgacaaa	ctccagaacg	gttgtctttt	gtttcgaaaa	gccaagggtcc	420
cttataaattg	ccctccattt	tgtgtcacct	atttaagcaa	aaaattgaaa	gtttactaac	480
ctttcattaa	agagaaataa	caatattata	aaaagcgctt	aaagctcaca	cgcgccagg	540
gggagccgag	ctcctcgaga	agttaagatt	atatgaataa	ctaaatacta	aatagaaatg	600
taaatacagt	gagaacaaaa	caaaaaaaa	cgaacagaga	aactaaatcc	acattaatgt	660
agagtcttat	ctattagaaa	atgcaaaactc	caactaaatg	ggaaaacaga	taacctcttt	720
tatttttttt	taatgtttga	tattcgagtc	tttttctttt	gttaggttta	tattcatcat	780
ttcaatgaat	aaaagaagct	tcttattttg	gttgcaaa	atgaaaaaaa	aggatttttt	840
catacttcta	aagcttcaat	tataaccaa	aattttataa	atgaagagaa	aaaatctagt	900
agtatcaagt	taaacctatt	cctttgccct	cggacgagtg	ctggggcgct	ggtttccact	960
atcggcgagt	acttctacac	agccatcggt	ccagacggcc	gcgcttctgc	ggcgattttg	1020
tgtacgccc	acagtcccgg	ctccggatcg	gacgattg	tcgcatcgac	cctgcgcca	1080
agctgcatca	tcgaaattgc	cgtaaccaaa	gctctgatag	agttggtcaa	gaccaatg	1140
gagcatatac	gcccggagcc	gcggcgatcc	tgcaagctcc	ggatgcctcc	gctcgaagta	1200
gcgctctgc	tgctccatac	aagccaacca	cggcctccag	agaagatgt	tggcgacctc	1260
gtattgggaa	tccccgaaca	tcgcctcgct	ccagtcaatg	accgctgtta	tcgcccatt	1320
gtccgtcagg	acattgtttg	agccgaaatc	cgctgacg	aggtgcccga	cttcggggca	1380
gtcctcgcc	caaagcatca	gctcatcgag	agcctgcg	acggacgcac	tgacggtgtc	1440
gtccatcaca	gtttgccagt	gatacacatg	gggatcagca	atcgcgcata	tgaaatcacg	1500
ccatgtagt	tattgaccga	ttccttgccg	tccgaatggg	ccgaaccgc	tcgtctggct	1560
aagatcgcc	gcagcgatcg	catccatggc	ctccgcgacc	ggctgcagaa	cagcgggcag	1620
ttcggtttca	ggcaggctct	gcaacgtgac	acctgtgca	cggcgggaga	tgcaataggt	1680
caggctctcg	ctgaattccc	caatgtcaag	cacttccgga	atcgggagcg	cggccgatgc	1740
aaagtgccga	taaacataac	gatctttgta	gaaaccatcg	gcgcagctat	ttaccgcgag	1800
gacatatcca	cgccctccta	catcgaagct	gaaagcacga	gattcttcgc	cctccgagag	1860
ctgcatcagg	tcggagacgc	tgctgaactt	ttcgatcaga	aacttctcga	cagacgtcgc	1920
ggtgagttca	ggctttttca	tttttaattg	tacttctctt	gcagttagg	aactataatg	1980
taactcaaaa	taagattaaa	caaactaaaa	taaaaagaag	ttatacagaa	aaaccatat	2040
aaaccagtac	taatccataa	taataataca	caaaaaaact	atcaaataaa	accagaaaac	2100
agattgaata	gaaaaatttt	ttcgatctcc	ttttatatcc	aaaattcgat	atatgaaaaa	2160
gggaactctc	agaaaatcac	caaatcaatt	taattagatt	tttcttttcc	ttctagcggt	2220
ggaaagaaaa	atttttcttt	ttttttttag	aaatgaaaaa	tttttgccgt	aggaatcacc	2280
gtataaaccc	tgtataaacg	ctactctggt	cacctgtgta	ggctatgatt	gaccagtggt	2340
tcattgttat	tgcgagagag	cgggagaaaa	gaaccgatac	aagagatcca	tgctggtata	2400
gttgtctgtc	caacactttg	atgaacttgt	aggacgatga	tgtgtattac	tagtgtcgac	2460
gctcgtccaa	cgccggcgga	cctcttttaa	ttctgtgtga	acccgtacat	gccccaaaata	2520
ggggcggggt	tacacagaat	atataacatc	gtaggtgtct	gggtgaacag	tttattcctg	2580
gcattccacta	aatataatgg	agcccgcttt	ttaagctggc	atccagaaaa	aaaaagaatc	2640
ccagcaccaa	aatattgttt	tcttcaccaa	ccatcagttc	ataggtccat	tccttctgag	2700
caactacaga	gaacaggggc	acaaacaggc	aaaaaacggg	cacaacctca	atggagtgtg	2760
gcaacctgcc	tggagtaaat	gatgacacaa	ggcaattgac	ccacgcagtg	atctatctca	2820
ttttcttaca	ccttctatta	ccttctgtct	tctctgattt	ggaaaaagct	gaaaaaaaag	2880
gttgaaacca	gttccctgaa	attattcccc	tacttgacta	ataagtatat	aaagacggta	2940
ggtattgatt	gtaattctgt	aaatctattt	cttaaacttc	ttaaattcta	cttttatagg	3000
tagtcttttt	tttagtttta	aaacaccaag	aacttagttt	cgatccccgc	gtgcttggtc	3060
ggccgtatga	cagatgtagt	aatagtatcc	gccgcaagaa	cagcagttgg	aaagtgttga	3120
ggctctcttg	caaagattcc	agcccctgaa	ttaggagctg	ttgttataaa	agccgcactt	3180
gaaagggcag	gtgtgaagcc	tgaacaagtc	agtgaagtca	taatgggtca	agttttaact	3240
gccggctcag	gtcaaaaccc	agccagacag	gctgctatta	aaagctggtt	accggcaatg	3300
gttcccgcga	tgactattaa	caaagtttgt	ggttccggcc	ttaaagcagt	gatgttagct	3360
gctaacgcaa	taatggctgg	ggatgctgaa	atagtatgct	ccggaggaca	agagaatatg	3420
agtgcagccc	cacacgtttt	accgggctcc	agagatggat	tccgtatggg	tgacgctaag	3480
ttagttgata	ctatgatagt	agatggacta	tgggatgtct	ataaccaata	tcacatgggt	3540
attacagccg	aaaacgtggc	gaaagaatat	gggattacga	gagaagcaca	ggatgagttc	3600
gccgtgggta	gtcaaaaata	ggcggaggcg	gctcaaaaag	ccggtaaatt	tgatgaggaa	3660
atagtactctg	tccttatacc	acagagaaaa	ggagatccgg	ttgcctttta	aaccgatag	3720
ttgtcgagac	aaggcgccac	attagacagc	atgtctggtt	tgaaacctgc	ttttgataag	3780
gccgggaccg	tgaccgctgc	taatgcgtca	ggactaaacg	atggagctgc	ggcgggtggt	3840
gttatgtctg	ctgctaaagc	aaaagaatta	gggttaactc	cattagccac	tatcaaatct	3900

[0005]

tatgctaacg	cgggggtgga	cccaaaagtg	atgggaatgg	gacctgttcc	agccagtaag	3960
agggcggttat	ctagggccga	atggactcct	caagacttgg	atttaatgga	aattaatgaa	4020
gcatttgccg	cacaggcggt	agctgtccac	caacagatgg	gttgggatac	aagtaagggtc	4080
aatgttaatg	gaggtgcaat	cgccattgggt	cacccaattg	gtgctgccc	atgtagaatt	4140
ttagttaccc	tactgcatga	gatgaagagg	cgtgatgcaa	agaaaggctt	agcttcggtt	4200
tgtatcggtg	gtggaatggg	tgtggcatta	gcagtcgagc	gtaaataaaa	cctgcaggcc	4260
gcgagcgccg	attaagtga	tttactttta	atcttgcatt	taaataaatt	ttctttttat	4320
agctttatga	cttagtttca	atztatatac	tatttttaatg	acattttcga	ttcattgatt	4380
gaaagctttg	tgttttttct	tgatgcgcta	ttgcattggt	cttgtctttt	tcgccacatg	4440
taatatctgt	agtagatacc	tgatacattg	tggatgctga	gtgaaatttt	agttaataat	4500
ggaggcgctc	tttaataatt	tggggatatt	ggcttaacgc	gatcgccgac	gccgcccatt	4560
gcagcagga	ttatcgtaat	acgtaatagt	tgaaaatctc	aaaaatgtgt	gggtcattac	4620
gtaaataatg	ataggaatgg	gattctttcta	tttttccttt	ttccattcta	gcagccgtcg	4680
ggaaaacgtg	gcacctctc	tttcgggctc	aattggagtc	acgctgccgt	gagcatcctc	4740
tctttccata	tctaacaact	gagcacgtaa	ccaatggaaa	agcatgagct	tagcgttgct	4800
ccaaaaaagt	attggatggg	taataccatt	tgtctgttct	cttctgactt	tgactcctca	4860
aaaaaaaaaa	atctacaatc	aacagatcgc	ttcaattacg	ccctcacaaa	aacttttttc	4920
cttcttcttc	gcccacgtta	aattttatcc	ctcatgttgt	ctaacggatt	tctgcacttg	4980
atztatattata	aaaagacaaa	gacataatac	ttctctatca	atttcagtta	ttgttcttcc	5040
ttgctgttatt	cttctgttct	tcttttttct	ttgtcatata	taaccaaggc	ggccgctggc	5100
gagggagata	tgaactctc	aactaaactt	tgttgggtgtg	gtattaaagg	aagacttagg	5160
ccgcaaaagc	aacaacaatt	acacaataca	aacttgcaaa	tgactgaact	aaaaaaacaa	5220
aagaccgctg	aacaaaaaac	cagacctcaa	aatgtcggta	ttaaagggtat	ccaaattttac	5280
atcccaactc	aatgtgtcaa	ccaatctgag	ctagagaaat	ttgatggcgt	ttctcaaggt	5340
aaatacacaa	ttggtctggg	ccaaaccaac	atgtcttttg	tcaatgacag	agaagataatc	5400
tactcgatgt	ccctaactgt	tttgtctaag	ttgatcaaga	gttacaacat	cgacaccaac	5460
aaaattggta	gattagaagt	cggtagtgaa	actctgattg	acaagtccaa	gtctgtcaag	5520
tctgtcttga	tgcaattggt	tgggtgaaaac	actgacgtcg	aaggatttga	cacgcttaatt	5580
gcctgttacg	gtggtaccaa	cgctgtgttc	aactctttga	actggattga	atctaacgca	5640
tgggatggta	gagacgccat	tgtagtgtgc	ggtgatattg	ccatctacga	taagggtgcc	5700
gcaagaccaa	ccggtggtgc	cggtactggt	gctatgtgga	tcggtcctga	tgctccaatt	5760
gtatttgact	ctgtaagagc	ttctttacatg	gaacacgcct	acgattttta	caagccagat	5820
ttcaccagcg	aatactctta	cgtcgatggg	catttttcat	taacttggtta	cgtaacaggct	5880
cttgatcaag	tttacaagag	ttattccaag	aaggctattt	ctaaagggtt	ggttagcgat	5940
cccgttggtt	cggatgcttt	gaacgttttg	aaatatttcg	actacaacgt	tttccatggt	6000
ccaacctgta	aattggtcac	aaaatcatat	ggttagattac	tatataacga	tttcagagcc	6060
aatcctcaat	tgttcccaga	agttgacgcc	gaattagcta	ctcgcgatta	tgacgaatct	6120
taaccgata	agaacattga	aaaaactttt	gttaattgtg	ctaagccatt	ccacaaagag	6180
agagttgccc	aatactttgat	tgttccaaca	aacacaggta	acatgtacac	cgcatctggt	6240
tatgccgcct	ttgcatctct	attaaactat	gttggatctg	acgacttaca	aggcaagcgt	6300
gttgggttat	tttcttacgg	ttccgggtta	gctgcatctc	tatattcttg	caaaattggt	6360
ggtgacgtcc	aacatattat	caaggaatta	gatattacta	acaaattagc	caagagaatc	6420
accgaaactc	caaaggatta	cgaagctgcc	atcgaattga	gagaaaatgc	ccatttgaag	6480
aagaacttca	aactcaagg	ttccattgag	cattttgcaa	gtggtgttta	ctactttgac	6540
aacatcgatg	acaaatttag	aagatcttac	gatgttaaaa	aataatcttc	ccccatcgat	6600
tgcattcttg	tgaaccccct	tcataaatgc	tttatttttt	tggcagcctg	cttttttttag	6660
ctctcattta	atagagtagt	tttttaattct	atatactagg	aaaactcttt	atttaataac	6720
aatgatataat	atataatatt	tttttataaa	gaattgtata	tctatattta	taacacaata	6780
aatctaattct	caactttttt	cttttaaagtt	aagcccaacc	gatttttttt	ctcataaggc	6840
gcgccacggt	cgtgcggatt	aaagcttttg	attaagcctt	ctagtccaaa	aaacacgttt	6900
ttttgtcatt	tattttcattt	tcttagaata	gttttagttta	ttcattttat	agtcacgaat	6960
gttttatgat	tctatatagg	gttgcaaaca	agcatttttc	attttatggt	aaaacaattt	7020
caggtttacc	ttttattctg	cttgtggtga	cgcgtgtatc	cgcccgtctt	tttggtcacc	7080
catgtattta	attgcataaa	taattcttaa	aagtggagct	agtcagcccc	tcagccccc	7140
tagcgtcgat	aaactaatga	ttttaaatcg	ttaaaaaaat	atgcgaattc	tgtggatcga	7200
acacaggacc	tccagataac	ttgaccgaag	ttttttcttc	agtctggcgc	tctcccaact	7260
gagctaaatc	cgcttactat	ttgttatcag	ttcccttcat	atctacatag	aatagggttaa	7320
gtattttatt	agttgccaga	agaactactg	atagttggga	atatttggtg	aataatgaag	7380
attgggtgaa	taatttgata	attttgagat	tcaattgtta	atcaatgtta	caatattatg	7440
tatacagagt	atactagaag	ttctcttcgg	agatcttgaa	gttcacaaaa	gggaatcgat	7500
atctctacat	aatattatca	ttacttcttc	cccatcttat	atctgtcatt	cattattgat	7560
tatgatcaat	gcaataatga	ttggtagttg	ccaaacattt	aatacgaatc	tctgtaatat	7620
ttctatgaat	aattatcaca	gcaacgttca	attatcttca	attccggtgt	ttaaacccca	7680
gcgcctggcg	gg					7692

<210> 13

<211> 3627

[0006]

<212> DNA

<213> 人工序列

<220>

<223> 载体pAM1466

<400> 13

```

gcacttttcg gggaaatgtg cgcggaaccc ctatttgttt atttttctaa atacattcaa 60
atatgtatcc gctcatgaga caataaccct gataaatgct tcaataatat tgaaaaagga 120
agagtatgag tattcaacat ttccgtgtcg cccttattcc cttttttgcg gcattttgcc 180
ttcctgtttt tgctcaccca gaaacgctgg tgaaagttaa agatgctgaa gatcagttgg 240
gtgcacgagt ggggttacatc gaactggatc tcaacagcgg taagatcctt gagagttttc 300
gccccgaaga acgtttttcca atgatgagca cttttaagt tctgctatgt ggcgcggtat 360
tatcccgat tgacgcggg caagagcaac tcggtcgccg catacactat tctcagaatg 420
acttggttga gtactcacca gtcacagaaa agcatcttac ggatggcatg acagtaagag 480
aattatgcag tgctgccata accatgagtg ataacactgc ggccaactta cttctgacaa 540
cgatcggagg accgaaggag ctaaccgctt ttttgcacaa catgggggat catgtaactc 600
gccttgatcg ttgggaaccg gagctgaatg aagccatacc aaacgacgag cgtgacacca 660
cgatgcctgt agcaatggca acaacgttgc gcaaactatt aactggcgaa ctacttactc 720
tagcttcccc gcaacaatta atagactgga tggaggcgga taaagttgca ggaccacttc 780
tgcgctcgcg ccttccggct ggctggttta ttgctgataa atctggagcc ggtgagcgtg 840
ggcttcgcyg tatcattgca gcaactgggc cagatggtta gccctcccgt atcgtagtta 900
tctacacgac ggggagtcag gcaactatgg atgaacgaaa tagacagatc gctgagatag 960
gtgcctcact gattaagcat tggttaactgt cagaccaagt ttactcatat atactttaga 1020
ttgattttaa acttcatttt taattttaaaa ggatctaggt gaagatcctt tttgataatc 1080
tcatgaccaa aatcccttaa cgtgagtttt cgttccactg agcgtcagac cccgtagaaa 1140
agatcaaagg atcttcttga gatccttttt ttctgcgctg aatctgctgc ttgcaaacaa 1200
aaaaaccacc gctaccagcg gtggtttgtt tgccggatca agagctacca actctttttc 1260
cgaaggtaac tggcttcagc agagcgcaga taccaaatac tgtccttcta gtgtagccgt 1320
agttaggcca ccacttcaag aactctgtag caccgcctac atacctcgct ctgctaattc 1380
tgttaccagt ggctgtctgc agtggcgata agtcgtgtct taccgggttg gactcaagac 1440
gatatgtacc ggataaggcg cagcggctcg gctgaacggg gggttcgtgc acacagccca 1500
gcttgagcga aacgacctac accgaactga gatacctaca gcgtgagcta tgagaaagcg 1560
ccacgcttcc cgaagggaga aaggcggaca ggtatccggt aagcggcagg gtcggaacag 1620
gagagcgcac gagggagctt ccagggggaa acgcctggta tctttatagt cctgtcgggt 1680
ttcgccacct ctgacttgag cgtcgatttt tgtgatgtc gtcagggggg cggagcctat 1740
ggaaaaacgc cagcaacgcg gcctttttac ggttccctgg cttttgctgg ctttttgctc 1800
acatgttctt tcctgcgtta tcccctgatt ctgtggataa ccgtattacc gcctttgagt 1860
gagctgatac cgctcgccgc agccgaacga ccgagcggcc gccagcgagc tcagtgagcg 1920
aggaagcgga agagcgccca atacgcaaac cgctctccc cgcgcttg cggattcatt 1980
aatgcagctg gcacgacagg tttcccgact ggaaagcggg cagtgcgcgc aacgcaatta 2040
atgtgagtta gctcactcat taggcacccc aggccttaca ctttatgctt ccggctcgta 2100
tgtgtgtggg aattctgagc ggataacaat ttacacagg aaacagctat gacctgagt 2160
ttaaacatga gcaaaggcga agaactgttc acgggcgttg taccgatcct ggtggaactg 2220
gacggggatg tgaatgggca caagttttca gtgagcggcg aaggagaagg cgatgcgacc 2280
tacggcaaac tgaccctgaa attcattttg accaccggta aacttcccgt gccgtggccc 2340
accctggtga ccaccttttg ctatggcgta cagtgtctcg cgcgttacc ggatcacatg 2400
aaacgccaag acttctttta gagcgctatg ccagagggt acgtccagga acgaccata 2460
ttcttcaaag acgacggcaa ctacaagac cgcgctgaag tcaagtttga agggacacg 2520
ctggtgaacc gtattgagct gaagggcatc gacttcaagg aggacgggaa catcctgggc 2580
cataagctgg agtacaatta caacagccac aacgtgtata tcatggcgga caagcagaag 2640
aacggcatca aggtcaactt caagatccgg cacaacatcg aggatggcag cgtgcagctg 2700
gcggatcatt atcaacagaa cccccgatt ggcgatggac cggctgtgct gcccgataat 2760
cattacctga gtaccagag cgccctgagc aaggaccga atgagaagcg tgatcacatg 2820
gtactgtctg aatttgtgac cgcggttggc atcaccacg gcatggatga actgtataaa 2880
taaggtaccg cggccgcccgt ctccggggac agacgtctcg ccccttaggg tccatgcagt 2940
tggcttcgat ggtctctttt ttataggtcg agtaccaatt cgccctatag tgagtctgat 3000
tacaattcac tggccgtcgt tttacaacgt cgtgactggg aaaaccctgg cgttacccaa 3060
cttaatcgcc ttgcagcaca tcccctttc gccagctggc gtaatagcga agaggccgc 3120
accgatcgcc cttcccaaca gttgcgcagc ctgaatggcg aatgggacgc gccctgtagc 3180
ggcgcattaa gcgcggcggg tgtggtggtt acgcgcagcg tgaccgctac acttgccagc 3240
gccctagcgc ccgctccttt cgctttcttc ccttcttttc tcgccacgtt cgccggtttt 3300
ccccgtcaag ctctaaatcg ggggtccctt ttagggttcc gatttagtgc tttacggcac 3360
ctcgaaccca aaaaacttga ttagggatg ggttccagta gtgggccatc gccctgatg 3420
acggtttttc gccctttgac gttggagtc acgttcttta atagtggact cttgttccaa 3480
actggaacaa cactcaacc tatctcggtc tattcttttg atttataagg gattttgccg 3540
atttcggcct attggttaaa aaatgagctg atttaacaaa aatttaacgc gaattttaac 3600

```

[0007]

aaaatatttaa cgcttacaat ttaggtg

3627

<210> 14
<211> 915
<212> DNA
<213> 人工序列

<220>
<223> 合成的DNA片段1040

<400> 14
taagtagttg accatagcta tggaagggtct cacggaaatg ttgaatactc atcaattgcc 60
tttttcaata ttattgaagc atttatcagg gttattgtct catgagcggg tacatatattg 120
aatgtatttta gaaaaataaa caaatagggg ttccgcgcac atttccccga aaagtgccac 180
ctgacgtcta agaaaccatt attatcatga cattaaccta taaaaatagg cgtatcacga 240
ggcccttttca tctcgcgcgt ttccggtgatg acggtgaaaa cctctgacac atgcagctcc 300
cggagacagt cacagcttgt ctgtaagcgg atgccgggag cagacaagcc cgtcagggcg 360
cgtcagcggg tgttgccggg tgtcggggct ggtaaaacga cggccagtat taaccctcac 420
taaagggaac tcgaggctct tcacgctcgt ccaacgccgg cggaccttgg atgactcgt 480
agccgcccac tacgcaaac gcctctcccc gcgcgttggc cgattcatta atgcagctgg 540
cacgacaggt ttcccgactg gaaagcgggc agtgagcgca acgcaattaa tgtgagttag 600
ctcactcatt aggcacccca ggcttttacac tttatgcttc cggctcgtat gttgtgtgga 660
attgtgagcg gataacaatt tcacacagga aacagctatg agcaaaggcg aagaactgtt 720
cacgggcgtt gtaccgatcc tgggtggaact ggacggggat gtgaatgggc acaagttttc 780
agttagcggc gaaggagaag gcgatgcgac ctacggcaaa ctgaccctga aattcatttg 840
caccaccggt aaacttcccc tgccgtggcc caccctggtg accaccttg gagacgccac 900
gtacatggct tattg 915

<210> 15
<211> 882
<212> DNA
<213> 人工序列

<220>
<223> 合成的DNA片段1041

<400> 15
taagtagttg accatagcta tggaagggtct ccctttggct atggcgtaca gtgcttcgcg 60
cgttaccggg atcacatgaa acgccacgac ttctttaaga gcgctatgcc agagggctac 120
gtccaggaac gcaccatatt cttcaaagac gacggcaact acaagacgcg cgctgaagtc 180
aagtttgaag gggacacgct ggtgaaccgt attgagctga agggcatcga cttcaaggag 240
gacgggaaca tcctgggcca taagctggag tacaattaca acagccacaa cgtgtatatc 300
atggcggaca agcagaagaa cggcatcaag gtcaacttca agatccggca caacatcgag 360
gatggcagcg tgcagctggc ggatcattat caacagaaca ccccgattgg cgtatggaccg 420
gtgctgctgc ccgataatca ttacctgagt acccagagcg ccctgagcaa ggacccgaat 480
gagaagcgtg atcacatggt actgctggaa tttgtgaccg cggctggcat caccacggc 540
atggatgaac tgtataaata acatatggag tctgccatcg gtgttttaac cccagcgcct 600
ggcgggtgaa gagcgagctc ccgctgagca ataactagcg tcatagctgt ttcctgggtc 660
gttcggctgc ggcgagcggg atcagctcac tcaaaggcgg taatacgggt atccacagaa 720
tcaggggata acgcaggaaa gaacatgtga gcaaaaggcc agcaaaaggc caggaaccgt 780
aaaaaggccg cgttgctggc gtttttccat aggtccgccc cccctgacga gcatcacaaa 840
aatcgacgct caagtcagag acgccacgta catggcttat tg 882

<210> 16
<211> 787
<212> DNA
<213> 人工序列

<220>
<223> 合成的DNA片段1042

<400> 16
taagtagttg accatagcta tggaagggtct caagtcagag gtggcgaaac ccgacaggac 60
tataaagata ccaggcgttt ccccttgga gctccctcgt gcgctctcct gttccgacct 120
tgccgcttac ccgatacctg tccgcctttc tcccttcggg aagcgtggcg ctttctcata 180
gctcacgctg taggtatctc agttcgggtg aggtcgttcg ctccaagctg ggctgtgtgc 240
acgaaccccc cgttcagccc gaccgctgcg ccttatccgg taactatcgt cttgtgtcca 300

[0008]

```

acccggttaag acacgactta tcgccactgg cagcagccac tggtaacagg attagcagag 360
cgaggtatgt aggcggtgct acagagttct tgaagtgggt gcctaactac ggctacacta 420
gaagaacagt atttggtatc tgcgctctgc tgaagccagt taccttcgga aaaagagttg 480
gtagctcttg atccggcaaa caaaccaccg ctggtagcgg tggttttttt gtttgcaagc 540
agcagattac gcgcagaaaa aaaggatctc aagaagatcc tttgatcttt tctacggggt 600
ctgacgctca gtggaacgaa aactcacgtt aagggatttt ggtcatgaga ttatcaaaaa 660
ggatcttcac ctagatcctt ttaaattaaa aatgaagttt taaatcaatc taaagtatat 720
atgagtaaac ttggtcgcac gcttaccaat gcttaatcag tggagacgcc acgtacatgg 780
cttattg 787

```

<210> 17

<211> 36

<212> DNA

<213> 人工序列

<220>

<223> 引物 JCB158-17A

<400> 17

cgttcatcca tagttgcctg actgcccgtc gtgtag 36

<210> 18

<211> 36

<212> DNA

<213> 人工序列

<220>

<223> 引物 JCB158-17B

<400> 18

ctacacgacg ggcagtcagg caactatgga tgaacg 36

<210> 19

<211> 59

<212> DNA

<213> 人工序列

<220>

<223> 引物 JCB158-17C

<400> 19

atatgagtaa acttggtcgc atgcttacca atgcttaatc agtgaggcac ctatctcag 59

<210> 20

<211> 52

<212> DNA

<213> 人工序列

<220>

<223> 引物 JCB158-17D

<400> 20

ttgaaaaagg caattgatga gtattcaaca tttccgtgtc gcccttattc cc 52

<210> 21

<211> 39

<212> DNA

<213> 人工序列

<220>

<223> 引物 L012

<400> 21

ccggttaacta tcgtcttgat tccaacccgg taagacacg 39

<210> 22

<211> 39

[0009]

<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 L013	
<400> 22	
cgtgtctttac cgggttggaa tcaagacgat agttaccgg	39
<210> 23	
<211> 38	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 L054	
<400> 23	
cgaaactaag ttcttgggtgt tttaaaacta aaaaaaag	38
<210> 24	
<211> 35	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 L057	
<400> 24	
ggttatatat gacaaaagaa aaagaagaac agaag	35
<210> 25	
<211> 29	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 L109	
<400> 25	
atgaaactct ctactaaact ttgttggtg	29
<210> 26	
<211> 28	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 L110	
<400> 26	
atgagaaaaa aaatcggttg ggcttaac	28
<210> 27	
<211> 31	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 L176	
<400> 27	
gactagctcc acttttaaga attatttatg c	31
<210> 28	
<211> 38	
<212> DNA	

[0010]

<213> 人工序列	
<220>	
<223> 引物 L185	
<400> 28 gtgaatttac tttaaattctt gcattttaaat aaatttttc	38
<210> 29	
<211> 30	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 L186	
<400> 29 aagccaatat ccccaaaatt attaagagcg	30
<210> 30	
<211> 54	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 VH296-235-55-Leu2 12-1 F	
<400> 30 gctcacacgc ggccaggggg agcccgttga gccattagta tcaatttgct tacc	54
<210> 31	
<211> 57	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 VH296-235-55-Leu2 12-1 R	
<400> 31 agggtccgccg gcgttggacg agcgaggcgc ctgattcaag aaatatcttg accgcag	57
<210> 32	
<211> 27	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 L219	
<400> 32 ctccaagctg acataaatcg cactttg	27
<210> 33	
<211> 39	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 L220	
<400> 33 tttaagcgct ttttataata ttgttatttc tctttaatg	39
<210> 34	
<211> 37	
<212> DNA	
<213> 人工序列	

[0011]

<220>		
<223>	引物 L221	
<400>	34	
	ataaactaat gatttttaat cgttaaaaaa atatgcg	37
<210>	35	
<211>	28	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物 L222	
<400>	35	
	gaattgaaga taattgaacg ttgctgtg	28
<210>	36	
<211>	28	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物 L224	
<400>	36	
	cttttaattc tgctgtaacc cgtacatg	28
<210>	37	
<211>	24	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物 L225	
<400>	37	
	tgacagcagg attatcgtaa tacg	24
<210>	38	
<211>	27	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物 L226	
<400>	38	
	atgacagatg tagtaatagt atccgcc	27
<210>	39	
<211>	27	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物 L227	
<400>	39	
	ttatttacgc tcgactgcta atgccac	27
<210>	40	
<211>	26	
<212>	DNA	
<213>	人工序列	

[0012]

<220>	
<223> 引物 L229	
<400> 40	
atgaaaaatt gtgtcatcgt cagtgc	26
<210> 41	
<211> 32	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 L230	
<400> 41	
ttaattcaat ctttcaatca ccatcgcaat tc	32
<210> 42	
<211> 26	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 L235	
<400> 42	
atgaccatcg gtattgacaa gattag	26
<210> 43	
<211> 38	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 L236	
<400> 43	
ttaattacga taagatctta cagtattggt aatcgag	38
<210> 44	
<211> 35	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 L248	
<400> 44	
taaagctttt gattaagcct tctagtccaa aaaac	35
<210> 45	
<211> 28	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 S000	
<400> 45	
gacggcacgg ccacgcgttt aaaccgcc	28
<210> 46	
<211> 28	
<212> DNA	
<213> 人工序列	
<220>	

[0013]

<223> 引物 S001		
<400> 46		
ggcggtttaa acgcgtggcc gtgccgtc	28	
<210> 47		
<211> 24		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物 S002		
<400> 47		
gctcacacgc ggccaggggg agcc	24	
<210> 48		
<211> 24		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物 S003		
<400> 48		
ggctccccct ggccgcgtgt gagc	24	
<210> 49		
<211> 24		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物 S004		
<400> 49		
cgctcgtcca acgccggcgg acct	24	
<210> 50		
<211> 24		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物 S005		
<400> 50		
aggtccgccg gcgttgacg agcg	24	
<210> 51		
<211> 24		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物 S006		
<400> 51		
atccccgcgt gcttgccgg ccgt	24	
<210> 52		
<211> 24		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物 S007		

[0014]

<400> 52 acggccggcc aagcacgcgg ggat	24
<210> 53 <211> 24 <212> DNA <213> 人工序列	
<220> <223> 引物 s008	
<400> 53 aacctgcagg ccgcgagcgc cgat	24
<210> 54 <211> 24 <212> DNA <213> 人工序列	
<220> <223> 引物 s009	
<400> 54 atcggcgctc gcggcctgca ggtt	24
<210> 55 <211> 24 <212> DNA <213> 人工序列	
<220> <223> 引物 s010	
<400> 55 aacgcgatcg ccgacgccgc cgat	24
<210> 56 <211> 24 <212> DNA <213> 人工序列	
<220> <223> 引物 s011	
<400> 56 atcggcggcg tcggcgatcg cgtt	24
<210> 57 <211> 24 <212> DNA <213> 人工序列	
<220> <223> 引物 s012	
<400> 57 aaggcggccg ctggcgaggg agat	24
<210> 58 <211> 24 <212> DNA <213> 人工序列	
<220> <223> 引物 s013	

[0015]

<400> 58 atctccctcg ccagcggccg cctt	24
<210> 59 <211> 24 <212> DNA <213> 人工序列	
<220> <223> 引物 S014	
<400> 59 aaggcgcgcc acggtcgtgc ggat	24
<210> 60 <211> 24 <212> DNA <213> 人工序列	
<220> <223> 引物 S015	
<400> 60 atccgcacga ccgtggcgcg cctt	24
<210> 61 <211> 24 <212> DNA <213> 人工序列	
<220> <223> 引物 S016	
<400> 61 agcccctcag cccccctagc gtcg	24
<210> 62 <211> 24 <212> DNA <213> 人工序列	
<220> <223> 引物 S017	
<400> 62 cgacgctagg ggggctgagg ggct	24
<210> 63 <211> 28 <212> DNA <213> 人工序列	
<220> <223> 引物 S018	
<400> 63 cggtgtttta accccagcgc ctggcggg	28
<210> 64 <211> 28 <212> DNA <213> 人工序列	
<220> <223> 引物 S019	
<400> 64	

[0016]

cccgccaggc gctggggttt aaacaccg	28
<210> 65	
<211> 37	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 S027	
<400> 65	
tggtatgactc gctagcgcg aacgcaatta atgtgag	37
<210> 66	
<211> 39	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 S028	
<400> 66	
tggcagactc catatgctat gcggcatcag agcagattg	39
<210> 67	
<211> 42	
<212> DNA	
<213> 人工序列e	
<220>	
<223> 引物 S036	
<400> 67	
taccgggga tcctctagcg tcgacctgca ggcattgcaag ct	42
<210> 68	
<211> 42	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 S037	
<400> 68	
agcttgcatg cctgcaggtc gacgctagag gatccccggg ta	42
<210> 69	
<211> 54	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 J036	
<400> 69	
gccgcaaaaa aggggaataag ggcgacacgg aaatgttgaa tactcatcaa ttgc	54
<210> 70	
<211> 42	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 J037	
<400> 70	
cggttttcgc cacctctgac ttgagcgctg atttttgtga tg	42

[0017]

<210> 71
<211> 42
<212> DNA
<213> 人工序列

<220>
<223> 引物 J038

<400> 71
catcacaaaa atcgacgctc aagtcagagg tggcgaaacc cg 42

<210> 72
<211> 55
<212> DNA
<213> 人工序列

<220>
<223> 引物 J039

<400> 72
ctgagatagg tgcctcactg attaagcatt ggtaagcatg cgaccaagtt tactc 55

<210> 73
<211> 58
<212> DNA
<213> 人工序列

<220>
<223> 引物 J018

<400> 73
gacggcacgg ccacgcgttt aaaccgcctt ggatggatac gctagccgcc caatacgc 58

<210> 74
<211> 58
<212> DNA
<213> 人工序列

<220>
<223> 引物 J019

<400> 74
gctcacacgc ggccaggggg agccttggat ggatggatac gctagccgcc caatacgc 58

<210> 75
<211> 57
<212> DNA
<213> 人工序列

<220>
<223> 引物 J020

<400> 75
cgctcgtcca acgccggcgg accttggatg gatggatacg ctagccgccc aatacgc 57

<210> 76
<211> 57
<212> DNA
<213> 人工序列

<220>
<223> 引物 J021

<400> 76
atccccgcgt gcttggccgg ccgttggatg gatggatacg ctagccgccc aatacgc 57

[0018]

<210> 77
<211> 57
<212> DNA
<213> 人工序列

<220>
<223> 引物 J022

<400> 77
aacctgcagg ccgcgagcgc cgattggatg gatggatacg ctagccgccc aatacgc 57

<210> 78
<211> 57
<212> DNA
<213> 人工序列

<220>
<223> 引物 J023

<400> 78
aacgcgatcg ccgacgccgc cgattggatg gatggatacg ctagccgccc aatacgc 57

<210> 79
<211> 57
<212> DNA
<213> 人工序列

<220>
<223> 引物 J024

<400> 79
aaggcggccg ctggcgaggg agattggatg gatggatacg ctagccgccc aatacgc 57

<210> 80
<211> 56
<212> DNA
<213> 人工序列

<220>
<223> 引物 J025

<400> 80
aaggcgcgcc acggtcgtgc ggatggatgg atggatacgc tagccgcca atacgc 56

<210> 81
<211> 58
<212> DNA
<213> 人工序列

<220>
<223> 引物 J026

<400> 81
agccccctcag cccccctagc gtcgttggat ggatggatac gctagccgcc caatacgc 58

<210> 82
<211> 56
<212> DNA
<213> 人工序列

<220>
<223> 引物 J029

<400> 82
acggccggcc aagcacgcgg ggattggcag gataccatat gttatttata cagttc 56

<210> 83

[0019]

<211> 56
<212> DNA
<213> 人工序列

<220>
<223> 引物 J030

<400> 83
atcggcgctc gcggcctgca ggtttggcag gataccatat gttatttata cagttc 56

<210> 84
<211> 56
<212> DNA
<213> 人工序列

<220>
<223> 引物 J031

<400> 84
atcggcggcg tcggcgatcg cgtttggcag gataccatat gttatttata cagttc 56

<210> 85
<211> 56
<212> DNA
<213> 人工序列

<220>
<223> 引物 J032

<400> 85
atctccctcg ccagcggccg cctttggcag gataccatat gttatttata cagttc 56

<210> 86
<211> 56
<212> DNA
<213> 人工序列

<220>
<223> 引物 J033

<400> 86
atccgcacga ccgtggcgcg cctttggcag gataccatat gttatttata cagttc 56

<210> 87
<211> 56
<212> DNA
<213> 人工序列

<220>
<223> 引物 J034

<400> 87
cgacgctagg ggggctgagg ggcttggcag gataccatat gttatttata cagttc 56

<210> 88
<211> 50
<212> DNA
<213> 人工序列

<220>
<223> 引物 J055

<400> 88
aaagggaaact cgaggctctt cagacggcac ggccacgcgt ttaaaccgcc 50

<210> 89
<211> 46

[0020]

<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 J056	
<400> 89	
aaaggggaact cgaggctctt cagctcacac gcggccaggg ggagcc	46
<210> 90	
<211> 46	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 J057	
<400> 90	
aaaggggaact cgaggctctt cacgctcgtc caacgccggc ggacct	46
<210> 91	
<211> 46	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 J058	
<400> 91	
aaaggggaact cgaggctctt caatccccgc gtgcttgGCC ggccgt	46
<210> 92	
<211> 46	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 J059	
<400> 92	
aaaggggaact cgaggctctt caaacctgca ggccgcgagc gccgat	46
<210> 93	
<211> 46	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 J060	
<400> 93	
aaaggggaact cgaggctctt caaacgcat cgccgacgcc gccgat	46
<210> 94	
<211> 46	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 J061	
<400> 94	
aaaggggaact cgaggctctt caaaggcggc cgctggcgag ggagat	46
<210> 95	
<211> 46	
<212> DNA	

[0021]

<213> 人工序列	
<220>	
<223> 引物 J062	
<400> 95 aaaggggaact cgaggctctt caaaggcgcg ccacggtcgt gcggat	46
<210> 96 <211> 46 <212> DNA <213> 人工序列	
<220>	
<223> 引物 J063	
<400> 96 aaaggggaact cgaggctctt caagcccctc agcccccta gcgtcg	46
<210> 97 <211> 46 <212> DNA <213> 人工序列	
<220>	
<223> 引物 J064	
<400> 97 ctcagcgggga gctcgctctt caggctcccc ctggccgcgt gtgagc	46
<210> 98 <211> 46 <212> DNA <213> 人工序列	
<220>	
<223> 引物 J065	
<400> 98 ctcagcgggga gctcgctctt caagggtccgc cggcgttgga cgagcg	46
<210> 99 <211> 46 <212> DNA <213> 人工序列	
<220>	
<223> 引物 J066	
<400> 99 ctcagcgggga gctcgctctt caacggccgg ccaagcacgc ggggat	46
<210> 100 <211> 46 <212> DNA <213> 人工序列	
<220>	
<223> 引物 J067	
<400> 100 ctcagcgggga gctcgctctt caatcggcgc tcgcggcctg caggtt	46
<210> 101 <211> 46 <212> DNA <213> 人工序列	

[0022]

