



(10) 授权公告号 CN 110312792 B

(45) 授权公告日 2024. 04. 09

(21) 申请号 201680092098.5

(22) 申请日 2016.12.14

(65) 同一申请的已公布的文献号
申请公布号 CN 110312792 A

(43) 申请公布日 2019.10.08

(85) PCT国际申请进入国家阶段日
2019.08.14

(86) PCT国际申请的申请数据
PCT/EP2016/081077 2016.12.14

(87) PCT国际申请的公布数据
W02018/108272 EN 2018.06.21

(73) 专利权人 瓦赫宁根大学
地址 荷兰瓦赫宁根
专利权人 科学技术基金会

(72) 发明人 约翰·万德奥斯特
理查德·范克拉嫩堡
艾勒克·芬纳·博斯马
扬尼斯·莫加科斯

(74) 专利代理机构 北京安信方达知识产权代理有限公司 11262

专利代理人 王玮玮 郑霞

(51) Int.Cl.
C12N 9/22 (2006.01)
C12N 15/113 (2006.01)

(56) 对比文件
WO 2016186946 A1, 2016.11.24
WO 2016179038 A1, 2016.11.10
CN 101641441 A, 2010.02.03
CN 1526013 A, 2004.09.01
吴彩云等. 基因组编辑新技术: CRISPR/Cas 系统在生物基因组学中的研究进展.《植物生理学报》. 2015, (第12期),
NCBI.type II CRISPR RNA-guided endonuclease Cas9 [Geobacillus stearothermophilus].《NCBI Reference Sequence: WP_064213580.1》. 2016,

审查员 崔力方

权利要求书8页 说明书44页
序列表18页 附图16页

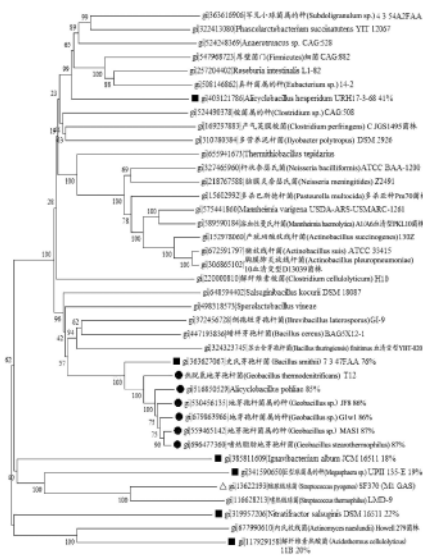
(54) 发明名称

热稳定的Cas9核酸酶

(57) 摘要

热稳定的Cas9核酸酶。本发明涉及遗传工程领域,并且更特别地涉及核酸编辑和基因组修饰。本发明提供了分离的Cas蛋白或其多肽片段,所述分离的Cas蛋白或其多肽片段具有SEQ ID NO:1的氨基酸序列或与SEQ ID NO:1具有至少77%同一性的序列。Cas蛋白或多肽能够在30℃和100℃的范围中(包括端点)的温度结合、裂解、标记或修饰双链靶多核苷酸。本发明还提供了编码所述Cas9核酸酶的分离的核酸分子、表达载体和宿主细胞。本发明还提供了被Cas蛋白或多肽识别的PAM序列。本文所公开的Cas9核酸酶提供了用于在升高的温度进行遗传工程的新型工具,并且在嗜热生物体(特别地,微生物)的遗传操作

中具有特别的价值。



1. 至少一种靶向RNA分子和Cas蛋白用于结合、裂解、标记或修饰包含靶核酸序列的双链靶多核苷酸的用途,其中:

所述双链靶多核苷酸包含靶核酸链和非靶核酸链,所述靶核酸链包含所述靶核酸序列,所述非靶核酸链包含与所述靶核酸序列互补的前间区核酸序列;

所述Cas蛋白由SEQ ID NO:1的氨基酸序列组成;

所述至少一种靶向RNA分子识别所述靶核酸序列;

所述非靶核酸链还包含与所述前间区核酸序列的3'末端直接地相邻的前间区相邻基序(PAM)序列,其中所述PAM序列包含5'-NNNNCNN-3',其中所述用途不用于人类或动物治疗。

2. 至少一种靶向RNA分子和Cas蛋白在制备用于结合、裂解、标记或修饰包含靶核酸序列的双链靶多核苷酸的试剂盒中的用途,其中:

所述双链靶多核苷酸包含靶核酸链和非靶核酸链,所述靶核酸链包含所述靶核酸序列,所述非靶核酸链包含与所述靶核酸序列互补的前间区核酸序列;

所述Cas蛋白由SEQ ID NO:1的氨基酸序列组成;

所述至少一种靶向RNA分子识别所述靶核酸序列;

所述非靶核酸链还包含与所述前间区核酸序列的3'末端直接地相邻的前间区相邻基序(PAM)序列,其中所述PAM序列包含5'-NNNNCNN-3'。

3. 根据权利要求1或2所述的用途,其中所述结合、裂解、标记或修饰发生在20℃和100℃之间的温度。

4. 根据权利要求1或2所述的用途,其中包含所述靶核酸序列的所述双链靶多核苷酸被所述Cas蛋白裂解。

5. 根据权利要求3所述的用途,其中包含所述靶核酸序列的所述双链靶多核苷酸被所述Cas蛋白裂解。

6. 根据权利要求1或2所述的用途,其中包含所述靶核酸序列的所述靶核酸链是双链DNA,并且所述用途引起包含所述靶核酸序列的所述双链靶多核苷酸中的双链断裂。

7. 根据权利要求3所述的用途,其中包含所述靶核酸序列的所述靶核酸链是双链DNA,并且所述用途引起包含所述靶核酸序列的所述双链靶多核苷酸中的双链断裂。

8. 根据权利要求1或2所述的用途,其中包含所述靶核酸序列的所述双链靶多核苷酸是双链DNA,所述Cas蛋白缺乏切割所述双链DNA的能力,并且所述用途引起所述双链靶多核苷酸的基因沉默。

9. 根据权利要求3所述的用途,其中包含所述靶核酸序列的所述双链靶多核苷酸是双链DNA,所述Cas蛋白缺乏切割所述双链DNA的能力,并且所述用途引起所述双链靶多核苷酸的基因沉默。

10. 根据权利要求1或2所述的用途,其中所述PAM序列包含5'-NNNNCNNA-3'、5'-CNNNCNN-3'、5'-NNNCCNN-3'、5'-NNCNCNN-3'、5'-NNNNCCN-3'和/或5'-NCNNCNN-3'。

11. 根据权利要求3所述的用途,其中所述PAM序列包含5'-NNNNCNNA-3'、5'-CNNNCNN-3'、5'-NNNCCNN-3'、5'-NNCNCNN-3'、5'-NNNNCCN-3'和/或5'-NCNNCNN-3'。

12. 根据权利要求1或2所述的用途,其中所述PAM序列包含5'-CCCCCNA-3' [SEQ ID NO:10]。

13. 根据权利要求3所述的用途, 其中所述PAM序列包含5' -CCCCCNA-3' [SEQ ID NO: 10]。

14. 根据权利要求12所述的用途, 其中所述PAM序列包含5' -CCCCCAA-3' [SEQ ID NO: 11]。

15. 根据权利要求13所述的用途, 其中所述PAM序列包含5' -CCCCCAA-3' [SEQ ID NO: 11]。

16. 根据权利要求1或2所述的用途, 其中所述Cas蛋白从热脱氮地芽孢杆菌(*Geobacillus thermodenitrificans*) T12可获得。

17. 根据权利要求3所述的用途, 其中所述Cas蛋白从热脱氮地芽孢杆菌T12可获得。

18. 根据权利要求1或2所述的用途, 其中所述靶向RNA分子包含crRNA和tracrRNA。

19. 根据权利要求3所述的用途, 其中所述靶向RNA分子包含crRNA和tracrRNA。

20. 根据权利要求1或2所述的用途, 其中所述至少一种靶向RNA分子的长度是在35-200个核苷酸残基的范围中。

21. 根据权利要求3所述的用途, 其中所述至少一种靶向RNA分子的长度是在35-200个核苷酸残基的范围中。

22. 根据权利要求1或2所述的用途, 其中所述靶核酸序列的长度是从15个至32个核苷酸残基。

23. 根据权利要求3所述的用途, 其中所述靶核酸序列的长度是从15个至32个核苷酸残基。

24. 根据权利要求1或2所述的用途, 其中所述Cas蛋白作为蛋白复合体的一部分被提供, 所述蛋白复合体包含至少一种另外的蛋白, 所述至少一种另外的蛋白选自解旋酶、核酸酶、解旋酶-核酸酶、DNA甲基化酶、组蛋白甲基化酶、乙酰基转移酶、磷酸酶、激酶、转录(共)活化物、转录阻遏物、DNA结合蛋白、DNA结构蛋白、标志物蛋白、报告物蛋白、荧光蛋白、配体结合蛋白、信号肽、亚细胞定位序列、抗体表位或亲和纯化标签。

25. 根据权利要求3所述的用途, 其中所述Cas蛋白作为蛋白复合体的一部分被提供, 所述蛋白复合体包含至少一种另外的蛋白, 所述至少一种另外的蛋白选自解旋酶、核酸酶、解旋酶-核酸酶、DNA甲基化酶、组蛋白甲基化酶、乙酰基转移酶、磷酸酶、激酶、转录(共)活化物、转录阻遏物、DNA结合蛋白、DNA结构蛋白、标志物蛋白、报告物蛋白、荧光蛋白、配体结合蛋白、信号肽、亚细胞定位序列、抗体表位或亲和纯化标签。

26. 根据权利要求24所述的用途, 其中所述Cas蛋白或另外的蛋白包含与所述Cas蛋白或蛋白复合体的N-末端和/或C-末端融合或连接的至少一种另外的蛋白。

27. 根据权利要求25所述的用途, 其中所述Cas蛋白或另外的蛋白包含与所述Cas蛋白或蛋白复合体的N-末端和/或C-末端融合或连接的至少一种另外的蛋白。

28. 根据权利要求24所述的用途, 其中所述另外的蛋白是绿色荧光蛋白(GFP)。

29. 根据权利要求25-27中任一项所述的用途, 其中所述另外的蛋白是绿色荧光蛋白(GFP)。

30. 根据权利要求24所述的用途, 其中所述Cas蛋白的天然活性被失活并且所述Cas蛋白与标志物蛋白或包含核酸酶结构域的蛋白连接。

31. 根据权利要求25-27中任一项所述的用途, 其中所述Cas蛋白的天然活性被失活并

且所述Cas蛋白与标志物蛋白或包含核酸酶结构域的蛋白连接。

32. 根据权利要求30所述的用途,其中所述核酸酶结构域是FokI核酸酶结构域。

33. 根据权利要求31所述的用途,其中所述核酸酶结构域是FokI核酸酶结构域。

34. 根据权利要求24所述的用途,其中所述双链靶多核苷酸是dsDNA,所述至少一种另外的蛋白是核酸酶或解旋酶-核酸酶,并且所述修饰是在期望的基因座处的单链断裂或双链断裂。

35. 根据权利要求25-27中任一项所述的用途,其中所述双链靶多核苷酸是dsDNA,所述至少一种另外的蛋白是核酸酶或解旋酶-核酸酶,并且所述修饰是在期望的基因座处的单链断裂或双链断裂。

36. 根据权利要求24所述的用途,其中所述双链靶多核苷酸是dsDNA,并且所述另外的蛋白选自DNA修饰酶、转录活化物或转录阻遏物,并且所述结合、裂解、标记或修饰引起基因表达的修饰。

37. 根据权利要求25-27中任一项所述的用途,其中所述双链靶多核苷酸是dsDNA,并且所述另外的蛋白选自DNA修饰酶、转录活化物或转录阻遏物,并且所述结合、裂解、标记或修饰引起基因表达的修饰。

38. 根据权利要求24所述的用途,其中所述结合、裂解、标记或修饰发生在细菌细胞内。

39. 根据权利要求25-27中任一项所述的用途,其中所述结合、裂解、标记或修饰发生在细菌细胞内。

40. 根据权利要求1或2所述的用途,其中所述结合、裂解、标记或修饰引起在期望的位置处修饰或缺失和/或插入期望的核苷酸序列,和/或其中所述结合、裂解、标记或修饰引起使在期望的基因座处的基因表达沉默。

41. 一种结合、裂解、标记或修饰双链靶多核苷酸的方法,其中所述双链靶多核苷酸包含靶核酸链和非靶核酸链,所述靶核酸链包含靶核酸序列,所述非靶核酸链包含与所述靶核酸序列互补的前间区核酸序列,所述方法包括:

a. 设计至少一种靶向RNA分子,其中所述靶向RNA分子识别所述靶核酸链中的所述靶核酸序列,并且所述非靶核酸链还包含与所述前间区核酸序列的3'末端直接地相邻的前间区相邻基序(PAM)序列,其中所述PAM序列包含5'-NNNCCNN-3';

b. 形成包含所述靶向RNA分子和Cas蛋白的核糖核蛋白复合体,其中分离的所述Cas蛋白由SEQ ID NO:1的氨基酸序列组成;和

c. 所述核糖核蛋白复合体结合、裂解、标记或修饰所述双链靶多核苷酸,其中所述方法不用于人类或动物治疗。

42. 根据权利要求41所述的方法,其中所述结合、裂解、标记或修饰发生在20℃和100℃之间的温度。

43. 根据权利要求41或权利要求42所述的方法,其中包含所述靶核酸序列的所述双链靶多核苷酸被所述Cas蛋白裂解。

44. 根据权利要求41或权利要求42所述的方法,其中所述双链靶多核苷酸是双链DNA,并且所述方法引起所述双链靶多核苷酸中的双链断裂。

45. 根据权利要求41或权利要求42所述的方法,其中包含所述靶核酸序列的所述双链靶多核苷酸是双链DNA,所述Cas蛋白缺乏切割所述双链DNA的能力,并且所述方法引起所述

双链靶多核苷酸的基因沉默。

46. 根据权利要求41或权利要求42所述的方法, 其中所述PAM序列包含5' -NNNNCNNA-3'、5' -CNNNCNN-3'、5' -NNNCCNN-3'、5' -NNCNCNN-3'、5' -NNNNCCN-3' 和/或5' -NCNNCNCN-3'。

47. 根据权利要求41或权利要求42所述的方法, 其中所述PAM序列包含5' -CCCCCCNA-3' [SEQ ID NO:10]。

48. 根据权利要求47所述的方法, 其中所述PAM序列包含5' -CCCCCCAA-3' [SEQ ID NO:11]。

49. 根据权利要求41或权利要求42所述的方法, 其中所述Cas蛋白从热脱氮地芽孢杆菌T12可获得。

50. 根据权利要求41或权利要求42所述的方法, 其中所述靶向RNA分子包含crRNA和tracrRNA。

51. 根据权利要求41或权利要求42所述的方法, 其中所述至少一种靶向RNA分子的长度是在35-200个核苷酸残基的范围中。

52. 根据权利要求41或权利要求42所述的方法, 其中所述靶核酸序列的长度是从15个至32个核苷酸残基。

53. 根据权利要求41或权利要求42所述的方法, 其中所述Cas蛋白作为蛋白复合体的一部分被提供, 所述蛋白复合体包含至少一种另外的蛋白, 所述至少一种另外的蛋白选自解旋酶、核酸酶、解旋酶-核酸酶、DNA甲基化酶、组蛋白甲基化酶、乙酰基转移酶、磷酸酶、激酶、转录(共)活化物、转录阻遏物、DNA结合蛋白、DNA结构蛋白、标志物蛋白、报告物蛋白、荧光蛋白、配体结合蛋白、信号肽、亚细胞定位序列、抗体表位或亲和纯化标签。

54. 根据权利要求53所述的方法, 其中所述Cas蛋白或另外的蛋白包含与所述Cas蛋白或蛋白复合体的N-末端和/或C-末端融合或连接的至少一种另外的蛋白。

55. 根据权利要求53所述的方法, 其中所述至少一种另外的蛋白是绿色荧光蛋白(GFP)。

56. 根据权利要求54所述的方法, 其中所述至少一种另外的蛋白是绿色荧光蛋白(GFP)。

57. 根据权利要求53所述的方法, 其中所述Cas蛋白的天然活性被失活并且所述Cas蛋白与标志物蛋白或包含核酸酶结构域的蛋白连接。

58. 根据权利要求54所述的方法, 其中所述Cas蛋白的天然活性被失活并且所述Cas蛋白与标志物蛋白或包含核酸酶结构域的蛋白连接。

59. 根据权利要求57或权利要求58所述的方法, 其中所述核酸酶结构域是FokI核酸酶结构域。

60. 根据权利要求53所述的方法, 其中所述双链靶多核苷酸是dsDNA, 并且所述另外的蛋白选自DNA修饰酶、转录活化物或转录阻遏物, 并且所述结合、裂解、标记或修饰引起基因表达的修饰。

61. 根据权利要求53所述的方法, 其中所述双链靶多核苷酸是dsDNA, 所述至少一种另外的蛋白是核酸酶或解旋酶-核酸酶, 并且所述修饰是在期望的基因座处的单链断裂或双链断裂。

62. 根据权利要求54所述的方法, 其中所述双链靶多核苷酸是dsDNA, 所述至少一种另

外的蛋白是核酸酶或解旋酶-核酸酶,并且所述修饰是在期望的基因座处的单链断裂或双链断裂。

63. 根据权利要求53所述的方法,其中所述双链靶多核苷酸是dsDNA,并且所述另外的蛋白选自DNA修饰酶、转录活化物或转录阻遏物,并且所述结合、裂解、标记或修饰引起基因表达的修饰。

64. 根据权利要求54所述的方法,其中所述双链靶多核苷酸是dsDNA,并且所述另外的蛋白选自DNA修饰酶、转录活化物或转录阻遏物,并且所述结合、裂解、标记或修饰引起基因表达的修饰。

65. 根据权利要求53所述的方法,其中所述结合、裂解、标记或修饰发生在细菌细胞内。

66. 根据权利要求54所述的方法,其中所述结合、裂解、标记或修饰发生在细菌细胞内。

67. 根据权利要求41所述的方法,其中所述结合、裂解、标记或修饰引起在期望的位置处修饰或缺失和/或插入期望的核苷酸序列,和/或其中所述结合、裂解、标记或修饰引起使在期望的基因座处的基因表达沉默。

68. 一种经转化的细胞,所述经转化的细胞具有包含靶核酸序列的双链靶多核苷酸,其中所述双链靶多核苷酸包含靶核酸链和非靶核酸链,所述靶核酸链包含所述靶核酸序列,所述非靶核酸链包含与所述靶核酸序列互补的前间区核酸序列,所述细胞包含:

成簇的规律地间隔的短回文重复 (CRISPR) 相关的 (Cas) 蛋白,所述成簇的规律地间隔的短回文重复 (CRISPR) 相关的 (Cas) 蛋白由SEQ ID NO:1的氨基酸序列组成;

至少一种靶向RNA分子,所述至少一种靶向RNA分子识别所述靶核酸链中的所述靶核酸序列,其中所述非靶核酸链还包含与所述前间区核酸序列的3'末端直接地相邻的前间区相邻基序 (PAM) 序列,其中所述PAM序列包含5'-NNNCNN-3';和

表达载体,所述表达载体包含编码所述Cas蛋白和所述靶向RNA分子中的至少一种的核酸,其中所述经转化的细胞不是人类胚胎干细胞。

69. 根据权利要求68所述的经转化的细胞,其中所述Cas蛋白和靶向RNA分子使得能够结合、裂解、标记或修饰所述细胞中的所述双链靶多核苷酸,并且所述结合、裂解、标记或修饰发生在20℃和100℃之间的温度。

70. 根据权利要求68或权利要求69所述的经转化的细胞,其中包含所述靶核酸序列的所述靶核酸链被所述Cas蛋白裂解。

71. 根据权利要求69所述的经转化的细胞,其中包含所述靶核酸序列的所述双链靶多核苷酸是双链DNA,并且所述结合、裂解、标记或修饰引起所述双链靶多核苷酸中的双链断裂。

72. 根据权利要求69所述的经转化的细胞,其中包含所述靶核酸序列的所述双链靶多核苷酸是双链DNA,所述Cas蛋白缺乏切割所述双链DNA的能力,并且所述结合、裂解、标记或修饰引起所述双链靶多核苷酸的基因沉默。

73. 根据权利要求68或权利要求69所述的经转化的细胞,其中所述PAM序列包含5'-NNNCNNA-3'、5'-CNNNCNN-3'、5'-NNCCNN-3'、5'-NCCNCNN-3'、5'-NNNNCCN-3'和/或5'-NCNNCNN-3'。

74. 根据权利要求68或权利要求69所述的经转化的细胞,其中所述PAM序列包含5'-CCCCCNNA-3' [SEQ ID NO:10]。

75. 根据权利要求74所述的经转化的细胞,其中所述PAM序列包含5'-CCCCCAA-3' [SEQ ID NO:11]。

76. 根据权利要求68或权利要求69所述的经转化的细胞,其中所述Cas蛋白从热脱氮地芽孢杆菌T12可获得。

77. 根据权利要求68或权利要求69所述的经转化的细胞,其中所述细胞是原核细胞。

78. 根据权利要求68或权利要求69所述的经转化的细胞,其中所述细胞是真核细胞。

79. 根据权利要求68或权利要求69所述的经转化的细胞,其中所述靶向RNA分子包含crRNA和tracrRNA。

80. 根据权利要求68或权利要求69所述的经转化的细胞,其中所述至少一种靶向RNA分子的长度是在35-200个核苷酸残基的范围中。

81. 根据权利要求68或权利要求69所述的经转化的细胞,其中所述靶核酸序列的长度是从15个至32个核苷酸残基。

82. 根据权利要求69所述的经转化的细胞,其中所述Cas蛋白作为蛋白复合体的一部分被提供,所述蛋白复合体包含至少一种另外的蛋白,所述至少一种另外的蛋白选自解旋酶、核酸酶、解旋酶-核酸酶、DNA甲基化酶、组蛋白甲基化酶、乙酰基转移酶、磷酸酶、激酶、转录(共)活化物、转录阻遏物、DNA结合蛋白、DNA结构蛋白、标志物蛋白、报告物蛋白、荧光蛋白、配体结合蛋白、信号肽、亚细胞定位序列、抗体表位或亲和纯化标签。

83. 根据权利要求82所述的经转化的细胞,其中所述Cas蛋白或另外的蛋白包含与所述Cas蛋白或蛋白复合体的N-末端和/或C-末端融合或连接的至少一种另外的蛋白。

84. 根据权利要求82或权利要求83所述的经转化的细胞,其中所述至少一种另外的蛋白是绿色荧光蛋白(GFP)。

85. 根据权利要求82或权利要求83所述的经转化的细胞,其中所述Cas蛋白的天然活性被失活并且所述Cas蛋白与标志物蛋白或包含核酸酶结构域的蛋白连接。

86. 根据权利要求85所述的经转化的细胞,其中所述核酸酶结构域是FokI核酸酶结构域。

87. 根据权利要求82或权利要求83所述的经转化的细胞,其中所述双链靶多核苷酸是dsDNA,所述至少一种另外的蛋白是核酸酶或解旋酶-核酸酶,并且所述修饰是在期望的基因座处的单链断裂或双链断裂。

88. 根据权利要求82所述的经转化的细胞,其中所述双链靶多核苷酸是dsDNA,并且所述另外的蛋白选自DNA修饰酶、转录活化物或转录阻遏物,并且所述结合、裂解、标记或修饰引起基因表达的修饰。

89. 根据权利要求82或权利要求83所述的经转化的细胞,其中所述Cas蛋白从表达载体表达。

90. 根据权利要求69所述的经转化的细胞,其中所述结合、裂解、标记或修饰引起在期望的位置处修饰或缺失和/或插入期望的核苷酸序列,和/或其中所述结合、裂解、标记或修饰引起使在期望的基因座处的基因表达沉默。

91. 一种核蛋白复合体,所述核蛋白复合体包含Cas蛋白、识别双链靶多核苷酸中的靶核酸序列的至少一种靶向RNA分子和所述双链靶多核苷酸,其中

所述Cas蛋白由SEQ ID NO:1的氨基酸序列组成;

所述双链靶多核苷酸包含靶核酸链和非靶核酸链,所述靶核酸链包含所述靶核酸序列,所述非靶核酸链包含与所述靶核酸序列互补的前间区核酸序列和与所述前间区核酸序列的3'末端直接地相邻的前间区相邻基序(PAM)序列,其中所述PAM序列包含5'-NNNNCNN-3'。

92.根据权利要求91所述的核蛋白复合体,其中所述核蛋白复合体在20℃和100℃之间的温度存在。

93.根据权利要求91或权利要求92所述的核蛋白复合体,其中包含所述靶核酸序列的所述双链靶多核苷酸被所述Cas蛋白裂解。

94.根据权利要求91或权利要求92所述的核蛋白复合体,其中包含所述靶核酸序列的所述双链靶多核苷酸是双链DNA,并且结合、裂解、标记或修饰引起所述双链靶多核苷酸中的双链断裂。

95.根据权利要求91或权利要求92所述的核蛋白复合体,其中包含所述靶核酸序列的所述双链靶多核苷酸是双链DNA,所述Cas蛋白缺乏切割所述双链DNA的能力,并且所述核蛋白复合体的存在引起所述双链靶多核苷酸的基因沉默。

96.根据权利要求91或权利要求92所述的核蛋白复合体,其中所述PAM序列包含5'-NNNNCNNA-3'、5'-CNNNCNN-3'、5'-NNNCCNN-3'、5'-NNCNCNN-3'、5'-NNNNCCN-3'和/或5'-NCNNCNCNN-3'。

97.根据权利要求91或权利要求92所述的核蛋白复合体,其中所述PAM序列包含5'-CCCCCNA-3' [SEQ ID NO:10]。

98.根据权利要求97所述的核蛋白复合体,其中所述PAM序列包含5'-CCCCCAA-3' [SEQ ID NO:11]。

99.根据权利要求91或权利要求92所述的核蛋白复合体,其中所述Cas蛋白从热脱氮地芽孢杆菌T12可获得。

100.根据权利要求91或权利要求92所述的核蛋白复合体,其中所述核蛋白复合体是在原核细胞中。

101.根据权利要求91或权利要求92所述的核蛋白复合体,其中所述核蛋白复合体是在真核细胞中。

102.根据权利要求91或权利要求92所述的核蛋白复合体,其中所述靶向RNA分子包含crRNA和tracrRNA。

103.根据权利要求91或权利要求92所述的核蛋白复合体,其中所述至少一种靶向RNA分子的长度是在35-200个核苷酸残基的范围中。

104.根据权利要求91或权利要求92所述的核蛋白复合体,其中所述靶核酸序列的长度是从15个至32个核苷酸残基。

105.根据权利要求91或权利要求92所述的核蛋白复合体,其中所述Cas蛋白作为蛋白复合体的一部分被提供,所述蛋白复合体包含至少一种另外的蛋白,所述至少一种另外的蛋白选自解旋酶、核酸酶、解旋酶-核酸酶、DNA甲基化酶、组蛋白甲基化酶、乙酰基转移酶、磷酸酶、激酶、转录(共)活化物、转录阻遏物、DNA结合蛋白、DNA结构蛋白、标志物蛋白、报告物蛋白、荧光蛋白、配体结合蛋白、信号肽、亚细胞定位序列、抗体表位或亲和纯化标签。

106.根据权利要求105所述的核蛋白复合体,其中所述Cas蛋白或另外的蛋白包含与所

述Cas蛋白或蛋白复合体的N-末端和/或C-末端融合或连接的至少一种另外的蛋白。

107. 根据权利要求105所述的核蛋白复合体, 其中所述至少一种另外的蛋白是绿色荧光蛋白(GFP)。

108. 根据权利要求105所述的核蛋白复合体, 其中所述Cas蛋白的天然活性被失活并且所述Cas蛋白与标志物蛋白或包含核酸酶结构域的蛋白连接。

109. 根据权利要求107所述的核蛋白复合体, 其中所述Cas蛋白的天然活性被失活并且所述Cas蛋白与标志物蛋白或包含核酸酶结构域的蛋白连接。

110. 根据权利要求108或权利要求109所述的核蛋白复合体, 其中所述核酸酶结构域是FokI核酸酶结构域。

111. 根据权利要求105所述的核蛋白复合体, 其中所述核酸是dsDNA, 所述至少一种另外的蛋白是核酸酶或解旋酶-核酸酶, 并且所述双链靶多核苷酸具有在期望的基因座处的单链断裂或双链断裂。

112. 根据权利要求106所述的核蛋白复合体, 其中所述核酸是dsDNA, 所述至少一种另外的蛋白是核酸酶或解旋酶-核酸酶, 并且所述双链靶多核苷酸具有在期望的基因座处的单链断裂或双链断裂。

113. 根据权利要求105所述的核蛋白复合体, 其中所述核酸是dsDNA, 并且所述另外的蛋白选自DNA修饰酶、转录活化物或转录阻遏物, 并且所述核蛋白复合体的形成引起基因表达的修饰。

114. 根据权利要求106所述的核蛋白复合体, 其中所述核酸是dsDNA, 并且所述另外的蛋白选自DNA修饰酶、转录活化物或转录阻遏物, 并且所述核蛋白复合体的形成引起基因表达的修饰。

115. 根据权利要求91或权利要求92所述的核蛋白复合体, 其中所述核蛋白的形成引起在期望的位置处修饰或缺失和/或插入期望的核苷酸序列, 和/或其中所述核蛋白复合体的形成引起使在期望的基因座处的基因表达沉默。

热稳定的Cas9核酸酶

发明领域

[0001] 本发明涉及遗传工程 (genetic engineering) 领域, 并且更特别地涉及核酸编辑和基因组修饰。本发明涉及呈核酸酶的形式遗传工程工具, 所述核酸酶可以被配置用于遗传物质的序列指导的位点特异性结合、切口、切割和修饰; 还涉及对遗传物质的序列特异性位点发挥活性 (特别地, 核酸酶活性) 的核糖核蛋白、和用于作为标志物使用的修饰的核酸酶和核糖核蛋白。因此, 本发明还涉及用于核酸酶和指导RNA在细胞中的递送和表达的相关的表达构建体。此外, 本发明涉及体外或体内的核酸的序列特异性编辑和被用于实现所述编辑的方法。本发明涉及的特定领域是嗜热生物体 (特别地, 微生物) 的遗传操作。

[0002] 发明背景

[0003] 在2007年首次证明CRISPR-Cas是在许多细菌和大多数古核生物 (archaea) 中的适应性免疫系统 (Barrangou等人, 2007, Science 315:1709-1712, Brouns等人, 2008, Science 321:960-964)。基于功能和结构标准, 迄今已经表征了三种类型的CRISPR-Cas系统, 大多数CRISPR-Cas系统使用小RNA分子作为指导 (guide) 以靶向互补的DNA序列 (Makarova等人, 2011, Nat Rev Microbiol 9:467-477; Van der Oost等人, 2014, Nat Rev Microbiol 12:479-492)。

[0004] 在Doudna/Charpentier实验室的最近的一项研究中, 对II型CRISPR-Cas系统的效应酶 (Cas9) 的进行了彻底的表征, 包括证明设计的CRISPR RNA指导 (具有特定间隔区序列) 的引入靶向了质粒上的互补序列 (前间区 (protospacers)), 引起该质粒的双链断裂 (Jinek等人, 2012, Science 337:816-821)。继Jinek等人, 2012之后, Cas9被用作基因组编辑的工具。

[0005] Cas9已经被用于工程化一系列真核细胞 (例如鱼、植物、人类) 的基因组 (Charpentier和Doudna, 2013, Nature 495:50-51)。

[0006] 另外, 通过选择专门的重组事件, Cas9已经被用于改进细菌中的同源重组的产率 (Jiang等人, 2013, Nature Biotechnol 31:233-239)。为了实现这一点, 将毒性片段 (靶向构建体) 与携带期望的改变的救援片段 (编辑构建体, 携带点突变或缺失) 共转染。靶向构建体由Cas9与设计的CRISPR的组合和抗生素抗性标志物组成, 定义了宿主染色体上的期望重组的位点; 在对应的抗生素的存在下, 选择靶向构建体在宿主染色体中的整合。仅当编辑构建体与宿主染色体上的CRISPR靶位点发生另外的重组时, 宿主才可以逃脱自身免疫问题。因此, 在抗生素的存在下, 仅期望的 (无标志物的) 突变体能够存活和生长。还呈现了选择用于后续从染色体去除整合的靶向构建体的相关的策略, 生成真正的无标志物突变体。

[0007] 在最近几年中, 已经建立了CRISPR-Cas介导的基因组编辑构成用于遗传工程的有用的工具。已经建立了原核CRISPR系统作为适应性免疫系统为其宿主服务 (Jinek等人, 2012, Science 337:816-821), 并且可以被用于快速和有效的遗传工程 (例如, Mali等人, 2013, Nat Methods 10:957-963), 仅需要指导序列的修饰以靶向感兴趣的序列。

[0008] 然而, 对开发应用于遗传研究和基因组编辑领域的在多种实验条件下具有改进的序列特异性核酸检测、裂解和操作的剂存在持续需求。特别地, 当前可得的序列特异性基因

组编辑工具,包括Cas9,不适用于在所有条件或生物体中使用,例如,序列特异性核酸酶是相对地热敏感的,并且因此不适用于在严格嗜热微生物(其能够在41℃与122℃之间生长,并且能够在从>45℃至80℃的温度范围中最佳地生长,其中极端嗜热菌(hyperthermophiles)能够在80℃以上最佳生长),例如,在工业发酵中使用的微生物或用于在升高的温度进行的体外实验室过程的微生物中使用。

[0009] 迄今,不存在关于在嗜热菌(thermophiles)中的活性Cas9蛋白的实验证据。基于由Chylinski等人(2014;Nucleic Acids Research 42:6091-6105)针对细菌中Cas9的存在的比较基因组筛选,发现II-C型CRISPR-Cas系统仅存在于所有细菌基因组的约3.3%中。在嗜热细菌中,基于统计学分析,II型系统的代表性不足(P=0.0019)。另外,然而,在古核生物中尚未发现II型系统,这可能是由于古核生物中不存在RNA酶III蛋白(参与II型系统)。Chylinski等人(2014;Nucleic Acids Research 42:6091-6105)确实描述了II型CRISPR-Cas系统的分类和演化,特别地,鉴定出了表现出这些系统的两个物种,然而这些物种在55℃最大地生长,并且没有表现出严格嗜热生长,其中最佳生长温度为60℃-80℃,而极端嗜热菌能够在80℃以上最佳地生长。

[0010] 尽管CRISPR-Cas系统在细菌基因组中的稀有性,并且特别是仅在具有45℃以下的最佳生长温度的细菌(非古核生物)中已经发现Cas9的事实,本发明人已经出乎意料地发现了几种能够在升高的温度进行基因组编辑的热稳定的Cas9变体。本发明人还已经发现了优化的前间区相邻基序(protospacer adjacent motif;PAM)序列,该优化的前间区相邻基序(PAM)序列与热稳定的Cas9变体一起起作用,以使基因组编辑能够在宽范围的温度中,包括在升高的温度进行。这些Cas9核酸酶和通过相关的PAM序列的知识设计的RNA分子,提供了用于在升高的温度进行遗传工程的新型工具,并且在嗜热生物体(特别地,微生物)的遗传操作中具有特别的价值。

[0011] 发明概述

[0012] 因此,本发明提供了分离的成簇的规律地间隔的短回文重复(clustered regularly interspaced short palindromic repeat;CRISPR)相关(Cas)蛋白或多肽,所述分离的成簇的规律地间隔的短回文重复(CRISPR)相关(Cas)蛋白或多肽包含:

[0013] a.氨基酸基序EKDGKYYC[SEQ ID NO:2];和/或

[0014] b.氨基酸基序 $X_1X_2CTX_3X_4$ [SEQ ID NO:3],其中 X_1 独立地选自异亮氨酸、甲硫氨酸或脯氨酸, X_2 独立地选自缬氨酸、丝氨酸、天冬酰胺或异亮氨酸, X_3 独立地选自谷氨酸或赖氨酸,并且 X_4 是丙氨酸、谷氨酸或精氨酸之一;和/或

