(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 114206376 A
(43)申请公布日 2022.03.18

(21)申请号 202080050148.X

(22)申请日 2020.07.10

(30)优先权数据
62/873,108 2019.07.11 US
62/873,118 2019.07.11 US
62/895,406 2019.09.03 US
62/895,422 2019.09.03 US

(85)PCT国际申请进入国家阶段日
2022.01.10

(86)PCT国际申请的申请数据
PCT/US2020/041714 2020.07.10

(87)PCT国际申请的公布数据
WO2021/007563 EN 2021.01.14

(71)申请人 阿伯生物技术公司
地址 美国马萨诸塞州

(72)发明人 D·A·斯科特 D·R·程
W·X·严 T·M·迪托马索

(74)专利代理机构 北京市中咨律师事务所 11247
代理人 胡志君 黄革生

(51)Int.Cl.
A61K 38/46(2006.01)
A61K 48/00(2006.01)
C07K 19/00(2006.01)
C12N 9/22(2006.01)
C12N 15/00(2006.01)
C12N 15/52(2006.01)

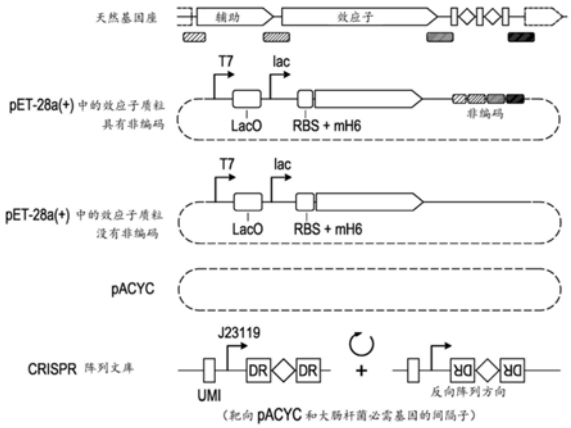
权利要求书4页 说明书64页
序列表104页 附图58页

(54)发明名称

新型CRISPR DNA靶向酶及系统

(57)摘要

本披露描述了用于以靶向方式操纵核酸的新型系统、方法和组合物。本披露描述了用于靶向修饰核酸的非天然存在的工程化CRISPR系统、组分和方法。每种系统包括一起靶向核酸的一种或多种蛋白质组分和一种或多种核酸组分。



CN 114206376 A

1.一种工程化的非天然存在的CLUST.133120成簇规律间隔短回文重复序列(CRISPR)-Cas系统,该系统包括:

(a)CRISPR相关蛋白或编码该CRISPR相关蛋白的核酸,其中该CRISPR相关蛋白包含与SEQ ID NO:1-29中任一个中列出的氨基酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的氨基酸序列;以及

(b)RNA指导物,该RNA指导物包含直接重复序列和能够与靶核酸杂交的间隔子序列;

其中该CRISPR相关蛋白能够结合该RNA指导物并且能够修饰与该间隔子序列互补的该靶核酸序列。

2.如权利要求1所述的系统,其中该CRISPR相关蛋白包含至少一个RuvC结构域或至少一个拆分型RuvC结构域。

3.如权利要求1或2所述的系统,其中该直接重复序列包含与SEQ ID NO:30-45、77-94或122-138中任一个中列出的核苷酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的核苷酸序列。

4.如权利要求1-3中任一项所述的系统,其中该CRISPR相关蛋白是与SEQ ID NO:1、SEQ ID NO:25、SEQ ID NO:26、SEQ ID NO:28或SEQ ID NO:29中列出的氨基酸序列具有至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)同一性的蛋白质。

5.如权利要求1-4中任一项所述的系统,其中该CRISPR相关蛋白能够识别原型间隔子相邻基序(PAM)序列,其中该PAM序列包括作为5'-TTN-3'、5'-YYN-3'、5'-HHN-3'、5'-YKN-3'或5'-HBN-3'列出的核酸序列,其中N是任何核苷酸,Y是C或T,K是G或T,B是G、T或C,并且H是A、C或T。

6.如权利要求1-5中任一项所述的系统,其中该RNA指导物的该间隔子序列包含约14个核苷酸至约50个核苷酸。

7.如权利要求6所述的系统,其中该RNA指导物的该间隔子序列包含20至35个核苷酸。

8.如权利要求1-7中任一项所述的系统,其中该CRISPR相关蛋白包含催化残基(例如,天冬氨酸或谷氨酸)。

9.如权利要求1-8中任一项所述的系统,其中该CRISPR相关蛋白切割该靶核酸。

10.如权利要求1-9中任一项所述的系统,其中该CRISPR相关蛋白进一步包含肽标签、荧光蛋白、碱基编辑结构域、DNA甲基化结构域、组蛋白残基修饰结构域、定位因子、转录修饰因子、光门控因子、化学诱导型因子或染色质可视化因子。

11.如权利要求1-10中任一项所述的系统,其中编码该CRISPR相关蛋白的该核酸经密码子优化以在细胞中表达。

12.如权利要求1-11中任一项所述的系统,其中编码该CRISPR相关蛋白的该核酸可操作地连接至启动子。

13.如权利要求1-12中任一项所述的系统,其中编码该CRISPR相关蛋白的该核酸在载体中。

14.如权利要求13所述的系统,其中该载体包括逆转录病毒载体、慢病毒载体、噬菌体

载体、腺病毒载体、腺相关载体或单纯疱疹载体。

15.如权利要求1-14中任一项所述的系统，其中该靶核酸是DNA分子。

16.如权利要求1-15中任一项所述的系统，其中该靶核酸包含PAM序列。

17.如权利要求1-16中任一项所述的系统，其中该CRISPR相关蛋白包含非特异性核酸酶活性。

18.如权利要求1-17中任一项所述的系统，其中通过该CRISPR相关蛋白和RNA指导物对该靶核酸的识别导致对该靶核酸的修饰。

19.如权利要求18所述的系统，其中对该靶核酸的该修饰是双链切割事件。

20.如权利要求18所述的系统，其中对该靶核酸的该修饰是单链切割事件。

21.如权利要求18所述的系统，其中对该靶核酸的该修饰导致插入事件。

22.如权利要求18所述的系统，其中对该靶核酸的该修饰导致缺失事件。

23.如权利要求18-22中任一项所述的系统，其中对该靶核酸的该修饰导致细胞毒性或细胞死亡。

24.如权利要求1-16中任一项所述的系统，该系统进一步包括供体模板核酸。

25.如权利要求24所述的系统，其中该供体模板核酸是DNA分子。

26.如权利要求24所述的系统，其中该供体模板核酸是RNA分子。

27.如权利要求1-26中任一项所述的系统，其中该系统不包括tracrRNA。

28.如权利要求1-27中任一项所述的系统，其中该CRISPR相关蛋白是自加工的。

29.如权利要求1-28中任一项所述的系统，其中该系统存在于递送组合物中，该递送组合物包含纳米颗粒、脂质体、外来体、微泡或基因枪。

30.如权利要求1-28中任一项所述的系统，该系统在细胞内。

31.如权利要求30所述的系统，其中该细胞是真核细胞，例如哺乳动物细胞，例如人细胞。

32.如权利要求30所述的系统，其中该细胞是原核细胞。

33.一种基因修饰的细胞，其中该细胞包含：

(a)CRISPR相关蛋白或编码该CRISPR相关蛋白的核酸，其中该CRISPR相关蛋白包含与SEQ ID NO:1-29中任一个中列出的氨基酸序列至少80％(例如，81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的氨基酸序列；以及

(b)RNA指导物，该RNA指导物包含直接重复序列和能够与靶核酸杂交的间隔子序列。

34.如权利要求33所述的基因修饰的细胞，其中该CRISPR相关蛋白是与SEQ ID NO:1、SEQ ID NO:25、SEQ ID NO:26、SEQ ID NO:28或SEQ ID NO:29中列出的氨基酸序列具有至少80％(例如，81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)同一性的蛋白质。

35.如权利要求33或34所述的基因修饰的细胞，其中该CRISPR相关蛋白能够识别PAM序列，该PAM序列包括作为5'-TTN-3'、5'-YYN-3'、5'-HHN-3'、5'-YKN-3'或5'-HBN-3'列出的核酸序列，其中N是任何核苷酸，Y是C或T，K是G或T，B是G、T或C，并且H是A、C或T。

36.如权利要求33-35中任一项所述的基因修饰的细胞，其中该直接重复序列包含与SEQ ID NO:30-45、77-94或122-138中任一个中列出的核苷酸序列至少80％(例如，81％、

82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的核苷酸序列。

37.如权利要求33-36中任一项所述的基因修饰的细胞,其中该间隔子序列包含约14个核苷酸至约50个核苷酸。

38.如权利要求37所述的基因修饰的细胞,其中该间隔子序列包含20至35个核苷酸。

39.如权利要求33-38中任一项所述的基因修饰的细胞,其中该细胞不包含tracrRNA。

40.如权利要求33-39中任一项所述的基因修饰的细胞,其中该细胞是真核细胞,例如哺乳动物细胞,例如人细胞。

41.如权利要求33-39中任一项所述的基因修饰的细胞,其中该细胞是原核细胞。

42.一种使如权利要求1-32中任一项所述的系统与细胞中的靶核酸结合的方法,该方法包括:

(a)提供该系统;以及

(b)将该系统递送至该细胞,

其中该细胞包含该靶核酸,其中该CRISPR相关蛋白结合该RNA指导物,并且其中该间隔子序列结合该靶核酸。

43.如权利要求42所述的方法,其中该细胞是真核细胞,例如哺乳动物细胞,例如人细胞。

44.一种修饰靶核酸的方法,该方法包括向该靶核酸递送工程化的非天然存在的CRISPR-Cas系统,该系统包括:

(a)CRISPR相关蛋白或编码该CRISPR相关蛋白的核酸,其中该CRISPR相关蛋白包含与SEQ ID NO:1-29中任一个中列出的氨基酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的氨基酸序列;以及

(b)RNA指导物,该RNA指导物包含直接重复序列和能够与该靶核酸杂交的间隔子序列;

其中该CRISPR相关蛋白能够结合该RNA指导物;并且

其中通过该CRISPR相关蛋白和RNA指导物对该靶核酸的识别导致对该靶核酸的修饰。

45.如权利要求44所述的方法,其中该CRISPR相关蛋白是与SEQ ID NO:1、SEQ ID NO:25、SEQ ID NO:26、SEQ ID NO:28或SEQ ID NO:29中列出的氨基酸序列具有至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)同一性的蛋白质。

46.如权利要求44或45所述的方法,其中该CRISPR相关蛋白能够识别PAM序列,该PAM序列包括作为5'-TTN-3'、5'-YYN-3'、5'-HHN-3'、5'-YKN-3'或5'-HBN-3'列出的核酸序列,其中N是任何核苷酸,Y是C或T,K是G或T,B是G、T或C,并且H是A、C或T。

47.如权利要求44-46中任一项所述的方法,其中该直接重复序列包含与SEQ ID NO:30-45、77-94或122-138中任一个中列出的核苷酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的核苷酸序列。

48.如权利要求44-47中任一项所述的方法,其中该间隔子序列包含约14个核苷酸至约50个核苷酸。

49.如权利要求48所述的方法,其中该间隔子序列包含20至35个核苷酸。

50.如权利要求44-49中任一项所述的方法,其中该系统不包括tracrRNA。

51.如权利要求44-50中任一项所述的方法,其中该靶核酸是DNA分子。

52.如权利要求44-51中任一项所述的方法,其中该靶核酸包含PAM序列。

53.如权利要求44-52中任一项所述的方法,其中该CRISPR相关蛋白包含非特异性核酸酶活性。

54.如权利要求44-53中任一项所述的方法,其中对该靶核酸的该修饰是双链切割事件。

55.如权利要求44-53中任一项所述的方法,其中对该靶核酸的该修饰是单链切割事件。

56.如权利要求44-53中任一项所述的方法,其中对该靶核酸的该修饰导致插入事件。

57.如权利要求44-53中任一项所述的方法,其中对该靶核酸的该修饰导致缺失事件。

58.如权利要求54-57中任一项所述的方法,其中对该靶核酸的该修饰导致细胞毒性或细胞死亡。

59.一种编辑靶核酸的方法,该方法包括使该靶核酸与如权利要求1-32中任一项所述的系统接触。

60.一种修饰靶核酸的表达的方法,该方法包括使该靶核酸与如权利要求1-32中任一项所述的系统接触。

新型CRISPR DNA靶向酶及系统

[0001] 相关申请的交叉引用

[0002] 本申请要求以下的优先权:于2019年7月11日提交的美国临时申请62/873,108、于2019年7月11日提交的美国临时申请62/873,118、于2019年9月3日提交的美国临时申请62/895,406、以及于2019年9月3日提交的美国临时申请62/895,422,将这些临时申请中的每一个的全部内容特此通过援引并入。

[0003] 序列表

[0004] 本申请含有已以ASCII格式以电子方式提交且特此通过援引以其全文并入的序列表。创建于2020年7月10日的所述ASCII副本被命名为A2186-7022WO_SL.txt,并且大小是246,736字节。

技术领域

[0005] 本披露涉及用于使用新型成簇规律间隔短回文重复序列(CRISPR)和CRISPR相关(Cas)基因来基因组编辑和调节基因表达的系统和方法。

背景技术

[0006] 基因组测序技术和分析最近进展对自然界许多不同领域的生物活动(范围从原核生物合成途径到人类病理)的遗传基础产生了重要的见解。为了充分理解和评价所产生的巨量信息,需要在基因组和表观基因组操纵的序列技术的规模、效率和易用性方面进行相应的提高。这些新型技术将加速许多领域(包括生物技术、农业和人类治疗学)的新型应用的发展。

[0007] 成簇规律间隔短回文重复序列(CRISPR)和CRISPR相关(Cas)基因(统称为CRISPR-Cas或CRISPR/Cas系统)是古细菌和细菌中针对外来遗传元件而防御特定物种的适应性免疫系统。CRISPR-Cas系统包括一组极其多样的蛋白质效应子、非编码元件以及基因座架构,其中一些实例已经被工程化并且适于产生重要的生物技术进展。

[0008] 参与宿主防御的系统的组分包括一种或多种能够修饰核酸的效应子蛋白和负责将一种或多种效应子蛋白靶向至噬菌体核酸上的特定序列的RNA指导元件。RNA指导物由CRISPR RNA(crRNA)构成,并且可能需要另外的反式激活RNA(tracrRNA)以使得能够通过一种或多种效应子蛋白进行靶向核酸操纵。crRNA由负责蛋白质与crRNA结合的直接重复和与所期望核酸靶序列互补的间隔子序列组成。CRISPR系统可以通过修饰crRNA的间隔子序列来重新编程以靶向替代性DNA或RNA靶标。

[0009] CRISPR-Cas系统可以广义地分为两类:1类系统由多种效应子蛋白构成,该多种效应子蛋白一起围绕crRNA形成复合物,并且2类系统由一种效应子蛋白构成,该一种效应子蛋白与RNA指导物复合以靶向核酸底物。2类系统的单亚基效应子组成为工程化和应用转换提供了更简单的组分集合,并且迄今为止一直是可编程效应子的重要来源。然而,除了当前的通过其独特特性实现了新型应用的CRISPR-Cas系统(诸如较小效应子和/或具有独特PAM序列要求的效应子)之外,仍然需要用于修饰核酸和多核苷酸(即,DNA、RNA或任何杂合体、

衍生物或修饰物)的另外的可编程效应子和系统。

发明内容

[0010] 本披露提供了用于新型单效应子2类CRISPR-Cas系统的非天然存在的工程化系统和组合物,首先将这些CRISPR-Cas系统从基因组数据库中计算鉴定出且随后对其进行工程化和实验验证。特别地,对这些CRISPR-Cas系统的组分的鉴定允许其在非天然环境中,例如在除这些系统最初发现于其中的那些之外的细菌中或在真核细胞(诸如哺乳动物细胞,例如人细胞)中使用。这些新的效应子在序列和功能上与现有2类CRISPR效应子的直系同源物和同源物相比是不同的。

[0011] 在一方面,本披露提供了工程化的非天然存在的CLUST.200916成簇规律间隔短回文重复序列(CRISPR)-Cas系统,这些系统包括:CRISPR相关蛋白,其中该CRISPR相关蛋白包含与SEQ ID NO:1-29中任一个中列出的氨基酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的氨基酸序列;以及RNA指导物,该RNA指导物包含直接重复序列和能够与靶核酸杂交的间隔子序列;其中该CRISPR相关蛋白能够结合该RNA指导物并且能够修饰与该间隔子序列互补的该靶核酸序列。在一方面,本披露提供了工程化的非天然存在的CLUST.200916成簇规律间隔短回文重复序列(CRISPR)-Cas系统,这些系统包括:CRISPR相关蛋白或编码该CRISPR相关蛋白的核酸,其中该CRISPR相关蛋白包含与SEQ ID NO:1-29中任一个中列出的氨基酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的氨基酸序列;以及RNA指导物或编码该RNA指导物的核酸,该RNA指导物包含直接重复序列和能够与靶核酸杂交的间隔子序列;其中该CRISPR相关蛋白能够结合该RNA指导物并且能够修饰与该间隔子序列互补的该靶核酸序列。

[0012] 在本文所述的任一系统的一些实施例中,该CRISPR相关蛋白包含至少一个(例如,一个、两个或三个)RuvC结构域或至少一个拆分型RuvC结构域。

[0013] 在本文所述的任一系统的一些实施例中,该直接重复序列包含与SEQ ID NO:30-45、77-94或122-138中任一个中列出的核苷酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的核苷酸序列。

[0014] 在一些实施例中,该CRISPR相关蛋白包含与SEQ ID NO:1的氨基酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的氨基酸序列,并且该直接重复序列包含与SEQ ID NO:30、122、77或78的核苷酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的核苷酸序列。在一些实施例中,该CRISPR相关蛋白包含与SEQ ID NO:2的氨基酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的氨基酸序列,并且该直接重复序列包含与SEQ ID NO:31或123的核苷酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同

的核苷酸序列。在一些实施例中，该CRISPR相关蛋白包含与SEQ ID NO:3的氨基酸序列至少80％(例如，81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的氨基酸序列，并且该直接重复序列包含与SEQ ID NO:31、123、79或80的核苷酸序列至少80％(例如，81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的核苷酸序列。在一些实施例中，该CRISPR相关蛋白包含与SEQ ID NO:4的氨基酸序列至少80％(例如，81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的氨基酸序列，并且该直接重复序列包含与SEQ ID NO:32或124的核苷酸序列至少80％(例如，81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的核苷酸序列。在一些实施例中，该CRISPR相关蛋白包含与SEQ ID NO:5的氨基酸序列至少80％(例如，81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的氨基酸序列，并且该直接重复序列包含与SEQ ID NO:32或124的核苷酸序列至少80％(例如，81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的核苷酸序列。在一些实施例中，该CRISPR相关蛋白包含与SEQ ID NO:6的氨基酸序列至少80％(例如，81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的氨基酸序列，并且该直接重复序列包含与SEQ ID NO:33或125的核苷酸序列至少80％(例如，81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的核苷酸序列。在一些实施例中，该CRISPR相关蛋白包含与SEQ ID NO:7的氨基酸序列至少80％(例如，81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的氨基酸序列，并且该直接重复序列包含与SEQ ID NO:34或126的核苷酸序列至少80％(例如，81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的核苷酸序列。在一些实施例中，该CRISPR相关蛋白包含与SEQ ID NO:8的氨基酸序列至少80％(例如，81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的氨基酸序列，并且该直接重复序列包含与SEQ ID NO:35或127的核苷酸序列至少80％(例如，81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的核苷酸序列。在一些实施例中，该CRISPR相关蛋白包含与SEQ ID NO:9的氨基酸序列至少80％(例如，81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的氨基酸序列，并且该直接重复序列包含与SEQ ID NO:35或127的核苷酸序列至少80％(例如，81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的核苷酸序列。在一些实施例中，该CRISPR相关蛋白包含与SEQ ID NO:10的氨基酸序列至少80％(例如，81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的氨基酸序列，并且该直接重复序列包含与SEQ ID NO:36或128的核苷酸序列至少80％(例如，81％、82％、83％、84％、85％、

86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的核苷酸序列。在一些实施例中,该CRISPR相关蛋白包含与SEQ ID NO:11的氨基酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的氨基酸序列,并且该直接重复序列包含与SEQ ID NO:36或128的核苷酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的核苷酸序列。在一些实施例中,该CRISPR相关蛋白包含与SEQ ID NO:12的氨基酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的氨基酸序列,并且该直接重复序列包含与SEQ ID NO:37或129的核苷酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的核苷酸序列。在一些实施例中,该CRISPR相关蛋白包含与SEQ ID NO:13的氨基酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的氨基酸序列,并且该直接重复序列包含与SEQ ID NO:38或130的核苷酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的核苷酸序列。在一些实施例中,该CRISPR相关蛋白包含与SEQ ID NO:14的氨基酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的氨基酸序列,并且该直接重复序列包含与SEQ ID NO:38或130的核苷酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的核苷酸序列。在一些实施例中,该CRISPR相关蛋白包含与SEQ ID NO:15的氨基酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的氨基酸序列,并且该直接重复序列包含与SEQ ID NO:38或130的核苷酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的核苷酸序列。在一些实施例中,该CRISPR相关蛋白包含与SEQ ID NO:16的氨基酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的氨基酸序列,并且该直接重复序列包含与SEQ ID NO:39或131的核苷酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的核苷酸序列。在一些实施例中,该CRISPR相关蛋白包含与SEQ ID NO:17的氨基酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的氨基酸序列,并且该直接重复序列包含与SEQ ID NO:31或123的核苷酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的核苷酸序列。在一些实施例中,该CRISPR相关蛋白包含与SEQ ID NO:18的氨基酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的氨基酸序列,并且该直接重复序列

包含与SEQ ID NO:40或132的核苷酸序列至少80％(例如,81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的核苷酸序列。在一些实施例中,该CRISPR相关蛋白包含与SEQ ID NO:19的氨基酸序列至少80％(例如,81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的氨基酸序列,并且该直接重复序列包含与SEQ ID NO:41或133的核苷酸序列至少80％(例如,81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的核苷酸序列。在一些实施例中,该CRISPR相关蛋白包含与SEQ ID NO:20的氨基酸序列至少80％(例如,81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的氨基酸序列,并且该直接重复序列包含与SEQ ID NO:41或133的核苷酸序列至少80％(例如,81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的核苷酸序列。在一些实施例中,该CRISPR相关蛋白包含与SEQ ID NO:21的氨基酸序列至少80％(例如,81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的氨基酸序列,并且该直接重复序列包含与SEQ ID NO:39或131的核苷酸序列至少80％(例如,81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的核苷酸序列。在一些实施例中,该CRISPR相关蛋白包含与SEQ ID NO:22的氨基酸序列至少80％(例如,81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的氨基酸序列,并且该直接重复序列包含与SEQ ID NO:39、131、81或82的核苷酸序列至少80％(例如,81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的核苷酸序列。在一些实施例中,该CRISPR相关蛋白包含与SEQ ID NO:23的氨基酸序列至少80％(例如,81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的氨基酸序列,并且该直接重复序列包含与SEQ ID NO:42、134、83或84的核苷酸序列至少80％(例如,81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的核苷酸序列。在一些实施例中,该CRISPR相关蛋白包含与SEQ ID NO:24的氨基酸序列至少80％(例如,81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的氨基酸序列,并且该直接重复序列包含与SEQ ID NO:43、135、85或86的核苷酸序列至少80％(例如,81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的核苷酸序列。在一些实施例中,该CRISPR相关蛋白包含与SEQ ID NO:25的氨基酸序列至少80％(例如,81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的氨基酸序列,并且该直接重复序列包含与SEQ ID NO:43、135、87或88的核苷酸序列至少80％(例如,81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的核苷酸序列。在一些实施例中,该CRISPR相关蛋白包含与SEQ ID NO:26的氨基酸序列至少80％(例如,81％、82％、83％、84％、85％、

86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的氨基酸序列,并且该直接重复序列包含与SEQ ID NO:44、136、89或90的核苷酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的核苷酸序列。在一些实施例中,该CRISPR相关蛋白包含与SEQ ID NO:27的氨基酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的氨基酸序列,并且该直接重复序列包含与SEQ ID NO:45或137的核苷酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的核苷酸序列。在一些实施例中,该CRISPR相关蛋白包含与SEQ ID NO:28的氨基酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的氨基酸序列,并且该直接重复序列包含与SEQ ID NO:46、138、91或92的核苷酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的核苷酸序列。在一些实施例中,该CRISPR相关蛋白包含与SEQ ID NO:29的氨基酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的氨基酸序列,并且该直接重复序列包含与SEQ ID NO:45、137、93或94的核苷酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的核苷酸序列。

[0015] 在本文所述的任一系统的一些实施例中,该CRISPR相关蛋白是与SEQ ID NO:1(CLUST.200916 3300013232)、SEQ ID NO:26(CLUST.200916 SRR6837570)、SEQ ID NO:28(CLUST.200916SRR6837575)、SEQ ID NO:29(CLUST.200916 SRR6837577)或SEQ ID NO:25(CLUST.200916 SRR6837569)中列出的氨基酸序列具有至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)同一性的蛋白质。

[0016] 在本文所述的任一系统的一些实施例中,该CRISPR相关蛋白能够识别原型间隔子相邻基序(PAM)序列,其中该PAM序列包括核酸序列,包括作为5'-TTN-3'、5'-YYN-3'、5'-HHN-3'、5'-YKN-3'或5'-HBN-3'列出的核酸序列,其中N是任何核苷酸(例如,A、G、T或C),Y是C或T,K是G或T,B是G、T或C,并且H是A、C或T。

[0017] 在一些实施例中,该CRISPR相关蛋白包含与SEQ ID NO:1的氨基酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的氨基酸序列,并且该PAM序列包含作为5'-TTN-3'、5'-YYN-3'或5'-HHN-3'列出的核酸序列。在一些实施例中,该CRISPR相关蛋白包含与SEQ ID NO:22的氨基酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)相同的氨基酸序列,并且该PAM序列包含作为5'-TTN-3'、5'-YYN-3'或5'-HHN-3'列出的核酸序列。在一些实施例中,该CRISPR相关蛋白包含与SEQ ID NO:24的氨基酸序列至少80%(例如,81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、

97％、98％、99％或100％)相同的氨基酸序列，并且该PAM序列包含作为5’-TTN-3’或5’-YKN-3’列出的核酸序列。在一些实施例中，该CRISPR相关蛋白包含与SEQ ID NO:25的氨基酸序列至少80％(例如，81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的氨基酸序列，并且该PAM序列包含作为5’-TTN-3’列出的核酸序列。在一些实施例中，该CRISPR相关蛋白包含与SEQ ID NO:26的氨基酸序列至少80％(例如，81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的氨基酸序列，并且该PAM序列包含作为5’-TTN-3’、5’-YYN-3’或5’-HBN-3’列出的核酸序列。在一些实施例中，该CRISPR相关蛋白包含与SEQ ID NO:28的氨基酸序列至少80％(例如，81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的氨基酸序列，并且该PAM序列包含作为5’-TTN-3’或5’-YKN-3’列出的核酸序列。在一些实施例中，该CRISPR相关蛋白包含与SEQ ID NO:29的氨基酸序列至少80％(例如，81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的氨基酸序列，并且该PAM序列包含作为5’-TTN-3’或5’-HHN-3’列出的核酸序列。

[0018] 在本文所述的任一系统的一些实施例中，该RNA指导物的该间隔子序列包含约14个核苷酸至约50个核苷酸。在本文所述的任一系统的一些实施例中，该RNA指导物的该间隔子序列包含20至35个核苷酸。

[0019] 在本文所述的任一系统的一些实施例中，该CRISPR相关蛋白包含催化残基(例如，天冬氨酸或谷氨酸)。在本文所述的任一系统的一些实施例中，该CRISPR相关蛋白切割该靶核酸。在本文所述的任一系统的一些实施例中，该CRISPR相关蛋白进一步包含肽标签、荧光蛋白、碱基编辑结构域、DNA甲基化结构域、组蛋白残基修饰结构域、定位因子、转录修饰因子、光门控因子、化学诱导型因子或染色质可视化因子。

[0020] 在本文所述的任一系统的一些实施例中，编码该CRISPR相关蛋白的该核酸经密码子优化以在细胞，例如真核细胞，例如哺乳动物细胞，例如人细胞中表达。在本文所述的任一系统的一些实施例中，编码该CRISPR相关蛋白的该核酸可操作地连接至启动子。在本文所述的任一系统的一些实施例中，编码该CRISPR相关蛋白的该核酸在载体中。在一些实施例中，该载体包括逆转录病毒载体、慢病毒载体、噬菌体载体、腺病毒载体、腺相关载体或单纯疱疹载体。

[0021] 在本文所述的任一系统的一些实施例中，该靶核酸是DNA分子。在本文所述的任一系统的一些实施例中，该靶核酸包含PAM序列。

[0022] 在本文所述的任一系统的一些实施例中，该CRISPR相关蛋白具有非特异性核酸酶活性。

[0023] 在本文所述的任一系统的一些实施例中，通过该CRISPR相关蛋白和RNA指导物对该靶核酸的识别导致对该靶核酸的修饰。在本文所述的任一系统的一些实施例中，对该靶核酸的该修饰是双链切割事件。在本文所述的任一系统的一些实施例中，对该靶核酸的该修饰是单链切割事件。在本文所述的任一系统的一些实施例中，对该靶核酸的该修饰导致插入事件。在本文所述的任一系统的一些实施例中，对该靶核酸的该修饰导致缺失事件。在本文所述的任一系统的一些实施例中，对该靶核酸的该修饰导致细胞毒性或细胞死亡。

[0024] 在本文所述的任一系统的一些实施例中，该系统进一步包括供体模板核酸。在本文所述的任一系统的一些实施例中，该供体模板核酸是DNA分子。在本文所述的任一系统的一些实施例中，其中该供体模板核酸是RNA分子。

[0025] 在本文所述的任一系统的一些实施例中，该系统不包括tracrRNA。在本文所述的任一系统的一些实施例中，该系统任选地包括tracrRNA。在本文所述的任一系统的一些实施例中，该CRISPR相关蛋白是自加工的。

[0026] 在本文所述的任一系统的一些实施例中，该系统存在于递送组合物中，该递送组合物包含纳米颗粒、脂质体、外来体、微泡或基因枪。

[0027] 在本文所述的任一系统的一些实施例中，这些系统在细胞内。在一些实施例中，该细胞是真核细胞。在一些实施例中，该细胞是哺乳动物细胞。在一些实施例中，该细胞是人细胞。在一些实施例中，该细胞是原核细胞。

[0028] 在另一方面，本披露提供了一种基因修饰的细胞，其中该细胞包含：CRISPR相关蛋白，其中该CRISPR相关蛋白包含与SEQ ID NO:1-29中任一个中列出的氨基酸序列至少80％(例如，81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的氨基酸序列；以及RNA指导物，该RNA指导物包含直接重复序列和能够与靶核酸杂交的间隔子序列。在另一方面，本披露提供了一种基因修饰的细胞，其中该细胞包含：CRISPR相关蛋白或编码该CRISPR相关蛋白的核酸，其中该CRISPR相关蛋白包含与SEQ ID NO:1-29中任一个中列出的氨基酸序列至少80％(例如，81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的氨基酸序列；以及RNA指导物或编码该RNA指导物的核酸，该RNA指导物包含直接重复序列和能够与靶核酸杂交的间隔子序列。

[0029] 在本文所述的任一细胞的一些实施例中，该CRISPR相关蛋白包含至少一个(例如，一个、两个或三个)RuvC结构域或至少一个拆分型RuvC结构域。

[0030] 在本文所述的任一细胞的一些实施例中，该CRISPR相关蛋白是与SEQ ID NO:1、SEQ ID NO:25、SEQ ID NO:26、SEQ ID NO:28或SEQ ID NO:29中列出的氨基酸序列具有至少80％(例如，81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)同一性的蛋白质。

[0031] 在本文所述的任一细胞的一些实施例中，该CRISPR相关蛋白能够识别PAM序列，该PAM序列包括作为5'-TTN-3'、5'-YYN-3'、5'-HHN-3'、5'-YKN-3'或5'-HBN-3'列出的核酸序列，其中N是任何核苷酸(例如，A、G、T或C)，Y是C或T，K是G或T，B是G、T或C，并且H是A、C或T。

[0032] 在本文所述的任一细胞的一些实施例中，该直接重复序列包含与SEQ ID NO:30-45、77-94或122-138中任一个中列出的核苷酸序列至少80％(例如，81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的核苷酸序列。

[0033] 在本文所述的任一细胞的一些实施例中，该间隔子序列包含约14个核苷酸至约50个核苷酸。在本文所述的任一细胞的一些实施例中，该间隔子序列包含20至35个核苷酸。

[0034] 在本文所述的任一细胞的一些实施例中，该CRISPR相关蛋白包含催化残基(例如，天冬氨酸或谷氨酸)。在本文所述的任一细胞的一些实施例中，该CRISPR相关蛋白切割该靶核酸。在本文所述的任一细胞的一些实施例中，该CRISPR相关蛋白进一步包含肽标签、荧光

蛋白、碱基编辑结构域、DNA甲基化结构域、组蛋白残基修饰结构域、定位因子、转录修饰因子、光门控因子、化学诱导型因子或染色质可视化因子。

[0035] 在本文所述的任一细胞的一些实施例中，编码该CRISPR相关蛋白的该核酸经密码子优化以在细胞，例如真核细胞，例如哺乳动物细胞，例如人细胞中表达。在本文所述的任一细胞的一些实施例中，编码该CRISPR相关蛋白的该核酸可操作地连接至启动子。在本文所述的任一细胞的一些实施例中，编码该CRISPR相关蛋白的该核酸在载体中。在一些实施例中，该载体包括逆转录病毒载体、慢病毒载体、噬菌体载体、腺病毒载体、腺相关载体或单纯疱疹载体。

[0036] 在本文所述的任一细胞的一些实施例中，该细胞不包含tracrRNA。在本文所述的任一细胞的一些实施例中，该细胞任选地包括tracrRNA。在本文所述的任一细胞的一些实施例中，该CRISPR相关蛋白是自加工的。

[0037] 在本文所述的任一细胞的一些实施例中，该细胞是真核细胞。在本文所述的任一细胞的一些实施例中，该细胞是哺乳动物细胞。在本文所述的任一细胞的一些实施例中，该细胞是人细胞。在本文所述的任一细胞的一些实施例中，该细胞是原核细胞。

[0038] 在本文所述的任一细胞的一些实施例中，该靶核酸是DNA分子。在本文所述的任一细胞的一些实施例中，该靶核酸包含PAM序列。

[0039] 在本文所述的任一细胞的一些实施例中，该CRISPR相关蛋白具有非特异性核酸酶活性。

[0040] 在本文所述的任一细胞的一些实施例中，通过该CRISPR相关蛋白和RNA指导物对该靶核酸的识别导致对该靶核酸的修饰。在本文所述的任一细胞的一些实施例中，对该靶核酸的该修饰是双链切割事件。在本文所述的任一细胞的一些实施例中，对该靶核酸的该修饰是单链切割事件。在本文所述的任一细胞的一些实施例中，对该靶核酸的该修饰导致插入事件。在本文所述的任一细胞的一些实施例中，对该靶核酸的该修饰导致缺失事件。在本文所述的任一细胞的一些实施例中，对该靶核酸的该修饰导致细胞毒性或细胞死亡。

[0041] 在另一方面，本披露提供了一种使本文所述的系统与细胞中的靶核酸结合的方法，该方法包括：(a)提供该系统；以及(b)将该系统递送至该细胞，其中该细胞包含该靶核酸，其中该CRISPR相关蛋白结合该RNA指导物，并且其中该间隔子序列结合该靶核酸。在一些实施例中，该细胞是真核细胞，例如哺乳动物细胞，例如人细胞。

[0042] 在另一方面，本披露提供了修饰靶核酸的方法，该方法包括向该靶核酸递送工程化的非天然存在的CRISPR-Cas系统，该系统包括：CRISPR相关蛋白，其中该CRISPR相关蛋白包含与SEQ ID NO:1-29中任一个中列出的氨基酸序列至少80％(例如，81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的氨基酸序列；以及RNA指导物，该RNA指导物包含直接重复序列和能够与该靶核酸杂交的间隔子序列；其中该CRISPR相关蛋白能够结合该RNA指导物；并且其中通过该CRISPR相关蛋白和RNA指导物对该靶核酸的识别导致对该靶核酸的修饰。在另一方面，本披露提供了修饰靶核酸的方法，该方法包括向该靶核酸递送工程化的非天然存在的CRISPR-Cas系统，该系统包括：CRISPR相关蛋白或编码该CRISPR相关蛋白的核酸，其中该CRISPR相关蛋白包含与SEQ ID NO:1-29中任一个中列出的氨基酸序列至少80％(例如，81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、

96％、97％、98％、99％或100％)相同的氨基酸序列;以及RNA指导物或编码该RNA指导物的核酸,该RNA指导物包含直接重复序列和能够与该靶核酸杂交的间隔子序列;其中该CRISPR相关蛋白能够结合该RNA指导物;并且其中通过该CRISPR相关蛋白和RNA指导物对该靶核酸的识别导致对该靶核酸的修饰。

[0043] 在本文所述的任一方法的一些实施例中,该CRISPR相关蛋白是与SEQ ID NO:1、SEQ ID NO:25、SEQ ID NO:26、SEQ ID NO:28或SEQ ID NO:29中列出的氨基酸序列具有至少80％(例如,81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)同一性的蛋白质。

[0044] 在本文所述的任一方法的一些实施例中,该CRISPR相关蛋白能够识别PAM序列,该PAM序列包括作为5'-TTN-3'、5'-YYN-3'、5'-HHN-3'、5'-YKN-3'或5'-HBN-3'列出的核酸序列,其中N是任何核苷酸,Y是C或T,K是G或T,B是G、T或C,并且H是A、C或T。

[0045] 在本文所述的任一方法的一些实施例中,该直接重复序列包含与SEQ ID NO:30-45、77-94或122-138中任一个中列出的核苷酸序列至少80％(例如,81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％)相同的核苷酸序列。

[0046] 在本文所述的任一方法的一些实施例中,该间隔子序列包含约14个核苷酸至约50个核苷酸。在本文所述的任一方法的一些实施例中,该间隔子序列包含20至35个核苷酸。

[0047] 在本文所述的任一方法的一些实施例中,该CRISPR相关蛋白包含催化残基(例如,天冬氨酸或谷氨酸)。在本文所述的任一方法的一些实施例中,该CRISPR相关蛋白切割该靶核酸。在本文所述的任一方法的一些实施例中,该CRISPR相关蛋白进一步包含肽标签、荧光蛋白、碱基编辑结构域、DNA甲基化结构域、组蛋白残基修饰结构域、定位因子、转录修饰因子、光门控因子、化学诱导型因子或染色质可视化因子。

[0048] 在本文所述的任一方法的一些实施例中,编码该CRISPR相关蛋白的该核酸经密码子优化以在细胞,例如真核细胞,例如哺乳动物细胞,例如人细胞中表达。在本文所述的任一方法的一些实施例中,编码该CRISPR相关蛋白的该核酸可操作地连接至启动子。在本文所述的任一方法的一些实施例中,编码该CRISPR相关蛋白的该核酸在载体中。在一些实施例中,该载体包括逆转录病毒载体、慢病毒载体、噬菌体载体、腺病毒载体、腺相关载体或单纯疱疹载体。

[0049] 在本文所述的任一方法的一些实施例中,该细胞不包含tracrRNA。在本文所述的任一方法的一些实施例中,该细胞任选地包括tracrRNA。

[0050] 在本文所述的任一方法的一些实施例中,该靶核酸是DNA分子。在本文所述的任一方法的一些实施例中,该靶核酸包含PAM序列。

[0051] 在本文所述的任一方法的一些实施例中,该CRISPR相关蛋白具有非特异性核酸酶活性。

[0052] 在本文所述的任一方法的一些实施例中,对该靶核酸的该修饰是双链切割事件。在本文所述的任一方法的一些实施例中,对该靶核酸的该修饰是单链切割事件。在本文所述的任一方法的一些实施例中,对该靶核酸的该修饰导致插入事件。在本文所述的任一方法的一些实施例中,对该靶核酸的该修饰导致缺失事件。在本文所述的任一方法的一些实施例中,对该靶核酸的该修饰导致细胞毒性或细胞死亡。

[0053] 在另一方面，本披露提供了一种编辑靶核酸的方法，该方法包括使该靶核酸与本文所述的系统接触。在另一方面，本披露提供了一种修饰靶核酸的表达的方法，该方法包括使该靶核酸与本文所述的系统接触。在另一方面，本披露提供了一种使有效载荷核酸在靶核酸的位点处靶向插入的方法，该方法包括使该靶核酸与本文所述的系统接触。在另一方面，本披露提供了一种从靶核酸处的位点靶向切除有效载荷核酸的方法，该方法包括使该靶核酸与本文所述的系统接触。在另一方面，本披露提供了一种在识别DNA靶核酸后非特异性降解单链DNA的方法，该方法包括使该靶核酸与本文所述的系统接触。

[0054] 在另一方面，本披露提供了一种检测样品中的靶核酸的方法，该方法包括：(a)使该样品与本文所述的系统和经标记的报告核酸接触，其中该间隔子序列与该靶核酸的杂交引起该经标记的报告核酸的切割；以及(b)测量由该经标记的报告核酸的切割产生的可检测信号，从而检测该样品中该靶核酸的存在。

[0055] 在本文所提供的任一系统或方法的一些实施例中，该接触包括直接接触或间接接触。在本文所提供的任一系统或方法的一些实施例中，间接接触包括在这样的条件下施用编码本文所述的RNA指导物或CRISPR相关蛋白的一种或多种核酸，这些条件允许产生该RNA指导物和/或CRISPR相关蛋白。在本文所提供的任一系统或方法的一些实施例中，接触包括体内接触或体外接触。在本文所提供的任一系统或方法的一些实施例中，使靶核酸与该系统接触包括在这样的条件下使包含该核酸的细胞与该系统接触，这些条件允许该CRISPR相关蛋白和该指导RNA到达该靶核酸。在本文所提供的任一系统或方法的一些实施例中，使体内细胞与该系统接触包括在这样的条件下将该系统施用至包括该细胞的受试者，这些条件允许该CRISPR相关蛋白和该指导RNA到达该细胞或在该细胞中产生。

[0056] 在另一方面，本披露提供了一种本文所提供的系统，用于在以下体外或离体方法中使用：(a)靶向并编辑靶核酸；(b)在识别该核酸后非特异性降解单链核酸；(c)在识别双链靶标的间隔子互补链后靶向该双链靶标的非间隔子互补链并使其产生切口；(d)靶向并切割双链靶核酸；(e)检测样品中的靶核酸；(f)特异性编辑双链核酸；(g)对双链核酸进行碱基编辑；(h)在细胞中诱导基因型特异性或转录状态特异性的细胞死亡或休眠；(i)在双链核酸靶标中产生插入缺失；(j)将序列插入双链核酸靶标中；或(k)缺失或倒置双链核酸靶标中的序列。

[0057] 本文所述的效应子提供了另外的特征，这些另外的特征包括但不限于1)新型核酸编辑特性和控制机制，2)更小的大小以便在递送策略中具有更大的通用性，3)基因型触发的细胞过程诸如细胞死亡，和4)可编程的RNA指导的DNA插入、切除和转移，以及5)先前存在的免疫力通过非人共生来源产生的分化谱。参见例如实例1-5和图1A-37。将本文所述的新型DNA靶向系统添加至基因组和表观基因组操纵的技术工具箱中使得能够广泛应用于特定的编程扰动。

[0058] 本发明的其他特征和优点通过以下详细描述以及权利要求将会变得清楚。

附图说明

[0059] 这些图是表示被称为CLUST.200916的蛋白质聚簇的分析结果的一系列示意图。

[0060] 图1A是实例2中所述的体内阴性选择筛选测定的组分的示意图。设计了CRISPR阵列文库，这些阵列文库包括侧接有两个DR并由J23119表达、从pACYC184或大肠杆菌必需基

因的两条链上均匀取样的非代表性间隔子。

[0061] 图1B是实例2中所述的体内阴性选择筛选工作流的示意图。将CRISPR阵列文库克隆至效应子质粒中。将效应子质粒和非编码质粒转化至大肠杆菌中，之后进行生长以便阴性选择赋予对来自pACYC184或大肠杆菌必需基因的转录物的干扰的CRISPR阵列。使用效应子质粒的靶向测序来鉴定耗减的CRISPR阵列。可以进一步进行小型RNA测序以鉴定成熟crRNA和潜在tracrRNA要求。

[0062] 图2A是示出了CLUST.200916效应子的RuvC和锌指结构域的示意图，该示意图是基于表7中所示序列的共有序列。

[0063] 图2B是示出了用于代表性CLUST.200916基因座的保守效应子(e_A)和CRISPR阵列元件的示意性序列表示。

[0064] 图2C是示出了用于CLUST.200916的CRISPR直接重复元件的实例的多个序列比对的一系列序列。图2C按照出现的顺序分别披露了SEQ ID NO:162、30、41、41、31、31、31-32、32-35、35-36、36、40、39、39、39、39、39、39、163、163、43、43、46、44、42、45、45、164、164、164、38、38、38和37。

[0065] 图2D是CLUST.200916效应子蛋白的系统发育树的示意图。

[0066] 图2E是CLUST.200916效应子蛋白的多序列比对的示意图，其中指示了RuvC结构域的保守催化残基的位置和锌指结构域的保守催化残基的位置。

[0067] 图3是CLUST.200916 3300013232(SEQ ID NO:1中列出的效应子)的图，该图示出了在具有非编码序列的情况下，工程化组合物针对靶向pACYC184的间隔子和直接重复转录取向的耗减活性程度。在呈“正向”取向(5’-CAAC…AGAC-[间隔子]-3’)的直接重复的情况下和在呈“反向”取向(5’-GTCT…GTTG-[间隔子]-3’)的直接重复的情况下的耗减程度分别以实线和虚线描绘。

[0068] 图4A是按pACYC184质粒上的位置示出了CLUST.200916 3300013232(具有非编码序列)的耗减靶标和非耗减靶标的密度的图形表示。图4B是按大肠杆菌菌株E.Cloni上的位置示出了CLUST.200916 3300013232(具有非编码序列)的耗减靶标和非耗减靶标的密度的图形表示。单独地并与注释基因的取向相关地示出了顶链和底链上的靶标。谱带的幅度指示耗减程度，其中较轻的谱带接近于3的命中阈值。两种梯度是示出了相对转录物丰度的RNA测序的热图。

[0069] 图5是作为CLUST.200916 3300013232(具有非编码序列)的PAM序列的预测的E.Cloni中侧接耗减靶标的序列的WebLogo。

[0070] 图6是CLUST.200916 3300013232(SEQ ID NO:1中列出的效应子)的图，该图示出了在没有非编码序列的情况下，工程化组合物针对靶向pACYC184的间隔子和直接重复转录取向的耗减活性程度。在呈“正向”取向(5’-CAAC…AGAC-[间隔子]-3’)的直接重复的情况下和在呈“反向”取向(5’-GTCT…GTTG-[间隔子]-3’)的直接重复的情况下的耗减程度分别以实线和虚线描绘。

[0071] 图7A是按pACYC184质粒上的位置示出了CLUST.200916 3300013232(没有非编码序列)的耗减靶标和非耗减靶标的密度的图形表示。图7B是按大肠杆菌菌株E.Cloni上的位置示出了CLUST.200916 3300013232(没有非编码序列)的耗减靶标和非耗减靶标的密度的图形表示。单独地并与注释基因的取向相关地示出了顶链和底链上的靶标。谱带的幅度指

示耗减程度,其中较轻的谱带接近于3的命中阈值。梯度是示出了相对转录物丰度的RNA测序的热图。

[0072] 图8是作为CLUST.200916 3300013232(没有非编码序列)的PAM序列的预测的E.Cloni中侧接耗减靶标的序列的WebLogo。

[0073] 图9是CLUST.200916 SRR6837570(SEQ ID NO:26中列出的效应子)的图,该图示出了在具有非编码序列的情况下,工程化组合物针对靶向pACYC184的间隔子和直接重复转录取向的耗减活性程度。在呈“正向”取向(5’-GTTC…GCGC-[间隔子]-3’)的直接重复的情况下和在呈“反向”取向(5’-GCGC…GAAC-[间隔子]-3’)的直接重复的情况下的耗减程度分别以实线和虚线描绘。

[0074] 图10A是按pACYC184质粒上的位置示出了CLUST.200916 SRR6837570(具有非编码序列)的耗减靶标和非耗减靶标的密度的图形表示。图10B是按大肠杆菌菌株E.Cloni上的位置示出了CLUST.200916 SRR6837570(具有非编码序列)的耗减靶标和非耗减靶标的密度的图形表示。单独地并与注释基因的取向相关地示出了顶链和底链上的靶标。谱带的幅度指示耗减程度,其中较轻的谱带接近于3的命中阈值。梯度是示出了相对转录物丰度的RNA测序的热图。

[0075] 图11是作为CLUST.200916 SRR6837570(具有非编码序列)的PAM序列的预测的E.Cloni中侧接耗减靶标的序列的WebLogo。

[0076] 图12是CLUST.200916 SRR6837570(SEQ ID NO:26中列出的效应子)的图,该图示出了在没有非编码序列的情况下,工程化组合物针对靶向pACYC184的间隔子和直接重复转录取向的耗减活性程度。在呈“正向”取向(5’-GTTC…GCGC-[间隔子]-3’)的直接重复的情况下和在呈“反向”取向(5’-GCGC…GAAC-[间隔子]-3’)的直接重复的情况下的耗减程度分别以实线和虚线描绘。

[0077] 图13A是按pACYC184质粒上的位置示出了CLUST.200916 SRR6837570(没有非编码序列)的耗减靶标和非耗减靶标的密度的图形表示。图13B是按大肠杆菌菌株E.Cloni上的位置示出了CLUST.200916 SRR6837570(没有非编码序列)的耗减靶标和非耗减靶标的密度的图形表示。单独地并与注释基因的取向相关地示出了顶链和底链上的靶标。谱带的幅度指示耗减程度,其中较轻的谱带接近于3的命中阈值。梯度是示出了相对转录物丰度的RNA测序的热图。

[0078] 图14是作为CLUST.200916 SRR6837570(没有非编码序列)的PAM序列的预测的E.Cloni中侧接耗减靶标的序列的WebLogo。

[0079] 图15是CLUST.200916 SRR6837575(SEQ ID NO:28中列出的效应子)的图,该图示出了在具有非编码序列的情况下,工程化组合物针对靶向pACYC184的间隔子和直接重复转录取向的耗减活性程度。在呈“正向”取向(5’-CCAT…AGAC-[间隔子]-3’)的直接重复的情况下和在呈“反向”取向(5’-GTCT…ATGG-[间隔子]-3’)的直接重复的情况下的耗减程度分别以实线和虚线描绘。

[0080] 图16A是按pACYC184质粒上的位置示出了CLUST.200916 SRR6837575(具有非编码序列)的耗减靶标和非耗减靶标的密度的图形表示。图16B是按大肠杆菌菌株E.Cloni上的位置示出了CLUST.200916 SRR6837575(具有非编码序列)的耗减靶标和非耗减靶标的密度的图形表示。单独地并与注释基因的取向相关地示出了顶链和底链上的靶标。谱带的幅度

指示耗减程度,其中较轻的谱带接近于3的命中阈值。梯度是示出了相对转录物丰度的RNA测序的热图。

[0081] 图17是作为CLUST.200916 SRR6837575(具有非编码序列)的PAM序列的预测的E.Cloni中侧接耗减靶标的序列的WebLogo。

[0082] 图18是CLUST.200916 SRR6837575(SEQ ID NO:28中列出的效应子)的图,该图示出了在没有非编码序列的情况下,工程化组合物针对靶向pACYC184的间隔子和直接重复转录取向的耗减活性程度。在呈“正向”取向(5’-CCAT…AGAC-[间隔子]-3’)的直接重复的情况下和在呈“反向”取向(5’-GTCT…ATGG-[间隔子]-3’)的直接重复的情况下的耗减程度分别以实线和虚线描绘。

[0083] 图19A是按pACYC184质粒上的位置示出了CLUST.200916 SRR6837575(没有非编码序列)的耗减靶标和非耗减靶标的密度的图形表示。图19B是按大肠杆菌菌株E.Cloni上的位置示出了CLUST.200916 SRR6837575(没有非编码序列)的耗减靶标和非耗减靶标的密度的图形表示。单独地并与注释基因的取向相关地示出了顶链和底链上的靶标。谱带的幅度指示耗减程度,其中较轻的谱带接近于3的命中阈值。梯度是示出了相对转录物丰度的RNA测序的热图。

[0084] 图20是作为CLUST.200916 SRR6837575(没有非编码序列)的PAM序列的预测的E.Cloni中侧接耗减靶标的序列的WebLogo。

[0085] 图21是CLUST.200916 SRR6837577(SEQ ID NO:29中列出的效应子)的图,该图示出了在具有非编码序列的情况下,工程化组合物针对靶向pACYC184的间隔子和直接重复转录取向的耗减活性程度。在呈“正向”取向(5’-GTCG…CGAC-[间隔子]-3’)的直接重复的情况下和在呈“反向”取向(5’-GTCG…CGAC-[间隔子]-3’)的直接重复的情况下的耗减程度分别以实线和虚线描绘。

[0086] 图22A是按pACYC184质粒上的位置示出了CLUST.200916 SRR6837577(具有非编码序列)的耗减靶标和非耗减靶标的密度的图形表示。图22B是按大肠杆菌菌株E.Cloni上的位置示出了CLUST.200916 SRR6837577(具有非编码序列)的耗减靶标和非耗减靶标的密度的图形表示。单独地并与注释基因的取向相关地示出了顶链和底链上的靶标。谱带的幅度指示耗减程度,其中较轻的谱带接近于3的命中阈值。梯度是示出了相对转录物丰度的RNA测序的热图。

[0087] 图23是作为CLUST.200916 SRR6837577(具有非编码序列)的PAM序列的预测的E.Cloni中侧接耗减靶标的序列的WebLogo。

[0088] 图24是CLUST.200916 SRR6837577(SEQ ID NO:29中列出的效应子)的图,该图示出了在没有非编码序列的情况下,工程化组合物针对靶向pACYC184的间隔子和直接重复转录取向的耗减活性程度。在呈“正向”取向(5’-GTCG…CGAC-[间隔子]-3’)的直接重复的情况下和在呈“反向”取向(5’-GTCG…CGAC-[间隔子]-3’)的直接重复的情况下的耗减程度分别以实线和虚线描绘。

[0089] 图25A是按pACYC184质粒上的位置示出了CLUST.200916 SRR6837577(没有非编码序列)的耗减靶标和非耗减靶标的密度的图形表示。图25B是按大肠杆菌菌株E.Cloni上的位置示出了CLUST.200916 SRR6837577(没有非编码序列)的耗减靶标和非耗减靶标的密度的图形表示。单独地并与注释基因的取向相关地示出了顶链和底链上的靶标。谱带的幅度

指示耗减程度,其中较轻的谱带接近于3的命中阈值。梯度是示出了相对转录物丰度的RNA测序的热图。

[0090] 图26是作为CLUST.200916 SRR6837577(没有非编码序列)的PAM序列的预测的E.Cloni中侧接耗减靶标的序列的WebLogo。

[0091] 图27是CLUST.200916 SRR6837569(SEQ ID NO:25中列出的效应子)的图,该图示出了在具有非编码序列的情况下,工程化组合物针对靶向pACYC184的间隔子和直接重复转录取向的耗减活性程度。在呈“正向”取向(5’-GTCT…CAGG-[间隔子]-3’)的直接重复的情况下和在呈“反向”取向(5’-CCTG…AGAC-[间隔子]-3’)的直接重复的情况下的耗减程度分别以实线和虚线描绘。

[0092] 图28A是按pACYC184质粒上的位置示出了CLUST.200916 SRR6837569(具有非编码序列)的耗减靶标和非耗减靶标的密度的图形表示。图28B是按大肠杆菌菌株E.Cloni上的位置示出了CLUST.200916 SRR6837569(具有非编码序列)的耗减靶标和非耗减靶标的密度的图形表示。单独地并与注释基因的取向相关地示出了顶链和底链上的靶标。谱带的幅度指示耗减程度,其中较轻的谱带接近于3的命中阈值。梯度是示出了相对转录物丰度的RNA测序的热图。

[0093] 图29是作为CLUST.200916 SRR6837569(具有非编码序列)的PAM序列的预测的E.Cloni中侧接耗减靶标的序列的WebLogo。

[0094] 图30是CLUST.200916 SRR6837569(SEQ ID NO:25中列出的效应子)的图,该图示出了在没有非编码序列的情况下,工程化组合物针对靶向pACYC184的间隔子和直接重复转录取向的耗减活性程度。在呈“正向”取向(5’-GTCT…CAGG-[间隔子]-3’)的直接重复的情况下和在呈“反向”取向(5’-CCTG…AGAC-[间隔子]-3’)的直接重复的情况下的耗减程度分别以实线和虚线描绘。

[0095] 图31A是按pACYC184质粒上的位置示出了CLUST.200916 SRR6837569(没有非编码序列)的耗减靶标和非耗减靶标的密度的图形表示。图31B是按大肠杆菌菌株E.Cloni上的位置示出了CLUST.200916 SRR6837569(没有非编码序列)的耗减靶标和非耗减靶标的密度的图形表示。单独地并与注释基因的取向相关地示出了顶链和底链上的靶标。谱带的幅度指示耗减程度,其中较轻的谱带接近于3的命中阈值。梯度是示出了相对转录物丰度的RNA测序的热图。

[0096] 图32是作为CLUST.200916 SRR6837569(没有非编码序列)的PAM序列的预测的E.Cloni中侧接耗减靶标的序列的WebLogo。

[0097] 图33是示出了双链DNA靶底物的制备和标记的示意图。

[0098] 图34A示出了SEQ ID NO:1的效应子对双链DNA靶底物(靶标A;SEQ ID NO:57)的切割;SEQ ID NO:58的非靶标底物(非靶标B)是阴性对照。图34B示出了SEQ ID NO:28的效应子对双链DNA靶底物(靶标C;SEQ ID NO:59)的切割;SEQ ID NO:60的非靶标底物(非靶标D)是阴性对照。图34C示出了SEQ ID NO:26的效应子对双链DNA靶底物(靶标E;SEQ ID NO:61)的切割;SEQ ID NO:62的非靶标底物(非靶标F)是阴性对照。图34D示出了SEQ ID NO:27的效应子对双链DNA靶底物(靶标G;SEQ ID NO:63)的切割;SEQ ID NO:64的非靶标底物(非靶标H)是阴性对照。图34E示出了SEQ ID NO:25的效应子对双链DNA靶底物(靶标I;SEQ ID NO:65)的切割;SEQ ID NO:66的非靶标底物(非靶标J)是阴性对照。

[0099] 图35是实施例4中所述的用于测量CLUST.200916效应子活性的荧光耗减测定的示意图。

[0100] 图36A示出了SEQ ID NO:28的效应子针对靶标1(SEQ ID NO:67)、靶标2(SEQ ID NO:68)、靶标3(SEQ ID NO:69)、靶标4(SEQ ID NO:70)和靶标5(SEQ ID NO:71)的GFP耗减比(非靶标/靶标)的绘图。图36B示出了SEQ ID NO:25的效应子针对靶标7(SEQ ID NO:72)、靶标8(SEQ ID NO:73)、靶标9(SEQ ID NO:74)、靶标10(SEQ ID NO:75)和靶标11(SEQ ID NO:76)的GFP耗减比(非靶标/靶标)的绘图。图36A和图36B中的耗减比值是根据在12小时的时段内获取的测量值计算的。

[0101] 图37示出了SEQ ID NO:24、SEQ ID NO:28和SEQ ID NO:25的效应子在HEK293细胞中的AAVS1靶基因座处诱导的插入缺失。

具体实施方式

[0102] 天然多样性的CRISPR-Cas系统包括广泛范围的可用于可编程生物技术的活性机制和功能性元件。在自然界中,这些系统能够有效地防御外来DNA和病毒,同时提供自我与非自我的区别以避免自我靶向。在工程化环境中,这些系统提供了分子技术的多样性工具箱,并且定义了靶向空间的边界。本文所述的方法已被用于发现单亚基2类效应子系统内的另外机制和参数,这些另外机制和参数扩展了RNA可编程核酸操纵的能力。

[0103] 除非另有定义,否则本文所用的所有技术和科学术语均具有与本发明所属领域普通技术人员所通常理解的相同含义。虽然类似于或等同于本文所述的那些的方法和材料可以用于本发明的实践或测试中,但是下面描述了合适的方法和材料。本文所提及的所有出版物、专利申请、专利和其他参考文献均通过援引以其全文并入。在有矛盾的情况下,将以本说明书(包括定义)为准。此外,材料、方法和实例仅为说明性的,而并非旨在为限制性的。根据专利法中的标准惯例,申请人使用过渡性短语“包含”、“基本上由……组成”或“由……组成”保留替代性地要求保护任何所披露的发明的权利。

[0104] 除非上下文另有明确规定,否则如本文所用,单数形式“一种/一个(a/an)”和“该”包括复数指示物。例如,对“一种核酸”的提及意指一种或多种核酸。

[0105] 应指出,术语如“优选地”、“合适地”、“通常”和“典型地”在本文中不用于限制所要求保护的发明的范围或者暗示某些特征对于所要求保护的发明的结构或功能是关键、必要的或甚至重要的。相反,这些术语仅仅旨在强调在本发明的特定实施例中可以利用或可以不利用的替代性或其另外的特征。

[0106] 出于描述和限定本发明的目的,应指出术语“基本上”在本文中用于表示可以归因于任何定量比较、值、测量值或其他表示的固有不确定性程度。术语“基本上”在本文中还用于表示定量表示可以不同于所述参考而不导致讨论中的主题的基本功能的改变的程度。

[0107] 如本文所用,术语“CRISPR-Cas系统”是指参与CRISPR效应子的表达或指导其活性的核酸和/或蛋白质,包括编码CRISPR效应子的序列、RNA指导物以及来自CRISPR基因座的其他序列和转录物。

[0108] 如本文可互换使用的术语“CRISPR相关蛋白”、“CRISPR-Cas效应子”、“CRISPR效应子”、“效应子”、“效应子蛋白”、“CRISPR酶”等是指完成酶活性或结合由RNA指导物指定的核酸上的靶位点的蛋白质。在一些实施例中,CRISPR效应子具有核酸内切酶活性、切口酶活性

和/或核酸外切酶活性。

[0109] 如本文所用,术语“RNA指导物”、“指导RNA”、“gRNA”和“指导序列”是指有利于将本文所述的效应子靶向至靶核酸(诸如DNA和/或RNA)的任何RNA分子。示例性“RNA指导物”包括但不限于crRNA,以及与tracrRNA和/或调节子RNA杂交或融合的crRNA。在一些实施例中,RNA指导物包括融合成单RNA分子或者作为单独的RNA分子的crRNA和tracrRNA两者。在一些实施例中,RNA指导物包括融合成单RNA分子或者作为单独的RNA分子的crRNA和调节子RNA。在一些实施例中,RNA指导物包括融合成单RNA分子或者作为单独的RNA分子的crRNA、tracrRNA和调节子RNA。

[0110] 如本文所用,术语“CRISPR RNA”或“crRNA”是指包含由CRISPR效应子用于特异性识别核酸序列的指导序列的RNA分子。典型地,crRNA含有介导靶标识别的序列和与tracrRNA形成双链体的序列。crRNA可以包含与tracrRNA杂交的序列。进而,crRNA:tracrRNA双链体可以结合CRISPR效应子。如本文所用,术语“前crRNA”是指包含DR-间隔子-DR序列的未加工RNA分子。如本文所用,术语“成熟crRNA”是指前crRNA的加工形式;成熟crRNA可以包含DR-间隔子序列,其中DR是前crRNA的DR的截短形式和/或间隔子是前crRNA的间隔子的截短形式。

[0111] 如本文使用,术语“CRISPR效应子复合物”、“效应子复合物”或“监视复合物”是指含有CRISPR效应子和RNA指导物的复合物。CRISPR效应子复合物可以进一步包含一种或多种辅助蛋白。一种或多种辅助蛋白可以是非催化性和/或非靶标结合的。

[0112] 如本文所用,术语“反式激活crRNA”或“tracrRNA”是指这样的RNA分子,该RNA分子包含形成为CRISPR效应子结合特定靶核酸所需的结构和/或序列基序的序列。

[0113] 如本文所用,术语“CRISPR阵列”是指包含CRISPR重复和间隔子的核酸(例如,DNA)区段,该核酸区段从第一个CRISPR重复的第一个核苷酸开始并以最后一个(末端)CRISPR重复的最后一个核苷酸结束。典型地,CRISPR阵列中的每个间隔子位于两个重复之间。如本文所用,术语“CRISPR重复”、“CRISPR直接重复”和“直接重复”是指多个短的直接重复序列,这些直接重复序列在CRISPR阵列内显示出很小的序列变化或不显示序列变化。

[0114] 如本文所述的术语“调节子RNA”是指调节(例如,增加或降低)CRISPR效应子或包含CRISPR效应子的核蛋白复合物的活性的任何RNA分子。在一些实施例中,调节子RNA调节CRISPR效应子或包含CRISPR效应子的核蛋白复合物的核酸酶活性。

[0115] 如本文所用,术语“靶核酸”是指这样的核酸,该核酸包含与RNA指导物中间隔子的全部或部分互补的核苷酸序列。在一些实施例中,靶核酸包含基因。在一些实施例中,靶核酸包含非编码区(例如,启动子)。在一些实施例中,靶核酸是单链的。在一些实施例中,靶核酸是双链的。如本文所用,“转录活性位点”是指在被主动转录的核酸序列中的位点。

[0116] 如本文所用,术语“激活的CRISPR效应子复合物”、“激活的CRISPR复合物”和“激活的复合物”是指能够修饰靶核酸的CRISPR效应子复合物。在一些实施例中,激活的CRISPR复合物能够在激活的CRISPR复合物与靶核酸结合后修饰靶核酸。在一些实施例中,激活的CRISPR复合物与靶核酸的结合导致另外的切割事件,诸如伴RNA切割。

[0117] 如本文所用,术语“切割事件”是指核酸(诸如DNA和/或RNA)中的断裂。在一些实施例中,切割事件是指由本文所述CRISPR系统的核酸酶产生的靶核酸中的断裂。在一些实施例中,切割事件是双链DNA断裂。在一些实施例中,切割事件是单链DNA断裂。在一些实施例

中,切割事件是指伴核酸中的断裂。

[0118] 如本文所用,术语“伴核酸”是指被激活的CRISPR复合物非特异性切割的核酸底物。如本文中关于CRISPR效应子使用的术语“伴DNA酶活性”是指激活的CRISPR复合物的非特异性DNA酶活性。如本文中关于CRISPR效应子使用的术语“伴RNA酶活性”是指激活的CRISPR复合物的非特异性RNA酶活性。

[0119] 如本文所用,术语“供体模板核酸”是指这样的核酸分子,该核酸分子可以用于在本文所述的CRISPR效应子修饰了靶核酸之后对靶序列或靶近侧序列作出模板化改变。在一些实施例中,供体模板核酸是双链核酸。在一些实施例中,供体模板核酸是单链核酸。在一些实施例中,供体模板核酸是线性的。在一些实施例中,供体模板核酸是环状的(例如,质粒)。在一些实施例中,供体模板核酸是外源核酸分子。在一些实施例中,供体模板核酸是内源核酸分子(例如,染色体)。

[0120] 如本文所用,术语“多核苷酸”、“核苷酸”、“寡核苷酸”和“核酸”可以互换使用以指包括DNA、RNA、其衍生物或其组合的核酸。可以使用本领域技术人员熟知的方法来构建根据本发明的基因表达构建体和重组细胞。这些方法包括体外重组DNA技术、合成技术、体内重组技术和聚合酶链反应(PCR)技术。参见例如如以下文献中所述的技术:Maniatis等人,1989,MOLECULAR CLONING:A LABORATORY MANUAL[分子克隆:实验室手册],Cold Spring Harbor Laboratory[冷泉港实验室],New York[纽约];Ausubel等人,1989,CURRENT PROTOCOLS IN MOLECULAR BIOLOGY[分子生物学实验指南],Greene Publishing Associates and Wiley Interscience[格林出版协会和威利跨学科出版社],New York[纽约];以及PCR Protocols:A Guide to Methods and Applications[PCR方案:方法和应用指南](Innis等人.,1990,Academic Press[学术出版社],San Diego,Calif.[加利福尼亚州圣迭哥])。

[0121] 术语“基因修饰”或“基因工程化”广义地是指对细胞的基因组或核酸的操纵。同样,术语“基因工程化的”和“工程化的”是指包含经操纵的基因组或核酸的细胞。基因修饰的方法包括例如异源基因表达、基因或启动子插入或缺失、核酸突变、改变基因表达或失活、酶工程化、定向演化、基于知识的设计、随机诱变方法、基因改组和密码子优化。

[0122] 术语“重组体”表明核酸、蛋白质或细胞是基因修饰、工程化或重组的产物。一般来讲,术语“重组体”是指含有源自多种来源的遗传物质或由其编码的核酸、蛋白质或细胞。如本文所用,术语“重组体”还可以用于描述包含突变的核酸或蛋白质(包括内源性核酸或蛋白质的突变形式)的细胞。术语“重组细胞”和“重组宿主”可以互换使用。在一些实施例中,重组细胞包含本文所披露的CRISPR效应子。CRISPR效应子可以经密码子优化以在重组细胞中表达。在一些实施例中,本文所披露的重组细胞进一步包含RNA指导物。在一些实施例中,本文所披露的重组细胞的RNA指导物包含tracrRNA。在一些实施例中,本文所披露的重组细胞包含调节子RNA。在一些实施例中,重组细胞可以是原核细胞,诸如大肠杆菌细胞。在一些实施例中,重组细胞是真核细胞,诸如哺乳动物细胞,包括人细胞。

[0123] CLUST.200916的鉴定

[0124] 本申请涉及在本文中称为“CLUST.200916”的新型蛋白质家族的鉴定、工程化和用途。如图2A中所示,CLUST.200916的蛋白质包含RuvC结构域(表示为RuvC I、RuvC II和RuvC III)和锌指结构域。如表6中所示,CLUST.200916的效应子的大小范围为从约650个氨基酸

至约850个氨基酸。因此,CLUST.200916的效应子小于如下面所示的本领域已知的效应子。参见例如表1。

[0125] 表1.已知CRISPR-Cas系统效应子的大小。

[0126]

效应子	大小(aa)
StCas9	1128
SpCas9	1368
SaCas9	1053
FnCpf1	1300
AsCpf1	1307
LbCpf1	1246
C2c1	1127(平均值)
CasX	982(平均值)
CasY	1189(平均值)
C2c2	1232(平均值)

[0127] 使用计算方法和算法鉴定CLUST.200916的效应子以搜索和鉴定展现出与某些其他特征的强共现模式的蛋白质。在某些实施例中,这些计算方法涉及鉴定与CRISPR阵列很靠近地共现的蛋白质。本文所披露的方法还可用于鉴定天然很靠近其他特征(非编码特征和蛋白质编码特征(例如,细菌基因座的非编码区域中的噬菌体序列片段;或CRISPR Cas1蛋白质))出现的蛋白质。应理解,本文所述的方法和计算可以在一个或多个计算设备上执行。

[0128] 从基因组或宏基因组数据库获得基因组序列的集合。数据库包括短读段、或重叠群水平数据、或组装的支架、或生物体的完整基因组序列。同样,数据库可以包括来自原核生物或真核生物的基因组序列数据,或可以包括来自宏基因组环境样品的数据。数据库储存库的实例包括国家生物技术信息中心(National Center for Biotechnology Information)(NCBI)RefSeq、NCBI GenBank、NCBI全基因组鸟枪(WGS)和联合基因组研究所(Joint Genome Institute)(JGI)综合微生物基因组(IMG)。

[0129] 在一些实施例中,施加最小大小要求以选择具有指定的最小长度的基因组序列数据。在某些示例性实施例中,最小重叠群长度可以是100个核苷酸、500nt、1kb、1.5kb、2kb、3kb、4kb、5kb、10kb、20kb、40kb或50kb。

[0130] 在一些实施例中,从完整的或选择的基因组序列数据集合中提取已知的或预测的蛋白质。在一些实施例中,已知或预测的蛋白质获取自提取源数据库提供的编码序列(CDS)注释。在一些实施例中,通过应用计算方法从核苷酸序列中鉴定蛋白质来确定预测的蛋白质。在一些实施例中,使用GeneMark套件来从基因组序列预测蛋白质。在一些实施例中,使用Prodigal来从基因组序列预测蛋白质。在一些实施例中,可以在相同的序列数据集合上使用多个蛋白质预测算法,其中所得的蛋白质集被去重复。

[0131] 在一些实施例中,从基因组序列数据中鉴定CRISPR阵列。在一些实施例中,使用PILER-CR来鉴定CRISPR阵列。在一些实施例中,使用CRISPR识别工具(CRT)来鉴定CRISPR阵列。在一些实施例中,通过鉴定重复最小次数(例如,2、3或4次)的核苷酸基序的启发式方法来鉴定CRISPR阵列,其中重复基序的连续出现之间的间隔不超过指定长度(例如,50、100或

150个核苷酸)。在一些实施例中,可以在相同的序列数据集上使用多个CRISPR阵列鉴定工具,其中所得的CRISPR阵列集被去重复。

[0132] 在一些实施例中,鉴定出与CRISPR阵列很靠近的蛋白质(在本文中称为"CRISPR近侧蛋白质聚簇")。在一些实施例中,靠近性被定义为核苷酸距离,并且可以在20kb、15kb或5kb内。在一些实施例中,靠近性被定义为蛋白质与CRISPR阵列之间的开放阅读框(ORF)的数目,并且某些示例性距离可以是10、5、4、3、2、1或0个ORF。然后将被鉴定为与CRISPR阵列很靠近的蛋白质分组到同源蛋白质聚簇中。在一些实施例中,使用blastclust来形成CRISPR近侧蛋白质聚簇。在某些其他实施例中,使用mmseqs2来形成CRISPR近侧蛋白质聚簇。

[0133] 为了建立CRISPR近侧蛋白质聚簇的成员之间的强共现模式,可以在先前编译的已知和预测的蛋白质的完整集合上执行对蛋白质聚簇的每个成员的BLAST搜索。在一些实施例中,可以使用UBLAST或mmseqs2来搜索类似蛋白质。在一些实施例中,可以仅对家族中的代表性蛋白质子集进行搜索。

[0134] 在一些实施例中,通过度量对CRISPR近侧蛋白质聚簇进行排序或过滤,以确定共现。一个示例性度量是蛋白质聚簇中的元件数目与达到某个E值阈值的BLAST匹配数目的比率。在一些实施例中,可以使用恒定的E值阈值。在其他实施例中,E值阈值可以由蛋白质聚簇的距离最远成员确定。在一些实施例中,将蛋白质的总体集合聚簇,并且共现度量是CRISPR近侧蛋白质聚簇的元件数目与一个或多个包含的总体聚簇的元件数目的比率。

[0135] 在一些实施例中,基于聚簇中蛋白质的天然存在的基因座结构,使用人工审查过程来评价工程化系统的潜在功能性和最小的组分集合。在一些实施例中,蛋白质聚簇的图形表示可以有助于人工审查,并且可以包含信息,该信息包括成对序列相似性、系统发育树、源生物体/环境、预测的功能性结构域和基因座结构的图形描绘。在一些实施例中,基因座结构的图形描绘可以针对具有高表示的附近蛋白质家族过滤。在一些实施例中,可以通过相关的附近蛋白质的数目与一个或多个包含的总体聚簇的一个或多个大小的比率来计算表示。在某些示例性实施例中,蛋白质聚簇的图形表示可以包含对天然存在的基因座的CRISPR阵列结构的描绘。在一些实施例中,蛋白质聚簇的图形表示可以包含保守的直接重复的数目相对于推定CRISPR阵列的长度的描绘,或者独特的间隔子序列的数目相对于推定CRISPR阵列的长度的描绘。在一些实施例中,蛋白质聚簇的图形表示可以包含对预测新的CRISPR-Cas系统并鉴定其组分的推定效应子与CRISPR阵列共现的各种度量的描绘。

[0136] CLUST.200916的池化筛选

[0137] 为了有效地验证本文鉴定的工程化CLUST.200916 CRISPR-Cas系统的活性、机制和功能参数,在大肠杆菌中使用了池化筛选方法,如实例2中所述。首先,根据CLUST.200916 CRISPR-Cas系统的保守蛋白和非编码元件的计算鉴定,使用DNA合成和分子克隆将单独的组分组装成单一人工表达载体,在一个实施例中,该单一人工表达载体是基于pET-28a+骨架。在第二实施例中,将效应子和非编码元件在mRNA转录物上转录,并且使用不同的核糖体结合位点来翻译单独效应子。

[0138] 其次,用含有靶向第二质粒pACYC184的非天然间隔子的未加工crRNA的文库替换天然crRNA和靶向间隔子。将此crRNA文库克隆至包含效应子和非编码元件的载体骨架(例如,pET-28a+)中,并且随后将文库连同pACYC184质粒靶标转化至大肠杆菌中。因此,每个所

得的大肠杆菌细胞含有不超过一个靶向阵列。在替代性实施例中，含有非天然间隔子的未加工crRNA的文库另外靶向大肠杆菌必需基因，这些大肠杆菌必需基因从诸如在以下文献中所述的那些的资源中提取：Baba等人(2006)Mol.Syst.Biol.[分子系统生物学]2:2006.0008；和Gerdes等人(2003)J.Bacteriol.[细菌学杂志]185(19):5673-84，将这些文献中的每一个的全部内容通过援引并入本文。在此实施例中，新型CRISPR-Cas系统的破坏必需基因功能的阳性靶向活性导致细胞死亡或生长停滞。在一些实施例中，可以将必需基因靶向间隔子与pACYC184靶标组合。

[0139] 第三，将大肠杆菌在抗生素选择下生长。在一个实施例中，使用三重抗生素选择：卡那霉素(用于确保成功转化含有工程化CRISPR效应子系统的pET-28a+载体)以及氯霉素和四环素(用于确保成功共转化pACYC184靶标载体)。由于pACYC184通常赋予对氯霉素和四环素的抗性，因此在抗生素选择下，靶向质粒的新型CRISPR-Cas系统的阳性活性将消除主动表达crRNA文库的效应子、非编码元件和特异性活性元件的细胞。典型地，在转化后12-14h分析存活细胞的群体。在一些实施例中，在转化后6-8h、在转化后8-12h、在转化后长达24h或在转化后超过24h进行对存活细胞的分析。与较早时间点相比，在较晚时间点检查存活细胞的群体产生与非活性crRNA相比耗减的信号。

[0140] 在一些实施例中，使用双重抗生素选择。撤除氯霉素或四环素以去除选择压力可以提供关于靶向底物、序列特异性和效力的新型信息。例如，对选择或未选择的基因中dsDNA的切割可以导致大肠杆菌中的阴性选择，其中观察到选择的基因和未选择的基因两者的耗减。如果CRISPR-Cas系统干扰转录或翻译(例如，通过结合或通过转录物切割)，则将仅观察到对于选择的抗性基因中而非选择的抗性基因中的靶标的选择。

[0141] 在一些实施例中，仅使用卡那霉素来确保成功转化包含工程化CRISPR-Cas系统的pET-28a+载体。此实施例适用于含有靶向大肠杆菌必需基因的间隔子的文库，因为除了卡那霉素之外，不需要另外的选择来观察生长变化。在此实施例中，去除氯霉素和四环素依赖性，并且它们在文库中的靶标(如果有的话)提供了关于靶向底物、序列特异性和效力的阳性或阴性信息的另外来源。

[0142] 由于pACYC184质粒包含一组可能影响CRISPR-Cas系统活性的不同特征和序列，因此将来自池化筛选的活性crRNA映射至pACYC184上提供了活性模式，该活性模式可以提示不同的活性机制和功能参数。这样，为在异源原核物种中重构新型CRISPR-Cas系统所需的特征可以得到更全面的测试和研究。

[0143] 本文所述的体内池化筛选的关键优点包括：

[0144] (1)通用性-质粒设计允许表达多种效应子和/或非编码元件；文库克隆策略使得计算预测的crRNA的两个转录方向都能表达；

[0145] (2)活性机制和功能参数的全面测试-评价多样性干扰机制，包括核酸切割；检查特征诸如转录、质粒DNA复制的共现；以及可以使用crRNA文库的侧接序列来可靠地确定复杂度等价为4N的PAM；

[0146] (3)敏感性-pACYC184是低拷贝质粒，能够对CRISPR-Cas活性具有高敏感性，这是由于即使适度的干扰率也可以消除质粒编码的抗生素抗性；以及

[0147] (4)效率-优化分子生物学步骤，使得实现速度更快且通量更大的RNA测序，并且蛋白质表达样品可以直接从筛选中的存活细胞中获得。

[0148] 将本文所述的新型CLUST.200916 CRISPR-Cas家族使用体内池化筛选进行评价，以评价其操作元件、机制和参数，以及其在其内源细胞环境之外的工程化系统中具有活性和重新编程的能力。

[0149] CRISPR效应子活性和修饰

[0150] 在一些实施例中，CLUST.200916的CRISPR效应子和RNA指导物形成可以包含其他组分的“二元”复合物。在结合与RNA指导物中的间隔子序列互补的核酸底物(即，序列特异性底物或靶核酸)后，二元复合物被激活。在一些实施例中，序列特异性底物是双链DNA。在一些实施例中，序列特异性底物是单链DNA。在一些实施例中，序列特异性底物是单链RNA。在一些实施例中，序列特异性底物是双链RNA。在一些实施例中，序列特异性要求RNA指导物(例如，crRNA)中的间隔子序列与靶底物完全匹配。在其他实施例中，序列特异性要求RNA指导物(例如，crRNA)中的间隔子序列与靶底物的部分(连续或非连续)匹配。

[0151] 在一些实施例中，二元复合物在结合靶底物后变为激活的。在一些实施例中，激活的复合物展现出“多次周转”活性，由此在作用于(例如，切割)靶底物后，激活的复合物保持在激活状态。在一些实施例中，激活的二元复合物展现出“单周转”活性，由此在作用于靶底物后，二元复合物恢复到非活性状态。在一些实施例中，激活的二元复合物展现出非特异性(即，“伴”)切割活性，由此该复合物切割非靶标核酸。在一些实施例中，非靶标核酸是DNA分子(例如，单链或双链DNA)。在一些实施例中，非靶标核酸是RNA分子(例如，单链或双链RNA)。

[0152] 在一些实施例中，可以将本文所述的CRISPR效应子融合至一个或多个肽标签，包括His-标签、GST-标签、FLAG-标签或myc-标签。在一些实施例中，可以将本文所述的CRISPR效应子融合至可检测部分，诸如荧光蛋白(例如，绿色荧光蛋白或黄色荧光蛋白)。在一些实施例中，将本披露的CRISPR效应子和/或辅助蛋白融合至肽或非肽部分，该肽或非肽部分允许蛋白质进入或定位至组织、细胞或细胞区域。例如，本披露的CRISPR效应子可以包含核定位序列(NLS)，诸如SV40(猿猴病毒40)NLS、c-Myc NLS或其他合适的单颗粒NLS。可以将NLS融合至CRISPR效应子的N末端和/或C末端，并且可以将其单独融合(即，单一NLS)或串联(例如，2、3、4等个NLS的链)。

[0153] 在一些实施例中，将至少一个核输出信号(NES)附接至编码CRISPR效应子的核酸序列。在一些实施例中，附接C末端和/或N末端NLS或NES以用于在真核细胞(例如，人细胞)中的最佳表达和核靶向。

[0154] 在将标签融合至CRISPR效应子的那些实施例中，这种标签可以有利于CRISPR效应子的基于亲和力或基于电荷的纯化，该纯化例如通过利用固定的亲和力或离子交换试剂的液相色谱或珠分离进行。作为非限制性实例，本披露的重组CRISPR效应子包含多组氨酸(His)标签，并且为了纯化，将其装载至包含固定化金属离子的色谱柱上(例如，由固定在树脂上的螯合配体螯合的Zn^{2+}、Ni^{2+}、Cu^{2+}离子，该树脂可以是单独制备的树脂或可商购获得的树脂或即用型柱，诸如由马萨诸塞州马尔堡(Marlborough)的通用医疗生命科学公司(GE Healthcare Life Sciences)市售的HisTrap FF柱)。在装载步骤后，任选地例如使用一种或多种合适的缓冲溶液冲洗柱，并且然后使用合适的洗脱缓冲液洗脱加His标签的蛋白质。替代性地或另外地，如果本披露的重组CRISPR效应子利用FLAG-标签，则可以使用业内已知的免疫沉淀方法纯化这种蛋白质。对于本领域技术人员来说，本披露的加标签的CRISPR效

应子或辅助蛋白的其他合适的纯化方法将是明显的。

[0155] 本文所述的蛋白质(例如,CRISPR效应子或辅助蛋白)可以作为核酸分子或者多肽递送或使用。当使用核酸分子时,编码CRISPR效应子的核酸分子可以是密码子优化的。核酸可以经密码子优化以用于在任何目的生物体(特别是人细胞或细菌)中使用。例如,核酸可以经密码子优化以用于任何非人真核生物(包括小鼠、大鼠、兔、狗、家畜或非人灵长类动物)。密码子使用表是易于获得的,例如在www.kazusa.orjp/codon/上可获得的“密码子使用数据库(Codon Usage Database)”中,并且这些表可以按多种方式进行改编。参见Nakamura等人Nucl.Acids Res.[核酸研究]28:292(2000),将该文献通过援引以其全文并入本文。用于密码子优化特定序列以在特定宿主细胞中表达的计算机算法也是可得的,诸如基因制造(Gene Forge)(Aptagen公司;宾夕法尼亚州雅各布斯(Jacobus,PA))。

[0156] 在一些情况下,本披露的编码用于在真核生物(例如,人或其他哺乳动物细胞)细胞中表达的CRISPR效应子的核酸包含一个或多个内含子,即一个或多个非编码序列,该一个或多个非编码序列在第一端(例如,5’端)包含剪接供体序列并且在第二端(例如,3’端)包含剪接受体序列。任何合适的剪接供体/剪接受体可以用于本披露的各种实施例中,包括但不限于猿猴病毒40(SV40)内含子、β-珠蛋白内含子和合成内含子。替代性地或另外地,本披露的编码CRISPR效应子或辅助蛋白的核酸可以在DNA编码序列的3’端包含转录终止信号,诸如聚腺苷酸化(聚A)信号。在一些情况下,聚A信号很靠近或相邻于内含子(诸如SV40内含子)定位。

[0157] 去激活/失活的CRISPR效应子

[0158] 可以将本文所述的CRISPR效应子修饰为具有减弱的核酸酶活性,例如与野生型CRISPR效应子相比,至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少97%或100%的核酸酶失活。核酸酶活性可以通过本领域已知的若干种方法,例如将突变引入至蛋白质的核酸酶结构域中来减弱。在一些实施例中,鉴定出核酸酶活性的催化残基,并且这些氨基酸残基可以被不同的氨基酸残基(例如,甘氨酸或丙氨酸)取代以减弱核酸酶活性。

[0159] 失活的CRISPR效应子可以包含一个或多个功能性结构域或与一个或多个功能性结构域相关联(例如,经由融合蛋白、接头肽、“GS”接头等)。这些功能性结构域可以具有各种活性,例如甲基化酶活性、去甲基化酶活性、转录激活活性、转录阻遏活性、转录释放因子活性、组蛋白修饰活性、RNA切割活性、DNA切割活性、核酸结合活性和开关活性(例如,光诱导的)。在一些实施例中,功能性结构域是Krüppel相关框(KRAB)、VP64、VP16、Fok1、P65、HSF1、MyoD1和生物素-APEX。

[0160] 一个或多个功能性结构域在失活的CRISPR效应子上的定位允许功能性结构域的正确空间取向,从而以所归属的功能效应影响靶标。例如,如果功能性结构域是转录激活因子(例如,VP16、VP64或p65),则转录激活因子被放置成允许其影响靶标的转录的空间取向。同样,转录阻遏子被定位成影响靶标的转录,并且核酸酶(例如,Fok1)被定位成切割或部分切割靶标。在一些实施例中,功能性结构域位于CRISPR效应子的N末端。在一些实施例中,功能性结构域位于CRISPR效应子的C末端。在一些实施例中,将失活的CRISPR效应子修饰为包含在N末端处的第一功能性结构域和在C末端处的第二功能性结构域。

[0161] 拆分型酶

[0162] 本披露还提供了一种本文所述的CRISPR效应子的拆分形式。CRISPR效应子的拆分形式可能对于递送是有利的。在一些实施例中,CRISPR效应子被拆分为酶的两个部分,该两个部分一起基本上构成功能性CRISPR效应子。

[0163] 拆分可以以一个或多个催化结构域不受影响的方式进行。CRISPR效应子可以起核酸酶的作用,或者可以是失活的酶,这些失活的酶基本上是具有很少或没有催化活性(例如,由于其催化结构域中的一个或多个突变)的RNA结合蛋白。

[0164] 在一些实施例中,核酸酶裂片和α-螺旋裂片表达为单独的多肽。尽管这些裂片本身并不相互作用,但RNA指导物将它们招募至三元复合物中,该三元复合物重现全长CRISPR效应子的活性,并且催化位点特异性DNA切割。使用经修饰的RNA指导物通过阻止二聚化来消除拆分型酶的活性,从而允许诱导型二聚化系统的开发。拆分型酶描述于例如Wright等人."Rational design of a split-Cas9 enzyme complex[拆分型Cas9酶复合物的合理设计],"Proc.Natl.Acad.Sci.[美国国家科学院院刊],112.10(2015):2984-2989中,将该文献通过援引以其全文并入本文。

[0165] 在一些实施例中,可以例如通过采用雷帕霉素敏感性二聚化结构域将拆分型酶融合至二聚化配偶体。这允许产生化学诱导型CRISPR效应子用于CRISPR效应子活性的时间控制。因此,CRISPR效应子可以通过拆分成两个片段而成为化学诱导型,并且雷帕霉素敏感性二聚化结构域可以用于CRISPR效应子的受控重组。

[0166] 拆分点典型地是经由计算机模拟设计的并且克隆至构建体中。在此过程中,可以将突变引入至拆分型酶,并且可以去除非功能性结构域。在一些实施例中,拆分型CRISPR效应子的两个部分或片段(即,N末端和C末端片段)可以形成包含野生型CRISPR效应子序列的例如至少70%、至少80%、至少90%、至少95%或至少99%的完全CRISPR效应子。

[0167] 自激活或自失活酶

[0168] 本文所述的CRISPR效应子可以被设计为自激活或自失活的。在一些实施例中,CRISPR效应子是自失活的。例如,可以将靶序列引入至CRISPR效应子编码构建体中。因此,CRISPR效应子可以切割靶序列以及编码该酶的构建体,从而自失活其表达。构建自失活CRISPR系统的方法描述于例如Epstein等人,"Engineering a Self-Inactivating CRISPR System for AAV Vectors[工程化用于AAV载体的自失活CRISPR系统],"Mol.Ther.[分子疗法],24(2016):S50中,将该文献通过援引以其全文并入本文。