<220>	
<223> 引物 J068	
<400> 101	
ctcagcggga gctcgctctt caatcggcgg cgtcggcgat cgcgtt	46
<210> 102	
<211> 46	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 J069	
<400> 102	
ctcagcggga gctcgctctt caatctccct cgccagcggc cgcctt	46
<210> 103	
<211> 46	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 J070	
<400> 103	
ctcagcggga gctcgctctt caatccgcac gaccgtggcg cgcctt	46
<210> 104	
<211> 46	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 J071	
<400> 104	
ctcagcggga gctcgctctt cagcagccta ggggggctga ggggct	46
<210> 105	
<211> 50	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 J072	
<400> 105	
ctcagcggga gctcgctctt caccgccag gcgctggggt ttaaaccacg	50
<210> 106	
<211> 57	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 J073	
<400> 106	
ggctccccct ggccgcgtgt gagcttggca ggataccata tgttatttat acagttc	57
<210> 107	
<211> 57	
<212> DNA	
<213> 人工序列	

[0023]

<220>
 <223> 引物 J074

 <400> 107
 aggtccgccg gcggttgacg agcggttgca ggataccata tggtatttat acagttc 57

 <210> 108
 <211> 58
 <212> DNA
 <213> 人工序列

 <220>
 <223> 引物 J075

 <400> 108
 cccgccaggc gctgggggttt aaacaccgtt ggcaggatac catatgttat ttatacag 58

 <210> 109
 <211> 20
 <212> DNA
 <213> 人工序列

 <220>
 <223> 引物 K162

 <400> 109
 tgaagagcga gctcccgcgtg 20

 <210> 110
 <211> 24
 <212> DNA
 <213> 人工序列

 <220>
 <223> 引物 K163

 <400> 110
 tgaagagcct cgagttccct ttag 24

 <210> 111
 <211> 8139
 <212> DNA
 <213> 人工序列

 <220>
 <223> Phase I-A 缝合产物

 <400> 111
 gacggcacgg ccacgcgttt aaaccgccca gatggaatcc cttccataga gagaaggagc 60
 aagcaactga cccaatattg actgccactg gacctgaaga catgcaacaa agtgcaagca 120
 tagtggggcc ttcttccaat gctaattcgg tcaactgccac tgctgctacg gaaaaccaac 180
 ctaaagggtat taacttcttc actataagaa aatcacacga gcgcccgac gatgtctctg 240
 tttaaatggc gcaagttttc cgctttgtta tatatattta taccctttc ttctctccc 300
 tgcaatataa tagtttaatt ctaatatata taatatccta tattttcttc atttaccggc 360
 gcactctcgc ccgaacgacc tcaaaatgtc tgctacattc ataataacca aaagtcata 420
 actttttttt ttgaacctga atatataatc atcacatgtc actgctggc cttgccgacc 480
 agcgtatata atctcgatag ttggtttccc gttctttcca ctcccgtcgc tcacacgcgg 540
 ccagggggag ccgagctcct cgagaagtta agattatatg aataactaaa tactaaatag 600
 aaatgtaaat acagtgaaga caaaacaaaa aaaaacgaac agagaaacta aatccacatt 660
 aattgagagt tctatctatt agaaaatgca aactccaact aaatgggaaa acagataacc 720
 tcttttattt ttttttaatg tttgatattc gagtcttttt cttttgtag gtttatattc 780
 atcatttcaa tgaataaaaag aagcttctta ttttggttgc aaagaatgaa aaaaaaggat 840
 tttttcatat ttctaaagct tcaattataa ccaaaaattt tataaatgaa gagaaaaaat 900
 ctagttagat caagttaaac ctattccttt gccctcggac gagtgcctgg gcgtcggttt 960
 ccactatcgg cgagtaactt tacacagcca tcggtccaga cggccgcgct tctgcgggag 1020
 atttggtgac gcccgacagt cccggctccg gatcggacga ttgcgtcgca tcgaccctgc 1080
 gcccaagctg catcatcgaa attgccgtca accaagctct gatagagttg gtcaagacca 1140

[0024]

atgctggagca	tatacgcccg	gagccgaggc	gatacctgcaa	gctccggatg	cctccgctcg	1200
aagtagcgcg	tctgctgctc	catacaagcc	aaccacggcc	tccagaagaa	gatgttggcg	1260
acctcggtatt	gggaatcccc	gaacatcgcc	tcgctccagt	caatgaccgc	tgttatgctg	1320
ccattgtccg	tcaggacatt	gttggagccg	aaatccgcgt	gcacgaggtg	ccggacttcg	1380
gggcagtcct	cggcccaaaag	catcagctca	tcgagagcct	gcgcgacgga	cgcactgacg	1440
gtgtcggtcca	tcacagtttg	ccagtgtatac	acatggggat	cagcaatcgc	gcatatgaaa	1500
tcacgccatg	tagtgtattg	accgattcct	tgcggtccga	atgggccgaa	cccgtctctc	1560
tggctaagat	cggccgcagc	gatcgcatcc	atggcctccg	cgaccggctg	cagaacagcg	1620
ggcagttcgg	tttcaggcag	gtcttgcaac	gtgacaccct	gtgcacggcg	ggagatgcaa	1680
taggtcaggc	tctcgctgaa	ttccccaatg	tcaagcactt	ccggaatcgg	gagcgcggcc	1740
gatgcaaagt	gcccataaac	ataacgatct	ttgtagaaac	catcggcgca	gctattttacc	1800
cgcaggacat	atccacgccc	tcctacatcg	aagctgaaag	cacgagattc	ttcgcccttc	1860
gagagctgca	tcaggctcga	gacgtgtctg	aacttttcga	tcagaaactt	ctcgacagac	1920
gtcgcgggtga	gttcaggcct	tttcatTTTT	aatgttactt	ctcttgcaat	tagggaaacta	1980
taatgttaact	caaaaataaga	ttaaacaaac	taaaataaaa	agaagtata	cagaaaaacc	2040
cataataaacc	agtactaatc	cataataata	atacacaaaa	aaactatcaa	ataaaaaccag	2100
aaaacagatt	gaatagaaaa	attttttcga	tctcctttta	tattcaaaat	tcgatatatg	2160
aaaaagggaa	ctctcagaaa	atcaccaaat	caattttaatt	agatttttct	tttctttcta	2220
gcgttggaata	gaaaaatttt	tctttttttt	tttagaaatg	aaaaattttt	gccgtaggaa	2280
tcaccgtata	aaccctgtat	aaacgctact	ctgttcacct	gtgtaggcta	tgattgacct	2340
agtgttcatt	gttattgcga	gagagcggga	gaaaagaacc	gatacaagag	atccatgctg	2400
gtatagttgt	ctgtccaaca	ctttgatgaa	cttgtaggac	gatgatgtgt	attactatgc	2460
tcgacgctcg	tccaacgccc	gcggaacctg	ttatgacaat	tacaacaaca	gaattctttc	2520
tatatatgca	cgaacttgta	atatggaaga	aattatgacg	tacaaactat	aaagtaaata	2580
ttttacgtaa	cacatggtgc	tgttgtgctt	ctttttcaag	agaataccaa	tgacgtatga	2640
ctaagtttag	gatttaaatgc	aggtgacgga	cccattcttt	aaacgattta	tatcagtggc	2700
ttccaaattg	ttaggttttg	ttggttcctg	aggtttcctg	ttgtgggtca	tatgactttg	2760
aaccaaattg	ccggtctgta	gggcagcaca	taaggataat	tcacctgcca	agacggcaca	2820
ggcaactatt	cttgctaatt	gacgtgcgtt	ggtaccagga	gcggtagcat	gtgggcctct	2880
tacacctaat	aagtccaaca	tgccaccttg	tggttctaga	acagtaccac	caccgatggt	2940
acctacttctg	atggatggca	tgatacggga	aattctcaaa	tcaccgtcca	cttctttcat	3000
caatgttata	cagttggaac	tttcgacatt	tttgtcagga	tcttgtccta	atgccaaaga	3060
aacagctgtc	actaaattag	ctgcatgtgc	gttaaatcca	ccaacagacc	cagccattgc	3120
agatccaacc	aaattcttag	caatgttcaa	ctcaaccaat	gcggaacat	cactttttta	3180
cacttttctg	acaacatcac	caggaatagt	agcttctgct	acgacactct	taccacgacc	3240
ttcgatccag	ttgatggcag	ctggtttttt	gtcggtagac	tagttaccag	aaacggagac	3300
aacctccata	cttctccagc	catactcttc	taccatttgc	tttaatgagt	attcgacacc	3360
cttagaaaatc	atattcatac	ccatttcgct	accagtagtt	gttctaaatc	tcatgaagag	3420
taaatctcct	gctagacaag	tttgaatatg	ttgcagacgt	gcaaactctg	atgtagagtt	3480
aaaagctttt	tttaattgct	tttgtccctc	ttctgagtct	aaccatatct	tacaggcacc	3540
agatcttttc	aaagtgggga	aacggactac	tgggcctctt	gtcataccat	ccttagttaa	3600
aacagtgtgt	gcaccaccgc	cagcattgat	tgctttacag	ccacgcatgg	cagaagctac	3660
caacaacccc	tctgtagtgt	ccattgtgat	atgataagat	gtaccatcga	taaccaaggg	3720
gcctataaca	ccaacgggca	aaggcatgta	acctataaca	ttttcacaac	aagcgccaaa	3780
tacgcggctc	tagtcataat	ttttatatgg	taaacgatca	gatgctaata	caggagcttc	3840
tgccaaaatt	gaaagagcct	tcctacgtac	cgcaaccgct	ctcgtagtat	cacctaatTT	3900
tttctccaaa	ggttacaaaag	gtaacttacc	gtgaataacc	aaggcagcga	cctctttggt	3960
cttcaattgt	tttgtatttc	cactacttaa	ttaatgcttc	aattcttcta	aaggacgtat	4020
tttcttatcc	aagctttcaa	tatcgcgga	atcatcttcc	tcactagatg	atgaaggctc	4080
tgatgagctc	gattgctcag	atgataaact	tttgactttc	gatccagaaa	tgactgtttt	4140
attgggttaa	actggtgtag	aagccttttg	tacaggagca	gtaaaagact	tcttggtgac	4200
ttcagttctc	accaattgat	ctgcagccat	atccccgcgt	gcttgccggg	ccgttacttt	4260
ttttttggat	ggacgcaaaag	aagtttaata	atcataattac	atggcaatac	caccatatac	4320
atatccatat	ctaactttac	ttatatgttg	tgaaaatgta	aagagcccca	ttatcttagc	4380
ctaaaaaaac	cttctctttg	gaactttcag	taatacgctt	aactgctcat	tgctatatgt	4440
aagtacggat	tagaagccgc	cgagcgggcg	acagccctcc	gacggaagac	tctcctccgt	4500
gcgtcctggt	cttcaccggt	cgctgttcctg	aaacgcagat	gtgcctcgcg	ccgcactgct	4560
ccgaacaata	aagattctac	aatactagct	tttatgggta	tgaagaggaa	aaattggcag	4620
taacctggcc	ccacaaacct	tcaaatcaac	gaatcaaat	aacaaccata	ggataataat	4680
gcgattagtt	ttttagcctt	atctctgggg	taattaatca	gcgaagcgat	gattttttgat	4740
ctattaacag	atatataaat	gcaaaagctg	cataaccact	ttactaata	ctttcaacat	4800
tttcgggttg	tattacttct	tattcaaatg	tcataaaagt	atcaacaaaa	aattgttaat	4860
atactcttat	acttaacctg	caggccgcga	gcgccgatat	gtctcagaac	gtttacattg	4920
tatcgactgc	cagaacccca	attggttcat	tccagggttc	tctatcctcc	aagacagcag	4980
tggaattggg	tgctgtttgct	ttaaaaggcg	ccttggtctaa	ggttcagaa	ttggatgcac	5040
ccaaggattt	tgacgaaatt	atTTTTtggt	acgttctttc	tgccaatttg	ggccaagctc	5100
cggccagaca	agttgctttg	gctgccgggt	tgagtaatca	tatcgttgca	agcacagtta	5160

[0025]

acaaggctctg	tgcatccgct	atgaaggcaa	tcatttttggg	tgctcaatcc	atcaaagtgtg	5220
gtaatgctga	tggtgtcgta	gctgggtggtt	gtgaatctat	gactaacgca	ccatactaca	5280
tgccagcagc	ccgtgcggtt	gccaaatttg	gccaaactgt	tcttggtgat	ggtgtcgaaa	5340
gagatgggtt	gaacgatgcg	tacgatggtc	tagccatggg	tgtaacgca	gaaaagtgtg	5400
cccgtgattg	ggatattact	agagaacaac	aagacaattt	tgccatcgaa	tcctaccaa	5460
aatctcaaaa	atctcaaaa	gaaggtaaat	tcgacaatga	aattgtacct	gttaccatta	5520
agggatattag	aggtaagcct	gatactcaag	tcacgaagga	cgaggaaacct	gctagattac	5580
acgttgaaaa	attgagatct	gcaaggactg	ttttccaaaa	agaaaacggt	actgttactg	5640
ccgctaacgc	ttctccaatc	aacgatgggtg	ctgcagccgt	catcttgggt	tccgaaaaag	5700
ttttgaagga	aaagaatttg	aagccttttg	ctattatcaa	agggtggggt	gaggccgctc	5760
atcaaccagc	gatttttaca	tgggctccat	ctcttgaggt	tccaaaggct	ttgaaacatg	5820
ctggcatcga	agacatcaat	tctgttgatt	actttgaatt	caatgaagcc	ttttcggttg	5880
tcggtttgggt	gaacactaag	attttgaagc	tagaccatc	taaggttaat	gtatatgggtg	5940
gtgctgttgc	tctaggtcac	ccattgggtt	gttctgggtg	tagagtgggt	gttacactgc	6000
tatccatctt	acagcaagaa	ggaggttaaga	tcgggtgttg	cgccatttgt	aatgggtgggtg	6060
gtgggtgcttc	ctctattgtc	attgaaaaga	tatgattacg	ttctgcgatt	ttctcatgat	6120
ctttttcata	aaatacataa	atatataaat	ggctttatgt	ataacaggca	taattttaag	6180
ttttatttgc	gattcatcgt	ttttcaggta	ctcaaacgct	gagggtgtg	ttttgactta	6240
cttttccgcc	ttggcaagct	ggccgggtga	tacttgcaca	agttccacta	attactgaca	6300
tttgtgggtat	taactcgttt	gactgctcta	caattgtagg	atgttaatca	atgtccttggc	6360
tgccataacgc	gatcgccgac	gccgccgata	tgagaaaaaa	aatcggttgg	gcttaacttt	6420
aaagaaaaaa	gttgagatta	gattttattgt	gttataaata	tagatataca	attcttttata	6480
aaaaaaatat	atatatatat	cattgttatt	aaataaagag	ttttcctagt	atatagatta	6540
aaaaactact	ctattaaatg	agagctaaaa	aaagcaggct	gccaaaaaaa	taaagcattt	6600
atgaaggggg	ttcagcaaga	tgcaatcgat	gggggaagat	tattttttta	catcgtaaga	6660
tcttctaaat	ttgtcatcga	tgttgggtcaa	gtagtaaaca	ccactttgca	aatgtcfaat	6720
ggaaccttga	ggtttgaagt	tcttcttcaa	atgggcattt	tctctcaatt	cgatggcagc	6780
ttcgtaatcc	tttggagttt	cgggtgattct	cttggctaatt	ttgttagtaa	tatctaattc	6840
cttgataata	tgttggacgt	caccaacaat	tttgcaagaa	tatagagatg	cagctaaacc	6900
ggaaccgtaa	gaaaataaac	caacacgctt	gccttgtaag	tcgtcagatc	caacatagtt	6960
taatagagat	gcaaaggcgg	cataaacaga	tgccgtgtac	atgttacctg	tgtttgttgg	7020
aacaatacaa	gattgggcaa	ctctctcttt	gtggaatggc	ttagcaacat	taacaaaagt	7080
tttttcaatg	ttcttatcgg	ttaaagattc	gtcataatcg	cgagttagcta	attcggcgctc	7140
aacttctggg	aacaattgag	gattggctct	gaaatcgtaa	tatagtaatc	taccgtatga	7200
ttttgtgacc	aatttacagg	ttggaacatg	gaaaacgttg	tagtcgaaat	atttcaaaac	7260
gttcaaagca	tccgaaccag	cgggatcgct	aaccaaccct	ttagaaatag	ccttcttggga	7320
ataactcttg	taaacttgat	caagagcctt	gacgtaacaa	gttaatgaaa	aatgaccatc	7380
gacgtaagga	tattcgctgg	tgaaatctgg	cttgtaaaaa	tcgtaggcgt	gttccatgta	7440
agaagctctt	acagagtcaa	atacaattgg	agcatcagga	ccgatccaca	tagcaacagt	7500
accggcacca	ccggttggtc	ttgcggcacc	cttatcgtag	atggcaatat	caccgcaaac	7560
tacaatggcg	tctctaccat	cccattgcgtt	agattcaatc	cagttcaaa	agttgaacaa	7620
cgcgttggta	ccaccgtaac	aggcatttaag	cgtgtcaata	ccttcgacgt	cagtgttttc	7680
accaaacaat	gcatcaaga	cagacttgac	agacttggac	ttgtcaatca	gagtttcagt	7740
accgacttct	aatctaccaa	ttttgttgggt	gtcgtgttg	taactcttga	tcaacttaga	7800
caaaacagtt	agggacatcg	agtagatatc	ttctctgtca	ttgacaaaag	acatgttgggt	7860
ttggcccaga	ccaattgtgt	atttaccttg	agaaacgcca	tcaaatctt	ctagctcaga	7920
ttggttgaca	cattgagttg	ggatgtaaat	ttggatacct	ttaataccga	cattttgagg	7980
tctggttttt	tgttcagcgg	tcttttggtt	ttttagttca	gtcatttgca	agttgttatt	8040
gtgtaattgt	gtttgctttt	gcggcctaag	tcttctttta	ataccacacc	aacaaagttt	8100
agtagagagt	ttcataaggc	ggccgctggc	gaggagagt			8139

<210> 112

<211> 2962

<212> DNA

<213> 人工序列

<220>

<223> Phase I-B 缝合产物

<400> 112

aaggcggccg	ctggcgaggg	agataagtat	agaggatat	taacaatttt	ttgttgatac	60
ttttatgaca	tttgaataag	aagtaatata	aaccgaaaat	gttgaaagta	ttagttaaag	120
tggttatgca	gcttttgc	ttatatatct	gttaatatagat	caaaaatcat	cgttcgctg	180
attaattacc	ccagaaataa	ggctaaaaaa	gtaaatcgcat	tattatccta	tggttgttaa	240
tttgattcgt	tgatttgaag	gtttgtgggg	ccaggttact	gccaaatttt	cctcttcata	300
accataaaag	ctagtattgt	agaatcttta	ttgttcggag	cagtgcggcg	cgaggacat	360
ctgcgtttca	ggaacgcgac	cggatgaagac	caggacgcac	ggaggagagt	cttccgctcg	420

[0026]


```

agggctgtcg cccgctcggc ggctttcta atcgtacttca atatatgcaat gagcagttaa 480
gcgtattact gaaagttcca aagagaagg ttttttaggc taagataatg gggctcttta 540
catttccaca acatataagt aagattagat atggatatgt atatggtggt attgccatgt 600
aatatgatta ttaaacttct ttgctccat ccaaaaaaa agtaaaggcg cgccacggtc 660
gtgcggatag ggctgcagat caattggtga agactgaagt caccaagaag tcttttactg 720
ctcctgtaca aaaggcttct acaccagttt taaccaataa aacagtcatt tctggatcga 780
aagtcaaaag tttatcatct gcgcaatcga gctcatcagg accttcatca tctagttagg 840
aagatgattc ccgcgatatt gaaagcttgg ataagaaaat acgtccttta gaagaattag 900
aagcattatt aagtagtgga aatacaaaac aattgaagaa caaagaggtc gctgccttgg 960
ttattcacgg taagttacct ttgtacgctt tggagaaaaa attaggtgat actacgagag 1020
cggttgcggg acgtaggaag gctctttcaa ttttggcaga agctcctgta ttacatctg 1080
atcgtttacc atataaaaa tatgactacg accgcgtatt tggcgcttgt tgtgaaaatg 1140
ttatagggtta catgcctttg cccgttgggt ttataggccc cttggttatc gatggtacat 1200
cttatcatat accaatggca actacagagg gttgtttggg agcttctgcc atgctgtggc 1260
gtaaggcaat caatgctggc ggtggtgcaa caactgtttt aactaaggat ggtatgacaa 1320
gaggcccagt agtccgtttc ccaactttga aaagatctgg tgcctgtaag atatggttag 1380
actcagaaga gggacaaaac gcaattaaaa aagcttttaa ctctacatca agatttgac 1440
gtctgcaaca tattcaaaact tgtctagcag gagatttact cttcatgaga tttagaacaa 1500
ctactggtga cgcaatgggt atgaatatga tttctaaggg tgtcgaatac tcattaaagc 1560
aaatggtaga agagtatggc tgggaagata tggaggttgt ctccgtttct ggtaactact 1620
gtaccgacaa aaaaccagct gccatcaact ggcaggaagg tcgtggttag agtgtcgtcg 1680
cagaagctac tattcctggg gatgttgctg gaaaagtgtt aaaaagtgat gtttccgcat 1740
tgggttaggt gaacattgct aagaatttgg ttggatctgc aatggctggg tctgttgggt 1800
gatttaacgc acatgcagct aatttagtga cagctgtttt cttggcatta ggacaagatc 1860
ctgcacaaaa tgtcgaaagt tccaactgta taacattgat gaaagaagtg gacggtgatt 1920
tgagaatttc cgtatccatg ccatccatcg aagtaggtac catcggtggg ggtactgttc 1980
tagaaccaca aggtgccatg ttggacttat taggtgtaag aggccacat gctaccgctc 2040
ctggtaccaa cgcagctcaa ttagcaagaa tagttgcctg tgccgtcttg gcaggtgaat 2100
tatccttatg tgctgcccta gcagccggcc atttggttca aagtcatatg acccacaaca 2160
ggaaacctgc tgaaccaaca aaacctaaca atttggacgc cactgatata aatcgtttga 2220
aagatgggtc cgtcacctgc attaaatcct aaacttagtc atacgtcatt ggtattctct 2280
tgaaaaagaa gcacaacagc accatgtgtt acgtaaaaata tttactttat agtttgtacg 2340
tcataatttc ttccatatta caagtctgtg catatataga aagaattctg ttgttgaat 2400
tgtcataact agccctcag ccccttagc gtcgaagcat cttgccctgt gcttggcccc 2460
cagtgcagcg aacgttataa aaacgaatac tgagtatata tctatgtaaa acaaccatat 2520
catttcttgt tctgaacttt gtttacctaa ctagttttaa atttcccttt ttcgtgcatg 2580
cgggtgttct tatttattag catactacat ttgaaatatc aaatttcctt agtagaaaag 2640
tgagagaagg tgcactgaca caaaaaataa aatgctacgt ataactgtca aaactttgca 2700
gcagcgggca tccttccatc atagcttcaa acatattagc gttcctgatc ttcatacccg 2760
tgctcaaaat gatcaaacaa actgttattg ccaagaaata aacgcaaggc tgccttcaa 2820
aactgatcca ttagatcctc atatcaagct tcctcataga acgccaatt acaataagca 2880
tgttttgctg ttatcaccgg gtgatagggt tgctcaacca tggaggttag catgcggtgt 2940
ttaaacccca gcgcctggcg gg

```

<210> 113
 <211> 9664
 <212> DNA
 <213> 人工序列

<220>
 <223> Phase II 完全缝合产物

```

<400> 113
gacggcacgg ccacgcgttt aaaccgccta ggataattat actctatttc tcaacaagta 60
attggttgtt tggccgagcg gtctaaggcg cctgattcaa gaaatatctt gaccgcagtt 120
aactgtggga atactcaggt atcgtaagat gcaagagttc gaatctctta gcaaccatta 180
tttttttctt caacataacg agaacacaca gggcgctat cgacagaaat caaattcgat 240
gactggaaat tttttgttaa tttagaggt cgcctgacgc atataccttt ttcaactgaa 300
aaattgggag aaaaaggaaa ggtgagagcg ccggaaccgg cttttcatat agaatagaga 360
agcgttcatg actaaatgct tgcatacaca tacttgaagt tgacaatatt atttaaggac 420
ctattgtttt ttccaatagg tggtagcaa tcgtcttact ttctaacttt tcttaccttt 480
tacatttcag caatatatat atatatatat caaggatata ccattctagc tcacacgctg 540
ccagggggag cctcgacact agtaatacac atcatcgctc tacaagttca tcaaatgttt 600
gagacacaaa ctataccagc atggatctct tgtatcgggt cttttctccc gctctctcgc 660
aataacaatg aacactgggt caatcatagc ctacacaggt gaacagagta gcgtttatc 720
agggtttata cgggtgattc tacggcaaaa atttttcatt tctaaaaaaa aaaagaaaaa 780
tttttctttc caacgctaga aggaaaagaa aaatctaatt aaattgattt ggtgattttc 840

```

[0027]

tgagagtcc	ctttttcata	tatcgaattt	tgaatataaa	aggagatcga	aaaaatTTTT	900
ctattcaatc	tgTTTTctgg	TTTTatttga	tagTTTTttt	gtgtattatt	attatggatt	960
agtactgggt	tatatgggtt	tttctgtata	acttctTTTT	atttttagttt	gtttaatctt	1020
atTTtgagtt	acattatagt	tccttaactg	caagagaagt	aacattaaaa	atgaccactc	1080
ttgacgacac	ggcttaccgg	taccgcacca	gtgtcccggt	ggacgccgag	gccatcgagg	1140
cactggatgg	gtccttcacc	accgacaccg	tcttccgctg	caccgccacc	ggggacggct	1200
tcaccctgcg	ggaggtgccc	gtggaccgcg	ccctgaccaa	ggtgttcccc	gacgacgaat	1260
cggacgacga	atcggacgcc	ggggaggacg	gcgacccgga	ctcccgagcg	ttcgtcgcgt	1320
acggggacga	cggcgacctg	gcgggcttcg	tggtcgtctc	gtactccggc	tggaaaccgcc	1380
ggctgaccgt	cgaggacatc	gaggtcgccc	cggagcaccg	ggggcacggg	gtcgggcgcg	1440
cgttgattgg	gctcgcgacg	gagttcgccc	gcgagcgggg	cgccgggcac	ctctgctgtg	1500
aggtcaccaa	gctcaacgca	cggcgatcc	acgcgtaccg	gcggatgggg	ttcacctctt	1560
gcggcctgga	caccgccctg	tacgacggca	ccgcctcgga	cggcgagcag	gcgctctaca	1620
tgagcatgcc	ctgcccctga	gtttaacttg	atactactag	atTTTTtctc	ttcattttata	1680
aaatTTTTtg	ttataattga	agcttttagaa	gtatgaaaaa	atcctTTTTt	ttcattcttt	1740
gcaacaaaaa	taagaagctt	cttttattca	ttgaaatgat	gaatataaac	ctaacaaaaa	1800
aaaaagactc	gaatatcaaa	cattaaaaaa	aaatataaga	ggttatctgt	tttcccattt	1860
agttggagtt	tgcattttct	aatagataga	actctcaatt	aatgtggatt	tagtttctct	1920
gttcgtTTTT	TTTTgttttg	ttctcactgt	atTTacattt	ctatttagta	tttagttatt	1980
catataatct	taacttctcg	aggagctccg	ctcgtccaac	gccggcggac	ctattcgcg	2040
tggaaggac	cttgtggagg	aatatgaagt	tgatagtcta	aagggaattga	atggctattt	2100
aactgattat	gagtcaatgt	atcaaggata	ctatggctcg	cttaaatctt	attctgtctt	2160
cgaagctga	attgatacta	cgaaaaaatt	TTTTttgttt	ctctttctat	ctttattaca	2220
taaaacttca	tacacagtta	agattaaaaa	caactaataa	ataatgccta	tcgcaaatta	2280
gcttatgaag	tccatggtaa	attcgtgttt	cctggcaata	atagatcgct	aatttgttgc	2340
tttgtggtag	TTTTattttc	aaataattgg	aaatactagg	atTTgatttt	aagatcttta	2400
ttcaaatTTT	ttgcgcttaa	caaacagcag	ccagtcctcc	ccaagtctgt	ttcaaatgtc	2460
tcgtaactaa	aatcatcttg	caatttcttt	ttgaaactgt	caatttgtct	ttgagtaatg	2520
tctcttcgta	acaaagtcaa	agagcaaccg	ccgccaccag	caccggtaa	ttttgtggag	2580
ccaatttcta	aatcatcgct	cagattttta	ataagttcta	atccaggatg	agaaacaccg	2640
attgagacaa	gcagtccatg	atTTattctt	atcaattcca	atagttgttc	atacagttca	2700
ttattagttt	ctacagcttc	gtcatcggtg	cttttaccat	tacttaactt	agtcagtatg	2760
tctaagcctt	tgagggcaca	ttcacccatg	gcacttagaa	ttggcttcat	aacttcagga	2820
aatttctcgg	tgaccaacac	acgaacgcga	gcaacaagat	cttttgtaga	ccttggaaatt	2880
ctagtatagg	ttaggatcat	tggaaatggc	gggaaatcat	ctaagaactt	aaaattgttt	2940
gtgtttattg	ttccattatg	tgagtctttt	tcaaatagca	gggcattacc	ataagtggcc	3000
acagcgttat	ctattcctga	aggggtaccg	tgaatacact	ttcacctat	gaaggcccat	3060
tgattcacta	ttgctttatc	gttttctgac	agcttttcca	agtcattaga	tcctattaac	3120
cccccaagt	aggccatagc	taaggccagt	gatacagaaa	tagaggcgct	tgagcccaac	3180
ccagcaccga	tgggtaaagt	agacttttaa	gaaaacttaa	tattcttggc	atgggggcat	3240
aggcaacaaa	acatatacag	gaaacaaaac	gctgcatggt	agtggaagga	ttcggatagt	3300
tgagctaaca	acggatccaa	aagactaacg	agttcctgag	acaagccatc	ggtggcttgt	3360
tgagccttgg	ccaatttttg	ggagtttact	tgatgctcgg	tgatggcatt	gaaatcattg	3420
atggaccact	tatgattaaa	gctaattgtc	gggaagtcca	attcaatagt	atctggtgca	3480
gatgactcgc	ttatttagcag	gtaggttctc	aacgcagaca	cactagcagc	gacggcaggc	3540
ttgttgtaga	cagcagagtg	ttcaccaaaa	ataataacct	ttcccgggtg	agaagttaag	3600
aacggtaatg	acataatccc	gcgtgcttgg	ccggccgtta	cttttttttt	ggatggacgc	3660
aaagaagttt	ataatcata	ttacatggca	atgaccacat	atacatatcc	atatataatc	3720
ttacttatat	gttggtgaaa	tgtaaagagc	cccattatct	tagcctaaaa	aaaccttctc	3780
tttggaactt	tcagtaatac	gcttaactgc	tcattgctat	attgaagtac	ggattagaag	3840
ccgcccagcg	ggcgacagcc	ctccgacgga	agactctcct	ccgtgcgtcc	tggtcttcac	3900
cggtcgcgtt	cctgaaacgc	agatgtgcct	cgcgcgcac	tgctccgaac	aataaagatt	3960
ctacaatact	agcttttatg	gttatgaaga	ggaaaaattg	gcagtaacct	ggccccacaa	4020
accttcaaat	caacgaatca	aattaacaac	cataggataa	taatgcgatt	agtttttttag	4080
ccttattttct	ggggtaatta	atcagcgaag	cgatgatttt	tgatctatta	acagatatata	4140
aaatgcaaaa	gctgcataac	cacttttaact	aatactttca	acatttttcg	tttgatttac	4200
ttcttattca	aatgtcataa	aagtatcaac	aaaaaattgt	taatatacct	ctatacttaa	4260
cctgcaggcc	gcgagcgccg	atatgtcaga	gttgagagcc	ttcagtgtcc	cagggaaagc	4320
gttactagct	ggtggaatatt	tagttttaga	tccgaaatat	gaagcatttg	tagtcggatt	4380
atcggcaaga	atgcatgctg	tagcccatcc	ttacggttca	ttgcaagagt	ctgataagtt	4440
tgaagtgcgt	gtgaaaagta	aacaatttta	agatggggag	tggctgtacc	atataagtcc	4500
taaaactggc	ttcattcctg	tttcgatagg	cggatctaag	aaccctttca	ttgaaaaagt	4560
tatcgctaac	gtatttagct	actttaagcc	taacatggac	gactactgca	atagaaactt	4620
gttcgttatt	gtatgtttct	ctgatgtgc	ctaccattct	caggaggaca	gcgttaccca	4680
acatcggtgg	aacagaagat	tgagttttca	ttcgcacaga	attgaagaag	ttcccaaac	4740
agggctgggc	tcctcggcag	gtttagtcac	agtttttaact	acagcttttg	cctccttttt	4800
tgtatcggac	ctggaaaata	atgtagacaa	atatagagaa	gttattcata	atttatcaca	4860

[0028]

agttgctcat	tgtcaagctc	agggtaaaat	tggaagcggg	tttgatgtag	cggcggcagc	4920
atatggatct	atcagatata	gaagattccc	acccgcatta	atctctaatt	tgccagatat	4980
tggaagtgtc	acttacggca	gtaaaactggc	gcattttggtt	aatgaagaag	actggaatat	5040
aacgattaaa	agtaaccatt	taccttcggg	attaacttta	tggaatggcg	atattaagaa	5100
tggttcagaa	acagtaaaac	tggtccagaa	ggtaaaaaat	tggtatgatt	cgcataatgcc	5160
ggaaagcttg	aaaatatata	cagaactcga	tcattgcaa	tctagattta	tggaatggact	5220
atctaaacta	gatcgcttac	acgagactca	tgacgattac	agcgatcaga	tatttgagtc	5280
tcttgagagg	aatgactgta	cctgtcaaaa	gtatcctgag	atcacagaag	ttagagatgc	5340
agttgccaca	attagacgtt	ccttttagaaa	aataactaaa	gaatctgggt	ccgatatcga	5400
acctcccgtg	caaactagct	tattggatga	ttgccagacc	ttaaaaggag	ttcttacttg	5460
cttaatacct	gggtctgggt	gttatgcagc	cattgcatgt	attgctaagc	aagatgttga	5520
tcttagggct	caaaccgtg	atgacaaaag	attttctaag	gttcaatggc	tggatgtaac	5580
tcaggctgac	tggggtgtta	ggaaagaaaa	agatccggaa	acttatcttg	ataaataact	5640
taaggtagat	aatagtgggt	catgtgacat	cctttataaat	gtgaagtttg	aagtgaaccg	5700
gcttaacatc	taaccattca	tcttccgata	gtacttgaaa	ttgttccttt	cggcggcatg	5760
ataaaattcg	tttaattgggt	acaagctata	catactagga	tgaggatggt	actgcaaacg	5820
ataaataaac	tttctagata	tataacttta	gtcattttaa	atataaataa	agtgcgtgtt	5880
agcttgaaag	tgtgactata	cgcatcgcc	gacgccgcg	atgatgtgta	ttactagtgt	5940
cgacgacagc	attcgcccag	tatttttttt	attctacaaa	ccttctataa	tttcaaagta	6000
tttacataat	tctgtatcag	tttaatcacc	ataatatcgt	tttctttgtt	tagtgcaatt	6060
aatttttctt	attgttactt	cgggcctttt	tctgttttat	gagctatttt	ttccgtcatc	6120
cttccggatc	cagattttca	gcttcatctc	cattgtgtgt	ctacgtaatg	cacgccatca	6180
ttttaagaga	ggaaggcggc	cgctggcgag	ggagatatga	agctactgtc	ttctatcgaa	6240
caagcatgct	atatttgccg	acttaaaaag	ctcaagtgtc	ccaaagaaaa	accgaagtgc	6300
gccaaagtgt	tgaagaacaa	ctgggagtgt	cgctactctc	ccaaaaccaa	aagggtctccg	6360
ctgactaggg	cacatctgac	agaagtggaa	tcaaggctag	aaagactgga	acagctattt	6420
ctactgattt	ttcctcgaga	agaccttgac	atgattttga	aaatggattc	tttacaggat	6480
ataaaaagcat	tgtaaacagg	attattttgt	caagataatg	tgaataaaga	tgccgtcaca	6540
gatagattgg	cttcagtggg	gactgatatg	cctctaacat	tgagacagca	tagaataagt	6600
gcgacatcat	catcggaaga	gagtagtaac	aaaggctcaa	gacagttgac	tgtatcgatt	6660
gactcggcag	ctcatcatga	taactccaca	attccgttgg	attttatgcc	cagggatgca	6720
cttcatggat	ttgatgtgtc	tgaagaggat	gacatgtcgg	atggcctggc	cttctgaaa	6780
acggacccca	acaataatgg	gttctttggc	gacggttctc	tcttatgtat	tcttcgatct	6840
attggcttta	aaccggaaaa	ttacacgaac	tctaacgtta	acaggctccc	gaccatgatt	6900
acggatagat	acacgttggc	ttctagatcc	acaacatccc	gtttacttca	aagttatctc	6960
aataattttc	acccctactg	ccctatcgtg	cactcaccga	cgctaattgat	gttggtataat	7020
aaccagattg	aaatcgcgct	gaaggatcaa	tggcaaatcc	tttttaactg	catattagcc	7080
attggagcct	gggttataga	gggggaatct	actgatatag	atgtttttta	ctatcaaaat	7140
gctaaatctc	atttgacgag	caaggctctc	gagtcagggt	ccataatttt	ggtgacagcc	7200
ctacatcttc	tgtcgcgata	tacacagtgg	aggcagaaaa	caaatactag	ctataatttt	7260
cacagctttt	ccataagaat	ggccatatca	ttgggcttga	atagggacct	cccctcgtcc	7320
ttcagtata	gcagcattct	ggaacaaaga	cgccgaattt	ggtggctctg	ctactcttgg	7380
gagatccaat	tttccctgct	ttatggctga	tccatccagc	tttctcagaa	tacaatctcc	7440
ttcccttctt	ctgtcgacga	tgtgcagcgt	accacaacag	gtcccaccat	atatcatggc	7500
atcattgaaa	cagcaaggct	cttacaagtt	ttcacaaaaa	tctatgaact	agacaaaaca	7560
gtaactgcag	aaaaaagtcc	tatatgtgca	aaaaaatgct	tgatgatttg	taatgagatt	7620
gaggagggtt	cgagacaggc	accaaagttt	ttacaaatgg	atatttccac	caccgctcta	7680
accaatttgt	tgaaggaaca	cccttggcta	tcccttacaa	gattcgaact	gaagtggaaa	7740
cagttgtctc	ttatcattta	tgtattaaaga	gattttttca	ctaattttac	ctagaaaaag	7800
tcacaactag	aacaggatca	aaatgatcat	caaagttatg	aagttaaacg	atgtccatc	7860
atgttaagcg	atgcagcaca	agaactgtt	atgtctgtaa	gtagctatat	ggacaatcat	7920
aatgtcaccc	catatttttg	ctggaattgt	tcttattact	gtttcaatgc	agtcctagta	7980
cccataaaga	ctctactctc	aaactcaaaa	tcgaatgctg	agaataacga	gaccgcacaa	8040
ttattacaac	aaattaacac	tgttctgatg	ctattaaaaa	aactggccac	ttttaaaaatc	8100
cagacttggt	aaaaatacat	tcaagtactg	gaagagggtat	gtgcgccgtt	tctgttatca	8160
cagtgtgcaa	tcccattacc	gcatatcagt	tataacaata	gtaatggtag	cgccattaaa	8220
aatattgtcg	gttctgcaac	tatcgcccaa	taccctactc	ttccggagga	aatgttcaac	8280
aatatcagtg	ttaaatatgt	ttctcctggc	tcagttaggg	cttcacctgt	gccattgaaa	8340
tcaggagcaa	gtttcagtga	tctagtcaag	ctgttatcta	accgtccacc	ctctcgtaac	8400
tctccagtga	caataccaag	aagcacacct	tcgcatcgct	cagtcacgcc	ttttctaggg	8460
caacagcaac	agctgcaatc	attagtgtca	ctgaccccg	ctgctttgtt	tgggtggcgcc	8520
aatttttaatc	aaagtgggaa	tattgtctgat	agctcattgt	ccttcacttt	cactaacagt	8580
agcaacggct	cgaacctcat	aacaactcaa	acaaattctc	aagcgctttc	acaaccaatt	8640
gcctcctcta	acgttcagta	taacttcatg	aataatgaaa	tcacggctag	taaaattgat	8700
gatggtaata	attcaaaaac	actgtcacct	ggttggacgg	accaaactgc	gtataacgcg	8760
tttggaatca	ctacagggat	gtttaataacc	actacaatgg	atgatgtata	taactatcta	8820
ttcgtatgat	aagatacccc	accaaacc	aaaaaagagt	aaaatgaatc	gtagatactg	8880