[0015] c.氨基酸基序 X_5LXX_6IE [SEQ ID NO:4],其中 X_5 独立地选自甲硫氨酸或苯丙氨酸,并且 X_6 独立地选自组氨酸或天冬酰胺;和/或

[0016] d.氨基酸基序 X_7VYSX_8K [SEQ ID NO:5],其中 X_7 是谷氨酸或异亮氨酸,并且 X_8 是色氨酸、丝氨酸或赖氨酸之一;和/或

[0017] e.氨基酸基序 $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO:6],其中 X_9 是丙氨酸或谷氨酸, X_{10} 是谷氨酰胺或赖氨酸, X_{11} 是精氨酸或丙氨酸, X_{12} 是天冬酰胺或丙氨酸,并且 X_{13} 是赖氨酸或丝氨酸。

[0018] 在本发明的上下文中的多肽可以被视作全长Cas蛋白的片段。这类片段可以是失活的并且以与结合、编辑和/或切割遗传物质不直接地相关的方式和目的使用,例如用于测

定中的标准物或产生抗体等。

[0019] 然而,在优选的实施方案中,Cas蛋白或多肽是有功能的并且当与至少一种靶向RNA分子和包含被靶向RNA分子识别的靶核酸序列的多核苷酸缔合时能够在20℃与100℃的范围中(包括端点)的温度进行裂解、结合、标记或修饰。优选地,Cas蛋白或多肽是有功能的,并且能够在50℃和70℃的范围中的温度,例如55℃或60℃的温度进行所述裂解、结合、标记或修饰。

[0020] 在特定实施方案中,本发明可以提供包含氨基酸基序EKDGKYYC[SEQ ID NO:2]的Cas蛋白或多肽。在其他实施方案中,Cas蛋白或多肽还可以包含氨基酸基序 $X_1X_2CTX_3X_4$ [SEQ ID NO:3],其中 X_1 独立地选自异亮氨酸、甲硫氨酸或脯氨酸, X_2 独立地选自缬氨酸、丝氨酸、天冬酰胺或异亮氨酸, X_3 独立地选自谷氨酸或赖氨酸,并且 X_4 是丙氨酸、谷氨酸或精氨酸之一。

[0021] 在其他实施方案中,本文定义的Cas蛋白或多肽还可以另外地包含氨基酸基序 X_5LXX_6IE [SEQ ID NO:4],其中 X_5 独立地选自甲硫氨酸或苯丙氨酸,并且 X_6 独立地选自组氨酸或天冬酰胺。

[0022] 在其他实施方案中,本文定义的Cas蛋白或多肽还可以另外地包含氨基酸基序 X_7VYSX_8K [SEQ ID NO:5],其中 X_7 是谷氨酸或异亮氨酸,并且 X_8 是色氨酸、丝氨酸或赖氨酸之一。

[0023] 在其他实施方案中,本文定义的Cas蛋白或多肽还可以另外地包含氨基酸基序 $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO:6],其中 X_9 是丙氨酸或谷氨酸, X_{10} 是谷氨酰胺或赖氨酸, X_{11} 是精氨酸或丙氨酸, X_{12} 是天冬酰胺或丙氨酸,并且 X_{13} 是赖氨酸或丝氨酸。

[0024] 根据本发明,可以理解的是,本发明的Cas蛋白或多肽可以包含单独的或以组合方式的SEQ ID NO:2至6的任何基序。以下总结了可以表征本发明的Cas蛋白或多肽的基序的每个组合:

[0025] EKDGKYYC[SEQ ID NO:2]。

[0026] EKDGKYYC[SEQ ID NO:2];和 $X_1X_2CTX_3X_4$ [SEQ ID NO:3],其中 X_1 独立地选自异亮氨酸、甲硫氨酸或脯氨酸, X_2 独立地选自缬氨酸、丝氨酸、天冬酰胺或异亮氨酸, X_3 独立地选自谷氨酸或赖氨酸,并且 X_4 是丙氨酸、谷氨酸或精氨酸之一。

[0027] EKDGKYYC[SEQ ID NO:2];和 $X_1X_2CTX_3X_4$ [SEQ ID NO:3],其中 X_1 独立地选自异亮氨酸、甲硫氨酸或脯氨酸, X_2 独立地选自缬氨酸、丝氨酸、天冬酰胺或异亮氨酸, X_3 独立地选自谷氨酸或赖氨酸,并且 X_4 是丙氨酸、谷氨酸或精氨酸之一;和 X_5LXX_6IE [SEQ ID NO:4],其中 X_5 独立地选自甲硫氨酸或苯丙氨酸,并且 X_6 独立地选自组氨酸或天冬酰胺。

[0028] EKDGKYYC[SEQ ID NO:2];和 $X_1X_2CTX_3X_4$ [SEQ ID NO:3],其中 X_1 独立地选自异亮氨酸、甲硫氨酸或脯氨酸, X_2 独立地选自缬氨酸、丝氨酸、天冬酰胺或异亮氨酸, X_3 独立地选自谷氨酸或赖氨酸,并且 X_4 是丙氨酸、谷氨酸或精氨酸之一;和 X_5LXX_6IE [SEQ ID NO:4],其中 X_5 独立地选自甲硫氨酸或苯丙氨酸,并且 X_6 独立地选自组氨酸或天冬酰胺;和 X_7VYSX_8K [SEQ ID NO:5],其中 X_7 是谷氨酸或异亮氨酸,并且 X_8 是色氨酸、丝氨酸或赖氨酸之一。

[0029] EKDGKYYC[SEQ ID NO:2];和 $X_1X_2CTX_3X_4$ [SEQ ID NO:3],其中 X_1 独立地选自异亮氨酸、甲硫氨酸或脯氨酸, X_2 独立地选自缬氨酸、丝氨酸、天冬酰胺或异亮氨酸, X_3 独立地选自谷氨酸或赖氨酸,并且 X_4 是丙氨酸、谷氨酸或精氨酸之一;和 X_5LXX_6IE [SEQ ID NO:4],其中

X_5 独立地选自甲硫氨酸或苯丙氨酸,并且 X_6 独立地选自组氨酸或天冬酰胺;和 X_7 VYSX₈K[SEQ ID NO:5],其中 X_7 是谷氨酸或异亮氨酸,并且 X_8 是色氨酸、丝氨酸或赖氨酸之一;和 X_9 FYX₁₀X₁₁REQX₁₂KEX₁₃[SEQ ID NO:6],其中 X_9 是丙氨酸或谷氨酸, X_{10} 是谷氨酰胺或赖氨酸, X_{11} 是精氨酸或丙氨酸, X_{12} 是天冬酰胺或丙氨酸,并且 X_{13} 是赖氨酸或丝氨酸。

[0030] EKDGKYYC[SEQ ID NO:2];和 X_1 X₂CTX₃X₄[SEQ ID NO:3],其中 X_1 独立地选自异亮氨酸、甲硫氨酸或脯氨酸, X_2 独立地选自缬氨酸、丝氨酸、天冬酰胺或异亮氨酸, X_3 独立地选自谷氨酸或赖氨酸,并且 X_4 是丙氨酸、谷氨酸或精氨酸之一;和 X_5 LKX₆IE[SEQ ID NO:4],其中 X_5 独立地选自甲硫氨酸或苯丙氨酸,并且 X_6 独立地选自组氨酸或天冬酰胺;和 X_9 FYX₁₀X₁₁REQX₁₂KEX₁₃[SEQ ID NO:6],其中 X_9 是丙氨酸或谷氨酸, X_{10} 是谷氨酰胺或赖氨酸, X_{11} 是精氨酸或丙氨酸, X_{12} 是天冬酰胺或丙氨酸,并且 X_{13} 是赖氨酸或丝氨酸。

[0031] EKDGKYYC[SEQ ID NO:2];和 X_1 X₂CTX₃X₄[SEQ ID NO:3],其中 X_1 独立地选自异亮氨酸、甲硫氨酸或脯氨酸, X_2 独立地选自缬氨酸、丝氨酸、天冬酰胺或异亮氨酸, X_3 独立地选自谷氨酸或赖氨酸,并且 X_4 是丙氨酸、谷氨酸或精氨酸之一;和 X_7 VYSX₈K[SEQ ID NO:5],其中 X_7 是谷氨酸或异亮氨酸,并且 X_8 是色氨酸、丝氨酸或赖氨酸之一;和 X_9 FYX₁₀X₁₁REQX₁₂KEX₁₃[SEQ ID NO:6],其中 X_9 是丙氨酸或谷氨酸, X_{10} 是谷氨酰胺或赖氨酸, X_{11} 是精氨酸或丙氨酸, X_{12} 是天冬酰胺或丙氨酸,并且 X_{13} 是赖氨酸或丝氨酸。

[0032] EKDGKYYC[SEQ ID NO:2];和 X_5 LKX₆IE[SEQ ID NO:4],其中 X_5 独立地选自甲硫氨酸或苯丙氨酸,并且 X_6 独立地选自组氨酸或天冬酰胺;和 X_7 VYSX₈K[SEQ ID NO:5],其中 X_7 是谷氨酸或异亮氨酸,并且 X_8 是色氨酸、丝氨酸或赖氨酸之一;和 X_9 FYX₁₀X₁₁REQX₁₂KEX₁₃[SEQ ID NO:6],其中 X_9 是丙氨酸或谷氨酸, X_{10} 是谷氨酰胺或赖氨酸, X_{11} 是精氨酸或丙氨酸, X_{12} 是天冬酰胺或丙氨酸,并且 X_{13} 是赖氨酸或丝氨酸。

[0033] EKDGKYYC[SEQ ID NO:2];和 X_5 LKX₆IE[SEQ ID NO:4],其中 X_5 独立地选自甲硫氨酸或苯丙氨酸,并且 X_6 独立地选自组氨酸或天冬酰胺。

[0034] EKDGKYYC[SEQ ID NO:2];和 X_7 VYSX₈K[SEQ ID NO:5],其中 X_7 是谷氨酸或异亮氨酸,并且 X_8 是色氨酸、丝氨酸或赖氨酸之一。

[0035] EKDGKYYC[SEQ ID NO:2];和 X_9 FYX₁₀X₁₁REQX₁₂KEX₁₃[SEQ ID NO:6],其中 X_9 是丙氨酸或谷氨酸, X_{10} 是谷氨酰胺或赖氨酸, X_{11} 是精氨酸或丙氨酸, X_{12} 是天冬酰胺或丙氨酸,并且 X_{13} 是赖氨酸或丝氨酸。

[0036] EKDGKYYC[SEQ ID NO:2];和 X_5 LKX₆IE[SEQ ID NO:4],其中 X_5 独立地选自甲硫氨酸或苯丙氨酸,并且 X_6 独立地选自组氨酸或天冬酰胺;和 X_7 VYSX₈K[SEQ ID NO:5],其中 X_7 是谷氨酸或异亮氨酸,并且 X_8 是色氨酸、丝氨酸或赖氨酸之一。

[0037] EKDGKYYC[SEQ ID NO:2];和 X_5 LKX₆IE[SEQ ID NO:4],其中 X_5 独立地选自甲硫氨酸或苯丙氨酸,并且 X_6 独立地选自组氨酸或天冬酰胺;和 X_9 FYX₁₀X₁₁REQX₁₂KEX₁₃[SEQ ID NO:6],其中 X_9 是丙氨酸或谷氨酸, X_{10} 是谷氨酰胺或赖氨酸, X_{11} 是精氨酸或丙氨酸, X_{12} 是天冬酰胺或丙氨酸,并且 X_{13} 是赖氨酸或丝氨酸。

[0038] EKDGKYYC[SEQ ID NO:2];和 X_7 VYSX₈K[SEQ ID NO:5],其中 X_7 是谷氨酸或异亮氨酸,并且 X_8 是色氨酸、丝氨酸或赖氨酸之一;和 X_9 FYX₁₀X₁₁REQX₁₂KEX₁₃[SEQ ID NO:6],其中 X_9 是丙氨酸或谷氨酸, X_{10} 是谷氨酰胺或赖氨酸, X_{11} 是精氨酸或丙氨酸, X_{12} 是天冬酰胺或丙氨酸,并且 X_{13} 是赖氨酸或丝氨酸。

[0039] $X_1X_2CTX_3X_4$ [SEQ ID NO:3], 其中 X_1 独立地选自异亮氨酸、甲硫氨酸或脯氨酸, X_2 独立地选自缬氨酸、丝氨酸、天冬酰胺或异亮氨酸, X_3 独立地选自谷氨酸或赖氨酸,并且 X_4 是丙氨酸、谷氨酸或精氨酸之一;和 X_5LX_6IE [SEQ ID NO:4], 其中 X_5 独立地选自甲硫氨酸或苯丙氨酸,并且 X_6 独立地选自组氨酸或天冬酰胺。

[0040] $X_1X_2CTX_3X_4$ [SEQ ID NO:3], 其中 X_1 独立地选自异亮氨酸、甲硫氨酸或脯氨酸, X_2 独立地选自缬氨酸、丝氨酸、天冬酰胺或异亮氨酸, X_3 独立地选自谷氨酸或赖氨酸,并且 X_4 是丙氨酸、谷氨酸或精氨酸之一;和 X_5LX_6IE [SEQ ID NO:4], 其中 X_5 独立地选自甲硫氨酸或苯丙氨酸,并且 X_6 独立地选自组氨酸或天冬酰胺;和 X_7VYSX_8K [SEQ ID NO:5], 其中 X_7 是谷氨酸或异亮氨酸,并且 X_8 是色氨酸、丝氨酸或赖氨酸之一。

[0041] $X_1X_2CTX_3X_4$ [SEQ ID NO:3], 其中 X_1 独立地选自异亮氨酸、甲硫氨酸或脯氨酸, X_2 独立地选自缬氨酸、丝氨酸、天冬酰胺或异亮氨酸, X_3 独立地选自谷氨酸或赖氨酸,并且 X_4 是丙氨酸、谷氨酸或精氨酸之一;和 X_5LX_6IE [SEQ ID NO:4], 其中 X_5 独立地选自甲硫氨酸或苯丙氨酸,并且 X_6 独立地选自组氨酸或天冬酰胺;和 X_7VYSX_8K [SEQ ID NO:5], 其中 X_7 是谷氨酸或异亮氨酸,并且 X_8 是色氨酸、丝氨酸或赖氨酸之一;和 $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO:6], 其中 X_9 是丙氨酸或谷氨酸, X_{10} 是谷氨酰胺或赖氨酸, X_{11} 是精氨酸或丙氨酸, X_{12} 是天冬酰胺或丙氨酸,并且 X_{13} 是赖氨酸或丝氨酸。

[0042] $X_1X_2CTX_3X_4$ [SEQ ID NO:3], 其中 X_1 独立地选自异亮氨酸、甲硫氨酸或脯氨酸, X_2 独立地选自缬氨酸、丝氨酸、天冬酰胺或异亮氨酸, X_3 独立地选自谷氨酸或赖氨酸,并且 X_4 是丙氨酸、谷氨酸或精氨酸之一;和 X_7VYSX_8K [SEQ ID NO:5], 其中 X_7 是谷氨酸或异亮氨酸,并且 X_8 是色氨酸、丝氨酸或赖氨酸之一;和 $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO:6], 其中 X_9 是丙氨酸或谷氨酸, X_{10} 是谷氨酰胺或赖氨酸, X_{11} 是精氨酸或丙氨酸, X_{12} 是天冬酰胺或丙氨酸,并且 X_{13} 是赖氨酸或丝氨酸。

[0043] $X_1X_2CTX_3X_4$ [SEQ ID NO:3], 其中 X_1 独立地选自异亮氨酸、甲硫氨酸或脯氨酸, X_2 独立地选自缬氨酸、丝氨酸、天冬酰胺或异亮氨酸, X_3 独立地选自谷氨酸或赖氨酸,并且 X_4 是丙氨酸、谷氨酸或精氨酸之一;和 X_7VYSX_8K [SEQ ID NO:5], 其中 X_7 是谷氨酸或异亮氨酸,并且 X_8 是色氨酸、丝氨酸或赖氨酸之一。

[0044] $X_1X_2CTX_3X_4$ [SEQ ID NO:3], 其中 X_1 独立地选自异亮氨酸、甲硫氨酸或脯氨酸, X_2 独立地选自缬氨酸、丝氨酸、天冬酰胺或异亮氨酸, X_3 独立地选自谷氨酸或赖氨酸,并且 X_4 是丙氨酸、谷氨酸或精氨酸之一;和 $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO:6], 其中 X_9 是丙氨酸或谷氨酸, X_{10} 是谷氨酰胺或赖氨酸, X_{11} 是精氨酸或丙氨酸, X_{12} 是天冬酰胺或丙氨酸,并且 X_{13} 是赖氨酸或丝氨酸。

[0045] X_5LX_6IE [SEQ ID NO:4], 其中 X_5 独立地选自甲硫氨酸或苯丙氨酸,并且 X_6 独立地选自组氨酸或天冬酰胺;和 X_7VYSX_8K [SEQ ID NO:5], 其中 X_7 是谷氨酸或异亮氨酸,并且 X_8 是色氨酸、丝氨酸或赖氨酸之一;和 $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO:6], 其中 X_9 是丙氨酸或谷氨酸, X_{10} 是谷氨酰胺或赖氨酸, X_{11} 是精氨酸或丙氨酸, X_{12} 是天冬酰胺或丙氨酸,并且 X_{13} 是赖氨酸或丝氨酸。

[0046] X_5LX_6IE [SEQ ID NO:4], 其中 X_5 独立地选自甲硫氨酸或苯丙氨酸,并且 X_6 独立地选自组氨酸或天冬酰胺;和 X_7VYSX_8K [SEQ ID NO:5], 其中 X_7 是谷氨酸或异亮氨酸,并且 X_8 是色氨酸、丝氨酸或赖氨酸之一。

[0047] $X_5LX_6IE[SEQ ID NO:4]$, 其中 X_5 独立地选自甲硫氨酸或苯丙氨酸, 并且 X_6 独立地选自组氨酸或天冬酰胺; 和 $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}[SEQ ID NO:6]$, 其中 X_9 是丙氨酸或谷氨酸, X_{10} 是谷氨酰胺或赖氨酸, X_{11} 是精氨酸或丙氨酸, X_{12} 是天冬酰胺或丙氨酸, 并且 X_{13} 是赖氨酸或丝氨酸。

[0048] $X_7VYSX_8K[SEQ ID NO:5]$, 其中 X_7 是谷氨酸或异亮氨酸, 并且 X_8 是色氨酸、丝氨酸或赖氨酸之一; 和 $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}[SEQ ID NO:6]$, 其中 X_9 是丙氨酸或谷氨酸, X_{10} 是谷氨酰胺或赖氨酸, X_{11} 是精氨酸或丙氨酸, X_{12} 是天冬酰胺或丙氨酸, 并且 X_{13} 是赖氨酸或丝氨酸。

[0049] 在另一个方面中, 本发明提供了分离的Cas蛋白或其多肽片段, 所述分离的Cas蛋白或其多肽片段具有SEQ ID NO:1的氨基酸序列或与SEQ ID NO:1具有至少77%同一性的序列。优选地, Cas蛋白或多肽能够在20°C和100°C的范围中(包括端点)的温度进行结合、裂解、标记或修饰。优选地, Cas蛋白或多肽能够在50°C和70°C之间的范围中的温度, 例如55°C或60°C的温度进行所述裂解、结合、标记或修饰。优选地, Cas蛋白或多肽能够在30°C和80°C之间的范围中的温度、在37°C和78°C之间的温度, 优选地在55°C以上的温度; 更优选地在55°C和80°C之间的温度; 甚至更优选地在55°C和65°C之间的温度或在60°C和65°C之间的温度进行所述裂解、结合、标记或修饰。

[0050] 本发明还提供了本文提供的靶向RNA分子和Cas蛋白或多肽用于结合、裂解、标记或修饰包含靶核酸序列的靶多核苷酸的用途。靶向RNA分子识别多核苷酸的靶核酸链上的靶核酸序列。

[0051] 包含靶核酸序列的靶多核苷酸可以是双链的, 并且因此包含靶核酸链和非靶核酸链, 该靶核酸链包含所述靶核酸序列, 该非靶核酸链包含前间区核酸序列。前间区核酸序列与靶核酸序列基本上互补, 并且在双链靶多核苷酸中与其配对。非靶核酸链还可以包含与前间区序列的3'末端直接地相邻的前间区相邻基序(PAM)序列。PAM序列的长度可以是至少6个、7个或8个核苷酸。优选地, PAM序列在第五个位置中具有胞嘧啶。优选地, PAM序列包含序列5'-NNNNC-3', 使得PAM序列从5'-末端以5'-NNNNC-3'开始。另外地或可选地, PAM序列可以在第八个位置中具有腺嘌呤, 使得PAM序列包含序列5'-NNNNNNA-3', 并且PAM序列从5'-末端以5'-NNNNNNA-3'开始。另外地或可选地, PAM序列可以在第一个、第二个、第三个、第四个和第六个位置的一个或多个位置中具有胞嘧啶, 使得PAM序列从5'-末端以5'-CNNNN-3'、5'-NCNNN-3'、5'-NNCNN-3'、5'-NNNCN-3'、和/或5'-NNNNNC-3'开始。优选地, PAM序列包含5'-CCCCCNA-3' [SEQ ID NO:10], 使得PAM序列从5'-末端以5'-CCCCCNA-3' [SEQ ID NO:10]开始, 并且进一步优选地, PAM序列包含5'-CCCCCAA-3' [SEQ ID NO:11], 使得PAM序列从5'-末端以5'-CCCCCAA-3' [SEQ ID NO:11]开始。其他优选的PAM序列包括5'-ATCCCCAA-3' [SEQ ID NO:21]和5'-ACGGCCAA-3' [SEQ ID NO:22]。

[0052] 优选地, Cas蛋白或多肽能够在40°C至80°C的范围中(包括端点)的温度, 优选地在45°C至80°C的范围中(包括端点)的温度, 并且进一步优选地在50°C至80°C的范围中(包括端点)的温度进行结合、裂解、标记或修饰。例如, 结合、裂解、标记或修饰发生在45°C、46°C、47°C、48°C、49°C、50°C、51°C、52°C、53°C、54°C、55°C、56°C、57°C、58°C、59°C、60°C、61°C、62°C、63°C、64°C、65°C、66°C、67°C、68°C、69°C、70°C、71°C、72°C、73°C、74°C、75°C、76°C、77°C、78°C、79°C或80°C的温度。更优选地, Cas蛋白或多肽能够在55°C至65°C的范围中的温度进行结合、裂解、标记或修饰。在优选的方面中, 本发明的Cas蛋白或多肽片段可以包含与

SEQ ID NO:1具有至少75%同一性;优选地至少85%;更优选地至少90%;甚至更优选地至少95%同一性的氨基酸序列。

[0053] Cas蛋白或多肽可以与识别靶核酸链上的靶核酸序列的靶向RNA分子组合使用,其中非靶核酸序列具有与在非靶链上的前间区序列的3'末端直接地相邻的PAM序列,如本文所公开的。因此,PAM序列可以包含序列5'-NNNNC-3',并且Cas蛋白可以在20℃和100℃的范围中(包括端点)的温度,优选地在30℃和90℃的范围中(包括端点)的温度,在37℃和78℃的范围中(包括端点)的温度、在40℃和80℃的范围中(包括端点)的温度、在50℃和70℃的范围中(包括端点)的温度、或在55℃和65℃的范围中(包括端点)的温度结合、裂解、标记或修饰靶链。优选地,PAM序列从5'-末端以5'-NNNNC-3'开始,并且Cas蛋白可以在20℃和100℃的范围中(包括端点)的温度,优选地在30℃和90℃的范围中(包括端点)的温度,在37℃和78℃的范围中(包括端点)的温度、在40℃和80℃的范围中(包括端点)的温度、在50℃和70℃的范围中(包括端点)的温度、或在55℃和65℃的范围中(包括端点)的温度结合、裂解、标记或修饰靶链。优选地,PAM序列从5'-末端以5'-NNNNNNA-3'开始,并且Cas蛋白可以在20℃和100℃的范围中(包括端点)的温度,优选地在30℃和90℃的范围中(包括端点)的温度,在37℃和78℃的范围中(包括端点)的温度、在40℃和80℃的范围中(包括端点)的温度、在50℃和70℃的范围中(包括端点)的温度、或在55℃和65℃的范围中(包括端点)的温度结合、裂解、标记或修饰靶链。进一步优选地,PAM序列从5'-末端以5'-NNNNNCNA-3'开始,并且Cas蛋白可以在20℃和100℃的范围中(包括端点)的温度,优选地在30℃和90℃的范围中(包括端点)的温度,在37℃和78℃的范围中(包括端点)的温度、在40℃和80℃的范围中(包括端点)的温度、在50℃和70℃的范围中(包括端点)的温度、或在55℃和65℃的范围中(包括端点)的温度结合、裂解、标记或修饰靶链。

[0054] 更特别地,本发明的Cas蛋白或多肽可以包含与SEQ ID NO:1具有如下同一性百分比的氨基酸序列:至少60%、至少61%、至少62%、至少63%、至少64%、至少65%、至少66%、至少67%、至少68%、至少69%、至少70%、至少71%、至少72%、至少73%、至少74%、至少75%、至少76%、至少77%、至少78%、至少79%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、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%、至少99%、至少99.5%或至少99.8%。同一性百分比可以是至少89%。同一性百分比可以是至少90%。优选地,同一性百分比将是至少95%,例如98%。

[0055] 与SEQ ID NO:1的氨基酸序列同一性百分比根据在选定的比较窗口中的序列共享的相同位置的数目,考虑到需要被引入用于两个序列的最佳比对的空位的数目和每一个空位的长度是可确定的。

[0056] 本发明的Cas蛋白或多肽片段可以根据参考序列SEQ ID NO:1和如由序列同一性百分比定义的任何上文提及的其百分比变体二者,单独地或与任何上文提及的氨基酸基序(即SEQ ID NO:2和/或SEQ ID NO:3和/或SEQ ID NO:4和/或SEQ ID NO:5和/或SEQ ID NO:6)组合作为基本特征来表征。

[0057] 本发明提供了如本文提供的靶向RNA分子和本发明的Cas蛋白或多肽用于结合、裂解、标记或修饰包含靶核酸序列的靶核酸链的用途。优选地,所述结合、裂解、标记或修饰发生在本文所公开的温度,例如20℃和100℃之间的温度。本发明还提供了结合、裂解、标记或

修饰靶核酸链中的靶核酸序列的方法,该方法包括设计如本文提供的靶向RNA分子和形成包含该靶向RNA分子和本发明的Cas蛋白或多肽的核糖核蛋白复合体。优选地,核糖核蛋白复合体在本文所公开的温度,例如在37℃和100℃之间的温度结合、裂解、标记或修饰靶核酸序列。

[0058] 本发明的用途和方法可以在体内例如在细菌细胞中进行,并且本发明的核蛋白可以在体内例如在细菌细胞中形成和使用。可选地,本发明的用途和方法可以在体外进行,并且本发明的核蛋白可以在体外形成和使用。本发明的Cas蛋白可以以分离的形式被提供,例如当在体外使用时或当通过转染添加至细胞时,Cas蛋白可以例如在通过编码Cas蛋白的核酸瞬时或稳定转化细胞后被异源表达,靶向RNA分子可以在通过编码RNA分子的核酸瞬时或稳定转化细胞后从表达载体转录,和/或RNA分子可以以分离的形式提供,例如当在体外使用时或当通过转染添加至细胞时。在优选的实施方案中,在编码Cas蛋白或多肽的核酸在宿主细胞基因组中稳定整合后,Cas蛋白或多肽从宿主细胞的基因组被表达。因此,使用用于将蛋白或核酸分子添加至细胞(否则在细胞中,该蛋白或核酸分子不存在)的任何人工或人为的方法,Cas蛋白和/或RNA分子可以被添加至体内或体外环境。

[0059] 包含靶核酸序列的多核苷酸可以被Cas蛋白裂解,并且任选地裂解可以是DNA裂解。包含靶序列的靶核酸链可以是双链DNA,并且方法或用途可以引起包含靶核酸序列的多核苷酸中的双链断裂。包含靶核酸序列的多核苷酸可以是双链DNA,Cas蛋白可以缺乏切割双链DNA的能力,并且用途或方法可以引起多核苷酸的基因沉默。

[0060] Cas蛋白或多肽可以以250nM或更低的浓度,例如以200nM或更低、150nM或更低、100nM或更低、50nM或更低、25nM或更低、10nM或更低、5nM或更低、1nM或更低或0.5nM或更低的浓度被提供用于本发明的方法、用途和核蛋白。可选地,Cas蛋白或多肽可以以至少0.5nM、至少1nM、至少5nM、至少10nM、至少25nM、至少50nM、至少100nM、至少150nM、至少200nM或至少250nM的浓度被提供。本发明的PAM序列可以在第八个位置中具有腺嘌呤,使得PAM序列包含序列5'-NNNNNNA-3',并且Cas蛋白或多肽的浓度可以是100nM或更低、50nM或更低、25nM或更低、10nM或更低、5nM或更低、1nM或更低或0.5nM或更低。PAM序列可以包含序列5'-NNNNCNNA-3',并且Cas蛋白或多肽的浓度可以是100nM或更低、50nM或更低、25nM或更低、10nM或更低、5nM或更低、1nM或更低或0.5nM或更低。PAM序列可以包含序列5'-CCCCCNA-3' [SEQ ID NO:10],并且Cas蛋白或多肽的浓度可以是100nM或更低、50nM或更低、25nM或更低、10nM或更低、5nM或更低、1nM或更低或0.5nM或更低。

[0061] 此外,本发明提供了编码任何上文提及的本发明的蛋白或多肽的核酸。核酸可以是分离的或呈表达构建体的形式。

[0062] 在本发明的所有上文提及的方面中,氨基酸残基可以被保守或非保守取代。保守氨基酸取代是指其中氨基酸残基被取代为具有类似的化学特性(例如,电荷或疏水性)的其他氨基酸残基并且因此不改变所得的多肽的功能特性的那些氨基酸取代。

[0063] 类似地,本领域普通技术人员将理解的是,核酸序列可以被保守或非保守取代,而不影响多肽的功能。保守修饰的核酸是被取代为编码氨基酸序列的相同或功能上相同的变体的核酸的那些核酸。本领域的读者将理解的是,核酸中的每一个密码子(除了AUG和UGG;通常,分别为甲硫氨酸或色氨酸的唯一密码子)可以被修饰以产生功能上相同的分子。因此,编码本发明的多肽的多核苷酸或多肽的每一个沉默变体(即同义密码子)隐含在每一个

描述的多肽序列中。

[0064] 本发明提供了经转化的细胞,该经转化的细胞具有在双链靶多核苷酸中的靶核酸序列,所述细胞包含如本文提供的Cas蛋白或多肽和至少一种如本文提供的靶向RNA分子、和表达载体,所述表达载体包含编码所述Cas蛋白和所述靶向RNA分子中的至少一种的核酸。如本文所公开的,Cas蛋白和靶向RNA分子可以使得结合、裂解、标记或修饰靶序列能够在升高的温度或在例如37℃和100℃之间的温度范围在经转化的细胞中发生,或允许结合、裂解、标记或修饰靶序列在升高的温度或在例如37℃和100℃之间的温度范围在经转化的细胞中发生。本发明还提供了结合、裂解、标记或修饰细胞中的靶核酸的方法,该方法包括1)用表达载体转化、转染或转导细胞,所述表达载体包含编码本发明的Cas蛋白或多肽的核苷酸序列和编码本发明的靶向RNA分子的核苷酸序列;或2)用表达载体和另外的表达载体转化、转染或转导细胞,所述表达载体包含编码本发明的Cas蛋白或多肽的核苷酸序列,所述另外的表达载体包含编码本发明的靶向RNA分子的核苷酸序列;或3)用表达载体转化、转染或转导细胞,该表达载体包含编码本发明的Cas蛋白或多肽的核苷酸序列,和将如本文提供的靶向RNA分子递送至细胞或递送到细胞中。Cas蛋白或多肽可以例如在编码Cas蛋白或多肽的核苷酸序列稳定整合到基因组中后从经转化的细胞的基因组表达。

[0065] 本发明还提供包含一种或更多种试剂的试剂盒,所述试剂盒用于实施本发明的用途和方法,或用于生成本发明的经转化的细胞或核蛋白复合体,所述试剂盒包含:本发明的Cas蛋白或多肽或包含编码本发明的Cas蛋白或多肽的核酸序列的表达载体;和/或本发明的靶向RNA分子或包含编码本发明的靶向RNA分子的核酸序列的表达载体。试剂盒还可以包含用于实施本发明的说明书,例如用于如何设计根据本发明的靶向RNA分子的说明书。

[0066] RNA指导(RNA Guides)和靶序列

[0067] 本发明的Cas蛋白允许在升高的温度对靶核酸进行序列特异性结合、裂解、加标签、标记或修饰。靶核酸可以是DNA(单链或双链)、RNA或合成的核酸。本发明的特别有用的应用是通过与一种或更多种指导RNA(gRNA)呈复合体形式的本发明的一种或更多种Cas蛋白对基因组DNA进行序列特异性靶向和修饰,所述指导RNA与基因组DNA的靶向序列互补地结合。因此,靶核酸优选地是双链DNA。这类靶向可以在体外或体内进行。优选地,这类靶向在体内进行。以该方式,本发明的Cas蛋白可以被用于靶向和修饰位于细胞的基因组DNA中的特定DNA序列。设想了Cas系统可以被用于修饰多种细胞类型和/或不同的生物体中的基因组。

[0068] gRNA,也被称为靶向RNA分子,识别多核苷酸靶链上的靶核酸序列。RNA分子可以被设计为识别双链靶多核苷酸中的靶序列,其中非靶链包含与前间区序列的3'末端直接地相邻的前间区相邻基序(PAM)序列。本文公开了以最佳方式与本发明的Cas蛋白和多肽一起起作用的PAM序列。在具有这些PAM序列知识的情况下,gRNA可以被设计用于与本发明的Cas蛋白和多肽一起跨越本发明的温度范围和增加的温度使用。

[0069] 因此,本发明提供了核糖核蛋白复合体,所述核糖核蛋白复合体包含如上文描述的本发明的Cas蛋白或多肽,并且还包含至少一种RNA分子,所述至少一种RNA分子由于其识别靶多核苷酸中的特定核苷酸序列而具有靶向功能。本发明还提供了至少一种靶向RNA分子和Cas蛋白或多肽用于结合、裂解、标记或修饰靶核酸链的用途,和使用本发明的核糖核蛋白或核蛋白结合、裂解、标记或修饰靶核酸链中的靶核酸序列的方法,以及具有Cas蛋白

或多肽和靶向RNA分子的经转化的细胞。根据本文提供的PAM序列,靶多核苷酸还可以包含与前间区序列的3'末端直接地相邻的定义的PAM序列。PAM序列的长度可以是6个、7个或8个核酸,或更长,优选地长度是8个核酸。优选地,RNA分子是单链RNA分子,例如CRISPR RNA (crRNA),并且例如通过杂交与tracrRNA缔合。靶向RNA可以是crRNA和tracrRNA的嵌合体。上文提及的RNA分子可以具有与靶核苷酸序列具有至少90%同一性或互补性的核糖核苷酸序列。任选地,RNA分子具有与靶核苷酸序列具有至少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%、至少99%或100%同一性或互补性的核糖核苷酸序列。优选的靶核苷酸序列是DNA。

[0070] 在优选的方面中,本发明提供了如上文描述的核糖核蛋白复合体,其中至少一种靶向RNA分子沿其长度与靶DNA序列基本上互补。

[0071] 靶向RNA分子可以在核蛋白复合体中与靶序列结合或缔合,使得包含靶序列的靶多核苷酸和在非靶链上的PAM序列,可以与本发明的核蛋白复合体缔合并因此形成本发明的核蛋白复合体的一部分。

