(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 111705365 B
(45)授权公告日 2024.12.17

(21)申请号 202010437901.2
(22)申请日 2015.02.11
(65)同一申请的已公布的文献号
申请公布号 CN 111705365 A
(43)申请公布日 2020.09.25
(30)优先权数据
61/938,608 2014.02.11 US
(62)分案原申请数据
201580018819.3 2015.02.11
(73)专利权人 科罗拉多州立大学董事会(法人团体)
地址 美国科罗拉多州
(72)发明人 R·T·吉尔 安德鲁·加斯特
(74)专利代理机构 北京安信方达知识产权代理有限公司 11262
专利代理师 武晶晶
(51)Int.Cl.
C40B 40/02(2006.01)
C12N 15/10(2006.01)
C12N 9/22(2006.01)
(56)对比文件
WO 2013176772 A1,2013.11.28
审查员 胡嘉倩

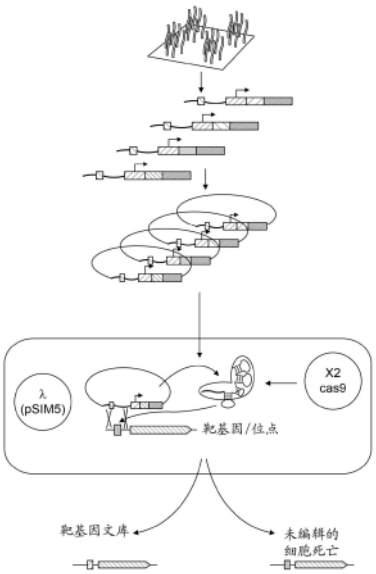
权利要求书4页 说明书19页
序列表（电子公布） 附图21页

(54)发明名称

CRISPR支持的多路基因组工程化

(57)摘要

本文描述了用于理性、多路操作开放阅读框(例如,蛋白质文库中)中的染色体或者细胞或细胞群体中染色体的任何区段的方法和载体,其中使用不同的CRISPR系统。



CN 111705365 B

1.一种核酸分子,所述核酸分子包含:

(i)编辑盒,所述编辑盒包含(a)与细胞的靶区域中的核酸序列同源并且(b)包含相对于所述靶区域的至少一个核苷酸中的突变和前间隔区邻近基序突变的区域;和

(ii)编码至少一种指导RNA的区域,所述至少一种指导RNA包含:(a)与所述靶区域的一部分互补的区域;以及(b)募集CRISPR关联的核酸酶的区域;其中所述编辑盒和所述编码至少一种指导RNA的区域共价连接。

2.根据权利要求1所述的核酸分子,其中通过所述至少一种指导RNA,包含前间隔区邻近基序突变的细胞赋予对核酸酶介导的编辑的免疫力。

3.根据权利要求1或2所述的核酸分子,其中至少一个核苷酸中的突变在所述细胞的靶区域的密码子中。

4.根据权利要求1或2所述的核酸分子,其中至少一个核苷酸中的突变包括相对于所述靶区域的一个或更多个置换。

5.根据权利要求1或2所述的核酸分子,其中至少一个核苷酸中的突变包括相对于所述靶区域的一个或更多个插入。

6.根据权利要求5所述的核酸分子,其中一个或更多个插入的长度为至少50个核苷酸。

7.根据权利要求1或2所述的核酸分子,其中至少一个核苷酸中的突变包括相对于所述靶区域的一个或更多个缺失。

8.根据权利要求1或2所述的核酸分子,其中所述核酸分子还包含用于通过所述核酸分子鉴定细胞编辑的反式条形码。

9.根据权利要求1或2所述的核酸分子,其中所述核酸分子是RNA载体的一部分。

10.一种文库,所述文库是根据权利要求1-9中任一项所述的核酸分子的文库,其中所述文库中的指导RNA靶向细胞的至少两个不同的靶区域。

11.根据权利要求10所述的文库,其中由所述文库编码的指导RNA靶向细胞的至少十个不同的靶区域。

12.根据权利要求10所述的文库,其中所述文库被配置为在全基因组规模下产生数以千计的突变。

13.一种核酸分子,所述核酸分子编码RNA分子,所述RNA分子包含:

(i)第一区段,所述第一区段包含指导RNA序列,所述指导RNA序列包含(X)与靶核酸的靶区域内的靶位点互补的区域及(Y)募集CRISPR关联的核酸酶的区域;和

(ii)第二区段,所述第二区段包含(a)与所述靶区域的一部分同源并且(b)包含相对于所述靶区域的至少一个核苷酸中的突变和前间隔区邻近基序突变的序列;

其中所述第一区段和第二区段共价连接。

14.根据权利要求13所述的核酸分子,其中通过所述指导RNA,所述第二区段包含赋予对核酸酶介导的编辑的免疫力的位点。

15.根据权利要求13或14所述的核酸分子,其中所述核酸分子是DNA载体的一部分。

16.根据权利要求13或14所述的核酸分子,其中至少一个核苷酸中的突变位于所述靶核酸的靶区域的密码子中。

17.根据权利要求13或14所述的核酸分子,其中相对于所述靶区域的至少一个核苷酸中的突变包括相对于所述靶区域的一个或更多个置换。

18.根据权利要求13或14所述的核酸分子,其中相对于所述靶区域的至少一个核苷酸中的突变包括相对于所述靶区域的一个或更多个插入。

19.根据权利要求18所述的核酸分子,其中一个或更多个插入的长度为至少50个核苷酸。

20.根据权利要求13或14所述的核酸分子,其中相对于所述靶区域的至少一个核苷酸中的突变包括相对于所述靶区域的一个或更多个缺失。

21.根据权利要求13或14所述的核酸分子,其中所述核酸分子还包含用于鉴定相对于所述靶区域的所述至少一个核苷酸中的所述突变的反式条形码。

22.根据权利要求13或14所述的核酸分子,其中所述靶核酸是靶细菌核酸或靶酵母核酸。

23.根据权利要求13或14所述的核酸分子,其中所述靶核酸是靶哺乳动物核酸。

24.一种文库,所述文库包含多于一个根据权利要求13-23中任一项所述的核酸分子。

25.根据权利要求24所述的文库,其中由文库编码的指导RNA靶向细胞的至少两个不同的靶区域。

26.根据权利要求24所述的文库,其中由文库编码的指导RNA靶向细胞的至少十个不同的靶区域。

27.根据权利要求24所述的文库,其中所述核酸分子通过基于阵列的合成来合成。

28.一种基因组工程化的方法,包括:

(a)将来自权利要求1-9中任一项的核酸分子引入第一细胞群体,从而产生包含所述核酸分子的第二细胞群体;和

(b)在表达CRISPR关联的核酸酶的条件下维持所述第二细胞群体,其中所述CRISPR关联的核酸酶在所述核酸分子上编码、在第二核酸分子上编码或在所述第二细胞群体的基因组DNA上编码,从而产生第三细胞群体,所述第三细胞群体包含所述核酸分子中的突变和前间隔区邻近基序突变。

29.根据权利要求28所述的方法,还包括:

(c)在适合细胞存活的条件下培养(b)的产物,从而产生包含终点文库的活细胞;

(d)获得(c)中产生的活细胞;和

(e)通过测序鉴定至少一个核苷酸中的突变。

30.根据权利要求28所述的方法,所述方法还包括筛选或选择具有感兴趣的表型的细胞。

31.根据权利要求29所述的方法,其中测序包括对所述编辑盒进行测序。

32.根据权利要求28所述的方法,所述方法还包括诱导不包含所述前间隔区邻近基序突变的细胞的死亡。

33.根据权利要求28所述的方法,其中所述细胞是原核细胞。

34.根据权利要求28所述的方法,其中所述细胞是真核细胞。

35.根据权利要求28所述的方法,其中所述细胞是酵母细胞、藻细胞、植物细胞或哺乳动物细胞。

36.根据权利要求28所述的方法,其中所述核酸分子的靶区域包含在基因组基因座内。

37.根据权利要求28所述的方法,其中所述核酸分子的靶区域包含在染色体外的基因

座内。

38.根据权利要求28所述的方法,所述方法以多路方式进行,从而产生突变的基因组的文库。

39.根据权利要求38所述的方法,其中所述文库包含至少两个突变的基因组。

40.根据权利要求39所述的方法,其中所述文库包含至少8,000个不同的突变。

41.根据权利要求40所述的方法,其中所述文库包含至少27,000种不同的理性设计的寡核苷酸。

42.根据权利要求1-2和13-14中任一项所述的核酸分子,其中所述CRISPR关联的核酸酶是Cas9核酸酶。

43.根据权利要求10所述的文库,其中所述CRISPR关联的核酸酶是Cas9核酸酶。

44.根据权利要求24所述的文库,其中所述CRISPR关联的核酸酶是Cas9核酸酶。

45.根据权利要求28所述的方法,其中所述CRISPR关联的核酸酶是Cas9核酸酶。

46.一种产生具有基因组编辑的细胞的文库的方法,所述方法包括:提供多于一个合成的编辑寡核苷酸,各自单独包含:

(i)编辑盒,所述编辑盒包含(a)与细胞的靶区域中的核酸序列同源并且(b)包含相对于所述靶区域的至少一个核苷酸中的突变和前间隔区邻近基序突变的区域;和

(ii)编码至少一种指导RNA的区域,所述至少一种指导RNA包含:(a)与所述靶区域的一部分互补的区域;以及(b)募集CRISPR关联的核酸酶的区域;其中所述编辑盒和所述编码至少一种指导RNA的区域共价连接;将所述多于一个合成的编辑寡核苷酸和载体骨架转化到包含所述靶区域的细胞群体中,

在表达CRISPR关联的核酸酶的条件下维持所述细胞,其中CRISPR关联的核酸酶在细胞内的基因组DNA上编码或在转化到细胞中的多核苷酸上编码,及

富集包含相对于所述靶区域的至少一个核苷酸中的突变的细胞。

47.根据权利要求46所述的方法,其中所述细胞是原核细胞。

48.根据权利要求46所述的方法,其中所述细胞是真核细胞。

49.根据权利要求46-48中任一项所述的方法,其中所述合成的编辑寡核苷酸通过基于阵列的合成来合成,并且任选地被扩增。

50.根据权利要求46-48中任一项所述的方法,其中至少一个核苷酸中的突变包括相对于所述靶区域的一个或更多个核苷酸置换、缺失或插入。

51.一种细胞,所述细胞包含权利要求10-23中任一项的核酸分子。

52.一种细胞,所述细胞包含:

(A)一种核酸,所述核酸包含:

(i)编码至少一个编辑盒的核酸序列,所述至少一个编辑盒包含

(a)与细胞中核酸的靶区域同源的区域,

(b)相对于所述靶区域的至少一个核苷酸中的突变,和

(c)第一前间隔区邻近基序突变;和

(ii)编码至少一种指导RNA的核酸序列,所述至少一种指导RNA包含:

(a)与所述靶区域的一部分互补的区域;和

(b)与CRISPR关联的核酸酶相互作用的区域,

其中所述至少一个编辑盒和所述至少一种指导RNA共价连接;和

(B)CRISPR关联的核酸酶。

53.根据权利要求52所述的细胞,其中所述编辑盒用作用于随后测序的反式条形码。

54.根据权利要求52或53所述的细胞,其中所述靶区域在非编码区域内。

55.根据权利要求52或53所述的细胞,其中所述靶区域位于细胞内的感兴趣基因内。

56.根据权利要求52或53所述的细胞,其中至少一个核苷酸中的突变位于细胞内的感兴趣的基因的至少一个密码子内。

57.根据权利要求52或53所述的细胞,其中所述细胞中的所述核酸的靶区域包含在基因组基因座内。

58.根据权利要求52或53所述的细胞,其中所述细胞中的所述核酸的靶区域包含在染色体外的基因座内。

59.根据权利要求52或53所述的细胞,其中至少一个核苷酸中的突变通过对所述编辑盒进行测序来鉴定。

60.根据权利要求46所述的方法,其中所述CRISPR关联的核酸酶是Cas9核酸酶。

61.根据权利要求52或53所述的细胞,其中所述CRISPR关联的核酸酶是Cas9核酸酶。

CRISPR支持的多路基因组工程化

[0001] 本申请是申请日为2015年2月11日、申请号为201580018819.3、发明名称为“CRISPR支持的多路基因组工程化”的中国专利申请(PCT申请号为PCT/US2015/015476)的分案申请。

[0002] 相关申请

[0003] 根据35U.S.C.§119,本申请要求2014年2月11日提交的美国临时申请61/938,608的权益,该申请的全部教导均通过引用并入本文。

背景技术

[0004] 大的DNA构建体的理性操作是当前合成生物学和基因组工程工作的主要挑战。近年来,已经开发了多种技术来应对这一挑战并提高能生成突变的特异性和速度。此外,适应性突变是进化的核心驱动因素,但其丰度和对细胞表型的相对贡献即使在大部分充分研究的生物中也知之甚少。这在很大程度上可以归因于与观察和重建这些基因型以及将其存在与感兴趣的表型进行关联相关的技术挑战。例如,依赖于随机诱变的基因组编辑方法导致包含许多突变的复杂的基因型,每一个突变的相对贡献难以解析。而且,由于缺乏关于各个突变的信息,难以确定等位基因之间的上位相互作用(epistatic interactions)。

发明内容

[0005] 成簇规律间隔短回文重复序列(Clustered Regularly Interspersed Short Palindromic Repeat)(CRISPR)存在于许多细菌基因组中,并已被发现在适应性细菌免疫中发挥重要作用。这些阵列的转录产生CRISPR RNA,它引导CRISPR/cas复合物序列特异性结合至细胞中的DNA靶标,导致基因阻抑或DNA裂解。这些复合物的特异性允许针对菌株工程化的新型体内应用。

[0006] 本文描述了理性、多路操作开放阅读框内的染色体(例如,以生成蛋白质文库)或在染色体的任何区段中的多个基因的方法,其中使用不同的CRISPR系统。这些方法提供了比先前可用的方法更有效的组合基因组工程化。

[0007] 扩展CRISPR的多路化(multiplexing)能力提出了当前的一个技术挑战,并且将能够使用这些系统生成高通量格式的理性文库。这样的进展对于力图重构复杂的遗传网络以实现最佳生产的代谢和蛋白质工程领域具有广泛深远的意义。

[0008] 所述方法包括向细胞中引入CRISPR系统的组件,包括CRISPR关联的核酸酶Cas9和序列特异性指导RNA(gRNA),从而利用CRISPR系统诱导序列指导的双链断裂的能力导致序列指导的双链断裂。可将CRISPR系统的组件,包括CRISPR关联的核酸酶Cas9和序列特异性指导RNA(gRNA),引入至细胞中,在一个或多个载体如质粒上编码。DNA重组工程盒(cassette)或编辑寡核苷酸可被理性地设计为包含位于靶基因座内的所需突变以及可被CRISPR系统识别的在靶基因座外部的常见位置中的突变。所述方法可以用于许多应用,包括改变感兴趣的途径。

[0009] 在一个实施方案中,所述方法是一种基因组工程化方法,其包括:(a)向细胞中引

入载体,该载体编码:(i)编辑盒,其包含与所述细胞中核酸的靶区域同源并且包含相对于该靶区域的至少一个核苷酸的突变(被称为所需突变)的区域,如相对于该靶区域的、在至少一个密码子中的至少一个核苷酸的突变,和前间隔区邻近基序(protospacer adjacent motif)(PAM)突变;(ii)启动子;以及(iii)至少一个指导RNA(gRNA),该gRNA包含:(a)与所述靶区域的一部分互补的区域(RNA);以及(b)募集Cas9核酸酶的区域(RNA),从而产生包含所述载体的细胞;(b)使包含所述载体的细胞维持在表达Cas9的条件下,其中Cas9核酸酶在所述载体上编码、在第二载体上编码或在细胞的基因组上编码,导致产生包含所述载体并且不包含所述PAM突变的细胞和包含所述载体和PAM突变的细胞;(c)在适合于细胞活力的条件下培养(b)的产物,从而产生活细胞;(d)获得在(c)中产生的活细胞;以及(e)对在(d)中获得的至少一个活细胞的载体的编辑寡核苷酸进行测序,并鉴定所述至少一个密码子的突变。

[0010] 在另一个实施方案中,所述方法是一种通过可追踪CRISPR富集重组工程化的基因组工程化方法,其包括:(a)向第一细胞群体中引入载体,该载体编码:(i)至少一个编辑盒,其包含:(a)与核酸的靶区域同源并且包含相对于该靶区域的至少一个核苷酸的突变的区域,如相对于该靶区域在至少一个密码子中的至少一个核苷酸的突变,和(b)前间隔区邻近基序(PAM)突变;(ii)至少一个启动子;以及(iii)至少一个指导RNA(gRNA),其包含:(a)与所述靶区域的一部分互补的区域(RNA),以及(b)募集Cas9核酸酶的区域(RNA),从而产生包含所述载体的第二细胞群体;(b)使所述第二细胞群体维持在表达Cas9核酸酶的条件下,其中所述Cas9核酸酶在所述载体、第二载体上或在所述第二细胞群体的细胞的基因组上编码,导致在不包含PAM突变的细胞中的DNA裂解和这类细胞的死亡;(c)获得在(b)中产生的活细胞;以及(d)通过对所述第二细胞群体的至少一个细胞的载体的编辑寡核苷酸进行测序,鉴定所述至少一个密码子的突变。

[0011] 以上实施方案中的任意一个可以进一步包括合成和/或获得编辑寡核苷酸群体。任意一个实施方案可以进一步包括扩增所述编辑寡核苷酸群体。在任意一个所述实施方案中,所述载体可以进一步包含间隔区、至少两个引发位点,或间隔区和至少两个引发位点两者。在一些实施方案中,所述编辑盒包含靶区域,该靶区域包含PAM突变的100个核苷酸内的至少一个密码子的突变。

[0012] 还描述了一种载体,其包含:

[0013] (i)编辑盒,其包含与细胞中核酸的靶区域同源并且包含相对于该靶区域的至少一个核苷酸的突变(被称为所需突变)的区域,和前间隔区邻近基序(PAM)突变;

[0014] (ii)启动子;以及

[0015] (iii)至少一个指导RNA(gRNA),其包含:(a)与所述靶区域的一部分互补的区域(RNA);以及(b)募集Cas9核酸酶的区域(RNA)。

[0016] 进一步的实施方案为一种载体,其包含:

[0017] (i)编辑盒,其包含与细胞中核酸的靶区域同源并且包含相对于该靶区域的、在至少一个密码子中的至少一个核苷酸的突变(被称为所需突变)的区域,和前间隔区邻近基序(PAM)突变;

[0018] (ii)启动子;以及

[0019] (iii)至少一个指导RNA(gRNA),其包含:(a)与所述靶区域的一部分互补的区域

(RNA);以及(b)募集Cas9核酸酶的区域(RNA)。

[0020] 进一步的实施方案为一种载体,其包含:

[0021] (i)至少一个编辑盒,其包含:(a)与核酸的靶区域同源并且包含相对于该靶区域的至少一个核苷酸的突变的区域,以及(b)前间隔区邻近基序(PAM)突变;