[0029]

aaaaaccccg	caagttcact	tcaactgtgc	atcgtgcacc	atctcaattt	ctttcattta	8940
tacatcgttt	tgccttcttt	tatgtaacta	tactcctcta	agtttcaatc	ttggccatgt	9000
aacctctgat	ctatagaatt	ttttaaatga	ctagaattaa	tgcccatctt	ttttttggac	9060
ctaaattctt	catgaaaata	tattacgagg	gcttattcag	aagcttcgct	caaaggcgcg	9120
ccacggtcgt	gcggataaag	atttctcttt	tttatgatat	ttgtacataa	actttataaa	9180
tgaaattcat	aatagaaacg	acacgaaatt	acaaaatgga	atatgttcat	agggtagacg	9240
aaactatata	cgcaatctac	atacatttat	caagaaggag	aaaaaggagg	atgtaaagga	9300
atacaggtaa	gcaaattgat	actaatggct	caacgtgata	aggaaaaaga	attgcacttt	9360
aacattaata	ttgacaagga	ggagggcacc	acacaaaaag	ttaggtgtaa	cagaaaatca	9420
tgaaactatg	attcctaatt	tatatatttg	aggattttct	ctaaaaaaa	aaaaatacaa	9480
caaataaaaa	acactcaatg	acctgacct	ttgatggagt	ttaagtcaat	accttcttga	9540
accattttccc	ataatggtga	aagttccctc	aagaatttta	ctctgtcaga	aacggcctta	9600
acgacgtagt	cgacctcttc	ttcagtacta	aatctacggt	gtttaaaccc	cagcgcctgg	9660
cggg						9664

<210> 114

<211> 4380

<212> DNA

<213> 人工序列

<220>

<223> Phase III-A 缝合产物

<400> 114

gacggcacgg	ccacgcgttt	aaaccgcca	aagtgcagct	cagagcccc	agcaccagta	60
ttagaggta	taatgggctg	cgaagcctgc	taaaatgcag	tggaggccgt	gtaccctttg	120
ccaaattggc	tattggaatc	ggcagagaac	ctgggtcccg	ttctagagac	cctgcgagcg	180
tgtcccgggtg	ggttctggga	gctctaactc	cgcaggaact	acaaaccttg	cttacacaga	240
gtgaacctgc	tgcctggcgt	gctctgactc	agtacatttc	atagcccatc	ttcaacaaca	300
ataccgactt	accatcctat	ttgctttgcc	ctttttcttt	tccactgcac	tttgcctcgg	360
aaggcggtat	cggttttggg	tttagtgcct	aaacgagcag	cgagaacacg	accacgggct	420
ataataatgg	aaagtttagga	caggggcaaa	gaataagagc	acagaagaag	agaaaagacg	480
aaaagcagaa	gcggaaaacg	tatacacgtc	acatatcaca	cacacacgct	cacacgcggc	540
cagggggagc	ctcgacacta	gtaatacaca	tcatcgctct	acaagtctcat	caaagtgttg	600
gacagacaac	tataccagca	tggatctctt	gtatcgggtc	ttttctcccg	ctctctcgca	660
ataacaatga	acactgggtc	aatcatagcc	tacacaggtg	aacagagtag	cgttttataca	720
gggttttatac	ggtgattcct	acggcaaaaa	tttttctatt	ctaaaaaaa	aaagaaaaat	780
ttttctttcc	aacgctagaa	ggaaaagaaa	aatctaatta	aattgatttg	gtgattttct	840
gagagtctcc	ttttctatat	atcgaatttt	gaatataaaa	ggagatcgaa	aaaatttttc	900
tattcaatct	gttttctggt	tttatttgat	agtttttttg	tgtattatta	ttatggatta	960
gtactgggtt	atatgggttt	ttctgtataa	cttcttttta	ttttagtttg	tttaattctta	1020
ttttgagtta	cattatagtt	ccctaactgc	aagagaagta	acattaaaaa	tgggtaagga	1080
aaagactcac	gtttcgaggc	cgcatataaa	ttccaacatg	gatgctgatt	tatatgggta	1140
taaattgggt	cgcgataatg	tcgggcaatc	aggtgcgaca	atctatcgat	tgtatgggaa	1200
gcccgatgcg	ccagagtgtg	ttctgaaaca	tggcaaaggt	agcgttgcca	atgatgttac	1260
agatgagatg	gtcagactaa	actggctgac	ggaattttatg	cctcttccga	ccatcaagca	1320
ttttatccgt	actcctgatg	atgcatgggt	actcaccact	gcgatccccg	gcaaaacagc	1380
attccaggta	ttagaagaat	atcctgattc	aggtgaaaaa	attgttgatg	cgctggcagc	1440
gttcctgctc	cgggttcatt	cgattcctgt	ttgtaattgt	ccttttaaca	gcgatcgctg	1500
atttcgtctc	gctcaggcgc	aatcacgaat	gaataacggt	ttggttgatg	cgagtgattt	1560
tgatgacgag	cgtaatggct	ggcctgttga	acaagtctgg	aaagaaatgc	ataagctttt	1620
gccatttcta	ccggattcag	tcgtcactca	tgggtgatttc	tcacttgata	accttatttt	1680
tgacgagggg	aaattaatat	gttgatttga	tgttgagaca	gtcggaatcg	cagaccgata	1740
ccaggatctt	gccatcctat	ggaactgcct	cggtagattt	tctccttcat	tacagaaacg	1800
gctttttcaa	aaatatggta	ttgataatcc	tgatatgaat	aaattgcagt	ttcattttgat	1860
gctcgatgag	tttttctaag	tttaacttga	tactactaga	ttttttctct	tcattttataa	1920
aatttttggg	tataattgaa	gcttttagaag	tatgaaaaaa	tccttttttt	tcatttctttg	1980
caacccaaa	aagaagcttc	ttttattcat	tgaatgatg	aatataaacc	taacaaaaga	2040
aaaagactcg	aatatcaaac	attaaaaaaa	aataaaaagag	gttatctggt	ttcccattta	2100
gttgaggatt	gcatttttcta	atagatagaa	ctctcaatta	atgtggattt	agtttctctg	2160
ttcgtttttt	tttgttttgt	tctcactgta	tttacatttc	tatttagtat	ttagttattc	2220
atataatctt	aacttctcga	ggagctccgc	tcgtccaacg	ccggcggacc	tcttgtgcta	2280
agtgggtgctg	ttagacagct	acgaataaag	aaattccgaa	gcatgtaggg	aggtcatgat	2340
atgaaaaagc	aaaagagttag	gcatcaaaaa	gtttctcatt	caagtgttaa	ctgctgttaa	2400
aatgaagata	tttataaaat	gaagcttggg	ggttccgacc	aataccgtag	ggaaacgtaa	2460
attagctatt	gtaaaaaaag	gaaaagaaaa	gaaaagaaaa	atgttacata	tcgaattgat	2520
cttatttcctt	tggtagacca	gtctttgcgt	caatcaaaga	ttcgtttgtt	tcttgtgggc	2580

[0030]

ctgaaccgac	ttgagttaaa	atcactctgg	caacatcctt	ttgcaactca	agatccaatt	2640
cacgtgcagt	aaagttagat	gattcaaatt	gatggttgaa	agcctcaagc	tgctcagtag	2700
taaatttctt	gtcccatcca	ggaacagagc	caaacaattt	atagataaat	gcaaagagtt	2760
tcgactcatt	ttcagctaag	tagtacaaca	cagcatttgg	acctgcatca	aacgtgtatg	2820
caacgattgt	ttctccgtaa	aactgattaa	tggtgtggca	ccaactgatg	atacgccttg	2880
aagtgtcatt	catgtagaat	attggaggga	aagagtccaa	acatgtggca	tggaagagat	2940
tggaatccat	cattgtttcc	tttgcaaagg	tggcgaaatc	tttttcaaca	atggcctttac	3000
gcatgacttc	aaatctcttt	ggtacgacat	gttcaattct	ttcttttaaat	agttcggagg	3060
ttgccacggt	caattgcata	ccctgagttg	aactcacatc	cttttttaata	tcgctgacaa	3120
ctaggacaca	agctttcatc	tgaggccagt	cagagctgtc	tgcgatttgt	actgccatgg	3180
aatcatgacc	atcttcagct	tttcccattt	cccaggccac	gtatccgcca	aacaacgatc	3240
tacaagctga	accagacccc	tttcttgcta	ttctagatat	ttctgaagtt	gactgtggta	3300
attggtataa	cttagcaatt	gcagagacca	atgcagcaaa	gccagcagcg	gaggaagcta	3360
aaccagctgc	tgtaggaaag	ttattttcgg	agacaatgtg	gagtttccat	tgagataatg	3420
tgggcaatga	ggcgtccttc	gattccattt	cctttcttaa	ttggcgtagg	tcgcgacagc	3480
aattttgagt	cttttcattg	tcgatgctgt	gtggttctcc	atttaaccac	aaagtgtcgc	3540
gttcaaactc	aggtgcagta	gccgcagagg	tcaacgttct	gaggtcatct	tgcatataag	3600
tcactgatat	ggacgaattg	gtgggcagat	tcaacttcgt	gtcccttttc	ccccataact	3660
taagggttgc	gatgttgacg	ggtgcggtaa	cggatgctgt	gtaaacggtc	atatccccgc	3720
gtgcttggcc	ggcgtttact	tttttttttg	atggacgcaa	agaagtttaa	taatcatatt	3780
acatggcaat	accaccatat	acatatccat	atctaatctt	acttatatgt	tgtggaaatg	3840
taagagagcc	cattatctta	gcctaaaaaa	accttctctt	tggaactttc	agtaatacgc	3900
ttactgtctc	attgctatat	tgaagtacgg	attagaagcc	gccgagcggg	cgacagccct	3960
ccgacggaag	actctctctc	gtgcgtctct	gtcttcaccg	gtcgcgttcc	tgaaacgcag	4020
atgtgcctcg	cgccgcactg	ctccgaacaa	taaagattct	acaatactag	cttttatggg	4080
tatgaagagg	aaaaattggc	agtaaccttg	ccccacaaac	cttcaaatca	acgaatcaaa	4140
ttaacaacca	taggataata	atgcgattag	tttttttagc	ttatttctgg	ggaattaat	4200
cagcgaagcg	atgatttttg	atctattaac	agatatataa	atgcaaaagc	tgcataacca	4260
ctttaactaa	tactttcaac	attttcgggt	tgtattactt	cttattcaaa	tgtcataaaa	4320
gtatcaacaa	aaaattgtta	atatacctct	atacttaacc	tgaggccgcg	gagcgccgat	4380

<210> 115

<211> 4284

<212> DNA

<213> 人工序列

<220>

<223> Phase III-B 缝合产物

<400> 115

aacctgcagg	ccgcgagcgc	cgatatggct	tcagaaaaag	aaattaggag	agagagattc	60
ttgaacgttt	cccctaaatt	agtagaggaa	ttgaacgcac	cgcttttggc	ttacggtatg	120
cctaaggaag	catgtgactg	gtatgccac	tcattgaact	acaacactcc	aggcggtaag	180
ttaaatagag	gtttgtccgt	tgtggacacg	tatgctattc	tctccaacaa	gaccgttgaa	240
caattggggc	aagaagaata	cgaaaagggt	gctattctag	gttgggtgcat	tgagttgttg	300
caggcttact	ctttggtcgc	cgatgatatg	atggacaagt	ccattaccag	aagaggccaa	360
ccatgttggt	acaaggttcc	tgaagttggg	gaaattgcca	tcaatgacgc	attcatgtta	420
gaggctgcta	tctacaagct	tttgaaatct	cacttcagaa	acgaaaaata	ctacatgatg	480
atcaccgaat	tgttccatga	agtcaccttc	caaaccgaat	tgggccaatt	gatggactta	540
atcactgcac	ctgaagacaa	agtcgacttg	agtaagttct	ccctaaagaa	gcactccttc	600
atagttactt	tcaagactgc	ttactattct	ttctacttgc	ctgtcgcatt	ggctatgtac	660
gttgccggta	tcacagatga	aaaggatttg	aaacaagcca	gagatgtctt	gattccattg	720
ggtgaatatt	tccaaattca	agatgactac	ttagactgct	tcggtacccc	agaacagatc	780
ggtaagatcg	gtacagatat	ccaagataac	aaatgttctt	gggtaatcaa	caaggcatta	840
gaacttgctt	ccgcagaaca	aagaaagact	ttagacgaaa	attacggtaa	gaaggactca	900
gtcgcagaag	ccaaatgcaa	aaagattttc	aatgacttga	aaatcgacca	ggtataccac	960
gaatatgaag	agtctgttgc	caaggatttg	aaggccaaga	tctcccaagt	cgacgagtct	1020
cgtaggcttca	aagccgacgt	cttaactgcg	tttttgaaca	agggtttaca	gagaagtaaa	1080
taagactgga	agcgttcaat	cgataaaaaa	tggaatacag	attagataag	gacctgttat	1140
aagaaatata	tacttccact	ataatatagt	ataagcttac	agataattat	ctcttgattt	1200
accggttacac	gtgactaaag	gacgcttttt	ctcagccaat	gaaagtgaag	aaaaacttga	1260
tcggaaatata	cgggtagtac	gtaaaaggaa	cttgagccac	cccccaaaaa	tttaccata	1320
taataatag	aaaacaacg	acctcaacgc	cgatcgccga	cgccgccgat	cacggtagca	1380
acaagaatat	agcacgagcc	gcgaagttca	tttcgttact	tttgatatcg	ctcacacta	1440
ttgcgaagcg	cttcagtga	aaaatcataa	ggaaaagttg	taaatattat	tggtagtatt	1500
cgtttggttaa	agtagagggg	gtaatttttc	ccctttattt	tgttcataca	ttcttaaaat	1560

[0031]

gctttgcctc	tccttttggga	aagctatact	tcggagcact	gttgagcgaa	ggctcattag	1620
atataatttc	tgtcattttc	cttaacccaa	aaataaggga	aagggtccaa	aaagcgctcg	1680
gacaactggt	gaccgtgatc	cgaaggactg	gctatacagt	gttcacaaaa	tagccaagct	1740
gaaaataatg	tgtagctatg	ttcagttagt	ttggctagca	aagatataaa	agcaggctcg	1800
aaatatttat	gggcatttat	atgcagagca	tcaacctgat	aaaaaaaaag	cggccgctgg	1860
cgagggagat	atgactgccg	acaacaatag	tatgccccat	ggtgcagtat	ctagttacgc	1920
caaattagtg	caaaaccaa	cacctgaaga	cattttggaa	gagtttcctg	aaattattcc	1980
attacaacaa	agacctaata	cccgatctag	tgagacgtca	aatgacgaaa	gcggagaaac	2040
atgtttttct	ggtcatgatg	aggagcaaat	taagttaatg	aatgaaaatt	gtattgtttt	2100
ggattgggac	gataatgcta	ttggtgccgg	taccaagaaa	gtttgtcatt	taatggaaaa	2160
tattgaaaag	ggttttactac	atcgtgcatt	ctccgtcttt	attttcaatg	aacaaggatga	2220
attactttta	caacaaagag	ccactgaaaa	aataactttc	cctgatcttt	ggactaacac	2280
atgctgctct	catccactat	gtattgatga	cgaattaggt	ttgaagggtg	agctagacga	2340
taagattaag	ggcgctatta	ctgcggcggt	gagaaaacta	gatcatgaat	taggtattcc	2400
agaagatgaa	actaagacaa	ggggtaagtt	tcacttttta	aacagaatcc	attacatggc	2460
accaagcaat	gaaccatggg	gtgaacatga	aattgattac	atcctatttt	ataagatcaa	2520
cgttaaagaa	aacttgactg	tcaacccaaa	cgtcaatgaa	gttagagact	tcaaatgggt	2580
ttcaccaaat	gatttgaaaa	ctatgtttgc	tgacccaagt	tacaagttaa	cgccttgggt	2640
taagattatt	tgcgagaatt	acttattcaa	ctggtgggag	caattagatg	acctttctga	2700
agtggaaaaat	gacaggcaaa	ttcatagaat	gctataacaa	cgctcaata	atataggcta	2760
cataaaaaatc	ataataactt	tgttatcata	gcaaaatgtg	atataaaacg	tttcatttca	2820
cctgaaaaat	agtaaaaaata	ggcgacaaaa	atccttagta	atatgtaaac	tttattttct	2880
ttattttattt	acagaactct	gaatatacat	tgattgttca	catttttttt	ttctcttctc	2940
aattttccctt	gatttatattc	aaaaggttat	tggcctcttg	aatgtttccc	actgatgaag	3000
gcgcgccacg	gtcgtgcgga	tgtattccaa	tgagaatcgc	tagaaatgct	ttaccagaac	3060
tagactactt	gtcgcagatc	acttttgaac	tgtatgagag	tacggatgct	tctggtcaaa	3120
aatcgcatct	cattagactg	aaaatgtctc	ctgggtgtca	tactcgagat	ccgttagatg	3180
ttcaattaga	tgacaggcat	tatatagttt	gtattccaaa	gatttccctg	acgaagcatt	3240
tggatatgga	ctacgttcaa	cagaaattga	gaaacaaatt	taccagggtc	attatgcctc	3300
cgaaattttac	accagtaaac	attacgagcc	ccaacttgag	tttccagaaa	cgcaaaacca	3360
gaagaaagtc	ggtatctggt	gagaagttga	agcttccctg	ctcgtccgga	tcttcatcat	3420
ctacctccgt	taacaagaca	ttagattagt	gatcacaccc	aatttttaat	ttagcaacc	3480
aaaataaata	agtattttact	caactttttt	ttaataaaaa	aaaacttaat	tgaattttgc	3540
tcgcgatctt	taggtccggg	gttttcgttg	aacccttaga	cgagcaaatt	agcgccataa	3600
ggatatacgt	cagagcacat	taattagtg	catataccta	tataaagagc	aaccttctcc	3660
gatagacttg	taattttatct	tattttcattt	cctaacactt	tggtcgaaga	agagggataa	3720
aaacagacga	aagccccctca	gcccccttag	cgctgatggg	aaagctatta	caattggcat	3780
tgcattccgg	cgagatgaag	gcagctttga	agctgaagtt	ttgcagaaca	ccgctattct	3840
ccatctatga	tcagtccacg	tctccatctc	tcttgactg	tttcgaactg	ttgaacttga	3900
cctccagatc	gtttgtctgct	gtgatcagag	agctgcattc	agaattgaga	aactgtgtta	3960
ctctcttttta	tttgattttta	agggtcttgg	ataccatcga	agacgatatg	tccatcgaac	4020
acgattttgaa	aatgactttg	ttgcgtcact	tccacgagaa	attgtttgta	actaaatgga	4080
gtttcgacgg	aaatgcccc	gatgtgaagg	acagagccgt	tttgacagat	ttcgaatcga	4140
ttctttattga	attccacaaa	ttgaaaccag	aatatcaaga	agtcattcaag	gagatcacgc	4200
agaaaatggg	taatggtatg	gccgactaca	tcttagatga	aaattacaac	ttgaatcggt	4260
gtttaaaccc	cagcgcctgg	cggg				4284

<210> 116

<211> 6344

<212> DNA

<213> 人工序列

<220>

<223> Phase I 标记循环缝合产物

<400> 116

gacggcacgg	ccacgcgttt	aaaccgcccc	gatggaatcc	cttccataga	gagaaggagc	60
aagcaactga	cccaatatgg	actgccactg	gacctgaaga	catgcaacaa	agtgaagca	120
tagtggggcc	ttctttccaa	gctaattccg	tcactgccac	tgctgctacg	gaaaaccaac	180
ctaaagggtat	taactttctt	actataagaa	aatcacacga	gcgcccggac	gatgtctctg	240
tttaaattggc	gcaagttttc	cgctttgtaa	tatatattta	tacccctttc	ttctctcccc	300
tgcaatataa	tagtttaatt	ctaataattaa	taatatccta	tattttcttc	atttaccggc	360
gcactctcgc	ccgaacgacc	tcaaaatgtc	tgctacattc	ataataacca	aaagctcata	420
actttttttt	ttgaacctga	atatatatatc	atcacatgtc	actgctggtc	cttgcgcacc	480
agcgtataca	atctcgatag	ttggtttccc	gttctttcca	ctcccgtcgc	tcacacgcgg	540
ccaggggggag	cctcactatt	attccataag	atgatcatta	gcattacgtt	caaaacgagt	600
acaataaact	taagtaataa	cacgagccat	atgaccatta	gcaagatgac	aagcaagtta	660

[0032]

agaccaatca	gcttccatca	tagcatcagc	ttaacgttca	ccattaataa	gagtagaaat	720
ttcaccttca	agacaataac	gatttttcgtg	gtaataactg	atataattaa	attgaagctc	780
taatttgtga	gttttagtata	catgcattta	cttataatac	agtttttttag	ttttgctggc	840
cgcatcttct	caaataatgct	tcccagcctg	cttttctgt	acgttcaccc	tctaccttag	900
catcccttcc	ctttgcaaat	agtcctcttc	caacaataat	aatgtcagat	cctgtagaga	960
ccacatcatc	cacggttcta	tactgttgac	ccaatgctgc	tcccttgtca	tctaaacca	1020
caccgggtgt	cataatcaac	caatcgtaac	cttcatctct	tccacccatg	tctctttgag	1080
caataaagcc	gataacaaaa	tctttgtcac	tcttcgcaat	gtcaacagta	cccttagtat	1140
attctccagt	agctagggag	cccttgcag	acaattctgc	taacatcaaa	aggcctctag	1200
gttcccttgt	tacttcttcc	gccgcctgct	tcaaaccgct	aacaatacct	gggcccacca	1260
caccgtgtgc	attcgtaatg	tctgcccatt	ctgctattct	gtatacacc	gcagagtact	1320
gcaatttgac	tgtattacca	atgtcagcaa	attttctgtc	ttcgaagagt	aaaaaattgt	1380
acttggcgga	taatgccttt	agcggcttaa	ctgtgccctc	catggaaaaa	tcagtcaaga	1440
tatccacatg	tgtttttagt	aaacaaattt	tgggacctaa	tgcttcaact	aactccagta	1500
attccttggg	ggtacgaaca	tccaatgaag	cacacaagtt	tgtttgcttt	tcgtgcatga	1560
tattaaatag	cttggcagca	acaggactag	gatgagttag	agcacgttcc	ttatatgtag	1620
ctttcgacat	gatttatctt	cgtttcttgc	aggtttttgt	tctgtgcagt	tgggttaaga	1680
atactgggca	atttcatgtt	tcttcaacac	cacatatgcg	tatatatacc	aatctaagtc	1740
tgtgtctcct	ccttcgttct	tccttctgct	cggagattac	cgaatcaaaa	aaatttcaaa	1800
gaaaccggaa	tcaaaaaaaaa	gaacaaaaaa	aaaaaagatg	aattgaaaag	ctttatggac	1860
cctgaaacca	ctcactatta	ttccataaga	tgactcattag	cattacgttc	aaaacgagta	1920
caaataactt	aagtaataac	acgagccata	tgaccattag	caagatgaca	agcaagttaa	1980
gaccaatcag	cttccatcat	agcatcagct	taacgttcac	cattaataag	agtagaaatt	2040
tcaccttcaa	gacaataacg	attttctgtc	gctcgtccaa	cgccggcgga	cctgacggt	2100
gcaacaagaa	tatagcacga	gccgcggagt	tcatttctgt	acttttgata	tcactcacaa	2160
ctattgcgaa	gcgcttcagt	gaaaaaatca	taaggaaaag	ttgtaaatat	tattggtagt	2220
attcgtttgg	taaagtagag	ggggttaatt	ttccccttta	ttttgttcat	acattcttaa	2280
attgctttgc	ctctcctttt	ggaaagctat	acttcggagc	actgttgagc	gaaggctcat	2340
tagatataat	ttctgtcatt	ttccttaacc	caaaaataag	ggaaagggtc	caaaaagcgc	2400
tcggacaact	gttgaccgtg	atccgaagga	ctggctatac	agtgttcaca	aaatagccaa	2460
gctgaaaata	atgtgtagct	atgttcagtt	agtttggtta	gcaaagatat	aaaagcaggt	2520
cggaaatatt	tattgggcatt	attatgcaga	gcatcaacat	gataaaaaaa	aacagttgaa	2580
tattccctca	aaaatccccg	cgtgcttggc	cgccgtaaat	taataatgtc	aactttgcct	2640
atttcttctg	tgtcattttc	ctcttctaca	tcaccattag	tcgtggacga	caaagtctca	2700
accaagcccg	acgttatcag	acatacaatg	aatttcaatg	cttctatttg	gggagatcaa	2760
ttcttgacct	atgatgagcc	tgaagattta	gttatgaaga	aacaattagt	ggaggaatta	2820
aaagaggaag	ttaagaagga	attgataact	atcaaagggt	caaatgagcc	catgcagcat	2880
gtgaaattga	ttgaattaat	tgatgctgtt	caacgtttag	gtatagctta	ccattttgaa	2940
gaagagatcg	aggaagcttt	gcaacatata	catgttacct	atggtgaaca	gtgggtggat	3000
aaggaaaatt	tacagagtat	ttcattgttg	ttcaggttgt	tgctgcaaca	gggctttaac	3060
gtctcctctg	gcgttttcaa	agactttatg	gacgaaaaag	gtaaattcaa	agagtcttta	3120
tgaactgatg	cacaaggaat	attagcctta	tatgaagctg	catttatgag	ggttgaagat	3180
gaaacctatc	tagacaattgc	tttggaattc	acaaaagttc	atttagatat	catagcaaaa	3240
gacccatctt	gcgattcttc	attgctgaca	caaattccatc	aagccttaaa	acaaccttta	3300
agaaggagat	tagcaaggat	tgaagcatta	cattacatgc	caatctacca	acaggaaaca	3360
tctcatgatg	aagtattgtt	gaaattagcc	aagttggatt	tcagtgtttt	gcagtctatg	3420
cataaaaaag	aattgtcaca	tatctgtaag	tggtggaaag	atttagattt	acaaaataag	3480
ttaccttatg	tacgtgatcg	tggtgtcgaa	ggctacttct	ggatattgtc	acataatact	3540
gagccacaac	acgctagaac	aagaatgttt	ttgatgaaaa	catgcatgtg	gttagtagtt	3600
ttggacgata	cttttgataa	ttatggaaca	tacgaagaat	tggaagattt	tactcaagcc	3660
gtcgagagat	ggcttatctc	atgcttagat	atgttgcccg	aatatatgaa	attaatctac	3720
caagaattag	tcaatttgca	tgtggaaatg	gaagaatctt	tggaagagga	gggaaagacc	3780
tatcagattc	attacgttaa	ggagatggct	aaagaattag	ttcgtaatta	cttagtagaa	3840
gcaagatggg	tgaaggaagg	ttatatgcct	acttttagaag	aatacatgtc	tgtttctatg	3900
gttactggta	cttatggttt	gatgattgca	aggtcctatg	ttggcagagg	agacattgtt	3960
actgaagaca	cattcaaatg	ggtttctagt	taccaccta	ttattaaagc	ttcctgtgta	4020
atagtaagat	taatggacga	tattgtatct	cacaaggaag	aacaagaaag	aggacatgtg	4080
gcttcatcta	tagaatgtta	ctctaaagaa	tcaggtgctt	ctgaagagga	agcatgtgaa	4140
tatattagta	ggaaagttga	ggatgcctgg	aaagtaatca	atagagaatc	tttgcgtcca	4200
acagccgttc	ccctcccttt	gttaatgcc	gcaataaact	tagctagaat	gtgtgaggtc	4260
ttgtactctg	ttaatgatgg	ttttactcat	gctgaggggtg	acatgaaatc	ttatatgaag	4320
tccttcttctg	ttcatcctat	ggtcggttga	gctagctaag	atccgctcta	accgaaaagg	4380
aaggagttag	acaacctgaa	gtctaggtcc	ctatttattt	ttttatagtt	atggttagtat	4440
taagaacgtt	atttatattt	caaattttct	tttttttctt	gtacagacgc	gtgtacgcag	4500
gtaacattat	actgaaaacc	ttgcttgaga	aggttttggg	acgctcgaag	aacctgcagg	4560
ccgcgagcgc	cgatagttat	gacaattaca	acaacagaat	tctttctata	tatgcacgaa	4620
cttghtaatat	ggaagaaatt	atgacgtaca	aactataaag	taaatatttt	acgtaacaca	4680

[0033]

tggtgctggt	gtgcttcttt	ttcaagagaa	taccaatgac	gtatgactaa	gtttaggatt	4740
taatgcaggt	gacggacca	tctttcaaac	gatttatatc	agtggcgctc	aaattgttag	4800
gttttgttgg	ttcagcaggt	ttcctgttgt	gggtcatatg	actttgaacc	aaatggccgg	4860
ctgctagggc	agcacataag	gataattcac	ctgccaagac	ggcacaggca	actattcttg	4920
cctaattgacg	tgcgttggtg	ccaggagcgg	tagcatgtgg	gcctcttaca	cctaataagt	4980
ccaacatggc	accttgtgg	tctagaacag	taccaccacc	gatggtacct	acttcgatgg	5040
atggcatgga	tacggaaatt	ctcaaatac	cgtccacttc	tttcatcaat	gttatacagt	5100
tggaaactttc	gacattttgt	gcaggatctt	gtcctaattg	caagaaaaca	gctgtcacta	5160
aatttagctgc	atgtgctgta	aatccaccaa	cagaccagc	cattgcagat	ccaaccaa	5220
tcttagcaat	gttcaactca	accaatgcgg	aaacatcact	ttttaacact	tttctgacaa	5280
catcaccagg	aatagtagct	tctgacga	cactcttacc	acgaccttcg	atccagttga	5340
tggcagctgg	ttttttgtcg	gtacagtagt	taccagaaac	ggagacaacc	tccatatctt	5400
cccagccata	ctcttctacc	atttgcttta	atgagtattc	gacaccctta	gaaatcatat	5460
tcataccat	tgcgtcacca	gtagttgttc	taaatctcat	gaagagtaaa	tctcctgcta	5520
gacaagtttg	aatatgttgc	agacgtgcaa	atcttgatgt	agagttaaaa	gcttttttaa	5580
ttgctgtttg	tccctcttct	gagtctaacc	atatcttaca	ggcaccagat	cttttcaaa	5640
ttgggaaacg	gactactggg	cctcttgcta	taccatcctt	agttaaaaca	gttggtgcac	5700
caccgccagc	attgattgcc	ttacagccac	gcatggcaga	agctaccaa	caaccctctg	5760
tagttgccat	tggatatatga	taagatgtac	catcgataac	caaggggcct	ataacaccaa	5820
cgggcaaagg	catgtaacct	ataacatttt	cacaacaagc	gccaaatagc	cggtcgtagt	5880
cataattttt	atatggtaaa	cgatcagatg	ctaatacagg	agcttctg	aaaattgaaa	5940
gagccttcct	accgtaccga	accgtctc	tagtatcacc	taattttttc	tccaaagcgt	6000
acaaaggtaa	cttaccgtga	ataaccaagg	cagcgacctc	tttgttcttc	aattgttttg	6060
tatttccact	acttaataat	gcttctaatt	cttctaaagg	acgtattttc	ttatccaagc	6120
tttcaatatc	gcgggaatca	tcttcctcac	tagatgatga	aggtcctgat	gagctcgatt	6180
gcgcagatga	taaacttttg	actttcgatc	cagaaatgac	tgttttattg	gttaaaactg	6240
gtgtagaagc	cttttgtaca	ggagcagtaa	aagacttctt	ggtgacttca	gtcttcacca	6300
attggtctgc	agccatcggt	gtttaaaccc	cagcgctg	cggg		6344

<210> 117

<211> 6184

<212> DNA

<213> 人工序列

<220>

<223> Phase II 标记循环缝合产物

<400> 117

gacggcacgg	ccacgcgttt	aaaccgccta	ggataattat	actctatttc	tcaacaagta	60
attggttggt	tggccgagcg	gtctaaggcg	cctgattcaa	gaaatatctt	gaccgcagtt	120
aactgtggga	atactcaggt	atcgtaagat	gcaagagttc	gaatctctta	gcaaccatta	180
tttttttctt	caacataacg	agaacacaca	ggggcgctat	cgcacagaat	caaattcgat	240
gactggaaat	tttttgttaa	tttcagaggt	cgcctgacgc	atataccttt	ttcaactgaa	300
aaattgggag	aaaaaggaaa	ggtgagagcg	cgggaaccgg	cttttcatat	agaatagaga	360
agcgttcag	actaaatgct	tgcatacaca	tacttgaagt	tgacaatatt	atttaaggac	420
ctattgtttt	ttccaatagg	tggttagcaa	tcgtcttact	ttctaacttt	tcttaccttt	480
tacatttcag	caatatatat	atatatat	caaggatata	ccattctagc	tcacacgcgg	540
ccagggggag	cctactatt	attccataag	atgatcatta	gcattacgtt	caaaacagat	600
acaaataact	taagtaataa	cacgagccat	atgaccatta	gcaagatgac	aagcaagtta	660
agaccaatca	gcttccatca	tagcatcagc	ttaacgttca	ccattaataa	gagtagaaat	720
ttcaccttca	agacaataac	gattttcgtg	gtaataactg	atataattaa	attgaagctc	780
taattttgtga	gttttagtata	catgcattta	cttataatac	agtttttttag	ttttgctggc	840
cgcattcttct	caaatatgct	tcccagcctg	cttttctgta	acgttcaccc	tctaccttag	900
catcccttcc	ctttgcaaat	agtcctcttc	caacaataat	aatgtcagat	cctgtagaga	960
ccacatcatc	cacggtttcta	tactgttgac	ccaatgcgtc	tcccttgta	tctaaaccca	1020
caccgggtgt	cataatcaac	caatcgtaac	cttcatctct	tccacccatg	tctcttttag	1080
caataaagcc	gataacaaaa	tctttgtcac	tcttcgcaat	gtcaacagta	cccttagtat	1140
attctccagt	agctagggag	cccttgcatg	acaattctgc	taacatcaaa	aggcctctag	1200
gttcctttgt	tacttcttcc	gccgcctgct	tcaaaccgct	aacaatacct	gggccacca	1260
caccgtgtgc	attcgtaattg	tctgcccatt	ctgtatttct	gtataacccc	gcagagtact	1320
gcaatttgac	tgtattacca	atgtcagcaa	attttctgtc	ttcgaagagt	aaaaaattgt	1380
acttggcgga	taatgccttt	agcggcttaa	ctgtgccttc	catggaaaaa	tcagtcaaga	1440
tatccacatg	tgtttttagt	aaacaaattt	tgggacctaa	tgcttcaact	aactccagta	1500
attccttggt	ggttagcaaca	tccaatgaag	cacacaagtt	tgtttgcttt	tcgtgtatga	1560
tattaaatag	cttggcagca	acaggacatg	gatgagtgc	agcacgttcc	ttatgtatag	1620
ctttcgacat	gatttatctt	cgtttctctg	aggtttttgt	tctgtgcagt	tgggttaaga	1680
atactgggca	atttcatggt	tcttcaacac	cacatatg	tatatatacc	aatctaagtc	1740

[0034]

tgtgtctcctt	ccttcgttct	tccttctgtc	cggagattac	cgaatcaaaa	aaatttcaaa	1800
gaaaccggaa	tcaaaaaaaa	gaacaaaaaa	aaaaaagatg	aattgaaaag	ctttatggac	1860
cctgaaacca	ctcactatta	ttccataaga	tgatcattag	cattacgttc	aaaacgagta	1920
caaataactt	aagtaataac	acgagccata	tgaccattag	caagatgaca	agcaagttaa	1980
gaccaatcag	cttccatcat	agcatcagct	taacgttcac	cattaataag	agtagaaatt	2040
tcaccttcaa	gacaataacg	attttcgtgc	gctcgtccaa	cgccggcgga	cctgacggta	2100
gcaacaagaa	tatagcacga	gccgcgaggt	tcatttcgtt	acttttgata	tcactcacia	2160
ctattgcgaa	gcgcttcagt	gaaaaaatca	taaggaaaag	ttgtaaatat	tattggtagt	2220
attcgtttgg	taaagtagag	ggggtaattt	ttccccttta	ttttgttcat	acattcttaa	2280
attgctttgc	ctctcctttt	ggaaagctat	acttcggagc	actgttgagc	gaaggctcat	2340
tagataattt	ttctgtcatt	ttccttaacc	gaaaaggggt	gaaaaggggt	caaaaagcgc	2400
tcggacaact	gttgaccgtg	atccgaagga	ctggctatac	agtgttcaca	aaatagccaa	2460
gctgaaaata	atgtgtagct	atgttcagtt	agtttggtta	gcaaagatat	aaaagcaggt	2520
cggaaatatt	tatgggcatt	attatgcaga	gcatcaacat	gataaaaaaa	aacagttgaa	2580
tattccctca	aaaatccccg	cgtgcttggc	cggccgtaat	taataatgtc	aactttgcct	2640
atttcttctg	tgctattttc	ctcttctaca	tcaccattag	tcgtggacga	caaagtctca	2700
accaagcccc	acgttatcat	acatacaatg	attttcaatg	cttctatttg	gggagatcaa	2760
ttcttgacct	atgatgagcc	tgaagattta	gttatgaaga	aacaattagt	ggaggaatta	2820
aaagaggaag	ttaagaagga	attgataact	atcaaagggt	caaatgagcc	catgcagcat	2880
gtgaaattga	ttgaattaat	tgatgctggt	caacgtttag	gtatagctta	ccattttgaa	2940
gaagagatcg	aggaagcttt	gcaacatata	cattgttacct	atggtgaaca	gtgggtggat	3000
aaggaataatt	tacagagtat	ttcattgtgg	ttcaggttgt	tgctgcaaca	gggctttaac	3060
gtctcctctg	gcgttttcaa	agactttatg	gacgaaaaag	gtaaattcaa	agagtcttta	3120
tgcaatgatg	cacaaggaat	attagcctta	tatgaagctg	catttatgag	ggttgaagat	3180
gaaaccatct	tagacaatgc	tttggaattc	acaaaagttc	atntagatat	catagcaaaa	3240
gacccatctt	gcgattcttc	attgctgata	caaattccatc	aagccttaaa	acaaccttta	3300
agaaggagat	tagcaaggat	tgaagcatta	cattacatgc	caatctacca	acaggaacaa	3360
tctcatgatg	aagtattggt	gaaattagcc	aagttggatt	tcagtgtttt	gcagtctatg	3420
cataaaaagg	aattgtcaca	tatctgtaag	tggtggaaag	atntagattt	acaaaataag	3480
ttaccttatg	tacgtgatcg	tggtgtcgaa	ggctacttct	ggatattgtc	catatactat	3540
gagccacaac	acgctagaac	aagaatgttt	ttgatgaaaa	catgcatgtg	gttagtagtt	3600
ttggacgata	cttttgataa	ttatggaaca	tacgaagaat	tggagatttt	tactcaagcc	3660
gtcagagagat	ggcttatctc	atgcttagat	atgttgcccg	aatatatgaa	attatctac	3720
caagaattag	tcaatttgca	tggtgaaatg	gaagaatctt	tggaaaagga	gggaaagacc	3780
tatcagattc	attacgttaa	ggagatggct	aaagaattag	ttcgtaatga	cttagtagaa	3840
gcaagatggt	tgaaggaagg	ttatatgcct	actttagaag	aatacatgtc	tgtttctatg	3900
gttactggta	cttatggttt	gatgattgca	aggtcctatg	ttggcagagg	agacattggt	3960
actgaagaca	cattcaaatg	ggtttctagt	taccaccta	ttattaaagc	ttcctgtgta	4020
atagtaagat	taatggacga	tattgtatct	cacaaggaag	aacaagaaag	aggacatgtg	4080
gcttcatcta	tagaatgtta	ctctaaagaa	tcaggtgctt	ctgaagagga	agcatgtgaa	4140
tatattagta	ggaaagtga	ggatgcctgg	aaagtaatca	atagagaatc	tttgcgtcca	4200
acagccgttc	ccttcccttt	gttaattgcca	gcaataaaact	tagctagaat	gtgtgaggtc	4260
ttgtactctg	ttaatgatgg	ttttactcat	gctgagggtg	acatgaaatc	ttatatgaag	4320
tccttcttcg	ttcatcctat	ggtcgtttga	gctagctaa	atccgctcta	accgaaaagg	4380
aaggagttag	acaacctgaa	gtctagggtc	ctattttattt	ttttatagtt	atgtagtat	4440
taagaacggt	atttatatatt	caaatttttc	ttttttttct	gtacagacgc	gtgtacgcat	4500
gtaacattat	actgaaaacc	ttgcttgaga	aggtttttgg	acgctcgaag	aacctgcagg	4560
ccgcgagcgc	cgtaattcgc	gggtggaagg	acctgtgga	ggaatatgaa	gtttagatgt	4620
caaaggggatt	gaaatggctat	ttacttgatt	atgagtcatt	gtatcaagga	tactatggct	4680
tgcttaaatt	tcattctgtc	ttcgaaagct	gaattgatac	tacgaaaaat	ttttttttgt	4740
ttctctttct	atctttatta	cataaaactt	catacacagt	taagattaaa	aacaactaat	4800
aaataatgcc	tatcgcaaat	tagcttatga	agtccatggt	aaattcgtgt	ttcctggcaa	4860
taatagatcg	tcaatttggt	gctttgtggt	agtttttattt	tcaataaatt	ggaatactag	4920
ggatttgatt	tttagatctt	tattcaaatt	ttttgcgctt	aacaaacagc	agccagttcc	4980
acccaagtct	gtttcfaatg	tctcgtaact	aaaatcatct	tgcaattttct	ttttgaaact	5040
gtcaatttgc	tcttgagtaa	tgctctctcg	taacaaagtc	aaagagcaac	cgccgccacc	5100
agcaccggta	agttttgtgg	agccaattct	caaatcatcg	ctcagatttt	taataagttc	5160
taatccagga	tgagaaacac	cgattgagac	aagcagtcga	tgattttattc	ttatcaattc	5220
caatagttgt	tcatacagtt	cattattagt	ttctacagcc	tcgtcatcgg	tgcttttaca	5280
tttacttaac	ttagtcatga	tctctaagcc	ttgtagggca	cattcaccca	tggcatctag	5340
aattggcttc	ataacttcag	gaaattttctc	ggtgaccaac	acacgaacgc	gagcaacaag	5400
atcttttgta	gaccttgga	ttctagtata	ggttaggatc	attggaatgg	ctgggaaatc	5460
atctaagaac	ttaaaattgt	ttgtgtttat	tgttccatta	tgtagatcct	tttcaaatag	5520
cagggcatta	ccataagtgg	ccacagcgtt	atctattcct	gaaggggtac	cgtaaataga	5580
cttttcacct	atgaaggccc	attgattcac	tatatgctta	tcgttttctg	acagcttttc	5640
caagtcatta	gatcctatta	accccccaa	gtaggccata	gctaaggcca	gtgatacaga	5700
aatagaggcg	cttgagccca	accagcacc	gatgggtaaa	gtagacttta	aagaaaactt	5760

