



(12)发明专利申请

(10)申请公布号 CN 108026536 A

(43)申请公布日 2018.05.11

(21)申请号 201680030913.5

(22)申请日 2016.05.27

(30)优先权数据

62/168,355 2015.05.29 US

62/296,853 2016.02.18 US

(85)PCT国际申请进入国家阶段日

2017.11.28

(86)PCT国际申请的申请数据

PCT/US2016/034812 2016.05.27

(87)PCT国际申请的公布数据

W02016/196361 EN 2016.12.08

(71)申请人 北卡罗来纳州立大学

地址 美国北卡罗来纳州

(72)发明人 R·巴伦古 K·M·泽勒

(74)专利代理机构 中国国际贸易促进委员会专利商标事务所 11038

代理人 刘晓东

(51)Int.Cl.

C12N 15/63(2006.01)

C12N 15/10(2006.01)

A01N 37/46(2006.01)

A01N 63/02(2006.01)

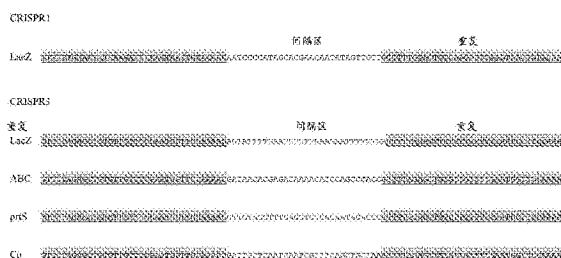
权利要求书6页 说明书66页 附图34页

(54)发明名称

使用CRISPR核酸筛选细菌、古细菌、藻类和酵母的方法

(57)摘要

本申请涉及使用CRISPR核酸来筛选细菌、古细菌、藻类和/或酵母中的必需基因和非必需基因以及可消耗基因组岛、来杀灭细菌、古细菌、藻类和/或酵母、来鉴定一种或多种基因的表型、和/或来筛选细菌、古细菌、藻类和/或酵母中的基因组大小减小和/或基因缺失。



1. 一种针对必需基因、非必需基因和/或可消耗基因组岛对细菌细胞群体进行筛选的方法,包括:

向所述细菌细胞群体中引入异源核酸构建体,所述构建体包含CRISPR阵列,所述CRISPR阵列包含(5'至3')重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述重复-间隔区-重复序列或所述至少一个重复-间隔区序列的间隔区包含与所述群体中的细菌细胞的基因组中的靶标区实质上互补的核苷酸序列,从而产生转化的细菌细胞的群体;以及

确定所述转化的细菌细胞的群体中是否存在缺失,其中所述转化的细菌细胞的群体中存在缺失表明所述靶标区包含在非必需基因和/或可消耗基因组岛内,而所述群体中不存在缺失意味着所述靶标区包含在必需基因内。

2. 一种针对必需基因、非必需基因和/或可消耗基因组岛对细菌、古细菌、藻类或酵母细胞群体进行筛选的方法,包括:

向所述细菌、古细菌、藻类或酵母细胞群体中引入(a)包含反式编码的CRISPR(tracr)核酸的异源核酸构建体,(b)异源核酸构建体,所述构建体包含CRISPR阵列,所述CRISPR阵列包含(5'至3')重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述重复-间隔区-重复序列或所述至少一个重复-间隔区序列的间隔区包含与所述群体中的细菌、古细菌、藻类或酵母细胞的基因组(染色体和/或质粒)中的靶标区实质上互补的核苷酸序列,和(c)Cas9多肽或包含编码Cas9多肽的多核苷酸的异源核酸构建体,从而产生转化的细菌、古细菌、藻类或酵母细胞的群体;以及

确定所述转化的细菌、古细菌、藻类或酵母细胞的群体中是否存在缺失,其中所述转化的细菌、古细菌、藻类或酵母细胞的群体中存在缺失意味着所述靶标区包含在非必需基因和/或可消耗基因组岛内,而所述转化的细菌、古细菌、藻类或酵母细胞的群体中不存在缺失意味着所述靶标区包含在必需基因内。

3. 一种杀灭细菌细胞群体内的一种或多种细菌细胞的方法,包括:

向所述细菌细胞群体中引入异源核酸构建体,所述构建体包含CRISPR阵列(crRNA、crDNA),所述CRISPR阵列包含(5'至3')重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述重复-间隔区-重复序列或所述至少一个重复-间隔区序列的间隔区包含与所述群体中的细菌细胞的基因组中的靶标区实质上互补的核苷酸序列,从而杀灭所述细菌细胞群体内包含所述靶标区的一种或多种细菌细胞。

4. 一种杀灭细菌、古细菌、藻类或酵母细胞群体内的一种或多种细胞的方法,包括:

向所述细菌、古细菌、藻类或酵母细胞群体中引入(a)包含反式编码的CRISPR(tracr)核酸的异源核酸构建体,(b)包含CRISPR阵列的异源核酸构建体,所述CRISPR阵列包含(5'至3')重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述重复-间隔区-重复序列或所述至少一个重复-间隔区序列的间隔区包含与所述群体中的细菌、古细菌、藻类或酵母细胞的基因组(染色体和/或质粒)中的靶标区实质上互补的核苷酸序列,和(c)Cas9多肽和/或包含编码Cas9多肽的多核苷酸的异源核酸构建体,从而杀灭细菌、古细菌、藻类或酵母细胞群体内在其基因组中包含所述靶标区的一种或多种细胞。

5. 根据权利要求3或权利要求4所述的方法,所述靶标区在必需基因或非必需基因内。

6. 一种鉴定与细菌基因相关联的表型的方法,包括:

向细菌细胞群体中引入异源核酸构建体,所述构建体包含CRISPR阵列,所述CRISPR阵列包含(5'至3')重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述至少一个重复-间隔区序列和所述重复-间隔区-重复序列的间隔区包含与所述群体中的细菌细胞的基因组中的靶标区实质上互补的核苷酸序列,其中所述靶标区包含编码多肽或功能性核酸的开放阅读框的至少一部分,从而杀灭包含所述靶标区的细胞并产生不含所述靶标区的转化细菌细胞的群体;以及

分析所述群体的表型。

7.根据权利要求6所述的方法,包括在分析之前,从所述转化细菌细胞的群体培养独立的细菌集落;以及

分析所述独立集落的表型。

8.一种鉴定细菌、古细菌、藻类或酵母基因的表型的方法,包括:

向细菌、古细菌、藻类或酵母细胞群体中引入(a)包含反式编码的CRISPR(tracr)核酸的异源核酸构建体,(b)包含CRISPR阵列的异源核酸构建体,所述CRISPR阵列包含(5'至3')重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述重复-间隔区-重复序列或所述至少一个重复-间隔区序列的间隔区包含与所述群体中的细菌、古细菌、藻类或酵母细胞的基因组(染色体和/或质粒)中的靶标区实质上互补的核苷酸序列,和(c)Cas9多肽和/或包含编码Cas9多肽的多核苷酸的异源核酸构建体,从而杀灭包含所述靶标区的细菌、古细菌或酵母细胞并产生不含所述靶标区的转化的细菌、古细菌、藻类或酵母细胞的群体;以及

分析所述转化的细菌、古细菌、藻类或酵母细胞的群体的表型,和/或(i)由所述转化的细菌、古细菌、藻类或酵母细胞的群体培养独立细菌、古细菌、藻类或酵母集落,以及(ii)分析所述独立集落的表型。

9.根据权利要求1、3、6或7中任一项所述的方法,其中所述CRISPR阵列为I型、II型、III型、IV型、V型CRISPR阵列。

10.根据权利要求9所述的方法,其中所述重复-间隔区-重复序列或所述至少一个重复-间隔区序列包含与来自野生型I型CRISPR阵列、野生型II型CRISPR阵列、野生型III型CRISPR阵列、野生型IV型CRISPR阵列或野生型V型CRISPR阵列的重复相同的重复。

11.根据权利要求2、4、5或8中任一项的方法,其中所述重复-间隔区-重复序列或所述至少一个重复-间隔区序列包含与来自野生型II型CRISPR阵列的重复相同的重复。

12.根据权利要求1至11中任一项所述的方法,其中所述靶标区是随机选择的或是特定选择的。

13.根据权利要求12所述的方法,其中随机选择的靶标区选自与细菌、古细菌或酵母基因组中的PAM序列相邻的任意至少10个连续核苷酸。

14.根据权利要求12所述的方法,其中特定选择的靶标区选自包含与细菌、古细菌、藻类或酵母基因组中的PAM序列相邻的至少10个连续核苷酸的基因、开放阅读框或推定的开放阅读框。

15.根据权利要求2、4、5、8、11至14中任一项所述的方法,其中所述异源核酸构建体包含反式编码的CRISPR(tracr)核酸并且所述包含CRISPR阵列的异源核酸构建体包含在CRISPR向导结构(gRNA、gDNA)中,所述向导结构任选还包含具有编码Cas9多肽的多核苷酸

的异源核酸构建体。

16. 根据权利要求15所述的方法, 其中所述CRISPR向导结构与启动子有效连接。

17. 一种从细菌细胞群体中选择基因组大小减小的一个或多个细菌细胞的方法, 包括:

向细菌细胞群体中引入异源核酸构建体, 所述构建体包含CRISPR阵列, 所述CRISPR阵列包含(5'至3')重复-间隔区-重复序列或至少一个重复-间隔区序列, 其中所述重复-间隔区-重复序列或所述至少一个重复-间隔区序列的间隔区包含与所述群体中的一个或多个细菌细胞的基因组中的靶标区实质上互补的核苷酸序列, 其中杀灭包含所述靶标区的细胞, 从而从所述细菌细胞群体中选择不含所述靶标区并且基因组大小减小的一种或多种细菌细胞。

18. 一种从细菌细胞群体中选择基因组大小减小的一种或多种细菌细胞的方法, 包括:
向细菌细胞群体中引入:

(a) (i) 一种或多种异源核酸构建体, 所述构建体包含与所述细菌细胞的基因组中存在的至少300个连续核苷酸具有至少80%同一性的核苷酸序列, 或者(ii) 两种或更多种异源核酸构建体, 所述构建体包含至少一个转座子, 从而产生转基因细菌细胞的群体, 所述转基因细菌细胞在整合进所述基因组中的所述一种或多种异源核酸构建体与所述基因组中存在的所述至少300个连续核苷酸之间、或在整合进所述基因组中的第一转座子与第二转座子之间包含非天然的同源重组位点; 和

(b) 异源核酸构建体, 所述构建体包含CRISPR阵列, 所述CRISPR阵列包含(5'至3')重复-间隔区-重复序列或至少一个重复-间隔区序列, 其中所述重复-间隔区-重复序列或所述至少一个重复-间隔区序列的间隔区包含与所述群体中的一种或多种细菌细胞的基因组中的靶标区实质上互补的核苷酸序列, 其中所述靶标区位于引入进所述基因组中的所述一种或多种异源核酸构建体与所述基因组中存在的所述至少300个连续核苷酸之间和/或位于所述第一转座子与所述第二转座子之间, 以及杀灭包含所述靶标区的细胞, 从而从所述转基因细菌细胞群体中选择不含所述靶标区并且基因组大小减小的一种或多种细菌细胞。

19. 一种从细菌、古细菌、藻类或酵母细胞群体中选择基因组大小减小的一种或多种细菌、古细菌、藻类或酵母细胞的方法, 包括:

向细菌、古细菌、藻类或酵母细胞群体中引入:

(a) 包含反式编码的CRISPR(tracr)核酸的异源核酸构建体,

(b) 异源核酸构建体, 所述构建体包含CRISPR阵列, 所述CRISPR阵列包含重复-间隔区-重复序列或至少一个重复-间隔区序列, 其中所述至少一个重复-间隔区序列和所述至少一个重复-间隔区-重复序列的间隔区包含与所述群体中的细菌、古细菌、藻类或酵母细胞的基因组(染色体和/或质粒)中的靶标区实质上互补的核苷酸序列, 和

(c) Cas9多肽和/或包含编码Cas9多肽的多核苷酸的异源核酸构建体, 其中杀灭包含所述靶标区的细胞, 从而从所述细菌、古细菌、藻类或酵母细胞群体中选择不含所述靶标区并且基因组大小减小的一种或多种细菌、古细菌、藻类或酵母细胞。

20. 一种从细菌、古细菌或酵母细胞群体中选择基因组大小减小的一种或多种细菌、古细菌、藻类或酵母细胞的方法, 包括:

向细菌、古细菌、藻类或酵母细胞群体中引入:

(a) (i) 一种或多种异源核酸构建体, 所述构建体包含与所述细菌、古细菌、藻类或酵母

细胞的基因组中存在的至少300个连续核苷酸具有至少80%同一性的核苷酸序列,或者 (i i) 两种或更多种异源核酸构建体,所述构建体包含至少一个转座子,从而产生转基因细菌、古细菌、藻类或酵母细胞的群体,所述转基因细菌、古细菌、藻类或酵母细胞在整合进所述基因组中的所述一种或多种异源核酸构建体与所述基因组中存在的所述至少300个连续核苷酸之间、或在整合进所述基因组中的第一转座子与第二转座子之间包含非天然的同源重组位点;和

(b) (i) 包含反式编码的CRISPR (tracr) 核酸的异源核酸构建体, (ii) 异源核酸构建体,所述构建体包含CRISPR阵列,所述CRISPR阵列包含重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述至少一个重复-间隔区序列和所述至少一个重复-间隔区-重复序列的间隔区包含与所述群体中的一种或多种细菌、古细菌、藻类或酵母细胞的基因组(染色体和/或质粒)中的靶标区实质上互补的核苷酸序列,和 (i i i) Cas9多肽和/或包含编码Cas9多肽的多核苷酸的异源核酸构建体,其中所述靶标区位于掺入进所述基因组中的所述一种或多种异源核酸构建体与所述基因组中存在的所述至少300个连续核苷酸之间和/或所述第一转座子与所述第二转座子之间,以及杀灭包含所述靶标区的细胞,从而从所述转基因细菌、古细菌、藻类或酵母细胞群体中选择不含所述靶标区并且基因组大小减小的一种或多种细菌、古细菌、藻类或酵母细胞。

21. 一种在细菌群体中鉴定在其基因组中具有缺失的至少一种分离株的方法,包括:

向细菌细胞群体中引入异源核酸构建体,所述构建体包含CRISPR阵列,所述CRISPR阵列包含(5'至3')重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述重复-间隔区-重复序列或所述至少一个重复-间隔区序列的间隔区包含与所述群体中的一种或多种细菌细胞的基因组中的靶标区实质上互补的核苷酸序列,其中杀灭包含所述靶标区的细胞,从而产生不含所述靶标区的转化的细菌细胞的群体;以及

从所述转化的细菌细胞的群体培养独立的细菌集落,从而从所述转化的细菌的群体中鉴定在其基因组中具有缺失的至少一种分离株。

22. 一种在细菌群体中鉴定在其基因组中具有缺失的至少一种分离株的方法,包括:

向所述细菌细胞群体中引入:

(a) (i) 一种或多种异源核酸构建体,所述构建体包含与所述细菌细胞的基因组中存在的至少300个连续核苷酸具有至少80%同一性的核苷酸序列,或者 (i i) 两种或更多种异源核酸构建体,所述构建体包含至少一个转座子,从而产生转基因细菌细胞的群体,所述转基因细菌细胞在整合进所述基因组中的所述一种或多种异源核酸构建体与所述基因组中存在的所述至少300个连续核苷酸之间、或整合进所述基因组中的第一转座子与第二转座子之间包含非天然的同源重组位点;和

(b) 异源核酸构建体,所述构建体包含CRISPR阵列,所述CRISPR阵列包含(5'至3')重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述重复-间隔区-重复序列或所述至少一个重复-间隔区序列的间隔区包含与所述群体中的一种或多种细菌细胞的基因组中的靶标区实质上互补的核苷酸序列,其中所述靶标区位于引入进所述基因组中的所述一种或多种异源核酸构建体与所述基因组中存在的所述至少300个连续核苷酸之间和/或所述第一转座子与所述第二转座子之间,以及杀灭包含所述靶标区的细胞,从而产生不含所述靶标区的转化细菌细胞的群体;以及

从所述转化细菌细胞的群体培养独立的细菌集落,从而从所述细菌的群体中鉴定在其基因组中具有缺失的至少一种分离株。

23. 一种在细菌、古细菌、藻类或酵母细胞的群体中鉴定在其基因组中具有缺失的至少一种分离株的方法,包括:

向细菌、古细菌、藻类或酵母细胞群体中引入:

(a) 包含反式编码的CRISPR (tracr) 核酸的异源核酸构建体,

(b) 异源核酸构建体,所述构建体包含CRISPR阵列,所述CRISPR阵列包含(5'至3')重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述重复-间隔区-重复序列或所述至少一个重复-间隔区序列的间隔区包含与所述群体中的细菌、古细菌、藻类或酵母细胞的基因组(染色体和/或质粒)中的靶标区实质上互补的核苷酸序列,和

(c) Cas9多肽或包含编码Cas9多肽的多核苷酸的异源核酸构建体,其中杀灭包含所述靶标区的细胞,从而产生不含所述靶标区的转化细菌、古细菌、藻类或酵母细胞的群体;以及

从所述转化细菌、古细菌、藻类或酵母细胞的群体培养独立细菌、古细菌或酵母集落,从而从所述转化细菌、古细菌、藻类或酵母细胞的群体中鉴定在其基因组中具有缺失的至少一种分离株。

24. 一种在细菌、古细菌、藻类或酵母细胞的群体中鉴定在其基因组中具有缺失的至少一种分离株的方法,包括:

向所述细菌、古细菌、藻类或酵母细胞群体中引入:

(a) (i) 一种或多种异源核酸构建体,所述构建体包含与所述细菌、古细菌、藻类或酵母细胞的基因组中存在的至少300个连续核苷酸具有至少80%同一性的核苷酸序列,或者(ii) 两种或更多种异源核酸构建体,所述构建体包含至少一个转座子,从而产生转基因细菌、古细菌、藻类或酵母细胞的群体,所述转基因细菌、古细菌、藻类或酵母细胞在整合进所述基因组中的所述一种或多种异源核酸构建体与所述基因组中存在的所述至少300个连续核苷酸之间、或整合进所述基因组中的第一转座子与第二转座子之间包含非天然的同源重组位点;和

(b) (i) 包含反式编码的CRISPR (tracr) 核酸的异源核酸构建体, (ii) 异源核酸构建体,所述构建体包含CRISPR阵列,所述CRISPR阵列包含(5'至3')重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述重复-间隔区-重复序列或所述至少一个重复-间隔区序列的间隔区包含与所述群体中的一种或多种细菌、古细菌、藻类或酵母细胞的基因组(染色体和/或质粒)中的靶标区实质上互补的核苷酸序列,和(iii) Cas9多肽和/或包含编码Cas9多肽的多核苷酸的异源核酸构建体,其中所述靶标区位于掺入进所述基因组中的所述一种或多种异源核酸构建体与所述基因组中存在的所述至少300个连续核苷酸之间和/或所述第一转座子与所述第二转座子之间,以及杀灭包含所述靶标区的细胞,从而产生不含所述靶标区的转化细菌、古细菌、藻类或酵母细胞的群体;并且

从所述转化细菌、古细菌、藻类或酵母细胞的群体培养独立的细菌、古细菌或酵母集落,从而从所述群体中鉴定在其基因组中具有缺失的至少一种分离株。

25. 根据权利要求17、18、21或22中任一项所述的方法,其中所述CRISPR阵列为I型、II型、III型、IV型、V型CRISPR阵列。

26. 根据权利要求25所述的方法,其中所述重复-间隔区-重复序列或所述至少一个重复-间隔区序列包含与来自野生型I型CRISPR阵列、野生型II型CRISPR阵列、野生型III型CRISPR阵列、野生型IV型CRISPR阵列或野生型V型CRISPR阵列的重复相同的重复。

27. 根据权利要求19、20、23或24中任一项的方法,其中所述重复-间隔区-重复序列或所述至少一个重复-间隔区序列包含与来自野生型II型CRISPR阵列的重复相同的重复。

28. 根据权利要求15至27中任一项所述的方法,其中所述靶标区是随机选择的或是特定选择的。

29. 根据权利要求28所述的方法,其中随机选择的靶标区选自与细菌、古细菌或酵母基因组中的PAM序列相邻的任意至少10个连续核苷酸。

30. 根据权利要求28所述的方法,其中特定选择的靶标区选自包含与细菌、古细菌、藻类或酵母基因组中的PAM序列相邻的至少10个连续核苷酸的基因、开放阅读框或推定的开放阅读框。

31. 根据权利要求3至5或9至15中任一项所述的方法,其中所述引入的CRISPR阵列与待杀灭的所述一种或多种细菌细胞中的CRISPR-Cas系统相容,但与所述细菌细胞群体的至少一种或多种细菌细胞的CRISPR Cas系统不相容。

使用CRISPR核酸筛选细菌、古细菌、藻类和酵母的方法

[0001] 优先权声明

[0002] 本申请根据35 U.S.C. §119 (e) 要求于2015年5月29日提交的美国临时专利申请No. 62/168,355和于2016年2月18日提交的美国临时专利申请No. 62/296,853的权益,将这两个专利申请每一者的全部内容以引用的方式并入本文。

技术领域

[0003] 本申请涉及使用CRISPR核酸来筛选细菌、古细菌、藻类和/或酵母中的必需基因和非必需基因以及可消耗基因组岛、来杀灭细菌、古细菌、藻类和/或酵母、来鉴定一种或多种基因的表型、和/或来筛选细菌、古细菌、藻类和/或酵母中的基因组大小减小和/或基因缺失。

背景技术

[0004] 成簇的规律间隔的短回文重复 (Clustered Regularly Interspaced Short Palindromic Repeats, CRISPR) 与相关的序列 (cas) 一起构成了CRISPR-Cas系统,该系统在许多细菌中赋予获得性免疫力。CRISPR介导的免疫通过吸收来自侵入性遗传元件诸如质粒和噬菌体的DNA作为新“间隔区”来形成。

[0005] CRISPR-Cas系统由被高变序列间隔开的短DNA重复的阵列组成,阵列侧接cas基因,该系统通过序列特异性的靶向和干扰,提供了对抗侵入性遗传元件的获得性免疫力 (Barrangou等人,2007.Science.315:1709-1712;Brouns等人,2008.Science 321:960-4; Horvath和Barrangou.2010.Science.327:167-70;Marraffini和Sontheimer.2008.Science.322:1843-1845;Bhaya等人,2011.Annu.Rev.Genet.45:273-297;Terns和Terns.2011.Curr.Opin.Microbiol.14:321-327;Westra等人,2012.Annu.Rev.Genet.46:311-339;Barrangou R.2013.RNA.4:267-278)。通常,侵入性DNA序列作为新的“间隔区”获得 (Barrangou等人,2007.Science.315:1709-1712),每个这种新“间隔区”与CRISPR重复配对并作为新的重复-间隔区单元插入CRISPR基因座中。随后,该重复-间隔区阵列转录为长的前CRISPR RNA (pre-crRNA) (Brouns等人,2008.Science 321:960-4),该前CRISPR RNA被加工成小的干扰性CRISPR RNA (crRNA),其驱动序列特异性识别。具体而言,crRNA将核酸酶引向互补的靶标以进行由Cas核酸内切酶介导的序列特异性核酸切割 (Garneau等人,2010.Nature.468:67-71;Haurwitz等人,2010.Science.329:1355-1358;Saprunauskas等人,2011.核酸Res.39:9275-9282;Jinek等人,2012.Science.337:816-821;Gasiunas等人,2012.Proc.Natl.Acad.Sci.109:E2579-E2586;Magadan等人,2012.PLoS One.7:e40913;Karvelis等人,2013.RNA Biol.10:841-851)。这种普遍的系统存在于几乎一半的细菌 (~46%) 和绝大部分古细菌 (~90%) 中。

[0006] 概括地说,存在两大类 (Makarova等人,Nat Rev Microbiol.13:722-736 (2015)) 的CRISPR-Cas系统,根据cas基因含量、cas操纵子结构、Cas蛋白序列以及加工步骤,这两种主要类别涵盖五种主要类型和16种不同的亚型 (Makarova等人,Biol Direct.6:

38 (2011); Makarova 和 Koonin *Methods Mol Biol.* 1311:47-75 (2015); Barrangou, R. *基因组 Biology* 16:247 (2015)。在 I 型和 III 型中, 专门的 Cas 内切核酸酶加工前 crRNA, 其然后组装成能够识别并切割与该 crRNA 互补的核酸的大的多 Cas 蛋白质复合体。I 型系统是最常见和普遍的系统, 其以级联驱动和 PAM 依赖的方式靶向 DNA, 从而通过利用标志蛋白 (signature protein) Cas3 破坏靶标核酸。在 II 型 CRISPR-Cas 系统中涉及不同的过程。这里, 通过其中反式激活 crRNA (tracrRNA) 与 crRNA 的重复区杂交的机制对前 CRNA 进行加工。RNA 酶 III 切割该杂交的 crRNA-tracrRNA, 接下来的第二个事件是移除每个间隔区的 5' 端, 产生保持与 tracrRNA 和 Cas9 结合的成熟 crRNA。该成熟复合体然后定位与该复合体中的间隔区序列互补的靶标 dsDNA 序列 (“原间隔区 (protospacer)” 序列) 并切开两条链。II 型系统中的复合体进行的靶标识别和切割不仅需要 crRNA-tracrRNA 复合体上的间隔区序列与靶标 “原间隔区” 序列之间互补的序列, 还需要位于原间隔区序列 3' 端的原间隔区相邻基序 (protospacer adjacent motif, PAM) 序列。

发明内容

[0007] 本发明的一个方面提供了一种针对必需基因、非必需基因和/或可消耗基因组岛对细菌细胞群体进行筛选的方法, 包括: 向所述细菌细胞群体中引入异源核酸构建体, 所述构建体包含 CRISPR 阵列, 所述 CRISPR 阵列包含 (5' 至 3') 重复-间隔区-重复序列或至少一个重复-间隔区序列, 其中所述重复-间隔区-重复序列或所述至少一个重复-间隔区序列的间隔区包含与所述群体中的细菌细胞的基因组中的靶标区实质上互补的核苷酸序列, 从而产生转化的细菌细胞的群体; 确定所述转化的细菌、古细菌、藻类或酵母细胞的群体中是否存在缺失, 其中所述转化的细菌、古细菌或酵母细胞的群体中存在缺失意味着所述靶标区包含在非必需基因和/或可消耗基因组岛内, 而所述转化的细菌、古细菌、藻类或酵母细胞的群体中不存在缺失意味着所述靶标区包含在必需基因内。可用于本发明的 CRISPR 阵列可以是 I 型、II 型、III 型、IV 型或 V 型 CRISPR 阵列。

[0008] 本发明的第二个方面提供了一种针对必需基因、非必需基因和/或可消耗基因组岛对细菌、古细菌、藻类或酵母细胞群体进行筛选的方法, 包括: 向所述细菌、古细菌、藻类或酵母细胞群体中引入 (a) 包含反式编码的 CRISPR (tracr) 核酸的异源核酸构建体, (b) 异源核酸构建体, 所述构建体包含 CRISPR 阵列, 所述 CRISPR 阵列包含 (5' 至 3') 重复-间隔区-重复序列或至少一个重复-间隔区序列, 其中所述重复-间隔区-重复序列或所述至少一个重复-间隔区序列的间隔区包含与所述群体中的细菌、古细菌、藻类或酵母细胞的基因组 (染色体和/或质粒) 中的靶标区实质上互补的核苷酸序列, 和 (c) Cas9 多肽或包含编码 Cas9 多肽的多核苷酸的异源核酸构建体, 从而产生转化的细菌、古细菌、藻类或酵母细胞的群体; 以及确定所述转化的细菌、古细菌、藻类或酵母细胞的群体中是否存在缺失, 其中所述转化的细菌、古细菌或酵母细胞的群体中存在缺失意味着所述靶标区包含在非必需基因和/或可消耗基因组岛内, 而所述转化的细菌、古细菌、藻类或酵母细胞的群体中不存在缺失意味着所述靶标区包含在必需基因内。

[0009] 本发明的第三个方面提供了一种杀灭细菌细胞群体内的一种或多种细菌细胞的方法, 包括: 向所述细菌细胞群体中引入异源核酸构建体, 所述构建体包含 CRISPR 阵列 (crRNA、crDNA), 所述 CRISPR 阵列包含 (5' 至 3') 重复-间隔区-重复序列或至少一个重复-间

隔区序列,其中所述重复-间隔区-重复序列或所述至少一个重复-间隔区序列的间隔区包含与所述群体中的细菌细胞的基因组中的靶标区实质上互补的核苷酸序列,从而杀灭所述细菌细胞群体内包含所述靶标区的一种或多种细菌细胞。可用于本发明的CRISPR阵列可以是I型、II型、III型、IV型或V型CRISPR阵列。

[0010] 本发明的第四个方面提供了一种杀灭细菌、古细菌、藻类或酵母细胞群体内的一种或多种细胞的方法,包括:向所述细菌、古细菌、藻类或酵母细胞群体中引入(a)包含反式编码的CRISPR(tracr)核酸的异源核酸构建体,(b)异源核酸构建体,所述构建体包含CRISPR阵列,所述CRISPR阵列包含(5'至3')重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述重复-间隔区-重复序列或所述至少一个重复-间隔区序列的间隔区包含与所述群体中的细菌、古细菌、藻类或酵母细胞的基因组(染色体和/或质粒)中的靶标区实质上互补的核苷酸序列,和(c)Cas9多肽和/或包含编码Cas9多肽的多核苷酸的异源核酸构建体,从而杀灭细菌、古细菌、藻类或酵母细胞群体内在其基因组中包含所述靶标区的一种或多种细胞。

[0011] 本发明的第五个方面提供了一种鉴定与细菌基因相关联的表型的方法,包括:向细菌细胞群体中引入异源核酸构建体,所述构建体包含CRISPR阵列,所述CRISPR阵列包含(5'至3')重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述至少一个重复-间隔区序列和所述重复-间隔区-重复序列的间隔区包含与所述群体中的细菌细胞的基因组中的靶标区实质上互补的核苷酸序列,其中所述靶标区包含编码多肽或功能性核酸的开放阅读框的至少一部分,从而杀灭包含所述靶标区的细胞并产生不含所述靶标区的转化细菌细胞的群体;以及分析所述群体的表型。可用于本发明的CRISPR阵列可以是I型、II型、III型、IV型或V型CRISPR阵列。

[0012] 本发明的第六个方面提供了一种鉴定细菌、古细菌、藻类或酵母基因的表型的方法,包括:向细菌、古细菌、藻类或酵母细胞群体中引入(a)包含反式编码的CRISPR(tracr)核酸的异源核酸构建体,(b)异源核酸构建体,所述构建体包含CRISPR阵列,所述CRISPR阵列包含(5'至3')重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述重复-间隔区-重复序列或所述至少一个重复-间隔区序列的间隔区包含与所述群体中的细菌、古细菌、藻类或酵母细胞的基因组(染色体和/或质粒)中的靶标区实质上互补的核苷酸序列,和(c)Cas9多肽和/或包含编码Cas9多肽的多核苷酸的异源核酸构建体,从而杀灭包含所述靶标区的细菌、古细菌或酵母细胞并产生不含所述靶标区的转化的细菌、古细菌、藻类或酵母细胞的群体;以及分析转化的细菌、古细菌、藻类或酵母细胞的群体的表型,和/或(i)从所述转化的细菌、古细菌、藻类或酵母细胞的群体培养独立的细菌、古细菌、藻类或酵母集落,以及(ii)分析所述独立集落的表型。

[0013] 本发明的第七个方面提供了一种从细菌细胞群体中选择基因组大小减小的一种或多种细菌细胞的方法,包括:向细菌细胞群体中引入异源核酸构建体,所述构建体包含CRISPR阵列,所述CRISPR阵列包含(5'至3')重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述重复-间隔区-重复序列或所述至少一个重复-间隔区序列的间隔区包含与所述群体中的一种或多种细菌细胞的基因组中的靶标区实质上互补的核苷酸序列,其中杀灭包含所述靶标区的细胞,从而从所述细菌细胞群体中选择不含所述靶标区并且基因组大小减小的一种或多种细菌细胞。可用于本发明的CRISPR阵列可以是I型、II型、III型、IV

型或V型CRISPR阵列。

[0014] 本发明的第八个方面提供了一种从细菌细胞群体中选择基因组大小减小的一种或多种细菌细胞的方法,包括:向细菌细胞群体中引入(a) (i) 一种或多种异源核酸构建体,所述构建体包含与所述细菌细胞的基因组中存在的至少300个连续核苷酸具有至少80%同一性的核苷酸序列,或者(ii) 两种或更多种异源核酸构建体,所述构建体包含至少一个转座子,从而产生转基因细菌细胞的群体,所述转基因细菌细胞在整合进所述基因组中的所述一种或多种异源核酸构建体与所述基因组中存在的所述至少300个连续核苷酸之间、或在整合进所述基因组中的第一转座子与第二转座子之间包含非天然的同源重组位点;和(b) 异源核酸构建体,所述构建体包含CRISPR阵列,所述CRISPR阵列包含(5'至3') 重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述重复-间隔区-重复序列或所述至少一个重复-间隔区序列的间隔区包含与所述群体中的一种或多种细菌细胞的基因组中的靶标区实质上互补的核苷酸序列,其其中所述靶标区位于引入进所述基因组中的所述一种或多种异源核酸构建体与所述基因组中存在的所述至少300个连续核苷酸之间和/或所述第一转座子与所述第二转座子之间,以及杀灭包含所述靶标区的细胞,从而从所述转基因细菌细胞群体中选择不含所述靶标区并且基因组大小减小的一种或多种细菌细胞。可用于本发明的CRISPR阵列可以是I型、II型、III型、IV型或V型CRISPR阵列。

[0015] 本发明的第九个方面提供了一种从细菌、古细菌、藻类或酵母细胞群体中选择基因组大小减小的一种或多种细菌、古细菌、藻类或酵母细胞的方法,包括:向细菌、古细菌、藻类或酵母细胞群体中引入(a) 包含反式编码的CRISPR (tracr) 核酸的异源核酸构建体, (b) 异源核酸构建体,所述构建体包含CRISPR阵列,所述CRISPR阵列包含重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述至少一个重复-间隔区序列和所述至少一个重复-间隔区-重复序列的间隔区包含与所述群体中的细菌、古细菌、藻类或酵母细胞的基因组(染色体和/或质粒)中的靶标区实质上互补的核苷酸序列,和(c) Cas9多肽和/或包含编码Cas9多肽的多核苷酸的异源核酸构建体,其中杀灭包含所述靶标区的细胞,从而从所述细菌、古细菌、藻类或酵母细胞群体中选择不含所述靶标区并且基因组大小减小的一种或多种细菌、古细菌、藻类或酵母细胞。

[0016] 本发明的第十个方面提供了一种从细菌、古细菌或酵母细胞群体中选择基因组大小减小的一种或多种细菌、古细菌、藻类或酵母细胞的方法,包括:向细菌、古细菌、藻类或酵母细胞群体中引入:(a) (i) 一种或多种异源核酸构建体,所述构建体包含与所述细菌、古细菌、藻类或酵母细胞的基因组中存在的至少300个连续核苷酸具有至少80%同一性的核苷酸序列,或者(ii) 两种或更多种异源核酸构建体,所述构建体包含至少一个转座子,从而产生转基因细菌、古细菌、藻类或酵母细胞的群体,所述转基因细菌、古细菌、藻类或酵母细胞在整合进所述基因组中的所述一种或多种异源核酸构建体与所述基因组中存在的所述至少300个连续核苷酸之间、或整合进所述基因组中的第一转座子与第二转座子之间包含非天然的同源重组位点;和(b) (i) 包含反式编码的CRISPR (tracr) 核酸的异源核酸构建体, (ii) 异源核酸构建体,所述构建体包含CRISPR阵列,所述CRISPR阵列包含重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述至少一个重复-间隔区序列和所述至少一个重复-间隔区-重复序列的间隔区包含与所述群体中的一种或多种细菌、古细菌、藻类或酵母细胞的基因组(染色体和/或质粒)中的靶标区实质上互补的核苷酸序列,和(iii) Cas9多

肽和/或包含编码Cas9多肽的多核苷酸的异源核酸构建体,其中所述靶标区位于掺入进所述基因组中的所述一种或多种异源核酸构建体与所述基因组中存在的所述至少300个连续核苷酸之间和/或所述第一转座子与所述第二转座子之间,以及杀灭包含所述靶标区的细胞,从而从所述转基因细菌、古细菌、藻类或酵母细胞群体中选择不含所述靶标区并且基因组大小减小的一种或多种细菌、古细菌、藻类或酵母细胞。

[0017] 本发明的第十一方面提供了一种在细菌群体中鉴定在其基因组中具有缺失的至少一种分离株的方法,包括:向细菌细胞群体中引入异源核酸构建体,所述构建体包含CRISPR阵列,所述CRISPR阵列包含(5'至3')重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述重复-间隔区-重复序列或所述至少一个重复-间隔区序列的间隔区包含与所述群体中的一种或多种细菌细胞的基因组中的靶标区实质上互补的核苷酸序列,其中杀灭包含所述靶标区的细胞,从而产生不含所述靶标区的转化细菌细胞的群体;以及从所述转化细菌细胞的群体培养独立的细菌集落,从而从所述转化细菌的群体中鉴定在其基因组中具有缺失的至少一种分离株。可用于本发明的CRISPR阵列可以是I型、II型、III型、IV型或V型CRISPR阵列。

[0018] 本发明的第十二方面提供了一种在细菌群体中鉴定在其基因组中具有缺失的至少一种分离株的方法,包括:向所述细菌细胞群体中引入(a) (i) 一种或多种异源核酸构建体,所述构建体包含与所述细菌细胞的基因组中存在的至少300个连续核苷酸具有至少80%同一性的核苷酸序列,或者(ii) 两种或更多种异源核酸构建体,所述构建体包含至少一个转座子,从而产生转基因细菌细胞的群体,所述转基因细菌细胞在整合进所述基因组中的所述一种或多种异源核酸构建体与所述基因组中存在的所述至少300个连续核苷酸之间、或整合进所述基因组中的第一转座子与第二转座子之间包含非天然的同源重组位点;和(b) 异源核酸构建体,所述构建体包含CRISPR阵列,所述CRISPR阵列包含(5'至3')重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述重复-间隔区-重复序列或所述至少一个重复-间隔区序列的间隔区包含与所述群体中的一种或多种细菌细胞的基因组中的靶标区实质上互补的核苷酸序列,其中所述靶标区位于引入进所述基因组中的所述一种或多种异源核酸构建体与所述基因组中存在的所述至少300个连续核苷酸之间和/或所述第一转座子与所述第二转座子之间,以及杀灭包含所述靶标区的细胞,从而产生不含所述靶标区的转化细菌细胞的群体;以及从所述转化细菌细胞的群体培养独立的细菌集落,从而从所述细菌的群体中鉴定在其基因组中具有缺失的至少一种分离株。可用于本发明的CRISPR阵列可以是I型、II型、III型、IV型或V型CRISPR阵列。

[0019] 本发明的第十三方面提供了一种在细菌、古细菌、藻类或酵母细胞群体中鉴定在其基因组中具有缺失的至少一种分离株的方法,包括:向细菌、古细菌、藻类或酵母细胞群体中引入:(a) 包含反式编码的CRISPR(tracr)核酸的异源核酸构建体,(b) 异源核酸构建体,所述构建体包含CRISPR阵列,所述CRISPR阵列包含(5'至3')重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述重复-间隔区-重复序列或所述至少一个重复-间隔区序列的间隔区包含与所述群体中的细菌、古细菌、藻类或酵母细胞的基因组(染色体和/或质粒)中的靶标区实质上互补的核苷酸序列,和(c) Cas9多肽或包含编码Cas9多肽的多核苷酸的异源核酸构建体,其中杀灭包含所述靶标区的细胞,从而产生不含所述靶标区的转化细菌、古细菌、藻类或酵母细胞的群体;以及从所述转化细菌、古细菌、藻类或酵母细胞的群

体培养独立的细菌、古细菌或酵母集落,从而从所述转化细菌、古细菌、藻类或酵母细胞的群体中鉴定在其基因组中具有缺失的至少一种分离株。

[0020] 本发明的第十四方面提供了一种在细菌、古细菌、藻类或酵母群体中鉴定在其基因组中具有缺失的至少一种分离株的方法,包括:向所述细菌、古细菌、藻类或酵母细胞群体中引入(a) (i) 一种或多种异源核酸构建体,所述构建体包含与所述细菌、古细菌、藻类或酵母细胞的基因组中存在的至少300个连续核苷酸具有至少80%同一性的核苷酸序列,或者(ii) 两种或更多种异源核酸构建体,所述构建体包含至少一个转座子,从而产生转基因细菌、古细菌、藻类或酵母细胞的群体,所述转基因细菌、古细菌、藻类或酵母细胞在整合进所述基因组中的所述一种或多种异源核酸构建体与所述基因组中存在的所述至少300个连续核苷酸之间、或整合进所述基因组中的第一转座子与第二转座子之间包含非天然的同源重组位点;和(b) (i) 包含反式编码的CRISPR (tracr) 核酸的异源核酸构建体, (ii) 异源核酸构建体,所述构建体包含CRISPR阵列,所述CRISPR阵列包含(5'至3') 重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述重复-间隔区-重复序列或所述至少一个重复-间隔区序列的间隔区包含与所述群体中的一种或多种细菌、古细菌、藻类或酵母细胞的基因组(染色体和/或质粒)中的靶标区实质上互补的核苷酸序列,和(iii) Cas9多肽和/或包含编码Cas9多肽的多核苷酸的异源核酸构建体,其中所述靶标区位于掺入进所述基因组中的所述一种或多种异源核酸构建体与所述基因组中存在的所述至少300个连续核苷酸之间和/或所述第一转座子与所述第二转座子之间,以及杀灭包含所述靶标区的细胞,从而产生不含所述靶标区的转化细菌、古细菌、藻类或酵母细胞的群体;并且从所述转化细菌、古细菌、藻类或酵母细胞的群体培养独立细菌、古细菌或酵母集落,从而从所述群体中鉴定在其基因组中具有缺失的至少一种分离株。