[0072] 因此,与本发明的Cas蛋白缔合的RNA指导的序列的改变允许Cas蛋白被编程为在与指导RNA互补的位点处标记或切割双链DNA。

[0073] 优选地,在本发明的核糖核蛋白复合体中的至少一种靶向RNA分子的长度是在35个至135个残基的范围中,任选地在以下范围中:35个至134个残基、35个至133个残基、35个至132个残基、35个至131个残基、35个至130个残基、35个至129个残基、35个至128个残基、35个至127个残基、35个至126个残基、35个至125个残基、35个至124个残基、35个至123个残基、35个至122个残基、35个至121个残基、35个至120个残基、35个至119个残基、35个至118个残基、35个至117个残基、35个至116个残基、35个至115个残基、35个至114个残基、35个至113个残基、35个至112个残基、35个至111个残基、35个至110个残基、35个至109个残基、35个至108个残基、35个至107个残基、35个至106个残基、35个至105个残基、35个至104个残基、35个至103个残基、35个至102个残基、35个至101个残基、35个至100个残基、35个至99个残基、35个至98个残基、35个至97个残基、35个至96个残基、35个至95个残基、35个至94个残基、35个至93个残基、35个至92个残基、35个至91个残基、35个至90个残基、35个至89个残基、35个至88个残基、35个至87个残基、35个至86个残基、35个至85个残基、35个至84个残基、35个至83个残基、35个至82个残基、35个至81个残基、35个至80个残基、35个至79个残基、35个至78个残基、35个至77个残基、35个至76个残基、35个至75个残基、35个至74个残基、35个至73个残基、35个至72个残基、35个至71个残基、35个至70个残基、35个至69个残基、35个至68个残基、35个至67个残基、35个至66个残基、35个至65个残基、35个至64个残基、35个至63个残基、35个至62个残基、35个至61个残基、35个至60个残基、35个至59个残基、35个至58个残基、35个至57个残基、35个至56个残基、35个至55个残基、35个至54个残基、35个至53个残基、35个至52个残基、35个至51个残基、35个至50个残基、35个至49个残基、35个至48个残基、35个至47个残基、35个至46个残基、35个至45个残基、35个至44个残基、35个至43个残基、35个至42个残基、35个至41个残基、35个至40个残基、35个至39个残基、35个至38个残基、35个至37个残基、35个至36个残基或35个残基。优选地,至少一种RNA分子的长度是在以下范围中:36个至174个残基、37个至173个残基、38个至172个残基、39个至171个残基、40个至170个残基、41个至169个残基、42个至168个残基、43个至167个残基、

44个至166个残基、45个至165个残基、46个至164个残基、47个至163个残基、48个至162个残基、49个至161个残基、50个至160个残基、51个至159个残基、52个至158个残基、53个至157个残基、54个至156个残基、36个至74个残基、37个至73个残基、38个至72个残基、39个至71个残基、40个至70个残基、41个至69个残基、42个至68个残基、43个至67个残基、44个至66个残基、45个至65个残基、46个至64个残基、47个至63个残基、48个至62个残基、49个至61个残基、50个至60个残基、51个至59个残基、52个至58个残基、53个至57个残基、54个至56个残基。

[0074] 在优选的方面中,本发明提供了核糖核蛋白复合体,其中至少一种RNA分子的互补部分是至少30个残基长。可选地,至少一种RNA分子的互补部分可以是31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41个、42个、43个、44个、45个、46个、47个、48个、49个、50个、51个、52个、53个、54个、55个、56个、57个、58个、59个、60个、61个、62个、63个、64个、65个、66个、67个、68个、69个、70个、71个、72个、73个、74个或75个残基长。

[0075] 靶向RNA分子将优选地需要对靶核酸序列的高特异性和亲和力。在1 μ M至1pM,优选地1nM至1pM;更优选地1pM-100pM的范围中的解离常数(K_d)是期望的,如可以通过非变性凝胶电泳(native gel electrophoresis)或可选地等温滴定量热法、表面等离子体共振或基于荧光的滴定方法来确定。亲和力可以使用电泳迁移率变动测定(EMSA)来确定,所述电泳迁移率变动测定也被称为凝胶阻滞测定(参见Semenova等人(2011)PNAS108:10098-10103)。

[0076] 靶向RNA分子优选地在从原核生物自然界已知作为CRISPR RNA(crRNA)分子的分子上建模(modeled)。crRNA分子的结构已经被建立并且在Jore等人,2011,Nature Structural&Molecular Biology 18:529-537中更详细地解释。简而言之,I-E型的成熟crRNA通常是61个核苷酸长,并且由8个核苷酸的5'“手柄”区域、32个核苷酸的“间隔区”序列和21个核苷酸的3'序列组成,所述21个核苷酸的3'序列形成具有四核苷酸环的发夹(图5)。I型系统不同于II型(Cas9)系统,并且不同的系统的详细内容在Van der Oost 2014Nat Rev Micr 12:479-492中被描述。在II型(Cas9)系统中,存在不同的处理机制,使用第二种RNA(tracrRNA)和两种核糖核酸酶。II型中的成熟的crRNA保持与tracrRNA的片段附接而非发夹(图5)。然而,本发明中使用的RNA不必严格地遵循天然存在的crRNA的设计被设计,无论是长度、区域还是特定的RNA序列。但明确的是,用于在本发明中使用的RNA分子可以基于公共数据库中的或新发现的基因序列信息来设计,并且然后人工地制备,例如通过完全或部分化学合成。本发明的RNA分子还可以被设计并通过在遗传修饰的细胞或无细胞表达系统中表达的方式来产生,并且该选择可以包括一些或所有的RNA序列的合成。

[0077] II型(Cas9)中的crRNA的结构和要求也已经在Jinek等人,2012同上中被描述。在I型中,存在形成间隔区序列的5'末端的所谓的“种子(SEED)”部分并且其5'侧翼为8个核苷酸的5'手柄。Semenova等人(2011,PNAS108:10098-10103)已经发现种子序列的所有残基都应该与靶序列互补,尽管对于在位置6的残基可以容忍错配(图5)。在II型中,存在位于间隔区的3'末端处的10-12个核苷酸的种子(图5)(由Van der Oost 2014同上综述)。类似地,当设计和制备针对靶基因座(即序列)处的本发明的核糖核蛋白复合体的RNA组分时,可以应用于II型种子序列的必要匹配和错配规则。

[0078] 因此,本发明包括检测和/或定位靶核酸分子中的单个碱基变化的方法,所述方法

包括使核酸样品与如上文描述的本发明的核糖核蛋白复合体或与如上文描述的本发明的Cas蛋白或多肽和单独的靶向RNA组分接触,并且其中靶向RNA的序列(包括当在核糖核蛋白复合体中时)使得通过在例如8个核苷酸残基的连续序列的位置6处的单个碱基变化区分正常等位基因和突变等位基因。

[0079] 不希望受特定理论束缚,可以被用于制备本发明的核糖核蛋白复合体的靶向RNA组分的设计规则涉及双链靶多核苷酸中的所谓的“PAM”(前间区相邻基序)序列。在大肠杆菌(*E. coli*)的I-E型系统中,PAM序列可以是保守的核苷酸残基三联体,诸如5'-CTT-3'、5'-CAT-3'、5'-CCT-3'、5'-CAC-3'、5'-TTT-3'、5'-ATT-3'、和5'-AWG-3',其中W是A、T或U。在I型中,位于靶向链中的PAM序列通常在对应于种子的5'的位置处。然而,在II型中,PAM位于靠近crRNA间隔区的3'末端的置换链或非靶链上的另一末端处,在对应于种子的3'的位置处(图5)(Jinek等人,2012,同上)。对于酿脓链球菌(*Streptococcus pyogenes*)Cas9,PAM序列具有保守的核苷酸残基对5'-NGG-3'。最近,已经表征了不同的Cas9变体(IIA型和IIC型)(Ran等人,2015Nature 520:186-191),并且已经揭示了PAM(参见Ran等人,2015,同上)。目前建立的Cas9 PAM包括:IIA型5'-NGGNNN-3'(酿脓链球菌)、5'-NGTNNN-3'(巴氏链球菌(*Streptococcus pasteurianus*))、5'-NNGGAAN-3'(嗜热链球菌(*Streptococcus thermophilus*))、5'-NNGGGNN-3'(金黄色葡萄球菌(*Staphylococcus aureus*))、和IIC型5'-NGGNNN-3'(白喉棒杆菌(*Corynebacterium diphtheriae*))、5'-NNGGGTN-3'(红嘴鸥弯曲杆菌(*Campylobacter lari*))、5'-NNNCATN-3'(Parvibaculum lavamentivorans)、5'-NNNNGTA-3'(灰色奈瑟菌(*Neisseria cinerea*))。热脱氮地芽孢杆菌(*Geobacillus thermodenitrificans*)T12的Cas9(本发明)属于IIC型(Ran等人,2015,同上)。发明人已经出乎意料地发现,用于与本发明一起使用的PAM序列的选择可以影响其中本发明的Cas蛋白和多肽将与靶序列相互作用的一种或更多种温度。特别地,本发明人已经发现优选8聚体(8-mer)PAM序列以跨越宽的温度范围赋予活性,其中胞嘧啶在靶序列3'末端后的第5个位置中,和/或腺嘌呤在第8个位置中。在前间区序列的3'末端后在PAM序列的第1个、第2个、第3个、第4个和/或第6个位置中也存在对于胞嘧啶的偏好。

[0080] 在本发明的实施方案中,靶向RNA分子可以具有在35-200个残基范围中的长度。在优选的实施方案中,与期望的核酸序列互补的并且被用于靶向期望的核酸序列的RNA的部分是从15个至32个残基长。在天然存在的crRNA的情况中,这将对应于如例如在Semenova等人(2011同上)中显示出的间隔区部分。

[0081] 本发明的核糖核蛋白复合体可以具有包含位于RNA序列5'的来源于CRISPR重复的8个残基的靶向组分,该RNA序列与DNA靶序列具有实质的互补性。与DNA靶序列具有互补性的RNA序列将被理解为对应于在crRNA的情况中作为间隔区序列。RNA的5'侧翼序列将被认为对应于crRNA的5'手柄;如例如在Semenova等人(2011同上)中显示出的。

[0082] 本发明的核糖核蛋白复合体可以具有位于与DNA靶序列具有互补性的靶向RNA序列3'的发夹和四核苷酸环形成序列,即位于对应于在crRNA中的间隔区序列侧翼的3'手柄的3';例如,如在Semenova等人(2011同上)中显示出的。

[0083] 不希望受特定理论的束缚,在优选的核糖核蛋白复合体和双链靶多核苷酸中,不与核糖核蛋白复合体的靶向RNA配对的非靶核酸链可以包含直接地3'相邻的PAM序列,该PAM序列选自5'-NNNNCNA-3'、5'-CNNNCNN-3'、5'-NNNCCNN-3'、5'-NNCNCNN-3'、5'-

NNNNCCN-3' 和 5'-NCNNCCN-3' 的一种或更多种。优选地, PAM 序列可以选自 5'-NNNNC-3'、5'-NNNNCNNA-3'、5'-CNNNC-3'、5'-CNNNCNNA-3'、5'-NCNNC-3'、5'-NCNNCNNA-3'、5'-NNCNC-3'、5'-NNCNCNNA-3'、5'-NNNCC-3'、5'-NNNCCNNA-3'、5'-NNNNCC-3'、5'-NNNNCCNNA-3'、5'-CCNNC-3'、5'-CCNNCNNA-3'、5'-CNCNC-3'、5'-CNCNCNNA-3'、5'-CNNCCN-3'、5'-CNNCCNNA-3'、5'-CNNNCC-3'、5'-CNNNCCNNA-3'、5'-CCNCNC-3'、5'-CCNCNCNNA-3'、5'-CCNCCN-3'、5'-CCNCCNNA-3'、5'-CCNNCC-3'、5'-CCNNCCNNA-3'、5'-CCCCC-3' [SEQ ID NO:12]、5'-CCCCCNNA-3' [SEQ ID NO:13]、5'-CCCCCC-3' [SEQ ID NO:14]、5'-CCCCCCNNA-3' [SEQ ID NO:10]、5'-NCCNC-3'、5'-NCCNCNNA-3'、5'-NCCCC-3'、5'-NCCCCNNA-3'、5'-NCCCCC-3' [SEQ ID NO:15]、5'-NCCCCCNA-3' [SEQ ID NO:16]、5'-NNCCC-3'、5'-NNCCCNNA-3'、5'-NNCCCC-3'、5'-NNCCCCNNA-3'、5'-NNNCCC-3'、和 5'-NNNCCNNA-3'。PAM 序列可以是 5'-CNCCCCAC-3' [SEQ ID NO:17]、5'-CCCCCAG-3' [SEQ ID NO:18]、5'-CCCCCAA-3' [SEQ ID NO:11]、5'-CCCCCAT-3' [SEQ ID NO:19]、5'-CCCCCAC-3' [SEQ ID NO:20]、5'-ATCCCCAA-3' [SEQ ID NO:21]、或 5'-ACGGCCAA-3' [SEQ ID NO:22]。优选地, PAM 序列将是序列 5'-NNNNCNNA-3'。然而, 将理解的是, 取决于期望的应用和/或 Cas 蛋白或多肽的浓度, 可以使用核苷酸的其他组合。这些序列对应于在天然存在的 crRNA 的情况中被称为“前间区相邻基序”或“PAM”的序列。在 IIC 型 CRISPR/Cas 系统中, 这些 PAM 序列促进 Cascade/crRNA 复合体与其 dsDNA 靶稳定相互作用, 以确保 crRNA 对靶序列的高度的特异性-在天然系统靶中和因此也优选地对于根据本发明的 RNA。优选地, 与前间区直接地相邻的序列将不是 5'-NNNCATN-3'。

[0084] 本文提供的本发明的 PAM 序列包含本文所公开的序列, 例如作为 6 聚体、7 聚体或 8 聚体序列。6 聚体、7 聚体或 8 聚体序列可以直接地从在非靶链上的前间区序列的 3' 开始, 而在前间区序列和 PAM 序列的 5' 末端之间不具有另外的核酸间隔, 该前间区序列与靶向 RNA 结合的序列互补。然而, 将理解的是, 在 6 聚体、7 聚体或 8 聚体序列的 3' 末端处, 可以存在形成 PAM 序列的一部分的另外的核酸。另外地或可选地, 非靶链可以包含在 PAM 序列的 3' 的另外的核酸。

[0085] 本发明的核蛋白复合体可以包含本发明的核糖核蛋白复合体和核糖核蛋白与其缔合的核酸的靶核酸链。

[0086] 结合、裂解、标记和修饰的温度

[0087] 本发明的 Cas 蛋白的活性, 例如核酸酶活性的温度范围, 包括最佳温度范围, 显著地高于已知的 Cas9 蛋白的活性的温度范围。此外, 在其中本发明的 Cas 蛋白保持活性的范围的上限比已知的 Cas9 蛋白保持活性的范围的上限高得多。较高的最佳温度和功能范围在高温度的遗传工程中提供了显著的优点, 并且因此, 例如, 在编辑嗜热生物体的基因组中提供了显著的优点, 嗜热生物体中的许多在升高的温度进行的一系列工业、农业和制药过程中具有效用。因此, 本发明的方法、用途、核蛋白和经转化的细胞可以是在工业过程中有用的, 例如为代谢工程目的提供基因组编辑。本发明的 PAM 序列 (与非靶链中的前间区序列直接地相邻) 的存在, 改进了 Cas 蛋白和多肽对靶序列的特异性, 并且支持 Cas 蛋白和多肽在更高温度和跨越更大功能温度范围的用途。

[0088] 有益地, 本发明的 Cas 蛋白或多肽能够在从 20℃ 至 100℃ 的温度进行核酸结合、裂解、标记或修饰, 但在升高的温度, 例如在 41℃ 和 122℃ 之间的温度, 优选地在 50℃ 和 100℃ 之间的温度是特别有用的。本发明的 Cas 蛋白和多肽能够结合、裂解、标记或修饰 DNA、RNA 和

合成的核酸。本发明的Cas蛋白或多肽还可以在例如20℃至50℃的范围中的温度提供用于核酸酶活性、基因编辑和核酸标记应用的可操作性。

[0089] 在本文包括温度范围的情况下,预期端点被包括在所公开的温度范围中,即该范围是“包括性的”。例如,当陈述在20℃和100℃之间的范围中的温度存在活性时,20℃和100℃的温度被包括在所述范围中。

[0090] 优选地,当与合适的gRNA(指导RNA,也称为靶向RNA分子)缔合时(所述gRNA识别一个或多个多核苷酸分子中的待被结合、裂解、标记或修饰的靶序列),本发明的Cas蛋白或多肽在50℃至100℃的范围中的温度,任选地在55℃至100℃、60℃至100℃、65℃至100℃、70℃至100℃、75℃至100℃、80℃至100℃、85℃至100℃、90℃至100℃、95℃至100℃的温度对靶序列进行结合、裂解、标记或修饰。更优选地,本发明的Cas蛋白在51℃至99℃、52℃至98℃、53℃至97℃、54℃至96℃、55℃至95℃、56℃至94℃、57℃至93℃、58℃至92℃、59℃至91℃、60℃至90℃、61℃至89℃、62℃至88℃、63℃至87℃、64℃至86℃、65℃至85℃、66℃至84℃、67℃至83℃、68℃至82℃、69℃至81℃、70℃至80℃、71℃至79℃、72℃至78℃、73℃至77℃、74℃至76℃的范围中的温度、或在75℃的温度裂解、标记或修饰核酸。优选地,本发明的Cas蛋白在60℃至80℃、61℃至79℃、62℃至78℃、63℃至77℃、64℃至76℃、60℃至75℃、60℃至70℃的范围中的温度结合、裂解、标记或修饰核酸。最佳地,本发明的Cas蛋白在60℃至65℃的范围中的温度,优选地在65℃结合、裂解、标记或修饰核酸。

[0091] 靶向RNA分子可以被设计用于与本发明的Cas蛋白和多肽一起使用,其中靶向RNA分子与靶链中的靶序列结合,并且非靶链还包含本文提供的紧邻前间区序列3'的PAM序列。PAM序列可以包含5'-NNNNNNNA-3',优选地5'-NNNNCNNA-3',例如5'-CCCCCNA-3'[SEQ ID NO:10]或5'-CCCCCAA-3'[SEQ ID NO:11],并且本发明的用途、方法、经转化的细胞和核蛋白可以跨越从55℃至65℃的温度范围,优选地跨越从50℃至70℃、从40℃至65℃、从45℃至75℃、从37℃至78℃和/或从20℃至80℃的温度范围,提供靶链的结合、裂解、标记和/或修饰。

[0092] 在本发明的所有方面中,Cas蛋白或多肽可以获得自或来源于细菌、古核生物或病毒;或可选地可以从头合成。在优选的实施方案中,本发明的Cas蛋白或多肽来源于嗜热原核生物体,其可以被分类为古核生物或细菌,但优选地是细菌。更优选地,本发明的Cas蛋白或多肽将来源于嗜热细菌。本文中,术语嗜热应当理解为意指能够在相对地高的温度存活和生长,例如,在本发明的情况中,能够在41℃和122℃(106°F和252°F)之间的温度进行核酸裂解、结合或修饰。优选地,本发明的Cas蛋白或多肽可以从一种或更多种嗜热细菌分离,并且将在60℃以上发挥功能。优选地,本发明的Cas蛋白或多肽可以从一种或更多种嗜热细菌分离,并且将在60℃至80℃的范围中并且最佳地在60℃与65℃之间的范围中发挥功能。在优选的实施方案中,本发明的Cas蛋白或多肽来源于地芽孢杆菌属的种(*Geobacillus* sp.)。更优选地,本发明的Cas蛋白来源于热脱氮地芽孢杆菌。甚至更优选地,本发明的Cas蛋白来源于热脱氮地芽孢杆菌T12。本发明的Cas蛋白或多肽可以来源于病毒。

[0093] 功能部分

[0094] 有益地,可以使用本发明的Cas蛋白、多肽和核糖核蛋白复合体以序列特异性方式靶向任何多核苷酸序列的能力来以某种方式修饰靶核酸,例如通过裂解靶核酸和/或标记靶核酸和/或修饰靶核酸。因此将理解的是,可以随Cas蛋白或多肽提供另外的蛋白以实现

这一点。因此,本发明的Cas蛋白或多肽还可以包含至少一个功能部分,和/或本发明的Cas蛋白、多肽或核糖核蛋白复合体可以作为蛋白复合体的一部分被提供,所述蛋白复合体包含至少一种另外的蛋白。在优选的方面中,本发明提供了Cas蛋白、多肽或核糖核蛋白复合体,其中Cas蛋白或至少一种另外的蛋白还包含至少一个功能部分。至少一个功能部分可以与Cas蛋白融合或连接。优选地,至少一个功能部分可以通过在天然或人工蛋白表达系统中表达而与Cas蛋白翻译地融合。可选地,至少一个功能部分可以通过化学合成步骤与Cas蛋白共价地连接。优选地,至少一个功能部分与Cas蛋白的N-末端和/或C-末端;优选地C-末端融合或连接。

[0095] 期望地,至少一个功能部分将是蛋白。它可以是异源蛋白,或可选地可以是Cas蛋白来源于其的细菌物种天然的。至少一个功能部分可以是蛋白;任选地选自解旋酶、核酸酶、解旋酶-核酸酶、DNA甲基化酶、组蛋白甲基化酶、乙酰基转移酶、磷酸酶、激酶、转录(共)活化物、转录阻遏物、DNA结合蛋白、DNA结构蛋白(DNA structuring protein)、标志物蛋白、报告物蛋白、荧光蛋白、配体结合蛋白、信号肽、亚细胞定位序列、抗体表位或亲和纯化标签。

[0096] 在特别地优选的方面中,本发明提供了Cas蛋白、多肽或核糖核蛋白复合体,其中至少一个功能部分是标志物蛋白,例如GFP。

[0097] 核酸酶活性

[0098] 本发明的Cas核糖核蛋白在本文所公开的温度,优选地在升高的温度,例如在50℃和100℃之间的温度,具有核酸结合、裂解、标记或修饰活性。本发明的核糖核蛋白可以能够结合、裂解、标记或修饰DNA、RNA或合成的核酸。在优选的方面中,本发明的Cas核糖核蛋白能够以序列特异性方式裂解DNA,特别是双链DNA。

[0099] 本发明的Cas蛋白、多肽或核糖核蛋白可以具有多于一个核酸酶结构域。位点特异性核酸酶可以允许在沿DNA的链的选定的位置处生成双链断裂(DSB)。在靶宿主细胞中,这使DSB能够在基因组中的特定的预先选定的位置处生成。通过位点特异性核酸酶创建这类断裂促使内源细胞修复机制被重新使用,以在感兴趣的基因组中的期望的位置处插入、缺失或修饰DNA。

[0100] Cas蛋白或多肽分子的一个或更多个核酸酶活性位点可以被失活,例如,以允许与Cas蛋白或多肽连接或融合的另一个功能部分(例如核酸酶结构域诸如FokI核酸酶)的活性。

[0101] 因此,尽管本发明的Cas蛋白、多肽和核糖核蛋白可以具有内源核酸酶活性的事实,但对于某些应用,可以期望使Cas蛋白的天然核酸酶活性失活并且提供Cas蛋白或核糖核蛋白复合体,其中天然Cas9核酸酶活性被失活并且Cas蛋白与至少一个功能部分连接。这类应用中的一种是通过补充天然的Cas9核酸酶活性来降低错误靶向事件的发生率。这可以期望通过使Cas蛋白或核糖核蛋白复合体的天然Cas9核酸酶活性失活和提供优选地与Cas蛋白融合的异源核酸酶来实现。因此,本发明提供了Cas蛋白或核糖核蛋白复合体,其中至少一个功能部分是核酸酶结构域,优选地是FokI核酸酶结构域。在特别地优选的方面中,与FokI核酸酶结构域融合的本发明的Cas蛋白或核糖核蛋白复合体作为蛋白复合体的一部分被提供,优选地包含与FokI核酸酶结构域融合的本发明的另一个Cas蛋白或核糖核蛋白复合体,并且其中两个复合体靶向靶基因组DNA的相对链。

[0102] 对于一些应用,可以期望完全地减弱Cas蛋白、多肽或核糖核蛋白的核酸酶活性,例如在其中Cas蛋白或核糖核蛋白复合体被用于识别和修饰核酸中的特定靶序列,例如用于将其标记为诊断测试的一部分的应用中。在此类应用中,Cas蛋白的核酸酶活性可以被失活,并且与Cas蛋白融合的功能部分可以是蛋白;任选地选自解旋酶、核酸酶、解旋酶-核酸酶、DNA甲基化酶、组蛋白甲基化酶、乙酰基转移酶、磷酸酶、激酶、转录(共)活化物、转录阻遏物、DNA结合蛋白、DNA结构蛋白、标志物蛋白、报告物蛋白、荧光蛋白、配体结合蛋白、信号肽、亚细胞定位序列、抗体表位或亲和纯化标签。

[0103] 在优选的方面中,缺乏核酸酶活性的催化失活或“死亡”的Cas蛋白或多肽(dCas)可以与靶核酸序列结合,并且从而在空间上阻遏该序列的活性。例如,可以设计与基因的启动子或外显子序列互补的靶向RNA,使得dCas和靶向RNA与基因的结合在空间上阻遏基因序列的转录起始或延伸,从而阻遏基因的表达。可选地,本文描述的方法和用途可以使用为切口酶(nickase)的gtCas9的修饰的核酸酶变体。切口酶可以通过gtCas9核酸酶的HNH或RuvC催化结构域的任一个中的突变来创建。这已经针对酿脓链球菌Cas9(spCas)显示出,spCas9-突变体D10A和H840A,它们分别地具有无活性的RuvC或HNH核酸酶结构域。这两种突变的组合导致催化死亡的Cas9变体(Standage-Beier,K.等人,2015,ACS Synth.Biol.4,1217-1225;Jinek,M.等人,2012,Science 337,816-821;Xu,T.等人,2015,Appl.Environ.Microbiol.81,4423-4431)。基于序列同源性(图3),这些残基可以是在gtCas9中的D8(在图3中的D17)和D581或H582(图3)。

[0104] 在特别地优选的方面中,本发明提供了Cas蛋白或核糖核蛋白复合体,其中Cas蛋白的核酸酶活性被失活,并且至少一个功能部分是标志物蛋白,例如GFP。以该方式,特异性地靶向感兴趣的核酸序列并且使用生成光信号的标志物使其可视化可以是可能的。合适的标志物可以包括例如荧光报告物蛋白,例如绿色荧光蛋白(GFP)、黄色荧光蛋白(YFP)、红色荧光蛋白(RFP)、青色荧光蛋白(CFP)或mCherry。这类荧光报告物基因提供了用于蛋白表达的可视化的合适的标志物,因为其表达可以通过荧光测量简单地且直接地测定。可选地,报告物核酸可以编码发光蛋白,诸如萤光素酶(例如萤火虫萤光素酶)。可选地,报告物基因可以是可以被用于生成光信号的显色酶,例如显色酶(诸如 β -半乳糖苷酶(LacZ)或 β -葡萄糖醛酸糖苷酶(Gus))。被用于测量表达的报告物还可以是抗原肽标签。其他报告物或标志物是本领域中已知的,并且可以适当地使用它们。

[0105] 因为标志物可以被可视化,在其中靶核酸是RNA,特别地mRNA的某些实施方案中,特别地在由标志物生成的光信号与表达产物的量成正比的情况,通过检测和定量由标志物提供的光信号来定量基因的转录活性可以是可能的。因此,在本发明的优选的实施方案中,本发明的Cas蛋白或核糖核蛋白可以被用于测定感兴趣的基因的表达产物。

[0106] 在一个方面中,本文描述的gtCas9可以被用于微生物细胞中的同源重组(HR)介导的基因组修饰方法中。这类方法涉及HR和位点定向的gtCas9活性,由此通过gtCas9活性,反选择(counter selection)发生,去除不具有由HR引入的期望的修饰的微生物。

[0107] 因此,本文提供的方法和用途允许同源重组的过程在第一步期间被支持,使得微生物基因组可以被修饰以具有期望的突变,以及在第二步期间被支持,在该第二步中未修饰的细胞可以被gtCas9核糖核酸酶复合体靶向以将DSDB引入未修饰的细胞的基因组中。由于在大多数微生物中不存在有效的非同源末端连接(NHEJ)修复机制,DSDB通常地导致细胞

死亡。因此,这些方法和用途总体上增加了具有期望的突变的微生物细胞的群体,同时消除了任何未修饰的微生物细胞。优选地,这类方法和用途被用于基本上不具有内源NHEJ修复机制的微生物中。可选地,所述方法和用途可以被应用于具有内源NHEJ修复机制的微生物。本文描述的方法和用途可以被应用于具有内源NHEJ修复机制,但其中NHEJ修复机制被有条件地降低或NHEJ活性被敲除的微生物。

[0108] 本文提供的方法和用途可以使用同源重组多核苷酸的序列,该序列具有与指导RNA的至少一个错配,使得指导RNA不再能够识别修饰的基因组。这意味着gtCas9核糖核酸酶复合体将不识别修饰的基因组。因此,不可以通过gtCas9核糖核酸酶复合体引入DSDB,并且因此修饰的细胞将存活。然而,具有未修饰的基因组的细胞仍将具有与指导RNA的实质的互补性,并且因此可以被gtCas9核糖核酸酶复合体位点特异性地裂解。

[0109] 在本发明的方法和用途的另一个方面中,其中防止gtCas9核糖核酸酶复合体起作用以裂解微生物基因组的方式与其说是以修饰或消除被该指导靶向的序列,不如说是修饰或消除被gtCas9核糖核酸酶复合体需要的PAM。PAM被修饰或消除,以使gtCas9核糖核酸酶复合体对特异性切割位点不敏感(blind)。因此,本发明的方法和用途可以包括使用同源重组多核苷酸的序列的那些方法和用途,该同源重组多核苷酸的序列不包括被gtCas9核糖核酸酶复合体识别的PAM序列。因此,不可以通过gtCas9核糖核酸酶复合体引入DSDB,并且因此HR修饰的细胞将存活。然而,未修饰的细胞仍将被gtCas9核糖核酸酶复合体及其指导所识别,并且因此被位点特异性地裂解。

[0110] 因此,本文提供了依赖于HR以修饰微生物的基因组的方法和用途。优选地,上游侧翼区(f flank)和下游侧翼区的长度各自是0.5千碱基(kb)至1.0kb。然而,使用更大或更短片段的重组也是可能的。同源重组多核苷酸还可以包含在上游和下游侧翼区域之间的多核苷酸序列。该多核苷酸序列可以例如包含将被引入微生物基因组中的修饰。

[0111] 尽管同源重组依赖于与靶区域具有实质互补性的上游和下游侧翼区,但也可以容纳错配。因此,在一些实施方案中,已知同源重组发生在与上游和下游侧翼区具有广泛同源性的DNA区段之间。在可选的实施方案中,上游和下游侧翼区具有与靶区域的完全互补性。上游和下游侧翼区的尺寸不必是相同的。然而,在一些实例中,上游和下游侧翼区的尺寸是相同的。同源重组的效率将根据侧翼区的最小片段长度的同源重组的可能性而变化。然而,即使同源重组过程是低效的,有益地,本文描述的方法将相对于未修饰的微生物细胞选择具有期望的修饰的任何微生物细胞。同源重组也允许产生包括完整基因簇的大的缺失(例如50kb或更大)。同源重组也被用于重组工程,重组工程是众所周知的允许在较小片段(45-100nt)内进行重组的方法。本文描述的方法和用途可以任选地还包含至少另一种同源重组多核苷酸;或包含编码同源重组多核苷酸的序列的多核苷酸,所述同源重组多核苷酸具有与包含微生物基因组中的靶的第二靶区域基本上互补的序列。

[0112] 在优选的实施方案中,本文描述的方法和用途使用同源重组多核苷酸,即DNA。在一些实施方案中,DNA是单链的。在其他实施方案中,DNA是双链的。在另外的实施方案中,DNA是双链的并且是质粒携带的。

[0113] 本文提供的方法和用途中的HR可以被用于从微生物基因组去除多核苷酸序列。可选地,本文提供的方法和用途中的HR可以被用于将一个或多个基因或所述基因的一个或多个片段插入微生物基因组中。作为另外的可选选择,本文提供的方法和用途中的HR可

可以被用于修饰或替换微生物基因组中的至少一个核苷酸。因此,本文提供的方法和用途可以被用于任何期望种类的基因组修饰。

[0114] 可选地,本文描述的gtCas9可以被用于微生物细胞中的HR介导的基因组修饰方法中,由此gtCas9活性在微生物细胞中引入DSDB并且可以诱导细胞HR,如对于spCas9已经显示出的(Jiang等人(2013) Nature Biotech, 31, 233-239; Xu等人(2015) Appl Environ Microbiol, 81, 4423-4431; Huang等人(2015) Acta Biochimica et Biophysica Sinica, 47, 231-243)。

[0115] 可选地,同源重组可以通过重组工程来促进,例如通过将寡核苷酸引入到表达编码RecT或 β 蛋白的基因的微生物细胞中来促进,如由Mougiakos等人((2016), Trends Biotechnol. 34:575-587)综述的。在另外的实施方案中,Cas9可以与多重自动化基因组工程(Multiplex Automated Genome Engineering) (MAGE) 进行组合,如由Ronda等人((2016), Sci. Rep. 6:19452.) 例示的。

[0116] 自始至终,本发明的Cas蛋白的参考序列可以被定义为编码氨基酸序列的核苷酸序列。例如,在SEQ ID NO:2至6中定义的基序的氨基酸序列还包括编码该氨基酸序列的所有核酸序列。

[0117] 因此,本发明还提供了分离的核酸分子,所述分离的核酸分子编码Cas蛋白,所述Cas蛋白包含:

[0118] a. 氨基酸基序EKDGKYYC[SEQ ID NO:2];和/或

[0119] b. 氨基酸基序 $X_1X_2CTX_3X_4$ [SEQ ID NO:3], 其中 X_1 独立地选自异亮氨酸、甲硫氨酸或脯氨酸, X_2 独立地选自缬氨酸、丝氨酸、天冬酰胺或异亮氨酸, X_3 独立地选自谷氨酸或赖氨酸,并且 X_4 是丙氨酸、谷氨酸或精氨酸之一;和/或

[0120] c. 氨基酸基序 X_5LXX_6IE [SEQ ID NO:4], 其中 X_5 独立地选自甲硫氨酸或苯丙氨酸,并且 X_6 独立地选自组氨酸或天冬酰胺;和/或

[0121] d. 氨基酸基序 X_7VYSX_8K [SEQ ID NO:5], 其中 X_7 是谷氨酸或异亮氨酸,并且 X_8 是色氨酸、丝氨酸或赖氨酸之一;和/或

[0122] e. 氨基酸基序 $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO:6], 其中 X_9 是丙氨酸或谷氨酸, X_{10} 是谷氨酰胺或赖氨酸, X_{11} 是精氨酸或丙氨酸, X_{12} 是天冬酰胺或丙氨酸,并且 X_{13} 是赖氨酸或丝氨酸;

[0123] 其中,当与至少一种靶向RNA分子和包含被靶向RNA分子识别的靶核酸序列的多核苷酸缔合时,Cas蛋白能够在50°C与100°C之间进行DNA结合、裂解、标记或修饰。

[0124] 在另一个方面中,本发明还提供了分离的核酸分子,所述分离的核酸分子编码成簇的规律地间隔的短回文重复(CRISPR)相关(Cas)蛋白,所述成簇的规律地间隔的短回文重复(CRISPR)相关(Cas)蛋白具有SEQ ID NO:1的氨基酸序列或与SEQ ID NO:1具有至少77%同一性的序列。