[0035]

aatattcttg	gcatgggggc	ataggcaaac	aaacatatac	aggaaacaaa	acgctgcatg	5820
gtagtgaag	gattcggata	gttgagctaa	caacggatcc	aaaagactaa	cgagttcctg	5880
agacaagcca	tcggtggctt	gttgagcctt	ggccaatttt	tgggagttta	cttgatcctc	5940
ggtgatggca	ttgaaatcat	tgatggacca	cttatgatta	aagctaattg	ccgggaagtc	6000
caattcaata	gtatctggtg	cagatgactc	gcttatttagc	aggtaggttc	tcaacgcaga	6060
cacactagca	gcgacggcag	gcttgttgta	cacagcagag	tgttcaccaa	aaataataac	6120
ctttcccggg	gcagaagtta	agaacggtaa	tgacatcggt	gtttaaaccc	cagcgcctgg	6180
cggg						6184

<210> 118

<211> 6042

<212> DNA

<213> 人工序列

<220>

<223> Phase III 标记循环缝合产物

<400> 118

gacggcacgg	ccacgcgttt	aaaccgccaa	aagtgcagct	cagagccccc	agcaccagta	60
ttagagggtca	taatgggctg	cgaagcctgc	taaaatgcag	tggaggccgt	gtaccctttg	120
ccaaattggc	tattggaatc	ggcagagaac	ctgggtcccg	ttctagagac	cctgcgagcg	180
tgtcccgggtg	ggttctggga	gctctaactc	cgcaggaact	acaaaccttg	cttacacaga	240
gtgaacctgc	tgcttggtg	gctctgactc	agtacatttc	atagcccatc	ttcaacaaca	300
ataccgactt	accatcctat	ttgctttgcc	ctttttcttt	tccactgcac	tttgcacg	360
aaggcggttat	cggtttttggg	tttagtgctt	aaacgagcag	cgagaacacg	accacgggct	420
atataaatgg	aaagttagga	caggggcaaa	gaataagagc	acagaagaag	agaaaagacg	480
aaaagcagaa	gcggaataac	tatacacgtc	acatatcaca	cacacacgct	cacacgcggc	540
caggggggagc	ctcactatta	ttccataaga	tgatcattag	cattacgttc	aaaacgagta	600
caaataactt	aagtaataac	acgagccata	tgaccattag	caagatgaca	agcaagttaa	660
gaccaatcag	cttccatcat	agcatcagct	taacggtcac	cattaataag	agtagaaatt	720
tcaccttcaa	gacaataacg	atcttctgtg	taataactga	tataattaaa	ttgaagctct	780
aatctgtgag	tttagtatac	atgcatttac	ttataataca	gttttttagt	tttgctggcg	840
gcattcttctc	aaatatgctt	cccagcctgc	ttttctgtaa	cgttcacctc	ctaccttagc	900
atcccttccc	tttgcaataa	gtcctcttcc	aacaataata	atgtcagatc	ctgtagagac	960
cacatcatcc	acgggttctat	actgttgacc	caatgcgtct	cccttgctat	ctaaaccac	1020
accgggtgtc	ataatcaacc	aatcgtaacc	ttcatctctt	ccacccatgt	ctctttgagc	1080
aataaagccg	ataacaaaat	ctttgtcact	cttcgcaatg	tcaacagtac	ccttagtata	1140
ttctccagta	gctagggagc	ccttgcatga	caattctgct	aacatcaaaa	ggcctctagg	1200
ttcctttggt	acttcttccg	ccgcctgctt	caaaccgcta	acaataacctg	ggccccacc	1260
accgtgtgca	ttcgtaattg	ctgcccattc	tgctattctg	tataaccccg	cagagtactg	1320
caatttgact	gtattacca	tgtcagcaaa	ttttctgtct	tcgaagagta	aaaaattgta	1380
cttggcggat	aatgccttta	gcggcttaac	tgtgccctcc	atggaaaaat	cagtcaagat	1440
atccacatgt	gttttttagta	aacaaatttt	gggaccta	gcttcaacta	actccagtaa	1500
ttccttggtg	gtacgaacat	ccaatgaagc	acacaagttt	gtttgctttt	cgtgcatgat	1560
attaaatagc	ttggcagcaa	caggactagg	atgagtagca	gcacgttcct	tatatgtagc	1620
tttcgacatg	atctatcttc	gtttcctgca	ggtttttggt	ctgtgcagtt	gggttaagaa	1680
tactgggcaa	tttcatgttt	cttcaacacc	acatatgcgt	atatatacca	atctaagtct	1740
gtgctctctc	cttcgttctt	ccttctgtct	ggagattacc	gaatcaaaaa	aatttcaag	1800
aaaccggaat	caaaaaaaag	aacaaaaaaa	aaaaagatga	attgaaaagc	tttatggacc	1860
ctgaaaccac	tcactattat	tccataagat	gatcattagc	attacgttca	aaacgagtac	1920
aaataactta	agtaataaca	cgagccatat	gaccattagc	aagatgacaa	gcaagttaag	1980
accaatcagc	ttccatcata	gcacagctt	aacgttcacc	attaataaga	gtagaaattt	2040
caccttcaag	acaataacga	ttttcgtgcg	ctcgtccaac	gccggcggac	ctgacggtag	2100
caacaagaat	atagcacgag	ccgcggagtt	catttcgtta	cttttgatat	cactcacaa	2160
tattgcaag	cgcttcagtg	aaaaaatcat	aaggaaaagt	tgtaaatatt	attggtagta	2220
ttcgtttggt	aaagtagagg	gggtaatttt	tcccctttat	tttgttcata	cattcttaaa	2280
ttgctttgcc	tctccttttg	gaaagctata	cttcggagca	ctgttgagcg	aaggctcatt	2340
agatatattt	tctgtcattt	tccttaaccc	aaaaataagg	gaaagggtcc	aaaaagcgct	2400
cggacaactg	ttgaccgtga	tccgaaggac	tggctataca	gtgttcacaa	aatagccaag	2460
ctgaaaataa	tgtgtagcta	tgttcagtta	gtttggctag	caaagatata	aaagcaggtc	2520
ggaaatattt	atgggcatta	ttatgcagag	catcaacatg	ataaaaaaaa	acagttgaat	2580
attccctcaa	aaatccccgc	gtgcttgccc	ggcgcgaatt	aataatgtca	actttgccta	2640
tttcttctgt	gtcatttttc	tcttctacat	caccattagt	cgtggacgac	aaagtctcaa	2700
ccagcccga	cgttatcaga	catacaatga	atttcaatgc	ttctatttgg	ggagatcaat	2760
tcttgaccta	tgatgagcct	gaagatttag	ttatgaagaa	acaatttagt	gaggaattaa	2820
aagaggaagt	taagaaggaa	ttgataacta	tcaaagggtc	aaatgagccc	atgcagcatg	2880
tgaattgat	tgaattaat	gatgctgttc	aacgttttag	tatagcttac	cattttgaag	2940

[0036]

aagagatcga	ggaagctttg	caacatatatac	atgttaccta	tggtgaacag	tggttgata	3000
aggaaaattt	acagagtatt	tcattgtggt	tcaggttgtt	gcgtcaacag	ggctttaacg	3060
tctcctctgg	cgttttcaaa	gactttatgg	acgaaaaagg	taaattcaaa	gagtccttat	3120
gcaatgatgc	acaaggaata	ttagccttat	atgaagctgc	atztatgagg	gttgaagatg	3180
aaaccatctt	agacaatgct	ttggaattca	caaaagttca	tttagataatc	atagcaaaag	3240
acccatcttg	cgattcttca	ttgcgtacac	aaatccatca	agccttaaaa	caacctttaa	3300
gaaggagatt	agcaaggatt	gaagcattac	attacatgcc	aatctacca	caggaaacat	3360
ctcatgatga	agtattgttg	aaattagcca	agttggattt	cagtgttttg	cagtctatgc	3420
ataaaaagga	attgtcacat	atctgtaagt	ggttgaaaga	tttagattta	caaaaataagt	3480
taccttatgt	acgtgatcgt	gttgctgaag	gctacttctg	gatattgtcc	atatactatg	3540
agccacaaca	cgtagaaca	agaatgtttt	tgatgaaaac	atgcatgttg	ttagtatttt	3600
tggaacgatac	ttttgataat	tatggaacat	acgaagaatt	ggagattttt	actcaagccg	3660
tcgagagatg	gtctatctca	tgcttagata	tggtgcccga	atatatgaaa	ttaatctacc	3720
aagaattagt	caatttgcat	gtggaaatgg	aagaatcttt	ggaaaaggag	ggaaagacct	3780
atcagattca	ttacgttaag	gagatggcta	aagaattagt	tcgtaattac	ttagtagaag	3840
caagatgggt	gaaggaagggt	tatatgccta	ctttagaaga	atacatgtct	gtttctatgg	3900
ttactggtac	ttatggtttg	atgattgcaa	ggtcctatgt	tggcagagga	gacattgtta	3960
ctgaagacac	attcaaatgg	gtttctagtt	acccacctat	tattaaagct	tcctgtgtaa	4020
tagtaagatt	aatggacgat	attgtatctc	acaaggaaga	acaagaaaga	ggacatgtgg	4080
cttcatctat	agaatgttac	tctaaagaat	caggtgcttc	tgaagaggaa	gcatgtgaat	4140
atattagttag	gaaagttgag	gatgcctgga	aagtaatcaa	tagagaatct	ttgcgtccaa	4200
cagccgttcc	cttccctttg	ttaatgccag	caataaactt	agctagaatg	gtgtaggctc	4260
tgtactctgt	taatgatgggt	tttactcatg	ctgagggtga	catgaaatct	tatatgaagt	4320
ccttcttctg	tcactctatg	gtcgtttgag	ctagctaaga	tccgctctaa	ccgaaaagga	4380
aggagttaga	caacctgaag	tctaggtccc	tatttatattt	tttatagtta	tgtagtattt	4440
aagaacgtta	tttatatttc	aaatttttct	tttttttctg	tacagacg	gtgtacgc	4500
taacattata	ctgaaaacct	tgcttgagaa	ggttttggga	cgctcgaaga	acctgcaggc	4560
cgcgagcgcc	gatcttgtgc	taagtgggtg	tgtagacag	ctacgaataa	ggaaattccg	4620
aagcatgtag	ggaggtcatg	atatgaaaaa	gcaaaagagt	aggcatcaaa	aagtttctca	4680
ttcaagtgggt	aactgctgtt	aaaattaaga	tatttataaa	ttgaagcttg	gtcgttccga	4740
ccaataccgt	agggaaacgt	aaattagcta	ttgtaaaaaa	aggaaaagaa	aagaaaagaa	4800
aaatgttaca	tctgaatttg	atcttattcc	tttggtagac	cagtctttgc	gtcaatcaaa	4860
gattcgtttg	ttcttgtgtg	gcctgaaccg	acttgagtta	aaatcactct	ggcaacatcc	4920
ttttgcaact	caagatccaa	ttcacgtgca	gtaaagttag	atgattcaaa	ttgatgggtg	4980
aaagcctcaa	gctgctcagt	agtaaatctc	ttgtcccatc	caggaaacaga	gccaaacaat	5040
ttatagataa	atgcaaagag	tttcgactca	ttttcagcta	agtagtacia	cacagcattt	5100
ggacctgc	caaacgtgta	tgcaacgatt	gtttctccgt	aaaactgatt	aatggtgtgg	5160
caccaactga	gtatacgctt	ggaagtgtca	ttcatgtaga	atattggagg	gaaagagtcc	5220
aaacatgtgg	catggaaaga	gttggaatcc	atcattgttt	cctttgcaaa	ggtggcgaaa	5280
tctttttcaa	caatggcttt	acgcatgact	tcaaactctc	ttggtacgac	atgttcaatt	5340
ctttcttttaa	atagttcgga	ggttgccacg	gtcaattgca	taccctgagt	ggaactcaca	5400
tcttttttaa	tatcgctgac	aactaggaca	caagctttca	tctgaggcca	gtcagagctg	5460
tctgcgattt	gtactgccat	ggaatcatga	ccatcttcag	cttttcccat	ttccaggcc	5520
acgtatccgc	caaacaacga	tctacaagct	gaaccagacc	cctttcttgc	tattctagat	5580
atitctgaag	ttgactgtgg	taattggtat	aacttagcaa	ttgcagagac	caatgcagca	5640
aagccagcag	cggaggaagc	taaaccagct	gctgtaggaa	agttattttc	ggagacaatg	5700
tggaagttcc	attgagataa	tggtggcaat	gaggcgctct	tcgattccat	ttcctttctt	5760
aattggcgta	ggtcgcgag	acaattttga	gttctttcat	tgctgatgct	gtgtggttct	5820
ccatttaacc	acaaagtgtc	gcgttcaaac	tcaggtgcag	tagccgcaga	ggtcaacgtt	5880
ctgaggtcat	cttgcgataa	agtcactgat	atggacgaat	tggtgggcag	attcaacttc	5940
gtgtcccttt	tcccccaata	cttaagggtt	gcgatgttga	cgggtgcggt	aacggatgct	6000
gtgtaaaccg	tcacgtgtgt	ttaaacccca	gcgcctggcg	gg		6042

<210> 119

<211> 5213

<212> DNA

<213> 人工序列

<220>

<223> STE5 敲除缝合产物

<400> 119

gacggcacgg	ccacgcgttt	aaaccgccac	gaagtgactg	acagaatact	gacatcagct	60
gatttctcat	agagctgttt	ctctgataac	acgttgtttg	aacatcgaca	agatgaaaat	120
ctagaagtat	caagtttctt	ttaaagggat	atataacaga	ttctaaaact	gacagaaata	180
tttcgagtga	agaagaagcg	ttaaataattg	gatctttccg	cagttctact	ctgatacatt	240
tttgaagtag	gagagtcatt	tagaaggcgt	attgtctaat	agtagaaagc	aggcctgtgc	300

[0037]

acatgaatta	attaaaaaat	ataaaggtag	tgatttagacg	acacatgtcc	ataggtaacc	360
tgtcataatt	ttgaacaatt	tcccttcttt	tctttttttt	ttttgggtgc	ggcgatatgt	420
agcttggttaa	tttacacatc	atgtactttt	ctgcatcaaa	atatgaaagg	cgatagtagc	480
taaagaaaaat	accgagaatt	tcctcgaaaa	gttgacgaca	aaagaaaggc	ataaaaaagt	540
aatttgaaaa	tattttaaaa	ctgttttaac	ccatctagca	tccgcgctaa	aaaaggaaga	600
tacaggatac	agcggaaaca	actttttaagc	tcacacgcgg	ccagggggag	ccatgcgtcc	660
atctttacag	tcctgtctta	ttgttcttga	tttgtgcccc	gtaaaatact	gttacttggg	720
tctggcgagg	tattggatag	ttccttttta	taaaggccat	gaagcttttt	ctttccaatt	780
tttttttttt	cgtcattata	gaaatcatta	cgaccgagat	tcccgggtaa	taactgatat	840
aatttaaattg	aagctctaatt	ttgtgagttt	agtatacatg	catttactta	taatacagtt	900
ttttagtttt	gctggccgca	tcttctcaaa	tatgcttccc	agcctgcctt	tctgtaacgt	960
tcaccctcta	ctttagcatc	ccttcccttt	gcaaatagtc	ctcttccaac	aataataatg	1020
tcagatcctg	tagagaccac	atcatccacg	gttctatact	gttgacccaa	tgcgtctccc	1080
ttgtcatcta	aaccacacac	gggtgtcata	atcaaccaat	cgtaaccttc	atctcttcca	1140
cccattgtctc	tttgagcaat	aaagccgata	acaaaatctt	tgctgcctct	cgcaatgtca	1200
acagatccct	tagtatattc	tccagtagct	agggagccct	tgcatgacaa	ttctgctaac	1260
atcaaaaaggc	ctctaggttc	ctttgttact	tcttccgccc	cctgcttcaa	accgctaaca	1320
atacctgggc	ccaccacacc	gtgtgcattc	gtaatgtctg	cccattctgc	tattctgtat	1380
acacccgcag	agtactgcaa	tttgactgta	ttaccaatgt	cagcaaatit	tctgtcttcg	1440
aagagtaaaa	aattgtactt	ggcggataat	gccttttagcg	gcttaactgt	gccctccatg	1500
gaaaaatacag	tcaagatatc	cacatgtgtt	tttagtaaac	aaattttggg	acctaagtct	1560
tcaactaact	ccagtaattc	cttgggtgta	cgaacatcca	atgaagcaca	caagtgtgtt	1620
tgcttttctgt	gcatgatatt	aaatagcttg	gcagcaacag	gactaggatg	agtagcagca	1680
cgttccttat	atgtagcttt	cgacatgatt	tatcttcgtt	tcctgcaggt	ttttgttctg	1740
tgagttggg	ttagaatac	tgggcaatit	catgtttctt	caacaccaca	tatgcgtata	1800
tataccaatc	taagtctgtg	ctccttctct	cgtttcttct	tctgctcgga	gattaccgaa	1860
tcaaaaaaat	ttcaagaaa	cgggaatcaa	aaaaaagaac	aaaaaaaaaa	aagatgaatt	1920
gaaaagcttt	atggaccctg	aaaccacagc	cacattaacc	ttctttgatg	gtcaaaactt	1980
atccttcacc	ataaatatgc	ctcgaaaaaa	aggtaattaa	catatataga	attacattat	2040
ttatgaaata	tcatactat	ctcttagcat	ctttaatcct	tttctacatc	agataacttc	2100
ggtttgttat	catcgtctgt	attgtcatca	attggcgcag	tagcctcaat	ttcaacgtcg	2160
tttgactctg	gtgtttgttc	atgtgcagat	ccatgagatg	atgaaccgct	cgccaacgc	2220
cgggcgacct	cttttaattc	tgctgtaacc	gcatactgcc	caaaataggg	ggcgggttac	2280
acagaatata	taacatcgta	gggtgtctggg	tgaacagttt	attcctggca	tccactaaat	2340
ataatggagc	ccgcttttta	agctggcatc	cagaaaaaaa	agaatccca	gcacaaaat	2400
attgttttct	tcaccaacca	tcagttcata	ggtccattct	cttagcgcaa	ctacagagaa	2460
caggggcaca	aacaggcaaa	aaacgggcac	aacctcaatg	gagtgatgca	acctgcctgg	2520
agtaaatgat	gacacaaggc	aattgaccca	cgcattgtatc	tatctcattt	tcttacacct	2580
tctattacct	tctgctctct	ctgatttggg	aaaagctgaa	aaaaaagggt	gaaaccagtt	2640
ccctgaaatt	attcccctac	ttgactaata	agtatataaa	gacggtaggt	attgattgta	2700
attctgtaaa	tctatttctt	aaacttctta	aattctactt	ttatagttag	tctttttttt	2760
agttttaaaa	accagaagac	ttagtttctg	ataaacacac	ataaacaaac	aaattaataa	2820
tggtctcgaga	ccaattggg	aagactgaag	tcaccaagaa	gtccttttact	gtctctgtac	2880
aaaaggcttc	tacaccagtt	ttaaccaata	aaacagtcac	ttctggatcg	aaagtcaaaa	2940
gtttatcatc	tgcgcaatcg	agctcatcag	gaccttcac	atctagttag	gaagatgatt	3000
cccgcgatat	tgaaagcttg	gataagaaaa	tacgtccttt	agaagaatta	gaagcattat	3060
taagtagtgg	aaatacaaaa	caattgaaga	acaaagaggt	cgctgccttg	gttattcacg	3120
gtaagttagc	ttgttagcgt	ttggagaaaa	aattagtgta	tactacgaga	gcggttgccg	3180
tacgtaggaa	ggctctttca	attttggcag	aagctcctgt	attagcatct	gacgttttct	3240
catataaaaa	ttatgactac	gaccgcgtat	ttggcgcttg	ttgtgaaaat	gttatagggt	3300
acatgccttt	gcccgttggg	gttataggcc	ccttgggtat	cgatggtaca	tcttatcata	3360
taccaatggc	aactacagag	ggttgttttg	tagcttctgc	catgcgtggc	tgtaaggcaa	3420
tcaatgctgg	cgggtgtgca	acaactgttt	taactaagga	tggtatgaca	agaggccag	3480
tagtccggtt	cccaactttg	aaaagatctg	gtgcctgtaa	gatatggtta	gactcagaag	3540
agggacaaaa	cgcaattaaa	aaagctttta	actctacatc	aagatttgca	cgtctgcaac	3600
atattcaaac	ttgtctagca	ggagatttac	tcttcatgag	atttagaaca	actactggtg	3660
acgcaatggg	tatgaatatg	atcttctaagg	gtgtcgaata	ctcattaaag	caaagtgtag	3720
aagagtatgg	ctgggaagat	atggaggttg	tctccgtttc	tggttaactac	tgtaccgaca	3780
aaaaaccagc	tgccatcaac	tggatcgaag	gtcgtggtaa	gagtgtcgtc	gcagaagcta	3840
ctattcctgg	tgatgttgtc	agaaaagtgt	taaaaagtga	tgtttccgca	ttggttgagt	3900
tgaacattgc	taagaatttg	gttggatctg	caatggctgg	gtctgttggg	ggatttaacg	3960
cacatgcagc	taatttagtg	acagctgttt	tcttggcatt	aggacaagat	cctgcacaaa	4020
atgtcgaaag	ttccaactgt	ataacattga	tgaaagaagt	ggacggtgat	ttgagaattt	4080
cgtatccat	gccatccatc	gaagttaggt	cgaatcgggtg	tggtactgtt	ctagaaccac	4140
aagggtgccat	gttggactta	ttaggtgtta	gaggccaca	tgctaccgct	cctggtacca	4200
acgcacgtca	attagcaaga	atagttgcct	gtgccgtctt	ggcaggtgaa	ttatccttat	4260
gtgctgccct	agcagccggc	catttgggtc	aaagtcatat	gaccacaaac	aggaaacctg	4320

[0038]

ctgaaccaac	aaaacctaac	aatttgagcg	ccactgatat	aaatcgtttg	aaagatgggt	4380
ccgtcacctg	cattaaatcc	taaacttagt	catacgcat	tggtattctc	ttgaaaaaga	4440
agcacaacag	caccatgtgt	tacgtaaaat	atttacttta	tagtttgtag	gtcataattt	4500
cttccatatt	acaagttcgt	gcataatag	aaagaattct	gttggtgtaa	ttgtcataac	4560
taacctgcag	gccgcgagcg	ccgatagtat	acactaaatt	ttatgcaata	ataaaaaaga	4620
agcatcccgc	caaacgtttc	gtaactacat	attggttacct	agtttgattc	cgtagaatttg	4680
aagtggacgc	agttcttctt	agtctttcaa	gttcatagta	aatagacatt	acccaaagat	4740
ctctttcctg	tcttgatctt	gccatgaaaa	ccattgactt	gccgctcaca	cctaactttt	4800
tagtgaaatg	gatcttatct	gtgttacttg	gaacattaga	gttattcaag	tcagcttccg	4860
gagcagaagg	tgggtcaata	ttattatctt	gtctaccata	atcttgcgta	tattgctttt	4920
cattcccctt	cttctgtaag	cgattgctag	aaagggcacg	ccttggtcca	aaccaaagcg	4980
taaagcatcg	cgaactctcg	tcttctacag	atcgccatcc	gtccccataa	actcgtggga	5040
gagcgtggga	tccatagttt	atttcgtcaa	aagtacgatc	tctctgcaga	aggtctagtt	5100
ctgttggtgt	cccagagtat	aggtagcaat	cgtagaatcg	aatggttaca	taatgagcat	5160
actccaaaac	ttccttagca	aaaccgggtg	tttaaacccc	agcgcctggc	ggg	5213

<210> 120

<211> 5417

<212> DNA

<213> 人工序列

<220>

<223> IME1 敲除缝合产物

<400> 120

gacggcacgg	ccacgcgttt	aaaccgccat	ccaattcctc	tattatatgc	atacattttt	60
ttgtttgatt	atctatgctg	tttataatat	cgtatatagt	taatgaataa	ccctacaaca	120
cgaagggcag	taatataatc	tgattctcag	ttgaaatttc	aaatttttca	agctgggtctt	180
actcggcagg	taggaacttc	ccagtgggtc	tgagttttct	ctccggaagg	tatttatcat	240
tatgcacgac	aagtgcgatt	caatcgaaag	attataaatt	tcgcatgtaa	caacatctga	300
taaaaaaaaa	attaaaaaaaa	aataatctaa	atgatatgta	tgggttaaaa	aggatgtatg	360
gatggagggt	tggcataaaa	atgaaaggca	gaagatgggc	ggctaaaaag	gtatagatat	420
cgtagctcac	cgtcacaaaa	atcactcata	gaatgcacta	tcacccctac	aataccatac	480
cagccgcaag	aaaaaaaaga	aaaaaaatca	attcatatca	tatattatct	atatcatgct	540
gttctttccg	ccacgctcac	acgcggccag	ggggagcccc	ttgagccatt	agtatcaatt	600
tgcttacctg	tattccttta	ctatcctcct	ttttctcctt	cttgataaat	gtatgtagat	660
tgcgtatata	gtttcgtcta	ccctatgaac	atattccatt	ttgtaatttc	gtgtcgtttc	720
tattatgaat	ttcattttata	aagtttatgt	acaaatatca	taaaaaaaga	gaatcttttt	780
aagcaaggat	tttcttaact	tcttcggcga	cagcatcacc	gacttcgggtg	gtactgttgg	840
aaccacctaa	atcaccagtt	ctgatacctg	catccaaaac	ctttttaact	gcattctcaa	900
tggccttacc	ttcttcaggc	aagttcaatg	acaatttcaa	catcattgca	gcagacaaga	960
tagtggcgat	aggggtcaacc	ttattctttg	gcaaatcttg	agcagaaccg	tggcatgggt	1020
cgtacaacc	aaatgcgggtg	ttcttgcttg	gcaagaggcg	caaggacgca	gatggcaaca	1080
aacccaagga	acctgggata	acggaggcct	catcgagat	gatataacca	aacatgttgc	1140
tggtgattat	aataccattt	agggtgggtg	ggttcttaac	taggatcatg	gcggcagaat	1200
caatcaattg	atgttgaacc	ttcaatgtag	ggaattcggt	cttgatgggt	tcctccacag	1260
tttttctcca	taatcttgaa	gaggccaaaa	gattagcttt	atccaaggac	caaataaggca	1320
atggtggctc	atgttgtagg	gccatgaaag	cgccattctt	tgtgattctt	tgcattcttg	1380
gaacgggtga	ttgttcaacta	tcccaagcga	caccatcacc	atcgtcttcc	tttctcttcc	1440
caaagtaaat	acctccact	aattctctga	caacaacgaa	gtcagtacct	ttagcaaat	1500
gtggcttgat	tggagataag	tctaaaagag	agtcggatgc	aaagttacat	ggtcttaagt	1560
tggcgtagaa	ttgaagttct	ttacggattt	ttagtaaac	ttgttcagg	ctaactac	1620
cgttaccaca	tttaggacca	gccacagcac	ctaacaaaac	ggcatcaacc	ttcttgagg	1680
cttccagcgc	ctcatctgga	agtgggacac	ctgtagcatc	gatagcagca	ccaccaatta	1740
aatgattttc	gaaatcgaa	ttgacatttg	aacgaacatc	agaaatagct	ttaagaacct	1800
taatggcttc	ggctgtgatt	tcttgaccaa	cgtggtcacc	tggcaaaacg	acgatcttct	1860
taggggcaga	cataggggca	gacattagaa	tggtatatcc	ttgaaatata	tatatatatt	1920
gctgaaatgt	aaaaggtaag	aaaagttaga	aagtaagacg	attgctaacc	acctattgga	1980
aaaaacaata	ggctcttaaa	taattattgt	aacttcaagt	attgtgatgc	aagcatttag	2040
tcatgaacgc	ttctctattc	tatatgaaaa	gccggttccg	gcctctcacc	tttctttttt	2100
ctcccaattt	ttcagttgaa	aaaggtatat	gcgtcaggcg	acctctgaaa	ttaacaaaaa	2160
atttccagtc	atcgaatttg	attctgtgcg	atagcgcccc	tgtgtgttct	cgttatgttg	2220
aggaaaaaaa	taatggttgc	taagagattc	gaactcttgc	atcttacgat	acctgagtat	2280
ttccacagtt	acctgcggtc	aagatatatt	ttgaatcagg	cgcctcgctc	gtccaacgac	2340
ggcggacctc	ttttaattct	gctgtaaccc	gtacatgccc	aaaaataggg	gcgggttaca	2400
cagaatatat	aacatcgtag	gtgtctgggt	gaacagttta	ttcctggcat	ccactaaata	2460
taatggagcc	cgctttttta	gctggcatcc	agaaaaaaa	agaatccag	caccaaata	2520

[0039]

ttgttttctt	caccaaccat	cagttcatag	gtccattctc	ttagcgcaac	tacagagaac	2580
aggggcacaa	acaggcaaaa	aacgggcaca	acctcaatgg	agtgatgcaa	cctgcctgga	2640
gtaaattgatg	acacaaggca	attgaccac	gcattgtatct	atctcatttt	cttacacctt	2700
ctattacctt	ctgtctcttc	tgattttgaa	aaagctgaaa	aaaaagggtt	aaaccagttc	2760
cctgaaatta	ttcccctact	tgactaataa	gtatataaag	acggtaggta	ttgattgtaa	2820
ttctgtaaat	ctattttcta	aactttctta	attctacttt	tatagttagt	ctttttttta	2880
gttttaaaac	accaagaact	tagtttcgaa	taaacacaca	taaacaaaca	aattaataat	2940
gtcaactttg	cctattttct	ctgtgtcatt	ttcctcttct	acatcaccat	tagtcgtgga	3000
cgacaaagtc	tcaaccaagc	ccgacgttat	cagacatata	atgaatttca	atgcttctat	3060
ttggggagat	caattcttga	cctatgatga	gcctgaagat	ttagttatga	agaaacaatt	3120
agtggaggaa	ttaaagagg	aagtttaaga	ggaattgata	actatcaaag	gttcaaata	3180
gccccatgag	catgtgaaat	tgattgaatt	aattgatgct	gttcaacgtt	taggtatagc	3240
ttaccatttt	gaagaagaga	tcgaggaagc	tttgcaacat	atacatgtta	cctatgggtg	3300
acagtgggtg	gataaggaaa	atttacagag	tatttcattg	tggttcaggt	tggtgcgtca	3360
acagggtctt	aacgtctcct	ctggcggttt	caaagacttt	atggacgaaa	aaggtaaatt	3420
caaagagtc	ttatgcaatg	atgcacaagg	aataattagcc	ttatatgaag	ctgcatttat	3480
gaggggttga	gatgaaacca	tcttagacaa	tgcttttgaa	ttcacaaaag	ttcattttaga	3540
tatcatagca	aaagacccat	cttgcgattc	ttcattgctg	acacaaatcc	atcaagcctt	3600
aaaacaacct	ttaagaagga	gatttagcaag	gattgaagca	ttacattaca	tgccaatcta	3660
ccaacaggaa	acatctcatg	atgaagtatt	gttgaaatta	gccaaagtgg	atttcagtgt	3720
tttgacgtct	atgcataaaa	aggaattgtc	acatatctgt	aagtgggtga	aagatttaga	3780
tttacaaaat	aagttacctt	atgtacgtga	tcgtgttgct	gaaggctact	tctggatatt	3840
gtccatatac	tatgagccac	aacacgctag	aacaagaatg	tttttgatga	aaacatgcat	3900
gtggttagta	gttttgacg	atacttttga	taattatgga	acatacgaag	aattggagat	3960
ttttactcaa	gccgtcgaga	gatggtctat	ctcatgctta	gatatgttgc	ccgaatatat	4020
gaaattaatc	taccaagaat	tagtcaattt	gcattgtgga	atggaagaat	ctttggaaaa	4080
ggagggaaag	acctatcaga	ttcattacgt	taaggagatg	gctaaagaat	tagttcgtaa	4140
ttacttagta	gaagcaagat	ggttgaagga	aggttatatg	cctacttttag	aagaatacat	4200
gtctgtttct	atggttactg	gtacttatgg	tttgatgatt	gcaaggctct	atgttggcag	4260
aggagacatt	gttactgaag	acacattcaa	atgggtttct	agttaccac	ctattattaa	4320
agcttctctg	gtaatatgta	gattaatgga	cgatattgta	tctcacaaag	aagaacaaga	4380
aagaggacat	gtggcttcac	ctatagaatg	ttactctaaa	gaatcagggt	cttctgaaga	4440
ggaagcatgt	gaatatatta	gtaggaaagt	tgaggatgcc	tggaagtaaa	tcaatagaga	4500
atctttgcgt	ccaacagccg	ttcccttccc	tttgttaatg	ccagcaataa	acttagctag	4560
aatgtgtgag	gtcttgtact	ctgttaatga	tggttttact	catgctgagg	gtgacatgaa	4620
atcttatatg	aagtccttct	tcgttcattc	tatggtcggt	tgagctagct	aagatccgct	4680
ctaaccgaaa	aggaaggagt	tagacaacct	gaagcttagg	tccctattta	tttttttata	4740
gttatgttag	tattaagaac	gttattttata	tttcaaattt	ttcttttttt	tctgtacaga	4800
cgcggtgtacg	catgtaacat	tatactgaaa	accttgcttg	agaaggtttt	gggacgctcg	4860
aagaacctgc	aggccgagag	cgccgatctc	gaaaagtaact	acaatcttcc	cccttccctc	4920
aaaatatatc	cattcacact	catttcttta	tttccattgt	ctcactcaaa	ttgctaagaa	4980
tttggtgtatt	tgcataatata	tatatattat	ataggtatat	atatgtattc	aatgtctcaa	5040
gttccatgac	ataataccgt	agcggttatta	gccttatcgt	atgtcgcgat	gggaaaggag	5100
attcggtttta	atcttgaaaa	accttcgtag	cgaataatgc	gacataaatc	ttgagagagt	5160
acatcaccaa	attcactttg	ttaaaccgca	ccatcggtgc	ttgcattctt	attccttttg	5220
cctacactaa	aattatttagc	atcttcttaa	atgagctcca	gtgaagacga	agacgacaag	5280
ttcttgtatg	gttccgactc	cgaatttagca	ctaccttcat	ctaaacgatc	aagagatgat	5340
gaagcagacg	caggtgcgtc	cagtaatcct	gatatagtta	aaaggcaaac	ggtgttttaa	5400
ccccagcgcc	tggcggg					5417

<210> 121

<211> 1737

<212> DNA

<213> 人工序列

<220>

<223> *Artemisia annua*的 β -法尼烯合成酶, 密码子优化以在
*Saccharomyces cerevisiae*中表达

<400> 121

ggatccatgt	caactttgccc	tattttcttct	gtgtcatttt	cctcttctac	atcaccatta	60
gtcgtggacg	acaaagtctc	aaccaagccc	gacgttatca	gacatacaat	gaatttcaat	120
gcttctattt	ggggagatca	attcttgacc	tatgatgagc	ctgaagattt	agttatgaag	180
aaacaattag	ttgaggaatt	aaaagaggaa	gttaagaagg	aattgataac	tatcaaagggt	240
tcaaatgagc	ccgtgcagca	tgtgaaattg	attgaattaa	ttgatgctgt	tcaacgttta	300
ggtatagctt	accattttga	agaagagatc	gaggaagctt	tgcaacatat	acatgttacc	360
tatggtgaac	agtgggtgga	taaggaaaat	ttacagagta	tttcattgtg	gttcagggtg	420

[0040]

ttgcgtcaac	agggcctttaa	cgtctcctct	ggcgtttttca	aagactttat	ggacgaaaaa	480
ggtaaattca	aagagtcttt	atgcaatgat	gcacaaggaa	tattagcctt	atatgaagct	540
gcatttatga	gggttgaaga	tgaaccatc	ttagacaatg	ctttggaatt	cacaaaagtt	600
catttagata	tcatagcaaa	agaccatct	tgcgattctt	cattgcgtac	acaaatccat	660
caagccttaa	aacaaccttt	aagaaggaga	ttagcaagga	ttgaagcatt	acattacatg	720
ccaatctacc	aacaggaaac	atctcatgat	gaagtattgt	tgaaattagc	caagttggat	780
ttcagtgttt	tgcagtctat	gcataaaaag	gaattgtcac	atatctgtaa	gtggtggaaa	840
gatttagatt	tacaaaataa	gttaccttat	gtacgtgatc	gtgttgctga	aggctacttc	900
tggatattgt	ccatatacta	tgagccacaa	cacgctagaa	caagaatggt	tttgatgaaa	960
acatgcatgt	ggttagtagt	tttggacgat	acttttgata	attatggaac	atacgaagaa	1020
ttggagattt	ttactcaagc	cgctcagaga	tggtctatct	catgcttaga	tatgttgccc	1080
gaatatatga	aattaatcta	ccaagaatta	gtcaatttgc	atgtggaaat	ggaagaatct	1140
ttgaaaagg	agggaaagac	ctatcagatt	cattacgtta	aggagatggc	taaagaatta	1200
gttcgtaatt	acttagtaga	agcaagatgg	ttgaaggaa	gttatatgcc	tacttttagaa	1260
gaatacatgt	ctgtttctat	ggttactggg	acttatgggt	tgatgattgc	aaggctcctat	1320
gttggcagag	gagacattgt	tactgaagac	acattcaaat	gggtttctag	ttaccacctt	1380
attattaaag	cttcctgtgt	aatagtaaga	ttaatggacg	atattgtatc	tcacaaggaa	1440
gaacaagaaa	gaggacatgt	ggcttcatct	atagaatgtt	actctaaaga	atcaggtgct	1500
tctgaagagg	aagcatgtga	atatattagt	aggaaagttg	aggatgcctg	gaaagtaatc	1560
aatagagaat	ctttgcgtcc	aacagccgtt	cccttcctt	tgtaaatgcc	agcaataaac	1620
ttagctagaa	tgtgtgaggt	cttgtactct	gttaatgatg	gttttactca	tgctgagggg	1680
gacatgaaat	cttatatgaa	gtccttcttc	gttcaccta	tggtcgtttg	actcgag	1737