[0021] 本文还提供了包含本发明的核酸构建体、核酸阵列、核酸分子和/或核苷酸序列的表达盒、细胞和试剂盒。

[0022] 下面在本发明的具体实施方式中更详细地描述本发明的这些方面及其它方面。

附图说明

[0023] 图1示出了用于靶向每个复合转座子的SthCRISPR1阵列和SthCRISPR 3阵列的序列。

[0024] 图2示出了用于构建靶向质粒的剪接重叠延伸(splicing overlap extension, SOE)法的示意图。

[0025] 图3A-3E示出了必需基因、插入序列和基因组岛的图谱。(A) 推定的必需开放阅读框(顶行)、插入序列(第二行)和推定的基因组岛(第三行)的位置和分布。通过对嗜热链球菌(*Streptococcus thermophilus*) LMD-9的基因组内各家族的可转座元件进行基因作图来鉴定CRISPR-Cas(第四行)介导的缺失的潜在靶标。推定的基因组岛及对应于每一者的原间隔区/PAM组合的遗传组织结构。(B) 基因组岛1,编码寡肽转运系统。(C) 基因组岛2,含有细胞包膜蛋白酶PrtS。(D) 基因组岛3,编码ATP酶铜外排蛋白。(E) 编码包含Lac操纵子的选定开放阅读框的基因组岛。

[0026] 图4提供了在嗜热链球菌LMD-9的整个基因组中分布的转座子编码序列的系统树图。用Geneious®软件生成比对。用www-is.biotoul.fr赋给家族名称。字母对应于图5中

每个家族的比对。

[0027] 图5A-5D示出了嗜热链球菌LMD-9基因组中存在的每个主要IS家族的转座子编码序列的比对(使用Geneious®软件)。不同的家族表现出长度和核苷酸同一性的不同保守水平。(A) Sth6转座子由于某些拷贝中的内部缺失而呈高度多态性和明显简并性。相比之下,IS1167(B)和IS1191(C)具有更少的拷贝并且在长度和同一性保持高保真性。IS1193(D)具有高保真性拷贝但在长度方面展现出最大的家族内多样性。

[0028] 图6A-6C示出了由CRISPR-Cas9进行的DNA靶向的结构基础和表观细胞毒性。用于靶向lacZ的SthCRISPR1的间隔区序列(A)和SthCRISPR3的间隔区序列(B)。Cas9侦询DNA并可逆的结合至PAM序列,通过形成tracrRNA::crRNA双链体在靶标处发生Cas9的稳定化。激活Cas9引起RuvC域和HNH域同时切割每条链,如黑色楔形物所指示的。对照及自我靶向质粒的电穿孔后回收转化株(C)。在全部针对对每种测试质粒,在独立转化实验(n=4)中筛选的平均克隆数±SD。

[0029] 图7A-7D示出了Lac⁻克隆的基因组测序和表型分析。序列数据表明使用CRISPR3系统靶向lacZ的(A)5'端和(B)阳离子结合残基编码序列而独立生成的两种突变体中不存在编码lacZ的染色体区段。缺失的大小范围为长101,865-102,146bp,占嗜热链球菌基因组的大约5.5%。(C)大缺失菌株相比于野生型在半合成Elliker培养基中的生长,表示为三次独立的生物重复实验的平均值±SD OD_{600nm}。(D)嗜热链球菌菌株在脱脂乳中的酸化能力。

[0030] 图8A-8E提供了插入序列(IS)之间的重组事件的描述。(A)通过对从转化株回收的gDNA进行PCR分析而产生的大缺失扩增子的电泳图像。使用在推定缺失位点上游和下游的IS1193元件旁侧的引物进行筛选。用Δ指示的泳道扩增自在对lacZ进行CRISPR-Cas介导的靶向后回收的Lac⁻克隆的gDNA,而WT扩增自野生型。(B)通过对对应于上游IS元件(灰色)或下游IS元件的单核苷酸多态性进行基因作图来确定预测重组位点的序列。根据在两个IS元件中保守的序列预测了三个位点(浅灰色)。绘出的位点代表来自独立克隆的基因型并且可代表在九个不同重组位点处观察到的Lac⁻表型。在每个基因组岛的缺失界点处类似地存在嵌合IS元件足迹。(C)预计在编码lacZ的基因岛的染色体缺失期间发生重组的IS的示意图。(D)用于确认缺失的由基因组岛1、2和3旁侧的引物生成的扩增子。(E)由内部引物生成扩增子,用于确认在每种CRISPR诱导缺失培养物中不存在野生型序列。用Δ指示的泳道扩增自在CRISPR-Cas介导的靶向后回收的克隆的gDNA,而WT是野生型。

[0031] 图9提供了致死的靶标并示出了所限定的基因座用于评估基于II型CRISPR-Cas系统通过靶向嗜热链球菌LMD-9的基因组致死的用途。测试了两个正交II型系统(CRISPR1和CRISPR3);CRISPR1靶标为深灰色,CRISPR3靶标为浅灰色。选择特定的遗传特征来测试(i)基因间区(INT)、(ii)可移动遗传元件(ISSth7、oppC-GEI1、prtS-GEI2、copA-GEI3、cI、lacZ-GEI4、epsU)、(iii)必需基因(dltA、ltaS)、(iv)复制弧的极点(oriC、xerS)以及DNA的正链对比负链(外部靶标对比内部靶标)。

[0032] 图10示出了通过靶向图9中所限定的区域而实现的基于CRISPR的致死。参照非靶向质粒对照pORI28的转化计算CFU(细胞形成单位)的对数降低。对于所测试的所有靶标,致死率范围为降低2个对数至降低3个对数,而与染色体位置、编码序列或必需性无关。ISSth7-插入序列元件;ltaS-脂磷壁酸合酶;prtS-基因组岛2;INT-基因间区;dltA-D-丙氨酸连接酶;rheB-chi位点缺陷基因座;oppC-基因组岛1;comS-chi位点稠密基因座;xerS-复

制终点;copA-基因组岛3;cI-前噬菌体残余物;OriC-复制起点;Cas9-CRISPR3Cas9编码序列;epsU-表多糖盒。

[0033] 图11示出了CRISPR介导的基因组岛缺失菌株的转录谱。

[0034] 图12示出了基因组岛缺失菌株GEI1、GEI2、GEI3和GEI4的log₂转化RNA测序读数覆盖度。

[0035] 图13示出了基因组岛缺失菌株表达值(X-轴)对野生型表达值(Y-值)的XY曲线图。对于每种基因组岛缺失菌株(GEI1-GEI4),靶标基因组岛(黑色)每一者上所编码的基因的表达最少。GEI1中编码的基因在左上分图中示出,GEI2中编码的基因在右上分图中示出,GEI3中编码的基因在左下分图中示出,GEI4中编码的基因在右下分图中示出。

[0036] 图14A-14B示出了引入编码CRISPR阵列(II型系统)的外源噬菌体、质粒或噬菌粒来增补内源系统用于嗜热链球菌中的程序化细胞死亡。

[0037] 图15.干酪乳杆菌(*Lactobacillus casei*)的示例性II型向导结构。顶部的结构是预测的向导结构。左下图是通过RNA测序确认的正确双向向导结构crRNA:tracrRNA。右下图是预测的人工单向向导结构的实例。

[0038] 图16提供了加氏乳杆菌(*Lactobacillus gasseri*)的示例性II型向导结构。顶部的结构是预测的向导结构。左下图是通过RNA测序确认的正确双向向导结构crRNA:tracrRNA。右下图是预测的人工单向向导结构的实例。

[0039] 图17提供了戊糖乳杆菌(*Lactobacillus pentosus*)的示例性II型向导结构。顶部的结构是预测的向导结构。左下图是通过RNA测序确认的正确双向向导结构crRNA:tracrRNA。右下图是预测的人工单向向导结构的实例。

[0040] 图18提供了詹氏乳杆菌(*Lactobacillus jensenii*)的示例性II型向导结构。左分图是通过RNA测序确认的正确双向向导结构crRNA:tracrRNA。右分图提供了预测的人工单向向导结构的实例。

[0041] 图19示出了质粒转化的结果,该质粒含有与天然加氏乳杆菌II型CRISPR阵列中最高转录的crRNA匹配的原间隔区。从左到右,四种不同的质粒被转化进加氏乳杆菌:空的pTRK563载体、具有正确原间隔区但具有不正确PAM的构建体、具有正确PAM但具有不在该阵列中的原间隔区的构建体以及具有正确原间隔区且具有证明了最大干扰靶向和细胞死亡的PAM的构建体。所报道的值代表三次独立重复的平均值±SEM。

[0042] 图20示出了质粒的转化,该质粒含有与天然戊糖乳杆菌II型CRISPR阵列中最高转录的crRNA匹配的原间隔区。从左到右,四种不同的质粒被转化进戊糖乳杆菌:具有正确原间隔区但具有不正确PAM的构建体(Lpe4 ctGttt)、具有正确PAM但具有不在该阵列中的原间隔区的构建体(Lpe8 noSPCR)、空pTRK563载体(pTRK563)以及具有正确原间隔区和正确PAM的质粒(Lpe1 gttaat)。所报道的值代表三次独立重复的平均值±SEM。

[0043] 图21提供了干酪乳杆菌的示例性I型CRISPR-Cas向导结构。所提供的序列是干酪乳杆菌NCK 125中存在的天然I型前导序列和重复。该人工阵列含有靶向宿主基因组中的16s rDNA基因的间隔区。

[0044] 图22示出了质粒的转化,该质粒含有与天然詹氏乳杆菌II型CRISPR阵列中最高转录的crRNA匹配的原间隔区。从左到右,四种不同的质粒被转化进詹氏乳杆菌:空的pTRK563载体、具有正确原间隔区但具有不正确PAM的构建体、具有正确PAM但具有不在该阵

列中的原间隔区的构建体以及具有正确原间隔区且具有证明了最大干扰靶向和细胞死亡的PAM的构建体。

[0045] 图23示出了干酪乳杆菌NCK 125中使用天然I型系统进行的靶向自杀。在16s rDNA基因中设计了两个靶标。使用该生物体中天然的间隔区序列预测PAM 5'-YAA-3'。将含有天然I型前导序列、重复和所选择的间隔区的人工阵列克隆进pTRK870中。所引入的构建体包括空质粒(pTRK563)和两种不同的人工阵列:一种含有靶向16s基因中的+链的单间隔区(1-2 alt),另一种阵列含有靶向+链的初始间隔区并且含有靶向16s基因中的-链的额外间隔区(1,2-3)。所报道的值代表三次独立重复的平均值 $\pm$ SEM。

具体实施方式

[0046] 现在将在下文参照附图和实施例描述本发明,其中示出了本发明的实施方案。这些描述并非意图成为可实施本发明的所有不同方式或可给本发明增加的所有特征的详细目录。例如,相对于一个实施方案所示例的特征可并入其它实施方案中,并且相对于一个具体实施方案示例的特征可从该实施方案删除。因而,本发明设想,在本发明的一些实施方案中,可排除或忽略本文所述的任何特征或特征组合。此外,根据本公开,针对本文所提出的各种实施方案的许多变型和添加对于本领域技术人员将是显而易见的,其不脱离本发明。从而,下面的描述旨在示例本发明的一些具体的实施方案,而不是穷举性地说明所有排列、组合物及其变型形式。

[0047] 除非另外定义,否则本文使用的所有科技术语的含义与本发明所属领域的普通技术人员通常理解的一致。在本文的描述中所用的术语仅仅是为了描述具体的实施方案而无意于限制本发明。

[0048] 本文所引用的所有出版物、专利申请、专利和其它参考文献以引用的方式全文并入,用于与所述参考文献所存在的句子和/或段落相关的教导。

[0049] 除非上下文中另外指明,否则特别意图是本文所述的本发明的各种特征可以以任何组合使用。此外,本发明还设想,在本发明的一些实施方案中,可排除或忽略本文所述的任何特征或特征组合。举例来说,如果本说明书声称组合物包含组分A、B和C,则特别意图是A、B或C任一者或它们的组合可以单独地或以任何组合忽略或排除。

[0050] 如在本发明的具体实施方式和随附的权利要求书中所用的,单数形式“一个”、“一种”和“所述”旨在还包括复数形式,除非上下文明确地指出并非如此。

[0051] 另外,如本文所用的,“和/或”是指并且涵盖所列相关各项中的一者或多者的任何和全部可能的组合,在理解为备选(“或”)时缺少组合。

[0052] 如本文所用,当提及可测量值诸如剂量或时间周期等时,术语“约”是指所指定量的 $\pm 20\%$ 、 $\pm 10\%$ 、 $\pm 5\%$ 、 $\pm 1\%$ 、 $\pm 0.5\%$ 或甚至 $\pm 0.1\%$ 的波动。

[0053] 如本文所用,短语诸如“介于X与Y之间”以及“介于约X与Y之间”应该理解为包括X和Y。如本文所用,短语诸如“介于约X与Y之间”意指“介于约X与约Y之间”,并且短语诸如“约X至Y”意指“约X至约Y”。

[0054] 本文所用的术语“包括”、“包含”指明存在所述特征、整数、步骤、操作、元件和/或组分,但并不排除存在或增加一种或多种其它特征、整数、步骤、操作、元件、组分和/或它们的组合。

[0055] 如本文所用的,连接词“基本上由...组成”意指权利要求的范围应理解为涵盖权利要求中所述的指定材料或步骤以及不实质影响要求权利保护的发明的基本特征和新颖特征的那些材料或步骤。因而,术语“基本上由...组成”在用于本发明的权利要求中时并非意图应理解为等价于“包含”。

[0056] “Cas9核酸酶”是指CRISPR Cas系统中催化双链DNA切割的一大组内切核酸酶。这些多肽是本领域众所周知的并且许多它们的结构(序列)得到表征(参见例如W02013/176772;W0/2013/188638)。催化双链DNA切割的结构域是RuvC结构域和HNH结构域。RuvC结构域负责使(-)链产生切口,而HNH结构域负责使(+)链产生切口(参见例如Gasiunas等人,PNAS109(36):E2579-E2586(2012年9月4日))。

[0057] 如本文所用,“嵌合”是指其中有至少两种组分源于不同来源(如不同生物体、不同编码区)的核酸分子或多肽。

[0058] 本文所用的“互补”可以意指与比较的核苷酸序列具有100%的互补性或同一性,或者其可以意指少于100%的互补性(如,约70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%等的互补性)。

[0059] 本文所用的术语“互补的”或“互补性”是指在碱基配对容许的盐和温度条件下多核苷酸的自然结合。例如,序列“A-G-T”结合至互补序列“T-C-A”。两个单链分子之间的互补性可以是“部分的”,其中仅某些核苷酸结合,或者当所述单链分子之间存在总的互补性时该互补性可以完全的。核酸链之间的互补程度对核酸链之间的杂交效率和强度具有重大的影响。

[0060] 如本文所用,“接触”、“使...接触”、“接触的”以及它们的语法上的变型形式,是指将所需反应的各组分在适于进行所述所需反应(如,整合、转化、筛选、选择、杀灭、鉴定、扩增等等)的条件下放在一起。适于进行这类反应的方法和条件是本领域众所周知的(参见例如,Gasiunas等人,(2012) Proc. Natl. Acad. Sci. 109:E2579-E2586;M.R.Green和J.Sambrook(2012) Molecular Cloning: A Laboratory Manual. 第4版, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY)。

[0061] 本文所用的“缺失”可包括遗传物质的丧失或缺失,包括但不限于一部分染色体或质粒缺失、基因或一部分基因从染色体或质粒缺失。在一些实施方案中,缺失可包含一个基因或不止一个基因。在一些实施方案中,缺失还可包括可编码小的非编码RNA的非蛋白质编码区的丢失。在一些实施方案中,缺失可包括丢失整个质粒或丢失整个可移动的遗传元件。在一些实施方案中,可移动遗传元件的丢失可定义为例如下不能复制或存留。

[0062] 在一些实施方案中,本发明的噬菌粒可包含来自I型CRISPR-Cas系统、II型CRISPR-Cas系统、III型CRISPR-Cas系统、IV型CRISPR-Cas系统和/或V型CRISPR-Cas系统的CRISPR阵列(参见Makarova等人,Nature Reviews Biotechnology 13:722736(2015))。

[0063] 因而,在一些实施方案中,除了I型crRNA之外,本发明的质粒还可包含I型多肽和/或I型Cascade多肽(即I型CRISPR-Cas系统)。

[0064] 如本文所用,“I型多肽”是指Cas3多肽、Cas3'多肽、Cas3"多肽、它们的融合变体中的任一者和I型规律间隔的成簇短回文重复(CRISPR)相关抗病毒防御复合体(“Cascade”)多肽中的任一者或多者。因而,术语“I型多肽”是指构成I-A型CRISPR-Cas系统、I-B型

CRISPR-Cas系统、I-C型CRISPR-Cas系统、I-D型CRISPR-Cas系统、I-E型CRISPR-Cas系统、I-F型CRISPR-Cas系统和/或I-U型CRISPR-Cas系统的多肽。每种I型CRISPR-Cas系统包含至少一种Cas3多肽。Cas3多肽通常包含解旋酶结构域和HD结构域二者。然而,在一些I型CRISPR-Cas系统中,解旋酶和HD结构域存在于独立的多肽Cas3'和Cas3"中。具体地讲,Cas3'编码解旋酶结构域而Cas3"编码HD结构域。因此,因为两个结构域都是Cas3功能所需的,I型的亚型要么编码Cas3 (I-C、I-D、I-E、I-F、I-U),要么编码Cas3'和Cas3" (I-A、I-B)。

[0065] 如本文所用,“I型Cascade多肽”是指涉及前crRNA的加工并随后结合至I型CRISPR-Cas系统中的靶标DNA的多肽复合体。这些多肽包括但不限于I型I-A、I-B、I-C、I-D、I-E和I-F亚型的Cascade多肽。I-A型Cascade多肽的非限制性实例包括Cas7 (Csa2)、Cas8a1 (Csx13)、Cas8a2 (Csx9)、Cas5、Csa5、Cas6a、Cas3'和/或Cas3"。I-B型Cascade多肽的非限制性实例包括Cas6b、Cas8b (Csh1)、Cas7 (Csh2)和/或Cas5。I-C型Cascade多肽的非限制性实例包括Cas5d、Cas8c (Csd1)和/或Cas7 (Csd2)。I-D型Cascade多肽的非限制性实例包括Cas10d (Csc3)、Csc2、Csc1和/或Cas6d。I-E型Cascade多肽的非限制性实例包括Cse1 (CasA)、Cse2 (CasB)、Cas7 (CasC)、Cas5 (CasD)和/或Cas6e (CasE)。I-F型Cascade多肽的非限制性实例包括Cys1、Cys2、Cas7 (Cys3)和/或Cas6f (Csy4)。I-U型Cascade多肽的非限制性实例包括Cas8c、Cas7、Cas5、Cas6和/或Cas4。

[0066] 在一些实施方案中,除了II型crRNA以外,本发明的质粒可还包含II型CRISPR-Cas系统。II型CRISPR-Cas系统包括三种亚型:II-A型、II-B型和II-C型,除了适应性多肽Cas1、Cas2和任选的Csn2和/或Cas4以外,它们每一者都还包含多结构域蛋白质Cas9。大多数II型基因座还编码tracrRNA。包含示例性的II型CRISPR-Cas系统的生物体包括嗜肺性军团病杆菌巴黎菌株 (*Legionella pneumophila* str. Paris)、嗜热链球菌CNRZ1066和乳糖奈瑟菌 (*Neisseria lactamica*) 020-06。

[0067] 在另外的实施方案中,除了III型crRNA以外,本发明的质粒可还包含III型CRISPR-Cas系统。与I型CRISPR-Cas系统类似,在III型系统中加工和干扰由多蛋白CRISPR RNA (crRNA)-效应蛋白复合体介导 (Makarova等人, *Nature Reviews Biotechnology* 13: 722736 (2015)) -I型中的“CASCADE”以及III型中的“Csm”或“Cmr”。因而,在一些实施方案中,III型CRISPR-Cas系统可包含Csm复合体 (如,III-A型Csm)和/或Cmr复合体 (如,III-B型Cmr)以及任选地Cas6多肽。在代表性的实施方案中,Csm复合体可包含Cas10 (或Csm1)、Csm2、Csm3、Csm4、Csm5和Csm6多肽,而Cmr复合体可包含Cmr1、Cas10 (或Cmr2)、Cmr3、Cmr4、Cmr5和Cmr6多肽。除了Csm复合体或Cmr复合体外,III型CRISPR-Cas系统可还包含Cas7多肽。已经表征了III型CRISPR-Cas系统的四种亚型III-A、III-B、III-C、III-D。在一些实施方案中,III-A型CRISPR-Cas系统包含Cas6、Cas10、Csm2、Cas7 (Csm3)、Cas5 (Csm4)、Cas7 (Csm5)和Csm6多肽。在一些实施方案中,III-B型CRISPR-Cas系统包含Cas7 (Cmr1)、Cas10、Cas5 (Cmr3)、Cas7 (Cmr4)、Cmr5、Cas6和Cas7 (Cmr6)多肽。在一些实施方案中,III型CCRISPR-Cas系统包含Cas7 (Cmr1)、Cas7 (Cmr6)、Cas10、Cas7 (Cmr4)、Cmr5和Cas5 (Cmr3)多肽。在一些实施方案中,III-D型CRISPR-Cas系统包含Cas10、Cas7 (Csm3)、Cas5 (Cs×10)、Csm2、Cas7 (Csm3)和a111473多肽。

[0068] 在一些实施方案中,除了IV型crRNA以外,本发明的质粒可还包含IV型CRISPR-Cas系统。IV型CRISPR-Cas系统可包含Csf4多肽 (dinG)和/或Csf1、Cas7 (Csf2)和/或Cas5

(csf3) 多肽。(Makarova等人,Nature Reviews Microbiology 13:722-736 (2015))。

[0069] 在一些实施方案中,除了V型crRNA以外,本发明的质粒还包含V型CRISPR-Cas系统。V型CRISPR-Cas系统可包含Cpf1多肽和/或Cas1、Cas2和/或Cas4多肽。(Makarova等人,Nature Reviews Microbiology 13:722-736 (2015))。

[0070] 本发明的核苷酸序列的“片段”或“部分”应该理解为意指相对于参照核酸或核苷酸序列而言长度减少(如,减少1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20个或更多个核苷酸)并且包含与参照核酸或核苷酸序列相同或实质上相同(如,70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%同一性)的邻接核苷酸的核苷酸序列、基本上由该邻接核苷酸的核苷酸序列组成和/或由该邻接核苷酸的核苷酸序列组成的核苷酸序列。根据本发明的这种核酸片段或部分在适当的情况下可包括在较大的多核苷酸中作为其组分。因而,杂交至(与...杂交,以及其其它的语法变型形式)例如靶标DNA(如基因组中的靶标区)的至少一部分是指杂交至与该靶标DNA的一定长度的毗邻核苷酸相同或基本相同的核苷酸序列。在一些实施方案中,重复-间隔区序列或重复-间隔区-重复序列的重复可包含野生型CRISPR基因座的重复序列或合成CRISPR阵列的重复序列的片段,其中所述重复序列片段保留重复序列在CRISPR阵列中与tracr核酸杂交的功能。

[0071] 在一些实施方案中,本发明可包含Cas9、Cas3、Cas3'、Cas3"或Cpf1核酸酶的功能片段。Cas9的功能片段保留天然Cas9核酸酶的一种或多种活性,包括但不限于HNH核酸酶活性、RuvC核酸酶活性、DNA、RNA和/或PAM识别和结合活性。Cas9核酸酶的功能片段可由Cas9多核苷酸的片段编码。Cas3、Cas3'或Cas3"功能片段保留天然Cas9核酸酶的一种或多种活性,包括但不限于切口酶活性、核酸外切酶活性、DNA结合和/或RNA结合。Cas3、Cas3'或Cas3"核酸酶的功能片段可分别由Cas3、Cas3'或Cas3"多核苷酸的片段编码。

[0072] 如本文所用,术语“基因”是指能够用于产生mRNA、反义RNA、RNAi(miRNA、siRNA、shRNA)、抗微小RNA反义寡脱氧核苷酸(AMO)等的核酸分子。基因可能能够或不能够用于产生功能蛋白质或基因产物。基因可包括编码区和非编码区(如内含子、调控元件、启动子、增强子、终止序列和/或5'和3'非翻译区)二者。基因可以是“分离的”,所谓“分离的”意指这样的核酸,该核酸实质上或基本上不含通常与处于天然状态的该核酸相关联存在的组分。这类组分包括其它细胞材料、来自重组制备的培养基和/或化学合成该核酸时所用的多种化学品。

[0073] 本文所用的术语“基因组”包括生物体的染色体/核基因组以及任何线粒体和/或质粒基因组。

[0074] 本文所用“发夹序列”是包含发夹的(如形成一个或多个发夹结构的)核苷酸序列。发夹(如,茎-环结构、折回结构)是指具有二级结构的核酸分子,该二级结构包括形成双链的互补核苷酸的区域,这些互补核苷酸在任一侧侧接单链区。这类结构是本领域众所周知的。如本领域已知的,双链区可在碱基配对中包含一些错配,或者可以是完美匹配的。在本公开的一些实施方案中,核酸构建体的发夹序列可位于tracr核酸的3'端。

[0075] “异源”或“重组”核苷酸序列是与其被引入的宿主细胞不天然相关的核苷酸序列,包括天然存在的核苷酸序列的非天然存在的多拷贝。

[0076] 具有同源性的不同核酸或蛋白质在本文称为“同源物”。术语同源物包括来自相同和其它物种的同源序列以及来自相同和其它物种的直系同源序列。“同源性”是指两个或更多个核酸和/或氨基酸序列之间在位置同一性(即,序列相似性或同一性)的百分比方面的相似性水平。同源性还指不同核酸或蛋白质之间类似功能特性的概念。因而,本发明的组合物和方法还包含本发明的核苷酸序列和多肽序列的同源物。本文所用的“直系同源物”是指在物种形成期间从共同祖先基因产生的不同物种中的同源核苷酸序列和/或氨基酸序列。本发明的核苷酸序列的同源物与所述本发明核苷酸序列具有实质性的序列同一性(如至少约70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%和/或100%)。因而,例如,I型、II型、III型、IV型或V型多核苷酸或多肽的同源物可与任何已知的或后来鉴定的I型、II型、III型、IV型或V型多核苷酸或多肽中的任一者具有约70%同源性或更高同源性。

[0077] 如本文所用的,杂交、与...杂交以及其语法变型形式是指两条完全互补的核苷酸序列或其中可存在一些错配碱基对的实质上互补的序列的结合。杂交条件是本领域众所周知的,并且基于核苷酸序列的长度和核苷酸序列之间的互补程度而变化。在一些实施方案中,杂交条件可以是高严格性,或者它们可以是中严格性或低严格性,这取决于互补性的量和待杂交的序列的长度。为了核苷酸序列之间的杂交而构成低、中和高严格性的条件是本领域众所周知的(参见例如,Gasiunas等人,(2012) *Proc. Natl. Acad. Sci.* 109:E2579-E2586; M.R.Green和J.Sambrook (2012) *Molecular Cloning: A Laboratory Manual*. 第4版, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY)。

[0078] 如本文所用,术语“增加”、“升高”、“增强”(以及它们的语法变型形式)描述与对照相比提高至少约25%、50%、75%、100%、150%、200%、300%、400%、500%或更多。

[0079] “天然”或“野生型”核酸、核苷酸序列、多肽或氨基酸序列是指天然存在的或内源的核酸、核苷酸序列、多肽或氨基酸序列。因而,例如,“野生型mRNA”是该生物体中天然存在的或其内源的mRNA。“同源”核酸序列是与其被引入的宿主细胞天然相关的核苷酸序列。

[0080] 另外如本文所用,术语“核酸”、“核酸分子”、“核酸构建体”、“核苷酸序列”和“多核苷酸”是指RNA或DNA,该RNA或DNA是线性或分支的、单链或双链的、或它们的杂合体。该术语还涵盖RNA/DNA杂合体。当以合成方式制备dsRNA时,不太常见的碱基,诸如肌苷、5-甲基胞嘧啶、6-甲基腺嘌呤、次黄嘌呤等等也可用于反义、dsRNA和核酶配对。例如,含有尿苷和胞苷的C-5丙炔类似物的多核苷酸已显示以高亲和力结合RNA,并且是基因表达的有效反义抑制剂。也可以进行其它修饰,诸如对磷酸二酯骨架或RNA的核糖糖基中的2'-羟基的修饰。本公开的核酸构建体可以是DNA或RNA,但优选是DNA。因此,尽管本发明的核酸构建体可以根据预期用途以DNA的形式描述和使用,但它们也可以以RNA的形式描述和使用。

[0081] 如本文所用,术语“核苷酸序列”是指核苷酸的杂聚物或这些核苷酸从核酸分子的5'至3'端的序列,并且包括DNA或RNA分子(包括cDNA、DNA片段或部分、基因组DNA、合成的(例如,化学合成的)DNA、质粒DNA、mRNA和反义RNA,它们任一者都可以是单链或双链的)。术语“核苷酸序列”、“核酸”、“核酸分子”、“寡核苷酸”和“多核苷酸”在本文中也可互换使用,指核苷酸的杂聚物。除非另有说明,否则本文提供的核酸分子和/或核苷酸序列在本文中以5'至3'方向从左至右给出,并且使用如美国序列法规37CFR§§1.821-1.825和世界知识产权

组织(WIPO)标准ST.25中说明的用于代表核苷酸身份的标准代码来代表。

[0082] 如本文所用,术语“序列同一性百分比”或“同一性百分比”是指当最佳比对两个序列时参照(“查询”)多核苷酸分子(或其互补链)的线性多核苷酸序列中与测试(“主题”)多核苷酸分子(或其互补链)相比相同核苷酸的百分比。在一些实施方案中,“同一性百分比”可以指氨基酸序列中相同氨基酸的百分比。

[0083] “原间隔区序列”是指靶标双链DNA,并且尤其是指靶标DNA(如,或者基因组中的靶标区)中与CRISPR重复-间隔区序列、CRISPR重复-间隔区-重复序列和/或CRISPR阵列的间隔区序列完全互补或实质上互补的部分。

[0084] 如本文所用,术语“减少”、“减小”、“抑制”和“降低”(以及它们的语法变型形式)描述例如与对照相比降低至少约5%、10%、15%、20%、25%、35%、50%、75%、80%、85%、90%、95%、97%、98%、99%或100%。在具体实施方案中,所述减小导致所测组分(如细胞群体或基因组大小)没有或基本上没有(即,不显著的量,例如小于约10%或甚至小于约5%)可检测的活性或量。因而,例如,减小的基因组大小可意指与对照相比基因组大小减小至少约5%、10%、15%、20%、25%、35%、50%、75%、80%、85%、90%、95%、97%、98%、99%或100%。

[0085] 本文所用的对照可以是例如,尚未用本发明的异源核酸构建体转化的细菌、古细菌、藻类或酵母细胞的群体。在一些实施方案中,对照可以是细菌、古细菌、藻类或酵母细胞的野生型群体,或者其可以用这样的异源构建体转化的细菌、古细菌或酵母细胞群体,该异源构建体包含CRISPR阵列,该CRISPR阵列包含重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述间隔区包含的核苷酸序列不与所述群体的细菌、古细菌或酵母细胞的基因组中的靶标区互补(即,非自我靶向/“加扰(scrambled)间隔区”)。在另外的方面,对照可以是例如,细菌、古细菌、藻类或酵母细胞的野生型群体,或者用这样的异源构建体转化的细菌、古细菌、藻类或酵母细胞群体,该异源构建体包含CRISPR阵列,该CRISPR阵列包含重复-间隔区-重复序列或至少一个重复-间隔区序列,其中间隔区包含的核苷酸序列与所述群体的细菌、古细菌、藻类或酵母细胞的基因组中位置不与原间隔区相邻基序(PAM)相邻的靶标区实质上互补。

[0086] 本文所用的“重复序列”是指例如,野生型CRISPR基因座的任何重复序列或合成CRISPR阵列的被“间隔区序列”分隔(如,本发明的重复-间隔区或重复-间隔区-重复序列)的重复序列。可用于本发明的重复序列可以是任何已知的或后来鉴定的CRISPR基因座的重复序列。因此,在一些实施方案中,重复-间隔区序列或重复-间隔区-重复包含与来自野生型II型CRISPR阵列的重复实质上相同(如至少约70%同一性(如至少约70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高))的重复。在一些实施方案中,重复序列与来自野生型I型CRISPR阵列、野生型II型CRISPR阵列、野生型III型CRISPR阵列、野生型IV型CRISPR阵列或野生型V型CRISPR阵列的重复100%相同。在另外的实施方案中,可用于本发明的重复序列可包含具有部分重复的核苷酸序列,该部分重复是I型crRNA、II型crRNA、III型crRNA、IV型crRNA或V型crRNA任一者的CRISPR基因座或合成CRISPR阵列的重复序列的连续核苷酸的片段或部分。

[0087] 如本文所用,I型、II型、III型、IV型或V型CRISPR-Cas系统的“CRISPR阵列”是指核

酸构建体,该核酸构建体从5'至3'包含重复-间隔区-重复序列或从5'至3'包含至少一个重复-间隔区序列(如约1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24或25个重复-间隔区序列,和其中的任何范围或值)。当不止一个重复-间隔区包含在CRISPR阵列中时,前面的(5'至3')重复-间隔区的间隔区可连接至后面的重复-间隔区的重复(如,第一重复-间隔区序列的间隔区连接至第二重复-间隔区的重复)。在一些实施方案中,CRISPR阵列可包含被间隔区隔开的两个重复(或两个部分重复)(如,重复-间隔区-重复序列)。

[0088] 如本文所用,“序列同一性”是指两个最佳比对的多核苷酸或肽序列在组分(例如,核苷酸或氨基酸)的整个比对窗口内不变的程度。“同一性”可容易地通过已知的方法计算,所述方法包括但不限于以下中描述的那些:Computational Molecular Biology(Lesk, A.M.(编辑))Oxford University Press,New York(1988);Biocomputing:Informatics and Genome Projects(Smith,D.W.(编辑))Academic Press,New York(1993);Computer Analysis of Sequence Data,Part I(Griffin,A.M.和Griffin,H.G.(编辑))Humana Press,New Jersey(1994);Sequence Analysis in Molecular Biology(von Heinje,G.(编辑))Academic Press(1987);以及Sequence Analysis Primer(Gribskov,M.和Devereux,J.(编辑))Stockton Press,New York(1991)。

[0089] 本文所用的“间隔区序列”是与靶标DNA(即基因组或“原间隔区序列”中的靶标区,其邻近原间隔区相邻基序(PAM)序列)互补的核苷酸序列。间隔区序列可与靶标DNA完全互补或实质上互补(如至少约70%互补(如约70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高))。在代表性的实施方案中,间隔区与靶标DNA具有100%互补性。在另外的实施方案中,间隔区序列的3'区与靶标DNA的互补性为100%,但间隔区的5'区中互补性低于100%,因此间隔区序列与靶标DNA的总体互补性低于100%。因而,例如,20个核苷酸的间隔区序列(种子序列)的3'区中前7、8、9、10、11、12、13、14、15、16等个核苷酸可能与靶标DNA 100%互补,而该间隔区序列5'区中的其余核苷酸与靶标DNA实质上互补(如至少约70%互补)。在一些实施方案中,间隔区序列的前7至12个核苷酸可能与靶标DNA 100%互补,而该间隔区序列5'区中的其余核苷酸与靶标DNA实质上互补(如至少约70%互补)。在其它实施方案中,间隔区序列的前7至10个核苷酸可能与靶标DNA 100%互补,而该间隔区序列5'区中的其余核苷酸与靶标DNA实质上互补(如至少约70%互补)。在代表性的实施方案中,间隔区序列的前7个核苷酸(该种子序列内)可能与靶标DNA 100%互补,而该间隔区序列5'区中的其余核苷酸与靶标DNA实质上互补(如至少约70%互补)。

[0090] 如本文所用,“靶标DNA”、“靶标区”或“基因组中的靶标区”是指生物体的基因组中与重复-间隔区序列或重复-间隔区-重复序列的间隔区序列完全互补或实质上互补(如至少70%互补(如,70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高))。在一些实施方案中,靶标区可以长度约为10至约40个连续核苷酸,位置紧邻生物体基因组(如,I型CRISPR-Cas系统和II型CRISPR-Cas系统)中的PAM序列(位置紧邻靶标区3'的PAM)。在一些实施方案中,如I型系统中,PAM在原间隔区的另一侧(5'

端)。对于III型系统没有已知的PAM。Makarova等人描述了CRISPR系统的所有类别、类型和亚型的命名(Nature Reviews Microbiology 13:722-736 (2015))。R.Barrangou (Genome Biol.16:247 (2015))中描述了向导结构和PAM。

[0091] 在一些实施方案中,可用于本发明的靶标区位于必需基因或非必需基因内。

[0092] 在代表性的实施方案中,可以随机选择或可以具体选择靶标区。在一些实施方案中,随机选择的靶标区可以选自位置紧邻细菌、古细菌、藻类或酵母细胞基因组中的PAM序列的至少10个连续核苷酸(如,10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50个等等,以及其中的任何范围或值)。在一些实施方案中,靶标区可以是位置紧邻细菌、古细菌、藻类或酵母细胞基因组中的原间隔区相邻基序(PAM)序列的约10至约20个连续核苷酸、约10至约30个连续核苷酸和/或约10至约40个连续核苷酸等等,或其中的任何范围或值。在一些实施方案中,具体选择靶标区可包括选择在一种或多种细菌、古细菌、藻类或酵母细胞的基因组中选择彼此相距约每100个核苷酸至约每1000个核苷酸、约每100个核苷酸至约每2000个、约每100个核苷酸至约每3000个、约每100个核苷酸至约每4000个和/或约每100个核苷酸至约每5000个核苷酸等的两个或更多个靶标区。在特定的实施方案中,具体选择靶标区可包括具体地从基因、开放阅读框、推定的开放阅读框或基因间区选择靶标区,所述靶标区包含位置紧邻细菌、古细菌、藻类或酵母细胞基因组中的PAM序列的至少约10至约40个连续核苷酸。

[0093] 本文所用的“反式激活CRISPR(tracr)核酸”或“tracr核酸”是指任何tracr RNA(或其编码DNA)。tracr核酸从5'至3'包括下茎区、上茎区、凸环、连接发夹和末端发夹(参见Briner等人,(2014)Molecular Cell.56(2):333-339)。反式激活CRISPR(tracr)核酸作用在于与成熟或不成熟crRNA的重复部分杂交、将Cas9蛋白募集至靶标位点,并且可通过诱导结构重排而帮助Cas9的催化活性。tracrRNA分子的功能性组成在上面给出。tracrRNA的序列对CRISPR-Cas系统是特异性的并且可以变化。任何tracr核酸(已知的或后来鉴定的)均可用于本发明。

[0094] 如本文所用,在两个核酸分子、核苷酸序列或蛋白质序列的语境中,短语“实质上相同”或“实质同一性”是指这样的两个或更多个序列或子序列,当使用以下序列比较算法之一或通过目视检查测量而针对最大对应性比较和比对时,具有至少约70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%和/或100%的核苷酸或氨基酸残基同一性。在本发明的一些实施方案中,在序列内长度至少约50个残基至约150个残基的区域上存在实质同一性。因而,在本发明的一些实施方案中,在序列内长度至少约3至约15个(如,长3、4、5、6、7、8、9、10、11、12、13、14、15个残基等,或其中的任何值或任何范围)、长度至少约5至约30、至少约10至约30、至少约16至约30、至少约18至至少约25、至少约18、至少约22、至少约25、至少约30、至少约40、至少约50、约60、约70、约80、约90、约100、约110、约120、约130、约140、约150或更多个残基以及其中的任何范围的区域上存在实质同一性。在代表性的实施方案中,序列可在至少约22个核苷酸上实质上相同。在一些特定的实施方案中,序列可在至少约150个残基上实质上相同。在一些实施方案中,本发明的序列可在至少约16个核苷酸至约25个核苷酸上具有约70%至约100%的同一性。在一些实施方案中,

本发明的序列可在至少约16个核苷酸至约25个核苷酸上具有约75%至约100%的同一性。在另外的实施方案中,本发明的序列可在至少约16个核苷酸至约25个核苷酸上具有约80%至约100%的同一性。在另外的实施方案中,本发明的序列可在至少约7个核苷酸至约25个核苷酸上具有约80%至约100%的同一性。在一些实施方案中,本发明的序列可在约至少约18个核苷酸上具有约70%的同一性。在其它实施方案中,序列可在约至少约22个核苷酸上具有约85%的同一性。在其它实施方案中,序列可在约至少约16个核苷酸上具有100%的同一性。在另一个实施方案中,序列在整个长度的编码区上实质上相同。此外,在示例的实施方案中,实质上相同的核苷酸或多肽序列行使实质上相同的功能(如,crRNA、tracrRNA、重复序列、Cas9核酸酶(切口酶、DNA、RNA和/或PAM识别和结合)、Cas3、Cas3'、Cas3"或任何其它CRISPR-Cas多核苷酸或多肽的功能或活性)。

[0095] 对于序列比较,通常一条序列充当与测试序列进行比较的参照序列。当使用序列比较算法时,将测试和参照序列输入计算机,如果必要的话,指定子序列坐标,并指定序列算法程序参数。然后,序列比较算法基于指定的程序参数计算测试序列相对于参照序列的序列同一性百分比。

[0096] 用于比对比较窗口的序列最佳比对是本领域技术人员众所周知的,并且可以通过以下工具进行:诸如Smith和Waterman的局部同源性算法,Needleman和Wunsch的同源性比对算法,Pearson和Lipman的相似性搜索方法,任选这些算法的计算机化执行,诸如作为**GCG®WisconsinPackage®** (Accelrys Inc., San Diego, CA)的一部分可供利用的GAP、BESTFIT、FASTA和TFASTA。测试序列和参照序列的比对区段的“同一性分数”是两个比对序列共享的相同组分的数目除以参照序列区段中组分(即整个参照序列或参照序列的较小限定部分)的总数。序列同一性百分比表示为同一性分数乘以100。一条或多条多核苷酸序列的比较可以与全长多核苷酸序列或其部分或更长的多核苷酸序列进行。对于本发明而言,还可以使用用于翻译的核苷酸序列的BLASTX版本2.0和用于多核苷酸序列的BLASTN版本2.0来确定“同一性百分比”。

[0097] 用于进行BLAST分析的软件可以通过美国国家生物技术信息中心(National Center for Biotechnology Information)公开获得。该算法涉及首先通过在询问序列中鉴定长度为W的字段而鉴定高得分序列对(HSP),当与数据库中相同长度的字段比对时,其或者匹配或者满足一些正数阈值得分T。T称为邻近字段得分阈值(Altschul等人,1990)。这些初始邻近字段命中充当种子用于起始寻找含有它们的更长HSP的搜索。然后沿每条序列在两个方向延伸字段命中,只要可以增加累积比对分数。对于核苷酸序列,使用参数M(对匹配残基对的奖励分数;总是>0)和N(错配残基的罚分;总是<0)来计算累积分数。对于氨基酸序列,用打分矩阵计算累积分数。当累积比对分数从其最大实现值下降量X、累积分数由于累积一个或多个负得分残基比对而达到0或低于0、或达到任一条序列的末端时,字段命中在每一方向的延伸即停止。BLAST算法参数W、T和X决定了比对的敏感度和速度。BLASTN程序(用于核苷酸序列)使用以下作为默认值:字长(W)为11,期望值(E)为10,截止值为100,M=5,N=-4以及两条链的比较。对于氨基酸序列,BLASTP程序使用以下作为默认值:字长(W)为3,期望值(E)为10,以及BLOSUM62打分矩阵(参见Henikoff&Henikoff, Proc.Natl.Acad.Sci.USA 89:10915(1989))。