[0125] 在另一个方面中,本发明还提供了分离的核酸分子,所述分离的核酸分子还包含编码在翻译后与Cas蛋白融合的肽的至少一种核酸序列。

[0126] 在另一个方面中,本发明还提供了分离的核酸分子,其中与编码Cas蛋白的核酸分子融合的至少一种核酸序列编码选自以下的蛋白:解旋酶、核酸酶、解旋酶-核酸酶、DNA甲基化酶、组蛋白甲基化酶、乙酰基转移酶、磷酸酶、激酶、转录(共)活化物、转录阻遏物、DNA

结合蛋白、DNA结构蛋白、标志物蛋白、报告物蛋白、荧光蛋白、配体结合蛋白、信号肽、亚细胞定位序列、抗体表位或亲和纯化标签。

[0127] 表达载体

[0128] 本发明的核酸可以被分离。然而,为了核酸感测构建体(nucleic acid sensing construct)的表达可以在选定的细胞中进行,编码Cas蛋白或核糖核蛋白的多核苷酸序列将优选地提供于表达构建体中。在一些实施方案中,编码Cas蛋白或核糖核蛋白的多核苷酸将作为合适的表达载体的一部分被提供。在某些实施方案中,本发明的表达载体(具有或不具有编码在表达后将与Cas蛋白融合的氨基酸残基的核苷酸序列)还可以包含编码如上文定义的靶向RNA分子的核苷酸序列。因此,这类表达载体可以在适当的宿主中被使用以生成可以靶向期望的核苷酸序列的本发明的核糖核蛋白复合体。可选地,编码如上文定义的靶向RNA分子的核苷酸序列可以在单独的表达载体中被提供,或可选地可以通过其他手段被递送至靶细胞。

[0129] 合适的表达载体将根据受体细胞而变化,并且合适地可以掺入使表达能够在靶细胞中进行并且优选地促进高水平的表达的调控元件。这类调控序列可以能够例如在起始、准确性、速率、稳定性、下游加工、和迁移率(mobility)方面影响基因或基因产物的转录或翻译。

[0130] 这类元件可以包括,例如,强和/或组成型启动子、5'和3'UTR、转录和/或翻译增强子、转录因子或蛋白结合序列、起始位点和终止序列、核糖体结合位点、重组位点、聚腺苷酸化序列、有义或反义序列、确保转录的正确起始的序列和任选地确保宿主细胞中的转录的终止和转录物稳定的多聚A信号。调控序列可以是植物、动物、细菌、真菌或病毒来源的,并且优选地可以来源于与宿主细胞相同的生物体。清楚地,适当的调控元件将根据感兴趣的宿主细胞而变化。例如,促进在原核宿主细胞中诸如在大肠杆菌中的高水平表达的调控元件可以包括pLac、T7、P(Bla)、P(Cat)、P(Kat)、trp或tac启动子。促进在真核宿主细胞中的高水平表达的调控元件可以包括酵母中的AOX1或GAL1启动子,或CMV启动子或SV40启动子、CMV增强子、SV40增强子、单纯疱疹(Herpes simplex)病毒VIP16转录活化物或动物细胞中的珠蛋白内含子的内含物。在植物中,组成型高水平表达可以使用例如玉米(Zea mays)泛素1启动子或花椰菜花叶病毒的35S和19S启动子来获得。

[0131] 合适的调控元件可以是组成型的,由此它们在大多数环境条件或发育阶段(发育阶段特异性或诱导型)下指导表达。优选地,启动子是诱导型的,以响应于环境、化学或发育信号(cues)诸如温度、光照、化学品、干旱和其他刺激物来指导表达。合适地,可以选择允许感兴趣的蛋白在特定发育阶段或响应于细胞外或细胞内条件、信号或外部施加的刺激物而表达的启动子。例如,存在用于在大肠杆菌中使用的一系列启动子,该启动子在生长的特定阶段(例如osmY稳定期启动子)或响应于特定刺激物(例如HtpG热激启动子)提供高水平表达。

[0132] 合适的表达载体可以包含编码允许在合适的宿主细胞中和/或在特定条件下选择所述载体的可选择的标志物的另外的序列。

[0133] 本发明还包括修饰细胞中的靶核酸的方法,所述方法包括用如上文描述的任何表达载体转染、转化或转导细胞。转染、转化或转导的方法是本领域技术人员熟知的类型。在使用一种表达载体生成本发明的核糖核蛋白复合体的表达的情况下,并且当靶向RNA被直

接地添加至细胞时,那么可以使用相同的或不同的转染、转化或转导的方法。类似地,当使用一种表达载体生成本发明的核糖核蛋白复合体的表达时,并且当另一种表达载体用于通过表达原位生成靶向RNA时,那么可以使用相同的或不同的转染、转化或转导的方法。

[0134] 在其他实施方案中,编码Cas蛋白或多肽的mRNA被引入细胞中,使得Cascade复合体在细胞中被表达。将Cas蛋白复合体引导至期望的靶序列的靶向RNA也被引入细胞中,无论是与mRNA同时地、单独地或顺序地引入到细胞中,使得在细胞中形成必需的核糖核蛋白复合体。

[0135] 因此,本发明还提供了修饰(即裂解、加标签、修饰、标记或结合)靶核酸的方法,所述方法包括使核酸与如上文定义的核糖核蛋白复合体接触。

[0136] 另外,本发明还包括修饰靶核酸的方法,所述方法包括使核酸与除了如上文定义的靶向RNA分子以外的如上文定义的Cas蛋白或多肽接触。

[0137] 根据以上方法,靶核酸的修饰因此可以在体外和在无细胞的环境中进行。在无细胞环境中,靶核酸、Cas蛋白和靶向RNA分子的各自的添加可以是同时的、顺序的(根据需要以任何顺序)或单独地。因此,以下是可能的:靶核酸和靶向RNA同时地被添加至反应混合物,并且然后在稍后阶段,本发明的Cas蛋白或多肽单独地被添加。

[0138] 等同地,靶核酸的修饰可以在体内进行,即在细胞中原位进行,无论是分离的细胞还是作为多细胞组织、器官或生物体的一部分。在整个组织和器官的情况中和在生物体的情况中,该方法可以期望地在体内进行,或可选地可以通过以下进行:从整个组织、器官或生物体分离细胞,根据该方法用核糖核蛋白复合体处理细胞,并且随后将用核糖核蛋白复合体处理的细胞返回至其之前的位置或不同的位置,无论是在相同的还是不同的生物体中。

[0139] 在这些实施方案中,核糖核蛋白复合体或Cas蛋白或多肽要求递送到细胞中的适当形式。这类合适的递送系统和方法是本领域技术人员熟知的,并且包括但不限于细胞质或核显微注射。在优选的递送方式中,使用腺相关病毒(AAV);该递送系统在人类中不引起疾病,并且已经在欧洲被批准用于临床使用。

[0140] 因此,本发明提供了修饰靶核酸的方法,所述方法包括使核酸与以下接触:

[0141] a. 如上文定义的核糖核蛋白复合体;或

[0142] b. 如上文定义的蛋白或蛋白复合体和如上文定义的RNA分子。

[0143] 在其他方面中,本发明提供了修饰细胞中的靶核酸的方法,所述方法包括用包含编码如上文定义的核糖核蛋白复合体的核苷酸序列的表达载体转化、转染或转导细胞;或可选地用包含编码如上文定义的蛋白或蛋白复合体的核苷酸序列的表达载体和包含编码如上文定义的靶向RNA分子的核苷酸序列的另外的表达载体转化、转染或转导细胞。

[0144] 在另外的方面中,本发明提供了修饰细胞中的靶核酸的方法,所述方法包括用包含编码如上文定义的蛋白或蛋白复合体的核苷酸序列的表达载体转化、转染或转导细胞,并且然后将如上文定义的靶向RNA分子递送到细胞中。

[0145] 在其中指导(即靶向)RNA(gRNA)分子和Cas蛋白或多肽被单独地而非作为核糖核蛋白复合体的一部分被提供的实施方案中,gRNA分子要求递送到细胞中的适当形式,无论是与Cas蛋白或蛋白复合体同时地、单独地还是顺序地递送。将RNA引入到细胞中的这类形式是本领域技术人员熟知的,并且可以包括通过常规转染方法的体外或离体递送。可以各

自使用物理方法,诸如显微注射和电穿孔、以及钙共沉淀、和商购可得的阳离子聚合物和脂质、和细胞穿透肽、细胞穿透(基因枪(biolistic))颗粒。例如,病毒(特别地优选的是AAV)可以被用作递送媒介物,无论是递送至细胞质和/或细胞核,例如通过本发明的Cas蛋白复合体或本发明的核糖核蛋白复合体与病毒颗粒的(可逆的)融合。

[0146] 在另一个方面中,本发明提供了修饰靶核酸的方法,其中至少一个功能部分是标志物蛋白或报告物蛋白,并且所述标志物蛋白或报告物蛋白与靶核酸缔合;优选地其中标志物是荧光蛋白,例如绿色荧光蛋白(GFP)。

[0147] 在上文提及的修饰靶核酸的方法中,功能部分可以是标志物,并且标志物与靶核酸缔合;优选地其中标志物是蛋白;任选地荧光蛋白,例如绿色荧光蛋白(GFP)、黄色荧光蛋白(YFP)、红色荧光蛋白(RFP)或mCherry。无论是在体外、离体还是体内,然后本发明的方法可以被用于直接地可视化核酸分子中的靶基因座,优选地呈更高级结构的形式,诸如超螺旋的质粒或染色体、或单链靶核酸诸如mRNA。靶基因座的直接可视化可以使用电子显微术或荧光显微术。然而,将理解的是,在本发明的方法的情况中,其他种类的标记物可以被用作标志物,包括可以是小分子的有机染料分子、放射性标记物和自旋标记物。

[0148] 在其中靶核酸是dsDNA的用于修饰靶核酸的本发明的方法中,功能部分可以是核酸酶或解旋酶-核酸酶,并且修饰优选地是在期望的基因座处的单链断裂或双链断裂。以该方式,独特的序列特异性DNA切割可以通过使用与核糖核蛋白复合体融合的合适的功能部分来工程化。最终的核糖核蛋白复合体的RNA组分的选定的序列提供了用于功能部分的作用所需的序列特异性。

[0149] 因此,本发明还提供了在细胞中在期望的基因座处非同源末端连接dsDNA分子以从dsDNA分子去除至少一部分核苷酸序列;任选地以敲除一个基因或更多个基因的功能的方法,其中该方法包括使用如上文描述的任何修饰靶核酸的方法制备双链断裂。

[0150] 本发明还提供了将核酸同源重组到细胞中期望的基因座处的dsDNA分子中以修饰现有的核苷酸序列或插入期望的核苷酸序列的方法,其中该方法包括使用如上文描述的任何修饰靶核酸的方法在期望基因座处制备双链断裂。

[0151] 因此,本发明还提供了修饰生物体中的基因表达的方法,该方法包括根据上文描述的任何方法修饰靶核酸序列,其中核酸是dsDNA,并且功能部分选自DNA修饰酶(例如甲基化酶或乙酰基转移酶)、转录活化物或转录阻遏物。

[0152] 本发明另外地提供了修饰生物体中的基因表达的方法,该方法包括根据上文描述的任何方法修饰靶核酸序列,其中核酸是mRNA,并且功能部分是核糖核酸酶;任选地选自内切核酸酶、3'外切核酸酶或5'外切核酸酶。

[0153] 靶核酸可以是DNA、RNA或合成的核酸。优选地,靶核酸是DNA;优选地dsDNA。

[0154] 然而,靶核酸可以是RNA;优选地mRNA。可选地,因此,本发明还提供了修饰靶核酸的方法,其中靶核酸是RNA。

[0155] 在另一个方面中,本发明提供了修饰靶核酸的方法,其中核酸是dsDNA,至少一个功能部分是核酸酶或解旋酶-核酸酶,并且修饰是在期望的基因座处的单链断裂或双链断裂。

[0156] 在另一个方面中,本发明提供了修饰细胞中的靶核酸的方法,其中修饰引起在期望的基因座处的基因表达的沉默;并且其中所述方法包括以下步骤:

[0157] a. 制备dsDNA分子中的双链断裂;和

[0158] b. 通过非同源末端连接 (NHEJ) 修复细胞中的dsDNA分子。

[0159] 在另一个方面中,本发明提供了修饰细胞中的靶核酸的方法;其中现有的核苷酸序列被修饰或缺失和/或期望的核苷酸序列被插入在期望的位置处,其中所述方法包括以下步骤:

[0160] a. 制备在期望的基因座处的双链断裂;和

[0161] b. 通过同源重组修复细胞中的dsDNA分子。

[0162] 在另一个方面中,本发明提供了修饰细胞中的基因表达的方法,所述方法包括如上文描述地修饰靶核酸序列;其中核酸是dsDNA,并且功能部分选自DNA修饰酶(例如甲基化酶或乙酰基转移酶)、转录活化物或转录阻遏物。

[0163] 在另一个方面中,本发明提供了修饰细胞中的基因表达的方法,所述方法包括如上文描述地修饰靶核酸序列,其中核酸是mRNA,并且功能部分是核糖核酸酶;任选地选自内切核酸酶、3' 外切核酸酶或5' 外切核酸酶。

[0164] 在另一个方面中,本发明提供了如上文描述的修饰靶核酸的方法,其中该方法在45°C和100°C之间的温度进行。优选地,该方法在50°C或在50°C以上的温度进行。更优选地,该方法在55°C和80°C之间的温度进行。最佳地,该方法在60°C和65°C之间的温度进行。可选地,该方法可以在20°C和45°C之间的温度进行。更优选地,在30°C和45°C之间的温度进行。甚至更优选地,在37°C和45°C之间的温度进行。

[0165] 在上文描述的修饰靶核酸的任何方法中,细胞可以是原核细胞,或可选地可以是真核细胞。

[0166] 宿主细胞

[0167] 有益地,本发明具有宽的适用性,并且本发明的宿主细胞可以来源于可以被培养的任何遗传上易处理的生物体。因此,本发明提供了通过如上文描述的方法转化的宿主细胞。本发明提供了经转化的细胞,该经转化的细胞具有在双链靶多核苷酸中的靶核酸序列,所述细胞包含如本文提供的Cas蛋白或多肽和如本文提供的至少一种靶向RNA分子、和包含编码所述Cas蛋白和所述靶向RNA分子中的至少一种的核酸的表达载体。

[0168] 适当的宿主细胞可以是原核细胞或真核细胞。特别地,可以选择通常地被使用的宿主细胞用于根据本发明的用途,所述通常被使用的宿主细胞包括遗传上可获得的并且可以被培养的原核细胞或真核细胞,例如原核细胞、真菌细胞、植物细胞和动物细胞,包括人类细胞(但不是胚胎干细胞)。优选地,宿主细胞将选自原核细胞、真菌细胞、植物细胞、原生生物细胞(protozoan cell)或动物细胞。用于根据本发明的用途的优选的宿主细胞通常来源于通常表现出高生长速率、易于被培养和/或转化、显示出短的世代时间的物种、已经建立了与它们相关的遗传资源的物种或已经被选择、修饰或合成以用于在特定条件下的异源蛋白的最佳表达的物种。在其中感兴趣的蛋白最终被用于特定的工业、农业、化学或治疗情况中的本发明的优选实施方案中,可以基于期望的特定条件或感兴趣的蛋白将被部署在其中的细胞情况来选择适当的宿主细胞。优选地,宿主细胞将是原核细胞。在优选的实施方案中,宿主细胞是细菌细胞。宿主细胞可以是例如大肠杆菌(*Escherichia coli*; *E. coli*)细胞。优选地,宿主细胞将是嗜热细菌的细胞。

[0169] 本文描述的本发明的方法和用途可以被用于修饰细菌细胞的基因组。在特定实施

方案中,细菌是嗜热细菌,优选地细菌选自:嗜酸硫杆菌属(*Acidithiobacillus*)的种,包括喜温嗜酸硫杆菌(*Acidithiobacillus caldus*);气芽孢杆菌属(*Aeribacillus*)的种,包括苍白气芽孢杆菌(*Aeribacillus pallidus*);脂环酸芽孢杆菌属(*Alicyclobacillus*)的种,包括酸热脂环酸芽孢杆菌(*Alicyclobacillus acidocaldarius*)、酸土脂环酸芽孢杆菌(*Alicyclobacillus acidoterrestris*)、环庚基脂环酸杆菌I(*Alicyclobacillus cycloheptanicus*I)、*Alicyclobacillus hesperidum*;厌氧芽孢杆菌属(*Anoxybacillus*)的种,包括热解蛋白厌氧芽孢杆菌(*Anoxybacillus caldiproteolyticus*)、黄嗜热厌氧芽孢杆菌(*Anoxybacillus flavithermus*)、*Anoxybacillus rupiensis*、*Anoxybacillus tepidamans*;芽孢杆菌属(*Bacillus*)的种,包括热容芽孢杆菌(*Bacillus caldolyticus*)、*Bacillus caldotenax*、*Bacillus caldovelox*、凝结芽孢杆菌(*Bacillus coagulans*)、克劳氏芽孢杆菌(*Bacillus clausii*)、地衣芽孢杆菌(*Bacillus licheniformis*)、甲醇芽孢杆菌(*Bacillus methanolicus*)、史氏芽孢杆菌(*Bacillus smithii*),包括史氏芽孢杆菌 ET138、枯草芽孢杆菌(*Bacillus subtilis*)、热堆肥芽孢杆菌(*Bacillus thermocopriae*)、*Bacillus thermolactis*、热噬淀粉芽孢杆菌(*Bacillus thermoamylovorans*)、*Bacillus thermoleovorans*、热杆菌属(*Caldibacillus*)的种,包括*Caldibacillus debilis*;热解纤维素菌属(*Caldicellulosiruptor*)的种,包括*Caldicellulosiruptor bescii*、热液热解纤维素菌(*Caldicellulosiruptor hydrothermalis*)、*Caldicellulosiruptor kristjanssonii*、*Caldicellulosiruptor kronotskyensis*、*Caldicellulosiruptor lactoaceticus*、*Caldicellulosiruptor obsidiansis*、*Caldicellulosiruptor owensensis*、*Caldicellulosiruptor saccharolyticus*、梭菌属(*Clostridium*)的种,包括*Clostridium clariflavum*、*Clostridium straminisolvens*、*Clostridium tepidiprofundum*、*Clostridium thermobutyricum*、热纤梭菌(*Clostridium thermocellum*)、热产琥珀酸梭菌(*Clostridium thermosuccinogenes*)、*Clostridium thermopalmarum*;奇球菌属(*Deinococcus*)的种,包括*Deinococcus cellulosilyticus*、*Deinococcus deserti*、*Deinococcus geothermalis*、*Deinococcus murrayi*、耐辐射奇球菌(*Deinococcus radiodurans*);*Defluviitalea*的种,包括*Defluviitalea phaphyphila*、脱硫肠状菌属(*Desulfotomaculum*)的种,包括*Desulfotomaculum carboxydivorans*、致黑脱硫肠状菌(*Desulfotomaculum nigrificans*)、*Desulfotomaculum salinum*、*Desulfotomaculum solfataricum*;硫酸还原菌属(*Desulfurella*)的种,包括*Desulfurella acetivorans*;除硫杆菌属(*Desulfurobacterium*)的种,包括*Desulfurobacterium thermolithotrophum*;地芽孢杆菌属(*Geobacillus*)的种,包括*Geobacillus icigianus*、*Geobacillus caldooxysilyticus*、*Geobacillus jurassicus*、*Geobacillus galactosidasius*、嗜热地芽孢杆菌(*Geobacillus kaustophilus*)、*Geobacillus lituanicus*、嗜热脂肪地芽孢杆菌(*Geobacillus stearothermophilus*)、地下地芽孢杆菌(*Geobacillus subterraneus*)、*Geobacillus thermantarcticus*、*Geobacillus thermocatenulatus*、热脱氮地芽孢杆菌、*Geobacillus thermoglucosidans*、*Geobacillus thermoleovorans*、*Geobacillus toebii*、*Geobacillus uzenensis*、*Geobacillus vulcanii*、*Geobacillus zalihiae*;产氢杆菌属(*Hydrogenobacter*)的种,包括嗜热产氢杆菌(*Hydrogenobacter thermophiles*);*Hydrogenobaculum*的种,包括*Hydrogenobaculum*

acidophilum; Ignavibacterium 的种, 包括 Ignavibacterium album; 乳杆菌属 (Lactobacillus) 的种, 包括保加利亚乳杆菌 (Lactobacillus bulgaricus)、德氏乳杆菌 (Lactobacillus delbrueckii)、Lactobacillus ingluviei、耐热乳杆菌 (Lactobacillus thermotolerans); 海栖热菌属 (Marinithermus) 的种, 包括热液海栖热菌属 (Marinithermus hydrothermalis); 穆尔氏菌属 (Moorella), 包括热醋穆尔氏菌 (Moorella thermoacetica); 海洋栖热菌属 (Oceanithermus) 的种, 包括 Oceanithermus desulfurans、Oceanithermus profundus; 类芽孢杆菌属 (Paenibacillus) 的种, 包括类芽孢杆菌属 J2 种、Paenibacillus marinum、Paenibacillus thermoaerophilus; Persephonella 的种, 包括 Persephonella guaymasensis、Persephonella hydrogeniphila、Persephonella marina; 嗜热盐菌属 (Rhodothermus) 的种, 包括海洋嗜热盐菌 (Rhodothermus marinus)、Rhodothermus obamensis、Rhodothermus profundus; 硫化杆菌属 (Sulfobacillus), 包括嗜酸硫化杆菌 (Sulfobacillus acidophilus); Sulfurihydrogenibium 的种, 包括 Sulfurihydrogenibium azorense、Sulfurihydrogenibium kristjanssonii、Sulfurihydrogenibium rodmanii、Sulfurihydrogenibium yellowstonense、共生杆菌 (Symbiobacterium) 的种, 包括嗜热共生杆菌 (Symbiobacterium thermophilum)、Symbiobacterium toebii; 热厌氧杆菌属 (Thermoanaerobacter) 的种, 包括 Thermoanaerobacter brockii、产乙醇热厌氧杆菌 (Thermoanaerobacter ethanolicus)、Thermoanaerobacter italicus、Thermoanaerobacter kivui、Thermoanaerobacter marianensis、Thermoanaerobacter mathranii、Thermoanaerobacter pseudoethanolicus、威吉利热厌氧杆菌 (Thermoanaerobacter wiegelii); 嗜热厌氧杆菌属 (Thermoanaerobacterium) 的种, 包括耐酸嗜热厌氧杆菌 (Thermoanaerobacterium aciditolerans)、Thermoanaerobacterium aotearoense、产乙醇嗜热厌氧杆菌 (Thermoanaerobacterium ethanolicus)、Thermoanaerobacterium pseudoethanolicus、解糖热厌氧杆菌 (Thermoanaerobacterium saccharolyticum)、Thermoanaerobacterium thermosaccharolyticum、Thermoanaerobacterium xylanolyticum、热杆菌属 (Thermobacillus) 的种, 包括 Thermobacillus composti、Thermobacillus xylanilyticus; 发状菌属 (Thermocrinis) 的种, 包括 Thermocrinis albus、Thermocrinis ruber; Thermodesulfatator 的种, 包括 Thermodesulfatator atlanticus、Thermodesulfatator autotrophicus、Thermodesulfatator indicus; 热脱硫杆菌属 (Thermodesulfobacterium) 的种, 包括 Thermodesulfobacterium commune、Thermodesulfobacterium hydrogeniphilum; 热脱硫菌属 (Thermodesulfobium) 的种, 包括 Thermodesulfobium narugense; 热脱硫弧菌属 (Thermodesulfovibrio) 的种, 包括 Thermodesulfovibrio aggregans、Thermodesulfovibrio thiophilus、Thermodesulfovibrio yellowstonii; 热腔菌属 (Thermosipho) 的种, 包括非洲栖热腔菌 (Thermosipho africanus)、Thermosipho atlanticus、Thermosipho melanesiensis; 热孢菌属 (Thermotoga) 的种, 包括海栖热孢菌 (Thermotoga maritima)、Thermotoga neopolitana、热孢菌属 RQ7 种 (Thermotoga sp. RQ7); 热弧菌属 (Thermovibrio) 的种, 包括 Thermovibrio ammonificans、Thermovibrio ruber; Thermovirga 的种, 包括 Thermovirga lienii 和栖热菌属 (Thermus) 的种, 包括水生栖热菌 (Thermus aquaticus)、Thermus

caldophilus、黄栖热菌 (*Thermus flavus*)、*Thermus scotoductus*、嗜热栖热菌 (*Thermus thermophilus*)；那不勒斯硫杆菌 (*Thiobacillus neapolitanus*)。

[0170] 在另一个方面中,本文描述的方法或用途可以被用于修饰是嗜温的细菌。在优选的实施方案中,细菌选自:嗜酸硫杆菌属的种,包括喜温嗜酸硫杆菌;放线杆菌属 (*Actinobacillus*) 的种,包括产琥珀酸放线杆菌 (*Actinobacillus succinogenes*);厌氧螺菌属 (*Anaerobiospirillum*) 的种,包括产琥珀酸厌氧螺菌 (*Anaerobiospirillum succiniciproducens*);芽孢杆菌属的种,包括嗜碱芽孢杆菌 (*Bacillus alcalophilus*)、解淀粉芽孢杆菌 (*Bacillus amyloliquefaciens*)、环状芽孢杆菌 (*Bacillus circulans*)、蜡样芽孢杆菌 (*Bacillus cereus*)、克劳氏芽孢杆菌 (*Bacillus clausii*)、坚硬芽孢杆菌 (*Bacillus firmus*)、耐盐芽孢杆菌 (*Bacillus halodurans*)、灿烂芽孢杆菌 (*Bacillus lautus*)、迟缓芽孢杆菌 (*Bacillus lentus*)、地衣芽孢杆菌 (*Bacillus licheniformis*)、巨大芽孢杆菌 (*Bacillus megaterium*)、短小芽孢杆菌 (*Bacillus pumilus*)、枯草芽孢杆菌、苏云金芽孢杆菌 (*Bacillus thuringiensis*); *Basfia* 的种,包括 *Basfia succiniciproducens*;短芽孢杆菌属 (*Brevibacillus*) 的种,包括短短芽孢杆菌 (*Brevibacillus brevis*);侧孢短芽孢杆菌 (*Brevibacillus laterosporus*);梭菌属的种,包括丙酮丁醇梭菌 (*Clostridium acetobutylicum*)、*Clostridium autoethanogenum*、*Clostridium beijerinckii*、*Clostridium carboxidivorans*、解纤维素梭菌 (*Clostridium cellulolyticum*)、*Clostridium ljungdahlii*、巴斯德梭菌 (*Clostridium pasteurianum*)、产气荚膜梭菌 (*Clostridium perfringens*)、*Clostridium ragsdalei*、*Clostridium saccharobutylicum*、*Clostridium saccharoperbutylacetonium*、棒杆菌属 (*Corynebacterium*) 的种,包括谷氨酸棒杆菌 (*Corynebacterium glutamicum*);脱亚硫酸菌属 (*Desulfitobacterium*) 的种,包括脱卤脱亚硫酸菌 (*Desulfitobacterium dehalogenans*)、*Desulfitobacterium hafniense*;脱硫肠状菌属的种,包括醋酸氧化脱硫肠状菌 (*Desulfotomaculum acetoxidans*)、*Desulfotomaculum gibsoniae*、还原脱硫肠状菌 (*Desulfotomaculum reducens*)、瘤胃脱硫肠状菌 (*Desulfotomaculum ruminis*)、肠杆菌属 (*Enterobacter*) 的种,包括阿氏肠杆菌 (*Enterobacter asburiae*);肠球菌属 (*Enterococcus*) 的种,包括粪肠球菌 (*Enterococcus faecalis*);埃希氏菌属 (*Escherichia*) 的种,包括大肠杆菌 (*Escherichia coli*);乳杆菌属的种,包括嗜酸乳杆菌 (*Lactobacillus acidophilus*)、嗜淀粉乳杆菌 (*Lactobacillus amylophilus*)、食淀粉乳杆菌 (*Lactobacillus amylovorus*)、动物乳杆菌 (*Lactobacillus animalis*)、*Lactobacillus arizonensis*、*Lactobacillus bavaricus*、短乳杆菌 (*Lactobacillus brevis*)、布氏乳杆菌 (*Lactobacillus buchneri*)、保加利亚乳杆菌、干酪乳杆菌 (*Lactobacillus casei*)、棒状乳杆菌 (*Lactobacillus coryniformis*)、卷曲乳杆菌 (*Lactobacillus crispatus*)、弯曲乳杆菌 (*Lactobacillus curvatus*)、德氏乳杆菌、发酵乳杆菌 (*Lactobacillus fermentum*)、格氏乳杆菌 (*Lactobacillus gasseri*)、瑞士乳杆菌 (*Lactobacillus helveticus*)、约氏乳杆菌 (*Lactobacillus johnsonii*)、戊糖乳杆菌 (*Lactobacillus pentosus*)、植物乳杆菌 (*Lactobacillus plantarum*)、罗伊氏乳杆菌 (*Lactobacillus reuteri*)、鼠李糖乳杆菌 (*Lactobacillus rhamnosus*)、沙克乳杆菌 (*Lactobacillus sakei*)、唾液乳杆菌 (*Lactobacillus salivarius*)、*Lactobacillus*

sanfriscensis;曼氏杆菌属(Mannheimia)的种,包括产琥珀酸曼氏杆菌(Mannheimia succiniciproducens);类芽孢杆菌属的种,包括蜂房类芽孢杆菌(Paenibacillus alvei)、北京类芽孢杆菌(Paenibacillus beijingensis)、Paenibacillus borealis、Paenibacillus dauci、Paenibacillus durus、Paenibacillus graminis、幼虫类芽孢杆菌(Paenibacillus larvae)、缓病类芽孢杆菌(Paenibacillus lentimorbus)、浸麻类芽孢杆菌(Paenibacillus macerans)、胶质类芽孢杆菌(Paenibacillus mucilaginosus)、Paenibacillus odorifer、多粘类芽孢杆菌(Paenibacillus polymyxa)、星孢类芽孢杆菌(Paenibacillus stellifer)、土地类芽孢杆菌(Paenibacillus terrae)、乌鲁木齐类芽孢杆菌(Paenibacillus wulumuqiensis);片球菌属(Pediococcus)的种,包括乳酸片球菌(Pediococcus acidilactici)、Pediococcus clausenii、耐乙醇片球菌(Pediococcus ethanolidurans)、戊糖片球菌(Pediococcus pentosaceus);鼠伤寒沙门氏菌(Salmonella typhimurium);芽孢乳杆菌属(Sporolactobacillus)的种,包括菊糖芽孢乳杆菌(Sporolactobacillus inulinus)、左旋乳酸芽孢乳杆菌(Sporolactobacillus laevolacticus);金黄色葡萄球菌(Staphylococcus aureus);链球菌属(Streptococcus)的种,包括无乳链球菌(Streptococcus agalactiae)、牛链球菌(Streptococcus bovis)、类马链球菌(Streptococcus equisimilis)、粪链球菌(Streptococcus faecalis)、变异链球菌(Streptococcus mutans)、口腔链球菌(Streptococcus oralis)、肺炎链球菌(Streptococcus pneumonia)、酿脓链球菌、唾液链球菌(Streptococcus salivarius)、嗜热链球菌(Streptococcus thermophilus)、远缘链球菌(Streptococcus sobrinus)、乳房链球菌(Streptococcus uberis);链霉菌属(Streptomyces)的种,包括不产色链霉菌(Streptomyces achromogenes)、阿维链霉菌(Streptomyces avermitilis)、天蓝色链霉菌(Streptomyces coelicolor)、灰色链霉菌(Streptomyces griseus)、变铅青链霉菌(Streptomyces lividans)、微小链霉菌(Streptomyces parvulus)、委内瑞拉链霉菌(Streptomyces venezuelae)、酒红链霉菌(Streptomyces vinaceus);四联球菌属(Tetragenococcus)的种,包括嗜盐四联球菌(Tetragenococcus halophilus)和发酵单胞菌属(Zymomonas)的种,包括运动发酵单胞菌(Zymomonas mobilis)。

[0171] 在另外的方面中,本文定义的方法或用途可以被用于修饰酵母或真菌的基因组。在特定实施方案中,真菌物种是嗜温的,优选地真菌选自:曲霉属(Aspergillus)的种,包括,但不限于,构巢曲霉(Aspergillus nidulans)、黑曲霉(Aspergillus niger)、土曲霉(Aspergillus terreus)、米曲霉(Aspergillus oryzae)和土曲霉,更优选地曲霉属的种是构巢曲霉或黑曲霉。可选地,嗜温真菌的种可以是念珠菌属(Candida)的种。

[0172] 本发明还涉及使用如本文定义的方法修饰嗜热的酵母或真菌的种,优选地该真菌或酵母选自:曲霉属的种,包括烟曲霉(Aspergillus fumigatus)、构巢曲霉、土曲霉、杂色曲霉(Aspergillus versicolor);Canariomyces的种,包括Canariomyces thermophile;毛壳菌属 Chaetomium)的种,包括Chaetomium mesopotamicum、嗜热毛壳菌(Chaetomium thermophilum);念珠菌属的种,包括Candida bovina、Candida slooffii、Candida thermophila、热带念珠菌(Candida tropicalis)、克柔氏念珠菌(Candida krusei)(=东方伊萨酵母(Issatchenkia orientalis));尾孢壳菌属(Cercophora)的种,包括Cercophora coronate、Cercophora septentrionalis;Coonemeria的种,包括Coonemeria

aegyptiaca;棒囊孢壳菌属(*Corynascus*)的种,包括嗜热棒囊孢壳菌(*Corynascus thermophiles*);地霉属(*Geotrichum*)的种,包括白地霉(*Geotrichum candidum*);克鲁维酵母属(*Kluyveromyces*)的种,包括脆壁克鲁维酵母(*Kluyveromyces fragilis*)、马克斯克鲁维酵母(*Kluyveromyces marxianus*);畸枝霉属(*Malbranchea*)的种,包括樟绒枝霉(*Malbranchea cinnamomea*)、*Malbranchea sulfurea*; *Melanocarpus*的种,包括 *Melanocarpus albomyces*; *Myceliophthora*的种,包括 *Myceliophthora fergusii*、*Myceliophthora thermophila*; *Mycothermus*的种,包括 *Mycothermus thermophiles*(=嗜热柱霉属(*Scytalidium thermophilum*)/嗜热圆酵母(*Torula thermophila*)); *Myriococcum*的种,包括 *Myriococcum thermophilum*;拟青霉属(*Paecilomyces*)的种,包括嗜热拟青霉(*Paecilomyces thermophila*); *Remersonia*的种,包括 *Remersonia thermophila*;根毛霉属(*Rhizomucor*)的种,包括微小根毛霉(*Rhizomucor pusillus*)、牛根毛霉(*Rhizomucor tauricus*);酵母菌属(*Saccharomyces*)的种,包括酿酒酵母(*Saccharomyces cerevisiae*)、裂殖酵母属(*Schizosaccharomyces*)的种,包括粟酒裂殖酵母(*Schizosaccharomyces pombe*)、柱霉属(*Scytalidium*)的种,包括嗜热柱霉属(*Scytalidium thermophilum*); *Sordaria*的种,包括 *Sordaria thermophila*;子囊菌属(*Thermoascus*)的种,包括耐热子囊菌(*Thermoascus aurantiacus*)、*Thermoascus thermophiles*; *Thermomucor*的种,包括 *Thermomucor indicae-seudaticae*和嗜热丝孢菌属(*Thermomyces*)的种,包括 *Thermomyces ibadanensis*、绵毛嗜热丝孢菌(*Thermomyces lanuginosus*)。

