



(12)发明专利申请

(10)申请公布号 CN 111448313 A

(43)申请公布日 2020.07.24

(21)申请号 201880073647.3

M.博卢利-耶加内 L.M.迈尔

(22)申请日 2018.11.16

(74)专利代理机构 中国专利代理(香港)有限公司 72001

(30)优先权数据

62/587029 2017.11.16 US

62/693690 2018.07.03 US

代理人 郭慧 林毅斌

(85)PCT国际申请进入国家阶段日

2020.05.12

(51)Int.Cl.

C12N 15/10(2006.01)

C12N 9/22(2006.01)

(86)PCT国际申请的申请数据

PCT/US2018/061680 2018.11.16

(87)PCT国际申请的公布数据

W02019/099943 EN 2019.05.23

(71)申请人 阿斯利康(瑞典)有限公司

地址 瑞典南泰利耶

(72)发明人 M.马尔斯卡

A.塔赫里-加赫法罗基 F.卡尔松

权利要求书11页 说明书68页 附图96页

(54)发明名称

用于改善基于Cas9的敲入策略的有效性的组合物和方法

(57)摘要

本披露提供了一种非天然存在的CRISPR-Cas系统,该系统包含:能够产生粘性末端的Cas9效应蛋白(stiCas9),以及与该stiCas9形成复合物并且包含引导序列的引导多核苷酸,其中该引导序列与真核细胞中的靶序列杂交,但不与细菌细胞中的序列杂交,并且其中该复合物在自然界中不存在。本披露还提供了一种将目的序列引入细胞的染色体中的方法。最后,本披露提供了一种使用无缝诱变修饰一个或多个核苷酸的方法。

1. 一种非天然存在的CRISPR-Cas系统,该CRISPR-Cas系统包含:
 - a) 能够产生粘性末端的Cas9效应蛋白(stiCas9);和
 - b) 与该stiCas9形成复合物并且包含引导序列的引导多核苷酸,其中该引导序列能够与真核细胞中的靶序列杂交,但不与细菌细胞中的序列杂交;
其中该复合物在自然界中不存在。
2. 一种非天然存在的CRISPR-Cas系统,该CRISPR-Cas系统包含:
 - a) 能够产生粘性末端并且包含核定位序列(NLS)的Cas9效应蛋白(stiCas9);和
 - b) 与该stiCas9形成复合物并且包含引导序列的引导多核苷酸;
其中该复合物在自然界中不存在。
3. 一种非天然存在的CRISPR-Cas系统,该CRISPR-Cas系统包含:
 - a) 编码能够产生粘性末端的Cas9效应蛋白(stiCas9)的一个或多个核苷酸序列;和
 - b) 编码引导多核苷酸的核苷酸序列,该引导多核苷酸与该stiCas9形成复合物并且包含引导序列,其中该引导序列能够与真核细胞中的靶序列杂交,但不与细菌细胞中的序列杂交;
其中该复合物在自然界中不存在。
4. 一种非天然存在的CRISPR-Cas系统,该CRISPR-Cas系统包含:
 - a) 编码能够产生粘性末端的Cas9效应蛋白(stiCas9)的一个或多个核苷酸序列;和
 - b) 编码引导多核苷酸的核苷酸序列,该引导多核苷酸与该stiCas9形成复合物并包含引导序列;
其中(a)和(b)的核苷酸序列在真核启动子的控制下,并且其中该复合物在自然界中不存在。
5. 如权利要求1至4中任一项所述的CRISPR-Cas系统,其中该引导多核苷酸包含tracrRNA序列。
6. 如权利要求1至4中任一项所述的CRISPR-Cas系统,该CRISPR-Cas系统进一步包含含有tracrRNA序列的单独的多核苷酸。
7. 如权利要求6所述的CRISPR-Cas系统,其中该引导多核苷酸、tracrRNA序列和该stiCas9能够形成复合物,并且其中该复合物在自然界中不存在。
8. 一种非天然存在的CRISPR-Cas系统,该系统包含一个或多个载体,该一个或多个载体包含:
 - a) 可操作地连接至一个或多个核苷酸序列的调节元件,该核苷酸序列编码能够产生粘性末端的Cas9效应蛋白(stiCas9);和
 - b) 与该stiCas9形成复合物并包含引导序列的引导多核苷酸,其中该引导序列能够与真核细胞中的靶序列杂交,但不与细菌细胞中的序列杂交;
其中该复合物在自然界中不存在。
9. 一种非天然存在的CRISPR-Cas系统,该系统包含一个或多个载体,该一个或多个载体包含:
 - a) 可操作地连接至一个或多个核苷酸序列的调节元件,该核苷酸序列编码能够产生粘性末端的Cas9效应蛋白(stiCas9),其中该调节元件是真核调节元件;和
 - b) 与该stiCas9形成复合物并且包含引导序列的引导多核苷酸;

其中该复合物在自然界中不存在。

10. 如权利要求8或权利要求9所述的非天然存在的载体,其中该引导多核苷酸进一步包含tracrRNA序列。

11. 如权利要求9或权利要求10所述的非天然存在的载体,该载体进一步包含含有tracrRNA序列的核苷酸序列。

12. 如权利要求1至11中任一项所述的系统,其中该复合物能够在前间区序列邻近基序(PAM)的10个核苷酸内的位点处裂解。

13. 如权利要求1至12中任一项所述的系统,其中该复合物能够在前间区序列邻近基序(PAM)的5个核苷酸内的位点处裂解。

14. 如权利要求1至13中任一项所述的系统,其中该复合物能够在前间区序列邻近基序(PAM)的3个核苷酸内的位点处裂解。

15. 如权利要求1至14中任一项所述的系统,其中该靶序列是前间区序列邻近基序(PAM)的5' 并且该PAM包含富含3' G的基序。

16. 如权利要求1至15中任一项所述的系统,其中该靶序列是前间区序列邻近基序(PAM)的5' ,并且该PAM序列是NGG,其中N是A、C、G或T。

17. 如权利要求1至16中任一项所述的系统,其中这些粘性末端包含具有3至40个核苷酸的单链多核苷酸突出端。

18. 如权利要求1至17中任一项所述的系统,其中这些粘性末端包含具有4至20个核苷酸的单链多核苷酸突出端。

19. 如权利要求1至18中任一项所述的系统,其中这些粘性末端包含具有5至15个核苷酸的单链多核苷酸突出端。

20. 如权利要求1至19中任一项所述的系统,其中该stiCas9源自具有II-B型CRISPR系统的细菌物种。

21. 如权利要求1至20中任一项所述的系统,其中该stiCas9包含与SEQ ID NO:10-97或192-195中的任一个具有至少95%同一性的结构域。

22. 如权利要求1至21中任一项所述的系统,其中该stiCas9包含以1E-5的E值截止值匹配TIGR03031蛋白家族的结构域。

23. 如权利要求1至22中任一项所述的系统,其中该stiCas9包含以1E-10的E值截止值匹配该TIGR03031蛋白家族的结构域。

24. 如权利要求23所述的系统,其中该细菌物种是嗜肺军团菌、新凶手弗朗西斯菌、 γ 变形菌HTCC5015、人粪便副萨特氏菌、华德萨特氏菌、硫磺单胞菌属物种SC ADC、瘤胃杆菌属物种RM87、伯克霍尔德氏菌目细菌1_1_47、拟杆菌门口腔分类群274菌株F0058、产琥珀酸沃廉氏菌、伯克霍尔德氏菌目细菌YL45、嗜淀粉瘤胃杆菌、弯曲杆菌属物种P0111、弯曲杆菌属物种RM9261、拉尼尔弯曲杆菌菌株RM8001、拉尼尔弯曲杆菌菌株P0121、鼠毛滴虫、伦敦军团菌、沙姆盐弧菌、钩端螺旋体属物种分离株FW.030、钩里特拉氏菌属物种分离株NORP46、内生单胞菌属物种S-B4-1U、喜盐泰米尔纳德菌、需钠弧菌、斯氏弓形杆菌、蜃楼弗朗西斯氏菌、西班牙弗朗西斯氏菌或嗜盐副内生单胞菌。

25. 如权利要求24所述的系统,其中该靶序列是前间区序列邻近基序(PAM)的5' ,并且该PAM序列是YG,其中Y是嘧啶,并且该stiCas9源自细菌物种新凶手弗朗西斯菌。

26. 如权利要求1至25中任一项所述的系统,其中该stiCas9包含一个或多个核定位信号。

27. 如权利要求1至26中任一项所述的系统,其中该真核细胞是动物或人类细胞。

28. 如权利要求1至27中任一项所述的系统,其中该真核细胞是人类细胞。

29. 如权利要求1至26中任一项所述的系统,其中该真核细胞是植物细胞。

30. 如权利要求1至29中任一项所述的系统,其中该引导序列与同向重复序列连接。

31. 一种递送颗粒,该递送颗粒包含如权利要求1至30中任一项所述的系统。

32. 如权利要求31所述的递送颗粒,其中该stiCas9和该引导多核苷酸以复合物存在。

33. 如权利要求32所述的递送颗粒,其中该复合物进一步包含含有tracrRNA序列的多核苷酸。

34. 如权利要求32或22所述的递送颗粒,该递送颗粒进一步包含脂质、糖、金属或蛋白质。

35. 一种囊泡,该囊泡包含如权利要求1至30中任一项所述的系统。

36. 如权利要求35所述的囊泡,其中该stiCas9和该引导多核苷酸以复合物存在。

37. 如权利要求36所述的囊泡,该囊泡进一步包含:含有tracrRNA序列的多核苷酸。

38. 如权利要求35至37中任一项所述的囊泡,其中该囊泡是外泌体或脂质体。

39. 如权利要求5至9中任一项所述的系统,其中对编码该stiCas9的一个或多个核苷酸序列进行密码子优化以在真核细胞中表达。

40. 如权利要求5至30或39中任一项所述的系统,其中该编码Cas9效应蛋白的核苷酸序列和该引导多核苷酸在单个载体上。

41. 如权利要求5至30或39中任一项所述的系统,其中该编码Cas9效应蛋白的核苷酸序列和该引导多核苷酸是单个核酸分子。

42. 一种病毒载体,该病毒载体包含如权利要求5至30或39至41中任一项所述的系统。

43. 如权利要求42所述的病毒载体,其中该病毒载体是腺病毒、慢病毒或腺相关病毒载体。

44. 一种包含CRISPR-Cas系统的真核细胞,该真核细胞包含

a) 能够产生粘性末端的Cas9效应蛋白(stiCas9),以及

b) 与该stiCas9形成复合物并且包含引导序列的引导多核苷酸,其中该引导序列能够与真核细胞中的靶序列杂交;

其中该复合物在自然界中不存在。

45. 一种包含CRISPR-Cas系统的真核细胞,该CRISPR-Cas系统包含能够产生粘性末端的Cas9效应蛋白(stiCas9),其中该Cas9效应蛋白源自具有II-B型CRISPR系统的细菌物种。

46. 一种用于在真核细胞中提供靶序列的位点特异性修饰的方法,该方法包括:

a) 将以下引入到该细胞中:

i. 能够产生粘性末端的Cas9效应蛋白(stiCas9);和

ii 与该stiCas9形成复合物并包含引导序列的引导多核苷酸,其中该引导序列能够与真核细胞中的靶序列杂交,但不与细菌细胞中的序列杂交;

其中该复合物在自然界中不存在;以及

b) 用该Cas9效应蛋白和该引导多核苷酸在该靶序列中产生粘性末端;以及

- c)
- i. 将这些粘性末端连接在一起, 或者
- ii 将目的多核苷酸序列 (SoI) 与这些粘性末端连接;
从而修饰该靶序列。
47. 一种用于在真核细胞中提供靶序列的位点特异性修饰的方法, 该方法包括:
- a) 将以下引入到该细胞中:
- i. 编码能够产生粘性末端的Cas9效应蛋白 (stiCas9) 的核苷酸序列; 和
- ii 与该stiCas9形成复合物并包含引导序列的引导多核苷酸, 其中该引导序列能够与真核细胞中的靶序列杂交, 但不与细菌细胞中的序列杂交;
其中该复合物在自然界中不存在; 以及
- b) 用该Cas9效应蛋白和该引导多核苷酸在该靶序列中产生粘性末端; 以及
- c)
- i. 将这些粘性末端连接在一起, 或者
- ii 将目的多核苷酸序列 (SoI) 与这些粘性末端连接;
从而修饰该靶序列。
48. 如权利要求46或47所述的方法, 其中该引导多核苷酸进一步包含tracrRNA序列。
49. 如权利要求46或47所述的方法, 该方法进一步包括将包含tracrRNA序列的多核苷酸引入该细胞中。
50. 如权利要求49所述的方法, 其中该引导多核苷酸、tracrRNA序列与该stiCas9能够形成复合物, 并且其中该复合物在自然界中不存在。
51. 如权利要求46至50中任一项所述的方法, 其中该复合物能够在前间区序列邻近基序 (PAM) 的10个核苷酸内的位点处裂解。
52. 如权利要求46至51中任一项所述的方法, 其中该复合物能够在前间区序列邻近基序 (PAM) 的5个核苷酸内的位点处裂解。
53. 如权利要求46至52中任一项所述的方法, 其中该复合物能够在前间区序列邻近基序 (PAM) 的3个核苷酸内的位点处裂解。
54. 如权利要求46至53中任一项所述的方法, 其中该靶序列是前间区序列邻近基序 (PAM) 的5' 并且该PAM包含富含3' G的基序。
55. 如权利要求46至54中任一项所述的方法, 其中该靶序列是PAM的5', 并且该PAM序列是NGG, 其中N是A、C、G或T。
56. 如权利要求46至55中任一项所述的方法, 其中这些粘性末端包含具有3至40个核苷酸的单链多核苷酸突出端。
57. 如权利要求46至56中任一项所述的方法, 其中这些粘性末端包含具有4至20个核苷酸的单链多核苷酸突出端。
58. 如权利要求46至57中任一项所述的方法, 其中这些粘性末端包含具有5至15个核苷酸的单链多核苷酸突出端。
59. 如权利要求46至58中任一项所述的方法, 其中该stiCas9源自具有II-B型CRISPR系统的细菌物种。
60. 如权利要求46至59中任一项所述的方法, 其中该真核细胞是动物或人类细胞。

61. 如权利要求46至60中任一项所述的方法,其中该真核细胞是人类细胞。
62. 如权利要求46至59中任一项所述的方法,其中该真核细胞是植物细胞。
63. 如权利要求46至62中任一项所述的方法,其中该修饰是该靶序列的至少一部分的缺失。
64. 如权利要求46至62中任一项所述的方法,其中该修饰是该靶序列的突变。
65. 如权利要求46至62中任一项所述的方法,其中该修饰是将目的序列插入该靶序列中。
66. 如权利要求46至65中任一项所述的方法,该方法进一步包括引入核酸外切酶以去除由该stiCas9产生的突出端。
67. 如权利要求66所述的方法,其中该核酸外切酶是Cas4、Artemis或TREX2。
68. 如权利要求67所述的方法,其中该Cas4源自具有II-B型CRISPR系统的细菌物种。
69. 如权利要求46至68中任一项所述的方法,其中将编码该复合物的组分的多核苷酸引入一个或多个载体上。
70. 一种将目的序列 (SoI) 引入细胞的染色体中的方法,其中该染色体包含含有区域1和区域2的靶序列 (TSC),该方法包括将以下引入该细胞中:
- 包含靶序列的载体 (TSV),该TSV包含区域2和区域1以及该SoI;
 - 能够在该TSC中产生粘性末端的第一Cas9-核酸内切酶二聚体,其中该第一Cas9-核酸内切酶二聚体的第一单体在该TSC的区域1处裂解,并且该第一Cas9-核酸内切酶二聚体的第二单体在该TSC的区域2处裂解;以及
 - 能够在该TSV中产生粘性末端的第二Cas9-核酸内切酶二聚体,其中该第二Cas9-核酸内切酶二聚体的第一单体在该TSV的区域2处裂解,并且该第二Cas9-核酸内切酶二聚体的第二单体在该TSV的区域1处裂解;
- 其中将 (a) 的载体、(b) 的第一Cas9-核酸内切酶二聚体和 (c) 的第二Cas9-核酸内切酶二聚体引入该细胞中导致该SoI被插入该细胞的染色体中。
71. 一种将目的序列 (SoI) 引入细胞的染色体中的方法,其中该染色体包含含有区域1和区域2的靶序列 (TSC),该方法包括将以下引入该细胞中:
- 包含靶序列的载体 (TSV),该TSV包含区域2和区域1以及该SoI,其中该载体包含粘性末端;
 - 能够在该TSC中产生粘性末端的第一Cas9-核酸内切酶二聚体,其中该Cas9-核酸内切酶二聚体的第一单体在该TSC的区域1处裂解,并且该Cas9-核酸内切酶二聚体的第二单体在该TSC的区域2处裂解;
- 其中将 (a) 的载体和 (b) 的第一Cas9-核酸内切酶二聚体引入该细胞中导致该SoI被插入该细胞的染色体中。
72. 如权利要求70或权利要求71所述的方法,其中该第一和第二Cas9-核酸内切酶二聚体是相同的。
73. 如权利要求70或权利要求71所述的方法,其中该第一和第二Cas9-核酸内切酶二聚体是不同的。
74. 如权利要求70至73中任一项所述的方法,该方法进一步包括将第一引导多核苷酸引入该细胞中,该第一引导多核苷酸与该第一Cas9-核酸内切酶二聚体的第一单体形成复

合物并且包含第一引导序列,其中该第一引导序列与该包含区域1的TSC杂交,但不与该载体杂交。

75.如权利要求70至73中任一项所述的方法,该方法进一步包括将第一引导多核苷酸引入该细胞中,该第一引导多核苷酸与该第一Cas9-核酸内切酶二聚体的第一单体形成复合物并包含第一引导序列,其中该第一引导序列与该TSC和该TSV杂交。

76.如权利要求70至75中任一项所述的方法,该方法进一步包括将第二引导多核苷酸引入该细胞中,该第二引导多核苷酸与该第一Cas9-核酸内切酶二聚体的第二单体形成复合物并包含第二引导序列,其中该第二引导序列与该包含区域2的TSC杂交,但不与载体杂交。

77.如权利要求70至75中任一项所述的方法,该方法进一步包括将第二引导多核苷酸引入该细胞中,该第二引导多核苷酸与该第一Cas9-核酸内切酶二聚体的第二单体形成复合物并包含第二引导序列,其中该第二引导序列与该TSC和该TSV杂交。

78.如权利要求70至77中任一项所述的方法,该方法进一步包括将第三引导多核苷酸引入该细胞中,该第三引导多核苷酸与该第二Cas9-核酸内切酶二聚体的第一单体形成复合物并包含第三引导序列,其中该第三引导序列与该包含区域2的TSV杂交,但不与该染色体杂交。

79.如权利要求70至78所述的方法,该方法进一步包括将第三引导多核苷酸引入该细胞中,该第三引导多核苷酸与该第二Cas9-核酸内切酶二聚体的第一单体形成复合物并包含第三引导序列,其中该第三引导序列与该TSC和该TSV杂交。

80.如权利要求70至79中任一项所述的方法,该方法进一步包括将第四引导多核苷酸引入该细胞中,该第四引导多核苷酸与该第二Cas9-核酸内切酶二聚体的第二单体形成复合物并包含第四引导序列,其中该第四引导序列与该包含区域1的TSV杂交,但不与染色体杂交。

81.如权利要求70至80中任一项所述的方法,该方法进一步包括将第四引导多核苷酸引入该细胞中,该第四引导多核苷酸与该第二Cas9-核酸内切酶二聚体的第二单体形成复合物并包含第四引导序列,其中该第四引导序列与TSC和TSV杂交。

82.如权利要求70至81中任一项所述的方法,该方法包括将该第一、第二、第三和第四引导多核苷酸引入该细胞中。

83.如权利要求70至82中任一项所述的方法,该方法进一步包括将包含tracrRNA序列的多核苷酸引入该细胞中。

84.如权利要求70至83中任一项所述的方法,其中该第一Cas9-核酸内切酶二聚体的第一单体和第二单体中的核酸内切酶是IIS型核酸内切酶。

85.如权利要求70至83中任一项所述的方法,其中该第二Cas9-核酸内切酶二聚体的第一单体和第二单体中的核酸内切酶是IIS型核酸内切酶。

86.如权利要求70至85中任一项所述的方法,其中该第一Cas9-核酸内切酶二聚体和该第二Cas9-核酸内切酶二聚体中的核酸内切酶是IIS型核酸内切酶。

87.如权利要求70至86中任一项所述的方法,其中该第一Cas9-核酸内切酶二聚体和该第二Cas9-核酸内切酶二聚体中的核酸内切酶独立地选自下组,该组由以下组成: BbvI、BgcI、BfuAI、BmpI、BspMI、CspCI、FokI、MboII、MmeI、NmeAIII和PleI。

88. 如权利要求70至87中任一项所述的方法,其中该第一Cas9-核酸内切酶二聚体和该第二Cas9-核酸内切酶二聚体中的核酸内切酶是FokI。

89. 如权利要求70至88中任一项所述的方法,其中将该第一和第二Cas9-核酸内切酶二聚体作为编码该第一和第二Cas9-核酸内切酶二聚体的多核苷酸引入该细胞中。

90. 如权利要求89所述的方法,其中该编码该第一和第二Cas9-核酸内切酶二聚体的多核苷酸在一个载体上。

91. 如权利要求89所述的方法,其中该编码该第一和第二Cas9-核酸内切酶二聚体的多核苷酸在多于一个载体上。

92. 如权利要求70至91中任一项所述的方法,其中该第一、第二Cas9-核酸内切酶二聚体或两者包含经修饰的Cas9。

93. 如权利要求92所述的方法,其中该第一、第二Cas9-核酸内切酶二聚体或两者包含无催化活性的Cas9。

94. 如权利要求93所述的方法,其中该第一、第二Cas9-核酸内切酶二聚体或两者中的核酸内切酶是FokI。

95. 如权利要求92所述的方法,其中该第一、第二Cas9-核酸内切酶二聚体或两者包含具有切口酶活性的Cas9。

96. 如权利要求95所述的方法,其中该第一、第二Cas9-核酸内切酶二聚体或两者中的核酸内切酶是FokI。

97. 如权利要求92所述的方法,其中该Cas9-核酸内切酶二聚体相对于野生型Cas9在Cas9中包含单个氨基酸取代。

98. 如权利要求97所述的方法,其中该第一、第二Cas9-核酸内切酶二聚体或两者中的核酸内切酶是FokI。

99. 如权利要求97或98所述的方法,其中该单个氨基酸取代是D10A或H840A。

100. 如权利要求97或98所述的方法,其中该单个氨基酸取代是D10A。

101. 如权利要求97或98所述的方法,其中该单个氨基酸取代是H840A。

102. 如权利要求92所述的方法,其中该Cas9-核酸内切酶二聚体相对于野生型Cas9包含双氨基酸取代。

103. 如权利要求102所述的方法,其中该双氨基酸取代是D10A和H840A。

104. 如权利要求97所述的方法,其中该野生型Cas9源自酿脓链球菌、金黄色葡萄球菌、伪中间型葡萄球菌、南极游动球菌、血链球菌、嗜热链球菌、变形链球菌、球团科里氏杆菌、香肠乳杆菌、光冈氏链小杆菌、鼠李糖乳杆菌、两歧双歧杆菌、北原酒球菌、果聚糖杆菌、大芬戈尔德菌、非典型韦荣氏球菌、穆氏索罗菌、氨基酸球菌属物种D21、尤里真杆菌、灵巧粪球菌、具核梭杆菌、龈沟产线菌、杜尔丹尼嗜脓菌或齿垢密螺旋体。

105. 如权利要求70至104中任一项所述的方法,其中这些粘性末端包括5' 突出端。

106. 如权利要求70至104中任一项所述的方法,其中这些粘性末端包括3' 突出端。

107. 如权利要求70至106中任一项所述的方法,其中该第一、第二Cas9-核酸内切酶二聚体或两者产生包含具有3至40个核苷酸的单链多核苷酸的粘性末端。

108. 如权利要求70至106中任一项所述的方法,其中该第一、第二Cas9-核酸内切酶二聚体或两者产生包含具有4至30个核苷酸的单链多核苷酸的粘性末端。

109. 如权利要求70至106中任一项所述的方法,其中该第一、第二Cas9-核酸内切酶二聚体或两者产生包含具有5至20个核苷酸的单链多核苷酸的粘性末端。

110. 如权利要求70至109中任一项所述的方法,其中在插入时,该染色体中的靶序列和该质粒中的靶序列没有被重建。

111. 如权利要求70至110中任一项所述的方法,其中该细胞是真核细胞。

112. 如权利要求70至111中任一项所述的方法,其中该细胞是动物或人类细胞。

113. 如权利要求70至112中任一项所述的方法,其中该细胞是植物细胞。

114. 如权利要求70至113中任一项所述的方法,其中将(a)的载体、(b)的第一Cas9-核酸内切酶二聚体、(c)的第二Cas9-核酸内切酶二聚体或其组合通过递送颗粒、囊泡或病毒载体引入该细胞中。

115. 如权利要求70至114中任一项所述的方法,其中将(a)的载体、(b)的第一Cas9-核酸内切酶二聚体、(c)的第二Cas9-核酸内切酶二聚体或其组合通过递送颗粒引入该细胞中。

116. 如权利要求115所述的方法,其中这些递送颗粒包含脂质、糖、金属或蛋白质。

117. 如权利要求70至114中任一项所述的方法,其中将(a)的载体、(b)的第一Cas9-核酸内切酶二聚体、(c)的第二Cas9-核酸内切酶二聚体或其组合通过囊泡引入该细胞中。

118. 如权利要求117所述的方法,其中这些囊泡是外泌体或脂质体。

119. 如权利要求70至113中任一项所述的方法,其中将能够或表达(b)、(c)或其组合的多核苷酸通过病毒载体引入该细胞中。

120. 如权利要求70至113中任一项所述的方法,其中(a)的载体是病毒载体。

121. 如权利要求119或120所述的方法,其中该病毒载体是腺病毒、慢病毒或腺相关病毒载体。

122. 如权利要求70至121中任一项所述的方法,其中该第一Cas9-核酸内切酶二聚体的第一单体与该第一引导多核苷酸形成复合物,并且该第一Cas9-核酸内切酶二聚体的第二单体与该第二引导多核苷酸形成复合物。

123. 如权利要求70至122中任一项所述的方法,其中该第二Cas9-核酸内切酶二聚体的第一单体与该第三引导多核苷酸形成复合物,并且该第二Cas9-核酸内切酶二聚体的第二单体与该第四引导多核苷酸形成复合物。

124. 如权利要求70至121中任一项所述的方法,其中该第一Cas9-核酸内切酶二聚体的第一单体与该第一引导多核苷酸序列和tracrRNA序列形成复合物,并且该第一Cas9-核酸内切酶二聚体的第二单体与该第二引导多核苷酸序列和tracrRNA序列形成复合物。

125. 如权利要求70至122中任一项所述的方法,其中该第二Cas9-核酸内切酶二聚体的第一单体与该第三引导多核苷酸序列和tracrRNA序列形成复合物,并且该第二Cas9-核酸内切酶二聚体的第二单体与该第四引导多核苷酸序列和tracrRNA序列形成复合物。

126. 如权利要求70至125中任一项所述的方法,其中该第一、第二Cas9-核酸内切酶二聚体或两者包含核定位信号。

127. 如权利要求70至126中任一项所述的方法,其中该细胞包含干细胞或干细胞系。

128. 一种修饰细胞中的靶多核苷酸序列中一个或多个核苷酸的方法,该方法包括:

a) 将包含插入盒(IC)的载体引入该细胞中,该IC沿5'至3'方向包含:

i. 与该靶多核苷酸序列的一部分同源的第一区域，
ii. 包含该靶多核苷酸序列中一个或多个核苷酸的突变的第二区域，
iii. 第一核酸酶结合位点，
iv. 编码标记基因的多核苷酸序列，
v. 第二核酸酶结合位点
vi. 包含该靶多核苷酸序列中一个或多个核苷酸的突变的第三区域，以及
vii. 与该靶多核苷酸序列的一部分同源的第四区域，其中该第一区域和该第四区域与它们在该靶多核苷酸序列中的相应部分95%-100%相同；

b) 通过同源重组将该IC插入该靶多核苷酸序列以产生第一经修饰的靶多核苷酸；

c) 选择表达该标记基因的细胞；

d) 使该第一经修饰的靶多核苷酸经受位点特异性核酸酶处理以产生具有粘性末端的第二经修饰的靶多核苷酸；以及

e) 使该具有粘性末端的第二经修饰的靶多核苷酸经受连接酶处理，其中该连接酶在该第二区域和该第三区域处连接这些粘性末端以产生连接的经修饰的靶核酸，当与该靶多核苷酸序列相比时，该连接的经修饰的靶核酸包含一个或多个经修饰的核苷酸。

129. 如权利要求128所述的方法，其中在(c)之后，从该细胞中分离出该第一经修饰的靶核酸。

130. 如权利要求128或129所述的方法，其中该位点特异性核酸酶对于该细胞而言是外源性的。

131. 如权利要求128至130中任一项所述的方法，其中该连接酶对于该细胞而言是外源性的。

132. 如权利要求128所述的方法，其中在(c)之后，该第一经修饰的靶蛋白在该细胞中。

133. 如权利要求132所述的方法，其中将该位点特异性核酸酶作为编码该位点特异性核酸酶的多核苷酸引入该细胞中。

134. 如权利要求132或133所述的方法，其中将该连接酶作为编码连接酶的多核苷酸引入该细胞中。

135. 如权利要求128至134中任一项所述的方法，其中该位点特异性核酸酶是重组位点特异性核酸酶。

136. 如权利要求128至135中任一项所述的方法，其中该连接酶是重组连接酶。

137. 如权利要求128至136中任一项所述的方法，其中该位点特异性核酸酶是Cas9效应蛋白。

138. 如权利要求137所述的方法，其中该Cas9效应蛋白是II-B型Cas9。

139. 如权利要求128至131中任一项所述的方法，其中该位点特异性核酸酶是Cas9-核酸内切酶融合蛋白。

140. 如权利要求139所述的方法，其中该Cas9-核酸内切酶融合蛋白中的核酸内切酶是IIS型核酸内切酶。

141. 如权利要求139所述的方法，其中该Cas9-核酸内切酶融合蛋白中的核酸内切酶是FokI。

142. 如权利要求139至141中任一项所述的方法，其中该Cas9-核酸内切酶融合蛋白包

含经修饰的Cas9。

143. 如权利要求142所述的方法,其中该经修饰的Cas9包含无催化活性的Cas9。

144. 如权利要求143所述的方法,其中该核酸内切酶是FokI。

145. 如权利要求142所述的方法,其中该Cas9-核酸内切酶融合蛋白包含具有切口酶活性的Cas9,并且该核酸内切酶是FokI。

146. 如权利要求143所述的方法,其中该Cas9-核酸内切酶融合蛋白包含具有D10A取代的Cas9。

147. 如权利要求143所述的方法,其中该Cas9-核酸内切酶融合蛋白包含具有H840A取代的Cas9。

148. 如权利要求128所述的方法,其中该位点特异性核酸酶是Cas9、Cpf1或Cas9-FokI。

149. 如权利要求128所述的方法,其中该位点特异性核酸酶是Cpf1效应蛋白。

150. 如权利要求128至149中任一项所述的方法,其中(d)的第二经修饰的靶多核苷酸的粘性末端包含5'突出端。

151. 如权利要求128至149中任一项所述的方法,其中(d)的第二经修饰的靶多核苷酸的粘性末端包含3'突出端。

152. 如权利要求128至151中任一项所述的方法,其中该位点特异性核酸酶能够产生包含具有3至40个核苷酸的单链多核苷酸的粘性末端。

153. 如权利要求128至151中任一项所述的方法,其中该核酸酶能够产生包含具有4至30个核苷酸的单链多核苷酸的粘性末端。

154. 如权利要求128至151中任一项所述的方法,其中该核酸酶能够产生包含具有5至20个核苷酸的单链多核苷酸的粘性末端。

155. 如权利要求128至154中任一项所述的方法,其中该靶多核苷酸序列处于质粒中。

156. 如权利要求128至155中任一项所述的方法,其中该靶多核苷酸序列处于染色体中。

157. 一种与stiCas9蛋白形成复合物的工程化的引导RNA,该工程化的引导RNA包含:

a) 能够与真核细胞中的靶序列杂交的引导序列;和

b) 能够与该Cas9蛋白结合的tracrRNA序列,其中该tracrRNA与天然存在的tracrRNA序列的差异为至少10个核苷酸,

其中该工程化的引导RNA改善了该Cas9蛋白的核酸酶效率。

158. 如权利要求157所述的工程化的引导RNA,其中该tracrRNA序列比天然存在的tracrRNA少至少10个核苷酸。

159. 如权利要求157所述的工程化的引导RNA,其中该tracrRNA序列比天然存在的tracrRNA多至少10个核苷酸。

160. 如权利要求157所述的工程化的引导RNA,其中该引导序列与SEQ ID NO:104-125或196-199中的任一个具有至少90%的序列同一性。

161. 如权利要求157所述的工程化的引导RNA,其中该tracrRNA序列与SEQ ID NO:148-171中的任一个具有至少90%的序列同一性。

162. 如权利要求157所述的工程化的引导RNA,其中该引导RNA序列与SEQ ID NO:172-191中的任一个具有至少90%的序列同一性。

163. 如权利要求157至159中任一项所述的工程化的引导RNA,其中该tracrRNA包含该tracrRNA的茎环中的一个或多个修饰。

164. 如权利要求163所述的工程化的引导RNA,其中该修饰包含该茎环的延长。

165. 如权利要求163所述的工程化的引导RNA,其中该修饰包含该茎环的缩短。

166. 如权利要求163所述的工程化的引导RNA,其中该修饰包含该茎环中的一个或多个核苷酸取代。

167. 如权利要求157至166中任一项所述的工程化的引导RNA,其中该Cas9蛋白的改善的核酸酶效率是通过生化测定、测序测定和/或亲和力测试确定的。

168. 一种CRISPR-Cas系统,该系统包含如权利要求157至163中任一项所述的工程化的引导RNA。

169. 一种工程化的Cas9-引导RNA复合物,该复合物包含如图40B中所示的Cas9、引导序列和tracrRNA序列的任何组合。

170. 如权利要求163所述的CRISPR-Cas系统,其中该系统不包含单独的多核苷酸上的tracrRNA序列。

171. 一种产生与Cas9蛋白结合的工程化的引导RNA的方法,该方法包括:

- a. 提供能够与真核细胞中的靶序列杂交的引导序列;
- b. 将天然存在的tracrRNA序列通过从该tracrRNA序列中除去至少十个核苷酸来进行修饰,以形成经修饰的tracrRNA序列;以及
- c. 将该引导序列与该经修饰的tracrRNA序列连接以产生该工程化的引导RNA。

172. 一种非天然存在的CRISPR-Cas系统,该CRISPR-Cas系统包含:

- a) 能够产生粘性末端的Cas9效应蛋白(stiCas9);和
- b) 与该stiCas9形成复合物并且包含引导序列的引导RNA,其中该引导序列能够与真核细胞中的靶序列杂交,但不与细菌细胞中的序列杂交;
其中该复合物在自然界中不存在,并且
其中该系统不包含单独的多核苷酸上的tracrRNA序列。

用于改善基于Cas9的敲入策略的有效性的组合物和方法

[0001] 序列表

[0002] 本申请含有已以ASCII格式以电子方式提交且特此通过引用以其整体并入的序列表。所述ASCII副本创建于2018年11月16日,名称为0098-0002W01_SL.txt并且大小为1,105,014字节。

技术领域

[0003] 本披露提供了一种非天然存在的CRISPR-Cas系统,该系统包含:能够产生粘性末端的Cas9效应蛋白(stiCas9),以及与该stiCas9形成复合物并且包含引导序列的引导多核苷酸,其中该引导序列与真核细胞中的靶序列杂交,但不与细菌细胞中的序列杂交,并且其中该复合物在自然界中不存在。

背景技术

[0004] 成簇的规律间隔的短回文重复序列(CRISPR)和CRISPR相关(Cas)系统是由Ishino在大肠杆菌(*E. coli*)中首次发现的原核生物免疫系统(Ishino等人, *Journal of Bacteriology* [细菌学杂志] 169 (12):5429-5433 (1987),通过引用以其整体并入本文)。该免疫系统通过以序列特异性方式靶向病毒和质粒的核酸来提供针对病毒和质粒的免疫。另请参见Soret等人,“CRISPR—a widespread system that provides acquired resistance against phages in bacteria and archaea [CRISPR—在细菌和古细菌中可提供对噬菌体的获得性抗性的广泛存在的系统]”, *Nature Reviews Microbiology* [自然微生物学综述] 6 (3):181-186 (2008),通过引用以其整体并入本文。CRISPR-Cas系统已被分为三种主要类型:I型、II型和III型。单独类型的主要定义特征是使用的不同cas基因以及它们编码的相应蛋白质。cas1和cas2基因显现在三种主要类型中通用,而cas3、cas9和cas10被认为分别特定于I型、II型和III型系统。参见,例如,Barrangou和Marraffini,“CRISPR-Cas systems:prokaryotes upgrade to adaptive immunity [CRISPR-Cas系统:原核生物的适应性免疫升级]”, *Cell* [细胞] 54 (2):234-244 (2014),通过引用以其整体并入本文。

[0005] 该免疫系统包含两个主要阶段:第一个是采集,第二个是干扰。第一阶段包括切割入侵病毒和质粒的基因组,并将其区段整合到该生物体的CRISPR基因座中。整合到基因组中的这些区段称为前间区序列,有助于保护该生物体免受相同病毒或质粒的后续攻击。第二阶段包括攻击入侵的病毒或质粒。该阶段依赖于将前间区序列转录为RNA,经过一些加工后,该RNA与入侵病毒或质粒的DNA中的互补序列杂交,同时还与有效裂解DNA的蛋白质或蛋白质复合物缔合。

[0006] 根据细菌物种的不同,CRISPR RNA加工的过程有所不同。例如,最初在细菌酿脓链球菌(*Streptococcus pyogenes*)中描述的II型系统中,转录的RNA与反式激活RNA(tracrRNA)配对,之后被RNA酶III裂解以形成单个CRISPR-RNA(crRNA)。在被Cas9核酸酶结合后,crRNA被进一步加工以产生成熟的crRNA。该crRNA/Cas9复合物随后结合包含与捕获区域互补的序列(称为前间区序列)的DNA。然后,该Cas9蛋白以位点特异性的方式裂解DNA

的两条链,形成双链断裂(DSB)。这提供了基于DNA的“记忆”,导致病毒或质粒DNA在重复暴露和/或感染后迅速降解。关于原生CRISPR系统已经有全面的综述(参见,例如,Barrangou和Marraffini,2014)。

[0007] 自最初发现以来,多个研究小组围绕CRISPR系统在基因工程中的潜在应用进行了大量的研究,包括基因编辑(Jinek等人,“A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity[适应性细菌免疫中的可编程双-RNA引导DNA核酸内切酶]”,*Science*[科学]337(6096):816-821(2012);Cong等人,“Multiplex genome engineering using CRISPR/Cas systems[使用CRISPR/Cas系统的多重基因组工程]”,*Science*[科学]339(6121):819-823(2013);以及Mali等人,“RNA-guided human genome engineering via Cas9[通过Cas9进行RNA引导的人类基因组工程]”,*Science*[科学]339(6121):823-826(2013);将其各自通过引用以其整体并入本文)。一个重大发展是利用嵌合RNA靶向Cas9蛋白,这是围绕CRISPR阵列中与tracrRNA融合的单单元设计的。这会创建一个称为小引导RNA(gRNA)的单个RNA种类,其中该前间区序列区域中序列的修饰可以位点特异性地靶向该Cas9蛋白。已经开展了大量的工作来了解嵌合RNA与靶位点之间碱基配对相互作用的性质及其对错配的耐受性,这对于预测和评估脱靶效应是高度相关的(参见,例如,Fu等人,“Improving CRISPR-Cas nucleases using truncated guide RNAs[使用截短的引导RNA改进CRISPR-Cas核酸酶]”,*Nature Biotechnology*[自然生物技术]32(3):279-284(2014),包括支持材料,通过引用以其整体并入本文)。

[0008] CRISPR-Cas9基因编辑系统已成功用于许多种生物体和细胞系中,既为了使用野生型Cas9蛋白诱导DSB形成,又为了使用称为Cas9n/Cas9 D10A的突变型蛋白切割一条DNA单链(参见,例如,Mali等人,2013和Sander和Joung,“CRISPR-Cas systems for editing, regulating and targeting genomes[用于编辑、调节和靶向基因组的多种CRISPR-Cas系统]”,*Nature Biotechnology*[自然生物技术]32(4):347-355(2014),将其各自通过引用以其整体并入本文)。虽然DSB的形成导致可能破坏基因功能的小插入和缺失(插缺(indel))的产生,但Cas9n/Cas9 D10A切口酶避免了插缺的产生(通过非同源末端连接进行修复的结果),同时刺激了内源同源重组机制。因此,该Cas9n/Cas9 D10A切口酶可用于将DNA区域高保真地插入基因组。

[0009] 除基因组编辑外,CRISPR系统还具有许多其他应用,包括调节基因表达、基因回路构建和功能基因组学等(在Sander和Joung,2014中综述)。

[0010] 本文引用了各种出版物,这些出版物的披露内容通过引用以其整体并入本文。

发明内容

[0011] 在一些实施例中,本披露提供了一种非天然存在的CRISPR-Cas系统,该系统包含:能够产生粘性末端的Cas9效应蛋白(stiCas9),以及与该stiCas9形成复合物并且包含引导序列的引导多核苷酸,其中该引导序列与真核细胞中的靶序列杂交,但不与细菌细胞中的序列杂交,其中该复合物在自然界中不存在。

[0012] 在一些实施例中,本披露提供了一种非天然存在的CRISPR-Cas系统,该系统包含:能够产生粘性末端并包含核定位序列(NLS)的Cas9效应蛋白(stiCas9),以及与该stiCas9形成复合物并包含引导序列的引导多核苷酸,其中该复合物在自然界中不存在。

[0013] 在一些实施例中,本披露提供了一种非天然存在的CRISPR-Cas系统,该系统包含:编码能够产生粘性末端的Cas9效应蛋白(stiCas9)的一个或多个核苷酸序列,以及编码引导多核苷酸的核苷酸序列,该引导多核苷酸与该stiCas9形成复合物并包含引导序列,其中该引导序列与真核细胞中的靶序列杂交,但不与细菌细胞中的序列杂交,并且其中该复合物在自然界中不存在。

[0014] 在一些实施例中,本披露提供了一种非天然存在的CRISPR-Cas系统,该系统包含:(a) 编码能够产生粘性末端的Cas9效应蛋白(stiCas9)的一个或多个核苷酸序列,以及(b) 编码引导多核苷酸的核苷酸序列,该引导多核苷酸与该stiCas9形成复合物并包含引导序列,其中(a)和(b)中的核苷酸序列在真核启动子的控制下,并且其中该复合物在自然界中不存在。

[0015] 在一些实施例中,本披露的CRISPR-Cas系统进一步包含含有tracrRNA序列的多核苷酸。在一些实施例中,该CRISPR-Cas系统的引导多核苷酸、tracrRNA序列与stiCas9能够形成复合物,并且该复合物在自然界中不存在。

[0016] 在一些实施例中,本披露提供了一种非天然存在的CRISPR-Cas系统,该系统包含一个或多个载体,该一个或多个载体包含:可操作地连接至编码能够产生粘性末端的Cas9效应蛋白(stiCas9)的一个或多个核苷酸序列的调节元件,以及与该stiCas9形成复合物并且包含引导序列的引导多核苷酸,其中该引导序列与真核细胞中的靶序列杂交,但不与细菌细胞中的序列杂交,其中该复合物在自然界中不存在。

[0017] 在一些实施例中,本披露提供了一种非天然存在的CRISPR-Cas系统,该系统包含一个或多个载体,该一个或多个载体包含:可操作地连接至编码能够产生粘性末端的Cas9效应蛋白(stiCas9)的一个或多个核苷酸序列的调节元件(其中该调节元件是真核调节元件),以及与stiCas9形成复合物并包含引导序列的引导多核苷酸序列,其中该复合物在自然界中不存在。

[0018] 在一些实施例中,该引导多核苷酸进一步包含tracrRNA序列。在一些实施例中,本披露的非天然存在的载体还包含含有tracrRNA序列的核苷酸序列。

[0019] 在CRISPR-Cas系统的一些实施例中,该复合物能够在前间区序列邻近基序(PAM)的10个核苷酸内的位点处裂解。在CRISPR-Cas系统的一些实施例中,该复合物能够在前间区序列邻近基序(PAM)的5个核苷酸内的位点处裂解。在CRISPR-Cas系统的一些实施例中,该复合物能够在前间区序列邻近基序(PAM)的3个核苷酸内的位点处裂解。

[0020] 在CRISPR-Cas系统的一些实施例中,该靶序列是前间区序列邻近基序(PAM)的5',并且该PAM包含富含3' G的基序。在CRISPR-Cas系统的多个实施例中,该靶序列是前间区序列邻近基序(PAM)的5',并且该PAM序列是NGG,其中N是A、C、G或T。

[0021] 在CRISPR-Cas系统的一些实施例中,这些粘性末端包含具有3至40个核苷酸的单链多核苷酸突出端。在CRISPR-Cas系统的一些实施例中,这些粘性末端包含具有4至20个核苷酸的单链多核苷酸突出端。在CRISPR-Cas系统的一些实施例中,这些粘性末端包含具有5至10个核苷酸的单链多核苷酸突出端。

[0022] 在CRISPR-Cas系统的一些实施例中,该stiCas9源自具有II-B型CRISPR系统的细菌物种。在CRISPR-Cas系统的一些实施例中,该stiCas9包含与SEQ ID NO:10-97或192-195中的任一个具有至少80%同一性、85%同一性、90%同一性或95%同一性的结构域。在一些

实施例中,该stiCas9包含以 $1E-5$ 的E值截止值匹配TIGR03031蛋白家族的结构域。在一些实施例中,该stiCas9包含以 $1E-10$ 的E值截止值匹配TIGR03031蛋白家族的结构域。

[0023] 在CRISPR-Cas系统的一些实施例中,该stiCas9所源自的细菌物种是嗜肺军团菌(*Legionella pneumophila*)、新凶手弗朗西斯菌(*Francisella novicida*)、 γ 变形菌HTCC5015、人粪便副萨特氏菌(*Parasutterella excrementihominis*)、华德萨特氏菌(*Sutterella wadsworthensis*)、硫磺单胞菌属物种(*Sulfurospirillum* sp.) SCADC、瘤胃杆菌属物种(*Ruminobacter* sp.) RM87、伯克霍尔德里氏菌目(*Burkholderiales*)细菌1_1_47、拟杆菌门(*Bacteroidetes*)口腔分类群274菌株F0058、产琥珀酸沃廉氏菌(*Wolinella succinogenes*)、伯克霍尔德里氏菌目细菌YL45、嗜淀粉瘤胃杆菌(*Ruminobacter amylophilus*)、弯曲杆菌属物种(*Campylobacter* sp.) P0111、弯曲杆菌属物种RM9261、拉尼尔弯曲杆菌(*Campylobacter lanienae*)菌株RM8001、拉尼尔弯曲杆菌菌株P0121、鼠毛滴虫(*Turicimonas muris*)、伦敦军团菌(*Legionella londiniensis*)、沙姆盐弧菌(*Salinivibrio sharmensis*)、钩端螺旋体属物种(*Leptospira* sp.)分离株FW.030、莫里特拉氏菌属物种(*Moritella* sp.)分离株NORP46、内生单胞菌属物种(*Endozoicomonas* sp.) S-B4-1U、喜盐泰米尔纳德菌(*Tamilnaduibacter salinus*)、需钠弧菌(*Vibrio natriegens*)、斯氏弓形杆菌(*Arcobacter skirrowii*)、蜃楼弗朗西斯氏菌(*Francisella philomiragia*)、西班牙弗朗西斯氏菌(*Francisella hispaniensis*)或嗜盐副内生单胞菌(*Parendozaicomonas haliclona*)。

[0024] 在CRISPR-Cas系统的一些实施例中,靶序列是前间区序列邻近基序(PAM)的5',并且该PAM序列是YG,其中Y是嘧啶,并且该stiCas9源自细菌物种新凶手弗朗西斯菌(*F.novicida*)。

[0025] 在CRISPR-Cas系统的一些实施例中,该stiCas9包含一个或多个核定位信号。在CRISPR-Cas系统的一些实施例中,该真核细胞是动物或人类细胞。在CRISPR-Cas系统的一些实施例中,该真核细胞是人类细胞。在CRISPR-Cas系统的一些实施例中,该真核细胞是植物细胞。

[0026] 在CRISPR-Cas系统的一些实施例中,该引导序列与同向重复序列连接。

[0027] 在一些实施例中,递送颗粒包含本披露的CRISPR-Cas系统。在一些实施例中,该stiCas9和该引导多核苷酸在递送颗粒内以复合物存在。

[0028] 在一些实施例中,该引导多核苷酸进一步包含tracrRNA序列。在一些实施例中,该递送颗粒内的复合物还包含含有tracrRNA序列的多核苷酸。

[0029] 在一些实施例中,该递送颗粒还包含脂质、糖、金属或蛋白质。

[0030] 在一些实施例中,囊泡包含本披露的CRISPR-Cas系统。

[0031] 在一些实施例中,该stiCas9和该引导多核苷酸在囊泡内以复合物存在。

[0032] 在一些实施例中,该囊泡内的复合物还包含含有tracrRNA序列的多核苷酸。在一些实施例中,该囊泡是外泌体或脂质体。

[0033] 在CRISPR-Cas系统的一些实施例中,对该编码stiCas9的一个或多个核苷酸序列进行密码子优化以在真核细胞中表达。

[0034] 在CRISPR-Cas系统的一些实施例中,该编码Cas9效应蛋白的核苷酸和该引导多核苷酸在单个载体上。

[0035] 在CRISPR-Cas系统的一些实施例中,该编码Cas9效应蛋白的核苷酸和该引导多核苷酸是单个核酸分子。

[0036] 在一些实施例中,病毒载体包含本披露的CRISPR-Cas系统。在一些实施例中,该病毒载体是腺病毒、慢病毒或腺相关病毒载体。

[0037] 在一些实施例中,本披露提供了包含CRISPR-Cas系统的真核细胞,该CRISPR-Cas系统包含:能够产生粘性末端的Cas9效应蛋白(stiCas9),以及与该stiCas9形成复合物并且包含引导序列的引导多核苷酸,其中该引导序列能够与真核细胞中的靶序列杂交,其中该复合物在自然界中不存在。

[0038] 在一些实施例中,本披露提供了包含CRISPR-Cas系统的真核细胞,该CRISPR-Cas系统包含能够产生粘性末端的Cas9效应蛋白(stiCas9),其中该Cas9效应蛋白源自具有II-B型CRISPR系统的细菌物种。

[0039] 在一些实施例中,本披露提供了一种用于在真核细胞中提供靶序列的位点特异性修饰的方法,该方法包括:(1)将以下引入到细胞中:(a)能够产生粘性末端的Cas9效应蛋白(stiCas9),和(b)与stiCas9形成复合物并且包含引导序列的引导多核苷酸,其中该引导序列能够与真核细胞中的靶序列杂交,但不与细菌细胞中的序列杂交,其中该复合物在自然界中不存在;(2)用Cas9效应蛋白和引导多核苷酸在靶序列中产生粘性末端;以及(3)将(a)粘性末端连接在一起,或(b)将目的多核苷酸序列(SoI)连接到该粘性末端,从而修饰靶序列。

[0040] 在一些实施例中,本披露提供了一种用于在真核细胞中提供靶序列的位点特异性修饰的方法,该方法包括:(1)将以下引入到细胞中:(a)编码能够产生粘性末端的Cas9效应蛋白(stiCas9)的核苷酸序列,和(b)与stiCas9形成复合物并且包含引导序列的引导多核苷酸,其中该引导序列能够与真核细胞中的靶序列杂交,但不与细菌细胞中的序列杂交,其中该复合物在自然界中不存在;(2)用Cas9效应蛋白和引导多核苷酸在靶序列中产生粘性末端;以及(3):将(a)粘性末端连接在一起,或(b)将目的多核苷酸序列(SoI)连接到该粘性末端,从而修饰靶序列。

[0041] 在一些实施例中,用于在真核细胞中提供靶序列的位点特异性修饰的方法还包括将含有tracrRNA序列的多核苷酸引入细胞中。

[0042] 在该方法的一些实施例中,该引导多核苷酸、tracrRNA序列与stiCas9能够形成复合物,并且其中该复合物在自然界中不存在。

[0043] 在该方法的一些实施例中,该复合物能够在前间区序列邻近基序(PAM)的10个核苷酸内的位点处裂解。在该方法的一些实施例中,该复合物能够在前间区序列邻近基序(PAM)的5个核苷酸内的位点处裂解。在该方法的一些实施例中,该复合物能够在前间区序列邻近基序(PAM)的3个核苷酸内的位点处裂解。

[0044] 在该方法的一些实施例中,该靶序列是前间区序列邻近基序(PAM)的5',并且该PAM包含富含3' G的基序。在该方法的一些实施例中,该靶序列是PAM的5',并且该PAM序列是NGG,其中N是A、C、G或T。

[0045] 在该方法的一些实施例中,这些粘性末端包含具有3至40个核苷酸的单链多核苷酸突出端。在该方法的一些实施例中,这些粘性末端包含具有4至20个核苷酸的单链多核苷酸突出端。在该方法的一些实施例中,这些粘性末端包含具有5至10个核苷酸的单链多核苷酸突出端。

酸突出端。

[0046] 在该方法的一些实施例中,该stiCas9源自具有II-B型CRISPR系统的细菌物种。

[0047] 在该方法的一些实施例中,该真核细胞是动物或人类细胞。在该方法的一些实施例中,该真核细胞是人类细胞。在该方法的一些实施例中,该真核细胞是植物细胞。

[0048] 在该方法的一些实施例中,该修饰是靶序列的至少一部分的缺失。在该方法的多个实施例中,该修饰是靶序列的突变。在该方法的一些实施例中,该修饰是将目的序列插入靶序列。

[0049] 在一些实施例中,该方法进一步包括引入核酸外切酶以去除由stiCas9产生的突出端。

[0050] 在该方法的一些实施例中,该核酸外切酶是Cas4、Artemis或TREX4。在该方法的一些实施例中,该Cas4源自具有II-B型CRISPR系统的细菌物种。

[0051] 在该方法的一些实施例中,将编码该复合物的组分的多核苷酸引入一个或多个载体上。

[0052] 在一些实施例中,本披露涉及将目的序列(SoI)引入细胞的染色体中的方法,其中该染色体包含含有区域1和区域2的靶序列(TSC),该方法包括将以下引入细胞中:

[0053] (a) 包含靶序列的载体(TSV),该TSV包含区域2和区域1以及该SoI;

[0054] (b) 能够在TSC中产生粘性末端的第一Cas9-核酸内切酶二聚体,其中该第一Cas9-核酸内切酶二聚体的第一单体在该TSC的区域1处裂解,并且该第一Cas9-核酸内切酶二聚体的第二单体在该TSC的区域2处裂解;以及

[0055] (c) 能够在TSV中产生粘性末端的第二Cas9-核酸内切酶二聚体,其中该第二Cas9-核酸内切酶二聚体的第一单体在该TSV的区域2处裂解,并且该第二Cas9-核酸内切酶二聚体的第二单体在该TSV的区域1处裂解;

[0056] 其中引入(a)的载体、(b)的第一Cas9-核酸内切酶二聚体和(c)的第二Cas9-核酸内切酶二聚体导致SoI被插入细胞的染色体中。

[0057] 在一些实施例中,本披露涉及将目的序列(SoI)引入细胞的染色体中的方法,其中该染色体包含含有区域1和区域2的靶序列(TSC),该方法包括将以下引入细胞中:

[0058] (a) 包含靶序列的载体(TSV),该TSV包含区域2和区域1以及该SoI,其中该载体包含粘性末端;

[0059] (b) 能够在TSC中产生粘性末端的第一Cas9-核酸内切酶二聚体,其中该第一Cas9-核酸内切酶二聚体的第一单体在该TSC的区域1处裂解,并且该第一Cas9-核酸内切酶二聚体的第二单体在该TSC的区域2处裂解;

[0060] 其中引入(a)的载体和(b)的第一Cas9-核酸内切酶二聚体导致SoI被插入细胞的染色体中。

[0061] 在一些实施例中,第一和第二Cas9-核酸内切酶二聚体是相同的。在一些实施例中,第一和第二Cas9-核酸内切酶二聚体是不同的。

[0062] 在一些实施例中,该方法进一步包括将第一引导多核苷酸引入细胞中,该第一引导多核苷酸与第一Cas9-核酸内切酶二聚体的第一单体形成复合物并且包含第一引导序列,其中该第一引导序列与包含区域1的TSC杂交,但不与该载体杂交。

[0063] 在一些实施例中,该方法进一步包括将第一引导多核苷酸引入细胞中,该第一引

导多核苷酸与第一Cas9-核酸内切酶二聚体的第一单体形成复合物并包含第一引导序列，其中该第一引导序列与TSC和TSV杂交。

[0064] 在一些实施例中，该方法进一步包括将第二引导多核苷酸引入细胞中，该第二引导多核苷酸与第一Cas9-核酸内切酶二聚体的第二单体形成复合物并包含第二引导序列，其中该第二引导序列与包含区域2的TSC杂交，但不与载体杂交。

[0065] 在一些实施例中，该方法进一步包括将第二引导多核苷酸引入细胞中，该第二引导多核苷酸与第一Cas9-核酸内切酶二聚体的第二单体形成复合物并包含第二引导序列，其中该第二引导序列与TSC和TSV杂交。

[0066] 在一些实施例中，该方法进一步包括将第三引导多核苷酸引入细胞中，该第三引导多核苷酸与第二Cas9-核酸内切酶二聚体的第一单体形成复合物并包含第三引导序列，其中该第三引导序列与包含区域2的TSV杂交，但不与该染色体杂交。

[0067] 在一些实施例中，该方法进一步包括将第三引导多核苷酸引入细胞中，该第三引导多核苷酸与第二Cas9-核酸内切酶二聚体的第一单体形成复合物并包含第三引导序列，其中该第三引导序列与TSC和TSV杂交。

[0068] 在一些实施例中，该方法进一步包括将第四引导多核苷酸引入细胞中，该第四引导多核苷酸与第二Cas9-核酸内切酶二聚体的第二单体形成复合物并包含第四引导序列，其中该第四引导序列与包含区域1的TSV杂交，但不与染色体杂交。

[0069] 在一些实施例中，该方法进一步包括将第四引导多核苷酸引入细胞中，该第四引导多核苷酸与第二Cas9-核酸内切酶二聚体的第二单体形成复合物并包含第四引导序列，其中该第四引导序列与TSC和TSV杂交。

[0070] 在一些实施例中，该方法包括将第一、第二、第三和第四引导多核苷酸引入细胞中。

[0071] 在一些实施例中，该方法进一步包括将包含tracrRNA序列的多核苷酸引入细胞中。

[0072] 在一些实施例中，该第一Cas9-核酸内切酶二聚体的第一单体和第二单体中的核酸内切酶是IIS型核酸内切酶。在一些实施例中，该第二Cas9-核酸内切酶二聚体的第一单体和第二单体中的核酸内切酶是IIS型核酸内切酶。

[0073] 在一些实施例中，第一Cas9-核酸内切酶二聚体和第二Cas9-核酸内切酶二聚体中的核酸内切酶是IIS型核酸内切酶。在一些实施例中，第一Cas9-核酸内切酶二聚体和第二Cas9-核酸内切酶二聚体中的核酸内切酶独立地选自下组，该组由以下组成：BbvI、BgcI、BfuAI、BmpI、BspMI、CspCI、FokI、MboII、MmeI、NmeAIII和PleI。在一些实施例中，第一Cas9-核酸内切酶二聚体和第二Cas9-核酸内切酶二聚体中的核酸内切酶是FokI。在一些实施例中，将第一和第二Cas9-核酸内切酶二聚体作为编码第一和第二Cas9-核酸内切酶二聚体的多核苷酸引入细胞中。

[0074] 在一些实施例中，编码第一和第二Cas9-核酸内切酶二聚体的多核苷酸在一个载体上。在一些实施例中，编码第一和第二Cas9-核酸内切酶二聚体的多核苷酸在多于一个载体上。

[0075] 在一些实施例中，第一、第二Cas9-核酸内切酶二聚体或两者包含经修饰的Cas9。在一些实施例中，第一、第二Cas9-核酸内切酶二聚体或两者包含无催化活性的Cas9。在一

些实施例中,第一、第二Cas9-核酸内切酶二聚体或两者中的核酸内切酶是FokI。在一些实施例中,第一、第二Cas9-核酸内切酶二聚体或两者包含具有切口酶活性的Cas9。在一些实施例中,第一、第二Cas9-核酸内切酶二聚体或两者中的核酸内切酶是FokI。

[0076] 在一些实施例中,该Cas9-核酸内切酶二聚体相对于野生型Cas9在Cas9中包含单个氨基酸取代。在一些实施例中,第一、第二Cas9-核酸内切酶二聚体或两者中的核酸内切酶是FokI。在一些实施例中,该单个氨基酸取代是D10A或H840A。在一些实施例中,该单个氨基酸取代是D10A。在一些实施例中,该单个氨基酸取代是H840A。在一些实施例中,该Cas9-核酸内切酶二聚体相对于野生型Cas9包含双氨基酸取代。在一些实施例中,该双氨基酸取代是D10A和H840A。

[0077] 在一些实施例中,该野生型Cas9源自酿脓链球菌 (*Streptococcus pyogenes*)、金黄色葡萄球菌 (*Staphylococcus aureus*)、伪中间型葡萄球菌 (*Staphylococcus pseudintermedius*)、南极游动球菌 (*Planococcus antarcticus*)、血链球菌 (*Streptococcus sanguinis*)、嗜热链球菌 (*Streptococcus thermophilus*)、变形链球菌 (*Streptococcus mutans*)、球团科里氏杆菌 (*Coribacterium glomerans*)、香肠乳杆菌 (*Lactobacillus farciminis*)、光冈氏链小杆菌 (*Catenibacterium mitsuokai*)、鼠李糖乳杆菌 (*Lactobacillus rhamnosus*)、两歧双歧杆菌 (*Bifidobacterium bifidum*)、北原酒球菌 (*Oenococcus kitahara*)、果聚糖杆菌 (*Fructobacillus fructosus*)、大芬戈尔德菌 (*Fingoldia magna*)、非典型韦荣氏球菌 (*Veillonella atypica*)、穆氏索罗菌 (*Solobacterium moorei*)、氨基酸球菌属物种 (*Acidaminococcus* sp.) D21、尤里真杆菌 (*Eubacterium yurri*)、灵巧粪球菌 (*Coproccoccus catus*)、具核梭杆菌 (*Fusobacterium nucleatum*)、龈沟产线菌 (*Filifactor alocis*)、杜尔丹尼嗜脲菌 (*Peptoniphilus duerdenii*) 或齿垢密螺旋体 (*Treponema denticola*)。

[0078] 在一些实施例中,这些粘性末端包括5' 突出端。在一些实施例中,这些粘性末端包括3' 突出端。在一些实施例中,第一、第二Cas9-核酸内切酶二聚体或两者产生包含具有3至40个核苷酸的单链多核苷酸的粘性末端。在一些实施例中,第一、第二Cas9-核酸内切酶二聚体或两者产生包含具有4至20个核苷酸的单链多核苷酸的粘性末端。在一些实施例中,第一、第二Cas9-核酸内切酶二聚体或两者产生包含具有5至15个核苷酸的单链多核苷酸的粘性末端。

[0079] 在该方法的一些实施例中,在插入时,染色体中的靶序列和质粒中的靶序列没有被重建。

[0080] 在一些实施例中,该细胞是真核细胞。在一些实施例中,该细胞是动物或人类细胞。在一些实施例中,该细胞是植物细胞。

[0081] 在将目的序列 (SoI) 引入细胞的染色体中的方法的一些实施例中,将 (a) 的载体、(b) 的第一Cas9-核酸内切酶二聚体、(c) 的第二Cas9-核酸内切酶二聚体或其组合通过递送颗粒、囊泡或病毒载体引入细胞中。在一些实施例中,将 (a) 的载体、(b) 的第一Cas9-核酸内切酶二聚体、(c) 的第二Cas9-核酸内切酶二聚体或其组合通过递送颗粒引入细胞中。在一些实施例中,这些递送颗粒包含脂质、糖、金属或蛋白质。

[0082] 在将目的序列 (SoI) 引入细胞的染色体中的方法的一些实施例中,将 (a) 的载体、(b) 的第一Cas9-核酸内切酶二聚体、(c) 的第二Cas9-核酸内切酶二聚体或其组合通过囊泡

引入细胞中。在一些实施例中,这些囊泡是外泌体或脂质体。

[0083] 在将目的序列(SoI)引入细胞的染色体中的方法的一些实施方案中,将能够表达(a)的载体、(b)的第一Cas9-核酸内切酶二聚体、(c)的第二Cas9-核酸内切酶二聚体的多核苷酸或其组合通过病毒载体引入细胞中。在一些实施例中,(a)的载体是病毒载体。在一些实施例中,该病毒载体是腺病毒、慢病毒或腺相关病毒。

[0084] 在一些实施例中,第一Cas9-核酸内切酶二聚体的第一单体与第一引导多核苷酸形成复合物,并且第一Cas9-核酸内切酶二聚体的第二单体与第二引导多核苷酸形成复合物。在一些实施例中,第二Cas9-核酸内切酶二聚体的第一单体与第三引导多核苷酸形成复合物,并且第二Cas9-核酸内切酶二聚体的第二单体与第四引导多核苷酸形成复合物。在一些实施例中,第一Cas9-核酸内切酶二聚体的第一单体与第一引导多核苷酸序列和tracrRNA序列形成复合物,并且第一Cas9-核酸内切酶二聚体的第二单体与第二引导多核苷酸序列和tracrRNA序列形成复合物。在一些实施例中,第二Cas9-核酸内切酶二聚体的第一单体与第三引导多核苷酸序列和tracrRNA序列形成复合物,并且第二Cas9-核酸内切酶二聚体的第二单体与第四引导多核苷酸序列和tracrRNA序列形成复合物。在一些实施例中,第一、第二Cas9-核酸内切酶二聚体或两者包含核定位信号。

[0085] 在将目的序列(SoI)引入细胞的染色体中的方法的一些实施例中,该细胞包括干细胞或干细胞系。

[0086] 在一些实施例中,本披露涉及修饰细胞中靶多核苷酸序列中一个或多个核苷酸的方法,该方法包括:

[0087] (a) 将包含插入盒(IC)的载体引入细胞,该IC在5'至3'方向上包含

[0088] (i) 与靶多核苷酸序列的一部分同源的第一区域,

[0089] (ii) 第二区域,该区域包含靶多核苷酸序列中一个或多个核苷酸的突变,

[0090] (iii) 第一核酸酶结合位点,

[0091] (iv) 编码标记基因的多核苷酸序列,

[0092] (v) 第二核酸酶结合位点,

[0093] (vi) 第三区域,该区域包含靶多核苷酸序列中一个或多个核苷酸的突变,以及

[0094] (vii) 与靶多核苷酸序列的一部分同源的第四区域,其中第一区域和第四区域与靶多核苷酸序列95%-100%相同;

[0095] (b) 通过同源重组将IC插入靶多核苷酸序列以产生第一经修饰的靶多核苷酸;

[0096] (c) 选择表达该标记基因的细胞;

[0097] (d) 使该第一经修饰的靶多核苷酸经受位点特异性核酸酶处理以产生具有粘性末端的第二经修饰的靶多核苷酸;以及

[0098] (e) 使该具有粘性末端的第二经修饰的靶多核苷酸经受连接酶处理,其中该连接酶在该第二区域和该第三区域处连接这些粘性末端以产生连接的经修饰的靶核酸,当与该靶多核苷酸序列相比时,该连接的经修饰的靶核酸包含一个或多个经修饰的核苷酸。

[0099] 在修饰细胞中的靶多核苷酸序列中一个或多个核苷酸的方法的一些实施例中,在(c)之后,从细胞中分离出第一经修饰的靶核酸。

[0100] 在一些实施例中,该位点特异性核酸酶对于该细胞而言是外源性的。在一些实施例中,该连接酶对于该细胞而言是外源性的。在一些实施例中,在(c)之后,该第一经修饰的

靶蛋白在该细胞中。在一些实施例中,将该位点特异性核酸酶作为编码位点特异性核酸酶的多核苷酸引入细胞中。在一些实施例中,将该连接酶作为编码连接酶的多核苷酸引入细胞中。

[0101] 在一些实施例中,该位点特异性核酸酶是重组位点特异性核酸酶。在一些实施例中,该连接酶是重组连接酶。在一些实施例中,该位点特异性核酸酶是Cas9效应蛋白。在一些实施例中,该Cas9效应蛋白是II-B型Cas9。在一些实施例中,该位点特异性核酸酶是Cas9-核酸内切酶融合蛋白。在一些实施例中,该Cas9-核酸内切酶融合蛋白中的核酸内切酶是IIS型核酸内切酶。在一些实施例中,该Cas9-核酸内切酶融合蛋白中的核酸内切酶是FokI。

[0102] 在一些实施例中,该Cas9-核酸内切酶融合蛋白包含经修饰的Cas9。在一些实施例中,该经修饰的Cas9包含无催化活性的Cas9。在一些实施例中,该无催化活性的Cas9与FokI核酸内切酶融合。

[0103] 在一些实施例中,该Cas9-核酸内切酶融合蛋白包含具有切口酶活性的Cas9,并且该核酸内切酶是FokI。在一些实施例中,该Cas9-核酸内切酶融合蛋白包含具有D10A取代的Cas9。在一些实施例中,该Cas9-核酸内切酶融合蛋白包含具有H840A取代的Cas9。

[0104] 在一些实施例中,该位点特异性核酸酶是Cpf1效应蛋白。在一些实施例中,该位点特异性核酸酶是Cas9、Cpf1或Cas9-FokI。

[0105] 在修饰细胞中的靶多核苷酸序列中一个或多个核苷酸的方法的一些实施例中,(d)的第二经修饰的靶多核苷酸的粘性末端包含5'突出端。在一些实施例中,(d)的第二经修饰的靶多核苷酸的粘性末端包含3'突出端。在一些实施例中,该位点特异性核酸酶能够产生包含具有3至40个核苷酸的单链多核苷酸的粘性末端。在一些实施例中,该核酸酶能够产生包含具有4至20个核苷酸的单链多核苷酸的粘性末端。在一些实施例中,该核酸酶能够产生包含具有5至15个核苷酸的单链多核苷酸的粘性末端。

[0106] 在修饰细胞中的靶多核苷酸序列中的一个或多个核苷酸的方法的一些实施例中,该靶多核苷酸序列处于质粒中。在一些实施例中,该靶多核苷酸序列处于染色体中。

[0107] 在一些实施例中,本披露涉及与stiCas9蛋白形成复合物的工程化引导RNA,该RNA包含:(a)能够与真核细胞中的靶序列杂交的引导序列;以及(b)能够与Cas9蛋白结合的tracrRNA序列,其中该tracrRNA与天然存在的tracrRNA序列的差异为至少10个核苷酸,其中该工程化引导RNA提高了该Cas9蛋白的核酸酶效率。在一些实施例中,该tracrRNA序列比天然存在的tracrRNA少至少10个核苷酸。在一些实施例中,该tracrRNA序列比天然存在的tracrRNA多至少10个核苷酸。在一些实施例中,该引导序列与SEQ ID NO:104-125或196-199中的任一个具有至少90%的序列同一性。在一些实施例中,该tracrRNA序列与SEQ ID NO:148-171中的任一个具有至少90%的序列同一性。在一些实施例中,该引导RNA具有与SEQ ID NO:172-191中的任一个至少90%的序列同一性。

[0108] 在一些实施例中,本披露涉及包含如本文所述的工程化引导RNA的CRISPR-Cas系统。在一些实施例中,该系统不包含tracrRNA序列。

[0109] 在一些实施例中,本披露涉及工程化的Cas9-引导RNA复合物,该复合物包含如图40B中所示的Cas9、引导序列和tracrRNA序列的任何组合。在一些实施例中,本披露涉及产生与Cas9蛋白结合的工程化引导RNA的方法,包括:(a)提供能够与真核细胞中的靶序列杂

交的引导序列；(b) 将天然存在的tracrRNA序列通过从该tracrRNA序列中除去至少十个核苷酸来进行修饰，以形成经修饰的tracrRNA序列；以及(c) 将引导序列与经修饰的tracrRNA序列连接以产生工程化引导RNA。在一些实施例中，本披露涉及一种非天然存在的CRISPR-Cas系统，该系统包含：(a) 能够产生粘性末端的Cas9效应蛋白(stiCas9)；以及(b) 与该stiCas9形成复合物并且包含引导序列的引导RNA，其中该引导序列能够与真核细胞中的靶序列杂交，但不与细菌细胞中的序列杂交；其中该复合物在自然界中不存在，并且其中该系统不包含tracrRNA序列。

附图说明

[0110] 图1是通过Cas9进行的修复的不同机制的示意图。图1a描绘基因敲除。图1b描绘碱基编辑。图1c描绘通过非同源末端连接(NHEJ)途径进行的基因敲入。图1d描绘通过同源重组(HDR)途径进行的基因敲入。

[0111] 图2是通过Cas9进行的基因插入的不同机制的示意图。左侧显示同源重组(HDR)。右侧显示非同源末端连接(NHEJ)。

[0112] 图3是使用不同的Cas9效应蛋白进行基因插入的结果的示意图和描绘。图3a-b显示了由Cas9介导的产生平端的基因插入。图3c-d显示了由Cas9介导的产生突出端(即“粘性末端”)的基因插入。图3的下图描绘了使用同源非依赖性靶向插入(HITI)，通过3a-3f中的不同Cas9蛋白实现的基因插入频率。

[0113] 图4在Shmakov等人,Nature Reviews Microbiology[自然微生物学综述]15:169-182(2017)中描述过。图4A是不同类型的CRISPR系统和具有每种类型的CRISPR系统的代表性细菌物种的系统发育树。图4B显示II型和V型CRISPR系统的特写，其中箭头指示包含cas4基因的操纵子。

[0114] 图5在Chylinski等人,Nucleic Acids Research[核酸研究]42(10):6091-6105(2014)中描述过。图5A-D描绘了II型CRISPR系统的系统发育树。图5E显示了与II型CRISPR系统的每个亚家族相关的不同签名基因。

[0115] 图6A描绘了使用来自新凶手弗朗西斯菌(*Francisella novicida*)的Cas9蛋白进行DNA裂解获得的结果。比较了用来自新凶手弗朗西斯菌(*Francisella novicida*)的Cas9和来自酿脓链球菌(*Streptococcus pyogenes*)的Cas9靶向的工程化HEK293细胞系中基因组基因座的突变签名。图6A按出现顺序分别披露了SEQ ID NO 204-205和284。图6B-C是II型CRISPR系统的系统发育树。选择用于体外确认的Cas9蛋白用斜体指示。

[0116] 图7是如美国专利号9,567,608中所述的使用锌指核酸酶(ZFN)的用于基因插入的ObLiGaRe方法的示意图。

[0117] 图8是如Sakuma等人,Nature Protocols[自然实验手册]11(1):118-133(2016)所述的用于基因插入的Cas9-PiTCH方法的示意图。

[0118] 图9是三种不同的Cas9-FokI融合蛋白的示意图。图9a:失灭酶活性的Cas9(deadCas9)与FokI的融合；图9b:具有D10A突变的Cas9(Cas9n^{D10A})与FokI的融合；图9c:具有H840A的Cas9与(Cas9n^{H840A})与FokI的融合。图9a-c披露了SEQ ID NO:206。

[0119] 图10是由图9和10中的不同Cas9-FokI融合蛋白产生的不同DNA断裂的示意图。图10按出现顺序分别将SEQ ID NO:206披露为“TCCCCTCCACCCCACAGTGGGGCCACTAGGGACAGGAT

TGGTGACAGAAAAGCCCCATCCTTAGGCCT”以及将裂解的序列披露为SEQ ID NO 285-289。

[0120] 图11是由Cas9n^{D10A}-FokI产生的裂解位点的示意图。图11披露了SEQ ID NO:206。

[0121] 图12是使用Cas9n^{D10A}-FokI的基因插入方法的示意图。gRNA: 引导RNA; PAM: 前间区序列邻近基序。图12均按出现顺序分别披露了SEQ ID NO:206-208的“基因组”序列, SEQ ID NO:209-211的“载体”序列和SEQ ID NO:212的“敲入”序列。

[0122] 图13是由Cas9n^{H840A}-FokI产生的裂解位点的示意图。图13披露了SEQ ID NO:206。

[0123] 图14是使用Cas9n^{H840A}-FokI的基因插入方法的示意图。gRNA: 引导RNA; PAM: 前间区序列邻近基序。图14均按出现顺序分别披露了SEQ ID NO:206和213-214的“基因组”序列, SEQ ID NO:215-217的“载体”序列和SEQ ID NO:218的“敲入”序列。

[0124] 图15-18涉及实例1中阐述的实验。

[0125] 图15是使用Cas9n^{D10A}-FokI (图15) 和Cas9n^{H840A}-FokI (图15) 的基因插入方法的示意图。图15a-b披露了SEQ ID NO:206。

[0126] 图16描绘了靶位点(AAVS1基因座)。“计划A”是指使用Cas9n^{D10A}-FokI的基因插入方法;“计划B”是指使用Cas9n^{H840A}-FokI的基因插入方法。图16披露了SEQ ID NO:219。

[0127] 图17显示了使用Cas9n^{D10A}-FokI的基因插入方法产生的代表性所得序列。图17按出现顺序分别披露了SEQ ID NO220-235。

[0128] 图18显示了使用Cas9n^{H840A}-FokI的基因插入方法产生的代表性所得序列。图18按出现顺序分别披露了SEQ ID NO236-258。

[0129] 图19-22涉及实例2中阐述的实验。

[0130] 图19显示了用于靶向AAVS1基因座的一组10个引导RNA (gRNA) 的设计。

[0131] 图20是“供体”质粒的质粒图,该质粒包含使用图20中的多个gRNA插入AAVS1基因座的基因。

[0132] 图21是选择包含正确插入的基因的细胞(mCherry+细胞)的程序的示意图。

[0133] 图22显示了使用不同长度间区序列的基因插入频率的结果。

[0134] 图23-24涉及实例3中阐述的实验。

[0135] 图23是含有待插入SERPINA1基因座中的基因的“供体”质粒的质粒图。

[0136] 图24是使用deadCas9-FokI的基因插入方法的示意图。图24披露了SEQ ID NO:206。

[0137] 图25是如实例2-4中所述的用于靶向基因插入的不同方法的效率的比较。

[0138] 图26-29涉及实例4中阐述的实验。

[0139] 图26是无缝诱变的示意图。

[0140] 图27是无缝诱变的第一步的示意图:使用同源臂将含有抗性标记的盒重组为靶序列。

[0141] 图28是整合到靶序列中的盒的示意图:两侧是核酸酶结合位点和核酸酶切割位点的抗性标记。

[0142] 图29是无缝诱变的第二步骤的示意图:在切割位点处核酸酶消化(图28所示)并随后连接,导致抗性标记的去除和无缝产生的突变。

[0143] 图30包括来自各种测序细菌的Cas9蛋白的氨基酸序列,这些细菌包括:嗜肺军团菌、新凶手弗朗西斯菌、 γ 变形菌HTCC5015、人粪便副萨特氏菌、华德萨特氏菌、硫磺单胞菌

属物种SCADC、瘤胃杆菌属物种RM87、伯克霍尔德氏菌目细菌1_1_47、拟杆菌门口腔分类群274菌株F0058和产琥珀酸沃廉氏菌。(SEQ ID NOS:10-80)

[0144] 图31包括来自各种测序细菌的Cas9蛋白的氨基酸序列,这些细菌包括:伯克霍尔德氏菌目细菌、弯曲杆菌属物种、鼠毛滴虫、沙姆盐弧菌、钩端螺旋体属物种、莫里特拉氏菌属物种、内生单胞菌属物种、喜盐泰米尔纳德菌、需钠弧菌、嗜淀粉瘤胃杆菌、萨迦弧菌(*Vibrio sagaiensis*)、豕弓形杆菌(*Arcobacter porcimus*)、脱硫杆菌属物种(*Desulfotomaculum* sp.)、琥珀酸单胞菌属物种(*Succinatimonas* sp.) (SEQ ID NO:81-97)。

[0145] 图32包括在来自MH0245_GL0161830_1的Cas9蛋白上具有在实例8中阐述的实验中使用的引导RNA序列、tracrRNA序列和crRNA序列的核苷酸序列(SEQ ID NO:101-103)。

[0146] 图33A显示了由II-B型Cas9蛋白产生的示例性4-核苷酸5'突出端。图33A披露了SEQ ID NO:259。图33B显示了示例性II-B型cas操纵子。cas9、cas1、cas2和cas4基因用箭头表示。CRISPR阵列标记在该操纵子的下游。

[0147] 图34涉及实例7中阐述的实验。图34A显示了电泳凝胶图像,该电泳凝胶图像证明了来自新凶手弗朗西斯菌的Cas9蛋白(FnCas9)的体外核酸酶活性。图34B显示了桑格(Sanger)测序图,该测序图指示FnCas9产生具有5'突出端的粘性末端。图34B按出现顺序分别披露了SEQ ID NO 204-205和284。图34C显示了酿脓链球菌Cas9蛋白(SPyCas9)和FnCas9之间的突变模式的RIMA比较。

[0148] 图35-36涉及实例8中阐述的实验。

[0149] 图35A显示了电泳凝胶图像,该电泳凝胶图像证明了来自序列肠道宏基因组MH0245的Cas9蛋白(MHCas9)的体外核酸酶活性。图35B显示了桑格测序图,该测序图指示MHCas9产生具有5'突出端的粘性末端。图35B按出现顺序分别披露了SEQ ID NO 260-262。图35C显示了电泳凝胶图像,该电泳凝胶图像证明了HEK293-REMINDEL细胞中通过Ce111测定确认的MHCas9活性。

[0150] 图36A显示了来自MHCas9的crRNA和tracrRNA的序列。图36A披露了SEQ ID NO:263。图36B显示了该crRNA/tracrRNA二级结构的示意图。图36C显示来自硫磺单胞菌属物种SCADC的Cas9蛋白(ssCas9)、来自产琥珀酸沃廉氏菌的Cas蛋白(WsCas9)、来自嗜肺军团菌的Cas9蛋白(LpCas9)、来自新凶手弗朗西斯菌的Cas9蛋白(FnCas9)以及来自MH0245的Cas9蛋白(MHCas9)的截短的系统发育树。

[0151] 图37是由来自如本文所述的各种细菌物种的Cas9蛋白的氨基酸序列产生的系统发育树。使用MUSCLE算法、CLC基因组学工作台v.9进行了序列比对。

[0152] 图38是由来自弯曲杆菌属的各种细菌物种的Cas9蛋白的氨基酸序列产生的系统发育树。使用MUSCLE算法、CLC基因组学工作台v.9进行了序列比对。

[0153] 图39包括本文所述的各种Cas9蛋白的crRNA的核苷酸序列(SEQ ID NO:104-147)

[0154] 图40A包括本文所述的各种Cas9蛋白的tracrRNA的核苷酸序列(SEQ ID NO:148-171)。

[0155] 图40B包括Cas9蛋白、crRNA(+)、crRNA(-)和tracrRNA的各种组合。

[0156] 图41A-T示出了通过实例9中描述的方法设计的各种sgRNA(也称为“嵌合gRNA”),包括sgRNA的序列(SEQ ID NO:172-191)。图41A还披露了如SEQ ID NO:264的发夹序列。

[0157] 图42A-L示出了实例9中描述的sgRNA的优化和修剪,以及用于进一步修饰的可能

靶位点。图42A按出现顺序分别披露了SEQ ID NO 265-266。图42B按出现顺序分别披露了SEQ ID NO 267-268。图42C按出现顺序分别披露了SEQ ID NO 269和173。图42D按出现顺序分别披露了SEQ ID NO 270-271。图42E按出现顺序分别披露了SEQ ID NO 178和272。图42F按出现顺序分别披露了SEQ ID NO 179和273。图42G按出现顺序分别披露了SEQ ID NO 180和274。图42H按出现顺序分别披露了SEQ ID NO 176和275。图42I按出现顺序分别披露了SEQ ID NO 174和276。图42J按出现顺序分别披露了SEQ ID NO 191和277。图42K按出现顺序分别披露了SEQ ID NO 184和278。图42L按出现顺序分别披露了SEQ ID NO 279-280。

[0158] 图43示出II-B型CRISPR-Cas系统的双向表达构建体。如插图所示,顶部链表达不包含tracrRNA的单-引导RNA的crRNA和间区序列。底部链表达包含tracrRNA的双-引导RNA的crRNA和间区序列。图43按出现顺序分别披露了SEQ ID NO 137、281和191。

[0159] 图44显示了本文所述的Cas9蛋白的单-引导RNA支架的预测二级结构。图44按出现顺序分别披露了SEQ ID NO 137、139、282、122、110、129、120、124和104。

[0160] 图45概括地描述了四个不同的工程化RNA,以及用MHCas9对每个的切割效率。

[0161] 图46证明了用三种不同的Cas9系统SpyCas9、C11Cas9和MHCas9对长度19、20、21、22和23的引导RNA的切割效率和功能性。

[0162] 图47包括来自各种测序细菌的Cas9蛋白的氨基酸序列,这些细菌包括:斯氏弓形杆菌、蜃楼弗朗西斯氏菌、西班牙弗朗西斯氏菌和嗜盐副内生单胞菌(SEQ ID NO:192-195)。

[0163] 图48包括本文所述的各种Cas9蛋白的crRNA的核苷酸序列(SEQ ID NO:196-203)。

[0164] 图49涉及实例11。图49A显示了确定Cas9蛋白的PAM序列的示例性方法。图49A披露了SEQ ID NO:283。图49B显示了通过图49A所示方法确定的SpCas9(顶部)和MHCas9(底部)的优选PAM序列。

[0165] 图50和51涉及实例12。

[0166] 图50A显示了精确修复的Cas9切割的示意图。图50B显示了Cas9切割的示意图,加上诸如TREX2或Artemis的核酸外切酶的末端加工,导致修复不精确和修饰增加。

[0167] 图51A显示了用于测试在三种不同的引导RNA的情况下将末端加工酶(FnCas4或TREX2)添加至各种Cas9(SpCas9、FnCas9、C11Cas9或MHCas9)的效果的方法的概览。图51B显示了每种Cas9蛋白与模拟末端加工酶FnCas4或TREX2以及在三种引导RNA的情况下的结果。

[0168] 图52和53涉及实例13。

[0169] 图52A、52B和52C分别显示了由SpCas9、C11Cas9或MHCas9在当所有三种Cas9蛋白按相同序列切割时产生的不同类型的突变。图52A-C披露了SEQ ID NO:290。

[0170] 图53A显示了切割与引导RNA复合的双链DNA序列的II-A型Cas9蛋白的RuvC和HNH结构域的示意图,该切割产生平端或单核苷酸突出端。图53B显示了切割与引导RNA复合的双链DNA序列的II-B型Cas9蛋白的RuvC和HNH结构域的示意图,该切割产生带有3-或4-核苷酸突出端的黏性末端。

具体实施方式

[0171] CRISPR-Cas9系统由于能够形成靶向的双链断裂而被广泛用于基因编辑。已知Cas9蛋白在裂解后会产生平端,这些平端与插入和/或修饰靶序列的粘性末端相比,特异性

较低。本文描述了能够产生粘性末端的Cas9蛋白,也称为stiCas9。本文描述了使用stiCas9蛋白用于插入和/或修饰靶序列的优点。

[0172] 本披露提供了非天然存在的CRISPR-Cas系统;包含CRISPR-Cas系统的真核细胞;提供靶序列的位点特异性修饰的方法;将目的序列引入细胞的染色体的方法;以及修饰细胞中的靶多核苷酸序列中一个或多个核苷酸的方法。

[0173] 定义

[0174] 如本文所用,“一个/一种(a或an)”可以意指一个/种或多个/种。如本文说明书和一项或多项权利要求中所用,当与单词“包括”结合使用时,单词“一个/种”可以意指一个/种或多于一个/种。如本文所用,“另一个/种”可以意指至少第二个/种或更多个/种。

[0175] 在整个本申请中,术语“约”用于指示值包括被采用以确定值的方法/装置的误差的固有变化,或者研究受试者之间存在的变化。典型地,该术语意指涵盖近似于或小于1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%或20%变化,这取决于具体情况。

[0176] 权利要求中使用术语“或”用于意指“和/或”,除非明确指示仅指替代方案或替代方案是相互排斥的,尽管本披露支持仅是指替代方案和“和/或”的定义。

[0177] 如在本说明书和一项和多项权利要求中所使用的,词语“包含”(和任何形式的包含,诸如“包含有”和“包含着”),“具有”(和任何形式的具有,诸如“具有着”和“有”),“包括”(以及任何形式的包括,诸如“包括有”和“包括着”)或“含有”(以及任何形式的含有,诸如“含有着”和“含”)是包扩性的或开放性的,并且不排除另外的未列举的元素或方法步骤。考虑到本说明书中讨论的任何实施例可以相对于本披露的任何方法、系统、宿主细胞、表达载体和/或组合物实施。此外,本披露的组合物、系统、宿主细胞和/或载体可用于实现本披露的方法和蛋白质。

[0178] 术语“例如(for example)”及其对应的缩写“例如(e.g.)”(无论是否斜体)的使用意味着所列举的特定术语是本披露的代表性实例和实施例,并不旨在限于所引用或列举的特定实例,除非另有明确说明。

[0179] “核酸”、“核酸分子”、“核苷酸”、“核苷酸序列”、“寡核苷酸”或“多核苷酸”意指包含共价连接的核苷酸的聚合化合物。术语“核酸”包括核糖核酸(RNA)和脱氧核糖核酸(DNA),两者都可以是单链或双链的。DNA包括但不限于互补DNA(cDNA)、基因组DNA、质粒或载体DNA和合成DNA。在一些实施例中,本披露提供了编码本文披露的任何一种多肽的多核苷酸,例如,本披露涉及编码Cas蛋白或其变体的多核苷酸。

[0180] “基因”是指编码多肽的核苷酸的组装,并且包括cDNA和基因组DNA核酸分子。“基因”还指可以充当编码序列之前(5'非编码序列)和之后(3'非编码序列)的调节序列的核酸片段。

[0181] 当单链形式的核酸分子可在适合的温度和溶液离子强度条件下退火至另一核酸分子上时,核酸分子与该另一核酸分子(诸如cDNA、基因组DNA或RNA)“可杂交”或“杂交”。杂交和洗涤条件是熟知的,并在Sambrook等人,Molecular Cloning:A Laboratory Manual, Second Edition[分子克隆:实验室手册,第二版],冷泉港实验室出版社,冷泉港(1989),特别是第11章和其中的表11.1(通过引用全文并入本文)中进行了例示。温度和离子强度条件决定了杂交的“严格性”。可以调节严格性条件以筛选中度相似的片段(诸如来自远亲关系

生物体的同源序列)到高度相似的片段(诸如复制来自近亲关系生物体的功能酶的基因)。为了初步筛选同源核酸,可以使用对应于55℃的 T_m 的低严格性杂交条件,例如,5XSSC、0.1%SDS、0.25%牛奶和不含甲酰胺;或30%甲酰胺,5XSSC,0.5%SDS。中等严格性杂交条件对应于较高的 T_m ,例如,40%甲酰胺和5X或6XSSC。高严格性杂交条件对应于最高的 T_m ,例如,50%甲酰胺,5X或6XSSC。杂交需要两个核酸包含互补序列,尽管取决于杂交的严格性,碱基之间可能存在错配。

[0182] 术语“互补的”用于描述能够彼此杂交的核苷酸碱基之间的关系。例如,对于DNA,腺苷与胸腺嘧啶互补,而胞嘧啶与鸟嘌呤互补。因此,本披露还包括与本文披露或使用的完整序列互补的分离核酸片段以及那些基本上相似的核酸序列。

[0183] DNA“编码序列”是双链DNA序列,当置于适当调节序列的控制下时,该双链DNA序列在体外细胞中或体内被转录并翻译成多肽。“合适的调节序列”是指位于编码序列上游(5'非编码序列)、内部或下游(3'非编码序列)的核苷酸序列,并且该核苷酸序列影响转录、RNA加工或稳定性或相关编码序列的翻译。调节序列可包括启动子、翻译前导序列、内含子、聚腺苷酸化识别序列、RNA加工位点、效应子结合位点和茎环结构。该编码序列的边界由5'(氨基)末端处的起始密码子和3'(羧基)末端处的翻译终止密码子确定。编码序列可以包括但不限于原核序列、来自mRNA的cDNA、基因组DNA序列,甚至是合成DNA序列。如果编码序列旨在真核细胞中表达,则多腺苷酸化信号和转录终止序列通常存在于该编码序列的3'端。

[0184] “开放阅读框”的缩写为ORF,意指包含翻译起始信号或起始密码子(诸如ATG或AUG)以及终止密码子并且可能被翻译成多肽序列的一段核酸序列(DNA、cDNA或RNA)。

[0185] 术语“同源重组”是指将外来DNA序列插入到另一个DNA分子中,例如,将载体插入染色体中。优选地,该载体靶向特定的染色体位点以进行同源重组。对于特定的同源重组,该载体将包含与染色体序列具有同源性的足够长的区域,以允许载体与染色体的互补结合和将载体掺入染色体。更长的同源性区域和更大程度的序列相似性可提高同源重组的效率。

[0186] 根据本文的披露内容,可以使用本领域已知的方法来扩增多核苷酸。一旦建立了合适的宿主系统和生长条件,就可以大量扩增和制备重组表达载体。如本文所述,可以使用的表达载体包括但不限于以下载体或其衍生物:人类或动物病毒,诸如牛痘病毒或腺病毒;昆虫病毒,诸如杆状病毒;酵母载体;噬菌体载体(例如, λ),以及质粒和粘粒DNA载体。

[0187] 如本文所用,“启动子”,“启动子序列”或“启动子区”是指能够结合RNA聚合酶并涉及起始下游编码或非编码序列的转录的DNA调节区/序列。在本披露的一些实例中,启动子序列包括转录起始位点并向上游延伸以包括以高于背景可检测水平起始转录所使用的最少数目的碱基或元件。在一些实施例中,启动子序列包括转录起始位点,以及负责RNA聚合酶结合的蛋白质结合结构域。真核启动子通常但并非总是包含多个“TATA”盒和“CAT”盒。各种启动子,包括诱导型启动子,可用于驱动本披露的各种载体。

[0188] “载体”是用于将核酸克隆和/或转移至宿主细胞中的任何工具。载体可以是可能与另一个DNA区段附接的复制子,这样以便产生附接区段的复制。“复制子”是任意一种遗传因子(例如,质粒、噬菌体、粘粒、染色体、病毒),它充当DNA体内复制的自动单元,即在其自我控制下能复制。在本披露的一些实施例中,载体是附加型载体,该载体在许多个细胞世代后,通过例如不对称分配,从细胞群中去除/丢失。术语“载体”包括用于在体外、离体或体内

将该核酸引入细胞中的病毒性和非病毒工具。可以使用本领域熟知的大量载体,以便操作核酸,将响应元件和启动子整合至基因中等等。可能的载体包括例如质粒或经修饰的病毒,包括例如噬菌体诸如 λ 衍生物,或质粒诸如PBR322或pUC质粒衍生物,或Bluescript载体。例如,将对应于响应元件和启动子的DNA片段插入合适的载体中能伴随着将合适的DNA片段连接至具有互补结合端的选定载体。可替代地,DNA分子的末端可以被酶催化修饰或者任意位点通过将核苷酸序列(接头)连接至该DNA末端中而产生。此类载体可进行工程化处理,以包含提供对细胞进行选择的选择性标记基因,这些细胞将标记掺入到细胞基因组中。此类标记允许鉴定和/或选择宿主细胞,这些宿主细胞掺入和表达该标记所编码的蛋白质。

[0189] 病毒载体,特别是逆转录病毒载体,已经用于细胞以及活体动物的许多种基因递送应用中。可以使用的病毒载体包括但不限于逆转录病毒、腺相关病毒、痘病毒、杆状病毒、牛痘、单纯性疱疹、艾巴氏病毒、腺病毒、双生病毒和花椰菜花叶病毒载体。非病毒载体包括但不限于质粒、脂质体、带电脂质(细胞转染素)、DNA-蛋白质复合物和生物聚合物。除核酸外,载体还可包含一个或多个调节区和/或用于选择、测量和监测核酸转移结果(转移至哪个组织、表达持续时间等)的选择性标记。

[0190] 可以通过熟知的方法将载体引入希望的宿主细胞中,这些方法包括但不限于转染、转导、细胞融合和脂质转染。载体可包含各种调节元件,包括启动子。在一些实施例中,载体设计可以基于由Mali等人“Cas9 as a versatile tool for engineering biology [Cas9作为工程生物学的多用途工具]”,Nature Methods[自然方法]10:957-63(2013)设计的多个构建体。在一些实施例中,本披露提供了包含本文所述的任何多核苷酸的表达载体,例如,包含编码Cas蛋白或其变体的多核苷酸的表达载体。在一些实施例中,本披露提供了包含编码Cas9蛋白或其变体的多核苷酸的表达载体。