<210> 122

<211> 7348

<212> DNA

<213> 人工序列

<220>

<223> 载体 pAM178

<400> 122

tcgcgcgttt	cgggtgatgac	ggtgaaaacc	tctgacacat	gcagctcccg	gagacgggtca	60
cagcttgtct	gtaagcggat	gccgggagca	gacaagcccg	tcagggcgcg	tcagcgggtg	120
ttggcgggtg	tcggggctgg	cttaactatg	cggcatcaga	gcagattgta	ctgagagtgc	180
accatatcga	ctacgtcgta	aggccgtttc	tgacagagta	aaattcttga	gggaactttc	240
accattatgg	gaaatgcttc	aagaaggat	tgacttaaac	tccatcaaat	ggtcagggtca	300
ttgagtggtt	tttatttggt	gtattttttt	tttttttagag	aaaatcctcc	aatatcaaat	360
taggaatcgt	agtttcatga	ttttctgtta	cacctaaact	tttgtgtggt	gccctcctcc	420
ttgtcaatat	taatgtttaa	gtgcaattct	ttttccttat	cacgttgagc	cattagtatc	480
aatttgctta	cctgtattcc	tttactatcc	tcctttttct	ccttcttgat	aaatgtatgt	540
agattgcgta	tatagtttct	tctaccctat	gaacatatcc	cattttgtaa	ttcgtgtctg	600
tttctattat	gaatttcatt	tataaagatt	atgtacaaat	atcataaaaa	aagagaatct	660
ttttaagcaa	ggattttctt	aacttcttcg	gcgacagcat	caccgacttc	ggtgggtactg	720
ttggaaccac	ctaaatcacc	agttctgata	cctgcatcca	aaaccttttt	aactgcatct	780
tcaatggcct	taccttcttc	aggcaagttc	aatgacaatt	tcaacatcat	tgcagcagac	840
aagatagttg	cgatagggtc	aaccttattc	tttggcaaat	ctggagcaga	accgtggcat	900
ggttcgtaca	aaccaaatgc	ggtgttcttg	tctggcaaa	aggccaagga	cgcagatggc	960
aacaaaccca	aggaacctgg	gataacggag	gcttcatcgg	agatgatatc	accaaactatg	1020
ttgctggtga	ttataatacc	atttaggtgg	gttgggttct	taactaggat	catggcggca	1080
gaatcaatca	attgatgttg	aaccttcaat	gtagggaatt	cgttcttgat	ggtttcctcc	1140
acagtttttc	tccataatct	tgaagaggcc	aaaagattag	ctttatccaa	ggaccaataa	1200
ggcaatggtg	gctcatgttg	tagggccatg	aaagcggcca	ttcttgatgt	ttcttgactt	1260
tctggaacgg	tgtattgttc	actatcccaa	gcgacaccat	caccatcgtc	ttcctttctc	1320
ttaccaaaag	aaatacctcc	cactaattct	ctgacaacaa	cgaagtcagt	accttttagca	1380
aattgtggct	tgattggaga	taagtctaaa	agagagtcgg	atgcaaagtt	acatggtctt	1440
aagttggcgt	acaattgaag	ttctttacgg	attttttagta	aaccttggtc	aggtctaaca	1500
ctaccggtac	cccatttagg	accagccaca	gcacctaaca	aaacggcatc	aaccttcttg	1560
gaggcttcca	gcgcctcatc	tggaagtgga	acacctgtag	catcgatagc	agcaccacca	1620
attaaatgat	tttcgaaatc	gaacttgaca	ttggaacgaa	catcagaat	agctttaaga	1680
accttaatgg	cttcggctgt	gatttcttga	ccaacgtggg	cacctggcaa	aacgacgatc	1740
ttcttagggg	cagacattac	aatggtatat	ccttgaaata	tatataaaaa	aaggcgcctt	1800
agaccgctcg	gccaaacaac	caattacttg	ttgagaaata	gagtataatt	atcctataaa	1860
tataacgttt	ttgaacacac	atgaacaagg	aaagtacagga	caattgattt	tgaagagaat	1920
gtggattttg	atgtaattgt	tgggattcca	tttttaataa	ggcaataata	ttaggtatgt	1980
ggatatacta	gaagttcttc	tcgaccgtcg	atatgcgggtg	tgaaataaccg	cacagatgcg	2040
taaggagaaa	ataccgcatc	aggaaattgt	aaacgttaat	attttgtaa	aattcgcgtt	2100

[0041]

aaattttt	gt	taaatcagct	catttttt	taa	ccaataggcc	gaaatcggca	aaatccctta	2160
taaatcaaaa	gaatagaccg	agatagggtt	gagtggtgtt	ccagtttggg	acaagagtcc	2220		
actattaaag	aacgtggact	ccaacgtcaa	agggcgaaaa	accgtctatc	agggcgatgg	2280		
cccactacgt	gaaccatcac	cctaatacaag	ttttttgggg	tcgaggtgcc	gtaaagcact	2340		
aaatcggaa	cctaaaggga	gccccgatt	tagagcttga	cggggaaagc	cggcgaacgt	2400		
ggcgagaaag	gaaggggaaga	aagcgaaagg	agcgggcgt	agggcgctgg	caagtgtagc	2460		
ggtcacgctg	cgcgtaacca	ccacaccgc	cgcgctta	gcgccgtac	agggcgctc	2520		
gcgccattcg	ccattcaggc	tgcgcaactg	ttgggaagg	cgatcggtgc	gggcctctc	2580		
gctattacgc	cagctgaatt	ggagcgacct	catgctatac	ctgagaaagc	aacctgacct	2640		
acaggaaaga	gttactcaag	aataagaatt	ttcgttttaa	aacctaagag	tcactttaaa	2700		
atttgtat	acttattttt	tttataaact	atttaataat	aaaaatcata	aatacaaga	2760		
aattcgctta	tttagaagtg	tcaacaacgt	atctaccaac	gatttgacc	ttttccatct	2820		
tttcgtaaat	ttctggcaag	gtagacaagc	cgacaacctt	gattggagac	ttgaccaa	2880		
ctctggcgaa	gaattgttaa	ttaagagctc	agatcttatc	gtcgtcatcc	ttgtaatcca	2940		
tcgatactag	tgcggccgcc	cttttagtgag	ggttgaattc	gaattttcaa	aaattcttac	3000		
tttttttttg	gatggacgca	aagaagttaa	ataactatat	tacatggcat	taccaccata	3060		
tacatatcca	tatacatatc	catatcta	cttacttata	tgttgtggaa	atgtaaagag	3120		
ccccattatc	ttagcctaaa	aaaaccttct	ctttggaact	ttcagtaata	cgcttaactg	3180		
ctcattgcta	tattgaagta	cggattagaa	gccgccgagc	gggtgacagc	cctccgaagg	3240		
aagactctcc	tccgtgcgtc	ctcgtcttca	cgggtcgcgt	tcctgaaacg	cagatgtgcc	3300		
tcgcgccgca	ctgctccgaa	caataaagat	tctacaatac	tagcttttat	ggtatgaag	3360		
aggaaaaatt	ggcagtaacc	tggccccaca	aaccttcaaa	tgaacgaatc	aaattaacaa	3420		
ccataggatg	ataatgcgat	tagtttttta	gccttatttc	tggggtaatt	aatcagcgaa	3480		
gcgatgattt	ttgatctatt	aacagatata	taaatacaaa	aactgcataa	ccactttaac	3540		
taatactttc	aacattttcg	gtttgtatta	cttcttattc	aaatgtaata	aaagtatcaa	3600		
caaaaaattg	ttataatacc	tctatacttt	aacgtcaagg	agaaaaaacc	cggatccgt	3660		
aatacgactc	actatagggc	cggggcgctg	acatggaaca	gaagttgatt	tccgaagaag	3720		
acctcgagta	agcttggtac	cgcggctagc	taagatccgc	tctaaccgaa	aaggaggag	3780		
ttagacaacc	tgaagtctag	gtccctattt	atttttttat	agttatgtta	gtattaagaa	3840		
cgttattttat	atttcaaat	tttctttttt	ttctgtacag	acgcgtgtac	gcagtgaaca	3900		
ttatactgaa	aaccttgctt	gagaagggtt	tgggacgctc	gaagatccag	ctgcattaat	3960		
gaatcggcca	acgcgcgggg	agaggcggtt	tgcgtattgg	gcgctcttcc	gcttccctcg	4020		
tcactgactc	gtcgcgtcgc	gtcgttcggc	tgcggcgagc	ggatcacgct	cactcaaaag	4080		
cggtaatatc	gttatccaca	gaatcagggg	ataacgcagg	aaagaacatg	tgagcaaaag	4140		
gccagcaaaa	ggccaggaac	cgtaaaaagg	ccgcgttgct	ggcgtttttc	cataggctcc	4200		
gcccccttga	cgagcatcac	aaaaatcgac	gctcaagtca	gaggtggcga	aaccgcagag	4260		
gactataaag	ataccaggcg	tttccccctg	gaagctccct	cgtgcgctct	cctgttccga	4320		
ccctgccgct	taccggatac	ctgtccgcct	tttcccttcc	gggaagcggtg	gcgctttctc	4380		
atagctcacg	ctgtaggtat	ctcagttcgg	tgtaggtcgt	tcgctccaag	ctgggctgtg	4440		
tgcacgaacc	ccccgttcag	cccgaccgct	gcgccttattc	cggtaaactat	cgctttgagt	4500		
ccaacccggt	aagacacgac	ttatcgccac	tggcagcagc	cactggtaac	aggattagca	4560		
gagcgaggta	tgtaggcggt	gctacagagt	tcttgaagtg	gtggcctaac	tacggctaca	4620		
ctagaaggac	acttattgggt	atctgcgctc	tgtgtaagcc	agttaccttc	ggaaaaagag	4680		
ttggtagctc	ttgatccggc	aaacaaacca	ccgctggtag	cggtggtttt	tttgtttgca	4740		
agcagcagat	tacgcgcaga	aaaaaaggat	ctcaagaaga	tcctttgatc	ttttctacgg	4800		
ggctctgacgc	tcagtggaa	gaaaactcac	gttaagggat	tttggtcatg	agattatcaa	4860		
aaaggatctt	cacctagatc	cttttaaat	aaaaatgaag	ttttaaatca	atctaagta	4920		
tatatagta	aaccttgggt	gacagttacc	aatgctttaa	cagtgaaggca	cttatctcag	4980		
cgatctgtct	atttcggttca	tccatagttg	cctgactccc	cgctcgttag	ataactacga	5040		
tacgggaggg	cttaccatct	ggccccagtg	ctgcaatgat	accgcgagac	ccacgctcac	5100		
cggctccaga	tttatcagca	ataaaccagc	cagccggaag	ggccgagcgc	agaagtggtc	5160		
ctgcaacttt	atccgcctcc	atccagtcta	ttaatgtgtg	ccgggaagct	agagtaagta	5220		
gttcgccagt	taatagtttg	cgcaacgttg	ttgccattgc	tacaggcatc	gtggtgtcac	5280		
gctcgtcggt	tgggtatggc	tcattcagct	ccggttccca	acgatcaagg	cgagttacat	5340		
gatcccccat	gttgtgcaaa	aaagcgggtta	gctccttcgg	tcctccgatc	gttgtcagaa	5400		
gtaagttggc	cgcagtgtta	tactcatgg	ttatggcagc	actgcataat	tctcttactg	5460		
tcatgccatc	cgtaagatgc	ttttctgtga	ctggtgagta	ctcaaccaag	tcattctgag	5520		
aatagtgtat	gcggcgaccg	agttgctctt	gccccgcgtc	aatacgggat	aataccgcgc	5580		
cacatagcag	aactttaaaa	gtgctcatca	ttggaaaacg	ttcttcgggg	cgaaaactct	5640		
caaggatctt	accgctgttg	agatccagtt	cgatgtaacc	cactcgtgca	cccaactgat	5700		
cttcagcatc	ttttactttc	accagcgttt	ctgggtgagc	aaaaacagga	aggcaaaatg	5760		
ccgcaaaaaa	gggaataagg	gcgacacgga	aatgttgaa	actcatactc	ttcctttttc	5820		
aatattattg	aagcatttat	cagggttatt	gtctcatgag	cggatacata	tttgaaatga	5880		
tttagaaaaa	taaacaaata	ggggttccgc	gcacatttcc	ccgaaaagtg	ccacctgaac	5940		
gaagcatctg	tgcttcattt	tgtagaacia	aaatgcaacg	cgagagcgct	aatttttcaa	6000		
acaaagaatc	tgagctgcat	ttttacagaa	cagaaatgca	acgcgaaagc	gctattttac	6060		
caacgaagaa	tctgtgcttc	atttttgtaa	aacaaaaatg	caacgcgaga	gcgctaattt	6120		

[0042]


```

ttcaaacaaa gaatctgagc tgcattttta cagaacagaa atgcaacgcg agagcgctat 6180
tttaccaaca aagaatctat acttcttttt tgttctacaa aaatgcatcc cgagagcgct 6240
atTTTTttaa caaagcatct tagattactt tttttctcct ttgtgcgctc tataatgcag 6300
tctcttgata acttttttgca ctgtagggtc gttaagggtta gaagaaggct actttgggtg 6360
ctattttctc ttccataaaa aaagcctgac tccacttccc gcgtttactg attactagcg 6420
aagctgcggg tgcatttttt caagataaag gcatccccga ttatattcta taccgatgtg 6480
gattgcgcat actttgtgaa cagaaagtga tagcgttgat gattcttcat tggtcagaaa 6540
attatgaacg gtttcttcta ttttgtctct atatactacg tataggaaat gtttacattt 6600
tcgtattggt ttcgattcac tctatgaata gttcttacta caattttttt gtctaaagag 6660
taatactaga gataaacata aaaaatgtag aggtcagagt tagatgcaag ttcaaggagc 6720
gaaagggtga tgggtagggt atatagggat atagcacaga gatatatagc aaagagatac 6780
ttttgagcaa tgtttgtgga agcgggtatt gcaatatttt agtagctcgt tacagtccgg 6840
tgcgtttttg gttttttgaa agtgcgtctt cagagcgctt ttggttttca aaagcgctct 6900
gaagttccta tactttctag agaataggaa cttcggaata ggaacttcaa agcgtttccg 6960
aaaacgagcg cttccgaaaa tgcaacgcga gctgcgcaca tacagctcac tgttcacgct 7020
gcacctatat ctgctgtgtt cctgtatata tatatacatg agaagaacgg catagtgcgt 7080
gtttatgctt aaatgcgtac ttatatgcgt ctatttatgt aggatgaaag gtagtctagt 7140
acctcctgtg atattatccc attccatgcg gggtatcgta tgcttccttc agcactaccc 7200
tttagctgtt ctatatgctg ccactcctca attggattag tctcatcctt caatgctatc 7260
atttcctttg atattggatc atactaagaa accattatta tcatgacatt aacctataaa 7320
aataggcgta tcacgaggcc ctttcgctc

```

<210> 123

<211> 746

<212> DNA

<213> 人工序列

<220>

<223> URA3 敲除构建体

<400> 123

```

tgcgaggcat atttatggtg aaggataagt tttgaccatc aaagaagggtt aatgtggctg 60
tggtttcagg gtccataaag cttttcaatt catctttttt ttttttggtc ttttttttga 120
ttccggtttc tttgaaattt ttttgattcg gtaatctccg agcagaagga agaacgaagg 180
aaggagcaca gacttagatt ggtatatata cgcatatgtg gtgttgaaga aacatgaaat 240
tgcccagtat tcttaaccca actgcacaga acaaaaaacct gcaggaaacg gctcacacgc 300
ggccaggggg agccctgtat tataagtaaa tgcatgtata ctaaactcac aaattagagc 360
ttcaattttaa ttatatcagt tattaccctg gaatctcggt cgtaatgatt tctataatga 420
cgaaaaaaaa aaaattggaa agaaaaagct tcatggcctt tataaaaaagg aactatccaa 480
tacctgcga gaaccaagta acagtatttt acggggcaca aatcaagaac aataagacag 540
gactgtaaag atggacgcat tgaactccaa agaacaacaa gagttccaaa aagtagtgga 600
acaaaagcaa atgaaggatt tcatgcgttt gtactctaatt ctggtagaaa gatgtttcac 660
agactgtgtc aatgacttca caacatcaaa gctaaccaat aaggaacaaa catgcatcat 720
gaagtgtcga gaaaagttct tgaagc

```

<210> 124

<211> 27

<212> DNA

<213> 人工序列

<220>

<223> 引物 JU-106-168-110-ERG9 CDS-f

<400> 124

atgggaaagc tattacaatt ggcattg

27

<210> 125

<211> 34

<212> DNA

<213> 人工序列

<220>

<223> 引物 JU-107-168-110-ERG9 CDS-r

<400> 125

attcaagttg taattttcat ctaagatgta gtcg

34

[0043]

<210> 126	
<211> 20	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 JU-108-168-110-ERG9 US-f	
<400> 126	
aaaagtgcag ctcagagccc	20
<210> 127	
<211> 37	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 JU-162-168-110-LEU2 DS-f	
<400> 127	
aaagattctc tttttttatg atatttgtac ataaact	37
<210> 128	
<211> 28	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 JU-163-168-110-LEU2 DS-r	
<400> 128	
tagatttagt actgaagagg aggtcgac	28
<210> 129	
<211> 37	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 JU-164-168-110-LEU2 US-f	
<400> 129	
taggataatt atactctatt tctcaacaag taattgg	37
<210> 130	
<211> 41	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 JU-165-168-110-LEU2 US-r	
<400> 130	
tagaatggta tatccttgaa atatatatat atatattgct g	41
<210> 131	
<211> 25	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 JU-169-168-110-URA3-f	
<400> 131	
gttcatcatc tcatggatct gcaca	25
<210> 132	

[0044]

<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 引物 JU-170-168-110-URA3-r

<400> 132
atgcgtccat ctttacagtc ctg 23

<210> 133
<211> 60
<212> DNA
<213> 人工序列

<220>
<223> 引物 JU-172-168-110-ERG9 US-r1

<400> 133
gtgtgtgtgt gatatgtgac gtgtatacgt tttccgcttc tgcttttcgt cttttctctt 60

<210> 134
<211> 25
<212> DNA
<213> 人工序列

<220>
<223> 引物 JU-218-168-130-GAL80US-F

<400> 134
cagatggaat cccttccata gagag 25

<210> 135
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 引物 JU-219-168-130-GAL80US-R

<400> 135
gacgggagtg gaaagaacgg 20

<210> 136
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 引物 JU-220-168-130-GAL80DS-F

<400> 136
aagcatcttg ccctgtgctt g 21

<210> 137
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 引物 JU-221-168-130-GAL80DS-R

<400> 137
catgctacct tccatggttg agc 23

<210> 138

[0045]

<211> 44	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 JU-284-275-31-GAL4-FIX-F2	
<400> 138	
ggattttatg cccagggatg cacttcatgg atttgattgg tctg	44
<210> 139	
<211> 44	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 JU-285-275-31-GAL4-FIX-R2	
<400> 139	
cagaccaatc aaatccatga agtgcacccc tgggcataaa atcc	44
<210> 140	
<211> 29	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 JU-286-275-31-GAL4-F	
<400> 140	
atgaagctac tgtcttctat cgaacaagc	29
<210> 141	
<211> 24	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 JU-287-275-31-GAL4-R	
<400> 141	
tgagcgaagc ttctgaataa gccc	24
<210> 142	
<211> 50	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 KB454-266-53	
<400> 142	
gacggcacgg ccacgcgttt aaaccgccat ccaattcctc tattatatgc	50
<210> 143	
<211> 41	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 KB455-266-53	
<400> 143	
ggctccccct ggccgcgtgt gagcgtggcg gaaagaacag c	41
<210> 144	
<211> 50	

[0046]

<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 KB457-266-53	
<400> 144	
cccgccaggc gctgggggttt aaacaccgtt tgccttttaa ctatcagg	50
<210> 145	
<211> 54	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 KMH14-276-4-接头9.ERG12.rev	
<400> 145	
cccgccaggc gctgggggttt aaacaccgat gtcattaccg ttcttaactt ctgc	54
<210> 146	
<211> 49	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 KMH15-276-4-接头9.ERG19.rev	
<400> 146	
cccgccaggc gctgggggttt aaacaccgat gaccgtttac acagcatcc	49
<210> 147	
<211> 40	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 KMH33-276-21-URA3 5 prime fwd	
<400> 147	
tgcgaggcat atttatggtg aaggataagt ttgaccatc	40
<210> 148	
<211> 64	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 KMH34-276-21-URA3 5 prime rev	
<400> 148	
ggctccccct ggccgcgtgt gagccgtttc ctgcagggtt ttgttctgtg cagttgggtt	60
aaga	64
<210> 149	
<211> 73	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 KMH35-276-21-URA3 3 prime fwd	
<400> 149	
gctcacacgc ggccaggggg agccctgtat tataagtaaa tgcattgtata ctaaactcac	60
aaattagagc ttc	73
<210> 150	

[0047]

<211> 40	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 KMH35-276-21-URA3 3 prime rev	
<400> 150	
gcttcaagaa cttttctgag cacttcatga tgcattgtttg	40
<210> 151	
<211> 44	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 KMH46-276-43-ERG12接头4.fwd	
<400> 151	
aacctgcagg ccgcgagcgc cgatattcgc ggggtggaagg acct	44
<210> 152	
<211> 49	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 KMH47-276-43-ERG19接头4.fwd	
<400> 152	
aacctgcagg ccgcgagcgc cgatcttctg ctaagtgggtg ctgttagac	49
<210> 153	
<211> 59	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 KMH5-276-1-接头3.FS(Kozak).fwd	
<400> 153	
atccccgcgt gcttggccgg ccgtaattaa taatgtcaac tttgcctatt tcttctgtg	59
<210> 154	
<211> 46	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 KMH7-276-1-接头4.TCYC1.rev	
<400> 154	
tacggcgctc gcggcctgca gggtcttcga gcgtcccaaa accttc	46
<210> 155	
<211> 53	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 KMH81-276-116-TDH3.rev.thMG1	
<400> 155	
ggtctgcagc cattattaat ttgtttgttt atgtgtgttt attcgaaact aag	53
<210> 156	
<211> 56	

[0048]

<212> DNA
<213> 人工序列

<220>
<223> 引物 KMH82-276-116-tHMG1.fwd.TDH3

<400> 156
cgaataaaca cacataaaca aacaaattaa taatggctgc agaccaattg gtgaag 56

<210> 157
<211> 56
<212> DNA
<213> 人工序列

<220>
<223> 引物 KMH8-276-1-接头4.tHMG1.fwd

<400> 157
aacctgcagg ccgcgagcgc cgatagttat gacaattaca acaacagaat tctttc 56

<210> 158
<211> 56
<212> DNA
<213> 人工序列

<220>
<223> 引物 KMH91-276-116-TDH3.rev.FS

<400> 158
taggcaaagt tgacattatt aatttgtttg tttatgtgtg tttattcgaa actaag 56

<210> 159
<211> 54
<212> DNA
<213> 人工序列

<220>
<223> 引物 KMH92-276-116-FS.fwd.TDH3

<400> 159
aaacacacat aaacaaacaa attaataatg tcaactttgc ctatttcttc tgtg 54

<210> 160
<211> 48
<212> DNA
<213> 人工序列

<220>
<223> 引物 KMH9-276-1-接头9.tHMG1.rev

<400> 160
cccgccaggc gctgggggtt aaacaccgat ggctgcagac caattggt 48

<210> 161
<211> 47
<212> DNA
<213> 人工序列

<220>
<223> 引物 KMH93-276-130-3 prime IME.接头4.fwd

<400> 161
aacctgcagg ccgcgagcgc cgatctcgaa aagtactaca atcttcc 47

<210> 162
<211> 44
<212> DNA

[0049]

<213> 人工序列	
<220>	
<223> 引物 PW-91-093-CPK422-G	
<400> 162	
gatgtgtatt actagtgtcg acgacagcat tcgcccagta tttt	44
<210> 163	
<211> 24	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 TRIX_K_0142	
<400> 163	
gtattccaat gagaatcgct agaa	24
<210> 164	
<211> 23	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 TRIX_K_0143	
<400> 164	
ttcgtctggt tttatccctc ttc	23
<210> 165	
<211> 20	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 TRIX_K_131	
<400> 165	
cctctcttaa aatgatggcg	20
<210> 166	
<211> 30	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 TRIX_L_034	
<400> 166	
gacggtagca acaagaatat agcacgagcc	30
<210> 167	
<211> 36	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 TRIX_L_035	
<400> 167	
ttttgagggg atattcaact gttttttttt atcatg	36
<210> 168	
<211> 34	
<212> DNA	
<213> 人工序列	

[0050]

<220>
<223> 引物 TRIX_L_036

<400> 168
ttttttatca tgttgatgct ctgcataata atgc 34

<210> 169
<211> 34
<212> DNA
<213> 人工序列

<220>
<223> 引物 TRIX_L_053

<400> 169
tttgtttggtt tatgtgtggtt tattcgaaac taag 34

<210> 170
<211> 27
<212> DNA
<213> 人工序列

<220>
<223> 引物 TRIX_L_106

<400> 170
atgtctcaga acgtttacat tgtatcg 27

<210> 171
<211> 27
<212> DNA
<213> 人工序列

<220>
<223> 引物 TRIX_L_107

<400> 171
aggcagccaa gacattgatt aacatcc 27

<210> 172
<211> 29
<212> DNA
<213> 人工序列

<220>
<223> 引物 TRIX_L_109

<400> 172
atgaaactct ctactaaact ttgttggtg 29

<210> 173
<211> 28
<212> DNA
<213> 人工序列

<220>
<223> 引物 TRIX_L_110

<400> 173
atgagaaaaa aaatcggttg ggcttaac 28

<210> 174
<211> 26
<212> DNA
<213> 人工序列

[0051]

<220>		
<223>	引物 TRIX_L_112	
<400>	174	
	atgtcattac cgttcttaac ttctgc	26
<210>	175	
<211>	25	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物 TRIX_L_113	
<400>	175	
	attcgcgggt ggaaggacct tgtgg	25
<210>	176	
<211>	27	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物 TRIX_L_115	
<400>	176	
	atgaccgttt acacagcatc cgttacc	27
<210>	177	
<211>	25	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物 TRIX_L_116	
<400>	177	
	cttgtgctaa gtggtgctgt tagac	25
<210>	178	
<211>	25	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物 TRIX_L_118	
<400>	178	
	atgtcagagt tgagagcctt cagtg	25
<210>	179	
<211>	23	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物 TRIX_L_119	
<400>	179	
	agtgcacact ttcaagctaa cac	23
<210>	180	
<211>	27	
<212>	DNA	
<213>	人工序列	
<220>		

[0052]

<223> 引物 TRIX_L_121	
<400> 180 atgactgccg acaacaatag tatgccc	27
<210> 181 <211> 26 <212> DNA <213> 人工序列	
<220> <223> 引物 TRIX_L_122	
<400> 181 catcagtggg aaacattcaa gaggcc	26
<210> 182 <211> 30 <212> DNA <213> 人工序列	
<220> <223> 引物 TRIX_L_124	
<400> 182 atggcttcag aaaaagaaat taggagagag	30
<210> 183 <211> 30 <212> DNA <213> 人工序列	
<220> <223> 引物 TRIX_L_125	
<400> 183 tgaggtcggt gcttttccta ttattatatg	30
<210> 184 <211> 28 <212> DNA <213> 人工序列	
<220> <223> 引物 TRIX_L_193	
<400> 184 tcgacactag taatacacat catcgccc	28
<210> 185 <211> 28 <212> DNA <213> 人工序列	
<220> <223> 引物 TRIX_L_194	
<400> 185 gagctcctcg agaagttaag attatatg	28
<210> 186 <211> 26 <212> DNA <213> 人工序列	
<220> <223> 引物 TRIX_L_194	

[0053]

<400> 186 atggctgcag atcaattggt gaagac	26
<210> 187 <211> 32 <212> DNA <213> 人工序列	
<220> <223> 引物 TRIX_L_233	
<400> 187 agttatgaca attacaacaa cagaattctt tc	32
<210> 188 <211> 32 <212> DNA <213> 人工序列	
<220> <223> 引物 GW-52-84 pAM326 BamHI	
<400> 188 taataaggat ccatgtcaac ttgcctatt tc	32
<210> 189 <211> 32 <212> DNA <213> 人工序列	
<220> <223> 引物 GW-52-84 pAM326 NheI	
<400> 189 ttatagctag ctcaaacgac cataggatga ac	32
<210> 190 <211> 26 <212> DNA <213> 人工序列	
<220> <223> 引物 TRIX_L_266	
<400> 190 tacttttttt ttggatggac gcaaag	26
<210> 191 <211> 29 <212> DNA <213> 人工序列	
<220> <223> 引物 TRIX_L_267	
<400> 191 aagtatagag gtatattaac aatttttttg	29
<210> 192 <211> 30 <212> DNA <213> 人工序列	
<220> <223> 引物 TRIX_RN017	

[0054]

<400> 192 acgaagtgac tgacagaata ctgacatcag	30
<210> 193 <211> 30 <212> DNA <213> 人工序列	
<220> <223> 引物 TRIX_RN018	
<400> 193 ttaaaagttg tttccgctgt atcctgtatc	30
<210> 194 <211> 30 <212> DNA <213> 人工序列	
<220> <223> 引物 TRIX_RN019	
<400> 194 agtatacact aaatTTTTatg caataataaa	30
<210> 195 <211> 30 <212> DNA <213> 人工序列	
<220> <223> 引物 TRIX_RN020	
<400> 195 ggTTTTgcta aggaagTTTT ggagtatgct	30
<210> 196 <211> 26 <212> DNA <213> 人工序列	
<220> <223> 引物 TRIX_Z025	
<400> 196 cacgaaaatc gttattgtct tgaagg	26
<210> 197 <211> 52 <212> DNA <213> 人工序列	
<220> <223> 引物 TRIX_Z026	
<400> 197 gctttatgga ccctgaaacc actcactatt attccataag atgatacatta gc	52
<210> 198 <211> 54 <212> DNA <213> 人工序列	
<220> <223> 引物 TRIX_Z027	
<400> 198	

[0055]

gcttcaattt aattatatca gttattacca cgaaaatcgt tattgtcttg aagg	54
<210> 199	
<211> 30	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 TRIX_Z028	
<400> 199	
tcactattat tccataagat gatcattagc	30
<210> 200	
<211> 22	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 TRIX_Z033	
<400> 200	
gtgggtttcag ggtccataaa gc	22
<210> 201	
<211> 28	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 TRIX_Z034	
<400> 201	
gtaataactg atataattaa attgaagc	28
<210> 202	
<211> 39	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 TRIX_Z035	
<400> 202	
ctgttgacat tgcgaagagt gacaaagatt ttgttatcg	39
<210> 203	
<211> 39	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 TRIX_Z036	
<400> 203	
cgataacaaa atctttgtca ctcttcgcaa tgtcaacag	39
<210> 204	
<211> 45	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物 VH228-235-7-URA3LOF3RYSE12-1F	
<400> 204	
gctcacacgc ggccaggggg agcctcacta ttattccata agatg	45

[0056]

<210> 205
 <211> 46
 <212> DNA
 <213> 人工序列

<220>
 <223> 引物 VH229-235-7-URA3LOF3RYSE12-1R

<400> 205
 aggtccgccg gcgttggacg agcgcacgaa aatcgttatt gtcttg 46

<210> 206
 <211> 198
 <212> DNA
 <213> 人工序列

<220>
 <223> 合成的DNA片段

<400> 206
 cacgaaaatc gttattgtct tgaagggtgaa atttctactc ttattaatgg tgaacgttaa 60
 gctgatgcta tgatggaagc tgattgggtc taacttgctt gtcattcttg taatgggtcat 120
 atggctcggt ttattactta agttatttgg actcgttttg aacgtaatgc taatgatcat 180
 cttatggaat aatagtga 198

<210> 207
 <211> 2733
 <212> DNA
 <213> 人工序列

<220>
 <223> pRYSE 输入载体 2

<400> 207
 gtaaaacgac ggccagtatt aaccctcact aaagggaact cgaggctctt cagctcacac 60
 gcggccaggg ggagcctggc agactccata tgctatgcgg catcagagca gattgtactg 120
 agagtgcacc atatgcggtg tgaataaccg cacagatgcg taaggagaaa ataccgcac 180
 aggcgccatt cgccattcag gctgcgcaac tgttgggaag ggcgatcggt gcgggcctct 240
 tcgctattac gccagctggc gaaaggggga tgtgctgcaa ggcgattaag ttgggtaacg 300
 ccagggtttt cccagtcacg acgttgtaaa acgacggcca gtgaattcga gctcgggtacc 360
 cggggatcct ctgagctcga cctgcaggca tgcaagcttg gcgtaatcat ggcatagct 420
 gtttcctgtg tgaaattgtt atccgctcac aattccacac aacatacgag ccggaagcat 480
 aaagtgtaaa gcctgggggt cctaattgagt gagctaactc acattaattg cgttgcgcgc 540
 tagcgagtca tccacgctcg tccaacgccg gcggaccttg aagagcgagc tcccgcgtgag 600
 caataactag cgtcatagct gtttcctggg tcgttcggct gcggcgagcg gtatcagctc 660
 actcaaaagg gtaatacgg ttatccacag aatcagggga taacgcagga aagaacatgt 720
 gagcaaaagg ccagcaaaag gccaggaacc gtaaaaaggc gcggttgctg gcgtttttcc 780
 ataggctccg cccccctgac gagcatcaca aaaatcgacg ctcaagtcag aggtggcgaa 840
 acccgacagg actataaaga taccaggcgt ttccccctgg aagctccctc gtgcgctctc 900
 ctgttcgcac cctgccgctt acccgatacc tgtccgcctt tctcccttcg ggaagcgtgg 960
 cgctttctca tagctcacgc tgtaggtatc tcagttcggg gtaggtcgtt cgctccaagc 1020
 tgggctgtgt gcacgaaccc cccgttcagc ccgaccgctg cgccttatcc ggtaactatc 1080
 gtcttgattc caaccgggta agacacgact tatcgccact ggcagcagcc actggtaaca 1140
 ggattagcag agcgaggtat gtaggcggtg ctacagagtt cttgaagtgg tggcctaact 1200
 acggctacac tagaagaaca gtatttggtg tctgcgctct gctgaagcca gttaccttcg 1260
 gaaaaagagt tggtagctct tgatccggca aacaaaccac cgctggtagc ggtgggtttt 1320
 ttgtttgcaa gcagcagatt acgcgcagaa aaaaaggatc tcaagaagat cctttgatct 1380
 tttctacggg gtctgacgct cagtggaaacg aaaactcacg ttaagggtatt ttgggtcatga 1440
 gattatcaaa aaggatcttc acctagatcc ttttaaatta aaaatgaagt tttaaatcaa 1500
 tctaaagtat atatgagtaa cttggctcga tgcttaacca tgcttaatca gtgaggcacc 1560
 tatctcagcg atctgtctat ttcgttcacg catagttgcc tgactgcccg tcgtgtagat 1620
 aactacgata cgggagggtc taccatctgg cccagtgctt gcaatgatac cgcgagaccc 1680
 acgctcaccg gctccagatt tatcagcaat aaaccagcca gccggaaggg ccgagcgcag 1740
 aagtggctct gcaactttat ccgcctccat ccagctcatt aattgttgcc gggaagctag 1800
 agtaagtagt tcgccagtta atagtttgcg caacgttggt gccattgcta caggcatcgt 1860
 ggtgtcacgc tcgtcgtttg gtatggcttc attcagctcc ggttcccaac gatcaaggcg 1920

[0057]

agttacatga	tccccatgt	tgtgcaaaaa	agcggttagc	tccttcggtc	ctccgatcgt	1980
tgtcagaagt	aagttggccg	cagtgttatc	actcatggtt	atggcagcac	tgcataattc	2040
tcttactgtc	atgccatccg	taagatgctt	ttctgtgact	ggtgagtact	caaccaagtc	2100
attctgagaa	tagtgtatgc	ggcgaccgag	ttgctcttgc	ccggcgtcaa	tacgggataa	2160
taccgcgcca	catagcagaa	ctttaaaagt	gctcatcatt	ggaaaacgtt	cttcggggcg	2220
aaaactctca	aggatcttac	cgctgttgag	atccagttcg	atgtaaccca	ctcgtgcacc	2280
caactgatct	tcagcatctt	ttactttcac	cagcgtttct	gggtgagcaa	aaacaggaag	2340
gcaaaatgcc	gcaaaaaagg	gaataagggc	gacacggaaa	tgttgaatac	tcatacaattg	2400
cctttttcaa	tattattgaa	gcattttatca	gggttattgt	ctcatgagcg	gttacatatt	2460
tgaatgtatt	tagaaaaata	aacaaatagg	ggttccgcgc	acatttcccc	gaaaagtgcc	2520
acctgacgtc	taagaaacca	ttattatcat	gacattaacc	tataaaaaata	ggcgtatcac	2580
gaggcccttt	catctcgcgc	gtttcgggtg	tgacggtgaa	aacctctgac	acatgcagct	2640
cccggagaca	gtcacagctt	gtctgtaagc	ggatgccggg	agcagacaag	cccgtcaggg	2700
cgcgctcagc	ggtgttgccg	ggtgtcgggg	ctg			2733

<210> 208

<211> 2733

<212> DNA

<213> 人工序列

<220>

<223> pRYSE 输入载体 3

<400> 208

gtaaaacgac	ggccagttatt	aaccctcact	aaagggaaact	cgaggctctt	cacgctcgtc	60
caacgccggc	ggaccttggc	agactccata	tgctatgcgg	catcagagca	gattgtactg	120
agagtgcacc	atatgcggtg	tgaaataccg	cacagatgcg	taaggagaaa	ataccgcatac	180
aggcgccatt	cgccattcag	gctgcgcaac	tggttggaag	ggcgatcggg	gcgggcctct	240
tcgctattac	gccagctggc	gaaaggggga	tgtgctgcaa	ggcgattaag	ttgggtaacg	300
ccagggtttt	cccagtcacg	acgttgtaaa	acgacggcca	gtgaattcga	gctcggtagc	360
cggggatcct	ctagcgtcga	cctgcaggca	tgcaagcttg	gcgtaatcat	ggtcatagct	420
gtttcctgtg	tgaaatttgt	atccgctcac	aattccacac	aacatacgag	ccggaagcat	480
aaagtgtaaa	gcctgggggtg	cctaattgag	gagctaactc	acattaattg	aggtggcgca	540
tagcgagtca	tccaatcccc	gcgtgcttgg	ccggccggtg	aagagcgagc	tcccgcgtgag	600
caataactag	cgatcatagct	gtttcctggg	tcgttcggct	gcggcgagcg	gtatcagctc	660
actcaaaggc	ggtaatacgg	ttatccacag	aatcagggga	taacgcagga	aagaacatgt	720
gagcaaaaag	ccagcaaaaag	gccaggaacc	gtaaaaaaggc	cgcgttgctg	gcgtttttcc	780
ataggctccg	cccccttgac	gagcatcaca	aaaatcgcag	ctcaagtcat	aggtggcgca	840
acccgacagg	actataaaga	taccaggcgt	ttccccctgg	aagctccctc	gtgcgctctc	900
ctgttccgac	cctgccgctt	acccgatacc	tgtccgcctt	tctcccttcg	ggaagcgtgg	960
cgctttctca	tagctcacgc	tgtaggtatc	tcagttcggg	gtaggtcggt	cgctccaagc	1020
tgggctgtgt	gcacgaaccc	cccgttcagc	ccgaccgctg	cgcttatcc	ggttaactatc	1080
gtcttgattc	gaaccggta	agacacgact	tatcgccact	ggcagcagcc	actggttaaca	1140
ggattagcag	agcgagggtat	gtaggcggtg	ctacagagtt	cttgaagtgg	tggcctaact	1200
acggctacac	tagaagaaca	gtatttggtg	tctgcgctct	gctgaagcca	gttaccttcg	1260
gaaaaagagt	tggtagctct	tgatccggca	aacaaaccac	cgctggtagc	ggtggttttt	1320
ttgtttgcaa	gcagcagatt	acgcgagaa	aaaaaggatc	tcaagaagat	cctttgatct	1380
tttctacggg	gtctgacgct	cagtggaaac	aaaactcacg	ttaagggtat	ttggtcatga	1440
gattatcaaaa	aaggatcttc	acctagatcc	ttttaaatta	aaaatgaagt	tttaaatcaa	1500
tctaaagtat	atatgagtaa	cttggtcgca	tgcttaccaa	tgcttaatca	gtgaggcacc	1560
tatctcagcg	atctgtctat	ttcgttcac	catagttgcc	tgactgccc	tcgtgtagat	1620
aactacgata	cgggagggtc	taccatcttg	ccccagtgct	gcaatgatac	cgcgagaccc	1680
acgctcaccg	gctccagatt	tatcagcaat	aaaccagcca	gccggaagg	ccgagcgag	1740
aagtgggtcct	gcaactttat	ccgctccat	ccagttctat	aattgttgcc	gggaagctag	1800
agtaagtagt	tcgccagtta	atagtttgcg	caacgttgtt	gccattgcta	caggcatcgt	1860
ggtgtcacgc	tcgtcgtttg	gtatggcttc	attcagctcc	ggttcccaac	gatcaaggcg	1920
agttacatga	tccccatgt	tgtgcaaaaa	agcggttagc	tccttcggtc	ctccgatcgt	1980
tgtcagaagt	aagttggccg	cagtgttatc	actcatggtt	atggcagcac	tgcataattc	2040
tcttactgtc	atgccatccg	taagatgctt	ttctgtgact	ggtgagtact	caaccaagtc	2100
attctgagaa	tagtgtatgc	ggcgaccgag	ttgctcttgc	ccggcgtcaa	tacgggataa	2160
taccgcgcca	catagcagaa	ctttaaaagt	gctcatcatt	ggaaaacgtt	cttcggggcg	2220
aaaactctca	aggatcttac	cgctgttgag	atccagttcg	atgtaaccca	ctcgtgcacc	2280
caactgatct	tcagcatctt	ttactttcac	cagcgtttct	gggtgagcaa	aaacaggaag	2340
gcaaaaatgcc	gcaaaaaagg	gaataagggc	gacacggaaa	tgttgaatac	tcatacaattg	2400
cctttttcaa	tattattgaa	gcattttatca	gggttattgt	ctcatgagcg	gttacatatt	2460
tgaatgtatt	tagaaaaata	aacaaatagg	ggttccgcgc	acatttcccc	gaaaagtgcc	2520
acctgacgtc	taagaaacca	ttattatcat	gacattaacc	tataaaaaata	ggcgtatcac	2580