[0022] (ii)至少一个启动子;以及

[0023] (iii)至少一个指导RNA(gRNA),其包含:(a)与靶区域的一部分互补的区域(RNA),以及(b)募集Cas9核酸酶的区域(RNA)。

[0024] 所述载体的另一个实施方案为一种载体,其包含:

[0025] (i)至少一个编辑盒,其包含:(a)与核酸的靶区域同源并且包含相对于该靶区域的、在至少一个密码子中的至少一个核苷酸的突变的区域,以及(b)前间隔区邻近基序(PAM)突变;

[0026] (ii)至少一个启动子;以及

[0027] (iii)至少一个指导RNA(gRNA),其包含:(a)与靶区域的一部分互补的区域(RNA),和(b)募集Cas9核酸酶的区域(RNA)。

[0028] 在任意一个所述实施方案中,所述载体可以进一步包含间隔区;至少两个引发位点;或间隔区和至少两个引发位点。在突变是在至少一个密码子中的至少一个核苷酸的突变的那些载体中,所述编辑盒所述突变可以在例如PAM突变的100个核苷酸内。

[0029] 还描述了一种包含通过本文所述方法产生的细胞群体的文库。细胞群体的文库可以包含具有本文所述的任何载体的细胞。例如,细胞群体可以包含一种载体,该载体包含:

[0030] (i)编辑盒,其包含与细胞中核酸的靶区域同源并且包含相对于该靶区域的至少一个核苷酸的突变(被称为所需突变)的区域,和前间隔区邻近基序(PAM)突变;

[0031] (ii)启动子;以及

[0032] (iii)至少一个指导RNA(gRNA),其包含:(a)与所述靶区域的一部分互补的区域(RNA);以及(b)募集Cas9核酸酶的区域(RNA)。

[0033] 在进一步的实施方案中,细胞群体可以包含一种载体,该载体包含:

[0034] (i)编辑盒,其包含与细胞中核酸的靶区域同源并且包含相对于该靶区域的、在至少一个密码子中的至少一个核苷酸的突变(被称为所需突变)的区域,和前间隔区邻近基序(PAM)突变;

[0035] (ii)启动子;以及

[0036] (iii)至少一个指导RNA(gRNA),其包含:(a)与所述靶区域的一部分互补的区域(RNA);以及(b)募集Cas9核酸酶的区域(RNA)。

[0037] 在进一步的实施方案中,所述方法是一种CRISPR辅助的理性蛋白质工程化(组合基因组工程化)方法,该方法包括:

[0038] (a)通过将以下物质引入至第一细胞群体中(例如通过共转化)来构建包含重组DNA如重组染色体或质粒中的重组DNA的供体文库:(i)一个或多个编辑寡核苷酸,如理性设计的寡核苷酸,其使第一单个前间隔区邻近基序(PAM)的缺失与邻近PAM的基因(邻近基因)中的至少一个密码子的突变偶联,和(ii)指导RNA(gRNA),其靶向位于染色体的开放阅读框的5'侧的核苷酸序列,从而产生包含第一细胞群体的供体文库,该第一细胞群体包含具有靶向密码子突变的重组染色体;

[0039] (b)例如通过重组染色体的PCR扩增来扩增在(a)中构建的供体文库,其使用来自所述编辑寡核苷酸的合成特征,并同时在所述基因的3'末端处并入第二PAM缺失(终点(destination)PAM缺失),从而使靶向密码子突变与终点PAM缺失直接偶联如共价偶联,并产生携带所述终点PAM缺失和靶向密码子突变的重获(retrieved)的供体文库;以及

[0040] (c)将携带终点PAM缺失和靶向密码子突变的供体文库与终点gRNA质粒引入(例如,共转化)至第二细胞群体中,从而产生包含靶向密码子突变的终点文库,该第二细胞群体通常为稚细胞群体。

[0041] 所述第一细胞群体和所述第二细胞群体(例如,稚细胞群体)通常是这样的群体:其中细胞全部为相同的类型,并且可以是原核生物或真核生物,例如但不限于细菌、哺乳动物细胞、植物细胞、昆虫细胞。

[0042] 在一些实施方案中,所述方法进一步包括使所述终点文库维持在产生蛋白质的条件下。

[0043] 在一些实施方案中,所述第一细胞表达具有Cas9核酸酶活性的多肽。在一些实施方案中,在诱导型启动子的控制下表达具有Cas9核酸酶活性的多肽。

[0044] 在一些实施方案中,所述编辑寡核苷酸与存在于所述第一细胞中的(一个、一个或多个、至少一个)靶核酸互补。在一些实施方案中,所述编辑寡核苷酸靶向所述第一细胞中的多于一个靶位点或基因座。在一些实施方案中,所述编辑寡核苷酸的核酸序列[所需密码子]包含相对于靶核酸的一个或多个置换、缺失、插入,或置换、缺失和插入的任意组合。在一些实施方案中,理性地设计所述编辑寡核苷酸;在进一步的实施方案中,通过随机诱变或通过使用简并引物寡核苷酸产生所述编辑寡核苷酸。在一些实施方案中,所述编辑寡核苷酸来源于核酸的集合(文库)。

[0045] 在一些实施方案中,所述gRNA在质粒上编码。在一些实施方案中,通过转化,例如通过编辑寡核苷酸和指导(g)RNA的共转化,将所述编辑寡核苷酸和所述gRNA引入至所述第一细胞中。在一些实施方案中,将所述编辑寡核苷酸和所述gRNA依次引入至所述第一细胞中。在其他实施方案中,将所述编辑寡核苷酸和所述gRNA同时引入至所述第一细胞中。

[0046] 在一些实施方案中,重获供体文库进一步包括(a)筛选细胞以供并入所述编辑寡核苷酸,和(b)选择被确认已并入所述编辑寡核苷酸的细胞。在一些实施方案中,重获供体文库进一步包括处理重获的供体文库。

[0047] 在一些实施方案中,所述终点细胞/稚细胞表达具有Cas9核酸酶活性的多肽。在一些实施方案中,在诱导型启动子的控制下表达具有Cas9核酸酶活性的多肽。

[0048] 还描述了一种CRISPR辅助的理性蛋白质工程化方法,其包括:

[0049] (a)将(i)包含编辑寡核苷酸的合成dsDNA编辑盒和(ii)表达靶向正好在感兴趣的基因上游的基因组序列的指导RNA(gRNA)的载体,在发生所述编辑寡核苷酸的多路重组工程化和通过gRNA的选择性富集的条件下引入(例如,共转化)至第一细胞群体中,从而产生供体文库;

[0050] (b)扩增具有缺失与感兴趣的基因的3'端邻近的前间隔区邻近基序(PAM)(终点PAM)的寡核苷酸的供体文库,从而产生扩增的供体文库,该扩增的供体文库包含已缺失终点PAM(具有3'PAM缺失)的dsDNA编辑盒、理性密码子突变和P1位点;

[0051] (c)用酶如限制酶(例如,BsaI)处理所述扩增的供体文库以去除P1位点;以及

[0052] (d)用在(c)中处理的扩增的供体文库和终点gRNA共转化稚细胞群体,从而产生共转化细胞群体,该共转化细胞群体包含已经缺失终点PAM(具有3'PAM缺失)的dsDNA编辑盒、理性密码子突变和终点gRNA。

[0053] 在所描述的全部实施方案中,突变可以是所需的任何类型,如一个或多个插入、缺失、置换,或者前述突变中两个或三个的任意组合(例如,插入和缺失;插入和置换;缺失和置换;置换和插入;插入、缺失和置换)。插入、缺失和置换可以是任何数目的核苷酸的插入、缺失和置换。它们可以在密码子(编码区)中和/或在非编码区中。

附图说明

[0054] 图1A和1B显示了CRISPR辅助的理性蛋白质工程化(CARPE)的概览。图1A显示了供体文库构建的示意图。使合成dsDNA编辑盒与表达靶向在感兴趣的基因上游的基因组序列的指导RNA(gRNA)的载体共转化。该共转化经由编辑寡核苷酸的多路重组工程化生成供体文库,该编辑寡核苷酸通过gRNA选择性地富集。随后使用使得与所述基因的3'端邻近的PAM(终点PAM)发生突变(缺失)的寡核苷酸扩增供体文库。图1B显示了最终蛋白质文库生成的示意图。用BsaI处理供体文库以去除P1位点,并将具有3' PAM缺失和理性密码子突变的dsDNA盒的文库与终点gRNA共转化以生成最终蛋白质文库。

[0055] 图2显示了来自galK供体文库构建产生的克隆的DNA序列,证实了以高效率并入编辑寡核苷酸的P1特征以及在靶向密码子位置(下划线)处的突变。P1的序列由SEQ ID NO:1提供。

[0056] 图3A显示了引物设计。图3B显示了相对于引物数目的预期密度。

[0057] 图4A显示了接头与构建体结果。图4B显示了与基于乳液PCR的追踪相关的10个编辑(edit)。

[0058] 图5是用于代谢工程化的理性蛋白质编辑的示意图。

[0059] 图6是生成CRISPR富集的理性蛋白质文库的示意图。

[0060] 图7是CARPE的设置和证明的示意图。

[0061] 图8显示了迭代(iterative)CRISPR共选择的策略。

[0062] 图9显示了使用CARPE的多路蛋白质工程化的策略。

[0063] 图10显示了使用CARPE构建galK供体文库。

[0064] 图11A显示了使用CARPE的基于CRISPR的多路编辑的示意图。图11B显示了使用通过可追踪CRISPR富集重组工程化的基因组工程化(GEn-TraCER),基于CRISPR的多路编辑的示意图。

[0065] 图12显示了代表性的GEn-TraCER载体(构建体),其包含用于编辑galK的密码子24的编辑盒、启动子和间隔区。

[0066] 图13显示了使用GEn-TraCER的galK编辑的结果。上面的图显示了来自已经用galK密码子24编辑GEn-TraCER载体转化的细胞的染色体和载体(质粒)的DNA测序结果,其表明该载体上的编辑盒(寡核苷酸)可以作为允许高效追踪所需基因组编辑(突变)的"反式条形码"被测序。下面的图示出了细胞的DNA测序色谱,其显示出未编辑的野生型表型(红色)。该方法允许鉴定具有既携带未编辑的野生型等位基因又携带编辑的/突变的等位基因的多个染色体的细胞。

[0067] 图14A-14C显示了GEn-TraCER的示意图。图14A显示了设计组件的概览。GEn-TraCER盒含有指导RNA(gRNA)序列,以靶向细胞基因组中的特定位点并引起dsDNA裂解。与靶区域互补的同源性区域使得PAM和其他附近的所需位点发生突变。将经历重组的细胞选择性地富集为高丰度。对载体中的GEn-TraCER编辑盒的测序使得能够追踪基因组编辑/突变。图14B显示了针对密码子145处的大肠杆菌galK基因的示例性编辑盒设计。用最近的可用PAM突变删除PAM,该突变可以针对在最近的可用PAM位置处的同义变化而进行。这实现了在PAM缺失位点处具有1-2个核苷酸的“沉默疤痕(silent scar)”的诱变。图14C显示了可以使用基于阵列的合成方法合成GEn-TraCER盒,从而实现至少10^4-10^6个盒的平行合成,以用于系统靶向和同时评价在全基因组规模下对数以千计的突变的适合性。

[0068] 图15A显示了GEn-TraCER载体的概况。图15B显示了用于在大肠杆菌galK基因中生成Y145*突变的代表性GEn-TraCER的一部分,其中该PAM突变和突变的密码子相隔17个核苷酸。代表性GEn-TraCER的该部分的核酸序列由SEQ ID NO:28提供,而反向互补序列由SEQ ID NO:33提供。

[0069] 图16A-16C显示了GEn-TraCER设计的对照。图16A显示了编辑盒的大小对所述方法的效率的影响。图16B显示了PAM突变/缺失与所需突变之间的距离对所述方法的效率的影响。图16C显示了MutS系统的存在与否对所述方法的效率的影响。

具体实施方式

[0070] 细菌和古细菌CRISPR系统已成为用于精确基因组编辑的强大的新工具。已经在体外特别好地表征了来自酿脓链球菌(Streptococcus pyogenes)的II型CRISPR系统,并且已经建立了简单的设计规则以便对其双链DNA(dsDNA)结合活性重编程序(Jinek等人Science(2012)337(6096):816-821)。CRISPR介导的基因组编辑方法在很多种生物中的使用在文献中快速积累,该生物包括细菌(Cong等人Science(2013)339(6121):819-823)、酿酒酵母(DiCarlo等人Nucleic Acids Res.(2013)41:4336-4343)、秀丽隐杆线虫(Waaijers等人Genetics(2013)195:1187-1191)和各种哺乳动物细胞系(Cong等人Science(2013)339(6121):819-823;Wang等人Cell(2013)153:910-918)。像其他基于内切核酸酶的基因组编辑技术,如锌指核酸酶(ZFN)、归巢核酸酶和TALENS一样,CRISPR系统介导精确基因组编辑的能力源自于靶标识别的高度特异性的性质。例如,来自大肠杆菌和酿脓链球菌系统的I型CRISPR系统要求CRISPR RNA(crRNA)与14-15碱基对识别靶标之间有完全的互补性,这表明自然地采用了CRISPR系统的免疫功能(Jinek等人Science(2012)337(6096):816-821;Brouns等人Science(2008)321:960-964;Semenova等人PNAS(2011)108:10098-10103)。

[0071] 本文描述了基因组编辑方法,其采用内切核酸酶,如由cas9基因编码的Cas9核酸酶,以在DNA如基因组DNA中进行定向基因组进化/产生变化(缺失、置换、添加)。cas9基因可以获自任何来源,如细菌,如细菌酿脓链球菌。cas9的核酸序列和/或Cas9的氨基酸序列可以相对于天然存在的cas9和/或Cas9的序列发生突变;突变可以是例如一个或多个插入、缺失、置换或前述的两个或三个的任意组合。在这样的实施方案中,所得到的突变的Cas9可以具有相对于天然存在的Cas9增强或降低的核酸酶活性。

[0072] 图1A、1B和11A显示了被称为CRISPR辅助的理性蛋白质工程化(CARPE)的CRISPR介导的基因组编辑方法。CARPE是一个两阶段的构建过程,它依赖于将来自单链DNA(ssDNA)或

双链DNA(dsDNA)编辑盒的定向突变直接并入基因组中的“供体”和“终点”文库的生成。在供体构建的第一阶段(图1A)中,将理性设计的编辑寡核苷酸与指导RNA(gRNA)共转化至细胞中,该gRNA杂交至/靶向靶DNA序列,如位于开放阅读框的5’侧的序列或其他感兴趣的序列。CARPE的关键创新是编辑寡核苷酸的设计,它使单个前间隔区邻近基序(PAM)的缺失或突变与邻近基因中的一个或多个所需密码子的突变偶联,从而使得能够在单一转化中生成整个供体文库。随后使用来自编辑寡核苷酸的合成特征,通过扩增重组染色体,例如通过PCR反应,来重获供体文库;同时在所述基因的3’末端处并入第二PAM缺失或突变。因此,该方法将密码子靶向突变与PAM缺失直接共价偶联。在CARPE的第二阶段(图1B)中,将携带终点PAM缺失/突变和靶向突变(一个或多个核苷酸如在一个或多个密码子中的一个或多个核苷酸的所需突变)的PCR扩增的供体文库与终点gRNA载体共转化至稚细胞中,以生成表达理性设计的蛋白质文库的细胞群体。

[0073] 在CRISPR系统中,CRISPR反式激活(tracrRNA)和间隔区RNA(crRNA)引导靶区域的选择。如本文所用的,靶区域是指细胞或细胞群体的核酸中需要至少一个核苷酸的突变如在至少一个密码子(一个或多个密码子)中的至少一个核苷酸的突变的任何基因座。靶区域可以是例如基因组基因座(靶基因组序列)或染色体外的基因座。tracrRNA和crRNA可被表达为单一的嵌合RNA分子,其被称为单一指导RNA、指导RNA或gRNA。gRNA的核酸序列包含第一核酸序列(也被称为第一区域,它与靶区域的区域互补)和第二核酸序列(也被称为第二区域,它形成茎环结构并用来将Cas9募集至靶区域)。在一些实施方案中,gRNA的第一区域与在靶基因组序列上游的区域互补。在一些实施方案中,gRNA的第一区域与靶区域的至少一部分互补。gRNA的第一区域可以与靶基因组序列完全互补(100%互补)或包含一个或多个错配,条件是它与靶基因组序列充分互补以特异性地杂交/引导和募集Cas9。在一些实施方案中,gRNA的第一区域为至少15、16、17、18、19、20、21、22、23、24、25、26、27、28、29个或至少30个核苷酸的长度。在一些实施方案中,gRNA的第一区域为至少20个核苷酸的长度。在一些实施方案中,由第二核酸序列形成的茎环结构为至少50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、7、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100个核苷酸的长度。在特定的实施方案中,该茎环结构为80至90或82至85个核苷酸的长度,并且在进一步的特定实施方案中,形成茎环结构的gRNA的第二区域为83个核苷酸的长度。

[0074] 在一些实施方案中,使用CARPE方法引入至第一细胞中的(供体文库的)gRNA的序列与引入至第二细胞/稚细胞中的(终点文库的)gRNA的序列相同。在一些实施方案中,将多于一个gRNA引入至第一细胞群体和/或第二细胞群体。在一些实施方案中,所述多于一个gRNA分子包含与多于一个靶区域互补的第一核酸序列。

[0075] 在CARPE方法中,用于在所述方法中使用的也被称为编辑寡核苷酸的双链DNA盒可以获自或来源于许多来源。例如,在一些实施方案中,dsDNA盒来源于已经通过非同源性随机重组(NRR)多样化的核酸文库;这样的文库被称为NRR文库。在一些实施方案中,例如通过基于阵列的合成来合成所述编辑寡核苷酸。所述编辑寡核苷酸的长度可能依赖于获得该编辑寡核苷酸所使用的方法。在一些实施方案中,所述编辑寡核苷酸为约50-200个核苷酸、75-150个核苷酸或80-120个核苷酸的长度。

[0076] 编辑寡核苷酸包含(a)与细胞的核酸的靶区域同源并且包含相对于该靶区域的至

少一个密码子的突变(被称为所需突变)的区域,和(b)前间隔区邻近基序(PAM)突变。该PAM突变可以是一个或多个核苷酸的任何插入、缺失或置换,其使PAM的序列发生突变,使得其不再被CRISPR系统识别。包含这样的PAM突变的细胞可以被称为对CRISPR介导的杀伤“有免疫力”。相对于靶区域的序列的所需突变可以是在靶区域的至少一个密码子处的一个或多个核苷酸的插入、缺失和/或置换。