[0098] 除了计算序列同一性百分比之外,BLAST算法还可进行两条序列之间相似性的统

计析(参见例如Karlin&Altschul, Proc. Nat'l. Acad. Sci. USA 90:5873-5787 (1993))。BLAST算法提供的相似性的一种量度是最小总和概率(P(N)),其提供了两条核苷酸或氨基酸序列之间的匹配偶然发生的概率的指示。例如,如果测试核苷酸序列与参照核苷酸序列的比较中的最小总和概率小于约0.1至小于约0.001,则认为测试核酸序列与参照序列相似。因此,在本发明的一些实施方案中,测试核苷酸序列与参照核苷酸序列比较中的最小总和概率小于约0.001。

[0099] 当两条核苷酸序列在严格条件下彼此杂交时,也可认为这两条序列实质上互补。在一些代表性实施方案中,视为实质上互补的两条核苷酸序列在高严格条件下彼此杂交。

[0100] 在核酸杂交实验诸如DNA印迹和RNA印迹杂交的语境中,“严格杂交条件”和“严格杂交洗涤条件”是序列依赖性的,并且在不同的环境参数下是不同的。核酸杂交的广泛指导见于:Tijssen Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes,第1部分第2章“Overview of principles of hybridization and the strategy of nucleic acid probe assays”Elsevier, New York (1993)。通常,将高严格的杂交和洗涤条件选择为比该具体序列在限定的离子强度和pH下的热解链点(T_m)低约5℃。

[0101] T_m 是50%的目标序列与完全匹配的探针杂交的温度(在限定的离子强度和pH下)。将非常严格条件选择为等于特定探针的 T_m 。在DNA或RNA印迹中具有多于100个互补残基的互补核苷酸序列在滤膜上杂交的严格杂交条件的实例是在42℃下具有1mg肝素的50%甲酰胺,其中杂交过夜进行。高严格的洗涤条件的实例是0.15M NaCl,在72℃下,持续约15分钟。严格洗涤条件的实例是0.2xSSC洗涤,在65℃下,持续15分钟(对于SSC缓冲液的描述,参见Sambrook(出处同上))。经常,高严格洗涤之前是低严格洗涤以除去背景探针信号。对于双链体(例如超过100个核苷酸的双链体)的中等严格洗涤的实例是1x SSC,在45℃下,持续15分钟。对于双链体(例如超过100个核苷酸的双链体)的低严格洗涤的实例是4-6x SSC,在40℃下,持续15分钟。对于短探针(例如,约10至50个核苷酸),严格条件通常涉及在pH 7.0至8.3的小于约1.0M Na离子、通常约0.01至1.0M Na离子浓度(或其它盐)的盐浓度,且温度通常为至少约30℃。严格条件也可以通过添加去稳定剂诸如甲酰胺来实现。通常,在特定杂交测定中信噪比是不相关探针信噪比观测值的2倍(或更高倍)指示检测到特异性杂交。在严格条件下不彼此杂交的核苷酸序列,如果它们编码的蛋白质实质上相同,则仍然实质上相同。这可以例如当使用遗传密码允许的最大密码子简并性来产生核苷酸序列的拷贝时发生。

[0102] 以下是可用于克隆与本发明的参照核苷酸序列实质上相同的同源核苷酸序列的杂交/洗涤条件的集合的实例。在一个实施方案中,参照核苷酸序列与“测试”核苷酸序列在以下条件下杂交:50℃下在7%十二烷基硫酸钠(SDS)、0.5M NaPO₄、1mM EDTA中,洗涤条件为50℃下在2X SSC、0.1%SDS中。在另一个实施方案中,参照核苷酸序列与“测试”核苷酸序列在以下条件下杂交:50℃下在7%十二烷基硫酸钠(SDS)、0.5M NaPO₄、1mM EDTA中(洗涤条件为50℃下在1X SSC、0.1%SDS中);或者50℃下在7%十二烷基硫酸钠(SDS)、0.5M NaPO₄、1mM EDTA中(洗涤条件为50℃下在0.5X SSC、0.1%SDS中)。在另外的实施方案中,参照核苷酸序列与“测试”核苷酸序列在以下条件下杂交:50℃下在7%十二烷基硫酸钠(SDS)、0.5M NaPO₄、1mM EDTA中(洗涤条件为50℃下在0.1X SSC、0.1%SDS中);或者50℃下

在7%十二烷基硫酸钠(SDS)、0.5M NaPO₄、1mM EDTA中(洗涤条件为65℃下在0.1X SSC、0.1%SDS中)。

[0103] 可对本发明的任何核苷酸序列和/或异源核酸构建体进行密码子优化用于在任何所关注的物种中表达。密码子优化是本领域众所周知的,并且涉及使用物种特异性密码子使用表针对密码子使用偏好修饰核苷酸序列。基于对所关注物种的最高表达基因进行的序列分析产生密码子使用表。当在核中表达核苷酸序列时,基于对所关注物种的高度表达的核基因的序列分析产生密码子使用表。通过将物种特异性密码子使用表与天然多核苷酸序列中存在的密码子相比较来确定对核苷酸序列的修饰。如本领域中所理解的,核苷酸序列的密码子优化导致核苷酸序列与天然核苷酸序列具有小于100%的同一性(如,70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%等等),但仍编码与初始的天然核苷酸序列编码的多肽具有相同功能的多肽。因此,在本发明的代表性实施方案中,可对本发明的核苷酸序列和/或异源核酸构建体进行密码子优化用于在任何所关注的特定物种中表达。

[0104] 在一些实施方案中,本发明的异源或重组核酸分子、核苷酸序列和/或多肽是“分离的”。“分离的”核酸分子、“分离的”核苷酸序列或“分离的”多肽是这样的核酸分子、核苷酸序列或多肽,其通过人工脱离其天然环境存在,因此不是自然的产物。分离的核酸分子、核苷酸序列或多肽可以以纯化形式存在,其至少部分地与天然存在的生物体或病毒的至少一些其它组分(例如通常发现与所述多核苷酸相关联的细胞或病毒结构组分或其它多肽或核酸)分离。在代表性实施方案中,分离的核酸分子、分离的核苷酸序列和/或分离的多肽的纯度为至少约1%、5%、10%、20%、30%、40%、50%、60%、70%、80%、90%、95%或更高。

[0105] 在其它实施方案中,分离的核酸分子、核苷酸序列或多肽可存在于非天然环境中,诸如例如重组宿主细胞中。因此,例如,关于核苷酸序列,术语“分离的”意指其与其天然存在于其中的染色体和/或细胞是分离的。如果多核苷酸与其天然存在于其中的染色体和/或细胞分离并然后插入遗传背景、与其不天然存在于其中的染色体和/或细胞(例如,不同的宿主细胞,不同的调节序列和/或基因组中与在自然界中发现的不同位置),则其也是分离的。因此,异源核酸构建体、核苷酸序列及其编码的多肽是“分离的”,因为通过人工,它们脱离其天然环境存在,因此不是自然产物,然而,在一些实施方案中,可将它们引入并存在于重组宿主细胞中。

[0106] 在一些实施方案中,本发明的异源或重组核酸构建体是“合成的”。“合成的”核酸分子、“合成的”核苷酸序列或“合成的”多肽是自然界中不存在而是通过人工生成并因而不是自然产物的核酸分子、核苷酸序列或多肽。

[0107] 在任何本文所述实施方案中,本发明的核苷酸序列和/或异源核酸构建体可以与用于在多种生物体细胞中表达的各种启动子和其它调节元件有效关联。因此,在代表性实施方案中,本发明的核酸构建体可进一步包含与一个或多个核苷酸序列有效连接的一个或多个启动子。

[0108] 如本文所用,所谓“有效连接”或“有效关联”意指所示元件在功能上彼此相关,并且通常在物理上也是相关的。因此,本文所用的术语“有效连接”或“有效关联”是指单一核酸分子上功能上相关联的核苷酸序列。因此,有效连接至第二核苷酸序列的第一核苷酸序

列意指当第一核苷酸序列与第二核苷酸序列处于功能关系时的情况。例如,如果启动子实现所述核苷酸序列的转录或表达,则启动子与核苷酸序列有效关联。本领域技术人员将理解,控制序列(例如启动子)不需要与其有效关联的核苷酸序列相邻,只要控制序列发挥功能以引导其表达即可。因此,例如,间插的非翻译、但转录的序列可以存在于启动子与核苷酸序列之间,并且该启动子仍然可视为与该核苷酸序列“有效连接”。

[0109] “启动子”是控制或调节与启动子有效关联的核苷酸序列(即,编码序列)的转录的核苷酸序列。编码序列可以编码多肽和/或功能性RNA。通常,“启动子”是指含有RNA聚合酶II的结合位点并指导转录起始的核苷酸序列。一般而言,相对于相应编码序列的编码区的起点,启动子存在于5'端或上游。启动子区域可以包含充当基因表达的调节子的其它元件。这些包括TATA框共有序列以及通常CAAT框共有序列(Breathnach和Chambon, (1981) *Annu.Rev.Biochem.* 50:349)。在植物中,CAAT盒可以被AGGA盒替代(Messing等人, (1983) in *Genetic Engineering of Plants*, T.Kosuge, C.Meredith和A.Hollaender (编辑), Plenum Press, 第211-227页)。

[0110] 启动子可以包括例如组成型、诱导型、时序调节型、发育调节型、化学调节型、组织优选型和/或组织特异型启动子,用于制备异源核酸构建体(即,“嵌合基因”或“嵌合多核苷酸”)。这些各种类型的启动子是本领域已知的。

[0111] 启动子的选择将根据表达的时间和空间需求以及还根据待转化的宿主细胞而变化。用于许多不同生物体的启动子是本领域众所周知的。基于本领域存在的广泛知识,可以针对所关注的特定宿主生物体选择适当的启动子。因此,例如,关于模型生物体中高度组成型表达的基因上游的启动子了解很多,并且这样的知识可以容易获得并视情况而定在其它系统中实施。

[0112] 在一些实施方案中,本发明的核酸构建体可以是“表达盒”或可以包含于表达盒内。如本文所用,“表达盒”意指包含所关注核苷酸序列的异源核酸构建体(如,本发明的核酸构建体(如,合成的tracr核酸构建体、合成的CRISPR核酸构建体、合成的CRISPR阵列、嵌合的核酸构建体;编码所关注的多肽、I型多肽、II型多肽、III型多肽、IV型多肽和/或V型多肽的核苷酸序列)),其中所述核苷酸序列与至少一个控制序列(如,启动子)有效关联。因此,本发明的一些方面提供了设计用于表达本发明的核苷酸序列的表达盒。

[0113] 包含所关注的核苷酸序列的表达盒可以是嵌合的,意指其组分中的至少一种相对于其至少一种其它组分是异源性的。表达盒也可以是天然存在、但已经以可用于异源表达的重组形式获得的表达盒。

[0114] 表达盒还可以任选包含在所选的宿主细胞中有功能的转录和/或翻译终止区(即,终止区)。各种转录终止子可供用于表达盒中,并且负责终止超出所关注的异源核苷酸序列的转录以及正确mRNA多聚腺苷酸化。终止区对于转录起始区可以是天然的,对于有效连接的所关注核苷酸序列可以是天然的,对于宿主细胞可以是天然的,或者可以源自另一来源(即,对于启动子、所关注的核苷酸序列、宿主或它们的任何组合是外来的或异源的)。

[0115] 表达盒还可以包含用于选择标记的核苷酸序列,所述选择标记可以用于选择转化的宿主细胞。如本文所用,“选择标记”意指一种核苷酸序列,其在表达时赋予表达该标记的宿主细胞以不同的表型,因此允许将这样的转化细胞与不具有该标记的细胞区分开。这样的核苷酸序列可以编码可选择或可筛选的标记,这取决于该标记是否赋予可以通过化学手

段,诸如通过使用选择剂(如抗生素等)选择的性状,或者取决于该标记是否仅仅是可以通过观察或测试(诸如通过筛选(例如,荧光))鉴别的性状。当然,合适的选择标记的许多实例是本领域已知的,并且可以用于本文所述的表达盒。

[0116] 除了表达盒以外,本文所述的核酸分子和核苷酸序列可以与载体结合使用。术语“载体”是指用于将一个核酸(或多个核酸)转移、递送或引入细胞中的组合物。载体包含具有待转移、递送或引入的核苷酸序列的核酸分子。用于转化宿主生物体的载体是本领域众所周知的。一般类别的载体的非限制性实例包括但不限于病毒载体、质粒载体、噬菌体载体、噬菌粒载体、粘粒载体、福斯质粒载体、噬菌体、人工染色体或农杆菌二元载体,其为双链或单链线性或环状形式,其可以是或可以不是自我转移的或可移动的。本文所定义的载体可以通过整合进细胞基因组或染色体外存在(例如,具有复制起点的自主复制质粒)来转化原核或真核宿主。另外包括的是穿梭载体,所谓穿梭载体意指天然或通过设计能够在两种不同宿主生物体中复制的DNA载体,所述宿主生物体可以选自放线菌及相关物种、细菌和真核生物(例如,高等植物、哺乳动物、酵母或真菌细胞)。在一些代表性实施方案中,载体中的核酸处于适当的启动子或用于在宿主细胞中转录的其它调节元件的控制下并与之有效连接。载体可以是在多种宿主中发挥功能的双功能表达载体。在基因组DNA的情况下,其可以含有其自身启动子或其它调节元件,并且在cDNA的情况下,其可以在适当的启动子或用于在宿主细胞中表达的其它调节元件的控制下。因此,本发明的核酸分子和/或表达盒可以包含于本文所述和本领域已知的载体中。

[0117] 在所关注的多核苷酸的语境中,“引入”(及其语法变型形式)意指将所关注的多核苷酸递呈至宿主生物体或所述生物体的细胞(例如,宿主细胞),其方式使得该多核苷酸进入细胞内部。若要引入不止一个多核苷酸,则可将这些多核苷酸组装为单一多核苷酸或核酸构建体的部分,或组装为独立的多核苷酸或核酸构建体,并且可位于相同或不同的表达构建体或转化载体上。因此,这些多核苷酸可以在单个转化事件中、在分开的转化/转染事件中引入细胞,或者例如,可以通过常规育种方案将它们掺入生物体中。因此,在一些方面,可以将本发明的一种或多种核酸构建体单个地或组合地引入宿主生物体或所述宿主生物体的细胞中。在细胞群体的语境中,“引入”意指使该群体与本发明的异源核酸构建体在一定条件下接触,在所述条件下本发明的异源核酸构建体进入该群体的一个或多个细胞的内部,由此转化该群体的所述一个或多个细胞。

[0118] 本文所用的术语“转化”或“转染”是指将异源核酸引入细胞中。细胞的转化可以是稳定的或瞬时的。因此,在一些实施方案中,用本发明的核酸构建体稳定转化宿主细胞或宿主生物体。在其它实施方案中,用本发明的核酸构建体瞬时转化宿主细胞或宿主生物体。因而,在代表性的实施方案中,可将本发明的异源核酸构建体稳定和/或瞬时地引入细胞中。

[0119] 在多核苷酸的语境中,“瞬时转化”意指多核苷酸被引入细胞中并且不整合至细胞的基因组中。

[0120] 在多核苷酸的语境中,所谓“稳定引入”或“被稳定地引入”意指所引入的多核苷酸稳定地整合至细胞的基因组中,因此该细胞稳定地转化有该多核苷酸。

[0121] 如本文所用,“稳定转化”或“稳定转化的”意指核酸构建体被引入细胞并整合至细胞的基因组中。因此,整合的核酸构建体能够被其子代,更具体地,被多个连续世代的子代遗传。如本文所用,“基因组”还包括核、质粒和质粒基因组,因此可包括将核酸构建体整合

进例如质粒或线粒体基因组中。本文所用的稳定转化也可以指保持在染色体外,例如作为微型染色体或质粒的转基因。

[0122] 瞬时转化可以通过例如酶联免疫吸附测定 (ELISA) 或蛋白质印迹 (其可以检测由引入生物体中的一种或多种转基因编码的肽或多肽的存在) 检测。细胞的稳定转化可以通过例如, 用与引入生物体 (如细菌、古细菌、酵母、藻类等) 的转基因的核苷酸序列特异性杂交的核酸序列对所述细胞的基因组DNA进行DNA印迹杂交测定法来检测。细胞的稳定转化可以通过例如, 用与引入植物或其它生物体的转基因的核苷酸序列特异性杂交的核酸序列对所述细胞的RNA进行RNA印迹杂交测定法来检测。细胞的稳定转化还可以通过例如以下方式检测: 采用与转基因的靶标序列杂交的特异性引物序列进行聚合酶链式反应 (PCR) 或本领域众所周知的其它扩增反应, 从而导致转基因序列的扩增, 可以根据标准方法检测。转化也可以通过本领域众所周知的直接测序和/或杂交方案来检测。

[0123] 因此, 在一些实施方案中, 核苷酸序列、构建体、表达盒可以瞬时表达和/或它们可以稳定地掺入宿主生物体的基因组中。

[0124] 可以通过本领域技术人员已知的任何方法将本发明的异源核酸构建体引入细胞中。在本发明的一些实施方案中, 细胞的转化包括核转化。在其它实施方案中, 本发明的异源核酸构建体可通过常规育种技术引入细胞。

[0125] 用于转化真核生物和原核生物两者的程序是本领域众所周知和常规的, 并且在整个领域中有描述 (参见例如, Jiang等人, 2013. *Nat. Biotechnol.* 31:233-239; Ran等人, *Nature Protocols* 8:2281-2308 (2013))。

[0126] 因此, 可以以本领域众所周知的多种方式将核苷酸序列引入宿主生物体或其细胞中。本发明的方法不依赖于将一种或多种核苷酸序列引入生物体的具体方法, 只要它们进入生物体的至少一个细胞的内部即可。若要引入不止一种核苷酸序列, 则可将它们组装为单一核酸构建体的部分, 或组装为独立的核酸构建体, 并且可位于相同或不同的核酸构建体上。因此, 可以在单个转化事件中或在分开的转化事件中将核苷酸序列引入所关注的细胞中, 或者备选地, 若相关, 可以作为育种方案的一部分将核苷酸序列掺入生物体中。

[0127] 可移动遗传元件 (MGE) 给细菌带来基因组稳定性的持续挑战, 通过水平基因转移促进进化。术语MGE涵盖质粒、噬菌体、可转座元件、基因组岛以及许多其它特化遗传元件 (1)。MGE涵盖赋予高播散速率、对宿主的适应优势以及基因组稳定性, 从而导致它们在细菌基因组中几乎普遍存在的基因。为了应付掠夺性噬菌体和自私遗传元件的持久威胁, 细菌已经进化出靶向外源遗传元件的先天免疫系统和获得性免疫系统。先天免疫包括细胞壁修饰、限制/修饰系统和非生产性噬菌体感染 (2)。成簇的规律间隔短回文重复 (CRISPR) 和CRISPR相关基因 (Cas) 是细菌中靶向侵入的遗传元件的获得性免疫系统 (3)。CRISPR-Cas介导的免疫力依赖于截然不同的分子过程, 这些分子过程被分为获取 (acquisition)、表达 (expression) 和干扰 (interference) (3)。获取通过对外来遗传元件的分子“采样”进行, 来自所述外来遗传元件的短序列 (称为间隔区) 以极化的方式被整合进CRISPR阵列中 (4)。CRISPR阵列的表达是组成型的并且可由前面的前导序列内的启动子元件诱导 (5-6)。干扰由相应的转录物引起, 所述转录物在每个重复序列处经选择性加工, 从而形成CRISPR RNA (crRNA), crRNA引导Cas蛋白对与所述间隔区互补的靶标DNA进行序列特异性识别和切割 (7)。CRISPR-Cas技术已经应用于株系分型和检测 (8-10)、针对可移动遗传元件的天然/工

程化免疫力的利用(11)、在多样背景中进行可编程基因组编辑(12)、转录控制(13-14)和确定复合系中微生物群体的操纵(15)。

[0128] 多种CRISPR系统是本领域已知的。例如,参见Makarova等人,其描述了CRISPR系统的所有类别、类型和亚型的命名(Nature Reviews Microbiology 13:722-736(2015));还可参见R.Barrangou(Genome Biol.16:247(2015))。

[0129] 尽管先前描述了多种生物体中对应于CRISPR阵列的序列特征(16-17),但嗜热链球菌是其中特异性的cas基因和CRISPR阵列组分的作用被阐明的第一个微生物(4)。嗜热链球菌是一种非病原性嗜热革兰氏阳性细菌,其作为起始培养物用于酸奶和多种奶酪的互养生产中使乳糖分解代谢为乳酸(18)。嗜热链球菌编码多达四种CRISPR-Cas系统,其中两种(SthCRISPR1和SthCRISPR3)被分类为II-A型系统,它们先天在获取和干扰方面是有活性的(4,19)。因此,对嗜热链球菌及其噬菌体进行的基因组分析建立了CRISPR-Cas系统用于噬菌体/DNA保护的可能机制。对嗜热链球菌中CRISPR-Cas系统的研究导致了对间隔区起源的生物信息分析(4,20)、原间隔区相邻基序(PAM)序列的发现(19;21)、噬菌体-宿主动力学的理解(22-23)、Cas9内切核酸酶活性的证明(7,24-25),以及最近地,支配II型系统的功能和正交性的tracrRNA结构基序的确定(26)。嗜热链球菌的基因组分析揭示了通过丧失致病性链球菌中存在的糖分解代谢基因和毒力基因实现的对牛奶的进化适应(18)。嗜热链球菌还经历大量获取小生境相关基因,诸如编码包括冷休克蛋白、抗铜蛋白、蛋白酶、细菌素和乳糖分解代谢蛋白的那些(18)。插入序列(IS)在嗜热链球菌基因组中非常普遍并且通过促进与乳制品适应基因相关的基因岛的播散而促成菌株之间的遗传异质性(18)。嗜热链球菌中同时存在MGE和功能性CRISPR-Cas系统表明,基因组稳态至少部分由这些动态力的相互作用支配。因而,嗜热链球菌构成了用于研究基因组岛的CRISPR-Cas靶向的遗传结果的理想宿主。

[0130] CRISPR-Cas系统最近已成为大量基因组编辑应用方面的研究的主题(12),但是大多数内源微生物系统的进化作用仍然未知(27)。关于拥有活性CRISPR-Cas系统超出防止外来DNA摄入以外的进化结果(7)、间隔区获取事件(4)和染色体自我靶向引起的突变(28-32)知之甚少。因而,本发明人寻求确定用内源II型CRISPR-Cas系统靶向整合的MGE的结果。在嗜热链球菌LMD-9中鉴定了四个基因组岛,长度范围为8至102kbp,总计大约132kbp,或总计基因组的7%。为了靶向基因组岛,将具有自我靶向间隔区的工程化CRISPR阵列的基于质粒的表达转化进嗜热链球菌LMD-9中。总体而言,我们的结果阐明了自我靶向事件的根本遗传结果,并显示CRISPR-Cas系统可在细菌群体水平上指导基因组进化。

[0131] 利用这些发现,本发明人已经开发了以下新方法:针对必需基因、非必需基因和/或可消耗基因组岛对细菌、古细菌、藻类或酵母细胞的群体进行筛选;杀灭细菌、古细菌、藻类或酵母细胞群体内的一种或多种细胞;鉴定细菌、古细菌、藻类或酵母基因的表型;从细菌、古细菌或酵母细胞群体中选择基因组大小减小的一种或多种细菌、古细菌、藻类或酵母细胞;和/或在细菌、古细菌、藻类或酵母群体中鉴定在其基因组中具有突变(缺失)的至少一种分离株。

[0132] 因而,在一个方面,本发明人已经开发了在群体中鉴定遗传变体的方法,该遗传变体具有改变的遗传内容,其使得它们能够避开靶向。这里,靶标序列已经被修饰,并且寻找具有该修饰的存活者。在一些方面,该修饰(即突变)是缺失。此外,如果靶标序列已经被修

饰,则该野生型基因型不是必需的。

[0133] 因此,在本发明的一个方面中,提供了一种针对必需基因、非必需基因和/或可消耗基因组岛对细菌细胞群体进行筛选的方法,包括:向所述细菌细胞群体中引入异源核酸构建体,所述构建体包含CRISPR阵列,所述CRISPR阵列包含(5'至3')重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述重复-间隔区-重复序列或所述至少一个重复-间隔区序列的间隔区包含与所述群体中的细菌细胞的基因组中的靶标区实质上互补的核苷酸序列,从而产生转化细菌细胞的群体;以及确定所述转化细菌细胞的群体中是否存在缺失,其中所述转化的细菌细胞的群体中存在缺失表明所述靶标区包含在非必需基因和/或可消耗基因组岛内,而所述群体中不存在缺失意味着所述靶标区包含在必需基因内。可用于本发明的CRISPR阵列可以是I型、II型、III型、IV型或V型CRISPR阵列。

[0134] 在另外的方面,本发明提供了一种针对必需基因、非必需基因和/或可消耗基因组岛对细菌、古细菌、藻类或酵母细胞群体进行筛选的方法,包括:向所述细菌、古细菌、藻类或酵母细胞群体中引入:(a)包含反式激活CRISPR(tracr)核酸的异源核酸构建体,(b)异源核酸构建体,所述构建体包含CRISPR阵列(crRNA、crDNA),所述CRISPR阵列包含(5'至3')重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述间隔区包含与所述群体中的细菌、古细菌、藻类或酵母细胞的基因组中的靶标区实质上互补的核苷酸序列,和(c)Cas9多肽或包含编码Cas9多肽的多核苷酸的异源核酸构建体,从而产生转化的细菌、古细菌、藻类或酵母细胞的群体;以及确定所述转化的细菌、古细菌、藻类或酵母细胞的群体中是否存在缺失,其中所述转化的细菌、古细菌或酵母细胞的群体中存在缺失意味着所述靶标区包含在非必需基因和/或可消耗基因组岛内,而所述转化的细菌、古细菌、藻类或酵母细胞的群体中不存在缺失意味着所述靶标区包含在必需基因内。

[0135] 在其它方面,提供了一种杀灭细菌细胞群体内的一种或多种细菌细胞的方法,包括:向所述细菌细胞群体中引入异源核酸构建体,所述构建体包含CRISPR阵列(crRNA、crDNA),所述CRISPR阵列包含重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述重复-间隔区-重复序列或至少一个重复-间隔区序列的间隔区包含与所述群体中的细菌细胞的基因组中的靶标区实质上互补的核苷酸序列,从而杀灭所述群体内包含所述靶标区的一种或多种细菌细胞。可用于本发明的CRISPR阵列可以是I型、II型、III型、IV型或V型CRISPR阵列。

[0136] 在另一个方面,提供了一种杀灭细菌、古细菌、藻类或酵母细胞群体内的一种或多种细胞的方法,该方法包括:向所述细菌、古细菌或酵母细胞群体中引入(a)包含反式激活CRISPR(tracr)核酸的异源核酸构建体,(b)异源核酸构建体,所述构建体包含CRISPR阵列(crRNA、crDNA),所述CRISPR阵列包含重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述至少一个重复-间隔区序列和重复-间隔区-重复序列的间隔区包含与所述群体中的细菌、古细菌、藻类或酵母细胞的基因组中的靶标区实质上互补的核苷酸序列,和(c)Cas9多肽和/或包含编码Cas9多肽的多核苷酸的异源核酸构建体,从而杀灭所述细菌、古细菌、藻类或酵母细胞群体内在其基因组中包含所述靶标区的一种或多种细胞。

[0137] 靶向细菌基因组的CRISPR RNA的转化可用于基于序列特异性而选择性杀灭细菌细胞以减除在遗传上不同的亚群,从而富集缺少靶标序列的细菌群体。这种差别可基于直系CRISPR-Cas系统在遗传相似的群体内的异质分布而出现。因而,在一些实施方案中,引入

进细胞群体中的CRISPR阵列可以与待杀灭的所述一种或多种细菌细胞中的CRISPR-Cas系统相容(即具有功能性),但与该群体内的至少一种或多种细菌细胞的CRISPR Cas系统不相容(即,不具功能性)。例如,大肠杆菌和肺炎克雷伯菌(*Klebsiella pneumoniae*)可展示I-E型或I-F型CRISPR-Cas系统;艰难梭状芽胞杆菌(*Clostridium difficile*)编码I-B型系统,而嗜热链球菌的不同菌株展示出II-A型系统和I-E型系统二者或仅展示出II-A型系统。取决于转化进细菌混合物中的具体CRISPR RNA,其可基于与其同源系统的功能相容性而特异性地靶向该群体的亚组。这可适用于含有内源CRISPR-Cas系统的各种各样的物种,例如但不限于:假单胞菌属物种(诸如:绿脓假单胞菌)、埃希杆菌属物种(诸如:大肠杆菌)、肠杆菌属物种(诸如:阴沟肠杆菌)、葡萄球菌属物种(诸如:金黄色葡萄球菌)、肠球菌属物种(诸如:粪肠球菌、屎肠球菌)、链霉菌属物种(诸如:索马里链霉菌)、链球菌属物种(诸如:酿脓链球菌)、弧菌属物种(诸如:霍乱弧菌)、耶尔森菌属物种(诸如:鼠疫耶尔森菌)、弗朗西斯菌属物种(诸如:土拉弗朗西斯菌、新凶手弗朗西斯菌)、芽孢杆菌属物种(诸如:炭疽芽孢杆菌、蜡样芽孢杆菌)、乳杆菌属物种(诸如:干酪乳杆菌、罗伊乳杆菌、嗜酸乳杆菌、鼠李糖乳杆菌)、伯克霍德菌属物种(诸如:鼻疽伯克霍德菌、类鼻疽伯克霍德菌)、克雷伯杆菌属物种(诸如:肺炎克雷伯杆菌)、志贺菌属物种(诸如:痢疾志贺菌、宋内志贺菌)、沙门菌属物种(诸如:肠道沙门菌)、疏螺旋体属物种(诸如:博氏疏螺旋体)、奈瑟菌属物种(诸如:脑膜炎奈瑟菌)、梭杆菌属物种(诸如:具核梭杆菌)、螺杆菌属物种(诸如:幽门螺杆菌)、衣原体属物种(诸如:沙眼衣原体)、拟杆菌属物种(诸如:脆弱拟杆菌)、巴尔通氏体属物种(诸如:五日巴尔通氏体)、博德特氏菌属物种(诸如:百日咳博德特氏菌)、布鲁氏杆菌属物种(诸如:流产布鲁氏杆菌)、弯曲杆菌属物种(诸如:空肠弯曲杆菌)、梭菌属物种(诸如:艰难梭菌)、双歧杆菌属物种(诸如:婴儿双歧杆菌)、嗜血杆菌属物种(诸如:流感嗜血杆菌)、李斯特菌属物种(诸如:单核细胞增多性李斯特菌)、军团杆菌属物种(诸如:嗜肺性军团病杆菌)、分枝杆菌属物种(诸如:结核分枝杆菌)、支原体属物种(诸如:肺炎支原体)、立克次体属物种(诸如:立克次体)、不动杆菌属物种(诸如:醋酸钙不动杆菌、鲍氏不动杆菌)、瘤胃球菌属物种(诸如:白色瘤胃球菌)、丙酸杆菌属物种(诸如:费氏丙酸杆菌)、棒状杆菌属物种(诸如:白喉棒状杆菌)、丙酸杆菌属物种(诸如:痤疮丙酸杆菌)、短杆菌属物种(诸如:碘短杆菌)、微球菌属物种(诸如:藤黄微球菌)和/或普氏菌属物种(诸如:组织普氏菌)。

[0138] CRISPR靶向可基于混合群体中的不同遗传内容而移除特定的细菌亚组。该论点的支持在实施例4、5中给出,其中从群体选择了Lac⁻细菌而移除了Lac⁺。Lac⁺和Lac⁻菌株之间的遗传差别在实施例8和10中给出,其中对存活克隆的测序表明,相比于参照野生型嗜热链球菌菌株,其遗传内容有高达5.5%的差别。CRISPR-靶向间隔区可因而通过靶向保守的或趋异的遗传序列来调节各种细菌相关性水平。因而,在一些实施方案中,群体中的细菌细胞可包含相同的CRISPR Cas系统,因而所引入的CRISPR阵列可作为整体在细菌群体中发挥功能,但是构成该细菌群体的不同菌株或物种的遗传内容足够不同而使得所引入的CRISPR阵列的靶标区仅存在于所述群体中要被杀灭的所述一种或多种细菌物种中。这可适用于含有内源CRISPR-Cas系统的各种各样的物种,例如但不限于:假单胞菌属物种(诸如:绿脓假单胞菌)、埃希杆菌属物种(诸如:大肠杆菌)、肠杆菌属物种(诸如:阴沟肠杆菌)、葡萄球菌属物种(诸如:金黄色葡萄球菌)、肠球菌属物种(诸如:粪肠球菌、屎肠球菌)、链霉菌属物种(诸如:索马里链霉菌)、链球菌属物种(诸如:酿脓链球菌)、弧菌属物种(诸如:霍乱弧菌)、

耶尔森菌属物种(诸如:鼠疫耶尔森菌)、弗朗西斯菌属物种(诸如:土拉弗朗西斯菌、新凶手弗朗西斯菌)、芽孢杆菌属物种(诸如:炭疽芽孢杆菌、蜡样芽孢杆菌)、乳杆菌属物种(诸如:干酪乳杆菌、罗伊乳杆菌、嗜酸乳杆菌、鼠李糖乳杆菌)、伯克霍德菌属物种(诸如:鼻疽伯克霍德菌、类鼻疽伯克霍德菌)、克雷伯杆菌属物种(诸如:肺炎克雷伯杆菌)、志贺菌属物种(诸如:痢疾志贺菌、宋内志贺菌)、沙门菌属物种(诸如:肠道沙门菌)、疏螺旋体属物种(诸如:博氏疏螺旋体)、奈瑟菌属物种(诸如:脑膜炎奈瑟菌)、梭杆菌属物种(诸如:具核梭杆菌)、螺杆菌属物种(诸如:幽门螺杆菌)、衣原体属物种(诸如:沙眼衣原体)、拟杆菌属物种(诸如:脆弱拟杆菌)、巴尔通氏体属物种(诸如:五日巴尔通氏体)、博德特氏菌属物种(诸如:百日咳博德特氏菌)、布鲁氏杆菌属物种(诸如:流产布鲁氏杆菌)、弯曲杆菌属物种(诸如:空肠弯曲杆菌)、梭菌属物种(诸如:艰难梭菌)、双歧杆菌属物种(诸如:婴儿双歧杆菌)、嗜血杆菌属物种(诸如:流感嗜血杆菌)、李斯特菌属物种(诸如:单核细胞增多性李斯特菌)、军团杆菌属物种(诸如:嗜肺性军团病杆菌)、分枝杆菌属物种(诸如:结核分枝杆菌)、支原体属物种(诸如:肺炎支原体)、立克次体属物种(诸如:立克次体)、不动杆菌属物种(诸如:醋酸钙不动杆菌、鲍氏不动杆菌)、瘤胃球菌属物种(诸如:白色瘤胃球菌)、丙酸杆菌属物种(诸如:费氏丙酸杆菌)、棒状杆菌属物种(诸如:白喉棒状杆菌)、丙酸杆菌属物种(诸如:痤疮丙酸杆菌)、短杆菌属物种(诸如:碘短杆菌)、微球菌属物种(诸如:藤黄微球菌)和/或普氏菌属物种(诸如:组织普氏菌)。

[0139] 使用本发明方法在群体中的杀灭程度可受特定群体对转化的顺从性以及靶标区是包含在非必需基因、必需基因还是可消耗基因组岛中的影响。细菌、古细菌或酵母细胞群体中的杀灭程度可例如因生物体、因属和种而变化。因此,如本文所用的,“杀灭”意指消除群体中2个对数或更多的细胞(1%的存活率或更低)。低于1个对数的杀灭是群体的少量减少;而2-3个对数的杀灭导致群体显著减少;超过3个对数的杀灭表明群体已实质上被根除。

[0140] 在另一个方面,提供了一种鉴定与细菌基因相关联的表型的方法,包括:向细菌细胞群体中引入异源核酸构建体,所述构建体包含CRISPR阵列(crRNA、crDNA),所述CRISPR阵列包含重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述至少一个重复-间隔区序列和所述重复-间隔区-重复序列的间隔区包含与所述群体中的细菌细胞的基因组中的靶标区实质上互补的核苷酸序列,其中所述靶标区包含编码多肽或功能性核酸的开放阅读框的至少一部分,从而杀灭包含所述靶标区的细胞并产生不含所述靶标区的转化细菌细胞的群体(即,存活细胞不包含所述靶标区);以及(i)分析所述细胞群体的表型,或者(ii)从所述转化细菌细胞的群体培养独立的细菌集落并分析所述独立集落的表型。可用于本发明的CRISPR阵列可以是I型、II型、III型、IV型或V型CRISPR阵列。

[0141] 在另一个方面,提供了一种鉴定细菌、古细菌、藻类或酵母基因的表型的方法,包括:向细菌、古细菌、藻类或酵母细胞群体中引入(a)包含反式激活CRISPR(tracr)核酸的异源核酸构建体,(b)异源核酸构建体,所述构建体包含CRISPR阵列(crRNA、crDNA),所述CRISPR阵列包含(5'至3')重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述间隔区包含与所述群体中的细菌、古细菌、藻类或酵母细胞的基因组中的靶标区实质上互补的核苷酸序列,和(c)Cas9多肽和/或包含编码Cas9多肽的多核苷酸的异源核酸构建体,从而杀灭包含所述靶标区的细菌、古细菌、藻类或酵母细胞并产生不含所述靶标区的转化的细菌、古细菌、藻类或酵母细胞的群体;以及(i)分析所述细胞群体的表型,和/或(ii)从

所述转化的细菌、古细菌、藻类或酵母细胞的群体培养独立细菌、古细菌或酵母集落；以及分析所述独立集落的表型。

[0142] 在一些实施方案中，所述分析包括用于鉴定和表征突变的PCR、光学基因组作图、基因组测序、限制性酶切作图和/或限制性酶切分析以及用于分析表型的互补分析和/或表型测定法。

[0143] 在本发明的一些实施方案中，测定群体中的杀灭或减少程度包括用于测定群体数目的任何方法，包括但不限于：(1) 将细胞进行平板接种并对集落进行计数，(2) 光密度法，(3) 显微镜计数法，(4) 最可能数法和/或(5) 亚甲蓝还原法。在一些实施方案中，可将16S rDNA测序用于概要描述混合群体的组成。这可以例如通过以下方式进行：整体从样品纯化DNA，并使用高通量技术进行全基因组鸟枪测序，或者例如通过PCR扩增16S基因并以相同方式对产物进行测序。然后可将序列通过计算赋给某一细菌类群。在其它实施方案中，还可将定量PCR方法用于定量细菌水平。这类技术是本领域众所周知的。例如，可设计qPCR的引物来从共享所述序列的生物体的株系、种、属或类群特异性扩增。因而，可将阈次数(ct)用于定量所述生物体或生物体类群。在另外的实施方案中，也可将靶标群体特异性的任何细菌活性(表型)用作度量标准来确定群体的耗竭。

[0144] 在另外的实施方案中，提供了一种从细菌细胞群体中选择基因组大小减小的一种或多种细菌细胞的方法，包括：向细菌细胞群体中引入异源核酸构建体，所述构建体包含CRISPR阵列(crRNA、crDNA)，所述CRISPR阵列包含(5'至3')重复-间隔区-重复序列或至少一个重复-间隔区序列，其中所述间隔区包含与所述群体中的一种或多种细菌细胞的基因组中的靶标区实质上互补的核苷酸序列，其中杀灭包含所述靶标区的细胞，从而从所述细菌细胞群体中选择不含所述靶标区并且基因组大小减小的一种或多种细菌细胞。可用于本发明的CRISPR阵列可以是I型、II型、III型、IV型或V型CRISPR阵列。

[0145] 在一些实施方案中，提供了一种从细菌细胞群体中选择基因组大小减小的一种或多种细菌细胞的方法，包括：向细菌细胞群体中引入：(a) (i) 一种或多种异源核酸构建体，所述构建体包含与所述细菌细胞的基因组中存在的至少300个连续核苷酸具有至少80%同一性的核苷酸序列，或者(ii) 两种或更多种异源核酸构建体，所述构建体包含至少一个转座子，从而产生转基因细菌细胞的群体，所述转基因细菌细胞在整合进基因组中的所述一种或多种异源核酸构建体与基因组中存在的所述至少300个连续核苷酸之间、或整合进基因组中的第一转座子与第二转座子之间包含非天然的同源重组位点；和(b) 异源核酸构建体，所述构建体包含CRISPR阵列(crRNA、crDNA)，所述CRISPR阵列包含(5'至3')重复-间隔区-重复序列或至少一个重复-间隔区序列，其中所述重复-间隔区-重复序列或至少一个重复-间隔区序列的间隔区包含与所述群体中的一种或多种细菌细胞的基因组中的靶标区实质上互补的核苷酸序列，其中所述靶标区位于引入进基因组中的所述一种或多种异源核酸构建体与基因组中存在的所述至少300个连续核苷酸之间和/或所述第一转座子与所述第二转座子之间，其中杀灭包含所述靶标区的细胞，而不包含所述靶标区的细胞存活，从而从所述转基因细菌细胞群体中选择不含所述靶标区并且基因组大小减小的一种或多种细菌细胞。可用于本发明的CRISPR阵列可以是I型、II型、III型、IV型或V型CRISPR阵列。

[0146] 如本领域众所周知的，可经由例如PCR扩增或通过设计的DNA合成产生转座子，并可经由任何转化方法引入。

[0147] 在一些实施方案中,本发明提供了一种从细菌、古细菌、藻类或酵母细胞群体中选择基因组大小减小的一种或多种细菌、古细菌、藻类或酵母细胞的方法,包括:向细菌、古细菌或酵母细胞群体中引入(a)包含反式激活CRISPR(tracr)核酸的异源核酸构建体,(b)异源核酸构建体,所述构建体包含CRISPR阵列(crRNA、crDNA),所述CRISPR阵列包含重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述至少一个重复-间隔区序列和所述至少一个重复-间隔区-重复序列的间隔区包含与所述群体中的细菌、古细菌、藻类或酵母细胞的基因组中的靶标区实质上互补的核苷酸序列,和(c)Cas9多肽和/或包含编码Cas9多肽的多核苷酸的异源核酸构建体,其中杀灭包含所述靶标区的细胞,从而从所述细菌、古细菌、藻类或酵母细胞群体中选择不含所述靶标区并且基因组大小减小的一种或多种细菌、古细菌、藻类或酵母细胞。