[0058]

gagggcccttt	catctcgcgc	gtttcgggtga	tgacgggtgaa	aacctctgac	acatgcagct	2640
cccggagaca	gtcacagctt	gtctgtaagc	ggatgccggg	agcagacaag	cccgtcaggg	2700
cgcgctcagcg	ggtgtttggcg	ggtgtcgggg	ctg			2733

<210> 209
 <211> 2733
 <212> DNA
 <213> 人工序列

<220>
 <223> pRYSE 输入载体 4

<400> 209						
gtaaaacgac	ggccagttatt	aaccctcact	aaagggaaact	cgaggctctt	caatccccgc	60
gtgcttggcc	ggccgtttggc	agactccata	tgctatgcgg	catcagagca	gattgtactg	120
agagtgcacc	atatgcgggtg	tgaaataccg	cacagatgcg	taaggagaaa	ataccgcac	180
aggcgccatt	cgccattcag	gctgcgcaac	tgttgggaag	ggcgatcgg	gcgggcctct	240
tcgctattac	gccagctggc	gaaaggggga	tgtgctgcaa	ggcgattaag	ttgggtaacg	300
ccagggtttt	cccagtcacg	acgtttgtaa	acgacggcca	gtgaattcga	gctcggtagc	360
cggggatcct	ctagcgtcga	cctgcaggca	tgcaagcttg	gcgtaatcat	ggtcatagct	420
gtttcctgtg	tgaaattgtt	atccgctcac	aattccacac	aacatacag	ccggaagcat	480
aaagtgtaaa	gcctgggggtg	cctaattagt	gagctaactc	acattaattg	cgttgcgcgc	540
tagcgagtca	tccaaacctg	caggccgcga	gcgccgattg	aagagcgagc	tccccgtgag	600
caataactag	cgatcatagct	gtttcctggg	tcgttcggct	gcggcgagcg	gtatcagctc	660
actcaaaggc	ggtaatacgg	ttatccacag	aatcagggga	taacgcagga	aagaacatgt	720
gagcaaaagg	ccagcaaaag	gccaggaacc	gtaaaaaggc	cgctttgctg	gcgtttttcc	780
ataggctccg	ccccctgac	gagcatcaca	aaaatcgacg	ctcaagtcag	aggtggcgaa	840
acccgacag	actataaaga	taccaggcgt	ttccccctgg	aagctccctc	gtgcgctctc	900
ctgttccgac	cctgccgctt	acccgatacc	tgtccgcctt	tctcccttcg	ggaagcgtgg	960
cgcttttctca	tagctcacgc	tgtaggtatc	tcagttcgg	gtaggtcgtt	cgctccaagc	1020
tggtgtgtgt	gcacgaaccc	cccgttcagc	ccgaccgctg	cgcttatcc	ggtaactatc	1080
gtcttgattc	caaccgggtc	agacacgact	tatcgccact	ggcagcagcc	actggtaaca	1140
ggattagcag	agcgaggtat	gtaggcgggtg	ctacagagtt	cttgaagtgg	tgccctaact	1200
acggctacac	tagaagaaca	gtatttggtc	tctgcgctct	gctgaagcca	gttaccttcg	1260
gaaaaagagt	tggtagctct	tgatccggca	aacaaaccac	cgctggtagc	gggtggtttt	1320
ttgtttgcaa	gcagcagatt	acgcgcagaa	aaaaaggatc	tcaagaagat	cctttgatct	1380
tttctacggg	gtctgacgct	cagtggaaacg	aaaactcacg	ttaagggatt	ttgggtcatga	1440
gattatcaaaa	aaggatcttc	acctagatcc	ttttaaatta	aaaatgaagt	tttaaatcaa	1500
tctaaagtat	atatgagtaa	cttggtcgca	tgcttaccaa	tgcttaatca	gtgaggcacc	1560
tatctcagcg	atctgtctat	ttcgttcatc	catagttgcc	tgactgcccg	tcgtgtagat	1620
aactacgata	cgggagggtc	taccatcttg	ccccagtgct	gcaatgatac	cgcgagaccc	1680
acgctcaccg	gctccagatt	tatcagcaat	aaaccagcca	gccggaagg	ccgagcgag	1740
aagtggctct	gcaactttat	ccgcctccat	ccagttctatt	aattgttgcc	gggaagctag	1800
agtaagtagt	tcgccagtta	atagtttgcg	caacgttggt	gccattgcta	caggcatcgt	1860
ggtgtcacgc	tcgtcgtttg	gtatggcttc	attcagctcc	ggttcccaac	gatcaaggcg	1920
agttacatga	tccccatgt	tgtgcaaaaa	agcggtttagc	tccttcgggtc	ctccgatcgt	1980
tgctcagaagt	aagttggccg	cagtgttatc	actcatggtt	atggcagcac	tgcataattc	2040
tcttactgtc	atgccatccg	taagatgctt	ttctgtgact	ggtgagtact	caaccaagtc	2100
attctgagaa	tgtgttatgc	ggcgaccgag	ttgtcttgct	ccggcgtaaa	tacgggataa	2160
taccgcgcca	catagcagaa	ctttaaaagt	gctcatcatt	ggaaaacgtt	cttcggggcg	2220
aaaactctca	aggatcttac	cgctgttgag	atccagttcg	atgtaaccca	ctcgtgcacc	2280
caactgatct	tcagcatctt	ttactttcac	cagcgtttct	gggtgagcaa	aaacaggaag	2340
gcaaaatgcc	gcaaaaaagg	gaataagggc	gacacggaaa	tggtgaatac	tcataaattg	2400
cctttttcaa	tattattgaa	gcatttatca	gggttattgt	ctcatgagcg	gttacatatt	2460
tgaatgtatt	tagaaaaata	aacaaatagg	ggttccgcgc	acattttccc	gaaaagtgcc	2520
acctgacgtc	taagaaacca	ttattatcat	gacattaacc	tataaaaaata	ggcgtatcac	2580
gagggcccttt	catctcgcgc	gtttcgggtga	tgacgggtgaa	aacctctgac	acatgcagct	2640
cccggagaca	gtcacagctt	gtctgtaagc	ggatgccggg	agcagacaag	cccgtcaggg	2700
cgcgctcagcg	ggtgtttggcg	ggtgtcgggg	ctg			2733

<210> 210
 <211> 2733
 <212> DNA
 <213> 人工序列

<220>
 <223> pRYSE 输入载体 5

[0059]

```

<400> 210
gtaaaacgac ggccagtatt aaccctcact aaaggggaact cgaggctctt caaacctgca 60
ggccgagcag gccgattggc agactccata tgctatgcgg catcagagca gattgtactg 120
agagtgcacc atatgcggtg tgaaataccg cacagatgcg taaggagaaa ataccgcac 180
aggcgccatt cgccattcag gctgcgcaac tgttgggaag ggcgatcggg gggggcctct 240
tcgctattac gccagctggc gaaaggggga tgtgctgcaa ggcgattaag ttgggtaacg 300
ccagggtttt cccagtcacg acgttgtaaa acgacggcca gtgaattcga gctcggtagc 360
cggggatcct ctagcgtcga cctgcaggca tgcaagcttg gcgtaatcat ggtcatagct 420
gtttcctgtg tgaaattggt atccgctcac aattccacac aacatacgag ccggaagcat 480
aaagtgtaaa gcctgggggt cctaattaggt gagctaactc acattaattg cgttgcgcgc 540
tagcgagtca tccaaacgcg atcgccgacg ccgcccattg aagagcgagc tcccgcgtgag 600
caataactag cgtcatagct gtttcctggg tcggttcggc ggcgcgagcg gtatcagctc 660
actcaaaggc ggtaatacgg ttatccacag aatcagggga taacgcagga aagaacatgt 720
gagcaaaagg ccagcaaaag gccaggaacc gtaaaaaggc cgcgttgctg gcgtttttcc 780
ataggctccg cccccctgac gagcatcaca aaaatcgacg ctcaagtcag aggtggcgaa 840
acccgacagg actataaaga taccaggcgt ttccccctgg aagctccctc gtgcgctctc 900
ctgttccgac ctgcccgttt acccgatacc tgtccgcctt tctcccttcg ggaagcgtgg 960
cgctttctca tagtcacgc tgtaggtatc tcagttcggg gttaggtcgtt cgctccaagc 1020
tgggctgtgt gcacgaaccc cccgttcagc ccgaccgctg cgccttatcc ggtaactatc 1080
gtcttgattc caaccggta agacacgact tatcgccact ggcagcagcc actggtaaca 1140
ggattagcag agcgaggtat gtaggcgggt ctacagagtt cttgaagtgg tggcctaact 1200
acggctacac tagaagaaca gtatttggtg tctgcgctct gctgaagcca gttaccttcg 1260
gaaaagagtg tgtagctctt tgatccggca acaaaaccac cgctggtagc ggtgggtttt 1320
ttgtttgcaa gcgacgagatt acgcgcagaa aaaaaggatc tcaagaagat cttttgatct 1380
tttctacggg gtctgacgct cagtggaaacg aaaactcacg ttaagggatt ttggtcatga 1440
gattatcaaa aaggatcttc acctagatcc ttttaaatta aaaatgaagt tttaaatcaa 1500
tctaaagtat atatgagtaa cttggtcgca tgcttaacca tgcttaacca gtgaggcacc 1560
tatctcagcg atctgtctat ttctgttcac catagttggc tgactgcccg tcgtgtagat 1620
aactacgata cgggagggct taccatctgg cccagtgctt gcaatgatac cgcgagacc 1680
acgctcacgg gctccagatt tatcagcaat aaaccagcca gccggaaggg ccgagcgtag 1740
aagtggctct gcaactttat ccgcctccat ccagctctatt aattgttgcc ggaagctag 1800
agtaagtagt tcgccagtta atagtttgcg caacgttggt gccattgcta caggcatcgt 1860
ggtgtcacgc tcgctggttg gtatggcttc attcagctcc ggttcccaac gatcaaggcg 1920
agttacatga tccccatgtt tgtgcaaaaa agcggtttagc tccttcggtc ctccgatcgt 1980
tgtcagaagt aagttggccg cagtgttatc actcatgggt atggcagcac tgcataattc 2040
tcttactgtc atgccatccg taagatgctt ttctgtgact ggtgagtact caaccaagtc 2100
attctgagaa tagtgtatgc ggcgaccgag ttgctcttgc ccggcgtaaa tacgggataa 2160
taccgcgcca catagcagaa ctttaaaagt gctcatcatt ggaaaacgtt cttcggggcg 2220
aaaactctca aggatcttac cgctgttgag atccagttcg atgtaaccca ctctgcacc 2280
caactgatct tcagcatctt ttacttttct cagcgtttct gggtagcaa aaacaggaag 2340
gcaaaatgcc gcaaaaaagg gaataaggcg gacacggaaa tggtgaatac tcatcaattg 2400
cctttttcaa tattattgaa gcatttatca gggttattgt ctcatgagcg gttacatatt 2460
tgaatgtatt tagaaaaata aacaaatagg ggttccgcgc acatttcccc gaaaagtgcc 2520
acctgacgtc taagaaacca ttattatcat gacattaacc tataaaaaata ggcgtatcac 2580
gaggcccttt catctcgcgc gtttcgggtg tgacggtgaa aacctctgac acatgcagct 2640
cccgagaca gtcacagctt gtctgtaagg gtagtcgggg agcagacaag ccgctcaggg 2700
cgcgtagcgc ggtgttggcg ggtgtcgggg ctg 2733

```

```

<210> 211
<211> 2733
<212> DNA
<213> 人工序列

```

```

<220>
<223> pRYSE 输入载体 6

```

```

<400> 211
gtaaaacgac ggccagtatt aaccctcact aaaggggaact cgaggctctt caaacgcgat 60
cgccgacgcc gccgattggc agactccata tgctatgcgg catcagagca gattgtactg 120
agagtgcacc atatgcggtg tgaaataccg cacagatgcg taaggagaaa ataccgcac 180
aggcgccatt cgccattcag gctgcgcaac tgttgggaag ggcgatcggg gggggcctct 240
tcgctattac gccagctggc gaaaggggga tgtgctgcaa ggcgattaag ttgggtaacg 300
ccagggtttt cccagtcacg acgttgtaaa acgacggcca gtgaattcga gctcggtagc 360
cggggatcct ctagcgtcga cctgcaggca tgcaagcttg gcgtaatcat ggtcatagct 420
gtttcctgtg tgaaattggt atccgctcac aattccacac aacatacgag ccggaagcat 480
aaagtgtaaa gcctgggggt cctaattaggt gagctaactc acattaattg cgttgcgcgc 540

```

[0060]

tagcgagtca	tccaaaggcg	gccgctggcg	agggagattg	aagagcgagc	tcccgtctag	600
caataactag	cgtcatagct	gtttcctggg	tcgttcggct	gcggcgagcg	gtatcagctc	660
actcaaaggc	ggtaatacgg	ttatccacag	aatcagggga	taacgcagga	aagaacatgt	720
gagcaaaagg	ccagcaaaag	gccaggaacc	gtaaaaaggc	cgcggttgctg	gcgtttttcc	780
ataggctccg	ccccctgac	gagcatcaca	aaaatcgacg	ctcaagtcag	aggtggcgaa	840
acccgacagg	actataaaga	taccaggcgt	ttccccctgg	aagctccctc	gtgcgctctc	900
ctgttccgac	cctgccgctt	acccgatacc	tgtccgcctt	tctcccttcg	ggaagcgtgg	960
cgctttctca	tagctcacgc	tgtaggtatc	tcagttcggg	gtaggtcggt	cgctccaagc	1020
tgggctgtgt	gcacgaaccc	cccgttcagc	ccgaccgctg	cgcttatcc	ggtaactatc	1080
gtcttgattc	caaccgggta	agacacgact	tatcgccact	ggcagcagcc	actggtaaca	1140
ggattagtag	agcgaggtat	gtaggcgggt	ctacagagtt	cttgaagtgg	tggcctaact	1200
acggctacac	tagaagaaca	gtatttggta	tctgcgctct	gctgaagcca	gttaccttcg	1260
gaaaaagagt	tggtagctct	tgatccggca	aacaaaccac	cgctggtagc	ggtggttttt	1320
ttgtttgcaa	gcagcagatt	acgcgcagaa	aaaaaggatc	tcaagaagat	cctttgatct	1380
tttctacggg	gtctgacgct	cagtggaaacg	aaaactcacg	ttaagggatt	ttgggtcatga	1440
gattatcaaa	aaggatcttc	acctagatcc	ttttaaatca	aaaatgaagt	tttaaatcaa	1500
tctaaagtat	atatgagtaa	cttggtcgca	tgcttaccac	tgcttaatca	gtgaggcacc	1560
tatctcagcg	atctgtctat	ttcgttcac	catagttgcc	tgactgccc	tcgtgtagat	1620
aactacgata	cgggagggct	taccatctgg	ccccagtgct	gcaatgatac	cgcgagaccc	1680
acgctcaccc	gctccagatt	tatcagcaat	aaaccagcca	gccggaagg	ccgagcgcag	1740
aagtggctct	gcaactttat	ccgcctccat	ccagtcctatt	aattgttgcc	gggaagctag	1800
agtaagtagt	tcgccagtta	atagtttgcg	caacgttggt	gccattgcta	caggcatcgt	1860
gggtgtcacgc	tcgtcgtttg	gtatggcttc	attcagctcc	ggttcccaac	gatcaaggcg	1920
agttacatga	tccccatgt	tgtgcaaaaa	agcggtttagc	tccttcggtc	ctccgatcgt	1980
tgtcagaagt	aagttggccg	cagtgttatc	actcatggtt	atggcagcac	tgcataattc	2040
tcttactgtc	atgccatccg	taagatgctt	ttctgtgact	ggtgagtact	caaccaagtc	2100
attctgagaa	tagtgtatgc	ggcgaccgag	ttgctcttgc	ccggcgtaaa	tacgggataa	2160
taccgcgcca	catagcagaa	ctttaaaagt	gctcatcatt	ggaaaacgtt	cttcggggcg	2220
aaaactctca	aggatcttac	cgctgttgag	atccagttcg	atgtaacca	ctcgtgcacc	2280
caactgatct	tcagcatctt	ttactttcac	cagcgtttct	gggtgagcaa	aaacaggaag	2340
gcaaaatgcc	gcaaaaaagg	gaataagggc	gacacggaaa	tgttgataac	tcataaattg	2400
cctttttcaa	tattattgaa	gcatttatcg	gggttattgt	ctcatgagcg	gttacatatt	2460
tgaattgtatt	tagaaaaata	aacaaatagg	ggttcgcgc	acatttcccc	gaaaagtgcc	2520
acctgacgtc	taagaaacca	ttattatcat	gacattaacc	tataaaaata	ggcgtatcac	2580
gaggcccttt	catctcgcgc	gtttcgggtg	tgacgggtga	aacctctgac	acatgcagct	2640
cccggagaca	gtcacagctt	gtctgtgaagc	ggatgccggg	agcagacaag	cccgtcaggg	2700
cgcgctcagcg	ggtgttgggc	ggtgtcgggg	ctg			2733

<210> 212

<211> 2733

<212> DNA

<213> 人工序列

<220>

<223> pRYSE 输入载体 7

<400> 212

gtaaaacgag	ggccagtatt	aaccctcact	aaagggaact	cgaggctctt	caaaggcggc	60
cgctggcgag	ggagattggc	agactccata	tgctatgcgg	catcagagca	gattgtactg	120
agagtgcacc	atatgcgggtg	tgaaataccg	cacagatgcg	taaggagaaa	ataccgcac	180
aggcgccatt	cgccattcag	gctgcgcaac	tgttgggaag	ggcgatcggt	gcgggcctct	240
tcgctattac	gccagctggc	gaaaggggga	tgtgctgcaa	ggcgattaa	ttgggtaacg	300
ccagggtttt	cccagtcacg	acgttgtaaa	acgacggcca	gtgaattcga	gctcggatcc	360
cggggatcct	ctagcgtcga	cctgcaggca	tgcaagcttg	gcgtaatcat	ggtcatagct	420
gtttcctgtg	tgaaattggt	atccgctcac	aattccacac	aacatacgag	ccggaagcat	480
aaagtgtaaa	gcctgggggtg	cctaattgag	gagctaactc	acattaattg	cgttgcgcgc	540
tagcgagtca	tccaaaggcg	cgccacgggtc	gtgcggattg	aagagcgagc	tcccgtctag	600
caataactag	cgtcatagct	gtttcctggg	tcgttcggct	gcggcgagcg	gtatcagctc	660
actcaaaggc	ggtaatacgg	ttatccacag	aatcagggga	taacgcagga	aagaacatgt	720
gagcaaaagg	ccagcaaaag	gccaggaacc	gtaaaaaggc	cgcggttgctg	gcgtttttcc	780
ataggctccg	ccccctgac	gagcatcaca	aaaatcgacg	ctcaagtcag	aggtggcgaa	840
acccgacagg	actataaaga	taccaggcgt	ttccccctgg	aagctccctc	gtgcgctctc	900
ctgttccgac	cctgccgctt	acccgatacc	tgtccgcctt	tctcccttcg	ggaagcgtgg	960
cgctttctca	tagctcacgc	tgtaggtatc	tcagttcggg	gtaggtcggt	cgctccaagc	1020
tgggctgtgt	gcacgaaccc	cccgttcagc	ccgaccgctg	cgcttatcc	ggtaactatc	1080
gtcttgattc	caaccgggta	agacacgact	tatcgccact	ggcagcagcc	actggtaaca	1140
ggattagtag	agcgaggtat	gtaggcgggtg	ctacagagtt	cttgaagtgg	tggcctaact	1200

[0061]

acggctacac	tagaagaaca	gtattttggt	tctgcgctct	gctgaagcca	gttaccttcg	1260
gaaaaagagt	tggtagctct	tgatccggca	aacaaaccac	cgctggtagc	ggtggttttt	1320
ttgtttgcaa	gcagcagatt	acgcgcagaa	aaaaaggatc	tcaagaagat	cctttgatct	1380
tttctacggg	gtctgacgct	cagtggaaacg	aaaactcacg	ttaagggatt	ttggtcatga	1440
gattatcaaa	aaggatcttc	acctagatcc	ttttaaatta	aaaatgaagt	tttaaataca	1500
tctaaagtat	atatgagtaa	cttggtcgca	tgcttaccaa	tgcttaatca	gtgaggcacc	1560
tatctcagcg	atctgtctat	ttcgttcac	catagttgcc	tgactgccc	tcgtgtagat	1620
aactacgata	cgggagggct	taccatctgg	ccccagtgct	gcaatgatac	cgcgagacc	1680
acgctcaccg	gctccagatt	tatcagcaat	aaaccagcca	gccggaagg	ccgagcgag	1740
aagtggctct	gcaactttat	ccgcctccat	ccagtctatt	aattgttgcc	gggaagctag	1800
agtaagtagt	tcgccagtta	atagtttgcg	caacgttggt	gccattgcta	caggcatcgt	1860
ggtgtcacgc	tcgtcggttg	gtatggcttc	attcagctcc	ggttcccaac	gatcaagcgc	1920
agttacatga	tccccatgt	tgtgcaaaaa	agcggttagc	tccttcggtc	ctccgatcgt	1980
tgtcagaagt	aagttggccg	cagtgttatc	actcatggtt	atggcagcac	tgcataattc	2040
tcttactgtc	atgccatccg	taagatgctt	ttctgtgact	ggtgagtact	caaccaagtc	2100
attctgagaa	tagtgtatgc	ggcgaccgag	ttgctcttgc	ccggcgtaaa	tacgggataa	2160
taccgcgcca	catagcagaa	ctttaaaagt	gtgctatcatt	ggaaaacgtt	cttcggggcg	2220
aaaactctca	aggatcttac	cgctgttgag	atccagttcg	atgtaacca	ctcgtgcacc	2280
caactgatct	tcagcatctt	ttactttcac	cagcgtttct	gggtgagcaa	aaacaggaag	2340
gcaaaatgcc	gcaaaaaagg	gaataagggc	gacacggaaa	tggtgaatac	tcatcaattg	2400
cctttttcaa	tattattgaa	gcatttatca	gggttattgt	ctcatgagcg	gttacatatt	2460
tgaatgtatt	tagaaaaata	aacaaatagg	ggttcgcgc	acatttcccc	gaaaagtgc	2520
acctgacgtc	taagaaacca	ttattatcat	gacattaacc	tataaaaata	ggcgtatcac	2580
gaggcccttt	catctcgcgc	gtttcgggtg	tgacggtgaa	aacctctgac	acatgcagct	2640
cccggagaca	gtcacagctt	gtctgtaagc	ggatgccggg	agcagacaag	cccgtcaggg	2700
cgcgctcagc	ggtgttgccg	ggtgtcgggg	ctg			2733

<210> 213

<211> 2733

<212> DNA

<213> 人工序列

<220>

<223> pRYSE 输入载体 8

<400> 213

gtaaaacgac	ggccagtatt	aaccctcact	aaagggaact	cgaggctctt	caaaggcgcg	60
ccacgggtcgt	gcggattggc	agactccata	tgctatgcgg	catcagagca	gattgtactg	120
agagtgcacc	atatgcgggtg	tgaaataccg	cacagatgcg	taaggagaaa	ataccgcatac	180
aggcgccatt	cgccattcag	gctgcgcaac	tggtgggaag	ggcgatcgg	gcgggcctct	240
tcgctattac	gccagctggc	gaaaggggga	tgtgctgcaa	ggcgattaag	ttgggttaacg	300
ccagggtttt	cccagtcacg	acgttgtaaa	acgacggcca	gtgaattcga	gctcgggtacc	360
cggggatacct	ctagcgtcga	cctgcaggca	tgcaagcttg	gcgtaatcat	ggtcatagct	420
gtttcctgtg	tgaaattggt	atccgctcac	aattccacac	aacatacgag	ccggaagcat	480
aaagtgtaaa	gcctgggggtg	cctaattgagt	gagctaactc	acattaattg	cgttgcgcgc	540
tagcgagtca	tccaagcccc	tcagcccccc	tagcgtcgtg	aagagcgagc	tcccgtgag	600
caataactag	cgtcatagct	gtttcctggg	tcgttcggct	gcggcgagcg	gtatcacgtc	660
actcaaaggc	ggtaatacgg	ttatccacag	tacagggga	taacgcagga	aagaacatgt	720
gagcaaaagg	ccagcaaaag	gccaggaacc	gtaaaaaggc	cgcgttgctg	gcgtttttcc	780
ataggctccg	ccccctgac	gagcatcaca	aaaatcgacg	ctcaagtcag	aggtggcgaa	840
acccgacagg	actataaaga	taccaggcgt	ttccccctgg	aagctccctc	gtgcgctctc	900
ctgttccgac	cctgccgctt	acccgatacc	tgtccgcctt	tctcccttcg	ggaagcgtgg	960
cgctttctca	tagctcacgc	tgtaggtatc	tcagttcgg	gtaggtcgtt	cgctccaagc	1020
tgggctgtgt	gcacgaaccc	cccgttcagc	cgcaccgctg	cgcttatcc	ggtaactatc	1080
gtcttgattc	caaccgggta	agacacgact	tatcgccact	ggcagcagcc	actggtaaca	1140
ggattagcag	agcgagggtat	gtaggcggtg	ctacagagtt	cttgaagtgg	tggcctaact	1200
acggctacac	tagaagaaca	gtattttggt	tctgcgctct	gctgaagcca	gttaccttcg	1260
gaaaaagagt	tggtagctct	tgatccggca	aacaaaccac	cgctggtagc	ggtggttttt	1320
ttgtttgcaa	gcagcagatt	acgcgcagaa	aaaaaggatc	tcaagaagat	cctttgatct	1380
tttctacggg	gtctgacgct	cagtggaaacg	aaaactcacg	ttaagggatt	ttggtcatga	1440
gattatcaaa	aaggatcttc	acctagatcc	ttttaaatta	aaaatgaagt	tttaaataca	1500
tctaaagtat	atatgagtaa	cttggtcgca	tgcttaccaa	tgcttaatca	gtgaggcacc	1560
tatctcagcg	atctgtctat	ttcgttcac	catagttgcc	tgactgccc	tcgtgtagat	1620
aactacgata	cgggagggct	taccatctgg	ccccagtgct	gcaatgatac	cgcgagacc	1680
acgctcaccg	gctccagatt	tatcagcaat	aaaccagcca	gccggaagg	ccgagcgag	1740
aagtggctct	gcaactttat	ccgcctccat	ccagtctatt	aattgttgcc	gggaagctag	1800
agtaagtagt	tcgccagtta	atagtttgcg	caacgttggt	gccattgcta	caggcatcgt	1860

[0062]

gggtgtcacgc	tcgttcgtttg	gatatggcttc	attcagctcc	ggttcccaac	gatcaaggcg	1920
agttacatga	tccccatgt	tgtgcaaaaa	agcggttagc	tccttcggtc	ctccgatcgt	1980
tgtcagaagt	aagttggccg	cagtgttatc	actcatgggt	atggcagcac	tgcataattc	2040
tcttactgtc	atgccatccg	taagatgctt	ttctgtgact	ggtagtact	caaccaagtc	2100
attctgagaa	tagtgtatgc	ggcgaccgag	ttgctcttgc	ccggcgtaa	tacgggataa	2160
taccgcgcca	catagcagaa	ctttaaaagt	gctcatcatt	ggaaaacggt	cttcggggcg	2220
aaaactctca	aggatcttac	cgctgttgag	atccagttcg	atgtaaccga	ctcgtgcacc	2280
caactgatct	tcagcatctt	ttactttcac	cagcgtttct	gggtgagcaa	aaacaggaag	2340
gcaaaatgcc	gcaaaaaagg	gaataagggc	gacacggaaa	tgttgaatac	tcatcaattg	2400
cctttttcaa	tattattgaa	gcatttatca	gggttattgt	ctcatgagcg	gttacatatt	2460
tgaatgtatt	tagaaaaata	aacaaatagg	ggttccgcgc	acatttcccc	gaaaagtgc	2520
acctgacgtc	taagaaacca	ttattatcat	gacattaacc	tataaaaaata	ggcgtatcac	2580
gaggcccttt	catctcgcgc	gtttcgggtga	tgacgggtgaa	aacctctgac	acatgcagct	2640
cccggagaca	gtcacagctt	gtctgtgaagc	ggatgccggg	agcagacaag	cccgtcaggg	2700
cgcgctcagcg	gggtgttgccg	gggtgtcgggg	ctg			2733

<210> 214

<211> 2737

<212> DNA

<213> 人工序列

<220>

<223> pRYSE 输入载体 9

<400> 214

gtaaaacgc	ggccagtatt	aacctcact	aaagggaact	cgaggctctt	cacgctcgtc	60
caacgcccgc	ggaccttggc	agactccata	tgctatgcgg	catcagagca	gattgtactg	120
agagtgcacc	atatgcgggtg	tgaaataccg	cacagatgcg	taaggagaaa	ataccgcatac	180
aggcgccatt	cgccattcag	gctgcgcaac	tggtgggaag	ggcgatcggg	gcgggcctct	240
tcgctattac	gccagctggc	gaaaggggga	tgtgctgcaa	ggcgattaag	ttgggttaacg	300
ccagggtttt	cccagtcacg	acgttgtaaa	acgacggcca	gtgaattcga	gctcgggtacc	360
cggggactct	ctagcgtcga	cctgcaggca	tgcaagcttg	gcgtaatcat	ggcatagct	420
gtttcctgtg	tgaaattggt	atccgctcac	aattccacac	aacatacgag	ccggaagcat	480
aaagtgtaaa	gcctgggggtg	cctaattgagt	gagctaactc	acattaattg	cgttgcgcgc	540
tagcgagtca	tccacgggtgt	ttaaacccca	gcgcctggcg	ggtgaagagc	gagctccgc	600
tgagcaataa	ctagcgtcat	agctgttttc	tggttcgctt	ggctgcggcg	agcggatatca	660
gctcactcaa	aggcggtaat	acggttatcc	acagaatcag	gggataacgc	aggaaagaac	720
atgtgagcaa	aaggccagca	aaaggccagg	aaccgtaaaa	aggccgcgtt	gctggcgctt	780
ttccataggc	tccgcccccc	tgacgagcat	cacaaaaatc	gacgctcaag	tcagagggtg	840
cgaaacccga	caggactata	aagataccag	gcgtttcccc	ctggaagctc	cctcgtgcgc	900
tctcctgttc	cgaccctgcc	gcttaccgga	tacctgtccg	cctttctccc	ttcgggaagc	960
gtggcgcttt	ctcatagctc	acgctgtagg	tatctcagtt	cggtgtaggt	cgttcgtcc	1020
aagctgggct	gtgtgcacga	acccccggt	cagctccgac	gctgcgcctt	atccggtaac	1080
tatcgtcttg	attccaaccc	ggtaagacac	gacttatcgc	cactggcagc	agccactgggt	1140
aacaggatta	gcagagcgag	gtatgtaggc	ggtgctacag	agttcttgaa	gtgggtggcct	1200
aactacggct	acactagaag	aacagtatct	ggtatctgcg	ctctgctgaa	gccagttacc	1260
ttcggaaaaa	gagttggtag	ctcttgatcc	ggcaaaacaa	ccaccgctgg	tagcgggtgg	1320
ttttttgttt	gcaagcagca	gattacgcgc	agaaaaaaag	gatctcaaga	agatcctttg	1380
atcttttcta	cggggtctga	cgctcagttg	aacgaaaact	cacgttaagg	gattttggct	1440
atgagattat	caaaaaggat	cttcacctag	atccttttta	attaaaaatg	aagttttaaa	1500
tcaatctaaa	gtatatatga	gtaacttggt	cgcatgctta	ccaatgctta	atcagtgagg	1560
cacctatctc	agcgatctgt	ctatttcggt	catccatagt	tgctgactg	cccgtcgtgt	1620
agataactac	gatacgggag	ggcttaccat	ctggccccag	tgctgcaatg	ataccgcgag	1680
acccacgctc	accggctcca	gatttatcag	caataaacca	gccagccgga	aggccgagc	1740
gcagaagtgg	tcctgcaact	ttatccgcct	ccatccagtc	tattaattgt	tgccgggaag	1800
ctagagtaag	tagttcgcca	gttaatagtt	tgcgcaacgt	tggtgccatt	gctacaggca	1860
tcgtgggtgtc	acgctcgtcg	tttggtatgg	cttcattcag	ctccggttcc	caacgatcaa	1920
ggcgagttac	atgatcccc	atgttgtgca	aaaaagcgg	tagctccttc	ggtcctccga	1980
tcgttgtcag	aagtaagttg	gccgcagtgt	tatcactcat	ggttatggca	gcactgcata	2040
attctcttac	tgtcatgcca	tccgtaagat	gcttttctgt	gactggtgag	tactcaacca	2100
agtcattctg	agaatagttg	atgcggcgac	cgagttgctc	ttgcccggcg	tcaatacggg	2160
ataataccgc	gccacatagc	agaactttta	aagtgtcat	cattggaaaa	cgttcttcgg	2220
ggcgaaaact	ctcaaggatc	ttaccgctgt	tgagatccag	ttcgaatgaa	cccactcgtg	2280
cccccaactg	tctttcagca	tcttttactt	tcaccagcgt	ttctgggtga	gcaaaaaacg	2340
gaaggcaaaa	tgccgcaaaa	aagggaataa	ggcgacacg	gaaatgttga	atactcatca	2400
attgcctttt	tcaatattat	tgaagcattt	atcagggtta	ttgtctcatg	agcggttaca	2460
tatttgaatg	tatttagaaa	aataaacaaa	taggggttcc	gcgcacattt	ccccgaaaag	2520

[0063]

tgccacctga	cgtctaagaa	accattatta	tcattgacatt	aacctataaa	aataggcgta	2580
tcacgaggcc	ctttcatctc	gcgcgtttcg	gtgatgacgg	tgaaaacctc	tgacacatgc	2640
agctcccgga	gacagtcaca	gcttgtctgt	aagcggatgc	cgggagcaga	caagcccgtc	2700
agggcgcgctc	agcgggtggt	ggcgggtgtc	ggggctg			2737

<210> 215

<211> 2737

<212> DNA

<213> 人工序列

<220>

<223> pRYSE 输入载体 10

<400> 215

gtaaaacgac	ggccagtatt	aaccctcact	aaagggaact	cgaggctctt	caatccccgc	60
gtgcttgcc	ggccgttgcc	agactccata	tgctatgcgg	catcagagca	gattgtactg	120
agagtgcacc	atatgcggtg	tgaaataccg	cacagatgcg	taaggagaaa	ataccgcatc	180
aggcgccatt	cgccattcag	gctgcgcaac	tggtgggaag	ggcgatcggt	gcgggcctct	240
tcgctattac	gccagctggc	gaaaggggga	tgtgctgcaa	ggcgattaag	ttgggtaacg	300
ccagggtttt	cccagtcacg	acgttgtaaa	acgacggcca	gtgaattcga	gctcggtagc	360
cggggatcct	ctagcgtcga	cctgcaggca	tgcaagcttg	gcgtaatcat	ggtcatagct	420
gtttcctgtg	tgaaattggt	atccgctcac	aattccacac	aacatacgag	ccggaagcat	480
aaagtgtaaa	gcctggggtg	cctaattgagt	gagctaactc	acattaattg	cgttgcgcgc	540
tagcgagtca	tccacggtgt	ttaaacccca	gcgcctggcg	ggtgaagagc	gagctccccg	600
tgagcaataa	ctagcgtcat	agctgtttcc	tgggtcggtc	ggctgcggcg	agcggtagtca	660
gctcactcaa	aggcggtaat	acggttatcc	acagaatcag	gggataacgc	aggaaagaac	720
atgtgagcaa	aaggccagca	aaaggccagg	aaccgtaaaa	aggccgcgtt	gctggcggtt	780
ttccataggc	tccgcccccc	tgacgagcat	cacaaaaatc	gacgctcaag	tcagagggtg	840
cgaaccgga	caggactata	aagataccag	gcgtttcccc	ctggaagctc	cctcgtgcgc	900
tctcctgttc	cgaccctgcc	gcttaccgga	tacctgtccg	cctttctccc	ttcggaagc	960
gtggcgcttt	ctcatagctc	acgctgtagg	tatctcagtt	cggtgtaggt	cgttcgctcc	1020
aagctgggct	gtgtgcacga	acccccggtt	cagccccgac	gctgcgcctt	atccggtaac	1080
tatcgtcttg	attccaaccc	ggtaagacac	gacttatcgc	cactggcagc	agccactggt	1140
aacaggatta	gcagagcgag	gtatgtaggc	ggtgctacag	agttcttgaa	gtggtggcct	1200
aactacggct	acactagaag	aacagtattt	ggtatctgcg	ctctgctgaa	gccagttacc	1260
ttcggaaaaa	gagttggtag	ctcttgatcc	ggcaaacaaa	ccaccgctgg	tagcggtagt	1320
ttttttgttt	gcaagcagca	gattacgcgc	agaaaaaaag	gatctcaaga	agatcctttg	1380
atcttttcta	cggggtctga	cgctcagttg	aacgaaaact	cacgttaagg	gattttggtc	1440
atgagattat	caaaaaggat	cttcacctag	atccttttaa	attaaaaatg	aagttttaaa	1500
tcaatctaaa	gtatatatga	gtaacttggt	cgcatgctta	ccaatgctta	atcagtgagg	1560
cacctatctc	agcgatctgt	ctatctcggt	catccatagt	tgccgtgactg	cccgtcgtgt	1620
agataactac	gatacgggag	ggcttaccat	ctggccccag	tgctgcaatg	ataccgcgag	1680
acccacgctc	accggctcca	gattttatcag	caataaacca	gccagccgga	agggccgagc	1740
gcagaagtgg	tcttgcgaact	ttatccgcct	ccatccagtc	tattaattgt	tgccgggaag	1800
ctagagtaag	tagttcgcca	gttaatagtt	tgcgcaacgt	tgttgccatt	gctacaggca	1860
tcgtgggtgc	acgctcgctg	tttggtatgg	cttcattcag	ctccgggttc	caacgatcaa	1920
ggcgagttac	atgatcccc	atgtttgtgca	aaaaagcgg	tagctccttc	ggtcctccga	1980
tcgtttgtcag	aagtaagttg	gccgcagttg	tatcactcat	ggttatggca	gactgcata	2040
attctcttac	tgcatgccca	tccgtaagat	gcttttctgt	gactggtgag	tactcaacca	2100
agtcattctg	agaatagttg	atgcggcgac	cgagttgtct	ttgcccggcg	tcaatacggg	2160
ataataccgc	gccacatagc	agaactttta	aagtgtcat	cattggaaaa	cgttcttcgg	2220
ggcgaaaact	ctcaaggatc	ttaccgctgt	tgagatccag	ttcgatgtaa	cccactcgtg	2280
cacccaactg	atcttcagca	tcttttactt	tcaccagcgt	ttctgggtga	gcaaaaacag	2340
gaaggcaaaa	tgccgcaaaa	aagggaataa	ggcgacacg	gaaatgttga	atactcatca	2400
attgcctttt	tcaatattat	tgaagcattt	atcagggtta	ttgtctcatg	agcggttaca	2460
tatttgaaatg	tatttagaaa	aataaaca	taggggttcc	gcgcacattt	ccccgaaaag	2520
tgccacctga	cgtctaagaa	accattatta	tcattgacatt	aacctataaa	aataggcgta	2580
tcacgaggcc	ctttcatctc	gcgcgtttcg	gtgatgacgg	tgaaaacctc	tgacacatgc	2640
agctcccgga	gacagtcaca	gcttgtctgt	aagcggatgc	cgggagcaga	caagcccgtc	2700
agggcgcgctc	agcgggtggt	ggcgggtgtc	ggggctg			2737