[0191] 术语“质粒”是指额外的染色体元件,该元件通常携带不参与细胞的中央代谢的基因,并且通常呈环状双链DNA分子的形式。此类元件可以是来源于任何来源的单链或双链DNA或RNA的线性、环状或超螺旋自主复制序列、基因组整合序列、噬菌体或核苷酸序列,其中许多核苷酸序列已连接或重组到独特结构中,该独特结构能够将针对选定基因产物的启动子片段和DNA序列连同适当的3'未翻译序列引入到细胞中。

[0192] 如本文所用,“转染”意指将外源核酸分子(包括载体)引入细胞。“转染的”细胞在细胞内部包含外源核酸分子,而“转化的”细胞是其中细胞内的外源核酸分子诱导细胞表型变化的细胞。转染的核酸分子可以整合到宿主细胞的基因组DNA中和/或可以被细胞暂时或长时间地维持在染色体外。表达外源核酸分子或片段的宿主细胞或生物体被称为“重组”、“转化”或“转基因”生物体。在一些实施例中,本披露提供了包含本文所述的任何表达载体(例如,包含编码Cas蛋白或其变体的多核苷酸的表达载体)的宿主细胞。在一些实施例中,本披露提供了包含表达载体的宿主细胞,该表达载体包含编码Cas9蛋白或其变体的多核苷酸。

[0193] 术语“肽”、“多肽”和“蛋白质”在本文中可互换使用,是指任何长度的氨基酸的聚合形式,这些聚合形式可以包括编码和非编码的氨基酸、经化学或生物化学修饰或衍生的氨基酸和具有经修饰的肽骨架的多肽。

[0194] 如本文所用,“氨基酸”是指包含羧基(-COOH)和氨基(-NH₂)的化合物。“氨基酸”是指天然和非天然(即合成)氨基酸。天然氨基酸及其三个字母和一个字母的缩写包括:丙氨

酸(Ala;A);精氨酸(Arg,R);天冬酰胺(Asn;N);天冬氨酸(Asp;D);半胱氨酸(Cys;C);谷氨酰胺(Gln;Q);谷氨酸(Glu;E);甘氨酸(Gly;G);组氨酸(His;H);异亮氨酸(Ile;I);亮氨酸(Leu;L);赖氨酸(Lys;K);蛋氨酸(Met;M);苯丙氨酸(Phe;F);脯氨酸(Pro;P);丝氨酸(Ser;S);苏氨酸(Thr;T);色氨酸(Trp;W);酪氨酸(Tyr;Y);以及缬氨酸(Val;V)。

[0195] “氨基酸取代”是指包含一个或多个野生型或天然存在的氨基酸被相对于该野生型或天然存在的氨基酸不同的氨基酸在该氨基酸残基处取代的多肽或蛋白质。被取代的氨基酸可以是合成或天然存在的氨基酸。在一些实施例中,被取代的氨基酸是选自下组的天然存在的氨基酸,该组由以下组成:A、R、N、D、C、Q、E、G、H、I、L、K、M、F、P、S、T、W、Y和V。取代突变体可以使用缩写体系进行描述。例如,第五(第5)个氨基酸残基被取代的取代突变可以缩写为“X5Y”,其中X是被代替的野生型或天然存在的氨基酸,5是该蛋白质或多肽的氨基酸序列内氨基酸残基的位置,Y是被取代的或非野生型或非天然存在的氨基酸。

[0196] “分离的”多肽、蛋白质、肽或核酸是已经从其天然环境中去除的分子。还应理解,“分离的”多肽、蛋白质、肽或核酸可以与赋形剂(诸如稀释剂)或佐剂一起配制,并且仍被认为是分离的。

[0197] 当用于指核酸分子、肽、多肽或蛋白质时,术语“重组”意指未知在自然界中存在的遗传物质的新组合或由其产生。重组分子可以通过重组技术领域中的任何熟知的技术来产生,包括但不限于聚合酶链式反应(PCR)、基因剪接(例如,使用限制性核酸内切酶)以及核酸分子、肽或蛋白质的固相合成。

[0198] 当用于指多肽或蛋白质时,术语“结构域”意指蛋白质中独特的功能和/或结构单元。结构域有时负责特定的功能或相互作用,有助于蛋白质的整体作用。在多种生物学背景下可以存在结构域。在具有不同功能的蛋白质中可以找到相似的结构域。可替代地,具有低序列同一性(即,小于约50%、小于约40%、小于约30%、小于约20%、小于约10%、小于约5%或小于约1%的序列同一性)的结构域可能具有相同的功能。在一些实施例中,Cas9结构域以 $1E-5$ 的E值截止值匹配TIGR03031蛋白家族。在一些实施例中,Cas9结构域以 $1E-10$ 的E值截止值匹配TIGR03031蛋白家族。在一些实施例中,Cas9结构域是RuvC结构域。在一些实施例中,Cas9结构域是HNH结构域。

[0199] 如本文所用,术语“序列相似性”或“相似性百分比%”是指核酸序列或氨基酸序列之间的同一性或一致程度。如本文所用,“序列相似性”是指如下的核酸序列,其中一个或多个核苷酸碱基的改变导致一个或多个氨基酸的取代,但不影响该DNA序列编码的蛋白质的功能性质。“序列相似性”还指核酸的修饰,诸如基本上不影响所得转录物的功能性质的一个或多个核苷酸碱基的缺失或插入。因此,应当理解,本披露不仅仅涵盖特定的示例性序列。所提出的多种修饰中的每一种正好在本领域的常规技术范围内,如所编码产物生物学活性的保留测定一样。

[0200] 此外,技术人员认识到,本披露涵盖的相似序列也由其严格条件下与本文示例的序列杂交的能力来定义。本披露的相似核酸序列是DNA序列与本文所披露的核酸的DNA序列具有至少70%、至少80%、至少90%、至少95%或至少99%相同的那些核酸。本披露的相似核酸序列是DNA序列与本文所披露的核酸的DNA序列具有约70%、至少约70%、约75%、至少约75%、约80%、至少约80%、约85%、至少约85%、约90%、至少约90%、约95%、至少约95%、约99%、至少约99%或约100%相同的那些核酸。

[0201] 如本文所用,“序列相似性”是指两个或更多个氨基酸序列,其中大于约40%的氨基酸是相同的,或者大于约60%的氨基酸是功能上相同的。功能相同或功能相似的氨基酸具有化学上相似的侧链。例如,可以根据功能相似性按照以下方式对氨基酸进行分组:

[0202] 带正电的侧链:Arg、His、Lys;

[0203] 带负电的侧链:Asn、Glu;

[0204] 极性、不带电的侧链:Ser、Thr、Asn、Gln;

[0205] 疏水性侧链:Ala、Val、Ile、Leu、Met、Phe、Tyr、Trp;

[0206] 其他:Cys、Gly、Pro。

[0207] 在一些实施例中,本披露的相似氨基酸序列具有至少40%、至少50%、至少60%、至少70%、至少80%、至少90%或至少99%相同的氨基酸。

[0208] 在一些实施例中,本披露的相似氨基酸序列具有至少60%、至少70%、至少80%、至少90%或至少95%功能相同的氨基酸。在一些实施例中,本披露的相似氨基酸序列具有约40%、至少约40%、约45%、至少约45%、约50%、至少约50%、约55%、至少约55%、约60%、至少约60%、约65%、至少约65%、约70%、至少约70%、约75%、至少约75%、约80%、至少约80%、约85%、至少约85%、约90%、至少约90%、约95%、至少约95%、约97%、至少约97%、约98%、至少约98%、约99%、至少约99%或约100%相同的氨基酸。

[0209] 在一些实施例中,本披露的相似氨基酸序列具有约60%、至少约60%、约65%、至少约65%、约70%、至少约70%、约75%、至少约75%、约80%、至少约80%、约85%、至少约85%、约90%、至少约90%、约95%、至少约95%、约97%、至少约97%、约98%、至少约98%、约99%、至少约99%或约100%功能相同的氨基酸。

[0210] 使用本领域常规方法,例如BLAST、MUSCLE、Clustal (包括ClustalW和ClustalX) 和T-Coffee (包括变体,诸如像M-Coffee、R-Coffee和Espresso),通过序列比对确定序列相似性。

[0211] 在核酸序列或氨基酸序列的背景下,术语“序列同一性”或“同一性百分比%”是指当该序列在指定的比较窗上比对时,比较的序列中相同的残基的百分比。在一些实施例中,仅将两个或更多个序列的特定部分进行比对以确定序列同一性。在一些实施例中,仅将两个或更多个序列的特定结构域进行比对以确定序列相似性。比较窗可以是至少10至超过1000个残基、至少20至约1000个残基或至少50至500个残基的区段,在其中可以比对和比较这些序列。用于确定序列同一性的比对方法是熟知的,并且可以使用公开可用的数据库(诸如BLAST)来进行。当指氨基酸序列时,“同一性百分比”或“同一性%”可以通过本领域已知的方法确定。例如,在一些实施例中,使用Karlin和Altschul,Proceedings of the National Academy of Sciences USA[美国国家科学院院刊]87:2264-2268(1990)的算法,按Karlin和Altschul,Proceedings of the National Academy of Sciences USA[美国国家科学院院刊]90:5873-5877(1993)所改良的,确定两个氨基酸序列的“同一性百分比”。这种算法被并入BLAST程序中,例如,Altschul等人,Journal of Molecular Biology[分子生物学杂志],215:403-410(1990)中描述的BLAST+或NBLAST和XBLAST程序。可采用诸如像XBLAST程序等的程序,分数=50,字长=3,执行BLAST蛋白质搜索,以获得与本披露的蛋白质分子同源的氨基酸序列。在两个序列之间存在空位的情况下,可以利用如Altschul等人,Nucleic Acids Research[核酸研究]25(17):3389-3402(1997)中描述的空位BLAST程序。

当利用BLAST程序和空位BLAST程序时,可以使用相应程序(例如,XBLAST和NBLAST)的缺省参数。

[0212] 在一些实施例中,多肽或核酸分子与参考多肽或核酸分子(或参考多肽或核酸分子的片段)分别具有70%、至少70%、75%、至少75%、80%、至少80%、85%、至少85%、90%、至少90%、95%、至少95%、97%、至少97%、98%、至少98%、99%或至少99%或100%的序列同一性。在一些实施例中,多肽或核酸分子与参考多肽或核酸分子(或参考多肽或核酸分子的片段)分别具有约70%、至少约70%、约75%、至少约75%、约80%、至少约80%、约85%、至少约85%、约90%、至少约90%、约95%、至少约95%、约97%、至少约97%、约98%、至少约98%、约99%、至少约99%或与约100%的序列同一性。

[0213] CRISPR-Cas系统

[0214] 在一些实施例中,本披露提供了一种非天然存在的CRISPR-Cas系统,该系统包含:(a)能够产生粘性末端的Cas9效应蛋白(“粘性末端Cas9”或“stiCas9”);以及(b)与该stiCas9形成复合物并且包含引导序列的引导多核苷酸,其中该引导序列与真核细胞中的靶序列杂交,但不与细菌细胞中的序列杂交;其中该复合物在自然界中不存在。

[0215] 通常,CRISPR或CRISPR-Cas系统的特征在于促进靶序列位点处CRISPR复合物形成的元件(在内源性CRISPR系统的背景下也称为前间区序列)。在CRISPR复合物形成的背景下,“靶序列”是指设计引导多核苷酸靶向的序列,例如,它们具有互补性,其中靶序列和引导多核苷酸之间的杂交促进了CRISPR复合物的形成。引导多核苷酸中与靶序列互补的对于裂解活性可能重要的节段在本文中称为引导序列。靶序列可以包含任何多核苷酸,诸如DNA或RNA多核苷酸,并且可能位于目的靶基因座内。在一些实施例中,靶序列位于细胞的细胞核或细胞质中。在一些实施例中,该靶序列位于染色体(TSC)上。在一些实施例中,该靶序列位于载体(TSV)上。

[0216] 如本文所述,Cas蛋白是CRISPR-Cas系统的组分,其可尤其用于基因组编辑、基因调节、基因回路构建和功能基因组学。尽管Cas1和Cas2蛋白显现对所有目前鉴定的CRISPR系统通用,但Cas3、Cas9和Cas10蛋白被认为分别对I型、II型和III型CRISPR系统具有特异性。

[0217] 在围绕CRISPR-Cas9系统(II型系统)的首次出版之后,已经在一系列细菌物种中鉴定出Cas9变体,并且已经对许多变体进行了功能性表征。参见,例如,Chylinski等人,“Classification and evolution of type II CRISPR-Cas systems[II型CRISPR-Cas系统的分类与进化]”,Nucleic Acids Research[核酸研究]42(10):6091-6105(2014),Ran等人,“In vivo genome editing using Staphylococcus aureus Cas9[使用金黄色葡萄球菌Cas9进行的体内基因组编辑]”,Nature[自然]520(7546):186-91(2015),以及Esvelt等人,“Orthogonal Cas9 proteins for RNA-guided gene regulation and editing[用于RNA引导基因调节和编辑的正交Cas9蛋白]”,Nature Methods[自然方法]10(11):1116-1121(2013),将其各自通过引用以其整体并入本文。

[0218] 本披露涵盖II型CRISPR-Cas系统的新型效应蛋白,其中Cas9是示例性效应蛋白。因此,术语“Cas9”、“Cas9蛋白”和“Cas9效应蛋白”是可互换的,并且在本文中用于描述当在CRISPR-Cas9系统中使用时能够提供粘性末端的效应蛋白。在一些实施例中,术语Cas9是指II-B型Cas9。在一些实施例中,术语Cas9是指工程化的Cas9变体,诸如像deadCas9-FokI、

Cas9n^{D10A}-FokI和Cas9n^{H840A}-FokI。

[0219] 在一些实施例中,在原核或真核细胞中的Cas9效应蛋白对于体外、体内或离体应用是功能性的。

[0220] 术语Cas9效应蛋白可以指具有Cas9样功能的效应蛋白,通常具有RuvC和HNH核酸酶结构域。在一些实施例中,Cas9效应蛋白的RuvC结构域和HNH结构域各自裂解双链靶DNA的一条链。因此,例如,如果RuvC结构域和HNH结构域在同一位置裂解每条链,则裂解的结果将是具有平端的双链靶DNA。若该RuvC结构域和HNH结构域在不同位置裂解每条链(即以某一“偏移”切割),则裂解的结果将是具有突出端的双链靶DNA。在多个实施例中,该stiCas9蛋白的RuvC和HNH结构域以3-核苷酸偏移切割。在多个实施例中,该stiCas9蛋白的RuvC和HNH结构域以4-核苷酸偏移切割。在多个实施例中,该stiCas9蛋白的RuvC和HNH结构域以5-核苷酸偏移切割。在多个实施例中,该stiCas9蛋白的RuvC和HNH结构域以约1、约2、约3、约4、约5、约6、约7、约8、约9、约10、约11、约12、约13、约14、约15、约16、约17、约18、约19、约20、约21、约22、约23、约24、约25、约26、约27、约28、约29、约30、约31、约32、约33、约34、约35、约36、约37、约38、约39或约40个核苷酸的偏移切割。

[0221] 在一些实施例中,术语Cas9效应蛋白是指具有RuvC结构域和HNH结构域的Cas9,其中该RuvC结构域和HNH结构域在双链靶DNA的每条链上的不同位置裂解。在一些实施例中,该Cas9效应蛋白的RuvC结构域在自PAM的约-10、约-9、约-8、约-7或约-6个核苷酸处裂解双链靶DNA的一条链(其可以称为例如“非靶链”),而Cas9效应蛋白的HNH结构域在自PAM的约-5、约-4、约-3、约-2或约-1个核苷酸处裂解双链靶DNA的另一条链(可以称为例如“靶链”)。

[0222] 在一些实施例中,该RuvC结构域在自PAM约-8个核苷酸处裂解双链靶DNA的一条链。在一些实施例中,该RuvC结构域在自PAM约-7个核苷酸处裂解双链靶DNA的一条链。在一些实施例中,该RuvC结构域在自PAM约-6个核苷酸处裂解双链靶DNA的一条链。在一些实施例中,该HNH结构域在自PAM约-4个核苷酸处裂解双链靶DNA的一条链。在一些实施例中,该HNH结构域在自PAM约-3个核苷酸处裂解双链靶DNA的一条链。在一些实施例中,该HNH结构域在自PAM约-2个核苷酸处裂解双链靶DNA的一条链。

[0223] 在一些实施例中,术语Cas9效应蛋白是指通过HMMER搜索,特别是hmmscan程序(HMMER版本3.1b2)鉴定的具有TIGR03031蛋白家族的Cas9。本披露还涉及与II型CRISPR-Cas系统相关的效应蛋白的鉴定和工程化。在一些实施例中,该效应蛋白包含单亚基效应子模块。在一些实施例中,将野生型Cas9效应子或Cas9蛋白的工程化版本与一个或多个功能结构域(诸如像核定位信号(NLS)和FokI核酸酶)融合。本披露涵盖预测新的II-B型CRISPR-Cas系统并鉴定其中组分的计算方法和算法。

[0224] 在一些实施例中,鉴定新型II-B型CRISPR-Cas基因座的计算方法包括以下描述的方法和先前在Shmakov等人,Nature Reviews Microbiology[自然微生物学综述]15,169-182(2017)中描述的方法。CRISPR-Cas基因座在给定核苷酸序列中的存在和位置可以通过在TBLASTN中使用已知Cas蛋白中的一种的蛋白序列作为种子(例如Cas1),使用例如0.01的E值截止针对核苷酸序列来鉴定。鉴定CRISPR-Cas基因座的存在和位置的另一种方法是通过使用默认参数的程序(诸如像CRISPRfnder或PILER-CR)在核苷酸序列中搜索CRISPR阵列。鉴定CRISPR-Cas基因座后,就可以提取包括CRISPR-Cas基因座上游和下游10kbp的序列。可以用使用默认参数的诸如GeneMark或MetaGeneMark的软件鉴定提取的核苷酸序列中

基因的存在。然后将鉴定的基因翻译成蛋白质序列,并针对具有已知功能的蛋白质(即,Cas1、Cas2、Cas4、Cas9等)的数据库使用同源性搜索(诸如RPS-BLAST、BLAST或HMMR)为其添加注释以指出对其预测的功能。

[0225] 研究了用上述方法鉴定的CRISPR-Cas基因座以探索同一CRISPR-Cas基因座中是否同时存在Cas9和Cas4蛋白,因为这些基因座很可能包含IIB型Cas9。为了进一步增加IIB型Cas9的可能性,使用hmmscan搜索Cas9蛋白是否属于TIGRFAM:TIGR03031家族。

[0226] 在一些实施例中,鉴定新型II-B型CRISPR-Cas基因座的方法包括在与Cas4蛋白质相同的基因座中鉴定Cas9蛋白。在一些实施例中,鉴定新型II-B型CRISPR-Cas基因座的方法包括将公开可用的宏基因组基因目录翻译成氨基酸序列,用TIGR03031蛋白家族谱扫描每个氨基酸序列以鉴定高于预定截止E值(诸如像 $1E-5$ 至 $1E-10$)的匹配。

[0227] TIGRFAM是蛋白质家族的集合,其具有组织管理的多序列比对、隐马尔可夫模型和设计为通过序列同源性支持蛋白质的自动功能鉴定的相关信息等特征。应用于序列比对的隐马尔可夫模型(HMM)是指连续列的蛋白质多序列比对的统计模型。典型地,蛋白质谱HMM是根据每个氨基酸的基于位置的评分、沿序列长度的插入和缺失,从组织管理的多序列比对中开发出来的。评分以零散的信息和E值形式报告。低于“信任的截止值”或“信任的限值”的E值(诸如像0.001)将被视为阳性“命中”或阳性鉴定。因此,鉴定具有低E值截止值的序列可能属于特定的蛋白质家族。在一些实施例中,该E值截止值为 $1E-10$ 。在一些实施例中,该E值截止值为 $1E-5$ 。在一些实施例中,信任的截止E值为至少 $1E-10$ 、至少 $1E-9$ 、至少 $1E-8$ 、至少 $1E-7$ 、至少 $1E-6$ 、至少 $1E-5$ 、至少 $1E-4$ 、至少 $1E-3$ 、至少 $1E-2$ 或至少 $1E-1$ 。

[0228] 在一些实施例中,所有预测的蛋白质编码基因的鉴定是通过将经鉴定的基因与Cas蛋白特异性谱进行比较,并根据NCBI保守结构域数据库(CDD)对其添加注释来实现的,该NCBI保守域数据库是由古老结构域和全长蛋白质的充分注释的多序列比对模型的集合组成的蛋白质注释资源。这些可用作位置特异性评分矩阵(PSSM),用于经由RPS-BLAST快速鉴定蛋白序列中的保守结构域。CDD内容包括NCBI组织管理的结构域,这些结构域使用3D结构信息来显式定义结构域边界并提供对序列/结构/功能关系的洞察,以及从多个外部源数据库(Pfam、SMART、COG、PRK、TIGRFAM)导入的结构域模型。蛋白质数据库描述于例如Finn等人,Nucleic Acids Research[核酸研究]数据库第44期:D279-D285(2016);Letunic等人,Nucleic Acids Research[核酸研究],doi:gkx922(2017);Tatusov等人,Science[科学]278(5338):631-637(1997);和Haft等人,Nucleic Acids Research[核酸研究]数据库第41期:D387-D395(2013),其中各自以其全文并入本文。

[0229] 在一些实施例中,使用HMMER(或HMMER的任何版本,诸如HMMER2或HMMER3)鉴定了新型II-B型CRISPR-Cas基因座以搜索保守结构域。HMMER是免费的常用软件包,用于序列分析、鉴定同源蛋白或核苷酸序列以及序列比对。HMMER实施了称为剖面隐马尔可夫模型的概率模型。HMMER可以与剖面数据库(profile database)(诸如Pfam、SMART、COG、PRK或TIGRFAM)一起使用。HMMER还可与查询序列一起使用,例如,针对数据库(即,phmmer)搜索蛋白质查询序列或进行迭代搜索(即,Jackhmmmer)。在一些实施例中,通过搜索特定蛋白家族中特定结构域的存在来鉴定新型II-B型CRISPR-Cas基因座。在一些实施例中,该TIGRFAM蛋白家族是TIGRFAM:TIGR03031。在一些实施例中,该特异性结构域以至少 $1E-10$ 、至少 $1E-9$ 、至少 $1E-8$ 、至少 $1E-7$ 、至少 $1E-6$ 、至少 $1E-5$ 、至少 $1E-4$ 、至少 $1E-3$ 、至少 $1E-2$ 或至少 $1E-1$ 的E值

截止值匹配TIGR03031蛋白家族。在一些实施例中,该特定结构域与本文鉴定的TIGR03031结构域的任何一个具有至少60%、至少70%、至少80%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或约100%的序列相似性。在一些实施例中,该特定结构域与SEQ ID NO:10-97或192-195中的任何一个具有至少60%、至少70%、至少80%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或约100%的序列相似性。在一些实施例中,该特定结构域与SEQ ID NO:10-97或192-195中的任何一个具有至少40%、至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或约100%的序列同一性。

[0230] 在一些实施例中,该stiCas9源自具有II-B型CRISPR系统的细菌物种。在一些实施例中,该II-B型CRISPR系统包括cas4基因。如本文所讨论,CRISPR系统被划分为I型、II型和III型。所有II型CRISPR系统都包含cas操纵子上的Cas1、cas2和cas9基因。II型CRISPR系统进一步划分为II-A型、II-B型和II-C型。在一些实施例中,通过cas操纵子上cas4基因的存在来鉴定II-B型CRISPR系统。在II-A型或II-C型CRISPR系统中找不到cas4基因。

[0231] II型CRISPR系统也可以根据单个cas基因的序列,例如cas 9的序列和/或结构域进行分类。蛋白质结构域可以通过保守的序列或保守的基序进行鉴定,并分为家族、超家族和亚家族。例如,可以根据PFAM或TIGRFAM对蛋白质域进行分类。因此,可以鉴定Cas蛋白并用蛋白结构域进行分类。例如,II-A型Cas9蛋白,包括来自酿脓链球菌的Cas9,属于TIGR01865TIGRFAM蛋白家族。相反,II-B型Cas9蛋白属于TIGR03031TIGRFAM蛋白家族。

[0232] 因此,在一些实施例中,本披露的stiCas9包含与SEQ ID NO:10-97或192-195中的任一个具有至少95%的序列相似性的结构域。在一些实施例中,本披露的stiCas9包含与SEQ ID NO:10-97或192-195中的任何一个具有至少10%、至少20%、至少30%、至少40%、至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或约100%的序列相似性的结构域。在一些实施例中,本披露的stiCas9包含以至少1E-10、至少1E-9、至少1E-8、至少1E-7,至少1E-6,至少1E-5、至少1E-4、至少1E-3、至少1E-2或至少1E-1的E值截止值匹配TIGR03031蛋白家族的结构域。

[0233] 在一些实施例中,该II-B型Cas9源自具有II-B型CRISPR系统的任何物种。在一些实施例中,该II-B型Cas9源自以下细菌物种:嗜肺军团菌、新凶手弗朗西斯菌、 γ 变形菌HTCC5015、人粪便副萨特氏菌、华德萨特氏菌、硫磺单胞菌属物种SCADC、瘤胃杆菌属物种RM87、伯克霍尔德氏菌目细菌1_1_47、拟杆菌门口腔分类群274菌株F0058、产琥珀酸沃廉氏菌、伯克霍尔德氏菌目细菌YL45、嗜淀粉瘤胃杆菌、弯曲杆菌属物种P0111、弯曲杆菌属物种RM9261、拉尼尔弯曲杆菌菌株RM8001、拉尼尔弯曲杆菌菌株P0121、鼠毛滴虫、伦敦军团菌、沙姆盐弧菌、钩端螺旋体属物种分离株FW.030、莫里特拉氏菌属物种分离株NORP46、内生单胞菌属物种S-B4-1U、喜盐泰米尔纳德菌、需钠弧菌、斯氏弓形杆菌、蜃楼弗朗西斯氏菌、西班牙弗朗西斯氏菌或嗜盐副内生单胞菌。

[0234] 在一些实施例中,术语Cas9是指包含嗜肺军团菌Cas9蛋白的氨基酸序列的多肽。在一些实施例中,术语Cas9是指包含新凶手弗朗西斯菌Cas9蛋白的氨基酸序列的多肽。在一些实施例中,术语Cas9是指包含 γ 变形菌HTCC5015Cas9蛋白的氨基酸序列的多肽。在一些实施例中,术语Cas9是指包含人粪便副萨特氏菌Cas9蛋白的氨基酸序列的多肽。在一些实施例中,术语Cas9是指包含华德萨特氏菌Cas9蛋白的氨基酸序列的多肽。在一些实施例

中,术语Cas9是指包含硫磺单胞菌属物种SCADC Cas9蛋白的氨基酸序列的多肽。在一些实施例中,术语Cas9是指包含瘤胃杆菌属物种RM87 Cas9蛋白的氨基酸序列的多肽。在一些实施例中,术语Cas9是指包含伯克霍尔德氏菌目细菌1_1_47 Cas9蛋白的氨基酸序列的多肽。在一些实施例中,术语Cas9是指包含拟杆菌门口腔分类群274菌株F0058 Cas9蛋白的氨基酸序列的多肽。在一些实施例中,术语Cas9是指包含产琥珀酸沃廉氏菌Cas9蛋白的氨基酸序列的多肽。在一些实施例中,术语Cas9是指包含伯克霍尔德氏菌目细菌YL45Cas9蛋白的氨基酸序列的多肽。在一些实施例中,术语Cas9是指包含嗜淀粉瘤胃杆菌菌株DSM1361Cas9蛋白的氨基酸序列的多肽。在一些实施例中,术语Cas9是指包含弯曲杆菌属物种P0111Cas9蛋白的氨基酸序列的多肽。在一些实施例中,术语Cas9是指包含弯曲杆菌属物种RM9261Cas9蛋白的氨基酸序列的多肽。在一些实施例中,术语Cas9是指包含拉尼尔弯曲杆菌RM8001菌株Cas9蛋白的氨基酸序列的多肽。在一些实施例中,术语Cas9是指包含拉尼尔弯曲杆菌P0121菌株Cas9蛋白的氨基酸序列的多肽。在一些实施例中,术语Cas9是指包含鼠毛滴虫Cas9蛋白的氨基酸序列的多肽。在一些实施例中,术语Cas9是指包含伦敦军团菌Cas9蛋白的氨基酸序列的多肽。在一些实施例中,术语Cas9是指包含沙姆盐弧菌Cas9蛋白的氨基酸序列的多肽。在一些实施例中,术语Cas9是指包含钩端螺旋体属物种分离株FW.030 Cas9蛋白的氨基酸序列的多肽。在一些实施例中,术语Cas9是指包含莫里特拉氏菌属物种分离株NORP46 Cas9蛋白的氨基酸序列的多肽。在一些实施例中,术语Cas9是指包含内生单胞菌属物种S-B4-1U Cas9蛋白的氨基酸序列的多肽。在一些实施例中,术语Cas9是指包含喜盐泰米尔纳德菌Cas9蛋白的氨基酸序列的多肽。在一些实施例中,术语Cas9是指包含需钠弧菌Cas9蛋白的氨基酸序列的多肽。在一些实施例中,术语Cas9是指包含斯氏弓形杆菌Cas9的氨基酸序列的多肽。在一些实施例中,术语Cas9是指包含蜃楼弗朗西斯氏菌Cas9的氨基酸序列的多肽。在一些实施例中,术语Cas9是指包含西班牙弗朗西斯氏菌Cas9的氨基酸序列的多肽。在一些实施例中,术语Cas9是指包含嗜盐副内生单胞菌Cas9的氨基酸序列的多肽。在一些实施例中,术语Cas9是指来自宏基因组序列目录的Cas9多肽。在一些实施例中,术语Cas9是指包含SEQ ID NO:10-97或192-195中的任一个的多肽。参见图30,SEQ ID NO:10-80;图31,SEQ ID NO:81-97;以及图47,SEQ ID NO:192-195。

[0235] 在一些实施例中,该stiCas9蛋白包含与SEQ ID NO:10-97或192-195中任一个的氨基酸序列具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或约100%同一性的序列的结构域。在一些实施例中,该stiCas9蛋白与SEQ ID NO:10-97或192-195中任一个的氨基酸序列至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或约100%相同。

[0236] 如本文所用,术语“粘性末端(cohesive end)”、“交错末端”或“粘性末端(sticky end)”是指具有长度不等的链的核酸片段。与“平端”相反,粘性末端是通过核酸(典型地为DNA)进行交错切割而产生的。黏性或粘性末端具有突出的单链的链(这些链具有不成对的核苷酸)或突出端,例如,3'或5'突出端。每个突出端可以与另一个互补的突出端进行退火以形成碱基对。两个互补的粘性末端可以通过相互作用诸如氢键结合在一起退火。退火的粘性末端的稳定性取决于成对的突出端的解链温度。两个互补的粘性末端可以通过化学或酶促连接(例如,通过DNA连接酶)连接在一起。

[0237] 先前已知Cas9蛋白会产生带有平端的双链DNA断裂(参见,例如,Jinek等人,2012)。本披露提供了能够产生粘性末端的Cas9蛋白,在本文中也称为“stiCas9”或“粘性Cas9”。具有粘性末端的DNA片段在其他应用中比平端具有优势,诸如像将核酸插入片段之间并将片段重新连接在一起。具有平端的DNA序列不提供插入核酸的特异性,即核酸可以插入任一平端。另一方面,粘性末端将仅与互补粘性末端配对,因此使转基因具有优选取向的整合成为可能。在一些实施例中,粘性末端通过非同源末端连接和微同源性介导的末端连接方法促进DNA的插入。

[0238] 在一些实施例中,由该stiCas9产生的粘性末端包含具有3至40个核苷酸的单链多核苷酸突出端。在一些实施例中,由该stiCas9产生的粘性末端包含具有4至20个核苷酸的单链多核苷酸突出端。在一些实施例中,由该stiCas9产生的粘性末端包含具有5至15个核苷酸的单链多核苷酸突出端。在一些实施例中,由该stiCas9产生的粘性末端包含具有3、4、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、31、32、33、34、35、36、37、38、39或40个核苷酸的单链多核苷酸突出端。在一些实施例中,由该stiCas9产生的粘性末端是5'突出端。在一些实施例中,由该stiCas9产生的粘性末端是3'突出端。

[0239] 本文所述的组合物和方法可包含引导多核苷酸。在一些实施例中,该引导多核苷酸是RNA分子。与CRISPR-Cas组分结合并将其靶向靶DNA内特定位置的RNA分子在本文中称为“引导RNA”、“gRNA”或“小引导RNA”,在本文中也可能称为“靶向DNA的RNA”。引导多核苷酸,例如引导RNA,包含至少两个核苷酸区段:至少一个“DNA结合区段”和至少一个“多肽结合区段”。“”是指分子的一个部分、节区段或区域,例如,引导多核苷酸分子的核苷酸的连续伸展段。除非另有明确定义,否则“区段”的定义不限于特定数目的总碱基对。

[0240] 在一些实施例中,该引导多核苷酸的DNA结合区段与真核细胞中的靶序列杂交,但不与细菌细胞中的序列杂交。如本文所用,细菌细胞中的序列是指细菌生物体原生的多核苷酸序列,即天然存在的细菌多核苷酸序列或细菌起源的序列。例如,该序列可以是细菌染色体或细菌质粒,或者是在细菌细胞中天然存在的任何其他多核苷酸序列。

[0241] 在一些实施例中,该引导多核苷酸的多肽结合区段与Cas9结合。在一些实施例中,该引导多核苷酸的多肽结合区段与stiCas9结合。

[0242] 在一些实施例中,该引导多核苷酸是10至150个核苷酸。在一些实施例中,该引导多核苷酸是20至120个核苷酸。在一些实施例中,该引导多核苷酸是30至100个核苷酸。在一些实施例中,该引导多核苷酸是40至80个核苷酸。在一些实施例中,该引导多核苷酸是50至60个核苷酸。在一些实施例中,该引导多核苷酸是10至35个核苷酸。在一些实施例中,该引导多核苷酸是15至30个核苷酸。在一些实施例中,该引导多核苷酸是20至25个核苷酸。

[0243] 可以将引导多核苷酸(例如,引导RNA)作为分离的分子(例如,RNA分子)引入靶细胞中,或使用包含编码引导多核苷酸(例如,引导RNA)的DNA的表达载体将其引入细胞中。

[0244] 引导多核苷酸(例如,引导RNA)的“DNA结合区段”(或“DNA靶向序列”)包含与靶DNA内的特定序列互补的核苷酸序列。

[0245] 本披露的引导多核苷酸(例如,引导RNA)可以包括多肽结合序列/区段。引导多核苷酸(例如,引导RNA)的多肽结合区段(或“蛋白结合序列”)与本披露的Cas蛋白的多核苷酸结合结构域相互作用。此类多肽结合区段或序列是本领域技术人员已知的,例如,美国专利

申请公开2014/0068797、2014/0273037、2014/0273226、2014/0295556、2014/0295557、2014/0349405、2015/0045546、2015/0071898、2015/0071899和2015/0071906中披露的那些,这些公开的披露内容以其全文并入本文。

[0246] 在本披露的一些实施例中,该stiCas9和引导多核苷酸可以形成复合物。“复合物”是两个或更多个缔合的核酸和/或多肽的组。在一些实施例中,当复合物的所有组分一起存在时形成复合物,即自组装复合物。在一些实施例中,通过复合物的不同组分之间的化学相互作用(诸如像氢键结合)形成复合物。在一些实施例中,通过stiCas9对引导多核苷酸的二级结构识别,该引导多核苷酸与该stiCas9形成复合物。在一些实施例中,stiCas9蛋白是无活性的,即,不展现出核酸酶活性,直到它与引导多核苷酸形成复合物。引导RNA的结合诱导stiCas9的构象变化,以将stiCas9从无活性形式转化为活性(即,催化活性)形式。在本披露的多个实施方案中,该stiCas9和引导多核苷酸的复合物在自然界中不存在。

[0247] 在一些实施例中,本披露提供了一种非天然存在的CRISPR-Cas系统,该系统包含:能够产生粘性末端并包含核定位信号(NLS)的Cas9效应蛋白(stiCas9),以及与stiCas9形成复合物并包含引导序列的引导多核苷酸,其中该复合物在自然界中不存在。

[0248] 在一些实施例中,该stiCas9包含一个或多个核定位信号。“核定位信号”或“核定位序列”(NLS)是对蛋白质“加标签”以通过核转运导入细胞核的氨基酸序列,即,具有NLS的蛋白质被转运至细胞核。典型地,NLS包含暴露在蛋白质表面的带正电荷的Lys或Arg残基。示例性核定位序列包括但不限于来自以下的NLS:SV40大T抗原,w,EGL-13,c-Myc以及TUS蛋白。在一些实施例中,该NLS包含PKKKRKV(SEQ ID NO:1)序列。在一些实施例中,该NLS包含AVKRPAATKKAGQAKKKLD(SEQ ID NO:2)序列。在一些实施例中,该NLS包含PAAKRVKLD(SEQ ID NO:3)序列。在一些实施例中,该NLS包含MSRRRKNPTKLSNAKLAKEVEN(SEQ ID NO:4)序列。在一些实施例中,该NLS包含KLKIKRPVK(SEQ ID NO:5)序列。其他核定位序列包括但不限于hnRNP A1的酸性M9结构域、酵母转录抑制子Mat α 2中的序列KIPIK(SEQ ID NO:6)和PY-NLS。

[0249] 在一些实施例中,本披露提供了一种非天然存在的CRISPR-Cas系统,该系统包含:(a)编码能够产生粘性末端的Cas9效应蛋白(stiCas9)的一个或多个核苷酸;以及(b)编码引导多核苷酸的核苷酸序列,该引导多核苷酸与该stiCas9形成复合物并且包含引导序列,其中该引导序列与真核细胞中的靶序列杂交,但不与细菌细胞中的序列杂交,并且其中该复合物在自然界中不存在。

[0250] 在一些实施例中,该stiCas9蛋白由一种或多种多核苷酸编码。在一些实施例中,该多核苷酸是DNA。在一些实施例中,该多核苷酸是RNA。

[0251] 在一些实施例中,该stiCas9由源自嗜肺军团菌Cas9蛋白的一种或多种多核苷酸编码。在一些实施例中,该stiCas9由源自新凶手弗朗西斯菌Cas9蛋白的一种或多种多核苷酸编码。在一些实施例中,该stiCas9由源自 γ 变形菌HTCC5015 Cas9蛋白的一种或多种多核苷酸编码。在一些实施例中,该stiCas9由源自人粪便副萨特氏菌Cas9蛋白的一种或多种多核苷酸编码。在一些实施例中,该stiCas9由源自华德萨特氏菌Cas9蛋白的一种或多种多核苷酸编码。在一些实施例中,该stiCas9由源自硫磺单胞菌属物种SCADC Cas9蛋白的一种或多种多核苷酸编码。在一些实施例中,该stiCas9由源自瘤胃杆菌属物种RM87Cas9蛋白的一种或多种多核苷酸编码。在一些实施例中,该stiCas9由源自伯克霍尔德氏菌目细菌1_1_

47 Cas9蛋白的一种或多种多核苷酸编码。在一些实施例中,该stiCas9由源自拟杆菌门口腔分类群274菌株F0058 Cas9蛋白的一种或多种多核苷酸编码。在一些实施例中,该stiCas9由源自产琥珀酸沃廉氏菌Cas9蛋白的一种或多种多核苷酸编码。在一些实施例中,该stiCas9由源自伯克霍尔德里氏菌目细菌YL45 Cas9蛋白的一种或多种多核苷酸编码。在一些实施例中,该stiCas9由源自嗜淀粉瘤胃杆菌菌株DSM 1361 Cas9蛋白的一种或多种多核苷酸编码。在一些实施例中,该stiCas9由源自弯曲杆菌属物种P0111 Cas9蛋白的一种或多种多核苷酸编码。在一些实施例中,该stiCas9由源自弯曲杆菌属物种RM9261 Cas9蛋白的一种或多种多核苷酸编码。在一些实施例中,该stiCas9由源自拉尼尔弯曲杆菌RM8001菌株Cas9蛋白的一种或多种多核苷酸编码。在一些实施例中,该stiCas9由源自拉尼尔弯曲杆菌P0121菌株Cas9蛋白的一种或多种多核苷酸编码。在一些实施例中,该stiCas9由源自鼠毛滴虫Cas9蛋白的一种或多种多核苷酸编码。在一些实施例中,该stiCas9由源自矽伦敦军团菌Cas9蛋白的一种或多种多核苷酸编码。在一些实施例中,该stiCas9由源自沙姆盐弧菌Cas9蛋白的一种或多种多核苷酸编码。在一些实施例中,该stiCas9由源自钩端螺旋体属物种分离株FW.030 Cas9蛋白的一种或多种多核苷酸编码。在一些实施例中,该stiCas9由源自莫里特拉氏菌属物种分离株NORP46Cas9蛋白的一种或多种多核苷酸编码。在一些实施例中,该stiCas9由源自内生单胞菌属物种S-B4-1U Cas9蛋白的一种或多种多核苷酸编码。在一些实施例中,该stiCas9由源自喜盐泰米尔纳德菌Cas9蛋白的一种或多种多核苷酸编码。在一些实施例中,该stiCas9由源自需钠弧菌Cas9蛋白的一种或多种多核苷酸编码。在一些实施例中,该stiCas9由源自斯氏弓形杆菌Cas9蛋白的一种或多种多核苷酸编码。在一些实施例中,该stiCas9由源自蜃楼弗朗西斯氏菌Cas9蛋白的一种或多种多核苷酸编码。在一些实施例中,该stiCas9由源自西班牙弗朗西斯氏菌Cas9蛋白的一种或多种多核苷酸编码。在一些实施例中,该stiCas9由源自嗜盐副内生单胞菌Cas9蛋白的一种或多种多核苷酸编码。

[0252] 在一些实施例中,本披露的stiCas9包含以至少1E-10、至少1E-9、至少1E-8、至少1E-7,至少1E-6,至少1E-5、至少1E-4、至少1E-3、至少1E-2或至少1E-1的E值截止值匹配TIGR03031蛋白家族的结构域。

[0253] 在一些实施例中,该CRISPR-Cas系统的引导多核苷酸由核苷酸序列编码。在一些实施例中,该核苷酸序列是DNA。在一些实施例中,该引导多核苷酸是引导RNA。在一些实施例中,该引导多核苷酸的引导序列是DNA靶向序列。

[0254] 在一些实施例中,编码stiCas9的核苷酸序列是密码子优化的序列。密码子优化序列的一个实例在这种情况下是针对在真核生物(例如,人类)中表达(即,针对在人类中表达进行优化)或针对如本文讨论的另一种真核生物、动物、或哺乳动物进行优化的序列;参见例如W0 2014/093622中的SaCas9人密码子优化序列作为密码子优化序列的实例(根据本领域知识和本披露,一个或多个密码子优化编码核酸分子,尤其是对于效应蛋白(例如Cas9)而言,在技术人员的范围内)。其他实例也是可能的,并且针对除人类之外的宿主物种的密码子优化或针对具体器官的密码子优化是已知的。在一些实施例中,对编码DNA/RNA靶向Cas蛋白的酶编码序列进行密码子优化以在特定细胞(诸如真核细胞)中表达。这些真核细胞可以是特定生物的细胞或源自特定生物的细胞,该特定生物诸如植物或哺乳动物,包括但不限于本文讨论的人类或非人类真核生物或动物或哺乳动物,例如,小鼠、大鼠、兔、狗、家畜或非人类的哺乳动物或灵长类动物。在一些实施例中,排除了可能导致人类或动物遭

受痛苦而对其没有任何实质性医学益处的用于修饰人类的种系遗传同一性的方法和/或用于修饰动物的遗传同一性的方法,以及由此类方法产生的动物。一般而言,密码子优化是指通过用在宿主细胞的基因中更常使用或者最常使用的密码子代替天然序列的至少一个密码子(例如约或多于约1、2、3、4、5、10、15、20、25、50个、或更多个密码子同时维持该天然氨基酸序列而修饰核酸序列以便增强在目的宿主细胞中的表达的方法。不同物种对特定氨基酸的某些密码子展现出特定的偏好。密码子偏好性(生物体之间密码子使用的差异)通常与信使RNA(mRNA)的翻译效率相关联,而信使RNA(mRNA)的翻译效率又被认为尤其取决于翻译的密码子的特性和特定转移RNA(tRNA)分子的可用性。细胞中所选tRNA的优势通常反映了肽合成中最常使用的密码子。因此,可以基于密码子优化来定制基因以在给定生物体中实现最佳基因表达。密码子使用表可以容易地获得,例如,在“密码子使用数据库”(www.kazusa.or.jp/codon/)上可获得,这些表可以通过多种方式进行修改。参见Nakamura等人,“Codon usage tabulated from the international DNA sequence databases: status for the year 2000[根据国际DNA序列数据库的密码子使用表:2000年状态]”, *Nucleic Acids Research*[核酸研究]28:292(2000)。用于密码子优化特定的序列以便在特定的宿主细胞中表达的计算机算法也是可行的。在一些实施例中,编码DNA/RNA靶向Cas蛋白的序列中的一个或多个密码子(例如,1、2、3、4、5、10、15、20、25、50或更多个或所有密码子)对应于特定氨基酸最常使用的密码子。关于酵母中的密码子使用,参考在线酵母基因组数据库(www.yeastgenome.org/community/codon_usage.shtml),或Bennetzen和Hall,“Codon selection in yeast[酵母中的密码子选择]”, *Journal of Biological Chemistry*[生物化学杂志],257(6):3026-31(1982)。至于包括藻类在内的植物中的密码子使用,参考Campbell和Gowri,“Codon usage in higher plants, green algae, and cyanobacteria[高等植物、绿藻和蓝细菌中的密码子使用]”, *Plant Physiology*[植物生理学]92(1):1-11(1990);以及Murray等人,“Codon usage in plant genes[植物基因中密码子的使用]”, *Nucleic Acids Research*[核酸研究]17(2):477-98(1989);或Morton,“Selection on the codon bias of chloroplast and cyanelle genes in different plant and algal lineages[不同植物和藻类谱系中叶绿体和蓝色小体基因的密码子偏好性的选择]”, *Molecular Evolution*[分子进化]46(4):449-59(1998)。在一些实施例中,将SEQ ID NO:10-97或192-195中的一个或多个进行了密码子优化。

[0255] 在一些实施例中,对编码stiCas9的核苷酸序列进行密码子优化以在真核细胞中表达。在一些实施例中,对编码stiCas9的核苷酸序列进行密码子优化以在动物细胞中表达。在一些实施例中,对编码stiCas9的核苷酸序列进行密码子优化以在人类细胞中表达。对编码stiCas9的核苷酸序列进行密码子优化以在植物细胞中表达。密码子优化是对密码子进行调整来匹配表达宿主的tRNA丰度,以提高重组或异源蛋白质表达的产量和效率。密码子优化方法是本领域的常规方法,可以使用软件程序执行,诸如像集成DNA技术公司(Integrated DNA Technologies)的密码子优化工具、Entelechon的密码子使用表分析工具、GENEMAKER的Blue Heron软件、Aptagen的Gene Forge软件、DNA Builder软件、通用密码子使用分析软件、公开可用的OPTIMIZER软件以及金斯瑞的OptimumGene算法。

[0256] 在一些实施例中,本披露的CRISPR-Cas系统进一步包含tracrRNA。“tracrRNA”或反式激活CRISPR-RNA与前体crRNA或前体CRISPR-RNA形成RNA双链体,然后被RNA特异性核

糖核酸酶RNA酶III裂解,形成crRNA/tracrRNA杂交体。在一些实施例中,该引导RNA包含crRNA/tracrRNA杂交体。在一些实施例中,该引导RNA的tracrRNA组分激活Cas9蛋白。

[0257] 在本披露的一些实施例中,stiCas9、引导多核苷酸和tracrRNA能够形成复合物。在一些实施例中,该stiCas9、引导多核苷酸和tracrRNA的复合物在自然界中不存在。

[0258] 在一些实施例中,本披露提供了一种非天然存在的CRISPR-Cas系统,该系统包含一个或多个载体,该一个或多个载体包含:(a)可操作地连接至一个或多个核苷酸序列的调节元件,所述核苷酸序列编码能够产生粘性末端的Cas9效应蛋白(stiCas9);(b)与stiCas9形成复合物并且包含引导序列的引导多核苷酸,其中该引导序列能够与真核细胞中的靶序列杂交,但不与细菌细胞中的序列杂交;其中该复合物在自然界中不存在。本领域技术人员应理解,包含“与stiCas9形成复合物并且包含引导序列的引导多核苷酸”的载体还将包括包含可被转录成引导多核苷酸的多核苷酸序列的载体。例如,可以转录DNA载体以产生引导RNA序列。

[0259] 在一些实施例中,本披露提供了一种非天然存在的CRISPR-Cas系统,该系统包含一个或多个载体,该一个或多个载体包含:可操作地连接至编码能够产生粘性末端的Cas9效应蛋白(stiCas9)的一个或多个核苷酸序列的调节元件(其中该调节元件是真核调节元件),以及与stiCas9形成复合物并包含引导序列的引导多核苷酸序列,其中该复合物在自然界中不存在。

[0260] 在一些实施例中,该调节元件是启动子。在一些实施例中,该调节元件是细菌启动子。在一些实施例中,该调节元件是病毒启动子。在一些实施例中,该调节元件是真核生物调节元件,即,真核生物启动子。在一些实施例中,该真核生物调节元件是哺乳动物启动子。

[0261] “可操作地连接”意指目的核苷酸,即编码Cas9蛋白的核苷酸,以允许核苷酸序列表达的方式与调节元件连接。因此,在一些实施例中,该载体是表达载体。

[0262] 在一些实施例中,包含CRISPR-Cas系统的载体的引导多核苷酸由核苷酸序列编码。在一些实施例中,该核苷酸序列是DNA。在一些实施例中,该引导多核苷酸是引导RNA。在一些实施例中,该引导多核苷酸的引导序列是DNA靶向序列。

[0263] 在一些实施例中,该stiCas9和引导多核苷酸能够形成复合物。在一些实施例中,该stiCas9和引导多核苷酸的复合物在自然界中不存在。

[0264] 在一些实施例中,该载体还包含含有tracrRNA序列的核苷酸序列。在一些实施例中,该引导RNA包含crRNA/tracrRNA杂交体。在一些实施例中,该引导RNA的tracrRNA组分激活Cas9蛋白。

[0265] 在一些实施例中,本文所述的CRISPR-Cas系统能够在前间区序列邻近基序的10个核苷酸内的位点处裂解。前间区序列邻近基序,或PAM,是2-6个碱基对的核苷酸序列,位于与引导RNA互补的区域的一个核苷酸内。当Cas9蛋白被激活时(例如,通过与引导多核苷酸形成复合物),它通过与匹配其PAM序列的序列结合来搜索靶DNA。参见,例如,Sternberg等人,“DNA interrogation by the CRISPRRNA-guided endonuclease Cas9[通过CRISPRRNA引导的核酸内切酶Cas9的DNA审查]”,Nature[自然]507(7490):62-67(2014),将其通过引用以其整体并入本文。一旦用合适的PAM识别了潜在的靶序列,并且引导RNA与靶区域正确配对后,Cas9的核酸酶结构域(即,RuvC和HNH结构域)就会切割靶DNA。

[0266] 在一些实施例中,本披露的Cas9蛋白的RuvC和HNH结构域各自切割靶DNA序列的一

条链。在多个实施例中，stiCas9蛋白的RuvC和HNH结构域的切割位点是偏移的，即，每个结构域在其靶DNA相应链上的不同位置处切割，从而造成突出端。在多个实施例中，该stiCas9蛋白的RuvC和HNH结构域以3-核苷酸偏移切割。在多个实施例中，该stiCas9蛋白的RuvC和HNH结构域以4-核苷酸偏移切割。在多个实施例中，该stiCas9蛋白的RuvC和HNH结构域以5-核苷酸偏移切割。在多个实施例中，该stiCas9蛋白的RuvC和HNH结构域以约1、约2、约3、约4、约5、约6、约7、约8、约9、约10、约11、约12、约13、约14、约15、约16、约17、约18、约19、约20、约21、约22、约23、约24、约25、约26、约27、约28、约29、约30、约31、约32、约33、约34、约35、约36、约37、约38、约39或约40个核苷酸的偏移切割。

[0267] 在一些实施例中，本披露的Cas9效应蛋白的RuvC和HNH结构域在双链靶DNA的每条链上的不同位置裂解。在一些实施例中，该Cas9效应蛋白的RuvC结构域在自PAM的约-10、约-9、约-8、约-7或约-6个核苷酸处裂解双链靶DNA的一条链（其可以称为例如“非靶链”），而Cas9效应蛋白的HNH结构域在自PAM的约-5、约-4、约-3、约-2或约-1个核苷酸处裂解双链靶DNA的另一条链（可以称为例如“靶链”）。

[0268] 在一些实施例中，该RuvC结构域在自PAM约-8个核苷酸处裂解双链靶DNA的一条链。在一些实施例中，该RuvC结构域在自PAM约-7个核苷酸处裂解双链靶DNA的一条链。在一些实施例中，该RuvC结构域在自PAM约-6个核苷酸处裂解双链靶DNA的一条链。在一些实施例中，该HNH结构域在自PAM约-4个核苷酸处裂解双链靶DNA的一条链。在一些实施例中，该HNH结构域在自PAM约-3个核苷酸处裂解双链靶DNA的一条链。在一些实施例中，该HNH结构域在自PAM约-2个核苷酸处裂解双链靶DNA的一条链。

[0269] 在本披露的一些实施例中，包含stiCas9和引导多核苷酸的复合物能够在前间区序列邻近基序（PAM）的10个核苷酸内的位点处裂解。在一些实施例中，包含stiCas9和引导多核苷酸的复合物能够在PAM的5个核苷酸内的位点处裂解。在一些实施例中，包含stiCas9和引导多核苷酸的复合物能够在PAM的3个核苷酸内的位点处裂解。在一些实施例中，该PAM在靶序列的下游（即，3' 方向）。在一些实施例中，该PAM在靶序列的上游（即，5' 方向）。在一些实施例中，该PAM位于靶序列内。

[0270] 不同的细菌物种识别不同的PAM序列。鉴定本披露的Cas9蛋白的优选PAM序列的一种方法在图49A中示出，并且包含例如产生与靶序列相邻的各种PAM序列的质粒文库，使该质粒文库与Cas9蛋白接触，然后对该质粒文库进行测序，以确定哪些PAM序列已被“耗尽”（即，在测序结果中未检测到）。“耗尽的”PAM序列是被Cas9蛋白识别和影响（即裂解）的序列。

[0271] 例如，酿脓链球菌的Cas9识别的PAM序列是5' -NGG-3'，其中N是任何核苷酸。不同的PAM与脑膜炎奈瑟氏菌（*Neisseria meningitidis*）、齿垢密螺旋体和嗜热链球菌的Cas9蛋白相关。对新凶手弗朗西斯菌的Cas9蛋白已进行了工程化以识别PAM 5' -YG-3'，其中Y是嘧啶。

[0272] 在一些实施例中，该PAM包含富含3' G的基序。在一些实施例中，该PAM序列是NGG，其中N是A、C、T、U或G。在一些实施例中，该PAM序列是NGA，其中N是A、C、T、U或G。在一些实施例中，该PAM序列是YG，其中Y是嘧啶（即，C、T或U）。

[0273] 在一些实施例中，该靶序列是PAM的5'，并且该PAM包含富含3' G的基序。在一些实施例中，该靶序列是PAM的5'，并且该PAM序列是NGG，其中N是A、C、T、U或G。在一些实施例中，

该靶序列是PAM的5'，该PAM序列是YG，其中Y是嘧啶，并且stiCas9源自细菌物种新凶手弗朗西斯菌。

[0274] 在一些实施例中，该stiCas9包含一个或多个核定位信号。“核定位信号”或“核定位序列” (NLS) 是对蛋白质“加标签”以通过核转运导入细胞核的氨基酸序列，即具有NLS的蛋白质被转运至细胞核。典型地，NLS包含暴露在蛋白质表面的带正电荷的Lys或Arg残基。示例性核定位序列包括但不限于来自以下的NLS：SV40大T抗原、核质蛋白、EGL-13、c-Myc和TUS蛋白。在一些实施例中，该NLS包含PKKKRKV (SEQ ID NO:1) 序列。在一些实施例中，该NLS包含AVKRPAATKKAGQAKKKKLD (SEQ ID NO:2) 序列。在一些实施例中，该NLS包含PAAKRVKLD (SEQ ID NO:3) 序列。在一些实施例中，该NLS包含MSRRRKANPTKLSNAKLAKEVEN (SEQ ID NO:4) 序列。在一些实施例中，该NLS包含KLKIKRPVK (SEQ ID NO:5) 序列。其他核定位序列包括但不限于hnRNP A1的酸性M9结构域、酵母转录抑制子Mata2中的序列KIPIK (SEQ ID NO:6) 和PY-NLS。

[0275] 在一些实施例中，本披露的引导多核苷酸具有与真核细胞中的靶序列杂交的引导序列。在一些实施例中，该真核细胞是动物或人类细胞。在一些实施例中，该真核细胞是人类或啮齿动物或牛细胞系或者细胞株。此类细胞、细胞系或细胞株的实例包括但不限于小鼠骨髓瘤 (NS0) 细胞系、中国仓鼠卵巢 (CHO) 细胞系、HT1080、H9、HepG2、MCF7、MDBK Jurkat、NIH3T3、PC 12、BHK (幼仓鼠肾细胞)、VERO、SP2/0、YB2/0、Y0、C127、L细胞、COS (例如COS1和COS7)、QC1-3、HEK-293、VERO、PER.C6、HeLa、EB1、EB2、EB3、溶瘤或杂交瘤细胞系。在一些实施例中，该真核细胞是CHO细胞系。在一些实施例中，该真核细胞是CHO细胞。在一些实施例中，该细胞是CHO-K1细胞、CHO-K1 SV细胞、DG44 CHO细胞、DUXB 11 CHO细胞、CHOS、CHO GS敲除细胞、CHO FUT8 GS敲除细胞、CHOZN或CHO衍生的细胞。CHO GS敲除细胞 (例如，GSKO细胞) 是例如CHO-K1 SV GS敲除细胞。CHO FUT8敲除细胞是例如Potelligent®CHOK1 SV (龙沙生物公司 (Lonza Biologies, Inc.))。真核细胞也可以是鸟类细胞、细胞系或细胞株，诸如像EBx®细胞、EB14、EB24、EB26、EB66或EBv13。

[0276] 在一些实施例中，该真核细胞是人类细胞。在一些实施例中，该人类细胞是干细胞。干细胞可以是例如多能干细胞，包括胚胎干细胞 (ESC)、成年干细胞、诱导多能干细胞 (iPSC)、组织特异性干细胞 (例如，造血干细胞) 和间充质干细胞 (MSC)。在一些实施例中，该人类细胞是本文描述的任何细胞的分化形式。在一些实施例中，该真核细胞是源自培养物中任何原代细胞的细胞。

[0277] 在一些实施例中，该真核细胞是肝细胞，诸如人肝细胞、动物肝细胞，或非实质细胞。例如，该真核细胞可以是可培养代谢合格的人肝细胞、可培养感应诱导合格的人肝细胞、可培养Qualyst Transporter Certified™人肝细胞、悬浮合格的人肝细胞 (包括10-供体和20-供体合并的肝细胞)、人肝库普弗细胞、人肝星状细胞、狗肝细胞 (包括单个和合并的比格犬肝细胞)、小鼠肝细胞 (包括CD-1和C57BI/6肝细胞)、大鼠肝细胞 (包括Sprague-Dawley、Wistar Han和Wistar肝细胞)、猴肝细胞 (包括食蟹猴或恒河猴肝细胞)、猫肝细胞 (包括家养短毛猫肝细胞) 和兔肝细胞 (包括新西兰白兔肝细胞)。

[0278] 在一些实施例中，该真核细胞是植物细胞。例如，该植物细胞可以是诸如木薯、玉米、高粱、小麦或水稻的农作物的细胞。该植物细胞可以是藻类、树木或蔬菜的细胞。该植物细胞可以是单子叶植物或双子叶植物的细胞，或者可以是农作物或谷物植物、生产植物、水

果或蔬菜的细胞。例如,该植物细胞可以是树木的细胞,该树木为例如柑橘属果树,诸如橘子树、葡萄柚树或柠檬树;桃树或油桃树;苹果树或梨树;坚果树,诸如杏仁树或胡桃树或开心果树;茄属植物,例如,马铃薯、芸苔属(*Brassica*)属植物、莴苣属(*Lactuca*)植物;菠菜属(*Spinacia*)植物;辣椒属(*Capsicum*)植物;棉花、烟草、芦笋、胡萝卜、卷心菜、西兰花、花椰菜、番茄、茄子、胡椒、莴苣、菠菜、草莓、蓝莓、覆盆子、黑莓、葡萄、咖啡、可可等。

[0279] 在一些实施例中,将CRISPR-Cas系统的引导多核苷酸与同向重复序列连接。同向重复序列或DR序列是CRISPR基因座中的重复序列阵列,由短伸展段的非重复序列(间区序列)隔开。该间区序列靶向靶序列上的前间区序列邻近基序(PAM)。当转录CRISPR基因座的非编码部分(即,引导多核苷酸和tracrRNA),转录物在DR序列上被裂解成多个短crRNA,这些crRNA包含单个的间区序列,这些间区序列将Cas9核酸酶引导至PAM。在一些实施例中,该DR序列是RNA。在一些实施例中,该DR序列由核酸编码。在一些实施例中,该DR序列与引导多核苷酸连接。在一些实施例中,该DR序列与引导多核苷酸的引导序列连接。在一些实施例中,该DR序列包含二级结构。在一些实施例中,该DR序列包含茎环结构。在一些实施例中,该DR序列为10至20个核苷酸。在一些实施例中,该DR序列为至少16个核苷酸。在一些实施例中,该DR序列为至少16个核苷酸并且包含单个茎环。在一些实施例中,该DR序列包含RNA适配体。在一些实施例中,该DR中的二级结构或茎环被核酸酶识别以用于裂解。在一些实施例中,该核酸酶是核糖核酸酶。在一些实施例中,该核酸酶是RNA酶III。

[0280] 用于递送CRISPR-Cas系统的各种工具是本领域已知的。在一些实施例中,本披露的CRISPR-Cas系统由递送颗粒递送。递送颗粒是包含颗粒的生物递送系统或配制品。如本文所定义,“颗粒”是最大直径约100微米(μm)的实体。在一些实施例中,该颗粒的最大直径为约10 μm 。在一些实施例中,该颗粒的最大直径为约2000纳米(nm)。在一些实施例中,该颗粒的最大直径为约1000nm。在一些实施例中,该颗粒的最大直径为约900nm、约800nm、约700nm、约600nm、约500nm、约400nm、约300nm、约200nm或约100nm。在一些实施例中,该颗粒的直径为约25nm至约200nm。在一些实施例中,该颗粒的直径为约50nm至约150nm。在一些实施例中,该颗粒的直径为约75nm至约100nm。

[0281] 递送颗粒可以以任何形式提供,包括但不限于:固体、半固体、乳液或胶体颗粒。在一些实施例中,该递送颗粒是基于脂质的系统、脂质体、胶束、微囊泡、外来体或基因枪。在一些实施例中,该递送颗粒包含CRISPR-Cas系统。在一些实施例中,该递送颗粒包含CRISPR-Cas系统,该CRISPR-Cas系统包含stiCas9和引导多核苷酸。在一些实施例中,该递送颗粒包含CRISPR-Cas系统,该CRISPR-Cas系统包含stiCas9和引导多核苷酸,其中stiCas9和引导多核苷酸以复合物存在。在一些实施例中,该递送颗粒包含CRISPR-Cas系统,该CRISPR-Cas系统包含stiCas9、引导多核苷酸和含有tracrRNA的多核苷酸。在一些实施例中,该递送颗粒包含CRISPR-Cas系统,该CRISPR-Cas系统包含stiCas9、引导多核苷酸和tracrRNA。

[0282] 在一些实施例中,该递送颗粒还包含脂质、糖、金属或蛋白质。在一些实施例中,该递送颗粒是脂质包膜。例如,Su等人,“*In vitro and in vivo mRNA delivery using lipid-enveloped pH-responsive polymer nanoparticles*[使用脂质包裹的pH响应聚合物纳米颗粒的体外和体内mRNA递送]”,*Molecular Pharmacology*[分子药理学]8(3):774-784(2011)中描述了使用脂质包膜或包含脂质的递送颗粒进行的mRNA递送。

[0283] 在一些实施例中,该递送颗粒是基于糖的颗粒,例如,GalNAc。基于糖的颗粒描述于WO 2014/118272和Nair等人,Journal of the American Chemical Society[美国化学学会杂志]136(49):16958-16961(2014),将其各自通过引用以其整体并入本文。

[0284] 在一些实施例中,该递送颗粒是纳米颗粒。本披露涵盖的纳米颗粒可以以不同形式提供,例如,作为固体纳米颗粒(例如,金属,诸如银、金、铁、钛)、非金属、基于脂质的固体、聚合物、纳米颗粒或其组合。可以制备金属、电介质和半导体纳米颗粒以及混合结构(例如,核-壳纳米颗粒)。由半导体材料制成的纳米粒子如果足够小(典型地小于10nm),可以量化电子能级,则也可以将其标记为量子点。这样的纳米级颗粒在生物医学应用中用作药物载体或显像剂,并且可以加以调整以适用于本披露中的类似用途。

[0285] 在美国专利公开号2011/0293703、2012/0251560和2013/0302401以及美国专利号5,543,158、5,855,913、5,895,309、6,007,845和8,709,843中进一步描述了递送颗粒的制备,将其各自通过引用以其整体并入本文。

[0286] 在一些实施例中,囊泡包含本披露的CRISPR-Cas系统。“囊泡”是细胞内具有被脂质双层包围的流体的小结构。在一些实施例中,本披露的CRISPR-Cas系统由囊泡递送。在一些实施例中,该囊泡包含stiCas9和引导多核苷酸。在一些实施例中,该囊泡包含stiCas9和引导多核苷酸,其中stiCas9和引导多核苷酸以复合物存在。在一些实施例中,该囊泡包含CRISPR-Cas系统,该CRISPR-Cas系统包含stiCas9、引导多核苷酸和含有tracrRNA的多核苷酸。在一些实施例中,该囊泡包含CRISPR-Cas系统,该CRISPR-Cas系统包含stiCas9、引导多核苷酸和tracrRNA。

[0287] 在一些实施例中,包含stiCas9和引导多核苷酸的囊泡是外泌体或脂质体。在一些实施例中,该囊泡是外泌体。在一些实施例中,该外泌体用于递送本披露的CRISPR-Cas系统。外泌体是内源性纳米囊泡(即,直径为约30nm至约100nm),其转运RNA和蛋白质,并且可将RNA递送至大脑和其他靶器官。例如,Alvarez-Erviti等人,Nature Biotechnology[自然生物技术]29:341(2011),El-Andaloussi等人,Nature Protocols[自然实验手册]7:2112-2116(2012),以及Wahlgren等人,Nucleic Acids Research[核酸研究]40(17):e130(2012)描述了用于将内源性生物材料递送至靶器官的工程化外泌体,将其各自通过引用以其整体并入本文。

[0288] 在一些实施例中,包含stiCas9和引导多核苷酸的囊泡是脂质体。在一些实施例中,该脂质体用于递送本披露的CRISPR-Cas系统。脂质体是具有至少一个脂质双层的球形囊泡结构,并且可以用作营养物和药物施用的媒介物。脂质体通常由磷脂(特别是磷脂酰胆碱)以及其他脂质(诸如蛋磷脂酰乙醇胺)组成。脂质体的类型包括但不限于多层囊泡、小单层囊泡、大单层囊泡和耳蜗囊泡。参见,例如,Spuch和Navarro,“Liposomes for Targeted Delivery of Active Agents against Neurodegenerative Diseases (Alzheimer's Disease and Parkinson's Disease)[用于抗神经退行性疾病(阿尔茨海默病和帕金森病)活性药物的靶向递送的脂质体]”,Journal of Drug Delivery[药物递送杂志]2011,文章ID 469679(2011)。例如,Morrissey等人,Nature Biotechnology[自然生物技术]23(8):1002-1007(2005),Zimmerman等人,Nature Letters[自然快报]441:111-114(2006),以及Li等人,Gene Therapy[基因疗法]19:775-780(2012)描述了用于递送诸如CRISPR-Cas组分的生物材料的脂质体,将其各自通过引用以其整体并入本文。

[0289] 在一些实施例中,编码Cas9和引导多核苷酸的核苷酸在单个载体上。在一些实施例中,编码Cas9的核苷酸、引导多核苷酸(或可以转录成引导多核苷酸的核苷酸)和tracrRNA在单个载体上。在一些实施例中,编码Cas9的核苷酸、引导多核苷酸(或可以转录成引导多核苷酸的核苷酸)、tracrRNA和同向重复序列在单个载体上。在一些实施例中,该载体是表达载体。在一些实施例中,该载体是哺乳动物表达载体。在一些实施例中,该载体是人类表达载体。在一些实施例中,该载体是植物表达载体。

[0290] 在一些实施例中,编码Cas9和引导多核苷酸的核苷酸是单个核酸分子。在一些实施例中,编码Cas9、引导多核苷酸和tracrRNA的核苷酸是单个核酸分子。在一些实施例中,编码Cas9、引导多核苷酸、tracrRNA和同向重复序列的核苷酸是单个核酸分子。在一些实施例中,该单个核酸分子是表达载体。在一些实施例中,该单个核酸分子是哺乳动物表达载体。在一些实施例中,该单个核酸分子是人类表达载体。在一些实施例中,该单个核酸分子是植物表达载体。

[0291] 在一些实施例中,病毒载体包含本披露的CRISPR-Cas系统。在一些实施例中,本披露的CRISPR-Cas系统由病毒载体递送。在一些实施例中,该病毒载体包含stiCas9和引导多核苷酸。在一些实施例中,该病毒载体包含stiCas9和引导多核苷酸,其中该stiCas9和引导多核苷酸以复合物存在。在一些实施例中,该病毒载体包含CRISPR-Cas系统,该CRISPR-Cas系统包含stiCas9、引导多核苷酸和含有tracrRNA的多核苷酸。在一些实施例中,病毒载体包含CRISPR-Cas系统,该CRISPR-Cas系统包含stiCas9、引导多核苷酸和tracrRNA。在一些实施例中,该病毒载体是腺病毒、慢病毒或腺相关病毒载体。本文提供了病毒载体的实例。

[0292] 在一些实施例中,腺相关病毒(AAV)和/或慢病毒载体可以用作包含本文所述的CRISPR-Cas系统的元件的病毒载体。在本披露的一些实施例中,该Cas蛋白由病毒载体转导的细胞在细胞内表达。

[0293] 对于许多治疗策略,包括本披露所设想的那些,可能仅需要瞬时表达Cas蛋白。其结果是,在本披露的一些实施例中,使用非整合型病毒载体将Cas蛋白递送至细胞中。在其他实施例中,需要延长时间的CRISPR-Cas系统组分的表达-例如,当用于永久整合到靶细胞基因组中的基因回路中时。这样的应用已经在Agustín-Pavón等人,“Synthetic biology and therapeutic strategies for the degenerating brain[用于退化性脑的合成生物学和治疗策略]”,Bioessays[生物学分析]36(10):979-990(2014)中进行了讨论,通过引用以其整体并入本文。

[0294] 在一些实施例中,本披露的Cas蛋白和方法用于离体基因编辑,诸如CAR-T型疗法。这些实施例可能涉及对来自人类供体的细胞的修饰。在这些情况下,也可以使用病毒载体;但是,还有其他选择可以直接将Cas蛋白(连同体外转录的引导RNA和供体DNA)转染到培养的细胞中。

[0295] 在一些实施例中,本披露提供了包含CRISPR-Cas系统的真核细胞,该CRISPR-Cas系统包含:(a)能够产生粘性末端的Cas9效应蛋白(stiCas9),以及(b)与该stiCas9形成复合物并且包含引导序列的引导多核苷酸,其中该引导序列能够与真核细胞中的靶序列杂交,其中该复合物在自然界中不存在。在一些实施例中,该真核细胞包含含有本披露的CRISPR-Cas系统的载体。

[0296] 在一些实施例中,该真核细胞是动物或人类细胞。在一些实施例中,该真核细胞是

动物细胞。在一些实施例中,该真核细胞是人类细胞,包括人类干细胞。在一些实施例中,该真核细胞是植物细胞。本文提供了各种类型的真核细胞的实例。

[0297] 在一些实施例中,本披露提供了包含CRISPR-Cas系统的真核细胞,该CRISPR-Cas系统包含能够产生粘性末端的Cas9效应蛋白(stiCas9),其中该Cas9效应蛋白源自具有II-B型CRISPR系统的细菌物种。在一些实施例中,该真核细胞包含stiCas9,该stiCas9包含以至少1E-10、至少1E-9、至少1E-8、至少1E-7,至少1E-6,至少1E-5、至少1E-4、至少1E-3、至少1E-2或至少1E-1的E值截止值匹配TIGR03031蛋白家族的结构域。在一些实施例中,该真核细胞stiCas9,该stiCas9包含与SEQ ID NO:10-97或192-195中的任何一个具有至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%的序列相似性的多肽序列。在一些实施例中,该真核细胞包含stiCas9,该stiCas9包含与SEQ ID NO:10-97或192-195中的任何一个具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或约100%序列同一性的多肽序列。

[0298] 在一些实施例中,本披露的Cas9蛋白是包含一个或多个异源蛋白结构域(例如,除Cas9蛋白之外约或至少约1、2、3、4、5、6、7、8、9或10个或更多个结构域)的融合蛋白的一部分。Cas9融合蛋白可包含任何其他蛋白序列,以及任选地任何两个结构域之间的接头序列。可以与Cas9蛋白融合的蛋白质结构域的实例包括但不限于:具有以下一项或多项活性的表位标签、报告基因序列和蛋白质结构域:甲基化酶活性、脱甲基酶活性、转录激活活性、转录抑制活性、转录释放因子活性、组蛋白修饰活性、RNA裂解活性和核酸结合活性。表位标签的非限制性实例包括:组氨酸(His)标签、V5标签、FLAG标签、流感病毒血凝素(HA)标签、Myc标签、VSV-G标签和硫氧还蛋白(Trx)标签。报告基因的实例包括但不限于谷胱甘肽-S-转氨酶(GST)、辣根过氧化物酶(HRP)、氯霉素乙酰转移酶(CAT)、 β -半乳糖苷酶、 β -葡萄糖醛酸酶、萤光素酶、绿色荧光蛋白(GFP)、HcRed、DsRed、青色荧光蛋白(CFP)、黄色荧光蛋白(YFP)、自发荧光蛋白(包括蓝色荧光蛋白(BFP))和mCherry。在一些实施例中,Cas9蛋白与结合DNA分子或结合其他细胞分子的蛋白或蛋白片段融合,包括但不限于:麦芽糖结合蛋白(MBP)、S标签、Lex A DNA结合结构域(DBD)、GAL4 DNA结合结构域和单纯疱疹病毒(HSV)BP16蛋白。可形成包含Cas9蛋白的融合蛋白的一部分的其他结构域在通过引用以其全文并入本文的US20110059502中描述。在一些实施例中,加标签的Cas9蛋白用于鉴定靶序列的位置。

[0299] 在一些实施例中,Cas9蛋白可形成诱导型系统的组分。该系统的可诱导性质允许使用某种形式的能量对基因编辑或基因表达进行时空控制。这种能量形式可以包括但不限于:电磁辐射、声能、化学能和热能。诱导型系统的非限制性实例包括:四环素诱导型启动子(Tet-On或Tet-Off)、小分子双杂交转录激活系统(FKBP、ABA等)或光诱导系统(光敏色素、LOV结构域或隐色素)。在一些实施例中,该Cas9蛋白是光诱导型转录效应子(LITE)的一部分,以序列特异性方式指导转录活性的变化。光的成分可以包括Cas9蛋白、光响应性细胞色素异二聚体(例如,来自拟南芥(*Arabidopsis thaliana*))和转录激活/抑制结构域。诱导型DNA结合蛋白及其使用方法的其他实例在国际申请公开号W02014/018423和W0 2014/093635;美国专利号8,889,418和8,895,308;以及美国专利公开号2014/0186919、2014/0242700、2014/0273234和2014/0335620中提供;将其各自通过引用以其整体并入本文。

[0300] 位点特异性修饰的方法

[0301] 在一些实施例中,本披露提出了一种用于在真核细胞中提供靶序列的位点特异性修饰的方法,该方法包括:(1)将以下引入到细胞中:(a)能够产生粘性末端的Cas9效应蛋白(stiCas9),和(b)与stiCas9形成复合物并且包含引导序列的引导多核苷酸,其中该引导序列能够与真核细胞中的靶序列杂交,但不与细菌细胞中的序列杂交,其中该复合物在自然界中不存在;(2)用Cas9效应蛋白和引导多核苷酸在靶序列中产生粘性末端;以及(3):将(a)粘性末端连接在一起,或(b)将目的多核苷酸序列(SoI)连接到该粘性末端,从而修饰靶序列。

[0302] 靶序列的修饰涵盖单核苷酸取代、多核苷酸取代、核酸的插入(即,敲入)和缺失(即,敲除)、移码突变和其他核酸修饰。

[0303] 在一些实施例中,该修饰是靶序列的至少一部分的缺失。可以在两个不同位点裂解靶序列并产生互补的粘性末端,并且这些互补的粘性末端可以重新连接,从而去除两个位点之间的序列部分。

[0304] 在一些实施例中,该修饰是靶序列的突变。真核细胞中的位点特异性诱变是通过使用位点特异性核酸酶来实现的,该酶可促进含有目的突变的外源多核苷酸模板(也称为“供体多核苷酸”或“供体载体”)的同源重组。在一些实施例中,目的序列(SoI)包含目的突变。

[0305] 在一些实施例中,该修饰是将目的序列(SoI)插入靶序列中。该SoI可以作为外源多核苷酸模板引入。在一些实施例中,该外源多核苷酸模板包含粘性末端。在一些实施例中,该外源多核苷酸模板包含与靶序列中的粘性末端互补的粘性末端。

[0306] 外源多核苷酸模板可以具有任何合适的长度,诸如约或至少约10、15、20、25、50、75、100、150、200、250、500或1000或更多个核苷酸的长度。在一些实施例中,该外源多核苷酸模板与包含靶序列的多核苷酸的一部分互补。当进行最佳比对时,该外源多核苷酸模板与靶序列的一个或多个核苷酸(例如,约或至少约1、5、10、15、20、25、30、35、40、45、50、60、70、80、90或100或更多个核苷酸)重叠。在一些实施例中,当对外源多核苷酸模板和包含靶序列的多核苷酸进行最佳比对时,外源多核苷酸模板最接近的核苷酸在距离靶序列约1、5、10、15、20、25、50、75、100、200、300、400、500、100、1500、2000、2500、5000、10000个或更多个核苷酸内。

[0307] 在一些实施例中,该外源多核苷酸是DNA,例如,DNA质粒、细菌人工染色体(BAC)、酵母人工染色体(YAC)、病毒载体、单链或双链DNA的线性片段、寡核苷酸、PCR片段、裸核酸或与诸如脂质体的递送媒介物复合的核酸。

[0308] 在一些实施例中,使用细胞的内源DNA修复途径将该外源多核苷酸插入靶序列。内源性DNA修复途径包括非同源末端连接(NHEJ)途径、微同源介导的末端连接(MMEJ)途径和同源定向修复(HDR)途径。NHEJ、MMEJ和HDR途径可修复双链DNA断裂。在NHEJ中,修复DNA中的断裂不需要同源模板。NHEJ修复可能容易出错,但是当DNA断裂包含相容的突出端时,错误会减少。NHEJ和MMEJ是在机理上截然不同的DNA修复途径,它们各自涉及DNA修复酶的不同子集。与可能既精确又容易出错的NHEJ不一样,MMEJ总是容易出错,并且会导致在修复位点处的缺失和插入。MMEI相关的缺失归因于双链断裂两侧的微同源性(2-10个碱基对)。相比之下,HDR需要同源模板来直接进行修复,但是HDR修复典型地具有高保真度,并且不易出错。在一些实施例中,利用NHEJ和MMEJ修复的容易出错的性质在靶序列中引入非特异性核

核苷酸取代。在一些实施例中,stiCas9以促进HDR修复的方式切割靶序列。

[0309] 在修复过程中,可以将包含SoI的外源多核苷酸模板引入该靶序列。在一些实施例中,将包含侧接上游序列和下游序列的SoI的外源多核苷酸模板引入细胞中,其中该上游和下游序列与靶序列中整合位点的任一侧具有序列相似性。在一些实施例中,该包含SoI的外源多核苷酸包含例如突变基因。在一些实施例中,该外源多核苷酸包含对于该细胞而言是内源性或外源性的序列。在一些实施例中,该SoI包含编码蛋白质的多核苷酸,或非编码序列,诸如像微小RNA。在一些实施例中,该SoI可操作地与调节元件连接。在一些实施例中,该SoI是调节元件。在一些实施例中,该SoI包含抗性盒,例如,赋予对抗生素抗性的基因。在一些实施例中,该SoI包含野生型靶序列的突变。在一些实施例中,该SoI通过产生移码突变或核苷酸取代来破坏或校正靶序列。在一些实施例中,该SoI包含标记。将标记引入靶序列可以便于筛选靶向的整合。在一些实施例中,该标记是限制性位点、荧光蛋白或选择性标记。在一些实施例中,将该SoI作为包含SoI的载体引入。

[0310] 选择该外源多核苷酸模板中的上游和下游序列以促进靶序列和外源多核苷酸之间的同源重组。该上游序列是与用于整合的靶向位点的上游序列(靶序列)具有序列相似性的核酸序列。类似地,该下游序列是与用于整合的靶位点的下游序列具有序列相似性的核酸序列。因此,在一些实施例中,通过在上游和下游序列处的同源重组,将包含SoI的外源多核苷酸模板插入靶序列中。在一些实施例中,该外源多核苷酸模板中的上游和下游序列与靶向的基因组序列的上游和下游序列分别具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、分别至少97%、至少98%、至少99%或100%的序列同一性。在一些实施例中,该上游或下游序列具有约20至2000个碱基对、或约50至1750个碱基对、或约100至1500个碱基对、或约200至1250个碱基对、或约300至1000个碱基对、或约400至约750个碱基对、或约500至600个碱基对。在一些实施例中,该上游或下游序列具有约50、约100、约250、约500、约1000、约1250、约1500、约1750、约2000、约2250或约2500个碱基对。

[0311] 在一些实施例中,靶序列中的修饰是细胞中靶序列表达的失活。例如,在CRISPR复合物与靶序列结合后,靶序列即失活,使得该序列不被转录,不产生编码的蛋白质,或者该序列不像野生型序列那样起作用。例如,蛋白质或微小RNA编码序列可能会失活,从而不产生蛋白质。

[0312] 在一些实施例中,调节序列可能会失活,使得它不再作为调节序列起作用。调节序列的实例包括启动子、转录终止子、增强子和本文所述的其他调节元件。失活的靶序列可以包括缺失突变(即,一个或多个核苷酸的缺失)、插入突变(即,一个或多个核苷酸的插入)或无义突变(即,用另一个核苷酸取代一个单核苷酸从而引入终止密码子)。在一些实施例中,靶序列的失活导致靶序列的“敲除”。

[0313] 在一些实施例中,该stiCas9和引导多核苷酸形成复合物,并且该引导多核苷酸与待修饰的靶序列杂交。在一些实施例中,该stiCas9在与引导多核苷酸杂交的靶序列中产生粘性末端。

[0314] 在该方法的多个实施例中,由stiCas9产生的粘性末端包含具有3至40个核苷酸的单链多核苷酸突出端。在一些实施例中,由该stiCas9产生的粘性末端包含具有4至20个核苷酸的单链多核苷酸突出端。在一些实施例中,由该stiCas9产生的粘性末端包含具有5至15个核苷酸的单链多核苷酸突出端。在一些实施例中,由该stiCas9产生的粘性末端是5'突

出端。

[0315] 在该方法的多个实例中,stiCas9源自具有II-B型CRISPR系统的细菌物种。如本文所讨论的,II-B型Cas9蛋白属于TIGR03031 TIGRFAM蛋白家族。因此,在一些实施例中,本披露的stiCas9包含以 $1E-5$ 谱截止值匹配TIGR03031蛋白家族的结构域。在一些实施例中,本披露的stiCas9包含以 $1E-10$ 谱截止值匹配TIGR03031蛋白家族的结构域。在一些实施例中,本披露的stiCas9包含以至少 $1E-10$ 、至少 $1E-9$ 、至少 $1E-8$ 、至少 $1E-7$,至少 $1E-6$,至少 $1E-5$ 、至少 $1E-4$ 、至少 $1E-3$ 、至少 $1E-2$ 或至少 $1E-1$ 的E值截止值匹配TIGR03031蛋白家族的结构域。

[0316] 在该方法的多个实施例中,II-B型Cas9源自具有II-B型CRISPR系统的任何物种。在一些实施例中,该II-B型Cas9源自以下细菌物种:嗜肺军团菌、新凶手弗朗西斯菌、 γ 变形菌HTCC5015、人粪便副萨特氏菌、华德萨特氏菌、硫磺单胞菌属物种SCADC、瘤胃杆菌属物种RM87、伯克霍尔德氏菌目细菌1_1_47、拟杆菌门口腔分类群274菌株F0058、产琥珀酸沃廉氏菌、伯克霍尔德氏菌目细菌YL45、嗜淀粉瘤胃杆菌、弯曲杆菌属物种P0111、弯曲杆菌属物种RM9261、拉尼尔弯曲杆菌菌株RM8001、拉尼尔弯曲杆菌菌株P0121、鼠毛滴虫、伦敦军团菌、沙姆盐弧菌、钩端螺旋体属物种分离株FW.030、莫里特拉氏菌属物种分离株NORP46、内生单胞菌属物种S-B4-1U、喜盐泰米尔纳德菌、需钠弧菌、斯氏弓形杆菌、蜃楼弗朗西斯氏菌、西班牙弗朗西斯氏菌或嗜盐副内生单胞菌。

[0317] 在该方法的多个实施例中,该引导多核苷酸是引导RNA。在一些实施例中,该引导多核苷酸包含至少两个核苷酸区段:至少一个“DNA结合区段”或“引导序列”和至少一个“多肽结合区段”。在一些实施例中,该引导多核苷酸的DNA结合区段与真核细胞中的靶序列杂交,但不与细菌细胞中的序列杂交。在一些实施例中,该引导多核苷酸的多肽结合区段与Cas9结合。在一些实施例中,该引导多核苷酸的多肽结合区段与stiCas9结合。

[0318] 在该方法的多个实施例中,该引导多核苷酸是10至35个核苷酸。在一些实施例中,该引导多核苷酸是15至30个核苷酸。在一些实施例中,该引导多核苷酸是20至25个核苷酸。

[0319] 在该方法的多个实施例中,该stiCas9和引导多核苷酸能够形成复合物。在一些实施例中,当复合物的所有组分一起存在时形成复合物,即自组装复合物。在一些实施例中,通过复合物的不同组分之间的化学相互作用(诸如像氢键结合)形成复合物。在一些实施例中,通过stiCas9对引导多核苷酸的二级结构识别,该引导多核苷酸与该stiCas9形成复合物。在一些实施例中,stiCas9蛋白是无活性的,即,不展现出核酸酶活性,直到它与引导多核苷酸形成复合物。引导RNA的结合诱导stiCas9的构象变化,以将stiCas9从无活性形式转化为活性(即,催化活性)形式。在该方法的多个实例中,该stiCas9和引导多核苷酸的复合物在自然界中不存在。

[0320] 在该方法的实施例中,由stiCas9产生的这些粘性末端被连接在一起(即,以化学方式连接在一起)。连接可以例如通过DNA连接酶(诸如T4连接酶或DNA连接酶IV)进行。在一些实施例中,使用引入一个或多个核苷酸取代的易错连接酶将这些粘性末端连接在一起。在一些实施例中,将目的多核苷酸序列(SoI)与这些粘性末端连接。在一些实施例中,该SoI包含目的突变。

[0321] 在该方法的多个实施例中,在SoI中产生与在靶序列中产生的粘性末端互补的粘性末端。在一些实施例中,SoI中的粘性末端由stiCas9产生。在一些实施例中,使用细胞的内源性DNA修复途径将SoI连接到粘性末端中。本文描述了多个内源性DNA修复途径。

[0322] 在一些实施例中,本披露提供了一种用于在真核细胞中提供靶序列的位点特异性修饰的方法,该方法包括:(1)将以下引入到细胞中:(a)编码能够产生粘性末端的Cas9效应蛋白(stiCas9)的核苷酸序列,和(b)与stiCas9形成复合物并且包含引导序列的引导多核苷酸,其中该引导序列能够与真核细胞中的靶序列杂交,但不与细菌细胞中的序列杂交,其中该复合物在自然界中不存在;(2)用Cas9效应蛋白和引导多核苷酸在靶序列中产生粘性末端;以及(3):将(a)粘性末端连接在一起,或(b)将目的多核苷酸序列(SoI)连接到该粘性末端,从而修饰靶序列。

[0323] 在该方法的多个实例中,该stiCas9由核苷酸序列编码。在一些实施例中,该核苷酸是DNA。在一些实施例中,该stiCas9蛋白包含与SEQ ID NO:10-97或192-195中任一项的核苷酸序列具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或约100%同一性的序列的结构域。

[0324] 在该方法的多个实例中,本披露的CRISPR-Cas系统进一步包含tracrRNA。在一些实施例中,该引导RNA包含crRNA/tracrRNA杂交体。在一些实施例中,该引导RNA的tracrRNA组分激活Cas9蛋白。在该方法的多个实施例中,该stiCas9、引导多核苷酸和tracrRNA能够形成复合物。在一些实施例中,该stiCas9、引导多核苷酸和tracrRNA的复合物在自然界中不存在。

[0325] 在该方法的多个实施例中,包含stiCas9和引导多核苷酸的复合物能够在前间区序列邻近基序(PAM)的10个核苷酸内的位点处裂解。在一些实施例中,包含stiCas9和引导多核苷酸的复合物能够在PAM的5个核苷酸内的位点处裂解。在一些实施例中,包含stiCas9和引导多核苷酸的复合物能够在PAM的3个核苷酸内的位点处裂解。在一些实施例中,该PAM在靶序列的下游(即,3'方向)。在一些实施例中,该PAM在靶序列的上游(即,5'方向)。在一些实施例中,该PAM位于靶序列内。

[0326] 在该方法的多个实施例中,该PAM包含富含3' G的基序。在一些实施例中,该PAM序列是NGG,其中N是A、C、T、U或G。在一些实施例中,该PAM序列是NGA,其中N是A、C、T、U或G。在一些实施例中,该PAM序列是YG,其中Y是嘧啶(即,C、T或U)。在该方法的多个实施例中,该靶序列是PAM的5',并且该PAM包含富含3' G的基序。在一些实施例中,该靶序列是PAM的5',并且该PAM序列是NGG,其中N是A、C、T、U或G。

[0327] 在该方法的多个实施例中,该真核细胞是动物或人类细胞。在一些实施例中,该真核细胞是动物细胞。在一些实施例中,该真核细胞是人类细胞,包括人类干细胞。在一些实施例中,该真核细胞是植物细胞。本文提供了各种类型的真核细胞的实例。在该方法的多个实施例中,通过递送颗粒将stiCas9和引导多核苷酸引入真核细胞中。在该方法的多个实施例中,通过囊泡将stiCas9和引导多核苷酸引入真核细胞中。在该方法的多个实施例中,通过载体将stiCas9和引导多核苷酸引入真核细胞中。在该方法的多个实施例中,通过病毒载体将stiCas9和引导多核苷酸引入真核细胞中。在该方法的多个实施例中,将编码包含stiCas9和引导多核苷酸的复合物的组分的多核苷酸引入一个或多个载体上。本文提供了载体的实例和载体递送入细胞(例如,转染)的方法。

[0328] 在一些实施例中,本披露的方法进一步包括将核酸外切酶引入真核细胞中以去除由stiCas9产生的突出端。在一些实施例中,该核酸外切酶是5'至3'核酸外切酶。在一些实施例中,该核酸外切酶是3'至5'核酸外切酶。在一些实施例中,在该方法的连接步骤之前添

加核酸外切酶。在一些实施例中,添加核酸外切酶以替代该方法的连接步骤。5'至3'核酸外切酶的非限制性实例包括: λ 核酸外切酶、RecJ、核酸外切酶V、核酸外切酶VIII、T5核酸外切酶、T7核酸外切酶、Artemis和Cas4。3'至5'核酸外切酶的非限制性实例包括:TREX1、TREX2、Werner综合征(WRN)蛋白、p53、MRE11、RAD1、RAD9、APE1和VDJP蛋白。在一些实施例中,该核酸外切酶是Cas4、Artemis或TREX2。

[0329] Cas4、Artemis、TREX2或其他类似的核酸外切酶的引入可以在连接之前对粘性末端进行末端加工,从而减少精确连接的机会,并因此提高诱变效率,与内源性DNA修复酶竞争以将修复偏向一个其他修复途径(例如,NHEJ或MMEJ),并调节突变模式。例如,Cas4、Artemis或TREX2可以通过与内源性末端加工酶竞争来提高诱变效率,从而促进易于出错的修复。Cas4、Artemis或TREX2还可以通过延长单链突出端来促进HDR修复。Cas4、Artemis或TREX2的其他作用可能涉及,例如,将突变模式改变为更理想的插入缺失。

[0330] 位点特异性基因插入的方法(ObLiGaRe 2.0)

[0331] 在一些实施例中,本披露提供了基于美国专利号9,567,608中描述的ObLiGaRe方法的衍生,将目的序列(SoI)引入细胞的染色体中的方法。ObLiGaRe(专性连接门控重组)反映了拉丁语动词obligare(头对头连接)的词源意义。它广泛适用于不同的细胞系,并为基因工程提供了另一种方法。鉴于美国专利号9,567,608采用锌指核酸酶来靶向和裂解靶序列,本文的披露内容提供了第一Cas9-核酸内切酶二聚体(例如,Cas9-FokI)和第二Cas9-核酸内切酶二聚体的用途。本文所述用于位点特异性基因插入的方法被非正式地简称为“ObLiGaRe 2.0”,以使其与美国专利号9,567,608中描述的ObLiGaRe方法相区分。

[0332] 在一些实施例中,本披露提供了一种将目的序列(SoI)引入细胞的染色体中的方法,其中该染色体包含含有区域1和区域2的靶序列(TSC),该方法包括将以下引入细胞中:(a)包含靶序列的载体(TSV),该TSV包含区域2和区域1以及该SoI;(b)能够在TSC中产生粘性末端的第一Cas9-核酸内切酶二聚体,其中该第一Cas9-核酸内切酶二聚体的第一单体在该TSC的区域1处裂解,并且该第一Cas9-核酸内切酶二聚体的第二单体在该TSC的区域2处裂解;以及(c)能够在TSV中产生粘性末端的第二Cas9-核酸内切酶二聚体,其中该第二Cas9-核酸内切酶二聚体的第一单体在该TSV的区域2处裂解,并且该第二Cas9-核酸内切酶二聚体的第二单体在该TSV的区域1处裂解,并且其中引入(a)的载体、(b)的第一Cas9-核酸内切酶二聚体和(c)的第二Cas9-核酸内切酶二聚体导致SoI被插入细胞的染色体中。

[0333] 在一些实施例中,本披露涉及将目的序列(SoI)引入细胞的染色体中的方法,其中该染色体包含含有区域1和区域2的靶序列(TSC),该方法包括将以下引入细胞中:(a)包含靶序列的载体(TSV),该TSV包含区域2和区域1以及该SoI,其中该载体包含粘性末端;以及(b)能够在TSC中产生粘性末端的第一Cas9-核酸内切酶二聚体,其中该第一Cas9-核酸内切酶二聚体的第一单体在该TSC的区域1处裂解,并且该第一Cas9-核酸内切酶二聚体的第二单体在该TSC的区域2处裂解;其中引入(a)的载体和(b)的第一Cas9-核酸内切酶二聚体导致SoI被插入细胞的染色体中。

[0334] 本披露的方法在无需载体(或“供体质粒”)同源性的情况下提供了有效且精确的基因靶向。本披露的方法提供了使用非同源末端连接(NHEJ)或微同源介导的末端连接(MMEJ)途径的位点特异性基因插入的策略。载体中裂解位点(即区域1和区域2)的设计和位置足以实现载体在基因组位点(即细胞染色体(TSC)中的靶序列)中的裂解位点(即区域1和

区域2)中的精确的末端连接。

[0335] 在一些实施例中,该TSV是环状载体,即,质粒。在一些实施例中,该TSV是线性化的载体或线性DNA,诸如像PCR产物,或者裂解后具有与TSC互补的末端的退火寡核苷酸双链体。在一些实施例中,该TSV包括粘性末端。在一些实施例中,该TSV中的这些粘性末端由Cas9-核酸内切酶二聚体产生。在一些实施例中,该TSV中的粘性末端在将TSV引入细胞中之前产生。在一些实施例中,该TSV中的粘性末端在将TSV引入细胞中之后产生。

[0336] 在一些实施例中,染色体上的靶序列(TSC)以5'至3'的方式包含区域1和区域2。如本文所用,序列的方向性(例如5'至3')是指读取双链DNA序列的“编码”链或“正义”链(典型地表示为双链DNA序列的顶部链)时的方向。

[0337] 图12描绘本披露的一个实施例。在图12中,该TSC由“基因组”框(左)中的序列表示,并且包括:“编码”链(显示为顶部链)上的区域1和区域2(其一部分与区域1重叠)。

[0338] 如图12的“基因组”框中所示,在区域1的上游(即,相对于编码链5')和“非编码”或“反义”DNA链(显示为底部链)上,存在第一PAM序列。该非编码链包含与第一引导多核苷酸(“gRNA1”)杂交的区域。gRNA1与第一PAM序列的上游序列(即,相对于非编码链5')杂交。该gRNA1杂交序列包括区域1的一部分,并且另外包括区域1以外的几个核苷酸。如箭头方向所示,gRNA1与靶序列的非编码链杂交。

