(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 114269912 A
(43)申请公布日 2022.04.01

(21)申请号 202080050922.7

(22)申请日 2020.06.12

(30)优先权数据
62/861,582 2019.06.14 US
62/872,157 2019.07.09 US
62/873,602 2019.07.12 US
62/873,606 2019.07.12 US

(85)PCT国际申请进入国家阶段日
2022.01.13

(86)PCT国际申请的申请数据
PCT/US2020/037585 2020.06.12

(87)PCT国际申请的公布数据
WO2020/252378 EN 2020.12.17

(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

(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.
C12N 9/22(2006.01)
C12N 15/113(2006.01)
C12N 15/10(2006.01)
C12N 15/90(2006.01)
C07K 14/195(2006.01)

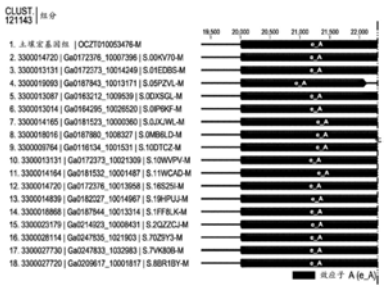
权利要求书3页 说明书115页
序列表373页 附图28页

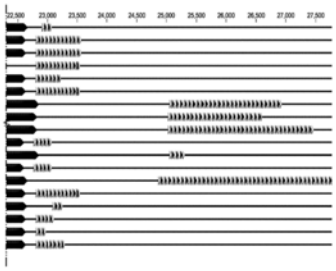
(54)发明名称

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诸如DNA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 114269912 A

1.一种工程化的非天然存在的CLUST.121143、CLUST.196682或CLUST.089537成簇规律间隔短回文重复序列(CRISPR)-Cas系统,该系统包括:

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;以及

CRISPR相关蛋白或编码该CRISPR相关蛋白的核酸,其中该CRISPR相关蛋白包含与表2、表8或表12中提供的氨基酸序列至少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;

其中该CRISPR相关蛋白能够结合该RNA指导物并且能够靶向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,其中该系统是CLUST.121143CRISPR-Cas系统并且该CRISPR相关蛋白包含与表2中提供的氨基酸序列至少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。

3.如权利要求2所述的系统,其中该CRISPR相关蛋白是CLUST.121143 3300014839效应子蛋白。

4.如权利要求2或3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),并且该靶核酸包含PAM,该PAM包括核酸序列5'-TTG-3'。

5.如权利要求2-4中任一项所述的系统,其中该直接重复序列包含与表3中提供的核苷酸序列至少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。

6.如权利要求2-5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22个核苷酸至约40个核苷酸。

7.如权利要求2-6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26个核苷酸至约35个核苷酸。

8.如权利要求1所述的系统,其中该系统是CLUST.196682CRISPR-Cas系统并且该CRISPR相关蛋白包含与表8中提供的氨基酸序列至少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。

9.如权利要求8所述的系统,其中该CRISPR相关蛋白是CLUST.196682 3300025638效应子蛋白。

10.如权利要求8或9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),并且该靶核酸包含PAM,该PAM包括核酸序列5'-CG-3'。

11.如权利要求8-10中任一项所述的系统,其中该直接重复序列包含与表9中提供的核苷酸序列至少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。

12.如权利要求8-11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17个核苷酸至约44个核苷酸。

13.如权利要求8-12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26至38个核苷酸。

14.如权利要求1所述的系统，其中该系统是CLUST.089537CRISPR-Cas系统并且该CRISPR相关蛋白包含与表12中提供的氨基酸序列至少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。

15.如权利要求14所述的系统，其中该CRISPR相关蛋白是CLUST.089537CAAACX010000652效应子蛋白。

16.如权利要求14或15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)，并且该靶核酸包含PAM，该PAM包括核酸序列5’-TTC-3’。

17.如权利要求14-16中任一项所述的系统，其中该直接重复序列包含与表13中提供的核苷酸序列至少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。

18.如权利要求14-17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20个核苷酸至约40个核苷酸。

19.如权利要求14-18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25至37个核苷酸。

20.如权利要求1-19中任一项所述的系统，其中该CRISPR相关蛋白包含至少一个(例如，一个、两个或三个)RuvC结构域。

21.如权利要求1-20中任一项所述的系统，其中该CRISPR相关蛋白包含拆分型RuvC结构域。

22.如权利要求1-21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)。

23.如权利要求1-22中任一项所述的系统，其中该CRISPR相关蛋白切割该靶核酸。

24.如权利要求1-23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。

25.如权利要求1-24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。

26.如权利要求1-25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。

27.如权利要求1-26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。

28.如权利要求27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。

29.如权利要求1-28中任一项所述的系统，其中该靶核酸是DNA。

30.如权利要求1-29中任一项所述的系统，其中该靶核酸包含PAM。

31.如权利要求1-30中任一项所述的系统，其中该CRISPR相关蛋白具有非特异性核酸酶活性。

32.如权利要求1-31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靶向导致该靶核酸中的修饰。

33.如权利要求32所述的系统,其中该靶核酸中的该修饰是双链切割事件。

34.如权利要求32所述的系统,其中该靶核酸中的该修饰是单链切割事件。

35.如权利要求32所述的系统,其中该靶核酸中的该修饰导致插入事件。

36.如权利要求32所述的系统,其中该靶核酸中的该修饰导致缺失事件。

37.如权利要求32-36中任一项所述的系统,其中该修饰导致细胞毒性或细胞死亡。

38.如权利要求1-37中任一项所述的系统,该系统进一步包括供体模板核酸。

39.如权利要求38所述的系统,其中该供体模板核酸是DNA。

40.如权利要求38所述的系统,其中该供体模板核酸是RNA。

41.如权利要求1-40中任一项所述的系统,其中该RNA指导物包括tracrRNA、调节子RNA或两者。

42.如权利要求1-41中任一项所述的系统,其中该系统进一步包括tracrRNA。

43.如权利要求1-42中任一项所述的系统,其中该系统进一步包括调节子RNA。

44.如权利要求1-43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。

45.如权利要求1-43中任一项所述的系统,该系统在细胞内。

46.如权利要求45所述的系统,其中该细胞是真核细胞。

47.如权利要求45所述的系统,其中该细胞是原核细胞。

48.一种使如权利要求1-47中任一项所述的系统与细胞中的该靶核酸结合的方法,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。

49.一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1-47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6月14日提交的美国临时申请62/861,582、于2019年7月9日提交的美国临时申请62/872,157、于2019年7月12日提交的美国临时申请62/873,606、以及于2019年7月12日提交的美国临时申请62/873,602,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6月12日的所述ASCII副本被命名为A2186-7021WO_SL.txt,并且大小是901,689字节。

技术领域

[0005] 本披露涉及用于控制涉及序列靶向和核酸编辑的基因表达的系统、方法和组合物,这些系统、方法和组合物使用与成簇规律间隔短回文重复序列(CRISPR)及其组分相关的载体系统。

背景技术

[0006] 基因组测序技术和分析的最近应用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基因测序技术所产生的巨量信息,需要在基因组和表观基因组操纵的技术的规模、效率和易用性方面进行相应的提高。这些新型基因组和表观基因组工程化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相关联(Cas)基因,统称为CRISPR-Cas或CRISPR/Cas系统,目前已被理解为对细菌和古细菌提供抗噬菌体感染的免疫。原核生物适应性免疫的CRISPR-Cas系统是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。

[0008] 参与宿主防御的系统的组分包括一种或多种能够修饰DNA或RNA的效应子蛋白和负责将这些蛋白活性靶向至噬菌体DNA或RNA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类系统由单效应子蛋白构成,该单效应子蛋白与crRNA复合以靶向DNA或RNA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。因此,新型2类系统的发现、工程化和优化可能导致广泛且强大的可编程技术用于基因组工程化和其他方面。

[0010] 本申请中的任何文献的引用或鉴定并不承认这个文献可供使用作为本发明的现有技术。

[0011] CRISPR-Cas系统是古细菌和细菌中针对外来遗传元件而防御物种的适应性免疫系统。以CRISPR-Cas9为例的2类CRISPR-Cas系统的表征和工程化为基因组编辑和其他方面的多样广泛生物技术应用铺平了道路。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

[0012] 本披露提供了针对新的单效应子2类CRISPR-Cas系统的非天然存在的工程化系统和组合物,以及用于从基因组数据库进行计算鉴定、将天然基因座开发成工程化系统、以及实验验证和应用转换的方法。这些新的效应子在序列上与现有2类CRISPR效应子的直系同源物和同源物是不同的,而且还具有独特的结构域组织。它们提供了另外的特征,这些另外的特征包括但不限于1)新型DNA/RNA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。

[0013] 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。

[0014] 本披露涉及新的CRISPR-Cas系统,该系统包括新发现的酶和用于产生最小系统的其他组分,这些最小系统可以在非天然环境中(例如,在除该系统最初发现于其中的那些之外的细菌中)使用。

[0015] 在一方面,本披露提供了工程化的非天然存在的CLUST.121143成簇规律间隔短回文重复序列(CRISPR)-Cas系统,这些系统包括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;以及CRISPR相关蛋白或编码该CRISPR相关蛋白的核酸,其中该CRISPR相关蛋白包含与表2中提供的氨基酸序列(例如,SEQ ID NO:1-17)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;其中该CRISPR相关蛋白能够结合该RNA指导物并且能够靶向与该间隔子序列互补的该靶核酸序列。在一些实施例中,该CRISPR相关蛋白具有至少一个(例如,一个、两个或三个)RuvC结构域。

[0016] 在本文所述的任一系统的一些实施例中,该CRISPR相关蛋白是CLUST.1211433300014839效应子蛋白。在一些实施例中,该CRISPR相关蛋白能够识别原型间隔子相邻基序(PAM),并且该靶核酸包含PAM,该PAM包括核酸序列5'-TTG-3'。

[0017] 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表3中提供的核苷酸序列至少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表3中提供的核苷酸序列。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22至约40个核苷酸(例如,26至35个核苷酸)。

[0018] 在本文所提供的任一系统的某些实施例中，该靶核酸是DNA。在本文所述的任一系统的一些实施例中，该靶核酸包含原型间隔子相邻基序(PAM)。

[0019] 在本文所提供的任一系统的某些实施例中，通过该CRISPR相关蛋白和RNA指导物对该靶核酸的靶向导致该靶核酸中的修饰(例如，单链或双链切割事件)。在一些实施例中，该修饰导致缺失事件。在一些实施例中，该修饰导致插入事件。在一些实施例中，该修饰导致细胞毒性。

[0020] 在一些实施例中，该CRISPR相关蛋白具有非特异性(即，“伴(collateral)”)核酸酶(例如，DNA酶或DNA酶)活性。在本文所提供的任一系统的某些实施例中，该系统进一步包括供体模板核酸(例如，DNA或RNA)。

[0021] 在本文所提供的任一系统的一些实施例中，该系统在细胞(例如，真核细胞(例如，哺乳动物细胞)或原核细胞(例如，细菌细胞))内。

[0022] 在本文所提供的任一系统的一些实施例中，该RNA指导物包括tracrRNA、调节子RNA或两者。在本文所提供的任一系统的一些实施例中，该系统包括tracrRNA。在本文所提供的任一系统的一些实施例中，该系统包括调节子RNA。在本文所提供的任一系统的一些实施例中，该系统不包括tracrRNA。

[0023] 在本文所提供的任一系统的一些实施例中，该CRISPR相关蛋白包含拆分型RuvC结构域。在一些实施例中，该CRISPR相关蛋白包含催化残基(例如，天冬氨酸或谷氨酸)。在一些实施例中，该CRISPR相关蛋白切割该靶核酸。在一些实施例中，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。在一些实施例中，编码该CRISPR相关蛋白的该核酸经密码子优化以在细胞中表达。在一些实施例中，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。在一些实施例中，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。

[0024] 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，其中该方法包括：(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。

[0025] 在另一方面，本披露提供了靶向和编辑靶核酸的方法，这些方法包括使该靶核酸与本文所述的任一系统接触。在另一方面，本披露提供了编辑靶核酸的方法，这些方法包括使该靶核酸与本文所述的任一系统接触。

[0026] 在一方面，本披露提供了工程化的非天然存在的CLUST.196682成簇规律间隔短回文重复序列(CRISPR)-Cas系统，这些系统包括：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；以及CRISPR相关蛋白或编码该CRISPR相关蛋白的核酸，其中该CRISPR相关蛋白包含与表8中提供的氨基酸序列至少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；其中该CRISPR相关蛋白能够结合该RNA指导物并且能够靶向与该间隔子序列互补的该靶核酸序列。

[0027] 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。

[0028] 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表9中提供的核苷酸序列至少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表9中提供的核苷酸序列。

[0029] 在本文所述的任一系统的一些实施例中,该CRISPR相关蛋白是CLUST.196682 3300025638效应子蛋白。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),并且该靶核酸包含PAM,该PAM包括核酸序列5'-CG-3'。

[0030] 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17个核苷酸至约44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26至38个核苷酸。

[0031] 在本文所述的任一系统的一些实施例中,该靶核酸是DNA。在本文所述的任一系统的一些实施例中,该靶核酸包含PAM。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。

[0032] 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靶向导致该靶核酸中的修饰。在本文所述的任一系统的一些实施例中,该靶核酸中的该修饰是双链切割事件或单链切割事件。在本文所述的任一系统的一些实施例中,该靶核酸中的该修饰导致插入事件或缺失事件。在本文所述的任一系统的一些实施例中,该修饰导致细胞毒性或细胞死亡。

[0033] 在本文所述的任一系统的一些实施例中,这些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或RNA。

[0034] 在本文所述的任一系统的一些实施例中,该RNA指导物包括tracrRNA、调节子RNA或两者。在本文所述的任一系统的一些实施例中,这些系统进一步包括tracrRNA和/或调节子RNA。在本文所提供的任一系统的一些实施例中,该系统不包括tracrRNA。

[0035] 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。在本文所述的任一系统的一些实施例中,这些系统在真核细胞或原核细胞内。

[0036] 在本文所提供的任一系统的一些实施例中,该CRISPR相关蛋白包含拆分型RuvC结构域。在一些实施例中,该CRISPR相关蛋白包含催化残基(例如,天冬氨酸或谷氨酸)。在一些实施例中,该CRISPR相关蛋白切割该靶核酸。在一些实施例中,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。在一些实施例中,编码该CRISPR相关蛋白的该核酸经密码子优化以在细胞中表达。在一些实施例中,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。在一些实施例中,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。

[0037] 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,其中该方法包括:(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。

[0038] 在一方面，本披露提供了靶向和编辑靶核酸的方法，这些方法包括使该靶核酸与本文所述的任一系统接触。在一方面，本披露提供了编辑靶核酸的方法，这些方法包括使该靶核酸与本文所述的任一系统接触。

[0039] 在一方面，本披露提供了使有效载荷核酸在靶核酸的位点处靶向插入的方法，这些方法包括使该靶核酸与本文所述的系统接触。

[0040] 在一方面，本披露提供了工程化的非天然存在的CLUST.089537成簇规律间隔短回文重复序列(CRISPR)-Cas系统，这些系统包括：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；以及CRISPR相关蛋白或编码该CRISPR相关蛋白的核酸，其中该CRISPR相关蛋白包含与表12中提供的氨基酸序列至少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；其中该CRISPR相关蛋白能够结合该RNA指导物并且能够靶向与该间隔子序列互补的该靶核酸序列。

[0041] CLUST.089537和CLUST.085589可以互换使用。

[0042] 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。

[0043] 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表13中提供的核苷酸序列至少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表13中提供的核苷酸序列。

[0044] 在本文所述的任一系统的一些实施例中，该CRISPR相关蛋白是CLUST.089537CAAACX010000652效应子蛋白。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)，并且该靶核酸包含PAM，该PAM包括核酸序列5’-TTC-3’。

[0045] 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20个核苷酸至约40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25至37个核苷酸。

[0046] 在本文所述的任一系统的一些实施例中，该靶核酸是DNA。在本文所述的任一系统的一些实施例中，该靶核酸包含PAM。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。

[0047] 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靶向导致该靶核酸中的修饰。例如，在本文所述的任一系统中，该靶核酸中的该修饰是双链切割事件。在本文所述的任一系统的其他实施例中，该靶核酸中的该修饰是单链切割事件、插入事件或缺失事件。在本文所述的任一系统的一些实施例中，该修饰导致细胞毒性或细胞死亡。

[0048] 在本文所述的任一系统的一些实施例中，这些系统进一步包括供体模板核酸。例如，该供体模板核酸可以是DNA或RNA。

[0049] 在本文所述的任一系统的一些实施例中，该RNA指导物包括tracrRNA、调节子RNA或两者。在本文所述的任一系统的一些实施例中，这些系统进一步包括tracrRNA和/或调节子RNA。在本文所提供的任一系统的一些实施例中，该系统不包括tracrRNA。

[0050] 在本文所述的任一系统的一些实施例中，这些系统在细胞，例如真核细胞或原核细胞内。

[0051] 在本文所提供的任一系统的一些实施例中，该CRISPR相关蛋白包含拆分型RuvC结构域。在一些实施例中，该CRISPR相关蛋白包含催化残基(例如，天冬氨酸或谷氨酸)。在一些实施例中，该CRISPR相关蛋白切割该靶核酸。在一些实施例中，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。在一些实施例中，编码该CRISPR相关蛋白的该核酸经密码子优化以在细胞中表达。在一些实施例中，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。在一些实施例中，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。

[0052] 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，其中该方法包括：(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。

[0053] 在一方面，本披露提供了靶向和编辑靶核酸的方法，这些方法包括使该靶核酸与本文所述的任一系统接触。在一方面，本披露提供了编辑靶核酸的方法，这些方法包括使该靶核酸与本文所述的任一系统接触。

[0054] 在本文所提供的任一系统的一些实施例中，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相关蛋白和指导RNA到达该靶核酸。在本文所提供的任一系统的一些实施例中，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相关蛋白和指导RNA到达该细胞或在该细胞中产生。

附图说明

[0055] 图是表示各种蛋白质聚簇的基因座分析结果的一系列示意图以及核酸和氨基酸序列。

[0056] 图1是示出了用于代表性CLUST.121143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。

[0057] 图2是示出了用于CLUST.121143的CRISPR直接重复元件的实例的多序列比对的一系列序列。图2按照出现的顺序分别披露了SEQ ID NO:825、101、101、101、101、101、101、106、106、106、106、102、103、103、104、104、104、107、107、107、107、107、107、108、108、108、108、108、108、108、110、109和105。

[0058] 图3是CLUST.121143效应子蛋白的系统发育树的示意图。

[0059] 图4是CLUST.121143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;并且指示了锌指结构域的保守催化残基。

[0060] 图5是示出了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的图。为了定量耗减,针对每种直接重复和间隔子将富集比计算为$R_{处理的}/R_{输入}$。将归一化输入读段计数计算为:

[0061] $R_{输入}$=含DR+间隔子的读段#/总读段

[0062] 其中读段计数从在转化前对表达CLUST.121143效应子和相关crRNA的质粒DNA文库的下一代测序获得。将归一化处理的读段计数计算为:

[0063] $R_{处理的}$=(1+含DR+间隔子的读段#)/总读段#

[0064] 其中读段计数从在抗生素筛选之后对自表达CLUST.121143效应子和相关crRNA的存活细胞提取的质粒DNA的下一代测序获得。强耗减靶标具有小于1/10的富集。在呈“正向”取向(5’-CTTT…AGAG-[间隔子]-3’)的直接重复的情况下和在呈“反向”取向(5’-CTCT…AAAG-[间隔子]-3’)的直接重复的情况下CLUST.121143 3300014839的耗减程度描绘于图中。

[0065] 图6是示出了阴性对照的耗减活性程度的图,其中在编码CLUST.121143效应子的序列从效应子质粒中缺失的情况下重复筛选。

[0066] 图7A和7B是按pACYC184质粒和大肠杆菌菌株E.Cloni上的位置示出了CLUST.121143 3300014839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。图7A和7B两者披露了SEQ ID NO:826。

[0067] 图8是侧接CLUST.121143 3300014839的耗减靶标的序列的weblogo。图8披露了SEQ ID NO:826。

[0068] 图9是示出了用于代表性CLUST.196682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。

[0069] 图10是示出了用于CLUST.196682的CRISPR直接重复元件的实例的多序列比对的一系列序列。图10按照出现的顺序分别披露了SEQ ID NO:827、714、714、714、733、701、704、705、705、703、734、706、743、720、720、740、740、731、731、742、742、828、732、730、730和741。

[0070] 图11是CLUST.196682效应子蛋白的系统发育树的示意图。

[0071] 图12是CLUST.196682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;并且指示了锌指结构域的保守催化残基。

[0072] 图13A和13B是按pACYC184质粒和大肠杆菌菌株E.Cloni上的位置示出了CLUST.196682 3300025638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。图13A和13B两者披露了SEQ ID NO:829。

[0073] 图14是侧接CLUST.196682 3300025638的耗减靶标的序列的weblogo。图14披露了SEQ ID NO:829。

[0074] 图15是示出了用于代表性CLUST.089537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。

[0075] 图16是示出了用于CLUST.089537的CRISPR直接重复元件的实例的多序列比对的一系列序列。图16按照出现的顺序分别披露了SEQ ID NO:830、402、429、403、419、431、436、432、441、404、455、446、447、411、414、415、440、437、438、430、412、425、448、452、406、407、422、442、443、424、409、413、420、418、426、444、416、445、434、401、410、449、454、417、427、

423、450、428、439、435、453、421、408、405、433和451。

[0076] 图17是CLUST.089537效应子蛋白的系统发育树的示意图。

[0077] 图18是CLUST.089537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；并且指示了锌指结构域的保守催化残基。

[0078] 图19是示出了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的图。为了定量耗减，针对每种直接重复和间隔子将富集比计算为$R_{处理的}/R_{输入}$。将归一化输入读段计数计算为：

[0079] $R_{输入}$＝含DR+间隔子的读段#/总读段

[0080] 其中读段计数从在转化前对表达CLUST.089537效应子和相关crRNA的质粒DNA文库的下一代测序获得。将归一化处理的读段计数计算为：

[0081] $R_{处理的}$＝(1+含DR+间隔子的读段#)/总读段#

[0082] 其中读段计数从在抗生素筛选之后对自表达CLUST.089537效应子和相关crRNA的存活细胞提取的质粒DNA文库的下一代测序获得。强耗减靶标具有小于1/10的富集。在呈“正向”取向(5’-GTGC…ACAG-[间隔子]-3’)的直接重复的情况下和在呈“反向”取向(5’-CTGT…GCAC-[间隔子]-3’)的直接重复的情况下CLUST.089537CAAACX010000652的耗减程度描绘于图中。

[0083] 图20A和20B是按pACYC184质粒和大肠杆菌菌株E.Cloni上的位置示出了CLUST.089537CAAACX010000652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。图20A和20B两者披露了SEQ ID NO:831。

[0084] 图21是侧接CLUST.089537CAAACX010000652的耗减靶标的序列的weblogo。图21披露了SEQ ID NO:831。

具体实施方式

[0085] CRISPR-Cas防御系统的广泛天然多样性包含了广泛范围的可用于可编程生物技术的活性机制和功能元件。在一个自然系统中，这些机制和参数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。在工程化系统中，相同的机制和参数也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。例如，系统Cas9和Cas13a具有典型的DNA和RNA核酸内切酶活性，并且它们的靶向空间分别由靶向的DNA上的原型间隔子相邻基序(PAM)和靶向的RNA上的原型间隔子侧接位点(PFS)限定。

[0086] 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可以扩展RNA可编程核酸操纵的能力。

[0087] 在一方面，本披露涉及使用计算方法和算法来搜索和鉴定在天然存在的基因组序列内展现出与某些其他特征强共现模式的新型蛋白质家族。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家族。然而，本文所披露的方法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。

[0088] 在一些实施例中，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)。

[0089] 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。

[0090] 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合被去重复。

[0091] 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合上使用多个CRISPR阵列鉴定工具，其中所得的CRISPR阵列集合被去重复。

[0092] 在一些实施例中，鉴定与CRISPR阵列很靠近的蛋白质。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来形成蛋白质聚簇。在某些其他实施例中，使用mmseqs2来形成蛋白质聚簇。

[0093] 为了建立蛋白质聚簇的成员与CRISPR阵列之间的强共现模式，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家族的每个成员的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。

[0094] 在一些实施例中，通过度量对很靠近CRISPR阵列内的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相关蛋白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。

[0095] 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。

[0096] 在一些实施例中，RNA指导物与本文所述的CRISPR相关蛋白之间的复合物是已结合或已修饰靶核酸的激活的CRISPR复合物。

[0097] 定义

[0098] 本文所用的术语“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DNA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。

[0099] 如本文所用，术语“CRISPR-Cas系统”是指参与CRISPR-Cas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-Cas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。

[0100] 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”或“CRISPR直接重复”或“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。

[0101] 如本文所用，术语“CRISPR RNA”或“crRNA”是指包含由CRISPR效应子用于特异性靶向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:tracrRNA双链体结合CRISPR效应子。如本文所用，术语“供体模板核酸”是指这样的核酸分子，该核酸分子可以在本文所述的CRISPR酶改变了靶核酸之后由一种或多种细胞蛋白质用于改变靶核酸的结构。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)。

[0102] 如本文所用，术语“CRISPR-Cas效应子”、“CRISPR效应子”、“效应子”、“CRISPR相关蛋白”或“CRISPR酶”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、核酸外切酶活性、转座酶活性和/或切除活性。

[0103] 如本文所用，术语“指导RNA”或“gRNA”是指能够指导具有核酸酶活性的CRISPR效应子以靶向并且切割指定靶核酸的RNA分子。

[0104] 如本文所用，术语“RNA指导物”是指促进本文所述的蛋白质靶向至靶核酸的任何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融合的crRNA。在一些实施例中，RNA指导物包括crRNA和tracrRNA两者。在一些实施例中，RNA指导

物包括crRNA和调节子RNA。在一些实施例中,RNA指导物包括crRNA、tracrRNA和调节子RNA。

[0105] 如本文所述的术语“调节子RNA”是指调节(例如,增加或降低)CRISPR-Cas效应子或包含CRISPR-Cas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-Cas效应子或包含CRISPR-Cas效应子的核蛋白复合物的核酸酶活性。

[0106] 如本文所用,术语“靶核酸”是指待通过本文所述的CRISPR系统修饰的特定核酸序列。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。

[0107] 如本文所用,术语“反式激活crRNA”或“tracrRNA”是指这样的RNA,该RNA包含形成为CRISPR效应子结合指定靶核酸所需的结构的序列。

[0108] 如本文所用,“转录活性位点”是指包含启动子区域的核酸序列中起始且主动发生转录的位点。

[0109] 如本文关于CRISPR酶所用,术语“伴RNA酶活性”是指在CRISPR酶已修饰特异性靶向的核酸之后该酶的非特异性RNA酶活性。

[0110] 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。

[0111] 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。

[0112] 池化筛选

[0113] 为了有效地验证工程化的新型CRISPR-Cas系统的活性,并且同时以无偏的方式评价不同的活性机制和功能参数,在大肠杆菌中使用了新的池化筛选方法。首先,根据新型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单一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。

[0114] 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文库克隆至含有蛋白效应子和非编码元件的载体骨架(例如,pET-28a+)中,并且然后随后将文库连同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间隔子。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,必需基因靶向间隔子可与pACYC184靶组合以增加测定的另一维度。

[0115] 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-Cas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。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。在一些实施例中，使用双重抗生素选择。例如，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。在一些实施例中，仅使用卡那霉素来确保成功转化含有工程化CRISPR-Cas效应子系统的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。

[0116] 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，该活性模式可以以广泛的与假设无关的方式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。

[0117] 本文所述的体内池化筛选的关键优点包括：

[0118] (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；

[0119] (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DNA或RNA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；

[0120] (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

[0121] (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。

[0122] 将本文所述的新型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，以及其在其天然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。

[0123] 具有RuvC结构域的2类CRISPR-Cas效应子

[0124] 在一方面，本披露提供了在本文中称为CLUST.121143的2类CRISPR-Cas系统。此2类CRISPR-Cas系统含有具有RuvC结构域的分离的CRISPR相关蛋白。

[0125] 在一方面，本披露提供了在本文中称为CLUST.196682的2类CRISPR-Cas系统。此2类CRISPR-Cas系统含有具有RuvC结构域的分离的CRISPR相关蛋白。

[0126] 在一方面，本披露提供了在本文中称为CLUST.089537的2类CRISPR-Cas系统。此2类CRISPR-Cas系统含有具有RuvC结构域的分离的CRISPR相关蛋白。

[0127] 在一些实施例中，CRISPR相关蛋白和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。

[0128] 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。

[0129] 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(例如，单链或双链DNA)。在一些实施例中，非靶标核酸是RNA(例如，单链或双链RNA)。

[0130] CRISPR酶修饰

[0131] 去激活/失活的CRISPR酶

[0132] 当本文所述的CRISPR酶具有核酸酶活性时，可以将CRISPR酶修饰为具有减弱的核酸酶活性，例如与野生型CRISPR酶相比，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。

[0133] 失活的CRISPR酶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。

[0134] 一个或多个功能性结构域在失活的CRISPR酶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酶的N末端。在一些实施例中，功能性结构域位于CRISPR酶的C末端。在一些实施例中，将失活的CRISPR酶修饰为包含在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。

[0135] 拆分型酶

[0136] 本披露还提供了一种本文所述的CRISPR酶的拆分形式。CRISPR酶的拆分形式可能对于递送是有利的。在一些实施例中，CRISPR酶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酶。

[0137] 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酶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。

[0138] 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，但指导RNA将它们招募至三元复合物中，该三元复合物重现全长CRISPR酶的活性，并且催化位点特异性DNA切割。使用经修饰的指导RNA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,Addison V.等人."Rational design of a split-Cas9 enzyme complex[拆分型Cas9酶复合物的合理设计],"Proc.Nat'l.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。

[0139] 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酶用于CRISPR酶活性的时间控制。因此,CRISPR酶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酶的受控重组。

[0140] 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酶的两个部分或片段(即,N末端和C末端片段)可以形成包含野生型CRISPR酶序列的例如至少70%、至少80%、至少90%、至少95%或至少99%的完全CRISPR酶。

[0141] 自激活或自失活酶

[0142] 本文所述的CRISPR酶可以被设计为自激活或自失活的。在一些实施例中,CRISPR酶是自失活的。例如,可以将靶序列引入至CRISPR酶编码构建体中。因此,CRISPR酶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,Benjamin E.和David V.Schaffer."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。

[0143] 在一些其他实施例中,在弱启动子(例如,7SK启动子)控制下表达的另外指导RNA可以靶向编码CRISPR酶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酶、指导RNA和靶向编码CRISPR酶的核酸的指导RNA的载体转染细胞可以导致编码CRISPR酶的核酸的有效破坏并降低CRISPR酶的水平,从而限制基因组编辑活性。

[0144] 在一些实施例中,CRISPR酶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酶开关可以通过在编码CRISPR酶的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,Moe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)。

[0145] 诱导型CRISPR酶

[0146] CRISPR酶可以是诱导型的,例如,光诱导型的或化学诱导型的。此机制允许激活CRISPR酶中的功能性结构域。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酶中(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酶中。雷帕霉素是形成融合复合物从而激活CRISPR酶所需的(参见例如,Zetsche,Volz和Zhang,“A split-Cas9 architecture for inducible genome editing and transcription modulation[用于诱导型基因组编辑和转录调节的拆分型Cas9架构],”Nature Biotech.[自然生物技术],33.2(2015):139-142)。

[0147] 此外,CRISPR酶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,Stephen J.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)。

[0148] 诱导型CRISPR酶和诱导型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。

[0149] 功能性突变

[0150] 可以将各种突变或修饰引入至如本文所述的CRISPR酶中,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酶修饰以识别不同的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酶可以识别例如5’-NGG-3’、5’-YG-3’、5’-TTTN-3’或5’-YTN-3’PAM,其中“Y”是嘧啶并且“N”是任何核碱基。

[0151] 在一些实施例中,将至少一个核定位信号(NLS)附接至编码CRISPR酶的核酸序列。在一些实施例中,将至少一个核输出信号(NES)附接至编码CRISPR酶的核酸序列。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。

[0152] 在一些实施例中,使本文所述的CRISPR酶在一个或多个氨基酸残基处突变以改变一种或多种功能活性。例如,在一些实施例中,使CRISPR酶在一个或多个氨基酸残基处突变以改变其解旋酶活性。在一些实施例中,使CRISPR酶在一个或多个氨基酸残基处突变以改变其核酸酶活性(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酶在一个或多个氨基酸残基处突变以改变其与指导RNA功能上相关联的能力。在一些实施例中,使CRISPR酶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。

[0153] 在一些实施例中,本文所述的CRISPR酶能够切割靶核酸分子。在一些实施例中,CRISPR酶切割靶核酸分子的两条链。然而,在一些实施例中,使CRISPR酶在一个或多个氨基酸残基处突变以改变其切割活性。例如,在一些实施例中,CRISPR酶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酶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单链(即,切口酶活性)。在一些实施例中,CRISPR酶能够切割靶核酸的与跟指导RNA杂交的链互补的链。在一些实施例中,CRISPR酶能够切割靶核酸的与指导RNA杂交的链。

[0154] 在一些实施例中,本文所述的CRISPR酶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,核酸酶活性和与指导RNA功能上相互作用的能力)。截短型CRISPR酶可以有利地与具有负载限制的递送系统组合使用。

[0155] 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%相同的氨基酸序列。

[0156] 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。

[0157] 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。

[0158] 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。

[0159] 指导RNA修饰

[0160] 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、9或13中任一个的直接重复序列包含这样的序列,该序列在表3、9或13中任一个中的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。

[0161] 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)。

[0162] 间隔子长度

[0163] 在一些实施例中,CRISPR-Cas系统是CLUST.121143CRISPR-Cas系统。指导RNA的间隔子长度可以在从约15至50个核苷酸的范围内。在一些实施例中,指导RNA的间隔子长度是至少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、至少22个核苷酸、至少23个核苷酸、至少24个核苷酸、至少25个核苷酸、至少26个核苷酸或至少27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33个核苷酸(例如,30、31、32或33个核苷酸)、从33至36个核苷酸(例如,33、34、35或36个核苷酸)、从36至40个核苷酸(例如,36、37、38、39或40个核苷酸)、从40至45个核苷酸(例如,40、41、42、43、44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。在一些实施例中,间隔子长度是在22至40个核苷酸之间,或是从26至35个核苷酸。在一些实施例中,指导RNA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,指导RNA的直接重复长度是19个核苷酸。

[0164] 在一些实施例中,CRISPR-Cas系统是CLUST.196682CRISPR-Cas系统。指导RNA的间隔子长度可以在从约15至50个核苷酸的范围内。在一些实施例中,指导RNA的间隔子长度是至少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。在一些实施例中,指导RNA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,指导RNA的直接重复长度是19个核苷酸。

[0165] 在一些实施例中,CRISPR-Cas系统是CLUST.089537 CRISPR-Cas系统。指导RNA的间隔子长度可以在从约15至50个核苷酸的范围内。在一些实施例中,指导RNA的间隔子长度是至少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。在一些实施例中,指导RNA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,指导RNA的直接重复长度是19个核苷酸。

[0166] 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)。

[0167] 指导RNA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RNA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指导RNA短5%、10%、20%、30%、40%或50%。指导RNA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)。

[0168] 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CRISPR酶、和指导RNA(gRNA),其中gRNA包含死指导序列,由此gRNA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。

[0169] 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。

[0170] 诱导型指导物

[0171] 指导RNA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。

[0172] 在一些实施例中,指导RNA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,将该两个文献通过援引以其全文并入本文。

[0173] 化学修饰

[0174] 可以将化学修饰应用于指导RNA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。

[0175] 可以将广泛多种的修饰应用于化学合成的指导RNA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。

[0176] 在一些实施例中，指导RNA包含一个或多个硫代磷酸酯修饰。在一些实施例中，指导RNA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。

[0177] 这些化学修饰的概述可见于例如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；WO2016205764；以及US 8795965 B2中；将这些文献中的每一个通过援引以其全文并入。

[0178] 序列修饰

[0179] 本文所述的指导RNA、tracrRNA和crRNA的序列和长度可以是优化的。在一些实施例中，指导RNA的优化长度可以通过鉴定tracrRNA和/或crRNA的加工形式，或者通过指导RNA、tracrRNA、crRNA和tracrRNA四元环的经验长度研究来确定。

[0180] 指导RNA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，指导RNA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，将这些文献通过援引以其全文并入本文。

[0181] 指导物:靶序列匹配要求

[0182] 在经典的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％。在一些实施例中，互补程度是100％。指导RNA的长度可以是约5、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、35、40、45、50、75、或更多个核苷酸。

[0183] 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%。

[0184] 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。

[0185] CRISPR系统的使用方法

[0186] 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。

[0187] 在另一方面，本披露提供了使有效载荷核酸在靶核酸的位点处靶向插入或切除的方法，其中这些方法包括使靶核酸与本文所述的任一系统接触。

[0188] 在另一方面，本披露提供了在识别DNA靶核酸后非特异性降解单链DNA的方法，该方法包括使靶核酸与本文所述的任一系统接触。

[0189] 在另一方面，本披露提供了检测样品中的靶核酸(例如，DNA或RNA)的方法，这些方法包括使样品与本文所述的任一系统和经标记的报告核酸接触，其中crRNA与靶核酸的杂交引起经标记的报告核酸的切割，以及测量由经标记的报告核酸的切割产生的可检测信号，从而检测样品中靶核酸的存在。

[0190] 在又另一方面，本披露提供了治疗有需要的受试者的病症或疾病的方法，这些方法包括向受试者施用本文所述的任一系统，其中间隔子序列与跟病症或疾病相关联的靶核酸的22至40个核苷酸互补；其中CRISPR相关蛋白与RNA指导物缔合形成复合物；其中复合物结合靶核酸序列，该靶核酸序列与间隔子序列的核苷酸互补；并且其中在复合物与靶核酸序列结合后，CRISPR相关蛋白切割靶核酸，从而治疗受试者的病症或疾病。

[0191] 在一方面，本披露提供了治疗有需要的受试者的病症或疾病的方法，这些方法包括向受试者施用本文所述的任一系统，其中间隔子序列与跟病症或疾病相关联的靶核酸的17至44个核苷酸互补；其中CRISPR相关蛋白与RNA指导物缔合形成复合物；其中复合物结合靶核酸序列，该靶核酸序列与间隔子序列的核苷酸互补；并且其中，在复合物与靶核酸序列结合后，CRISPR相关蛋白切割靶核酸，从而治疗受试者的病症或疾病。

[0192] 在一方面，本披露提供了治疗有需要的受试者的病症或疾病的方法，这些方法包括向受试者施用本文所述的任一系统，其中间隔子序列与跟病症或疾病相关联的靶核酸的20至40个核苷酸互补；其中CRISPR相关蛋白与RNA指导物缔合形成复合物；其中复合物结合靶核酸序列，该靶核酸序列与间隔子序列的核苷酸互补；并且其中在复合物与靶核酸序列结合后，CRISPR相关蛋白切割靶核酸，从而治疗受试者的病症或疾病。

[0193] 在另一方面，本披露提供了用于将本文所披露的任一系统作为药物(例如，用于在治疗或预防病症或疾病中使用)使用的方法。

[0194] 在一些实施例中，病症或疾病是感染性疾病或癌症(例如，维尔姆斯瘤(Wilms' tumor)、尤文肉瘤(Ewing sarcoma)、神经内分泌瘤、胶质母细胞瘤、神经母细胞瘤、黑色素瘤、皮肤癌、乳腺癌、结肠癌、直肠癌、前列腺癌、肝癌、肾癌、胰腺癌、肺癌、胆道癌、宫颈癌、子宫内膜癌、食管癌、胃癌、头颈癌、甲状腺髓样癌、卵巢癌、胶质瘤、淋巴瘤、白血病、骨髓瘤、急性成淋巴细胞性白血病、急性骨髓性白血病、慢性淋巴细胞性白血病、慢性骨髓性白血病、霍奇金氏淋巴瘤、非霍奇金氏淋巴瘤或尿膀胱癌)。

[0195] 在又另一方面，本披露提供了用于在编辑靶核酸，例如靶向和编辑靶核酸的方法(例如，体外或离体方法)中使用本文所披露的任一系统的方法；在识别DNA靶核酸后非特异性降解单链DNA；在识别双链靶DNA的间隔子互补链后靶向该双链靶DNA的非间隔子互补链并使其产生切口；靶向并切割双链靶DNA；检测样品中的靶核酸；特异性编辑双链核酸；对双链核酸进行碱基编辑；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；在双链靶DNA中产生插入缺失；将序列插入至双链靶DNA中；或者缺失或倒置双链靶DNA中的序列。

[0196] DNA/RNA检测

[0197] 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。

[0198] 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。

[0199] 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。

[0200] 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中，将该文献通过援引以其全文并入本文。

[0201] 核酸的跟踪和标记

[0202] 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WO2017070605中;将这些文献中的每一个通过援引以其全文并入本文。

[0203] 高通量筛选

[0204] 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酶转染的克隆。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中,将该文献通过援引以其全文并入本文。

[0205] 工程化微生物

[0206] 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。

[0207] 在一些实施例中,可以将靶向噬菌体序列的指导RNA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。

[0208] 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,将该两个文献通过援引以其全文并入本文。

[0209] 在植物中的应用

[0210] 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。

[0211] 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 A1中;将该两个文献通过援引以其全文并入本文。

[0212] 基因驱动

[0213] 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中,将该文献通过援引以其全文并入本文。

[0214] 池化筛选

[0215] 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。用本文所述的指导RNA(gRNA)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中,将该文献通过援引以其全文并入本文。

[0216] 饱和诱变("过度攻击(Bashing)")

[0217] 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,池化的指导RNA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。

[0218] 治疗性应用

[0219] 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)、可以通过核酸酶活性(例如,Pcsk9靶向、杜氏(Duchenne)肌营养不良(DMD)、BCL11a靶向)治疗的疾病。

[0220] 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。

[0221] 在本文所述的一些治疗方法中，病症或疾病选自下组，该组由以下组成：囊性纤维化、杜氏肌营养不良、贝克(Becker)肌营养不良、α-1-抗胰蛋白酶缺乏、庞贝病(Pompe Disease)、肌强直性营养不良、亨廷顿病、脆性X综合征、弗里德赖希共济失调(Friedreich's ataxia)、肌萎缩性侧索硬化症、额颞叶痴呆、遗传性慢性肾病、高脂血症、高胆固醇血症、莱伯氏先天性黑蒙(Leber Congenital Amaurosis)、镰状细胞病和β地中海贫血。

[0222] 在本文所述的方法(和用于在此类方法中使用的组合物)的一些实施例中，病症或疾病是癌症或感染性疾病。

[0223] 在一些实施例中，病症或疾病是癌症，并且其中癌症选自下组，该组由以下组成：维尔姆斯瘤、尤文肉瘤、神经内分泌瘤、胶质母细胞瘤、神经母细胞瘤、黑色素瘤、皮肤癌、乳腺癌、结肠癌、直肠癌、前列腺癌、肝癌、肾癌、胰腺癌、肺癌、胆道癌、宫颈癌、子宫内膜癌、食管癌、胃癌、头颈癌、甲状腺髓样癌、卵巢癌、胶质瘤、淋巴瘤、白血病、骨髓瘤、急性成淋巴细胞性白血病、急性骨髓性白血病、慢性淋巴细胞性白血病、慢性骨髓性白血病、霍奇金氏淋巴瘤、非霍奇金氏淋巴瘤和尿膀胱癌。

[0224] 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。

[0225] 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)。

[0226] 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，细胞的分子机器将在修复和/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系统可以用于改变靶核酸，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PCT公开号WO 2016094874 A1中，将该文献的全部内容通过援引明确并入本文。

[0227] 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。

[0228] 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 A1中,将该两个文献通过援引以其全文并入本文。本领域技术人员将理解如何使用新的CRISPR系统治疗这些疾病。

[0229] 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。

[0230] 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。

[0231] 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序列的合适指导RNA靶向病毒RNA。

[0232] 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。

[0233] 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。

[0234] CRISPR系统的递送

[0235] 通过本披露和本领域中的知识,本文所述的CRISPR系统、或其组分、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、病毒递送载体)来递送。新的CRISPR酶和/或任何RNA(例如,指导RNA)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。可以将蛋白质和一种或多种指导RNA包装至一种或多种载体,例如质粒或病毒载体中。

[0236] 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,该多种因素诸如载体选择、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、所寻求的

转化/修饰的类型等。

[0237] 在某些实施例中，递送是经由腺病毒进行的，这些腺病毒可以是在含有至少$1x10^5$个腺病毒颗粒(也称为颗粒单位，pu)的单剂量下。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。递送方法和剂量例如描述于WO 2016205764 A1和美国专利号8,454,972 B2中，将该两个文献通过援引以其全文并入本文。

[0238] 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酶的序列；(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。

[0239] 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和美国专利号5,593,972、5,589,466和5,580,859中；将这些文献中的每一个通过援引以其全文并入本文。

[0240] 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。

[0241] 将新的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酶。在一些实施例中，将CRISPR酶和/或指导RNA偶联至一种或多种CPP以便将它们有效地转运在细胞(例如，植物原生质体)内。在一些实施例中，将CRISPR酶和/或一种或多种指导RNA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。

[0242] 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 A1中；将这些文献中的每一个通过援引以其全文并入本文。

[0243] 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。

[0244] 实例

[0245] 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。

[0246] 实例1-对CLUST.121143CRISPR-Cas系统的最小组分的鉴定(图1-4)

[0247] 此蛋白质家族描述了与从淡水环境收集的未培养宏基因组序列中发现的CRISPR系统相关联的大型单效应子(图3)。CLUST.121143效应子包括下表1和2中详述的蛋白质实例。这些系统的直接重复序列的实例示出于表3中。

[0248] 含有此效应子复合物的天然存在的基因座的实例描绘于图1中,该图表明对于CLUST.121143基因座,效应子蛋白与CRISPR阵列共现。在双向15kb窗口内没有鉴定出与效应子蛋白或CRISPR阵列共现的其他大蛋白质家族。

[0249] 针对CLUST.121143的直接重复序列展现出靠近3'端的5'-TGCAAAAG-3'或5'-TTTCAAAG-4'的保守基序(图2)。

[0250] 图3是CLUST.121143效应子的系统发育树的示意图,该示意图显示该家族展现出序列多样性。

[0251] 图4是CLUST.121143效应子蛋白的多序列比对的示意图,该示意图揭示出由颜色品红色、蓝色和红色指示的RuvC结构域的保守催化残基的位置。

[0252] • 任选地,本文所述的CLUST.121143 CRISPR系统包括如表4中详述的具有DR同源性的反式激活RNA(tracrRNA)和包含在表5中详述的DR同源性基因座中的完整tracrRNA。任选地,该系统包括作为表6中列出的非编码序列的子集的tracrRNA。

[0253] • 任选地,本文所述的CLUST.121143 CRISPR系统包括由表6中列出的非编码序列(或其片段)编码的RNA调节子。

[0254] 表1.代表性CLUST.121143效应子蛋白

[0255]

物种	效应子登录号	间隔子#	cas1	cas2	效应子大小
水生-淡水(3300013006|Ga0164294_10029128)	3300013006|Ga0164294_10029128_3|M	11	N	N	878
水生-淡水(3300013087|Ga0163212_1009539)	3300013087|Ga0163212_1009539_3|M	6	N	N	877
水生-淡水(3300013131|Ga0172373_10014249)	3300013131|Ga0172373_10014249_3|P	11	N	N	877
水生-淡水(3300013131|Ga0172373_10021309)	3300013131|Ga0172373_10021309_4|M	4	N	N	867
水生-淡水(3300013132|Ga0172372_10046622)	3300013132|Ga0172372_10046622_2|M	4	N	N	867
水生-淡水(3300014720|Ga0172376_10007396)	3300014720|Ga0172376_10007396_6|M	11	N	N	877
水生-淡水(3300014720|Ga0172376_10013958)	3300014720|Ga0172376_10013958_4|M	4	N	N	867
水生-淡水(3300018790|Ga0187842_1006374)	3300018790|Ga0187842_1006374_2|M	11	N	N	878
水生-淡水(3300018815|Ga0187845_1009203)	3300018815|Ga0187845_1009203_1|M	11	N	N	878
水生-淡水(3300018868|Ga0187844_10013314)	3300018868|Ga0187844_10013314_2|M	11	N	N	878
水生-淡水(3300019093|Ga0187843_10013171)	3300019093|Ga0187843_10013171_2|M	11	N	N	708

[0256]

水生-淡水（3300023179|Ga0214923_10008431）	3300023179|Ga0214923_10008431_2|P	2	N	N	885
水生-淡水（3300027730|Ga0247833_1032983）	3300027730|Ga0247833_1032983_1|M	2	N	N	878
水生-淡水（3300028114|Ga0247835_1021903）	3300028114|Ga0247835_1021903_1|P	4	N	N	878
水生-淡水（3300028553|Ga0247839_1034275）	3300028553|Ga0247839_1034275_1|M	2	N	N	878
水生-淡水（3300028569|Ga0247843_1031164）	3300028569|Ga0247843_1031164_2|P	5	N	N	878
水生-淡水-沼泽（3300014164|Ga0181532_10001487）	3300014164|Ga0181532_10001487_14|M	3	Y	Y	951
水生-淡水-沼泽（3300014164|Ga0181532_10001487）	3300014164|Ga0181532_10001487_14|P	3	Y	Y	1000
水生-淡水-沼泽（3300014165|Ga0181523_10000360）	3300014165|Ga0181523_10000360_35|M	26	Y	Y	951
水生-淡水-沼泽（3300014165|Ga0181523_10000360）	3300014165|Ga0181523_10000360_33|P	26	Y	Y	1000
水生-淡水-沼泽（3300014200|Ga0181526_10000193）	3300014200|Ga0181526_10000193_31|M	26	Y	Y	951
水生-淡水-沼泽（3300014200|Ga0181526_10000193）	3300014200|Ga0181526_10000193_30|P	26	Y	Y	1000
水生-淡水-淡水和沉积物（3300027720|Ga0209617_10001817）	3300027720|Ga0209617_10001817_3|M	7	N	N	878
水生-淡水-泥炭地（3300009764|Ga0116134_1001531）	3300009764|Ga0116134_1001531_13|P	34	Y	Y	941
水生-淡水-泥炭地（3300018009|Ga0187884_10009678）	3300018009|Ga0187884_10009678_2|M	8	Y	Y	941
水生-淡水-泥炭地（3300018014|Ga0187860_1004616）	3300018014|Ga0187860_1004616_12|M	34	Y	Y	941
水生-淡水-泥炭地（3300018016|Ga0187880_1008327）	3300018016|Ga0187880_1008327_2|M	22	Y	Y	941
水生-淡水-泥炭地（3300018017|Ga0187872_10010132）	3300018017|Ga0187872_10010132_5|M	6	Y	Y	941
水生-淡水-泥炭地（3300018022|Ga0187864_10004559）	3300018022|Ga0187864_10004559_12|P	34	Y	Y	941
水生-淡水-泥炭地（3300018040|Ga0187862_10006434）	3300018040|Ga0187862_10006434_12|P	34	Y	Y	941
土壤宏基因组（OCZT010053476）	OCZT010053476_2|M	2	N	N	887
陆生-土壤-沼地（3300014839|Ga0182027_10014967）	3300014839|Ga0182027_10014967_5|M	49	Y	Y	888

[0257] 表2.代表性CLUST.121143效应子蛋白的氨基酸序列

[0258]

```
>3300013006|Ga0164294_10029128_3|M
[水生-淡水]
MNHKSTVKASVRHLTRHALQKPGEESDPGETTKTLKFRVDAKGHTRTRLDQIFTAQRQFVRDLLNKLEGMWDKDPQRFIRMVES
SASDPFESKTSCYGWMMTKFLTGSDLPEHLSRSAATASVATLSGNMKSFVTRRENVLKDMAVVIERNMENCEAGLKDLCAELKI
DRFKSPPSVDLESINAKTVEKYNEWVAAARVWCNLILAQKHKIPRGDAFFPTRLKAYPGFPGSVIHKEATSFATAIKELRKCLA
DLKQSHQKLFGALSNEEWSLVLERFPAPEGTGGKGMRHQLGKTLAVLGKKHPEWKPRQYIEELVAGLERGADSMAKHLENRQFS
DRPAAIKLINLLNNACVVLMEPLRVKSLYKEYFEQDTPRRNAFGQARGSLAEPTDQADSLQILGFSLMDEVRGICRYDGFLAFA
```

[0259]

GDKEKDQWAFFFDPDGPGGMGIADPVKHPKGSTDSGYIGFGTMGGSRKRAQAAEKRLVRERLWIRRDRKPLQLPLHFGTRQGRE
YLWNFDHGLKKSSDPWMLTNGRLLRIYPPNRPQFSDYFLTITLQKDTPPLAISSGEKLVGVDRGEQNPASFAVVDAKGRLLERG
NVGEKYTAKIRQFAEMKSDLQRTAGGYTRKLRAKERNTASALGGLVMRELLDLLVRHPGALVFERLSSGITTRGGRGTLMGNMH
YERILSGVEGKLAEAGCYKAPSAAKYRKGITPFLRFVNPAYTSATCSACGEVFSSSWYKKTAATISKESPGSWSVTLPNGKKLQ
LPCSYTSYMRGRGEKTWNTAERLEELFKDRTWSDLPPSTRKTVEGLVARLTHHRPKRGEFVCPSCGHKDDADLQAALNIARKSI
WVSSLPPRQAKQSEGEDRRSTEEPWRVWYKKQIMEKWS (SEQ ID NO: 1)
>3300013087|Ga0163212_1009539_3|M
[水生-淡水]
MIHKSTVKASVRHLTRHALDKPGEGDPGETTKTLKFRIDAKGHTRTRLDQIFSAQRQFVRDLLNKLEGMWDKDPERFIRMARSS
PSEPFEKKTSCAGWMLTKFLTGSDLPKHLSRQSARATLSPLSRNMKSFITRRKNVLKDMAVVIGRNRDNWDSGLKELCAELKVD
QFKGPPSVDPDKLNAKTVEKYNEWVAAARVWCNLILAQKHKIPRGDAFFPTRLKAYPGFPGSVIHKEAVPFSTAIEELRKCLAD
LKRSQLKLFDGLTAEEWVLVLERFPATEGEGGKGMRHQLGKTLAVLGKKHSDWKPRQLVEELVAGLERGADSLVNHLKDRDFSD
RPAAIKLINLLNNACVVLMEPLRVKNLYREYFEQDTPRRNAFGQARGSLAEPTDQADSLEIEGFSLKKGDAGSCNFDGLLAVTP
KDGKDQWSFFFDPEGPGGMGIAEITKHPRGSTTTEHLSFGSSGGSRKKAQASEKKLVRERLWIRRDRKPLQLPLQFGTRQGREY
LWNFDHGLKKSSDPWLLANGRLLRVYPINRPQFSQYNLTITLQKDTPPLALSAGEMLIGVDRGEVNPASFAVVDAKGKLLKCGN
VGEKYTAKIRQLAEMKSDLQRTAGGYTRKLRAKERNTASALGGLVIRELLDLLVRHPGDLVFERLTSGITTRGGRGTLMGNMHY
ERIMSGVEGKLAEAGCYKVPRATKFRSKDTRFVRFVHPAYTSATCSACGDVFSSGWYKKTAATIAKESPGSWSVTLPSGKKIQL
PSSYTYYMRGRGEKTWNTAERLEELFKGRKWSDLPPSTRKTVEGLVARLIHHRPKRGEFVCPSCGHKDDADLQAALNIARKSIW
VRSLPPRQAKQSEGEDRRSTEEPWREWYKKQIREKWA (SEQ ID NO: 2)
>3300013131|Ga0172373_10014249_3|P
[水生-淡水]
MIHKSTVKASVRHLTRHALDKPGEGDPGETTKTLKFRIDAKDHTRTRLDQIFNAQRQFVRDLLNKLEGMWDKDPERFIRMVRSS
ASEPFEKKTSCAGWMLTKFLTGSDLPEHLSRSSATASVATLSGNMKSFITRRENVLKDMAVVIERNRENWDSGLKELCAELKID
QFKAPPSVDPDKLNAKTVEKYNEWVAAARVWCNLILAQKHKIPRGDAFFPTRLKAYPGFPGSVIHKEGIAFDTAIKELRKCLAD
LKRSHLKLFDGLTVEEWGLVLERFPAPEGEGGKGMRHQLGKTLAVLGKKHSDWKPRQLAEELVTGLERGADSLANHLKDRDFSD
RPAAIKLINLLNNACVVLMEPLRVKNLYREYFEQDTPRRNAFGQARGSLAEPTDQADSLQIMGFAIRKTPEGHSGFDGFLAVSQ
KEGKDIWSFFFDPDGPGGMGLCDASKHPRGSTTTDQIGFGTSGGSRKRAQAAEKRLVRERLWIRRDRKPLQLPLQFGTRQGREY
LWNFDHGLKKSSDPWMLTNGRLLRIYPPNRPQFSQYYLTIALQKDTPPPALTAGEMLIGVDRGEQNPASYAVTDGKGKLLAKGN
VGESYSAKIRQFAEMKSEMQRTAGGYTRKLRAKERNTASALGGLVIRELLDLLVRHPGALVFERLTSGITTRGGRGTLMGNMHY
ERILSGVEGKLAEAGCYKAPSAAKYRKGITPFLRFVNPAYTSATCSACGDVFSSGWYRKAAATISRESSGSWSVTLPSGKKVQL
PSSYTSYMRGRGEKTWNTAERLEELFKGRKWSDLPPSTRKTVDGLVARLIHHRPKRGEFVCPSCAHKDDADLQAALNIARKSIW
VSSLPHRQAKQSEGEDRRSTEEPWREWYKKRIREKWA (SEQ ID NO: 3)
>3300013131|Ga0172373_10021309_4|M
[水生-淡水]
MIHKSTVKASVRHLTRHALDKPGEGDPGKPTTKTLKFRIDAKGHTRTRLDQIFTAQNHFRRDLYIKLEGMWDKDPESFISMVNS
SPSKPFQKKTSCSGWIETKFVTGVALPEHLSRETAAATAVAVGNSLTATFERRKNVISQLAKNIEENRSVWRGNLKELCSELKI
TPFKAPPSIHPAKINAKTVEKYNEWVAAARIWCNLILVQKHRVPRNSACLPVRLKHYPGFPGSVIFKKGASLATAIKEFRKCLA
GLKRSQLKLFDGLTAEEWGLVIDRFPAPEGEGGKGMRHQLGKTLAVLGKKHPDWKPRQLVEELVAGLERGADSLVNHLKDRDFS
DRPAAIKLINLLNNACVVLMEPLRVKNLYRELFEVGTPRRDAFGQARGALAEPTDQSEGLEIRGFSIRTNRPPIGHSSFNGFLA
VRQTEGKDEWSFFFDPDGPGGMGLSGLPPAGSIPTDYIGFGSTGGSRKKAQASEKKLVRERLWIRRDRNPLQLPLQFGKRQGRE
YLWNFDHGLKKSSDPWMLANGRLLRVYPPNQPQFSDYFLTITLQKDTPPLALSAGEMLIGVDQGEVNPASYAVTDAKGKLLTRG
NVGESYSAKIRQFAEMKSAEQRTAGGYTRKLRSKERNTASACGGLVIRELLDLLVRYPGALVFERLSSGIATRGGRGTLMSNMH
YQRITGGIEKKLSEAGCHRITDFKSNFSPFLRMVQPHYTSATCSNCGEVFDSDWYEKLTATIRKSVSGSWVVTLPDGKELWLQG
REEIPWNPTDRLEEIFKGRTFPDLTSRSQKIVQGIVGRLVAFRPKRGEFVCPCCGHTDDADLQAALNIARKSIWLSTLPKRSKK
ETLAEARGSTKPAWKKWYQNRISENWC (SEQ ID NO: 4)
>3300013132|Ga0172372_10046622_2|M
[水生-淡水]

MIHKSTVKASVRHLTRHALDKPGEGDPGKPTTKTLKFRIDAKGHTRTRLDQIFTAQNHFRRDLYIKLEGMWDKDPESFISMVNS
SPSKPFQKKTSCSGWIETKFVTGVALPEHLSRETAAATAVAVGNSLTATFERRKNVISQLAKNIEENRSVWRGNLKELCSELKI
TPFKAPPSIHPAKINAKTVEKYNEWVAAARIWCNLILVQKHRVPRNSACLPVRLKHYPGFPGSVIFKKGASLATAIKEFRKCLA
GLKRSQLKLFDGLTAEEWGLVIDRFPAPEGEGGKGMRHQLGKTLAVLGKKHPDWKPRQLVEELVAGLERGADSLVNHLKDRDFS
DRPAAIKLINLLNNACVVLMEPLRVKNLYRELFEVGTPRRDAFGQARGALAEPTDQSEGLEIRGFSIRTNRPPIGHSSFNGFLA
VRQTEGKDEWSFFFDPDGPGGMGLSGLPPAGSIPTDYIGFGSTGGSRKKAQASEKKLVRERLWIRRDRNPLQLPLQFGKRQGRE
YLWNFDHGLKKSSDPWMLANGRLLRVYPPNQPQFSDYFLTITLQKDTPPLALSAGEMLIGVDQGEVNPASYAVTDAKGKLLTRG
NVGESYSAKIRQFAEMKSAEQRTAGGYTRKLRSKERNTASACGGLVIRELLDLLVRYPGALVFERLSSGIATRGGRGTLMSNMH
YQRITGGIEKKLSEAGCHRITDFKSNFSPFLRMVQPHYTSATCSNCGEVFDSDWYEKLTATIRKSVSGSWVVTLPDGKELWLQG
REEIPWNPTDRLEEIFKGRTFPDLTSRSQKIVQGIVGRLVAFRPKRGEFVCPCCGHTDDADLQAALNIARKSIWLSTLPKRSKK
ETLAEARGSTKPAWKKWYQNRISENWC (SEQ ID NO: 4)
>3300014720|Ga0172376_10007396_6|M
[水生-淡水]
MIHKSTVKASVRHLTRHALDKPGEGDPGETTKTLKFRIDAKDHTRTRLDQIFNAQRQFVRDLLNKLEGMWDKDPERFIRMVRSS
ASEPFEKKTSCAGWMLTKFLTGSDLPEHLSRSSATASVATLSGNMKSFITRRENVLKDMAVVIERNRENWDSGLKELCAELKID
QFKAPPSVDPDKLNAKTVEKYNEWVAAARVWCNLILAQKHKIPRGDAFFPTRLKAYPGFPGSVIHKEGIAFDTAIKELRKCLAD
LKRSHLKLFDGLTVEEWGLVLERFPAPEGEGGKGMRHQLGKTLAVLGKKHSDWKPRQLAEELVTGLERGADSLANHLKDRDFSD
RPAAIKLINLLNNACVVLMEPLRVKNLYREYFEQDTPRRNAFGQARGSLAEPTDQADSLQIMGFAIRKTPEGHSGFDGFLAVSQ
KEGKDIWSFFFDPDGPGGMGLCDASKHPRGSTTTDQIGFGTSGGSRKRAQAAEKRLVRERLWIRRDRKPLQLPLQFGTRQGREY
LWNFDHGLKKSSDPWMLTNGRLLRIYPPNRPQFSQYYLTIALQKDTPPPALTAGEMLIGVDRGEQNPASYAVTDGKGKLLAKGN
VGESYSAKIRQFAEMKSEMQRTAGGYTRKLRAKERNTASALGGLVIRELLDLLVRHPGALVFERLTSGITTRGGRGTLMGNMHY
ERILSGVEGKLAEAGCYKAPSAAKYRKGITPFLRFVNPAYTSATCSACGDVFSSGWYRKAAATISRESSGSWSVTLPSGKKVQL
PSSYTSYMRGRGEKTWNTAERLEELFKGRKWSDLPPSTRKTVDGLVARLIHHRPKRGEFVCPSCAHKDDADLQAALNIARKSIW
[0260] VSSLPHRQAKQSEGEDRRSTEEPWREWYKKRIREKWA (SEQ ID NO: 3)
>3300014720|Ga0172376_10013958_4|M
[水生-淡水]
MIHKSTVKASVRHLTRHALDKPGEGDPGKPTTKTLKFRIDAKGHTRTRLDQIFTAQNHFRRDLYIKLEGMWDKDPESFISMVNS
SPSKPFQKKTSCSGWIETKFVTGVALPEHLSRETAAATAVAVGNSLTATFERRKNVISQLAKNIEENRSVWRGNLKELCSELKI
TPFKAPPSIHPAKINAKTVEKYNEWVAAARIWCNLILVQKHRVPRNSACLPVRLKHYPGFPGSVIFKKGASLATAIKEFRKCLA
GLKRSQLKLFDGLTAEEWGLVIDRFPAPEGEGGKGMRHQLGKTLAVLGKKHPDWKPRQLVEELVAGLERGADSLVNHLKDRDFS
DRPAAIKLINLLNNACVVLMEPLRVKNLYRELFEVGTPRRDAFGQARGALAEPTDQSEGLEIRGFSIRTNRPPIGHSSFNGFLA
VRQTEGKDEWSFFFDPDGPGGMGLSGLPPAGSIPTDYIGFGSTGGSRKKAQASEKKLVRERLWIRRDRNPLQLPLQFGKRQGRE
YLWNFDHGLKKSSDPWMLANGRLLRVYPPNQPQFSDYFLTITLQKDTPPLALSAGEMLIGVDQGEVNPASYAVTDAKGKLLTRG
NVGESYSAKIRQFAEMKSAEQRTAGGYTRKLRSKERNTASACGGLVIRELLDLLVRYPGALVFERLSSGIATRGGRGTLMSNMH
YQRITGGIEKKLSEAGCHRITDFKSNFSPFLRMVQPHYTSATCSNCGEVFDSDWYEKLTATIRKSVSGSWVVTLPDGKELWLQG
REEIPWNPTDRLEEIFKGRTFPDLTSRSQKIVQGIVGRLVAFRPKRGEFVCPCCGHTDDADLQAALNIARKSIWLSTLPKRSKK
ETLAEARGSTKPAWKKWYQNRISENWC (SEQ ID NO: 4)
>3300018790|Ga0187842_1006374_2|M
[水生-淡水]
MNHKSTVKASVRHLTRHALQKPGEESDPGETTKTLKFRVDAKGHTRTRLDQIFTAQRQFVRDLLNKLEGMWDKDPQRFIRMVES
SASDPFESKTSCYGWMMTKFLTGSDLPEHLSRSAATASVATLSGNMKSFVTRRENVLKDMAVVIERNMENCEAGLKDLCAELKI
DRFKSPPSVDLESINAKTVEKYNEWVAAARVWCNLILAQKHKIPRGDAFFPTRLKAYPGFPGSVIHKEATSFATAIKELRKCLA
DLKQSHQKLFGALSNEEWSLGLERFPAPEGTGGKGMRHQLGKTLAVLGKKHPEWKPRQYIEELVAGLERGADSMAKHLENRQFS
DRPAAIKLINLLNNACVVLMEPLRVKSLYKEYFEQDTPRRNAFGQARGSLAEPTDQADSLQILGFSLMDEVRGICRYDGFLAFA
GDKEKDQWAFFFDPDGPGGMGIADPVKHPKGSTDSGYIGFGTMGGSRKRAQAAEKRLVRERLWIRRDRKPLQLPLHFGTRQGRE
YLWNFDHGLKKSSDPWMLTNGRLLRIYPPNRPQFSDYFLTITLQKDTPPLAISSGEKLVGVDRGEQNPASFAVVDAKGRLLERG
NVGEKYTAKIRQFAEMKSDLQRTAGGYTRKLRAKERNTASALGGLVMRELLDLLVRHPGALVFERLSSGITTRGGRGTLMGNMH