[0077] 仅为了说明的目的,以下参考细菌基因描述了CARPE方法。所述方法可以应用于任何感兴趣的基因,包括来自包括细菌和古细菌在内的任何原核生物或来自任何真核生物的基因,包括酵母和哺乳动物(包括人)基因。对大肠杆菌基因组中的galK基因进行CARPE方法,这部分地是由于针对该基因的活性测定的可用性。使用BW23115亲代菌株和pSIM5载体(Datta等人Gene(2008)379:109-115)进行所述方法以介导重组工程化。将cas9基因克隆至pBTBX-2骨架中,处于pBAD启动子的控制下,以允许通过添加阿拉伯糖控制裂解活性。使用来自NNK文库的dsDNA盒对选择性并入合成dsDNA盒(127bp)的能力进行评估,该NNK文库由简并引物和/或由作为27,000个成员的文库的一部分合成的理性设计的寡核苷酸(oligo)经由微阵列技术构建。在这两种情况下,将寡核苷酸设计为使galK基因产物的活性位点残基发生突变。基于用针对galK基因座的引物获得的扩增子大小的变化,验证了供体菌株文库的高度有效的恢复(recovery)。对来自NRR文库的这些菌落PCR产物的测序表明,来自dsDNA盒的合成引发位点(P1)以约90-100%的效率并入。这表明可以在不依靠通常在其他基于重组工程化的编辑方法中使用的易错mutS敲除菌株(Costantino等人PNAS(2003)100:15748-15753;Wang等人Nature(2009)460:894-898)的情况下,以高效率生成这些文库。密码子突变的效率有所降低(约20%),这可能是由于在等位基因替换过程中的mutS校正。终点文库中克隆的初步评估表明,当构建的两个阶段都在$mutS^{+}$背景下进行时,最终密码子编辑效率约为10%。

[0078] 使用不将PAM与密码子突变共价连接而是依赖于它们在复制过程中彼此的接近性的替代方案,完成与其他近来发表的可共选择编辑方案的比较(Wang等人Nat.Methods(2012)9:591-593)。在这些非共价的实验中,使用与上述相同的编辑寡核苷酸,并使用靶向相同供体/终点PAM位点的ssDNA寡核苷酸努力针对其插入进行共选择。产生的突变体的菌落筛选揭示了PAM突变体的恢复的高效率。然而,对于dsDNA编辑盒的插入似乎没有强大的共选择。这可能是由于PAM缺失寡核苷酸与生成相当大的染色体缺失的编辑盒在相对重组工程化效率上的巨大差异。

[0079] 可以例如通过在转移到野生型供体菌株之前在mutS缺乏的菌株中进行供体构建以试图防止在供体构建阶段中突变的损失,来评估改善CARPE方法的最终编辑效率的能力。此外,可以例如通过对包括dxs、metA和folA在内的多个必需基因使用CARPE来评估CARPE方法的通用性。已经使用描述的gRNA设计策略有效地靶向必需基因。结果还表明,尽管在供体文库创建过程中发生了基因破坏,但可以在重组工程化后1-3小时内有效地构建并重获供体文库。

[0080] 本文还提供了使用CRISPR介导的系统的可追踪精确基因组编辑方法,其被称为通过可追踪CRISPR富集重组工程化的基因组工程化(GEn-TraCER)。GEn-TraCER方法使用编码编辑盒和gRNA的单个载体获得高效率的编辑/突变。当与平行DNA合成如基于阵列的DNA合成一起使用时,GEN-TraCER提供了数以千计的精确编辑/突变的单步生成,并使得通过对载

体上的编辑盒进行测序而不是通过对细胞的基因组(基因组DNA)进行测序来定位突变成为可能。所述方法在蛋白质和基因组工程化应用方面以及对于突变如实验室进化实验中鉴定的突变的重建具有广泛应用。

[0081] GEn-TraCER方法和载体将包含所需突变和PAM突变的编辑盒与编码单一载体上的gRNA的基因组合起来,这使得在单一反应中生成突变的文库成为可能。如图11B所示,所述方法包括将包含含有所需突变和PAM突变的编辑盒的载体引入至细胞或细胞群体中。在一些实施方案中,引入载体的细胞还编码Cas9。在一些实施方案中,随后将编码Cas9的基因引入至细胞或细胞群体中。在该细胞或细胞群体中包括Cas9和gRNA的CRISPR系统的表达得到激活;gRNA将Cas9募集至靶区域,在靶区域中发生dsDNA裂解。不希望受到任何特定理论的束缚,与靶区域互补的编辑盒的同源区域使得PAM和靶区域的一个或多个密码子发生突变。该细胞群体中没有整合PAM突变的细胞经历由Cas9介导的dsDNA裂解引起的未编辑的细胞死亡。该细胞群体中整合了PAM突变的细胞不经历细胞死亡;它们保持存活并且选择性地富集为高丰度。获得活细胞并提供靶向突变的文库。

[0082] 使用GEn-TraCER的可追踪基因组编辑方法包括:(a)将编码至少一个编辑盒、启动子和至少一个gRNA的载体引入至细胞或细胞群体中,从而产生包含该载体的细胞或细胞群体(第二细胞群体);(b)使第二细胞群体维持在表达Cas9的条件下,其中Cas9核酸酶在该载体、第二载体上或在第二细胞群体的细胞的基因组上编码,导致第二细胞群体中不包含PAM突变的细胞的DNA裂解和死亡,而第二细胞群体中包含PAM突变的细胞存活;(c)获得活细胞;以及(d)对第二细胞群体的至少一个细胞中的载体的编辑盒进行测序,以鉴定至少一个密码子的突变。

[0083] 在一些实施方案中,还将编码cas9的单独载体引入至细胞或细胞群体中。可使用本领域已知的任何方法或技术将载体引入至细胞或细胞群体中。例如,可以通过标准方案如转化(包括化学转化和电穿孔)、转导和粒子轰击引入载体。

[0084] 编辑盒包含(a)区域,该区域识别(杂交)细胞或细胞群体中核酸的靶区域,与该细胞的核酸的靶区域同源并且包含相对于该靶区域的、在至少一个密码子中的至少一个核苷酸的突变(被称为所需突变),和(b)前间隔区邻近基序(PAM)突变。该PAM突变可以是一个或多个核苷酸的任何插入、缺失或置换,其使PAM的序列发生突变,使得突变的PAM(PAM突变)不被CRISPR系统识别。包含如PAM突变的细胞可以被称为对CRISPR介导的杀伤“有免疫力”。相对于靶区域的序列的所需突变可以是在靶区域的至少一个密码子处的一个或多个核苷酸的插入、缺失和/或置换。在一些实施方案中,所述编辑盒上PAM突变与所需突变之间的距离为至少5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30个核苷酸。在一些实施方案中,PAM突变位于距离编辑盒的末端至少9个核苷酸的位置处。在一些实施方案中,所需突变位于距离编辑盒的末端至少9个核苷酸的位置处。

[0085] 在一些实施方案中,相对于靶区域的序列的所需突变为核酸序列的插入。插入靶区域中的核酸序列可以为任意长度。在一些实施方案中,插入的核酸序列为至少50、100、150、200、250、300、350、400、450、500、550、600、650、700、750、800、850、900、950、1000、1100、1200、1300、1400、1500、1600、1700、1800、1900或至少2000个核苷酸的长度。在核酸序列插入靶区域中的实施方案中,所述编辑盒包含长度为至少30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59或至少60个核

苷酸并且与靶区域同源的区域。

[0086] 术语“GEn-TraCER盒”可以用来指编辑盒、启动子、间隔区序列和编码gRNA的基因的至少一部分。在一些实施方案中,在GEn-TraCER盒上编码gRNA的基因的部分编码gRNA中与靶区域互补的部分。在一些实施方案中,gRNA中与靶区域互补的部分为至少15、16、17、18、19、20、21、22、23、24、25、26、27、28、29或至少30个核苷酸的长度。在一些实施方案中,gRNA中与靶区域互补的部分为24个核苷酸的长度。在一些实施方案中,GEn-TraCER盒进一步包含至少两个引发位点。在一些实施方案中,所述引发位点可以用于通过例如PCR来扩增GEn-TraCER盒。在一些实施方案中,使用gRNA中与靶区域互补的部分作为引发位点。

[0087] 在GEn-TraCER方法中,用于在所述方法中使用的编辑盒和GEn-TraCER盒可以获自或来源于许多来源。例如,在一些实施方案中,例如通过基于阵列的合成来合成所述编辑盒。在一些实施方案中,例如通过基于阵列的合成来合成所述GEn-TraCER盒。所述编辑盒和/或GEn-TraCER盒的长度可依赖于获得编辑盒和/或GEn-TraCER盒所使用的方法。在一些实施方案中,所述编辑盒为约50-300个核苷酸、75-200个核苷酸或80-120个核苷酸的长度。在一些实施方案中,所述GEn-TraCER盒为约50-300个核苷酸、75-200个核苷酸或80-120个核苷酸的长度。

[0088] 在一些实施方案中,所述方法还包括例如通过基于阵列的合成获得GEn-TraCER盒以及构建所述载体。构建载体的方法是本领域普通技术人员已知的,并且可以包括将GEn-TraCER盒连接至载体中。在一些实施方案中,在构建载体之前例如通过PCR来扩增GEn-TraCER盒或GEn-TraCER盒的亚组(库)。

[0089] 在表达Cas9的条件下维持或培养包含所述载体并且还编码Cas9的细胞或细胞群体。可以控制Cas9表达。本文描述的方法包括使细胞维持在Cas9表达被激活的条件下,这导致产生Cas9。表达Cas9的具体条件将取决于诸如用来调节Cas9表达的启动子的性质等因素。在一些实施方案中,在诱导物分子如阿拉伯糖的存在下,Cas9表达受到诱导。当包含编码Cas9的DNA的细胞或细胞群体在诱导物分子的存在下时,发生Cas9的表达。在一些实施方案中,在阻抑物分子的存在下,Cas9表达受到抑制。当包含编码Cas9的DNA的细胞或细胞群体在不存在阻抑Cas9表达的分子时,发生Cas9的表达。

[0090] 从经历由Cas9介导的杀伤引起的未编辑的细胞死亡的细胞中获得或分离出保持存活的细胞群体的细胞;这可以通过例如将所述细胞群体涂布在培养表面上以允许活细胞生长(该活细胞随后可用于评估)来完成。

[0091] 使用GEn-TraCER方法,通过对所述群体的活细胞(整合了PAM突变的细胞)中载体上的编辑盒进行测序,与PAM突变偶联的所需突变是可追踪的。这允许在不需要对细胞的基因组进行测序的情况下容易地鉴定突变。所述方法包括对编辑盒进行测序以鉴定一个或多个密码子的突变。可以对作为所述载体的组件的编辑盒进行测序,或在从所述载体中分离出编辑盒和任选扩增之后进行测序。可以使用本领域已知的任何测序方法如通过Sanger测序进行测序。

[0092] 本文所述的方法可以在CRISPR系统可起作用(例如,靶向和裂解DNA)的任何类型的细胞(包括原核细胞和真核细胞)中进行。在一些实施方案中,该细胞是细菌细胞,如埃希氏菌属的种(Escherichia spp.)(例如,大肠杆菌)。在其他实施方案中,该细胞是真菌细胞,如酵母细胞,例如,酵母属的种(Saccharomyces spp.)。在其他实施方案中,该细胞是藻

细胞、植物细胞、昆虫细胞或哺乳动物细胞,包括人细胞。

[0093] “载体”是包含所需序列或有待递送至细胞或在细胞中表达的序列的多种核酸中的任意一种。例如,可以通过限制和连接或通过重组将所需序列包含在载体中。载体通常由DNA组成,尽管RNA载体也是可用的。载体包括但不限于:质粒、粘粒(fosmid)、噬菌粒、病毒基因组和人工染色体。

[0094] 在GEN-TraCER方法中有用的载体包含至少一个如本文所述的编辑盒、启动子和至少一个编码gRNA的基因。在一些实施方案中,载体上包含多于一个编辑盒(例如2、3、4、5、6、7、8、9、10个或更多个编辑盒)。在一些实施方案中,所述多于一个编辑盒与不同的靶区域同源(例如,存在不同的编辑盒,每一个编辑盒与不同的靶区域同源)。可替代地或除此之外,所述载体可以包含编码多于一个gRNA(例如,2、3、4、5、6、7、8、9、10个或更多个gRNA)的多于一个基因。在一些实施方案中,所述多于一个gRNA含有与不同靶区域的一部分互补的区域(例如,存在不同的gRNA,每一个gRNA与不同靶区域的一部分互补)。

[0095] 在一些实施方案中,将包含至少一个编辑盒、启动子和编码gRNA的一部分的基因的GEn-TraCER盒连接至编码gRNA的另一部分的载体中。经连接,来自GEn-TraCER盒的gRNA的部分与gRNA的另一部分连接并形成功能性gRNA。

[0096] 启动子和编码gRNA的基因可操作地连接。在一些实施方案中,所述方法包括引入编码Cas9的第二载体。在这样的实施方案中,该载体可以进一步包含与编码Cas9的基因可操作地连接的一个或多个启动子。如本文所用的,“可操作地”连接意指启动子影响或调节诸如编码gRNA的基因或编码Cas9的基因等基因的编码DNA的转录。该启动子可以是天然启动子(在引入载体的细胞中存在的启动子)。在一些实施方案中,该启动子是诱导型或阻抑型启动子(调节该启动子以允许诸如编码gRNA的基因或编码Cas9的基因等基因的诱导型或阻抑型转录),如通过分子(例如,诱导物或阻抑物)的存在或不存在调节的启动子。表达gRNA所需的启动子的性质可以根据物种或细胞类型而变化,并且将被本领域普通技术人员所认识到。

[0097] 在一些实施方案中,所述方法包括在引入包含至少一个如本文所述的编辑盒、启动子和至少一个gRNA的载体之前或同时,将编码Cas9的单独载体引入至细胞或细胞群体中。在一些实施方案中,将编码Cas9的基因整合到细胞或细胞群体的基因组中。可在引入包含至少一个如本文所述的编辑盒、启动子和至少一个gRNA的载体之前,或在引入包含至少一个如本文所述的编辑盒、启动子和至少一个gRNA的载体之后,将编码Cas9的DNA整合到细胞基因组中。或者,核酸分子,如编码Cas9的DNA,可以由整合到基因组中的DNA表达。在一些实施方案中,将编码Cas9的基因整合到细胞的基因组中。

[0098] 在本文所述的GEn-TraCER方法中有用的载体可以进一步包含间隔区序列、两个或更多个引发位点,或者间隔区序列和两个或更多个引发位点两者。在一些实施方案中,位于GEn-TraCER盒侧翼的引发位点的存在允许编辑盒、启动子和gRNA核酸序列的扩增。

[0099] 本发明提供了包括但不限于以下实施方案:

[0100] 1.一种基因组工程化方法,其包括:

[0101] (a)向细胞中引入载体,该载体编码:

[0102] (i)编辑盒,其包含与所述细胞中核酸的靶区域同源并且包含相对于该靶区域的至少一个核苷酸的突变(被称为所需突变)的区域,和前间隔区邻近基序(PAM)突变;

[0103] (ii)启动子;以及

[0104] (iii)至少一个指导RNA(gRNA),其包含:

[0105] (a)与所述靶区域的一部分互补的区域(RNA);以及

[0106] (b)募集Cas9核酸酶的区域(RNA),从而产生包含所述载体的细胞;

[0107] (b)使包含所述载体的细胞维持在表达Cas9的条件下,其中Cas9核酸酶在所述载体上编码、在第二载体上编码或在所述细胞的基因组上编码,导致产生包含所述载体并且不包含所述PAM突变的细胞和包含所述载体和所述PAM突变的细胞;

[0108] (c)在适合于细胞活力的条件下培养(b)的产物,从而产生活细胞;

[0109] (d)获得在(c)中产生的活细胞;以及

[0110] (e)对在(d)中获得的至少一个活细胞的载体的编辑寡核苷酸进行测序,并鉴定至少一个密码子的突变。

[0111] 2.一种通过可追踪CRISPR富集重组工程化的基因组工程化方法,其包括:

[0112] (a)向第一细胞群体中引入载体,该载体编码:

[0113] (i)至少一个编辑盒,其包含:(a)与核酸的靶区域同源并且包含相对于该靶区域的突变的区域,和(b)前间隔区邻近基序(PAM)突变;

[0114] (ii)至少一个启动子;以及

[0115] (iii)至少一个指导RNA(gRNA),其包含:(a)与所述靶区域的一部分互补的区域(RNA),和(b)募集Cas9核酸酶的区域(RNA),从而产生包含所述载体的第二细胞群体;

[0116] (b)使所述第二细胞群体维持在表达Cas9核酸酶的条件下,其中所述Cas9核酸酶在所述载体、第二载体上或在所述第二细胞群体的细胞的基因组上编码,导致在不包含所述PAM突变的细胞中的DNA裂解和这类细胞的死亡;

[0117] (c)获得在(b)中产生的活细胞;以及

[0118] (d)通过对所述第二细胞群体的至少一个细胞的载体的编辑寡核苷酸进行测序,鉴定至少一个密码子的突变。

[0119] 3.根据实施方案1或实施方案2所述的方法,其进一步包括合成和/或获得编辑寡核苷酸的群体。

[0120] 4.根据实施方案1至3中任一项所述的方法,其进一步包括扩增编辑寡核苷酸的群体。

[0121] 5.根据实施方案1-4中任一项所述的方法,其中所述载体进一步包含间隔区和/或至少两个引发位点。

[0122] 6.根据实施方案1-5中任一项所述的方法,其中所述编辑盒包含靶区域,该靶区域包含所述PAM突变的100个核苷酸内的至少一个密码子的突变。

[0123] 7.一种CRISPR辅助的理性蛋白质工程化(组合基因组工程化)方法,其包括:

[0124] (a)通过将以下物质引入至第一细胞群体中(用以下物质共转化第一细胞群体)来构建包含重组DNA如重组染色体或质粒中的重组DNA的供体文库:(i)一个或多个编辑寡核苷酸,如理性设计的寡核苷酸,其使第一单个前间隔区邻近基序(PAM)的缺失与邻近PAM的基因(邻近基因)中的至少一个密码子的突变偶联,和(ii)指导RNA(gRNA),其靶向位于染色体的开放阅读框的5'侧的核苷酸序列,从而产生包含第一细胞群体的供体文库,该第一细胞群体包含具有靶向密码子突变的重组染色体;

[0125] (b)通过重组DNA(如重组染色体)的扩增(如PCR扩增)重获在(a)中构建的供体文库,其使用来自所述编辑寡核苷酸的合成特征,并同时在所述基因的3'末端处并入第二PAM缺失(终点PAM缺失),从而使靶向密码子突变与终点PAM缺失直接共价偶联,并产生携带所述终点PAM缺失和靶向密码子突变的重获的供体文库;以及