[0339] 如图12的“基因组”框中所示,在区域2的下游(即,相对于编码链3')和编码链上,存在第二PAM序列。该编码链包含与第二引导多核苷酸(“gRNA2”)杂交的区域。gRNA2与第二PAM序列的上游序列(即,相对于编码链5')杂交。该gRNA2杂交序列包括区域2的一部分,并且另外包括区域2以外的几个核苷酸。如箭头方向所示,gRNA2与靶序列的编码链杂交。

[0340] 在一些实施例中,载体上的靶序列(TSV)以5'至3'的方式包含区域2,紧接着后是区域1以及SoI。图12描绘本披露的一个实施例。在图12中,该TSV由“载体”框(右)中的序列表示,并且包括:“编码”链上的区域2,随后是区域1(两个区域之间没有任何重叠)。

[0341] 如图12的“载体”框中所示,在区域2的上游(即,相对于编码链5')和在“非编码”区域,存在第三PAM序列。该非编码链包含与第三引导多核苷酸(gRNA3)杂交的区域。gRNA3与第三PAM序列的上游序列(即,相对于非编码链5')杂交。该gRNA3杂交序列包括区域2的一部分,并且另外包括区域2以外的几个核苷酸。如箭头方向所示,gRNA3与靶序列的非编码链杂交。

[0342] 如图12的“载体”框中所示,在区域1的下游(即,相对于编码链3')和编码链上,存在第四PAM序列。该编码链包含与第四引导多核苷酸(“gRNA4”)杂交的区域。gRNA4与第四PAM序列的上游序列(即,相对于编码链5')杂交。该gRNA4杂交序列包括区域1的一部分,并且另外包括区域1以外的几个核苷酸。如箭头方向所示,gRNA4与靶序列的编码链杂交。

[0343] 图14描绘本披露的另一个实施例。图14与图14相似,不同的是在TSC上的区域1和区域2之间存在几个核苷酸的间隙,并且在TSV上的区域2和区域1之间存在几个核苷酸的间隙。然而,图14和图12中这些区域相对于彼此的排列以及引导多核苷酸的方向性相同。

[0344] 因此,在一些实施例中,染色体上的靶序列(即,TSC)包括区域1和区域2,其中区域1的一部分与区域2的一部分重叠。在其他实施例中,该TSC包含区域1和区域2,其中区域1和区域2被一个或多个核苷酸分隔。在一些实施例中,区域1和区域2有1、2、3、4、5、6、7、8、9、10或更多个核苷酸重叠。在一些实施例中,区域1和区域2被1、2、3、4、5、6、7、8、9、10或更多个

核苷酸分隔。

[0345] 在一些实施例中,载体上的靶序列(即,TSV)包含区域2和区域1,其中区域2紧接在区域1之前,而两者之间没有任何核苷酸。在其他实施方案中,该TSV包含区域2和区域1,其中区域2和区域1被1个或更多个核苷酸分隔。在一些实施例中,区域2和区域1被1、2、3、4、5、6、7、8、9、10或更多个核苷酸分隔。

[0346] 在该方法的多个实施例中,Cas9-核酸内切酶二聚体在靶序列中产生粘性末端。如本文所述,Cas9蛋白产生核酸中的位点特异性断裂。在一些实施例中,Cas9蛋白产生DNA中的位点特异性双链断裂。Cas9靶向核酸中特定序列的能力(即,位点特异性)是通过Cas9与引导多核苷酸(例如,引导RNA)复合而实现的,其中该引导多核苷酸与指定序列杂交。因此,包含Cas9和引导多核苷酸的复合物具有至少两种不同的功能:(1)特异性靶向核酸序列,和(2)在靶向的核酸序列处或附近产生断裂的核酸酶活性。在一些实施例中,Cas9-引导多核苷酸复合物经过修饰,使得它仅执行两种功能之一。在一些实施例中,Cas9经过修饰以去除核酸酶活性,但是保留与引导多核苷酸复合的能力,使得Cas9仍可以靶向特定核酸序列。

[0347] 如本文所述,野生型Cas9是包含核酸结合结构域(该结构域与引导多核苷酸相互作用)和裂解结构域(该结构域裂解靶核酸)的单体蛋白。在某些情况下,使用二聚核酸酶(即,直到靶序列存在二聚体的两个单体之前才有活性的核酸酶)更有利,以实现更高的靶向特异性。天然存在的核酸酶(诸如像Cas9)的结合结构域和裂解结构域,以及可以融合以产生核酸酶结合特异性靶位点的模块结合结构域与裂解结构域是本领域技术人员所熟知的。例如,RNA可编程核酸酶(例如,Cas9)的结合结构域或具有非活性DNA裂解结构域的Cas9蛋白可以用作与希望的靶位点特异性结合的结合结构域(例如,结合gRNA以直接与靶位点结合),并与裂解结构域(例如,核酸内切酶FokI的裂解结构域)融合或缀合,以产生裂解靶位点的工程化核酸酶。Cas9-FokI融合蛋白进一步描述于例如美国专利公开号2015/0071899和Guilinger等人,“Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification[无催化活性的Cas9与FokI核酸酶的融合改善了基因组修饰的特异性]”,Nature Biotechnology[自然生物技术] 32:577-582(2014),将其各自通过引用以其整体并入本文。

[0348] 在一些实施例中,该工程化的核酸酶可识别回文、双链靶位点,例如,双链DNA靶位点。许多天然存在的核酸酶的靶位点,诸如像天然存在的DNA限制性核酸酶,是本领域技术人员熟知的。在一些实施例中,DNA核酸酶,诸如像EcoRI、HindIII或BamHI,可识别长度为4至10个碱基对的回文、双链DNA靶位点,并在该靶位点内的特定位置切割这两条DNA链的每条链。在一些实施例中,核酸内切酶对称地切割双链核酸靶位点,即,在同一位置切割两条链,使得末端包含碱基配对的核苷酸,在本文中也称为平端。在一些实施例中,核酸内切酶不对称地切割双链核酸靶位点,即在不同位置切割每条链,使得末端包含未配对的核苷酸,即,粘性末端或突出端。在一些实施例中,这些突出端是5'-突出端,即未配对的核苷酸形成DNA链的5末端。在一些实施例中,这些突出端是3'-突出端,即未配对的核苷酸形成DNA链的3末端。突出端可以“粘附”于(即,连接)包含互补的未配对核苷酸的其他双链DNA分子末端。

[0349] 在一些实施例中,提供了包含两个结构域的融合蛋白:与(ii)核酸酶结构域融合或连接的(i)RNA可编程核酸酶(例如,Cas9蛋白或其片段)结构域。例如,在一些实施例中,该Cas9蛋白(例如融合蛋白的Cas9结构域)包含核酸酶灭活的Cas9(例如,缺乏DNA裂解活性

的Cas9;“dCas9”),该Cas9保留了RNA(gRNA)结合活性,因此能够结合与gRNA互补的靶位点。在一些实施例中,与核酸酶灭活的Cas9结构域融合的核酸酶是需要二聚化(例如,核酸酶的两个单体结合在一起)以便裂解靶核酸(例如,DNA)的任何核酸酶。在一些实施例中,与核酸酶灭活的Cas9融合的核酸酶是FokI DNA裂解结构域的单体,从而产生称为Cas9-FokI的Cas9变体。该FokI DNA裂解结构域是已知的,并且在多个实施例中对应于FokI的氨基酸388-583(NCBI登录号J04623)。在一些实施例中,该FokI DNA裂解结构域对应于FokI的氨基酸300-583、320-583、340-583或360-583。(另请参见Wah等人,“Structure of FokI has implications for DNA cleavage[FokI的结构对DNA裂解有影响]”,Proceedings of the National Academy of Sciences USA[美国国家科学院院刊]95(18):10564-9(1996);Li等人,“TAL nucleases(TALNs):hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain[TAL核酸酶(TALN):由TAL效应子和FokI DNA裂解结构域组成的杂交蛋白]”,Nucleic Acids Research[核酸研究]39(1):359-72(2011);Kim等人,“Hybrid restriction enzymes:zinc finger fusions to FokI cleavage domain[杂交限制性酶:与FokI裂解结构域的锌指融合]”,Proceedings of the National Academy of Sciences USA[美国国家科学院院刊]93:1156-1160(1996);将其各自通过引用以其整体并入本文。)

[0350] 在一些实施例中,提供了Cas9-核酸内切酶融合蛋白的二聚体,例如,Cas9-FokI的二聚体。例如,在一些实施例中,该Cas9-FokI融合蛋白与其自身形成二聚体以介导靶核酸的裂解。在一些实施例中,该Cas9-核酸内切酶融合蛋白或其二聚体与一种或多种gRNA相缔合。在一些实施例中,因为该二聚体包含两种融合蛋白,每种融合蛋白含有具有gRNA结合活性的Cas9结构域,所以使用与核酸靶标的两个不同区域互补的两个不同的gRNA序列靶向靶核酸。参见,例如,图10和11。因此,在一些实施例中,该靶核酸的裂解在两种融合蛋白与靶核酸结合(例如,如由gRNA:靶核酸碱基配对所指定的)前不会发生,并且这些核酸酶结构域二聚化(例如,FokI DNA裂解结构域;作为基于融合蛋白的Cas9:gRNA结构域的结合而使它们邻近的结果)并且裂解靶核酸,例如,在结合的Cas9融合蛋白之间的区域中。这通过图10和图11所示的示意图来例示。这种方法代表了相对野生型Cas9和其他Cas9变体(诸如切口酶)的显著改进(Ran等人,“Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity[通过RNA引导的CRISPR Cas9进行双切口以实现增强的基因组编辑特异性]”,Cell[细胞]154:1380-1389(2013);Mali等人,“CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering[用于靶标特异性筛选的CAS9转录激活剂和用于协作基因组工程的配对切口酶]”,Nature Biotechnology[自然生物技术]31:833-838(2013)),该方法不需要核酸酶结构域的二聚化即可裂解核酸。这些切口酶变体可以诱导裂解或在单个切口酶与核酸结合后产生切口,这可以在靶上和脱靶位点处进行,并且已知切口可以诱导突变形。由于本文提供的变体需要彼此邻近的两个Cas9变体的结合以诱导靶核酸裂解,从而降低了脱靶裂解的机会。在一些实施例中,与核酸酶结构域融合的Cas9变体(例如,Cas9-FokI)具有比野生型Cas9或其他Cas9变体(例如,切口酶)的中靶:脱靶修饰比高至少2倍、至少5倍、至少10倍、至少20倍、至少30倍、至少40倍、至少50倍、至少60倍、至少70倍、至少80倍、至少90倍、至少100倍、至少110倍、至少120倍、至少130倍、至少140倍、至少150倍、至少175倍、至少200倍、至少250倍或更高的中靶:脱靶修饰比。在一些实施例中,与核酸酶

结构域融合的Cas9变体(例如,Cas9-FokI)具有比野生型Cas9或其他Cas9变体的中靶:脱靶修饰比高约60至180倍之间、约80至160倍之间、约100至150倍之间或约120至140倍之间的中靶:脱靶修饰比。确定中靶:脱靶修饰比的方法是已知的。在一些实施例中,通过测量某些基因中已知Cas9脱靶位点的修饰的数目或量来确定中靶:脱靶修饰比。例如,CLTA、EMX和VEGF基因的Cas9脱靶位点是已知的,可以测量这些位点的修饰并在测试蛋白和对照品之间进行比较。从分离自用特定Cas9蛋白或变体处理的细胞(例如,HEK293)的基因组DNA扩增靶位点及其相应的已知脱靶位点。然后通过高通量测序分析这些修饰。含有潜在的基因组脱靶位点中的两个或更多个碱基对的插入或缺失的序列,并且靶gRNA处理样品中与对照gRNA处理样品中相比数目显著更多的序列(p值<0.005,Fishers精确检验)被认为是Cas9核酸酶诱导的基因组修饰。

[0351] 在一些实施例中,本披露的方法提供了Cas9-核酸内切酶的二聚体,该二聚体包含第一Cas9-核酸内切酶单体和第二Cas9-核酸内切酶单体。在该方法的多个实施例中,Cas9-核酸内切酶的核酸内切酶是IIS型核酸内切酶。在一些实施例中,第一Cas9-核酸内切酶二聚体中的第一单体的核酸内切酶是IIS型核酸内切酶。在一些实施例中,第一Cas9-核酸内切酶二聚体中的第二单体的核酸内切酶是IIS型核酸内切酶。在一些实施例中,第一Cas9-核酸内切酶二聚体中的第一单体和第二单体的核酸内切酶是IIS型核酸内切酶。在一些实施例中,第二Cas9-核酸内切酶二聚体中的第一单体的核酸内切酶是IIS型核酸内切酶。在一些实施例中,第二Cas9-核酸内切酶二聚体中的第二单体的核酸内切酶是IIS型核酸内切酶。在一些实施例中,第二Cas9-核酸内切酶二聚体中的第一单体和第二单体的核酸内切酶是IIS型核酸内切酶。在一些实施例中,第一Cas9-核酸内切酶二聚体和第二Cas9-核酸内切酶二聚体中的核酸内切酶是IIS型核酸内切酶。

[0352] 内切核酸酶或限制性内切酶传统上根据亚基组成、裂解位置、序列特异性和辅因子需求而分为四种类型。然而,氨基酸测序已发现限制性内切酶具有极其丰富的多样性,并揭示了在分子水平上有四种以上的不同类型。

[0353] “IIS型”核酸内切酶是那些像FokI和AlwI那样在其识别序列之外裂解至一侧的酶。IIS型限制酶具有中等大小,长度为400-650个氨基酸,并且它们可识别连续且不对称的序列。它们包含两个不同的结构域,一个结构域用于DNA结合,另一个结构域用于DNA裂解。它们被认为主要作为单体与DNA结合,但通过相邻酶分子的裂解结构域的二聚化来协同裂解DNA。因此,某些IIS型酶对包含多个识别位点的DNA分子的活性更高。IIS型核酸内切酶的非限制性实例包括:AcuI、AlwI、BaeI、BbsI、BbvI、BccI、BceAI、BcgI、BciVI、BcoDI、BfuAI、BmrI、BpmI、BpuEI、BsaI、BsaXI、BseRI、BsgI、BsmAI、BsmBI、BsmFI、BsmI、BspCNI、BspMI、BspQI、BsrDI、BsrI、BtgZI、BtsCI、BtsI、CspCI、EarI、EciI、FauI、FokI、HgaI、HphI、HpyAV、MboII、MlyI、MmeI、MnlI、NmeAIII、PleI、SapI和SfaNI。在一些实施例中,第一Cas9-核酸内切酶二聚体和第二Cas9-核酸内切酶二聚体中的核酸内切酶独立地选自下组,该组由以下组成:BbvI、BgcI、BfuAI、BmpI、BspMI、CspCI、FokI、MboII、MmeI、NmeAIII和PleI。在一些实施例中,第一Cas9-核酸内切酶二聚体和第二Cas9-核酸内切酶二聚体中的核酸内切酶是FokI。由FokI进行的DNA裂解仅在两个FokI单体二聚化时发生。DNA的FokI裂解产生具有4个碱基对突出端的粘性末端。

[0354] Cas9-核酸内切酶融合蛋白中的核酸内切酶也可以是工程化的FokI核酸酶,例如,

工程化的FokI二聚体。在一些实施例中,该工程化的FokI二聚体是强制性异二聚体,即,需要两个不同的单体来形成功能性(催化活性)二聚体。

[0355] 在一些实施例中,第一和第二Cas9-核酸内切酶二聚体是相同的。在一些实施例中,第一和第二Cas9-核酸内切酶二聚体是不同的。

[0356] 在一些实施例中,本方法提供了该第一、第二Cas9-核酸内切酶二聚体或两者,这些二聚体包含经修饰的Cas9。在一些实施例中,该经修饰的Cas9是无催化活性的Cas9(“deadCas9”)。在一些实施例中,第一、第二Cas9-核酸内切酶二聚体或两者包含无催化活性的Cas9。无催化活性的Cas9不能裂解DNA(即,Cas9的裂解结构域失活);然而,它们通过与引导多核苷酸(例如,引导RNA)形成复合物而保留了靶向核酸序列的能力。无催化活性的Cas9已在本领域进行了描述,例如,Jinek等人(2012)和Qi等人,“Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression[将CRISPR重新用作基因表达的序列特异性控制的RNA引导平台]”,*Cell*[细胞]152(5):1173-1183(2013)。在一些实施例中,无催化活性的Cas9相对于野生型Cas9包含双氨基酸取代。在一些实施例中,该Cas9-核酸内切酶二聚体包含相对于野生型Cas9的双氨基酸取代。在一些实施例中,该双氨基酸取代是D10A和H840A。在一些实施例中,第一、第二Cas9-核酸内切酶二聚体或两者中的核酸内切酶是FokI,并且第一、第二Cas9-核酸内切酶二聚体或两者中的Cas9是无催化活性的Cas9(“deadCas9-FokI”)。在一些实施例中,第一、第二Cas9-核酸内切酶二聚体或两者中的核酸内切酶是FokI,并且第一、第二Cas9-核酸内切酶二聚体或两者中的Cas9包含D10A/H840A双氨基酸取代。

[0357] 在一些实施例中,该经修饰的Cas9是具有切口酶活性的Cas9(“Cas9切口酶”或“Cas9n”)。在一些实施例中,第一、第二Cas9-核酸内切酶二聚体或两者包含具有切口酶活性的Cas9。Cas9切口酶仅能够裂解双链DNA的一条链(即,“切口”该DNA)。例如,Cho等人,“Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases[CRISPR/Cas衍生RNA引导的核酸内切酶和切口酶的脱靶效应分析]”,*Genome Research*[基因组研究]24:132-141(2013),Ran等人(*Cell*[细胞],2013)和Mali等人(*Nature Biotechnology*[自然生物技术],2013)描述了Cas9切口酶。在一些实施例中,Cas9切口酶包含相对于野生型Cas9的单个氨基酸取代。在一些实施例中,该Cas9-核酸内切酶二聚体包含相对于野生型Cas9的单个氨基酸取代。在一些实施例中,该单个氨基酸取代是D10A(“Cas9n^(D10A)”)。在一些实施例中,该单个氨基酸取代是H840A(“Cas9n^(H840A)”)。在一些实施例中,第一、第二Cas9-核酸内切酶二聚体或两者中的核酸内切酶是FokI,并且第一、第二Cas9-核酸内切酶二聚体或两者中的Cas9是Cas9切口酶。在一些实施例中,第一、第二Cas9-核酸内切酶二聚体或两者中的核酸内切酶是FokI,并且第一、第二Cas9-核酸内切酶二聚体或两者中的Cas9包含D10A单氨基酸取代(“Cas9n^(D10A)-FokI”)。在一些实施例中,第一、第二Cas9-核酸内切酶二聚体或两者中的核酸内切酶是FokI,并且第一、第二Cas9-核酸内切酶二聚体或两者中的Cas9包含H8410A单氨基酸取代(“Cas9n^(H840A)-FokI”)。

[0358] 在一些实施例中,该野生型Cas9源自酿脓链球菌、金黄色葡萄球菌、伪中间型葡萄球菌、南极游动球菌、血链球菌、嗜热链球菌、变形链球菌、球团科里氏杆菌、香肠乳杆菌、光冈氏链小杆菌、鼠李糖乳杆菌、两歧双歧杆菌、北原酒球菌、果聚糖杆菌、大芬戈尔德菌、非

典型韦荣氏球菌、穆氏索罗菌、氨基酸球菌属物种D21、尤里真杆菌、灵巧粪球菌、具核梭杆菌、颞沟产线菌、杜尔丹尼嗜脲菌或齿垢密螺旋体。

[0359] 在一些实施例中,由该Cas9-核酸内切酶产生的粘性末端包含5'突出端。在一些实施例中,由该Cas9-核酸内切酶产生的粘性末端包含3'突出端。在一些实施例中,第一、第二Cas9-核酸内切酶二聚体或两者产生包含具有3至40个核苷酸的单链多核苷酸的粘性末端。在一些实施例中,第一、第二Cas9-核酸内切酶二聚体或两者产生包含具有4至30个核苷酸的单链多核苷酸的粘性末端。在一些实施例中,第一、第二Cas9-核酸内切酶二聚体或两者产生包含具有5至20个核苷酸的单链多核苷酸的粘性末端。在一些实施例中,第一、第二Cas9-核酸内切酶二聚体或两者产生包含具有约5个核苷酸、约10个核苷酸、约15个核苷酸、约20个核苷酸、约25个核苷酸或约30个核苷酸的单链多核苷酸的粘性末端。在一些实施例中,deadCas9-FokI二聚体产生包含4-核苷酸5'突出端的粘性末端。在一些实施例中,Cas9n^(D10A)-FokI二聚体产生包含27-核苷酸5'突出端的粘性末端。在一些实施例中,Cas9^(H840A)-FokI二聚体产生包含23-核苷酸3'-突出端的粘性末端。

[0360] 在该方法的多个实施例中,该目的序列(SoI)由供体质粒组成。该供体质粒可以具有任何合适的长度,诸如约或至少约10、15、20、25、50、75、100、150、200、250、500或1000或更多个核苷酸的长度。在一些实施例中,该供体质粒与包含TSC的染色体的一部分互补。当进行最佳比对时,该供体质粒模板与TSC的一个或多个核苷酸(例如,约或至少约1、5、10、15、20、25、30、35、40、45、50、60、70、80、90或100或更多个核苷酸)重叠。在一些实施例中,当对该供体质粒模板和包含TSC的染色体进行最佳比对时,该供体质粒最接近的核苷酸在距离TSC约1、5、10、15、20、25、50、75、100、200、300、400、500、100、1500、2000、2500、5000、10000个或更多个核苷酸内。

[0361] 在一些实施例中,该SoI是DNA,例如,DNA质粒、细菌人工染色体(BAC)、酵母人工染色体(YAC)、病毒载体、线性DNA片段、PCR片段、裸露的核酸,或与诸如脂质体的递送媒介物复合的核酸。

[0362] 在一些实施例中,使用细胞的内源DNA修复途径将SoI插入TSC中。在一些实施例中,使用非同源末端连接(NHEJ)修复途径的组分将SoI插入TSC中。在修复过程中,可以将包含SoI的供体质粒引入TSC中。

[0363] 在一些实施例中,将包含侧接上游序列和下游序列的SoI的供体质粒引入细胞中,其中该上游和下游序列与TSC中整合位点的任一例具有序列相似性。在一些实施例中,该包含SoI的外源多核苷酸包含例如突变基因。在一些实施例中,该外源多核苷酸包含对于该细胞而言是内源性或外源性的序列。在一些实施例中,该SoI包含编码蛋白质的多核苷酸,或非编码序列,诸如像微小RNA。在一些实施例中,该SoI可操作地与调节元件连接。在一些实施例中,该SoI是调节元件。在一些实施例中,该SoI包含抗性盒,例如,赋予对抗生素抗性的基因。在一些实施例中,该SoI包含野生型靶序列的突变。在一些实施例中,该SoI通过产生移码突变或核苷酸取代来破坏靶序列。在一些实施例中,该SoI包含标记。将标记引入靶序列可以便于筛选靶向的整合。在一些实施例中,该标记是限制性位点、荧光蛋白或选择性标记。在一些实施例中,将该SoI作为包含SoI的载体引入。

[0364] 选择该外源多核苷酸模板中的上游和下游序列以促进靶序列和外源多核苷酸之间的同源重组。该上游序列是与用于整合的靶向位点的上游序列(靶序列)具有序列相似性

的核酸序列。类似地,该下游序列是与用于整合的靶位点的下游序列具有序列相似性的核酸序列。因此,在一些实施例中,通过在上游和下游序列处的同源重组,将包含SoI的外源多核苷酸模板插入靶序列中。在一些实施例中,该外源多核苷酸模板中的上游和下游序列与靶向的基因组序列的上游和下游序列分别具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、分别至少97%、至少98%、至少99%或100%的序列同一性。在一些实施例中,该上游或下游序列具有约20至2000个碱基对、或约50至1750个碱基对、或约100至1500个碱基对、或约200至1250个碱基对、或约300至1000个碱基对、或约400至约750个碱基对、或约500至600个碱基对。在一些实施例中,该上游或下游序列具有约50、约100、约250、约500、约1000、约1250、约1500、约1750、约2000、约2250或约2500个碱基对。

[0365] 在一些实施例中,在插入SoI时,染色体中的靶序列和质粒中的靶序列没有被重建。即,在一些实施例中,染色体中的所得序列(即,来自SoI插入的所得序列)不与第一、第二、第三或第四引导多核苷酸中的任何一个杂交。因此,在一些实施例中,包含SoI的染色体中的所得序列不易被第一或第二Cas9-核酸内切酶二聚体或第一或第二Cas9-核酸内切酶的二聚体中的任何单体裂解。如图13和图15所例示的,所得到的“敲入”序列(“预期的5'连接”)是与“基因组”和“载体”序列不同的序列,并且该“敲入”序列不具有可与gRNA1、gRNA2、gRNA3或gRNA4中任何一个杂交的序列。

[0366] 在一些实施例中,本披露的方法进一步包括将第一引导多核苷酸引入细胞中,该第一引导多核苷酸与第一Cas9-核酸内切酶二聚体的第一单体形成复合物并且包含第一引导序列,其中该第一引导序列与包含区域1的TSC杂交,但不与该载体杂交。如图13和15所例示的,第一引导序列(显示为“gRNA1”)与基因组靶DNA的非编码链上的区域1的一部分以及区域1之外的几个核苷酸结合。gRNA1不与基因组或载体中的任何其他序列杂交。在一些实施例中,第一引导多核苷酸通过与Cas9的结合结构域相互作用而与第一Cas9-核酸内切酶二聚体的第一单体形成复合物。

[0367] 在一些实施例中,本披露的方法进一步包括将第二引导多核苷酸引入细胞中,该第二引导多核苷酸与第一Cas9-核酸内切酶二聚体的第二单体形成复合物并包含第二引导序列,其中该第二引导序列与包含区域2的TSC杂交,但不与载体杂交。如图13和15所例示的,第二引导序列(显示为“gRNA2”)与基因组靶DNA的编码链上的区域2的一部分结合。gRNA2不与基因组或载体中的任何其他序列杂交。在一些实施例中,第二引导多核苷酸通过与Cas9的结合结构域相互作用而与第一Cas9-核酸内切酶二聚体的第二单体形成复合物。

[0368] 在一些实施例中,本披露的方法进一步包括将第三引导多核苷酸引入细胞中,该第三引导多核苷酸与第二Cas9-核酸内切酶二聚体的第一单体形成复合物并包含第三引导序列,其中该第三引导序列与包含区域2的TSV杂交,但不与该基因组杂交。如图13和15所例示的,第三引导序列(显示为“gRNA3”)与载体中靶DNA的非编码链上的区域2的一部分以及区域2之外的几个核苷酸结合。gRNA3不与基因组或载体中的任何其他序列杂交。在一些实施例中,第三引导多核苷酸通过与Cas9的结合结构域相互作用而与第二Cas9-核酸内切酶二聚体的第一单体形成复合物。

[0369] 在一些实施例中,本披露的方法进一步包括将第四引导多核苷酸引入细胞中,该第四引导多核苷酸与第二Cas9-核酸内切酶二聚体的第二单体形成复合物并包含第四引导序列,其中该第四引导序列与包含区域1的TSC杂交,但不与该基因组杂交。如图13和15所例

示的,第四引导序列(显示为“gRNA4”)与载体中靶DNA的编码链上的区域1的一部分结合。gRNA4不与基因组或载体中的任何其他序列杂交。在一些实施例中,第四引导多核苷酸通过与Cas9的结合结构域相互作用而与第二Cas9-核酸内切酶二聚体的第二单体形成复合物。

[0370] 在一些实施例中,引导多核苷酸能够与TSC和TSV两者结合。因此,在一些实施例中,该方法进一步包括将第一引导多核苷酸引入细胞中,该第一引导多核苷酸与第一Cas9-核酸内切酶二聚体的第一单体形成复合物并包含第一引导序列,其中该第一引导序列与TSC和TSV杂交。

[0371] 在一些实施例中,该方法进一步包括将第二引导多核苷酸引入细胞中,该第二引导多核苷酸与第一Cas9-核酸内切酶二聚体的第二单体形成复合物并包含第二引导序列,其中该第二引导序列与TSC和TSV杂交。

[0372] 在一些实施例中,该方法进一步包括将第三引导多核苷酸引入细胞中,该第三引导多核苷酸与第二Cas9-核酸内切酶二聚体的第一单体形成复合物并包含第三引导序列,其中该第三引导序列与TSC和TSV杂交。

[0373] 在一些实施例中,该方法进一步包括将第四引导多核苷酸引入细胞中,该第四引导多核苷酸与第二Cas9-核酸内切酶二聚体的第二单体形成复合物并包含第四引导序列,其中该第四引导序列与TSC和TSV杂交。

[0374] 在一些实施例中,第一、第二、第三和/或第四引导多核苷酸是相同的。在一些实施例中,第一、第二、第三和/或第四引导多核苷酸是不同的。

[0375] 在一些实施例中,本披露的方法包括将第一、第二、第三和第四引导多核苷酸引入细胞中。在一些实施例中,第一Cas9-核酸内切酶二聚体的第一单体与第一引导多核苷酸形成复合物,并且第一Cas9-核酸内切酶二聚体的第二单体与第二引导多核苷酸形成复合物。在一些实施例中,第二Cas9-核酸内切酶二聚体的第一单体与第三引导多核苷酸形成复合物,并且第二Cas9-核酸内切酶二聚体的第二单体与第四引导多核苷酸形成复合物。

[0376] 在一些实施例中,第一Cas9-核酸内切酶二聚体的第一单体与第一引导多核苷酸形成复合物,第一Cas9-核酸内切酶二聚体的第二单体与第二引导多核苷酸形成复合物,第二Cas9-核酸内切酶二聚体的第一单体与第三引导多核苷酸形成复合物,并且第二Cas9-核酸内切酶二聚体的第二单体与第四引导多核苷酸形成复合物。在一些实施例中,第一和第二引导多核苷酸将第一Cas9-核酸内切酶二聚体导向细胞染色体上的靶序列,第三和第四引导多核苷酸将第二Cas9-核酸内切酶二聚体导向引入细胞中的载体上的靶序列。

[0377] 在一些实施例中,本披露的方法进一步包括将tracrRNA引入细胞中。在一些实施例中,该引导多核苷酸包含crRNA/tracrRNA杂交体。在一些实施例中,该引导多核苷酸的tracrRNA组分激活Cas9-核酸内切酶的Cas9。在一些实施例中,Cas9-核酸内切酶、引导多核苷酸和tracrRNA能够形成复合物。在一些实施例中,该复合物包含Cas9-核酸内切酶、两个引导多核苷酸和两个tracrRNA序列。在一些实施例中,Cas9-核酸内切酶、引导多核苷酸和tracrRNA的复合物在自然界中不存在。

[0378] 在一些实施例中,第一Cas9-核酸内切酶二聚体的第一单体与第一引导多核苷酸序列和tracrRNA序列形成复合物,并且第一Cas9-核酸内切酶二聚体的第二单体与第二引导多核苷酸序列和tracrRNA序列形成复合物。在一些实施例中,第二Cas9-核酸内切酶二聚体的第一单体与第三引导多核苷酸序列和tracrRNA序列形成复合物,并且第二Cas9-核酸

内切酶二聚体的第二单体与第四引导多核苷酸序列和tracrRNA序列形成复合物。

[0379] 在一些实施例中,第一Cas9-核酸内切酶二聚体的第一单体与第一引导多核苷酸和tracrRNA形成复合物,第一Cas9-核酸内切酶二聚体的第二单体与第二引导多核苷酸和tracrRNA形成复合物,第二Cas9-核酸内切酶二聚体的第一单体与第三引导多核苷酸和tracrRNA形成复合物,并且第二Cas9-核酸内切酶二聚体的第二单体与第四引导多核苷酸和tracrRNA形成复合物。在一些实施例中,第一引导多核苷酸和tracrRNA以及第二引导多核苷酸和tracrRNA将第一Cas9-核酸内切酶二聚体导向细胞染色体上的靶序列,第三引导多核苷酸和tracrRNA以及第四引导多核苷酸和tracrRNA将第二Cas9-核酸内切酶二聚体导向引入细胞中的载体上的靶序列。

[0380] 在该方法的多个实施例中,将TSV、第一和/或第二Cas9-核酸内切酶二聚体作为编码第一和第二Cas9-核酸内切酶二聚体的一种或多种多核苷酸引入细胞。在一些实施例中,对编码TSV、第一和/或第二Cas9-核酸内切酶二聚体的多核苷酸进行密码子优化以在真核细胞中表达。在一些实施例中,对编码TSV、第一和/或第二Cas9-核酸内切酶二聚体的多核苷酸进行密码子优化以在哺乳动物细胞中表达。本文描述了密码子优化方法和技术。

[0381] 在一些实施例中,将TSV、第一和/或第二Cas9-核酸内切酶二聚体作为单个核酸分子引入细胞中。在一些实施例中,编码TSV、第一和/或第二Cas9-核酸内切酶二聚体的多核苷酸在单个载体上。在一些实施例中,编码第一和第二Cas9-核酸内切酶二聚体、一个或多个引导多核苷酸以及一个或多个tracrRNA序列的多核苷酸在单个载体上。在一些实施例中,该载体是表达载体。在一些实施例中,该载体是真核表达载体。在一些实施例中,该载体是哺乳动物表达载体。在一些实施例中,该载体是人类表达载体。在一些实施例中,该载体是植物表达载体。

[0382] 在一些实施例中,编码TSV、第一和/或第二Cas9-核酸内切酶二聚体的多核苷酸在多于一个载体上。在一些实施例中,编码TSV、第一和/或第二Cas9-核酸内切酶二聚体、一个或多个引导多核苷酸以及一个或多个tracrRNA序列的多核苷酸在多于一个载体上。在一些实施例中,这些载体是表达载体。在一些实施例中,这些载体是真核表达载体。在一些实施例中,这些载体是哺乳动物表达载体。在一些实施例中,这些载体是人类表达载体。在一些实施例中,这些载体是植物表达载体。

[0383] 在该方法的多个实施例中,该细胞是真核细胞。在一些实施例中,该真核细胞是动物或人类细胞。在一些实施例中,该真核细胞是人类或啮齿动物或牛细胞系或者细胞株。此类细胞、细胞系或细胞株的实例包括但不限于小鼠骨髓瘤(NS0)细胞系、中国仓鼠卵巢(CHO)细胞系、HT1080、H9、HepG2、MCF7、MDBK Jurkat、NIH3T3、PCI2、BHK(幼仓鼠肾细胞)、VERO、SP2/0、YB2/0、Y0、C127、L细胞、COS(例如COS1和COS7)、QC1-3、HEK-293、VERO、PER.C6、HeLA、EB1、EB2、EB3、溶瘤或杂交瘤细胞系。在一些实施例中,该真核细胞是CHO细胞系。在一些实施例中,该真核细胞是CHO细胞。在一些实施例中,该细胞是CHO-K1细胞、CHO-K1 SV细胞、DG44 CHO细胞、DUXB11 CHO细胞、CHOS、CHO GS敲除细胞、CHO FUT8 G_s敲除细胞、CHOZN或CHO衍生的细胞。CHO GS敲除细胞(例如,GSKO细胞)是例如CHO-K1 SV GS敲除细胞。CHO FUT8敲除细胞是例如Potelligent®CHOK1 SV(龙沙生物公司)。真核细胞也可以是鸟类细胞、细胞系或细胞株,诸如像EBx®细胞、EB14、EB24、EB26、EB66或EBv13。

[0384] 在一些实施例中,该真核细胞是人类细胞。在一些实施例中,该人类细胞是干细

胞。干细胞可以是例如多能干细胞,包括胚胎干细胞(ESC)、成年干细胞、诱导多能干细胞(iPSC)、组织特异性干细胞(例如,造血干细胞)和间充质干细胞(MSC)。在一些实施例中,该人类细胞是本文描述的任何细胞的分化形式。在一些实施例中,该真核细胞是源自培养物中任何原代细胞的细胞。在一些实施例中,该细胞是干细胞或干细胞系。

[0385] 在一些实施例中,该真核细胞是肝细胞,诸如人肝细胞、动物肝细胞,或非实质细胞。例如,该真核细胞可以是可培养代谢合格的人肝细胞、可培养感应诱导合格的人肝细胞、可培养Qualyst Transporter Certified™人肝细胞、悬浮合格的人肝细胞(包括10-供体和20-供体合并的肝细胞)、人肝库普弗细胞、人肝星状细胞、狗肝细胞(包括单个和合并的比格犬肝细胞)、小鼠肝细胞(包括CD-1和C57BI/6肝细胞)、大鼠肝细胞(包括Sprague-Dawley、Wistar Han和Wistar肝细胞)、猴肝细胞(包括食蟹猴或恒河猴肝细胞)、猫肝细胞(包括家养短毛猫肝细胞)和兔肝细胞(包括新西兰白兔肝细胞)。

[0386] 在一些实施例中,该真核细胞是植物细胞。例如,该植物细胞可以是诸如木薯、玉米、高粱、小麦或水稻的农作物的细胞。该植物细胞可以是藻类、树木或蔬菜的细胞。该植物细胞可以是单子叶植物或双子叶植物的细胞,或者可以是农作物或谷物植物、生产植物、水果或蔬菜的细胞。例如,该植物细胞可以是树木的细胞,该树木为例如柑橘属果树,诸如橘子树、葡萄柚树或柠檬树;桃树或油桃树;苹果树或梨树;坚果树,诸如杏仁树或胡桃树或开心果树;茄属植物,即,马铃薯;芸苔属植物,莴苣属植物;菠菜属植物;辣椒属植物;棉花、烟草、芦笋、胡萝卜、卷心菜、西兰花、花椰菜、番茄、茄子、胡椒、莴苣、菠菜、草莓、蓝莓、覆盆子、黑莓、葡萄、咖啡、可可等。

[0387] 在该方法的多个实施例中,通过递送颗粒、囊泡或病毒载体将能够在TSC中产生粘性末端的第一Cas9-核酸内切酶二聚体和能够在TSV中产生粘性末端的第二Cas9-核酸内切酶二聚体引入细胞。

[0388] 在一些实施例中,该TSV、第一和/或第二Cas9-核酸内切酶二聚体通过递送颗粒被递送到细胞中。本文提供了递送颗粒的实例。在一些实施例中,该递送颗粒是基于脂质的系统、脂质体、胶束、微囊泡、外来体或基因枪。在一些实施例中,该递送颗粒包含该Cas9-核酸内切酶二聚体的全部两个单体。在一些实施例中,该递送颗粒包含全部两种Cas9-核酸内切酶二聚体的全部两个单体。在一些实施例中,该递送颗粒包含Cas9-核酸内切酶和引导多核苷酸。在一些实施例中,该递送颗粒包含Cas9-核酸内切酶和引导多核苷酸,其中Cas9-核酸内切酶和引导多核苷酸以复合物存在。在一些实施例中,该递送颗粒包含编码Cas9-核酸内切酶的多核苷酸、编码引导多核苷酸的多核苷酸和包含tracrRNA的多核苷酸。在一些实施例中,该递送颗粒包含Cas9-核酸内切酶、引导多核苷酸和tracrRNA。在一些实施例中,该递送颗粒包含第一和/或第二Cas9-核酸内切酶二聚体,第一、第二、第三和/或第四引导多核苷酸以及tracrRNA。在一些实施例中,该递送颗粒包含编码一个或多个Cas9-内切核酸酶的多核苷酸,编码第一、第二、第三和/或第四引导多核苷酸的多核苷酸,以及编码tracrRNA的多核苷酸。

[0389] 在一些实施例中,该递送颗粒还包含脂质、糖、金属或蛋白质。在一些实施例中,该递送颗粒是脂质包膜。在一些实施例中,该递送颗粒是基于糖的颗粒,例如,GalNAc。在一些实施例中,该递送颗粒是纳米颗粒。本文描述了纳米颗粒的实例。在美国专利公开号2011/0293703、2012/0251560和2013/0302401以及美国专利号5,543,158、5,855,913、5,895,

309,6,007,845和8,709,843中进一步描述了递送颗粒的制备,将其各自通过引用以其整体并入本文。

[0390] 在一些实施例中,该TSV,第一和/或第二Cas9-核酸内切酶二聚体通过囊泡被递送到细胞中。“囊泡”是细胞内具有被脂质双层包围的流体的小结构。本文提供了囊泡的实例。在一些实施例中,该囊泡包含该Cas9-核酸内切酶二聚体的全部两个单体。在一些实施例中,该囊泡包含全部两种Cas9-核酸内切酶二聚体的全部两个单体。在一些实施例中,该囊泡包含Cas9-核酸内切酶和引导多核苷酸。在一些实施例中,该囊泡包含Cas9-核酸内切酶和引导多核苷酸,其中该Cas9-核酸内切酶和引导多核苷酸以复合物存在。在一些实施例中,该囊泡包含编码Cas9-核酸内切酶的多核苷酸、编码引导多核苷酸的多核苷酸和包含tracrRNA的多核苷酸。在一些实施例中,该囊泡包含Cas9-核酸内切酶、引导多核苷酸和tracrRNA。在一些实施例中,该囊泡包含第一和/或第二Cas9-核酸内切酶二聚体、第一、第二、第三和/或第四引导多核苷酸以及tracrRNA。在一些实施例中,该囊泡包含编码一个或多个Cas9-核酸内切酶的多核苷酸,编码第一、第二、第三和/或第四引导多核苷酸的多核苷酸,以及编码tracrRNA的多核苷酸。

[0391] 在一些实施例中,该囊泡是外泌体或脂质体。在一些实施例中,第一和/或第二Cas9-核酸内切酶二聚体通过外泌体递送到细胞中。外泌体是内源性纳米囊泡(即,直径为约30nm至约100nm),可转运RNA和蛋白质,并且可将RNA递送至大脑和其他靶器官。例如,Alvarez-Erviti等人,Nature Biotechnology[自然生物学]29:341(2011),El-Andaloussi等人,Nature Protocols[自然实验手册]7:2112-2116(2012),以及Wahlgren等人,Nucleic Acids Research[核酸研究]40(17):e130(2012)中描述了用于将内源性生物材料递送至靶器官的工程化外泌体,将其各自通过引用以其整体并入本文。

[0392] 在一些实施例中,该TSV,第一和/或第二Cas9-核酸内切酶二聚体通过脂质体被递送到细胞中。脂质体是具有至少一个脂质双层的球形囊泡结构,并且可以用作营养物和药物施用的媒介物。脂质体通常由磷脂(特别是磷脂酰胆碱)以及其他脂质(诸如蛋磷脂酰乙醇胺)组成。脂质体的类型包括但不限于多层囊泡、小单层囊泡、大单层囊泡和耳蜗囊泡。参见,例如,Spuch和Navarro,“Liposomes for Targeted Delivery of Active Agents against Neurodegenerative Diseases (Alzheimer's Disease and Parkinson's Disease) [用于抗神经退行性疾病(阿尔茨海默病和帕金森病)活性药物的靶向递送的脂质体]”,Journal of Drug Delivery[药物递送杂志]2011,文章ID 469679(2011)。例如,Morrissey等人,Nature Biotechnology[自然生物技术]23(8):1002-1007(2005),Zimmerman等人,Nature Letters[自然快报]441:111-114(2006),以及Li等人,Gene Therapy[基因疗法]19:775-780(2012)描述了用于递送诸如CRISPR-Cas组分的生物材料的脂质体,将其各自通过引用以其整体并入本文。

[0393] 在该方法的多个实施例中,该TSV、第一和/或第二Cas9-核酸内切酶二聚体通过病毒载体被递送到细胞中。在一些实施例中,该病毒载体包含该Cas9-核酸内切酶二聚体的全部两个单体。在一些实施例中,该病毒载体包含全部两种Cas9-核酸内切酶二聚体的全部两个单体。在一些实施例中,该病毒载体包含TSV。在一些实施例中,该病毒载体包含Cas9-核酸内切酶和引导多核苷酸。在一些实施例中,该病毒载体包含Cas9-核酸内切酶和引导多核苷酸,其中该Cas9-核酸内切酶和引导多核苷酸以复合物存在。在一些实施例中,该病毒载

体包含编码Cas9-核酸内切酶的多核苷酸、编码引导多核苷酸的多核苷酸和包含tracrRNA的多核苷酸。在一些实施例中,该病毒载体包含第一和/或第二Cas9-核酸内切酶二聚体,第一、第二、第三和/或第四引导多核苷酸以及tracrRNA。在一些实施例中,该病毒载体包含编码一个或多个Cas9-内切核酸酶的多核苷酸,编码第一、第二、第三和/或第四引导多核苷酸的多核苷酸,以及编码tracrRNA的多核苷酸。在一些实施例中,该病毒载体包含TSV,编码一个或多个Cas9-内切核酸酶的多核苷酸,编码第一、第二、第三和/或第四引导多核苷酸的多核苷酸以及编码tracrRNA的多核苷酸。

[0394] 在一些实施例中,该病毒载体是腺病毒、慢病毒或腺相关病毒载体。本文提供了病毒载体的实例。使用腺相关病毒(AAV)和慢病毒载体的病毒转导(可以局部、靶向或全身施用)已被用作体内基因疗法的递送方法。在本披露的多个实施例中,该Cas蛋白由转导的细胞在细胞内表达。

[0395] 在一些实施例中,第一、第二Cas9-核酸内切酶二聚体或两者包含核定位信号。在一些实施例中,第一Cas9-核酸内切酶二聚体的第一、第二单体或两者包含核定位信号。在一些实施例中,第二Cas9-核酸内切酶二聚体的第一、第二单体或两者包含核定位信号。在一些实施例中,第一、第二Cas9-核酸内切酶二聚体或两者的第一、第二单体或两者均包含核定位信号。本文描述了核定位信号(“NLS”)。示例性核定位序列包括但不限于来自以下的NLS:SV40大T抗原、核质蛋白、EGL-13、c-Myc和TUS蛋白。在一些实施例中,该NLS包含PKKKRKV(SEQ ID NO:1)序列。在一些实施例中,该NLS包含AVKRPAATKKAGQAKKKLD(SEQ ID NO:2)序列。在一些实施例中,该NLS包含PAAKRVKLD(SEQ ID NO:3)序列。在一些实施例中,该NLS包含MSRRRK ANPTKL SENAKKL AKEVEN(SEQ ID NO:4)序列。在一些实施例中,该NLS包含KLIKIKRPVK(SEQ ID NO:5)序列。其他核定位序列包括但不限于hnRNP A1的酸性M9结构域、酵母转录抑制子Mata2中的序列KIPIK(SEQ ID NO:6)和PY-NLS。

[0396] 无缝诱变的方法

[0397] 在一些实施例中,本披露提供了一种无缝修饰细胞中靶多核苷酸序列中一个或多个核苷酸的方法。“无缝诱变”是指定点诱变(即,一个或多个核苷酸的取代、缺失或插入)而没有任何其他附近变化,诸如存在用于引入突变的可选择基因。在蛋白质编码区域进行诱变的无缝DNA工程是有利的,因为在诱变步骤中引入的任何外来序列都可能干扰蛋白质表达。本披露提供了使用两步选择/反向选择策略的无缝诱变,这种策略首先涉及在靶位点插入诸如伴有反向可选择基因的抗生素抗性基因之类的可选择盒。然后,随后通过对通常涉及小分子(诸如链霉素或糖)施用的反向可选择基因进行选择,用希望的序列将该盒无缝替换。反向选择性标记的流行选项包括sacB、rpsL,以及可以在正确的宿主背景下同时针对其进行选择的标记,包括galK、thyA和tolC。先前的无缝诱变方法描述于例如Wang等人,“Improved seamless mutagenesis by recombineering using ccdB for counterselection[通过使用ccdB进行反向选择的重组进行改进的无缝诱变]”,Nucleic Acids Research[核酸研究]42(5):e37(2014);Zhang等人,“A new logic for DNA engineering using recombination in Escherichia coli[在大肠杆菌中使用重组进行DNA工程的新逻辑]”,Nature Genetics[自然遗传学]20(2):123-128(1998);Westenberg等人,“Counter-selection recombineering of the baculovirus genome:a strategy for seamless modification of repeat-containing BACs[杆状病毒基因组的反向选择重组:

含重复序列的BAC的无缝修饰策略”], *Nucleic Acids Research* [核酸研究] 38:e166 (2010); Wong等人, “Efficient and seamless DNA recombineering using a thymidylate synthase A selection system in *Escherichia coli* [在大肠杆菌中使用胸苷酸合酶A选择系统进行高效无缝的DNA重组工程]”, *Nucleic Acids Research* [核酸研究] 33:e59 (2005), 将其各自通过引用以其整体并入本文。

[0398] 在一些实施例中, 本披露提供了一种修饰细胞中靶多核苷酸序列中一个或多个核苷酸的方法, 该方法包括: (1) 将包含插入盒(IC)的载体引入细胞中, 该IC沿5'至3'方向包含: (a) 与靶多核苷酸序列的一部分同源的第一区域, (b) 包含靶多核苷酸序列中的一个或多个核苷酸的突变的第二区域, (c) 第一核酸酶结合位点, (d) 编码标记基因的多核苷酸序列, (e) 第二核酸酶结合位点, (f) 包含靶多核苷酸序列中的一个或多个突变中的一个突变的第三区域, 以及 (g) 与靶多核苷酸序列的一部分同源的第四区域, 其中第一区域和第四区域与它们在该靶多核苷酸序列中相应部分95%-100%相同; (2) 通过同源重组将IC插入靶多核苷酸序列以产生第一经修饰的靶多核苷酸; (3) 选择表达该标记基因的细胞; (4) 使该第一经修饰的靶多核苷酸经受位点特异性核酸酶处理以产生具有粘性末端的第二经修饰的靶多核苷酸; 以及 (5) 使该具有粘性末端的第二经修饰的靶多核苷酸经受连接酶处理, 其中该连接酶在该第二区域和该第三区域处连接这些粘性末端以产生连接的经修饰的靶核酸, 当与该靶多核苷酸序列相比时, 该连接的经修饰的靶核酸包含一个或多个经修饰的核苷酸。

[0399] 在一些实施例中, 靶多核苷酸序列中一个或多个核苷酸的修饰是核苷酸取代, 即, 单个核苷酸取代或多个核苷酸取代。靶多核苷酸序列中一个或多个核苷酸的修饰可导致由多核苷酸编码的多肽序列的改变。靶多核苷酸序列中一个或多个核苷酸的修饰还可导致细胞中下游多核苷酸序列的表达的失活。例如, 下游序列失活, 使得该序列不被转录, 不产生编码的蛋白质, 或者该序列不像野生型序列那样起作用。在一些实施例中, 该靶多核苷酸序列是调节序列。在一些实施例中, 调节序列可能会失活, 使得它不再作为调节序列起作用。本文描述了调节序列的实例。

[0400] 通过无缝诱变修饰细胞中靶多核苷酸序列中的一个或多个核苷酸的方法利用到插入盒。在一些实施例中, 该插入盒(IC)存在于载体上。本文提供了载体的实例。本文所述的IC包括:

[0401] (i) 与该靶多核苷酸序列的一部分同源的第一区域,

[0402] (ii) 其包含靶多核苷酸序列中一个或多个核苷酸的突变的第二区域,

[0403] (iii) 第一核酸酶结合位点,

[0404] (iv) 编码标记基因的多核苷酸序列,

[0405] (v) 第二核酸酶结合位点

[0406] (vi) 包含靶多核苷酸序列中一个或多个核苷酸的突变的第三区域, 以及

[0407] (vii) 与靶多核苷酸序列的一部分同源的第四区域, 其中第一区域和第四区域与它们在该靶多核苷酸序列中的相应部分95%-100%相同。

[0408] 图28显示了示例性IC。在图28中, 该IC在5'至3' (相对于双链DNA的“顶部”或“编码”链) 方向包含: 第一核酸酶切割位点、第一核酸酶结合位点、抗性标记、第二核酸酶结合位点和第二核酸酶切割位点。第一和第二核酸酶切割位点包含靶多核苷酸序列内的希望的

核苷酸突变。

[0409] 如图27所示,“同源臂”(“HA”)存在于第一核酸酶切割位点的上游和第二核酸酶切割位点的下游。该“同源臂”包括与靶多核苷酸序列的一部分同源的区域。在一些实施例中,与靶多核苷酸序列的一部分同源的IC的第一区域包含第一核酸酶切割位点上游的HA。在一些实施例中,与靶多核苷酸序列的一部分同源的IC的第四区域包含第二核酸酶切割位点下游的HA。

[0410] 在一些实施例中,该IC包含与靶多核苷酸序列的一部分同源的第一区域。在一些实施例中,该IC包含与靶多核苷酸序列的一部分同源的第四区域。在一些实施例中,该IC中的第一和第四区域与它们在该靶多核苷酸序列中的相应部分分别具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、分别至少97%、至少98%、至少99%或100%的序列同一性。在一些实施例中,该IC中的第一和第四区域的HA具有约10至5000个碱基对、约20至2000个碱基对、或约50至1750个碱基对、或约100至1500个碱基对、或约200至1250个碱基对、或约300至1000个碱基对、或约400至约750个碱基对、或约500至600个碱基对。在一些实施例中,该IC中的第一和第四区域的HA具有约5、约10、约20、约30、约40、约50、约100、约250、约500、约1000、约1250、约1500、约1750、约2000、约2250或约2500个碱基对。

[0411] 在一些实施例中,该IC包含第二区域,该第二区域包含靶多核苷酸序列中一个或多个核苷酸的突变。在一些实施例中,该IC包含第三区域,该第三区域包含靶多核苷酸序列中一个或多个核苷酸的突变。如图28和29所示,该核酸酶切割位点包含靶多核苷酸序列内一个或多个核苷酸的突变。在一些实施例中,该核酸酶切割位点是任何合适的核酸酶的裂解位点。例如,该核酸酶切割位点可以是限制性内切酶的裂解位点,诸如像HindIII、BamHI、EcoRI、BbvI、FokI、MmeI等。在一些实施例中,该IC的第二区域包含含有希望的突变的第一核酸酶切割位点。在一些实施例中,该IC的第三区域包含含有希望的突变的第二核酸酶切割位点。在一些实施例中,该IC的第二和第三区域相同或基本相同。

[0412] 在一些实施例中,该IC包含第一和第二核酸酶结合位点。该核酸酶结合位点可以是任何合适的核酸酶的结合位点。例如,限制酶、锌指核酸酶、TALEN(转录激活剂样核酸内切酶)或Cas9的核酸酶结合位点。例如,如果该核酸酶是Cas9,则可以将引导RNA设计为与PAM上游(即,相对于相关DNA链的5')的任何序列杂交。因此,在一些实施例中,该核酸酶结合位点在PAM的上游。在一些实施例中,第一和第二核酸酶结合位点相同或基本相同。

[0413] 在一些实施例中,该IC包含编码标记基因的多核苷酸。“标记”基因用于确定核酸序列是否已被成功插入靶序列。标记基因可以是选择性标记(例如,抗性或选择标记)或可筛选标记(例如,荧光或比色标记)。

[0414] 抗性/选择标记的非限制性实例包括:抗生素抗性基因(例如,氨苄青霉素抗性基因、卡那霉素抗性基因等)和其他抗生素抗性基因;营养缺陷型标记(例如URA3、HIS3)和/或其他宿主细胞选择标记;促进向供体核酸内插入的核酸,例如,转座酶和反向重复序列,诸如用于转位到支原体基因组中;支持宿主细胞中复制和分离的核酸,诸如自主复制序列(ARS)或着丝粒序列(CEN)。

[0415] 可筛选标记将使含有标记基因的细胞看起来不同。可筛选标记的非限制性实例包括:绿色荧光蛋白(GFP)及其变体(例如,黄色荧光蛋白、红色荧光蛋白等); β -葡萄糖醛酸

酶,用于GUS分析中通过将其染成蓝色来检测细胞;以及X-gal,用于本领域技术人员熟知的蓝/白筛选中。

[0416] 表达标记基因的细胞的选择方法因所使用的标记而异。例如,若使用抗生素抗性标记,则选择包括在含有抗生素的培养基中培养细胞群并收集存活的细胞。若使用可筛选标记(诸如GFP),则选择涉及收集绿色的细胞。细胞收集可以,例如,通过从培养板手动挑选菌落,或通过使用流式细胞仪进行分选(例如,荧光激活细胞分选(FACS))来进行。

[0417] 在无缝诱变的方法的多个实施例中,该方法的第一步包括将包含IC的载体引入细胞中。可以使用本领域常规方法(诸如像转染、转导、细胞融合和脂质转染)将该载体引入细胞中。本文进一步描述了将载体引入细胞中。

[0418] 在无缝诱变的方法的多个实施例中,该方法的第二步包括通过同源重组将IC插入靶多核苷酸序列中以产生第一经修饰的靶多核苷酸。如图27所示,通过同源重组将抗性盒插入靶多核苷酸序列中(如“GATC”序列任一侧的叉所示)。如本文所述,对于特定的同源重组,该载体将包含与染色体序列具有同源性的足够长的区域(即,IC中的第一和第四区域),以允许载体与染色体的互补结合和掺入染色体。如本文所述,更长的同源性区域和更大程度的序列相似性可提高同源重组的效率。

[0419] 在无缝诱变的方法的多个实施例中,该方法的第三步包括选择表达标记基因的细胞。如本文所述,选择表达标记基因的细胞的方法取决于选择标记。本文描述了选择方法以及各种类型的标记基因。

[0420] 在无缝诱变的方法的多个实施例中,该方法的第四步包括使第一经修饰的靶多核苷酸(即,从上述步骤(2)产生的第一经修饰的靶多核苷酸)经过位点特异性核酸酶处理以产生具有粘性末端的第二经修饰的靶多核苷酸。在一些实施例中,粘性末端在IC的第二和第三区域中。位点特异性核酸酶可以是产生粘性末端的任何位点特异性核酸酶,包括但不限于限制性内切酶、本文所述的Cas9-核酸内切酶或本文所述的stiCas9。在一些实施例中,该核酸酶产生包含粘性末端的双链DNA断裂。在一些实施例中,该位点特异性核酸酶对于该细胞而言是外源性的,即,该位点特异性核酸酶不是天然存在于该细胞中。在一些实施例中,将该位点特异性核酸酶引入细胞中。在一些实施例中,将该位点特异性核酸酶作为编码位点特异性核酸酶的多核苷酸引入细胞中。本文描述了引入多核苷酸(诸如像载体)的方法,包括例如转染、转导、细胞融合和脂质转染。在一些实施例中,该位点特异性核酸酶是重组位点特异性核酸酶。如本文所述,重组蛋白质是指不是产生它们的细胞所固有的蛋白质,或具有由已知自然界中不存在的遗传物质的新组合产生的序列的蛋白质,诸如像由引入细胞中的外源性核酸表达的蛋白质。在一些实施例中,该重组位点特异性核酸酶由非细胞固有核酸表达。

[0421] 在一些实施例中,该位点特异性核酸酶是Cas9效应蛋白。本文描述了Cas9蛋白。在一些实施例中,该Cas9效应蛋白是II-B型Cas9。本文描述了II-B型Cas9蛋白,并且这些II-B型Cas9蛋白能够产生粘性末端。如本文所述,鉴定了II-B型CRISPR系统,尤其是通过cas操纵子上cas4基因的存在来鉴定,并且II-B Cas9蛋白属于TIGR03031 TIGRFAM蛋白家族。因此,在一些实施例中,该位点特异性核酸酶是TIGR03031 TIGRFAM蛋白家族的核苷酸。在一些实施例中,该位点特异性核酸酶包含以 $1E-5$ 的E值截止值匹配TIGR03031蛋白家族的结构域。在一些实施例中,该位点特异性核酸酶包含以 $1E-10$ 的E值截止值匹配TIGR03031蛋白家

族的结构域。在多个细菌物种中发现了II-B型CRISPR系统,诸如像嗜肺军团菌、新凶手弗朗西斯菌、 γ 变形菌HTCC5015、人粪便副萨特氏菌、华德萨特氏菌、硫磺单胞菌属物种SCADC、瘤胃杆菌属物种RM87、伯克霍尔德氏菌目细菌1_1_47、拟杆菌门口腔分类群274菌株F0058、产琥珀酸沃廉氏菌、伯克霍尔德氏菌目细菌YL45、嗜淀粉瘤胃杆菌、弯曲杆菌属物种P0111、弯曲杆菌属物种RM9261、拉尼尔弯曲杆菌菌株RM8001、拉尼尔弯曲杆菌菌株P0121、鼠毛滴虫、伦敦军团菌、沙姆盐弧菌、钩端螺旋体属物种分离株FW.030、莫里特拉氏菌属物种分离株NORP46、内生单胞菌属物种S-B4-1U、喜盐泰米尔纳德菌、需钠弧菌、斯氏弓形杆菌、蜃楼弗朗西斯氏菌、西班牙弗朗西斯氏菌或嗜盐副内生单胞菌。

[0422] 在一些实施例中,该位点特异性核酸酶是Cas9-核酸内切酶融合蛋白。本文描述了Cas9-核酸内切酶蛋白。在一些实施例中,该Cas9-核酸内切酶融合蛋白包含Cas9的DNA-靶向结构域和核酸内切酶的核酸酶结构域。在一些实施例中,该Cas9-核酸内切酶融合蛋白中的核酸内切酶是IIS型核酸内切酶。本文提供了IIS型核酸内切酶的实例,这些实例包括:BbvI、BgcI、BfuAI、BmpI、BspMI、CspCI、FokI、MboII、MmeI、NmeAIII和PleI。在一些实施例中,该Cas9-核酸内切酶融合蛋白中的核酸内切酶是FokI。由FokI进行的DNA裂解仅在两个FokI单体二聚化时发生。DNA的FokI裂解产生具有4个碱基对突出端的粘性末端。

[0423] 在一些实施例中,该Cas9-核酸内切酶融合蛋白包含经修饰的Cas9。本文中描述了经修饰的Cas9,并且包含无催化活性的Cas9和具有切口酶活性的Cas9。在一些实施例中,该经修饰的Cas9是无催化活性的Cas9(“deadCas9”)。无催化活性的Cas9不能裂解DNA(即,Cas9的裂解结构域失活);然而,它们通过与引导多核苷酸(例如,引导RNA)形成复合物而保留了靶向核酸序列的能力。本文描述了无催化活性的Cas9。在一些实施例中,无催化活性的Cas9相对于野生型Cas9包含双氨基酸取代。在一些实施例中,该双氨基酸取代是D10A和H840A。在一些实施例中,该Cas9-核酸内切酶融合蛋白包含无催化活性的Cas9,并且该核酸内切酶是FokI。

[0424] 在一些实施例中,该经修饰的Cas9是具有切口酶活性的Cas9(“Cas9切口酶”或“Cas9n”)。Cas9切口酶仅能够裂解双链DNA的一条链(即,“切口”该DNA)。本文描述了Cas9切口酶。在一些实施例中,Cas9切口酶包含相对于野生型Cas9的单个氨基酸取代。在一些实施例中,该单个氨基酸取代是D10A(“Cas9n^(D10A)”)。在一些实施例中,该单个氨基酸取代是H840A(“Cas9n^(H840A)”)。在一些实施例中,该Cas9-核酸内切酶融合蛋白包含具有切口酶活性的Cas9,并且该核酸内切酶是FokI。在一些实施例中,该Cas9-核酸内切酶融合蛋白包含具有D10A突变的Cas9,并且该核酸内切酶是FokI。在一些实施例中,该Cas9-核酸内切酶融合蛋白包含具有H840A突变的Cas9,并且该核酸内切酶是FokI。

[0425] 在一些实施例中,该位点特异性核酸酶是Cpf1。Cpf1(着丝粒和启动子因子1)是在CRISPR/Cpf1系统中发现的能够产生粘性末端的单RNA引导的核酸内切酶。CRISPR/Cpf1系统类似于CRISPR/Cas9系统。但是,Cas9和Cpf1之间存在几个显著差异。Cpf1不利用tracrRNA。Cpf1蛋白识别与Cas9不同的PAM序列。Cpf1的PAM序列是富含5' T的基序,诸如像5'-TTTN-3',其中N是A、T、C或G。Cpf1在与Cas9不同的位点进行裂解。Cas9在与PAM相邻的序列裂解,而Cpf1在远离PAM的序列裂解。Cpf1蛋白在例如国外专利公开GB 1506509.7、美国专利号9,580,701、美国专利公开2016/0208243和Zetsche等人,“Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System”[Cpf1是2类CRISPR-Cas系统的

单RNA引导核酸内切酶”],Cell[细胞]163(3):759-771(2015)中进一步描述,将其各自通过引用以其整体并入本文。

[0426] 在一些实施例中,该位点特异性核酸酶是Cas9、Cpf1或Cas9-FokI。

[0427] 在一些实施例中,由该位点特异性核酸酶产生的粘性末端包含5'突出端。在一些实施例中,由该位点特异性核酸酶产生的粘性末端包含3'突出端。在一些实施例中,该位点特异性核酸酶产生包含具有3至40个核苷酸的单链多核苷酸的粘性末端。在一些实施例中,该位点特异性核酸酶产生包含具有4至30个核苷酸的单链多核苷酸的粘性末端。在一些实施例中,该位点特异性核酸酶产生包含具有5至20个核苷酸的单链多核苷酸的粘性末端。在一些实施例中,该位点特异性核酸酶产生包含具有约5个核苷酸、约10个核苷酸、约15个核苷酸、约20个核苷酸、约25个核苷酸或约30个核苷酸的单链多核苷酸的粘性末端。在一些实施例中,deadCas9-FokI二聚体产生包含4-核苷酸5'突出端的粘性末端。在一些实施例中,Cas9n^(D10A)-FokI二聚体产生包含27-核苷酸5'突出端的粘性末端。在一些实施例中,Cas9^(H840A)-FokI二聚体产生包含23-核苷酸3'-突出端的粘性末端。

[0428] 在该方法的多个实施例中,该方法的第五步包括使该具有粘性末端的第二经修饰的靶多核苷酸经受连接酶处理,其中该连接酶在该第二区域和该第三区域处连接这些粘性末端以产生连接的经修饰的靶核酸,当与该靶多核苷酸序列相比时,该连接的经修饰的靶核酸包含一个或多个经修饰的核苷酸。连接酶是通过形成化学键来催化两个或更多个核酸片段的结合的酶。在一些实施例中,连接酶通过催化磷酸二酯键的形成将两个或更多个DNA片段连接在一起。可以使用任何合适的连接酶,并且可以由本领域技术人员确定该合适的连接酶。连接酶的非限制性实例包括:大肠杆菌连接酶、来自噬菌体T4的T4 DNA连接酶、DNA连接酶I、DNA连接酶II、DNA连接酶III、DNA连接酶IV,以及热稳定的连接酶,诸如Ampligase®DNA连接酶。连接酶可以连接平端或粘性末端。在一些实施例中,该连接酶连接粘性末端。在一些实施例中,该连接酶需要ATP以便连接DNA片段。

[0429] 在一些实施例中,该连接酶对于该细胞而言是外源性的,即,该连接酶不是天然存在于该细胞中。在一些实施例中,将该连接酶引入这些细胞中。在一些实施例中,将该连接酶作为编码连接酶的多核苷酸引入细胞中。本文描述了引入多核苷酸(诸如像载体)的方法。在一些实施例中,该连接酶是重组连接酶,即,从非细胞固有的核酸表达的连接酶。

[0430] 在一些实施例中,当与靶多核苷酸序列相比时,连接的经修饰的靶核酸包含一个或多个经修饰的核苷酸,但不包含标记基因或靶多核苷酸序列上游或下游的任何其他核苷酸,即,该靶多核苷酸序列产生无缝突变。

[0431] 在该方法的多个实施例中,在第三步之后,从细胞中分离出第一经修饰的靶核酸。从细胞中分离核酸的方法是本领域熟知的方法,并且包括例如苯酚/氯仿萃取、低pH/高盐条件下沉淀和固相萃取。可以使用可商购的用于分离核酸的试剂盒,诸如凯杰公司(QIAGEN)小量制备试剂盒、伯乐公司(Bio-Rad)QuantumPrep®小量制备试剂盒和酶研究公司(Zymo Research)ZYMOPURE质粒小量制备试剂盒。

[0432] 在该方法的多个实施例中,在第三骤之后,第一经修饰的靶核酸在该细胞中,即,没有从细胞中分离该核酸。在一些实施例中,该方法的步骤(1)-(5)在同一细胞内进行。在一些实施例中,该方法的组分被引入细胞中。在一些实施例中,将包含插入盒、位点特异性核酸酶和连接酶的载体引入细胞中。本文描述了将载体和蛋白质引入细胞的方法,这些方

法包含例如经由递送颗粒、囊泡和/或载体(包括病毒载体)的递送。

[0433] 在该方法的多个实施例中,该靶多核苷酸序列处于质粒中。本文描述了多种质粒及其实例。在一些实施例中,含有靶多核苷酸序列的质粒是天然细菌质粒(即,在细菌细胞中天然存在的质粒)。在一些实施例中,含有靶多核苷酸序列的质粒是引入细胞中的外源质粒。在一些实施例中,该细胞是细菌细胞。在一些实施例中,该质粒是工程化的质粒。在一些实施例中,对质粒中一个或多个核苷酸的修饰导致经修饰的细胞行为。该经修饰的行为可以是经修饰的蛋白质的表达、一种或多种蛋白质的更高或更低水平的表达、对抗生素的增加的抗性 or 易感性、对小分子和/或蛋白质的改变的反应、小分子和/或蛋白质的产生的改变等。

[0434] 在该方法的多个实施例中,该靶多核苷酸序列处于染色体中。该染色体可以是原核染色体或真核染色体。在一些实施例中,该染色体是真核细胞的染色体。在一些实施例中,该染色体是人类细胞的染色体。在一些实施例中,该染色体是动物细胞的染色体。在一些实施例中,该染色体是植物细胞的染色体。在一些实施例中,对染色体中一个或多个核苷酸的修饰导致经修饰的细胞行为。该经修饰的行为可以是经修饰的蛋白质的表达、一种或多种蛋白质的更高或更低水平的表达、对抗生素的增加的抗性 or 易感性、对小分子和/或蛋白质的改变的反应、小分子和/或蛋白质的产生的改变等。

[0435] 工程化的引导RNA (sgRNA)

[0436] 在一些实施例中,本披露提供了与st iCas9蛋白形成复合物的工程化的引导RNA,该工程化的引导RNA包含:(a)能够与真核细胞中的靶序列杂交的引导序列;以及(b)能够与Cas9蛋白结合的tracrRNA序列,其中该tracrRNA与天然存在的tracrRNA序列的差异为至少10个核苷酸,其中该工程化引导RNA提高了该Cas9蛋白的核酸酶效率。

[0437] 如本文所述,在一些实施例中,引导多核苷酸(例如,引导RNA)与Cas9蛋白形成复合物,即,在一些实施例中,引导多核苷酸与Cas9结合。在一些实施例中,该引导多核苷酸的DNA结合区段与真核细胞中的靶序列杂交,但不与细菌细胞中的序列杂交。

[0438] 在一些实施例中,该引导多核苷酸是10至150个核苷酸。在一些实施例中,该引导多核苷酸是20至120个核苷酸。在一些实施例中,该引导多核苷酸是30至100个核苷酸。在一些实施例中,该引导多核苷酸是40至80个核苷酸。在一些实施例中,该引导多核苷酸是50至60个核苷酸。在一些实施例中,该引导多核苷酸是10至35个核苷酸。在一些实施例中,该引导多核苷酸是15至30个核苷酸。在一些实施例中,该引导多核苷酸是20至25个核苷酸。

[0439] 可以将引导多核苷酸作为分离的分子(例如,RNA分子)引入靶细胞中,或使用包含编码引导多核苷酸的DNA的表达载体将其引入细胞中。

[0440] 天然存在的CRISPR系统利用crRNA和tracrRNA,其中crRNA包含与靶序列互补的区域, tracrRNA与Cas9蛋白结合并且与crRNA杂交。crRNA/tracrRNA杂交体形成RNA二级结构,该RNA二级结构可允许crRNA部分与靶序列结合以及tracrRNA部分与Cas9蛋白结合。RNA二级结构的非限制性实例包括螺旋、茎环和假结。在一些实施例中,该Cas9蛋白识别crRNA/tracrRNA杂交体中的至少一个茎环用于结合。

[0441] 在工程化的CRISPR-Cas系统诸如像本披露的CRISPR-Cas系统中,利用既可以与靶序列互补又可以与Cas9蛋白结合的单个引导多核苷酸可能是有利的。因此,在一些实施例中,本披露提供了非天然存在的CRISPR-Cas系统,该系统包含能够产生粘性末端的Cas9效

应蛋白(stiCas9);以及与stiCas9形成复合物并包含引导序列的引导多核苷酸,其中该引导序列能够与真核细胞中的靶序列杂交,但不与细菌细胞中的序列杂交;其中该复合物在自然界中不存在,并且其中该系统不包含tracrRNA。在一些实施例中,该引导多核苷酸形成至少一种二级结构。在一些实施例中,该至少一种二级结构是茎环、螺旋或假结中的一种。

[0442] 为改善对Cas9蛋白的结合亲和力和/或增加对靶序列的靶向效率,优化本文所述的工程化的引导多核苷酸可能是有利的。参见,例如,Dang等人,Genome Biology[基因组生物学]16:280(2015);Nowak等人,Nucleic Acids Res[核酸研究]44(20):9555-9564(2016);以及Vejnar等人,Cold Spring Harb Protoc[冷泉港实验手册],doi:10.1101/pdb.top090894(2016)。在一些实施例中,该工程化的引导多核苷酸,例如,引导RNA,比天然存在的crRNA和tracrRNA的组合更短。在一些实施例中,该工程化的引导RNA比天然存在的crRNA和tracrRNA的组合短至少5个核苷酸、短至少6个核苷酸、短至少7个核苷酸、短至少8个核苷酸、短至少9个核苷酸、短至少10个核苷酸、至少短11个核苷酸、短至少12个核苷酸、短至少13个核苷酸、短至少14个核苷酸、短至少15个核苷酸、短至少16个核苷酸、短至少17个核苷酸、短至少18个核苷酸、短至少19个核苷酸、短至少20个核苷酸、短至少21个核苷酸、短至少22个核苷酸、短至少23个核苷酸、短至少24个核苷酸、短至少25个核苷酸、短至少26个核苷酸、短至少27个核苷酸、短至少28个核苷酸、短至少29个核苷酸或短至少30个核苷酸。

[0443] 在一些实施例中,该tracrRNA序列比天然存在的tracrRNA序列短至少5个核苷酸、短至少6个核苷酸、短至少7个核苷酸、短至少8个核苷酸、短至少8个核苷酸、短至少9个核苷酸、短至少10个核苷酸、至少短11个核苷酸、短至少12个核苷酸、短至少13个核苷酸、短至少14个核苷酸、短至少15个核苷酸、短至少16个核苷酸、短至少17个核苷酸、短至少18个核苷酸、短至少19个核苷酸、短至少20个核苷酸、短至少21个核苷酸、短至少22个核苷酸、短至少23个核苷酸、短至少24个核苷酸、短至少25个核苷酸、短至少26个核苷酸、短至少27个核苷酸、短至少28个核苷酸、短至少29个核苷酸或短至少30个核苷酸。

[0444] 在一些实施例中,该工程化的引导多核苷酸比天然存在的crRNA和tracrRNA的组合短5个核苷酸至40个核苷酸、短6个核苷酸至40个核苷酸、短7个核苷酸至40个核苷酸、短8个核苷酸至40个核苷酸、短9个核苷酸至40个核苷酸、短10个核苷酸至40个核苷酸、短11个核苷酸至40个核苷酸、短12个核苷酸至40个核苷酸、短13个核苷酸至40个核苷酸、短14个核苷酸至40个核苷酸、短15个核苷酸至40个核苷酸、短16个核苷酸至40个核苷酸、短17个核苷酸至40个核苷酸、短18个核苷酸至40个核苷酸、短19个核苷酸至40个核苷酸、短20个核苷酸至40个核苷酸、短21个核苷酸至40个核苷酸、短22个核苷酸至40个核苷酸、短23个核苷酸至40个核苷酸、短24个核苷酸至40个核苷酸、短25个核苷酸至40个核苷酸、短26个核苷酸至40个核苷酸、短27个核苷酸至40个核苷酸、短28个核苷酸至40个核苷酸、短29个核苷酸至40个核苷酸、短30个核苷酸至40个核苷酸、短31个核苷酸至40个核苷酸、短32个核苷酸至40个核苷酸、短33个核苷酸至40个核苷酸、短34个核苷酸至40个核苷酸、短35个核苷酸至40个核苷酸、短36个核苷酸至40个核苷酸、短37个核苷酸至40个核苷酸、短38个核苷酸至40个核苷酸或短39个核苷酸至40个核苷酸。

[0445] 在一些实施例中,该工程化的tracrRNA比天然存在的tracrRNA短5个核苷酸至40个核苷酸、短6个核苷酸至40个核苷酸、短7个核苷酸至40个核苷酸、短8个核苷酸至40个核

核苷酸、短9个核苷酸至40个核苷酸、短10个核苷酸至40个核苷酸、短11个核苷酸至40个核苷酸、短12个核苷酸至40个核苷酸、短13个核苷酸至40个核苷酸、短14个核苷酸至40个核苷酸、短15个核苷酸至40个核苷酸、短16个核苷酸至40个核苷酸、短17个核苷酸至40个核苷酸、短18个核苷酸至40个核苷酸、短19个核苷酸至40个核苷酸、短20个核苷酸至40个核苷酸、短21个核苷酸至40个核苷酸、短22个核苷酸至40个核苷酸、短23个核苷酸至40个核苷酸、短24个核苷酸至40个核苷酸、短25个核苷酸至40个核苷酸、短26个核苷酸至40个核苷酸、短27个核苷酸至40个核苷酸、短28个核苷酸至40个核苷酸、短29个核苷酸至40个核苷酸、短30个核苷酸至40个核苷酸、短31个核苷酸至40个核苷酸、短32个核苷酸至40个核苷酸、短33个核苷酸至40个核苷酸、短34个核苷酸至40个核苷酸、短35个核苷酸至40个核苷酸、短36个核苷酸至40个核苷酸、短37个核苷酸至40个核苷酸、短38个核苷酸至40个核苷酸或短39个核苷酸至40个核苷酸。

[0446] 在一些实施例中,该工程化的引导多核苷酸,例如,引导RNA,比天然存在的crRNA和tracrRNA的组合更长。在一些实施例中,该工程化的引导RNA比天然存在的crRNA和tracrRNA的组合长至少5个核苷酸、长至少6个核苷酸、长至少7个核苷酸、长至少8个核苷酸、长至少8个核苷酸、长至少9个核苷酸、长至少10个核苷酸、至少长11个核苷酸、长至少12个核苷酸、长至少13个核苷酸、长至少14个核苷酸、长至少15个核苷酸、长至少16个核苷酸、长至少17个核苷酸、长至少18个核苷酸、长至少19个核苷酸、长至少20个核苷酸、长至少21个核苷酸、长至少22个核苷酸、长至少23个核苷酸、长至少24个核苷酸、长至少25个核苷酸、长至少26个核苷酸、长至少27个核苷酸、长至少28个核苷酸、长至少29个核苷酸或长至少30个核苷酸。

[0447] 在一些实施例中,该tracrRNA序列比天然存在的tracrRNA序列长至少5个核苷酸、长至少6个核苷酸、长至少7个核苷酸、长至少8个核苷酸、长至少8个核苷酸、长至少9个核苷酸、长至少10个核苷酸、至少长11个核苷酸、长至少12个核苷酸、长至少13个核苷酸、长至少14个核苷酸、长至少15个核苷酸、长至少16个核苷酸、长至少17个核苷酸、长至少18个核苷酸、长至少19个核苷酸、长至少20个核苷酸、长至少21个核苷酸、长至少22个核苷酸、长至少23个核苷酸、长至少24个核苷酸、长至少25个核苷酸、长至少26个核苷酸、长至少27个核苷酸、长至少28个核苷酸、长至少29个核苷酸或长至少30个核苷酸。

[0448] 在一些实施例中,该工程化的引导多核苷酸比天然存在的crRNA和tracrRNA的组合长5个核苷酸至40个核苷酸、长6个核苷酸至40个核苷酸、长7个核苷酸至40个核苷酸、长8个核苷酸至40个核苷酸、长9个核苷酸至40个核苷酸、长10个核苷酸至40个核苷酸、长11个核苷酸至40个核苷酸、长12个核苷酸至40个核苷酸、长13个核苷酸至40个核苷酸、长14个核苷酸至40个核苷酸、长15个核苷酸至40个核苷酸、长16个核苷酸至40个核苷酸、长17个核苷酸至40个核苷酸、长18个核苷酸至40个核苷酸、长19个核苷酸至40个核苷酸、长20个核苷酸至40个核苷酸、长21个核苷酸至40个核苷酸、长22个核苷酸至40个核苷酸、长23个核苷酸至40个核苷酸、长24个核苷酸至40个核苷酸、长25个核苷酸至40个核苷酸、长26个核苷酸至40个核苷酸、长27个核苷酸至40个核苷酸、长28个核苷酸至40个核苷酸、长29个核苷酸至40个核苷酸、长30个核苷酸至40个核苷酸、长31个核苷酸至40个核苷酸、长32个核苷酸至40个核苷酸、长33个核苷酸至40个核苷酸、长34个核苷酸至40个核苷酸、长35个核苷酸至40个核苷酸、长36个核苷酸至40个核苷酸、长37个核苷酸至40个核苷酸、长38个核苷酸至40个核苷酸

或长39个核苷酸至40个核苷酸。

[0449] 在一些实施例中,该工程化的tracrRNA比天然存在的tracrRNA长5个核苷酸至40个核苷酸、长6个核苷酸至40个核苷酸、长7个核苷酸至40个核苷酸、长8个核苷酸至40个核苷酸、长9个核苷酸至40个核苷酸、长10个核苷酸至40个核苷酸、长11个核苷酸至40个核苷酸、长12个核苷酸至40个核苷酸、长13个核苷酸至40个核苷酸、长14个核苷酸至40个核苷酸、长15个核苷酸至40个核苷酸、长16个核苷酸至40个核苷酸、长17个核苷酸至40个核苷酸、长18个核苷酸至40个核苷酸、长19个核苷酸至40个核苷酸、长20个核苷酸至40个核苷酸、长21个核苷酸至40个核苷酸、长22个核苷酸至40个核苷酸、长23个核苷酸至40个核苷酸、长24个核苷酸至40个核苷酸、长25个核苷酸至40个核苷酸、长26个核苷酸至40个核苷酸、长27个核苷酸至40个、长28个核苷酸至40个核苷酸、长29个核苷酸至40个核苷酸、长30个核苷酸至40个核苷酸、长31个核苷酸至40个核苷酸、长32个核苷酸至40个核苷酸、长33个核苷酸至40个核苷酸、长34个核苷酸至40个核苷酸、长35个核苷酸至40个核苷酸、长36个核苷酸至40个核苷酸、长37个核苷酸至40个核苷酸、长38个核苷酸至40个核苷酸或长39个核苷酸至40个核苷酸。

[0450] 在一些实施例中,该工程化的引导多核苷酸与天然存在的crRNA和tracrRNA的组合的差异为至少一个核苷酸,使得该工程化的引导多核苷酸的结合亲和力和/或靶向效率高于天然存在的crRNA/tracrRNA杂交体的结合亲和力和/或靶向效率。在一些实施例中,该工程化的引导多核苷酸与crRNA/tracrRNA杂交体的差异为至少2个、至少3个、至少4个、至少5个、至少6个、至少7个、至少8个、至少9个、至少10个、至少11个、至少12个、至少13个、至少14个、至少15个、至少16个、至少17个、至少18个、至少19个、至少20个、至少21个、至少22个、至少23个、至少24个、至少25个、至少26个、至少27个、至少28个、至少29个、或至少30个核苷酸。在一些实施例中,该工程化的tracrRNA与天然存在的tracrRNA的差异为至少2个、至少3个、至少4个、至少5个、至少6个、至少7个、至少8个、至少9个、至少10个、至少11个、至少12个、至少13个、至少14个、至少15个、至少16个、至少17个、至少18个、至少19个、至少20个、至少21个、至少22个、至少23个、至少24个、至少25个、至少26个、至少27个、至少28个、至少29个、或至少30个核苷酸。

[0451] 在一些实施例中,对天然存在的tracrRNA进行修饰以改善Cas9蛋白的核酸酶效率。在一些实施例中,该修饰在tracrRNA的茎环中进行。在一些实施例中,该修饰是茎环的延长。在一些实施例中,该修饰是茎环的缩短。在一些实施例中,该修饰是茎环中的一个或多个核苷酸取代。在一些实施例中,该修饰是对如图41所示的茎环进行的。

[0452] 在一些实施例中,使用工程化的引导RNA,将Cas9蛋白的核酸酶效率改善了至少约30%、至少约40%、至少约50%、至少约60%、至少约70%、至少约80%、至少约90%、或至少约100%。在一些实施例中,使用工程化的引导RNA,将Cas9蛋白的核酸酶效率改善了至少约两倍、至少约三倍、至少约四倍、至少约五倍、至少约六倍、至少约七倍、至少约八倍、至少约九倍或至少约十倍。

[0453] 可以测量Cas9蛋白的核酸酶效率,例如,以便将与天然存在的引导RNA复合的Cas9蛋白的核酸酶效率和与本文所述的工程化的引导RNA复合的Cas9蛋白相比较。在一些实施例中,该测量方法是生化测定,诸如像测量针对线性或圆形模板的体外Cas9核酸酶活性率。在一些实施例中,该测量方法使用例如下一代测序、T7核酸内切酶I测定和/或细胞测定来

测量Cas9蛋白的靶向效率。在一些实施例中，测量方法是使用例如BIACORE系统进行的在Cas9蛋白和tracrRNA之间的亲和力测试。

[0454] 在一些实施例中，该引导序列与SEQ ID NO:104-125或196-199中的任一个具有至少90%的序列同一性。在一些实施例中，该tracrRNA序列与SEQ ID NO:148-171中的任一个具有至少90%的序列同一性。在一些实施例中，该引导RNA具有与SEQ ID NO:172-191中的任一个至少90%的序列同一性。

[0455] 在一些实施例中，该工程化的引导RNA或该引导RNA的crRNA部分与SEQ ID NO:104-125或196-199中的任一个具有至少90%的序列同一性。在一些实施例中，该引导RNA或该引导RNA的crRNA部分与SEQ ID NO:104-125或196-199中的任一个具有至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%的序列同一性。

[0456] 在一些实施例中，工程化的引导多核苷酸的蛋白质结合区段或tracrRNA序列与SEQ ID NO:102和148-171中的任一个具有至少90%的序列同一性。在一些实施例中，该工程化的引导多核苷酸的蛋白质结合区段与SEQ ID NO:102或148-171中的任一个具有至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%的序列同一性。

[0457] 在一些实施例中，本披露提供了Cas9蛋白的工程化引导多核苷酸，该工程化的引导多核苷酸与SEQ ID NO:172-191中的任一个具有至少90%的序列同一性。在一些实施例中，该工程化的引导多核苷酸与SEQ ID NO:172-191中的任一个具有至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%的序列同一性。

[0458] 可以使用生物信息学工具结合生化验证设计本文所述的引导多核苷酸。设计引导多核苷酸的示例性方法如下：(1) 使用蛋白质BLAST找到相关的CRISPR操纵子；(2) 搜索已经在基因组中注释的crRNA，或使用例如CRISPR-Finder对该CRISPR添加注释；(3) 使用比对工具，例如，CLC基因组学工作台(凯杰公司)，确定tracrRNA的可能位置；(4) 在与该crRNA相似的区域附近搜索TATAA盒；(5) 测试该crRNA的二级结构以及在比对过程中发现的所有可能的tracrRNA，并选择组成希望的二级结构的crRNA/tracrRNA杂交体；以及(6) 修剪该crRNA和tracrRNA，以产生短引导RNA(sgRNA)。例如，可以将本文所述的crRNA和tracrRNA序列组合以产生sgRNA。在一些实施例中，该crRNA和tracrRNA序列如表1中所示组合，以产生sgRNA。

[0459] 表1. Cas9蛋白的短引导RNA序列(sgRNA)

[0460]

Cas9蛋白	crRNA SEQ ID NO	tracrRNA SEQ ID NO
LpCas9	104	148
SsCas9	105	149
WsCas9	106	150
BbCas9	107	151

PeCas9	108	152
SwCas9	109	153
RaCas9	110	154
Csp1Cas9	111	155
Csp2Cas9	112	156
Cl1Cas9	113	157
Cl2Cas9	114	158
MH0245Cas9	115	159
FnCas9	116	160
GpCas9	117	161;162
TmCas9	118	163
LlCas9	119	164
SshCas9	120	165
Lept.Cas9	121	166
莫里特拉氏菌属Cas9	122	167
ExCas9	123	168
TsCas9	124	169
VnCas9	125	170;171

[0461] 实例

[0462] 实例1-在AAVS1基因座处进行靶向基因插入

[0463] 该实例使用如本文披露的无缝诱变 (ObLiGaRe 2.0系统) 验证基因向AAVS1基因座中的插入。

[0464] 如图12和14所示产生了两个Cas9n-FokI变体,即Cas9n^{D10A}和Cas9n^{H840A}。如图13和15所示产生了两个供体载体,这两个供体载体含有SA-2A-Puro选择盒上游的ObLiGaRe 2.0靶位点(在图中表示为区域2和区域1)。该供体载体的大小为6kb。如图16所示,基于AAVS1基因座设计了ObLiGaRe 2.0靶位点。

[0465] 将编码Cas9n-FokI变体、4种单独克隆的引导RNA (gRNA) 和相应供体载体之一的质粒共转染到HEK293细胞中。嘌呤霉素抗性盒(供体质粒上的目的基因)的基因组插入在图15中示意性地示出。

[0466] 选择具有嘌呤霉素抗性的细胞,并收集嘌呤霉素抗性细胞的基因组DNA并且对其进行连接PCR。对这些PCR产物进行TOPQ克隆并通过桑格测序来测序以确定这些连接处的精度。

[0467] 使用Cas9n^{D10A}-FokI进行基因插入的5'连接的序列在图17中示出。使用Cas9n^{H840A}-FokI进行基因插入的5'连接的序列在图18中示出。因此,使用ObLiGaRe2.0系统将转基因盒成功敲入AAVS1基因座,在预期的接头上具有很高的精度。

[0468] 实例2-评估不进行抗生素选择的靶向插入的效率,以及间区序列长度对基因插入效率的影响

[0469] 在该实例中,使用不需要抗生素选择的实验设置测试间区序列长度(两个gRNA之间的偏移序列)对基因插入效率的影响。

[0470] 选择AAVS1-Exon2基因座作为靶位点。如图19所示,设计并克隆了靶向10个靶位点的所需gRNA(间区序列长度不同)。因此,如图20所示产生了含有设计的ObLiGaRe 2.0靶位点和mCherry的10个供体载体(在EF1a启动子的控制下)。

[0471] 将编码Cas9n^{H840A}-FokI和2AGFP的质粒、gRNA中的2种和供体载体共转染到HEK293细胞中。按如下进行选择:首先通过FACS针对GFP表达对细胞进行分选,指示引入了活性Cas9n-FokI。然后,使细胞传代至少10次,然后通过FACS针对mCherry表达进行分选,指示mCherry插入靶位点。图21中示出了该示意图。

[0472] 具有mCherry的细胞百分比对比间区序列长度(以碱基对指示)的结果示于图22中。17bp的间区序列长度指示mCherry插入效率最高(约20%)。因此,未施加抗生素选择的情况下用ObLiGaRe 2.0实现了高效率的转基因插入。

[0473] 实例3-不同基因插入方法的效率的比较

[0474] 在该实例中,比较了使用ObLiGaRe(使用锌指核酸酶)和ObLiGaRe 2.0的基因插入。

[0475] 使用ObLiGaRe基因插入将基因插入AAVS1-int1基因座中。将使用Cas9n-FokI变体的ObLiGaRe 2.0与2种或4种gRNA一起使用,靶向AAVS1-int1和SERPINA1-内含子1基因座中的三个位点。还测试了使用deadCas9-FokI的ObLiGaRe 2.0。如实例2所述开展实验程序(无抗生素选择,并且基于mCherry阳性细胞的FACS测量进行细胞选择)。该SERPINA1基因座的供体质粒显示在图23中。图24显示了使用deadCas9-FokI进行供体质粒上目的基因的基因组插入。

[0476] 图25显示了针对测试的每种基因插入方法获得的结果。这些结果是在一个实验中从三个独立的生物学复制中获得的。误差条指示S.E.M.。AAVS1-int1基因座上基于锌指核酸酶的ObLiGaRe(“AAVS1-int-ZFN”)和Cas9n^{D10A}-FokI(“AAVS1-int-C9nF-A”)的效率相当。不同基因座之间ObLiGaRe 2.0效率的差异可能归因于gRNA的效率。通过评估靶位点和不同间区序列长度的组合实现了较高的基因插入效率。

[0477] 实例4-无缝诱变

[0478] 在该实例中,描述了本文的披露内容中提供的用于无缝诱变的一般方法。无缝诱变的希望结果在图26中显示,其中在靶位点进行突变而不改变靶中的任何序列。

[0479] 图27显示了该方法的步骤1。将侧接同源臂的抗性盒引入具有靶序列的细胞中,并通过同源重组插入靶区域中。选择含有抗性盒的细胞。

[0480] 图28显示了该抗性盒的特写。核酸酶切割位点和核酸酶结合位点存在于抗性盒的两侧。能够产生突出端的核酸酶(诸如Cpf1或Cas9)在核酸酶切割位点裂解,产生包含希望的点突变的突出端。

[0481] 图29显示了该方法的步骤2。体外或体内连接使用由核酸酶产生的相容突出端以去除抗性盒。从而插入了点突变而不会留下任何“疤痕”,即,任何额外的序列。实例5中描述了用于核酸消化和连接的方案。

[0482] 实例5-使用Cpf1进行无缝诱变的方案

[0483] 在该实例中,核酸消化和连接如下进行:

[0484] 消化

[0485] 1. 在无RNA酶的0.5mL试管中一起添加:

- [0486] 1 μ L Cas910X缓冲液
- [0487] 1 μ L Cpf1蛋白(10 μ g/ μ L)
- [0488] 1 μ L gRNA
- [0489] 最多10 μ L不含RNA酶的H₂O(这一量由步骤3中添加的DNA量确定)。
- [0490] 2.在室温下孵育5分钟。
- [0491] 3.添加2-2.5 μ g待切割的质粒DNA(这一体积将随浓度而变化;相应地调整步骤1中的水的量)。
- [0492] 4.在37 $^{\circ}$ C下孵育2小时。
- [0493] 5.消化后,在150V下用1.5%琼脂糖凝胶进行凝胶电泳。
- [0494] 凝胶提取
- [0495] 6.从凝胶上切割具有适当长度的DNA。
- [0496] 7.使用凝胶提取试剂盒(例如,来自凯杰公司的试剂盒)从凝胶中提取DNA。
- [0497] 8.在NANODROP上测量DNA浓度。
- [0498] 连接
- [0499] 9.在PCR管中一起添加:
- [0500] 25-30ng质粒DNA(这一体积将随浓度而变化)
- [0501] 1 μ L DTT
- [0502] 1 μ L 10X T4连接酶缓冲液
- [0503] 1 μ L T4连接酶
- [0504] 最多10 μ L H₂O
- [0505] 10.在16 $^{\circ}$ C下孵育2小时。
- [0506] 11.使用10 μ L进行转化。
- [0507] 转化
- [0508] 12.通过将其置于冰上10分钟解冻从-80 $^{\circ}$ C的冰箱中取出的NEB 10 β 细胞(新英格兰生物实验室(NEW ENGLAND BIOLABS))。每个小瓶含有50 μ L(足以进行3次转化)。解冻SOC培养基。
- [0509] 13.将10 μ L的连接反应添加到1.5mL EPPENDORF管中,并置于冰上冷却。
- [0510] 14.解冻后,将15 μ L的NEB10 β 细胞添加到连接反应中。
- [0511] 15.置于冰上30分钟。42 $^{\circ}$ C水浴加温。
- [0512] 16.将细胞在42 $^{\circ}$ C水浴中放置30秒钟,然后置于冰上2分钟,以使其热休克。
- [0513] 17.向这些细胞中添加300 μ L的SOC培养基,并在37 $^{\circ}$ C下孵育45分钟。
- [0514] 18.在板的1/3上铺板100 μ L细胞,或在整个板上铺板300 μ L细胞;板上含有适当的抗生素。
- [0515] 实例6-Cas9体外消化方案
- [0516] 在该实例中,通过Cas9按如下进行底物DNA的体外消化(对于30 μ L反应):
- [0517] 1.按以下顺序在室温下组装反应:
- [0518] 20 μ L无核酸酶的水
- [0519] 3 μ L 10X Cas9核酸酶反应缓冲液
- [0520] 3 μ L 300nM sgRNA(30nM终浓度)

[0521] 1 μ L 1 μ M Cas9核酸酶(约30nM终浓度)

[0522] 在25 $^{\circ}$ C下预孵育10分钟,然后添加:

[0523] 3 μ L 30nM底物DNA

[0524] 2.充分混合并在微量离心管中脉冲旋转(pulse-spin)。

[0525] 3.在37 $^{\circ}$ C下孵育15分钟。

[0526] 4.向每个样品中添加1 μ L蛋白酶K。充分混合并在微量离心管中脉冲旋转。

[0527] 5.在室温下孵育10分钟。

[0528] 6.继续进行片段分析。

[0529] 实例7-Cas9裂解后的DNA修复谱的分析

[0530] 在该实例中,通过搜索操纵子中cas4的存在,使用计算分析来鉴定II-B型Cas9操纵子。选择来自新凶手弗朗西斯菌的Cas9蛋白(FnCas9)进行生产。如图34A所示,核酸酶活性在体外裂解测定中得到证实。裂解产物的桑格测序揭示,FnCas9在体外产生5'粘性末端,如图34B所示。在HEK293人类细胞系中验证了该蛋白质表达构建体。使用RIMA比较FnCas9和来自酿脓链球菌的Cas9蛋白(SpyCas9)情形下的突变模式,如图34C所示。