[0169] 在一些其他实施例中,在弱启动子(例如,7SK启动子)控制下表达的另外RNA指导物可以靶向编码CRISPR效应子的核酸序列以阻止和/或阻断其表达(例如,通过阻止核酸的转录和/或翻译)。用表达CRISPR效应子、RNA指导物和靶向编码CRISPR效应子的核酸的RNA指导物的载体转染细胞可以导致编码CRISPR效应子的核酸的有效破坏并降低CRISPR效应子的水平,从而限制基因组编辑活性。

[0170] 在一些实施例中,CRISPR效应子的基因组编辑活性可以通过哺乳动物细胞中的内源RNA特征(例如,miRNA)来调节。CRISPR效应子开关可以通过在编码CRISPR效应子的mRNA的5-UTR中使用miRNA互补序列来实现。开关选择性地且有效地响应靶细胞中的miRNA。因此,开关可以通过感测异质细胞群体内的内源miRNA活性来差异地控制基因组编辑。因此,开关系统可以为基于细胞内miRNA信息的细胞类型选择性基因组编辑和细胞工程化提供框架(Hirosawa等人"Cell-type-specific genome editing with a microRNA-responsive

CRISPR-Cas9 switch[利用微RNA响应性CRISPR-Cas9开关的细胞类型特异性基因组编辑],"Nucl.Acids Res.[核酸研究],2017年7月27日;45(13):e118)。

[0171] 诱导型CRISPR效应子

[0172] CRISPR效应子可以是诱导型的,例如,光诱导型的或化学诱导型的。此机制允许激活CRISPR效应子中的功能性结构域。光诱导性可以通过本领域已知的各种方法来实现,例如通过设计这样的融合复合物来实现,其中将CRY2PHR/CIBN配对用于拆分型CRISPR效应子中(参见例如,Konermann等人,"Optical control of mammalian endogenous transcription and epigenetic states[哺乳动物内源转录和表观遗传状态的光学控制],"Nature[自然],500.7463(2013):472)。化学诱导性可以例如通过设计这样的融合复合物来实现,其中将FKBP/FRB(FK506结合蛋白/FKBP雷帕霉素结合结构域)配对用于拆分型CRISPR效应子中。雷帕霉素是形成融合复合物从而激活CRISPR效应子所需的(参见例如,Zetsche等人,"A split-Cas9 architecture for inducible genome editing and transcription modulation,"Nature Biotech.[用于诱导型基因组编辑和转录调节的拆分型Cas9架构],33.2(2015):139-142)。

[0173] 此外,CRISPR效应子的表达可以通过诱导型启动子,例如四环素或强力霉素控制的转录激活(Tet-On和Tet-Off表达系统)、激素诱导型基因表达系统(例如,蜕皮激素诱导型基因表达系统)和阿拉伯糖诱导型基因表达系统来调节。当作为RNA递送时,RNA靶向效应子蛋白的表达可以经由核糖开关来调节,该核糖开关可以感测小分子如四环素(参见例如,Goldfless等人,"Direct and specific chemical control of eukaryotic translation with a synthetic RNA-protein interaction[通过合成的RNA-蛋白质相互作用直接和特异性地化学控制真核翻译],"Nucl.Acids Res.[核酸研究],40.9(2012):e64-e64)。

[0174] 诱导型CRISPR效应子和诱导型CRISPR系统的各种实施例描述于例如US 8871445、US 20160208243和WO 2016205764中,将这些文献中的每一个通过援引以其全文并入本文。

[0175] 功能性突变

[0176] 可以将各种突变或修饰引入至如本文所述的CRISPR效应子中,以改善特异性和/或稳健性。在一些实施例中,鉴定出识别原型间隔子相邻基序(PAM)的氨基酸残基。可以进一步将本文所述的CRISPR效应子修饰以识别不同的PAM,例如通过用其他氨基酸残基取代识别PAM的氨基酸残基。在一些实施例中,CRISPR效应子可以识别例如5'-TTN-3',其中"N"是任何核碱基。

[0177] 在一些实施例中,可以使本文所述的CRISPR效应子在一个或多个氨基酸残基处突变以修饰一种或多种功能活性。例如,在一些实施例中,使CRISPR效应子在一个或多个氨基酸残基处突变以修饰其解旋酶活性。在一些实施例中,使CRISPR效应子在一个或多个氨基酸残基处突变以修饰其核酸酶活性(例如,核酸内切酶活性或核酸外切酶活性)。在一些实施例中,使CRISPR效应子在一个或多个氨基酸残基处突变以修饰其与RNA指导物功能上相关联的能力。在一些实施例中,使CRISPR效应子在一个或多个氨基酸残基处突变以改变其与靶核酸功能上相关联的能力。

[0178] 在一些实施例中,本文所述的CRISPR效应子能够切割靶核酸分子。在一些实施例中,CRISPR效应子切割靶核酸分子的两条链。然而,在一些实施例中,使CRISPR效应子在一个或多个氨基酸残基处突变以修饰其切割活性。例如,在一些实施例中,CRISPR效应子可以

包含增加CRISPR效应子切割靶核酸的能力的一个或多个突变。在另一实例中，在一些实施例中，CRISPR效应子可以包含使该酶不能切割靶核酸的一个或多个突变。在其他实施例中，CRISPR效应子可以包含一个或多个突变，由此使得该酶能够切割靶核酸的链(即，切口酶活性)。在一些实施例中，CRISPR效应子能够切割靶核酸的与跟RNA指导物杂交的链互补的链。在一些实施例中，CRISPR效应子能够切割靶核酸的与RNA指导物杂交的链。

[0179] 在一些实施例中，将本文所披露的CRISPR效应子的一个或多个残基突变为精氨酸部分。在一些实施例中，将本文所披露的CRISPR效应子的一个或多个残基突变为甘氨酸部分。在一些实施例中，基于本文所披露的CRISPR效应子的系统发育比对的共有残基对本文所披露的CRISPR效应子的一个或多个残基进行突变。

[0180] 在一些实施例中，本文所述的CRISPR效应子可以经工程化以包含一个或多个氨基酸残基的缺失，以减小酶的大小，同时保留一个或多个所期望的功能活性(例如，核酸酶活性和与RNA指导物功能上相互作用的能力)。截短型CRISPR效应子可以有利地与具有负载限制的递送系统组合使用。

[0181] 在一方面，本披露提供了与本文所述的核酸序列至少10％、15％、20％、25％、30％、35％、40％、45％、50％、55％、60％、65％、70％、75％、80％、85％、90％、91％、92％、93％、94％、95％、96％、97％、98％或99％相同，同时保持图2A中所示的结构域架构的核酸序列。在另一方面，本披露还提供了与本文所述的氨基酸序列至少10％、15％、20％、25％、30％、35％、40％、45％、50％、55％、60％、65％、70％、75％、80％、85％、90％、91％、92％、93％、94％、95％、96％、97％、98％或99％相同，同时保持图2A中所示的结构域架构的氨基酸序列。

[0182] 在一些实施例中，核酸序列具有至少一部分(例如，至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、20、30、40、50、60、70、80、90或100个核苷酸，例如，连续或非连续核苷酸)与本文所述的序列相同。在一些实施例中，核酸序列具有至少一部分(例如，至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、20、30、40、50、60、70、80、90或100个核苷酸，例如，连续或非连续核苷酸)与本文所述的序列不同。

[0183] 在一些实施例中，氨基酸序列具有至少一部分(例如，至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、20、30、40、50、60、70、80、90或100个氨基酸残基，例如，连续或非连续氨基酸残基)与本文所述的序列相同。在一些实施例中，氨基酸序列具有至少一部分(例如，至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、20、30、40、50、60、70、80、90或100个氨基酸残基，例如，连续或非连续氨基酸残基)与本文所述的序列不同。

[0184] 为了确定两个氨基酸序列、或两个核酸序列的同一性百分比，出于最佳比对的目的将序列比对(例如，可以在第一和第二氨基酸或核酸序列中的一者或两者中引入空位以用于最佳比对，并且出于比较目的可以忽略非同源序列)。一般来讲，出于比较目的而比对的参考序列的长度应是参考序列长度的至少80％，并且在一些实施例中是参考序列长度的至少90％、95％或100％。然后比较在相应氨基酸位置或核苷酸位置处的氨基酸残基或核苷酸。当第一序列中的位置被与在第二序列中的相应位置相同的氨基酸残基或核苷酸占据时，则这些分子在那个位置是相同的。两个序列之间的同一性百分比是这些序列共享的相同位置的数目的函数，考虑了空位数目和每个空位的长度，需要引入它们以用于两个序列的最佳比对。出于本披露的目的，序列的比较和两个序列之间的一致性百分比的确定可以

使用具有12的空位罚分、4的空位延伸罚分和5的移码空位罚分的Blossum 62评分矩阵来完成。

[0185] RNA指导物和RNA指导物修饰

[0186] 在一些实施例中，本文所述的RNA指导物包含尿嘧啶(U)。在一些实施例中，本文所述的RNA指导物包含胸腺嘧啶(T)。在一些实施例中，本文所述的RNA指导物的直接重复序列包含尿嘧啶(U)。在一些实施例中，本文所述的RNA指导物的直接重复序列包含胸腺嘧啶(T)。在一些实施例中，根据表3或8的直接重复序列包含这样的序列，该序列在表3或8中相应序列中指示为胸腺嘧啶的一个或多个位置中包含尿嘧啶。

[0187] 在一些实施例中，直接重复仅包含在内源CRISPR阵列中重复的序列的一个拷贝。在一些实施例中，直接重复是相邻于(例如，侧接)内源CRISPR阵列中发现的一个或多个间隔子序列的全长序列。在一些实施例中，直接重复是相邻于(例如，侧接)内源CRISPR阵列中发现的一个或多个间隔子序列的全长序列的一部分(例如，加工部分)。

[0188] 间隔子长度

[0189] RNA指导物的间隔子长度可以在从约14至50个核苷酸的范围内。RNA指导物的间隔子长度可以在从约20至35个核苷酸的范围内。在一些实施例中，RNA指导物的间隔子长度是至少15个核苷酸、至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸或至少22个核苷酸。在一些实施例中，间隔子长度是从15至17个核苷酸、从15至23个核苷酸、从16至22个核苷酸、从17至20个核苷酸、从20至24个核苷酸(例如，20、21、22、23或24个核苷酸)、从23至25个核苷酸(例如，23、24或25个核苷酸)、从24至27个核苷酸、从27至30个核苷酸、从30至45个核苷酸(例如，30、31、32、33、34、35、40或45个核苷酸)、从30或35至40个核苷酸、从41至45个核苷酸、从45至50个核苷酸或更长。对应于本申请的成熟crRNA的大约间隔子长度示出于表2中。在一些实施例中，表2中鉴定出的间隔子长度是本申请的成熟crRNA的优选间隔子长度。在一些实施例中，本申请的RNA指导物(前crRNA或成熟crRNA)的优选间隔子长度是约24个核苷酸。

[0190] 表2.成熟crRNA的示例性间隔子长度。

[0191]

效应子	间隔子长度
3300013232(SEQ ID NO:1)	16或17个核苷酸
3300027824(SEQ ID NO:3)	30至35个核苷酸
SRR6837557(SEQ ID NO:23)	12个核苷酸
SRR6837562(SEQ ID NO:24)	11至14个核苷酸
SRR6837569(SEQ ID NO:25)	11至13个核苷酸
SRR6837570(SEQ ID NO:26)	17个核苷酸
SRR6837575(SEQ ID NO:28)	11个核苷酸
SRR6837577(SEQ ID NO:29)	14个核苷酸

[0192] 在一些实施例中，RNA指导物的直接重复长度是至少16个核苷酸，或是从16至20个核苷酸(例如，16、17、18、19或20个核苷酸)。在一些实施例中，RNA指导物的直接重复长度是约22至36个核苷酸。示例性全长直接重复序列(例如，前crRNA或未加工crRNA的直接重复序列)和成熟crRNA的直接重复序列(例如，加工crRNA的直接重复序列)示出于表3中。还参见表8。

[0193] 表3.前crRNA和成熟crRNA序列的示例性直接重复序列。

[0194]

效应子	前 crRNA 直接重复序列	成熟 crRNA 直接重复序列
3300013232（SEQ ID NO:1）	CAACCTAAACGATGGCTCGATTCGTCGAGAC（SEQ ID NO: 77）	CTAAACGATGGCTCGATTCGTCGAGAC（SEQ ID NO: 78）
3300027824（SEQ ID NO: 3）	GTAGAAGACCTCGCTGATTGCTCGGTGCGCCGAGAC（SEQ ID NO: 79）	TGATTGCTCGGTGCGCCGAGA（SEQ ID NO: 80）
3300027908（SEQ ID NO: 22）	CTTTCAAGACTAATAGATTGCTCCTTACGAGGAGAC（SEQ ID NO: 81）	CTAATAGATTGCTCCTTACGAGGAGAC（SEQ ID NO: 82）

[0195]

SRR6837557（SEQ ID NO: 23）	GTCGAGACCGATGACGAGTGCGCGGTGCGCCGCGAC（SEQ ID NO: 83）	TGACGAGTGCGCGGTGCGCCGCGAC（SEQ ID NO: 84）
SRR6837562（SEQ ID NO: 24）	CCTGCAAGGGATCCAAATTGCTCTGTTCGCAGAGAC（SEQ ID NO: 85）	TCCAAATTGCTCTGTTCGCAGAGAC（SEQ ID NO: 86）
SRR6837569（SEQ ID NO: 25）	CCTGCAAGGGATCCAAATTGCTCTGTTCGCAGAGAC（SEQ ID NO: 87）	ATCCAAATTGCTCTGTTCGCAGAGAC（SEQ ID NO: 88）
SRR6837570（SEQ ID NO: 26）	GCGCCAACGACCTCTGATTGTCCGGTACGCCGGAAC（SEQ ID NO: 89）	TCTGATTGTCCGGTACGCCGGAAC（SEQ ID NO: 90）
SRR6837575（SEQ ID NO: 28）	CCATCAATGGATCCAAATTGCTCTGTACGCAGAGAC（SEQ ID NO: 91）	ATCCAAATTGCTCTGTACGCAGAGAC（SEQ ID NO: 92）
SRR6837577（SEQ ID NO: 29）	GTCGAGAGCGATGACGAGTGCGCGGTGCGCCGCGAC（SEQ ID NO: 93）	TGACGAGTGCGCGGTGCGCCGCGAC（SEQ ID NO: 94）

[0196] 如本文所用，术语“原型间隔子相邻基序”或“PAM”是指相邻于与包含效应子和RNA指导物的复合物结合的靶序列的DNA序列。在一些实施例中，PAM是酶活性所需的。如本文所用，术语“相邻”包括复合物的RNA指导物与紧邻PAM的靶序列特异性结合、相互作用或缔合的情况。在此类情况下，在靶序列与PAM之间没有核苷酸。术语“相邻”还包括在与靶向部分结合的靶序列与PAM之间存在少数(例如，1、2、3、4或5个)核苷酸的情况。在一些实施例中，对应于本申请的效应子的PAM示出于表4中。如本文所用，N可以各自是任何核苷酸(例如，A、G、T或C)或其子集(例如Y(C或T)、K(G或T)、B(G、T或C)、H(A、C或T))。例如，在一些实施例中，5’-TTN-3’的PAM序列是指5’-TTT-3’或5’-TTG-3’的PAM序列。

[0197] 表4.对应于本发明的效应子的PAM序列。

[0198]

效应子	PAM 序列
3300013232（SEQ ID NO:1）	5’-TTN-3’

[0199]

	5’-YYN-3’ 5’-HHN-3’
3300027908（SEQ ID NO: 22）	5’-TTN-3’ 5’-YYN-3’ 5’-HHN-3’
SRR6837562（SEQ ID NO: 24）	5’-TTN-3’ 5’-YKN-3’
SRR6837569（SEQ ID NO: 25）	5’-TTN-3’
SRR6837570（SEQ ID NO: 26）	5’-TTN-3’ 5’-YYN-3’ 5’-HBN-3’
SRR6837575（SEQ ID NO: 28）	5’-TTN-3’ 5’-YKN-3’
SRR6837577（SEQ ID NO: 29）	5’-TTN-3’ 5’-HHN-3’

[0200] 在一些实施例中，RNA指导物进一步包含tracrRNA。在一些实施例中，不需要tracrRNA(例如，tracrRNA是任选的)。在一些实施例中，tracrRNA是表9中所示的非编码序列的一部分。例如，在一些实施例中，任选的tracrRNA是表5的序列。

[0201] 表5.示例性tracrRNA序列。

[0202]

效应子	TracrRNA
3300013232（SEQ ID NO: 1）	GCTGGATTGATTCCTGCGGTGTAGAATAGCATAACCAGGGTCCGCTGGACTATCCAGCAACGACGGCTCGATACGTCGAGCCGATGGAGAAGTCATGCCAAAGATCAAGAA（SEQ ID NO: 95） TCGATGCTGGATTGATTCCTGCGGTGTAGAATAGCATAACCAGGGTCCGCTGGACTATCCAGCAACGACGGCTCGATACGTCGAGCCGATGGAGAAGTCATGCCAAAGATCA（SEQ ID NO: 96） ATTGATTCCTGCGGTGTAGAATAGCATAACCAGGGTCCGCTGGACTATCCAGCAACGACGGCTCGATACGTCGAGCCGATGGAGAAGTCATGCCAAAG（SEQ ID NO: 97） ATAGCATAACCAGGGTCCGCTGGACTATCCAGCAACGACGGCTCGATACGTCGAGCCGATGGAGAAGTCATGCCAAAGAT（SEQ ID NO: 98） TAGCATAACCAGGGTCCGCTGGACTATCCAGCAACGACGGCTCGATACGTCGAGCCGATGGAGAAGTCATGCCAAAGA（SEQ ID NO: 99）

[0203]

	GCCAAAGATCAAGAAACCGACTGAGATTTCCCTGCTACGCAAGGAGGTGTTCCCTGA (SEQ ID NO: 100) ATTCTACACCGCAGGAATCAATCCAGCATCGAGTAGACCGTTGC (SEQ ID NO: 101) ATTCTACACCGCAGGAATCAATCCAGCATCGAGTAGACCGTTGCA (SEQ ID NO: 102)
3300027908 (SEQ ID NO: 22)	CTTGTAAGAAATTCGTTCGAGAAAATGAAATTCCTGTCGCTATAACGGCCCAGTCGAT GCCGAAACCAACTTGCGAGCGCTCGGGCGACGCTAAAAAGCCTGTCCGTGCTCGCAAG GCTAAAGCACCGGAATTTCAC (SEQ ID NO: 103) GGCTTGTAAGAAATTCGTTCGAGAAAATGAAATTCCTGTCGCTATAACGGCCCAGTCG ATGCCGAAACCAACTTGCGAGCGCTCGGGCGACGCTAAAAAGCCTGTCCGTGCTCGCA AGGCTAAAGCACCGGAATTT (SEQ ID NO: 104)
SRR6837557 (SEQ ID NO: 23)	TGTCGAGAGCGATGACGAGTGCGCGGTGCGCCGCGACCAGCTCTGTGTGTAGTTGACC CGTAGTCAGGACAAGGAGAGGTCGAGAGCGATGACGAGTGCGCGGTGCGCCGCGACGT GCTGTTGATCTGATAGACGGGG (SEQ ID NO: 105) TCGAGAGCGATGACGAGTGCGCGGTGCGCCGCGACGTGCTGTTGATCTGATAGACG (SEQ ID NO: 106) GGTTCCGAATTTCTCGGTGGCTGTCGAGAGCGATGACGAGTGCGCGGTGCGCCGCGAC CAGCTCTGTGTGTAGTTGACCCG (SEQ ID NO: 107) GGTTCCGAATTTCTCGGTGGCTGTCGAGAGCGATGACGAGTGCGCGGTGCGCCGCGAC CAGCTCTGTGTGTAGTTGACC (SEQ ID NO: 108)
SRR6837570 (SEQ ID NO: 26)	TGCTACAGGCAATCACAAACAGCGGGAGACGAACAATGACTCTGGCCGAGCTGCGCGA CAAATACTTCTACAAGATCAA (SEQ ID NO: 109) CTACAGGCAATCACAAACAGCGGGAGACGAACAATGACTCTGGCCGAGCTGCGCGACA AATACTTCTACAAGATC (SEQ ID NO: 110) GAGACGAACAATGACTCTGGCCGAGCTGCGCGACAAATACTTCTACAAGATCAAGTTC CGCAAGATCGATCTCAGGCAAGCCGGCAAGATCCT (SEQ ID NO: 111) GGGAGACGAACAATGACTCTGGCCGAGCTGCGCGACAAATACTTCTACAAGATCAAGT TCCGCAAGATCGATCTCAGGCAAGCCGGCAAGATCCTCAAG (SEQ ID NO: 112)

[0204]

	GCGGCCGGTGCTCGCAGACGAGCTGCTAGGTCTTTGAAAATTGAATAGATTGTAATGGTGAGTTGC (SEQ ID NO: 113) CGGCCGGTGCTCGCAGACGAGCTGCTAGGTCTTTGAAAATTGAATAGATTGTAATGGTGAGTTGC (SEQ ID NO: 114)
SRR6837575 (SEQ ID NO: 28)	CGTCTACACCGAACGTAAAATCTTAGAAAACGTCATGAAGGAGGCGCAATACCTATCGTGCG (SEQ ID NO: 115)
SRR6837577 (SEQ ID NO: 29)	GGTTGGCGTCAGTGAAAACTGAATAGACGAAGATACGAGCCGCTCGCAGGGTGGAGCGGAAGATCTCGATCTG (SEQ ID NO: 116) GCTGGTTGGCGTCAGTGAAAACTGAATAGACGAAGATACGAGCCGCTCGCAGGGTGGAGCGGAAGATCTCGATCTGG (SEQ ID NO: 117) CCCGCATGTCGCCAGAACAGACGCACAAGAAGTTCGTGAAGATCGTCGAGTCCGAGGGCACGACCAAGGTCGCCGCCCGTCTCGGCTGCTCCG (SEQ ID NO: 118) CCCCGCATGTCGCCAGAACAGACGCACAAGAAGTTCGTGAAGATCGTCGAGTCCGAGGGCACGACCAAGGTCGCCGCCCGTCTCGGCTGCTCCGTGGCGC (SEQ ID NO: 119) AGGTGCGAGCATGCCCGACAAGCAGACGCCCAAGGACACCAAGGACAAGCCCGAGAGCCCCG (SEQ ID NO: 120) AGGTGCGAGCATGCCCGACAAGCAGACGCCCAAGGACACCAAGGACAAGCCCGAGAGCCCC (SEQ ID NO: 121)

[0205] RNA指导序列可以按这样的方式进行修饰，该方式允许形成CRISPR复合物并成功结合靶标，而同时不允许成功的核酸酶活性(即，没有核酸酶活性/没有引起插入缺失)。这些经修饰的指导序列被称为“死指导物”或“死指导序列”。这些死指导物或死指导序列对于核酸酶活性而言可以是催化上失活的或构象上失活的。死指导序列典型地比导致活性切割的对应指导序列短。在一些实施例中，死指导物比具有核酸酶活性的对应RNA指导物短5%、10%、20%、30%、40%或50%。RNA指导物的死指导序列可以具有从13至15个核苷酸的长度(例如，13、14或15个核苷酸的长度)、从15至19个核苷酸的长度、或从17至18个核苷酸的长度(例如，17个核苷酸的长度)。

[0206] 因此，在一方面，本披露提供了非天然存在的或工程化的CRISPR系统，这些系统包括如本文所述的功能性CLUST.200916 CRISPR效应子、和RNA指导物，其中RNA指导物包含死指导序列，由此RNA指导物能够与靶序列杂交，使得将CRISPR系统指导至细胞中目的基因组基因座而没有可检测的切割活性。对死指导物的详细描述是描述于例如WO 2016094872中，将该文献通过援引以其全文并入本文。

[0207] 诱导型RNA指导物

[0208] RNA指导物可以作为诱导型系统的组分产生。这些系统的诱导型性质允许对基因编辑或基因表达的时空控制。在一些实施例中，用于诱导型系统的刺激包括例如电磁辐射、

声能、化学能和/或热能。

[0209] 在一些实施例中,RNA指导物的转录可以通过诱导型启动子,例如四环素或强力霉素控制的转录激活(Tet-On和Tet-Off表达系统)、激素诱导型基因表达系统(例如,蜕皮激素诱导型基因表达系统)和阿拉伯糖诱导型基因表达系统来调节。诱导型系统的其他实例包括例如小分子双杂交转录激活系统(FKBP,ABA等)、光诱导型系统(光敏色素、LOV结构域或隐花色素)或光诱导型转录效应子(LITE)。这些诱导型系统描述于例如WO 2016205764和US 8795965中,将这些文献中的每一个通过援引以其全文并入本文。

[0210] 化学修饰

[0211] 可以将化学修饰应用于RNA指导物的磷酸骨架、糖和/或碱基。骨架修饰(诸如硫代磷酸酯)修饰磷酸骨架上的电荷并且有助于寡核苷酸的递送和核酸酶抗性(参见例如,Eckstein,“Phosphorothioates,essential components of therapeutic oligonucleotides[硫代磷酸酯,治疗性寡核苷酸的基本组分],”Nucl.Acid Ther.[核酸疗法],24(2014),第374-387页);糖的修饰(诸如2’-O-甲基(2’-OMe)、2’-F和锁核酸(LNA))增强碱基配对和核酸酶抗性两者(参见例如,Allerson等人“Fully 2‘-modified oligonucleotide duplexes with improved in vitro potency and stability compared to unmodified small interfering RNA[与未修饰的小干扰RNA相比,完全2’修饰的寡核苷酸双链体具有改善的体外效力和稳定性],”J.Med.Chem.[医药化学杂志],48.4(2005):901-904)。化学修饰的碱基(诸如2-硫尿苷或N6-甲基腺苷等)可以允许更强或更弱的碱基配对(参见例如,Bramsen等人,“Development of therapeutic-grade small interfering RNAs by chemical engineering[通过化学工程化开发治疗级小分子干扰RNA],”Front.Genet.[遗传学前沿],2012年8月20日;3:154)。此外,RNA适于5’端和3’端两者与多种功能性部分缀合,该多种功能性部分包括荧光染料、聚乙二醇或蛋白质。

[0212] 可以将广泛多种的修饰应用于化学合成的RNA指导物分子。例如,用2’-OMe修饰寡核苷酸以改善核酸酶抗性可以改变沃森-克里克碱基配对的结合能。此外,2’-OMe修饰可以影响寡核苷酸与转染试剂、蛋白质或细胞中任何其他分子相互作用的方式。这些修饰的效果可以通过经验测试来确定。

[0213] 在一些实施例中,RNA指导物包含一个或多个硫代磷酸酯修饰。在一些实施例中,RNA指导物包含一个或多个锁核酸以用于增强碱基配对和/或增加核酸酶抗性。

[0214] 这些化学修饰的概述可以见于例如Kelley等人,“Versatility of chemically synthesized guide RNAs for CRISPR-Cas9 genome editing[用于CRISPR-Cas9基因组编辑的化学合成的指导RNA的通用性],”J.Biotechnol.[生物技术杂志]2016年9月10日;233:74-83;WO 2016205764;以及US 8795965中;将这些文献中的每一个通过援引以其全文并入。

[0215] 序列修饰

[0216] 本文所述的RNA指导物、tracrRNA和crRNA的序列和长度可以是优化的。在一些实施例中,RNA指导物的优化长度可以通过鉴定tracrRNA和/或crRNA的加工形式,或者通过RNA指导物、tracrRNA、crRNA和tracrRNA四元环的经验长度研究来确定。

[0217] RNA指导物还可以包含一个或多个适配体序列。适配体是可以结合特定靶分子的寡核苷酸或肽分子。适配体可以对基因效应子、基因激活子或基因阻遏子具有特异性。在一

些实施例中,适配体可以对蛋白质具有特异性,该蛋白质又特异性于并招募/结合特定基因效应子、基因激活子或基因阻遏子。效应子、激活子或阻遏子可以以融合蛋白的形式存在。在一些实施例中,RNA指导物具有两个或更多个对相同衔接蛋白具有特异性的适配体序列。在一些实施例中,两个或更多个适配体序列对不同的衔接蛋白具有特异性。衔接蛋白可以包括例如MS2、PP7、Qβ、F2、GA、fr、JP501、M12、R17、BZ13、JP34、JP500、KU1、M11、MX1、TW18、VK、SP、FI、ID2、NL95、TW19、AP205、ΦCb5、ΦCb8r、ΦCb12r、ΦCb23r、7s和PRR1。因此,在一些实施例中,适配体选自特异性结合如本文所述的任一种衔接蛋白的结合蛋白。在一些实施例中,适配体序列是MS2环。适配体的详细描述可以见于例如Nowak等人,"Guide RNA engineering for versatile Cas9 functionality[针对通用性Cas9功能的指导RNA工程化],"Nucl.Acid.Res.[核酸研究],2016年11月16日;44(20):9555-9564;和WO 2016205764中,将这些文献中的每一个通过援引以其全文并入本文。

[0218] 指导物:靶序列匹配要求

[0219] 在CRISPR系统中,指导序列与其相应的靶序列之间的互补程度可以是约50%、60%、75%、80%、85%、90%、95%、97.5%、99%或100%。为了减少脱靶相互作用,例如为了减少指导物与具有低互补性的靶序列相互作用,可以将突变引入至CRISPR系统,使得CRISPR系统可以区分具有大于80%、85%、90%或95%互补性的靶序列和脱靶序列。在一些实施例中,互补程度是从80%至95%、例如约83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%或95%(例如,区分具有18个核苷酸的靶标与具有1、2或3个错配的18个核苷酸的脱靶)。因此,在一些实施例中,指导序列与其相应的靶序列之间的互补程度大于94.5%、95%、95.5%、96%、96.5%、97%、97.5%、98%、98.5%、99%、99.5%或99.9%。在一些实施例中,互补程度是100%。

[0220] 在本领域已知,不需要完全的互补性,前提是有足够的互补性发挥作用。可以通过引入错配(例如,间隔子序列与靶序列之间的一个或多个错配,诸如1或2个错配(包括沿着间隔子/靶标的错配的位置))来利用对切割效率的调节。典型地,错配(例如,双错配)位于越中心(即,不在3'或5'端处);切割效率受到的影响越大。因此,通过选择沿着间隔子序列的错配位置,可以调节切割效率。例如,如果期望靶标的小于100%的切割率(例如,在细胞群体中),则间隔子序列中可以引入间隔子序列与靶序列之间的1或2个错配。

[0221] CRISPR系统的使用方法

[0222] 本文所述的CRISPR系统具有广泛多种的效用,包括修饰(例如,缺失、插入、易位、失活或激活)多种细胞类型中的靶多核苷酸。CRISPR系统在以下中具有广谱的应用:例如DNA/RNA检测(例如,特异性高敏感性酶促报告子解锁(SHERLOCK))、跟踪和标记核酸、富集测定(从背景中提取所期望序列)、检测循环肿瘤DNA、制备下一代文库、药物筛选、疾病诊断和预后以及治疗各种遗传障碍。

[0223] DNA/RNA检测

[0224] 在一方面,本文所述的CRISPR系统可以用于DNA/RNA检测中。单效应子RNA指导的DNA酶可以用CRISPR RNA(crRNA)重新编程,以为特异性单链DNA(ssDNA)感测提供平台。激活的V型单效应子DNA指导的DNA酶在识别其DNA靶标后参与附近的非靶向ssDNA的"伴"切割。此crRNA编程的伴切割活性允许CRISPR系统通过经标记的ssDNA的非特异性降解来检测特定DNA的存在。

[0225] 伴ssDNA活性可以在DNA检测应用中与报告子组合，这些检测应用诸如称为DNA核酸内切酶靶向CRISPR反式报告子(DETECTR)方法的方法，该方法实现了对DNA检测的渺摩尔敏感性(参见例如，Chen等人，Science[科学]，360(6387)：436-439，2018)，将该文献通过援引以其全文并入本文。使用本文所述的酶的一个应用是在体外环境中降解非特异性ssDNA。也可以将连接荧光团和淬灭剂的“报告子”ssDNA分子与未知DNA样品(单链或者双链)一起添加至体外系统中。在识别未知DNA片段中的靶序列后，效应子复合物切割报告子ssDNA，从而产生荧光读出。

[0226] 在其他实施例中，SHERLOCK方法(特异性高敏感性酶促报告子解锁)还提供基于报告子ssDNA的核酸扩增和伴切割的具有渺摩尔(或单分子)敏感性的体外核酸检测平台，从而允许实时检测靶标。在SHERLOCK中使用CRISPR的方法详细描述于例如Gootenberg等人“Nucleic acid detection with CRISPR-Cas13a/C2c2[使用CRISPR-Cas13a/C2c2进行核酸检测]，”Science[科学]，356(6336)：438-442(2017)中，将该文献通过援引以其全文并入本文。

[0227] 在一些实施例中，本文所述的CRISPR系统可以用于多重复用的误差稳健性荧光原位杂交(MERFISH)中。这些方法描述于例如Chen等人，“Spatially resolved，highly multiplexed RNA profiling in single cells[在单一细胞中进行空间分辨的高度复用的RNA谱分析]，”Science[科学]，2015年4月24日；348(6233)：aaa6090，将该文献通过援引以其全文并入本文。

[0228] 核酸的跟踪和标记

[0229] 细胞过程依赖于蛋白质、RNA和DNA之间的分子相互作用网络。准确检测蛋白质-DNA和蛋白质-RNA的相互作用是理解此类过程的关键。体外邻近标记技术采用亲和标签与报告基团(例如，可光激活基团)的组合，以在体外标记目的蛋白或RNA附近的多肽和RNA。在UV辐照后，可光激活基团与跟加标签的分子很靠近的蛋白质和其他分子反应，从而标记它们。随后可以回收和鉴定经标记的相互作用分子。RNA靶向效应子蛋白例如可以用于将探针靶向至选择的RNA序列。这些应用还可以应用于疾病或难以培养的细胞类型的体内成像的动物模型中。核酸的跟踪和标记方法描述于例如US 8795965；WO 2016205764；和WO 2017070605中；将这些文献中的每一个通过援引以其全文并入本文。

[0230] 高通量筛选

[0231] 本文所述的CRISPR系统可以用于制备下一代测序(NGS)文库。例如，为了产生具有成本效益的NGS文库，可以将CRISPR系统用于破坏靶基因的编码序列，并且可以同时通过下一代测序(例如，在离子激流(Ion Torrent)PGM系统上)筛选CRISPR效应子转染的克隆。关于如何制备NGS文库的详细描述可见于例如Bell等人，“A high-throughput screening strategy for detecting CRISPR-Cas9 induced mutations using next-generation sequencing[用于使用下一代测序检测CRISPR-Cas9诱导的突变的高通量筛选策略]，”BMC Genomics[BMC基因组学]，15.1(2014)：1002，将该文献通过援引以其全文并入本文。

[0232] 工程化细胞

[0233] 微生物(例如，大肠杆菌、酵母和微藻)广泛用于合成生物学。合成生物学的发展具有广泛的效用，包括各种临床应用。例如，可编程的CRISPR系统可以用于拆分具有用于靶向细胞死亡的毒性结构域的蛋白质，例如使用癌症相关联RNA作为靶转录物。此外，涉及蛋白

质-蛋白质相互作用的途径可以在使用例如具有适当效应子诸如激酶或酶的融合复合物的合成生物系统中受到影响。

[0234] 在一些实施例中,可以将靶向噬菌体序列的RNA指导序列引入至微生物中。因此,本披露还提供了针对噬菌体感染“接种”微生物(例如,生产菌株)的方法。

[0235] 在一些实施例中,本文所提供的CRISPR系统可以用于工程化微生物,例如以改善产率或改善发酵效率。例如,本文所述的CRISPR系统可以用于工程化微生物(诸如酵母),以从可发酵糖产生生物燃料或生物聚合物,或降解源自作为可发酵糖来源的农业废弃物的植物来源木质纤维素。更具体地,本文所述的方法可以用于修饰生物燃料生产所需的内源基因的表达和/或修饰可能干扰生物燃料合成的内源基因。这些工程化微生物的方法描述于例如Verwaal等人,“CRISPR/Cpf1 enables fast and simple genome editing of Saccharomyces cerevisiae[CRISPR/Cpf1实现对酿酒酵母的快速且简单的基因组编辑],”Yeast[酵母],2017年9月8日doi:10.1002/yea.3278;以及Hlavova等人,“Improving microalgae for biotechnology—from genetics to synthetic biology[改善生物技术中的微藻-从遗传学到合成生物学],”Biotechnol.Adv.[生物技术进展],2015年11月1日;33:1194-203中,将这些文献中的每一个通过援引以其全文并入本文。

[0236] 在一些实施例中,本文所提供的CRISPR系统可以用于工程化真核细胞或真核生物。例如,本文所述的CRISPR系统可以用于工程化真核细胞,这些真核细胞不限于植物细胞、真菌细胞、哺乳动物细胞、爬行动物细胞、昆虫细胞、禽细胞、鱼细胞、寄生虫细胞、节肢动物细胞、无脊椎动物细胞、脊椎动物细胞、啮齿动物细胞、小鼠细胞、大鼠细胞、灵长类动物细胞、非人灵长类细胞或人细胞。在一些实施例中,真核细胞在体外培养物中。在一些实施例中,真核细胞是体内的。在一些实施例中,真核细胞是离体的。

[0237] 基因驱动

[0238] 基因驱动是一种现象,在这种现象中,特定的基因或一组基因的遗传被有利地偏向。本文所述的CRISPR系统可以用于构建基因驱动。例如,CRISPR系统可以被设计为靶向并破坏基因的特定等位基因,从而引起细胞拷贝第二等位基因以固定序列。因为拷贝,第一等位基因将转换成第二等位基因,增加了第二等位基因传给后代的机会。关于如何使用本文所述的CRISPR系统以构建基因驱动的详细方法描述于例如Hammond等人,“A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae[靶向疟蚊载体冈比亚按蚊中雌性繁殖的CRISPR-Cas9基因驱动系统],”Nat.Biotechnol.[自然生物技术],2016年1月;34(1):78-83,将该文献通过援引以其全文并入本文。

[0239] 池化筛选

[0240] 如本文所述,池化CRISPR筛选是用于鉴定参与诸如细胞增殖、药物抗性和病毒感染等生物学机制的基因的有力工具。用本文所述的RNA指导物编码载体文库成批转导细胞,并且在应用选择性激发之前和之后测量gRNA的分布。池化CRISPR筛选对于影响细胞存活和增殖的机制工作良好,它们可以扩展至测量单独基因的活性(例如,通过使用工程化报告细胞系)。阵列CRISPR筛选(其中一次只靶向一个基因)使得能够使用RNA-seq作为读出。在一些实施例中,本文所述的CRISPR系统可以用于单细胞CRISPR筛选中。关于池化CRISPR筛选的详细描述可见于例如Datlinger等人,“Pooled CRISPR screening with single-cell

transcriptome read-out[具有单细胞转录组读出的池化CRISPR筛选],"Nat.Methods.[自然方法],2017年3月;14(3):297-301,将该文献通过援引以其全文并入本文。

[0241] 饱和诱变("过度攻击(Bashing)")

[0242] 本文所述的CRISPR系统可以用于原位饱和诱变。在一些实施例中,池化的RNA指导物文库可以用于对特定基因或调控元件执行原位饱和诱变。此类方法可以揭示这些基因或调控元件(例如,增强子)的关键的最小特征和离散脆弱性。这些方法描述于例如Canver等人,"BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis[通过Cas9介导的原位饱和诱变进行BCL11A增强子解析],"Nature[自然],2015年11月12日;527(7577):192-7中,将该文献通过援引以其全文并入本文。

[0243] 治疗性应用

[0244] 在一些实施例中,本文所述的CRISPR系统可以用于编辑靶核酸以修饰靶核酸(例如,通过插入、缺失或突变一个或多个氨基酸残基)。例如,在一些实施例中,本文所述的CRISPR系统包括包含所期望核酸序列的外源供体模板核酸(例如,DNA分子或RNA分子)。在解析到用本文所述的CRISPR系统诱导的切割事件后,细胞的分子机器可以在修复和/或解决切割事件中利用外源供体模板核酸。替代性地,细胞的分子机器可以在修复和/或解决切割事件中利用内源模板。在一些实施例中,本文所述的CRISPR系统可以用于修饰靶核酸,从而导致插入、缺失和/或点突变。在一些实施例中,插入是无痕插入(即,在解析到切割事件后将预期的核酸序列插入至靶核酸中未导致另外的非预期核酸序列)。供体模板核酸可以是双链或单链核酸分子(例如,DNA或RNA)。设计外源供体模板核酸的方法描述于例如WO 2016094874中,将该文献的全部内容通过援引明确并入本文。

[0245] 在另一方面,本披露提供了本文所述的系统在选自下组的方法中的用途,该组由以下组成:RNA序列特异性干扰;RNA序列特异性基因调控;筛选RNA、RNA产物、lncRNA、非编码RNA、核RNA或mRNA;诱变;抑制RNA剪接;荧光原位杂交;育种;诱导细胞休眠;诱导细胞周期停滞;减少细胞生长和/或细胞增殖;诱导细胞无变应性;诱导细胞凋亡;诱导细胞坏死;诱导细胞死亡;或诱导程序化细胞死亡。

[0246] 本文所述的CRISPR系统可以具有各种治疗性应用。在一些实施例中,新的CRISPR系统可以用于治疗各种疾病和障碍,例如遗传障碍(例如,单基因疾病)或可以通过核酸酶活性(例如,Pcsk9靶向或BCL11a靶向)治疗的疾病。在一些实施例中,将在此所述的方法用于治疗受试者,诸如哺乳动物,诸如人患者。哺乳动物受试者也可以是驯养哺乳动物,诸如狗、猫、马、猴、兔、大鼠、小鼠、牛、山羊或绵羊。

[0247] 这些方法可以包括具有感染性的病症或疾病,并且其中感染原选自下组,该组由以下组成:人类免疫缺陷病毒(HIV)、单纯疱疹病毒-1(HSV1)和单纯疱疹病毒-2(HSV2)。

[0248] 在一方面,本文所述的CRISPR系统可以用于治疗由RNA、毒性RNA和/或突变型RNA(例如,剪接缺陷或截短)的过表达引起的疾病。例如,毒性RNA的表达可以与核内含物的形成以及脑、心脏或骨骼肌的迟发型退行性变化相关联。在一些实施例中,障碍是肌强直性营养不良。在肌强直性营养不良中,毒性RNA的主要致病作用是螯合结合蛋白并且损害选择性剪接的调控(参见例如,Osborne等人,"RNA-dominant diseases[RNA显性疾病],"Hum.Mol.Genet.[人类分子遗传学],2009年4月15日;18(8):1471-81)。肌强直性营养不良(肌营养不良性肌强直症(DM))对于遗传学家而言是特别受关注的,因为它产生了极其广泛

范围的临床特征。DM的经典形式(现称为DM 1型(DM1))是由编码细胞溶质蛋白激酶的基因DMPK的3'-非翻译区(UTR)中CTG重复的扩增引起的。如本文所述的CRISPR系统可以靶向过表达的RNA或毒性RNA(例如,DMPK基因)或者DM1骨骼肌、心脏或脑中的任何错误调控的选择性剪接。

[0249] 本文所述的CRISPR系统还可以靶向影响引起各种疾病的RNA依赖性功能的反式作用突变,这些各种疾病例如像普拉德威利综合征(Prader Willi syndrome)、脊髓性肌萎缩症(SMA)和先天性角化不良。可以使用本文所述的CRISPR系统治疗的疾病列表汇总于Cooper等人,"RNA and disease[RNA和疾病],"Cell[细胞],136.4(2009):777-793和WO 2016205764中,将这些文献中的每一个通过援引以其全文并入本文。

[0250] 本文所述的CRISPR系统也可以用于治疗各种tau蛋白病,包括例如原发性和继发性tau蛋白病,诸如原发性年龄相关tau蛋白病(PART)/神经元纤维缠结(NFT)突出的老年性痴呆(其中NFT与阿尔茨海默病(AD)中看到的那些相似,但没有斑块)、拳击员痴呆(慢性创伤性脑病)和进展性核上性麻痹。tau蛋白病的可用列表和治疗这些疾病的方法描述于例如WO 2016205764中,将该文献通过援引以其全文并入本文。

[0251] 本文所述的CRISPR系统也可以用于靶向破坏可能引起剪接缺陷和疾病的顺式作用剪接密码的突变。这些疾病包括例如由缺失SMN1基因引起的运动神经元退行性疾病(例如,脊髓性肌萎缩症)、杜氏(Duchenne)肌营养不良(DMD)、与染色体17有关联的额颞叶痴呆和帕金森综合征(FTDP-17)以及囊性纤维化。

[0252] 本文所述的CRISPR系统可以进一步用于抗病毒活性、特别是抗RNA病毒。效应子蛋白可以使用经选择以靶向病毒RNA序列的合适RNA指导物靶向病毒RNA。

[0253] 此外,可以使用体外RNA感测测定来检测特定RNA底物。RNA靶向效应子蛋白可以用于活细胞中的基于RNA的感测。应用的实例是通过感测例如疾病特异性RNA进行的诊断。

[0254] 本文所述的CRISPR系统的治疗性应用的详细描述可见于例如US 8795965、EP 3009511、WO 2016205764和WO 2017070605中,将这些文献中的每一个通过援引以其全文并入本文。

[0255] 在植物中的应用

[0256] 本文所述的CRISPR系统在植物中具有广泛多种的效用。在一些实施例中,CRISPR系统可以用于工程化植物基因组(例如,改善产量、制备具有所期望翻译后修饰的产物、或引入用于生产工业产物的基因)。在一些实施例中,CRISPR系统可以用于将所期望的性状引入至植物(例如,对基因组进行或不进行可遗传修饰),或调控植物细胞或整株植物中内源基因的表达。

[0257] 在一些实施例中,CRISPR系统可以用于鉴定、编辑和/或沉默编码特定蛋白质(例如,过敏性蛋白质(例如,花生、大豆、扁豆、豌豆、四季豆和绿豆中的过敏性蛋白质))的基因。关于如何鉴定、编辑和/或沉默编码蛋白质的基因的详细描述是描述于例如Nicolaou等人,"Molecular diagnosis of peanut and legume allergy[花生和豆类过敏的分子诊断],"Curr.Opin.Allergy Clin.Immunol.[过敏和临床免疫学新见],11(3):222-8(2011)和WO 2016205764中,将这些文献中的每一个通过援引以其全文并入本文。

[0258] CRISPR系统的递送

[0259] 通过本披露和本领域中的知识,本文所述的CRISPR系统、其组分、其核酸分子或者

编码或提供其组分的核酸分子可以通过各种递送系统诸如载体(例如,质粒或病毒递送载体)来递送。本文所披露的CRISPR效应子和/或任何RNA(例如,RNA指导物)可以使用合适的载体,例如质粒或病毒载体来递送,这些病毒载体诸如腺相关病毒(AAV)、慢病毒、腺病毒和其他病毒载体或其组合。可以将效应子和一种或多种RNA指导物包装至一种或多种载体,例如质粒或病毒载体中。

[0260] 在一些实施例中,通过例如以下方式将载体(例如,质粒或病毒载体)递送至目的组织:肌肉内注射、静脉内施用、经皮施用、鼻内施用、口服施用或粘膜施用。这种递送可以是经由单剂量或者多剂量进行的。本领域技术人员应理解的是,本文有待递送的实际剂量可以在很大程度上根据多种因素而变化,该多种因素包括但不限于载体选择、靶细胞、生物体、组织、有待治疗的受试者的一般状况、所寻求的转化/修饰的程度、施用途径、施用方式和所寻求的转化/修饰的类型。

[0261] 在某些实施例中,递送是经由腺病毒进行的,这些腺病毒可以是含有至少$1x10^5$个腺病毒颗粒(也称为颗粒单位,pu)的一个剂量。在一些实施例中,剂量优选地是至少约$1x10^6$个颗粒、至少约$1x10^7$个颗粒、至少约$1x10^8$个颗粒和至少约$1x10^9$个颗粒的腺病毒。递送方法和剂量描述于例如WO 2016205764和US 8454972中,将这些文献中的每一个通过援引以其全文并入本文。

[0262] 在一些实施例中,递送是经由质粒进行的。剂量可以是足以引发应答的质粒数量。在一些情况下,质粒组合物中质粒DNA的合适量可以是从约0.1至约2mg。质粒通常包括(i)启动子;(ii)可操作地连接至启动子、编码核酸靶向CRISPR效应子的序列;(iii)选择性标记物;(iv)复制起点;以及(v)在(ii)的下游并可操作地连接至(ii)的转录终止子。质粒还可以编码CRISPR复合物的RNA组分,但是这些组分中的一种或多种可以代替地编码在不同的载体上。施用频率在医学或兽医学从业者(例如,医师、兽医师)或本领域技术人员的范围之内。

[0263] 在另一个实施例中,递送是经由脂质体或脂质转染配制品等进行的,并且可以通过本领域技术人员已知的方法制备。此类方法描述于例如WO 2016205764、US 5593972、US 5589466和US 5580859中,将这些文献中的每一个通过援引以其全文并入本文。

[0264] 在一些实施例中,递送是经由纳米颗粒或外来体进行的。例如,外来体已显示在递送RNA方面是特别有用的。

[0265] 将本文所述的CRISPR系统的一种或多种组分引入至细胞的另外手段是通过使用细胞穿透肽(CPP)。在一些实施例中,将细胞穿透肽连接至CRISPR效应子。在一些实施例中,将CRISPR效应子和/或RNA指导物偶联至一种或多种CPP以便转运至细胞(例如,植物原生质体)中。在一些实施例中,将CRISPR效应子和/或一种或多种RNA指导物由偶联至用于细胞递送的一种或多种CPP的一个或多个环状或非环状DNA分子编码。

[0266] CPP是来源于蛋白质或嵌合序列的少于35个氨基酸的短肽,这些短肽能够以非受体依赖性方式将生物分子跨细胞膜转运。CPP可以是阳离子肽、具有疏水性序列的肽、两亲性肽、具有富含脯氨酸且抗微生物的序列的肽和嵌合肽或二分肽。CPP的实例包括例如Tat(其是1型HIV病毒复制所需的核转录激活蛋白)、穿膜肽、卡波西成纤维细胞生长因子(FGF)信号肽序列、整合素β3信号肽序列、多精氨酸肽Args序列、富含鸟嘌呤的分子转运蛋白和甜箭肽。CPP和使用它们的方法描述于例如Hällbrink 等人,“Prediction of cell-

penetrating peptides[细胞穿透肽的预测],"Methods Mol.Biol.[分子生物学方法],2015;1324:39-58;Ramakrishna等人,"Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA[通过细胞穿透肽介导的Cas9蛋白和指导RNA递送进行的基因破坏],"Genome Res.[基因组研究],2014年6月;24(6):1020-7;以及WO 2016205764,将这些文献中的每一个通过援引以其全文并入本文。

[0267] 用于本文所述的CRISPR系统的各种递送方法还描述于例如US 8795965、EP 3009511、WO 2016205764和WO 2017070605中,将这些文献中的每一个通过援引以其全文并入本文。

[0268] 实例

[0269] 本发明在以下实例中进一步描述,这些实例不限制权利要求中所述的本发明的范围。

[0270] 实例1-对CLUST.200916 CRISPR-Cas系统的组分的鉴定

[0271] 使用上文所述的计算方法鉴定此蛋白质家族。CLUST.200916系统包括与从环境收集的未培养宏基因组序列中发现的CRISPR系统相关联的单效应子,这些环境不限于废水、淡水沼泽森林土壤、淡水沉积物、甲壳类动物、微生物垫、根围和土壤环境(表6)。示例性CLUST.200916效应子包括下表6和7中所示的那些。这些系统的直接重复序列和间隔子长度的实例示出于表8中。任选地,该系统包括包含在表9中列出的非编码序列的tracrRNA。

[0272] 表6.代表性CLUST.200916效应子蛋白

[0273]

来源	效应子登录号	间隔子#	效应子大小	SEQ ID NO
废水-工业废水-沉积物	3300013232|Ga0170573_11130214_1|M	7	830	1
水生-淡水-沼泽森林土壤	3300004152|Ga0062386_100000828_7|M	17	781	2
水生-淡水-沼泽森林土壤	3300027824|Ga0209040_10001283_7|P	17	781	3
水生-淡水-沉积物	3300031746|Ga0315293_10002498_12|M	5	790	4
水生-淡水-沉积物	3300031885|Ga0315285_10000472_48|P	3	790	5
水生-淡水-沉积物	3300032053|Ga0315284_10070270_2|M	6	790	6
水生-淡水-沉积物	3300032397|Ga0315287_10017659_12|M	3	782	7
水生-海生-虫穴	3300032136|Ga0316201_10000001_22|M	9	774	8
水生-海生-虫穴	3300032136|Ga0316201_10000001_31|P	9	769	9
甲壳类动物宏基因组	SRR3139690_618635_49|M	3	769	10
甲壳类动物宏基因组	SRR3139691_122305_24|P	3	769	11
微生物垫宏基因组	SRR6448207_105533_15|M	2	812	12
植物-根面-柳枝稷根围	3300005841|Ga0068863_100056305_4|P	4	708	13
根围宏基因组	SRR5578837_7398444_24|M	6	666	14
根围宏基因组	SRR5578837_7398444_30|P	6	708	15
土壤宏基因组	SRR3984957_953281_100|P	2	717	16
土壤宏基因组	SRR5271166_784995_23|P	17	781	17
土壤宏基因组	SRR7094831_6289921_1|M	2	769	18
土壤宏基因组	SRR8554505_3408779_2|M	4	805	19
土壤宏基因组	SRR8554505_3408779_1|P	4	801	20
陆生-土壤-森林土壤	3300001471|JGI12712J15308_10000506_8|P	2	717	21
陆生-土壤-森林土壤	3300027908|Ga0209006_10000286_53|P	2	717	22
废水宏基因组	SRR6837557_1806432_3|M	6	749	23
废水宏基因组	SRR6837562_732703_2|P	3	773	24
废水宏基因组	SRR6837569_616344_2|P	3	773	25
废水宏基因组	SRR6837570_202287_11|M	6	774	26
废水宏基因组	SRR6837571_88368_1|P	3	749	27
废水宏基因组	SRR6837575_1071878_1|M	2	767	28

[0274]

废水宏基因组	SRR6837577_75829_4|M	5	749	29

[0275] 表7.代表性CLUST.200916效应子蛋白的氨基酸序列

[0276]

>3300013232|Ga0170573_11130214_1|M
[废水-工业废水-沉积物]
MPKIKKPTEISLLRKEVFPDLHFAKDRMRAASLVLKNEGREAAIEYLRVNHEDKPPNFMPPAKTPYVALSRPLEQWPIAQASIA
IQKYIFGLTKDEFSATKKLLYGDKSTPNTESRKRWFEVTGVPNFGYMSAQGLNAIFSGALARYEGVVQKVENRNKKRFEKLSEK
NQLLIEEGQPVKDYVPDTAYHTPETLQKLAENNHVRVEDLGDMIDRLVHPPGIHRSIYGYQQVPPFAYDPDNPKGIILPKAYAG
YTRKPHDIIEAMPNRLNIPEGQAGYIPEHQRDKLKKGGRVKRLRTTRVRVDATETVRAKAEALNAEKARLRGKEAILAVFQIEE
DWALIDMRGLLRNVYMRKLIAAGELTPTTLLGYFTETLTLDPRRTEATFCYHLRSEGALHAEYVRHGKNTRELLLDLTKDNEKI
ALVTIDLGQRNPLAAAIFRVGRDASGDLTENSLEPVSRMLLPQAYLDQIKAYRDAYDSFRQNIWDTALASLTPEQQRQILAYEA
YTPDDSKENVLRLLLGGNVMPDDLPWEDMTKNTHYISDRYLADGGDPSKVWFVPGPRKRKKNAPPLKKPPKPRELVKRSDHNIS
HLSEFRPQLLKETRDAFEKAKIDTERGHVGYQKLSTRKDQLCKEILNWLEAEAVRLTRCKTMVLGLEDLNGPFFNQGKGKVRGW
VSFFRQKQENRWIVNGFRKNALARAHDKGKYILELWPSWTSQTCPKCKHVHADNRHGDDFVCLQCGARLHADAEVATWNLAVVA
IQGHSLPGPVREKSNDRKKSGSARKSKKANESGKVVGAWAAQATPKRATSKKETGTARNPVYNPLETQASCPAP（SEQ ID NO:
1）

>3300004152|Ga0062386_100000828_7|M
[水生-淡水-沼泽森林土壤]
MRQPAEKTAFQVFRQEVIGTQKLSGGDAKTAGRLYKQGKMEAAREWLLKGARDDVPPNFQPPAKCLVVAVSHPFEEWDISKTNH
DVQAYIYAQPLQAEGHLNGLSEKWEDTSADQHKLWFEKTGVPDRGLPVQAINKIAKAAVNRAFGVVRKVENRNEKRRSRDNRIA
EHNRENGLTEVVREAPEVATNADGFLLHPPGIDPSILSYASVSPVPYNSSKHSFVRLPEEYQAYNVEPDAPIPQFVVEDRFAIP
PGQPGYVPEWQRLKCSTNKHRRMRQWSNQDYKPKAGRRAKPLEFQAHLTRERAKGALLVVMRIKEDWVVFDVRGLLRNVEWRKV
LSEEAREKLTLKGLLDLFTGDPVIDTKRGIVTFLYKAEITKILSKRTVKTKNARDLLLRLTEPGEDGLRREVGLVAVDLGQTHP
IAAAIYRIGRTSAGALESTVLHRQGLREDQKEKLKEYRKRHTALDSRLRKEAFETLSVEQQKEIVTVSGSGAQITKDKVCNYLG
VDPSTLPWEKMGSYTHFISDDFLRRGGDPNIVHFDRQPKKGKVSKKSQRIKRSDSQWVGRMRPRLSQETAKARMEADWAAQNEN
EEYKRLARSKQELARWCVNTLLQNTRCITQCDEIVVVIEDLNVKSLHGKGAREPGWDNFFTPKTENRWFIQILHKTFSELPKHR
GEHVIEGCPLRTSITCPACSYCDKNSRNGEKFVCVACGATFHADFEVATYNLVRLATTGMPMPKSLERQGGGEKAGGARKARKK
AKQVEKIVVQANANVTMNGASLHSP（SEQ ID NO: 2）

>3300027824|Ga0209040_10001283_7|P
[水生-淡水-沼泽森林土壤]
MRQPAEKTAFQVFRQEVIGTQKLSGGDAKTAGRLYKQGKMEAAREWLLKGARDDVPPNFQPPAKCLVVAVSHPFEEWDISKTNH
DVQAYIYAQPLQAEGHLNGLSEKWEDTSADQHKLWFEKTGVPDRGLPVQAINKIAKAAVNRAFGVVRKVENRNEKRRSRDNRIA
EHNRENGLTEVVREAPEVATNADGFLLHPPGIDPSILSYASVSPVPYNSSKHSFVRLPEEYQAYNVEPDAPIPQFVVEDRFAIP
PGQPGYVPEWQRLKCSTNKHRRMRQWSNQDYKPKAGRRAKPLEFQAHLTRERAKGALLVVMRIKEDWVVFDVRGLLRNVEWRKV
LSEEAREKLTLKGLLDLFTGDPVIDTKRGIVTFLYKAEITKILSKRTVKTKNARDLLLRLTEPGEDGLRREVGLVAVDLGQTHP
IAAAIYRIGRTSAGALESTVLHRQGLREDQKEKLKEYRKRHTALDSRLRKEAFETLSVEQQKEIVTVSGSGAQITKDKVCNYLG
VDPSTLPWEKMGSYTHFISDDFLRRGGDPNIVHFDRQPKKGKVSKKSQRIKRSDSQWVGRMRPRLSQETAKARMEADWAAQNEN
EEYKRLARSKQELARWCVNTLLQNTRCITQCDEIVVVIEDLNVKSLHGKGAREPGWDNFFTPKTENRWFIQILHKTFSELPKHR
GEHVIEGCPLRTSITCPACSYCDKNSRNGEKFVCVACGATFHADFEVATYNLVRLATTGMPMPKSLERQGGGEKAGGARKARKK
AKQVEKIVVQANANVTMNGASLHSP（SEQ ID NO: 3）

>3300031746|Ga0315293_10002498_12|M
[水生-淡水-沉积物]

MTPSPQIARLVETPLAAALKAHHPGKKFRSDYLKKAGKILKDQGVEAAMAHLDGKDQAEPPNFKPPAKCRIVARSREFSEWPIV
KASVEIQKYIYGLTLEERKACDPGKSSASHKAWFAKTGVNTFGYSSVQGFNLIFGHTLGRYDGVLVKTENLNKKRAEKNERFRA
KALAEGRAEPVCPPLVTATNDTGQDVTLEDGRVVRPGQLLQPPGINPNIYAYQQVSPKAYVPGIIELPEEFQGYSRDPNAVILP
LVPRDRLSIPKGQPGYVPEPHREGLTGRKDRRMRRYYETERGTKLKRPPLTAKGRADKANEALLVVVRIDSDWVVMDVRGLLRN
ARWRRLVSKEGITLNGLLDLFTGDPVLNPKDCSVSRDTGDPVNDPRHGVVTFCYKLGVVDVCSKDRPIKGFRTKEVLERLTSSG
TVGMVSIDLGQTNPVAAAVSRVTKGLQAETLETFTLPDDLLGKVRAYRAKTDRMEEGFRRNALRKLTAEQQAEITRYNDATEQQ
AKALVCSTYGIGPEEVPWERMTSNTTYISDHILDHGGDPDTVFFMATKRGQNKPTLHKRKDKAWGQKFRPAISVETRLARQAAE
WELRRASLEFQKLSVWKTELCRQAVNYVMERTKKRTQCDVIIPVIEDLPVPLFHGSGKRDPGWANFFVHKRENRWFIDGLHKAF
SELGKHRGIYVFEVCPQRTSITCPKCGHCDPDNRDGEKFVCLSCQATLHADLDVATTNLVRVALTGKVMPRSERSGDAQTPGPA
RKARTGKIKGSKPTSAPQGATQTDAKAHLSQTGV (SEQ ID NO: 4)

>3300031885|Ga0315285_10000472_48|P
[水生-淡水-沉积物]
MTPSPQIARLVETPLAAALKAHHPGKKFRSDYLKKAGKILKDQGVEAAMAHLDGKDQAEPPNFKPPAKCRIVARSREFSEWPIV
KASVEIQKYIYGLTLEERKACDPGKSSASHKAWFAKTGVNTFGYSSVQGFNLIFGHTLGRYDGVLVKTENLNKKRAEKNERFRA
KALAEGRAEPVCPPLVTATNDTGQDVTLEDGRVVRPGQLLQPPGINPNIYAYQQVSPKAYVPGIIELPEEFQGYSRDPNAVILP
LVPRDRLSIPKGQPGYVPEPHREGLTGRKDRRMRRYYETERGTKLKRPPLTAKGRADKANEALLVVVRIDSDWVVMDVRGLLRN
ARWRRLVSKEGITLNGLLDLFTGDPVLNPKDCSVSRDTGDPVNDPRHGVVTFCYKLGVVDVCSKDRPIKGFRTKEVLERLTSSG
TVGMVSIDLGQTNPVAAAVSRVTKGLQAETLETFTLPDDLLGKVRAYRAKTDRMEEGFRRNALRKLTAEQQAEITRYNDATEQQ
AKALVCSTYGIGPEEVPWERMTSNTTYISDHILDHGGDPDTVFFMATKRGQNKPTLHKRKDKAWGQKFRPAISVETRLARQAAE
WELRRASLEFQKLSVWKTELCRQAVNYVMERTKKRTQCDVIIPVIEDLPVPLFHGSGKRDPGWANFFVHKRENRWFIDGLHKAF
SELGKHRGIYVFEVCPQRTSITCPKCGHCDPDNRDGEKFVCLSCQATLNADLDVATTNLVRVALTGKVMPRSERSGDAQTPGPA
RKARTGKIKGSKPTSAPQGATQTDAKAHLSQTGV (SEQ ID NO: 5)

[0277]

>3300032053|Ga0315284_10070270_2|M
[水生-淡水-沉积物]
MTPSPQIARLVETPLAAALKAHHPGKKFRSDYLKKAGKILKDQGVEAAMAHLDGKDQAEPPNFKPPAKCRIVARSREFSEWPIV
KASVEIQKYIYGLTLEERKACDPGKSSASHKAWFAKTGVNTFGYSSVQGFNLIFGHTLGRYDGVLVKTENLNKKRAEKNERFRA
KALAEGRAEPVCPPLVTATNDTGQDVTLEDGRVVRPGQLLQPPGINPNIYAYQQVSPKAYVPGIIELPEEFQGYSRDPNAVILP
LVPRDRLSIPKGQPGYVPEPHREGLTGRKDRRMRRYYETERGTKLKRPPLTAKGRADKANEALLVVVRIDSDWVVMDVRGLLRN
ARWRRLVSKEGITLNGLLDLFTGDPVLNPKDCSVSRDTGDPVNDPRHGVVTFCYKLGVVDVCSKDRPIKGFRTKEVLERLTSSG
TVGMVSIDLGQTNPVAAAVSRVTKGLQAETLETFTLPDDLLGKVRAYRAKTDRMEEGFRRNALRKLTAEQQAEITRYNDATEQQ
AKALVCSTYGIGPEEVPWERMTSNTTYISDHILDHGGDPDTVFFMATKRGQNKPTLHKRKDKAWGQKFRPAISVETRLARQAAE
WELRRASLEFQKLSVWKTELCRQAVNYVMERTKKRTQCDVIIPVIEDLPVPLFHGSGKRDPGWANFFVHKRENRWFIDGLHKAF
SELGKHRGIYVFEVCPQRTSITCPKCGHCDPDNRDGEKFVCLSCQATLNADLDVATTNLVRVALTGKVMPRSERSGDAQTPGPA
RKARTGKIKGSKPTSAPQGATQTDAKAHLSQTGV (SEQ ID NO: 6)

>3300032397|Ga0315287_10017659_12|M
[水生-淡水-沉积物]
MKTEKPKTALTLLREEVFPGKKYRLDVLKEAGKKLSTKGREATIEFLTGKDEERPQNFQPPAKTSIVAQSRPFDQWPIVQVSLA
VQKYIYGLTQSEFEANKKALYGETGKAISTESRRAWFEATGVDNFGFTAAQGINPIFSQAVARYEGVIKKVENRNEKKLKKLTK
KNLLRLESGEEIEDFEPEATFNEEGRLLQPPGANPNIYCYQQISPRIYDPSDPKGVILPQIYAGYDRKPEDIISAGVPNRLAIP
EGQPGYIPEHQRAGLKTQGRIRCRASVEAKARAAILAVVHLGEDWVVLDLRGLLRNVYWRKLASPGTLTLKGLLDFFTGGPVLD
ARRGIATFSYTLKSAAAVHAENTYKGKGTREVLLKLTENNSVALVTVDLGQRNPLAAMIARVSRTSQGDLTYPESVEPLTRLFL
PDPFLEEVRKYRSSYDALRLSIREAAIASLTPEQQAEIRYIEKFSAGDAKKNVAEVFGIDPTQLPWDAMTPRTTYISDLFLRMG
GDRSRVFFEVPPKKAKKAPKKPPKKPAGPRIVKRTDGMIARLREIRPRLSAETNKAFQEARWEGERSNVAFQKLSVRRKQFART
VVNHLVQTAQKMSRCDTVVLGIEDLNVPFFHGRGKYQPGWEGFFRQKKENRWLINDMHKALSERGPHRGGYVLELTPFWTSLRC

PKCGHTDSANRDGDDFVCVKCGAKLHSDLEVATANLALVAITGQSIPRPPREQSSGKKSTGTARMKKTSGETQGKGSKACVSEA
LNKIEQGTARDPVYNPLNSQVSCPAP (SEQ ID NO: 7)

>3300032136|Ga0316201_10000001_22|M
[水生-海生-虫穴]
MYNPDMKKPNNIRRIREEHFEGLCFGKDVLTKAGKIYEKDGEEAAIDFLMGKDEEDPPNFKPPAKTTIVAQSRPFDQWPIYQVS
QAVQERVFAYTEEEFNASKEALFSGDISSKSRDFWFKTNNISDQGIGAQGLNTILSHAFSRYSGVIKKVENRNKKRLKKLSKKN
QLKIEEGLEILEFKPDSAFNENGLLAQPPGINPNIYGYQAVTPFVFDPDNPGDVILPKQYEGYSRKPDDIIEKGPSRLDIPKGQ
PGYVPEHQRKNLKKKGRVRLYRRTPPKTKALASILAVLQIGKDWVLFDMRGLLRSVYMREAATPGQISAKDLLDTFTGCPVLNT
RTGEFTFCYKLRSEGALHARKIYTKGETRTLLTSLTSENNTIALVTVDLGQRNPAAIMISRLSRKEELSEKDIQPVSRRLLPDR
YLNELKRYRDAYDAFRQEVRDEAFTSLCPEHQEQVQQYEALTPEKAKNLVLKHFFGTHDPDLPWDDMTSNTHYIANLYLERGGD
PSKVFFTRPLKKDSKSKKPRKPTKRTDASISRLPEIRPKMPEDARKAFEKAKWEIYTGHEKFPKLAKRVNQLCREIANWIEKEA
KRLTLCDTVVVGIEDLSLPPKRGKGKFQETWQGFFRQKFENRWVIDTLKKAIQNRAHDKGKYVLGLAPYWTSQRCPACGFIHKS
NRNGDHFKCLKCEALFHADSEVATWNLALVAVLGKGITNPDSKKPSGQKKTGTTRKKQIKGKNKGKETVNVPPTTQEVEDIIAF
FEKDDETVRNPVYKPTGT (SEQ ID NO: 8)

>3300032136|Ga0316201_10000001_31|P
[水生-海生-虫穴]
MKKPNNIRRIREEHFEGLCFGKDVLTKAGKIYEKDGEEAAIDFLMGKDEEDPPNFKPPAKTTIVAQSRPFDQWPIYQVSQAVQE
RVFAYTEEEFNASKEALFSGDISSKSRDFWFKTNNISDQGIGAQGLNTILSHAFSRYSGVIKKVENRNKKRLKKLSKKNQLKIE
EGLEILEFKPDSAFNENGLLAQPPGINPNIYGYQAVTPFVFDPDNPGDVILPKQYEGYSRKPDDIIEKGPSRLDIPKGQPGYVP
EHQRKNLKKKGRVRLYRRTPPKTKALASILAVLQIGKDWVLFDMRGLLRSVYMREAATPGQISAKDLLDTFTGCPVLNTRTGEF
TFCYKLRSEGALHARKIYTKGETRTLLTSLTSENNTIALVTVDLGQRNPAAIMISRLSRKEELSEKDIQPVSRRLLPDRYLNEL
KRYRDAYDAFRQEVRDEAFTSLCPEHQEQVQQYEALTPEKAKNLVLKHFFGTHDPDLPWDDMTSNTHYIANLYLERGGDPSKVF
FTRPLKKDSKSKKPRKPTKRTDASISRLPEIRPKMPEDARKAFEKAKWEIYTGHEKFPKLAKRVNQLCREIANWIEKEAKRLTL
CDTVVVGIEDLSLPPKRGKGKFQETWQGFFRQKFENRWVIDTLKKAIQNRAHDKGKYVLGLAPYWTSQRCPACGFIHKSNRNGD
HFKCLKCEALFHADSEVATWNLALVAVLGKGITNPDSKKPSGQKKTGTTRKKQIKGKNKGKETVNVPPTTQEVEDIIAFFEKDD
ETVRNPVYKPTGT (SEQ ID NO: 9)

>SRR3139690_618635_49|M
[甲壳类动物宏基因组]
MEKSNTRKVIDEHFKGLLFRKDILQKAGKIYKKEGEEATISFLMGKDEEAPPNFQPPAKTSIVAQSRPFNQWPIYQVSEAIQKR
VFGYTEDEFYAQKKALFGEGGASSKSRDAWFKANGISDRGIVAQGLNMILGHAFARYEGVIQKVENRNKKRLDKLSKKNQLRVK
EGLEVYEFTPESAFIDGSGLLAQPPGISPNIYGYQAIAPFVFDPDDPRDIVLPKEYEGYSRKPDDIIEKGPNRLDIPKGQPGYV
PEHQRSGLKKGGRVWLYRRATTRAKALASILGVLQIGEDWVLFDMRGLLRNAYMRKALTPGKASARDLLGTFTEYPVLNARTGE
FTFCYKLRSGGSLYARQVYKKGKTREILTELTSEGKTIALVTVDLGQRNPVAAMVARVSRDGELSESCIDPVSRFLLPEYYARQ
IQKYRDDFDAFRQEVWDEAFASMPPEYQEQIRQYEAYTPDQAKSLVLKHFFGDEVSLDDLPWEKMTSNTCYISNLYIKRGGDPS
RVTFTPSPGKNSKKPRKPVKRTDSGISRLPEVRPGLPKDTRDAFEEAKWDVYRGHEKFPKLAKRVNQLCREIANWLEKEAGRIT
LCDTVVFGIEDMGAKFCGKGKGKFQETWEGFFRQKSENRWVMNLLKSSIHMRAHDKGRYVLELAPFYTSQRCPKCGYIHKNNRK
GDRFECLSCGALLHADSEVATWNLAVVAILGKALKKPSLKCEKSSGQKKARTSRKIQIKVGNKAETSSSPQENGEVLAPPEENS
GTSRDPVYNPSGT (SEQ ID NO: 10)

>SRR3139691_122305_24|P
[甲壳类动物宏基因组]
MEKSNTRKVIDEHFKGLLFRKDILQKAGKIYKKEGEEATISFLMGKDEEAPPNFQPPAKTSIVAQSRPFNQWPIYQVSEAIQKR
VFGYTEDEFYAQKKALFGEGGASSKSRDAWFKANGISDRGIVAQGLNMILGHAFARYEGVIQKVENRNKKRLDKLSKKNQLRVK
EGLEVYEFTPESAFIDGSGLLAQPPGISPNIYGYQAIAPFVFDPDDPRDIVLPKEYEGYSRKPDDIIEKGPNRLDIPKGQPGYV

[0278]

PEHQRSGLKKGGRVWLYRRATTRAKALASILGVLQIGEDWVLFDMRGLLRNAYMRKALTPGKASARDLLGTFTEYPVLNARTGE
FTFCYKLRSGGSLYARQVYKKGKTREILTELTSEGKTIALVTVDLGQRNPVAAMVARVSRDGELSESCIDPVSRFLLPEYYARQ
IQKYRDDFDAFRQEVWDEAFASMPPEYQEQIRQYEAYTPDQAKSLVLKHFFGDEVSLDDLPWEKMTSNTCYISNLYIKRGGDPS
RVTFTPSPGKNSKKPRKPVKRTDSGISRLPEVRPGLPKDTRDAFEEAKWDVYRGHEKFPKLAKRVNQLCREIANWLEKEAGRIT
LCDTVVFGIEDMGAKFCGKGKGKFQETWEGFFRQKSENRWVMNLLKSSIHMRAHDKGRYVLELAPFYTSQRCPKCGYIHKNNRK
GDRFECLSCGALLHADSEVATWNLAVVAILGKALKKPSLKCEKSSGQKKARTSRKIQIKVGNKAETSSSPQENGEVLAPPEENS
GTSRDPVYNPSGT (SEQ ID NO: 11)

>SRR6448207_105533_15|M
[微生物垫宏基因组]
MDMLDTETNYATETPSQQQDYSPKPPKKDRRAPKGFSKKARPEKKPPKPITLFTQKHFSGVRFLKRVIRDASKILKLSESRTIT
FLEQAIERDGSAPPDVTPPVHNTIMAVTRPFEEWPEVILSKALQKHCYALTKKIKIKTWPKKGPGKKCLAAWSARTKIPLIPGQ
VQATNGLFDRIGSIYDGVEKKVTNRNANKKLEYDEAIKEGRNPAVPEYETAYNIDGTLINKPGYNPNLYITQSRTPRLITEADR
PLVEKILWQMVEKKTQSRNQARRARLEKAAHLQGLPVPKFVPEKVDRSQKIEIRIIDPLDKIEPYMPQDRMAIKASQDGHVPYW
QRPFLSKRRNRRVRAGWGKQVSSIQAWLTGALLVIVRLGNEAFLADIRGALRNAQWRKLLKPDATYQSLFNLFTGDPVVNTRTN
HLTMAYREGVVDIVKSRSFKGRQTREHLLTLLGQGKTVAGVSFDLGQKHAAGLLAAHFGLGEDGNPVFTPIQACFLPQRYLDSL
TNYRNRYDALTLDMRRQSLLALTPAQQQEFADAQRDPGGQAKRACCLKLNLNPDEIRWDLVSGISTMISDLYIERGGDPRDVHQ
QVETKPKGKRKSEIRILKIRDGKWAYDFRPKIADETRKAQREQLWKLQKASSEFERLSRYKINIARAIANWALQWGRELSGCDI
VIPVLEDLNVGSKFFDGKGKWLLGWDNRFTPKKENRWFIKVLHKAVAELAPHRGVPVYEVMPHRTSMTCPACHYCHPTNREGDR
FECQSCHVVKNTDRDVAPYNILRVAVEGKTLDRWQAEKKPQAEPDRPMILIDNQES (SEQ ID NO: 12)

>3300005841|Ga0068863_100056305_4|P
[植物-根面-柳枝稷根围]

[0279]

MSKTKELNDYQEALARRLPGVRHQKSVRRAARLVYDRQGEDAMVAFLDGKEVDEPYTLQPPAKCHILAVSRPIEEWPIARVTMA
VQEHVYALPVHEVEKSRPETTEGSRSAWFKNSGVSNHGVTHAQTLNAILKNAYNVYNGVIKKVENRNAKKRDSLAAKNKSRERK
GLPHFKADPPELATDEQGYLLQPPSPNSSVYLVQQHLRTPQIDLPSGYTGPVVDPRSPIPSLIPIDRLAIPPGQPGYVPLHDRE
KLTSNKHRRMKLPKSLRAQGALPVCFRVFDDWAVVDGRGLLRHAQYRRLAPKNVSIAELLELYTGDPVIDIKRNLMTFRFAEAV
VEVTARKIVEKYHNKYLLKLTEPKGKPVREIGLVSIDLNVQRLIALAIYRVHQTGESQLALSPCLHREILPAKGLGDFDKYKSK
FNQLTEEILTAAVQTLTSAQQEEYQRYVEESSHEAKADLCLKYSITPHELAWDKMTSSTQYISRWLRDHGWNASDFTQITKGRK
KVERLWSDSRWAQELKPKLSNETRRKLEDAKHDLQRANPEWQRLAKRKQEYSRHLANTVLSMAREYTACETVVIAIENLPMKGG
FVDGNGSRESGWDNFFTHKKENRWMIKDIHKALSDLAPNRGVHVLEVNPQYTSQTCPECGHRDKANRDPIQRERFCCTHCGAQR
HADLEVATHNIAMVATTGKSLTGKSLAPQRLQEAAE (SEQ ID NO: 13)

>SRR5578837_7398444_24|M
[根围宏基因组]
MVAFLDGKEVDEPYTLQPPAKCHILAVSRPIEEWPIARVTMAVQEHVYALPVHEVEKSRPETTEGSRSAWFKNSGVSNHGVTHA
QTLNAILKNAYNVYNGVIKKVENRNAKKRDSLAAKNKSRERKGLPHFKADPPELATDEQGYLLQPPSPNSSVYLVQQHLRTPQI
DLPSGYTGPVVDPRSPIPSLIPIDRLAIPPGQPGYVPLHDREKLTSNKHRRMKLPKSLRAQGALPVCFRVFDDWAVVDGRGLLR
HAQYRRLAPKNVSIAELLELYTGDPVIDIKRNLMTFRFAEAVVEVTARKIVEKYHNKYLLKLTEPKGKPVREIGLVSIDLNVQR
LIALAIYRVHQTGESQLALSPCLHREILPAKGLGDFDKYKSKFNQLTEEILTAAVQTLTSAQQEEYQRYVEESSHEAKADLCLK
YSITPHELAWDKMTSSTQYISRWLRDHGWNASDFTQITKGRKKVERLWSDSRWAQELKPKLSNETRRKLEDAKHDLQRANPEWQ
RLAKRKQEYSRHLANTVLSMAREYTACETVVIAIENLPMKGGFVDGNGSRESGWDNFFTHKKENRWMIKDIHKALSDLAPNRGV
HVLEVNPQYTSQTCPECGHRDKANRDPIQRERFCCTHCGAQRHADLEVATHNIAMVATTGKSLTGKSLAPQRLQEAAE (SEQ ID
NO: 14)

>SRR5578837_7398444_30|P
[根围宏基因组]

MSKTKELNDYQEALARRLPGVRHQKSVRRAARLVYDRQGEDAMVAFLDGKEVDEPYTLQPPAKCHILAVSRPIEEWPIARVTMA
VQEHVYALPVHEVEKSRPETTEGSRSAWFKNSGVSNHGVTHAQTLNAILKNAYNVYNGVIKKVENRNAKKRDSLAAKNKSRERK
GLPHFKADPPELATDEQGYLLQPPSPNSSVYLVQQHLRTPQIDLPSGYTGPVVDPRSPIPSLIPIDRLAIPPGQPGYVPLHDRE
KLTSNKHRRMKLPKSLRAQGALPVCFRVFDDWAVVDGRGLLRHAQYRRLAPKNVSIAELLELYTGDPVIDIKRNLMTFRFAEAV
VEVTARKIVEKYHNKYLLKLTEPKGKPVREIGLVSIDLNVQRLIALAIYRVHQTGESQLALSPCLHREILPAKGLGDFDKYKSK
FNQLTEEILTAAVQTLTSAQQEEYQRYVEESSHEAKADLCLKYSITPHELAWDKMTSSTQYISRWLRDHGWNASDFTQITKGRK
KVERLWSDSRWAQELKPKLSNETRRKLEDAKHDLQRANPEWQRLAKRKQEYSRHLANTVLSMAREYTACETVVIAIENLPMKGG
FVDGNGSRESGWDNFFTHKKENRWMIKDIHKALSDLAPNRGVHVLEVNPQYTSQTCPECGHRDKANRDPIQRERFCCTHCGAQR
HADLEVATHNIAMVATTGKSLTGKSLAPQRLQEAAE (SEQ ID NO: 15)

>SRR3984957_953281_100|P
[土壤宏基因组]
MIKPTVSQFLTPGFKLIRNHSRTAGLKLKNEGEEACKKFVRENEIPKDECPNFQGGPAIANIIAKSREFTEWEIYQSSLAIQEV
IFTLPKDKLPEPILKEEWRAQWLSEHGLDTVPYKEAAGLNLIIKNAVNTYKGVQVKVDNKNKNNLAKINRKNEIAKLNGEQEIS
FEEIKAFDDKGYLLQKPSPNKSIYCYQSVSPKPFITSKYHNVNLPEEYIGYYRKSNEPIVSPYQFDRLRIPIGEPGYVPKWQYT
FLSKKENKRRKLSKRIKNVSPILGIICIKKDWCVFDMRGLLRTNHWKKYHKPTDSINDLFDYFTGDPVIDTKANVVRFRYKMEN
GIVNYKPVREKKGKELLENICDQNGSCKLATVDVGQNNPVAIGLFELKKVNGELTKTLISRHPTPIDFCNKITAYRERYDKLES
SIKLDAIKQLTSEQKIEVDNYNNNFTPQNTKQIVCSKLNINPNDLPWDKMISGTHFISEKAQVSNKSEIYFTSTDKGKTKDVMK
SDYKWFQDYKPKLSKEVRDALSDIEWRLRRESLEFNKLSKSREQDARQLANWISSMCDVIGIENLVKKNNFFGGSGKREPGWDN
FYKPKKENRWWINAIHKALTELSQNKGKRVILLPAMRTSITCPKCKYCDSKNRNGEKFNCLKCGIELNADIDVATENLATVAIT
AQSMPKPTCERSGDAKKPVRARKAKAPEFHDKLAPSYTVVLREAV (SEQ ID NO: 16)

[0280]

>SRR5271166_784995_23|P
[土壤宏基因组]
MRQPAEKTAFQVFRQEVIGTQKLSGGDAKTAGRLYKQGKMEAAREWLLKGARDDVPPNFQPPAKCLVVAVSHPFEEWDISKTNH
DVQAYIYAQPLQAEGHLNGLSEKWEDTSADQHKLWFEKTGVPDRGLPVQAINKIAKAAVNRAFGVVRKVENRNEKRRSRDNRIA
EHNRENGLTEVVREAPEVATNADGFLLHPPGIDPSILSYASVSPVPYNSSKHSFVRLPEEYQAYNVEPDAPIPQFVVEDRFAIP
PGQPGYVPEWQRLKCSTNKHRRMRQWSNQDYKPKAGRRAKPLEFQAHLTRERAKGALLVVMRIKEDWVVFDVRGLLRNVEWRKV
LSEEAREKLTLKGLLDLFTGDPVIDTKRGIVTFLYKAEITKILSKRTVKTKNARDLLLRLTEPGEDGLRREVGLVAVDLGQTHP
IAAAIYRIGRTSAGALESTVLHRQGLREDQKEKLKEYRKRHTALDSRLRKEAFETLSVEQQKEIVTVSGSGAQITKDKVCNYLG
VDPSTLPWEKMGSYTHFISDDFLRRGGDPNIVHFDRQPKKGKVSKKSQRIKRSDSQWVGRMRPRLSQETAKARMEADWAAQNEN
EEYKRLARSKQELARWCVNTLLQNTRCITQCDEIVVVIEDLNVKSLHGKGAREPGWDNFFTPKTENRWFIQILHKTFSELPKHR
GEHVIEGCPLRTSITCPACSYCDKNSRNGEKFVCVACGATFHADFEVATYNLVRLATTGMPMPKSLERQGGGEKAGGARKARKK
AKQVEKIVVQANANVTMNGASLHSP (SEQ ID NO: 17)

>SRR7094831_6289921_1|M
[土壤宏基因组]
MEKHKTKLSIIMKEFFPGERFPKNVLMQIGKKITNNKDGKETIDVKEKEDVVSFLTGKGSKKLLDFQPPAKALIVAKSRPFEEW
PIYQASKIFQEYIYGLPHNQLSIPGTSKSEHKLWLEKIGLNIGTYKDVQGLNLIFRHTKNIYEGVIKKVENKNKKNKEKIEIKN
KFEKEHGFLLTPFEEETAFDDNGKLKNPPGINNSIYCYSQVSPEATKSTTKLDNVPSIYLGYYRDIDTNIKIEYINRLSIPKGD
PGYIPLWQHELLSKKENNTRRQRKWYSNNRMKRVKRKGVSKYSDEQINQARLQDAILGKISIGEDWVLFDMRGLLRNLHWRKLV
PSQGFSPKEILEQFTGDPVIDPVRNVITFIYKDGLAHKEEIVLTKKAPDLLCKLTLNNPIGIVSIDVGQTHPQSAKFSLLKLED
DKLVAECKDRQFLPDYLLNKLFAYRERSDQLRGEINQLAMQSLSEEHQKEFNDLKIENDPTAVRIRIEKQLGIDFNNLPINDMI
YDRTTYIADAYLSIPGVDKLLVMLGTSSKKKYDSRIVKDFFKKVSKEAREALKVAFGEIQKLHPGYKKLSKSLQQWARECVNFT
HKYANKITGCTNIVFVIENLKNIRKRNGSGKRAKGYDNFFVYKKENRWVMNALQKAYIDLATHKGINIIEIQAARTSITCPKCN
CQDKNNRKGDQFNCVKCNHQANTDLEIATDNIELVALNGKGMPKIDCERSSGEENAVGARKGKKTRKIKEIQETDKNIKMDNAG
GDLLKNNRSQTAA (SEQ ID NO: 18)

[0281]

>SRR8554505_3408779_2|M
[土壤宏基因组]
MDPGMTITEGKPQEEPTELAKLLREHYPGKKLSKKDLLMAGKILAGKVAKTRPEGLVEWLADKVVDEPPNFSPPAKANIVAMSR
PFEEWPIAKATLAIESYVFGMTVDERNRLCPKETEEDRDEWFRVTGVSNYGFTSAQGLNHIFKNAFNTFDGVVTRGSNANEKKR
KEIEAQNEKRAERGEAPVPFEPRPVLTEDGHLVHPPGSKSGLRLNKIQLYQQVTDKGRGFRGQVELPLEYEDYVRDPEAPIPFG
VPRDRLCIPEGEPGYVPEWQRPLLSTKKKRRRRGWGPAGPEQVRAKAKAALLWYLPLGDDWVVLDVRGLLRNVRWRGLAPEGLS
LNGLMELFTGYPIIHNKTGDVTFKFCPEVAGVRSHEPLKKAEGRKLLLHLTKPRGEFHPRVGMVAIDLGQTNPAAFSVSRLHQV
EVEREVEVKRKLPDGETEKVTELRMVGEVRQTILSTHFLPDELVAEIKRYRNLYSAMNERHHAEAVLRLPKEAQDVYHAWQNFS
ADEAKRQLCAKYGLDPAKMPWDQMTSSTMHISKAVLAAGHDRSEVHRMVKKKGKKEREVKKRDRAWVSDFKPTLPEEVWRPKRE
HLLDIQRESEEYRRLTIFKQQLVRRCVNHVVKVGERESQCEKLVFAIEDLNVKGFFFGKGKNLPGWEGFFQHKRENRFIIREFH
RAFSELGPHRGYYVLEANPGYTSCTCPDCRHPDPVSRNGERFKCTRCGATHHADSEVATYNIAQVAIMGKALPRPKKQKKPKRE
RSGAVKKAETARKRNGRKSNGKGGQRQEAPLLRPPVRGTAREPVANASC (SEQ ID NO: 19)

>SRR8554505_3408779_1|P
[土壤宏基因组]
MTITEGKPQEEPTELAKLLREHYPGKKLSKKDLLMAGKILAGKVAKTRPEGLVEWLADKVVDEPPNFSPPAKANIVAMSRPFEE
WPIAKATLAIESYVFGMTVDERNRLCPKETEEDRDEWFRVTGVSNYGFTSAQGLNHIFKNAFNTFDGVVTRGSNANEKKRKEIE
AQNEKRAERGEAPVPFEPRPVLTEDGHLVHPPGSKSGLRLNKIQLYQQVTDKGRGFRGQVELPLEYEDYVRDPEAPIPFGVPRD
RLCIPEGEPGYVPEWQRPLLSTKKKRRRRGWGPAGPEQVRAKAKAALLWYLPLGDDWVVLDVRGLLRNVRWRGLAPEGLSLNGL
MELFTGYPIIHNKTGDVTFKFCPEVAGVRSHEPLKKAEGRKLLLHLTKPRGEFHPRVGMVAIDLGQTNPAAFSVSRLHQVEVER
EVEVKRKLPDGETEKVTELRMVGEVRQTILSTHFLPDELVAEIKRYRNLYSAMNERHHAEAVLRLPKEAQDVYHAWQNFSADEA
KRQLCAKYGLDPAKMPWDQMTSSTMHISKAVLAAGHDRSEVHRMVKKKGKKEREVKKRDRAWVSDFKPTLPEEVWRPKREHLLD
IQRESEEYRRLTIFKQQLVRRCVNHVVKVGERESQCEKLVFAIEDLNVKGFFFGKGKNLPGWEGFFQHKRENRFIIREFHRAFS
ELGPHRGYYVLEANPGYTSCTCPDCRHPDPVSRNGERFKCTRCGATHHADSEVATYNIAQVAIMGKALPRPKKQKKPKRERSGA
VKKAETARKRNGRKSNGKGGQRQEAPLLRPPVRGTAREPVANASC (SEQ ID NO: 20)