[0126] (c)将携带所述终点PAM缺失和靶向密码子突变的重获的供体文库与终点gRNA质粒共转化至稚细胞群体中,从而产生包含靶向密码子突变的终点文库。

[0127] 8.根据实施方案7所述的方法,其进一步包括使所述终点文库维持在产生蛋白质的条件下。

[0128] 9.根据实施方案7或实施方案8所述的方法,其中所述第一细胞群体表达具有Cas9核酸酶活性的多肽。

[0129] 10.根据实施方案9所述的方法,其中在诱导型启动子的控制下表达具有Cas9核酸酶活性的多肽。

[0130] 11.根据实施方案10所述的方法,其中所述编辑寡核苷酸与存在于第一细胞中的靶核酸互补。

[0131] 12.根据实施方案11所述的方法,其中所述编辑寡核苷酸靶向所述第一细胞中的多于一个靶位点或基因座。

[0132] 13.根据实施方案7-12中任一项所述的方法,其中所述编辑寡核苷酸的核酸序列包含相对于所述靶核酸的一个或多个置换、缺失、插入。

[0133] 14.根据实施方案7-13中任一项所述的方法,其中理性地设计所述编辑寡核苷酸。

[0134] 15.根据实施方案7-14中任一项所述的方法,其中使用简并引物寡核苷酸产生所述编辑寡核苷酸。

[0135] 16.根据实施方案7所述的方法,其中所述编辑寡核苷酸来源于核酸的集合(文库)。

[0136] 17.根据实施方案7所述的方法,其中所述gRNA在质粒上编码。

[0137] 18.根据实施方案7所述的方法,其中通过转化将所述编辑寡核苷酸和所述gRNA引入至所述第一细胞。

[0138] 19.根据实施方案7所述的方法,其中将所述编辑寡核苷酸和所述gRNA依次引入至所述第一细胞。

[0139] 20.根据实施方案7所述的方法,其中将所述编辑寡核苷酸和所述gRNA同时引入至所述第一细胞。

[0140] 21.根据实施方案7所述的方法,其中重获所述供体文库进一步包括以下步骤:(a)筛选细胞以供并入所述编辑寡核苷酸,以及(b)选择被确认已并入所述编辑寡核苷酸的细胞。

[0141] 22.根据实施方案7所述的方法,其中重获所述供体文库进一步包括处理重获的供体文库。

[0142] 23.根据实施方案7所述的方法,其中终点细胞/稚细胞表达具有Cas9核酸酶活性的多肽。

[0143] 24.根据实施方案23所述的方法,其中在诱导型启动子的控制下表达具有Cas9核酸酶活性的多肽。

[0144] 25.一种CRISPR辅助的理性蛋白质工程化方法,其包括:

[0145] (a)将(i)包含编辑寡核苷酸的合成dsDNA编辑盒和(ii)表达靶向正好在感兴趣的基因上游的基因组序列的指导RNA(gRNA)的载体,在发生所述编辑寡核苷酸的多路重组工程化和通过gRNA的选择性富集的条件下,引入(共转化)至第一细胞群体中,从而产生供体文库;

[0146] (b)扩增具有缺失与感兴趣的基因的3'端邻近的前间隔区邻近基序(PAM)(终点PAM)的寡核苷酸的供体文库,从而产生扩增的供体文库,该扩增的供体文库包含具有3'PAM缺失的dsDNA编辑盒、理性密码子突变和P1位点;

[0147] (c)用BsaI处理所述扩增的供体文库以去除所述P1位点;以及

[0148] (d)用在(c)中产生的扩增的供体文库和终点gRNA共转化稚细胞群体,从而产生共转化细胞群体,该共转化细胞群体包含具有3'PAM缺失的dsDNA编辑盒、理性密码子突变和终点gRNA。

[0149] 26.根据实施方案25所述的方法,其进一步包括使在(d)中产生的共转化细胞群体维持在发生蛋白质表达的条件下。

[0150] 27.根据实施方案25或26所述的方法,其中第一细胞表达具有Cas9核酸酶活性的多肽。

[0151] 28.根据实施方案27所述的方法,其中在诱导型启动子的控制下表达具有Cas9核酸酶活性的多肽。

[0152] 29.根据实施方案25所述的方法,其中所述编辑寡核苷酸与存在于第一细胞中的靶核酸互补。

[0153] 30.根据实施方案29所述的方法,其中所述编辑寡核苷酸靶向所述第一细胞中的多于一个靶位点或基因座。

[0154] 31.根据实施方案25-30中任一项所述的方法,其中所述编辑寡核苷酸的核酸序列包含相对于所述靶核酸的一个或多个置换、缺失、插入。

[0155] 32.根据实施方案25-31中任一项所述的方法,其中理性地设计所述编辑寡核苷酸。

[0156] 33.根据实施方案25-31中任一项所述的方法,其中使用简并引物寡核苷酸产生所述编辑寡核苷酸。

[0157] 34.根据实施方案25所述的方法,其中所述编辑寡核苷酸来源于核酸的集合(文库)。

[0158] 35.根据实施方案25所述的方法,其中所述gRNA在质粒上编码。

[0159] 36.根据实施方案25所述的方法,其中通过转化将所述编辑寡核苷酸和所述gRNA引入至所述第一细胞。

[0160] 37.根据实施方案25所述的方法,其中将所述编辑寡核苷酸和所述gRNA依次引入至所述第一细胞群体。

[0161] 38.根据实施方案25所述的方法,其中将所述编辑寡核苷酸和所述gRNA同时引入至所述第一细胞。

[0162] 39.根据实施方案25所述的方法,其中重获所述供体文库进一步包括以下步骤:(a)筛选细胞以供并入所述编辑寡核苷酸,以及(b)选择被确认已并入所述编辑寡核苷酸的

细胞。

[0163] 40.根据实施方案25所述的方法,其中重获所述供体文库进一步包括处理重获的供体文库。

[0164] 41.根据实施方案25所述的方法,其中终点细胞/稚细胞表达具有Cas9核酸酶活性的多肽。

[0165] 42.根据实施方案38所述的方法,其中在诱导型启动子的控制下表达具有Cas9核酸酶活性的多肽。

[0166] 43.一种载体,其包含:

[0167] (i)编辑盒,其包含与细胞中核酸的靶区域同源并且包含相对于该靶区域的至少一个核苷酸的突变(被称为所需突变)的区域,和前间隔区邻近基序(PAM)突变;

[0168] (ii)启动子;以及

[0169] (iii)至少一个指导RNA(gRNA),其包含:

[0170] (a)与所述靶区域的一部分互补的区域(RNA);以及

[0171] (b)募集Cas9核酸酶的区域(RNA)。

[0172] 44.一种载体,其包含:

[0173] (i)编辑盒,其包含与细胞中核酸的靶区域同源并且包含相对于该靶区域的、在至少一个密码子中的至少一个核苷酸的突变(被称为所需突变)的区域,和前间隔区邻近基序(PAM)突变;

[0174] (ii)启动子;以及

[0175] (iii)至少一个指导RNA(gRNA),其包含:

[0176] (a)与所述靶区域的一部分互补的区域(RNA);以及

[0177] (b)募集Cas9核酸酶的区域(RNA)。

[0178] 45.一种载体,其包含:

[0179] (i)至少一个编辑盒,其包含:(a)与核酸的靶区域同源并且包含相对于该靶区域的至少一个核苷酸的突变的区域,和(b)前间隔区邻近基序(PAM)突变;

[0180] (ii)至少一个启动子;以及

[0181] (iii)至少一个指导RNA(gRNA),其包含:(a)与所述靶区域的一部分互补的区域(RNA),和(b)募集Cas9核酸酶的区域(RNA)。

[0182] 46.一种载体,其包含:

[0183] (i)至少一个编辑盒,其包含:(a)与核酸的靶区域同源并且包含相对于该靶区域的、在至少一个密码子中的至少一个核苷酸的突变的区域,和(b)前间隔区邻近基序(PAM)突变;

[0184] (ii)至少一个启动子;以及

[0185] (iii)至少一个指导RNA(gRNA),其包含:(a)与所述靶区域的一部分互补的区域(RNA),和(b)募集Cas9核酸酶的区域(RNA)。

[0186] 47.根据实施方案43至46中任一项所述的载体,其中所述载体进一步包含间隔区;至少两个引发位点;或间隔区和至少两个引发位点。

[0187] 48.根据实施方案44、46或47中任一项所述的载体,其中所述编辑盒包含靶区域,该靶区域包含PAM突变的100个核苷酸内的至少一个密码子的突变。

[0188] 49.一种文库,其包含通过实施方案1-48中任一项所述的方法产生的细胞群体。

[0189] 50.一种包含细胞群体的文库,该细胞群体包含实施方案43-48中任一项所述的载体。

[0190] 51.一种包含载体的细胞群体,该载体包含:

[0191] (i)编辑盒,其包含与细胞中核酸的靶区域同源并且包含相对于该靶区域的至少一个核苷酸的突变(被称为所需突变)的区域,和前间隔区邻近基序(PAM)突变;

[0192] (ii)启动子;以及

[0193] (iii)至少一个指导RNA(gRNA),其包含:

[0194] (a)与所述靶区域的一部分互补的区域(RNA);以及

[0195] (b)募集Cas9核酸酶的区域(RNA)。

[0196] 52.一种包含载体的细胞群体,其包含:

[0197] (i)编辑盒,其包含与细胞中核酸的靶区域同源并且包含相对于该靶区域的、在至少一个密码子中的至少一个核苷酸的突变(被称为所需突变)的区域,和前间隔区邻近基序(PAM)突变;

[0198] (ii)启动子;以及

[0199] (iii)至少一个指导RNA(gRNA),其包含:

[0200] (a)与所述靶区域的一部分互补的区域(RNA);以及

[0201] (b)募集Cas9核酸酶的区域(RNA)。

[0202] 53.一种基因组工程化方法,其包括:

[0203] (a)向细胞中引入载体,该载体编码:

[0204] (i)编辑盒,其包含与所述细胞中核酸的靶区域同源并且包含相对于该靶区域的、在至少一个密码子中的至少一个核苷酸的突变(被称为所需突变)的区域,和前间隔区邻近基序(PAM)突变;

[0205] (ii)启动子;以及

[0206] (iii)至少一个指导RNA(gRNA),其包含:

[0207] (a)与所述靶区域的一部分互补的区域(RNA);以及

[0208] (b)募集Cas9核酸酶的区域(RNA),从而产生包含所述载体的细胞;

[0209] (b)使包含所述载体的细胞维持在表达Cas9的条件下,其中Cas9核酸酶在所述载体上编码、在第二载体上编码或在所述细胞的基因组上编码,导致产生包含所述载体并且不包含所述PAM突变的细胞和包含所述载体和PAM突变的细胞;

[0210] (c)在适合于细胞活力的条件下培养(b)的产物,从而产生活细胞;

[0211] (d)获得在(c)中产生的活细胞;以及

[0212] (e)对在(d)中获得的至少一个活细胞的载体的编辑寡核苷酸进行测序,并鉴定至少一个密码子的突变。

[0213] 实施例

[0214] 实施例1:使用CARPE方法编辑galK

[0215] 对大肠杆菌基因组中的半乳糖激酶基因galK实施CARPE方法;有许多试验可用来评估基因产物的活性。使用大肠杆菌BW23115亲代菌株和pSIM5载体进行实验(Datta等人.Gene(2008)379:109-115)以介导重组工程化。将编码Cas9的基因克隆至pBTBX-2骨架中,

处于pBAD启动子的控制下,以允许通过向培养基中添加阿拉伯糖来控制Cas9裂解活性。

[0216] 首先,测试选择性地并入合成dsDNA盒(127bp)的能力。该合成dsDNA盒来源于NNR文库,该NNR文库由简并引物或由作为27,000个成员的文库的一部分合成的理性设计的寡核苷酸经由微阵列技术构建。在这两种情况下,将寡核苷酸设计为使galK基因产物的活性位点残基发生突变,以及含有合成引发位点P1(SEQ ID NO:1)。基于使用针对galK基因座的引物通过菌落PCR获得的扩增子大小的变化,验证了供体菌株文库的高度有效的恢复。对来自NNR文库的菌落PCR产物的测序表明,来自dsDNA盒的合成引发位点(P1)以约90-100%的效率并入(图2)。这一惊人且出乎意料的结果表明,可以在不依靠通常在其他基于重组工程化的编辑方法中使用的易错mutS缺乏的菌株(Costantino等人PNAS(2003)100:15748-15753;Wang等人Nature(2009)460:894-898)的情况下,以高效率生成文库。然而,密码子突变的效率有所降低(约20%),这可能是由于在等位基因替换过程中的mutS校正。在此项工作中,当构建的两个阶段都在mutS+背景下进行时,最终密码子编辑效率约为10%。

[0217] 为了提高CARPE方法的最终编辑效率和通用性,可以在转移到mutS+供体菌株之前在mutS缺乏的菌株中进行供体构建,以试图防止在供体构建阶段中突变的损失。

[0218] 实施例2:使用CARPE方法靶向必需基因

[0219] 为了测试CARPE方法的通用性,如上所述对包括dxs、metA和folA在内的多种必需基因使用该方法。可以使用gRNA设计策略靶向必需基因(图3)。

[0220] 来自靶向dxs基因的CARPE实验的数据也表明,尽管在供体文库创建过程中发生基因破坏,但在重组工程化后1-3小时内有效构建并重获供体文库是可能的。

[0221] 实施例3:使用CARPE方法调整异戊烯醇的生产

[0222] 经由细菌生产寻求用于工业生产的更好的生物燃料,需要进行现有技术基因组设计、工程化和筛选所需产物的能力。先前我们证明了单独改变大肠杆菌基因组中每个基因的表达水平的能力(Warner等人Nat.Biotechnol(2010)28:856-862)。该方法(被称为可追踪多路重组工程化(TRMR))产生了约8000个基因组修饰的细胞的文库(约4000个过表达的基因和约4000个敲减的基因)。随后在不同条件下筛查该文库,这使得能够更深入地了解基因产物的活性并在这些选择下产生表现更好的菌株。TRMR允许改变两个水平的蛋白质表达(过表达的和敲减的)但不能实现开放阅读框(ORF)的修饰。在此,我们的目的是产生ORF修饰的大文库,并对整个代谢途径进行工程化以最优化生物燃料的生产。

[0223] 产生这样的理性设计(相比于随机诱变)的文库的主要困难是将所需突变插入靶细胞中的效率。重组工程化——在大肠杆菌中进行基因组修饰的典型方法——使用来自λ噬菌体的重组基因来促进外来DNA向宿主基因组中的插入。然而,该方法具有低效率,并且可通过添加抗生素抗性基因随后选择(如在TRMR中)或通过递归诱导重组事件(即,通过MAGE(Wang等人Nature(2008)460:894-898))来克服。本文所述的CARPE方法提高了重组工程化效率,包括使用CRISPR系统从群体中去除所有非重组细胞。CRISPR是近来发现的细菌和古细菌针对入侵噬菌体和质粒的基于RNA的适应性防御机制(Bhaya等人Ann.Rev.of Genetics(2011)45:273-297)。该系统经历巨大工程化,以使得能够使用两个质粒来允许序列引导的双链断裂,一个质粒编码CRISPR相关的核酸酶Cas9,而第二个质粒编码序列特异性指导RNA(gRNA),gRNA将Cas9引导至其独特的位置(Qi等人Cell(2013)45:273-297)。CARPE方法利用CRISPR系统以序列依赖性方式诱导DNA断裂并因此导致细胞死亡的能力。我

们产生了DNA重组工程化盒,它除了ORF内的所需突变以外还包含在被CRISPR机制所针对的基因的开放阅读框外部的常见位置中的突变。这种将所需突变与由于PAM突变/缺失而避免CRISPR介导的死亡相连接/偶联的方法使得能够显著富集在全部细胞群体内的工程化细胞。

[0224] 使用DXS途径进一步阐释所述方法。DSX途径导致产生异戊烯焦磷酸(IPP),IPP导致萜类和类萜化合物的生物合成。有趣的是,如果添加所需基因,IPP还可以是番茄红素或异戊烯醇的前体。番茄红素使得细菌菌落呈红色,并因此可容易地筛选,而异戊烯醇被认为是具有比乙醇更高的能量密度和更低的水混溶性的'第二代'生物燃料。选择以下三种蛋白质进行工程化:1)DSX,所述途径的第一个限速酶,2)IspB,其将代谢流从DXS途径转移,和3)NudF,已证明其在大肠杆菌和枯草芽孢杆菌(B.subtilis)中均将IPP转化为异戊烯醇(Withers等人App.Environ.Microbiol(2007)73:6277-6283;Zheng等人Biotechnol.for biofuels(2013)6:57))。将采用针对菌落颜色量化开发的新的图像分析工具针对增加的番茄红素产量来筛选编码DXS和IspB的基因的突变。NudF活性通过经GC/MS测量异戊烯醇水平而直接测定,并且通过将会用作生物传感器的异戊烯醇营养缺陷型细胞而间接测定。该方法提供了以高准确度和效率将大的突变文库理性工程化至大肠杆菌基因组中的能力,以及产生高产率异戊烯醇的菌株。

[0225] 实施例4:使用GEn-TraCER方法编辑galK

[0226] 使用GEn-TraCER方法编辑galK基因,galK基因已用作在大肠杆菌中重组工程化的模式系统(Yu等人2000)。构建的第一GEn-TraCER盒被设计为在galK的密码子24处引入终止密码子代替框内PAM,被称为galK_Q24(图12)。使用自定义的python脚本设计构建体和载体以高通量地生成必要突变。

[0227] 使用环形聚合酶克隆(CPEC)法将对照盒克隆至由Qi等人Cell(2013)描述的gRNA载体中。采用以下引物使骨架线性化:CCAGAAATCATCCTTAGCGAAAGCTAAGGAT(SEQ ID NO:29)和GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCT(SEQ ID NO:30)。

[0228] GenTRACER盒作为gblock定购,并使用以下引物进行扩增:ATCACGAGGCAGAATTTCAGATAAAAAAAAATCCTTAGCTTTCGCTAAGGATGATTTCTGG(SEQ ID NO:31),ACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC(SEQ ID NO:32)。

[0229] 使用CPEC将组件缝合在一起,并将组件转化至大肠杆菌中以生成载体。使用合并的寡核苷酸文库以多路的方式执行该程序,其中克隆效率为10^4-10^5CFU/μg的数量级。

[0230] 在30℃下,在含有50μg/mL卡那霉素和34μg/mL氯霉素的LB中,使携带pSIM5(λ-RED质粒)和X2-cas9质粒的大肠杆菌MG1655细胞生长至对数中期(0.4-0.7OD)。在42℃下诱导pSIM5载体的重组工程化功能15min,随后在冰上放置10min。随后通过沉淀并用10mL冷却H_2O洗涤2次而使细胞成为电感受态细胞。将细胞用100ng的GEn-TraCER质粒(也编码羧苄青霉素抗性)转化并在37℃下恢复3hr。将50-100μL的细胞接种到含有50μg/mL卡那霉素和100μg/ml羧苄青霉素的合适的培养基中,以选择性地富集CRISPR编辑的菌株。在补充有半乳糖的MacConkey琼脂上使用红色/白色筛选,计算galK基因的编辑效率。

[0231] 基于MacConkey琼脂上的筛选,观察到galK_Q24*设计有约100%的编辑效率。有趣的是,不同于需要错配修复敲除来达到高效率的寡核苷酸介导的重组工程化方法(Li等人2003;Sawitzke等人2011;Wang等人2011),在错配修复机制完整或不完整的情况下在菌株