[0148] 在其它实施方案中,提供了一种从细菌、古细菌、藻类或酵母细胞群体中选择基因组大小减小的一种或多种细菌、古细菌、藻类或酵母细胞的方法,包括:向细菌、古细菌、藻类或酵母细胞群体中引入:(a)(i)一种或多种异源核酸构建体,所述构建体包含与所述细菌、古细菌、藻类或酵母细胞的基因组中存在的至少300个连续核苷酸具有至少80%同一性的核苷酸序列,或者(ii)两种或更多种异源核酸构建体,所述构建体包含至少一个转座子,从而产生转基因细菌、古细菌、藻类或酵母细胞的群体,所述转基因细菌、古细菌、藻类或酵母细胞在整合进所述基因组中的所述一种或多种异源核酸构建体与所述基因组中存在的所述至少300个连续核苷酸之间、或整合进所述基因组中的第一转座子与第二转座子之间包含非天然的同源重组位点;和(b)(i)包含反式激活CRISPR(tracr)核酸的异源核酸构建体,(ii)异源核酸构建体,所述构建体包含CRISPR阵列(crRNA、crDNA),所述CRISPR阵列包含重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述至少一个重复-间隔区序列或所述至少一个重复-间隔区-重复序列的间隔区包含与所述群体中的一种或多种细菌、古细菌、藻类或酵母细胞的基因组中的靶标区实质上互补的核苷酸序列,和(iii)Cas9多肽和/或包含编码Cas9多肽的多核苷酸的异源核酸构建体,其中所述靶标区位于掺入进基因组中的所述一种或多种异源核酸构建体与基因组中存在的所述至少300个连续核苷酸之间和/或所述第一转座子与第二转座子之间,其中杀灭包含所述靶标区的细胞,而不包含所述靶标区的细胞存活,从而从所述转基因细菌、古细菌、藻类或酵母细胞群体中选择不含所述靶标区并且基因组大小减小的一种或多种细菌、古细菌、藻类或酵母细胞。

[0149] 在一些方面,所述减小的基因组大小可以是与对照相比减小。在一些方面中,对照可以是细菌、古细菌、藻类或酵母细胞的野生型群体,或者其可以用这样的异源构建体转化的细菌、古细菌、藻类或酵母细胞群体,该异源构建体包含CRISPR阵列(如,I型、II型、III型、IV型或V型CRISPR阵列),该CRISPR阵列包含重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述重复-间隔区-重复序列或所述至少一个重复-间隔区序列的间隔区包含的核苷酸序列不与所述群体的细菌、古细菌、藻类或酵母细胞的基因组中的靶标区互补(即,非自我靶向/“加扰间隔区”)。在另外的方面中,对照可以用异源构建体转化的细菌、古细菌、藻类或酵母细胞群体,该异源构建体包含CRISPR阵列,该CRISPR阵列包含重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述重复-间隔区-重复序列或所述至少一个重复-间隔区序列的间隔区包含的核苷酸序列与所述群体的细菌、古细菌、藻类或酵母细胞的基因组中的靶标区实质上互补但缺少原间隔区相邻基序(proto-spacer adjacent

motif,PAM)。

[0150] 在一些实施方案中,提供了一种在细菌群体中鉴定在其基因组中具有缺失(如染色体和/或质粒缺失)的至少一种分离株的方法,包括:向细菌细胞群体中引入异源核酸构建体,所述构建体包含CRISPR阵列(crRNA、crDNA),所述CRISPR阵列包含(5'至3')重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述重复-间隔区-重复序列或所述至少一个重复-间隔区序列的间隔区包含与所述群体中的一种或多种细菌细胞的基因组中的靶标区实质上互补的核苷酸序列,以及杀灭包含所述靶标区的细胞,从而产生不含所述靶标区的转化细菌细胞的群体;以及从所述转化细菌细胞的群体培养独立的细菌集落,从而从所述转化细菌的群体中鉴定在其基因组中具有缺失的至少一种分离株。可用于本发明的CRISPR阵列可以是I型、II型、III型、IV型或V型CRISPR阵列。

[0151] 在另外的实施方案中,提供了一种在细菌群体中鉴定在其基因组中具有缺失的至少一种分离株的方法,包括:向所述细菌细胞群体中引入:(a) (i) 一种或多种异源核酸构建体,所述构建体包含与所述细菌细胞的基因组中存在的至少300个连续核苷酸具有至少80%同一性的核苷酸序列,或者(ii) 两种或更多种异源核酸构建体,所述构建体包含至少一个转座子,从而产生转基因细菌细胞的群体,所述转基因细菌细胞在整合进所述基因组中的所述一种或多种异源核酸构建体与所述基因组中存在的所述至少300个连续核苷酸之间、或整合进所述基因组中的第一转座子与第二转座子之间包含非天然的同源重组位点;和b) 异源核酸构建体,所述构建体包含CRISPR阵列(crRNA、crDNA),所述CRISPR阵列包含重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述间隔区包含与所述群体中的一种或多种细菌细胞的基因组中的靶标区实质上互补的核苷酸序列,其中所述靶标区位于引入进所述基因组中的所述一种或多种异源核酸构建体与所述基因组中存在的所述至少300个连续核苷酸之间和/或所述第一转座子与第二转座子之间,以及杀灭包含所述靶标区的细胞[而不包含所述靶标区的细胞存活],从而产生不含所述靶标区的转化细菌细胞的群体;以及从所述转化细菌细胞的群体培养独立的细菌集落,从而从所述细菌的群体中鉴定在其基因组中具有缺失的至少一种分离株。可用于本发明的CRISPR阵列可以是I型、II型、III型、IV型或V型CRISPR阵列。

[0152] 在另外的实施方案中,提供了一种在细菌、古细菌、藻类或酵母群体中鉴定在其基因组中具有缺失的至少一种分离株的方法,包括:向细菌、古细菌、藻类或酵母细胞群体中引入:(a) 包含反式激活CRISPR(tracr)核酸的异源核酸构建体;(b) 异源核酸构建体,所述构建体包含CRISPR阵列(crRNA、crDNA),所述CRISPR阵列包含重复-间隔区-重复序列或至少一个重复-间隔区序列,其中所述重复-间隔区-重复序列或至少一个重复-间隔区序列的间隔区包含与所述群体中的细菌、古细菌、藻类或酵母细胞的基因组(如染色体、线粒体和/或质粒基因组)中的靶标区实质上互补的核苷酸序列;和(c) Cas9多肽或包含编码Cas9多肽的多核苷酸的异源核酸构建体,其中杀灭包含所述靶标区的细胞,从而产生不含所述靶标区的转化细菌、古细菌、藻类或酵母细胞的群体;以及从所述转化细菌、古细菌、藻类或酵母细胞的群体培养独立的细菌、古细菌或酵母集落,从而从所述转化细菌、古细菌、藻类或酵母细胞的群体中鉴定在其基因组中具有缺失的至少一种分离株。

[0153] 在另外的实施方案中,本发明提供了一种在细菌、古细菌、藻类或酵母群体中鉴定在其基因组中具有缺失的至少一种分离株的方法,包括:向所述细菌、古细菌、藻类或酵母

细胞群体中引入：(a) (i) 一种或多种异源核酸构建体，所述构建体包含与所述细菌、古细菌、藻类或酵母细胞的基因组中存在的至少300个连续核苷酸具有至少80%同一性的核苷酸序列，或者(ii) 两种或更多种异源核酸构建体，所述构建体包含至少一个转座子，从而产生转基因细菌、古细菌、藻类或酵母细胞的群体，所述转基因细菌、古细菌、藻类或酵母细胞在整合进所述基因组中的所述一种或多种异源核酸构建体与所述基因组中存在的所述至少300个连续核苷酸之间、或整合进所述基因组中的第一转座子与第二转座子之间包含非天然的同源重组位点；和(b) (i) 包含反式激活CRISPR(tracr)核酸的异源核酸构建体，(ii) 异源核酸构建体，所述构建体包含CRISPR阵列(crRNA、crDNA)，所述CRISPR阵列包含重复-间隔区-重复序列或至少一个重复-间隔区序列，其中所述间隔区包含与所述群体中的一种或多种细菌、古细菌、藻类或酵母细胞的基因组中的靶标区实质上互补的核苷酸序列；和(iii) Cas9多肽和/或包含编码Cas9多肽的多核苷酸的异源核酸构建体，其中所述靶标区位于掺入进所述基因组中的所述一种或多种异源核酸构建体与所述基因组中存在的所述至少300个连续核苷酸之间和/或所述第一转座子与所述第二转座子之间，以及杀灭包含所述靶标区的细胞，而不包含所述靶标区的细胞存活，从而产生不含所述靶标区的转化细菌、古细菌、藻类或酵母细胞的群体；并且从所述转化细菌、古细菌、藻类或酵母细胞的群体培养独立的细菌、古细菌或酵母集落，从而从所述群体中鉴定在其基因组中具有缺失的至少一种分离株。

[0154] 在一些实施方案中，可通过减小基因组大小或通过缺失需要高能量输入来转录和翻译的选定基因(编码多肽或功能性核酸(如，转录调节子))来增加适合性/生长速率。因而，在一些实施方案中，提供了增加细菌、古细菌、藻类或酵母细胞群体的适合性或生长速率的方法，包括：在如本文所述的群体的细菌、古细菌、藻类或酵母细胞的基因组中选定减小的基因组大小(如，选定基因组的一部分缺少)和/或缺失。在一些实施方案中，缺失可包含一个基因或不止一个基因。因此，通过减小基因组大小或缺失特定的一个基因或多个基因，群体的细胞不再消耗能量在基因组缺少的部分或缺失的一种基因或多种基因的转录/翻译上，从而与仍包含基因组的所述部分和/或所述一种基因或多种基因的对照群体相比具有降低的能量需求和增加的适合性。

[0155] 在其它实施方案中，提供了增加从细菌、古细菌、藻类或酵母细胞群体产生的产物的量的方法，包括通过在如本文所述的细菌、古细菌、藻类或酵母细胞的基因组中选定缺失来增加所述细胞的适合性或生长速率。在一些实施方案中，所述产物可包括但不限于抗生素、次级代谢物、维生素、蛋白质、酶、酸和药物。

[0156] 在一些实施方案中，可用于本发明的CRISPR阵列(crRNA、crDNA)可以是来自任何I型CRISPR-Cas系统、II型CRISPR-Cas系统、III型CRISPR-Cas系统、IV型CRISPR-Cas系统或V型CRISPR-Cas系统的阵列。

[0157] 就前述实施方案而言，可将包含tracr核酸的异源核酸构建体和包含CRISPR阵列的异源核酸构建体包含在同一构建体(如，表达盒或载体)或不同构建体中并引入。在特定的实施方案中，可将包含tracr核酸的异源核酸构建体和包含CRISPR阵列的异源核酸构建体包含在单个构建体(如，表达盒和/或载体)中，该单个构建体可任选还包含编码Cas9多肽的多核苷酸。在一些实施方案中，可将包含tracr核酸的异源核酸构建体和包含CRISPR阵列的异源核酸构建体与单个启动子和/或与分开的启动子有效连接。

[0158] 在一些实施方案中,可将包含反式激活CRISPR (tracr) 核酸的异源核酸构建体和包含CRISPR阵列(crRNA、crDNA) 的异源核酸构建体包含在CRISPR向导结构(gRNA、gDNA) 中。在一些实施方案中,可将CRISPR向导结构与启动子有效连接。

[0159] 在一些实施方案中,可用于本发明的Cas9多肽与Cas9核酸酶的氨基酸序列具有至少70%的同一性(如约70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%等)。可用于本发明的示例性Cas9核酸酶可以是已知在CRISPR-Cas系统中催化DNA切割的任何Cas9核酸酶。如本领域已知的,这种Cas9核酸酶包含HNH基序和RuvC基序(参见例如W02013/176772;W0/2013/188638)。在一些实施方案中,Cas9核酸酶的功能性片段可用于本发明。

[0160] CRISPR-Cas系统以及Cas9核酸酶的分组是本领域众所周知的并且包括例如,嗜热链球菌CRISPR 1 (Sth CR1) 组的Cas9核酸酶、嗜热链球菌CRISPR 3 (Sth CR3) 组的Cas9核酸酶、布氏乳杆菌 (*Lactobacillus buchneri*) CD034 (Lb) 组的Cas9核酸酶和鼠李糖乳杆菌GG (Lrh) 组的Cas9核酸酶。另外的Cas9核酸酶包括但不限于弯曲乳杆菌CRL 705的那些。可用于本发明的其它另外的Cas9核酸酶包括但不限于来自动物乳杆菌 (*Lactobacillus animalis*) KCTC 3501和香肠乳杆菌 (*Lactobacillus farciminis*) WP 010018949.1的Cas9。

[0161] 此外,在特定的实施方案中,Cas9核酸酶可由针对包含靶标DNA的生物体进行了密码子优化的核苷酸序列编码。在其它实施方案中,Cas9核酸酶可包含至少一种核定位序列。

[0162] 在一些实施方案中,可用于本发明的I型多肽与Cas3、Cas3'核酸酶、Cas3"核酸酶、它们的融合变体的氨基酸序列具有至少70%的同一性(如约70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%等)。在一些实施方案中,可用于本发明的I型Cascade多肽与Cas7 (Csa2)、Cas8a1 (Csx13)、Cas8a2 (Csx9)、Cas5、Csa5、Cas6a、Cas6b、Cas8b (Csh1)、Cas7 (Csh2)、Cas5、Cas5d、Cas8c (Csd1)、Cas7 (Csd2)、Cas10d (Csc3)、Csc2、Csc1、Cas6d、Cse1 (CasA)、Cse2 (CasB)、Cas7 (CasC)、Cas5 (CasD)、Cas6e (CasE)、Cys1、Cys2、Cas7 (Cys3)、Cas6f (Csy4)、Cas6和/或Cas4的氨基酸序列具有至少70%同一性(如约70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%等等)。

[0163] I型CRISPR-Cas系统是本领域众所周知的,并且包括例如:闪烁古生球菌 (*Archaeoglobus fulgidus*) 包含示例性的I-A型CRISPR-Cas系统,克鲁氏梭菌 (*Clostridium kluyveri*) DSM 555包含示例性的I-B型CRISPR-Cas系统,*Bacillus halodurans* C-125包含示例性的I-C型CRISPR-Cas系统,蓝丝菌属物种 (*Cyanothece* sp.) PCC 802包含示例性的I-D型CRISPR-Cas系统,大肠杆菌K-12包含示例性的I-E型CRISPR-Cas系统,硫还原地杆菌 (*Geobacter sulfurreducens*) 包含示例性的I-U型CRISPR-Cas系统,假结核耶尔森菌 (*Yersinia pseudotuberculosis*) YPIII包含示例性的I-F型CRISPR-Cas系统。

[0164] 在一些实施方案中,可用于本发明的II型多肽与Cas9的氨基酸序列具有至少70%的同一性(如约70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、

82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%等)。II型CRISPR-Cas系统是本领域众所周知的并且包括例如嗜肺性军团病杆菌巴黎菌株、嗜热链球菌CNRZ1066和乳糖奈瑟菌020-06。

[0165] 在一些实施方案中,可用于本发明的III型多肽与Cas6、Cas10(或Csm1)、Csm2、Csm3、Csm4、Csm5和Csm6、Cmr1、Cas10(或Cmr2)、Cmr3、Cmr4、Cmr5和Cmr6、Cas7、Cas10、Cas7(Csm3)、Cas5(Csm4)、Cas7(Csm5)、Csm6、Cas7(Cmr1)、Cas5(Cmr3)、Cas7(Cmr4)、Cas7(Cmr6)、Cas7(Cmr6)、Cmr5、Cas5(Cmr3)、Cas5(Cs×10)、Csm2、Cas7(Csm3)和all1473的氨基酸序列具有至少70%同一性(如约70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%等等)。III型CRISPR-Cas系统是本领域众所周知的,并且包括例如:表皮葡萄球菌(*Staphylococcus epidermidis*) RP62A,其包含示例性的III-A型CRISPR-Cas系统,激烈火球菌(*Pyrococcus furiosus*) DSM 3638,其包含示例性的III-B型CRISPR-Cas系统,热自养甲烷热杆菌(*Methanothermobacter thermautotrophicus*) Delta H菌株,其包含示例性的III-C型CRISPR-Cas系统,以及玫瑰弯菌属(*Roseiflexis*) 物种Rs-1,其包含示例性的III-D型CRISPR-Cas系统。

[0166] 在一些实施方案中,可用于本发明的IV型多肽与Csf4(dinG)、Csf1、Cas7(Csf2)和/或Cas5(csf3)的氨基酸序列具有至少70%的同一性(如约70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%等)。IV型CRISPR-Cas系统是本领域众所周知的,例如,氧化亚铁嗜酸硫杆菌(*Acidithiobacillus ferrooxidans*) ATCC 23270包含示例性的IV型CRISPR-Cas系统。

[0167] 在一些实施方案中,可用于本发明的V型多肽与Cpf1、Cas1、Cas2或Cas4的氨基酸序列具有至少70%的同一性(如约70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%等)。V型CRISPR-Cas系统是本领域众所周知的,并且包括例如,似新凶手弗朗西斯菌(*Francisella cf. novicida*) Fx1包含示例性的V型CRISPR-Cas系统。

[0168] 本发明还提供了包含本发明的核酸构建体、核酸阵列、核酸分子和/或核苷酸序列的表达盒和载体,它们可用于本公开的方法。

[0169] 在另外的方面,可将本发明的核酸构建体、核酸阵列、核酸分子和/或核苷酸序列引入宿主生物体的细胞中。可以使用可用于本发明的任何细胞/宿主生物体。示例性的宿主生物体包括细菌、古细菌、藻类和真菌(如酵母)。

[0170] 现在将参照下面的实施例描述本发明。应当理解,这些实施例无意于限制本发明的权利要求书的范围,而是旨在示例某些实施方案。技术人员能够想到的对示例方法的任何变型均意图落入本发明的范围内。

[0171] 实施例

[0172] 实施例1. 细菌菌株

[0173] 所有细菌菌株均在表1中列出。将细菌培养物冷藏于含有25%甘油(体积/体积)的适当培养基中并于-80℃下保存。使嗜热链球菌在37℃静态有氧条件下于补充有1%牛肉浸

膏(重量/体积)和1.9%(重量/体积) β -甘油磷酸盐(西格玛公司(Sigma))肉汤的Elliker培养基(Difco公司)中增殖48小时,或在含有1.5%(重量/体积)琼脂(Difco公司)的固体培养基上37℃下厌氧温育增殖48小时。适当时,将2 μ g/mL红霉素(Em)和5 μ g/mL氯霉素(Cm)(西格玛公司)的浓度用于嗜热链球菌中的质粒选择。将大肠杆菌EC1000在37℃下于Luria-Bertani(Difco公司)肉汤中,或在补充有1.5%琼脂的脑心浸液(BHI)(Difco公司)的固体培养基上有氧增殖。适当时,针对重组大肠杆菌用40 μ g/mL卡那霉素(Kn)和150 μ g/mL的Em维持大肠杆菌的抗生素选择。通过用1%乳糖、1.5%琼脂和作为pH指示剂的0.04%溴甲酚紫补充合成的Elliker培养基来定量评估对嗜热链球菌衍生物的 β -半乳糖苷酶活性筛选。

[0174] 实施例2.DNA分离和克隆

[0175] 所有试剂盒、酶和试剂均根据制造商的说明书使用。按以前的描述进行DNA纯化和克隆(41)。简而言之,从嗜热链球菌纯化基因组DNA时采用ZR真菌/细菌微量制备试剂盒(ZR Fungal/Bacterial MiniPrep kit)(Zymo公司)。使用凯杰离心微量制备试剂盒(Qiagen Spin miniprep kit)(凯杰公司(Qiagen))从大肠杆菌分离质粒DNA。用PFU HS II DNA聚合酶(Stratagene公司)进行DNA的高保真PCR扩增。用Choice-Taq Blue聚合酶(Denville公司)进行常规的PCR。用于PCR扩增的引物购自Integrated DNA Technologies公司(艾奥瓦州克拉尔维尔市(Coralville,IA))。用Zymoclean DNA凝胶回收试剂盒(Zymo公司)进行DNA从琼脂糖凝胶的提取。限制性核酸内切酶获自Roche Molecular Biochemicals公司。用New England Biolabs quick T4连接酶进行连接。由Davis Sequencing公司(加利福尼亚州戴维斯市)进行测序。按以前的描述制备冷藏的氯化铷感受态大肠杆菌(41)。构建具有lacZ靶向阵列的质粒,各依次由以下组成:(1)SthCRISPR1或SthCRISPR3特异性的天然前导序列、(2)CRISPR 1或CRISPR 3特异性的天然重复、(3)lacZ的5'端特异性的间隔区序列、(4)另一天然重复(图1)。为了工程改造每种质粒,订购了作为延伸寡聚物的上文所列的序列特征(表2),通过重叠延伸PCR利用剪接组合(42)并克隆进pORI28中(图2)。

[0176] 实施例3.CRISPR间隔区的选择和设计

[0177] 染色体切割的可编程特异性取决于所需的靶标等位基因独特的间隔区序列的选择。特异性还因必需的PAM而变得复杂,PAM是必须邻近靶标序列中的原间隔区的短保守序列(21,43)。因而,间隔区选择和设计的严格标准是共有PAM序列的位置以及与外来基因组座位的偶然序列同一性。通过首先限定lacZ的有义链和反义链中所有推定PAM序列的位置来约束推定的原间隔区。在3,081nt基因中,存在22个CRISPR1(AGAAW)和39个CRISPR3(GGNG)PAM位点,这些位点与它们通过生物信息分析而推断出的共有序列具有同一性(21)。在鉴定潜在的间隔区后,将完整的原间隔区、种子和PAM序列对嗜热链球菌LMD-9的基因组进行BLAST分析,以防止额外的对非特异性基因座的靶向。CRISPR1的间隔区和CRISPR3的间隔区在序列及对应的PAM位点上是不相干的,但是被设计为靶向lacZ的5'端,从而导致间隔6nt的预定切割位点。因此,每个质粒上的前导序列、重复和间隔区代表了分别对CRISPR1或CRISPR3独特的直系特征。为了评估靶标基因座依赖性突变,产生额外的CRISPR3质粒,其具有针对金属阳离子结合残基的间隔区,该残基对于 β -半乳糖苷酶活性是必需的。将含有非自我靶向间隔区的CRISPR1阵列质粒用作对照来量化自我靶向的致死性。

[0178] 实施例4.转化

[0179] 根据以前描述的方法将质粒电穿孔进含有温度敏感性辅助质粒pTRK669的感受态

嗜热链球菌中(44)。简而言之,将嗜热链球菌的过夜培养物以1% (体积/体积) 接种进50mL的Elliker培养基中,该培养基补充有1%牛肉浸膏、1.9%β-甘油磷酸盐和Cm选择剂。当培养物达到OD_{600nm}为0.3时,添加青霉素G以实现10μg/mL的最终浓度,以便促进电穿孔效率(45)。通过离心收集细胞并在10mL冷电穿孔缓冲液(1M蔗糖和3.5mM MgCl₂)中洗涤3次。将细胞在电穿孔缓冲液中浓缩100倍并将40μL的该悬浮液等分进0.1mm电穿孔杯中。将每份悬浮液与700ng质粒合并。电穿孔条件设定为2,500V、25μF电容和2000Ω电阻。记录时间常数,其范围为4.4至4.6ms。立刻将悬浮液与950μL回收培养基合并并在37℃下温育8小时。将细胞悬浮液涂布接种于选择培养基上并用培养基洗涤电穿孔杯以确保回收细胞。

[0180] 实施例5.β-半乳糖苷酶表型确认

[0181] 最初通过将集落重新划线接种于半合成Elliker培养基上针对β-半乳糖苷酶缺陷表型对由CRISPR1和CRISPR3二者产生的转化株进行筛选,该培养基补充有作为市售碳水化合物来源的1%乳糖。通过进行米勒测定法(邻-硝基苯基-β-D-半乳糖苷(ONPG))来确认β-半乳糖苷酶活性的丧失(46)。简而言之,使培养物在5mL培养基中增殖至晚对数期(OD_{600nm}=1.2)并通过离心(4,000x g 10分钟)收集。洗涤细胞并重新悬浮于0.5mL磷酸盐缓冲盐水(Gibco-Invitrogen公司)中。将每份悬浮液与100μL的0.1mm玻璃珠(Biospec公司)合并,然后在Mini-Beadbeater(Biospec公司)中进行五个60s匀浆循环。然后将样品离心(15,000x g,5分钟)以除去碎片和完整的细胞。将细胞裂解物(10μL等分试样)与600μL底物溶液(60mM Na₂HPO₄;40mM NaH₂PO₄;1mg/mL ONPG;2.7μL/mLβ-巯基乙醇)合并并在室温下温育5分钟,此时添加700μL终止溶液(1M NaCO₃)。记录420nm下的吸光度并将β-半乳糖苷酶的活性报告为米勒单位,按先前的描述计算(46)。

[0182] 实施例6.生长和活性评估

[0183] 通过将培养物在乳糖缺乏的半合成Elliker培养基中继代培养12代来预先调理以供生长测定法。将新鲜的培养基以1% (体积/体积) 接种过夜培养物并在37℃下静态温育。每小时监控OD₆₀₀直至培养物达到稳定期。通过用过夜培养物接种脱脂乳至10⁸cfu/mL并在42℃下温育来评估牛奶的酸化。随后使用Mettler Toledo Seven Easy pH计和Accumet探针监测pH。脱脂乳获自NCSU乳品厂并在80℃下巴氏灭菌30分钟。

[0184] 实施例7.可消耗基因组区的鉴定

[0185] 基于i) IS元件的位置、取向和核苷酸同一性以及ii) 必需ORF的位置,进行用于CRISPR-Cas靶向的可移动和可消耗基因座的计算机预测。在枯草芽孢杆菌(*Bacillus subtilis*)中,通过测定全基因组基因敲除的致死性而鉴定了271个必需的开放阅读框(33)。使用BLASTp搜索工具在用于氨基酸序列的默认打分矩阵下针对与来自枯草芽孢杆菌的每种必需基因同源的同源物查询嗜热链球菌基因组。在嗜热链球菌中鉴定了与约239种必需ORF同源的同源物,所有这些同源物均是染色体编码的(表4)。容易地鉴定了涉及保守的细胞过程(包括DNA复制/内稳态、翻译机制以及核心代谢途径)的蛋白质。根据发酵细菌的代谢谱,未观察到对应于细胞色素生物合成/呼吸作用同源物的同源物。用SnapGene软件将每个推定的必需ORF基因作图至参照基因组,从而便于使它们在嗜热链球菌LMD-9中的位置和分布可视化(图3A)。

[0186] 通过使用Geneious®软件比对转座子编码序列来将嗜热链球菌基因组内的IS元件分组(图4)。根据在IS元件数据库(www-is.biotoul.fr/)内的BLAST分析确定家族命名。

为了预测重组介导的染色体区段切除潜能,将相关IS元件的相对位置作图至嗜热链球菌基因组中(图3A)。IS1193和Sth6IS家族的元件在基因组中出现最频繁并且普遍存在于肺炎链球菌(*Streptococcus pneumoniae*)和变形链球菌(*Streptococcus mutans*)中(34)。尽管IS1193元件普遍存在,许多这些基因座显示为展现出一定多态性和简并性的小片段,但也存在具有高水平序列同一性的数个拷贝(图5A)。相比之下,Sth6家族展现出相当的多态性和高简并性,其中一些拷贝带有显著的内部缺失(图5B)。IS1167和IS1191元件频率较低,但在基因组中鉴定的拷贝之间展现出接近完全的保真性(图5C和图5D)。基于嗜热链球菌的IS1167和IS1191元件的长度和序列的保守性,以及它们与牛奶适应基因的相对接近度,我们推测基因组中最近获得了这些保守的/高保真转座子。

[0187] 通过组合预测的必需ORF和IS元件的位置,鉴定了旁侧为高保真性IS元件的可消耗基因组岛(图3A)(表3)。第一个基因组岛含有嗜热链球菌LMD-9独特的操纵子,编码具有未知特异性的推定APT依赖性寡核苷酸转运系统(图3B)(35)。第二个带有细胞包膜蛋白酶PrtS,其促成嗜热链球菌的快速酸化表型(图3C)(36)。要注意的是,虽然prtS在嗜热链球菌基因组中不是普遍存在的,但已经证明利用自然感受态编码prtS的基因组岛在菌株之间是可以转移的(36)。第三个基因组岛含有推定的ATP依赖性铜外排蛋白并且在每个测序的嗜热链球菌菌株中存在(图3D)。第四个基因组岛目前就长度(102kbp)和基因含量而言是最大的,有102种预测的开放阅读框,包括lac操纵子(图3E)。该基因组岛存在于全部的嗜热链球菌菌株中,但是特定基因含量和长度在各菌株之间变化。为了确定用两种内源II型系统靶向大基因组岛的结果,针对lacZ编码序列产生了重复-间隔区阵列(图3E)并克隆进pORI28(图2)中。因为其大小、在嗜热链球菌菌株中的遍在性以及能够基于 β -半乳糖苷酶阴性表型筛选lacZ突变的能力,选择第四个基因组岛用于CRISPR-Cas靶向。

[0188] 实施例8.lacZ的CRISPR-Cas靶向选择大缺失事件

[0189] 在II型系统中,Cas9侦询DNA并可逆地结合至PAM序列,其中通过形成tracrRNA::crRNA双链体在靶标处发生Cas9的激活(37),最终导致dsDNA切割(图6A和图6B)(25)。图14A和图14B是示出了用于增补内源CRISPR系统进行靶向杀灭的通用方法的示意图。具体地讲,这些图14A和图14B示出了用于增补嗜热链球菌中的内源II型系统进行靶向杀灭的方法。因而,在嗜热链球菌中,通过设计侧接天然重复的基因组靶向间隔区序列,利用CRISPR-Sth1(A)或CRISPR-Sth3(B)II型系统实现了程序化细胞死亡,该侧接天然重复的基因组靶向间隔区序列的表达由天然或合成启动子驱动。转录的重复-间隔区阵列通过宿主编码的RNA酶III和Cas9加工而产生成熟的crRNA,其募集Cas9至基因组而引发双链DNA切割,从而导致细胞死亡。

[0190] 用引发由CRISPR-Cas系统进行的染色体自我靶向的质粒转化显示出细胞毒性,该细胞毒性由与非自我靶向质粒相比存活转化株的相对减少来测量(15,29)。靶向嗜热链球菌中的lacZ基因导致回收的转化株减少约2.5个对数(图6C),接近转化效率的极限。双链DNA裂口(DSB)对生物体的存活带来巨大威胁。相应的修复途径通常需要末端切除来修复平末端DNA。Cas9实现的内切核苷酸裂解进一步加剧了DSB引起的突变的压力,因为靶标基因座恢复至野生型不规避后续的CRISPR靶向。对乳酸菌内的间隔区起点进行的鉴定表明,22%的间隔区展现出与自身的互补性,并且相应的基因座发生改变,可能有利于天然存在的自我靶向事件的存活(28)。

[0191] 为了确定靶标基因座是否响应Cas9诱导的切割而发生突变,首先针对 β -半乳糖苷酶活性的丧失对转化子进行筛选。将活性缺陷的克隆在lacZ基因座处进行基因型分析。在经测序的任何克隆中未观察到因经典末端连接或替代性末端连接引起的突变,也未观察到任何自发的单核苷酸多态性。不存在单核苷酸多态性可能归因于低转化效率外加低点突变发生率,并且不存在Ku和连接酶IV同源物与不存在非同源末端连接相关(38)。PCR筛选表明,不存在野生型lacZ,但是PCR扩增子不对应天然的lacZ基因座,而是,在另一基因座处扩增了侧接IS元件的序列。为了研究负责 β -半乳糖苷酶活性丧失的基因型,对两个克隆进行了单分子实时测序(Single Molecule Real Time sequencing):一个由靶向lacZ的5'端的CRISPR3产生,一个由靶向编码 β -半乳糖苷酶催化所需的离子结合口袋的序列的CRISPR3产生(图7A和图7B)。采用该测序策略是因为其长的读段长度可以规避将读段可靠地映射至正确座位的困难,该困难是因为基因组中有大量的IS元件(35)。用Geneious软件将读段映射至参照基因组序列,表明不存在编码lacZ开放阅读框的大区段(约102kbp)(图7A和图7B)。两种经测序的菌株均确认了大缺失边界的重现性,并且显示,发生的缺失不依赖于用于靶向的lacZ间隔区或CRISPR-Cas系统。然而,测序数据并未可靠地展示该缺失的精确界点。

[0192] 缺失的102kbp区段占嗜热链球菌1.86Mbp基因组的大约5.5%。该区域含有102个推定的开放阅读框(STER_1278-1379),编码ABC转运蛋白、双组分调节系统、细菌素合成基因、噬菌体相关蛋白、乳糖分解代谢基因以及不具有注释功能的数个隐形基因(35)。通过测量OD_{600nm}随时间的变化而在肉汤培养物中评估该缺失对生长表型的影响(图7C)。缺失克隆看起来具有更长的迟缓期、更低最终OD($p < 0.01$)并且在迟缓期期间展现出显著更长的世代时间,平均值为103分钟,相比之下野生型为62分钟($p < 0.001$)。尽管缺失衍生物每个世代要复制的基因组少5.5%,并且不需要耗费资源用于被除去的ORF的转录或翻译,但未观察到适合性相对于野生型的明显增加。 β -半乳糖苷酶活性是乳酸菌工业应用的标志性特征并且是食品系统通过酸化进行保存所必需的。因此通过监测pH对lacZ缺陷型嗜热链球菌菌株酸化牛奶的能力进行评估(图7D)。可以预测到,与野生型中所观察到的快速酸化表型形成鲜明对比,在实验期间缺失菌株不能酸化牛奶。

[0193] 实施例9. 通过同源IS元件之间的重组发生基因组缺失

[0194] 为了研究缺失的机理,测定了所述区段两侧的核苷酸序列。在界点处观察到的唯一同源序列是两个截短的IS1193插入序列,它们全局在727bp上展现出91%的核苷酸序列同一性。因此,设计了这两个IS元件两侧的引物对来扩增展现出缺失的存活克隆的基因组DNA。每种缺失菌株展现出具有预测大小(约1.2kb)的强条带,确认了大的基因组缺失事件(图8A)。有意思的是,在野生型中观察到对应于该染色体缺失的微弱扩增子,从而表明该区域可能在野生型群体中以低的比率被天然地从基因组切除。对20个克隆进行界点扩增子的测序,所述克隆通过CRISPR3进行染色体自我靶向而产生。该基因座的基因型分析表明在每个克隆中存在一个嵌合IS元件,此外还揭示了每个克隆的该嵌合元件内从上游元件向下游序列的转换(图8B)。所观察到的缺失的大小范围为101,865-102,146bp。转换的确切基因座在所述克隆内是可变的,但不是随机的,暗示该缺失机制的潜在偏向。嗜热链球菌带有编码为RecA(STER_0077)的典型重组机制,RecA是AddAB同源物,起RecD家族的双ATP依赖性DNA外切核酸酶(STER_1681和STER_1682)和解旋酶(STER_1742)的作用。旁侧IS元件之间的高核苷酸同一性以及嗜热链球菌进行位点特异性重组的能力(4)确认了RecA介导的重组介导

基因组区段切除的潜能(图8C)。

[0195] 接下来,评价CRISPR-Cas靶向帮助分离具有相同遗传架构的每个基因座的缺失的能力。为此,产生三种CRISPR3重复-间隔区阵列并克隆进pORI28中,一种靶向第一基因座中的寡核苷酸转运蛋白基因,一种靶向来自第二基因座的prtS,另一种靶向来自第三基因座的ATP酶铜外排基因(图5)。为了筛选缺失,设计每个基因座处的IS元件旁侧的引物来扩增每个缺失界点(图8D)。还在每种情形中通过设计每个基因组岛的内部引物来确认野生型基因座的不存在(图8E)。用靶向质粒转化后,分离每个基因座处的缺失并确认野生型的不存在。缺失界点扩增子的测序确认了单个嵌合IS元件足迹得到保留,从而表明每个基因座处的缺失机制相同。有意思的是,IS元件旁侧的引物也从野生型gDNA发生了扩增,进一步表明在每个基因座处自然出现的群体异质性是由于自发的基因组缺失。这些结果暗示,序列特异性Cas9切割可选择缺少靶向所必需的原间隔区和PAM组合的变体。因而,利用CRISPR-Cas靶向作为已丢失那些基因组岛的微生物变体的强力选择方法,可分离自发的基因组缺失。

[0196] 实施例10. 群体筛选

[0197] 在该研究中,将嗜热链球菌中所带的天然IIA型系统重新用于限定大基因组岛的自发缺失。通过独立地靶向嗜热链球菌中的四个基因组岛,生成了总体缺少总计基因组的7%的稳定突变体。对缺失界点的表征表明IS依赖性重组机制促成了群体异质性,并且揭示了范围为8至102kbp的缺失事件。对嵌合IS元件的精确基因作图表明,天然的重组事件可能是造成嗜热链球菌中的大染色体缺失的原因并且可能潜在地被用于靶向基因组编辑。

[0198] 我们的结果证明,野生型克隆被从群体中移除,除而不具有CRISPR-Cas靶向特征的突变体存留。因而,适应性基因组岛得以鉴定和验证,显示了内源Cas9进行的精确靶向可被用于分离混合群体中的大缺失变体。

[0199] 细菌的基因组进化通过水平基因转移、内在突变和基因组重建发生。嗜热链球菌菌株的基因组测序和对比分析已揭示了明显的基因组衰减,也表明通过小生境特异性基因获得发生了对富营养食品环境的适应(18;35)。嗜热链球菌基因组中存在包括整合型和接合型元件、原噬菌体和IS元件在内的MGE指示了针对乳制品环境的快速进化(38-39)。可移动的遗传元件特征有利于基因获得,以及相反地,有利于非必需序列的失活或丢失。因此,MGE赋予了基因组可塑性,作为增加适合性或改变生态生活方式的手段。我们的结果强烈表明,这些元件的CRISPR-Cas靶向可影响染色体重排和内稳态。这与靶向必需特征的实验形成对比,后者导致了具有失活的CRISPR-Cas机制的变体的选择(Jiang 2013)。必需开放阅读框的突变不是规避CRISPR-Cas靶向的可行手段,因而仅具有失活CRISPR-Cas系统的那些克隆得以保留。通过设计,对预测为高度可变且可消耗的遗传元件进行靶向证明了具有改变的基因座的变体是存活的,在自我靶向事件期间保持活性的CRISPR-Cas系统。

[0200] 尽管IS元件在细菌基因组中几乎遍在分布,但它们仍是神秘的遗传实体,主要是由于它们在功能方面的多样性和可塑性(34)。我们的结果表明,有可能通过分析它们的位置、取向和序列保守性来预测相关IS元件之间的重组(图4和图5)。然后可将CRISPR-Cas靶向用于经验性地验证每个预测基因座处的群体异质性,同时增加低发生率突变体的回收率。MGE在乳酸菌、尤其是嗜热链球菌中的高度普遍性与它们在这些高适应性细菌通过基因组进化形成物种的过程中的作用相一致(39-40)。此外,利用CRISPR-Cas靶向来回收基因组缺失突变体可有利于具有未知功能的基因的表型表征。展现出编码lac操纵子的102kb基因

组岛的缺失的突变体相对于野生型具有显著增加的世代时间,并且实现较低的最终OD。在其中存在102个预测开放阅读框的情况下,有可能额外的表型受到影响并且许多这些基因不具有注释的功能。考虑到小生境特异性基因诸如prtS的工业相关性,该方法允许对基因组岛编码基因如何促成对在牛奶中的生长的适应性进行直接评估。此外,这是在水平获得的性状的自然基因组和生态背景中,因为它们有可能作为离散基因组岛获得。这些结果确立了新的方法,可用于将细菌中的自我靶向CRISPR-Cas9系统应用于研究转座、DNA修复机理和基因组可塑性。

[0201] CRISPR-Cas系统通常通过干扰遗传元件来限制遗传多样性,但是获得的MGE也可以为宿主细菌提供适应性优势。因而,在有CRISPR-Cas靶向的情况下仍保持基因组整合的MGE的有益效果是基因组内稳态的重要推动力。总体而言,我们的结果确定,可将GEI的计算机预测与CRISPR-Cas靶向相结合来分离展现出大基因组缺失的克隆。每个缺失界点处的嵌合插入序列足迹指示所有四个基因组岛具有共同的缺失机理。与基因座具有同一性的自我靶向间隔区的高度普遍性,加之实验证明了基因组改变,表明CRISPR-Cas自我靶向可显著地促成细菌的基因组进化(28;30)。总体而言,对CRISPR-Cas诱导的大缺失的研究证实该方法是对没有注释的基因簇评估其必要性和功能,以及基于通过可转座元件发生的染色体缺失来限定最小细菌基因组的一种快速有效的手段。

[0202] 图9示出了限定的基因座,其用于评估基于II型CRISPR-Cas系统通过靶向嗜热链球菌LMD-9的基因组的致死性。进行该分析的方法是本领域已知的。参见Selle和Barrangou PNAS.112(26):8076-8081(2015)。

[0203] 测试了两个正交II型系统(CRISPR1和CRISPR3);CRISPR1靶标为深灰色,CRISPR3靶标为浅灰色。选择特定的遗传特征来测试(i)基因间区(INT)、(ii)可移动遗传元件(IS、GEI1-GEI3、PRO、lacZ、EPS)、(iii)必需基因(dltA、LTA)、(iv)复制弧的极点(ORI、TER)以及DNA的正链对负链(外部靶标对内部靶标)。

[0204] 图10示出了通过靶向图9中所限定的区域而实现的基于CRISPR的致死性。参照非靶向质粒对照pORI28的转化计算CFU的对数降低。对于所测试的所有靶标,致死率范围为降低2个对数至降低3对数,而与染色体位置、编码序列或必需性无关。

[0205] 图11示出了CRISPR介导的基因组岛缺失菌株的转录谱。对CRISPR靶向基因组岛1-4后幸存的细胞进行回收和基因型分析,鉴定了缺少每个实验中所靶向的基因组岛的稳定的独立突变体。随后,使这些细胞增殖并对它们的总RNA进行分离和测序。采用该方法,通过将测序读段映射至参照基因组而生成了转录谱。在每种情形中,缺少预测基因组岛座位的测序读段还表明靶标遗传实体的丢失,同时对整个剩余基因组的核心基因的表达具有极低影响。