[0261]

YERILSGVEGKLAEAGCYKAPSAAKYRKGITPFLRFVNPAYTSATCSACGEVFSSSWYKKTAATISKESPGSWSVTLPNGKKLQ
LPCSYTSYMRGRGEKTWNTAERLEELFKDRTWSDLPPSTRKTVEGLVARLTHHRPKRGEFVCPSCGHKDDADLQAALNIARKSI
WVSSLPPRQAKQSEGEDRRSTEEPWRVWYKKQIMEKWS (SEQ ID NO: 5)
>3300018815|Ga0187845_1009203_1|M
[水生-淡水]
MNHKSTVKASVRHLTRHALQKPGEESDPGETTKTLKFRVDAKGHTRTRLDQIFTAQRQFVRDLLNKLEGMWDKDPQRFIRMVES
SASDPFESKTSCYGWMMTKFLTGSDLPEHLSRSAATASVATLSGNMKSFVTRRENVLKDMAVVIERNMENCEAGLKDLCAELKI
DRFKSPPSVDLESINAKTVEKYNEWVAAARVWCNLILAQKHKIPRGDAFFPTRLKAYPGFPGSVIHKEATSFATAIKELRKCLA
DLKQSHQKLFGALSNEEWGLVLERFPAPEGTGGKGMRHQLGKTLAVLGKKHPEWKPRQYIEELVAGLERGADSMAKHLENRQFS
DRPAAIKLINLLNNACVVLMEPLRVKSLYKEYFEQDTPRRNAFGQARGSLAEPTDQADSLQILGFSLMDEVRGICRYDGFLAFA
GDKEKDQWAFFFDPDGPGGMGIADPVKHPKGSTDSGYIGFGTMGGSRKRAQAAEKRLVRERLWIRRDRKPLQLPLHFGTRQGRE
YLWNFDHGLKKSSDPWMLTNGRLPRIYPPNRPQFSDYFLTITLQKDTPPLAISSGEKLVGVDRGEQNPASFAVVDAKGRLLERG
NVGEKYTAKIRQFAEMKSDLQRTAGGYTRKLRAKERNTASALGGLVMRELLDLLVRHPGALVFERLSSGITTRGGRGTLMGNMH
YERILSGVEGKLAEAGCYKAPSAAKYRKGITPFLRFVNPAYTSATCSACGEVFSSSWYKKTAATISKESPGSWSVTLPNGKKLQ
LPCSYTSYMRGRGEKTWNTAERLEELFKDRTWSDLPPSTRKTVEGLVARLTHHRPKRGEFVCPSCGHKDDADLQAALNIARKSI
WVSSLPPRQAKQSEGEDRRSTEEPWRVWYKKQIMEKWS (SEQ ID NO: 6)
>3300018868|Ga0187844_10013314_2|M
[水生-淡水]
MNHKSTVKASVRHLTRHALQKPGEESDPGETTKTLKFRVDAKGHTRTRLDQIFTAQRQFVRDLLNKLEGMWDKDPQRFIRMVES
SASDPFESKTSCYGWMMTKFLTGSDLPEHLSRSAATASVATLSGNMKSFVTRRENVLKDMAVVIERNMENCEAGLKDLCAESKI
DRFKSPPSVDLESINAKTVEKYNEWVAAARVWCNLILAQKHKIPRGDAFFPTRLKAYPGFPGSVIHKEATSFATAIKELRKCLA
DLKQSHQKLFGALSNEEWSLVLERFPAPEGTGGKGMRHQLGKTLAVLGKKHPEWKPRQYIEELVAGLERGADSMAKHLENRQFS
DRPAAIKLINLLNNACVVLMEPLRVKSLYKEYFEQDTPRRNAFGQARGSLAEPTDQADSLQILGFSLMDEVRGICRYDGFLAFA
GDKEKDQWAFFFDPDGPGGMGIADPVKHPKGSTDSGYIGFGTMGGSRKRAQAAEKRLVRERLWIRRDRKPLQLPLHFGTRQGRE
YLWNFDHGLKKSSDPWMLTNGRLLRIYPPNRPQFSDYFLTITLQKDTPPLAISSGEKLVGVDRGEQNPASFAVVDAKGRLLERG
NVGEKYTAKIRQFAEMKSDLQRTAGGYTRKLRAKERNTASALGGLVMRELLDLLVRHPGALVFERLSSGITTRGGRGTLMGNMH
YERILSGVEGKLAEAGCYKAPSAAKYRKGITPFLRFVNPAYTSATCSACGEVFSSSWYKKTAATISKESPGSWSVTLPNGKKLQ
LPCSYTSYMRGRGEKTWNTAERLEELFKDRTWSDLPPSTRKTVEGLVARLTHHRPKRGEFVCPSCGHKDDADLQAALNIARKSI
WVSSLPPRQAKQSEGEDRRSTEEPWRVWYKKQIMEKWS (SEQ ID NO: 7)
>3300019093|Ga0187843_10013171_2|M
[水生-淡水]
MNHKSTVKASVRHLTRHALQKPGEESDPGETTKTLKFRVDAKGHTRTRLDQIFTAQRQFVRDLLNKLEGMWDKDPQRFIRMVES
SASDPFESKTSCYGWMMTKFLTGSDLPEHLSRSAATASVATLSGNMKSFVTRRENVLKDMAVVIERNMENCEAGLKDLCAELKI
DRFKSPPSVDLESINAKTVEKYNEWVAAARVWCNLILAQKHKIPRGDAFFPTRLKAYPGFPGSVIHKEATYFATAIKELRKCLA
DLKQSHQKLFGALSNEEWSLVLERFPAPEGTGGKGMRHQLGKTLAVLGKKHPEWKPRQYIEELVAGLERGADSMAKHLENRQFS
DRPAAIKLINLLNNACVVLMEPLRVKSLYKEYFEQDTPRRNAFGQARGSLAEPTDQADSLQILGFSLMDEVRGICRYDGFLAFA
GDKEKDQWAFFFDPDGPGGMGIADPVKHPKGSTDSGYIGFGTMGGSRKRAQAAEKRLVRERLWIRRDRKPLQLPLHFGTRQGRE
YLWNFDHGLKKSSDPWMLTNGRLLRIYPHNRPQFSDYFLTITLQKDTPPLAISSGEKLVGVDRGEQNPASFAVVDAKGRLLERG
NVGEKYTAKIRQFAEMKSDLQRTAGGYTRKLRAKERNTASALGGLVMRELLDLLVRHPGALVFERLSSGITTRGGRGTLMGNMH
YERILSGLRASWPRQVATRHPVPPSTAKESRRSCDL (SEQ ID NO: 8)
>3300023179|Ga0214923_10008431_2|P
[水生-淡水]
MLNPAEMQMPSNSIKSIKASVRELYNHALNKQGDGEPGETTKTLKFKIDPSPENQIQLTKIFDAQRAFLRDVLQKLEGMWDKHP
EKIEEMANCSGSKPFEKKTSCYGWLLTKFLTGSDLPENLSRSAATASIGTLAGNLKSFLTRRKNVLKEMANTVSANAKLWDEDL
PIKCKEMDLDLFKAPPKIKVESLNAKTVDEYNEWVASARVWCNLILVQKNKVPRSDACFPKRLKAYPGFPFSVKHKETAKITDA
VKSLQDCLCELKKAHLKAFDSLTEVEWQSVLERFPSPKEVDGKGTRRLIGKTLAALHKKNPNWKPVDIVKELIAGMERSIQSLV
KHLEVREFTDRKAAIKLMNHMNHACVFLLEPLRVRNDYKGLYEKDTPRRNAFGEARGSLSEPTDEADSLQIIGFSLNERGGTRY

DGILAYSGEPGSQKWAFFFDPLGPEGMGLADEGKNPARTTLLPWVGFGTTGGARKKAQAKAKALVRGKIWVSPKRKPMALPLLF
GARQGREYLWNFDRGLKSSSTPWMLTNGRILRIFPYQKPELAEFYLTLTLQKDNPPLALSLGEHLIGVDRGEENPASYALVDAK
GRLVERGHIGQEYSQRIKTFAKEKAELQRSLGGYTRRLRAKERNIASALGGQVIRELMNLLVAKKGALVFENLSSGIVTRGGRG
TLMGNMHYERILSGIEQKLAEAGCYEVPNAKKYRKGVSPFVRFVNPSYTSSTCSKCGEVYSRAWYEKIAPSLERKGAGIWAVKL
GEKHFTLPLSYSYFLRGRGEQHKKTDERLTEILKGKDWAKLSATSKKEIVGLLQSKLVAYRPKRGKFNCPSCEHEDDADLQAAL
NIARKTIWVSTLPPKKAKSGDQDFRGVGQELWAKWYKDRVHKNWA (SEQ ID NO: 9)
>3300027730|Ga0247833_1032983_1|M
[水生-淡水]
MNHKNTVKASVRELTRHALQKPGEESEPGETTKTLKFRVDAKGHTRTRLDQIFTAQRQFVRDLLTKLEGMWDKDPQRFIRMVES
SASDPFESKTSCYGWMMTKFLTGSDLPEHLSRSAATASVATLSGNMKSFVTRRENVLKDMAVVIERNMENCEAGLKDLCAELKI
DRFKSPPSVDLESINAKTVEKYNEWVAAARVWCNLILAQKHKIPRGDAFFPTRLKAYPGFPGSVIHKEATSFATAIKELRKCLA
DLKQSHQKLFDALSTEEWSLVLEKFPAPEGTGGKGMRHQLGKTLAVLGKKHPEWKPRQYIEELVAGLERGADSMAKHLKNRQFS
DRPAAIKLINLLNNACVVLMEPLRVKSLYKEYFEQDTPRRNAFGQARGSLAEPTDQADSLQILGFSLMDEVRGVCRYDGFLAFT
GDKGKDQWAFFFDPDGPGGMGIADPVKHPKGSSDSGYIGFGTMGGSRKRAQAAEKRLVRERLWIRRDRKPLQLPLHFGTRQGRE
YLWNFDHGLKKSSDPWMLTNGRLLRVYPPNRPQLSDYFLTITLQKDTPPLAISSGDKLVGVDRGEQNPASFAVVDDKGRLLERG
NVGEKYTAKIRQFAEMKSDLQRTAGGYTRKLRAKERNTASALGGLVMRELLDLLVRHPGALVFERLSSGITTRGGRGTLMGNMH
YERILSGVEGKLAEAGCYKAPSAAKYRKGITPFLRFVNPAYTSATCSACGEVFSSGWYKKTAATISKESLGSWSVTLPNGKKLQ
LPSSYTSYMRGRGEKTWNTAERLEELFQDRTWSDLPPSTRKTVEGLVARLTHHRPKRGEFVCSSCGHKDDADLQAALNIARKSI
WVSSLPPRQAKQSEGEDRRSTEEPWRVWYKKQIMGKWS (SEQ ID NO: 10)
>3300028114|Ga0247835_1021903_1|P
[水生-淡水]
MNHKNTVKASVRELTRHALQKPGEESEPGETTKTLKFRVDAKGHTRTRLDQIFTAQRQFVRDLLTKLEGMWDKDPQRFIRMVES
SASDPFESKTSCYGWMMTKFLTGSDLPEHLSRSAATASVATLSGNMKSFVTRRENVLKDMAVVIERNMESFEAGLKDLCAELKI
[0262] DRFKSPPSVDLESINAKTVEKYNEWVAAARVWCNLILAQKHKIPRGDAFFPTRLKAYPGFPGSVIHKEATSFATAIKELRKCLA
DLKQSHQKLFDALSTEEWSLVLEKFPAPEGTGGKGMRHQLGKTLAVLGKKHPEWKPRQYIEELVAGLERGADSMAKHLKNRQFS
DRPAAIKLINLLNNACVVLMEPLRVKSLYKEYFEQDTPRRNAFGQARGSLAEPTDQADSLQILGFSLMDEVRGVCRYDGFLAFT
GDKGKDQWAFFFDPDGPGGMGIADPVKHPKGSSDSGYIGFGTMGGSRKRAQAAEKRLVRERLWIRRDRKPLQLPLHFGTRQGRE
YLWNFDHGLKKSSDPWMLTNGRLLRVYPPNRPQLSDYFLTITLQKDTPPLAISSGDKLVGVDRGEQNPASFAVVDDKGRLLERG
NVGEKYTAKIRQFAEMKSDLQRTAGGYTRKLRAKERNTASALGGLVMRELLDLLVRHPGALVFERLSSGITTRGGRGTLMGNMH
YERILSGVEGKLAEAGCYKAPSAAKYRKGITPFLRFVNPAYTSATCSACGEVFSSGWYKKTAATISKESLGSWSVTLPNGKKLQ
LPSSYTSYMRGRGEKTWNTAERLEELFQDRTWSDLPPSTRKTVEGLVARLTHHRPKRGEFVCSSCGHKDDADLQAALNIARKSI
WVSSLPPRQAKQSEGEDRRSTEEPWRVWYKKQIMGKWS (SEQ ID NO: 11)
>3300028553|Ga0247839_1034275_1|M
[水生-淡水]
MNHKNTVKASVRELTRHALQKPGEESEPGETTKTLKFRVDAKGHTRTRLDQIFTAQRQFVRDLLTKLEGMWDKDPQRFIRMVES
SASDPFESKTSCYGWMMTKFLTGSDLPEHLSRSAATASVATLSGNMKSFVTRRENVLKDMAVVIERNMENCEAGLKDLCAELKI
DRFKSPPSVDLESINAKTVEKYNEWVAAARVWCNLILAQKHKIPRGDAFFPTRLKAYPGFPGSVIHKEATSFATAIKELRKCLA
DLKQSHQKLFDALSTEEWSLVLEKFPAPEGTGGKGMRHQLGKTLAVLGKKHPEWKPRQYIEELVAGLERGADSMAKHLKNRQFS
DRPAAIKLINLLNNACVVLMEPLRVKSLYKEYFEQDTPRRNAFGQARGSLAEPTDQADSLQILGFSLMDEVRGVCRYDGFLAFT
GDKGKDQWAFFFDPDGPGGMGIADPVKHPKGSSDSGYIGFGTMGGSRKRAQAAEKRLVRERLWIRRDRKPLQLPLHFGTRQGRE
YLWNFDHGLKKSSDPWMLTNGRLLRVYPPNRPQLSDYFLTITLQKDTPPLAISSGDKLVGVDRGEQNPASFAVVDDKGRLLERG
NVGEKYTAKIRQFAEMKSDLQRTAGGYTRKLRAKERNTASALGGLVMRELLDLLVRHPGALVFERLSSGITTRGGRGTLMGNMH
YERILSGVEGKLAEAGCYKAPSAAKYRKGITPFLRFVNPAYTSATCSACGEVFSSGWYKKTAATISKESLGSWSVTLPNGKKLQ
LPSSYTSYMRGRGEKTWNTAERLEELFQDRTWSDLPPSTRKTVEGLVARLTHHRPKRGEFVCSSCGHKDDADLQAALNIARKSI
WVSSLPPRQAKQSEGEDRRSTEEPWRVWYKKQIMGKWS (SEQ ID NO: 10)
>3300028569|Ga0247843_1031164_2|P
[水生-淡水]

[0263]

MNHKNTVKASVRELTRHALQKPGEESEPGETTKTLKFRVDAKGHTRTRLDQIFTAQRQFVRDLLTKLEGMWDKDPQRFIRMVES
SASDPFESKTSCYGWMMTKFLTGSDLPEHLSRSAATASVATLSGNMKSFVTRRENVLKDMAVVIERNMESFEAGLKDLCAELKI
DRFKSPPSVDLESINAKTVEKYNEWVAAARVWCNLILAQKHKIPRGDAFFPTRLKAYPGFPGSVIHKEATSFATAIKELRKCLA
DLKQSHQKLFDALSTEEWSLVLEKFPAPEGTGGKGMRHQLGKTLAVLGKKHPEWKPRQYIEELVAGLERGADSMAKHLKNRQFS
DRPAAIKLINLLNNACVVLMEPLRVKSLYKEYFEQDTPRRNAFGQARGSLAEPTDQADSLQILGFSLMDEVRGVCRYDGFLAFT
GDKGKDQWAFFFDPDGPGGMGIADPVKHPKGSSDSGYIGFGTMGGSRKRAQAAEKRLVRERLWIRRDRKPLQLPLHFGTRQGRE
YLWNFDHGLKKSSDPWMLTNGRLLRVYPPNRPQLSDYFLTITLQKDTPPLAISSGDKLVGVDRGEQNPASFAVVDDKGRLLERG
NVGEKYTAKIRQFAEMKSDLQRTAGGYTRKLRAKERNTASALGGLVMRELLDLLVRHPGALVFERLSSGITTRGGRGTLMGNMH
YERILSGVEGKLAEAGCYKAPSAAKYRKGITPFLRFVNPAYTSATCSACGEVFSSGWYKKTAATISKESLGSWSVTLPNGKKLQ
LPSSYTSYMRGRGEKTWNTAERLEELFQDRTWSDLPPSTRKTVEGLVARLTHHRPKRGEFVCSSCGHKDDADLQAALNIARKSI
WVSSLPPRQAKQSEGEDRRSTEEPWRVWYKKQIMGKWS (SEQ ID NO: 11)

>3300014164|Ga0181532_10001487_14|M

[水生-淡水-沼泽]

MPQTIELKLFMNSIHKSVRSLLSLGLSRSASASPQTITRTVKFKINTDIRPELIPVLNRHFDAFEKFRRKVLDELEAWWNKKPD
DFTSMVRATAKERFQDKSSCYGWLYRHFLSGKDLPAGLTREAANSVLWSLKGGLKSFLTRRENVAKELEENYTTNANLWNESII
RIIRDHKLDKPTPPTAPPTVNFKKLTDDQIEKYNGWVGIVRMWGNLILHQQLVQKGLVDREDAKLPRRLKGYPGFPGSQRYLQG
ASLADSLKRLDDITGKFAKDSEKRFAKITSEHWGAVLERFFPPEAKEDETRRRPRTVKQMAAERLHYLCKHNPDWKPEQTAKEV
LAGLFRLADKLHAHLKNNSPHDRRAVIKLANVLNVAAVFSLEPLRASGDYIACFEADEPRRQAFADLRGALAQPGDDTVAIEVT
GFSLSSNSPQYCGLLACREIKPDCDEWAFFYTLDNQDLKIFAKPGPRPPRDFQPTNWHGFATKGSDDEKEIVRGTVWLPTEISK
TKPALALPLAFGARQGREYLWHFDRDLRTRNEWTPLNARLLRVMPAAQESESEKERALRANRADFYVTITFLRLAPPIASVKSK
ELIGVDRGEVVPAAFAVVGATGRFLAKPSRLREFMEYLRQLDNWEKLNLAHPENARENERKPRWEARPDDGFGFIAPEYRDQQR
EFNKLKRELQQTKGGYTRWLRAKERNRARALGGQVMRELLMLCATHKAPLVFESLGSGLATRGGKGTMMSQMQYSRVFTSVEQK
LAEAGLCAIARQPKFRKWDNGFIKLVGPACTSATCSVCGQVHSTSFYATIAATLKESSKGNWQVTLPNGKVRRLPKEYGYWLRG
KGEQTANTDERIHQLLNKPFAELSTTNHNALISLLKSRWLPFRPDQATFQCLACGYSVNADVQAALNIARKHLFRIGREKKSDE
ENEAARRKVMGEWETWYRQKLASVWNK (SEQ ID NO: 12)

>3300014164|Ga0181532_10001487_14|P

[水生-淡水-沼泽]

MGLGRTFCLGHNLARSTKRREKEDECGNFRGWQNHVGFRLGHRLRIARRMPQTIELKLFMNSIHKSVRSLLSLGLSRSASASPQ
TITRTVKFKINTDIRPELIPVLNRHFDAFEKFRRKVLDELEAWWNKKPDDFTSMVRATAKERFQDKSSCYGWLYRHFLSGKDLP
AGLTREAANSVLWSLKGGLKSFLTRRENVAKELEENYTTNANLWNESIIRIIRDHKLDKPTPPTAPPTVNFKKLTDDQIEKYNG
WVGIVRMWGNLILHQQLVQKGLVDREDAKLPRRLKGYPGFPGSQRYLQGASLADSLKRLDDITGKFAKDSEKRFAKITSEHWGA
VLERFFPPEAKEDETRRRPRTVKQMAAERLHYLCKHNPDWKPEQTAKEVLAGLFRLADKLHAHLKNNSPHDRRAVIKLANVLNV
AAVFSLEPLRASGDYIACFEADEPRRQAFADLRGALAQPGDDTVAIEVTGFSLSSNSPQYCGLLACREIKPDCDEWAFFYTLDN
QDLKIFAKPGPRPPRDFQPTNWHGFATKGSDDEKEIVRGTVWLPTEISKTKPALALPLAFGARQGREYLWHFDRDLRTRNEWTP
LNARLLRVMPAAQESESEKERALRANRADFYVTITFLRLAPPIASVKSKELIGVDRGEVVPAAFAVVGATGRFLAKPSRLREFM
EYLRQLDNWEKLNLAHPENARENERKPRWEARPDDGFGFIAPEYRDQQREFNKLKRELQQTKGGYTRWLRAKERNRARALGGQV
MRELLMLCATHKAPLVFESLGSGLATRGGKGTMMSQMQYSRVFTSVEQKLAEAGLCAIARQPKFRKWDNGFIKLVGPACTSATC
SVCGQVHSTSFYATIAATLKESSKGNWQVTLPNGKVRRLPKEYGYWLRGKGEQTANTDERIHQLLNKPFAELSTTNHNALISLL
KSRWLPFRPDQATFQCLACGYSVNADVQAALNIARKHLFRIGREKKSDEENEAARRKVMGEWETWYRQKLASVWNK (SEQ ID
NO: 13)

>3300014165|Ga0181523_10000360_35|M

[水生-淡水-沼泽]

MPQTIELKLFMNSIHKSVRSLLSLGLSRSASASPQTITRTVKFKINTDIRPELIPVLNRHFDAFEKFRRKVLDELEAWWNKKPD
DFTSMVRATAKERFQDKSSCYGWLYRHFLSGKDLPAGLTREAANSVLWSLKGGLKSFLTRRENVAKELEENYTTNANLWNESII
RIIRDHKLDKPTPPTAPPTVNFKKLTDDQIEKYNGWVGIVRMWGNLILHQQLVQKGLVDREDAKLPRRLKGYPGFPGSQRYLQG
ASLADSLKRLDDITGKFAKDSEKRFAKITSEHWGAVLERFFPPEAKEDETRRRPRTVKQMAAERLHYLCKHNPDWKPEQTAKEV
LAGLFRLADKLHAHLKNNSPHDRRAVIKLANVLNVAAVFSLEPLRASGDYIACFEADEPRRQAFADLRGALAQPGDDTVAIEVT

GFSLSSNSPQYCGLLACREIKPDCDEWAFFYTLDNQDLKIFAKPGPRPPRDFQPTNWHGFATKGSDDEKEIVRGTVWLPTEISK
TKPALALPLAFGARQGREYLWHFDRDLRTRNEWTPLNARLLRVMPAAQESESEKERALRANRADFYVTITFLRLAPPIASVKSK
ELIGVDRGEVVPAAFAVVGATGRFLAKPSRLREFMEYLRQLDNWEKLNLAHPENARENERKPRWEARPDDGFGFIAPEYRDQQR
EFNKLKRELQQTKGGYTRWLRAKERNRARALGGQVMRELLMLCATHKAPLVFESLGSGLATRGGKGTMMSQMQYSRVFTSVEQK
LAEAGLCAIARQPKFRKWDNGFIKLVGPACTSATCSVCGQVHSTSFYATIAATLKESSKGNWQVTLPNGKVRRLPKEYGYWLRG
KGEQTANTDERIHQLLNKPFAELSTTNHNALISLLKSRWLPFRPDQATFQCLACGYSVNADVQAALNIARKHLFRIGREKKSDE
ENEAARRKVMGEWETWYRQKLASVWNK (SEQ ID NO: 12)

>3300014165|Ga0181523_10000360_33|P

[水生-淡水-沼泽]

MGLGRTFCLGHNLARSTKRREKEDECGNFRGWQNHVGFRLGHRLRIARRMPQTIELKLFMNSIHKSVRSLLSLGLSRSASASPQ
TITRTVKFKINTDIRPELIPVLNRHFDAFEKFRRKVLDELEAWWNKKPDDFTSMVRATAKERFQDKSSCYGWLYRHFLSGKDLP
AGLTREAANSVLWSLKGGLKSFLTRRENVAKELEENYTTNANLWNESIIRIIRDHKLDKPTPPTAPPTVNFKKLTDDQIEKYNG
WVGIVRMWGNLILHQQLVQKGLVDREDAKLPRRLKGYPGFPGSQRYLQGASLADSLKRLDDITGKFAKDSEKRFAKITSEHWGA
VLERFFPPEAKEDETRRRPRTVKQMAAERLHYLCKHNPDWKPEQTAKEVLAGLFRLADKLHAHLKNNSPHDRRAVIKLANVLNV
AAVFSLEPLRASGDYIACFEADEPRRQAFADLRGALAQPGDDTVAIEVTGFSLSSNSPQYCGLLACREIKPDCDEWAFFYTLDN
QDLKIFAKPGPRPPRDFQPTNWHGFATKGSDDEKEIVRGTVWLPTEISKTKPALALPLAFGARQGREYLWHFDRDLRTRNEWTP
LNARLLRVMPAAQESESEKERALRANRADFYVTITFLRLAPPIASVKSKELIGVDRGEVVPAAFAVVGATGRFLAKPSRLREFM
EYLRQLDNWEKLNLAHPENARENERKPRWEARPDDGFGFIAPEYRDQQREFNKLKRELQQTKGGYTRWLRAKERNRARALGGQV
MRELLMLCATHKAPLVFESLGSGLATRGGKGTMMSQMQYSRVFTSVEQKLAEAGLCAIARQPKFRKWDNGFIKLVGPACTSATC
SVCGQVHSTSFYATIAATLKESSKGNWQVTLPNGKVRRLPKEYGYWLRGKGEQTANTDERIHQLLNKPFAELSTTNHNALISLL
KSRWLPFRPDQATFQCLACGYSVNADVQAALNIARKHLFRIGREKKSDEENEAARRKVMGEWETWYRQKLASVWNK (SEQ ID
NO: 13)

>3300014200|Ga0181526_10000193_31|M

[0264]

[水生-淡水-沼泽]

MPQTIELKLFMNSIHKSVRSLLSLGLSRSASASPQTITRTVKFKINTDIRPELIPVLNRHFDAFEKFRRKVLDELEAWWNKKPD
DFTSMVRATAKERFQDKSSCYGWLYRHFLSGKDLPAGLTREAANSVLWSLKGGLKSFLTRRENVAKELEENYTTNANLWNESII
RIIRDHKLDKPTPPTAPPTVNFKKLTDDQIEKYNGWVGIVRMWGNLILHQQLVQKGLVDREDAKLPRRLKGYPGFPGSQRYLQG
ASLADSLKRLDDITGKFAKDSEKRFAKITSEHWGAVLERFFPPEAKEDETRRRPRTVKQMAAERLHYLCKHNPDWKPEQTAKEV
LAGLFRLADKLHAHLKNNSPHDRRAVIKLANVLNVAAVFSLEPLRASGDYIACFEADEPRRQAFADLRGALAQPGDDTVAIEVT
GFSLSSNSPQYCGLLACREIKPDCDEWAFFYTLDNQDLKIFAKPGPRPPRDFQPTNWHGFATKGSDDEKEIVRGTVWLPTEISK
TKPALALPLAFGARQGREYLWHFDRDLRTRNEWTPLNARLLRVMPAAQESESEKERALRANRADFYVTITFLRLAPPIASVKSK
ELIGVDRGEVVPAAFAVVGATGRFLAKPSRLREFMEYLRQLDNWEKLNLAHPENARENERKPRWEARPDDGFGFIAPEYRDQQR
EFNKLKRELQQTKGGYTRWLRAKERNRARALGGQVMRELLMLCATHKAPLVFESLGSGLATRGGKGTMMSQMQYSRVFTSVEQK
LAEAGLCAIARQPKFRKWDNGFIKLVGPACTSATCSVCGQVHSTSFYATIAATLKESSKGNWQVTLPNGKVRRLPKEYGYWLRG
KGEQTANTDERIHQLLNKPFAELSTTNHNALISLLKSRWLPFRPDQATFQCLACGYSVNADVQAALNIARKHLFRIGREKKSDE
ENEAARRKVMGEWETWYRQKLASVWNK (SEQ ID NO: 12)

>3300014200|Ga0181526_10000193_30|P

[水生-淡水-沼泽]

MGLGRTFCLGHNLARSTKRREKEDECGNFRGWQNHVGFRLGHRLRIARRMPQTIELKLFMNSIHKSVRSLLSLGLSRSASASPQ
TITRTVKFKINTDIRPELIPVLNRHFDAFEKFRRKVLDELEAWWNKKPDDFTSMVRATAKERFQDKSSCYGWLYRHFLSGKDLP
AGLTREAANSVLWSLKGGLKSFLTRRENVAKELEENYTTNANLWNESIIRIIRDHKLDKPTPPTAPPTVNFKKLTDDQIEKYNG
WVGIVRMWGNLILHQQLVQKGLVDREDAKLPRRLKGYPGFPGSQRYLQGASLADSLKRLDDITGKFAKDSEKRFAKITSEHWGA
VLERFFPPEAKEDETRRRPRTVKQMAAERLHYLCKHNPDWKPEQTAKEVLAGLFRLADKLHAHLKNNSPHDRRAVIKLANVLNV
AAVFSLEPLRASGDYIACFEADEPRRQAFADLRGALAQPGDDTVAIEVTGFSLSSNSPQYCGLLACREIKPDCDEWAFFYTLDN
QDLKIFAKPGPRPPRDFQPTNWHGFATKGSDDEKEIVRGTVWLPTEISKTKPALALPLAFGARQGREYLWHFDRDLRTRNEWTP
LNARLLRVMPAAQESESEKERALRANRADFYVTITFLRLAPPIASVKSKELIGVDRGEVVPAAFAVVGATGRFLAKPSRLREFM
EYLRQLDNWEKLNLAHPENARENERKPRWEARPDDGFGFIAPEYRDQQREFNKLKRELQQTKGGYTRWLRAKERNRARALGGQV

[0265]

MRELLMLCATHKAPLVFESLGSGLATRGGKGTMMSQMQYSRVFTSVEQKLAEAGLCAIARQPKFRKWDNGFIKLVGPACTSATC
SVCGQVHSTSFYATIAATLKESSKGNWQVTLPNGKVRRLPKEYGYWLRGKGEQTANTDERIHQLLNKPFAELSTTNHNALISLL
KSRWLPFRPDQATFQCLACGYSVNADVQAALNIARKHLFRIGREKKSDEENEAARRKVMGEWETWYRQKLASVWNK (SEQ ID
NO: 13)
>3300027720|Ga0209617_10001817_3|M
[水生-淡水-淡水和沉积物]
MNHKSTVKASVRHLTRHALQKPGEESDPGETTKTLKFRVDAKAHTRTRLDQIFTAQRQFVRDLLNKLEGMWDKDPQRFIRMVES
SASDPFESKTSCYGWMMTKFLTGSDLPEHLSRSAATASVATLSGNMKSFVTRRENVLKDMAVVIERNMENCEAGLKDLCAELKI
DRFKSPPSVDLESINAKTVEKYNEWVAAARVWCNLILAQKHKIPRGDAFFPTRLKAYPGFPGSVIHKEATSFATAIKELRKCLA
DLKQSHQKLFDALSNEEWSLVLERFPAPEGTGGKGMRHQLGKTLAVLGKKHPEWKPRQFIEELVAGLERGADSMAKHLENRQFS
DRPAAIKLINLLNNACVVLMEPLRVKSLYKEYFEQDTPRRNAFGQARGSLAEPTDQADSLQILGFSLMDEVRGICRYDGFLAFA
GDKGKDQWAFFFDPDGPGGMGIADPVKHPKGSSDSGYIGFGTMGGSRKRAQAAEKRLVRERLWIRRDRKPLQLPLHFGTRQGRE
YLWNFDHGLKKSSDPWMLTNGRLLRIYPPNRPQFSDYFLTITLQKDTPPLAISSGEKLVGVDRGEQNPASFAVVDAKGRLLERG
NVGEKYTAKIRQFAEMKSDLQRTAGGYTRKLRAKERNTASALGGLVMRELLDLLVRHPGALVFERLSSGITTRGGRGTLMGNMH
YERILSGVEGKLAEAGCYKAPSAAKYRKGITPFLRFVNPAYTSATCSACGEVFSSSWYKKTAATISKESPGSWSVTLPNGKKLQ
LPCSYTSYMRGRGEKTWNTAERLEELFKDRTWSDLPPSTRKTVEGLVARLTHHRPKRGEFVCPSCGHKDDADLQAALNIARKSI
WVSSLPPRQAKQSEGEDRRSTEEPWRVWYKKQIMEKWS (SEQ ID NO: 14)
>3300009764|Ga0116134_1001531_13|P
[水生-淡水-泥炭地]
MKAKYRSNSIHASLRQLLELGLSKSSSAEPQKITRTAKFKINTDIRPDLVPVLNRHFDAFEKFRRKVLDELEARWNKDQKSFQA
MVQCSAKEPYQKKSSCYAWLDTHFITEAKAVLDLPRKPATSLLYNLSGGLKGFLTRRETVTEDIQKRFNDNLREWNGDLSQLAS
DLKAPLPPAPPNLDFENLTERAIERYNDWVGRTRAWCNLILVQQKKVERRDACLPRYLKGYPGFFGSQRYATTASLAENLKKLE
QEAREQSKKAPTRFAKLSPEIWTAIQERFSPPEGHEAGEKRRPRTAHQTVCLRFAALRAAHPEWTPAQLAGEILAGIFHGTEKL
KKHLAANGFEDRRAVIKLANLYNVVVAFSLDPIRAAGNYISFYKEETPKRNAFGDVRGGLHQPSDESAAIEIMGFGLQKESGKP
LYNGLLVCKKSEKEHDDSWAFLYCHTEGQMFELANEKAKLRGKLLTDWTGFASRGGSRKKAEASAKQLVRGRVWISEKTPPTVM
PLAFGSRQGREYLWHFDRDLREKNEWVLGNGRLLRIMPPGRPNAADFYLTITLERQAPPVAEFIASKLIGIDRGEAVPAAYALI
DREGRWLVNQKRFQEFLKALGTWHDELDEWWKKNKNTPINDRVKRPRLQWTSELEDGFGFVASEYREQQADFNDQKRELQRTQG
GYTRWLRSKERNRARALGGEVTRAVLSLASEHRSPLVLEKLGSALATRGGKGTMMSQMQYERMLTTLEQKLAEVGLYAMPSAPK
YRKLTNGFINFAAPHYTSSTCSSCGQVHDSIFYNELAKKIVRPNDSKWQVTLPSGQTRTLPQTYSYWLKGKGEQTKQTNERLNE
LLGERTVTQLSKSSYMTLVSLLKNSWLPYRPQQADFCCLNCGYTTNADVQGALNIARKFLFRSERGKKTDDAGEEGEAKRGKYK
DDWQTWYRQKLEKVWCK (SEQ ID NO: 15)
>3300018009|Ga0187884_10009678_2|M
[水生-淡水-泥炭地]
MKAKYRSNSIHASLRQLLELGLSKSSSAEPQKITRTAKFKINTDIRPDLVPVLNRHFDAFEKFRRKVLDELEARWNKDQKSFQA
MVQCSAKEPYQKKSSCYAWLDTHFITEAKAVLDLPRKPATSLLYNLSGGLKGFLTRRETVTEDIQKRFNDNLREWNGDLSQLAS
DLKAPLPPAPPNLDFENLTERAIERYNDWVGRTRAWCNLILVQQKKVERRDACLPRYLKGYPGFFGSQRYATTASLAENLKKLE
QEAREQSKKAPTRFAKLSPEIWTAIQERFSPPEGHEAGEKRRPRTAHQTVCLRFAALRAAHPEWTPAQLAGEILAGIFHGTEKL
KKHLAANGFEDRRAVIKLANLYNVVVAFSLDPIRAAGNYISFYKEETPKRNAFGDVRGGLHQPSDESAAIEIMGFGLQKESGKP
LYNGLLVCKKSEKEHDDSWAFLYCHTEGQMFELANEKAKLRGKLLTDWTGFASRGGSRKKAEASAKQLVRGRVWISEKTPPTVM
PLAFGSRQGREYLWHFDRDLREKNEWVLGNGRLLRIMPPGRPNAADFYLTITLERQAPPVAEFIASKLIGIDRGEAVPAAYALI
DREGRWLVNQKRFQEFLKALGTWHDELDEWWKKNKNTPINDRVKRPRLQWTSELEDGFGFVASEYREQQADFNDQKRELQRTQG
GYTRWLRSKERNRARALGGEVTRAVLSLASEHRSPLVLEKLGSALATRGGKGTMMSQMQYERMLTTLEQKLAEVGLYAMPSAPK
YRKLTNGFINFAAPHYTSSTCSSCGQVHDSIFYNELAKKIVRPNDSKWQVTLPSGQTRTLPQTYSYWLKGKGEQTKQTNERLNE
LLGERTVTQLSKSSYMTLVSLLKNSWLPYRPQQADFCCLNCGYTTNADVQGALNIARKFLFRSERGKKTDDAGEEGEAKRGKYK
DDWQTWYRQKLEKVWCK (SEQ ID NO: 15)
>3300018014|Ga0187860_1004616_12|M
[水生-淡水-泥炭地]

[0266]

MKAKYRSNSIHASLRQLLELGLSKSSSAEPQKITRTAKFKINTDIRPDLVPVLNRHFDAFEKFRRKVLDELEARWNKDQKSFQA
MVQCSAKEPYQKKSSCYAWLDTHFITEAKAVLDLPRKPATSLLYNLSGGLKGFLTRRETVTEDIQKRFNDNLREWNGDLSQLAS
DLKAPLPPAPPNLDFENLTERAIERYNDWVGRTRAWCNLILVQQKKVERRDACLPRYLKGYPGFFGSQRYATTASLAENLKKLE
QEAREQSKKAPTRFAKLSPEIWTAIQERFSPPEGHEAGEKRRPRTAHQTVCLRFAALRAAHPEWTPAQLAGEILAGIFHGTEKL
KKHLAANGFEDRRAVIKLANLYNVVVAFSLDPIRAAGNYISFYKEETPKRNAFGDVRGGLHQPSDESAAIEIMGFGLQKESGKP
LYNGLLVCKKSEKEHDDSWAFLYCHTEGQMFELANEKAKLRGKLLTDWTGFASRGGSRKKAEASAKQLVRGRVWISEKTPPTVM
PLAFGSRQGREYLWHFDRDLREKNEWVLGNGRLLRIMPPGRPNAADFYLTITLERQAPPVAEFIASKLIGIDRGEAVPAAYALI
DREGRWLVNQKRFQEFLKALGTWHDELDEWWKKNKNTPINDRVKRPRLQWTSELEDGFGFVASEYREQQADFNDQKRELQRTQG
GYTRWLRSKERNRARALGGEVTRAVLSLASEHRSPLVLEKLGSALATRGGKGTMMSQMQYERMLTTLEQKLAEVGLYAMPSAPK
YRKLTNGFINFAAPHYTSSTCSSCGQVHDSIFYNELAKKIVRPNDSKWQVTLPSGQTRTLPQTYSYWLKGKGEQTKQTNERLNE
LLGERTVTQLSKSSYMTLVSLLKNSWLPYRPQQADFCCLNCGYTTNADVQGALNIARKFLFRSERGKKTDDAGEEGEAKRGKYK
DDWQTWYRQKLEKVWCK (SEQ ID NO: 15)
>3300018016|Ga0187880_1008327_2|M
[水生-淡水-泥炭地]
MKAKYRSNSIHASLRQLLELGLSKSSSAEPQKITRTAKFKINTDIRPDLVPVLNRHFDAFEKFRRKVLDELEARWNKDQKSFQA
MVQCSAKEPYQKKSSCYAWLDTHFITEAKAVLDLPRKPATSLLYNLSGGLKGFLTRRETVTEDIQKRFNDNLREWNGDLSQLAS
DLKAPLPPAPPNLDFENLTERAIERYNDWVGRTRAWCNLILVQQKKVERRDACLPRYLKGYPGFFGSQRYATTASLAENLKKLE
QEAREQSKKAPTRFAKLSPEIWTAIQERFSPPEGHEAGEKRRPRTAHQTVCLRFAALRAAHPEWTPAQLAGEILAGIFHGTEKL
KKHLAANGFEDRRAVIKLANLYNVVVAFSLDPIRAAGNYISFYKEETPKRNAFGDVRGGLHQPSDESAAIEIMGFGLQKESGKP
LYNGLLVCKKSEKEHDDSWAFLYCHTEGQMFELANEKAKLRGKLLTDWTGFASRGGSRKKAEASAKQLVRGRVWISEKTPPTVM
PLAFGSRQGREYLWHFDRDLREKNEWVLGNGRLLRIMPPGRPNAADFYLTITLERQAPPVAEFIASKLIGIDRGEAVPAAYALI
DREGRWLVNQKRFQEFLKALGTWHDELDEWWKKNKNTPINDRVKRPRLQWTSELEDGFGFVASEYREQQADFNDQKRELQRTQG
GYTRWLRSKERNRARALGGEVTRAVLSLASEHRSPLVLEKLGSALATRGGKGTMMSQMQYERMLTTLEQKLAEVGLYAMPSAPK
YRKLTNGFINFAAPHYTSSTCSSCGQVHDSIFYNELAKKIVRPNDSKWQVTLPSGQTRTLPQTYSYWLKGKGEQTKQTNERLNE
LLGERTVTQLSKSSYMTLVSLLKNSWLPYRPQQADFCCLNCGYTTNADVQGALNIARKFLFRSERGKKTDDAGEEGEAKRGKYK
DDWQTWYRQKLEKVWCK (SEQ ID NO: 15)
>3300018017|Ga0187872_10010132_5|M
[水生-淡水-泥炭地]
MKAKYRSNSIHASLRQLLELGLSKSSSAEPQKITRTAKFKINTDIRPDLVPVLNRHFDAFEKFRRKVLDELEARWNKDQKSFQA
MVQCSAKEPYQKKSSCYAWLDTHFITEAKAVLDLPRKPATSLLYNLSGGLKGFLTRRETVTEDIQKRFNDNLREWNGDLSQLAS
DLKAPLPPAPPNLDFENLTERAIERYNDWVGRTRAWCNLILVQQKKVERRDACLPRYLKGYPGFFGSQRYATTASLAENLKKLE
QEAREQSKKAPTRFAKLSPEIWTAIQERFSPPEGHEAGEKRRPRTAHQTVCLRFAALRAAHPEWTPAQLAGEILAGIFHGTEKL
KKHLAANGFEDRRAVIKLANLYNVVVAFSLDPIRAAGNYISFYKEETPKRNAFGDVRGGLHQPSDESAAIEIMGFGLQKESGKP
LYNGLLVCKKSEKEHDDSWAFLYCHTEGQMFELANEKAKLRGKLLTDWTGFASRGGSRKKAEASAKQLVRGRVWISEKTPPTVM
PLAFGSRQGREYLWHFDRDLREKNEWVLGNGRLLRIMPPGRPNAADFYLTITLERQAPPVAEFIASKLIGIDRGEAVPAAYALI
DREGRWLVNQKRFQEFLKALGTWHDELDEWWKKNKNTPINDRVKRPRLQWTSELEDGFGFVASEYREQQADFNDQKRELQRTQG
GYTRWLRSKERNRARALGGEVTRAVLSLASEHRSPLVLEKLGSALATRGGKGTMMSQMQYERMLTTLEQKLAEVGLYAMPSAPK
YRKLTNGFINFAAPHYTSSTCSSCGQVHDSIFYNELAKKIVRPNDSKWQVTLPSGQTRTLPQTYSYWLKGKGEQTKQTNERLNE
LLGERTVTQLSKSSYMTLVSLLKNSWLPYRPQQADFCCLNCGYTTNADVQGALNIARKFLFRSERGKKTDDAGEEGEAKRGKYK
DDWQTWYRQKLEKVWCK (SEQ ID NO: 15)
>3300018022|Ga0187864_10004559_12|P
[水生-淡水-泥炭地]
MKAKYRSNSIHASLRQLLELGLSKSSSAEPQKITRTAKFKINTDIRPDLVPVLNRHFDAFEKFRRKVLDELEARWNKDQKSFQA
MVQCSAKEPYQKKSSCYAWLDTHFITEAKAVLDLPRKPATSLLYNLSGGLKGFLTRRETVTEDIQKRFNDNLREWNGDLSQLAS
DLKAPLPPAPPNLDFENLTERAIERYNDWVGRTRAWCNLILVQQKKVERRDACLPRYLKGYPGFFGSQRYATTASLAENLKKLE
QEAREQSKKAPTRFAKLSPEIWTAIQERFSPPEGHEAGEKRRPRTAHQTVCLRFAALRAAHPEWTPAQLAGEILAGIFHGTEKL
KKHLAANGFEDRRAVIKLANLYNVVVAFSLDPIRAAGNYISFYKEETPKRNAFGDVRGGLHQPSDESAAIEIMGFGLQKESGKP

LYNGLLVCKKSEKEHDDSWAFLYCHTEGQMFELANEKAKLRGKLLTDWTGFASRGGSRKKAEASAKQLVRGRVWISEKTPPTVM
PLAFGSRQGREYLWHFDRDLREKNEWVLGNGRLLRIMPPGRPNAADFYLTITLERQAPPVAEFIASKLIGIDRGEAVPAAYALI
DREGRWLVNQKRFQEFLKALGTWHDELDEWWKKNKNTPINDRVKRPRLQWTSELEDGFGFVASEYREQQADFNDQKRELQRTQG
GYTRWLRSKERNRARALGGEVTRAVLSLASEHRSPLVLEKLGSALATRGGKGTMMSQMQYERMLTTLEQKLAEVGLYAMPSAPK
YRKLTNGFINFAAPHYTSSTCSSCGQVHDSIFYNELAKKIVRPNDSKWQVTLPSGQTRTLPQTYSYWLKGKGEQTKQTNERLNE
LLGERTVTQLSKSSYMTLVSLLKNSWLPYRPQQADFCCLNCGYTTNADVQGALNIARKFLFRSERGKKTDDAGEEGEAKRGKYK
DDWQTWYRQKLEKVWCK (SEQ ID NO: 15)
>3300018040|Ga0187862_10006434_12|P
[水生-淡水-泥炭地]
MKAKYRSNSIHASLRQLLELGLSKSSSAEPQKITRTAKFKINTDIRPDLVPVLNRHFDAFEKFRRKVLDELEARWNKDQKSFQA
MVQCSAKEPYQKKSSCYAWLDTHFITEAKAVLDLPRKPATSLLYNLSGGLKGFLTRRETVTEDIQKRFNDNLREWNGDLSQLAS
DLKAPLPPAPPNLDFENLTERAIERYNDWVGRTRAWCNLILVQQKKVERRDACLPRYLKGYPGFFGSQRYATTASLAENLKKLE
QEAREQSKKAPTRFAKLSPEIWTAIQERFSPPEGHEAGEKRRPRTAHQTVCLRFAALRAAHPEWTPAQLAGEILAGIFHGTEKL
KKHLAANGFEDRRAVIKLANLYNVVVAFSLDPIRAAGNYISFYKEETPKRNAFGDVRGGLHQPSDESAAIEIMGFGLQKESGKP
LYNGLLVCKKSEKEHDDSWAFLYCHTEGQMFELANEKAKLRGKLLTDWTGFASRGGSRKKAEASAKQLVRGRVWISEKTPPTVM
PLAFGSRQGREYLWHFDRDLREKNEWVLGNGRLLRIMPPGRPNAADFYLTITLERQAPPVAEFIASKLIGIDRGEAVPAAYALI
DREGRWLVNQKRFQEFLKALGTWHDELDEWWKKNKNTPINDRVKRPRLQWTSELEDGFGFVASEYREQQADFNDQKRELQRTQG
GYTRWLRSKERNRARALGGEVTRAVLSLASEHRSPLVLEKLGSALATRGGKGTMMSQMQYERMLTTLEQKLAEVGLYAMPSAPK
YRKLTNGFINFAAPHYTSSTCSSCGQVHDSIFYNELAKKIVRPNDSKWQVTLPSGQTRTLPQTYSYWLKGKGEQTKQTNERLNE
LLGERTVTQLSKSSYMTLVSLLKNSWLPYRPQQADFCCLNCGYTTNADVQGALNIARKFLFRSERGKKTDDAGEEGEAKRGKYK
DDWQTWYRQKLEKVWCK (SEQ ID NO: 15)
>OCZT010053476_2|M
[土壤宏基因组]

[0267]

MSKEPRKNSTHASLRQLLELGLSKSSSAEPQKITRTVKFKIHTLARPELVAPLNKHFNAFDQFREKVLAELESWWDKDPDSFTH
MVKCSTKLKFEEKSSCYAWLYKKFLTGAELPPELSRDAANAMLDNLGGGLKSFLTRRATVSLEIKKRYEQNKEEWEKELSLIAK
ERNLKLPEVPREIDFNNLTAKKIDFYNEWVGLVRMWCNLILIQKEKLTRENACLPRYLKSYPGFPGSQRFRKPVTIEQTLAELE
TIAREQAGKASQRFADLKPEVWQAIQERFSPPKKEEAETTPRPRTASQTVSRRFTALIQANPNWKPQQIAEEILAGIFRAADKL
KKHLVTTKYTDRRATIKFANLFNVAVVFAREPFRARGDYVASYEADIARQKAFGDLRGALHQPSDESSAIQISGFSIKDGGTPN
YNGLLVCKSEANGKDSWAFLYCHQEGQIFQLADEKAKLRGKKLSEWTGFASRGGSRKKAEASPKQLVKSSVWVSEKNPPTVLPL
AFGTRLGREYLWHFDRNLRTKSEWVLGNGRLLRVMPPGRTDLAEFYLAITMERETPPVASVVANKLIGIDRGEAVPASYAVVDG
EERLLDSGRIAENYREEQRAFNKSKRELQSERGGYTRTLRSKERNRARALGGEVTRKLLSLAAEHCAPLVLENLNSSLATRGGK
NTMMSQMQYERVFTALEQKLAEAGLYSLPSAPKFRKGDNAFIKLVGPAYTSQTCSACGHVHWAEFYEKILDTLAPCGEKQWQVT
VEGKARTLPESYTYWLRGKGEQTKNTHERIIELLKDRPVSKLNKTDRKSLVSLFKNRWMPYRPKQADFQCVVCGHKMNADEQGA
LNIARKFLFASGRGKKSGEMTEAERRKVRGEWEKWYKEKLATVWKKQ (SEQ ID NO: 16)
>3300014839|Ga0182027_10014967_5|M
[陆生-土壤-沼地]
MSKEYRPDSIHASLRHLLELGLSKTGSSEPQKITRTVKFKVHTLARPELVAPLNQHFNAFDEFRHKVLADLESWWDKAPDSFTR
MVKSSTKAKYEEKSSCYAWLYRKYLTGAELPTELSRDAANAMLDNLAGSLKSFLTRRATVSLEIKKRYEQNREEWEKELSLVAK
EKNLTLPKPPPEIDFDHLNVMAIGLYNEWVGLVRMWCNLILIQKEKLTRENACLPRYLKSYPGFPGSQRFRKPVTIEQTFGELE
AAARDQAGNATNRFANVKPEVWQAIEERFKPPAKDEVEIEARPRSRTASQTVARRLAALVKANPNWTPDQLAVEILAGIFRAAD
KLQKHLAATEYTDHRATIKLANLFNAAVVFAREPFRASGDYVASYGADMARQKAFGDLRGALHQPSDESSAIQITGFSVNEDGS
PNYNGLLVCKPEADGKDSWAFLYCHQEGQTFQLADEKAKLRGKKLSEWTGFASRGGSRKKAEASAKQLVKGRVWVSEKKPPTVL
PLAFGTRMGREYLWHFDRNLRTKSAWVLGNGRLLRVMPPARPDLAEFYLTITMEREAPPVASVVANKLIGIDRGEAVPAAYAVV
DSEGRLLDSGRIAEKYREEQRAFNKSKRELQSERGGYTRTLRSKERNRARALGGEVTRRLLSLAAEHRAPLVLENLNSSLATRG
GKNTMMSQMQYERVFTALEQKLAEAGLYSLPSAPKYRKGDNAFIKLVGPAYTSQTCSACGHVHSSEFYDQLTDTLAPCGDKQWR
VTVEGKPRILPESYTYWLRGKGDQTRNTNERIVELLKDRPISKLTKTDRKSLVSLVKNRWMQYRPTQADFQCVVCGHKMNADEQ
GALNIARKFLFALGRGKKSGEMTEAERRKVRGEWEKWYKEKLGTVWRK (SEQ ID NO: 17)

[0268] 表3.代表性CLUST.121143直接重复的核苷酸序列

[0269]

CLUST.121143 效应子蛋白登录号	直接重复核苷酸序列
3300013006|Ga0164294_10029128_3|M (SEQ ID NO: 1)	GGTTTGCTGACCTCGATTGTTGGAGGAGTGCAAAAG (SEQ ID NO: 101)
3300013087|Ga0163212_1009539_3|M (SEQ ID NO: 2)	GGTTTCGCACTCCCTCATTGTTGGTGGTGTGCAAAAG (SEQ ID NO: 102)
3300013131|Ga0172373_10014249_3|P (SEQ ID NO: 3)	GGTTTGCTCCGGTCCTTGTTGGTAGGTGTGCAAAAG (SEQ ID NO: 103)
3300013131|Ga0172373_10021309_4|M (SEQ ID NO: 4)	GGTTAGCTCCCATCCTTGTTGGGTGGTTTGCAAAAG (SEQ ID NO: 104)
3300013132|Ga0172372_10046622_2|M (SEQ ID NO: 4)	GGTTAGCTCCCATCCTTGTTGGGTGGTTTGCAAAAG (SEQ ID NO: 104)
3300014720|Ga0172376_10007396_6|M (SEQ ID NO: 3)	GGTTTGCTCCGGTCCTTGTTGGTAGGTGTGCAAAAG (SEQ ID NO: 103)
3300014720|Ga0172376_10013958_4|M (SEQ ID NO: 4)	GGTTAGCTCCCATCCTTGTTGGGTGGTTTGCAAAAG (SEQ ID NO: 104)
3300018790|Ga0187842_1006374_2|M (SEQ ID NO: 5)	GGTTTGCTGACCTCGATTGTTGGAGGAGTGCAAAAG (SEQ ID NO: 101)
3300018815|Ga0187845_1009203_1|M (SEQ ID NO: 6)	GGTTTGCTGACCTCGATTGTTGGAGGAGTGCAAAAG (SEQ ID NO: 101)
3300018868|Ga0187844_10013314_2|M (SEQ ID NO: 7)	GGTTTGCTGACCTCGATTGTTGGAGGAGTGCAAAAG (SEQ ID NO: 101)
3300019093|Ga0187843_10013171_2|M (SEQ ID NO: 8)	GGTTTGCTGACCTCGATTGTTGGAGGAGTGCAAAAG (SEQ ID NO: 101)
3300023179|Ga0214923_10008431_2|P (SEQ ID NO: 9)	TTGCCTCCTGTGGGGTTGTATGTGAAG (SEQ ID NO: 105)
3300027730|Ga0247833_1032983_1|M (SEQ ID NO: 10)	GGTTTGCTGACCTCGATTGTTGGAGGTGTGCAAAAG (SEQ ID NO: 106)
3300028114|Ga0247835_1021903_1|P (SEQ ID NO: 11)	GGTTTGCTGACCTCGATTGTTGGAGGTGTGCAAAAG (SEQ ID NO: 106)
3300028553|Ga0247839_1034275_1|M (SEQ ID NO: 10)	GGTTTGCTGACCTCGATTGTTGGAGGTGTGCAAAAG (SEQ ID NO: 106)
3300028569|Ga0247843_1031164_2|P (SEQ ID NO: 11)	GGTTTGCTGACCTCGATTGTTGGAGGTGTGCAAAAG (SEQ ID NO: 106)
3300014164|Ga0181532_10001487_14|M (SEQ ID NO: 12)	GTGAGAACCGCTGCGGTTCGCGGGGATGGTTTCAAAG (SEQ ID NO: 107)
3300014164|Ga0181532_10001487_14|P (SEQ ID NO: 13)	GTGAGAACCGCTGCGGTTCGCGGGGATGGTTTCAAAG (SEQ ID NO: 107)
3300014165|Ga0181523_10000360_35|M (SEQ ID NO: 12)	GTGAGAACCGCTGCGGTTCGCGGGGATGGTTTCAAAG (SEQ ID NO: 107)
3300014165|Ga0181523_10000360_33|P (SEQ ID NO: 13)	GTGAGAACCGCTGCGGTTCGCGGGGATGGTTTCAAAG (SEQ ID NO: 107)
3300014200|Ga0181526_10000193_31|M (SEQ ID NO: 12)	GTGAGAACCGCTGCGGTTCGCGGGGATGGTTTCAAAG (SEQ ID NO: 107)

[0270]

3300014200|Ga0181526_10000193_30|P (SEQ ID NO: 13)	GTGAGAACCGCTGCGGTTCGCGGGGATGGTTTCAAAG (SEQ ID NO: 107)
3300027720|Ga0209617_10001817_3|M (SEQ ID NO: 14)	GGTTTGCTGACCTCGATTGTTGGAGGAGTGCAAAAG (SEQ ID NO: 101)
3300009764|Ga0116134_1001531_13|P (SEQ ID NO: 15)	TTCGGAATCGCTGACGCTCGCGGGGCTGGTTTCAAAG (SEQ ID NO: 108)
3300018009|Ga0187884_10009678_2|M (SEQ ID NO: 15)	TTCGGAATCGCTGACGCTCGCGGGGCTGGTTTCAAAG (SEQ ID NO: 108)
3300018014|Ga0187860_1004616_12|M (SEQ ID NO: 15)	TTCGGAATCGCTGACGCTCGCGGGGCTGGTTTCAAAG (SEQ ID NO: 108)
3300018016|Ga0187880_1008327_2|M (SEQ ID NO: 15)	TTCGGAATCGCTGACGCTCGCGGGGCTGGTTTCAAAG (SEQ ID NO: 108)
3300018017|Ga0187872_10010132_5|M (SEQ ID NO: 15)	TTCGGAATCGCTGACGCTCGCGGGGCTGGTTTCAAAG (SEQ ID NO: 108)
3300018022|Ga0187864_10004559_12|P (SEQ ID NO: 15)	TTCGGAATCGCTGACGCTCGCGGGGCTGGTTTCAAAG (SEQ ID NO: 108)
3300018040|Ga0187862_10006434_12|P (SEQ ID NO: 15)	TTCGGAATCGCTGACGCTCGCGGGGCTGGTTTCAAAG (SEQ ID NO: 108)
OCZT010053476_2|M (SEQ ID NO: 16)	GTTTGCATGATGACGTTTGCGGGGTTGGTTTCAAAG (SEQ ID NO: 109)
3300014839|Ga0182027_10014967_5|M (SEQ ID NO: 17)	CTCTGACGCGCTGACGTTCGCGGGGTTGGTTTCAAAG (SEQ ID NO: 110)

[0271] 表4.代表性CLUST.121143系统的含直接重复同源性的区域

[0272]

效应子登录号	同源区
3300018040|Ga0187862_10006434_12|P	CCGCGAG(SEQ ID NO:200)

[0273] 表5.代表性CLUST.121143系统的含直接重复同源性的基因座序列

[0274]

>3300018040|Ga0187862_10006434_12|P
TGTTCCGTTCCGAACGAGGCAAAAAGACTGACGACGCAGGCGAAGAAGGTGAGGCAAAGCGTGGTAAATACAAAGACGACTGGC
AGACTTGGTATCGCCAAAAGTTGGAAAAGGTCTGGTGCAAGTAATCTGACATCGCCCATTCGAGACCCGCGAGCATGACATGGA
GACCGACGATTTCCTTCCTATCTACTCGCTCGCGGAGTGGCATTATTGCCCGCGCTCGTCTTTTCTGTCGTGGTTTGGTGCAGA
GCGTGAGGATCGCGTCACCCCGGCTTACCAGAAAATGCGCGAAGTCCACGCGGTG (SEQ ID NO: 201)

[0275] 表6.代表性CLUST.121143系统的非编码序列

[0276]