中没有影响。

[0232] 随后通过Sanger测序验证染色体和载体序列。

[0233] 如所预期的,载体中设计的突变反映在染色体上(图13),这表明突变存在于两个位置中并且该质粒充当反式作用条形码(反式条形码)或基因组编辑的记录。

[0234] 所述设计适合于通过生成由同义PAM突变(图14B,ΔPAM)组成的“沉默可选择疤痕”,在基因组规模上理性诱变蛋白质编码框,以针对Cas9介导的裂解“免疫”该细胞,而使翻译产物未受到干扰。我们推断沉默疤痕可以允许以高效率共选择在密码子附近的编辑或其他感兴趣的特征。评估同源性臂长和galK中PAM突变/缺失与所需突变之间的距离的影响,并比较其效率(图16B)。当用相同的PAM编辑将同源性臂长从80个核苷酸延伸至100个核苷酸时,观察到在galK位置145处的突变效率显著增加(分别为约5%和45%)。

[0235] 实施例5:使用GEn-TraCER方法重建突变

[0236] 使用自定义的自动化设计软件将GEn-TraCER方法扩展至基因组规模,该软件允许用简单的用户输入定义来靶向基因组周围的位点。通过重建来自近来报道的大肠杆菌热适应研究(Tenaillon等人2012)的全部非同义点突变来测试所述方法。该研究表征了在来自独立繁殖的菌株的115个隔离株中发生的全套突变。该数据集提供了突变的多样化来源,该来源的单独适合性影响进一步阐明了这一复杂表型的机械基础。在密码子使用和ΔPAM中,在可能的情况下,这些突变中的每一个均以2倍冗余度重建,以便在下游适合性分析中实现PAM和靶密码子突变的统计校正。

[0237] 实施例6:使用GEn-TraCER方法调整遗传相互作用

[0238] 通过在大肠杆菌基因组中每个基因的上游整合被环境因素(氧水平、碳源、应力)动态调节的启动子来生成启动子重置(rewiring)文库。使用GEn-TraCER方法生成菌株,该菌株具有可能有利于例如对用于生产的感兴趣化学品的耐受性的重置的基因型。

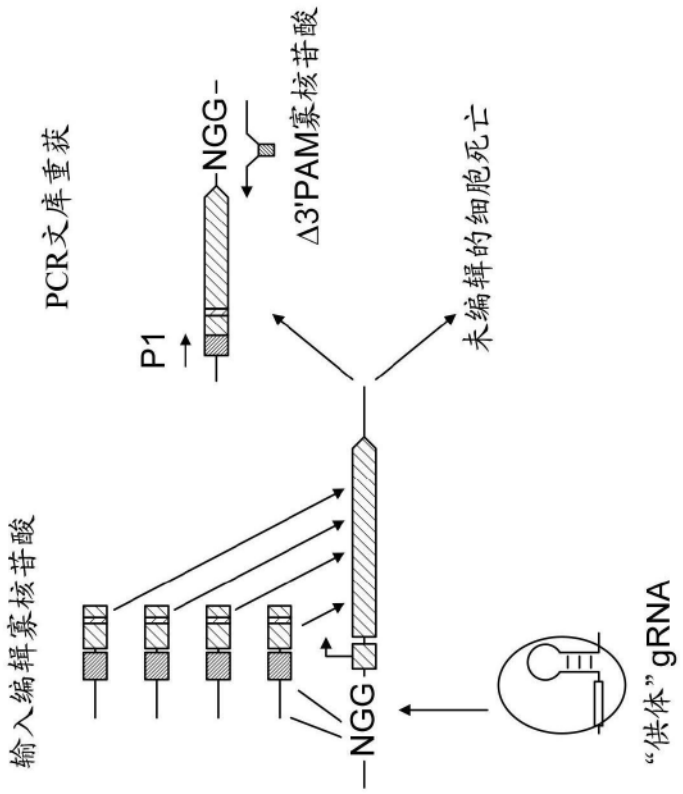

输入编辑寡核苷酸
PCR文库重获
P1
NGG
Δ3'PAM寡核苷酸
未编辑的细胞死亡
NGG
"供体"gRNA

图1A

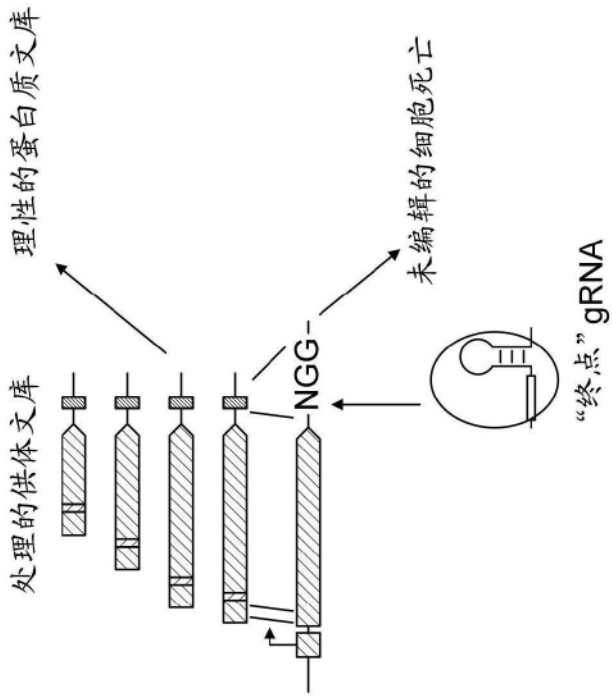

处理的供体文库
理性的蛋白质文库
NGG
未编辑的细胞死亡
"终点"gRNA

图1B

克隆
P1
密码子
F_F01 CCGTGGATCCTAGGCTGGTCTC ... GCCTGGC TAA GTGAATT SEQ ID NO:2
F_F02 CCGTGGATCCTAGGCTGGTCTC ... AGCAAAA ACA GGTATTA SEQ ID NO:3
F_F03 CCGTGGATCCTAGGCTGGTCTC ... AAACAGG TAT TAAAGAG SEQ ID NO:4
F_F04 CCGTGGATCCTAGGCTGGTCTC ... GCCTGGC CGC GTGAATT SEQ ID NO:5
差的读取质量
F_F05
F_F06 CCGTGGATCCTAGGCTGGTCTC ... GCCTGGC CGC GTGAATT SEQ ID NO:5
F_F07 CCGTGGATCCTAGGCTGGTCTC ... CCAGTTT CAA GGCTGTA SEQ ID NO:6
F_F08 CCGTGGATCCTAGGCTGGTCTC ... CCAGTTT GTA GGCTGTA SEQ ID NO:7
差的读取质量
F_F09
F_F10 CCGTGGATCCTAGGCTGGTCTC ... CCAGTTT GTA GGCTGTA SEQ ID NO:7
F_G01 CCGTGGATCCTAGGCTGGTCTC ... CTTCAAA CGT ACCCTGG SEQ ID NO:8
F_G02 CCGTGGATCCTAGGCTGGTCTC ... CCAGTTT GTA GGCTGTA SEQ ID NO:7
F_G03 CCGTGGATCCTAGGCTGGTCTC ... GCCTGGC CGC GTGAATT SEQ ID NO:5
F_G04 CCGTGGATCCTAGGCTGGTCTC ... CCAGTTT GTA GGCTGTA SEQ ID NO:7
F_G05 CCGTGGATCCTAGGCTGGTCTC ... CTTCAAA CGT ACCCTGG SEQ ID NO:8
F_G06 CCGTGGATCCTAGGCTGGTCTC ... GCCTGGC CGC GTGAATT SEQ ID NO:5
F_G07 CCGTGGATCCTAGGCTGGTCTC ... AGCCAAA AAT AGGTATT SEQ ID NO:9
F_G08 CCGTGGATCCTAGGCTGGTCTC ... GCCTGGC CGC GTGAATT SEQ ID NO:10
F_G09 CCGTGGATCCTAGGCTGGTCTC ... CTTCAAA AGG ACCCTGG SEQ ID NO:11
F_G10 CCGTGGATCCTAGGCTGGTCTC ... CTTCAAA ACA ACCCTGG SEQ ID NO:12

图2

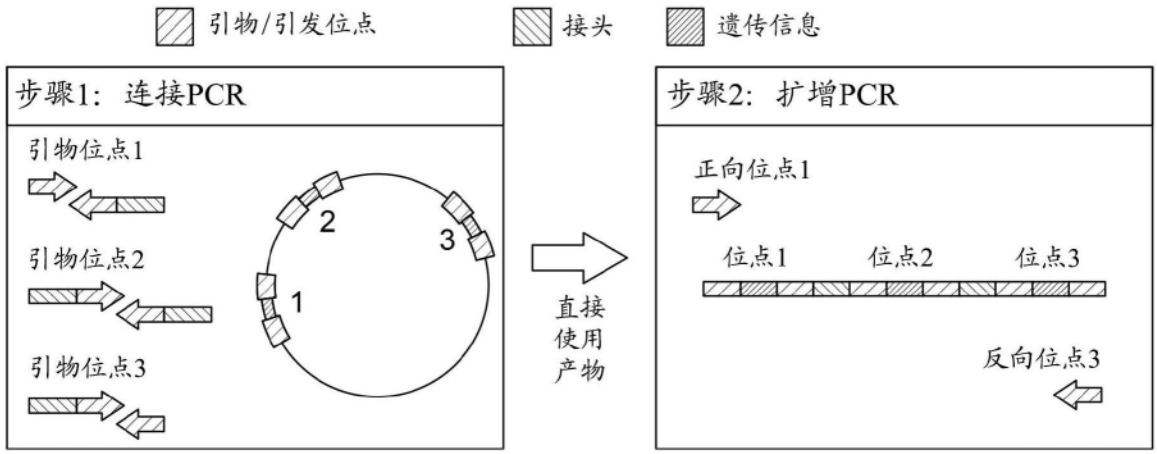

引物/引发位点
接头
遗传信息
步骤1：连接PCR
引物位点1
引物位点2
引物位点3
1
2
3
直接
使用
产物
步骤2：扩增PCR
正向位点1
位点1
位点2
位点3
反向位点3

图3A

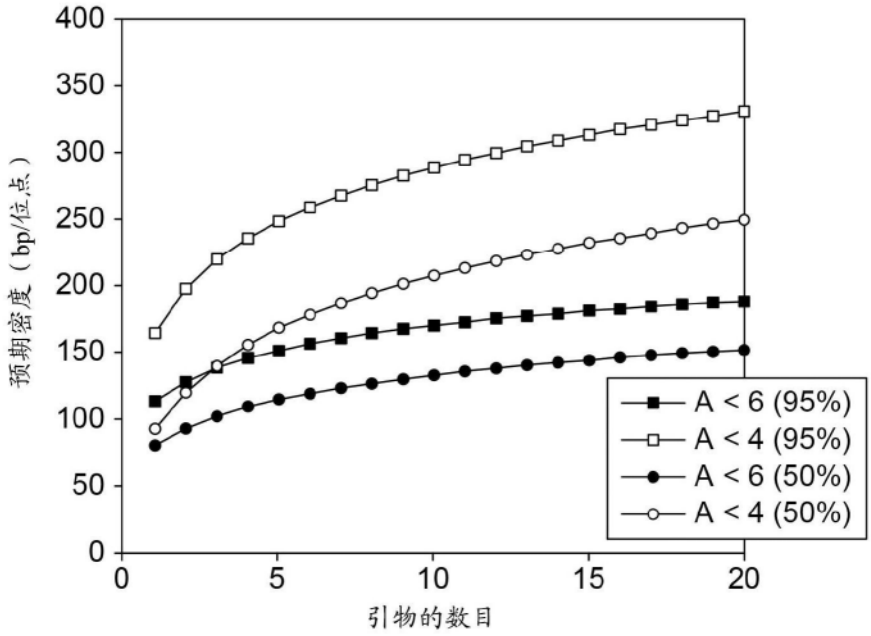

预期密度（bp/位点）
引物的数目
A < 6 (95%)
A < 4 (95%)
A < 6 (50%)
A < 4 (50%)
0
50
100
150
200
250
300
350
400
0
5
10
15
20

图3B

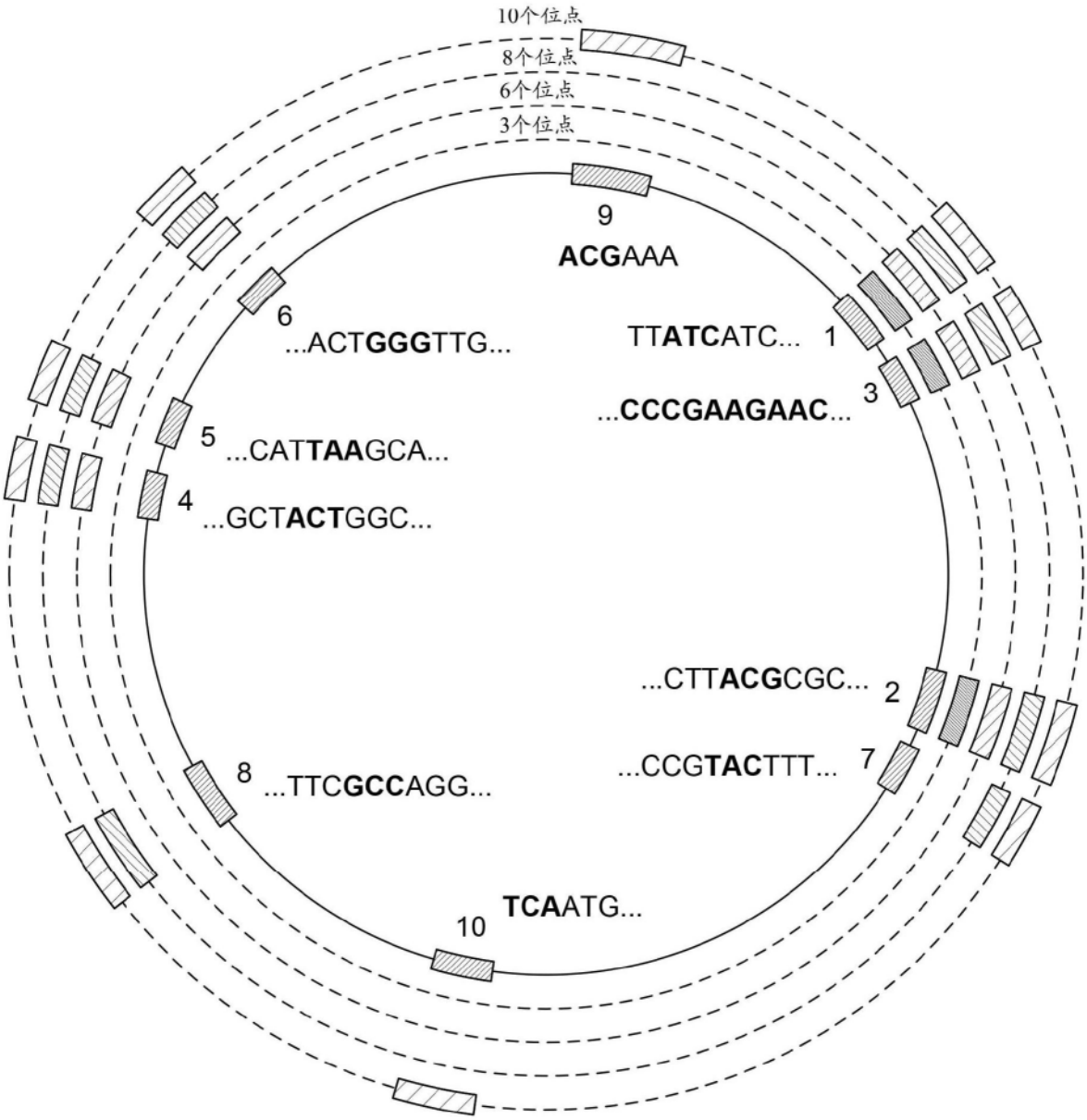

10个位点
8个位点
6个位点
3个位点
9
ACGAAA
6
...ACTGGGTTG...
TTATCATC... 1
3
...CCCGAAGAAC...
5
...CATTAAGCA...
4
...GCTACTGGC...
...CTTACGCGC... 2
...CCGTACTTT... 7
8
...TTCGCCAGG...
TCAATG...
10