>3300001471|JGI12712J15308_10000506_8|P
[陆生-土壤-森林土壤]
MIKPTVSQFLTPGFKLIRNHSRTAGLKLKNEGEEACKKFVRENEIPKDECPNFQGGPAIANIIAKSREFTEWEIYQSSLAIQEV
IFTLPKDKLPEPILKEEWRAQWLSEHGLDTVPYKEAAGLNLIIKNAVNTYKGVQVKVDNKNKNNLAKINRKNEIAKLNGEQEIS
FEEIKAFDDKGYLLQKPSPNKSIYCYQSVSPKPFITSKYHNVNLPEEYIGYYRKSNEPIVSPYQFDRLRIPIGEPGYVPKWQYT
FLSKKENKRRKLSKRIKNVSPILGIICIKKDWCVFDMRGLLRTNHWKKYHKPTDSINDLFDYFTGDPVIDTKANVVRFRYKMEN
GIVNYKPVREKKGKELLENICDQNGSCKLATVDVGQNNPVAIGLFELKKVNGELTKTLISRHPTPIDFCNKITAYRERYDKLES
SIKLDAIKQLTSEQKIEVDNYNNNFTPQNTKQIVCSKLNINPNDLPWDKMISGTHFISEKAQVSNKSEIYFTSTDKGKTKDVMK
SDYKWFQDYKPKLSKEVRDALSDIEWRLRRESLEFNKLSKSREQDARQLANWISSMCDVIGIENLVKKNNFFGGSGKREPGWDN
FYKPKKENRWWINAIHKALTELSQNKGKRVILLPAMRTSITCPKCKYCDSKNRNGEKFNCLKCGIELNADIDVATENLATVAIT
AQSMPKPTCERSGDAKKPVRARKAKAPEFHDKLAPSYTVVLREAV (SEQ ID NO: 21)

>3300027908|Ga0209006_10000286_53|P
[陆生-土壤-森林土壤]
MIKPTVSQFLTPGFKLIRNHSRTAGLKLKNEGEEACKKFVRENEIPKDECPNFQGGPAIANIIAKSREFTEWEIYQSSLAIQEV
IFTLPKDKLPEPILKEEWRAQWLSEHGLDTVPYKEAAGLNLIIKNAVNTYKGVQVKVDNKNKNNLAKINRKNEIAKLNGEQEIS
FEEIKAFDDKGYLLQKPSPNKSIYCYQSVSPKPFITSKYHNVNLPEEYIGYYRKSNEPIVSPYQFDRLRIPIGEPGYVPKWQYT
FLSKKENKRRKLSKRIKNVSPILGIICIKKDWCVFDMRGLLRTNHWKKYHKPTDSINDLFDYFTGDPVIDTKANVVRFRYKMEN
GIVNYKPVREKKGKELLENICDQNGSCKLATVDVGQNNPVAIGLFELKKVNGELTKTLISRHPTPIDFCNKITAYRERYDKLES
SIKLDAIKQLTSEQKIEVDNYNNNFTPQNTKQIVCSKLNINPNDLPWDKMISGTHFISEKAQVSNKSEIYFTSTDKGKTKDVMK

SDYKWFQDYKPKLSKEVRDALSDIEWRLRRESLEFNKLSKSREQDARQLANWISSMCDVIGIENLVKKNNFFGGSGKREPGWDN
FYKPKKENRWWINAIHKALTELSQNKGKRVILLPAMRTSITCPKCKYCDSKNRNGEKFNCLKCGIELNADIDVATENLATVAIT
AQSMPKPTCERSGDAKKPVRARKAKAPEFHDKLAPSYTVVLREAV (SEQ ID NO: 22)

>SRR6837557_1806432_3|M
[废水宏基因组]
MPDKQTPKDTKDKPESPVSAFLKKHFPGKHFFGHAGTLARLLKTKGEEVARDYAAKKVRDEKLDFRPPAKCQIVAWSRDFSEWP
IARASATIQQHVSGLTKEDFERFDPGKSKAAHDAWFQESGVDCHGYRHVQGLNLIFANARDYYEGVVKKVENKNAQRRRRVEAL
NARRAEEGEEPIPLDVEESPFGEDGRLAHPPGVNPSIYVYQAVSPRPLKKSDLETVVLPPAYAGYDRDPSAPIPVMGDRLSIPE
GQRGHVPAWQRDQLSPDKHRRMRAWYSAANTKPKPGRTSVPDAAAIERARAEGALLVVIRIGEDWVVLDARGLLRNARWRRIAD
KEISLDGLLDLFTGDPVIDSKRNVVTFIYKAEHATATSRKVVHRKASRKALLDMTSPGEDGLPREVALASVDLGQTNAAAVRYA
RVHREGDDITSECLVRELLPDEISRDIARYRAASDRMEAEIREAAIAGLPEPMQAEVRAADASSPEAARAAVVALVGDGLPWEK
MSSATYHISDALVALGRGREAYLLSKSKDGEEKSVQRSDYGWSRHLRPRLSEETRKAMNEAVWSIKDAHEGYQKLSRRKTEIGR
RAANHVVRRLRKLAKTDKVAIAVEDLNVRMFHGGGSRSTGWDNFFVAKRENLWFVQVLHKSFCDLALHRGEVVIEVDTARTSQT
CPACGHCDPKNRSSVDREVFRCVVCGRTFHADLEVATFNIERVALTGESMPKGEEGARERGGGGKSRGGARGRNKLK (SEQ ID
NO: 23)

>SRR6837562_732703_2|P
[废水宏基因组]
MSEITDLLKANFKGKTFKSADMRMAGRILKKSGAQAVIKYLSDKGAVDPPDFRPPAKCNIIAQSRPFDEWPICKASMAIQQHIY
GLTKNEFDESSPGTSSASHEQWFAKTGVDTHGFTHVQGLNLIFQHAKKRYEGVIKKVENYNEKERKKFEGINERRSKEGMPLLE
PRLRTAFGDDGKFAEKPGVNPSIYLYQQTSPRPYDKTKHPYVHAPFELKEITTIPTQDDRLKIPFGAPGHVPEKHRSQLSMAKH
KRRRAWYALSQNKPRPPKDGSKGRRSVRDLADLKAASLADAIPLVSRVGFDWVVIDGRGLLRNLRWRKLAHEGMTVEEMLGFFS
GDPVIDPRRNVATFIYKAEHATVKSRKPIGGAKRAREELLKATASSDGVIRQVGLISVDLGQTNPVAYEISRMHQANGELVAEH
LEYGLLNDEQVNSIQRYRAAWDSMNESFRQKAIESLSMEAQDEIMQASTGAAKRTREAVLTMFGPNATLPWSRMSSNTTCISDA
LIEVGKEEETNFVTSNGPRKRTDAQWAAYLRPRVNPETRALLNQAVWDLMKRSDEYERLSKRKLEMARQCVNFVVARAEKLTQC
NNIGIVLENLVVRNFHGSGRRESGWEGFFEPKRENRWFMQVLHKAFSDLAQHRGVMVFEVHPAYSSQTCPACRYVDPKNRSSED
RERFKCLKCGRSFNADREVATFNIREIARTGVGLPKPDCERSRDVQTPGTARKSGRSLKSNKNPSEPKHVLRSKTRSNITSTLS
QNEPLATDQKTAPKTGP (SEQ ID NO: 24)

>SRR6837569_616344_2|P
[废水宏基因组]
MSEITDLLKANFKGKTFKSADMRMAGRILKKSGAQAVIKYLSDKGAVDPPDFRPPAKCNIIAQSRPFDEWPICKASMAIQQHIY
GLTKNEFDESSPGTSSASHEQWFAKTGVDTHGFTHVQGLNLIFQHAKKRYEGVIKKVENYNEKERKKFEGINERRSKEGMPLLE
PRLRTAFGDDGKFAEKPGVNPSIYLYQQTSPRPYDKTKHPYVHAPFELKEITTIPTQDDRLKIPFGAPGHVPEKHRSQLSMAKH
KRRRAWYALSQNKPRPPKDGSKGRRSVRDLADLKAASLADAIPLVSRVGFDWVVIDGRGLLRNLRWRKLAHEGMTVEEMLGFFS
GDPVIDPRRNVATFIYKAEHATVKSRKPIGGAKRAREELLKATASSDGVIRQVGLISVDLGQTNPVAYEISRMHQANGELVAEH
LEYGLLNDEQVNSIQRYRAAWDSMNESFRQKAIESLSMEAQDEIMQASTGAAKRTREAVLTMFGPNATLPWSRMSSNTTCISDA
LIEVGKEEETNFVTSNGPRKRTDAQWAAYLRPRVNPETRALLNQAVWDLMKRSDEYERLSKRKLEMARQCVNFVVARAEKLTQC
NNIGIVLENLVVRNFHGSGRRESGWEGFFEPKRENRWFMQVLHKAFSDLAQHRGVMVFEVHPAYSSQTCPACRYVDPKNRSSED
RERFKCLKCGRSFNADREVATFNIREIARTGVGLPKPDCERSRGVQTTGTARNPGRSLKSNKNPSEPKHVLRSKTRSNITSTLS
QNEPLATDQKTAPKTGP (SEQ ID NO: 25)

>SRR6837570_202287_11|M
[废水宏基因组]
MTLAELRDKYFYKIKFRKIDLRQAGKILKREGEEAARRYLDEQRESPPEGNFCPPAKCQIVGWSRPVGEWPISIASANMQQYVY
DLPKVERDKMTKFDLTSEEYAVWFAQTGIDNAGYTHVQGLNKAFKNAFRTYDGVIKKVANRNEKRKLKAEKAAERALLRGREPE

[0282]

VFVPEEALDERGFLKEKPGINRSIWTYQQVSPRPYDPTRDLKIKEKLAQRRGRSEPVAYADRLAIPEGQPGHVPQWQRDAGLLS
ANKHRRMRAHYSWHNNKPRPNRKTSRTAEECRDLGAPEAILAVIEIGEDWLAVDLRGLLRSAYYRRILSPKEVPTAAELLKLFT
GDPTIDPVREVVTFIYKEDVVPVLSTKPLRERQGLKKILDLTAPVNGVRDFIAIASIDLGVTNPAAMAYSRVRQTAAGGFDIEE
LAREFLPAAVLDQLAHHRQQWDEMEDRFRQQAVRALPEADQAECEAVFGHTGDQYAADIARALQLDGAALPWAQMSSRTTYITD
ALLARGGSPATYHTFVHSERKKGKKKKGKKKKGQEGAKPELKIKEKPSDFDWAYDFARKQLSKEVRERFNKALWEIKRTSPDYA
RMSKQKRDLGRQVANHVVKQARKLAGTQTLVVVVENINVVFFHGTGKRPVGWHQCFLPKKENRWFVQTIHKALTEIAMHKGIYV
IEVSPYYTSLHCPKCEHIDSGNRCGEQFRCLKCGYTAHADLEVAPYNIRLVALRGAGLRKVETEAEVEAVDEAPAAASVEKPRK
RRKSASGEAATFEAAPLA (SEQ ID NO: 26)

>SRR6837571_88368_1|P
[废水宏基因组]
MPDKQTPKDTKDKPESPVSAFLKKHFPGKHFFGHAGTLARLLKTKGEEVARDYAAKKVRDEKLDFRPPAKCQIVAWSRDFSEWP
IARASATIQQHVSGLTKEDFERFDPGKSKAAHDAWFQESGVDCHGYRHVQGLNLIFANARDYYEGVVKKVENKNAQRRRRVEAL
NARRAEEGEEPIPLDVEESPFGEDGRLAHPPGVNPSIYVYQAVSPRPLKKSDLETVVLPPAYAGYDRDPSAPIPVMGDRLSIPE
GQRGHVPAWQRDQLSPDKHRRMRAWYSAANTKPKPGRTSVPDAAAIERARAEGALLVVIRIGEDWVVLDARGLLRNARWRRIAD
KEISLDGLLDLFTGDPVIDSKRNVVTFIYKAEHATATSRKVVHRKASRKALLDMTSPGEDGLPREVALASVDLGQTNAAAVRYA
RVHREGDDITSECLVRELLPDEISRDIARYRAASDRMEAEIREAAIAGLPEPMQAEVRAADASSPEAARAAVVALVGDGLPWEK
MSSATYHISDALVALGRGREAYLLSKSKDGEEKSVQRSDYGWSRHLRPRLSEETRKAMNEAVWSIKDAHEGYQKLSRRKTEIGR
RAANHVVRRLRKLAKTDKVAIAVEDLNVRMFHGGGSRSTGWDNFFVAKRENRWFVQVLHKSFCDLALHRGEVVIEVDPARTSQT
CPACGHCDPKNRSSVDREVFRCVVCGRTFHADLEVATFNIERVALTGESMPKGEEGARERGGGGKSRGGARGRNKLK (SEQ ID
NO: 27)

>SRR6837575_1071878_1|M
[0283] [废水宏基因组]
MSEITDLLKANFKGKTFKSADMRMAGRILKKSGAQAVIKYLSDKGAVDPPDFRPPAKCNIIAQSRPFDEWPICKASMAIQQHIY
GLTKNEFDESSPGTSSASHEQWFAKTGVDTHGFTHVQGLNLIFQHAKKRYEGVIKKVENYNEKERKKFEGINERRSKEGMPLLE
PRLRTAFGDDGKFAEKPGVNPSIYLYQQTSPRPYDKTKHPYVHAPFELKEITTIPTQDDRLKIPFGAPGHVPEKHRSQLSMAKH
KRRRAWYALSQNKPRPPKDGSKGRRSVRDLADLKAASLADAIPLVSRVGFDWVVIDGRGLLRNLRWRKLAHEGMTVEEMLGFFS
GDPVIDPRRNVATFIYKAEHATVKSRKPIGGAKRAREELLKATASSDGVIRQVGLISVDLGQTNPVAYEISRMHQANGELVAEH
LEYGLLNDEQVNSIQRYRAAWDSMNESFRQKAIESLSMEAQDEIMQASTGAAKRTREAVLTMFGPNATLPWSRMSSNTTCISDA
LIEVGKEEETNFVTSNGPRKRTDAQWAAYLRPRVNPETRALLNQAVWDLMKRSDEYERLSKRKLEMARQCVNFVVARAEKLTQC
NNIGIVLENLVVRNFHGSGRRESGWEGFFEPKRENRWFMQVLHKAFSDLAQHRGVMVFEVHPAYSSQTCPACRYVDPKNRSSED
RERFKCLKCGRSFNADREVATFNIREIARTGVGLPKPDRERSRDVQTPGTARKSGRSLKSQDNPSEPKRVLQSKTRKKITSTET
QNEPLATDLKT (SEQ ID NO: 28)

>SRR6837577_75829_4|M
[废水宏基因组]
MPDKQTPKDTKDKPESPVSAFLKKHFPGKHFFGHAGTLARLLKTKGEEVARDYAAKKVRDEKLDFRPPAKCQIVAWSRDFSEWP
IARASATIQQHVSGLTKEDFERFDPGKSKAAHDAWFQESGVDCHGYRHVQGLNLIFANARDYYEGVVKKVENKNAQRRRRVEAL
NARRAEEGEEPIPLDVEESPFGEDGRLAHPPGVNPSIYVYQAVSPRPLKKSDLETVVLPPAYAGYDRDPSAPIPVMGDRLSIPE
GQRGHVPAWQRDQLSPDKHRRMRAWYSAANTKPKPGRTSVPDAAAIERARAEGALLVVIRIGEDWVVLDARGLLRNARWRRIAD
KEISLDGLLDLFTGDPVIDSKRNVVTFIYKAEHATATSRKVVHRKASRKALLDMTSPGEDGLPREVALASVDLGQTNAAAVRYA
RVHREGDDITSECLVRELLPDEISRDIARYRAASDRMEAEIREAAIAGLPEPMQAEVRAADASSPEAARAAVVALVGDGLPWEK
MSSATYHISDALVALGRGREAYLLSKSKDGEEKSVQRSDYGWSRHLRPRLSEETRKAMNEAVWSIKDAHEGYQKLSRRKTEIGR
RAANHVVRRLRKLAKTDKVAIAVEDLNVRMFHGGGSRSTGWDNFFVAKRENRWFVQVLHKSFCDLALHRGEVVIEVDPARTSQT
CPACGHCDPKNRSSVDREVFRCVVCGRTFHADLEVATFNIERVALTGESMPKGEEGARERGGGGKSRGGARGRNKLK (SEQ ID
NO: 29)

[0284] 表8.代表性CLUST.200916直接重复和间隔子长度的核苷酸序列

[0285]

CLUST.164091 效应子蛋白登录号	直接重复核苷酸序列	间隔子长度
3300013232|Ga0170573_11130214_1|M（SEQ ID NO: 1）	CAACCTAAACGATGGCTCGATTCGTCGAGAC（SEQ ID NO: 30） GTCTCGACGAATCGAGCCATCGTTTAGGTTG（SEQ ID NO: 122）	25-38
3300004152|Ga0062386_100000828_7|M（SEQ ID NO: 2）	GTAGAAGACCTCGCTGATTGCTCGGTGCGCCGAGAC（SEQ ID NO: 31） GTCTCGGCGCACCGAGCAATCAGCGAGGTCTTCTAC（SEQ ID NO: 123）	34-39
3300027824|Ga0209040_10001283_7|P（SEQ ID NO: 3）	GTAGAAGACCTCGCTGATTGCTCGGTGCGCCGAGAC（SEQ ID NO: 31） GTCTCGGCGCACCGAGCAATCAGCGAGGTCTTCTAC（SEQ ID NO: 123）	34-39
3300031746|Ga0315293_10002498_12|M（SEQ ID NO: 4）	GGTTGAACCCTCAACAGATTGCTCGGTAAGCCGAGAC（SEQ ID NO: 32） GTCTCGGCTTACCGAGCAATCTGTTGAGGGTTCAACC（SEQ ID NO: 124）	35-43
3300031885|Ga0315285_10000472_48|P（SEQ ID NO: 5）	GGTTGAACCCTCAACAGATTGCTCGGTAAGCCGAGAC（SEQ ID NO: 32） GTCTCGGCTTACCGAGCAATCTGTTGAGGGTTCAACC（SEQ ID NO: 124）	35-36
3300032053|Ga0315284_10070270_2|M（SEQ ID NO: 6）	GTTGAACCCTCAACAGATTGCTCGGTAAGCCGAGAC（SEQ ID NO: 33） GTCTCGGCTTACCGAGCAATCTGTTGAGGGTTCAAC（SEQ ID NO: 125）	36-44
3300032397|Ga0315287_10017659_12|M（SEQ ID NO: 7）	CTTGAAATCCTGTCAGATTGCTCCCTTCGGGGAGAC（SEQ ID NO: 34） GTCTCCCCGAAGGGAGCAATCTGACAGGATTTCAAG（SEQ ID NO: 126）	36-37
3300032136|Ga0316201_10000001_22|M（SEQ ID NO: 8）	GCTGGAAGACTCAATGATGGCTCCTTACGAGGAGAC（SEQ ID NO: 35） GTCTCCTCGTAAGGAGCCATCATTGAGTCTTCCAGC（SEQ ID NO: 127）	30-46
3300032136|Ga0316201_10000001_31|P（SEQ ID NO: 9）	GCTGGAAGACTCAATGATGGCTCCTTACGAGGAGAC（SEQ ID NO: 35） GTCTCCTCGTAAGGAGCCATCATTGAGTCTTCCAGC（SEQ ID NO: 127）	30-46
SRR3139690_618635_49|M（SEQ ID NO: 10）	CTTAGAATACTCAATGATGGCTCCTTACGAGGAGAC（SEQ ID NO: 36） GTCTCCTCGTAAGGAGCCATCATTGAGTATTCTAAG（SEQ ID NO: 128）	36-37
SRR3139691_122305_24|P（SEQ ID NO: 11）	CTTAGAATACTCAATGATGGCTCCTTACGAGGAGAC（SEQ ID NO: 36） GTCTCCTCGTAAGGAGCCATCATTGAGTATTCTAAG（SEQ ID NO: 128）	36-37
SRR6448207_105533_15|M（SEQ ID NO: 12）	ATCGGCAGCTGGTCCACCTTGG（SEQ ID NO: 37） CCAAGGTGGACCAGCTGCCGAT（SEQ ID NO: 129）	20-22
3300005841|Ga0068863_100056305_4|P（SEQ ID NO: 13）	GTCTCTCCGTAGAGAGCAATCGTTATCCATTGAGAG（SEQ ID NO: 38） CTCTCAATGGATAACGATTGCTCTCTACGGAGAGAC（SEQ ID NO: 130）	36-38
SRR5578837_7398444_24|M（SEQ ID NO: 14）	GTCTCTCCGTAGAGAGCAATCGTTATCCATTGAGAG（SEQ ID NO: 38） CTCTCAATGGATAACGATTGCTCTCTACGGAGAGAC（SEQ ID NO: 130）	36-46
SRR5578837_7398444_30|P（SEQ ID NO: 15）	GTCTCTCCGTAGAGAGCAATCGTTATCCATTGAGAG（SEQ ID NO: 38） CTCTCAATGGATAACGATTGCTCTCTACGGAGAGAC（SEQ ID NO: 130）	36-46
SRR3984957_953281_100|P（SEQ ID NO: 16）	CTTTCAAGACTAATAGATTGCTCCTTACGAGGAGAC（SEQ ID NO: 39） GTCTCCTCGTAAGGAGCAATCTATTAGTCTTGAAAG（SEQ ID NO: 131）	36-38
SRR5271166_784995_23|P（SEQ ID NO: 17）	GTAGAAGACCTCGCTGATTGCTCGGTGCGCCGAGAC（SEQ ID NO: 31） GTCTCGGCGCACCGAGCAATCAGCGAGGTCTTCTAC（SEQ ID NO: 123）	34-39
SRR7094831_6289921_1|M（SEQ ID NO: 18）	GTAGAAAGGTTTACTAATTGCTCCTTACGAGGAGAC（SEQ ID NO: 40） GTCTCCTCGTAAGGAGCAATTAGTAAACCTTTCTAC（SEQ ID NO: 132）	39
SRR8554505_3408779_2|M（SEQ ID NO: 19）	CCCGGAAGACCAGATGATGGCTCGATCAGTCGAGAC（SEQ ID NO: 41） GTCTCGACTGATCGAGCCATCATCTGGTCTTCCGGG（SEQ ID NO: 133）	35-50
SRR8554505_3408779_1|P（SEQ ID NO: 20）	CCCGGAAGACCAGATGATGGCTCGATCAGTCGAGAC（SEQ ID NO: 41） GTCTCGACTGATCGAGCCATCATCTGGTCTTCCGGG（SEQ ID NO: 133）	35-50
3300001471|JGI12712J15308	CTTTCAAGACTAATAGATTGCTCCTTACGAGGAGAC（SEQ ID NO: 39）	36-38

[0286]

_10000506_8|P (SEQ ID NO: 21)	GTCTCCTCGTAAGGAGCAATCTATTAGTCTTGAAAG (SEQ ID NO: 131)	
3300027908|Ga0209006_10000286_53|P(SEQ ID NO: 22)	CTTTCAAGACTAATAGATTGCTCCTTACGAGGAGAC (SEQ ID NO: 39) GTCTCCTCGTAAGGAGCAATCTATTAGTCTTGAAAG (SEQ ID NO: 131)	36-38
SRR6837557_1806432_3|M (SEQ ID NO: 23)	GTCGAGACCGATGACGAGTGCGCGGTGCGCCGCGAC (SEQ ID NO: 42) GTCGCGGCGCACCGCGCACTCGTCATCGGTCTCGAC (SEQ ID NO: 134)	36-48
SRR6837562_732703_2|P (SEQ ID NO: 24)	CCTGCAAGGGATCCAAATTGCTCTGTTCGCAGAGAC (SEQ ID NO: 43) GTCTCTGCGAACAGAGCAATTTGGATCCCTTGCAGG (SEQ ID NO: 135)	36-42
SRR6837569_616344_2|P (SEQ ID NO: 25)	CCTGCAAGGGATCCAAATTGCTCTGTTCGCAGAGAC (SEQ ID NO: 43) GTCTCTGCGAACAGAGCAATTTGGATCCCTTGCAGG (SEQ ID NO: 135)	38-39
SRR6837570_202287_11|M (SEQ ID NO: 26)	GCGCCAACGACCTCTGATTGTCCGGTACGCCGGAAC (SEQ ID NO: 44) GTTCCGGCGTACCGGACAATCAGAGGTCGTTGGCGC (SEQ ID NO: 136)	36-38
SRR6837571_88368_1|P (SEQ ID NO: 27)	GTCGAGAGCGATGACGAGTGCGCGGTGCGCCGCGAC (SEQ ID NO: 45) GTCGCGGCGCACCGCGCACTCGTCATCGCTCTCGAC (SEQ ID NO: 137)	38-41
SRR6837575_1071878_1|M (SEQ ID NO: 28)	CCATCAATGGATCCAAATTGCTCTGTACGCAGAGAC (SEQ ID NO: 46) GTCTCTGCGTACAGAGCAATTTGGATCCATTGATGG (SEQ ID NO: 138)	36-37
SRR6837577_75829_4|M (SEQ ID NO: 29)	GTCGAGAGCGATGACGAGTGCGCGGTGCGCCGCGAC (SEQ ID NO: 45) GTCGCGGCGCACCGCGCACTCGTCATCGCTCTCGAC (SEQ ID NO: 137)	36-50

[0287] 表9.代表性CLUST.200916系统的非编码序列

[0288]

>3300005841|Ga0068863_100056305_4|P (SEQ ID NO: 13 的效应子)
GAGGATAAGCTCGCGTTCGGCCGCCAGCTCTTCTGGGCTCATGTCGGCGAGCTTGTGGGCTACCACAGCGAACGGCTTGCTCAT
GGTCTGCAAGTGTTCTGGCAGGTGGGCGTACGTGAAAAATTGGAGCGTGTGGTTTGTGTCAATCATGACGTTGACTACACCGAC
GTGACATCTCTGTGCCTTTGGTTTTGGACGGGCACCGGTGGTGTAGGGTTGGACAGCAGCCGTGTCGCAGCGGGGGAGGTCTTG
GGGAGAATGGTTCTCTTTGGGGCCTTTTCTATTGGTGGTCTTGGTTCGCCATCCTCGTCCATACCAACCCTCCAAGTATTTCCC
GCGGAGGATAACACGAAATGCTGCGATGGCGTCTCTCCGTAGAGAGCAATCGTTACTCGTTGAGATATGTACAGCAACTTAGAC
ACTGGGTGCTGGGCCAGCTGTCTCCCCGTAGGGAGCAATCGTGCGCGAACACGATCACTTGTCTCGTACGGCTACCACGGCTCC
AGCAGAAGCTCAATCATACTCACCACATACACCGCAGATTGCGAGAGGGACGGAAGAAAAACCATGAACTGTGAAAAATAGTAC
AAAGATTCTGTGCAGATTAAACTACTGAGCTTTTGATATTATTGGTTTTCGAGCGGGTAACAGCCTTTGTGCACAACCAAACCA
CTCGCATTTCAAACCTTCGTATGGCCCAGGAGCGGTGTAATGCAGCTTCGGTTGCCCACCGAGGTGGGACTTCAGGAGCAAGGA
CATGTCGAAAACCAAAGAACTCAATGACTATCAAGAAGCGCTAGCGCGTCGCCTTCCTGGTGTCCGACACCAAAAGTCGGTACG
CCGAGCTGCGCGACTCGTCTACGACAGACAAGGGGAAGATGCCATGGTCGCCTTCCTCGACGGCAAACGTGAGCGTTTTTGCTG
CACCCACTGCGGAGCACAGCGACACGCTGACCTCGAAGTGGCGACACACAACATCGCCATGGTAGCCACCACAGGCAAATCTCT
AACAGGCAAATCGTTGGCGCCACAACGACTCCAGGAAGCTGCAGAGTAACGATTGCTCCCTATG (SEQ ID NO: 47)

>3300027824|Ga0209040_10001283_7|P (SEQ ID NO: 3 的效应子)
CTTGATGCCGTCGACGACGAAGCTCGCCGGCAGCGCCCGGCGCTCGCCCTCGGGGTACTACTGGCTCGGCCAGCCGCCCCAGAA
GAGAACGAAATTCCCACGGACAGTGACTTCGTAGAGATCGGCGTGCGTAGGGCCTTCGGCGGGCATGGAGAACTCCCTGGTGAA
GACGCCTTTAGGAGCTAGGCGGCAGGATTGCCGCTGTTTTCACGCTTGTGCGTGACCTCGACCTCGAACGCGGAGGACACGCTA
ACACGTAAGAATCTAACTTGGCAAGTACAGGCTTTCTTTTCTTGATCCCGAATTAGATGCGTAGTAGCCGTGAAATAGGAGACA
AGATCCTCATGCGTCAGCCCGCGGAGAAGACCGCGTTCCAAGTCTTTCGTCAGGAAGTGATCGGGACACAGAAACTATCGGGAG
GCGATGCCAAGACTGCCGGACGGCTCTACAAGCAGGGAAAGATGGAAGCAGCACGTGAGTGGTTGCTCAAAGGT (SEQ ID NO: 48)

>3300013232|Ga0170573_11130214_1|M (SEQ ID NO: 1 的效应子)

CGCATCTGCAACGGTCTACTCGATGCTGGATTGATTCCTGCGGTGTAGAATAGCATAACCAGGGTCCGCTGGACTATCCAGCAA
CGACGGCTCGATACGTCGAGCCGATGGAGAAGTCATGCCAAAGATCAAGAAACCGACTGAGATTTCCCTGCTACGCAAGGAGGT
GTTCCCTGACCTCCACTTTGCCAAGGACCGGATGCGTGCTGCATCCTTGGTCCTTAAGAATGAGGGCCGGGAAGCCGCCATCGA
GTACCTCCGGGTGAAT (SEQ ID NO: 49)

>3300027908|Ga0209006_10000286_48|M (SEQ ID NO: 22 的效应子)
GGAATGAGTGGAGAAGGAGTGGTTTGTAAAGGTACTTGGAATAAGAAAAAGAATAGACCAAACATGTTCAAGATTAAGTCTAAT
GCTTGGATTGGAAGACTTAAAGAATACTGCAACGGTAATATGGACCTAATGGCCAGACTACTATAATGTTTGATCTTAAAGCAG
AAATTGTATAAATAATGATATCAAAAATGATTAAGCCTACGGTAAGTCAGTTTCTTACGCCTGGATTTAAGTTAATTCGGAATC
ACTCAAGAACTGCTGGTCTAAAACTAAAAAATGAAGGCGAAGAGGCTTGTAAGAAATTCGTTCGAGAAAATGAAATTCCTGTCG
CTATAACGGCCCAGTCGATGCCGAAACCAACTTGCGAGCGCTCGGGCGACGCTAAAAAGCCTGTCCGTGCTCGCAAGGCTAAAG
CACCGGAATTTCACGATAAATTGGCACCAAGCTATACGGTTGTGCTAAGGGAGGCTGTTTGAAGACATAACCGCTCGCAAACCG
AACGATAAATATGCGAAATCATTGAAGAATGGCGAGGCAT (SEQ ID NO: 50)

>SRR6837562_732703_2|P (SEQ ID NO: 24 的效应子)
GTCGCGCCGAGGCCCATTGGCTTTTATGACTATGGCTCTTTCGACATGGCTCGATGGGAGGACTTCATCGAAAAATGGATACCG
GAGCAGTTGCCAACGTCGTGGGATCACCTCATGGATGACTTGGAGATCGCGGATGCCGCACGATAGGTATTGCGCCTCCTTCAT
GACGTTTTCTAAGATTTTACGTTCGGTGTAGACGTGAGTATGTCCGAAATCACTGATCTGCTCAAAGCCAATTTCAAAGGGAAG
ACATTCAAAAGTGCAGACATGAGGATGGCTGGGAGGATCTTCAAGAAAAGCGGCGCGCAGGCCGTTATCAAATACTTGTCGGAC
AAGGGGGCAGTTGATCCGCCTAAGTCCGGCAGGTCCTTGAAATCAAACAAGAATCCGAGCGAGCCCAAACACGTTTTGCGGTCA
AAAACACGTAGCAACATAACATCTACGTTGTCGCAAAATGAACCGCTCGCAACGGACCAAAAAACCGCTCCAAAAACAGGACCT
TAGACCCCACGCCCCGCGGTTCCCGCTTTT (SEQ ID NO: 51)

[0289] >SRR6837557_1806432_3|M (SEQ ID NO: 23 的效应子)
CGCCCTCGCGACGCGCATCCCGGGGTTTCGTTTTCCGGAGACGATGCAGCTCACCTGCGCCACGGAGCAGCCGAGACGGGCGGC
GACCTTGGTCGTGCCCTCGGACTCGACGATCTTCACGAACTTCTTGTGCGTCTGTTCTGGCGACATGCGGGGTAGCTTGCGACG
GTTTAGGTCGCCTGTCAAGCGAGCTGGCGAAAAGCTGTTGACAGGAGCATGGGCTTTGGCTAAGGTGCGAGCATGCCCGACAAG
CAGACGCCCAAGGACACCAAGGACAAGCCCGAGAGCCCCGTCTCCGCCTTCCTCAAGAAGCACTTCCCCGGGAAGCATTTCTTC
GGGCATGCCGGGACGCTCGCCCGCCTCCTCAAGACCAAGGGTGAGGAGGTTGCGTTCCACGCCGACCTCGAAGTGGCGACTTTC
AACATCGAGCGGGTGGCGCTGACGGGCGAGTCGATGCCGAAAGGCGAGGAGGGCGCACGCGAGCGGGGAGGTGGTGGGAAGAGC
AGGGGAGGCGCTCGCGGGCGAAATAAATTGAAATAGTTCGCGAAATCGGGTAAGCTGGTTGGCGTCAGTGAAAACTGAATAGAC
GAAGATACGAGCCGCTCGCAGGGTGGAGCGGAAGATCTCGATCTGGTTCCGAATTTCTCGGTGGCTGTCGAGAGCGATGACGAG
TGCGCGGTGCGCCGCGACCAGCTCTGTGTGTAGTTGACCCGTAGTCAGGACAAGGAGAGGTCGAGAGCGATGACGAGTGCGCGG
TGCGCCGCGACGTGCTGTTGATCTGATAGACGGGGGTGGGCACCTGCGCGGTGTCGAGAGCGATGACGAGTGCGCGGTGCGCCG
CGACTTTCGTACAGAGCCTCCGGACGGCGTCCGGATGCCGGAGTCGAGAGCGATGACGAGTGCGCGGTGCGCCGCGACGCCGTC
GCCTCGAAGGCGGGCGCGGGTGAACTACTTCGTCAAGACCGATGACGATTGCGGTGCGCCGCGACGGCTTGCGCGAGGACTTTT
TGGTCGGCTTCTTGCCGTCGTCAAGACCGATGACGAGTGCGCGGTGCGCCGCGACGACTTGGCCTTCGCCTCGGTCAGGGCGCG
TCGGTCGACAATTTCAGTGCAATGGCTCGGTACACCGGGACTATGATTGTTCGGTACGCCGAGACACGGTCGGCCAGCTGGTTA
TCTCACGTCCCTCACCACCCCCACGTACGCCATCATCTCCCTCATCCTCACCTCCCACCCGTGGTTCTCCGCCGCGAGCCTCGC
CCCCCGCTCGGCGAT (SEQ ID NO: 52)

>SRR6837575_1071878_1|M (SEQ ID NO: 28 的效应子)
GTCGCGCCGAGGCCCATTGGCTTTTATGACTATGGCTCTTTCGACATGGCTCGATGGGAGGACTTCATCGAAAAATGGGTACTG
GAGCAGTTGCCAACGTCGTGGGATCACCTCATGGATGACTTGGAGATCGCGGATGCCGCACGATAGGTATTGCGCCTCCTTCAT
GACGTTTTCTAAGATTTTACGTTCGGTGTAGACGTGAGTATGTCCGAAATCACTGATCTGCTCAAAGCCAATTTCAAAGGGAAG
ACATTCAAAAGTGCAGACATGAGGATGGCTGGGAGGATCTTCAAGAAAAGCGGCGCGCAGGCCGTTATCAAATACTTGTCGGAC
AAGGGGGCAGTTGATCCGCCTACCCCGGGGACCGCTCGCAAGTCCGGCAGGTCCTTGAAATCACAAGATAATCCGAGCGAGCCC

AAACGCGTTTTGCAGTCAAAAACACGTAAGAAAATCACATCTACGGAGACACAAAACGAACCGCTCGCAACGGACCTAAAAACC
TAAGTGAAAACAGGGGCCTAGACACACATCCAAGGATGTCGCAAAGGGCCTCGTCGAGATCCCAAG (SEQ ID NO: 53)

>SRR6837570_202287_11|M (SEQ ID NO: 26 的效应子)
CTGCTGCGCGCGTACCCGCTCTGGGAGAGTCGCGACTGCGGTCAGCACGACTCGCGCGGCGAGGGCGTGGAGGGCTGCTATTGC
GCGCGACAAGATCTGAAGCACCTGATCGGCTGGCCGGAGCCGAAGCAAAAAAGCTCCTGCAAATGATCTATCTCTGTTGACTTC
GCATCAGACGATCACTAGATTGCTACAGGCAATCACAAACAGCGGGAGACGAACAATGACTCTGGCCGAGCTGCGCGACAAATA
CTTCTACAAGATCAAGTTCCGCAAGATCGATCTCAGGCAAGCCGGCAAGATCCTCAAGAGAGAGGGCGAGGAAGCGGCTCGCCG
CTATCTTGACGAGCAGCGGGAGTCGCCGCCCGAGGGCCGCGGCGCGGGGCTGCGGAAGGTCGAGACTGAGGCCGAGGTCGAGGC
GGTCGACGAGGCCCCCGCGGCCGCGAGCGTGGAGAAGCCGCGCAAGCGGCGCAAGAGCGCGAGCGGCGAGGCGGCGACTTTCGA
GGCCGCACCGCTCGCGTAGTCTAAGATCGCGTTCTCGCAAGAGAATGCGAGCACCGGCCGCGAACACCGCTCTAGTTGTGCGAA
CGTTCGCGGCCGGTGCTCGCAGACGAGCTGCTAGGTCTTTGAAAATTGAATAGATTGTAATGGTGAGTTGCCCGAGCCCGCACT
CGCGGCGCGCTACCTCTTCGCGCCAACGACC (SEQ ID NO: 54)

>SRR6837577_75829_4|M (SEQ ID NO: 29 的效应子)
[0290] CGCCCTCGCGACGCGCATCCCGGGGTTTCGTTTTCCGGAGACGATGCAGCTCACCTGCGCCACGGAGCAGCCGAGACGGGCGGC
GACCTTGGTCGTGCCCTCGGACTCGACGATCTTCACGAACTTCTTGTGCGTCTGTTCTGGCGACATGCGGGGTAGCTTGCGACG
GTTTAGGTCGCCTGTCAAGCGAGCTGGCGAAAAGCTGTTGACAGGAGCATGGGCTTTGGCTAAGGTGCGAGCATGCCCGACAAG
CAGACGCCCAAGGACACCAAGGACAAGCCCGAGAGCCCCGTCTCCGCCTTCCTCAAGAAGCACTTCCCCGGGAAGCATTTCTTC
GGGCATGCCGGGACGCTCGCCCGCCTCCTCAAGACCAAGGGTGAGGAGGTTGCGTTCCACGCCGACCTCGGGGTGGCGACTTTC
AACATCGAGCGGGTGGCGCTGACGGGCGAGTCGATGCCGAAAGGCGAGGAGGGCGCACGCGAGCGGGGAGGTGGTGGGAAGAGC
AGGGGAGGCGCTCGCGGGCGAAATAAATTGAAATAGTTCGCGAAATCGGGTAAGCTGGTTGGCGTCAGTGAAAACTGAATAGAC
GAAGATACGAGCCGCTCGCAGGGTGGAGCGGAAGATCTCGATCTGGTTCCGAATTTCTCGGTGGCT (SEQ ID NO: 55)

>SRR6837569_616344_2|P (SEQ ID NO: 25 的效应子)
GTCGCGCCGAGGCCCATTGGCTTTTATGACTATGGCTCTTTCGACATGGCTCGATGGGAGGACTTCATCGAAAAATGGGTGCCG
GAGCAGTTGCCAACGTCGTGGGATCACCTCATGGATGACTTGGAGATCGCGGATGCCGCACGATAGGTATTGCGCCTCCTTCAT
GACGTTTTCTAAGATTTTACGTTCGGTGTAGACGTGAGTATGTCCGAAATCACTGATCTGCTCAAAGCCAATTTCAAAGGGAAG
ACATTCAAAAGTGCAGACATGAGGATGGCTGGGAGGACCCTCAAGAAAAGCGGCGCGCAGGCCGTTATCAAATACTTGTCGGAC
AAGGGGGCGGTTGATCCGCCTAATCCCGGCAGGTCCTTGAAATCAAACAAGAATCCGAGCGAGCCCAAACACGTTTTGCGGTCA
AAAACACGTAGCAACATAACATCTACGTTGTCGCAAAATGAACCGCTCGCAACGGACCAAAAAACCGCTCCAAAAACAGGACCT
TAGACCCCACGCCGGAAGTCGCT (SEQ ID NO: 56)

[0291] 实例2-工程化CLUST.200916 CRISPR-Cas系统的功能验证

[0292] 鉴定了CLUST.200916 CRISPR-Cas系统的组分，选择以下基因座用于功能验证：1)来自命名为3300013232的宏基因组来源的基因座(SEQ ID NO:1)、2)来自命名为SRR6837570的宏基因组来源的基因座(SEQ ID NO:26)、3)来自命名为SRR6837575的宏基因组来源的基因座(SEQ ID NO:28)、4)来自命名为SRR6837577的宏基因组来源的基因座(SEQ ID NO:29)和5)来自命名为SRR6837569的宏基因组来源的基因座(SEQ ID NO:25)。

[0293] DNA合成和效应子文库克隆

[0294] 为了测试示例性CLUST.200916 CRISPR-Cas系统的活性，使用pET28a(+)载体设计和合成系统。简而言之，合成编码如表7中所示的CLUST.200916 3300013232效应子(SEQ ID NO:1)、CLUST.200916 SRR6837570效应子(SEQ ID NO:26)、CLUST.200916 SRR6837575效应子(SEQ ID NO:28)、CLUST.200916 SRR6837577效应子(SEQ ID NO:29)和CLUST.200916 SRR6837569效应子(SEQ ID NO:25)的大肠杆菌密码子优化的核酸序列并且将其单独地克隆至来源于pET-28a(+)(EMD密理博公司(EMD Millipore))的定制表达系统中。载体包含在lac启动子控制下编码CLUST.200916效应子的核酸和大肠杆菌核糖体结合序列。载体还包含在CLUST.200916效应子的开放阅读框后、用于由J23119启动子驱动的CRISPR阵列文库的

接受位点。用于CLUST.200916 3300013232效应子(SEQ ID NO:1)的非编码序列在SEQ ID NO:49中列出,用于CLUST.200916 SRR6837570效应子(SEQ ID NO:26)的非编码序列在SEQ ID NO:54中列出,用于CLUST.200916 SRR6837575效应子(SEQ ID NO:28)的非编码序列在SEQ ID NO:53中列出,用于CLUST.200916 SRR6837577效应子(SEQ ID NO:29)的非编码序列在SEQ ID NO:55中列出,并且用于CLUST.200916 SRR6837569效应子(SEQ ID NO:25)的非编码序列在SEQ ID NO:56中列出,如表9中所示。测试另外的条件,其中将CLUST.200916效应子在没有非编码序列的情况下单独地克隆至pET28a(+)中。参见图1A。

[0295] 计算设计含有"重复-间隔子-重复"序列的寡核苷酸文库合成(OLS)池,其中"重复"表示CRISPR阵列中发现的与效应子相关联的共有直接重复序列,并且"间隔子"表示拼接pACYC184质粒和或大肠杆菌必需基因的序列。特别地,用于CLUST.200916 3300013232效应子(SEQ ID NO:1)的重复序列在SEQ ID NO:30中列出,用于CLUST.200916 SRR6837570效应子(SEQ ID NO:26)的重复序列在SEQ ID NO:44中列出,用于CLUST.200916 SRR6837575效应子(SEQ ID NO:28)的重复序列在SEQ ID NO:46中列出,用于CLUST.200916 SRR6837577效应子(SEQ ID NO:29)的重复序列在SEQ ID NO:45中列出,并且用于CLUST.200916 SRR6837569效应子(SEQ ID NO:25)的重复序列在SEQ ID NO:43中列出,如表8中所示。间隔子长度由内源CRISPR阵列中发现的间隔子长度的模式决定。将重复-间隔子-重复序列附加有限制性位点,从而使得能够将片段双向克隆至前面提及的CRISPR阵列文库受体位点中;和独特的PCR引发位点,以使得能够从更大的池中特异性扩增特定重复-间隔子-重复文库。

[0296] 接下来,使用金门组装(Golden Gate assembly)方法将重复-间隔子-重复文库克隆至质粒中。简而言之,首先使用独特的PCR引物将每种重复-间隔子-重复从OLS池(安捷伦基因组学(Agilent Genomics))中扩增出并且对其使用BsaI预线性化质粒骨架以减少潜在背景。将两种DNA片段用Ampure XP(贝克曼库尔特公司(Beckman Coulter))纯化,之后添加至金门组装主混合物(新英格兰实验室(New England Biolabs))并且根据制造商的说明书孵育。将金门反应物进一步纯化并且浓缩以使得在细菌筛选的随后步骤中实现最大转化效率。

[0297] 按照卢西根公司(Lucigen)推荐的方案使用基因Pulser Xcell®(伯乐公司(Bio-rad))将含有不同的重复-间隔子-重复元件和CRISPR效应子的质粒文库电穿孔至E.Cloni电感受态大肠杆菌(卢西根公司)中。将文库与纯化的pACYC184质粒一起共转染或直接转染至含有pACYC184的E.Cloni电感受态大肠杆菌(卢西根公司)中,铺板至BioAssay®培养皿(赛默飞世尔公司(Thermo Fisher))中的含有氯霉素(飞世尔公司)、四环素(阿法埃莎公司(Alfa Aesar))和卡那霉素(阿法埃莎公司)的琼脂上,并且在37℃下孵育10-12小时。在估计大约的菌落计数以确保细菌板上的足够文库表示之后,收获细菌,并且使用QIAprep Spin Miniprep®试剂盒(凯杰公司(Qiagen))提取质粒DNA,以产生"输出"文库。通过使用含有与依诺米那(Illumina)测序化学兼容的条形码和位点的定制引物执行PCR,从转化前"输入文库"和收获后"输出文库"产生条形码化的下一代测序文库,然后将该文库池化并加载至Nextseq 550(依诺米那公司)上以评价效应子。对于每种筛选执行至少两个独立的生物学平行测定以确保一致性。参见图1B。

[0298] 细菌筛选测序分析

[0299] 使用依诺米那bcl2fastq对筛选输入和输出文库的下一代测序数据进行解复用。每个样品的所得的fastq文件中的读段包含用于筛选质粒文库的CRISPR阵列元件。使用CRISPR阵列的直接重复序列来确定阵列取向，并且将间隔子序列映射至来源(pACYC184或E.Cloni)或阴性对照序列(GFP)以确定相应的靶标。对于每个样品，对给定质粒文库中每个独特阵列元件(r_a)的读段的总数目进行计数并且如下归一化：(r_a+1)/所有文库阵列元件的总读段。耗减得分通过将给定阵列元件的归一化输出读段除以归一化输入读段来计算。

[0300] 为了鉴定导致酶活性和细菌细胞死亡的特定参数，使用下一代测序(NGS)来定量和比较输入和输出质粒文库的PCR产物中单独CRISPR阵列(即，重复-间隔子-重复)的表示。将阵列耗减比定义为归一化输出读段计数除以归一化输入读段计数。如果耗减比小于描绘为图3、图6、图9、图12、图15、图18、图21、图24和图27中的“命中阈值”虚线的0.3(超过3倍耗减)，则阵列被认为是“强耗减”的。当计算跨生物学平行测定的阵列耗减比时，取给定CRISPR阵列跨所有实验的最大耗减比值(即强耗减阵列必须在所有生物学平行测定中强耗减)。针对每个间隔子靶标产生矩阵，该矩阵包括阵列耗减比和以下特征：靶链、转录物靶向、ORI靶向、靶序列基序、侧接序列基序和靶二级结构。研究了此矩阵中不同特征解释CLUST.200916系统靶耗减的程度。

[0301] 图3、图9、图15、图21和图27通过针对给定靶标绘制筛选输出相对于筛选输入中测序读段的归一化比率示出了在具有非编码序列的情况下工程化组合物的干扰活性的程度。针对每种DR转录取向绘制结果。在组合物的功能性筛选中，与活性RNA指导物复合的活性效应子将干扰pACYC184赋予大肠杆菌对氯霉素和四环素的抗性的能力，从而导致细胞死亡和池内间隔子元件的耗减。初始DNA文库(筛选输入)相对于存活转化的大肠杆菌(筛选输出)的深度测序结果的比较表明了实现活性的可编程CRISPR系统的特定靶序列和DR转录取向。筛选还表明效应子复合物仅在一种DR取向的情况下具有活性。如此，筛选表明CLUST.200916 3300013232效应子在DR的“正向”取向(5’-CAAC…AGAC-[间隔子]-3’)上具有活性(图3)，CLUST.200916 SRR6837570效应子在DR的“反向”取向(5’-GCGC…GAAC-[间隔子]-3’)上具有活性(图9)，CLUST.200916 SRR6837575效应子在DR的“正向”取向(5’-CCAT…AGAC-[间隔子]-3’)上具有活性(图15)，CLUST.200916 SRR6837577效应子在DR的“正向”取向(5’-GTCG…CGAC-[间隔子]-3’)上具有活性(图21)；并且CLUST.200916 SRR6837569效应子在DR的“反向”取向(5’-CCTG…AGAC-[间隔子]-3’)上具有活性(图27)。

[0302] 图4A和图4B描绘了分别靶向pACYC184和大肠杆菌E.Cloni必需基因的CLUST.200916 3300013232效应子(加上非编码序列)的强耗减靶标的位置。图10A和图10B描绘了分别靶向pACYC184和大肠杆菌E.Cloni必需基因的CLUST.200916 SRR6837570效应子(加上非编码序列)的强耗减靶标的位置。图16A和图16B描绘了分别靶向pACYC184和大肠杆菌E.Cloni必需基因的CLUST.200916 SRR6837575效应子(加上非编码序列)的强耗减靶标的位置。图22A和图22B描绘了分别靶向pACYC184和大肠杆菌E.Cloni必需基因的CLUST.200916 SRR6837577效应子(加上非编码序列)的强耗减靶标的位置。图28A和图28B描绘了分别靶向pACYC184和大肠杆菌E.Cloni必需基因的CLUST.200916 SRR6837569效应子(加上非编码序列)的强耗减靶标的位置。

[0303] 分析耗减靶标的侧接序列以确定具有非编码序列的CLUST.200916 效应子的PAM

序列。针对CLUST.200916效应子3300013232、SRR6837570、SRR6837575、SRR6837577和SRR6837569(具有非编码序列)的PAM序列的WebLogo表示(Crooks等人,Genome Research[基因组研究]14:1188-90,2004)分别示出于图5、图11、图17、图23和图29中,其中“20”位置对应于相邻于靶标的5’端的核苷酸。

[0304] 此外,图6、图12、图18、图24和图30显示CLUST.200916效应子在不存在非编码序列的情况下保留活性。与图3一致,CLUST.200916 3300013232效应子(没有非编码序列)在DR的“正向”取向(5’-CAAC…AGAC-[间隔子]-3’)上具有活性(图6)。与图9一致,CLUST.200916 SRR6837570效应子在DR的“反向”取向(5’-GCGC…GAAC-[间隔子]-3’)上具有活性(图12)。与图15一致,CLUST.200916 SRR6837575效应子在DR的“正向”取向(5’-CCAT…AGAC-[间隔子]-3’)上具有活性(图18)。与图21一致,CLUST.200916 SRR6837577效应子在DR的“正向”取向(5’-GTCG…CGAC-[间隔子]-3’)上具有活性(图24)。与图27一致,CLUST.200916 SRR6837569效应子在DR的“反向”取向(5’-CCTG…AGAC-[间隔子]-3’)上具有活性(图30)。

[0305] 图7A和图7B描绘了分别靶向pACYC184和大肠杆菌E.Cloni必需基因的CLUST.200916 3300013232效应子(没有非编码序列)的强耗减靶标的位置。图13A和图13B描绘了分别靶向pACYC184和大肠杆菌E.Cloni必需基因的CLUST.200916 SRR6837570效应子(加上非编码序列)的强耗减靶标的位置。图19A和图19B描绘了分别靶向pACYC184和大肠杆菌E.Cloni必需基因的CLUST.200916 SRR6837575效应子(加上非编码序列)的强耗减靶标的位置。图25A和图25B描绘了分别靶向pACYC184和大肠杆菌E.Cloni必需基因的CLUST.200916 SRR6837577效应子(加上非编码序列)的强耗减靶标的位置。图31A和图31B描绘了分别靶向pACYC184和大肠杆菌E.Cloni必需基因的CLUST.200916 SRR6837569效应子(加上非编码序列)的强耗减靶标的位置。

[0306] 分析耗减靶标的侧接序列以确定没有非编码序列的CLUST.200916效应子的PAM序列。针对CLUST.200916效应子3300013232、SRR6837570、SRR6837575、SRR6837577和SRR6837569(没有非编码序列)的PAM序列的WebLogo表示分别示出于图8、图14、图20、图26和图32中,其中“20”位置对应于相邻于靶标的5’端的核苷酸。

[0307] 实例3-用CLUST.200916效应子进行的双链DNA切割

[0308] 此实例展示出通过CLUST.200916效应子进行的双链DNA(dsDNA)切割。

[0309] 针对以下五种CLUST.200916效应子合成包含DR-间隔子-DR序列的RNA指导序列:SEQ ID NO:1、SEQ ID NO:28、SEQ ID NO:26、SEQ ID NO:27和SEQ ID NO:25的效应子。将RNA指导物的每种间隔子产生为与DNA靶序列的一条链具有互补性。例如,对于SEQ ID NO:1的效应子,将SEQ ID NO:139的RNA指导物的间隔子序列设计为与靶标A(SEQ ID NO:57)具有互补性,但与非靶标B(SEQ ID NO:58)不具有互补性,该非靶标B因此用作阴性对照。对应于SEQ ID NO:1、SEQ ID NO:28、SEQ ID NO:26、SEQ ID NO:27和SEQ ID NO:25的效应子的靶标、非靶标对照和RNA指导物序列示出于表10中。表10中的加粗序列(例如,SEQ ID NO:57、SEQ ID NO:59、SEQ ID NO:61、SEQ ID NO:63和SEQ ID NO:65的靶序列中的加粗序列)对应于与表10中的前crRNA(例如,分别地SEQ ID NO:139、SEQ ID NO:140、SEQ ID NO:141、SEQ ID NO:142和SEQ ID NO:143的前crRNA序列)结合的序列。

[0310] 表10.针对CLUST.200916效应子的靶标和RNA指导物序列。

[0311]

效应子	靶序列	前 crRNA 序列	非靶标对照序列
SEQ ID NO: 1	靶标 A: TCCATGTCTCGTTATACGC TGTGGTTCGCCAACAAATA CCTTACTACTCGTCACGCC ACG***CCATCGTGGTGCCAGA*** ***TCTGTCACAGGTAACT***GAA CACCTGGAAACTACTAATA CCTTTTGACAGCTAGCTCA GTCCTAGGTATAAT （SEQ ID NO: 57）	CAACCTAAACGATGGCTCG ATTCGTCGAGACAACGAAA GTGGCCTCTGGCGAAGCGG GCGGCACAACCAAAACGAT GGCTCGATTCGTCGAGACA GTTACCTGTGACAGATCTG GCACCACGATGGCAACCAA AACGATGGCTCGATTCGTC GAGAC （SEQ ID NO: 139）	非靶标 B: TCCATGTCTCGTTATACGC TGTGGTTCGCCAACGCAAG AATACTACTACTGCAACCT GTACTAACGTATGGTGACT TAACTCGTCTGGATCCTAC AACAGTAACGCACTACTAG AATACTTTGACAGCTAGCT CAGTCCTAGGTATAAT （SEQ ID NO: 58）
SEQ ID NO: 28	靶标 C: TCCATGTCTCGTTATACGC TGTGGTTCGCCAACACTCT	CCATCAATGGATCCAAATT GCTCTGTACGCAGAGACAG CGCGCCGCGGTACATCTTC	非靶标 D: TCCATGTCTCGTTATACGC TGTGGTTCGCCAACGCACC

[0312]

	CATAACTACTGCGTGAAGT GCT***TTCCGAACGCTTCGCC ACGCTCTCCTTTGATGTAC A***GAAAACTCAGCGACTACT CTCTCATATTGACAGCTAG CTCAGTCCTAGGTATAAT (SEQ ID NO: 59)	TTGTTAACTTTTTGACCAT CAATGGATCCAAATTGCTC TGTACGCAGAGACTGTACA TCAAAGGAGAGCGTGGCGA AGCGTTCGGAACCATCAAT GGATCCAAATTGCTCTGTA CGCAGAGAC (SEQ ID NO: 140)	CCCCCGACTACTCGGTAGC GTCGATGTTCTGCTGCCGT TGCCGGGGCGTCACAATAT TGCGAATGCGCTGGCACTA CTCCCCCCCCGTTGACAGCT AGCTCAGTCCTAGGTATAA T (SEQ ID NO: 60)
SEQ ID NO: 26	靶标 E: TCCATGTCTCGTTATACGC TGTGGTTCGCCAACACTGC ATACACTACTGTGACCGAC GGT***GTTGTAACTGACGAAA TTCACTACCTGTCTGCTAT C***GAAGAAGGCAACACTACT CTGCATACTTGACAGCTAG CTCAGTCCTAGGTATAAT (SEQ ID NO: 61)	GCGCCAACGACCTCTGATT GTCCGGTACGCCGGAACGA GGAGCGCAGTCACCAAAAC TTGTCCTTTCAGTTTGCGC CAACGACCTCTGATTGTCC GGTACGCCGGAACGATAGC AGACAGGTAGTGAATTTCG TCAGTTACAACGCGCCAAC GACCTCTGATTGTCCGGTA CGCCGGAAC (SEQ ID NO: 141)	非靶标 F: TCCATGTCTCGTTATACGC TGTGGTTCGCCAACGCAAG AAAGTCACTACTACTGAAA GCGATCATCTCTGATGTGA ACGCTTCCGACGAAGATCG TTGGAACGCTGTTCTACTA CTAGAAAGTCTTGACAGCT AGCTCAGTCCTAGGTATAA T (SEQ ID NO: 62)
SEQ ID NO: 27	靶标 G: TCCATGTCTCGTTATACGC TGTGGTTCGCCAACATGAG ATGGACTACTTGTTTGGTC GTG***CTGCGCGTGAAAGCCG TATTGAAAGCCTCCATGCC GAGCGT***GAAGTGCTTTCCA CTACTTGAGATGGTTGACA GCTAGCTCAGTCCTAGGTA TAAT (SEQ ID NO: 63)	GTCGAGAGCGATGACGAGT GCGCGGTGCGCCGCGACGA ACCGAGGTAACTGGCTTGG AGGAGCGCAGTCACCAAAA CGTCGAGAGCGATGACGAG TGCGCGGTGCGCCGCGACA CGCTCGGCATGGAGGCTTT CAATACGGCTTTCACGCGC AGGTCGAGAGCGATGACGA GTGCGCGGTGCGCCGCGAC (SEQ ID NO: 142)	非靶标 H: TCCATGTCTCGTTATACGC TGTGGTTCGCCAACGCAGT GGGACCACTACTGAAGGTG ATCATGCCGTTTTCAGTTT CAACGTGGTATGCACCAAT GTGCTGGGTAATGCCGCCC GACTACTGTGGGACCTTGA CAGCTAGCTCAGTCCTAGG TATAAT (SEQ ID NO: 64)
SEQ ID NO: 25	靶标 I: TCCATGTCTCGTTATACGC TGTGGTTCGCCAACATACT GGCTACTACTATGGAAATA CGC***GGGTCCGTACCTGGAG GAAATGAAGTCGCCGCATG GCTA***AAACGAAGACCGACT ACTTACTGGCTTTGACAGC TAGCTCAGTCCTAGGTATA AT (SEQ ID NO: 65)	CCTGCAAGGGATCCAAATT GCTCTGTTCGCAGAGACTG GCGCGCAAGCCGAATGCCA AAGTGGTTTATATGCACTC CTGCAAGGGATCCAAATTG CTCTGTTCGCAGAGACTAG CCATGCGGCGACTTCATTT CCTCCAGGTACGGACCCCC TGCAAGGGATCCAAATTGC TCTGTTCGCAGAGAC (SEQ ID NO: 143)	非靶标 J: TCCATGTCTCGTTATACGC TGTGGTTCGCCAACGCACA ACGCTGACTACTCCGCGTC CGGCAGTTTTGCCAGCCAG CGGCGGCCCACTTCATCGT CGGCGTTAATAATCGCCTA CTACTCAACGCTGTTGACA GCTAGCTCAGTCCTAGGTA TAAT (SEQ ID NO: 66)

[0313] 使用体外转录(IVT)制备针对所有效应子的RNA指导序列。使用具有T7启动子的商业上合成的寡核苷酸模板(IDT)通过PCR制备用于IVT反应的双链DNA。通过以下方式执行IVT:将双链DNA模板与T7 RNA聚合酶(HiScribe T7快速高产率RNA试剂盒新英格兰生物学实验室公司(NEB))一起孵育,之后用DNA酶(赛默飞世尔科技公司)处理以去除DNA模板。使用RNA制备型试剂盒(左莫研究公司(Zymo Research))净化IVT产物。

[0314] 使用经IR800标记的正向引物和未标记的反向引物经由PCR产生经标记的dsDNA靶标和非靶标底物。所得的PCR产物在间隔子互补链上包含IR800标记,如图33中所示。使用SPRI珠(安捷伦公司)纯化这些底物,并且经由nanodrop分光光度计(赛默飞世尔科技公司)

测量浓度。

[0315] 在包含25mM Tris pH 8.0、50mM NaCl、10mM MgCl2、1mM DTT的缓冲液中进行切割测定。通过以下方式来形成复合的RNP(效应子加上相应的RNA指导物):将每种纯化的效应子与来自表1的RNA指导物以1:2的比率一起孵育。然后将复合的RNP添加至100nM dsDNA底物并且孵育。将反应物用RNA酶混合物处理并且孵育,之后与蛋白酶K一起孵育。为了检测dsDNA切割,在15%TBE-尿素凝胶上分析来自反应的DNA产物。将凝胶在荧光数字成像系统(立卡生物科学公司(LI-COR Biosciences))上针对IR800荧光成像。

[0316] 如图34A、图34B、图34C、图34D和图34E中所示,分别在SEQ ID NO:1、SEQ ID NO:28、SEQ ID NO:26、SEQ ID NO:27和SEQ ID NO:25的效应子RNP的情况下观察到针对dsDNA靶标的靶标特异性切割。指示出了全长靶标/非靶标谱带和切割谱带。对于效应子RNP中的每一种,切割与效应子RNP浓度正相关,如图34A、图34B、图34C、图34D和图34E的泳道5-7中所示。在不存在RNA指导物的情况下(图34A、图34B、图34C、图34D和图34E的泳道2-4)或在不存在效应子RNP的情况下(图34A、图34B、图34C、图34D和图34E的泳道1)未观察到可检测的切割活性。此外,对于与非靶标dsDNA一起孵育的效应子RNP未观察到可检测的切割活性,如图34A、图34B、图34C、图34D和图34E的泳道12-14中所示。例如,在针对靶标A设计的RNA指导物的情况下,在非靶标B中未观察到可检测的切割(图34A的泳道12-14)。

[0317] 此实例因此显示,SEQ ID NO:1、SEQ ID NO:28、SEQ ID NO:26、SEQ ID NO:27和SEQ ID NO:25的效应子具有核酸酶活性并且催化靶标特异性dsDNA切割。

[0318] 实例4-通过CLUST.200916效应子靶向GFP

[0319] 此实例描述了使用荧光耗减测定(FDA)来测量CLUST.200916效应子的活性。

[0320] 在此测定中,被设计为靶向GFP的活性CRISPR系统结合并切割编码GFP的双链DNA区域,从而导致GFP荧光的耗减。FDA测定涉及体外转录和翻译,从而允许从编码CLUST.200916效应子的DNA模板和含有具有直接重复(DR)-间隔子-直接重复(DR)、在T7启动子下的前crRNA序列的DNA模板产生RNP;间隔子靶向GFP。在同一单锅反应中,还作为靶标和荧光报告子两者产生GFP和RFP(图35)。靶GFP质粒序列在SEQ ID NO:144中列出,并且RFP质粒序列在SEQ ID NO:145中列出。使用帝肯公司(TECAN)Infinite F Plex板读取器在37℃下每20min测量GFP和RFP荧光值,持续12h。由于未靶向RFP,因此其荧光不受影响,并且因此用作内部信号对照。

[0321] SEQ ID NO:144

[0322] ccccttgtattactgtttatgtaagcagacaggatgcgtccggcgtagaggatcgagatctcCAAAAAATGGCTGTTTTTGAAAAAAATTCTAAAGGTTGTTTTACGACAGACGATAACAGGGTTgaaataattttgtttaacttaagaaggagATTTAAATatgAAAATCGAAGAAGGTAAAGGTCACCATCACCATCACCACggatccatgacggcattgacggaaggtgcaaaactgtttgagaaagagatcccgtatatcaccgaactggaaggcgacgtcgaaggtatgaaatttatcattaaaggcgagggtaccggtgacgcgaccacgggtaccattaaagcgaaatacatctgcactacgggcgacctgccggtcccgtgggcaaccctggtgagcaccctgagctacggtgttcagtgtttcgccaagtacccgagccacatcaaggatttctttaagagcgccatgccggaaggttatacccaagagcgtaccatcagcttcgaaggcgacggcgtgtacaagacgcgtgctatggttacctacgaacgcggttctatctacaatcgtgtcacgctgactggtgagaactttaagaaagacggtcacattctgcgtaagaacgttgcattccaatgcccgccaagcattctgtatattctgcctgacaccgttaacaatggcatccgcgttgagttcaaccaggcgtacgatattgaaggtgtgaccgaaaaactggtta

ccaaatgcagccaaatgaatcgtccgttggcgggctccgcggcagtgcatatcccgcgttatcatcacattaccta
ccacaccaaactgagcaaagaccgcgacgagcgccgtgatcacatgtgtctggtagaggtcgtgaaagcggttgat
ctggacacgtatcagTAATAAaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcat
aacccctttggggcctctaaacgggtcttgaggggtttttttgctgaaaggaggaactatatccggCTTCCTCGCTCA
CTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCAC
AGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCG
TTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAA
ACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCC
GCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTC
AGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTAT
CCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGAT
TAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACA
GTATITGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAA
CCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCC
TTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGggtggcacttttcggggaaatgtgcgcg
gaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgaattaattcttagaaaaactcat
cgagcatcaaatgaaactgcaatttattcatatcaggattatcaataccatatttttgaaaaagccgtttctgtaa
tgaaggagaaaactcaccgaggcagttccataggatggcaagatcctggtatcggtctgcgattccgactcgtcca
acatcaatacaacctattaatttcccctcgtcaaaaataaggttatcaagtgagaaatcaccatgagtgacgactg
aatccggtgagaatggcaaaagtttatgcatttctttccagacttgttcaacaggccagccattacgctcgtcatc
aaaatcactcgcatcaaccaaaccgttattcattcgtgattgcgcctgagcgagacgaaatacgcgatcgctgtta
aaaggacaattacaaacaggaatcgaatgcaaccggcgcaggaacactgccagcgcatcaacaatattttcacctg
aatcaggatattcttctaatacctggaatgctgttttcccggggatcgcagtggtgagtaaccatgcatcatcagg
agtacggataaaatgcttgatggtcggaagaggcataaattccgtcagccagtttagtctgaccatctcatctgta
acatcattggcaacgctacctttgccatgtttcagaaacaactctggcgcatcgggcttcccatacaatcgataga
ttgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaa
tcgcggcctagagcaagacgtttcccgttgaatatggctcataaca

[0323] SEQ ID NO:145

[0324] ccccttgtattactgtttatgtaagcagacaggatgcgtccggcgtagaggatcgagatctcCAAAAA
ATGGCTGTTTTTGAAAAAAATTCTAAAGGTTGTTTTACGACAGACGATAACAGGGTTgaaataatttcgtttaact
ttaagaaggagATTTAAATatgAAAATCGAAGAAGGTAAAGGTCACCATCACCATCACCACggatccaTGGTCAGC
AAGGGGGAGGAAGACAATATGGCTATTATCAAGGAATTCATGCGCTTCAAGGTGCATATGGAAGGAAGCGTGAATG
GACACGAATTCGAGATCGAAGGCGAGGGGGAGGGTCGCCCTTATGAAGGCACACAAACAGCTAAACTGAAAGTGAC
GAAGGGAGGGCCGCTTCCCTTCGCTTGGGACATTCTTTCACCCCAGTTCATGTATGGTTCAAAGGCTTATGTCAAG
CACCCGGCGGACATTCCAGACTACTTAAAATTGTCGTTCCCCGAGGGGTTTAAATGGGAACGCGTTATGAATTTCG
AGGATGGGGGAGTCGTAACGGTTACCCAGGACAGTAGCCTGCAGGATGGCGAGTTCATCTACAAAGTGAAATTGCG
CGGGACGAACTTCCCTAGCGATGGGCCAGTCATGCAGAAGAAAACGATGGGATGGGAAGCGTCATCCGAGCGCATG
TATCCTGAAGATGGTGCTTTAAAAGGTGAGATCAAGCAGCGTTTGAAACTGAAGGACGGGGGCCATTATGATGCTG
AAGTTAAAACGACATATAAGGCCAAGAAGCCAGTTCAACTGCCAGGGGCTTATAATGTTAATATTAAATTAGACAT

TACGAGCCATAATGAAGATTACACGATTGTCGAGCAATACGAGCGCGCAGAAGGACGCCACTCAACGGGGGGCATG
GACGAGCTGTACAAGTAAaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcataac
cccttggggcctctaaacgggtcttgaggggtttttgctgaaaggaggaactatatccggCTTCCTCGCTCACTG
ACTCCCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGA
ATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTG
CTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACC
CGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCT
TACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGT
TCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGGACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCG
GTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAG
CAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTA
TTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCA
CCGCTCGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTT
GATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGggtggcacttttcggggaaatgtgcgcggaa
cccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgaattaattcttagaaaaactcatcga
gcatcaaatgaaactgcaatttattcatatcaggattatcaataccatatttttgaaaaagccgtttctgtaatga
aggagaaaactcaccgaggcagttccataggatggcaagatcctggtatcggtctgcgattccgactcgtccaaca
tcaatacaacctattaatttcccctcgtcaaaaataaggttatcaagtgagaaatcaccatgagtgacgactgaat
ccggtgagaatggcaaaagtttatgcatttctttccagacttgttcaacaggccagccattacgctcgtcatcaaa
atcactcgcatcaaccaaaccgttattcattcgtgattgcgcctgagcgagacgaaatacgcgatcgctgttaaaa
ggacaattacaaacaggaatcgaatgcaaaccggcgcaggaacactgccagcgcatcaacaatattttcacctgaa
tcaggatattcttctaatacctggaatgctgttttcccggggatcgcagtggtgagtaaccatgcatcatcaggag
tacggataaaatgcttgatggtcggaagaggcataaattccgtcagccagtttagtctgaccatctcatctgtaac
atcattggcaacgctacctttgccatgtttcagaaacaactctggcgcatcgggcttcccatacaatcgatagatt
gtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatc
gcggcctagagcaagacgtttcccgttgaatatggctcataaca

[0325] 针对SEQ ID NO:28和SEQ ID NO:25的效应子中的每一种设计5种GFP靶标(加上1种非靶标)。直接重复(DR)序列和间隔子长度针对每种效应子有所变化。用于FDA测定的RNA指导物序列、靶序列和非靶标对照序列列出于表11中。示出于表11中的前crRNA序列进一步包含在5'端处的T7启动子和对RNA的3'端加帽的发夹基序以确保RNA不被存在于体外转录和翻译混合物中的核酸酶降解。

[0326] 表11.用于FDA测定的RNA指导物和靶序列。

[0327]

靶标	测试的效应子	PAM 序列	前 crRNA 序列	靶序列
靶标 1	SEQ ID NO: 28	TTG	gaaattaatacgactcactatagCCATCAATGGATCCAAATTGCTCTGTACGCAGAGACagaaagagatcccgtatatcaccgaactggaaggcgC	agaaagagatcccgtatatcaccgaactggaaggcg (SEQ ID NO: 67)

[0328]

			CATCAATGGATCCAAATTGCTCTG TACGCAGAGACctaacccctctct aaacggaggggttt (SEQ ID NO: 146)	
靶标 2	SEQ ID NO: 28	TTG	gaaattaatacgactcactatagC CATCAATGGATCCAAATTGCTCTG TACGCAGAGACgtgtggtaggtaa tgtgatgataacgcgggatatgcC CATCAATGGATCCAAATTGCTCTG TACGCAGAGACctaacccctctct aaacggaggggttt (SEQ ID NO: 147)	gtgtggtaggtaat gtgatgataacgcg ggatatgc (SEQ ID NO: 68)
靶标 3	SEQ ID NO: 28	TTT	gaaattaatacgactcactatagC CATCAATGGATCCAAATTGCTCTG TACGCAGAGACaagagcgccatgc cggaaggttatacccaagagcgtC CATCAATGGATCCAAATTGCTCTG TACGCAGAGACctaacccctctct aaacggaggggttt (SEQ ID NO: 148)	aagagcgccatgcc ggaaggttataccc aagagcgt (SEQ ID NO: 69)
靶标 4	SEQ ID NO: 28	TTT	gaaattaatacgactcactatagC CATCAATGGATCCAAATTGCTCTG TACGCAGAGACaagaaagacggtc acattctgcgtaagaacgttgcaC CATCAATGGATCCAAATTGCTCTG TACGCAGAGACctaacccctctct aaacggaggggttt (SEQ ID NO: 149)	aagaaagacggtca cattctgcgtaaga acgttgca (SEQ ID NO: 70)
靶标 5	SEQ ID NO: 28	TTT	gaaattaatacgactcactatagC CATCAATGGATCCAAATTGCTCTG TACGCAGAGACgagaaagagatcc cgtatatcaccgaactggaaggcC CATCAATGGATCCAAATTGCTCTG TACGCAGAGACctaacccctctct aaacggaggggttt (SEQ ID NO: 150)	gagaaagagatccc gtatatcaccgaac tggaaggc (SEQ ID NO: 71)
非靶标 6	SEQ ID NO: 28		gaaattaatacgactcactatagG TATAACGACCCTGCGAAGTGGGGT GTAACTTCGACgtgattgcgcctg agcgagacgaaatacgcgatcgcG TATAACGACCCTGCGAAGTGGGGT GTAACTTCGACctaacccctctct aaacggaggggttt (SEQ ID NO: 151)	
靶标 7	SEQ ID NO: 25	TTG	gaaattaatacgactcactatagC CTGCAAGGGATCCAAATTGCTCTG TTCGCAGAGACacggaaggtgcaa	acggaaggtgcaaa actgtttgagaaag agatcccgtat

[0329]

			aactgtttgagaaagagatcccgt atCCTGCAAGGGATCCAAATTGCT CTGTTCGCAGAGACctaaccccctc tctaaacggaggggttt (SEQ ID NO: 152)	(SEQ ID NO: 72)
靶标 8	SEQ ID NO: 25	TTG	gaaattaatacgactcactatagC CTGCAAGGGATCCAAATTGCTCTG TTCGCAGAGACagaaagagatccc gtatatcaccgaactggaaggcga cgCCTGCAAGGGATCCAAATTGCT CTGTTCGCAGAGACctaaccccctc tctaaacggaggggttt (SEQ ID NO: 153)	agaaagagatcccg tatatcaccgaact ggaaggcgacg (SEQ ID NO: 73)
靶标 9	SEQ ID NO: 25	TTT	gaaattaatacgactcactatagC CTGCAAGGGATCCAAATTGCTCTG TTCGCAGAGACaatgataaatttc ataccttcgacgtcgccttccagt tcCCTGCAAGGGATCCAAATTGCT CTGTTCGCAGAGACctaaccccctc tctaaacggaggggttt (SEQ ID NO: 154)	aatgataaatttca taccttcgacgtcg ccttccagttc (SEQ ID NO: 74)
靶标 10	SEQ ID NO: 25	TTT	gaaattaatacgactcactatagC CTGCAAGGGATCCAAATTGCTCTG TTCGCAGAGACaagagcgccatgc cggaaggttatacccaagagcgta ccCCTGCAAGGGATCCAAATTGCT CTGTTCGCAGAGACctaaccccctc tctaaacggaggggttt (SEQ ID NO: 155)	aagagcgccatgcc ggaaggttataccc aagagcgtacc (SEQ ID NO: 75)
靶标 11	SEQ ID NO: 25	TTT	gaaattaatacgactcactatagC CTGCAAGGGATCCAAATTGCTCTG TTCGCAGAGACgagaaagagatcc cgtatatcaccgaactggaaggcg acCCTGCAAGGGATCCAAATTGCT CTGTTCGCAGAGACctaaccccctc tctaaacggaggggttt (SEQ ID NO: 156)	gagaaagagatccc gtatatcaccgaac tggaaggcgac (SEQ ID NO: 76)
非靶标 12	SEQ ID NO: 25		gaaattaatacgactcactatagG TATAACGACCCTGCGAAGTGGGGT GTAACTTCGACgtgattgcgcctg agcgagacgaaatacgcgatcgct gtGTATAACGACCCTGCGAAGTGG GGTGTAACTTCGACctaaccccctc tctaaacggaggggttt (SEQ ID NO: 157)	

[0330] 将GFP信号针对RFP信号归一化，然后在每个时间点获取三个技术平行测定的平均荧光。然后通过以下方式来计算GFP荧光耗减：将与非GFP靶向RNA指导物(其代替地靶向卡那霉素抗性基因而不耗减非GFP信号)一起孵育的效应子的GFP信号除以与GFP靶向RNA指导物一起孵育的效应子的GFP信号。所得的值在图36A和图36B中称为“耗减”。

[0331] 耗减为一或大约一(例如，10RFU/10RFU＝1)表明在关于非GFP靶向前crRNA和GFP

靶向前crRNA的GFP耗减方面存在很小或不存在差异。耗减大于一(例如,10RFU/5RFU＝2)表明在关于非GFP靶向前crRNA和GFP靶向前crRNA的GFP耗减方面存在差异。GFP信号的耗减表明,效应子形成功能性RNP并通过在GFP编码区内引入双链DNA切割来干扰GFP的产生。GFP耗减的程度在很大程度上与CLUST.200916效应子的特异性活性相关。

[0332] 图36A和36B示出了分别通过SEQ ID NO:28和SEQ ID NO:25的效应子形成的RNP的针对GFP靶标(针对SEQ ID NO:28的靶标1-5和针对SEQ ID NO:25的7-11)中的每一种每20分钟测量的耗减曲线。在每种靶标下,与SEQ ID NO:28的效应子和SEQ ID NO:25的效应子形成的RNP的耗减值大于一。

[0333] 这表明CLUST.200916效应子形成能够干扰GFP产生的功能性RNP。

[0334] 实例5-通过CLUST.200916效应子靶向哺乳动物基因

[0335] 此实例描述了通过CLUST.200916效应子进行的对哺乳动物AAVS1靶标的插入缺失评估,这些效应子通过瞬时转染引入至哺乳动物细胞中。

[0336] 将SEQ ID NO:24、SEQ ID NO:28和SEQ ID NO:25的效应子克隆至pcda3.1骨架(英杰公司(Invitrogen))中。然后将质粒大量制备并且稀释至1μg/μL。对于RNA指导物制备,通过含有靶序列支架的超聚体(ultramer)和U6启动子衍生编码RNA指导物的dsDNA片段。将超聚体重悬于pH 7.5的10mM Tris·HCl中至最终原液浓度为100μM。随后再次使用10mM Tris·HCl将工作原液稀释至10μM以充当用于PCR反应的模板。将RNA指导物的扩增在具有以下组分的50μL反应物中进行:0.02μl前面提及的模板、2.5μl正向引物、2.5μl反向引物、25μL新英格兰生物学实验室公司HiFi聚合酶和20μl水。循环条件是:1x(98℃下30s)、30x(98℃下10s,67℃下15s)、1x(72℃下2min)。将PCR产物用1.8X SPRI处理净化并且归一化至25ng/μL。针对测试的AAVS1靶基因座TGGCCTGGGTCACCTCTACGGCTG(SEQ ID NO:158)所制备的RNA指导物序列示出于表12中。

[0337] 表12.用于用CLUST.200916效应子进行的瞬时转染的RNA指导物。

[0338]

效应子	前 crRNA
SEQ ID NO: 24	CCTGCAAGGGATCCAAATTGCTCTGTTCGCAGAGACTGGCCTGGGTCACCTCTACG GCTGCCTGCAAGGGATCCAAATTGCTCTGTTCGCAGAGAC (SEQ ID NO: 159)
SEQ ID NO: 28	CCATCAATGGATCCAAATTGCTCTGTACGCAGAGACTGGCCTGGGTCACCTCTACG GCTGCCATCAATGGATCCAAATTGCTCTGTACGCAGAGAC (SEQ ID NO: 160)
SEQ ID NO: 25	CCTGCAAGGGATCCAAATTGCTCTGTTCGCAGAGACTGGCCTGGGTCACCTCTACG GCTGCCTGCAAGGGATCCAAATTGCTCTGTTCGCAGAGAC (SEQ ID NO: 161)

[0339] 在转染前大约16小时,将100μl在DMEM/10%FBS+青霉素/链霉素中的25,000个HEK293T细胞铺板至96孔板中的每个孔中。在转染的当天,细胞是70%-90%汇合的。对于待转染的每个孔,制备0.5μlLipofectamine 2000和9.5μl Opti-MEM的混合物,并且然后在室温下孵育5-20分钟(溶液1)。在孵育之后,将lipofectamine:OptiMEM混合物添加至最多至10μL的含有182ng效应子质粒和14ng RNA指导物以及水的单独混合物(溶液2)。对于阴性对照,在溶液2中不包含RNA指导物。将溶液1和溶液2通过向上和向下吸移进行混合,且然后在室温下孵育25分钟。在孵育后,将20μL溶液1和溶液2混合物逐滴添加至含有细胞的96孔板的每个孔。在转染后72小时,通过以下方式使细胞胰蛋白酶化:向每个孔的中心添加10μL TrypLE并且孵育大约5分钟。然后将100μL D10培养基添加至每个孔并且混合以重悬细胞。

然后将细胞以500g旋转沉降10分钟，并且弃去上清液。将QuickExtract缓冲液添加至原始细胞悬浮液体积的量的1/5。将细胞在65℃下孵育15分钟，在68℃下孵育15分钟，并且在98℃下孵育10分钟。

[0340] 通过两轮PCR制备用于下一代测序的样品。使用第一轮(PCR1)来扩增根据靶标的特定基因组区域。通过柱纯化对PCR1产物进行纯化。进行第2轮PCR(PCR2)以添加依诺米那衔接子和索引。然后将反应物池化并且通过柱纯化进行纯化。用150次循环NextSeq v2.5中等或高输出试剂盒进行测序运行。

[0341] 图37示出了在用SEQ ID NO:24、SEQ ID NO:28和SEQ ID NO:25的效应子转染后HEK293T细胞中AAVS1靶基因座的插入缺失百分比。对于每种效应子，点反映了在两个生物学平行测定中测量的插入缺失百分比，并且条反映了在两个生物学平行测定中测量的平均插入缺失百分比。对于SEQ ID NO:24、SEQ ID NO:28和SEQ ID NO:25的效应子中的每一种，插入缺失百分比高于由虚线指示的阴性对照的插入缺失百分比。

[0342] 此实例表明CLUST.200916效应子在哺乳动物细胞中具有核酸酶活性。

[0343] 其他实施例

[0344] 应理解，虽然已经结合其具体实施方式对本发明进行了描述，但前面的描述旨在说明而非限制本发明的范围，本发明的范围由所附权利要求书的范围限定。其他方面、优点和修改在以下权利要求的范围内。

序列表

<110> 阿伯生物技术公司（ARBOR BIOTECHNOLOGIES, INC.）
<120> 新型 CRISPR DNA 靶向酶及系统
<130> A2186-7022WO
<140>
<141>
<150> 62/895,406
<151> 2019-09-03
<150> 62/895,422
<151> 2019-09-03
<150> 62/873,108
<151> 2019-07-11
<150> 62/873,118
<151> 2019-07-11
<160> 164
<170> PatentIn 3.5 版
<210> 1
<211> 830
<212> PRT
<213> 未知
<220>
<223> 未知的描述：
废水-工业废水-沉积物序列

[0001]

<400> 1

```
Met Pro Lys Ile Lys Lys Pro Thr Glu Ile Ser Leu Leu Arg Lys Glu
1               5                  10                  15
Val Phe Pro Asp Leu His Phe Ala Lys Asp Arg Met Arg Ala Ala Ser
            20                  25                  30
Leu Val Leu Lys Asn Glu Gly Arg Glu Ala Ala Ile Glu Tyr Leu Arg
        35                  40                  45
Val Asn His Glu Asp Lys Pro Pro Asn Phe Met Pro Pro Ala Lys Thr
    50                  55                  60
Pro Tyr Val Ala Leu Ser Arg Pro Leu Glu Gln Trp Pro Ile Ala Gln
65                  70                  75                  80
Ala Ser Ile Ala Ile Gln Lys Tyr Ile Phe Gly Leu Thr Lys Asp Glu
                85                  90                  95
Phe Ser Ala Thr Lys Lys Leu Leu Tyr Gly Asp Lys Ser Thr Pro Asn
            100                 105                 110
Thr Glu Ser Arg Lys Arg Trp Phe Glu Val Thr Gly Val Pro Asn Phe
        115                 120                 125
Gly Tyr Met Ser Ala Gln Gly Leu Asn Ala Ile Phe Ser Gly Ala Leu
    130                 135                 140
Ala Arg Tyr Glu Gly Val Val Gln Lys Val Glu Asn Arg Asn Lys Lys
145                 150                 155                 160
```

[0002]

```
Arg Phe Glu Lys Leu Ser Glu Lys Asn Gln Leu Leu Ile Glu Glu Gly
                165                 170                 175
Gln Pro Val Lys Asp Tyr Val Pro Asp Thr Ala Tyr His Thr Pro Glu
            180                 185                 190
Thr Leu Gln Lys Leu Ala Glu Asn Asn His Val Arg Val Glu Asp Leu
        195                 200                 205
Gly Asp Met Ile Asp Arg Leu Val His Pro Pro Gly Ile His Arg Ser
    210                 215                 220
Ile Tyr Gly Tyr Gln Gln Val Pro Pro Phe Ala Tyr Asp Pro Asp Asn
225                 230                 235                 240
Pro Lys Gly Ile Ile Leu Pro Lys Ala Tyr Ala Gly Tyr Thr Arg Lys
                245                 250                 255
Pro His Asp Ile Ile Glu Ala Met Pro Asn Arg Leu Asn Ile Pro Glu
            260                 265                 270
Gly Gln Ala Gly Tyr Ile Pro Glu His Gln Arg Asp Lys Leu Lys Lys
        275                 280                 285
Gly Gly Arg Val Lys Arg Leu Arg Thr Thr Arg Val Arg Val Asp Ala
    290                 295                 300
Thr Glu Thr Val Arg Ala Lys Ala Glu Ala Leu Asn Ala Glu Lys Ala
305                 310                 315                 320
Arg Leu Arg Gly Lys Glu Ala Ile Leu Ala Val Phe Gln Ile Glu Glu
                325                 330                 335
Asp Trp Ala Leu Ile Asp Met Arg Gly Leu Leu Arg Asn Val Tyr Met
            340                 345                 350
Arg Lys Leu Ile Ala Ala Gly Glu Leu Thr Pro Thr Thr Leu Leu Gly
        355                 360                 365
Tyr Phe Thr Glu Thr Leu Thr Leu Asp Pro Arg Arg Thr Glu Ala Thr
    370                 375                 380
Phe Cys Tyr His Leu Arg Ser Glu Gly Ala Leu His Ala Glu Tyr Val
385                 390                 395                 400
Arg His Gly Lys Asn Thr Arg Glu Leu Leu Leu Asp Leu Thr Lys Asp
                405                 410                 415
Asn Glu Lys Ile Ala Leu Val Thr Ile Asp Leu Gly Gln Arg Asn Pro
            420                 425                 430
Leu Ala Ala Ala Ile Phe Arg Val Gly Arg Asp Ala Ser Gly Asp Leu
        435                 440                 445
Thr Glu Asn Ser Leu Glu Pro Val Ser Arg Met Leu Leu Pro Gln Ala
    450                 455                 460
Tyr Leu Asp Gln Ile Lys Ala Tyr Arg Asp Ala Tyr Asp Ser Phe Arg
465                 470                 475                 480
Gln Asn Ile Trp Asp Thr Ala Leu Ala Ser Leu Thr Pro Glu Gln Gln
                485                 490                 495
Arg Gln Ile Leu Ala Tyr Glu Ala Tyr Thr Pro Asp Asp Ser Lys Glu
            500                 505                 510
```

[0003]

Asn Val Leu Arg Leu Leu Leu Gly Gly Asn Val Met Pro Asp Asp Leu
515 520 525
Pro Trp Glu Asp Met Thr Lys Asn Thr His Tyr Ile Ser Asp Arg Tyr
530 535 540
Leu Ala Asp Gly Gly Asp Pro Ser Lys Val Trp Phe Val Pro Gly Pro
545 550 555 560
Arg Lys Arg Lys Lys Asn Ala Pro Pro Leu Lys Lys Pro Pro Lys Pro
565 570 575
Arg Glu Leu Val Lys Arg Ser Asp His Asn Ile Ser His Leu Ser Glu
580 585 590
Phe Arg Pro Gln Leu Leu Lys Glu Thr Arg Asp Ala Phe Glu Lys Ala
595 600 605
Lys Ile Asp Thr Glu Arg Gly His Val Gly Tyr Gln Lys Leu Ser Thr
610 615 620
Arg Lys Asp Gln Leu Cys Lys Glu Ile Leu Asn Trp Leu Glu Ala Glu
625 630 635 640
Ala Val Arg Leu Thr Arg Cys Lys Thr Met Val Leu Gly Leu Glu Asp
645 650 655
Leu Asn Gly Pro Phe Phe Asn Gln Gly Lys Gly Lys Val Arg Gly Trp
660 665 670
Val Ser Phe Phe Arg Gln Lys Gln Glu Asn Arg Trp Ile Val Asn Gly
675 680 685
Phe Arg Lys Asn Ala Leu Ala Arg Ala His Asp Lys Gly Lys Tyr Ile
690 695 700
Leu Glu Leu Trp Pro Ser Trp Thr Ser Gln Thr Cys Pro Lys Cys Lys
705 710 715 720
His Val His Ala Asp Asn Arg His Gly Asp Asp Phe Val Cys Leu Gln
725 730 735
Cys Gly Ala Arg Leu His Ala Asp Ala Glu Val Ala Thr Trp Asn Leu
740 745 750
Ala Val Val Ala Ile Gln Gly His Ser Leu Pro Gly Pro Val Arg Glu
755 760 765
Lys Ser Asn Asp Arg Lys Lys Ser Gly Ser Ala Arg Lys Ser Lys Lys
770 775 780
Ala Asn Glu Ser Gly Lys Val Val Gly Ala Trp Ala Ala Gln Ala Thr
785 790 795 800
Pro Lys Arg Ala Thr Ser Lys Lys Glu Thr Gly Thr Ala Arg Asn Pro
805 810 815
Val Tyr Asn Pro Leu Glu Thr Gln Ala Ser Cys Pro Ala Pro
820 825 830