>3300028569|Ga0247843_1031164_2|P
CACAATATTTCTTAAATGTCGAGAAAATAATATTTTACACCTATCCATATGCCAAATTACAAAAATGAATGTTTTACCTCATAA
AGAGTAAAAAGTGCTCCCCCGTCACAAAACATTAGTTTGGTGACTGGTTGTTTTTAACCACCAACTTCTCTGTCCAATTTATTG
GGGTTTTGTGAATCCCTATTCCAAGGACCGTTTTTGTACTACCCCCTAATTTACAGTTATCCTTATTATGAACTGGAATAGCTC
CCTTAAAAGTCCGATTATAACTGCTTGCTTAATACGATGAATCACAAAAATACCGTTAAAGCTCCTGTTCGCGAACTTACCCGG
CATGCATTGCAGAAACCCGGGGAGGAGAGTGAACCGGGAGAAACAACAAAGACTCTTAAGTTCCGTGTGGATGCCAAGGGACAT
ACCCGAACCCGACTCGACCAGGCAGCACTGAACATTGCTCGGAAAAGCATCTGGGTCAGTTCCCTTCCACCCCGTCAGGCAAAG
CAATCGGAAGGTGAAGACCGGCGTTCCACCGAGGAACCATGGCGCGTTTGGTACAAGAAGCAGATCATGGGAAAGTGGTCTTAA

[0277]

AAAATCCCTTGCCACCCGGCTATCATAGGATCAAATTAACCCCTTAGTATTTTCTTTGCACAACGTCGACTGCCCTGCCAAGGA
GCGGCTGATAACCAACAAGTTCTTTTACATTTGAGGTGCAAACCGTTCTTGTGATTTTTTTATAAAAGGATGCACCCGAAAGGG
TTGGCTATACCCACATACTGCCCGCTTGCATTATGT (SEQ ID NO: 202)
>3300013006|Ga0164294_10029128_3|M
GTCTACTTTCACGGCAACGTGAAAGACTTGGAACCCAAACCCGGTTCGGTCACACCGATCCAGAAAACGGCCGCGGGGCTGATG
TTTCTTGTCGATGCGATCATGCTGGGACTCGCTCTTTATTTGATCCTTCGCTAACGCGAACCGATCGATAAGGATTAAGTCGCA
CTACCCCTGCATCGATTCCCGTTTTTTCGATTCCCTTCCAGGTATTTAATTTTTACTTCTCAACGAGGTATTTGGCTTGGCATA
GTCTCCTCTGCGCCCCCTGCCCCCTGTCCCCTGCCCCCTTTCACTTCCAGCCTACCAATTCCCAGTGTTCAAAACGTCGGTGTT
TTGGCGTTTATCCTTTCCTGAGATAACGCCCAGGACTGTTTTTGTCCCACCTCCTAATTTACACTTATCCTTATTATGAACTGG
AATAGCTCCGTTAAAAGTCCGATTATGACTGCTTGCTTAATACGATGAATCACAAAAGCACTGTCAAAGCCTCTGTTCGCCATC
TTACTCGGCATGCATTGCAGAAACCCGGGGAGGAGAGTGATCCGGGAGAAACGACAAAGACTTTAAAGTTCCGGGTAGATGCCA
AAGGACATACCCGAACCCGACTCGACCAGGCAGCACTGAACATTGCCCGCAAAAGCATCTGGGTCAGCTCCCTTCCACCCCGTC
AGGCAAAGCAATCGGAAGGTGAAGACCGGCGTTCCACCGAGGAACCATGGCGCGTTTGGTACAAAAAGCAGATCATGGAAAAGT
GGTCTTAAAAAATCCCTTGCCACCCGGCTATCATAGGATCATTTTAGTCTCTTAGTATTTTCTTTGCACAACGTCGACTGCCCT
GCCAAGGAGCGGCTGACAACCAACAAGTTCTTTTACATTTGAGGTGCAAACCGTTCTTGTGATTTTTTTACAAAAGGATGCACC
CGAAAGGGTTGGCCATACCC (SEQ ID NO: 203)
>3300018868|Ga0187844_10013314_2|M
ACGGCAACGTGAAAGACTTGGAACCCAAACCCGGTTCGGTCACACCGATCCAGAAAACGGCCGCGGGGCTGATGTTTCTTGTCG
ATGCGATCATGCTGGGACTCGCTCTTTATTTGATCCTTCGCTAACGCGAACCGATCGATAAGGATTAAGTCGCACTACCCCTGC
ATCGATTCCCGTTTTTTCGATTCCCTTCCAGGTATTTAATTTTTACTTCTCAACGAGGTATTTGGCTTGGCATGGTCTTCTCTG
CGCCCCCTGCCCCCTGTCCCCTGCCCCCTTTCACTTCCAGCCTACCAATTCCCAGTGTTCAAAACGTCGGTGTTTTGGCGTTTA
TCCTTTCCTGAGATAACGCCCAGGACTGTTTTTGTCCCACCTCCTAATTTACACTTATCCTTATTATGAACTGGAATAGCTCCG
TTAAAAGTCCGATTATGACTGCTTGCTTAATACGATGAATCACAAAAGCACTGTCAAAGCCTCTGTTCGCCATCTTACTCGGCA
TGCATTGCAGAAACCCGGGGAGGAGAGTGATCCGGGAGAAACGACAAAGACTTTAAAGTTCCGGGTAGATGCCAAAGGACATAC
CCGAACCCGACTCGACCAGGCAGCACTGAACATTGCCCGCAAAAGCATCTGGGTCAGCTCCCTTCCACCCCGTCAGGCAAAGCA
ATCGGAAGGTGAAGACCGGCGTTCCACCGAGGAACCATGGCGCGTTTGGTACAAAAAGCAGATCATGGAAAAGTGGTCTTAAAA
AATCCCTTGCCACCCGGCTATCATAGGATCATTTTAGTCTCTTAGTATTTTCTTTGCACAACGTCGACTGCCCTGCCAAGGAGC
GGCTGACAACCAACAAGTTCTTTTACATTTGAGGTGCAAACCGTTCTTGTGATTTTTTTACAAAAGGATGCACCCGAAAGGGTT
GGCCATACCC (SEQ ID NO: 204)
>3300023179|Ga0214923_10008431_2|P
TCAAAACCAGCCGTAGCTGGCGGAAAATATGCTTGGGATTGTCATTGGTATGTGGCCAATTCCTTTCCCACACGGCGTGCTTGT
CCACAGGATTTTAACCCTGCCTGTTTTTTTGATATCAAACCTGATATCGGTCATTTCTCCTCGATACTCTGAAACCAATCAGAA
ACACCCTTCCCAGGATGCTGAGTTAAGAACCACACGCTAACTTATCGTTTGGCTGAAAGGGATTTGAACCCTTTTCTTATTATA
TTGCTGAATACAGAGAAATTCTAAGAAATCTTAAAATATATTTTACCTGCAAACAGCACAAAGCACTAATAGCCAGTTATATGC
ACACCTAGTAGGATTGAATCTTATTAATACAACCTAGGACAGCAAATGACCCACCATATGAGTTATTATTGTGGAATGAAGAAT
TACCAAACTTGTGGAAAGTCGATATCTGGGGATACGATCAATGCTTAATCCAGCTGAAATGCAAATGCCCTCAAATAGTATAAA
ATCGATCAAAACTTCAGTACGAGAGCTGTACAACCACGCCTTGAACAAACAGGGCGATGGTGAGCCAGGAGAGACAACCAAGAC
ACTTAAGTTCAAGATCGATCCACAAGCGGCGCTGAACATTGCCAGGAAGACGATCTGGGTCAGTACCCTACCTCCTAAAAAAGC
CAAGAGCGGTGATCAGGACTTTCGTGGTGTAGGACAAGAGCTGTGGGCGAAATGGTACAAGGATAGGGTTCACAAAAACTGGGC
ATAAAATAGCACTTGTCTGAATTGGGATGGGTTAATACATTGAAATCTTAGTATTTACTTTGCATAGCGTTGCCAACCTTGCCC
AAGGATGAGTCAACGCCCCATAGAAGGCAGTTCATTAACACGCAAAAGGCACACCGCCGTAGTAAGTAGGGGGCAGCAACCACA
GATGGTGAGATGCCTCCTATGGGGTCATATGTGAAGTTCATGTTGTCTCCCATCTTGGACCCATAGCCGTTGGTGAGTTGCCTC
CTGTGGGGCCATATGTGAAGGAGGCCAATTCTGCTCGATTGTGATCCAATGTGATGGTGAGTTGCCTCCTGTGGGGTTGTATGT
GAAGGTCAATGTGGAGCGATTGCTGGGGGTATTCAGTTGGTGATTTGCCTCCTATGGGGCTATGTGTGAAGGGATGCGATCCTG
CAACATTCAAGCCTAGTCAGTTGCAAGCGTTTTCTGCAAATGATAGAGATCTGTTGGTGGTTTGCCTTCTGTGGGGCTATGTGT
GAAGGATGTCGAAGCGCCATTGCTCGGCATTGATGCGTGTGTGATCTGGTTTCTGTGGAGTTATGTGTGAAGGCTCGCTCTCAT
AGTTCTTTGGCTTGGCACTCATAGCACTTTTCCGTTTAAACCAGCAAAAAATCGCCGCACTAGATAATGCAACATAATTTGTAT

[0278]

AAATTATCCAAATAAATGCCTGTTTTCTCAACACAAGAAAAACAGCCTATCCTTTCATACTTATCTGTTTCTCTCTGTGCATCA
ATTAAATATATAATAGGATCTTAATCGTATAATAGCTTAATCCAGAGTAACCTCCTATTAGAAATCCGCTGCTCTAGTCCCCTG
AGCTACGGGCGCAGCACGTAGTTTTGTAGCAGAAAATTGTCAGAGAGGCAGGGATTATGGTGGAGTAAGAAGGGGGTTAAGGGA
TGATGGGATTAAGGGGAAATCTGAAAAAGGGGTTAAGGGATTCAGGGGTTAAGGAATTAAGGAAGAGACAAATCGATGACTGAC
AAGACGTTTAAGGTTTTTTAGGACGGAGGAGGAGGCCGAGGTCGGGGAGCTTGAGCAGTTTGGCGACAAGGAAGTAGGCGAGTG
ACCCGAGAAGAACGAGCGCAAAGGTTCCACCGATACGAACAGGGATGGGTTGGCCCATGGGTTGAATGAATTGGAGGCCTATAA
GTAGGGCTCCAATTAAGGTGACGCAGCAGAGAAGTATCTTTCCGACTCCTTGGAGGAAGTGAGGTTCGAAGTAGCTGCCTACGT
GCCGTTGGGAGGCCCACACCAGCTGCAGAATATTCAGCGTCGCCACAGAGGCTGTGGTGAGAGTGAGTCCTAAAATGCCGAGGT
TGTAGACCCGAATAAGAATGTAGTTGAAGGTGAGGTTGAGAAAAATTCCCGTAAGGCTAACGCGAAGCGGAATCCCCGGGCGAT
CGATAGCCGAGAAGGCGGGGGCG (SEQ ID NO: 205)
>3300028553|Ga0247839_1034275_1|M
CCTCATAAAGAGTAAAAAGTGCTCCCCCGTCACAAAACATTAGTTTGGTGACTGGTTGTTTTTAACCACCAACTTCTCTGTCCA
ATTTATTGGGGTTTTGTGAATCCCTATTCCAAGGACCGTTTTTGTACTACCCCCTAATTTACAGTTATCCTTATTATGAACTGG
AATAGCTCCCTTAAAAGTCCGATTATAACTGCTTGCTTAATACGATGAATCACAAAAATACCGTTAGAGCTTCTGTTCGCGAAC
TTACCCGGCATGCATTGCAGAAACCCGGGGAGGAGAGTGAACCGGGAGAAACAACAAAGACTCTTAAGTTCCGTGTGGATGCCA
AGGGACATACCCGAACCCGACTCGACCAGGCAGCACTGAACATTGCTCGGAAAAGCATCTGGGTCAGTTCCCTTCCACCCCGTC
AGGCAAAGCAATCGGAAGGTGAAGACCGGCGTTCCACCGAGGAACCATGGCGCGTTTGGTACAAGAAGCAGATCATGGGAAAGT
GGTCTTAAAAAATCCCTTGCCACCCGGCTATCATAGGATCAAATTAACCCCTTAGTATTTTCTTTGCACAACGTCGACTGCCCT
GCCAAGGAGCGGCTGATAACCAACAAGTTCTTTTACATTTGAGGTGCAAACCGTTCTTGTGATTTTTTTATAAAAGGATGCACC
CGAAAGGGTTGGCTATACCCAAGGATTCCCAGCCACCCATCTGACGGACTGGTTTGCTGACCTCG (SEQ ID NO: 206)
>3300014839|Ga0182027_10014967_5|M
TGCCGGGCGGCGGGGAAAGCGTATGATTTGCGCCGTCACCAGAAAACAAACGAAGGCACAAAATGGGGCGGAAAAGAAAAGGCA
TCGGGCAGAAAGAGCAAGCCCAGCCGGATTTGGGTCCGGCCGGGCTTGTCCAAGAAACGCTTTTGAACAGGTGTATCCGCATGG
AGGATAAGTCCATATTCAGCTTCCCTTGCTTGACCACTTTACTTTACAACACTTCGAATTTATGAGCAAAGAATACCGACCGGA
TTCGATTCACGCCTCGCTTCGCCATCTGCTTGAACTTGGCTTGAGCAAAACCGGTTCGAGTGAACCGCAAAAGATCACGCGCAC
AGTTAAGTTCAAGGTACACACGCTTGCCCGACCGGAACTGGTGCAAGGCGCATTGAACATCGCCCGCAAGTTCCTTTTCGCCTT
GGGACGCGGGAAGAAATCCGGCGAGATGACCGAAGCGGAGCGGCGCAAAGTCAGGGGCGAGTGGGAAAAATGGTATAAGGAAAA
ACTCGGAACCGTTTGGAGAAAGTGATAAGTCTGGATTTTCGCCCATTGCGGGCCTGATATGGAAAACGATGATTTTCTGCCCAT
CTACTCGCTGGCGGAATGGCATTATTGCCCGCGCTCGGCGTTTCTCTCCTGGTTCGGCGCTGAACGCGAAGACCGCGTGACGCC
GGCTTACCAGAAGATGCGTGAAGTCCATGCCGTGAGCAACCATGTGGCCTTCCGCGGCGAACTCGAACGGGGCGGTTGGCTGGG
CGGAAAAAACGCGCTGCTCATTTACCCGCTTCACGATGACGACCTTGAAAATGTGGAGCGGTTCGGGGCGGTGCGCGTCTGGCA
GGCCACATTTGAACTTCTTTAAGAAAACGCTTGCGTCTGTCCTGCTAATGGCGTAAAAAGGCCCCGTTAGTATTTACTTTGATA
CGATCTGGCGTGCCTTTCCGCCCAACGGATTCGACACGACCCACGAACTTCAGCGTGTCTGAGGAAACCTCAAGCGGTGGATGA
GCCGCTGCTTGTCGCGATCCTTGATAAGTCAAACGGAGCGAGTCGGTTAGGTTTGGAAAAAGCCTTGAACCGTTCTCAGCGGTT
TCCTATGCGTAACTTGCCTAAAATCAACACGAGTTTGCGCGTTAGCCGCTGCGACAGCAGCGGCTCACGCGCAAACTCGTGTGA
TCCAAGGGCCGAAGA (SEQ ID NO: 207)
>3300018040|Ga0187862_10006434_12|P
GCGCATCGCCGAGTTCGGCTTCTTGGGCGTGCGCGTCATGACCTGAATGCACACCCCGCGTCGAAACGGCGAGCCTTCCAAGGC
TGGCGACTTGGACTTGTATTTAACTTTGCGTCGGCCCTGGCGGACCAGTTGATTGATAGTCGGCATTGCGTCTTAAAAATGGAA
CGGGGAATATAACAAACCAAATCCGGCTTGCAACAACTATTTCGCCCGATTTGCGGTCTTAACTCATCGTAGCGGCGGTCGGCA
GACCGCCGCATTTCTTCTGGAAATTGCGCTGTCCGGCACTTTGCCGGGAGCCGCTGCGCCCGTTTTGCAGAAGTCTCGTTTTAA
TCCTTTTGGAAAGCACCGCGAAGCCAGCTTCAACTACCGCTTGCCGCCGCCGGGACACGGGCGTAGCATCACTCCACAATCAGC
CATCAGGAAAATGGGTATGAAAATCTCTTTGGGGATTTCATCATGAAAGCAAAGTATCGGTCTAACTCAATTCATGCCTCGCTC
CGGCAGCTTTTGGAGCTTGGCTTAAGCAAGAGCAGTTCGGCAGAACCACAGAAAATCACGCGCACTGCCAAGTTCAAGATTAAT
ACCGACATCCGGCCGGATTTGGTTTTAAACATTGCCCGAAAGTTTTTGTTCCGTTCCGAACGAGGCAAAAAGACTGACGACGCA
GGCGAAGAAGGTGAGGCAAAGCGTGGTAAATACAAAGACGACTGGCAGACTTGGTATCGCCAAAAGTTGGAAAAGGTCTGGTGC
AAGTAATCTGACATCGCCCATTCGGGACCCGCGAGCATGACATGGAGACTGACGATTTCCTTCCTATCTACTCGCTCGCGGAGT

[0279]

GGCATTATTGCCCGCGCTCGTCTTTTCTGTCGTGGTTTGGTGCAGAGCGTGAGGATCGCGTCACCCCGGCTTACCAGAAAATGC
GCGAAGTCCACGCGGTGAGCAAT (SEQ ID NO: 208)
>3300027720|Ga0209617_10001817_3|M
AAACCCAACGGCATAAAGAGGATCAGGAGTTTTGGGTGCCCTTATTATGAGATAAAGAATAAGAATACCCCATACGATATTCAT
CCATTTAAGCAGACGTTGCACTTGGGATATTCTGTCACCTGCGTTGACCATTTGGCAACGTGCCATTATCGATGACAAGGGAGA
GGTATACCCCCAAATGGATATGCCTGTTATAAAACATTAGTTTGGTGACTTGGCATTTTTAGCCACTTTTTCCCTGTCCAATTT
TAGGGGGTTTTGTGAATCCCTATTCCAAGGACTGTTTTTGTCCCACCTCCTAATTTACAGTTATCCTTATTATGAACTGGAATA
GCTCCGTTAAAGGTCCGATTATGACTTCTTGCTTAATACGATGAATCACAAAAGCACCGTCAAGGCTTCTGTTCGCCATCTTAC
CCGGCATGCATTGCAGAAACCCGGGGAGGAGAGTGATCCGGGAGAAACGACAAAGACTTTAAAGTTCCGGGTAGATGCCAAAGC
ACATACCCGAACCCGACTCGACCAGGCAGCACTGAACATTGCCCGCAAAAGCATCTGGGTCAGCTCCCTTCCACCCCGTCAGGC
AAAGCAATCGGAAGGTGAAGACCGGCGTTCCACCGAGGAACCATGGCGCGTTTGGTACAAAAAGCAGATCATGGAAAAGTGGTC
TTAAAAAATCCCTTGCCACCCGGCTATCATAGGATCATTTTAGTCTCTTAGTATTTTCTTTGCACAACGTCGACTGCCCTGCCA
AGGAGCGGCTGACAACCAACAAGTTCTTTTACATTTGAGGTGCAAACCGTTCTTGTGATTTTTTTACAAAAGGATGCACCCGAA
AGGGTTGGCCATACCC (SEQ ID NO: 209)
>OCZT010053476_2|M
CTGAAATTCCCGGCCATCAATCCCGCCAACGCCGCCGAGCGCGATGAATTTCGTCTGAATACCTTCCCACCGGACAACCACACC
GCGCCCGCGCCGCCGACTCCGCCCGCGACACCGAAGCCGTAAAATCAAATTTGGAAAACGCAAAGCGGCCTTTGGGAATGGAGG
CCGCTTTTTGTTTTGGATTGTTCCTGTTCACGAAATCCGCTATTCAGGCTACAAATCTAAAAGGGTAAATATGAGCAAAGAGCC
TCGGAAAAATTCGACTCACGCCTCGCTTCGCCAGTTGCTTGAACTTGGTTTGAGCAAAAGCAGTTCGGCAGAGCCACAGAAGAT
CACGCGCACGGTGAAGTTCAAGATTCACACGCTTGCACGACCGGAACTCGTTGGCGCGTTGAACATCGCCCGCAAATTCCTGTT
CGCCTCGGGACGCGGAAAAAAGTCCGGTGAAATGACCGAAGCCGAGCGGCGCAAAGTGCGGGGCGAGTGGGAAAAGTGGTATAA
GGAAAAACTTGCGACCGTCTGGAAAAAACAATGAAAGGCTTGCGTCAGTTCTGCGAAAGGCGTAAACAGGCGGCGTTAGTATTT
ACTTTGATACGATCTAGCGTGCCTTTCCGTCCAACGGATTCGACACGACCCGTGAACTTCAGCGTGCAAAAGGAAATCACAAGT
GGTGGTTGACCGCTGCCTGTAGCGATCCTTGAAAACTTAAACCGAGCGAATCAGTTAGGTTCGGAAAGAAGCCCGAACTGGTTT
CAAAGTCAATCGCGTCAATGATTTCCTGCAAACCATAAACGGGTCTAAACGGGTCAGTCCTAACTTT (SEQ ID NO: 210)
>3300014720|Ga0172376_10007396_6|M
TGCCCCAGCCCCTAACTCAAGTTCTGTAAAAACATTATTCGAAAGAATAAATCCCCGATCAGAGGATTGCCCAACAATAGTGCC
TTGGTTGTTTATATCGTAACAGGCGCTCCACGTATGGGCGGAACTAAGCACCCCCACGGAACTCATGCTTTTATTTCTGTACAC
GAACCCCCTGTCACCAAGCGAACTTGAGGAGGCACCAACCACAACACCCGAGTCGTTAATGCTGTAGCTTCGACTGGAATTATT
CGTGTAGGTGTACTCGCCTCCGAACCACGATGTCCAAGTAACAACCTGAGGACCAAGAGTCCTAAGATTTACCAGCTTGTAAGT
CGGGGCTTGCGCATTTCCGTTTTCGCAAATTACGAGCGAAGCAATTAAAACGCAGAGCCACCGAACCATATGATTATTCAACAC
CTATTTTACTTTTTTTCAACTAGGGAAAACCGGCTTTTTAACTACTATCCCTTTTGTCCTAAGCATTAACTTTTGCCTCAGAAA
CTGTCCTGATCACTAAACTAATGGTTTGTGACCGGCTGTGTTTTGGGCGTTCCGACTCTGTGGGTTTCATTGGGGTTTTGTGAC
TGCCCCCTCCTAATACCGTATTTGTACCACCCCCTACCTTAGGGTTGGCCCTGTTTTGAACTGGAATACGTACCTTAATCGTCT
GATTATAACCCTTCGGTTAAAACGATGATTCATAAGAGTACCGTTAAAGCATCGGTTCGCCACCTGACCCGCCATGCGTTGGAC
AAACCCGGCGAGGGCGACCCGGGTGAAACAACAAAAACCCTAAAGTTCCGGATTGATGCCAAAGATCACACCCGCACCCGGCTC
GACCAGCAGGCAGCCCTAAACATTGCCCGAAAAAGCATCTGGGTCAGTTCCTTGCCGCACCGGCAGGCAAAACAGTCGGAGGGA
GAAGACCGGCGTTCCACTGAGGAGCCGTGGCGGGAGTGGTACAAGAAGCGGATCTGGGAGAAGTGGGCTTAAAAAAAAGCCCTT
GCCACCCGGTCCCCGCAGGATCATTTTAGTCTCTTAGTATTTTCTTTGCACAGCGTTGACTCCCCTGCCAAGGGTGCGGCCAAC
AATCCAACAAGTTCTTTCGCAACAGGGGGCAAACCCCGTTCAATTTTGCTTTAAGGAACTAAAGCCCATGGCGTTAGGTATAGC
T (SEQ ID NO: 211)
>3300014720|Ga0172376_10013958_4|M
AAATACGGTTCCGTCAAATAAGTTGCCTGAAGAGTCACTGGCGTTCCCTGAAAATTCGTATTTTCCAACCAAGCCATAATTGAG
GTTAACCACGGCAGACTTCGCTGTTGAACCCATCAACAAGATGGCTAATGCGACCACCCGATTCATATATTAATTAAACCATGT
TCTTTTGTTTTTTCAATGTCCCACAGGCTCTTTTTTTGACTAAATAAAATATATCCCTTTTGGTATATCCTTGGTTGGTTTCGT
ATTATCGCATAAAACAAATGGTTTGTGACCGGTCATGTTTTGAGCGTTCCATCTCTGTAGGTTTCATTGGGGTTTTGTGATTAC
CCCCTCCTAATACCGTATTTGTACCACCCCCTACCTTAGGGTTAGCCCTGTTTTGAACTGGAATACGTACCTTAATCGTCTGAT

[0280]

TATAACCCTTCGGTTAAAACGATGATTCATAAGAGTACTGTTAAAGCATCCGTTCGCCACCTGACCCGCCATGCTCTGGACAAA
CCCGGCGAGGGCGATCCGGGTAAACCAACAACAAAAACCCTGAAGTTCCGGATTGATGCCAAGGGGCACACCCGCACCCGGCTC
GACCAAGCCGCTTTAAATATAGCTCGCAAGAGTATCTGGCTGAGTACTTTGCCCAAAAGATCCAAAAAGGAAACACTGGCTGAA
GCACGTGGGTCCACAAAACCAGCGTGGAAAAAGTGGTATCAAAATCGTATCAGCGAGAATTGGTGTTAAAAATACTTGTCTTAC
ACTCGAAAATATTATAAAAAGCAGTCTTAGTATTTTCTTAGCAACACGTTGTCTGTCCTCGCCAAGGAACGGTCAACAATCCAA
CAAGTTCTTTCGCAACGGGGGGGCAAACCAGAAAATTTTCCCTTAAATTTGAATCTTAGGACTTTAGATACTGCCATGCGCGGA
TACTACAAAACCAAATGCATATTTTTCGTA (SEQ ID NO: 212)
>3300013087|Ga0163212_1009539_3|M
GGCAGGGTCAAAGTCATCATGGCTTGGAAGTTGGACAGAATCGGCCGCTCGACCCAGCACTTGGCCGGAATCATCAATCGACTC
CTTAAGGCGGATGTTGGCCTCGTGGTTCCCTCTCAGGGTATTGATACCTCGGCCGGCAGCATGAATCCTGCCAGTAAGCTCCAG
TTGAACGTTCTGGCCGCAGTCGCTGAGTTTGAGAGGGACCTTATCCGCGAGAGGACGAAGGCCGGCATGAAGGCAGCTAAGGCC
AGAGGCGTTAAATTCGGCCGTGAAACGGTGGTAACCGAAGCCCAGCGAATCCTAGCCAAAGAGATTCTTCAGCTTAATCCGGAT
ACTTCGGTGACCAGCCTGTCAAAGCGGTTGGGCTTGTCTTTGGGAACGGCGAGCAGACTTAAGACGATTTTGGGAGGTAGCTGC
TAGTACCGCCAATGTCCTATCCACCCGCTTACGATTGACCCAGTTTTGAACTGGAATACGCACCTTAATCGTCTGATTATAACC
CTTCGGTTAACACGATGATTCATAAGAGTACCGTTAAAGCATCGGTTCGCCACCTGACCCGCCATGCTCTGGACAAACCCGGCG
AGGGCGATCCGGGTGAAACAACAAAAACCCTGAAGTTCCGCATTGATGCCAAAGGTCACACCCGCACCCGGCTCGATCAGCAGG
CCGCCCTGAACATCGCTCGTAAAAGCATCTGGGTCAGATCATTGCCGCCCCGACAGGCAAAGCAGTCGGAGGGCGAGGATCGGC
GTTCCACTGAGGAGCCATGGAGGGAGTGGTATAAGAAGCAGATCCGGGAGAAGTGGGCTTAAAAAAACCCTTGCCACCCGGTCC
CCCTAGGATCATCTTGGTCCCTTAGTATTTTCTTTGTACAGCGTTGACTCCCCTGCCAAGGGTGCGGCCAACAACCAACAAGTT
TTTTCTCAACGAGGGGTGTTATAACCCTGAAATTTCGACAGGGTGCTGAACTGTAGGGTTTTGGCGATAAC(SEQ ID NO: 213)
>3300014200|Ga0181526_10000193_31|M
CACCACGCCGTCTTCGATGGCATTGAACAACGTCTCCAGAAACTTCCGTTCCCGGGCCAATCGTTGGACAACCGTTTGCAACCC
TTCGGCATCCAGGCGTCCGATCCTGCCCAGGACTTTATCGAGAAAAGTCGATATTACCGCAGCCATAATGATGATACCCGCCAA
TGCGCAACCACCGCGCCAATTTGACGCACCCAAAGTATTCAAGCAATTCTGATTAGGCAAGCATTGCTTTCAATTTGAAAGGCC
GGAAAAAACCGTGAAAAATTCGGGGTTTTTCAATTAAACAAGAAAGTTGTCAAGGGGAAAATTGCGCAAAAAAAACAGTCAAAA
TTTCGGCATTTTTGACCTAAGCAATTGCAAACCAAACACTTACGAATTTCCACCTGCTTGCTGCCCATGACTGCAAAATGGTAG
CTTGGGACTGATGGACGAATGCGACTGATGGTCTGGCAACCAGCCAGTGGTCAAGGCGGGTTTCTGCCAGATGGACAGATAGGA
TTGTAATCATAAGACGATTCGGGCAAAACCAGGGCGGACAGGGAAAGCAGGAATTGGACGGTGGGGCGAGATTCCATCGAGCCG
TAAAATCGAGTCGGAGCACAGCCCGTCACGTTGCAGGATTAATGGTGAAAATGGGTCTTGGCAGGACGTTCTGTTTGGGCCATA
ATCTTGCCCGCTCCACAAAACGAAGGGAAAAAGAAGATGAGTGCGGAAACTTCAGAGGCTGGCAAAACCATGTGGGCTTTCGAC
TTGGGCACAGACTGCGAATTGCGCGTCGAATGCCGCAAACCATCGAATTGAAACTATTTATGGATTCAATCCACAAATCCGTTC
GTTCGCTGCTATCGCTTGGCCTGTCGCGTAGCGCTTCGGCAAGTCCGCAAACAATTACGCGCACAGTAAAATTCAAGATCAATA
CCGACATTCGGCAAGCCGCGCTCAATATCGCCAGAAAGCATCTGTTTCGCATTGGCCGCGAAAAAAAATCAGATGAGGAGAATG
AAGCCGCACGACGCAAAGTTATGGGTGAGTGGGAAACTTGGTATCGGCAGAAACTCGCCAGCGTTTGGAATAAATAATTCACCC
TTAGCCATTTTGACTGTGAGCGAGATATGGAGATCGACGATTTCCTTCCAATCTATTCGCTCGCGGAGTGGCATTATTGCCCAC
GCTCGTCTTTTCTGTCGTGGTTCGGCGCGGAACGTGAGGATCGCGTCACCCCGGCTTACCAGAAGATGCGTGAAGTCCATGCCG
TGAGCAATCATGTGGCGTTTCGGGGCGCATTGGAACGCGGCGGCTGGCTGGATGGCAAACACGCGCTGCTGGTTTATCCCTTGC
ATGATGACGACCTTGAAAATGTGGAACGTTACGGCGCGGTACGGGTGTGGCAGGCAACGTTTGAATTTCTTTGAATTTAGGCTT
GCAGTTGGATGGGAGCAGGCATAATACTTCTCCCGTTAGCTGTCTCCTCGGAGGCAGATTTACTTTGATTCTGTGCGCCTTTAC
GCCCCACGTATTTGACGCACTTCGCGAACTTCAGCGGTTAACGGAAACCGCCAGCGGTGGATGGACGATTCACCCGGCGTTCCT
TGACAACACAGTCCGTAGCAGTTGGTTATATACGAAACGGACTAAAAGCGGTTTGCCATGCGTAACTTGTTGCCAGACAATGGA
GTCATCATCTCGGCAGAAATTTTGCATTTAGGTCTTGATGCAAATATTTGGACGGCAGGGCGTGAAGACAGATCAGCATCGCGC
TTGGCCAGGCCC (SEQ ID NO: 214)
>3300019093|Ga0187843_10013171_2|M
CTCTCTGGATTGGGTTGGTCTTCGCCACGATGGTCTACTTTCACGGCAACGTGAAAGACTTGGAACCCAAACCCGGTTCGGTCA
CACCGATCCAGAAAACGGCCGCGGGGCTGATGTTTCTTGTCGATGCGATCATGCTGGGACTCGCTCTTTATTTGATCCTTCGCT
AACGCGAACCGATCGATAAGGATTAAGTCGCACTACCCCTGCATCGATTCCCGTTTTTTCGATTCCCTTCCAGGTATTTAATTT

[0281]

TTACTTCTCAACGAGGTATTTGGCTTGGCATGGTCTTCTCTGCGCCCCCTGCCCCCTGTCCCCTGCCCCCTTTCACTTCCAGCC
TACCAATTCCCAGTGTTCAAAACGTCGGTGTTTTGGCGTTTATCCTTTCCTGAGATAACGCCCAGGACTGTTTTTGTCCCACCT
CCTAATTTACACTTATCCTTATTATGAACTGGAATAGCTCCGTTAAAAGTCCGATTATGACTGCTTGCTTAATACGATGAATCA
CAAAAGCACTGTCAAAGCCTCTGTTCGCCATCTTACTCGGCATGCATTGCAGAAACCCGGGGAGGAGAGTGATCCGGGAGAAAC
GACAAAGACTTTAAAGTTCCGGGTAGATGCCAAAGGACATACCCGAACCCGACTCGACACCCGGGGTGGTCGGGGAACATTGAT
GGGCAACATGCACTATGAGCGAATTTTGTCTGGGTTGAGGGCAAGTTGGCCGAGGCAGGTTGCTACAAGGCACCCAGTGCCGCC
AAGTACCGCAAAGGAATCACGCCGTTCTTGCGATTTGTAAATCCAGCATATACCAGTGCCACCTGTTCAGCTTGTGGTGAGGTG
TTCAGTTCTAGTTGGTATAAGAAAACCGCAGCCACAATCAGTAAGGAATCACCCGGTTCATGGTCAGTGACGCTTCCCAACGGA
AAAAAACTTCAGTTGCCATGCTCCTACACCTCATACATGCGTGGGCGCGGTGAAAAGACATGGAACACCGCTGAACGGCTGGAG
GAACTTTTCAAGGATCGAACATGGTCAGATCTTCCCCCCAGCACCCGAAAAACAGTAGAGGGATTGGTTGCTCGCCTGACTCAT
CATCGCCCCAAACGGGGTCAGGCAGCACTGAACATTGCCCGCAAAAGCATCTGGGTCAGCTCCCTTCCACCCCGTCAGGCAAAG
CAATCGGAAGGTGAAGACCGGCGTTCCACCGAGGAACCATGGCGCGTTTGGTACAAAAAGCAGATCATGGAAAAGTGGTCTTAA
AAAATCCCTTGCCACCCGGCTATCATAGGATCATTTTAGTCTCTTAGTATTTTCTTTGCACAACGTCGACTGCCCTGCCAAGGA
GCGGCTGACAACCAACAAGTTCTTTTACATTTGAGGTGCAAACCGTTCTTGTGATTTTTTTACAAAAGGATGCACCCGAAAGGG
TTGGCCATACCC (SEQ ID NO: 215)

[0282] 实例2-对反式激活RNA元件的鉴定

[0283] 除了效应子蛋白和crRNA之外，本文所述的一些CRISPR系统还可以包括激活稳健性酶活性的另外小型RNA，将其称为反式激活RNA(tracrRNA)。此类tracrRNA典型地包括与crRNA杂交的互补区。crRNA-tracrRNA杂交体形成具有效应子的复合物，该复合物导致可编程酶活性的激活。

[0284] 通过针对与crRNA的直接重复部分同源的短序列基序搜索侧接CRISPR阵列的基因组序列来鉴定tracrRNA序列。搜索方法包括对完整直接重复(DR)或DR子序列的精确或简并匹配。例如，可以将长度n个核苷酸的DR分解成一组重叠6-10nt k聚体。可以将这些k聚体与侧接CRISPR基因座的序列比对，并且可以将具有1或更多个k聚体对齐的同源性的区域鉴定为DR同源区以用于作为tracrRNA的实验验证。替代性地，可以针对完整DR或DR子序列以及来自侧接CRISPR系统的元件的基因组序列的短k聚体序列计算RNA共折叠自由能。可以将具有低最小自由能结构的侧接序列元件鉴定为DR同源区以用于作为tracrRNA的实验验证。

[0285] tracrRNA元件常常很靠近CRISPR相关基因或CRISPR阵列出现。作为搜索DR同源区以鉴定tracrRNA元件的替代方案，可以通过克隆或基因合成来分离侧接CRISPR相关蛋白或CRISPR阵列的非编码序列以用于tracrRNA的直接实验验证。

[0286] 可以使用CRISPR系统的宿主生物的小型RNA测序或在非天然物种中异源性表达的合成序列进行tracrRNA元件的实验验证。可以使用小型RNA序列自起源基因组基因座的比对来鉴定含有DR同源区和完整tracrRNA元件的典型模式化加工的表达RNA产物。

[0287] 通过RNA测序鉴定的完整tracrRNA候选物可以通过以下方式来进行体外或体内验证：在与tracrRNA候选物组合或不组合的情况下表达crRNA和效应子，并且监测效应子酶活性的激活。

[0288] 在工程化构建体中，tracrRNA的表达可以由启动子驱动，这些启动子包括但不限于用于在哺乳动物细胞中表达的U6、U1和H1启动子或用于在细菌中表达的J23119启动子。

[0289] 在一些情况下，可以将tracrRNA与crRNA融合并且表达为单指导RNA。

[0290] 实例3-对酶活性的新型RNA调节子的鉴定

[0291] 除了效应子蛋白和crRNA之外，本文所述的一些CRISPR系统还可以包括激活或调节效应子活性的另外小型RNA，将其在本文中称为RNA调节子。

[0292] 预期RNA调节子很靠近CRISPR相关基因或CRISPR阵列出现。为了鉴定和验证RNA调节子,可以通过克隆或基因合成来分离侧接CRISPR相关蛋白或CRISPR阵列的非编码序列以用于直接实验验证。

[0293] 可以使用CRISPR系统的宿主生物的小型RNA测序或在非天然物种中异源性表达的合成序列进行RNA调节子的实验验证。可以使用小型RNA序列与起源基因组基因座的比对来鉴定含有DR同源区和模式化加工的表达RNA产物。

[0294] 通过RNA测序鉴定的候选RNA调节子可以通过以下方式来进行体外或体内验证:在与候选RNA调节子组合或不组合的情况下表达crRNA和效应子,并且监测效应子酶活性的改变。

[0295] 在工程化构建体中,RNA调节子可以由启动子驱动,这些启动子包括用于在哺乳动物细胞中表达的U6、U1和H1启动子或用于在细菌中表达的J23119启动子。

[0296] 在一些情况下,可以将RNA调节子与crRNA、tracrRNA或两者人工融合并且表达为单RNA元件。

[0297] 实例4-工程化CLUST.121143 CRISPR-Cas系统的功能验证(图5-8)

[0298] 鉴定了CLUST.121143 CRISPR-Cas系统的最小组分,我们从命名为3300014839的宏基因组源选择一个基因座用于功能验证。

[0299] DNA合成和效应子文库克隆

[0300] 为了测试示例性CLUST.121143 CRISPR-Cas系统的活性,我们设计并合成含有pET28a(+)载体的系统。简而言之,合成编码CLUST.121143 3300014839效应子(表2中提供的氨基酸序列)的大肠杆菌密码子优化的核酸序列(金斯瑞公司(Genscript))并且将其克隆至来源于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.121143 3300014839效应子蛋白的核酸和大肠杆菌核糖体结合序列。载体还包含在CLUST.121143 3300014839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。

[0301] 我们计算设计了含有“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质粒以及大肠杆菌必需基因的序列。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。

[0302] 我们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,我们首先使用独特的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。我们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。

[0303] 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Cas蛋白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,我们从转化前"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。

[0304] 细菌筛选测序分析

[0305] 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。

[0306] 为了鉴定导致酶活性和细菌细胞死亡的特定参数,我们使用下一代测序(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。我们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0.1(超过10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,我们取给定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)。我们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。我们研究了此矩阵中不同特征解释CLUST.121143系统靶标耗减的程度,从而得到了单一筛选内功能参数的广泛调查。

[0307] 图5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。在每种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。

[0308] 图6示出了阴性对照的干扰活性的缺乏,其中在电穿孔至大肠杆菌中前已经将CLUST.121143效应子蛋白的编码区从pET-28a(+)来源的质粒中缺失。

[0309] 图7A-B描绘了靶向pACYC184和大肠杆菌E.Cloni必需基因的CLUST.121143 3300014839效应子的强耗减靶标的位置。值得注意的是,强耗减靶标的位置似乎分散在整个潜在靶标空间。

[0310] 图8描绘了侧接耗减靶标的序列的weblogo，其指示出TTG的突出5’PAM。

[0311] 实例5-对CLUST.196682 CRISPR-Cas系统的最小组分的鉴定(图9-12)

[0312] 此蛋白质家族描述了与在Fervidibacteria、盐盒菌属(Haloarcula)、盐环境杆菌属(Halohasta)、盐红菌属(Halorubrum)、盐碱球菌属(Natronoccus)物种以及从淡水、热泉、盐湖、沉积物、土壤和废水环境收集的未培养宏基因组序列中发现的CRISPR系统相关联的大型单效应子(表7)。CLUST.196682效应子包括下表7和8中详述的蛋白质实例。这些系统的直接重复序列的实例示出于表9中。

[0313] • 含有此效应子复合物的天然存在的基因座的实例描绘于图9中，该图表明对于CLUST.196682基因座，效应子蛋白与CRISPR阵列共现。在双向15kb窗口内没有鉴定出与效应子蛋白或CRISPR阵列共现的其他大蛋白质家族。

[0314] • CLUST.196682的直接重复序列展现出多个保守基序，包括靠近5’端的5’-GYYYGNKRKNNAC-3’(SEQ ID NO:832)、或中心的5’-YCCRCGCGCGCGNGGG-3’(SEQ ID NO:833)，其中Y是指C或T或U，K是指G或T或U，R是指A或G，并且N是指任何核碱基(图10)。

[0315] • 图11是实例CLUST.196682效应子的系统发育树的示意图，该示意图显示该家族展现出序列多样性。

[0316] • 图12是CLUST.196682效应子蛋白的多序列比对的示意图，该示意图揭示出由颜色品红色、蓝色和红色指示的RuvC结构域的保守催化残基的位置。

[0317] • 任选地，该系统包括作为表10中列出的非编码序列的子集的tracrRNA。

[0318] 表7.代表性CLUST.196682效应子蛋白

[0319]

物种	效应子登录号	间隔子#	cas1	cas2	效应子大小
废水-厌氧消化器-厌氧消化器污泥（3300025638|Ga0208198_1021643）	3300025638|Ga0208198_1021643_2|P	2	N	N	481
水生-温泉-热泉（3300005300|Ga0072500_136962）	3300005300|Ga0072500_136962_4|M	33	N	N	512
激活的污泥宏基因组（CAAACJ010042499）	CAAACJ010042499_2|P	2	N	N	481
激活的污泥宏基因组（SRR5007116）	SRR5007116_megahit_k177_300528_2|M	6	N	N	481
激活的污泥宏基因组（SRR5007140）	SRR5007140_megahit_k177_707372_2|P	7	N	N	481
激活的污泥宏基因组（SRR5007457）	SRR5007457_megahit_k177_513017_2|M	13	N	N	481
激活的污泥宏基因组（SRR5007461）	SRR5007461_megahit_k177_311051_7|P	4	N	N	419
水生-淡水-淡水湖沉积物（3300027896|Ga0209777_10000273）	3300027896|Ga0209777_10000273_30|P	2	N	N	341
水生-淡水-淡水湖沉积物（3300027896|Ga0209777_10000273）	3300027896|Ga0209777_10000273_32|M	2	N	N	353
水生-淡水-沉积物（3300013127|Ga0172365_10030480）	3300013127|Ga0172365_10030480_1|M	3	N	N	373
水生-淡水-沉积物（3300013128|Ga0172366_10002984）	3300013128|Ga0172366_10002984_1|M	3	N	N	373
水生-淡水-沉积物（3300013129|Ga0172364_10030775）	3300013129|Ga0172364_10030775_1|P	3	N	N	373
水生-淡水-沉积物（3300013130|Ga0172363_10005240）	3300013130|Ga0172363_10005240_1|P	3	N	N	373
水生-淡水-沉积物（3300013133|Ga0172362_10039269）	3300013133|Ga0172362_10039269_3|M	3	N	N	373
水生-淡水-沉积物（3300032070|Ga0315279_10000130）	3300032070|Ga0315279_10000130_51|M	8	Y	N	342
水生-淡水-沉积物（3300032118|Ga0315277_10001250）	3300032118|Ga0315277_10001250_38|M	2	N	N	368
水生-海生-深部地下（3300024429|Ga0209991_10029076）	3300024429|Ga0209991_10029076_1|P	2	N	N	343
水生-海生-海生沉积物（3300014886|Ga0180300_10007429）	3300014886|Ga0180300_10007429_2|M	2	Y	Y	365
水生-非海生盐水和碱性-超盐水（2140918017|Ga0065697_1000127）	SEJ25361.1	2	N	N	382
水生-非海生盐水和碱性-盐水湖（3300005935|Ga0075125_10033396）	3300005935|Ga0075125_10033396_1|P	5	N	N	432
水生-非海生盐水和碱性-盐水湖（3300025736|Ga0207997_1027507）	3300025736|Ga0207997_1027507_2|M	2	N	N	432
水生-温泉（3300007203|Ga0099832_1115789）	3300007203|Ga0099832_1115789_2|M	5	N	N	521
水生-温泉-热泉（3300000083|Ga0072500_119562）	3300000083|Ga0072500_119562_4|M	5	N	N	546
水生-温泉-热泉（3300008000|Ga0105162_1004089）	3300008000|Ga0105162_1004089_4|M	7	N	N	543

[0320]

水生-温泉-热泉（3300008000|Ga0105162_1004089）	3300008000|Ga0105162_1004089_4|P	7	N	N	530
水生-温泉-热泉（3300022544|Ga0212120_1002806）	3300022544|Ga0212120_1002806_5|M	7	N	N	543
水生-温泉-热泉（3300022544|Ga0212120_1002806）	3300022544|Ga0212120_1002806_5|P	7	N	N	530
水生-温泉-热泉（3300025063|Ga0209493_1003385）	3300025063|Ga0209493_1003385_3|M	16	N	N	512
水生-温泉-热泉（3300025090|Ga0209080_1004127）	3300025090|Ga0209080_1004127_5|M	7	N	N	490
水生-温泉-热泉和微生物垫飘带（3300025440|Ga0209741_1005495）	3300025440|Ga0209741_1005495_1|M	2	N	N	521
水生-温泉-沉积物（3300028675|Ga0272445_1001731）	3300028675|Ga0272445_1001731_11|M	6	N	N	516
水生-温泉-沉积物（3300029977|Ga0272449_1037372）	3300029977|Ga0272449_1037372_1|P	3	N	N	516
发酵宏基因组（SRR1562008）	SRR1562008_megahit_k177_216982_5|P	2	N	N	417
发酵宏基因组（UOPD01011681）	UOPD01011681_4|M	2	N	N	417
Fervidibacteria 细菌 JGI MDM2（2645728127|Ga0097870_126）	2645728127|Ga0097870_126_5|M	20	N	N	512
淡水宏基因组（SRR5272259）	SRR5272259_megahit_k177_4486292_121|P	2	N	N	341
淡水宏基因组（SRR5272259）	SRR5272259_megahit_k177_4486292_118|M	2	N	N	353
淡水宏基因组（SRR5273149）	SRR5273149_megahit_k177_25824_1|M	14	N	N	514
淡水宏基因组（SRR5273182）	SRR5273182_megahit_k177_57442_1|M	15	N	N	514
淡水宏基因组（SRR5273183）	SRR5273183_megahit_k177_82677_3|M	3	N	N	454
淡水沉积物宏基因组（SRR6481359）	SRR6481359_megahit_k177_932521_3|M	3	N	N	373
真菌-菌丝-切叶（attine）蚂蚁真菌园（3300012677|Ga0153928_1011778）	3300012677|Ga0153928_1011778_2|M	3	N	N	332
方形盐盒菌（Haloarcula quadrata）（NZ_RBWW01000003）	WP_121304574.1	2	Y	N	384
利氏盐环境杆菌（Halohasta litchfieldiae）tADL（2775507274|2775622397）	SEJ25361.1	2	N	N	382
盐红菌属物种阿特里特（Atlit）-26R（NZ_QPLT01000012）	WP_121597521.1	2	N	N	384
特班齐赫盐红菌（Halorubrum tebenquichense）DSM 14210（AOJD01000077）	WP_006630662.1	4	N	N	384
热泉宏基因组（DRR082460）	DRR082460_megahit_k177_106880|P	7	N	N	535
热泉宏基因组（DRR082460）	DRR082460_megahit_k177_106880|M	7	N	N	537
热泉宏基因组（OGCH01000195）	OGCH01000195_3|M	3	N	N	546
热泉宏基因组（SRR4027810）	SRR4027810_megahit_k177_26210_1|M	17	N	N	512
热泉宏基因组（SRR5247126）	SRR5247126_megahit_k177_220968_4|P	7	N	N	532
实验室富集限定的培养基-有氧富集培养基（3300027941|Ga0207719_100292）	3300027941|Ga0207719_100292_34|M	2	N	N	384
实验室富集限定的培养基-有氧富集培	3300027941|Ga0207719_100292_35|P	2	N	N	376

[0321]

养基(3300027941|Ga0207719_100292)					
填埋物宏基因组（SRR6484323）	SRR6484323_megahit_k177_106899_4|P	3	N	N	429
填埋物宏基因组（SRR6484325）	SRR6484325_megahit_k177_1345066_6|P	2	N	N	429
填埋物宏基因组（SRR6484787）	SRR6484787_megahit_k177_1514834_8|M	4	N	N	468
填埋物宏基因组（SRR8542065）	SRR8542065_megahit_k177_1115558_2|M	3	N	N	429
填埋物宏基因组（SRR8542065）	SRR8542065_megahit_k177_821416_8|M	4	N	N	440
宏基因组（FZPS01000148）	FZPS01000148_2153|M	2	N	N	417
解淀粉嗜盐碱球菌（Natronococcus amylolyticus）DSM 10524（AOIB01000014）	WP_005554257.1	2	N	N	389
泥炭宏基因组（SRR7004341）	SRR7004341_megahit_k177_2665443_5|P	2	N	N	363
盐湖宏基因组（SRR8526178）	SRR8526178_megahit_k177_763700_3|P	5	N	N	432
沉积物宏基因组（OVXO01016275）	OVXO01016275_2|M	3	N	N	358
沉积物宏基因组（SRR7067823）	SRR7067823_megahit_k177_107253_1|P	7	N	N	512
沉积物宏基因组（SRR7067986）	SRR7067986_megahit_k177_69094_7|P	45	N	N	512
沉积物宏基因组（SRR8554444）	SRR8554444_megahit_k177_504614_2|M	2	N	N	351
沉积物宏基因组（SRR8554446）	SRR8554446_megahit_k177_83048_1|M	8	N	N	516
土壤宏基因组（SRR6201998）	SRR6201998_megahit_k177_1568738_2|M	4	N	N	359
土壤宏基因组（SRR8554511）	SRR8554511_megahit_k177_4248910_4|P	3	N	N	457
土壤宏基因组（SRR8554511）	SRR8554511_megahit_k177_4248910_4|M	3	N	N	437
固体废物-填埋物-填埋物沥滤液（3300014204|Ga0172381_10015437）	3300014204|Ga0172381_10015437_8|M	4	N	N	468
固体废物-填埋物-填埋物沥滤液（3300014205|Ga0172380_10102486）	3300014205|Ga0172380_10102486_2|M	2	N	N	429
陆生-土壤（3300031566|Ga0307378_10048523）	3300031566|Ga0307378_10048523_2|M	2	N	N	419
废水宏基因组（SRR6837575）	SRR6837575_megahit_k177_854481_7|M	5	N	N	469
废水宏基因组（SRR6837576）	SRR6837576_megahit_k177_1219670_1|P	5	N	N	469
废水-厌氧消化器-厌氧消化器污泥（3300009716|Ga0116191_1022210）	3300009716|Ga0116191_1022210_3|P	4	N	N	419

[0322] 表8.代表性CLUST.196682效应子蛋白的氨基酸序列

[0323]

>3300025638|Ga0208198_1021643_2|P
[废水-厌氧消化器-厌氧消化器污泥]
MKATTITKSLRFWIDEPPERCRLLYGVSDELTDAYNAILEYWETDVRKVAVEATEAAIEAYHVWKDAPKETRGDKPKWWSADTA
TRAAFLRVGVAATSTIRTRLVDNMVQGEWLRDVKTYMSRRFKDHIAAGIRHRAKGVVISLDQSTMKPDVQLGDDGRWTIRCALW
AQGTEESARNNDRWLLSPYDRKRQTWTIANLLEADQHGEVRGVKLVPPKPGAPAGKRRWSAMITVTLPVESSVLDEERPNVAGV
DMGLTHFAVYSCPARNTFEFVSSRELQAAMEKARRRRRGVPRKRSTALGQKLGRRQDALCRLTARRLIDCCRRDRVGTLRVEDL
TGIRDQGSDDADRNFALGARFPYYKLQTYLEQAAASAGVRLEKVQPAGTSQTCSRCAVRDPESRDGKRFVCRHCGYKGDADLNA
ANNIASGRFRRSIRPSTQPAGRIPSAAPDCTGEAVKASDVDAELQPARTPAACGGDPYAGE (SEQ ID NO: 601)
>3300005300|Ga0072500_136962_4|M
[水生-温泉-热泉]
MEKVILTITSGVEFLQPQDEQLVIKAAKWYTEARKALLYNALQKGEEGVKQAASTLSTLNFPPAPHSRLPDLLTRDVGGTLQAL
KERREGYAQDFSRWWEEICRVANEKGSQAVSVLKPHVNKLSSLIHQLSWNSPEWDEEAQEVFKELKQIWAVVRQKDHERLKIKW
RAEVPPVPRPPRTPKFQRLVFNADGRVSGGSRSNALQIEEREDGDLYLRLQLRDPERVAKKGSYQDVWLRLVIEDEERKQMLKN

[0324]

FAPKYLQLRQDGKRWIAHIVIERENEFEVPQYVYGVDIGARKYLLAAVAIPLNPALPKPPNIAVPSSKFWQIVERVEAEWRKWQ
QLRTGKRDKEGRVVLKKRKMAWRVLKRKRRMRREVTKSMIAEAVNTFIKSLERPCAIVIEQIKNLHLSPTLQDRKSRWRQRRWA
YRFLLEMLRYKAEWEGIRVIEVNPAYTSQTCPRCGKRDRNARQGNLFQCTKCRFQHNADFVAAYNITVRGLMKLRTRGKVDWKV
LTGKPRVL (SEQ ID NO: 602)
>CAAACJ010042499_2|P
[激活的污泥宏基因组]
MKATTITKSLRFWIDEPPERCRLLYGVSDELTDAYNAILEYWETDVRKVAVEATEAAIEAYHVWKDAPKETRGDKPKWWSADTA
TRAAFLRVGVAATSTIRTRLVDNMVQGEWLRDVKTYMSRRFKDHIAAGIRHRAKGVVISLDQSTMKPDVQLGDDGRWTIRCALW
AQGTEESARNNDRWLLSPYDRKRQTWTIANLLEADQHGEVRGVKLVPPKPGAPAGKRRWSAMITVTLPVESSVLDEERPNVAGV
DMGLTHFAVYSCPARNTFEFVSSRELQAAMEKARRRRRGVPRKRSTALGQKLGRRQDALCRLTARRIIDCCRRDRVGTLRVEDL
TGIRDQGSDDADRNFALGARFPYYKLQTYLEQAAASAGVRLEKVQPAGTSQTCSRCAVRDPESRDGKRFECRHCGYKGDADLNA
ANNIASGRFRRSIRPSTQPAGRIPSAAPDCTGEAVKASDVDAELQPARTPAACGGDPYAGE (SEQ ID NO: 603)
>SRR5007116_megahit_k177_300528_2|M
[激活的污泥宏基因组]
MKATTTTKSLRFWIDEPPERCRLLYGVSDELTAAYNAILEYWETDVRSVAVEVTEAAQAAYHVWKDAPKETRGDKPKWWSADTA
TRAAFLRVGVAATSTIRTRLVDNMVQGEWLRDVKTYMSRRFKDHIAAGIRHRAKGVVISLDQSTMKPDVQLGDDGRWTIRCALW
AQGTEESARNNDRWRLSPYDRRRQTWTIANLLEADQHGEVRGVKLVPPKPGAPAGKRRWSAMITVTLPVESSVLDEERPNVAGV
DMGLTHFAVYSCPARNTFEFVSSRELQAAMEKARRRRRGVPRKRSTALGQKLGRRQDALCRLTARRLIDCCLRDRVGTLRVEDL
TGIRDQGSDDPDRNFALGARFPYYKLQTYLEQAAASAGVRLEKVQPAGTSQTCSRCAVRDPESRDGKRFVCRHCGYKGDADLNA
ANNIASGRFRRSIRPSTQPAGRIPSAAPDCTGEAVEASNVGAELQPAREPVAGGGDPYAGE (SEQ ID NO: 604)
>SRR5007140_megahit_k177_707372_2|P
[激活的污泥宏基因组]
MKATTITKSLRFWIDEPPERCRLLYGVSDELTDAYNAILEYWETDVRKVAVEATEAAIEAYHVWKDAPKETRGDKPKWWSADTA
TRAAFLRVGVAATSTIRTRLVDNMVQGEWLRDVKTYMSRRFKDHIAAGIRHRAKGVVISLDQSTMKPDVQLGDDGRWTIRCALW
AQGTEESARNNDRWLLSPYDRKRQTWTIANLLEADQHGEVRGVKLVPPKPGAPAGKRRWSAMITVTLPVESSVLDEERPNVAGV
DMGLTHFAVYSCPARNTFEFVSSRELQAAMEKARRRRRGVPRKRSTALGQKLGRRQDALCRLTARRIIDCCRRDRVGTLRVEDL
TGIRDQGSDDADRNFALGARFPYYKLQTYLEQAAASAGVRLEKVQPAGTSQTCSRCAVRDPESRDGKRFECRHCGYKGDADLNA
ANNIASGRFRRSIRPSTQPAGRIPSAAPDCTGEAVKASDVDAELQPARTPAACGGDPYAGE (SEQ ID NO: 603)
>SRR5007457_megahit_k177_513017_2|M
[激活的污泥宏基因组]
MKATTITKSLRFWIDEPPERCRLLYGVSDELTDAYNAILEYWETDVRKVAVEATEAAIEAYHVWKDAPKETRGDKPKWWSADTA
TRAAFLRVGVAATSTIRTRLVDNMVQGEWLRDVKTYMSRRFKDHIAAGIRHRAKGVVISLDQSTMKPDVQLGDDGRWTIRCALW
AQGTEESARNNDRWLLSPYDRKRQTWTIANLLEADQHGEVRGVKLVPPKPGAPAGKRRWSAMITVTLPVESSVLDEERPNVAGV
DMGITHFAVYSCPARNTFEFVSSRELQAAMEKARRRRRGVPRKRSTALGQKLGRRQDALCRLTARRIIDCCRRDRVGTLRVEDL
TGIRDQGSDDADRNFALGARFPYYKLQTYLEQAAASAGVRLEKVQPAGTSQTCSRCAVRDPESRDGKRFECRHCGYKGDADLNA
ANNIASGRFRRSIRPSTQPAGRIPSAAPDCTGEAVKASDVDAELQPARTPAACGGDPYAGE (SEQ ID NO: 605)
>SRR5007461_megahit_k177_311051_7|P
[激活的污泥宏基因组]
MIKTVTFYIAEMPREDKRLCDRLCIDLGAAGARMVREWFRACDETPPGAAIDWPEVQRRGVAVGKAMVEAAGYRTELGNDLARA
EFFARFKMEFKKVLKGERASLPFPGRAPFIFLRMDPGHETAIIEQRDDAWYMRNPRFYPGGDSNEANRREWRLKPVDRKGWAEN
ILEAAEKFACVRIMPDRKKPGSWVVKITVHLPDALERKLVEEHECWAGIDLGINCPAVLSIPDRDKGGFVRFFGMDKYTALWAK
LNRYEERKRALNRVGQQHAARDLRAKITGLRQYINEQISREIIDLCLRMHVTGIRMEDLKGLRTEGTGKLKYWPRFHLTTRIEQ
KAREVGLAFEQIRAAGTSQTCSVCGHRHRDNRHGSDFLCLRCGFARHADVNAANNIARGMQDFDLPAGSAAPGRALRLELSAQ
(SEQ ID NO: 606)
>3300027896|Ga0209777_10000273_30|P
[水生-淡水-淡水湖沉积物]
MTATKQQILQTEYDNLQHFLQTGEDLGVYSANKQQAKRYYKTIKPEHQYPLSIRNDLIQVQHKPNSIAEYWLKIPVKGRRGGLW

[0325]