[0531] 实例8-Cas9处理后DNA切割谱的分析

[0532] 如实例7中所述,来自新凶手弗朗西斯菌的II-B型Cas9变体(FnCas9)被证明在哺乳动物细胞中形成多个具有低编辑效率的粘性末端。测试了II-B型Cas9家族的其他成员是否产生粘性末端。从测序的肠道宏基因组MH0245鉴定了一个新的Cas9变体(MHCas9)。图33显示了为MHCas9设计的引导RNA、tracrRNA和crRNA的序列。体外测定显示,MHCas9能够裂解DNA片段,如图35A所示。桑格测序揭示MHCas9在体外产生5'突出端,如图35B所示。此外,进行细胞测定以验证MHCas9在HEK293-REMINDEL人类细胞系中也起作用,如图35C所示。

[0533] 图36A显示了来自MHCas9的crRNA/tracrRNA的序列。图36B显示了crRNA/tracrRNA(指示二级结构)的一种方案。图36C中的截短的系统发育树显示了MHCas9与其他II-B型Cas9的比对,包括来自硫磺单胞菌属物种SCADC的Cas9蛋白(ssCas9)、来自产琥珀酸沃廉氏菌的Cas蛋白(WsCas9)、来自嗜肺军团菌的Cas9蛋白(LpCas9)和FnCas9。如该系统发育树所示,FnCas9和MHCas9差异很大。然而,实例7和本实例中描述的实验结果显示MHCas9和FnCas9具有相同的裂解机制。

[0534] 实例9-sgRNA的设计

[0535] 在该实例中,描述了关于设计sgRNA的方法学:

[0536] 1.使用蛋白BLAST(NCBI,blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins)查找相关的CRISPR操纵子。对于该搜索中出现的每个物种,选择一个RefSeq进行进一步分析。在各种输入和不同设置下运行几次BLAST。

[0537] 2.检查已添加注释的CRISPR RNA(crRNA)。若不存在,使用CRISPR-Finder(crspr.i2bc.paris-saclay.fr/Server/)对crRNA添加注释。

[0538] 3.使用CLC基因组学工作台v.9.5(凯杰公司)中的“创建比对”来找到tracrRNA的可能位置。将该crRNA的两条链与Cas4和CRISPR重复序列之间的序列进行比对。

[0539] 4.在显示与crRNA有相似性的区域附近寻找TATAA盒。

[0540] 5.使用所有可能的tracrRNA(在比对中找到的)测试crRNA的二级结构,并选择能形成理想结构的二级结构。

[0541] 6. 修剪该crRNA和tracrRNA,以生成短引导RNA (sgRNA)。

[0542] 图41A-T示出了通过本文描述的方法设计的各种sgRNA。图42A-L示出了通过修剪对sgRNA (也称为“嵌合gRNA”) 进行的优化,以及用于进一步修饰的可能靶位点。

[0543] 实例10-经修饰的sgRNA的体外消化测定

[0544] 如图45所概括的通过去除各种核苷酸对四种不同的引导RNA进行工程化(引导-1、引导-2、引导-3、引导-4)。然后在体外消化测定中将该经修饰的引导RNA与原始引导RNA进行比较。图45证明一些修饰改善了MHCas9的消化效率。

[0545] 在三种不同的Cas9系统中进一步研究了引导RNA的长度:SpyCas9、C11Cas9和MHCas9。制备了长度为19-23的引导RNA,然后将新的Cas9变体和工程化的引导RNA转染到报告细胞系中,并进行Surveyor™核酸酶测定(集成DNA技术公司,伊利诺伊州斯科基)。图46证明了新的Cas9变异C11和MH的体外切割效率和功能性。

[0546] 实例11-MHCas9的PAM序列

[0547] 使用图49A中示意性示出的方法研究了MHCas9的优选PAM序列。生成了涵盖各种PAM序列组合和靶标裂解位点的64个质粒的合并文库。使用SpCas9和MHCas9分别消化该文库。使用该质粒的正向和反向引物扩增含有靶标裂解位点和该PAM的区域,然后通过下一代测序对扩增的区域进行测序。含有SpCas9或MHCas9的优选PAM序列的质粒被消化,因此未进行扩增或测序。另一方面,含有SpCas9或MHCas9的非优选PAM序列的这些质粒没有被消化并且可以被扩增。

[0548] 图49B显示了SpCas9和MHCas9的“耗尽的”PAM序列的结果。与SpCas9相比,MHCas9对“NGG”PAM序列的偏好较不严格。

[0549] 实例12-将Cas9蛋白与核酸外切酶偶联

[0550] 将通过II-B型Cas9蛋白进行的裂解结合末端加工核酸外切酶以提高编辑效率。图50示出了该方法的示意图。如图50A所示,由II-B型Cas9进行的裂解产生的突出端可以被细胞精确地修复以回复到原始序列,从而当希望插入-缺失或取代修饰时限制编辑效率。在图50B中,在被II-B型Cas9裂解之后,引入了末端加工核酸外切酶Artemis或TREX2,该核酸外切酶进一步加工II-B型Cas9切割位点处的裂解的突出端。这些加工的末端的细胞修复导致相对于原始序列的修复不精确(即,增加的插入-缺失或取代修饰的数目),从而提高了编辑效率。

[0551] 为了测试Cas9与核酸外切酶的结合的效果,测试了具有或不具有末端加工酶的II-B型Cas9在人类细胞系中的活性。图51A显示了实验程序的示意性概览。将编码各种II-B型Cas9蛋白(FnCas9、C11Cas9、MHCas9)和II-A型SpCas9的质粒连同编码末端加工酶FnCas4或TREX2的质粒以及编码三种不同的引导RNA序列的质粒引入HEK293细胞中。转染72小时后,收获HEK293细胞的基因组DNA,并通过下一代测序进行分析。

[0552] 结果显示于图51B中。用对照质粒转染的细胞仅显示背景水平的修饰(归因于测序中的自然变异)。在存在或不存在末端加工酶的情况下,FnCas9、MHCas9和SpCas9均显示出不同量的基因组修饰。通常,相对于没有末端加工酶,用末端加工酶引入Cas9显示增加数目的修饰。

[0553] 实例13-Cas9蛋白的突变模式分析

[0554] 对由不同Cas9进行的切割进行了突变模式分析。用SpCas9、C11Cas9或MHCas9及其

相应的引导RNA转染HEK293细胞。72小时后裂解细胞,提取基因组DNA并进行下一代扩增子测序。使用生物信息学工具分析测序读数,以量化检测到的经修饰读数中每个突变的相对频率。

[0555] 结果显示于图52中。图52A、52B和52C分别显示了使用SpCas9、C11Cas9和MHCas9诱导切割后相同靶序列的突变模式。该靶序列显示在每个图的顶部。这些结果指示,在使用不同的Cas9蛋白诱导切割后,在同一基因座处的突变模式是不同的,指示对于不同的Cas9s,核酸酶活性的模式不同。

[0556] 核酸酶活性差异的一个非限制性假设可能是II-A型和II-B型Cas9蛋白之间的RuvC和HNH核酸酶结构域构型不同。如图53中所示,II-A型Cas9(图A)指示其RuvC和HNH结构域的切割位点相同(例如,NGG PAM序列上游大约3个核苷酸),这导致了多个平端或单个核苷酸突出端。另一方面,II-B型Cas9(图B)指示RuvC和HNH的切割位点偏移(例如,在NGG PAM序列的上游分别大约7个和3个核苷酸),这导致了多个“粘性”末端,即,3-4个核苷酸的突出端。

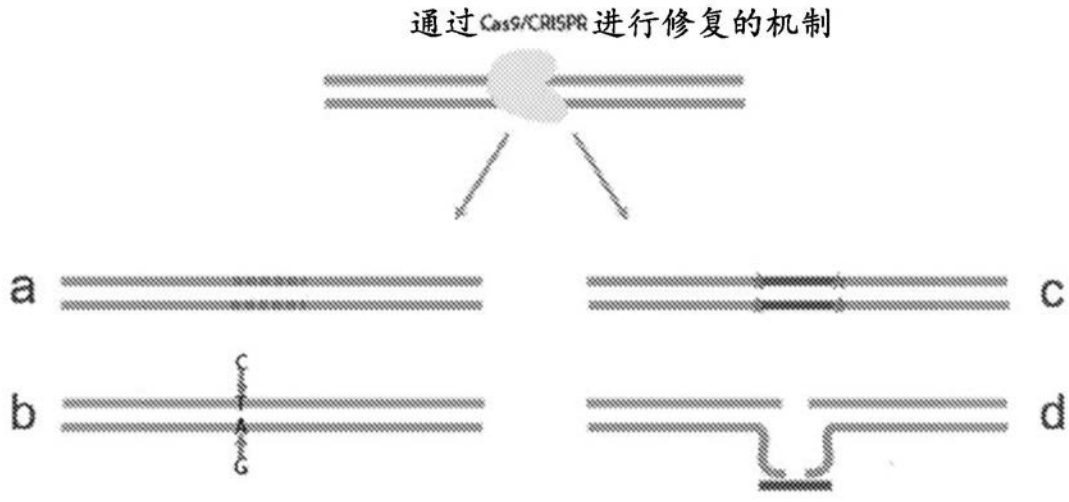


图1

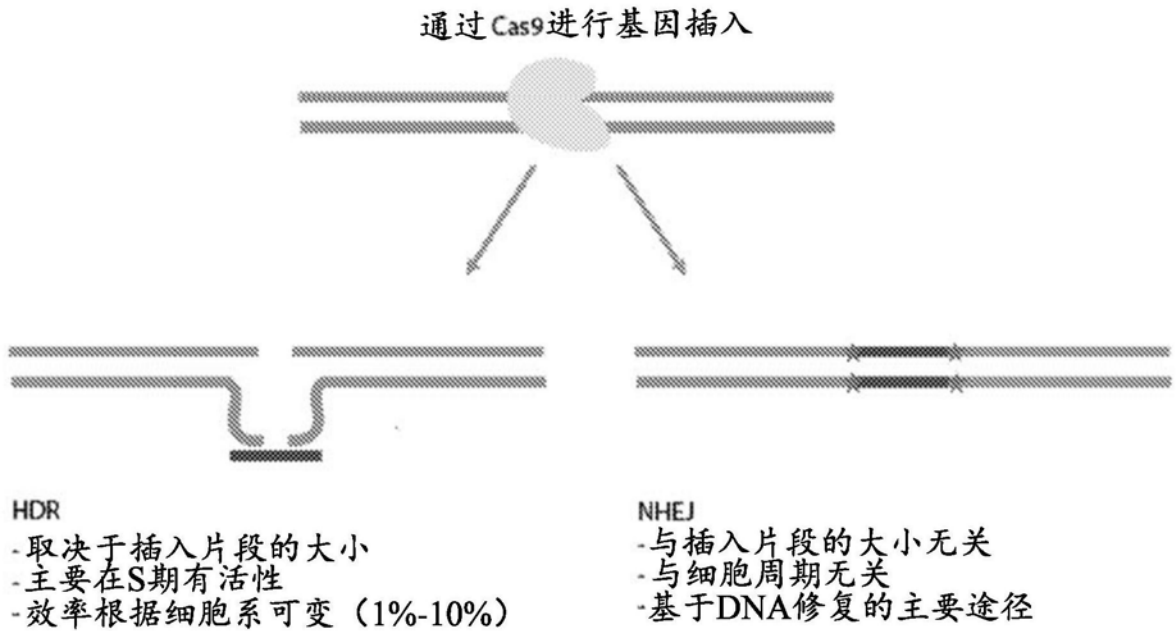


图2

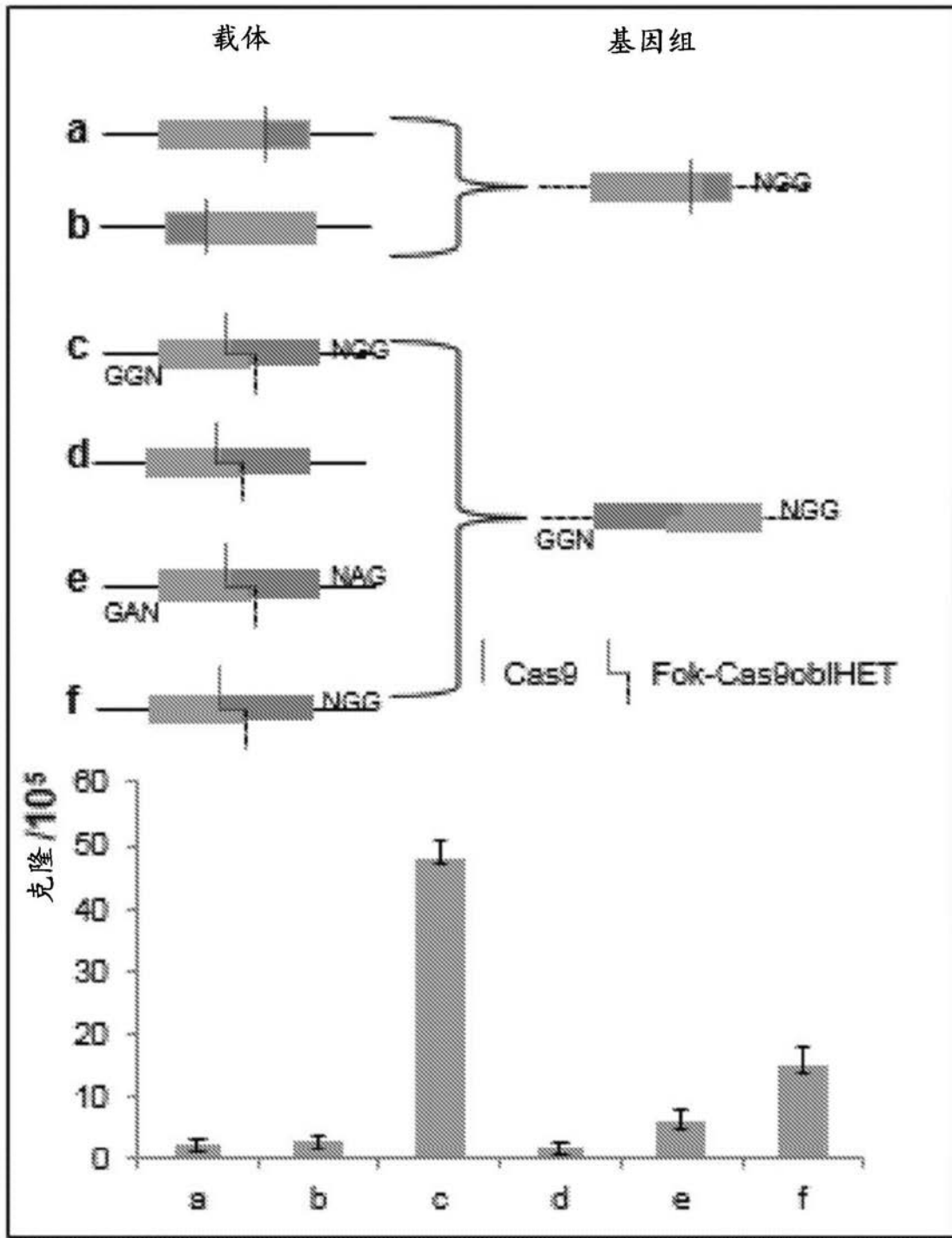


图3

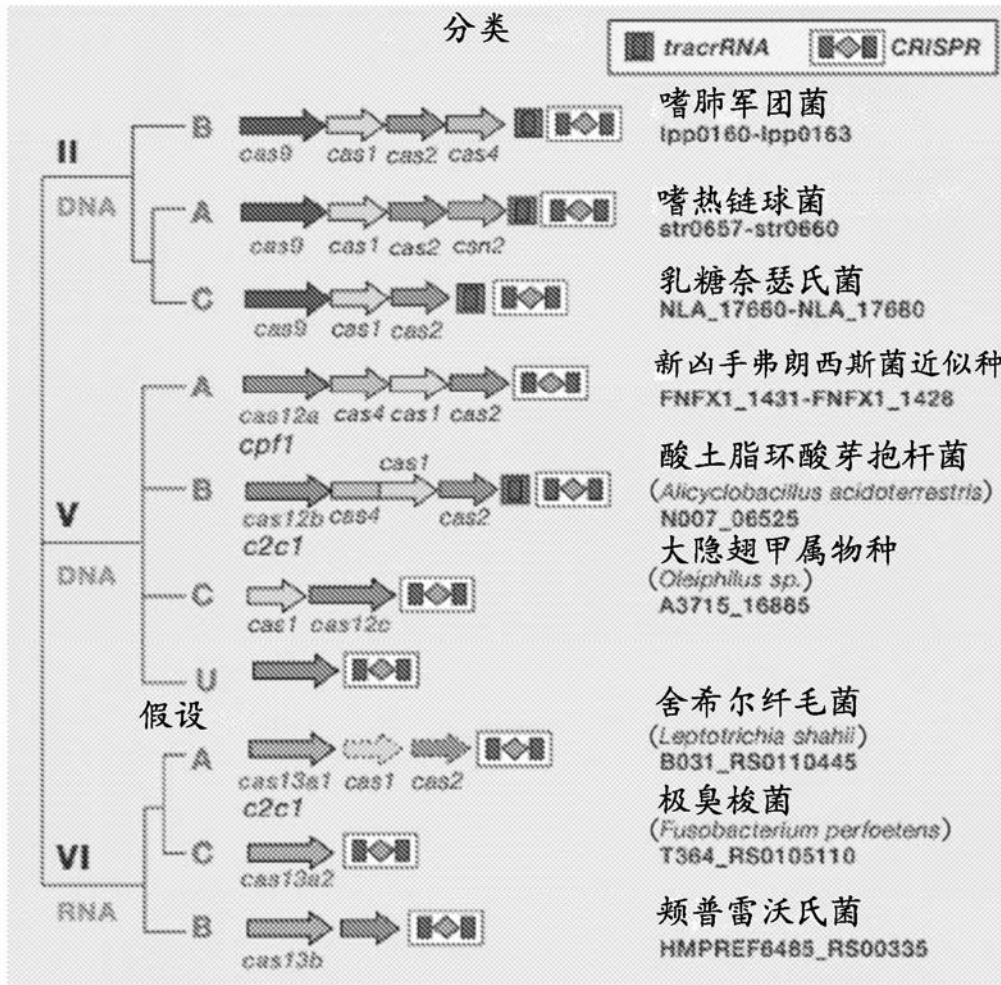


图4A

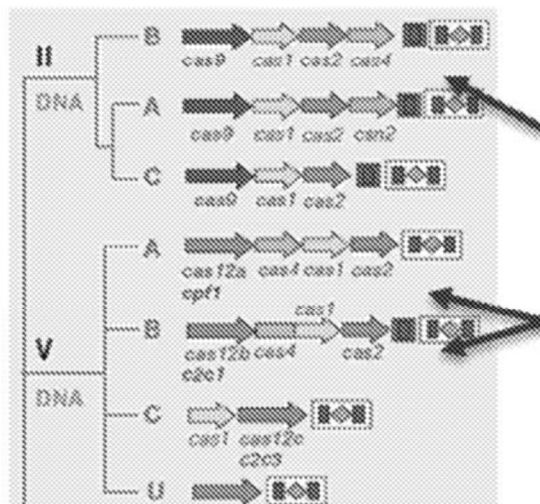
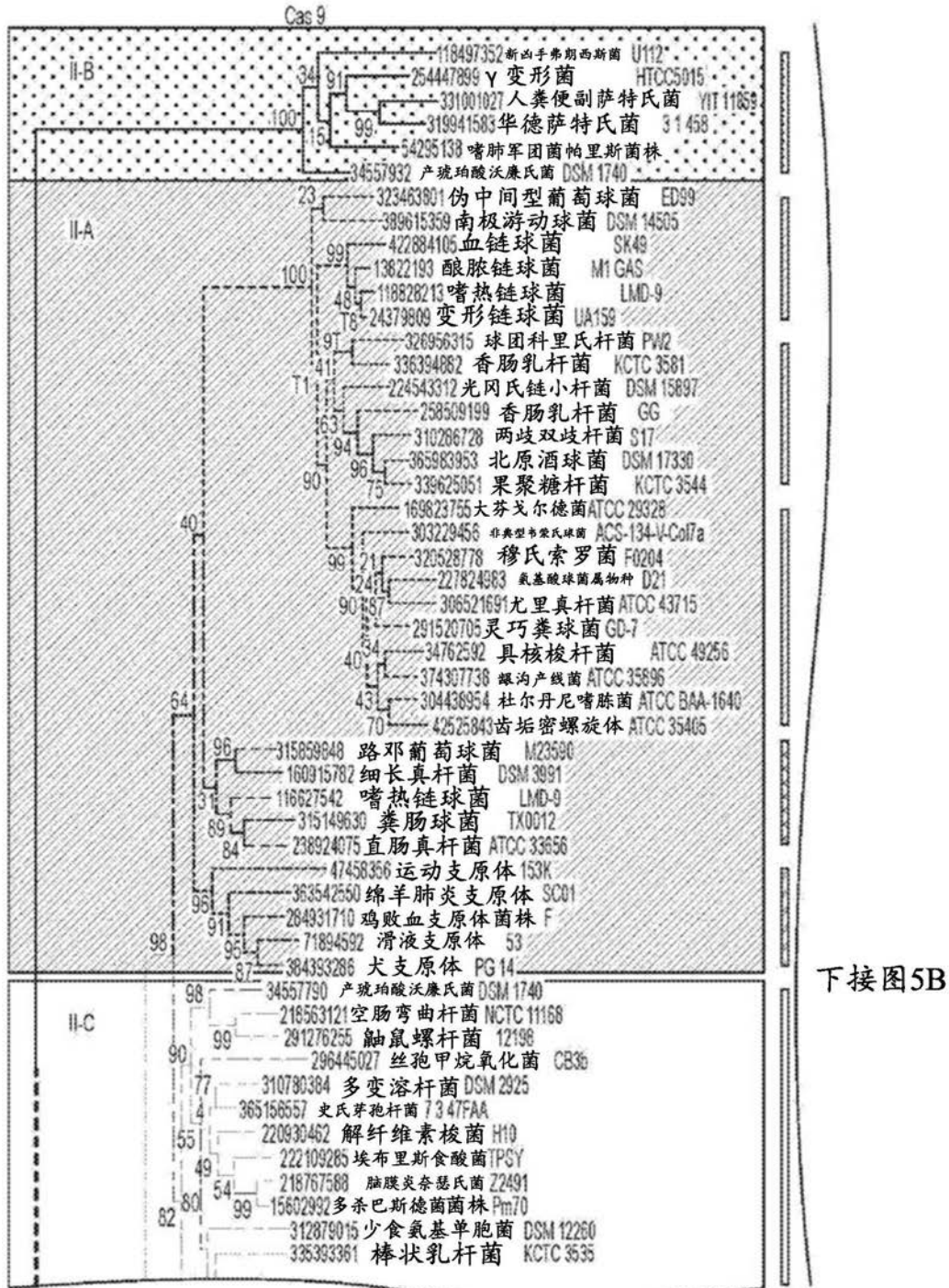


图4B



下接图5C

下接图5B

图 5A

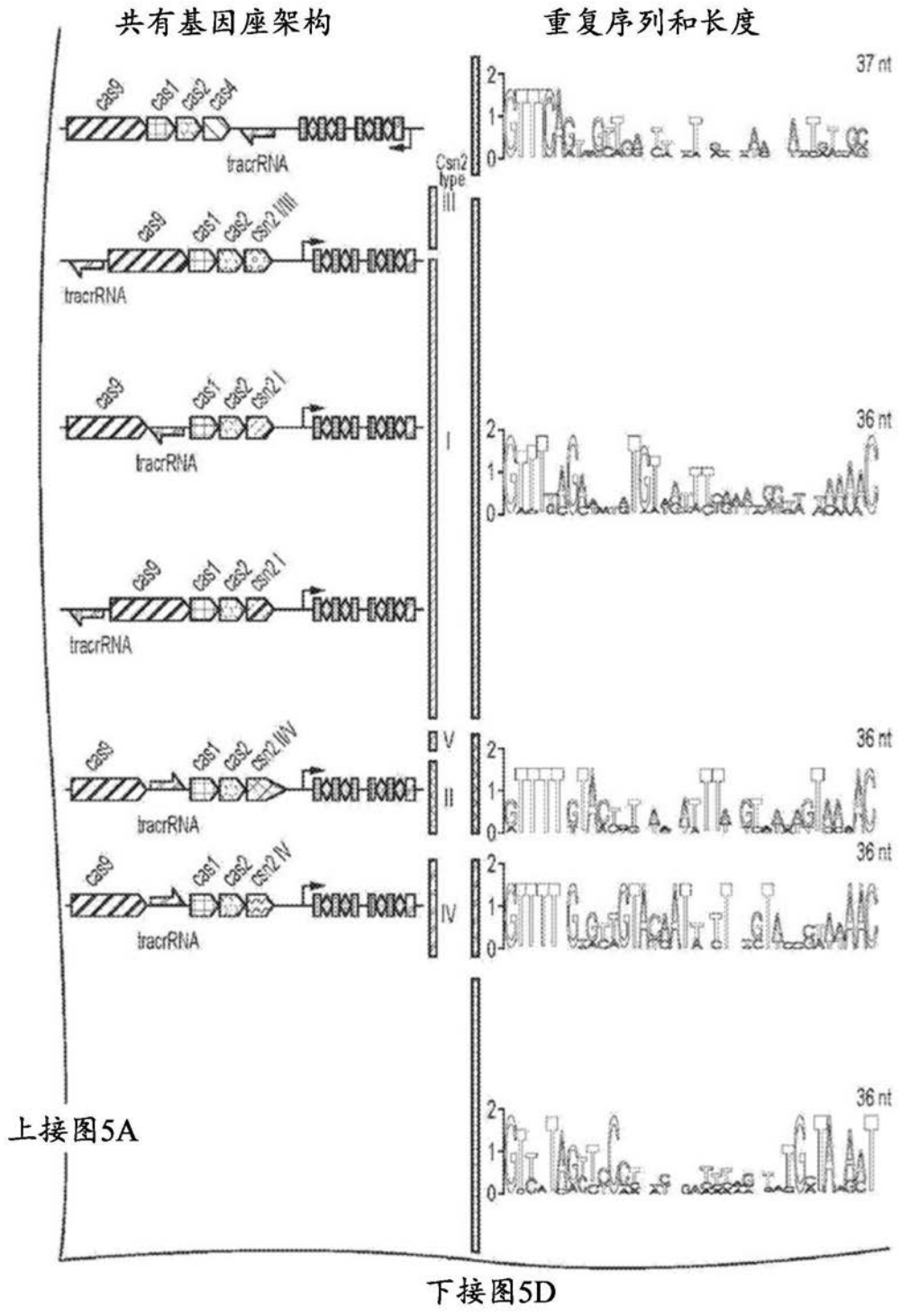


图 5B

上接图5A

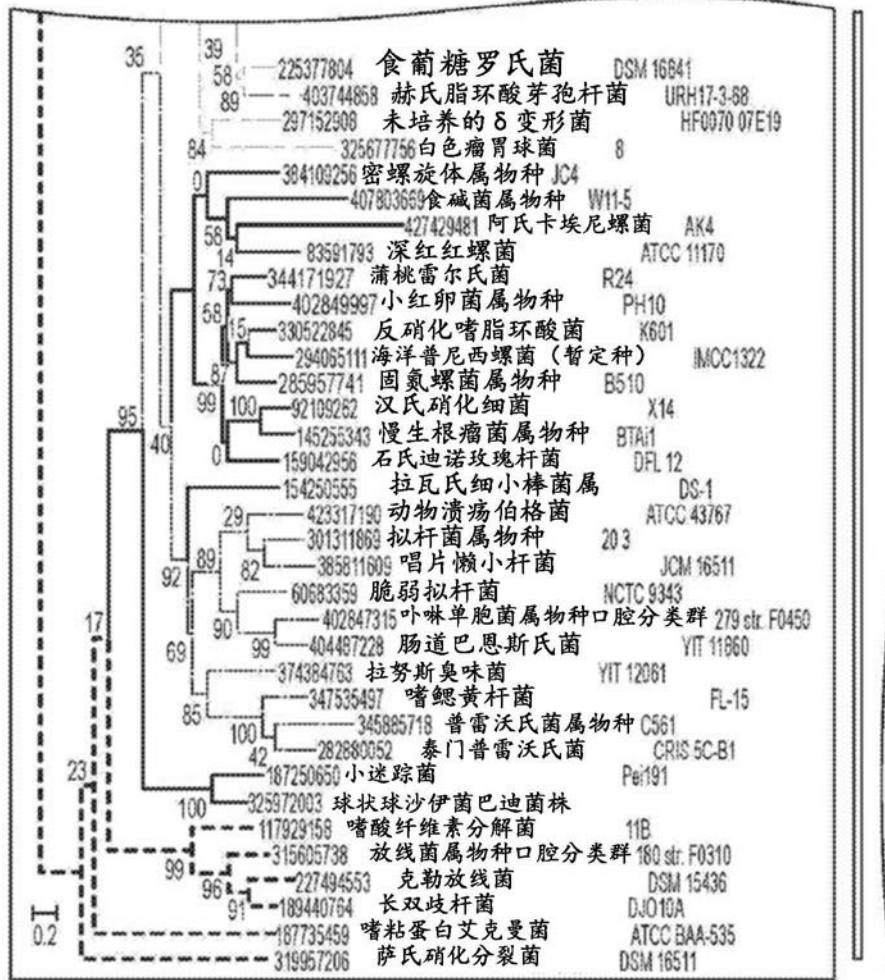


图 5C

上接图5B

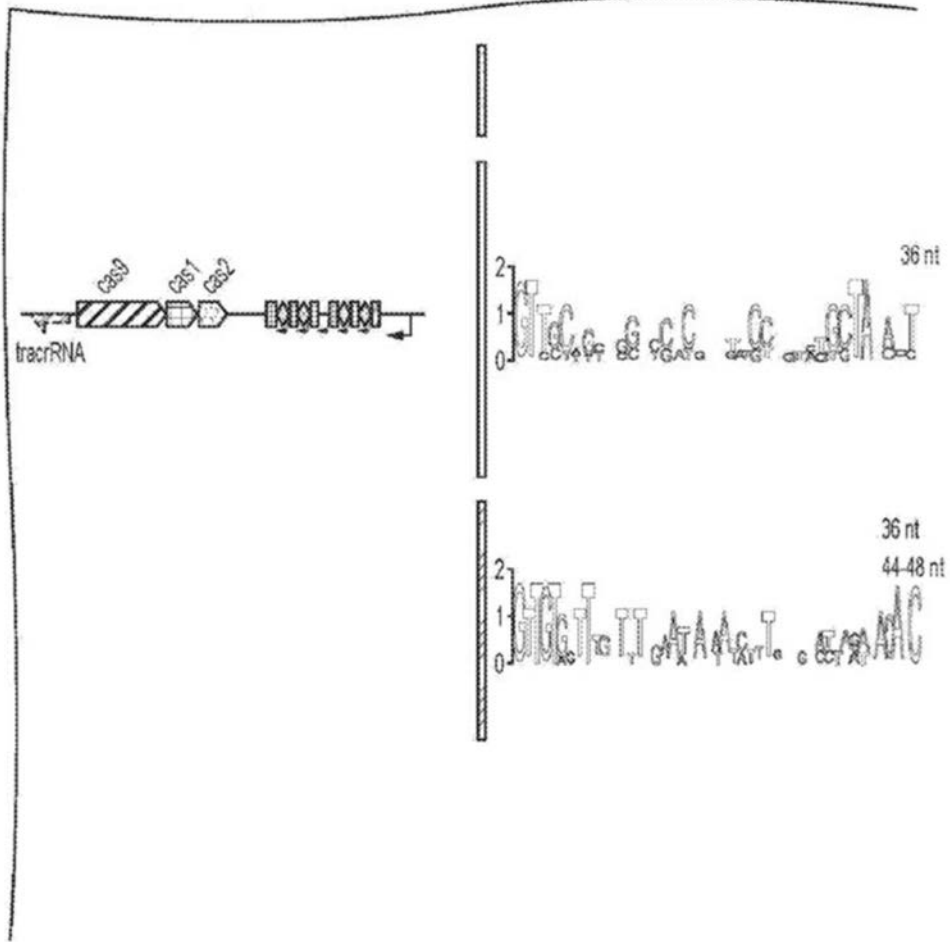


图 5D

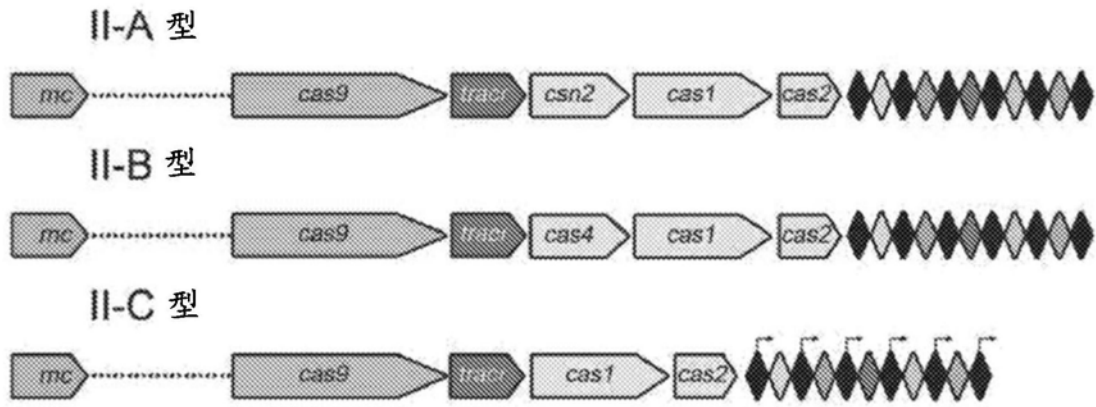


图 5E

图5

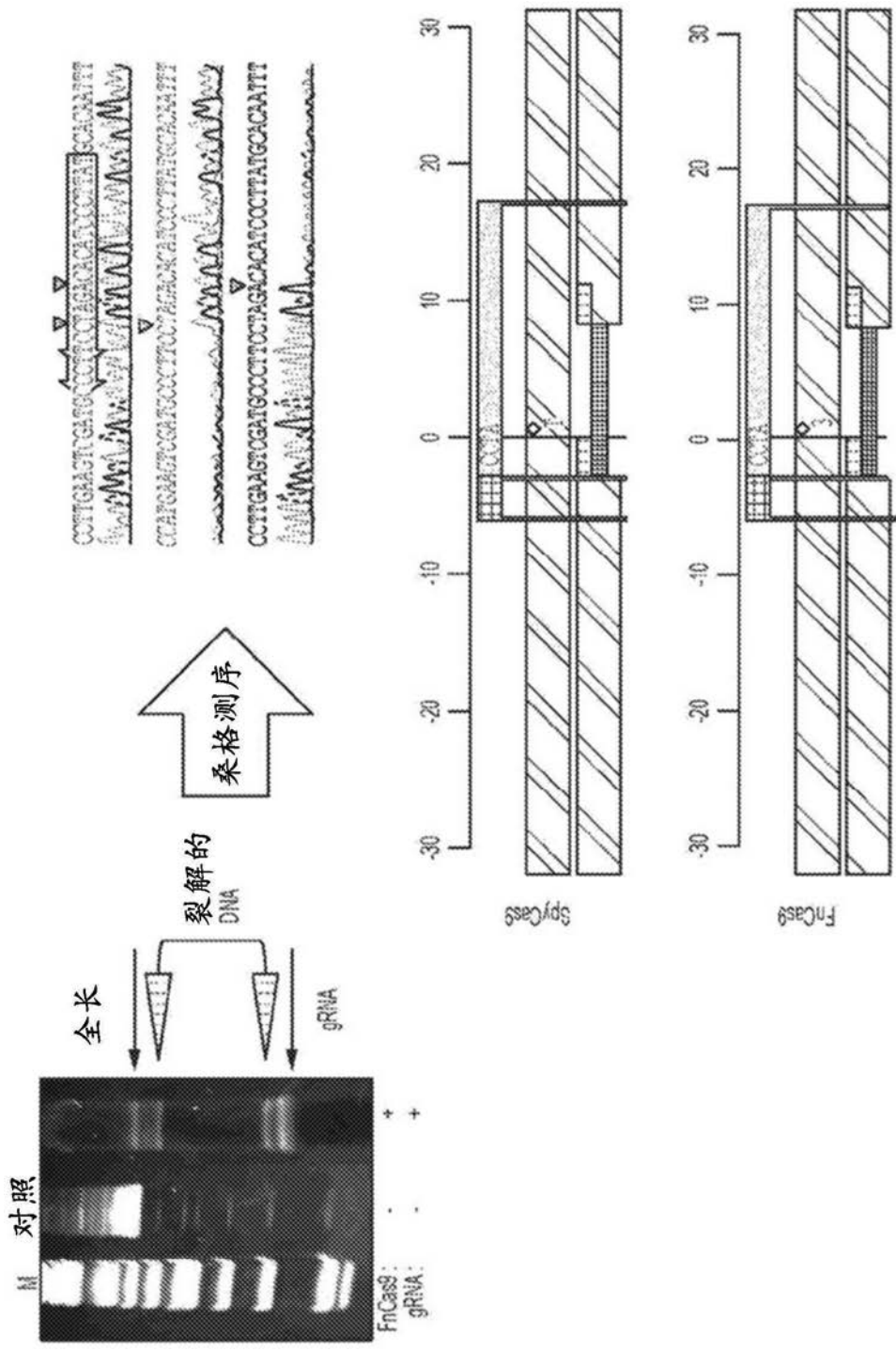


图 6A

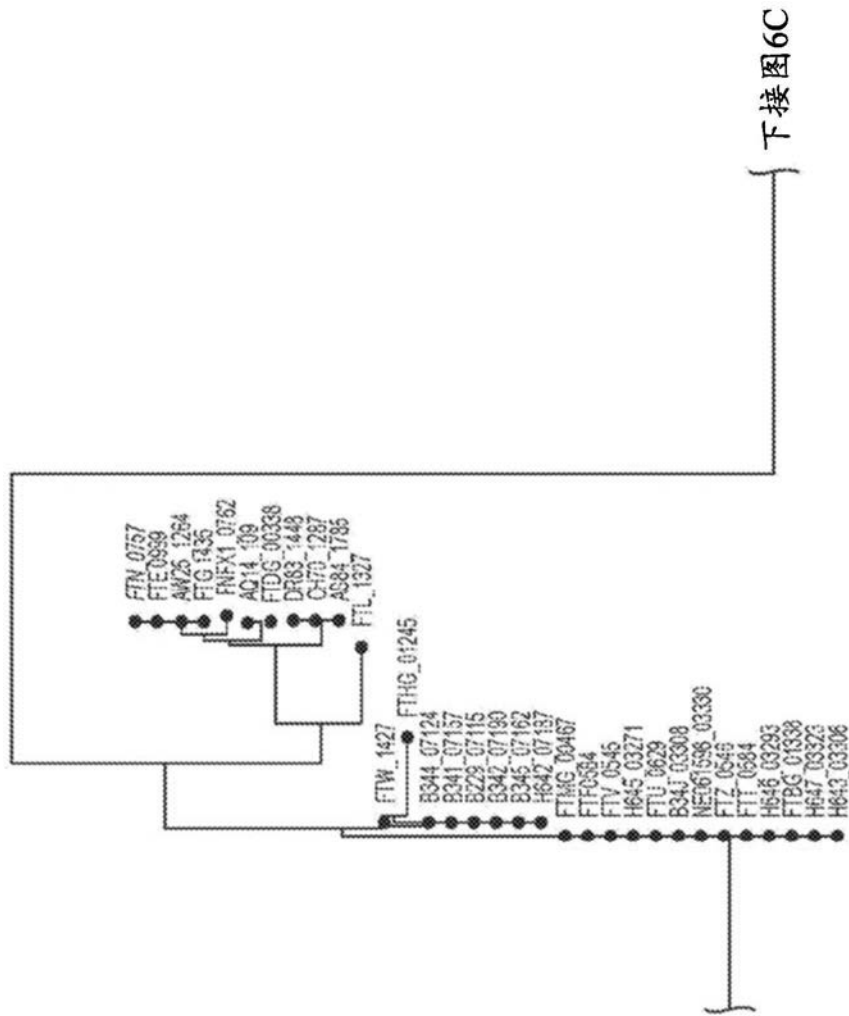


图 6B

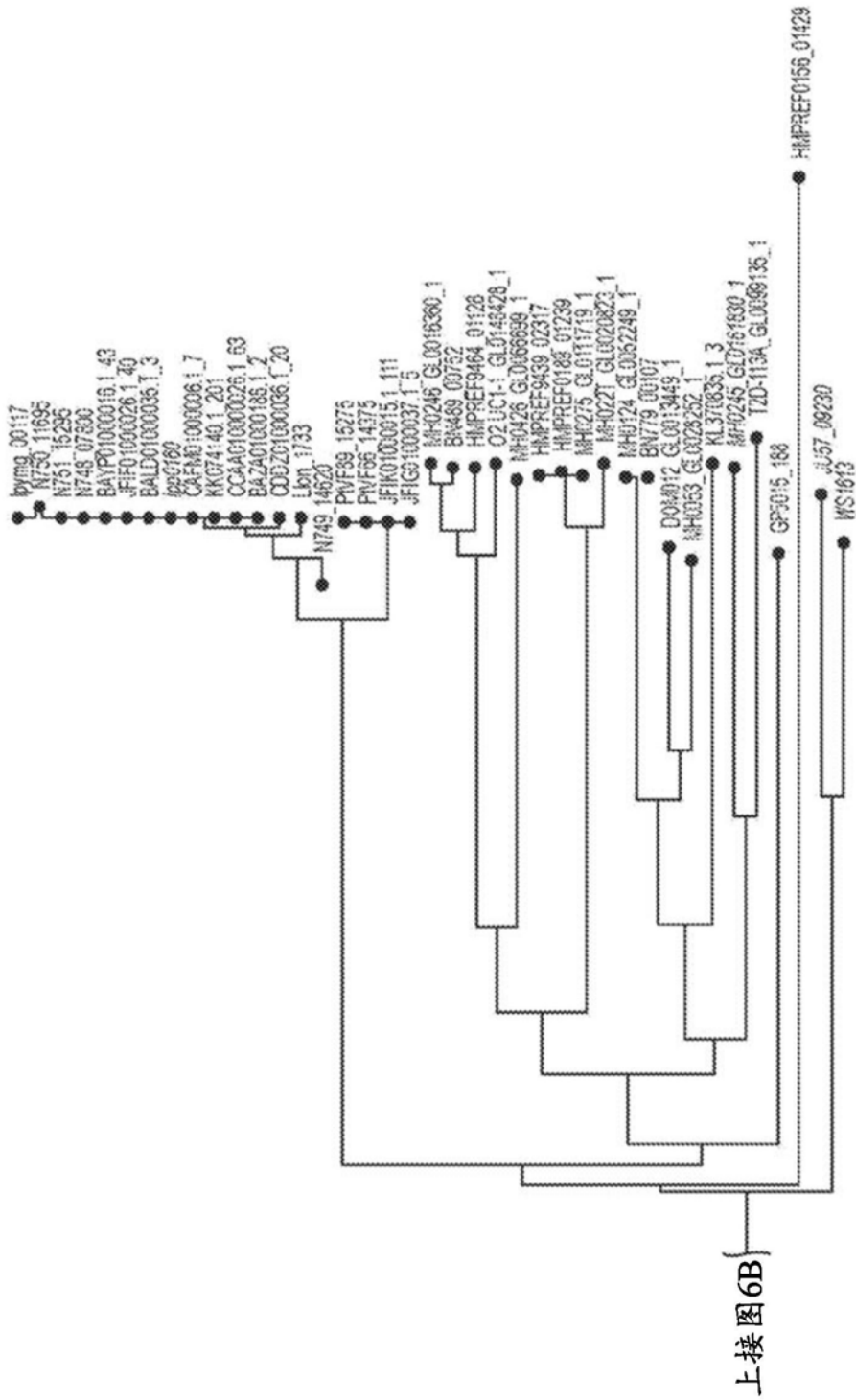


图 6C

图6

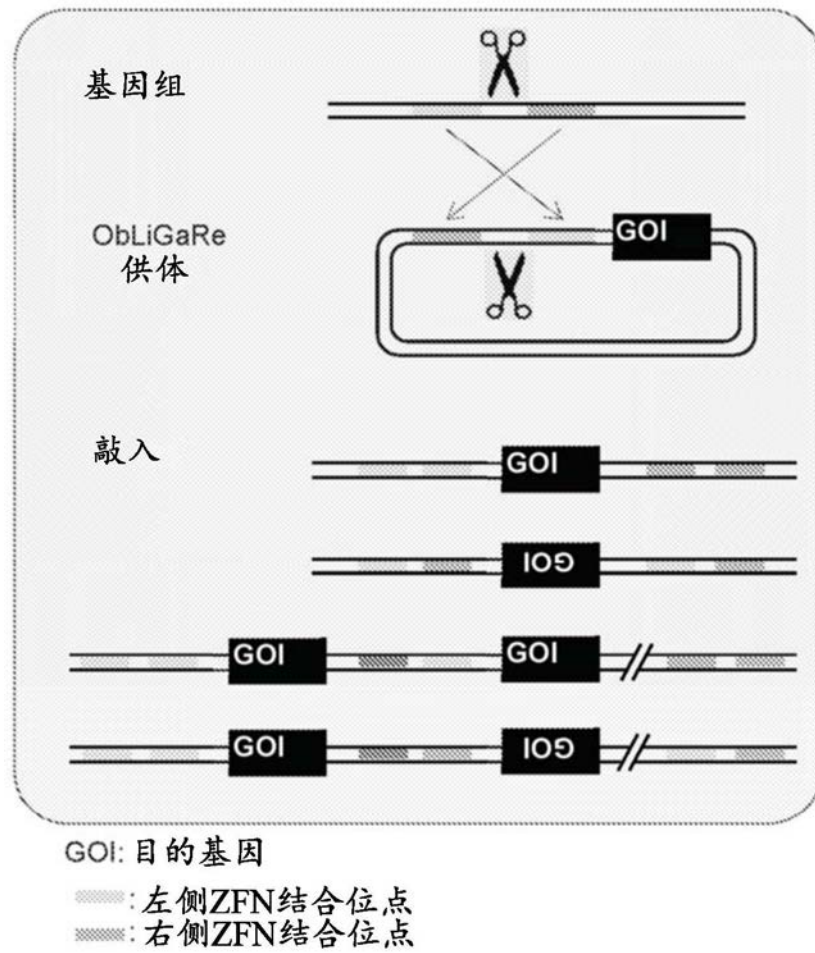


图7

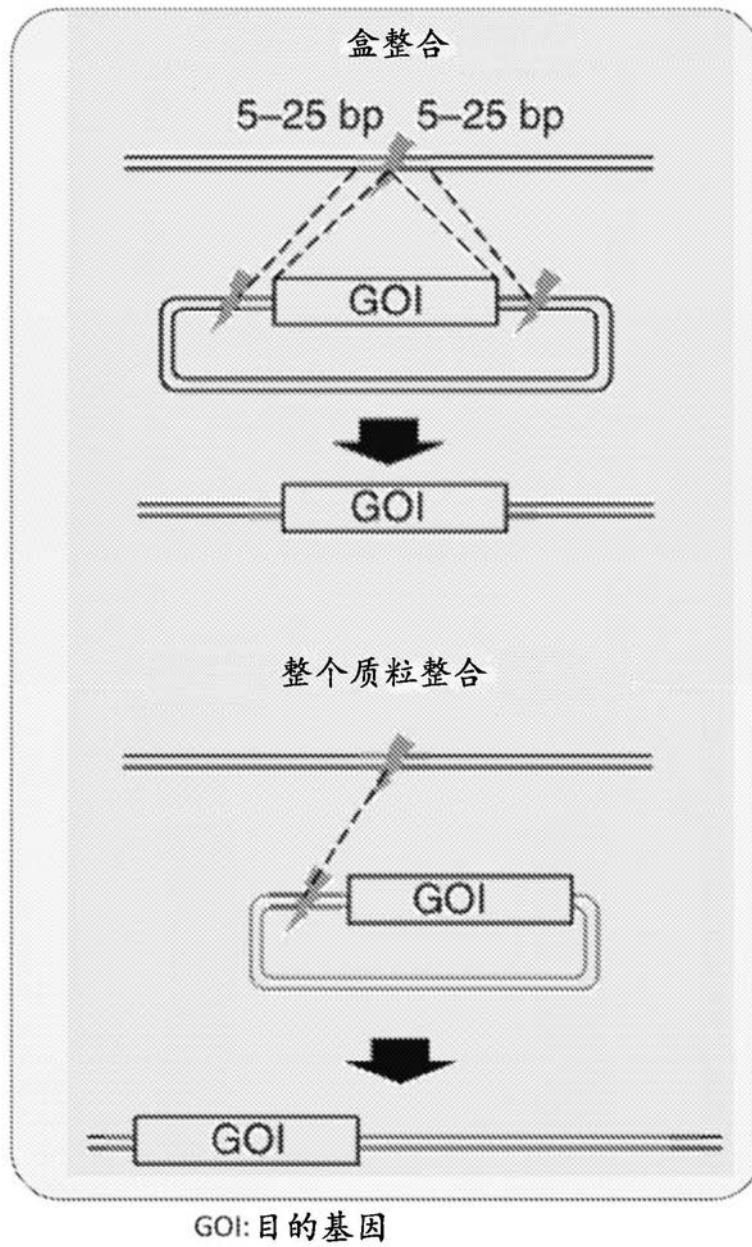


图8

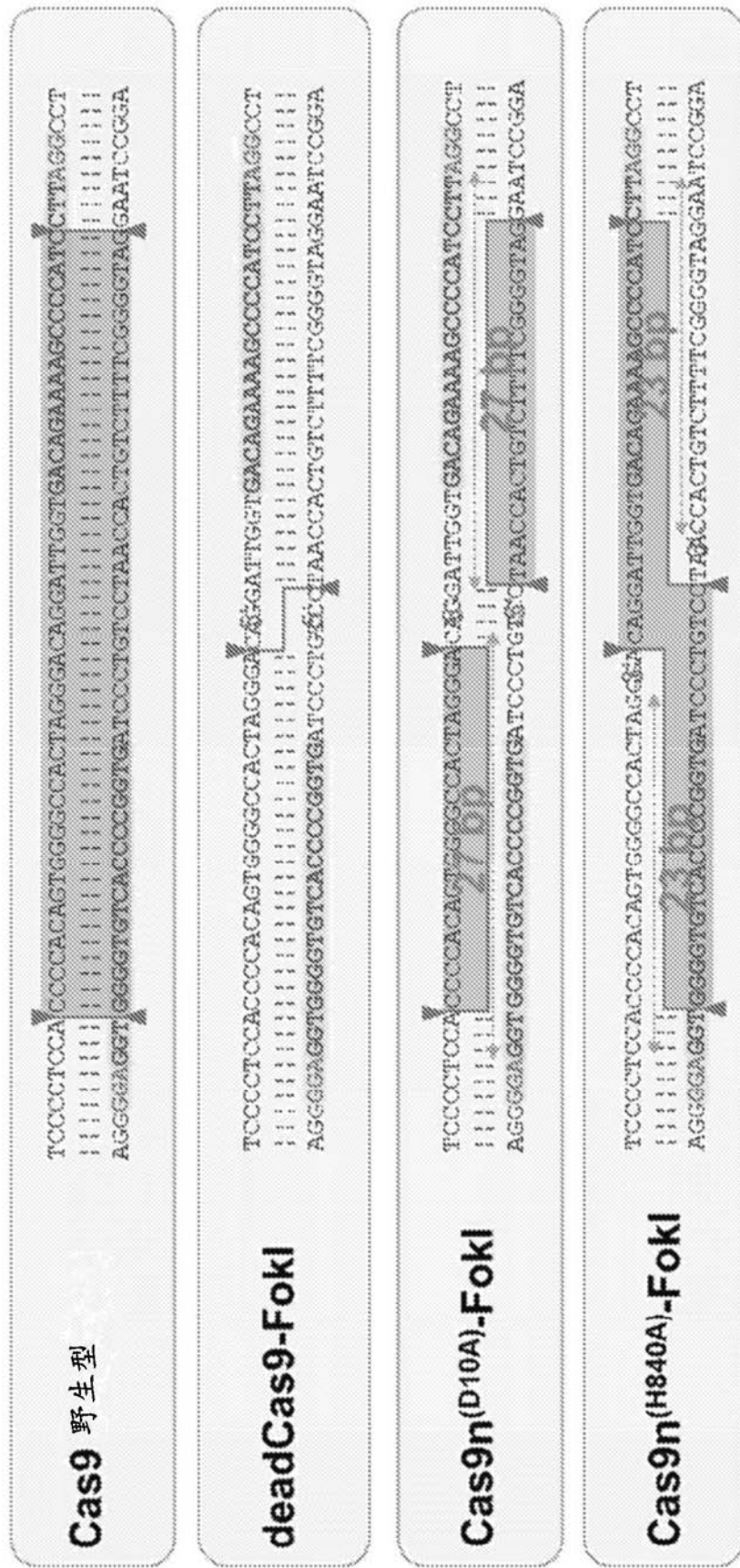


图10

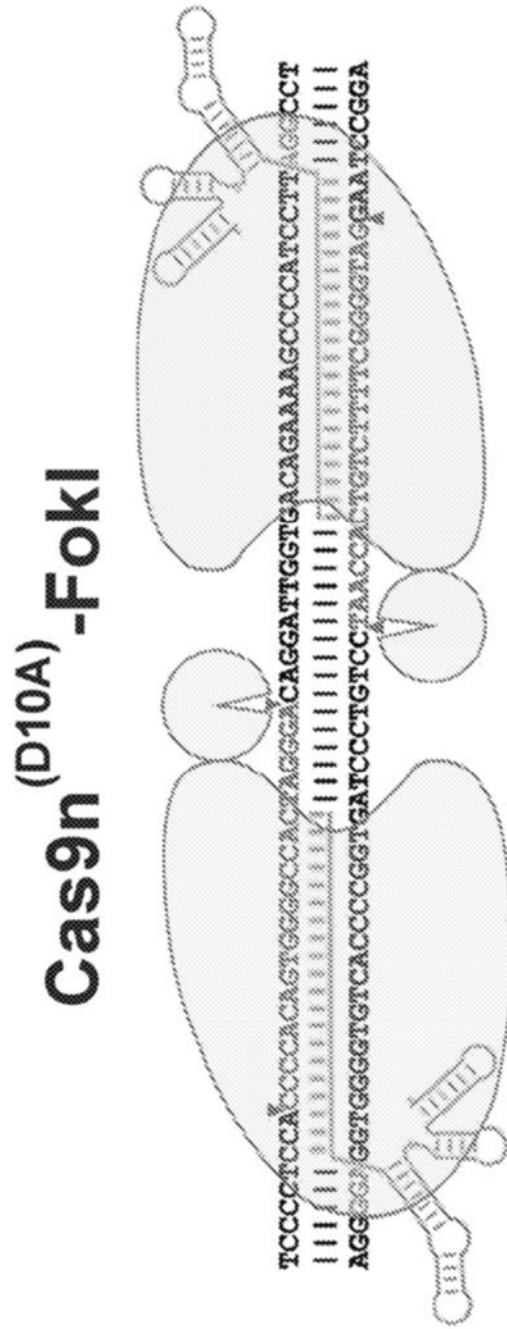


图11

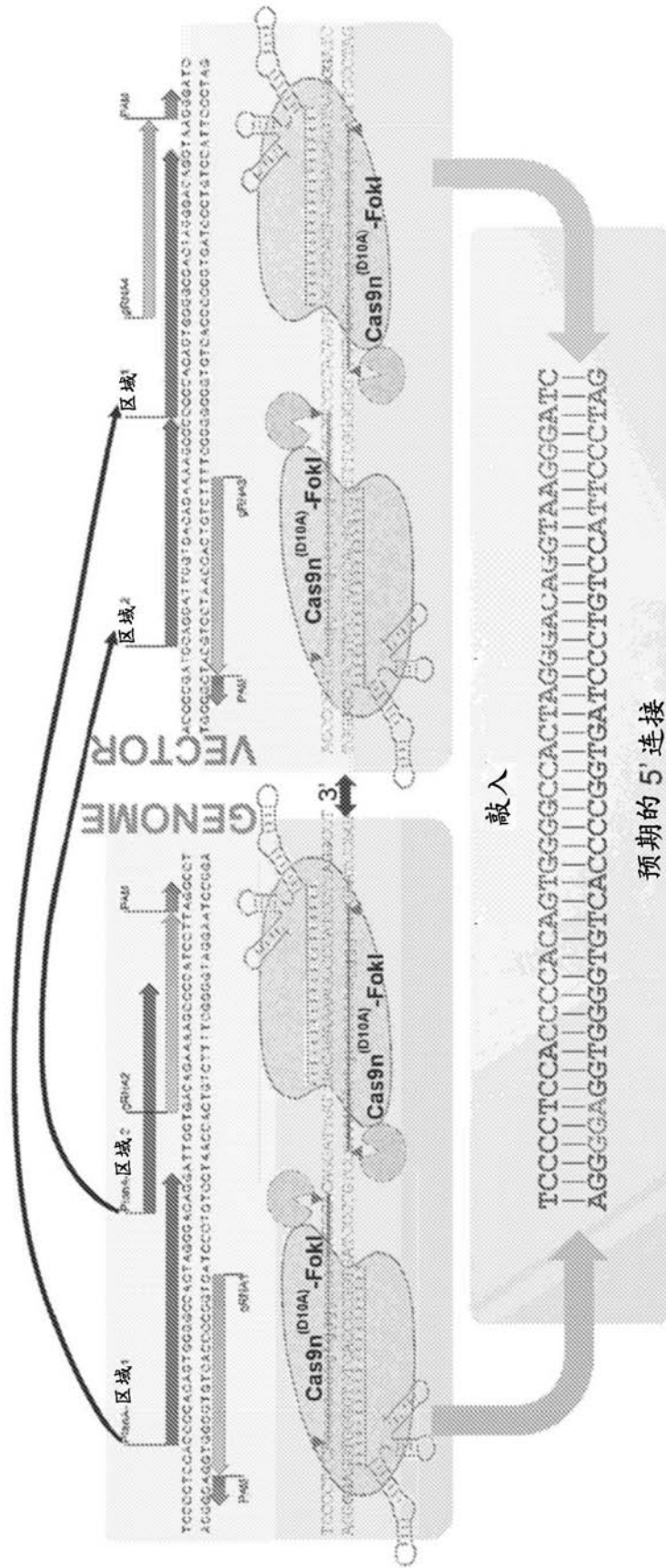


图12

Cas9n (H840A)-FokI

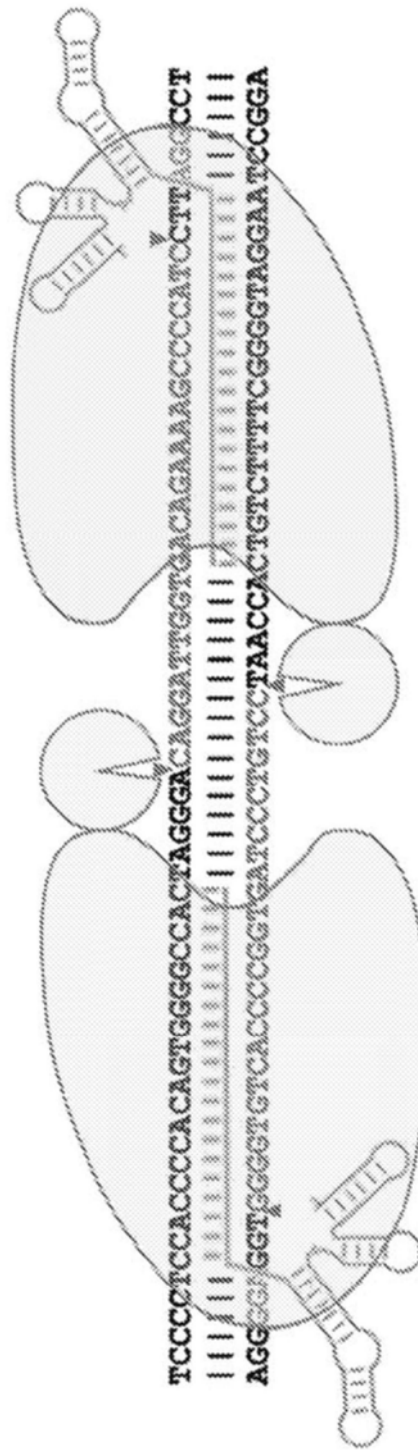


图13

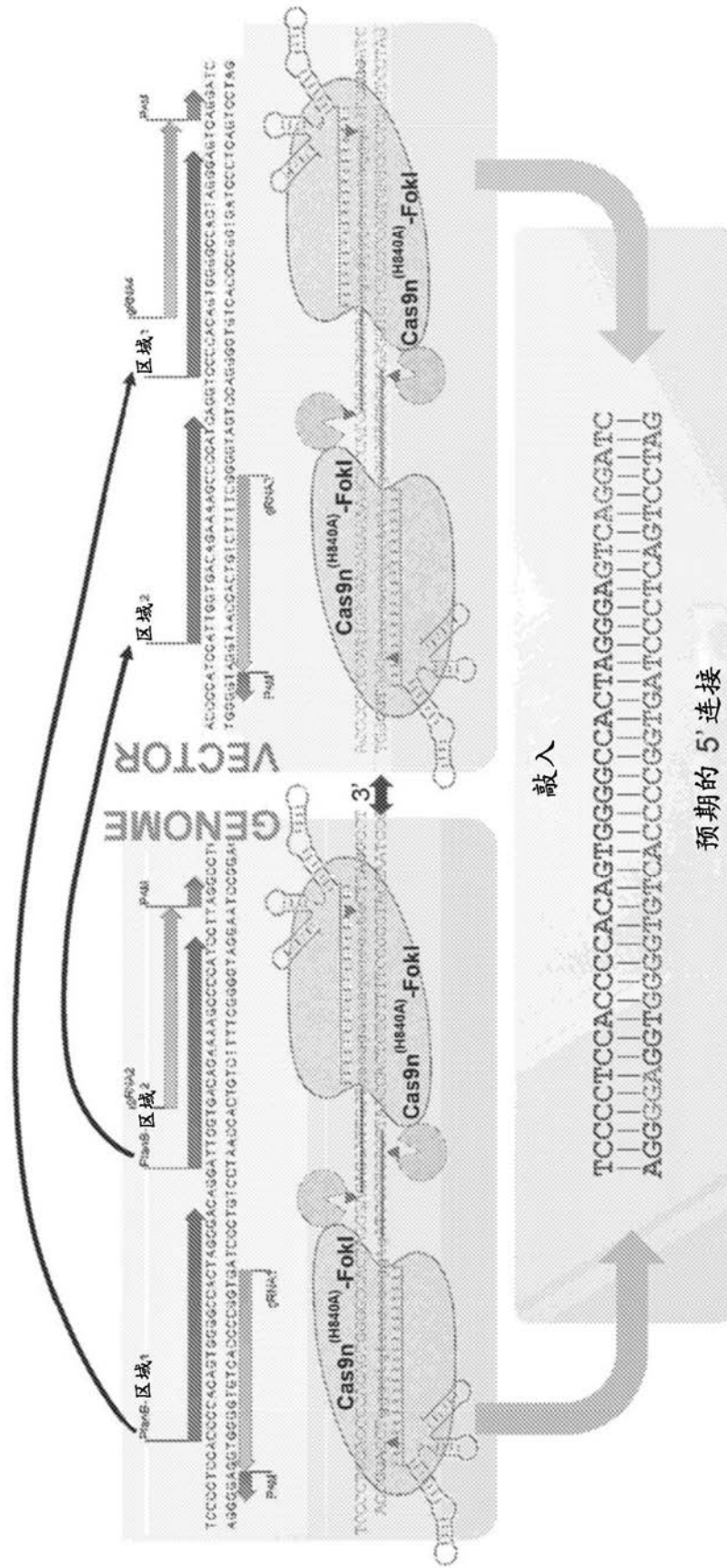


图14

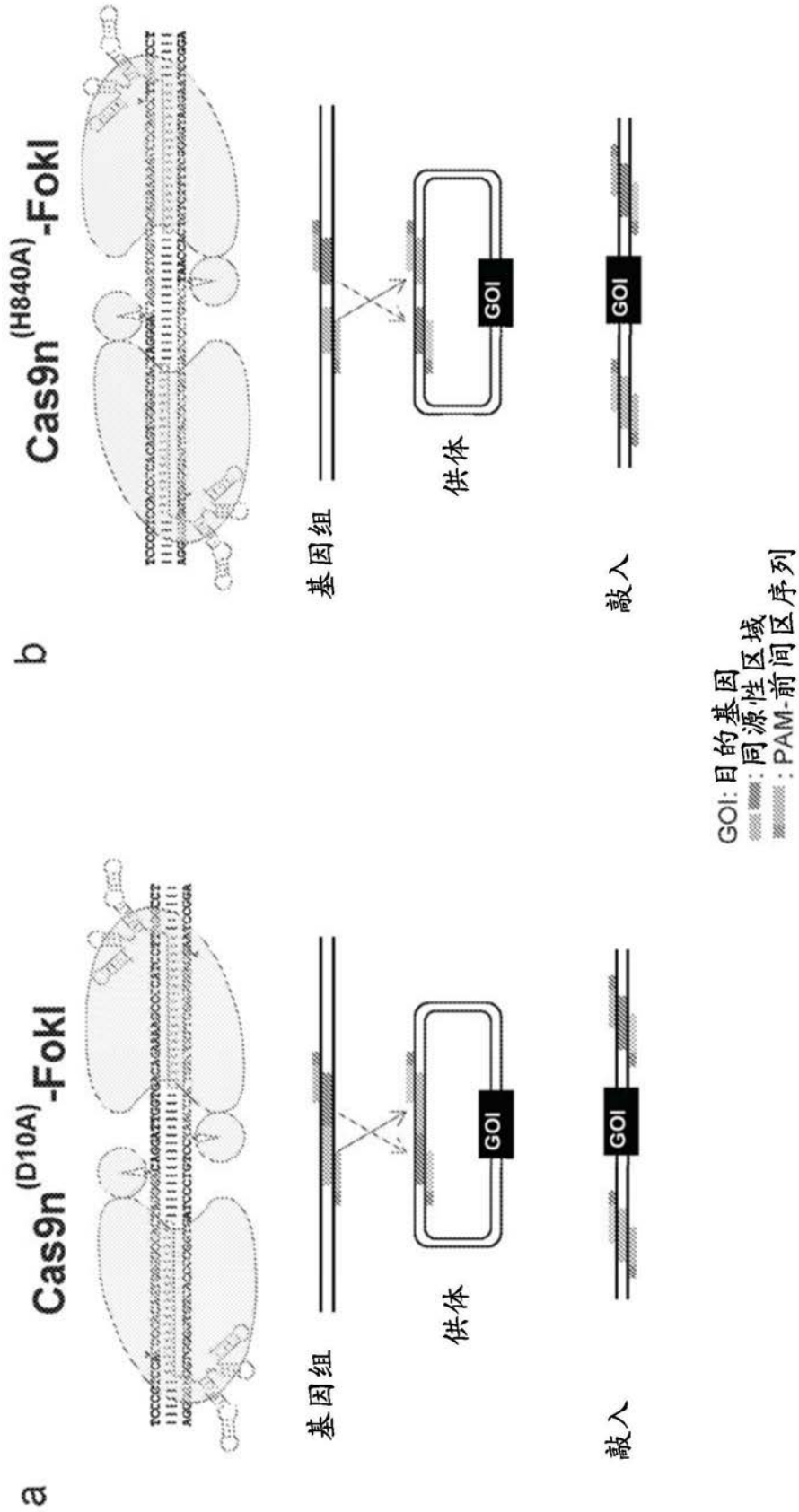


图15

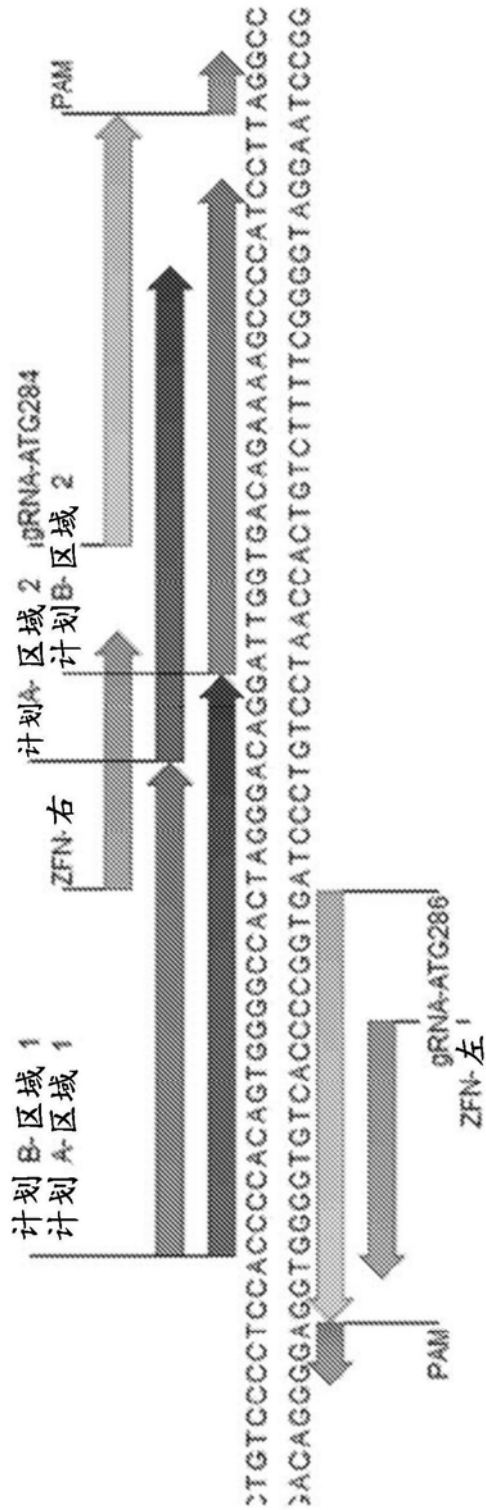
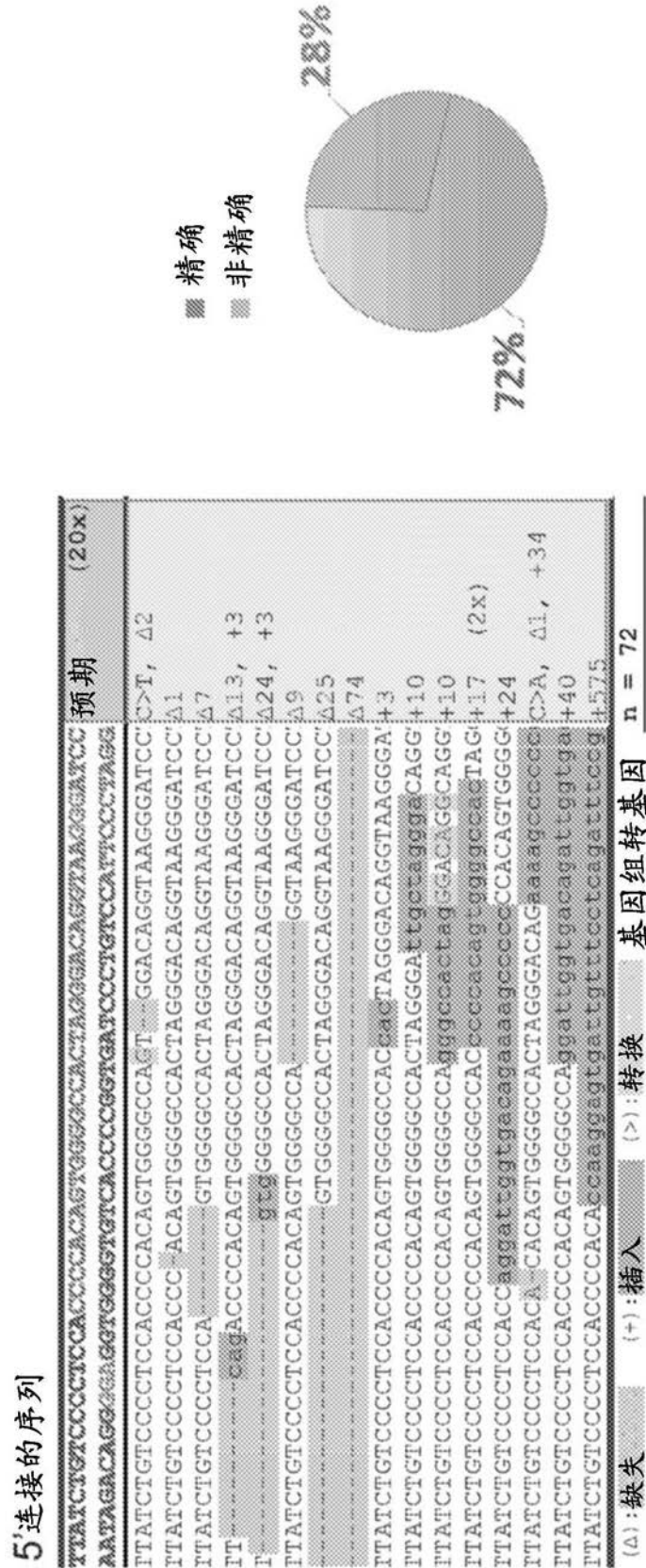


图16



■ 精确
 ■ 非精确

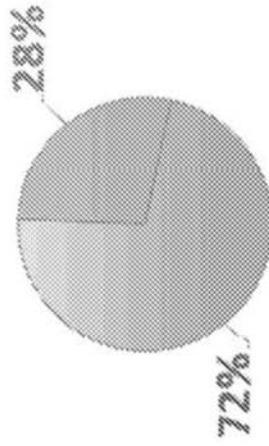


图17

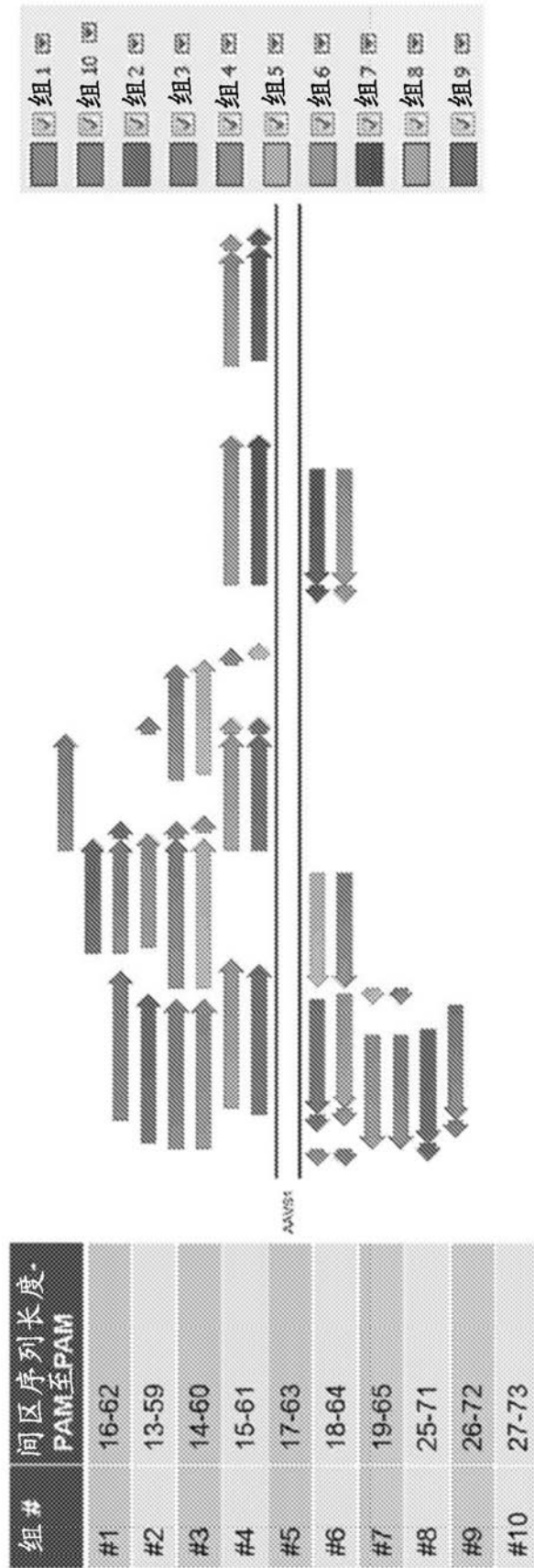


图19

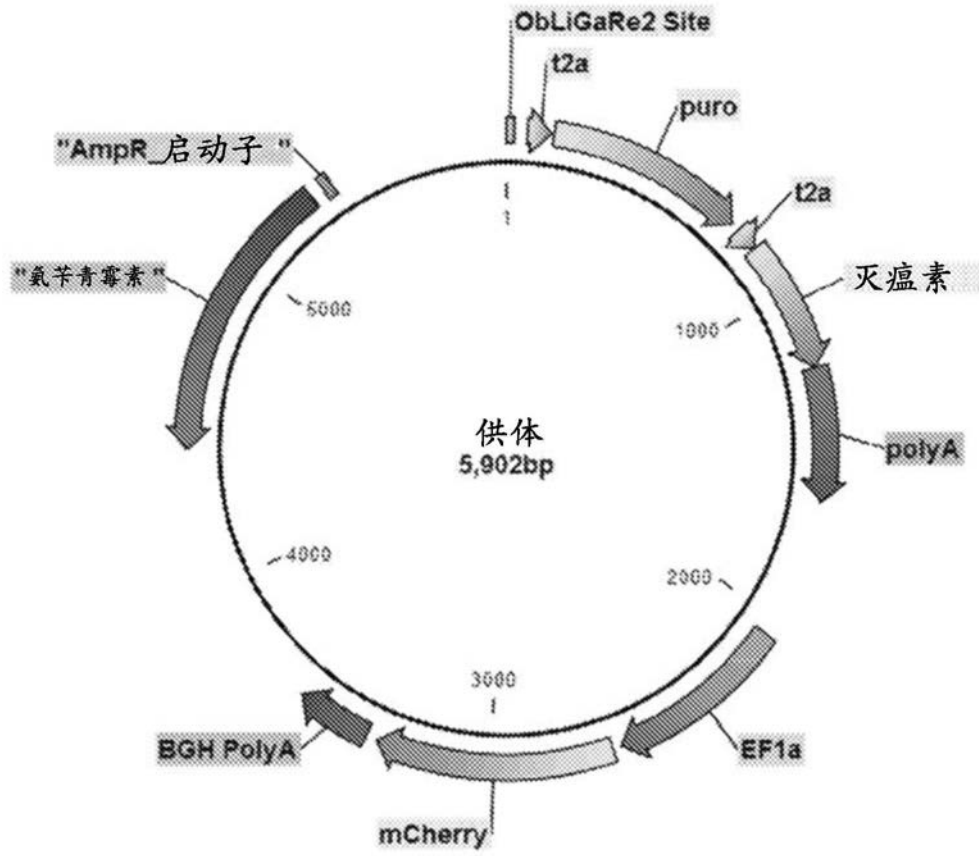


图20

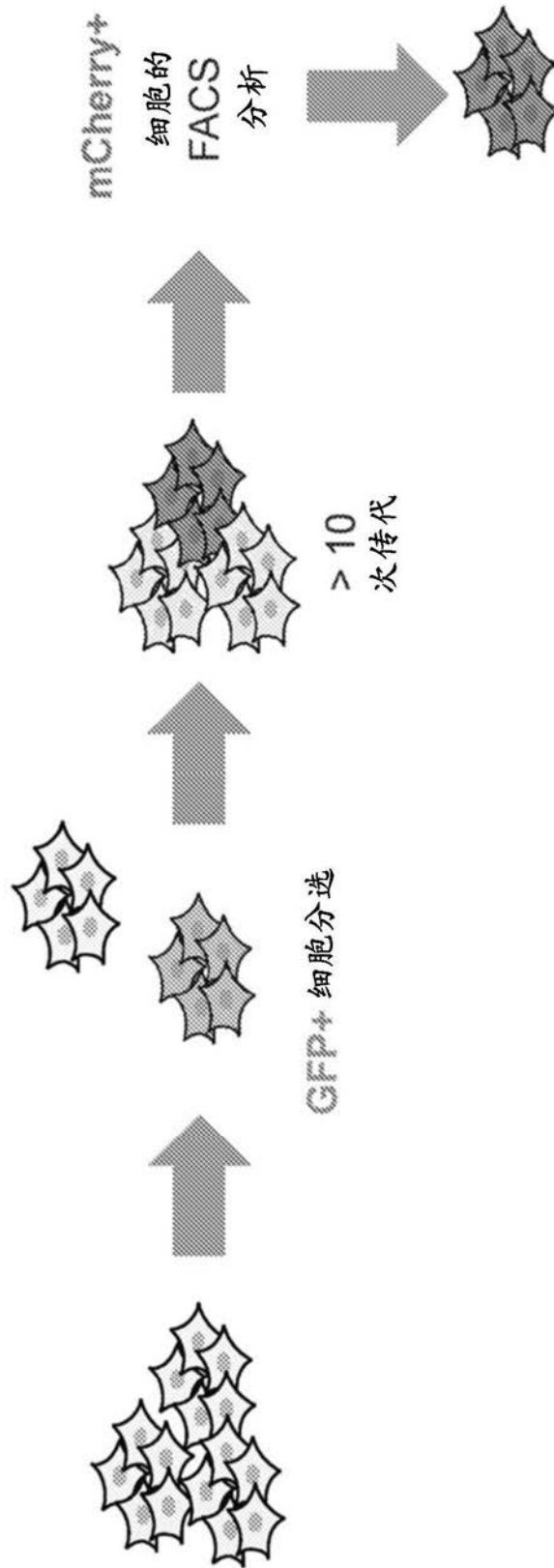


图21

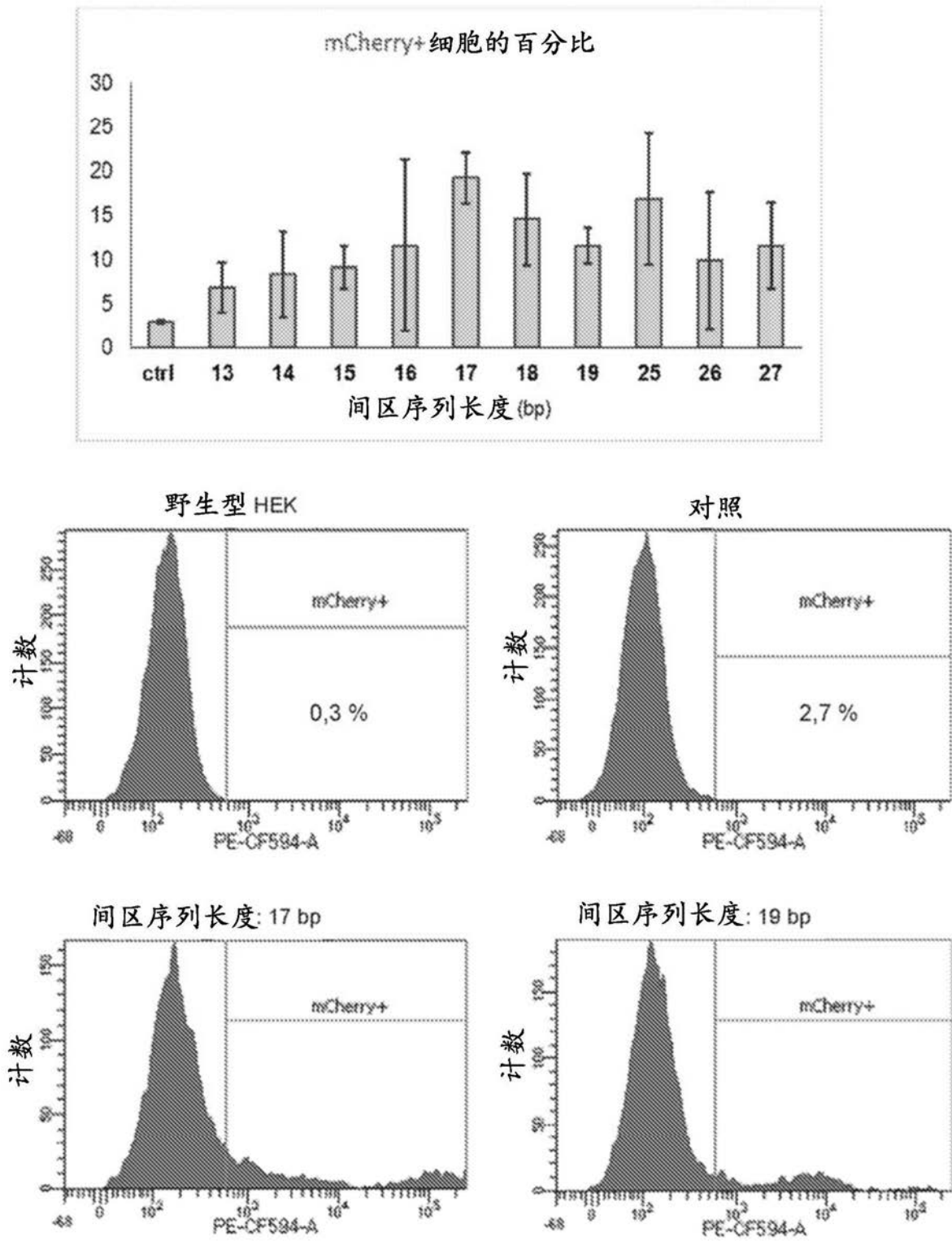


图22

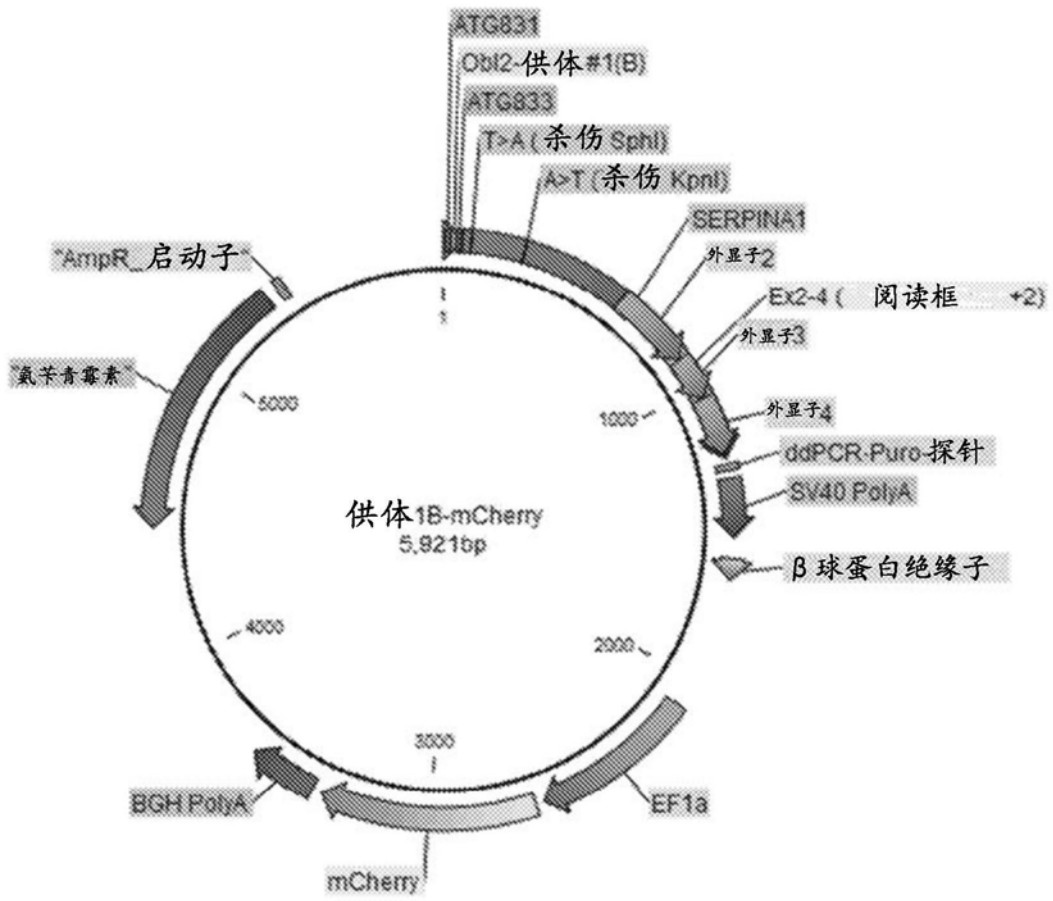


图23

deadCas9-FokI

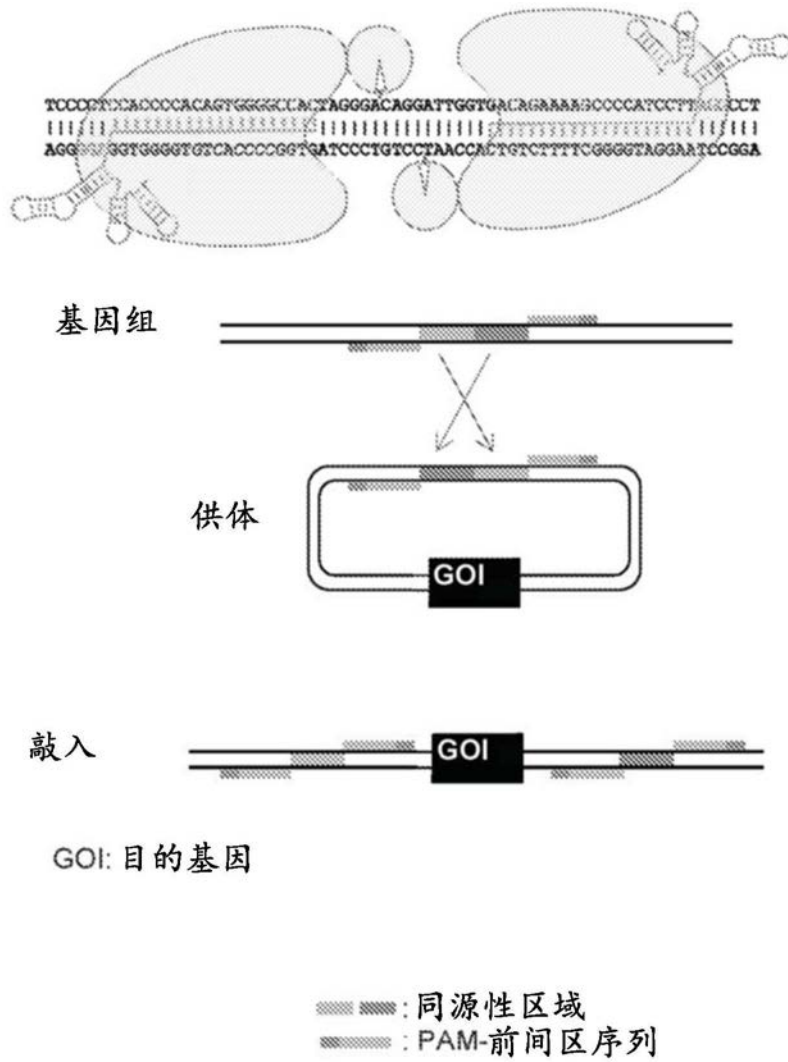


图24

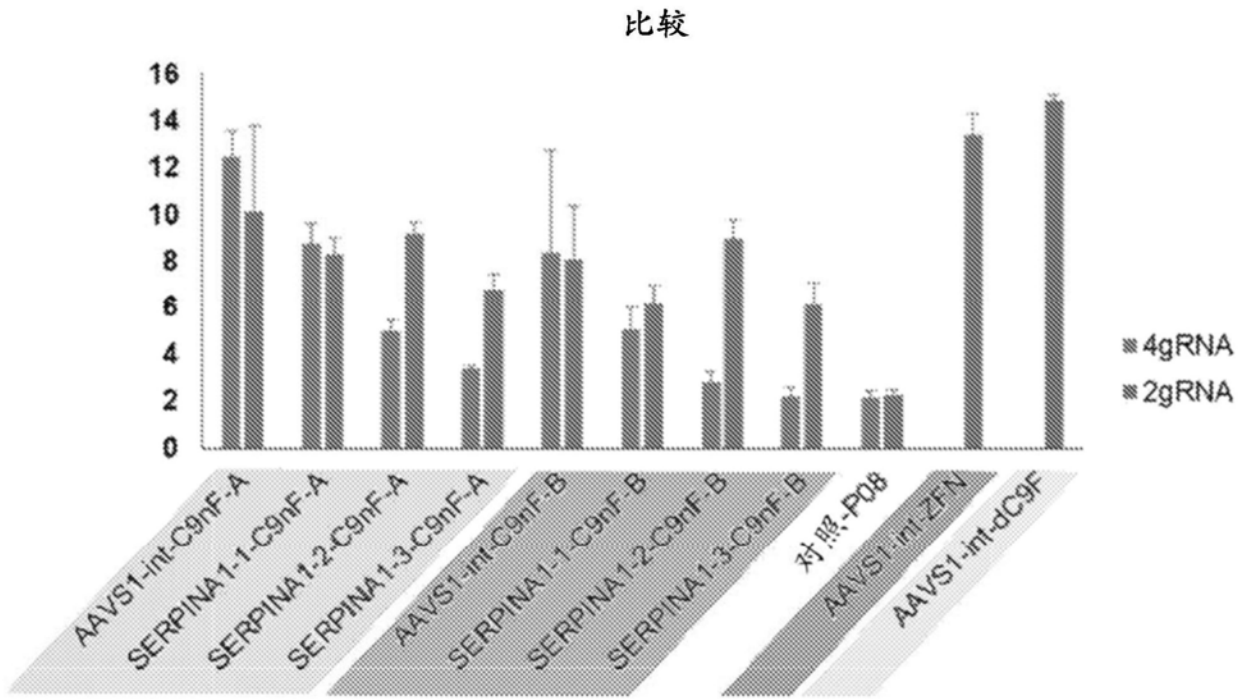


图25



图26

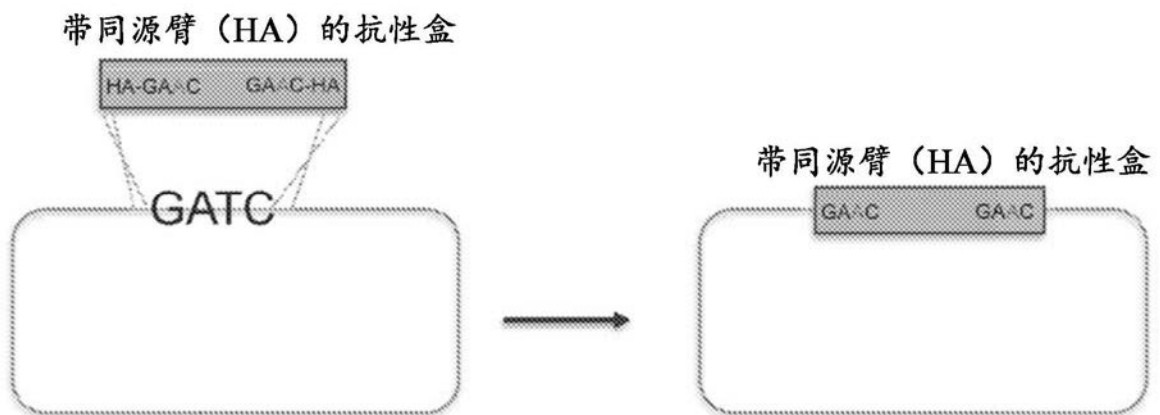


图27



图28

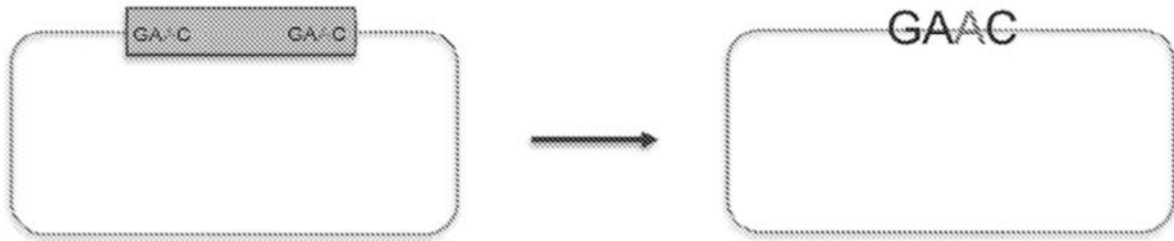


图29

(1). 华德萨特氏菌 3_1_45B Cas9 蛋白(SwCas9)的氨基酸序列 (SEQ ID NO: 10)。

MTQSERRFSCSIGIDMGAKYTGVFYALFDREELPTNLNSKAMTLVMPETGPRYVQAQRTAVRHRLRGQKRYTLARKL
 AFLVVDDMIKKQEKRLTDEEWKRGREALSGLLKRRGYSRPNADGEDLTPLENVRADVFAAHPAFSTYFSEVRS LAEQ
 WEEFTANISNVEKFLGDPNI PADKEFI EFAVAEGLIDKTEKKAYQSALSTLRANANVLTGLRQMGHKPRSEYFKAIE
 ADLKKSRLAKINEAFGGAERLARLLGNLSNLQLRAERWYFNAPDIMKDRGWEPDRFKKTLVRAFKFFHHPAKDQNKQ
 HLELIKQIENSEDIIETLCTLDPNRTI PPYEDQNNRRPPLDQTL LLSPEKLT RQYGEIWKTSARLTSAEPTLAPAA
 EILERSTDRKSRVAVNGHEPLPTLAYQLSYALQRAFDRSKALDPYALRALAAGSKSNKLT SARTALENCIGGQNVKT
 FLDCARRYREADDAKVGWFDNADGLLERSDLHPPMKKKILPLLVANILQTDETTGGQKFLDEIWRKQIKGRETVAS
 RCARIETVRKSFSGGFNIAYN TAQYREVNKLPRNAQDKELLTIRDRVAETADFI AANLGLSDEQKRKFANPFSLAQF
 YTLIETEVS GFSATT LAVHLENAWRMTIKDAVINGETVRAAQCSRLPAETARPF DGLVRRRLVDRQAW EIAKRVSTDI
 QSKVDFSN GIVDVSIFVEENKFEFSASVADLKKNRVVKDKMLSEAEKLETRWLIKNERIKKASRGTC PYTGDRLAEG
 GEIDHILPRSLIKDARGIVFNAEPNLIYASSRGNQLKKNQRYSLSDLKANYRNEIFKTSNIAAITAEIEDVVTKLQ
 THRLKFFDLLNEHEQDCVRHALFLDDGSEARDAVLELLATQRTRVNGTQI WMIKNLANKI REELQNWCKTTNNRHL
 FQAAATNVSDAKNLRKLQAQNPDFEKPDIQPIASHSIDALCSFAVGSADAERDQNGFDYLDGKTVLGLYPQSEVI
 HLQAKPQEEKSHFDSVAIFKEGIYAEQFLPIFTLNEKIWI GYETLNAKGERCGAIEVSGKQPKELLEMLAPFFNKPV
 GDLSAHATYRILKKPAYEFLAKAALQPLSAEEKRLAALLDALRYCTSRSKLSMSLFMAANGKSLKKREDVLKPKLFQL
 KVELKGEKSFKLNGLSLTPVKQDWLRICDSPELADAFGKPCSAD ELT SKLARIWKRPMVRDLAHAPVRREFSLPAID
 NPSGGFRIRRTNLFGNELYQVHAINAKKYRGFASAGSNVDWSKGI LFNELQHENLTECGGRFIT SADVTMSEWRKV
 VAEDNLSIWIAPGTEGRRYVRVETTFIQASHWFEQSVENWAITSPLSLPASFKVDKPAEFQKAVGTELSSELLGQPRS
 EIFIENVGNAKHIRFWYIVVSSNKKMNESYNNVSKS