[0173] 在上文提及的列表中,以粗体标识的微生物已经被发现是特别合适的/适用于本发明的用途中。

[0174] 本发明的一些优选实施方案包括一种或更多种嗜热微生物,所述嗜热微生物选自:嗜热的杆菌(*bacilli*),包括气芽孢杆菌属、脂环酸芽孢杆菌属、厌氧芽孢杆菌属、芽孢杆菌属、地芽孢杆菌属、类芽孢杆菌属的种;嗜热的梭菌(*clostridia*),包括厌氧杆菌属(*Anaerobacter*)、厌氧杆菌属(*Anaerobacterium*)、热解纤维素菌属、梭菌属(*Clostridium*)、穆尔氏菌属、嗜热厌氧杆菌属(*Thermoanaerobacter*)、嗜热厌氧杆菌属(*Thermoanaerobacterium*)、栖热分枝菌属(*Thermobrachium*)、*Thermohalobacter*的种或一种或更多种嗜热的乳杆菌的种和嗜温细菌(所述嗜温细菌选自芽孢杆菌属的种、大肠杆菌和乳杆菌的种)。

[0175] 附图简述

[0176] 现在将参考具体实施方案并且参考附图详细地描述本发明,在附图中:

[0177] 图1示出了Cas9蛋白序列的邻接树。包括基于pBLAST或PSI-BLAST与T12菌株具有40%以上的序列相似性的所有序列,以及目前良好表征的序列(酿脓链球菌(*S. pyogenes*)、嗜热链球菌(*S. thermophiles*)和内氏放线菌(*A. naeslundii*)),以及当这些低于40%同一性时所有目前鉴定的嗜热序列。对于所有的嗜热序列,在菌株名称后指示出与T12的同一性百分比。在物种名称前指示出基因标识符(gi)编号。图例:实心圆形:嗜热(最佳温度为60℃以上)Cas9序列,实心方形:耐热(最佳温度<50℃)Cas9序列,空心三角形:来自嗜温来源的目前最常用于基因组编辑目的的Cas9序列;无标志:嗜温Cas9。在节点处的值代表1000个重复的自展值(bootstrap values);比例尺代表估计的氨基酸取代/位点。

[0178] 图2示出了Cas9基因序列的邻接树。在基因水平的同一性非常差;使用来自与被用于蛋白比对的那些生物体相同的生物体的序列进行基因比对。在物种名称前指示出基因标识符(gi)编号。图例:实心圆形:嗜热(最佳温度为60℃以上)Cas9序列,实心方形:耐热(最佳温度<50℃)Cas9序列,空心三角形:来自嗜温来源的目前最常用于基因组编辑目的的Cas9序列;无标志:嗜温Cas9。在节点处的值代表1000个重复的自展值。

[0179] 图3示出了gtCas9(SEQ ID NO:1)(II-C型)与良好表征的II-C型(内氏放线菌/‘ana’;SEQ ID NO:8)和II-A型(酿脓链球菌/‘pyo’;SEQ ID NO:9和嗜热链球菌)Cas9序列的蛋白序列比对。重要的活性位点残基非常保守,并且用黑色箭头指示出。如对于Ana-Cas9和Pyo-Cas9描述的蛋白结构域(Jinek等人,2014,Science 343:1247997)用阴影框和类似地着色的字母指示出。已经确定了对于酿脓链球菌II-A型系统的PAM识别结构域,但对于任何II-C型系统未确定PAM识别结构域,并且因此仅在酿脓链球菌序列中指示出PAM识别结构域。

[0180] 图4示出了内氏放线菌Cas9(Cas9-Ana)的蛋白结构(Jinek等人,2014)。gtCas9属于相同的II-C型CRISPR系统并且可以鉴定活性位点残基。

[0181] 图5示出了crRNA指导的互补dsDNA的靶向的比较。碱基配对用虚线指示出。RNA以黑色描绘,DNA用灰色描绘。在crRNA间隔区和靶前间区之间的碱基配对用加粗的黑色虚线指示出,在DNA链之间和在RNA链之间的碱基配对用加粗的灰色虚线指示出。指示出crRNA的5'末端。应注意,I型中的PAM(小白色框)位于靶链(前间区)的下游,而在II型中它则位于置换链上的另一端处。同样地,种子(与靶DNA链开始碱基配对处的指导的预测序列,并且在此处不允许错配)位于PAM附近,并且因此在I型和II型中不同(Van der Oost,2014,同上)。上图示出了大肠杆菌的I型Cascade系统的示意图。crRNA具有内部间隔区(灰色框,允许靶识别的31-32nt),侧翼为8nt 5'手柄和由茎环结构(发夹)组成的29nt 3'手柄(Jore2011同上)。下图示出了酿脓链球菌的II型Cas9系统的示意图。crRNA与tracrRNA碱基配对,允许通过RNA酶III(相对黑色三角形)处理。另外地,crRNA的5'末端被RNA酶(黑色三角形)修剪(trimmed),通常地产生20nt间隔区。应注意,合成的环可以被引入以连接crRNA和tracrRNA,产生单指导RNA(sgRNA)(Jinek等人,2012同上)。

[0182] 图6示出了热脱氮地芽孢杆菌T12 IIc型CRISPR系统的序列的比对。

[0183] 图7示出了获得的六个单击中(hit),以提供了对于gtCas9的计算机模拟PAM预测。

[0184] 图8示出了组合图7中例示的比对的结果的weblogo。使用weblogo.berkeley.edu生成该weblogo。

[0185] 图9示出了在60℃用纯化的gtCas9靶向质粒的体外裂解测定的结果。质粒包括PAM序列的特定8个核苷酸长的序列变体。

[0186] 图10示出了使用具有CCCCCAA[SEQ ID NO:11]PAM序列的靶向质粒研究gtCas9浓度的作用的体外测定的结果。

[0187] 图11示出了在一定范围的温度使用靶向质粒的体外测定的结果,该靶向质粒具有CCCCCAA[SEQ ID NO:11]PAM序列。

[0188] 图12示出了使用gtCas9和8nt PAM序列对史氏芽孢杆菌ET138细胞进行体内基因组编辑的结果,即筛选平板上的史氏芽孢杆菌ET138细胞的菌落的生长或不存在,如在实施例9中解释的。在图12中用箭头指示出菌落。

[0189] 图13示出了对于其中缺失pyrF基因的菌落的PCR筛选的结果。用构建体3 (阴性对照) 转化史氏芽孢杆菌ET138细胞后生成菌落。筛选了15个菌落,但没有一个显示出缺失基因型-2.1kb条带大小,而是全部显示出野生型-2.9kb条带大小,如在实施例9中解释的。

[0190] 图14示出了对于其中缺失pyrF基因的菌落的PCR筛选的结果。用构建体1 (PAM序列ATCCCCAA[SEQ ID NO:21]) 转化史氏芽孢杆菌ET138细胞后生成菌落。筛选了20个菌落,并且一个菌落显示出缺失基因型-2.1kb条带大小,而剩余的菌落显示出野生型-2.9kb条带大小和缺失基因型-2.1kb条带大小二者,如在实施例9中解释的。没有观察到仅野生型的基因型。

[0191] 以下是根据本发明使用的Cas蛋白的多核苷酸和氨基酸序列。

[0192] [SEQ ID NO:1]热脱氮地芽孢杆菌T12 Cas9蛋白AA序列

MKYKIGLDIGITSIGWAVINLDIPRIEDLGVRIFDRAENPKTGESLALPRRLARSARRR
LRRRKHRLERIRRLFVREGILTKEELNKLFEKKHEIDVWQLRVEALDRKLNNDELARI
LLHLAKRRGFRSNRKSERTNKENSTMLKHIEENQSILSSYRTVAEMVVKDPKFS
LHKRNKEDNYTNTVARDDLEREIKLIFAKQREYGNIVCTEAFEHEYISIWASQRPFASK
DDIEKKVGFCCTFEPKEKRAPKATYTFQSFTVWEHINKLRLVSPGGIRALTDDERRLIY
KQAFHKNKITFHDVRTLLNLPDDTRFKGLLYDRNTTLKENEKVRFLGAYHKIRKAI
DSVYGKGAAKSFRPIDFDTFGYALTMFKDDTDIRSILRNEYEQNGKRMEENLADKVY
DEELIEELLNLSFSKFGHLSLKALRNILPYMEQGEVYSTACERAGYTFTGPKKKQKT
VLLPNIPPIANPVVMRALTQARKVNVNIIKKYGSPVSIHIELARELSQSFDERRKMQK
EQEGNRKKNETAIRQLVEYGLTLNPTGLDIVKFKLWSEQNGKCAYSLQPIEIERLLE
PGYTEVDHVIPYSRSLDDSYTNKVLVLTKENREKGNRTPAEYLGLGSERWQQFETF
VLTNKQFSKKKRDRLLRLHYDENEENEFKNRNLNDTRYISRFLANFIREHLKFADSD
DKQKVYTVNGRITAHLSRWNFNKNREESNLHHAVDAAIVACTTPSDIARVTAFYQ
RREQNKELSKKTDPPQFPQPWPHFADELQARLSKNPKESIKALNLGNYDNEKLESL
QPVFVSRMPKRSITGAAHQETLRRYIGIDERSGKIQTVVKKKLSEIQLDKTGHFPMY
GKESDPRTYEAIRQRLLHNNDPKKAQFEPLYKPKNGELGPIIRTIKIIDTTNQVIPL
NDGKTVAYNSNIVRVDFEKGDKYYCVPIYTIDMMKGILPNKAIEPNKPYSEWKEMT
EDYTFRFSLYPNDLIRIEFPREKTIKTAVGEEIKIKDLFAYYQTIDSSNGGLSLVSHDN
NFSLRSIGSRSLKRFKEYQVDVLGNIYKVRGEKRVGVASSSHSKAGETIRPL*

[0195] [SEQ ID NO:7]热脱氮地芽孢杆菌T12 Cas9 DNA序列

ATGAAGTATAAAATCGGTCTTGATATCGGCATTACGTCTATCGGTTGGGCTGTC
ATTAATTTGGACATTCCTCGCATCGAAGATTTAGGTGTCCGCATTTTTGACAGAG
CGGAAAACCCGAAAACCGGGGAGTCACTAGCTCTTCCACGTCGCCTCGCCCGC
TCCGCCCCGACGTCGTCTGCGGCGTCGAAACATCGACTGGAGCGCATTCGCC
GCCTGTTCTGTCGCGAAGGAATTTTAACGAAGGAAGAGCTGAACAAGCTGTTT
GAAAAAAGCACGAAATCGACGTCTGGCAGCTTCGTGTTGAAGCACTGGATCG
AAACTAAATAACGATGAATTAGCCCGCATCCTTCTTCATCTGGCTAAACGGCG
TGGATTTAGATCCAACCGCAAGAGTGAGCGCACCAACAAAGAAAACAGTACGAT
GCTCAAACATATTGAAGAAAACCAATCCATTCTTTCAAGTTACCGAACGGTTGCA
GAAATGGTTGTCAAGGATCCGAAATTTTCCCTGCACAAGCGTAATAAAGAGGAT
AATTACACCAACACTGTTGCCCGCGACGATCTTGAACGGGAAATCAAACGATT
TTCGCCAAACAGCGCGAATATGGGAACATCGTTTGCACAGAAGCATTGTAACAC
GAGTATATTTCCATTTGGGCATCGCAACGCCCTTTTGCTTCTAAGGATGATATC
GAGAAAAAAGTCGGTTTCTGTACGTTTGAGCCTAAAGAAAAACGCGCGCCAAAA
GCAACATACACATTCCAGTCCTTCACCGTCTGGGAACATATTAACAACTTCGT
CTTGTCTCCCCGGGAGGCATCCGGGCACTAACCGATGATGAACGTCGTCTTAT
ATACAAGCAAGCATTTTATAAAAAATAAAATCACCTTCCATGATGTTTGAACATTG
CTTAACCTTGCCTGACGACACCCGTTTTAAAGGTCTTTTATATGACCGAAACACCA
CGCTGAAGGAAAATGAGAAAGTTTCGCTTCCTTGAACCTCGGCGCCTATCATAAAA
TACGGAAAGCGATCGACAGCGTCTATGGCAAAGGAGCAGCAAAATCATTTTCGT
CCGATTGATTTTGATACATTTGGCTACGCATTAACGATGTTTAAAGACGACACCG
ACATTGCGAGTTACTTGCGAAACGAATACGAACAAAATGGAAAACGAATGGAAA
ATCTAGCGGATAAAGTCTATGATGAAGAATTGATTGAAGAACTTTTAACTTATC
GTTTTCTAAGTTTGGTCATCTATCCCTTAAAGCGCTTCGCAACATCCTTCCATAT

[0196]

ATGGAACAAGGCGAAGTCTACTCAACCGCTTGTGAACGAGCAGGATATACATTT
ACAGGGCCAAAGAAAAACAGAAAACGGTATTGCTGCCGAACATTCCGCCGAT
CGCCAATCCGGTCGTCATGCGCGCACTGACACAGGCACGCAAAGTGGTCAATG
CCATTATCAAAAAGTACGGGCTCACCGGTCTCCATCCATATCGAACTGGCCCCG
GAACTATCACAATCCTTTGATGAACGACGTAAAATGCAGAAAGAACAGGAAGGA
AACCGAAAGAAAAACGAAACTGCCATTCCGCCAACTTGTTGAATATGGGCTGACG
CTCAATCCAACCTGGGCTTGACATTGTGAAATTCAACTATGGAGCGAACAAAAAC
GGAAAATGTGCCTATTCACTCCAACCGATCGAAATCGAGCGGTTGCTCGAACCA
GGCTATACAGAAGTCGACCATGTGATTCCATACAGCCGAAGCTTGACGATAG
CTATACCAATAAAGTTCTTGTGTTGACAAAGGAGAACCGTGAAAAAGGAAACCG
CACCCCAGCTGAATATTTAGGATTAGGCTCAGAACGTTGGCAACAGTTCGAGAC
GTTTGTCTTGACAAATAAGCAGTTTTTCGAAAAAGAAGCGGGATCGACTCCTTCG
GCTTCATTACGATGAAACGAAGAAAATGAGTTTAAAAATCGTAATCTAAATGAT
ACCCGTTATATCTCACGCTTCTTGGCTAACTTTATTCCGGAACATCTCAAATTCG
CCGACAGCGATGACAAACAAAAAGTATACACGGTCAACGGCCGATTACCGCC
CATTTACGCAGCCGTTGGAATTTTAACAAAAACCGGGAAGAATCGAATTTGCAT
CATGCCGTCGATGCTGCCATCGTCGCCTGCACAACGCCGAGCGATATCGCCCCG
AGTCACCGCCTTCTATCAACGGCGCGAACAAAAACAAAGAACTGTCCAAAAAGAC
GGATCCGCAGTTTCCGCAGCCTTGCGCCGCACTTTGCTGATGAACTGCAGGCGC
GTTTATCAAAAAATCCAAAGGAGAGTATAAAAGCTCTCAATCTTGAAATTATGA
TAACGAGAAACTCGAATCGTTGCAGCCGGTTTTTGTCTCCCGAATGCCGAAGC
GGAGCATAACAGGAGCGGCTCATCAAGAAACATTGCGGCGTTATATCGGCATC
GACGAACGGAGCGGAAAAATACAGACGGTCGTCAAAAAGAACTATCCGAGAT
CCAACCTGGATAAAACAGGTCATTTCCCAATGTACGGGAAAGAAAGCGATCCAAG
GACATATGAAGCCATTCGCCAACGGTTGCTTGAACATAACAATGACCCAAAAAA
GGCGTTTCAAGAGCCTCTGTATAAACCGAAGAAGAACGGAGAACTAGGTCCTAT
CATCCGAACAATCAAAATCATCGATACGACAAATCAAGTTATTCCGCTCAACGAT
GGCAAAACAGTCGCCTACAACAGCAACATCGTGCGGGTCGACGTCTTTGAGAA
AGATGGCAAATATTATTGTGTCCCTATCTATACAATAGATATGATGAAAGGGATC
TTGCCAAACAAGGCGATCGAGCCGAACAAACCGTACTCTGAGTGGAAGGAAAT
GACGGAGGACTATACATTCCGATTCAGTCTATACCCAAATGATCTTATCCGTATC
GAATTTCCCCGAGAAAAACAATAAAGACTGCTGTGGGGGAAGAAATCAAAATT
AAGGATCTGTTCGCCTATTATCAAACCATCGACTCCTCCAATGGAGGGTTAAGT
TTGGTTAGCCATGATAACAACCTTTTCGCTCCGCAGCATCGGTTCAAGAACCCTC
AAACGATTGAGAAATACCAAGTAGATGTGCTAGGCAACATCTACAAAGTGAGA
GGGGAAAAGAGAGTTGGGGTGGCGTCATCTTCTATTGAAAGCCGGGGAAAC
TATCCGTCCGTTATAA

[0197] 详细描述

[0198] 实施例1:热脱氮地芽孢杆菌的分离

[0199] 在搜索能够在厌氧条件下降解木质纤维素基质的嗜热菌的 ± 500 个分离株的文库期间出乎意料地发现了热脱氮地芽孢杆菌。首先建立了 ± 500 个分离株的文库,在通过在纤维素和木聚糖上的分离进行几轮选择后,文库被缩小至110个分离株。该110个分离株的文库仅由地芽孢杆菌属分离株组成,其中热脱氮地芽孢杆菌代表了文库的79%。

[0200] 分离的热脱氮地芽孢杆菌菌株已经被命名为“T12”。来自热脱氮地芽孢杆菌T12的Cas9蛋白已经被命名为“gtCas9”。

[0201] 实施例2:定义热脱氮地芽孢杆菌中的Cas9的基本共有序列

[0202] 进行以下数据库搜索和比对:

[0203] 在内部BLAST服务器上进行pBLAST和InBLAST,其中使用热脱氮地芽孢杆菌T12的蛋白或基因序列作为查询序列。该数据库最近一次更新为2014年5月,并且因此不包含最近添加的地芽孢杆菌属基因组,但没有使用通常的在线BLAST以防止T12序列的公开。在BLAST搜索中发现的大于40%的序列同一性被包括在图1中。

[0204] 为了包括更近期的序列数据,在NCBI网站上使用地芽孢杆菌属MAS1的序列(与gtCas9最密切地相关)进行PSI-BLAST(Johnson等人,2008Nucleic Acids Res.36(网络服务器期号):W5-9)。进行连续两轮的PSI-BLAST,其中仅使用满足以下标准的序列用于下一轮:在第一轮中的最小序列覆盖率为96%,并且在第二轮和第三轮中的最小序列覆盖率为97%,最小同一性为40%,每个物种仅一个菌株。

[0205] 将从PSI-BLAST所得的序列以及在PSI-BLAST中未出现的、来自内部服务器pBLAST的与T12具有多于40%同一性的序列与目前良好表征的嗜温序列和所有目前鉴定的嗜热序列一起比对,此外如果这些序列更加疏远地相关,则从其构建邻接树(参见图1)。在Mega6中使用ClustalW进行比对,之后使用邻接方法构建树,并且使用1000个重复进行自展分析。

[0206] 当使用地芽孢杆菌属的种MAS1作为查询序列进行BLASTn时,仅地芽孢杆菌属的种JF8 Cas9被鉴定出具有88%同一性,指示出在基因水平的同源性非常小。图2是ClustalW比对的Cas9基因序列的邻接树。

[0207] 通过使用具有默认设置的BL0SUM62在CloneManager中比对热脱氮地芽孢杆菌T12、内氏放线菌和酿脓链球菌的蛋白序列来进一步分析它们的蛋白结构域同源性(参见图3)。

[0208] 实施例3:鉴定对于CAS9的功能至关重要的核心氨基酸基序和赋予嗜热Cas9核酸酶的热稳定性的那些核心氨基酸基序

[0209] 在图1中提供了以上描述的比对的蛋白序列的同一性百分比。gtCas9属于II-C型。研究最充分并且最近结晶的结构的II-C系统来自内氏放线菌(Jinek等人,2014,Science 343:1247997)。该蛋白序列显示出与gtCas9仅20%同一性,但可以被用于估计高度保守的残基。在分析中还包括了两个良好表征的II-A型系统(酿脓链球菌和嗜热链球菌)(Jinek等人,2014,Science 343:1247997;Nishimasu等人,2014,Cell 156:935-949)。在图3中示出了这四个蛋白序列的比对;图4示出了如对于内氏放线菌(‘Ana-Cas9’)确定的蛋白结构(Jinek等人,2014,Science 343:1247997)。来自T12的Cas9(gtCas9)和来自内氏放线菌的Cas9的长度是高度类似的(内氏放线菌为1101aa,gtCas9为1082aa),并且预期gtCas9具有类似的蛋白结构,但这仍待确定,因为与cas9-Ana的总体序列同一性仅为20%。由Jinek等人(Jinek等人,2014,Science 343:1247997)描述的在来自内氏放线菌和酿脓链球菌的

Cas9中的所有活性位点残基可以在gtCas9中被鉴定出(参见图3)。已经确定了对于酿脓链球菌II-A型系统的PAM结合结构域,但对于任何II-C型系统未确定PAM结合结构域,并且因此仅在酿脓链球菌序列中指示出PAM结合结构域。此外,PAM识别位点不仅在CRISPR系统之间而且在包含相同的系统的物种之间变化很大。

[0210] 实施例4:热脱氮地芽孢杆菌gtCas9的PAM序列的确定

[0211] 已经建立了原核CRISPR系统作为适应性免疫系统为其宿主服务(Jinek等人,2012,Science 337:816-821),并且可以被用于快速和有效的遗传工程(Mali等人,2013,Nat Methods 10:957-963.)。

[0212] Cas9蛋白作为用于II型CRISPR系统的序列特异性核酸酶发挥功能(Makarova等人,2011,Nat Rev Micro 9:467-477)。由与重复区域连接的“间隔区”(靶)组成的小crRNA分子是CRISPR基因座的转录和加工产物。“间隔区”天然起源于噬菌体的基因组和移动遗传元件,但它们也可以被设计为在遗传工程过程期间靶向特定核苷酸序列(Bikard等人,2013,Nucleic Acids Research 41:7429-7437)。Cas9使用crRNA分子作为用于鉴定其DNA靶的指导。间隔区区域与靶向的用于裂解的DNA区域,“前间区”相同(Brouns等人,2012,Science 337:808-809)。由Cas9对靶的识别需要紧邻前间区的PAM(前间区相邻基序)(Jinek等人,2012,Science 337:816-821)。

[0213] 为了对II型系统进行体外或体内PAM确定研究,有必要进行计算机模拟预测该系统的CRISPR阵列,即表达tracrRNA的模块。使用CRISPR阵列用于crRNA模块的鉴定。表达tracrRNA的序列位于Cas9侧翼的500bp的窗口中,或在Cas基因和CRISPR基因座之间(Chylinski,K.,等人(2014)Classification and evolution of type IICRISPR-Cas systems.Nucleic Acids Res.42,6091-6105)。tracrRNA应该由与CRISPR阵列的直接重复具有高水平的互补性的5'-序列、随后是不少于两个茎环结构的预测的结构和Rho非依赖性转录终止信号组成(Ran,F.A.,等人(2015)In vivo genome editing using Staphylococcus aureus Cas9.Nature 520,186-191)。然后,crRNA和tracrRNA分子可以被用于设计嵌合sgRNA模块.sgRNA的5'-末端由截短的20nt长的间隔区组成,随后是CRISPR阵列的16-20nt长的截短的重叠。该重叠之后是对应的截短的反向重叠和tracrRNA模块的茎环.sgRNA的重叠部分和反向重叠部分通常地由GAAA接头连接(Karvelis,T.,等人(2015)Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements.Genome Biol.16,253)。

[0214] 使用T12染色体的反义链来转录热脱氮地芽孢杆菌T12 IIc型CRISPR系统的cas基因(cas9之后是cas1和cas2基因)。cas2基因之后是100bp长的DNA片段,该DNA片段在转录后形成具有多个环的RNA结构。该结构显然地充当转录终止子。

[0215] 具有11个重复和10个间隔区序列的CRISPR阵列位于转录终止序列的上游,并且阵列的前导区位于阵列的5'末端处。转录为tracrRNA的DNA基因座预期是在cas9基因的下游。cas9基因的直接下游的325bp长的序列与来自CRISPR阵列的36bp长的重复的比对揭示,在tracrRNA基因座中存在36bp长的序列,其几乎与该重复相同(如在图6中示出的)。该结果使我们得出结论,tracrRNA基因座的转录方向应该与CRISPR阵列的转录方向相反。因此,tracrRNA的5'末端将与crRNA的3'末端互补,引起Cas9所需要的双RNA分子的形成。

[0216] 实施例5:用随机化的PAM的靶生成

[0217] 使用热脱氮地芽孢杆菌T12基因组DNA作为模板通过PCR扩增来自热脱氮地芽孢杆菌T12菌株的CRISPRII基因座的两个不同的间隔区。使用两对简并引物用于每一个间隔区的扩增：

[0218] 首先，使用引起在“前间区”片段的上游引入六个随机核苷酸的一对，导致产生具有随机化的PAM序列的前间区的池。

[0219] 其次，使用引起在“前间区”片段的下游引入六个随机核苷酸的一对，导致产生具有随机化的PAM序列的前间区的池。

[0220] 将产生的片段连接至pNW33n载体，产生“前间区”构建体的4个池，每一个池具有6个核苷酸长的PAM的所有可能的4096种不同的组合。使用组装的DNA用于转化热脱氮地芽孢杆菌T12细胞。将细胞铺板在氯霉素选择培养基上，并且将汇集来自每一个前间区池的多于 2×10^6 个细胞。从池提取质粒DNA，对靶区域进行PCR扩增，并且将产物送出用于深度测序。具有最少读段的PAM将被认为是有活性的，并且将仅用包含具有这些PAM的间隔区的pNW33n构建体重复该过程。热脱氮地芽孢杆菌T12的降低的转化效率将证实PAM的活性。

[0221] 实施例6：对于gtCas9的PAM序列的体外确定

[0222] pRham:cas9_{gt}载体的构建

[0223] 使用BG6927和BG6928引物，从热脱氮地芽孢杆菌T12基因组PCR扩增cas9_{gt}基因，并且将其与pRham C-His Kan载体 (Lucigen) 组合在一种混合物中。根据所提供的方案，使用该混合物用于转化E.cloni热感受态细胞。将来自转化混合物的100μl混合物铺板在LB+50卡那霉素平板上，用于在37℃过夜生长。从形成的E.cloni::pRham:cas9_{gt}单菌落中随机地选择3个单菌落并且接种在包含50μg/ml卡那霉素的10ml LB培养基中。通过添加无菌甘油至来自每种培养物的1ml培养物直到最终浓度为20% (v/v)，从培养物制备甘油贮存物 (glycerol stock)。在-80℃储存甘油贮存物。根据“GeneJET Plasmid Miniprep Kit” (Thermoscientific) 方案，将来自每种培养物的剩余的9ml培养物用于质粒分离。将质粒送出用于cas9_{gt}的序列验证，并且一个质粒被验证包含具有正确序列的基因。对应的培养物被进一步用于gtCas9的异源表达和纯化。

[0224] gtCas9在E.cloni::pRham:cas9_{gt}载体中的异源表达

[0225] E.cloni::pRham:cas9_{gt}预培养物用对应的甘油贮存物接种10ml LB+50卡那霉素后来制备。在37℃和180rpm过夜生长后，使用来自预培养物的2ml预培养物接种200ml的LB+50卡那霉素培养基。将E.cloni::pRham:cas9_{gt}培养物在37℃、180rpm培养，直至OD₆₀₀为0.7。然后，通过添加L-鼠李糖至0.2% w/v的最终浓度来诱导gtCas9表达。允许表达进行持续8h，之后以4700rpm、4℃离心培养物持续10分钟以收获细胞。弃去培养基，并且将沉淀的细胞储存在-20℃或根据以下方案用于无细胞提取物 (CFE) 的制备：

[0226] 1. 将沉淀物重悬浮在20ml声处理缓冲液 (20mM磷酸钠缓冲液 (pH=7.5)、100mM NaCl、5mM MgCl₂、5% (v/v) 甘油、1mM DTT) 中

[0227] 2. 通过声处理破碎1ml的细胞 (8个30秒的脉冲，在脉冲之间在冰上冷却持续20秒)

[0228] 3. 以35000g、4℃离心持续15分钟，以使不可溶的部分沉淀

[0229] 4. 取出上清液并且将其储存在4℃或冰上

[0230] 对于gtCas9的PAM文库靶向sgRNA模块的设计和构建

[0231] 在进行计算机模拟确定在热脱氮地芽孢杆菌T12菌株的基因组中的表达tracrRNA

的DNA模块后(见以上实施例4),设计了表达单指导(sg)RNA的DNA模块,所述DNA模块将CRISPR/Cas9系统的crRNA模块和tracrRNA模块组合在单个分子中。在sgRNA的5'末端处的间隔区被设计为与质粒文库的前间区互补,并且该模块被设置在T7启动子的转录控制下。pT7_sgRNA DNA模块由Baseclear合成,并且被接纳在pUC57载体中,形成pUC57:pT7_sgRNA载体。用载体转化DH5 α 感受态大肠杆菌细胞(NEB),并且将转化混合物铺板在包含100 μ g/ml氨苄青霉素的LB琼脂平板上。将平板在37 $^{\circ}$ C孵育过夜。将三个形成的单菌落接种在包含100 μ g/ml氨苄青霉素的10ml LB培养基中。通过添加无菌甘油至来自每种培养物的1ml培养物直到最终浓度为20%(v/v),从培养物制备甘油贮存物。在-80 $^{\circ}$ C储存甘油贮存物。根据“GeneJET Plasmid Miniprep Kit”(Thermoscientific)方案,将来自每种培养物的剩余的9ml培养物用于质粒分离。分离的质粒用作扩增pT7_sgRNA模块的PCR模板。使用引物BG6574和BG6575获得218bp长的pT7_sgRNA DNA模块(其中前18bp对应于pT7)。将完整的PCR混合物在1.5%琼脂糖凝胶上运行。根据“ZymocleanTM Gel DNA Recovery Kit”方案,切除和纯化具有期望尺寸的条带。

[0232] 使用“HiScribeTM T7 High Yield RNA Synthesis Kit”(NEB)进行体外转录(IVT)。使用纯化的pT7_sgRNA DNA模块作为模板。将IVT混合物与等体积的RNA加样染料(NEB)混合,并且在70 $^{\circ}$ C加热持续15分钟以破坏二级结构。将热处理的IVT混合物在变性尿素-PAGE上运行,并将所得聚丙烯酰胺凝胶在包含10 μ l的SYBR Gold(Invitrogen)的100ml 0.5 \times TBE缓冲液中洗涤(embaptised)持续10分钟,用于染色目的。切下在期望尺寸(200nt)处的条带,并且根据以下RNA纯化方案纯化sgRNA:

[0233] 1.用解剖刀切割RNA凝胶片段,并且添加1ml的RNA洗脱缓冲液,置于室温过夜。

[0234] 2.将330 μ l等分试样分到新的1.5ml管中。

[0235] 3.添加3倍体积(990 μ l)的预冷(-20 $^{\circ}$ C)100% EtOH。

[0236] 4.在-20 $^{\circ}$ C孵育持续60分钟。

[0237] 5.在室温,在微量离心机中以13000rpm离心持续20分钟。

[0238] 6.去除EtOH,用1ml 70% EtOH洗涤沉淀物。

[0239] 7.在室温,在微量离心机中以13000rpm离心持续5分钟。

[0240] 8.去除990 μ l的上清液。

[0241] 9.将剩余EtOH在55 $^{\circ}$ C在热混合器中蒸发持续15分钟至20分钟。

[0242] 10.将沉淀物重悬浮在20 μ l MQ中,储存在-20 $^{\circ}$ C。

[0243] 7nt长的PAM文库的设计与构建、和文库的线性化

[0244] PAM文库的设计和构建基于pNW33n载体。将20bp长的前间区引入至载体,其3'侧侧翼为7个简并核苷酸长的序列;简并序列用作PAM,并且当前间区的侧翼为正确的PAM时,那么它就可以被装载sgRNA的Cas9识别为靶并且被裂解。根据以下方案制备PAM文库:

[0245] 1.通过使单链DNA寡聚物1(BG6494)和2(BG6495)退火来制备SpPAM双链DNA插入物

[0246] I. 10 μ l 10 \times NEBuffer 2.1

[0247] II. 1 μ l 50 μ M寡聚物1(\sim 1.125 μ g)

[0248] III. 1 μ l 50 μ M寡聚物2(\sim 1.125 μ g)

[0249] IV. 85 μ l MQ

[0250] V.将混合物在94 $^{\circ}$ C孵育持续5min,并且以0.03 $^{\circ}$ C/秒的速率冷却至37 $^{\circ}$ C

[0251] 2.添加1 μ l Klenow 3'->5' exo-聚合酶(NEB)至每个退火的寡聚物混合物并且然后添加2.5 μ l的10 μ M dNTP。在37 $^{\circ}$ C孵育持续1h,并且然后在75 $^{\circ}$ C孵育持续20min。

[0252] 3.添加2 μ l的HF-BamHI和2 μ l的BspHI限制性酶至46 μ l的退火混合物。在37 $^{\circ}$ C孵育持续1h。该过程将导致SpPAMbb插入物具有粘性末端。使用Zymo DNA清洗和浓缩试剂盒(Zymo Research)清洗创建的插入物。

[0253] 4.用HF-BamHI和BspHI(NEB)消化pNW33n,并且使用Zymo DNA清洗和浓缩试剂盒(Zymo Research)纯化具有粘性末端的3,400bp长的线性

[0254] pNW33nbb片段。

[0255] 5.根据所提供的方案,使用NEB T4连接酶将50ng的pNW33nBB与11ng的SPPAMbb插入物进行连接。使用Zymo DNA清洗和浓缩试剂盒(Zymo Research)纯化连接混合物。

[0256] 6.转化DH10b电感受态细胞(200 μ l细胞,用500ng的DNA)。在SOC培养基(在800 μ l SOC中200 μ l细胞)中回收细胞持续1小时,并且然后用回收的细胞接种50ml的LB+12.5 μ g/ml 氯霉素。在37 $^{\circ}$ C和180rpm,孵育培养物过夜。

[0257] 7.使用JetStar 2.0maxiprep试剂盒(GENOMED)从培养物分离质粒DNA。

[0258] 8.根据所提供的方案,使用SapI(NEB)限制性处理用于使分离的质粒线性化。

[0259] PAM确定反应的设计和执行

[0260] 设置以下裂解反应,用于将gtCas9诱导的dsDNA断裂引入至PAM文库成员,所述PAM文库成员包含在靶向的前间区的3'末端的下游的正确PAM:

[0261] 1.每反应2.5 μ g的E.cloni::pRham:cas9_{gt} CFE

[0262] 2.sgRNA至30nM最终浓度

[0263] 3.每反应200ng的线性化的PAM文库

[0264] 4.2 μ l的裂解缓冲液(100mM磷酸钠缓冲液(pH=7.5)、500mM NaCl、25mM MgCl₂、25% (v/v)甘油、5mM DTT)

[0265] 5.MQ水,直到20 μ l最终体积

[0266] 反应在60 $^{\circ}$ C孵育持续1h,并且在加入4 μ l的6 \times 凝胶加样染料(NEB)后终止。然后将反应混合物加样至1%琼脂糖凝胶。凝胶在100V经历1h和15min长的电泳,并且然后在包含10 μ l的SYBR Gold染料(ThermoFisher)的100ml 0.5 \times TAE缓冲液中孵育持续30min。根据所提供的方案,在用蓝光使DNA条带可视化后,从凝胶上切下与成功地裂解的和包含PAM的DNA片段对应的条带,并且使用“ZymocleanTMGel DNA Recovery Kit”进行凝胶纯化。

[0267] 对包含PAM的gtCas9裂解的DNA片段加标签用于测序

[0268] Cas9诱导的DNA断裂通常被引入在前间区的第3个和第4个核苷酸之间,靠近PAM序列。因此,不可能设计可以PCR扩增裂解的DNA片段的包含PAM的部分的一对引物,以进一步测序和确定PAM序列。为了该目标,使用了5步的过程:

[0269] 第1步:用Taq聚合酶接A尾(A-Tailing)