VAIRPHQQIPDNAEFLESKIRRKRKGWFANIVIQKQIEPIRTKNILAVDLGEHVIATVCGSFDNQRPRFYGRDVRGIRRHYAWL
RKRLGEKKLLGVIRRIKDTEKRKVNDTLHKISRSIVDLAQEHNATIVLGELQGIRNHAKGRRMNRIVSNMPYLKLSQMIEYKAE
WLGIPVIKMQEHDTSKTCSRCNEYGKRPTQGQFRCDTCHRTSFNADYNGALNILKRATSYMDIAGAVSEPARDWGNDDQTSEAT
QLFEW (SEQ ID NO: 607)
>3300027896|Ga0209777_10000273_32|M
[水生-淡水-淡水湖沉积物]
MAQKTITAKIVAMTATKQQILQTEYDNLQHFLQTGEDLGVYSANKQQAKRYYKTIKPEHQYPLSIRNDLIQVQHKPNSIAEYWL
KIPVKGRRGGLWVAIRPHQQIPDNAEFLESKIRRKRKGWFANIVIQKQIEPIRTKNILAVDLGEHVIATVCGSFDNQRPRFYGR
DVRGIRRHYAWLRKRLGEKKLLGVIRRIKDTEKRKVNDTLHKISRSIVDLAQEHNATIVLGELQGIRNHAKGRRMNRIVSNMPY
LKLSQMIEYKAEWLGIPVIKMQEHDTSKTCSRCNEYGKRPTQGQFRCDTCHRTSFNADYNGALNILKRATSYMDIAGAVSEPAR
DWGNDDQTSEATQLFEW (SEQ ID NO: 608)
>3300013127|Ga0172365_10030480_1|M
[水生-淡水-沉积物]
MANRELVVVMQVQKVVKGKVVGLTHAKRQILDHEYEGLQRFLQGDSNVELYSANKQQAKRFYKTVKPEREYPLSIRKDLIRVER
RDTKIARYWARIPAKLRRGGVWVAVKPHCEITDDLEICESKLLRRNGEYFLHLAVKREVELNTSYSSVLAIDLGIRHVACSVDT
RSGETHFHGRELRRVRGHYFHLRRRLGEKKLLKVIKRIGARESRITNDQLHKIAKAIVDEAERLNALIAIGNPKGIRRNGKGRR
FNRKLNSWPFWKLRQYIEYKANWRGIAVVEVNEAYTSQRCWRCGTIGTRSGTHHGLFECPRCGLRENADRNGAFNIGRRALGQV
SKVGAVVSQPVTGAVPFSELGSVVQCPTPEATLFKGW (SEQ ID NO: 609)
>3300013128|Ga0172366_10002984_1|M
[水生-淡水-沉积物]
MANRELVVVMQVQKVVKAKVVGLTHAKRQILDHEYEGLQRFLQGDSNVELYSANKQQAKRFYKTVKPEREYPLSIRKDLIRVER
RDTKIARYWARIPAKLRRGGVWVAVKPHCEITDDLEICESKLLRRNGEYFLHLAVKREVELNTSYSSVLAIDLGIRHVACSVDT
RSGETHFHGRELRRVRGHYFHLRRRLGEKKLLKVIKRIGARESRITNDQLHKIAKAIVDEAERLNALIAIGNPKGIRRNGKGRR
FNRKLNSWPFWKLRQYIEYKANWRGIAVVEVNEAYTSQRCWRCGTIGTRSGTHHGLFECPRCGLRENADRNGAFNIGRRALGQV
SKVGAVVSQPVTGAVPFSELGSVVQCPTPEATLFKGW (SEQ ID NO: 610)
>3300013129|Ga0172364_10030775_1|P
[水生-淡水-沉积物]
MANRELVVVMQVQKVVKGKVVGLTHAKRQILDHEYEGLQRFLQGDSNVELYSANKQQAKRFYKTVKPEREYPLSIRKDLIRVER
RDTKIARYWARIPAKLRRGGVWVAVKPHCEITDDLEICESKLLRRNGEYFLHLAVKREVELNTSYSSVLAIDLGIRHVACSVDT
RSGETHFHGRELRRVRGHYFHLRRRLGEKKLLKVIKRIGARESRITNDQLHKIAKAIVDEAERLNALIAIGNPKGIRRNGKGRR
FNRKLNSWPFWKLRQYIEYKANWRGIAVVEVNEAYTSQRCWRCGTIGTRSGTHHGLFECPRCGLRENADRNGAFNIGRRALGQV
SKVGAVVSQPVTGAVPFSELGSVVQCPTPEATLFKGW (SEQ ID NO: 609)
>3300013130|Ga0172363_10005240_1|P
[水生-淡水-沉积物]
MANRELVVVMQVQKVVKGKVVGLTHAKRQILDHEYEGLQRFLQGDSNVELYSANKQQAKRFYKTVKPEREYPLSIRKDLIRVER
RDTKIARYWARIPAKLRRGGVWVAVKPHCEITDDLEICESKLLRRNGEYFLHLAVKREVELNTSYSSVLAIDLGIRHVACSVDT
RSGETHFHGRELRRVRGHYFHLRRRLGEKKLLKVIKRIGARESRITNDQLHKIAKAIVDEAERLNALIAIGNPKGIRRNGKGRR
FNRKLNSWPFWKLRQYIEYKANWRGIAVVEVNEAYTSQRCWRCGTIGTRSGTHHGLFECPRCGLRENADRNGAFNIGRRALGQV
SKVGAVVSQPVTGAVPFSELGSVVQCPTPEATLFKGW (SEQ ID NO: 609)
>3300013133|Ga0172362_10039269_3|M
[水生-淡水-沉积物]
MANRELVVVMQVQKVVKGKVVGLTHAKRQILDHEYEGLQRFLQGDSNVELYSANKQQAKRFYKTVKPEREYPLSIRKDLIRVER
RDTKIARYWARIPAKLRRGGVWVAVKPHCEITDDLEICESKLLRRNGEYFLHLAVKREVELNTSYSSVLAIDLGIRHVACSVDT
RSGETHFHGRELRRVRGHYFHLRRRLGEKKLLKVIKRIGARESRITNDQLHKIAKAIVDEAERLNALIAIGNPKGIRRNGKGRR
FNRKLNSWPFWKLRQYIEYKANWRGIAVVEVNEAYTSQRCWRCGTIGTRSGTHHGLFECPRCGLRENADRNGAFNIGRRALGQV
SKVGAVVSQPVTGAVPFSELGSVVQCPTPEATLFKGW (SEQ ID NO: 609)
>3300032070|Ga0315279_10000130_51|M

[0326]

[水生-淡水-沉积物]
MILQKTVNAKLFALTRTKEQKISREYWNFQKALHGEDAPLYSATKQQAKRKNAWNGDCEYPLVLRRDCFGVKELGLKFARWWAK
IPVFGGSIWCPVQLPYVQEPLLKGNIRETKLVRRKKGWFLRITVQQEIELDVPCASVVGVDFGERHIATSVEWQGRFQNPIFYG
AEVRGIRRHYAWLRKRLGNKKALRTIKKVGGKENRKVNAVLHNISRAIVDRAKELSAVILLGAPNGGSMRRKALSKGKRFTRIV
YSMPYYRLRQFIEYKAVQEGILVVNANEEWTSKTCSFCGEISDRPRQDLIVCPHGHRLNADINACRNLIKRFIDHWLIDGAILA
SPKLPH (SEQ ID NO: 611)
>3300032118|Ga0315277_10001250_38|M
[水生-淡水-沉积物]
MKSQKIIKCKITYLTNIKAQQLETEHHNLQNLLQLEREGLDFLPIYQNIPLHSANKQQALRFYKRISTDKIYAISIRNDIIKVE
LKPNNKISKYWCKIPVKQRYGGIWVAIKPHKKDIDFSQYKLGESKLFKKKNEWWIYITIIREVQIKDSYSNILGIDLGSKVIAT
VCGSFYNQRPTFYGREVRGIRRHYQYIRTELGKKKLLKKIKQLSDKEHRIVNDLLHKISKQIVEKAEQTDSYIVLGDLKGIRNS
TKNKGRNFRRIVSNMPYLKLTQFIEYKAKEKGIKIIRISEKGTSKICSKCGYEDKGNRKSQGLFKCKQCGFELNADANGARNIL
KFSDGYTLPERAMSELALNSNQLEKPTCFSGG (SEQ ID NO: 612)
>3300024429|Ga0209991_10029076_1|P
[水生-海生-深部地下]
MLVEKTVKAKVIAPTNRKQELLDREYQSFQYTLYGGEANLYSATKQQAERFLRKIKTPKHRHYPLILRRDVIRLERQETKVAPY
WFKFPVFGVRGGIWLGLKPHCDIPVDCSLREAKLIRRKNAWFIHLTIQKEVPESIIDEHHILAVDLGERYLATVCGNNGTNPQF
LGKEVRGIRRHYAWIRKRLGERKLLNTTRKVGHTEQKKVNDICHQISRRIVNEARETDSVIILGDLKGIRKNAKGKRMNRIVSN
MPYYKLTQFIKYKAAWDGVPVLVIPEAYTSKGCHRCGSLGSRPYRGLFLCHSCKLHYNADLNGARNILKRALDYMSSAGAALAQ
PVTQSLN (SEQ ID NO: 613)
>3300014886|Ga0180300_10007429_2|M
[水生-海生-海生沉积物]
MKVKKTIKAKILELRKGKEELLKREYENWQRYLRGDKTVPLYSATKQQADRLLKRLKGRLKPNKEYPMILRRDVYRANTKLTPY
WLKISIYGVRGGINVPIKTHEPITEDMVCREAKIIRKGDEWFVYITVEKEVEERNPKSILAIDLGIRWMATTVNSNNPKPNFYG
KELRKVKGHYFYLRRSLTLKKAYKTIKKIGLKERRVVNDILHKISRAIVNEALENDSMIVLGNLKGIRRNGKGRRFNRKLNNGF
PYYRLSQFIEYKAKWLGIKVIKISEKNTSKLCHKCGEEGVRVGSSFKCPNCGYSCHADYNGAMNILKRAMGLVSVAGAGLTQPR
TRYDEARRPKNRESPHLSGKSVKSYSFKF (SEQ ID NO: 614)
>SEJ25361.1
[水生-非海生盐水和碱性-超盐水]
MKQTAEKTVVAKIFTNDLTQSKLDAINHEYEAFQRYIRGDEEADLYSATKQCADAYVDTDNLRDDHDYPWFIRNDVFDVGVDVE
QHDTDLTDWWLKLPVSQVWGGVNVPIQPHDEIPDEADVKDSKIVRVNGDYYAHLTVQQEVEVSEEYGGVIGVDFGVRWVATSVA
LPSRNTEFYGREIRRIRRHHHDLRTRLQEKGAYGTLQSVKVKEYRQVNDRLHKISRAIVEEAKECDALIVVGDLEGIQNQDLGA
EMNRRLHSMPHHTLKTYIEYKAMFAGIAVKSVNEYMTSQTCWRCGEQEATTRKGQGQHLCHECGLDDNADKNGATNIAKKGLGK
DIACPLSSLGAVCEPVLGPSGADQASATVCLRADLEAPASTKRTVR (SEQ ID NO: 615)
>3300005935|Ga0075125_10033396_1|P
[水生-非海生盐水和碱性-盐水湖]
MSLKTLTFWPIGMSAKYRRIFDNIRKQLGAAQREMVKAWLRLALAEDAYGIDPGKRAGREVFGDNEYSNDLLNNLAYADAHENF
KQQWKRVRSGNATSVSFSDAAANVLFLMPPSGGHHATVEHIDGNWWITRLKVFGGCTKADRERNDQFRWQLKGVSRFGGKYEKD
LMDGAQRICQVRVTPSRKDPSKYVVQVTVEVPDLQTVGARTPVCAGIDLGINCPVVLSIPDKGFVQFMGRERDEYPRLWEKLRE
YEDRKRRLNRAGKRRAARDLRPRITNVRKYINEAISKDVADTCRALGVTHLRMEDLKGLSRHKSMTGKLRYWPRYHLQERIERK
CCEVGIEVERIKPAGTSQTCSVCGHREKANRNGAQFVCIECGYQQHADLNAANNIARGMQTFGPCAKSSASDDAIGLRFDSEGQ
ESALGQAKRLFA (SEQ ID NO: 616)
>3300025736|Ga0207997_1027507_2|M
[水生-非海生盐水和碱性-盐水湖]
MSLKTLTFWPIGMSAKYRRIFDNIRKQLGAAQREMVKAWLRLALAEDAYGIDPGKRAGREVFGDNEYSNDLLNNLAYADAHENF
KQQWKRVRSGNATSVSFSDAAANVLFLMPPSGGHHATVEHIDGNWWITRLKVFGGCTKADRERNDQFRWQLKGVSRFGGKYEKD
LMDGAQRICQVRVTPSRKDPSKYVVQVTVEVPDLQTVGARTPVCAGIDLGINCPVVLSIPDKGFVQFMGRERDEYPRLWEKLRE

YEDRKRRLNRAGKRRAARDLRPRITNVRKYINEAISKDVADTCRALGVTHLRMEDLKGLSRHKSMTGKLRYWPRYHLQERIERK
CCEVGIEVERIKPAGTSQTCSVCGHREKANRNGAQFVCIECGYQQHADLNAANNIARGMQTFGPCAKSSASDDAIGLRFDSEGQ
ESALGQAKRLFA (SEQ ID NO: 616)

>3300007203|Ga0099832_1115789_2|M

[水生-温泉]

MKKVVITVTTGVEFLQPQDEQIVRKAAEWYTKAREALFAHALQQGKVKEAASTLRTLNFPMAPHSRLPDLLTRDVGGILKALKE
KREGYLGEFKRWWEEDCRVANEKGSQAVADLKTHVGKLLNLVHQLPWNSPVWDGEAQEAFKNLKEAWAVVRQKDHEQLKIRWRA
EVSPLRRPPRTPKFQRLVFNADGRVSGGSKSEALQIEEREDGDLYLRLQLRDPEGVVGGGAYQEVWLRLVVEDEERKQMLRDFA
PKYLQLRQDGKRWIAHIVIEREIKFGLKGIMWTAGDYLYGVDIGARKYLLAAVAVPLKPGLPKLPNIAIPSGKFWQAVERVEAE
WRKWQRLRTGKKDKEGKVVLKKRTAWRVLKRKRRMRREVAKNMIAEAVNTFINFIKGDNVDIRLWLAIEQIKNLNLSPAFQDRK
SRWRQRRWAYRLLLEMLRYKAEWEGIRVIEVNPAHTSQICPRCGKKDGKARQGNRFCCTKCKFQHNADFVAAYNIAVRALKKLG
MREEIGWKVLTGKPRVL (SEQ ID NO: 617)

>3300000083|Ga0072500_119562_4|M

[水生-温泉-热泉]

MKHLVSNAQVNEKNEARLTLKCGLEFENAEGEELVKRAAEWYCKARIFLLSYALKNNKTRVQKIASLLKSEGFPSPPHPRLIDL
LQRTITPAIKNMRRLRRRYPKTLRAWFDRVKKFAEEKGSDAAERLQPYVEDLLQLVKNLPINSPAWDTDAQEKWKELKKLWGKV
RESDWKKGIVWKFQIPNPKKPPSTPAFRCADSSNIFVTADYRMSGKNARNSYDMKTDSEENLWVLVRLDDVDGDVKVRKDVWIK
LCVDKSRYDLLRIGKITEIKILRNRDGRWEGRIVVKMNAGYNSPQVVVGVDVGASEHLIAAVGIPLDGSNNNFLLSVPASYFWK
RVEQAERLWRRWQRLYTGRTSEKKKAPILPRRRKAYQILKRNRNWRREVTKHCIFHAVKKFVQLLPPNSVVVVENLKDIPPNVW
KEDKDKRFRRWVYAFLLFTLKYKCQQRKIAVIAVSPHNTSQTCPRCGVVDPKSRGLWKDGELISRKHFCCTQCKFTHNADFVAA
HNIAVLGAKALGITLPESWRALRGEVQYVIPYQNMPKVRKSW (SEQ ID NO: 618)

>3300008000|Ga0105162_1004089_4|M

[水生-温泉-热泉]

[0327] MNNKQVTFGGLFLMALQKKRVRRTIVAGVEIESAEEENLVTEAAKWYLQAKQKLLQYAIEKRKERITAPARILKDELKLPEPPH
SRLPDELAQVINGVLKALKKRRTEYLRRFQDWWNKVVEVGNQHGLESYRAMKPKVEALISSVPNNPFDPTWDSDIQKRFAEVKE
IWSKIRLKDSNERNFRWRYHVPPLVKPPKTPEFKLLGFDAGARVSGGGKAKTLTIKSQDGEIFLRLLLRQPGVSNHGRGSYREI
WLHIILEDESRKALLLSDKVQAKFIRLRRRTDKWEAHITVEKEVSFEKPRVIYGIDIGMKDALLFCVPIFIDSQETTETASEGK
KLIVVEAGQFWHRLEVEDAQWRKWQFRRSGKVKEGIKGKRMVWRILKRRRRRRREVTKHLIHQIVNALISQLQRPCAVAIERLK
GLRNKPHLTDKRSQWKFHRWAYRFLLETLRYKLEWEGIMIIEVPPYGTSQTCPRCSHRDKRSRDGRRFRCTKCKFQHNADAAGA
FNIAMRAVEQLQSSRDLQEVSIAESAWKILTGKPMVLKY (SEQ ID NO: 619)

>3300008000|Ga0105162_1004089_4|P

[水生-温泉-热泉]

MALQKKRVRRTIVAGVEIESAEEENLVTEAAKWYLQAKQKLLQYAIEKRKERITAPARILKDELKLPEPPHSRLPDELAQVING
VLKALKKRRTEYLRRFQDWWNKVVEVGNQHGLESYRAMKPKVEALISSVPNNPFDPTWDSDIQKRFAEVKEIWSKIRLKDSNER
NFRWRYHVPPLVKPPKTPEFKLLGFDAGARVSGGGKAKTLTIKSQDGEIFLRLLLRQPGVSNHGRGSYREIWLHIILEDESRKA
LLLSDKVQAKFIRLRRRTDKWEAHITVEKEVSFEKPRVIYGIDIGMKDALLFCVPIFIDSQETTETASEGKKLIVVEAGQFWHR
LEVEDAQWRKWQFRRSGKVKEGIKGKRMVWRILKRRRRRRREVTKHLIHQIVNALISQLQRPCAVAIERLKGLRNKPHLTDKRS
QWKFHRWAYRFLLETLRYKLEWEGIMIIEVPPYGTSQTCPRCSHRDKRSRDGRRFRCTKCKFQHNADAAGAFNIAMRAVEQLQS
SRDLQEVSIAESAWKILTGKPMVLKY (SEQ ID NO: 620)

>3300022544|Ga0212120_1002806_5|M

[水生-温泉-热泉]

MNNKQVTFGGLFLMALQKKRVRRTIVAGVEIESAEEENLVTEAAKWYLQAKQKLLQYAIEKRKERITAPARILKDELKLPEPPH
SRLPDELAQVINGVLKALKKRRTEYLRRFQDWWNKVVEVGNQHGLESYRAMKPKVEALISSVPNNPFDPTWDSDIQKRFAEVKE
IWSKIRLKDSNERNFRWRYHVPPLVKPPKTPEFKLLGFDAGARVSGGGKAKTLTIKSQDGEIFLRLLLRQPGVSNHGRGSYREI
WLHIILEDESRKALLLSDKVQAKFIRLRRRTDKWEAHITVEKEVSFEKPRVIYGIDIGMKDALLFCVPIFIDSQETTETASEGK
KLIVVEAGQFWHRLEVEDAQWRKWQFRRSGKVKEGIKGKRMVWRILKRRRRRRREVTKHLIHQIVNALISQLQRPCAVAIERLK
GLRNKPHLTDKRSQWKFHRWAYRFLLETLRYKLEWEGIMIIEVPPYGTSQTCPRCSHRDKRSRDGRRFRCTKCKFQHNADAAGA

FNIAMRAVEQLQSSRDLQEVSIAESAWKILTGKPMVLKY (SEQ ID NO: 619)

>3300022544|Ga0212120_1002806_5|P

[水生-温泉-热泉]

MALQKKRVRRTIVAGVEIESAEEENLVTEAAKWYLQAKQKLLQYAIEKRKERITAPARILKDELKLPEPPHSRLPDELAQVING
VLKALKKRRTEYLRRFQDWWNKVVEVGNQHGLESYRAMKPKVEALISSVPNNPFDPTWDSDIQKRFAEVKEIWSKIRLKDSNER
NFRWRYHVPPLVKPPKTPEFKLLGFDAGARVSGGGKAKTLTIKSQDGEIFLRLLLRQPGVSNHGRGSYREIWLHIILEDESRKA
LLLSDKVQAKFIRLRRRTDKWEAHITVEKEVSFEKPRVIYGIDIGMKDALLFCVPIFIDSQETTETASEGKKLIVVEAGQFWHR
LEVEDAQWRKWQFRRSGKVKEGIKGKRMVWRILKRRRRRRREVTKHLIHQIVNALISQLQRPCAVAIERLKGLRNKPHLTDKRS
QWKFHRWAYRFLLETLRYKLEWEGIMIIEVPPYGTSQTCPRCSHRDKRSRDGRRFRCTKCKFQHNADAAGAFNIAMRAVEQLQS
SRDLQEVSIAESAWKILTGKPMVLKY (SEQ ID NO: 620)

>3300025063|Ga0209493_1003385_3|M

[水生-温泉-热泉]

MEKVILTITSGVEFLQPQDEQLVIKAAKWYTEARKALLYNALQKGEEGVKQAASTLSTLNFPPAPHSRLPDLLTRDVGGTLQAL
KERREGYAQDFSRWWEEICRVANEKGSQAVSVLKPHVNKLSSLIHQLSWNSPEWDEEAQEVFKELKQIWAVVRQKDHERLKIKW
RAEVPPVPRPPRTPKFQRLVFNADGRVSGGSRSNALQIEEREDGDLYLRLQLRDPERVAKKGSYQDVWLRLVIEDEERKQMLKN
FAPKYLQLRQDGKRWIAHIVIERENEFEVPQYVYGVDIGARKYLLAAVAIPLNPALPKPPNIAVPSSKFWQIVERVEAEWRKWQ
QLRTGKRDKEGRVVLKKRKMAWRVLKRKRRMRREVTKSMIAEAVNTFIKSLERPCAIVIEQIKNLHLSPTLQDRKSRWRQRRWA
YRFLLEMLRYKAEWEGMRVIEVNPAYTSQTCPRCGERDRNARQGNLFRCTKCRFQHNADFVAAYNITVRGLMKLRTRGRIDWKI
LTGKPRVL (SEQ ID NO: 621)

>3300025090|Ga0209080_1004127_5|M

[水生-温泉-热泉]

MLQYAIEKRKERITAPARILKDELKLPEPPHSRLPDELAQVINGVLKALKKRRTEYLRRFQDWWNKVVEVGNQHGLESYRAMKP
KVEALISSVPNNPFDPTWDSDIQKRFAEVKEIWSKIRLKDSNERNFRWRYHVPPLVKPPKTPEFKLLGFDAGARVSRGGKAKTL
TIKSQDGEIFLRLLLRQPGVSNHGRGSYREIWLHIILEDESRKALLLSDKVQAKFIRLRRRTDKWEAHITVEKEVSFEKPRVIY
GIDIGMKDALLFCVPIFIDSQETTETASEGKKLIVVEAGQFWHRLEVEDAQWRKWQFRRSGKVKEGIKGKRMVWRILKRRRRRR
REVTKHLIHQIVNALISQLQRPCAVAIERLKGLRNKPHLTDKRSQWKFHRWAYRFLLETLRYKLEWEGIMIIEVPPYGTSQTCP
RCSHRDKRSRDGRRFRCTKCKFQHNADAAGAFNIAMRAVEQLQSSRDLQEVSIAESAWKILTGKPMVLKY (SEQ ID NO: 622)

[0328]

>3300025440|Ga0209741_1005495_1|M

[水生-温泉-热泉和微生物垫飘带]

MEKVVITITTGVEFLQPQDEQIVRKAAEWYTKAREALFDHALQQGKVKKAASTLRTLNLPMAPHSRLPDLLTRDVGGILKALKE
KRKGYLREFKRWWKEVCKVANEKGSQAVAVLKPHVDKLLNLVRQLPWNSPVWDGEAQEAFKNLKEAWAVVRQKDHEQLKIKWRA
EVSPLRRPPRTPKFQRLVFNADGRVSGGSKSEALQIEEREDGDLYLRLQLRDPEGVVGGGAYQEVWLRLVVEDEERKQMLRDFA
PKYLQLRQDGKRWIAHIVIEREIKFGLKGIMWTAGDYLYGVDIGARKYLLAAVAVPLKPGLPKLPNIAIPSGKFWQAVERVEAE
WRKWQRLRTGKKDKEGKVVLKKRTAWRVLKRKRHMRREVAKNMIAEAVNTFINFIKGDNVDIRLWLAIEQIKNLNLSPAFQDRK
SRWRQRRWAYRLLLEMLRYKAEWEGIRVIEVNPAHTSQICPRCGKKDGKARQGNRFCCTECKFQHNADFVAAYNIAVRALKKLG
MREEIGWKVLTGKPRVL (SEQ ID NO: 623)

>3300028675|Ga0272445_1001731_11|M

[水生-温泉-沉积物]

MKKVVITITTGVEFLQPQDEQIVRKAAEWYTKAREALFAHALQQQQGKVEKAASTLRTLNFPMAPHSRLPDLLTRDVGGILKAL
KEKREGYLGEFKRWWEEVCRVANEKGSQAVAALKTHVDKLLNLVHQLPWNSPVWDGEAQEAFKNLKEAWAVVRQKDHEQLKIKW
RAEVSPLRRPPRTPKFQRLVFNADGRVSGGSKSEALQIEEREDGDLYLRLQLRDPEGVVGGGAYQEVWLRLVVEDEERKQMLRD
FAPKYLQLRQDGKRWIAHIVIEREIKFGKPDYLYGVDIGARKYLLAAVAVPLKPDLPKLPNIAIPSGKFWQAVERVEAEWRKWQ
RLRTGKKDKEGRVVLKKRTAWRVLKRKRRMRREVAKNMIAEAVNTFINFIKGDNVDIRLWLAIEQIKNLNLSPAFQDRKSRWRQ
RRWAYRLLLEMLRYKAEWEGIRVIEVNPAHTSQICPRCGKKDGKARQGNRFCCTKCKFQHNADFVAAYNIAVRALKKLGMREEI
GWKVLTGKPKVL (SEQ ID NO: 624)

>3300029977|Ga0272449_1037372_1|P

[水生-温泉-沉积物]

MKKVVITITTGVEFLQPQDEQIVRKAAEWYTKAREALFAHALQQQQGKVEKAASTLRTLNFPMAPHSRLPDLLTRDVGGILKAL
KEKREGYLGEFKRWWEEVCRVANEKGSQAVAALKTHVDKLLNLVHQLPWNSPVWDGEAQEAFKNLKEAWAVVRQKDHEQLKIRW
RAEVSPLRRPPRTPKFQRLVFNADGRVSGGSKSEALQIEEKEDGDLYLRLQLRDPEGIVGGGAYQEVWLRLVVEDEERKQMLRD
FAPKYLQLRQDGERWIAHIVIEREIKFGKPDYLYGVDIGARKYLLAAVAVPLKPDLPKLPNIAIPSGKFWQAVERVEAEWRKWQ
RLRTGKKDKEGRVVLKKRTAWRVLKRKRRMRREVAKNMIAEAVNTFINFIKGDNVDIRLWLAIEQIKNLNLSPAFQDRKSRWRQ
RRWAYRLLLEMLRYKAEWEGIRVIEVNPAHTSQICPRCGKKDGKARQGNRFCCTKCKFQHNADFVAAYNIAVRALKKLGMREEI
GWKVLTGKPKVL (SEQ ID NO: 625)

>SRR1562008_megahit_k177_216982_5|P

[发酵宏基因组]

MIKTATFYLADMPQEDKRLCQLLCDELGAAGSAMVREWYRACDEVDGERINWGDVQARGVQVGKQAFADRGYRGELVNDLQRAE
FFARFKTEFKKIAAGERASLPFPGSAPFIFLRMDPGHEVASVVNEGGTWYLCNPHFYPGGSRSEDNRREWRLKDVGRNGWGYAI
LAEAEKFACIRIMPDRKKRGSWVVKITVHLPDAIARREVEKQEVWAGIDLGLNCPAVLSIPDHPDGGFVRFLGNDQYPRLWKRL
DDYEERKRVLNRAGKREAARDLRAKITGVRQHINELISREIVDLCLAMRVTGLRMEDLKGLGVEGSGRLKFWPRFHLVTRIEQK
CGEVGLTFEKIKPAGTSQTCSSCGHRDKDNRNGPDFVCLRCGFARHADVNAANNIARGMDDFDLPTGASAPTKARRLEIAS
(SEQ ID NO: 626)

>UOPD01011681_4|M

[发酵宏基因组]

MIKTATFYLADMPQEDKRLCQLLCDELGAAGSAMVREWYRACDEVDGERINWGDVQARGVQVGKQAFADRGYRGELVNDLQRAE
FFARFKTEFKKIAAGERASLPFPGSAPFIFLRMDPGHEVASVVNEGGTWYLCNPHFYPGGSRSEDNRREWRLKDVGRNGWGYAI
LAEAEKFACIRIMPDRKKRGSWVVKITVHLPDAIARREVEKQEVWAGIDLGLNCPAVLSIPDHPDGGFVRFLGNDQYPRLWKRL
DDYEERKRVLNRAGKREAARDLRAKITGVRQHINELISREIVDLCLAMRVTGLRMEDLKGLGVEGSGRLKFWPRFHLVTRIEQK
CGEVGLTFEKIKPAGTSQTCSSCGHRDKDNRNGPDFVCLRCGFARHADVNAANNIARGMDDFDLPTGASAPTKARRLEIAS
(SEQ ID NO: 626)

[0329]

>2645728127|Ga0097870_126_5|M

[Fervidibacteria 细菌 JGI MDM2]

MEKVILTITSGVEFLQPQDEQLVIKAAKWYTEARKALLYNALQKGEEGVKQAASTLSTLNFPPAPHSRLPDLLTRDVGGTLQAL
KERREGYAQDFSRWWEEICRVANEKGSQAVSVLKPHVNKLSSLIHQLSWNSPEWDEEAQEVFKELKQIWGVVKQKDHERLKIKW
CAEVPPVPRPPRTPKFQRLVFNADGRVSGGSRSNALQIEEREDGDLYLRLQLRDPERVAKKGSYQDVWLRLVIEDEERKQMLKN
FAPKYLQLRQDGKRWIAHIVIERENEFEVPQYVYGVDIGARKYLLAAVAIPLNPALPKPPNIAVPSSKFWQIVERVEAEWRKWQ
QLRTGKRDKEGRVVLKKRKMAWRVLKRKRRMRREVTKSMIAEAVNTFIKSLERPCAIVIEQIKNLHLSPTLQDRKSRWRQRRWA
YRFLLEMLRYKAEWEGMRVIEVNPAYTSQTCPRCGERDRNARQGNLFRCTKCRFQHNADFVAAYNITVRGLMKLRTRGRIDWKI
LTGKPRVL (SEQ ID NO: 627)

>SRR5272259_megahit_k177_4486292_121|P

[淡水宏基因组]

MTATKQQILQTEYDNLQHFLQTGEDLGVYSANKQQAKRYYKTIKPEHQYPLSIRNDLIQVQHKPNSIAEYWLKIPVKGRRGGLW
VAIRPHQQIPDNAEFLESKIRRKRKGWFANIVIQKQIEPIRTKNILAVDLGEHVIATVCGSFDNQRPRFYGRDVRGIRRHYAWL
RKRLGEKKLLGVIRRIKDTEKRKVNDTLHKISRSIVDLAQEHNATIVLGELQGIRNHAKGRRMNRIVSNMPYLKLSQMIEYKAE
WLGIPVIKMQEHDTSKTCSRCNEYGKRPTQGQFRCDTCHRTSFNADYNGALNILKRATSYMDIAGAVSEPARDWGNDDQTSEAT
QLFEW (SEQ ID NO: 607)

>SRR5272259_megahit_k177_4486292_118|M

[淡水宏基因组]

MAQKTITAKIVAMTATKQQILQTEYDNLQHFLQTGEDLGVYSANKQQAKRYYKTIKPEHQYPLSIRNDLIQVQHKPNSIAEYWL
KIPVKGRRGGLWVAIRPHQQIPDNAEFLESKIRRKRKGWFANIVIQKQIEPIRTKNILAVDLGEHVIATVCGSFDNQRPRFYGR
DVRGIRRHYAWLRKRLGEKKLLGVIRRIKDTEKRKVNDTLHKISRSIVDLAQEHNATIVLGELQGIRNHAKGRRMNRIVSNMPY
LKLSQMIEYKAEWLGIPVIKMQEHDTSKTCSRCNEYGKRPTQGQFRCDTCHRTSFNADYNGALNILKRATSYMDIAGAVSEPAR
DWGNDDQTSEATQLFEW (SEQ ID NO: 608)

>SRR5273149_megahit_k177_25824_1|M

[0330]

[淡水宏基因组]
MEKVVITITTGVEFLQPQDEQIVRKAAEWYTKAREALFAHALQQGKVKKAASTLRTLNLPMAPHSRLPDLLTRDVGGILKALKE
KREGYLREFKRWWEEVCRVANEKGSQAVAALKTHVDKLLNLVHQIPWNSPVWDGEAQEAFKNLKEAWAVVRQKDHEQLKIKWRA
EVSPLRRPPRKPKFQRLVFNADGRVSGGSRAEALQIEEREEGDLYLRLRLRDPEGVVGGGAYQEVWLRLVIEDEGRKQMLRDFA
PKYLQLRQDGKRWIAHIVIEREIKFGEPDYLYGVDIGARKYLLAAVAVPLKPDLPKLPNIAIPSGEFWQAVERVEAEWRKWQRL
RTGKKDKEGRVVLKKKRMAWRVLKRKRRMRREVAKNMIAEAVNTFINFIKRGNVNSGLAIEQIKNLNLSPTLQDRKSRWRQRRW
AYRLLLEMLRYKAEWEGIRVIEVNPTHTSQICPRCRKKDGKARQGNRFCCTKCHFQHNADFVAAYNIAMRGLRKLGMRGEIGWK
VLTGKPRVLQ (SEQ ID NO: 628)
>SRR5273182_megahit_k177_57442_1|M
[淡水宏基因组]
MEKVVITITTGVEFLQPQDEQIVRKAAEWYTKAREALFAHALQQGKVKKAASTLRTLNLPMAPHSRLPDLLTRDVGGILKALKE
KREGYLREFKRWWEEVCRVANEKGSQAVAALKTHVDKLLNLVHQIPWNSPVWDGEAQEAFKNLKEAWAVVRQKDHEQLKIKWRA
EVSPLRRPPRKPKFQRLVFNADGRVSGGSRAEALQIEEREEGDLYLRLRLRDPEGVVGGGAYQEVWLRLVIEDEGRKQMLRDFA
PKYLQLRQDGKRWIAHIVIEREIKFGEPDYLYGVDIGARKYLLAAVAVPLKPDLPKLPNIAIPSGEFWQAVERVEAEWRKWQRL
RTGKKDKEGRVVLKKKRMAWRVLKRKRRMRREVAKNMIAEAVNTFINFIKRGNVNSGLAIEQIKNLNLSPTLQDRKSRWRQRRW
AYRLLLEMLRYKAEWEGIRVIEVNPTHTSQICPRCRKKDGKARQGNRFCCTKCHFQHNADFVAAYNIAMRGLRKLGMRGEIGWK
VLTGKPRVLQ (SEQ ID NO: 628)
>SRR5273183_megahit_k177_82677_3|M
[淡水宏基因组]
MAPHSRLPDLLTRDVGGILKALKEKREGYLREFKRWWEEVCRVANEKGSQAVAALKTHVDKLLNLVHQIPWNSPVWDGEAQEAF
KNLKEAWAVVRQKDHEQLKIKWRAEVSPLRRPPRKPKFQRLVFNADGRVSGGSRAEALQIEEREEGDLYLRLRLRDPEGVVGGG
AYQEVWLRLVIEDEGRKQMLRDFAPKYLQLRQDGKRWIAHIVIEREIKFGEPDYLYGVDIGARKYLLAAVAVPLKPDLPKLPNI
AIPSGEFWQAVERVEAEWRKWQRLRTGKKDKEGRVVLKKKRMAWRVLKRKRRMRREVAKNMIAEAVNTFINFIKRGNVNSGLAI
EQIKNLNLSPTLQDRKSRWRQRRWAYRLLLEMLRYKAEWEGIRVIEVNPTHTSQICPRCRKKDGKARQGNRFCCTKCHFQHNAD
FVAAYNIAMRGLRKLGMRGEIGWKVLTGKPRVLQ (SEQ ID NO: 629)
>SRR6481359_megahit_k177_932521_3|M
[淡水沉积物宏基因组]
MANRELVVVMQVQKVVKGKVVGLTHAKRQILDHEYEGLQRFLQGDSNVELYSANKQQAKRFYKTVKPEREYPLSIRKDLIRVER
RDTKIARYWARIPAKLRRGGVWVAVKPHCEITDDLEICESKLLRRNGEYFLHLAVKREVELNTSYSSVLAIDLGIRHVACSVDT
RSGETHFHGRELRRVRGHYFHLRRRLGEKKLLKVIKRIGARESRITNDQLHKIAKAIVDEAERLNALIAIGNPKGIRRNGKGRR
FNRKLNSWPFWKLRQYIEYKANWRGIAVVEVNEAYTSQRCWRCGTIGTRSGTHHGLFECPRCGLRENADRNGAFNIGRRALGQV
SKVGAVVSQPVTGAVPFSELGSVVQCPTPEATLFKGW (SEQ ID NO: 609)
>3300012677|Ga0153928_1011778_2|M
[真菌-菌丝-切叶蚂蚁真菌园]
MKITVKAKVTDDSKNKLGKIESEFQNYQVYLQSGCTTRVNLYSATKQQAQRALKKIKQPNPKKQYPVILRRDLVDIQQDIKFEG
VYWIKIPVYPNSINIRIKTSCKYDLTQHSIRESKILKIQNVWYIFITIEKQDERIIRADNVLAVDLGCKNIAVTVNTINTKPNF
YGAAIRGIRGFYFNLRRKLGEKKASDKIKLLKNIEFLQVNHELHKISKAIVEEAKRTNAVIVLGKLKGIRKNIKGSRRIRRLIN
NFPYYRLVQYIKYKAAWHSIKTLEISEADTSNTCFNCRTKDKKARKTQGLFQCGNCSIKTNADYNGAMNILQRGVGYFQL (SEQ
ID NO: 630)
>WP_121304574.1
[方形盐盒菌]
MKRNVSTTVRAKLHSLTNRKTDLLAREYEAFQTEVHGGDADLYSATDQQASKVQRQKYPNPGTEQPVVLRNDVFDVAYDEDTVL
SSWWVKVPVYDPEKGQGNSIWCPAHVPHKDEQLVREGNLRDSELVRSDGDWYVHLVVKRSVTVQDEYDDVLAIDMGARWVATCA
FLSDRKTTFYGEEVRRIREHYKQLRKSIGKAKPRQGQQVVERIGDAEARKVDDRLHKIARSIVEDAEERNAVIVVGDLGGIRKD
NDKGRYVNDKTHKMPFARLLNYIEYKAHGAGIDVQLVEEYDTSKTCNRCACEGVRTTQGRFECPECGLDDNADKNGALNIGKRA
LGKFLKPLSEAGAVLAQSETQVIVQRDDEPANLSVSVDSTPSGGTPRL (SEQ ID NO: 631)
>SEJ25361.1

[0331]

[利氏盐环境杆菌 tADL]
MKQTAEKTVVAKIFTNDLTQSKLDAINHEYEAFQRYIRGDEEADLYSATKQCADAYVDTDNLRDDHDYPWFIRNDVFDVGVDVE
QHDTDLTDWWLKLPVSQVWGGVNVPIQPHDEIPDEADVKDSKIVRVNGDYYAHLTVQQEVEVSEEYGGVIGVDFGVRWVATSVA
LPSRNTEFYGREIRRIRRHHHDLRTRLQEKGAYGTLQSVKVKEYRQVNDRLHKISRAIVEEAKECDALIVVGDLEGIQNQDLGA
EMNRRLHSMPHHTLKTYIEYKAMFAGIAVKSVNEYMTSQTCWRCGEQEATTRKGQGQHLCHECGLDDNADKNGATNIAKKGLGK
DIACPLSSLGAVCEPVLGPSGADQASATVCLRADLEAPASTKRTVR (SEQ ID NO: 615)
>WP_121597521.1
[盐红菌属物种阿特里特-26R]
MQRNVSTTVRVKLHSLTNRKADLLAREYEAFQTAVHGGDADLYSATDQQASKVQRQKDPNPDTEQPVVLRNDVFDVAYDEDTVL
SSWWIKVPVYDPEQGRGNSIWCPAHVPYKDEQLVREGDVRDSELVRRDGNWYVHLVVRRSVTVQDEYDDVLAIDMGARWVATCA
FLSDRKTAFYGEEVRRIREHYKQLRKSIGKAKPRQGQQVVKRIGDAEARKVDERLHKIARQIVEDAEERNAVIVVGDLGGIRKD
NDKGRYVNDKTHKMPFARLLNYIEYKAHDAGIDVQLTNEYDTSQTCNRCACKGVRETQGRFECPECGLDDNADKNGALNIGKRA
LGKFSKPLSEAGAVLAQPETQVIVPRDDEPANLSVSVGSTPSGGTPRL (SEQ ID NO: 632)
>WP_006630662.1
[特班齐赫盐红菌 DSM 14210]
MQRNVSTTVRVKLHSLTNRKSDMLAREYEAFQREVHGGDADLYSATKQQASKVQQQKDPNPDTEQPVVLRNDVLDIAHDKDTVL
SSWWMKVPVYDPERGQGNSIWCPAHVPHKDETLVQEGDIRDSELVRRDGDWYVHLVVKRSVTIPDEYDDVLAIDMGARWVATCA
FLSDRDTTFYGEEVRRLREHYKQLRKSIGKAKPRQGQQVVERIGDAEARKVDDRLHKIARHIVEDAEERNTVIVVGDLGGICRD
NDKGRYINDKIHKMPFARLLNYIEYKAHGAGIDVVLVEEYDTSKTCNRCACEGVRETQGRFECPECGLDDNADKNGALNIGKRA
SGKFSKPLSEAGAVLAQPKTQVIVERDEEPANLSISVGSTPSGGTPRL (SEQ ID NO: 633)
>DRR082460_megahit_k177_106880|P
[热泉宏基因组]
MRRRKPQQEGKEVQRTIICGVKFVAHEDEQLITESAKWYLEAKQKLLRYALQQEKVSISNPATVLKGLSVPAPPHSRLPDELSQ
AVSGILKALKKRRSEYLEKFQNWWQEVVRVSNKHGLQSFIPLKPKVEAFLNALANVKSWADPAWDGELQRRFEELKDMWAVVRQ
RDSNERRFRWRLHVPPLPKPPKTPELRLLGFDAGVRVSGGSRAQTLQIEQRNNDLFLRLLLARPGEKSKAGQYREVWLQLVLED
DLRKNWLLSPDIEAKFVRLRQIEDGKWQAHITVRKKVAKSEDFNIVYGIDIGLKRALLYCVPVATANAKPENWKELRLNKFIVP
AGEFWHRVIVEDERYRKWQRRFKGKKTEGIKGKRMAWRILKRLRKRRRKVTELLIHRVVNSLLENLMKASETAGFAIAIERLKD
LRNKPHLTSGKAQWKFHRWAYRFLLNTLRYKAEWEGVTVIEVNPAGTSQTCPRCGYKSRESRHDRLFRCRRCKFQHNADAVAAF
NIAKRVIEMKNGQVNQDAWKVLTGEWNVLKY (SEQ ID NO: 634)
>DRR082460_megahit_k177_106880|M
[热泉宏基因组]
MQVRRRKPQQEGKEVQRTIICGVKFVAHEDEQLITESAKWYLEAKQKLLRYALQQEKVSISNPATVLKGLSVPAPPHSRLPDEL
SQAVSGILKALKKRRSEYLEKFQNWWQEVVRVSNKHGLQSFIPLKPKVEAFLNALANVKSWADPAWDGELQRRFEELKDMWAVV
RQRDSNERRFRWRLHVPPLPKPPKTPELRLLGFDAGVRVSGGSRAQTLQIEQRNNDLFLRLLLARPGEKSKAGQYREVWLQLVL
EDDLRKNWLLSPDIEAKFVRLRQIEDGKWQAHITVRKKVAKSEDFNIVYGIDIGLKRALLYCVPVATANAKPENWKELRLNKFI
VPAGEFWHRVIVEDERYRKWQRRFKGKKTEGIKGKRMAWRILKRLRKRRRKVTELLIHRVVNSLLENLMKASETAGFAIAIERL
KDLRNKPHLTSGKAQWKFHRWAYRFLLNTLRYKAEWEGVTVIEVNPAGTSQTCPRCGYKSRESRHDRLFRCRRCKFQHNADAVA
AFNIAKRVIEMKNGQVNQDAWKVLTGEWNVLKY (SEQ ID NO: 635)
>OGCH01000195_3|M
[热泉宏基因组]
MKHLVSNAQVNEKNEARLTLKCGLEFENAEGEELVKRAAEWYCKARIFLLSYALKNNKTRVQKIASLLKSEGFPSPPHPRLIDL
LQRTITPAIKNMRRLRRRYPKTLRAWFDRVKKFAEEKGSDAAERLQPYVEDLLQLVKNLPINSPAWDTDAQEKWKELKKLWGKV
RESDWKKGIVWKFQIPNPKKPPSTPAFRCADSSNIFVTADYRMSGKNARNSYDMKTDSEENLWVLVRLDDVDGDVKVRKDVWIK
LCVDKSRYDLLRIGKITEIKILRNRDGRWEGRIVVKMNAGYNSPQVVVGVDVGASEHLIAAVGIPLDGSNNNFLLSVPASYFWK
RVEQAERLWRRWQRLYTGRTSEKKKAPILPRRRKAYQILKRNRNWRREVTKHCIFHAVKKFVQLLPPNSVVVVENLKDIPPNVW
KEDKDKRFRRWVYAFLLFTLKYKCQQRKIAVIAVSPHNTSQTCPRCGVVDPKSRGLWKDGELISRKHFCCTQCEFTHNADFVAA
HNIAVLGAKALGITLPESWRALRGEVQYVIPYQNMPKVRKSW (SEQ ID NO: 636)

[0332]

>SRR4027810_megahit_k177_26210_1|M
[热泉宏基因组]
MEKVILTITSGVEFLQPQDEQLVIKAAKWYTEARKALLYNALHKGEEGVKQAASTLSTLNFPPAPHSRLPDLLTRDVGGTLQAL
KERREGYAQDFSRWWEEICRVANEKGSQAVSVLKPHVNKLSSLIHQLSWNSPEWDEEAQEVFKELKQIWAVVRQKDHERLKIKW
RAEVPPVPRPPRTPKFQRLVFNADGRVSGGSRSNALQIEEREDGDLYLRLQLRDPERVAKKGSYQDVWLRLVIEDEERKQMLKN
FAPKYLQLRQDGKRWIAHIVIERENEFEVPQYVYGVDIGARKYLLAAVAIPLNPALPKPPNIAVPSSKFWQIVERVEAEWRKWQ
QLRTGKRDKEGRVVLKKRKMAWRVLKRKRRMRREVTKSMIAEAVNTFIKSLERPCAIVIEQIKNLHLSPTLQDRKSRWRQRRWA
YRFLLEMLRYKAEWEGIRVIEVNPAYTSQTCPRCGKRDRNARQGNLFQCTKCRFQHNADFVAAYNITVRGLMKLRTRGRIDWKI
LTGKPRVL (SEQ ID NO: 637)
>SRR5247126_megahit_k177_220968_4|P
[热泉宏基因组]
MEKVVITITTGVEFLQPQDEQIVRKAAEWYTKAREALFDHALQQGKVKKAASTLRTLNLPMAPHSRLPDLLTRDVGGILKALKE
KRKGYLREFKRWWKEVCKVANEKGSQAVAVLKPHVDKLLNLVRQLPWNSPVWDGEAQEAFKNLKEAWAVVRQKDHEQLKIKWRA
EVSPLRRPPRTPKFQRLVFNADGRVSGGSKSEALQIEEREDGDLYLRLQLRDPEGVVGGGAYQEVWLRLVVEDEERKQMLRDFA
PKYLQLRQDGKRWIAHIVIEREIKFGLKGIMWTAGDYLYGVDIGARKYLLAAVAVPLKPGLPKLLESQGDNVGIRHIAIPSGKF
WQAVERVEAEWRKWQRLRTGKKDKEGRVVLKKKRMAWRVLKRKRRMRREVAKNMLAKAVNTFILDFIKRNNVDIWLAIEQIKNL
NLSPTLQDRKSRWRQRRWAYRLLLEMLRYKAEWEGIRVIEVNPAHTSQICPRCGKKDGKARQGNRFCCTKCHFQHNADFVAADN
IAMRGLRKVGMRGEIGWKVLTGKPRVLQ (SEQ ID NO: 638)
>3300027941|Ga0207719_100292_34|M
[实验室富集限定的培养基-有氧富集培养基]
MQQTVSTTVRVKLPSLTNRKADLLAREYEAFQTEVHGGDADLYSATKQQASKVRRQKDPNPDTEQPVVLRNDVFDVSYDEDTVL
SSWWVKVPVYDPERGRGNSIWCPAHVPHKDEQLVQEGDVRDSELVRRDGDWYIHLVVKRSVTVQDEYDDVLAIDMGARWVATCA
FLSDHKTTFYGEEVRRIREHYKQLRKSIGKAKPRQGQQVVERIGDAEARKVDDRLHKIARQIVGDAEERNAVIVVGDLGGIRKD
NDKGRYVNDKTHKMPFARLLNYIEYKAHDAGIDLQLVDEYDTSQTCNRCACEGVRATQGRFECPGCGLDDNADKNGALNIGKRA
LGKFSRPLSEAGAVLAQSETQVIVHRDEEPANLSVSVDSTPSGGTPRL (SEQ ID NO: 639)
>3300027941|Ga0207719_100292_35|P
[实验室富集限定的培养基-有氧富集培养基]
MRVKLPSLTNRKADLLAREYEAFQTEVHGGDADLYSATKQQASKVRRQKDPNPDTEQPVVLRNDVFDVSYDEDTVLSSWWVKVP
VYDPERGRGNSIWCPAHVPHKDEQLVQEGDVRDSELVRRDGDWYIHLVVKRSVTVQDEYDDVLAIDMGARWVATCAFLSDHKTT
FYGEEVRRIREHYKQLRKSIGKAKPRQGQQVVERIGDAEARKVDDRLHKIARQIVGDAEERNAVIVVGDLGGIRKDNDKGRYVN
DKTHKMPFARLLNYIEYKAHDAGIDLQLVDEYDTSQTCNRCACEGVRATQGRFECPGCGLDDNADKNGALNIGKRALGKFSRPL
SEAGAVLAQSETQVIVHRDEEPANLSVSVDSTPSGGTPRL (SEQ ID NO: 640)
>SRR6484323_megahit_k177_106899_4|P
[填埋物宏基因组]
MALKTLTFWPLEIPPEHKRIFDDIRKQLGNAQREMVKAWLKLALAAEPYGITPGEQAGRVLFGETDYSNGLLKSLAYTDAHDSF
KEQWKGVLKGDATSVSFPDEASGVVLLVGSGQATVEKIDGEWWVTRLLVFGGRTAEDRSRNDQFRWQLKGVSRFGGHYEKSVMD
RAERICQVRVLPARKDRGKYVVQVTVEVPDEQIERTAIPVCAGIDLGINCPVVLSIPDKGFVRFMGRERDEYPRLWEKLREYED
RKRTLNRAGKRRAARDLRPRLTNIRKYINEAISKDVADTCRAMRVTTLRMENLKGLSFHKSGTGKLRHWPRYHLQERIEQKCGE
VGIEVERIKPAGTSQTCSMCGHRDKGNRNGSDFICLACGYRQHADLNAANNIARGMQEFGPCVQSSASDNTIELRFDGAGQEPA
LGRAMRLFA (SEQ ID NO: 641)
>SRR6484325_megahit_k177_1345066_6|P
[填埋物宏基因组]
MALKTLTFWPLEIPPEHKRIFDDIRKQLGNAQREMVKAWLKLALAAEPYGITPGEQAGRVLFGETDYSNGLLKSLAYTDAHDSF
KEQWKGVLKGDATSVSFPDEASGVVLLVGSGQATVEKIDGEWWVTRLLVFGGRTAEDRSRNDQFRWQLKGVSRFGGHYEKSVMD
RAERICQVRVLPARKDRGKYVVQVTVEVPDEQIERTAIPVCAGIDLGINCPVVLSIPDKGFVRFMGRERDEYPRLWEKLREYED
RKRTLNRAGKRRAARDLRPRLTNIRKYINEAISKDVADTCRAMRVTTLRMENLKGLSFHKSGTGKLRHWPRYHLQERIEQKCGE
VGIEVERIKPAGTSQTCSMCGHRDKGNRNGSDFICLACGYRQHADLNAANNIARGMQEFGPCVQSSASDNTIELRFDGAGQEPA

[0333]

LGRAMRLFA (SEQ ID NO: 641)

>SRR6484787_megahit_k177_1514834_8|M

[填埋物宏基因组]

MDTAANETTPKALCFWLGALSDAEFKLLHELSRQLARCHATFVQAALEYQFAADRGSAPEVIAQTGPNKGKPSSDPRETYAKRA
FQAEFAKAPFMRKFGDTYCRQVSGKDFKRQWGLIRTGQKSSVSFGPGSHANIDILTGRQAAIEHDADGWWLCGLGITLEDRKER
PGRRWALIPQTNGGAAVKAKQAARMLRQLEEAVRTPQVRLVWRDGKRGGSKGWEAQVVCHMPVKVAKPAEPVVCGIDVGMRCLL
VASIPERRKARFIGGNTFGKLWHEIERHGKRRSVLNKAHKRHAAAAKSDKVSRIRRHACELASRQLIDWCVYHHVSTIRMEDLS
GIRDRCSEDGATEQQRTWNKRLAHWPWYTLQQRIEQKAAAVNIAVEKINPHLTSQTCSACGHVDKGNREGAQFRCLSCGFECHA
DLNAANNIAGRDPATDFDTPDRAAVRSDGPKSRPQAGRASRLVAASTP (SEQ ID NO: 642)

>SRR8542065_megahit_k177_1115558_2|M

[填埋物宏基因组]

MALKTLTFWPLEIPPEHKRIFDDIRKQLGNAQREMVKAWLKLALAAEPYGITPGEQAGRVLFGETDYSNGLLKSLAYTDAHDSF
KEQWKGVLKGDATSVSFPDEASGVVLLVGSGQATVEKIDGEWWVTRLLVFGGRTAEDRSRNDQFRWQLKGVSRFGGHYEKSVMD
RAERICQVRVLPARKDRGKYVVQVTVEVPDEQIERTAIPVCAGIDLGINCPVVLSIPDKGFVRFMGRERDEYPRLWEKLREYED
RKRTLNRAGKRRAARDLRPRLTNIRKYINEAISKDVADTCRAMRVTTLRMENLKGLSFHKSGTGKLRHWPRYHLQERIEQKCGE
VGIEVERIKPAGTSQTCSMCGHRDKGNRNGSDFICLACGYRQHADLNAANNIARGMQEFGPCVQSSASDNTIELRFDGAGQEPA
LGRAMRLFA (SEQ ID NO: 641)

>SRR8542065_megahit_k177_821416_8|M

[填埋物宏基因组]

MLGQVKTVKFHITGVTRDQGEALTGSLRALGAAARTAVKTALMEVADKDLHGVKDAYWQGVQVRAEQSAREAFKETGYSPAILN
EYWNAFGFTGFKIRWSQVRKGERTSIPFPGDAPFLQFTNGATIQVAETEDGYCIRRIRIYGLVGNKPDTRNEVEWALKPVGRGR
HAAWGIEQLRRANKICAVRIHKAERGWIAAVSVRIQAEAVAASRPERWAGIDLGLNSPAVLSVPDANFSRFYGRDLYSGLWRSL
DEYEARKRELVRAGKKQAARDLDDRISAKRKHINELISRQIVDDCRRLRVTGIRLEDLKGIQQFATGRLKHWPRHNLALRIEQK
AAEVGIETEMVRPAYTSLTCSHCGHMEKDNRSRENVSLFLCRRCGFSRHADLNGANNIARKLERFGIEEPNLSVFERAEQPRLH
RLGATTGPEHLEGRLTEAAT (SEQ ID NO: 643)

>FZPS01000148_2153|M

[宏基因组]

MIKTATFYIADMPQEDKRLCERLCAELGAAGATMVREWYRACDEVTGERINWPEVQKRGVQVGRQAFAERGYRAELVNDLQRAE
FFARFKMEFKKIQAGERASLPFPGSAPFIFLRMDPGKEKATLVQEDGTWYLRDPHFYPGGGRNEDNKREWRLKDVGRNGWGYAI
LAEAEKAACIRIMPDRKKRGSWVVKITVHLPDAALRKEVERQEVWAGIDLGLNCPAVLSIPDHPDGGFVRFFGNDEYARLWQRL
KEYEDRKRRLNRAGKRRAARDLRAKITGVRQHINELISREIIDLCIAMRVTGLRMEDLKGLGADGSGKLKFWPRFHLVTRIEQK
AHEAGLEFVKIKPAGTSQTCSACGYRDKDNRNGSDFTCKRCGFSRHADVNAANNIARGMEEFSLPTGATAPNKAQRLVIAS
(SEQ ID NO: 644)

>WP_005554257.1

[解淀粉嗜盐碱球菌 DSM 10524]

MHREVTTTVRVKLHSLTERKARLIEREYTAFQDAVHGGDASLYSATKQQAGKVRSNKNPREDTDQPVVLRNDCITIEHDEDTVL
SSWWFKLPIYNPKRERGDSIWVPVHVPEKDTHLLTDEYIRDSELVHRDGEWYVHLVCKRSVAVADEYDDVLAVDMGAKWIAVST
FLSDRDTQFHGAEVRRVREHYKQLRKSIGKAKVRQGAQVIERIGDKESWTVEHELHQVANELIARARERNAVIVFGDMTGLRFD
NDEGRYVNDKTHKMPYAKLANILTYKSHLDGRECIPVSEYDTSVTCWRCGSQNTSRGAQGRVECCDCGLDDNADKNGASNIGKR
AVGKDIQSPLSTVGAAVAQPETQVVLDGTSGEMEPANSPKDVGLTLSEGSPRL (SEQ ID NO: 645)

>SRR7004341_megahit_k177_2665443_5|P

[泥炭宏基因组]

MSLVKVKKCYYCKVIEDLTNTKLVALQQEYDNLQHYLQTKEDLGLYSANKQQANRFYKVIKRGKKYPLSIRNDLIRIEHKPSTI
AEYWIRIPVKSVKGGLWIGLIKPYEPIPTDAKICESKLYKKDNSWFLDLVVEKDIPEKLQYQNVIGIDMGIRHIATSVELANGL
TKFYGKDLNRVRGHYFWLRRKLGIKKAIDTIKKIGDHEKRIANDIIHNISRDIVNQAIESNAMIVLGDIKHLRRRKKNTRGKKF
NRKLAGFQYCKLTQYITYKAALAGIKVIQVSQARTSQYCRKCYQKGVRKIQGLFTCAKCGEENADRNAAFNIIYRALGYISKVG
VTVNIPNARTFASVDRNAMMTREATCF (SEQ ID NO: 646)

[0334]

>SRR8526178_megahit_k177_763700_3|P
[盐湖宏基因组]
MSLKTLTFWPIGMSAKYRRIFDNIRKQLGAAQREMVKAWLRLALAEDAYGIDPGKRAGREVFGDNEYSNDLLNNLAYADAHENF
KQQWKRVRSGNATSVSFSDAAANVLFLMPPSGGHHATVEHIDGNWWITRLKVFGGCTKADRERNDQFRWQLKGVSRFGGKYEKD
LMDGAQRICQVRVTPSRKDPSKYVVQVTVEVPDLQTVGARTPVCAGIDLGINCPVVLSIPDKGFVQFMGRERDEYPRLWEKLRE
YEDRKRRLNRAGKRRAARDLRPRITNVRKYINEAISKDVADTCRALGVTHLRMEDLKGLSRHKSMTGKLRYWPRYHLQERIERK
CCEVGIEVERIKPAGTSQTCSVCGHREKANRNGAQFVCIECGYQQHADLNAANNIARGMQTFGPCAKSSASDDAIGLRFDSEGQ
ESALGQAKRLFA (SEQ ID NO: 616)
>OVXO01016275_2|M
[沉积物宏基因组]
MLVRKVIKAKIMDEANSRKLKALEKEYNNFQRALRGEDVKLYSATRQQAERLLRKIRKQRGKLKQKEYPLILRNDVYDVRITNN
KLSKYWIKIPVYGVRGGIKLPIKPHCEIDASWKFKEAKVIRKENGWFVHITVEKEIEIKKEYENVIAIDLGIRNIATVVSTADG
KPRFYGKELRRIRGHYFHLRRRLAKLKKFKAIKKIGNKERRVVDDLLHKISREIVNRAKEENAVIVLGDLKRIRKQNRGRKFNR
KLNTFPYYRLSKFIEYKALWEGIAVIKVNEANTSKICSRCGSQGLRRNGKFVCSKCGVELNADYNGALNILKRALGYMSKAGAD
LTQPFQSHFGLILTMKKRYGRD (SEQ ID NO: 647)
>SRR7067823_megahit_k177_107253_1|P
[沉积物宏基因组]
MEKVILTITSGVEFLQPQDEQLVIKAAKWYTEARKALLYNALQKGEEGVKQAASTLSTLNFPPAPHSRLPDLLTRDVGGTLQAL
KERREGYAQDFSRWWEEICRVANEKGSQAVSVLKPHVNKLSSLIHQLSWNSPEWDEEAQEVFKELKQIWAVVRQKDHERLKIKW
RAEVPPVPRPPRTPKFQRLVFNADGRVSGGSRSNALQIEEREDGDLYLRLQLRDPERVAKKGSYQDVWLRLVIEDEERKQMLKN
FAPKYLQLRQDGKRWIAHIVIERENEFEVPQYVYGVDIGARKYLLAAVAIPLNPALPKPPNIAVPSSKFWQIVERVEAEWRKWQ
QLRTGKRDKEGRVVLKKRKMAWRVLKRKRRMRREVTKSMIAEAVNTFIKSLERPCAIVIEQIKNLHLSPTLQDRKSRWRQRRWA
YRFLLEMLRYKAEWEGIRVIEVNPAYTSQTCPRCGERDRNARQWNLFRCTKCRFQHNADFVAAYNITVRGLMKLRTRGRIDWKI
LTGKPRVL (SEQ ID NO: 648)
>SRR7067986_megahit_k177_69094_7|P
[沉积物宏基因组]
MEKVILTVTSGVEFLQPQDEQLVIKAAKWYTEARKALLYNALQKGEEGVKQAASTLSTLNFPPAPHSRLPDLLTRDVGGTLQAL
KERREGYAQDFSRWWEEICRVANEKGSQAVSVLKPHVNKLSSLIHRLSWNSPEWDEEAQEVFKELKQIWGVVKQKDHERLKIKW
CAEVPPVPRPPRTPKFQRLVFNADGRVSGGSRSNALQIEEREDGDLYLRLQLRDPERVAKKGSYQDVWLRLVIEDEERKQMLKN
FAPKYLQLRQDGKRWIAHIVIERENEFEVPQYVYGVDIGARKYLLAAVAIPLNPALPKPPNIAVPSSKFWQIVERVEAEWRKWQ
QLRTGKRDKEGRVVLKKRKMAWRVLKRKRRMRREVTKSMIAEAVNTFIKSLERPCAIVIEQIKNLHLSPTLQDRKSRWRQRRWA
YRFLLEMLRYKAEWEGIRVIEVNPAYTSQTCPRCGERDRNARQGNLFRCTKCRFQHNADFVAAYNITVRGLMKLRTRGRIDWKI
LTGKPRVL (SEQ ID NO: 649)
>SRR8554444_megahit_k177_504614_2|M
[沉积物宏基因组]
MVAVKTIKAKIIAPTRRKAEILEKEYQGFQDILHGQQAEIYSRTRDQALRVLRRIKAPKDRHYPLILHRASIRLERRDTKIAQY
WFRFPCHGVRGGIWLAIKPHCQIPEGCSLRETKLIKRNNAWFLYIIIKKEVPEPVINPDRIVAVDLGERYLATVCGGEGIRPQF
LGKEARGIRRHYAWLRRRLGELKLLRTIKKVGNTERRKINTICHQISRKIVDLAKQTGSTIVLGDLKGIRQRARGRRMNRIVSG
MPYYKLTQFIKYKAAWESVPVLIVSERDTSKTCHRCGAIGTRPYQGLFLCQACGLRYSADLNGAKNIMKRAREYISLAGAAFSQ
PMTRALAGNPLMETG (SEQ ID NO: 650)
>SRR8554446_megahit_k177_83048_1|M
[沉积物宏基因组]
MKKVVITITTGVEFLQPQDEQIVRKAAEWYTKAREALFAHALQQQQGKVEKAASTLRTLNFPMAPHSRLPDLLTRDVGGILKAL
KEKREGYLGEFKRWWEEVCRVANEKGSQAVAALKTHVDKLLNLVHQLPWNSPVWDGEAQEAFKNLKEAWAVVRQKDHEQLKIRW
RAEVSPLRRPPRTPKFQRLVFNADGRVSGGSKSEALQIEEKEDGDLYLRLQLRDPEGIVGGGAYQEVWLRLVVEDEERKQMLRD
FAPKYLQLRQDGERWIAHIVIEREIKFGKPDYLYGVDIGARKYLLAAVAVPLKPDLPKLPNIAIPSGKFWQAVERVEAEWRKWQ
RLRTGKKDKEGRVVLKKRTAWRVLKRKRRMRREVAKNMIAEAVNTFINFIKGDNVDIRLWLAIEQIKNLNLSPAFQDRKSRWRQ

[0335]