(2). 土拉弗朗西斯菌土拉亚种 WY96-3418 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 11)

MNVKILPIAIDL DVKNTGVFSAFYQKGTSL EKLDNKNKGVYELSKDSYTL LMNNRTARRHQRRGIDRKQLVKRFLFKL
 VWTEQLNLEWDKDTQQAISFLFNRRGFSFITDGYSTEYLNIVPEQVKAILMDI FDDYNGEDDLDSYLKLATEQESKI
 SEIYNKLMQKILEFKLRKLCTDIKDDKVSTKTLKEITSYEFELLADYLANYSESLKIQKFSYTDKQGNLKELSYYHH
 DKYNIQEF LKRHATINDEILD TLLTDDFDIWNFNFEKFD FDKNEEKLQ NQEDKDHTQAHLHFFVFAVNKI KSEMASG
 GRHRSQYFQEITNVLDENNHQEGY LKNFCENLHNK KYSNL SVKNLVN LVGNLSNLELKLPLRKYFNDKI HAKADHWDE
 QKFTET YCHWILGEWRVGVK DQDKKDGAKYSYKDL CNELKQKVT KAGLIDF LLELDP CRTI PPYLDNNNRKPPKCQS
 LILNPKFLDNQYPNWQYQLQELKQLS IQDYLD S FETDLKDLKSKDQPYFVEYKSSNQMASGQRDYKDL DARI LQ
 FIFDRVKASDELL LNEIYFQAKK LKQKASSELEKLESSK KLDEVIANSQLSQILKSQHTNGIFEQGTFLHLVCKYYK
 QRQRARDSRLYIMPEYRYDKKLDKYNNTGRFDDNNQLLTYCNHKPRQKRYQLLNDLAGVLQVSRNQLLS SVEEFQ
 AQRVGEI SKSQDEQI FEWLKSKFKIASYCKAAVEMQKQYRGTLKNAIQTAI FRQSENINKNKNTGNQQALS ENSK
 DVKSLTADEKLLKLIENIAKASQKIGESLGLNDKQIKKFNSIYSFAQIQQIAFAERKGNANTCAVCSADNAHRMQQ
 IKITELVEDNKDNIILSAKAQRLPAIPTRIVD GAVKKMATILAKNIVDDNWQNIKQVLSAKHQHLHIPIITESNAF EF
 EPALADVKGKSLKDRRKKALERISPENIFKDKNRIKEFAKGISAYSANLTDGDFDGAKEELDHIIPRSHKKYGT L
 NDEANLICVTRDDNKNI FAIDTSKWF EIE TSPDLRDI G VATIQYKIDNNSRPKVRVKLDYVIDDDSKINYFMNHSLL
 KSRYPDKVL EILKQSTIIEFESSGFNKTIKEMLGMTLAGIYNETSNN

(3). 土拉弗朗西斯菌土拉亚种 MA00-2987 Cas9蛋白的氨基酸序列 (SEQ ID NO: 12)

MNVKILPIAIDL DVKNTGVFSAFYQKGTSL EKLDNKNKGVYELSKDSYTL LMNNRTARRHKRRGIDRKQLVKRFLFKL
 VWTEQLNLEWDKDTQQAISFLFNRRGFSFITDGYSTEYLNIVPEQVKAILMDI FDDYNGEDDLDSYLKLATEQESKI
 SEIYNKLMQKILEFKLRKLCTDIKDDKVSTKTLKEITSYEFELLADYLANYSESLKIQKFSYTDKQGNLKELSYYHH
 DKYNIQEF LKRHATINDEILG TLLTDDFDIWNFNFEKFD FDKNEEKLQ NQEDKDHTQAHLHFFVFAVNKI KSEMASG
 GRHRSQYFQEITNVLDENNHQEGY LKNFCENLHNK KYSNL SVKNLVN LVGNLSNLELKLPLRKYFNDKI HAKADHWDE
 QKFTET YCHWILGEWRVGVK DQDKKDGAKYSYKDL CNELKQKVT KAGLIDF LLELDP CRTI PPYLDNNNRKPPKCQS
 LILNPKFLDNQYPNWQYQLQELKQLS IQDYLD S FETDLKDLKSKDQPYFVEYKSSNQMASGQRDYKDL DARI LQ
 FIFDRVKASDELL LNEIYFQAKK LKQKASSELEKLESSK KLDEVIANSQLSQILKSQHTNGIFEQGTFLHLVCKYYK
 QRQRARDSRLYIMPEYRYDKKLDKYNNTGRFDDNNQLLTYCNHKPRQKRYQLLNDLAGVLQVSRNQLLS SVEEFQ
 AQRVGEI SKSQDEQI FEWLKSKFKIASYCKAAVEMQKQYRGTLKNAIQTAI FRQSENINKNKNTGNQQALS ENSKDV
 KSLTADEKLLKLIENIAKASQKIGESLGLNDKQIKKFNSIYSFAQIQQIAFAKRKGNANTCAVCSADNAHRMQQIK
 ITELVEDNKDNIILSAKAQRLPAIPTRIVD GAVKKMATILAKNIVDDNWQNIKQVLSAKHQHLHIPIITESNAF EF
 ALADVKGKSLKDRRKKALERISPENIFKDKNRIKEFAKGISAYSANLTDGDFDGAKEELDHIIPRSHKKYGT LND

EANLICVTRDDNKNI FAIDTSKWFIEIETPSDLRDIGVATI QYKI DNNSRPKVRVKLDYVIDDDSKINYFMNHSLLKS
RYPDKVLEILKQSTII EFESSGFNKTI KEMLGMTLAGIYNETSNN

(4). 土拉弗朗西斯菌土拉亚种菌株 FSC 198 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 13)

MNVKILPIAIDLVDKNTGVFSAFYQKGTSLKLDNKNKGVYELSKDSYTLMMNRTARRHKRRGIDRKQLVKRFLFKL
VWTEQLNLEWDKDTQQAISFLFNRRGFSFITDGYSTEYLNIVPEQVKAILMDIFDDYNGEDDLDSYLKLATEQESKI
SEIYNKLMQKILEFKLRKLCCTDIKDDKVSTKTLKEITSYEFELLADYLANYSESLKTQKFSYTDKQGNLKELSYYHH
DKYNIQEFELKRHATINDEILGTLTDDFDIWNFNFEKFDKNEEKLQNEQEDKDHTQAHLHHFVFAVVKIKSEMASG
GRHRSQYFQEIITNVLDENNHQEGYLKNCENLHNKYSNLSVKNLVNLVGNLSNLELKLPLRKYFNDKIHAKADHWDE
QKFTETYCHWILGEWRVGVKDDKKGAKYSYKDLNCLKQKVTKAGLIDFLELDP CRTIPPYLDNNNRKPPKCQS
LILNPKFLDNQYPNWQQYLQELKQLS IQDYLDSEFETDLKDLKSSKDQPYFVEYKSSNQMASGQRDYKDL DARI LQ
FIFDRVKASDELLLEIYFQAKKQKASSELEKLESSKLDDEVIANSQLSQILKSQHTNGIFEQGTFLHLVCKYYK
QRQRARDSRLYIMPEYRYDKKLDKYNNTGRFDDNNQLLTYCNHKPPQKRYQLLNDLAGVQLVSRNQLLSVVEWFQQ
AQRVGEISKSQDEQIFEWLKSFKIASYCKAAVEMQKQYRGTLKNAIQTAIFRQSENINKNKNTGNQQALSSENSKDV
KSLTADEKLLKLIENIAKASQKIGESLGLNDKQIKKFNSIYSAQIQQIAFAKRKGNANTCAVCSADNAHRMQQIK
ITELVEDNKDNIILSAKAQRLPAIPTRIVDGAVKMATILAKNIVDDNWQNIKQVLSAKHQHLPIITESNAFEFEP
ALADVKGKSLKDRRKKALERISPENIFKDKNRRIKEFAKGISAYSGANLTDGDFDGAKEELDHIIPRSHKKYGTLDN
EANLICVTRDDNKNI FAIDTSKWFIEIETPSDLRDIGVATI QYKI DNNSRPKVRVKLDYVIDDDSKINYFMNHSLLKS
RYPDKVLEILKQSTII EFESSGFNKTI KEMLGMTLAGIYNETSNN

(5). 华德萨特氏菌 CAG:135 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 14)

MRPISMKDRGWEPDRFKKTLVRAFKFFHPAKDQNKQHLELIKQIENSEDI IETLCTLDPNRTIPPYEDQNNRRPPLD
QTLLELSPKLRQYGEIWKTSARLTSDEPTLAPAAEILERS TDRKSRVAVNGHEPQPTLAYQLSYALQRAFDRSKA
LDPYALRALAAGSKSNKLT SARTALENCIGGQNVKFLDCARRYREADDAKVLWFDNAEGLLERSDLHP PMKKKI
LPLLVANILQTDETTQKFLDEIWRKQIKGRETVASRCARIETVRKSFSGGFNIAYNQAQYREVNKLPRNAQDKELL
IIRDRVAETADFIANLGLSDEQKRKFANPFLAQFYTLIETEVSGFSATTLAVHLENAWRMTIKDAVINGETVRAA
QCSRLPAETARPF DGLVRRLLVDRQAWELAKRASTDIQSKVDFSNQIVDVSIFVEENKFEFSASVADLKKNKRVKDKM
LSEAELKLETRWLNKNERIKKASRETCPYTGDRLAEGGEIDHILPRSLIKDARGIVFNAEPNLIYASSRGNQLKKNQR
YLSKLDPDYLNKVFKTSNIAAITAEIEDVVTKLQQT HRLKFFDLNEHEQDCVRHALFLDDGSEARNAVLELLATQ
RRTRVNGTQIWMIKSLANKI REELQDWCRTTNNRHLHFQAAATDVSTAKNLRLKLAQNQPNFEKPDIIQPIASHSIDAL
CSFAVGSADAERDQDGFYLDGKAVLGLYPQSCVIRLQAKPQEEKSHFDSVAIFKEGIYAEQFLPIFTLNGRIWIG
YETLDSKGERCGAIEVSGKKPEELLSMLAPFFNKPVGDLSAHATYRIQKKPAYEFLAKAALQPLSAEEKRLAALLDA
LRYCTSRKSLKSFMAANGKSLKKREDLLKPKFLQKVELKGEKAFKLNGLSLTPMKQDWLRI CNSPELADAFGNPC
SADELTSKLAQIWKRPVMRDLAHAPVRREFSLPVIDNPSGGFRI RRTNLFGNELYQVHAINAKKYRGFASADSNDVW
ANGILFNEHQHENLTECCGRFIT SADVTPMSEWRKVVTEDNLRWIAPGTEGRRYVRVETTFIQASHWFEQSVENWA
ITSPSLPASFKVDKPAEFQKAVGTELSLGLQPRSEIFIESVGNAKHIRFWYIVVSSNKKMNESYNNASKS

(6). 土拉弗朗西斯菌土拉亚种 831 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 15)

MNVKILPIAIDLVDKNTGVFSAFYQKGTSLKLDNKNKGVYELSKDSYTLMMNRTARRHQRRGIDRKQLVKRFLFKL
VWTEQLNLEWDKDTQQAISFLFNRRGFSFITDGYSTEYLNIVPEQVKAILMDIFDDYNGEDDLDSYLKLATEQESKI
SEIYNKLMQKILEFKLRKLCCTDIKDDKVSTKTLKEITSYEFELLADYLANYSESLKIQKFSYTDKQGNLKELSYYHH
DKYNIQEFELKRHATINDEILD TLLTDDFDIWNFNFEKFDKNEEKLQNEQEDKDHTQAHLHHFVFAVVKIKSEMASG
GRHRSQYFQEIITNVLDENNHQEGYLKNCENLHNKYSNLSVKNLVNLVGNLSNLELKLPLRKYFNDKIHAKADHWDE
QKFTETYCHWILGEWRVGVKDDKKGAKYSYKDLNCLKQKVTKAGLIDFLELDP CRTIPPYLDNNNRKPPKCQS
LILNPKFLDNQYPNWQQYLQELKQLS IQDYLDSEFETDLKDLKSSKDQPYFVEYKSSNQMASGQRDYKDL DARI LQ
FIFDRVKASDELLLEIYFQAKKQKASSELEKLESSKLDDEVIANSQLSQILKSQHTNGIFEQGTFLHLVCKYYK
QRQRARDSRLYIMPEYRYDKKLDKYNNTGRFDDNNQLLTYCNHKPPQKRYQLLNDLAGVQLVSRNQLLSVVEWFQQ
AQRVGEISKSQDEQIFEWLKSFKIASYCKAAVEMQKQYRGTLKNAIQTAIFRQSENINKNKNTGNQQALSSENSK
DVKSLTADEKLLKLIENIAKASQKIGESLGLNDKQIKKFNSIYSAQIQQIAFAERKGNANTCAVCSADNAHRMQQ
IKITELVEDNKDNIILSAKAQRLPAIPTRIVDGAVKMATILAKNIVDDNWQNIKQVLSAKHQHLPIITESNAFEFEP
EPALADVKGKSLKDRRKKALERISPENIFKDKNRRIKEFAKGISAYSGANLTDGDFDGAKEELDHIIPRSHKKYGTLDN
NDEANLICVTRDDNKNI FAIDTSKWFIEIETPSDLRDIGVATI QYKIDNNSRPKVRVKLDYVIDDDSKINYFMNHSLL
KSRYDPDKVLEILKQSTII EFESSGFNKTI KEMLGMTLAGIYNETSNN

(7). 嗜肺军团菌嗜肺亚种菌株 ATCC 43703 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 16)。

MESSQILSPIGIDLGGKFTGVCLSHLEAFAELPNHANTKYSVILIDHNNFQLSQAQRRA TRHRVRNKKRNQFVKRVA
 LQLFQHI LSRDLNAKEETALCHYLNNRGYTYVDTDLDEYIKDETTINLLKELLPSSEHNFIDWFLQKMQSSEFRKI
 LVSKVEEKDDKELKNAVKNIKNFITGF EKNSVEGHRHRKRVYFENIKSDITKDNQLDSIKKKIPSVCLSNLLGHLSN
 LQWKNLHRYLAKNPKQFDEQTFGNEFLRMLKNFRHLKGSQESLAVRNLIQQLEQSQDYISILEKTPPEITIPPYEAR
 TNTGMEKDQSLLLNPEKLNLYPNWRNLI PGIIDAHFPLEKDEHTKLRDRKRIISPSKQDEKRDSYILQRYLDL NK
 KIDKFKIKKQLSFLGQGKLPANLIETQKEMETHFNSSLVSVLIQIASAYNKEREDAAQGIWFDNAFRLCELSNINP
 PRKQKILPLLVGAILSEDFINNKDKWAKFKNFWNTHKIGRTSLKSKCKEIEEARKNSGNAFKIDYEEALNHPEHSNN
 KALIKIIQTI PDI IQAIQSHLGHND SQALIYHNPFSLSQLYTILETKRDGFHKNCVAVTCENYWR SQKTEIDPEISY
 ASRLPADSVRPFDFGVLARMMQRLAYEIAMAKWEQIKHIPDNSSLLIPIYLEQNRFEFEESFKKIKGSSSDKTLEQAI
 EKQNIQWEEKFQRIINASMNICPYKGASIGGQGEIDHIYPRSLSKKHFGVIFNSEANLIYCSSQGNREKKEEHYLL E
 HLSPLYLKHQFGTDNVDIKNFI SQNVANI KKYISFHLLTPEQQKAARHALFLDYDDEAFKTIKFLMSQQKARVNG
 TQKFLGKQIMEFLSTLADSKQLQLEFSIKQITAEVVDHRELLSKQEPKLVKSQQQSPSHAI DATLTMSIGLKEFP
 QFSQELDN SWFINHLM PDEVHLNPVRSKEKYNKPNISSTPLFKDSL YAERFIPVWVKGETFAIGFSEKDLFEIKPSN
 KEKLFLLKTYSTKNPGESLQELQAKSKAKWLYFPINKTLALEFLHHYFHKEIVTPDDTTVCHFINSLRYTTKKEI
 TVKILKEPMPVLSVKFESSKKNVLSFKHTIALPATKDWERLFNHPN FLALKANPAPNPKEFNEFIRKYFLSDNNPN
 SDIPNNGHNIKPKQKHKA VRKVFSLPVI PGNAGTMMRI RRDKNKGQPLYQLQTI DDTTPSMGIQINEDRLVKQEVLM DA
 YKTRNLSTIDGINNSEGQAYATFDNWLTLPVSTFKPEI IKLEMKPHSKTRRYIRITQSLADFIKTIDEALMIKPSDS
 IDDP LNMPNEIVCKNKLFGNELKPRDGKMKIVSTGKIVTYEFESDSTPQWIQTLYVTQLKKQP

(8). 嗜肺军团菌菌株 Ymg289 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 17).

MESSQILSPIGIDLGGKFTGVCLSHLEAFAELPNHANTKYSVILIDHNNFQLSQAQRRA TRHRVRNKKRNQFVKRVA
 LQLFQHI LSRDLNAKEETALCHYLNNRGYTYVDTDLDEYIKDETTINLLKELLPSSEHNFIDWFLQKMQSSEFRKI
 LVSKVEEKDDKELKNAVKNIKNFITGF EKNSVEGHRHRKRVYFENIKSDITKDNQLDSIKKKIPSVCLSNLLGHLSN
 LQWKNLHRYLAKNPKQFDEQTFGNEFLRMLKNFRHLKGSQESLAVRNLIQQLEQSQDYISILEKTPPEITIPPYEAR
 TNTGMEKDQSLLLNPEKLNLYPNWRNLI PGIIDAHFPLEKDEHTKLRDRKRIISPSKQDEKRDSYILQRYLDL NK
 KIDKFKIKKQLSFLGQGKLPANLIETQKEMETHFNSSLVSVLIQIASAYNKEREDAAQGIWFDNAFRLCELSNINP
 PRKQKILPLLVGAILSEDFINNKDKWAKFKI FWNTHKIGRTSLKSKCKEIEEARKNSGNAFKIDYEEALNHPEHSNN
 KALIKIIQTI PDI IQAIQSHLGHND SQALIYHNPFSLSQLYTILETKRDGFHKNCVAVTCENYWR SQKTEIDPEISY
 ASRLPADSVRPFDFGVLARMMQRLAYEIAMAKWEQIKHIPDNSSLLIPIYLEQNRFEFEESFKKIKGSSSDKTLEQAI
 EKQNIQWEEKFQRIINASMNICPYKGASIGGQGEIDHIYPRSLSKKHFGVIFNSEVNLIYCSSQGNREKKEEHYLL E
 HLSPLYLKHQFGTDNVDIKNFI SQNVANI KKYISFHLLTPEQQKAARHALFLDYDDEAFKTIKFLMSQQKARVNG
 TQKFLGKQIMEFLSTLADSKQLQLEFSIKQITAEVVDHRELLSKQEPKLVKSQQQSPSHAI DATLTMSIGLKEFP
 QFSQELDN SWFINHLM PDEVHLNPVRSKEKYNKPNISSTPLFKDSL YAERFIPVWVKGETFAIGFSEKDLFEIKPSN
 KEKLFLLKTYSTKNPGESLQELQAKSKAKWLYFPINKTLALEFLHHYFHKEIVTPDDTTVCHFINSLRYTTKKEI
 TVKILKEPMPVLSVKFESSKKNVLSFKHTIALPATKDWERLFNHPN FLALKANPAPNPKEFNEFIRKYFLSDNNPN
 SDIPNNGHNIKPKQKHKA VRKVFSLPVI PGNAGTMMRI RRDKNKGQPLYQLQTI DDTTPSMGIQINEDRLVKQEVLM DA
 YKTRNLSTIDGINNSEGQAYATFDNWLTLPVSTFKPEI IKLEMKPHSKTRRYIRITQSLADFIKTIDEALMIKPSDS
 IDDP LNMPNEIVCKNKLFGNELKPRDGKMKIVSTGKIVTYEFESDSTPQWIQTLYVTQLKKQP

(9). 土拉弗朗西斯菌新凶手亚种 D9876 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 18).

MNFKILPIAIDLGVKNTGVVSAFYQKGTSLERLDNKNKGVYELSKDSYTL LMNRTARRHQRRGIDRKQLVKKL FKL
 IWTEQLNLEWDKDTQQAISFLFNRRGFSFITDGYSP EYLNIVPEQVKA I LMDI FDDYNGEDDLDSYLKLATEQESKI
 SEIYNKLMQKILEFKLRKLC TDIKDDKVSTKTLKEITSYEFELLADYLANYESLKTQKFSYTDKQGNLKELSYHH
 DKYNIQEF LKRHATINDEILD TLLTDDLDIWNFNFEKDFDKNEEKLQNQEDKDHTQAHLHHFVFAVNKIKSEMASG
 GRHRGQYFQEITNVLDENNHQEGYLKNFCENLHNKYSNLSVKNLVNLI GNLSNLELKLPRKYFNDKIHAKADHWDE
 QKFTET YCHWILGEWRVGLKQDKKEGAKYSYKDLCELKQKVTNAGLIDFLELDP CRTIPPYLDNNNRKPKPCQS
 LILNPKFLDNQYPNWQQYLQELKQLIQDYLESFETDLKVLKSSKDQPYFVEYKSSNQQIASGQRDYKDL DARI LQ
 FIFDRVKASDELLNEIYFQAKLQKASSELEKLESSKLDDEVIANSQLSKILKSQHTNGIFEQGTFLHLVCKYYK
 QRQRARDSRLYIMPEYRYDKKLDKYNNTGRFDDDNQLLTYCNHKPRQKRYQLLNDLAGVLQVSPNFLKDKIGSDDDL
 FISKWLVEHIRGFKKACEDSLKIQKDNRGLLNHKINIARNTKGKCEKEIFNLI CKIEGSEDKKGNYKHGLAYELGVL
 LFGEPNEASKPEFDRKIKKFNSIYSFAQIQQIAFAERKGNANTCAVCSADNAHRMQQIKITEPVEDNKDKIILSAKA
 QRLPAIPTRIVDGAVKMATILAKNIVDDNWQNIKQVLSAKHQLHIPIT ESNAFEPALADVKGKSLKDRRKKAL
 ERISPENIFKDKNNRIKEFAKGISAYSGANLTDGDFDGAKEELDHIIPRSHKKYGTLNDEANLIVTRDDNKKNKGNR
 IFCLHDLANNYK LKQFETTDLEIEKKIADTIWDASKKDFKFGNYRSFINLTPQEQA FRHALFLADENPIKQAVIR
 AINNRNRTFVNGTQRYFAEVLANNIYLRAKKENLNTDKISFDYFGIPTIGNGRGIAEIRQLYEKVDSDIQVYAKGDK
 PQASYSHLIDAMLAFCIAA DEHRNDGSI GLEIDKNYSLYPLDKNTGEVFTKDI FSQIKITDNEFSDKKLVRKKATEG
 FNTHRQMTRDGIYAENYLPILIHKELDEVRKGYTWKNSEEIKIFKGKKYDIQQLNNLLYCLKFVDKPI SIDIQISTL

EELRNILTTNNAATAEYYYYINLKTQKLHEYYIENYNTALGYKKYTKEMEFRLSLAYRSERVKI KSIDDVKQVLDKD
SNFIIGKITLFPKKEWQRLYREWQNTTIKDDYEFLKSFFNVKSI TKLHKVRKDFSLPISTNEGKFLVCRKTDWNNF
IYQILNDSDSRVDGTFKPFIPAFDISKNEIVEAIDSFTSKNIFWLPKNIELQKVDNKNIFAIDTSKWFEVETPSDLR
DIGVATIYQKIDNNSRPKVRVKLDYVIDDDSKINYFTNHSLLKSRYPDKVLEILKQSTIIEFESSGFNKTIKEMLGM
TLAGIYNETSNN

(10). 嗜肺军团菌菌株 Leg01/11 Cas9蛋白的氨基酸序列 (SEQ ID NO: 19).

MESSQILSPIGIDLGGKFTGVCLSHLEAFAELPNHANTKYSVILIDHNNFQLSQAQRATRHRVRNKKRNQFVKRVA
LQLFQHILSRDLNAKEETALCHYLNNRGYTYVDTDLDEYIKDETTINLLKELLPSSEHNFI DWFLQKMQSSEFRKI
LVSKVEEKDDKELKNAVKNIKNFITGF EKNSVEGHRHRKRVYFENIKSDITKDNQLDSIKKKIPSVCLSNLLGHLSN
LQWKNLHRYLAKNPKQFDEQTFGNEFLRMLKNFRHLKGSQESLAVRNLIQQLEQSQDYISILEKTPPEITIPPEAR
TNTGMEKDQSLLLNPEKLNLYPNWRNLI PGI DAHPFLEKDEHTKLRDRKRIISPSKQDEKRDSYILQRYLDLKN
KIDKFKIKKQLSFLGQGKQLPANLIETQKEMETHFNSSLVSVLIQIASAYNKEREDAAQGIWFDNAFSLCELSNINP
PRKQKILPLLVGAILSEDFINNKDKWAKFKI FWNTHKIGRTSLKSKCKEIEEARKNSGNAFKIDYEEALNHPEHSNN
KALIKIIQTI PDI IQAIQSHLGHNSQALI YHNPFSLSQLYTILETKRDGFHKNCVAVTCENYWRSSQKTEIDPEISY
ASRLPADSVRPFDFGLARMMQRLAYEIAMAKWEQIKHIPDNSSLLIPIYLEQNRFEFEESFKKIKGSSSDKTLEQAI
EKQNIQWEEKFQRI INASMNICPYKGASIGGQGEIDHIYPRSLSKKHFGVIFNSEVNLIYCSSQGNREKKEEHYLL
HLSPLYLKHQFGTDNVDIKNFISQNVANI KKYISFHLLTPEQQKAARHALFLDYDDEAFKTIKFLMSQQKARVNG
TQKFLGKQIMEFLSTLADSKQLQLEFSIKQITAEVHDHRELLSKQEPKLVKSRQQSFP SHAI DATLTMSIGLKEFP
QFSQELDNSWFINHLMPEVHLNPVRSKEKYNKPNISSTPLFKDSLAEYERFIPVWVKGETFAIGFSEKDLFEIKPSN
KEKLFLLKTYSTKNPGESELQELQAKSKAKWLYFPINKTLALEFLHYYFHKEIVTPDDTTVCHFINSLRYTTKESI
TVKILKEPMPVLSVKFESSKKNVLSFKHTIALPATKDWERLFNHPN FLALKANPAPNPKFENEFIRKYFLSDNNPN
SDIPNNGHNIKPKQKHKAVRKFVSLPVI PGNAGTMMRI RRDKNKGQPLYQLQTI DDTTPSMGIQINEDRLVKQEVLMDA
YKTRNLSTIDGINNSEGQAYATFDNWLTLPVSTFKPEIKLEMKPHSKTRRYIRITQSLADFIKTIDEALMIKPSDS
IDDP LNMPNEIVCKNKLFGNELKPRDGKMKIVSTGKI VTYEFESDSTPQWIQTLYVTQLKKQP

(11). 嗜肺军团菌菌株 121004 Cas9蛋白的氨基酸序列 (SEQ ID NO: 20).

MESSQILSPIGIDLGGKFTGVCLSHLEAFAELPNHANTKYSVILIDHNNFQLSQAQRATRHRVRNKKRNQFVKRVA
LQLFQHILSRDLNAKEETALCHYLNNRGYTYVDTDLDEYIKDETTINLLKELLPSSEHNFI DWFLQKMQSSEFRKI
LVSKVEEKDDKELKNAVKNIKNFITGF EKNSVEGHRHRKRVYFENIKSDITKDNQLDSIKKKIPSVCLSNLLGHLSN
LQWKNLHRYLAKNPKQFDEQTFGNEFLRMLKNFRHLKGSQESLAVRNLIQQLEQSQDYISILEKTPPEITIPPEAR
TNTGMEKDQSLLLNPEKLNLYPNWRNLI PGI DAHPFLEKDEHTKLRDRKRIISPSKQDEKRDSYILQRYLDLKN
KIDKFKIKKQLSFLGQGKQLPANLIETQKEMETHFNSSLVSVLIQIASAYNKEREDAAQGIWFDNAFSLCELSNINP
PRKQKILPLLVGAILSEDFINNKDKWAKFKI FWNTHKIGRTSLKSKCKEIEEARKNSGNAFKIDYEEALNHPEHSNN
KALIKIIQTI PDI IQAIQSHLGHNSQALI YHNPFSLSQLYTILETKRDGFHKNCVAVTCENYWRSSQKTEIDPEISY
ASRLPADSVRPFDFGLARMMQRLAYEIAMAKWEQIKHIPDNSSLLIPIYLEQNRFEFEESFKKIKGSSSDKTLEQAI
EKQNIQWEEKFQRI INASMNICPYKGASIGGQGEIDHIYPRSLSKKHFGVIFNSEVNLIYCSSQGNREKKEEHYLL
HLSPLYLKHQFGTDNVDIKNFISQNVANI KKYISFHLLTPEQQKAARHALFLDYDDEAFKTIKFLMSQQKARVNG
TQKFLGKQIMEFLSTLADSKQLQLEFSIKQITAEVHDHRELLSKQEPKLVKSRQQSFP SHAI DATLTMSIGLKEFP
QFSQELDNSWFINHLMPEVHLNPVRSKEKYNKPNISSTPLFKDSLAEYERFIPVWVKGETFAIGFSEKDLFEIKPSN
KEKLFLLKTYSTKNPGESELQELQAKSKAKWLYFPINKTLALEFLHYYFHKEIVTPDDTTVCHFINSLRYTTKESI
TVKILKEPMPVLSVKFESSKKNVLSFKHTIALPATKDWERLFNHPN FLALKANPAPNPKFENEFIRKYFLSDNNPN
SDIPNNGHNIKPKQKHKAVRKFVSLPVI PGNAGTMMRI RRDKNKGQPLYQLQTI DDTTPSMGIQINEDRLVKQEVLMDA
YKTRNLSTIDGINNSEGQAYATFDNWLTLPVSTFKPEIKLEMKPHSKTRRYIRITQSLADFIKTIDEALMIKPSDS
IDDP LNMPNEIVCKNKLFGNELKPRDGKMKIVSTGKI VTYEFESDSTPQWIQTLYVTQLKKQP

(12). 嗜肺军团菌血清组1菌株 TUM 13947 Cas9蛋白的氨基酸序列 (SEQ ID NO: 21).

MESSQILSPIGIDLGGKFTGVCLSHLEAFAELPNHANTKYSVILIDHNNFQLSQAQRATRHRVRNKKRNQFVKRVA
LQLFQHILSRDLNAKEETALCHYLNNRGYTYVDTDLDEYIKDETTINLLKELLPSSEHNFI DWFLQKMQSSEFRKI
LVSKVEEKDDKELKNAVKNIKNFITGF EKNSVEGHRHRKRVYFENIKSDITKDNQLDSIKKKIPSVCLSNLLGHLSN
LQWKNLHRYLAKNPKQFDEQTFGNEFLRMLKNFRHLKGSQESLAVRNLIQQLEQSQDYISILEKTPPEITIPPEAR
TNTGMEKDQSLLLNPEKLNLYPNWRNLI PGI DAHPFLEKDEHTKLRDRKRIISPSKQDEKRDSYILQRYLDLKN
KIDKFKIKKQLSFLGQGKQLPANLIETQKEMETHFNSSLVSVLIQIASAYNKEREDAAQGIWFDNAFSLCELSNINP
PRKQKILPLLVGAILSEDFINNKDKWAKFKI FWNTHKIGRTSLKSKCKEIEEARKNSGNAFKIDYEEALNHPEHSNN
KALIKIIQTI PDI IQAIQSHLGHNSQALI YHNPFSLSQLYTILETKRDGFHKNCVAVTCENYWRSSQKTEIDPEISY
ASRLPADSVRPFDFGLARMMQRLAYEIAMAKWEQIKHIPDNSSLLIPIYLEQNRFEFEESFKKIKGSSSDKTLEQAI

EKQNIQWEEKFQRI INASMNICPYKGASIGGQGEIDHIYPRSLSKKHFGVIFNSEANLIYCSSQGNREKKEEHYLL
 HLSPLYLKHQFGT DNVS DIKNFISQNVANI KKYISFHLLTPEQQKAARHALFLDYDDEAFKTIKFLMSQQKARVNG
 TQKFLGKQIMEFLSTLADSKQLQLEFSIKQITAEVVDHRELLSKQEPKLVKSQQQSFP SHAIDATLTMSIGLKEFP
 QFSQELDNWSFINHLMPEVHLNPNVRSKEKYNKPNISSTPLFKDSLAEERFIPVWVKGETFAIGFSEKDLFEIKPSN
 KEKLFLLKTYSTKNPGESLQELQAKSKAKWLYFPINKTLALEFLHHYFHKEIVTPDDTTVCHFINSLRYYTKKESI
 TVKILKEPMPVLSVKFESSKKNVLGSKHTIALPATKDWERLFNHPN FLALKANPAPNPKEFNEFIRKYFLSDNNPN
 SDIPNNGHNIKPKQKHKAVRKVFLPVI PGNAGTMMRI RRDKNKGQPLYQLQTI DDTPSMGIQINEDRLVKQEVLMDA
 YKTRNLSTIDGINNSEGQAYATFDNWLTLPVSTFKPEI IKLEMKPHSKTRRYIRITQSLADFIKTIDEALMIKPSDS
 IDDP LNMPNEIVCKNKLFGNELKPRDGKMKIVSTGKIVTYEFESDSTPQWIQTLYVTQLKKQP

(13). 嗜肺军团菌血清组1分离株 ID120371 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 22)。

MESSQILSPIGIDLGGKFTGVCLSHLEAFAELPNHANTKYSVILIDHNNFQLSQAQRATRHRVRNKKRNQFVKRVA
 LQLFQHILSRDLNAKEETALCHYLNNRGYTYVDTDLDEYIKDETTINLLKELLPSSEHNFIWDFLQKMQSSEFRKI
 LVSKVEEKDDKELKNAVKNIKNFITGF EKNSVEGHRHRKRVYFENIKSDITKDNQLDSIKKKIPSVCLSNLLGHLSN
 LOWKNLHRYLAKNPKQFDEQTFGNEFLRMLKNFRHLKGSQESLAVRNLIQQLEQSQDYISILEKTPPEITIPPEAR
 TNTGMEKDQSLLLNPEKLNLYPNWRNLI PGIIDAHPFLEKDLHTKLRDRKRIISPSKQDEKRDSYILQRYLDLNLK
 KIDKFKIKKQLSFLGQGGKLPANLIETQKEMETHFNSLLVSVLIQIASAYNKEREDAAQGIWFDNAFRLCELSNINP
 PRKQKILPLLVGAILSEDFINNKDKWAKFKNFWNTHKIGRTSLKSKCKEIEEARKNSGNAFKIDYEEALNHPEHSNN
 KALIKIIQTI PDI IQAIQSHLGHND SQALI YHNPFSLSQLYTILETKRDGFHKNCVAVTCENYWRSQKTEIDPEISY
 ASRLPADSVRPF DGVLARMMQRLAYEIAMAKWEQIKHIPDNSSLLIPIYLEQNRFEFEESFKKIKGGSSDKTLEQAI
 EKQNIQWEEKFQRI INASMNICPYKGASIGGQGEIDHIYPRSLSKKHFGVIFNSEANLIYCSSQGNREKKEEHYLL
 HLSPLYLKHQFGT DNVS DIKNFISQNVANI KKYISFHLLTPEQQKAARHALFLDYDDEAFKTIKFLMSQQKARVNG
 TQKFLGKQIMEFLSTLADSKQLQLEFSIKQITAEVVDHRELLSKQEPKLVKSQQQSFP SHAIDATLTMSIGLKEFP
 QFSQELDNWSFINHLMPEVHLNPNVRSKEKYNKPNISSTPLFKDSLAEERFIPVWVKGETFAIGFSEKDLFEIKPSN
 KEKLFLLKTYSTKNPGESLQELQAKSKAKWLYFPINKTLALEFLHHYFHKEIVTPDDTTVCHFINSLRYYTKKESI
 TVKILKEPMPVLSVKFESSKKNVLGSKHTIALPATKDWERLFNHPN FLALKANPAPNPKEFNEFIRKYFLSDNNPN
 SDIPNNGHNIKPKQKHKAVRKVFLPVI PGNAGTMMRI RRDKNKGQPLYQLQTI DDTPSMGIQINEDRLVKQEVLMDA
 YKTRNLSTIDGINNSEGQAYATFDNWLTLPVSTFKPEI IKLEMKPHSKTRRYIRITQSLADFIKTIDEALMIKPSDS
 IDDP LNMPNEIVCKNKLFGNELKPRDGKMKIVSTGKIVTYEFESDSTPQWIQTLYVTQLKKQP

(14). 嗜肺军团菌菌株 Nagoya-1 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 23)。

MESSQILSPIGIDLGGKFTGVCLSHLEAFAELPNHANTKYSVILIDHNNFQLSQAQRATRHRVRNKKRNQFVKRVA
 LQLFQHILSRDLNAKEETALCHYLNNRGYTYVDTDLDEYIKDETTINLLKELLPSSEHNFIWDFLQKMQSSEFRKI
 LVSKVEEKDDKELKNAVKNIKNFITGF EKNSVEGHRHRKRVYFENIKSDITKDNQLDSIKKKIPSVCLSNLLGHLSN
 LOWKNLHRYLAKNPKQFDEQTFGNEFLRMLKNFRHLKGSQESLAVRNLIQQLEQSQDYISILEKTPPEITIPPEAR
 TNTGMEKDQSLLLNPEKLNLYPNWRNLI PGIIDAHPFLEKDLHTKLRDRKRIISPSKQDEKRDSYILQRYLDLNLK
 KIDKFKIKKQLSFLGQGGKLPANLIETQKEMETHFNSLLVSVLIQIASAYNKEREDAAQGIWFDNAFRLCELSNINP
 PRKQKILPLLVGAILSEDFINNKDKWAKFKNFWNTHKIGRTSLKSKCKEIEEARKNSGNAFKIDYEEALNHPEHSNN
 KALIKIIQTI PDI IQAIQSHLGHND SQALI YHNPFSLSQLYTILETKRDGFHKNCVAVTCENYWRSQKTEIDPEISY
 ASRLPADSVRPF DGVLARMMQRLAYEIAMAKWEQIKHIPDNSSLLIPIYLEQNRFEFEESFKKIKGGSSDKTLEQAI
 EKQNIQWEEKFQRI INASMNICPYKGASIGGQGEIDHIYPRSLSKKHFGVIFNSEANLIYCSSQGNREKKEEHYLL
 HLSPLYLKHQFGT DNVS DIKNFISQNVANI KKYISFHLLTPEQQKAARHALFLDYDDEAFKTIKFLMSQQKARVNG
 TQKFLGKQIMEFLSTLADSKQLQLEFSIKQITAEVVDHRELLSKQEPKLVKSQQQSFP SHAIDATLTMSIGLKEFP
 QFSQELDNWSFINHLMPEVHLNPNVRSKEKYNKPNISSTPLFKDSLAEERFIPVWVKGETFAIGFSEKDLFEIKPSN
 KEKLFLLKTYSTKNPGESLQELQAKSKAKWLYFPINKTLALEFLHHYFHKEIVTPDDTTVCHFINSLRYYTKKESI
 TVKILKEPMPVLSVKFESSKKNVLGSKHTIALPATKDWERLFNHPN FLALKANPAPNPKEFNEFIRKYFLSDNNPN
 SDIPNNGHNIKPKQKHKAVRKVFLPVI PGNAGTMMRI RRDKNKGQPLYQLQTI DDTPSMGIQINEDRLVKQEVLMDA
 YKTRNLSTIDGINNSEGQAYATFDNWLTLPVSTFKPEI IKLEMKPHSKTRRYIRITQSLADFIKTIDEALMIKPSDS
 IDDP LNMPNEIVCKNKLFGNELKPRDGKMKIVSTGKIVTYEFESDSTPQWIQTLYVTQLKKQP

(15). 土拉弗朗西斯菌新凶手亚种 U112 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 24)。

MNFKILPIAIDLGVKNTGVVSAFYQKGTSLERLDNKNKGVYELSKDSYTLMMNRTARRHQRRGIDRKQLVKRFLK
 IWTEQLNLEWDKDTQQAISFLFNRRGFSFITDGYSP EYLNIVPEQVKA ILMDFDDYNGEDDLDSYLKLATEQESKI
 SEIYNKLMQKILEFKLMKLC TDI KDDKVSTKTLKEITSYEFELLADYLANYESLKTQKFSYTDKQGNLKELSYHH
 DKYNIQEF LKRHATINDRI LDTLLTDDLDIWNFNFEKFDKNEEKLQNQEDKDHIQAHLHFFVFAVNKIKSEMASG
 GRHRSQYFQ EITNVLDENNHQEGYLNFCENLHNKYSNLSVKNLVNLI GNLSNLELKLPRKYFNDKIHAKADHWDE

QKFTETECHWILGEWRVGVKQDKKDGAKYSYKDLNELKQKVTKAGLVDLLELDP CRTI PPYLDNNNRKPPKCQS
 LILNPKFLDNQYPNWQQYLQELKKLQSIQNYLDSFETDLKVLKSSKDQPYFVEYKSSNQI ASGQRDYKDL DARI LQ
 FIFDRVKASDELLLENIYFQAKKQKASSELEKLESSKKLDEVIANSQLSQILKSQHTNGIFEQGTFLHLVCKYYK
 QRQRARDSRLYIMPEYRYDKKLHKYNNTRGFDDNQLLTYCNHKPRQKRYQLLNDLAGVLQVSPNFLKDKIGSDDDL
 FISKWLVEHIRGFKKACEDSLKIQKDNRGLLNHKIN IARNTKGKCEKEIFNLICKIEGSEDKKGNKYKHLAYELGVL
 LFGEPNEASKPEFDRKI KKFNSIYSFAQIQQIAFAERKGNANTCAVCSADNAHRMQQIKITEPVEDNKDKI ILSAKA
 QRLPAIPTRIVDGA VKKMATI LAKNIVDDNWQNI KQVLSAKHQ LHIPIITESNAFEFEPALADVKGKSLKDRRKKAL
 ERISPENIFKDKNNRIKEFAKGISAYSGANLTDGDFDGAKEELDHIIPRSHKKYGTLNDEANLIVTRGDNKKNKGNR
 IFCLRDLDANLYKQKFETDDLEIEKKIADTIWDANKKDFKFGNYRSFINLTPQEQA FRHALFLADENPIKQAVIR
 AINNRNRTFVNGTQRYFAEVLANNIYLRAKKENLNTDKISFDYFGIPTIGNGRGIAEIRQLYEKVDSDIQAYAKGDK
 PQASYSHLIDAMLAFCAADEHRNDG SIGLEIDKNYSLYPLDKNTGEVFTKDI FSQIKITDNEFSDKKLVKKAIEG
 FNTHRQMT RDGIYAENYLPILIHKELNEVRKGYTWKNSEEI KIFKGKDYDIQQLNNLVYCLK FVDKPI SIDIQI STL
 EELRNILTTNNIAATAEY YYYINLKTQKLHEYYIENYNTALGYK KYSKEMEF LRS LAYRSERVKIKSIDDVKQVLDKD
 SNFIIIGKITL PFKKEWQRLYREWQNTTIKDDYEF LKSFVVKSI TKLHKVRKDFSLPISTNEGKFLV KRKTWDNNF
 IYQILNDSDSRADGTPFIPAFDISKNEIVEAIDSFTSKNI FWLPKNI ELQKVDNKNIFAIDTSKWF EVETPSDLR
 DIGIATI QYKIDNNSRPKVRVKLDYVIDDDSKIN YFMNHSLLKSRYPDKVLEILKQSTIIEFESSGFNKTIKEM LGM
 KLAGIYNETSNN

(16). 土拉弗朗西斯菌新凶手亚种菌株 FAI Cas9 蛋白的氨基酸序列 (SEQ ID NO: 25)。

MNVKILPIAIDLGVKNTGVFSAFYQKGTSLERLDNKNKGVYELSKDSYTL LMNNRTARRHQRRGIDRKQLVKRFLFKL
 IWTEQLNLEWDKDTQQAISFLFNRRGFSFITDGYSTEYLNIVPEQVKAILMDIFDDYNGEDDLDSYLKLATEQESKI
 SEIYNKLMQKILEFKLRKLCTDIKDDKVESTKLTKEITSYEFELLADYLANYSEILKTQKFSYTDKQGNLKELSYHH
 DKYNIQEFLKRHATINDEILDTLLTDDLDIWNFNFEKFDKNEEKLNQNEQEDKDHTQAHLHHFVFAVNKIKSEMASG
 GRHRGQYFQEITNVLDENNHQEGYLKNCENLHNKYSNLSVKNLVNLI GNLSNLELKLPRKYFNDKIHAKADHWDE
 QKFTETECHWILGEWRVGVKQDKKDGAKYSYKDLNELKQKVTKAGLVDLLELDP CRTI PPYLDNNNRKPPKCQS
 LILNPKFLDNQYPNWQQYLQELKKLQSIQNYLESFETDLKVLKSSKDQPYFVEYKSSNQI ASGQRDYKDL DARI LQ
 FIFDRVKASDELLLENIYFQAKKQKASSELEKLESSKKLDEVIANSQLSQILKSQHTNGIFEQGTFLHLVCKYYK
 QRQRARDSRLYIMPEYRYDKKLDKYNNTGRFDDNQLLTYCNHKPRQKRYQLLNDLAGVLQVSPNFLKDKIGSDDDL
 FISKWLVEHIRGFKKACEDSLKIQKDNRGLLNHKIN IARNTKGKCEKEIFNLICKIEGSEDKKGNKYKHLAYELGVL
 LFGEPNEASKPEFDRKI KKFNSIYSFAQIQQIAFAERKGNANSCAVCSADNAHRMQQIKITEPVEDNKDKI ILSAKA
 QRLPAIPTRIVDGA VKKMATI LAKNIVDDNWQNI KQVLSAKHQ LHIPIITESNAFEFEPALADVKGKSIKDRRKKAL
 ERISPENIFKDKNNRIKEFAKGISAYSGDNLASGDFDGAKEELDHIIPRSHKKYGTLNDEANLIVTRDDNKNKGNR
 IFCLHDLANNYKQKFETDDLEIEKKIADTIWDASKKDFKFGNYRSFINLTPQEQA FRHALFLADENPIKQAVIR
 AINNRNRTFVNGTQRYFAEVLANNIYLRAKKENLNTDKISFDYFGIPTIGNGRGIAEIRQLYEKVDSDIQAYAKGDK
 PQASYSHLIDAMLAFCAADEHRNDG SIGLEIDKNYSLYPLDKNTGEVFTKDI FSQIKITNNEFSDKKLVKKAIEG
 FNTHRQMT RDGIYAENYLPILIHKELNEVRKGYTWGNSEELKIHKGKDYDIQQLNNLVYCLK FVDKPI SIDIQI STL
 EELRNILTTNNIATTAEY YYYINLKTQKLHEYYIENYNTALGYK KYTKEMEF LRS LAYRSERVKIKSIDDVKQVLDKD
 SNFIIIGKITL PFKKEWQRLYREWQNTTIKDDYEF LKSFVVKSI TKLHKVRKDFSLPISTNEGKFLV KRKTWDNNF
 IYQILNDSDSRVDGTPFIPAFDISKNEIVEAIDSFTSKNI FWLPKNI ELQKVDNKNIFAIDTSKWF EVETPSDLR
 DIGVATI QYKIDNNSRPKVRVKLDYVIDDDSKIN YFTNHSLLKSRYPDKVLEILKQSTIIEFESSGFNKTIKEM LGM
 TLAGIYNETSNN

(17). 土拉弗朗西斯菌土拉亚种 80700075 contig_04 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 26)。

MNVKILPIAIDLVDKNTGVFSAFYQKGTSLKLDNKNKGVYELSKDSYTL LMNNRTARRHKRRGIDRKQLVKRFLFKL
 VWTEQLNLEWDKDTQQAISFLFNRRGFSFITDGYSTEYLNIVPEQVKAILMDIFDDYNGEDDLDSYLKLATEQESKI
 SEIYNKLMQKILEFKLRKLCTDIKDDKVESTKLTKEITSYEFELLADYLANYSESILKTQKFSYTDKQGNLKELSYHH
 DKYNIQEFLKRHATINDEILGTLLTDDFDIWNFNFEKFDKNEEKLNQNEQEDKDHTQAHLHHFVFNKIKSEMASG
 GRHRGQYFQEITNVLDENNHQEGYLKNCENLHNKYSNLSVKNLVNLVGNLSNLELKLPRKYFNDKIHAKADHWDE
 QKFTETECHWILGEWRVGVKQDKKDGAKYSYKDLNELKQKVTKAGLIDFLELDP CRTI PPYLDNNNRKPPKCQS
 LILNPKFLDNQYPNWQQYLQELKKLQSIQDYLDSFETDLKDLKSSKDQPYFVEYKSSNQMASGQRDYKDL DARI LQ
 FIFDRVKASDELLLENIYFQAKKQKASSELEKLESSKKLDEVIANSQLSQILKSQHTNGIFEQGTFLHLVCKYYK
 QRQRARDSRLYIMPEYRYDKKLDKYNNTGRFDDNQLLTYCNHKPRQKRYQLLNDLAGVLQVSRNQLLSVVEWFQQ
 AQRVGEISKSQDEQIFEWLKSFKIASYCKAAVEMQKQYRGTLKNAIQTAIFRQSENINKNKNTGNQQALS ENSKDV
 KSLTADEKLLKLIENIAKASQKIGESLGLNDKQIKKFNSIYSFAQIQQIAFAKRKGNANTCAVCSADNAHRMQQIK
 ITELVEDNKDNII ILSAKAQRLPAIPTRIVDGA VKKMATI LAKNIVDDNWQNI KQVLSAKHQ LHIPIITESNAFEFEP
 ALADVKGKSLKDRRKKALERISPENIFKDKNNRIKEFAKGISAYSGANLTDGDFDGAKEELDHIIPRSHKKYGT LND

EANLICVTRDDNKNI FAIDTSKWFEIETPSDLRDIGVATI QYKIDNNSRPKVRVKLDYVIDDDSKINYFMNHSLLKS
RYPDKVLEILKQSTIIEFESSGFNKTIKEMLGMTLAGIYNETSNN

(18). 土拉弗朗西斯菌全北美区亚种 LVS Cas9 蛋白的氨基酸序列 (SEQ ID NO: 27)。

MNVKILPIAIDLVDKNTGVFSAFYQKGTSLKLDNKNKGKVEYELSKDSYTLMMNRTAQRHQRRGIDRKQLVKRFLFKL
VWTEQLNLEWDKDTQQAISFLFNRRGFSFITDGYSSTEYLNIVPEQVKAILMDIFDDYNGEDDLDLSYLKLA TEQESKI
SEIYNKLMQKILEFKLRKLCTDIKDDKVSTKTLKEITSYEFELLADYLANYESLKTQKFSYTDKQGNLKELSYYHH
DKYNIQEFLLKRHATINDEILD TLLTDDDFDIWNFNFEKFDKNEEKLQSQEDKDHTQAYFHFFVFAVNKIKSEMASG
GRHRSQYFQEITNVLDENNHQEGYLKNFCENLHNKKYSNLSVKNLNLVGNLSNLELKLPRKYFNDKNLIIGMSKSL
QKLIATGY

(19). 嗜肺军团菌菌株 Leg01/53 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 28)。

METHFNSSLVSVLIQIASAYNKEREDAAQGIWFDNAFSLCELSNINPPRKQKILPLLVGAILSEDFINNKDKWAKFK
IFWNTHKIGRTSLKSKCKEIEEARKNSGNAFKIDYEEALNHPEHSNNKALIKI IQTIPIIQA IQSHLGHNDSQALI
YHNPFSLSQLYTIILETKRDGFHKNCVAVTCENYWR SQTEIDPEISYASRLPADSVRPF DGVLARMMQRLAYEIAMA
KWEQIKHIPDNSSLLIPIYLEQNRFEFEESFKKIKGSSSDKTLEQAI EKQNIQWEEKFQRI INASMNICPYKGASIG
GQGEIDHIYPRSLSKKHFGVIFNSEVNLIYCSSQGNREKKEEHYLLHLSPLYLKHQFGTDNVSDIKNFISQNVANI
KKYISPHLLTPEQQKAARHALFLDYDDEAFKTIITKFLMSQQKARVNGTQKFLGKQIMEFLSTLADSKQLQLEFSIKQ
ITAEVVDHRELLSKQEPKLVKSRQSFPSHAIDATLTMSIGLKEFPQFSQELDNSWFINHLMPEVHLNPNVRSKEK
YNKPNISSTPLFKDLSYAEFPIVWVKGETFAIGFSEKDLFEIKPSNKEKLFLLKTYSTKNPGESLQELQAKSKAK
WLYFPINKTLALEFLHHYFHKEIVTPDDTTVCHFINSLRYTTKESITVKILKEPMPVLSVKFESSKKNVLSGFKHT
IALPATKDWERLFNHPNFLALKANPAPNPKFENEFIRKYFLSDNNPNSDI PNNGHNI KPQKHKAVRKFVSLPVI PGN
AGTMMRIRRKDNKGQPLYQLQTI DDTPSMGIQINEDRLVKQEVLM DAYKTRNLSTIDGINNSEGQAYATFDNWLTLP
VSTFKPEI IKLEMKPHSKTRRYIRITQSLADFIKTI DEALMIKPSDSIDDP LNMPNEIVCKNKLFGNELKPRDGMK
IVSTGKIVTYEFESDSTPQWIQTLYVTQLKKQP

(20). 土拉弗朗西斯菌新凶手亚种 GA99-3548 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 29)。

MNFKILPIAIDLGVKNTGVFSAFYQKGTSLERLDNKNKGKVEYELSKDSYTLMMNRTARRHQRRGIDRKQLVKKLFLK
IWTEQLNLEWDKDTQQAISFLFNRRGFSFITDGYSPEYLNIVPEQVKAILMDIFDDYNGEDDLDLSYLKLA TEQESKI
SEIYNKLMQKILEFKLRKLCTDIKDDKVSTKTLKEITSYEFELLADYLANYESLKTQKFSYTDKQGNLKELSYYHH
DKYNIQEFLLKRHATINDEILD TLLTDDLDIWNFNFEKFDKNEEKLQSQEDKDHTQAHLHFFVFAVNKIKSEMASG
GRHRGQYFQEITNVLDENNHQEGYLKNFCENLHNKKYSNLSVKNLNLIGNLSNLELKLPRKYFNDKIHAKADHWDE
QKFTETYCHWILGEWRVGLKDQDKKEGAKYSYKDLNCELKQKVTNAGLIDFLELDP CRTI PPYLDNNNRKPPKQCS
LILNPKFLDNQYPNWQQYLQELKQLQIQDYLESFETDLKVLKSSKDQPYFVEYKSSNQI ASGQRDYKDL DARI LQ
FIFDRVKASDELLNEIYFQAKKLLKQKASSELEKLESSKLLDEVIANSQLSKILKSQHTNGIFEQGTFLHLVCKYYK
QRQRARDSRLYIMPEYRYDKKLDKYNNTGRFDDDDNQLLTYCNHKPRQKRYQLLNDLAGVLQVSPNFLKDKIGSDDDL
FISKWLVEHIRGFKKACEDSLKIQKDNRGLLNHKINIARNTKKGCEKEIFNLICKIEGSEDKKGNKYGHLAYELGVL
LFGEPNEASKPEFDRKIKKFNSIYSFAQIQQIAFAERKGNANTCAVCSADNAHRMQQIKITEPVEDNKDKI ILSAKA
QRLPAIPTRIVDGA VKMATILAKNIVDDNWQNIKQVLSAKHQHLHPIITESNAFEFEPALADVKGKSLKDRRKKAL
ERISPENIFKDKNNRIKEFAKGISAYSANLTDGDFDGAKEELDHIIPRSHKKYGTLNDEANLICVTRDDNKKNKGNR
IFCLHDLANNYKLLKQFETTDLEIEKKIADTIWDASKKDFKFGNYRSFINLTPQEQA FRHALFLADENPIKQAVIR
AINNRNRTFVNGTQRYFAEVLANNIYLRAKKENLNTDKISFDYFGIPTIGNRGIAEIRQLYEKVDSDIQVYAKGDK
PQASYSHLIDAMLAFCIAADEHRNDGSIGLEIDKNYSLYPLDKNTGEVFTKDI FSQIKITDNEFSDKKLVRRKATEG
FNTHRQMT RDGIYAENYLPILIHKELDEVRKGYTWKNSEEIKIFKGKKYDIQQLNLLYCLKFVDKPI SIDIQI STL
EELRNILTTNNAATAEYYYINLKTQKLHEYYIENYNTALGYKKYTKEMEFLRS LAYRSE RVKIKSIDDVKQVLDKD
SNFII GKITL PFKKEWQRLYREWQNTTIKDDYEF LKSFFNVKSITKLHKKVRKDFSLPISTNEGKFLVRRKTWDNNF
IYQILNDSDSRVDGTPKPIPAFDISKNEIVEAIDSFTSKNI FWLPKNI ELQKVDNKNIFAIDTSKWFEVETPSDLR
DIGVATI QYKIDNNSRPKVRVKLDYVIDDDSKINYFTNHSLLKSRYPDVKVLEILKQSTIIEFESSGFNKTIKEMLGM
TLAGIYNETSNN

(21). 嗜肺军团菌菌株 ATCC35289 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 30)。

MESSQILSPIGIDLGGKFTGVCLSHLEAFAELPNHANTKYSVILIDHNNFQLSQAQRRA TRHRVRNKKRNQFVKRVA
LQLFQHLISRD LNAKEETALCHYLNRRGYTYVDTDLDEYIKDETTINLLKELLPSESEHNFIDWFLQKMQSSEFRKI
LVSKVEEKDDKELKNAVKNIKNFITGF EKNSVEGHRHRKVYFENIKSDITKDNQLDSIKKKIPSVCLSNLLGHLSN
LQWKNLHRYLAKNPKQFDEQTFGNEFLRMLKNFRHLKGSQESLAVRNLIQQLEQSQDYISILEKTPPEITI PPYEAR
TNTGMEKDDQSLLLNPEKLNLYPNWRNLI PGIIDAHPFLEKDL EHTKLRDRKRI ISPSKQDEKRDSYILQRYLDL NK
KIDKFKIKKQLSFLGQKQLPANLIETQKEMETHFNSSLVSVLIQIASAYNKEREDAAQGIWFDNAFSLCELSNINP

PRKQKILPLLVGAILSEDFINNKDKWAKFKI FWNTHKIGRTSLKSKCKEIEEARKNSGNAFKIDYEEALNHPEHSNN
 KALIKIIQTI PDI IQAI QSHLGHNDSQALI YHNPFSLSQLYTI LETKRDFHKNCAVAVTCENYWRSQKTEIDPEISY
 ASRLPADSVRPFDFGLARMQRLAYEIAMAKWEQIKHIPDNSSLLIPIIYLEQNRFEFEESFKKIKGSSSDKTLEQAI
 EKQNIQWEEKFQRI INASMNICPYKGASIGGQGEIDHIYPRSLSKKHFGVIFNSEVNLIYCSSQGNREKKEEHYLLE
 HLSPLYLKHQFGTDNVDIKNFI SQNVANIKKYISFHLLTPEQQAARHALFLDYDDEAFKTIKFLMSQQKARVNG
 TQKFLGKQIMEFLSTLADSKQLQLEFSIKQITAAEVVDHRELLSKQEPKLVKSRQQSFP SHAI DATLTMSIGLKEFP
 QFSQELDNSWFINHLMPDEVHLNPVRSKEKYNKPNISSTPLFKDSL YAE RFI PVWVKGETFAIGFSEKDLFEIKPSN
 KEKLFLLKTYSTKNPGESLQELQAKSKAKWLYFPINKTLALEFLHHYFHKEIVTPDDTTVCHFINS LRYTTKKEI
 TVKILKEPMPVLSVKFESSKKNVLSFKHTIALPATKDWERLFNHPN FLALKANPAPNPKEFNEFIRKYFLSDNNPN
 SDIPNNGHNIKPKQKHKAVRKVFSLPVIPGNAGTMMRI RRDKNKGQPLYQLQTI DDT PSMGIQINEDRLVKQEVLMDA
 YKTRNLSTIDGINNSEGQAYATFDNWLTLPVSTFKPEI IKLEMKPHSKTRRYIRITQSLADFIKTIDEALMIKPSDS
 IDDPLNMPNEIVCKNKLFGNELKPRDGMKIVSTGKIVTYEFESDSTPQWIQTLYVTQLKKQP

(22). 土拉弗朗西斯菌 TIGB03 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 31)。

MNVKILPIAIDLVDKNTGVFSAFYQKGTSLKLDNKNKGVYELSKDSYTLMMNRTARRHKRRGIDRKQLVKRFLFKL
 VWTEQLNLEWDKDTQQAISFLFNRRGFSFITDGYSTEYLNIVPEQVKAILMDIFDDYNGEDDLDLSYLKLATEQESKI
 SEIYNKLMQKILEFKLRKLCTDIKDDKVSTKTLKEITSYEFELLADYLANYSES LKTQKFSYTDKQGNLKELSYHH
 DKYNIQEFLLKRHATINDEILGTLTDDFDIWNFNFEKFDKNEEKLNQEDKDHTQAHLHHFVAVNKKIKSEMASG
 GRHRSQYFQEI TNVLDENNHQEGYLKNFCENLHNKYSNLSVKNLVNLVGNLSNLELKLPRKYFNDKIHAKADHWDE
 QKFTETYCHWILGEWRVGVKDDKDGAKYSYKDLNCKELKQVTKAGLIDFLELDP CRTI PPYLDNNNRKPPKQCS
 LILNPKFLDNQYPNWQQYLQELKQLS IQDYLD SFETDLKDLKSKDQPYFVEYKSSNQMASGQRDYKDL DARI LQ
 FIFDRVKASDELLNEIYFQAKKLLKQKASSELEKLESSKLLDEVIANSQLSQILKSQHTNGIFEQGTFLHLVCKYYK
 QRQRARDSRLYIMPEYRYDKKLDKYNNTGRFDDNNQLLTYCNHKPRQKRYQLLNDLAGVLQVSRNQLLSVVEWFQQ
 AQRVGEISKSQDEQIFEWLKSFKIASYCKAAVEMQKQYRGT LKNAIQTAI FRQSENINKNKNTGNQQALS ENSKDV
 KSLTADEKLLKLIENIAKASQKIGESLGLNDKQIKKFNSIYSFAQIQQIAFAKRKGNANTCAVCSADNAHRMQQIK
 ITELVEDNKDNIILSAKAQRLPAIPTRIVDGA VKKMATILAKNIVDDNWQNIKQVLSAKHQHLHIPIITESNAFEFEP
 ALADVKGKSLKDRRKKALERISPENIFKDKNNRIKEFAKGISAYSANLTDGDFDGAKEELDHIIPRSHKKYGT LND
 EANLICVTRDDNKNI FAIDTSKWFIEITPSDLRDI GVATI QYKIDNNSRPKVRVKLDYVIDDDSKINYFMNHSLLS
 RYPDKVLEILKQSTIEFESSGFNKTIKEMLGMTLAGIYNETSNN

(23). 土拉弗朗西斯菌土拉亚种 70102010 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 32)。

MNVKILPIAIDLVDKNTGVFSAFYQKGTSLKLDNKNKGVYELSKDSYTLMMNRTARRHQRRGIDRKQLVKRFLFKL
 VWTEQLNLEWDKDTQQAISFLFNRRGFSFITDGYSTEYLNIVPEQVKAILMDIFDDYNGEDDLDLSYLKLATEQESKI
 SEIYNKLMQKILEFKLRKLCTDIKDDKVSTKTLKEITSYEFELLADYLANYSES LKIQKFSYTDKQGNLKELSYHH
 DKYNIQEFLLKRHATINDEILDTLLTDDFDIWNFNFEKFDKNEEKLNQEDKDHTQAHLHHFVAVNKKIKSEMASG
 GRHRSQYFQEI TNVLDENNHQEGYLKNFCENLHNKYSNLSVKNLVNLVGNLSNLELKLPRKYFNDKIHAKADHWDE
 QKFTETYCHWILGEWRVGVKDDKDGAKYSYKDLNCKELKQVTKAGLIDFLELDP CRTI PPYLDNNNRKPPKQCS
 LILNPKFLDNQYPNWQQYLQELKQLS IQDYLD SFETDLKDLKSKDQPYFVEYKSSNQMASGQRDYKDL DARI LQ
 FIFDRVKASDELLNEIYFQAKKLLKQKASSELEKLESSKLLDEVIANSQLSQILKSQHTNGIFEQGTFLHLVCKYYK
 QRQRARDSRLYIMPEYRYDKKLDKYNNTGRFDDNNQLLTYCNHKPRQKRYQLLNDLAGVLQVSRNQLLSVVEWFQQ
 AQRVGEISKSQDEQIFEWLKSFKIASYCKAAVEMQKQYRGT LKNAIQTAI FRQSENINKNKNTGNQQALS ENSK
 DVKSLTADEKLLKLIENIAKASQKIGESLGLNDKQIKKFNSIYSFAQIQQIAFAERKGNANTCAVCSADNAHRMQQ
 IKITELVEDNKDNIILSAKAQRLPAIPTRIVDGA VKKMATILAKNIVDDNWQNIKQVLSAKHQHLHIPIITESNAFEF
 EPALADVKGKSLKDRRKKALERISPENIFKDKNNRIKEFAKGISAYSANLTDGDFDGAKEELDHIIPRSHKKYGT LND
 NDEANLICVTRDDNKNI FAIDTSKWFIEITPSDLRDI GVATI QYKIDNNSRPKVRVKLDYVIDDDSKINYFMNHSLL
 KSRYDPKVLEILKQSTIEFESSGFNKTIKEMLGMTLAGIYNETSNN

(24). 硫磺单胞菌属物种 SCADC Cas9 蛋白 (SSCas9) 的氨基酸序列 (SEQ ID NO: 33)。

MASLISLDLGGKNTGFFSFTAKDVSIIIDTFQSGTIIYDES FVLSQVARRGKRHGKRNRLRNTLVKRLFLLLLQKHYG
 LSLDFLPDEILGLFNKRGYTYASFEIKEDEKENLES DILKEFLNDKLN YTIQNDDVEEFLNQIASNEETFKNYKDK
 FENLFGASTHQPKQIELIDEIKKDKLEKEDAKELLDGLKVIKKIIDEFHKQQNQG NLPRAKYFEELYLEI EYNLKI Q
 KFFTCNHLHINDMQYLIGNLSNYQLKELRRYFNDEAMAKEDFWSC EKLHRITWRFIQSWHPKSPEDKQRQKENLANL
 KTKSIIIEFLT TTNPLMTIPPYDDMNRGAVKQCTLRLN E EYLDKHLPHWR AIAHTLASETQKENLEGVTVKGYSEDS
 TLLHRI LDTSSIIDPYRLRSDEIDSYCDVLT KDNAFALKKFAKEYYQLVKEKVRTGIWTKDDDMFKKCDHNPPHKNN
 QIHNLVAGILGKPIAKERFEAFENELWNVKGFNKLLS SYCKNIEEFRKSNGNL FKQIVELGEDKEVQKYQKELNEWV
 RKIGEFFNIETPYRARFNLF SMAQLHTIIDTTRSGFNATCKWCSCENQYRASTRIEIDEQTGEITTNANCQRLPAD
 TQRPFSGKIERIYDKLGYEIAKVKAKELEGIKEDTIDLKII LEQNAFAYEESI RNAKIKNANAKAKKALEEAQKRGL

KNIEDKTKRIKDFSNSICPYCGQSLGEDGEIDHILSRSYTLKKYDTVFNSEGNLLYVHQKCNQAKLAKTDYSLQDLK
 IDISQKWIEEQIATIKTYKTFVLTQEQQKAFKYALFLDNSNEAYQKVISWLRTDQSSRVNGTQKYLAKKIQEKLKA
 MFPAKTFNFEFILADANDVHDLRIKAYQLPEKPKDSKQETYSHTIDAVMSLVSVWDKVLPKTEKPTKEDILKFNANVE
 NWSALNNEFLTCKGSANQKIEEMIQANDFGQKNMRQVFSKPIFKDESIGERYKPFVRYHNQFYI GYPISIKDGYDMQ
 HCQAMISKNDISRVEEILKDTSLCTLLKEKNGIKLYSINKQSINELSNQFFNLNYQNLNDAQKKKSELAEFVINHCK
 YVVKTSVINAPQFIDKDSMKPYPFYKDWQKFHEAYKKELEDAEPKTKKDNGLVYDI SGI DDFWTEFCKKYFGIKTK
 DNRNKARKVFSIVALTSAPGTVFRIKRKTPKGHIYQATAIDNQQISGDYANVLLAGNSKTLALAGQKPPSSDLKKELS
 VKESKDIRDIKLEPSRFFKEGFDCRGI EIVNKT SATIKNFPLTKIDKKIKKLI FKTLEFKKDGKRQKQKTSISLKE
 KNTMQETLKKLLKDSIKVTIRDGSI SGI EISKKT VNF TLPFKSEN LAKLLDDKRPAATKKAGQAKKKKGSLEYPYDV
 PDYAYPYDVPDYAYPYDVPDYALE

(25). 嗜肺军团菌 130b Cas9 蛋白的氨基酸序列 (SEQ ID NO: 34)。

MESSQILSPIGIDLGKFTGVCLSHLEAFAELPNHANTKYSVILIDHNNFQLSQAQRRAHRVRNKRNRQFVKRVA
 LQLFQHILSRDLNAKEETALCHYLNNRGYTYVDTDLDEYIKDETTINLLKELLPSSEHNFI DWF LQKMQSSEFRKI
 LVSKVEEKDDKELKNAVNKIKNFITGF EKNSVEGHRHRKVFENIKSDITKDNQLDSIKKKIPSVCLSNLLGHLSN
 LQWKNLHRYLAKNPKQFDEQTFGNEFLRMLKNFRHLKGSQESLAVRNLIQQLEQSQDYISILEKTPPEITIPPEAR
 TNTGMEKQSLLLNPEKLNLYPNWRNLI PGII DAHPFLEKDL EHTKLRDRKRI ISPSKQDEKRDSYILQRYLDL NK
 KIDKFKIKKQLSFLGQGKQLPANLIETQKEMETHFNSLVSVLIQIASAYNKEREDAAQGIWFDNAFSLCELSNINP
 PRKQKILPLLVGAILSEDFINNKDKWAKFKNFWNTHKIGRTSLKSKCKEIEEARKNSGNAFKIDYEEALNHPEHSNN
 KALIKIIQTI PDI IQAI QSHLGHNDSQALIYHNPFSLSQLYTI LETKRDGFHKNCVAVTCENYWRSQKTEIDPEISY
 ASRLPADSVRPF DGV LARMQRLAYEIAMAKWEQIKHIPDNSSLLIPIYLEQNRFEFEESFKKIKGGSSDKTLEQAI
 EKQNIQWEEKFQRIINASMNICPYKGASIGGQGEIDHIYPRSLSKKHFGVIFNSEANLIYCSSQGNREKKEEHYLL E
 HSLPLYLKHQFGT DNVS DIKNFI SQNVANI KKYI SFHLLTPEQQAKARHALFLDYDDEAFKTI TKFLMSQQKARVNG
 TQKFLGKQIMEFLSTLADSKQLLEFSIKQITAEVVDHRELLSKQEPKLVKSRQQSFP SHAI DATL TMSI GLKEFP
 QFSQELDNSWFINHLMFDEVHLNPNVRSKEKYNKPNISSTPLFKDSL YAERFIPVWVKGETFAIGFSEKDLFEIKPSN
 KEKLFLLKTYSTKNPGESLQELQAKSKAKWLYFPINKTLALEFLHHYFHKEIVTPDDTTVCHFINS LRYTTKKEI
 TVKILKEPMPVLSVKFESSKKNVLSFKHTIALPATKDWERLFNHPN FLALKANPAPNPKEFNEFIRKYFLSDNNPN
 SDIPNNGHNIKPQKHAVRKFVSLPVI PGNAGTMMRI RRDKNKGQPLYQLQTIDDTPSMGIQINEDRLVKQEVLM DA
 YKTRNLSTIDGINNSEGQAYATFDNWLTLPVSTFKPEI IKLEMKPHSKTRRYIRITQSLADFIKTI DEALMIKPSDS
 IDDP LNPNEIVCKNKLFGNELKPRDGKMKIVSTGKIVTYEFESDSTPQWIQTLYVTQLKKQP

(26). 土拉弗朗西斯菌土拉亚种 70001275 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 35)。

MNVKILPIAIDLGVKNTGVVSAFYQKGTSLERLDNKNKGKVEYELSKDSYTL LLMNRTARRHQRRGIDRKQLVKRFLK
 VWTEQLNLEWDKDTQQAISFLFNRRGFSFITDGYSPEYLNIVPEQVKAILMDIFDDYNGEDDLDLSYLK LATEQESKI
 SEIYNKLMQKILEFKLMLKCTDIKDDKVSTKTLKEITSYEFELLADYLANYESLKTQKFSYTDKQGNL KELSYYHH
 DKYNIQEF LKRHATINDRI LDTLLTDDLDIWNFNFEKFDKNEEKLNQEDKDHIQAHLHHFVFAVNKIKSEMASG
 GRHRSQYFQEITNVLDENNHQEGYLKNFCENLHNKKYSNLSVKNLVNLI GNLSNLELKLPRKYFNDKIHAKADHWDE
 QKFTET YCHWILGEWRVGVKDDKDGAKYSYKDL CNELKQKVTKAGLIDFLELDP CRTI PPYLDNNNRKPPKCQS
 LILNPKFLDNQYPNWQQYLQELKQLS IQDYLD SFETDLKDLKSSKDQPYFVEYKSSNQMASGQRDYKDL DARI LQ
 FIFDRVKASDELLNEIYFQAKK LKQKASSELEKLESSKKLDEVIANSQLSQILKSQHTNGIFEQGTFLHLVCKY YK
 QRQRARDSRLYIMPEYRYDKL DKYNNTGRFDDNNQLLYC NHKPRQKRYQLLNDLAGVLQVSRNQLLSSVEEWFQ
 AQRVGEISKSQDEQIFEWLKSFKIASYCKAAVEMQKQYRGT LKNAIQTAI FRQSENI NKNKNKNTGNQQALSEN SK
 DVKSLTAD EKKLLKLIENIAKASQKIGESLGLNDKQIKKFNSIYSFAQIQQIAFAERKGNANTCAVCSADNAHRMQQ
 IKITELVEDNKN ILSAKAQR LPAIPTRIVDGA VKKMATILAKNIVDDNWQNIKQVLSAKHQLHIP IITESNAFEF
 EPALADVKGKSLKDRRKALERISPENIFKDKNRIKEFAKGISAYS GANLTDGDFDGAKEELDHIIPRSHK KYGT
 NDEANLICVTRDDNKNI FAIDTSKWFEIETPSDLRDI G VATIQYKIDNNSRPKVRVKLDYVIDDSDKIN YFMNHSLL
 KSRYPDKVLEILKQSTIEFESSGFNKTIKEMLGMTLAGIYNETSNN

(27). 土拉弗朗西斯菌新凶手亚种 FTE Cas9 蛋白的氨基酸序列 (SEQ ID NO: 36)。

MNFKILPIAIDLGVKNTGVVSAFYQKGTSLERLDNKNKGKVEYELSKDSYTL LLMNRTARRHQRRGIDRKQLVKRFLK
 IWTEQLNLEWDKDTQQAISFLFNRRGFSFITDGYSPEYLNIVPEQVKAILMDIFDDYNGEDDLDLSYLK LATEQESKI
 SEIYNKLMQKILEFKLMLKCTDIKDDKVSTKTLKEITSYEFELLADYLANYESLKTQKFSYTDKQGNL KELSYYHH
 DKYNIQEF LKRHATINDRI LDTLLTDDLDIWNFNFEKFDKNEEKLNQEDKDHIQAHLHHFVFAVNKIKSEMASG
 GRHRSQYFQEITNVLDENNHQEGYLKNFCENLHNKKYSNLSVKNLVNLI GNLSNLELKLPRKYFNDKIHAKADHWDE
 QKFTET YCHWILGEWRVGVKDDKDGAKYSYKDL CNELKQKVTKAGLIDFLELDP CRTI PPYLDNNNRKPPKCQS
 LILNPKFLDNQYPNWQQYLQELKQLS IQNYLDS FETDLKVLKSSKDQPYFVEYKSSNQI ASGQRDYKDL DARI LQ
 FIFDRVKASDELLNEIYFQAKK LKQKASSELEKLESSKKLDEVIANSQLSQILKSQHTNGIFEQGTFLHLVCKY YK

QRQRARDSRLYIMPEYRYDKLHKYNNTRGFDDNQLLTYCNHKPRQKRYQLLNDLAGVLQVSPNFLKDKIGSDDDL
 FISKWLVEHIRGFKKACEDSLKIQKDNRGLLNHKNINAFNTKGKCEKEIFNLICKIEGSEDKKGNKYKHGLAYELGVL
 LFGEPNEASKPEFDRKI KKFNSIYSFAQIQQIAFAERKGNANTCAVCSADNAHRMQQIKITEPVEDNKDKIILSAKA
 QRLPAIPTRIVDGAVKMATIILAKNIVDDNWQNIKQVLSAKHQHLHPIITESNAFEFEPALADVKGKSLKDRRKKAL
 ERISPENIFKDKNNRIKEFAKGISAYSGANLTDGDFDGAKEELDHIIPRSHKKYGTLNDEANLICVTRGDNKNKGNR
 IFCLRLADNYKQKFETDDLEIEKKIADTIWDANKKDFKFGNYRSFINLTPQEQKAFRHALFLADENPIKQAVIR
 AINNRNRTFVNGTQRYFAEVLANNIYLRAKKENLNTDKISFDYFGIPTIGNGRGIAEIRQLYEKVDSDIQAYAKGDK
 PQASYSHLIDAMLAFCAIADEHRNDGSGLEIDKNYSLYPLDKNTGEVFTKDIIFSQIKITDNEFSDDKLVKKAIEG
 FNTHRQMTRDGIYAENYLPILIHKELNEVRKGYTWKNSEEIKIFKGKKYDIQQLNNLVYCLKFVDPKISIDIQISTL
 EELRNILTTNNIAATAEYYYINLKTQKLHEYYIENYNTALGYKKYSKEMEFRLSLAYRSEVVKIKSIDDVKQVLDKD
 SNFIIGKITLFPKKEWQRLYREWQNTTIKDDYEFLLKSFNVKSIITKLHKKVRKDFSLPISTNEGKFLVKKRKTWDNNF
 IYQILNDSDSRADGTKPFIPAFDISKNEIVEAII DSFTSKNIFWLPKNIELQKVDNKNI FAIDTSKWFEVETPSDLR
 DIGIATIYQKIDNNSRPKVRVKLDYVIDDDSKINYFMNHSLLKSRYPDKVLEILKQSTII EFESSGFNKTIKEMLGM
 KLAGIYNETSNN

(28). 嗜肺军团菌弗雷泽亚种菌株 ATCC 35251 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 37)。

MESSQLLSPIGIDLGGKFTGVCLSHLESFAELPNHANTKYSVILMDHNNFQLSQAQRRITRHRVRNKKRNQFVKRVA
 LQLFQHMLSRELNVKEEIALCHYLNNRGYTYVDTDLDEYLKDETTVNLKELLPPESEHNFIDWLLQKMQSSKFREI
 LGSRVEEKKDDKELKNAVKNIKNFITGFEEKNSVEGHRHRKVYFENIKSDITNDNQLDGIKKKIPSVCLSNLLGHLN
 LQWKNLHRYLAKNPKQFDEQTFGNEFLRMLKNFKHLKGSQESLAVRDLIQOLEQSKDYISILEKTPPEITIPPEEAR
 TNTGMEKDQSLLLNPEKLNLYPKWRNLIPGII DAHPFLEKDLEHTKLRDRKRIISPSKQDEKRDSYILQRYLDLKN
 KIDKFKIKKQLSFLGQKQLPANLIETHKEMETHFSSSLVSALIQIASAYNKEREDAAQGIWFDNAFSLCELSINP
 PRKQKILPLLVGAILSEDFINNKDKWAKFKSFWNTHKIGRTLLKYKCKEIEEARKNSGNAFKIDYEEALNHPEHSNN
 KALTKIQTIPDI IQAIQSHLGHNDSQLIYNNPFSLSQLYTILETKRDGFHKNCVAVTCENYWRSQKTEIDPEISY
 ASRLPADSVRPFDFGVLARMMQRLAYEIAMAKWEQIKHIPDNSSLLIPIYLEQNRFEFEESFKKIKGISRDKTLEQAI
 EKQNI RWEKQRI INASMNVCYKGTLIGGQGEIDHIYPRSLSKKHFGTIFNSEANLIYCSSQGNREKKEEHYLL
 HLSPIYLKHQFGTDNVDIKNFI SQNVANIKKYISFHLPTPEQQAARHALFLDYDDEAFKTVTKFLMSQQKARVNG
 TQKFLGKQIMEFLSTLADSKQLQLEFSIKQINAEVHNHRELLSKQEPKLAQSQQSFP SHAI DATLTMGIGLKEFP
 QFSQELDNSWFINHLMPEVHLN PVQSKEKYNKPNISSIPLFKDSLYAERFIPVWVKGETFAIGFSEKDLFEIKPSN
 KEKLFLLKTYSTKNPGESLQELQAIKAKWLYFPINKTLALEFLHHYFHKETVTPDDTTVCHFINSLRYTCKESI
 TVKILKESMPVLSVRFESSKKNVLGTFKHTIALPATKDWERLFNHPNFLALKANPNPNPKEFNEFIKKYFLSENNPN
 SDIPNNDHTIKPKHKTVRKVFLPVI PGNAGTMMRI RRDKNKGQPLYQLQTIIDTTPSMGIQINEGKLVKQEVLMDA
 YKTRNLSTIDGINNSERQAYAAFNDWLTLPVSTFKPEI IKLEMKPHSKTRRYIRITQSLADFIKTIDEALMIKPSDS
 IDDP LNMPNEIVCKNKLFGNELKPRDGMKIVSTGKTVTYEFESDSTPQWIKLYVTQLKHKP

(29). 土拉弗朗西斯菌土拉亚种 80700103 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 38)

MNVKILPIAIDLVDKNTGVSAFYQKGTSLKELDNKNGKVEYSKDSYTLMMNRTARRHQRRGIDRKQLVKRFLFKL
 VWTEQLNLEWDKDTQQAISFLNRRGFSFITDGYSTEYLNIVPEQVKAILMDIFDDYNGEDDLDLSYLKLATEQESKI
 SEIYNKLMQKILEFKLRKLCETDIKDDKVSTKLKEITSYEFELLADYLANYESLKIQKFSYTDKQGNLKELSYHH
 DKYNIQEFLEKRRHATINDEILDITLTDDFDIWNFNFEKDFDKNEEKLNQEDKDHDTQAHLHHFVFAVANKIKSEMASG
 GRHRSQYFQEITNVLDENNHQEGYLKNCENLHNNKYSNLSVKNLVNLVGNLSNLELKLPLRKYFNDKIHAKADHWDE
 QKFTETYCHWILGEWRVGVKQDKKDGAKYSYKDLCELKQKVTKAGLIDFLELDP CRTIPPYLDNNRKPKPCQS
 LILNPKFLDNQYPNWQQYLQELKQLS IQDYLDSFETDLKDLKSSKDQPYFVEYKSSNQMASGQRDYKDL DARI LQ
 FIFDRVKASDELLLENIYFQAKLQKASSELEKLESSKKLDEVIANSQLSQILKSQHTNGIFEQGTFLHLVCKYYK
 QRQRARDSRLYIMPEYRYDKLHKYNNTRGFDDNQLLTYCNHKPRQKRYQLLNDLAGVLQVSRNQLLSVVEWFQQ
 AQRVGEISKSQDEQIFEWLKSFKIASYCKAAVEMQKQYRGT LKNAIQTAI FRQSENINKNKNKNTGNQQALSNSK
 DVKSLTADEKLLKLIENIAKASQKIGESLGLNDKQIKKFNSIYSFAQIQQIAFAERKGNANTCAVCSADNAHRMQQ
 IKITELVEDNKDNIILSAKAQRLPAIPTRIVDGAVKMATIILAKNIVDDNWQNIKQVLSAKHQHLHPIITESNAFEF
 EPALADVKGKSLKDRRKKALERISPENIFKDKNNRIKEFAKGISAYSGANLTDGDFDGAKEELDHIIPRSHKKYGT
 LNDEANLICVTRDDNKNI FAIDTSKWFEIETPSDLRDI G VATIYQKIDNNSRPKVRVKLDYVIDDDSKINYFMNHSLL
 KSRYPDKVLEILKQSTII EFESSGFNKTIKEMLGMTLAGIYNETSNN

(30). 嗜肺军团菌亚种菌株 PtVF89/2014 Contig_56 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 39)。

MESSQLLSPIGIDLGGKFTGVCLSHLESFAELPNHANTKYSVILMDHNNFQLSQAQRRITRHRVRNKKRNQFVKRVA
 LQLFQHMLSRELNVKEEIALCHYLNNRGYTYVDTDLDEYLKDETTVNLKELLPPESEHNFIDWLLQKMQSSKFREI

LGSRVEEKKDDKELKNAVKNIKNFITGFEKNSVEGHRHRKVFYFENIKSDITNDNQLDGIKKKIPSVCLSNLLGHLSN
 LQWKNLHRYLAKNPKQFDEQTFGNEFLRMLKNFKHLKGSQESLAVRDLIQQLEQSKDYISILEKTPPEITIPPEAR
 TNTGMEKDQSLLLNPEKLNLYPKWRNLIPGIIDAHPFLEKDLEHTKLRDRKRLISPSKQDEKRDSYILQRYLDLNLK
 KIDKFKIKKQLSFLGQGKQLPANLIETHKEMETHFSSSLVSALIQIASAYNKEREDAAQGIWFDNAFSLCELSNINP
 PRKQKILPLLVGAILSEDFINNKDKWAKFKSFWNTHKIGRTLLKYCKEIEEARKNSGNAFKIDYEEALNHPEHSNN
 KALTKI IQTI PDI IQAIQSHLGHNDSQALIYNNPFSLSQLYTILETKRDGFHKNCVAVTCENYWRSQKTEIDPEISY
 ASRLPADSVRPFDFGLARMMQRLAYEIAMAKWEQIKHIPDNSSLLIPIYLEQNRFEFEESFKKIKGISRDKTLEQAI
 EKQNI RWEKQFQRI INASMNVC PYKGT LIGGQGEIDHIYPRSLSKKHFGTIFNSEANLIYCSSQGNREKKEEHYLL
 HLSPIYLKHQFGTDNVDIKNFISQNVANIKKYISFHLLTPEQQKAARHALFLDYDDEAFKTVT KFLMSQQKARVNG
 TQKFLGKQIMEFLSTLADSKQLQLEFSIKQINAEVHNHRELLSKQEPKLAQSQQSFPSHAIDATLTMGIGLKEFP
 QFSQELDNSWFINHMLPDEVHLN PVQSKEKYNKPNISSIPLFKDSLAEYERFIPVWVKGETFAIGFSEKDLFEIKPSN
 KEKLFLLKTYSTKNPGESLQELQAI SKAKWLYFPINKTLALEFLHHYFHKETVTPDDTTVCHFINSLRYTTKKEI
 TVKILKESMPVLSVRFESSKKNVLGTFKHTIALPATKDWERLFNHPN FLALKANPNPNPKEFNEFIKKYFLSENNPN
 SDIPNNDHTIKPKQKHTVRKVFSLPVI PGNAGTMMRI RRDKNKGQPLYQLQTI DDTPSMGIQINEGKLVKQEVLMDA
 YKTRNLSTIDGINNSERQAYAAFNDWLTLPVSTFKPEIKKLEMKPHSKTRRYIRITQSLADFIKTIDEALMIKPSDS
 IDDP LNMPNEIVCKNKLFGNELKPRDGKMKIVSTGKTVTYEFESDSTPQWIQKLYVTQLKKHP

(31). 嗜肺军团菌菌株 HO4020049 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 40)。

MESSQILSPIGIDLGGKFTGVCLSHLEAFELPNHANTKYSVILIDHNNFQLSQAQRRA TRHRVRNKKRNQFVKRVA
 LQLFQHILSRDLNAKEETALCHYLNNRGTYVDTDLDEYIKDETTINLLKELLPSESEHNFI DWFLQKMQSSSEFRKI
 LVSKVEEKKDDKELKNAVKNIKNFITGFEKNSVEGHRHRKVFYFENIKSDITKDNQLDSIKKKIPSVCLSNLLGHLSN
 LQWKNLHRYLAKNPKQFDEQTFGNEFLRMLKNFKHLKGSQEF LAVRDLIQQLEQSKDYISILEKTPPEITIPPEAR
 TNTGMEKDQSLLLNPEKLNLYPNWRNLMPGIIDAHPFLEKDLEHTKLRDRKRLISPSKQDEKRDSYILQRYLDLNLK
 KIDKFKIKKQLSFLGQGKQLPANLIETQKEMETHFNSSLVSVLIQIASAYNKEREDAAQGIWFDNAFSLCELSNINP
 PRKQKILPLLVGAILSEDFINNKDKWAKFKDFWNTHKIGRTSLKSKCKEIEEARKNSGNAFKIDYEEALNHPEHSNN
 KTLIKI IQTI PDI IQAIQSHLGHNDSQALIYHNPFSLSQLYTILETKRDGFHKNCVAVTCENYWRSQKTEIDPEISY
 ASRLPADSVRPFDFGLARMMQRLAYEIAMAKWEQIKHIPDNSSLLIPIYLEQNRFEFEESFKKIKGSSSDKTLEQAI
 EKQNIQWEEKFQRI INASMNICPYKGASIGGQGEIDHIYPRSLSKKHFGVIFNSEANLIYCSSQGNREKKEEHYLL
 HLSPLYLKHQFGTDNVDIKNFISQNVANIKKYISFHLLTPEQQKASRHALFLDYDDEAFKITKFLMSQQKARVNG
 TQKFLGKQIMEFLSTLADSKQLQLEFSIKQITAEVVDHRELLSKQEPKLVKSRQSFPSHAIDATLTMISIGLKEFP
 QFSQELDNSWFINHMLPDEVHLN PVRSKEKYNKPNISSIPLFKDSLAEYERFIPVWVKGETFAIGFSEKDLFEIKPSN
 KEKLFLLKTYSTKNPGESLQELQAKSKAKWLYFPINKTLALEFLHHYFHKEIVTPDDTTVCHFINSLRYTTKKEI
 TVKILKEPMPVLSVKFESSKKNVLGTFKHTIALPATKDWERLFNHPN FLALKANPAPNPKEFNEFIRKYFLSDNNPN
 SDIPNNGHNIKPKQKHAVRKVFSLPVI PGNAGTMMRI RRDKNKGQPLYQLQTI DDTPSMGIQINEDRLVKQEVLMDA
 YKTRNLSTIDGINNSEGQAYATFDNWLTPVSTFKPEIKKLEMKPHSKTRRYIRITQSLADFIKTIDEALMIKPSDS
 IDDP LNMPNEIVCKNKLFGNELKPRDGKMKIVSTGKIVTYEFESDSTPQWIQTLVVTQLKKHP

(32). 嗜肺军团菌亚种菌株 PtVF89/2014 Contig_45 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 41)。

MESSQLLSPIGIDLGGKFTGVCLSHLESFAELPNHANTKYSVILMDHNNFQLSQAQRRI TRHRVRNKKRNQFVKRVA
 LQLFQHMLSRRELVKKEIALCHYLNNRGTYVDTDLDEYIKDETTVNLLKELLPPESEHNFI DWLLQKMQSSSKFREI
 LGSRVEEKKDDKELKNAVKNIKNFITGFEKNSVEGHRHRKVFYFENIKSDITNDNQLDGIKKKIPSVCLSNLLGHLSN
 LQWKNLHRYLAKNPKQFDEQTFGNEFLRMLKNFKHLKGSQESLAVRDLIQQLEQSKDYISILEKTPPEITIPPEAR
 TNTGMEKDQSLLLNPEKLNLYPKWRNLIPGIIDAHPFLEKDLEHTKLRDRKRLISPSKQDEKRDSYILQRYLDLNLK
 KIDKFKIKKQLSFLGQGKQLPANLIETHKEMETHFSSSLVSALIQIASAYNKEREDAAQGIWFDNAFSLCELSNINP
 PRKQKILPLLVGAILSEDFINNKDKWAKFKSFWNTHKIGRTLLKYCKEIEEARKNSGNAFKIDYEEALNHPEHSNN
 KALTKI IQTI PDI IQAIQSHLGHNDSQALIYNNPFSLSQLYTILETKRDGFHKNCVAVTCENYWRSQKTEIDPEISY
 ASRLPADSVRPFDFGLARMMQRLAYEIAMAKWEQIKHIPDNSSLLIPIYLEQNRFEFEESFKKIKGISRDKTLEQAI
 EKQNI RWEKQFQRI INASMNVC PYKGT LIGGQGEIDHIYPRSLSKKHFGTIFNSEANLIYCSSQGNREKKEEHYLL
 HLSPIYLKHQFGTDNVDIKNFISQNVANIKKYISFHLLTPEQQKAARHALFLDYDDEAFKTVT KFLMSQQKARVNG
 TQKFLGKQIMEFLSTLADSKQLQLEFSIKQINAEVHNHRELLSKQEPKLAQSQQSFPSHAIDATLTMGIGLKEFP
 QFSQELDNSWFINHMLPDEVHLN PVQSKEKYNKPNISSIPLFKDSLAEYERFIPVWVKGETFAIGFSEKDLFEIKPSN
 KEKLFLLKTYSTKNPGESLQELQAI SKAKWLYFPINKTLALEFLHHYFHKETVTPDDTTVCHFINSLRYTTKKEI
 TVKILKESMPVLSVRFESSKKNVLGTFKHTIALPATKDWERLFNHPN FLALKANPNPNPKEFNEFIKKYFLSENNPN
 SDIPNNDHTIKPKQKHTVRKVFSLPVI PGNAGTMMRI RRDKNKGQPLYQLQTI DDTPSMGIQINEGKLVKQEVLMDA
 YKTRNLSTIDGINNSERQAYAAFNDWLTLPVSTFKPEIKKLEMKPHSKTRRYIRITQSLADFIKTIDEALMIKPSDS
 IDDP LNMPNEIVCKNKLFGNELKPRDGKMKIVSTGKTVTYEFESDSTPQWIQKLYVTQLKKHP