<210> 2
<211> 781
<212> PRT
<213> 未知

<220>

<223> 未知的描述:
水生-淡水-沼泽森林土壤序列

<400> 2

[0004]

```
Met Arg Gln Pro Ala Glu Lys Thr Ala Phe Gln Val Phe Arg Gln Glu
1               5                  10                  15
Val Ile Gly Thr Gln Lys Leu Ser Gly Gly Asp Ala Lys Thr Ala Gly
            20                  25                  30
Arg Leu Tyr Lys Gln Gly Lys Met Glu Ala Ala Arg Glu Trp Leu Leu
        35                  40                  45
Lys Gly Ala Arg Asp Asp Val Pro Pro Asn Phe Gln Pro Pro Ala Lys
    50                  55                  60
Cys Leu Val Val Ala Val Ser His Pro Phe Glu Glu Trp Asp Ile Ser
65                  70                  75                  80
Lys Thr Asn His Asp Val Gln Ala Tyr Ile Tyr Ala Gln Pro Leu Gln
                85                  90                  95
Ala Glu Gly His Leu Asn Gly Leu Ser Glu Lys Trp Glu Asp Thr Ser
            100                 105                 110
Ala Asp Gln His Lys Leu Trp Phe Glu Lys Thr Gly Val Pro Asp Arg
        115                 120                 125
Gly Leu Pro Val Gln Ala Ile Asn Lys Ile Ala Lys Ala Ala Val Asn
    130                 135                 140
Arg Ala Phe Gly Val Val Arg Lys Val Glu Asn Arg Asn Glu Lys Arg
145                 150                 155                 160
Arg Ser Arg Asp Asn Arg Ile Ala Glu His Asn Arg Glu Asn Gly Leu
                165                 170                 175
Thr Glu Val Val Arg Glu Ala Pro Glu Val Ala Thr Asn Ala Asp Gly
            180                 185                 190
Phe Leu Leu His Pro Pro Gly Ile Asp Pro Ser Ile Leu Ser Tyr Ala
        195                 200                 205
Ser Val Ser Pro Val Pro Tyr Asn Ser Ser Lys His Ser Phe Val Arg
    210                 215                 220
Leu Pro Glu Glu Tyr Gln Ala Tyr Asn Val Glu Pro Asp Ala Pro Ile
225                 230                 235                 240
Pro Gln Phe Val Val Glu Asp Arg Phe Ala Ile Pro Pro Gly Gln Pro
                245                 250                 255
Gly Tyr Val Pro Glu Trp Gln Arg Leu Lys Cys Ser Thr Asn Lys His
            260                 265                 270
Arg Arg Met Arg Gln Trp Ser Asn Gln Asp Tyr Lys Pro Lys Ala Gly
        275                 280                 285
Arg Arg Ala Lys Pro Leu Glu Phe Gln Ala His Leu Thr Arg Glu Arg
    290                 295                 300
Ala Lys Gly Ala Leu Leu Val Val Met Arg Ile Lys Glu Asp Trp Val
305                 310                 315                 320
```

Val Phe Asp Val Arg Gly Leu Leu Arg Asn Val Glu Trp Arg Lys Val
325 330 335
Leu Ser Glu Glu Ala Arg Glu Lys Leu Thr Leu Lys Gly Leu Leu Asp
340 345 350
Leu Phe Thr Gly Asp Pro Val Ile Asp Thr Lys Arg Gly Ile Val Thr
355 360 365
Phe Leu Tyr Lys Ala Glu Ile Thr Lys Ile Leu Ser Lys Arg Thr Val
370 375 380
Lys Thr Lys Asn Ala Arg Asp Leu Leu Leu Arg Leu Thr Glu Pro Gly
385 390 395 400
Glu Asp Gly Leu Arg Arg Glu Val Gly Leu Val Ala Val Asp Leu Gly
405 410 415
Gln Thr His Pro Ile Ala Ala Ala Ile Tyr Arg Ile Gly Arg Thr Ser
420 425 430

Ala Gly Ala Leu Glu Ser Thr Val Leu His Arg Gln Gly Leu Arg Glu
435 440 445
Asp Gln Lys Glu Lys Leu Lys Glu Tyr Arg Lys Arg His Thr Ala Leu
450 455 460
Asp Ser Arg Leu Arg Lys Glu Ala Phe Glu Thr Leu Ser Val Glu Gln
465 470 475 480
Gln Lys Glu Ile Val Thr Val Ser Gly Ser Gly Ala Gln Ile Thr Lys
485 490 495
Asp Lys Val Cys Asn Tyr Leu Gly Val Asp Pro Ser Thr Leu Pro Trp
500 505 510
Glu Lys Met Gly Ser Tyr Thr His Phe Ile Ser Asp Asp Phe Leu Arg
515 520 525
Arg Gly Gly Asp Pro Asn Ile Val His Phe Asp Arg Gln Pro Lys Lys
530 535 540
Gly Lys Val Ser Lys Lys Ser Gln Arg Ile Lys Arg Ser Asp Ser Gln
545 550 555 560
Trp Val Gly Arg Met Arg Pro Arg Leu Ser Gln Glu Thr Ala Lys Ala
565 570 575
Arg Met Glu Ala Asp Trp Ala Ala Gln Asn Glu Asn Glu Glu Tyr Lys
580 585 590
Arg Leu Ala Arg Ser Lys Gln Glu Leu Ala Arg Trp Cys Val Asn Thr
595 600 605
Leu Leu Gln Asn Thr Arg Cys Ile Thr Gln Cys Asp Glu Ile Val Val
610 615 620
Val Ile Glu Asp Leu Asn Val Lys Ser Leu His Gly Lys Gly Ala Arg
625 630 635 640
Glu Pro Gly Trp Asp Asn Phe Phe Thr Pro Lys Thr Glu Asn Arg Trp
645 650 655
Phe Ile Gln Ile Leu His Lys Thr Phe Ser Glu Leu Pro Lys His Arg

[0005]

660 665 670
Gly Glu His Val Ile Glu Gly Cys Pro Leu Arg Thr Ser Ile Thr Cys
675 680 685
Pro Ala Cys Ser Tyr Cys Asp Lys Asn Ser Arg Asn Gly Glu Lys Phe
690 695 700
Val Cys Val Ala Cys Gly Ala Thr Phe His Ala Asp Phe Glu Val Ala
705 710 715 720
Thr Tyr Asn Leu Val Arg Leu Ala Thr Thr Gly Met Pro Met Pro Lys
725 730 735
Ser Leu Glu Arg Gln Gly Gly Gly Glu Lys Ala Gly Gly Ala Arg Lys
740 745 750
Ala Arg Lys Lys Ala Lys Gln Val Glu Lys Ile Val Val Gln Ala Asn
755 760 765
Ala Asn Val Thr Met Asn Gly Ala Ser Leu His Ser Pro
770 775 780

<210> 3
<211> 781
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水-沼泽森林土壤序列

[0006]

<400> 3
Met Arg Gln Pro Ala Glu Lys Thr Ala Phe Gln Val Phe Arg Gln Glu
1 5 10 15
Val Ile Gly Thr Gln Lys Leu Ser Gly Gly Asp Ala Lys Thr Ala Gly
20 25 30
Arg Leu Tyr Lys Gln Gly Lys Met Glu Ala Ala Arg Glu Trp Leu Leu
35 40 45
Lys Gly Ala Arg Asp Asp Val Pro Pro Asn Phe Gln Pro Pro Ala Lys
50 55 60
Cys Leu Val Val Ala Val Ser His Pro Phe Glu Glu Trp Asp Ile Ser
65 70 75 80
Lys Thr Asn His Asp Val Gln Ala Tyr Ile Tyr Ala Gln Pro Leu Gln
85 90 95
Ala Glu Gly His Leu Asn Gly Leu Ser Glu Lys Trp Glu Asp Thr Ser
100 105 110
Ala Asp Gln His Lys Leu Trp Phe Glu Lys Thr Gly Val Pro Asp Arg
115 120 125
Gly Leu Pro Val Gln Ala Ile Asn Lys Ile Ala Lys Ala Ala Val Asn
130 135 140
Arg Ala Phe Gly Val Val Arg Lys Val Glu Asn Arg Asn Glu Lys Arg
145 150 155 160
Arg Ser Arg Asp Asn Arg Ile Ala Glu His Asn Arg Glu Asn Gly Leu

[0007]

```
                165                 170                 175
Thr Glu Val Val Arg Glu Ala Pro Glu Val Ala Thr Asn Ala Asp Gly
            180                 185                 190
Phe Leu Leu His Pro Pro Gly Ile Asp Pro Ser Ile Leu Ser Tyr Ala
        195                 200                 205
Ser Val Ser Pro Val Pro Tyr Asn Ser Ser Lys His Ser Phe Val Arg
    210                 215                 220
Leu Pro Glu Glu Tyr Gln Ala Tyr Asn Val Glu Pro Asp Ala Pro Ile
225                 230                 235                 240
Pro Gln Phe Val Val Glu Asp Arg Phe Ala Ile Pro Pro Gly Gln Pro
                245                 250                 255
Gly Tyr Val Pro Glu Trp Gln Arg Leu Lys Cys Ser Thr Asn Lys His
            260                 265                 270
Arg Arg Met Arg Gln Trp Ser Asn Gln Asp Tyr Lys Pro Lys Ala Gly
        275                 280                 285
Arg Arg Ala Lys Pro Leu Glu Phe Gln Ala His Leu Thr Arg Glu Arg
    290                 295                 300
Ala Lys Gly Ala Leu Leu Val Val Met Arg Ile Lys Glu Asp Trp Val
305                 310                 315                 320
Val Phe Asp Val Arg Gly Leu Leu Arg Asn Val Glu Trp Arg Lys Val
                325                 330                 335
Leu Ser Glu Glu Ala Arg Glu Lys Leu Thr Leu Lys Gly Leu Leu Asp
            340                 345                 350
Leu Phe Thr Gly Asp Pro Val Ile Asp Thr Lys Arg Gly Ile Val Thr
        355                 360                 365
Phe Leu Tyr Lys Ala Glu Ile Thr Lys Ile Leu Ser Lys Arg Thr Val
    370                 375                 380
Lys Thr Lys Asn Ala Arg Asp Leu Leu Leu Arg Leu Thr Glu Pro Gly
385                 390                 395                 400
Glu Asp Gly Leu Arg Arg Glu Val Gly Leu Val Ala Val Asp Leu Gly
                405                 410                 415
Gln Thr His Pro Ile Ala Ala Ala Ile Tyr Arg Ile Gly Arg Thr Ser
            420                 425                 430
Ala Gly Ala Leu Glu Ser Thr Val Leu His Arg Gln Gly Leu Arg Glu
        435                 440                 445
Asp Gln Lys Glu Lys Leu Lys Glu Tyr Arg Lys Arg His Thr Ala Leu
    450                 455                 460
Asp Ser Arg Leu Arg Lys Glu Ala Phe Glu Thr Leu Ser Val Glu Gln
465                 470                 475                 480
Gln Lys Glu Ile Val Thr Val Ser Gly Ser Gly Ala Gln Ile Thr Lys
                485                 490                 495
Asp Lys Val Cys Asn Tyr Leu Gly Val Asp Pro Ser Thr Leu Pro Trp
            500                 505                 510
Glu Lys Met Gly Ser Tyr Thr His Phe Ile Ser Asp Asp Phe Leu Arg
```

[0008]

```
          515                 520                 525
Arg Gly Gly Asp Pro Asn Ile Val His Phe Asp Arg Gln Pro Lys Lys
      530                 535                 540
Gly Lys Val Ser Lys Lys Ser Gln Arg Ile Lys Arg Ser Asp Ser Gln
545                 550                 555                 560
Trp Val Gly Arg Met Arg Pro Arg Leu Ser Gln Glu Thr Ala Lys Ala
                  565                 570                 575
Arg Met Glu Ala Asp Trp Ala Ala Gln Asn Glu Asn Glu Glu Tyr Lys
              580                 585                 590
Arg Leu Ala Arg Ser Lys Gln Glu Leu Ala Arg Trp Cys Val Asn Thr
          595                 600                 605
Leu Leu Gln Asn Thr Arg Cys Ile Thr Gln Cys Asp Glu Ile Val Val
      610                 615                 620
Val Ile Glu Asp Leu Asn Val Lys Ser Leu His Gly Lys Gly Ala Arg
625                 630                 635                 640
Glu Pro Gly Trp Asp Asn Phe Phe Thr Pro Lys Thr Glu Asn Arg Trp
                  645                 650                 655
Phe Ile Gln Ile Leu His Lys Thr Phe Ser Glu Leu Pro Lys His Arg
              660                 665                 670
Gly Glu His Val Ile Glu Gly Cys Pro Leu Arg Thr Ser Ile Thr Cys
          675                 680                 685
Pro Ala Cys Ser Tyr Cys Asp Lys Asn Ser Arg Asn Gly Glu Lys Phe
      690                 695                 700
Val Cys Val Ala Cys Gly Ala Thr Phe His Ala Asp Phe Glu Val Ala
705                 710                 715                 720
Thr Tyr Asn Leu Val Arg Leu Ala Thr Thr Gly Met Pro Met Pro Lys
                  725                 730                 735
Ser Leu Glu Arg Gln Gly Gly Gly Glu Lys Ala Gly Gly Ala Arg Lys
              740                 745                 750
Ala Arg Lys Lys Ala Lys Gln Val Glu Lys Ile Val Val Gln Ala Asn
          755                 760                 765
Ala Asn Val Thr Met Asn Gly Ala Ser Leu His Ser Pro
      770                 775                 780
<210> 4
<211> 790
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
      水生-淡水-沉积物序列
<400> 4
Met Thr Pro Ser Pro Gln Ile Ala Arg Leu Val Glu Thr Pro Leu Ala
1                   5                  10                  15
Ala Ala Leu Lys Ala His His Pro Gly Lys Lys Phe Arg Ser Asp Tyr
```

[0009]

```
                20                  25                  30
Leu Lys Lys Ala Gly Lys Ile Leu Lys Asp Gln Gly Val Glu Ala Ala
        35                  40                  45
Met Ala His Leu Asp Gly Lys Asp Gln Ala Glu Pro Pro Asn Phe Lys
    50                  55                  60
Pro Pro Ala Lys Cys Arg Ile Val Ala Arg Ser Arg Glu Phe Ser Glu
65                  70                  75                  80
Trp Pro Ile Val Lys Ala Ser Val Glu Ile Gln Lys Tyr Ile Tyr Gly
                85                  90                  95
Leu Thr Leu Glu Glu Arg Lys Ala Cys Asp Pro Gly Lys Ser Ser Ala
            100                 105                 110
Ser His Lys Ala Trp Phe Ala Lys Thr Gly Val Asn Thr Phe Gly Tyr
        115                 120                 125
Ser Ser Val Gln Gly Phe Asn Leu Ile Phe Gly His Thr Leu Gly Arg
    130                 135                 140
Tyr Asp Gly Val Leu Val Lys Thr Glu Asn Leu Asn Lys Lys Arg Ala
145                 150                 155                 160
Glu Lys Asn Glu Arg Phe Arg Ala Lys Ala Leu Ala Glu Gly Arg Ala
                165                 170                 175
Glu Pro Val Cys Pro Pro Leu Val Thr Ala Thr Asn Asp Thr Gly Gln
            180                 185                 190
Asp Val Thr Leu Glu Asp Gly Arg Val Val Arg Pro Gly Gln Leu Leu
        195                 200                 205
Gln Pro Pro Gly Ile Asn Pro Asn Ile Tyr Ala Tyr Gln Gln Val Ser
    210                 215                 220
Pro Lys Ala Tyr Val Pro Gly Ile Ile Glu Leu Pro Glu Glu Phe Gln
225                 230                 235                 240
Gly Tyr Ser Arg Asp Pro Asn Ala Val Ile Leu Pro Leu Val Pro Arg
                245                 250                 255
Asp Arg Leu Ser Ile Pro Lys Gly Gln Pro Gly Tyr Val Pro Glu Pro
            260                 265                 270
His Arg Glu Gly Leu Thr Gly Arg Lys Asp Arg Arg Met Arg Arg Tyr
        275                 280                 285
Tyr Glu Thr Glu Arg Gly Thr Lys Leu Lys Arg Pro Pro Leu Thr Ala
    290                 295                 300
Lys Gly Arg Ala Asp Lys Ala Asn Glu Ala Leu Leu Val Val Val Arg
305                 310                 315                 320
Ile Asp Ser Asp Trp Val Val Met Asp Val Arg Gly Leu Leu Arg Asn
                325                 330                 335
Ala Arg Trp Arg Arg Leu Val Ser Lys Glu Gly Ile Thr Leu Asn Gly
            340                 345                 350
Leu Leu Asp Leu Phe Thr Gly Asp Pro Val Leu Asn Pro Lys Asp Cys
        355                 360                 365
Ser Val Ser Arg Asp Thr Gly Asp Pro Val Asn Asp Pro Arg His Gly
```

[0010]

```
    370                 375                 380
Val Val Thr Phe Cys Tyr Lys Leu Gly Val Val Asp Val Cys Ser Lys
385                 390                 395                 400
Asp Arg Pro Ile Lys Gly Phe Arg Thr Lys Glu Val Leu Glu Arg Leu
                405                 410                 415
Thr Ser Ser Gly Thr Val Gly Met Val Ser Ile Asp Leu Gly Gln Thr
            420                 425                 430
Asn Pro Val Ala Ala Ala Val Ser Arg Val Thr Lys Gly Leu Gln Ala
        435                 440                 445
Glu Thr Leu Glu Thr Phe Thr Leu Pro Asp Asp Leu Leu Gly Lys Val
    450                 455                 460
Arg Ala Tyr Arg Ala Lys Thr Asp Arg Met Glu Glu Gly Phe Arg Arg
465                 470                 475                 480
Asn Ala Leu Arg Lys Leu Thr Ala Glu Gln Gln Ala Glu Ile Thr Arg
                485                 490                 495
Tyr Asn Asp Ala Thr Glu Gln Gln Ala Lys Ala Leu Val Cys Ser Thr
            500                 505                 510
Tyr Gly Ile Gly Pro Glu Glu Val Pro Trp Glu Arg Met Thr Ser Asn
        515                 520                 525
Thr Thr Tyr Ile Ser Asp His Ile Leu Asp His Gly Gly Asp Pro Asp
    530                 535                 540
Thr Val Phe Phe Met Ala Thr Lys Arg Gly Gln Asn Lys Pro Thr Leu
545                 550                 555                 560
His Lys Arg Lys Asp Lys Ala Trp Gly Gln Lys Phe Arg Pro Ala Ile
                565                 570                 575
Ser Val Glu Thr Arg Leu Ala Arg Gln Ala Ala Glu Trp Glu Leu Arg
            580                 585                 590
Arg Ala Ser Leu Glu Phe Gln Lys Leu Ser Val Trp Lys Thr Glu Leu
        595                 600                 605
Cys Arg Gln Ala Val Asn Tyr Val Met Glu Arg Thr Lys Lys Arg Thr
    610                 615                 620
Gln Cys Asp Val Ile Ile Pro Val Ile Glu Asp Leu Pro Val Pro Leu
625                 630                 635                 640
Phe His Gly Ser Gly Lys Arg Asp Pro Gly Trp Ala Asn Phe Phe Val
                645                 650                 655
His Lys Arg Glu Asn Arg Trp Phe Ile Asp Gly Leu His Lys Ala Phe
            660                 665                 670
Ser Glu Leu Gly Lys His Arg Gly Ile Tyr Val Phe Glu Val Cys Pro
        675                 680                 685
Gln Arg Thr Ser Ile Thr Cys Pro Lys Cys Gly His Cys Asp Pro Asp
    690                 695                 700
Asn Arg Asp Gly Glu Lys Phe Val Cys Leu Ser Cys Gln Ala Thr Leu
705                 710                 715                 720
His Ala Asp Leu Asp Val Ala Thr Thr Asn Leu Val Arg Val Ala Leu
```

725 730 735
Thr Gly Lys Val Met Pro Arg Ser Glu Arg Ser Gly Asp Ala Gln Thr
740 745 750
Pro Gly Pro Ala Arg Lys Ala Arg Thr Gly Lys Ile Lys Gly Ser Lys
755 760 765
Pro Thr Ser Ala Pro Gln Gly Ala Thr Gln Thr Asp Ala Lys Ala His
770 775 780
Leu Ser Gln Thr Gly Val
785 790

<210> 5
<211> 790
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水-沉积物序列
<400> 5

[0011]

Met Thr Pro Ser Pro Gln Ile Ala Arg Leu Val Glu Thr Pro Leu Ala
1 5 10 15
Ala Ala Leu Lys Ala His His Pro Gly Lys Lys Phe Arg Ser Asp Tyr
20 25 30
Leu Lys Lys Ala Gly Lys Ile Leu Lys Asp Gln Gly Val Glu Ala Ala
35 40 45
Met Ala His Leu Asp Gly Lys Asp Gln Ala Glu Pro Pro Asn Phe Lys
50 55 60
Pro Pro Ala Lys Cys Arg Ile Val Ala Arg Ser Arg Glu Phe Ser Glu
65 70 75 80
Trp Pro Ile Val Lys Ala Ser Val Glu Ile Gln Lys Tyr Ile Tyr Gly
85 90 95
Leu Thr Leu Glu Glu Arg Lys Ala Cys Asp Pro Gly Lys Ser Ser Ala
100 105 110
Ser His Lys Ala Trp Phe Ala Lys Thr Gly Val Asn Thr Phe Gly Tyr
115 120 125
Ser Ser Val Gln Gly Phe Asn Leu Ile Phe Gly His Thr Leu Gly Arg
130 135 140
Tyr Asp Gly Val Leu Val Lys Thr Glu Asn Leu Asn Lys Lys Arg Ala
145 150 155 160
Glu Lys Asn Glu Arg Phe Arg Ala Lys Ala Leu Ala Glu Gly Arg Ala
165 170 175
Glu Pro Val Cys Pro Pro Leu Val Thr Ala Thr Asn Asp Thr Gly Gln
180 185 190
Asp Val Thr Leu Glu Asp Gly Arg Val Val Arg Pro Gly Gln Leu Leu
195 200 205
Gln Pro Pro Gly Ile Asn Pro Asn Ile Tyr Ala Tyr Gln Gln Val Ser

[0012]

```
    210                 215                 220
Pro Lys Ala Tyr Val Pro Gly Ile Ile Glu Leu Pro Glu Glu Phe Gln
225                 230                 235                 240
Gly Tyr Ser Arg Asp Pro Asn Ala Val Ile Leu Pro Leu Val Pro Arg
                245                 250                 255
Asp Arg Leu Ser Ile Pro Lys Gly Gln Pro Gly Tyr Val Pro Glu Pro
            260                 265                 270
His Arg Glu Gly Leu Thr Gly Arg Lys Asp Arg Arg Met Arg Arg Tyr
        275                 280                 285
Tyr Glu Thr Glu Arg Gly Thr Lys Leu Lys Arg Pro Pro Leu Thr Ala
    290                 295                 300

Lys Gly Arg Ala Asp Lys Ala Asn Glu Ala Leu Leu Val Val Val Arg
305                 310                 315                 320
Ile Asp Ser Asp Trp Val Val Met Asp Val Arg Gly Leu Leu Arg Asn
                325                 330                 335
Ala Arg Trp Arg Arg Leu Val Ser Lys Glu Gly Ile Thr Leu Asn Gly
            340                 345                 350
Leu Leu Asp Leu Phe Thr Gly Asp Pro Val Leu Asn Pro Lys Asp Cys
        355                 360                 365
Ser Val Ser Arg Asp Thr Gly Asp Pro Val Asn Asp Pro Arg His Gly
    370                 375                 380
Val Val Thr Phe Cys Tyr Lys Leu Gly Val Val Asp Val Cys Ser Lys
385                 390                 395                 400
Asp Arg Pro Ile Lys Gly Phe Arg Thr Lys Glu Val Leu Glu Arg Leu
                405                 410                 415
Thr Ser Ser Gly Thr Val Gly Met Val Ser Ile Asp Leu Gly Gln Thr
            420                 425                 430
Asn Pro Val Ala Ala Ala Val Ser Arg Val Thr Lys Gly Leu Gln Ala
        435                 440                 445
Glu Thr Leu Glu Thr Phe Thr Leu Pro Asp Asp Leu Leu Gly Lys Val
    450                 455                 460
Arg Ala Tyr Arg Ala Lys Thr Asp Arg Met Glu Glu Gly Phe Arg Arg
465                 470                 475                 480
Asn Ala Leu Arg Lys Leu Thr Ala Glu Gln Gln Ala Glu Ile Thr Arg
                485                 490                 495
Tyr Asn Asp Ala Thr Glu Gln Gln Ala Lys Ala Leu Val Cys Ser Thr
            500                 505                 510
Tyr Gly Ile Gly Pro Glu Glu Val Pro Trp Glu Arg Met Thr Ser Asn
        515                 520                 525
Thr Thr Tyr Ile Ser Asp His Ile Leu Asp His Gly Gly Asp Pro Asp
    530                 535                 540
Thr Val Phe Phe Met Ala Thr Lys Arg Gly Gln Asn Lys Pro Thr Leu
545                 550                 555                 560
```

[0013]

His Lys Arg Lys Asp Lys Ala Trp Gly Gln Lys Phe Arg Pro Ala Ile
565 570 575
Ser Val Glu Thr Arg Leu Ala Arg Gln Ala Ala Glu Trp Glu Leu Arg
580 585 590
Arg Ala Ser Leu Glu Phe Gln Lys Leu Ser Val Trp Lys Thr Glu Leu
595 600 605
Cys Arg Gln Ala Val Asn Tyr Val Met Glu Arg Thr Lys Lys Arg Thr
610 615 620
Gln Cys Asp Val Ile Ile Pro Val Ile Glu Asp Leu Pro Val Pro Leu
625 630 635 640
Phe His Gly Ser Gly Lys Arg Asp Pro Gly Trp Ala Asn Phe Phe Val
645 650 655
His Lys Arg Glu Asn Arg Trp Phe Ile Asp Gly Leu His Lys Ala Phe
660 665 670
Ser Glu Leu Gly Lys His Arg Gly Ile Tyr Val Phe Glu Val Cys Pro
675 680 685
Gln Arg Thr Ser Ile Thr Cys Pro Lys Cys Gly His Cys Asp Pro Asp
690 695 700
Asn Arg Asp Gly Glu Lys Phe Val Cys Leu Ser Cys Gln Ala Thr Leu
705 710 715 720
Asn Ala Asp Leu Asp Val Ala Thr Thr Asn Leu Val Arg Val Ala Leu
725 730 735
Thr Gly Lys Val Met Pro Arg Ser Glu Arg Ser Gly Asp Ala Gln Thr
740 745 750
Pro Gly Pro Ala Arg Lys Ala Arg Thr Gly Lys Ile Lys Gly Ser Lys
755 760 765
Pro Thr Ser Ala Pro Gln Gly Ala Thr Gln Thr Asp Ala Lys Ala His
770 775 780
Leu Ser Gln Thr Gly Val
785 790

<210> 6
<211> 790
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水-沉积物序列
<400> 6

Met Thr Pro Ser Pro Gln Ile Ala Arg Leu Val Glu Thr Pro Leu Ala
1 5 10 15
Ala Ala Leu Lys Ala His His Pro Gly Lys Lys Phe Arg Ser Asp Tyr
20 25 30
Leu Lys Lys Ala Gly Lys Ile Leu Lys Asp Gln Gly Val Glu Ala Ala
35 40 45

[0014]

```
Met Ala His Leu Asp Gly Lys Asp Gln Ala Glu Pro Pro Asn Phe Lys
    50                  55                  60
Pro Pro Ala Lys Cys Arg Ile Val Ala Arg Ser Arg Glu Phe Ser Glu
65                  70                  75                  80
Trp Pro Ile Val Lys Ala Ser Val Glu Ile Gln Lys Tyr Ile Tyr Gly
                85                  90                  95
Leu Thr Leu Glu Glu Arg Lys Ala Cys Asp Pro Gly Lys Ser Ser Ala
            100                 105                 110
Ser His Lys Ala Trp Phe Ala Lys Thr Gly Val Asn Thr Phe Gly Tyr
        115                 120                 125
Ser Ser Val Gln Gly Phe Asn Leu Ile Phe Gly His Thr Leu Gly Arg
    130                 135                 140
Tyr Asp Gly Val Leu Val Lys Thr Glu Asn Leu Asn Lys Lys Arg Ala
145                 150                 155                 160
Glu Lys Asn Glu Arg Phe Arg Ala Lys Ala Leu Ala Glu Gly Arg Ala
                165                 170                 175
Glu Pro Val Cys Pro Pro Leu Val Thr Ala Thr Asn Asp Thr Gly Gln
            180                 185                 190
Asp Val Thr Leu Glu Asp Gly Arg Val Val Arg Pro Gly Gln Leu Leu
        195                 200                 205
Gln Pro Pro Gly Ile Asn Pro Asn Ile Tyr Ala Tyr Gln Gln Val Ser
    210                 215                 220
Pro Lys Ala Tyr Val Pro Gly Ile Ile Glu Leu Pro Glu Glu Phe Gln
225                 230                 235                 240
Gly Tyr Ser Arg Asp Pro Asn Ala Val Ile Leu Pro Leu Val Pro Arg
                245                 250                 255
Asp Arg Leu Ser Ile Pro Lys Gly Gln Pro Gly Tyr Val Pro Glu Pro
            260                 265                 270
His Arg Glu Gly Leu Thr Gly Arg Lys Asp Arg Arg Met Arg Arg Tyr
        275                 280                 285
Tyr Glu Thr Glu Arg Gly Thr Lys Leu Lys Arg Pro Pro Leu Thr Ala
    290                 295                 300
Lys Gly Arg Ala Asp Lys Ala Asn Glu Ala Leu Leu Val Val Val Arg
305                 310                 315                 320
Ile Asp Ser Asp Trp Val Val Met Asp Val Arg Gly Leu Leu Arg Asn
                325                 330                 335
Ala Arg Trp Arg Arg Leu Val Ser Lys Glu Gly Ile Thr Leu Asn Gly
            340                 345                 350
Leu Leu Asp Leu Phe Thr Gly Asp Pro Val Leu Asn Pro Lys Asp Cys
        355                 360                 365
Ser Val Ser Arg Asp Thr Gly Asp Pro Val Asn Asp Pro Arg His Gly
    370                 375                 380
Val Val Thr Phe Cys Tyr Lys Leu Gly Val Val Asp Val Cys Ser Lys
385                 390                 395                 400
```

[0015]

Asp Arg Pro Ile Lys Gly Phe Arg Thr Lys Glu Val Leu Glu Arg Leu
405 410 415
Thr Ser Ser Gly Thr Val Gly Met Val Ser Ile Asp Leu Gly Gln Thr
420 425 430
Asn Pro Val Ala Ala Ala Val Ser Arg Val Thr Lys Gly Leu Gln Ala
435 440 445
Glu Thr Leu Glu Thr Phe Thr Leu Pro Asp Asp Leu Leu Gly Lys Val
450 455 460
Arg Ala Tyr Arg Ala Lys Thr Asp Arg Met Glu Glu Gly Phe Arg Arg
465 470 475 480
Asn Ala Leu Arg Lys Leu Thr Ala Glu Gln Gln Ala Glu Ile Thr Arg
485 490 495
Tyr Asn Asp Ala Thr Glu Gln Gln Ala Lys Ala Leu Val Cys Ser Thr
500 505 510
Tyr Gly Ile Gly Pro Glu Glu Val Pro Trp Glu Arg Met Thr Ser Asn
515 520 525
Thr Thr Tyr Ile Ser Asp His Ile Leu Asp His Gly Gly Asp Pro Asp
530 535 540
Thr Val Phe Phe Met Ala Thr Lys Arg Gly Gln Asn Lys Pro Thr Leu
545 550 555 560
His Lys Arg Lys Asp Lys Ala Trp Gly Gln Lys Phe Arg Pro Ala Ile
565 570 575
Ser Val Glu Thr Arg Leu Ala Arg Gln Ala Ala Glu Trp Glu Leu Arg
580 585 590
Arg Ala Ser Leu Glu Phe Gln Lys Leu Ser Val Trp Lys Thr Glu Leu
595 600 605
Cys Arg Gln Ala Val Asn Tyr Val Met Glu Arg Thr Lys Lys Arg Thr
610 615 620
Gln Cys Asp Val Ile Ile Pro Val Ile Glu Asp Leu Pro Val Pro Leu
625 630 635 640
Phe His Gly Ser Gly Lys Arg Asp Pro Gly Trp Ala Asn Phe Phe Val
645 650 655
His Lys Arg Glu Asn Arg Trp Phe Ile Asp Gly Leu His Lys Ala Phe
660 665 670
Ser Glu Leu Gly Lys His Arg Gly Ile Tyr Val Phe Glu Val Cys Pro
675 680 685
Gln Arg Thr Ser Ile Thr Cys Pro Lys Cys Gly His Cys Asp Pro Asp
690 695 700
Asn Arg Asp Gly Glu Lys Phe Val Cys Leu Ser Cys Gln Ala Thr Leu
705 710 715 720
Asn Ala Asp Leu Asp Val Ala Thr Thr Asn Leu Val Arg Val Ala Leu
725 730 735
Thr Gly Lys Val Met Pro Arg Ser Glu Arg Ser Gly Asp Ala Gln Thr
740 745 750

Pro Gly Pro Ala Arg Lys Ala Arg Thr Gly Lys Ile Lys Gly Ser Lys
755 760 765
Pro Thr Ser Ala Pro Gln Gly Ala Thr Gln Thr Asp Ala Lys Ala His
770 775 780
Leu Ser Gln Thr Gly Val
785 790

<210> 7
<211> 782
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水-沉积物序列
<400> 7

[0016]

Met Lys Thr Glu Lys Pro Lys Thr Ala Leu Thr Leu Leu Arg Glu Glu
1 5 10 15
Val Phe Pro Gly Lys Lys Tyr Arg Leu Asp Val Leu Lys Glu Ala Gly
20 25 30
Lys Lys Leu Ser Thr Lys Gly Arg Glu Ala Thr Ile Glu Phe Leu Thr
35 40 45
Gly Lys Asp Glu Glu Arg Pro Gln Asn Phe Gln Pro Pro Ala Lys Thr
50 55 60
Ser Ile Val Ala Gln Ser Arg Pro Phe Asp Gln Trp Pro Ile Val Gln
65 70 75 80
Val Ser Leu Ala Val Gln Lys Tyr Ile Tyr Gly Leu Thr Gln Ser Glu
85 90 95
Phe Glu Ala Asn Lys Lys Ala Leu Tyr Gly Glu Thr Gly Lys Ala Ile
100 105 110
Ser Thr Glu Ser Arg Arg Ala Trp Phe Glu Ala Thr Gly Val Asp Asn
115 120 125
Phe Gly Phe Thr Ala Ala Gln Gly Ile Asn Pro Ile Phe Ser Gln Ala
130 135 140
Val Ala Arg Tyr Glu Gly Val Ile Lys Lys Val Glu Asn Arg Asn Glu
145 150 155 160
Lys Lys Leu Lys Lys Leu Thr Lys Lys Asn Leu Leu Arg Leu Glu Ser
165 170 175
Gly Glu Glu Ile Glu Asp Phe Glu Pro Glu Ala Thr Phe Asn Glu Glu
180 185 190
Gly Arg Leu Leu Gln Pro Pro Gly Ala Asn Pro Asn Ile Tyr Cys Tyr
195 200 205
Gln Gln Ile Ser Pro Arg Ile Tyr Asp Pro Ser Asp Pro Lys Gly Val
210 215 220
Ile Leu Pro Gln Ile Tyr Ala Gly Tyr Asp Arg Lys Pro Glu Asp Ile
225 230 235 240

[0017]

```
Ile Ser Ala Gly Val Pro Asn Arg Leu Ala Ile Pro Glu Gly Gln Pro
                245                 250                 255
Gly Tyr Ile Pro Glu His Gln Arg Ala Gly Leu Lys Thr Gln Gly Arg
            260                 265                 270
Ile Arg Cys Arg Ala Ser Val Glu Ala Lys Ala Arg Ala Ala Ile Leu
        275                 280                 285
Ala Val Val His Leu Gly Glu Asp Trp Val Val Leu Asp Leu Arg Gly
    290                 295                 300
Leu Leu Arg Asn Val Tyr Trp Arg Lys Leu Ala Ser Pro Gly Thr Leu
305                 310                 315                 320
Thr Leu Lys Gly Leu Leu Asp Phe Phe Thr Gly Gly Pro Val Leu Asp
                325                 330                 335
Ala Arg Arg Gly Ile Ala Thr Phe Ser Tyr Thr Leu Lys Ser Ala Ala
            340                 345                 350
Ala Val His Ala Glu Asn Thr Tyr Lys Gly Lys Gly Thr Arg Glu Val
        355                 360                 365
Leu Leu Lys Leu Thr Glu Asn Asn Ser Val Ala Leu Val Thr Val Asp
    370                 375                 380
Leu Gly Gln Arg Asn Pro Leu Ala Ala Met Ile Ala Arg Val Ser Arg
385                 390                 395                 400
Thr Ser Gln Gly Asp Leu Thr Tyr Pro Glu Ser Val Glu Pro Leu Thr
                405                 410                 415
Arg Leu Phe Leu Pro Asp Pro Phe Leu Glu Glu Val Arg Lys Tyr Arg
            420                 425                 430
Ser Ser Tyr Asp Ala Leu Arg Leu Ser Ile Arg Glu Ala Ala Ile Ala
        435                 440                 445
Ser Leu Thr Pro Glu Gln Gln Ala Glu Ile Arg Tyr Ile Glu Lys Phe
    450                 455                 460
Ser Ala Gly Asp Ala Lys Lys Asn Val Ala Glu Val Phe Gly Ile Asp
465                 470                 475                 480
Pro Thr Gln Leu Pro Trp Asp Ala Met Thr Pro Arg Thr Thr Tyr Ile
                485                 490                 495
Ser Asp Leu Phe Leu Arg Met Gly Gly Asp Arg Ser Arg Val Phe Phe
            500                 505                 510
Glu Val Pro Pro Lys Lys Ala Lys Lys Ala Pro Lys Lys Pro Pro Lys
        515                 520                 525
Lys Pro Ala Gly Pro Arg Ile Val Lys Arg Thr Asp Gly Met Ile Ala
    530                 535                 540
Arg Leu Arg Glu Ile Arg Pro Arg Leu Ser Ala Glu Thr Asn Lys Ala
545                 550                 555                 560
Phe Gln Glu Ala Arg Trp Glu Gly Glu Arg Ser Asn Val Ala Phe Gln
                565                 570                 575
Lys Leu Ser Val Arg Arg Lys Gln Phe Ala Arg Thr Val Val Asn His
            580                 585                 590
```

Leu Val Gln Thr Ala Gln Lys Met Ser Arg Cys Asp Thr Val Val Leu
595 600 605
Gly Ile Glu Asp Leu Asn Val Pro Phe Phe His Gly Arg Gly Lys Tyr
610 615 620
Gln Pro Gly Trp Glu Gly Phe Phe Arg Gln Lys Lys Glu Asn Arg Trp
625 630 635 640
Leu Ile Asn Asp Met His Lys Ala Leu Ser Glu Arg Gly Pro His Arg
645 650 655
Gly Gly Tyr Val Leu Glu Leu Thr Pro Phe Trp Thr Ser Leu Arg Cys
660 665 670
Pro Lys Cys Gly His Thr Asp Ser Ala Asn Arg Asp Gly Asp Asp Phe
675 680 685
Val Cys Val Lys Cys Gly Ala Lys Leu His Ser Asp Leu Glu Val Ala
690 695 700
Thr Ala Asn Leu Ala Leu Val Ala Ile Thr Gly Gln Ser Ile Pro Arg
705 710 715 720
Pro Pro Arg Glu Gln Ser Ser Gly Lys Lys Ser Thr Gly Thr Ala Arg
725 730 735
Met Lys Lys Thr Ser Gly Glu Thr Gln Gly Lys Gly Ser Lys Ala Cys
740 745 750
Val Ser Glu Ala Leu Asn Lys Ile Glu Gln Gly Thr Ala Arg Asp Pro
755 760 765
[0018] Val Tyr Asn Pro Leu Asn Ser Gln Val Ser Cys Pro Ala Pro
770 775 780

<210> 8
<211> 774
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-海生-虫穴序列
<400> 8
Met Tyr Asn Pro Asp Met Lys Lys Pro Asn Asn Ile Arg Arg Ile Arg
1 5 10 15
Glu Glu His Phe Glu Gly Leu Cys Phe Gly Lys Asp Val Leu Thr Lys
20 25 30
Ala Gly Lys Ile Tyr Glu Lys Asp Gly Glu Glu Ala Ala Ile Asp Phe
35 40 45
Leu Met Gly Lys Asp Glu Glu Asp Pro Pro Asn Phe Lys Pro Pro Ala
50 55 60
Lys Thr Thr Ile Val Ala Gln Ser Arg Pro Phe Asp Gln Trp Pro Ile
65 70 75 80
Tyr Gln Val Ser Gln Ala Val Gln Glu Arg Val Phe Ala Tyr Thr Glu
85 90 95

[0019]

```
Glu Glu Phe Asn Ala Ser Lys Glu Ala Leu Phe Ser Gly Asp Ile Ser
            100                 105                 110
Ser Lys Ser Arg Asp Phe Trp Phe Lys Thr Asn Asn Ile Ser Asp Gln
        115                 120                 125
Gly Ile Gly Ala Gln Gly Leu Asn Thr Ile Leu Ser His Ala Phe Ser
    130                 135                 140
Arg Tyr Ser Gly Val Ile Lys Lys Val Glu Asn Arg Asn Lys Lys Arg
145                 150                 155                 160
Leu Lys Lys Leu Ser Lys Lys Asn Gln Leu Lys Ile Glu Glu Gly Leu
                165                 170                 175
Glu Ile Leu Glu Phe Lys Pro Asp Ser Ala Phe Asn Glu Asn Gly Leu
            180                 185                 190
Leu Ala Gln Pro Pro Gly Ile Asn Pro Asn Ile Tyr Gly Tyr Gln Ala
        195                 200                 205
Val Thr Pro Phe Val Phe Asp Pro Asp Asn Pro Gly Asp Val Ile Leu
    210                 215                 220
Pro Lys Gln Tyr Glu Gly Tyr Ser Arg Lys Pro Asp Asp Ile Ile Glu
225                 230                 235                 240
Lys Gly Pro Ser Arg Leu Asp Ile Pro Lys Gly Gln Pro Gly Tyr Val
                245                 250                 255
Pro Glu His Gln Arg Lys Asn Leu Lys Lys Lys Gly Arg Val Arg Leu
            260                 265                 270
Tyr Arg Arg Thr Pro Pro Lys Thr Lys Ala Leu Ala Ser Ile Leu Ala
        275                 280                 285
Val Leu Gln Ile Gly Lys Asp Trp Val Leu Phe Asp Met Arg Gly Leu
    290                 295                 300
Leu Arg Ser Val Tyr Met Arg Glu Ala Ala Thr Pro Gly Gln Ile Ser
305                 310                 315                 320
Ala Lys Asp Leu Leu Asp Thr Phe Thr Gly Cys Pro Val Leu Asn Thr
                325                 330                 335
Arg Thr Gly Glu Phe Thr Phe Cys Tyr Lys Leu Arg Ser Glu Gly Ala
            340                 345                 350
Leu His Ala Arg Lys Ile Tyr Thr Lys Gly Glu Thr Arg Thr Leu Leu
        355                 360                 365
Thr Ser Leu Thr Ser Glu Asn Asn Thr Ile Ala Leu Val Thr Val Asp
    370                 375                 380
Leu Gly Gln Arg Asn Pro Ala Ala Ile Met Ile Ser Arg Leu Ser Arg
385                 390                 395                 400
Lys Glu Glu Leu Ser Glu Lys Asp Ile Gln Pro Val Ser Arg Arg Leu
                405                 410                 415
Leu Pro Asp Arg Tyr Leu Asn Glu Leu Lys Arg Tyr Arg Asp Ala Tyr
            420                 425                 430
Asp Ala Phe Arg Gln Glu Val Arg Asp Glu Ala Phe Thr Ser Leu Cys
        435                 440                 445
```

[0020]

Pro Glu His Gln Glu Gln Val Gln Gln Tyr Glu Ala Leu Thr Pro Glu
450 455 460
Lys Ala Lys Asn Leu Val Leu Lys His Phe Phe Gly Thr His Asp Pro
465 470 475 480
Asp Leu Pro Trp Asp Asp Met Thr Ser Asn Thr His Tyr Ile Ala Asn
485 490 495
Leu Tyr Leu Glu Arg Gly Gly Asp Pro Ser Lys Val Phe Phe Thr Arg
500 505 510
Pro Leu Lys Lys Asp Ser Lys Ser Lys Lys Pro Arg Lys Pro Thr Lys
515 520 525
Arg Thr Asp Ala Ser Ile Ser Arg Leu Pro Glu Ile Arg Pro Lys Met
530 535 540
Pro Glu Asp Ala Arg Lys Ala Phe Glu Lys Ala Lys Trp Glu Ile Tyr
545 550 555 560
Thr Gly His Glu Lys Phe Pro Lys Leu Ala Lys Arg Val Asn Gln Leu
565 570 575
Cys Arg Glu Ile Ala Asn Trp Ile Glu Lys Glu Ala Lys Arg Leu Thr
580 585 590
Leu Cys Asp Thr Val Val Val Gly Ile Glu Asp Leu Ser Leu Pro Pro
595 600 605
Lys Arg Gly Lys Gly Lys Phe Gln Glu Thr Trp Gln Gly Phe Phe Arg
610 615 620
Gln Lys Phe Glu Asn Arg Trp Val Ile Asp Thr Leu Lys Lys Ala Ile
625 630 635 640
Gln Asn Arg Ala His Asp Lys Gly Lys Tyr Val Leu Gly Leu Ala Pro
645 650 655
Tyr Trp Thr Ser Gln Arg Cys Pro Ala Cys Gly Phe Ile His Lys Ser
660 665 670
Asn Arg Asn Gly Asp His Phe Lys Cys Leu Lys Cys Glu Ala Leu Phe
675 680 685
His Ala Asp Ser Glu Val Ala Thr Trp Asn Leu Ala Leu Val Ala Val
690 695 700
Leu Gly Lys Gly Ile Thr Asn Pro Asp Ser Lys Lys Pro Ser Gly Gln
705 710 715 720
Lys Lys Thr Gly Thr Thr Arg Lys Lys Gln Ile Lys Gly Lys Asn Lys
725 730 735
Gly Lys Glu Thr Val Asn Val Pro Pro Thr Thr Gln Glu Val Glu Asp
740 745 750
Ile Ile Ala Phe Phe Glu Lys Asp Asp Glu Thr Val Arg Asn Pro Val
755 760 765
Tyr Lys Pro Thr Gly Thr
770

<210> 9

<211> 769

<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-海生-虫穴序列
<400> 9

[0021]

```
Met Lys Lys Pro Asn Asn Ile Arg Arg Ile Arg Glu Glu His Phe Glu
1               5                  10                  15
Gly Leu Cys Phe Gly Lys Asp Val Leu Thr Lys Ala Gly Lys Ile Tyr
            20                  25                  30
Glu Lys Asp Gly Glu Glu Ala Ala Ile Asp Phe Leu Met Gly Lys Asp
        35                  40                  45
Glu Glu Asp Pro Pro Asn Phe Lys Pro Pro Ala Lys Thr Thr Ile Val
    50                  55                  60
Ala Gln Ser Arg Pro Phe Asp Gln Trp Pro Ile Tyr Gln Val Ser Gln
65                  70                  75                  80
Ala Val Gln Glu Arg Val Phe Ala Tyr Thr Glu Glu Glu Phe Asn Ala
                85                  90                  95
Ser Lys Glu Ala Leu Phe Ser Gly Asp Ile Ser Ser Lys Ser Arg Asp
            100                 105                 110
Phe Trp Phe Lys Thr Asn Asn Ile Ser Asp Gln Gly Ile Gly Ala Gln
        115                 120                 125
Gly Leu Asn Thr Ile Leu Ser His Ala Phe Ser Arg Tyr Ser Gly Val
    130                 135                 140
Ile Lys Lys Val Glu Asn Arg Asn Lys Lys Arg Leu Lys Lys Leu Ser
145                 150                 155                 160
Lys Lys Asn Gln Leu Lys Ile Glu Glu Gly Leu Glu Ile Leu Glu Phe
                165                 170                 175
Lys Pro Asp Ser Ala Phe Asn Glu Asn Gly Leu Leu Ala Gln Pro Pro
            180                 185                 190
Gly Ile Asn Pro Asn Ile Tyr Gly Tyr Gln Ala Val Thr Pro Phe Val
        195                 200                 205
Phe Asp Pro Asp Asn Pro Gly Asp Val Ile Leu Pro Lys Gln Tyr Glu
    210                 215                 220
Gly Tyr Ser Arg Lys Pro Asp Asp Ile Ile Glu Lys Gly Pro Ser Arg
225                 230                 235                 240
Leu Asp Ile Pro Lys Gly Gln Pro Gly Tyr Val Pro Glu His Gln Arg
                245                 250                 255
Lys Asn Leu Lys Lys Lys Gly Arg Val Arg Leu Tyr Arg Arg Thr Pro
            260                 265                 270
Pro Lys Thr Lys Ala Leu Ala Ser Ile Leu Ala Val Leu Gln Ile Gly
        275                 280                 285
Lys Asp Trp Val Leu Phe Asp Met Arg Gly Leu Leu Arg Ser Val Tyr
    290                 295                 300
```

[0022]

```
Met Arg Glu Ala Ala Thr Pro Gly Gln Ile Ser Ala Lys Asp Leu Leu
305                 310                 315                 320
Asp Thr Phe Thr Gly Cys Pro Val Leu Asn Thr Arg Thr Gly Glu Phe
                325                 330                 335
Thr Phe Cys Tyr Lys Leu Arg Ser Glu Gly Ala Leu His Ala Arg Lys
            340                 345                 350
Ile Tyr Thr Lys Gly Glu Thr Arg Thr Leu Leu Thr Ser Leu Thr Ser
        355                 360                 365
Glu Asn Asn Thr Ile Ala Leu Val Thr Val Asp Leu Gly Gln Arg Asn
    370                 375                 380
Pro Ala Ala Ile Met Ile Ser Arg Leu Ser Arg Lys Glu Glu Leu Ser
385                 390                 395                 400
Glu Lys Asp Ile Gln Pro Val Ser Arg Arg Leu Leu Pro Asp Arg Tyr
                405                 410                 415
Leu Asn Glu Leu Lys Arg Tyr Arg Asp Ala Tyr Asp Ala Phe Arg Gln
            420                 425                 430
Glu Val Arg Asp Glu Ala Phe Thr Ser Leu Cys Pro Glu His Gln Glu
        435                 440                 445
Gln Val Gln Gln Tyr Glu Ala Leu Thr Pro Glu Lys Ala Lys Asn Leu
    450                 455                 460
Val Leu Lys His Phe Phe Gly Thr His Asp Pro Asp Leu Pro Trp Asp
465                 470                 475                 480
Asp Met Thr Ser Asn Thr His Tyr Ile Ala Asn Leu Tyr Leu Glu Arg
                485                 490                 495
Gly Gly Asp Pro Ser Lys Val Phe Phe Thr Arg Pro Leu Lys Lys Asp
            500                 505                 510
Ser Lys Ser Lys Lys Pro Arg Lys Pro Thr Lys Arg Thr Asp Ala Ser
        515                 520                 525
Ile Ser Arg Leu Pro Glu Ile Arg Pro Lys Met Pro Glu Asp Ala Arg
    530                 535                 540
Lys Ala Phe Glu Lys Ala Lys Trp Glu Ile Tyr Thr Gly His Glu Lys
545                 550                 555                 560
Phe Pro Lys Leu Ala Lys Arg Val Asn Gln Leu Cys Arg Glu Ile Ala
                565                 570                 575
Asn Trp Ile Glu Lys Glu Ala Lys Arg Leu Thr Leu Cys Asp Thr Val
            580                 585                 590
Val Val Gly Ile Glu Asp Leu Ser Leu Pro Pro Lys Arg Gly Lys Gly
        595                 600                 605
Lys Phe Gln Glu Thr Trp Gln Gly Phe Phe Arg Gln Lys Phe Glu Asn
    610                 615                 620
Arg Trp Val Ile Asp Thr Leu Lys Lys Ala Ile Gln Asn Arg Ala His
625                 630                 635                 640
Asp Lys Gly Lys Tyr Val Leu Gly Leu Ala Pro Tyr Trp Thr Ser Gln
                645                 650                 655
```

Arg Cys Pro Ala Cys Gly Phe Ile His Lys Ser Asn Arg Asn Gly Asp
660 665 670
His Phe Lys Cys Leu Lys Cys Glu Ala Leu Phe His Ala Asp Ser Glu
675 680 685
Val Ala Thr Trp Asn Leu Ala Leu Val Ala Val Leu Gly Lys Gly Ile
690 695 700
Thr Asn Pro Asp Ser Lys Lys Pro Ser Gly Gln Lys Lys Thr Gly Thr
705 710 715 720
Thr Arg Lys Lys Gln Ile Lys Gly Lys Asn Lys Gly Lys Glu Thr Val
725 730 735
Asn Val Pro Pro Thr Thr Gln Glu Val Glu Asp Ile Ile Ala Phe Phe
740 745 750
Glu Lys Asp Asp Glu Thr Val Arg Asn Pro Val Tyr Lys Pro Thr Gly
755 760 765
Thr

<210> 10
<211> 769
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
甲壳类动物宏基因组序列

[0023]

<400> 10
Met Glu Lys Ser Asn Thr Arg Lys Val Ile Asp Glu His Phe Lys Gly
1 5 10 15
Leu Leu Phe Arg Lys Asp Ile Leu Gln Lys Ala Gly Lys Ile Tyr Lys
20 25 30
Lys Glu Gly Glu Glu Ala Thr Ile Ser Phe Leu Met Gly Lys Asp Glu
35 40 45
Glu Ala Pro Pro Asn Phe Gln Pro Pro Ala Lys Thr Ser Ile Val Ala
50 55 60
Gln Ser Arg Pro Phe Asn Gln Trp Pro Ile Tyr Gln Val Ser Glu Ala
65 70 75 80
Ile Gln Lys Arg Val Phe Gly Tyr Thr Glu Asp Glu Phe Tyr Ala Gln
85 90 95
Lys Lys Ala Leu Phe Gly Glu Gly Gly Ala Ser Ser Lys Ser Arg Asp
100 105 110
Ala Trp Phe Lys Ala Asn Gly Ile Ser Asp Arg Gly Ile Val Ala Gln
115 120 125
Gly Leu Asn Met Ile Leu Gly His Ala Phe Ala Arg Tyr Glu Gly Val
130 135 140
Ile Gln Lys Val Glu Asn Arg Asn Lys Lys Arg Leu Asp Lys Leu Ser
145 150 155 160

[0024]

Lys Lys Asn Gln Leu Arg Val Lys Glu Gly Leu Glu Val Tyr Glu Phe
165 170 175
Thr Pro Glu Ser Ala Phe Ile Asp Gly Ser Gly Leu Leu Ala Gln Pro
180 185 190
Pro Gly Ile Ser Pro Asn Ile Tyr Gly Tyr Gln Ala Ile Ala Pro Phe
195 200 205
Val Phe Asp Pro Asp Asp Pro Arg Asp Ile Val Leu Pro Lys Glu Tyr
210 215 220
Glu Gly Tyr Ser Arg Lys Pro Asp Asp Ile Ile Glu Lys Gly Pro Asn
225 230 235 240
Arg Leu Asp Ile Pro Lys Gly Gln Pro Gly Tyr Val Pro Glu His Gln
245 250 255
Arg Ser Gly Leu Lys Lys Gly Gly Arg Val Trp Leu Tyr Arg Arg Ala
260 265 270
Thr Thr Arg Ala Lys Ala Leu Ala Ser Ile Leu Gly Val Leu Gln Ile
275 280 285
Gly Glu Asp Trp Val Leu Phe Asp Met Arg Gly Leu Leu Arg Asn Ala
290 295 300
Tyr Met Arg Lys Ala Leu Thr Pro Gly Lys Ala Ser Ala Arg Asp Leu
305 310 315 320
Leu Gly Thr Phe Thr Glu Tyr Pro Val Leu Asn Ala Arg Thr Gly Glu
325 330 335
Phe Thr Phe Cys Tyr Lys Leu Arg Ser Gly Gly Ser Leu Tyr Ala Arg
340 345 350
Gln Val Tyr Lys Lys Gly Lys Thr Arg Glu Ile Leu Thr Glu Leu Thr
355 360 365
Ser Glu Gly Lys Thr Ile Ala Leu Val Thr Val Asp Leu Gly Gln Arg
370 375 380
Asn Pro Val Ala Ala Met Val Ala Arg Val Ser Arg Asp Gly Glu Leu
385 390 395 400
Ser Glu Ser Cys Ile Asp Pro Val Ser Arg Phe Leu Leu Pro Glu Tyr
405 410 415
Tyr Ala Arg Gln Ile Gln Lys Tyr Arg Asp Asp Phe Asp Ala Phe Arg
420 425 430
Gln Glu Val Trp Asp Glu Ala Phe Ala Ser Met Pro Pro Glu Tyr Gln
435 440 445
Glu Gln Ile Arg Gln Tyr Glu Ala Tyr Thr Pro Asp Gln Ala Lys Ser
450 455 460
Leu Val Leu Lys His Phe Phe Gly Asp Glu Val Ser Leu Asp Asp Leu
465 470 475 480
Pro Trp Glu Lys Met Thr Ser Asn Thr Cys Tyr Ile Ser Asn Leu Tyr
485 490 495
Ile Lys Arg Gly Gly Asp Pro Ser Arg Val Thr Phe Thr Pro Ser Pro
500 505 510

[0025]

Gly Lys Asn Ser Lys Lys Pro Arg Lys Pro Val Lys Arg Thr Asp Ser
515 520 525
Gly Ile Ser Arg Leu Pro Glu Val Arg Pro Gly Leu Pro Lys Asp Thr
530 535 540
Arg Asp Ala Phe Glu Glu Ala Lys Trp Asp Val Tyr Arg Gly His Glu
545 550 555 560
Lys Phe Pro Lys Leu Ala Lys Arg Val Asn Gln Leu Cys Arg Glu Ile
565 570 575
Ala Asn Trp Leu Glu Lys Glu Ala Gly Arg Ile Thr Leu Cys Asp Thr
580 585 590
Val Val Phe Gly Ile Glu Asp Met Gly Ala Lys Phe Cys Gly Lys Gly
595 600 605
Lys Gly Lys Phe Gln Glu Thr Trp Glu Gly Phe Phe Arg Gln Lys Ser
610 615 620
Glu Asn Arg Trp Val Met Asn Leu Leu Lys Ser Ser Ile His Met Arg
625 630 635 640
Ala His Asp Lys Gly Arg Tyr Val Leu Glu Leu Ala Pro Phe Tyr Thr
645 650 655
Ser Gln Arg Cys Pro Lys Cys Gly Tyr Ile His Lys Asn Asn Arg Lys
660 665 670
Gly Asp Arg Phe Glu Cys Leu Ser Cys Gly Ala Leu Leu His Ala Asp
675 680 685
Ser Glu Val Ala Thr Trp Asn Leu Ala Val Val Ala Ile Leu Gly Lys
690 695 700
Ala Leu Lys Lys Pro Ser Leu Lys Cys Glu Lys Ser Ser Gly Gln Lys
705 710 715 720
Lys Ala Arg Thr Ser Arg Lys Ile Gln Ile Lys Val Gly Asn Lys Ala
725 730 735
Glu Thr Ser Ser Ser Pro Gln Glu Asn Gly Glu Val Leu Ala Pro Pro
740 745 750
Glu Glu Asn Ser Gly Thr Ser Arg Asp Pro Val Tyr Asn Pro Ser Gly
755 760 765
Thr

<210> 11
<211> 769
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
甲壳类动物宏基因组序列
<400> 11
Met Glu Lys Ser Asn Thr Arg Lys Val Ile Asp Glu His Phe Lys Gly
1 5 10 15

[0026]

Leu Leu Phe Arg Lys Asp Ile Leu Gln Lys Ala Gly Lys Ile Tyr Lys
20 25 30
Lys Glu Gly Glu Glu Ala Thr Ile Ser Phe Leu Met Gly Lys Asp Glu
35 40 45
Glu Ala Pro Pro Asn Phe Gln Pro Pro Ala Lys Thr Ser Ile Val Ala
50 55 60
Gln Ser Arg Pro Phe Asn Gln Trp Pro Ile Tyr Gln Val Ser Glu Ala
65 70 75 80
Ile Gln Lys Arg Val Phe Gly Tyr Thr Glu Asp Glu Phe Tyr Ala Gln
85 90 95
Lys Lys Ala Leu Phe Gly Glu Gly Gly Ala Ser Ser Lys Ser Arg Asp
100 105 110
Ala Trp Phe Lys Ala Asn Gly Ile Ser Asp Arg Gly Ile Val Ala Gln
115 120 125
Gly Leu Asn Met Ile Leu Gly His Ala Phe Ala Arg Tyr Glu Gly Val
130 135 140
Ile Gln Lys Val Glu Asn Arg Asn Lys Lys Arg Leu Asp Lys Leu Ser
145 150 155 160
Lys Lys Asn Gln Leu Arg Val Lys Glu Gly Leu Glu Val Tyr Glu Phe
165 170 175
Thr Pro Glu Ser Ala Phe Ile Asp Gly Ser Gly Leu Leu Ala Gln Pro
180 185 190
Pro Gly Ile Ser Pro Asn Ile Tyr Gly Tyr Gln Ala Ile Ala Pro Phe
195 200 205
Val Phe Asp Pro Asp Asp Pro Arg Asp Ile Val Leu Pro Lys Glu Tyr
210 215 220
Glu Gly Tyr Ser Arg Lys Pro Asp Asp Ile Ile Glu Lys Gly Pro Asn
225 230 235 240
Arg Leu Asp Ile Pro Lys Gly Gln Pro Gly Tyr Val Pro Glu His Gln
245 250 255
Arg Ser Gly Leu Lys Lys Gly Gly Arg Val Trp Leu Tyr Arg Arg Ala
260 265 270
Thr Thr Arg Ala Lys Ala Leu Ala Ser Ile Leu Gly Val Leu Gln Ile
275 280 285
Gly Glu Asp Trp Val Leu Phe Asp Met Arg Gly Leu Leu Arg Asn Ala
290 295 300
Tyr Met Arg Lys Ala Leu Thr Pro Gly Lys Ala Ser Ala Arg Asp Leu
305 310 315 320
Leu Gly Thr Phe Thr Glu Tyr Pro Val Leu Asn Ala Arg Thr Gly Glu
325 330 335
Phe Thr Phe Cys Tyr Lys Leu Arg Ser Gly Gly Ser Leu Tyr Ala Arg
340 345 350
Gln Val Tyr Lys Lys Gly Lys Thr Arg Glu Ile Leu Thr Glu Leu Thr
355 360 365

[0027]

```
Ser Glu Gly Lys Thr Ile Ala Leu Val Thr Val Asp Leu Gly Gln Arg
    370                 375                 380
Asn Pro Val Ala Ala Met Val Ala Arg Val Ser Arg Asp Gly Glu Leu
385                 390                 395                 400
Ser Glu Ser Cys Ile Asp Pro Val Ser Arg Phe Leu Leu Pro Glu Tyr
                405                 410                 415
Tyr Ala Arg Gln Ile Gln Lys Tyr Arg Asp Asp Phe Asp Ala Phe Arg
            420                 425                 430
Gln Glu Val Trp Asp Glu Ala Phe Ala Ser Met Pro Pro Glu Tyr Gln
        435                 440                 445
Glu Gln Ile Arg Gln Tyr Glu Ala Tyr Thr Pro Asp Gln Ala Lys Ser
    450                 455                 460
Leu Val Leu Lys His Phe Phe Gly Asp Glu Val Ser Leu Asp Asp Leu
465                 470                 475                 480
Pro Trp Glu Lys Met Thr Ser Asn Thr Cys Tyr Ile Ser Asn Leu Tyr
                485                 490                 495
Ile Lys Arg Gly Gly Asp Pro Ser Arg Val Thr Phe Thr Pro Ser Pro
            500                 505                 510
Gly Lys Asn Ser Lys Lys Pro Arg Lys Pro Val Lys Arg Thr Asp Ser
        515                 520                 525
Gly Ile Ser Arg Leu Pro Glu Val Arg Pro Gly Leu Pro Lys Asp Thr
    530                 535                 540
Arg Asp Ala Phe Glu Glu Ala Lys Trp Asp Val Tyr Arg Gly His Glu
545                 550                 555                 560
Lys Phe Pro Lys Leu Ala Lys Arg Val Asn Gln Leu Cys Arg Glu Ile
                565                 570                 575
Ala Asn Trp Leu Glu Lys Glu Ala Gly Arg Ile Thr Leu Cys Asp Thr
            580                 585                 590
Val Val Phe Gly Ile Glu Asp Met Gly Ala Lys Phe Cys Gly Lys Gly
        595                 600                 605
Lys Gly Lys Phe Gln Glu Thr Trp Glu Gly Phe Phe Arg Gln Lys Ser
    610                 615                 620
Glu Asn Arg Trp Val Met Asn Leu Leu Lys Ser Ser Ile His Met Arg
625                 630                 635                 640
Ala His Asp Lys Gly Arg Tyr Val Leu Glu Leu Ala Pro Phe Tyr Thr
                645                 650                 655
Ser Gln Arg Cys Pro Lys Cys Gly Tyr Ile His Lys Asn Asn Arg Lys
            660                 665                 670
Gly Asp Arg Phe Glu Cys Leu Ser Cys Gly Ala Leu Leu His Ala Asp
        675                 680                 685
Ser Glu Val Ala Thr Trp Asn Leu Ala Val Val Ala Ile Leu Gly Lys
    690                 695                 700
Ala Leu Lys Lys Pro Ser Leu Lys Cys Glu Lys Ser Ser Gly Gln Lys
705                 710                 715                 720
```

Lys Ala Arg Thr Ser Arg Lys Ile Gln Ile Lys Val Gly Asn Lys Ala
725 730 735
Glu Thr Ser Ser Ser Pro Gln Glu Asn Gly Glu Val Leu Ala Pro Pro
740 745 750
Glu Glu Asn Ser Gly Thr Ser Arg Asp Pro Val Tyr Asn Pro Ser Gly
755 760 765
Thr

<210> 12
<211> 812
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
微生物垫宏基因组序列
<400> 12

[0028]

Met Asp Met Leu Asp Thr Glu Thr Asn Tyr Ala Thr Glu Thr Pro Ser
1 5 10 15
Gln Gln Gln Asp Tyr Ser Pro Lys Pro Pro Lys Lys Asp Arg Arg Ala
20 25 30
Pro Lys Gly Phe Ser Lys Lys Ala Arg Pro Glu Lys Lys Pro Pro Lys
35 40 45
Pro Ile Thr Leu Phe Thr Gln Lys His Phe Ser Gly Val Arg Phe Leu
50 55 60
Lys Arg Val Ile Arg Asp Ala Ser Lys Ile Leu Lys Leu Ser Glu Ser
65 70 75 80
Arg Thr Ile Thr Phe Leu Glu Gln Ala Ile Glu Arg Asp Gly Ser Ala
85 90 95
Pro Pro Asp Val Thr Pro Pro Val His Asn Thr Ile Met Ala Val Thr
100 105 110
Arg Pro Phe Glu Glu Trp Pro Glu Val Ile Leu Ser Lys Ala Leu Gln
115 120 125
Lys His Cys Tyr Ala Leu Thr Lys Lys Ile Lys Ile Lys Thr Trp Pro
130 135 140
Lys Lys Gly Pro Gly Lys Lys Cys Leu Ala Ala Trp Ser Ala Arg Thr
145 150 155 160
Lys Ile Pro Leu Ile Pro Gly Gln Val Gln Ala Thr Asn Gly Leu Phe
165 170 175
Asp Arg Ile Gly Ser Ile Tyr Asp Gly Val Glu Lys Lys Val Thr Asn
180 185 190
Arg Asn Ala Asn Lys Lys Leu Glu Tyr Asp Glu Ala Ile Lys Glu Gly
195 200 205
Arg Asn Pro Ala Val Pro Glu Tyr Glu Thr Ala Tyr Asn Ile Asp Gly
210 215 220

[0029]

```
Thr Leu Ile Asn Lys Pro Gly Tyr Asn Pro Asn Leu Tyr Ile Thr Gln
225                 230                 235                 240
Ser Arg Thr Pro Arg Leu Ile Thr Glu Ala Asp Arg Pro Leu Val Glu
                245                 250                 255
Lys Ile Leu Trp Gln Met Val Glu Lys Lys Thr Gln Ser Arg Asn Gln
            260                 265                 270
Ala Arg Arg Ala Arg Leu Glu Lys Ala Ala His Leu Gln Gly Leu Pro
        275                 280                 285
Val Pro Lys Phe Val Pro Glu Lys Val Asp Arg Ser Gln Lys Ile Glu
    290                 295                 300
Ile Arg Ile Ile Asp Pro Leu Asp Lys Ile Glu Pro Tyr Met Pro Gln
305                 310                 315                 320
Asp Arg Met Ala Ile Lys Ala Ser Gln Asp Gly His Val Pro Tyr Trp
                325                 330                 335
Gln Arg Pro Phe Leu Ser Lys Arg Arg Asn Arg Arg Val Arg Ala Gly
            340                 345                 350
Trp Gly Lys Gln Val Ser Ser Ile Gln Ala Trp Leu Thr Gly Ala Leu
        355                 360                 365
Leu Val Ile Val Arg Leu Gly Asn Glu Ala Phe Leu Ala Asp Ile Arg
    370                 375                 380
Gly Ala Leu Arg Asn Ala Gln Trp Arg Lys Leu Leu Lys Pro Asp Ala
385                 390                 395                 400
Thr Tyr Gln Ser Leu Phe Asn Leu Phe Thr Gly Asp Pro Val Val Asn
                405                 410                 415
Thr Arg Thr Asn His Leu Thr Met Ala Tyr Arg Glu Gly Val Val Asp
            420                 425                 430
Ile Val Lys Ser Arg Ser Phe Lys Gly Arg Gln Thr Arg Glu His Leu
        435                 440                 445
Leu Thr Leu Leu Gly Gln Gly Lys Thr Val Ala Gly Val Ser Phe Asp
    450                 455                 460
Leu Gly Gln Lys His Ala Ala Gly Leu Leu Ala Ala His Phe Gly Leu
465                 470                 475                 480
Gly Glu Asp Gly Asn Pro Val Phe Thr Pro Ile Gln Ala Cys Phe Leu
                485                 490                 495
Pro Gln Arg Tyr Leu Asp Ser Leu Thr Asn Tyr Arg Asn Arg Tyr Asp
            500                 505                 510
Ala Leu Thr Leu Asp Met Arg Arg Gln Ser Leu Leu Ala Leu Thr Pro
        515                 520                 525
Ala Gln Gln Gln Glu Phe Ala Asp Ala Gln Arg Asp Pro Gly Gly Gln
    530                 535                 540
Ala Lys Arg Ala Cys Cys Leu Lys Leu Asn Leu Asn Pro Asp Glu Ile
545                 550                 555                 560
Arg Trp Asp Leu Val Ser Gly Ile Ser Thr Met Ile Ser Asp Leu Tyr
                565                 570                 575
```

Ile Glu Arg Gly Gly Asp Pro Arg Asp Val His Gln Gln Val Glu Thr
580 585 590
Lys Pro Lys Gly Lys Arg Lys Ser Glu Ile Arg Ile Leu Lys Ile Arg
595 600 605
Asp Gly Lys Trp Ala Tyr Asp Phe Arg Pro Lys Ile Ala Asp Glu Thr
610 615 620
Arg Lys Ala Gln Arg Glu Gln Leu Trp Lys Leu Gln Lys Ala Ser Ser
625 630 635 640
Glu Phe Glu Arg Leu Ser Arg Tyr Lys Ile Asn Ile Ala Arg Ala Ile
645 650 655
Ala Asn Trp Ala Leu Gln Trp Gly Arg Glu Leu Ser Gly Cys Asp Ile
660 665 670
Val Ile Pro Val Leu Glu Asp Leu Asn Val Gly Ser Lys Phe Phe Asp
675 680 685
Gly Lys Gly Lys Trp Leu Leu Gly Trp Asp Asn Arg Phe Thr Pro Lys
690 695 700
Lys Glu Asn Arg Trp Phe Ile Lys Val Leu His Lys Ala Val Ala Glu
705 710 715 720
Leu Ala Pro His Arg Gly Val Pro Val Tyr Glu Val Met Pro His Arg
725 730 735
Thr Ser Met Thr Cys Pro Ala Cys His Tyr Cys His Pro Thr Asn Arg
740 745 750
[0030]
Glu Gly Asp Arg Phe Glu Cys Gln Ser Cys His Val Val Lys Asn Thr
755 760 765
Asp Arg Asp Val Ala Pro Tyr Asn Ile Leu Arg Val Ala Val Glu Gly
770 775 780
Lys Thr Leu Asp Arg Trp Gln Ala Glu Lys Lys Pro Gln Ala Glu Pro
785 790 795 800
Asp Arg Pro Met Ile Leu Ile Asp Asn Gln Glu Ser
805 810

<210> 13
<211> 708
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
植物-根面-柳枝稷根围序列
<400> 13
Met Ser Lys Thr Lys Glu Leu Asn Asp Tyr Gln Glu Ala Leu Ala Arg
1 5 10 15
Arg Leu Pro Gly Val Arg His Gln Lys Ser Val Arg Arg Ala Ala Arg
20 25 30
Leu Val Tyr Asp Arg Gln Gly Glu Asp Ala Met Val Ala Phe Leu Asp
35 40 45

[0031]

```
Gly Lys Glu Val Asp Glu Pro Tyr Thr Leu Gln Pro Pro Ala Lys Cys
    50                  55                  60
His Ile Leu Ala Val Ser Arg Pro Ile Glu Glu Trp Pro Ile Ala Arg
65                  70                  75                  80
Val Thr Met Ala Val Gln Glu His Val Tyr Ala Leu Pro Val His Glu
                85                  90                  95
Val Glu Lys Ser Arg Pro Glu Thr Thr Glu Gly Ser Arg Ser Ala Trp
            100                 105                 110
Phe Lys Asn Ser Gly Val Ser Asn His Gly Val Thr His Ala Gln Thr
        115                 120                 125
Leu Asn Ala Ile Leu Lys Asn Ala Tyr Asn Val Tyr Asn Gly Val Ile
    130                 135                 140
Lys Lys Val Glu Asn Arg Asn Ala Lys Lys Arg Asp Ser Leu Ala Ala
145                 150                 155                 160
Lys Asn Lys Ser Arg Glu Arg Lys Gly Leu Pro His Phe Lys Ala Asp
                165                 170                 175
Pro Pro Glu Leu Ala Thr Asp Glu Gln Gly Tyr Leu Leu Gln Pro Pro
            180                 185                 190
Ser Pro Asn Ser Ser Val Tyr Leu Val Gln Gln His Leu Arg Thr Pro
        195                 200                 205
Gln Ile Asp Leu Pro Ser Gly Tyr Thr Gly Pro Val Val Asp Pro Arg
    210                 215                 220
Ser Pro Ile Pro Ser Leu Ile Pro Ile Asp Arg Leu Ala Ile Pro Pro
225                 230                 235                 240
Gly Gln Pro Gly Tyr Val Pro Leu His Asp Arg Glu Lys Leu Thr Ser
                245                 250                 255
Asn Lys His Arg Arg Met Lys Leu Pro Lys Ser Leu Arg Ala Gln Gly
            260                 265                 270
Ala Leu Pro Val Cys Phe Arg Val Phe Asp Asp Trp Ala Val Val Asp
        275                 280                 285
Gly Arg Gly Leu Leu Arg His Ala Gln Tyr Arg Arg Leu Ala Pro Lys
    290                 295                 300
Asn Val Ser Ile Ala Glu Leu Leu Glu Leu Tyr Thr Gly Asp Pro Val
305                 310                 315                 320
Ile Asp Ile Lys Arg Asn Leu Met Thr Phe Arg Phe Ala Glu Ala Val
                325                 330                 335
Val Glu Val Thr Ala Arg Lys Ile Val Glu Lys Tyr His Asn Lys Tyr
            340                 345                 350
Leu Leu Lys Leu Thr Glu Pro Lys Gly Lys Pro Val Arg Glu Ile Gly
        355                 360                 365
Leu Val Ser Ile Asp Leu Asn Val Gln Arg Leu Ile Ala Leu Ala Ile
    370                 375                 380
Tyr Arg Val His Gln Thr Gly Glu Ser Gln Leu Ala Leu Ser Pro Cys
385                 390                 395                 400
```

[0032]

Leu His Arg Glu Ile Leu Pro Ala Lys Gly Leu Gly Asp Phe Asp Lys
405 410 415
Tyr Lys Ser Lys Phe Asn Gln Leu Thr Glu Glu Ile Leu Thr Ala Ala
420 425 430
Val Gln Thr Leu Thr Ser Ala Gln Gln Glu Glu Tyr Gln Arg Tyr Val
435 440 445
Glu Glu Ser Ser His Glu Ala Lys Ala Asp Leu Cys Leu Lys Tyr Ser
450 455 460
Ile Thr Pro His Glu Leu Ala Trp Asp Lys Met Thr Ser Ser Thr Gln
465 470 475 480
Tyr Ile Ser Arg Trp Leu Arg Asp His Gly Trp Asn Ala Ser Asp Phe
485 490 495
Thr Gln Ile Thr Lys Gly Arg Lys Lys Val Glu Arg Leu Trp Ser Asp
500 505 510
Ser Arg Trp Ala Gln Glu Leu Lys Pro Lys Leu Ser Asn Glu Thr Arg
515 520 525
Arg Lys Leu Glu Asp Ala Lys His Asp Leu Gln Arg Ala Asn Pro Glu
530 535 540
Trp Gln Arg Leu Ala Lys Arg Lys Gln Glu Tyr Ser Arg His Leu Ala
545 550 555 560
Asn Thr Val Leu Ser Met Ala Arg Glu Tyr Thr Ala Cys Glu Thr Val
565 570 575
Val Ile Ala Ile Glu Asn Leu Pro Met Lys Gly Gly Phe Val Asp Gly
580 585 590
Asn Gly Ser Arg Glu Ser Gly Trp Asp Asn Phe Phe Thr His Lys Lys
595 600 605
Glu Asn Arg Trp Met Ile Lys Asp Ile His Lys Ala Leu Ser Asp Leu
610 615 620
Ala Pro Asn Arg Gly Val His Val Leu Glu Val Asn Pro Gln Tyr Thr
625 630 635 640
Ser Gln Thr Cys Pro Glu Cys Gly His Arg Asp Lys Ala Asn Arg Asp
645 650 655
Pro Ile Gln Arg Glu Arg Phe Cys Cys Thr His Cys Gly Ala Gln Arg
660 665 670
His Ala Asp Leu Glu Val Ala Thr His Asn Ile Ala Met Val Ala Thr
675 680 685
Thr Gly Lys Ser Leu Thr Gly Lys Ser Leu Ala Pro Gln Arg Leu Gln
690 695 700
Glu Ala Ala Glu
705

<210> 14
<211> 666
<212> PRT
<213> 未知

<220>

<223> 未知的描述:
根围宏基因组序列

<400> 14

[0033]

Met Val Ala Phe Leu Asp Gly Lys Glu Val Asp Glu Pro Tyr Thr Leu
1 5 10 15
Gln Pro Pro Ala Lys Cys His Ile Leu Ala Val Ser Arg Pro Ile Glu
20 25 30
Glu Trp Pro Ile Ala Arg Val Thr Met Ala Val Gln Glu His Val Tyr
35 40 45
Ala Leu Pro Val His Glu Val Glu Lys Ser Arg Pro Glu Thr Thr Glu
50 55 60
Gly Ser Arg Ser Ala Trp Phe Lys Asn Ser Gly Val Ser Asn His Gly
65 70 75 80
Val Thr His Ala Gln Thr Leu Asn Ala Ile Leu Lys Asn Ala Tyr Asn
85 90 95
Val Tyr Asn Gly Val Ile Lys Lys Val Glu Asn Arg Asn Ala Lys Lys
100 105 110
Arg Asp Ser Leu Ala Ala Lys Asn Lys Ser Arg Glu Arg Lys Gly Leu
115 120 125
Pro His Phe Lys Ala Asp Pro Pro Glu Leu Ala Thr Asp Glu Gln Gly
130 135 140
Tyr Leu Leu Gln Pro Pro Ser Pro Asn Ser Ser Val Tyr Leu Val Gln
145 150 155 160
Gln His Leu Arg Thr Pro Gln Ile Asp Leu Pro Ser Gly Tyr Thr Gly
165 170 175
Pro Val Val Asp Pro Arg Ser Pro Ile Pro Ser Leu Ile Pro Ile Asp
180 185 190
Arg Leu Ala Ile Pro Pro Gly Gln Pro Gly Tyr Val Pro Leu His Asp
195 200 205
Arg Glu Lys Leu Thr Ser Asn Lys His Arg Arg Met Lys Leu Pro Lys
210 215 220
Ser Leu Arg Ala Gln Gly Ala Leu Pro Val Cys Phe Arg Val Phe Asp
225 230 235 240
Asp Trp Ala Val Val Asp Gly Arg Gly Leu Leu Arg His Ala Gln Tyr
245 250 255
Arg Arg Leu Ala Pro Lys Asn Val Ser Ile Ala Glu Leu Leu Glu Leu
260 265 270
Tyr Thr Gly Asp Pro Val Ile Asp Ile Lys Arg Asn Leu Met Thr Phe
275 280 285
Arg Phe Ala Glu Ala Val Val Glu Val Thr Ala Arg Lys Ile Val Glu
290 295 300
Lys Tyr His Asn Lys Tyr Leu Leu Lys Leu Thr Glu Pro Lys Gly Lys
305 310 315 320

[0034]

```
Pro Val Arg Glu Ile Gly Leu Val Ser Ile Asp Leu Asn Val Gln Arg
                325                 330                 335
Leu Ile Ala Leu Ala Ile Tyr Arg Val His Gln Thr Gly Glu Ser Gln
            340                 345                 350
Leu Ala Leu Ser Pro Cys Leu His Arg Glu Ile Leu Pro Ala Lys Gly
        355                 360                 365
Leu Gly Asp Phe Asp Lys Tyr Lys Ser Lys Phe Asn Gln Leu Thr Glu
    370                 375                 380
Glu Ile Leu Thr Ala Ala Val Gln Thr Leu Thr Ser Ala Gln Gln Glu
385                 390                 395                 400
Glu Tyr Gln Arg Tyr Val Glu Glu Ser Ser His Glu Ala Lys Ala Asp
                405                 410                 415
Leu Cys Leu Lys Tyr Ser Ile Thr Pro His Glu Leu Ala Trp Asp Lys
            420                 425                 430
Met Thr Ser Ser Thr Gln Tyr Ile Ser Arg Trp Leu Arg Asp His Gly
        435                 440                 445
Trp Asn Ala Ser Asp Phe Thr Gln Ile Thr Lys Gly Arg Lys Lys Val
    450                 455                 460
Glu Arg Leu Trp Ser Asp Ser Arg Trp Ala Gln Glu Leu Lys Pro Lys
465                 470                 475                 480
Leu Ser Asn Glu Thr Arg Arg Lys Leu Glu Asp Ala Lys His Asp Leu
                485                 490                 495
Gln Arg Ala Asn Pro Glu Trp Gln Arg Leu Ala Lys Arg Lys Gln Glu
            500                 505                 510
Tyr Ser Arg His Leu Ala Asn Thr Val Leu Ser Met Ala Arg Glu Tyr
        515                 520                 525
Thr Ala Cys Glu Thr Val Val Ile Ala Ile Glu Asn Leu Pro Met Lys
    530                 535                 540
Gly Gly Phe Val Asp Gly Asn Gly Ser Arg Glu Ser Gly Trp Asp Asn
545                 550                 555                 560
Phe Phe Thr His Lys Lys Glu Asn Arg Trp Met Ile Lys Asp Ile His
                565                 570                 575
Lys Ala Leu Ser Asp Leu Ala Pro Asn Arg Gly Val His Val Leu Glu
            580                 585                 590
Val Asn Pro Gln Tyr Thr Ser Gln Thr Cys Pro Glu Cys Gly His Arg
        595                 600                 605
Asp Lys Ala Asn Arg Asp Pro Ile Gln Arg Glu Arg Phe Cys Cys Thr
    610                 615                 620
His Cys Gly Ala Gln Arg His Ala Asp Leu Glu Val Ala Thr His Asn
625                 630                 635                 640
Ile Ala Met Val Ala Thr Thr Gly Lys Ser Leu Thr Gly Lys Ser Leu
                645                 650                 655
Ala Pro Gln Arg Leu Gln Glu Ala Ala Glu
            660                 665
```

<210> 15
<211> 708
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
根围宏基因组序列
<400> 15

[0035]

```
Met Ser Lys Thr Lys Glu Leu Asn Asp Tyr Gln Glu Ala Leu Ala Arg
1               5                  10                  15
Arg Leu Pro Gly Val Arg His Gln Lys Ser Val Arg Arg Ala Ala Arg
            20                  25                  30
Leu Val Tyr Asp Arg Gln Gly Glu Asp Ala Met Val Ala Phe Leu Asp
        35                  40                  45
Gly Lys Glu Val Asp Glu Pro Tyr Thr Leu Gln Pro Pro Ala Lys Cys
    50                  55                  60
His Ile Leu Ala Val Ser Arg Pro Ile Glu Glu Trp Pro Ile Ala Arg
65                  70                  75                  80
Val Thr Met Ala Val Gln Glu His Val Tyr Ala Leu Pro Val His Glu
                85                  90                  95
Val Glu Lys Ser Arg Pro Glu Thr Thr Glu Gly Ser Arg Ser Ala Trp
            100                 105                 110
Phe Lys Asn Ser Gly Val Ser Asn His Gly Val Thr His Ala Gln Thr
        115                 120                 125
Leu Asn Ala Ile Leu Lys Asn Ala Tyr Asn Val Tyr Asn Gly Val Ile
    130                 135                 140
Lys Lys Val Glu Asn Arg Asn Ala Lys Lys Arg Asp Ser Leu Ala Ala
145                 150                 155                 160
Lys Asn Lys Ser Arg Glu Arg Lys Gly Leu Pro His Phe Lys Ala Asp
                165                 170                 175
Pro Pro Glu Leu Ala Thr Asp Glu Gln Gly Tyr Leu Leu Gln Pro Pro
            180                 185                 190
Ser Pro Asn Ser Ser Val Tyr Leu Val Gln Gln His Leu Arg Thr Pro
        195                 200                 205
Gln Ile Asp Leu Pro Ser Gly Tyr Thr Gly Pro Val Val Asp Pro Arg
    210                 215                 220
Ser Pro Ile Pro Ser Leu Ile Pro Ile Asp Arg Leu Ala Ile Pro Pro
225                 230                 235                 240
Gly Gln Pro Gly Tyr Val Pro Leu His Asp Arg Glu Lys Leu Thr Ser
                245                 250                 255
Asn Lys His Arg Arg Met Lys Leu Pro Lys Ser Leu Arg Ala Gln Gly
            260                 265                 270
Ala Leu Pro Val Cys Phe Arg Val Phe Asp Asp Trp Ala Val Val Asp
        275                 280                 285
```

[0036]

```
Gly Arg Gly Leu Leu Arg His Ala Gln Tyr Arg Arg Leu Ala Pro Lys
    290                 295                 300
Asn Val Ser Ile Ala Glu Leu Leu Glu Leu Tyr Thr Gly Asp Pro Val
305                 310                 315                 320
Ile Asp Ile Lys Arg Asn Leu Met Thr Phe Arg Phe Ala Glu Ala Val
                325                 330                 335
Val Glu Val Thr Ala Arg Lys Ile Val Glu Lys Tyr His Asn Lys Tyr
            340                 345                 350
Leu Leu Lys Leu Thr Glu Pro Lys Gly Lys Pro Val Arg Glu Ile Gly
        355                 360                 365
Leu Val Ser Ile Asp Leu Asn Val Gln Arg Leu Ile Ala Leu Ala Ile
    370                 375                 380
Tyr Arg Val His Gln Thr Gly Glu Ser Gln Leu Ala Leu Ser Pro Cys
385                 390                 395                 400
Leu His Arg Glu Ile Leu Pro Ala Lys Gly Leu Gly Asp Phe Asp Lys
                405                 410                 415
Tyr Lys Ser Lys Phe Asn Gln Leu Thr Glu Glu Ile Leu Thr Ala Ala
            420                 425                 430
Val Gln Thr Leu Thr Ser Ala Gln Gln Glu Glu Tyr Gln Arg Tyr Val
        435                 440                 445
Glu Glu Ser Ser His Glu Ala Lys Ala Asp Leu Cys Leu Lys Tyr Ser
    450                 455                 460
Ile Thr Pro His Glu Leu Ala Trp Asp Lys Met Thr Ser Ser Thr Gln
465                 470                 475                 480
Tyr Ile Ser Arg Trp Leu Arg Asp His Gly Trp Asn Ala Ser Asp Phe
                485                 490                 495
Thr Gln Ile Thr Lys Gly Arg Lys Lys Val Glu Arg Leu Trp Ser Asp
            500                 505                 510
Ser Arg Trp Ala Gln Glu Leu Lys Pro Lys Leu Ser Asn Glu Thr Arg
        515                 520                 525
Arg Lys Leu Glu Asp Ala Lys His Asp Leu Gln Arg Ala Asn Pro Glu
    530                 535                 540
Trp Gln Arg Leu Ala Lys Arg Lys Gln Glu Tyr Ser Arg His Leu Ala
545                 550                 555                 560
Asn Thr Val Leu Ser Met Ala Arg Glu Tyr Thr Ala Cys Glu Thr Val
                565                 570                 575
Val Ile Ala Ile Glu Asn Leu Pro Met Lys Gly Gly Phe Val Asp Gly
            580                 585                 590
Asn Gly Ser Arg Glu Ser Gly Trp Asp Asn Phe Phe Thr His Lys Lys
        595                 600                 605
Glu Asn Arg Trp Met Ile Lys Asp Ile His Lys Ala Leu Ser Asp Leu
    610                 615                 620
Ala Pro Asn Arg Gly Val His Val Leu Glu Val Asn Pro Gln Tyr Thr
625                 630                 635                 640
```

Ser Gln Thr Cys Pro Glu Cys Gly His Arg Asp Lys Ala Asn Arg Asp
645 650 655
Pro Ile Gln Arg Glu Arg Phe Cys Cys Thr His Cys Gly Ala Gln Arg
660 665 670
His Ala Asp Leu Glu Val Ala Thr His Asn Ile Ala Met Val Ala Thr
675 680 685
Thr Gly Lys Ser Leu Thr Gly Lys Ser Leu Ala Pro Gln Arg Leu Gln
690 695 700
Glu Ala Ala Glu
705