[0206] 此外,RNA测序数据支持了通过利用读段覆盖作图和转录值比较确定的缺失的边界,并且支持了利用同一数据集生成的对比转录组学确定的表型的辨别。具体而言,利用高通量RNA测序确认在预期缺失区不存在转录活性,如图12中所示。图12示出了基因组岛缺失菌株以及每个基因组岛菌株(GEI1-GEI4)的 $\log_2$ 转换RNA测序读段覆盖度,缺少预测基因组岛座位的测序读段还表明靶标遗传实体的丢失,同时对整个剩余基因组的核心基因的表达具有极低影响。

[0207] 图13还确认缺少缺失基因的转录活性。对于每种基因组岛缺失菌株(GEI1-GEI4),

靶标基因组岛(黑色)每一者上所编码的基因的表达最少。GEI1中编码的基因在左上分图中示出,GEI2中编码的基因在右上分图中示出,GEI3中编码的基因在左下分图中示出,GEI4中编码的基因在右下分图中示出。总体上,基因组岛缺失1和2对其它基因(灰色)的转录具有极低影响,而基因组岛3和4看起来影响不在所述基因组岛上编码的其它基因的转录。

[0208] 此外,将RNA测序数据用于比较不在缺失基因组岛(GEI4)上编码的基因(即仍存在于染色体中的其它基因)的转录水平,并鉴定与基因组缺失相关的表型。缺失菌株中差异化转录的基因表明,细胞过程受丢失的基因的影响或者对这些基因活性的丧失存在补偿。因而,可以推断与这些基因或基因组区域相关的途径。表5提供了在缺失菌株GEI4中鉴定的差异表达基因的列表。观察到差异化表达的许多基因涉及嗜热链球菌的生物合成能力,包括芳族氨基酸和嘌呤生物合成。

[0209] 实施例11.利用II型CRISPR-Cas系统靶向杀灭干酪乳杆菌

[0210] 图15中提供了干酪乳杆菌的示例性CRISPR-Cas II型向导结构。上面的分图提供了预测的向导结构,而左下分图示出了示例性的双向导结构,右下分图示出了示例性的单向导结构。

[0211] 实施例12.利用II型CRISPR-Cas系统靶向杀灭加氏乳杆菌

[0212] 图16中提供了用于靶向杀灭加氏乳杆菌的示例性II型向导结构。上面的分图提供了预测的向导结构,而左下分图提供了正确的双向导结构crRNA:tracrRNA(通过RNA测序确认),右下分图提供了示例性的预测的单向导结构。

[0213] 将质粒转化进加氏乳杆菌中,每种质粒携带如下的不同构建体:空pTRK563载体、具有正确的原间隔区但具有不正确的PAM的构建体、具有正确PAM但具有不在该阵列中的原间隔区的构建体以及具有正确原间隔区且具有PAM的构建体。结果示于图19中。具有正确原间隔区和正确PAM的质粒显示出显著更高的干扰靶向和细胞死亡。

[0214] 实施例13.利用II型CRISPR-Cas系统靶向杀灭戊糖乳杆菌

[0215] 图17中提供了用于靶向杀灭戊糖乳杆菌的示例性II型向导结构。顶部的分图示出了预测的向导结构。左下分图为正确的双向导结构crRNA:tracrRNA(通过RNA测序确认),右下分图为示例性的预测的人工单向导结构。

[0216] 将质粒转化进戊糖乳杆菌中,每种质粒携带如下的不同构建体:具有正确的原间隔区但具有不正确的PAM的构建体、具有正确PAM但具有不在该阵列中的原间隔区的构建体、空pTRK563载体以及具有正确原间隔区且具有正确PAM的构建体。结果示于图20中。具有正确原间隔区和正确PAM的质粒(Lpelgttaat)显示出显著更高的干扰靶向和细胞死亡。

[0217] 实施例14.利用II型CRISPR-Cas系统靶向杀灭詹氏乳杆菌

[0218] 图18提供了示例性的CRISPR-Cas II型向导结构。左边的分图为正确的双向导结构crRNA:tracrRNA(通过RNA测序确认),右边的分图为示例性的预测的人工单向导结构。

[0219] 将质粒转化进加詹氏乳杆菌中,每种携带如下的不同构建体:包含空pTRK563载体的构建体、具有正确的原间隔区但具有不正确的PAM的构建体、具有正确PAM但具有不在该阵列中的原间隔区的构建体、以及具有正确原间隔区且具有正确PAM的构建体。结果示于图22中。具有正确原间隔区和正确PAM的质粒显示出实质上更高的干扰靶向和细胞死亡。

[0220] 实施例15.利用I型CRISPR-Cas系统靶向杀灭干酪乳杆菌NCK 125

[0221] 图21提供了用于干酪乳杆菌的示例性I型CRISPR-Cas向导结构,其包含干酪乳杆

菌NCK 125中存在的天然I型前导序列和重复。使用该生物体中的天然间隔区序列预测PAM 5'-YAA-3'。该人工阵列含有靶向宿主基因组中的16s rDNA基因的间隔区。图22中提供了结果,其示出了空载体与以下两种不同人工阵列之间的显著降低:其中一种含有靶向16s基因中的+链的单间隔区(1-2alt),另一种阵列含有靶向+链的初始间隔区并且还含有靶向16s基因中的-链的额外间隔区(1,2-3)。

[0222] 本文所述的CRISPR-Cas系统可用于例如,(i)在医学干预的情形中靶向性地减少病原体(如,包括但不限于以下的病原体:真菌、线虫、原生动物(如疟疾)、绦虫、球虫类(微孢子目)、吸虫类、舌形虫类、棘头虫类、节肢动物类等等);(ii)用于保护消费品(食品系统、动物、农作物);(iii)用于从工业发酵过程(原料、加工设备、起始培养物)控制和/或移除不期望的生物体以及(iv)用于控制环境微生物复合系以影响生态系统和/或化学循环以及用于修复。表1.细菌菌株和质粒

[0223]

菌株名称	描述	初始参照
大肠杆菌 EC1000	pORI 质粒的宿主, 染色体 $repA^+$ (pWV01), Km^R pTRK935 的宿主	47
嗜热链球菌 LMD-9	野生型	40
具有 pTRK669 的嗜 热链球菌 LMD-9	野生型, pTRK669 赋予的 $RepA^+$ 和 Cm^R	本研究
质粒		
pORI28	宽范围非复制型载体, Em^R	47
pTRK669	Ts-辅助质粒 $repA^+$, Cm^R	44
pCRISPR1:: <i>lacZ</i>	pORI28:: CRISPR1-前导序列-RSR- <i>lacZ</i> N-端间隔区	本研究
pCRISPR3:: <i>lacZ</i>	pORI28:: CRISPR1-前导序列-RSR- <i>lacZ</i> 活性位点 间隔区	本研究
pCRISPR3:: ABC	pORI28:: CRISPR1-前导序列-RSR-ABC 间隔区	本研究
pCRISPR3:: Cu	pORI28:: CRISPR1-前导序列-RSR-Cu 外排间隔区	本研究
pCRISPR3:: <i>prts</i>	pORI28:: CRISPR1-前导序列-RSR- <i>prts</i> 间隔区	本研究
pCRISPR3:: <i>lacZ</i> pCRISPR1:: Non-self	pORI28:: CRISPR3-前导序列-RSR- <i>lacZ</i> N-端间隔区 pORI28:: CRISPR1-前导序列-RSR-非自我靶向间隔区	本研究 本研究

[0224]

表 2. 引物

引物名称	序列	功能
C1_N-term_F	CAAGAACAGTTATTGATTTTATAATCACTATGTGGGTATGAAAAATCTCAAAAAATCAT TTGAGGTTTTTGTACTCTCAAGATTAAAGTAACGTGTACAACATTAAAGAGATTGTCTT AACTT	SOE-PCR 的模板
C3_N-term	AGCGGATAACAAATTTCAAGTTTGGAAACCATTTCGAAACAACACACAGCTCTAAAACTCA GAAAAATTTCTTCAAGAGATTCAAAAATACTGTTTTTGGAAACCATTTCGAAACAACACACAGCT CTAAACCTCGTAGGATATCTTTTCTAC	SOE-PCR 的模板
C1_N-term_R	ACGGGATAACAAATTTCAAGTTTGTACAGTTACTTAAATCTTTGAGAGTACAAAAACAGG GGAGATGAAGTTAAGACAATCTCTTAATGT	SOE-PCR 的模板
C3_A-位点	AGCGGATAACAAATTTCAAGTTTGGAAACCATTTCGAAACAACACACAGCTCTAAAAACGAA GTCTTGGTCTTCCAAACCAGCTTGCTGTAGTTTTTGGAAACCATTTCGAAACAACACACAGCT CTAAACCTCGTAGGATATCTTTTCTAC	SOE-PCR 的模板
C3_ABC	AGCGGATAACAAATTTCAAGTTTGGAAACCATTTCGAAACAACACACAGCTCTAAACGATA ACACGAGATAAAACATCCAGCCACCGTTTTTGGAAACCATTTCGAAACAACACACAGCTCTA AAACctcgtaggatatcttttctac	
C3_prtS	AGCGGATAACAAATTTCAAGTTTGGAAACCATTTCGAAACAACACACAGCTCTAAACGTTG TAGCTTTGAGTCTGAGAAATACACGGCTTTTGGAAACCATTTCGAAACAACACACAGCTCTA	

[0225]

	<u>AAACctcgtaggatatcttttctac</u>	
C3_Cu	AGCGGATAACAATTTTCAGCTTTTGGAAACCATTTCGAAACAACACAGCTCTAAAAACGAT TGCTCAATCAATCGTTTTCAGCTGCTAAGTTTTTGGAAACCATTTCGAAACAACACAGCTC <u>TAAAACctcgtaggatatcttttctac</u>	
C3LF	AGCAGGGATCCTGGTAATAAGTATAGATAGTCTTG	从 gDNA 扩增 Sth3 前导序列
C3LR	CTCGTAGGATATCTTTTCTAC	从 gDNA 扩增 Sth3 前导序列
C1F	AGCAGGGATCCCAAGAACAGTTATTGATTTTATAATC	CRISPR1 SOE-PCR 正向引物
C3F	AGCAGGGATCCTGGTAATAAGTATAGATAGTCTTG	CRISPR3 SOE-PCR 正向引物
C1C3R	TGCTGGAGCTCGTGAAATTGTTATCCGCT	CRISPR1 和 CRISPR3 SOE-PCR 反向引物
1193F	TTGAACACTAGGAACCTCATA	缺失接口扩增
1193R	CGTAAGGTTTTTGATGACTCAAG	缺失接口扩增
pORI28F	TTGGTTGATAATGAACCTGTGCTG	对 pORI28 的 MCS 测序
pORI28R	TTGTTGTTTTTATGATTACAAAGTGA	对 pORI28 的 MCS 测序

[0226] 表3.推定的可消耗基因组岛和基因组小岛

[0227]

基因 组岛	ORF区	长度 kbp	GC 含量 %	注释基因	IS家族
1	STER_139-STER_148	7.81	37.1	寡肽转运蛋白	IS6
2	STER_840-STER_848	10.29	39.9	蛋白酶 PrtS	ISSt16/IS1 167
3	STER_881-STER_888	8.71	39.3	铜外排	IS1191
4	STER_1277-STER_1380	101.76	37.2	乳糖分解代谢, 2-成分 调节, 细菌素合成, ABC 转运蛋白	IS1193

[0228]

表 4. 嗜热链球菌中鉴定的约 239 个必需开放阅读框的同源物

基因组-部分	STER	起始	终止	方向		查询 覆盖 率	e-值	氨基酸	
CSTER	1	101	1465	+	染色体复制起始蛋白 DnaA	98%	4. 00E-124	酸 Id	枯草芽孢杆菌注释
CSTER	2	1620	2756	+	DNA 聚合酶 III 亚基 β	99%	1. 00E-88	39%	染色体复制起始蛋白 DnaA
CSTER	6	5818	6387	+	肽酰-tRNA 水解酶	97%	1. 00E-61	51%	DNA 聚合酶 III 亚基 β dnan
CSTER	9	10331	10702	+	细胞周期蛋白	94%	1. 00E-05	22%	细胞分裂蛋白 DivIC
CSTER	12	12122	13387	+	tRNA (Ile)-赖西丁合成酶, MesJ	92%	3. 00E-40	32%	tRNA (Ile)-赖西丁合酶 tils
CSTER	13	13469	14011	+	次黄嘌呤鸟嘌呤磷酸核糖基转移酶	99%	2. 00E-82	63%	次黄嘌呤鸟嘌呤磷酸核糖基转移酶 hprt
CSTER	40	25595	26425	+	细胞形状决定蛋白 MreC	94%	1. 00E-22	28%	细胞形状决定蛋白 MreC
CSTER	43	28617	29582	+	磷酸核糖焦磷酸激酶	97%	3. 00E-145	65%	磷酸核糖焦磷酸激酶 prs
CSTER	47	32842	33846	+	推定的甘油-3-磷酸酰基转移酶	94%	9. 00E-124	54%	磷酸酰基转移酶 PlsX
CSTER	48	33846	34091	+	酰基载体蛋白	76%	1. 00E-08	40%	酰基载体蛋白
CSTER	65	53036	54727	-	精氨酸-tRNA 合成酶	98%	6. 00E-52	28%	
CSTER	95	72798	77192	+	DNA 聚合酶 III PolC	99%	0. 00E+00	51%	DNA 聚合酶 III PolC 型
CSTER	105	82956	83723	+	30S 核糖体蛋白 S2	96%	1. 00E-118	69%	
CSTER	106	83841	84881	+	翻译延伸因子 Ts	97%	8. 00E-72	44%	

[0229]

基因组部分	STER	起始	终止	方向		查询覆盖率	e-值	氨基酸 Id	杜草芽孢杆菌注释
CSTER	117	97363	98706	+	半胱氨酰-tRNA 合成酶	99%	1.00E-178	54%	
CSTER	127	104406	104852	+	50S 核糖体蛋白 L13	99%	9.00E-57	57%	
CSTER	128	104880	105272	+	30S 核糖体蛋白 S9	100%	3.00E-59	68%	
CSTER	193	159363	160967	+	CTP 合成酶	99%	0.00E+00	68%	CTP 合酶 pyrg
CSTER	199	166439	166897	-	保守的假定蛋白	81%	2.00E-07	28%	蛋白 NrdI
CSTER	208	176203	176472	+	30S 核糖体蛋白 S15	100%	6.00E-38	63%	
CSTER	217	183028	184185	+	十一异戊二烯焦磷酸酯磷酸酶	98%	2.00E-68	38%	可能的十一异戊二烯焦磷酸酯 N-乙酰基葡萄糖胺 1-磷酸转移酶 tagO
CSTER	218	184346	185116	+	ABC 转运蛋白 ATP 酶	95%	8.00E-130	71%	植物蛋白 296 sufC
CSTER	219	185153	186415	+	假定蛋白	97%	4.00E-111	42%	FeS 簇组装蛋白 SufD
CSTER	220	186469	187701	+	推定的氨基酸转移酶 (第 V 类)	98%	0.00E+00	60%	半胱氨酸脱硫酶 SufS
CSTER	221	187688	188122	+	NifU 家族蛋白	95%	3.00E-48	51%	锌依赖的硫基转移酶 SufU
CSTER	245	208199	208993	+	磷脂酰胞苷酰转移酶	98%	3.00E-67	41%	磷脂酰胞苷酰转移酶
CSTER	247	210343	212205	+	脯氨酰-tRNA 合成酶	98%	0.00E+00	50%	
CSTER	252	215735	216022	+	辅伴侣蛋白 GroES (HSP10)	97%	1.00E-18	43%	
CSTER	253	216071	217690	+	伴侣蛋白 GroEL (HSP60 家族)	98%	0.00E+00	75%	

[0230]

基因组部分	STER	起始	终止	方向		查询覆盖率	e-值	氨基酸 Id	杜草芽孢杆菌注释
CSTER	261	224055	224204	+	50S 核糖体蛋白 L33	97%	9.00E-12	50%	
CSTER	262	224216	224392	+	前蛋白转位酶亚基 SecE	81%	1.40E+00	21%	前蛋白转位酶亚基 SecE
CSTER	268	227616	230117	+	亮氨酸-tRNA 合成酶	99%	0.00E+00	70%	
CSTER	273	232342	233796	+	烟酰胺磷酸核糖基转移酶	97%	0.00E+00	63%	烟酰胺磷酸核糖基转移酶 pncb
CSTER	274	233808	234629	+	NAD 合成酶	100%	9.00E-109	59%	NH(3)-依赖性 NAD(+) 合成酶 nade
CSTER	286	248252	249580	+	UDP-N-乙酰胞壁酰-丙氨酸连接酶	99%	3.00E-170	55%	UDP-N-乙酰胞壁酰-L-丙氨酸连接酶 murC
CSTER	302	260849	261664	+	谷氨酸消旋酶	94%	7.00E-83	48%	谷氨酸消旋酶 1 racE
CSTER	307	264328	265041	+	分离和集缩蛋白 A	93%	5.00E-37	39%	分离和集缩蛋白 A
CSTER	308	265034	265615	+	分离和集缩蛋白 B	94%	9.00E-36	40%	分离和集缩蛋白 B
CSTER	313	269788	270297	+	rRNA 甲基转移酶	96%	3.00E-58	51%	rRNA 甲基转移酶
CSTER	349	303983	305959	+	转酮醇酶	98%	0.00E+00	58%	转酮醇酶 tkt
CSTER	357	312857	314032	+	染色体复制起始/膜附着蛋白 DnaB	63%	3.00E-02	18%	复制起始和膜附着蛋白
CSTER	358	314036	314938	+	引发体蛋白 DnaI	99%	7.00E-58	35%	引发体蛋白 DnaI
CSTER	359	315044	316354	+	GTP-结合蛋白 EngA	100%	0.00E+00	67%	GTP 酶 Der
CSTER	368	321054	322394	-	丝氨酸-tRNA 合成酶	100%	0.00E+00	63%	

[0231]

基因组部分	STER	起始	终止	方向		查询覆盖率	e-值	氨基酸	注释
								赖氨酸	枯草芽孢杆菌注释
CSTER	376	329673	330116	+	保守的假定蛋白	89%	6.00E-33	43%	tRNA 苏氨酸甲酰腺苷生物合成蛋白 TsaE
CSTER	380	332545	333732	+	转录延伸因子 Ts	90%	2.00E-113	46%	转录终止/抗终止蛋白 NusA
CSTER	383	334363	337194	+	翻译起始因子 IF-2	99%	0.00E+00	55%	
CSTER	387	339864	341309	-	UDP-N-乙酰胞壁酰基丙氨酸-D-谷氨酸-2,6-二氨基庚二酸连接酶	92%	3.00E-52	29%	UDP-N-乙酰胞壁酰-L-丙氨酸-D-谷氨酸-2,6-二氨基庚二酸连接酶 murE
CSTER	419	364962	365894	+	推定的锰依赖性无机焦磷酸酶	99%	3.00E-125	58%	锰依赖性无机焦磷酸酶 ppac
CSTER	430	375355	375579	+	酰基载体蛋白	93%	5.00E-14	49%	酰基载体蛋白
CSTER	432	376667	377593	+	酰基载体蛋白 S-丙二酰转移酶	95%	3.00E-86	47%	丙二酰-CoA-酰基载体蛋白转酰基酶
					3-酮酰基-(酰基-载体-蛋白)还原酶	99%	6.00E-79	47%	3-氧化酰-[酰基-载体-蛋白]还原酶
CSTER	433	377606	378340	+	3-酮酰基-(酰基-载体-蛋白)还原酶	99%	6.00E-79	47%	FabG
CSTER	434	378401	379633	+	3-氧化酰-(酰基-载体-蛋白)合酶 II	99%	1.00E-133	48%	3-氧化酰-[酰基-载体-蛋白]合酶 2
					乙酰辅酶 A 羧化酶生物素羧基载体蛋白亚基	98%	9.00E-26	37%	乙酰辅酶 A 羧化酶的生物素羧基载体蛋白
CSTER	435	379637	380125	+	乙酰辅酶 A 羧化酶生物素羧基载体蛋白亚基	98%	9.00E-26	37%	生物素羧化酶 1
CSTER	437	380667	382037	+	乙酰辅酶 A 羧化酶生物素羧基载体蛋白亚基	98%	0.00E+00	60%	
CSTER	438	382043	382909	+	乙酰辅酶 A 羧化酶亚基 B	94%	8.00E-98	51%	乙酰辅酶 A 羧化酶转酰基酶亚基 B

[0232]

基因组-部分	STER	起始	终止	方向		查询 覆盖率	e-值	氨基酸	注释
CSTER	439	382906	383676	+	乙酰辅酶A羧化酶亚基 α	76%	2.00E-81	53%	杜草芽孢杆菌注释
CSTER	442	385819	387417	+	保守的假定蛋白	98%	0.00E+00	58%	乙酰辅酶A羧化酶亚基转移酶 α
CSTER	455	402156	402470	+	50S 核糖体蛋白 L21	99%	2.00E-44	66%	核糖核酸酶 Y ymda
CSTER	456	402507	402797	+	50S 核糖体蛋白 L27	95%	5.00E-47	79%	
CSTER	460	405038	405805	+	二氢吡啶二羧酸还原酶	98%	7.00E-95	53%	4-羟基-四氢吡啶二羧酸还原酶 dapb
CSTER	461	405802	407010	+	tRNA CCA-焦磷酸化酶	98%	6.00E-86	40%	CCA-添加酶
CSTER	475	422256	422813	+	核糖体再循环因子	100%	2.00E-68	56%	
CSTER	485	430547	432550	+	甲硫氨酰-tRNA 合成酶	98%	0.00E+00	58%	
CSTER	492	438748	439890	+	蛋白酶成熟蛋白前体	91%	6.00E-11	26%	折叠酶蛋白 Prsa
CSTER	493	440232	442850	+	丙氨酰-tRNA 合成酶	99%	0.00E+00	52%	
CSTER	513	460453	463104	+	缬氨酰-tRNA 合成酶	99%	0.00E+00	63%	
CSTER	523	469888	471168	+	细胞分裂蛋白 FtsW	93%	4.00E-59	34%	推定的脂质 II 翻转酶 FtsW
CSTER	524	471406	472602	+	延伸因子 Tu	99%	0.00E+00	76%	
CSTER	525	472851	473609	+	磷酸丙糖异构酶	97%	8.00E-109	62%	磷酸丙糖异构酶 tpia
CSTER	526	473848	474477	+	胸苷酸激酶	96%	4.00E-73	55%	胸苷酸激酶 tmk
CSTER	527	474486	475361	+	DNA 聚合酶 III 亚基 delta'	96%	8.00E-41	33%	DNA 聚合酶 III 亚基 delta'

[0233]

基因组-部分	STER	起始	终止	方向		查询 覆盖率	e-值	氨基酸 酸 Id	杜草芽孢杆菌注释
CSTER	539	484827	485744	+	甘氨酸-tRNA 合成酶, α 亚基	97%	4.00E-169	74%	
CSTER	540	486028	488064	+	甘氨酸-tRNA 合成酶, β 亚基	97%	0.00E+00	45%	
CSTER	567	512413	512916	+	50S 核糖体蛋白 L10	100%	7.00E-61	57%	
CSTER	568	512991	513359	+	50S 核糖体蛋白 L7/L12	100%	4.00E-38	67%	
CSTER	603	548926	550308	+	N-乙酰葡萄糖胺-1-磷酸尿苷转移酶	97%	3.00E-164	52%	双功能蛋白 G1mJ
CSTER	623	566278	566781	+	二氢叶酸还原酶	94%	8.00E-34	38%	二氢叶酸还原酶 dfra
CSTER	626	568210	568809	+	GTP 酶 EngB	96%	7.00E-92	63%	可能的 GTP-结合蛋白 EngB
CSTER	632	573613	574254	-	推定的甘油-3-磷酸酰基转移酶 PlsY	90%	3.00E-40	46%	甘油-3-磷酸酰基转移酶 plsY
CSTER	633	574392	576341	+	DNA 拓扑异构酶 IV 亚基 B	97%	0.00E+00	71%	DNA 拓扑异构酶 4 亚基 B
CSTER	634	576969	579434	+	DNA 拓扑异构酶 IV 亚基 A	95%	0.00E+00	54%	DNA 拓扑异构酶 4 亚基
CSTER	660	598871	599725	+	亚甲基四氢叶酸脱氢酶/ 亚甲基四氢叶酸环化水解酶	99%	2.00E-109	56%	双功能蛋白 FoID
CSTER	668	605570	606469	+	GTP 酶 Era	98%	5.00E-143	64%	GTP 酶 Era
CSTER	670	607313	607927	+	脱磷酸辅酶 A 激酶	95%	2.00E-50	45%	脱磷酸辅酶 A 激酶 coae
CSTER	672	609192	609338		50S 核糖体蛋白 L33	97%	1.00E-16	52%	
CSTER	684	620012	621316	+	烯醇酶	100%	0.00E+00	70%	烯醇酶 eno

[0234]

基因组_部分	STER	起始	终止	方向		查询 覆盖率	e-值	氨基酸 酸 Id	杜草芽孢杆菌注释
CSTER	731	664470	665690	-	推定的胞嘧啶-C5 特异性 DNA 甲基化酶	60%	5.00E-16	31%	可能的BsuMI 修饰甲基化酶YdiO
CSTER	733	666939	668429	-	赖氨酰-tRNA 合成酶 (II 类)	97%	0.00E+00	64%	
CSTER	761	694548	695582	+	DNA 聚合酶 III δ 亚基	84%	2.00E-37	33%	未表征的蛋白质 YqeN
CSTER	773	704704	706056	+	UDP-N-乙酰胞壁酰-L-丙氨酸-D-谷氨酰胺合成酶	98%	8.00E-147	49%	UDP-N-乙酰胞壁酰基丙氨酸-D-谷氨酰胺连接酶 murD
CSTER	774	706060	707130	+	N-乙酰葡糖胺基转移酶	95%	3.00E-30	27%	UDP-N-乙酰葡糖胺-N-乙酰胞壁酰-(五肽) 焦磷酸十一-异戊烯醇 N-乙酰葡糖胺转移酶 murG
CSTER	775	707140	708264	+	细胞分裂蛋白 FtsQ	79%	4.00E-15	26%	细胞分裂蛋白 DivIB
CSTER	776	708388	709764	+	细胞分裂蛋白 FtsA	84%	4.00E-103	44%	细胞分裂蛋白 FtsA
CSTER	777	709793	711115	+	细胞分裂蛋白 FtsZ	92%	2.00E-123	53%	细胞分裂蛋白 FtsZ
CSTER	783	714803	717592	+	异亮氨酰-tRNA 合成酶	99%	0.00E+00	58%	
CSTER	787	720193	720459	-	50S 核糖体蛋白 L31	100%	1.00E-22	50%	
CSTER	793	725315	726394	+	肽链释放因子 1	99%	6.00E-158	59%	
CSTER	796	727911	729161	+	丝氨酸羟甲基转移酶	97%	2.00E-171	60%	丝氨酸羟甲基转移酶 glyA

[0235]

基因组-部分	STER	起始	终止	方向	查询 覆盖率	e-值	氨基酸 Id	
CSTER	833	760678	762396	-	98%	0.00E+00	47%	杜草芽孢杆菌注释 磷酸葡萄糖变位酶 pgm
CSTER	850	783500	783736	-	94%	1.00E-13	46%	
CSTER	864	797961	799307	+	99%	4.00E-180	56%	
CSTER	903	833853	835661	+	100%	0.00E+00	59%	谷氨酰胺—果糖-6-磷酸氨基转移酶 [异构化] glms
CSTER	915	845349	846911	+	95%	2.00E-178	57%	信号识别颗粒蛋白
CSTER	919	848991	849845	+	98%	8.00E-101	50%	核糖体生物发生 GTP 酶 rbgA
CSTER	923	852605	854749	+	98%	0.00E+00	64%	DNA 拓扑异构酶 I
CSTER	994	915311	917623	+	99%	0.00E+00	54%	ATP-依赖性 DNA 解旋酶 PcrA
CSTER	1034	959843	961231	-	99%	5.00E-69	34%	2-氧代戊二酰脱氢酶复合体的二氢硫 辛酰赖氨酸残基琥珀酰转移酶组分 odhb
CSTER	1036	962412	963383	-	81%	3.00E-33	28%	丙酮酸脱氢酶 E1 组分亚基 α pdha
CSTER	1087	1007529	1007888	-	85%	2.00E-56	80%	
CSTER	1088	1007945	1008145	-	100%	1.00E-23	62%	

[0236]

基因组部分	STER	起始	终止	方向		查询 覆盖率	e-值	氨基酸 Id	杜草芽孢杆菌注释
CSTER	1089	1008184	1008714	-	翻译起始因子 3 (IF-3)	98%	2.00E-63	60%	
CSTER	1102	1017867	1018880	-	肽链释放因子 2	91%	2.00E-137	57%	
CSTER	1115	1029796	1031136	-	感受器组氨酸激酶	97%	4.00E-117	71%	转录调节蛋白 YycF
CSTER	1116	1031129	1031836	-	双组分应答调控蛋白	68%	2.00E-120	47%	感受器组氨酸激酶 YycG
CSTER	1123	1036555	1037826	-	UDP-N-乙酰葡萄糖胺 1-羧基乙烯基转移酶	96%	2.00E-172	59%	UDP-N-乙酰葡萄糖胺 1-羧基乙烯基转移酶 1 muraa
CSTER	1135	1049456	1050649	-	S-腺苷甲硫氨酸合成酶	97%	0.00E+00	67%	S-腺苷甲硫氨酸合成酶 metK
CSTER	1136	1050929	1051867	+	生物素-(乙酰辅酶A羧化酶)连接酶	92%	4.00E-44	33%	双功能连接酶/阻遏蛋白 BirA
CSTER	1138	1052266	1053918	-	DNA 聚合酶 III, γ /tau 亚基	93%	3.00E-137	44%	DNA 聚合酶 III 亚基 γ /tau DnaX
CSTER	1144	1056546	1056893	-	50S 核糖体蛋白 L19	100%	2.00E-61	77%	
CSTER	1164	1073225	1074244	-	6-磷酸果糖激酶	93%	4.00E-124	61%	ATP 依赖性 6-磷酸果糖激酶 pfka
CSTER	1165	1074337	1077447	-	DNA 聚合酶 III, α 亚基 DnaE	96%	0.00E+00	35%	DNA 聚合酶 III 亚基 α dnae
CSTER	1166	1077600	1078385	-	推定的翻译因子 (SUA5)	42%	3.00E-04	24%	yw1C 具有推定的 RNA 结合基序 TW 的未知保守蛋白
CSTER	1175	1085950	1086225	-	组蛋白样 DNA 结合蛋白	95%	4.00E-40	72%	DNA 结合蛋白 HU 1
CSTER	1182	1091629	1092501	-	香叶基香叶基焦磷酸合酶	89%	5.00E-58	44%	法呢酰基-二磷酸合酶 ispa yqld

[0237]

基因组部分	STER	起始	终止	方向	查询覆盖率	e-值	氨基酸	杜草芽孢杆菌注释
CSTER	1188	1095764	1096462	-	85%	1.00E-07	22%	DNA 复制蛋白 DnaD
CSTER	1193	1101272	1102201	-	99%	2.00E-105	50%	核糖核酸酶 Z rnz
CSTER	1197	1105450	1106877	-	88%	2.00E-57	38%	杆状决定蛋白 RodA
CSTER	1230	1131390	1132742	-	100%	0.00E+00	63%	磷酸葡萄糖胺变位酶 ybbt glmM
CSTER	1248	1154457	1156616	+	98%	0.00E+00	47%	核糖核苷酸二磷酸还原酶 α nrde
CSTER	1249	1156775	1157737	+	97%	8.00E-98	49%	核糖核苷酸二磷酸还原酶 β nrdf
CSTER	1256	1161831	1164284	-	98%	0.00E+00	62%	DNA 旋转酶亚基 A
CSTER	1267	1174041	1176449	-	99%	0.00E+00	47%	
CSTER	1270	1177285	1178328	-	98%	3.00E-161	62%	
CSTER	1271	1178950	1182483	-	99%	0.00E+00	37%	染色体分配蛋白 Smc
CSTER	1272	1182486	1183175	-	89%	4.00E-68	47%	核糖核酸酶 3
CSTER	1273	1183332	1184267	-	93%	5.00E-88	47%	4-羧基-四氢吡啶二羧酸合酶 dapA
CSTER	1274	1184605	1185681	-	99%	1.00E-143	58%	天冬氨酸-半醛脱氢酶 asd
CSTER	1382	1293065	1294063	-	97%	7.00E-113	50%	铁氧还蛋白—NADP 还原酶 2 yunc

[0238]

基因组-部分	STER	起始	终止	方向		查询 覆盖率	e-值	氨基酸 Id	杜草芽孢杆菌注释
CSTER	1383	1294065	1294784	-	tRNA (鸟嘌呤-N(1)-)-甲基转移酶	100%	6.00E-87	56%	tRNA (鸟嘌呤-N(1)-)-甲基转移酶
CSTER	1395	1303326	1304261	-	甲硫氨酸-tRNA 甲酰基转移酶	96%	5.00E-98	49%	
CSTER	1396	1304279	1306675	-	引发体装配蛋白 PriA	99%	0.00E+00	48%	引发体蛋白 N PriA
CSTER	1398	1307153	1307782	-	鸟苷酸激酶	98%	5.00E-87	59%	鸟苷酸激酶 gmk
CSTER	1399	1308015	1309406	-	细胞分裂蛋白 FtsY	97%	8.00E-121	55%	信号识别颗粒受体 FtsY
CSTER	1422	1330503	1331339	-	无机磷酸/ATP-NAD 激酶	100%	5.00E-78	44%	NAD 激酶 1 ppnk
CSTER	1425	1333019	1333990	+	磷酸核糖焦磷酸激酶	100%	4.00E-113	54%	磷酸核糖焦磷酸激酶 prs
CSTER	1426	1333994	1335106	+	推定的半胱氨酸脱羧酶	98%	2.00E-115	47%	推定的半胱氨酸脱羧酶 IscS
CSTER	1448	1355769	1356878	-	RNA 聚合酶 σ 因子 RpoD	93%	6.00E-167	70%	RNA 聚合酶 σ 因子 SigA
CSTER	1449	1356882	1358693	-	DNA 引发酶	68%	9.00E-86	37%	DNA 引发酶 DnaG
CSTER	1451	1359076	1359252	-	30S 核糖体蛋白 S21	98%	3.00E-29	89%	
CSTER	1464	1371306	1372619	-	GTP 酶 ObgE	100%	0.00E+00	67%	GTP 酶 ObgE
CSTER	1480	1387053	1389005	-	DNA 旋转酶亚基 B	100%	0.00E+00	68%	DNA 旋转酶亚基 B
CSTER	1497	1401241	1402143	-	UDP-N-乙酰烯醇式丙酮酸葡萄糖胺还原酶	98%	1.00E-66	40%	UDP-N-乙酰烯醇式丙酮酸葡萄糖胺还原酶 murB
CSTER	1506	1409996	1410268	-	30S 核糖体蛋白 S16	100%	5.00E-42	67%	

[0239]

基因组部分	STER	起始	终止	方向		查询覆盖率	e-值	氨基酸 Id	氨基酸注释
CSTER	1512	1415065	1416078	-	推定的脂酶	95%	1. 00E-96	51%	枯草芽孢杆菌注释
CSTER	1513	1416088	1418034	-	NAD-依赖性 DNA 连接酶	97%	0. 00E+00	55%	二酰基甘油激酶 dagK
CSTER	1516	1419659	1420519	-	甲硫氨酸氨肽酶	99%	6. 00E-56	37%	DNA 连接酶 LigA
CSTER	1519	1422447	1423709	-	UDP-N-乙酰葡萄糖胺 1-羧基乙烯基转移酶	96%	6. 00E-125	46%	UDP-N-乙酰葡萄糖胺 1-羧基乙烯基转移酶 1 muraa
CSTER	1522	1426828	1427583	-	1-酰基-sn-甘油-3-磷酸酰基转移酶	92%	1. 00E-29	32%	1-酰基-sn-甘油-3-磷酸酰基转移酶 plsC
CSTER	1534	1439747	1441120	-	UDP-N-乙酰胞壁酰-三肽-D-丙氨酸-D-丙氨酸连接酶	98%	2. 00E-98	38%	UDP-N-乙酰胞壁酰-三肽-D-丙氨酸-D-丙氨酸连接酶 murF
CSTER	1544	1447537	1448583	-	D-丙氨酸-丙氨酸合成酶 A	98%	2. 00E-95	42%	D-丙氨酸-D-丙氨酸连接酶 ddl
CSTER	1583	1483475	1484107	-	烟酸单核苷酸腺苷酰转移酶	98%	2. 00E-62	44%	烟酸核苷酸腺苷酰转移酶 nadd
CSTER	1584	1484211	1484528	-	推定的 RNA 结合蛋白	95%	1. 00E-24	50%	可能的 RNA 结合蛋白 YqeI
CSTER	1585	1484767	1485885	-	GTP 结合蛋白 YqeH	100%	6. 00E-163	59%	未表征的蛋白质 YqeH
CSTER	1590	1488021	1489463	-	天冬氨酸/谷氨酰-tRNA 酰基转移酶 亚基 B	99%	0. 00E+00	63%	

[0240]

基因组部分	STER	起始	终止	方向		查询覆盖率	e-值	氨基酸 Id	杜草芽孢杆菌注释
CSTER	1591	1489463	1490929	-	天冬氨酸/谷氨酰-tRNA 酰胺转移酶 亚基 A	99%	0.00E+00	59%	
CSTER	1592	1490929	1491231	-	天冬氨酸/谷氨酰-tRNA 酰胺转移酶 亚基 C	97%	1.00E-18	41%	
CSTER	1615	1511378	1512298	-	硫氧还蛋白还原酶	95%	3.00E-126	58%	硫氧还蛋白还原酶 trxb
CSTER	1665	1554189	1555211	-	磷酸-N-乙酰胞壁酰-五肽-转移酶	92%	1.00E-66	41%	磷酸-N-乙酰胞壁酰-五肽-转移酶 mray
CSTER	1666	1555213	1557480	-	推定的青霉素结合蛋白 2X	90%	9.00E-88	32%	青霉素结合蛋白 2B
CSTER	1667	1557484	1557804	-	细胞分裂蛋白 FtsL				
CSTER	1701	1591284	1592387	-	丙氨酸消旋酶	93%	1.00E-98	43%	丙氨酸消旋酶 1 alr
CSTER	1702	1592408	1592767	-	推定的 4'-磷酸泛酰巯基乙胺基转移酶	97%	6.00E-29	41%	全[酰基-载体-蛋白]合酶
CSTER	1705	1594955	1597504	-	前蛋白转位酶亚基 SecA	100%	0.00E+00	55%	蛋白质转位酶亚基 SecA
CSTER	1726	1616337	1616576	-	30S 核糖体蛋白 S18	100%	5.00E-32	65%	
CSTER	1727	1616618	1617136	-	单链 DNA 结合蛋白	93%	1.00E-41	56%	单链 DNA 结合蛋白 B
CSTER	1728	1617148	1617438	-	30S 核糖体蛋白 S6	96%	4.00E-40	61%	

[0241]

基因组-部分	STER	起始	终止	方向		查询 覆盖率	e-值	氨基酸 酸 Id	
CSTER	1745	1633808	1634821	-	推定的O-唾液酸糖蛋白内肽酶	97%	6.00E-130	55%	杜草芽孢杆菌注释 tRNA N6-腺苷苏氨酸酰基甲酰基转移酶 tsad
CSTER	1747	1635236	1635922	-	推定的糖蛋白内肽酶	100%	6.00E-44	38%	tRNA 苏氨酸酰基甲酰基腺苷生物合成蛋 白TsaB
CSTER	1749	1636421	1638103	+	mRNA 降解核糖核酸酶 J1/J2	99%	0.00E+00	62%	核糖核酸酶 J1 yqkc
CSTER	1755	1642501	1643700	-	磷酸甘油酸激酶	100%	2.00E-124	50%	磷酸甘油酸激酶 pgk
CSTER	1762	1648240	1650321	-	翻译延伸因子 G	100%	0.00E+00	77%	
CSTER	1763	1650542	1651012	-	30S 核糖体蛋白 S7	100%	1.00E-86	75%	
CSTER	1764	1651031	1651444	-	30S 核糖体蛋白 S12	98%	3.00E-80	86%	
CSTER	1770	1656078	1656950	-	核糖体相关的 GTP 酶	99%	5.00E-105	51%	推定的核糖体生物发生 GTP 酶 RsgA
CSTER	1776	1660099	1660413	-	推定的硫氧还蛋白	92%	1.00E-37	61%	硫氧还蛋白 trxa
CSTER	1787	1670058	1670192	-	50S 核糖体蛋白 L34	100%	1.00E-15	70%	
CSTER	1790	1673060	1673395	-	核糖核酸酶 P	95%	1.00E-31	47%	核糖核酸酶 P 蛋白组分
CSTER	1793	1677296	1678750	-	谷氨酰-tRNA 合成酶	100%	0.00E+00	55%	
CSTER	1797	1681106	1681795	-	50S 核糖体蛋白 L1	98%	4.00E-99	65%	50S 核糖体蛋白 L1
CSTER	1809	1690036	1691058	+	甘油-3-磷酸脱氢酶	94%	4.00E-122	53%	甘油-3-磷酸脱氢酶[NAD(P)+] gspA

[0242]

基因组部分	STER	起始	终止	方向		查询覆盖率	e-值	氨基酸 Id	杜草芽孢杆菌注释
CSTER	1813	1693298	1694431	-	金属依赖性酰胺酶/氨基酰化酶/羧肽酶	95%	6.00E-120	47%	N-乙酰二氨基庚二酸脱乙酰酶 ykur
CSTER	1814	1694510	1695208	-	推定的 2, 3, 4, 5-四氢吡啶-2-羧酸 N-琥珀酰转移酶	98%	4.00E-86	60%	2, 3, 4, 5-四氢吡啶-2, 6-二羧酸 N-乙酰转移酶 dapH
CSTER	1821	1699209	1699601	-	单链结合蛋白	93%	1.00E-25	40%	单链 DNA 结合蛋白 B
CSTER	1844	1713286	1716924	-	DNA-指导的 RNA 聚合酶亚基 β'	97%	0.00E+00	68%	DNA-指导的 RNA 聚合酶亚基 β
CSTER	1845	1717025	1720606	-	DNA-指导的 RNA 聚合酶亚基 β	98%	0.00E+00	71%	DNA-指导的 RNA 聚合酶亚基 β'
CSTER	1847	1723493	1724749	+	酪氨酸-tRNA 合成酶	99%	0.00E+00	58%	
CSTER	1876	1751169	1752050	-	果糖-二磷酸醛缩酶	100%	4.00E-83	45%	可能的果糖-二磷酸醛缩酶 fbaa
CSTER	1880	1755102	1755488	-	50S 核糖体蛋白 L17	100%	2.00E-62	73%	
CSTER	1881	1755506	1756444	-	DNA-指导的 RNA 聚合酶 α 亚基	99%	7.00E-139	62%	DNA-指导的 RNA 聚合酶亚基 α
CSTER	1882	1756493	1756876	-	30S 核糖体蛋白 S11	93%	7.00E-69	87%	
CSTER	1883	1756904	1757269	-	30S 核糖体蛋白 S13	100%	2.00E-58	73%	
CSTER	1884	1757290	1757403	-	50S 核糖体蛋白 L36	100%	4.00E-16	84%	
CSTER	1885	1757429	1757647	-	翻译起始因子 1 (IF-1)	100%	1.00E-38	78%	

[0243]

基因组-部分	STER	起始	终止	方向		查询 覆盖率	e-值	氨基酸 酸 Id	
CSTER	1886	1757765	1758421	-	腺苷酸激酶	99%	3.00E-89	59%	枯草芽孢杆菌注释
CSTER	1887	1758553	1759848	-	前蛋白转位酶亚基 SecY	99%	4.00E-140	49%	腺苷酸激酶 adk
CSTER	1888	1759865	1760305	-	50S 核糖体蛋白 L15	100%	1.00E-70	72%	蛋白转位酶亚基 SecY
CSTER	1889	1760433	1760615	-	50S 核糖体蛋白 L30	98%	3.00E-20	62%	
CSTER	1890	1760630	1761124	-	30S 核糖体蛋白 S5	93%	8.00E-70	76%	
CSTER	1891	1761143	1761499	-	50S 核糖体蛋白 L18	100%	4.00E-55	73%	
CSTER	1892	1761589	1762125	-	50S 核糖体蛋白 L6	100%	3.00E-69	60%	50S 核糖体蛋白 L6
CSTER	1893	1762252	1762650	-	30S 核糖体蛋白 S8	100%	1.00E-73	77%	
CSTER	1894	1762773	1762958	-	30S 核糖体蛋白 S14	100%	1.00E-31	77%	
CSTER	1895	1762976	1763518	-	50S 核糖体蛋白 L5	99%	3.00E-100	77%	
CSTER	1896	1763545	1763850	-	50S 核糖体蛋白 L24	99%	7.00E-38	64%	
CSTER	1897	1763931	1764299	-	50S 核糖体蛋白 L14	100%	5.00E-74	87%	
CSTER	1898	1764324	1764584	-	30S 核糖体蛋白 S17	97%	4.00E-46	84%	
CSTER	1899	1764612	1764818	-	50S 核糖体蛋白 L29	78%	1.00E-14	58%	
CSTER	1900	1764828	1765241	-	50S 核糖体蛋白 L16	93%	1.00E-76	83%	
CSTER	1901	1765245	1765898	-	30S 核糖体蛋白 S3	100%	1.00E-119	75%	

[0244]

基因组部分	STER	起始	终止	方向		查询 覆盖率	e-值	氨基酸 Id	枯草芽孢杆菌注释
CSTER	1902	1765911	1766255	-	50S 核糖体蛋白 L22	96%	7.00E-49	64%	
CSTER	1903	1766271	1766549	-	30S 核糖体蛋白 S19	100%	2.00E-54	83%	
CSTER	1904	1766643	1767476	-	50S 核糖体蛋白 L2	100%	5.00E-159	76%	50S 核糖体蛋白 L2
CSTER	1905	1767494	1767790	-	50S 核糖体蛋白 L23	94%	2.00E-31	60%	
CSTER	1906	1767790	1768413	-	50S 核糖体蛋白 L4	100%	9.00E-92	60%	50S 核糖体蛋白 L4
CSTER	1907	1768438	1769064	-	50S 核糖体蛋白 L3	99%	4.00E-115	75%	50S 核糖体蛋白 L3
CSTER	1908	1769181	1769489	-	30S 核糖体蛋白 S10	100%	1.00E-58	78%	
CSTER	1936	1795296	1795484	+	50S 核糖体蛋白 L28	100%	7.00E-26	69%	
CSTER	1948	1805428	1807179	-	天冬氨酸-tRNA 合成酶	98%	0.00E+00	57%	
CSTER	1950	1807719	1808999	-	组氨酸-tRNA 合成酶	99%	1.00E-140	49%	
CSTER	1953	1810651	1810833	+	50S 核糖体蛋白 L32	91%	1.00E-10	48%	
CSTER	1954	1810849	1810998	+	50S 核糖体蛋白 L33	100%	1.00E-14	55%	
CSTER	1973	1823718	1824329	-	30S 核糖体蛋白 S4	100%	1.00E-95	69%	
CSTER	1975	1824852	1826213	-	复制型 DNA 解旋酶	95%	8.00E-176	56%	复制型 DNA 解旋酶 DnaC
CSTER	1976	1826257	1826715	-	50S 核糖体蛋白 L9	100%	2.00E-39	55%	

[0245]

基因组部分	STER	起始	终止	方向		查询覆盖率	e-值	氨基酸 Id	枯草芽孢杆菌注释
CSTER	1979	1830931	1832052	-	tRNA (5-甲基氨(甲基-2-硫尿苷酸))-甲基转移酶	98%	9.00E-177	70%	tRNA 特异性 2-硫尿苷酸酶 MrraA
CSTER	1986	1838183	1838725	-	磷脂酰甘油磷酸合酶	93%	2.00E-45	52%	CDP-二酰基甘油-甘油-3-磷酸 3-磷酸脂酰转移酶 pgsA
CSTER	1992	1843910	1845391	-	肌醇-5-单磷酸脱氢酶	99%	0.00E+00	68%	肌醇-5'-单磷酸脱氢酶 guab
CSTER	1993	1845568	1846590	-	色氨酸-tRNA 合成酶 II	98%	8.00E-42	35%	

[0246]

表 5. 缺失菌株 GEI4 中鉴定的差异表达基因的列表.