图4A

10个位点

CGGAACCGGTATTGCAGCAGCTTTATCATCTGCCGCTGGACGGCGCACAAATCGC

GCTTAACGGTCAGGAAGAACCTCGTGAATCGCATCTCCGCAACGCCAATGACACT

CCGCCAGCAGAACGCGGCGTTGGTATGGTGTTCAGTCTTACGCGCTCTATCCCCA

CCTGTCAGTAGCAGAAACATGTCATTTGGCCTGAAACTGGCAAGCACGCCCGGT

AGCAGAGCCCGAGTATTACATCGAACTGGATTTCAACACAGCGGTAAGATCCTTGAG

AGTTTTCGCCCCCGAAGAACGTTTCTCAATGATGAGCACTTTTAAAGTTNTGCTATG

TGGCGCGGACGGTTTGGGCGAGGATGTTCGGGACTGAGCGTATGGAAGAGCACTT

GCGTTTTGCCGCCTGCTACTTGGCACACCTTCTGCTGGAACGGGGCGGATATGTTTT

TCAGATTGGGTTGCGACTGGTGGGCGACGGTTTTATGATTGCAGGTCCGCTGGGC

GGCTGGTGCATTTAAGCACTTCGACCGCTGGGTAGACGGTAAGATCAAATGCCGAT

TGTCAGTTTGGAGGGAGCAAGGGAACCAGATCAAACCAGAGCACTTCAAAAACTG

GGTTGAATGGGCGAAAGCCAATCANCTCGGTCTGGATCCCAACCATGCGTATCGTG

CCAAGTGCGTCCCAGACAACAAGTATATGAATTATCGGAGGTTGTCGATCAACTCG

ATAACCCGTACTTTGTTATGGTTACGTACCGATTTTCGAGGTGAATTATTTATTGG

CAGCCCCCGTTGTGGTGTCCTTCTGATGANGGGATTACAGGCACGCTTGAGCCAG

GACCTGGCGCGCGAGAAATTCCGCCAGGCGCAGGATGGTCACTTACCGACTTGCAA

CTTTANCCCGCTGTTCCCGAGGACGTTAACGCGCTGGTCGATGAGTTACAAAAGCTG

CTACACCATGACGCCTTGCATAGGAGCCACGAACCGCCATACNAGACATTTTGAG

GCATTTCAGTCAGTTGCTCAATGTACCTATACCCAGACCGTTCAGCTGGATATTA

SEQ ID NO:13

图4B

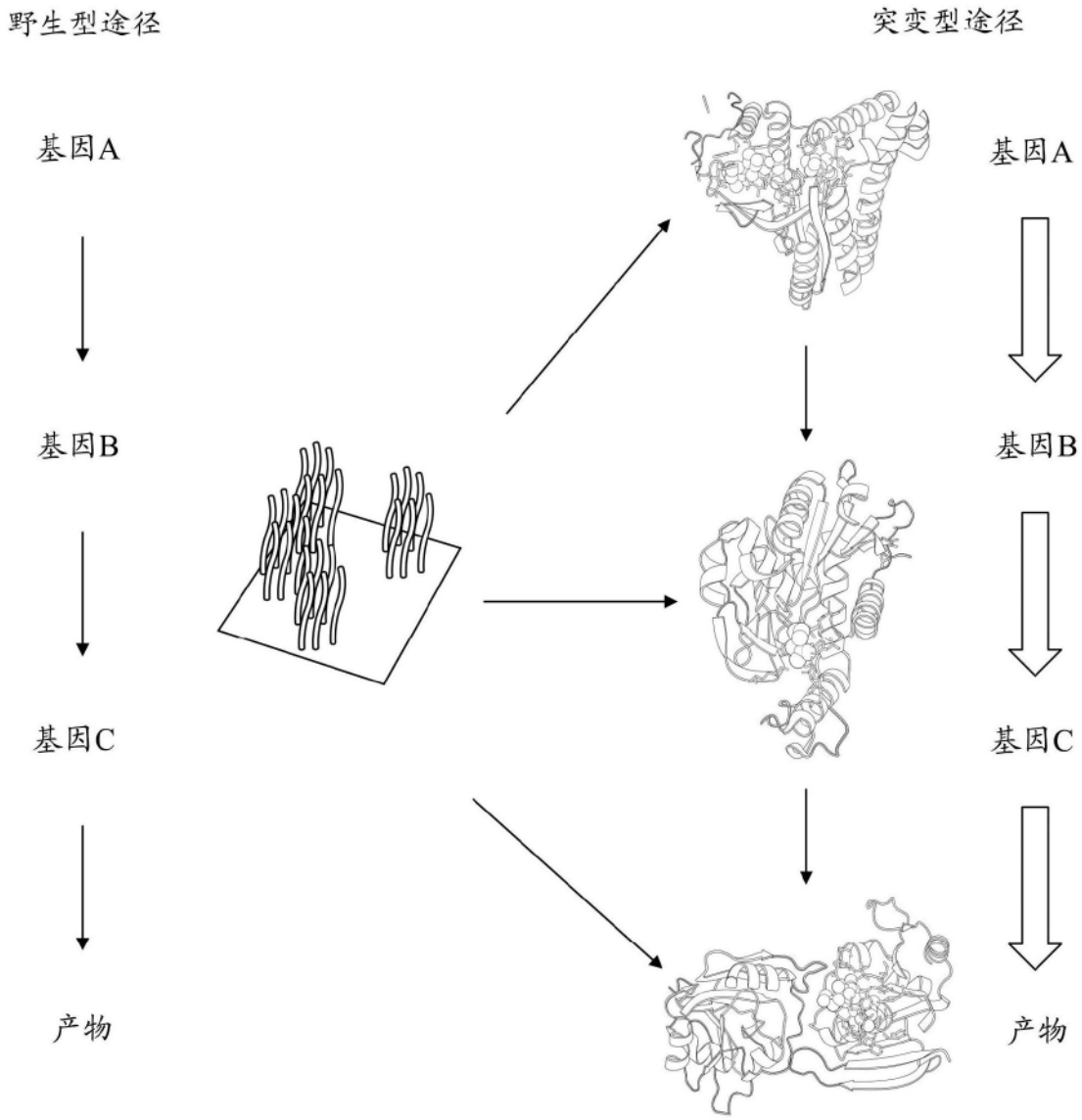


图5

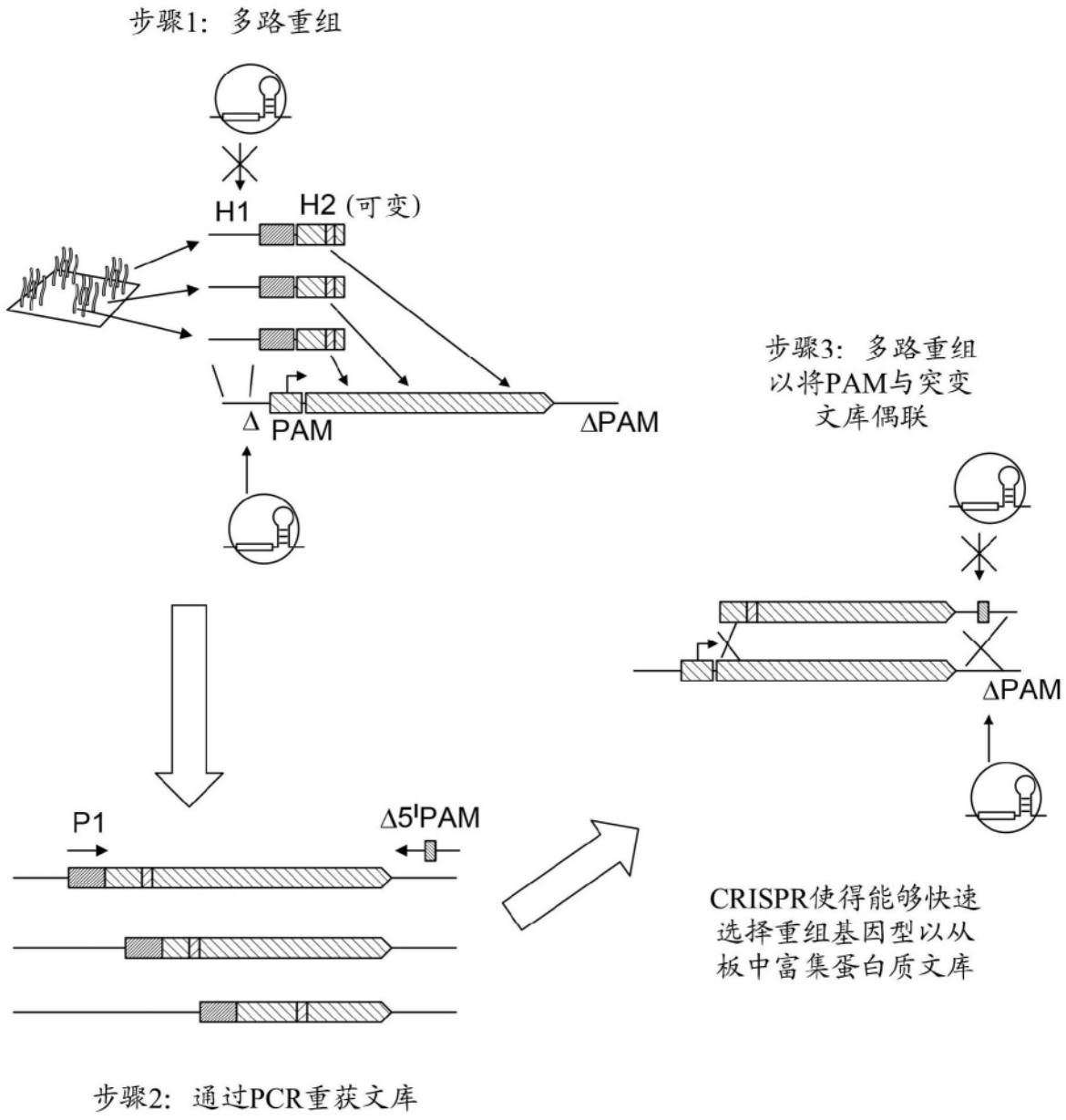

步骤1：多路重组
H1
H2 (可变)
Δ PAM
ΔPAM
步骤3：多路重组
以将PAM与突变
文库偶联
ΔPAM
P1
Δ5'PAM
CRISPR使得能够快速
选择重组基因型以从
板中富集蛋白质文库
步骤2：通过PCR重获文库

图6

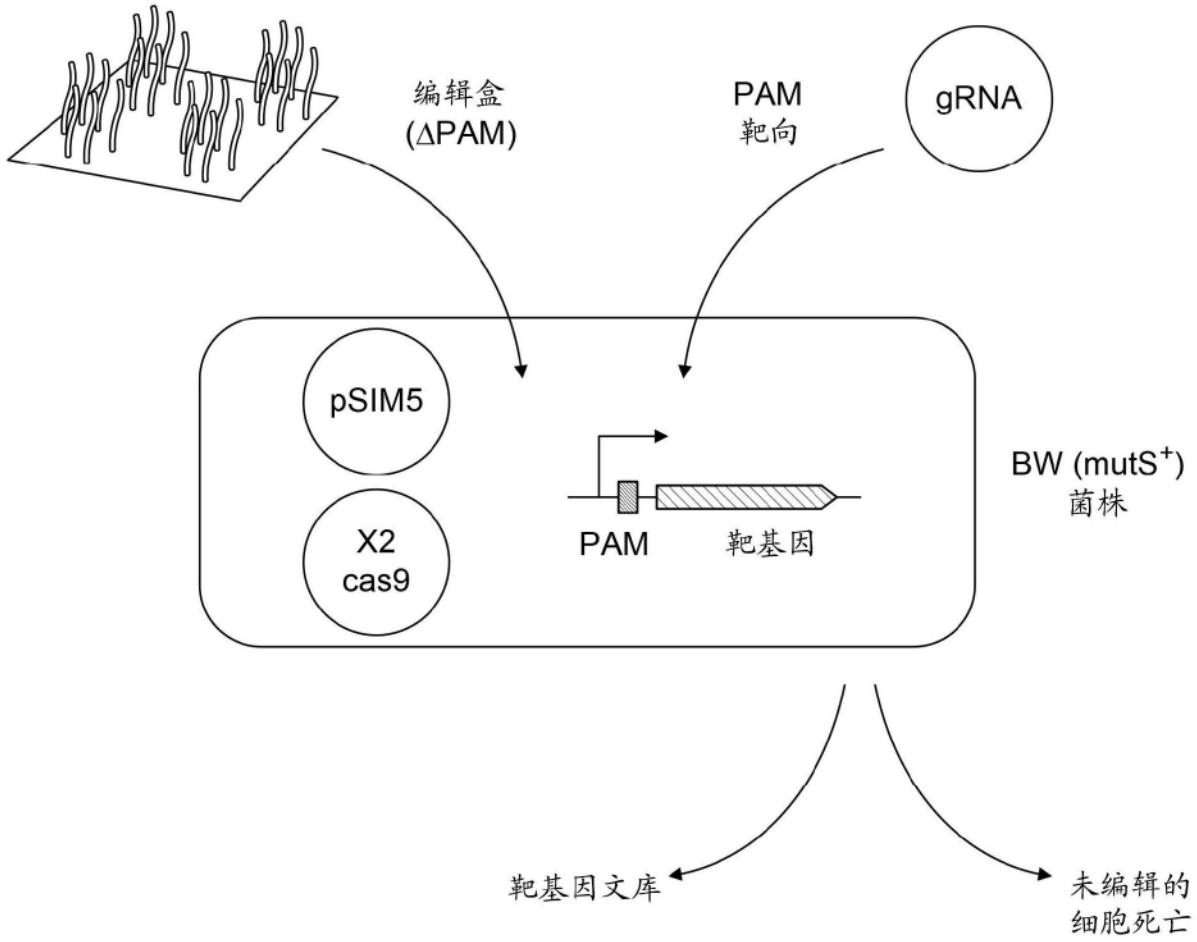

编辑盒
(ΔPAM)
PAM
靶向
gRNA
pSIM5
X2
cas9
PAM
靶基因
BW (mutS+)
菌株
靶基因文库
未编辑的
细胞死亡

图7

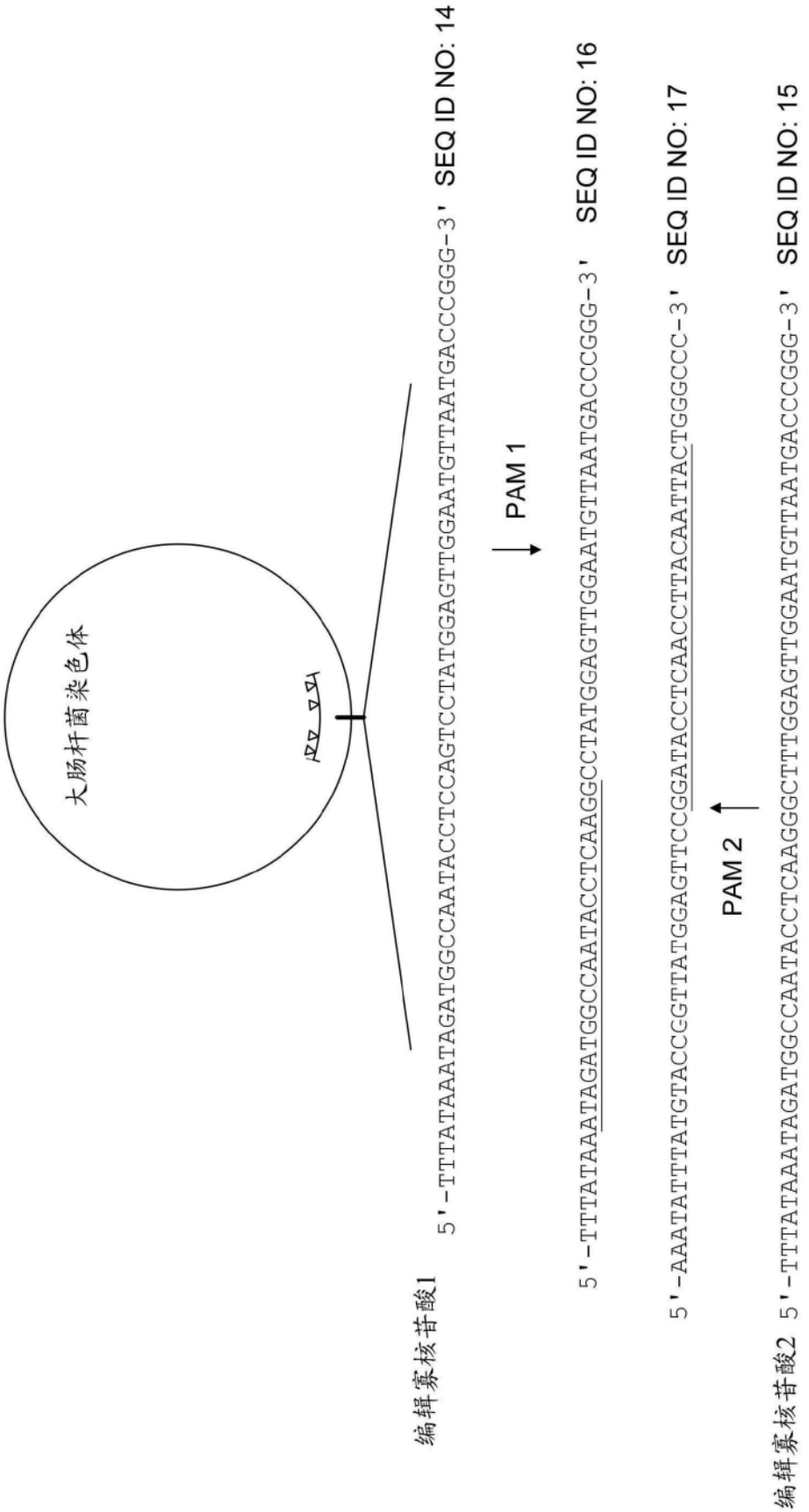

大肠杆菌染色体
靶基因
编辑寡核苷酸1
5'-TTTATAAATAGATGGCCAATACCTCCAGTCCTATGGAGTTGGAATGTTAATGACCCGGG-3' SEQ ID NO: 14
PAM 1
5'-TTTATAAATAGATGGCCAATACCTCAAGGCCTATGGAGTTGGAATGTTAATGACCCGGG-3' SEQ ID NO: 16
5'-AAATATTTATGTACCGGTTATGGAGTTCCGGATACCTCAACCTTACAATTACTGGGCCC-3' SEQ ID NO: 17
PAM 2
编辑寡核苷酸2 5'-TTTATAAATAGATGGCCAATACCTCAAGGGCTTTGGAGTTGGAATGTTAATGACCCGGG-3' SEQ ID NO: 15