<210> 216

<211> 2737

<212> DNA

<213> 人工序列

<220>

[0064]

<223> pRYSE 输入载体 11

<400> 216

```

gtaaaacgac ggccagttatt aaccctcact aaaggggaact cgaggctctt caaacctgca 60
ggccgcgagc gccgattggc agactccata tgctatgcgg catcagagca gattgtactg 120
agagtgcacc atatgcggtg tgaaataccg cacagatgcg taaggagaaa ataccgcatac 180
aggcgccatt cgccattcag gctgcgcaac tgttggaag ggcgatcggc gcgggcctct 240
tcgctattac gccagctggc gaaaggggga tgtgctgcaa ggcgattaag ttgggtaacg 300
ccagggtttt cccagtcacg acgttgtaaa acgacggcca gtgaattcga gctcggtagc 360
cggggatcct ctagcgtcga cctgcaggca tgcaagcttg gcgtaatcat ggtcatagct 420
gtttcctgtg tgaaattggt atccgctcac aattccacac aacatacagc ccggaagcat 480
aaagtgtaaa gcctggggtg cctaattgag gagctaactc acattaattg cgttgcgcgc 540
tagcgagtca tccacggtgt ttaaacccca gcgcctggcg ggtgaagagc gagtcccg 600
tgagcaataa ctacgctcat agctgtttcc tgggtcggtc ggctgcggcg agcggatatca 660
gctcactcaa aggcggtaat acggttatcc acagaatcag gggataacgc aggaagaac 720
atgtgagcaa aaggccagca aaaggccagg aaccgtaaaa aggccgcgtt gctggcggtt 780
ttccataggc tccgcccccc tgacgagcat cacaaaaatc gacgctcaag tcagaggtag 840
cgaaaccgga caggactata aagataccag gcgtttcccc ctggaagctc cctcgtgcgc 900
tctcctgttc cgaccctgcc gcttaccgga tacctgtccg ctttctccc ttcggaagc 960
gtggcgcttt ctcatagctc acgctgtagg tatctcagtt cgggttaggt cggtcgtcc 1020
aagctgggct gtgtgcacga acccccgtt cagcccgcct gctgcgcctt atccggaac 1080
tatcgtcttg attccaaccc ggtaagacac gacttatcgc cactggcagc agccactggt 1140
aacaggatta gcagagcgag gtatgtaggc ggtgctacag agttcttgaa gtgggtggcct 1200
aactacggct aactagaag aacagtattt ggtatctgcg ctctgctgaa gccagttacc 1260
ttcggaaaaa gagttggtag ctcttgatcc ggcaaacaaa ccaccgctgg tagcggtagt 1320
ttttttgttt gcaagcagca gattacgcgc agaaaaaaag gatctcaaga agatcctttg 1380
atcttttcta cggggtctga cgctcagtg aacgaaaact cacgttaagg gattttggtc 1440
atgagattat caaaaaggat cttcacctag atccttttaa attaaaaatg aagttttaaa 1500
tcaatctaaa gtatatatga gtaacttggc cgcatgctta ccaatgctta atcagttagg 1560
cacctatctc agcgatctgt ctatttcgtt catccatagt tgcctgactg cccgtcgtgt 1620
agataactac gatacgggag ggcttaccat ctggccccag tgctgcaatg ataccgcgag 1680
accacgcctc accggctcca gatttatcag caataaacca gccagccgga agggccgagc 1740
gcagaagtgg tcctgcaact ttatccgcct ccatccagtc tattaattgt tgccgggaag 1800
ctagagtaag tagttcgcca gttaatagtt tgcgcaacgt tgttgccatt gctacaggca 1860
tcgtgggtgc acgctcgtcg tttggtatgg cttcattcag ctccggttcc caacgatcaa 1920
ggcgagttac atgatcccc atgttggtga aaaaagcggc tagctccttc ggtcctccga 1980
tcgttgctcag aagtaagttg gccgcagtg tctactcat ggttatggca gactgcata 2040
attctcttac tgcctgcca tccgtaagat gcttttctgt gactggtgag tactcaacca 2100
agtcattctg agaatagtg atgcggcgac cgagttgctc ttgcccggcg tcaatacggg 2160
ataataccgc gccacatagc agaactttaa aagtgtcat cattgaaaa cgttcttcgg 2220
ggcgaaaact ctcaaggatc ttaccgctgt tgagatccag ttcatgtaa cccactcgtg 2280
caccacactg atcttcagca tcttttactt tcaccagcgt ttctgggtga gcaaaaacag 2340
gaaggcaaaa tgccgcaaaa aagggaataa gggcgacac gaaatgttga atactcata 2400
attgcctttt tcaatattat tgaagcattt atcagggtta ttgtctcatg agcggttaca 2460
tatttgaatg tatttagaaa aataaacaaa taggggttcc gcgcacattt ccccgaag 2520
tgccacctga cgtctaagaa accattatta tcatgacatt aacctataaa aataggcgta 2580
tcacgaggcc ctttcatctc gcgcgtttcg gtgatgacgg tgaaaacctc tgacacatgc 2640
agctcccgga gacagtcaca gcttgctgt aagcggatgc cgggagcaga caagcccgtc 2700
agggcgcgtc agcgggtggt ggccgggtgtc ggggctg 2737

```

<210> 217

<211> 2737

<212> DNA

<213> 人工序列

<220>

<223> pRYSE 输入载体 12

<400> 217

```

gtaaaacgac ggccagttatt aaccctcact aaaggggaact cgaggctctt caaacgcgat 60
cgccgacgcc gccgattggc agactccata tgctatgcgg catcagagca gattgtactg 120
agagtgcacc atatgcggtg tgaaataccg cacagatgcg taaggagaaa ataccgcatac 180
aggcgccatt cgccattcag gctgcgcaac tgttggaag ggcgatcggc gcgggcctct 240
tcgctattac gccagctggc gaaaggggga tgtgctgcaa ggcgattaag ttgggtaacg 300
ccagggtttt cccagtcacg acgttgtaaa acgacggcca gtgaattcga gctcggtagc 360
cggggatcct ctagcgtcga cctgcaggca tgcaagcttg gcgtaatcat ggtcatagct 420
gtttcctgtg tgaaattggt atccgctcac aattccacac aacatacagc ccggaagcat 480

```

[0065]

aaagtgtaaa	gcctgggggtg	cctaattgagt	gagctaactc	acattaattg	cgttgcgcgc	540
tagcgagtca	tccacgggtgt	ttaaacccca	gcgcctggcg	ggtgaagagc	gagctccgc	600
tgagcaataa	ctagcgatcat	agctgtttcc	tgggtcggtc	ggctgcggcg	agcggatatca	660
gctcactcaa	aggcggtaat	acggttatcc	acagaatcag	gggataacgc	aggaaagaac	720
atgtgagcaa	aaggccagca	aaaggccagg	aaccgtaaaa	aggccgcgtt	gctggcggtt	780
ttccataggc	tccgcccccc	tgacgagcat	cacaaaaatc	gacgctcaag	tcagagggtg	840
cgaaccgccg	caggactata	aagataccag	gcgtttcccc	ctggaagctc	cctcgtgcgc	900
tctcctgttc	cgaccctgcc	gcttaccgca	tacctgtccg	cctttctccc	ttcgggaagc	960
gtggcgcttt	ctcatagctc	acgctgtagg	tatctcagtt	cggtgtaggt	cgttcgctcc	1020
aagctgggct	gtgtgcacga	accccccggt	cagcccagacc	gctgcgcctt	atccggtaac	1080
tatcgtcttg	attccaaccc	ggtaagacac	gacttatcgc	cactggcagc	agccactggt	1140
aacaggatta	gcagagcgag	gtatgtaggc	ggtgtctacag	agttcttgaa	gtgggtggcct	1200
aactacggct	acactagaag	aacagtattt	ggtatctgcg	ctctgtctgaa	gccagttacc	1260
ttcggaaaaa	gagttggtag	ctcttgatcc	ggcaaacaaa	ccaccgctgg	tagcgggtggt	1320
ttttttgttt	gcaagcagca	gattacgcgc	agaaaaaaag	gatctcaaga	agatcctttg	1380
atcttttcta	cggggtctga	cgctcagttg	aacgaaaaact	cacgttaagg	gattttggct	1440
atgagattat	caaaaaggat	cttcacctag	atccttttaa	attaaaaatg	aagtttttaa	1500
tcaatctaaa	gtatatatga	gtaacttggt	cgcatgctta	ccaatgctta	atcagtgagg	1560
cacctatctc	agcgatctgt	ctatttcggt	catccatagt	tgccctgactg	cccgtcgtgt	1620
agataactac	gatacgggag	ggcttaccat	ctggccccag	tgctgcaatg	ataccgcgag	1680
acccacgctc	accggctcca	gattttatcag	caataaacca	gccagccgga	agggccgagc	1740
gcagaagtgg	tccgtcaact	ttatccgcct	ccatccagtc	tattaattgt	tgccgggaag	1800
ctagagtaag	tagttcgcca	gttaatatgtt	tgcgcaacgt	tgttgccatt	gctacaggca	1860
tcgtgggtgtc	acgctcgtcg	tttggtatgg	cttcattcag	ctccggttcc	caacgatcaa	1920
ggcgagttac	atgatcccc	atgtttgtgca	aaaaagcgg	tagctccttc	ggtcctccga	1980
tcgtttgtcag	aagtaagtgt	gccgcagttg	tatcactcat	ggttatggca	gcactgcata	2040
attctcttac	gtcatgccca	tccgtaagat	gcttttctgt	gactggtgag	tactcaacca	2100
agtcattctg	agaatagtgt	atgcggcgac	cgagttgtct	ttgcccggcg	tcaatacggg	2160
ataataccgc	gccacatagc	agaactttta	aagtgtctcat	cattggaaaa	cgttcttcgg	2220
ggcgaaaact	ctcaaggatc	ttaccgctgt	tgagatccag	ttcgatgtaa	cccactcgtg	2280
cacccaactg	atcttcagca	tcttttactt	tcaccagcgt	ttctgggtga	gcaaaaacag	2340
gaaggcaaaa	tgcgcgaaaa	aagggaataa	gggcgacacg	gaaatgttga	atactcatca	2400
attgcctttt	tcaatattat	tgaagcattt	atcagggtta	ttgtctcatg	agcggttaca	2460
tatttgaaatg	tatttagaaa	aataaacaaa	taggggttcc	gcgcacattt	ccccgaaaag	2520
tgccacctga	cgtctaagaa	accattatta	tcatgacatt	aacctataaa	aataggcgta	2580
tcacgaggcc	ctttcatctc	gcgcgtttcg	gtgatgacgg	tgaaaacctc	tgacacatgc	2640
agctcccgga	gacagtcaca	gcttgtctgt	aagcggatgc	cgggagcaga	caagcccgtc	2700
agggcgcgctc	agcgggtgtt	ggcgggtgtc	ggggctg			2737

<210> 218

<211> 2737

<212> DNA

<213> 人工序列

<220>

<223> pRYSE 输入载体 13

<400> 218

gtaaaaacgac	ggccagttatt	aaccctcact	aaagggaact	cgaggctctt	caaaggcggc	60
cgctggcgag	ggagattggc	agactccata	tgctatgcgg	catcagagca	gattgtactg	120
agagtgcacc	atatgcggtg	tgaaataccg	cacagatgcg	taaggagaaa	ataccgcata	180
aggcgccatt	cgccattcag	gctgcgcaac	tggtgggaag	ggcgatcgg	gccccctct	240
tcgctattac	gccagctggc	gaaaggggga	tgtgctgcaa	ggcgattaag	ttgggtaacg	300
ccagggtttt	cccagtcacg	acgttgtaaa	acgacggcca	gtgaattcga	gctcggtaac	360
cggggatcct	ctagcgctga	cctgcaggca	tgcaagcttg	gcgtaatcat	ggtcatagct	420
gtttcctgtg	tgaaattggt	atccgctcac	aattccacac	aacatacag	ccggaagcat	480
aaagtgtaaa	gcctgggggtg	cctaattgagt	gagctaactc	acattaattg	cgttgcgcgc	540
tagcgagtca	tccacgggtgt	ttaaacccca	gcgcctggcg	ggtgaagagc	gagctccgc	600
tgagcaataa	ctagcgatcat	agctgtttcc	tgggtcggtc	ggctgcggcg	agcggatatca	660
gctcactcaa	aggcggtaat	acggttatcc	acagaatcag	gggataacgc	aggaaagaac	720
atgtgagcaa	aaggccagca	aaaggccagg	aaccgtaaaa	aggccgcgtt	gctggcggtt	780
ttccataggc	tccgcccccc	tgacgagcat	cacaaaaatc	gacgctcaag	tcagagggtg	840
cgaaccgccg	caggactata	aagataccag	gcgtttcccc	ctggaagctc	cctcgtgcgc	900
tctcctgttc	cgaccctgcc	gcttaccgca	tacctgtccg	cctttctccc	ttcgggaagc	960
gtggcgcttt	ctcatagctc	acgctgtagg	tatctcagtt	cggtgtaggt	cgttcgctcc	1020
aagctgggct	gtgtgcacga	accccccggt	cagcccagacc	gctgcgcctt	atccggtaac	1080
tatcgtcttg	attccaaccc	ggtaagacac	gacttatcgc	cactggcagc	agccactggt	1140

[0066]


```

aacaggatta gcagagcgag gtatgtaggc ggtgctacag agttcttgaa gtggtggcct 1200
aactacggct acactagaag aacagtatctt ggtatctgag ctctgctgaa gccagttacc 1260
ttcggaaaaa gagttggtag ctcttgatcc ggcaaacaaa ccaccgctgg tagcggtggt 1320
ttttttgttt gcaagcagca gattacgcgc agaaaaaaag gatctcaaga agatcctttg 1380
atcttttcta cggggtctga cgctcagtg aacgaaaact cacgttaagg gattttggctc 1440
atgagattat caaaaaggat cttcacctag atccttttaa attaaaaatg aagttttaaa 1500
tcaatctaaa gtatatatga gtaacttggt cgcattgctta ccaatgctta atcagtggag 1560
cacctatctc agcgatctgt ctatttcgtt catccatagt tgcctgactg cccgtcgtgt 1620
agataactac gatacgggag ggcttaccat ctggccccag tgctgcaatg ataccgcgag 1680
acccacgctc accggctcca gatttatcag caataaacca gccagccgga agggccgagc 1740
gcagaagtgg tcctgcaact ttatccgcct ccatccagtc tattaattgt tgccgggaag 1800
ctagagtaag tagttcgcca gttaatatgt tgcgcaacgt tgttgccatt gctacaggca 1860
tcgtggtgtc acgctcgtcg ttggtatgg cttcattcag ctccggttcc caacgatcaa 1920
ggcgagttac atgatcccc atgttggtgca aaaaagcggg tagctccttc ggtcctccga 1980
tcgttgctcag aagtaagttg gccgcagtg tctactcat ggttatggca gactgcata 2040
attctcttac gtcatgcca tccgtaagat gcttttctgt gactggtgag tactcaacca 2100
attgcctttt tcaatattat tgaagcattt atcagggtta ttgtctcatg agcggttaca 2160
ataataccgc gccacatagc agaactttta aagtgtcat cattggaaaa cgttcttcgg 2220
ggcgaaaact ctcaaggatc ttaccgctgt tgagatccag ttcgatgtaa cccactcgtg 2280
cacccaactg atcttcagca tcttttactt tcaccagcgt ttctgggtga gcaaaaacag 2340
gaaggcaaaa tgccgcaaaa aagggaataa gggcgacacg gaaatgttga atactcatca 2400
ttgccttttt tcaatattat tgaagcattt atcagggtta ttgtctcatg agcggttaca 2460
tatttgaaatg tatttagaaa aataaacaaa taggggttcc gcgcacattt ccccgaaaag 2520
tgccacctga cgtctaagaa accattatta tcatgacatt aacctataaa aataggcgta 2580
tcacgaggcc ctttcatctc gcgcgtttcg gtgatgacgg tgaaaacctc tgacacatgc 2640
agctcccgga gacagtcaca gcttgtctgt aagcggatgc cgggagcaga caagcccgtc 2700
agggcgcgctc agcgggtgtt ggcgggtgtc ggggctg 2737

```

<210> 219

<211> 2737

<212> DNA

<213> 人工序列

<220>

<223> pRYSE 输入载体 14

<400> 219

```

gtaaaacgac ggccagtatt aaccctcact aaagggaact cgaggctctt caaaggcgcg 60
ccacggctgt gcggtattggc agactccata tgctatgcgg catcagagca gattgtactg 120
agagtgcacc atatgcggtg tgaataaccg cacagatgcg taaggagaaa ataccgcatac 180
aggcgccatt cgccattcag gctgcgcaac tgttggaag gccgatcggg gcgggcctct 240
tcgctattac gccagctggc gaaaggggga tgtgctgcaa ggcgattaag ttgggtaacg 300
ccagggtttt ccaggtcacg acgttgtaaa acgacggcca gtgaattcga gctcggttacc 360
cggggatcct ctagcgtcga cctgcaggca tgcaagcttg gcgtaatcat ggtcatagct 420
gtttcctgtg tgaaattggt atccgctcac aattccacac aacatacgag cgggaagcat 480
aaagtgtaaa gcctgggggt cctaattagt gagctaactc acattaattg cgttgcgcgc 540
tagcgagtca tccacgggtgt ttaaacccca gcgcctggcg ggtgaagagc gagctcccgc 600
tgagcaataa ctacgctcat agctgtttcc ttgggtcgctt ggctgcggcg agcggatca 660
gctcactcaa aggcggtaat acggttatcc acagaatcag gggataacgc aggaaagaac 720
atgtgagcaa aaggccagca aaaggccagg aaccgtaaaa aggccgcgtt gctggcggtt 780
ttccataggc tccgcccccc tgacgagcat cacaaaaatc gacgctcaag tcagagggtg 840
cgaaacccga caggactata aagataccag gcgtttcccc ctggaagctc cctcgtgcgc 900
tctcctgttc cgaccctgcc gcttaccgca tacctgtccg ctttctccc ttccgggaagc 960
gtggcgcttt ctcatagctc acgctgtagg tatctcagtt cgggtgtaggt cgttcgctcc 1020
aagctgggct gtgtgcacga acccccgtt cagcccagac gctgcgcctt atccggtaac 1080
tatcgtcttg attccaaccc ggtaagacac gacttatcgc cactggcagc agccactggt 1140
aacaggatta gcagagcgag gtatgtaggc ggtgctacag agttcttgaa gtggtggcct 1200
aactacggct acactagaag aacagtatctt ggtatctgag ctctgctgaa gccagttacc 1260
ttcggaaaaa gagttggtag ctcttgatcc ggcaaacaaa ccaccgctgg tagcggtggt 1320
ttttttgttt gcaagcagca gattacgcgc agaaaaaaag gatctcaaga agatcctttg 1380
atcttttcta cggggtctga cgctcagtg aacgaaaact cacgttaagg gattttggctc 1440
atgagattat caaaaaggat cttcacctag atccttttaa attaaaaatg aagttttaaa 1500
tcaatctaaa gtatatatga gtaacttggt cgcattgctta ccaatgctta atcagtggag 1560
cacctatctc agcgatctgt ctatttcgtt catccatagt tgcctgactg cccgtcgtgt 1620
agataactac gatacgggag ggcttaccat ctggccccag tgctgcaatg ataccgcgag 1680
acccacgctc accggctcca gatttatcag caataaacca gccagccgga agggccgagc 1740
gcagaagtgg tcctgcaact ttatccgcct ccatccagtc tattaattgt tgccgggaag 1800

```

[0067]

ctagagtaag	tagttcgcca	gttaatagtt	tgcgcaacgt	tgttgccatt	gctacaggca	1860
tcgtggtgtc	acgctcgtcg	tttggtatgg	cttcattcag	ctccggttcc	caacgatcaa	1920
ggcgagttac	atgatccccc	atgtttgtgca	aaaaagcgg	tagctccttc	ggctctccga	1980
tcgtttgtcag	aagtaagttg	gccgcagtg	tatcactcat	ggttatggca	gcactgcata	2040
attctcttac	tgtcatgcca	tccgtaagat	gcttttctgt	gactggtgag	tactcaacca	2100
agtcattctg	agaatagtgt	atgcggcgac	cgagttgctc	ttgcccggcg	tcaatacggg	2160
ataataccgc	gccacatagc	agaactttaa	aagtgtcat	cattggaaaa	cgttcttcgg	2220
ggcgaaaact	ctcaaggatc	ttaccgctgt	tgagatccag	ttcgatgtaa	cccactcgtg	2280
cacccaactg	atcttcagca	tcttttactt	tcaccagcgt	ttctgggtga	gcaaaaacag	2340
gaaggcaaaa	tgccgcaaaa	aagggaataa	ggcgacacg	gaaatgttga	atactcatca	2400
attgcctttt	tcaatattat	tgaagcattt	atcagggtta	ttgtctcatg	agcggttaca	2460
tatttgaatg	tatttagaaa	aataaaciaa	taggggttcc	gcgcacattt	ccccgaaaag	2520
tgccacctga	cgtctaagaa	accattatta	tcattgacatt	aacctataaa	aataggcgta	2580
tcacgaggcc	ctttcatctc	gcgcgtttcg	gtgatgacgg	tgaaaacctc	tgacacatgc	2640
agctcccggg	gacagtcaca	gcttgtctgt	aagcggatgc	cgaggacaga	caagcccgtc	2700
agggcgcgct	agcgggtgtt	ggcgggtgtc	ggggctg			2737

<210> 220

<211> 2737

<212> DNA

<213> 人工序列

<220>

<223> pRYSE 输入载体 15

<400> 220

gtaaaacgac	ggccagtatt	aaccctcact	aaagggaact	cgaggctctt	caagcccctc	60
agccccctta	gcgtcgtggc	agactccata	tgctatgcgg	catcagagca	gattgtactg	120
agagtgcacc	atatgcggtg	tgaaataccg	cacagatgcg	taaggagaaa	ataccgcata	180
agggcgccatt	cgccattcag	gctgcgcaac	tgttgggaag	ggcgatcggg	gcgggcctct	240
tcgctattac	gccagctggc	gaaaggggga	tgtgctgcaa	ggcgattaag	ttgggttaacg	300
ccaggggtttt	cccagtcacg	acgttgtaaa	acgacggcca	gtgaattcga	gctcggtacc	360
cggggactct	ctagcgtcga	cctgcaggca	tgcaagcttg	gcgtaatcat	ggtcatagct	420
gtttcctgtg	tgaaattgtt	atccgctcac	aattccacac	aacatacag	ccggaagcat	480
aaagtgtaaa	gcctgggggtg	cctaattgag	gagctaactc	acattaattg	cgttgcgcgc	540
tagcgagtca	tccacggtgt	ttaaacccca	gcgcctggcg	ggtgaagagc	gagctcccgc	600
tgagcaataa	ctagcgtcat	agctgttttc	tgggtcggtc	ggctgcggcg	agcggtatca	660
gctcactcaa	aggcggtaat	acggttatcc	acagaatcag	gggataacgc	aggaaagaac	720
atgtgagcaa	aaggccagca	aaaggccagg	aaccgtaaaa	aggccgcgtt	gctggcggtt	780
ttccataggc	tccgcccccc	tgacgagcat	cacaaaaatc	gacgctcaag	tcagaggtgg	840
cgaaacccga	caggactata	aagataccag	gcgtttcccc	ctggaagctc	cctcgtgcgc	900
tctcctgttc	cgaccctgcc	gcttaccgga	tactgttccg	cctttctccc	ttcgggaagc	960
gtggcgcttt	ctcatagctc	acgctgtagg	tatctcagtt	cggtgtaggt	cggtcgctcc	1020
aagctgggct	gtgtgcacga	acccccggtt	cagcccagac	gctgcgcctt	atccggtaac	1080
tatcgtcttg	attccaaccc	ggtaagacac	gacttatcgc	cactggcagc	agccactggt	1140
aacaggatta	gcagagcgag	gtatgtaggc	ggtgctacag	agttcttgaa	gtgggtggcct	1200
aactacggct	acactagaag	aacagtattt	ggtatctgcg	ctctgctgaa	gccagttacc	1260
ttcggaaaaa	gagtgtgtag	ctcttgatcc	ggcaaacaaa	ccaccgctgg	tagcgtgggt	1320
ttttttgttt	gcaagcagca	gattacgcgc	agaaaaaaag	gatctcaaga	agatcctttg	1380
atcttttcta	cggggtctga	cgctcagttg	aacgaaaact	cacgttaagg	gattttgggtc	1440
atgagattat	caaaaaggat	cttcacctag	atccttttaa	attaaaaatg	aagttttaaa	1500
tcaatctaaa	gtatatatga	gtaacttggt	cgcattgctta	ccaatgctta	atcagtggag	1560
cacctatctc	agcgatctgt	ctatttctgt	catccatagt	tgccctgactg	cccgtcgtgt	1620
agataactac	gatacgggag	ggcttaccat	ctggccccag	tgctgcaatg	ataccgcgag	1680
accacgcctc	accggctcca	gattttatcag	caataaacca	gccagccgga	agggccgagc	1740
gcagaagtgg	tcctgcaact	ttatccgcct	ccatccagtc	tattaattgt	tgccgggaag	1800
ctagagtaag	tagttcgcca	gttaatagtt	tgcgcaacgt	tgttgccatt	gctacaggca	1860
tcgtggtgtc	acgctcgtcg	tttggtatgg	cttcattcag	ctccggttcc	caacgatcaa	1920
ggcgagttac	atgatccccc	atgtttgtgca	aaaaagcgg	tagctccttc	ggctctccga	1980
tcgtttgtcag	aagtaagttg	gccgcagtg	tatcactcat	ggttatggca	gcactgcata	2040
attctcttac	tgtcatgcca	tccgtaagat	gcttttctgt	gactggtgag	tactcaacca	2100
agtcattctg	agaatagtgt	atgcggcgac	cgagttgctc	ttgcccggcg	tcaatacggg	2160
ataataccgc	gccacatagc	agaactttaa	aagtgtcat	cattggaaaa	cgttcttcgg	2220
ggcgaaaact	ctcaaggatc	ttaccgctgt	tgagatccag	ttcgatgtaa	cccactcgtg	2280
cacccaactg	atcttcagca	tcttttactt	tcaccagcgt	ttctgggtga	gcaaaaacag	2340
gaaggcaaaa	tgccgcaaaa	aagggaataa	ggcgacacg	gaaatgttga	atactcatca	2400
attgcctttt	tcaatattat	tgaagcattt	atcagggtta	ttgtctcatg	agcggttaca	2460

[0068]

tatttgaatg	tatttagaaa	aataaacaaa	taggggttcc	gcgcacattt	ccccgaaaag	2520
tgccacctga	cgtctaagaa	accattatta	tcatgacatt	aacctataaa	aataggcgta	2580
tcacgaggcc	ctttcatctc	gcgcgtttcg	gtgatgacgg	tgaaaacctc	tgacacatgc	2640
agctcccgga	gacagtcaca	gcttgtctgt	aagcggatgc	cgggagcaga	caagcccgtc	2700
agggcgcgtc	agcgggtggt	ggcgggtgtc	ggggctg			2737