操纵子	推定的途径/功能	调节子-效应子	基因	log ₂ 改变	p-值
STER_0390-STER_0393	半胱氨酸代谢	Cmbr/Homr/f/kr	STER_0390	-1.1	2.7E-23
			STER_0391	-1.3	2.2E-22
			STER_0392	-1.4	1.2E-32
			STER_0393	-1.2	2.8E-15
STER_1016-STER_1017	麦芽糊精代谢		STER_1016	-1.0	0
			STER_1017	-0.8	7.50E-38
STER_1548-STER_1555	芳族氨基酸生物合成	T-盒 RNA(Trp)	STER_1548	-1.3	2.3E-37
			STER_1549	-1.0	1.5E-25
			STER_1550	-1.3	7.9E-38
			STER_1551	-1.3	1.2E-35
			STER_1552	-1.3	0
			STER_1553	-1.4	0
			STER_1554	-1.3	0
			STER_1555	-1.6	7.8E-73
STER_1960-STER_1963	膜蛋白		STER_1960	-0.9	0
			STER_1961	-1.0	0
			STER_1962	-1.0	8.3E-317
			STER_1963	-1.1	3.2E-965
STER_0049-STER_0054	嘌呤生物合成	PurR	STER_0049	1.2	5.5E-34
			STER_0050	1.1	1.4E-29
			STER_0051	1.1	0
			STER_0052	1.1	0
			STER_0053	1.1	0
			STER_0054	1.1	6.5E-40
STER_0698-STER_0701	乙醇胺代谢		STER_0698	1.6	0
			STER_0700	2.1	9.5E-200
			STER_0701	2.2	7.6E-329
STER_1020-STER_1024	双精氨酸转位酶	TatA	STER_1020	1.3	5.3E-14
			STER_1021	1.3	8.8E-12
			STER_1022	1.4	3.3E-12
			STER_1023	1.0	9.6E-6
			STER_1024	0.9	0.0022
STER_1025-STER_1028	铁稳态	PerR	STER_1025	1.1	1.9E-09
			STER_1026	0.9	2.5E-6
			STER_1027	1.2	3E-12
			STER_1028	1.2	1.5E-17
STER_1405-STER_1409	ABC 肽转运	Cody	STER_1405	1.6	9.5E-90
			STER_1406	1.6	0
			STER_1407	1.6	3.7E-59
			STER_1408	1.5	0
			STER_1409	1.4	4.2E-54
STER_1821-STER_1823	应激		STER_1821	2.0	0
			STER_1822	2.2	5.4E-98
			STER_1823	2.0	0

[0247] 参考文献

- [0248] 1. Darmon E, Leach DF. (2014) Bacterial Genome Instability. *Microbiol. Mol. Biol. Rev.* 78, 1-39.
- [0249] 2. Labrie SJ, Samson JE 和 Moineau S. (2010) Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.* 8, 317-327.
- [0250] 3. Barrangou R, Marraffini LA (2014) CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. *Mol Cell* 54(2):234-244.
- [0251] 4. Barrangou R 等人 (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315(5819):1709-1712.
- [0252] 5. Brouns SJJ 等人 (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321(5891):960-964.
- [0253] 6. Young JC 等人 (2012) Phage-induced expression of CRISPR-associated proteins is revealed by shotgun proteomics in *Streptococcus thermophilus*. *PLoS ONE* 7(5):e38077.
- [0254] 7. Garneau JE 等人 (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468(7320):67-71.
- [0255] 8. Groenen PM, Bunschoten AE, Van Soolingen D, Van Embden JD (1993) Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. *Mol Microbiol* 10(5):1057-1065.
- [0256] 9. Yin S 等人 (2013) The evolutionary divergence of Shiga toxin-producing *Escherichia coli* is reflected in clustered regularly interspaced short palindromic repeat (CRISPR) spacer composition. *Appl Environ Microbiol* 79(18):5710-5720.
- [0257] 10. Liu F 等人 (2011) Novel virulence gene and clustered regularly interspaced short palindromic repeat (CRISPR) multilocus sequence typing scheme for subtyping of the major serovars of *Salmonella enterica* subsp. *enterica*. *Appl Environ Microbiol* 77(6):1946-1956.
- [0258] 11. Barrangou R, Horvath P (2012) CRISPR: new horizons in phage resistance and strain identification. *Annu Rev Food Sci Technol* 3, 143-162.
- [0259] 12. Sander JD 和 Joung JK. (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* 32, 347-355.
- [0260] 13. Bikard D 等人 (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res* 41(15):7429-7437
- [0261] 14. Qi LS 等人 (2013) Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* 152(5):1173-1183.
- [0262] 15. Goma AA 等人 (2014) Programmable Removal of Bacterial Strains by Use of Genome-Targeting CRISPR-Cas Systems. *mBio* 5(1):e00928-13.
- [0263] 16. Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A (1987) Nucleotide

sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J Bacteriol* 169(12):5429-5433.

[0264] 17. Jansen R, Embden JDA van, Gaastra W, Schouls LM (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* 43(6):1565-1575.

[0265] 18. Bolotin A 等人 (2004) Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. *Nat Biotechnol* 22(12):1554-1558.

[0266] 19. Horvath P 等人 (2008) Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J Bacteriol* 190(4):1401-1412.

[0267] 20. Bolotin A, Quinguis B, Sorokin A, Ehrlich SD (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 151(Pt 8):2551-2561.

[0268] 21. Deveau H 等人 (2008) Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J Bacteriol* 190(4):1390-1400.

[0269] 22. Paez-Espino D 等人 (2013) Strong bias in the bacterial CRISPR elements that confer immunity to phage. *Nat Commun* 4, 1430.

[0270] 23. Sun CL 等人 (2013) Phage mutations in response to CRISPR diversification in a bacterial population. *Environ Microbiol* 15(2):463-470.

[0271] 24. Sapranauskas R 等人 (2011) The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res* 39(21):9275-9282.

[0272] 25. Gasiunas G, Barrangou R, Horvath P, Siksnys V (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci USA* 109(39):E2579-2586.

[0273] 26. Briner AE 等人 (2014) Guide RNA Functional Modules Direct Cas9 Activity and Orthogonality. *Molecular Cell*. 56(2):333-339

[0274] 27. Bondy-Denomy J, Davidson AR. (2014) To acquire or resist: the complex biological effects of CRISPR-Cas systems. *Trends Microbiol*. 22, 218-225.

[0275] 28. Horvath P 等人 (2009) Comparative analysis of CRISPR loci in lactic acid bacteria genomes. *Int J Food Microbiol* 131(1):62-70.

[0276] 29. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. (2013a) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* 31(3):233-239.

[0277] 30. Vercoe RB 等人 (2013) Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape bacterial genomes and expel or remodel pathogenicity islands. *PLoS Genet* 9(4):e1003454.

[0278] 31. Oh JH, van Pijkeren JP. (2014) CRISPR-Cas9-assisted recombineering in

Lactobacillus reuteri. Nucleic Acids Res 10.1093/nar/gku623

[0279] 32. Selle K, Barrangou R. (2015) Harnessing CRISPR-Cas systems for bacterial genome editing. Trends Microbiol. In press

[0280] 33. Kobayashi K 等人 (2003) Essential Bacillus subtilis genes. Proc. Natl. Acad. Sci. U.S.A. 100, 4678-4683.

[0281] 34. Mahillon J, Chandler M. (1998) Insertion sequences. Microbiol Mol Biol Rev 62(3):725-774.

[0282] 35. Goh YJ, Goin C, O'Flaherty S, Altermann E, Hutkins R (2011) Specialized adaptation of a lactic acid bacterium to the milk environment; the comparative genomics of Streptococcus thermophilus LMD-9. Microbial Cell Factories 10 (Suppl 1):S22

[0283] 36. Dandoy D 等人 (2011) The fast milk acidifying phenotype of Streptococcus thermophilus can be acquired by natural transformation of the genomic island encoding the cell-envelope proteinase PrtS. Microb. Cell Fact. 10 Suppl 1, S21.

[0284] 37. Deltcheva E 等人 (2011) CRISPR RNA maturation by trans-activating small RNA and host factor RNase III. Nature 471, 602-607.

[0285] 38. Aravind L, Koonin EV (2001) Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. Genome Res 11(8):1365-1374.

[0286] 39. Koonin EV, Makarova KS (2007) Evolutionary genomics of lactic acid bacteria. J Bacteriol 189(4):1199-1208

[0287] 40. Makarova KS 等人 (2006) Comparative genomics of the lactic acid bacteria. Proc Natl Acad Sci USA 103(42):15611-15616.

[0288] 41. Goh YJ 等人 (2009) Development and application of a upp-based counterselective gene replacement system for the study of the S-layer protein SlpX of Lactobacillus acidophilus NCFM. Appl Environ Microbiol 75(10):3093-3105.

[0289] 42. Horton RM, Hunt HD, Ho SN, Pullen JK, Pease LR (1989) Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. Gene 77(1):61-68.

[0290] 43. Mojica FJM, Díez-Villaseñor C, García-Martínez J, Almendros C (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology 155(Pt3):733-740.

[0291] 44. Russell WM, Klaenhammer TR (2001) Efficient System for Directed Integration into the Lactobacillus acidophilus and Lactobacillus gasseri Chromosomes via Homologous Recombination. Appl Environ Microbiol 67(9):4361-4364.

[0292] 45. Wei M-Q 等人 (1995) An improved method for the transformation of

Lactobacillus strains using electroporation. Journal of Microbiological Methods 21 (1):97-109.

[0293] 46. Zhang X, Bremer H (1995) Control of the Escherichia coli rrnB p1 promoter strength by ppGpp. Journal of Biological Chemistry 270 (19):11181-11189.

[0294] 47. Law J 等人 (1995) A system to generate chromosomal mutations in Lactococcus lactis which allows fast analysis of targeted genes. J Bacteriol 177 (24):7011-7018.

[0295] 上文示例说明了本发明并且不应理解为对本发明的限制。本发明由下面的权利要求限定, 其中权利要求的等同物包括在权利要求中。

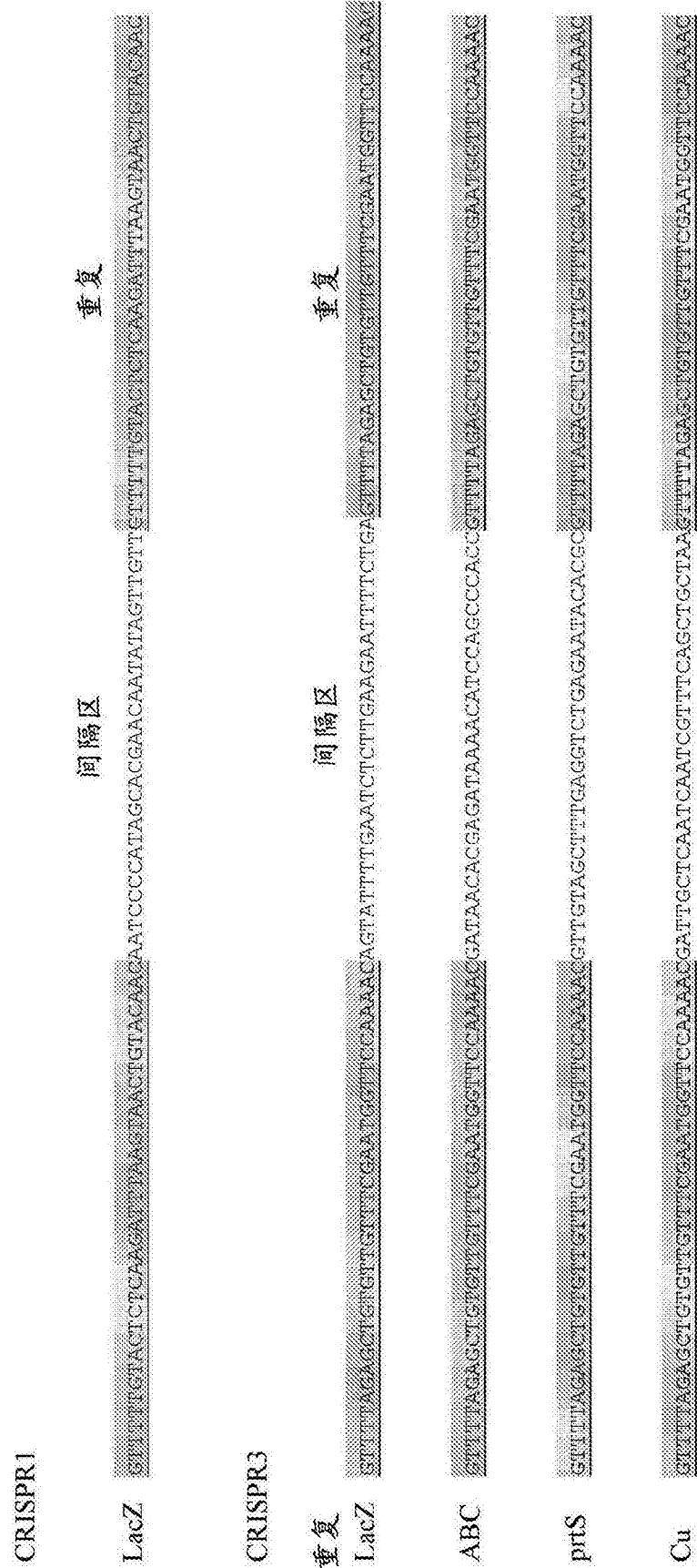


图1

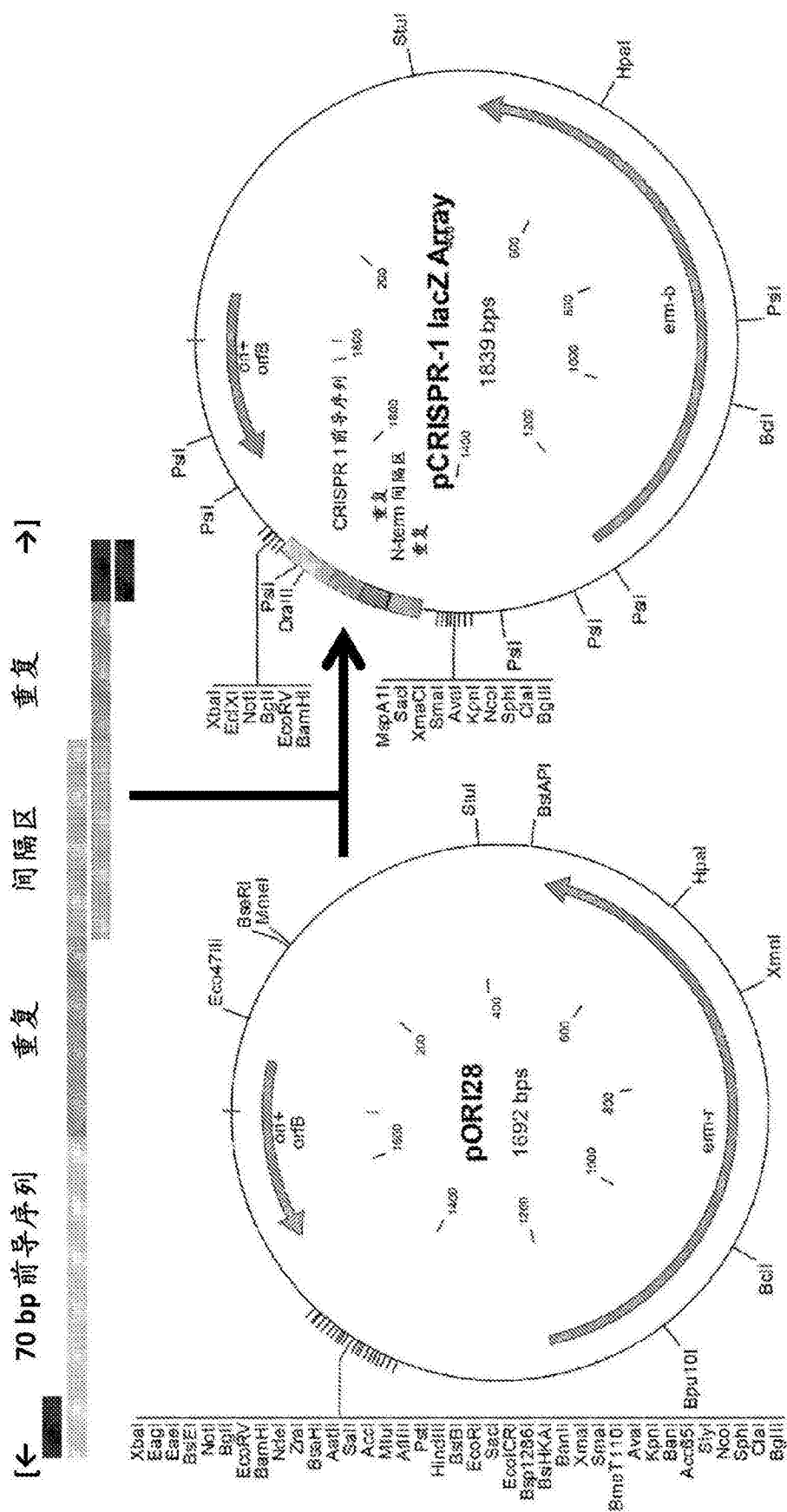


图2

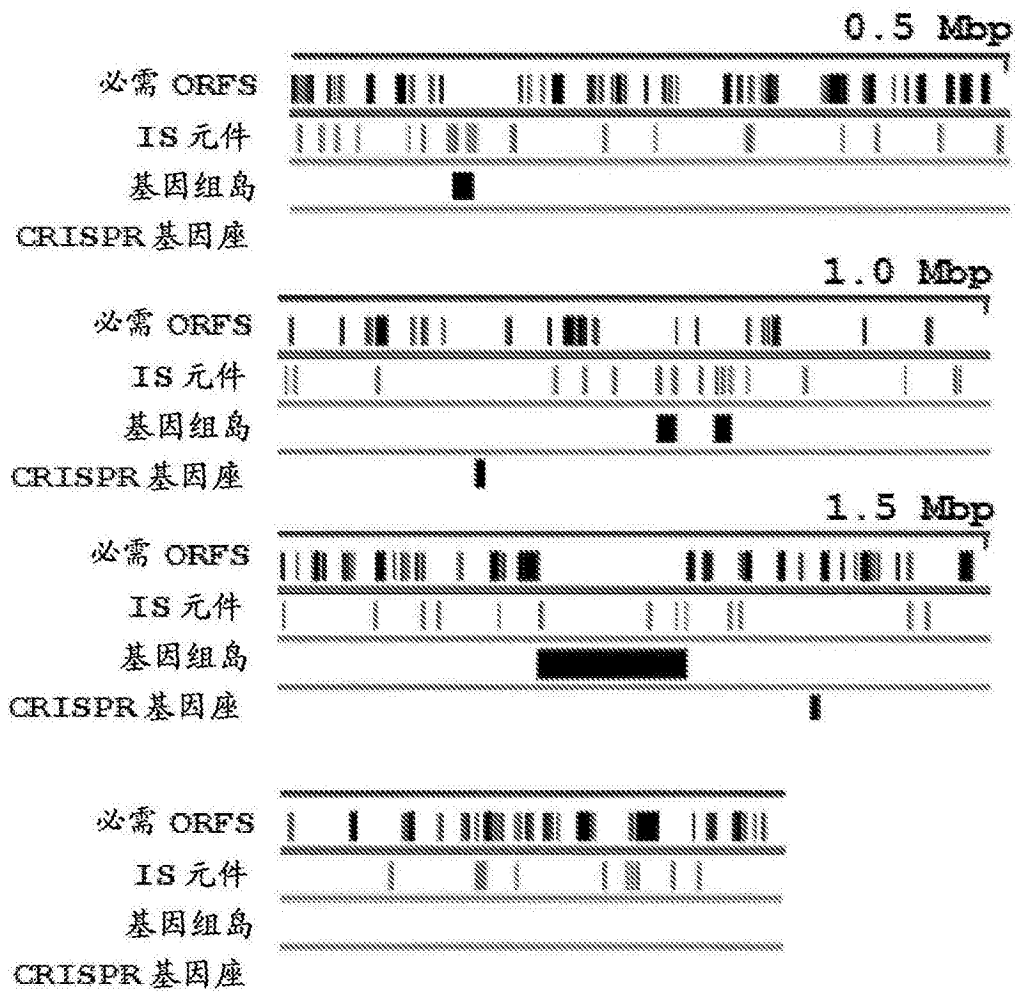


图3A

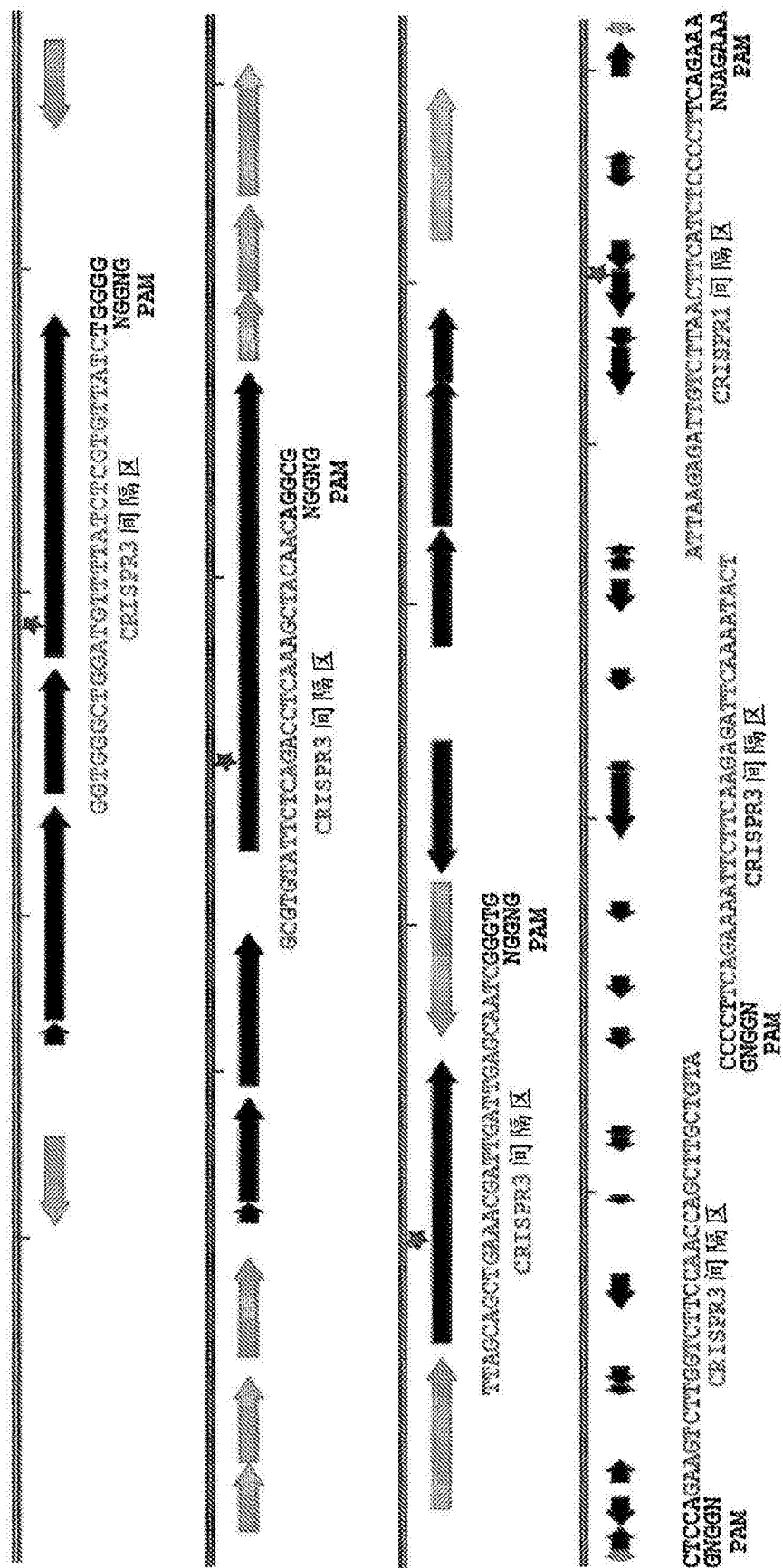


图3B

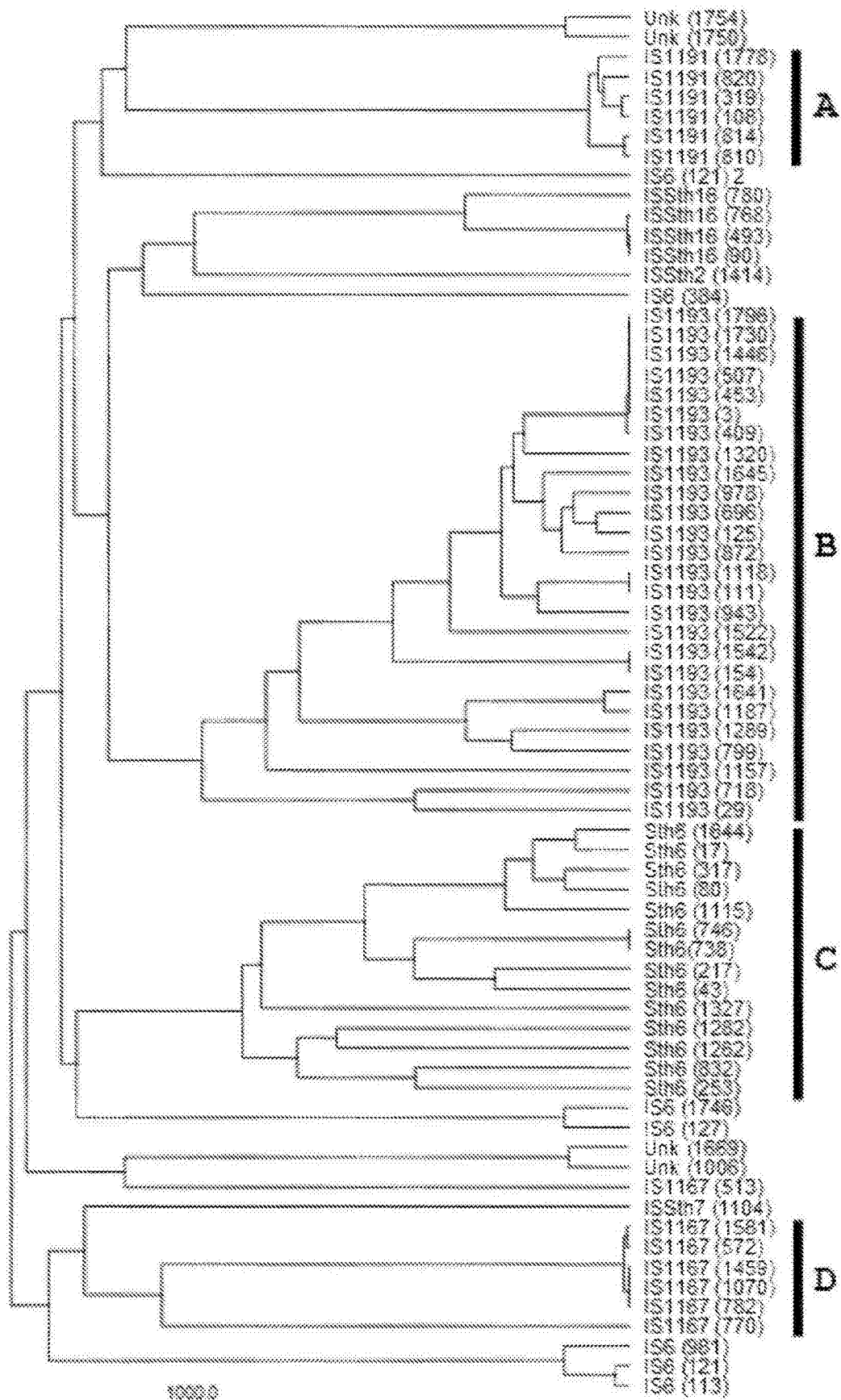


图4

IS1191

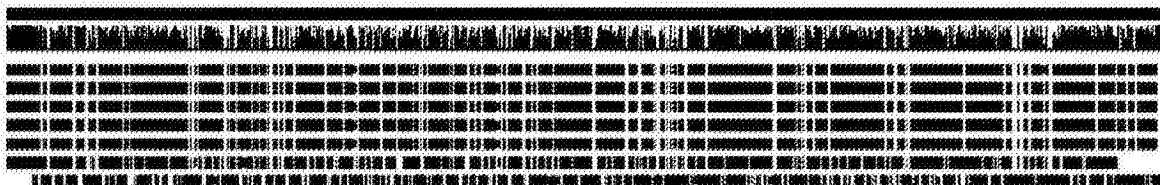


图5A

Sth6

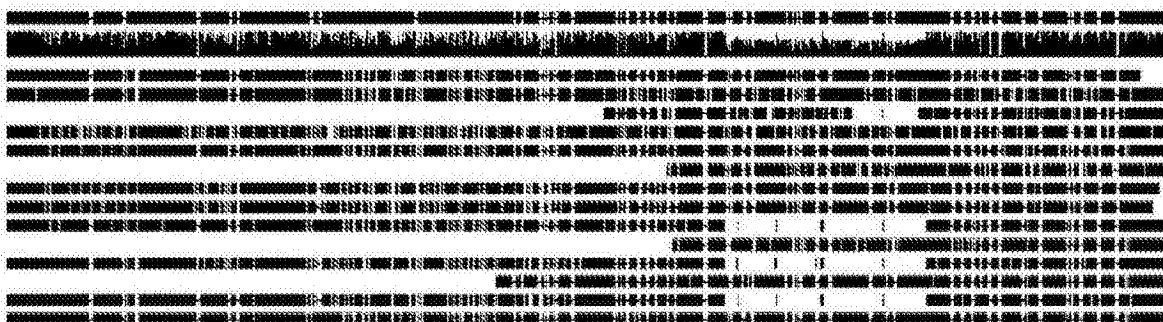


图5B

IS1193

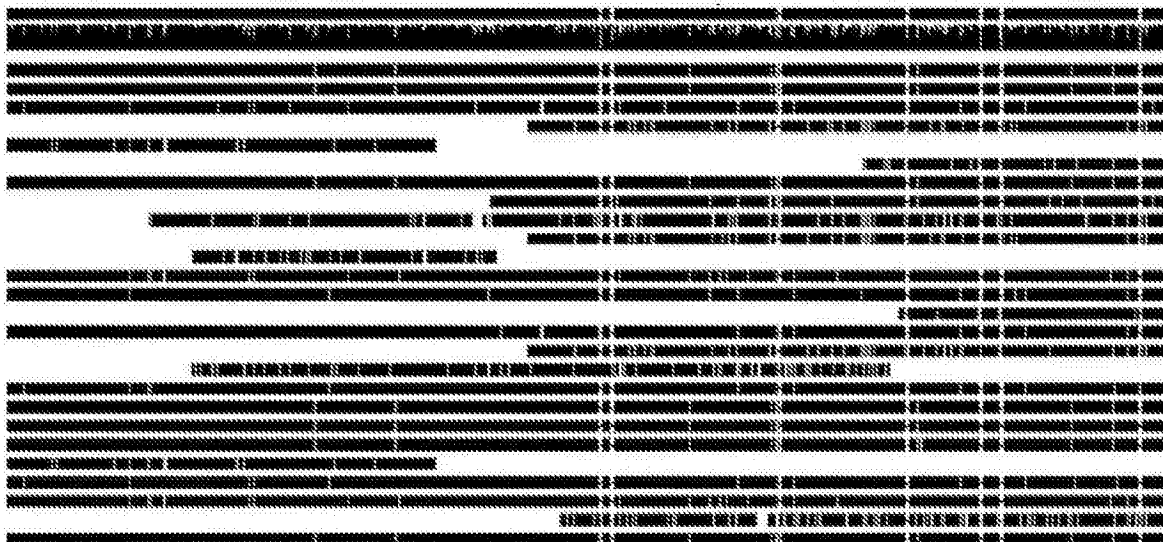


图5C

[illegible]