(33). 新凶手弗朗西斯菌 Cas9 蛋白 (FnCas9) 的氨基酸序列 (SEQ ID NO: 42)。

MNFKILPIAIDLGVKNTGVFSAFYQKGTSLERLDNKNKGVYELSKDSYTLMMNRTARRHQRRGIDRKQLVKRFLFKL
 IWTEQLNLEWDKDTQQAISFLFNRRGFSFITDGYSPYELNIVPEQVKAILMDIFDDYNGEDDLDLSYLKLATEQESKI
 SEIYNKLMQKILEFKLMLKCTDIKDDKVSTKTLKEITSYEFELLADYLANYSESLKTQKFSYTDKQGNLKELSYYHH
 DKYNIQEFLLKRHATINDRILDTLLTDDLDIWNFNFEKFDKNEEKLQNEQEDKDHIIQAHLHHFVFAVANKIKSEMASG
 GRHRSQYFQIEITNVLDENNHQEGYLKNFCENLHNKYSNLSVKNLVNLIIGNLSNLELKLPLRKYFNDKIHAKADHWDE
 QKFTETECHWILGEWRVGVKQDKKDGAKYSYKDLNCELKQKVTKAGLVDLLELDPCRTIPPYLDNNNRKPKCQS
 LILNPKFLDNQYPNWQQYLQELKQLS IQNYLDS FETDLKVLKSKDQPYFVEYKSSNQOIASGQRDYKDLDARI LQ
 FIFDRVKASDELLLEIYFQAKKQKASSELEKLESSKLLDEVIANSQLSQILKSQHTNGIFEQGTFLHLVCKYYK
 QRQRARDSRLYIMPEYRYDKKLHKYNNNTGRFDDDNQLLTYCNHKPRQKRYQLLNDLAGVLQVSPNFLKDKIGSDDDL
 FISKWLVEHIRGFKKACEDSLKIQKDNRGLLNHNKIN IARNTKKGCEKEIFNLI CKIEGSEDKKGNKYKHLAYELGVL
 LFGEPNEASKPEFDRKIKKFNISYFSAQIQQIAFAERKGNANTCAVCSADNAHRMQQIKITEPVEDNKDKIILSAKA
 QRLPAIPTRIVDGA VKKMATILAKNIVDDNWQNIKQVLSAKHQ LHIPIITESNAFEFEPALADVKGKSLKDRRKKAL
 ERISPENIFKDKNNRIKEFAKGISAYSGANLTDGDFDGAKEELDHIIPRSHKKYGTLNDEANLIVTRGDNKNKGNR
 IFCLRD LADNYK LKQFETDDLEIEKKIADTIWDANKKDFKFNYSRFINLTPQEQKAFRHALFLADENPIKQAVIR
 AINNRNRTFVNGTQRYFAEVLANNIYLRAKKENLNTDKISFDYFGIPTIGNRGIAEIRQLYEKVDSDIQAYAKGDK
 PQASYSHLIDAMLAFCIAA DEHRNDGSI GLEIDKNYSLYPLDKNTGEVFTKDI FSQIKITDNEFSDKKLVRKKAIEG
 FNTHRQMTRDGIYAENYLPILIHKELNEVRKGYTWKNSEEIKIFKGGKYDIQQLNNLVYCLKFDVKPISIDIQISTL
 EELRNILTTNIIAATAEYIIYINLKTQKLHEYYIENYNTALGYKKYSKEMEF LRS LAYRSERVKIKSIDDVQVLDKD
 SNFIIGKITL PFKKEWQRLYREWQNTTIKDDYEF LKSFNVKSI TKLHKKVRKDFSLPISTNEGKFLV KRKTWDNNF
 IYQILNDSDSRVDGTFKPIPAFDISKNEIVEAII DSFTSKNIFWLPKNI ELQKVDNKNIFAIDTSKWFVETPSDLR
 DIGVATI QYKIDNNSRPKVRVKLDYVIDDDSKINYFTNHSLLKSRYPDKVLEILKQSTTIEFESSGFNKTIKEMLG
 TLAGIYNETSNN

(34). 人粪便副萨特氏菌 Cas9 蛋白 (PcCas9) 的氨基酸序列 (SEQ ID NO: 43)。

MGKTHIIGVGLDLGGTYTGTFTSHPSDEAEHRDHSSAFTVVNSEKLSFSSSKSRTAVRHRVRSYKGFDLRRRLLLLV
 AEYQLLQKKQTLAPEERENLRIALSGYLKRRGYARTEAETDVSLES LDP SVFSSAPSFTNFFNDSEPLNIQWEAIA
 NSPETTKALNKELSGQKEADFKKYIKTSFPEYSAKEILAN YVEGRRAILDASKYIANLQSLGHKHSKYLSDI LQDM
 KRDSRITRLSEAFGSTDNLWRIIGNISNLQERAVRWYFNDAKFEQGGQQLDAVKLKNVLRALKYLRSDDEKWSASQ
 KQIIQSLEQSGDVLVDVLAGLDPDRITIPPYEDQNNRRPPEDQTLYLNPKALSSEYGEKWKSWANKFAGAYPLLTEDLT
 EILKNTDRKSRIKIRSDVLPDS DYRLAYILQRAFDRSIALDECSI RRTAEDFENG VVIKNEKLEDVLSGHQLEEFLE
 FANRYQETAKAKNGLWFPENALLERADLHPMPKNKILNVI VGOALGVSPAEGTDFIEEIWNSKVKGRSTVRSICNA
 IENERKTYGPIYFSEDYKFKVTALKEGKTEKELSKKFAAVIKVLMVSEVVPF IGKELRLSDEAQS KFDNLYSLAQLY
 NLIETERNGFSKVS LAAHLENARWMTMTDGS AQCCRLPADCVRPFDGFIRKAI DRNSWEVAKRIAEVKKSVDF TNG
 TVKIPVAIEANSFNFTASLTDLKYIQLKEQKLKLEDIQRNEENQEKRWLSKEERI RADSHGICAYTGRPLDDVGE
 IDHII PRSLTLKKSESIYNSEVNLI FVSAQGNQEKNNIYLLSNLAKNYLAAVFGTSDLSQITNEIESTVLQ LKAAG
 RLG YFDLLSEKERACARHALFLNSDSEARRAVIDVLGSRKASVNGTQAWFVRSIFSKVRQALAAWTQETGNELIFD
 AISVPAADSSEMRKRFAEYRPEFRKPKVQPVASHSIDAMCIYLAACSDPFKTKRMGSQ LAIYEPINF DNLTGSCQV
 IQNTPRNFSDKTNIANSPIFKETIYAERFLDIIVSRGEIFIGYPSNMPFEKPNRISIGGKDPF S ILSVLGAYLDKA
 PSSEKEKLTIRVVKNKAFELFSKVAGSKFTAEDKAAKILEALHFVTVKQDVAATVSDLIKSKKELSKDSIENLAK
 QKGLCKKVEYSSKEFKFKGSLIIPAAVEWGKVLWNVFKENTAEELK DENALRKALEAAWPS SFGTRNLHSAKRVFS
 LPVVATQSGAVRIRRKTAFGDFVYQSQDTNNLYSSFPVKNGKLDWSSPIIHPALQNRNLTAYGYRFVDHRSISMSE
 FREVYNKDDL MRIELAQGTSSRRYL RVEMPGEKFLAWFGENSISLGS SFKFSVSEVFDNKIYTENA EFTKFLPKPRE
 DNKHNGTIF FELVGRVIFNYIVGGAASSLKEIFSEAGKERS

(35). 瘤胃杆菌属物种 RM87 Cas9 蛋白 (RCas9) 的氨基酸序列 (SEQ ID NO 44)。

MGGKYTGFI SYTSDNLPDSENVKAAIIE MPDDGSVINYTVKNRTAVRHHLRANNRYKKARKLI FTILADTI KRNL TQ
 SEQEAISLMLRRGYTRLETEIDLEQLKECPVDLFFQCCPGLFTGDVPLYDQFINNCNDDTVVYVRESLSSIKETVS
 NLSDKTEKSMYGSALKTMVDFVDKFI DQKEFGHKHRT EYLANIRKDI QKDSRLNDLIK LISGEERLYNCIGNISNLQ
 LRALRWYFNDRGMGSKAVFDKFRKFDVWYRAYKYFHYTAGSNQKDNAEKVDIIKKNEFLKRIEQTENITGLLLAQD
 PVFTIPPYEDQNNRRPPVDQTL LLS PQSLDAHYPGWELWVQKFM DYGSGSYSSLSDELEQIVKLTDRSSRLQSAGG
 KQSDHIASYTEQKLIHSYYLQRLLDIT YNRGNPESQIRKWAQNPNSCSATNAVIKNILGESDIDSFFELANNYYMEV
 SLSKGLWSIVEHPLLEISGIHPPMKSKMLEELVAGVLNIRKGFNLERFKEVWNSTVKGKSTVRSLSYSIEKTRKKY
 GNCFKQYKNCRYEELKNKASEVHFSSDIKIDKEVVKVYEYTKNLSKFI GEKLELSTDQSKFANPFSLAQLYTIIL
 ENDVHGFSGNCKAVCLENTCRMQMSDKGNALCCRLPAESTRPF DGS LGNILD RQAYEIAKIKIAEIKELSALKNTKV
 KLSVLVEENSFEFSSSLSTIKKSLKAKQLREFALKAEKTEREKWKS KDERLKNDSHNICAYTGMALSDDLSENDHIV

SRSETKESYGTIYNSEFNLI RVSRRGNQLKGNNI YTLDNLNP KYLEAVFGTSDTAI ITTRI RETLNRLLI SNP NLNI
 DIMTEEERQCCRHALFTSRTGVPFNQI IQAIS KQYRVRVNGTQAWFVKNLI AKI DEGLREWKS INNITVEYNAYKID
 AGSTSEIRHDIGKLDARYEKKDI QPITSHAI DAVCLLGNASTQQEISEEISGNNEISVLSVVKELIKLIPDRFDIIR
 ISSLNFAEKNNPESKQLFKDTIYAEHFLCIMEKDDTVKVGDFDGSNSIQIVKGGKQLLSCLANYLES PKQNETTGR
 TYKVCKTAAFKL FERFYRGLVAEGSDEELAYLALMSLR YTTTLHQDVFSLLYDVKSKKFAAKKDI LKNLDIKI SFPSV
 LKLTDKASLVLPSFNEWAH IASIFEDYLGKKDEQESGYDKLQYLGTILNKTINKNHAKSKRVYSLPVVASPSGGV
 RIKRLTHTHGDIYQLVAANTPVTSISKGFASADNGDVLWDEAVMVDSYTTKNLTVINKKISGDKNFVSMDEQRLVYE
 SPDERVYMT PATDIRRGITIEQKFEKFVYCIDQQNIYKSFHDLPT EIKTDPKIFMERLNI DICGKPRGNLKI LSIGN
 IIKYSYMVESNRVMNQVYNNPIDKGTIFKSN

(36). 新凶手弗朗西斯菌近似种 Fx1 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 45)。

MNFKILPIAIDLGVKNTGVFSAFYQKGTSLERLDNKNKGVYELSKDSYTL LLMNRTARRHQRRGI DRKQLVKRFLKFL
 IWTEQLNLEWDKDTQQAISFLFNRRGFSFITDGYSP EYLNIVPEQVKAILMDIFDDYNGEDDLD SYLKLATEQESKI
 SEIYNKLMQKILEFKLRKLCTDIKDDKVSTKTLKEITSYEFELLADYLANYSEILKTQKFSYTDKQGNLKELSYHH
 DKYNIQEF LKRHATINI QILDTLLTDDLDIWNFNFEKFD FDKNEEKLQNEQEDKDHIQAHLHFFVFAVNKIKSEMASG
 GRHRSQYFQ EITNVLDENNHQEGYLKNFCENLHNKYSNLSVKNLVNLI GNLSNLELKLPLRKYFNDKI HAKADHWDE
 QKFTET YCHWILGEWRVGVKQDKKDGAKYSYKDL CNELKQKVT KAGLVD FLELDP CRTI PPYLDNNNRKPPKCQS
 LILNPKFLDNQYPNWQQYLQELKQLS IQNYLESFETDLKVLKSKDQPYFVEYRSSNQQIASGQRDYKDL DARI LQ
 FIFDRVKASDELLNEIYFQAKK LKQKASSELEKLES SKKLDEVIANSQLSQILKSQHTNGIFEQGTFLHLVCKYYK
 QRQRARDSRLYIMPEYRYDKKLHKYNN TGRFDDNQ LLYTCNHKPRQKRYQLLNDLAGVLQVSPNFLKDKIGSDDDL
 FISKWLVEHIRGFKKACEDSLKIQKDNRGLLNHKN IARNTK GKCEKEIFNLICKIEGSEDKKNYKHLAYELGVL
 LFGEPNEASKPEFDRKIKKFNSIYSFAQIQQIAFAERKGNANTCAVCSADNAHRMQQIKITEPVEDNKDKIILSAKA
 QRLPAITTRIVDGA VKKMATILAKNIVDDNWQNIKQVLSAKHQ LHIPIITESNAFEFEPALADVKGKSLKDRRKKAL
 ERISPENIFKDKNNRIKEFAKGISAYSGANLTDGDFDGAKEELDHIIPRSHKKYGT LND EANLICVTRGDNKNKGNR
 IFCLRDLANNYK LKQFETTDLEIEKKIADTIWDANKKDFKFGNYRSFINLTPQEQA FRHALFLADENPIKQAVIR
 AINNRRRTFVNGTQRYFAEVLANNIYLRAKKENLNTDKISFDYFGIPTIGNGRGIAEIRQLYEKVDSDIQAYAKGDK
 PQASYSHLIDAMLAF CIAA DEHRNDGSIGLEIDKNYSLYPLDKNTGEVFTKDI FSQIKITDNEFS DKKLVRK KATEG
 FNTHRQMT RDGIYAENYLPILIHKELDEVRKGYTWKNSEI KIFKGKKYDIQQLNLLYCLKFVDKPI SIDIQISTL
 EELRNILTTNNIAATAEY YINLKTQKLHEYYIENYNTALGYK KYTKEMEFLRS LAYRSERVKIKSIDDVKQVLDKD
 SNFIIGKITL PFKKEWQRLYREWQNI TIKDDYEF LKSFFNVKSI TKLHKKVRKDFSLPISTNEGKFLV KRKTW DNNF
 IYQILNDSDSRVDGTPFIPAFDISKNEIVEAII DSFTSKNI FWLPKNI ELQKVDNKNIFAIDTSKWFEVETPSDLR
 DIGVATIYQKIDNNSRPKVRVKLDYVIDDDSKIN YFTNHSLLKS RYPDKVLEILKQSTIIEFESSGFNKTIKEM LGM
 TLAGIYNETSNN

(37). 土拉弗朗西斯菌土拉亚种 8070075 contig00018 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 46)。

MNVKILPIAIDLGVKNTGVFSAFYQKGTSLERLDNKNKGVYELSKDSYTL LLMNRTARRHKRRGI DRKQLVKRFLKFL
 VWTEQLNLEWDKDTQQAISFLFNRRGFSFITDGYSTEYLNIVPEQVKAILMDIFDDYNGEDDLD SYLKLATEQESKI
 SEIYNKLMQKILEFKLRKLCTDIKDDKVSTKTLKEITSYEFELLADYLANYSESLKTQKFSYTDKQGNLKELSYHH
 DKYNIQEF LKRHATINDEILGTLLTDDFDIWNFNFEKFD FDKNEEKLQNEQEDKDHTQAHLHFFVFNKIKSEMASG
 GRHRSQYFQ EITNVLDENNHQEGYLKNFCENLHNKYSNLSVKNLVNLVGNLSNLELKLPLRKYFNDKI HAKADHWDE
 QKFTET YCHWILGEWRVGVKQDKKDGAKYSYKDL CNELKQKVT KAGLID FLELDP CRTI PPYLDNNNRKPPKCQS
 LILNPKFLDNQYPNWQQYLQELKQLS IQDYLD SFETDLKDLKSKDQPYFVEYKSSNQMASGQRDYKDL DARI LQ
 FIFDRVKASDELLNEIYFQAKK LKQKASSELEKLES SKKLDEVIANSQLSQILKSQHTNGIFEQGTFLHLVCKYYK
 QRQRARDSRLYIMPEYRYDKKLDKYNNTGRFDDNNQ LLYTCNHKPRQKRYQLLNDLAGVLQVSRNQLLS SVEEFQ
 AQRVGEISKSQDEQIFEWLKSFKIASYCKAAVEMQKQYRGT LKNAIQTAI FRQSENINKNKNTGNQQALS ENSKDV
 KSLTADEKLLKLIENI AKASQKIGESLGLNDKQIKKFNSIYSFAQIQQIAFAKRKGNANTCAVCSADNAHRMQQIK
 ITELVEDNKDNIILSAKAQRLPAIPTRIVDGA VKKMATILAKNIVDDNWQNIKQVLSAKHQ LHIPIITESNAFEFEP
 ALADVKGKSLKDRRKKALERISPENIFKDKNNRIKEFAKGISAYSGANLTDGDFDGAKEELDHIIPRSHKKYGT LND
 EANLICVTRDDNKNIFAIDTSKWFEIETPSDLR DIGVATIYQKIDNNSRPKVRVKLDYVIDDDSKIN YFMNHSLLKS
 RYPDKVLEILKQSTIIEFESSGFNKTIKEM LGM TLAGIYNETSNN

(38). 土拉弗朗西斯菌新凶手亚种 F6168 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 47)。

MNVKILPIAIDLGVKNTGVFSAFYQKGTSLERLDNKNKGVYELSKDSYTL LLMNRTARRHQRRGI DRKQLVKRFLKFL
 IWTEQLNLEWDKDTQQAISFLFNRRGFSFITDGYSP EYLNIVPEQVKAILMDIFDDYNGEDDLD SYLKLATEQESKI
 SEIYNKLMQKILEFKLRKLCTDIKDDKVSTKTLKEITSYEFELLADYLANYSEILKTQKFSYTDKQGNLKELSYHH
 DKYNIQEF LKRHATINDEILD TLLTDDLDIWNFNFEKFD FDKNEEKLQNEQEDKDHTQAHLHFFVFAVNKIKSEMASG

GRHRGQYFQEITNVLDENNHQEGYLKNCENLHNKYSNLSVKNLVNLIIGNLSNLELKPLRKYFNDKI HAKADHWDE
 QKFTETYCHWILGEWRVGVKDDKDGAKYSYKDLNCLKQKVTAGLVDFLLELDPCRTIPFYLDNNNRKPPKQCS
 LILNPKFLDNQYPNWQQYLQELKKLQSIQNYLESFETDLKVLKSSKDPYFVEYKSSNQI ASGQRDYKDL DARI LQ
 FIFDRVKASDELLNEIYFQAKKLLKQKASSELEKLESSKLLDEVIANSQLSQILKSQHTNGIFEQGTFLHLVCKYYK
 QRQRARDSRLYIMPEYRYDKKLDKYNNTGRFDDDNQLLYTCNHKPRQKRYQLLNDLAGVLQVSPNFKDKIGSDDDL
 FISKWLVEHIRGFKKACEDSLKIQKDNRGLLNHKINIARNTKKGCEKEIFNLICKIEGSEDKKGNYPKHGLAYELGVL
 LFGEPNEASKPEFDRKIKKFNSIYSAQIQQIAFAERKGNANTCAVCSADNAHRMQQIKITEPVEDNKDKIILSAKA
 QRLPAIPTRIVDGA VVKMATILAKNIVDDNWQNIKQVLSAKHQHLHIPIITESNAFEFEPALADVKGKSIKDRRKKAL
 ERISPENIFKDKNNRIKEFAKGISAYS GDNLASGDFDGAKEELDHIIPRSHKKYGTLNDEANLIVTRDDNKKNKGNR
 IFCLHDLANNYKQKQFETDDLEIEKKIADTIWDASKKDFKFGNYRSFINLTPQEQKAFRHALFLADENPIKQAVIR
 AINNRNRTFVNGTQRYFAEVLANNIYLRAKKENLNTDKISFDYFGIPTIGNGRGIVEIRQLYEKVDSDIQAYAKGDK
 PQASYSHLIDAMLAF CIAADEHRNDGSIGLEIDKNYSLYPLDKNTGEVFTKDI FSQIKITNNEFSDKKLVKRRKATEG
 FNTHRQMT RDGIYAENYLPILIHKELDEV RKGYTWGNSEELKIHKGKKYDIQQLNLVYCLKFDKISISIDIQISTL
 EELRNI LTTNNIATTA EYYYINLKTQKLHEYYIENYNTALGYKKYTKEMEFLRS LAYRSE RVKIKSIDDVQVLDKD
 SNFIIGKITLFPFKKEWQRLYREWQNTTIKDDYEF LKSF FNVKSIKHLKHKVRKDFSLPISTNEGKFLV KRKTWDNNF
 IYQILNDSDSRVDGTPKPFIPAFDISKNEIVEAII DSFTSKNIFWLPKNI ELQKVDNKNIFAIDTSKWFEVETPSDLR
 DIGVATI QYKIDNNSRPKVRV KLDYVIDDDSKINYFTNHSLLKSRYDPK VLEILKQSTIEFESSGFNKTIKEM LGM
 TLAGIYNETSNN

(39). 土拉弗朗西斯菌全北美区亚种 257 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 48)。
 LLNDLAGVLQVSRNQLLSSVEEWFQQAQRVGEISKSQDEQIFEWLKSFKIASYCKAAVEMQKQYRGT LKNAIQTAF
 RQSENINKNKNKNTGNQQALS ENSKDVKSLTADEKLLKLIENIAKASQKIGESLGLNDKQIKKFNSIYNFAQI
 QQIAFAERKGNANTCAVYSDNAHRMQQIKITELVEDNKDNIILSAKAQRLPAIPTRIVDGA VVKMATILAKNIVDD
 NWQNIKQVLSAKHQHLHIPIITESNAFEFEPALADVKGKSLKDRRKKALERISPENIFKDKNNRIKEFAKGISAYS
 GANLTDGDFDGAKEELDHIIPRSYKKGTLNDEVNLI CVTRDDNKKNKGRIFCLNDLANNYKQKQFETDDLEIEKKIA
 DTIWDASKQDFKFGNYRSFINLTPQEQKAFRYALFLADENS IKQAVIRAINNRNRTFVNGTQRYFAEVLANNIYLRA
 KKENLNTDKISFDYFGIPTIGNGRGIADIRQLYEKVDSDIQAYAKGDKPQASYSHLIDAMIAFCVAAD EHRNDGSIG
 LEIDKNYSLYPLDKNTGEVFTKDI FSQIKITDNEFSDKKLVKRRKATEGFNTHRQMT RDGIYAENYLPILIHKELDEV
 RKGYTWGNSEELKIHKGKKNYIQQLNIVYCLKFDKISISIDIQISTLEELRNI LTTNNIATAEYYYINLKNQKLH
 EYYIENYNTALDYKYSKEMEFLRS LAYRSE RVKIKSIDDVQVLDKDSNFIIGKITLFPFKKEWQRLYLEWQNTTIK
 DDYEF LKSF FNVKSIKHLKHKVRKDFSLPISTNEGKFLV KRKTWDNNFIYQILNDSDSRVDGTPKPFIPAFDISKNEI
 VEAII DSFTSKNIFWLPKNI ELQKVDKIFLL

(40). 伦敦军团菌菌株 ATCC 49505 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 49)。
 MESSQILSPIGIDLGGKFTGVCLSHLEAFAELPNHANTKYSVILIDHNNFQLSQAQRRA TRHRVRNKRKRNQFVKRVA
 LQLFQHLISRDLNAKEETALCHYLNRRGYTYVATDLDEYIKDETTINLLKELLPSESEHNFI DWFLQKMQSSEFRKI
 LVSKVEEKDDKELKNAVKNIKNFITGFEKNSVEGHRHRKVFENIKSDITKDNQLDSIKKKIPSVCLSNLLGHLSN
 LQWKNLHRYLAKNPKQFDEQTFGNEFLRMLKFNFRHLKGSQESLAIRNLIQQLEQSQDYISILEKTPPEITIPPEAR
 TNTGMEKDQSLLLNPEKLNLYPNWRNLI PGII DAHPFLEKDLEHTKLRDRKRIISPSKQDEKRDSYILQRYLDL NK
 KIDKFKIKKQLSFLGQGKLPANLIETQKEMETHFNSSLVSVLIQIASAYNKEREDAAQGIWFDNAFR LCELSNINP
 PRKQKILP LLVGA ILS EDFINNKDKWAKFKNFWNTHKIGRTSLKSKCKEIEEARKNSGNAFKIDYEEALNHPEHSNN
 KALIKI IQTI PDI IQAI QSHLGHNDSQALYHNPFSLSQLYTIETKRDGFHKNCVAVTCENYWR SQTEIDPEISY
 ASRLPADSVRPF DGV LARMMQRLAYEIAMAKWEQIKH I PDNSSLLIPIYLEQNRFEFEESFKKIKGSSDKTLEQAI
 EKQNIQWEEKFQRIINASMNICPYKGT SIGGQGEIDHIYPRSLSKKHFGVIFNSEANLIYCSSQGNREKKEEHYLL E
 HLSPLYLKHQFGTDNVSDIKNFISQNVANIKKYISFHLLTPEQQKAARHALFLDYDDEAFKITKFLMSQQKARVNG
 TQKFLGKQIIEFLSTLADSKQLQLEFSIKQITAEVVDHRELLSKQEPKLVKNRQSFPSHAIDATLTMSIGLKEFP
 QFSQELDNSWFINHLMPEVHNLN PVRSEKYNKPNISSTPLFKDSL YAERFIPVWVKGETFAIGFSEKDLFEIKPSN
 KEKLF TLLKTYSTKNPGESLQELQAKSKAKWLYFPINKTLALEFLHHYFHKEIVTPDDTIVCHFINS LRYTYKKE SI
 TVKILKEPMPVLSVKFESSKNVLSFKHTIALPATKDWERLFNHPN FLALKANPAPNPKEFNEFIRKYFLSGNNPN
 SDISNNGNMIKPKKHAVR KVFSLPVI PGNAGTMMRI RRDKNKGQPLYQLQTIDDTPSMGIQINEDRLVKQEVLM DA
 YKTRNLSTIDGINNSEGQAYATFDNWL TLPVTTFKPEI IKLEMKPHSKTRRYIRITQSLADFIKTIDEALMIKPSDS
 IDDP LNMPNEIVCKNKLFGNELKPRDGKMKIVSTGKIVTYEFESDSTPQWIQTLYVTQLKKQP

(41). 土拉弗朗西斯菌土拉亚种 NE061598 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 50)。
 MNVKILPIAIDL DVKNTGVFSAFYQKGTSLKELDNKNGKVYELSKDSYTL LMNNRTARRHKRRGIDRKQLVKRFLK L
 VWTEQLNLEWDKDTQQAISFLFNRRGFSFITDGYSTEYLNIVPEQVKAILMDIFDDYNGEDDLDLSYLK LATEQESKI
 SEIYNKLMQKILEFKLRKLC TDIKDDKVSTKTLKEITSYEFELLADYLANYESLKTQKFSYTDKQGNLKELSYHH

DKYNIQEFLLKRHATINDEILGTLTDDFDIWNFNFEKFDKNEEKLQNEQEDKDHTQAHLHHFVFNVNIKSEMASG
 GRHRSQYFQIEITNVLDENNHQEGYLKNCENLHNKYSNLSVKNLVNLVGNLSNLELKLPLRKYFNDKIHAKADHWDE
 QKFTETTYCHWILGEWRVGVKQDKKDGAKYSYKDLNCLKQKVTKAGLIDFLELDP CRTI PPYLDNNNRKPPKCQS
 LILNPKFLDNQYPNWQYQLQELKQLS IQDYLDSEFETDLKDLKSSKDQPYFVEYKSSNQOMASGQRDYKDL DARI LQ
 FIFDRVKASDELLLEIYFQAKKQKASSELEKLESSKKLDEVIANSQLSQILKSQHTNGIFEQGTFLHLVCKYYK
 QRQRARDSRLYIMPEYRYDKKLDKYNNTGRFDDNNQLLYCNHKKPRQKRYQLLNDLAGVLQVSRNQLLSVVEWFQQ
 AQRVGEISKSQDEQIFEWLKSFKIASYCKAAVEMQKQYRGTLNKAIQTAFRQSENINKNKNTGNQQQALSSENSKDV
 KSLTADEKLLKLIENIAKASQKIGESLGLNDKQIKKFNSIYSAQIQQIAFAKRRKGNANTCAVCSADNAHRMQQIK
 ITELVEDNKDNIILSAKAQRLPAIPTRIVDGAVKKMATILAKNIVDDNWQNIKQVLSAKHQHLHPIITESNAFEFEP
 ALADVKGKSLKDRRKKALERISPENIFKDKNNRIKEFAKGISAYSGANLTDGDFDGAKEELDHIIPRSHKKGTLND
 EANLICVTRDDNKNI FAIDTSKWFEIETPSDLRDIGVATI QYKIDNNSRPKVRVKLDYVIDDDSKINYFMNHSLLKS
 RYPDKVLEILKQSTII EFESSGFNKTIKEMLGMTLAGIYNETSNN

(42). 嗜肺军团菌弗雷泽亚种菌株 ATCC 33216 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 51)。

MESSQLSPIGIDLGGKFTGVCLSHLESFAELPNHANTKYSVILMDHNNFQLSQAQRRI TRHRVRNKKRNQFVKRVA
 LQLFQHMLSRELNVKEEIALCHYLNNRGYTYVDTDLDEYLKDETTVNLKELLPPSEEHNFIDWLLQKMQSSKFRFI
 LGSRVEEKKDDKELKNAVKNIKNFITGFENKNSVEGHRHRKRYFENIKSDITNDNQLDGIKKKIPSVCLSNLLGHLSN
 LQWKNLHRYLAKNPKQFDEQTFGNEFLRMLKNFKHLKGSQESLAVRDLIQQLEQSKDYISILEKTPPEITIPPEEAR
 TNTGMEKQDQSLNPEKLNLYPKWRNLIPGIDAHFPLEKDLKLEHTKLRDRKRLISPSKQDEKRDYSIILQRYLDLNLK
 KIDKFKIKKQSLFGLGQKQLPANLIETHKEMETHFSSSLVSALIQIASAYNKEREDAAQGIWFDNAFSLCELSNINP
 PRKQKILPLLVGAILSEDFINNKDKWAKFKSFWNTHKIGRTLLKYKCKEIEEARKNSGNAFKIDYEEALNHPEHSNN
 KALTKIIQTIQAIQSHLGHNSQALINYNPFSLSQLYTILETKRDGFHKNCVAVTCENYWRSSQKTEIDPEISY
 ASRLPADSVRPFDFGLARMMQRLAYEIAMAKWEQIKHIPDNSSLLIPIYLEQNRFEFEESFKKIKGISRDKTLEQAI
 EKQNI RWEKFKRI INASMNVC PYKGT LIGGQGEIDHIYPRSLSKKHFGTIFNSEANLIYCSSQGNREKKEEHLLE
 HLSPIYLKHQFGTDNVS DIKNFISQNVANI KKYISFHLTPEQQKAARHALFLDYDDEAFKTVTKFLMSQQKARVNG
 TQKFLGKQIMEFLSTLADSKQLLEFSIKQINAEVHNHRELLSKQEPKLAQSQQSFP SHAIDATLTMGIGLKEFP
 QFSQELDNSWFINHLMPEVHLNPNVQSKEKYNKPNISSIPLFKDSLAEERFIPVWVKGETFAIGFSEKDLFEIKPSN
 KEKLFLLKTYSTKNPGESLQELQAISSAKWLYFPINKTLALEFLHHYFHKETVTPDDTTVCHFINSLRYTTKESI
 TVKILKESMPVLSVRFESSKKNVLTGTFKHTIALPATKDWERLFNHPNFLALKANPNPNKPEFNEFIKKYFLSENNPN
 SDIPNNDHTIKPKQKHTVRKVFSLPVI PGNAGTMMRI RRRKDNKGQPLYQLQTI DDTTPSMGIQINEGKLVKQEVLMDA
 YKTRNLSTIDGINNSERQAYAAFNDWLTLPVSTFKPEIKLEMKPHSKTRRYIRITQSLADFIKTIDEALMIKPSDS
 IDDP LNMPNEIVCKNKLFGNELKPRDGKMKIVSTGKTVTYEFESDSTPQWIQKLYVTQLKHP

(43). 土拉弗朗西斯菌土拉亚种 MA00-2987 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 52)。

MNVKILPIAIDLVDKNTGVFSAFYQKGTSLKLDNKNKGVYELSKDSYTLMMNRTARRHKRRGIDRKQLVKRFLFKL
 VWTEQLNLEWDKDTQQAISFLFNRRGFSFITDGYSTEYLNIVPEQVKAILMDIFDDYNGEDDLSYLKLATEQESKI
 SEIYNKLMQKILEFKLRKLCCTDIKDDKVSTKTLKEITSYEFELLADYLANYSESLKTQKFSYTDKQGNLKELSYHH
 DKYNIQEFLLKRHATINDEILGTLTDDFDIWNFNFEKFDKNEEKLQNEQEDKDHTQAHLHHFVFNVNIKSEMASG
 GRHRSQYFQIEITNVLDENNHQEGYLKNCENLHNKYSNLSVKNLVNLVGNLSNLELKLPLRKYFNDKIHAKADHWDE
 QKFTETTYCHWILGEWRVGVKQDKKDGAKYSYKDLNCLKQKVTKAGLIDFLELDP CRTI PPYLDNNNRKPPKCQS
 LILNPKFLDNQYPNWQYQLQELKQLS IQDYLDSEFETDLKDLKSSKDQPYFVEYKSSNQOMASGQRDYKDL DARI LQ
 FIFDRVKASDELLLEIYFQAKKQKASSELEKLESSKKLDEVIANSQLSQILKSQHTNGIFEQGTFLHLVCKYYK
 QRQRARDSRLYIMPEYRYDKKLDKYNNTGRFDDNNQLLYCNHKKPRQKRYQLLNDLAGVLQVSRNQLLSVVEWFQQ
 AQRVGEISKSQDEQIFEWLKSFKIASYCKAAVEMQKQYRGTLNKAIQTAFRQSENINKNKNTGNQQQALSSENSKDV
 KSLTADEKLLKLIENIAKASQKIGESLGLNDKQIKKFNSIYSAQIQQIAFAKRRKGNANTCAVCSADNAHRMQQIK
 ITELVEDNKDNIILSAKAQRLPAIPTRIVDGAVKKMATILAKNIVDDNWQNIKQVLSAKHQHLHPIITESNAFEFEP
 ALADVKGKSLKDRRKKALERISPENIFKDKNNRIKEFAKGISAYSGANLTDGDFDGAKEELDHIIPRSHKKGTLND
 EANLICVTRDDNKNI FAIDTSKWFEIETPSDLRDIGVATI QYKIDNNSRPKVRVKLDYVIDDDSKINYFMNHSLLKS
 RYPDKVLEILKQSTII EFESSGFNKTIKEMLGMTLAGIYNETSNN

(44). 嗜肺军团菌菌株 Leg01/20 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 53)。

MESSQILSPIGIDLGGKFTGVCLSHLEAFAELPNHANTKYSVILIDHNNFQLSQAQRRA TRHRVRNKKRNQFVKRVA
 LQLFQHILSRDLNAKEETALCHYLNNRGYTYVDTDLDEYIKDETTINLLKELLPSESEHNFIDWFLQKMQSSSEFRKI
 LVSKVEEKKDDKELKNAVKNIKNFITGFENKNSVEGHRHRKRYFENIKSDITKDNQLDSIKKKIPSVCLSNLLGHLSN
 LQWKNLHRYLAKNPKQFDEQTFGNEFLRMLKNFRHLKGSQESLAVRNLIQQLEQSQDYISILEKTPPEITIPPEEAR

TNTGMEKQDQSLNPEKLNLYPNWRNLI PGI IDAHP FLEKDLHTKLRDRKRI ISPSKQDEKRDYSI LQRYLDL NK
KNR

(45). 伯克霍尔德里氏菌目细菌 1_1_47 Cas9 蛋白 (BBCas9) 的氨基酸序列 (SEQ ID NO: 54)。

MGKTHIIGVGLDLGGTYTGT FITSHPSDEAEHRDHSSAFTVVNSEKLSFSSKSRTAVRHRVRSYKGFDLRRRLLLIV
AEYQLLQKKQTLAPEERENLRIALSGYLKRRGYARTEAETDTSVLES LDPSVFSAPSFTNFFNDSEPLNIQWEAIA
NSPETTKALNKELSGQKEADFKKYIKTSFPPEYSAKEILANYVEGRRAILDASKYIANLQSLGHKHSKYLSDI LQDM
KRDSRITRLSEAFGSTDNLWRIIGNISNLQERAVRWYFNDAKFEQQEQQLDAVKKLVLRALKYLRSDDKWEWSAQ
KQIIQSLEQSGDVLVDLAGLDPDRITIPPYEDQNNRRPPEDQTLYLNPKALSSEYGEKWKSWANKFAGAYPLLTEDLT
EILKNTDRKSRIKIRSDVLPDSYRLAYILQRAFDRSIALDECSI RRTAEDFENGVVIKNEKLEDVLSGHQLEEFLE
FANRYYQETAKAKNGLWFPENALLERADLHPPMKNKILNVI VGOALGVSPAEGTDFIEEIWNSKVKGRSTVHSICNA
IENERKTYGPFSEDYKFKVTALKKEGKTEKELSKKFAAVIKVLMVSEVVPFIGKELRSLDEAQS KFDNPYSLAQLY
NLIETERNFGFSKVS LAAHLENAWRMTMTDGS AQCCRLPADCVRPFDFGFI RKAIDRNSWEVAKRIAEEVKKSVDFTN
TVKIPVAIEANSFNFTASLTDLKYIQLKEQKLLKKLEDIQRNEENQEKRWLSKEERI RADSHGICAYTGRPLDDVGE
IDHII PRSLTLKKSSESIYNSEVNLI FVSAQGNQEKNNIYLLSNLAKNYLAAVFGASDLSQITNEIESTVLQ LKAAG
RLGYFDLLSEKERACARHALFLNSDSEARRAVI DVLGSRKASVNGTQAWFVRSI FSKVRQALAAWTQETGNELI FD
AISVPAADSSEMRKRFAEYRPEFRKPKVQPVASHSIDAMCIYLAACSDPFKTKRMGSQ LAIYEPINF DNLTGSCQV
IQNTPRNFSDKTN IANSPIFKETIYAERFLDIIVSRGEI FIGYPSNMPFEEKPNRISI GGKDPFSLSVLGAYLDKA
PSSEKEKLTIRVVKNAFELFSKVAGSKFTAEDDAAKILEALHFVTVKQDVAATVSDLI KSKKELSKDSIENLAK
QKGC LKKVEYSSKEFKFKGSLIIPAAVEWGV LWNVFKENTAEELKDENALRKALEAAWPPSFGTRNLHLSKAKRVFS
LPVATQSGAVRIRRRTAFGDFVYQSQDTNNLYSSFPVKNGKLDWSSPIIHPALQNRNLTAYGYRFVDHRSISMSE
FREVYNKDDLMRIELAQGTSRRYL RVEMPGEKFLAWFGENSISLGSSFKFSVSEVFDNKIYTENAETFKFLPKPRE
DNKHNGTIFFEVGVPRVIFNYIVGGAASSLKEIFSEAGKERS

(46). 土拉弗朗西斯菌土拉亚种 SCHU S4 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 55)。

MNVKILPIAIDLGVKNTGVFSAFYQKGTSLKELDNKNGKVEYELSKDSYTL LMNRTARRHKRRGIDRKQLVKRFLK
VWTEQLNLEWDKDTQQAISFLFNRRGFSFITDGYS TEYLNIVPEQVKAILMDIFDDYNGEDDLD SYLKLATEQESKI
SEIYNKLMQKILEFKLRKLC TDIKDDKVSTKTLKEITSYEFELLADYLANYSES LKTQKFSYTDKQGNL KELSYYHH
DKYNIQEFLKRHATINDEILGTLLTDDFDIWNFNFEKFD FDKNEEKLNQEDKDHTQAHLHHFV FVFNKI KSEMASG
GRHRSQYFQEITNVLDENNHQEGYLKNFCENLHNKYSNL SVKNLVNLVGNLSNLEL KPLRKYFNDKIHAKADHWDE
QKFTETYCHWILGEWRVGVKDDKDGAKYSYKDL CNELKQKVTKAGLIDFLELDPCRTI PPYLDNNNRKPPKCQS
LILNPKFLDNQYPNWQQYLQELKQLS IQDYLDSFETDLKDLKSSKDQPYFVEYKSSNQ QMASGQRDYKDL DARI LQ
FIFDRVKASDELLNEIYFQAKK LKQKASSELEKLES SKKLDEVIANSQLSQILKSQHTNGIFEQGTFLHLVCKYK
QRQRARDSRLYIMPEYRYDKKLDKYNNTGRFDDNNQL LTYCNHKPRQKRYQLLNDLAGVLQVSRNQLLSSVEEWFQQ
AQRVGEISKSQDEQI FEWLKSFKIASYCKAAVEMQKQYRGTLLKNAIQTAIFRQSENINKNKNTGNQQALS ENSKDV
KSLTADEKLLKLIENIAKASQKIGESLGLNDKQIKKFNSIY SFAQIQQIAFAKRKGNANTCAVCSADNAHRMQQIK
ITELVEDNKDNIILSAKAQRLPAI PTRIVDGA VKMATILAKNIVDDNWQNIKQVLSAKHQLHIPIITESNAFEFEP
ALADVKGKSLKDRRKKALERISPENIFKDKNRIKEFAKGISAYS GANLTDGDFDGAKEELDHIIPRSHKKYGT LND
EANLICVTRDDNKNI FAIDTSKWF EIETPSDLRDI G VATIQYKIDNNSRPKVRVKLDYVIDDDSKINYFMNHSLLKS
RYPDKVLEILKQSTIIEFESSGFNKTIKEMLGMTLAGIYNETSNN

(47). 土拉弗朗西斯菌新凶手亚种 U112 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 56)。

MNFKILPIAIDLGVKNTGVFSAFYQKGTSLERLDNKNGKVEYELSKDSYTL LMNRTARRHQRRGIDRKQLVKRFLK
IWTEQLNLEWDKDTQQAISFLFNRRGFSFITDGYSPEYLNIVPEQVKAILMDIFDDYNGEDDLD SYLKLATEQESKI
SEIYNKLMQKILEFKLMKLC TDIKDDKVSTKTLKEITSYEFELLADYLANYSES LKTQKFSYTDKQGNL KELSYYHH
DKYNIQEFLKRHATINDRI LDTLLTDDLDIWNFNFEKFD FDKNEEKLNQEDKDHIQAHLHHFVFAVNKI KSEMASG
GRHRSQYFQEITNVLDENNHQEGYLKNFCENLHNKYSNL SVKNLVNLIGNLSNLEL KPLRKYFNDKIHAKADHWDE
QKFTETYCHWILGEWRVGVKDDKDGAKYSYKDL CNELKQKVTKAGLVD FLELDPCRTI PPYLDNNNRKPPKCQS
LILNPKFLDNQYPNWQQYLQELKQLS IQNYLDSFETDLKVLKSSKDQPYFVEYKSSNQ IASGQRDYKDL DARI LQ
FIFDRVKASDELLNEIYFQAKK LKQKASSELEKLES SKKLDEVIANSQLSQILKSQHTNGIFEQGTFLHLVCKYK
QRQRARDSRLYIMPEYRYDKK LHKYNNTGRFDDNQL LTYCNHKPRQKRYQLLNDLAGVLQVSPNFLKDKIGSDDDL
FISKWLVEHIRGFKKACEDSLKIQKDNRGLLNHKINIARNTKGKCEKEIFNLICKIEGSEDKKGNYKHGLAYELGVL
LFGEPNEASKPEFDRKIKKFNSIY SFAQIQQIAFAERKGNANTCAVCSADNAHRMQQIKITEPVEDNKDKIILSAKA
QRLPAI PTRIVDGA VKMATILAKNIVDDNWQNIKQVLSAKHQLHIPIITESNAFEFEPALADVKGKSLKDRRKKAL
ERISPENIFKDKNRIKEFAKGISAYS GANLTDGDFDGAKEELDHIIPRSHKKYGT LND EANLICVTRGDNKKNKGNR
IFCLRDLADNYK LKQFETDDLEIEKKIADTIWDANKKDFKFGNYRSFINLTPQEQA FRHALFLADENPIKQAVIR
AINNRNRTFVNGTQR YFAEVLANNIYLRAKKENLNTDKISFDYFGIPTIGNRGIAEIRQLYEKVDSDIQAYAKGDK

PQASYSHLIDAMLAFCIAADEHRNDGSIGLEIDKNYSLYPLDKNTGEVFTKDI FSQIKITDNEFSDKKLVKKAIEG
 FNTHRQMTTRDGIYAENYLPI LIHKELNEVRKGYTWKNSEEIKI FKGKKYDIQQLNNLVYCLKFVDPKISIDIQISTL
 EELRNILTTNNIAATAEYYYINLKTQKLHEYYIENYNTALGYKKYSKEMEFLRSLAYRSEVVKI KSIDVQVLDKD
 SNFIIGKITL PFKKEWQRLYREWQNTTI KDDYEF LKSFFNVKSITKLHKKVRKDFSLPI STNEGKFLVKKRTWDNNF
 IYQILNDSDSRADGTPFIPAFDISKNEIVEAII DSFTSKNIFWLPKNIELQKVDNKNIFAIDTSKWFEVETPSDLR
 DIGIATI QYKIDNNSRPKVRVKLDYVIDDDSKIN YFMNHSLLKS RYPDKVLEILKQSTII EFESSGFNKTI KEMLGM
 KLAGIYNETSNN

(48). 土拉弗朗西斯菌土拉亚种 79201237 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 57)。

MNVKILPIAIDLVDKNTGVFSAFYQKGTSLKLDNKNKGVYELSKDSYTL LMNRTARRHRRGIDRKQLVKRFLK
 VWTEQLNLEWDKDTQQAISFLFNRRGFSFITDGYSTEYLNIVPEQVKAILMDIFDDYNGEDDLSYLKLATEQESKI
 SEIYNKLMQKILEFKLRKLCCTDIKDDKVSTKTLKEITSYEFELLADYLANYSES LKTQKFSYTDKQGNLKELSYHH
 DKYNIQEFLKRHATINDEILGTLTLLTDDFDIWNFNFEKFDKNEEKLNQEDKDHTQAHLHHFVFNKIKSEMASG
 GRHRSQYFQEITNVLDENNHQEGYLKNFCENLHNKKYSNLSVKNLVNLVGNLSNLELKLPRKYFNDKIHAKADHWDE
 QKFTETECHWILGEWRVGVKDDKDGAKYSYKDLCELKQKVTKAGLIDFLELDP CRTI PLYDNNNRKPPKCQS
 LILNPKFLDNQYPNWQQYLQELKQLS IQDYLD SFETDLKDLKSSKDQPYFVEYKSSNQOMASGQRDYKDL DARI LQ
 FIFDRVKASDELLNEIYFQAKKLKQKASSELEKLESSKKLDEVIANSQLSQILKSQHTNGIFEQGTFLHLVCKYYK
 QRQRARDSRLYIMPEYRYDKKLDKYNNTGRFDDNNQLLYTCNHKPRQKRYQLLNDLAGVLQVSRNQLLSVVEWFQQ
 AQRVGEISKSQDEQIFEWLKSFKIASYCKAAVEMQKQYRGT LKNAIQTAI FRQSENINKNKNTGNQQALS ENSKDV
 KSLTADEKLLKLIENIAKASQKIGESLGLNDKQIKKFNSIYSFAQIQQIAFAKRKGNANTCAVCSADNAHRMQQIK
 ITELVEDNKDNIILSAKAQRLPAIPTRIVDGA VKKMATILAKNIVDDNWQNIKQVLSAKHQ LHIPIITESNAFEFEP
 ALADVKGKSLKDRRKKALERISPENIFKDKNRRIKEFAKGISAYSANLTDGDFDGAKEELDHIIPRSHKKYGT LND
 EANLICVTRDDNKNIFAIDTSKWFEIETPSDLRDI G VATIQYKIDNNSRPKVRVKLDYVIDDDSKIN YFMNHSLLKS
 RYPDKVLEILKQSTIEFESSGFNKTIKEMLGMTLAGIYNETSNN

(49). 土拉弗朗西斯菌土拉亚种 AS 713 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 58)。

MNVKILPIAIDLVDKNTGVFSAFYQKGTSLKLDNKNKGVYELSKDSYTL LMNRTARRHQRRGIDRKQLVKRFLK
 VWTEQLNLEWDKDTQQAISFLFNRRGFSFITDGYSTEYLNIVPEQVKAILMDIFDDYNGEDDLSYLKLATEQESKI
 SEIYNKLMQKILEFKLRKLCCTDIKDDKVSTKTLKEITSYEFELLADYLANYSES LKIQKFSYTDKQGNLKELSYHH
 DKYNIQEFLKRHATINDEILD TLLTDDFDIWNFNFEKFDKNEEKLNQEDKDHTQAHLHHFVFAVNKIKSEMASG
 GRHRSQYFQEITNVLDENNHQEGYLKNFCENLHNKKYSNLSVKNLVNLVGNLSNLELKLPRKYFNDKIHAKADHWDE
 QKFTETECHWILGEWRVGVKDDKDGAKYSYKDLCELKQKVTKAGLIDFLELDP CRTI PLYDNNNRKPPKCQS
 LILNPKFLDNQYPNWQQYLQELKQLS IQDYLD SFETDLKDLKSSKDQPYFVEYKSSNQOMASGQRDYKDL DARI LQ
 FIFDRVKASDELLNEIYFQAKKLKQKASSELEKLESSKKLDEVIANSQLSQILKSQHTNGIFEQGTFLHLVCKYYK
 QRQRARDSRLYIMPEYRYDKKLDKYNNTGRFDDNNQLLYTCNHKPRQKRYQLLNDLAGVLQVSRNQLLSVVEWFQQ
 AQRVGEISKSQDEQIFEWLKSFKIASYCKAAVEMQKQYRGT LKNAIQTAI FRQSENINKNKNTGNQQALS ENSK
 DVKSLTADEKLLKLIENIAKASQKIGESLGLNDKQIKKFNSIYSFAQIQQIAFAERKGNANTCAVCSADNAHRMQQ
 IKITELVEDNKDNIILSAKAQRLPAIPTRIVDGA VKKMATILAKNIVDDNWQNIKQVLSAKHQ LHIPIITESNAFEF
 EPALADVKGKSLKDRRKKALERISPENIFKDKNRRIKEFAKGISAYSANLTDGDFDGAKEELDHIIPRSHKKYGT L
 NDEANLICVTRDDNKNIFAIDTSKWFEIETPSDLRDI G VATIQYKIDNNSRPKVRVKLDYVIDDDSKIN YFMNHSLL
 KS RYPDKVLEILKQSTIEFESSGFNKTIKEMLGMTLAGIYNETSNN

(50). γ 变形菌 HTCC5015 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 59)。

MTKNYISPIAIDLGAFTGVALYQYLEGADCTQEVAKGLLVDDRGNVTWSQEGRRGKRHVQVRYKRRKMAKRLWL
 LDSEYGIKREEVTEPLLKFINGLLNRRGYTYISEEVDEESMNVSPLPFSEMMPDYFNSSAPLLEQLAKLLSDKNKLV
 RFRAEGKIPSNKNEFKLLD TALDGKYKDEKKESEAWGNILIASENVLKSTVDGHKSRSEYLANIKEDIKSNEELE
 KQISSKEIDGFYNLVGHLSNFQLRLLRKYFNDPNMSGVSYWDEKRLEKYFYQWVQGWHTKGGTDEAEKNIILKTKG
 APLKTLKSLADLTI PPYEDQNNRRPKCQSVLLSDEKLTMHYPKWKEWVGLVKQNDNAYLNENVTLANALHRIV
 ERSRSIDPYQLRLLSITDAEKRN DLAGYKRLKLSLSEVDEFLLLVKNIVDETKEAREGLWFETENKLFKCGKTP
 PRKEKLSLTL SVAVLGKNLS DDEQSSFIEEFWKS GTPKIERNRVGRWCRLASQVQKTYGVYLKEYGLQLHKLEAGK
 KLDDKPLALLYKNSGLIASKIGEALNIEPDEVSRFASPHSLAQIFNIEGDVAGFNKTCRACTYENIWRMQEEKVES
 LLTNQLLSEIHGERKVP LKSAMCTRLSADSTRPFDGQMASIEHIARKIAQHKAQINDVPKEFSIDIPIIIESNQF
 SFTAELEEKRGRGSAKAKKAKELGEKSKAGVWSKTERIKTSSEGICPYTGAPLGGSGEIDHIIPRSLTGRTKKTVF
 NSEANLIYCSSKGNHDKGNRVYVIEQLNDKYLKQFSTSDVNLIKKIKTTIQRFTEGGEKLRFSSELSREDQKAFR
 HALFVPELKS EVTSL LAVKNI TRVNGTQAWLAKKIASLLAEHLDKQGRDYTL SAHQIDPWSVSKQRKMLASAEPIWA
 KKDPQPAASHVVDVAVCTFLEALEQPHTASRLKTI SSTSFEKTGWRSAIIPDLIKVDALDRRPKYRRYINIGSTS LFKD
 GIYAERFLPILIDENGLMAGYDIDNSLKAKGADVVFESLSPFLLFKGEVGAQSLSDWQERIDGRYLYMSIDKVKAF

DYLQEKVGEKDIAAELLNSIHFTQRKTELRAKFSDDSGKKMKTLDAIRKSLKLTVTVNEIGKRKEKCGFSGTIGIPA
KSAWENLLDEPALLEYWGTKMPPQEIWEKVYRKHFRPNI PNQAHKVRKDFSLPVVDSVSGGFRVVRKTPNGYNYQL
LAIDGYSAVGFKKEGDNVDFKSPALVPQIAESKSVTPISSELVHLDKNEIVYFDEWRKIDISDSDLKQFVSSLELAP
GSQNRFYIRFTVDEDDQFERHFKSALRVNGIQDLDTVNKTFDWNREIPSLIIPRNLFLLETGQKITFEYIANGANA
EVKKAYSLRRA

(51). 土拉弗朗西斯菌土拉亚种 3571 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 60)。
MNVKILPIAIDLVDKNTGVFSAFYQKGTSLKLDNKNKGVYELSKDSYTLMMNNRTARRHQRRGIDRKQLVKRFLFKL
VWTEQLNLEWDKDTQQAISFLFNRRGFSFITDGYSTEYLNIVPEQVKAII LMDIFDDYNGEDDLDLSYLKLATEQESKI
SEIYNKLMQKILEFKLRKLCTDIKDDKVSTKTLKEITSYEFELLADYLANYESLKI QKFSYTDKQGNLKELSYYHH
DKYNIQEFLLKRHATINDEILDITLTDFFDIWNFNFEKFDKNEEKLQNEQEDKDHTQAHLHHFVFAVNKIKSEMASG
GRHRSQYFQEIITNVLDENNHQEGYLKNCENLHNKKYSNLVKNLVN LVGNLSNLELKLPLRKYFNDKIHAKADHWDE
QKFTETECHWILGEWRVGVKQDKKDGAKYSYKDLCELKQKVTKAGLIDFLELDP CRTI PPLYDNNNRKPPKCQS
LILNPKFLDNQYPNWQYQLQELKQLS IQDYLDSEFETDLKDLKSSKDQPYFVEYKSSNQMASGQRDYKDL DARI LQ
FIFDRVKASDELLLENIYFQAKKQKASSELEKLESSKKLDEVIANSSQLS QILKSQHTNGIFEQGTFLHLVCKYYK
QRQRARDSRLYIMPEYRYDKKLDKYNNTGRFDDNNQLLTYCNHKPRQKRYQLLNDLAGVLQVSRNQLLSSVEEFQO
AQRVGEISKSQDEQIFEWLKSFKIASYCKAAVEMQKQYRGT LKNAIQTAI FRQSENINKNKNKNTGNQQALS ENSK
DVKSLTADEKLLKLIENIAKASQKIGESLGLNDKQIKKFNSIYSFAQIQQIAFAERKGNANTCAVCSADNAHRMQQ
IKITELVEDNKDNIILSAKAQRLPAIPTRIVDGAVKMMATILAKNIVDDNWQNIKQVLSAKHQ LHIPIITESNAFEF
EPALADVKGKSLKDRRKKALERISPENIFKDKNRRIKEFAKGISAYSGANLTDGDFDGAKEELDHIIPRSHKKYGT L
NDEANLICVTRDDNKNI FAIDTSKWFEIETPSDLRDI G VATIQYKIDNNSRPKVRVKLDYVIDDDSKINYFMNHSLL
KSRYPDKVLEILKQSTIIEFESSGFNKTIKEMLGMTLAGIYNETSNN

(52). 土拉弗朗西斯菌土拉亚种 FSC033 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 61)。
MNVKILPIAIDLVDKNTGVFSAFYQKGTSLKLDNKNKGVYELSKDSYTLMMNNRTARRHQRRGIDRKQLVKRFLFKL
VWTEQLNLEWDKDTQQAISFLFNRRGFSFITDGYSTEYLNIVPEQVKAII LMDIFDDYNGEDDLDLSYLKLATEQESKI
SEIYNKLMQKILEFKLRKLCTDIKDDKVSTKTLKEITSYEFELLADYLANYESLKTQKFSYTDKQGNLKELSYYHH
DKYNIQEFLLKRHATINDEILGTLLTDDFFDIWNFNFEKFDKNEEKLQNEQEDKDHTQAHLHHFV FVVN KIKSEMASG
GRHRSQYFQEIITNVLDENNHQEGYLKNCENLHNKKYSNLVKNLVN LVGNLSNLELKLPLRKYFNDKIHAKADHWDE
QKFTETECHWILGEWRVGVKQDKKDGAKYSYKDLCELKQKVTKAGLIDFLELDP CRTI PPLYDNNNRKPPKCQS
LILNPKFLDNQYPNWQYQLQELKQLS IQDYLDSEFETDLKDLKSSKDQPYFVEYKSSNQMASGQRDYKDL DARI LQ
FIFDRVKASDELLLENIYFQAKKQKASSELEKLESSKKLDEVIANSSQLS QILKSQHTNGIFEQGTFLHLVCKYYK
QRQRARDSRLYIMPEYRYDKKLDKYNNTGRFDDNNQLLTYCNHKPRQKRYQLLNDLAGVLQVSRNQLLSSVEEFQO
AQRVGEISKSQDEQIFEWLKSFKIASYCKAAVEMQKQYRGT LKNAIQTAI FRQSENINKNKNKNTGNQQALS ENSKDV
KSLTADEKLLKLIENIAKASQKIGESLGLNDKQIKKFNSIYSFAQIQQIAFAERKGNANTCAVCSADNAHRMQQIK
ITELVEDNKDNIILSAKAQRLPAIPTRIVDGAVKMMATILAKNIVDDNWQNIKQVLSAKHQ LHIPIITESNAFEFEP
ALADVKGKSLKDRRKKALERISPENIFKDKNRRIKEFAKGISAYSGANLTDGDFDGAKEELDHIIPRSHKKYGT LND
EANLICVTRDDNKNI FAIDTSKWFEIETPSDLRDI G VATIQYKIDNNSRPKVRVKLDYVIDDDSKINYFMNHSLLKS
RYPDKVLEILKQSTIIEFESSGFNKTIKEMLGMTLAGIYNETSNN

(53). 土拉弗朗西斯菌新凶手亚种菌株 DPG 3A-IS Cas9 蛋白的氨基酸序列 (SEQ ID NO: 62)。
MNVKILPIAIDLGVKNTGVFSAFYQKGTSLERLDNKNKGVYELSKDSYTLMMNNRTARRHQRRGIDRKQLVKRFLFKL
IWTEQLNLEWDKDTQQAISFLFNRRGFSFITDGYSP EYLNIVPEQVKAII LMDIFDDYNGEDDLDLSYLKLATEQESKI
SEIYNKLMQKILEFKLRKLCTDIKDDKVSTKTLKEITSYEFELLADYLANYESILKTQKFSYTDKQGNLKELSYYHH
DKYNIQEFLLKRHATINDEILDITLTDLDIWNFNFEKFDKNEEKLQNEQEDKDHTQAHLHHFVFAVNKIKSEMASG
GRHRGQYFQEIITNVLDENNHQEGYLKNCENLHNKKYSNLVKNLVN LIGNLSNLELKLPLRKYFNDKIHAKADHWDE
QKFTETECHWILGEWRVGVKQDKKDGAKYSYKDLCELKQKVTKAGLVD FLELDP CRTI PPLYDNNNRKPPKCQS
LILNPKFLDNQYPNWQYQLQELKQLS IQNYLESFETDLKVLKSSKDQPYFVEYKSSNQIASGQRDYKDL DARI LQ
FIFDRVKASDELLLENIYFQAKKQKASSELEKLESSKKLDEVIANSSQLS QILKSQHTNGIFEQGTFLHLVCKYYK
QRQRARDSRLYIMPEYRYDKKLDKYNNTGRFDDNNQLLTYCNHKPRQKRYQLLNDLAGVLQVSPNFLKDKIGSDDDL
FISKWLVEHIRGFKKACEDSLKIQKDNRGLLNHNKINIARNTKGKCEKEIFNLICKIEGSEDKKNYKHGLAYELGVL
LFGEPNEASKPEFDRKIKKFNSIYSFAQIQQIAFAERKGNANSCAVCSADNAHRMQQIKITEPVEDNKDKIILSAKA
QRLPAIPTRIVDGAVKMMATILAKNIVDDNWQNIKQVLSAKHQ LHIPIITESNAFEFEPALADVKGKSIKDRRKKAL
ERISPENIFKDKNRRIKEFAKGISAYSGDNLASGDFDGAKEELDHIIPRSHKKYGT LND EANLICVTRDDNKKNKGNR
IFCLHDLANNYKQKQFETDDLEIEKKIADTIWDASKDFKFGNYRSFINLTPQEQA FRHALFLADENPIKQAVIR
AINNRNRTFVNGTQRYFAEVLANNIYLRAKKENLNTDKISFDYFGIPTIGNGRGIAEIRQLYEKVDSDIQAYAKGDK

PQASYSHLIDAMLAFCIAADEHRNDGSIGLEIDKNYSLYPLDKNTGEVFTKDI FSQIKITNNEFS DKKLVRKKATEG
FNTHRQMT RDGIYAENYLPILIHKELEVRKGYTWGNSEELKIHKGKKYDIQQLNNLVYCLKFDKSI SIDIQISTL
EELRNILTTNNIATTA EYYYINLKTQKLHEYYIENYNTALGYKKYTKEMEFLRSLAYRSEVKIKSIDDVQVLDKD
SNFIIGKITL PFKKEWQRLYREWQNTTIKDDYEF LKSFFNVKSITKLHKKVRKDFSLPISTNEGKFLVRRKTDWNNF
IYQILNDSDSRVDGTPFIPAFDISKNEIVEAII DSFTSKNIFWLPKNIELQKVDNKNIFAIDTSKWFEVETPSDLR
DIGVATI QYKIDNNSRPKVRV KLDYVIDDDSKINYFTNHSLLKS RYPDKVLEILKQSTII EFESSGFNKTIKEM LGM
TLAGIYNETSNN

(54). 琥珀酸单胞菌属物种 CAG:777 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 63)。

MSKKT VYSNIAIDLGGKYTGCVSYTSSILPSADDINAFIIEMPDDGNGIKYTVKDRTQKRHMIRALDRFKKARKLIY
LVISSVIHRELTADEKESISSLMKRRGYTRLEAEIDL DILKECPVDL FAS YIKDLFDESQSLYDQFNQKCCDLDSAK
AMKKALSSEELKQAMDSSELDKIEKKIYKSAFDMVKA AENIINTCEFGNKHRKQYLSDIKADINKDSRLADIVKSF
DGNEKLFKCVGNI SNLQLRALRWYFNDINMKG NPKWI PERFKK VWARAYKYFHYPKEDQ QKARELIK NLEI SNDVTS
LLCSIDPLDTIPFYEDQNNRNPPTDLTLLSPKALDKKYTNLWEKWA EKFAKLYPNLKYQLEDIVKISDRKSRI DVK
SSSYTQSKVFN SYVMQRLFDLAADKNEPITLLRSWVRDP SNSKLADTVKIIQSVVEDREDDFYKLAQSYNEVEMA
KSGLWSVIDDPLEISGIHPMKNKVISTLVGNVLGINNDFDYEFLEVWRAPIKGRKTLKSVCKDIEGLRKEYGNT
FKIEYDYANYS AKEKKE RLTA DQKKLVEINKIIEEISPLVRNSISQDPLSEIKFNNPFSFSQLYTLIETDTLGFSSN
CVAVNMENARMQFLG NHALCSRLPTESVRPFDGALGKIL ERQAYEIAKVSAEIKELDSLNETVINVGILIEQNK
FEFTESI AKIKKSKLEKIKKRSQSGVSRQLKKWENKEKRIKSASRGVCPY TGERISETNGEIDHII PRSFTKS SMG
TIFNSEANLIFCRLTGNRNK DENMYTLRELSPKYLN AVFGTANIVQIQQNI EGTVAKISSKNPRFLFDMMDQKEQDC
CRHALFMPHSAKAYQIVVSAMARSYSTRVNGTQSYFVNAIISKLKNE LSSWLKSNRNTLNFFAYKVDP SDVHSTRDDL
SKLNTLLKKQENQPITSHAI DALCVLASASTNERI SENISKESCLAKI SNTGNLLKLEPHEFKILRIEKKDFSDKND
DFSRLFNDTIYAEHFLPVMVMKDQVKI GFSWNSKEGNSIEIVKDQHKFLELLAPFFNKAI ESSCKFITYKINKKK
AFDFIHR YIHQ EVDVND CALIVKILES LRYTTVNTD VCDIYNDREKRFKSKSEILTNKNFELNFSLPKPFKILACKK
IYLP SKNEWSKITEKFS DYLGKDNIKNWKEI IYNSLNLNRKTTK KLNHAGTKRVYSLPI LKNASGGIRIKRYTDEME
EFYQLLEVNTPQTVKRKGFALSEKGEVMWNK PITVDSYIGK NLTLL EKNYKNVCIGDNFVPMKENRMVYKNEKTEVY
ISPSSDSRRLV IIKQPFDEFNKCLKNKYKNYLELPSTLVLSGDEIKIYVKNMKNDFIGNPRVETKNNKKTGKIVMLS
IGKLVRYCYMSTTCDSLMKTAYDCALS

(55). 拟杆菌门口腔分类群274菌株 F0058 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 64)。

MKNILGLDLGTNSIGWAYVKQAQNDSEKSEIAELGVRVNPLTVDEKTNFEKGP ISTNADRTLKRGARRNLQRYKLR
RQELIELLKDYQVISDKTILAEDGKETTFETLRLRVKSAKEQVSL EELARILLMINKKRGYKSSRKAKNEDEGQLID
GMEIARRLYDENLTTGQLSLQLLKEGKSLP DYYSRDLQHELDKVVNFQKQFHPDILTDEFYTSIQGKGQKATRESF
SSKYGIDTAENKGSREEKQLQAYQWRADATHKELTKEELAFVIT EINNINSSGYLGAI SDRSKELIFANETVGEY
LWKQIQNDPHTSLKNQVYRQDYFDEFERIWETQSGFYQILTESLKNELRDILIFYQRKLSQKALVSI CEFESKE
IPLKDKQGNPISKNGIPVTKLVGPKVAPKS SPLFQEFKIWQNLNNVLRPKGSRKRKVT TNQOSTLLEGQEPVFELN
LEQKELLFEELNLKGNLKT DYCLELIGYSAKEWEMNYPTLEGNKTNKALYEAYLKIIDSEGYDAKLLKVKTDKDDI
NLDDIKVSASEIKT MISDIFNELGINTKILDFDAELERKQFEAQASYRLWHLLYSYEGDDSKSGNELLYRLLLEDKFG
FKREHSKTLANVSLSDYGNLSTKAIRTIYPYIKENKFSTACELAGYRHSKLSISKEENENRTLKDTLDVLKKNLSLR
QPVVEKILNQMINLINELEKHSEKDEKGNIVKYFKFDEIHI ELARELKKNAEERRELD SKVREGKDRNEKIITILK
SEFGLPYPTKNDITRYRLYEELKDN GYKDLTYNTYI PREKLFSKEIDIEHII PKSR LFNDSFSNKTICYRQDNLNG
ERTAYDYIEGLGKTQLDEYLN RVENLYTNKFITKAKYRNLLKKESEI GDGFIERDLRETQYIAKKAKEMLLQITRHV
VSTTGSITDRLRDDWELVNMRELNL PKYKVLGLTEI QERKNGQTVEVITDWTKRNDHRHAMDALTVAF TKHSHIQ
YLNFLNARKDEKHKEYSNIMGIQQLETEKTTDKDGKDRRVFKTPIPNFRQVAK EHLSEILISHKTKNKVVT RATNKI
AGSKKQQETLS PRGQLHKETVYGKYRRYVQKEEKIGSKFDLQTI ERVANPLYKRL LTKRLEENGDPKKAFTGKNAL
AKSPIYLD ETQTQQLPEKVKLVWLETDFSI RKDITPDNFKDEKSIDKVLDEGIKRI LKDR LKSF DGDAKRAFTDLEQ
NPIWLNKEKGIAIKRVTISGVSN AEPLHTKKDHLGNEI PDDTGKPIPV DYSTGNHHIAIYRDENGNLQEKAVSFF
EAVTRANQGLPTVDKTYNQHLGWQFLFTMKQNMVFVFNKTEGDFPQEIDLDDPKNKKLISPNLFRVQKISTKNYFF
RHHLET SVETHNALNGV TYKSQGLGNGITDI I KVRINHLGDIVHIGEY

(56). 土拉弗朗西斯菌土拉亚种 8070069 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 65)。

MNVKILPIAIDL DVKNTGVFSAFYQKGTSL EKLDNKNKGVYELSKDSYTL LMNRTARRH KRRGIDRKQLVKRFLK
VWTEQLNLEWDKDTQQAISFLFNRRGFSFITDGYSTEYLNIVPEQVKAILMDI FDDYNGEDDLSYLKLATEQESKI
SEIYNKLMQKILEFKLRKLC TDIKDDKVSTKTLKEITSYEFELLADYLANYESLKTQKFSYTDKQGNL KELSYYHH
DKYNIQEF LKRHATINDEILGTLTDDFDIWNFNFEKFD FDKNEEKLQ NQEDKDHTQAHLH HFVFNKI KSEMASG
GRHRSQYFQEITNVL DENNHQEGY LKNFCENLHNKYSNLSVKNLVN LVGNLSNLEL KPLRKYFNDKIHAKADHWDE

QKFTETYCHWILGEWRVGVKQDKKDGAKYSYKDLNCLKQKVTKAGLIDFLELDFPCRTIPPYLDNNRKPKPCQS
LILNPKFLDNQYPNWQQYLQELKQLS IQDYLDSEFETDLKDLKSSKDQPYFVEYKSSNQMASGQRDYKDLDARILQ
FIFDRVKASDELLLEIYFQAKKQKASSELEKLESSKKLDEVIANSQLSQILKSQHTNGIFEQGTFLHLVCKYYK
QRQRARDSRLYIMPEYRYDKKLDKYNNTGRFDDNNQLTYCNHKPRQKRYQLLNDLAGVLQVSRNQLLSVVEWFQQ
AQRVGEISKSQDEQIFEWLKSFKIASYCKAAVEMQKQYRGLTKNAIQTAIFRQSENINKNKNTGNQQALSSENSKDV
KSLTADEKLLKLIENIAKASQKIGESLGLNDKQIKKFNISYFAQIQQIAFAKRKGNANTCAVCSADNAHRMQQIK
ITELVEDNKDNIILSAKAQRLPAIPTRIVDGVKMMATILAKNIVDDNWQNIKQVLSAKHQHLHPIITESNAFEFEP
ALADVKGKSLKDRRKKALERISPENIFKDKNNRIKEFAKGISAYSANLTDGDFDGAKEELDHIIPRSHKKYGTLDND
EANLICVTRDDNKNIIFAIDTSKWFIEIETPSDLRDIGVATIQYKIDNNSRPKVRVKLDYVIDDDSKINYFMNHSLLKS
RYPDKVLEILKQSTIIEFESSGFNKTIKEMLGMTLAGIYNETSNN

(57). 嗜肺军团菌嗜肺亚种 JCM 7571 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 66)。

MESSQILSPIGIDLGGKFTGVCLSHLEAFAELPNHANTKYSVILIDHNNFQLSQAQRATRHRVRNKKRNQFVKRVA
LQLFQHILSRDLNAKEETALCHYLNNRGTYVDTDLDEYIKDETTINLLKELLPSESEHNFIDWFLQKMQSSEFRKI
LVSKVEEKDDKELKNAVKNIKNFITGFENKSVGHRHRKRVYFENIKSDITKDNQLDSIKKKIPSVCLSNLLGHLSN
LQWKNLHRYLAKNPKQFDEQTFGNEFLRMLKNFRHLKGSQESLAVRNLIQQLEQSQDYISILEKTPPEITIPPEAR
TNTGMEKDQSLLLNPEKLNLYPNWRNLI PGI IDAHPLEKDLKLEHTKLRDRKRIISPSKQDEKRDYSIILQRYLDLNDK
KIDKFKIKKQLSFLGQKQLPANLIETQKEMETHFNSSILSVSLIQIASAYNKEREDAAQGIWFDNAFSLCELSNINP
PRKQKILPLLVGAILSEDFINNKDKWAKFKIFWNTHKIGRTSLKSKCKEIEEARKNSGNAFKIDYEEALNHPEHSNN
KALIKIIQTI PDI IQAIQSHLGHNSQALIHNPFSLSQLYTILETKRDFHKNCAVAVTCENYWRSSQKTEIDPEISY
ASRLPADSVRPFDFGVLARMMQRLAYEIAMAKWEQIKHIPDNSSLLIPIYLEQNRFEFEESFKKIKGSSSDKTLQAI
EKQNIQWEEKFQRIINASMNICPYKGASIGGQGEIDHIYPRSLSKKHFGVIFNSEVNLICYSSQGNREKKEEHYLL
HLSPLYLKHQFGTNDVSDIKNFIQNVANI KKYISFHLTPEQQKAARHALFLDYDDEAFKTIKFLMSQQKARVNG
TQKFLGKQIMEFLSTLADSKQLQLEFSIKQITAEVHDHRELLSKQEPKLVKSRQQSFP SHAI DATLTMSI GLKEFP
QFSQELDNSWFINHLMPEDEVHLPVRSKEKYNKPNISSTPLFKDSLVAERFIPVWVKGETFAIGFSEKDLFEIKPSN
KEKLFLLKTYSTKNPGESLQELQAKSKAKWLYFPINKTLALEFLHHYFHKEIVTPDDTTVCHFINSLRYTTKESI
TVKILKEPMPVLSVKFESSKKNVLSFKHTIALPATKDWERLFNHPN FLALKANPAPNPKFENEFIRKYFLSDNNPN
SDIPNNGHNIKPKQKHKAVRKFVSLPVI PGNAGTMMRI RRDKNKGQPLYQLQTI DDT PSMGI QINEDRLVKQEVLMDA
YKTRNLSTIDGINNSEGQAYATFDNWLTLPVSTFKPEI IKLEMKPHSKTRRYIRITQSLADFIKTIDEALMIKPSDS
IDDPLNMPNEIVCKNKLFGNELKPRDGKMKIVSTGKIVTYEFESDSTPQWIQTLYVTQLKKQP

(58). 产琥珀酸沃廉氏菌 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 67)。

MLVSPISVDLGGKNTGFFSFTDSLNSQSGTVIYDES FVLSQVGRRSKRHSKRNNLRNKLKRLFLLLIQEHHGLSI
DVLPADEIRGLFNKRGYTYAGFELDEKDKDALESDTLKEFLSEKQLS I DRSDVDFLNQIASNAESFKDYKKGFEAV
FASATHSPNKKLEKDELKSEYGENAKELLAGLRLVTEKILDEFDKQENQGNLPRAKYFEELGEYIATNEKVKSSFDS
NSLKLTDMTKLIIGNISNYQLKELRRYFNDKEMEKDIIWIPNKLHKITERFVRSWHPKNDADRQRAELMKDLKSKEI
MELLTTEPEVMTIPPYDDMNRRGAVKQTLRLNNEYLDKHLPNWRDIAKRLNHGKFNDDLADSTVKGYSEDSTLLHR
LLDTSKEIDIYELRGKKPNEELLVKTGLQSDANRLYGFQNYELIRQKVRAGIWPVKNKDDSLNLEDNSNMLKRCN
HNPPHKKNIHNLVAGILGVKLDEAKFAEFKELWSAKVGNKLSAYCKNIEELRKTGHNTFKIDIEELRKKDPAEL
SKEEKAKLRLTDDVILNEWSQKIANFFDIDDKHRQRFNNLFSMAQLHTVIDTPRSGFSSTCKRCTAENRFRSETAFY
NDETGEFHKKATATCQRLPADTQRPFSGKIERIYDKLGYELAKIKAKELEGMEAKEIKVPIILEQNAFEYEESLRKS
KTGSNDRVINSKDRDGGKLA KAKENAEDRLKDKDKRIKAFSSGICPYCGDTIGDDGEIDHILPRSHTLKIYGTVFN
PEGNLIYVHQKCNQAKADSIYKLSDIKAGVSAQWIEEQVANIKGYKTFVSVLSAEQQKAFRYALFLQNDNEAYKKVVD
WLRTDQSARVNGTQKYLAKKIQEKLTKMLPNKHLSEFELADATEVSELRRQYARQNPLLAKEKQAPSSHAIDAVM
AFVARYQKVKDGTPPNADEVAKLAMLDSWNPASNEPLTKGLSTNQKIEKMIKSGDYGQKNMREVFSGKIFGENAIG
ERYKPIVVQEGGYYIGYPATVKKGYELKNCKVVTSKNDIAKLEKIKKNQDLISLKENQYIKIFSNKQTI SELSNRY
FNMNYKNLVERDKEIVGLLEFIVENCYRYTKKVDVKFAPKYIHETKYPFYDDWRRFDEAWRYLQENQNKTSKDRFV
IDKSSLNEYYPDKNEYKLDVDTQPIWDDFCRWYFLDRYKTANDKKSIRIKARKTFSLLAESGVQGVKFRARKKIPT
GYAYQALPMDNNVIAGDYANILLEANSKTLVLPKSGISIEKQLDKKLDVIKKT DVRGLAIDNNSFFNADFDTGIR
LIVENTSVKVGNFPISAIKSAKRMIFRALFEKEKGRKKKTTISFKESGPVQDYLVKVLKIKVIQLRTDGSISNI
VVRKNAADFTLSFRSEHIQKLLK

(59). 嗜肺军团菌嗜肺亚种 JCM 7571 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 68)。

MESSQILSPIGIDLGGKFTGVCLSHLEAFAELPNHANTKYSVILIDHNNFQLSQAQRATRHRVRNKKRNQFVKRVA
LQLFQHILSRDLNAKEETALCHYLNNRGTYVDTDLDEYIKDETTINLLKELLPSESEHNFIDWFLQKMQSSEFRKI

LVSKVEEKKDDKELKNAVKNIKNFITGFEEKNSVEGHRHRKVFENIKSDITKDNQLDSIKKKI PSVCLSNLLGHLN
LQWKNLHRYLAKNPKQFDEQTFGNEFLRMLKNFRHLKGSQESLAVRNLIQQLEQSQDYISILEKTPPEITIPPYEAR
TNTGMEKQDQSLLLNPEKLNLYPNWRNLI PGII DAHP FLEKDLHTKLRDRKRIISPSKQDEKRDYSILQRYLDLNK
KIDKFKIKKQLSFLGQKQLPANLIETQKEMETHFNSSLVSVLIQIASAYNKEREDAAQGIWFDNAFSLCELSNINP
PRKQKILPLLVGAILSEDFINNKDKWAKFKI FWNTHKIGRTSLKSKCKEIEEARKNSGNAPKIDYEEALNHPEHSNN
KALIKIIQTI PDI IQAIQSHLGHNDQSALI YHNPFSLSQLYTIETKRDGFHKNCVAVTCENYWRSQKTEIDPEISY
ASRLPADSVRPFDFGLARMQRLAYEIAMAKWEQIKHII PDNSSLLIPIYLEQNRFEFEESFKKIKGSSSDKTEQAI
EKQNIQWEEKFQRI INASMNICPYKGASIGGQGEIDHIYPRSLSKKHFGVIFNSEVNLIYCSSQGNREKKEEHYLL
HLSPLYLKHQFGTNDVSDIKNFISQNVANIKKYSFHLTPEQQAARHALFLDYDDEAFKTIKFLMSQQKARVNG
TQKFLGKQIMEFLSTLADSKQLQLEFSIKQITAEVVDHRELLSKQEPKLVKSRQOSFPSHAIDATLTMSIGLKEFP
QFSQELDNWFINHMLPDEVHLN PVRSEKYNKPNISSTPLFKDSLYAERFIPVWVKGETFAIGFSEKDLFEIKPSN
KEKLFLLKTYSTKNPGESLQELQAKSKAKWLYFPINKTLALEFLHHYFHKIIVTPDDTTVCHFINSRLRYTKKESI
TVKILKEPMPVLSVKFESSKKNVLSFKHTIALPATKDWERLNFNPNFLALKANPAPNPKFENEFIRKYFLSDNNPN
SDIPNNGHNIKPKQKHKAVRKFVSLPVI PGNAGTMMRI RRRKDNKGQPLYQLQTI DDTPSMGIQINEDRLVKQEVLMDA
YKTRNLSTIDGINNSEGQAYATFDNWLTLPVSTFKPEIKLEMKPHSKTRRYIRITQSLADFIKTIDEALMIKPSDS
IDDPLNMPNEIVCKNKLFGNELKPRDGKMKIVSTGKIIVTYEFESDSTPQWIQTLYVTQLKKQP

(60). 土拉弗朗西斯菌土拉亚种 1378 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 69).
MNVKILPIAIDLVDKNTGVSAFYQKGTSLKLDNKNKGVYELSKDSYTLMLNNTARRHKRRGIDRKQLVKRFLFKL
VWTEQLNLEWDKDTQQAISFLFNRRGFSFITDGYSTEYLNIVPEQVKAILMDIFDDYNGEDDLDSYLKLATEQESKI
SEIYNKLMQKILEFKLRKLCTDIKDDKVSTKTLKEITSYEFELLADYLANYESLKTQKFSYTDKQGNLKELSYHH
DKYNIQEFKLRHATINDEILGTLTDDFDIWNFNFEKFDKNEEKLQNEQEDKDHTQAHLHHFVFNKIKSEMASG
GRHRSQYFQEIITNVLDENNHQEGYLKNCENLHNKYSNLSVKNLVNLVGNLSNLELKLPRKYFNDKIHAKADHWDE
QKFTETECHWILGEWRVGVKQDQDKKDGAKYSYKDLNKLKQKVTKAGLIDFLELDP CRTI PLYLDNNNRKPPKCQS
LILNPKFLDNQYPNWQQYLQELKQLS IQDYLDSEFETDLKDLKSKDQPYFVEYKSSNQMASQRDYKDLDARI LQ
FI FDRVKASDELLLENIYFQAKKQKASSELEKLESSKLLDEVIANSQLSQILKSQHTNGIFEQGTFLHLVCKYYK
QRQRARDSRLYIMPEYRYDKKLDKYNNTGRFDDNNQLLYTCNHKPRQKRYQLLNDLAGVLQVSRNQLLSVVEWFQQ
AQRVGEISKSQDEQIFEWLKSFKIASYCKAAVEMQKQYRGLTKNAIQTAIFRQSENINKNKNTGNQQALSSENSKDV
KSLTADEKLLKLIENIAKASQKIGESLGLNDKQIKKFNSIYSAFQIQI AFAKRKGNANTCAVCSADNAHRMQQIK
ITELVEDNKDNIILSAKAQRLPAIPTRIVDGAVKKMATILAKNIVDDNWQNIKQVLSAKHQHLHPIITESNAFEFEP
ALADVKGKSLKDRRKKALERISPENIFKDKNNRIKEFAKGISAYSGANLTDGDFDGAKEELDHIIPRSHKKYGTLND
EANLIVTRDDNKNI FAIDTSKWFEIETPSDLRDIGVATIYQKIDNNSRPKVRVKLDYVIDDDSKINYFMNHSLLKS
RYPDKVLEILKQSTIIEFESSGFNKTIKEMLGMTLAGIYNETSNN

(61). 土拉弗朗西斯菌 T10902 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 70).
MNVKILPIAIDLVDKNTGVSAFYQKGTSLKLDNKNKGVYELSKDSYTLMLNNTARRHKRRGIDRKQLVKRFLFKL
VWTEQLNLEWDKDTQQAISFLFNRRGFSFITDGYSTEYLNIVPEQVKAILMDIFDDYNGEDDLDSYLKLATEQESKI
SEIYNKLMQKILEFKLRKLCTDIKDDKVSTKTLKEITSYEFELLADYLANYESLKTQKFSYTDKQGNLKELSYHH
DKYNIQEFKLRHATINDEILGTLTDDFDIWNFNFEKFDKNEEKLQNEQEDKDHTQAHLHHFVFNKIKSEMASG
GRHRSQYFQEIITNVLDENNHQEGYLKNCENLHNKYSNLSVKNLVNLVGNLSNLELKLPRKYFNDKIHAKADHWDE
QKFTETECHWILGEWRVGVKQDQDKKDGAKYSYKDLNKLKQKVTKAGLIDFLELDP CRTI PLYLDNNNRKPPKCQS
LILNPKFLDNQYPNWQQYLQELKQLS IQDYLDSEFETDLKDLKSKDQPYFVEYKSSNQMASQRDYKDLDARI LQ
FI FDRVKASDELLLENIYFQAKKQKASSELEKLESSKLLDEVIANSQLSQILKSQHTNGIFEQGTFLHLVCKYYK
QRQRARDSRLYIMPEYRYDKKLDKYNNTGRFDDNNQLLYTCNHKPRQKRYQLLNDLAGVLQVSRNQLLSVVEWFQQ
AQRVGEISKSQDEQIFEWLKSFKIASYCKAAVEMQKQYRGLTKNAIQTAIFRQSENINKNKNTGNQQALSSENSKDV
KSLTADEKLLKLIENIAKASQKIGESLGLNDKQIKKFNSIYSAFQIQI AFAKRKGNANTCAVCSADNAHRMQQIK
ITELVEDNKDNIILSAKAQRLPAIPTRIVDGAVKKMATILAKNIVDDNWQNIKQVLSAKHQHLHPIITESNAFEFEP
ALADVKGKSLKDRRKKALERISPENIFKDKNNRIKEFAKGISAYSGANLTDGDFDGAKEELDHIIPRSHKKYGTLND
EANLIVTRDDNKNI FAIDTSKWFEIETPSDLRDIGVATIYQKIDNNSRPKVRVKLDYVIDDDSKINYFMNHSLLKS
RYPDKVLEILKQSTIIEFESSGFNKTIKEMLGMTLAGIYNETSNN

(62). MH0245_GL0161830_1 Cas9 蛋白 (MH0245Cas9) 的氨基酸序列 (SEQ ID NO: 71).
VCTVKAQGLYLFASQTEVLFVKLIHNRFNIIHIREHKMTVKSTLAVDMGGRYTGI FSYTTDSGFPKAKEARAYVLNMP
DNDALTY SMAARTQTRHRIRSRQRFVLARRLTYILIEGKLRKLSPREKEAIISSLLRRRGYSRLESELDLSVLQGV
SGFFKCFLPNFEDEENLLTQWTSITDGYLQNNSDSRRQIQIFLESSKDSKEFLT VVKSQHQDTKEYKNALKVMRDDA
ESMIEQSMFGHKHRRLYLEAIAQDIPRDSRLKPIIEAFSGVEKFHFFIGNLSNLQRLRALRWYFNDPSMKNNVFDKER
LKSVLVRAVQFFHYPKDLTQQRAEVLNAYEGATDILETLQTLNPELTI PPYEDQNNRRPPLDQTLWLS PRLLDQRYG

DTWEIWVQNLRLSRPLSKGIDENLDTLITLITDRKARLLERQSGRLIHYSQKLYHSYVLQRLDRTVENDAYLLKTLV
SSNRGNSNEIHQAQERLTRDLGSLQHIKKFLDFVRQYYDEVKAKRGLWFIVEKPLMERADIHPPMKNDSVILRLVGN
ILCVSDLVDFWTRKVKQGQSTVRSCTAI EKTRKEYGNSFNLYQRALYLQSKGKLSAEDKDFIKLQSNVLLVSD
VIAEALDIKEEQKKKFFANPFSLAQLYNI IETEKSGFI STTLAAVDENAWRNQLQKARCVCQLCADTVRPFDFGALRNI
LDRQAYEIAKLKAEELLSTELKNQTI DLVVLLESNQFAPSASLAEVKKSANTAAI RQKVAKAQKRQDDRWLSKDERI
KSASRGLCPYTGKNLGDKGEVDHI I PRSLSMNYMGSILNSEANLIYCSQEGNQLKLNGRKKLSDLADNYLKVVFVGT
DRGTICKYIEKVSVELTDAKIVQFELLDRSQDQAVRHALFLEDFSEARRRI I RLLGKINTARVNGTQAWFAKSFITK
LRELTKEWCANNQITLAFDLYRLDAQTVSQDYRKKFALINKDWAKPDDKKQPIASHAIDAFCVFAAAKDKRNIANVL
GVFDEVAEEQNLKTI AQLMPEVNLI SPKRKSIDKNEVGSRALMKEGIFAEHFLPI LVRGDDCRIGFDWSESGSVK
VKDADKLFVGLDGLLQSKRSVNGFETYTVDRIKAFELLHDVFI RPCSQKMLEQAEVLEKLHYITQNI SVTSVYDA
VNRQFKCREEILKDKDFDIKVDLGNRFSGAKGKITLPKREWEKLVNRSELKNI KDKLSDKGSEKTPDGETLIYDI
FRSIPVQKLSHKATRVRVWLSLPKI PSISSGVRIKRKDSNGNDIYQLYMLNDTKCKGFVVNEKGVIDWSDDLADLYKQ
PTLTILNGRYLKADQYVRMDQWYEVDCGRDDVIVKMCPTSGRRYIEITQSKQFEDWTGYISGSFWNYPVTIKLSS
QQIANFVKNSQMPLLGKPRSGQITVITLGNLTKYWCVESKNSMMNEAYQKAYLVHFNQ

(63). O2.UC1-1_GL0148428_1 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 72)。

MTQSERRFSCSIGIDMGAKYTGVFYALFDREELPTNLNSKAMTLVMPETGPRYVQAQRTAVRHRRLRGQKRYTLARKL
AFLVVDDMIKKQEKRLTDEEWKRGREALSGLLKRGRYSRPNADGEDLTPLENVRADVFAAHPAFSTYFREVRS LAEQ
WEEFTANIGNVEKFLGDPNI PADKEFI EFVAEGLIDKTEKKAYQSALSTMRANANVLTGLRQMGHKPRSEYFKAIE
ADLKKDSRLAKISEAFGGAKRLARLLGNLSNLQLRAERWYFNAPDIMKDRGWEPDRFKKTLVRAFKFFHFAKQDNKQ
HLELIKQIENSEDI IETLCTLDPNRTI PYPEDQNNRRPFLDQTLLESPEKLTROYGEIWKTSARLTSAEPTLAPAA
EILERSTDRKSRVAVNGHEPQPTLAYSALQRAFDRSKALDPYALRALAAGSKSNKLTSAARTALENCIGGQNVK
FLDCARRYREADDAKVGWFDNADGLLERSDLHPPMKKILPPLLVANILQTDETTQKFLDEIWRKQIKGRETVAS
RCARIETVRKSFGGYFNTAYNTAQYREVNKLPRNAQDKELTI RDRVAETADFIANMGLSDEQKRKFANPFSLAQF
YTLIETEVSFGFSATTLAVHLENARMTIKDAVINGETVRAAQCSRLPAETARPF DGLVRRLLVDRQAWEI AKPASTDI
QNKVDFSNQIVDVSIFVEENKFEFSASVADLKNQSHAKDTKKNKRTKNKMLSEAEKLETRWLNKNERIKKASRETCP
YTGQRLAEGGEIDHILPRSLIKDARGIVFNSEPNLI YVSSRGNQFKKQRYSLSNLHTDYLKIKFKTSNIAAITAEI
EDVVTKLQQTTHRLKFFDLLNEHEQDCVRHALFLDDGSEARNAVLELLATQRRTRVNGTQIWMIKNLANKIREELQDW
CRTTNNRHLHFQAAATDVSTAKNLRKLAQNQPDFEKPDIQPIASHSIDALCSFAVGSADAERDQDGFYLDGKAVLG
LYPQSCEVIHLQAKPQEEKSHFDSVAIFKEGIYAEQFLPIFTLNGKIWI GYETLDAKGERCGAIEVSGKQPEELTLM
LTPFFNKPVGDL SAHATYRIQKKPAYEFLAKAALQPLSAEKRLAALLDALRYCTSRKSLKSLFMAANGKSLKKQED
VLNPELFQLLFQLRVELKGEKAFKLNGLSKLPVEQDWIRICNSPELIDAFGKACTDAELTSKIARIWKR SATRDLAH
APVRREFSLPVIDKPSGGFRIRRTNLFQDELYQVHAINAKKYRGFASADCNVDSKGI LFNELQHENLTECGGRFII
SADETSMSDWRKVLSEDNLGIWIAPGTERRRYVRVETTFIQANHWFKQSVENWAITSPLSLPASFKIDKPAEFKKA
VGTLSSELLGQPRDKIFIENIVNAKRIRFWYTVRNSNKKMNESYNNASKS

(64). MH0275_GL0111719_1 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 73)。

MGKTHIIGVGLDLGGTYTGT FITSHPSDEAEHRDHSSAFTVVNSEKLSFSSKSRRTAVRHRVRSYKGFDLRRRLLLLV
AEYQLLQKKQTLAPEERENLRIALSGYLKRRGYARTEAETDTSVLETLDPSVFSAPSFTNFFNDSEPLNIQWEAIA
NSPETTKALNKELSGQKEADFKKYIKTSFPEYSAKEILANYVEGRRAILDASKYIANLQSLGHKHSKYLSIDLQDM
KRDSRITRLSEAFGSTDNLWRIIGNISNLQERAVRWYFNDAKFEQGEQLDAVKLKNVLRALKYLRSDDKWEASQ
KQIIQSLEQSGDVLVLAGLDPDRTI PYPEDQNNRRPPEQDTLYLNPKALSSEYGEKWKSWANKFAGAYPLLTEDLT
EILKNTDRKSRIKIRSDVLPDSYRLAYILQRAFDRSIALDECSI RRTAEDFENGVVIKNEKLEDVLSGHQLEEFLE
FANRYYQETAKAKNGLWFPENALLERADLHPPMKNKILNVI VGGALGVSPAEGTDFIEEIWNSKVKGRSTVRSICNA
IENERKTYGPFYSEDKFVKTKALKEGKTEKELSKKFAAVIKVLMVSEVVPF IGKELRLSDEAQSKFDNPFYSLAQLY
NLIETERNGFSKVS LAAHLENARMTMTDGSQAQCRLPADCVRPFDGFIKAI DRNSWEVAKRIAEVKKSVDFDFTNG
TVKIPVAIEANSFNFTASLTDLKYIQLKEQKLEKLEDIQRNEENQEKRWLSKEERI RADSHGICAYTGRPLDDVGE
IDHIIPRSLTLKKSESIYNSEVNLI FVSAQGNQEKNNIYLLSNLAKNYLA AVFGTSDLSQITNEIESTVLQKAA
RLGYFDLLSEKERACARHALFLNSDSEARRAVI DVLGSRKASVNGTQAWFVRSIFSKVRQALAAWTQETGNELIFD
AISVPAADSSEMRKRFAEYRPEFRKPKVQPVASHSIDAMCIYLAACSDPFKTKRMGSQALAIYEPINFDNLF TGSQCV
IQNTPRSFSKDNIANSPIFKETIYAERFLDIIVSRGEIFIGYPSNMPFEEKPNRISIGGKDPFISILSVLGAYLDKA
PSSEKEKLTIRVVKNKAFELFSKVAGSKFTAEDKAAKILEALHFVTVKQDVAATVSDLIKSKKELSKDSIENLAK
QKGCLKKVEYSSKEFKFKGSLIIPAAVEGWKVLWNVFKENTAEELK DENALRKALEAAWPS SFGTRNLHSAKRVFS
LPVVATQSGAVRIRRKTAFGDFVYQSQDTNNLYSSFPVKNGKLDWSSPIIHPALQNRNLTAYGYRFVDHRSISMSE
FREVYNKDDLMRI ELAQGTSSRRYLRVEMPGEKFLAWFGENISLGS SFKFSVSEVFDNKIYTENA EFTKFLPKPRE
DNKHNGTIF FELVGPVIFNYIVGGAASSLKEIFSEAGKERS