<210> 16
<211> 717
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
土壤宏基因组序列
<400> 16

[0037]

Met Ile Lys Pro Thr Val Ser Gln Phe Leu Thr Pro Gly Phe Lys Leu
1 5 10 15
Ile Arg Asn His Ser Arg Thr Ala Gly Leu Lys Leu Lys Asn Glu Gly
20 25 30
Glu Glu Ala Cys Lys Lys Phe Val Arg Glu Asn Glu Ile Pro Lys Asp
35 40 45
Glu Cys Pro Asn Phe Gln Gly Gly Pro Ala Ile Ala Asn Ile Ile Ala
50 55 60
Lys Ser Arg Glu Phe Thr Glu Trp Glu Ile Tyr Gln Ser Ser Leu Ala
65 70 75 80
Ile Gln Glu Val Ile Phe Thr Leu Pro Lys Asp Lys Leu Pro Glu Pro
85 90 95
Ile Leu Lys Glu Glu Trp Arg Ala Gln Trp Leu Ser Glu His Gly Leu
100 105 110
Asp Thr Val Pro Tyr Lys Glu Ala Ala Gly Leu Asn Leu Ile Ile Lys
115 120 125
Asn Ala Val Asn Thr Tyr Lys Gly Val Gln Val Lys Val Asp Asn Lys
130 135 140
Asn Lys Asn Asn Leu Ala Lys Ile Asn Arg Lys Asn Glu Ile Ala Lys
145 150 155 160
Leu Asn Gly Glu Gln Glu Ile Ser Phe Glu Glu Ile Lys Ala Phe Asp
165 170 175
Asp Lys Gly Tyr Leu Leu Gln Lys Pro Ser Pro Asn Lys Ser Ile Tyr
180 185 190
Cys Tyr Gln Ser Val Ser Pro Lys Pro Phe Ile Thr Ser Lys Tyr His
195 200 205

[0038]

```
Asn Val Asn Leu Pro Glu Glu Tyr Ile Gly Tyr Tyr Arg Lys Ser Asn
    210                 215                 220
Glu Pro Ile Val Ser Pro Tyr Gln Phe Asp Arg Leu Arg Ile Pro Ile
225                 230                 235                 240
Gly Glu Pro Gly Tyr Val Pro Lys Trp Gln Tyr Thr Phe Leu Ser Lys
                245                 250                 255
Lys Glu Asn Lys Arg Arg Lys Leu Ser Lys Arg Ile Lys Asn Val Ser
            260                 265                 270
Pro Ile Leu Gly Ile Ile Cys Ile Lys Lys Asp Trp Cys Val Phe Asp
        275                 280                 285
Met Arg Gly Leu Leu Arg Thr Asn His Trp Lys Lys Tyr His Lys Pro
    290                 295                 300
Thr Asp Ser Ile Asn Asp Leu Phe Asp Tyr Phe Thr Gly Asp Pro Val
305                 310                 315                 320
Ile Asp Thr Lys Ala Asn Val Val Arg Phe Arg Tyr Lys Met Glu Asn
                325                 330                 335
Gly Ile Val Asn Tyr Lys Pro Val Arg Glu Lys Lys Gly Lys Glu Leu
            340                 345                 350
Leu Glu Asn Ile Cys Asp Gln Asn Gly Ser Cys Lys Leu Ala Thr Val
        355                 360                 365
Asp Val Gly Gln Asn Asn Pro Val Ala Ile Gly Leu Phe Glu Leu Lys
    370                 375                 380
Lys Val Asn Gly Glu Leu Thr Lys Thr Leu Ile Ser Arg His Pro Thr
385                 390                 395                 400
Pro Ile Asp Phe Cys Asn Lys Ile Thr Ala Tyr Arg Glu Arg Tyr Asp
                405                 410                 415
Lys Leu Glu Ser Ser Ile Lys Leu Asp Ala Ile Lys Gln Leu Thr Ser
            420                 425                 430
Glu Gln Lys Ile Glu Val Asp Asn Tyr Asn Asn Asn Phe Thr Pro Gln
        435                 440                 445
Asn Thr Lys Gln Ile Val Cys Ser Lys Leu Asn Ile Asn Pro Asn Asp
    450                 455                 460
Leu Pro Trp Asp Lys Met Ile Ser Gly Thr His Phe Ile Ser Glu Lys
465                 470                 475                 480
Ala Gln Val Ser Asn Lys Ser Glu Ile Tyr Phe Thr Ser Thr Asp Lys
                485                 490                 495
Gly Lys Thr Lys Asp Val Met Lys Ser Asp Tyr Lys Trp Phe Gln Asp
            500                 505                 510
Tyr Lys Pro Lys Leu Ser Lys Glu Val Arg Asp Ala Leu Ser Asp Ile
        515                 520                 525
Glu Trp Arg Leu Arg Arg Glu Ser Leu Glu Phe Asn Lys Leu Ser Lys
    530                 535                 540
Ser Arg Glu Gln Asp Ala Arg Gln Leu Ala Asn Trp Ile Ser Ser Met
545                 550                 555                 560
```

Cys Asp Val Ile Gly Ile Glu Asn Leu Val Lys Lys Asn Asn Phe Phe
565 570 575
Gly Gly Ser Gly Lys Arg Glu Pro Gly Trp Asp Asn Phe Tyr Lys Pro
580 585 590
Lys Lys Glu Asn Arg Trp Trp Ile Asn Ala Ile His Lys Ala Leu Thr
595 600 605
Glu Leu Ser Gln Asn Lys Gly Lys Arg Val Ile Leu Leu Pro Ala Met
610 615 620
Arg Thr Ser Ile Thr Cys Pro Lys Cys Lys Tyr Cys Asp Ser Lys Asn
625 630 635 640
Arg Asn Gly Glu Lys Phe Asn Cys Leu Lys Cys Gly Ile Glu Leu Asn
645 650 655
Ala Asp Ile Asp Val Ala Thr Glu Asn Leu Ala Thr Val Ala Ile Thr
660 665 670
Ala Gln Ser Met Pro Lys Pro Thr Cys Glu Arg Ser Gly Asp Ala Lys
675 680 685
Lys Pro Val Arg Ala Arg Lys Ala Lys Ala Pro Glu Phe His Asp Lys
690 695 700
Leu Ala Pro Ser Tyr Thr Val Val Leu Arg Glu Ala Val
705 710 715

[0039]

<210> 17
<211> 781
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
土壤宏基因组序列
<400> 17
Met Arg Gln Pro Ala Glu Lys Thr Ala Phe Gln Val Phe Arg Gln Glu
1 5 10 15
Val Ile Gly Thr Gln Lys Leu Ser Gly Gly Asp Ala Lys Thr Ala Gly
20 25 30
Arg Leu Tyr Lys Gln Gly Lys Met Glu Ala Ala Arg Glu Trp Leu Leu
35 40 45
Lys Gly Ala Arg Asp Asp Val Pro Pro Asn Phe Gln Pro Pro Ala Lys
50 55 60
Cys Leu Val Val Ala Val Ser His Pro Phe Glu Glu Trp Asp Ile Ser
65 70 75 80
Lys Thr Asn His Asp Val Gln Ala Tyr Ile Tyr Ala Gln Pro Leu Gln
85 90 95
Ala Glu Gly His Leu Asn Gly Leu Ser Glu Lys Trp Glu Asp Thr Ser
100 105 110
Ala Asp Gln His Lys Leu Trp Phe Glu Lys Thr Gly Val Pro Asp Arg
115 120 125

[0040]

```
Gly Leu Pro Val Gln Ala Ile Asn Lys Ile Ala Lys Ala Ala Val Asn
    130                 135                 140
Arg Ala Phe Gly Val Val Arg Lys Val Glu Asn Arg Asn Glu Lys Arg
145                 150                 155                 160
Arg Ser Arg Asp Asn Arg Ile Ala Glu His Asn Arg Glu Asn Gly Leu
                165                 170                 175
Thr Glu Val Val Arg Glu Ala Pro Glu Val Ala Thr Asn Ala Asp Gly
            180                 185                 190
Phe Leu Leu His Pro Pro Gly Ile Asp Pro Ser Ile Leu Ser Tyr Ala
        195                 200                 205
Ser Val Ser Pro Val Pro Tyr Asn Ser Ser Lys His Ser Phe Val Arg
    210                 215                 220
Leu Pro Glu Glu Tyr Gln Ala Tyr Asn Val Glu Pro Asp Ala Pro Ile
225                 230                 235                 240
Pro Gln Phe Val Val Glu Asp Arg Phe Ala Ile Pro Pro Gly Gln Pro
                245                 250                 255
Gly Tyr Val Pro Glu Trp Gln Arg Leu Lys Cys Ser Thr Asn Lys His
            260                 265                 270
Arg Arg Met Arg Gln Trp Ser Asn Gln Asp Tyr Lys Pro Lys Ala Gly
        275                 280                 285
Arg Arg Ala Lys Pro Leu Glu Phe Gln Ala His Leu Thr Arg Glu Arg
    290                 295                 300
Ala Lys Gly Ala Leu Leu Val Val Met Arg Ile Lys Glu Asp Trp Val
305                 310                 315                 320
Val Phe Asp Val Arg Gly Leu Leu Arg Asn Val Glu Trp Arg Lys Val
                325                 330                 335
Leu Ser Glu Glu Ala Arg Glu Lys Leu Thr Leu Lys Gly Leu Leu Asp
            340                 345                 350
Leu Phe Thr Gly Asp Pro Val Ile Asp Thr Lys Arg Gly Ile Val Thr
        355                 360                 365
Phe Leu Tyr Lys Ala Glu Ile Thr Lys Ile Leu Ser Lys Arg Thr Val
    370                 375                 380
Lys Thr Lys Asn Ala Arg Asp Leu Leu Leu Arg Leu Thr Glu Pro Gly
385                 390                 395                 400
Glu Asp Gly Leu Arg Arg Glu Val Gly Leu Val Ala Val Asp Leu Gly
                405                 410                 415
Gln Thr His Pro Ile Ala Ala Ala Ile Tyr Arg Ile Gly Arg Thr Ser
            420                 425                 430
Ala Gly Ala Leu Glu Ser Thr Val Leu His Arg Gln Gly Leu Arg Glu
        435                 440                 445
Asp Gln Lys Glu Lys Leu Lys Glu Tyr Arg Lys Arg His Thr Ala Leu
    450                 455                 460
Asp Ser Arg Leu Arg Lys Glu Ala Phe Glu Thr Leu Ser Val Glu Gln
465                 470                 475                 480
```

[0041]

Gln Lys Glu Ile Val Thr Val Ser Gly Ser Gly Ala Gln Ile Thr Lys
485 490 495
Asp Lys Val Cys Asn Tyr Leu Gly Val Asp Pro Ser Thr Leu Pro Trp
500 505 510
Glu Lys Met Gly Ser Tyr Thr His Phe Ile Ser Asp Asp Phe Leu Arg
515 520 525
Arg Gly Gly Asp Pro Asn Ile Val His Phe Asp Arg Gln Pro Lys Lys
530 535 540
Gly Lys Val Ser Lys Lys Ser Gln Arg Ile Lys Arg Ser Asp Ser Gln
545 550 555 560
Trp Val Gly Arg Met Arg Pro Arg Leu Ser Gln Glu Thr Ala Lys Ala
565 570 575
Arg Met Glu Ala Asp Trp Ala Ala Gln Asn Glu Asn Glu Glu Tyr Lys
580 585 590
Arg Leu Ala Arg Ser Lys Gln Glu Leu Ala Arg Trp Cys Val Asn Thr
595 600 605
Leu Leu Gln Asn Thr Arg Cys Ile Thr Gln Cys Asp Glu Ile Val Val
610 615 620
Val Ile Glu Asp Leu Asn Val Lys Ser Leu His Gly Lys Gly Ala Arg
625 630 635 640
Glu Pro Gly Trp Asp Asn Phe Phe Thr Pro Lys Thr Glu Asn Arg Trp
645 650 655
Phe Ile Gln Ile Leu His Lys Thr Phe Ser Glu Leu Pro Lys His Arg
660 665 670
Gly Glu His Val Ile Glu Gly Cys Pro Leu Arg Thr Ser Ile Thr Cys
675 680 685
Pro Ala Cys Ser Tyr Cys Asp Lys Asn Ser Arg Asn Gly Glu Lys Phe
690 695 700
Val Cys Val Ala Cys Gly Ala Thr Phe His Ala Asp Phe Glu Val Ala
705 710 715 720
Thr Tyr Asn Leu Val Arg Leu Ala Thr Thr Gly Met Pro Met Pro Lys
725 730 735
Ser Leu Glu Arg Gln Gly Gly Gly Glu Lys Ala Gly Gly Ala Arg Lys
740 745 750
Ala Arg Lys Lys Ala Lys Gln Val Glu Lys Ile Val Val Gln Ala Asn
755 760 765
Ala Asn Val Thr Met Asn Gly Ala Ser Leu His Ser Pro
770 775 780

<210> 18
<211> 769
<212> PRT
<213> 未知
<220>
<223> 未知的描述:

土壤宏基因组序列

<400> 18

[0042]

Met Glu Lys His Lys Thr Lys Leu Ser Ile Ile Met Lys Glu Phe Phe
1 5 10 15
Pro Gly Glu Arg Phe Pro Lys Asn Val Leu Met Gln Ile Gly Lys Lys
20 25 30
Ile Thr Asn Asn Lys Asp Gly Lys Glu Thr Ile Asp Val Lys Glu Lys
35 40 45
Glu Asp Val Val Ser Phe Leu Thr Gly Lys Gly Ser Lys Lys Leu Leu
50 55 60
Asp Phe Gln Pro Pro Ala Lys Ala Leu Ile Val Ala Lys Ser Arg Pro
65 70 75 80
Phe Glu Glu Trp Pro Ile Tyr Gln Ala Ser Lys Ile Phe Gln Glu Tyr
85 90 95
Ile Tyr Gly Leu Pro His Asn Gln Leu Ser Ile Pro Gly Thr Ser Lys
100 105 110
Ser Glu His Lys Leu Trp Leu Glu Lys Ile Gly Leu Asn Ile Gly Thr
115 120 125
Tyr Lys Asp Val Gln Gly Leu Asn Leu Ile Phe Arg His Thr Lys Asn
130 135 140
Ile Tyr Glu Gly Val Ile Lys Lys Val Glu Asn Lys Asn Lys Lys Asn
145 150 155 160
Lys Glu Lys Ile Glu Ile Lys Asn Lys Phe Glu Lys Glu His Gly Phe
165 170 175
Leu Leu Thr Pro Phe Glu Glu Glu Thr Ala Phe Asp Asp Asn Gly Lys
180 185 190
Leu Lys Asn Pro Pro Gly Ile Asn Asn Ser Ile Tyr Cys Tyr Ser Gln
195 200 205
Val Ser Pro Glu Ala Thr Lys Ser Thr Thr Lys Leu Asp Asn Val Pro
210 215 220
Ser Ile Tyr Leu Gly Tyr Tyr Arg Asp Ile Asp Thr Asn Ile Lys Ile
225 230 235 240
Glu Tyr Ile Asn Arg Leu Ser Ile Pro Lys Gly Asp Pro Gly Tyr Ile
245 250 255
Pro Leu Trp Gln His Glu Leu Leu Ser Lys Lys Glu Asn Asn Thr Arg
260 265 270
Arg Gln Arg Lys Trp Tyr Ser Asn Asn Arg Met Lys Arg Val Lys Arg
275 280 285
Lys Gly Val Ser Lys Tyr Ser Asp Glu Gln Ile Asn Gln Ala Arg Leu
290 295 300
Gln Asp Ala Ile Leu Gly Lys Ile Ser Ile Gly Glu Asp Trp Val Leu
305 310 315 320
Phe Asp Met Arg Gly Leu Leu Arg Asn Leu His Trp Arg Lys Leu Val
325 330 335

[0043]

```
Pro Ser Gln Gly Phe Ser Pro Lys Glu Ile Leu Glu Gln Phe Thr Gly
            340                 345                 350
Asp Pro Val Ile Asp Pro Val Arg Asn Val Ile Thr Phe Ile Tyr Lys
        355                 360                 365
Asp Gly Leu Ala His Lys Glu Glu Ile Val Leu Thr Lys Lys Ala Pro
    370                 375                 380
Asp Leu Leu Cys Lys Leu Thr Leu Asn Asn Pro Ile Gly Ile Val Ser
385                 390                 395                 400
Ile Asp Val Gly Gln Thr His Pro Gln Ser Ala Lys Phe Ser Leu Leu
                405                 410                 415
Lys Leu Glu Asp Asp Lys Leu Val Ala Glu Cys Lys Asp Arg Gln Phe
            420                 425                 430
Leu Pro Asp Tyr Leu Leu Asn Lys Leu Phe Ala Tyr Arg Glu Arg Ser
        435                 440                 445
Asp Gln Leu Arg Gly Glu Ile Asn Gln Leu Ala Met Gln Ser Leu Ser
    450                 455                 460
Glu Glu His Gln Lys Glu Phe Asn Asp Leu Lys Ile Glu Asn Asp Pro
465                 470                 475                 480
Thr Ala Val Arg Ile Arg Ile Glu Lys Gln Leu Gly Ile Asp Phe Asn
                485                 490                 495
Asn Leu Pro Ile Asn Asp Met Ile Tyr Asp Arg Thr Thr Tyr Ile Ala
            500                 505                 510
Asp Ala Tyr Leu Ser Ile Pro Gly Val Asp Lys Leu Leu Val Met Leu
        515                 520                 525
Gly Thr Ser Ser Lys Lys Lys Tyr Asp Ser Arg Ile Val Lys Asp Phe
    530                 535                 540
Phe Lys Lys Val Ser Lys Glu Ala Arg Glu Ala Leu Lys Val Ala Phe
545                 550                 555                 560
Gly Glu Ile Gln Lys Leu His Pro Gly Tyr Lys Lys Leu Ser Lys Ser
                565                 570                 575
Leu Gln Gln Trp Ala Arg Glu Cys Val Asn Phe Thr His Lys Tyr Ala
            580                 585                 590
Asn Lys Ile Thr Gly Cys Thr Asn Ile Val Phe Val Ile Glu Asn Leu
        595                 600                 605
Lys Asn Ile Arg Lys Arg Asn Gly Ser Gly Lys Arg Ala Lys Gly Tyr
    610                 615                 620
Asp Asn Phe Phe Val Tyr Lys Lys Glu Asn Arg Trp Val Met Asn Ala
625                 630                 635                 640
Leu Gln Lys Ala Tyr Ile Asp Leu Ala Thr His Lys Gly Ile Asn Ile
                645                 650                 655
Ile Glu Ile Gln Ala Ala Arg Thr Ser Ile Thr Cys Pro Lys Cys Asn
            660                 665                 670
Cys Gln Asp Lys Asn Asn Arg Lys Gly Asp Gln Phe Asn Cys Val Lys
        675                 680                 685
```

Cys Asn His Gln Ala Asn Thr Asp Leu Glu Ile Ala Thr Asp Asn Ile
690 695 700
Glu Leu Val Ala Leu Asn Gly Lys Gly Met Pro Lys Ile Asp Cys Glu
705 710 715 720
Arg Ser Ser Gly Glu Glu Asn Ala Val Gly Ala Arg Lys Gly Lys Lys
725 730 735
Thr Arg Lys Ile Lys Glu Ile Gln Glu Thr Asp Lys Asn Ile Lys Met
740 745 750
Asp Asn Ala Gly Gly Asp Leu Leu Lys Asn Asn Arg Ser Gln Thr Ala
755 760 765
Ala

<210> 19
<211> 805
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
土壤宏基因组序列
<400> 19

[0044]

Met Asp Pro Gly Met Thr Ile Thr Glu Gly Lys Pro Gln Glu Glu Pro
1 5 10 15
Thr Glu Leu Ala Lys Leu Leu Arg Glu His Tyr Pro Gly Lys Lys Leu
20 25 30
Ser Lys Lys Asp Leu Leu Met Ala Gly Lys Ile Leu Ala Gly Lys Val
35 40 45
Ala Lys Thr Arg Pro Glu Gly Leu Val Glu Trp Leu Ala Asp Lys Val
50 55 60
Val Asp Glu Pro Pro Asn Phe Ser Pro Pro Ala Lys Ala Asn Ile Val
65 70 75 80
Ala Met Ser Arg Pro Phe Glu Glu Trp Pro Ile Ala Lys Ala Thr Leu
85 90 95
Ala Ile Glu Ser Tyr Val Phe Gly Met Thr Val Asp Glu Arg Asn Arg
100 105 110
Leu Cys Pro Lys Glu Thr Glu Glu Asp Arg Asp Glu Trp Phe Arg Val
115 120 125
Thr Gly Val Ser Asn Tyr Gly Phe Thr Ser Ala Gln Gly Leu Asn His
130 135 140
Ile Phe Lys Asn Ala Phe Asn Thr Phe Asp Gly Val Val Thr Arg Gly
145 150 155 160
Ser Asn Ala Asn Glu Lys Lys Arg Lys Glu Ile Glu Ala Gln Asn Glu
165 170 175
Lys Arg Ala Glu Arg Gly Glu Ala Pro Val Pro Phe Glu Pro Arg Pro
180 185 190

[0045]

```
Val Leu Thr Glu Asp Gly His Leu Val His Pro Pro Gly Ser Lys Ser
        195                 200                 205
Gly Leu Arg Leu Asn Lys Ile Gln Leu Tyr Gln Gln Val Thr Asp Lys
    210                 215                 220
Gly Arg Gly Phe Arg Gly Gln Val Glu Leu Pro Leu Glu Tyr Glu Asp
225                 230                 235                 240
Tyr Val Arg Asp Pro Glu Ala Pro Ile Pro Phe Gly Val Pro Arg Asp
                245                 250                 255
Arg Leu Cys Ile Pro Glu Gly Glu Pro Gly Tyr Val Pro Glu Trp Gln
            260                 265                 270
Arg Pro Leu Leu Ser Thr Lys Lys Lys Arg Arg Arg Arg Gly Trp Gly
        275                 280                 285
Pro Ala Gly Pro Glu Gln Val Arg Ala Lys Ala Lys Ala Ala Leu Leu
    290                 295                 300
Trp Tyr Leu Pro Leu Gly Asp Asp Trp Val Val Leu Asp Val Arg Gly
305                 310                 315                 320
Leu Leu Arg Asn Val Arg Trp Arg Gly Leu Ala Pro Glu Gly Leu Ser
                325                 330                 335
Leu Asn Gly Leu Met Glu Leu Phe Thr Gly Tyr Pro Ile Ile His Asn
            340                 345                 350
Lys Thr Gly Asp Val Thr Phe Lys Phe Cys Pro Glu Val Ala Gly Val
        355                 360                 365
Arg Ser His Glu Pro Leu Lys Lys Ala Glu Gly Arg Lys Leu Leu Leu
    370                 375                 380
His Leu Thr Lys Pro Arg Gly Glu Phe His Pro Arg Val Gly Met Val
385                 390                 395                 400
Ala Ile Asp Leu Gly Gln Thr Asn Pro Ala Ala Phe Ser Val Ser Arg
                405                 410                 415
Leu His Gln Val Glu Val Glu Arg Glu Val Glu Val Lys Arg Lys Leu
            420                 425                 430
Pro Asp Gly Glu Thr Glu Lys Val Thr Glu Leu Arg Met Val Gly Glu
        435                 440                 445
Val Arg Gln Thr Ile Leu Ser Thr His Phe Leu Pro Asp Glu Leu Val
    450                 455                 460
Ala Glu Ile Lys Arg Tyr Arg Asn Leu Tyr Ser Ala Met Asn Glu Arg
465                 470                 475                 480
His His Ala Glu Ala Val Leu Arg Leu Pro Lys Glu Ala Gln Asp Val
                485                 490                 495
Tyr His Ala Trp Gln Asn Phe Ser Ala Asp Glu Ala Lys Arg Gln Leu
            500                 505                 510
Cys Ala Lys Tyr Gly Leu Asp Pro Ala Lys Met Pro Trp Asp Gln Met
        515                 520                 525
Thr Ser Ser Thr Met His Ile Ser Lys Ala Val Leu Ala Ala Gly His
    530                 535                 540
```

[0046]

Asp Arg Ser Glu Val His Arg Met Val Lys Lys Lys Gly Lys Lys Glu
545 550 555 560
Arg Glu Val Lys Lys Arg Asp Arg Ala Trp Val Ser Asp Phe Lys Pro
565 570 575
Thr Leu Pro Glu Glu Val Trp Arg Pro Lys Arg Glu His Leu Leu Asp
580 585 590
Ile Gln Arg Glu Ser Glu Glu Tyr Arg Arg Leu Thr Ile Phe Lys Gln
595 600 605
Gln Leu Val Arg Arg Cys Val Asn His Val Val Lys Val Gly Glu Arg
610 615 620
Glu Ser Gln Cys Glu Lys Leu Val Phe Ala Ile Glu Asp Leu Asn Val
625 630 635 640
Lys Gly Phe Phe Phe Gly Lys Gly Lys Asn Leu Pro Gly Trp Glu Gly
645 650 655
Phe Phe Gln His Lys Arg Glu Asn Arg Phe Ile Ile Arg Glu Phe His
660 665 670
Arg Ala Phe Ser Glu Leu Gly Pro His Arg Gly Tyr Tyr Val Leu Glu
675 680 685
Ala Asn Pro Gly Tyr Thr Ser Cys Thr Cys Pro Asp Cys Arg His Pro
690 695 700
Asp Pro Val Ser Arg Asn Gly Glu Arg Phe Lys Cys Thr Arg Cys Gly
705 710 715 720
Ala Thr His His Ala Asp Ser Glu Val Ala Thr Tyr Asn Ile Ala Gln
725 730 735
Val Ala Ile Met Gly Lys Ala Leu Pro Arg Pro Lys Lys Gln Lys Lys
740 745 750
Pro Lys Arg Glu Arg Ser Gly Ala Val Lys Lys Ala Glu Thr Ala Arg
755 760 765
Lys Arg Asn Gly Arg Lys Ser Asn Gly Lys Gly Gly Gln Arg Gln Glu
770 775 780
Ala Pro Leu Leu Arg Pro Pro Val Arg Gly Thr Ala Arg Glu Pro Val
785 790 795 800
Ala Asn Ala Ser Cys
805

<210> 20
<211> 801
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
土壤宏基因组序列
<400> 20
Met Thr Ile Thr Glu Gly Lys Pro Gln Glu Glu Pro Thr Glu Leu Ala
1 5 10 15

[0047]

```
Lys Leu Leu Arg Glu His Tyr Pro Gly Lys Lys Leu Ser Lys Lys Asp
            20                  25                  30
Leu Leu Met Ala Gly Lys Ile Leu Ala Gly Lys Val Ala Lys Thr Arg
        35                  40                  45
Pro Glu Gly Leu Val Glu Trp Leu Ala Asp Lys Val Val Asp Glu Pro
    50                  55                  60
Pro Asn Phe Ser Pro Pro Ala Lys Ala Asn Ile Val Ala Met Ser Arg
65                  70                  75                  80
Pro Phe Glu Glu Trp Pro Ile Ala Lys Ala Thr Leu Ala Ile Glu Ser
                85                  90                  95
Tyr Val Phe Gly Met Thr Val Asp Glu Arg Asn Arg Leu Cys Pro Lys
            100                 105                 110
Glu Thr Glu Glu Asp Arg Asp Glu Trp Phe Arg Val Thr Gly Val Ser
        115                 120                 125
Asn Tyr Gly Phe Thr Ser Ala Gln Gly Leu Asn His Ile Phe Lys Asn
    130                 135                 140
Ala Phe Asn Thr Phe Asp Gly Val Val Thr Arg Gly Ser Asn Ala Asn
145                 150                 155                 160
Glu Lys Lys Arg Lys Glu Ile Glu Ala Gln Asn Glu Lys Arg Ala Glu
                165                 170                 175
Arg Gly Glu Ala Pro Val Pro Phe Glu Pro Arg Pro Val Leu Thr Glu
            180                 185                 190
Asp Gly His Leu Val His Pro Pro Gly Ser Lys Ser Gly Leu Arg Leu
        195                 200                 205
Asn Lys Ile Gln Leu Tyr Gln Gln Val Thr Asp Lys Gly Arg Gly Phe
    210                 215                 220
Arg Gly Gln Val Glu Leu Pro Leu Glu Tyr Glu Asp Tyr Val Arg Asp
225                 230                 235                 240
Pro Glu Ala Pro Ile Pro Phe Gly Val Pro Arg Asp Arg Leu Cys Ile
                245                 250                 255
Pro Glu Gly Glu Pro Gly Tyr Val Pro Glu Trp Gln Arg Pro Leu Leu
            260                 265                 270
Ser Thr Lys Lys Lys Arg Arg Arg Arg Gly Trp Gly Pro Ala Gly Pro
        275                 280                 285
Glu Gln Val Arg Ala Lys Ala Lys Ala Ala Leu Leu Trp Tyr Leu Pro
    290                 295                 300
Leu Gly Asp Asp Trp Val Val Leu Asp Val Arg Gly Leu Leu Arg Asn
305                 310                 315                 320
Val Arg Trp Arg Gly Leu Ala Pro Glu Gly Leu Ser Leu Asn Gly Leu
                325                 330                 335
Met Glu Leu Phe Thr Gly Tyr Pro Ile Ile His Asn Lys Thr Gly Asp
            340                 345                 350
Val Thr Phe Lys Phe Cys Pro Glu Val Ala Gly Val Arg Ser His Glu
        355                 360                 365
```

[0048]

```
Pro Leu Lys Lys Ala Glu Gly Arg Lys Leu Leu Leu His Leu Thr Lys
    370                 375                 380
Pro Arg Gly Glu Phe His Pro Arg Val Gly Met Val Ala Ile Asp Leu
385                 390                 395                 400
Gly Gln Thr Asn Pro Ala Ala Phe Ser Val Ser Arg Leu His Gln Val
                405                 410                 415
Glu Val Glu Arg Glu Val Glu Val Lys Arg Lys Leu Pro Asp Gly Glu
            420                 425                 430
Thr Glu Lys Val Thr Glu Leu Arg Met Val Gly Glu Val Arg Gln Thr
        435                 440                 445
Ile Leu Ser Thr His Phe Leu Pro Asp Glu Leu Val Ala Glu Ile Lys
    450                 455                 460
Arg Tyr Arg Asn Leu Tyr Ser Ala Met Asn Glu Arg His His Ala Glu
465                 470                 475                 480
Ala Val Leu Arg Leu Pro Lys Glu Ala Gln Asp Val Tyr His Ala Trp
                485                 490                 495
Gln Asn Phe Ser Ala Asp Glu Ala Lys Arg Gln Leu Cys Ala Lys Tyr
            500                 505                 510
Gly Leu Asp Pro Ala Lys Met Pro Trp Asp Gln Met Thr Ser Ser Thr
        515                 520                 525
Met His Ile Ser Lys Ala Val Leu Ala Ala Gly His Asp Arg Ser Glu
    530                 535                 540
Val His Arg Met Val Lys Lys Lys Gly Lys Lys Glu Arg Glu Val Lys
545                 550                 555                 560
Lys Arg Asp Arg Ala Trp Val Ser Asp Phe Lys Pro Thr Leu Pro Glu
                565                 570                 575
Glu Val Trp Arg Pro Lys Arg Glu His Leu Leu Asp Ile Gln Arg Glu
            580                 585                 590
Ser Glu Glu Tyr Arg Arg Leu Thr Ile Phe Lys Gln Gln Leu Val Arg
        595                 600                 605
Arg Cys Val Asn His Val Val Lys Val Gly Glu Arg Glu Ser Gln Cys
    610                 615                 620
Glu Lys Leu Val Phe Ala Ile Glu Asp Leu Asn Val Lys Gly Phe Phe
625                 630                 635                 640
Phe Gly Lys Gly Lys Asn Leu Pro Gly Trp Glu Gly Phe Phe Gln His
                645                 650                 655
Lys Arg Glu Asn Arg Phe Ile Ile Arg Glu Phe His Arg Ala Phe Ser
            660                 665                 670
Glu Leu Gly Pro His Arg Gly Tyr Tyr Val Leu Glu Ala Asn Pro Gly
        675                 680                 685
Tyr Thr Ser Cys Thr Cys Pro Asp Cys Arg His Pro Asp Pro Val Ser
    690                 695                 700
Arg Asn Gly Glu Arg Phe Lys Cys Thr Arg Cys Gly Ala Thr His His
705                 710                 715                 720
```

Ala Asp Ser Glu Val Ala Thr Tyr Asn Ile Ala Gln Val Ala Ile Met
725 730 735
Gly Lys Ala Leu Pro Arg Pro Lys Lys Gln Lys Lys Pro Lys Arg Glu
740 745 750
Arg Ser Gly Ala Val Lys Lys Ala Glu Thr Ala Arg Lys Arg Asn Gly
755 760 765
Arg Lys Ser Asn Gly Lys Gly Gly Gln Arg Gln Glu Ala Pro Leu Leu
770 775 780
Arg Pro Pro Val Arg Gly Thr Ala Arg Glu Pro Val Ala Asn Ala Ser
785 790 795 800
Cys

<210> 21
<211> 717
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
陆生-土壤-森林土壤序列
<400> 21

[0049]

Met Ile Lys Pro Thr Val Ser Gln Phe Leu Thr Pro Gly Phe Lys Leu
1 5 10 15
Ile Arg Asn His Ser Arg Thr Ala Gly Leu Lys Leu Lys Asn Glu Gly
20 25 30
Glu Glu Ala Cys Lys Lys Phe Val Arg Glu Asn Glu Ile Pro Lys Asp
35 40 45
Glu Cys Pro Asn Phe Gln Gly Gly Pro Ala Ile Ala Asn Ile Ile Ala
50 55 60
Lys Ser Arg Glu Phe Thr Glu Trp Glu Ile Tyr Gln Ser Ser Leu Ala
65 70 75 80
Ile Gln Glu Val Ile Phe Thr Leu Pro Lys Asp Lys Leu Pro Glu Pro
85 90 95
Ile Leu Lys Glu Glu Trp Arg Ala Gln Trp Leu Ser Glu His Gly Leu
100 105 110
Asp Thr Val Pro Tyr Lys Glu Ala Ala Gly Leu Asn Leu Ile Ile Lys
115 120 125
Asn Ala Val Asn Thr Tyr Lys Gly Val Gln Val Lys Val Asp Asn Lys
130 135 140
Asn Lys Asn Asn Leu Ala Lys Ile Asn Arg Lys Asn Glu Ile Ala Lys
145 150 155 160
Leu Asn Gly Glu Gln Glu Ile Ser Phe Glu Glu Ile Lys Ala Phe Asp
165 170 175
Asp Lys Gly Tyr Leu Leu Gln Lys Pro Ser Pro Asn Lys Ser Ile Tyr
180 185 190

[0050]

```
Cys Tyr Gln Ser Val Ser Pro Lys Pro Phe Ile Thr Ser Lys Tyr His
        195                 200                 205
Asn Val Asn Leu Pro Glu Glu Tyr Ile Gly Tyr Tyr Arg Lys Ser Asn
    210                 215                 220
Glu Pro Ile Val Ser Pro Tyr Gln Phe Asp Arg Leu Arg Ile Pro Ile
225                 230                 235                 240
Gly Glu Pro Gly Tyr Val Pro Lys Trp Gln Tyr Thr Phe Leu Ser Lys
                245                 250                 255
Lys Glu Asn Lys Arg Arg Lys Leu Ser Lys Arg Ile Lys Asn Val Ser
            260                 265                 270
Pro Ile Leu Gly Ile Ile Cys Ile Lys Lys Asp Trp Cys Val Phe Asp
        275                 280                 285
Met Arg Gly Leu Leu Arg Thr Asn His Trp Lys Lys Tyr His Lys Pro
    290                 295                 300
Thr Asp Ser Ile Asn Asp Leu Phe Asp Tyr Phe Thr Gly Asp Pro Val
305                 310                 315                 320
Ile Asp Thr Lys Ala Asn Val Val Arg Phe Arg Tyr Lys Met Glu Asn
                325                 330                 335
Gly Ile Val Asn Tyr Lys Pro Val Arg Glu Lys Lys Gly Lys Glu Leu
            340                 345                 350
Leu Glu Asn Ile Cys Asp Gln Asn Gly Ser Cys Lys Leu Ala Thr Val
        355                 360                 365
Asp Val Gly Gln Asn Asn Pro Val Ala Ile Gly Leu Phe Glu Leu Lys
    370                 375                 380
Lys Val Asn Gly Glu Leu Thr Lys Thr Leu Ile Ser Arg His Pro Thr
385                 390                 395                 400
Pro Ile Asp Phe Cys Asn Lys Ile Thr Ala Tyr Arg Glu Arg Tyr Asp
                405                 410                 415
Lys Leu Glu Ser Ser Ile Lys Leu Asp Ala Ile Lys Gln Leu Thr Ser
            420                 425                 430
Glu Gln Lys Ile Glu Val Asp Asn Tyr Asn Asn Asn Phe Thr Pro Gln
        435                 440                 445
Asn Thr Lys Gln Ile Val Cys Ser Lys Leu Asn Ile Asn Pro Asn Asp
    450                 455                 460
Leu Pro Trp Asp Lys Met Ile Ser Gly Thr His Phe Ile Ser Glu Lys
465                 470                 475                 480
Ala Gln Val Ser Asn Lys Ser Glu Ile Tyr Phe Thr Ser Thr Asp Lys
                485                 490                 495
Gly Lys Thr Lys Asp Val Met Lys Ser Asp Tyr Lys Trp Phe Gln Asp
            500                 505                 510
Tyr Lys Pro Lys Leu Ser Lys Glu Val Arg Asp Ala Leu Ser Asp Ile
        515                 520                 525
Glu Trp Arg Leu Arg Arg Glu Ser Leu Glu Phe Asn Lys Leu Ser Lys
    530                 535                 540
```

Ser Arg Glu Gln Asp Ala Arg Gln Leu Ala Asn Trp Ile Ser Ser Met
545 550 555 560
Cys Asp Val Ile Gly Ile Glu Asn Leu Val Lys Lys Asn Asn Phe Phe
565 570 575
Gly Gly Ser Gly Lys Arg Glu Pro Gly Trp Asp Asn Phe Tyr Lys Pro
580 585 590
Lys Lys Glu Asn Arg Trp Trp Ile Asn Ala Ile His Lys Ala Leu Thr
595 600 605
Glu Leu Ser Gln Asn Lys Gly Lys Arg Val Ile Leu Leu Pro Ala Met
610 615 620
Arg Thr Ser Ile Thr Cys Pro Lys Cys Lys Tyr Cys Asp Ser Lys Asn
625 630 635 640
Arg Asn Gly Glu Lys Phe Asn Cys Leu Lys Cys Gly Ile Glu Leu Asn
645 650 655
Ala Asp Ile Asp Val Ala Thr Glu Asn Leu Ala Thr Val Ala Ile Thr
660 665 670
Ala Gln Ser Met Pro Lys Pro Thr Cys Glu Arg Ser Gly Asp Ala Lys
675 680 685
Lys Pro Val Arg Ala Arg Lys Ala Lys Ala Pro Glu Phe His Asp Lys
690 695 700
Leu Ala Pro Ser Tyr Thr Val Val Leu Arg Glu Ala Val
705 710 715

[0051]

<210> 22
<211> 717
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
陆生-土壤-森林土壤序列
<400> 22
Met Ile Lys Pro Thr Val Ser Gln Phe Leu Thr Pro Gly Phe Lys Leu
1 5 10 15
Ile Arg Asn His Ser Arg Thr Ala Gly Leu Lys Leu Lys Asn Glu Gly
20 25 30
Glu Glu Ala Cys Lys Lys Phe Val Arg Glu Asn Glu Ile Pro Lys Asp
35 40 45
Glu Cys Pro Asn Phe Gln Gly Gly Pro Ala Ile Ala Asn Ile Ile Ala
50 55 60
Lys Ser Arg Glu Phe Thr Glu Trp Glu Ile Tyr Gln Ser Ser Leu Ala
65 70 75 80
Ile Gln Glu Val Ile Phe Thr Leu Pro Lys Asp Lys Leu Pro Glu Pro
85 90 95
Ile Leu Lys Glu Glu Trp Arg Ala Gln Trp Leu Ser Glu His Gly Leu
100 105 110

[0052]

```
Asp Thr Val Pro Tyr Lys Glu Ala Ala Gly Leu Asn Leu Ile Ile Lys
        115                 120                 125
Asn Ala Val Asn Thr Tyr Lys Gly Val Gln Val Lys Val Asp Asn Lys
    130                 135                 140
Asn Lys Asn Asn Leu Ala Lys Ile Asn Arg Lys Asn Glu Ile Ala Lys
145                 150                 155                 160
Leu Asn Gly Glu Gln Glu Ile Ser Phe Glu Glu Ile Lys Ala Phe Asp
                165                 170                 175
Asp Lys Gly Tyr Leu Leu Gln Lys Pro Ser Pro Asn Lys Ser Ile Tyr
            180                 185                 190
Cys Tyr Gln Ser Val Ser Pro Lys Pro Phe Ile Thr Ser Lys Tyr His
        195                 200                 205
Asn Val Asn Leu Pro Glu Glu Tyr Ile Gly Tyr Tyr Arg Lys Ser Asn
    210                 215                 220
Glu Pro Ile Val Ser Pro Tyr Gln Phe Asp Arg Leu Arg Ile Pro Ile
225                 230                 235                 240
Gly Glu Pro Gly Tyr Val Pro Lys Trp Gln Tyr Thr Phe Leu Ser Lys
                245                 250                 255
Lys Glu Asn Lys Arg Arg Lys Leu Ser Lys Arg Ile Lys Asn Val Ser
            260                 265                 270
Pro Ile Leu Gly Ile Ile Cys Ile Lys Lys Asp Trp Cys Val Phe Asp
        275                 280                 285
Met Arg Gly Leu Leu Arg Thr Asn His Trp Lys Lys Tyr His Lys Pro
    290                 295                 300
Thr Asp Ser Ile Asn Asp Leu Phe Asp Tyr Phe Thr Gly Asp Pro Val
305                 310                 315                 320
Ile Asp Thr Lys Ala Asn Val Val Arg Phe Arg Tyr Lys Met Glu Asn
                325                 330                 335
Gly Ile Val Asn Tyr Lys Pro Val Arg Glu Lys Lys Gly Lys Glu Leu
            340                 345                 350
Leu Glu Asn Ile Cys Asp Gln Asn Gly Ser Cys Lys Leu Ala Thr Val
        355                 360                 365
Asp Val Gly Gln Asn Asn Pro Val Ala Ile Gly Leu Phe Glu Leu Lys
    370                 375                 380
Lys Val Asn Gly Glu Leu Thr Lys Thr Leu Ile Ser Arg His Pro Thr
385                 390                 395                 400
Pro Ile Asp Phe Cys Asn Lys Ile Thr Ala Tyr Arg Glu Arg Tyr Asp
                405                 410                 415
Lys Leu Glu Ser Ser Ile Lys Leu Asp Ala Ile Lys Gln Leu Thr Ser
            420                 425                 430
Glu Gln Lys Ile Glu Val Asp Asn Tyr Asn Asn Asn Phe Thr Pro Gln
        435                 440                 445
Asn Thr Lys Gln Ile Val Cys Ser Lys Leu Asn Ile Asn Pro Asn Asp
    450                 455                 460
```

[0053]

Leu Pro Trp Asp Lys Met Ile Ser Gly Thr His Phe Ile Ser Glu Lys
465 470 475 480
Ala Gln Val Ser Asn Lys Ser Glu Ile Tyr Phe Thr Ser Thr Asp Lys
485 490 495
Gly Lys Thr Lys Asp Val Met Lys Ser Asp Tyr Lys Trp Phe Gln Asp
500 505 510
Tyr Lys Pro Lys Leu Ser Lys Glu Val Arg Asp Ala Leu Ser Asp Ile
515 520 525
Glu Trp Arg Leu Arg Arg Glu Ser Leu Glu Phe Asn Lys Leu Ser Lys
530 535 540
Ser Arg Glu Gln Asp Ala Arg Gln Leu Ala Asn Trp Ile Ser Ser Met
545 550 555 560
Cys Asp Val Ile Gly Ile Glu Asn Leu Val Lys Lys Asn Asn Phe Phe
565 570 575
Gly Gly Ser Gly Lys Arg Glu Pro Gly Trp Asp Asn Phe Tyr Lys Pro
580 585 590
Lys Lys Glu Asn Arg Trp Trp Ile Asn Ala Ile His Lys Ala Leu Thr
595 600 605
Glu Leu Ser Gln Asn Lys Gly Lys Arg Val Ile Leu Leu Pro Ala Met
610 615 620
Arg Thr Ser Ile Thr Cys Pro Lys Cys Lys Tyr Cys Asp Ser Lys Asn
625 630 635 640
Arg Asn Gly Glu Lys Phe Asn Cys Leu Lys Cys Gly Ile Glu Leu Asn
645 650 655
Ala Asp Ile Asp Val Ala Thr Glu Asn Leu Ala Thr Val Ala Ile Thr
660 665 670
Ala Gln Ser Met Pro Lys Pro Thr Cys Glu Arg Ser Gly Asp Ala Lys
675 680 685
Lys Pro Val Arg Ala Arg Lys Ala Lys Ala Pro Glu Phe His Asp Lys
690 695 700
Leu Ala Pro Ser Tyr Thr Val Val Leu Arg Glu Ala Val
705 710 715

<210> 23
<211> 749
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
废水宏基因组序列
<400> 23

Met Pro Asp Lys Gln Thr Pro Lys Asp Thr Lys Asp Lys Pro Glu Ser
1 5 10 15
Pro Val Ser Ala Phe Leu Lys Lys His Phe Pro Gly Lys His Phe Phe
20 25 30

[0054]

```
Gly His Ala Gly Thr Leu Ala Arg Leu Leu Lys Thr Lys Gly Glu Glu
        35                  40                  45
Val Ala Arg Asp Tyr Ala Ala Lys Lys Val Arg Asp Glu Lys Leu Asp
    50                  55                  60
Phe Arg Pro Pro Ala Lys Cys Gln Ile Val Ala Trp Ser Arg Asp Phe
65                  70                  75                  80
Ser Glu Trp Pro Ile Ala Arg Ala Ser Ala Thr Ile Gln Gln His Val
                85                  90                  95
Ser Gly Leu Thr Lys Glu Asp Phe Glu Arg Phe Asp Pro Gly Lys Ser
            100                 105                 110
Lys Ala Ala His Asp Ala Trp Phe Gln Glu Ser Gly Val Asp Cys His
        115                 120                 125
Gly Tyr Arg His Val Gln Gly Leu Asn Leu Ile Phe Ala Asn Ala Arg
    130                 135                 140
Asp Tyr Tyr Glu Gly Val Val Lys Lys Val Glu Asn Lys Asn Ala Gln
145                 150                 155                 160
Arg Arg Arg Arg Val Glu Ala Leu Asn Ala Arg Arg Ala Glu Glu Gly
                165                 170                 175
Glu Glu Pro Ile Pro Leu Asp Val Glu Glu Ser Pro Phe Gly Glu Asp
            180                 185                 190
Gly Arg Leu Ala His Pro Pro Gly Val Asn Pro Ser Ile Tyr Val Tyr
        195                 200                 205
Gln Ala Val Ser Pro Arg Pro Leu Lys Lys Ser Asp Leu Glu Thr Val
    210                 215                 220
Val Leu Pro Pro Ala Tyr Ala Gly Tyr Asp Arg Asp Pro Ser Ala Pro
225                 230                 235                 240
Ile Pro Val Met Gly Asp Arg Leu Ser Ile Pro Glu Gly Gln Arg Gly
                245                 250                 255
His Val Pro Ala Trp Gln Arg Asp Gln Leu Ser Pro Asp Lys His Arg
            260                 265                 270
Arg Met Arg Ala Trp Tyr Ser Ala Ala Asn Thr Lys Pro Lys Pro Gly
        275                 280                 285
Arg Thr Ser Val Pro Asp Ala Ala Ala Ile Glu Arg Ala Arg Ala Glu
    290                 295                 300
Gly Ala Leu Leu Val Val Ile Arg Ile Gly Glu Asp Trp Val Val Leu
305                 310                 315                 320
Asp Ala Arg Gly Leu Leu Arg Asn Ala Arg Trp Arg Arg Ile Ala Asp
                325                 330                 335
Lys Glu Ile Ser Leu Asp Gly Leu Leu Asp Leu Phe Thr Gly Asp Pro
            340                 345                 350
Val Ile Asp Ser Lys Arg Asn Val Val Thr Phe Ile Tyr Lys Ala Glu
        355                 360                 365
His Ala Thr Ala Thr Ser Arg Lys Val Val His Arg Lys Ala Ser Arg
    370                 375                 380
```

[0055]

```
Lys Ala Leu Leu Asp Met Thr Ser Pro Gly Glu Asp Gly Leu Pro Arg
385                 390                 395                 400
Glu Val Ala Leu Ala Ser Val Asp Leu Gly Gln Thr Asn Ala Ala Ala
                405                 410                 415
Val Arg Tyr Ala Arg Val His Arg Glu Gly Asp Asp Ile Thr Ser Glu
            420                 425                 430
Cys Leu Val Arg Glu Leu Leu Pro Asp Glu Ile Ser Arg Asp Ile Ala
        435                 440                 445
Arg Tyr Arg Ala Ala Ser Asp Arg Met Glu Ala Glu Ile Arg Glu Ala
    450                 455                 460
Ala Ile Ala Gly Leu Pro Glu Pro Met Gln Ala Glu Val Arg Ala Ala
465                 470                 475                 480
Asp Ala Ser Ser Pro Glu Ala Ala Arg Ala Ala Val Val Ala Leu Val
                485                 490                 495
Gly Asp Gly Leu Pro Trp Glu Lys Met Ser Ser Ala Thr Tyr His Ile
            500                 505                 510
Ser Asp Ala Leu Val Ala Leu Gly Arg Gly Arg Glu Ala Tyr Leu Leu
        515                 520                 525
Ser Lys Ser Lys Asp Gly Glu Glu Lys Ser Val Gln Arg Ser Asp Tyr
    530                 535                 540
Gly Trp Ser Arg His Leu Arg Pro Arg Leu Ser Glu Glu Thr Arg Lys
545                 550                 555                 560
Ala Met Asn Glu Ala Val Trp Ser Ile Lys Asp Ala His Glu Gly Tyr
                565                 570                 575
Gln Lys Leu Ser Arg Arg Lys Thr Glu Ile Gly Arg Arg Ala Ala Asn
            580                 585                 590
His Val Val Arg Arg Leu Arg Lys Leu Ala Lys Thr Asp Lys Val Ala
        595                 600                 605
Ile Ala Val Glu Asp Leu Asn Val Arg Met Phe His Gly Gly Gly Ser
    610                 615                 620
Arg Ser Thr Gly Trp Asp Asn Phe Phe Val Ala Lys Arg Glu Asn Leu
625                 630                 635                 640
Trp Phe Val Gln Val Leu His Lys Ser Phe Cys Asp Leu Ala Leu His
                645                 650                 655
Arg Gly Glu Val Val Ile Glu Val Asp Thr Ala Arg Thr Ser Gln Thr
            660                 665                 670
Cys Pro Ala Cys Gly His Cys Asp Pro Lys Asn Arg Ser Ser Val Asp
        675                 680                 685
Arg Glu Val Phe Arg Cys Val Val Cys Gly Arg Thr Phe His Ala Asp
    690                 695                 700
Leu Glu Val Ala Thr Phe Asn Ile Glu Arg Val Ala Leu Thr Gly Glu
705                 710                 715                 720
Ser Met Pro Lys Gly Glu Glu Gly Ala Arg Glu Arg Gly Gly Gly Gly
                725                 730                 735
```

Lys Ser Arg Gly Gly Ala Arg Gly Arg Asn Lys Leu Lys
740 745

<210> 24
<211> 773
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
废水宏基因组序列
<400> 24

[0056]

Met Ser Glu Ile Thr Asp Leu Leu Lys Ala Asn Phe Lys Gly Lys Thr
1 5 10 15
Phe Lys Ser Ala Asp Met Arg Met Ala Gly Arg Ile Leu Lys Lys Ser
20 25 30
Gly Ala Gln Ala Val Ile Lys Tyr Leu Ser Asp Lys Gly Ala Val Asp
35 40 45
Pro Pro Asp Phe Arg Pro Pro Ala Lys Cys Asn Ile Ile Ala Gln Ser
50 55 60
Arg Pro Phe Asp Glu Trp Pro Ile Cys Lys Ala Ser Met Ala Ile Gln
65 70 75 80
Gln His Ile Tyr Gly Leu Thr Lys Asn Glu Phe Asp Glu Ser Ser Pro
85 90 95
Gly Thr Ser Ser Ala Ser His Glu Gln Trp Phe Ala Lys Thr Gly Val
100 105 110
Asp Thr His Gly Phe Thr His Val Gln Gly Leu Asn Leu Ile Phe Gln
115 120 125
His Ala Lys Lys Arg Tyr Glu Gly Val Ile Lys Lys Val Glu Asn Tyr
130 135 140
Asn Glu Lys Glu Arg Lys Lys Phe Glu Gly Ile Asn Glu Arg Arg Ser
145 150 155 160
Lys Glu Gly Met Pro Leu Leu Glu Pro Arg Leu Arg Thr Ala Phe Gly
165 170 175
Asp Asp Gly Lys Phe Ala Glu Lys Pro Gly Val Asn Pro Ser Ile Tyr
180 185 190
Leu Tyr Gln Gln Thr Ser Pro Arg Pro Tyr Asp Lys Thr Lys His Pro
195 200 205
Tyr Val His Ala Pro Phe Glu Leu Lys Glu Ile Thr Thr Ile Pro Thr
210 215 220
Gln Asp Asp Arg Leu Lys Ile Pro Phe Gly Ala Pro Gly His Val Pro
225 230 235 240
Glu Lys His Arg Ser Gln Leu Ser Met Ala Lys His Lys Arg Arg Arg
245 250 255
Ala Trp Tyr Ala Leu Ser Gln Asn Lys Pro Arg Pro Pro Lys Asp Gly
260 265 270

[0057]

```
Ser Lys Gly Arg Arg Ser Val Arg Asp Leu Ala Asp Leu Lys Ala Ala
        275                 280                 285
Ser Leu Ala Asp Ala Ile Pro Leu Val Ser Arg Val Gly Phe Asp Trp
    290                 295                 300
Val Val Ile Asp Gly Arg Gly Leu Leu Arg Asn Leu Arg Trp Arg Lys
305                 310                 315                 320
Leu Ala His Glu Gly Met Thr Val Glu Glu Met Leu Gly Phe Phe Ser
                325                 330                 335
Gly Asp Pro Val Ile Asp Pro Arg Arg Asn Val Ala Thr Phe Ile Tyr
            340                 345                 350
Lys Ala Glu His Ala Thr Val Lys Ser Arg Lys Pro Ile Gly Gly Ala
        355                 360                 365
Lys Arg Ala Arg Glu Glu Leu Leu Lys Ala Thr Ala Ser Ser Asp Gly
    370                 375                 380
Val Ile Arg Gln Val Gly Leu Ile Ser Val Asp Leu Gly Gln Thr Asn
385                 390                 395                 400
Pro Val Ala Tyr Glu Ile Ser Arg Met His Gln Ala Asn Gly Glu Leu
                405                 410                 415
Val Ala Glu His Leu Glu Tyr Gly Leu Leu Asn Asp Glu Gln Val Asn
            420                 425                 430
Ser Ile Gln Arg Tyr Arg Ala Ala Trp Asp Ser Met Asn Glu Ser Phe
        435                 440                 445
Arg Gln Lys Ala Ile Glu Ser Leu Ser Met Glu Ala Gln Asp Glu Ile
    450                 455                 460
Met Gln Ala Ser Thr Gly Ala Ala Lys Arg Thr Arg Glu Ala Val Leu
465                 470                 475                 480
Thr Met Phe Gly Pro Asn Ala Thr Leu Pro Trp Ser Arg Met Ser Ser
                485                 490                 495
Asn Thr Thr Cys Ile Ser Asp Ala Leu Ile Glu Val Gly Lys Glu Glu
            500                 505                 510
Glu Thr Asn Phe Val Thr Ser Asn Gly Pro Arg Lys Arg Thr Asp Ala
        515                 520                 525
Gln Trp Ala Ala Tyr Leu Arg Pro Arg Val Asn Pro Glu Thr Arg Ala
    530                 535                 540
Leu Leu Asn Gln Ala Val Trp Asp Leu Met Lys Arg Ser Asp Glu Tyr
545                 550                 555                 560
Glu Arg Leu Ser Lys Arg Lys Leu Glu Met Ala Arg Gln Cys Val Asn
                565                 570                 575
Phe Val Val Ala Arg Ala Glu Lys Leu Thr Gln Cys Asn Asn Ile Gly
            580                 585                 590
Ile Val Leu Glu Asn Leu Val Val Arg Asn Phe His Gly Ser Gly Arg
        595                 600                 605
Arg Glu Ser Gly Trp Glu Gly Phe Phe Glu Pro Lys Arg Glu Asn Arg
    610                 615                 620
```

Trp Phe Met Gln Val Leu His Lys Ala Phe Ser Asp Leu Ala Gln His
625 630 635 640
Arg Gly Val Met Val Phe Glu Val His Pro Ala Tyr Ser Ser Gln Thr
645 650 655
Cys Pro Ala Cys Arg Tyr Val Asp Pro Lys Asn Arg Ser Ser Glu Asp
660 665 670
Arg Glu Arg Phe Lys Cys Leu Lys Cys Gly Arg Ser Phe Asn Ala Asp
675 680 685
Arg Glu Val Ala Thr Phe Asn Ile Arg Glu Ile Ala Arg Thr Gly Val
690 695 700
Gly Leu Pro Lys Pro Asp Cys Glu Arg Ser Arg Asp Val Gln Thr Pro
705 710 715 720
Gly Thr Ala Arg Lys Ser Gly Arg Ser Leu Lys Ser Asn Lys Asn Pro
725 730 735
Ser Glu Pro Lys His Val Leu Arg Ser Lys Thr Arg Ser Asn Ile Thr
740 745 750
Ser Thr Leu Ser Gln Asn Glu Pro Leu Ala Thr Asp Gln Lys Thr Ala
755 760 765
Pro Lys Thr Gly Pro
770

[0058]

<210> 25
<211> 773
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
废水宏基因组序列
<400> 25
Met Ser Glu Ile Thr Asp Leu Leu Lys Ala Asn Phe Lys Gly Lys Thr
1 5 10 15
Phe Lys Ser Ala Asp Met Arg Met Ala Gly Arg Ile Leu Lys Lys Ser
20 25 30
Gly Ala Gln Ala Val Ile Lys Tyr Leu Ser Asp Lys Gly Ala Val Asp
35 40 45
Pro Pro Asp Phe Arg Pro Pro Ala Lys Cys Asn Ile Ile Ala Gln Ser
50 55 60
Arg Pro Phe Asp Glu Trp Pro Ile Cys Lys Ala Ser Met Ala Ile Gln
65 70 75 80
Gln His Ile Tyr Gly Leu Thr Lys Asn Glu Phe Asp Glu Ser Ser Pro
85 90 95
Gly Thr Ser Ser Ala Ser His Glu Gln Trp Phe Ala Lys Thr Gly Val
100 105 110
Asp Thr His Gly Phe Thr His Val Gln Gly Leu Asn Leu Ile Phe Gln
115 120 125

[0059]

```
His Ala Lys Lys Arg Tyr Glu Gly Val Ile Lys Lys Val Glu Asn Tyr
    130                 135                 140
Asn Glu Lys Glu Arg Lys Lys Phe Glu Gly Ile Asn Glu Arg Arg Ser
145                 150                 155                 160
Lys Glu Gly Met Pro Leu Leu Glu Pro Arg Leu Arg Thr Ala Phe Gly
                165                 170                 175
Asp Asp Gly Lys Phe Ala Glu Lys Pro Gly Val Asn Pro Ser Ile Tyr
            180                 185                 190
Leu Tyr Gln Gln Thr Ser Pro Arg Pro Tyr Asp Lys Thr Lys His Pro
        195                 200                 205
Tyr Val His Ala Pro Phe Glu Leu Lys Glu Ile Thr Thr Ile Pro Thr
    210                 215                 220
Gln Asp Asp Arg Leu Lys Ile Pro Phe Gly Ala Pro Gly His Val Pro
225                 230                 235                 240
Glu Lys His Arg Ser Gln Leu Ser Met Ala Lys His Lys Arg Arg Arg
                245                 250                 255
Ala Trp Tyr Ala Leu Ser Gln Asn Lys Pro Arg Pro Pro Lys Asp Gly
            260                 265                 270
Ser Lys Gly Arg Arg Ser Val Arg Asp Leu Ala Asp Leu Lys Ala Ala
        275                 280                 285
Ser Leu Ala Asp Ala Ile Pro Leu Val Ser Arg Val Gly Phe Asp Trp
    290                 295                 300
Val Val Ile Asp Gly Arg Gly Leu Leu Arg Asn Leu Arg Trp Arg Lys
305                 310                 315                 320
Leu Ala His Glu Gly Met Thr Val Glu Glu Met Leu Gly Phe Phe Ser
                325                 330                 335
Gly Asp Pro Val Ile Asp Pro Arg Arg Asn Val Ala Thr Phe Ile Tyr
            340                 345                 350
Lys Ala Glu His Ala Thr Val Lys Ser Arg Lys Pro Ile Gly Gly Ala
        355                 360                 365
Lys Arg Ala Arg Glu Glu Leu Leu Lys Ala Thr Ala Ser Ser Asp Gly
    370                 375                 380
Val Ile Arg Gln Val Gly Leu Ile Ser Val Asp Leu Gly Gln Thr Asn
385                 390                 395                 400
Pro Val Ala Tyr Glu Ile Ser Arg Met His Gln Ala Asn Gly Glu Leu
                405                 410                 415
Val Ala Glu His Leu Glu Tyr Gly Leu Leu Asn Asp Glu Gln Val Asn
            420                 425                 430
Ser Ile Gln Arg Tyr Arg Ala Ala Trp Asp Ser Met Asn Glu Ser Phe
        435                 440                 445
Arg Gln Lys Ala Ile Glu Ser Leu Ser Met Glu Ala Gln Asp Glu Ile
    450                 455                 460
Met Gln Ala Ser Thr Gly Ala Ala Lys Arg Thr Arg Glu Ala Val Leu
465                 470                 475                 480
```

[0060]

Thr Met Phe Gly Pro Asn Ala Thr Leu Pro Trp Ser Arg Met Ser Ser
485 490 495
Asn Thr Thr Cys Ile Ser Asp Ala Leu Ile Glu Val Gly Lys Glu Glu
500 505 510
Glu Thr Asn Phe Val Thr Ser Asn Gly Pro Arg Lys Arg Thr Asp Ala
515 520 525
Gln Trp Ala Ala Tyr Leu Arg Pro Arg Val Asn Pro Glu Thr Arg Ala
530 535 540
Leu Leu Asn Gln Ala Val Trp Asp Leu Met Lys Arg Ser Asp Glu Tyr
545 550 555 560
Glu Arg Leu Ser Lys Arg Lys Leu Glu Met Ala Arg Gln Cys Val Asn
565 570 575
Phe Val Val Ala Arg Ala Glu Lys Leu Thr Gln Cys Asn Asn Ile Gly
580 585 590
Ile Val Leu Glu Asn Leu Val Val Arg Asn Phe His Gly Ser Gly Arg
595 600 605
Arg Glu Ser Gly Trp Glu Gly Phe Phe Glu Pro Lys Arg Glu Asn Arg
610 615 620
Trp Phe Met Gln Val Leu His Lys Ala Phe Ser Asp Leu Ala Gln His
625 630 635 640
Arg Gly Val Met Val Phe Glu Val His Pro Ala Tyr Ser Ser Gln Thr
645 650 655
Cys Pro Ala Cys Arg Tyr Val Asp Pro Lys Asn Arg Ser Ser Glu Asp
660 665 670
Arg Glu Arg Phe Lys Cys Leu Lys Cys Gly Arg Ser Phe Asn Ala Asp
675 680 685
Arg Glu Val Ala Thr Phe Asn Ile Arg Glu Ile Ala Arg Thr Gly Val
690 695 700
Gly Leu Pro Lys Pro Asp Cys Glu Arg Ser Arg Gly Val Gln Thr Thr
705 710 715 720
Gly Thr Ala Arg Asn Pro Gly Arg Ser Leu Lys Ser Asn Lys Asn Pro
725 730 735
Ser Glu Pro Lys His Val Leu Arg Ser Lys Thr Arg Ser Asn Ile Thr
740 745 750
Ser Thr Leu Ser Gln Asn Glu Pro Leu Ala Thr Asp Gln Lys Thr Ala
755 760 765
Pro Lys Thr Gly Pro
770

<210> 26
<211> 774
<212> PRT
<213> 未知
<220>
<223> 未知的描述:

废水宏基因组序列

<400> 26

[0061]

```
Met Thr Leu Ala Glu Leu Arg Asp Lys Tyr Phe Tyr Lys Ile Lys Phe
1               5                  10                  15
Arg Lys Ile Asp Leu Arg Gln Ala Gly Lys Ile Leu Lys Arg Glu Gly
            20                  25                  30
Glu Glu Ala Ala Arg Arg Tyr Leu Asp Glu Gln Arg Glu Ser Pro Pro
        35                  40                  45
Glu Gly Asn Phe Cys Pro Pro Ala Lys Cys Gln Ile Val Gly Trp Ser
    50                  55                  60
Arg Pro Val Gly Glu Trp Pro Ile Ser Ile Ala Ser Ala Asn Met Gln
65                  70                  75                  80
Gln Tyr Val Tyr Asp Leu Pro Lys Val Glu Arg Asp Lys Met Thr Lys
                85                  90                  95
Phe Asp Leu Thr Ser Glu Glu Tyr Ala Val Trp Phe Ala Gln Thr Gly
            100                 105                 110
Ile Asp Asn Ala Gly Tyr Thr His Val Gln Gly Leu Asn Lys Ala Phe
        115                 120                 125
Lys Asn Ala Phe Arg Thr Tyr Asp Gly Val Ile Lys Lys Val Ala Asn
    130                 135                 140
Arg Asn Glu Lys Arg Lys Leu Lys Ala Glu Lys Ala Ala Glu Arg Ala
145                 150                 155                 160
Leu Leu Arg Gly Arg Glu Pro Glu Val Phe Val Pro Glu Glu Ala Leu
                165                 170                 175
Asp Glu Arg Gly Phe Leu Lys Glu Lys Pro Gly Ile Asn Arg Ser Ile
            180                 185                 190
Trp Thr Tyr Gln Gln Val Ser Pro Arg Pro Tyr Asp Pro Thr Arg Asp
        195                 200                 205
Leu Lys Ile Lys Glu Lys Leu Ala Gln Arg Arg Gly Arg Ser Glu Pro
    210                 215                 220
Val Ala Tyr Ala Asp Arg Leu Ala Ile Pro Glu Gly Gln Pro Gly His
225                 230                 235                 240
Val Pro Gln Trp Gln Arg Asp Ala Gly Leu Leu Ser Ala Asn Lys His
                245                 250                 255
Arg Arg Met Arg Ala His Tyr Ser Trp His Asn Asn Lys Pro Arg Pro
            260                 265                 270
Asn Arg Lys Thr Ser Arg Thr Ala Glu Glu Cys Arg Asp Leu Gly Ala
        275                 280                 285
Pro Glu Ala Ile Leu Ala Val Ile Glu Ile Gly Glu Asp Trp Leu Ala
    290                 295                 300
Val Asp Leu Arg Gly Leu Leu Arg Ser Ala Tyr Tyr Arg Arg Ile Leu
305                 310                 315                 320
Ser Pro Lys Glu Val Pro Thr Ala Ala Glu Leu Leu Lys Leu Phe Thr
                325                 330                 335
```

[0062]

```
Gly Asp Pro Thr Ile Asp Pro Val Arg Glu Val Val Thr Phe Ile Tyr
            340                 345                 350
Lys Glu Asp Val Val Pro Val Leu Ser Thr Lys Pro Leu Arg Glu Arg
        355                 360                 365
Gln Gly Leu Lys Lys Ile Leu Asp Leu Thr Ala Pro Val Asn Gly Val
    370                 375                 380
Arg Asp Phe Ile Ala Ile Ala Ser Ile Asp Leu Gly Val Thr Asn Pro
385                 390                 395                 400
Ala Ala Met Ala Tyr Ser Arg Val Arg Gln Thr Ala Ala Gly Gly Phe
                405                 410                 415
Asp Ile Glu Glu Leu Ala Arg Glu Phe Leu Pro Ala Ala Val Leu Asp
            420                 425                 430
Gln Leu Ala His His Arg Gln Gln Trp Asp Glu Met Glu Asp Arg Phe
        435                 440                 445
Arg Gln Gln Ala Val Arg Ala Leu Pro Glu Ala Asp Gln Ala Glu Cys
    450                 455                 460
Glu Ala Val Phe Gly His Thr Gly Asp Gln Tyr Ala Ala Asp Ile Ala
465                 470                 475                 480
Arg Ala Leu Gln Leu Asp Gly Ala Ala Leu Pro Trp Ala Gln Met Ser
                485                 490                 495
Ser Arg Thr Thr Tyr Ile Thr Asp Ala Leu Leu Ala Arg Gly Gly Ser
            500                 505                 510
Pro Ala Thr Tyr His Thr Phe Val His Ser Glu Arg Lys Lys Gly Lys
        515                 520                 525
Lys Lys Lys Gly Lys Lys Lys Lys Gly Gln Glu Gly Ala Lys Pro Glu
    530                 535                 540
Leu Lys Ile Lys Glu Lys Pro Ser Asp Phe Asp Trp Ala Tyr Asp Phe
545                 550                 555                 560
Ala Arg Lys Gln Leu Ser Lys Glu Val Arg Glu Arg Phe Asn Lys Ala
                565                 570                 575
Leu Trp Glu Ile Lys Arg Thr Ser Pro Asp Tyr Ala Arg Met Ser Lys
            580                 585                 590
Gln Lys Arg Asp Leu Gly Arg Gln Val Ala Asn His Val Val Lys Gln
        595                 600                 605
Ala Arg Lys Leu Ala Gly Thr Gln Thr Leu Val Val Val Val Glu Asn
    610                 615                 620
Ile Asn Val Val Phe Phe His Gly Thr Gly Lys Arg Pro Val Gly Trp
625                 630                 635                 640
His Gln Cys Phe Leu Pro Lys Lys Glu Asn Arg Trp Phe Val Gln Thr
                645                 650                 655
Ile His Lys Ala Leu Thr Glu Ile Ala Met His Lys Gly Ile Tyr Val
            660                 665                 670
Ile Glu Val Ser Pro Tyr Tyr Thr Ser Leu His Cys Pro Lys Cys Glu
        675                 680                 685
```

His Ile Asp Ser Gly Asn Arg Cys Gly Glu Gln Phe Arg Cys Leu Lys
690 695 700
Cys Gly Tyr Thr Ala His Ala Asp Leu Glu Val Ala Pro Tyr Asn Ile
705 710 715 720
Arg Leu Val Ala Leu Arg Gly Ala Gly Leu Arg Lys Val Glu Thr Glu
725 730 735
Ala Glu Val Glu Ala Val Asp Glu Ala Pro Ala Ala Ala Ser Val Glu
740 745 750
Lys Pro Arg Lys Arg Arg Lys Ser Ala Ser Gly Glu Ala Ala Thr Phe
755 760 765
Glu Ala Ala Pro Leu Ala
770

<210> 27
<211> 749
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
废水宏基因组序列

[0063]

<400> 27
Met Pro Asp Lys Gln Thr Pro Lys Asp Thr Lys Asp Lys Pro Glu Ser
1 5 10 15
Pro Val Ser Ala Phe Leu Lys Lys His Phe Pro Gly Lys His Phe Phe
20 25 30
Gly His Ala Gly Thr Leu Ala Arg Leu Leu Lys Thr Lys Gly Glu Glu
35 40 45
Val Ala Arg Asp Tyr Ala Ala Lys Lys Val Arg Asp Glu Lys Leu Asp
50 55 60
Phe Arg Pro Pro Ala Lys Cys Gln Ile Val Ala Trp Ser Arg Asp Phe
65 70 75 80
Ser Glu Trp Pro Ile Ala Arg Ala Ser Ala Thr Ile Gln Gln His Val
85 90 95
Ser Gly Leu Thr Lys Glu Asp Phe Glu Arg Phe Asp Pro Gly Lys Ser
100 105 110
Lys Ala Ala His Asp Ala Trp Phe Gln Glu Ser Gly Val Asp Cys His
115 120 125
Gly Tyr Arg His Val Gln Gly Leu Asn Leu Ile Phe Ala Asn Ala Arg
130 135 140
Asp Tyr Tyr Glu Gly Val Val Lys Lys Val Glu Asn Lys Asn Ala Gln
145 150 155 160
Arg Arg Arg Arg Val Glu Ala Leu Asn Ala Arg Arg Ala Glu Glu Gly
165 170 175
Glu Glu Pro Ile Pro Leu Asp Val Glu Glu Ser Pro Phe Gly Glu Asp
180 185 190

[0064]

```
Gly Arg Leu Ala His Pro Pro Gly Val Asn Pro Ser Ile Tyr Val Tyr
        195                 200                 205
Gln Ala Val Ser Pro Arg Pro Leu Lys Lys Ser Asp Leu Glu Thr Val
    210                 215                 220
Val Leu Pro Pro Ala Tyr Ala Gly Tyr Asp Arg Asp Pro Ser Ala Pro
225                 230                 235                 240
Ile Pro Val Met Gly Asp Arg Leu Ser Ile Pro Glu Gly Gln Arg Gly
                245                 250                 255
His Val Pro Ala Trp Gln Arg Asp Gln Leu Ser Pro Asp Lys His Arg
            260                 265                 270
Arg Met Arg Ala Trp Tyr Ser Ala Ala Asn Thr Lys Pro Lys Pro Gly
        275                 280                 285
Arg Thr Ser Val Pro Asp Ala Ala Ala Ile Glu Arg Ala Arg Ala Glu
    290                 295                 300
Gly Ala Leu Leu Val Val Ile Arg Ile Gly Glu Asp Trp Val Val Leu
305                 310                 315                 320
Asp Ala Arg Gly Leu Leu Arg Asn Ala Arg Trp Arg Arg Ile Ala Asp
                325                 330                 335
Lys Glu Ile Ser Leu Asp Gly Leu Leu Asp Leu Phe Thr Gly Asp Pro
            340                 345                 350
Val Ile Asp Ser Lys Arg Asn Val Val Thr Phe Ile Tyr Lys Ala Glu
        355                 360                 365
His Ala Thr Ala Thr Ser Arg Lys Val Val His Arg Lys Ala Ser Arg
    370                 375                 380
Lys Ala Leu Leu Asp Met Thr Ser Pro Gly Glu Asp Gly Leu Pro Arg
385                 390                 395                 400
Glu Val Ala Leu Ala Ser Val Asp Leu Gly Gln Thr Asn Ala Ala Ala
                405                 410                 415
Val Arg Tyr Ala Arg Val His Arg Glu Gly Asp Asp Ile Thr Ser Glu
            420                 425                 430
Cys Leu Val Arg Glu Leu Leu Pro Asp Glu Ile Ser Arg Asp Ile Ala
        435                 440                 445
Arg Tyr Arg Ala Ala Ser Asp Arg Met Glu Ala Glu Ile Arg Glu Ala
    450                 455                 460
Ala Ile Ala Gly Leu Pro Glu Pro Met Gln Ala Glu Val Arg Ala Ala
465                 470                 475                 480
Asp Ala Ser Ser Pro Glu Ala Ala Arg Ala Ala Val Val Ala Leu Val
                485                 490                 495
Gly Asp Gly Leu Pro Trp Glu Lys Met Ser Ser Ala Thr Tyr His Ile
            500                 505                 510
Ser Asp Ala Leu Val Ala Leu Gly Arg Gly Arg Glu Ala Tyr Leu Leu
        515                 520                 525
Ser Lys Ser Lys Asp Gly Glu Glu Lys Ser Val Gln Arg Ser Asp Tyr
    530                 535                 540
```

[0065]

Gly Trp Ser Arg His Leu Arg Pro Arg Leu Ser Glu Glu Thr Arg Lys
545 550 555 560
Ala Met Asn Glu Ala Val Trp Ser Ile Lys Asp Ala His Glu Gly Tyr
565 570 575
Gln Lys Leu Ser Arg Arg Lys Thr Glu Ile Gly Arg Arg Ala Ala Asn
580 585 590
His Val Val Arg Arg Leu Arg Lys Leu Ala Lys Thr Asp Lys Val Ala
595 600 605
Ile Ala Val Glu Asp Leu Asn Val Arg Met Phe His Gly Gly Gly Ser
610 615 620
Arg Ser Thr Gly Trp Asp Asn Phe Phe Val Ala Lys Arg Glu Asn Arg
625 630 635 640
Trp Phe Val Gln Val Leu His Lys Ser Phe Cys Asp Leu Ala Leu His
645 650 655
Arg Gly Glu Val Val Ile Glu Val Asp Pro Ala Arg Thr Ser Gln Thr
660 665 670
Cys Pro Ala Cys Gly His Cys Asp Pro Lys Asn Arg Ser Ser Val Asp
675 680 685
Arg Glu Val Phe Arg Cys Val Val Cys Gly Arg Thr Phe His Ala Asp
690 695 700
Leu Glu Val Ala Thr Phe Asn Ile Glu Arg Val Ala Leu Thr Gly Glu
705 710 715 720
Ser Met Pro Lys Gly Glu Glu Gly Ala Arg Glu Arg Gly Gly Gly Gly
725 730 735
Lys Ser Arg Gly Gly Ala Arg Gly Arg Asn Lys Leu Lys
740 745

<210> 28
<211> 767
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
废水宏基因组序列
<400> 28
Met Ser Glu Ile Thr Asp Leu Leu Lys Ala Asn Phe Lys Gly Lys Thr
1 5 10 15
Phe Lys Ser Ala Asp Met Arg Met Ala Gly Arg Ile Leu Lys Lys Ser
20 25 30
Gly Ala Gln Ala Val Ile Lys Tyr Leu Ser Asp Lys Gly Ala Val Asp
35 40 45
Pro Pro Asp Phe Arg Pro Pro Ala Lys Cys Asn Ile Ile Ala Gln Ser
50 55 60
Arg Pro Phe Asp Glu Trp Pro Ile Cys Lys Ala Ser Met Ala Ile Gln
65 70 75 80

[0066]

Gln His Ile Tyr Gly Leu Thr Lys Asn Glu Phe Asp Glu Ser Ser Pro
85 90 95
Gly Thr Ser Ser Ala Ser His Glu Gln Trp Phe Ala Lys Thr Gly Val
100 105 110
Asp Thr His Gly Phe Thr His Val Gln Gly Leu Asn Leu Ile Phe Gln
115 120 125
His Ala Lys Lys Arg Tyr Glu Gly Val Ile Lys Lys Val Glu Asn Tyr
130 135 140
Asn Glu Lys Glu Arg Lys Lys Phe Glu Gly Ile Asn Glu Arg Arg Ser
145 150 155 160
Lys Glu Gly Met Pro Leu Leu Glu Pro Arg Leu Arg Thr Ala Phe Gly
165 170 175
Asp Asp Gly Lys Phe Ala Glu Lys Pro Gly Val Asn Pro Ser Ile Tyr
180 185 190
Leu Tyr Gln Gln Thr Ser Pro Arg Pro Tyr Asp Lys Thr Lys His Pro
195 200 205
Tyr Val His Ala Pro Phe Glu Leu Lys Glu Ile Thr Thr Ile Pro Thr
210 215 220
Gln Asp Asp Arg Leu Lys Ile Pro Phe Gly Ala Pro Gly His Val Pro
225 230 235 240
Glu Lys His Arg Ser Gln Leu Ser Met Ala Lys His Lys Arg Arg Arg
245 250 255
Ala Trp Tyr Ala Leu Ser Gln Asn Lys Pro Arg Pro Pro Lys Asp Gly
260 265 270
Ser Lys Gly Arg Arg Ser Val Arg Asp Leu Ala Asp Leu Lys Ala Ala
275 280 285
Ser Leu Ala Asp Ala Ile Pro Leu Val Ser Arg Val Gly Phe Asp Trp
290 295 300
Val Val Ile Asp Gly Arg Gly Leu Leu Arg Asn Leu Arg Trp Arg Lys
305 310 315 320
Leu Ala His Glu Gly Met Thr Val Glu Glu Met Leu Gly Phe Phe Ser
325 330 335
Gly Asp Pro Val Ile Asp Pro Arg Arg Asn Val Ala Thr Phe Ile Tyr
340 345 350
Lys Ala Glu His Ala Thr Val Lys Ser Arg Lys Pro Ile Gly Gly Ala
355 360 365
Lys Arg Ala Arg Glu Glu Leu Leu Lys Ala Thr Ala Ser Ser Asp Gly
370 375 380
Val Ile Arg Gln Val Gly Leu Ile Ser Val Asp Leu Gly Gln Thr Asn
385 390 395 400
Pro Val Ala Tyr Glu Ile Ser Arg Met His Gln Ala Asn Gly Glu Leu
405 410 415
Val Ala Glu His Leu Glu Tyr Gly Leu Leu Asn Asp Glu Gln Val Asn
420 425 430

[0067]

```
Ser Ile Gln Arg Tyr Arg Ala Ala Trp Asp Ser Met Asn Glu Ser Phe
        435                 440                 445
Arg Gln Lys Ala Ile Glu Ser Leu Ser Met Glu Ala Gln Asp Glu Ile
    450                 455                 460
Met Gln Ala Ser Thr Gly Ala Ala Lys Arg Thr Arg Glu Ala Val Leu
465                 470                 475                 480
Thr Met Phe Gly Pro Asn Ala Thr Leu Pro Trp Ser Arg Met Ser Ser
                485                 490                 495
Asn Thr Thr Cys Ile Ser Asp Ala Leu Ile Glu Val Gly Lys Glu Glu
            500                 505                 510
Glu Thr Asn Phe Val Thr Ser Asn Gly Pro Arg Lys Arg Thr Asp Ala
        515                 520                 525
Gln Trp Ala Ala Tyr Leu Arg Pro Arg Val Asn Pro Glu Thr Arg Ala
    530                 535                 540
Leu Leu Asn Gln Ala Val Trp Asp Leu Met Lys Arg Ser Asp Glu Tyr
545                 550                 555                 560
Glu Arg Leu Ser Lys Arg Lys Leu Glu Met Ala Arg Gln Cys Val Asn
                565                 570                 575
Phe Val Val Ala Arg Ala Glu Lys Leu Thr Gln Cys Asn Asn Ile Gly
            580                 585                 590
Ile Val Leu Glu Asn Leu Val Val Arg Asn Phe His Gly Ser Gly Arg
        595                 600                 605
Arg Glu Ser Gly Trp Glu Gly Phe Phe Glu Pro Lys Arg Glu Asn Arg
    610                 615                 620
Trp Phe Met Gln Val Leu His Lys Ala Phe Ser Asp Leu Ala Gln His
625                 630                 635                 640
Arg Gly Val Met Val Phe Glu Val His Pro Ala Tyr Ser Ser Gln Thr
                645                 650                 655
Cys Pro Ala Cys Arg Tyr Val Asp Pro Lys Asn Arg Ser Ser Glu Asp
            660                 665                 670
Arg Glu Arg Phe Lys Cys Leu Lys Cys Gly Arg Ser Phe Asn Ala Asp
        675                 680                 685
Arg Glu Val Ala Thr Phe Asn Ile Arg Glu Ile Ala Arg Thr Gly Val
    690                 695                 700
Gly Leu Pro Lys Pro Asp Arg Glu Arg Ser Arg Asp Val Gln Thr Pro
705                 710                 715                 720
Gly Thr Ala Arg Lys Ser Gly Arg Ser Leu Lys Ser Gln Asp Asn Pro
                725                 730                 735
Ser Glu Pro Lys Arg Val Leu Gln Ser Lys Thr Arg Lys Lys Ile Thr
            740                 745                 750
Ser Thr Glu Thr Gln Asn Glu Pro Leu Ala Thr Asp Leu Lys Thr
        755                 760                 765
<210> 29
<211> 749
```

<212> PRT
<213> 未知
<220>
<223> 未知的描述:
废水宏基因组序列
<400> 29

[0068]

```
Met Pro Asp Lys Gln Thr Pro Lys Asp Thr Lys Asp Lys Pro Glu Ser
1               5                  10                  15
Pro Val Ser Ala Phe Leu Lys Lys His Phe Pro Gly Lys His Phe Phe
            20                  25                  30
Gly His Ala Gly Thr Leu Ala Arg Leu Leu Lys Thr Lys Gly Glu Glu
        35                  40                  45
Val Ala Arg Asp Tyr Ala Ala Lys Lys Val Arg Asp Glu Lys Leu Asp
    50                  55                  60
Phe Arg Pro Pro Ala Lys Cys Gln Ile Val Ala Trp Ser Arg Asp Phe
65                  70                  75                  80
Ser Glu Trp Pro Ile Ala Arg Ala Ser Ala Thr Ile Gln Gln His Val
                85                  90                  95
Ser Gly Leu Thr Lys Glu Asp Phe Glu Arg Phe Asp Pro Gly Lys Ser
            100                 105                 110
Lys Ala Ala His Asp Ala Trp Phe Gln Glu Ser Gly Val Asp Cys His
        115                 120                 125
Gly Tyr Arg His Val Gln Gly Leu Asn Leu Ile Phe Ala Asn Ala Arg
    130                 135                 140
Asp Tyr Tyr Glu Gly Val Val Lys Lys Val Glu Asn Lys Asn Ala Gln
145                 150                 155                 160
Arg Arg Arg Arg Val Glu Ala Leu Asn Ala Arg Arg Ala Glu Glu Gly
                165                 170                 175
Glu Glu Pro Ile Pro Leu Asp Val Glu Glu Ser Pro Phe Gly Glu Asp
            180                 185                 190
Gly Arg Leu Ala His Pro Pro Gly Val Asn Pro Ser Ile Tyr Val Tyr
        195                 200                 205
Gln Ala Val Ser Pro Arg Pro Leu Lys Lys Ser Asp Leu Glu Thr Val
    210                 215                 220
Val Leu Pro Pro Ala Tyr Ala Gly Tyr Asp Arg Asp Pro Ser Ala Pro
225                 230                 235                 240
Ile Pro Val Met Gly Asp Arg Leu Ser Ile Pro Glu Gly Gln Arg Gly
                245                 250                 255
His Val Pro Ala Trp Gln Arg Asp Gln Leu Ser Pro Asp Lys His Arg
            260                 265                 270
Arg Met Arg Ala Trp Tyr Ser Ala Ala Asn Thr Lys Pro Lys Pro Gly
        275                 280                 285

Arg Thr Ser Val Pro Asp Ala Ala Ala Ile Glu Arg Ala Arg Ala Glu
```

[0069]

```
    290                 295                 300
Gly Ala Leu Leu Val Val Ile Arg Ile Gly Glu Asp Trp Val Val Leu
305                 310                 315                 320
Asp Ala Arg Gly Leu Leu Arg Asn Ala Arg Trp Arg Arg Ile Ala Asp
                325                 330                 335
Lys Glu Ile Ser Leu Asp Gly Leu Leu Asp Leu Phe Thr Gly Asp Pro
            340                 345                 350
Val Ile Asp Ser Lys Arg Asn Val Val Thr Phe Ile Tyr Lys Ala Glu
        355                 360                 365
His Ala Thr Ala Thr Ser Arg Lys Val Val His Arg Lys Ala Ser Arg
    370                 375                 380
Lys Ala Leu Leu Asp Met Thr Ser Pro Gly Glu Asp Gly Leu Pro Arg
385                 390                 395                 400
Glu Val Ala Leu Ala Ser Val Asp Leu Gly Gln Thr Asn Ala Ala Ala
                405                 410                 415
Val Arg Tyr Ala Arg Val His Arg Glu Gly Asp Asp Ile Thr Ser Glu
            420                 425                 430
Cys Leu Val Arg Glu Leu Leu Pro Asp Glu Ile Ser Arg Asp Ile Ala
        435                 440                 445
Arg Tyr Arg Ala Ala Ser Asp Arg Met Glu Ala Glu Ile Arg Glu Ala
    450                 455                 460
Ala Ile Ala Gly Leu Pro Glu Pro Met Gln Ala Glu Val Arg Ala Ala
465                 470                 475                 480
Asp Ala Ser Ser Pro Glu Ala Ala Arg Ala Ala Val Val Ala Leu Val
                485                 490                 495
Gly Asp Gly Leu Pro Trp Glu Lys Met Ser Ser Ala Thr Tyr His Ile
            500                 505                 510
Ser Asp Ala Leu Val Ala Leu Gly Arg Gly Arg Glu Ala Tyr Leu Leu
        515                 520                 525
Ser Lys Ser Lys Asp Gly Glu Glu Lys Ser Val Gln Arg Ser Asp Tyr
    530                 535                 540
Gly Trp Ser Arg His Leu Arg Pro Arg Leu Ser Glu Glu Thr Arg Lys
545                 550                 555                 560
Ala Met Asn Glu Ala Val Trp Ser Ile Lys Asp Ala His Glu Gly Tyr
                565                 570                 575
Gln Lys Leu Ser Arg Arg Lys Thr Glu Ile Gly Arg Arg Ala Ala Asn
            580                 585                 590
His Val Val Arg Arg Leu Arg Lys Leu Ala Lys Thr Asp Lys Val Ala
        595                 600                 605
Ile Ala Val Glu Asp Leu Asn Val Arg Met Phe His Gly Gly Gly Ser
    610                 615                 620
Arg Ser Thr Gly Trp Asp Asn Phe Phe Val Ala Lys Arg Glu Asn Arg
625                 630                 635                 640
Trp Phe Val Gln Val Leu His Lys Ser Phe Cys Asp Leu Ala Leu His
```

645 650 655
Arg Gly Glu Val Val Ile Glu Val Asp Pro Ala Arg Thr Ser Gln Thr
660 665 670
Cys Pro Ala Cys Gly His Cys Asp Pro Lys Asn Arg Ser Ser Val Asp
675 680 685
Arg Glu Val Phe Arg Cys Val Val Cys Gly Arg Thr Phe His Ala Asp
690 695 700
Leu Glu Val Ala Thr Phe Asn Ile Glu Arg Val Ala Leu Thr Gly Glu
705 710 715 720
Ser Met Pro Lys Gly Glu Glu Gly Ala Arg Glu Arg Gly Gly Gly Gly
725 730 735
Lys Ser Arg Gly Gly Ala Arg Gly Arg Asn Lys Leu Lys
740 745

<210> 30
<211> 31
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 30

[0070] caacctaaac gatggctcga ttcgtcgaga c 31

<210> 31
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 31
gtagaagacc tcgctgattg ctcggtgcgc cgagac 36

<210> 32
<211> 37
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 32
ggttgaaccc tcaacagatt gctcggtaag ccgagac 37

<210> 33
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 33
gttgaaccct caacagattg ctcggtaagc cgagac 36
<210> 34
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 34
cttgaaatcc tgtcagattg ctcccttcgg ggagac 36
<210> 35
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 35
[0071] gctggaagac tcaatgatgg ctccttacga ggagac 36
<210> 36
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 36
cttagaatac tcaatgatgg ctccttacga ggagac 36
<210> 37
<211> 22
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 37
atcggcagct ggtccacctt gg 22
<210> 38
<211> 36
<212> DNA