[0270] 接A尾是使用Taq聚合酶将非模板的腺嘌呤添加至双链DNA分子的平的3'末端的过程

[0271] 反应组分:

[0272] • gtCas9-裂解的和包含PAM的DNA片段-200ng

[0273] • 10 \times ThermoPol[®]缓冲液(NEB)-5 μ l

- [0274] • 1mM dATP-10 μ l
- [0275] • Taq DNA聚合酶 (NEB) -0.2 μ l
- [0276] • H₂O-直到50 μ l最终反应体积
- [0277] • 孵育时间-20min
- [0278] • 孵育温度-72℃
- [0279] 第2步:测序衔接子的构建
- [0280] 将两个互补的短ssDNA寡核苷酸磷酸化和退火,以形成用于来自第1步的DNA片段的PAM近端位点的测序衔接子。其中一种寡核苷酸在其3'末端处具有另外的胸腺嘧啶,以促进将衔接子连接至接A尾的片段。
- [0281] 衔接子寡核苷酸磷酸化(对于每个寡聚物,单独的磷酸化反应)
- [0282] • 100 μ M寡核苷酸原液(stock) -2 μ L
- [0283] • 10 \times T4 DNA连接酶缓冲液 (NEB) -2 μ L
- [0284] • 无菌MQ水-15 μ L
- [0285] • T4多核苷酸激酶 (NEB) -1 μ L
- [0286] • 孵育时间-60min
- [0287] • 孵育温度-37℃
- [0288] • T4 PNK失活-65℃持续20min
- [0289] 磷酸化的寡核苷酸的退火
- [0290] • 寡核苷酸1-来自对应的磷酸化混合物的5 μ L混合物
- [0291] • 寡核苷酸2-来自对应的磷酸化混合物的5 μ L混合物
- [0292] • 无菌MQ水-90 μ L
- [0293] • 将磷酸化的寡聚物在95℃孵育持续3分钟。在室温,缓慢地冷却反应持续~30min至1hr
- [0294] 第3步:gtCas9裂解的接A尾的片段与测序衔接子的连接
- [0295] 根据以下方案,使第1步和第2步的产物进行连接:
- [0296] • 10 \times T4 DNA连接酶缓冲液-2 μ L
- [0297] • 第1步的产物-50ng
- [0298] • 第2步的产物-4ng
- [0299] • T4 DNA连接酶-1 μ l
- [0300] • 无菌MQ水-至20 μ l
- [0301] • 孵育时间-10min
- [0302] • 孵育温度-20-25℃
- [0303] • 在65℃加热失活持续10min
- [0304] 第4步:150个核苷酸长的包含PAM片段的PCR扩增
- [0305] 使用来自第3步的连接混合物的5 μ l混合物作为用于使用Q5 DNA聚合酶 (NEB) 的PCR扩增的模板。具有来自第2步的胸腺嘧啶延伸的寡核苷酸被用作正向引物,并且反向引物被设计为在PAM序列下游的150个核苷酸处退火。
- [0306] 使用非gtCas9处理的PAM文库DNA作为模板来扩增相同的序列。两种PCR产物被凝胶纯化,并且送出用于Illumina HiSeq 2500配对末端测序 (paired-end sequencing)

(Baseclear)。

[0307] 测序结果的分析和候选PAM序列的确定

[0308] 在分析测序结果后,构建了以下频率矩阵。这些矩阵描绘了在gtCas9消化的和非消化的文库的每个PAM位置处的每个核苷酸的相对丰度:

非消化的	位置1	位置2	位置3	位置4	位置5	位置6	位置7
A	19.22	20.83	19.12	24.43	24.59	21.75	18.22
C	34.75	30	31.9	30.54	25.96	27.9	27.17
T	19.16	22.19	25.34	21.28	26.09	26	21.56
G	26.87	26.98	23.64	23.75	23.36	24.35	33.05

[0309]

消化的	位置1	位置2	位置3	位置4	位置5	位置6	位置7
A	10.63	18.65	14.6	14.49	3.36	8.66	27.54
C	66.22	49.59	56.82	60.35	92.4	62.26	34.94
T	8.09	11.21	19.12	12.15	2.35	14.66	5.58
G	15.05	20.54	9.45	13.01	1.89	14.43	31.94

[0310] 这些结果指示对于在第5个PAM位置具有胞嘧啶的靶的明显偏好和对于在前4个PAM位置具有胞嘧啶的靶的偏好。

[0311] 实施例7:对于gtCas9的计算机模拟PAM预测

[0312] 如果在基因组数据库中足够的前间区序列是可得的,PAM的计算机模拟预测是可能的。gtCas9 PAM的计算机模拟预测以以下开始:通过与在基因组数据库诸如GenBank中的序列比较,鉴定来自CRISPR阵列的间隔区在热脱氮地芽孢杆菌T12菌株的基因组中的击中。使用“CRISPR查找器”(http://crispr.u-psud.fr/Server/)工具以鉴定在T12中的候选CRISPR基因座。然后,将鉴定的CRISPR基因座输出加载到“CRISPR靶”(http://bioanalysis.otago.ac.nz/CRISPRTarget/crispr_analysis.html)工具中,该工具搜索选定的数据库并且提供具有匹配前间区的输出。然后,筛选这些前间区序列以确定独特的击中和与间隔区的互补性一例如,在种子序列中的错配被认为可能是假阳性击中并且被排除在进一步分析之外。与前间区序列和(整合的)质粒具有同一性的击中证明了获得的击中是真阳性。总的来说,该过程产生了6个单击中(图7)。随后,使用WebLogo(http://weblogo.berkeley.edu/logo.cgi)(CrooksGE、HonG、ChandoniaJM、Brenner SE WebLogo:A sequence logo generator,Genome Research,14:1188-1190,(2004))工具(图8),对剩余的独特前间区击中的侧翼区域(对于II型gtCas核酸酶的3')进行比对和比较以确定共有序列。

[0313] 计算机模拟结果与体外PAM鉴定实验结果(参见实施例6)相当,其中存在PAM序列的第5个残基的身份偏向为胞嘧啶。

[0314] 实施例8:对于gtCas9的8个核苷酸长的PAM序列的确定

[0315] 来自实施例8的计算机模拟数据表明gtCas9在第8个位置处具有对于腺苷的某种偏好,因此进行了进一步的PAM确定实验,其中还测试了PAM序列的第8个位置。这与嗜温侧孢短芽孢杆菌SSP360D4(Karvelis等人,2015年)Cas9 PAM序列的表征是一致的,发现嗜温侧孢短芽孢杆菌SSP360D4 Cas9 PAM序列在前间区的3'末端处的第5个和第8个位置之间延伸。

[0316] 用gtCas9试验了PAM的特定8个核苷酸长的序列变体:

[0317] 1) CNCCCCAC[SEQ ID NO:17]、

[0318] 2) CCCCCCAG[SEQ ID NO:18]、

[0319] 3) CCCCCCAA[SEQ ID NO:11]、

[0320] 4) CCCCCCAT[SEQ ID NO:19]、

[0321] 5) CCCCCCAC[SEQ ID NO:20]、

[0322] 6) NNNNTNNC(阴性对照PAM)。

[0323] 在60℃进行体外裂解测定后,如之前一样地(参见实施例6)用纯化的gtCas9和相同的sgRNA靶向这些(非线性化的)质粒,当CCCCCAA[SEQ ID NO:11]序列被用作PAM时,观察到增加的gtCas9裂解活性(图9)。然而,对于所有测试的PAM序列,裂解活性明确地是可检测的,即使对于阴性对照PAM序列,观察到微弱的裂解条带。不希望受限于特定的理论,可能的是,高gtCas9浓度的使用促成通过阴性对照观察到的裂解。通常地已经观察到,在体外测定中高Cas9浓度导致Cas9诱导的DNA裂解,而无严格的PAM需求。

[0324] 通常地,已知Cas9浓度会影响Cas9诱导的DNA裂解的效率(Cas9浓度越高,引起越高的Cas9活性)。这也是当使用具有CCCCCAA[SEQ ID NO:11]PAM序列的靶向的质粒和不同的gtCas9浓度进行体外测定时观察到的(图10)。

[0325] 如以上描述的,具有CCCCCAA[SEQ ID NO:11]PAM序列的靶向的质粒用于体外测定,在38℃和78℃之间的宽的温度范围中进行所述体外测定(图11)。出乎意料地,gtCas9在所有温度是有活性的,在40.1℃和64.9℃之间显示出最高的活性。

[0326] 因此,来自地芽孢杆菌属的种的Cas9的最佳温度范围比迄今已经被表征的Cas9蛋白的最佳温度范围高得多。类似地,其中来自地芽孢杆菌属的种的Cas9保持核酸酶活性的范围的上限比已知的Cas9蛋白的范围的上限高得多。较高的最佳温度和功能范围在高温度的遗传工程中提供了显著的优点,并且因此在编辑嗜热生物体的基因组中提供了显著的优点,其在升高的温度进行的一系列工业、农业和制药过程中具有效用。

[0327] 实施例9:用gtCas9和8个核苷酸长度的PAM序列对史氏芽孢杆菌ET138进行体内基因组编辑

[0328] 为了证实8个核苷酸的PAM也被gtCas9体内识别,设计了在55℃使在史氏芽孢杆菌ET138基因组中的pyrF基因缺失的实验。

[0329] 该方法依赖于提供同源重组模板构建体,在同源重组模板构建体中与靶(pyrF)基因上游和下游互补的区域被提供至史氏芽孢杆菌ET 138细胞。模板的引入允许同源重组的过程被用于将同源重组模板(不具有pyrF基因)引入到基因组中,使得它也代替在细胞的基因组中的WT pyrF基因。

[0330] 在同源重组构建体中包含gtCas9和sgRNA可以被用于将双链DNA断裂(DSDB)引入到包含WT pyrF的细菌基因组中。通常地,在细菌基因组中的DSDB引起细胞死亡。因此,识别WT pyrF中的序列的sgRNA可以引起DSDB和仅包含WT pyrF的细胞的死亡。DSDB的引入还取决于合适的PAM序列,该合适的PAM序列位于被gtCas9识别的前间区的3'末端的下游。

[0331] 使用pNW33n质粒作为骨架以克隆:

[0332] i) 在内部开发的葡萄糖阻遏启动子的控制下的cas9_{gt}基因;和

[0333] ii) 在史氏芽孢杆菌ET138的基因组中的pyrF基因的1kb上游和1kb下游区域作为

用于同源重组的模板,该同源重组将引起pyrF基因从史氏芽孢杆菌ET138的基因组的缺失;和

[0334] iii) 在组成型启动子的转录控制下的表达单指导RNA (sgRNA) 的模块。

[0335] 生成了三种单独的构建体,其中单指导RNA序列在前20个核苷酸处是不同的,所述前20个核苷酸对应于将gtCas9指导至在基因组中的其特异性DNA靶(也被称为间隔区)的序列。三种不同的间隔区序列被设计为靶向三种不同的候选前间区,其全部都在史氏芽孢杆菌ET138的pyrF基因中。这些构建体在本文中分别地被称为构建体1、构建体2和构建体3。

[0336] 三种不同的靶向的前间区在其3'末端具有以下候选PAM序列:

[0337] 1. TCCATTCC (根据体外测定的结果,阴性对照;被构建体编号3上编码的sgRNA靶向的前间区的3'-末端)

[0338] 2. ATCCCCAA (被构建体编号1上编码的sgRNA靶向的前间区的3'-末端; [SEQ ID NO:21])

[0339] 3. ACGGCCAA (被构建体编号2上编码的sgRNA靶向的前间区的3'-末端; [SEQ ID NO:22])

[0340] 在用三种构建体之一转化史氏芽孢杆菌ET 138细胞并且在筛选板上铺板后,获得以下结果:

[0341] 1. 当用靶向前间区(前间区在3'末端具有阴性对照TCCATTCC PAM序列)的构建体(构建体编号3)转化细胞时,转化效率不受影响(图12(A))。菌落的数目在与用pNW33n阳性对照构建体转化后的菌落的数目相同的范围中(图12(B))。15个菌落经历菌落PCR以筛选其中缺失pyrF基因的菌落,没有一个显示出缺失基因型-2.1kb的预期条带尺寸-,所有菌落是野生型-2.9kb的预期条带尺寸-(图13)。这指示测试的PAM事实上在体内没有

[0342] 被gtCas9识别。

[0343] 2. 当用构建体编号1转化细胞时,当与阳性对照(用pNW33n转化的细胞)相比时,仅获得几个菌落(图12(C))。20个菌落经历菌落PCR以筛选其中缺失pyrF基因的菌落。大多数(19个)菌落包含野生型和pyrF缺失基因型二者,而一个菌落具有pyrF缺失基因型(图14)。该结果指示,PAM序列ATCCCCAA [SEQ ID NO:21]在体内被gtCas9识别,因为没有观察到仅WT基因型。降低的转化效率也指示一部分的细胞群体已经被减少,其可以是可归因于由于通过gtCas9的成功靶向,通过DSDB对仅WT基因型细胞引起的细胞死亡。

[0344] 3. 当用构建体编号2转化细胞时,没有获得菌落(图12(D))。菌落的缺乏指示所有的细胞群体已经成功地被gtCas9靶向,其导致通过DSDB引起的细胞死亡。这表明ACGGCCAA [SEQ ID NO:22] PAM被gtCas9识别。

[0345] 这些结果指示,用以上提及的PAM序列,gtCas9在55°C在体内是有活性的,该结果与体外PAM确定结果一致。此外,它可以在相同的温度与质粒携带的同源重组模板组合用作基因组编辑工具。

[0346] 说明书的以下部分由编号的段落组成,这些段落仅仅提供了本文已经描述的本发明的陈述。在本部分中的编号的段落不是权利要求。权利要求在所附的题为“权利要求书”的部分阐述。

[0347] 1. 一种分离的成簇的规律地间隔的短回文重复 (CRISPR) 相关 (Cas) 蛋白或多肽,包含:

- [0348] a. 氨基酸基序EKDGKYYC[SEQ ID NO:2];和/或
- [0349] b. 氨基酸基序 $X_1X_2CTX_3X_4$ [SEQ ID NO:3], 其中 X_1 独立地选自异亮氨酸、甲硫氨酸或脯氨酸, X_2 独立地选自缬氨酸、丝氨酸、天冬酰胺或异亮氨酸, X_3 独立地选自谷氨酸或赖氨酸,并且 X_4 是丙氨酸、谷氨酸或精氨酸之一;和/或
- [0350] c. 氨基酸基序 X_5LXX_6IE [SEQ ID NO:4], 其中 X_5 独立地选自甲硫氨酸或苯丙氨酸,并且 X_6 独立地选自组氨酸或天冬酰胺;和/或
- [0351] d. 氨基酸基序 X_7VYSX_8K [SEQ ID NO:5], 其中 X_7 是谷氨酸或异亮氨酸,并且 X_8 是色氨酸、丝氨酸或赖氨酸之一;和/或
- [0352] e. 氨基酸基序 $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO:6], 其中 X_9 是丙氨酸或谷氨酸, X_{10} 是谷氨酰胺或赖氨酸, X_{11} 是精氨酸或丙氨酸, X_{12} 是天冬酰胺或丙氨酸,并且 X_{13} 是赖氨酸或丝氨酸;
- [0353] 其中, 当与至少一种靶向RNA分子和包含被靶向RNA分子识别的靶核酸序列的多核苷酸缔合时,Cas蛋白能够在50℃与100℃之间进行核酸裂解。
- [0354] 2. 一种分离的Cas蛋白或多肽片段, 该分离的Cas蛋白或多肽片段具有SEQ ID NO:1的氨基酸序列或与SEQ ID NO:1具有至少77%同一性的序列, 其中当与至少一种识别靶序列的RNA分子缔合时,Cas蛋白能够在50℃和100℃之间的温度结合、裂解、修饰或标记包含靶核酸序列的多核苷酸。
- [0355] 3. 如在编号的段落1或2中所述的Cas蛋白或多肽片段, 其中Cas蛋白或片段能够在50℃和75℃之间的温度, 优选地在60℃以上的温度;更优选地在60℃和80℃之间的温度;甚至更优选地在60℃和65℃之间的温度进行核酸结合、裂解、标记或修饰。
- [0356] 4. 如在编号的段落1至3的任一项中所述的Cas蛋白或多肽片段, 其中核酸结合、裂解、标记或修饰是DNA裂解。
- [0357] 5. 如在前述编号的段落的任一项中所述的Cas蛋白或多肽片段, 其中氨基酸序列包含SEQ ID NO:1的氨基酸序列或与SEQ ID NO:1具有至少77%同一性的序列。
- [0358] 6. 如在前述编号的段落的任一项中所述的Cas蛋白或多肽片段, 其中Cas蛋白从细菌、古核生物或病毒可获得。
- [0359] 7. 如在前述编号的段落的任一项中所述的Cas蛋白或多肽片段, 其中Cas蛋白从地芽孢杆菌属的种, 优选地从热脱氮地芽孢杆菌可获得。
- [0360] 8. 一种核糖核蛋白复合体, 所述核糖核蛋白复合体包含如在前述编号的段落的任一项中所述的Cas蛋白, 并且包含识别靶多核苷酸中的序列的至少一种靶向RNA分子。
- [0361] 9. 如在编号的段落8中所述的核糖核蛋白复合体, 其中靶向RNA分子包含crRNA和任选地tracrRNA。
- [0362] 10. 如在编号的段落7至9的任一项中所述的核糖核蛋白复合体, 其中至少一种RNA分子的长度是在35-135个核苷酸残基的范围中。
- [0363] 11. 如在编号的段落8或9中所述的核糖核蛋白复合体, 其中靶序列的长度是31个或32个核苷酸残基。
- [0364] 12. 如在编号的段落1至7的任一项中所述的Cas蛋白或多肽、或如在编号的段落8至11的任一项中所述的核糖核蛋白复合体, 其中Cas蛋白或多肽作为蛋白复合体的一部分被提供, 所述蛋白复合体包含至少一种另外的功能蛋白或非功能蛋白。

[0365] 13. 如在编号的段落12中所述的Cas蛋白、多肽或核糖核蛋白复合体,其中Cas蛋白或多肽和/或至少一种另外的蛋白还包含至少一个功能部分。

[0366] 14. 如在编号的段落13中所述的Cas蛋白或多肽、或核糖核蛋白复合体,其中至少一个功能部分与Cas蛋白、多肽或核糖核蛋白复合体的N-末端和/或C-末端、优选地N-末端融合或连接。

[0367] 15. 如在编号的段落13或14中所述的Cas蛋白或多肽、或核糖核蛋白复合体,其中至少一个功能部分是蛋白;任选地选自解旋酶、核酸酶、解旋酶-核酸酶、DNA甲基化酶、组蛋白甲基化酶、乙酰基转移酶、磷酸酶、激酶、转录(共)活化物、转录阻遏物、DNA结合蛋白、DNA结构蛋白、标志物蛋白、报告物蛋白、荧光蛋白、配体结合蛋白、信号肽、亚细胞定位序列、抗体表位或亲和纯化标签。

[0368] 16. 如在编号的段落15中所述的Cas蛋白或多肽、或核糖核蛋白复合体,其中Cas9核酸酶活性的天然活性被失活并且Cas蛋白与至少一个功能部分连接。

[0369] 17. 如在编号的段落15或16中所述的Cas蛋白或多肽、或核糖核蛋白复合体,其中至少一个功能部分是核酸酶结构域;优选地为FokI核酸酶结构域。

[0370] 18. 如在编号的段落15至17的任一项中所述的Cas蛋白或多肽、或核糖核蛋白复合体,其中至少一个功能部分是标志物蛋白,例如GFP。

[0371] 19. 一种编码Cas蛋白或多肽的分离的核酸分子,所述Cas蛋白或多肽包含:

[0372] a. 氨基酸基序EKDGKYYC[SEQ ID NO:2];和/或

[0373] b. 氨基酸基序 $X_1X_2CTX_3X_4$ [SEQ ID NO:3],其中 X_1 独立地选自异亮氨酸、甲硫氨酸或脯氨酸, X_2 独立地选自缬氨酸、丝氨酸、天冬酰胺或异亮氨酸, X_3 独立地选自谷氨酸或赖氨酸,并且 X_4 是丙氨酸、谷氨酸或精氨酸之一;和/或

[0374] c. 氨基酸基序 X_5LXX_6IE [SEQ ID NO:4],其中 X_5 独立地选自甲硫氨酸或苯丙氨酸,并且 X_6 独立地选自组氨酸或天冬酰胺;和/或

[0375] d. 氨基酸基序 X_7VYSX_8K [SEQ ID NO:5],其中 X_7 是谷氨酸或异亮氨酸,并且 X_8 是色氨酸、丝氨酸或赖氨酸之一;和/或

[0376] e. 氨基酸基序 $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO:6],其中 X_9 是丙氨酸或谷氨酸, X_{10} 是谷氨酰胺或赖氨酸, X_{11} 是精氨酸或丙氨酸, X_{12} 是天冬酰胺或丙氨酸,并且 X_{13} 是赖氨酸或丝氨酸;

[0377] 其中,当与至少一种靶向RNA分子和包含被该靶向RNA分子识别的靶核酸序列的多核苷酸缔合时,Cas蛋白或多肽能够在50℃与100℃之间进行DNA结合、裂解、标记或修饰。

[0378] 20. 一种分离的核酸分子,所述分离的核酸分子编码成簇的规律地间隔的短回文重复(CRISPR)相关(Cas)蛋白或其多肽片段,所述成簇的规律地间隔的短回文重复(CRISPR)相关(Cas)蛋白具有SEQ ID NO:1的氨基酸序列或与SEQ ID NO:1具有至少77%同一性的序列。

[0379] 21. 如在编号的段落19或20中所述的分离的核酸分子,所述分离的核酸分子还包含编码在翻译后与Cas蛋白或多肽融合的氨基酸序列的至少一种核酸序列。

[0380] 22. 如在编号的段落21中所述的分离的核酸分子,其中与编码Cas蛋白或多肽的核酸分子融合的至少一种核酸序列编码选自以下的蛋白:解旋酶、核酸酶、解旋酶-核酸酶、DNA甲基化酶、组蛋白甲基化酶、乙酰基转移酶、磷酸酶、激酶、转录(共)活化物、转录阻遏

物、DNA结合蛋白、DNA结构蛋白、标志物蛋白、报告物蛋白、荧光蛋白、配体结合蛋白、信号肽、亚细胞定位序列、抗体表位或亲和纯化标签。

[0381] 23.一种表达载体,所述表达载体包含如在编号的段落19至22的任一项中所述的核酸分子。

[0382] 24.如在编号的段落23中所述的表达载体,所述表达载体还包含编码至少一种靶向RNA分子的核苷酸序列。

[0383] 25.一种修饰靶核酸的方法,所述方法包括使所述核酸与以下接触:

[0384] a.如在编号的段落6至11的任一项中所述的核糖核蛋白复合体;或

[0385] b.如在编号的段落12至18的任一项中所述的蛋白或蛋白复合体和至少一种如在编号的段落6至11的任一项中定义的靶向RNA分子。

[0386] 26.一种修饰细胞中的靶核酸的方法,所述方法包括用编号的段落24的表达载体转化、转染或转导细胞;或可选地用编号的段落23的表达载体和包含编码如在编号的段落6至11的任一项中定义的靶向RNA分子的核苷酸序列的另外的表达载体转化、转染或转导细胞。

[0387] 27.一种修饰细胞中的靶核酸的方法,所述方法包括用编号的段落23的表达载体转化、转染或转导细胞,并且然后将如在编号的段落6至11的任一项中定义的靶向RNA分子递送至细胞或递送到细胞中。

[0388] 28.如在编号的段落25至28的任一项中所述的修饰靶核酸的方法,其中至少一个功能部分是标志物蛋白或报告物蛋白,并且该标志物蛋白或报告物蛋白与靶核酸缔合;优选地其中标志物是荧光蛋白,例如绿色荧光蛋白(GFP)。

[0389] 29.如在编号的段落25至28的任一项中所述的方法,其中靶核酸是DNA;优选地是dsDNA。

[0390] 30.如在编号的段落25至28的任一项中所述的方法,其中靶核酸是RNA。

[0391] 31.如在编号的段落29中所述的修饰靶核酸的方法,其中核酸是dsDNA,至少一个功能部分是核酸酶或解旋酶-核酸酶,并且修饰是在期望的基因座处的单链断裂或双链断裂。

[0392] 32.一种在期望的基因座处使基因表达沉默的方法,所述方法根据在编号的段落26、27、29或31的任一项中所述的方法中的任一种方法进行。

[0393] 33.一种在期望的位置处修饰或缺失和/或插入期望的核苷酸序列的方法,所述方法根据如在编号的段落26、27、29或31的任一项中所述的方法中的任一种方法进行。

[0394] 34.一种修饰细胞中的基因表达的方法,所述方法包括以如在编号的段落25至29的任一项中所述的方法修饰靶核酸序列;其中核酸是dsDNA,并且功能部分选自DNA修饰酶(例如甲基化酶或乙酰基转移酶)、转录活化物或转录阻遏物。

[0395] 35.一种修饰细胞中的基因表达的方法,所述方法包括以如编号的段落30所述的方法修饰靶核酸序列,其中核酸是mRNA并且功能部分是核糖核酸酶;任选地选自内切核酸酶、3'外切核酸酶或5'外切核酸酶。

[0396] 36.如在编号的段落25至35的任一项中所述的修饰靶核酸的方法,其中所述方法在50℃和100℃之间的温度进行。

[0397] 37.如在编号的段落36中所述的修饰靶核酸的方法,其中所述方法在60℃或60℃

以上的温度,优选地在60℃和80℃之间的温度,更优选地在60℃和65℃之间的温度进行。

[0398] 38.如在编号的段落25至37的任一项中所述的方法,其中细胞是原核细胞。

[0399] 39.如在编号的段落25至38的任一项中所述的方法,其中细胞是真核细胞。

[0400] 40.一种宿主细胞,所述宿主细胞通过如在编号的段落22至36的任一项中所述的方法转化。

[0001]	序列表			
[0002]	<110> 瓦赫宁根大学;科学技术基金会			
[0003]	<120> 热稳定的Cas9核酸酶			
[0004]	<130> P237231GB			
[0005]	<160> 22			
[0006]	<170> PatentIn 3.5版			
[0007]	<210> 1			
[0008]	<211> 1082			
[0009]	<212> PRT			
[0010]	<213> 热脱氮地芽孢杆菌(Geobacillus thermodenitrificans) T12			
[0011]	<400> 1			
[0012]	Met	Lys	Tyr	Lys
[0013]	1	5	10	15
[0014]	Ala	Val	Ile	Asn
[0015]	20	25	30	
[0016]	Ile	Phe	Asp	Arg
[0017]	35	40	45	
[0018]	Pro	Arg	Arg	Leu
[0019]	50	55	60	
[0020]	His	Arg	Leu	Glu
[0021]	65	70	75	80
[0022]	Thr	Lys	Glu	Glu
[0023]	85	90	95	
[0024]	Val	Trp	Gln	Leu
[0025]	100	105	110	
[0026]	Glu	Leu	Ala	Arg
[0027]	115	120	125	
[0028]	Ser	Asn	Arg	Lys
[0029]	130	135	140	
[0030]	Lys	His	Ile	Glu
[0031]	145	150	155	160
[0032]	Ala	Glu	Met	Val
[0033]	165	170	175	
[0034]	Lys	Glu	Asp	Asn
[0035]	180	185	190	
[0036]	Glu	Ile	Lys	Leu
[0037]	195	200	205	
[0038]	Cys	Thr	Glu	Ala
[0039]	210	215	220	
[0040]	Arg	Pro	Phe	Ala
[0041]	225	230	235	240

[0042]	Thr Phe Glu Pro Lys Glu Lys Arg Ala Pro Lys Ala Thr Tyr Thr Phe		
[0043]		245	250 255
[0044]	Gln Ser Phe Thr Val Trp Glu His Ile Asn Lys Leu Arg Leu Val Ser		
[0045]		260	265 270
[0046]	Pro Gly Gly Ile Arg Ala Leu Thr Asp Asp Glu Arg Arg Leu Ile Tyr		
[0047]		275	280 285
[0048]	Lys Gln Ala Phe His Lys Asn Lys Ile Thr Phe His Asp Val Arg Thr		
[0049]		290	295 300
[0050]	Leu Leu Asn Leu Pro Asp Asp Thr Arg Phe Lys Gly Leu Leu Tyr Asp		
[0051]		305	310 315 320
[0052]	Arg Asn Thr Thr Leu Lys Glu Asn Glu Lys Val Arg Phe Leu Glu Leu		
[0053]		325	330 335
[0054]	Gly Ala Tyr His Lys Ile Arg Lys Ala Ile Asp Ser Val Tyr Gly Lys		
[0055]		340	345 350
[0056]	Gly Ala Ala Lys Ser Phe Arg Pro Ile Asp Phe Asp Thr Phe Gly Tyr		
[0057]		355	360 365
[0058]	Ala Leu Thr Met Phe Lys Asp Asp Thr Asp Ile Arg Ser Tyr Leu Arg		
[0059]		370	375 380
[0060]	Asn Glu Tyr Glu Gln Asn Gly Lys Arg Met Glu Asn Leu Ala Asp Lys		
[0061]		385	390 395 400
[0062]	Val Tyr Asp Glu Glu Leu Ile Glu Glu Leu Leu Asn Leu Ser Phe Ser		
[0063]		405	410 415
[0064]	Lys Phe Gly His Leu Ser Leu Lys Ala Leu Arg Asn Ile Leu Pro Tyr		
[0065]		420	425 430
[0066]	Met Glu Gln Gly Glu Val Tyr Ser Thr Ala Cys Glu Arg Ala Gly Tyr		
[0067]		435	440 445
[0068]	Thr Phe Thr Gly Pro Lys Lys Lys Gln Lys Thr Val Leu Leu Pro Asn		
[0069]		450	455 460
[0070]	Ile Pro Pro Ile Ala Asn Pro Val Val Met Arg Ala Leu Thr Gln Ala		
[0071]		465	470 475 480
[0072]	Arg Lys Val Val Asn Ala Ile Ile Lys Lys Tyr Gly Ser Pro Val Ser		
[0073]		485	490 495
[0074]	Ile His Ile Glu Leu Ala Arg Glu Leu Ser Gln Ser Phe Asp Glu Arg		
[0075]		500	505 510
[0076]	Arg Lys Met Gln Lys Glu Gln Glu Gly Asn Arg Lys Lys Asn Glu Thr		
[0077]		515	520 525
[0078]	Ala Ile Arg Gln Leu Val Glu Tyr Gly Leu Thr Leu Asn Pro Thr Gly		
[0079]		530	535 540
[0080]	Leu Asp Ile Val Lys Phe Lys Leu Trp Ser Glu Gln Asn Gly Lys Cys		
[0081]		545	550 555 560
[0082]	Ala Tyr Ser Leu Gln Pro Ile Glu Ile Glu Arg Leu Leu Glu Pro Gly		
[0083]		565	570 575

[0084]	Tyr Thr Glu Val Asp His Val Ile Pro Tyr Ser Arg Ser Leu Asp Asp		
[0085]	580	585	590
[0086]	Ser Tyr Thr Asn Lys Val Leu Val Leu Thr Lys Glu Asn Arg Glu Lys		
[0087]	595	600	605
[0088]	Gly Asn Arg Thr Pro Ala Glu Tyr Leu Gly Leu Gly Ser Glu Arg Trp		
[0089]	610	615	620
[0090]	Gln Gln Phe Glu Thr Phe Val Leu Thr Asn Lys Gln Phe Ser Lys Lys		
[0091]	625	630	635
[0092]	Lys Arg Asp Arg Leu Leu Arg Leu His Tyr Asp Glu Asn Glu Glu Asn		
[0093]	645	650	655
[0094]	Glu Phe Lys Asn Arg Asn Leu Asn Asp Thr Arg Tyr Ile Ser Arg Phe		
[0095]	660	665	670
[0096]	Leu Ala Asn Phe Ile Arg Glu His Leu Lys Phe Ala Asp Ser Asp Asp		
[0097]	675	680	685
[0098]	Lys Gln Lys Val Tyr Thr Val Asn Gly Arg Ile Thr Ala His Leu Arg		
[0099]	690	695	700
[0100]	Ser Arg Trp Asn Phe Asn Lys Asn Arg Glu Glu Ser Asn Leu His His		
[0101]	705	710	715
[0102]	Ala Val Asp Ala Ala Ile Val Ala Cys Thr Thr Pro Ser Asp Ile Ala		
[0103]	725	730	735
[0104]	Arg Val Thr Ala Phe Tyr Gln Arg Arg Glu Gln Asn Lys Glu Leu Ser		
[0105]	740	745	750
[0106]	Lys Lys Thr Asp Pro Gln Phe Pro Gln Pro Trp Pro His Phe Ala Asp		
[0107]	755	760	765
[0108]	Glu Leu Gln Ala Arg Leu Ser Lys Asn Pro Lys Glu Ser Ile Lys Ala		
[0109]	770	775	780
[0110]	Leu Asn Leu Gly Asn Tyr Asp Asn Glu Lys Leu Glu Ser Leu Gln Pro		
[0111]	785	790	795
[0112]	Val Phe Val Ser Arg Met Pro Lys Arg Ser Ile Thr Gly Ala Ala His		
[0113]	805	810	815
[0114]	Gln Glu Thr Leu Arg Arg Tyr Ile Gly Ile Asp Glu Arg Ser Gly Lys		
[0115]	820	825	830
[0116]	Ile Gln Thr Val Val Lys Lys Lys Leu Ser Glu Ile Gln Leu Asp Lys		
[0117]	835	840	845
[0118]	Thr Gly His Phe Pro Met Tyr Gly Lys Glu Ser Asp Pro Arg Thr Tyr		
[0119]	850	855	860
[0120]	Glu Ala Ile Arg Gln Arg Leu Leu Glu His Asn Asn Asp Pro Lys Lys		
[0121]	865	870	875
[0122]	Ala Phe Gln Glu Pro Leu Tyr Lys Pro Lys Lys Asn Gly Glu Leu Gly		
[0123]	885	890	895
[0124]	Pro Ile Ile Arg Thr Ile Lys Ile Ile Asp Thr Thr Asn Gln Val Ile		
[0125]	900	905	910

[0126]	Pro Leu Asn Asp Gly Lys Thr Val Ala Tyr Asn Ser Asn Ile Val Arg
[0127]	915 920 925
[0128]	Val Asp Val Phe Glu Lys Asp Gly Lys Tyr Tyr Cys Val Pro Ile Tyr
[0129]	930 935 940
[0130]	Thr Ile Asp Met Met Lys Gly Ile Leu Pro Asn Lys Ala Ile Glu Pro
[0131]	945 950 955 960
[0132]	Asn Lys Pro Tyr Ser Glu Trp Lys Glu Met Thr Glu Asp Tyr Thr Phe
[0133]	965 970 975
[0134]	Arg Phe Ser Leu Tyr Pro Asn Asp Leu Ile Arg Ile Glu Phe Pro Arg
[0135]	980 985 990
[0136]	Glu Lys Thr Ile Lys Thr Ala Val Gly Glu Glu Ile Lys Ile Lys Asp
[0137]	995 1000 1005
[0138]	Leu Phe Ala Tyr Tyr Gln Thr Ile Asp Ser Ser Asn Gly Gly Leu
[0139]	1010 1015 1020
[0140]	Ser Leu Val Ser His Asp Asn Asn Phe Ser Leu Arg Ser Ile Gly
[0141]	1025 1030 1035
[0142]	Ser Arg Thr Leu Lys Arg Phe Glu Lys Tyr Gln Val Asp Val Leu
[0143]	1040 1045 1050
[0144]	Gly Asn Ile Tyr Lys Val Arg Gly Glu Lys Arg Val Gly Val Ala
[0145]	1055 1060 1065
[0146]	Ser Ser Ser His Ser Lys Ala Gly Glu Thr Ile Arg Pro Leu
[0147]	1070 1075 1080
[0148]	<210> 2
[0149]	<211> 8
[0150]	<212> PRT
[0151]	<213> 热脱氮地芽孢杆菌T12
[0152]	<400> 2
[0153]	Glu Lys Asp Gly Lys Tyr Tyr Cys
[0154]	1 5
[0155]	<210> 3
[0156]	<211> 6
[0157]	<212> PRT
[0158]	<213> 人工序列
[0159]	<220>
[0160]	<223> 嗜热Cas9的氨基酸基序
[0161]	<220>
[0162]	<221> misc_feature
[0163]	<222> (1) .. (1)
[0164]	<223> Xaa可以是Ile、Met或Pro的任一种
[0165]	<220>
[0166]	<221> misc_feature
[0167]	<222> (2) .. (2)