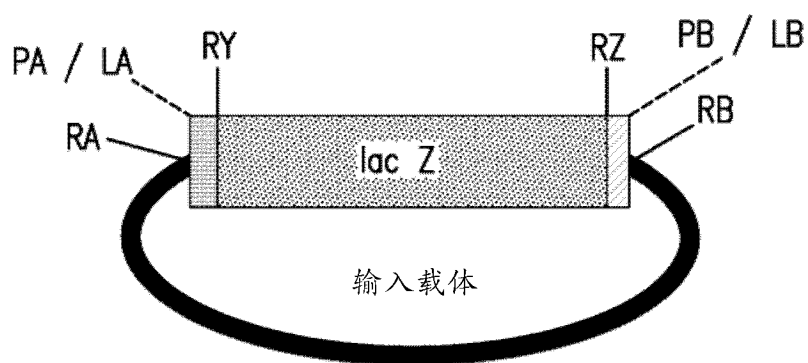


图 1A

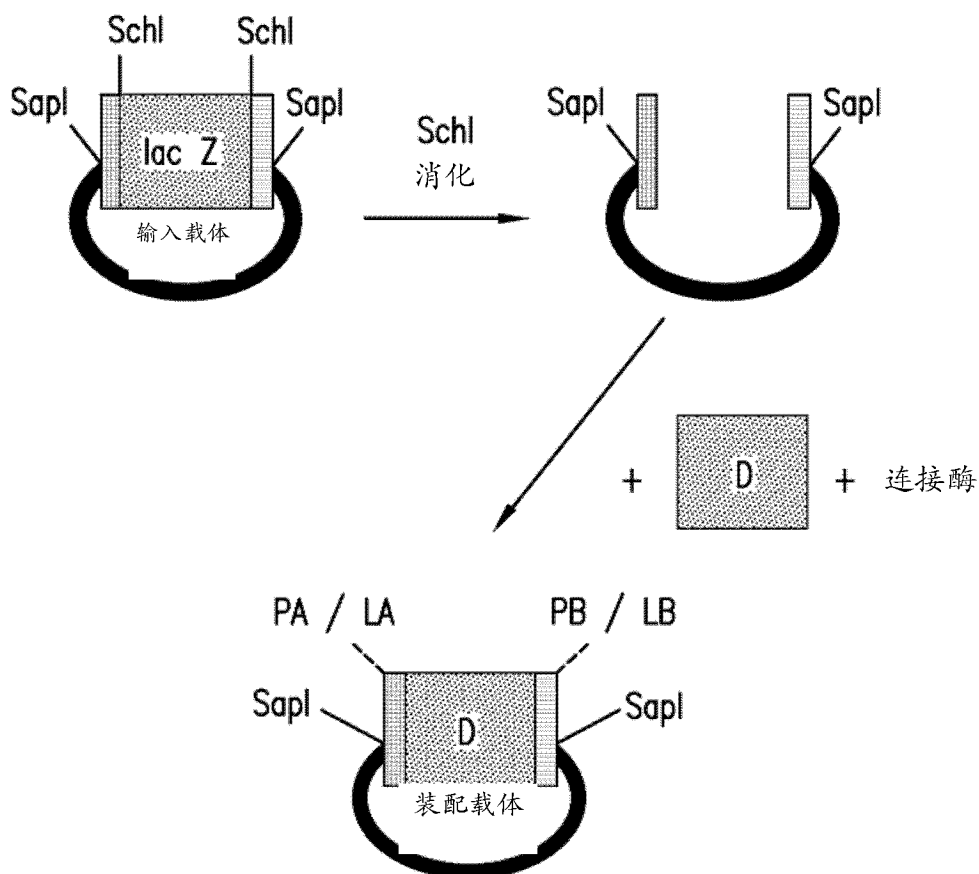


图 1B

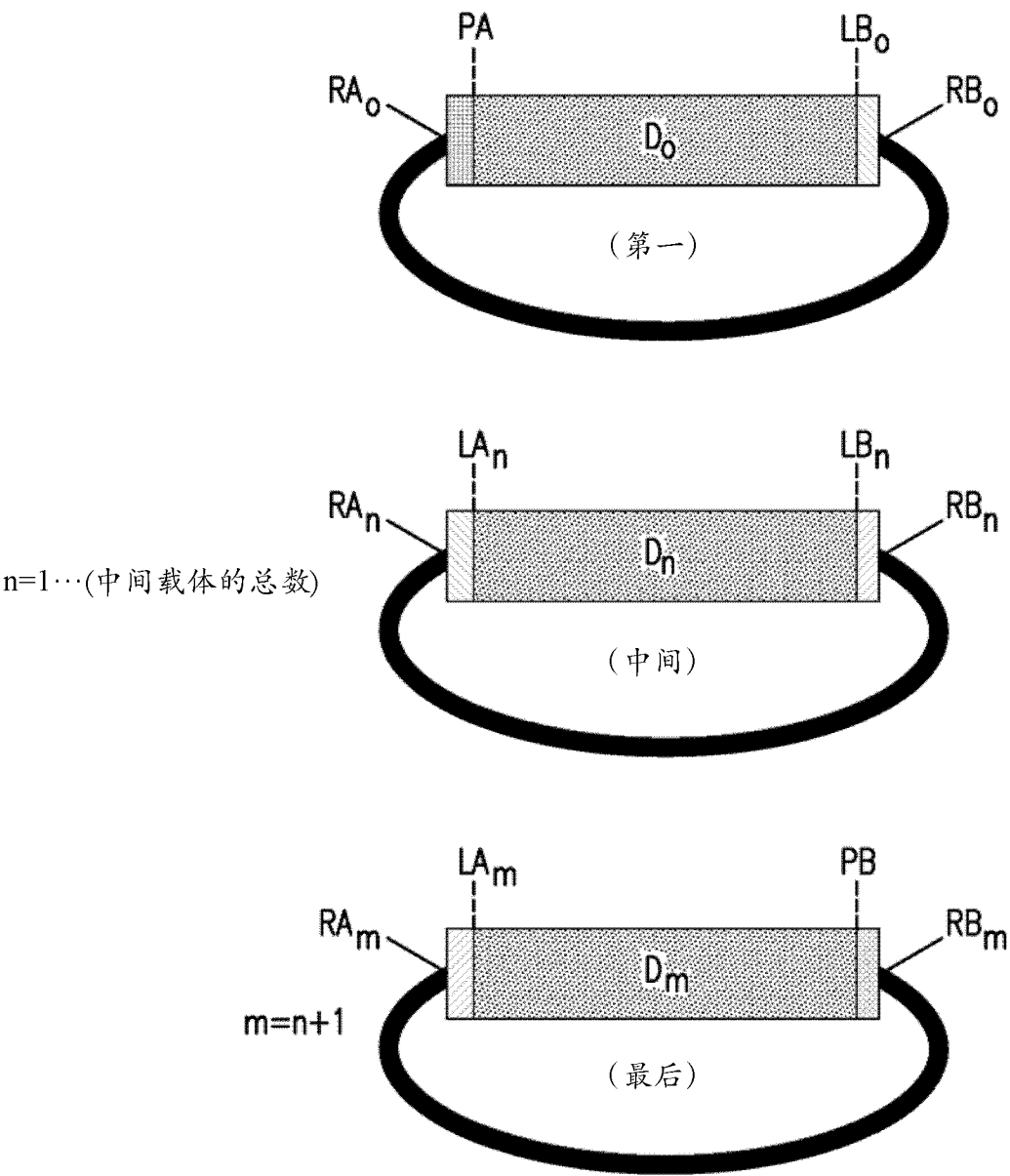


图 2

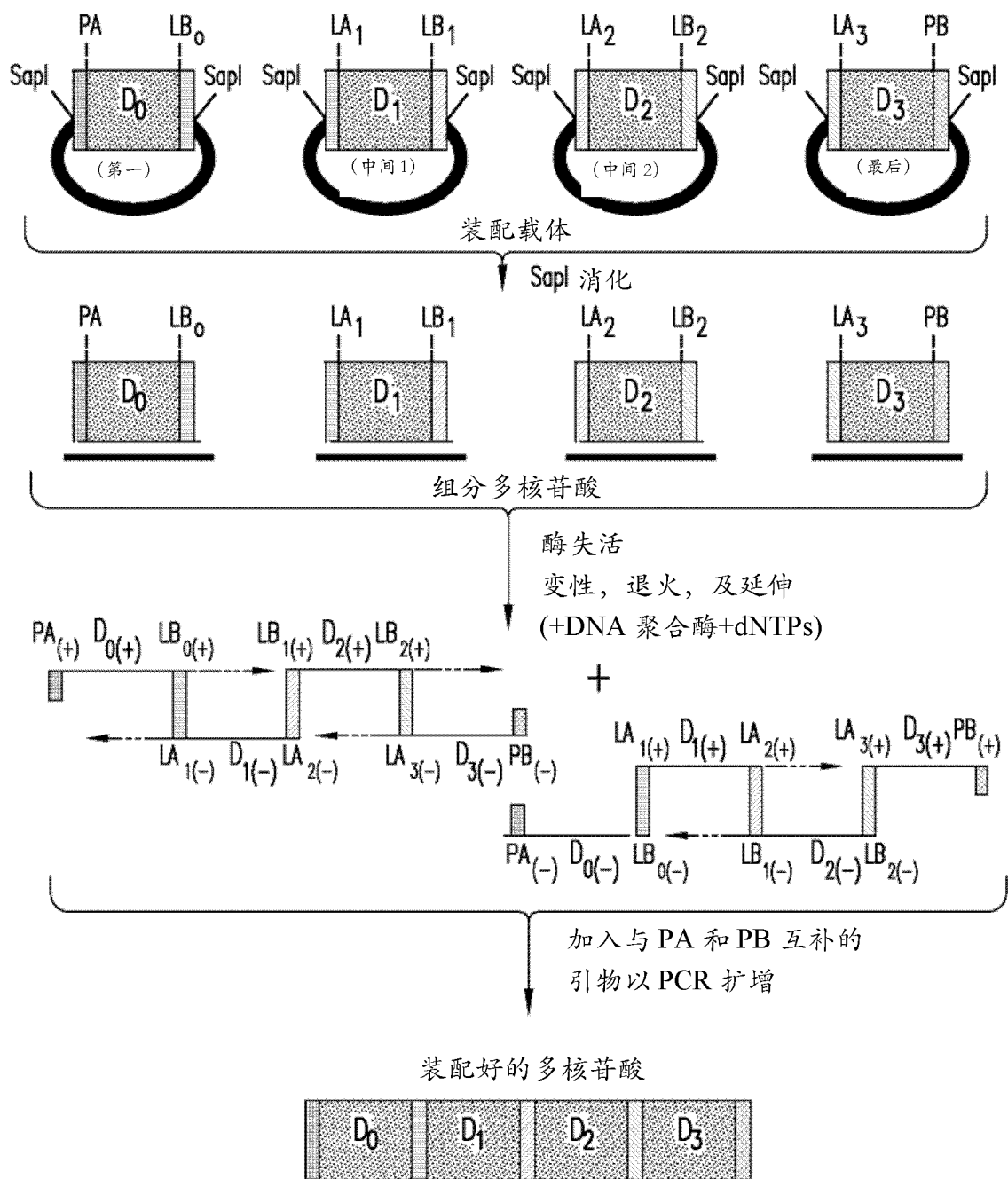


图 3

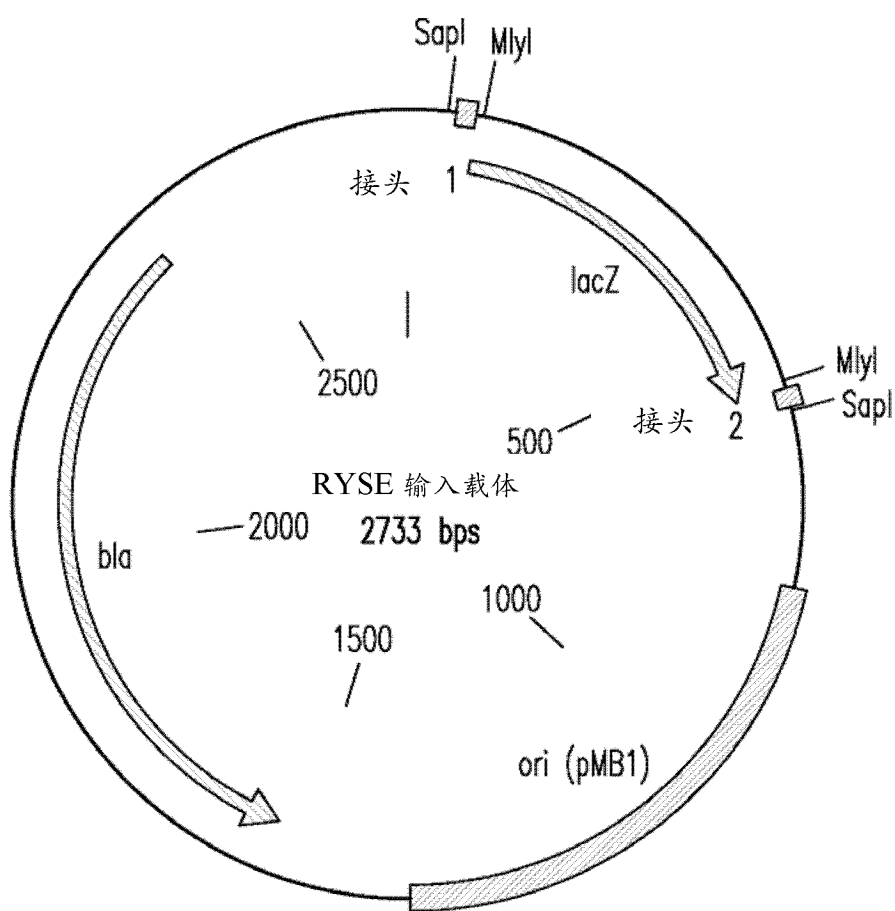


图 4

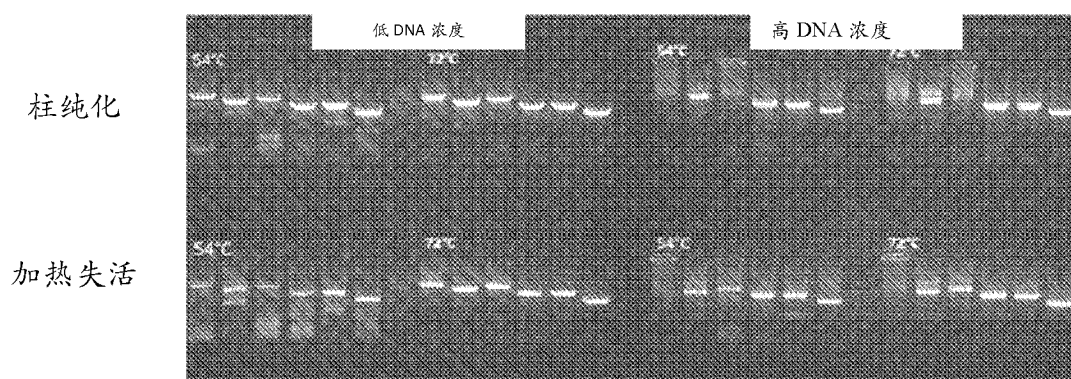


图 5

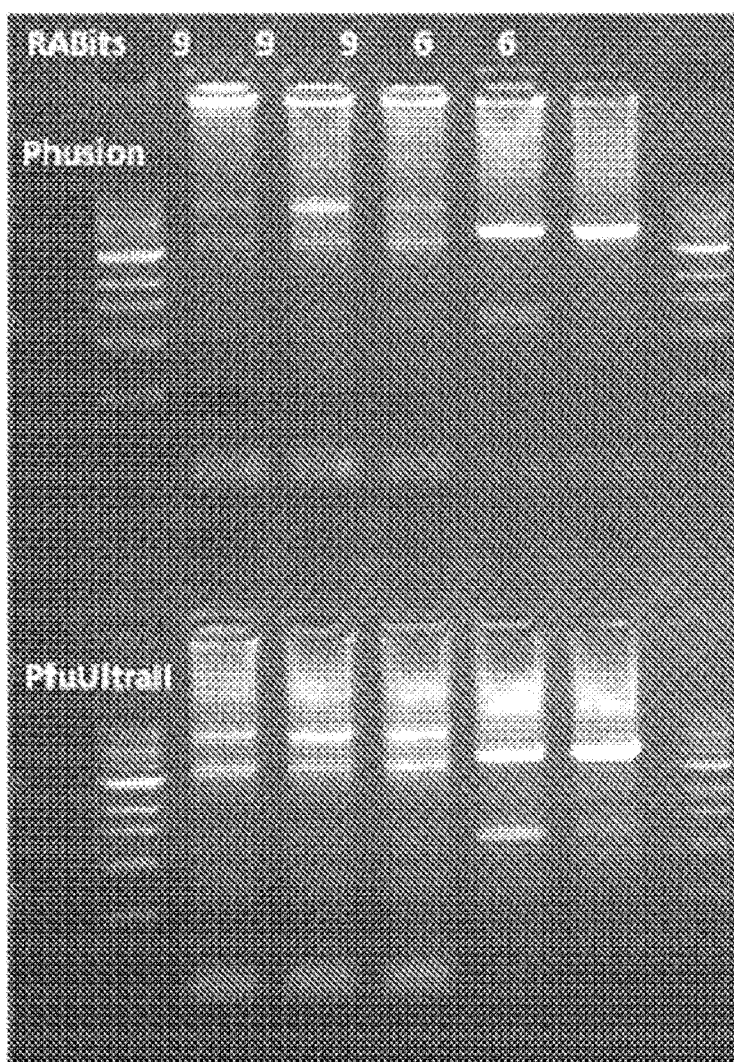


图 6

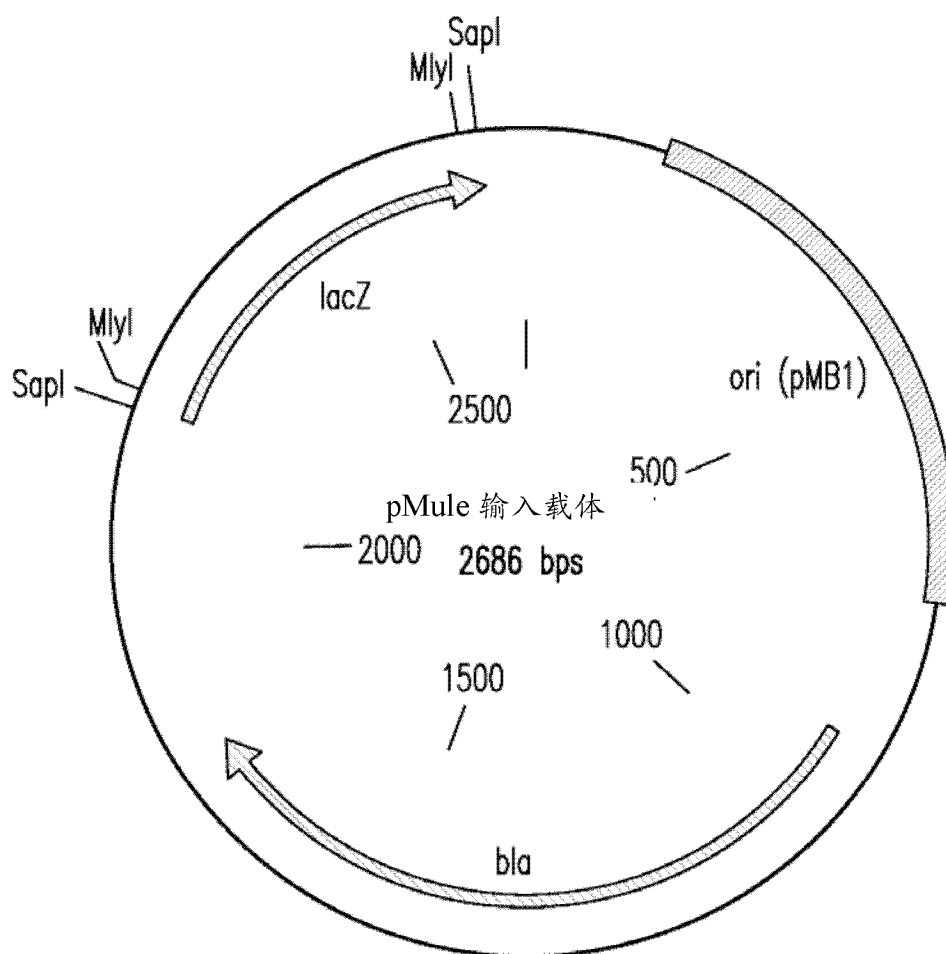


图 7

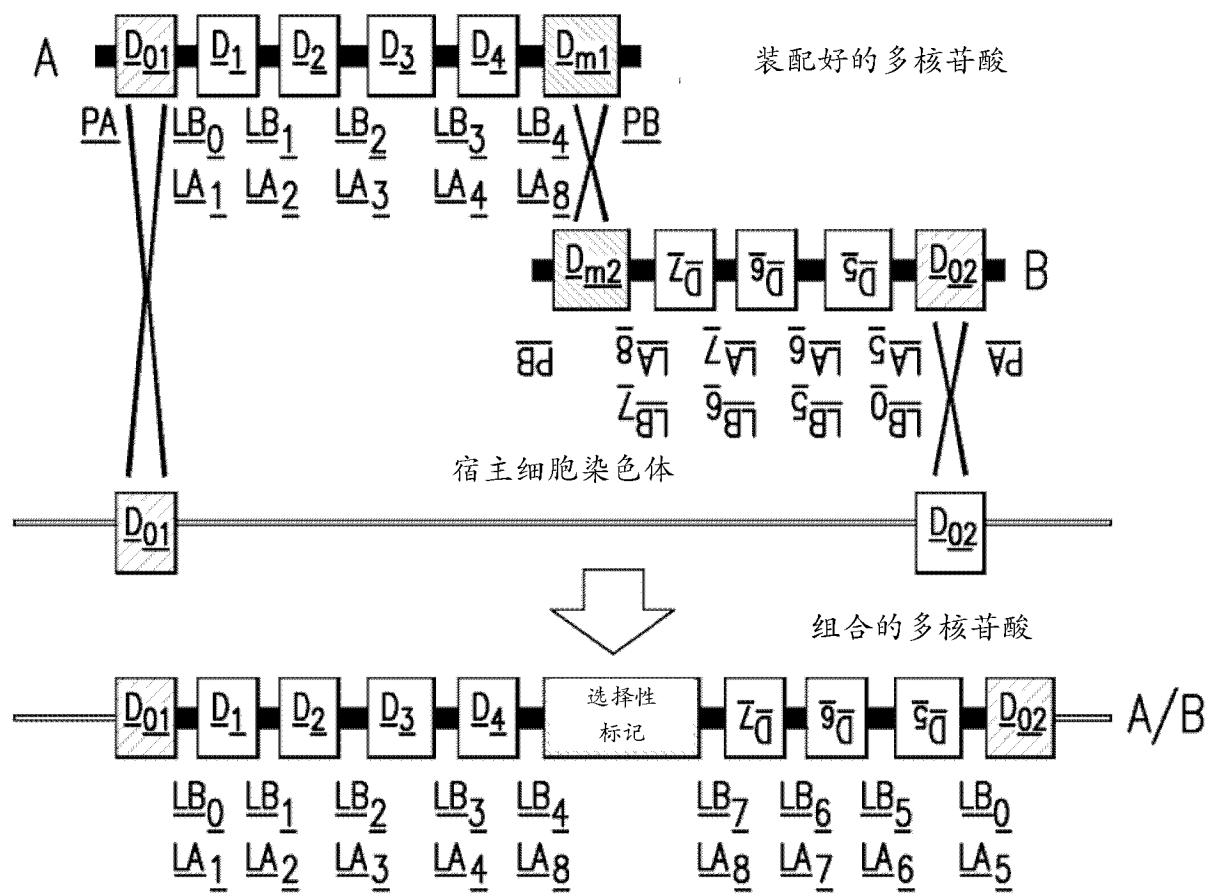


图 8

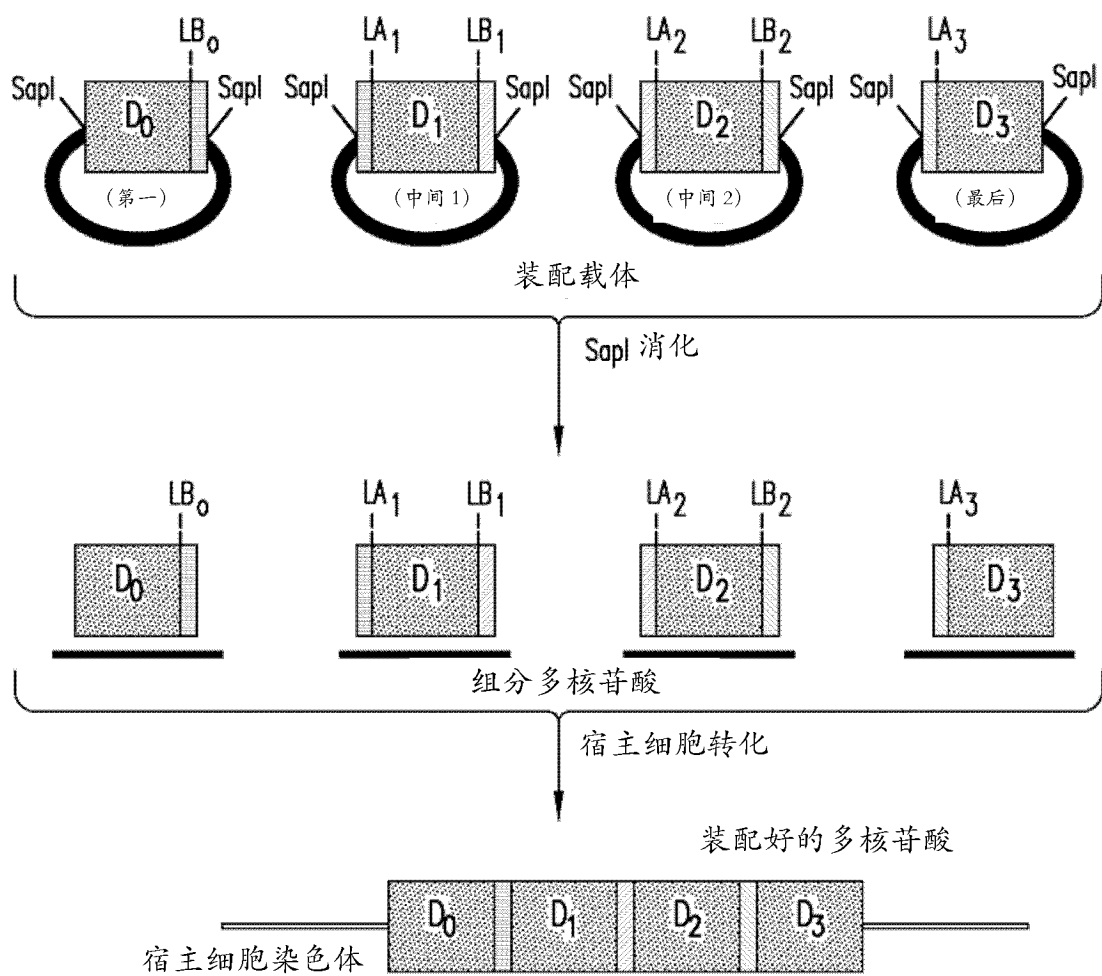


图 9

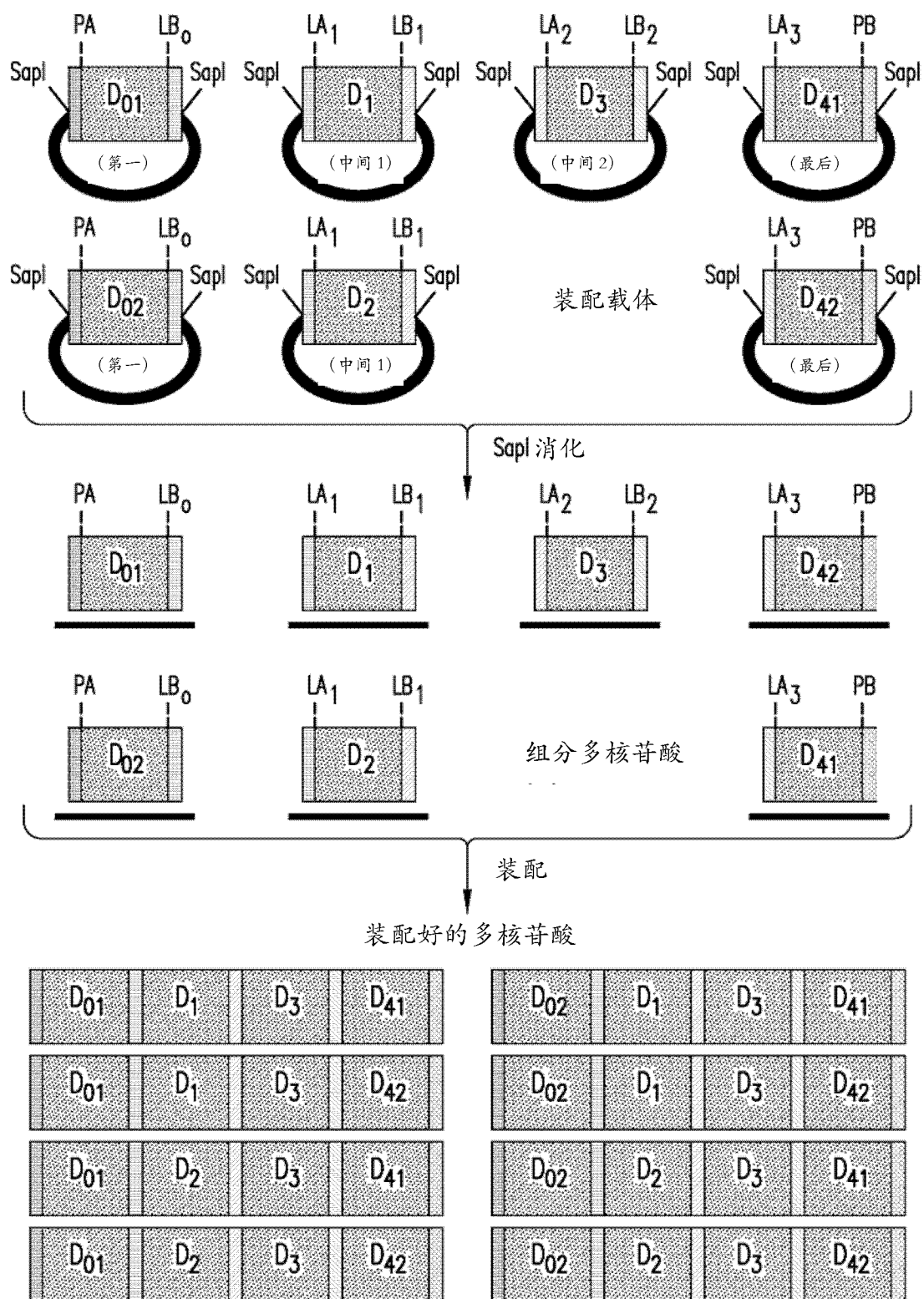


图 10

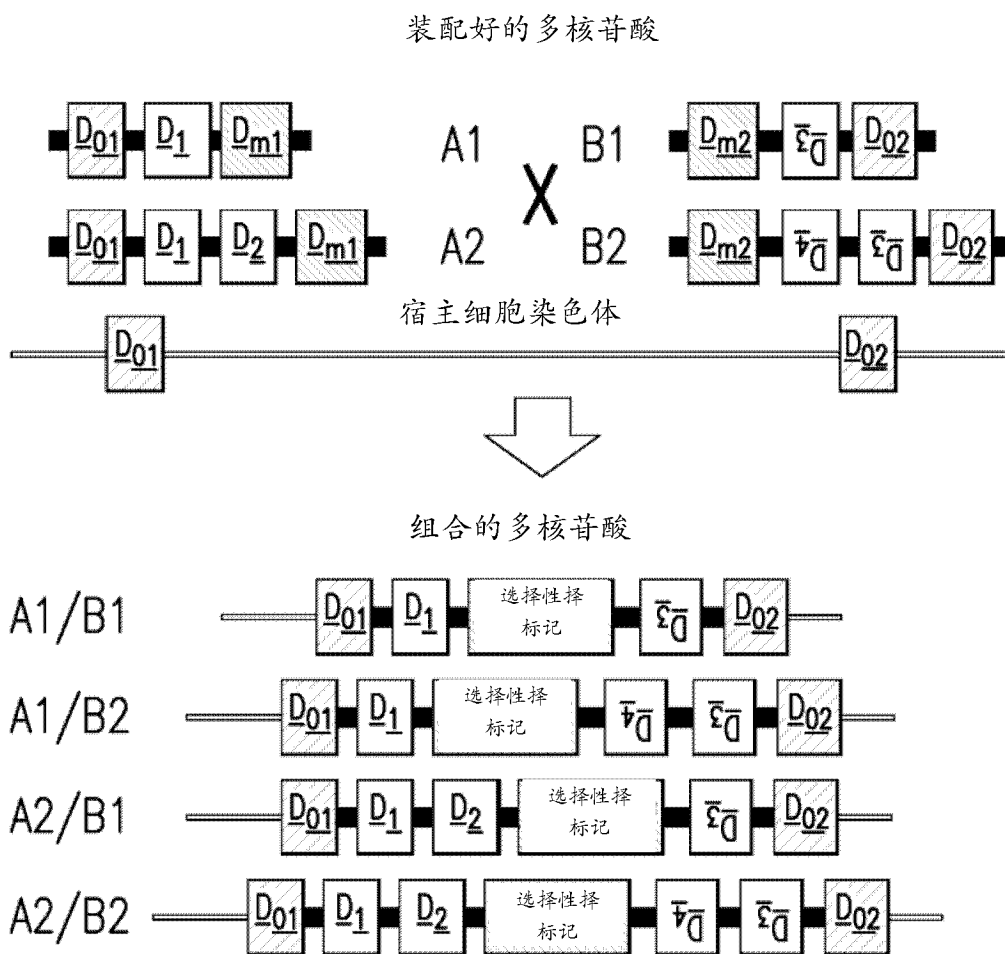


图 11

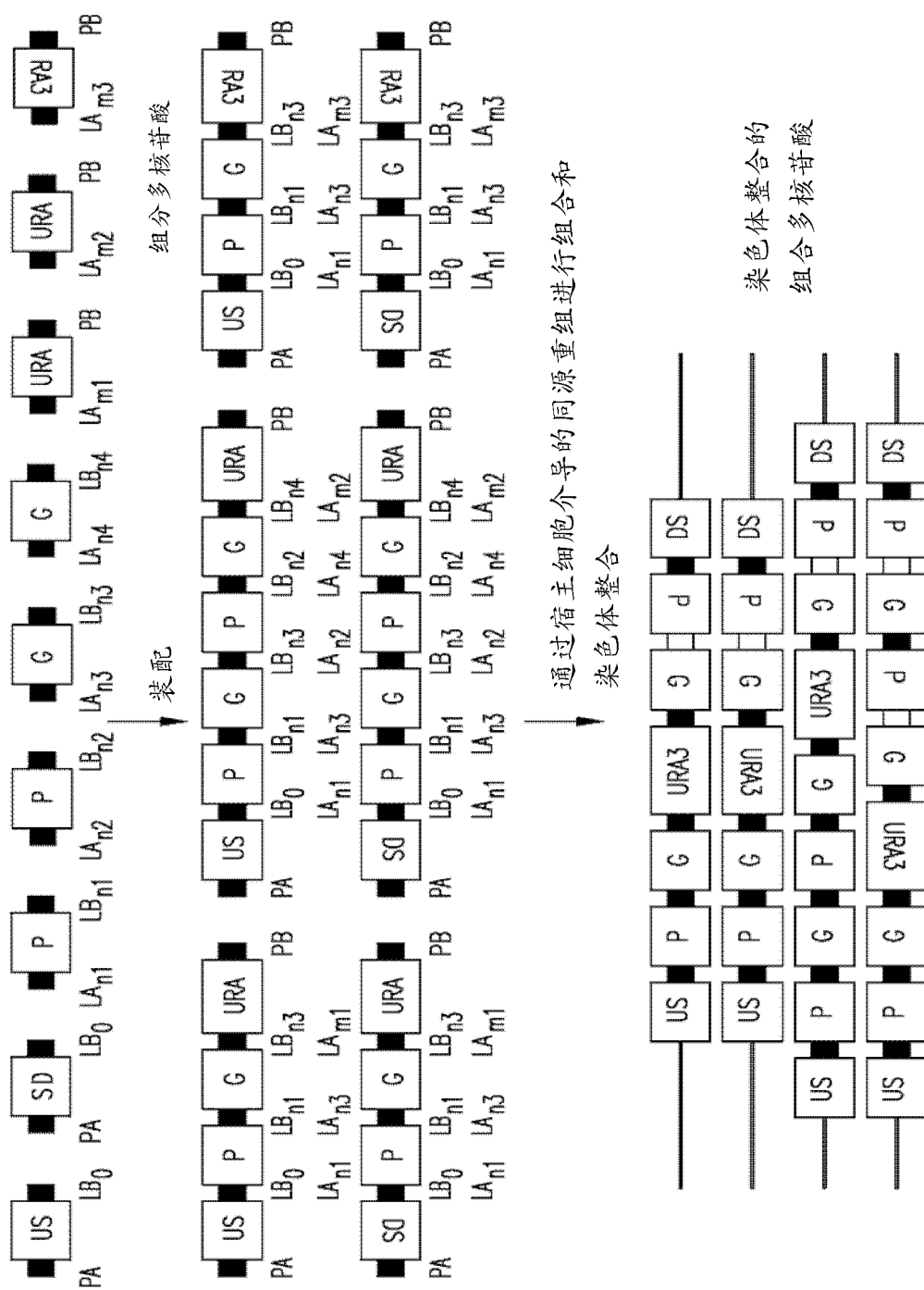


图 12A

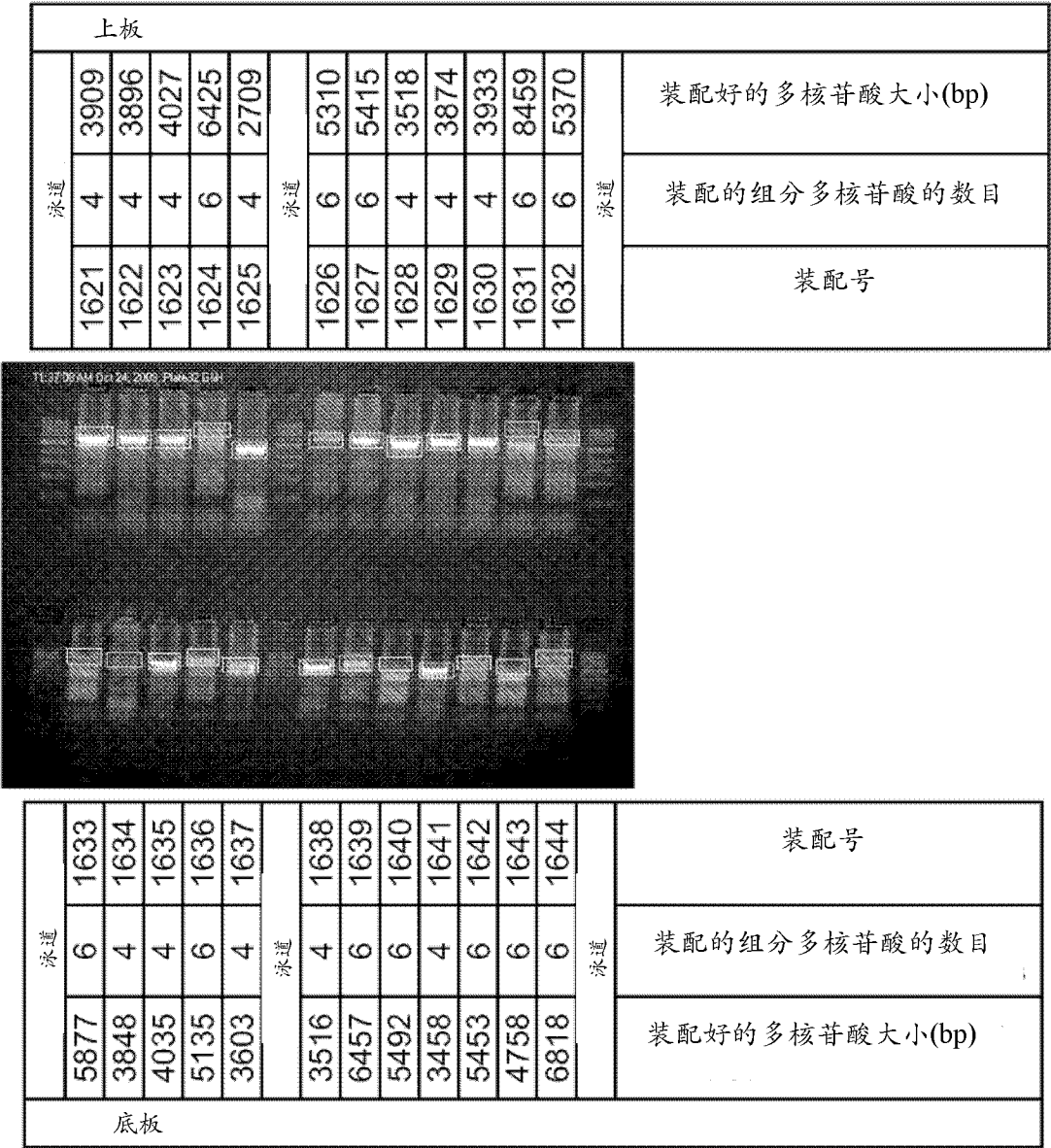


图 12B

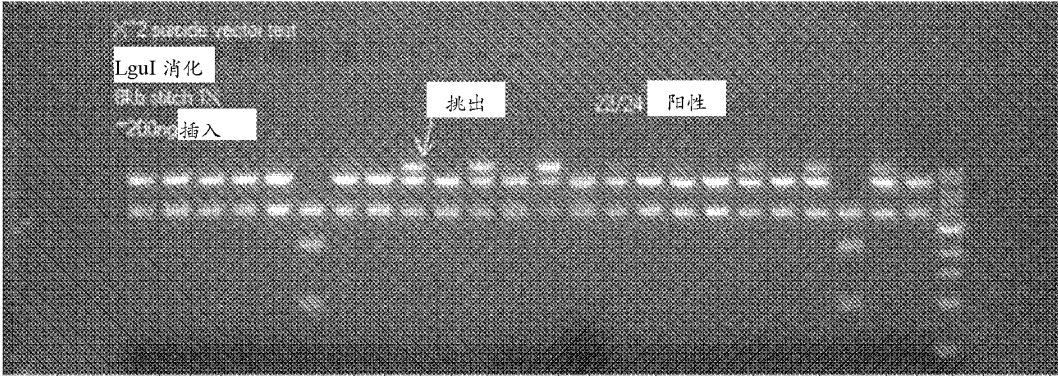


图 12C

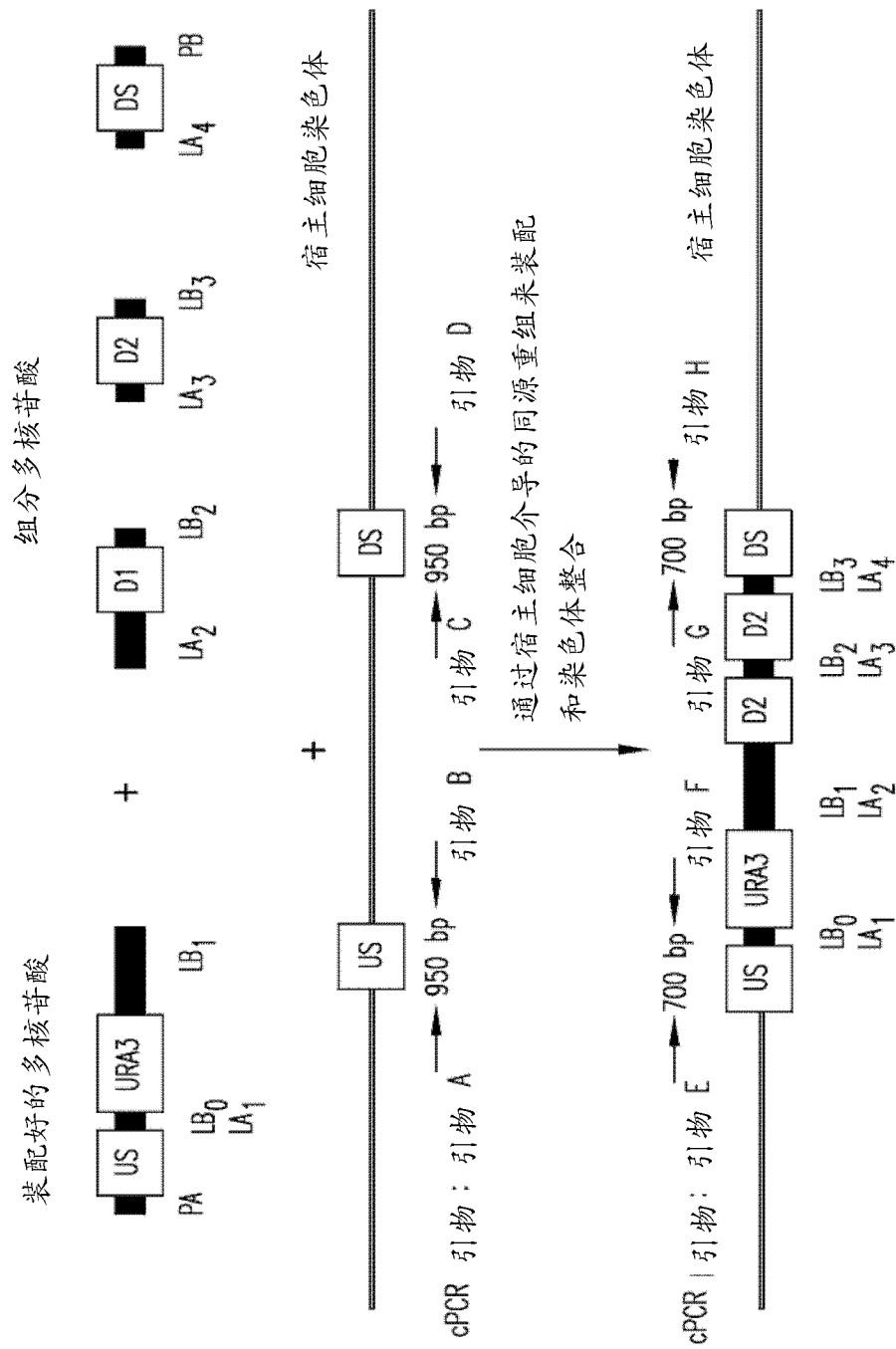


图 13A

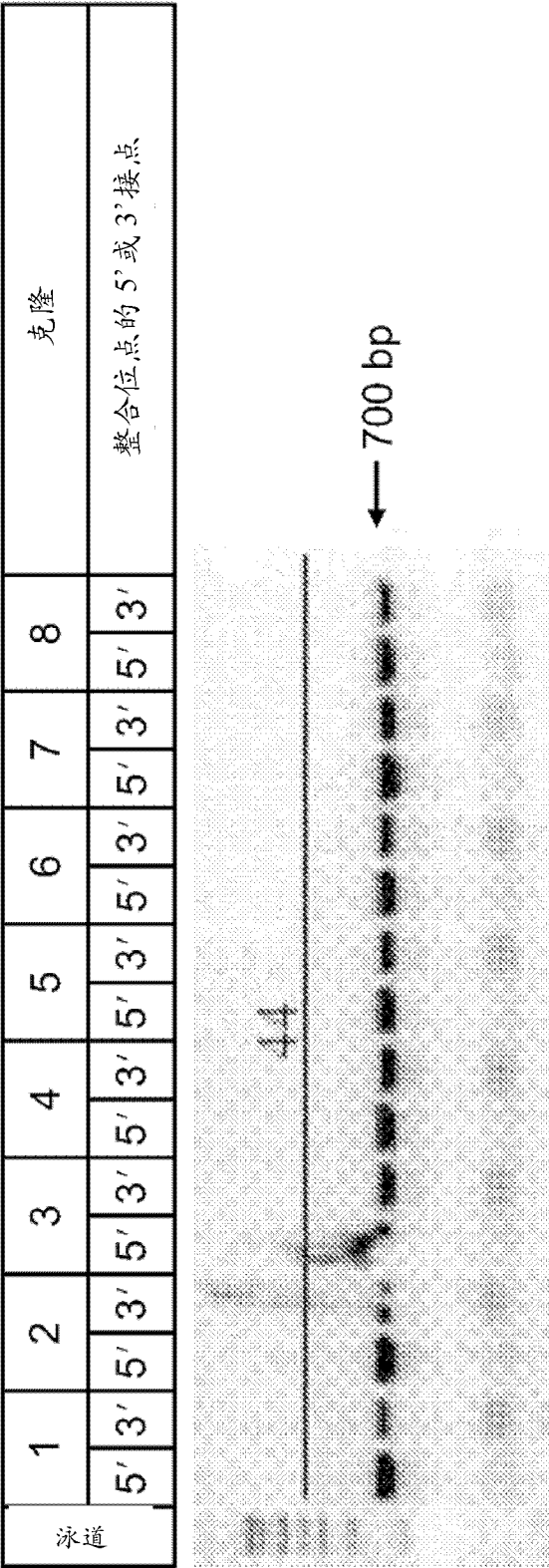


图 13B

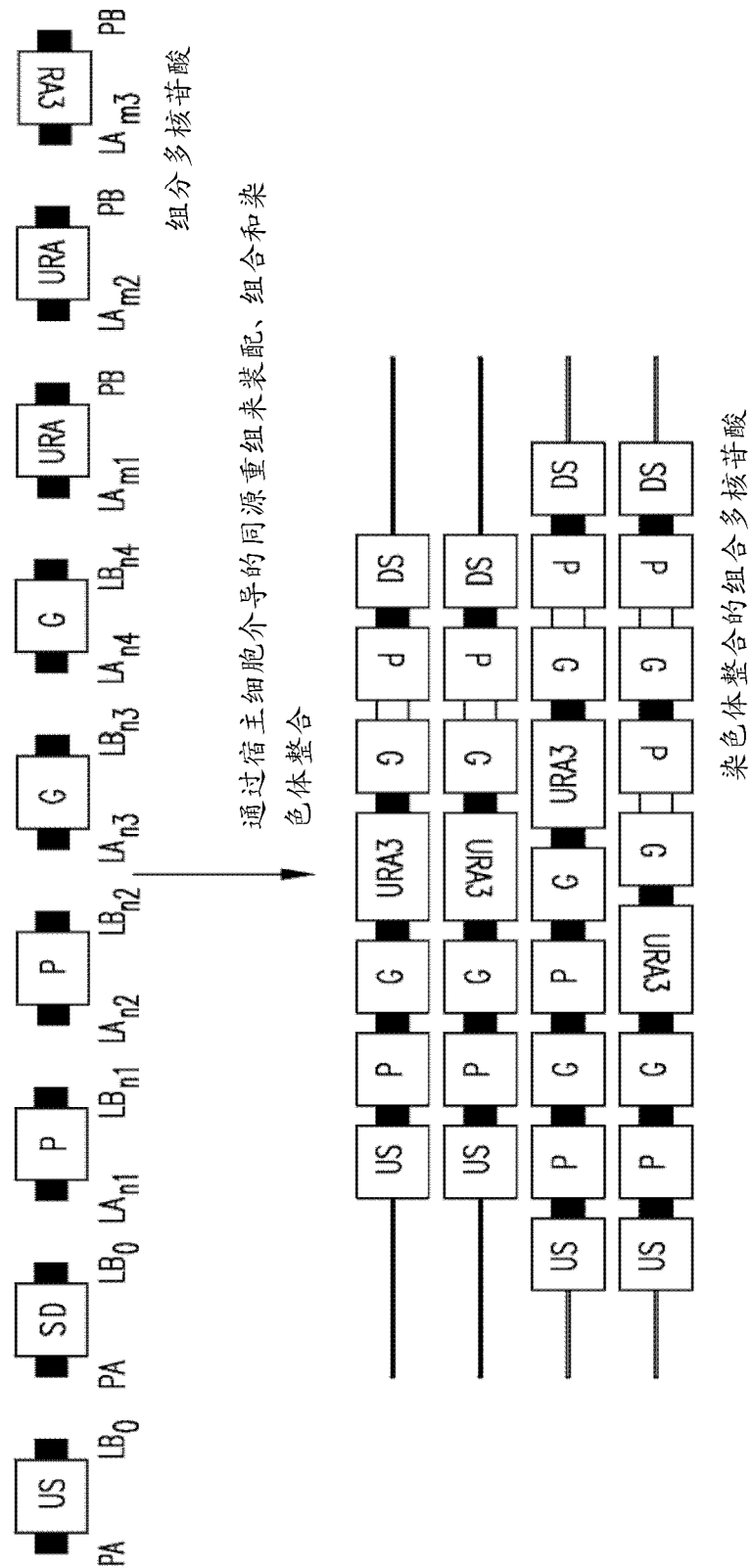


图 14