RRWAYRLLLEMLRYKAEWEGIRVIEVNPAHTSQICPRCGKKDGKARQGNRFCCTKCKFQHNADFVAAYNIAVRALKKLGMREEI
GWKVLTGKPKVL (SEQ ID NO: 625)
>SRR6201998_megahit_k177_1568738_2|M
[土壤宏基因组]
MKITVKAKVTDDSKNKLEKIESEFQNYQVYLQSGCTAIVNLYSATKQQTQRALKKIKQPNQKKQYPVILRRDLIDIQKDIKFKG
VYWMRIPVYPNSINIRIKTSCKYDLTEYSIRECKILKIRNEWYIFITIEKQDEPIIPPDNVLAVDLGCKNIAVTVNTINTKPNF
YGAALRGIRGFYFNLRRKLGEKKASNKIKSLKNIEFLQVNHELHKISKAIVEEAKRTNAVIVLGKLKGIRKNIKGSRRMRRLIN
NFPYYRLVQYIKYKAAWRSIRALEISEADTSNTCFNCRTKDKKARKTQGLFQCSNCSIRTNADYNGAMNILQRGVGILSTLGGF
LTYPEPSVIIDRNKMITKEPQML (SEQ ID NO: 651)
>SRR8554511_megahit_k177_4248910_4|P
[土壤宏基因组]
MIKTATFWLCEVTSDQRAVLVDISKQLGAAHRAMVCEWLRLAMAGEPYGENAGKQKAREVFADTPYSNGLVNNYSHADAHPRFK
KQWEAVLKGERSSVAFPGHAPFIYLGAPEQVSVELLDGGWHIVDPRLYIGDSKDNGRSITWQEGRSRNRSIAWRLDPVDNRRTG
WCLGLLRDAECVSCVRLMKSDRRPGRWLAKISVELAEAASARDKEEVELCAGIDLGLNCPVVLSVPDKGLLRFFGREGDMYPRL
WQKLRQYEDRKRAFNRAGHKREARDLRMNLTNVRQYINECVSKAVVDYCLANRVTKLRLENLKGLGVGATGKLKHWPRYHLQQR
IEQKGAAAGLEVEYVKAAGTSQTCSVCGHRHRDNRQRSQFECLACGYTAHADVNAANNIARNLETFGPVTGSRSAQEAPVGLRF
DWEGREPGLRGKSSRLQLDLGTQVSRSAEGGVINAAT (SEQ ID NO: 652)
>SRR8554511_megahit_k177_4248910_4|M
[土壤宏基因组]
MDISKQLGAAHRAMVCEWLRLAMAGEPYGENAGKQKAREVFADTPYSNGLVNNYSHADAHPRFKKQWEAVLKGERSSVAFPGHA
PFIYLGAPEQVSVELLDGGWHIVDPRLYIGDSKDNGRSITWQEGRSRNRSIAWRLDPVDNRRTGWCLGLLRDAECVSCVRLMKS
DRRPGRWLAKISVELAEAASARDKEEVELCAGIDLGLNCPVVLSVPDKGLLRFFGREGDMYPRLWQKLRQYEDRKRAFNRAGHK
REARDLRMNLTNVRQYINECVSKAVVDYCLANRVTKLRLENLKGLGVGATGKLKHWPRYHLQQRIEQKGAAAGLEVEYVKAAGT
SQTCSVCGHRHRDNRQRSQFECLACGYTAHADVNAANNIARNLETFGPVTGSRSAQEAPVGLRFDWEGREPGLRGKSSRLQLDL
GTQVSRSAEGGVINAAT (SEQ ID NO: 653)
>3300014204|Ga0172381_10015437_8|M
[固体废物-填埋物-填埋物沥滤液]
MDTAANETTPKALCFWLGALSDAEFKLLHELSRQLARCHATFVQAALEYQFAADRGSAPEVIAQTGPNKGKPSSDPRETYAKRA
FQAEFAKAPFMRKFGDTYCRQVSGKDFKRQWGLIRTGQKSSVSFGPGSHANIDILTGRQAAIEHDADGWWLCGLGITLEDRKER
PGRRWALIPQTNGGAAVKAKQAARMLRQLEEAVRTPQVRLVWRDGKRGGSKGWEAQVVCHMPVKVAKPAEPVVCGIDVGMRCLL
VASIPERRKARFIGGNTFGKLWHEIERHGKRRSVLNKAHKRHAAAAKSDKVSRIRRHACELASRQLIDWCVYHHVSTIRMEDLS
GIRDRCSEDGATEQQRTWNKRLAHWPWYTLQQRIEQKAAAVNIAVEKINPHLTSQTCSACGHVDKGNREGAQFRCLSCGFECHA
DLNAANNIAGRDPATDFDTPDRAAVRSDGPKSRPQAGRASRLVAASTP (SEQ ID NO: 642)
>3300014205|Ga0172380_10102486_2|M
[固体废物-填埋物-填埋物沥滤液]
MALKTLTFWPLEIPPEHKRIFDDIRKQLGNAQREMVKAWLKLALAAEPYGITPGEQAGRVLFGETDYSNGLLKSLAYTDAHDSF
KEQWKGVLKGDATSVSFPDEASGVVLLVGSGQATVEKIDGEWWVTRLLVFGGRTAEDRSRNDQFRWQLKGVSRFGGHYEKSVMD
RAERICQVRVLPARKDRGKYVVQVTVEVPDEQIERTAIPVCAGIDLGINCPVVLSIPDKGFVRFMGRERDEYPRLWEKLREYED
RKRTLNRAGKRRAARDLRPRLTNIRKYINEAISKDVADTCRAMRVTTLRMENLKGLSFHKSGTGKLRHWPRYHLQERIEQKCGE
VGIEVERIKPAGTSQTCSMCGHRDKGNRNGSDFICLACGYRQHADLNAANNIARGMQEFGPCVQSSASDNTIELRFDGAGQEPA
LGRAMRLFA (SEQ ID NO: 641)
>3300031566|Ga0307378_10048523_2|M
[陆生-土壤]
MSLKTVGLYLAEMPQEDKVVCKAICRDLGRAGARMVREWYRACDEVAPGEKIPWGEVQTRGVVVGKATVQEAGLRPELANEMQR
TEFFARFKTEFGKILRGEKASLPFPGDAPFIYLRLDPGNAAATIEKREGVWYLRDPHFYPGGGRNQDNKREWRLKGRSRDGWEY
RWLEAAEKLACVRIKPDHKKRGAWVALITLHVPQVDDRPGSPEAVWAGVDLGLNSPAVLSIPDHPEGGFVRFFGRDLYLRMWDR
LDQYEERKARLNRAGKRRAARDLRAKITGIRQHINELISREIVDLCRRLRVTGIRMEDLRGIQDAATGKLKHWPRYHLMMRIQQ

[0336]

KAGEAGLAFELVRAAGTSQTCSACGYREKGNREGALFHCLRCGFERHADVNAANNIARGMEEFDLPTGAQAMRPARRLELVAG
（SEQ ID NO: 654）
>SRR6837575_megahit_k177_854481_7|M
[废水宏基因组]
MTTDKQTVTKTLRFWLTDLSKEDFDLLHELSRELARCHRLFVQASLEYAFAGGTNTTERLRLARRVYQEEFGKSRFVKRFGNTY
TGEIADKDFKRQMGLIQRGQKSSVSFGPGSHANIDLLTEGLQCTLVKEGDGWYITGAALTLDEDKATRAARKWGIIPQTSGGAA
VKAKQHARTMRQIEQAERTSRVRLVWRDGKKGGRKGWEAQVVCRMPVTASKPSEPTVCGIDVGMHCLLVASIPTKDKATFIGGN
TFGELWHEIERHCERRTILNKAHKRHAAAAQSDKISRIRKHGCELASRQLIDWCMYHHVTTIRLEDLSGIRDRCSEDENSEKAR
TWNRRLSHWPWYQLQQRIEQKADAYGITVEKVNPHLTSQTCSSCGYANPANREGTQFKCQRCGFERHADLNAANNIARREPGVD
FDAPNTPGRAAVLSDGPKSRPKAGRSGRLVLSDNPVTPRPDTLQLDMVL（SEQ ID NO: 655）
>SRR6837576_megahit_k177_1219670_1|P
[废水宏基因组]
MTTDKQTVTKTLRFWLTDLSKEDFDLLHELSRELARCHRLFVQASLEYAFAGGTNTTERLRLARRVYQEEFGKSRFVKRFGNTY
TGEIADKDFKRQMGLIQRGQKSSVSFGPGSHANIDLLTEGLQCTLVKEGDGWYITGAALTLDEDKATRAARKWGIIPQTSGGAA
VKAKQHARTMRQIEQAERTSRVRLVWRDGKKGGRKGWEAQVVCRMPVTASKPSEPTVCGIDVGMHCLLVASIPTKDKATFIGGN
TFGELWHEIERHCERRTILNKAHKRHAAAAQSDKISRIRKHGCELASRQLIDWCMYHHVTTIRLEDLSGIRDRCSEDENSEKAR
TWNRRLSHWPWYQLQQRIEQKADAYGITVEKVNPHLTSQTCSSCGYANPANREGTQFKCQRCGFERHADLNAANNIARREPGVD
FDAPNTPGRAAVLSDGPKSRPKAGRSGRLVLSDNPVTPRPDTLQLDMVL（SEQ ID NO: 655）
>3300009716|Ga0116191_1022210_3|P
[废水-厌氧消化器-厌氧消化器污泥]
MIKTVTFYIAEMPREDKRLCDRLCIDLGAAGARMVREWFRACDETPPGAAIDWPEVQRRGVAVGKAMVEAAGYRTELGNDLARA
EFFARFKMEFKKVLKGERASLPFPGRAPFIFLRMDPGHETAIIEQRDDAWYMRNPRFYPGGDSNEANRREWRLKPVDRKGWAEN
ILEAAEKFACVRIMPDRKKPGSWVVKITVHLPDALERKLVEEHECWAGIDLGINCPAVLSIPDRDKGGFVRFFGMDKYTALWAK
LNRYEERKRALNRVGQQHAARDLRAKITGLRQYINEQISREIIDLCLRMHVTGIRMEDLKGLRTEGTGKLKYWPRFHLTTRIEQ
KAREVGLAFEQIRAAGTSQTCSVCGHRHRDNRHGSDFLCLRCGFARHADVNAANNIARGMQDFDLPAGSAAPGRALRLELSAQ
（SEQ ID NO: 606）

[0337] 表9.代表性CLUST.196682直接重复的核苷酸序列

[0338]

CLUST.196682_全部效应子蛋白登录号	直接重复核苷酸序列
3300025638|Ga0208198_1021643_2|P（SEQ ID NO: 601）	CTCTCCACGCGCGCGCGGGATGCGGG（SEQ ID NO: 701）
3300005300|Ga0072500_136962_4|M（SEQ ID NO: 602）	GCCTGCTGTCCACACTATCCCCCAAAAAC（SEQ ID NO: 702）
CAAACJ010042499_2|P（SEQ ID NO: 603）	GTGCTCTCCACGCGCGCGCGGGATGCGG（SEQ ID NO: 703）
SRR5007116_megahit_k177_300528_2|M（SEQ ID NO: 604）	GGTCTCTCCACGCGCGCGCGGGATGCGGG（SEQ ID NO: 704）
SRR5007140_megahit_k177_707372_2|P（SEQ ID NO: 603）	GTGCTCTCCACGCGCGCGCGGGATGCGGG（SEQ ID NO: 705）
SRR5007457_megahit_k177_513017_2|M（SEQ ID NO: 605）	GTGCTCTCCACGCGCGCGCGGGATGCGGG（SEQ ID NO: 705）
SRR5007461_megahit_k177_311051_7|P（SEQ ID NO: 606）	GTGCGGCCCACGCGCGCGCGGGGCAGAGG（SEQ ID NO: 706）
3300027896|Ga0209777_10000273_30|P（SEQ ID NO: 607）	TGGTGACCATGTTTTTGTTGAA（SEQ ID NO: 707）
3300027896|Ga0209777_10000273_32|M（SEQ ID NO: 608）	TGGTGACCATGTTTTTGTTGAA（SEQ ID NO: 707）

[0339]

3300013127|Ga0172365_10030480_1|M（SEQ ID NO: 609）	CCAGTCCAAGTTTCGATGAG（SEQ ID NO: 708）
3300013128|Ga0172366_10002984_1|M（SEQ ID NO: 610）	CCAGTCCAAGTTTCGATGAG（SEQ ID NO: 708）
3300013129|Ga0172364_10030775_1|P（SEQ ID NO: 609）	CCAGTCCAAGTTTCGATGAG（SEQ ID NO: 708）
3300013130|Ga0172363_10005240_1|P（SEQ ID NO: 609）	CCAGTCCAAGTTTCGATGAG（SEQ ID NO: 708）
3300013133|Ga0172362_10039269_3|M（SEQ ID NO: 609）	CCAGTCCAAGTTTCGATGAG（SEQ ID NO: 708）
3300032070|Ga0315279_10000130_51|M（SEQ ID NO: 611）	GGAAACAAAATAACAAACAA（SEQ ID NO: 709）
3300032118|Ga0315277_10001250_38|M（SEQ ID NO: 612）	TATAAGGTGTTTTGAAAATGA（SEQ ID NO: 710）
3300024429|Ga0209991_10029076_1|P（SEQ ID NO: 613）	GAGAAGGCAAAAAGAGAAGC（SEQ ID NO: 711）
3300014886|Ga0180300_10007429_2|M（SEQ ID NO: 614）	TTTCAATTCCCATTATATGGGATTTCTCTTTGAAACC（SEQ ID NO: 712）
3300005935|Ga0075125_10033396_1|P（SEQ ID NO: 616）	GTTGTCTCCACGCGCGCGAGGGCGGTTTG（SEQ ID NO: 714）
3300025736|Ga0207997_1027507_2|M（SEQ ID NO: 616）	GTTGTCTCCACGCGCGCGAGGGCGGTTTG（SEQ ID NO: 714）
3300007203|Ga0099832_1115789_2|M（SEQ ID NO: 617）	CTTTCAATCCCACAATGGTGCGATTGCTGCG（SEQ ID NO: 715）
3300000083|Ga0072500_119562_4|M（SEQ ID NO: 618）	AGTGCGGGGGGAACCTAGGAGGGCAGTGGGC（SEQ ID NO: 716）
3300008000|Ga0105162_1004089_4|M（SEQ ID NO: 619）	GTTCGCGAGAGACTTTGTGAATAACAAGC（SEQ ID NO: 717）
3300008000|Ga0105162_1004089_4|P（SEQ ID NO: 620）	GTTCGCGAGAGACTTTGTGAATAACAAGC（SEQ ID NO: 717）
3300022544|Ga0212120_1002806_5|M（SEQ ID NO: 619）	GTTCGCGAGAGACTTTGTGAATAACAAGC（SEQ ID NO: 717）
3300022544|Ga0212120_1002806_5|P（SEQ ID NO: 620）	GTTCGCGAGAGACTTTGTGAATAACAAGC（SEQ ID NO: 717）
3300025063|Ga0209493_1003385_3|M（SEQ ID NO: 621）	GCCTGCTGTCCACACTATCCCCCAAAAAC（SEQ ID NO: 702）
3300025090|Ga0209080_1004127_5|M（SEQ ID NO: 622）	GTTCGCGAGAGACTTTGTGAATAACAAGC（SEQ ID NO: 717）
3300025440|Ga0209741_1005495_1|M（SEQ ID NO: 623）	GCCTGATGTCCACATTATCCCCTTAAGAC（SEQ ID NO: 718）
3300028675|Ga0272445_1001731_11|M（SEQ ID NO: 624）	GCCTGATGTCCACATTATCCCCTTAAGAC（SEQ ID NO: 718）
3300029977|Ga0272449_1037372_1|P（SEQ ID NO: 625）	GCCTGATGTCCACATTATCCCCTTAAGACC（SEQ ID NO: 719）
SRR1562008_megahit_k177_216982_5|P（SEQ ID NO: 626）	CGTGCGATCCACGCGCGCGCGGGGCAGAGG（SEQ ID NO: 720）
UOPD01011681_4|M（SEQ ID NO: 626）	CGTGCGATCCACGCGCGCGCGGGGCAGAGG（SEQ ID NO: 720）
2645728127|Ga0097870_126_5|M（SEQ ID NO: 627）	GCCTGCTGTCCACACTATCCCCCAAAAAC（SEQ ID NO: 702）
SRR5272259_megahit_k177_4486292_121|P（SEQ ID NO: 607）	TGGTGACCATGTTTTTGTTGAA（SEQ ID NO: 707）

[0340]

SRR5272259_megahit_k177_4486292_118|M（SEQ ID NO: 608）	TGGTGACCATGTTTTTGTTGAA（SEQ ID NO: 707）
SRR5273149_megahit_k177_25824_1|M（SEQ ID NO: 628）	GCCTGCTGTCCACATTATCCCCTTAAGAC（SEQ ID NO: 721）
SRR5273182_megahit_k177_57442_1|M（SEQ ID NO: 628）	GCCTGCTGTCCACATTATCCCCTTAAGAC（SEQ ID NO: 721）
SRR5273183_megahit_k177_82677_3|M（SEQ ID NO: 629）	GCCTGCTGTCCACATTATCCCCTTAAGAC（SEQ ID NO: 721）
SRR6481359_megahit_k177_932521_3|M（SEQ ID NO: 609）	CCAGTCCAAGTTTCGATGAG（SEQ ID NO: 708）
3300012677|Ga0153928_1011778_2|M（SEQ ID NO: 630）	ATATTATGATAAAGCATTGGCTATAGATCCAA（SEQ ID NO: 722）
WP_121304574.1（SEQ ID NO: 631）	CAACAGAGTCACCAACAGC（SEQ ID NO: 723）
SEJ25361.1（SEQ ID NO: 615）	AGTAATTACTGAGTCATTACTC（SEQ ID NO: 724）
WP_121597521.1（SEQ ID NO: 632）	GCAGCTGGGGCTTTGGCGGTGTT（SEQ ID NO: 725）
WP_006630662.1（SEQ ID NO: 633）	ACACTTCGATAGAGGGGTGT（SEQ ID NO: 726）
DRR082460_megahit_k177_106880|P（SEQ ID NO: 634）	AGCTTACTATCCACAAAGTCTTCCGCGAAC（SEQ ID NO: 727）
DRR082460_megahit_k177_106880|M（SEQ ID NO: 635）	AGCTTACTATCCACAAAGTCTTCCGCGAAC（SEQ ID NO: 727）
OGCH01000195_3|M（SEQ ID NO: 636）	AGTGCGGGGGGAACCTAGGAGGGCAGTGGGC（SEQ ID NO: 716）
SRR4027810_megahit_k177_26210_1|M（SEQ ID NO: 637）	GCCTGCTGTCCACACTATCCCCCAAAAAC（SEQ ID NO: 702）
SRR5247126_megahit_k177_220968_4|P（SEQ ID NO: 638）	GCCTGATGTCCACATTATCCCCTTAAGAC（SEQ ID NO: 718）
3300027941|Ga0207719_100292_34|M（SEQ ID NO: 639）	GACCCTGAGCGGTGAAAGCCCTCGACCCGCTCGC（SEQ ID NO: 728）
3300027941|Ga0207719_100292_35|P（SEQ ID NO: 640）	GACCCTGAGCGGTGAAAGCCCTCGACCCGCTCGC（SEQ ID NO: 728）
SRR6484323_megahit_k177_106899_4|P（SEQ ID NO: 641）	CGTCGCACCTCGCGCGCGCGTAGGAGATGGTTTGG（SEQ ID NO: 729）
SRR6484325_megahit_k177_1345066_6|P（SEQ ID NO: 641）	GTCGCACCTCGCGCGCGCGTAGGAGATGGTTT（SEQ ID NO: 730）
SRR6484787_megahit_k177_1514834_8|M（SEQ ID NO: 642）	GTGCCGCCCACGGGCGCAGGGGTGGATTG（SEQ ID NO: 731）
SRR8542065_megahit_k177_1115558_2|M（SEQ ID NO: 641）	GTCGCACCTCGCGCGCGCGTAGGAGATGGTTTG（SEQ ID NO: 732）
SRR8542065_megahit_k177_821416_8|M（SEQ ID NO: 643）	GGTCGGCCCGCGCGCGCGAGGGGGGTTGG（SEQ ID NO: 733）
FZPS01000148_2153|M（SEQ ID NO: 644）	GTGCGACCCACGCGCGCGCGGGGCAGAGG（SEQ ID NO: 734）
WP_005554257.1（SEQ ID NO: 645）	GTCGTCTTCGTCGCCGATATCG（SEQ ID NO: 735）
SRR7004341_megahit_k177_2665443_5|P（SEQ ID NO: 646）	TAACCAGAATAGCGGTAAT（SEQ ID NO: 736）
SRR8526178_megahit_k177_763700_3|P（SEQ ID NO: 616）	GTTGTCTCCACGCGCGCGAGGGCGGTTTG（SEQ ID NO: 714）
OVXO01016275_2|M（SEQ ID NO: 647）	CTTTCAATCCCATTTTGGTCTGATTTTAAC（SEQ ID NO: 737）
SRR7067823_megahit_k177_107253_1|P（SEQ ID NO:	GCCTGCTGTCCACACTATCCCCCAAAAAC（SEQ ID NO: 702）

648）	
SRR7067986_megahit_k177_69094_7|P（SEQ ID NO: 649）	GCCTGCTGTCCACACTATCCCCCAAAAAC（SEQ ID NO: 702）
SRR8554444_megahit_k177_504614_2|M（SEQ ID NO: 650）	GTTGAGTGGCCACATAGTGGGAGCTGAAC（SEQ ID NO: 738）
SRR8554446_megahit_k177_83048_1|M（SEQ ID NO: 625）	GCCTGATGTCCACATTATCCCCTTAAGAC（SEQ ID NO: 718）
SRR6201998_megahit_k177_1568738_2|M（SEQ ID NO: 651）	GTATTATGATAAAGCATTGGCTAT（SEQ ID NO: 739）
SRR8554511_megahit_k177_4248910_4|P（SEQ ID NO: 652）	GCCGTGCTCTCCGCGCGCGCGTGGGCGGTTCG（SEQ ID NO: 740）
SRR8554511_megahit_k177_4248910_4|M（SEQ ID NO: 653）	GCCGTGCTCTCCGCGCGCGCGTGGGCGGTTCG（SEQ ID NO: 740）
3300014204|Ga0172381_10015437_8|M（SEQ ID NO: 642）	GTGCCGCCCACGGGCGCAGGGGTGGATTG（SEQ ID NO: 731）
3300014205|Ga0172380_10102486_2|M（SEQ ID NO: 641）	GTCGCACCTCGCGCGCGCGTAGGAGATGGTTT（SEQ ID NO: 730）
3300031566|Ga0307378_10048523_2|M（SEQ ID NO: 654）	GGCGCAACCCGCACGCGCGCGCGAGGGCAGAG（SEQ ID NO: 741）
SRR6837575_megahit_k177_854481_7|M（SEQ ID NO: 655）	GCAGAGCCCACGGGCGTGGGGGTGGATTG（SEQ ID NO: 742）
SRR6837576_megahit_k177_1219670_1|P（SEQ ID NO: 655）	GCAGAGCCCACGGGCGTGGGGGTGGATTG（SEQ ID NO: 742）
3300009716|Ga0116191_1022210_3|P（SEQ ID NO: 606）	TGCGGCCCACGCGCGCGCGGGGCAGAGG（SEQ ID NO: 743）

[0341]

[0342] 表10.代表性CLUST.196682系统的非编码序列

[0343]

```
>3300009716|Ga0116191_1022210_3|P
GCCTTCCGCGCGTGGCTGTCCTACGTGGGCGCGGAGGTGACGCCGGAGGAGCGGCGCAAGGCGGAGAGCGCGGCGGTGAACGGG
GAGTGCCCGCCACAACGCAGTTACGGGACTTCCCAAGTCTTTGATGGCTTTGAGAACCCAGCGGGCGCAGGGATGACCAACCCA
GAACACGCCGCGCAACGTCACCGCGACGGCGCGGGTATCCTCACCCGCCTGCTGACCGCGCACTGGCCGTCGTCGGCGCAGGAG
CAGCGCGCGCACTGGGAGGCGGTCATCGGCGAGGTGCCGGAGGGCTGGGATTGCACGAGGCCGCCGGGGGTGAGGGTTTCGCAG
GGGCGATCCCATTGGGGGCGCGAGAGAGGCCGCAATGGTTCTCATGCTGGCGTGTTGGACAAGTCCTCCACTTTCGCGAACGAG
GCATCGCACGACACCGCCCCATCGACTCCCGACGCCGCGCGCTTCGACGGCCTGCGCGGCTCCGTCGCCCCGCTCAAGCCCTTC
GCCTACCCCATCCTGCACGCGCAGAAGCCGCGCGAGGGCAGCTACCTCGCGAACCTGCGCGCGCATGGCGTCGGCGCGTGGAAT
GTGGAGGAGGCGAGGGTGCCGTTCAGTGAGGATGGTGTCGAGGGCTTCTGGCCGGGCGCCCGAAGTGGACGCGGCGACCTGCAG
CCCCCGTCGTGGGGAGCCACAGAAGGGGGGCAGCGCGATGACGCAGGCCGCACCCCCTCCAACCTCCTGTCCTACGGCGACCTG
CTCGGCGACCTGCAACGCTACTGCGACCTCGACGCCTGGTGCGAGCGCCTCGGCCTGCCCGACGCCGCCGCCGACCTGCTCGAA
GCGGGGTTGGTCTACGCGCCGAAGCCGAGCAGGGCGGAGAAAGAGGCCGGGCTGGAGGCAGACAAGCCGCCCGCGTTCGGCAAC
GATAAGGGCGATGGCCTTGGGCGGGGCATCAGCAACACGAGACAGGATTGGGCGCGGGGCAACACGCATCCCACCGCGAAGCCC
GCCGCGCTCTGCGCCTACCTCGTCGCCCTCGCCACCCGCGAGGGGCAGACCGTGCTTGACCCGTTCTGCGGGTCGGGCAGCACG
GGCGTCGCCTGCGCGCGCATCGGCAGGAACTTCGTCGGCGTCGAACTCGACGCCGACTTCGCCGCCATCGCACGAGCGCGGATC
GCACACGCCAAGACCACACGACCGGCTCAGGAGGCCCTCCCGCTATGACCGCGGACCTCCGCGACCGGAAGTTGACTCATTCCC
TCCATTCTGGTATACTGTCAGCGTTCCAGGCAAGGGACTCAGAGGAGGTCGTTTCAGTGTACGCTCGACAGTTTTGCCCCCGTG
TCTCCAGCGACCTCCTGTTCCCTTGCCTGGGCAGAGACGCTGTCACGGGGGCTGTTCGTGTACGCGCTTCGAGACAAGCCGACC
CGGCTGCGGTTGGGTTGGTTGCGGCTACTCTGGGCCGCGAGGCCGGGCGATCAGACCCGGCAAACCCGCGAGTACAAGTGTGGG
```

[0344]

CAGCATGGGGCGGGCAACTCTGCTCATCGTAGTGGGAGCTGGCAAGCATGATCACAGGAGGCAAGACTGATGATCAAGACCGTC
ACCTTCTATATCGCTGAAATGCCGAGGGAGGACAAGCGCCTCTGCGACCGCCTCTGCATCGACCTCGGGGCCGCCGGGGCGCGC
ATGGTGCGCGAGTGGTTCCGCGCCTGTGACGAAACCCCGCCCGGTGCGGCCTCAGACTTCCTCTGCTTGCGGTGCGGCTTCGCG
CGCCATGCCGACGTGAACGCCGCGAACAACATCGCGCGGGGGATGCAGGACTTCGATTTGCCAGCCGGGTCCGCCGCCCCTGGG
CGAGCGCTTCGGCTGGAGTTGAGCGCGCAATGAGACAATGCGGGCGGGCTGCGGTTCGCGCGTGTGCGGCGACTGTGCGCCGCG
AGGACGACGGTCAAGTTGGCGAGTCATTGTTATGCAATGAACACAGGCCTGCCCACGGGACGGATGAAGGGCGCGGTGCGCCTC
ACCGCCTGCGCGACCGTCAGCGACCTGGTGCGCGCGGGGCTGGTCGTCGTCGCGTGGGCCTACGAGAACGGCGACCTCGGGCTG
CGGGGGACGGAGGCGGCGGCGACTTTCTTTCGCGAGGGTTGA (SEQ ID NO: 800)
>3300012677|Ga0153928_1011778_2|M
AGAATGAAGATGCATTAATTTGGAAAGGTATAGCTTTCTTTACAATGAAGAAAACGTTATTTCTGACAATCGCAATAACAATCG
CTGTGGTGGCGGTAGCGTTACCTTTCGTAGCACCAATAATCGCAAAAGTAGCTAATGCTGGCGGTAGCGTTACCTTTCGTAGCA
CCAATAATCGCAAAAGTAGCTAATGCTGACGGTAATTAGTTTTGATAAATTGTCGGCCTATAAGGTCATGGGCATTAGCAATTT
ACAATATATTTGAAGCCTGAAAGCCCGCCTCCTTTAGGAGCGGGATGAAATGCTGACCGATATCTTAATAGCTCCAAAGCTCTA
GTATTTGTATGAGTCGAGGACTCTTCGAATGAAAATCACAGTTAAAGCAAAAGTCACAGATGACAGCAAAAACAAACTAGGAAA
AATAGAATCAGAGTTTCAAAACTATCAGGTATATTTGCAATCAGGATGTACCACAAGAGTTAATCTATATTCTGCGACGAAACA
ACAAGCACAAAACACATGCTTTAACTGTCGCACTAAAGATAAGAAAGCAAGAAAAACACAAGGATTGTTCCAATGTGGTAATTG
CAGCATTAAAACGAATGCTGATTATAATGGAGCAATGAACATCTTGCAACGAGGGGTTGGATACTTTCAACTCTAGTGAGCTTT
CTGACATACCCCGAACCGTCGGTGATTATAGACAGGAATAAAATGATCACGAAGGAACCCCACATGCTTTATCTCTGGGAGGAT
GTCAGAGGATCAAACGGCTAACAATACCATAGATAGGTTCTCCTCCGAAATGATGTCCTGTGTCTTTATCGTTTAGGCTAAGAG
TCATTGTGTTTCCCTTGGAAATCAATACTGATATATGCGCATTATTCCACATAGTTACACCATTTATTTTTACATCAGCAGTAC
CATTCAAAGATAAGCTATTGTCTGATGTCATTTGAACTAGTTGTGTTTTTCCTCCAATATTATCAATTTTGAAGTTTGTTATTT
GGTGAATATGTGGTCTTCTACCATCCATCGGTGATGCAAGGAAGTCTACACCAAAATTCGTAACGTTCCCTTTATGCACATTCA
AGTTCCATTCTCCAGATAAGATAACCTTAAAAGCATCAGTTATATTAAATCCAGATTCGGGAATTGTAATAATTAAAGAATTTA
TTGTTCCTATTGCAGAGAATGTATCCTTTGCTGGGATTGGAAAAAATACTGGTAGTGAAGTCACCGCTATTTTTTTAGCTGCAT
ATGCCTCATAGTAATAACTATGAGTATTGTTATTACTAGCGTTGATATCTACTACAACGACGATCCCTATTATTATCCA (SEQ
ID NO: 801)
>SRR6837576_megahit_k177_1219670_1|P
CTGCGGCCACCCTGCCCCGTGGACTCCCCCGAGGGGCAGCAGGTGCTGCGGGGCCTCTTTCGGCAGGCGTTCGCGGCGTTGCCG
CCGGAGGCGCTGGAGCGGTGCTGCGAGGCGATCACGGCGGAGGATGACGAGGACGCTGGTGGGTGATGCCCACACCCGCCACGC
GGCAGGCCCCCTGTCCGGCACATCCTCGACTGGGCAGGGGGCCAGGCACACACAGGACGCACCCGCCGCCGGCCTCCTTCCCCC
GCGGCGGTAGACGGACCCGGTCTTGCTCCCGCGCGGACCGGCGACGACAGGGGCCGGGGCTCAACTCTCACTCCCGGCCCCGGG
GCCCGAGCAACAGACTACAGACAGGAGTAACGGCAGTGACCACCGACAAGCAGACGGTAACCAAGACGCTCCGCTTCTGGCTCA
CCGACCTGAGCAAGGAGGACTTCGACCTCCTGCACAAGCTCAGCCGCGAACTGGCACGTTGCCACCGTCTGTTTGTCCAGGCGT
CCCTTGAGTACGCGTTCTTTGACGCACCCAACACACCCGGCAGGGCTGCGGTTCTGTCGGACGGGCCGAAATCACGGCCCAAGG
CGGGGCGGTCTGGGAGGTTGGTTCTGTCAGACAACCCCGTAACCCCGCGCCCAGACACACTGCAACTTGACATGGTGCTCTGAC
CTGTACAGGTTTCGCGAGATTCCTTGTAGGAGCGTGGTAGACATGCCGACATGCAGCAACTGCCTGGCCCTCACTGAGCGCATC
CGTCGACTCAAGGCTGACGAGGACGCGCACCTGGAACAGATCGCCGCCCTGGGAGCCCGGCTTGCGCGCAGCGAGGCCAAACTT
GCGCGGTCGCAGGCGCTGTGCCGGCGGCTGGCGAGCAACGTCGCCGCGCTGGGTGGC (SEQ ID NO: 802)
>OVXO01016275_2|M
CTCGTTGTGCCTATGACGGCCTGAATGACTTCCTC (SEQ ID NO: 803)
>3300013128|Ga0172366_10002984_1|M
TCGCTTAACCAAAGATGGCGAATAGAGAACTGGTGGTTGTGATGCAGGTGCAGAAAGTCGTCAAGGCAAAAGTCGTGGGGCTGA
CTCACGCTAAACGCCAGATACTCGACCACGAATACGAAGGGCTCCAGCGATTCTTGCAGGGGGATTCCAACGTTGAACTCGCCT
TCAACATAGGCAGACGAGCCTTGGGGCAAGTCTCCAAGGTAGGGGCTGTCGTGAGCCAGCCCGTAACTGGGGCAGTGCCGTTCA
GCGAACTCGGTAGCGTCGTTCAATGTCCGACCCCAGAAGCCACCCTTTTTAAAGGGTGGTAGTTCACGTGCGCGAGCGCCTGCG
GCGTTTAGTCCAATCTACTAATAGGTGCTCTCCCATCTCCCTAAATTTCCTCGCCGCTGGAGAACGCGGGGCATAGCGAAGAAC
TGAGCTGCCAGCGGTCAAGCTCTCACGAACTTTATCGTCATCGGGTTGCCACTCAATTGGTGCTTGGACAGCCCCAGTCCATGT

[0345]

TTCAATCAGATTCCACCCAATTGGAGCTTGCACCGTGCCAATCCAAGTTTCAATCAGTTGCCATGCGGCAGTGACCGTGACTGT
CCCAGTCCAGGTTTCGATGAGTTTCCATTCACCAGTGGCAGTGACTGTTCCAGTCCAAGTTTCAATCTCGTACCACTGAGCAGC
TGCTGGTGCTTGGACCGCCCCGGTCCATGACTCGGCGAGTTGCCATTCAGCCGGAGCACTTACTGTCCCAGTCCAAGAATCGAT
CTGCTGCCACTGTCCCGTGGTTGCATTAACAGTCCCGCTCCAAGACTCGACCTGACGCCACTGACCAGTGGTCGCGTTGACCGT
TCCAGTCCAACTCTCAATCTGCCGCCACTGTCCCGTGGTCGCGTTCACAGTCCCAGTCCAGCTCTCAATCTCGTACCACTGAAC
AGCTGCTGGTGCTTGGACCGTCCCGGCCCATGACTCGATGAGTTGCCATGCAGCAGGGGCGGTCACCGTGCCCGTCCAGCTCTC
GACGAGCTGCCACGCGGCCGGGGCGCTCACGGTGCCAGTCCATGTTTCGATGATCTGCCAGACTGCGGAAGCATTGACCGTCCC
CATCCACGACTCCACCAATTGCCATTCAGCTGGCGCCTGAACGGTTCCAGTCCACGTCTCTATGAGCTGCCACGCTGCAGGTGC
ATTAATAGTACCCGTCCATGTCTCGATCACTTGCCAGACGGCAGGCGCTGTCACGGTCCCAGTCCAGGTTTCGATTGTTTGCCA
AGCCGCGGGTGCGCTGACCGTTCCAGTCCAGGCCTCGATTAGCTGCCATTGGGCAGGAGGTGGAACCACCGTGAACTGCCAAGT
GTTTTCAGAAACATTCTCACCCGCAGCGTTGACCGCCCACACTCTCCAGTAGTAGGTCCCCGGGGTGGAGAAGTTGCGCTGGTA
GATGTTGTCGGTGTTGTCTTGGTAGTTGTCCACGAGCGGCGTGGAGAAGGTGGGCTCGTTGTCAATCTCGATCCTATGGTTCTC
CCCAAACCACCCGTTGTCCCACTCGAAGGTCACAGTTTCCCCGACCCCGAAGACAGAGTTGTTCTCTGGACTGTAGAGCCTGGG
CCTCTTCGGAAAGAATGTGTACTCGTATGTGAAAAAGATAGTACCGGTAAGAGCGGCGCCTGTTCCCATGTCGTATTCAAAGTT
GTAGGTCAGGTCGGCAAGGGTCCTATCTACAACAACGTCGATGTTGTCTTTGTTTACAGAGATGGATGTCAAGGCCTCCCCAGA
GCTGCCTGTGCTCATACTGACGGCTGAGCCCCAGTTTACCCCATCAGCTTAGATCCTGGTCTGTCGCGTTTCCGAAGCGGTTGC
ACTTCCTATCCAATATGCCGTGAGGCCAACGCTCTTCACAGTCTTTTTGTTTTCCGGGGCATAGAACGTCAGCTCCCCGCTGGT
CCAAGGGTCACCGGCAGCCCCGCTGGGTTTTTGGCCCGCGGTGAACATCGTGGTTTTCAGCTTTGTCGCGGCCGATGGGCGGAA
GTGATAGGTGATGTGCAAAAGGCCGCTTACCTCATAGAGGTCGGAGGCCGCAGCGCTCTGGATGTTGAGGTTCACATTGTCCCC
GTTCGACCAGTTCTCCCTCATGAAGGCGCTGATGTCGTGGAAGTAGTCGAACCCCTCTACATGCTCGCTGCCTATGTTGGTCGT
GTAAGCTCCCGAGGTCTGGGTTTTCGCCCCGACGCTCGCGTACACGCTCGTGCTGCCGTCACTAGTCAAGATGGTGTGGAAAAT
TTCAAAGTACGCGTTCTCGATGTTGTCCACGCGAGAGCTGTCCAGGTCCGTCAGATAAACGGGCGTCCCGAAGGCGAACAGGTT
GGCCGAGCCGTTGACGTTGGCGGCCCCCATGGCAAACGTGAGTGTTTTTATCTGGTCCGCGCTGTTGTCTGAATACTCATATGT
GACGTAAGTTTCACAGCTGAGCGCATCGTTTGTTTGTCCTCCGGTAGTAATTGTAATATCCTGATCAACGTTGTAGTTCCAGGA
AGCCGTCGGCCGGGAAATATACCAGAAATTGAGATCGTTGCCGGTACGGGCTTGCTCTGTCCAAGCATTTGTGGTGGCTGTATC
GCTTCCGATCTTGATGTTAAGAGAAAAGTCGGTTGAAGAGGCTTGGGAAGAGTGGGCATGGATTTCGTACCAGTTGTCCCTCAG
CGTGACCCCGTCTTCTGGGATTTCGATTCTGTGTGTGTAGGTCGCCGTGCCGCCAGCCCCTATCTCGGTGTTCGTGCTCCCCAC
AAAATACCTGACAGTCTTGATCTTCGTGTCCGCCAGCTCGTTCTCCTCATATGTGATGTACATCTTGGCCGTGTGGCTCTGGAT
ACCCGCGTCAGTGACGGTTAATTTTGGGTAGACGGCGGCCCCGTTTGTCCAGTATGTTGTAAAAGCAGAGGTAAAATCTGCGAC
CACCCATGCCGCCACGCTTTCCCCGCTGTTGGCGGTGTACTGGCCGGTCACTGGGTCTGCGTCTGGGTTCCGGTCGGTCCCCGC
GATGTTCAGGGTGACGGTCATATCTGTCACGTTGGCAGCAGTCGATGCCCCCGTCTGGACGTGATACATCACGAACACCTTCCT
GAACGGCCCAGCCTTGTCCGGGATGTAGATGGTTTTTGTCGGGGTGAACGAGTAGGCCGTGCCAGTGCCCAAAGCGGACGATGA
GGTTCCTTGCCTCGTATCCCATGAATATTGGATTGTGTTTATGGTCACTATATTTTCGTTGGGTTTGGTGGGGGAGGGAGTAGC
CACGGCTAGGGGCTGCATCGAGATAACGAAAACGCTGGCTAACATCATCGCTATCAATACGAATACAAAGATTTGTTTTGCCTC
CTTCCGCCCGAACTCTTCCCGCCCCAACCTAAACAAGCTACAACCCATCTCCCTCACCTGCTTGCCTCCACTTGGACTATGGAA
TGGGCGCGGCTCACGGCAGAGCCATCTCTCATCTCTCTCAGCAGGTTCGTGAAGTACTCAATCCCGCTCGCGGTGAGCTTGCAG
AGCTTTCGAGGCCTCCCCCGGGTTAAACGGGGCTC (SEQ ID NO: 804)
>UOPD01011681_4|M
CGCGTTCGTGCAGGGCTTCCTTGCGATCCGCGCAGACGTGAAGCACCCGCAGGCGGAGGATGGCGAGGACTCGTTCGGCGATCA
GTAGCGAGTGGTATTGACAAGCCGCCCCATATCATGTAAACTGAAAGTGTCCCGGCAAGGACTTGAGAGGAGGTCGCGTTCACA
ATGCACATCAATAGGATTGCCCCCGCGAAAGCGCGACCTCCGTTTCCTTGCCGGGACGAGAGCGTTGCTTTTCGCGGGGGTTTT
GCATGTACGCTTACAAGACGCCTCGAACCGGCTGCGGTTGGTTCGACTGCGGCTACTCTGAGCCGCGAGGCCGGGCAACGAGTC
GTCCGGCAAACCTGCGCGCGCCCCAAGCGTGCGATCCACGCGCGCGGGGCAGAGGACAACGGCACGTGACGACCGGTTCGACGT
GCGATCCACGCGCGCGCGGGGCAGAGGTAACCGGGTCAGAGTGTCGTACATAGGGCACCGTGCGATCCACGCGCGCGCGGGGCA
GAGGCAAGGCGGCCTGGGCGGCGCTGGAGAAGGCGGAGAGGCAGGCCTCGCAGCGTGCGATCCACGCGCGCGCGGGGCAGAGGA
GGAGAGATAGCTAAAGCAATGATAAAGACCGCGACATTCTATCTGGCCGACATGCCGCAGGAGGACAAGCGCCTGTGTCAACTC
CTCTGCGATGAACTCGGCGCGGCGGGCTCCGCGATGGTTCGCGAGTGGTATCGCGCCTGTGACGAAGTAGATGGCGAGCGTATC

[0346]

GGCCCCGACTTTGTTTGCCTGCGGTGCGGTTTCGCCCGCCATGCCGACGTGAACGCAGCGAACAACATTGCCCGTGGCATGGAT
GACTTCGACTTGCCGACGGGCGCAAGCGCGCCGACGAAGGCGCGGCGATTGGAGATTGCATCATGAGAAGTGGCTGTCGGGCTG
AGGTTTGACGGCTTGCGGTGATTGCGCACCGCGAGGATAGCCGATGGTTATCAAGTGCTGATACAGCATCGCAAGCGCAGATGA
TGGTGCAGAGATCGTGCTTGACGAAGACGAAGTGCAGAAGCTCAGGGAGGCGCTGGAGGAGCACGCGAAGAGCAGTCCGCAGGA
CTGACGCACACGCCGCACCTGCTGACGCGCATGGGGACGCACGGGGATGCAGAGCGCGCAGCGAGAGGTTGCCGGGATGCGCTT
GCAATCGACGAACGTTCGTGCTATAATGGAGGTGTCTGGAGAGACACGGCGCGGCTTTGCGTTCTATGGCCGCCCACCACAATC
TGACCGTGACCATATTGCCCTCGTGCAAACCTCCGCGCGTGTCTCTCCAGACTGCGTAGGCGGCCAGCGCGAGGGCTTTTCTTT
GCTCCGGGGTGATCCAGTTCGTGGCTAACCCAAACCACTACGTCAAACTCTTCCTCGCAACGTTCGACGGGCAGTCCGCCGATA
TGCAGGCGCTCACTATCGGCCAGAAGTACGTGTGGCTGGCGCTCCTCGCATTTGCCCGAGCCTACAGACCGTATGGCATTCTGA
ACGCAGGAGAGCACGAGGGGCGTCCTATGCCCGCGAGTGTCCGCCGCCTGGCTGACTGGGCTTCCGTTGGAGTCGGCACTGTGA
GTGAGGCGCTGGCAATCTATGAGAAGAGAGGAATGATAAAGCAGGATAACCGGCGCAGTGATCCGTGTTGGGAGATCGTCAACT
GGGAGAAGTGGCAGGTGTCTCACGGCCTCGAAGAGTGTTCGGGAGCCGAACAAGATGGTGTTCGGGAGGCGAACACACAGACTG
TAGCGCGTTCGGGAGATGAACAGCACTGTTCGCCATCTGAACGCCAGTGTTCGGCAAATGAACAGCAGTGTTCAGATGGCGAAC
ACCTATCTCCGGACGCAGAGTGTTCAGGAGGCGAACACAAACAGGTAGAAAGCCTAGGGGCGCAAGGGCTATTGGCCTTAGTAT
CTGACGAGCAGGGGTGTTCGCCTACCGAACACGATTCCCTAGACGTTAAGAGAACAACATTAACAGCAGAAGAAGGCACGGCTA
CTTCTGCTTCTCAAGTAGATCAGGCACGACAAATCCTCTGCACCCTTCTTCGCTTGCCCGATGGATCAATCGCGATTGATGCCG
CGCTTATGGCGGCCTCTCGTGTTCATCCGCGCGCCGCCGATGTCCTTGTCCAAGAGTGTGGAGCACTTGCGGCGCGGGAAGCTC
AAGGCGGCAAGCCGGTTGATCTACAATACCTCGCCTCGAATGTTCGCCAAGCGGCGGAGCGACTTGAGCAG(SEQ ID NO: 805)
>3300025638|Ga0208198_1021643_2|P
GTGCGGGTGTGGCGCTCGCACCGGGCTGAATATCAGGAGGCGAGCGGCATGCAGGAGGCGCTGATGGTCGTGGCTGGGATCGTG
ATCGCGCTCGCGGCGACCGGATGGGTCTGGTTCGCGCTGTGTATGCGGGAGGGCTAGATTGAAGGCAACGACCATCACAAAGAG
CCTCCGCTTCTGGATCGACGAACCGCCTGAGCGGTGCCGTCTCCTCTATGGCGTCAGCGATGAACTCACCGATGCATACAACGC
TATCCCCGAGTACTGGAAGACCGATGTCCGCAAGGTAGCCATCCGACCGTCTACACAGCCGGCGGGACGTATCCCGTCGGCCGC
GCCCGACTGCACGGGCGAGGCAGTGAAGGCGTCAGACGTTGACGCTGAGCTGCAACCTGCGCGCACGCCGGCAGCTTGTGGAGG
TGATCCATATGCCGGCGAGTGAGCGAGTGGAGAGGTGTGGGGCCCTGGTCTGTACGGGGTAG (SEQ ID NO: 806)
>3300014205|Ga0172380_10102486_2|M
GACCTTAATGCTGCGAATAACATCGCGCGGGGGATGCAGGAGTTTGGGCCTTGTGTCCAGTCCTCCGCGTCAGACAACACTATC
GAGCTGCGGTTCGATGGTGCGGGGCAAGAACCTGCCCTGGGGCGGGCAATGCGATTATTCGCATGAGTCCGCAAGTCTCTACAA
GCGAGTGCGACCCCGAGATGGTTTGGAGGCAACCAGTGACCATACAGGAGCGAGTCTACCAGACATTAGCCGCACGCAACATCC
GTCCATCAGCGCACCTGCTGCACGATATGGGCCACTACGGCGAGGCTCAGGCGCTCAGTCAGAACCCTGCCGAGAACCCGTTGC
AGCAGGCGCTCAGA (SEQ ID NO: 807)
>3300014204|Ga0172381_10015437_8|M
ATCTACCGTATGTGCCAGATAGTCCGCATAAAGCGCCTGCAGATGGAGCAGTTTCGGAGAGCGCAGGCCAGGGCGGAAGCGGAG
ATGGACTGGCGCCGTCAGGCGCATCGCAAATGGCAGCGGTTGCGCGAAGCGATGGACAGCCGGTAGTCGTCTTGCCGTGGACCT
CTCCCGCCGCGACTGACATGCCCGGAGGAAACTCTGGCCCCCCCCACACCCGGGGCTCCGATCACAGCCGGGGCCCCGGCCCAC
AGGCCGCACCCCCGCGTTGGCGGGGACAACGACATAAGAGGTAGGCCGTAGACGCAGAGGCACAGAGGACGCGTGACGAGGGGC
GCTGCGACAACTGCACGGGAAGCAGTGACGTAGCGCCCAACCCACCATAATCGAGGAGCAAGCGGACCGTGGATACTGCAGCGA
ACGAGACGACACCCAAGGCTCTGTGCTTCTGGCTCGGCGCTCTGAGTGATGCCGGATTCAAGCTCCTACATGAGCTGAGTCGCC
AACTGGCACGGTGCCATGCGACGTTCGTGCAGGCGGCCCTGGAGTATCAGGCAGACCTGAACGCTGCGAACAACATCGCCGGGC
GTGACCCGGCGACAGACTTTGACACACCCGATCGGGCTGCGGTTCGGTCGGACGGGCCGAAATCACGGCCCCAGGCGGGGCGGG
CGTCGCGACTGGTCGCGGCGTCAACCCCGTAACCCTGCGCACCTAGACATCGGTGTTGGCACATTTCTGGTCTGCGCAGCTTGC
ACGGAAAGCAGTTGACGTGGCGCCCCACCCAACACACAGGCCATTCCCCGCGTTAGCGGGTTGGCAGGCAAGGCGTGCAGGGGC
CGGGCTGGACACCCGGCCCCACTATTCCACAGGAGGCGGACGATGCCAGGCACGGAGACGGTGACGGTGACAATCCCTGTGGAC
TACGTGACGGCCCGCCGGATAGCCGAGCTGGAGAGGAAACTCTCCCAGGCCAGGGGCGAGGCCAACGACCAGGAGTGCGACGCC
ACCTTCTGGCGGGAGAGGTGCGAG (SEQ ID NO: 808)
>WP_005554257.1
GCGGAACCGTGCCGTTTCATCGTCGACGCGAAACACGACTAACCCGCGGCCGCCTGCGTGAAGCGCGGTTCGCTCGGGTAATCC

[0347]

CGACACCGGCTCCGTCGGCTCGATCGCCGCCGGGTCAACTACCTGAATCTCGTCGCTCATGTTCATTGTACCGATGGGGTCGGG
TTACAGAGGTTCGAGGAGAAAGCCCACGACTTCAGTCGTGGGATGAATCCGATCAAATACGACACAATCCATCCTCAATAGCCG
ACCGAAGGTTGATAGAATCATAACATCTAATATAGATTGACGTCTATATAACAGTATGCCACGGGGGTACAGCCGAGAGCGAAC
GTCGGTTCACAACCTCCACTACCACTTCGTGTGGTGTCCGAAGTATCGTAAACCGGTACTGACAGACGAGGTTGCCGACCGCCT
CGAACACCTCATCGAGGAGAAAGCCGACGAACTCGGACAGGTCAAGGGCTACACCTCGCGGAACCTTCGCGACGAGTTTGATTT
CGGCCTCCCGTCGCTGTGGACACGCTCGTACTTCGTCTCCTCGGCGGGCGAAGTGTCGAGTCAGACCATCGAGGAGTACATCGA
AGCACAGACTGGGCAATAGCGATGCACCGCGAAGTCACTACCACAGTACGTGTCAAACTCCACTCGCTCACCGAGCGGAAAGCC
CGCCTCATCGAACGCGAGTATACGGCGTTCCAAGACGCCGTTCACGGTGGCGACGCGAGCCTCTACTCCGCCACCAAACAACAG
GCGGACATTCAGAGCCCGCTATCCACGGTGGGGGCTGCCGTGGCTCAGCCCGAAACGCAGGTCGTACTCGACGGAACCAGCGGT
GAGATGGAACCTGCGAACTCCCCGAAAGACGTGGGCCTAACCCTCAGTGAGGGAAGTCCACGACTTTAGTCGTGGGAGAAGTCA
CGTCTGTCGGATGACCCACGGGCTCGCGGTTCGAGCCGTCGCTCCGCCGCTATACCTCCCAGGGTTCGCCGGTCGTCTCGCCGA
ACAGCTCCCGCATGAGCGTGACGATGCTTTCCGGCGGGAACTGGCCCCGTTCGGTGACGATCGCGTCGACGTGACGCGGCGGCG
TGACGTCGAACGCCGGGTTCTCCACCGTGAGTCCTCCTCGGAGTCGTCTATCTCGTCGTCGCCGTCGTCATCCGAGTCGTACTC
CTCTCCGTCGTACTGCTCGTCGTCTCCGTTTTCGCTCTCGTCTTCGTCGTCGGCCACCCCTTCGTTGTCATCCGCTTCGCCCTC
CTCATCTCCGTCGTACTGCTCGTCGTGGTCGTCCCTCTGGTCTTCGCTCTCCTCCTCACCGTCGTCGTCATCCTCGTCGTCCCC
ATCTCCGCGATCATCGTCCGCCGTTTCCTCACTTTCGTCCTCCTCTCCATCGTCTTCGGCATCGTCGCCGTTTTCTCCATCACC
TTCGTCACCCTCGTCGTCCTCGCCAGTCTCTTCGTCCCCATCATCCTCCTCGCGAAGCGCGACGAGACTGTCCTCGTGTTCGAC
GTAGAGCGTTTCGTCGCTGATACCCAGGACGTTGCCGACTGGCAGCTCCCCGGCTTGGATCGCCCAGCGCTCGCTTCCGTCGTC
CTTGTCGAGGGCGACGAGTTCGCCGTCGTATTCTCCGTACGTCGCATCCGTCCCCCTGAAGTAGACGTAGACGGTGTCGCCGGC
GATGATCGCGCGGCCGACGTTGCGAGACACGTCGACACCCCAGTCGTCCTCGTCGCCGTGATAGGCGCTTCCCGTGAGCTGGTC
GGCCTCGTACCCGCTGAGGTAGCGCTCCTTGCCGAGCGTCACCGAGATCCCCTCCGCCGACACCTTCGCAATATGCTCGCCGGT
CTGTGCGTCGAGGAGCACCCGGGCGTAGAAATCGACCCGGCTCACGACGACGGCGGTCGACGTCGCCCGTGCCTGGGCGCCTCC
GGTGTAAAAATTCGAATCCCCTCGCCAGACCTCATCACCCGTATCGGGGTCGAGTGCAACCGTTGCACCGCCTTCCAGATCCCC
CGCGCCGGCGTAGACGACGCCGTTCGCCGCGGCGGTCGTATCGAGGAAGTCGTACTCCTGTTCGCCCTCGTCATCGTGTTGGGT
TTCAGCGGTGAGCGTGTCGATCTCCCATCGCACCGACCCGTCGTCGGTCCCGAGGGCGTACAGCGTGCCGTTGACGACGACGAA
GACGCCTCCGTACGCGACGGTCTGACTCTGAATCGAATCCTCCGGCTCGAACTCGGTCTGCCACCGGACGCTTCCGTCGGCGAT
GTCGAGCGCGAACAGTTCGTCGCCGGTAACGTAGACGGTGTCGTCGACAACCGACGGCGCGTTCGCGTCGACGTCGGTGTTCTC
CCAGATAATCGAACCGTCGGCGACGTCGAGCGCGACGACTCCGTTCGCATTCGTCGCGTAGACGACGTCGTCGGCGACGGCCAC
GCCGGCGTCTCGTCCGAAGCCGAAGTCGACCGACCAGGCTGCCTCGGGCGCGTCGCCGTCGAGCGCGGTCCCGTCGGGGACGTA
CCTCGCGTGCCCGGCGTCGCCGTCGAACGACGGCCAGTCCTCGTCCATCCCTGGGTTCGGATCGCGGTCCGAATCCGGTAGATC
GTCGGTCGACTCCTCCCCGGCACCGATCGACGCGAGGGCTCCCCCTGCAGTGAGCGTTGCACCGGTCGCCAGCACCGAACGCCG
TCGCCACTTCGTCATACGGGTGTCGACGACAGAATATCGGATTACTATCCAGTTGCTACGGCTGACTTGTCCTTCCTGCAACTG
TTTCAGACGGACGAGCACCGAGCTTACTCGACGGAACGGAACCGACGGCATCGACCCTCGAAGACGGACTCGCGGGCTCGGGCG
AAGCGGCTTCCGGCCGTTACCACTCCTGGCCGGCGGGTTCGGCCTGCCGTGGTTCCCCCTCGTCCTCGGACGTCTTCTCCCCGC
GGACGACGAAGTACGCGATTGCGGCGACCACGCCCAGCGGGAAGAACAGGAACAACAGCCCGGCGGTGCCGAGCGCCCAGAG

(SEQ ID NO: 809)

>SEJ25361.1

GCAGTCGTTGCATCGATCTCCCGACTCGCAGCTGGCGAGATTTCGTATTCGGCCCGTCCATCCTCGGCAGTGAGGTTCTGTTTG
ATTAGGTCAGCCCGCTCGTTGAACTCGACGGCGGTGCTGGTGTCGGCGTCGTATTCCGGATCGCGGAAGTATTCCCGTGGCGTT
TCGGGTTTGCGGTAGGCGGCGTTGGGGTTGTCGCGGTTCCACTCTCTGTTTTCAGGGATCTCGGCGTCAAGCTGTTGGTGGAAG
AACTCGATAGCCGCTGCCATTGCCGCCTTGATCCGGTGTTCTTCGGGTGGGTCGACGGCGAGTTCGGCTTCGAGTTCCGAGTAG
TCTTCATCAAGATAGTCCTGTTTGGTCTTGGTGGGAGTCGTGTCTGTCTCGGTCCCTTCCCCCGTCTCTGATTCAGTCTCCGTC
GACTCGTCGGTGGGTGTTGCGGTGTCGTCTTCGTCGGCGTCGGCGTCTTGATCGTCGCTAGTTGCTGTGGTTTCGGCAGTCCCG
TCTGTGTTGGCCGCTGTCTGCGGTTCGTCTGTGTCGGTGGTAGGATGTATCGGCTTCGGCGTCCGATTGCTCACCAGCCGATTC
TGCGGCGTCTGTGTCAGCGGCAGTCTGTGCGGGGGCTTCGTCTGACGTGTTCGAGGTGTCGGCCCCATCACTTCCATTTTCACT
CTCAGCACTTTCCCCACCTTCCTCACCCGTGGTTGCTGGGCTGTCGTCTGCAGTCTCGGTGTCGATTGATTGCTCACCGTCTCC
GTTGTCGCCTCCATCATCGTCATCGTCGTTGTCGTCGGTGGTTTCACCGCCGAGTTTCCGGGTGATCTCTGGGCCCTCGTCTGT

[0348]

GGCGGGGTCGTTGATCACATACTGGCCGTCTGACTCAAACGCCGGAGAGTAGATCTCGCCTCGTGAGACTGCATCCTGGAGTAT
CTCTTCGGCTTTGCCGAGTGGCTTCGGGTCTGCTGAATAAGATTATCTCATGTATGGTGGTTGCGACTTTGTATGGCTTCGAAC
GCCTGAGTATATTATTTTTATTTGTTAATATTGCAGTTGTAATATAAATAAGACCATGAAGAATACGATACTAGAGATAATTAC
GACGCGTCGAAACCAACCAACACGGAAGCACCAGTATTGCTCTCGATCCGACGGATCAGGATACTGGCACCTCATCTATGAGTG
GGTAGACGGTTCGTGGCAGATCGTAAGCCGCGACCCACTCGACGAGATCGAGATCGACATCTCCAACTCGCGTAGTGATTGCAC
ACCAACCCCACTCGTTGAGTGTCCTGTGTGTGGTCGGACTGGACTTCCCGAACGGATCGCCCAGCACAACTGTACACCAAGGTC
GACACAGTAAGCTAGGAGTCAAAGATACAACCAATATAAACACACCAAGCCATTACCAACAAAAATATATGGTTACGGGTGTTG
TATAACTAGTACTAGTTGTTATTCACGCATCTCCGCTCGCTCTCGTAGGCTGTTTTCGGTGCTATCATAGAGTATAAATTGGCT
ATCGACTCATCTGGTGAGTTCCTCAGTCGACCGCCCGACGCGGCGATCAGTGGTGAGAGGAACAGCAAGACAGCAGAAGAGTGA
GATTGAGAACTAAGATAGAGCACCTCGCACCTCCAAACAGATGGACGACCAGTTCCGAACACGCTCGTGTTTCGTCGGAATCGG
CGCCACCGGTGCCGAGCGCATCGGTGACGGCATCGAACGTCGTGACGCAAGTGCCACGATGGTTCCGGACGAAGTTCCGGACCC
AGCCGACGCCGTCGTCGACTGCGTTGCCGCACCGAGGGCAGGTTGTACTCTCCCAGAACTGCTTGAGTCGCCGCTCGATGGCCG
CATCGAGCGGCTCCGTGTTGATTGTGATCACGCCCAGATCGACAAGCTGGCGAAGCATCCGGCCGAAGGCTGGATCTGAGGAGG
ACATACTCCCTCATTTGACACCGTAGACCGTAGCTATATGGACTTTCCACGTTTCTACATAAGAGCGCAACAGGTTAACCAACA
AATCTATCTTATACTCGTTTTGTTGGTTAACACAAGGCGCTCTTATGTGAACCACCTCGCGGTCAAGCCCCGAGGCACTCGCCT
TGCTTATCTGTAGACGTGGAAATGTTCAAAAAGTGGTTCGATACCCTCAAGAGATCACCAGCCGGCCAACATTCGACGTCCCAT
ACAGGTACTAAGTGACGCTAGCACTGATTGTCAGAAAAAAATTTATGAAACGGTGGTGTGGCCCTCCGATTTAACTCTTGCAAG
GCTAAGTCCCCGTCCTCAAGGAGCAACGCAGGGAGCGTAGCGACCGAGTAGCGCAGTAGGGCGGGGATACAGCCGACACGAAAT
TAGGCACCCCACGAAATCGCACACCCTAACCAAAGAATTACATGATTGTCACCCGATAGTGGAAGTGAACCCACAATGGAATAC
GACCTCGATACGGGAGCGCACTCAACGTACTCCCTGCACTACCACCTCGTCCTCGTCGTGAAATATCGACGCGAGGTGTTCACC
GAAGAACTTCATGTTCACCGAAGAACTTCAAGGGTTCTTCCGCGAGGTCGTGAGCGGGTTCGCTGACAACTACGGCGTCGAAAT
CAAAAACCTCGAAGCCGACCGAGGCCACGTTCACCTCCTGTTCAAAGCCACGCCCACAACCGACCTCACCAAATTCATCAACGT
CCTCAAAGGCGCATCGGCTCGTCGCATCCGCAATGAGTACGGCGAGCAAATCGGGGACAAACTGTGGGGTGACTCGTTCTGGGC
AGATTCGTACTGTCTCATCTCGACGGGACAGGTATCGCTGGACATGCTGAAAGAATACGTCGAGAACCAGCGGGAGCGTGACGA
CGAATGAAACAGACTGCCGAGAAAACGGTCGTGGCAAAAATCTTCACCAACGACCTAACGCAGTCGAAACTTGACGCCATCAAC
CACGAGTATGAGGCGTTCCAGCGATACATTCGTGGTGACGAGGAAGCAGATCTCTACTCGGCCACGAAATTGGGTAAGGACATC
GCGTGCCCACTATCCAGTCTCGGCGCAGTCTGTGAACCTGTGCTCGGACCGAGCGGTGCTGACCAAGCCTCTGCCACGGTATGC
TTGCGAGCCGACTTGGAAGCCCCTGCCTCAACGAAGCGAACCGTTAGGTGAGCGAAGTAGGCAGGGGTAGTTCACTTGAAATTG
CTGAAATGGTGGGGGGTGTTGAGGAGCCTAATATTGGTTCAAGACCCGCTCACCAACAAGACTAGGCGTTCGGTCGATCAGCAC
TGATGCGCTCAAAATCGATATCTGCGAATCGAGTCGTATCCTTGATAACTTCGACTGCGATCTCGGTATCAACTTGCAGCTGGT
ACTCCCATGCTGGCCCACCGATCGAACCTGCCGTCACCTTTTC (SEQ ID NO: 810)
>3300014886|Ga0180300_10007429_2|M
GGCAGGTTGTATTAGACGGATTTTATTTAAAATTGTTTCTGGTAGGTTGTTTTCTCTGGGTCGTTGTGAAGGTTGGAAAATCTG
TTAGGGGTTTATAAAGATCTTCACGGTTTGGTTGGTGGTTTAGAAAGTGTTAAAAATCTTATAGGTTTAGATTTCTTTACTGGG
TGGTGTGTGTTGAAGACTTTTGTGATCTATGATATCTCCGATGATCGGATGCGGGAGGGGTTGAGGGAGCATATTAGGAACTAT
GGTCTTCGGCGGGTTCAGTATAGTGGTTTCCTGGGGGAGATGAGTCCCCATGATCGGTTTGTCCTTTCTCGTGAGGTTGATCGT
GAGGACCTGAGGGCACTTTACAGCGATGTGATGAAAACCTTGCATGATGAGGCCTTTTATGAGGGCGTTTATAACCCAGAGGAG
GTCGAGTATGCTATAAAACAGGTTAAGAAGATGATTGAAGAATTTAAAGGAAGTTAAAACCGCATATTGTTAGGATTCTTCTTT
TTACCTATTCTCGGCGAGAAATCCTCCCACTTTAGTGGGGGGATGAATCGCCGTTTAAGTTTAAATATCAGAAATCTTATATTT
GTTTCGGGCGAGTCGCCCATGAAGGTTAAGAAGACAATTAAGGCTAAGATTCTTGAGCTTCGGAAGGGCAAGGAAGAGCTTTTG
AAACGTGAGTATGAGAATTGGCAACGCTACCTAAGAGGCGATAAAACGGTTCCGCTTTATTCTGCGACCAAGCAACAGGCAGAT
ATACTCAAGCGGGCTATGGGGCTAGTCTCCGTAGCAGGGGCTGGATTGACACAGCCCAGAACTCGGTATGATGAAGCCAGAAGG
CCGAAAAACCGAGAATCTCCCCACTTAAGTGGGAAGAGTGTCAAGAGTTATTCCTTCAAGTTCTAAGATGAGTTTCTCGCATAG
GTTGTAGGTCTTCTTGGCTCTGTTGGTGGCGTCCTCTGCATCGTTCTTGTCCATGACCTCGTTGGGTGGTGTGAGGGTCTTTTC
ATCCCCGTATAGGCTTGGCTCCCTTTTCTTGAAGAGGAGTTTGGAGGATTCGCGTAGGAAGTCTATCTCGGTTTTGAACCAGTC
TGGGAATCAGTCTCTGATGTCGGAGAGAAGGTCTGATACCTCGTGGATTTTGGGGTACTCGATT (SEQ ID NO: 811)
>WP_006630662.1