[0168] <223> Xaa可以是Val、Ser、Asn或Ile的任一种
[0169] <220>
[0170] <221> misc_feature
[0171] <222> (5) .. (5)
[0172] <223> Xaa可以是Glu或Lys的任一种
[0173] <220>
[0174] <221> misc_feature
[0175] <222> (6) .. (6)
[0176] <223> Xaa可以是Ala、Glu或Arg的任一种
[0177] <400> 3
[0178] Xaa Xaa Cys Thr Xaa Xaa
[0179] 1 5
[0180] <210> 4
[0181] <211> 6
[0182] <212> PRT
[0183] <213> 人工序列
[0184] <220>
[0185] <223> 嗜热Cas9的氨基酸基序
[0186] <220>
[0187] <221> misc_feature
[0188] <222> (1) .. (1)
[0189] <223> Xaa可以是Met或Phe的任一种
[0190] <220>
[0191] <221> misc_feature
[0192] <222> (4) .. (4)
[0193] <223> Xaa可以是His或Asn的任一种
[0194] <400> 4
[0195] Xaa Leu Lys Xaa Ile Glu
[0196] 1 5
[0197] <210> 5
[0198] <211> 6
[0199] <212> PRT
[0200] <213> 人工序列
[0201] <220>
[0202] <223> 嗜热Cas9的氨基酸基序
[0203] <220>
[0204] <221> misc_feature
[0205] <222> (1) .. (1)
[0206] <223> Xaa可以是Glu或Ile的任一种
[0207] <220>
[0208] <221> misc_feature
[0209] <222> (5) .. (5)

[0210] <223> Xaa可以是Trp、Ser或Lys的任一种
 [0211] <400> 5
 [0212] Xaa Val Tyr Ser Xaa Lys
 [0213] 1 5
 [0214] <210> 6
 [0215] <211> 12
 [0216] <212> PRT
 [0217] <213> 人工序列
 [0218] <220>
 [0219] <223> 嗜热Cas9的氨基酸基序
 [0220] <220>
 [0221] <221> misc_feature
 [0222] <222> (1) .. (1)
 [0223] <223> Xaa可以是Ala或Glu的任一种
 [0224] <220>
 [0225] <221> misc_feature
 [0226] <222> (4) .. (4)
 [0227] <223> Xaa可以是Gln或Lys的任一种
 [0228] <220>
 [0229] <221> misc_feature
 [0230] <222> (5) .. (5)
 [0231] <223> Xaa可以是Arg或Ala的任一种
 [0232] <220>
 [0233] <221> misc_feature
 [0234] <222> (9) .. (9)
 [0235] <223> Xaa可以是Asn或Ala的任一种
 [0236] <220>
 [0237] <221> misc_feature
 [0238] <222> (12) .. (12)
 [0239] <223> Xaa可以是Lys或Ser的任一种
 [0240] <400> 6
 [0241] Xaa Phe Tyr Xaa Xaa Arg Glu Gln Xaa Lys Glu Xaa
 [0242] 1 5 10
 [0243] <210> 7
 [0244] <211> 3249
 [0245] <212> DNA
 [0246] <213> 热脱氮地芽孢杆菌T12
 [0247] <400> 7
 [0248] atgaagtata aaatcgggtct tgatatcggc attacgtcta tcggttgggc tgtcattaat 60
 [0249] ttggacattc ctgcacatga agatttaggt gtccgcattt ttgacagagc ggaaaaccgc 120
 [0250] aaaaccgggg agtcactagc tcttccacgt cgcctcgccc gctccgccc acgtcgtctg 180
 [0251] cggcgctgca aacatcgact ggagcgcatt cgccgcctgt tcgtccgcga aggaatttta 240

[0252]	acgaaggaag agctgaacaa gctgtttgaa aaaaagcacg aaatcgacgt ctggcagctt	300
[0253]	cgtgttgaag cactggatcg aaaactaaat aacgatgaat tagcccgcat ctttcttcat	360
[0254]	ctggctaaac ggcgtggatt tagatccaac cgcaagagtg agcgcaccaa caaagaaaac	420
[0255]	agtacgatgc tcaaacatat tgaagaaaac caatccattc tttcaagtta ccgaacggtt	480
[0256]	gcagaaatgg ttgtcaagga tccgaaatth tccctgcaca agcgtataaa agaggataat	540
[0257]	tacaccaaca ctgttgcccg cgacgatctt gaacgggaaa tcaaactgat tttcgccaaa	600
[0258]	cagcgcgaat atgggaacat cgtttgcaca gaagcatttg aacacgagta tatttccatt	660
[0259]	tgggcatcgc aacgcccttt tgcttctaag gatgatatcg agaaaaaagt cggtttctgt	720
[0260]	acgtttgagc ctaaagaaaa acgcgcgcca aaagcaacat acacattcca gtccttcacc	780
[0261]	gtctgggaac atattaacaa acttcgtctt gtctccccgg gaggcattcg ggcactaacc	840
[0262]	gatgatgaac gtcgtcttat atacaagcaa gcatthcata aaaataaaat caccttccat	900
[0263]	gatgttcgaa cattgcttaa cttgcctgac gacacccgtt ttaaaggtct tttatatgac	960
[0264]	cgaaacacca cgctgaagga aaatgagaaa gttegtcttc ttgaactcgg cgcctatcat	1020
[0265]	aaaatacggg aagcgatcga cagcgtctat ggcaaaggag cagcaaaatc atttcgtccg	1080
[0266]	attgattttg atacatttgg ctacgcatta acgatgttta aagacgacac cgacattcgc	1140
[0267]	agttacttgc gaaacgaata cgaacaaaat ggaaaacgaa tggaaaatct agcggataaa	1200
[0268]	gtctatgatg aagaattgat tgaagaactt ttaaacttat cgttttctaa gtttggtcat	1260
[0269]	ctatccctta aagcgcttcg caacatcctt ccatatatgg aacaaggcga agtctactca	1320
[0270]	accgcttggt aacgagcagg atatacattt acagggccaa agaaaaaaca gaaaacggta	1380
[0271]	ttgtctccga acattccgcc gatcgccaat ccggtcgtca tgcgcgcact gacacaggca	1440
[0272]	cgcaaagtgg tcaatgccat tatcaaaaag tacggctcac cggcttccat ccatatcgaa	1500
[0273]	ctggcccggg aactatcaca atcctttgat gaacgacgta aaatgcagaa agaacaggaa	1560
[0274]	ggaaaccgaa agaaaaacga aactgccatt cgccaacttg ttgaatatgg gctgacgctc	1620
[0275]	aatccaactg ggcttgacat tgtgaaattc aaactatgga gcgaacaaaa cggaatgt	1680
[0276]	gcctattcac tccaaccgat cgaaatcgag cggttgctcg aaccaggcta tacagaagtc	1740
[0277]	gaccatgtga ttccatacag ccgaagcttg gacgatagct ataccaataa agttcttgtg	1800
[0278]	ttgacaaagg agaaccgtga aaaaggaaac cgcacccag ctgaatatth aggattaggc	1860
[0279]	tcagaacgtt ggcaacagtt cgagacgtth gtcttgacaa ataagcagtt ttcgaaaaag	1920
[0280]	aagcgggacg gactccttcg gcttcattac gatgaaaacg aagaaaatga gtttaaaaat	1980
[0281]	cgtaatctaa atgatacccg ttatatctca cgcttcttgg ctaactthtat tcgcgaacat	2040
[0282]	ctcaaatcgc ccgacagcga tgacaaacaa aaagtataca cgggtcaacgg ccgtattacc	2100
[0283]	gcccatttac gcagccgttg gaattthaac aaaaaccggg aagaatcgaa tttgcatcat	2160
[0284]	gccgtcgatg ctgccatcgt cgctgcaca acgccgagcg atatcgcccg agtcaccgcc	2220
[0285]	ttctatcaac ggcgcgaaca aaacaaagaa ctgtccaaaa agacggatcc gcagtttccg	2280
[0286]	cagccttggc cgcacttgc tgatgaactg caggcgcgtt tatcaaaaaa tccaaaggag	2340
[0287]	agtataaaag ctctcaatct tggaaattat gataacgaga aactcgaatc gttgcagccg	2400
[0288]	gtttttgtct cccgaatgcc gaagcggagc ataacaggag cggctcatca agaaacattg	2460
[0289]	cggcgttata tcggcatcga cgaacggagc ggaaaaatac agacggtcgt caaaaagaaa	2520
[0290]	ctatccgaga tccaactgga taaaacaggt catttcccaa tgtacgggaa agaaagcgat	2580
[0291]	ccaaggacat atgaagccat tcgccaacgg ttgcttgaac ataacaatga cccaaaaaag	2640
[0292]	gcgtttcaag agcctctgta taaaccgaag aagaacggag aactaggtcc tatcatccga	2700
[0293]	acaatcaaaa tcatcgatac gacaaatcaa gttattccgc tcaacgatgg caaacagtc	2760

[0294] gcctacaaca gcaacatcgt gcgggtcgac gtctttgaga aagatggcaa atattattgt 2820
 [0295] gtccctatct atacaataga tatgatgaaa gggatcttgc caaacaaggc gatcgagccg 2880
 [0296] aacaaaccgt actctgagtg gaaggaaatg acggaggact atacattccg attcagtcta 2940
 [0297] tacccaaata atcttatccg tatcgaattt ccccgagaaa aaacaataaa gactgctgtg 3000
 [0298] ggggaagaaa tcaaaattaa ggatctgttc gcctattatc aaaccatcga ctctccaat 3060
 [0299] ggagggttaa gtttggttag ccatgataac aacttttcgc tccgcagcat cggttcaaga 3120
 [0300] accctcaaac gattcgagaa ataccaagta gatgtgctag gcaacatcta caaagtgaga 3180
 [0301] ggggaaaaga gagttgggt ggcgtcatct tctcattcga aagccgggga aactatccgt 3240
 [0302] ccgttataa 3249
 [0303] <210> 8
 [0304] <211> 1045
 [0305] <212> PRT
 [0306] <213> 内氏放线菌 (Actinomyces naeslundii)
 [0307] <400> 8
 [0308] Met Trp Tyr Ala Ser Leu Met Ser Ala His His Leu Arg Val Gly Ile
 [0309] 1 5 10 15
 [0310] Asp Val Gly Thr His Ser Val Gly Leu Ala Thr Leu Arg Val Asp Asp
 [0311] 20 25 30
 [0312] His Gly Thr Pro Ile Glu Leu Leu Ser Ala Leu Ser His Ile His Asp
 [0313] 35 40 45
 [0314] Ser Gly Val Gly Lys Glu Gly Lys Lys Asp His Asp Thr Arg Lys Lys
 [0315] 50 55 60
 [0316] Leu Ser Gly Ile Ala Arg Arg Ala Arg Arg Leu Leu His His Arg Arg
 [0317] 65 70 75 80
 [0318] Thr Gln Leu Gln Gln Leu Asp Glu Val Leu Arg Asp Leu Gly Phe Pro
 [0319] 85 90 95
 [0320] Ile Pro Thr Pro Gly Glu Phe Leu Asp Leu Asn Glu Gln Thr Asp Pro
 [0321] 100 105 110
 [0322] Tyr Arg Val Trp Arg Val Arg Ala Arg Leu Val Glu Glu Lys Leu Pro
 [0323] 115 120 125
 [0324] Glu Glu Leu Arg Gly Pro Ala Ile Ser Met Ala Val Arg His Ile Ala
 [0325] 130 135 140
 [0326] Arg His Arg Gly Trp Arg Asn Pro Tyr Ser Lys Val Glu Ser Leu Leu
 [0327] 145 150 155 160
 [0328] Ser Pro Ala Asn Ala Asn Glu Ile Arg Lys Ile Cys Ala Arg Gln Gly
 [0329] 165 170 175
 [0330] Val Ser Pro Asp Val Cys Lys Gln Leu Leu Arg Ala Val Phe Lys Ala
 [0331] 180 185 190
 [0332] Asp Ser Pro Arg Gly Ser Ala Val Ser Arg Val Ala Pro Asp Pro Leu
 [0333] 195 200 205
 [0334] Pro Gly Gln Gly Ser Phe Arg Arg Ala Pro Lys Cys Asp Pro Glu Phe
 [0335] 210 215 220

[0336]	Gln Arg Phe Arg Ile Ile Ser Ile Val Ala Asn Leu Arg Ile Ser Glu		
[0337]	225	230	235 240
[0338]	Thr Lys Gly Glu Asn Arg Pro Leu Thr Ala Asp Glu Arg Arg His Val		
[0339]		245	250 255
[0340]	Val Thr Phe Leu Thr Glu Asp Ser Gln Ala Asp Leu Thr Trp Val Asp		
[0341]		260	265 270
[0342]	Val Ala Glu Lys Leu Gly Val His Arg Arg Asp Leu Arg Gly Thr Ala		
[0343]		275	280 285
[0344]	Val His Thr Asp Asp Gly Glu Arg Ser Ala Ala Arg Pro Pro Ile Asp		
[0345]	290	295	300
[0346]	Ala Thr Asp Arg Ile Met Arg Gln Thr Lys Ile Ser Ser Leu Lys Thr		
[0347]	305	310	315 320
[0348]	Trp Trp Glu Glu Ala Asp Ser Glu Gln Arg Gly Ala Met Ile Arg Tyr		
[0349]		325	330 335
[0350]	Leu Tyr Glu Asp Pro Thr Asp Ser Glu Cys Ala Glu Ile Ile Ala Glu		
[0351]		340	345 350
[0352]	Leu Pro Glu Glu Asp Gln Ala Lys Leu Asp Ser Leu His Leu Pro Ala		
[0353]		355	360 365
[0354]	Gly Arg Ala Ala Tyr Ser Arg Glu Ser Leu Thr Ala Leu Ser Asp His		
[0355]	370	375	380
[0356]	Met Leu Ala Thr Thr Asp Asp Leu His Glu Ala Arg Lys Arg Leu Phe		
[0357]	385	390	395 400
[0358]	Gly Val Asp Asp Ser Trp Ala Pro Pro Ala Glu Ala Ile Asn Ala Pro		
[0359]		405	410 415
[0360]	Val Gly Asn Pro Ser Val Asp Arg Thr Leu Lys Ile Val Gly Arg Tyr		
[0361]		420	425 430
[0362]	Leu Ser Ala Val Glu Ser Met Trp Gly Thr Pro Glu Val Ile His Val		
[0363]		435	440 445
[0364]	Glu His Val Arg Asp Gly Phe Thr Ser Glu Arg Met Ala Asp Glu Arg		
[0365]	450	455	460
[0366]	Asp Lys Ala Asn Arg Arg Arg Tyr Asn Asp Asn Gln Glu Ala Met Lys		
[0367]	465	470	475 480
[0368]	Lys Ile Gln Arg Asp Tyr Gly Lys Glu Gly Tyr Ile Ser Arg Gly Asp		
[0369]		485	490 495
[0370]	Ile Val Arg Leu Asp Ala Leu Glu Leu Gln Gly Cys Ala Cys Leu Tyr		
[0371]		500	505 510
[0372]	Cys Gly Thr Thr Ile Gly Tyr His Thr Cys Gln Leu Asp His Ile Val		
[0373]		515	520 525
[0374]	Pro Gln Ala Gly Pro Gly Ser Asn Asn Arg Arg Gly Asn Leu Val Ala		
[0375]	530	535	540
[0376]	Val Cys Glu Arg Cys Asn Arg Ser Lys Ser Asn Thr Pro Phe Ala Val		
[0377]	545	550	555 560

[0378]	Trp	Ala	Gln	Lys	Cys	Gly	Ile	Pro	His	Val	Gly	Val	Lys	Glu	Ala	Ile
[0379]					565					570					575	
[0380]	Gly	Arg	Val	Arg	Gly	Trp	Arg	Lys	Gln	Thr	Pro	Asn	Thr	Ser	Ser	Glu
[0381]					580					585					590	
[0382]	Asp	Leu	Thr	Arg	Leu	Lys	Lys	Glu	Val	Ile	Ala	Arg	Leu	Arg	Arg	Thr
[0383]					595					600					605	
[0384]	Gln	Glu	Asp	Pro	Glu	Ile	Asp	Glu	Arg	Ser	Met	Glu	Ser	Val	Ala	Trp
[0385]					610					615					620	
[0386]	Met	Ala	Asn	Glu	Leu	His	His	Arg	Ile	Ala	Ala	Ala	Tyr	Pro	Glu	Thr
[0387]					625					630					635	
[0388]	Thr	Val	Met	Val	Tyr	Arg	Gly	Ser	Ile	Thr	Ala	Ala	Ala	Arg	Lys	Ala
[0389]					645					650					655	
[0390]	Ala	Gly	Ile	Asp	Ser	Arg	Ile	Asn	Leu	Ile	Gly	Glu	Lys	Gly	Arg	Lys
[0391]					660					665					670	
[0392]	Asp	Arg	Ile	Asp	Arg	Arg	His	His	Ala	Val	Asp	Ala	Ser	Val	Val	Ala
[0393]					675					680					685	
[0394]	Leu	Met	Glu	Ala	Ser	Val	Ala	Lys	Thr	Leu	Ala	Glu	Arg	Ser	Ser	Leu
[0395]					690					695					700	
[0396]	Arg	Gly	Glu	Gln	Arg	Leu	Thr	Gly	Lys	Glu	Gln	Thr	Trp	Lys	Gln	Tyr
[0397]					705					710					715	
[0398]	Thr	Gly	Ser	Thr	Val	Gly	Ala	Arg	Glu	His	Phe	Glu	Met	Trp	Arg	Gly
[0399]					725					730					735	
[0400]	His	Met	Leu	His	Leu	Thr	Glu	Leu	Phe	Asn	Glu	Arg	Leu	Ala	Glu	Asp
[0401]					740					745					750	
[0402]	Lys	Val	Tyr	Val	Thr	Gln	Asn	Ile	Arg	Leu	Arg	Leu	Ser	Asp	Gly	Asn
[0403]					755					760					765	
[0404]	Ala	His	Thr	Val	Asn	Pro	Ser	Lys	Leu	Val	Ser	His	Arg	Leu	Gly	Asp
[0405]					770					775					780	
[0406]	Gly	Leu	Thr	Val	Gln	Gln	Ile	Asp	Arg	Ala	Cys	Thr	Pro	Ala	Leu	Trp
[0407]					785					790					795	
[0408]	Cys	Ala	Leu	Thr	Arg	Glu	Lys	Asp	Phe	Asp	Glu	Lys	Asn	Gly	Leu	Pro
[0409]					805					810					815	
[0410]	Ala	Arg	Glu	Asp	Arg	Ala	Ile	Arg	Val	His	Gly	His	Glu	Ile	Lys	Ser
[0411]					820					825					830	
[0412]	Ser	Asp	Tyr	Ile	Gln	Val	Phe	Ser	Lys	Arg	Lys	Lys	Thr	Asp	Ser	Asp
[0413]					835					840					845	
[0414]	Arg	Asp	Glu	Thr	Pro	Phe	Gly	Ala	Ile	Ala	Val	Arg	Gly	Gly	Phe	Val
[0415]					850					855					860	
[0416]	Glu	Ile	Gly	Pro	Ser	Ile	His	His	Ala	Arg	Ile	Tyr	Arg	Val	Glu	Gly
[0417]					865					870					875	
[0418]	Lys	Lys	Pro	Val	Tyr	Ala	Met	Leu	Arg	Val	Phe	Thr	His	Asp	Leu	Leu
[0419]					885					890					895	

[0420]	Ser Gln Arg His Gly Asp Leu Phe Ser Ala Val Ile Pro Pro Gln Ser		
[0421]		900	905 910
[0422]	Ile Ser Met Arg Cys Ala Glu Pro Lys Leu Arg Lys Ala Ile Thr Thr		
[0423]		915	920 925
[0424]	Gly Asn Ala Thr Tyr Leu Gly Trp Val Val Val Gly Asp Glu Leu Glu		
[0425]		930	935 940
[0426]	Ile Asn Val Asp Ser Phe Thr Lys Tyr Ala Ile Gly Arg Phe Leu Glu		
[0427]		945	950 955 960
[0428]	Asp Phe Pro Asn Thr Thr Arg Trp Arg Ile Cys Gly Tyr Asp Thr Asn		
[0429]		965	970 975
[0430]	Ser Lys Leu Thr Leu Lys Pro Ile Val Leu Ala Ala Glu Gly Leu Glu		
[0431]		980	985 990
[0432]	Asn Pro Ser Ser Ala Val Asn Glu Ile Val Glu Leu Lys Gly Trp Arg		
[0433]		995	1000 1005
[0434]	Val Ala Ile Asn Val Leu Thr Lys Val His Pro Thr Val Val Arg		
[0435]		1010	1015 1020
[0436]	Arg Asp Ala Leu Gly Arg Pro Arg Tyr Ser Ser Arg Ser Asn Leu		
[0437]		1025	1030 1035
[0438]	Pro Thr Ser Trp Thr Ile Glu		
[0439]		1040	1045
[0440]	<210> 9		
[0441]	<211> 1160		
[0442]	<212> PRT		
[0443]	<213> 酿脓链球菌 (Streptococcus pyogenes)		
[0444]	<400> 9		
[0445]	Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val		
[0446]	1	5	10 15
[0447]	Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe		
[0448]		20	25 30
[0449]	Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile		
[0450]		35	40 45
[0451]	Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu		
[0452]		50	55 60
[0453]	Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys		
[0454]	65	70	75 80
[0455]	Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser		
[0456]		85	90 95
[0457]	Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys		
[0458]		100	105 110
[0459]	His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr		
[0460]		115	120 125
[0461]	His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp		

[0462]	130	135	140
[0463]	Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His		
[0464]	145	150	155 160
[0465]	Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro		
[0466]	165	170	175
[0467]	Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Arg		
[0468]	180	185	190
[0469]	Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu		
[0470]	195	200	205
[0471]	Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe		
[0472]	210	215	220
[0473]	Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile		
[0474]	225	230	235 240
[0475]	Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp		
[0476]	245	250	255
[0477]	Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu		
[0478]	260	265	270
[0479]	Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr		
[0480]	275	280	285
[0481]	Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser		
[0482]	290	295	300
[0483]	Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys		
[0484]	305	310	315 320
[0485]	Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln		
[0486]	325	330	335
[0487]	Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr		
[0488]	340	345	350
[0489]	Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp		
[0490]	355	360	365
[0491]	Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly		
[0492]	370	375	380
[0493]	Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp		
[0494]	385	390	395 400
[0495]	Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr		
[0496]	405	410	415
[0497]	Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala		
[0498]	420	425	430
[0499]	His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr		
[0500]	435	440	445
[0501]	Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp		
[0502]	450	455	460
[0503]	Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe		

[0504]	465	470	475	480
[0505]	Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe			
[0506]		485	490	495
[0507]	Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu			
[0508]		500	505	510
[0509]	His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly			
[0510]		515	520	525
[0511]	Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly			
[0512]		530	535	540
[0513]	Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln			
[0514]	545	550	555	560
[0515]	Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile			
[0516]		565	570	575
[0517]	Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro			
[0518]		580	585	590
[0519]	Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu			
[0520]		595	600	605
[0521]	Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg			
[0522]		610	615	620
[0523]	Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys			
[0524]	625	630	635	640
[0525]	Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg			
[0526]		645	650	655
[0527]	Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys			
[0528]		660	665	670
[0529]	Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys			
[0530]		675	680	685
[0531]	Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp			
[0532]		690	695	700
[0533]	Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr			
[0534]	705	710	715	720
[0535]	Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp			
[0536]		725	730	735
[0537]	Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser			
[0538]		740	745	750
[0539]	Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg			
[0540]		755	760	765
[0541]	Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val			
[0542]		770	775	780
[0543]	Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe			
[0544]	785	790	795	800
[0545]	Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys			

[0546]		805		810		815
[0547]	Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser					
[0548]		820		825		830
[0549]	Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu					
[0550]		835		840		845
[0551]	Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile					
[0552]		850		855		860
[0553]	Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser					
[0554]	865		870		875	880
[0555]	Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly					
[0556]		885		890		895
[0557]	Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile					
[0558]		900		905		910
[0559]	Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser					
[0560]		915		920		925
[0561]	Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly					
[0562]		930		935		940
[0563]	Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile					
[0564]	945		950		955	960
[0565]	Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala					
[0566]		965		970		975
[0567]	Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys					
[0568]		980		985		990
[0569]	Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser					
[0570]		995		1000		1005
[0571]	Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys					
[0572]		1010		1015		1020
[0573]	Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys					
[0574]		1025		1030		1035
[0575]	Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln					
[0576]		1040		1045		1050
[0577]	His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe					
[0578]		1055		1060		1065
[0579]	Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu					
[0580]		1070		1075		1080
[0581]	Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala					
[0582]		1085		1090		1095
[0583]	Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro					
[0584]		1100		1105		1110
[0585]	Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr					
[0586]		1115		1120		1125
[0587]	Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser					

[0588]	1130	1135	1140
[0589]	Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly		
[0590]	1145	1150	1155
[0591]	Gly Asp		
[0592]	1160		
[0593]	<210> 10		
[0594]	<211> 8		
[0595]	<212> DNA		
[0596]	<213> 人工序列		
[0597]	<220>		
[0598]	<223> RNA、DNA或合成的核酸共有PAM序列		
[0599]	<220>		
[0600]	<221> misc_feature		
[0601]	<222> (7) .. (7)		
[0602]	<223> n是a、c、g、t或u		
[0603]	<400> 10		
[0604]	ccccccna 8		
[0605]	<210> 11		
[0606]	<211> 8		
[0607]	<212> DNA		
[0608]	<213> 人工序列		
[0609]	<220>		
[0610]	<223> DNA PAM序列		
[0611]	<400> 11		
[0612]	ccccccaa 8		
[0613]	<210> 12		
[0614]	<211> 5		
[0615]	<212> DNA		
[0616]	<213> 人工序列		
[0617]	<220>		
[0618]	<223> DNA PAM序列		
[0619]	<400> 12		
[0620]	cccccc 5		
[0621]	<210> 13		
[0622]	<211> 8		
[0623]	<212> DNA		
[0624]	<213> 人工序列		
[0625]	<220>		
[0626]	<223> RNA、DNA或合成的核酸共有PAM序列		
[0627]	<220>		
[0628]	<221> misc_feature		
[0629]	<222> (6) .. (7)		

[0630]	<223> n是a、c、g、t或u
[0631]	<400> 13
[0632]	cccccnna 8
[0633]	<210> 14
[0634]	<211> 6
[0635]	<212> DNA
[0636]	<213> 人工序列
[0637]	<220>
[0638]	<223> DNA PAM序列
[0639]	<400> 14
[0640]	cccccc 6
[0641]	<210> 15
[0642]	<211> 6
[0643]	<212> DNA
[0644]	<213> 人工序列
[0645]	<220>
[0646]	<223> RNA、DNA或合成的核酸共有PAM序列
[0647]	<220>
[0648]	<221> misc_feature
[0649]	<222> (1) .. (1)
[0650]	<223> n是a、c、g、t或u
[0651]	<400> 15
[0652]	nccccc 6
[0653]	<210> 16
[0654]	<211> 8
[0655]	<212> DNA
[0656]	<213> 人工序列
[0657]	<220>
[0658]	<223> RNA、DNA或合成的核酸共有PAM序列
[0659]	<220>
[0660]	<221> misc_feature
[0661]	<222> (1) .. (1)
[0662]	<223> n是a、c、g、t或u
[0663]	<220>
[0664]	<221> misc_feature
[0665]	<222> (7) .. (7)
[0666]	<223> n是a、c、g、t或u
[0667]	<400> 16
[0668]	nccccnna 8
[0669]	<210> 17
[0670]	<211> 8
[0671]	<212> DNA

[0672]	<213> 人工序列
[0673]	<220>
[0674]	<223> RNA、DNA或合成的核酸共有PAM序列
[0675]	<220>
[0676]	<221> misc_feature
[0677]	<222> (2) .. (2)
[0678]	<223> n是a、c、g、t或u
[0679]	<400> 17
[0680]	cnccccac 8
[0681]	<210> 18
[0682]	<211> 8
[0683]	<212> DNA
[0684]	<213> 人工序列
[0685]	<220>
[0686]	<223> DNA PAM序列
[0687]	<400> 18
[0688]	ccccccag 8
[0689]	<210> 19
[0690]	<211> 8
[0691]	<212> DNA
[0692]	<213> 人工序列
[0693]	<220>
[0694]	<223> DNA PAM序列
[0695]	<400> 19
[0696]	ccccccat 8
[0697]	<210> 20
[0698]	<211> 8
[0699]	<212> DNA
[0700]	<213> 人工序列
[0701]	<220>
[0702]	<223> DNA PAM序列
[0703]	<400> 20
[0704]	ccccccac 8
[0705]	<210> 21
[0706]	<211> 8
[0707]	<212> DNA
[0708]	<213> 史氏芽孢杆菌 (Bacillus smithii) ET138
[0709]	<400> 21
[0710]	atccccaa 8
[0711]	<210> 22
[0712]	<211> 8
[0713]	<212> DNA

-
- [0714] <213> 史氏芽孢杆菌ET138
[0715] <400> 22
[0716] acggccaa 8

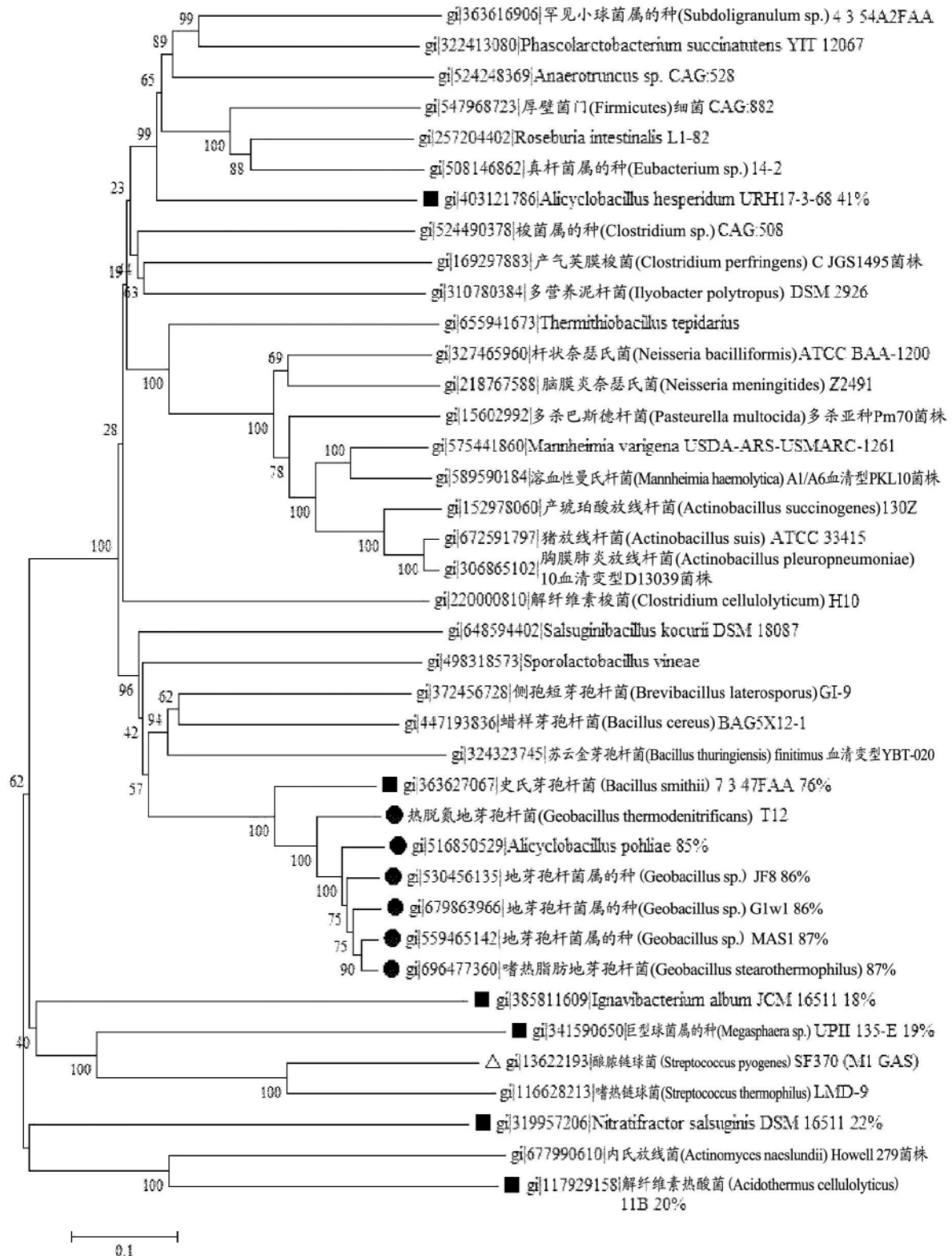


图1

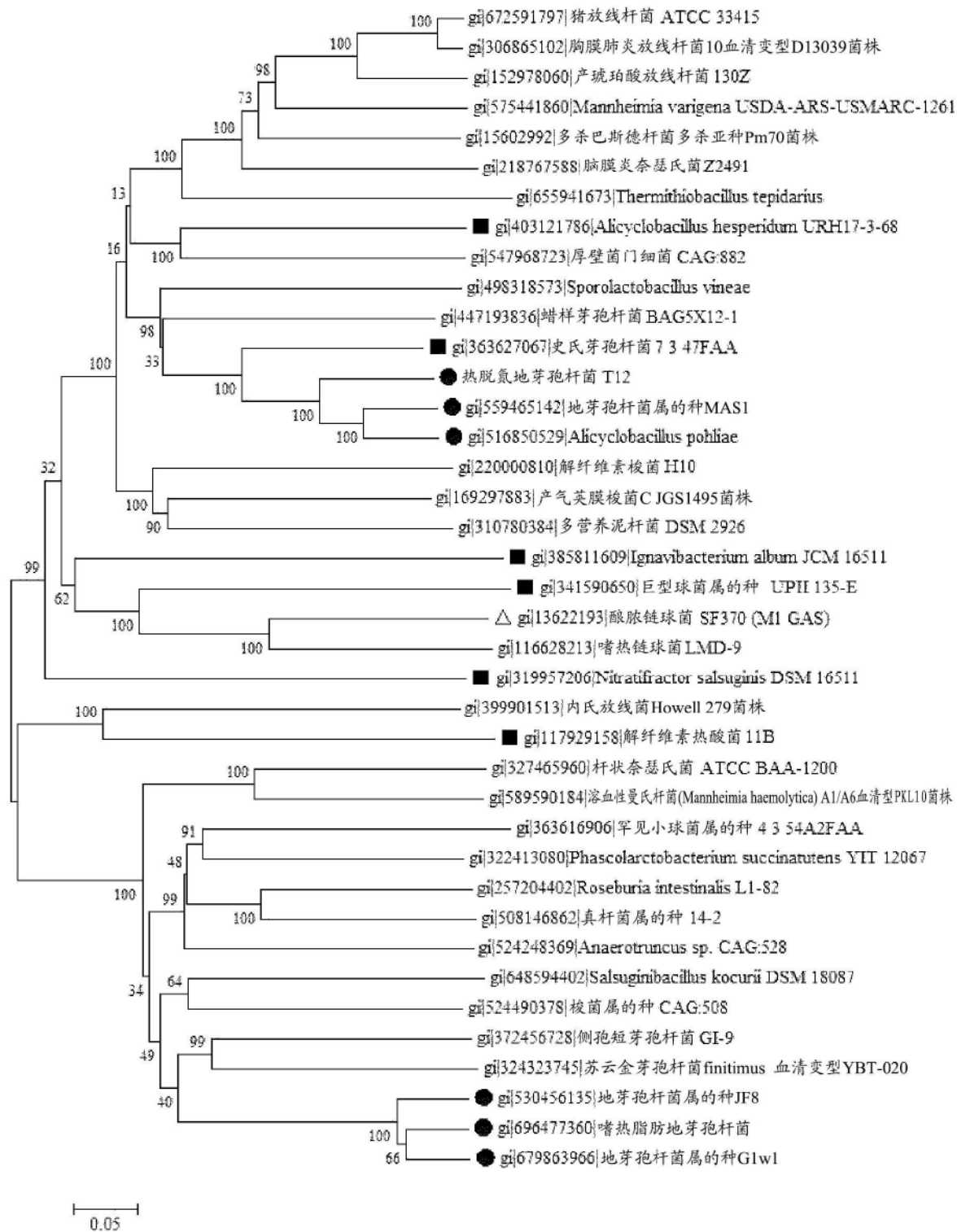


图2

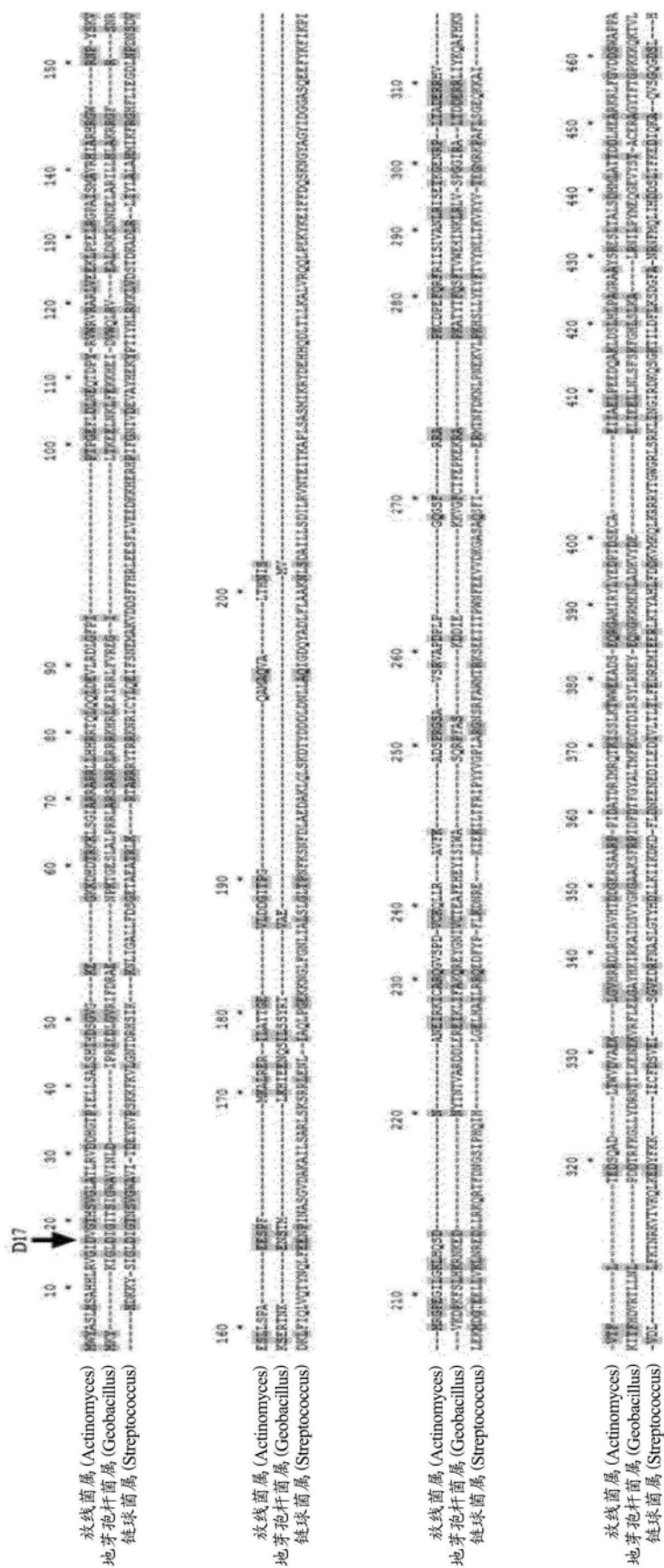
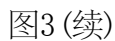