图5D

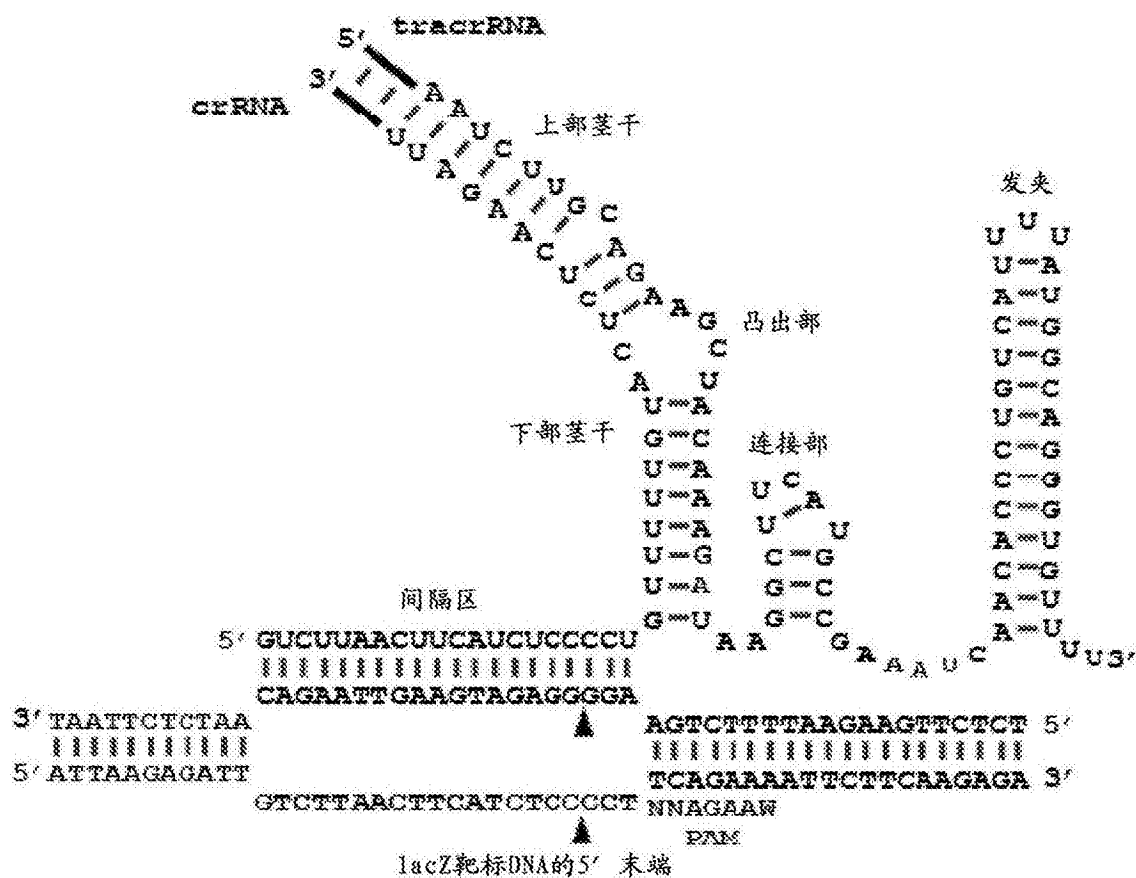


图6A

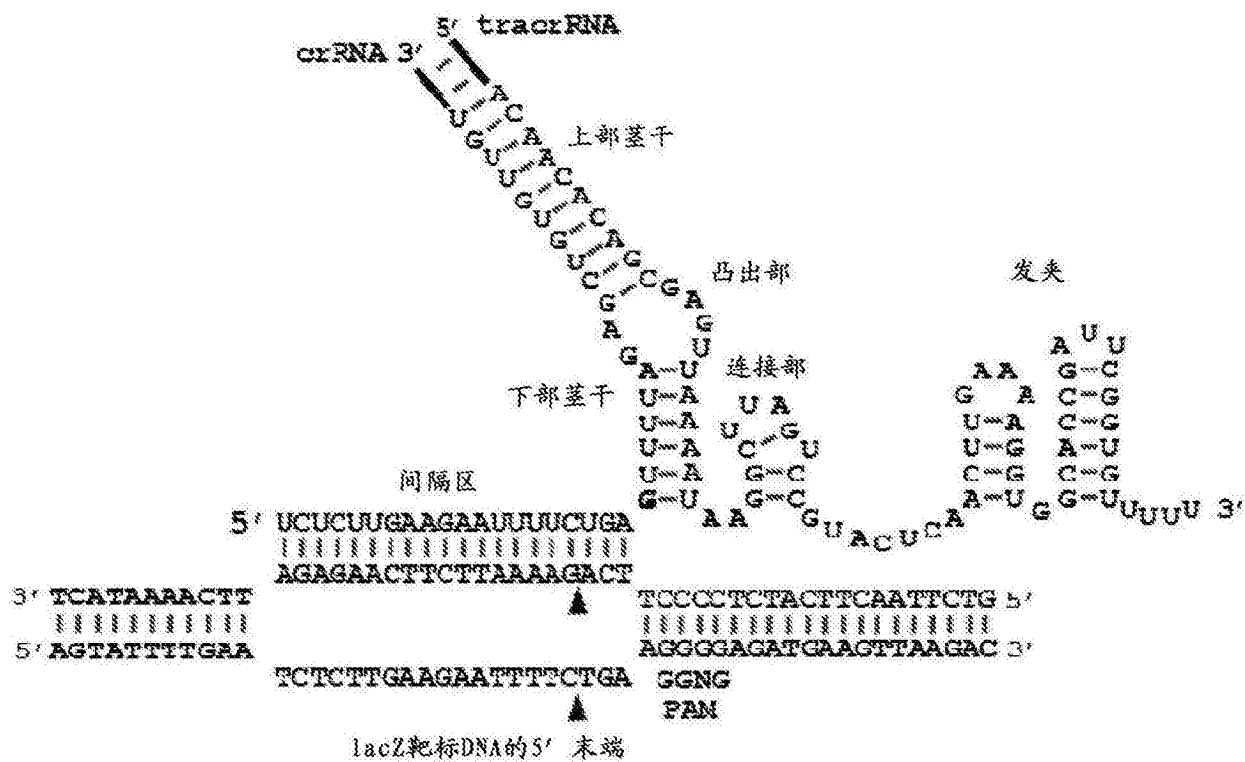


图6B

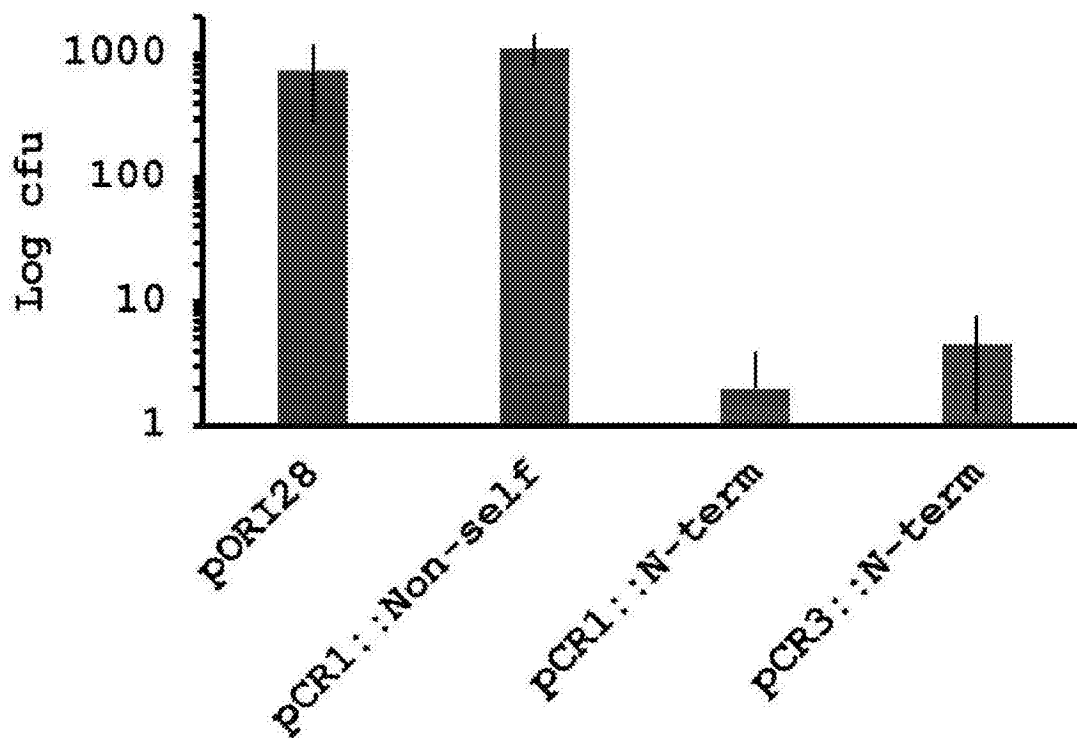


图6C

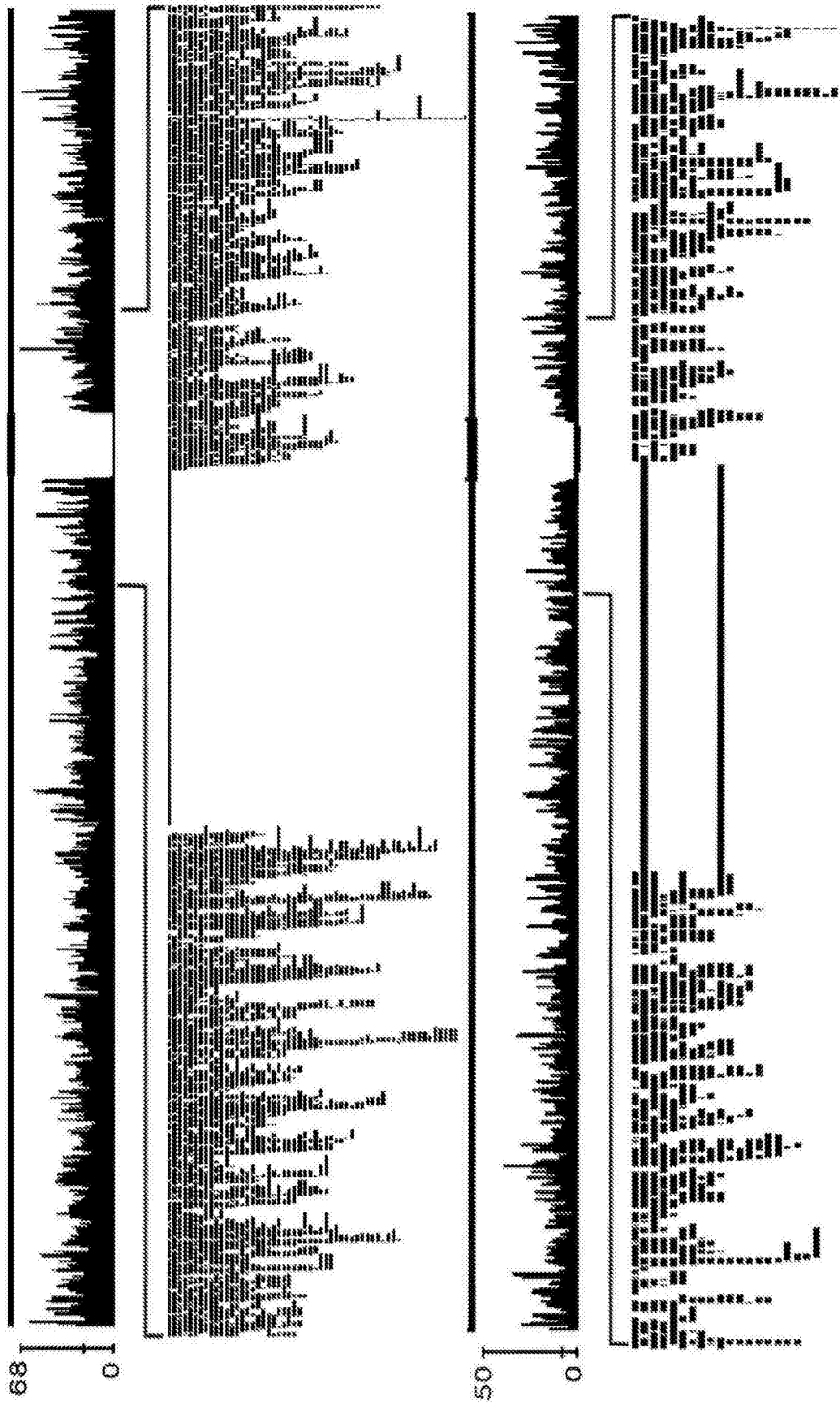


图7A

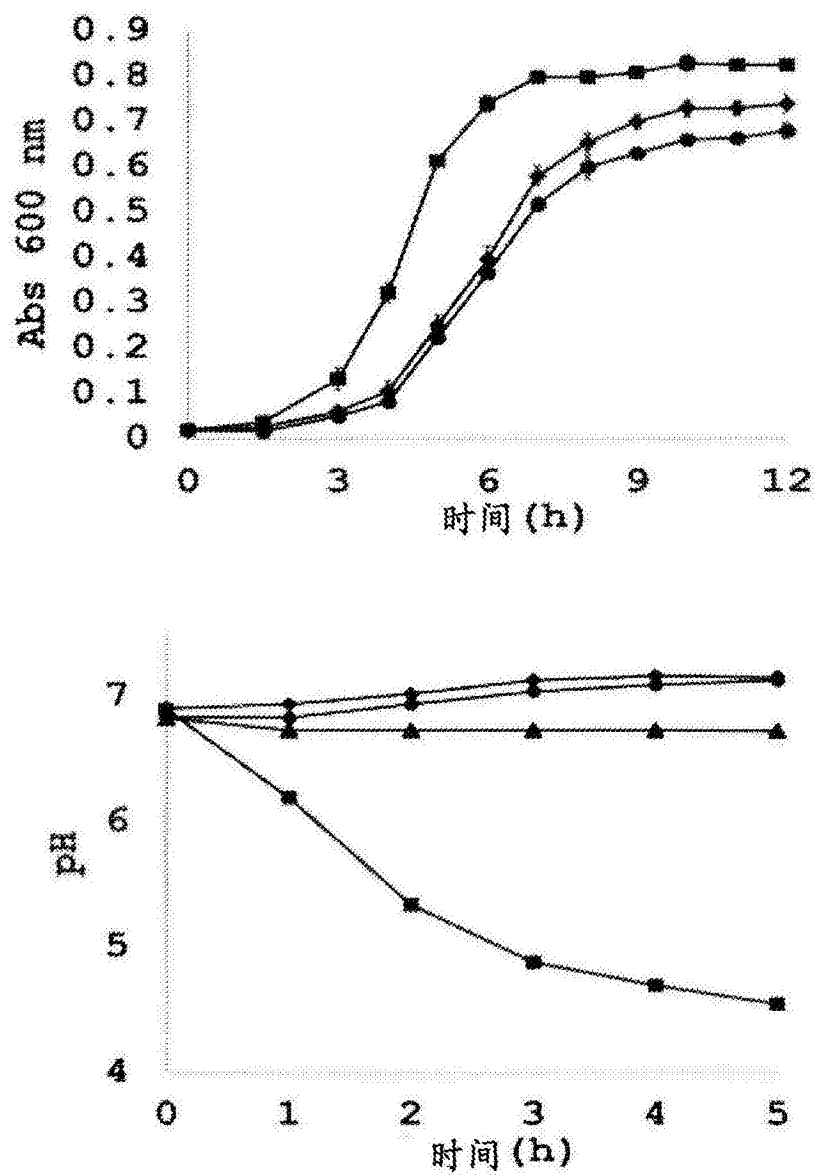


图7B

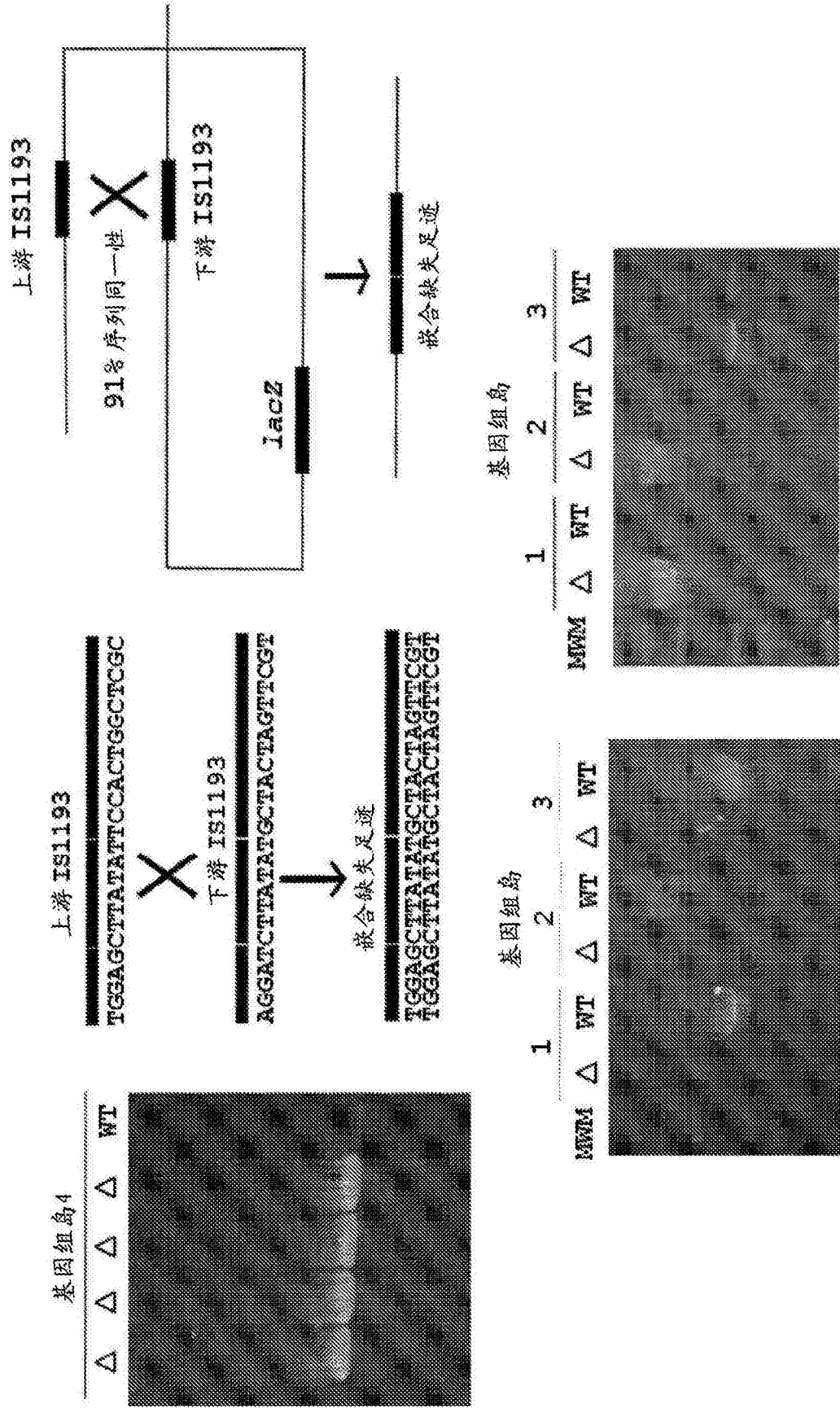


图8

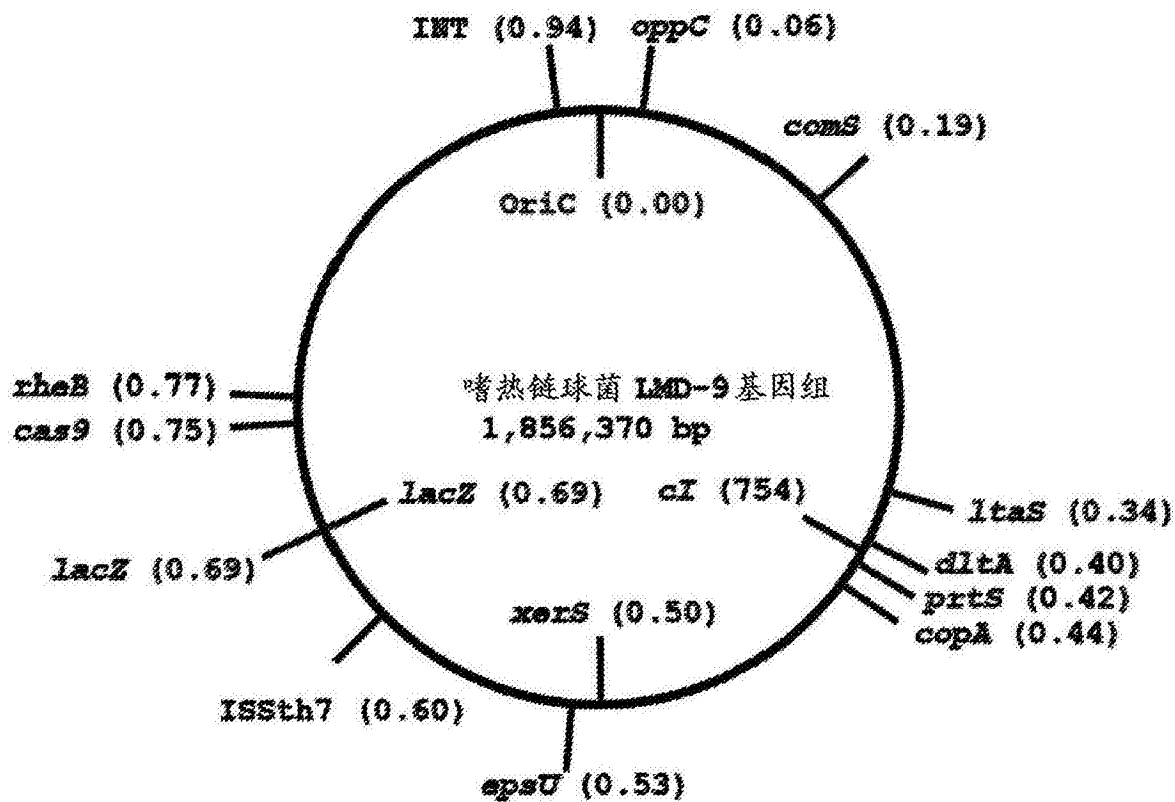


图9

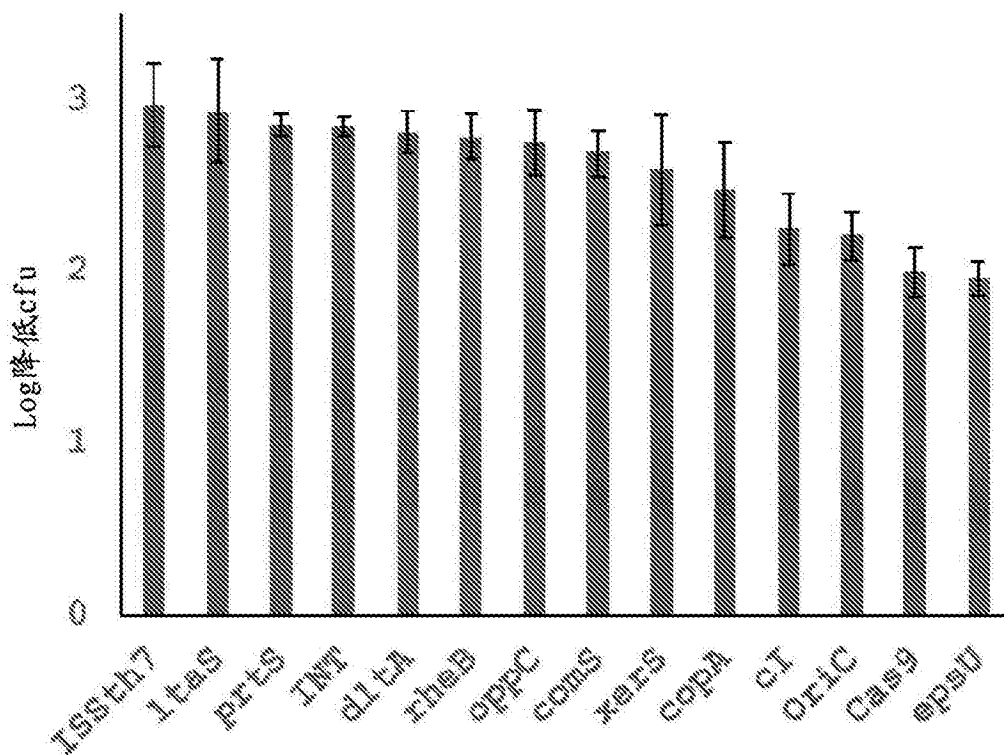


图10

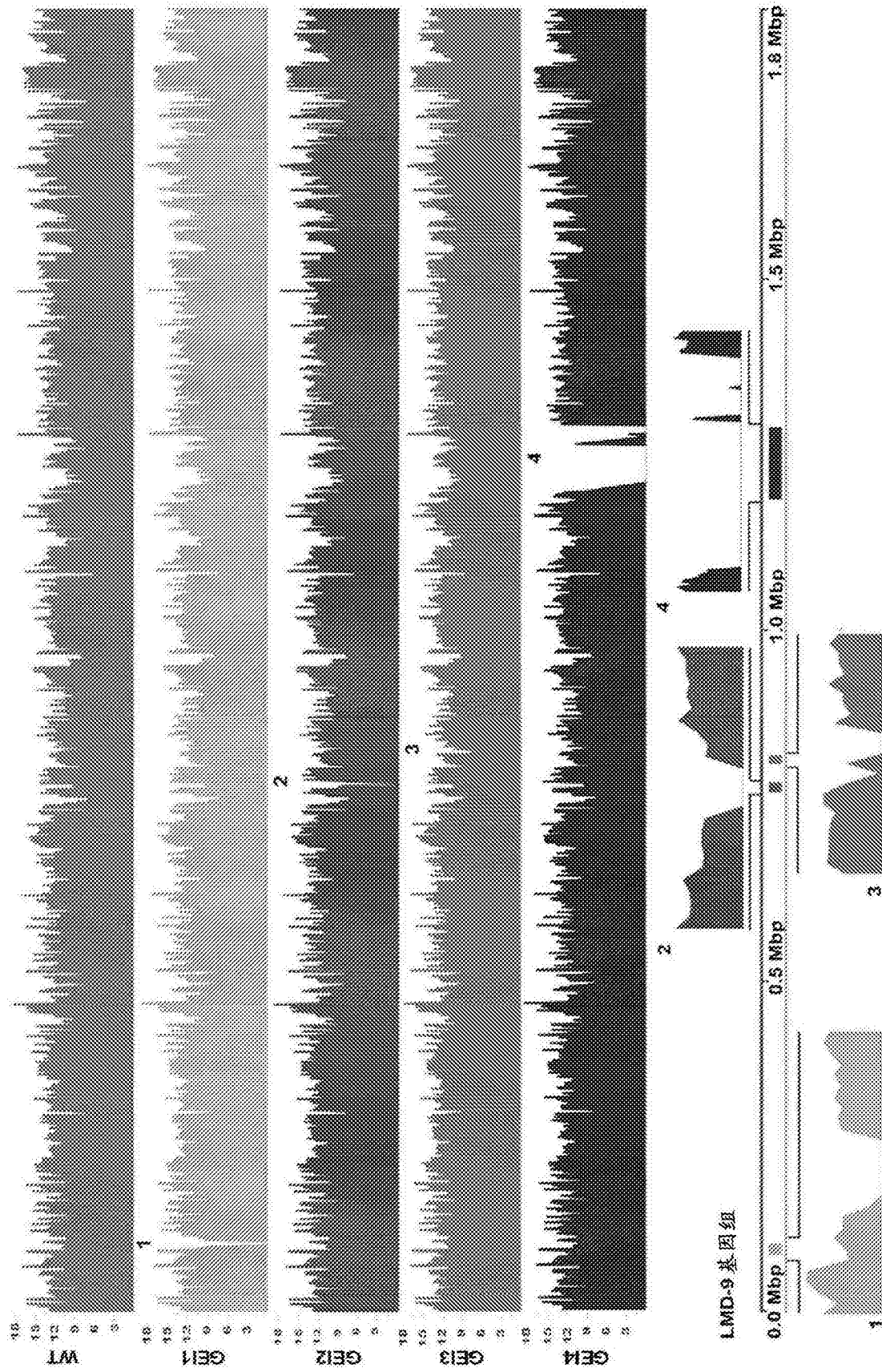


图12

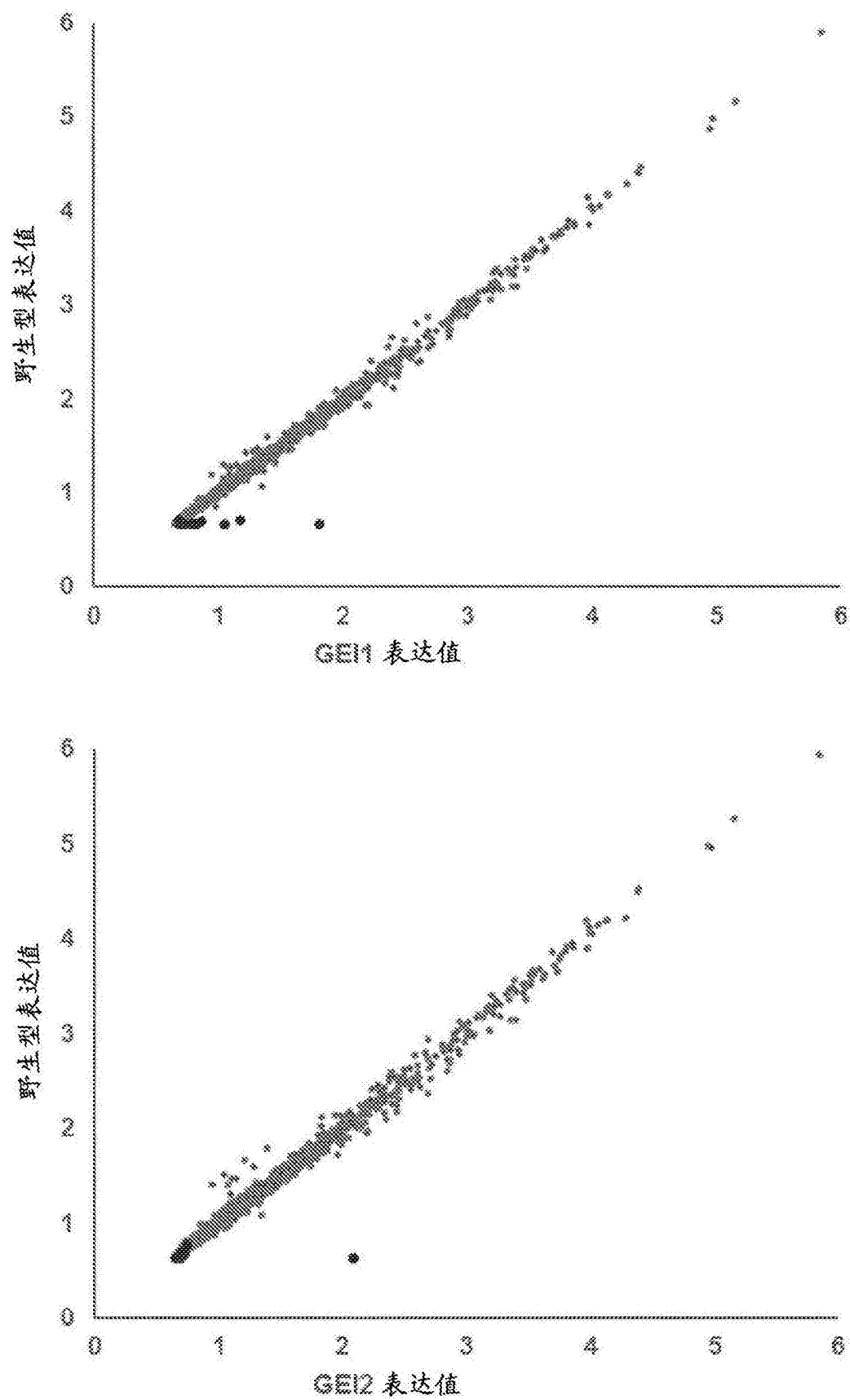


图13

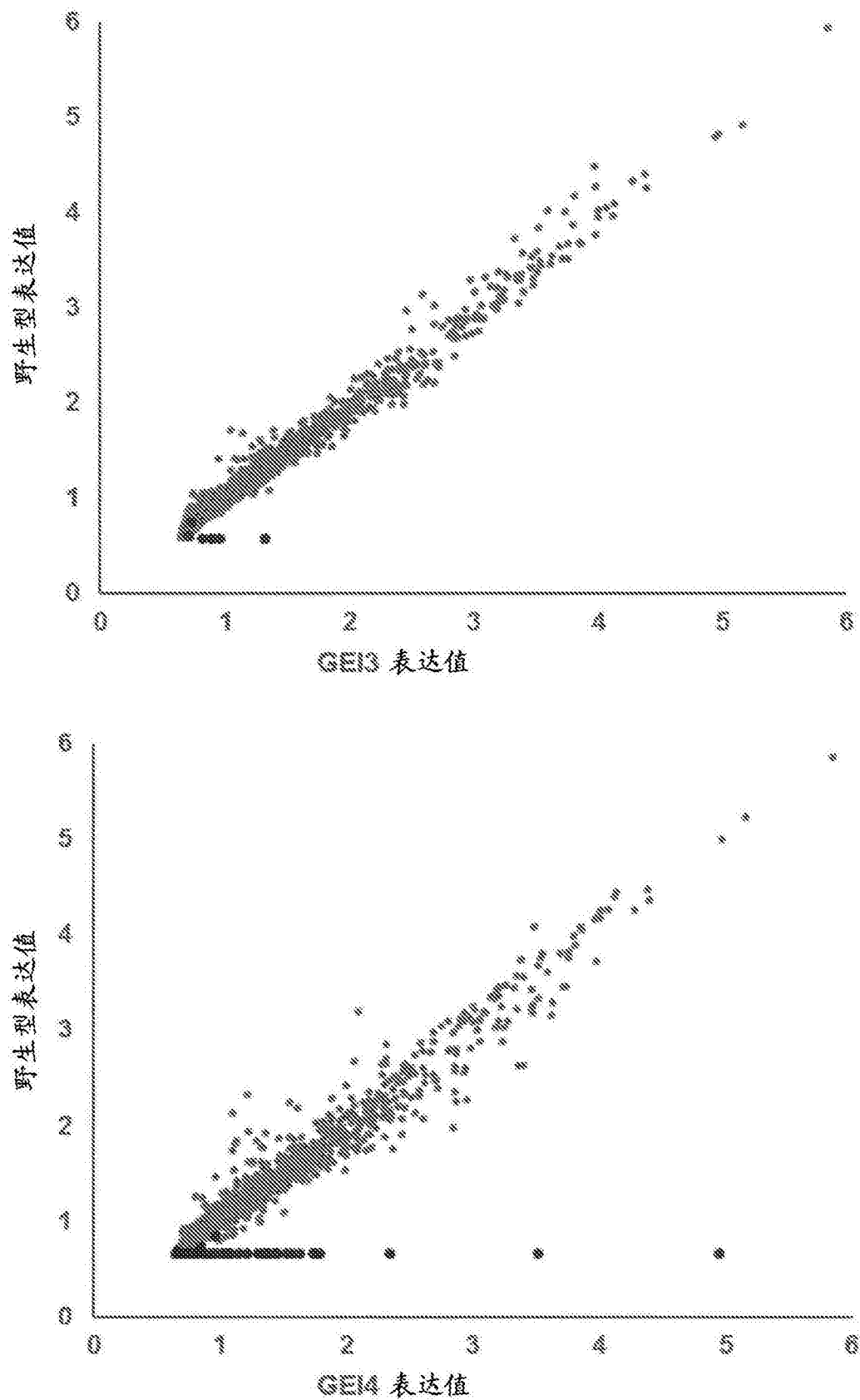


图13续

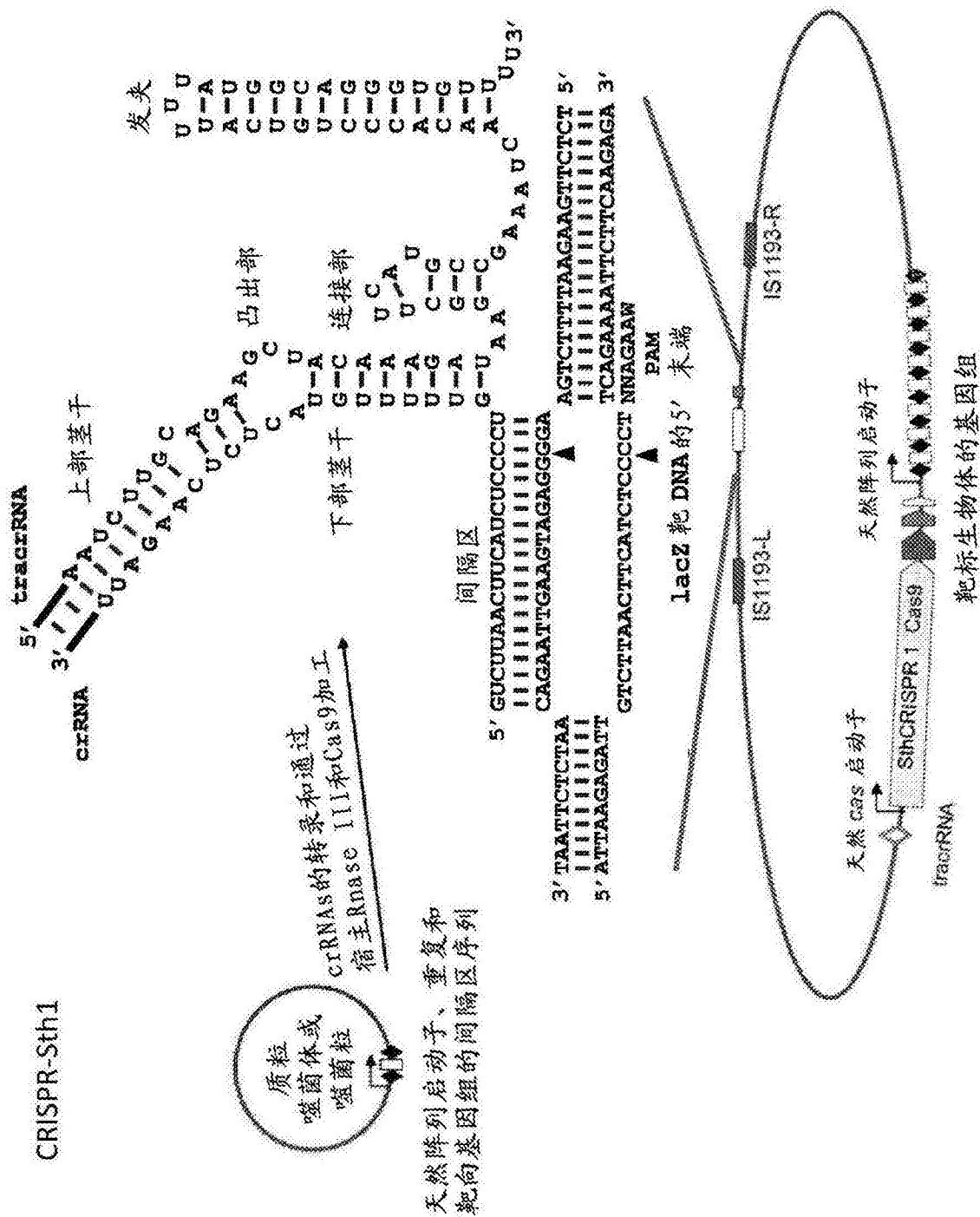


图14A

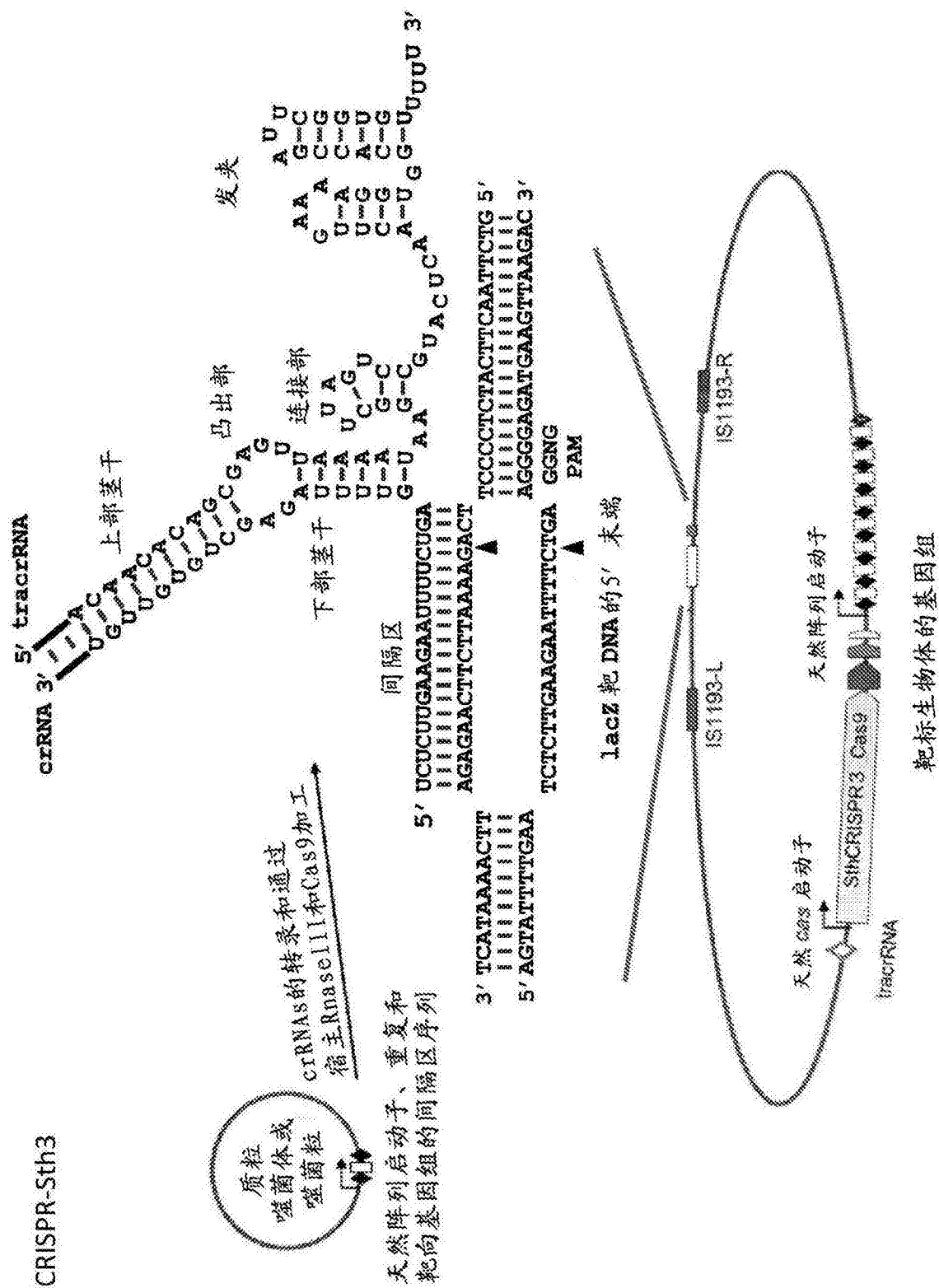


图14B

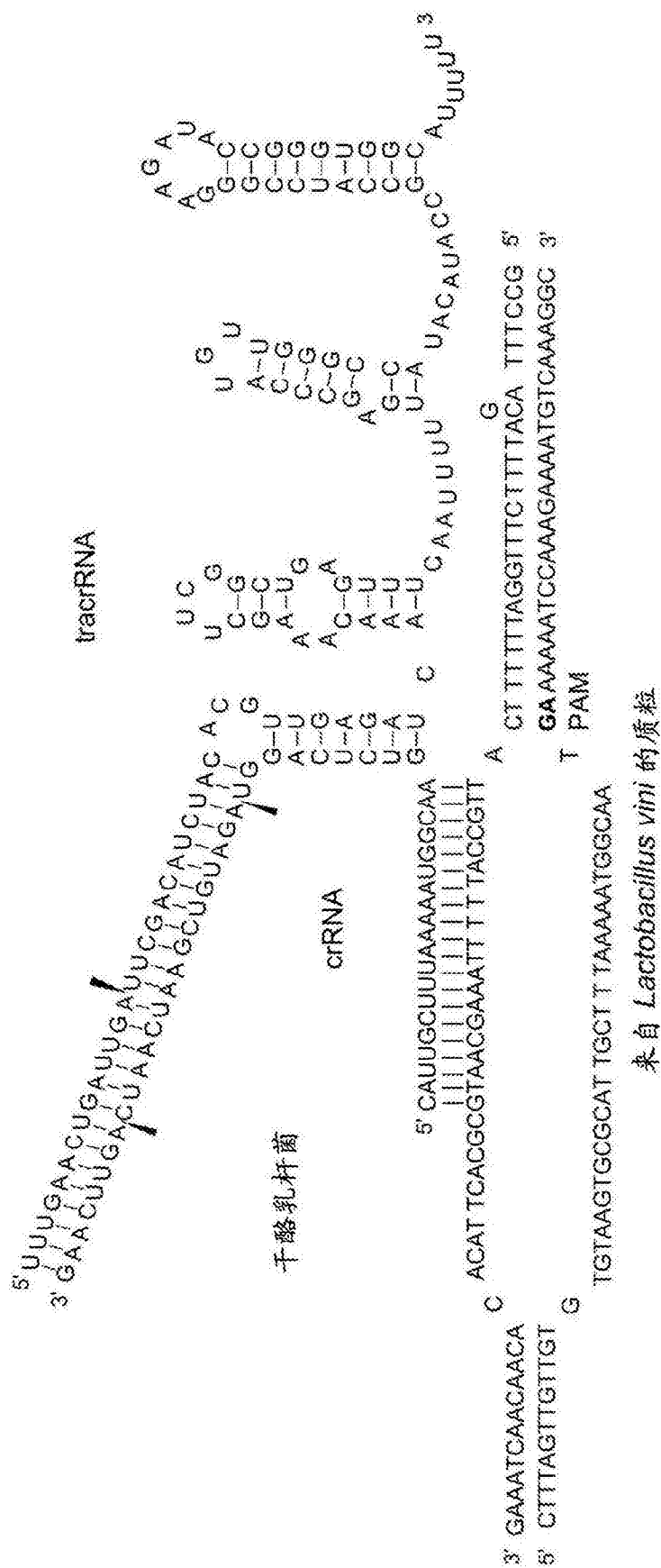


图15



图15续