[0349]

CCCTTCCGGTCAGTTGAACCGTAGATCGAAGCGGAAAATTAAGTCAAGTCCAAATGTTGGTCACCGTCGATACTGGCAGGTAGA
ACAGGTTTATAGCCGGATATTGATGTCTACACACCCCGTTCTAGTTATGGAAGCGACAGCACTGAACGCTACTTTCCTTCGAAG
AAATCATCTATTTGTGCGTTGACAACAGATTTGATTAGATCGCCTTCATTGTCTGCCTCTTCAATCCGGATGTCAGCTCGAAGA
GTTTCAGCAACCACACCCGGATCATCAACAAACGTGTACTCCTTGTGAACGACCTCGTCACGGATCTCGGCGAGTGCTTCGTCG
TCGACGCCGTGGTCGTGTAATCTCGATTCGAGGCGATCGATCGAATCCCGCGCGAGTGCCGCCTCCTCGTCCTCATCAGGCCGA
TACGTCTGTGGATCGCCCATGAAGTGGCCCATCCGTCGGTGGACTTGAACTTCGAGGAGTGTGGGTCCGTTCCTGTCCCGGGCT
CTGCCGATTGCGTCAGCGGCAGTTTCGTAGATGGCAACGGCGTCGTCGCGGTCGACGCGAACACCTGGCATATCAAACCCGGCT
GCTCTGTCTGACCCGTCTTCTACGTCCGTTACAGCGGTAGCAAGGCGAAAGCCCACCCGTTCACGGGTGGGATGAAGCCGACAA
CTGGGAATGTATCCACGTTCGTAGCCACGGCTTGGGGTTGATATTTGCGTAACATCTTTCGATGAGTTACTTCTATATACCGAC
GATGCCTCGCGGGTTCGACCGAGAACGTACCAACATCCACAAACTCCAGTACCACTTCGTCTGGTGTCCGAAGTACCGCAAGTC
GGTACTCAAAGGCGAGGTACGCCATCGACTCGAAGAACTCATCGAGGACAAAGCTAACGAACTCGACCAGGTCAAGGGCTACTC
CTCGCATCACCTCCGTGAGAAGTACGACTTTGGCTTGCCCTCGCTGTGGACGCGCTCGTACTTCGTCTCCAGTGGGGGTGACGT
CTCCAGCGAGGTCATCGAGGAGTATATCGACGCACAAGCGGGTGAATAACGATGCAACGGAACGTCTCAACCACGGTGCGGGTC
AAGCTCCACTCGCTGACCAACAGAAAATCCGATATGCTGGCCCGTGAGTACGAGGCGTTCCAACGCGAGGTTCACGGCGGGGAT
GCCGACCTCTACTCCGCGACGAAACAGCAAGCGGCCTCGGGCAAGTTTTCGAAACCGCTGTCCGAGGCGGGGGCTGTACTGGCA
CAGCCCAAAACGCAGGTCATCGTCGAACGTGACGAGGAACCTGCGAACCTCTCAATTTCTGTGGGCTCAACCCCCAGTGGAGGA
ACCCCACGGCTTTAGCCGTGGGAGGATGTCAGACGGTCTTTGGGCATGCTGATTGCCCAGTCATTATCCTCGACGACGAATACA
ACTGGTAGGTTGTGGACACTCGCGAAATTCAGCGATTCGAGGAACGCACCCTGATCGATCGCTCCCTCACCGAGGAACGCGACT
GCCACTGCGTCAGTATTGCGCTTTTTGGCCGCAAGCCCTGCCCCGGCTGCTGGTGGACATCCCTCAGCGATGATGCC (SEQ ID
NO: 812)
>3300024429|Ga0209991_10029076_1|P
CGAGAAAGCTAAGGAAGCCAGAGAAGCAGAGAAGCTAGCTCAGAAACAAGCCGAACTGGAGGCTAAAGAAAAGGCAAAGAGAGA
GGCCGAGGAGGCAAAAAGAGAGACTGAAGAAGCTAAGAAGACCAAAGCTGAAAGGATGCTAGCCGAGAAGGAAGCCGAGATAAG
CCGTCTGGAAGCGGCATTATCACGCGGAGATATACGACTGGGTAAAGTTATCTATCATTCCTGGAAGCTGGGCTCAGCACGAGC
CACTAAGGAAAAGGCCAGACGAGACGCCGAGAAAGCCAAGCAGGCCAAGGCGGCTCAGCAGTAGGCCGAGCGGGAAGCGGAGCG
GGGCGCTAAGGAGAAGGCGAAGCGAGAGGCTGAAGAGGCTAAGCAGGCACTACAATCGCTGATAGATTTTGACCAAGAGAAACA
AATTGAAGAATGTCTAAAAGAGGTGAAGAGCATTTGTGAAGAACTAAGGGCAGGGAAAGCTGAGACAAAGGAAGCAATACAAAG
GCTTTTAGATATTTCAGAAAGGATTCCCGAGTAGATATTCCCCAGCTAATATCAGTAGGCTGATGTGACTGGGCTTCATCCGTG
GCAACCTATTTGAGTTTGATAATTCACTACCTCAGTGTACATTCAGGCGGACTCCTAAAATTACAGAATCAGGCTGAGAAAGCC
CCCTGCTTCAGCGGGGGGATGAATCAGCCCTTCCCTCCTTTCCCAAAGTATACCAATATATGTTTATGGGACAAGAATATAGAC
ATAACAGACATTCGATCTCTTTGGTCAACTATCATCTCATCTGGATTCCGAGGCGTCGGAGAAAAGTATTGGTGAATCAGGTTA
AGGATAGACTTGATAAACTGCTTAGGCAAAAAGCTGAAGAGCTTGGTATAGGAGTGATAGCTTTGGAGATAAACCCAGACCATG
TTCATCTTTTTGTGAATGCTATGCCGGATTTGGCCATTCATCAAGTGGCCTTTCCTTATCGCGGTCTTTTTCTTTGCCATTCTT
GCAAACTCCACTATAATGCCGACCTTAATGGTGCCCGTAATATTCTAAAGAGGGCTTTGGACTATATGTCCTCAGCCGGGGCTG
CTTTGGCACAGCCCGTAACTCAGTCTCTGAACTGAGAATCCTCCGCCTTCAGACGGGGGAGTGTCAACTGCAGGAACATTACTG
GCAA (SEQ ID NO: 813)
>FZPS01000148_2153|M
GGGCAGAGAAGGTAGATGCAGTAGAGGCAGGAGAGGATTGGCCTACTCATGATTAAGACAGCGACATTCTACATCGCCGATATG
CCGCAGGAGGACAAGCGCCTCTGCGAACGCCTTTGCGCAGAACTCGGAGCGGCGGGCGCGACGATGGTTCGCGAGTGGTATCGC
GCCTGTGACGAAGTGACCGGCGAACGTATC (SEQ ID NO: 814)
>3300013133|Ga0172362_10039269_3|M
AAGCTGTTTCACATGGCAAGAGATTTTTGATTCGAGTCGTGCGGCCAGGTCGGGCTGATTGAAAAGAAGGAGATGGGTTTTTAT
CGCTAACTTTTACCTTAAAACTGCTTCACAACCGAAAATCGCTTAACCAAAGATGGCGAATAGAGAACTGGTGGTCGTGATGCA
GGTGCAGAAAGTCGTCAAGGGGAAGGTCGTGGGGCTGACTCACGCTAAACGCCAGATACTCGACCACGAATACGAAGGGCTCCA
GCGATTCTTGCAGGGGGATTCCAACGTTGAACTCGCCTTCAACATAGGCAGACGAGCCTTGGGGCAAGTCTCCAAGGTAGGGGC
TGTCGTGAGCCAGCCCGTAACTGGGGCAGTGCCGTTCAGCGAACTCGGTAGCGTCGTTCAATGTCCGACCCCAGAAGCCACCCT
TTTTAAAGGGTGGTAGTTCACGTGCGCGAGCGCCTGCGGCGTTTAGTCCAATCTACTAATAGGTGCTCTCCCATCTCCCTAAAT

[0350]

TTCCTCGCCGCTGGAGAACGCGGGGCATAGCGAAGAACTGGGCTGCCAGCGGTCAAGCTCTCACGAACTTTATCGTCATCGGGT
TGCCACTCAATTGGTGCTTGGACAGCCCCAGTCCATGTTTCAATCAGATTCCACCCAATTGGAGCTTGCACCGTGCCAATCCAA
GTTTCAATCAGTTGCCATGCGGCAGTGACCGTGACTGTCCCAGTCCAGGTTTCGATGAGTTTCCATTCACCAGTGGCAGTGACT
GTTCCAGTCCAAGTTTCAATCTCGTACCACTGAGCAGCTGCTGGTGCTTGGACCGCCCCGGTCCATGACTCGGCGAGTTGCCAT
TCAGCCGGAGCACTTACTGTCCCAGTCCAAGAATCGATCTGCTGCCACTGTCCCGTGGTTGCATTAACAGTCCCGCTCCAAGAC
TCGACCTGACGCCACTGACCAGTGGTCGCGTTGACCGTTCCAGTCCAACTCTCAATCTGCCGCCACTGTCCCGTGGTCGCGTTC
ACAGTCCCAGTCCAGCTCTCAATCTCGTACCACTGAACAGCTGCTGGTGCTTGGACCGTCCCGGCCCATGACTCGATGAGTTGC
CATGCAGCAGGGGCGGTCACCGTGCCCGTCCAGCTCTCGACGAGCTGCCACGCGGCCGGGGCGCTCACGGTGCCAGTCCATGTT
TCGATGATCTGCCAGACTGCGGAAGCATTGACCGTCCCCATCCACGACTCCACCAATTGCCATTCAGCTGGCGCCTGAACGGTT
CCAGTCCACGTCTCTATGAGCTGCCACGCTGCAGGTGCATTAATAGTACCCGTCCATGTCTCGATCACTTGCCAGACGGCAGGC
GCTGTCACGGTCCCAGTCCAGGTTTCGATTGTTTGCCAAGCCGCGGGTGCGCTGACCGTTCCAGTCCAAGTCTCGATTAGCTGC
CATTGGGCAGGAGGTGGAACCACCGTGAACTGCCAAGTGTTTTCAGAAACATTCTCACCCGCAGCGTTGACCGCCCACACTCTC
CAGTAGTAGGTCCCCGGGGTGGAGAAGTTGCGCTGGTAGATGTTGTCGGTGTTGTCTTGGTAGTTGTCCACGAGCGGCGTGGAG
AAGGTGGGCTCGTTGTCAATCTCGATCCTATGGTTCTCCCCAAACCACCCGTTGTCCCACTCGAAGGTCACAGTTTCCCCGACC
CCGAAGACAGAGTTGTTCTCTGGACTGTAGAGCCTGGGCCTCTTCGGAAAGAATGTGTACTCGTATGTGAAAAAGATGGCACCG
GTAAGAGCGGCGCCTGTTCCCATGTCGTATTCAAAGTTGTAGGTCAGGTCGGCAAGGGTCCTATCTACAACAACGTCGATGTTG
TCTTTGTTTACAGAGATGGATGTCAAGGCCTCCCCAGAGCTGCCTGTGCTCATACTGACGGCTGAGCCCCAGTTTACCCCATCA
GCTTGAATCCTGGTCTGTCGCGTTTCCGAAGCGGTTGCACTTCCTATCCAATATGCCGTGAGGCCAACGCTCTTCACAGTCTTT
TTGTTTTCCGGGGCATAGAACGTCAGCTCCCCGCTGGTCCAAGGGTCACCGGCAGCCCCGCTGGGTTTTTGGCCCGCGGTGAAC
ATCGTGGTTTTCAGCTTTGTCGCGGCCGATGGGCGGAAGTGATAGGTGATGTGCAAAAGGCCGCTTACCTCATAGAGGTCGGAG
GCCGCAGCGCTCTGGATGTTGAGGTTCACATTGTCCCCGTTCGACCAGTTCTCCCTCATGAAGGCGCTGATGTCGTGGAAGTAG
TCGAACCCCTCTACATGCTCGCTGCCTATGTTGGTCGTGTAAGCTCTCAAGGTCTGGGTTTTCGCCCCGACGCTCGCGTACACG
CTCGTGCTGCCGTCACTAGTCAAGATGGTGTGGAAAATTTCAAAGTACGCGTTCTCGATGTTGTCCACGCGAGAGCTGTCCAGG
TCCGTCAGATAAACGGGCGTCCCGAAGGCGAACAGGTTGGCCGAGCCGTTGACGTTGGCGGCCCCCATGGCAAACGTGACTGTT
TTTATCTGGTCCGCGCTGTTGTCTGAATACTCATATGTGACGTAAGTTTCACAGCTGAGCGCATCGTTTGTTTGTCCTCCGGTA
GTAATTGTAATATCCTGATCAACGTTGTAGTTCCAGGAAGCCGTCGGCCGGGAAATATACCAGAAATTGAGATCGTTGCCGGTA
CGGGCTTGCTCTGTCCAAGCATTTGTGGTGGCTGTATCGCTTCCGATCTTGATGTTAAGAGAAAAGTCGGTTGAAGAGGCTTGG
GAAGAGTGGGCATGGATTTCGTACCAGTTGTCCCTCAGCGTGACCCCGTCTTCTGGGATTTCGATTCTGTGTGTGTAGGTCGCC
GTGCCGCCAGCCCCTATCTCGGTGTTCGTGCTCCCCACAAAATACCTGACAGTCTTGATCTTCGTGTCCGCCAGCTCGTTCTCC
TCATATGTGATGTACATCTTGGCCGTGTGGCTCTGGGTATCCGCGTCAGTGACGGTTAATTTTGGGTAGACGGCGGCCCCGTTT
GTCCAGTATGTTGTAAAAGCAGAGGTAAAATCTGCGACCACCCATGCCGCCACGCTTTCCCCGCTGTTGGCGGTGTACTGGCCG
GTCACTGGGTCTGCGTCTGGGTTCCGGTCGGTCCCCGCGATGTTCAGGGTGACGGTCATATCTGTCACGTTGGCAGCAGTCGAT
GCCCCCGTCTGGACGTGATACATCACGAACACCTTCCTGAACGGCCCAGCCTTGTCCGGGATGTAGATGGTTTTTGTCGGGGTG
AACGAGTAGGCCGTGCCAGTGCCCAAAGCGGACGATGAGGTTCCTTGCCTCGTATCCCATGAATATTGGATTGTGTTTATGGTC
ACTATATTTTCGTTGGGTTTGGTGGGGGAGGGAGTAGCCACGGCTAGGGGCTGCATCGAGATAACGAAAACGCTGGCTAACATC
ATCGCTATCAATACGAATACAAAGATTTGTTTTGCCTCCTTCCGCCCGAACTCTTCCCGCCCCAACCTAAACAAGCTACAACCC
ATCTCCCTCACCTGCTTGCCTCCACTTGGACTATGGAATGGGCGCGGCTCACGGCAGAGCCATCTCTCATCTCTCTCAGCAGGT
TCGTGAAGTACTCAATCCCGCTCGCGGTGAGCTTGCAGAGCTTTCGAGGCCTCCCCCGGGTTAAACGGGGCTC (SEQ ID NO:
815)
>WP_121597521.1
GCGAAGCGAACGAACGAGGAGCGCAGCGAGTGTGCGCCGCCCTCGCGGCTGGGGCTTTGGCGGTGTTTTCTGTCGATCCGTTGG
CGGCTATTTATAAATGAGCAGCTGGGGCTTTGGCGGTGTTTTCTGTCGATCCGTTGGCGGCTATTTATAAATGAGCAGCTGGGG
CTTTGGCGGTGTTCCCCGCCGATCCAGAATCAGCAATTTATAAGCGAGCAGATAGGGATATAGCGGTGCTCTCTGCCATTCCGC
AGTCAGCCATTTATAAGCAGTCGGCGGTAGTTCAGTACGTCTCTTTCCCCCACCGCAGGTCGTCGTAGGTGGCCTGCTGGTCGC
TGCCGAGCCGCTCCTCGAACTTCCGGCGGAGCGCGGCCTTCTCGGTCGCGGAGATCCGCGCCAGCTCGTCCGCCTTCGAAACCG
GCTCCGGCGGGCCCAGCCCGCGACCGACGGCGACGTCGACGAGGCGCTTCGCCTGGTCGGGCGAGAACCACTCCCGCTCTAACG
ACAGCGCGACGACGAACTCCCCCTCCCCGAGCCGGTCCTGGGCCCGGGTCTTGAACGGGGCGGCGACGGCGACCTCCAGACTCA

[0351]

TGTGCGAATCGGGGGGCGCCGACGGCGGTAAACCTACCGGACCGCATCCCGCTTCAGGGCGGCCGCGACGTCCGGCTTATAGCG
GTAGCGAGGCGAAAGCCCACCCGTTTACGGGTGGGATGAAGTCGACAACTGGGAATCTATCCACGCTCGTAGCCACGGCTGGGG
TTGGATGCTTGTGCAACATATTTAGGTGAAGATACTCTATATACTGACGATGCCTCGTGGGTTCGACAGGGAACGTACCAACGT
CCACAAACTCCAGTACCACTTCATCTGGTGTTCGAAATACCGTAAATCGGTGCTTGAGGGCGAGGTACGCGACCGTCTCGAAGA
ACTCATCGAGGAGAAAGCCGACGAACTCGACCAGGTCAAGGGCTACTCCTCGCGCCGACTTCGGGACGAGTTCGACTTCGGCTT
GCCGTCGCTGTGGACGCGCTCGTACTTCGTCTCAAGCGCGGGCGACGTGTCCAGCGAAGTCATCGAGGAGTACATCGACACCCA
AGCAGGTGAATAGTCATGCAACGGAACGTCTCGACCACGGTGCGAGTCAAACTCCACTCGCTGACCAACAGGAAAGCCGACCTA
CTCGCCCGCGAGTATGAGGCGTTCCAGACCGCAGTTCACGGCGGAGACGCTGACCTCTACTCCGCGACCGACCAGCAGGCGGCC
TTGGGCAAGTTCTCGAAACCGCTGTCCGAGGCGGGGGCTGTACTGGCACAGCCCGAAACGCAGGTCATCGTCCCACGTGACGAC
GAACCTGCGAACCTCTCCGTGTCCGTGGGTTCAACCCCCAGTGGGGGAACCCCACGGCTTTAGCCGTGGGAGGATGTCAGGCCC
GGCAACCGTTATGTCGGGCCAGCGAGATGGCGCACGTATGAAGACCCGCCCGCTCGGCGACACCGGCCACGACAGCACCGTCAT
GACGTTCGGCGCGATCGCGCTCAACTGGCTCGAACAGGAGGGCGCGAACCAGCTGGTCGAACACGTCCTCGACTACGGGGTCAA
TCACTTCGACGTGGCCCCG (SEQ ID NO: 816)
>3300027896|Ga0209777_10000273_32|M
TCCGGGGTGTATCTTGCTTCTGGTTTGTATGTGAATGCAGGGACCATGTCTGCGAATGTGTCTGGGAATTGGATTGAGGCGCAT
GTGATGTCTGGTGCTTTGACCATCATCTCTGGTGGCGTGTTTGTCTCTGGTTCTATCACAGTGATTTCTGGTTTGATTGGTGCG
TATGTTTCTGGTTTTGTCACTGTGCAGAGTGGTGTGATTCAGGTTCTATCTGGTGTAGGTGTTCTTACTTCTGTTTCTGGGAAT
ATTGTTTCTATTGCTTCTGGGTTGTATGTGACTATTCAGTCTGGTGGTGCTGTTTCTGTTCCTATCACTTCTGGTTCTGTGATT
GTGTCTGGGAATATTGGTGTCTCTGGCTGGGTGAGTGTGCAGAGTGGTGTGATTCAGGTTCTATCTGGTGTAGGTGTTCTTACT
TCTGTGTCTGGGAATATTGTTTCTGTCGCTTCTGGGTTGTATCTTGAATCTGGTCAAGAGATTATCTGGCGTGGTGTGCAAGTC
TCTGGCACTGTCGCTGTCTCTGGTTTGATTGGTGTGCAATCTGGTGAGGTTCATGTGATGTCTGGTGCTTTGACCATCATCTCT
GGTGGCGTGTTTGTCTCTGGTTCTATCACAGTGATTTCTGGTTTGATTGGTGCGTATGTTTCTGGTTTTGTCACTGTGCAGAGT
GGTGTGATTCAGGTTCTGTCTGGTGCGGGTGTGGTTGCTTCTGTTTCTGGGAATATTGTTTCTGTTGCTTCTGGGTTGTATCTT
GCTTCTGGGTTGTATGTGAATGCGGGGACCATGTCTGCGAATGTGTCTGGGAACTGGGTTGAGACGCATGTGATGTCTGGGAAT
CTTCAGGTTTTACAGACATCTGGTTTTGTTGCTCTTCAATCTGGTGAAATTCATATAATGAGTGGTTCAGTGACTATACCCGCA
TTGACAACAGTTGTTGCCACCCCAGCAACAAACCCCATGCTGATTACCGGTCTTTCAGGAGGACAAATTTTGACTTCAAGTGTA
GTAATAAGCGCTATTGTGAAAGCAGTATCACTAAACAGTAGTGATATGTATATTGGTGGAGCATCGGGGGTAAATAGCCCTCCG
TATTCTGGTGGAGGGTTTGTGCTTGCTCCTGGTGAAGCTATTAATCTTGATGTCGCAAATATCGGGTATGTGCGGGCCTGTGCA
GTGATGAGTGGTGATAAAGTATCATGGATTGGGATAGTATAAGAGGCAAATATGGTTAAAGAATTTCTCTTTGAACAGTTTTAC
ACAGGGACAACAGGTGTCGATACAGCGTATGGAGTAAATTGGAGGGCAACGACATTCACGATAGGGACTGTTGGCGCGAAAGCG
TCTCATTATGCGACAAGCATAGGATGGCGAGGGCAAGATTTTCCAACCATTCGTTTCAGCGATAATGTTGAAACAGTCCTTGAG
GAAAATTTCAATATCTCCGACCAAGACTTGTTTTCTATATTTCGGGCAGAAAACAACATGGTAATTTAGATTGTATACTGCGGA
ACTTGACCTGCCCATAAATCTTAAATACCTATAGGTGTATATAAGATTTGTGGCACAGAAAACCATCACCGCAAAAATCGTAGC
AATGACCGCTACGAAACAACAAATACTCCAAACCGAGTATGACAACCTCCAGCATTTTCTACAAACTGGAGAAGACCTTGGAGT
ATACTCAGCGAACAAACAACAAGCCAAAAGAACAAGTTTCAATGCCGACTACAATGGTGCATTAAATATCTTGAAGAGGGCAAC
ATCCTACATGGATATTGCTGGGGCTGTGAGTGAACCAGCCCGTGACTGGGGCAACGACGACCAGACCTCAGAAGCCACCCAACT
TTTTGAGTGGTAGTTCACTAATGTAACTGGGAATAAGAAAATCATTGAAGTATTTGGTGAACGATGGCACACTGGGGAAAAGAT
TCCATATAGACGAACTGAGATTGGGAGGAAAAAAATATTCGAAGAAAATGGGTATGCTCTTTTGATAATATGGGATTATGAATT
AATTGATGAACAAAAAGTTATAGAAAAAATAAAAGATTTTGTCTAAATTAGACAAAATTAAATCCTTCCTGTTTCGAGTACTAT
TGTTGTTGTGGTTGTGATTCCCGAGTGAGTATATGTTTTAGTAACAGTTCCATTCACAGTGACGCTTTGACTCCAATATGAACC
AGCGAGTTCGATTCCACCCACGAGCAGATGAGCGTAGAG (SEQ ID NO: 817)
>WP_121304574.1
TACCCCAACACCGGCGCCAACAGAAGCACCAACTGCGACCCCGACGCCGACTCCAACTCCAGATCCAGTAATCGTTGAAGTCGG
CGAATGGGGGAAGATAGAGCAGAACGGTACAGTGGTTCGGTTTCGAGTGGTTAGCTATGACATCACGGAATCTGTTGAAGAAGC
CGATGGAGAGACAGAAGAGCCATCTGAAGGGTATAAATTTGTAGTTGCAGACGTACATATCTCTACTGATCCCGGCAATGAAGT
TACGATTCGAAAAGACCAGTGGTCGTTTACTGAACTTACCGATAAGAAGCGCGGCCCTGATGGCGTGACCAACGGAATTAGAGA
GCCGTTCCCAAACAAGACTGCATTACAGGGTGATTCGGTTCGAGGAACTATTGTTTGGGAAGTAAAATATCCGTCTGATGGGCA

[0352]

GTTCGAGGTGGAGATACATAAAAACAAAACAGGCAGGGCTGGTAAAGTGTTAATATAACAGAGTAGCTTGATAAGAGTTTTACA
TATTATCGTGGATAATTTATGCTCGAAAGGGGTTCAAAAACAGTTCTGCTCGGAGATAATTTTTGAGGGCTGACTGTAGTTGTG
GAAGTGGTCCATCGTGCCGGGTGATGCTCTGTATAATTATACCATGTAGCATCGCTAGGAGAAATTCAGCGGTTTCTTCGGGGG
AGAGAGTCACCGCCGAAGCGACCGGAGTTCCCGGCAGAAGGCCGGTGACGATGACTCGGCTATGGATGGGGCCTCTTTGATCGG
GTCACAAGACAGCCACGCTGATGCTGAACCCCAGCAGACGACGCGTGGAACGTATGCGTCTTGAAACCAACAGGCCGCTACACT
CGGCCTGAAATCCCGTGGTGGGACTCCTCCACGTTCACGCGGAGGAGGAGGTCAATTTGCCGATCTTGTTTGCACGCATTCAGT
GAGACAATGCGAGAAACCGCCCGGTTGCAGTCCAAACTACGAGAAGGTCGATACAATAGCGAGCGTATCCGGCGGTTGTACCGG
AAACAGACTCGTCGTCACGACACACCCGAGGAGCACTGTGTCATGACCTGCTGGAACAACCTGTACGCCGAGGGTATCAACATC
GTGTATATCGGTGACTTGACCGACGTGCTGGATACGCACTGGTCGGTCGAAAGCAACGCCAAGACCCAGAACTTTTGGGCGCTC
AAGAAGTTTCAGTCGATCACATCGAAAGCTCACAAACCCTTCCAGTAGAAGTTGAATCTGTGTCTACGGTGCCTGGCCAGTTCT
CGAAGTGAGACGGACAGTGAGTTTTGGCGGTGAGGAGTCGCTGTCGGCGGCGGATCGCCGCCGACGCGACAGTGTCGCCACTCA
CGCCCGCTGTCCTAGCTGGGACAACTACCTGAAGAACCGGTCGTCAAGTGGTTGGTTCGCTGGAACAGACCTGCGGACACCAAG
TGCCGTCCAAGCTGCGCGAAGCGTCATTTCGCAGAACTCGGTGAACGACCACTGTAAGAGGCGGCGCCCCCCGCGGCGGGGCGC
CGCCAGACTGCACATCCTCTCCAAGCAGACGCTTACTCTAACCCGATGTGATCGACTGAGTTTACAGCGGTAGCGAGGCGAAAG
CCCATCGGCTTTAGCCGTGGGATGAAGCCGACAACTGGGAATCTATCCACGCTTGTAGCCACGACTGGGGTTAGATGTTTGCGT
AACATCTTTAGCTGAGTGATATCTATATACTGACGATGCCTCGTGGGTTCGACAGGGAACGTACCAGCGTCCACAAACTCCAGT
ACCACTTCATCTGGTGTCCGAAATACCGTAAATCGGTGCTTGAGGGCGAGGTACGTACCCGCCTCGGGGAACTCATCGAAGAGA
AAGCCGACGAGTTCGGGCAGGTCAAGGGCTACTCCTCACGCCGACTTCGGGACGAGTTCGACTTCGGTTTGCCGTCGCTGTGGA
CGCGCTCGTACTTCGTCTCAAGTGCGGGCGACGTATCCAGCGAAGTCATCGAGGAGTACATCGACGCCCAAGCAGGTGAGTAGC
CATGAAACGGAACGTCTCAACCACGGTTCGGGCCAAACTCCACTCACTGACAAACAGAAAGACTGATCTGCTCGCCCGTGAGTA
TGAGGCGTTCCAGACTGAGGTACACGGTGGAGACGCTGACCTGTACTCTGCAACCGACCAGCAGGCGGCCTTGGGCAAGTTCTT
GAAACCGCTGTCCGAGGCGGGGGCTGTACTGGCACAGTCCGAAACGCAGGTCATCGTCCAACGCGACGACGAACCTGCGAACCT
CTCTGTTTCCGTGGATTCAACCCCCAGCGGGGGAACCCCACGGCTTTAGCCGTGTGGAGGATGTCAGAAGCGGTCTAGGAGAAT
GCCCCGGGGCTTGACCCCGGGGTTGAATCCGATAAGCAGGAGTACAAATCATTAAGTGAACATATCCCGAATCAAGAGACAGCG
ATGGAGTACAGTCACCGCTACCCCGCGTACCCAACACAAGAGGTGGCGGCTGAACTGGAACGACACATCGACGCTCATCGCCAA
GCGTACAACTACACCCTGTACAAGTACGAGAACGTGGACGCCGACAATATCGGTTCCGCGTACAAA (SEQ ID NO: 818)
>3300027941|Ga0207719_100292_34|M
ACCCTGAGCCTAGGCGTCGTCGGCGTGATCGCCGGCCCGCTCGCCATCGCCCTGCTCGTCGAAGTGGTCGAACTGATCACCGAC
GAACGGGTCAGCATCCAGCAGACCTTCCAGGAGAGCGCCACCGAGGCCGCCGGCGACGACGACTGAGCGGCGGCCGTACCGACG
CGAGTGCGGGCGGCGTTACAGCGGTAGCGAGGCGAAAGCTCACCCGTTTACGGGTGGGATGAAGCCGACAATTGGGAATCTATC
CACGTTCGTAGCCGCGGCTGGGGTTGGATACTTGCATAACATCTTTAGGTGAATAGTATCTATATACTGACGATGCCTCGTGGG
TTCGACAGGGAACGTACCAGCGTCCACAAACTCCAGTATCACTTCATTTGGTGTCCGAAGTACCGCAAATCGGTGCTTGAGGGT
GAGGTACGTGACCGTCTCGAAGAACTCATTGAGGAGAAAGGAGACGAACTCGGCCAGGTCAAGGGCTACTCCTCGCGCCACCTC
CGCGACGAGTTCGACTTCGGTCTTCCCTCGCTGTGGACGCGCTCGTACTTCGTCTCAACTGCGGGCGACGTATCCAGCGAAGTC
ATCGAGGAGTACATCGACGCCCAAGCAGGTGAGTAGTCTTGCAACAAACCGTCTCGACCACGGTGCGGGTCAAACTCCCCTCGC
TGACCAACAGGAAGGCAGACCTGCTCGCCCGTGAGTACGAGGCGTTCCAGACCGAAGTTCACGGCGGCGATGCTGACCTCTACT
CCGCGACTAAGCAACAGGCGGCCTTGGGCAAGTTCTCGAAACCGCTGTCCGAGGCGGGGGCTGTACTGGCACAGTCCGAAACGC
AGGTCATCGTCCACCGTGACGAGGAACCTGCGAACCTCTCCGTATCCGTGGATTCAACCCCCAGTGGGGGAACCCCACGGCTTT
AGCCGTGGGAGGATGTCAGTCGCCCGTCTCGGGTTCGCGCCCCGCGGCCTTCCACTCGGTCAGACTCCCCTCGTAGAAGGCGAG
GTTCTCGTACCCGAGGTGTCGGAGGACGACGTAGGTGTGACTGATCCGGCGAGCGGTGTTGCAGTACAGCAGAACGCGCCGGGT
CGCTGAGCGGGATATTCTCACTCTCCGCACCGCCCGCGAGAATAGGGCCGCGCTCAGTCGACCACGCAAGCGCGAGCGCGCCTC
GCCCTTTCAGTCCACCAGGACCGCACCGCTCCGCCCCGCACCGCCACCGCAGACAGCCTCACACCTCCCCAGCCTCCTCGCGTC
TCGTCGCTCGCGTGACTCGCTCCTCGGTGCTCGTCCCTCGCACGGTGCGTCGCGGCACGGAGGCCGCGACGGCGCGCGCCACCG
CAATCGATCGGTGTCGGAACGGCCGCGAACGCACCTCTCAAGAGCCTGCGGGCCGTGGCGATCGGTAATGAGCGACACCGGCGG
CCCACTCTCACCGGACAGACCGGACGCCGAGAGCGATTTCCGGGTGGACGCCCCCTTCGACCCCGCGGGCGACCAGCCCGAGGC
CATCGAGCAACTGGTCGACGGCTTCGAACGCCAGGGCCACGACAAACAG (SEQ ID NO: 819)
>SRR8526178_megahit_k177_763700_3|P

[0353]

ATGGTCACGGTTCGGATCAATGAAGTCACCAAGTCAAAGCTCGACAAGTTGTGCGAGGACAACGGCCTCAAGATGCGGTGGGTT
GTTGAACAGGCGATCCTTCAATATTGTGAGATGCACGAGTCGAAACATCGTCCAGAGGATGGGTAACTTATGTCACTCAAAACA
CTCACATTCTGGCCGATTGGTATGTCTGCAAAATACAGGCGGATATTTGACAACATCCGTAAGCAACTTGGGGCGGCGCAGCGC
GAAATGGTGAAAGCATGGCTACGGCTGGCGTTGGCAGAGGATGCTTATGGTGACCTTAATGCTGCGAATAACATCGCGCGGGGG
ATGCAGACGTTTGGGCCTTGTGCCAAGTCCTCTGCGTCAGACGACGCTATCGGGCTGCGGTTCGATAGCGAGGGGCAAGAATCT
GCCCTGGGGCAGGCGAAGCGATTGTTTGCTTGAGCTTGCAAGTTCTTGTGTGTGTGTGGGTATGAACTGCTGAGCGTTTGTGTA
GT (SEQ ID NO: 820)
>3300028675|Ga0272445_1001731_11|M
GTCGTAACGATGGCTGATGTGTTAGAGGGAGGGGCTGTGGGGATCAATGACAATAAAATGAACACTCTCGTATTAATGCATGCT
AACCCACTATGGGGCGAGATTGAGATAAAAACTCCCGCAGGGCGAACGGTTTGGCGTGCTCCTTAACCGTTGGTGACCTGAATG
TGGGAAACAATCTTGATTGGCGTTTGTGGGGCGCACTTTCCCGCTACCGTAGGGGCGACAAAAAGCAGCACTATTCTGGCTTTC
CGTCGCCATCATCGCTATCATCGCTCTCATAGCCATCAACACTTTGACCGTCCGCTGGCTCGGTCATCCGCCACAAAACTCATC
CACAACAATAGCACTGTCAGAGTGAAAGGACTTAGCAGCGTTCGTTAAAGCGCCGACCTGCAGCGCCGAAAGACACTTGCTAAG
CCTTACCGTCGGCTTGGTCACTTCCCGTTTCCCGTAGCGCGTGAAGCCAGAATGGTAAAGCCTAAAATGCCGTTTTTGCGTTGC
GACAACGCGCAAGTTAGCGTCAGCAGGCGCACATAGGCACAGGCGCACGGGCGTAGGCACGCAGGCGCACGCACGCGGGCACGG
GCGCGCTTTCAACGGGGATACCCGAGCGCGCAGGCGGGTCAGGCTGCCCGCCGACCTTAAAGTCAACCTTAAAGACTACAACCC
CCAAACCGACGCTATGTGCGCTATGCTTTGGGACGCTTATAAAGTTCAGGTGCTTAAGGAACTTGTTGGCGAGGATGAGGGCGA
CTACATTAAGTAAAGCCGTGAGGAGTTATTAGAGGCTTTGCGGGAACGGGAGCACAAGTATGGTTATAAAGCGCCCCTTAATGC
GAACTTAACAAAACCGTCCCGACCAGTTATTTCGGGTAAAATTTTAACCCATTTGCCCAAATTCCGCAACTTCGTTGCAGCAAA
CCTCCAGTCACCCCTTTCCCCTAAGGGGTTTGGCGCAAGCAATTCTTGAAATTCCTGCCAAAATCGGCAAAATTCAAAGGTCCC
GTTCTTTGATGCCTATCACGGCAGAATAAAGGGGGAGGTTGACGGCAAATGAAAAAAGTGGTCATAACCATCACTACTGGAGTA
GAGTTTCTACAACCTCAAGATGAGCAAATCGTGAGGAAGGCAGCGGAATGGTATACAAAGGCACGAGAAGCCCTTTTTGCCCAT
GCCCTCCAACAACAACAAGGTAAAGTTGAGGGAAACCGATTTTGTTGCACAAAGTGTAAATTTCAGCACAATGCGGACTTTGTG
GCGGCGTATAACATTGCCGTAAGGGCATTGAAGAAGTTAGGGATGAGGGAAGAGATAGGTTGGAAGGTATTGACAGGCAAGCCG
AAGGTGTTATAATGACGATGGCTCATGGGGAGCAGGCTCCCCATGGGCAGCGTGCCCGCCGAAAGGTGGGTGGAGGGCAGAGCC
TGTCTGCCTGTAAGCAGCACCTTCCCACATGGGGAGGTTGTCTGCAAGTTCACATTATCCCCTGACCCCAGCGGTGAAGAGCCG
CTGGGGTTTCAATTGTCTTAGCACCCTGATGTCCACATTAACCCCTTAAGACAGAAAACTTGCCCAAGCGAGGATTAGCACGAA
CTATTGCCTTGTAACGCTCCCCTTAGCAGTTGCACATTGCAAAGCAAGTTTCTTGGCATTAAGAGACTCTTTCAGTCCCACAAT
GGTTTGATTGCTACATTGCCTGCCAAGATAAAGCCCTTCCTTTCCATTATCACTTGTAAAGGACCAGTGTCATAATTCCTCAGG
GGTGAGGGAGATGAAGGTCATTATTGGGCGTGTAATGGAAGACACTTACACGGTCAAAAGTCAAGAACCCTTGCCACCAGGACA
GGAAGTCTTCATTGTAACTAAGATGGACGCTGAGGATGTCTATTACCTTGCAAAAACTTTGGCGTGG (SEQ ID NO: 821)

[0354] 实例6-工程化CLUST.196682 CRISPR-Cas系统的功能验证(图13A、13B和14)

[0355] 鉴定了CLUST.196682 CRISPR-Cas系统的最小组分,我们从命名为3300025638的宏基因组源选择一个基因座用于功能验证。

[0356] DNA合成和效应子文库克隆

[0357] 为了测试示例性CLUST.196682 CRISPR-Cas系统的活性,我们设计并合成含有pET28a(+)载体的系统。简而言之,合成编码CLUST.196682 3300025638效应子(表8中提供为SEQ ID NO:601的氨基酸序列)的大肠杆菌密码子优化的核酸序列(金斯瑞公司)并且将其克隆至来源于pET-28a(+)(EMD-密理博公司)的定制表达系统中。载体包含在lac启动子控制下编码CLUST.196682 3300025638效应子蛋白的核酸和大肠杆菌核糖体结合序列。载体还包含在CLUST.196682 3300025638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。

[0358] 我们计算设计了含有“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质粒以及大肠杆菌必需基因的序列。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。

[0359] 我们接下来使用金门组装方法将重复-间隔子-重复文库克隆至质粒中。简而言之,我们首先使用独特的PCR引物将每种重复-间隔子-重复从OLS池(安捷伦基因组学)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)纯化,之后添加至金门组装主混合物(新英格兰实验室)并且根据制造商的说明书孵育。我们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。

[0360] 按照卢西根公司推荐的方案使用基因Pulser Xcell®(伯乐公司)将含有不同的重复-间隔子-重复元件和Cas蛋白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)中的含有氯霉素(飞世尔公司)、四环素(阿法埃莎公司)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)提取质粒DNA,以产生"输出"文库。通过使用含有与依诺米那测序化学兼容的条形码和位点的定制引物执行PCR,我们从转化前"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。

[0361] 细菌筛选测序分析

[0362] 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。

[0363] 为了鉴定导致酶活性和细菌细胞死亡的特定参数,我们使用下一代测序(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。我们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0.1(超过10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,我们取给定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)。我们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。我们研究了此矩阵中不同特征解释CLUST.196682系统靶标耗减的程度,从而得到了单一筛选内功能参数的广泛调查。

[0364] 图13A-B描绘了靶向pACYC184和大肠杆菌E.Cloni必需基因的CLUST.196682 3300025638效应子的强耗减靶标的位置。值得注意的是,强耗减靶标的位置似乎分散在E.Cloni必需基因的整个潜在靶标空间。

[0365] 图14描绘了侧接耗减靶标的序列的weblogo，其指示出CG的突出5’PAM。

[0366] 实例7-对CLUST.089537 CRISPR-Cas系统的最小组分的鉴定(图15-18)

[0367] 此蛋白质家族描述了与在菲西芬氏菌目(Phycisphaerales)和浮霉菌门(Planctomycetes)生物以及从淡水、永久冻土、沉积物、土壤、厌氧和热泉环境收集的未培养宏基因组序列中发现的CRISPR系统相关联的大型单效应子(表11)。CLUST.089537效应子包括下表11和12中详述的蛋白质实例。这些系统的直接重复序列的实例示出于表13中。

[0368] • 含有此效应子复合物的天然存在的基因座的实例描绘于图15中，该图表明对于CLUST.089537基因座，效应子蛋白与CRISPR阵列共现。在双向15kb窗口内没有鉴定出与效应子蛋白或CRISPR阵列共现的其他大蛋白质家族。

[0369] • CLUST.089537的直接重复序列展现出多个保守基序，包括靠近3’端的5’-RBBNRBKGACAS-3’(SEQ ID NO:834)(图16)，其中R是指A或G，B是指C或G或T或U，K是指G或T或U，S是指C或G，并且N是指任何核碱基。

[0370] • 图17是CLUST.089537效应子的系统发育树的示意图，该示意图显示该家族展现出序列多样性。

[0371] • 图18是CLUST.089537效应子蛋白的多序列比对的示意图，该示意图揭示出由颜色品红色、蓝色和红色指示的RuvC结构域的保守催化残基的位置。

[0372] • 任选地，该系统包括作为表14中列出的非编码序列的子集的tracrRNA。

[0373] 表11.代表性CLUST.089537效应子蛋白

[0374]

物种	效应子登录号	间隔子#	cas1	cas2	效应子大小
生物反应器宏基因组（CAAACX010000652）	CAAACX010000652_28|M	7	Y	N	674
激活的污泥宏基因组（CAAACH010011569）	CAAACH010011569_2|M	3	Y	Y	759
激活的污泥宏基因组（CAAACI010012331）	CAAACI010012331_3|M	2	Y	Y	759
激活的污泥宏基因组（UZXW01002767）	UZXW01002767_11|M	2	Y	N	758
水生-淡水-淡水湖（3300009161|Ga0114966_10008307）	3300009161|Ga0114966_10008307_8|M	4	Y	N	728
水生-淡水-淡水湖（3300009161|Ga0114966_10008307）	3300009161|Ga0114966_10008307_7|P	4	Y	N	762
水生-淡水-淡水湖（3300010885|Ga0133913_10083255）	3300010885|Ga0133913_10083255_2|P	4	Y	N	762
水生-淡水-淡水湖（3300020109|Ga0194112_10008260）	3300020109|Ga0194112_10008260_1|P	2	Y	N	676
水生-淡水-淡水湖（3300020109|Ga0194112_10008260）	3300020109|Ga0194112_10008260_1|M	2	Y	N	678
水生-淡水-淡水湖（3300020222|Ga0194125_10021011）	3300020222|Ga0194125_10021011_3|M	4	Y	N	687
水生-淡水-淡水湖（3300027770|Ga0209086_10006474）	3300027770|Ga0209086_10006474_1|P	3	Y	N	762
水生-淡水-淡水微生物垫（3300015360|Ga0163144_10184575）	3300015360|Ga0163144_10184575_1|M	2	Y	N	779
水生-淡水-沉积物（3300013130|Ga0172363_10020541）	3300013130|Ga0172363_10020541_1|M	27	Y	N	761
水生-淡水-沉积物（3300013133|Ga0172362_10044703）	3300013133|Ga0172362_10044703_1|M	8	Y	N	654
水生-淡水-沉积物（3300031862|Ga0315280_10005430）	3300031862|Ga0315280_10005430_2|M	4	Y	N	710
水生-淡水-沉积物（3300031952|Ga0315294_10035890）	3300031952|Ga0315294_10035890_2|P	12	Y	N	691
水生-淡水-沉积物（3300031952|Ga0315294_10041214）	3300031952|Ga0315294_10041214_2|M	2	Y	N	741
水生-淡水-沉积物（3300031952|Ga0315294_10041214）	3300031952|Ga0315294_10041214_2|P	2	Y	N	745
水生-淡水-沉积物（3300031952|Ga0315294_10058002）	3300031952|Ga0315294_10058002_1|M	7	Y	N	794
水生-淡水-沉积物（3300032046|Ga0315289_10071835）	3300032046|Ga0315289_10071835_2|M	3	Y	N	727
水生-淡水-沉积物（3300032053|Ga0315284_10046226）	3300032053|Ga0315284_10046226_2|P	7	Y	N	740
水生-淡水-沉积物（3300032069|Ga0315282_10002412）	3300032069|Ga0315282_10002412_8|P	6	Y	N	744
水生-淡水-沉积物	3300032118|Ga0315277_10001015_12|P	3	Y	Y	796

[0375]

(3300032118|Ga0315277_10001015)					
水生-淡水-沉积物 (3300032118|Ga0315277_10012802)	3300032118|Ga0315277_10012802_3|M	6	Y	N	759
水生-海生-深部地下 (3300009499|Ga0114930_10003500)	3300009499|Ga0114930_10003500_21|M	2	Y	N	702
水生-海生-深部地下 (3300011118|Ga0114922_10000510)	3300011118|Ga0114922_10000510_27|M	3	Y	N	736
水生-非海生盐水和碱性-超盐水湖沉积物 (3300017960|Ga0180429_10052609)	3300017960|Ga0180429_10052609_1|M	3	Y	N	679
水生-非海生盐水和碱性-超盐水湖沉积物 (3300017963|Ga0180437_10045654)	3300017963|Ga0180437_10045654_1|M	13	Y	N	741
水生-非海生盐水和碱性-超盐水湖沉积物 (3300017971|Ga0180438_10003700)	3300017971|Ga0180438_10003700_2|P	16	Y	N	741
水生-非海生盐水和碱性-超盐水湖沉积物 (3300017991|Ga0180434_10065880)	3300017991|Ga0180434_10065880_1|P	4	Y	N	762
水生-非海生盐水和碱性-超盐水湖沉积物 (3300017992|Ga0180435_10008274)	3300017992|Ga0180435_10008274_7|P	14	Y	N	670
水生-非海生盐水和碱性-超盐水湖沉积物 (3300018080|Ga0180433_10080815)	3300018080|Ga0180433_10080815_1|M	4	Y	N	762
水生-温泉-热泉 (3300009503|Ga0123519_10002165)	3300009503|Ga0123519_10002165_1|M	4	Y	Y	792
生物反应器宏基因组 (CAAACB010000011)	CAAACB010000011_94|M	12	Y	N	789
生物反应器宏基因组 (SRR5251193)	SRR5251193_megahit_k177_201695_120|P	8	Y	N	679
生物反应器宏基因组 (SRR5251487)	SRR5251487_megahit_k177_199995_1|P	5	Y	N	679
生物反应器宏基因组 (SRR5576005)	SRR5576005_megahit_k177_1292294_1|M	8	Y	N	789
生物反应器宏基因组 (UOUK01008448)	UOUK01008448_2|M	3	Y	Y	757
生物反应器宏基因组 (UOUL01003058)	UOUL01003058_1|M	2	Y	Y	769
生物反应器-厌氧-厌氧沼气反应器 (3300009652|Ga0123330_1010394)	3300009652|Ga0123330_1010394_3|M	2	Y	Y	769
生物反应器-柳枝稷降解 (3300012949|Ga0153798_10025413)	3300012949|Ga0153798_10025413_2|P	13	Y	N	638
洞穴宏基因组 (SRR7811943)	SRR7811943_megahit_k177_406423_2|M	3	Y	N	778
淡水宏基因组 (SRR5818194)	SRR5818194_megahit_k177_802550_7|M	4	Y	N	728
淡水宏基因组 (SRR5818194)	SRR5818194_megahit_k177_802550_6|P	4	Y	N	762
淡水沉积物宏基因组 (SRR6481359)	SRR6481359_megahit_k177_1860565_4|M	6	Y	Y	743
热泉宏基因组 (SRR6049666)	SRR6049666_megahit_k177_256604_1|P	4	Y	N	792
热液喷口宏基因组 (SRR2080392)	SRR2080392_megahit_k177_207267_1|M	4	Y	N	785
热液喷口宏基因组 (SRR2080392)	SRR2080392_megahit_k177_207267_2|P	4	Y	N	742
实验室富集限定的培养基-离子液体和高固体富集的 (3300025517|Ga0207869_1001046)	3300025517|Ga0207869_1001046_11|M	6	Y	N	674
实验室富集限定的培养基-离子液体和高固体富集的 (3300025572|Ga0207864_1000213)	3300025572|Ga0207864_1000213_47|M	7	Y	N	674
实验室富集限定的培养基-离子液体和高	3300025572|Ga0207864_1014515_2|P	3	Y	N	679

[0376]

固体富集的（3300025572|Ga0207864_1014515）					
实验室富集限定的培养基-离子液体和高固体富集的（3300025572|Ga0207864_1014515）	3300025572|Ga0207864_1014515_3|M	3	Y	N	638
永久冻土宏基因组（SRR6505097）	SRR6505097_megahit_k177_377355_35|M	7	Y	N	784
永久冻土宏基因组（SRR6962291）	SRR6962291_megahit_k177_1754533_5|M	3	Y	N	722
永久冻土宏基因组（SRR6962291）	SRR6962291_megahit_k177_1754533_5|P	3	Y	N	762
永久冻土宏基因组（SRR6962291）	SRR6962291_megahit_k177_3367137_15|M	12	Y	N	735
永久冻土宏基因组（SRR6962291）	SRR6962291_megahit_k177_3367137_14|P	12	Y	N	760
菲西芬氏菌纲（Phycisphaerae）细菌 ST-NAGAB-D1（CP019791）	AQT69685.1	13	Y	N	747
菲西芬氏菌目细菌（PMBN01000097）	PMBN01000097_4|M	3	Y	N	784
菲西芬氏菌目细菌（PMDC01000375）	PMDC01000375_4|M	7	Y	N	784
菲西芬氏菌目细菌（PMMM01000072）	PMMM01000072_9|M	7	Y	N	784
浮霉菌门细菌 RBG_13_46_10（MHYD01000012）	QBM02859.1	19	Y	Y	743
沉积物宏基因组（SRR6963308）	SRR6963308_megahit_k177_1876783_1|M	8	Y	N	679
沉积物宏基因组（SRR6963352）	SRR6963352_megahit_k177_1225001_5|P	12	Y	N	762
沉积物宏基因组（SRR6963480）	SRR6963480_megahit_k177_1813084_1|M	5	Y	N	762
沉积物宏基因组（SRR6963491）	SRR6963491_megahit_k177_4304123_3|P	17	Y	N	670
沉积物宏基因组（SRR6963587）	SRR6963587_megahit_k177_2272343_29|M	16	Y	N	741
沉积物宏基因组（SRR6963591）	SRR6963591_megahit_k177_1621237_2|P	16	Y	N	741
沉积物宏基因组（SRR8554444）	SRR8554444_megahit_k177_608234_1|M	6	Y	Y	792
土壤宏基因组（SRR4030068）	SRR4030068_megahit_k177_2163203_10|M	2	Y	N	754
土壤宏基因组（SRR5487326）	SRR5487326_megahit_k177_254484_5|M	3	Y	Y	767
土壤宏基因组（SRR5487761）	SRR5487761_megahit_k177_404752_1|M	3	Y	N	767
土壤宏基因组（SRR8554485）	SRR8554485_megahit_k177_439309_3|P	8	Y	N	767
土壤宏基因组（SRR8554511）	SRR8554511_megahit_k177_4512460_1|M	2	Y	N	669
土壤宏基因组（SRR8554511）	SRR8554511_megahit_k177_1767604_11|P	8	Y	N	744
土壤宏基因组（SRR8554643）	SRR8554643_megahit_k177_4328113_1|P	3	Y	N	669
土壤宏基因组（SRR8554643）	SRR8554643_megahit_k177_1281129_3|P	5	Y	N	740
土壤宏基因组（SRR8554643）	SRR8554643_megahit_k177_1281129_3|M	5	Y	N	706
地下宏基因组（SRR6056591）	SRR6056591_megahit_k177_2733919_1|P	3	Y	N	736
陆生宏基因组（SRR6263259）	SRR6263259_megahit_k177_232832_6|M	2	Y	N	758
陆生-土壤（3300031539|Ga0307380_10018846）	3300031539|Ga0307380_10018846_3|P	5	Y	N	740
陆生-土壤（3300031539|Ga0307380_10018846）	3300031539|Ga0307380_10018846_4|M	5	Y	N	706
陆生-土壤（3300031539|Ga0307380_10072822）	3300031539|Ga0307380_10072822_2|M	3	Y	N	669
陆生-土壤（3300031539|Ga0307380_10072822）	3300031539|Ga0307380_10072822_2|P	3	Y	N	691
陆生-土壤（3300031566|Ga0307378_10072489）	3300031566|Ga0307378_10072489_1|M	4	Y	N	669
陆生-土壤	3300031566|Ga0307378_10072489_2|P	4	Y	N	691

[0377]

(3300031566|Ga0307378_10072489)					
陆生-土壤 (3300031670|Ga0307374_10070813)	3300031670|Ga0307374_10070813_2|P	6	Y	N	767
陆生-土壤 (3300031965|Ga0326597_10006968)	3300031965|Ga0326597_10006968_2|P	12	Y	N	698
陆生-土壤-堆肥 (3300017540|Ga0180217_1000254)	3300017540|Ga0180217_1000254_36|M	8	Y	N	687
陆生-土壤-沼地 (3300014494|Ga0182017_10000840)	3300014494|Ga0182017_10000840_9|P	7	Y	N	784
陆生-土壤-沼地 (3300014839|Ga0182027_10004505)	3300014839|Ga0182027_10004505_3|M	21	Y	N	735
陆生-土壤-沼地 (3300014839|Ga0182027_10004505)	3300014839|Ga0182027_10004505_2|P	21	Y	N	760
陆生-土壤-硬木森林土壤 (3300031715|Ga0307476_10010105)	3300031715|Ga0307476_10010105_2|P	3	Y	Y	768
陆生-土壤-硬木森林土壤 (3300031715|Ga0307476_10010105)	3300031715|Ga0307476_10010105_1|M	3	Y	Y	699

[0378] 表12.代表性CLUST.089537效应子蛋白的氨基酸序列

[0379]

>CAAACX010000652_28|M
[生物反应器宏基因组]
MGVKAWKAKIIVADPDLAAALDHTVRQWNLWAEQACIAIMRLRRGDYGDEGQAVWRFLSEKGTFGTSILDKITRVTGRGIKNDE
ERTIASNFLSRHGPFEPLRAARDALGPNFWYQFQRAVRERIESHDAKLRQWREDHAEWIETRRTWEEEHRAYMNVRPVIEAFEA
AEGQSGRRRGRWRRWLDFLSSHPELAAWRGGPATIVPLSKEELAACRRRRRRAVAEQARLFFEKNPELRALNREHDEYQEKFAR
PWAKRRNPDGFRHRPTFTLPHPDSHPDWPRFKGNAGWRDLSVPDRTVSLELLCPDGSRRWFNVRFVPDPRLASLRPATQPVRQG
NTTYHYVSDALGVAEQPVQPQGIKLILRPDGAYLSVSVSVLAPPCGFDVRQSAIEKYGSIWACRKYKQDHPDRPLVTCAVDLAV
RHNGALTVARDGEVFARRLLHNRFFLPGDDRERAMNIPSLLEIAAVKRQLRRARRQTGRIAPGKPSCARLQAHYRALCEDRYKK
LVAAIFAYARHHGAEVVIFEDLKNLLPNAVNERGVNRALQSWNRGAIVRFARQSAEDVGLRVATVPAFHTSIVCARCGALGCRF
EQTREGVRIAKLGPWFGCRACGRRIHADLNASENLHKVLIGTFPQVRRISRDPPVLSIADVRVDQREADRLAAAWLSELSVGPT
PF (SEQ ID NO: 301)
>CAAACH010011569_2|M
[激活的污泥宏基因组]
MQLWRTHCVFNQRLPALLKRLFAMRRGEVGGNEAQRQVYQRVAQFVLARDAKDSVDLLNAVSLRKRSANSAFKKKATISCNGQA
REVTGEEVFAEAVALASKGVFAYDKDDMRAGLPDSLFQPLTRDAVACMRSHEELVKTWEKEYQEWRDRKSEWEAEPEHARYLNL
RPKFEEGEAARGGRFRKRAERDHAYLDWLEANPQLAAWRGKAPPAVVPIDEAGKRRIARAKAWKQASVRAEEFWKRNPELHALH
KIHVQYLREFVRFKQRPTFTMPDPVHHPRWCLFNAPQTSPQGYRLLRLPQSRRTVGSVELRLLTGPSDGAGFPDAWVNVRFKAD
PRLAQLRPVKVPRTVTRGKNKGAKVEADGFRYYDDQLLIERDAQVSGLKLLFRDIRMAPFADKPIEDRLLSATPYLVFAVKIKD
EARTERAKAIRFDETSELTKSGKKQKSKTLPAGLVSVAVDLDTRGVGFLTRAVIGVPEIQQTHHGVRLLQSRYVAVGQVEARAS
GEAEWSPGPDLAHIARHKREIRRLRQLRGKPVKGERSHVRLQAHIDRMGEDRFKKAARKIVNEALRGSNPAAGDPYTRADVLLY
ESLETLLPDAERERGINRALLRWNRAKLIEHLKRMCDDAGIRHFPVSPFGTSQVCSKCGALGRRYSLARENGRAVIRFGWVERL
FACPNPECPGRRPDRPDRPFTCNSDHNASVNLHRVFALGDQAVAAFRALPPRDSTARTLAVKRVEDTLRPQLMRLHKLADAGVD
SPF (SEQ ID NO: 302)
>CAAACI010012331_3|M
[激活的污泥宏基因组]
MQLWRTHCVFNQRLPALLKRLFAMRRGEVGGNEAQRQVYQRVAQFVLARDAKDSVDLLNAVSLRKRSANSAFKKKATISCNGQA

REVTGEEVFAEAVALASKGVFAYDKDDMRAGLPDSLFQPLTRDAVACMRSHEELVATWKKEYQEWLDRKSEWEAEPEHALYLNL
RPKFEEGEAARGGRFRKRAERDHAYLDWLEANPQLAAWRGKAPPAVVPIDEAGKRRIARAKAWKQASVRAEEFWKRNPELHALH
KIHVQYLREFVRFKQRPTFTMPDPVHHPRWCLFNAPQTSPQGYRLLRLPQSRRTVGSVELRLLTGPSDGAGFPDAWVNVRFKAD
PRLAQLRPVKVPRTVTRGKNKGAKVEADGFRYYDDQLLIERDAQVSGLKLLFRDIRMAPFADKPIEDRLLSATPYLVFAVKIKD
EARTERAKAIRFDETSELTKSGKKQKPKTLPAGLVSVAVDLDTRGVGFLTRAVIGVPEIQQTHHGVRLLQSRYVAVGQVEARAS
GEAEWSPGPDLAHIARHKREIRRLRQLRGKPVKGERSHVRLQAHIDRMGEDRFKKAARKIVNEALRGSNPAAGDPYTRADVLLY
ESLETLLPDAERERGINRALLRWNRAKLIEHLKRMCDDAGIRHFPVSPFGTSQVCSKCGALGRRYSLARENGRAVIRFGWVERL
FACPNPECPGRRPDRPDRPFTCNSDHNASVNLHRVFALGDQAVAAFRALPPRDSTARTLAGKRVEDTLRPQLMRLHKLADAGVD
SPF（SEQ ID NO: 303）

>UZXW01002767_11|M

[激活的污泥宏基因组]