图3



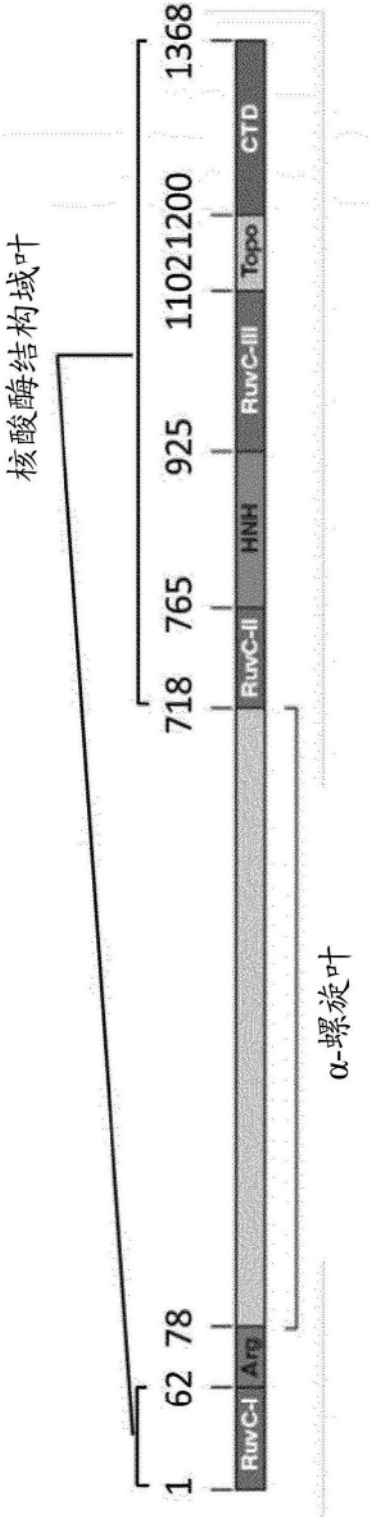
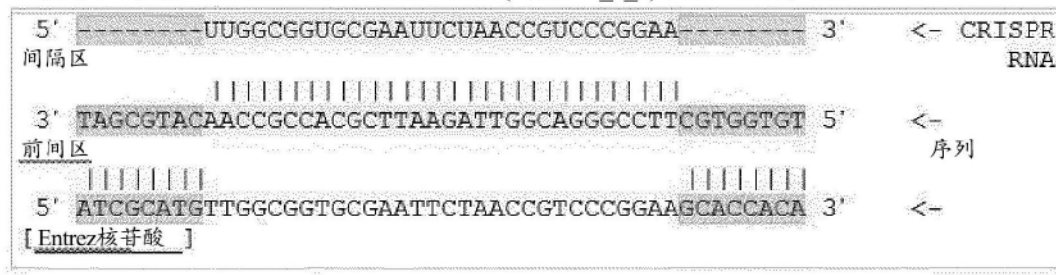


图4

tracrRNA t12 重复 gb	1 attgtttttccctcccatgcacaatagtttttatagtaaaaaagacottgacgtttttccgccaagggtcttc -----
tracrRNA t12 重复 gb	71 gttcgctaagagtggggaatgcccgaagaagcggcgcataggcgatccccaacgccaacgggtcagtcct -----
tracrRNA t12 重复 gb	141 gcctataggcagaaagcccttatcatagtaacccctgagatcattgctgtgtatataacccctattactataa 1 -----gtcatagttccctgagattatcgctgtgtataat-----
tracrRNA t12 重复 gb	211 taatgtttatatattgggaaaaatcaagtcctttttctatatatttttatacttttcatttcttgcattat -----
tracrRNA t12 重复 gb	281 gatgatgtgaggaggatagatttctgacaggagggttcacatcg -----

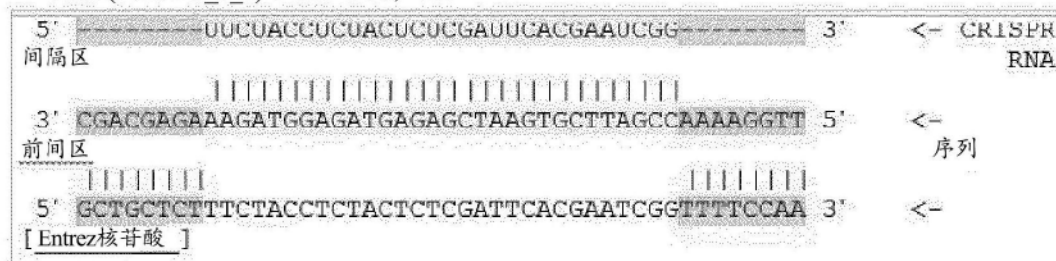
图6

1. 匹配至: DQ453159 地芽孢杆菌属 病毒E2, 完整基因组(DQ453159) 位置: 39980-39951
具有: 间隔区4 CRISPR No.4 间隔区 No.1 (间隔区4_4_1) 位置: 1-30, 链: +



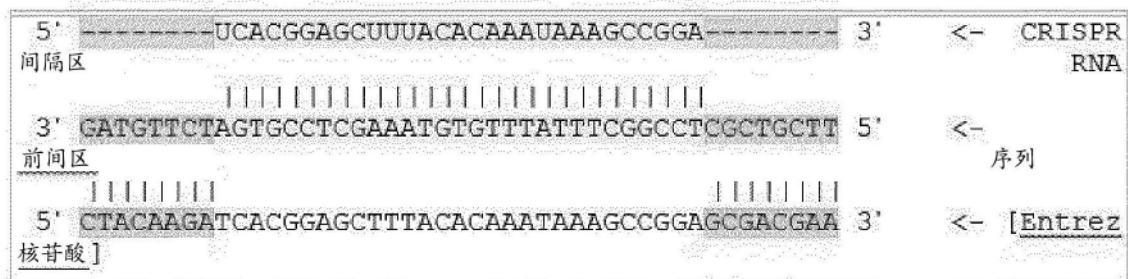
评分: 30

2. 匹配至: 香鱼海槽芽孢杆菌(Bacillus alveayuensis) 菌株24KAM51 LG50_053, 整个基因组
鸟枪法序列(NZ_JYCE01000053) 位置: 110918-110889, 具有: 间隔区9 CRISPR No.9 间隔区
No.1 (间隔区9_9_1) 位置: 1-30, 链: +



评分: 30

3. 匹配至: 嗜热厌氧芽孢杆菌(Anoxybacillus flavithermus)WK1, 完整基因组(CP000922) 位置:
701422-701450, 具有: 间隔区5 CRISPR No.5 间隔区 No.1 (间隔区5_5_1) 位置: 1-29, 链: -



评分: 29

4. 匹配至: 嗜热地芽孢杆菌(Geobacillus kaustophilus)菌株 Et23 LG51_086, 整个基因组鸟枪法序列
(NZ_JYCF01000086) 位置: 15565-15537, 具有: 间隔区5 CRISPR No.5 间隔区 No.1 (间隔区5_5_1)
位置: 1-29, 链: +

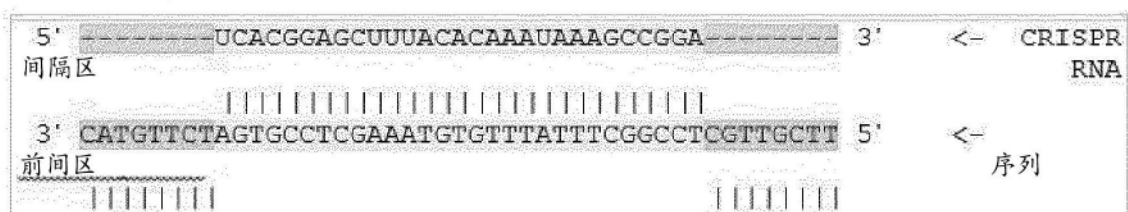


图7

5' GTACAAAGATCACGGAGCTTTACACAAATAAAGCCGGAGCAACGAA 3' <- [Entrez
核苷酸]

评分: 29

5. 匹配至: DQ453159 DQ453159 地芽孢杆菌属 病毒E2, 完整基因组(DQ453159) 位置: 6492-6463, 具有: 间隔区6 CRISPR No.6 间隔区 No.1 (间隔区6_6_1) 位置: 1-30, 链: +

5' -----CAACACCUUCCGCGCUGUCUCGUCUACUUU----- 3' <- CRISPR
间隔区 RNA
|||||
3' CGAAGAAAGTTGTGGAAGGCACGACAGAGTAGATGAAAAAATGCTT 5' <-
前间区 序列
|||||
5' GCTTCCTTTCAACACCTTCCGTGCTGTCTCATCTACTTTTTTACGAA 3' <-
[Entrez核苷酸]

评分: 26

6. 匹配至: 贝氏巴斯德菌(*Pasteurella bettyae*) CCUG 2042 重叠群(contig)00003, 整个基因组鸟枪法序列(NZ_AJSX01000041) 位置: 137780-137808, 具有: 间隔区7 CRISPR No.7 间隔区 No.1 (间隔区7_7_1) 位置: 1-30, 链: -

5' -----UUGAUUAGCAAUUUGACUUGGGAAUUUAGC----- 3' <- CRISPR
间隔区 RNA
|||||
3' TAATTGTAACTAATCGTTAACTGACGCCTTAAATCATTACGGTT 5' <-
前间区 序列
|||||
5' ATTAACATTTGATTAGCAATTTGACTGCGGAATTTAGTAATGCCAA 3' <-
[Entrez核苷酸]

图7 (续)

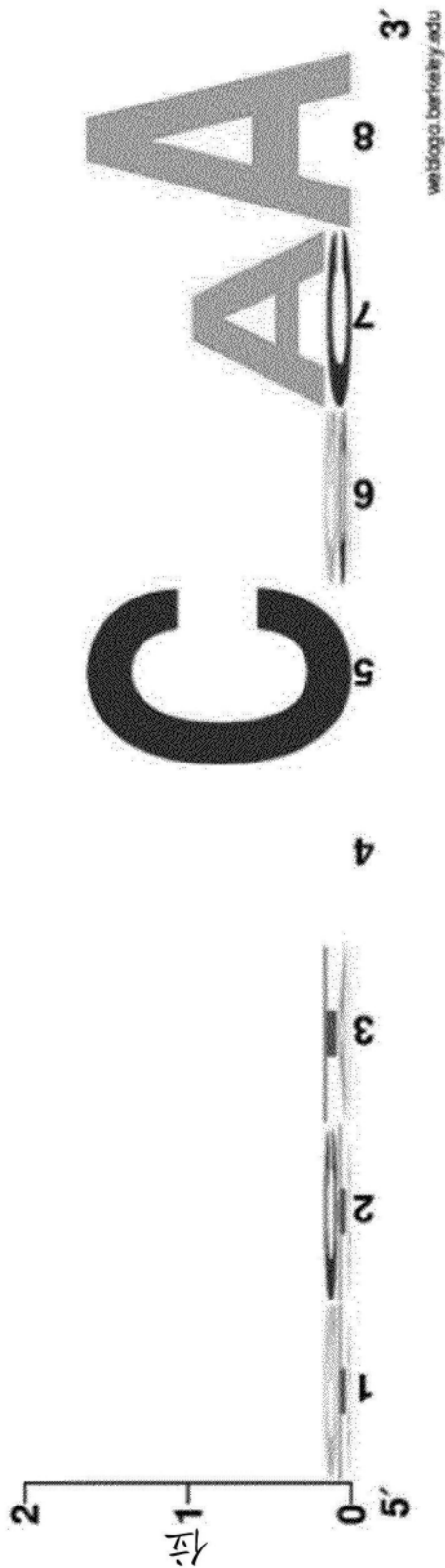


图8

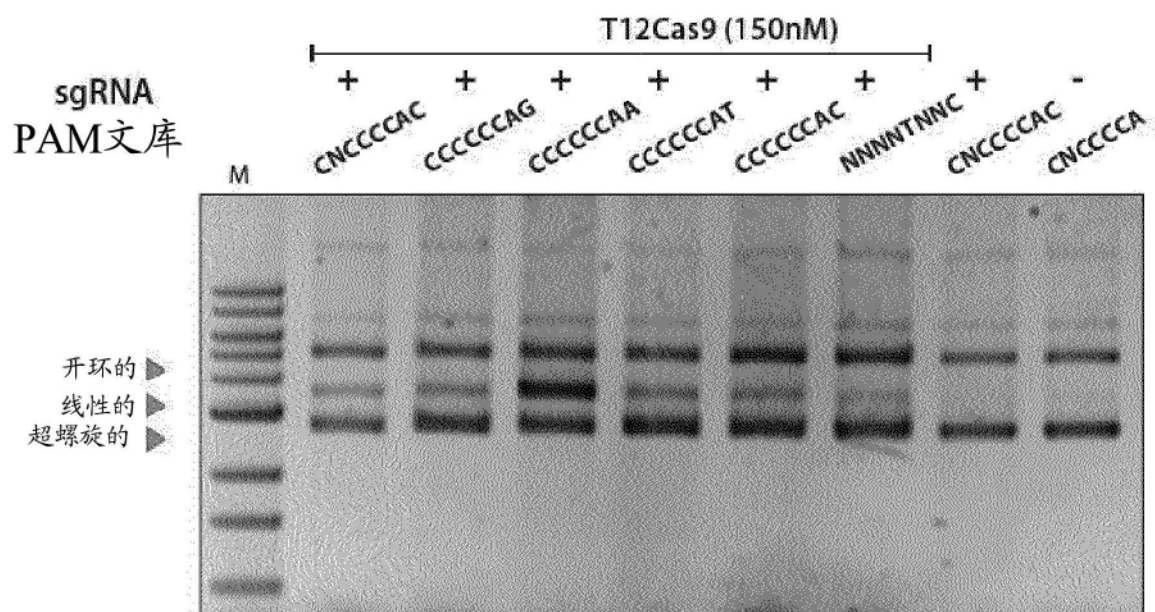


图9

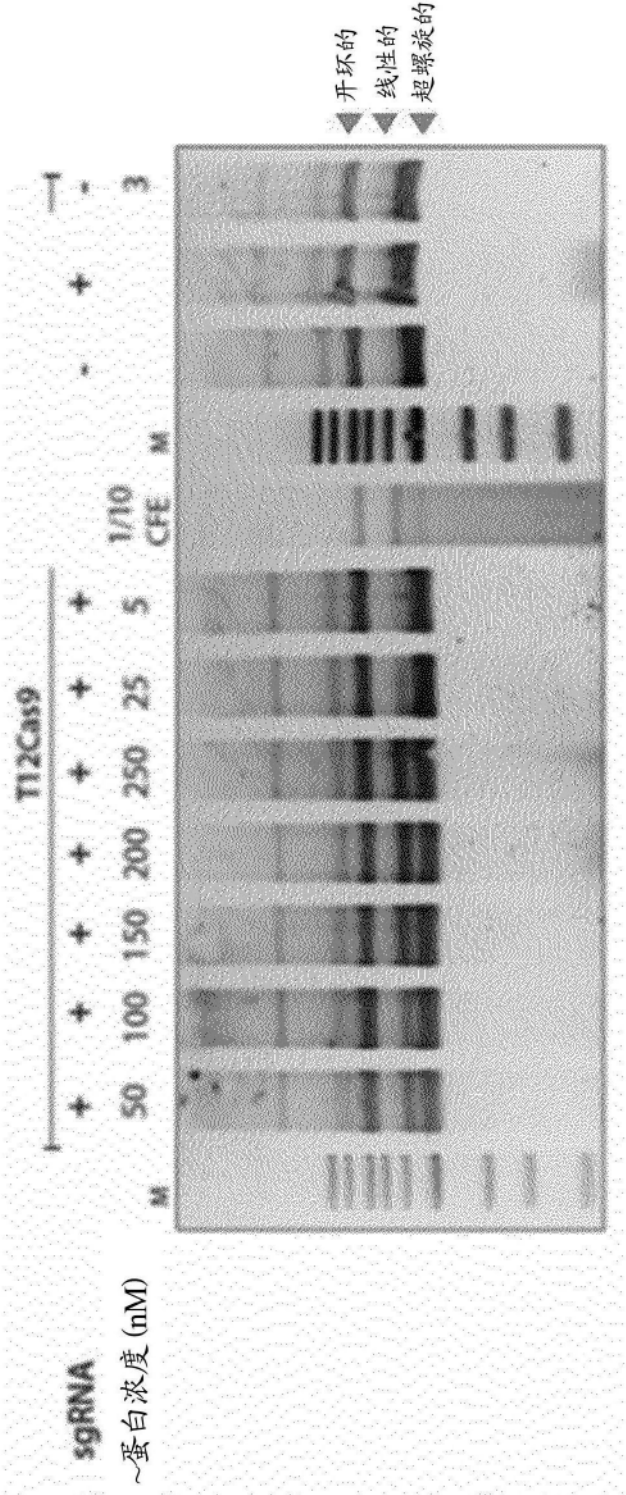


图10

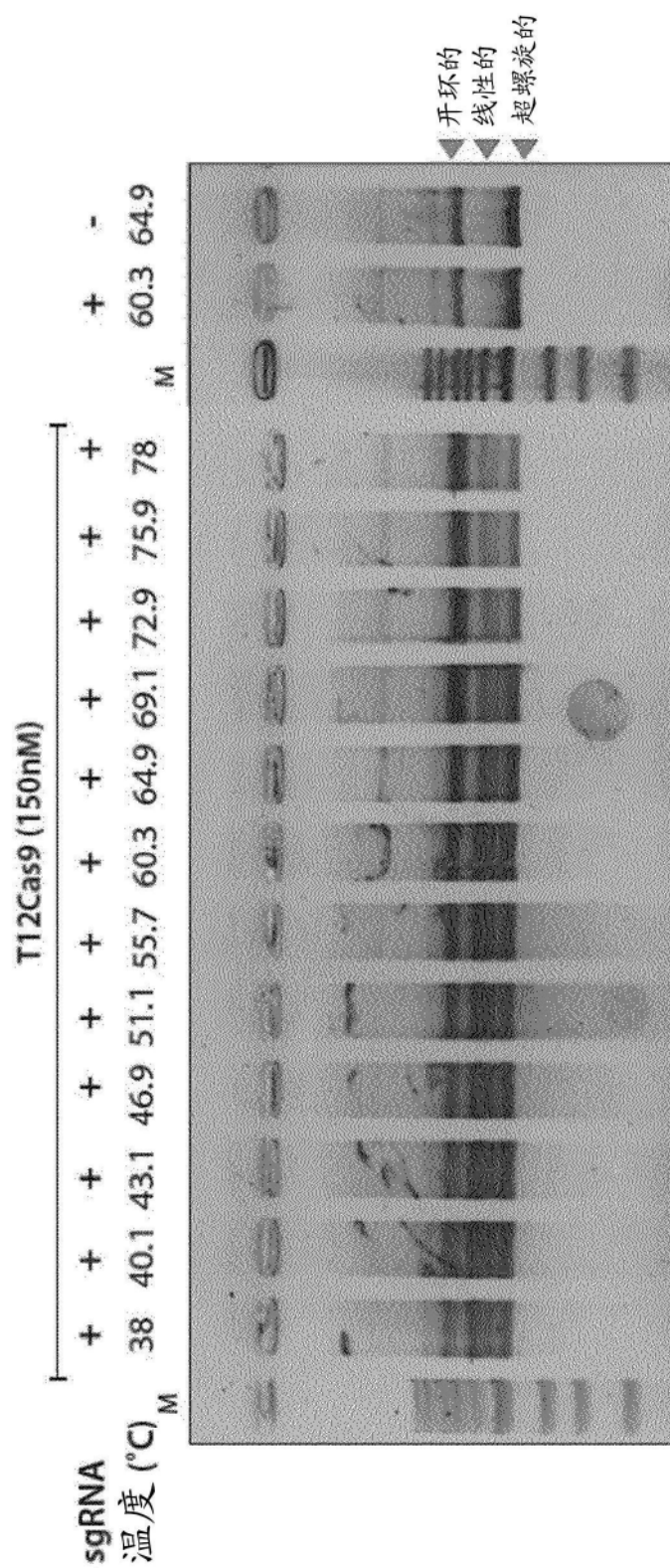
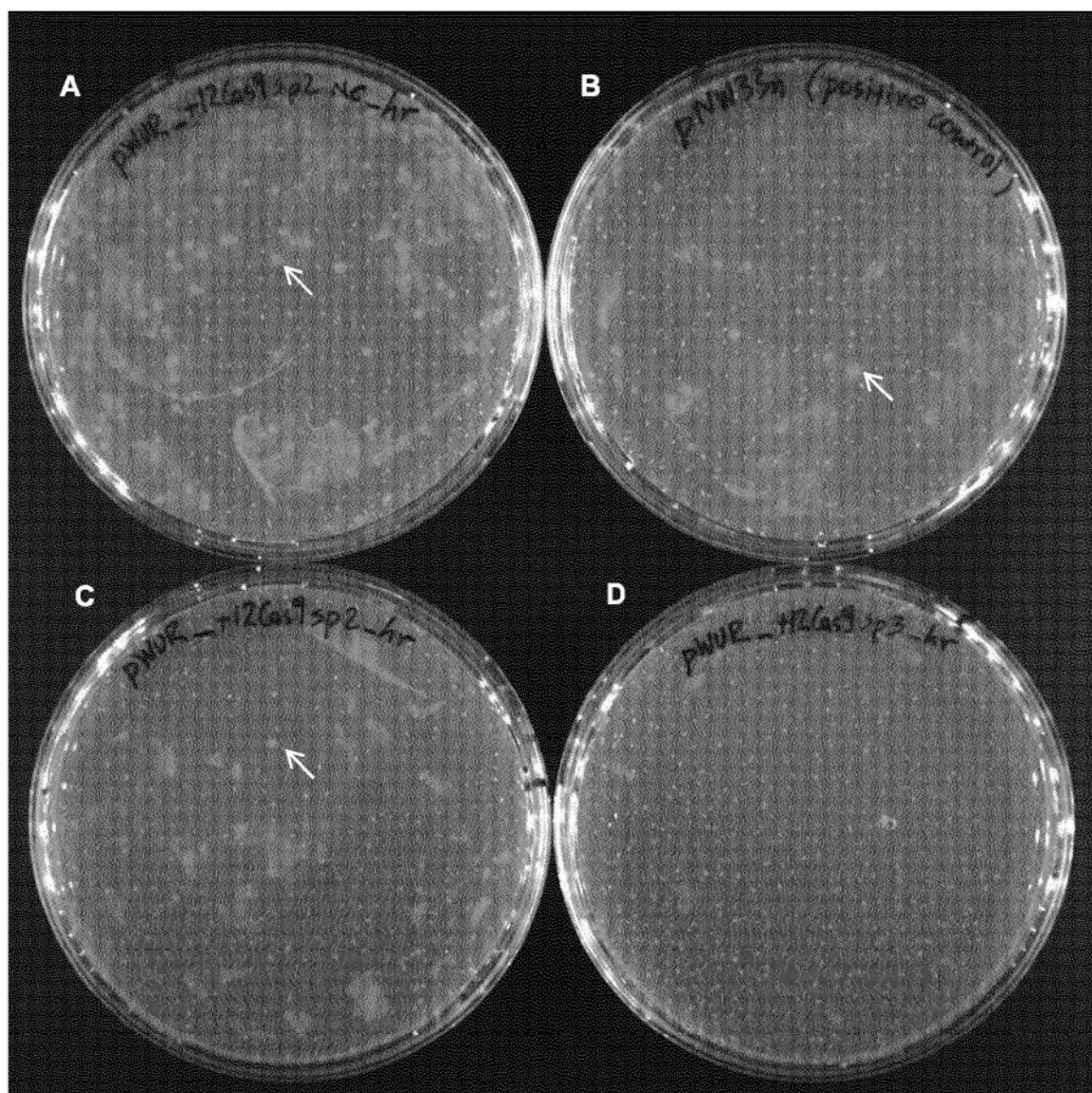


图11



菌落用箭头指示出。

图12

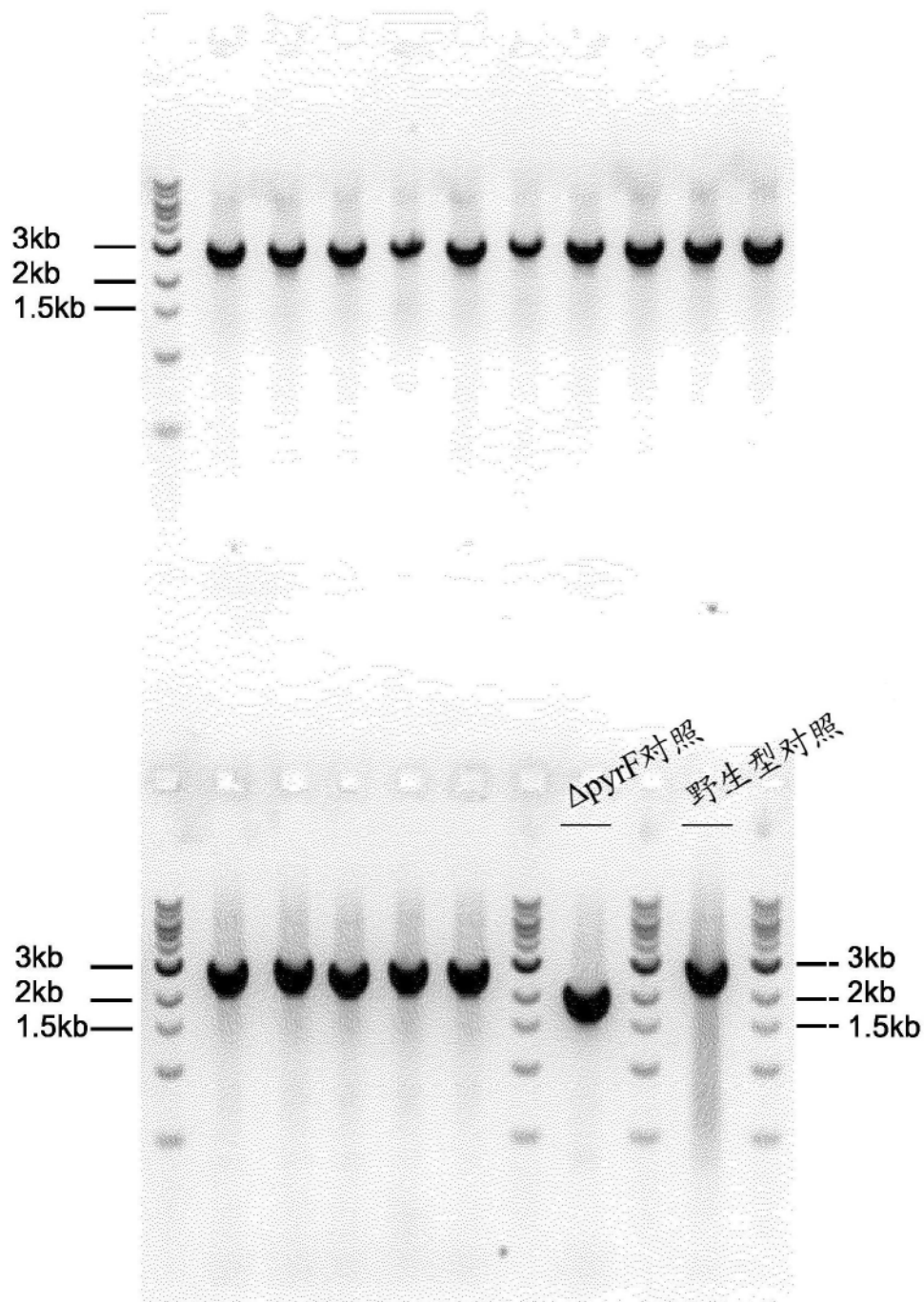


图13

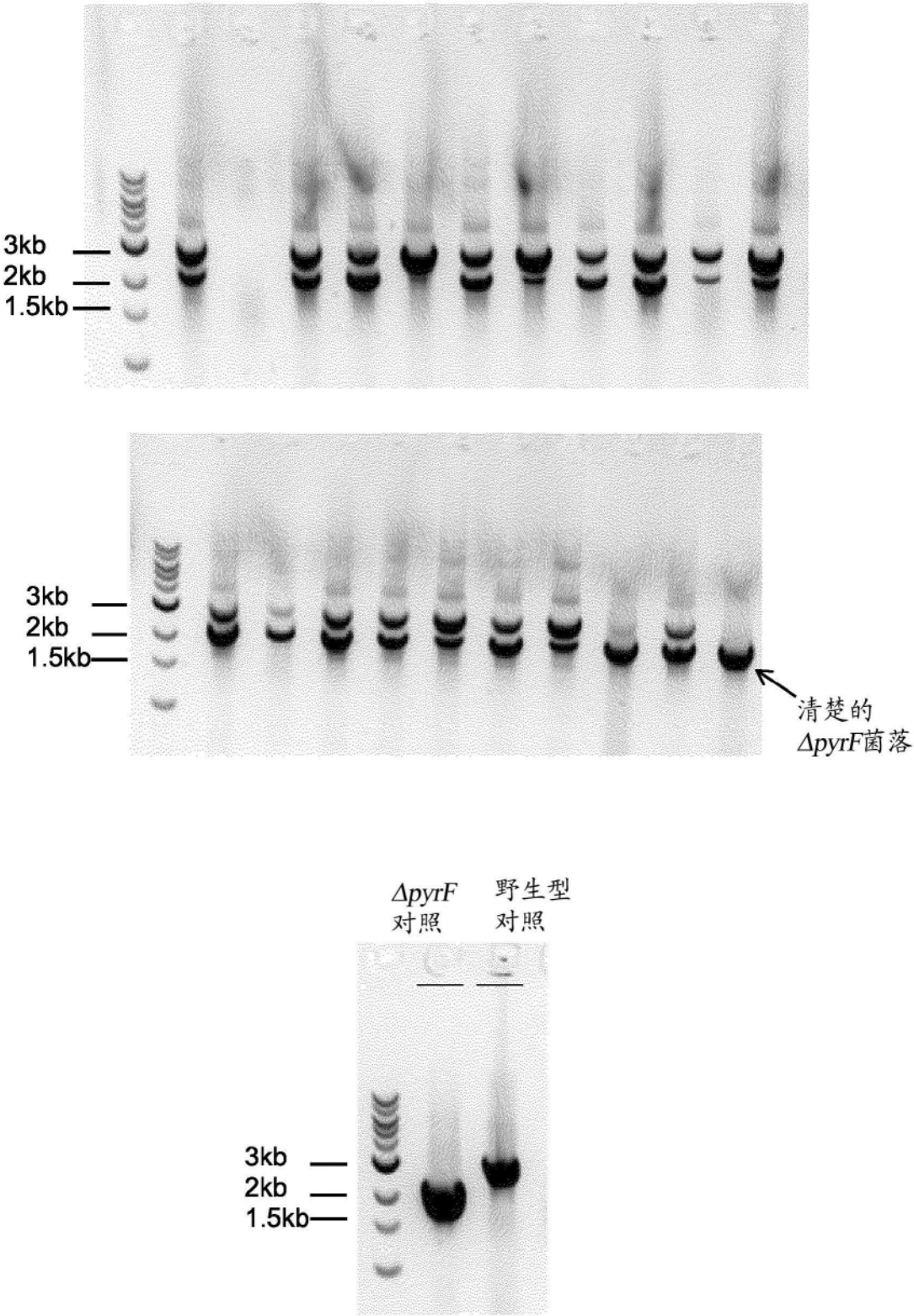


图14