(65). MH0221_GL0020823_1 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 74)。
 MGKTHIIGVGLDLGGTYTGTFFITSHPSDETEHGDHSSAFTVVNSEKLSFSSSKSRTAVRHRVRSYKGFDLRRRLLLLV
 AEYQLFQKKQALTPEERENLRIALSSYLKRRGYARTEAETDVSVLES LDP SVFSSAPGFTNFFNDSEPLNIQWKVIS
 NSPETARALNKAVSGQKEADFKKYVKTSFPEYSAKEILANYVEGRRAL LDASKYIANLQSLGHKHSKYLSDI LQDM
 KRDSRISRLSEAFSGSIDNLWRIIGNISNLQERAVRWYFNDAKFEQQEQQLDADKLNKLVILRALKYLRSDDKKWSASQ
 KQIIQSLEQSGDVLVDLAVLDPDRTIIPPYEDQNNRRPPEQDTLYLNPRALSSEYGDKWKSWANKFAAAAYPLTDDLA
 EILKNTDRKSRIKIRSDALPDSYRLTYILQRALDRSIALDECSI RR TAEDFENG VVIKNEKLEEVLSGHQLEEFLE
 FANRYYQETAKAKNGLWFPENALLERADLHPMPKNKI LNVI V GQALGVSPAKGTEFIEEIWNRRKVKGRSTVRSICNA
 IENERKTYGPYFSDDYKFVK TALKEGKTEKELSKKFASVIKVLKMNVEVVPFIGKELRLSDEAQS KFDN PYS LAQLY
 NLIETERNGFSKVS LAAHLENAWRMTMTD GSAQC CR L P ADCV R PFDGFI RKAIDRNSWEVAKRIAEVVKKSVDF TNG
 TVKIPVAVEANSFNFTASLTDLKYIQLKEQKLK K KLEDIQRNEENQEKRWLSKEERI RADSHGICAYTGRPLDDVGE
 IDHII PRSLT LK KSESIYNSEVNLIFVSAQGNQEKNNIYLLSNLAKNYLAAVFGTSDLNQITNEIESTVLQLKAAG
 RLG YFDLLSEKERACARHALFLNSDSEARRAVIDVLGSRRKASVNGTQAWFVRSIFS KVRQALAAWIQETGNELVFD
 AISVPAADSSSEMRKRFAEYRPEFRKAKVQPVASHSIDAMCIYLAACSDPFKTKRMGSQ LAIYEPINF DN LFTGSCQV
 VQNTPRSFTDKTNIANSPIFKETIYAERFLDIIVSRGEIFIGYPSNMPFEKKPNRISIGGKDPF S ILSVLGAYLNKA
 PSSEKEKLTIIYRVEKNKAFELLSKVAGSKFTAEBEDKAAKILEALHFVTVKQDVAATVSDLVKSKKELSKDSIEDLAK
 QKGCLKKVEYSSKEFKFKGSLIIP TAV EWGKVLWNVFKENTAEELKDENVLRKALEAAWPS SFGTRNLHSAKRVFS
 LPVVATQSGAVRIRRKTAFGNFVYQSQDTNNLYSSFPVKNKGLDWSSPIIHPALQNRNLTA YGYRFVDHRSISMSE
 FRVIYNKDDLRIELAQGTSSRRYLRVEMPGEKFLAWFGENSISLGSSFKFSVSEVFDNKIYTENA EFTKFLPKPRED
 NKHNGTIFPELVGPRVIFNYIVGGAASSLKEIFSEAGKERS

(66). MH0246_GL0016360_1 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 75)。
 MTQSERRFSCSIGIDMGAKYTG V FYALFDREELPTNLNSKAMTLVMPETGPRYVQAQRTAVRHLRGQKRYTLARKL
 TFLVVDDMIKKQEKRLTDEEWKRGREALSGLLKRRGYSRPNADGEDLTPLENVRADVFAAHPAFSTYFSEVRS LAEQ
 WEEFTANIGNVEKFLGDPNIPADKEFIEFAVAEGLIDKTEKKAYQSALSTMRANANVLTGLRQMGHKPRSEYFKAIE
 ADLKKDSRLAKIREAFGGAERLARLLGNLSNLQIRAE R WYFNAPDIMKDRGWEPDRFKKTLVRAFKFFHPAKDQNKQ
 HLELIKQIENSEDIIETLCTLDPNRTIIPPYEDQNNRRPPLDQTL LLSPEKLT RQYGEIWKTSARLT S SEPTLAPAA
 EILERSTDRKSRVAVNGHEPLPTLAYQLSYALQRAFDRSKALDPYALRALAAGSKSNKLT SARTALENCIIGGQVVEK
 FLDCARRYREADDAKVLWFDNADGLLERSDLHPPMKKKIIPLLVANI LQTDETT GQKFLDKIWRKQIKGRETVAS
 RCARIETVRKSFGGGFNIAYNTAQYRKVNKLPRNAQDKELLTIRDRVAETADFIAANLGLSDEQKRKFANPFSLAQF
 YTLIETEVS GFSATT LAIHLENAWRMTIKDAVINGETVRAAQCSRLPAETARPF DGLVRRLVDRQAW EIAKRAS TDI
 QSKVDFSNGIVDVSIFVEENKFEFSASVADLKKNKRVKDKMLSEAEKLET R WLNKNERIKKASRETCPYTGDR LAEG
 GEIDHILPRSLIKDARGIVFNAEPNLIYASSRGNQLKKNQRYLSKLDPDYLNKVFKTSNIAAITAEIEDVVTKLQQ
 THRLKFFDLLNEHEQDCVRHALFLDDGSEARNAVLELLATQRRTRVNGTQI WMIKSLANKI REELQDWCRTTNNRLH
 FQAAATDVSTAKNLRKLQAQNPFEKPDIQPIASHSIDALCSFAVGSADAERDQDGFYLDGKAVLGLYPQSC EVI
 RLQAKPQEEKSHFDSVAIFKEGIYAEQFLPIFTLNGRIWIGYETLDSKGERCGAIEVSGKKPEELLSMLAPFFNKPV
 GDLSAHATYRIQKKPAYEFLAKAALQPLSAEEKRLAALLDALRYCTS RKS LKSFFMAANGKSLKKREDLLKPKLFQL
 KVELKGEKAFKNGSLTLPKQDWLRICNSPELADAFGNPCSADELTSKLAQIWKRPVMRDLAHAPVRREFSLPVID
 NPSGGFRIRRTNLFGNELYQVHAINAKKYRGFASADSNVDWANGILFNELQHENLTECCGRFIT SADVTPMSEWRKV
 VTEDNLR IWIAPGTGRRYRVETTFIQASHWFEQSVENWAITSPLSLPASFKVDKPAEFQKAVGTELSSELLGQPRS
 EIFIESVGNAKHIRFWYIVVSSNKKMNESYNNASKS

(67). MH0124_GL0052249_1 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 76)。
 MSKKT VYSNIAIDLGGKYTGCVSYTSSILPSADDINAFIIE MPD DNGIKYTVKDRTQKRHMIRALDRFKKARKLIY
 LVISSVIHRELTADEKESISSLMKRRGYTRLEAEIDLDI LKECPVDLFASYIKDLFDESQS L YDQFNQKCCDLDSAK
 AMKKALSSEELKQAMDSS ELDKIEKKIYKSAFDMVKAENIINTCEFGNKHKQYLSDIKADINKDSRLADIVKSF
 DGNEKLFKCVGNI SNLQLRALRWYFN DINMKNPKWI PERFKVWARAYKYFHYPKEDQ QKARELIK NLEI SNDVTS
 LLC SIDPLDTIIPPYEDQNNRNPPTDLTLLS PKALDKKYTNLWEKWA EKFAKLYPNLKYQLEDIVKISDRKSRI DVK
 SSSSYTQSKVFN SYVMQRLFDLAADKNEPITLLRSWVRDP SNSKLADTVKIIQSVVEDREDDFYKLAQSYNEVEMA
 KSGLWSVIDDPLEISIGIHPMKNKVISTLVGNVLGINNDFDYEFLEVWRAPIKGRKTLKSVCKDIEGLRKEYGNT
 FKIEYDYANYS AKEKKEKRLTADQKKLVEINKIEEISPLVRNSISQDPLSEIKFNPF SFSQLYTLIETDTLGFSSN
 CVAVNMENNARMQFLLGNHALCSRLPTESVRPFDGALGKILERQAYEIAKVKS AEIKELDSLNETVINVGILIEQNK
 FEFTESI AKIKKSKLLEKIKKRSQSGVSRQLKKWENKEKRIKSASRGVCPYTG ER ISETNGEIDHII PRSFTKSMG
 TIFNSEANLIFCRLTGNRNKDENMYTLRELSPKYLN AVFGTANIVQIQNIEGTVAKISSKNPRFLFDMMDQKEQDC
 CRHALFMPHSKAYQIVVSAMARSYSTRVNGTQSYFVNAII SKLKNELSSWLKSNRNTLNFFAYKVDPSDVHSTRDDL
 SKLNTLLKKQENQPITSHAI DALCVLASASTNERISENISKESCLAKISNTGNLLKLEPHEFKILRIEKKDFSDKND
 DFSRKL FNDTIYAEHFLPVMVMKDQVKIGFSWNSKEGNSIEIVKDQHKFLELLAPFFNKAI E SCKFITYKINKKK

AFFDFIHRHYIHQEVVDVNDICALIVKILES LRYTTVNTDVCDIYNDREKRFKSKSEILTNKNFELNFSLPKPFKILACKK
IYLPSKNEWSKITEKFSYDLGKDNKIKWKEIIYNSLNLNRKTKKLNHAGTKRVYSLPILKNASGGIRIKRYTDEME
EFYQLLEVNTPQTVKRKGFALSEKGEVMWNKPITVDSYIGKNTLLEKNYKNVCIGDNFVPMKENRMVYKNEKTEVY
ISPSSDSRRLVIIKQPFDEFNKCLKNKYKNYLELPSTLVLSGDEIKIYVKNMKNDFIGNPRVETKNNKKTGKIVMLS
IGKLVRYCYMSTTCDLSLMKTAYDCALS

(68). T2D-113A_GL0099135_1 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 77)。

MKSTLAIDLGGRYTGVVSYTSEFGLPSADSLKAYVINTFNDELTVSVKSTQVRHRLRSQKRFLKARQLMTLIVQH
KVARNLSNSEKEALFGLLKRGRYSRYDSELDLSCLSIETEPFQSFIPEISEDESLLTQWESFKAGYSTGNVDCINI
VRSFASLPDAKQFKEHLKQYSLSKEEVKDYLTALSVLREDALLVEQQSLFGHKHRTQFFDDIYGDIPKDSRLKTI
DFGGLEPFYRTIGNICNLQRLRALRWYFNDPSMKNPYYDDKRLKSVLSRAYHFFHYPKKELEKCRDILSAIESAD
AFLASFDPKKTIIPPYEDQNNRKPPLDQTLLELPRALDLRYGSRWELWVNNLLRAPESKIDDDQLDVI
LASTDRKSRLQERQTGQIVHYTSKKLYNAYVLHRLLDRTLDNDPYSRLISKERIGQSNDIHEAMSRLTKDLSQHLKEFLEFARNY
YNEVELAKRGLWSIVQTPLEERADIHPPMKKGTIISRLINNVLCTENVSI SQFWHLRVKGTSTVKSICASIERTRKE
YGNFFNAEYKALAKHRVKEKLDADQKEFIKIADNVQLTSQRIQAALNLSSEQREKFNPFSLAQLYTIIEETEPNGF
IKTSVAVNDENAWRTNLAGHARCVRSSDTVRPFDGAVKKTLEQAYELAKLKFSELQNSVKNDCTVDVTLLESNE
FAFTSSMAEIKKSDKAKAIRNKADRAMKRQDQRWQNNRIKTASRGLCPYTGQPLGDNGEIDHIVRSASLMLYGD
ILNSEPNLIYCSQKGNQLKSNRIKTIQDLASNYLEKVFGTSNFSAVEELICSTVSRSLSDNDIVQFDSLQEQDAVR
HSLFLNSANPVNRRVRQALAKQFKSRVNGTQAWFARQFIKTLQELSANWCKTNNISLTFDVYRSLPSQSVQGDYRAKI
ATSAPELAKPDSAQPIGSHAI DALCVFGAALDNQKIAKRMGLDNLTLSDSLHLTALLPQSVDIINI KRKSLDFDKDD
LQSI SWFKETIYAEFLPIMVKKDETRIGFVSWNDQYS LKVKHGDILIKTLSSFLNEAKPVREYEIQTFTINREKAFE
LLHKVFTAPCTSEECQIADLLEGMYSYVTKTPI SDFDFDEKSR SFKLNKADKLF SIKIKLDRRFAGSGTIFLPCQKDW
ENEIKKINQVFTGQLSNSQNGQDVVLSMKKNLKRSAKHRRTRRVWSPPIVPSGLSAAIRLNKSHDGRNIYQLQMIK
NAKYKGLCVDNQEQIDWNQLLVADIYKQNSLTLRAERYLDSQKFI SMDHWLDDIDIGIPEVKVRMS PGEANRRHICIT
QSRQQFENWIGCTLGSMTMPATKKNLNEELANFIKNANSPI LGSFRDGVIKFARLGETITYYYYCSQSAN AQMKNAY
QKAYLASLAK

(69). MH0426_GL0066699_1 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 78)。

MQNLRHTAVGIDLGAKYTGVMIAIDTVPGISLRQEDLTAFTVMPDAEKMMLAQTQRTATRHRRLRGKKRFNLARRLTL
LVINEKLTAAAGVVLPEVRLRMI EAVNGLRRRGYSRIESEVNLQVLEDDLDPQCFACEDLLKSFSEATRS LMDQWE
ELTNDMGRI RDLDRVLQDQTKQDVKEAFKQGLEKEMLSAMEAVASVKEAAHDLVTQQFMGHKPRATYLQHLSD
KRGDSRLTKVIEAFGGVERFHHVIGNISNLQLRAERWYFNAPHMI KGDCWEPCLKQSVIRAFKYFHPSQEGNEKTR
LAELIRELESTDDILEALCTIDPQRTIPPYEDQNNRRFPVDQTLYLSPEKLNDRLGERWMIWAKSFAKADPCLEEL
DEILDLDLDRRSRIARDGTLVLSVLQYRFSYVLQVRVDRSKLLDFAI RALAHGAVLSASRHQLAMVIGGQHVDEFLA
FAAEYREIEGAKSGLWMPQTGSYLSRDLHPPMKKKILPILVGNVQKT DKEGERFMADIWNAKVGGRSTVRSVCA
VIEKTRKDNNGGDFNRLYQRALRRKLGQKLSKEDKVFALSDKVDVAEVIANTLGLSEDEAAKAFANPYSLSQLYNL
IETERDGFTSVSVAAHRENNWRMLTECTDKDGSKTWSANCSRLPADAARPF DGVVRRVIRKAWEI AKRVAESVRR
HVT FQNGTVEIPLFVEENRFEFSSALAEKHNAAARKKMLDRSEGQTKRWQSKI ERIKAASRGVCPYSGKAIKANGE
IDHII PRSLTKKNAGTIFNAEPNLIYVSQEGNQRKGNQGYRLSDLNATYLTAVFRTNSVSEAQATIEKTVTNLQKQG
RLVFFDLLSESEQDCVRHALFLEDFNPAKLAVLEMLGNQRKARVNGTQMWLVRCLVEKLGEEELASWCERTGNQIQFN
ATAVLPVSMQLREALGEVDATMKTQVQPVASHTIDAFCALAVGCDGLFVDEPDLAVPENMVALYPENCISIVQIQR
KPLEEKSQFGGVPLFKDGIYAEHYVPLFTLDGRIYAGYISKPS EDGVMFGAVEVCGKDPVQLLQTLSPYLTRPFEGE
VSKNESYRVDRLKAAALLAEVAHQCSSEQKVAAGVLEGLSYCTKRC PVI GAFWDSVKKKLV AEDVALKDDKFEVKQ
QERRPGHHVLLARNLSSEK PATETSGSDNGRSGIRGACNLSSEK PATETTSLFRTWCATT CNLSSES YFSKLRHL

(70). DOM012_GL0013449_1 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 79)。

VVHSKKIIHPQTEFIMNKKT VYSNIAIDLGGKYTGCVSYTSSILPSADDINAFIIEMPDDGNGIKYTVKDRTOQRH
VVRSLDRFKKARKLIYLVISSIIRRELTADENEAISSLMKRRGYTRLEAEIDL DILKDCPIDLFASYMKDLFDESQS
LYEQFNQKCCDLDSAKAMKALSSEELKQAMDSSELDEKIKYKSAFNTMVEAADNIINTCEFGNKHRKQYLSDIK
VDINKDSRLTDIVKAFASSEKLYRCIGNISNLQLRALRWYFNDI DMKGEKPKWISERFKKWARAYKYFHYPKEDQTK
ARELIKKI ESSEDAISLLSSIDPLKTIIPPYEDQNNRNPPTDLTLLSPKALDDKYADLWDAAEKFANLYPNLTYQL
KDIVKATDRKSRIEQKNSSSYTQSKI FNSYVIQRLDLSTNEDEPITLLRKWVRDPSNPNLARTVKIIQSVVEDHED
DFYKLAQSYNEVEKAKSGLWSVVEDP LLEISGIHPPMKKNKVI STLVGNV LGINNDFDYEFLEVWKSPIKGRKTLK
SICRDIEELRKDYGNTFKVEYDYAIYSAKDKKKT L TSDQKLV EYKTI EEISPLLWSNISQDPLSEIKFNNPFSLS
QLYTLIETDTLGFSSNCVAVNKENNARMQFLSGNCALCSRLPTESVRPFDGAVGKILERQAYEIAKAKSSEIKELDG
LNDAVINVGILIEQNKFEYSASLATIKNSKYKKIKRRAKSGVSRQLEQWKSKEQRIKDASRDLCYPYTGEPISLSN
GEIDHII PRSFTKASMGTI FNSEANLIFCSKDNQTKKENVYTLRDLSPNYLNEIFGLSDILKIQQEIECTVANVNE

KNPRFLFDMMPKEQVCCRHALFMPQSKAYRLVVSSLARSYSTRVNGTQSYFVKV I I SRLRNELSSWLKSNRNTLNF
FAYKVDPSDVHSTRDDLKLNALLKKQKNQPITSHAI DALCVLSAASTVEYI SENISENGKISFLSNIENLKKVVPN
QYRVMRIEKNHCLIKMFLNQVNSLKIRYMQSISLLL

(71). MH0053_GL0028252_1 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 80)。

MSKKTVYSNIAIDLGGKYTGCVSYTSSILPSADDFNAFI I EMPDDGNGI KYTVKDRTQKRHMVRS LDRFCKARKLIY
LVTASII RRELSSDEKEA ISSLMKRRGYTRLEAEIDL DILKECPVDLFTLYLNDLFDENQSLYDQFNQKCCDLDTAK
AMKKALASEELSIALKSSDLGKTEKTLYKSAFNTMVEAADNI INTCEFGNKHKQYLSDI RFDINKDSRLT DIVNGL
GNSEKLYRCIGNISNLQLRALRWYFNDISMKGNPKWI PEKFKKVWVRAYKYFHYPKEDQLKARALIDNLESREDATS
LLFSIDPLDTIPPYEDQNNRNPPDTLTL LLLSPKALDARYSKVWENWAEKFASLYPNLTYQLEDIVKNTDRKSRIELK
NSNSYTKSKVLHSYILQRLFDLSANKDEPVTLLRSWVKDP SNSKLLSTAKI IHSVVGEREDDFYKLAQSYNEVEMA
KSGLWTVVESP LLEISEIHPPMKNKVI STL VGNVLGINNDFDYKFLVWKSPI KGRKTLKSICRDI EELRKNYNT
FKVEYDYAIYSAKDKKTLTSDQKLV E IYKTIEEISPLLWSNI SQDSLSEIKFNPFSLSQLYTLIETDTLGFSSN
CVAVNKENARMQFLSGNCALCSRLPTESVRPFDGAVGKIL ERQAYEIAKAKSSEI KELDGLNDAVINVGILIEQNK
FEYSASLATIKNSKYKKKI QKRAKSGVSRQLEQWKSKEQRIKDASRDLCPTGEPISELSNGEIDHI I PRSFTKASM
GTIFNSEANLIFCSKDGNO TKENVYTLRDLS PKYLNAVFGTADSVKIKQEIDSTVAILSDKNPRLLFDMMNQREQD
CCRHALFMPHASKAYQIVVSALARSYSTRVNGTQSYFVKAI I SKLRNELSSWLKTHNSTLNFFAYKVDPSDVHSTREN
LSKSNPFLKKQENQPITSHAI DAMCVLSAASNVEISENISENGSIA

图30

(1) 伯克霍尔德氏菌目细菌 YL45 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 81)

MESSKYIFPVGIDMGAKYTGVMFLSHLDGVLPTAADATAMTIVNSDKIHYSMKHRTAVRHRVRSKKRFVLRLLNL
 IVDHQLMKAGFVLSAQEKSQLTEALSGYLKRRGYSRIETEIDLDSLNDVDPIVFMCPGFEDFFSSDIELLTQWNSL
 TADLSEVNKINSNKIPDVKTFKKYVKDNFPDYSSKESLERYLEARKTIIEAARQISELTSLGHLHREKYLDNI LHD
 MGRDSRLKKIGEAFFGGICRFWNLIGNLSNLQRLRALRWYFNDPWIEKVKPVETENSQPLNEQKLKDIWVRAYKYFHPD
 PENKARISDVIRALSKADDIDCLCELDCKSTIPPYEDQNNRRPPLDYTLYLNPAKLTEVYGDSSWRWSASRLVNADS
 DLQDNLDVILQNRDRKSRRTTQDSLDFRLSYILQRALDRSKVNDIFEIRKILSSPTLYSEAATRNLFFFKTLEEQNF
 GPFLDFARRYFYEVAKAQKGLWNEKESILERSDLHPKQKKVLDRLVANVLRVDEETAKRFIADVWNGKVKGNSTVK
 SICTRIEQLRKDRGSFAFNLEYQSALTKAALGNEKLDSEKEVVKLQDVSLTAAFIQEQLGLSDEQRTKFNPFSLA
 QLYTIIETERSGFTKITTAALLENAWRLTFKNGAAQCGLPADSVRPFDDGYMRRVLERNAREVGRKRIEIKRDTQD
 QGAEIVVPIILVEENKFNFTASLLELKRLELSDSRLKKKQEKIEKAGEAQKKKWLKSKKEERIKDDAKGICAYTGAPLNS
 EGEVDHII PRSKTIKNYSSIFNSEVNLIFVSRQGNAAKENLYFLKQLNSAYLTAVFGTSDFFSTIESKIESTVQQLI
 ENKRLGULDLLSQEERNVCRHALFLNPGSDARKEVLSVLGARRKTVVNGSQAWFI RTLRSYLENGLQNWLTETGNKL
 SFDAFSIPAVTVSSLRSLQGELENSMFKEDIQSVSSHSVDAMCVYAAACGDASAECVLHGDALLADSHSMRAAQLLK
 KLYPETCQVVQISPKSFSQKDRINNSAIFKDTIYAERFLNIYKKKEEVFVGFNDKKGSKNEKESLTVQVKGKPHLLL
 QILEPFFEKRNESKDEGTYRIDKTKAFQLMGKAGCALLSNDERSQLSLLNQLHYTTLLKQEVLTTELVDTASKKI KLKS
 KYQSNFKREFLLKLELRDRERQFSAGNIVYPSYKRWQELSEELSNFLTACPEGGLEEVNARLSSLWTTSEKSSVRH
 KVKRSMSLPMLASQSGAVRIRRKDFRDGYVYQSIDANQLYGFSSKDKGKILFNSPLPHPRLTTKNLKIYEYQYQPV
 NECVKMDEWKTVYEKEGVIVEHAPGTKARRYIRATLPFDLYLLWDKELQKAKDKSSLTLPKVI RETKDFFEEGIYK
 NEELAQFFPKPRENNKFTMNIELVGSTVKFSYVTA

(2) 弯曲杆菌属物种 P0111 ERR460672.12216_2_83.14 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 82)

MYELSLINVDMGAANNGVFI AITNENSI IYKNAFN IYFDKNTLTFKSDRTARRHARRSYDRDRFI LRLLNEILP IDGL
 SQSQIETIYGLFKNRGFNYHNIEFNESLDDDDVVEFLNGLDRYIFGLSKNKDEFEQIINEIVVDHNSDEIKKILDTQS
 QILINIDKENKKAIKASKNIQSLIQSICNEIDKNNKHRSYLYKDIKELIYKDCFEIIQKSDKFDNLDEFYNFVGNIS
 NFQTRILRRYFNNKYLDNINFDDAKLKKNI IRNINMEYITDIQNNKLNMLNSLKTSALEYLKSIDPVITIPPE
 NRKKNPQKCNLMLNIDKITPSLANLTYKILNSDDFTHIRRDENGQLISLNDINENNI AKYLQRVLDISKESLIDV
 ALYPRALDNNSKI FADTFRLNSDEIREFKEFARRYNEVDNAKGIITNDLLIPCGKNTPHKRNKNELISALFGKN
 ITKDDVINLEEFMLKKNIKGNRSYKGFRRDLNELKKS YQNSFYHNLNNDI INDKDIQTI LELYPKVAQNI NNHNSV
 EFKHPFDKNNLNTNINYLQGLGDI IYDEKNHGFLKTKCKHTLENLIRSGSQNAICTRLPANSARLINGKIEMYLNR
 AYEI STAIKPEILKDISYITINVMNKFSFENSASDLNLI SKRQKSKDMI CPYSGQKIDEINCEYDHILPRSKGLYN
 SRANLIPASSTANLQKGNQSYTLENLHQNYLNSIYKI IKVQNLDEFKNFIDTQMAKININKFTNFDNLNSIEQIALR
 HALFYKNSDSFTKALEILKLDRI KTHSNGTQKRFVNLIQKVKDKLNKLNQDI EFSVNFINAELVSAIRSELSKEDK
 ELQKAKIQDSSHCHIDASIVFYANSKLIKNSKGQREFKYNYNHI KPEYSNKMTMQSKKYLELNSNKISRKKLFDDG
 VYSLVYENANILKDKEANILLDLGLLHTKENGKVAITSDFKSGKFYI STHKVFDLLFKAFNEGDI KLLNKLKFLDN
 HLC SYIRKDI FAI IKDEDKSMFFNNKNEKVPDAKIKTKNIDKFYHILKANESKII EIKDGKNI LQHQEIKELFKE
 CFYTKQAKRSRNSRI IYSLPIKAIKSPNYIIRQNGGYAGLVNSDIATKTYIDLDSKNI IKIPFFSKNILPRKIADI
 INI IELKSKNIKQIYKLLVTKNLP SAI IKLEFII SQANRHNIEVEFDKSOIGDYNLLDQTSRDEFIDKYLKGEFKE
 LGEPDRKKITIIKDTKNSLIITYCTQTS AIHKKIMLDNLI DETSSS

(3) 弯曲杆菌属物种 RM9261 RM9261contig00003 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 83)

MNELSLINIDMGAANNGVFI AITNEDSIVYKNAFN IYFDKNTLTFKSDRTARRHARRSYDRDRFI LRLLNEILPINS
 SQSQKEMIYGLFKNRGFNYHNIEFNESLDDDDVVKFLNSLDGYIFGLSKNKDEFEQVLNEIVVDHNSDEIKEILETQS
 QILSNIDKKNKEAIKASKNIQSLQSI RNEIDKNNKHRSYLYKDIKELIYKDCFEIIQKSDKFDNLDEFYNFVGNIA
 NFQTRILRRYFNNKYLDNINFDDAKLKKNI IRNINMEYITDIQKSNKLNMLNSLKTSALEYLKSIDPVITIPPE
 NRKKNPQKCNLTLINIDKITPSLVNSTYKILNSDDFTHIRRDENGQLISINDI DENNI AKYLQRVLDISKESLMDT
 ALYPRALDNNSKI FADTFRLNSDEIREFKEFARRYNEVDSAKKGIITNDLLIPCGKNTPHKRNKNELISALFGKN
 ITKDDVI SLEEFMLKKNIKGNRGYKGFENLNELKKS YQNSFYHNLNNDI INDKDIQAI LELYPKVAQNI SNHNSIF
 EFKLPFDKSNLNTNINYLSQLGDI IYDEKNRGFLKTKCKHTLENLIRSGSQNAICTRLPNSARLINGKIEMYLNR
 AYEI STALKPEILQDISYISINVMNKFSFENSASDLNLI SKRQKAKEI CPYSGQKIDESNCEYDHILPRSKGLYN
 SRANLIPASSTANLQKSNQSYTLENLDQKYLDIYEKIKVENLDEFKNFIDTQMAKININKFTNFDNLSSLEQIALR
 HALFYKNSDSFAKALEILKLDRI KTHSNGTQKRFVNLIQKIKDRLNSSNQNL EFSVNFINAELVSAIRNELSKEDK
 ELQKAKIQDSSHCHIDASIVFYANSKLIINNSKGQREFKYDYNHI KPEYSNKITMQSKKYLELNSNKIARKKLFDDG

VYSLVYENANILKDKKEASILLDLGLLHTKENGKQVAITS DLKSSKFYISTHKVFDLLFKAFNDS DIKLLNKLKFLDD
 HLSFYIRKDI FAI IKDKDKSSMFFTNKNELKEPDTKI KTKNIDKFYHILKANESKII EIKDDKNILKHQEI KELFKE
 CFYTKEAKRSRNSRIIYSLPIKAIKSPNYIIRQNGGYAGLSNSAIATKTYIDLDSKNIIKIPFFSKNILPCKIADI
 INI IKS KSKNIKQIYKLLVIKNLPSAITKLEFII SQANRHDI EVEFDKSIQIGDYNLLDQTSRDEFIDKYLNGEFKEL
 LGKPRDKKITIIKDTKDSLIIKYCTQTS AIHQKIMLDNLI DETSSS

(4) 拉尼尔弯曲杆菌菌株RM8001 RM8001contig00026 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 84)

MKSL SINIDMGAINNGI FIATCQDDKI IDKNAFDI KFDK GKLNFSKI DRTLRRHTRRS FDRDRFVLRLLGEI LPLEN
 LSNNQKEKLFGLFKNRGFNYLVNFNFAEIDS DNLEILNALDRYF KDCRVLSDFENSISKI FNELENNEI LAF LDDQ
 AQI LNQAKEHKNI LNKNINAI LEFLES IKNESFKNNKHKRSY LKDI KNEIENS FKDI I LK LKI FKNSQEFYFIGN
 IANFQTRILRRYFNAKFDLVFDDAKLKKNI LRNINYLDFI STDEKANKNI LENLNAFSTLEFLKNTDPIV TIPPYE
 NAKNKKPQSCQALVINDEKI TPSLKDIAHKI LSN SNFECLKT DENGVLADNI KDNEIAKY LQRI LDI SKGI LDDKEL
 YPRLLEK DENAKI FAQEIRLNSVELKEFKDFAKKYDEVN LAKKGI IKNELLKPCCLHTPHKNKNKNALISALFLKS
 ITQKEIENLEDF FANNKISGNQTFKGFFERLSELKKRYQNRFFHILEHEPTKELEGI LKNYKVMKIVEHNSNFSF
 KAPFDDSTI STNFN YLTQLGDI I FDEKSRGFAKTCKFHTLENLI RSSAKSAIASRLASNSARLINGKI EMYLDRLAH
 EIIKKIDFTDI YNVQINVMNKNFNFESNLS KLGLAKERQKDEES FICPYSGAKI SQTNCEYDHI LPRSKGLYNSLAN
 LIPTSSTSNQIKSNTRYLLSNLDENYLKYI FAKTQTTDLNSFKAFIDENI SKI DENDFTNFKNLSFFEQIAFRHALF
 LDTDDNSFKKALRLLKLDTI KTHSNGTQKRFINI LIEKISKLTKSGI DLNFIDS KLVSSSTRKDLAQN DTSIAKQEIQ
 DSHSHCV DASIVFYLANSKI IKGQIEPDYDYFKDI YPKSDI I LPQSKQYLELNQKSQKS KKLFDSDSVYSLVYEGKF
 DEKQKKI LTQYELLKENG FVDTHKVFELI FKSFKDKDTIVLKD LKFLDNHLKSYIRKDI FSIFPNEKGELKEAKELK
 SATKNIDKFY TILSQNQDKI YTI KLDEKAKEKKT LQRNKVIELYKEI FYTKNAKPRARGKNRVTFSLPVYAKDAKYA
 IKRNSGFAGLKNKDIATKTYIDI DSREILKI PYYSTNVLP IIRISDILEILSLQSKNAKQIYKIKIQDNL PKQIVSLD
 FILSQANRHEVLVSFDKSKLEIQNINDDNGENTFIQKYLCDDFKEFLGKPRDNKIEI IKNDDI LAIKYCVESTSKA
 NKEIIVKALSDETPSA

(5) 拉尼尔弯曲杆菌菌株P0121 ERR460911.12227_1_93.7 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 85)

MKSL SINIDMGAINNGI FIATCQDDKI IDKNAFDI KFDK GKLNFSKI DRTLRRHTRRS FDRDRFVLRLLGEI LPLEN
 LSNNQKEKLFGLFKNRGFNYLVNFNFAEIDS DSLEILNALDGYF KDCRVLSDFENSISKI FNELENDEI LAF LDDQ
 AQI LNQAKEHKNI LNKNINAI LEFLES IKNESFKNNKHKRSY LKDI KNEIENS FKDI I LK LKI FKNSQEFYFIGN
 IANFQTRILRRYFNAKFDLVFDDAKLKKNI LRNINYLDFI STDEKANKSKILENLNAFSALEFLKNT EPIITIPPYE
 NAKNKKPQSCQALVINDEKI TPSLKDIAHKI LSN SNFECLKT DENGVLADNI KDNEIAKY LQRI LDI SKGI LDDKEL
 YPRLLEK DENAKI FAQEIRLNSVELKEFKDFAKKYDEVN LAKKGI IKNELLKPCCLHTPHKNKNKNALISALFLKS
 ITQKEIENLEDF FANNKISGNQTFKGFFERLSELKKRYQNRFFHILEHEPTKELEGI LKNYKVMKIVEHNSNFSF
 KAPFDDSTI STNFN YLTQLGDI I FDEKSRGFAKTCKFHTLENLI RSSAKSAIASRLASNSARLINGKI EMYLDRLAH
 EIIKKIDFTDI YNVQINVMNKNFNFESNLS KLGLAKERQKDEES FICPYSGAKI SQTNCEYDHI LPRSKGLYNSLAN
 LIPTSSTSNQIKSNTRYLLSNLDENYLKYI FAKTQTTDLNSFKAFIDENI SKI DENDFTNFKNLSFFEQIAFRHALF
 LDTDDNSFKKALRLLKLDTI KTHSNGTQKRFINI LIEKISKLTKSGI DLNFIDS KLVSSSTRKDLAQN DINIAKQEIQ
 DSHSHCV DASIVFYLANSKI IKGQIEPDYDYFKDI YPKSDI I LPQSKKYLELNQKSQKS KKLFDSDSVYSLVYEGKF
 DEKQKKI LTQYELLKENG FVDTHKVFELI FKSFKDKDTIVLKD LKFLDNHLKSYIRKDI FSIFPNEKGELKEAKELK
 SATKNIDKFY TILSQNQDKI YTI KLDEKAKEKKT LQRNKVIELYKEI FYTKNAKPRARGKNRVTFSLPVYAKDAKYA
 IRRNSGFAGLKNAN IATKTYINI DSIEILKI PYYSTNVLP IIRISDILEILSLQSKNAKQIYKIKIVDNL PKQIVSLD
 FILSQANRHEVLVSFDKSKLEIQN IHDDNDENAFIQKYLCDDFKEFLGKPENNKIEI IKNSDDI LAIKYRVDRSSKA
 NKEIIVKALSDEAPSA

(6) 鼠毛滴虫菌株YL45 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 86)

MESSKYI FVPGIDMGAKYTGVMFLSHLDGVLPTAADATAMTIVNSDKIHYSMKHRTAVRHRVRSKRFVILARLLNL
 IVDHQLMKAGFVLSAQEKSQLTEALSGYLKRRGYSRIETEIDLDSLNDVDPIVFAMCPGFEDFFSSDI ELLTQWNSL
 TADLSEVNKINSNKKIPDVKTFKKYVKDNFPDYSSKESLERYLEARKTI IEAARQISELTS LGHLHREKYLDNI LHD
 MGRDSRLKKIGEAFGGICRFWNLIGNLSNLQLRALRWYFNDPWI EKVKPVETENSQPLNEQKLKDIWVRAYKYFHPD
 PENKARISDVIRALSKADDI DC LCELDCSTI PYPYEDQNNRRPPLDYTL YLNP AKL TEVYGD SWRS WASRLVNADS
 DLQDNL DVI LQNRDRKSRRTTQDSLDFRLSYILQALDRSKVNDI FEIRKILSSPTLYSEAATRNL LLLFKTLEEQNF
 GPFLDFARRYFYEVAKAQGLWNEKESILERSDLHPKQKKKVLDRLVANVLRVDEETAKRFIADVWNGKVKGNSTVK
 SICTRIEQLRKDRGSAFNLEYQSALTKAALGNEKLDSEEKEYVKKQDVSLTAAFIGEQLGLSDEQRTK FANPFSLA
 QLYTII ETERSGFTKITTAALLENAWRLTFKNGAAQCGLPADSVRPF DGYMRRV LERNAREV GKRLIEKIKRDTQD
 QGAEIVVPI LVEENKFNFTASLLELKRLELSDSRLKKKQEKIEKAGEAQKKKWL SKKEERIKDDAKGICAYTGAPLNS

EGEVDHI I PRSKTI KNYSSI FNSEVNLI FVSRQGNAAKENLYFLKQLNSAYLTA VFGTSD FSTIESKIESTVQQLI
ENKRLGYLDLLSQEERNVCRHALFLNPGSDARKEVLSVLGARRKTVVNGSQAWFIRTLRSYLENGLQNWLTETGNKL
SFDAFSI PAVTVSSLRSQLGELENSMFKEDIQSVSSHSVDAMCVYAAACGDASAECVLHG DALLADSHSMRAAQLLK
KLYPETCQVVQISPKSFSQKDRINNSAIFKDTIYAERFLNIYKKKEEVFVGFNDKKSNEKESLTQVKGKPNHLLL
QILEPFFEKRNESKDEGTYRIDKTKAFQLMGKAGCALLSNDESRLSLLNQLHYTTLKQEVLTTELVDTASKKIKLKS
KYQSNFKREFLLKLELRDRERQFSAKGNIVYPSYKRWQELSEELSNFLTACPEGGLEEVNARLSSLWTTSEKSSVRH
KKVKRMSLPLMASQSGAVRIRRKDFRDGYVYQSIDANQLYGFSSKDKGKILFNSPLPHPRLTTKNLKIYEYQYQPV
NECVKMDDEWKTVEYKEGVIVEHAFPGTKARRYIRATLPFDLYLLWDELQKAKDKSSLTLP SKVI RETKDFFE EGIYK
NEELAQFFPKPRENNKFTMNI ELVGSTVKFSYVTA

(7) 沙姆盐弧菌菌株 CBH463 scaffold11.1 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 87)

MSDNALI FPVAIDLGAKTG VYAAGYSPNTHVGFNSDSVFKTG FVAEVDGQGTGYQLLQ TQRTSTRHARRSRTNRN
QAKRLLMLMLKCCLGFDTDKHQIAISHFINRRGYSYVETNIDEAAINEIASDNLGTFHVLSENGHEALSELLTQTM
PVDGLPLFDALQRI AEDKQQLITLQTELAQFTRQDKKQQRALLPLSKAVDFLIAEQSSGAKHREYFKNIRSDLKKL
SDHPERHCRRLFHALHHKKTGSDTFDSFYQLLCHINNF DLKLLNSILKRVKSTQSGEQIEVIVA EKFQKWLTKQW
RLSSANGADRLQOIRDVQHAWKDYRAQNPQSVIGFFLDTPPECTVPPYESHNTNRRPPSCQTLTLNPDLLSRDYPQWR
DWLAWLSSTPAMANAVQAYRTQLVAVESAKGNTLIQTDEL DARTLQFI FDSAQDKDPFKLNAIWGTWKKQRQQRQG
DEHDKTTDKLAKLLVQSALPASLTENFTSHVPVPGDFWHLINRYFQSRRRAREGRYFLHYDNHYPKQHRWQRDGLLV
MCQHRPRQLKHQATRDI AVLLGVPQTQLQQA VDNQCDDNKALFAKVKGLQAACRQAYQAQKAWGPD LKAMMVSDFE
LKKLAEKLAPLTA KVATQLALRDEQYDSFCERNQSVTFQAQLFQLVWGDRSGFAKTC PVC SADNANRMAGEDGQPHA
ARLNTLTMLRIDGGLKRL LTHQAHHIANRLWDHIEHIGQSADTITVPFVIEQNRFDFTENLPTL KGLNKPKPQPQKI
NPEQAKKDRIKHASQGLCPYSAGSSVGDNGDIDHII PRSQGYGVLNDEANLIYSSIKGNRQVKKDRQLTLDLNPDI
LKAQFGTADPIAITQQIETTLCGDDADRFSFGRYSQFAALS AEQQTAFRHALFLSAEHP LRMQV ISAMQHRNKAKVN
GTQRFMAQLLADI LTKKAKQKGLAARLKFDFYQV SANGQDESDTVALRRLMRDLHAGTKYDIQPF EKGKEKQHAYSH
VIDATLAFMIAAEKHQEGEGALNIHF DENVS IWGQVDENG VLS DALFPQIAVAPENLAPNVRVGPQ SNEHRAFRLAG
KKPHQVFSQPIFKQNALGLEFYDLAVIDGQLYKGFVEGEDKHFIKANEKVVDKQLNTFLYAVSHHYR LIQSSDIQV
YTADKHKLTALLFETLAAAKDATT FDSNAEANKLARWVCGKSAGSLYFYSTSAGLEEAPSVVKKQPN SRFRKQWLAH
YQSWQAKHPDTP EKNGMWII PTELEADWRQHCDWYLG RPTDDQQPDHRAVKS DTMRAETKSGAMALVRRQS QGKAI
YQLLALDTQKVPKDHSAALALASPNLVLLDKAIFTKGYQTKVEEQQLAGKTI ESSAFFDQAI SEQWQLDI STLQVA
IVSNTNVQITGLHKDWVAEHLI LSGEDSEKAWTKLTSIAI SDEAQVMDVFAKASLKAWLTAPCRNDKSVRVKLD DDDQ
VSLTLPYKQPQLRKLVD

(8) 钩端螺旋体属物种分离株 FW_030 FD_1000036_176051 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 88)

MREGEGFDMGKEVISP I AIDLGAKN TG VFAATYV PGEFLTNPDI I KEGYLLQMGDIQLSQADRTV KRHQLRGNRRKK
LVRRLFLI LKKEKFGINEFSKSTSEFLQGLLKRGRFSYISENIN EDVLKSADLEFFEFTQ GKITKKSPLLNQLFDI
SQNLYDAQSLLDCEI FQKNKKDIKSFLRTQEI EVNDDLLDQIESTKSYLETI IKSDIDGHKHRDYIKNISKDVEAK
WKSIQEFSSLNMEPLDFSNLLGHI SNLQLKPLRKYFNDFQMAEGNDY WDEERLGTI I RKWLKSLR PSEAKDIQNLR
LLSKIRSNESKII EILLKTNPELSIPPYEDMNNRRI PSCQTQYLHTPALNQLYPRWKI WTDKLVKDLTDGKELIADA
GFDDFRLDSELI I LPKVGESYKEIQDRKQARLLQ RVL DLSKKVDPYSIRSQSVETKKGLTKKEKSELAGAKDRDLK
LDKDSDAFIGMAKDYYNQLKTI RQGI ELS EDISL FHSCEEKTKKKGNQTESIRTI LGINNISDERLETIMDTVWKD
RLYGNTTLRSYCN SCEEARKKDGTAFI LTLSSVKKKKLNIN ITEDHEKKVSGLLDNMHLAVKKICEAFPEIQEISS
KRMERDWEKDYNYLPFANLFS LAQIFNILEGDTRGFSKICNQC IVENKWRGTIVRSGSNLDQGAIASRLNSDTIRQ
FDGVLRRMTDRLA FEIAYKKKEELKDINRNSKII IPI LTEMNRFENL DLGHYKLEARIASRMQKKG NQKNEQALER
LESFEDSKKDRIKEASLQICAITGKAI GENGQIDHII PRSHSKSNATI FNTEMNLIY LSTGGNVQKGAKFYSIADL
HPNYLQAQFGTKDKEQVQKQITEGIESILRIKRKVFRSYPDEQKILRHS LFIPELFP LVIPLL GEGIMAKANGTQA
YLARCLQKRLKDI LEKEGFKNFEFSIFHLRSEESI SDIRKELS NYKQDYSKPGKQQQAPGSHILD AVSVLVGSLGNP
NIQSI LKTDI FSEDKGMSKEEQAKSIEACLPSQIRILKI DLSLPKYRKNLGSKSLFKEGIFGERFVPLYLHKEMLF
IGFDTQMKSGLEIKNNPEKVF DALKPYFMIHNISYNQAE LSKSII RFPIDRKKAVRYLFD SVKQNTWNDVSEILQ
GLRYTVSKKDIASQIVDKNSKI QDKKEI LKETQKNIKVEI SKLELTGELLKGVFEYPGITD WNTFLKDFYKQKGIKT
GEKLNFEIYDKFCREYFNI I KQDKQH QKVRKMFSLPVIDNPSGGFRI KRKNYDGT EIQVATADDGYAAGFTVIDGN
VEIYNKKLGRALPSLEKAESISTLQNKNSSDDFVFMDDWREIDI SEVETLKRLCIQPNSLPRARIYAEVNTIDFKK
VCENIDSLSIYKIPQVLKLDNIKISENFKILKEPRGDIRVKLIDAKKGI LGIEYI IQSTDKAISELYNLGKPIGSP
YATPDSN

(9) 莫里特拉氏菌属物种分离株 NORP46 Contig_来源1382A_1689 Cas9蛋白的氨基酸序列 (SEQ ID NO: 89)

MKGLVLKII SP I AIDMGAKYTG VYLTQYEQGDVLENANPEGLVVMHTENMQYSQVDRTQRRHQTRTAKRRKMAKRL
WLVLAEHYNLSFKNCDVKERAINSLNRRGFTYLS EDVDEALLQEI DSSAFHLYFN SDI PAATNLFEVYQVITTSV

ETAKSFQNSDEYAI SVKDFSKRLEGDYKEQKKQLESFGKAFRQGVDSFINAEQDGHKIRSAAYLANIKKDLSEHEDYQ
 FLRTRLTVSI DAFSNFVGHISNLQLRVLRRYFNNDENMAYGDIWDFERLHKVYRNVSGMHYKDNTQDKQNVTLFSM
 RAQDIISTFEALDPIVSI PPFEDQNNRRPPKQSLKLSPTTMIKYLPKWHIIVPQLADDLLEARSILVKDGLLELSIK
 NTNDWAQCLQRILELSYDHPYLLRRWIGRKDQVGEHTKRAISNFERVIQQQHADFIVLAEIFYAETQAARNSIWF
 KENSQLELVCAQKTPHKANQFKELLSGTVNVNLTEDQKEQFEKLWFDNPRIGKRGIKGWCGFAAEMQKEYGNALKEQ
 INRNIWLNENDKKVENKKLIELNKAVESVAKI IADSMNIESKRFANVFDLAKLHNLVEKDQKGFSSNCRCTKDNW
 RGNIVENDEGKDAAFGLRLPADTTRPFDGLLARLIDKQAYQIAQAKVAQLLSEPTNEMVSVPLLI EENKFRFTSSLA
 EVKKNTKKRKDNEDKAKKREGLWLEKNGRI KEASKNICPYTDKPLYNGGEIDHII SRAESKGRKQTVFNHEANLIYC
 SVTNGIKDRQSYSLADLKPPLYTKQFETS NREKIKQRLIALCEPI IAKRELGSFGDLNEDTRQAIRHGLFVDEL RP
 KLI PLLHQSNKTRVNGTQAYLAKLILQKIKLLYKGP IEFVHYVEAE LASRLRVQLGNQYPEFAKPEHQPVASHVID
 AAMVMAAGLQPPHTQACLQETAGLDNNGDWLKA ILPTEIDVKRI EPKLFKNKQVHSQTLFKDGIYAEHVIPLVND E
 KFGVGFTADNVAWLKNQEAPVIWVWALQLYLNKSPETLDDVRKNI GNQTKVYSFNNTAVFELLYKVAQPCSD EEL
 LSADLLETLIYTSQKVKLQSALYDENNKTYKKITDIVNMKTTDNKGDFTIKLDVKSRYLKTMMRAFESSLIVYPAIA
 SWSNLEHDFVKAAGKKEAFDSDSYNRVIKGLFSTIKENTRQHKKVRKNWSLPKLS SPSGGYRARRKNADGSI IWQ
 LFAVEGLSASGFYIMIDGKIDFSEAGVAGIKAI DDKNLT SVKSRYKNIQLPVEPFDNWRVVTIPDKLKGKIKGILLS
 PNSKSRFRVRIEFSYQGFKENVLP L TDSIELVQAWSQIPAE LKMAKGKEWKASFGDLIGTPRSNLFTLVIGT SITLE
 YIVGSSNTVMKQAYQNGSLID

(10) 内生单胞菌属物种 S-B4-1U 分离株 SB41UT1 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 90)
 MRYISPI S IDFGAKYTG IHNMHYLEGEDPRGSVQNTGCVLVMDDDNLQLSQVGR TQKRHQSRNYKRQDLARRLLI VI
 LKDCFDQDVWQWPFVKNREFLFG LINRRGFNRFD DDSLIEIEFCDSLVSFESVLPDIAGYEGLEEWLQALAE NAAECRL
 ALQRLGRTRFLKTLPEIWKEEKKSI EATHKHI IHFLETSINASEEGLHRRKYL RNIAADII LEQHGRLASLFQTN
 SGITPDQFANLVGHI SNLQLRVLRRYFN DQNAKPNQNGQTD AUREDHF SKQFWRWIKSWHCDEQT KSRWHELIKFKKK
 GGSVFKMLTVQPELTI PPEYEDQNNRRPPKDKTLFLN PVALNKAYPLWREIVKRLVA AHPELSDHASTIAHMDAAL
 DQDCILLQRMLDRTSGLDLYKIRALAMGRPDRGNQLKVAS ENLRLRLNQHAEPFIQS IARPYLQIN EARQGLWVEE
 DSLLLCLNRTPQRKARQLPTLLGRIVGDLGSHGQEHIQSLKQQWSQRGVI KGARTPKGWCKFAAEAQKDFGNGLKET
 IRSYRQLQARIKAGDTEEKLLPEAKKLLALEAALPDVARFIAQGLGLAAENSARFANPFS LAQIYNLLES DPSPGF SH
 VSKPSAHENLWRMQYEQSV DGDKDCARASVLSADSIRPFDGMLSRI LEKQAYSIAKRK LKQIESHEMTSAKVRVPI LI
 EENRFQFAEDLQT IKKNTRKKKDAQKSCESDFIRFEDKEQRICKASKGMCAYTGEDLG RYGEFDHII PRSYTQKRTG
 AIFNSEANLI FVSSRGNQQRDQLYTLAE LDDRYLTSQFGSTNRQQIAQAVQDAVTPFLT KGE LIGFHNLGLQE QRM
 IRHALFV PGLRDQVLQLLAGQRKAKVNGTQAWLAKRI IHHIRQGASATIEFDI IKVPTDSVRFCR GVLASRFPAMAK
 QSPQPVASHAIDAALVSGSLLM E PAIARSLSLAPVD TTTQAEWLKGLLPETIDIQILRKKPL YRKDAPASSPLFKAG
 LYGERFLPLLVMNDEL RVGFTLPNSAKVLDHADNLYALLRPVLVCKDSHEL PSTLAELQESLALGQRQILNINROKA
 FAYLHEFGQSPTHNDNCLAI EALLYCVQKKI DAHINYKDKKTFKKDEDI LADKFFI FETHKNRPI FPRTVARI EK
 KDLTLPFKQAWVELCDTDELKNFLGTKQTL PNNFWRNFGSSL PPSIHNEHDHQHKKTRKVYS LPLKISEG SYRVRRT
 PEGASVWPQMSVGD FGAAGFAKI DGKTQWNKVVEI PHFAKSKNLHSLNRSLEKNPQS ITYFDEWRS L CALADALPNG
 VIKVACAPGTVGRPYIRFEMLESQFIKKINAASGSKSSWELPMEI PNVQLAGFDI FSCPLL PKPRTETKPKPKHK
 ARFLRLGKTVLIEYFASSKANKALQQAYEQGEPECI

(11) 喜盐泰米尔纳德菌菌株 Mi-7 Mi-7_contig-1 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 91)
 MCNLYLKAHSENDYLGESTMTKLVSP IAVDQGARYTGIYSLSYEPGVDPFLESKNAEVVVVDNDKLTLSMAPRRAGR
 HLRRGNQRFKLARLLTLVLREAFGVDPDQLDQREREFLFGLLKRRGFTFLGEEI DEEAI SGLDPPDLVKRLLRNCPF
 FESLES LWQSLSSDPGMCREFLQLSTGSDDSL DLETLDKRWLKKWLKDNVEEASDKETIERLTSGLVALR DGCRAVI
 RAKWDGHHPRHIYLENIAQDLKHQPEGNALADRI GLSTGELSHLIGNISNLPLRPLRRYFN DKRFSRNADR WYPERM
 ARQLTREWKGWQPDGVSKDIVQRSRELLKGLSNAADPVRFWKYTDPLQTI PPYERHSNRRPPKQSL LIDVTALGE
 RWPEWERWIAPLHKANPELSEGI DGFSPSISGPNEPRLAFLQRVLDRSTLFDPYGLRQLAVESRS HRAAQGKETLFR
 DLGDQHARGFLS FVTAYHREREQARSGFWDNQREGRL LKRCHTNPPHKGKILHLLLNRLLDGDI EDVEGFRDEITKT
 KIGRSTLKGIAQRASDVQKRYGNALNATLKEADLDKATYKRLNDLQALIEKAGPEI GRIAGCPENRRNRLGNVFILA
 QLFNLLFGDPSGFSSTCRSCSEENAWRSRSVPVPGEDREGWRMI PLTADSVRPF DGVVRAVDAQARAMARLKKQEL
 KKRDL PAGSFVSI PVLLEQNRFSFSEDLAAIKQRIAPSARHRNRQKAAREAEQEQI SRWRGKLDRI RDASHGICPYS
 GRDLNQVNGQIDHILPRAQSRKLDGTTWNHEANLIYCST SANVDKGETLKNLEDLDPGYLEAVFPQSLTSIRQWIA
 DRVKPLLARSGRMPSFVELE RDQSIALRHALFDEDLRPQVMELLRGQARATRVNGTQAWLARRFVEELRRSVPELTL
 SFSVNRIQPEDVGLMRESLSEIRPEFLKAPDQPLSSHAIDAAMVYAAASSLPRMESELHVPAP EPLAELHDLRLARLLP
 ERIDFTDVS RKPVYSLPRADSQTLFKD TYAERFLPLLVS SDGALRAGFTPENSVEVKPARKRSESTPAEIAI SRLD
 GFLRLRGARWSGDLSSLFASDASSDQWVALSVDRPLALARLQQVGRGEDELTD DALLGALKYLTNKKELRTAVVGT
 GSALMSWEKLN EIEKKT TSVS IKGAGLKS DGAKLSLPARSHWLRF AARLNEEGVLGSNPGWSDAEWRRFFQTVFP
 SAKRDRQH QKVRKVWSLPVVP GS PSLVVRHQRRTSRGQTVFQVASLHEFKHMGVVL RERKAEGVFPGYLRWSTNLAP

LAGVLQSPSQRMIRFDQWYPLDRPVSQCASADILSVSAAWGSGLRRSVEVTVSQELACRVFDVNSAWNLPANLKLHNN
ASELWPIEAIGWFRNNKVELRAITPSGVVLRVYQKDKGGKPIQQELEIRAIREEE

(12) 需钠弧菌菌株 CCUG 16373 染色体 1 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 92)
MYAASYPAGTTLLKDFSTDMVSKIAFAATSYPADKGGYVLLQNGRTANRHMSRNRTNRQAKKLFKTMLECLVGFDFVE
KHKEAIAHFLNRRGYSYVERQVSLLEDLAALTTEQLPLMRELTKLGHKGI VEEFELTDPIEALRCVASQGAFFQLEAL
IEAFKKLPVKQSKPMLALRDSAKLYLKDQTDGALHRSEYFENLAHDFNKLKHPERHCRRFFHALNDLKQAKHGYQK
DDVLRLLCHVSNFDFLKLNRILRALPESPTETLLNAVLTEHVSKWILKQWVLSSENGQARLQEI RELRRALKQHLLTL
KPNEIVSFLNTPPEQTI PPYESHTNRRPPSCQTLI LNPKCLDDAYKDWPNWLS ELIQASPRKSEIYSYREQLAALI
SNQEQLITTEELDARTLQLILDTSKAIDPFKLNCIWVSHLKKCHKQARSQDTS SVIHKLEEALKSSVLPFAFKQVD
MAEALQGSFWHLVNRYYQARRKCRDGRYFIPIDKTKQSRHKKRWQNDGKLLAMCRHRPKQIKHQALQDVRVLLGLNQ
QLPKKVVLNQNGEVLDLQAYLLQVKGKACDDARKAQKQWGAELNDNLTDTNTLIELTAKIQVLLPALSQHLQLDSDS
SSAAFIERNQSIYQFAQLYQLTWGDRSGFASTCPVCATDNAVRMADNSAKAQARLSTLSMHLIDGGLKRHLNHQAH
HIANRIWDKLLKPKALLKGDVMTPLILEQNRFDFAENLHMLKGTLPKPKGSAPDQFQPKVERIKQASHVFCPYDKQV
TSLGDDGDVDHII PRSGRYGTLDNEANLIYASRNGNRKIKGNRELTLSDLSDAYLEHQFATTDRAAIQHTIEQRLTG
EREIEFRFGRYRQFRALT PENQVAFRHALFLPADHPLRVLVI STLQHRSKSRVNGTQRFMAQLLADI FWQKAKELGK
QGNLAFDYFEVSSDPNNEASTVALRRHLESSTLASGFDLQPFQKQTYQPQAYSHVIDATLAFMLSVQAHPNEGAMR
IHLAKGETIWSQVDEGDVYLNATAFEQVVVLPQLSPVVEVKPRPHTTKVKQLLDGAKPHQVFNRRIFNQNALGLEFY
DFQVISGQLFRGWVVVDNGQAEFVKHSGDKSKPADKNLDKIQRVLELGLFRLQRIQGLPIYRANKQALLGYIFDALS
QAKVAPQPKQEQLVSWLLGKAGNLYYSSKELTKAPSVLKS KGASPFKQHWLN FYTGWEYRAQHPVKQSNVGL
EIASDEQANWQAYCETFLKWTPIQETNEVFQPAHAASRAFTMRAPGTPSGVPSLIRKLSAEGSKSYNVMIDINGVLG
KEQAPALALASPNVVLFRNVLQNGYMADTSPKQVLEATEIPVRAALNTENCNAGLDLESIKVVI EANKLHIKNIE
RQWFETNLLLENEVAEKPWQEKLRIFCSGQTKAYGDHAVAEDSFKKLLKRS AKHDKGMALKVEDDKVSVTLVLDKKV
TQKILEEITQN

(13) 嗜淀粉瘤胃杆菌菌株 DSM 1361 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 93)
MSNTYYSNIAIDMGKTYTGFI SYTSNDLPDSENVKAAI IEMPDGGSVINYTVKNRTAVRHHLRANNRYKKARKLI FT
ILADTIKRNLTSQEAEISSLMLRRGYTRLETEIDLEQLKECPVDLFFQCCPGLFTGDVPLYDQFINNCNDDTVVYVY
RESLSSIKETVSNLSDKTEKSMYGSALKTMVDFVDKFI DQKEFGHKHRT EYLANIRKDIQKDSRLNDLIK LISGEDR
LYNCIGNISNLQLRALRWYFNDRGMSGKAVFDKRFKDVWYRAYKYFHYTAGSNQKDNAEKVDI IKKKNFLKRIEQ
TENITDLLLAQDPVFTI PPYEDQNNRRPPVDQTL LLS PQSLDAHY PGKWELWVQKFM DYGSYSLS SDELEQIVKL
TDRSSRLQSAGGKQSDHIASYTEQKLIHSYLLQRLLDITYNRGNPESQIRKWAQNPNSCSATNAVIKNI LGESDIDS
FFELANNYYREVLSLKKGLWSIVEHPLLEISGHPMKS KMLEELVAGVLNIRKGFNLERFKEVWNSTVKGKSTVRS
LCSYIEKTRKKYGNCFKQ EYKNVCRYEELKNKASEVHFSSDIKIDKEVVKVY EYTKNLSKFI GEKLELSTDQSKFA
NPFSLAQYLTILET DVHGFSGNCKAVCLENTRCQM QSDKGNALREFCALPAESTRPF DGSGLNILD RQAYEIAKIAIE
IKELSALKNTKVKLSVLVEENSFEFSSSLSTIKKLSKAKQLRELFALKA EKTEREKWKSDGERKNDSHNICAYTGM A
LSDDLSENDHIVSRSETKESYGTIYNSEFNLI RVSRRGNQLKGNNIY TLDNLPKYLEAVFGSDTAIITTRIRETL
NRL LISNPNLNI DIMTEERQCCRHALFTSRTGVPFNQII QAISKQYRARVNGTQAWFVKNLIAKIDEGLEWKSIN
NITVEYNAYKIDAGSTSEIRHDI GKLDARYEKDIIQPI TSHAI DAVCLLGNASTQQEISEEISGNNEISVLS DVKEL
IKLIPDRFDIIRISSLNFAEKNNPESKQLFKDTIYAEHFLCIMEKDDTVKVGDFDGSNSIQIVKGGKQLLSCLANYL
ESPQNETTGFRTYKVCKTA AFKLFERFYRGLVAEGSDEELAYLALMSLR YTTLHQDVFSLLYDVKSKKFAAKKDIL
KNLDIKISFPSVLKKTDKASLVLP SFNEWAHIASIFEDYLGKKDEQESGYDKLQYLG TI LNKTINKNHAKSKRVY
SLPVVASPSGGVRIKRLTHTHGDIYQLVAANTPVTSISKGFASADNGDVLWDEAVMVD SYTTKNLTVINKKISGDKN
FVSMDEQRLVYESPDERVYMPATDIRRGITIEQKFEK FVYCIDQQNIYKSFHDLPT EIKTDPKIFMERLNIDICGK
PRGNLKI LSIGNIIKYSYMVESSNRVMNQVYNNPIDKGTIFKSN

(14) 萨迦弧菌 NBRC 104589 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 94)
MSTETKPWQDDDKLLAMCRHRPKQIKHQGLQDVRVLLGLNQQLPK EAVLGLDGKVDLGAYLLQVKGKQACDDARK
AQKQWGTELNDNLTDTNTL IKLTAKIKELLPALCQRLKLDKSSSSAFIERNQSIYHFAQLYQMAWGDRNGFASTCPV
CATDNAVRMADNSGQAQAVRLSTLSMRLIDGGLKRHLNHQAH HIANRIWDKLLKPKALLKGDVMTPLILEQNRFDFA
ENLHV LKGT LRSKGLSAPDQFQPKVERIKQASHNVCPYDKQTTKI GDDGDVDHII PRSSRYGTLDNEANLIYASRTG
NRQIKANRELTLANLSDAYLQSQFNKEFPEGNI DRDEIQR LIEQRLTGKRADIFSFGRYHQFRALS PENQVAFRHAL
FLPADHPLRALVISTLQHRSKNRVNGTQRFMAQLLADTLWQKAKELGKQGNLAFDYFEVSSDPNNEASTVALRRHLE
SSTLAVGFDLRPFQKQTTQS QHGYSHVIDATLAFMLSVQAHPNEGAMKIQLAENESIWSRIDDDGTYLNATAFEQVAV
LPKELSPLEVEVKPRATHTKVKQLLDGAKPHQVVS RRI FKKNALGIEFYDFKDIAGQLFRGWVTIDNGQAKFTKHS GD
KSKPADKNLDKIQRVLELGLFRLQRLQGLAIYRANKQAVLGYIFDALSQAKVAPQPNQE QSKLVSWLLGKAGKLYY
YTARKELKKAPSVLKS RGNPFQEEWLSFYTGWEQRSQLP I KKNKGMLEIASDEQANWQTYCEAF LNRIPIKENDQV

FQPAHTALRDFTRAPGTPSGVPSLIRKISGEGTKRYHIMDINTRVLGEEQAPALAITSANVVLFDRNIIQKGYMAD
ITPKQTLAIEIPIEAAALNSKNCQVAELDLKPIKVVIEANKLHINNIERQWFESNLLLKNEAVEKPPWEKSKI FCSG
QTNAEGDHSIAQDSFKLLKRS AKHDKGMTLKVTDVTVTLTLDKKVTQKIVENMTQD

(15) 豕弓形杆菌菌株 117434 AAX28 Cas9 蛋白的氨基酸序列 (SEQ ID NO: 95)

MGAQRLPADTQRPFSGKIERYIDKLGYEIAKIKAKELEKVAVKNIKIKIILEENSFEYEESIRSAKIKNAKKNANVS
LENSKKS YEKSLEDKASRIK NFGNNICAYCSKEIVNNDGEIDHILPRSYTLKNYGTVFNSEGNLLYVHQKCNQDKKD
KIYSLDKITAPISKEQIISIVESIKSYKFTLLSNEQQIAFKYALFLPHNDEAFKKVAEFLRTDQSARVNGTQKYLA
KKIQTKLNKMFENKKELDYEFILADSEVVSSLRKQYVKEEPLLAKPKDSKQASSHTIDAILSLVSVYKKNATLNVE
PNASEIVKYANIKSWSALENEFLTGGKSTNQKIENMIETESFTQKNMREVFNKSIFGENALGERFKPIVINSNDTIYI
GFPIKIKDGYSFENCKQITSKNDKHMVENIKSELVSKDKSTIVIYEINKNLFNELSNEYFNLNKYNLADNKKEQV
AMLELIIKFSKFYTKRANVIPSAPSYKKQIEEMKYPFYNEWKLFDEAWKHLMINKNKKGTCDRYVIDKKQENSHYKID
LNKGEYKLDTDNRSLWDDFCWKYFLDRYAKSESNSIRRKARKNFSLIAEGTPQGTMFRLKRKTPKGFVYQALPIDNK
QIAGDYSNILLHNNKGSISLSSNAKDLTKEFEIKVSKDIRDTKLHPAKFFKDDFNYGKIEVILNKTSATIKNFPL
DYFINSFIQKNEEFETNLNSLSFVKNSEEFVGIKSKLDQILKMTTRGDQKINQITINDNFIDFILPYKNTSKLLDE

(16) 脱硫杆菌属物种 PB-SRBI Cas9 蛋白的氨基酸序列 (SEQ ID NO: 96)

MENETYISPIAIDLGAHTGVYSLIYKAGSTLADVEEKSGVVYELDPQNNLLMVSRQKRHRNRNLDRRQMAKRLF
RLIWEKHFGQLQWDADAQAQSIGFLFNRRGFSFITEHYDIQALANFEEAWQELPRAFTDLIPVDLRNS SGLAEFIMAL
PRDEPLSNSCLCELYKETNKIKKELVFTQKSI ALAKACEAIGQGRGVPEERRKKE SPLASVSKWIVERWRDEGVCGL
EAVEPTGPVYNIAGFLTEHPSLAETVTATLPDYRAAQNESKQSMWNFAIDKFKLEEA EFERTGETTSEKQDFIRTHL
HHLAFAYTISDELRS GARHRARYFEEVEKVLGTIHRHG YLQRF CGS LQ QEHFAGLTPAKLTRLISHISNLEL KPL
RKYFNDKKHQGGFYWDEERLKRCS DWVLGQWRV DLAKNRDKGAGKPDYQKLRTVLQNYEGGIEFWIGHDPQLSI
PPYQDNNRNPPRCQSLLLNACFLDHAYPSWRRWLEALKD NAADHLGDL ETRLLGLESGKKNYSYFNQAKSGDQRKDS
QKLG MADLDARLFQFILD RRKDS DPLRLADIYGHAKRLRQLGWRT ELTEPEKMEQARHQKLAQTVLESGLPPDLKT
SPQFNQQDIFPAGSFLHLVCRYFKNRLRAREGR LFIHPDYQRTAHRGYQFRNRFI SENNLLRYCNL KPRQKRYQMVN
DVAGVLQVSPDRRVLVARQNHNDGSQSEAVFAWLKDFRGLQTACKNAADSQKEHRGLLKTLLAGDRALQRLQDRCT
QLSRLIAREICSGDEDPEARAQKFSSIFSAQLYAI AFSDRAGNASTCPVCSLDNSRRMEMVGEDQRAKAQRLPAIS
TRVIDGAVKRMARILGRKIANDRWPLLKQKLVQGT PVRVPIITESNRFEFEPALSRLKPGVKEKSIGKDTSYEDKRN
RIAEQGGGICPYTGQPVGNGEIDHII PRSSSFGTINDEANLI FATEQGNKAKGGQFYSLKNLSRTYKQGLFGTNTD
EQIAAWIRETLWDERRGRFRFGNYLSFINLGSDEQKAFRHALFLEDGDHVREQVVAASNRSR AFVNGTQRYFAEVL
ANEFYKEALDIGKERLISFDYFGVEATSTSRGDGVRDWRHYEEFY PGEFAPYRKKDGI SQHPYSHLLDAQIAFAIN
ADKHRGDGGLRLRIPDHMGLDALNMDTGEI LHDNIFHAIRVLP EQMTAVHVQRRRPSDNRCTHRAFT RDLFYADRYL
PILLKQEENG VILKAGFTWKNAEVFNKNQRLLPQLLPTQAPHYPARLSQELDKGSTAGTAFEQRLYQLLINTDY
FKEQIDRTGAIYLTINRKT LHELWLERHNTKNGT PFENKDHATLFEFCWGKLG YRTEKKHIATEDDLK KLEAKEGNF
TPPLTARKSCCRLNMTG

(17) 琥珀酸单胞菌属物种 CAG:777 Cas9蛋白的氨基酸序列 (SEQ ID NO: 97)

MSKKT VYSNIAIDLGGKYTGCVSYTSSILPSADDINAFIIE MPDDGNGIKYTVKDRTQKRHMI RALDRFKKARKLIY
LVISSVIHRELTAD EKESISSLMKRRGYTRLEAIEDLDILKECPVDL FASYIKDLFDESQSLYDQFNQKCCDLDSAK
AMKKALSSEELKQAMDS SELDKIEKKIYKSAFDMVKA AENIINTCEFGNKH RKQYLSDIKADINKDSRLADIVKSF
DGNEKLFKCVGNISNLQLRALRWYFNDINMKG NPKWI PERFKVWARAYKYFHYPKEDQ QKARELIK NLEISNDVTS
LLCSIDPLDTIPPYEDQNNRNPPDTLTL LLSPKALDKKYTNLWEKWA EKFAKLYPNLKYQLEDIVKISDRKSRI DVK
SSSYTQSKVFN SYVMQRLFDLAADKNEPITLLRSWVRDPSNSKLADTVKIIQSVVEDREDDFYKLAQSY YNEVEMA
KSGLWSVIDDPLLEISGIHPPMKNKVI STLGVNVLGINNDFDYEFLEVWRAP I KGRKTLKSVCKDIEGLRKEYGNT
FKIEYDYANYS AKEKERTADQKKLVEINKIIEEISPLVRNSISQDPLSEIKFNPF SFSQLYTLIETD TLGFSSN
CVAVNMENNARMQFL LGNHALCSRLPTESVRPFDGALGKILERQAYEIAKVSAEIKELDSLNETVINVGILIEQNK
FEFTESI AKIKKSKLLEKIKKRSQSGVSRQLKKWENKEKRIKSASRGVCPYTGERISETNGEIDHII PRSFTKSSMG
TIFNSEANLIFCRLTGNRNKDENMYTLRELSPKYLN AVFGTANIVQIQQNI EGTVAKISSKNPRFLFDMMDQKEQDC
CRHALFMPHSKAYQIVVSAMARS YSTRVNGTQSYFVNAIISKLKNE LSSWLKSNRNTLNFFAYKVDPSDVHSTRDDL
SKLNTLLKKQENQPITSHAI DALCVLASASTNERISENISKESCLAKI SNTGNLLKLEPHEFKILRIEKKDFSDKND
DFS RKL FNDTIYAEHFLPVMVMKDQVKIGFSWNS SKEGNSIEIVKDQHKFLELLAPFFNKAI ESSCKFIT YKINKKK
AFDFIHR YIHQEV DVND CALIVKILES LRYTTVNTDVCDIYNDREKRFKSKSEILTNKNFELNFSLPKPFKILACKK
IYLP SKNEWSKITEKFS DYLGDKNIKNWKEI IYNSLNLNRKTTKLNHAGTKRVYSLPILKNASGGIRIKRYTDEME
EFYQLLEVNTPQTVKRKGFAL SEKGEVMWNKPI TVDSYIGKNLTLLEKNYKNVCIGDNFVPMKENRMVYKNEKTEVY
ISPSSDSRRLV IIKQPFDEFNKCLKNKYKNYLELPSTLVLSGDEIKIYVKNMKNDFIGNPRVETKNNKKTGKIVMLS
IGKLVRYCYMSTTCDSLMKTAYDCALS

图31

(1). MH0245_GL0161830_1 的引导RNA序列 (N匹配靶序列) (SEQ ID NO: 101)
 NNNNNNNNNNNNNNNNNNNNNNNgtttcagttattcgttggatagtgtAcatatccaacgaatgaagtcacttttaagtgagctgaaatcactaaaaattaagattgaaccgggctactgactctgtcatcgggtttactta

(2). MH0245_GL0161830_1 的TracrRNA序列 (SEQ ID NO: 102)
 Acatatccaacgaatgaagtgcacttttaagtgagctgaaatcactaaaaattaagattgaaccgggctactgactctgtcatcgggtttactta

(3). MH0245_GL0161830_1 的crRNA序列 (SEQ ID NO: 103)
 NNNNNNNNNNNNNNNNNNNNNNNgtttcagttattcgttggatagtgt

图32

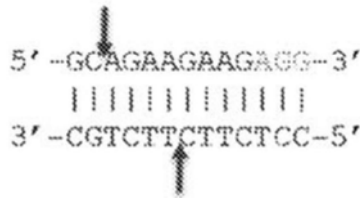


图33A



图33B

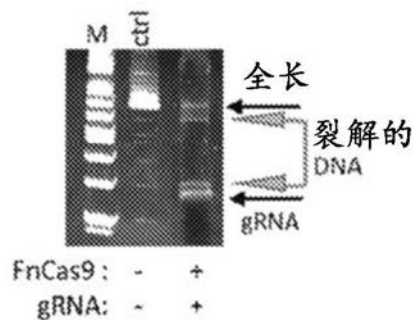


图34A

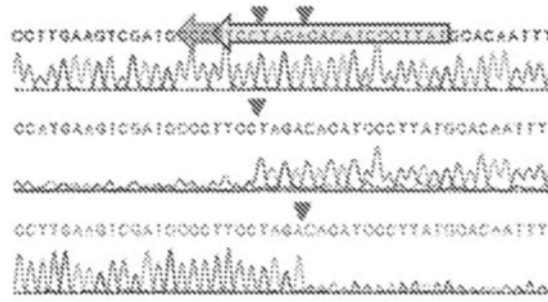


图34B

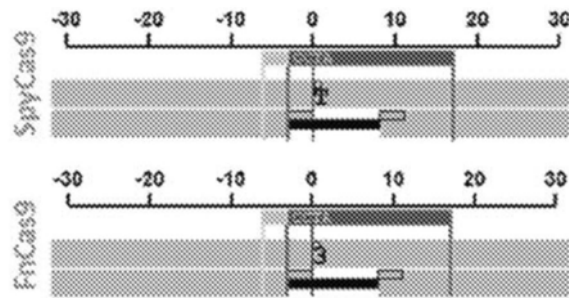


图34C

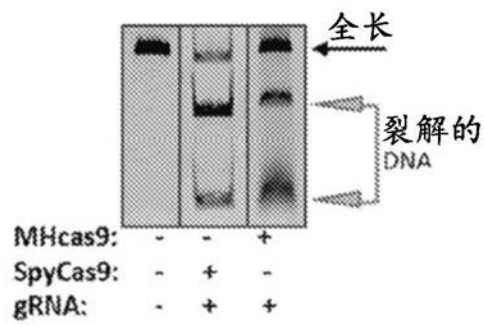


图35A

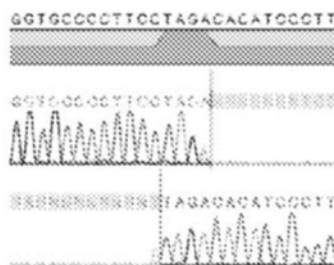


图35B

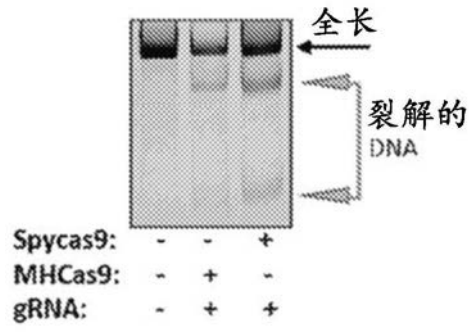


图35C

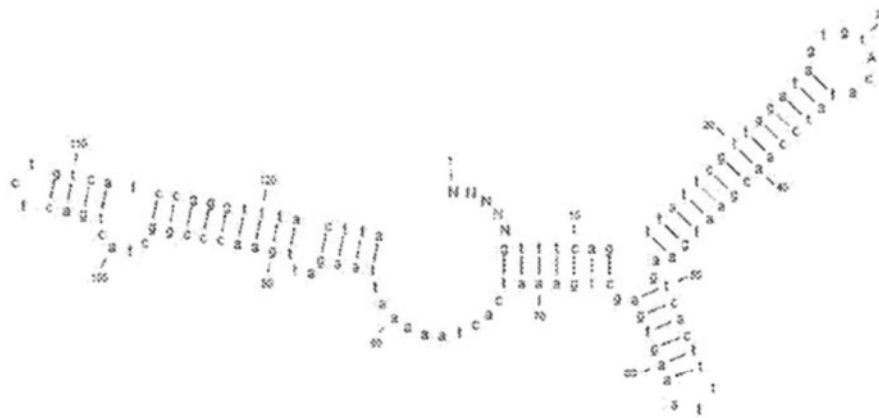


图36A

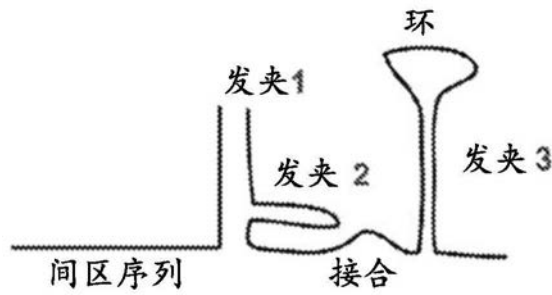


图36B

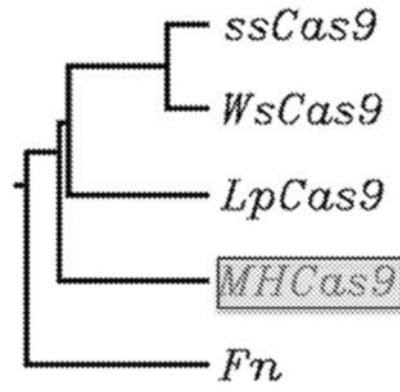


图36C

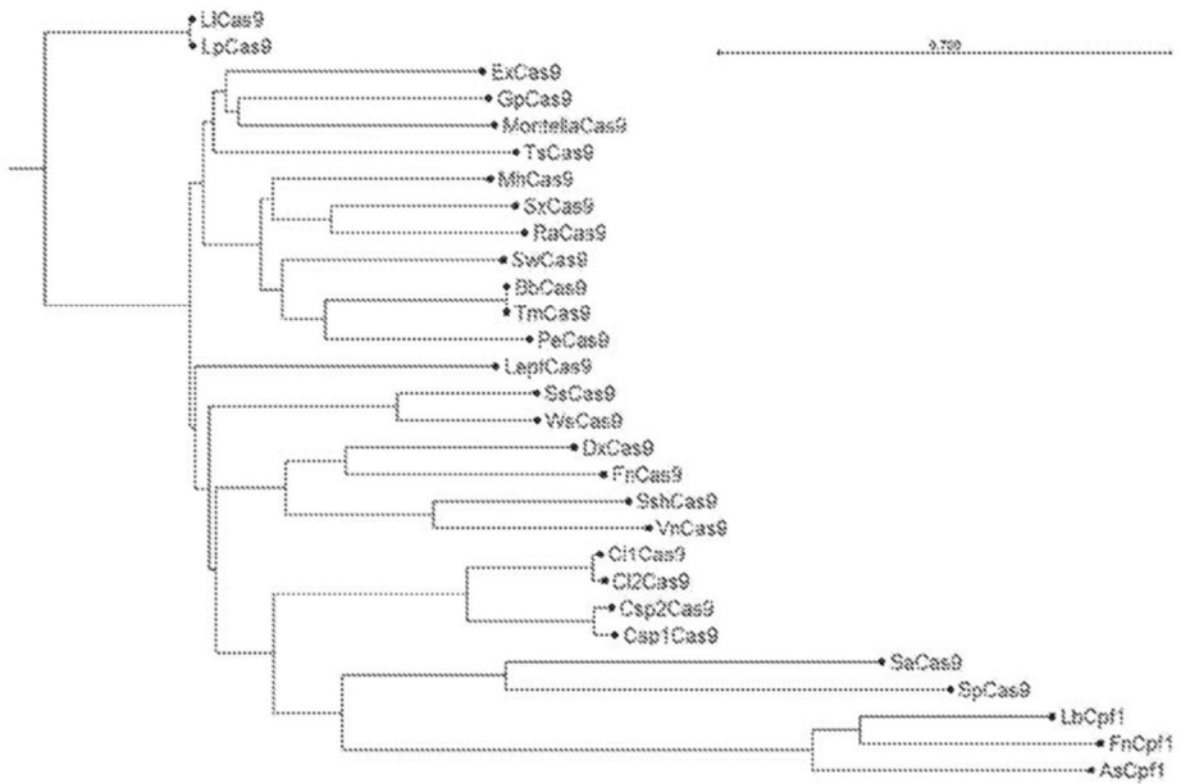


图37

- (1) 嗜肺军团菌 Cas9 (LpCas9) 的 crRNA (+) 序列 (SEQ ID NO: 104)
GTTTCAGTGGTTGGATTTTTAGATGAGGGATTATTGG
- (2) 嗜肺军团菌 Cas9 (LpCas9) 的 crRNA (-) 序列 (SEQ ID NO: 126)
CCAATAATCCCTCATCTAAAAATCCAACCACTGAAAC
- (3) 硫磺单胞菌属物种 SCADC Cas9 (SsCas9) 的 crRNA (+) 序列 (SEQ ID NO: 105)
GTTCCAGTTGCTAAGCGATGGTGTGTAAGTGTCTG
- (4) 硫磺单胞菌属物种 SCADC Cas9 (SsCas9) 的 crRNA (-) 序列 (SEQ ID NO: 127)
ACGACACTTTACAACACCATCGCTTAGCAACTGGAAC
- (5) 产琥珀酸沃廉氏菌 Cas9 (WsCas9) 的 crRNA (+) 序列 (SEQ ID NO: 106)
GTTTCACAGGCTAAGCGGATTTGCTATAAAGTGTTC
- (6) 产琥珀酸沃廉氏菌 Cas9 (WsCas9) 的 crRNA (-) 序列 (SEQ ID NO: 128)
GCAACACTTTATAGCAAATCCGCTTAGCCTGTGAAAC
- (7) 伯克霍尔德氏菌目细菌 YL45 Cas9 (BbCas9) 的 crRNA (+) 序列 (SEQ ID NO: 107)
GTTTCAATTGTACAACTTACATCCTTCAACTGAAAC
- (8) 伯克霍尔德氏菌目细菌 YL45 Cas9 (BbCas9) 的 crRNA (-) 序列 (SEQ ID NO: 129)
GTTTCAGTTGAAGGATGTAAGTTTGTACAATTGAAAC
- (9) 人粪便副萨特氏菌 Cas9 (PeCas9) 的 crRNA (+) 序列 (SEQ ID NO: 108)
GTTTCAGTTGTTATTGAAAGAATGTACGATTGAAGC
- (10) 人粪便副萨特氏菌 Cas9 (PeCas9) 的 crRNA (-) 序列 (SEQ ID NO: 130)
GCTTCAATCGTACATTCTTTCAATAACAACACTGAAAC
- (11) 华德萨特氏菌 Cas9 (SwCas9) 的 crRNA (+) 序列 (SEQ ID NO: 109)
GTTTCAGTGCTATAGCTCGTAGCGTTATGATCTTCGC
- (12) 华德萨特氏菌 Cas9 (SwCas9) 的 crRNA (-) 序列 (SEQ ID NO: 131)
GCGAAGATCATAACGCTACGAGCTATAGCACTGAAAC
- (13) 嗜淀粉瘤胃杆菌 Cas9 (RaCas9) 的 crRNA (+) 序列 (SEQ ID NO: 110)
GTTTCAGTTATATGTAGATACAGCTTATAATTGATAC
- (14) 嗜淀粉瘤胃杆菌 Cas9 (RaCas9) 的 crRNA (-) 序列 (SEQ ID NO: 132)
GTATCAATTATAAGCTGTATCTACATATAACTGAAAC
- (15) 弯曲杆菌属物种 P0111 Cas9 (Csp1Cas9) 的 crRNA (+) 序列 (SEQ ID NO: 111)
GTTTCAAACCGAGTTGATTTGTGGTATAATTTGCAAC
- (16) 弯曲杆菌属物种 P0111 Cas9 (Csp1Cas9) 的 crRNA (-) 序列 (SEQ ID NO: 133)
GTTGCAAATTATAACCACAAATCAACTCGGTTTGAAC
- (17) 弯曲杆菌属物种 RM9261 Cas9 (Csp2Cas9) 的 crRNA (+) 序列 (SEQ ID NO: 112)
GTTTCAAACCGAGTTGATTTGTGGTATAATTTGCAAC
- (18) 弯曲杆菌属物种 RM9261 Cas9 (Csp2Cas9) 的 crRNA (-) 序列 (SEQ ID NO: 134)
GTTGCAAATTATAACCACAAATCAACTCGGTTTGAAC

- (19) 拉尼尔弯曲杆菌菌株 RM8001 Cas9 (Cl1Cas9) 的 crRNA (+) 序列 (SEQ ID NO: 113)
ACAAATGCAAATTATACTACAGATTAACCTTTAGCGTTTGAAATCACCTAAAGGTGTTAACAGATCGGGGCAACTG
CCTTTGGCACCCCGGTATTATTC
- (20) 拉尼尔弯曲杆菌菌株 RM8001 Cas9 (Cl1Cas9) 的 crRNA (-) 序列 (SEQ ID NO: 135)
GTTGCAAATTATAACCACAATCCAACCTGGTTTGAAAC
- (21) 拉尼尔弯曲杆菌菌株 P0121 Cas9 (Cl2Cas9) 的 crRNA (+) 序列 (SEQ ID NO: 114)
GTTTCAAACCAAGTTGGATTGTGGTATAATTTGCAAC
- (22) 拉尼尔弯曲杆菌菌株 P0121 Cas9 (Cl2Cas9) 的 crRNA (-) 序列 (SEQ ID NO: 136)
GTTGCAAATTATAACCACAATCCAACCTGGTTTGAAAC
- (23) MH0245_GL0161830_1 Cas9 (MH0245Cas9) 的 crRNA (+) 序列 (SEQ ID NO: 115)
GTTTCAATTAACACTATCCAACGAATAACTGAAAC
- (24) MH0245_GL0161830_1 Cas9 (MH0245Cas9) 的 crRNA (-) 序列 (SEQ ID NO: 137)
GTTTCAGTTATTTCGTTGGATAGTGTTTAATTGAAAC
- (25) 土拉弗朗斯细菌新凶手亚种 Cas9 (FnCas9) 的 crRNA (+) 序列 (SEQ ID NO: 116)
GTTTCAGTTGCTGAATTATTTGGTAACTACTGTTAG
- (26) 土拉弗朗斯细菌新凶手亚种 Cas9 (FnCas9) 的 crRNA (-) 序列 (SEQ ID NO: 138)
CTAACAGTAGTTTACCAAATAATTCAGCAACTGAAAC
- (27) γ 变形菌 HTCC5015 Cas9 (GpCas9) 的 crRNA (+) 序列 (SEQ ID NO: 117)
GTTTCAGAGCTTATCCCAACAAACCAACAGCTGAAAC
- (28) γ 变形菌 HTCC5015 Cas9 (GpCas9) 的 crRNA (-) 序列 (SEQ ID NO: 139)
GTTTCAGCTGTTGGTTTGTGGGATAAGCTCTGAAAC
- (29) 鼠毛滴虫 Cas9 (TmCas9) 的 crRNA (+) 序列 (SEQ ID NO: 118)
TAGGAGACTTTGTAAAATCGAAATAAACCCCTTTTCCCATTTATTTAGCCCTCTTAAATTCTCTCGGGAAAGTATT
ATGACAACCTATGGATAGGCGACAA
- (30) 鼠毛滴虫 Cas9 (TmCas9) 的 crRNA (-) 序列 (SEQ ID NO: 140)
GTTTCAGTTGAAGGATGTAAGTTTGTACAATTGAAAC
- (31) 伦敦军团菌 Cas9 (LlCas9) 的 crRNA (+) 序列 (SEQ ID NO: 119)
GTTTCAGTGGTTGGATTTTTAGATGAGGGATTATTGG
- (32) 伦敦军团菌 Cas9 (LlCas9) 的 crRNA (-) 序列 (SEQ ID NO: 141)
CCAATAATCCCTCATCTAAAAATCCAACCACTGAAACT
- (33) 沙姆盐弧菌 Cas9 (SshCas9) 的 crRNA (+) 序列 (SEQ ID NO: 120)
GTTTCAGTGGCTGACTTGTGAGATAGGCTCTGAAGC
- (34) 沙姆盐弧菌 Cas9 (SshCas9) 的 crRNA (-) 序列 (SEQ ID NO: 142)
GCTTCAGAGCCTATCTCAACAAGTCAGCCACTGAAAC
- (35) 钩端螺旋体属物种分离株 FW.030 Cas9 (Lept.Cas9) 的 crRNA (+) 序列 (SEQ ID NO: 121)
GTTTCAGTTGATCGAATCTCAAGACTGAACATGATGG
- (36) 钩端螺旋体属物种分离株 FW.030 Cas9 (Lept.Cas9) 的 crRNA (-) 序列 (SEQ ID NO: 143)
CCATCATGTTTCAGTCTTGAGATTTCGATCAACTGAAAC

- (37) 莫里特拉氏菌属物种分离株 NORP46 Cas9 (莫里特拉氏菌属 Cas9) 的 crRNA (+) 序列 (SEQ ID NO: 122)
GTTTCATTACTACGTAAGTTTGATAAGCTAGTGATAC
- (38) 莫里特拉氏菌属物种分离株 NORP46 Cas9 (莫里特拉氏菌属 Cas9) 的 crRNA (-) 序列 (SEQ ID NO: 144)
GTATCACTAGCTTATCAAACCTTACGTAGTAATGAAAC
- (39) 内生单胞菌属物种 S-B4-1U Cas9 (ExCas9) 的 crRNA (+) 序列 (SEQ ID NO: 123)
CTTTCAGTTGCTACCCAGACGAAGTAAGCTTTGATTC
- (40) 内生单胞菌属物种 S-B4-1U Cas9 (ExCas9) 的 crRNA (-) 序列 (SEQ ID NO: 145)
GAATCAAAGCTTACTTTCGTCTGGGTAGCAACTGAAAG
- (41) 喜盐泰米尔纳德菌 Cas9 (TsCas9) 的 crRNA (+) 序列 (SEQ ID NO: 124)
GTTTCAGAGGTCAGTCATTTGGGATAGCCTTTGAAGC
- (42) 喜盐泰米尔纳德菌 Cas9 (TsCas9) 的 crRNA (-) 序列 (SEQ ID NO: 146)
GCTTCAAAGGCTATCCCAAATGACTGACCTCTGAAAC
- (43) 需钠弧菌 Cas9 (VnCas9) 的 crRNA (+) 序列 (SEQ ID NO: 125)
GTTTCAGTGGCTGACTTGTTGAGATAAGCTGTGGAAC
- (44) 需钠弧菌 Cas9 (VnCas9) 的 crRNA (-) 序列 (SEQ ID NO: 147)
GTTCCACAGCTTATCTCAACAAGTCAGCCACTGAAAC

图39

- (1) 嗜肺军团菌 Cas9 (LpCas9) 的tracrRNA序列(SEQ ID NO: 148)
ATCATCTAAATTTTCGATACCCTGAAATCAACAAAATTAAGATTGAATCGTTTTTCTATGCTCGTCTTAAT
AGCGAGCATATAACGATTTCAAAA
- (2) 硫磺单胞菌属物种 SCADC Cas9 (SsCas9)的tracrRNA 序列 (SEQ ID NO: 149)
TCACGACACTTTTACAACACGTATCGTTTCGATGCACTGAAATCATCGAAAAGATATATAGACCCGCCAACT
GCCTTAGGCATGGCGGGCTTCTTTCTTAAAA
- (3) 产琥珀酸沃廉氏菌 Cas9 (WsCas9)的tracrRNA序列(SEQ ID NO: 150)
GCAACACTTTTATAGCAAATCCGTTTCGATGCCCTTGAAATCATCAAAAAGATATAATAGACCCGCCCACTGTA
TTGTACATGGCGGGACTTTTTTA
- (4) 伯克霍尔德氏菌目细菌YL45 Cas9 (BbCas9) 的 tracrRNA 序列 (SEQ ID NO: 151)
ATAGGAGACTTTGTAAAATCGAAATAAACCCCTTTTCCCATTTATTTAGCCCTCTTAAATTTCTCTCGGGA
AAGTATTATGACAACCTATGGATAGGCCA
- (5) 人粪便副萨特氏菌 Cas9 (PeCas9) 的 tracrRNA 序列 (SEQ ID NO: 152)
TAGTTACATTCTTTCAACAACGAAATCGCCTTAGGGCGAGCTGAAATCAATTTGATTAAATATTAGATCCG
GCTACTGAGGCCTCTTGGCCTTATCCGGATTAACAAGGAGCCTCTATTTCGAGGCTT
- (6) 华德萨特氏菌 Cas9 (SwCas9)的tracrRNA序列(SEQ ID NO: 153)
ACATCGTAGGATACCGATTAGAGATCATAACGCTATGAGCTATAGGAAATCACCTTCGGGTGAGCTGAAAT
CCCCTAAAGCTAAGATTGAATCCGGCCACTATC
- (7) 嗜淀粉瘤胃杆菌 Cas9 (RaCas9)的tracrRNA序列(SEQ ID NO: 154)
AAAGTATCTACATATGAAGTCGCTTTTATGGCGAGCTGAAATCACCAAAGTTAAGATATAGACCAAGCAA
CTGTAATTTACATCTTGGTATCTTCAGTATCTGT
- (8) 弯曲杆菌属物种 P0111 Cas9 (Csp1Cas9)的tracrRNA序列(SEQ ID NO: 155)
ATACAAACGCAAATTATACTACAGATTAACCTTTTAGCGTTTGAAATCACCTAAAGGTGTTAACAGATCGG
GGCAACTGCCATATGGCACCCCGGTATC
- (9) 弯曲杆菌属物种 RM9261 Cas9 (Csp2Cas9)的tracrRNA序列(SEQ ID NO: 156)
ATACAAACGCAAATTATACTACAGATTAACCTTTTAGCGTTTGAAATCACCTAAAGGTGTTAACAGATCGG
GGCAACTGCCATATGGCACCCCGGTATC
- (10) 拉尼尔弯曲杆菌菌株 RM8001 Cas9 (Cl1Cas9)的 tracrRNA序列(SEQ ID NO: 157)
CAAATTATACCATAATCTAGCTTATAGCGTTTGAAATCACCAAGGTGTTTTTCATATCGGGGCAACTGCC
TTTGGCACCCCGGTTA
- (11) 拉尼尔弯曲杆菌菌株 P0121 Cas9 (Cl2Cas9)的tracrRNA序列(SEQ ID NO: 158)
CAAATTATACCATAATCTAGCTTATAGCGTTTGAAATCACCAAGGTGTTTTTCATATCGGGGCAACTGCC
TTTGGCACCCCGGTTATT
- (12) MH0245_GL0161830_1 Cas9 (MH0245Cas9)的tracrRNA 序列 (SEQ ID NO: 159)
TATCCAACGAATGAAGTCACTTTTAAAGTGAGCTGAAATCACTAAAAATTAAGATTGAACCCGGCTACTGA
CTCTGTCTATCCGGGTTTACTTATTTT
- (13) 土拉弗朗西斯菌新凶手亚种 Cas9 (FnCas9) 的 tracrRNA 序列 (SEQ ID NO: 160)
GTACCAATAATTAATGCTCTGTAATCATTTAAAGTATTTTGAACGGACCTCTGTTTGACACGTCTGAAT
AACTAAAA
- (14) γ 变形菌 HTCC5015 Cas9 (GpCas9) 的 tracrRNA (1) 序列 (SEQ ID NO: 161)