[0072]

<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 38
gtctctccgt agagagcaat cgttatccat tgagag 36
<210> 39
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 39
ctttcaagac taatagattg ctccttacga ggagac 36
<210> 40
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 40
gtagaaaggt ttactaattg ctccttacga ggagac 36
<210> 41
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 41
cccggaagac cagatgatgg ctcgatcagt cgagac 36
<210> 42
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 42
gtcgagaccg atgacgagtg cgcggtgcgc cgcgac 36
<210> 43
<211> 36

[0073]

<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 43
cctgcaaggg atccaaattg ctctgttcgc agagac 36
<210> 44
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 44
gcgccaacga cctctgattg tccggtacgc cggaac 36
<210> 45
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 45
gtcgagagcg atgacgagtg cgcggtgcgc cgcgac 36
<210> 46
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 46
ccatcaatgg atccaaattg ctctgtacgc agagac 36
<210> 47
<211> 1072
<212> DNA
<213> 未知
<220>
<223> 未知的描述：
植物-根面-柳枝稷根围序列
<400> 47
gaggataagc tcgcgttcgg ccgccagctc ttctgggctc atgtcggcga gcttgtgggc 60
taccacagcg aacggcttgc tcatggtctg caagtgttct ggcaggtggg cgtacgtgaa 120

aaattggagc gtgtggtttg tgtcaatcat gacgttgact acaccgacgt gacatctctg 180
tgcctttggt tttggacggg caccggtggt gtagggttgg acagcagccg tgtcgcagcg 240
ggggaggtct tggggagaat ggttctcttt ggggcctttt ctattggtgg tcttggttcg 300
ccatcctcgt ccataccaac cctccaagta tttcccgcgg aggataacac gaaatgctgc 360
gatggcgtct ctccgtagag agcaatcgtt actcgttgag atatgtacag caacttagac 420
actgggtgct gggccagctg tctccccgta gggagcaatc gtgcgcgaac acgatcactt 480
gtctcgtacg gctaccacgg ctccagcaga agctcaatca tactcaccac atacaccgca 540
gattgcgaga gggacggaag aaaaaccatg aactgtgaaa aatagtacaa agattctgtg 600
cagattaaac tactgagctt ttgatattat tggttttcga gcgggtaaca gcctttgtgc 660
acaaccaaac cactcgcatt tcaaaccttc gtatggccca ggagcggtgt aatgcagctt 720
cggttgccca ccgaggtggg acttcaggag caaggacatg tcgaaaacca aagaactcaa 780
tgactatcaa gaagcgctag cgcgtcgcct tcctggtgtc cgacaccaaa agtcggtacg 840
ccgagctgcg cgactcgtct acgacagaca aggggaagat gccatggtcg ccttcctcga 900
cggcaaacgt gagcgttttt gctgcaccca ctgcggagca cagcgacacg ctgacctcga 960
agtggcgaca cacaacatcg ccatggtagc caccacaggc aaatctctaa caggcaaatc 1020
gttggcgcca caacgactcc aggaagctgc agagtaacga ttgctcccta tg 1072

<210> 48
<211> 494
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
水生-淡水-沼泽森林土壤序列

[0074]

<400> 48
cttgatgccg tcgacgacga agctcgccgg cagcgcccgg cgctcgccct cggggtacta 60
ctggctcggc cagccgcccc agaagagaac gaaattccca cggacagtga cttcgtagag 120
atcggcgtgc gtagggcctt cggcgggcat ggagaactcc ctggtgaaga cgcctttagg 180
agctaggcgg caggattgcc gctgttttca cgcttgtgcg tgacctcgac ctcgaacgcg 240
gaggacacgc taacacgtaa gaatctaact tggcaagtac aggctttctt ttcttgatcc 300
cgaattagat gcgtagtagc cgtgaaatag gagacaagat cctcatgcgt cagcccgcgg 360
agaagaccgc gttccaagtc tttcgtcagg aagtgatcgg gacacagaaa ctatcgggag 420
gcgatgccaa gactgccgga cggctctaca agcagggaaa gatggaagca gcacgtgagt 480
ggttgctcaa aggt 494

<210> 49
<211> 268
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
废水-工业废水-沉积物序列

<400> 49
cgcatctgca acggtctact cgatgctgga ttgattcctg cggtgtagaa tagcataacc 60
agggtccgct ggactatcca gcaacgacgg ctcgatacgt cgagccgatg gagaagtcat 120
gccaaagatc aagaaaccga ctgagatttc cctgctacgc aaggaggtgt tccctgacct 180

ccactttgcc aaggaccgga tgcgtgctgc atccttggtc cttaagaatg agggccggga 240
agccgccatc gagtacctcc gggtgaat 268
<210> 50
<211> 544
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
陆生-土壤-森林土壤序列
<400> 50
ggaatgagtg gagaaggagt ggtttgtaaa ggtacttgga ataagaaaaa gaatagacca 60
aacatgttca agattaagtc taatgcttgg attggaagac ttaaagaata ctgcaacggt 120
aatatggacc taatggccag actactataa tgtttgatct taaagcagaa attgtataaa 180
taatgatatc aaaaatgatt aagcctacgg taagtcagtt tcttacgcct ggatttaagt 240
taattcggaa tcactcaaga actgctggtc taaaactaaa aaatgaaggc gaagaggctt 300
gtaagaaatt cgttcgagaa aatgaaattc ctgtcgctat aacggcccag tcgatgccga 360
aaccaacttg cgagcgctcg ggcgacgcta aaaagcctgt ccgtgctcgc aaggctaaag 420
caccggaatt tcacgataaa ttggcaccaa gctatacggt tgtgctaagg gaggctgttt 480
gaagacataa ccgctcgcaa accgaacgat aaatatgcga aatcattgaa gaatggcgag 540
gcat 544
<210> 51
<211> 534
[0075] <212> DNA
<213> 未知
<220>
<223> 未知的描述:
废水宏基因组序列
<400> 51
gtcgcgccga ggcccattgg cttttatgac tatggctctt tcgacatggc tcgatgggag 60
gacttcatcg aaaaatggat accggagcag ttgccaacgt cgtgggatca cctcatggat 120
gacttggaga tcgcggatgc cgcacgatag gtattgcgcc tccttcatga cgttttctaa 180
gattttacgt tcggtgtaga cgtgagtatg tccgaaatca ctgatctgct caaagccaat 240
ttcaaaggga agacattcaa aagtgcagac atgaggatgg ctgggaggat cttcaagaaa 300
agcggcgcgc aggccgttat caaatacttg tcggacaagg gggcagttga tccgcctaag 360
tccggcaggt ccttgaaatc aaacaagaat ccgagcgagc ccaaacacgt tttgcggtca 420
aaaacacgta gcaacataac atctacgttg tcgcaaaatg aaccgctcgc aacggaccaa 480
aaaaccgctc caaaaacagg accttagacc ccacgccccg cggttcccgc tttt 534
<210> 52
<211> 1275
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
废水宏基因组序列

<400> 52

cgccctcgcg acgcgcatcc cggggtttcg ttttccggag acgatgcagc tcacctgcgc 60
cacggagcag ccgagacggg cggcgacctt ggtcgtgccc tcggactcga cgatcttcac 120
gaacttcttg tgcgtctgtt ctggcgacat gcggggtagc ttgcgacggt ttaggtcgcc 180
tgtcaagcga gctggcgaaa agctgttgac aggagcatgg gctttggcta aggtgcgagc 240
atgcccgaca agcagacgcc caaggacacc aaggacaagc ccgagagccc cgtctccgcc 300
ttcctcaaga agcacttccc cgggaagcat ttcttcgggc atgccgggac gctcgcccgc 360
ctcctcaaga ccaagggtga ggaggttgcg ttccacgccg acctcgaagt ggcgactttc 420
aacatcgagc gggtggcgct gacgggcgag tcgatgccga aaggcgagga gggcgcacgc 480
gagcggggag gtggtgggaa gagcagggga ggcgctcgcg ggcgaaataa attgaaatag 540
ttcgcgaaat cgggtaagct ggttggcgtc agtgaaaact gaatagacga agatacgagc 600
cgctcgcagg gtggagcgga agatctcgat ctggttccga atttctcggt ggctgtcgag 660
agcgatgacg agtgcgcggt gcgccgcgac cagctctgtg tgtagttgac ccgtagtcag 720
gacaaggaga ggtcgagagc gatgacgagt gcgcggtgcg ccgcgacgtg ctgttgatct 780
gatagacggg ggtgggcacc tgcgcggtgt cgagagcgat gacgagtgcg cggtgcgccg 840
cgactttcgt acagagcctc cggacggcgt ccggatgccg gagtcgagag cgatgacgag 900
tgcgcggtgc gccgcgacgc cgtcgcctcg aaggcgggcg cgggtgaact acttcgtcaa 960
gaccgatgac gattgcggtg cgccgcgacg gcttgcgcga ggacttttg gtcggcttct 1020
tgccgtcgtc aagaccgatg acgagtgcgc ggtgcgccgc gacgacttgg ccttcgcctc 1080
ggtcagggcg cgtcggtcga caatttcagt gcaatggctc ggtacaccgg gactatgatt 1140
gttcggtacg ccgagacacg gtcggccagc tggttatctc acgtccctca ccacccccac 1200
gtacgccatc atctccctca tcctcacctc ccacccgtgg ttctccgccg cgagcctcgc 1260
ccccgctcg gcgat 1275

[0076]

<210> 53
<211> 570
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
废水宏基因组序列
<400> 53

gtcgcgccga ggcccattgg cttttatgac tatggctctt tcgacatggc tcgatgggag 60
gacttcatcg aaaaatgggt actggagcag ttgccaacgt cgtgggatca cctcatggat 120
gacttggaga tcgcggatgc cgcacgatag gtattgcgcc tccttcatga cgttttctaa 180
gattttacgt tcggtgtaga cgtgagtatg tccgaaatca ctgatctgct caaagccaat 240
ttcaaaggga agacattcaa aagtgcagac atgaggatgg ctgggaggat cttcaagaaa 300
agcggcgcgc aggccgttat caaatacttg tcggacaagg gggcagttga tccgcctacc 360
ccggggaccg ctcgcaagtc cggcaggtcc ttgaaatcac aagataatcc gagcgagccc 420
aaacgcgttt tgcagtcaaa aacacgtaag aaaatcacat ctacggagac acaaaacgaa 480
ccgctcgcaa cggacctaaa aacctaagtg aaaacagggg cctagacaca catccaagga 540
tgtcgcaaag ggcctcgtcg agatcccaag 570

<210> 54
<211> 703
<212> DNA

<213> 未知
<220>
<223> 未知的描述:
废水宏基因组序列
<400> 54
ctgctgcgcg cgtacccgct ctgggagagt cgcgactgcg gtcagcacga ctcgcgcggc 60
gagggcgtgg agggctgcta ttgcgcgcga caagatctga agcacctgat cggctggccg 120
gagccgaagc aaaaaagctc ctgcaaatga tctatctctg ttgacttcgc atcagacgat 180
cactagattg ctacaggcaa tcacaaacag cgggagacga acaatgactc tggccgagct 240
gcgcgacaaa tacttctaca agatcaagtt ccgcaagatc gatctcaggc aagccggcaa 300
gatcctcaag agagagggcg aggaagcggc tcgccgctat cttgacgagc agcgggagtc 360
gccgcccgag ggccgcggcg cggggctgcg gaaggtcgag actgaggccg aggtcgaggc 420
ggtcgacgag gccccccgcgg ccgcgagcgt ggagaagccg cgcaagcggc gcaagagcgc 480
gagcggcgag gcggcgactt tcgaggccgc accgctcgcg tagtctaaga tcgcgttctc 540
gcaagagaat gcgagcaccg gccgcgaaca ccgctctagt tgtgcgaacg ttcgcggccg 600
gtgctcgcag acgagctgct aggtctttga aaattgaata gattgtaatg gtgagttgcc 660
cgagcccgca ctcgcggcgc gctacctctt cgcgccaacg acc 703
<210> 55
<211> 654
<212> DNA
<213> 未知
<220>
[0077]
<223> 未知的描述:
废水宏基因组序列
<400> 55
cgccctcgcg acgcgcatcc cggggtttcg ttttccggag acgatgcagc tcacctgcgc 60
cacggagcag ccgagacggg cggcgacctt ggtcgtgccc tcggactcga cgatcttcac 120
gaacttcttg tgcgtctgtt ctggcgacat gcggggtagc ttgcgacggt ttaggtcgcc 180
tgtcaagcga gctggcgaaa agctgttgac aggagcatgg gctttggcta aggtgcgagc 240
atgcccgaca agcagacgcc caaggacacc aaggacaagc ccgagagccc cgtctccgcc 300
ttcctcaaga agcacttccc cgggaagcat ttcttcgggc atgccgggac gctcgcccgc 360
ctcctcaaga ccaagggtga ggaggttgcg ttccacgccg acctcggggt ggcgactttc 420
aacatcgagc gggtggcgct gacgggcgag tcgatgccga aaggcgagga gggcgcacgc 480
gagcggggag gtggtgggaa gagcagggga ggcgctcgcg ggcgaaataa attgaaatag 540
ttcgcgaaat cgggtaagct ggttggcgtc agtgaaaact gaatagacga agatacgagc 600
cgctcgcagg gtggagcgga agatctcgat ctggttccga atttctcggt ggct 654
<210> 56
<211> 527
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
废水宏基因组序列
<400> 56

gtcgcgccga ggcccattgg cttttatgac tatggctctt tcgacatggc tcgatgggag 60
gacttcatcg aaaaatgggt gccggagcag ttgccaacgt cgtgggatca cctcatggat 120
gacttggaga tcgcggatgc cgcacgatag gtattgcgcc tccttcatga cgttttctaa 180
gattttacgt tcggtgtaga cgtgagtatg tccgaaatca ctgatctgct caaagccaat 240
ttcaaaggga agacattcaa aagtgcagac atgaggatgg ctgggaggac cctcaagaaa 300
agcggcgcgc aggccgttat caaatacttg tcggacaagg gggcggttga tccgcctaat 360
cccggcaggt ccttgaaatc aaacaagaat ccgagcgagc ccaaacacgt tttgcggtca 420
aaaacacgta gcaacataac atctacgttg tcgcaaaatg aaccgctcgc aacggaccaa 480
aaaaccgctc caaaaacagg accttagacc ccacgccgga agtcgct 527

<210> 57
<211> 147
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸

<400> 57
tccatgtctc gttatacgct gtggttcgcc aacaaatacc ttactactcg tcacgccacg 60
ccatcgtggt gccagatctg tcacaggtaa ctgaacacct ggaaactact aatacctttt 120
gacagctagc tcagtcctag gtataat 147

<210> 58
<211> 149
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸

<400> 58
tccatgtctc gttatacgct gtggttcgcc aacgcaagaa tactactact gcaacctgta 60
ctaacgtatg gtgacttaac tcgtctggat cctacaacag taacgcacta ctagaatact 120
ttgacagcta gctcagtcct aggtataat 149

<210> 59
<211> 151
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸

<400> 59
tccatgtctc gttatacgct gtggttcgcc aacactctca taactactgc gtgaagtgct 60
ttccgaacgc ttcgccacgc tctcctttga tgtacagaaa actcagcgac tactctctca 120
tattgacagc tagctcagtc ctaggtataa t 151

<210> 60
<211> 153

[0078]

<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 60
tccatgtctc gttatacgct gtggttcgcc aacgcacccc cccgactact cggtagcgtc 60
gatgttctgc tgccgttgcc ggggcgtcac aatattgcga atgcgctggc actactcccc 120
cccgttgaca gctagctcag tcctaggtat aat 153
<210> 61
<211> 151
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 61
tccatgtctc gttatacgct gtggttcgcc aacactgcat acactactgt gaccgacggt 60
gttgtaactg acgaaattca ctacctgtct gctatcgaag aaggcaacac tactctgcat 120
acttgacagc tagctcagtc ctaggtataa t 151
<210> 62
<211> 153
[0079]
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 62
tccatgtctc gttatacgct gtggttcgcc aacgcaagaa agtcactact actgaaagcg 60
atcatctctg atgtgaacgc ttccgacgaa gatcgttgga acgctgttct actactagaa 120
agtcttgaca gctagctcag tcctaggtat aat 153
<210> 63
<211> 156
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 63
tccatgtctc gttatacgct gtggttcgcc aacatgagat ggactacttg tttggtcgtg 60
ctgcgcgtga aagccgtatt gaaagcctcc atgccgagcg tgaagtgctt tccactactt 120
gagatggttg acagctagct cagtcctagg tataat 156
<210> 64
<211> 158

<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 64
tccatgtctc gttatacgct gtggttcgcc aacgcagtgg gaccactact gaaggtgatc 60
atgccgtttt cagtttcaac gtggtatgca ccaatgtgct gggtaatgcc gcccgactac 120
tgtgggacct tgacagctag ctcagtccta ggtataat 158
<210> 65
<211> 154
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 65
tccatgtctc gttatacgct gtggttcgcc aacatactgg ctactactat ggaaatacgc 60
gggtccgtac ctggaggaaa tgaagtcgcc gcatggctaa aacgaagacc gactacttac 120
tggctttgac agctagctca gtcctaggta taat 154
<210> 66
<211> 156
[0080]
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 66
tccatgtctc gttatacgct gtggttcgcc aacgcacaac gctgactact ccgcgtccgg 60
cagttttgcc agccagcggc ggcccacttc atcgtcggcg ttaataatcg cctactactc 120
aacgctgttg acagctagct cagtcctagg tataat 156
<210> 67
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 67
agaaagagat cccgtatatc accgaactgg aaggcg 36
<210> 68
<211> 36
<212> DNA
<213> 人工序列

[0081]

<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 68
gtgtggtagg taatgtgatg ataacgcggg atatgc 36
<210> 69
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 69
aagagcgcca tgccggaagg ttatacccaa gagcgt 36
<210> 70
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 70
aagaaagacg gtcacattct gcgtaagaac gttgca 36
<210> 71
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 71
gagaaagaga tcccgtatat caccgaactg gaaggc 36
<210> 72
<211> 39
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 72
acggaaggtg caaaactgtt tgagaaagag atcccgtat 39
<210> 73
<211> 39
<212> DNA

[0082]

<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 73
agaaagagat cccgtatatc accgaactgg aaggcgacg 39
<210> 74
<211> 39
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 74
aatgataaat ttcatacctt cgacgtcgcc ttccagttc 39
<210> 75
<211> 39
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 75
aagagcgcca tgccggaagg ttatacccaa gagcgtacc 39
<210> 76
<211> 39
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 76
gagaaagaga tcccgtatat caccgaactg gaaggcgac 39
<210> 77
<211> 31
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 77
caacctaaac gatggctcga ttcgtcgaga c 31
<210> 78
<211> 27

[0083]

<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 78
ctaaacgatg gctcgattcg tcgagac 27
<210> 79
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 79
gtagaagacc tcgctgattg ctcggtgcgc cgagac 36
<210> 80
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 80
tgattgctcg gtgcgccgag a 21
<210> 81
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 81
ctttcaagac taatagattg ctccttacga ggagac 36
<210> 82
<211> 27
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 82
ctaatagatt gctccttacg aggagac 27
<210> 83

[0084]

<210> 36
<211> DNA
<212> 人工序列
<213>
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 83
gtcgagaccg atgacgagtg cgcggtgcgc cgcgac 36
<210> 84
<211> 25
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 84
tgacgagtgc gcggtgcgcc gcgac 25
<210> 85
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 85
cctgcaaggg atccaaattg ctctgttcgc agagac 36
<210> 86
<211> 25
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 86
tccaaattgc tctgttcgca gagac 25
<210> 87
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 87
cctgcaaggg atccaaattg ctctgttcgc agagac 36

[0085]

<210> 88
<211> 26
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 88
atccaaattg ctctgttcgc agagac 26
<210> 89
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 89
gcgccaacga cctctgattg tccggtacgc cggaac 36
<210> 90
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 90
tctgattgtc cggtacgccg gaac 24
<210> 91
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 91
ccatcaatgg atccaaattg ctctgtacgc agagac 36
<210> 92
<211> 26
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 92

atccaaattg ctctgtacgc agagac 26

<210> 93

<211> 36

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成
寡核苷酸

<400> 93

gtcgagagcg atgacgagtg cgcggtgcgc cgcgac 36

<210> 94

<211> 25

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成
寡核苷酸

<400> 94

tgacgagtgc gcggtgcgcc gcgac 25

<210> 95

<211> 111

<212> DNA

[0086]

<213> 人工序列

<220>

<223> 人工序列的描述：合成
多核苷酸

<400> 95

gctggattga ttcctgcggt gtagaatagc ataaccaggg tccgctggac tatccagcaa 60

cgacggctcg atacgtcgag ccgatggaga agtcatgcca aagatcaaga a 111

<210> 96

<211> 112

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成
多核苷酸

<400> 96

tcgatgctgg attgattcct gcggtgtaga atagcataac cagggtccgc tggactatcc 60

agcaacgacg gctcgatacg tcgagccgat ggagaagtca tgccaaagat ca 112

<210> 97

<211> 98

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成
寡核苷酸
<400> 97
attgattcct gcggtgtaga atagcataac cagggtccgc tggactatcc agcaacgacg 60
gctcgatacg tcgagccgat ggagaagtca tgccaaag 98
<210> 98
<211> 80
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 98
atagcataac cagggtccgc tggactatcc agcaacgacg gctcgatacg tcgagccgat 60
ggagaagtca tgccaaagat 80
<210> 99
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
[0087]
<400> 99
tagcataacc agggtccgct ggactatcca gcaacgacgg ctcgatacgt cgagccgatg 60
gagaagtcat gccaaaga 78
<210> 100
<211> 57
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 100
gccaaagatc aagaaaccga ctgagatttc cctgctacgc aaggaggtgt tccctga 57
<210> 101
<211> 44
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 101
attctacacc gcaggaatca atccagcatc gagtagaccg ttgc 44
<210> 102

[0088]

<210> 45
<211> 45
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 102
attctacacc gcaggaatca atccagcatc gagtagaccg ttgca 45
<210> 103
<211> 137
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 103
cttgtaagaa attcgttcga gaaaatgaaa ttcctgtcgc tataacggcc cagtcgatgc 60
cgaaaccaac ttgcgagcgc tcgggcgacg ctaaaaagcc tgtccgtgct cgcaaggcta 120
aagcaccgga atttcac 137
<210> 104
<211> 136
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 104
ggcttgtaag aaattcgttc gagaaaatga aattcctgtc gctataacgg cccagtcgat 60
gccgaaacca acttgcgagc gctcgggcga cgctaaaaag cctgtccgtg ctcgcaaggc 120
taaagcaccg gaattt 136
<210> 105
<211> 138
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 105
tgtcgagagc gatgacgagt gcgcggtgcg ccgcgaccag ctctgtgtgt agttgacccg 60
tagtcaggac aaggagaggt cgagagcgat gacgagtgcg cggtgcgccg cgacgtgctg 120
ttgatctgat agacgggg 138
<210> 106
<211> 56
<212> DNA

[0089]

<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 106
tcgagagcga tgacgagtgc gcggtgcgcc gcgacgtgct gttgatctga tagacg 56
<210> 107
<211> 81
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 107
ggttccgaat ttctcggtgg ctgtcgagag cgatgacgag tgcgcggtgc gccgcgacca 60
gctctgtgtg tagttgaccc g 81
<210> 108
<211> 79
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 108
ggttccgaat ttctcggtgg ctgtcgagag cgatgacgag tgcgcggtgc gccgcgacca 60
gctctgtgtg tagttgacc 79
<210> 109
<211> 79
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 109
tgctacaggc aatcacaaac agcgggagac gaacaatgac tctggccgag ctgcgcgaca 60
aatacttcta caagatcaa 79
<210> 110
<211> 75
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 110

ctacaggcaa tcacaaacag cgggagacga acaatgactc tggccgagct gcgcgacaaa 60
tacttctaca agatc 75
<210> 111
<211> 93
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 111
gagacgaaca atgactctgg ccgagctgcg cgacaaatac ttctacaaga tcaagttccg 60
caagatcgat ctcaggcaag ccggcaagat cct 93
<210> 112
<211> 99
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 112
gggagacgaa caatgactct ggccgagctg cgcgacaaat acttctacaa gatcaagttc 60
cgcaagatcg atctcaggca agccggcaag atcctcaag 99

[0090]

<210> 113
<211> 66
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 113
gcggccggtg ctcgcagacg agctgctagg tctttgaaaa ttgaatagat tgtaatggtg 60
agttgc 66
<210> 114
<211> 65
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 114
cggccggtgc tcgcagacga gctgctaggt ctttgaaaat tgaatagatt gtaatggtga 60
gttgc 65
<210> 115
<211> 62

<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 115
cgtctacacc gaacgtaaaa tcttagaaaa cgtcatgaag gaggcgcaat acctatcgtg 60
cg 62
<210> 116
<211> 73
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 116
ggttggcgtc agtgaaaact gaatagacga agatacgagc cgctcgcagg gtggagcgga 60
agatctcgat ctg 73
<210> 117
<211> 77
<212> DNA
[0091] <213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 117
gctggttggc gtcagtgaaa actgaataga cgaagatacg agccgctcgc agggtggagc 60
ggaagatctc gatctgg 77
<210> 118
<211> 93
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 118
cccgcatgtc gccagaacag acgcacaaga agttcgtgaa gatcgtcgag tccgagggca 60
cgaccaaggt cgccgcccgt ctcggctgct ccg 93
<210> 119
<211> 100
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成

多核苷酸

<400> 119

ccccgcatgt cgccagaaca gacgcacaag aagttcgtga agatcgtcga gtccgagggc 60

acgaccaagg tcgccgcccg tctcggctgc tccgtggcgc 100

<210> 120

<211> 62

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成

寡核苷酸

<400> 120

aggtgcgagc atgcccgaca agcagacgcc caaggacacc aaggacaagc ccgagagccc 60

cg 62

<210> 121

<211> 61

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成

寡核苷酸

[0092]

<400> 121

aggtgcgagc atgcccgaca agcagacgcc caaggacacc aaggacaagc ccgagagccc 60

c 61

<210> 122

<211> 31

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成

寡核苷酸

<400> 122

gtctcgacga atcgagccat cgtttaggtt g 31

<210> 123

<211> 36

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成

寡核苷酸

<400> 123

gtctcggcgc accgagcaat cagcgaggtc ttctac 36

<210> 124

<211> 37

[0093]

<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 124
gtctcggctt accgagcaat ctgttgaggg ttcaacc 37
<210> 125
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 125
gtctcggctt accgagcaat ctgttgaggg ttcaac 36
<210> 126
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 126
gtctccccga agggagcaat ctgacaggat ttcaag 36
<210> 127
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 127
gtctcctcgt aaggagccat cattgagtct tccagc 36
<210> 128
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 128
gtctcctcgt aaggagccat cattgagtat tctaag 36
<210> 129

<210> 22
<211> DNA
<212> 人工序列
<213>
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 129
ccaaggtgga ccagctgccg at 22
<210> 130
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 130
ctctcaatgg ataacgattg ctctctacgg agagac 36
<210> 131
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 131
gtctcctcgt aaggagcaat ctattagtct tgaaag 36
<210> 132
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 132
gtctcctcgt aaggagcaat tagtaaacct ttctac 36
<210> 133
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成寡核苷酸
<400> 133
gtctcgactg atcgagccat catctggtct tccggg 36

[0094]

[0095]

<210> 134
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 134
gtcgcggcgc accgcgcact cgtcatcggt ctcgac 36
<210> 135
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 135
gtctctgcga acagagcaat ttggatccct tgcagg 36
<210> 136
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 136
gttccggcgt accggacaat cagaggtcgt tggcgc 36
<210> 137
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 137
gtcgcggcgc accgcgcact cgtcatcgct ctcgac 36
<210> 138
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 138

gtctctgcgt acagagcaat ttggatccat tgatgg 36

<210> 139

<211> 157

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成
多核苷酸

<400> 139

caacctaaac gatggctcga ttcgtcgaga caacgaaagt ggcctctggc gaagcgggcg 60

gcacaaccaa aacgatggct cgattcgtcg agacagttac ctgtgacaga tctggcacca 120

cgatggcaac caaaacgatg gctcgattcg tcgagac 157

<210> 140

<211> 180

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成
多核苷酸

<400> 140

ccatcaatgg atccaaattg ctctgtacgc agagacagcg cgccgcggta catcttcttg 60

[0096] ttaacttttt gaccatcaat ggatccaaat tgctctgtac gcagagactg tacatcaaag 120

gagagcgtgg cgaagcgttc ggaaccatca atggatccaa attgctctgt acgcagagac 180

<210> 141

<211> 180

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成
多核苷酸

<400> 141

gcgccaacga cctctgattg tccggtacgc cggaacgagg agcgcagtca ccaaaacttg 60

tcctttcagt ttgcgccaac gacctctgat tgtccggtac gccggaacga tagcagacag 120

gtagtgaatt tcgtcagtta caacgcgcca acgacctctg attgtccggt acgccggaac 180

<210> 142

<211> 190

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成
多核苷酸

<400> 142

gtcgagagcg atgacgagtg cgcggtgcgc cgcgacgaac cgaggtaact ggcttggagg 60

agcgcagtca ccaaaacgtc gagagcgatg acgagtgcgc ggtgcgccgc gacacgctcg 120

gcatggaggc tttcaatacg gctttcacgc gcaggtcgag agcgatgacg agtgcgcggt 180
gcgccgcgac 190
<210> 143
<211> 186
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 143
cctgcaaggg atccaaattg ctctgttcgc agagactggc gcgcaagccg aatgccaaag 60
tggtttatat gcactcctgc aagggatcca aattgctctg ttcgcagaga ctagccatgc 120
ggcgacttca tttcctccag gtacggaccc cctgcaaggg atccaaattg ctctgttcgc 180
agagac 186
<210> 144
<211> 2774
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 144

[0097]

ccccttgtat tactgtttat gtaagcagac aggatgcgtc cggcgtagag gatcgagatc 60
tccaaaaaat ggctgttttt gaaaaaaatt ctaaaggttg ttttacgaca gacgataaca 120
gggttgaaat aattttgttt aactttaaga aggagattta aatatgaaaa tcgaagaagg 180
taaaggtcac catcaccatc accacggatc catgacggca ttgacggaag gtgcaaaact 240
gtttgagaaa gagatcccgt atatcaccga actggaaggc gacgtcgaag gtatgaaatt 300
tatcattaaa ggcgagggta ccggtgacgc gaccacgggt accattaaag cgaaatacat 360
ctgcactacg ggcgacctgc cggtcccgtg ggcaaccctg gtgagcaccc tgagctacgg 420
tgttcagtgt ttcgccaagt acccgagcca catcaaggat ttctttaaga gcgccatgcc 480
ggaaggttat acccaagagc gtaccatcag cttcgaaggc gacggcgtgt acaagacgcg 540
tgctatggtt acctacgaac gcggttctat ctacaatcgt gtcacgctga ctggtgagaa 600
ctttaagaaa gacggtcaca ttctgcgtaa gaacgttgca ttccaatgcc cgccaagcat 660
tctgtatatt ctgcctgaca ccgttaacaa tggcatccgc gttgagttca accaggcgta 720
cgatattgaa ggtgtgaccg aaaaactggt taccaaatgc agccaaatga atcgtccgtt 780
ggcgggctcc gcggcagtgc atatcccgcg ttatcatcac attacctacc acaccaaact 840
gagcaaagac cgcgacgagc gccgtgatca catgtgtctg gtagaggtcg tgaaagcggt 900
tgatctggac acgtatcagt aataaaaagc ccgaaaggaa gctgagttgg ctgctgccac 960
cgctgagcaa taactagcat aacccccttgg ggcctctaaa cgggtcttga ggggtttttt 1020
gctgaaagga ggaactatat ccggcttcct cgctcactga ctcgctgcgc tcggtcgttc 1080
ggctgcggcg agcggtatca gctcactcaa aggcggtaat acggttatcc acagaatcag 1140
gggataacgc aggaaagaac atgtgagcaa aaggccagca aaaggccagg aaccgtaaaa 1200
aggccgcgtt gctggcgttt ttccataggc tccgcccccc tgacgagcat cacaaaaatc 1260
gacgctcaag tcagaggtgg cgaaacccga caggactata aagataccag gcgtttcccc 1320

[0098]

ctggaagctc cctcgtgcgc tctcctgttc cgaccctgcc gcttaccgga tacctgtccg 1380
cctttctccc ttcgggaagc gtggcgcttt ctcatagctc acgctgtagg tatctcagtt 1440
cggtgtaggt cgttcgctcc aagctgggct gtgtgcacga acccccccgtt cagcccgacc 1500
gctgcgcctt atccggtaac tatcgtcttg agtccaaccc ggtaagacac gacttatcgc 1560
cactggcagc agccactggt aacaggatta gcagagcgag gtatgtaggc ggtgctacag 1620
agttcttgaa gtggtggcct aactacggct acactagaag aacagtattt ggtatctgcg 1680
ctctgctgaa gccagttacc ttcggaaaaa gagttggtag ctcttgatcc ggcaaacaaa 1740
ccaccgctgg tagcggtggt ttttttgttt gcaagcagca gattacgcgc agaaaaaaag 1800
gatctcaaga agatcctttg atcttttcta cggggtctga cgctcagtgg aacgaaaact 1860
cacgggtggc acttttcggg gaaatgtgcg cggaacccct atttgtttat ttttctaaat 1920
acattcaaat atgtatccgc tcatgaatta attcttagaa aaactcatcg agcatcaaat 1980
gaaactgcaa tttattcata tcaggattat caataccata tttttgaaaa agccgtttct 2040
gtaatgaagg agaaaactca ccgaggcagt tccataggat ggcaagatcc tggtatcggt 2100
ctgcgattcc gactcgtcca acatcaatac aacctattaa tttcccctcg tcaaaaataa 2160
ggttatcaag tgagaaatca ccatgagtga cgactgaatc cggtgagaat ggcaaaagtt 2220
tatgcatttc tttccagact tgttcaacag gccagccatt acgctcgtca tcaaaatcac 2280
tcgcatcaac caaaccgtta ttcattcgtg attgcgcctg agcgagacga aatacgcgat 2340
cgctgttaaa aggacaatta caaacaggaa tcgaatgcaa ccggcgcagg aacactgcca 2400
gcgcatcaac aatattttca cctgaatcag gatattcttc taatacctgg aatgctgttt 2460
tcccggggat cgcagtggtg agtaaccatg catcatcagg agtacggata aaatgcttga 2520
tggtcggaag aggcataaat tccgtcagcc agtttagtct gaccatctca tctgtaacat 2580
cattggcaac gctacctttg ccatgtttca gaaacaactc tggcgcatcg ggcttcccat 2640
acaatcgata gattgtcgca cctgattgcc cgacattatc gcgagcccat ttatacccat 2700
ataaatcagc atccatgttg gaatttaatc gcggcctaga gcaagacgtt tcccgttgaa 2760
tatggctcat aaca 2774

<210> 145
<211> 2771
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 145

ccccttgtat tactgtttat gtaagcagac aggatgcgtc cggcgtagag gatcgagatc 60
tccaaaaaat ggctgttttt gaaaaaaatt ctaaaggttg ttttacgaca gacgataaca 120
gggttgaaat aattttgttt aactttaaga aggagattta aatatgaaaa tcgaagaagg 180
taaaggtcac catcaccatc accacggatc catggtcagc aaggggggagg aagacaatat 240
ggctattatc aaggaattca tgcgcttcaa ggtgcatatg gaaggaagcg tgaatggaca 300
cgaattcgag atcgaaggcg agggggaggg tcgcccttat gaaggcacac aaacagctaa 360
actgaaagtg acgaagggag ggccgcttcc cttcgcttgg gacattcttt caccccagtt 420
catgtatggt tcaaaggctt atgtcaagca cccggcggac attccagact acttaaaatt 480
gtcgttcccc gaggggttta aatgggaacg cgttatgaat ttcgaggatg ggggagtcgt 540
aacggttacc caggacagta gcctgcagga tggcgagttc atctacaaag tgaaattgcg 600
cgggacgaac ttccctagcg atgggccagt catgcagaag aaaacgatgg gatgggaagc 660

[0099]

gtcatccgag cgcatgtatc ctgaagatgg tgctttaaaa ggtgagatca agcagcgttt 720
gaaactgaag gacgggggcc attatgatgc tgaagttaaa acgacatata aggccaagaa 780
gccagttcaa ctgccagggg cttataatgt taatattaaa ttagacatta cgagccataa 840
tgaagattac acgattgtcg agcaatacga gcgcgcagaa ggacgccact caacgggggg 900
catggacgag ctgtacaagt aaaaagcccg aaaggaagct gagttggctg ctgccaccgc 960
tgagcaataa ctagcataac cccttggggc ctctaaacgg gtcttgaggg gttttttgct 1020
gaaaggagga actatatccg gcttcctcgc tcactgactc gctgcgctcg gtcgttcggc 1080
tgcggcgagc ggtatcagct cactcaaagg cggtaatacg gttatccaca gaatcagggg 1140
ataacgcagg aaagaacatg tgagcaaaag gccagcaaaa ggccaggaac cgtaaaaagg 1200
ccgcgttgct ggcgtttttc cataggctcc gccccctga cgagcatcac aaaaatcgac 1260
gctcaagtca gaggtggcga aacccgacag gactataaag ataccaggcg tttccccctg 1320
gaagctccct cgtgcgctct cctgttccga ccctgccgct taccggatac ctgtccgcct 1380
ttctcccttc gggaagcgtg gcgctttctc atagctcacg ctgtaggtat ctcagttcgg 1440
tgtaggtcgt tcgctccaag ctgggctgtg tgcacgaacc ccccgttcag cccgaccgct 1500
gcgccttatc cggtaactat cgtcttgagt ccaacccggt aagacacgac ttatcgccac 1560
tggcagcagc cactggtaac aggattagca gagcgaggta tgtaggcggt gctacagagt 1620
tcttgaagtg gtggcctaac tacggctaca ctagaagaac agtatttggt atctgcgctc 1680
tgctgaagcc agttaccttc ggaaaaagag ttggtagctc ttgatccggc aaacaaacca 1740
ccgctggtag cggtggtttt tttgtttgca agcagcagat tacgcgcaga aaaaaaggat 1800
ctcaagaaga tcctttgatc ttttctacgg ggtctgacgc tcagtggaac gaaaactcac 1860
gggtggcact tttcggggaa atgtgcgcgg aacccctatt tgtttatttt tctaaataca 1920
ttcaaatatg tatccgctca tgaattaatt cttagaaaaa ctcatcgagc atcaaatgaa 1980
actgcaattt attcatatca ggattatcaa taccatattt ttgaaaaagc cgtttctgta 2040
atgaaggaga aaactcaccg aggcagttcc ataggatggc aagatcctgg tatcggtctg 2100
cgattccgac tcgtccaaca tcaatacaac ctattaattt cccctcgtca aaaataaggt 2160
tatcaagtga gaaatcacca tgagtgacga ctgaatccgg tgagaatggc aaaagtttat 2220
gcatttcttt ccagacttgt tcaacaggcc agccattacg ctcgtcatca aaatcactcg 2280
catcaaccaa accgttattc attcgtgatt gcgcctgagc gagacgaaat acgcgatcgc 2340
tgttaaaagg acaattacaa acaggaatcg aatgcaaccg gcgcaggaac actgccagcg 2400
catcaacaat attttcacct gaatcaggat attcttctaa tacctggaat gctgttttcc 2460
cggggatcgc agtggtgagt aaccatgcat catcaggagt acggataaaa tgcttgatgg 2520
tcggaagagg cataaattcc gtcagccagt ttagtctgac catctcatct gtaacatcat 2580
tggcaacgct acctttgcca tgtttcagaa acaactctgg cgcatcgggc ttcccataca 2640
atcgatagat tgtcgcacct gattgcccga cattatcgcg agcccattta tacccatata 2700
aatcagcatc catgttggaa tttaatcgcg gcctagagca agacgtttcc cgttgaatat 2760
ggctcataac a 2771

<210> 146
<211> 158
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 146

gaaattaata cgactcacta tagccatcaa tggatccaaa ttgctctgta cgcagagaca 60
gaaagagatc ccgtatatca ccgaactgga aggcgccatc aatggatcca aattgctctg 120
tacgcagaga cctaacccct ctctaaacgg aggggttt 158
<210> 147
<211> 158
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 147
gaaattaata cgactcacta tagccatcaa tggatccaaa ttgctctgta cgcagagacg 60
tgtggtaggt aatgtgatga taacgcggga tatgcccatc aatggatcca aattgctctg 120
tacgcagaga cctaacccct ctctaaacgg aggggttt 158
<210> 148
<211> 158
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 148

[0100]

gaaattaata cgactcacta tagccatcaa tggatccaaa ttgctctgta cgcagagaca 60
agagcgccat gccggaaggt tatacccaag agcgtccatc aatggatcca aattgctctg 120
tacgcagaga cctaacccct ctctaaacgg aggggttt 158
<210> 149
<211> 158
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 149
gaaattaata cgactcacta tagccatcaa tggatccaaa ttgctctgta cgcagagaca 60
agaaagacgg tcacattctg cgtaagaacg ttgcaccatc aatggatcca aattgctctg 120
tacgcagaga cctaacccct ctctaaacgg aggggttt 158

<210> 150
<211> 158
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸

<400> 150
gaaattaata cgactcacta tagccatcaa tggatccaaa ttgctctgta cgcagagacg 60
agaaagagat cccgtatatc accgaactgg aaggcccatc aatggatcca aattgctctg 120
tacgcagaga cctaacccct ctctaaacgg aggggttt 158
<210> 151
<211> 158
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 151
gaaattaata cgactcacta taggtataac gaccctgcga agtggggtgt aacttcgacg 60
tgattgcgcc tgagcgagac gaaatacgcg atcgcgtata acgaccctgc gaagtggggt 120
gtaacttcga cctaacccct ctctaaacgg aggggttt 158
<210> 152
<211> 161
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸

[0101]

<400> 152
gaaattaata cgactcacta tagcctgcaa gggatccaaa ttgctctgtt cgcagagaca 60
cggaaggtgc aaaactgttt gagaaagaga tcccgtatcc tgcaagggat ccaaattgct 120
ctgttcgcag agacctaacc cctctctaaa cggaggggtt t 161
<210> 153
<211> 161
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 153
gaaattaata cgactcacta tagcctgcaa gggatccaaa ttgctctgtt cgcagagaca 60
gaaagagatc ccgtatatca ccgaactgga aggcgacgcc tgcaagggat ccaaattgct 120
ctgttcgcag agacctaacc cctctctaaa cggaggggtt t 161
<210> 154
<211> 161
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸

<400> 154
gaaattaata cgactcacta tagcctgcaa gggatccaaa ttgctctgtt cgcagagaca 60
atgataaatt tcataccttc gacgtcgcct tccagttccc tgcaagggat ccaaattgct 120
ctgttcgcag agacctaacc cctctctaaa cggaggggtt t 161
<210> 155
<211> 161
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 155
gaaattaata cgactcacta tagcctgcaa gggatccaaa ttgctctgtt cgcagagaca 60
agagcgccat gccggaaggt tatacccaag agcgtacccc tgcaagggat ccaaattgct 120
ctgttcgcag agacctaacc cctctctaaa cggaggggtt t 161
<210> 156
<211> 161
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸

[0102]

<400> 156
gaaattaata cgactcacta tagcctgcaa gggatccaaa ttgctctgtt cgcagagacg 60
agaaagagat cccgtatatc accgaactgg aaggcgaccc tgcaagggat ccaaattgct 120
ctgttcgcag agacctaacc cctctctaaa cggaggggtt t 161
<210> 157
<211> 161
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
多核苷酸
<400> 157
gaaattaata cgactcacta taggtataac gaccctgcga agtggggtgt aacttcgacg 60
tgattgcgcc tgagcgagac gaaatacgcg atcgctgtgt ataacgaccc tgcgaagtgg 120
ggtgtaactt cgacctaacc cctctctaaa cggaggggtt t 161
<210> 158
<211> 24
<212> DNA
<213> 未知
<220>
<223> 未知的描述：
AAVS1 靶基因座序列

<400> 158
tggcctgggt cacctctacg gctg 24
<210> 159
<211> 96
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 159
cctgcaaggg atccaaattg ctctgttcgc agagactggc ctgggtcacc tctacggctg 60
cctgcaaggg atccaaattg ctctgttcgc agagac 96
<210> 160
<211> 96
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 160
ccatcaatgg atccaaattg ctctgtacgc agagactggc ctgggtcacc tctacggctg 60
[0103] ccatcaatgg atccaaattg ctctgtacgc agagac 96
<210> 161
<211> 96
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<400> 161
cctgcaaggg atccaaattg ctctgttcgc agagactggc ctgggtcacc tctacggctg 60
cctgcaaggg atccaaattg ctctgttcgc agagac 96
<210> 162
<211> 37
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述：合成
寡核苷酸
<220>
<221> 经修饰的碱基
<222> (28)..(28)
<223> a、 c、 t、 g、 未知或其他
<400> 162

gsyykmaabm ywhdhwgatk gckcbdtncg hbgagac 37

<210> 163

<211> 36

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成
寡核苷酸

<400> 163

[0104] cctgcgaaac cttttgattg ctcagtacgc tgagac 36

<210> 164

<211> 37

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成
寡核苷酸

<400> 164

ggatccaatc ctttttgatt gcccaattcg ttgggac 37

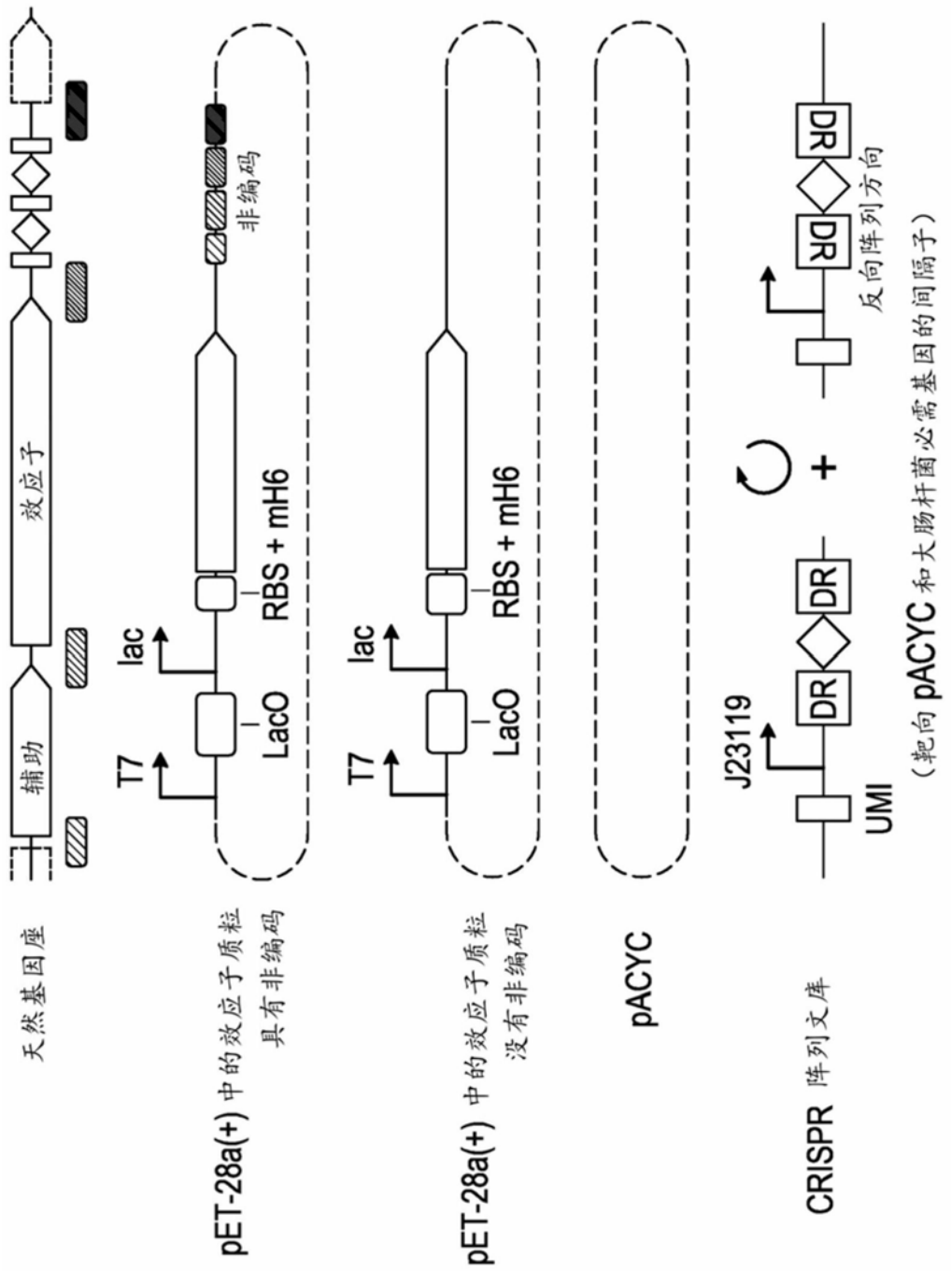

天然基因座
辅助
效应子
pET-28a(+) 中的效应子质粒
具有非编码
T7
lac
LacO
RBS + mH6
非编码
pET-28a(+) 中的效应子质粒
没有非编码
T7
lac
LacO
RBS + mH6
pACYC
CRISPR 阵列文库
J23119
DR
DR
UMI
+
DR
DR
反向阵列方向
(靶向 pACYC 和大肠杆菌必需基因的间隔子)

图1A

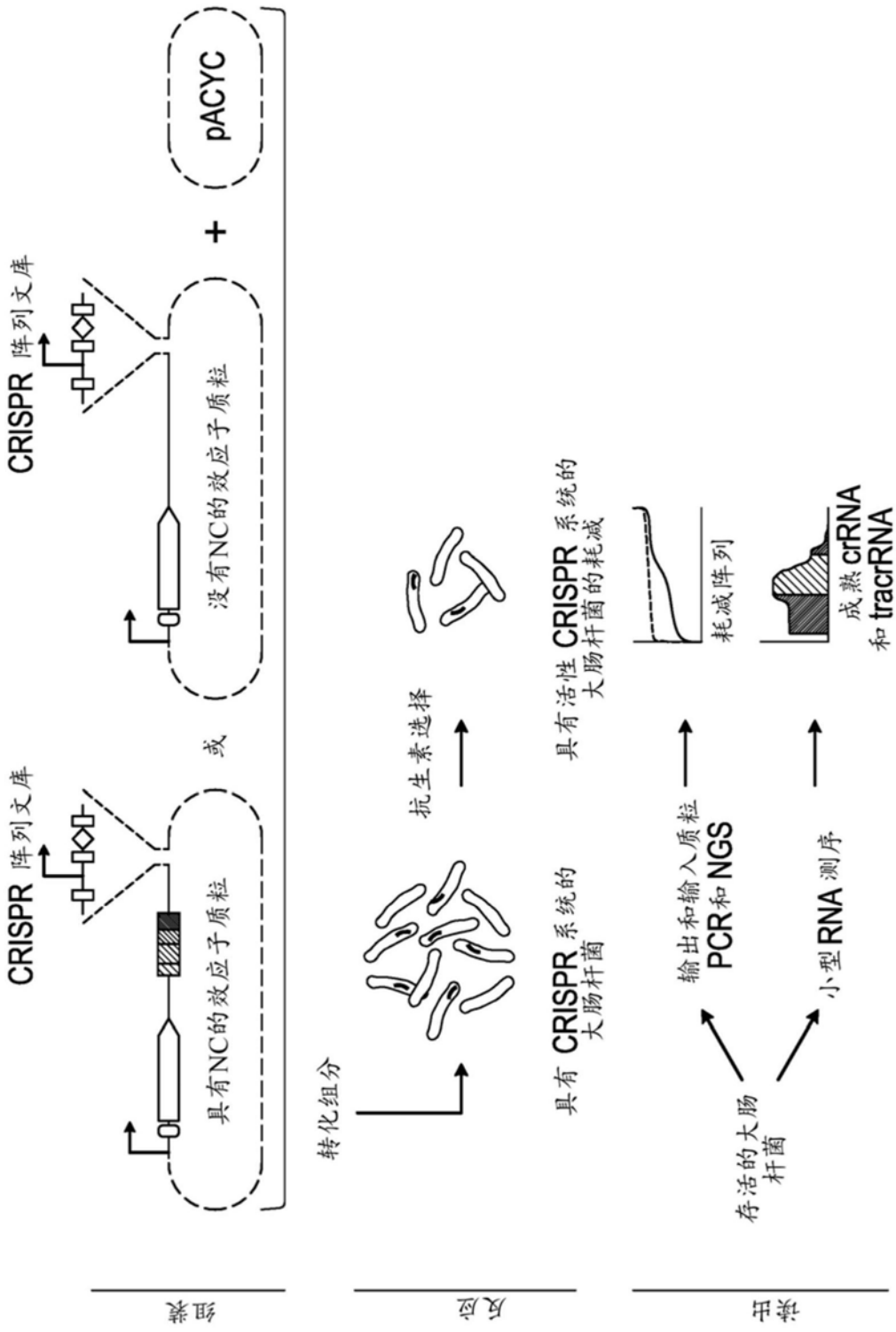

CRISPR 阵列文库
具有NC的效应子质粒
或
CRISPR 阵列文库
没有NC的效应子质粒
+
pACYC
转化组分
抗生素选择
具有 CRISPR 系统的
大肠杆菌
具有活性 CRISPR 系统的
大肠杆菌的耗减
存活的大肠
杆菌
输出和输入质粒
PCR和 NGS
耗减阵列
小型 RNA 测序
成熟 crRNA
和 tracrRNA
设计
构建
筛选
书籍

图1B

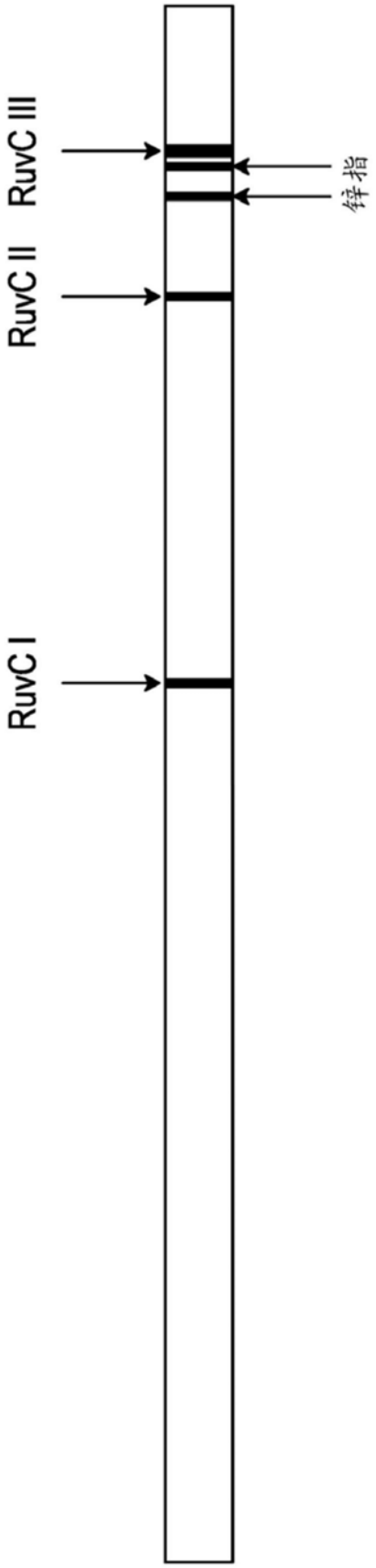


图2A

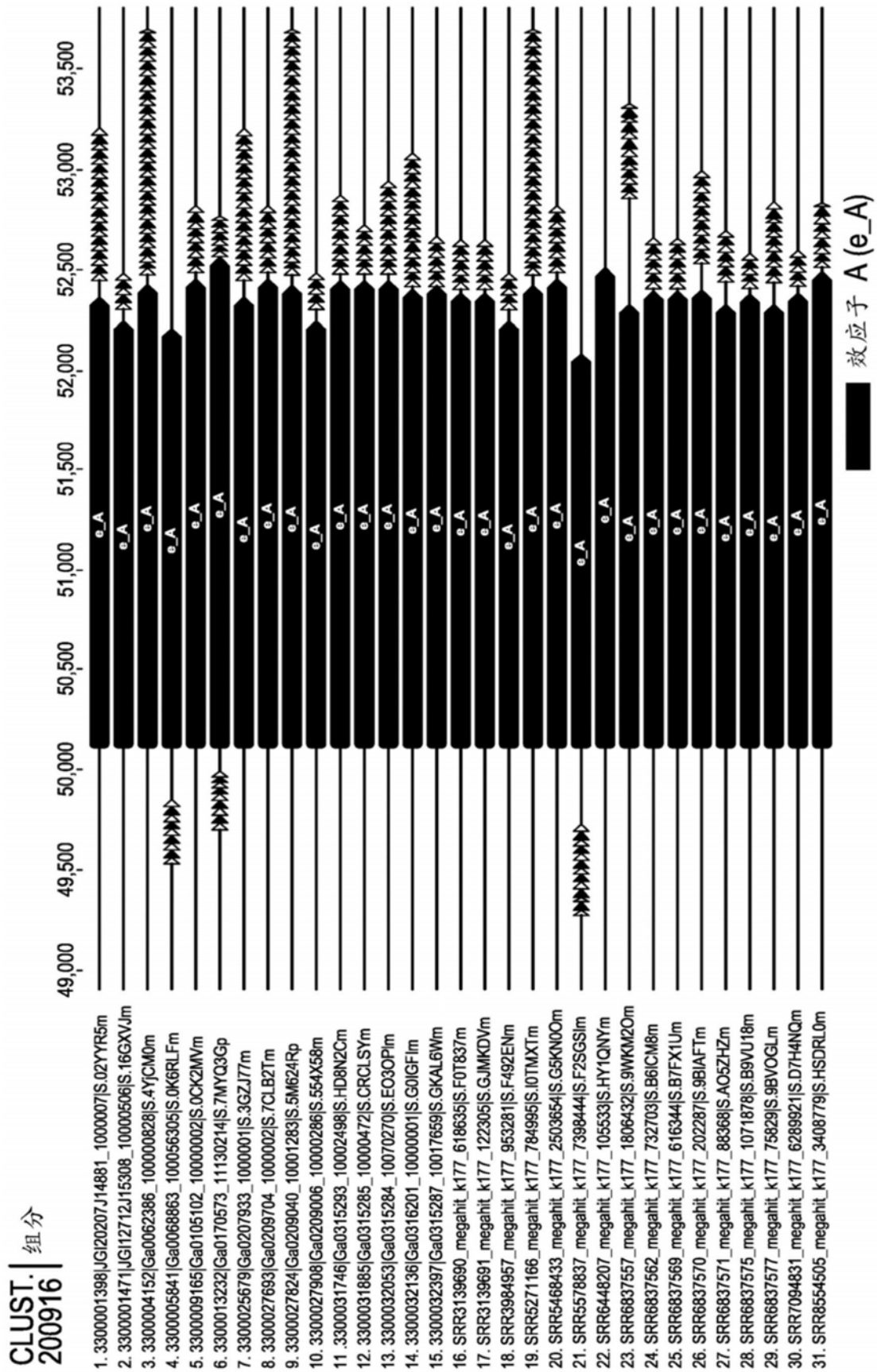

CLUST. 200916 | 组分
49,000
49,500
50,000
50,500
51,000
51,500
52,000
52,500
53,000
53,500
1. 3300001398|JGI20207J14881_1000007|S.02YYR5m
2. 3300001471|JGI12712J15308_10000506|S.16GXVJm
3. 3300004152|Ga0062386_100000828|S.4YjCM0m
4. 3300005841|Ga0068863_100056305|S.0K6RLFm
5. 3300009165|Ga0105102_10000002|S.0CK2MVm
6. 3300013232|Ga0170573_11130214|S.7MYQ3Gp
7. 3300025679|Ga0207933_1000001|S.3GZJ77m
8. 3300027693|Ga0209704_1000002|S.7CLB2Tm
9. 3300027824|Ga0209040_10001283|S.5M624Rp
10. 3300027908|Ga0209006_10000286|S.554X58m
11. 3300031746|Ga0315293_10002498|S.HD8N2Cm
12. 3300031885|Ga0315285_10000472|S.CRCLSYm
13. 3300032053|Ga0315284_10070270|S.EO3OPlm
14. 3300032136|Ga0316201_10000001|S.G0IGFlm
15. 3300032397|Ga0315287_10017659|S.GKAL6Wm
16. SRR3139690_megahit_k177_618635|S.F0T837m
17. SRR3139691_megahit_k177_122305|S.GJMKDVm
18. SRR3984957_megahit_k177_953281|S.F492ENm
19. SRR5271166_megahit_k177_784995|S.I0TMXTm
20. SRR5468433_megahit_k177_2503654|S.G5KN0Om
21. SRR5578837_megahit_k177_7398444|S.F2SGSlm
22. SRR6448207_megahit_k177_105533|S.HY1QNYm
23. SRR6837557_megahit_k177_1806432|S.9WKM2Om
24. SRR6837562_megahit_k177_732703|S.B6ICM8m
25. SRR6837569_megahit_k177_616344|S.B7FX1Um
26. SRR6837570_megahit_k177_202287|S.9BIAFTm
27. SRR6837571_megahit_k177_88368|S.AO5ZHZm
28. SRR6837575_megahit_k177_1071878|S.B9VU18m
29. SRR6837577_megahit_k177_75829|S.9BVOGLm
30. SRR7094831_megahit_k177_6289921|S.D7H4NQm
31. SRR8554505_megahit_k177_3408779|S.HSDRL0m
e_A
效应子 A (e_A)

图2B

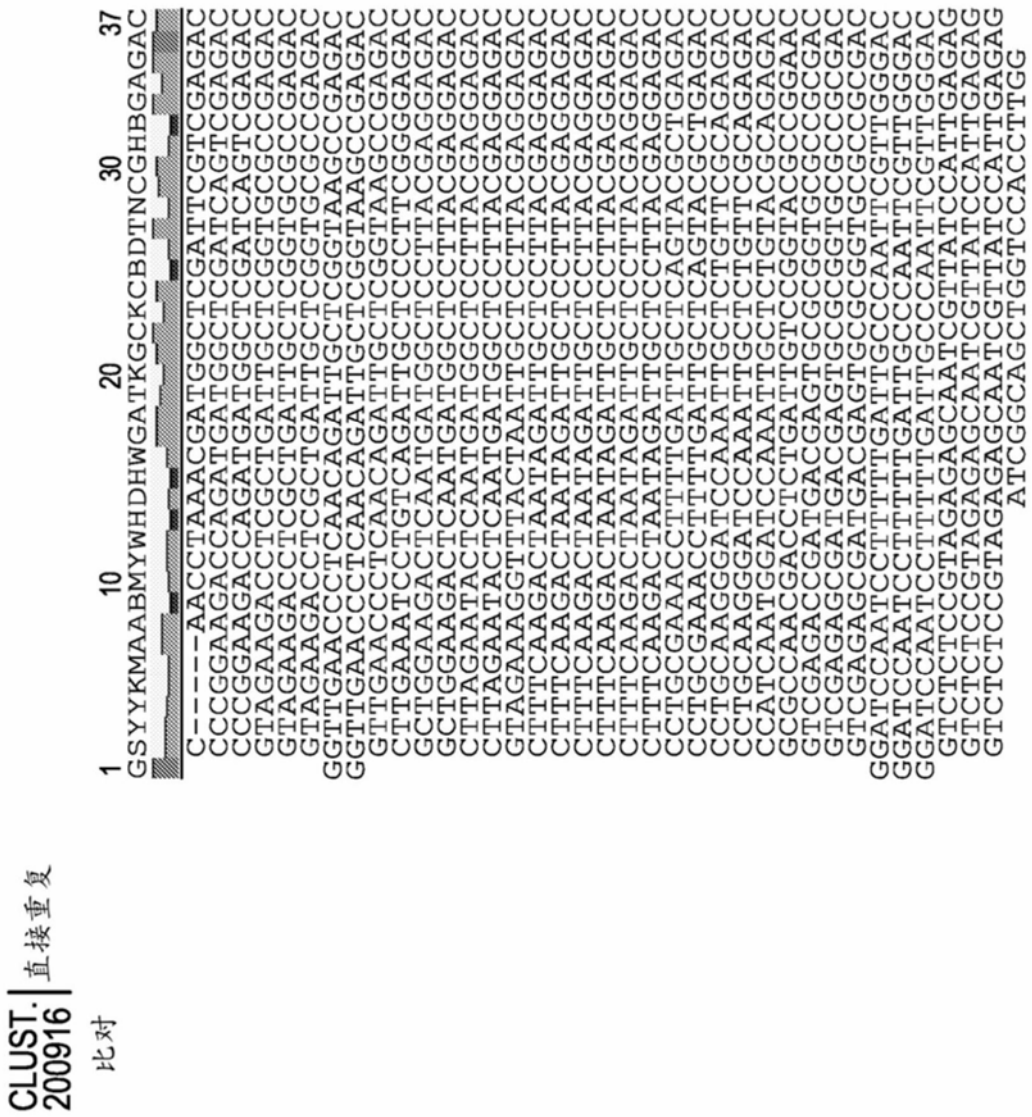

CLUST.
200916
直接重复
比对
1
10
20
30
37
GSYYKMAABMYWHDHWGATKGCKCBDTNCGHBGAC
C-----AACCTAAACGATGGCTCGATTCGTCGAGAC
CCCGGAAGACCAGATGATGGCTCGATCAGTCGAGAC
CCCGGAAGACCAGATGATGGCTCGATCAGTCGAGAC
GTAGAAGACCTCGCTGATTGCTCGGTGCGCCCGAGAC
GTAGAAGACCTCGCTGATTGCTCGGTGCGCCCGAGAC
GGTTGAACCCCTCAACAGATTGCTCGGTAAGCCCGAGAC
GGTTGAACCCCTCAACAGATTGCTCGGTAAGCCCGAGAC
GTTTGAAACTCCTGTCAGATTGCTCCCTTCGGGGAGAC
GCTTGGAAGACTCAATGATGGCTCCCTTACGAGGGAGAC
GCTTGGAAGACTCAATGATGGCTCCCTTACGAGGGAGAC
CTTTAGAATACTCAATGATGGCTCCCTTACGAGGGAGAC
CTTTAGAATACTCAATGATGGCTCCCTTACGAGGGAGAC
GTTTTCAAGAGTTTACTAATTGCTCCCTTACGAGGGAGAC
CTTTTCAAGACTAATAGATTGCTCCCTTACGAGGGAGAC
CTTTTCAAGACTAATAGATTGCTCCCTTACGAGGGAGAC
CTTTTCAAGACTAATAGATTGCTCCCTTACGAGGGAGAC
CTTTTCAAGACTAATAGATTGCTCCCTTACGAGGGAGAC
CCTTGCGAAACCTTTTTGATTGCTCAGTACGCTGAGAC
CCTGCAAGGGATCTCCAAATTGCTCGCTGAGAC
CCTGCAATGGGATCCCAAATTGCTCTGTTCGCAGAGAC
GCGCCAACGACCTCTGATTGTCCGGTACGCCGGAAC
GTCGAGACCGATGACGAGTGCGCGGTGCGCCGCGAC
GTCGAGAGCGATGACGAGTGCGCGGTGCGCCGCGAC
GGATCCAATCCTTTTTGATTGCCCAATTCGTTGGGAC
GGATCCAATCCTTTTTGATTGCCCAATTCGTTGGGAC
GTCTCTCCGTAGAGAGCAATCGTTATCCATTGAGAG
GTCTCTCCGTAGAGAGCAATCGTTATCCATTGAGAG
ATCGGCAGCTGGTCCACCTTGG

图2C

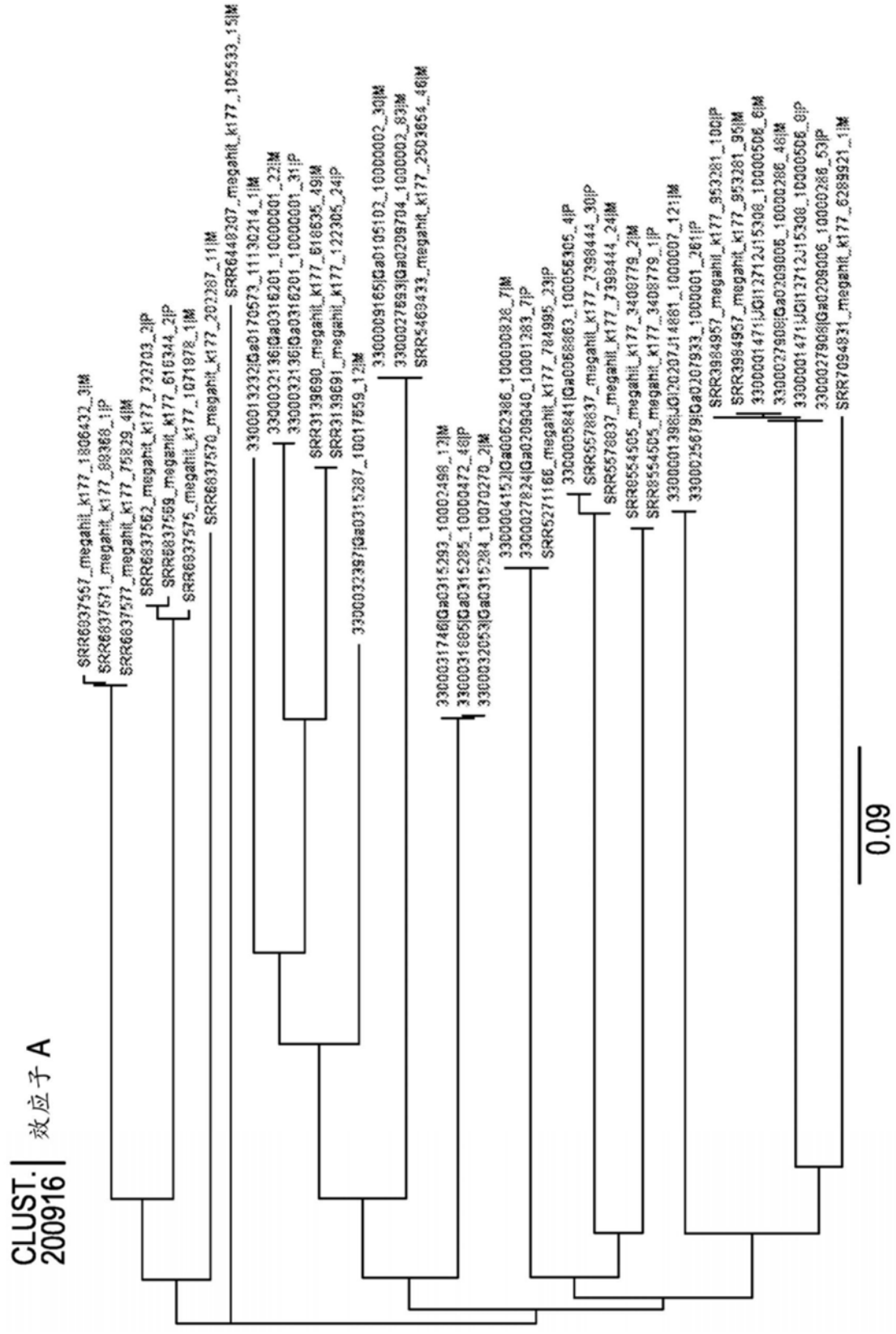

CLUST.
200916
效应子A
SRR6837557_megahit_k177_1806432_3kb
SRR6837571_megahit_k177_88368_1kP
SRR6837577_megahit_k177_75829_4kb
SRR6837562_megahit_k177_732703_2kP
SRR6837559_megahit_k177_616344_2kP
SRR6837575_megahit_k177_1071878_1kb
SRR6837570_megahit_k177_202287_11kb
SRR6448207_megahit_k177_105533_15kb
3300013232|Ga0170573_11130214_1kb
3300032136|Ga0316201_10000001_22kb
3300032136|Ga0316201_10000001_31kP
SRR3139690_megahit_k177_818635_49kb
SRR3139691_megahit_k177_122305_24kP
3300032397|Ga0315287_10017859_12kb
3300009165|Ga0105102_10000002_30kb
3300027893|Ga0209704_1000002_83kb
SRR5469433_megahit_k177_2503654_46kb
3300031746|Ga0315293_10002498_13kb
3300031685|Ga0315285_10000472_48kP
3300032053|Ga0315284_10070270_2kb
3300004152|Ga0062386_100000826_7kb
3300027824|Ga0209040_10001283_7kP
SRR5271166_megahit_k177_784995_23kP
3300005841|Ga0068863_100056305_4kP
SRR5578037_megahit_k177_7398444_30kP
SRR5578037_megahit_k177_7398444_24kb
SRR8554505_megahit_k177_3408779_2kb
SRR8554505_megahit_k177_3408779_1kP
3300001398|JGI20207J14881_1000007_121kb
3300025879|Ga0207933_1000001_261kP
SRR3984957_megahit_k177_953281_100kP
SRR3984957_megahit_k177_953281_95kb
3300001471|JGI12712J15308_10000506_6kb
3300027908|Ga0209005_10000286_46kb
3300001471|JGI12712J15308_10000506_6kP
3300027908|Ga0209086_10000288_53kP
SRR7094831_megahit_k177_6289921_1kb
0.09

图2D

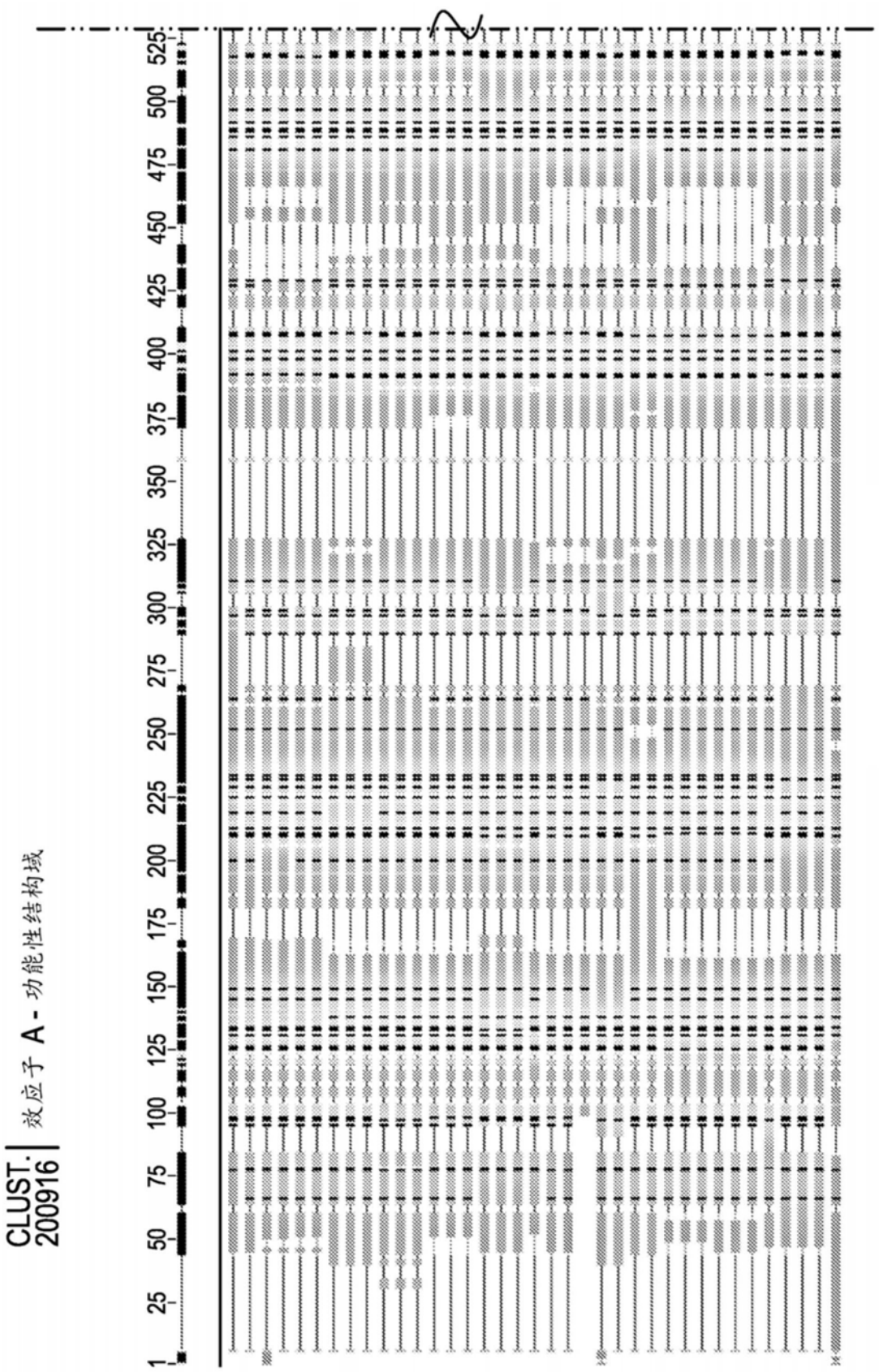

CLUST.
200916
效应子 A - 功能性结构域
1
25
50
75
100
125
150
175
200
225
250
275
300
325
350
375
400
425
450
475
500
525

图2E(部分1)

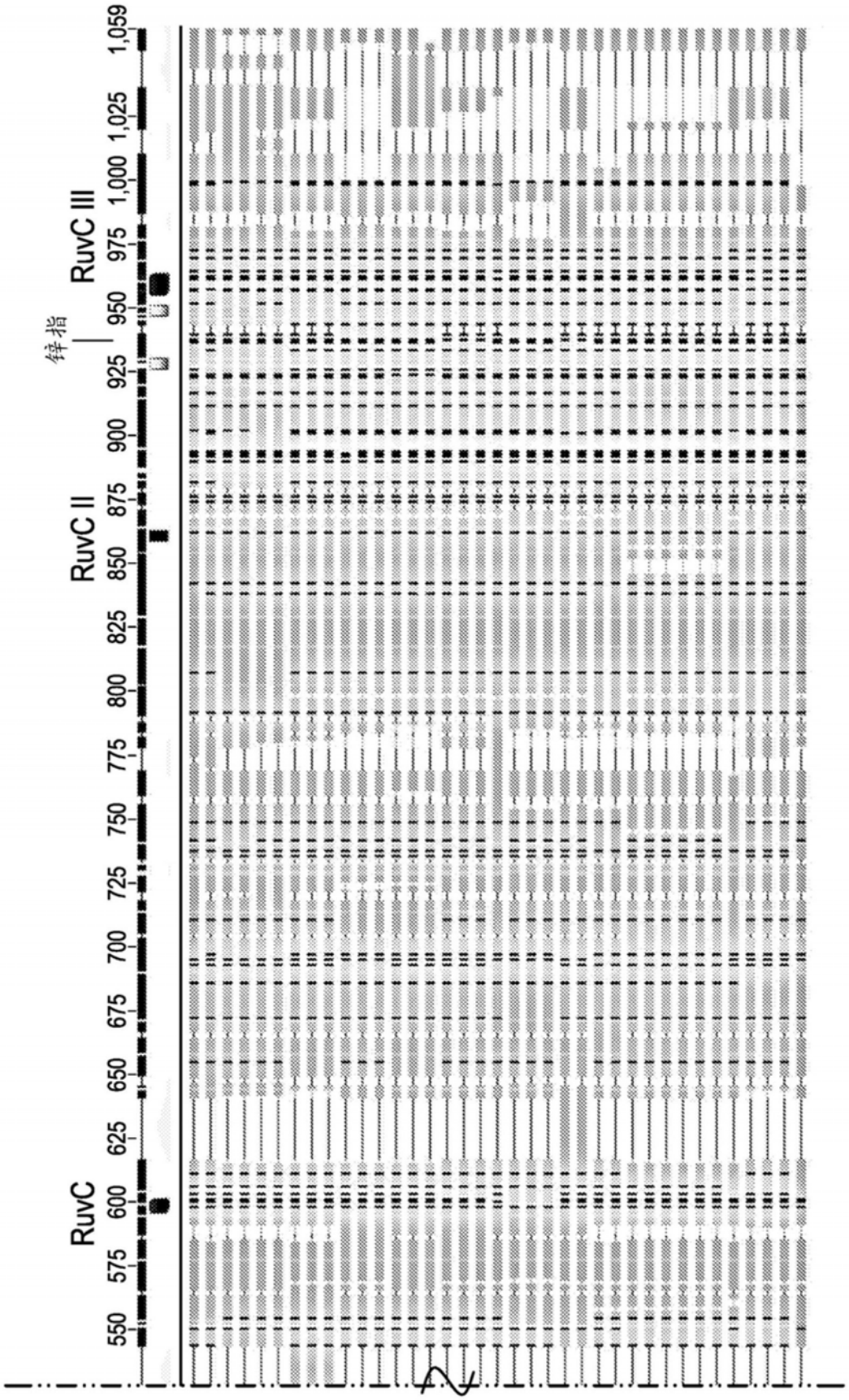

RuvC
RuvC II
RuvC III
锌指
550
575
600
625
650
675
700
725
750
775
800
825
850
875
900
925
950
975
1,000
1,025
1,059

图2E(部分2)

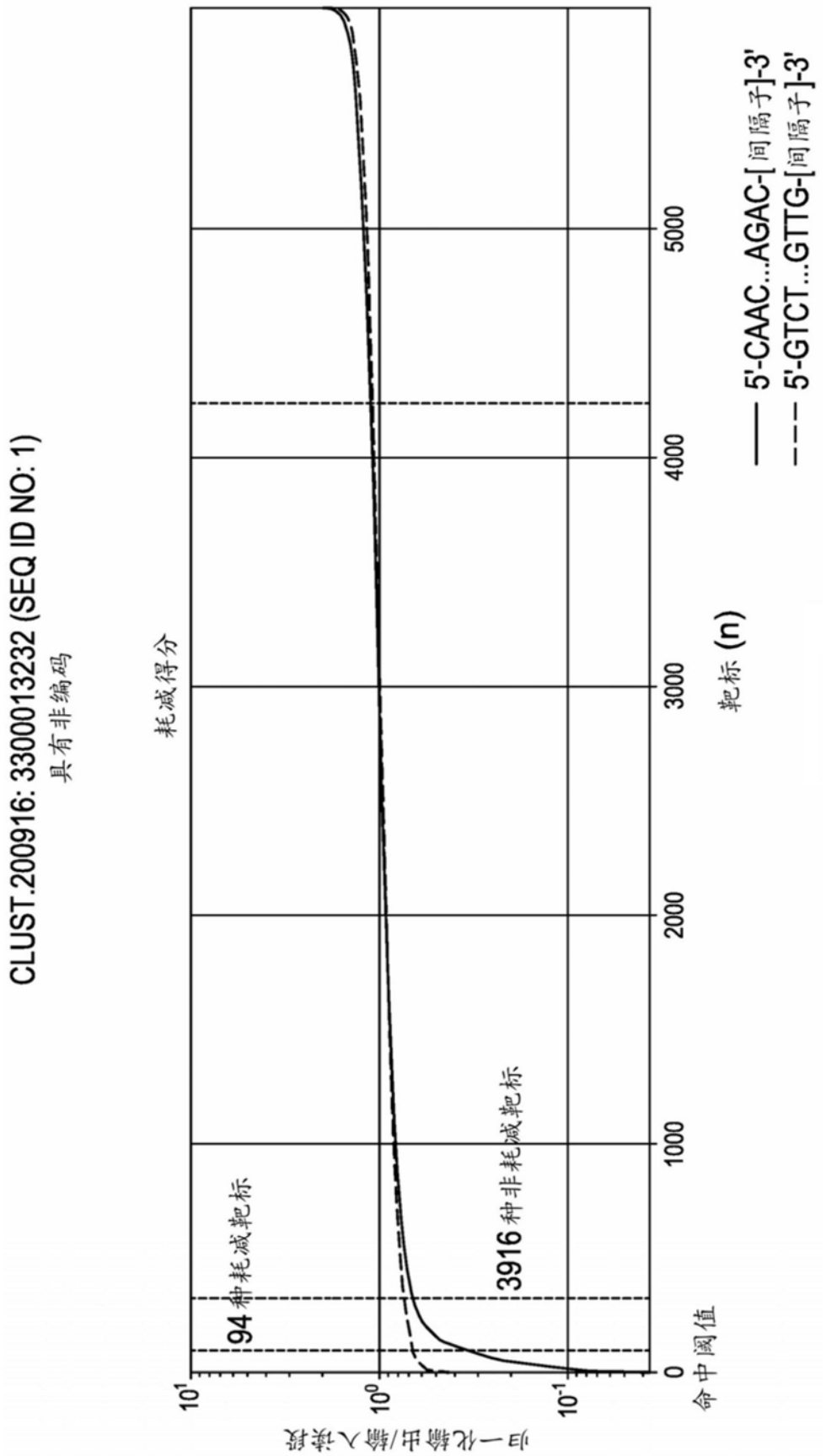

CLUST.200916: 3300013232 (SEQ ID NO: 1)
具有非编码
耗减得分
94 种耗减靶标
3916 种非耗减靶标
归一化输出/输入读段
命中阈值
靶标 (n)
5'-CAAC...AGAC-[间隔子]-3'
5'-GTCT...GTTG-[间隔子]-3'

图3

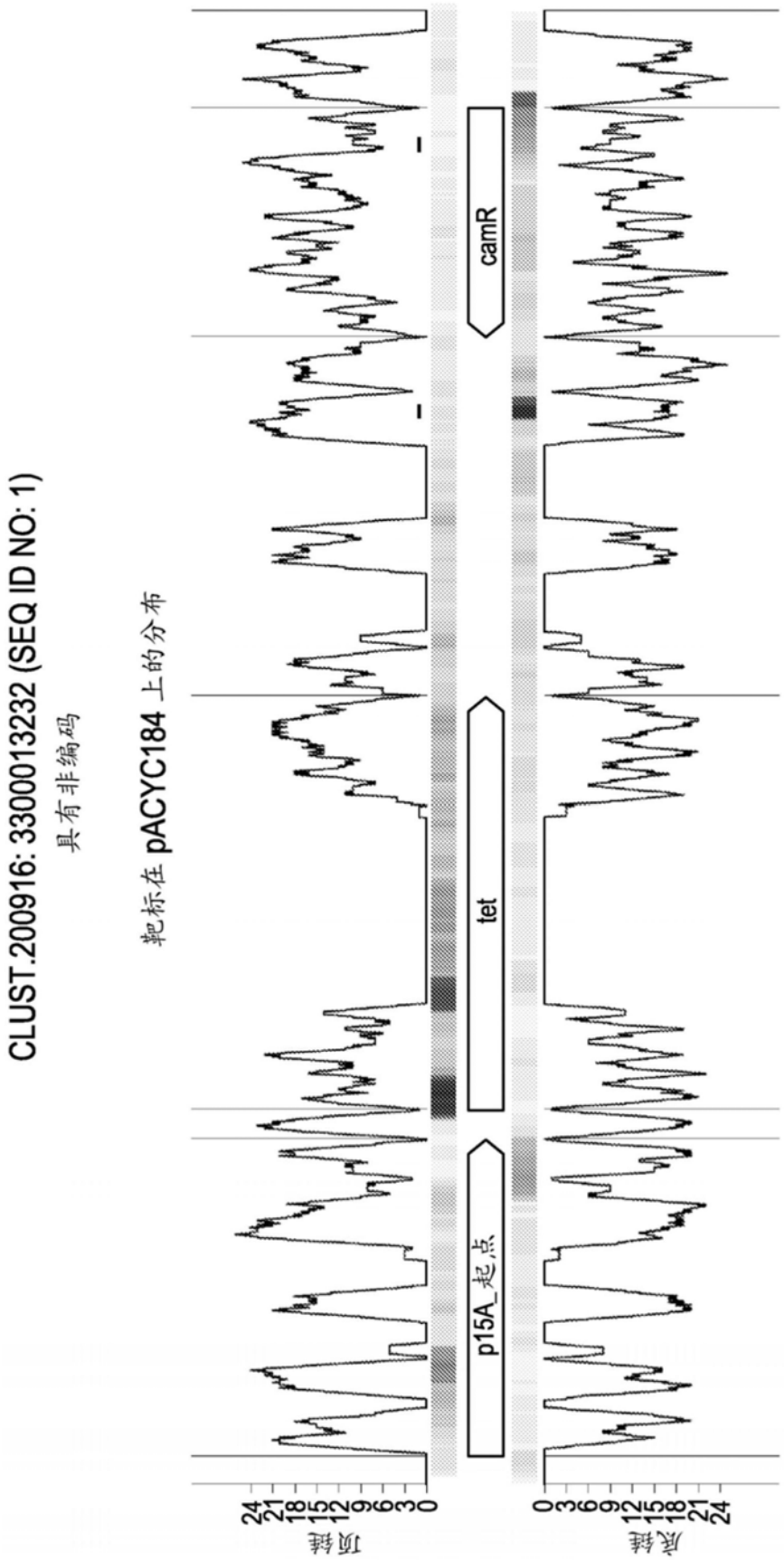

CLUST.200916: 3300013232 (SEQ ID NO: 1)
具有非编码
靶标在 pACYC184 上的分布
顶链
底链
24 21 18 15 12 9 6 3 0
0 3 6 9 12 15 18 21 24
p15A_起点
tet
camR