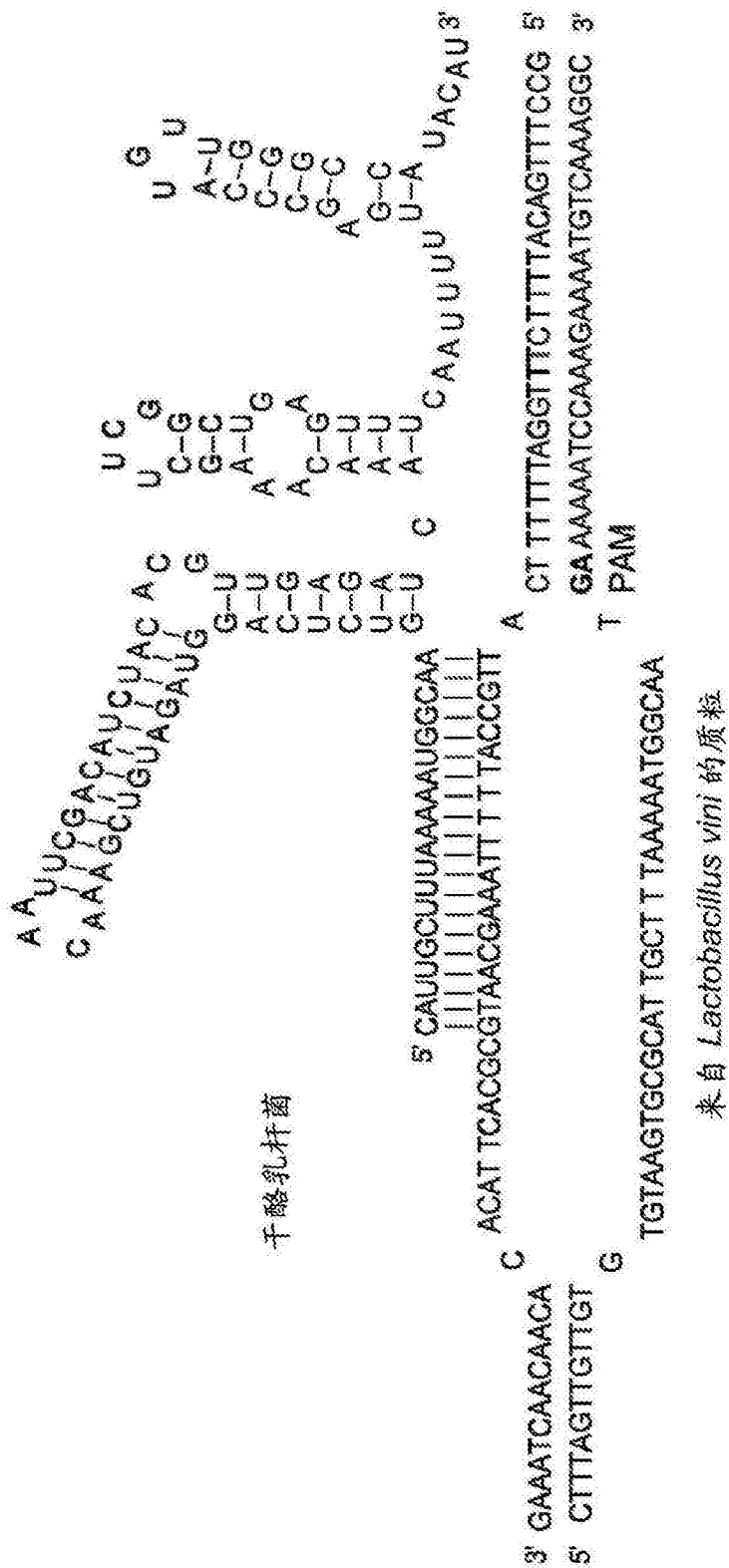


图15续

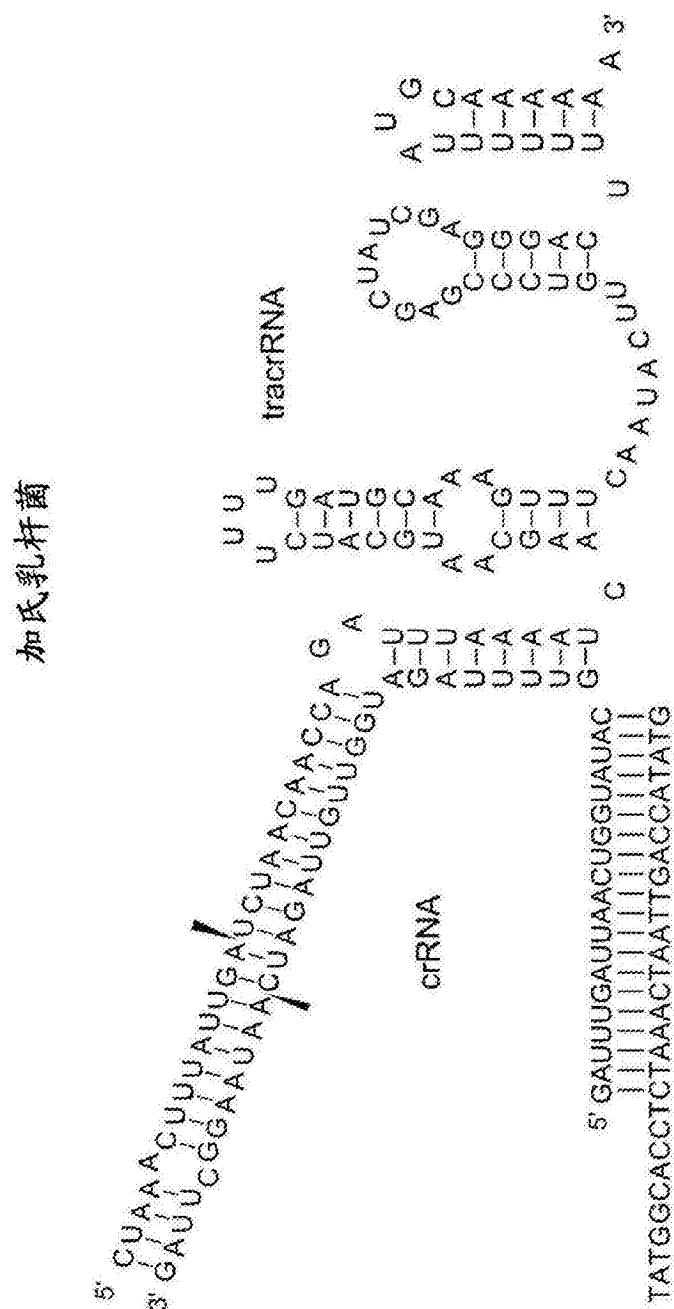


图16

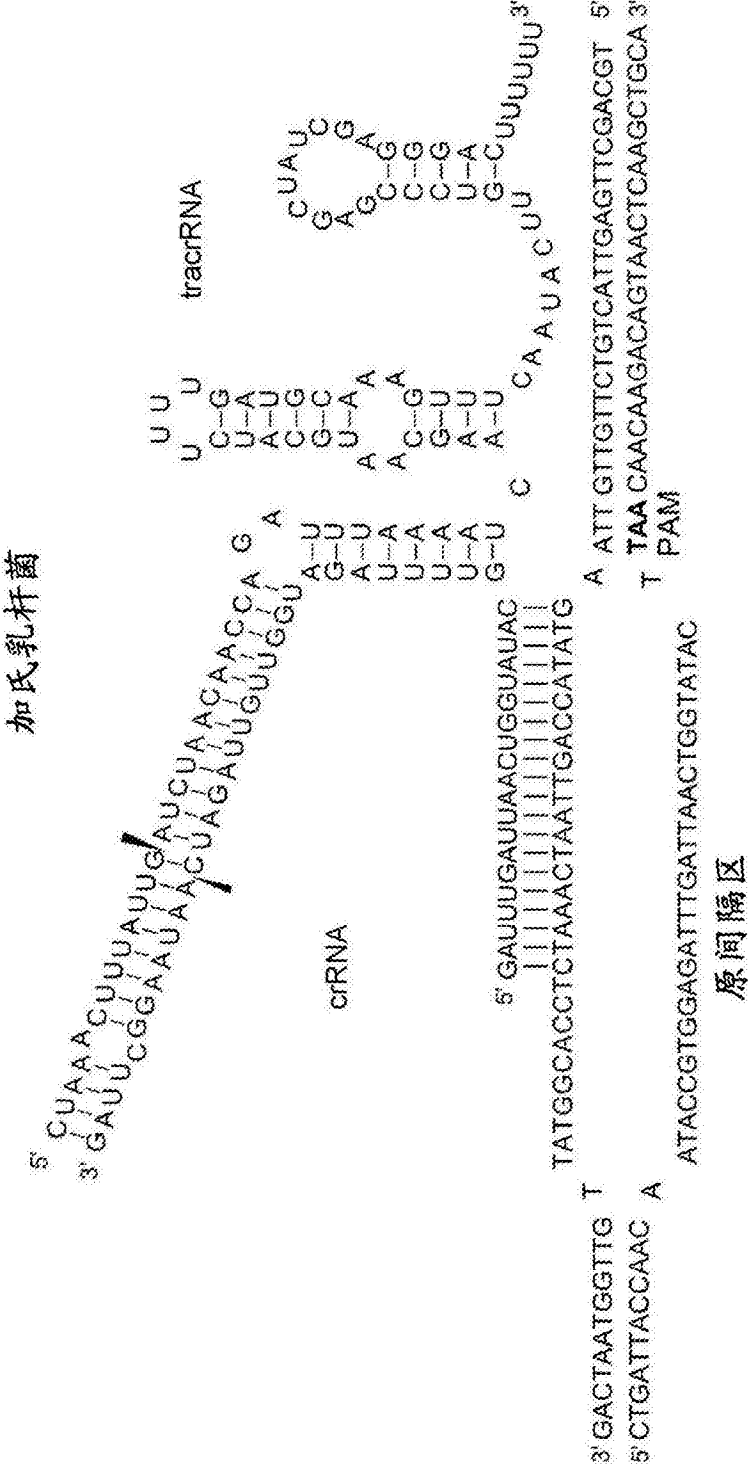


图16续

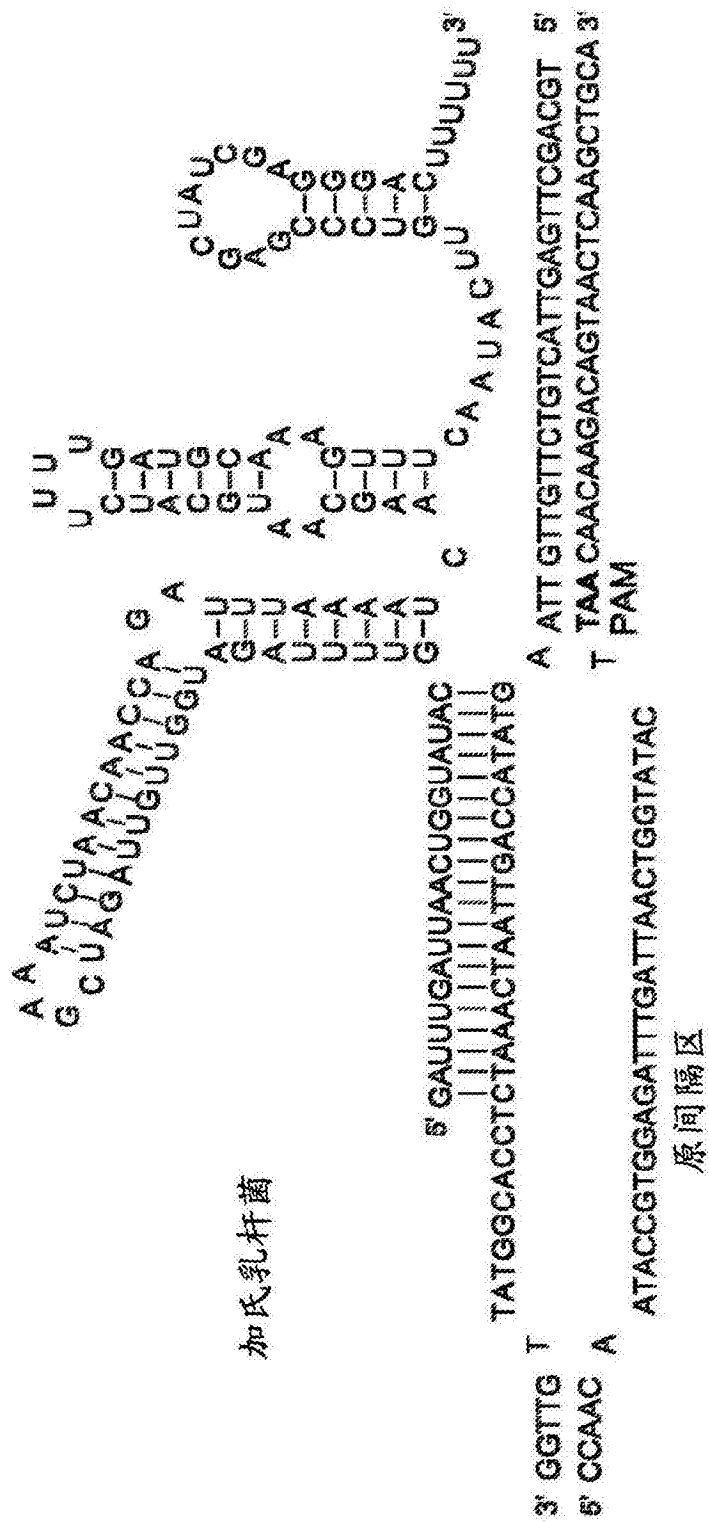


图16续

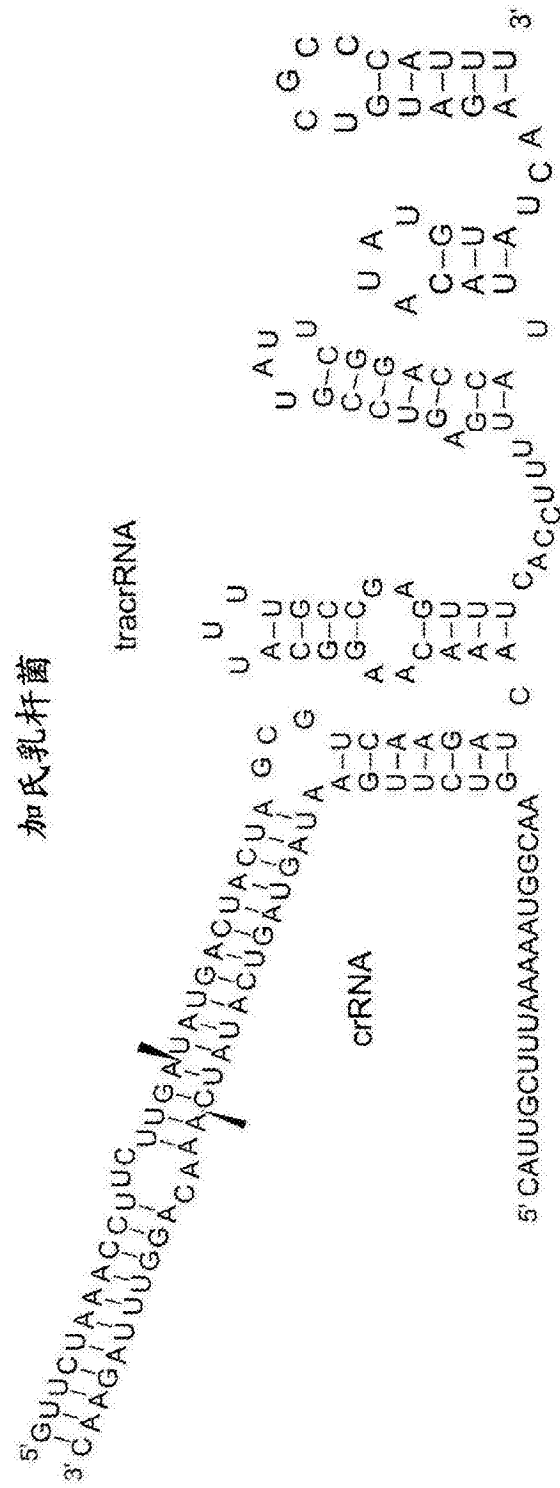


图17

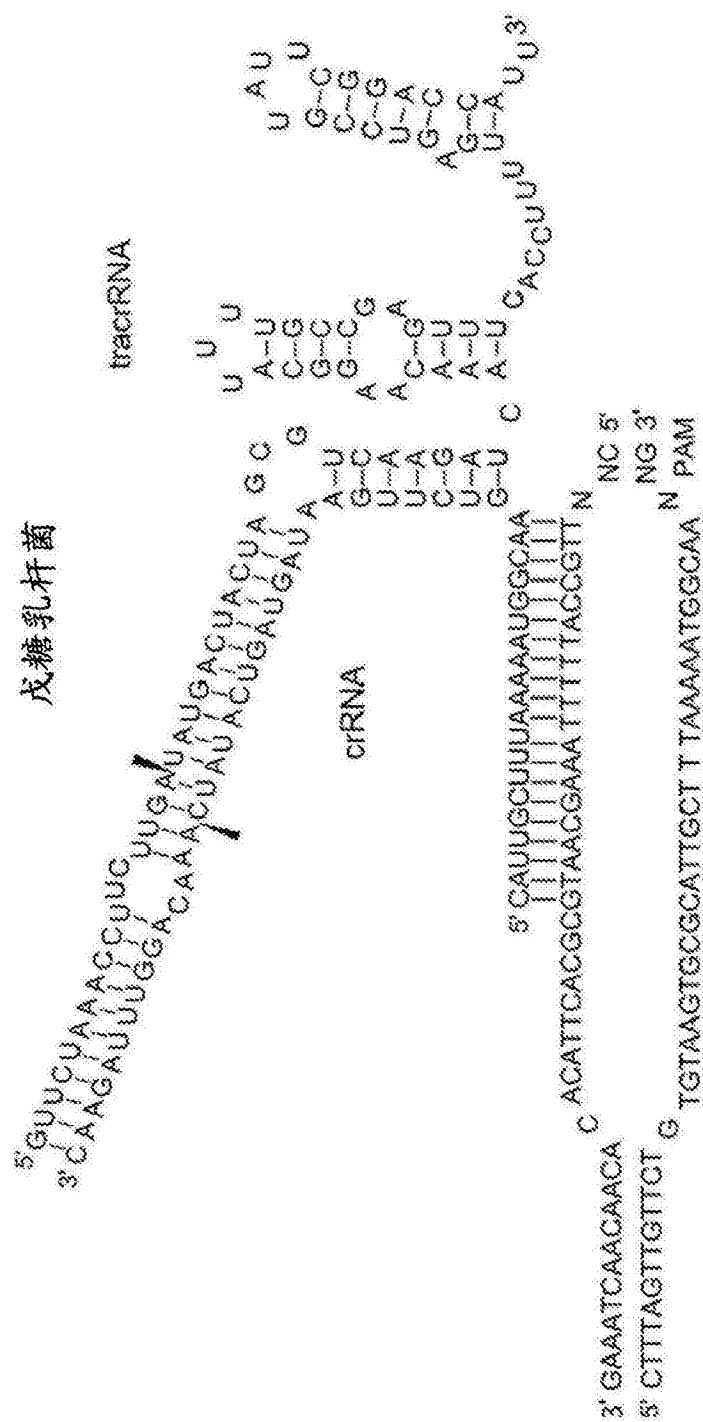


图17续

戊糖乳杆菌

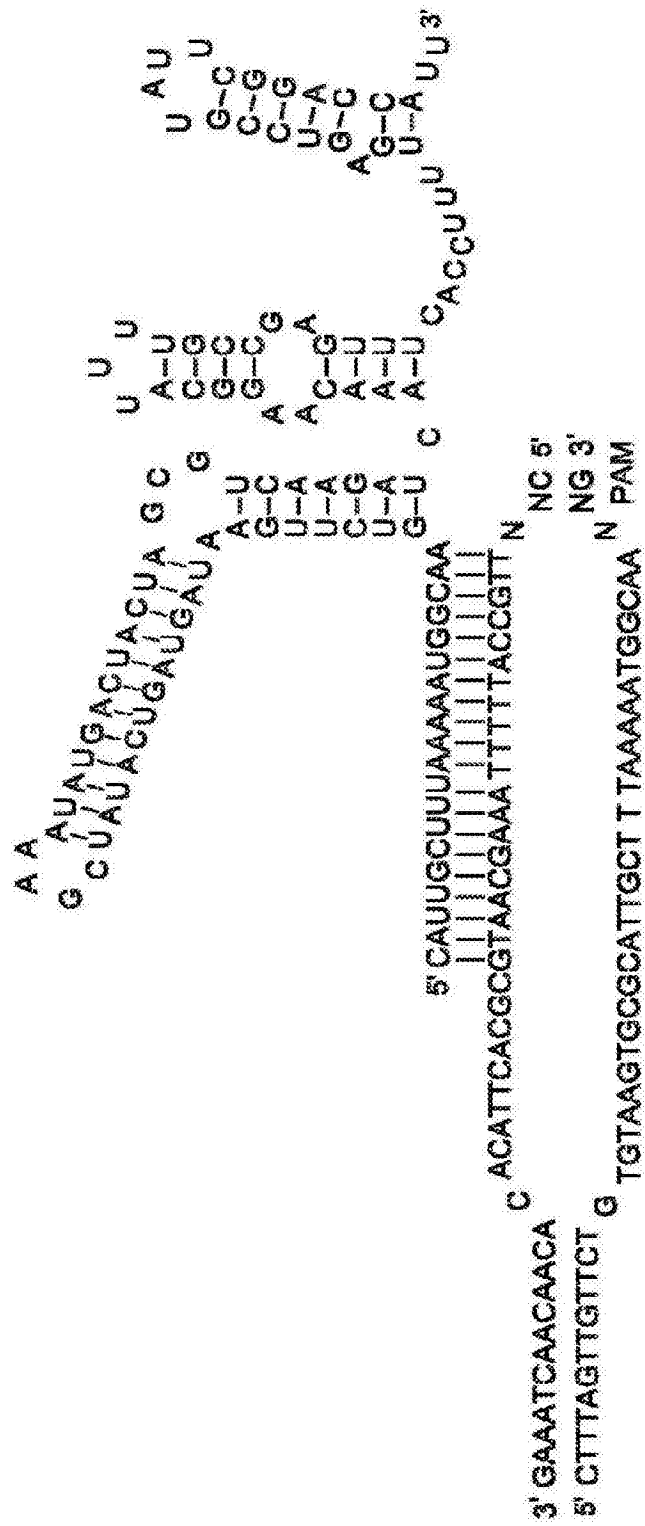


图17续

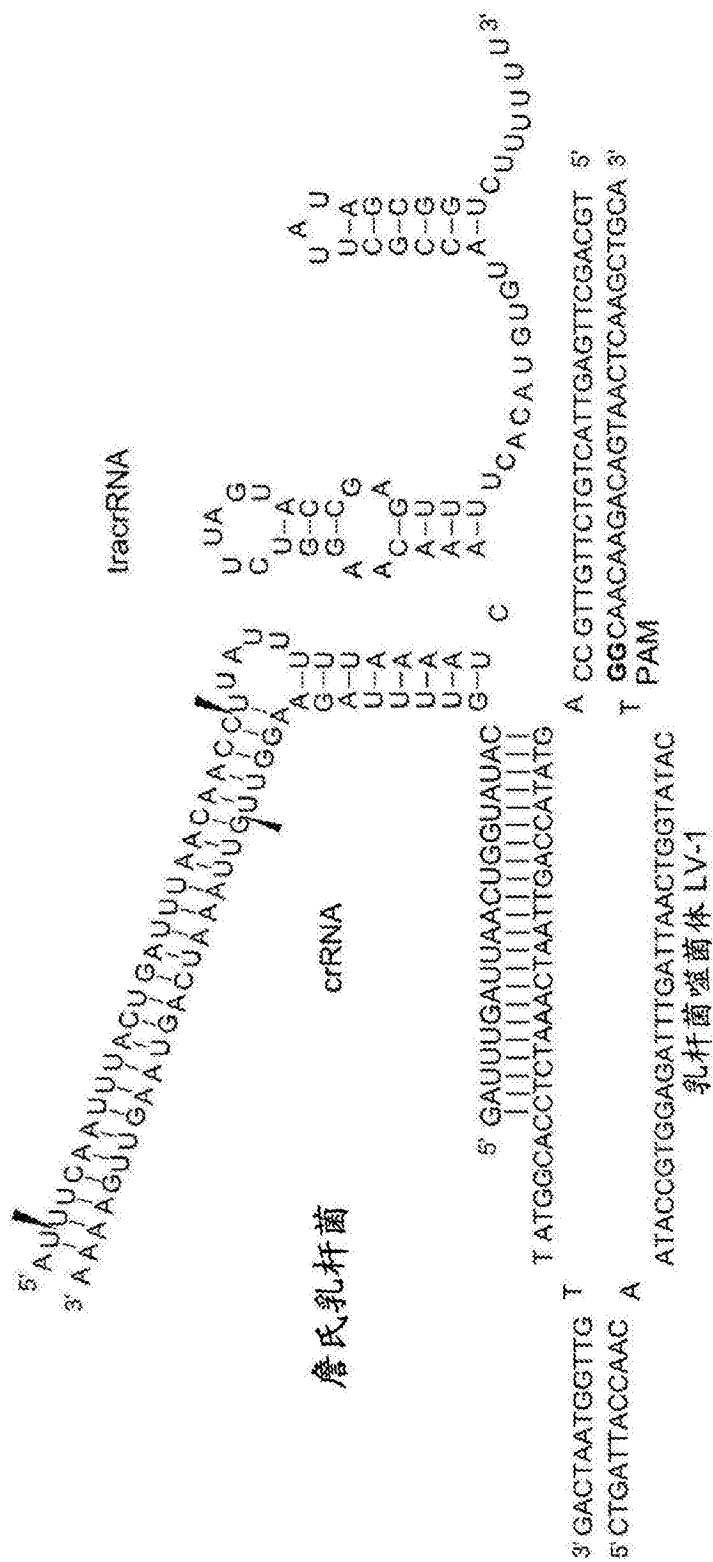


图18

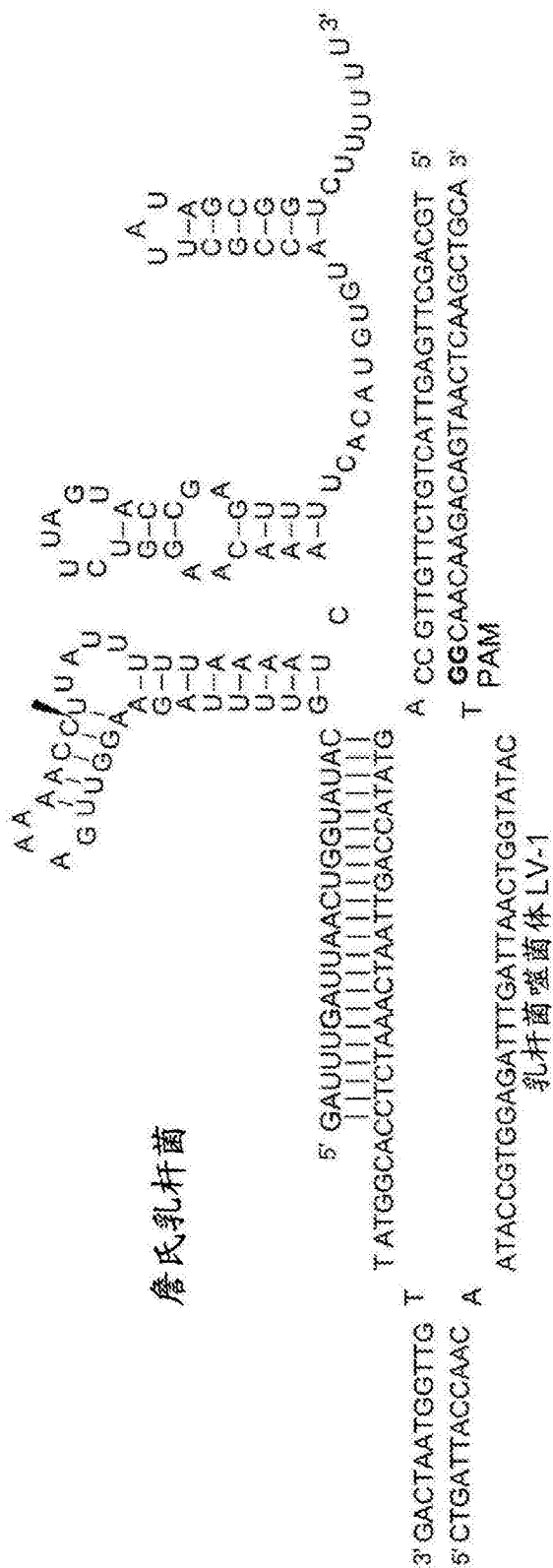


图18续

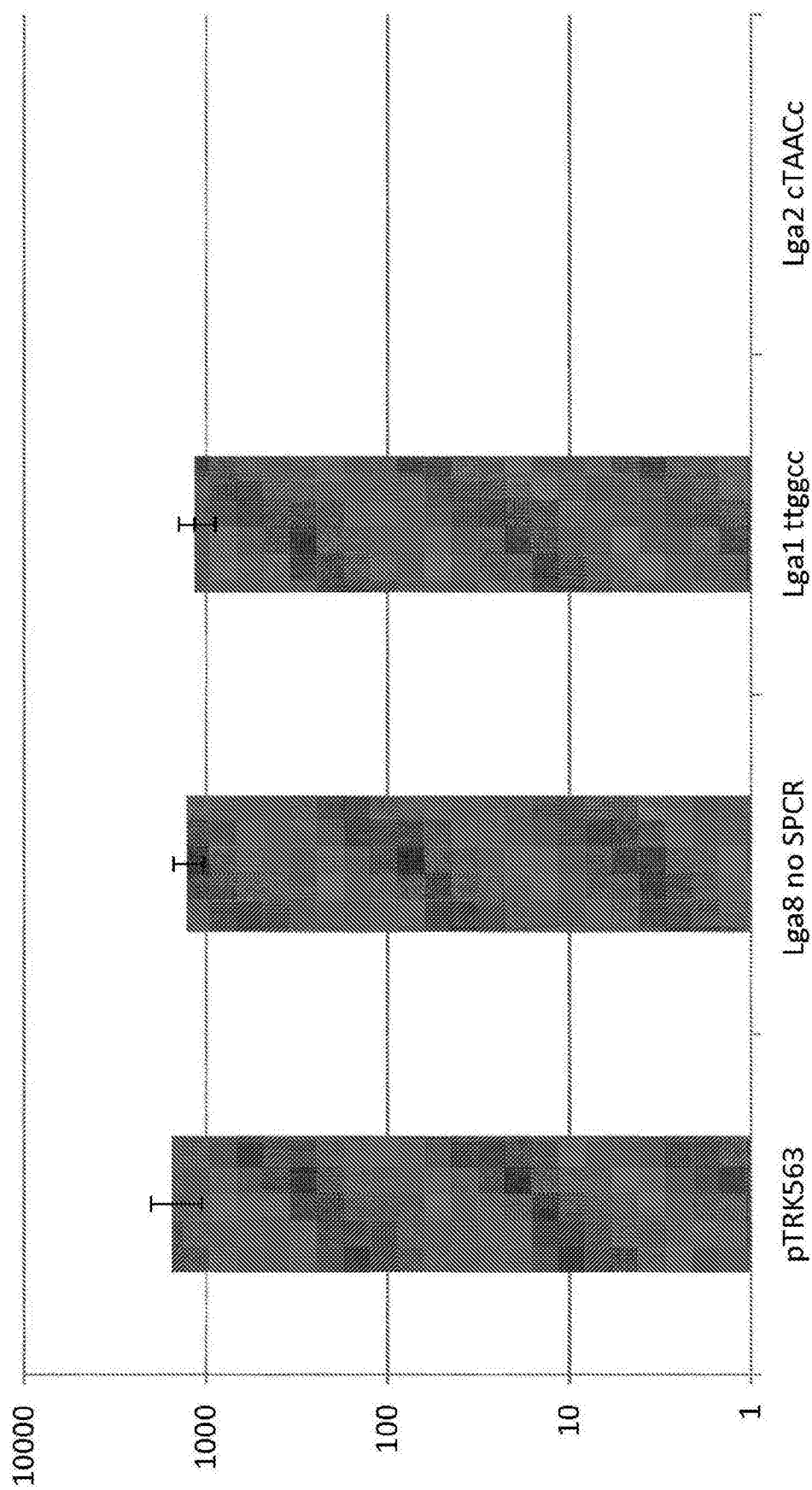


图19

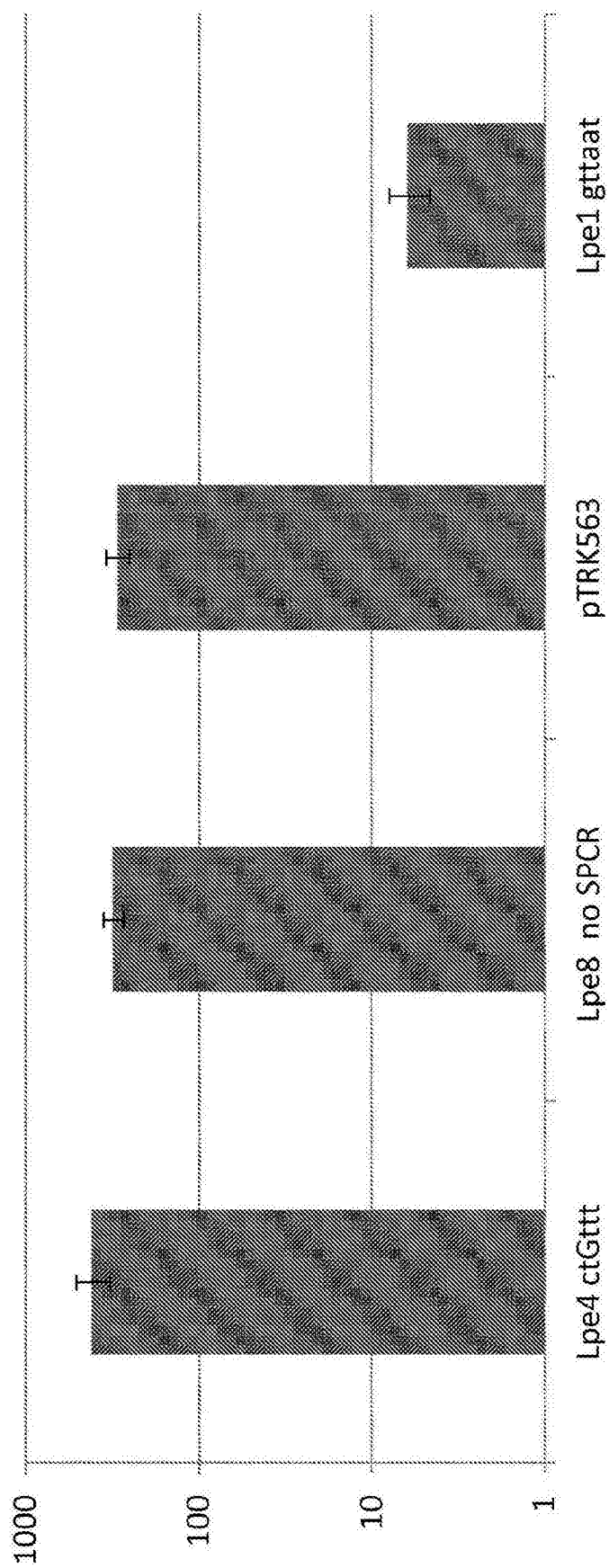


图20

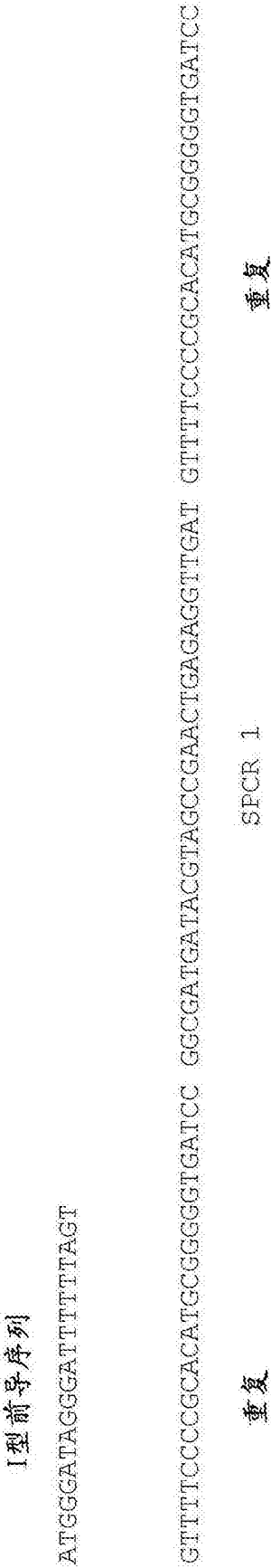


图21

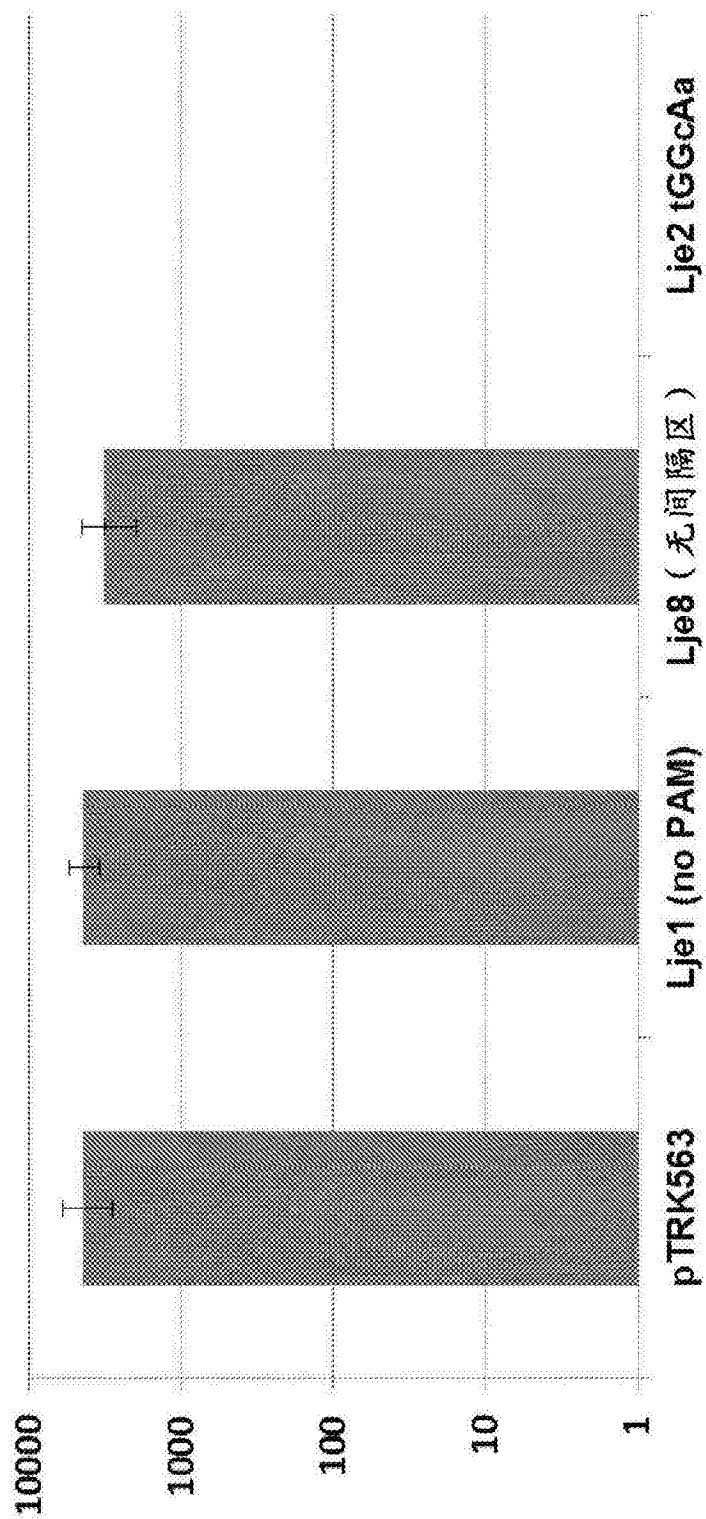


图22

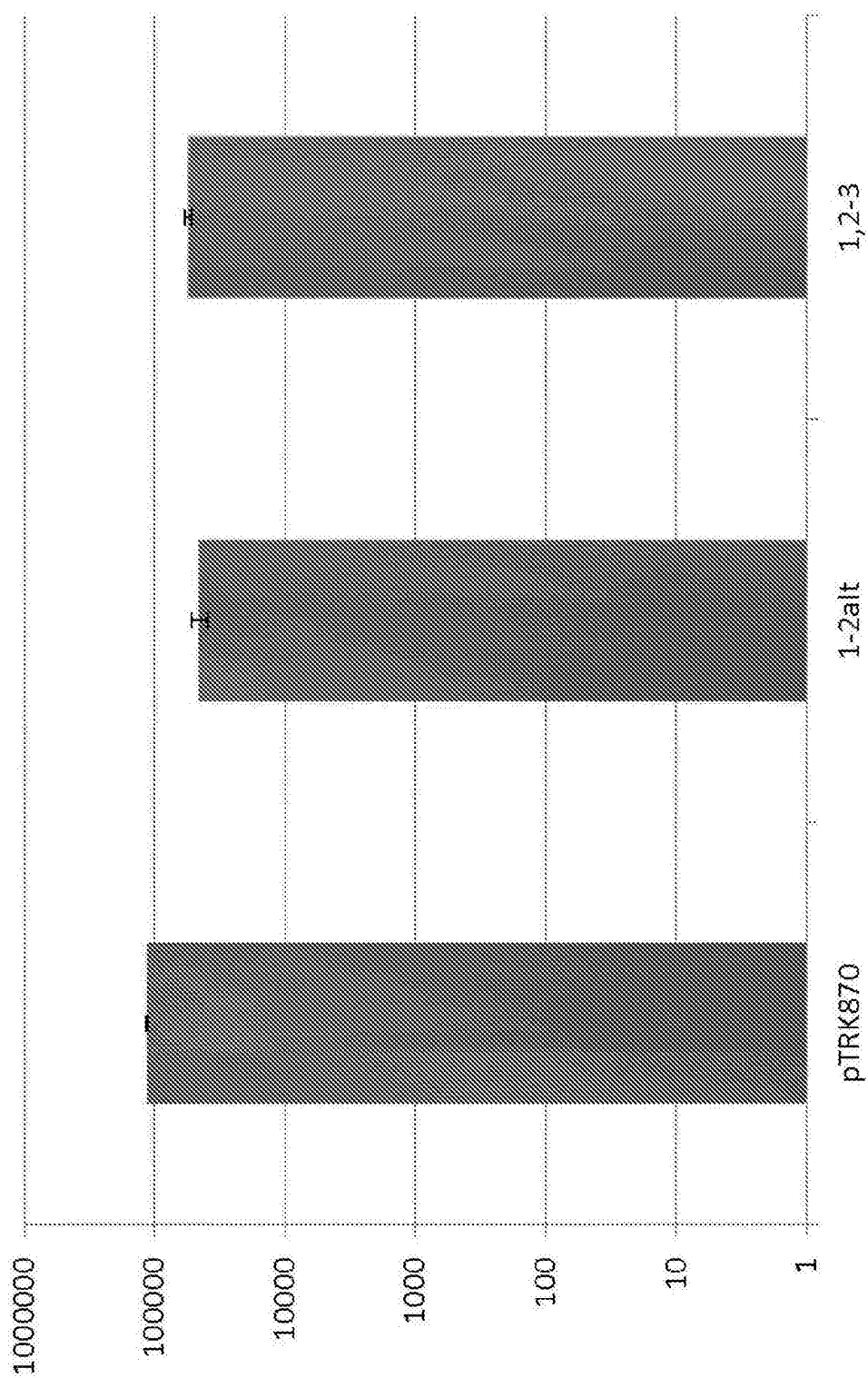


图23