图8

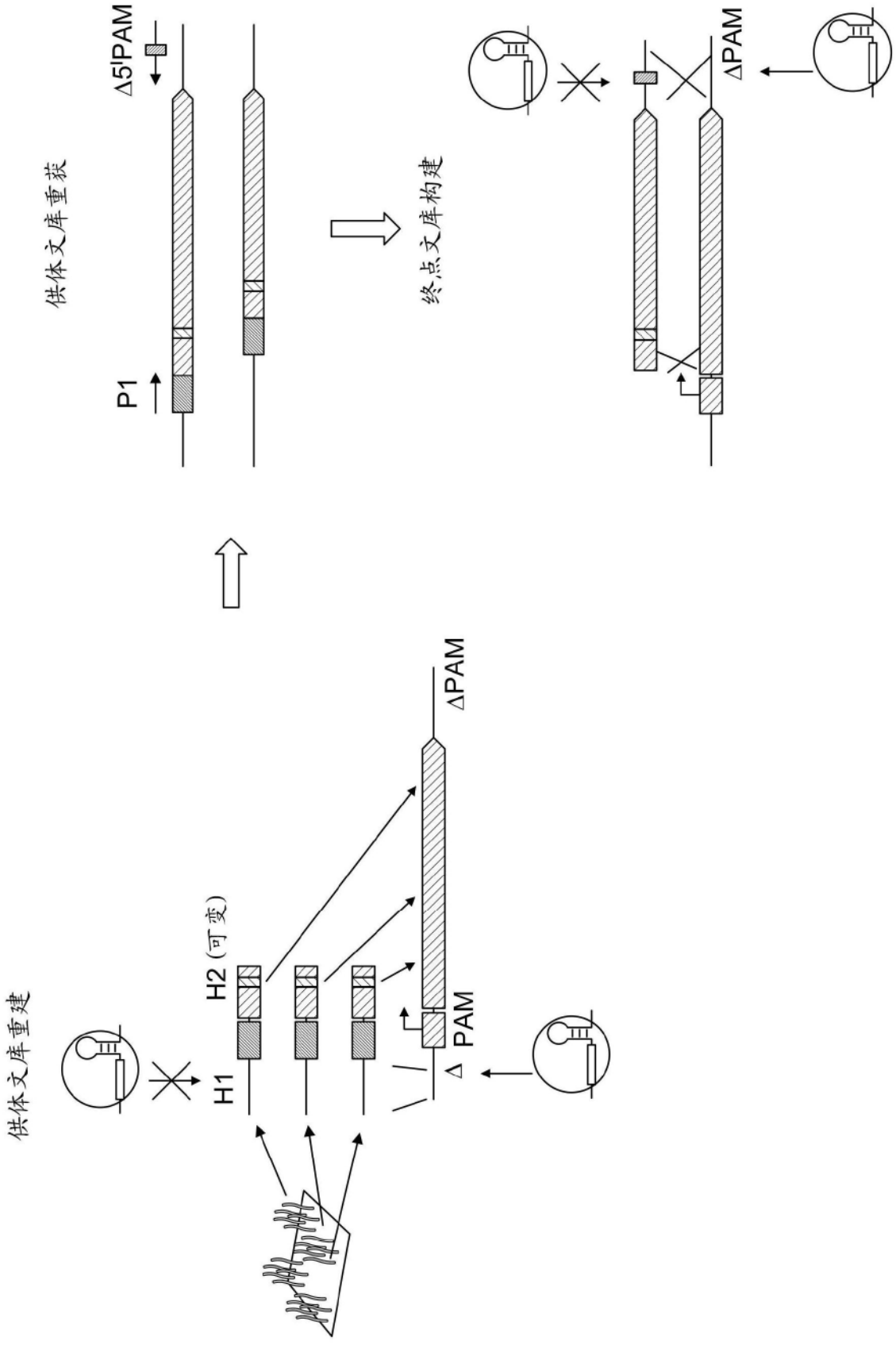


图9

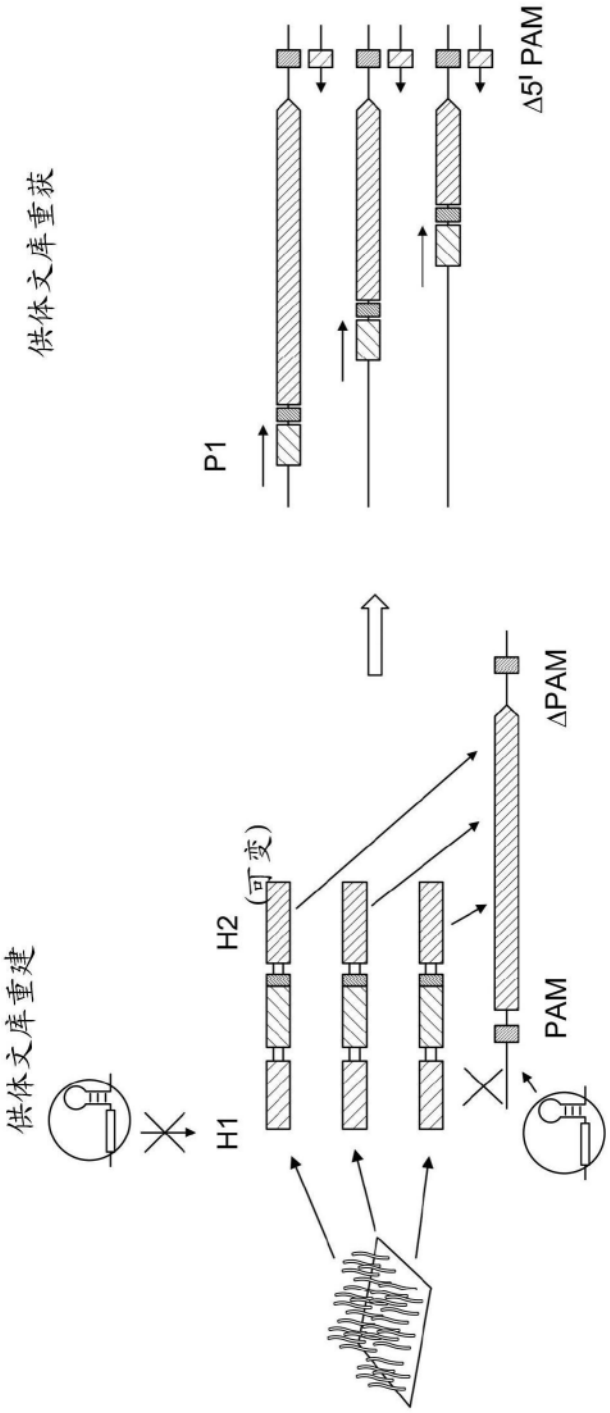


图10A

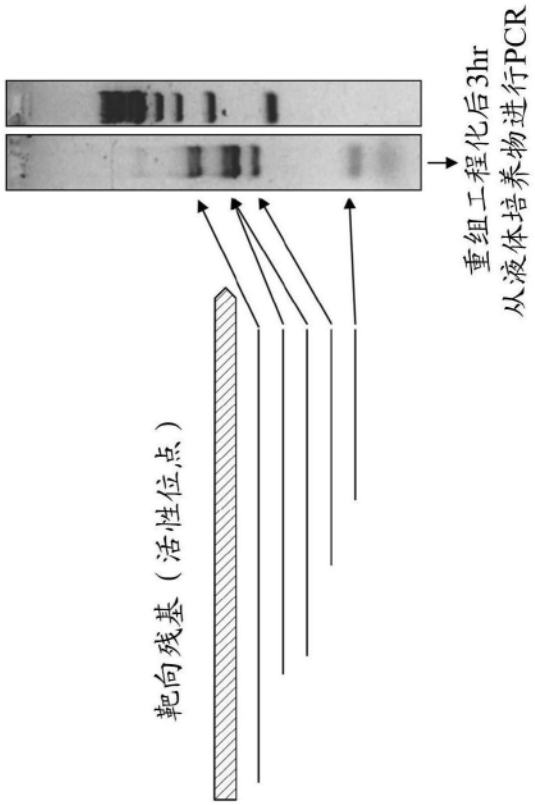


图10B

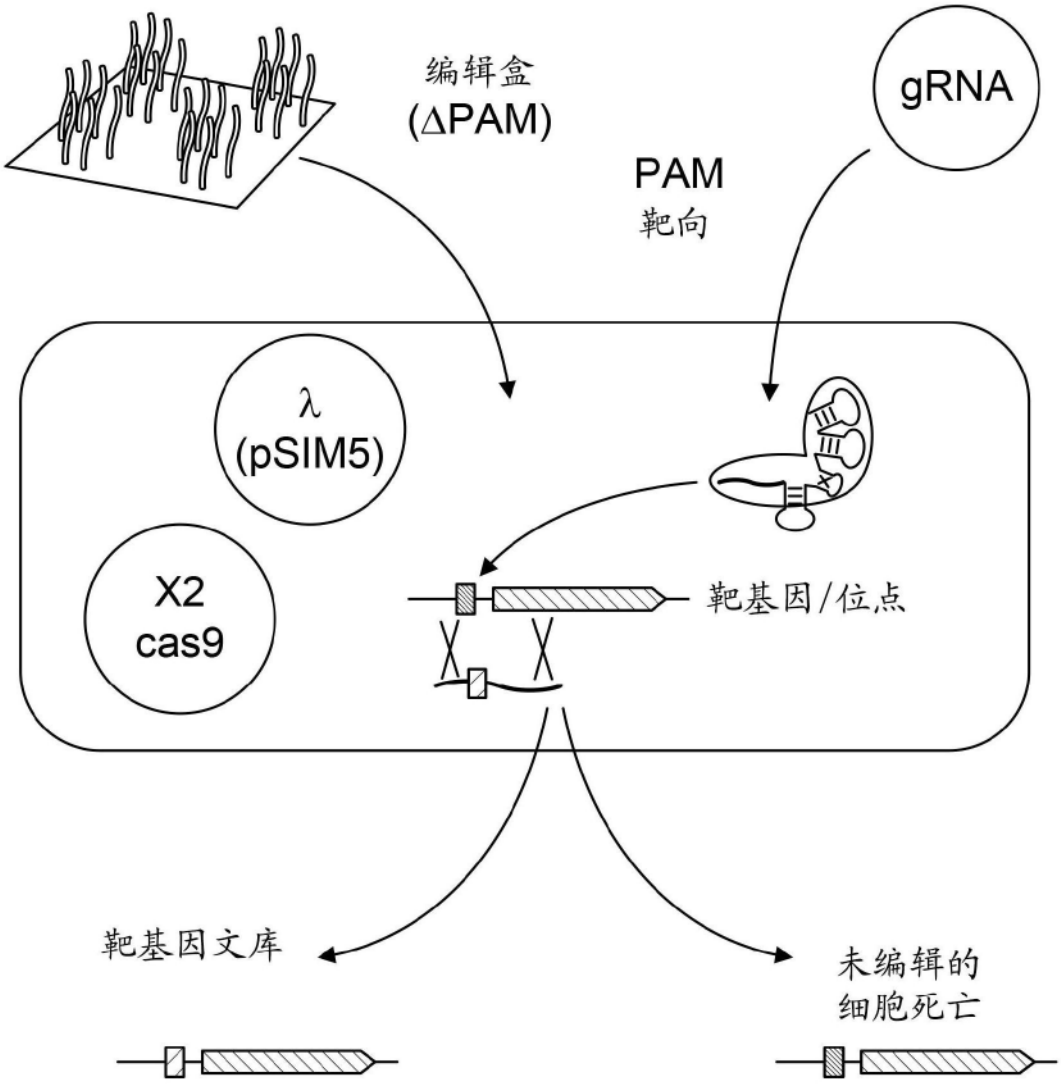


图11A

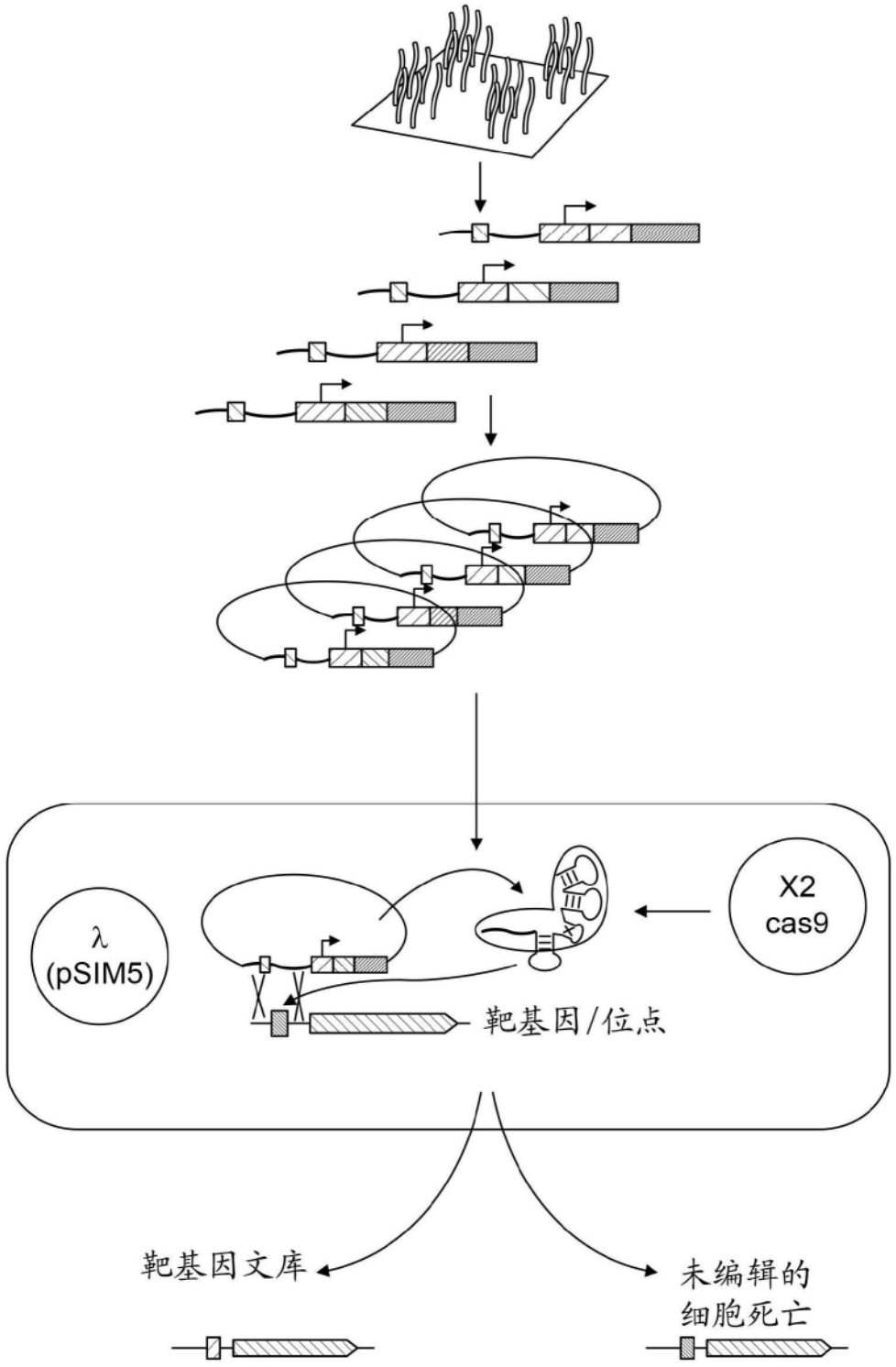


图11B

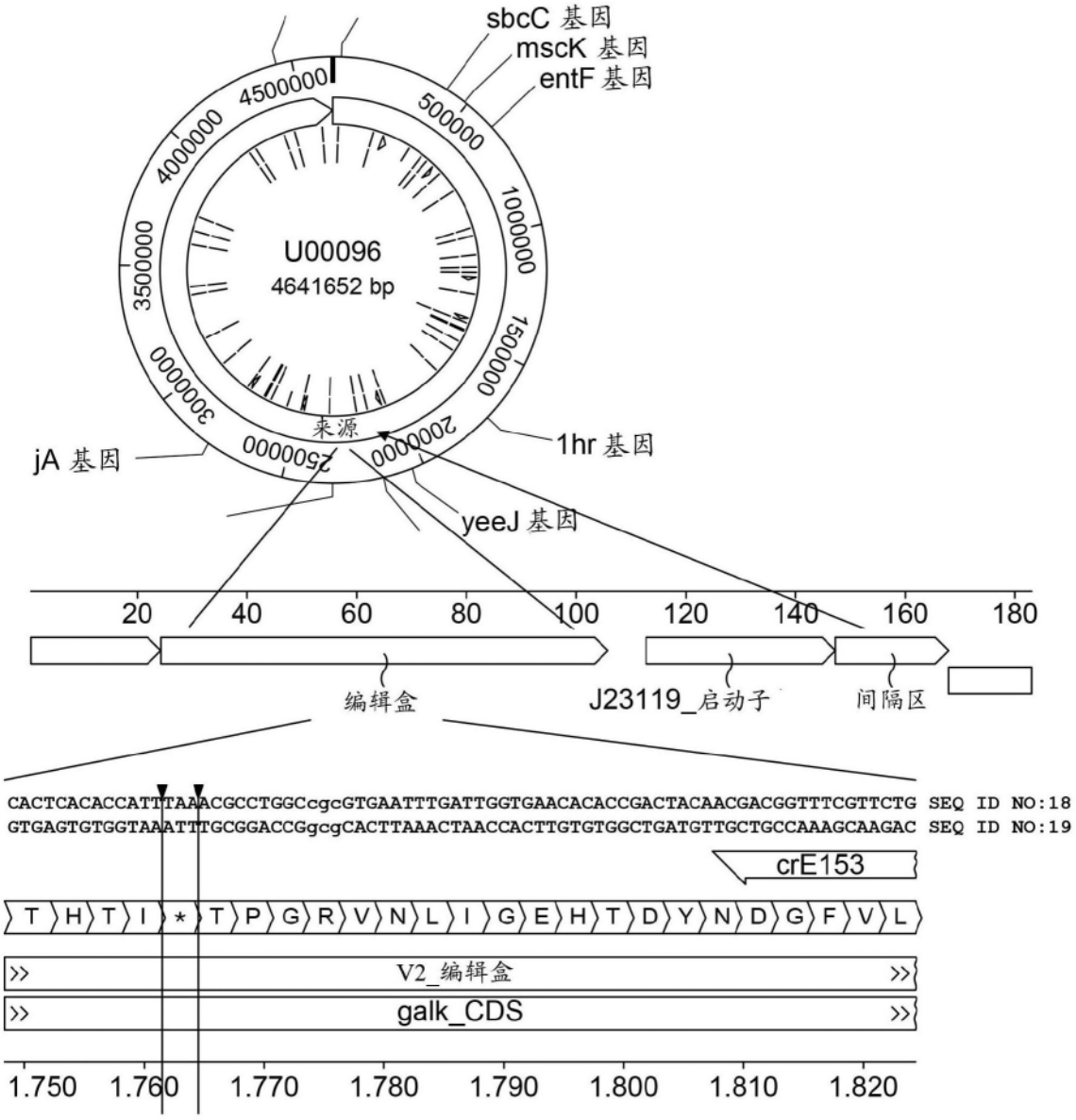

sbcC 基因
mscK 基因
entF 基因
4500000
500000
1000000
1500000
2000000
2500000
3000000
3500000
4000000
U00096
4641652 bp
来源
jA 基因
1hr 基因
yeeJ 基因
20
40
60
80
100
120
140
160
180
编辑盒
J23119_启动子
间隔区
CACTCACACCATTTAAACGCCTGGCcgcGTGAATTTGATTGGTGAACACACCGACTACAACGACGGTTTCGTTCTG SEQ ID NO:18
GTGAGTGTGGTAAATTTGCGGACCGgcgCACTTAAACTAACCACTTGTGTGGCTGATGTTGCTGCCAAAGCAAGAC SEQ ID NO:19
crE153
T H T I * T P G R V N L I G E H T D Y N D G F V L
V2_编辑盒
galk_CDS
1.750
1.760
1.770
1.780
1.790
1.800
1.810
1.820