图4A

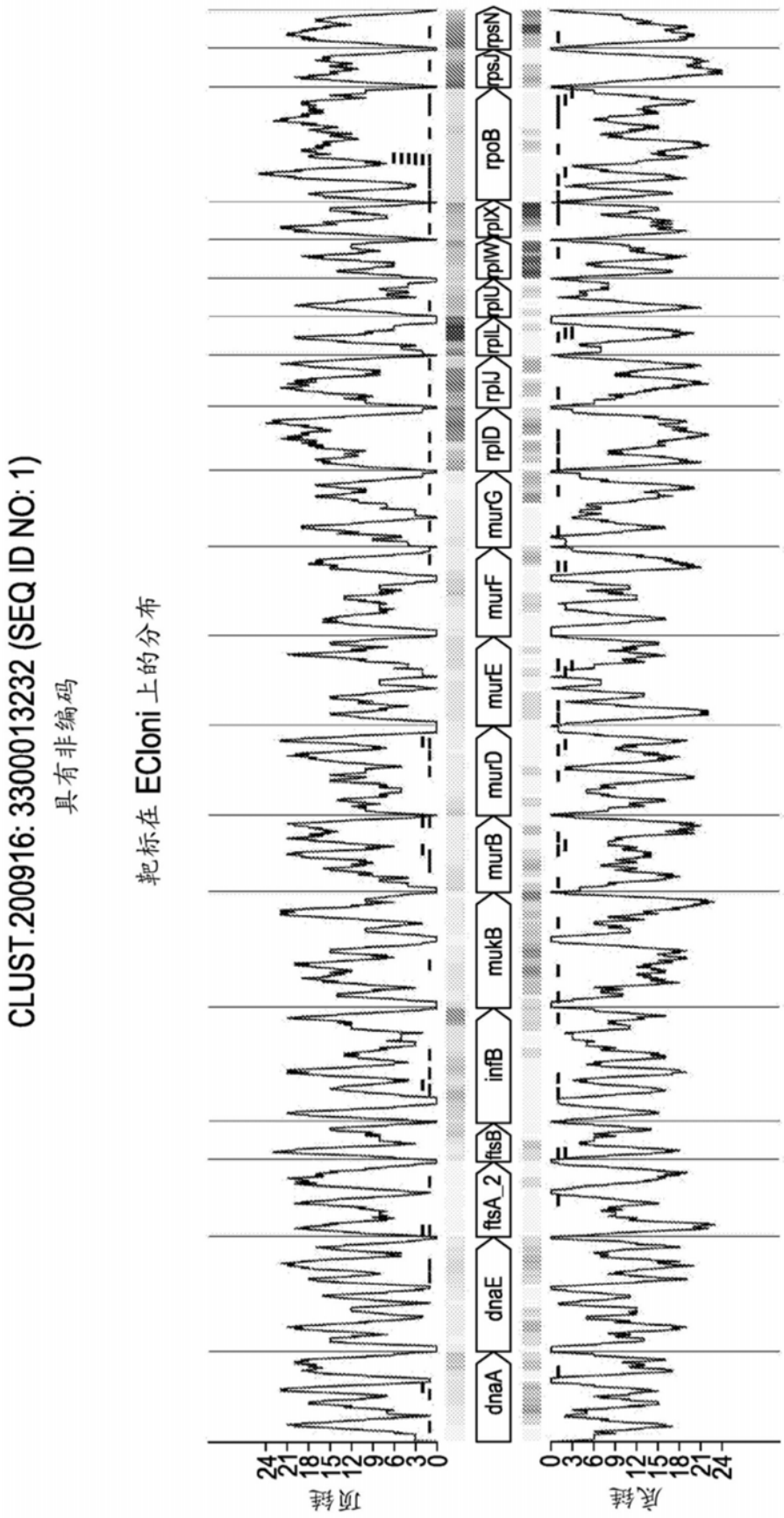

CLUST.200916: 3300013232 (SEQ ID NO: 1)
具有非编码
靶标在 ECloni 上的分布
dnaA
dnaE
ftsA_2
ftsB
infB
mukB
murB
murD
murE
murF
murG
rplD
rplJ
rplL
rplU
rplW
rplX
rpoB
rpsJ
rpsN
顶链
底链

图4B

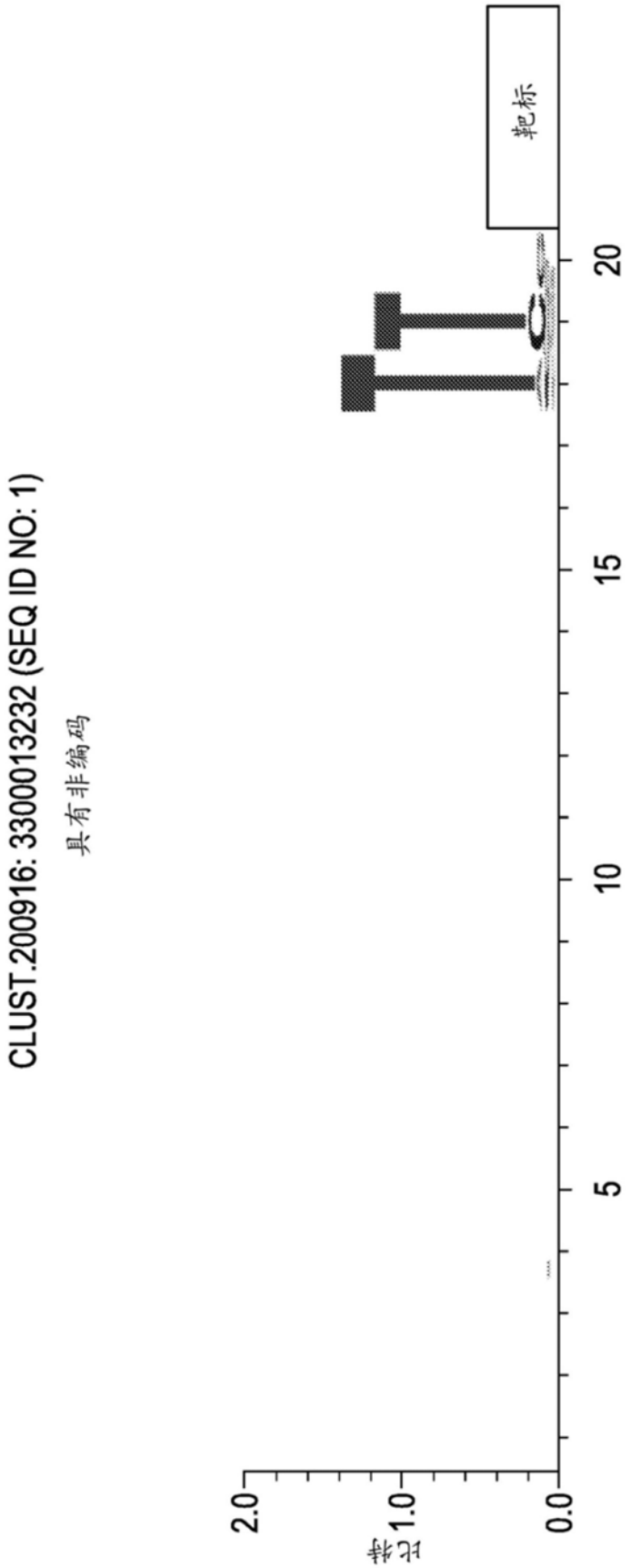

CLUST.200916: 3300013232 (SEQ ID NO: 1)
具有非编码
靶标
比特
2.0
1.0
0.0
5
10
15
20

图5

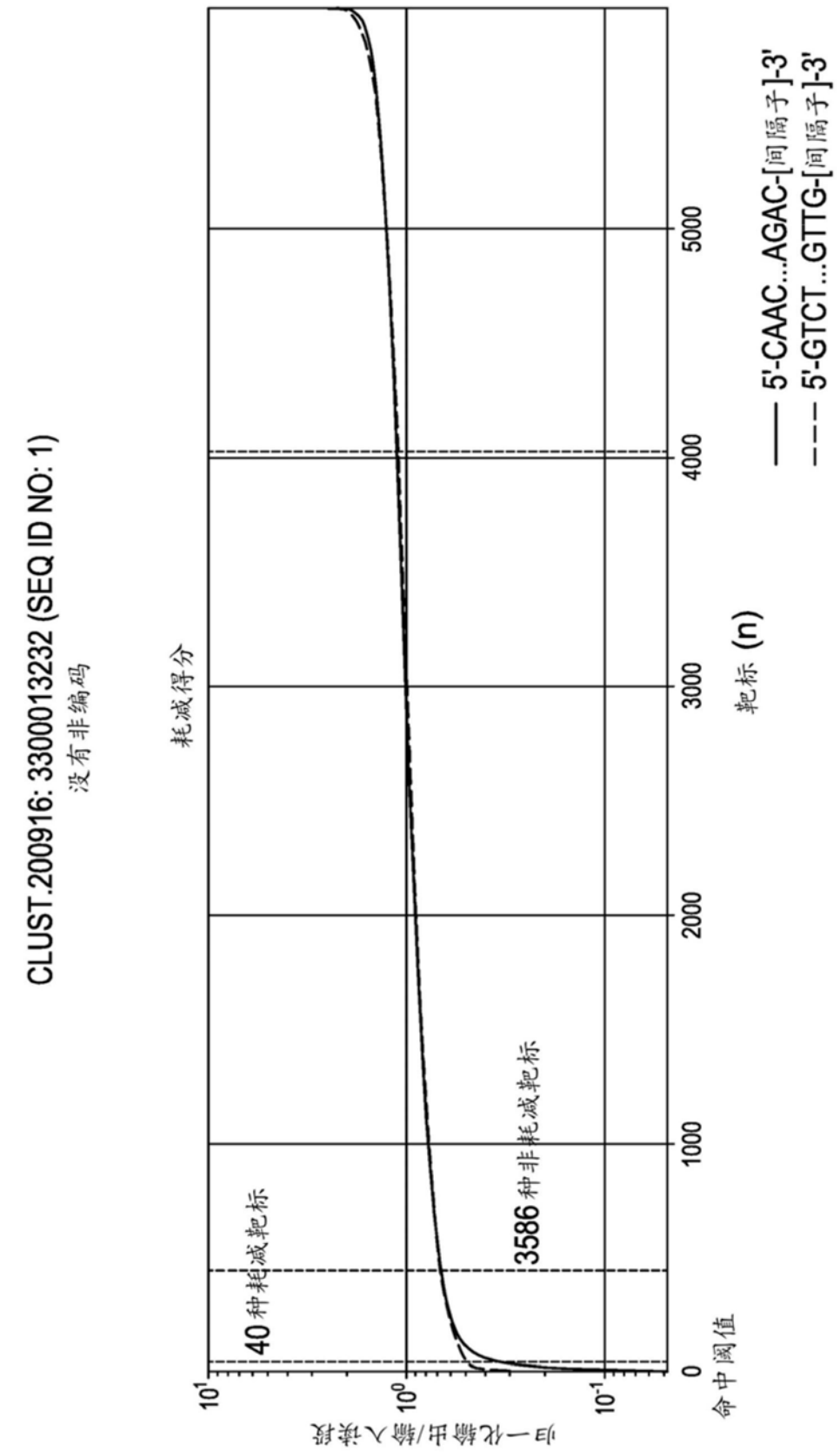

CLUST.200916: 3300013232 (SEQ ID NO: 1)
没有非编码
耗减得分
归一化输出/输入读段
10^1
10^0
10^{-1}
40 种耗减靶标
3586 种非耗减靶标
0
1000
2000
3000
4000
5000
命中阈值
靶标 (n)
5'-CAAC...AGAC-[间隔子]-3'
5'-GTCT...GTTG-[间隔子]-3'

图6

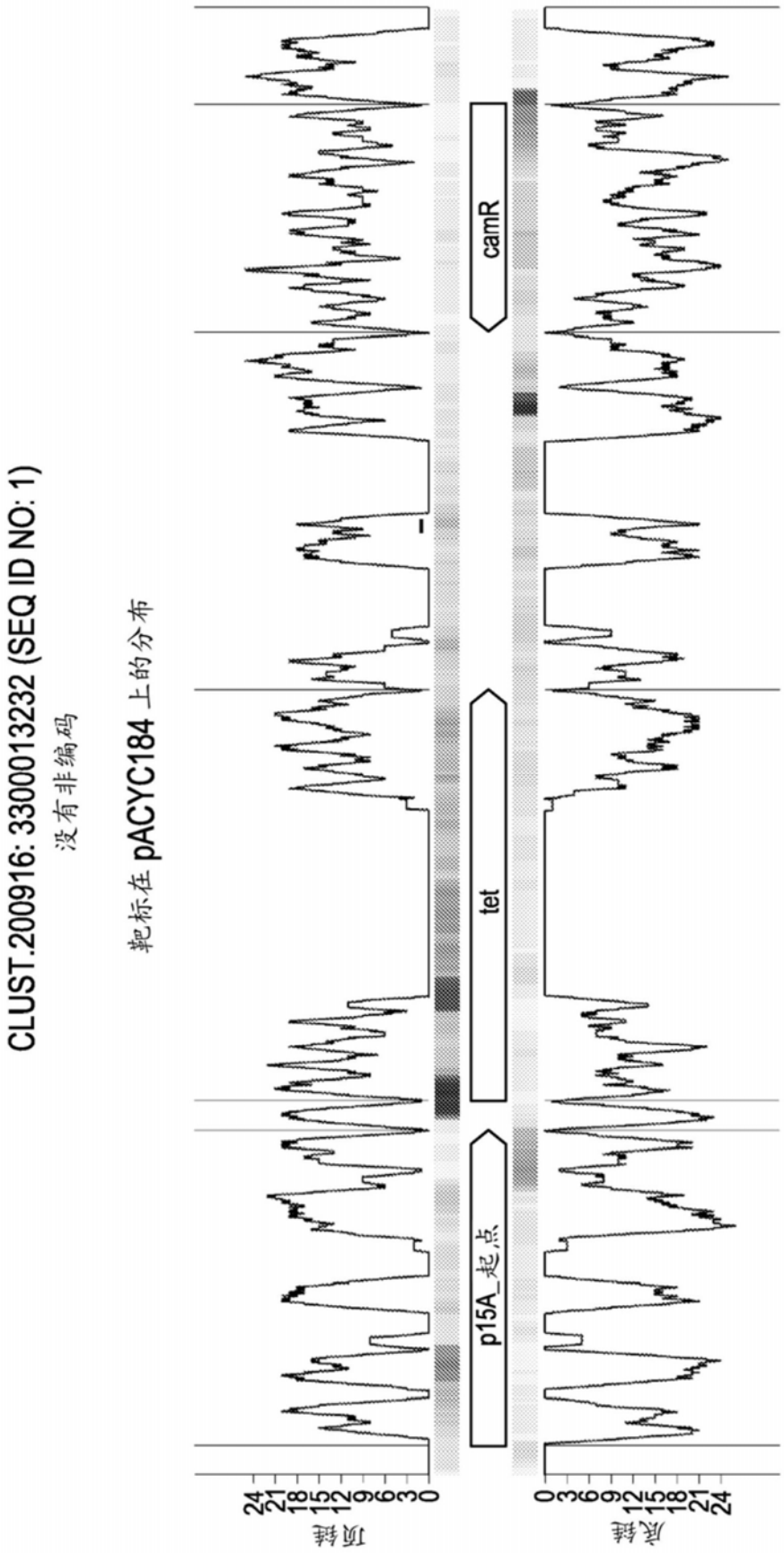

CLUST.200916: 3300013232 (SEQ ID NO: 1)
没有非编码
靶标在 pACYC184 上的分布
顶链
底链
24 21 18 15 12 9 6 3 0
0 3 6 9 12 15 18 21 24
p15A_起点
tet
camR

图7A

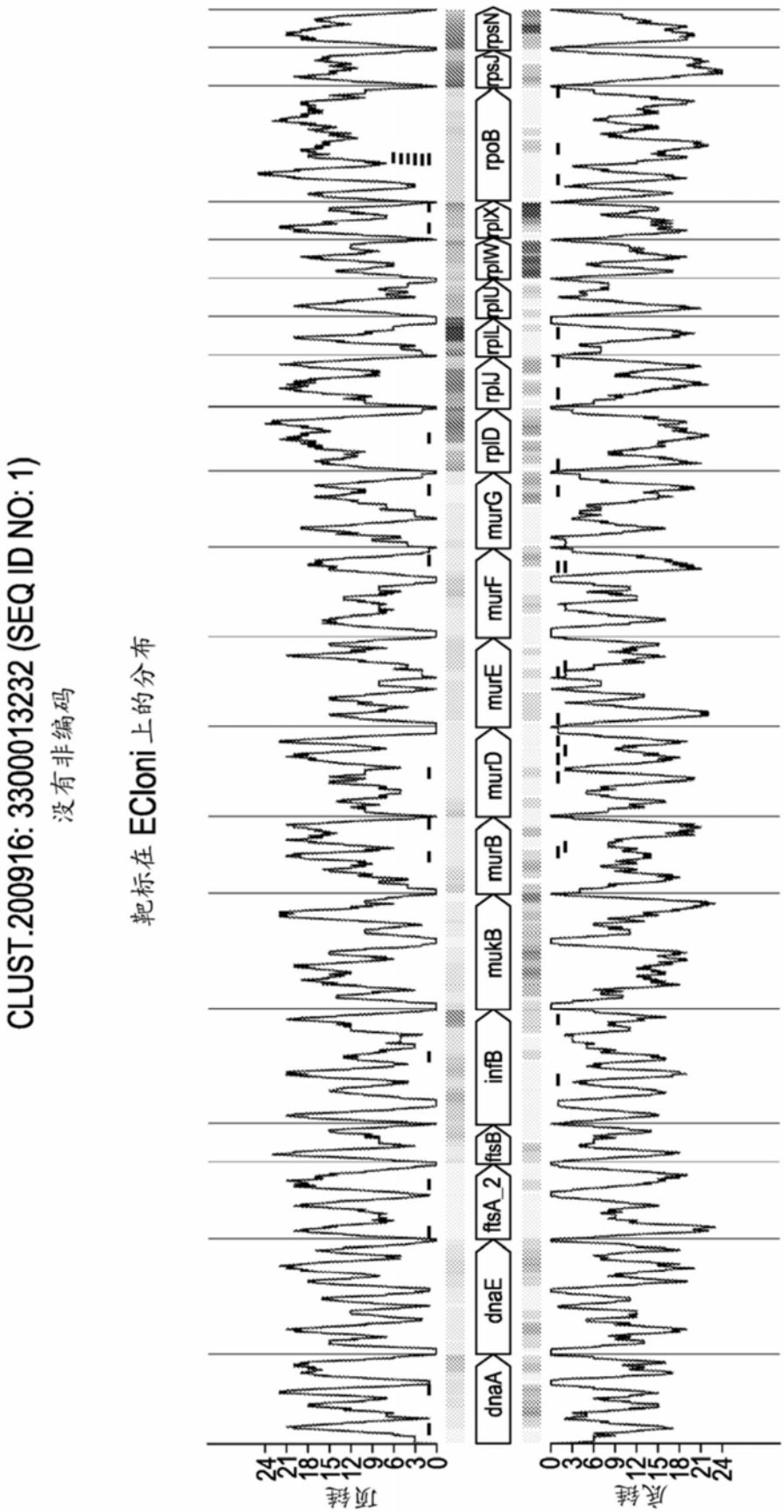

CLUST.200916: 3300013232 (SEQ ID NO: 1)
没有非编码
靶标在 ECIoni 上的分布
dnaA
dnaE
ftsA_2
ftsB
infB
mukB
murB
murD
murE
murF
murG
rplD
rplJ
rplL
rplU
rplW
rplX
rpoB
rpsJ
rpsN
顶链
底链
0 3 6 9 12 15 18 21 24

图7B

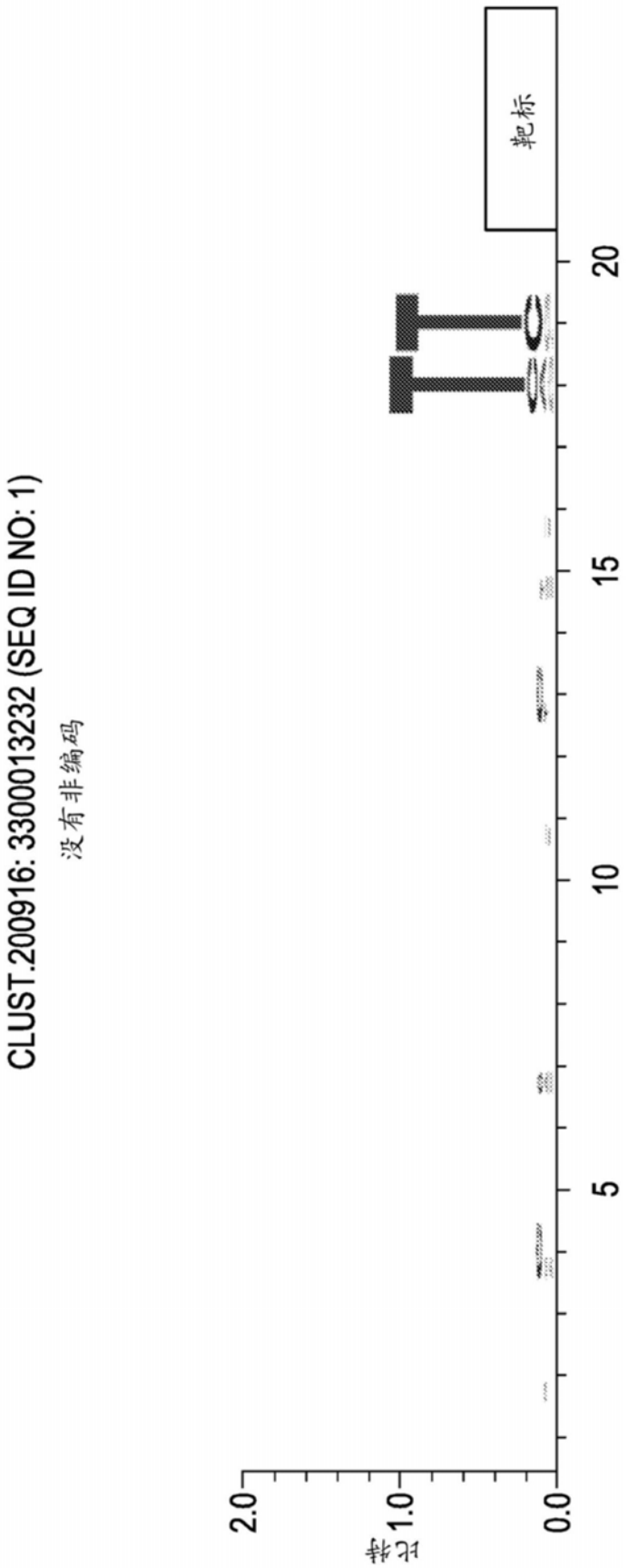

CLUST.200916: 3300013232 (SEQ ID NO: 1)
没有非编码
比特
2.0
1.0
0.0
5
10
15
20
靶标

图8

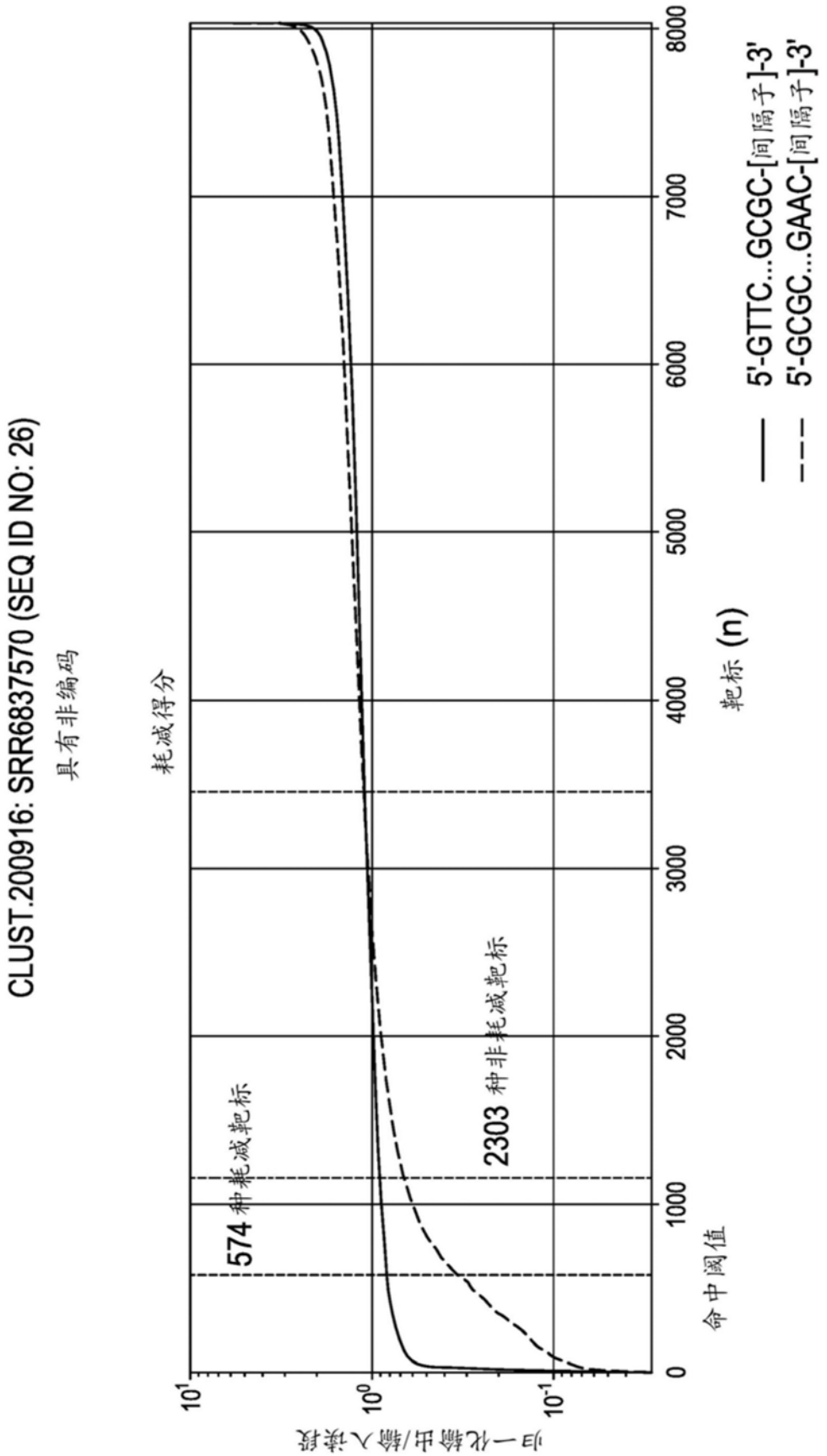

CLUST.200916: SRR6837570 (SEQ ID NO: 26)
具有非编码
耗减得分
574 种耗减靶标
2303 种非耗减靶标
归一化输出/输入读段
命中阈值
靶标 (n)
5'-GTTC...GCGC-[间隔子]-3'
5'-GCGC...GAAC-[间隔子]-3'

图9

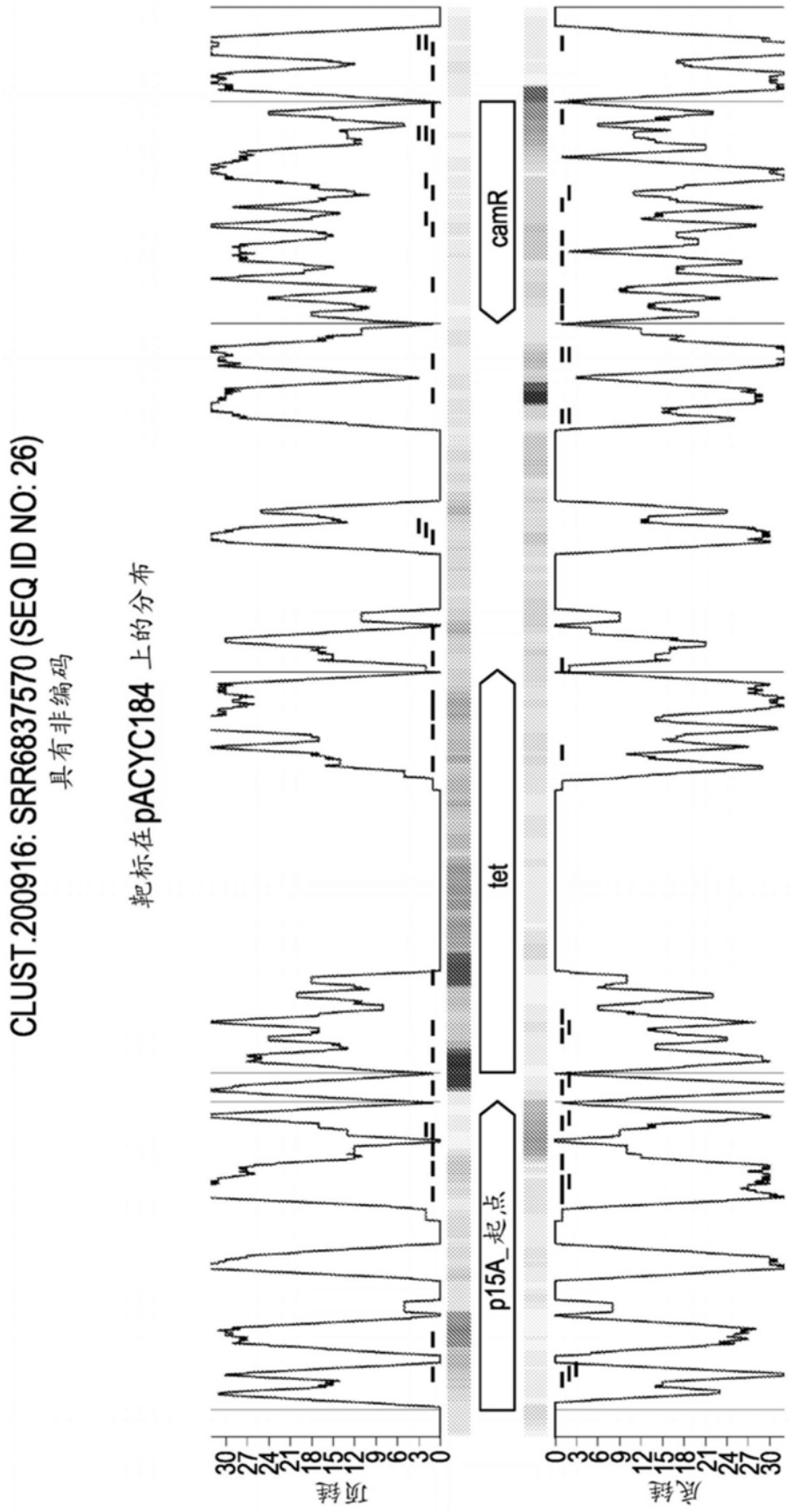

CLUST.200916: SRR6837570 (SEQ ID NO: 26)
具有非编码
靶标在pACYC184 上的分布
顶链
底链
30
27
24
21
18
15
12
9
6
3
0
p15A_起点
tet
camR

图10A

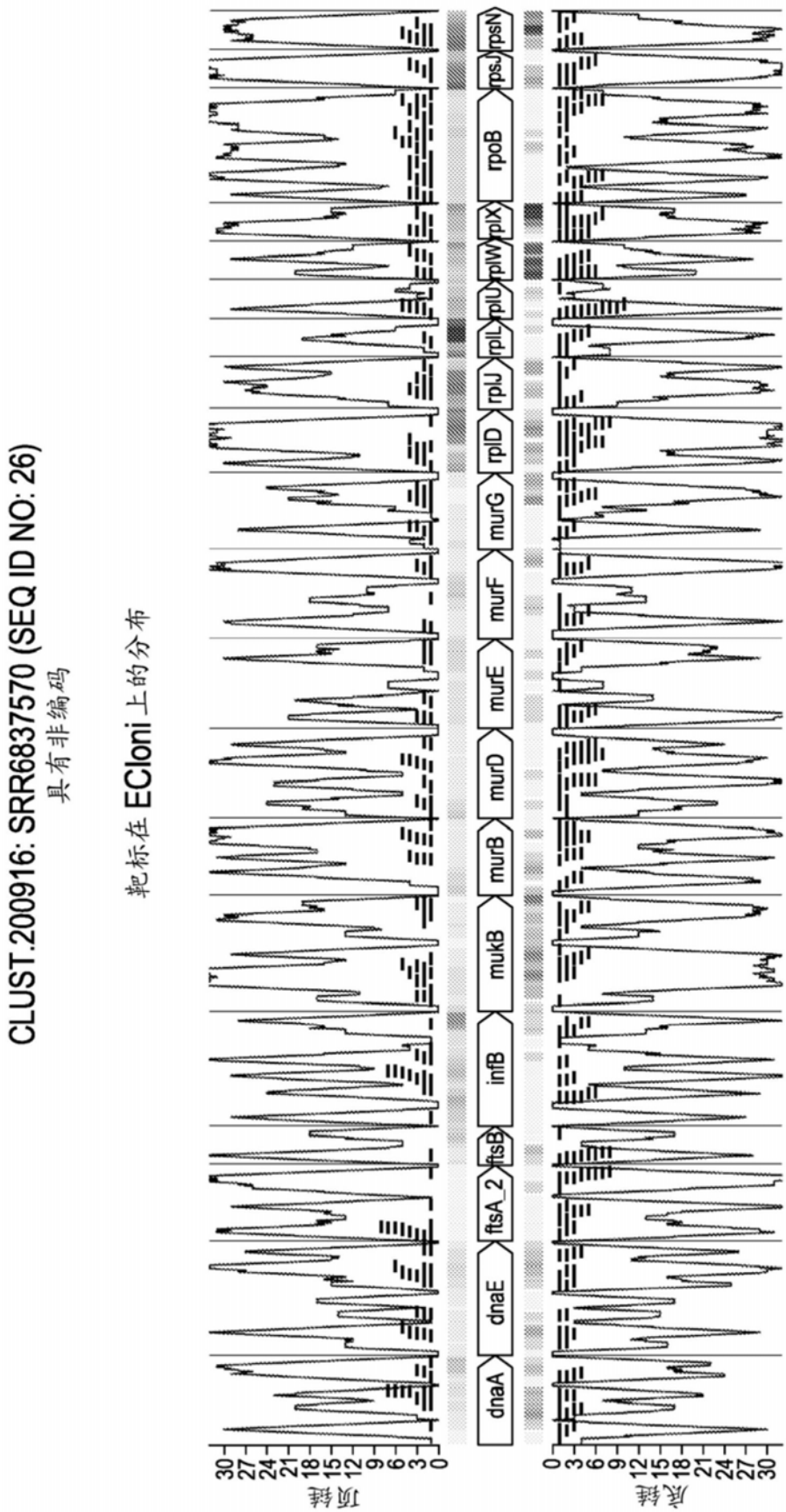

CLUST.200916: SRR6837570 (SEQ ID NO: 26)
具有非编码
靶标在 ECloni 上的分布
dnaA
dnaE
ftsA_2
ftsB
infB
mukB
murB
murD
murE
murF
murG
rplD
rplJ
rplL
rplU
rplW
rplX
rpoB
rpsJ
rpsN
顶链
底链

图10B

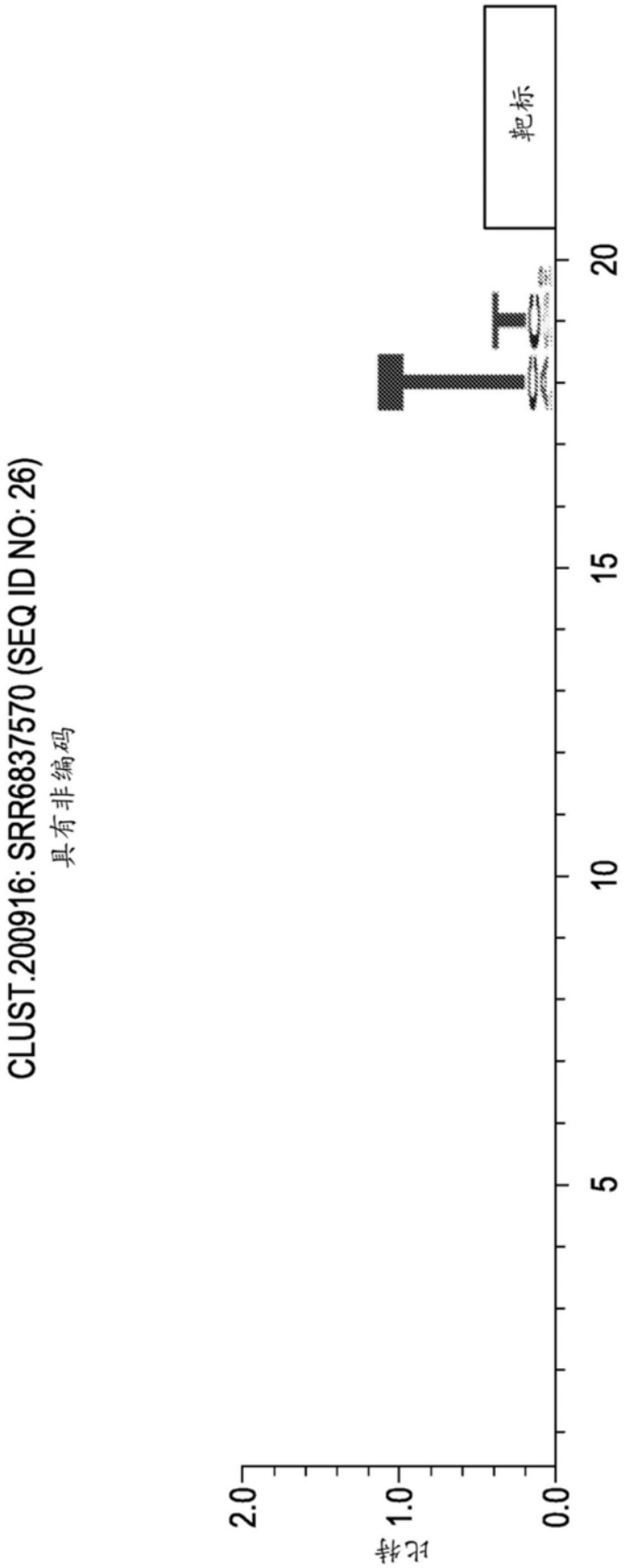

CLUST.200916: SRR6837570 (SEQ ID NO: 26)
具有非编码
靶标
比特
2.0
1.0
0.0
5
10
15
20

图11

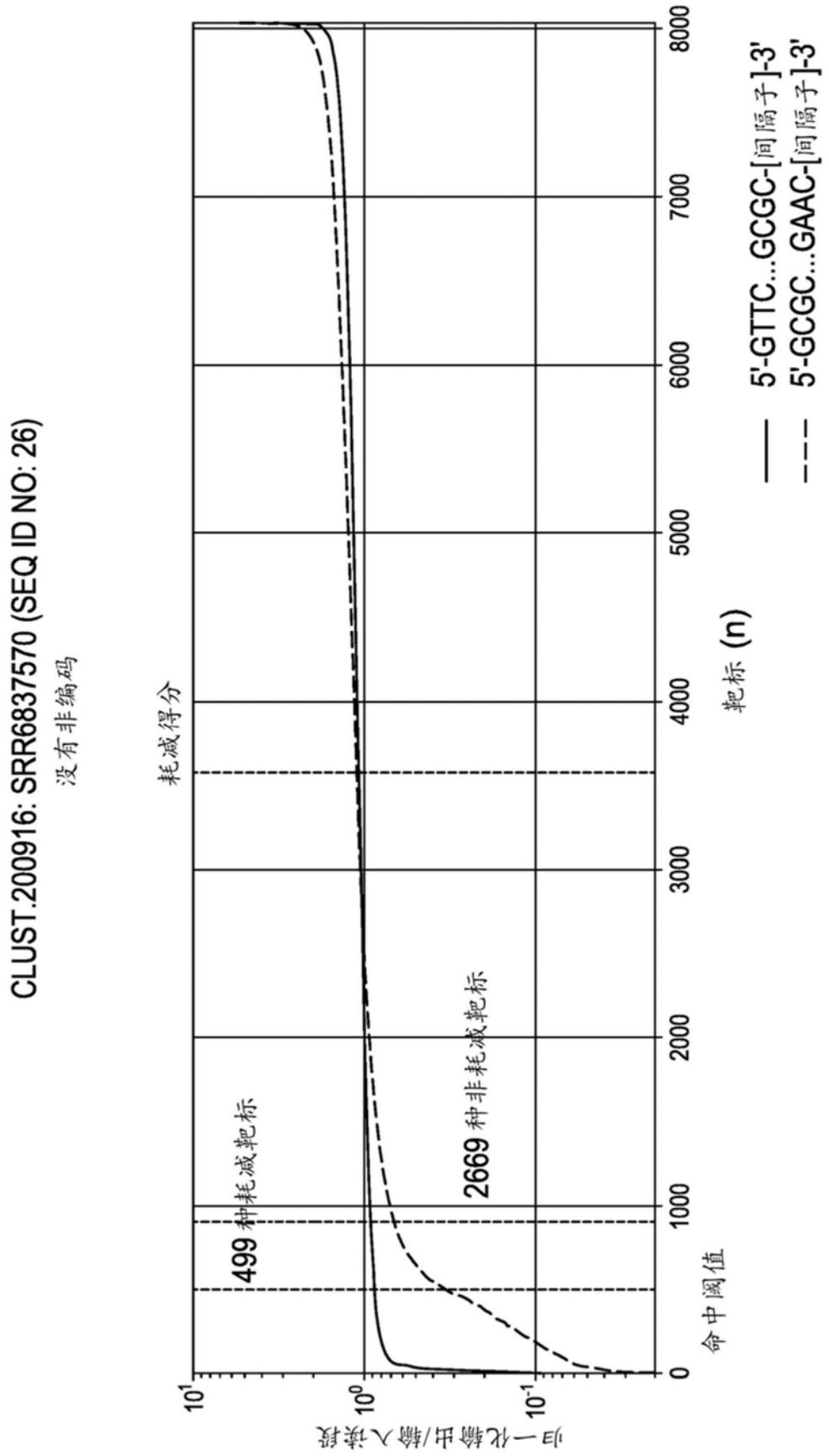

CLUST.200916: SRR6837570 (SEQ ID NO: 26)
没有非编码
耗减得分
499 种耗减靶标
2669 种非耗减靶标
归一化输出/输入读段
命中阈值
靶标 (n)
5'-GTTC...GCGC-[间隔子]-3'
5'-GCGC...GAAC-[间隔子]-3'

图12

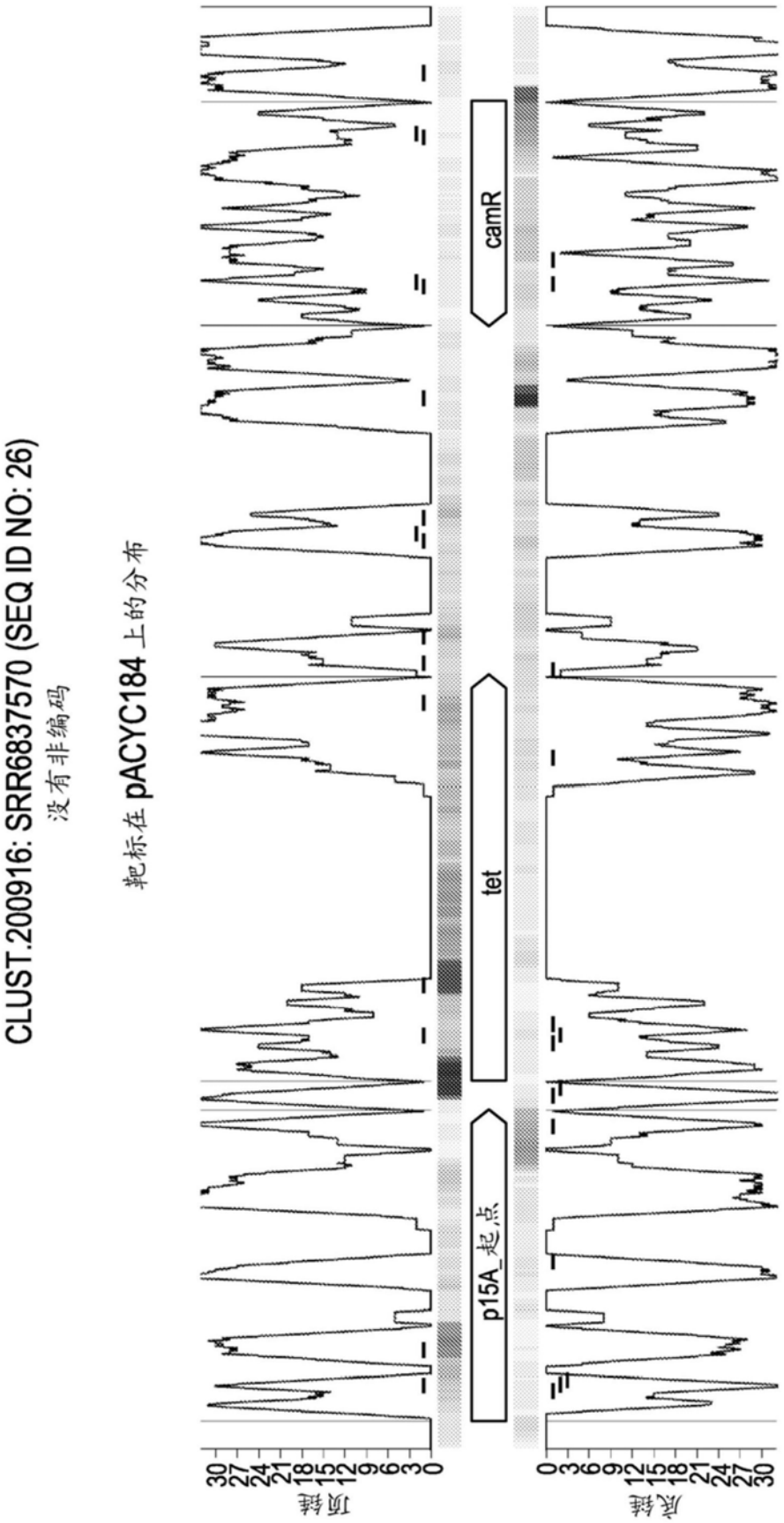

CLUST.200916: SRR6837570 (SEQ ID NO: 26)
没有非编码
靶标在 pACYC184 上的分布
顶链
底链
30 27 24 21 18 15 12 9 6 3 0
0 3 6 9 12 15 18 21 24 27 30
p15A_起点
tet
camR

图13A

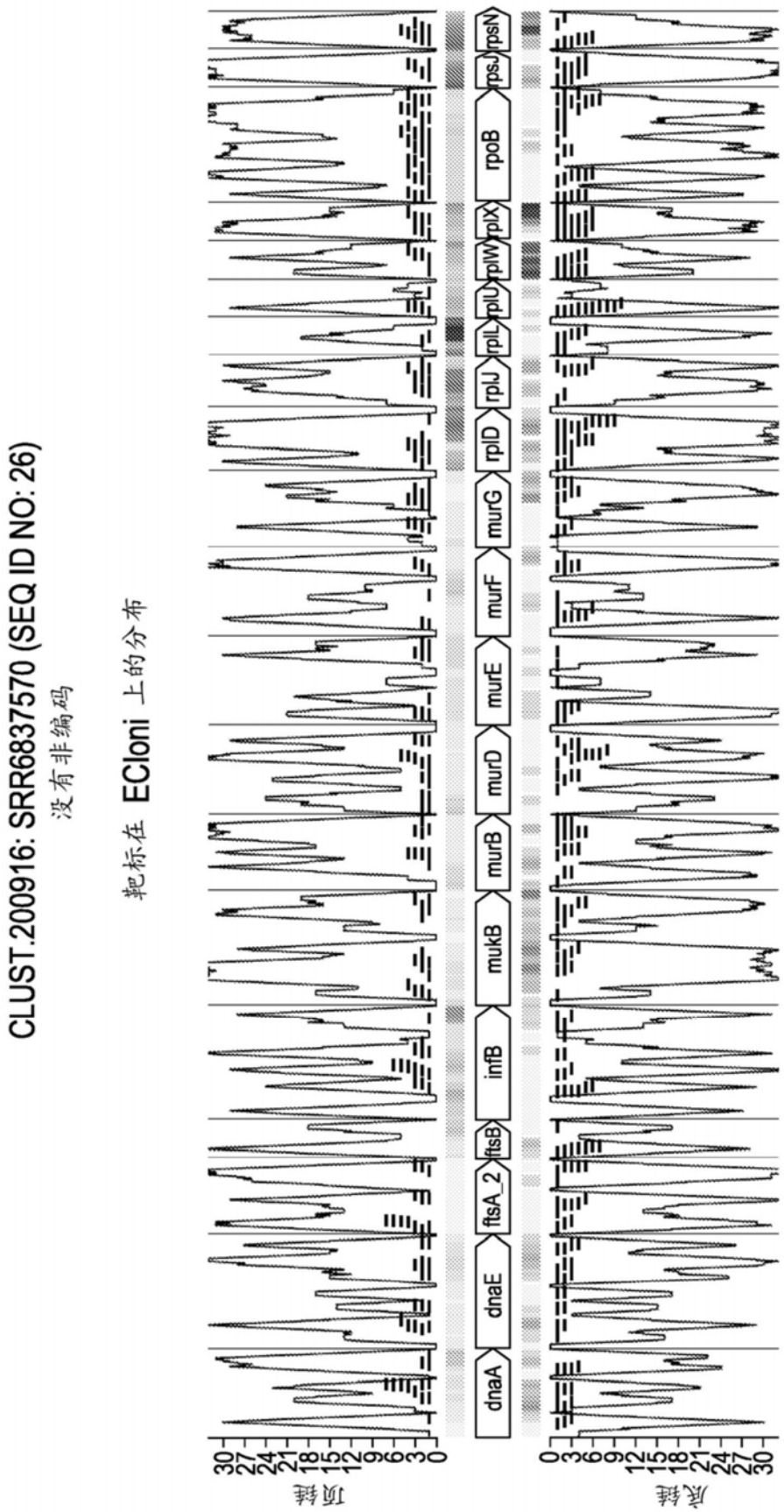

CLUST.200916: SRR6837570 (SEQ ID NO: 26)
没有非编码
靶标在 ECloni 上的分布
顶链
底链
dnaA
dnaE
ftsA_2
ftsB
infB
mukB
murB
murD
murE
murF
murG
rplD
rplJ
rplL
rplU
rplW
rplX
rpoB
rpsJ
rpsN

图13B

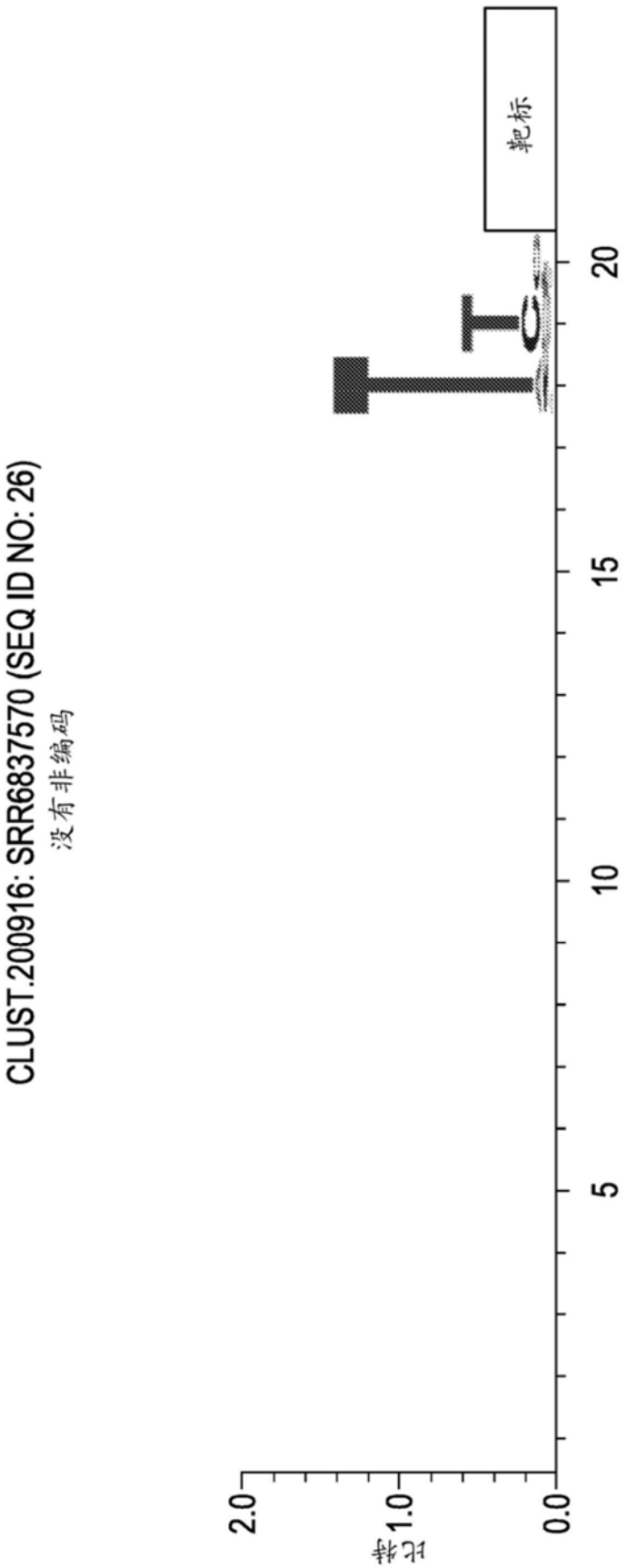

CLUST.200916: SRR6837570 (SEQ ID NO: 26)
没有非编码
比特
2.0
1.0
0.0
5
10
15
20
靶标

图14

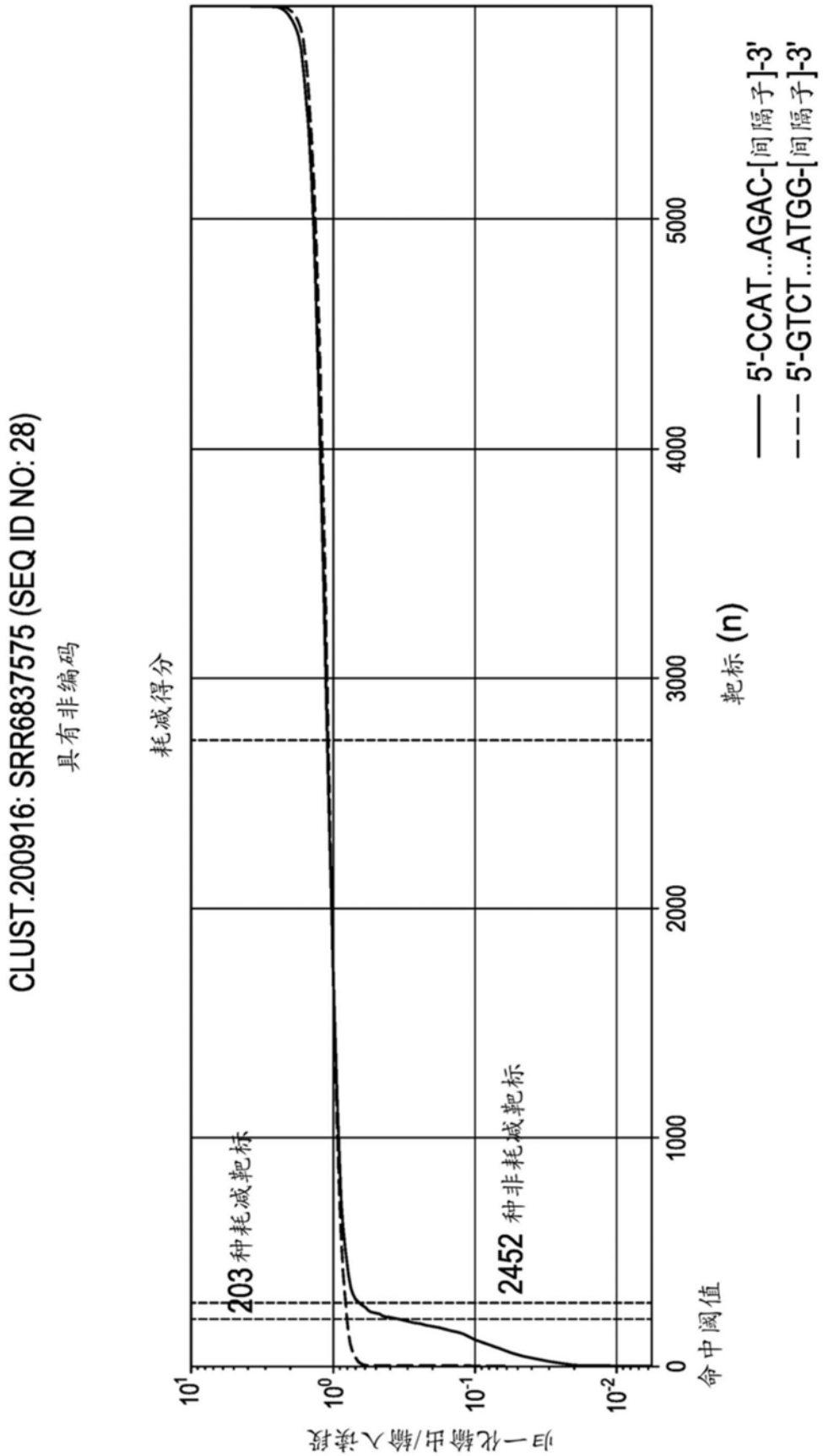

CLUST.200916: SRR6837575 (SEQ ID NO: 28)
具有非编码
耗减得分
203 种耗减靶标
2452 种非耗减靶标
归一化输出/输入读段
命中阈值
靶标 (n)
5'-CCAT...AGAC-[间隔子]-3'
5'-GTCT...ATGG-[间隔子]-3'

图15

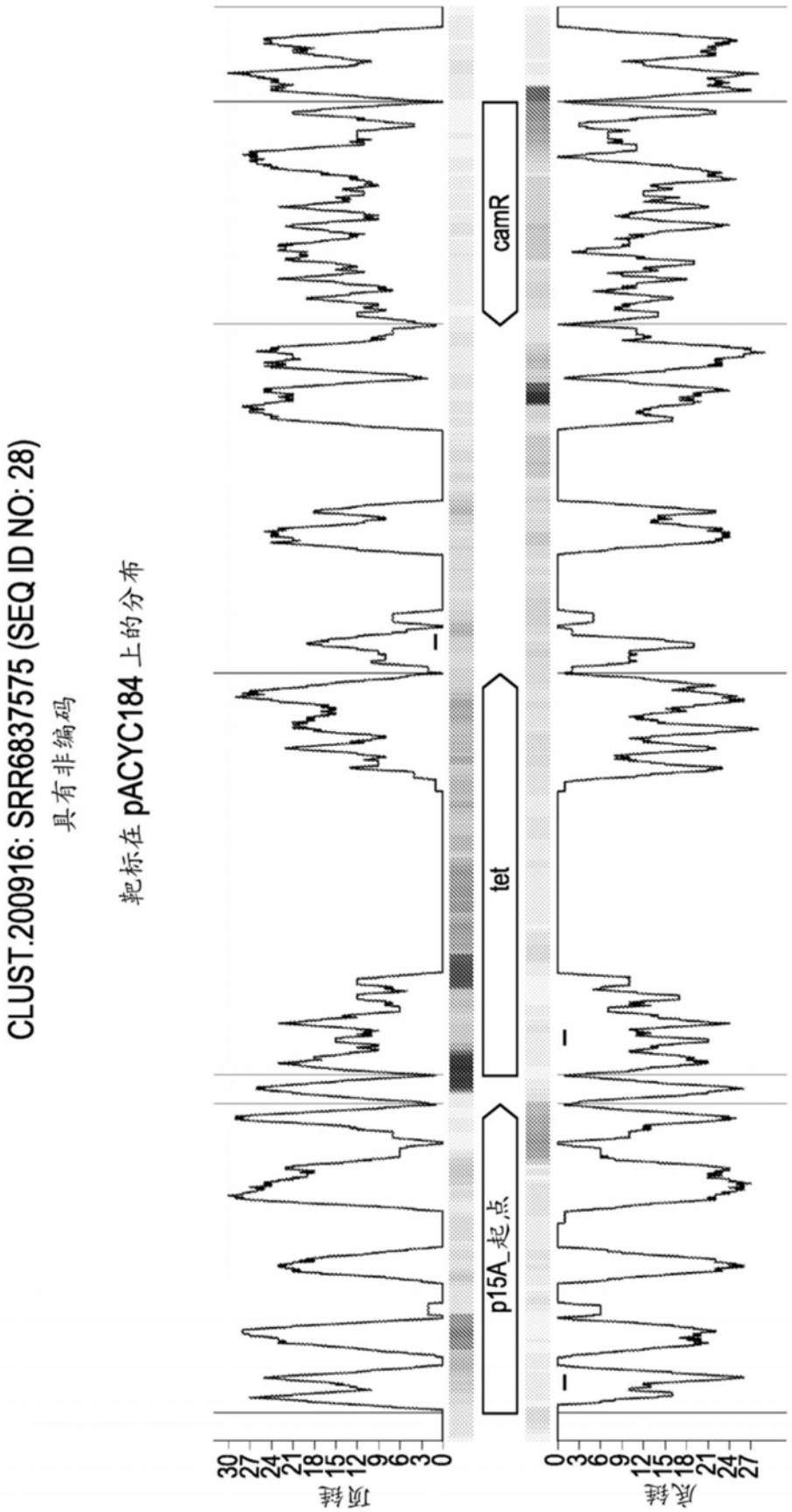

CLUST.200916: SRR6837575 (SEQ ID NO: 28)
具有非编码
靶标在 pACYC184 上的分布
顶链
底链
p15A_起点
tet
camR
30
27
24
21
18
15
12
9
6
3
0

图16A

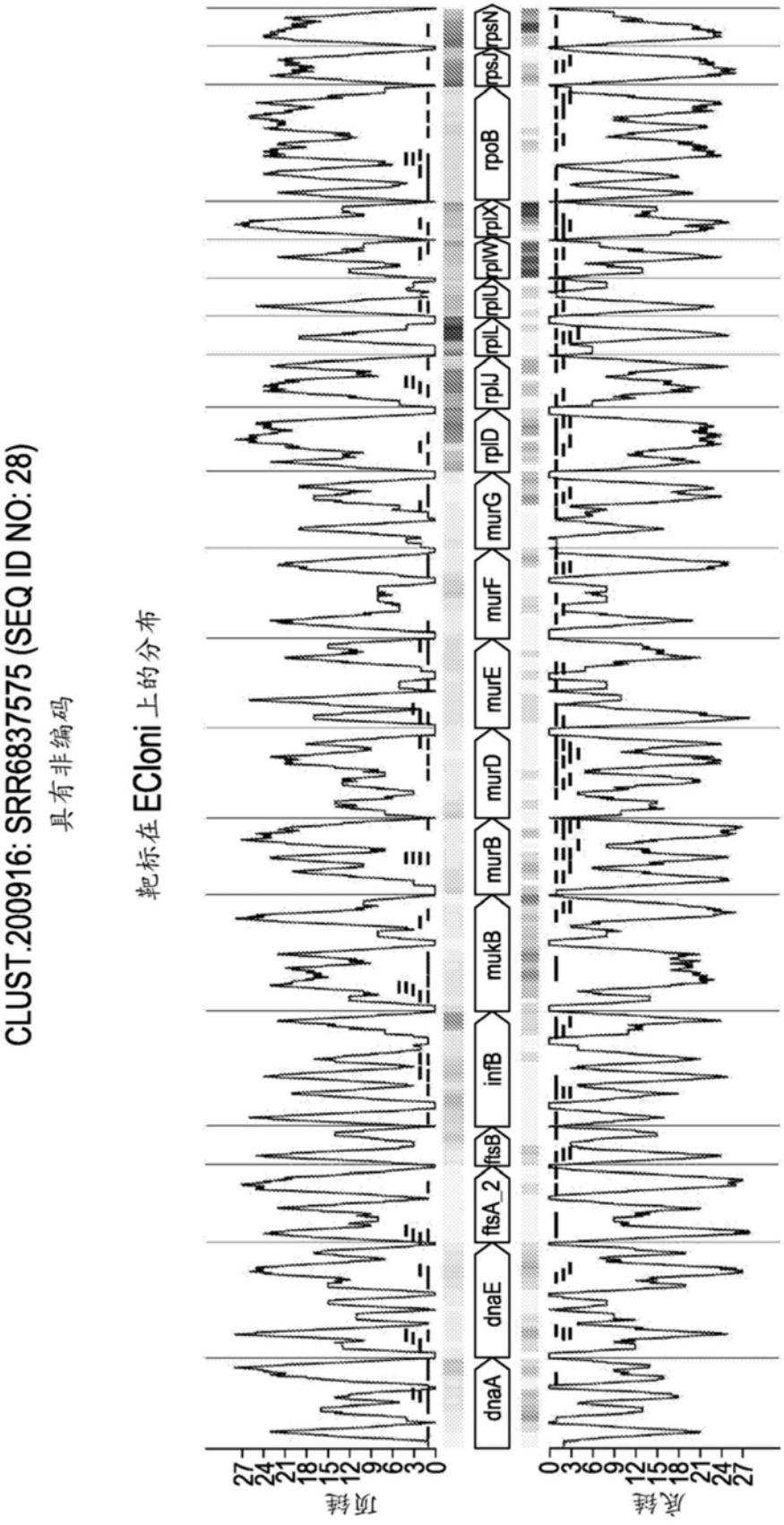

CLUST.200916: SRR6837575 (SEQ ID NO: 28)
具有非编码
靶标在 ECloni 上的分布
dnaA
dnaE
ftsA_2
ftsB
infB
mukB
murB
murD
murE
murF
murG
rplD
rplJ
rplL
rplU
rplW
rplX
rpoB
rpsJ
rpsN
顶链
底链
0 3 6 9 12 15 18 21 24 27
27 24 21 18 15 12 9 6 3 0

图16B

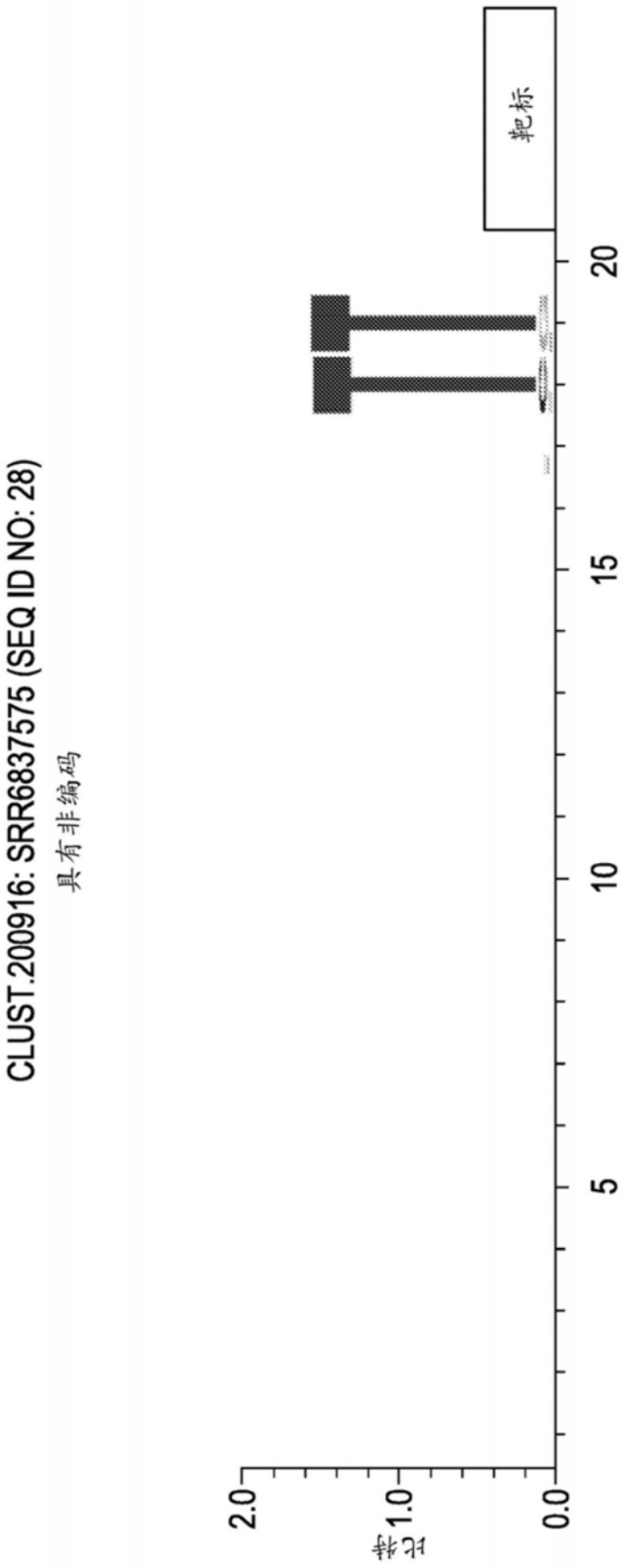

CLUST.200916: SRR6837575 (SEQ ID NO: 28)
具有非编码
靶标
2.0
1.0
0.0
比特
5
10
15
20

图17

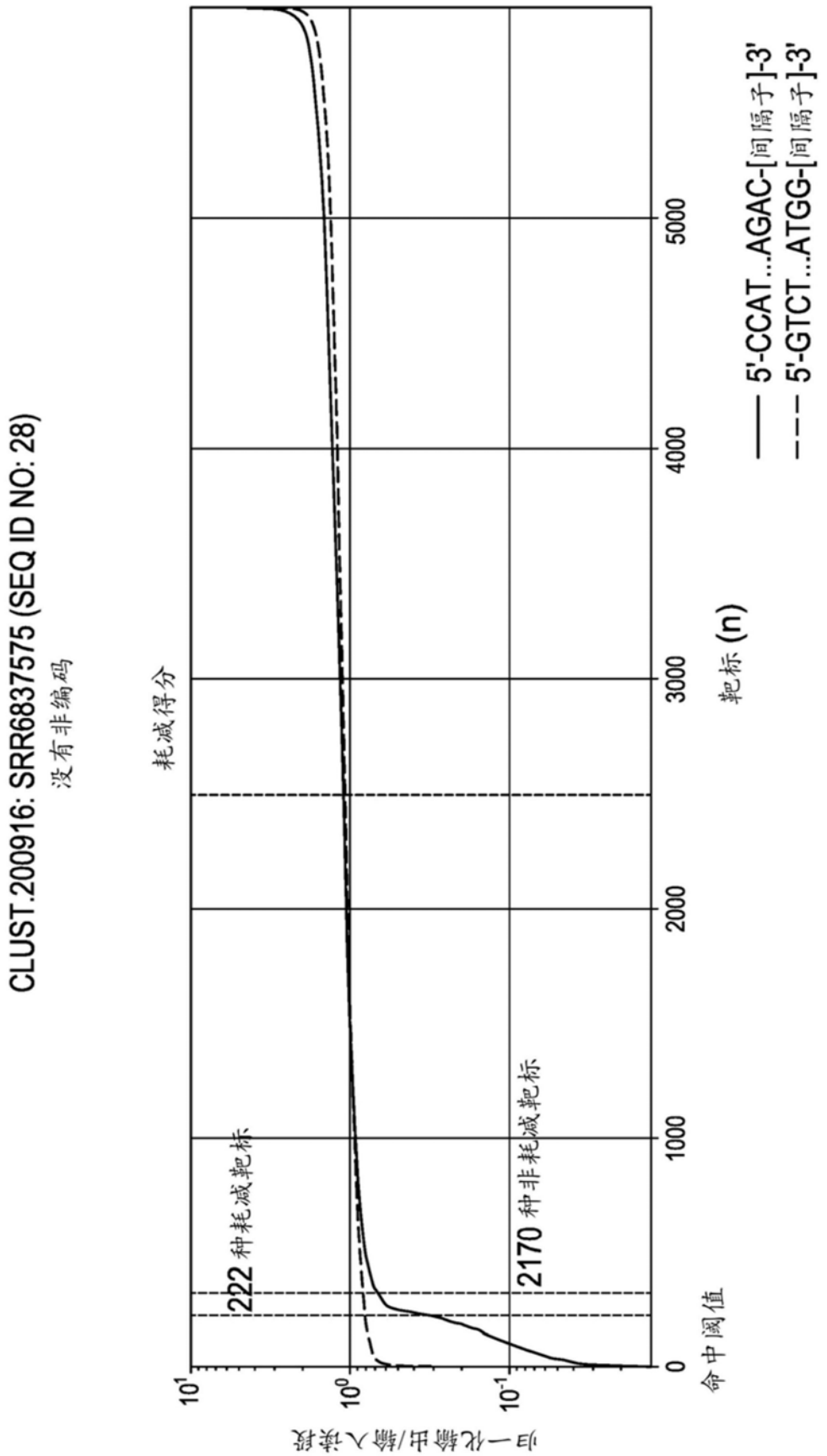

CLUST.200916: SRR6837575 (SEQ ID NO: 28)
没有非编码
耗减得分
222 种耗减靶标
2170 种非耗减靶标
归一化输出/输入读段
命中阈值
靶标 (n)
5'-CCAT...AGAC-[间隔子]-3'
5'-GTCT...ATGG-[间隔子]-3'

图18

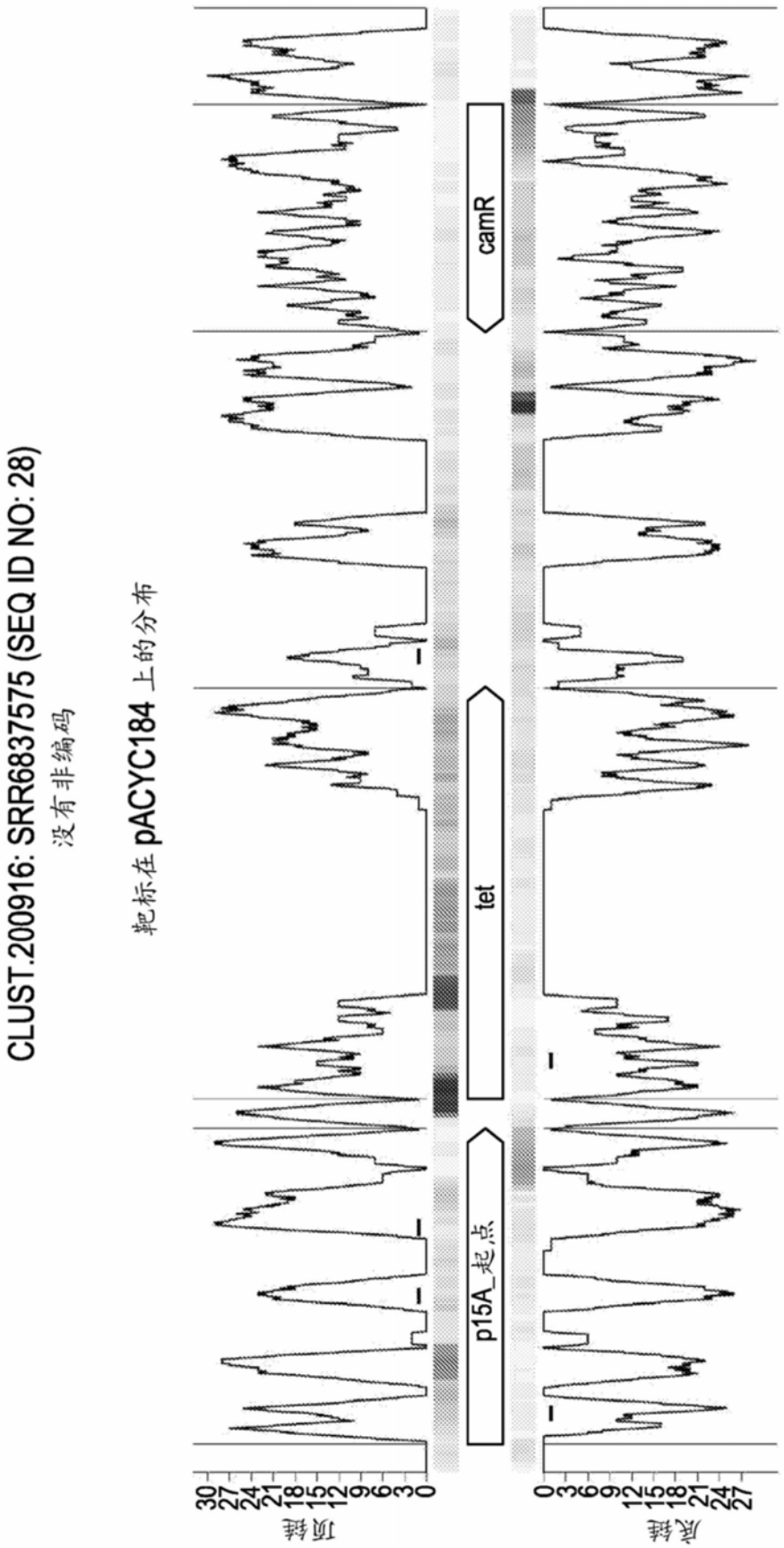

CLUST.200916: SRR6837575 (SEQ ID NO: 28)
没有非编码
靶标在 pACYC184 上的分布
顶链
底链
p15A_起点
tet
camR
30
27
24
21
18
15
12
9
6
3
0

图19A

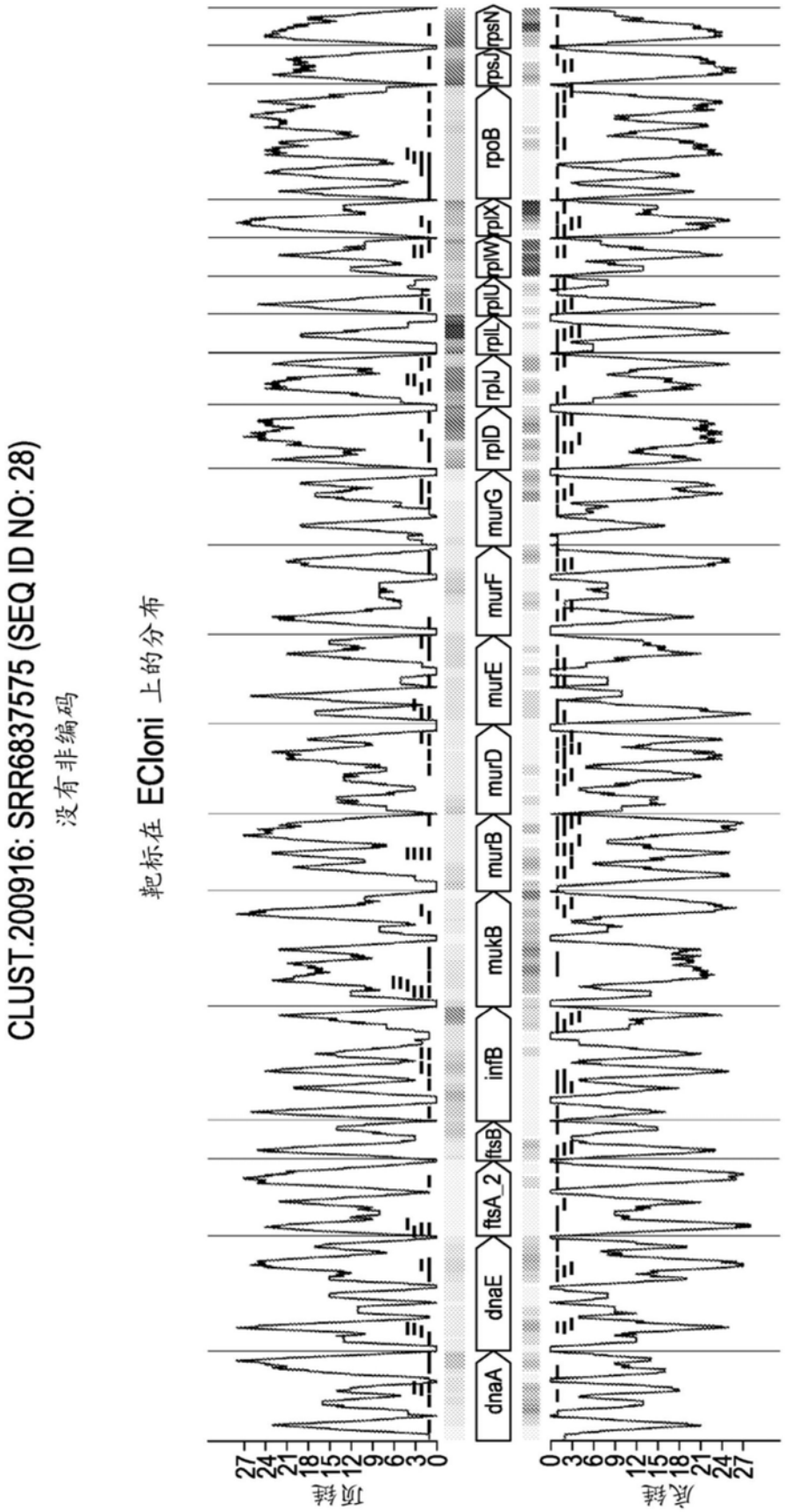

CLUST.200916: SRR6837575 (SEQ ID NO: 28)
没有非编码
靶标在 ECloni 上的分布
dnaA
dnaE
ftsA_2
ftsB
infB
mukB
murB
murD
murE
murF
murG
rplD
rplJ
rplL
rplU
rplW
rplX
rpoB
rpsJ
rpsN
顶链
底链
0 3 6 9 12 15 18 21 24 27

图19B

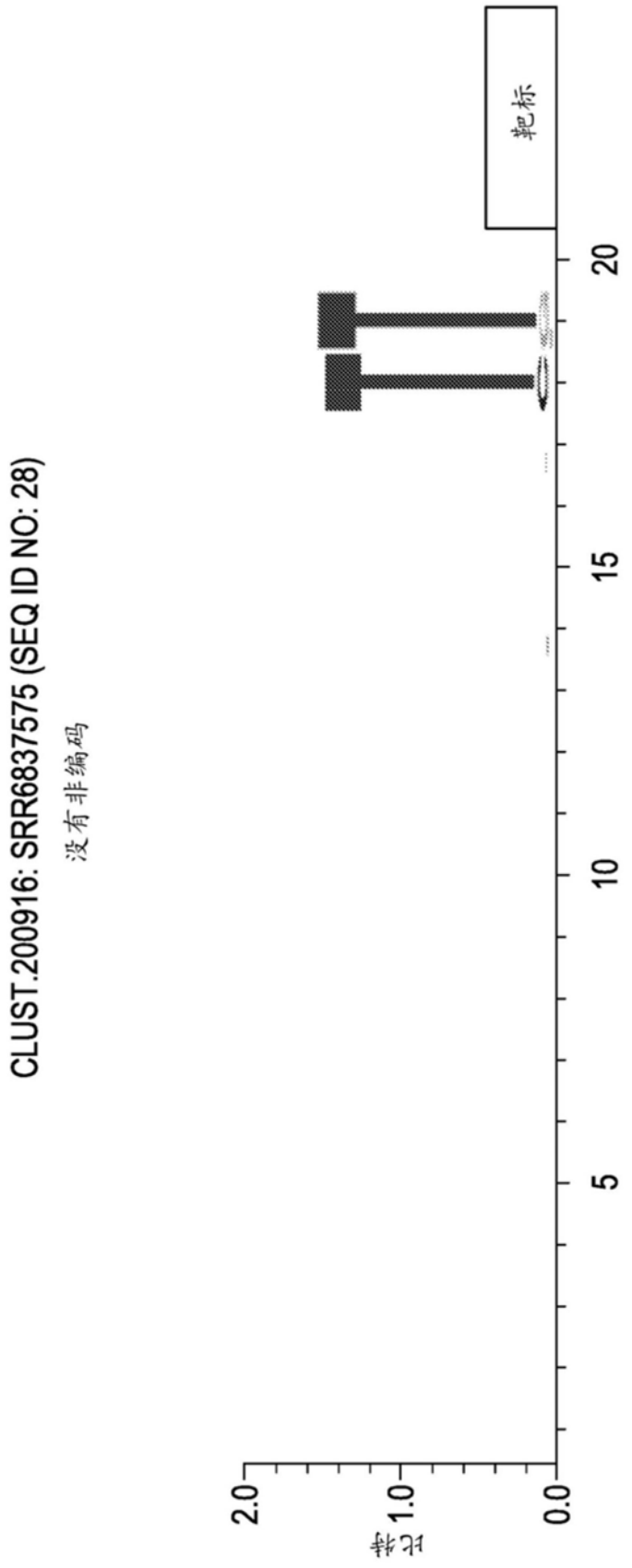

CLUST.200916: SRR6837575 (SEQ ID NO: 28)
没有非编码
比特
2.0
1.0
0.0
5
10
15
20
靶标

图20

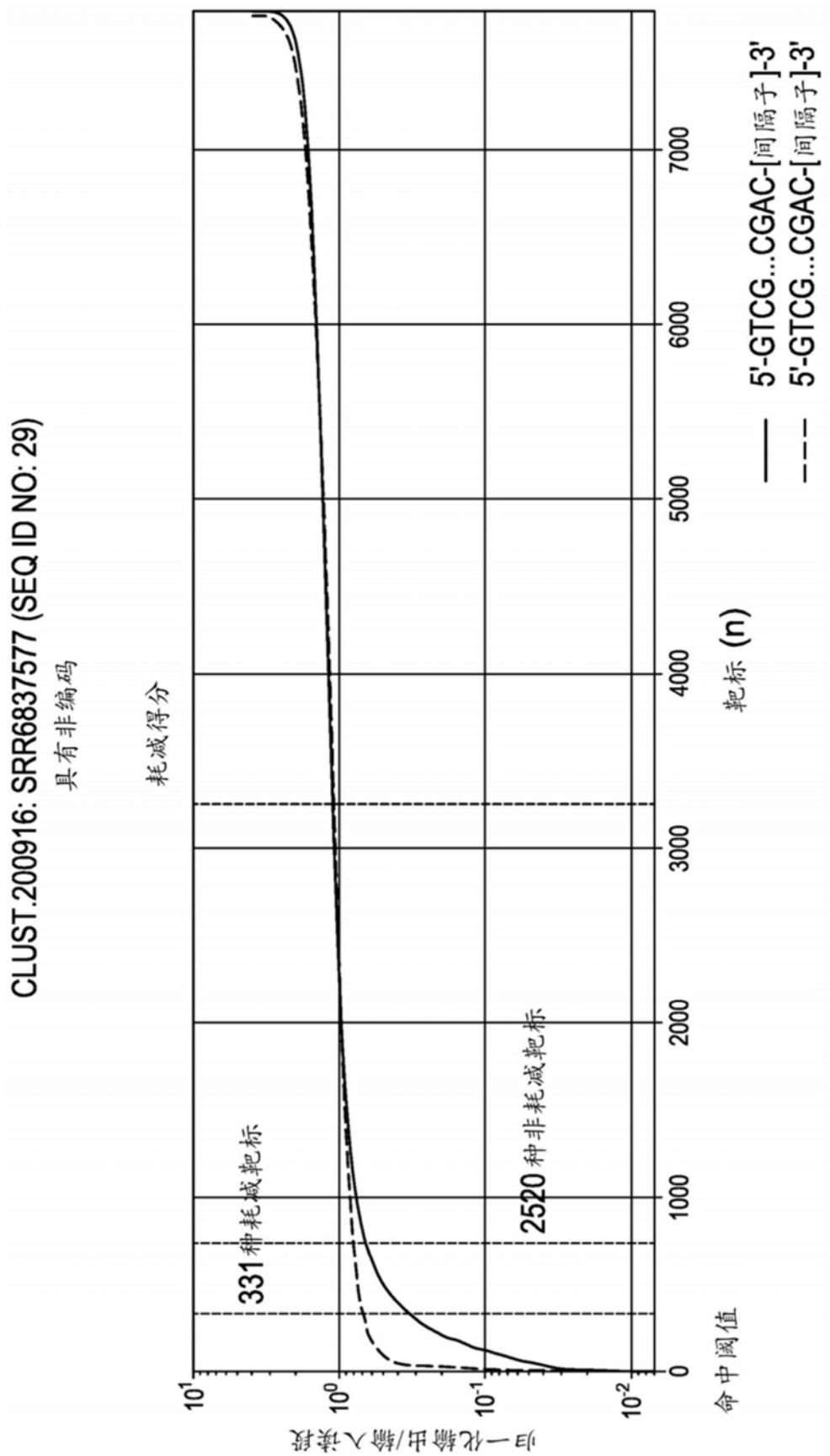

CLUST.200916: SRR6837577 (SEQ ID NO: 29)
具有非编码
耗减得分
331 种耗减靶标
2520 种非耗减靶标
归一化输出/输入读段
命中阈值
靶标 (n)
5'-GTCG...CGAC-[间隔子]-3'
5'-GTCG...CGAC-[间隔子]-3'

图21

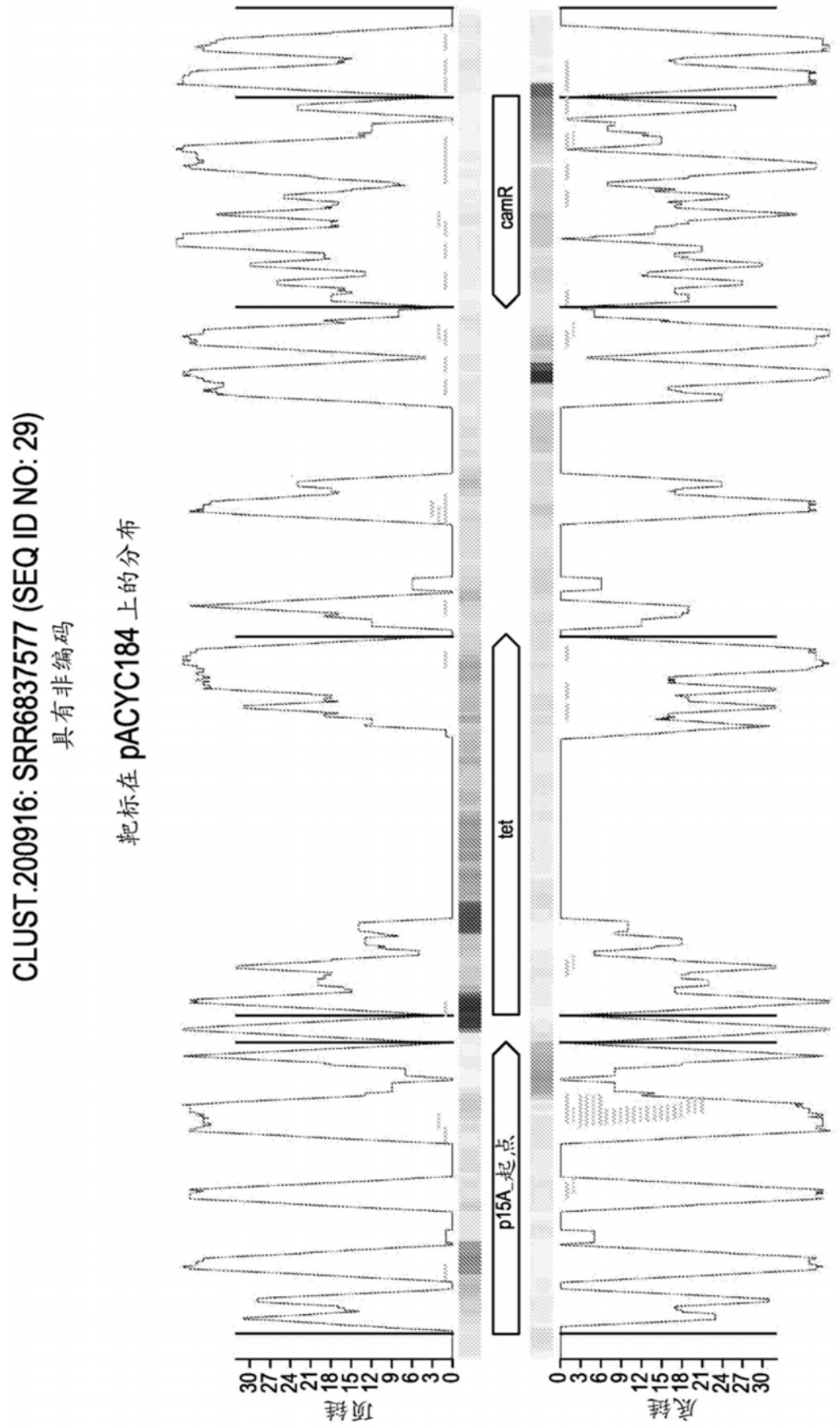

CLUST.200916: SRR6837577 (SEQ ID NO: 29)
具有非编码
靶标在 pACYC184 上的分布
camR
tet
p15A_起点
顶链
底链
30 27 24 21 18 15 12 9 6 3 0
0 3 6 9 12 15 18 21 24 27 30