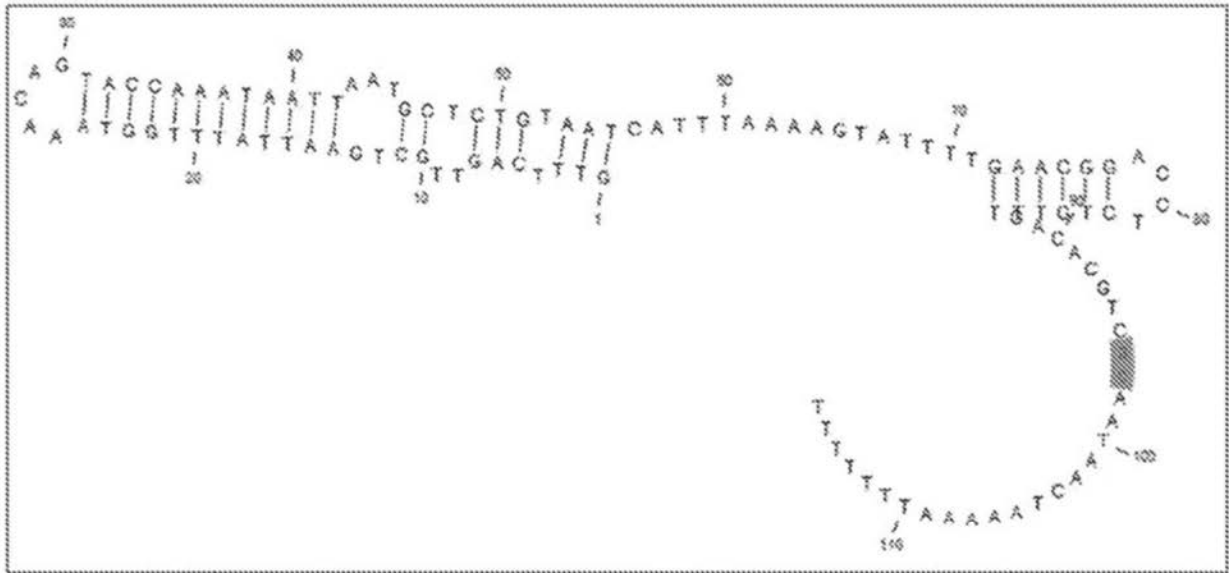
- TTCAGAATGCATCCCAACATTCTATACACTGAAATCATAGAAAATCACGTTTGTGGCCCGACCAACTGCTT
CGGCATGTCGGGTTTTTTTCATTTCTAAACC
- (15) **γ 变形菌 HTCC5015 Cas9 (GpCas9) 的 tracrRNA (2) 序列 (SEQ ID NO: 162)**
ATTTTCAGTGTATAGAATGTTGGGATGCATTCTGAATACGAAGCTATCACCCAGTTCGAAGAATTGCAAGAA
AGTGCCGTCACCTAACCAATTGCAATA
- (16) **鼠毛滴虫 Cas9 (TmCas9) 的 tracrRNA 序列 (SEQ ID NO: 163)**
TAGGAGACTTTGTAAAATCGAAATAAACCCTTTTCCCATTATTTAGCCCTCTTAAATTCTCTCGGGAA
AGTATTATGACAACCTATGGATAGGCGACAA
- (17) **伦敦军团菌 Cas9 (LlCas9) 的 tracrRNA 序列 (SEQ ID NO: 164)**
GGCAATAATTAATTAGAAATCATCTAAATTTTCGATACCCTGAAATCAACAAAATTAAGATTGAATCGTT
TTTCTATGCTCGTCTTAATAGCGAGCATATAACGATTTCAAATCTATCGTTACAGTGATTT
- (18) **沙姆盐弧菌 Cas9 (SshCas9) 的 tracrRNA 序列 (SEQ ID NO: 165)**
TCAGCGCCTATTTTTTTTCGCAAACCGCTTTGATGTCAGAAAATGGCGGTATTTTTCGCAACATGCTCCTTAG
ACAAAATAATCTTCATCTGTCATTGCAG
- (19) **钩端螺旋体属物种分离株FW.030 Cas9 (Lept.Cas9) 的 tracrRNA 序列 (SEQ ID NO: 166)**
GATAGACATTTTTCGAAAACAAGTCATAGTGTAAATGAGATTTCGTACAACCTGAAATCAATCAAATGTAAGA
TTGGCTGACCCTCTACAGTGCGTATGGTC
- (20) **莫里特拉氏菌属物种分离株NORP46 Cas9 (莫里特拉氏菌属 Cas9) 的 tracrRNA 序列 (SEQ ID NO: 167)**
AAGAGTATATTGGAAATATCAAACCTTACTATGTAATGAAATCAATTAAAATCAAGATTGATTGTCCCTCTG
CGCTAGCATGGGCAA
- (21) **内生单胞菌属物种S-B4-1U Cas9 (ExCas9) 的 tracrRNA 序列 (SEQ ID NO: 168)**
TACTTCGTTTTGGTATGCACTGAAATCACCAAAGTCGCGATAAGCGAACCAGCTCTCCAGACTATCTGGAG
CCCTCTGGAA
- (22) **喜盐泰米尔纳德菌 Cas9 (TsCas9) 的 tracrRNA 序列 (SEQ ID NO: 169)**
TGGACTAGGGTAAGAAATATCCCAAATGACATACCCTGAAATCACCAAAGGTAGAATGAAGATTCTGACA
CCCCACTGCATTGCGCACCGGGTGTTTTTTCATTTTCAGAAATCGGTTCGTT
- (23) **需钠弧菌 Cas9 (VnCas9) 的 tracrRNA (1) 序列 (SEQ ID NO: 170)**
TTATGTGGGGTTTTACGCTTACTTTTTTCGCAAATGCATCGGATGTCATCAAATCCCCTTTTTTTGCGAAC
ATTTCTTCCTATCGCACCTTTAGCGATCAT
- (24) **需钠弧菌 Cas9 (VnCas9) 的 tracrRNA (2) 序列 (SEQ ID NO: 171)**
CAAATGCATCGGATGTCATCAAATCCCCTTTTTTTCGGAACATTTCTTCCTATCGCACCTTTAGCGATCAT
TGCAATACCTTAGATACTCACATTGATCGC

图40A

Attorney Docket No.: 0098-0002WO1

Cas9	登录号	定义	SEQ ID Nos:			
			crRNA (+)	crRNA (-)	tracrRNA 1	tracrRNA 2
LpCas9	CP015931	嗜肺军团菌费城菌株	104	126	148	
SsCas9	JQK01000013	硫酸单胞菌属物种 SCADC contig_724	105	127	149	
WsCas9	BX571661	产琥珀酸沃廉氏菌	106	128	150	
BbCas9	NZ_CP015403	伯克霍尔德氏菌目细菌YL45	107	129	151	
PeCas9	NZ_GL883755	人粪便副萨氏菌 YIT 11859 Scfld238	108	130	152	
SwCas9	NZ_GL636545	华德萨氏菌 3_1_45B supercont1.6	109	131	153	
RaCas9	NZ_FOXF01000052	嗜淀粉瘤胃杆菌菌株 DSM 1361	110	132	154	
Csp1Cas9 (Cp0111Cas9)	NZ_MJKL01000014	弯曲杆菌属物种 P0111 ERR466672.12216 2_83.14	111	133	155	
Csp2Cas9 (Crm9261Cas9)	NZ_MJLP01000003	弯曲杆菌属物种 RM9261 RM9261contig00003	112	134	156	
Cl1Cas9 (Clrm8001Cas9)	NZ_MJJW01000026	拉尼尔弯曲杆菌菌株 RM8001 RM8001contig00026	113	135	157	
Cl2Cas9 (Clp0121Cas9)	NZ_MJJN01000007	拉尼尔弯曲杆菌菌株 P0121 ERR460911.12227 1_93.7	114	136	158	
MH0245Cas9	???	???	115	137	159	
FnCas9	NC_008601	土拉弗朗西斯菌新凶手亚种 U112	116	138	160	
GpCas9	NZ_DS990600	Y变形菌 HTCC5015 scf_1106129323647	117	139	161	162
TmCas9=BbCas9	NZ_NHMP01000007 REGION: 20000..80000	鼠毛滴虫菌株 YL45	118	140	163	
LjCas9	NZ_LNYK01000030	伦敦军团菌菌株 ATCC 49505 Lion_ctg030	119	141	164	
SshCas9	NZ_MUFC01000011	沙姆弧菌菌株 CBH463 支架 11.1	120	142	165	
LeptCas9	PEQN01000004	狗端螺旋体属物种分菌株 FW_030 FD_1000036_176051	121	143	166	
MoritellaCas9	NVWD01000026		122	144	167	
ExCas9	NZ_FWPT01000003	内生单胞菌属物种 S-B4-1U分菌株 SB41UT1	123	145	168	
TsCas9	NMPM01000001	喜盐泰米尔纳德菌菌株 Mi-7 contig-1	124	146	169	
VnCas9	NZ_CP016349	需钠弧菌菌株 CCUG 16373 染色体 1	125	147	170	171

图40B

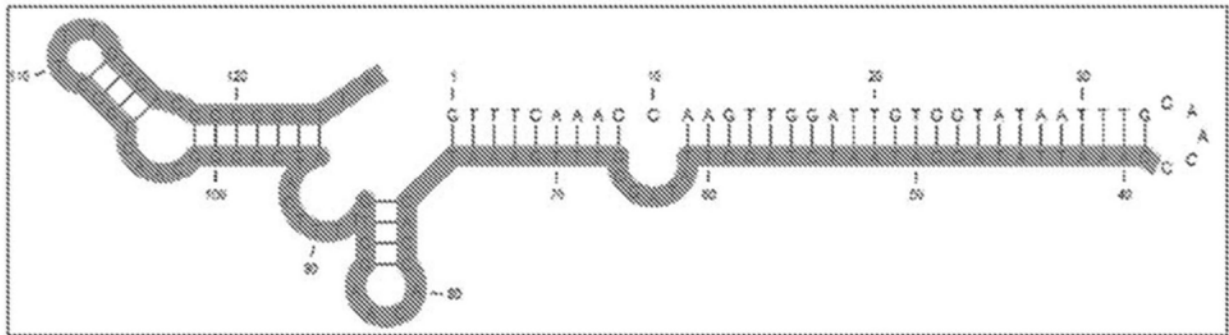


土拉弗朗西斯菌新凶手亚种

嵌合 gRNA (参考 : PMID 28387220) (SEQ ID NO: 172)

GTTTCAGTTGCTGAATTATTTGGTAAACAGTACCAAATAATTAATGCTCTGTAATCATTTAAAA
GTATTTTGAACGGACCTCTGTTTGACACGTCTGAATAACTAAAAATTTTTTTT

图41A



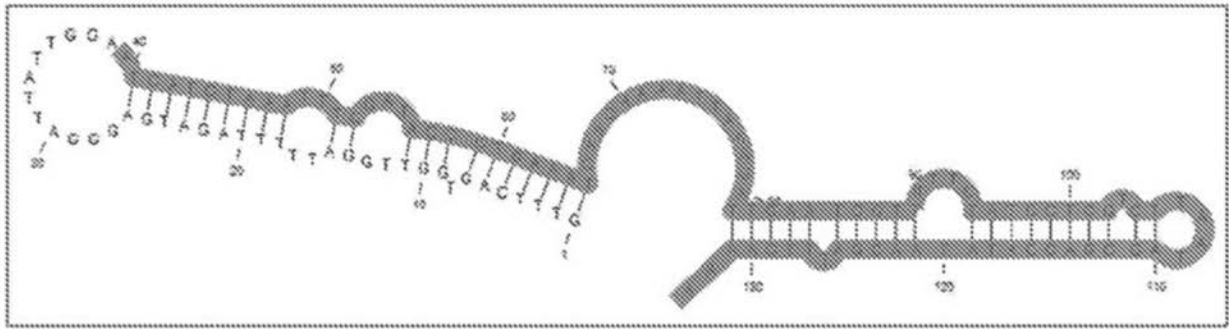
拉尼尔弯曲杆菌

RM8001

嵌合 gRNA (SEQ ID NO: 173)

GTTTCAAACCAAGTTGGATTGTGGTATAATTTGCAACCCAAATTATACCATAATCTAGCTTATA
GCGTTTCAAATCACCACAAGGTGTTTTTCATATCGGGGCAACTGCCTTTGGCACCCCGGTTA

图41B

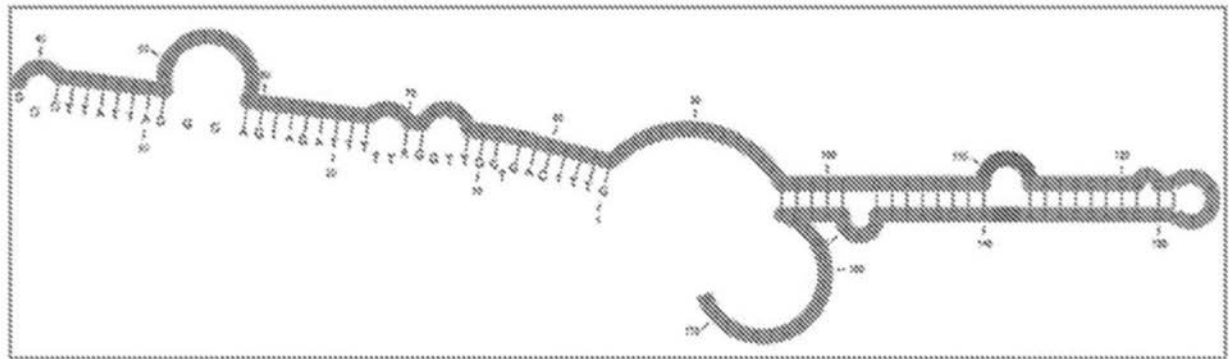


嗜肺军团菌巴黎菌株

嵌合 gRNA (SEQ ID NO: 174)

GTTTCAGTGGTTGGATTTT TAGATGAGGGATTATTGGAATCATCTAAATTTTCGATACCCTGAAA
TCAACAAAATTAAGATTGAATCGTTTTTCTATGCTCGTCTTAATAGCGAGCATATAACGATTT
CAAAA

图41C



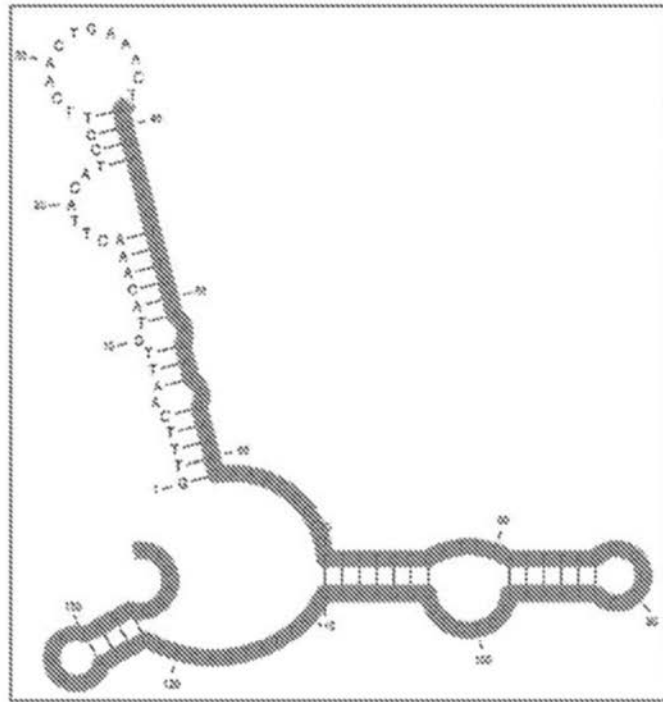
伦敦军团菌菌株

ATCC 49505

嵌合 gRNA (SEQ ID NO: 175)

GTTTCAGTGGTTGGATTTT TAGATGAGGGATTATTGGGGCAATAATTAATTAGAAATCATCT
AAATTTTCGATACCCTGAAATCAACAAAATTAAGATTGAATCGTTTTTCTATGCTCGTCTTAAT
AGCGAGCATATAACGATTTCAAATCTATCGTTACAGTGATTT

图41D



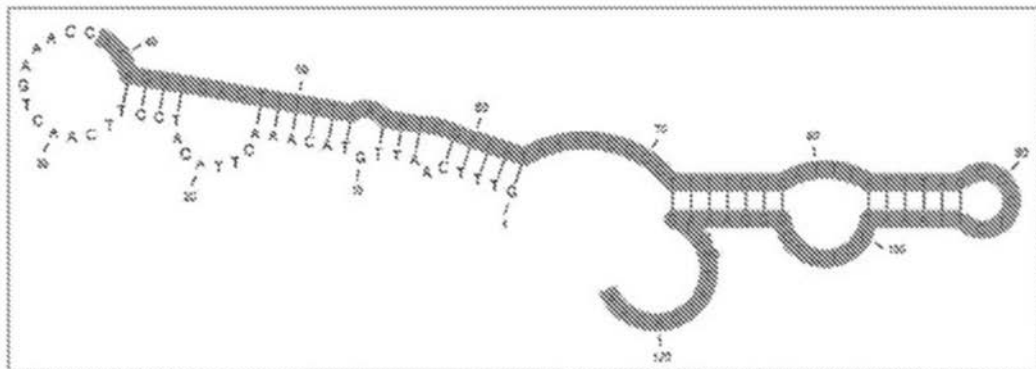
鼠毛滴虫菌株

YL45

嵌合 gRNA (SEQ ID NO: 176)

GTTTCAATTGTACAAACTTACATCCTTCAACTGAACTAGGAGACTTTGTAAAATCGAAATAAA
 CCCCTTTTCCCATTTATTTAGCCCTCTTAAATTCTCTCGGAAAGTATTATGACAACCTATGG
 ATAGGCGACAA

图41E

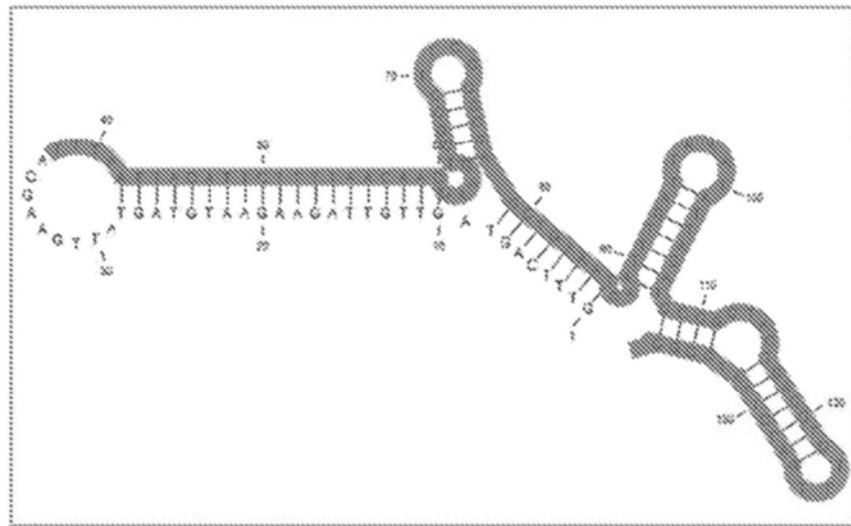


伯克霍尔德里氏菌目细菌YL45

嵌合 gRNA (SEQ ID NO: 177)

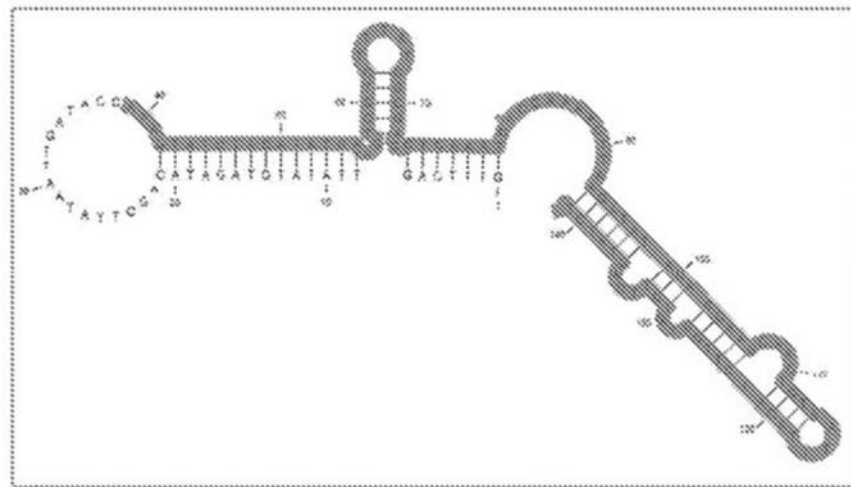
GTTTCAATTGTACAAACTTACATCCTTCAACTGAAACCATAGGAGACTTTGTAAAATCGAAATA
 AACCCCTTTTCCCATTTATTTAGCCCTCTTAAATTCTCTCGGAAAGTATTATGACAA

图41F



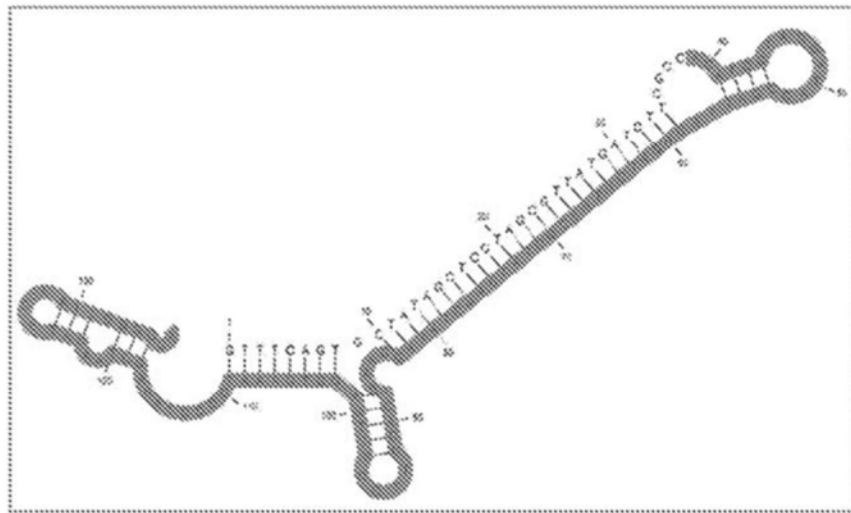
人粪便副萨特氏菌 YIT 11859
 嵌合 gRNA (SEQ ID NO: 178)
 GTTTCAGTAGTTGTTAGAGAATGTAGTATTGAAGCATTTAATTACATTCTTTTAACAACGAAG
 TCGCCTTCGGGCGAGCTGAAATCAATTTGATTAAATATTAGATCCGGCTACTGAGGTCTTTGAC
 CTTATCCGG

图41G



嗜淀粉瘤胃杆菌菌株 DSM 1361
 嵌合 gRNA (SEQ ID NO: 179)
 GTTTCAGTTATATGTAGATACAGCTTATAATTGATACCAAAGTATCTACATATGAAGTCGTCTT
 TATGGCGAGCTGAAATCACCAAAGTTAAGATATAGACCAAGCAACTGTAATTTACATCTTGGT
 ATCTTCAGTATCT

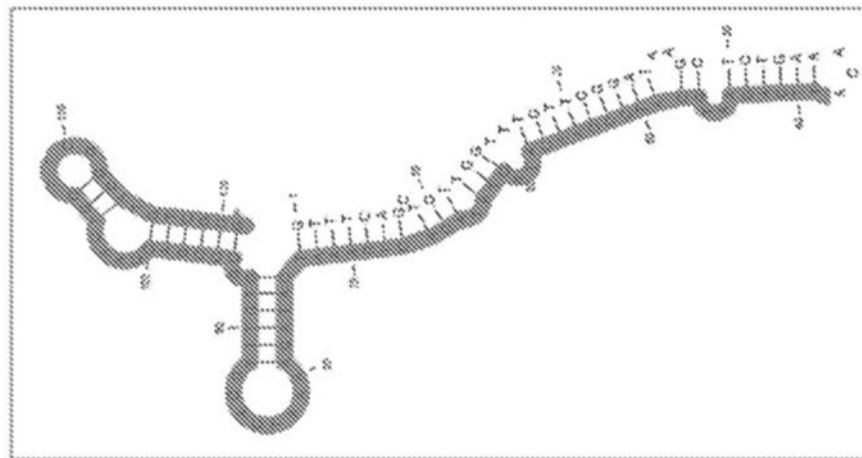
图41H



华德萨特氏菌 3_1_45B supercont1.6
嵌合 gRNA (SEQ ID NO: 180)

GTTTCAGTGCTATAGCTCGTAGCGTTATGATCTTCGCCACATCGTAGGATACCGATTAGAGATC
ATAACGCTATGAGCTATAGGAAATCACCTTCGGGTGAGCTGAAATCCCCTAAAGCTAAGATTGA
ATCCGGCCA

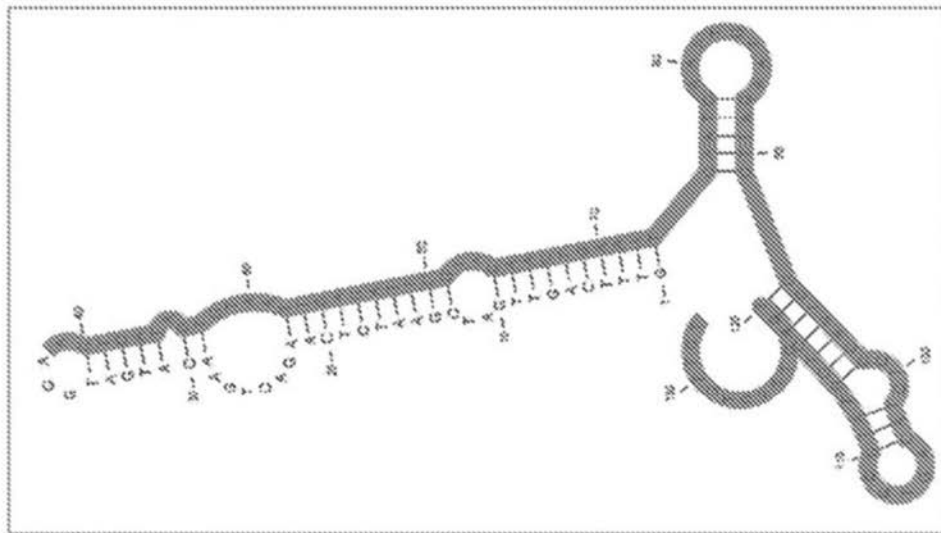
图41I



γ 变形菌 HTCC5015
嵌合 gRNA (SEQ ID NO: 181)

GTTTCAGCTGTTGGTTTGTGGGATAAGCTCTGAAACATTCAGAATGCATCCCAACATTCTATA
CACTGAAATCATAGAAAATCACGTTTGTGGCCGACCAACTGCTTCGGCATGTCGGG

图41J

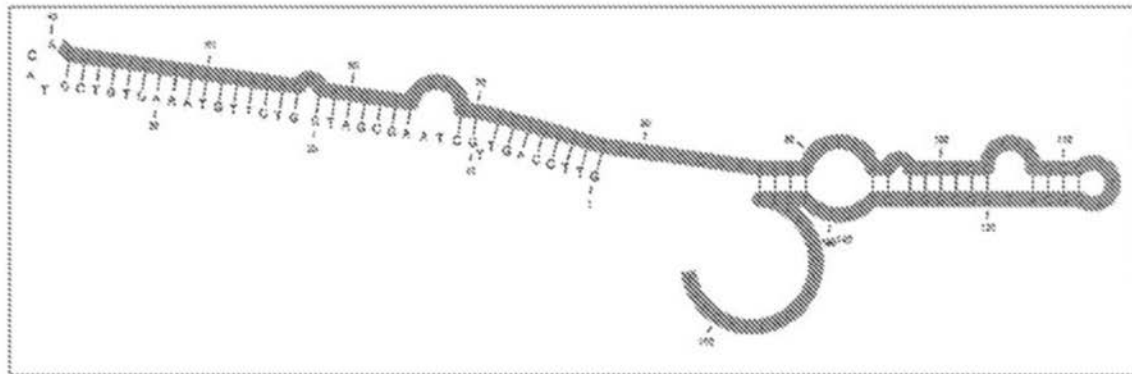


钩端螺旋体属物种分离株 FW.030 FD_1000036_176051

嵌合 gRNA (SEQ ID NO: 182)

GTTTCAGTTGATCGAATCTCAAGACTGAACATGATGGAAGTCATAGTGTAATTGAGATTCGTAC
AACTGAAATCAATCAAATGTAAGATTGGCTGACCCTCTACAGTGCATGGTCAGTAAGTTAGT
GTTCT

图41K

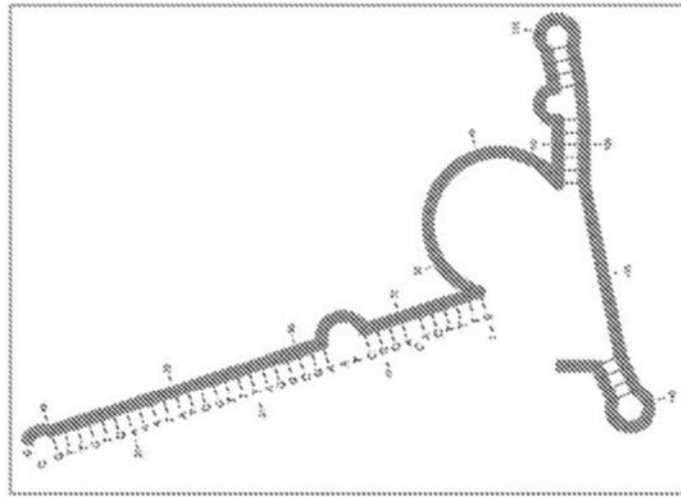


硫磺单胞菌属物种SCADC

嵌合 gRNA (SEQ ID NO: 183)

CGACACTTTACAACACGTATCGTTCGATGCACTGgAATCATCGAAAAGATATATAGACCCGCC
AACTGCCTTAGGCATGGCGGGCTTCTTTCTAAAACCTATCATTACAAT

图41L

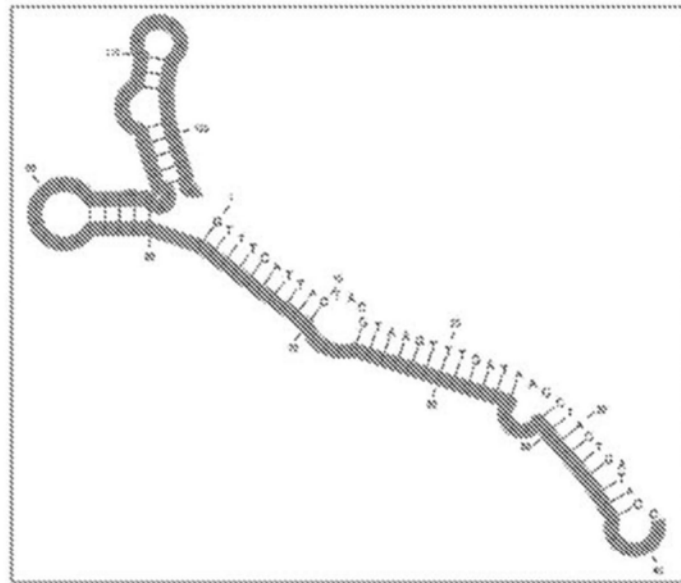


产琥珀酸沃廉氏菌 DSM 1740

嵌合 gRNA (SEQ ID NO: 184)

GTTTCACAGGCTAAGCGGATTTGCTATAAAGTGTTGCGGCAACACTTTATAGCAAATCCGTTTCG
ATGCCTTGAAATCATCAAAAAGATATAATAGACCCGCCCACTGTATTGTACATGGCGGGACTTT
TAAAACCTGTCATTGCAGT

图41M

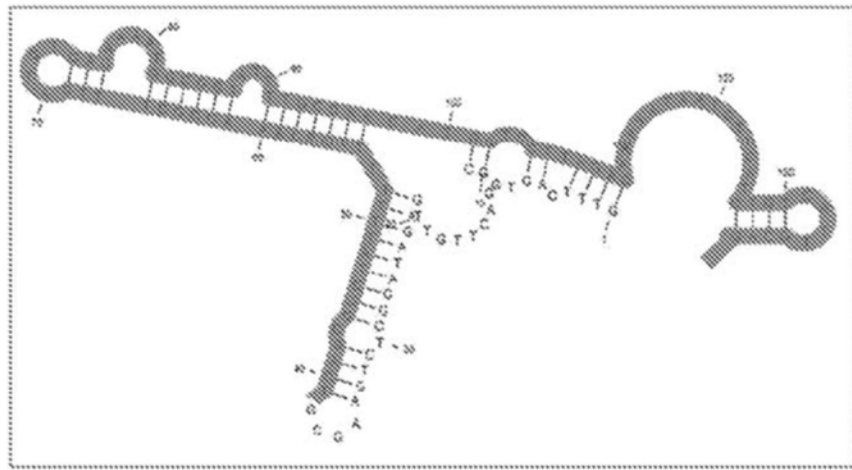


莫里特拉氏菌属物种分离株 NORP46

嵌合 gRNA (SEQ ID NO: 185)

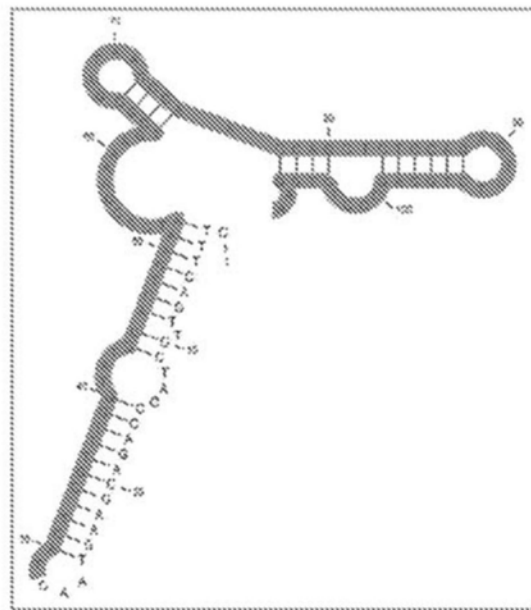
GTTTCATTACTACGTAAGTTTGATAAGCTAGTGATACCAAGAGTATATTGGAAATATCAAACCTT
ACTATGTAATGAAATCAATTAATAAATCAAGATTGATTGTCCCTCTGCGCTAGCATGGGCAA

图41N



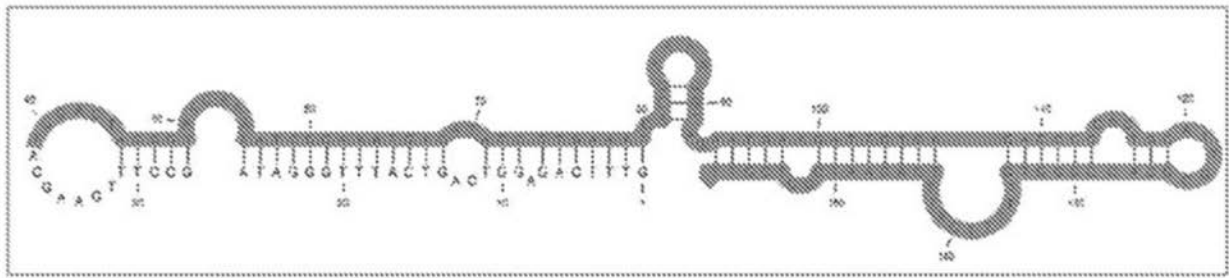
沙姆盐弧菌菌株 CBH463
 嵌合 gRNA (SEQ ID NO: 186)
 GTTTCAGTGGCTGACTTGTTGAGATAGGCTCTGAAGCGTCAGCGCCTATTTTTTTTCGCAAACCG
 CTTTGATGTCAGAAAATGGCGGTATTTTGCGAACATGCTCCTTAGACAAAATAATCTTCATCT
 GTCATTGCAG

图410



内生单胞菌属物种 S-B4-1U分离株SB41UT1
 嵌合 gRNA (SEQ ID NO: 187)
 CTTTCAGTTGCTACCCAGACGAAGTAAGTACTTCGTTTGGTATGCACTGAAATCACCAAAGTC
 GCGATAAGCGAACCAGCTCTCCAGACTATCTGGAGCCCTCTGGAA

图41P

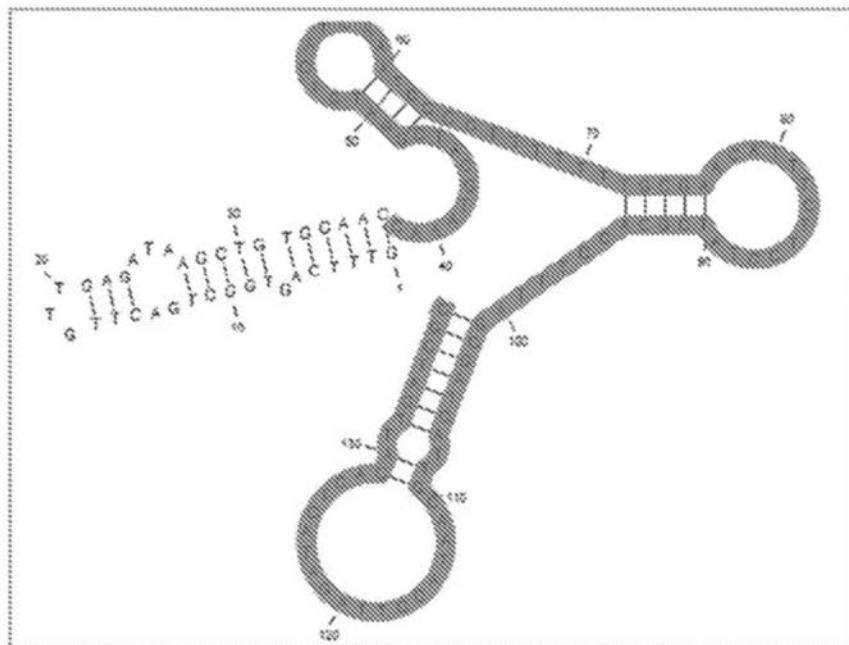


喜盐泰米尔纳德菌菌株 Mi-7

嵌合 gRNA (SEQ ID NO: 188)

GTTTCAGAGGTCAGTCATTTGGGATAGCCTTTGAAGCATGGACTAGGGTAAGAAATATCCCAA
 TGACATACCCTGAAATCACCAAAGGTAGAATGAAGATTCTGACACCCGCACTGCATTGCGCAC
 CGGGTGTTCATTCAGAATCGGTCGTT

图41Q

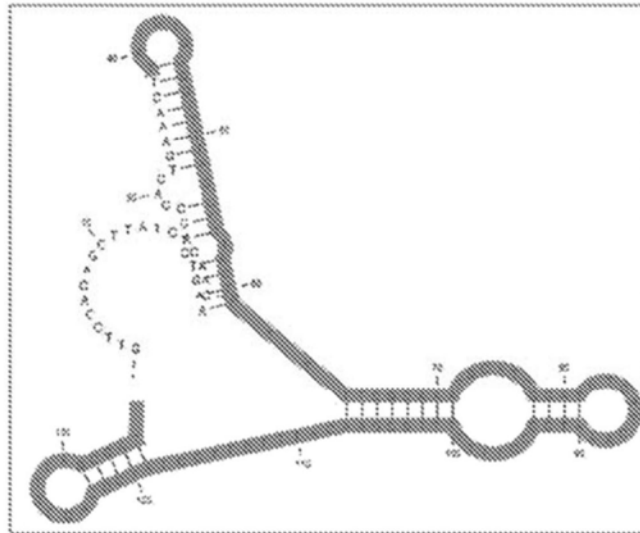


需钠弧菌菌株 CCUG 16373

嵌合 gRNA (SEQ ID NO: 189)

GTTTCAGTGGCTGACTTGTGAGATAAGCTGTGGAACCAAATGCATCGGATGTCATCAAATCCC
 CTTTTTTTCGGAACATTTCTTCCTATCGCACCTTTAGCGATCATTGCAATACCTTAGATACTCA
 CATTGATCGC

图41R

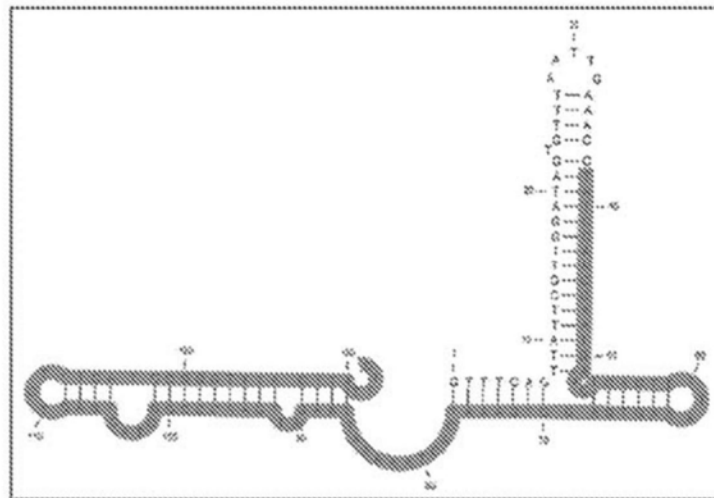


需钠弧菌菌株 CCUG 16373

嵌合 gRNA (SEQ ID NO: 190)

GTTCCACAGCTTATCTCAACAAGTCAGCCACTGAACTTATGTGGGGTTTTACGCTTACTTTTT
TCGCAAATGCATCGGATGTCATCAAATCCCCTTTTTTTCGAACATTTCTTCCTATCGCACCT
TAGCGAT

图41S



未培养的细菌，人类微生物群

嵌合 gRNA (SEQ ID NO: 191)

GTTTCAGTTATTTCGTTGGATAGTGTTTAATTGAAACCTATCCAACGAATGAAGTCACTTTTAAA
GTGAGCTGAAATCACTAAAAATTAAGATTGAACCCGGCTACTGACTCTGTCATCCGGGTTTACT
TATTTTC

图41T

Csp1Cas9 sgRNA 优化

初始修饰的核苷酸
未来修饰的核苷酸
修剪的核苷酸

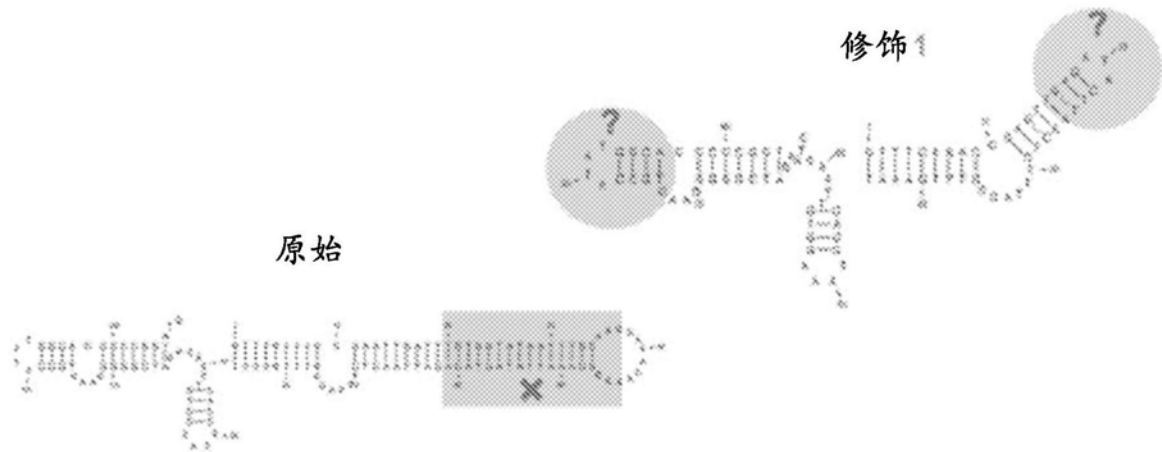


图42A

Csp2Cas9 sgRNA 优化

初始修饰的核苷酸
未来修饰的核苷酸
修剪的核苷酸

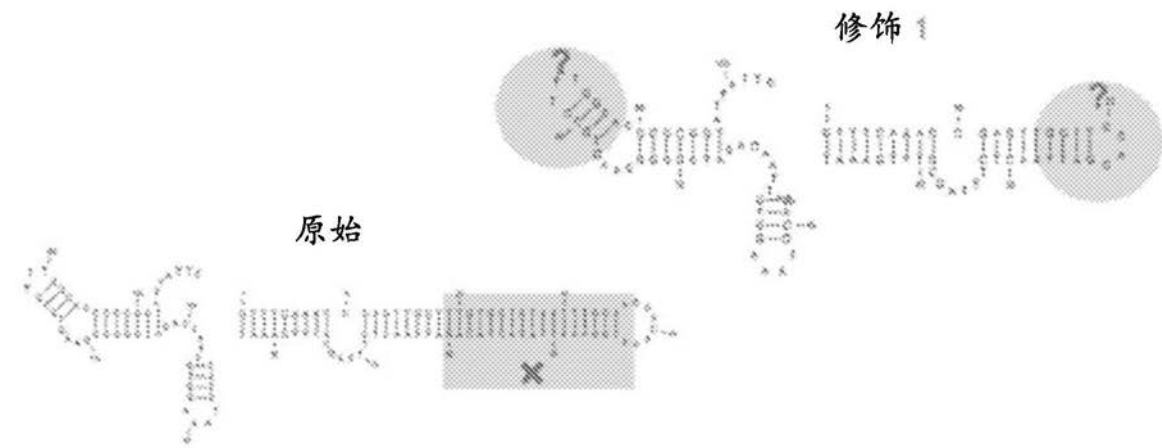


图42B

CI1Cas9 sgRNA 优化

× 初始修剪
? 未来修饰的核苷酸
■ 已经修饰的核苷酸

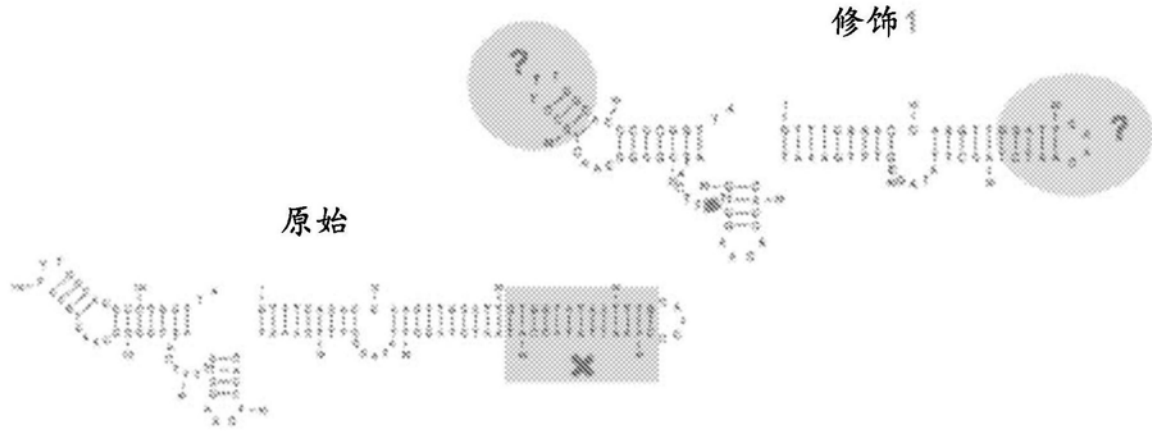


图42C

CI2Cas9 sgRNA 优化

× 初始修剪
? 未来修饰的核苷酸
■ 已经修饰的核苷酸

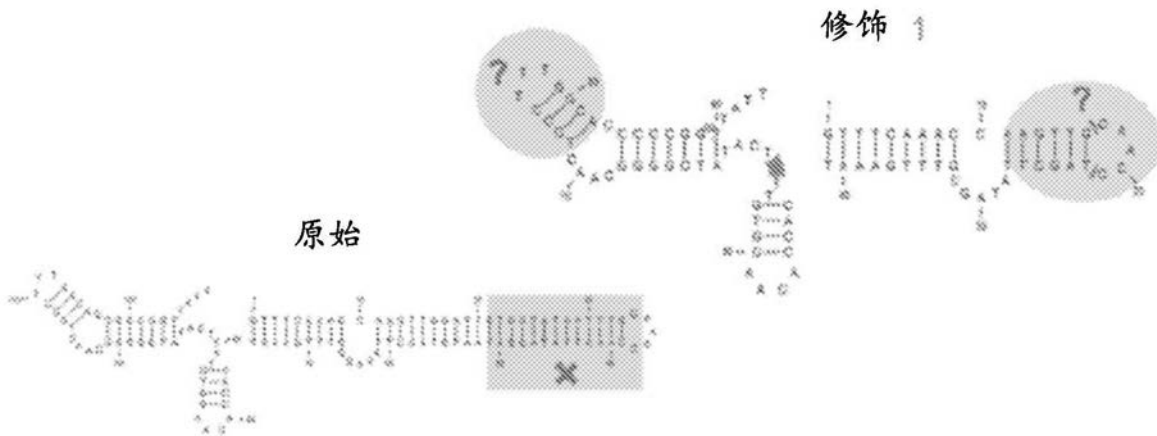


图42D

PeCas9 sgRNA 优化

原始

修饰 1

X = 初始修剪
? = 未来修饰
■ = 经修饰的核苷酸

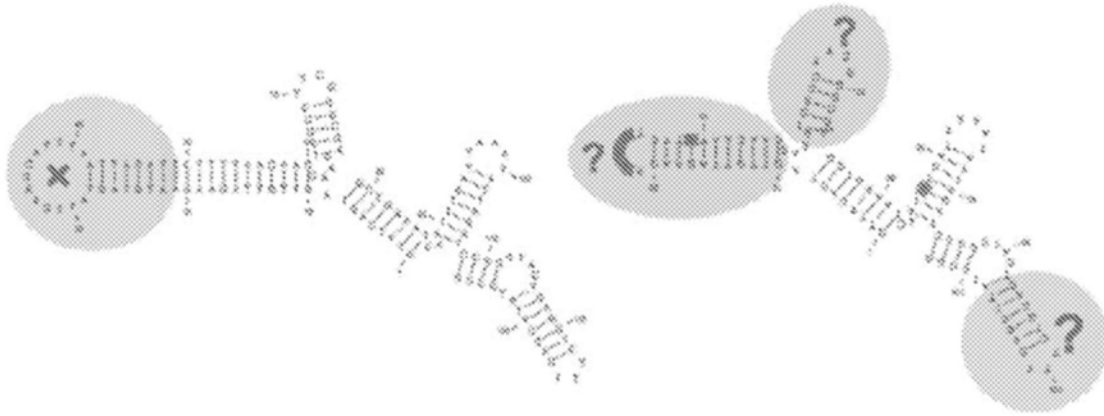


图42E

RaCas9 sgRNA 优化

原始

修饰 1

X = 初始修剪
? = 未来修饰
■ = 经修饰的核苷酸

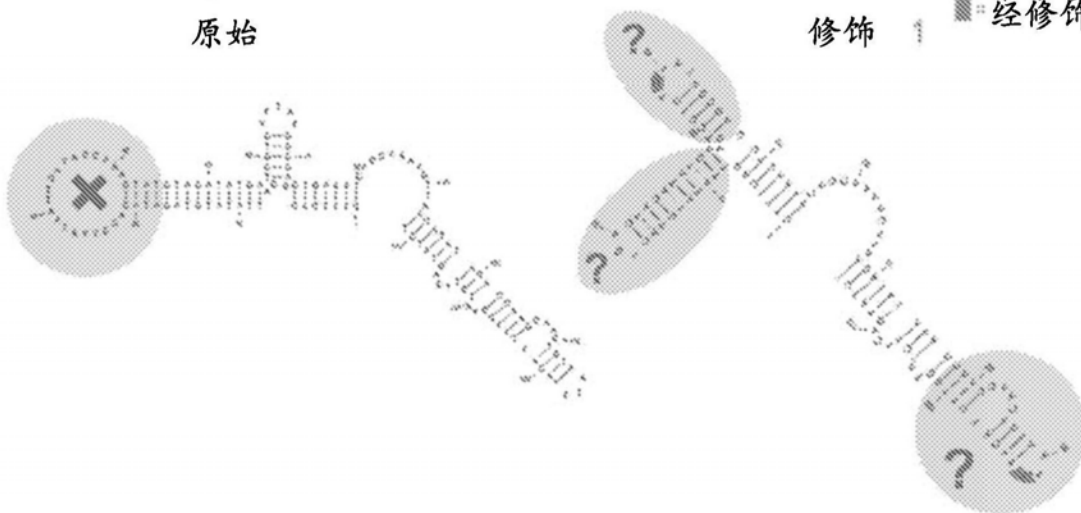


图42F

SwCas9 sgRNA 优化

原始

修饰 1
××× 初始修剪
未修饰的核苷酸
■ ■ ■ 未来修剪的核苷酸

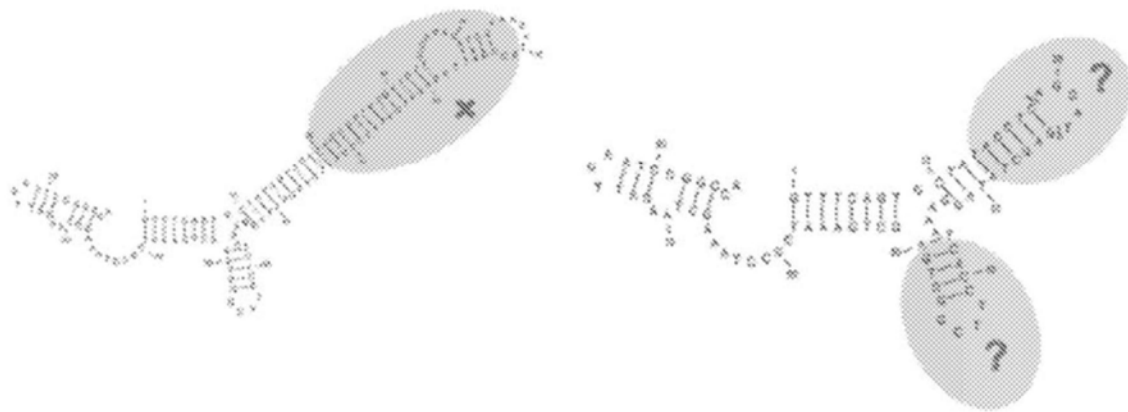


图42G

TmCas9=BbCas9 sgRNA 优化

原始

修饰 1
××× 初始修剪
未修饰的核苷酸
■ ■ ■ 未来修剪的核苷酸

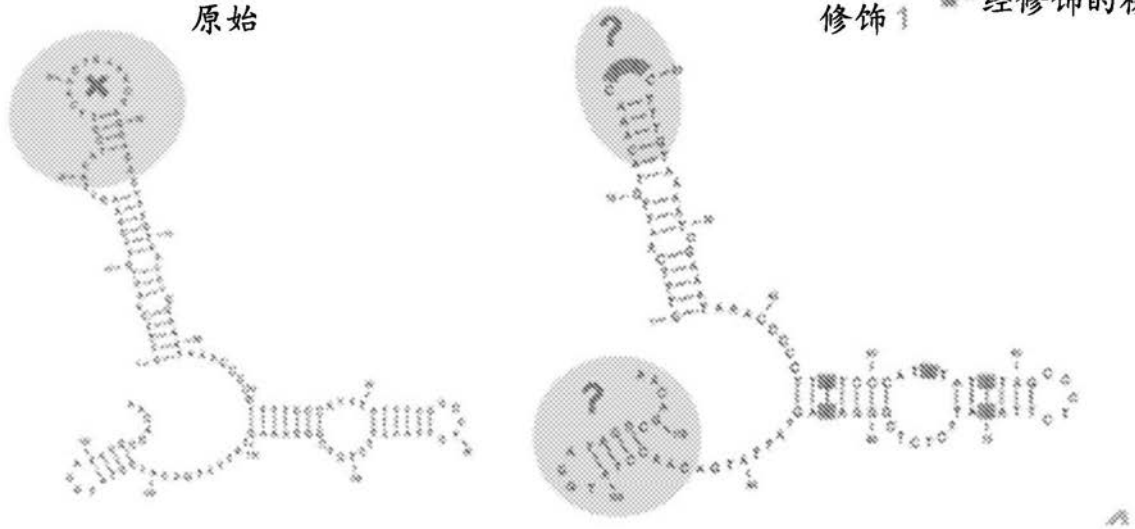


图42H

LpCas9 sgRNA 优化

原始

修饰 1
X 初始修剪
? 未来修饰
■ 经修饰的核苷酸

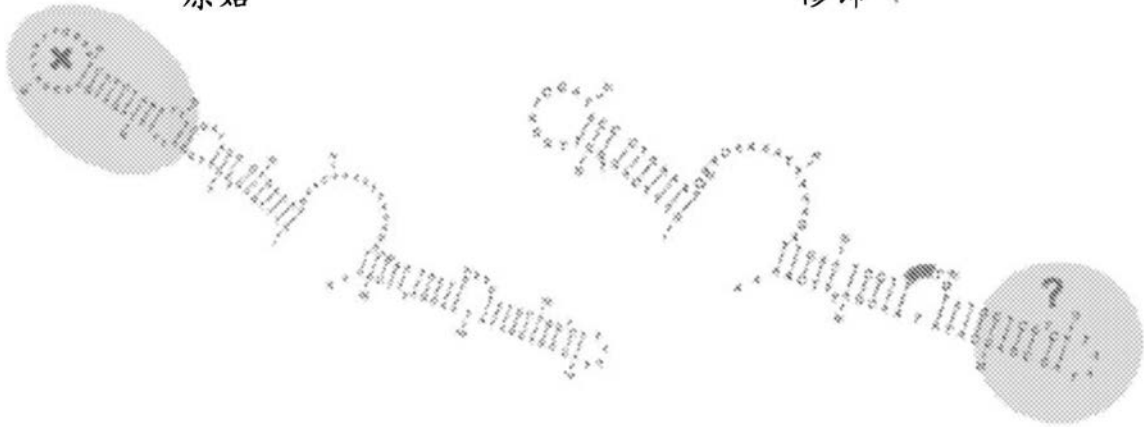


图42I

MhCas9 sgRNA 优化

原始

修饰 1
X 初始修剪
? 未来修饰
■ 经修饰的核苷酸

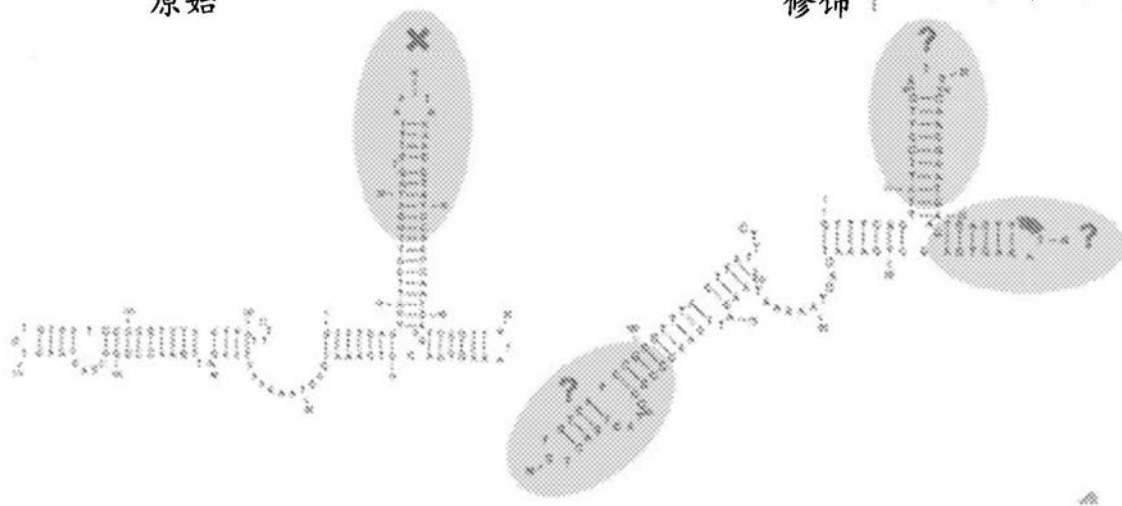


图42J

WsCas9 sgRNA 优化

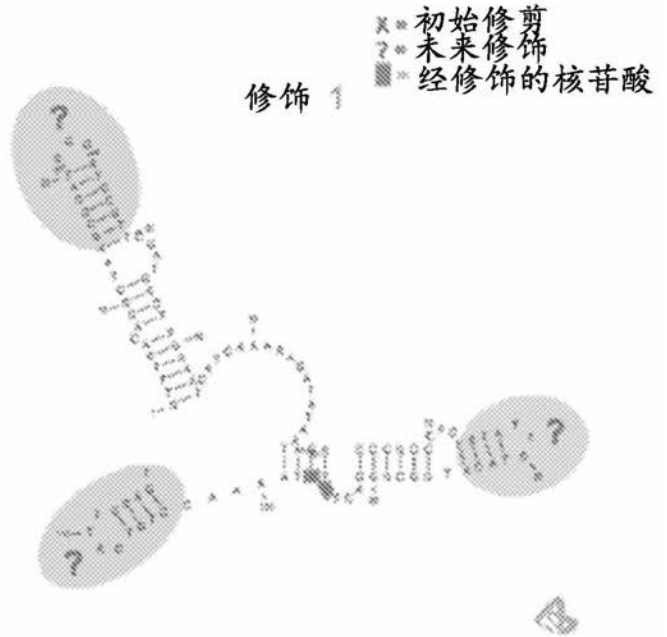
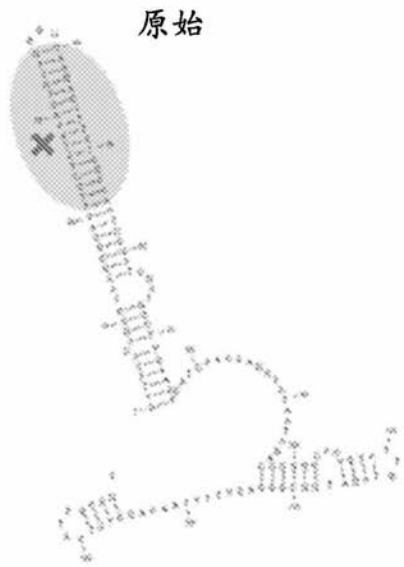


图42K

SsCas9 sgRNA 优化

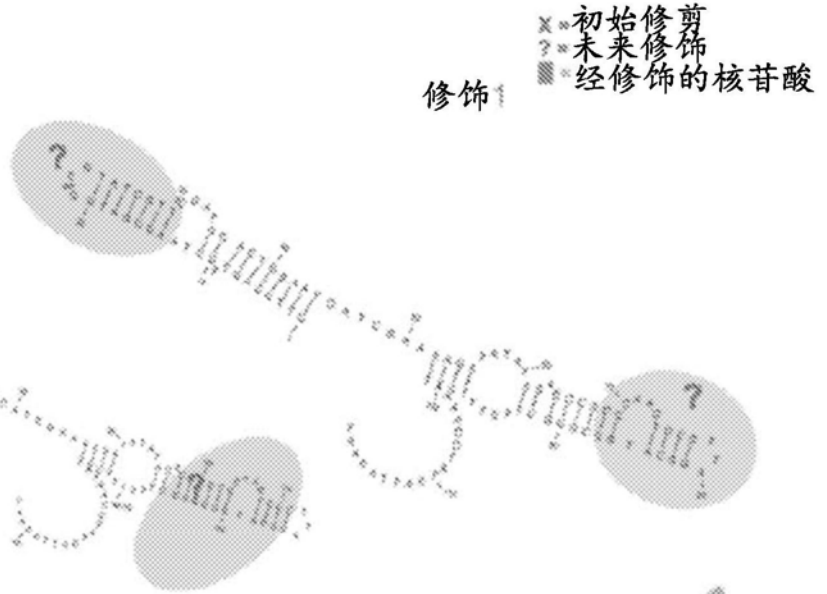
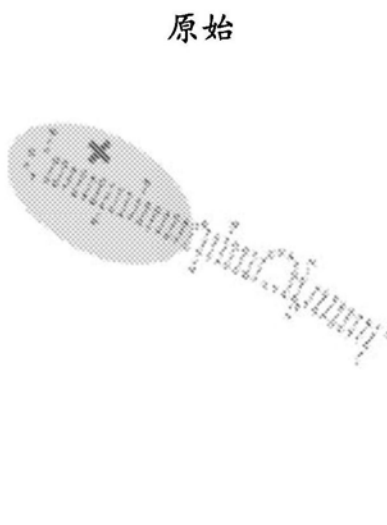


图42L

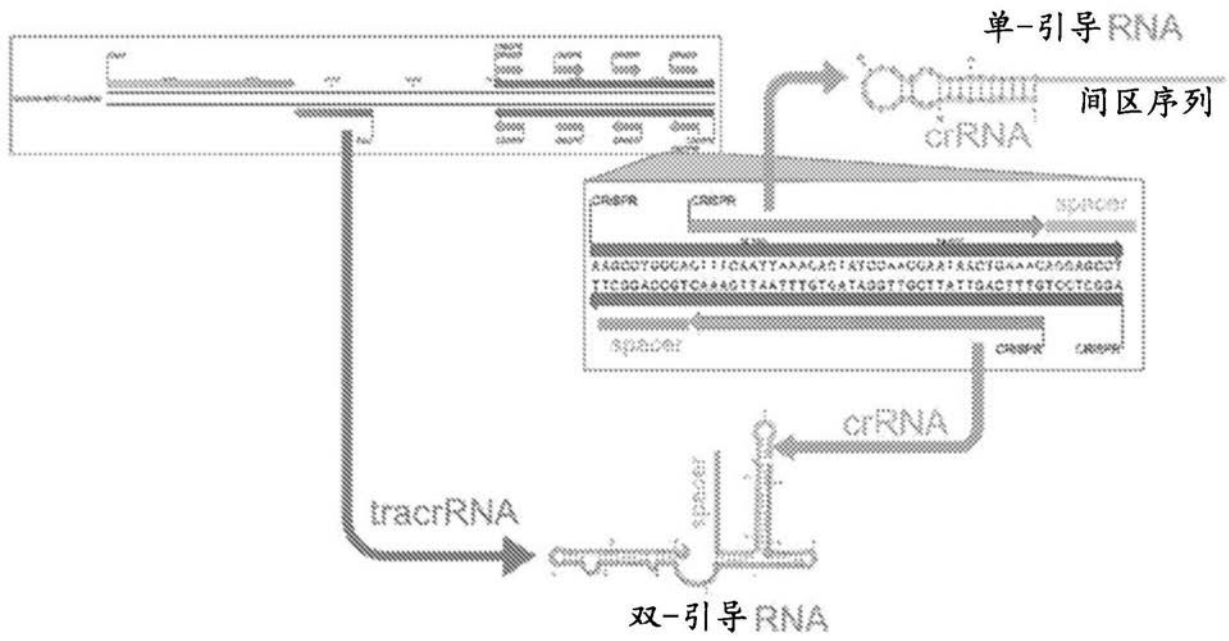


图43

单-引导RNA支架的二级结构的预测

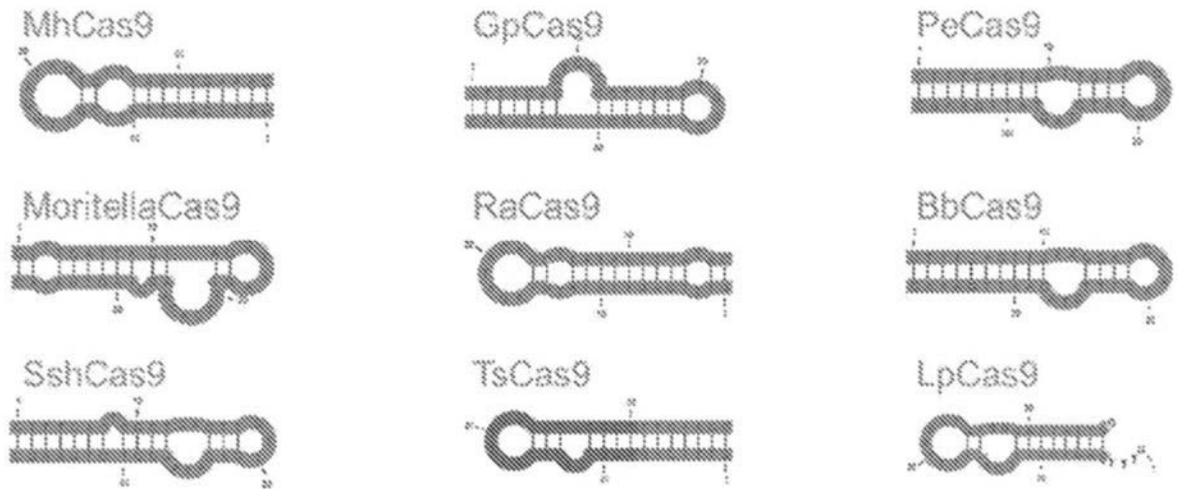


图44

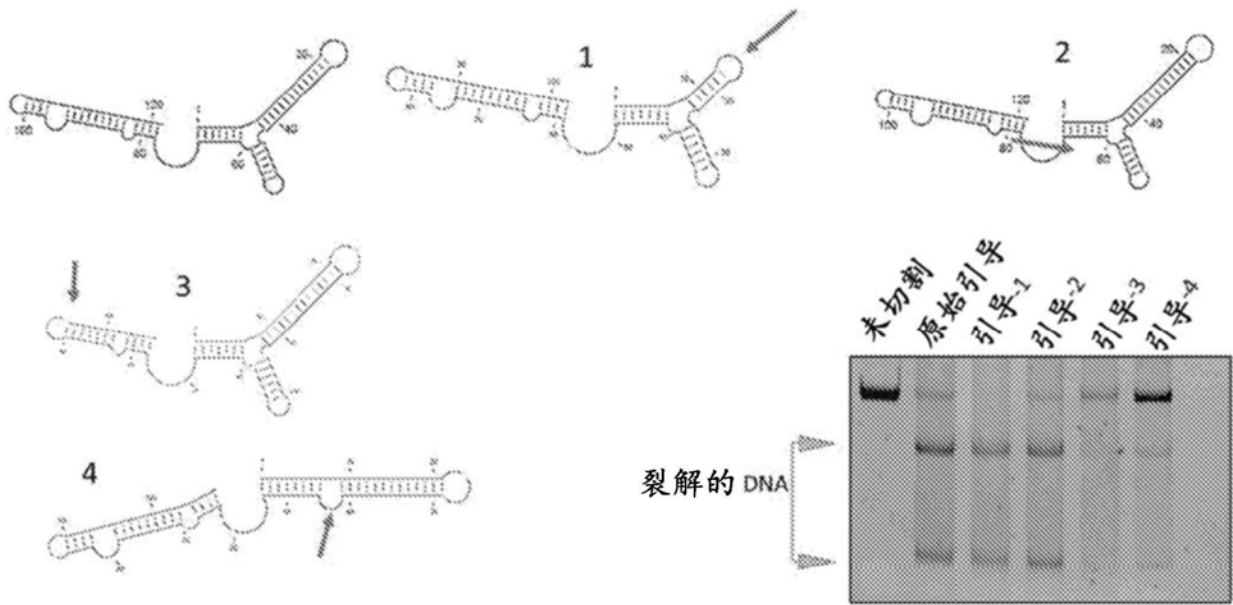


图45

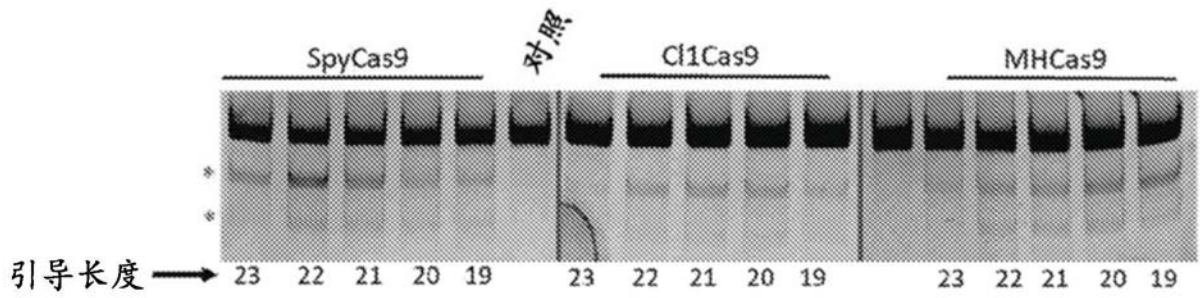


图46

(1). 斯氏弓形杆菌 Cas9 蛋白 (AsCas9) 的氨基酸序列 (SEQ ID NO: 192)

MNTTSIIISLDLGGKNTGFFSCTTNDNFNNLKNFQSGTIIYDESFILSQVNRGKRHTKRNNLRKKLVKRLFLILLKEH
 YKLKIKYLPDEILALFNKRGYTYASFELSNEEQESLSSNILKEFLNENLENFNISNDIEIEDFLNQIASNENAFKQY
 SKDFLNLYESSTFKPKNKIELKDEIKNSYEDKEEQKELFDGLKTVKKILEEFDKQQNQGNLPRAKYFLEIKEEIKTN
 SNIQNFLLKNSNLEEEKINNLIGNISNFQKELRRYFNDEKEMVQGDIIWENKLNKLVWRFITSWHPKKDETIKKNQDE
 LTSNLKNSKIIIEFLTQTNPNKTIIPPYDDMNNRGAVKQCQSLRLNKNYLDTHLPNWRKIANLANDSLKENLKNCTTNK
 SDIDLTLHRLLDLDTSSSIDSYKREYNIENYIDILGKDDSLKFKKFTQNYEYETITKKVRTGIWQKADNIFELCNHNP
 PYKNNQIHTLVSAIIGVEISDTKFKFEFEETLWNKKFPGNKKLVNYCKNIEEIRKRKGNLFKLYIEELKEIEKPDSEQK
 KDINLLKDELLELFWTDEIANFFKLDNIFKSRFSNHFSMAQLYTIETKRAFMSCTCKWCSAENSFRKTNTNIENFTLY
 DKFTGKLEDVIFDENIHIKVYENSNAQRLPADTQRPFSGKIERYIDKLGYEIAKIKAKELENTTEEKKIDLKIVLEQ
 NSFYEYESIRSAKIKNANAKAKKSLEDSKFFFEKSIEEKEKRIKNFNKICLYCNSEITTDGEIDHILPRSYTLKNY
 GTVFNSEGNLLYVHQKCNQSKGNKIYKLEDIKASMNVNEIEEQISKIKSYKTFTLLNQKQQAFAKFFALFLPNSSEAY
 KKVLFGLRDTQSSRVNGTQKYLAKKIQEKLIKMFQKEFDPEFI LASSSEDVSRRLKDYAKQNSILENQKIINNLHLV
 TQAVVALSSVFKKVTNLDLIPTSNKILEFIQINNIPIKPLGKIEKEQIGNLGNGSIFKDSIYQENFAPIYEFENRFY
 TGFLKASKTPDFKNCIEIKSSDFEQIKEYLKDEIKKENGLIIYQINKQKTLDFHNI SNNLSDELKLAQTLDFKFSYT
 LLKTSVENAPKILKDSKKTSTKIYGKTLILPIKDEWERFDKFLKELISQNI EYKEKDGKYQIENEILISICKDYFN
 IQNKNI RNKARKVFLPTVEAPSGGFRI RRKNHKNENIYQLVIGDGAHSVGFALENKKNINSKDVLLPTLVKSKNIS
 FKEPAKKELEQYIDMNNFRDLTKFVNLENINIFVAPASTNRPLVKVELDFNLLKNILKNENLNFLLELDNIKLEDEQK
 AIFDEIFNSCNFAPRDNKIKILKADS NRVLIEFSNSNKTIMKYDKAEDINI

(2). 蜃楼弗朗西斯氏菌 Cas9 蛋白 (FpCas9) 的氨基酸序列 (SEQ ID NO: 193)

MSIKILPIAIDLAKNTGVFSAFYDKGLKLES LITKHGMVYELSKDSYTLMLMSTRAKRHQRRGLDRKQLVKRFLFKL
 IWTEKLLDWDKDTQQTVSFLLNRRGFSFLEEKHDASYFEKFFIEAIELLPQQLRLGIDSEENYDLNSKILETLEND
 TFKDVFEEIESNIKQAKKNIVKANVINI IKS AIDDKVRMGEIKPTKQNKFNLDKVKSLWIVEDIFDSTTSAYIERVD
 NYSKNI FNYLNSLSSDQLEILKTNLPIIDLNEHKNVWVFNISNFNLDNEKLVKLEVDVSAHLHLLGYALQNIKK
 ELDEGRHRSKYFEVAVSLLLEQNHKEGYLRNFCQNLHIGKYKNLVKLVNLI GNLSNLELPLRKYFNDKAHAKA
 DYWDEKRFAYTCRWLIGEWVRVGVKDKDKKDDAKYSYKDLCELKLEVGTTQEVKANKSKGSVVEFLELDPCRTIP
 PYQDNNNRKPPKCSLILNPKYLDKKYPEWQDYLKQLKQSDVYQKYLEEFQDLKGLKTSKQDQYFVGTKSENKQIA
 SGQRDQKDL DARVLQFI FDRVKANDDLLNEIYSQAKKAKQNASNESEIARAVQNLDKTLAKSKLPSALKTKQIKGI
 FEQGTLLHLVCKYFKARQRARDSRLYIMPEYRYDKNLDKYNNTGRFDDDNQLLTYCNHKPRQKRYQLLNDLAGVLQV
 SNGQLLSVEKWFRAQPAETTKSQDEQIFEWLKAFKIASYCKAAVDVQKQYRGT LKNVQTAIFKQRLETIKKNKKA
 SQDEKALLEKYKNAKPLTADEKNLVKLVNISKASQKIGENLGLDNKQVSKFNSIYSFAQIQQIAFAERNGNANTCA
 VCSADNAHRMQISNDSAKAQRPAIPTRIVDGAVKKMATILARNIVDDNWDNIIKKTLENKEELHIPIITESNAFEFE
 PNLAEIKGTTKKKDRIS SQELFAKKEERIKKASNEISAYSGASLGGETELDHIIPRSHKKYGTLNDEANLICVTRDDN
 QKRGNTPYLNSLADKYKLLKFGTADNQEIEKYIADKIWDSEKQDFKFGNYRFINLSQEEQIAFRHALFLADENPI
 KQAVIKAIIDNRNRAVNGTQRYFAEVLANNIYLRAKKENLATNKLSDYFVGVETNSNGRGIADIRKLYEAIKSDIE
 AYAKGDKPQDSYSHLIDAMIAFCVADEHKNDGSI GLEIDETYSLYPLDENTGEVFKEDIFSQIKVADSEFRETCLA
 RRKAYTVETNIRQEIQSKGDFKKSIPYKIHIDLSAERFFPLLQISQDKWKYGFITLNSVDIPKKDIELIQEFLTKS
 KEYKDYTVWIVNKTQAQEFLINIGYVGADKQEQEIAKLLDKLSYTTIKKEITSILAKANKTPDVTVDALDVIDELVF
 KTDKKNKLIN FVKSEIILLPVYNDWIMLKKAMQKADMNQNLHDFLSMYFITTNLKNKHQKVRKNFSLPVVSTIGTIRI
 RRKSWDGSNIYQVTGDES LAKYGFDTVRPKTILSRNSVPIKYNGKEKEVLI EPLSWIILDIDDLTDTSIIRAQIK
 NRSAGQVSIKVELSSLDNICLKEDSWSGNIYAGDKDKPNTIKGDKFYIKKEDFYWLKDFSLMFDGSKTKASLSIDK
 IGNSYILEFVSSKGGKKEKQWLKNETPNS

(3). 西班牙弗朗西斯氏菌 Cas9 蛋白 (FhCas9) 的氨基酸序列 (SEQ ID NO: 194)

MNIKILPIAIDLAKNTGVFSAFYQKGTSLKESLDNKNKGVYELSKDSYTLMLMNRRTARRHQRRGIDRQKLVKRLFKL
 IWTQKLNLEWNKDTQQTISFLLNRRGFSFITDGYSPYLNIAPEPVKVI LMSILDDYNGEDDLDSYLQSATENDSKI
 DELYNKLLQKTFEFLKRLKCIDIKEDKVTTKTLKELSNTEFKLLANYLVYDRI LRTQKFSYTDKQGNLRELNYHH
 DKYNIQEFKRNIIINDVILEKLTDDLDIWNLFDFKDFEKNLEKLENQEDKDYLOTHLHHFVAVNKKISEMASGG
 RHRSYQFQEITNVLVENNHQEGYLKFNFCENLHNKKYSNLSVKNLVNLI GNLSNLELPLRKYFNDKIHAKADYWDEQ
 KFAETYSDWILGWRVGVAKDKDKKDGAKYSYKILCDELKQKVGINQDGIINQTKGDFVGFLELDPCRTIPPYLDNN
 NRKPPKCSLILNPKFLDNKYPNWQYQLQELKLLQTVQNYLGNFEIDLKDLKSSKEQPYFVKYKSSNQI IASGQRDY
 KDLDARVLQFI FDRVKASDELLELNEIYSHAKKQNI SSELEKLELCKKLDKI IANSQLSQILKSQHINGIFEQGT
 LHLVCKYKQRQRARDSRLYIMPEYRYDKKLDKYNNTGRFDDDNQLLTYCNHKPRQKRYQLLNDLAGVLQVSPNFLK

DKIGSDDDLFISKWLVEHIRGFKKACEDSLKIQKDNRGLLNHKINIARNTKGKCEKEIFNLICKIEGSEDKKGNKYKH
GLAYELGVLLFGEPNQASKLEFDRKIKKFNSIYSFAQIQQIAFAERKGNANTCAVCSADNAHRMQQIKVAKPVEGNK
DNIILSAKAQRLPAIPTRIVDGAVKKMATILARNIVDDNWDNIKQALSNNQQLHVPIITESNAFEFEPALADVKGKS
LKDKRKKALERINPENTFKDKNNRIKEFAKGISAYSGDNLANGDFDGAKEELDHIIPRAHKKYGTLNDEANLICVTR
EDNQNRGNKAVFLYDLKPNYKQFDTTDDLEIEKKIADTIWDASKQDFKFGNYRSFINLTPQEQAFRHALFLADE
NPIKKAVIRAINNRNRTFVNGTQRYFAEVLANNIYLRAKKENLATNRITFDYFGIETTNSNGRGVADVRKLYEKVDS
DIQAYAKGDKPQDSYSHLIDAMIAFCVAADDEHKNGGSGLEIDKNYSLYPLDKNTGEVFSKDI FSQIKIADNEFSDK
KLVRKKATEGFNTHRQMTDRDGIYGESYLPILIHKNLNEVRKGYNWENSEEIKIFKGKKYDIQQLNNLVYCLKFVDFP
ISIDIQITTLLEELRNILETNNISTTAEYYYINLKTQKLHEYYIENYNTALGYKKYTKEMEFLRS LAYRTERVKIKSI
DDVSMILAKDSNFKAGKIELPFKTEWQRLYLEWQNTTIKDNIEFLKSYFSVKNATKQHKKVRKDFSLPISTNEGKLL
VKRKTWDDNNFIYQILNDSDSRADGTPFFIPAFDISKNEIVETIISKFTSKNIFWLPKNIKLQKVDNKSIFAIDTSRW
FEVETPKDLIEIGVSTIQYKIDNNSRPKVRVKLDYVMDDDSKINYFMNHSLLKSRYDPKVLIEILKQSTIIEFESSGF
NKTIKEMLGMTLAGIYNETLNN

(4). 嗜盐副内生单胞菌Cas9 蛋白 (PhCas9) 的氨基酸序列 (SEQ ID NO: 195)

MRYISPI SIDFGAKYTG IHNMHYLEGEDPRGSVQNTGCVLVMDDDNLQLSQVGRTOQRHQS RNYKRODLARRLLI VI
LKDCFDQDVWQWPVKNREFLFG LINRRGFNRFDLDDSLIEIEFCDLSVFESVLPDIAGYEGLEEWLQALAEANAECRL
ALQRLGRTREFLKTLP EIWKEEKKSIEATHKHI IHFLETSINASEEGHLHRRKYL RNIAADI ILEQHGRLASLFQTN
SGITPDQFANLVGHI SNLQLRVLRRYFNDQNAKPQNGQTD AWRDHF SKQFWRWIKSWHCDEQTKSRWHEL IKFKKK
GGSVFKMLTVQPELTI PPYEDQNNRRPPKDKTLFLN PVALNKAYPLWREIVKRLVAAHPELDS DHASTIAHMDAAL
DQDCILLQRMLDRTSGLDLYKIRALAMGRPDRGNQLK VASENLRLRLNQHAEPFIQS IARPYLQINEARQGLWVEE
DSL LLLCLNRTPPQRKARQLPTLLGRIVGDLGSHGQEH I QSLKQQWSQRGVI KGARTPKGWCKFAAEAQKDFGNGLKET
IRSYRQLQARIKAGDTEEKLLPEAKKLLALEAALPDVARFIAQGLGLAAENSARFANPFSLAQIYNLLESDFSGF SH
VSKPSAHENLWRMQYEQSVDGKDCARASVLSADSIRPFDGMLSRI LEKQAYSIAKRKLKQIESHEMTSAKVRVPI LI
EENRFQFAEDLQTIKKNTRKKKDAQKSCESDFIRFEDKEQRI CKASKGMCAYTGEDLGRYGEFDHII PRSYTQKRTG
AIFNSEANLIFVSSRGNQQKRDQLYTLAELDDRYLTSQFGSTNRQQIAQAVQDAVTPFLTKGELIGFHNLGLQE QRM
IRHALFVPGLRDQVLQ LLAGQRKAKVNGTQAWLAKRI IHHIRQGASATIEFDI IKVPTDSVRFCRGV LASRFPAMAK
QSPQPVASHAIDAALVSGSLMEPAIARSLSLAPVDTTTQAEWLKGLLPETIDIQILRKKPL YRKDAPASSPLFKAG
LYGERFLPLLVMNDEL RVGFTLPNSAKVLDHADNLYALLRPVLVCKDSHELPSTLAELQESLALGQRQILNINRQKA
FAYLHEFGQSPTHNDNCLAI EALLYCVQKKI DAHIYNKDKKTFKKDEDI LADKFFI FETHKNNRPI FPRTVARI EK
KDLTL PFKQAWVELCDTDELKNFLGKQTL PNNFWRNFGSSLFPSIHNEHDHQHKTRKVYSLPLKISEG SYRVRHT
PEGASVWQRMSVGD FGAAGFAKIDGKTQWNKVEI PHFAKSKNLHSLNRSLEKNPQSITYFDEWRS L CALADALPNG
VIKVACAPGTVGRPYIRFEMLESQFIKKINAASGSKYKSSWELPMEI PNVQLAGFDI FSCPLLKPRPTETTKPKNKHK
ARFLRLGKTVLIEYFASSKANKALQQA YEQGEPECI

图47

- (1) 斯氏弓形杆菌 Cas9蛋白(AsCas9)的crRNA (+)序列(SEQ ID NO: 196)
GCGACACTTTACAACAATATCGCTTAGCTAATGAAAC
- (2) 斯氏弓形杆菌 Cas9蛋白(AsCas9)的crRNA (-)序列(SEQ ID NO: 200)
GTTTCATTAGCTAAGCGATATTGTTGTAAAGTGTTCGC
- (3) 蜃楼弗朗西斯氏菌 Cas9蛋白(FpCas9)的crRNA (+)序列(SEQ ID NO: 197)
CCAACAGTAGTATACCAAATATTCTAGCTACTGAAAC
- (4) 蜃楼弗朗西斯氏菌 Cas9蛋白(FpCas9)的crRNA (-)序列(SEQ ID NO: 201)
GTTTCAGTAGCTAGAATATTTGGTATACTACTGTTGG
- (5) 西班牙弗朗西斯氏菌 Cas9蛋白(FhCas9)的crRNA (+)序列(SEQ ID NO: 198)
CTAACAGCAGTATATCAAATAATCTAGCTACTGAAAC
- (6) 西班牙弗朗西斯氏菌 Cas9蛋白(FhCas9)的crRNA (-)序列(SEQ ID NO: 202)
TTTCAGTAGCTAGATTATTTGATATACTGCTGTTAG
- (7) 嗜盐副内生单胞菌 Cas9 (PhCas9)的crRNA (+)序列(SEQ ID NO: 199)
GAATCAAAGCTTACTTCGTCTGGGTAGCAACTGAAAG
- (8) 嗜盐副内生单胞菌 Cas9 (PhCas9)的crRNA (-)序列(SEQ ID NO: 203)
CTTTCAGTTGCTACCCAGACGAAGTAAGCTTTGATTC

图48

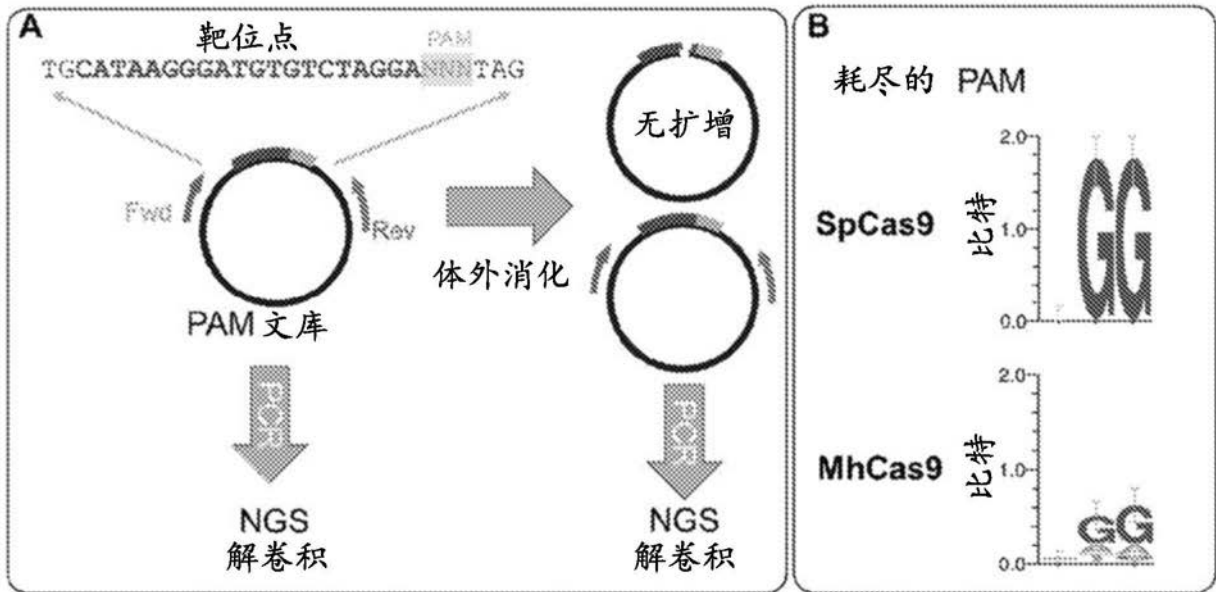


图49

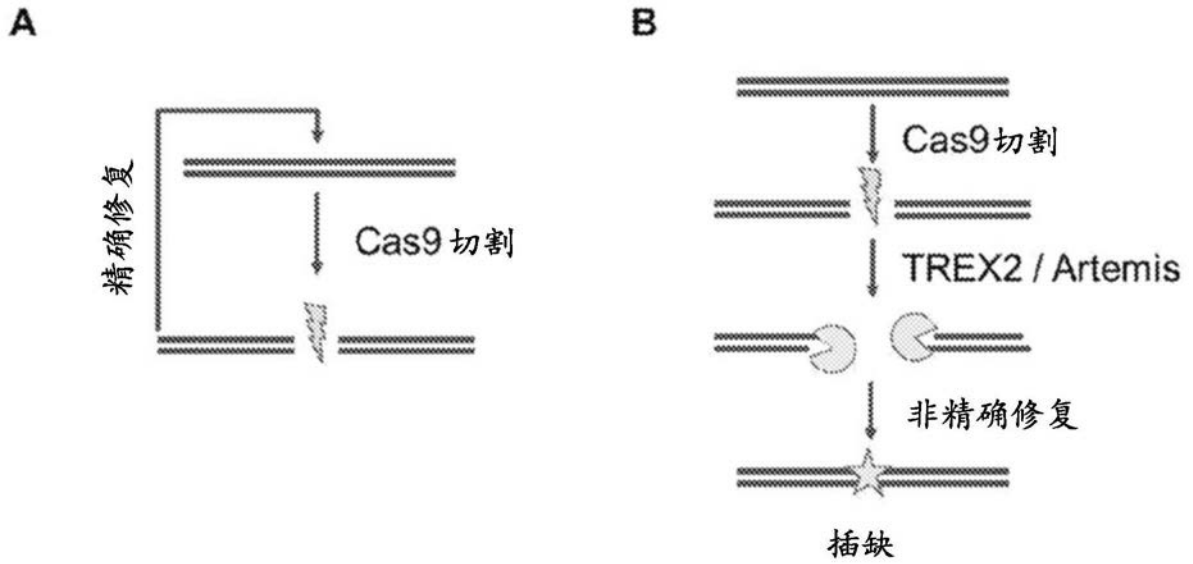


图50

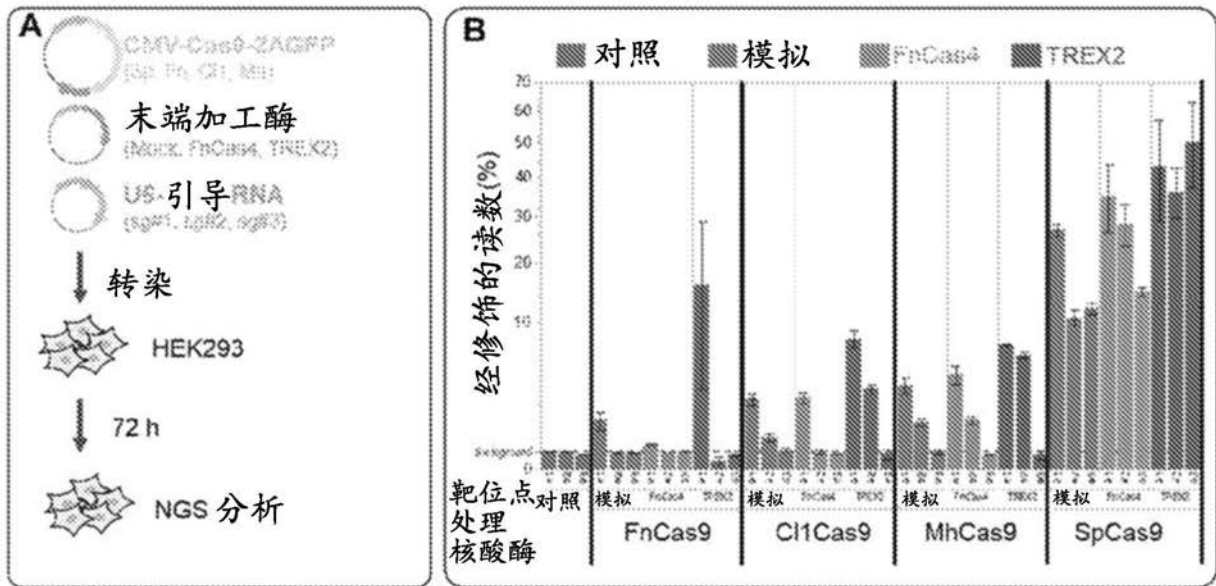


图51

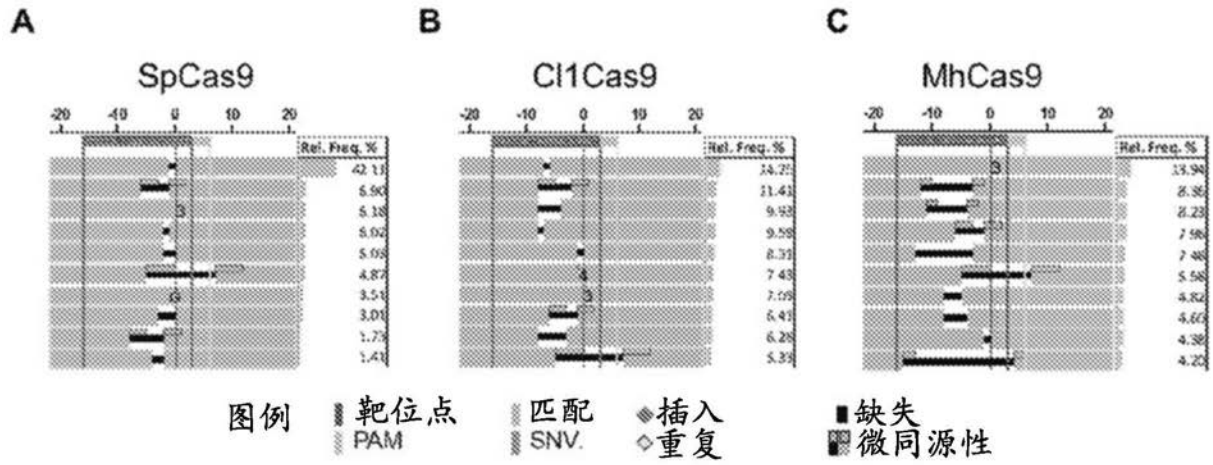
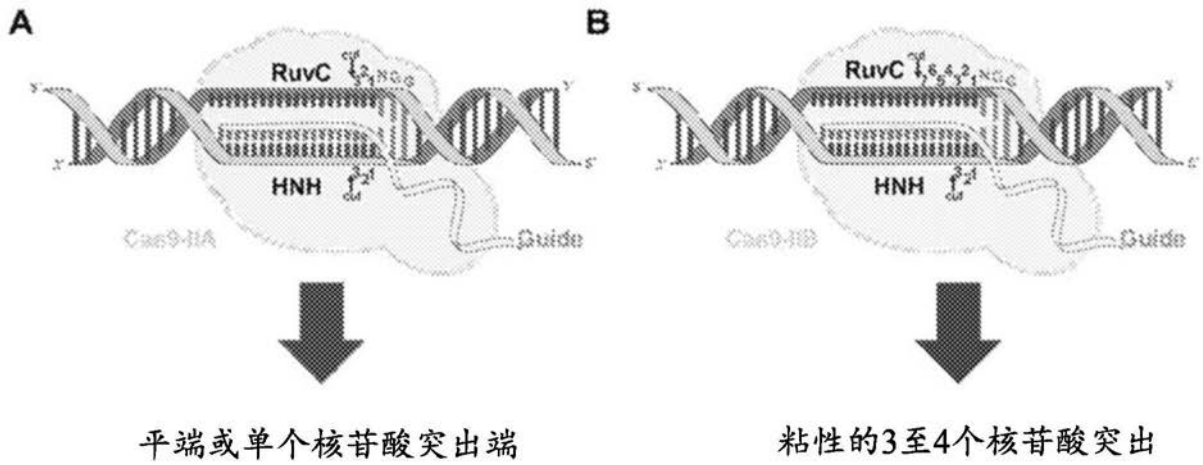


图52



4845-9328-0128, v. 1

图53