MSVRAIQAKVMCDRATLENLWRTHRVFNERVPRILSLLFRMRRGECGKTPAERQLYQRIGQFITNCSAQNADYLMNAVTMRGWK
PGSAKKYKIRVKDENGETVEISGESWADEAAALSAKGSILCDKEDLAGDLPGCMRQMLNREAVAIISGHDELVANWSTDHVQWL
AEKAKWEREDEHKQYLALRSKFDEFEKSIGGKVGKRRGRWHLYIDWLRANPQLAAWRGGQATVNDLPPDALARIKKAKPWKTRS
VEAEEFWKVNPELAALDKLHGYYEREFVRRRKTKRNPDGFDHRPTFTMPDPTNHPRWFVFNAPQTNPAGYADLNLPDNQGDSGS
IKLCLLTGEKGDATEYPSDWITAQFRADPRLSRFRRVKVATTVNRGKAKGQAKEKDGYEFFDAHLMQWRPAEISGVKLILRHIK
LSGDSSLKSATPYLVFTVSTDDLALTEKARKIEWSDTGHVTKTGKPRKKRTLPDGLVACAIDLGLRHAGFATLALYQDGRPHVL
RSRNIWLEPDGRGPELGHIGRHKRKIRKLRRLRGKPVKGEESHIGLQKHITHMGEDRFKKAARGIINFAWNADGTVATKGADPF
PRADVILIEKLEGFIPDAEKERGVNRALAAWNRGQLVNRLKELAIDAGYKARIFEILPHGTSQVCSRCGAMGRRYSIARDEESG
RPEIRFGWVEKLFACPKCHFSANSDHNASVNLHRKFLLGDAAVAAFFDWKNLPEEKRRGELDAIDEALRGYLRKMHKLQAEIET
PF（SEQ ID NO: 304）

>3300009161|Ga0114966_10008307_8|M

[水生-淡水-淡水湖]

[0380] MKLLFKLRRGKIGKNKQQRDDYKRWFEYVIYNGAKNADYNLNALSINNWTPNTAKKYITKLMAGKNKIHAKGLDLVVSRIKSLS
DGGILAFDKDKIFIEIPYSFSQLIIRDAVAGISGHDALVKIWKEDHKKWRDEKKSWEDKTENKQYLKLQHIFKQFSDTAGGVGS
KKRGRWFTYLCWLSKQSALANWRGNGSFIEPFGEENCGKPSCNKGFKSKLIKESKKFFDANPELFEIHKLHQFYEKNFIRPKAK
KHNIDGFEHPPTFTLPDPVRHPIWACYNAPQTNPIGYKDLVLPKKYSGKSAGSFALNLLTGDIIDGKYPKEWVPFKFIADKRLS
CFKGKIKVVNNKIRTEYIFEDPHLSKKRQATIGGIKIILKDIKIDNLGHLLSATPYVVFACDVDNEPISEQAKKLKSEDTGETG
NSGLPIRKRILPEGCLISCAVDLGIRNLGFATLAESGGIPSECTHGDIKLLRSYNLLLEDGPSLEKIAKQKKLLGKFRRLRGKP
IKGEKSHVFLQNHITKMGVDRFNKAARLIVNFALNTEKKKNVQGIEYPRADVIVLENLAGLIPDAAKRKGINKALMNWNRGHLV
KRIQEIAHDSGYLYRVYIMSPWGSSQVCCKCKGLGRRYSIIKDENGRQIQFGNGEKLFACANCRYMANADHNASINLHHIFKDK
DAVGFYKDYANKNEEEKQKIIDNLEAILRPHLQKMHFTTKKERSLDVPEHPLDVPF（SEQ ID NO: 305）

>3300009161|Ga0114966_10008307_7|P

[水生-淡水-淡水湖]

MGTRAIKAKIICDKTTLGYLWRSHCMYNQKLSEYMKLLFKLRRGKIGKNKQQRDDYKRWFEYVIYNGAKNADYNLNALSINNWT
PNTAKKYITKLMAGKNKIHAKGLDLVVSRIKSLSDGGILAFDKDKIFIEIPYSFSQLIIRDAVAGISGHDALVKIWKEDHKKWR
DEKKSWEDKTENKQYLKLQHIFKQFSDTAGGVGSKKRGRWFTYLCWLSKQSALANWRGNGSFIEPFGEENCGKPSCNKGFKSKL
IKESKKFFDANPELFEIHKLHQFYEKNFIRPKAKKHNIDGFEHPPTFTLPDPVRHPIWACYNAPQTNPIGYKDLVLPKKYSGKS
AGSFALNLLTGDIIDGKYPKEWVPFKFIADKRLSCFKGKIKVVNNKIRTEYIFEDPHLSKKRQATIGGIKIILKDIKIDNLGHL
LSATPYVVFACDVDNEPISEQAKKLKSEDTGETGNSGLPIRKRILPEGCLISCAVDLGIRNLGFATLAESGGIPSECTHGDIKL
LRSYNLLLEDGPSLEKIAKQKKLLGKFRRLRGKPIKGEKSHVFLQNHITKMGVDRFNKAARLIVNFALNTEKKKNVQGIEYPRA
DVIVLENLAGLIPDAAKRKGINKALMNWNRGHLVKRIQEIAHDSGYLYRVYIMSPWGSSQVCCKCKGLGRRYSIIKDENGRQIQ
FGNGEKLFACANCRYMANADHNASINLHHIFKDKDAVGFYKDYANKNEEEKQKIIDNLEAILRPHLQKMHFTTKKERSLDVPEH
PLDVPF（SEQ ID NO: 306）

>3300010885|Ga0133913_10083255_2|P

[水生-淡水-淡水湖]

MGTRAIKAKIICDKTTLGYLWRSHCMYNQKLSEYMKLLFKLRRGKIGKNKQQRDDYKRWFEYVIYNGAKNADYNLNALSINNWT

PNTAKKYITKLMAGKNKIHAKGLDLVVSRIKSLSDGGILAFDKDKIFIEIPYSFSQLIIRDAVAGISGHDALVKIWKEDHKKWR
DEKKSWEDKTENKQYLKLQHIFKQFSDTAGGVGSKKRGRWFTYLCWLSKQSALANWRGNGSFIEPFGEENCGKPSCNKGFKSKL
IKESKKFFDANPELFEIHKLHQFYEKNFIRPKAKKHNIDGFEHPPTFTLPDPVRHPIWACYNAPQTNPIGYKDLVLPKKYSGKS
AGSFALNLLTGDIIDGKYPKEWVPFKFIADKRLSCFKGKIKVVNNKIRTEYIFEDPHLSKKRQATIGGIKIILKDIKIDNLGHL
LSATPYVVFACDVDNEPISEQAKKLKSEDTGETGNSGLPIRKRILPEGCLISCAVDLGIRNLGFATLAESGGIPSECTHGDIKL
LRSYNLLLEDGPSLEKIAKQKKLLGKFRRLRGKPIKGEKSHVFLQNHITKMGVDRFNKAARLIVNFALNTEKKKNVQGIEYPRA
DVIVLENLAGLIPDAAKRKGINKALMNWNRGHLVKRIQEIAHDSGYLYRVYIMSPWGSSQVCCKCKGLGRRYSIIKDENGRQIQ
FGNGEKLFACANCRYMANADHNASINLHHIFKDKDAVGFYKDYANKNEEEKQKIIDNLEAILRPHLQKMHFTTKKERSLDVPEH
PLDVPF (SEQ ID NO: 306)
>3300020109|Ga0194112_10008260_1|P
[水生-淡水-淡水湖]
MVDPEARAALWRTHGVYNECLRFALRHMHKMKRGEADPRYAEIFGAIRNSQHATAVMEAVTDPRWKPKKDPKPWHAPAAALIAE
GKLLFDRWKELPGLPGRFHRKLWEGAYQAMDSHKELMAIWASAHEEWQKARAEWEAENPEYMKLRPALEAFEAEHGQAAKRRQR
WHKWLAFLQANPGLAAWRGGADVVNPISDDGKERLRRAKRSKRNKVESEEFFNANPELKELDKLHGYYQREFVRPWAKKRNPDG
FRHRPTFTEPSPEKHPFWYQFKKGETYKALDLQAGTISLQLLAPEGAKRAALWTTFAFRPDPRLRLIQKVNPPLNEKFTYAFAD
PAYRGQPARPAEIRGAKLIFRPAQPDGAVYLSFTCDMPDLHCRLSVGQDSIDRYTTKGTKWLVGRLREEPELAGHEPVTCAVDL
GLRHLGAATVRRDGKIVRARLLREQGDPGKGPGLPAIAGHKRALAKGRRKRGKPIAGEESFVELQRHTDKMGEDRFKKGARRIV
AFAHENGADLIILEKLAGLVPDAEKERGINRALVSWNRANLEKWVKQLANDAGMRVVEVFPHWTSQLCSRCGSMGARFSAKEGL
PAFEPVGKLFACPECGYIANADHNASVNLHKRFYGELPEVERPNKDKRVFRLTPRGQPTVEVDMADVEKRILARGRAEDLCRGK
PTPF (SEQ ID NO: 307)
>3300020109|Ga0194112_10008260_1|M
[水生-淡水-淡水湖]
MDMVDPEARAALWRTHGVYNECLRFALRHMHKMKRGEADPRYAEIFGAIRNSQHATAVMEAVTDPRWKPKKDPKPWHAPAAALI
AEGKLLFDRWKELPGLPGRFHRKLWEGAYQAMDSHKELMAIWASAHEEWQKARAEWEAENPEYMKLRPALEAFEAEHGQAAKRR
QRWHKWLAFLQANPGLAAWRGGADVVNPISDDGKERLRRAKRSKRNKVESEEFFNANPELKELDKLHGYYQREFVRPWAKKRNP
DGFRHRPTFTEPSPEKHPFWYQFKKGETYKALDLQAGTISLQLLAPEGAKRAALWTTFAFRPDPRLRLIQKVNPPLNEKFTYAF
ADPAYRGQPARPAEIRGAKLIFRPAQPDGAVYLSFTCDMPDLHCRLSVGQDSIDRYTTKGTKWLVGRLREEPELAGHEPVTCAV
DLGLRHLGAATVRRDGKIVRARLLREQGDPGKGPGLPAIAGHKRALAKGRRKRGKPIAGEESFVELQRHTDKMGEDRFKKGARR
IVAFAHENGADLIILEKLAGLVPDAEKERGINRALVSWNRANLEKWVKQLANDAGMRVVEVFPHWTSQLCSRCGSMGARFSAKE
GLPAFEPVGKLFACPECGYIANADHNASVNLHKRFYGELPEVERPNKDKRVFRLTPRGQPTVEVDMADVEKRILARGRAEDLCR
GKPTPF (SEQ ID NO: 308)
>3300020222|Ga0194125_10021011_3|M
[水生-淡水-淡水湖]
MSTKAFTARLIVESPEARDALWRTHRLYNECLRFVLRHMHKMKRGEADPRYGEIFGAIRNSQHATAVMEAVTDLRWKPKKDPKP
WHTPAAALIAEGKLLFDRWKELPGLPGRFHRKLWEGAYQAMDSHKELMALWVSAHQEWQKARAEWEAENPEYMKVRPALEAFER
EHGQAAKRRQRWHKWLAFMRADPRLAAWRGGKALANPISDQGQQRLRRARRNKRNKVESEEFFNANPELKELDKLHGYYQREYV
RPWAKKRNPDGFRHRPTFTEPSPEKHPFWYQFKKGETYKALDLQAGTISLQLLAPEGAKRAALWTTFAFRPDPRLRLIQKANPP
LKEKFTYAFADPAFPGQPARPAEIRGAKLIFRRAQPDGAIYLSFTCDVPDLDSRLSILQKSCDKYGAKWACKKAMSDLGGQQPV
TCAVDLGLRHLGAATVRRDGKIVRARLLREEDQPGKGPALPAIAGHKRALAKGRRKRGKPIAGEESFIELQRHTDKMGEDRFKK
GARRIVAFARENGADLIILEKLGGDRRGALIPDAERERGINRALVLWNRGNLAKWVKQLAADAGMRVVEVHPYMTSQLCSRCGA
MGARFSAKEGLPAFEPVGKLFACPDCGYLANADHNASVNLHKRFFGELPEVQRPNRDKRVFRVTPDDQPTVEVDMADVEKRILA
RGRADDLCRGKRTPF (SEQ ID NO: 309)
>3300027770|Ga0209086_10006474_1|P
[水生-淡水-淡水湖]
MGTRAIKAKIICDKTTLGYLWRSHCMYNQKLSEYMKLLFKLRRGKIGKNKQQRDDYKRWFEYVIYNGAKNADYNLNALSINNWT
PNTAKKYITKLMAGKNKIHAKGLDLVVSRIKSLSDGGILAFDKDKIFIEIPYSFSQLIIRDAVAGISGHDALVKIWKEDHKKWR
DEKKSWEDKTENKQYLKLQHIFKQFSDTAGGVGSKKRGRWFTYLCWLSKQSALANWRGNGSFIEPFGEENCGKPSCNKGFKSKL

[0381]

[0382]

IKESKKFFDANPELFEIHKLHQFYEKNFIRPKAKKHNIDGFEHPPTFTLPDPVRHPIWACYNAPQTNPIGYKDLVLPKKYSGKS
AGSFALNLLTGDIIDGKYPKEWVPFKFIADKRLSCFKGKIKVVNNKIRTEYIFEDPHLSKKRQATIGGIKIILKDIKIDNLGHL
LSATPYVVFACDVDNEPISEQAKKLKSEDTGETGNSGLPIRKRILPEGCLISCAVDLGIRNLGFATLAESGGIPSECTHGDIKL
LRSYNLLLEDGPSLEKIAKQKKLLGKFRRLRGKPIKGEKSHVFLQNHITKMGVDRFNKAARLIVNFALNTEKKKNVQGIEYPRA
DVIVLENLAGLIPDAAKRKGINKALMNWNRGHLVKRIQEIAHDSGYLYRVYIMSPWGSSQVCCKCKGLGRRYSIIKDENGRQIQ
FGNGEKLFACANCRYMANADHNASINLHHIFKDKDAVGFYKDYANKNEEEKQKIIDNLEAILRPHLQKMHFTTKKERSLDVPEH
PLDVPF (SEQ ID NO: 306)
>3300015360|Ga0163144_10184575_1|M
[水生-淡水-淡水微生物垫]
MECNREELQHLWRTHCAFNERLPSILSVLFKMRRGEWGDTEQIKKKMQAFAWGVLQLEAQNSYDPLIKAASGSARNSTARNALR
KIAAKASKAIDDETAKPSESQDAAKLAKARKTLLQCQEYESSFDLFAKGLTPLKDVKDALADLPNHIERKLFEESSACIRGHLA
LVENWENEHVEWLKKKKDWEADPEHRKYLDVRHRFDAFEEEAGGKAGKRRGRWHLYLDWLRKNPDLAAWRGGQSSIGEPDASAK
LRIQKAKPWKQRSVEAEEFWKVNPELSALNKLHGFYEREFIRRRKTKKHSDGFDHRPTFTLPHAVNHPRWFVFNAPQTSPKGYC
ELKLPGAPGAEGSLELLLLTGEKEMGVYPKKWTSLVFKADPRLSDFKPVTVKKNANKGKVKGQETEKPAYSFYDRQLKQHRAAQ
ISGVKLIFKQVKLNEDGTLKSAVPYLYFTCSVVDIAITQKAQDIEWSETGETTKTGKARKRKTIPNGLIACAVDLGLRNLGFAT
LGVLESGEVRVLRSRNLWIGVAEPSGAHAGRWAEGPDLAHIGAHKRNIKRLRRQRGKPVEGERSHVELQEHIDHMGEDRFKKSA
RAIINFALNVDQTIDKKTGKPFPRADVLVMEKLEGFIPDAEKERGINRALVSWNRGQLIERLKQLAEDSGFKKRLFEIHPAGTS
QVCSRCGALGRRYSIARDETGQPEIRFGWVEKLFACPNSKCGYRANADHNASVNLQRSFFLDGQAVQAFYDWKGKPEPERKTFF
AAMDGKLQPALRVEHKLDGPTPF (SEQ ID NO: 310)
>3300013130|Ga0172363_10020541_1|M
[水生-淡水-沉积物]
MSIRAIQARVECDRPTLERLWLTHRIFNERLPVILSFLFRMRRGECGDTVVERKMYHEIADFILTHSQEAPYLFNAVSIKGWKP
GTAVKQSPGSEWPAQAAQLSKEGKLLYDKSNVLGDLPGTLQQMVCRESAAIISGHHELVRMWEAEHAQWLEDKSAWEADAEHQK
YLALRPRFDAFEQEAGGPAGKRRGRWHLYLNWLKQNPDLAAWRGGKPEVIELSPAARRRVERAKPNRQRSVEAEEFWKGNPELS
ALDRLHGYYEREFVRRRKTKKNADGFKHRPTFTLPHPRIHPRWFVFNAPQTQPSGYADLVLPTKPGEEGSVRMLVINAEKANGK
YPSEWVRIRFHGDPRLCNFRAVSSTRTISRGKDKGKTKESPSYVFRDRHLGQDRPATISGIKLIFRTNTDRTPKAAYLYFACTI
DDLPQSDAAKSIQWTETGEITSKGKKRKNMKVPDGLVGCAVDLGIRNLGFATLARYADGQPQILRSRNIWVRRQEAKGAHTGRW
SAGADLAHIGEHKRTIRRLRRQRGKPVKDEKSHVGLQLHIDHMGEDRFKQAARAIINFALNADGDCNRKTGEVHPRADVLVLEN
LGGLIPDAEKEHGINRALVNWNRGQLVERIKEMAQDVGLKVLEVSPVGTSQVCCKCGSLGRRYSVVRNEDTGRLDIRFGQVEKL
FACPASDCPGREAGRPDRPFTCNADYNASANLHRCFLLRDQAVKTYRDWRSGGDSKKAKEERETKIRNVEDALLGPLRIMHGME
DAPPF (SEQ ID NO: 311)
>3300013133|Ga0172362_10044703_1|M
[水生-淡水-沉积物]
MDGDTWANEAAALSAQGVLLYDKHAVLGDLPAHLRQMVCRESVAILSGHDELVARWNKEHQEWLTAKAKWEDEEEHKKYLALRP
RFEAFEKEAGGKAGKRRGRWHLYFSWLKANPDLAAWRGKADSVALLSDKAKERIRKARPWKKRSVEAEEFWKVNPELQALDKLH
GYYEREFVRRRKTQRNPDGFHHRPTFTLPHFTRHPRWFVFNAPQTSPQGYDNLQLPAKPHAGGSVELLLLTGDKTNGEYPSQWI
SIGFQGDPRLSLFRKVKKQRTVNRGKDKGTQKESDAYEFRDLHLNRWRKAEISGIKLLFRLNPNGSPKAAYLCFTCSITDEPLS
SNARKIKWSETGETTKKGKKRKRKTLPDGLVTCAVDLGIRNTGFATLAVHDQGNVLIRRSRNLWAGWLGEDSKDRDRWMDGPVL
AHIAAHKRKIRFLRRKRGKPVKGETSHVELQTHIDHMGEDRFKRAARAIINFALNVGSECDKKTGEVYPRADVLVLENMDGLIP
DAEKEHGVNRALVAWNRGQLTKRIEEIAEDVGLKILFVSPMGTSQVCSKCGSLGRRYSIKHDDMIDRPDIQFGFVEKLFACPSC
QYMANADHNASVNLHRRMVLDDKAVVAYNSFREKNKTDQATEIERIENILRPRLRLKHSLEGSVPF (SEQ ID NO: 312)
>3300031862|Ga0315280_10005430_2|M
[水生-淡水-沉积物]
MVELYLRMRKGKHRETARKIAEIILKSPKNTAAAIMEGLTAKEWTSTRTTEWANLVRERRNAEGLLFDRHESVGSADGRNACCQ
RQQGPRKKPTDPSRLPVSSKFWRWLCDDACEVLKSNRGQMDDWRSEYREWRDEKTQWEKDHPEFMAFWNGLYQQFRQDCERAQI
EAQQAKGQKPTFRKRGGRQWGKRFERWHLWPEWIEQHPEILAWRGQAAAADFKPVPQQEKDALRKRFKRQDKYIGALLDKLRES
NSELAELDRLRRAYERQYVRFRRFPTLTLPDPDKHPRWMSLERDCLYKGFDPEKGTIKLSLVGENADGVWYMDWFDVQVQVDRR

[0383]

LKPSFRREVFGREGRLPPFIEGRVGRKLSQPAPSPAERQSGVAGAKLIFKQGAAHLTFTIIDQDRAPRVKPRKVADRRCPADNI
FDPQGDRIPLRVVAVDLGVRHVGGYAVAEGRFENGAWSVGYITKGIVSGPQVPERFSIDAHHRQLKWARRLRGKPVFDEESCAE
LQDHVTHMGEDRFKKAAHAIVELARRHDVHVILFENLKGFVPSACDERWANRTRSKMLRRKTVEMVKAQCHDFGFYCNEEISPK
LTSHICSKCFRPGWRFSMKAKQPWSEKQPRSKCRDFGYPAWDAGGHLFRCPHCGYRVNADINAAGNLVAKFFGLWPDEKLERKD
WVYRWRENGESRQFDAKAEFEKWSDDVRSRKAARSFPY (SEQ ID NO: 313)
>3300031952|Ga0315294_10035890_2|P
[水生-淡水-沉积物]
MSVKCFQAKVVCDTPEKREYLWLTHTLFHEALRALIPHLFKMRRGDYGADFKEIYEAITDNQNSFAKLEPITTPKAWRSKHADT
GDTKNRWADLCLAVNREGRILFDRTQTPFGFGSEFWRKVCEMAIQLIHSHDGLLADWRKERAEWLDNKKQWEAANPEYMAAKPT
FDAFEDEAGRLRGSRKRWLLYLEFLQKHPELAAWRGGEAKVAPLTPAERKQCRRPGDQFEIFWEKNPQLAELNRLDRDWRRNFA
RFKRRPTWTNPDPELHPAWFSFKRGQTYKDLDFARRTFNLNVLVGEDVKGRKGKWVTYEFRHDPRLLRLRPAEQPVKVGPNKFT
WLYSDPLLKKDRPAEVKGIKLVFRNRHPYLLFSVDIADEKSAPLTRPGFDDKRLKRKATAADIPDGTRILSIDFGQRSLAACSV
CEFIKGHPERPDAVFFMKLPGLKFGDIGRHEQTISRRSSRMFRAGPRSRESRHAPRGGTTFREFRRHVLKMKDDRYKKATRLIL
DAAVRYDAQVILVENLENYKPDLERSARENRARMRWNVQRIVEFLDKSAHPLGLRLWKVWPHWSSRFCSACGHPGARFTIPRKN
AWKRFYERRHGPIRKPVTEPGGQFFICSNPKCPRSTGIINADVNASLNLPRILADDLEKPTGQGKTRLWRGQEYNPKAIAEQCQ
QRLDEHFRSKADLAQQTPW (SEQ ID NO: 314)
>3300031952|Ga0315294_10041214_2|M
[水生-淡水-沉积物]
MFQARIECDRQTLDRLWLTHRVFNERLPAIMSVLFNMRRGESGRNPQERRLYQRMARFFLAMDSKNAEYLLHSVSIKGWTPNTA
LKMRAKVRDAHGAEVPVSGDSWAPAAAALSAKGRLLYEKAQVLGDLPATLRQMVCREAVAVISGHDELVARWEKAHADWLKDKA
RWEADAEHQQYLALRPRFEAFEAEAGGKAAKRRGRWHKYLSWLRANPDLAAWRGKAAPGVQDLSPQARRRVQRAGLKRRAAVEG
EEFWKANPELSALDRLHGYYERAFVRRRKTKKNPDGFDHRPTFTLPHATRHPRWFVFNAPQTSPSGYADLAFPEKRGATGSIRL
CLLTGDKAADGKYPHEWVPVRFLADPRLSAFRPVKRQRVIGRGKRKGEAAEAKSAYEYFDRHLGKTREARFSGVKLVFRLYPDG
KPKAAYLYFACELTDEPLSERAKAIQWTETGEVTRKGKKRKSKRLPDGLVSAAVDLGIRNLGFATLARKEDGAVQVLRSRNLWT
GPDLHHITRHKRRLRALRRDRGKPVAGESSHVRLQGHIDHMGEDRFKRAARAIVNFALNTDGARNGRGQAHPRADVLVLESMEG
LIPDAEKERGINRALVAWNRGHLVERIRDMALDAGLKVFEVSPFGTSQVCSRCGALGRRYSLVREAETGEQAVRFGFVEKLFAC
PACGYRANADHNASVNLHRRFILEDAAVKAYQDLYNRPESERKVALADMESRLLSRLGVEHGLAKPPPF (SEQ ID NO: 315)
>3300031952|Ga0315294_10041214_2|P
[水生-淡水-沉积物]
MSVRMFQARIECDRQTLDRLWLTHRVFNERLPAIMSVLFNMRRGESGRNPQERRLYQRMARFFLAMDSKNAEYLLHSVSIKGWT
PNTALKMRAKVRDAHGAEVPVSGDSWAPAAAALSAKGRLLYEKAQVLGDLPATLRQMVCREAVAVISGHDELVARWEKAHADWL
KDKARWEADAEHQQYLALRPRFEAFEAEAGGKAAKRRGRWHKYLSWLRANPDLAAWRGKAAPGVQDLSPQARRRVQRAGLKRRA
AVEGEEFWKANPELSALDRLHGYYERAFVRRRKTKKNPDGFDHRPTFTLPHATRHPRWFVFNAPQTSPSGYADLAFPEKRGATG
SIRLCLLTGDKAADGKYPHEWVPVRFLADPRLSAFRPVKRQRVIGRGKRKGEAAEAKSAYEYFDRHLGKTREARFSGVKLVFRL
YPDGKPKAAYLYFACELTDEPLSERAKAIQWTETGEVTRKGKKRKSKRLPDGLVSAAVDLGIRNLGFATLARKEDGAVQVLRSR
NLWTGPDLHHITRHKRRLRALRRDRGKPVAGESSHVRLQGHIDHMGEDRFKRAARAIVNFALNTDGARNGRGQAHPRADVLVLE
SMEGLIPDAEKERGINRALVAWNRGHLVERIRDMALDAGLKVFEVSPFGTSQVCSRCGALGRRYSLVREAETGEQAVRFGFVEK
LFACPACGYRANADHNASVNLHRRFILEDAAVKAYQDLYNRPESERKVALADMESRLLSRLGVEHGLAKPPPF (SEQ ID NO:
316)
>3300031952|Ga0315294_10058002_1|M
[水生-淡水-沉积物]
MSVRSIQARLHCDAKTLDDLWRTHRVFNERLPGINAILFKMRRGEFGGTPKEREMMQAFAWCVLRLDAKNSYDPLNKAAIGSAR
SASAVKAVNALIAQAKKTIAKPSAPQAMQDRGTLEEARKTMLFGPVCLDAFSKFATGYLPLKDLKDSLGDLPNAIGRRVLEEAC
ACLRGHFALAAKWQTEHADWLTEKAKWEADLEHQKYLGLRPRFEAFEKEAGGKAGKRRGRWHLYLAWLKANPDLAAWRGKADGV
TDLSKQARERIRRAKPRKQLSVEFDEFWKVNPEFQALNKLHGFYERRFSPPGKQQCGAKSKSLRKKKRNPDGFKHRPTFTLPHP
VRHPRWFVFNAPQTSPKGYWNLCLPTNPEAEGSVELLLLKGEKGADGKYPSDWVKVQYRGDPRLCLFRKVKKRRTVNRGKDKGA
EKEADAYEFRDAHLGLWRPTDISGVKLIFRFNHDGTPKAAYLYFTCNIKDVPLTANAKAIQWTETGEVTKKGKKRKRKTLPDGI

[0384]

LTCAVDLGIRNLAFATLAVHHEKGNLVIKRSRNLWVGWQEGGDAKGERWLDGPDLAHIAKHKRKIRWLRGKRGKPVKGESSHVE
LQTHIDHMGEDRFKRAARAIINFALNAQAACDRHGEVYPRADVLILENMDGLIPDAEKEHGINRALIAWNRGQLTKRIEEMAKD
VGLKIIPVSPMGTSQVCSRCGSLGRRYSIGRDEKTGQAEIRFGFVEKLFACPNSDCGYLANADHNASVNLHRRFALDDKAVAGY
AEHRVKGDDERTKEIADIEARLRPRLRLKHSLDGPVPF (SEQ ID NO: 317)
>3300032046|Ga0315289_10071835_2|M
[水生-淡水-沉积物]
MSARAFQAKVVVERPEDLQALRDLHRRFNEFLPPVIEALFKMKRGALGPDAQLLFSSMTSNQNAVAKLEPLTTPKPWKPRKKQQ
LDETWVQAAVRLREGRGILFDRFEAMPDVSSEFRRKVFEMAFACVDGHLQLVKRWRKEHEDWLKQKERWERGLVVTEEGEDEAS
AETRTWRERYMRLRPLFEEFEQRGVVRRNRRRWHEYIEFLSTHPAFAEFRPNALTDQERADLRDPIARFERLLEKNADLERLDN
LHGQYQDLIRPWARKRNPDGFKHRPTFTMPSVASHPQWFSFKKSATYKNLDLEKGVLALALGKEGAKAQWVTVRLATDPRLARF
RPAQRVGKGKRARSWVYLDPISGEEFDAEVRGLKLMFPKGIQGPAYLMFTVQLSDRPRQVVPRKDWTPPEGLLTLACDFGQRHL
GAITMCRYENGSPVQVPFTPAYLDHGRGRRTPSKPVSAWLPVIPGLRFDDIREHELEISARLSATYGDPRSRRQGGVGASRGRH
AAAGSRPFAQLRRHIDEMREVHFKKAAHLIVATALKNGAQVVIMEFLANYRPDLERSHRENRARMIWAVRRIQEFVEQTAALYG
IAVRTPSPYLTSQVCSACGSPGARFTRPSAMDWTRKPPGRKESWYEKRFGPKAKAVIEPGGPWFCCSNPNCPRPCRVIQADINA
SLNLHRRFYERFDLSWLPDGKPARKQFWGRLKGAIQAYLDQFDSLPRRPGATSHE (SEQ ID NO: 318)
>3300032053|Ga0315284_10046226_2|P
[水生-淡水-沉积物]
MATKTYRARFVNLTGRLDAALARTHHVFVECLRQLIGRYIEMRKGKYGPECRQLAEVILSRFNTFAHGVMDKLSRPNWESRLTD
DWVLLANSVCEAQGPLFDQHDGFAVVDGVTIHTKPHSKSEPVKNRLPVPAKFWHQVCDSASAYLQSNRALMEQWRNERKDWFKN
KTEWEQKHQEFMTFWNGMFKQFEDACEKQRLLSQQAAGQLPTAQKKQSRERGKRLGRWHLWYEWLVAHPEIIEWRGKAKATDFV
FVPDELKKIMSKHSRQDKFIPKHLDWLKENNPELKVLDDLRRTYVRNYLRFKRPPTLTLPSPKKHPYWFTLELDQFYKSPDFDK
GTVKVLLIDENDNGAWFFNWFDAGLLCDKRLKPSYRAEMFRNKGRYPPYIEGKVGRTLNRPAGSSKDRKAGYAGAKLVINGKRK
ELLFTVIEQDCPPSIKWTKVKHRRCSADNALSPDGERIPIKVMAIDLGIRHAGAYAIAEGTADDDIWQVKWLKKGIIDNSAMPD
LYQIRGHDRQLRRGRSKTGKAPKGERSFIDLQSHRTNMAQDRFKKAANAIINIARSYKVDVIIFEKLESLAPTAFDERWMNRQL
RDMNRRKIVDMVKQQAAEFGIKADDNVSPYLTSHVCSRCFKGGWRFSMKAKQPYKEKLSRKNCCDYGYPTWDAGGHLFRCPHCG
YSVNADINAACNVGAKFFGLWPATMSKIKREWVYTWDAENSEPFNAKEAFDKWAEDVKKHKALADSPF (SEQ ID NO: 319)
>3300032069|Ga0315282_10002412_8|P
[水生-淡水-沉积物]
MVTRALRARVIHPDATLLAALKRTHRVYNVLLRQMVELYLRMRKGKHRETARKIAEIILKSPKNTAAAIMEGLTAKEWTSTRTT
EWANLVRERRNAEGLLFDRHESVGSADGRNVCCQRQQGPRKKPTDPSRLPVSSKFWRWLCDDACEVLKSNRGQMDDWRRDYKEW
RNEKARWEKDHPEFMAFWNGLYQQFRQDCERAQIEAQQAKGQKPTFRKRGGRQWGKRFERWHLWPEWIEQHPEILAWRGQAAAA
DFKPVPQQEKDALRKRFKRQDKYIGALLDKLRESNSELAELDRLRRAYERQYVRFRRFPTLTLPDPDKHPRWMSLERDCLYKDF
DPEKGTIKLSLVGENADGVWYMDWFDVQVQVDRRLKPSFRREVFGREGRLPPFIEGRVGRKLSQPAPSPAERQSGVAGAKLIFK
QGAAHLTFTIIDQDRAPRVKPRKVADRRCPADNIFDPQGDRIPLRVVAVDLGVRHVGGYAVAEGRFENGAWSVGYITKGIVSGP
QVPERFSIDAHHRQLKWARRLRGKPVFDEESCAELQDHVTHMGEDRFKKAAHAIVELARRHDVHVILFENLKGFVPSARDERWA
NRTRTKMLRRKTVEMVKAQAGEFGFLCNEKISPKLTSHICSKCFRPGWRFSMKAKQPWSEKQPRSKCRDFGYPAWDAGGHLFRC
PHCGYRVNADINAAGNLVAKFFGLWPDEKLERKDWVYRWRENGESRQFDAKAEFEKWSNDVRSRKAAKSFPY (SEQ ID NO:
320)
>3300032118|Ga0315277_10001015_12|P
[水生-淡水-沉积物]
MATKSFEARVLCRDDPPDKKGEKYSAEQKKQFLWYTHQVFNESIRKVLPYVFKMKRGELGDEFQAIYYAISSSQDAVGKLEAIT
SLDWSSKKIGKSEPEKWKEILRYRELEKEFRQKLKDEEKNTKKFRKELQKKKKALANEIGQKDIWADAASIIQQKNNLLFDREK
ILPNLPSEFRRKIFEMCIQLIHGHQELVANWQDEHAEWLQEKDKFEEEHPDYMKVRPILEEFEKEAGKVKGSRIRWLKYLDYMS
SHPELAGWRGGEAKVEPLTSAERAKFRKPREHFEAFFKKNPELSALDSIHKQYQEKFARTQTKRTPNPDGFKHRPTFTLPSANL
HPAWYSFKGATDPTKGATYKLNFREDRRVDSFMNLKVLTDMEGEGRNPKGMITYDFEPDDRIRDFRYIGTTEKGKKAKGFVYYD
KILEKERPAKIQGIKLVFRPPRSDGVPYLIFSCSIEDEKSKIKIWQDKESEEGPDEARKGKKAKKIYPPELVTLAIDFGQRHLG
AITICKNNNGKPEQIKFIPAYPKRKQGKQTKPVSAWLAKVPGLTFNAIGMHEKEIRAGMHQRFQDPKSIRQAGEKEGRASKPRH

[0385]

IPETETSFAHLRDHIAGMKEDHYKKAANLIIKTALQNGAQVILLEDLKKYRPMLEKESRENRRRMQWAVRQTAEFLKKTAKPLG
LLIYDRINPAYTSSMCSVCGNPGARISLPSKRNWEKYYAEYYGKERRMIAVVGGQFFCCPVCKKIINADINASQNMHKVFYNTF
NWPPKMEKKEKKIFWESVREQAQNILDQKAEPKREEDISY (SEQ ID NO: 321)
>3300032118|Ga0315277_10012802_3|M
[水生-淡水-沉积物]
MSVRAIQARVECDRATLEHLWRTHRVFNERVPVVLSWLFKMRRGECGRTEEEKRLYRTIADFVLANSQNAEAILNAVSIPNWTP
TSARKLKAKTVLSDGTEKVSTGMDWADKAAELSAAGRLLFDKREMLFDLPGTLRQMVCRDAGAIISGHEELVAVWKKAHADWLK
AKADWESQEENKLYLNLRPEFEAFEQAAGGRAAKRRGRWHLYLKWLKDHPRLGGWRGGEPIVNELSEQARRRIQKAKPWKQRSV
EGEEFWRANPELAALDRLHGYYEREFIRRRKTKKNPDGFTHPPTFTVPHAVRHPRWFVFNAPQTSPIGYRNLVLPPHAGPEGTV
ELRLLTGDKTDGKHPGNWVKVRFRADPRLSDFRPAKRTRSFNRGRNKGQTTQSDAYEYFDHHLGLMRSAEIKGLKLIFRFNPDD
SLRAAYLYFTCDVGSIPFTERTRAIQWTETGEVTKKGKKRKTKKLPDGLIACAVDLGLRHLGFATRARYAGANPEILRSRNLWL
GREEQKGRHPGRLSPGADLRHIAHHKRNIRRLRRLRGKPVKSESSHVELQTHIDHMGEDRFKQGARAIINFALNVKKGPDTRTG
EVHPRADVLILEKLAGLLPDAEKERGINRALVAWNRGQLVKRVKEMAQDVGLKVFEVSPVGTSQVCSRCGALGRRYSIVPDDRT
GRPGFRFGWVEKLFACPNPDCRYRGNADHNASVNLHRRFMLEDAAVKAFLDWRAKTKKEQEEVLRRIEKALCADLLRMHGLEEP
TPF (SEQ ID NO: 322)
>3300009499|Ga0114930_10003500_21|M
[水生-海生-深部地下]
MSTKAIQAKVICDTWEHRYFLWITHKLFNDSVQFVIPKVFAMKRGELGDDFKLIYDSIRSNQDAIGKLEQITSLEAKSNRQEKW
VLAAKKIREEKNLLFDRKKELPGFSSEFRRKVFEMTIQKIHGHQELVDDWEKEHNAWLKEKEKFEKDHQDLMALWSEIEKFEKE
AGSLKGSKKRWYRYLEFLGTHPAFSTYRDGGPLSKEERAGLKRPGEHFELFFKKNPKLVPVDSAIKKYCDLLRPWAKKRNSDGF
KLKPTFTLPNPIVHPEWYSFKKNSTYRKLDLQKGEVELKIIPSEDIDKRDPKKYISYRFGADPRIKNLIPSNIEKIGREKYDHN
YFDPVLDKKRPAVIKGIKLIFKGLNRKRKLWAEYEREKNTDKKQLFAQWIDKLRPYFVFTLTLADEAKPVKLEWNKKEKHSVI
PKELVTMAIDVGQRDLGAVTIAREHNGVPVQIAFLPQYWKPLKNGNSAKPVKAWLVDIPGISFEALKKHEKEARGRERGAKARR
VKGENYAIGLRQHILHMKEDRYKKAGSILLRTAIKNKVHVIILEDLKKYKMNLERSKQENRQLMVWAIRKIQDFLELQAKPLGV
RVRSVWPHYTSQVCSACGEPGKRASLPSKSSWEKFYQKKYGEKRKWIVENGGKMFVCSNPNCVRKSRMINADINASLNLHKKFY
GILTYPPKTKEFKQQLTLDIQNYLDTILKF (SEQ ID NO: 323)
>3300011118|Ga0114922_10000510_27|M
[水生-海生-深部地下]
MATKSYRSRIVCTKAELDDLVRTHRVFNAGVRAMVDLYISMKKGTYGETAKEIARYLLDHSGTKGHDLMNPLTDRGGKDTRDDE
GMKLVRRHRQEHGILFPRHERICEIDGRKVRAGSSAKKEKSRDTLTISSIFWHFICNQAIEYLRSNNELMAAWREERGKWLRDK
EKFFAENPGFREFVEGYLAAFDAHAEKARQEFQTKAGQRAHARKRKKSGSGKRFSRWHLWCEWICDHPEIIEWRGKAKGEGFAF
LTAEEVRAIEEEFPKREDRRAGRMLDMLTRKNPELAEVEKLRKKYNNDYARFRRPPTLTLPSAKKHPCWYRLEKGKLYRNEDFE
KGTIEILVICHDEKEDAYYMDWRPFSIKPEPRLVPSHRARAFAAEGRYPPYKEGKVGNTLNRPAESAEERLAGLKGAKLILKPK
GNAQLVFTVVEQNEPLKAKFRKKAGATSSADRLDILDGNKATRVMGIDLGARHVGAYVITEGKKKGDGWKLAPLRKGFIDSPDL
PRLRSITRHENQLKKGRRKRGKPVKNERSFISLQGHRTKMSTDRFKKAANLIVEKAREHDVHIIVFEDLRRLKPTAYDERWLNR
QVRSMNHRHIFDLAKEMAPEFGIQVTDINPYNTSRLCSKCGFPGYRFSMKRKNPYREQAPRRECADFGYPVWDSGGHLFRCPHC
GYRVQADLNAAGNIARKFFGHFMDWECKEYTYRNAVQPSFDARKDFDQWASGVLKRRTMPDAPF (SEQ ID NO: 324)
>3300017960|Ga0180429_10052609_1|M
[水生-非海生盐水和碱性-超盐水湖沉积物]
MSQRALAVRLLVDSEDDRAALWRTHRIFNERLGWVLRQMHRMKRGEPDARYAEIFNAITTSQRATAALEAVTSLAWTGTKEDRW
CQLARALIQEGTLLFDREARLAGLSKEFRRKLFEAAFQMISGHRELLAQWERDHNEWQQARETWEAENPEYMAIRPALQRFEAE
HGQAAKRRARWHKWRAFLQGSSQLAGWRGGPQVVNEISADARRRIERARRNKRNKIEAEEFFRVNPELKELDAKHGYYQREFLR
PWAKRRNADGFKHRPTFTEPSAENHPFWFQFKKHETYKDLNLKVRSLRLKLLASDSAVPGSQWRDFRFRVDPRLLDLSEAQSPA
KIGIARFSLLFNDRALAMEDRPAEIRGAKLIFRPAQRDGRVYLSFTCDVADLPCRVPLKQESWDKYSAEWLCKKSREALDGEDP
VTCAVDLGIRRLGAASVRRDGKIVRARIIREEQQRGCGPRLAAIKNHKRALKAGRRKRGKPVRGEQSLVELQEHVDKMGRDRFK
KGARRIVNFAHQNHCDVIVMEKLAGLIPDAERERGINSALVNWNRGHLVKWVKQLANDAGMRVVEVYPHWTSQLCSRCGAMGAR
FSADHGHVAFEYVGSLFACADCGYVANADHNASVNLHRRFYGELPKVKKLREGVYELAKPDHAPKEVKLAELRSAVSVAAAAMC

RGQASPF (SEQ ID NO: 325)
>3300017963|Ga0180437_10045654_1|M
[水生-非海生盐水和碱性-超盐水湖沉积物]
MATKTYRARIVQISPRLGAALERTHYLFAEALNQMVRRYIDMRKGKYGQECKKLADIMLTRSNTFAHGVMDQLTREISTSGLTD
EWVDLARAVHKSHGPLFLQHDGFAEVDGVMLHTKRHGKVSPMPGRLPVNAKFWHQVCDSASAFLKSNNELLENWRQGRKEWLQE
KADWESANADFMRFWNGSYQEFEETCEIKRKESQALEGQKVVEKKRHRRDRGKRMSRWHLWYNWVITHPEIIEWRNRAKAAEFR
PVPPTVHKKIKNKYPRQDKHITKFLDWLKENNPELKALDSRRRFYVRNYLRFRRPPTLTLPSANKHPYWFTFELNQFYKNVDFD
KGSIQLLLVDETEEGSWFFRWFDAKIQCDPRLRPCVRAGVFATKGRYPPYIGGKVGNKLNRPAERPENRKAGYAGAKLVLNKRR
KELFFTVIEQDCPRKVSWKKTKQRTCRGDNVFAKDGSQIPLKVMAIDLGIRHIGAYVIAEACRNRGVWDISWQHKGITGRHAVP
GLKEIRQHDRQLRKARSQRGKPVQGEESFIELQDHRTNMADDRFKKAANTIVELARQYDVHLILFERLGNLSPTAFDERWMNRQ
LRDMNRRQIVDTVKLQAREFGIECLEDISPWMTSHICSRCFKPGWRFSIKRKNPYGERFTRKNCRDYGYPVWDTGGHLFRCPHC
EYKCNADINAACNVAAKFFGLWSGLDGLKRDQWVYTWQENQPKTFDAKKTFLEWALDVRNRKLMGDSPF (SEQ ID NO: 326)
>3300017971|Ga0180438_10003700_2|P
[水生-非海生盐水和碱性-超盐水湖沉积物]
MATKTYRARIVQISPRLGAALERTHYLFAEALNQMVRRYIDMRKGKYGQECKKLADIMLTRSNTFAHGVMDQLTREISTSGLTD
EWVDLARAVHKSHGPLFLQHDGFAEVDGVMLHTKRHGKVSPMPGRLPVNAKFWHQVCDSASAFLKSNNELLENWRQGRKEWLQE
KADWESANADFMRFWNGSYQEFEETCEIKRKESQALEGQKVVEKKRHRRDRGKRMSRWHLWYNWVITHPEIIEWRNRAKAAEFR
PVPPTVHKKIKNKYPRQDKHITKFLDWLKENNPELKALDSRRRFYVRNYLRFRRPPTLTLPSANKHPYWFTFELNQFYKNVDFD
KGSIQLLLVDETEEGSWFFRWFDAKIQCDPRLRPCVRAGVFATKGRYPPYIGGKVGNKLNRPAERPENRKAGYAGAKLVLNKRR
KELFFTVIEQDCPRKVSWKKTKQRTCRGDNVFAKDGSQIPLKVMAIDLGIRHIGAYVIAEACRNRGVWDISWQHKGITGRHAVP
GLKEIRQHDRQLRKARSQRGKPVQGEESFIELQDHRTNMADDRFKKAANTIVELARQYDVHLILFERLGNLSPTAFDERWMNRQ
LRDMNRRQIVDTVKLQAREFGIECLEDISPWMTSHICSRCFKPGWRFSIKRKNPYGERFTRKNCRDYGYPVWDTGGHLFRCPHC
EYKCNADINAACNVAAKFFGLWSGLDGLKRDQWVYTWQENQPKTFDAKKTFLEWALDVRNRKLMGDSPF (SEQ ID NO: 326)

[0386]

>3300017991|Ga0180434_10065880_1|P
[水生-非海生盐水和碱性-超盐水湖沉积物]
MSVKTIQAKLECDRETLETLWRTHTVFNEGLLPILALFFKMKRRETGSTEEEKDLYKRICEFILRTNASNPGYLLNSVCLENWK
PDSARKTKAEVVRDGKKEKISAEEWIDDLLNMKEESGLLFLKQELLNGLPIMFFEKLVAESKAILSAYEELQYQWEVDHNKWLK
EKEGWEKKHKEYLAVRSLFQKFEQEVKGKITKKRRRWHLYKDWMKRTPELVAWRGKETKVYGLTEEDREAIKKAPLKKKIGVEV
GLFFKRNPELSALDKLHSYYEEKFVRRGKKKRNCDGYEHRPTFTLPHSVKHPRWYVFSGPQTGNGYKNLNIPERKKDFGQLDLL
VLTDDKVDGEYPREYRTVKFRGDHRFLDIRGATRTRLVTKGKNKGQEVEEKFAYEMYDRYLKIWRPLKISGAKLAFRDIKINQD
GSLKSAVPYLWVTLNLEDLPISEKAKEIKWVETDEVKDNGDKRKKIQIPEGLITCAVNIGIRNLGFGTLAVGEKGKARILRSRN
IWIGEEEKTGKHKGRWQKGPTLRELRDHKQEIRRLRKLRGKVVKGERSHIKLQNHITNFADDRLKKAARKILNFALNVENHIDK
KTGKPYPRADVLLLEDLSTLVPTTKKRRDQNQLLIEFMHRGTLSKIEEMAKDAGLKVLRIKPGGISQVCPKCGGIGRRYEIRID
QETKKPDIHFTDVIDDIDCQKLFVCASCSYRGNGDHVASVNLHRKFYDWSVVDKYFTLLRSDNYKEQKKVIEEELREFLRKEHF
PQMALN (SEQ ID NO: 327)
>3300017992|Ga0180435_10008274_7|P
[水生-非海生盐水和碱性-超盐水湖沉积物]
MTIKSFEAKIVTNPSLRETLEYTHRVYNEDLKYMVEVLLAMRKGKHGDGYVEFWREHLQKAANPSGLLEPLTTPDGPKADAKSK
KSKIIRQLFKEHGMLFDRDERWPHWNARHRRHLFDDAKKWARQLDERLERWRSEHAQWLKEKEKFENENPLYMAVEPLFGGFFK
EFGQGGKRRERWAAYYHFLLAHPELANWRGQADAIVPLTPEEEAACLRDRRRGVKKRLDLFLEKNPELEALHKLDAKYGRNMVR
RSARHRNIDGFRHPPTMTLPHSRLHPEWITFQAGETYKDLDLNGGTVRLNILPPDAETSKLHSVEFRAEKRLRRITPRETPLRK
GRNEFNKNYRDAADRVHPARVKGIRLILRPRGAYLKITVELDRPESRTGLKQQTCDKYSSNWCYQEIVKRREGKPLKCLSLDLG
VRHLAVGVVVQETRANGDRETVIEGVRFFKEVGGRPFPSLDEIRRHKDRLREKRSQTRRAPKGRRTCIALQDHVTKMGQDRYRK
GAAAIVRYAVERDVDVILLEKLAGFIPDAAKQRGINKALMEWNRRKIADFVKLLAADHGISVVEVAPHHLSQLCSKCGELGRRF
SWKDGLPKFGTVEKLFVCPACGYQANSDYNAAVNLYKAFSGAYEERPKKAKKGVYKYRDRTICLKKLDGQLAERLTQEQTPF
(SEQ ID NO: 328)
>3300018080|Ga0180433_10080815_1|M

[水生-非海生盐水和碱性-超盐水湖沉积物]
MSVKTIQAKLECDRETLETLWRTHTVFNEGLLPILALFFKMKRRETGSTEEEKDLYKRICEFILRTNASNPGYLLNSVCLENWK
PDSARKTKAEVVRDGKKEKISAEEWIDDLLNMKEESGLLFLKQELLNGLPIMFFEKLVAESKAILSAYEELQYQWEVDHNKWLK
EKEGWEKKHKEYLAVRSLFQKFEQEVKGKITKKRRRWHLYKDWMKRTPELVAWRGKETKVYGLTEEDREAIKKAPLKKKIGVEV
GLFFKRNPELSALDKLHSYYEEKFVRRGKKKRNCDGYEHRPTFTLPHSVKHPRWYVFSGPQTGNGYKNLNIPERKKDFGQLDLL
VLTDDKVDGEYPREYRTVKFRGDHRFLDIRGATRTRLVTKGKNKGQEVEEKFAYEMYDRYLKIWRPLKISGAKLAFRDIKINQD
GSLKSAVPYLWFTLNLEDLPISEKAKEIKWVETDEVKDNGDKRKKIQIPEGLITCAVNIGIRNLGFGTLAVGEKGKARILRSRN
IWIGEEEKTGKHKGRWQKGPTLRELRDHKQEIRRLRKLRGKVVKGERSHIKLQNHITNFADDRLKKAARKILNFALNVENHIDK
KTGKPYPRADVLLLEDLSTLVPTTKKRRDQNQLLIEFMHRGTLSKIEEMAKDAGLKVLRIKPGGISQVCPKCGGIGRRYEIRID
QETKKPDIHFTDVIDDIDCQKLFVCASCSYRGNGDHVASVNLHRKFYDWSVVDKYFTLLRSDNYKEQKKVIEEELREFLRKEHF
PQMALN (SEQ ID NO: 329)
>3300009503|Ga0123519_10002165_1|M
[水生-温泉-热泉]
MATKSFEAKIVCKPDEKYTAEQKKQFLWFTHQVFNDGVRKVIPYVFKMKRGELGPEFQAIYYAITSSQDAIGKLEAVINPDWTS
GKIGKSDPNKWKELLKYQELEKGFRQRLKEEGIKSTKKFRKELEDEKKKLAKEIGQKDIWADAAAILRNKNLLLFNRDELLPNL
PSEFRRKIYEMTIQLIHGHQELVANWEDEHAEWLIEKDKWEEEHPEYMNVRPIFEKFEKEQGKVKGSRIRWLAYLDFLSSKPEL
ANWRGKAKETIPLTKEERAGFRKPGQHFAAFFNKNPELQELDRLHKEYQEKFARTQSKRTPHPDGFKHRPTFTLPDAMRHPVWY
SFKGATDPTKGSTYRNLDLENCTLDLKVLTAMEGEGRNPGGMIQYAFEPDERIKGFRYVGTTEKGKRAKGYIYYDPILEKERPA
KIQGIKLVFRPPRPDGTAYLIFSCQIEDEKPKIKIWKDKEEESPGEITKRKKTEVYPPELITLAIDFGQRHLGAITICKNNNGR
PEPIRFIPAYPKRRKDRESKPVSAWLAKIPGLTFNAVGMHEKEISAGMSRRFQDPKSIRQAGEKEGRKSKGQHIPETETPWAHL
REHIANMKEDHYKKAANLIIRTALQNGAQVILIENLRNYRPMLERTNLENRRRMQWAVRQTAKFLEDTARPLGLIVRQVSSAYT
SRFCSSCGHPGARVSLPGQKNWEKFYAEKYGKERKMIAVAGGQFFCCPACKKIINADINASLNMHKVFYQNFIWPGKIDKKDTK
NFIWQGKNYNWDQIADDVQSFLDQKAGIKKEDDIPY (SEQ ID NO: 330)

[0387]

>CAAACB010000011_94|M
[生物反应器宏基因组]
MSTRAIQARVVADKALLQKLWLTHRAFNEHLPHFSQTLFRMMRGELGDSDQDKAQWAGFVKGVLAQPSKNAYDCLIKATNQTTR
SDTTAAKALSKMEAAARKDIDALNAVPEAKRDPEKFAEASNRLEICRLYNAAFEEFRKGNRPLAHSRQQLDALPNFIERKVIEE
SVAILRSHLELVELWKAEWRQWSDDKAKWEADEVNRRYLAIRPIFEQFEESSGRSAHRSFRWKAYFEFLKSHPELAAWRGGAAV
VHPITDAGLRAIQRARLRKKDQIFAEEFFKANPELEALHRRHMEYEKRFVRRRKSKRSKAWITGFKNRPTFTLPDALRHPRWFV
FNAPQTAPAGYRKLRLPNTPGGEGSMELELLSGTMEKGEHKTVRTTIRFRGDPRLALFKPVMVKRKAIKGAIKGQETDKAAYVF
SDPQLHKDRPAKIGGAKLILRNIKLDEAGQLVSATPYIYFTVDVESPRWSPLAAAIKSVDDVPGGDGRKSRRRSLPDGLVVCTV
DLGIRNLGFATLARIEDGKPRILRSRNLWLGFEETGKHEGRWASGPDLTHISNHKQELRRLRRMRGRPVAGEESHAELQAHIDH
MAEDRFKKAARLITDFALNTDRNSPKGQETPYPSADILVVERLAGLVPQADKERGINRALIEWNRGQLVERLKLVTESVGLRLF
EIGAAGTSQVCSRCGALGRRYSIVPDPHPLFRPAPHVIRFSEVEKLFACPACGYRANADHNAAVNLQRRFALGETAIAAAMDLW
HRPKPEKEAMRKSIEAQLRSKLEIEHKLAEPLP (SEQ ID NO: 331)
>SRR5251193_megahit_k177_201695_120|P
[生物反应器宏基因组]
MGIKSWRVKLWVESPEIRATLEYCLSLWNTSAALLCRHVLAMRRGDFGPEGQAIWEHLKKTPHNFHVLYPISNGSGIEKTTRYY
PDARAFVNKYGYGALGPVRSASGCLWKQLTKAVRERTHSYDKTCEKWSEEHSKWLEERAKWEAEHPEYMAIRPLIDTFHAKHGV
TRGRRNRWALWLQFLMETPELVRWRNKTAKVVPLTDSELKRIRSKRRGAARREQALFYEKNPELAELNRIHGIYERRFARSAAK
RKNPDGFRHRPTFTLPSEAKYPDWPRFQAGEGWRDLDLDRRTICLKLPTDGGSARWVQVHFVPDPRLQRLARLKQPVKEGATRY
DYSLDMPSGHTVLAQPQEVRLIRRGSSYFATIRLSLHSAPSAIPVPQTAASSPAGMLRKIRTDAPELKLITCAVDLGVRHLGAA
TIAREGEIIGRRILHNRFLAPGADGGVINIPSLPQIAQVRRQIRCAQRRSGRISPKVAGCRHLWTHYRHLCEDRYKKAVAAIFG
FARAHGAHIVLFEDLKMLNPDSANERGVNRALQNWNRGRIVEFAKNTAEDAGLCVVTVPAYWTSRICARCNAFGVRFNQGQRRK
PGDDGHVSRVACELDFLGHWFLCPSCGRSVHADLNASENLHRVFLGTFPTVQTIQRQPRIYEINGQRVALDEIQQRAVEILASR
RAGDTPF (SEQ ID NO: 332)
>SRR5251487_megahit_k177_199995_1|P

[生物反应器宏基因组]
MGIKSWRVKLWVESPEIRATLEYCLSLWNTSAALLCRHVLAMRRGDFGPEGQAIWEHLKKTPHNFHVLYPISNGSGIEKTTRYY
PDARAFVNKYGYGALGPVRSASGCLWKQLTKAVRERTHSYDKTCEKWSEEHSKWLEERAKWEAEHPEYMAIRPLIDTFHAKHGV
TRGRRNRWALWLQFLMETPELVRWRNKTAKVVPLTDSELKRIRSKRRGAARREQALFYEKNPELAELNRIHGIYERRFARSAAK
RKNPDGFRHRPTFTLPSEAKYPDWPRFQAGEGWRDLDLDRRTICLKLPTDGGSARWVQVHFVPDPRLQRLARLKQPVKEGATRY
DYSLDMPSGHTVLAQPQEVRLIRRGSSYFATIRLSLHSAPSAIPVPQTAASSPAGMLRKIRTDAPELKLITCAVDLGVRHLGAA
TIAREGEIIGRRILHNRFLAPGADGGVINIPSLPQIAQVRRQIRCAQRRSGRISPKVAGCRHLWTHYRHLCEDRYKKAVAAIFG
FARAHGAHIVLFEDLKMLNPDSANERGVNRALQNWNRGRIVEFAKNTAEDAGLCVVTVPAYWTSRICARCNAFGVRFNQGQRRK
PGDDGHVSRVACELDFLGHWFLCPSCGRSVHADLNASENLHRVFLGTFPTVQTIQRQPRIYEINGQRVALDEIQQRAVEILASR
RAGDTPF (SEQ ID NO: 332)
>SRR5576005_megahit_k177_1292294_1|M
[生物反应器宏基因组]
MSTRAIQARVVADKALLQKLWLTHRAFNEHLPHFSQTLFRMMRGELGDSDQDKAQWAGFVKGVLAQPSKNAYDCLIKATNQTTR
SDTTAAKALSKMEAAARKDIDALNAVPEAKRDPEKFAEASNRLEICRLYNAAFEEFRKGNRPLAHSRQQLDALPNFIERKVIEE
SVAILRSHLELVELWKAEWRQWSDDKAKWEADEVNRRYLAIRPIFEQFEESSGRSAHRSFRWKAYFEFLKSHPELAAWRGGAAV
VHPITDAGLRAIQRARLRKKDQIFAEEFFKANPELEALHRRHMEYEKRFVRRRKSKRSKAWITGFKNRPTFTLPDALRHPRWFV
FNAPQTAPAGYRKLRLPNTPGGEGSMELELLSGTMEKGEHKTVRTTIRFRGDPRLALFKPVMVKRKAIKGAIKGQETDKAAYVF
SDPQLHKDRPAKIGGAKLILRNIKLDEAGQLVSATPYIYFTVDVESPRWSPLAAAIKSVDDVPGGDGRKSRRRSLPDGLVVCTV
DLGIRNLGFATLARIEDGKPRILRSRNLWLGFEETGKHEGRWASGPDLTHISNHKQELRRLRRMRGRPVAGEESHAELQAHIDH
MAEDRFKKAARLITDFALNTDRNSPKGQETPYPSADILVVERLAGLVPQADKERGINRALIEWNRGQLVERLKLVTESVGLRLF
EIGAAGTSQVCSRCGALGRRYSIVPDPHPLFRPAPHVIRFSEVEKLFACPACGYRANADHNAAVNLQRRFALGETAIAAAMDLW
HRPKPEKEAMRKSIEAQLRSKLEIEHKLAEPLP (SEQ ID NO: 331)
>UOUK01008448_2|M

[0388]

[生物反应器宏基因组]
MHLWRTHCVFNQRLPALLKRLFAMRRGEVGGNEAQRQVYQRVAQFVLARDAKDSVDLLNAVSLRKRSANSAFKKKATILRNGQR
REVTGEEVFAEAVALASKGVFAYDKADMRAGLPDSLFQPLTRDAVACMRSHEELVETWKKECREWLDRKSKWEAEPEHALYLKL
RPKFEKGEAERGGRFRKRAERDRAYLDWLEANPQLAAWRGKAPPAVVPIDEAGKRRIARAKAWKQASVRAEEFWKRNPELHALH
KIHVQYLREFVRFKQRPTFTMPDPVRHPRWCLFNAPQTSPQGYRLLRLPQSRRTVGSVELRLLTGPSDGAGFPDAWVNVRFKAD
PRLAQLRPVKVARTVTRGKNKGAKVEADGFRYYDDQLLIERDAQVSGVKLLFRDIRMAPFADKPIEDRLLSATPYLVFAVEIKD
EARTERAKAIRFDETSELTKSGKKRKTLPAGLVSVAVDLDTRGVGFLTRAVIGVPEIQQTHHGVRLLQSRYVAVGQVEARASGE
AEWSPGPDLAHIARHKREIRRLRQLRGKPVKGERSHVRLQAHIDRMGEDRFKKAARKIVNEALRGSNPAAGDRYTRADVLLYES
LETLLPDAERERGINRALLRWNRAKLIEHLKRMCDDAGIRHFPVSPFGTSQVCSKCGALGRRYSLARENGRAVIRFGWVERLFA
CPNPECPGRRPDRPDRPFTCNSDHNASVNLHRVFALGDQAVAAFRALPPRDSPARTLAVKRVEDTLRPQLMRVHKLADAGVDSP
F (SEQ ID NO: 333)
>UOUL01003058_1|M
[生物反应器宏基因组]
MHLWRTHCVFNQRLPALLKRLFAMRRGEVGGNEAQRQVYQRVAQFVLARDAKDSVDLLNAVSLRKRSANSAFKKKATISCNGQA
REVTGEEVFAEAVALASKGVFAYDKDDMRAGLPDSLFQPLTRDAVACMRSHEELVATWKKEYREWRDRKSEWEAEPEHALYLNL
RPKFEEGEAARGGRFRKRAERDHAYLDWLEANPQLAAWRGKAPPAVVPIDEAGKRRIARAKAWKQASVRAEEFWKRNPELHALH
KIHVQYLREFVRPRRTRRNKRREGFKQRPTFTMPDPVRHPRWCLFNAPQTSPQGYRLLRLPQSRRTVGSVELRLLTGPSDGAGF
PDAWVNVRFKADPRLAQLRPVKVPRTVTRGKNKGAKVEADGFRYYDDQLLIERDAQVSGVKLLFRDIRMAPFADKPIEDRLLSA
TPYLVFAVEIKDEARTERAKAIRFDETSELTKSGKKRKTLPAGLVSVAVDLDTRGVGFLTRAVIGVPEIQQTHHGVRLLQSRYV
AVGQVEARASGEAEWSPGPDLAHIARHKREIRRLRQLRGKLVKGERSHVRLQAHIDRMGEDRFKKAARKIVNEALRGSNPAAGD
PYTRADVLLYESLETLLPDAERERGINRALLRWNRAKLIEHLKRMCDDAGIRHFPVSPFGTSQVCSKCGALGRRYSLARENGRA
VIRFGWVERLFACPNPECPGRRPDRPDRPFTCNSDHNASVNLHRVFALGDQAVAAFRALAPRDSPARTLAVKRVEDTLRPQLMR
VHKLADAGVDSPF (SEQ ID NO: 334)
>3300009652|Ga0123330_1010394_3|M

[生物反应器-厌氧-厌氧沼气反应器]