图22A

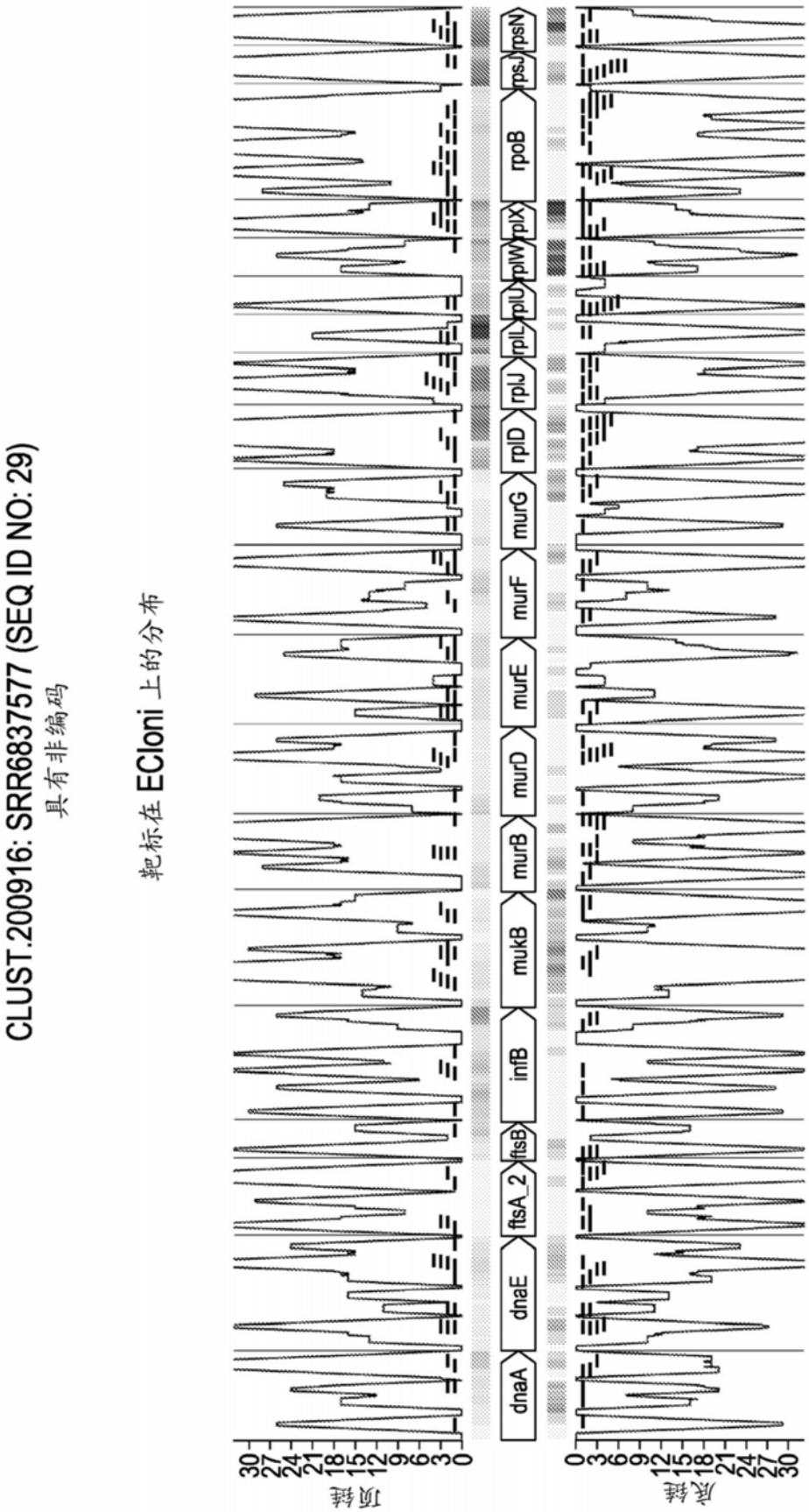

CLUST.200916: SRR6837577 (SEQ ID NO: 29)
具有非编码
靶标在 ECloni 上的分布
dnaA
dnaE
ftsA_2
ftsB
infB
mukB
murB
murD
murE
murF
murG
rplD
rplJ
rplL
rplU
rplW
rplX
rpoB
rpsJ
rpsN
顶链
底链
0 3 6 9 12 15 18 21 24 27 30
30 27 24 21 18 15 12 9 6 3 0

图22B

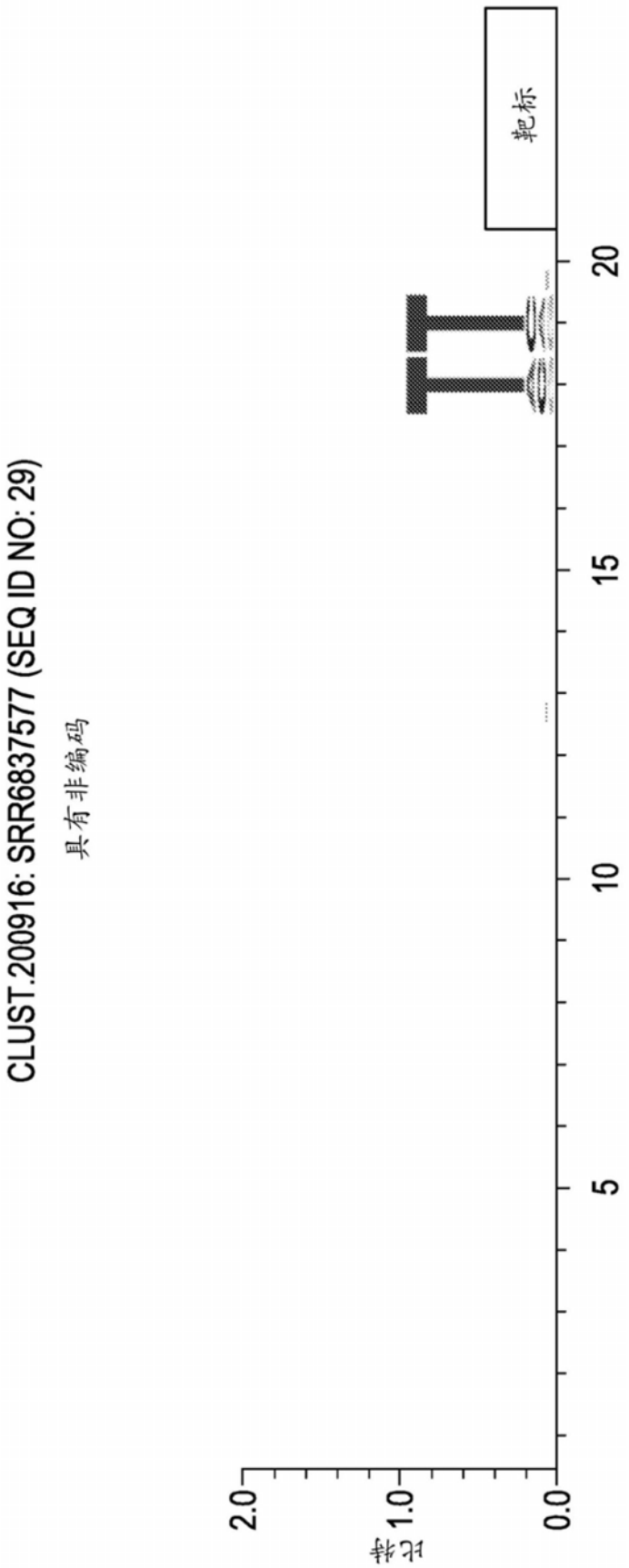

CLUST.200916: SRR6837577 (SEQ ID NO: 29)
具有非编码
比特
2.0
1.0
0.0
5
10
15
20
靶标

图23

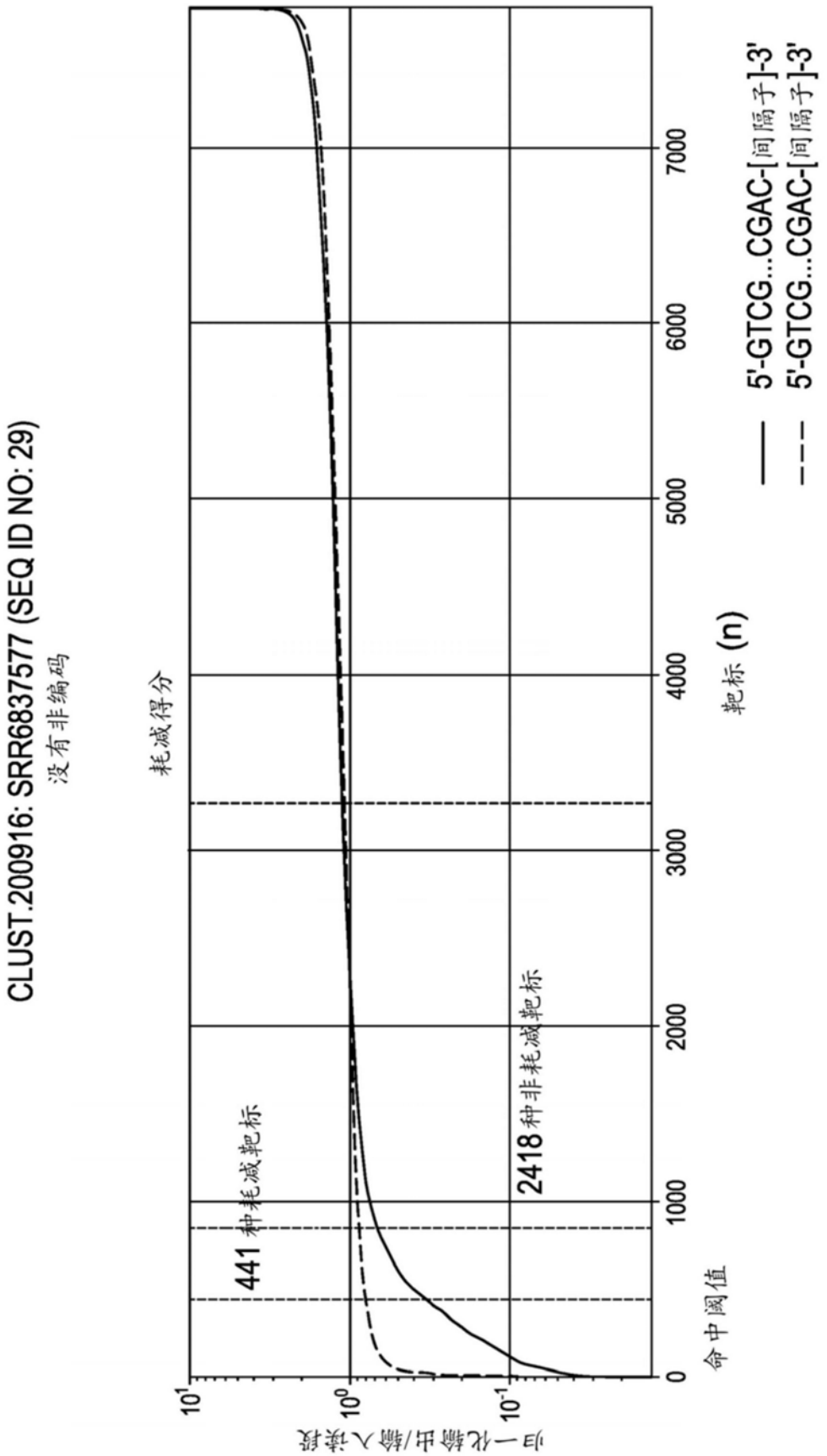

CLUST.200916: SRR6837577 (SEQ ID NO: 29)
没有非编码
耗减得分
441 种耗减靶标
2418 种非耗减靶标
归一化输出/输入读段
命中阈值
靶标 (n)
5'-GTCG...CGAC-[间隔子]-3'
5'-GTCG...CGAC-[间隔子]-3'

图24

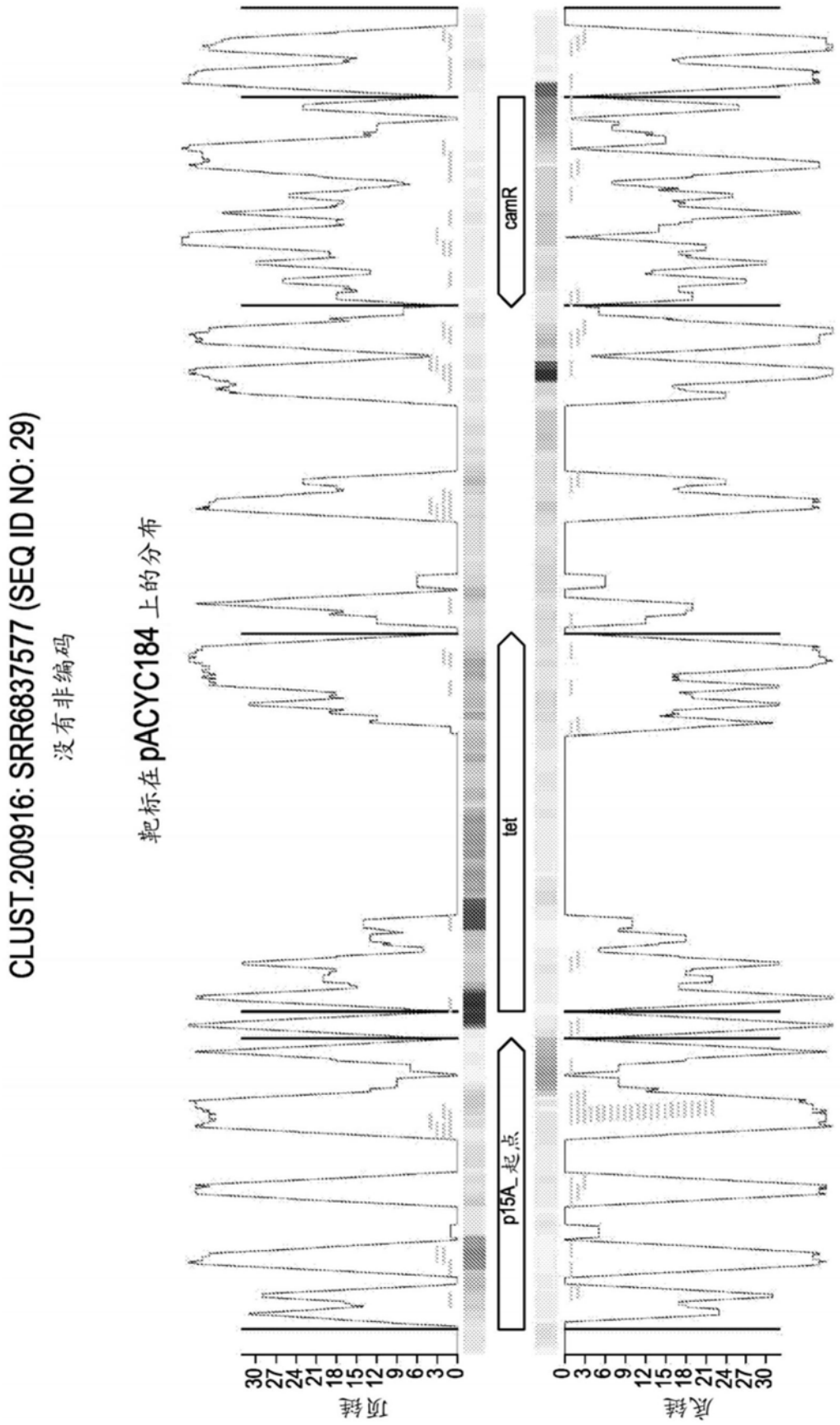

CLUST.200916: SRR6837577 (SEQ ID NO: 29)
没有非编码
靶标在 pACYC184 上的分布
p15A_起点
tet
camR
顶链
底链
0
3
6
9
12
15
18
21
24
27
30

图25A

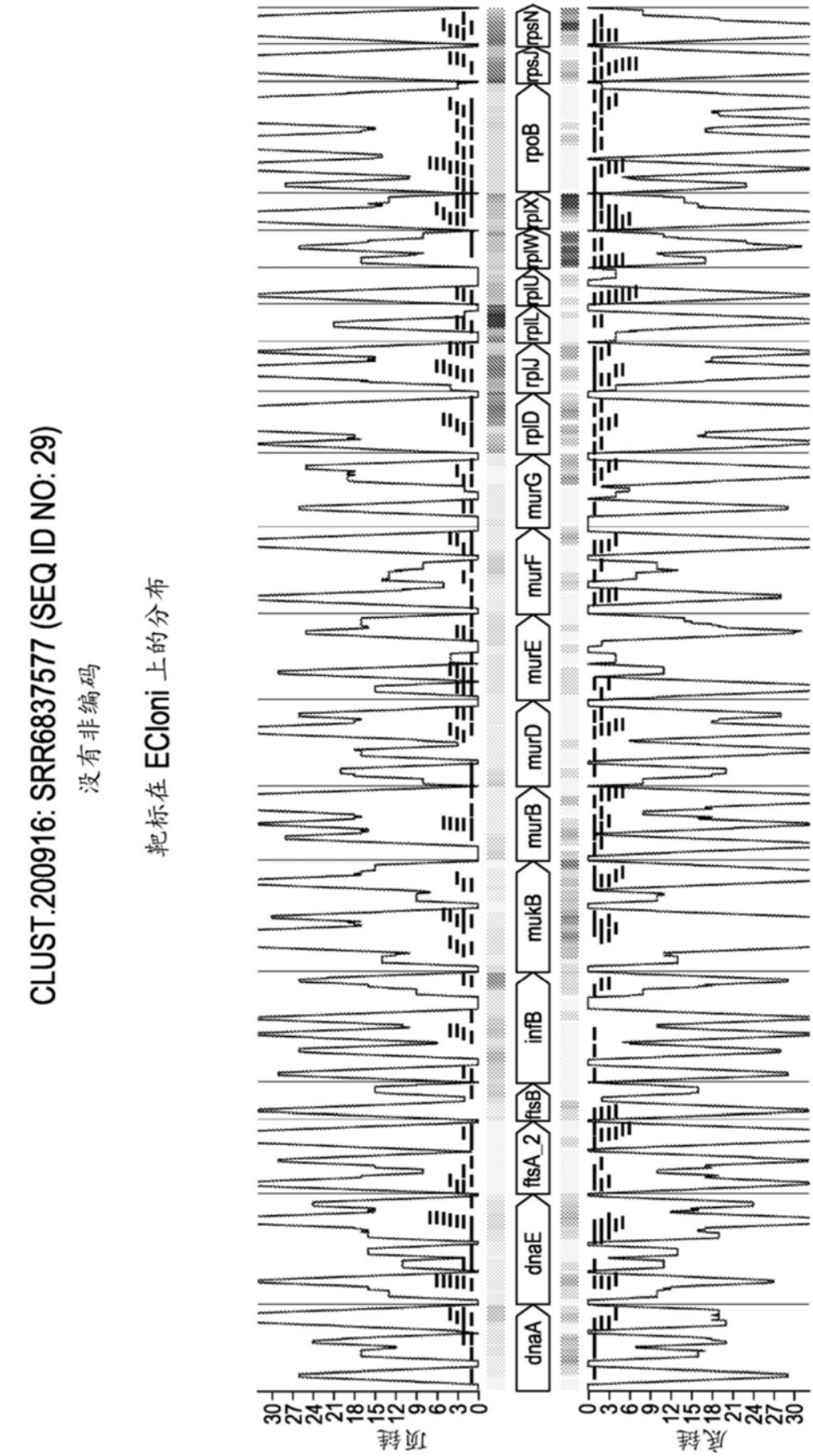

CLUST.200916: SRR6837577 (SEQ ID NO: 29)
没有非编码
靶标在 ECloni 上的分布
dnaA
dnaE
ftsA_2
ftsB
infB
mukB
murB
murD
murE
murF
murG
rplD
rplJ
rplL
rplU
rplW
rplX
rpoB
rpsJ
rpsN
顶链
底链
0 3 6 9 12 15 18 21 24 27 30

图25B

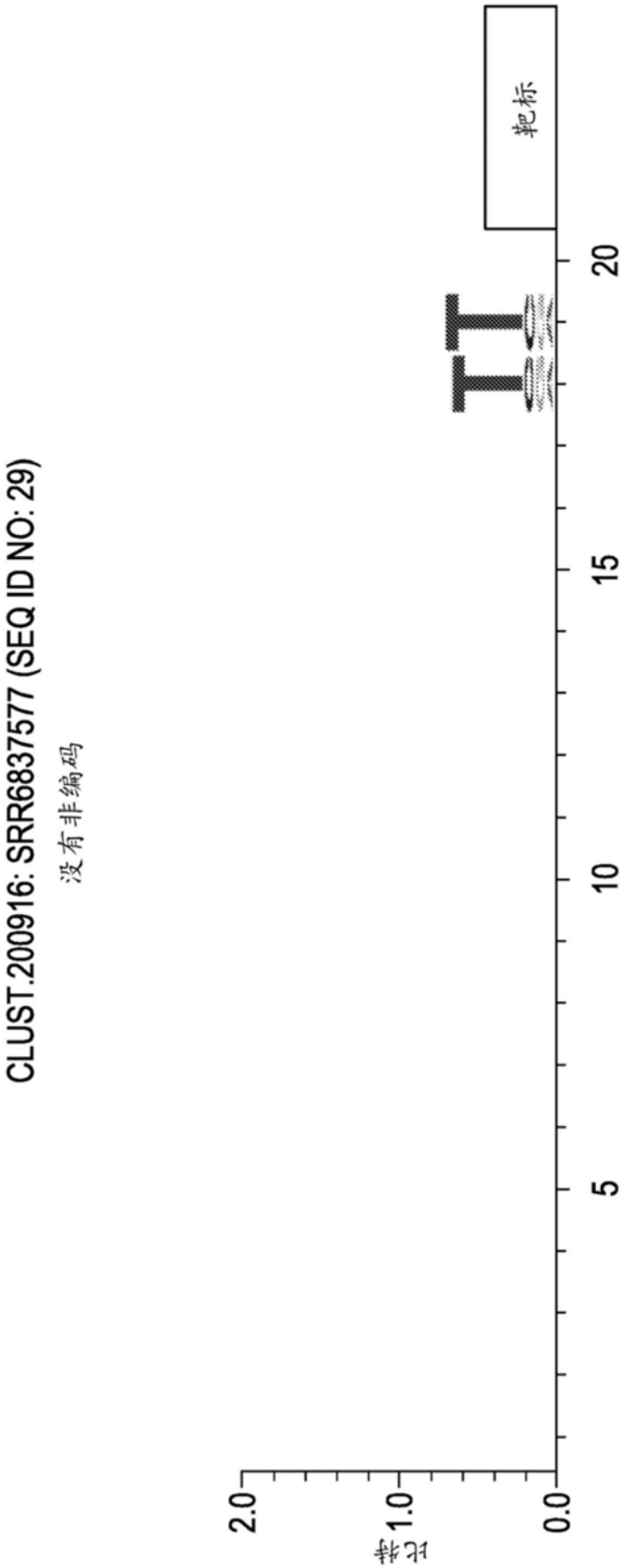

CLUST.200916: SRR6837577 (SEQ ID NO: 29)
没有非编码
比特
2.0
1.0
0.0
5
10
15
20
靶标

图26

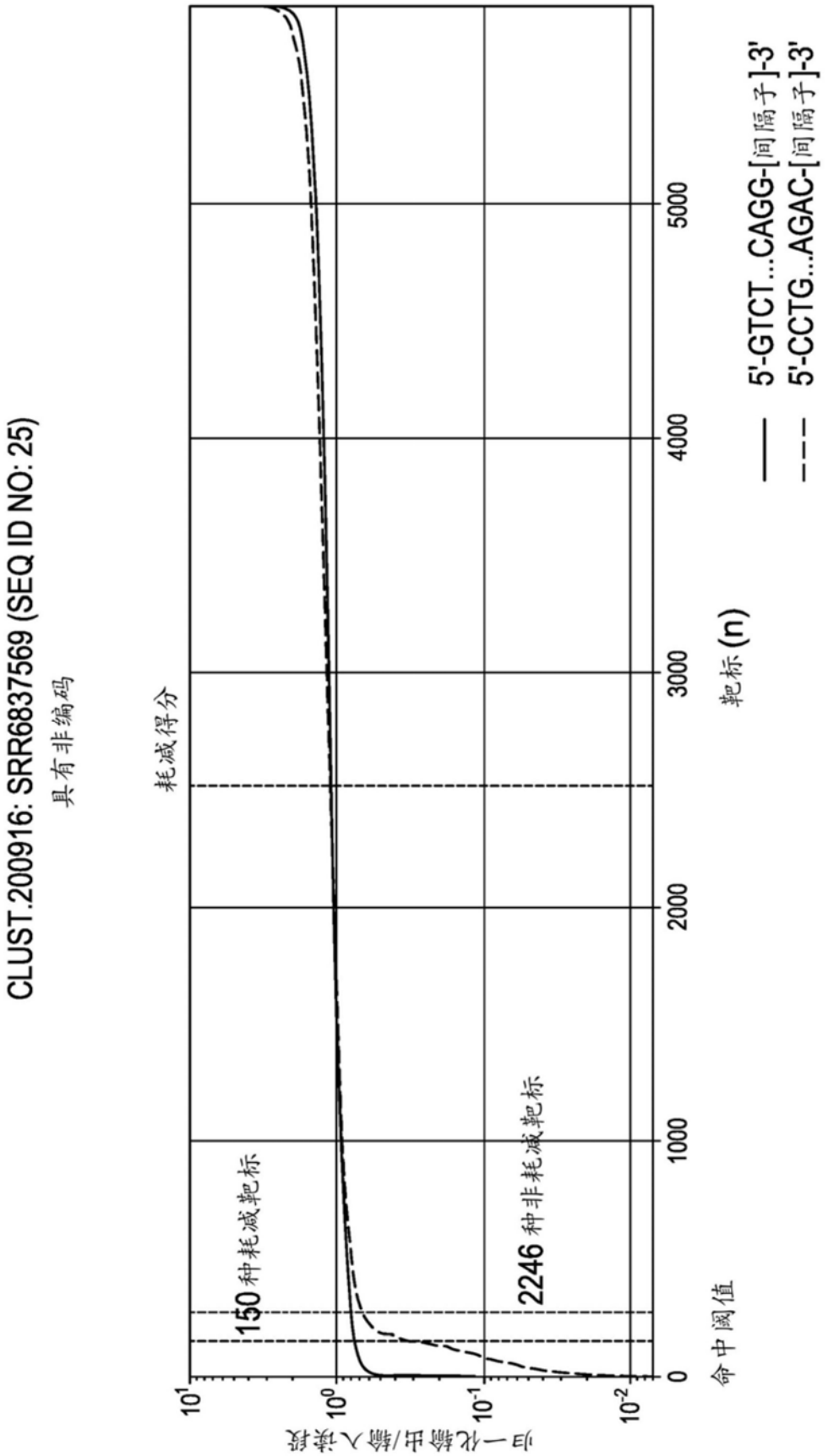

CLUST.200916: SRR6837569 (SEQ ID NO: 25)
具有非编码
耗减得分
150 种耗减靶标
2246 种非耗减靶标
归一化输出/输入读段
命中阈值
靶标 (n)
5'-GTCT...CAGG-[间隔子]-3'
5'-CCTG...AGAC-[间隔子]-3'

图27

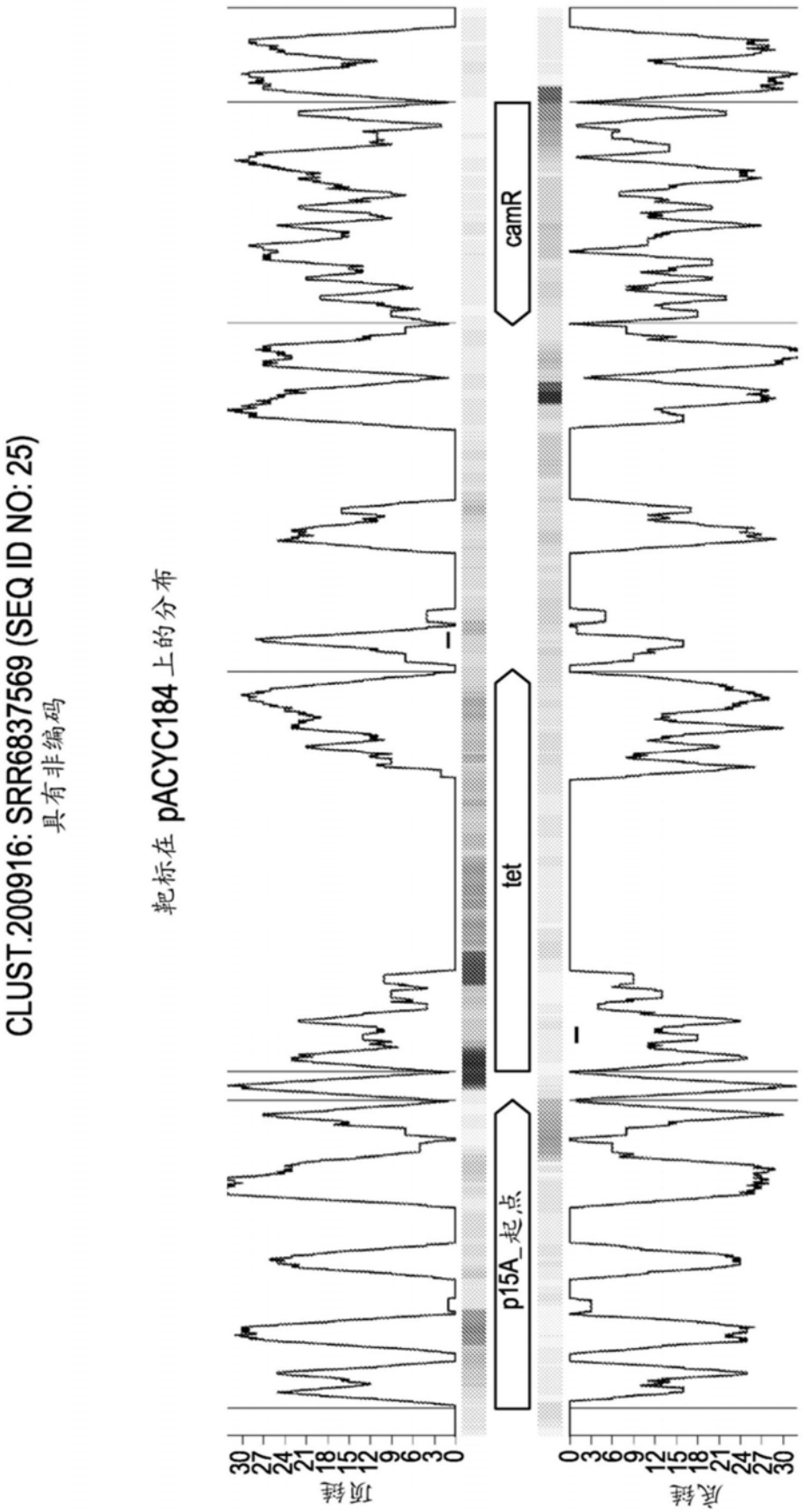

CLUST.200916: SRR6837569 (SEQ ID NO: 25)
具有非编码
靶标在 pACYC184 上的分布
camR
tet
p15A_起点
顶链
底链
30 27 24 21 18 15 12 9 6 3 0
0 3 6 9 12 15 18 21 24 27 30

图28A

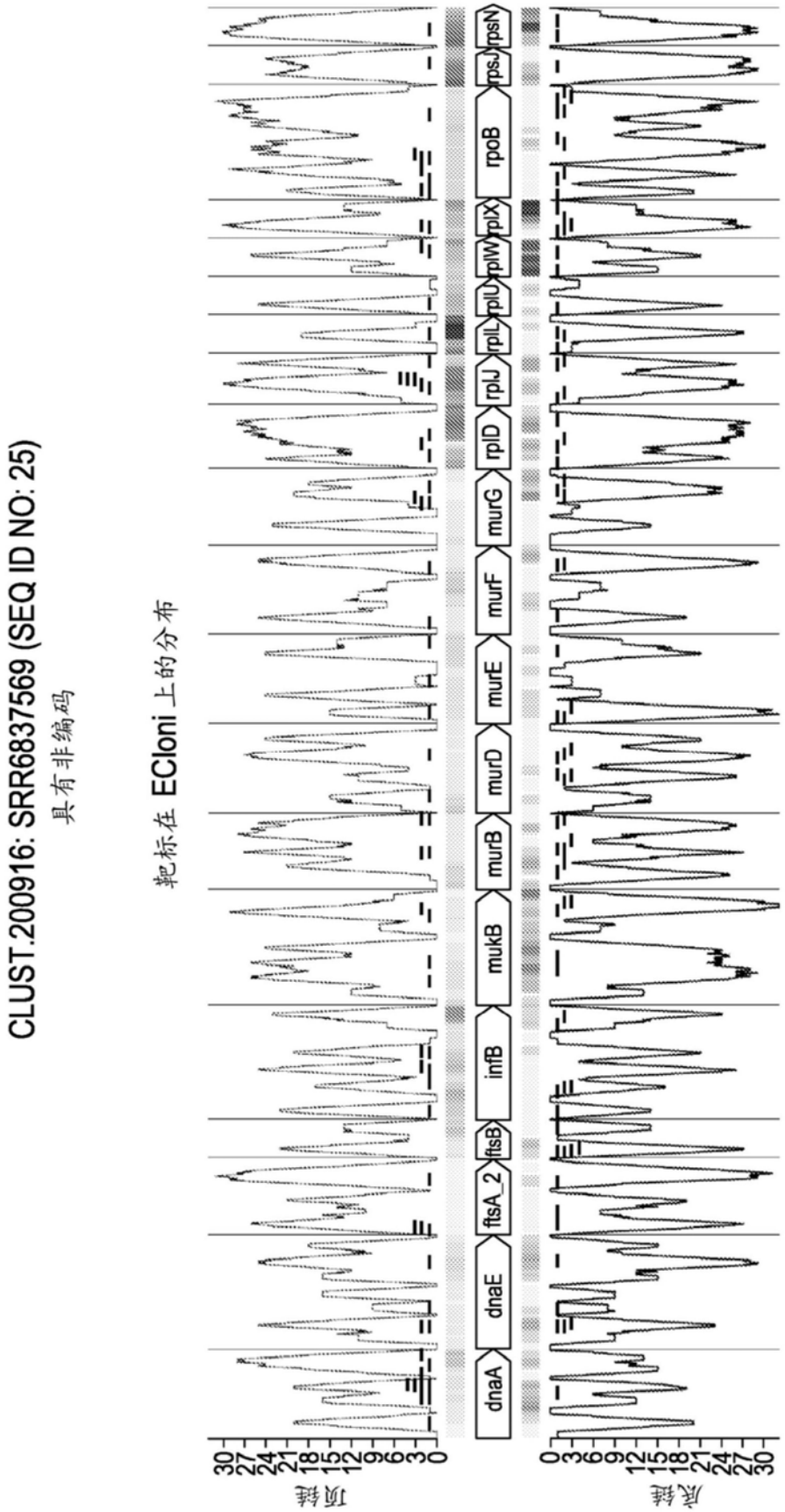

CLUST.200916: SRR6837569 (SEQ ID NO: 25)
具有非编码
靶标在 ECIoni 上的分布
顶链
底链
dnaA
dnaE
ftsA_2
ftsB
infB
mukB
murB
murD
murE
murF
murG
rplD
rplJ
rplL
rplU
rplW
rplX
rpoB
rpsJ
rpsN

图28B

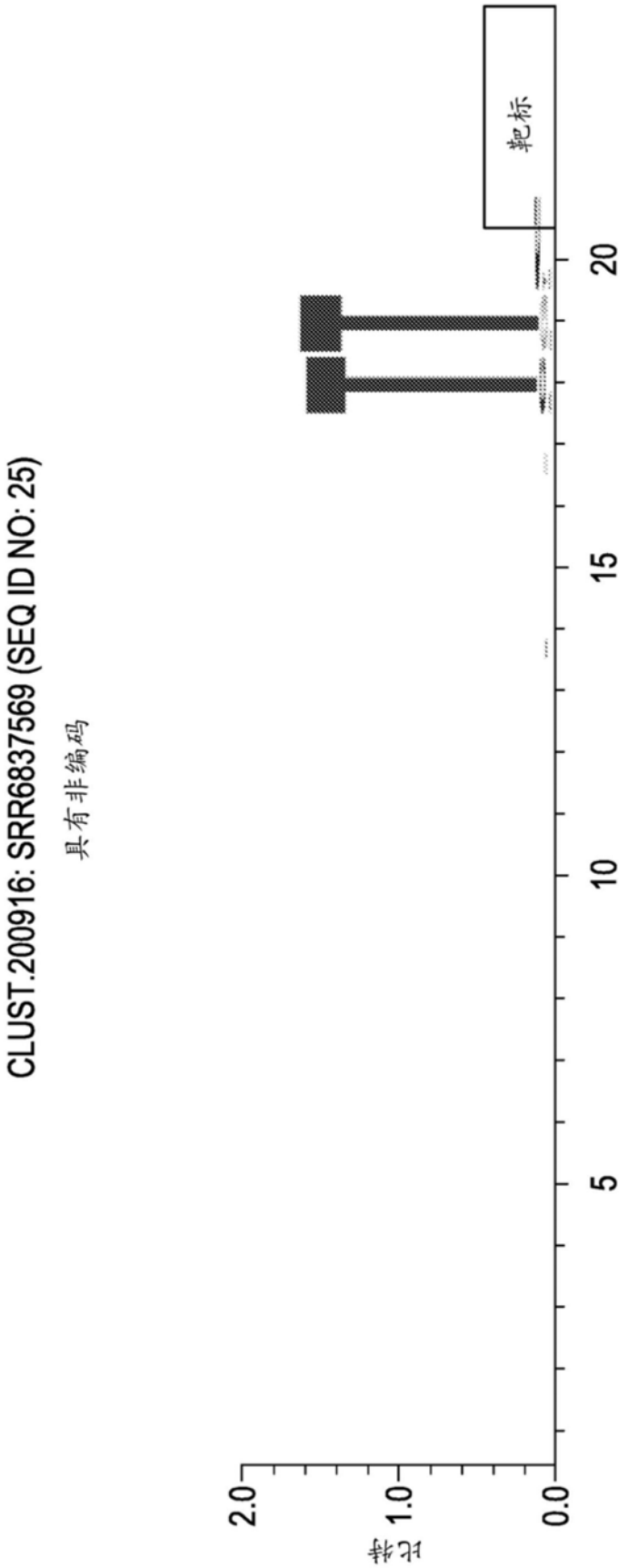

CLUST.200916: SRR6837569 (SEQ ID NO: 25)
具有非编码
比特
2.0
1.0
0.0
5
10
15
20
靶标

图29

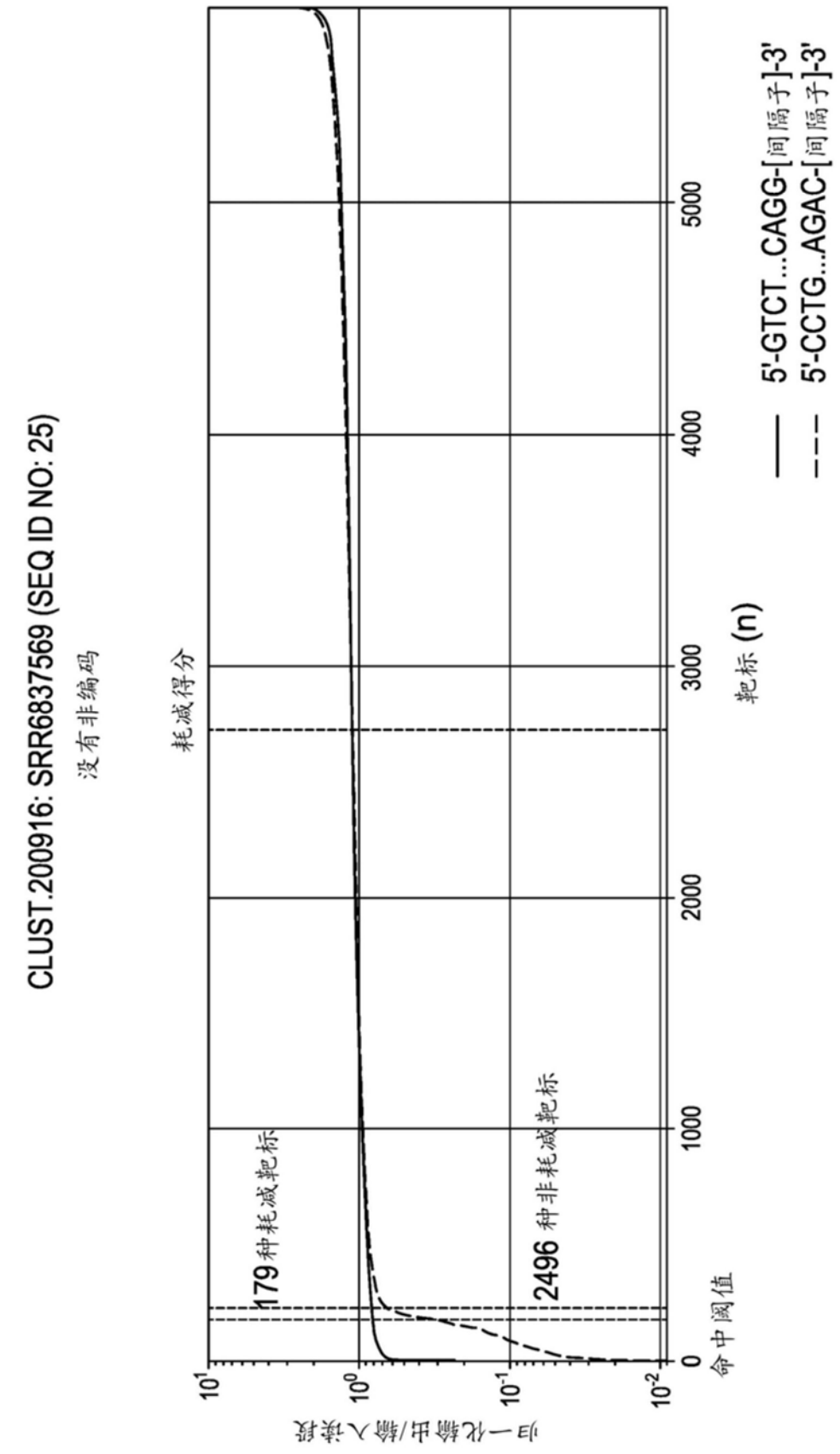

CLUST.200916: SRR6837569 (SEQ ID NO: 25)
没有非编码
耗减得分
179 种耗减靶标
2496 种非耗减靶标
归一化输出/输入读段
命中阈值
靶标 (n)
5'-GTCT...CAGG-[间隔子]-3'
5'-CCTG...AGAC-[间隔子]-3'

图30

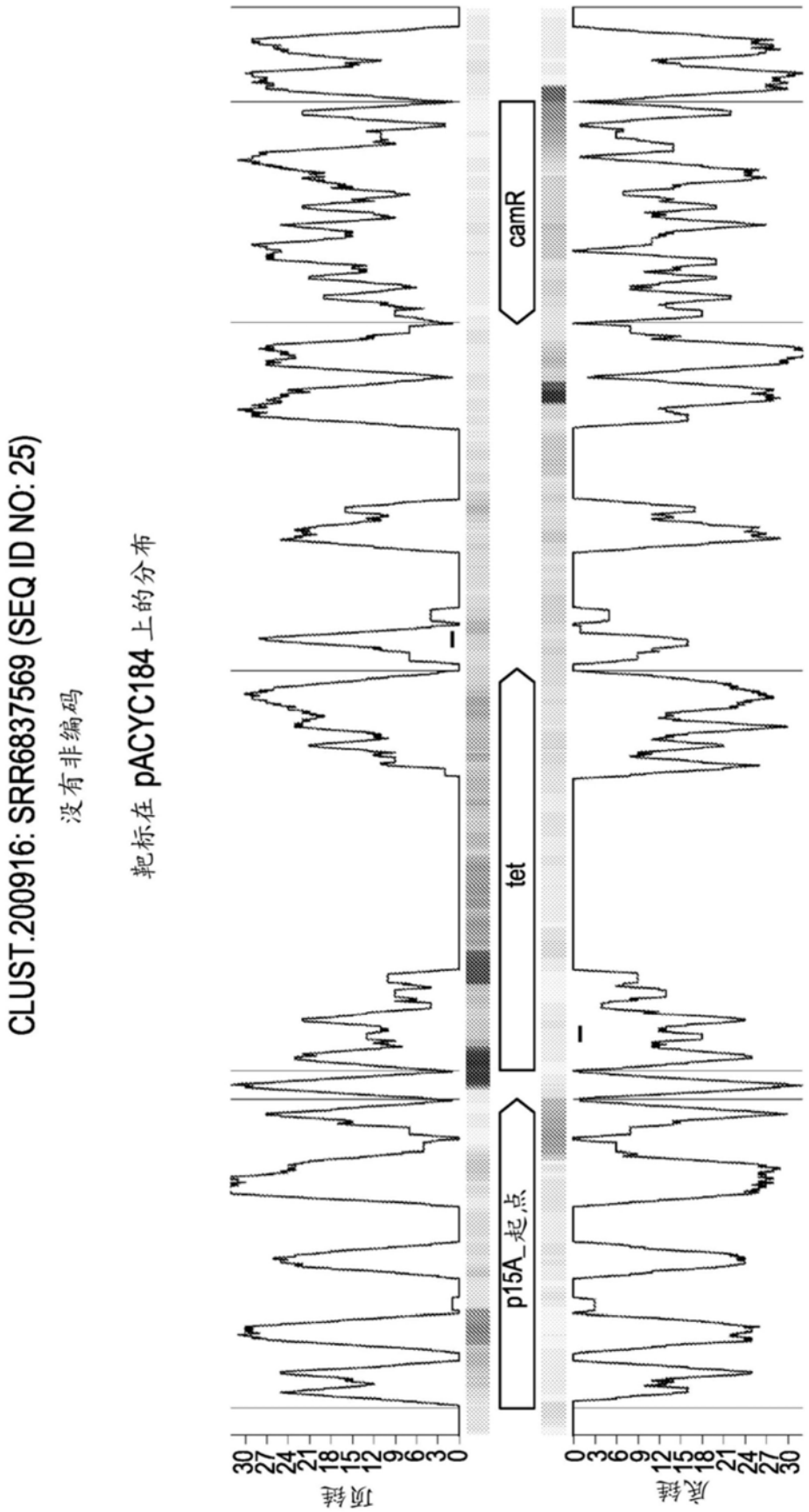

CLUST.200916: SRR6837569 (SEQ ID NO: 25)
没有非编码
靶标在 pACYC184 上的分布
顶链
底链
30 27 24 21 18 15 12 9 6 3 0
0 3 6 9 12 15 18 21 24 27 30
p15A_起点
tet
camR
–
–

图31A

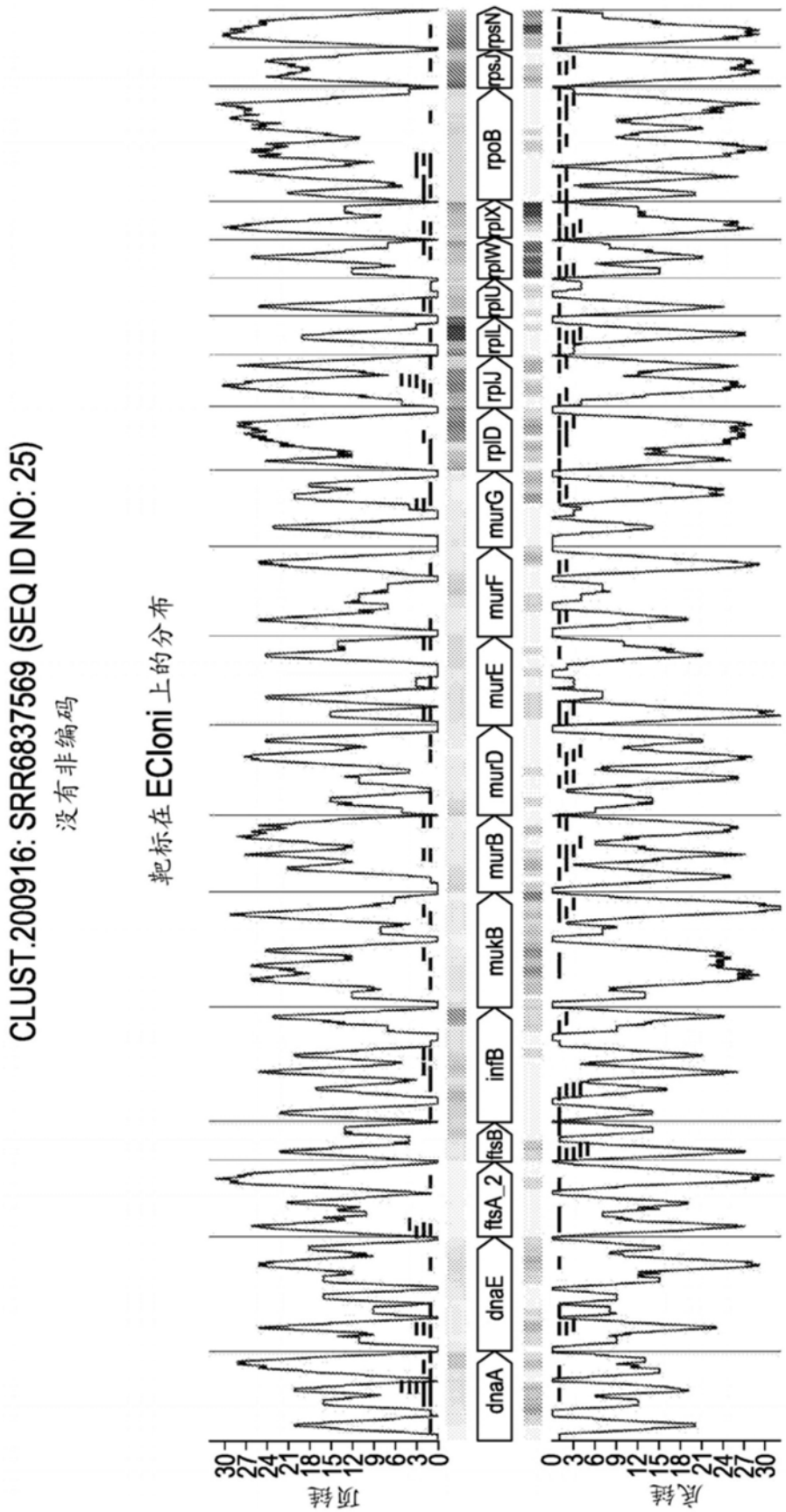

CLUST.200916: SRR6837569 (SEQ ID NO: 25)
没有非编码
靶标在 ECIoni 上的分布
dnaA
dnaE
ftsA_2
ftsB
infB
mukB
murB
murD
murE
murF
murG
rplD
rplJ
rplL
rplU
rplW
rplX
rpoB
rpsJ
rpsN
顶链
底链
0 3 6 9 12 15 18 21 24 27 30
30 27 24 21 18 15 12 9 6 3 0

图31B

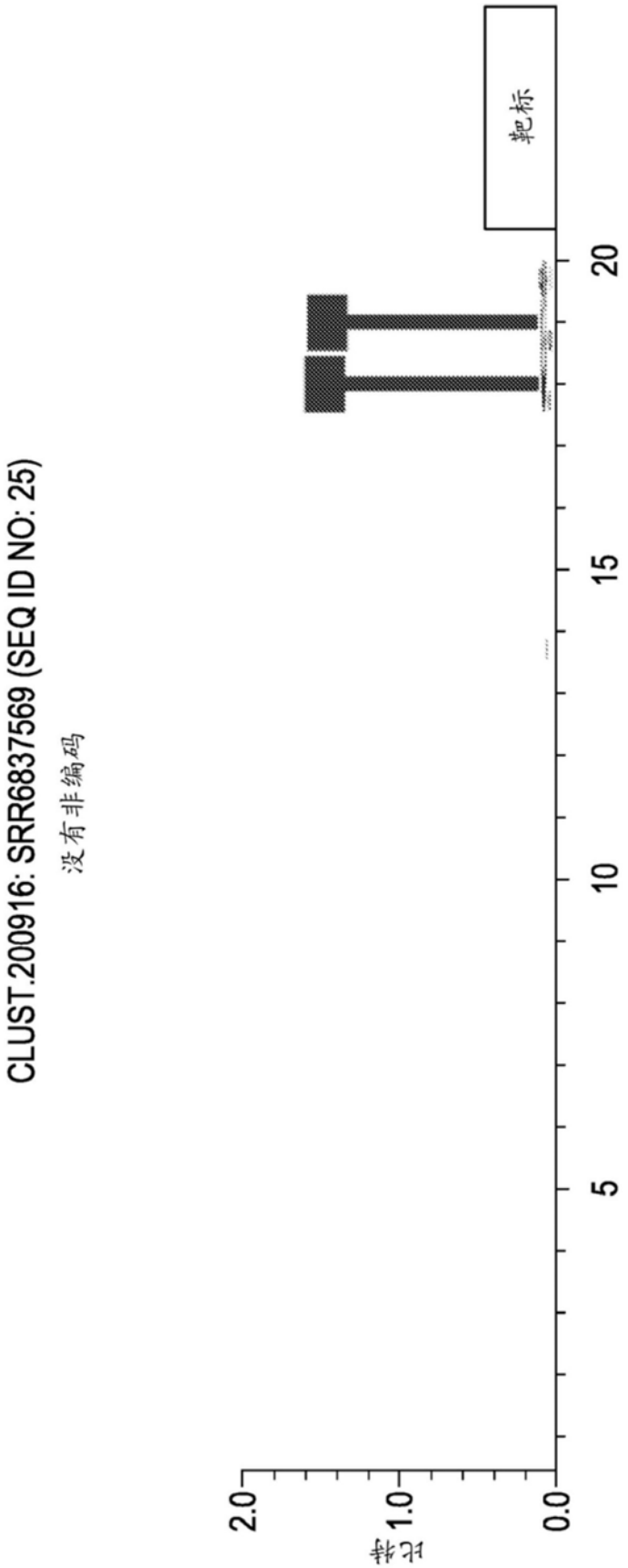

CLUST.200916: SRR6837569 (SEQ ID NO: 25)
没有非编码
比特
2.0
1.0
0.0
5
10
15
20
靶标

图32

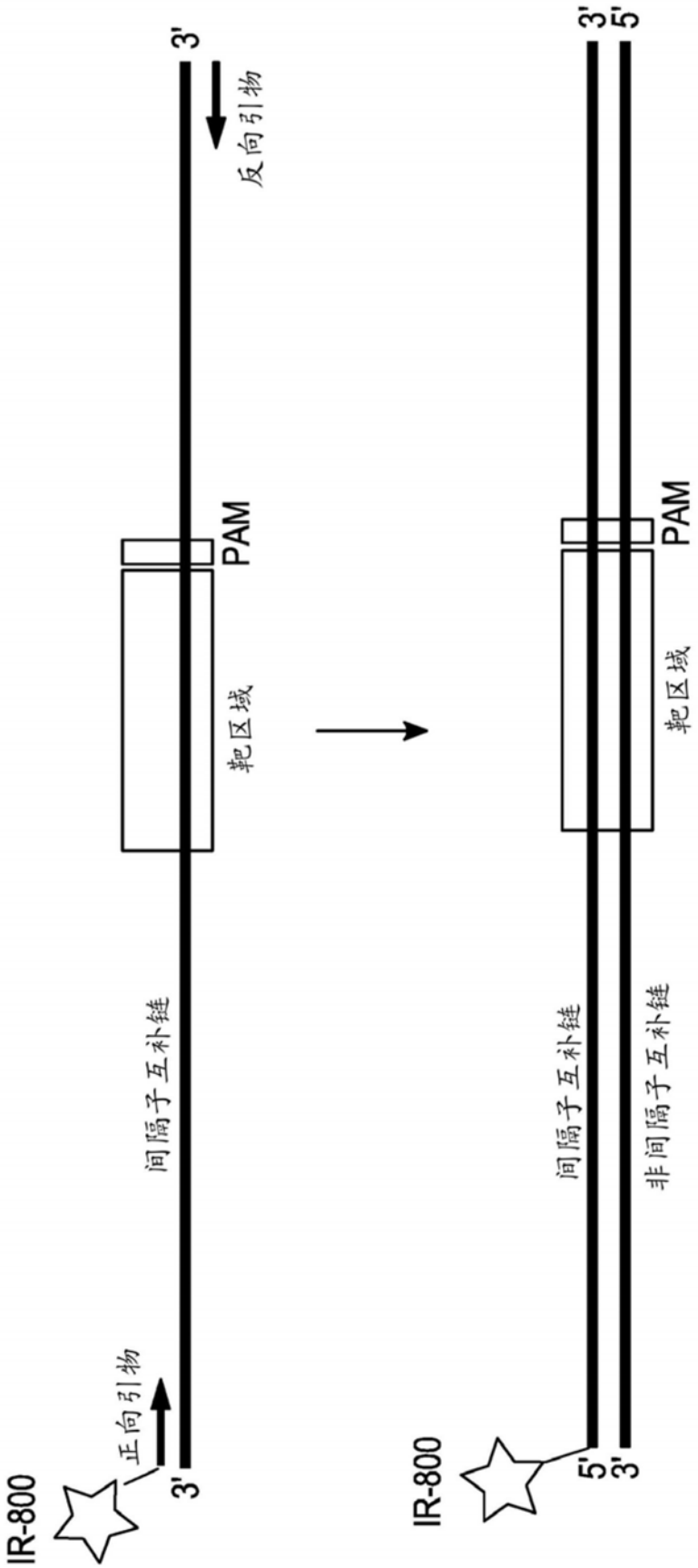

3'
反向引物
PAM
靶区域
间隔子互补链
正向引物
IR-800
3'
3'
5'
PAM
靶区域
间隔子互补链
非间隔子互补链
IR-800
5'
3'

图33

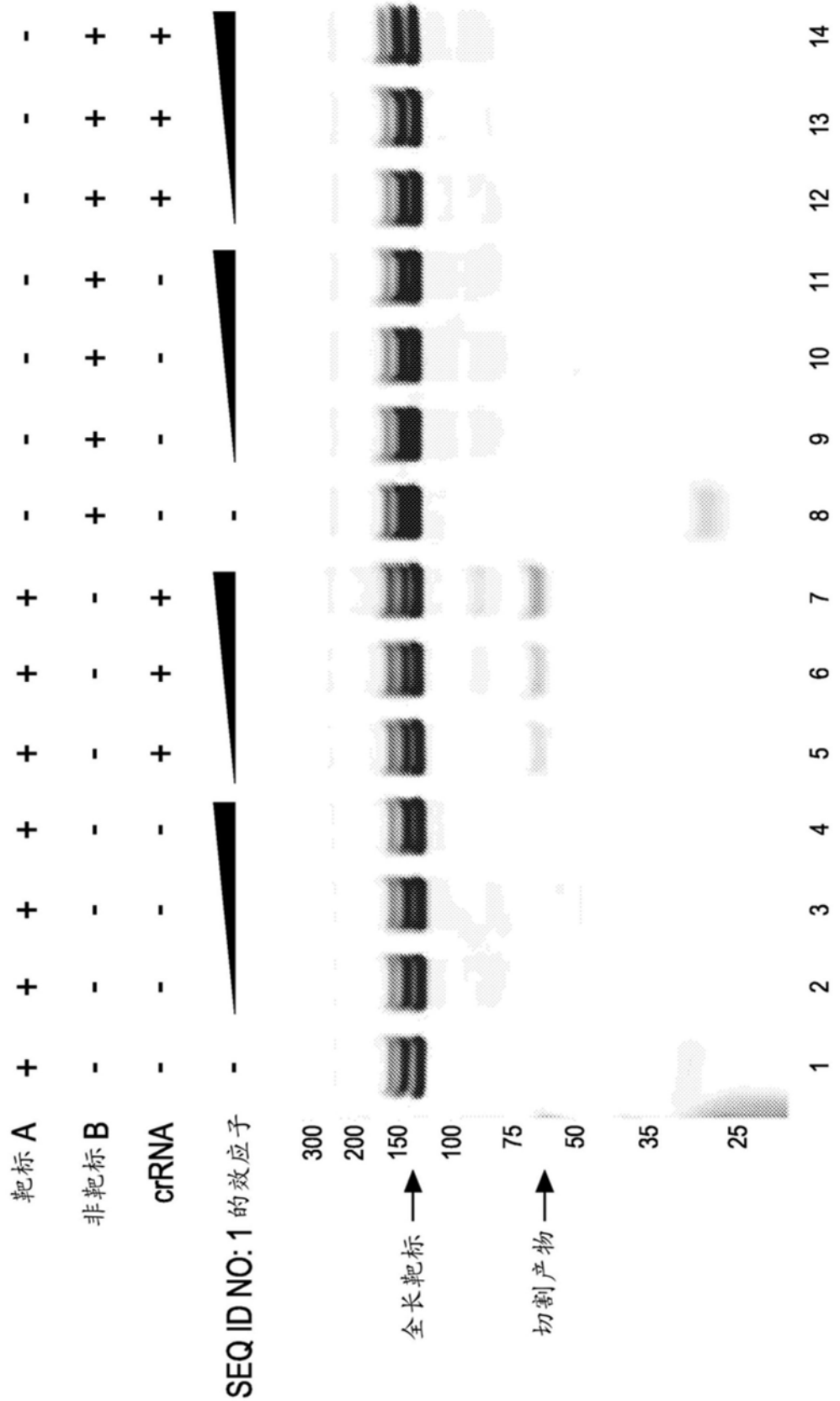

靶标 A
非靶标 B
crRNA
SEQ ID NO: 1 的效应子
300
200
150
100
75
50
35
25
全长靶标
切割产物
1
2
3
4
5
6
7
8
9
10
11
12
13
14

图34A

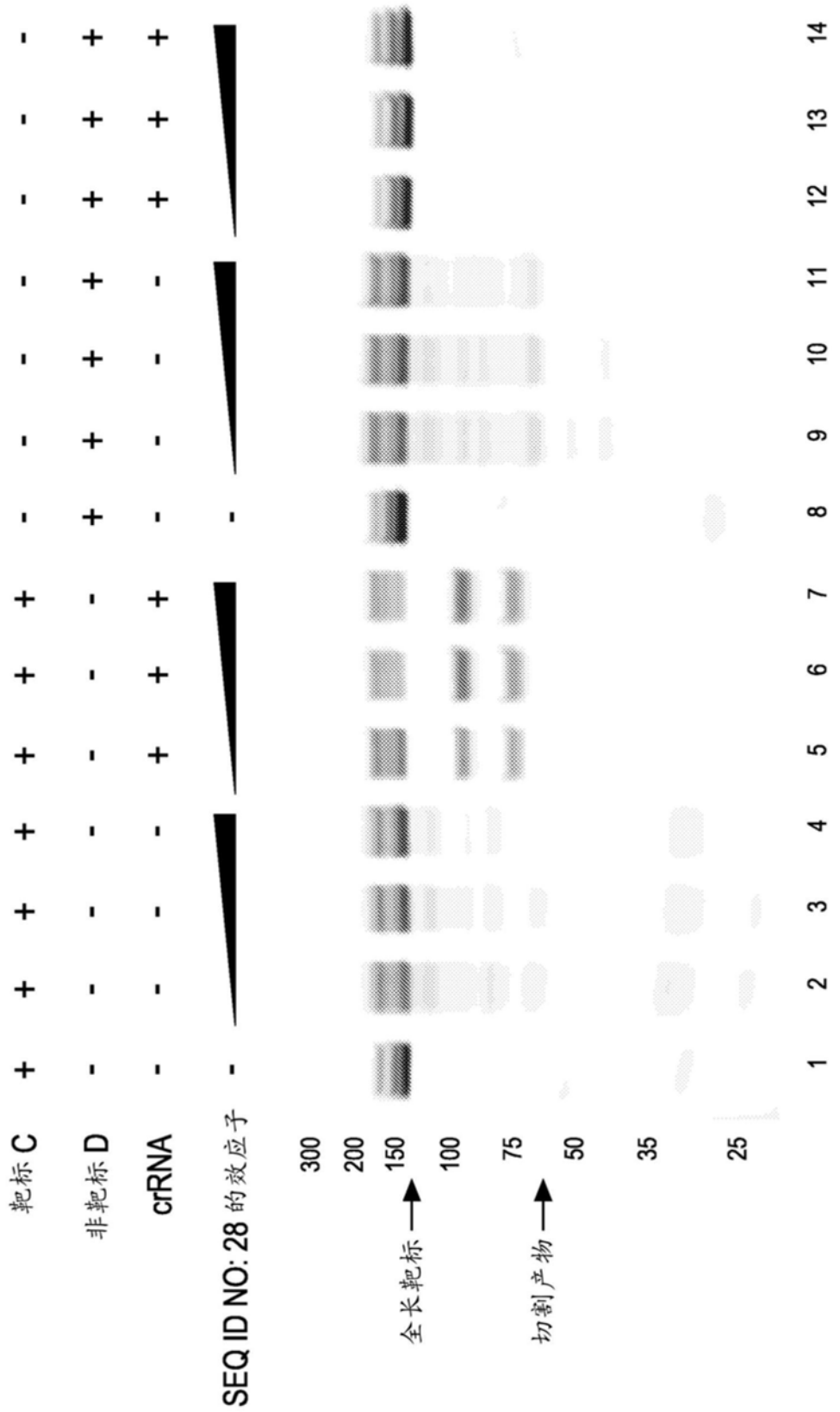

靶标 C
非靶标 D
crRNA
SEQ ID NO: 28 的效应子
300
200
150
100
75
50
35
25
全长靶标
切割产物
1 2 3 4 5 6 7 8 9 10 11 12 13 14

图34B

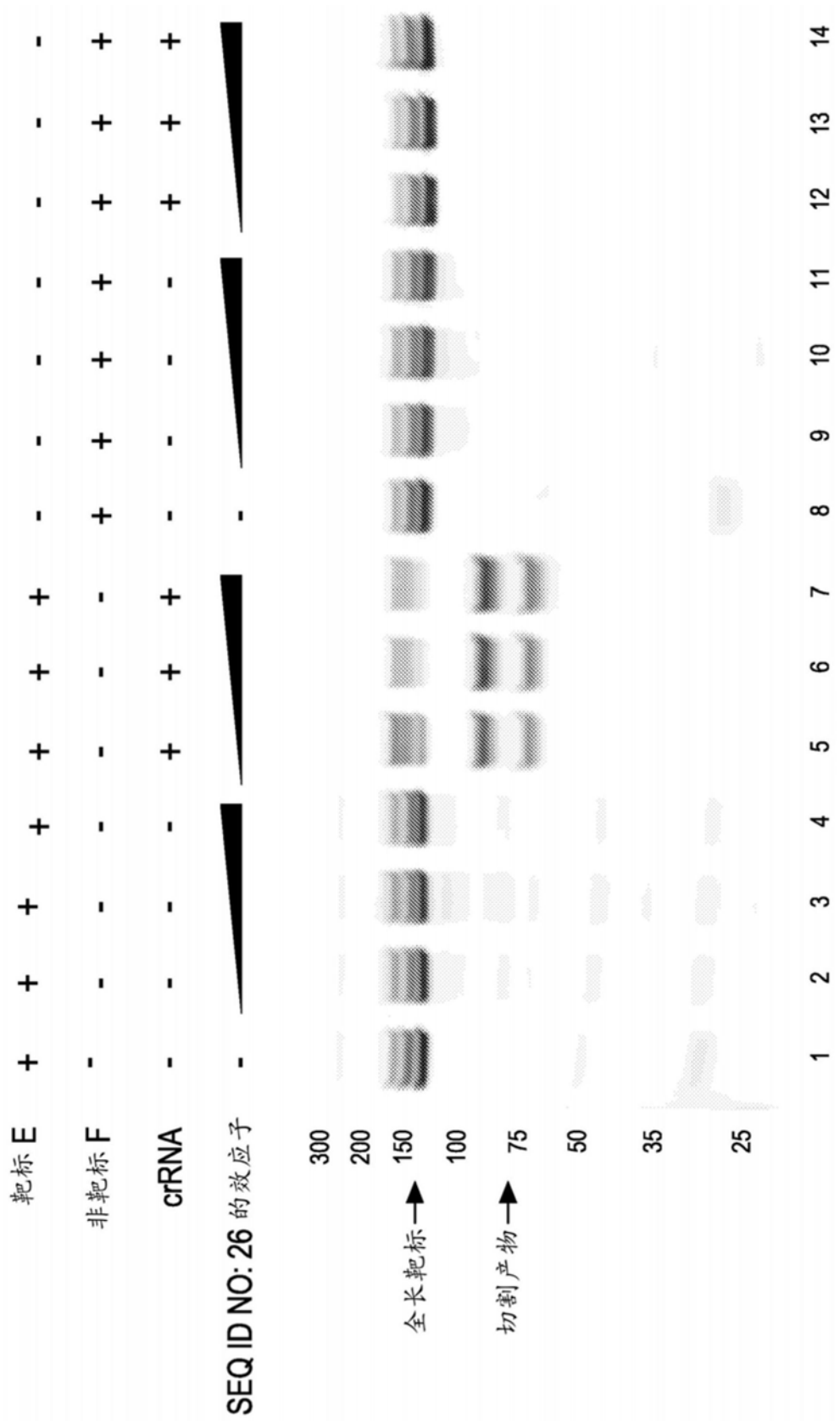

靶标 E
非靶标 F
crRNA
SEQ ID NO: 26 的效应子
300
200
150
100
75
50
35
25
全长靶标
切割产物
1 2 3 4 5 6 7 8 9 10 11 12 13 14

图34C

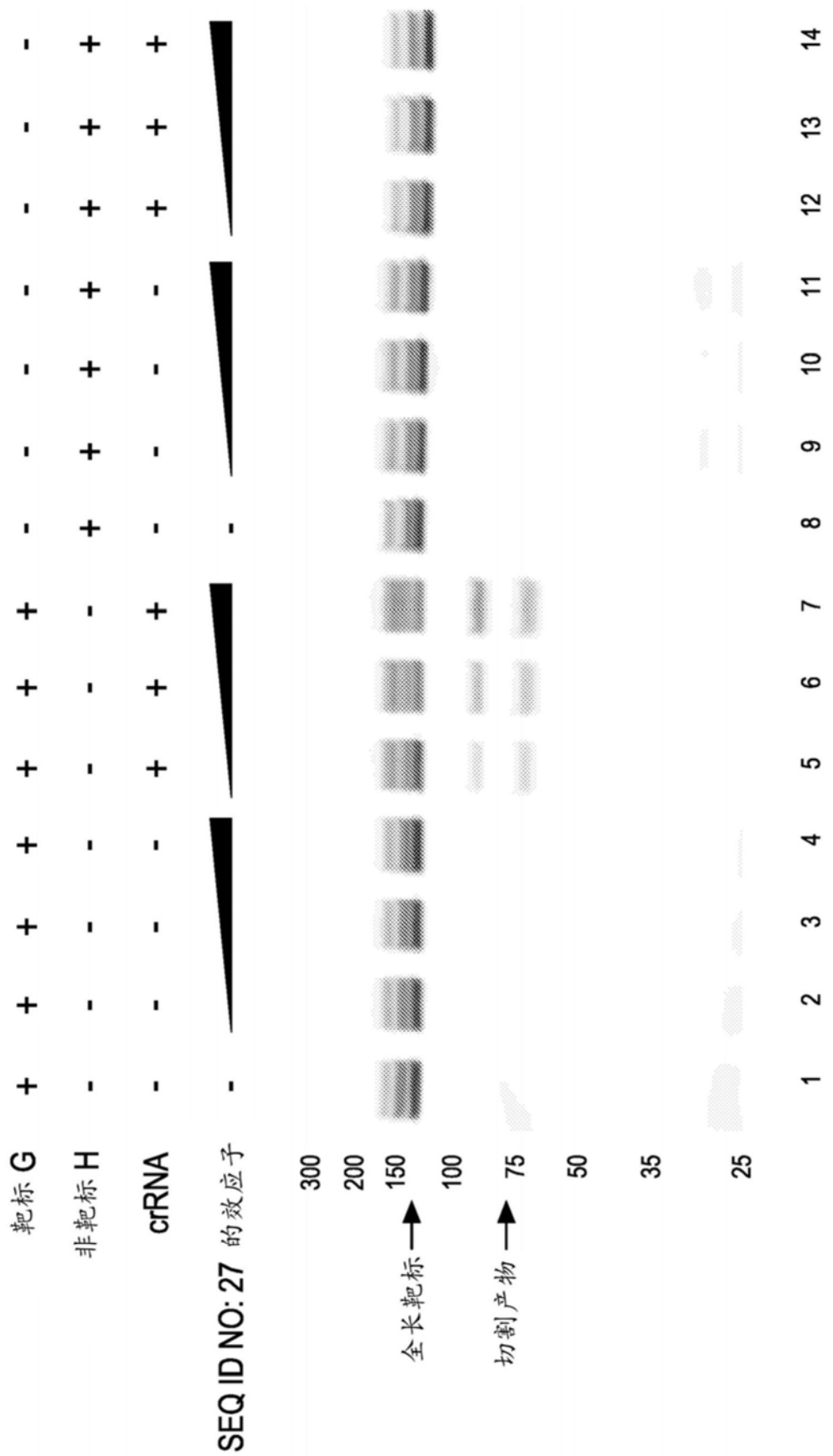

靶标 G
非靶标 H
crRNA
SEQ ID NO: 27 的效应子
300
200
150
100
75
50
35
25
全长靶标
切割产物
1 2 3 4 5 6 7 8 9 10 11 12 13 14

图34D

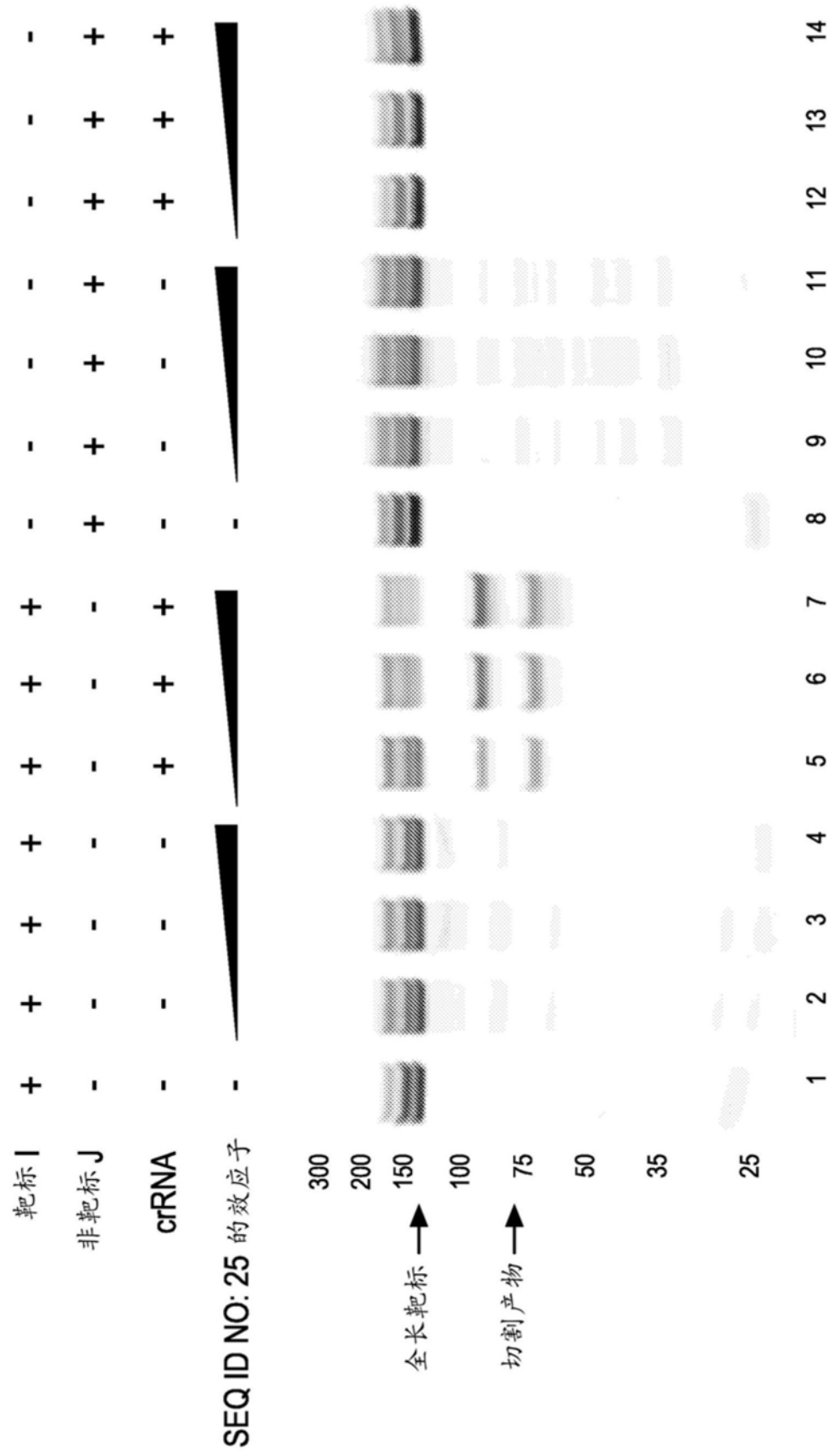

靶标 I
非靶标 J
crRNA
SEQ ID NO: 25 的效应子
300
200
150
100
75
50
35
25
全长靶标
切割产物
1 2 3 4 5 6 7 8 9 10 11 12 13 14

图34E

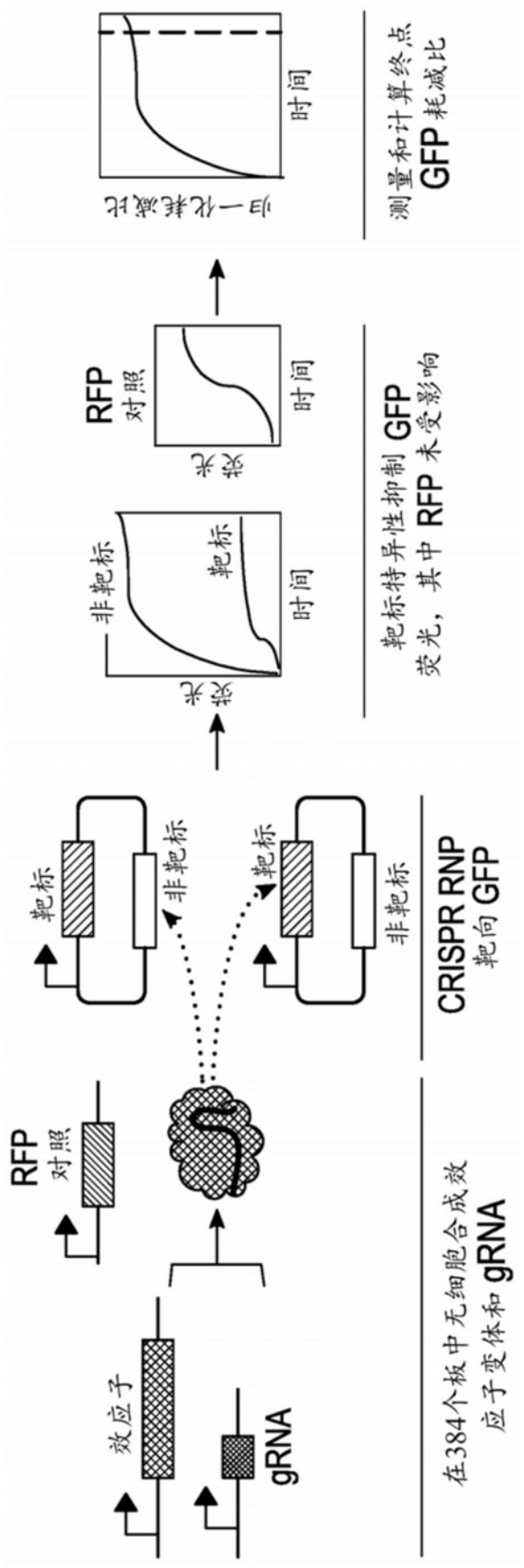

效应子
gRNA
RFP
对照
靶标
非靶标
非靶标
靶标
CRISPR RNP
靶向 GFP
在384个板中无细胞合成效
应子变体和 gRNA
荧光
时间
非靶标
靶标
荧光
时间
RFP
对照
荧光
时间
靶标特异性抑制 GFP
荧光，其中 RFP 未受影响
当一化荧光比
时间
测量和计算终点
GFP 耗减比

图35

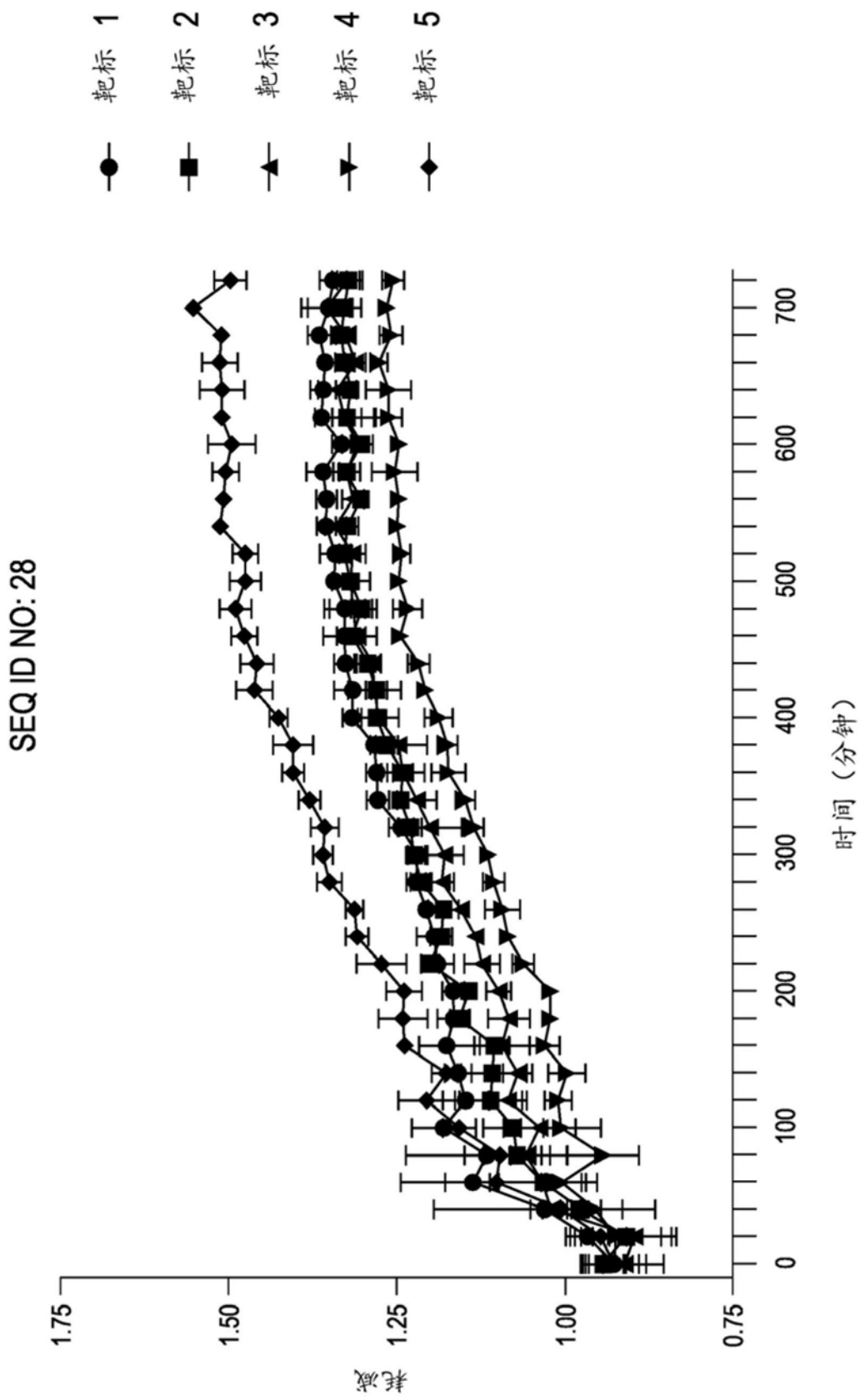

靶标 1
靶标 2
靶标 3
靶标 4
靶标 5
SEQ ID NO: 28
耗减
1.75
1.50
1.25
1.00
0.75
0
100
200
300
400
500
600
700
时间（分钟）

图36A

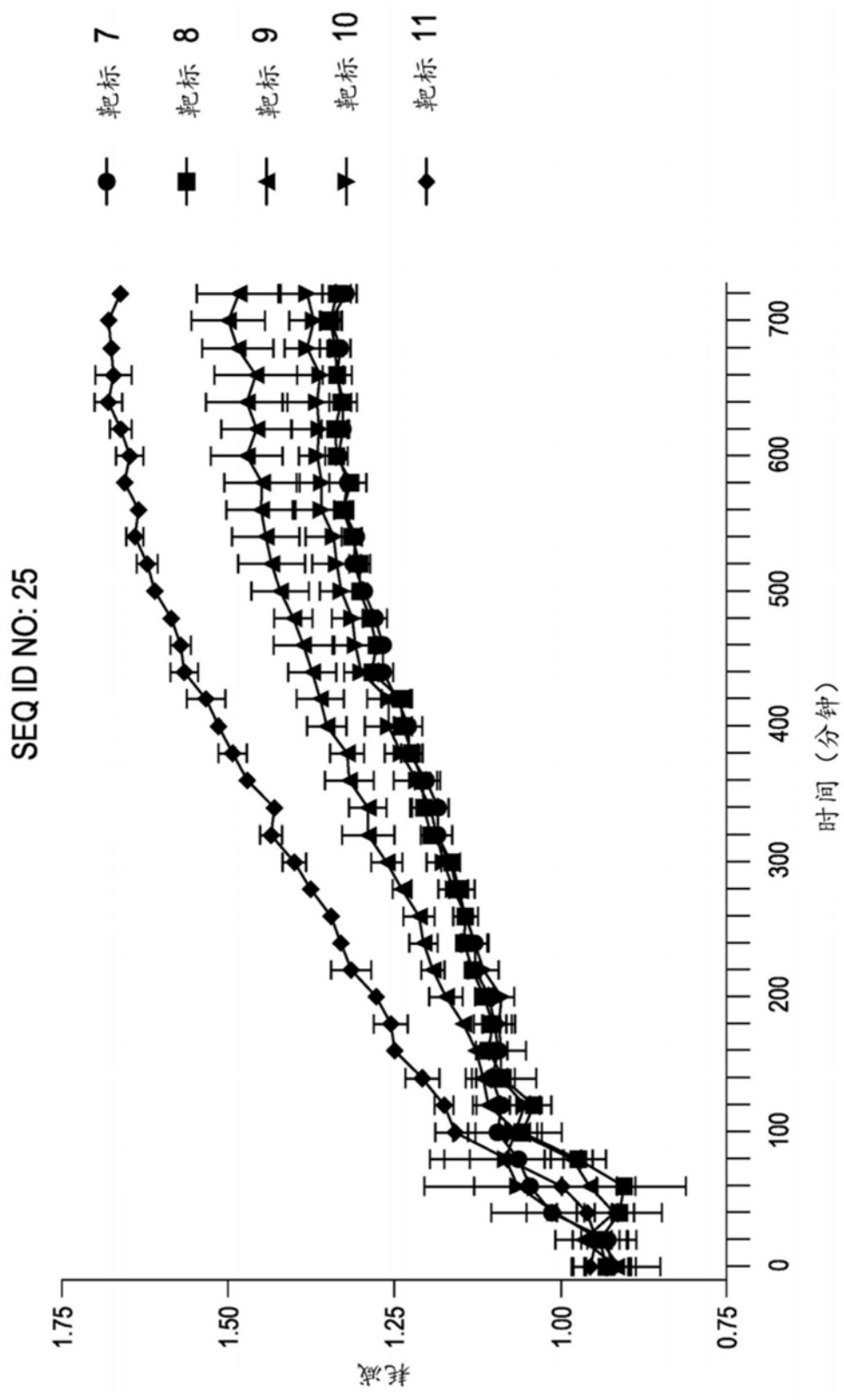

靶标 7
靶标 8
靶标 9
靶标 10
靶标 11
SEQ ID NO: 25
耗减
时间（分钟）
1.75
1.50
1.25
1.00
0.75
0
100
200
300
400
500
600
700

图36B

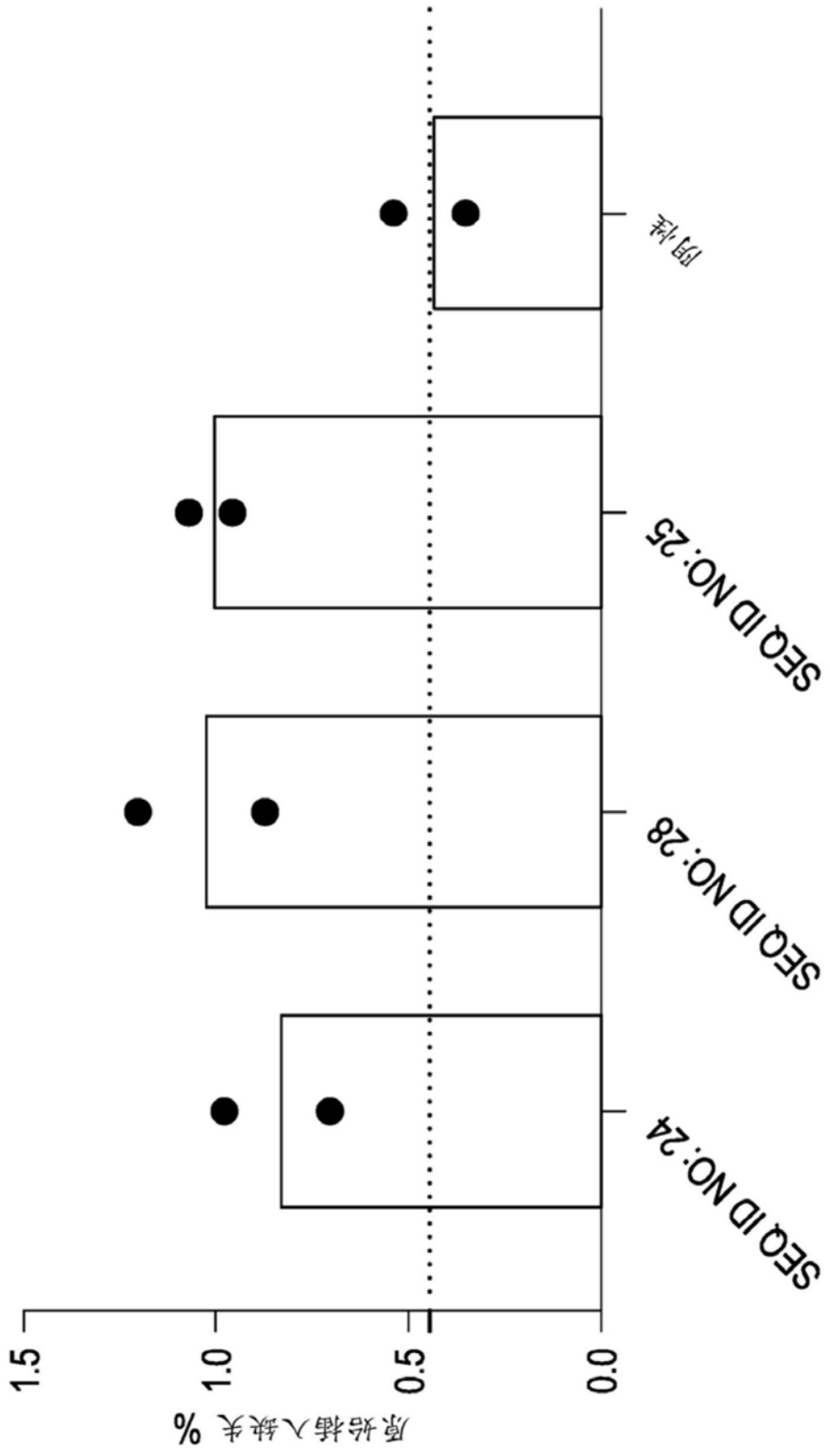


图37