图12

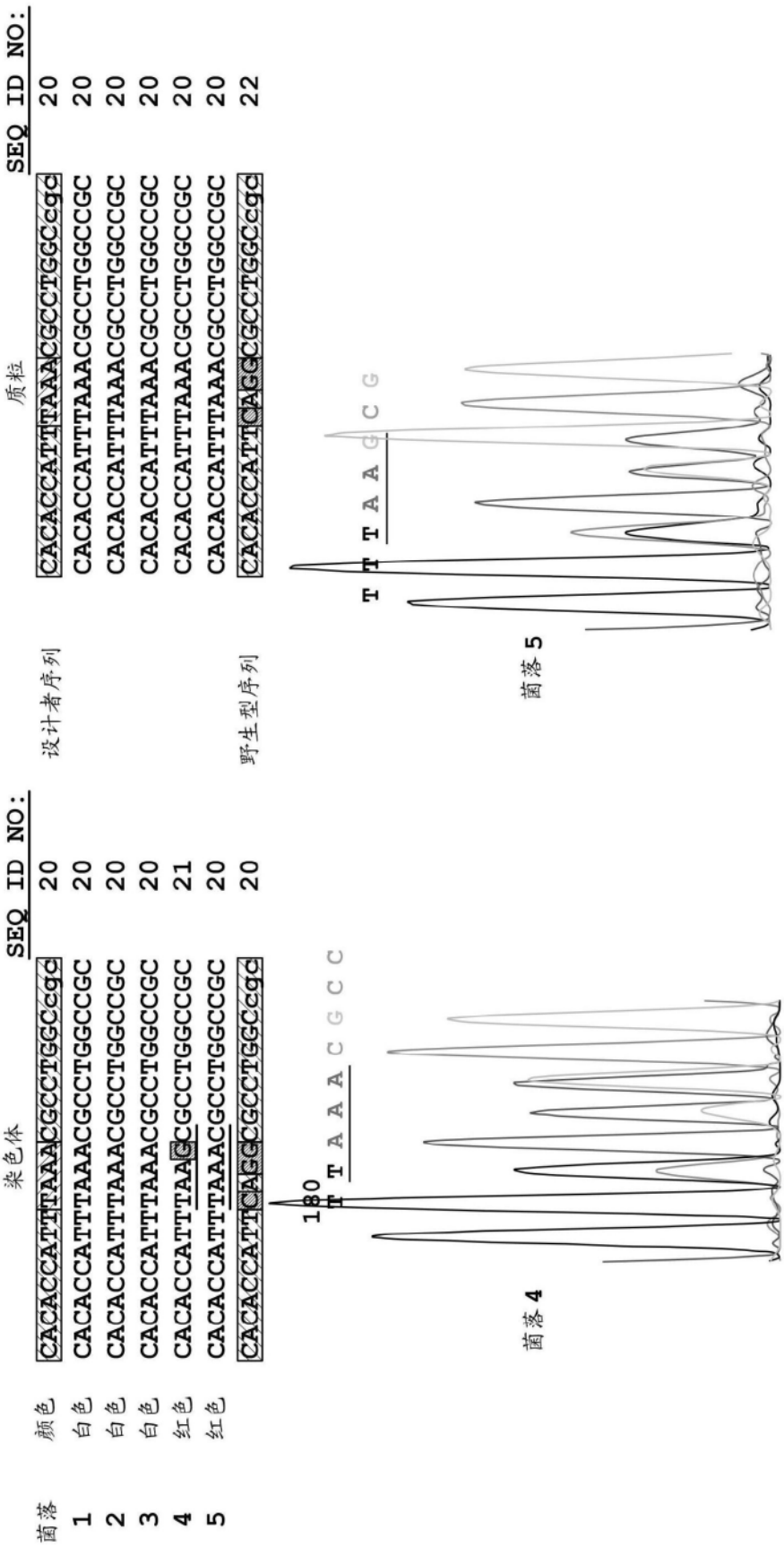

菌落 颜色 染色体 SEQ ID NO:
1 白色 CACACCATTTAAACGCCTGGCCGC 20
2 白色 CACACCATTTAAACGCCTGGCCGC 20
3 白色 CACACCATTTAAACGCCTGGCCGC 20
4 红色 CACACCATTTAAGCGCCTGGCCGC 21
5 红色 CACACCATTTAAACGCCTGGCCGC 20
CACACCATTCAGCGCCTGGCCgc 20
设计者序列
野生型序列
质粒 SEQ ID NO:
CACACCATTTAAACGCCTGGCCgc 20
CACACCATTTAAACGCCTGGCCGC 20
CACACCATTTAAACGCCTGGCCGC 20
CACACCATTTAAACGCCTGGCCGC 20
CACACCATTTAAACGCCTGGCCGC 20
CACACCATTTAAACGCCTGGCCGC 20
CACACCATTCAGCGCCTGGCCgc 22
180
T T A A A C G C C C
T T T A A G C C G
菌落 4
菌落 5

图13

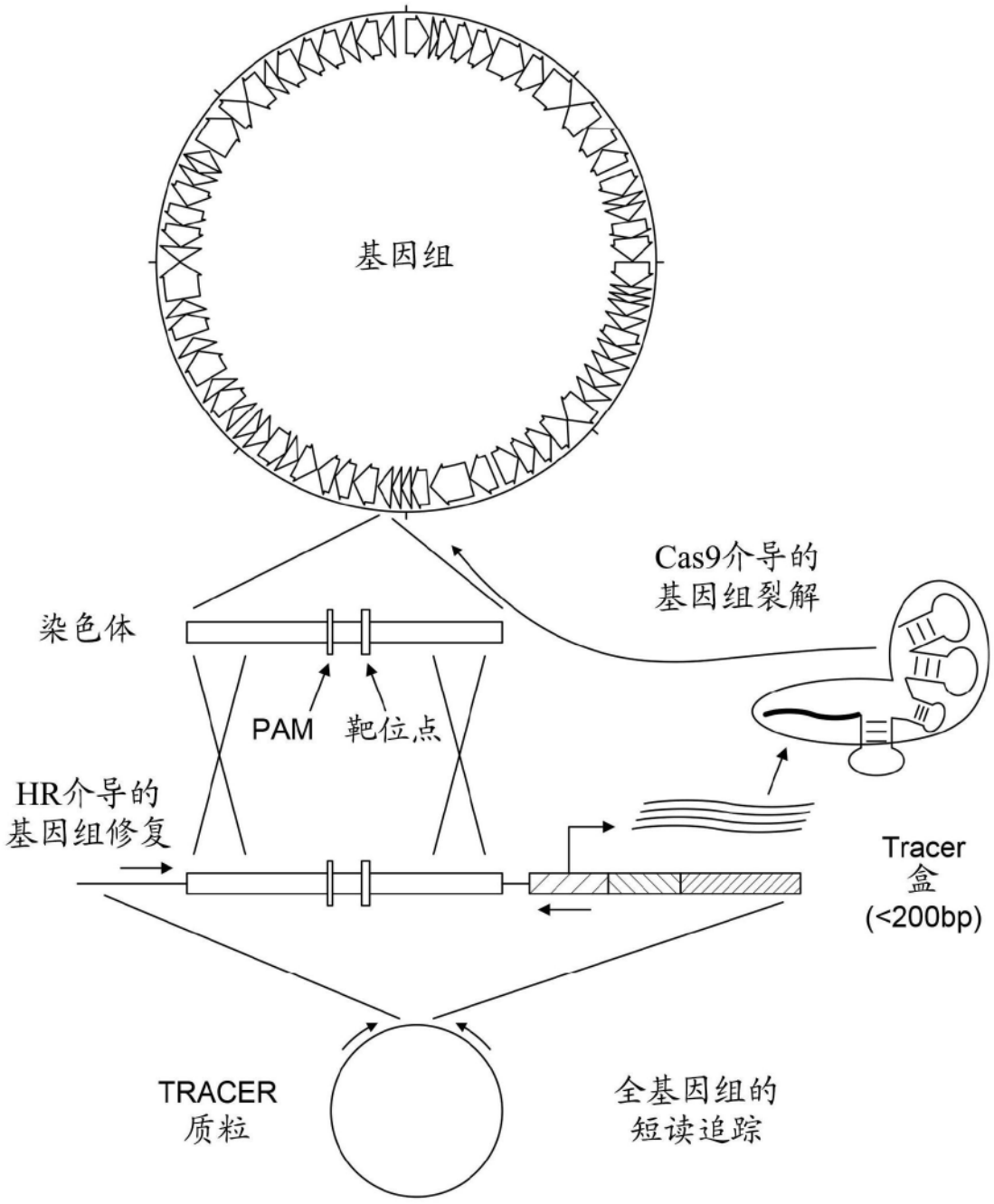

基因组
Cas9介导的
基因组裂解
染色体
PAM
靶位点
HR介导的
基因组修复
Tracer
盒
(<200bp)
TRACER
质粒
全基因组的
短读追踪

图14A

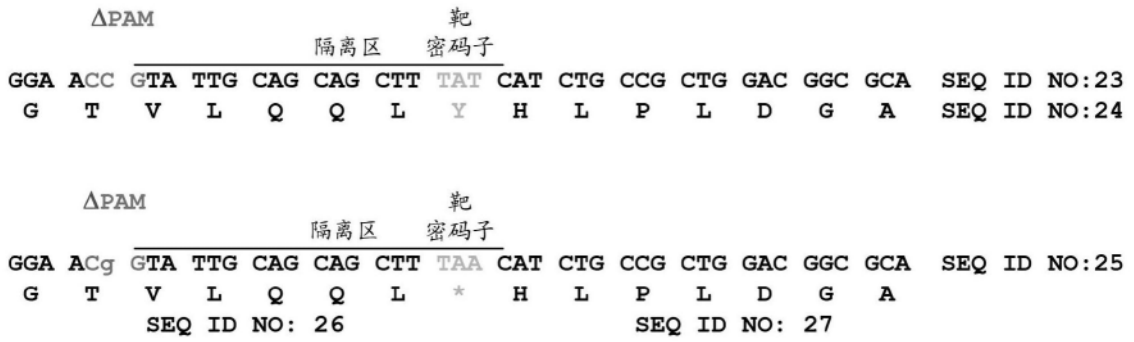

ΔPAM
靶
隔离区
密码子
GGA ACC GTA TTG CAG CAG CTT TAT CAT CTG CCG CTG GAC GGC GCA SEQ ID NO:23
G T V L Q Q L Y H L P L D G A SEQ ID NO:24
ΔPAM
靶
隔离区
密码子
GGA ACg GTA TTG CAG CAG CTT TAA CAT CTG CCG CTG GAC GGC GCA SEQ ID NO:25
G T V L Q Q L * H L P L D G A
SEQ ID NO: 26
SEQ ID NO: 27

图14B

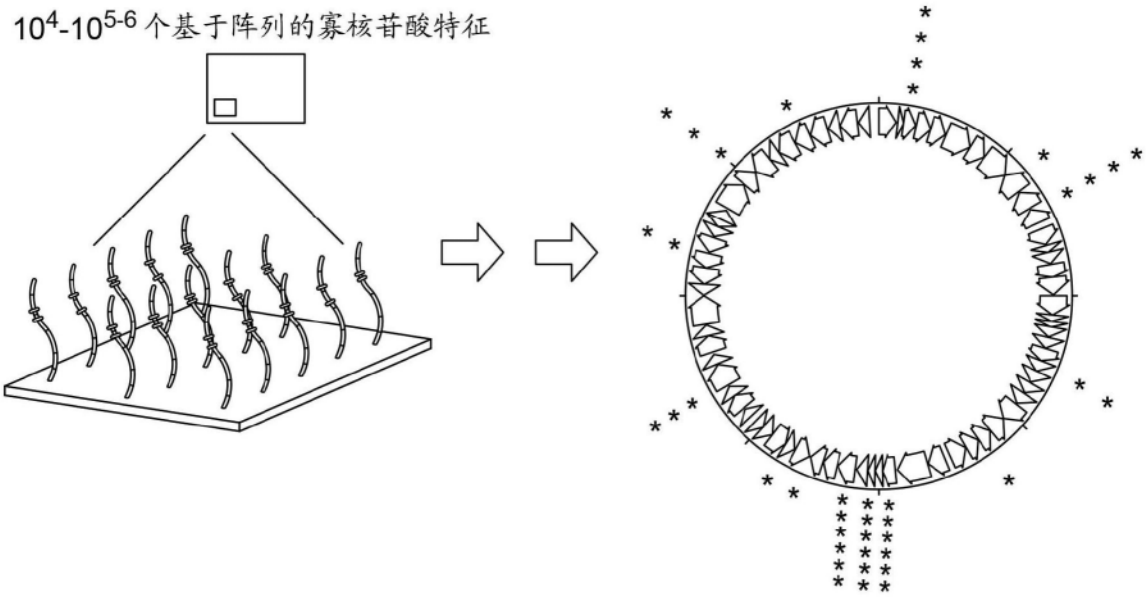
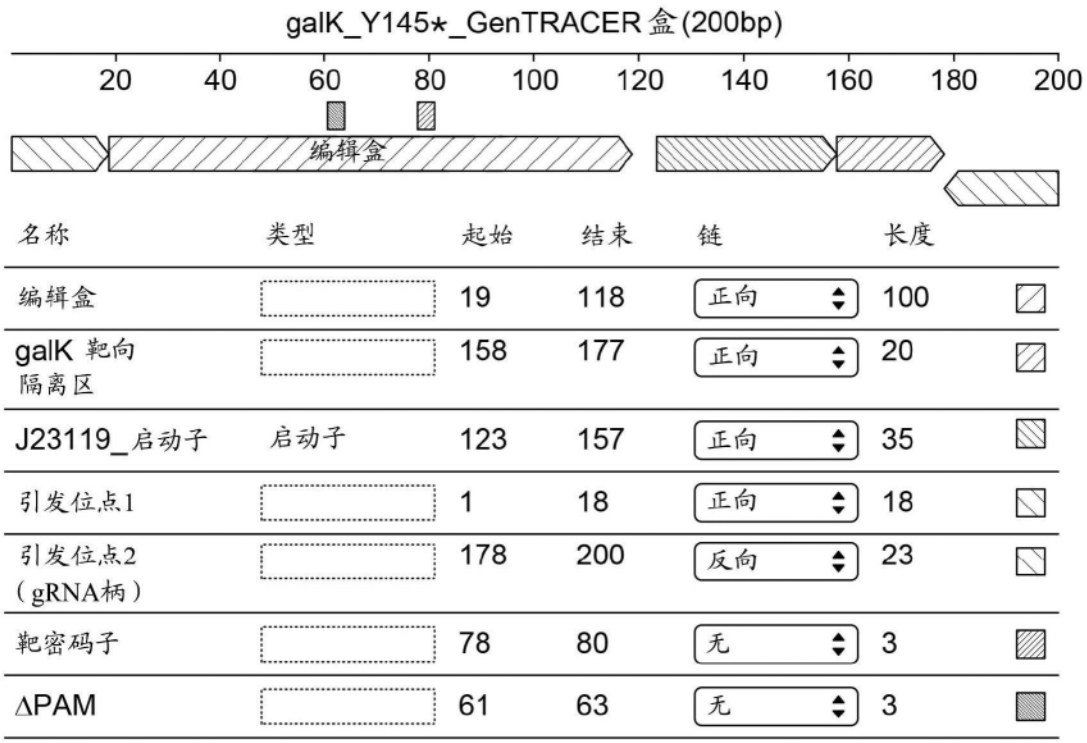

10^4-10^{5-6}个基于阵列的寡核苷酸特征

图14C



名称	类型	起始	结束	链	长度
编辑盒		19	118	正向	100
galK 靶向隔离区		158	177	正向	20
J23119_启动子	启动子	123	157	正向	35
引发位点1		1	18	正向	18
引发位点2（gRNA柄）		178	200	反向	23
靶密码子		78	80	无	3
ΔPAM		61	63	无	3

图15A

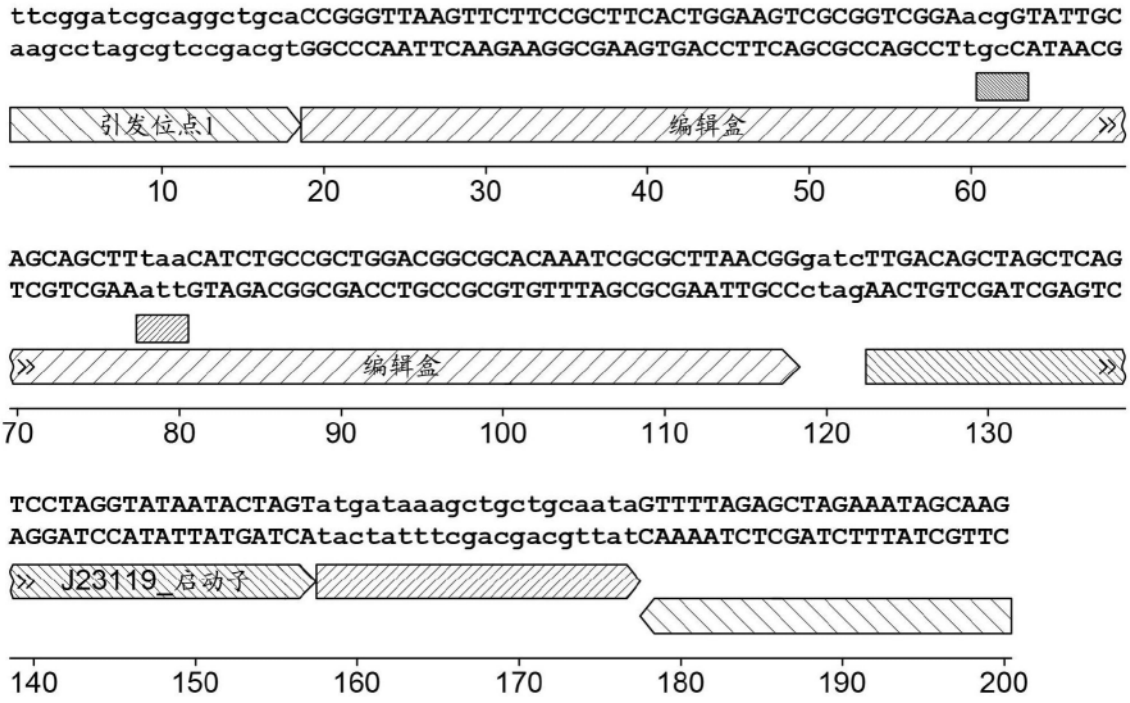


图15B

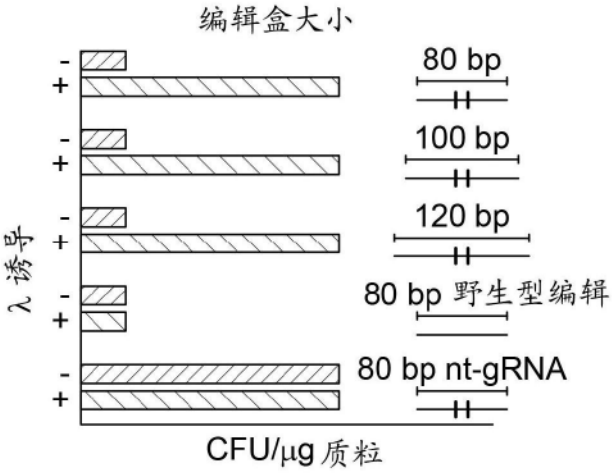

编辑盒大小
80 bp
100 bp
120 bp
80 bp 野生型编辑
80 bp nt-gRNA
入诱导
CFU/μg 质粒

图16A

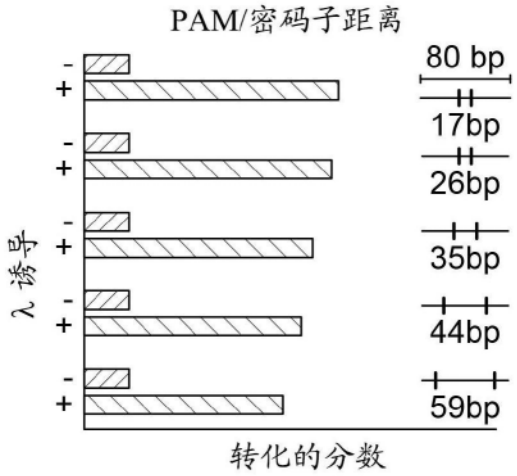

PAM/密码子距离
80 bp
17bp
26bp
35bp
44bp
59bp
入诱导
转化的分数

图16B

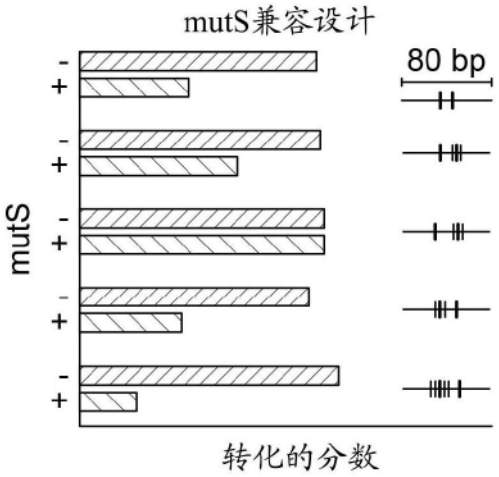


图16C