MHLWRTHCVFNQRLPALLKRLFAMRRGEVGGNEAQRQVYQRVAQFVLARDAKDSVDLLNAVSLRKRSANSAFKKKATISCNGQA
REVTGEEVFAEAVALASKGVFAYDKDDMRAGLPDSLFQPLTRDAVACMRSHEELVATWKKEYREWRDRKSEWEAEPEHALYLNL
RPKFEEGEAARGGRFRKRAERDHAYLDWLEANPQLAAWRRKAPPAVVPIDEAGKRRIARAKAWKQASVRAEEFWKRNPELHALH
KIHVQYLREFVRPRRTRRNKRREGFKQRPTFTMPDPVRHPRWCLFNAPQTSPQGYRLLRLPQSRRTVGSVELRLLTGPSDGAGF
PDAWVNVRFKADPRLAQLRPVKVPRTVTRGKNKGAKVEADGFRYYDDQLLIERDAQVSGVKLLFRDIRMAPFADKPIEDRLLSA
TPYLVFAVEIKDEARTERAKAIRFDETSELTKSGKKRKTLPAGLVSVAVDLDTRGVGFLTRAVIGVPEIQQTHHGVRLLQSRYV
AVGQVEARASGEAEWSPGPDLAHIARHKREIRRLRQLRGKPVKGERSHVRLQAHIDRMGEDRFKKAARKIVNEALRGSNPAAGD
PYTRADVLLYESLETLLPDAERERGINRALLRWNRAKLIEHLKRMCDDAGIRHFPVSPFGTSQVCSKCGALGRRYSLARENGRA
VIRFGWVERLFACPNPECPGRRPDRPDRPFTCNSDHNASVNLHRVFALGDQAVAAFRALAPRDSPARTLAVKRVEDTLRPQLMR
VHKLADAGVDSPF (SEQ ID NO: 335)

>3300012949|Ga0153798_10025413_2|P

[生物反应器-柳枝稷降解]

MRRGDFGPEGQAIWEHLKKTPHNFHVLYPISNGSGIEKTTGYYPDARAFVKKYGYGALGPVRSASGCLWYQLTKAVRERSHSYD
KTCEKWSEEHSKWLEERAKWEAEHPEYMAIRPLIDTFHAKHGVTRGRRNRWALWLQFLMETPELVRWRNKTAKVVPLTDSELKR
IRSKRRGAARREQALFYEKNPELAELNRIHGIYERRFARSAAKRKNPDGFRHRPTFTLPSEAKHPDWPRFQAGEGWRDLDLDRR
TICLRLPTDGGSARWVQVHFVPDPRLQRLARLKQPVKEGATRYDYSLDMPSGHTVLAQPQEVRLIRRGSSYFATIRLSLHSAPS
AIPVPQTAASSPAGMLRKIRTDAPELKLITCAVDLGVRHLGAATIAREGEIIGRRILHNRFLAPGAAGGVINIPSLPQIAQVRR
QIRCAQRRSGRISPKVAGCRHLWTHYRHLCEDRYKKAVAAIFGFARAHGAHIVLFEDLKMLNPDSANERGVNRALQNWNRGRIV
EFAKNTAEDAGLCVVTVPAYWTSRICARCNAFGVRFNQGQRRKPGDDGHVSRVACELNFLGHWFLCPSCGRSVHADLNASGNLH
RVFLGTFPTVQTIQRQPRIYEINGQRVALDEIQQRAVEILASRRAGDTPF (SEQ ID NO: 336)

>SRR7811943_megahit_k177_406423_2|M

[洞穴宏基因组]

[0389] MGVRAIRARIECDRETLLSLWRTHRVFNERLPSLIAKLFAMRRGEFGSNAAQRKLYERAVRFILARDSKDAPYLLNSVSIKGWK
HATALKMKARVPDGAGDVQEVTGDSWAAEAATLSANGTLAYDKETERDGLPDSLFQPLIRDAVAYISGHDELVRNWRKEHEDWL
KKRAEWETKDEHKKYLALRPQFDQFEEAVGGKAGKRRERWHLYLKWLRDNPDLAAWRGGAAVVRPLDARAQGRVARAKPWKQRS
VEAEEFWKANPELSALDRLHGFYEREFVRRRKTKRNPDGFDHRPTFTLPDPVLHPRWALFNAPQTSPSGYRALSLPMGRAGFGS
VELCLLTANRDGREWPRAWKSIRFRSDPRLAQFCEHTETTTFVKGKKKGQAKTRQSFSYTDNQLGGINRPAQISGAKLILKDLR
LNDDGSLKSATPYVVFACTIDDQPLTDRARAIQWSETGEVTKTGKKRKAKKLPEGLVACAVDLGVRNLGFLTVAVASDNPGDGV
RVLRSRNLWLGEFEGDGQWSPGPHLAHIGAHKREIKRLRRLRGKPVKGEESHIELQQHIDGMGKDRFKKAARAIVNFALNVEDR
ASPRTGEVYPRADVLIVEKLAGFIPDAERERGINRALVAWNRGQLVTRLKEVAKDAGLKLFEVHPAGTSQCCSKCGALGRRYSI
VRDDETRAPVIRFGWVEKLFACPNLNCAGRKSMHCDQPFVCNADHNASVNLHRRFVMDDKAVAAFFAWKARPDGEREQSVRSID
DALRPQLERIHGLSASPLETPF (SEQ ID NO: 337)

>SRR5818194_megahit_k177_802550_7|M

[淡水宏基因组]

MKLLFKLRRGKIGKNKQQRDDYKRWFEYVIYNGAKNADYNLNALSINNWTPNTAKKYITKLMAGKNKIHAKGLDLVVSRIKSLS
DGGILAFDKDKIFIEIPYSFSQLIIRDAVAGISGHDALVKIWKEDHKKWRDEKKSWEDKTENKQYLKLQHIFKQFSDTAGGVGS
KKRGRWFTYLCWLSKQSALANWRGNGSFIEPFGEENCGKPSCNKGFKSKLIKESKKFFDANPELFEIHKLHQFYEKNFIRPKAK
KHNIDGFEHPPTFTLPDPVRHPIWACYNAPQTNPIGYKDLVLPKKYSGKSAGSFALNLLTGDIIDGKYPKEWVPFKFIADKRLS
CFKGKIKVVNNKIRTEYIFEDPHLSKKRQATIGGIKIILKDIKIDNLGHLLSATPYVVFACDVDNEPISEQAKKLKSEDTGETG
NSGLPIRKRILPEGCLISCAVDLGIRNLGFATLAESGGIPSECTHGDIKLLRSYNLLLEDGPSLEKIAKQKKLLGKFRRLRGKP
IKGEKSHVFLQNHITKMGVDRFNKAARLIVNFALNTEKKKNVQGIEYPRADVIVLENLAGLIPDAAKRKGINKALMNWNRGHLV
KRIQEIAHDSGYLYRVYIMSPWGSSQVCCKCKGLGRRYSIIKDENGRQIQFGNGEKLFACANCRYMANADHNASINLHHIFKDK
DAVGFYKDYANKNEEEKQKIIDNLEAILRPHLQKMHFTTKKERSLDVPEHPLDVPF (SEQ ID NO: 305)

>SRR5818194_megahit_k177_802550_6|P

[淡水宏基因组]

MGTRAIKAKIICDKTTLGYLWRSHCMYNQKLSEYMKLLFKLRRGKIGKNKQQRDDYKRWFEYVIYNGAKNADYNLNALSINNWT

[0390]

PNTAKKYITKLMAGKNKIHAKGLDLVVSRIKSLSDGGILAFDKDKIFIEIPYSFSQLIIRDAVAGISGHDALVKIWKEDHKKWR
DEKKSWEDKTENKQYLKLQHIFKQFSDTAGGVGSKKRGRWFTYLCWLSKQSALANWRGNGSFIEPFGEENCGKPSCNKGFKSKL
IKESKKFFDANPELFEIHKLHQFYEKNFIRPKAKKHNIDGFEHPPTFTLPDPVRHPIWACYNAPQTNPIGYKDLVLPKKYSGKS
AGSFALNLLTGDIIDGKYPKEWVPFKFIADKRLSCFKGKIKVVNNKIRTEYIFEDPHLSKKRQATIGGIKIILKDIKIDNLGHL
LSATPYVVFACDVDNEPISEQAKKLKSEDTGETGNSGLPIRKRILPEGCLISCAVDLGIRNLGFATLAESGGIPSECTHGDIKL
LRSYNLLLEDGPSLEKIAKQKKLLGKFRRLRGKPIKGEKSHVFLQNHITKMGVDRFNKAARLIVNFALNTEKKKNVQGIEYPRA
DVIVLENLAGLIPDAAKRKGINKALMNWNRGHLVKRIQEIAHDSGYLYRVYIMSPWGSSQVCCKCKGLGRRYSIIKDENGRQIQ
FGNGEKLFACANCRYMANADHNASINLHHIFKDKDAVGFYKDYANKNEEEKQKIIDNLEAILRPHLQKMHFTTKKERSLDVPEH
PLDVPF (SEQ ID NO: 306)

>SRR6481359_megahit_k177_1860565_4|M

[淡水沉积物宏基因组]

MSVRSFQAKVECDKQIMEHLWRTHTVFNERLPVIIKILFQMKRGECGQNDKQKALYQNIAKSILDTNAQNADYLLNSVSIKGWK
PGAARKYNKGSFTWLDEAANLSSKGILVYDKDQVLGDIPLMMRQMICRQSVETIKSQQEILKNWENEYKEWLKKKEEWESQEDN
KKYLALRDKFEQFRKSYGGKKSKKRGRWHLYLKWLRDNPDLAAWRGGKAEINPISEKAQSRINKAKPNKKNAIELEEFLEANPE
IKKLDELHGYYERTFIRRRETRKGKELTGFKQKPTFTLPHSTTHPRWFVFNAPQTNPAGYRNLILPQKTGALGSLEMRLLNGDK
NNGDYPDDWINVRFKADPRLALMRGIKIQRIVRRGKDAGQTKESDSYKYFDKFLNKWRPATLSGVKLIFKFNPDKSVNTAYLYF
TCDITNEPLTETAKKIQWIETGEVTKKGKIRKKKVLPEGLISCAVDLGMRKGTTGFGTLCKYKNGKPYILRSRNLWIRYKEEKG
SHSGRWAEGPDLGHIAKHKREIRFLRRKRGKPIKGEKSHIKLQQHIDDMGEDRFKKAARAIVNFALNTENVAGKNGIYPRADIL
LLENLEGLIPDAERERGINRALAGWNRRHLVERIKEMAEDAGFRRRVFEIPPYGTSQVCSKCGSLGRRYTLGINEETGKREIHF
GYVEKLFACPNCGYCANADHNASVNLHRRFLIDDALKNYYEFKKLPEKKQKEAIAAIENSLKEKLSTIHKI(SEQ ID NO: 338)

>SRR6049666_megahit_k177_256604_1|P

[热泉宏基因组]

MATKSFEAKIVCKPDEKYTAEQKKQFLWFTHQVFNDGVRKVIPYVFKMKRGELGPEFQAIYYAITSSQDAIGKLEAVINPDWTS
GKIGKSDPNKWKELLKYQELEKGFRQRLKEEGIKSTKKFRKELEDEKKKLAKEIGQKDIWADAAAILRNKNLLLFNRDELLPNL
PSEFRRKIYEMTIQLIHGHQELVANWEDEHAEWLIEKDKWEEEHPEYMNVRPIFEKFEKEQGKVKGSRIRWLAYLDFLSSKPEL
ANWRGKAKETIPLTKEERAGFRKPGQHFAAFFNKNPELQELDRLHKEYQEKFARTQSKRTPHPDGFKHRPTFTLPDAMRHPVWY
SFKGATDPTKGSTYRNLDLENCTLDLKVLTAMEGEGRNPGGMIQYAFEPDERIKGFRYVGTTEKGKRAKGYIYYDPILEKERPA
KIQGIKLVFRPPRPDGTAYLIFSCQIEDEKPKIKIWKDKEEESPGEITKRKKTEVYPPELITLAIDFGQRHLGAITICKNNNGR
PEPIRFIPAYPKRRKDRESKPVSAWLAKIPGLTFNAVGMHEKEISAGMSRRFQDPKSIRQAGEKEGRKSKGQHIPETETPWAHL
REHIANMKEDHYKKAANLIIRTALQNGAQVILIENLRNYRPMLERTNLENRRRMQWAVRQTAKFLEDTARPLGLIVRQVSSAYT
SRFCSSCGHPGARVSLPGQKNWEKFYAEKYGKERKMIAVAGGQFFCCPACKKIINADINASLNMHKVFYQNFIWPGKIDKKDTK
NFIWQGKNYNWDQIADDVQSFLDQKAGIKKEDDIPY (SEQ ID NO: 330)

>SRR2080392_megahit_k177_207267_1|M

[热液喷口宏基因组]

MLVLPHGAHGAASSIICPFGIKNISAKWDSPEYHGQTYYIMPIMKRRVIQAKISDTDPDTRQKIHLTHQIFNEYVKQTIKLFLS
IMNGTYDHPAYKTNKELYRRWMASIIFAGSGGGSINYDYLINSICKTKSKRHWEGASIPEKVTIPTVNQITDQKGKKTAKIEKE
EHDFKSFSSVDQIVELAKQQIYLFDKSEAFPLPSQIISAVISEAIAIIKQHRTLITKWYEDNGDFETEQAEFLNSPRYKSFLKI
KYIFDAFEQENKLRAERKRSQRKRSLRWERWDKYVNLLIDNSELMQWREKNTTLTIPPALNDPAKNKKSRLGILLRANPELGAL
HRTYVETWKDIPPQRPTFCWPDINEHPRFFTASSPPTSPYAYQRLSIDHKKSKATVEFFLLDGKRDKNGRYIGNFYLLRVVFDP
RINRFEHNPNAITKKGNKKPNSYYFLDPRTKKKREAHIKGCMLFLKKQKTSNHTSAYLTFNYAIERSNISEQAKKVRYDEKAKG
WKAPIGAIVASVDMMYKEIAFVTIAEKIKPQEFTILASRKIKLHRNQDNGFKGPQAKQIEQHQISLRDRKRNLPRKQYPQKLQN
HLNNMREEFSKRSARRILDFVMNLNKTIKDSQDKIIEQVDVILLERLKGLKPQKKNPRGLNRVLSNKRRPKVVTAIQEGCEIAG
ISCREIMPHYTSQICSKCGKLGRRFYENRKDVHEVERKMSPFGGKLFICPHCGYQANAEFNATINLLNIVLLNKRMTPEDKTNL
SPAEKRNYYTQLDDLAMKKLSPQKSPVKV (SEQ ID NO: 339)

>SRR2080392_megahit_k177_207267_2|P

[热液喷口宏基因组]

MKRRVIQAKISDTDPDTRQKIHLTHQIFNEYVKQTIKLFLSIMNGTYDHPAYKTNKELYRRWMASIIFAGSGGGSINYDYLINS

[0391]

ICKTKSKRHWEGASIPEKVTIPTVNQITDQKGKKTAKIEKEEHDFKSFSSVDQIVELAKQQIYLFDKSEAFPLPSQIISAVISE
AIAIIKQHRTLITKWYEDNGDFETEQAEFLNSPRYKSFLKIKYIFDAFEQENKLRAERKRSQRKRSLRWERWDKYVNLLIDNSE
LMQWREKNTTLTIPPALNDPAKNKKSRLGILLRANPELGALHRTYVETWKDIPPQRPTFCWPDINEHPRFFTASSPPTSPYAYQ
RLSIDHKKSKATVEFFLLDGKRDKNGRYIGNFYLLRVVFDPRINRFEHNPNAITKKGNKKPNSYYFLDPRTKKKREAHIKGCML
FLKKQKTSNHTSAYLTFNYAIERSNISEQAKKVRYDEKAKGWKAPIGAIVASVDMMYKEIAFVTIAEKIKPQEFTILASRKIKL
HRNQDNGFKGPQAKQIEQHQISLRDRKRNLPRKQYPQKLQNHLNNMREEFSKRSARRILDFVMNLNKTIKDSQDKIIEQVDVIL
LERLKGLKPQKKNPRGLNRVLSNKRRPKVVTAIQEGCEIAGISCREIMPHYTSQICSKCGKLGRRFYENRKDVHEVERKMSPFG
GKLFICPHCGYQANAEFNATINLLNIVLLNKRMTPEDKTNLSPAEKRNYYTQLDDLAMKKLSPQKSPVKV (SEQ ID NO: 340)

>3300025517|Ga0207869_1001046_11|M

[实验室富集限定的培养基-离子液体和高固体富集的]

MGVKAWKAKIIVADPDLAAALDHTVRQWNLWAEQACIAIMRLRRGDYGDEGQAVWRFLSEKGTFGTSILDKITRVTGRGIKNDE
ERTIASNFLSRHGPFEPLRAARDALGPNFWYQFQRAVRERIESHDAKLRQWREDHAEWIETRRTWEEEHRAYMNVRPVIEAFEA
AEGQSGRRRGRWRRWLDFLSSHPELAAWRGGPATIVPLSKEELAACRRRRRRAVAEQARLFFEKNPELRALNREHDEYQEKFAR
PWAKRRNPDGFRHRPTFTLPHPDSHPDWPRFKGNAGWRDLSVPDRTVSLELLCPDGSRRWFNVRFVPDPRLASLRPATQPVRQG
NTTYHYVSDALGVAEQPVQPQGIKLILRPDGAYLSVSVSVLAPPCGFDVRQSAIEKYGSIWACRKYKQDHPDRPLVTCAVDLAV
RHNGALTVARDGEVFARRLLHNRFFLPGDDRERAMNIPSLLEIAAVKRQLRRARRQTGRIAPGKPSCARLQAHYRALCEDRYKK
LVAAIFAYARHHGAEVVIFEDLKNLLPNAVNERGVNRALQSWNRGAIVRFARQSAEDVGLRVATVPAFHTSIVCARCGALGCRF
EQTREGVRIAKLGPWFGCRACGRRIHADLNASENLHKVLIGTFPQVRRISRDPPVLSIADVRVDQREADRLAAAWLSELSVGPT
PF (SEQ ID NO: 301)

>3300025572|Ga0207864_1000213_47|M

[实验室富集限定的培养基-离子液体和高固体富集的]

MGVKAWKAKIIVADPDLAAALDHTVRQWNLWAEQACIAIMRLRRGDYGDEGQAVWRFLSEKGTFGTSILDKITRVTGRGIKNDE
ERTIASNFLSRHGPFEPLRAARDALGPNFWYQFQRAVRERIESHDAKLRQWREDHAEWIETRRTWEEEHRAYMNVRPVIEAFEA
AEGQSGRRRGRWRRWLDFLSSHPELAAWRGGPATIVPLSKEELAACRRRRRRAVAEQARLFFEKNPELRALNREHDEYQEKFAR
PWAKRRNPDGFRHRPTFTLPHPDSHPDWPRFKGNAGWRDLSVPDRTVSLELLCPDGSRRWFNVRFVPDPRLASLRPATQPVRQG
NTTYHYVSDAIGVAEQPVQPQGIKLILRPDGAYLSVSVSVLAPPCGFDVRQSAIEKYGSIWACRKYKQDHPDRPLVTCAVDLAV
RHNGALTVARDGEVFARRLLHNRFFLPGDDRERAMNIPSLLEIAAVKRQLRRARRQTGRIAPGKPSCARLQAHYRALCEDRYKK
LVAAIFAYARHHGAEVVIFEDLKNLLPNAVNERGVNRALQSWNRGAIVRFARQSAEDVGLRVATVPAFHTSIVCARCGALGCRF
EQTREGVRIAKLGPWFGCRACGRRIHADLNASENLHKVLIGTFPQVRRISRDPPVLSIADVRVDQREADRLAAAWLSELSVGPT
PF (SEQ ID NO: 341)

>3300025572|Ga0207864_1014515_2|P

[实验室富集限定的培养基-离子液体和高固体富集的]

MGIKSWRVKLWVESPEIRATLEYCLSLWNTSAALLCRHVLAMRRGDFGPEGQAIWEHLKKTPHNFHVLYPISNGSGIEKTTRYY
PDARAFVNKYGYGALGPVRSASGCLWKQLTKAVRERTHSYDKTCEKWSEEHSKWLEERAKWEAEHPEYMAIRPLIDTFHAKHGV
TRGRRNRWALWLQFLMETPELVRWRNKTAKVVPLTDSELKRIRSKRRGAARREQALFYEKNPELAELNRIHGIYERRFARSAAK
RKNPDGFRHRPTFTLPSEAKYPDWPRFQAGEGWRDLDLDRRTICLKLPTDGGSARWVQVHFVPDPRLQRLARLKQPVKEGATRY
DYSLDMPSGHTVLAQPQEVRLIRRGSSYFATIRLSLHSAPSAIPVPQTAASSPAGMLRKIRTDAPELKLITCAVDLGVRHLGAA
TIAREGEIIGRRILHNRFLAPGADGGVINIPSLPQIAQVRRQIRCAQRRSGRISPKVAGCRHLWTHYRHLCEDRYKKAVAAIFG
FARAHGAHIVLFEDLKMLNPDSANERGVNRALQNWNRGRIVEFAKNTAEDAGLCVVTVPAYWTSRICARCNAFGVRFNQGQRRK
PGDDGHVSRVACELDFLGHWFLCPSCGRSVHADLNASENLHRVFLGTFPTVQTIQRQPRIYEINGQRVALDEIQQRAVEILASR
RAGDTPF (SEQ ID NO: 332)

>3300025572|Ga0207864_1014515_3|M

[实验室富集限定的培养基-离子液体和高固体富集的]

MRRGDFGPEGQAIWEHLKKTPHNFHVLYPISNGSGIEKTTRYYPDARAFVNKYGYGALGPVRSASGCLWKQLTKAVRERTHSYD
KTCEKWSEEHSKWLEERAKWEAEHPEYMAIRPLIDTFHAKHGVTRGRRNRWALWLQFLMETPELVRWRNKTAKVVPLTDSELKR
IRSKRRGAARREQALFYEKNPELAELNRIHGIYERRFARSAAKRKNPDGFRHRPTFTLPSEAKYPDWPRFQAGEGWRDLDLDRR
TICLKLPTDGGSARWVQVHFVPDPRLQRLARLKQPVKEGATRYDYSLDMPSGHTVLAQPQEVRLIRRGSSYFATIRLSLHSAPS

[0392]

AIPVPQTAASSPAGMLRKIRTDAPELKLITCAVDLGVRHLGAATIAREGEIIGRRILHNRFLAPGADGGVINIPSLPQIAQVRR
QIRCAQRRSGRISPKVAGCRHLWTHYRHLCEDRYKKAVAAIFGFARAHGAHIVLFEDLKMLNPDSANERGVNRALQNWNRGRIV
EFAKNTAEDAGLCVVTVPAYWTSRICARCNAFGVRFNQGQRRKPGDDGHVSRVACELDFLGHWFLCPSCGRSVHADLNASENLH
RVFLGTFPTVQTIQRQPRIYEINGQRVALDEIQQRAVEILASRRAGDTPF (SEQ ID NO: 342)
>SRR6505097_megahit_k177_377355_35|M
[永久冻土宏基因组]
MGVRALRARIECDRETLLHLWRTHCAFNQRLLSILTKLFGMRRGELGETAARRALYQRVARFLLARDSKDAPYLLNSVSIQGWK
PATALKMKVRMPAPDGTLAEVSADTWAPEAAKLSAHGTLAYDKEAVRDGLPDSLFQPMLRDAVAYISGHIELARNWEAEHADWL
KAKAEWEEKDEHKKYLALRPRFDAFEQSVGGKAGKRRERWHLYLTWLRENPALAAWRGGPPAVHPLDAKATARVAKAKPWKKRS
VEADEFWKANPELRELDKLHGYYEREFVRRRKTKKNPDGFDHRPTFTLPHAVLHPRWTLFNAPQTSPQGYADLRLPQRPGTLGS
VKLRLLTGEKAPDGCPNEWVGVQFRADPRLSQFAERAEAQEIRKGKKKGTQKTVRVFRFVDQQLNCERPAQISGAKLIFKNVKL
NDDGSLKSATPYLVFACTIDNLPLTEHARAIQWSETGEVTKTGKKRKAKKLPTGLVTCAVDLGLRNLAFLTLATSGEGPQNNTH
DGIRVLRSRNQWLALEETAGTHPGRWSAGPDLAHIAAHKREIRRLRRLRGKPVKGQQSHVELQGHIDHMGQDRFKKAARAIVNF
ALNVDRAAGQSGQPYPRADLLIVEKLEGFIPDAARERGINRALVAWNRGQLVTRIKEVAADAGLKLFPVHPAGTSQCCSRCGAL
GRRYSIVRDQQTGVPAIRFGWVEKLFACPNRECPGRAPDKPDQPFTCNADHNASLNLHRRFVLDDRAVAAFFAWKNKTNAERER
ELTQMDQALRPELEKLHRLPLAELATPF (SEQ ID NO: 343)
>SRR6962291_megahit_k177_1754533_5|M
[永久冻土宏基因组]
MKRGDSGTTDAQKHLYSQIARFVLARSSRDAPYLMNSISIRGWTPNTAKKKMKATLMGPDGQEVETSGDDWADAAAALSAEGTL
LYDKSELLGDVPDTLRQVVVRECAAIISGHEELVRAWELSHEEWLEQKAKWESDPEHERYMVLRPQFETFEQSVGGKAGKRRGR
WQRYLDWLRSNPELAAWRGGPATVVELPAAARERVRKARPSKVRSVEAEEFWKVNPELFELDRLHGFYEREFVRRRKTKKHPDG
FDHRPTFTLPDPVRHPRWFVFNAPQTSPQGYRKLSLPKKNGDFGTIELLLLTGEKTDAKYPNGWNQVRFAADPRLSDFKQISKK
RQIRKGKTKGQETDSNAFTFLDRQISMERPAQISGAKLLFKKVRLKSDGALLSADPYLVFTCNITDVPFTDAARAIKWEDAGER
RKRAVIPAGLITCAVDLGIRHVGFATIAAKDEAHPEGLRVLRSRNVWMGHEEQKGRHPGRWSPGPELPHLASHKRSLRRLRRLR
GKPVKDESSHAELQGHITNMANDRFKKAAREIVNFALNVERKVDPRTGEPYSRADLLVVENLANLIPDSERERGINRALIEFNR
GHLVDRIAEVAKDCGLKLMKVSPVGTSQVCCRCGALGRRYSIQRDRDSGYADIRFGFVEALFACPHCGYRANADHNASINLHRR
FLLGEAAIRGFKEWQQGSRQERADTVRRIEESLVAPLRLMHRLEVADMPF (SEQ ID NO: 344)
>SRR6962291_megahit_k177_1754533_5|P
[永久冻土宏基因组]
MSTRAIQSRIQCDHQTLEHLWRTHVVFNRRLPPLISLLFRMKRGDSGTTDAQKHLYSQIARFVLARSSRDAPYLMNSISIRGWT
PNTAKKKMKATLMGPDGQEVETSGDDWADAAAALSAEGTLLYDKSELLGDVPDTLRQVVVRECAAIISGHEELVRAWELSHEEW
LEQKAKWESDPEHERYMVLRPQFETFEQSVGGKAGKRRGRWQRYLDWLRSNPELAAWRGGPATVVELPAAARERVRKARPSKVR
SVEAEEFWKVNPELFELDRLHGFYEREFVRRRKTKKHPDGFDHRPTFTLPDPVRHPRWFVFNAPQTSPQGYRKLSLPKKNGDFG
TIELLLLTGEKTDAKYPNGWNQVRFAADPRLSDFKQISKKRQIRKGKTKGQETDSNAFTFLDRQISMERPAQISGAKLLFKKVR
LKSDGALLSADPYLVFTCNITDVPFTDAARAIKWEDAGERRKRAVIPAGLITCAVDLGIRHVGFATIAAKDEAHPEGLRVLRSR
NVWMGHEEQKGRHPGRWSPGPELPHLASHKRSLRRLRRLRGKPVKDESSHAELQGHITNMANDRFKKAAREIVNFALNVERKVD
PRTGEPYSRADLLVVENLANLIPDSERERGINRALIEFNRGHLVDRIAEVAKDCGLKLMKVSPVGTSQVCCRCGALGRRYSIQR
DRDSGYADIRFGFVEALFACPHCGYRANADHNASINLHRRFLLGEAAIRGFKEWQQGSRQERADTVRRIEESLVAPLRLMHRLE
VADMPF (SEQ ID NO: 345)
>SRR6962291_megahit_k177_3367137_15|M
[永久冻土宏基因组]
MFNRRLPDLISLLFRMRRGECGETEQASRLYSQVASFILARGSKDAPYLLNSVSVRGWKPHTAKKMKARALGPDGQQIELAGED
WADPAAALSAEGRLLYDKTELLGDVPDTLRQMAARECVAIISGHEELVRAWEKAHEEWLASKAKWESDPENQFYLGLRARFDEF
EQSVGGKVGKRRGRWQRYLDWLRANPDLAAWRGGPAAVVELSAGAKERIRKAKPRKVRSVEAEEFWKVNSELAALDRLHGYYER
EFVRRRKTKKHPDGFDHRPTFTLPHPVRHPRWFVFNAPQTSPQGYRKLCLPSSGGHTGTVELLLLTGEKADNKYPSGWETVRFA
ADPRLSDFKAITKRRQVLKGKTKGQETDTKAFTFLDRQINLERPAQISGVKLIFKGVRQNHDGSLLSAHPYLVFTCNVADVPLT
DGAKKIKWEDTGERRKRVVVPAGLIACAVDLGIRHVGFATVATQDAESGGLHVLRSRNIWIGQEEEKGRNPGRWSPGPDLAHLS

[0393]

AHKRSLRHLRGLRGKPVKGENSHIELQDHITNMASDRFKKAAREIVNFALNADSKVDPRTGEPYPRADVLVLENLANLLPDSER
ERGINRALIEFNRGHLADRIVEVAKDCGLKVMMVSPVGTSQVCCRCGALGRRYSLRRHQDGGYTDIHFGFVEPLFACPSCGYRA
NSDHNASVNLHRRFRLGEAAIAKFKEWQKGTKPERDEVLRRIEQSLVGPLREMHRLAVADTPF (SEQ ID NO: 346)
>SRR6962291_megahit_k177_3367137_14|P
[永久冻土宏基因组]
MSTRALQSRIQCDHQTLEHLWRTHLVFNRRLPDLISLLFRMRRGECGETEQASRLYSQVASFILARGSKDAPYLLNSVSVRGWK
PHTAKKMKARALGPDGQQIELAGEDWADPAAALSAEGRLLYDKTELLGDVPDTLRQMAARECVAIISGHEELVRAWEKAHEEWL
ASKAKWESDPENQFYLGLRARFDEFEQSVGGKVGKRRGRWQRYLDWLRANPDLAAWRGGPAAVVELSAGAKERIRKAKPRKVRS
VEAEEFWKVNSELAALDRLHGYYEREFVRRRKTKKHPDGFDHRPTFTLPHPVRHPRWFVFNAPQTSPQGYRKLCLPSSGGHTGT
VELLLLTGEKADNKYPSGWETVRFAADPRLSDFKAITKRRQVLKGKTKGQETDTKAFTFLDRQINLERPAQISGVKLIFKGVRQ
NHDGSLLSAHPYLVFTCNVADVPLTDGAKKIKWEDTGERRKRVVVPAGLIACAVDLGIRHVGFATVATQDAESGGLHVLRSRNI
WIGQEEEKGRNPGRWSPGPDLAHLSAHKRSLRHLRGLRGKPVKGENSHIELQDHITNMASDRFKKAAREIVNFALNADSKVDPR
TGEPYPRADVLVLENLANLLPDSERERGINRALIEFNRGHLADRIVEVAKDCGLKVMMVSPVGTSQVCCRCGALGRRYSLRRHQ
DGGYTDIHFGFVEPLFACPSCGYRANSDHNASVNLHRRFRLGEAAIAKFKEWQKGTKPERDEVLRRIEQSLVGPLREMHRLAVA
DTPF (SEQ ID NO: 347)
>AQT69685.1
[菲西芬氏菌纲细菌 ST-NAGAB-D1]
MATKSYRARILTDSRLAAALDRTHVVFVESLKQMINTYLRMQNGKFGPDHKKLAQIMLSRSNTFAHGVMDQITRDQPTSTLDEE
WTDLARRIHKTTGPLFLQAERFATVKNRAIHTKSRGKVIPSPETLAVPAKFWHQVCDSASAYIRSNRELMQQWRKDRAAWLKDK
NEWQQKHPEFMQFYNGPYQNFLKLCDDDRITSQLAAEQQPTASKNNRPRKTGKRFARWHLWYKWLSENPEIIEWRNKASASDFK
TVTDDVRKQIITKYPQQNKYITRLLDWLEDNNPELKTLENLRRTYVKKFDSFKRPPTLTLPSPYRHPYWFTMELDQFYKKADFE
NGTIQLLLIDEDDDGNWFFNWMPASLKPDPRLVPSWRAETFETEGRFPPYLGGKIGKKLSRPAPTDAERKAGIAGAKLMIKNNR
SELLFTVFEQDCPPRVKWAKTKNRKCPADNAFSSDGKTRKPLRILSIDLGIRHIGAFALTQGTRNDSAWQTESLKKGIINSPSI
PPLRQVRRHDYDLKRKRRRHGKPVKGQRSNANLQAHRTNMAQDRFKKGASAIVSLAREHSADLILFENLHSLKFSAFDERWMNR
QLRDMNRRHIVELVSEQAPEFGITVKDDINPWMTSRICSNCNLPGFRFSMKKKNPYREKLPREKCTDFGYPVWEPGGHLFRCPH
CDHRVNADINAAANLANKFFGLGYWNNGLKYDAETKTFTVHTDKKTPPLIFKPRPQFDLWADSVKTRKQLGPDPF(SEQ ID NO:
348)
>PMBN01000097_4|M
[菲西芬氏菌目细菌]
MGVRALRARIECDRETLLHLWRTHCAFNQRLLSILTKLFGMRRGELGETAARRALYQRVARFLLARDSKDAPYLLNSVSIQGWK
PATALKMKVRMPAPDGTLAEVSADTWAPEAAKLSAHGTLAYDKEAVRDGLPDSLFQPMLRDAVAYISGHIELARNWEAEHADWL
KAKAEWEEKDEHKKYLALRPRFDAFEQSVGGKAGKRRERWHLYLTWLRENPALAAWRGGPPAVHPLDAKATARVAKAKPWKKRS
VEADEFWKANPELRELDKLHGYYEREFVRRRKTKKNPDGFDHRPTFTLPHAVLHPRWTLFNAPQTSPQGYADLRLPQRPGTLGS
VKLRLLTGEKAPDGCPNEWVGVQFRADPRLSQFAERAEAQEIRKGKKKGTQKTVRVFRFVDQQLNCERPAQISGAKLIFKNVKL
NDDGSLKSATPYLVFACTIDNLPLTEHARAIQWSETGEVTKTGKKRKAKKLPTGLVTCAVDLGLRNLAFLTLATSGEGPQNNTH
DGIRVLRSRNQWLALEETAGTHPGRWSAGPDLAHIAAHKREIRRLRRLRGKPVKGQQSHVELQGHIDHMGQDRFKKAARAIVNF
ALNVDRAAGQSGQPYPRADLLIVEKLEGFIPDAARERGINRALVAWNRGQLVTRIKEVAADAGLKLFPVHPAGTSQCCSRCGAL
GRRYSIVRDQQTGVPAIRFGWVEKLFACPNRECPGRAPDKPDQPFTCNADHNASLNLHRRFVLDDRAVAAFFAWKNKTNAERER
ELTQMDQALRPELEKLHRLPLAELATPF (SEQ ID NO: 343)
>PMDC01000375_4|M
[菲西芬氏菌目细菌]
MGVRALRARIECDRETLLHLWRTHCAFNQRLLSILTKLFGMRRGELGETAARRALYQRVARFLLARDSKDAPYLLNSVSIQGWK
PATALKMKVRMPAPDGTLAEVSADTWAPEAAKLSAHGTLAYDKEAVRDGLPDSLFQPMLRDAVAYISGHIELARNWEAEHADWL
KAKAEWEEKDEHKKYLALRPRFDAFEQSVGGKAGKRRERWHLYLTWLRENPALAAWRGGPPAVHPLDAKATARVAKAKPWKKRS
VEADEFWKANPELRELDKLHGYYEREFVRRRKTKKNPDGFDHRPTFTLPHAVLHPRWTLFNAPQTSPQGYADLRLPQRPGTLGS
VKLRLLTGEKAPDGCPNEWVGVQFRADPRLSQFAERAEAQEIRKGKKKGTQKTVRVFRFVDQQLNCERPAQISGAKLIFKNVKL
NDDGSLKSATPYLVFACTIDNLPLTEHARAIQWSETGEVTKTGKKRKAKKLPTGLVTCAVDLGLRNLAFLTLATSGEGPQNNTH

[0394]

DGIRVLRSRNQWLALEETAGTHPGRWSAGPDLAHIAAHKREIRRLRRLRGKPVKGQQSHVELQGHIDHMGQDRFKKAARAIVNF
ALNVDRAAGQSGQPYPRADLLIVEKLEGFIPDAARERGINRALVAWNRGQLVTRIKEVAADAGLKLFPVHPAGTSQCCSRCGAL
GRRYSIVRDQQTGVPAIRFGWVEKLFACPNRECPGRAPDKPDQPFTCNADHNASLNLHRRFVLDDRAVAAFFAWKNKTNAERER
ELTQMDQALRPELEKLHRLPLAELATPF (SEQ ID NO: 343)
>PMMM01000072_9|M
[菲西芬氏菌目细菌]
MGVRALRARIECDRETLLHLWRTHCAFNQRLLSILTKLFGMRRGELGETAARRALYQRVARFLLARDSKDAPYLLNSVSIQGWK
PATALKMKVRMPAPDGTLAEVSADTWAPEAAKLSAHGTLAYDKEAVRDGLPDSLFQPMLRDAVAYISGHIELARNWEAEHADWL
KAKAEWEEKDEHKKYLALRPRFDAFEQSVGGKAGKRRERWHLYLTWLRENPALAAWRGGPPAVHPLDAKATARVAKAKPWKKRS
VEADEFWKANPELRELDKLHGYYEREFVRRRKTKKNPDGFDHRPTFTLPHAVLHPRWTLFNAPQTSPQGYADLRLPQRPGTLGS
VKLRLLTGEKAPDGCPNEWVGVQFRADPRLSQFAERAEAQEIRKGKKKGTQKTVRVFRFVDQQLNCERPAQISGAKLIFKNVKL
NDDGSLKSATPYLVFACTIDNLPLTEHARAIQWSETGEVTKTGKKRKAKKLPTGLVTCAVDLGLRNLAFLTLATSGEGPQNNTH
DGIRVLRSRNQWLALEETAGTHPGRWSAGPDLAHIAAHKREIRRLRRLRGKPVKGQQSHVELQGHIDHMGQDRFKKAARAIVNF
ALNVDRAAGQSGQPYPRADLLIVEKLEGFIPDAARERGINRALVAWNRGQLVTRIKEVAADAGLKLFPVHPAGTSQCCSRCGAL
GRRYSIVRDQQTGVPAIRFGWVEKLFACPNRECPGRAPDKPDQPFTCNADHNASLNLHRRFVLDDRAVAAFFAWKNKTNAERER
ELTQMDQALRPELEKLHRLPLAELATPF (SEQ ID NO: 343)
>QBM02859.1
[浮霉菌门细菌 RBG_13_46_10]
MSVRSFQARVECDKQTMEHLWRTHKVFNERLPEIIKILFKMKRGECGQNDKQKSLYKSISQSILEANAQNADYLLNSVSIKGWK
PGTAKKYRNASFTWADDAAKLSSQGIHVYDKKQVLGDLPGMMSQMVCRQSVEAISGHIELTKKWEKEHNEWLKEKEKWESEDEH
KKYLDLREKFEQFEQSIGGKITKRRGRWHLYLKWLSDNPDFAAWRGNKAVINPLSEKAQIRINKAKPNKKNSVERDEFFKANPE
MKALDNLHGYYERNFVRRRKTKKNPDGFDHKPTFTLPHPTIHPRWFVFNKPKTNPEGYRKLILPKKAGDLGSLEMRLLTGEKNK
GNYPDDWISVKFKADPRLSLIRPVKGRRVVRKGKEQGQTKETDSYEFFDKHLKKWRPAKLSGVKLIFPDKTPKAAYLYFTCDIP
DEPLTETAKKIQWLETGDVTKKGKKRKKKVLPHGLVSCAVDLSMRRGTTGFATLCRYENGKIHILRSRNLWVGYKEGKGCHPYR
WTEGPDLGHIAKHKREIRILRSKRGKPVKGEESHIDLQKHIDYMGEDRFKKAARTIVNFALNTENAASKNGFYPRADVLLLENL
EGLIPDAEKERGINRALAGWNRRHLVERVIEMAKDAGFKRRVFEIPPYGTSQVCSKCGALGRRYSIIRENNRREIRFGYVEKLF
ACPNCGYCANADHNASVNLNRRFLIEDSFKSYYDWKRLSEKKQKEEIETIESKLMDKLCAMHKISRGSISK(SEQ ID NO: 349)
>SRR6963308_megahit_k177_1876783_1|M
[沉积物宏基因组]
MSQRALAVRLLVDSEDDRAALWRTHRIFNERLGWVLRQMHRMKRGEPDARYAEIFNAITTSQRATAALEAVTSLAWTGTKEDRW
CQLARALIQEGTLLFDREARLAGLSKEFRRKLFEAAFQMISGHRELLAQWERDHNEWQQARETWEAENPEYMAIRPALQRFEAE
HGQAAKRRARWHKWRAFLQGSSQLAGWRGGPQVVNEISADARRRIERARRNKRNKIEAEEFFRVNPELKELDAKHGYYQREFLR
PWAKRRNADGFKHRPTFTEPSAENHPFWFQFKKHETYKDLNLKVRSLRLKLLASDSAVPGSQWRDFRFRVDPRLLDLSEAQSPA
KIGIARFSLLFNDRALAMEDRPAEIRGAKLIFRPAQRDGRVYLSFTCDVADLPCRVPLKQESWDKYSAEWLCKKSREALDGEDP
VTCAVDLGIRRLGAASVRRDGKIVRARIIREEQQRGCGPRLAAIKNHKRALKAGRRKRGKPVRGEQSLVELQEHVDKMGRDRFK
KGARRIVNFAHQNHCDVIVMEKLAGLIPDAERERGINSALVNWNRGHLVKWVKQLANDAGMRVVEVYPHWTSQLCSRCGAMGAR
FSADHGHVAFEYVGSLFACADCGYVANADHNASVNLHRRFYGELPKVKKLREGVYELAKPDHAPKEVKLAELRSAVSVAAAAMC
RGQASPF (SEQ ID NO: 325)
>SRR6963352_megahit_k177_1225001_5|P
[沉积物宏基因组]
MSVKTIQAKLECDRETLETLWRTHTVFNEGLLPILALFFKMKRRETGSTEEEKDLYKRICEFILRTNASNPGYLLNSVCLENWK
PDSARKTKAEVVRDGKKEKISAEEWIDDLLNMKEESGLLFLKQELLNGLPIMFFEKLVAESKAILSAYEELQYQWEVDHNKWLK
EKEGWEKKHKEYLAVRSLFQKFEQEVKGKITKKRRRWHLYKDWMKRTPELVAWRGKETKVYGLTEEDREAIKKAPLKKKIGVEV
GLFFKRNPELSALDKLHSYYEEKFVRRGKKKRNCDGYEHRPTFTLPHSVKHPRWYVFSGPQTGNGYKNLNIPERKKDFGQLDLL
VLTDDKVDGEYPREYRTVKFRGDHRFLDIRGATRTRLVTKGKNKGQEVEEKFAYEMYDRYLKIWRPLKISGAKLAFRDIKINQD
GSLKSAVPYLWFTLNLEDLPISEKAKEIKWVETDEVKDNGDKRKKIQIPEGLITCAVNIGIRNLGFGTLAVGEKGKARILRSRN
IWIGEEEKTGKHKGRWQKGPTLRELRDHKQEIRRLRKLRGKVVKGERSHIKLQNHITNFADDRLKKAARKILNFALNVENHIDK

[0395]

KTGKPYPRADVLLLEDLSTLVPTTKKRRDQNQLLIEFMHRGTLSKIEEMAKDAGLKVLRIKPGGISQVCPKCGGIGRRYEIRID
QETKKPDIHFTDVIDDIDCQKLFVCASCSYRGNGDHVASVNLHRKFYDWSVVDKYFTLLRSDNYKEQKKVIEEELREFLRKEHF
PQMALN（SEQ ID NO: 329）
>SRR6963480_megahit_k177_1813084_1|M
[沉积物宏基因组]
MSVKTIQAKLECDRETLETLWRTHTVFNEGLLPILALFFKMKRRETGSTEEEKDLYKRICEFILRTNASNPGYLLNSVCLENWK
PDSARKTKAEVVRDGKKEKISAEEWIDDLLNMKEESGLLFLKQELLNGLPIMFFEKLVAESKAILSAYEELQYQWEVDHNKWLK
EKEGWEKKHKEYLAVRSLFQKFEQEVKGKITKKRRRWHLYKDWMKRTPELVAWRGKETKVYGLTEEDREAIKKAPLKKKIGVEV
GLFFKRNPELSALDKLHSYYEEKFVRRGKKKRNCDGYEHRPTFTLPHSVKHPRWYVFSGPQTGNGYKNLNIPERKKDFGQLDLL
VLTDDKVDGEYPREYRTVKFRGDHRFLDIRGATRTRLVTKGKNKGQEVEEKFAYEMYDRYLKIWRPLKISGAKLAFRDIKINQD
GSLKSAVPYLWFTLNLEDLPISEKAKEIKWVETDEVKDNGDKRKKIQIPEGLITCAVNIGIRNLGFGTLAVGEKGKARILRSRN
IWIGEEEKTGKHKGRWQKGPTLRELRDHKQEIRRLRKLRGKVVKGERSHIKLQNHITNFADDRLKKAARKILNFALNVENHIDK
KTGKPYPRADVLLLEDLSTLVPTTKKRRDQNQLLIEFMHRGTLSKIEEMAKDAGLKVLRIKPGGISQVCPKCGGIGRRYEIRID
QETKKPDIHFTDVIDDIDCQKLFVCASCSYRGNGDHVASVNLHRKFYDWSVVDKYFTLLRSDNYKEQKKVIEEELREFLRKEHF
PQMALN（SEQ ID NO: 329）
>SRR6963491_megahit_k177_4304123_3|P
[沉积物宏基因组]
MTIKSFEAKIVTNPSLRETLEYTHRVYNEDLKYMVEVLLAMRKGKHGDGYVEFWREHLQKAANPSGLLEPLTTPDGPKADAKSK
KSKIIRQLFKEHGMLFDRDERWPHWNARHRRHLFDDAKKWARQLDERLERWRSEHAQWLKEKEKFENENPLYMAVEPLFGGFFK
EFGQGGKRRERWAAYYHFLLAHPELANWRGQADAIVPLTPEEEAACLRDRRRGVKKRLDLFLEKNPELEALHKLDAKYGRNMVR
RSARHRNIDGFRHPPTMTLPHSRLHPEWITFQAGETYKDLDLNGGTVRLNILPPDAETSKLHSVEFRAEKRLRRITPRETPLRK
GRNEFNKNYRDAADRVHPARVKGIRLILRPRGAYLKITVELDRPESRTGLKQQTCDKYSSNWCYQEIVKRREGKPLKCLSLDLG
VRHLAVGVVVQETRANGDRETVIEGVRFFKEVGGRPFPSLDEIRRHKDRLREKRSQTRRAPKGRRTCIALQDHVTKMGQDRYRK
GAAAIVRYAVERDVDVILLEKLAGFIPDAAKQRGINKALMEWNRRKIADFVKLLAADHGISVVEVAPHHLSQLCSKCGELGRRF
SWKDGLPKFGTVEKLFVCPACGYQANSDYNAAVNLYKAFSGAYEERPKKAKKGVYKYRDRTICLKKLDGQLAERLTQEQTPF
（SEQ ID NO: 328）
>SRR6963587_megahit_k177_2272343_29|M
[沉积物宏基因组]
MATKTYRARIVQISPRLGAALERTHYLFAEALNQMVRRYIDMRKGKYGQECKKLADIMLTRSNTFAHGVMDQLTREISTSGLTD
EWVDLARAVHKSHGPLFLQHDGFAEVDGVMLHTKRHGKVSPMPGRLPVNAKFWHQVCDSASAFLKSNNELLENWRQGRKEWLQE
KADWESANADFMRFWNGSYQEFEETCEIKRKESQALEGQKVVEKKRHRRDRGKRMSRWHLWYNWVITHPEIIEWRNRAKAAEFR
PVPPTVHKKIKNKYPRQDKHITKFLDWLKENNPELKALDSRRRFYVRNYLRFRRPPTLTLPSANKHPYWFTFELNQFYKNVDFD
KGSIQLLLVDETEEGSWFFRWFDAKIQCDPRLRPCVRAGVFATKGRYPPYIGGKVGNKLNRPAERPENRKAGYAGAKLVLNKRR
KELFFTVIEQDCPRKVSWKKTKQRTCRGDNVFAKDGSQIPLKVMAIDLGIRHIGAYVIAEACRNRGVWDISWQHKGITGRHAVP
GLKEIRQHDRQLRKARSQRGKPVQGEESFIELQDHRTNMADDRFKKAANTIVELARQYDVHLILFERLGNLSPTAFDERWMNRQ
LRDMNRRQIVDTVKLQAREFGIECLEDISPWMTSHICSRCFKPGWRFSIKRKNPYGERFTRKNCRDYGYPVWDTGGHLFRCPHC
EYKCNADINAACNVAAKFFGLWSGLDGLKRDQWVYTWQENQPKTFDAKKTFLEWALDVRNRKLMGDSPF（SEQ ID NO: 326）
>SRR6963591_megahit_k177_1621237_2|P
[沉积物宏基因组]
MATKTYRARIVQISPRLGAALERTHYLFAEALNQMVRRYIDMRKGKYGQECKKLADIMLTRSNTFAHGVMDQLTREISTSGLTD
EWVDLARAVHKSHGPLFLQHDGFAEVDGVMLHTKRHGKVSPMPGRLPVNAKFWHQVCDSASAFLKSNNELLENWRQGRKEWLQE
KADWESANADFMRFWNGSYQEFEETCEIKRKESQALEGQKVVEKKRHRRDRGKRMSRWHLWYNWVITHPEIIEWRNRAKAAEFR
PVPPTVHKKIKNKYPRQDKHITKFLDWLKENNPELKALDSRRRFYVRNYLRFRRPPTLTLPSANKHPYWFTFELNQFYKNVDFD
KGSIQLLLVDETEEGSWFFRWFDAKIQCDPRLRPCVRAGVFATKGRYPPYIGGKVGNKLNRPAERPENRKAGYAGAKLVLNKRR
KELFFTVIEQDCPRKVSWKKTKQRTCRGDNVFAKDGSQIPLKVMAIDLGIRHIGAYVIAEACRNRGVWDISWQHKGITGRHAVP
GLKEIRQHDRQLRKARSQRGKPVQGEESFIELQDHRTNMADDRFKKAANTIVELARQYDVHLILFERLGNLSPTAFDERWMNRQ
LRDMNRRQIVDTVKLQAREFGIECLEDISPWMTSHICSRCFKPGWRFSIKRKNPYGERFTRKNCRDYGYPVWDTGGHLFRCPHC

EYKCNADINAACNVAAKFFGLWSGLDGLKRDQWVYTWQENQPKTFDAKKTFLEWALDVRNRKLMGDSPF (SEQ ID NO: 326)

>SRR8554444_megahit_k177_608234_1|M

[沉积物宏基因组]

MATKSFEAKIVCKPDEKYTAEQKKQFLWFTHQVFNDGVRKVIPYVFKMKRGELGPEFQAIYYAITSSQDAIGKLEAVINPDWTS
GKIGKSDPNKWKELLKYQELEKGFRQRLKEEGIKSTKKFRKELEDEKKKLAKEIGQKDIWADAAAILRNKNLLLFNRDELLPNL
PSEFRRKIYEMTIQLIHGHQELVANWEDEHAEWLIEKDKWEEEHPEYMNVRPIFEKFEKEQGKVKGSRIRWLAYLDFLSSKPEL
ANWRGKAKETIPLTKEERAGFRKPGQHFAAFFNKNPELQELDRLHKEYQEKFARTQSKRTPHPDGFKHRPTFTLPDAMRHPVWY
SFKGATDPTKGSTYRNLDLENCTLDLKVLTAMEGEGRNPGGMIQYAFEPDERIKGFRYVGTTEKGKRAKGYIYYDPILEKERPA
KIQGIKLVFRPPRPDGTAYLIFSCQIEDEKPKIKIWKDKEEESPGEITKRKKTEVYPPELITLAIDFGQRHLGAITICKNNNGR
PEPIRFIPAYPKRRKDRESKPVSAWLAKIPGLTFNAVGMHEKEISAGMSRRFQDPKSIRQAGEKEGRKSKGQHIPETETPWAHL
REHIANMKEDHYKKAANLIIRTALQNGAQVILIENLRNYRPMLERTNLENRRRMQWAVRQTAKFLEDTARPLGLIVRQVSSAYT
SRFCSSCGHPGARVSLPGQKNWEKFYAEKYGKERKMIAVAGGQFFCCPACKKIINADINASLNMHKVFYQNFIWPEKIDKKDTK
NFIWQGKNYNWDQIADDVQSFLDQKAGIKKEDDIPY (SEQ ID NO: 350)

>SRR4030068_megahit_k177_2163203_10|M

[土壤宏基因组]

MGVRAIKSKIICDKVVLTHLWRTHKVFNERLPLIISLLFKMRRGEAGETPERQDLYRRIAEFILMRSSKDAPYLFNSICIRNWK
PNSAKGYKVRLIDDGGVATETSGESWAEEAARLSAEGHLLYDKDDMVGDLPALLQQMVFRDSVSIISGHDGLVDLWKKDHAEWL
KRKADWEGREENKKYITMRPKFEEFEAAVGGKIGKRRGRWHLYHEWLRENPDLAGWRGGPPVVNPISTAGKKRIARAAPSKKMS
VEFEEFWKANPELQALDRLHGYYEREFVRRRKRKRNPDGFDHRPTFTLPHPVRHPRWLVFNAPQTNPVGYRNLILPTARSLRGS
VELQLVTGEKTEERYPHAWVPVGFKADPRLGDLRRVTAKSVYRKGKDKGKEKEVEAYEFLDRHLNRYRKAALSGAKLIFRLSPG
GSPKEAYLYFTLEVANEPLTDKAKAIKWTETGEVTKRGKARKRKTVSEGLVTCAVDLGIRNLGFATLAEYVDGVPRILRSRNLW
IGIEEEGGAHPGRWSEGPTLHHISEHKAKIRSLRRLRGKPVKGEESHLELQSHITSMGEDRFKRSARAIINFALNAKKEAGKKG
AYPRADVLLLERLEGMIPDAERGRGINRALMAWNRGQLVEQVKEVARDAGLKVFEIHPIGTSQVCSKCGSLGRRYSIGRNKEDG
EKEITFGWVEKLFACPCGYTANADHNASVNLHRRFLFDDGAIPGFLEWSKRDKSGREAMVARIEAALEPRLRAMHGLEPISV
(SEQ ID NO: 351)

[0396]

>SRR5487326_megahit_k177_254484_5|M

[土壤宏基因组]

MAVRAIQARLESDKNMLEHLWRTHKVFNERLPRIISILFKMRRGEYGTTDEDRALYKSIAKFVLGRNARDAPYLLNSISIKNWT
PSTAKKMKVTITTSDGSEQEVYGDTWANKAAELSSAGRLLYDKEEVLGDLPDSLRQIVVRECVAIISGHDELVKNWDKAHTEWL
KKKEEWEAKEEHKKYLAVRSHFDDFEKSVGGKAGKRRARWHKYLQWLRENPDLAGWRGGLAAVNELSKEAQERVRKAKPWRQRS
VEAEEFWKANPELQALDRLHGQYERDFVRRRKTKRNPDGFEHRPTFTLPHFLRHPRWLVFNAPQTSPAGYRNLKLPGKVGQYGE
VDLRLLTDEIRGAEFPVDWTRLRFLGDPRLADFKPVKIQKAATKGKVKGELRERPAYRFFDGELQMERPAQISGAKLIFKNVRL
DDGSLQSATPYLYFTCEAESLQWTENARKVKWNATGEVMKKGKKRGYRTLPDGLIACAVDLGIRNLGFATLARYENGVPRLLRS
RNLWIGHAEARGIHPGRWSEGPELAHLALHKRELRRLRSLRGDPIAGEESHIKLQQHITYFAEDRFKKAARAIVNFALNVKEFA
DKNTGEIYPRADVLVLENLGTLLPNAEREKGINRSLIEFNRGHLVDRVKELAKDVGLRVFEVSPVGTSQVCSRCGTLGRRYSIR
RSAETHRPEIQFGFVEKLFACPECGYRANADHNASVNLHRRLAQVPGAFAQWEELGKDQQSKRKRWEEIEAKLLPSLQKMHGLD
GAGKPHPFLSA (SEQ ID NO: 352)

>SRR5487761_megahit_k177_404752_1|M

[土壤宏基因组]

MAVRAIQARLESDKNMLEHLWRTHKVFNERLPRIISILFKMRRGEYGTTDEDRALYKSIAKFVLGRNARDAPYLLNSISIKNWT
PSTAKKMKVTITTSDGSEQEVYGDTWANKAAELSSAGRLLYDKEEVLGDLPDSLRQIVVRECVAIISGHDELVKNWDKAHTEWL
KKKEEWEAKEEHKKYLAVRSHFDDFEKSVGGKAGKRRARWHKYLQWLRENPDLAGWRGGLAAVNELSKEAQERVRKAKPWRQRS
VEAEEFWKANPELQALDRLHGQYERDFVRRRKTKRNPDGFEHRPTFTLPHFLRHPRWLVFNAPQTSPAGYRNLKLPGKVGQYGE
VDLRLLTDEIRGAEFPVDWTRLRFLGDPRLADFKPVKIQKAATKGKVKGELRERPAYRFFDGELQMERPAQISGAKLIFKNVRL
DDGSLQSATPYLYFTCEAESLQWTENARKVKWNATGEVMKKGKKRGYRTLPDGLIACAVDLGIRNLGFATLARYENGVPRLLRS
RNLWIGHAEARGIHPGRWSEGPELAHLALHKRELRRLRSLRGDPIAGEESHIKLQQHITYFAEDRFKKAARAIVNFALNVKEFA
DKNTGEIYPRADVLVLENLGTLLPNAEREKGINRSLIEFNRGHLVDRVKELAKDVGLRVFEVSPVGTSQVCSRCGTLGRRYSIR

RSAETHRPEIQFGFVEKLFACPECGYRANADHNASVNLHRRLAQVPGAFAQWEELGKDQQSKRKRWEEIEAKLLPSLQKMHGFD
GAGKPHPFLSA (SEQ ID NO: 353)
>SRR8554485_megahit_k177_439309_3|P
[土壤宏基因组]
MAVRAIQVRLESDKTLLEHLWRTHKVFNERLPMIISILFKMRRGECGTTDEERTLYKNIARFVLARSSKDAPYLLNSISIKDWT
PSTARKMKATITIPDGSEQEVCGDTWADKVAELSAAGRLLYSKQEVLGDLPDSLRQMIVRECVAIISGHDELVRNWDKSHTEWL
KKREEWEAKEEHKKYLAVRSRFDEFEKSVGGKAGKRRARWHKYLQWLRESPDLAGWRGSLAVVNELSREAQERVRKAKPWRQRG
VEAEEFWKANPELQALDRLHGQYERDFVRRGKTKRNPDGFEHRPTFTLPHFLRHPRWLVFNAPQTSPPGYRNLKLPGKAGQYGE
VELRLLTNEIRGAEFPIDWTTLRFLGDPRLADFKPVKIQKAATKGKVKGELRGRPAYRFFDRELQMERPAQISGAKLIFKNVRL
DDGPLQSATPYLYFACEAKSLRWTENARKVKWNATGEVTKKGKKRGYRTLPDGLIACAVDLGIRNLGFATLARYENGVPRLLRS
RNLWIGHAETRGRHPGRWCEGPELAHLALHKRELRRLRSLRGDPIAGEESHIELQQHITHFAEDRFKKAARAIVNFALNVKEFA
DKNTGEIYPRADVLVLENLGSLLPDAEREKGINRSLIEFNRGHLVDRVKEVAKDVGLRVFEVSPVGTSQVCSRCGTLGRRYSIR
RSAETHRPEIQFGFVEKLFACPECGYRANADHNASVNLHRRLAQVPGAFAQWEELGKDQQRKRKRWEEIEAKLLPSLQKMHGFD
GAGKPHPFLSA (SEQ ID NO: 354)
>SRR8554511_megahit_k177_4512460_1|M
[土壤宏基因组]
MCESPQQRAFLELTHRLFNESLRPLVPLLYAARKGRHGPDYQQILSTIKSGQGASAQVEAVSSLAARPGSRGGKEWTDLAKELV
ARGQILFDRHALMPNVHKEFRRKVFEMAFQIILSHDAKMKQWRTDHAKWLREKAEWERSYPEYLAVRPVIQEFEAREGQAGKRR
GRWHRWLDFLVLYPQLAAWRGGEPTVRQLTQQELQEARRSRRRAVPNTIDAFFAKNPELKELHGLHGEYQARFVRPWAKRRHQD
GFFLPPTFTLPSADRHPAWYSFKRGDTYRNLSLEQRTIELRVLKSADPNESGRHAYVQYQFVGDNRLGKLSPASKIVRSGRFPC
DLLVWDEGSDHARPARVQGAKLVIRKGQAYLYFSVYIEDAPSRLPVKQAQIDKYSPGWSINRIRGEPYDPPLRTMAIDLGIRHI
AAATVMENGNVLATRFLHSRPVSRVSGHVIEGIPTLPQISEMKQRLRRARRKFGKPLKGTMSCRRIQRHVTSMSEDRCKKSASA
IVDFGRAHRVDIILMERLEGLVPDAARERRINCGLINWNRGHLASWIRIIAEPYGIRVLELPPRWTSQVCSHCGSIGARFTCTR
GEMKISAGGKLFGCPTCAFEANADFNASLNLQRVFWGLFPQVSKAGRGRVQCNDRVIELASVNQAWENHWLRKHNSANDVF
(SEQ ID NO: 355)
>SRR8554511_megahit_k177_1767604_11|P
[土壤宏基因组]
MAAKSYQPKMSTKSYPSKVICGKSEFDDLRRTHRVFNAGVRKLVDTYIAMKKGAYGQTAVEIAMYLLGKKGSKGHDLMNPLTAE
DAKDTLDTEGMKLVRQYRRESGILFSRYEKIYEIDGKAVCAGSAKKGDSKGKLVISSIFWHFICNQAIEYLRSNNELMQAWRDD
RRKWLEEKQKFFDENAGFREFVQGYLADFDAHAEKMRQEFQTQAGQRAYLRKRNKSDSGKRFSRWHLWYEWICSHPEILEWRGR
AKAADFVLLTPDEVHEVEKQFPKRQDRRTGKMLDMLRRKNPELDEIEKLRKTYTNDFARFRRPPTLTLPSADKHPCWCRLEKDK
LYRKADFEKGTIELCVVCKDETDGEYDMKWRTFRLKPDPRLQPSHRAKAFAAEGRYPPYKEGKVGNTLNRPAESPDKRLAGLKG
AKLILKRNDNAQIVFTVVEQNEPLKAKFSKKEGAKSSADRLEILDGKEATRVMAVDLGARHVGAYVITEGKRSSDVWRLEHLKK
AFVDSPDLPDLRSITQHDWQLRKARRKRGKPVKNQRSFISLQDHRTKMSDDRFKKAANVIVENARRHSVHIIVFEDLRDLKPTA
YNERWLNRQVRGMNHRHIFDFTKKMAAEFGINVTDVRAYNTSRLCSKCGFPGYRFSMKQKSPYREKAPRKGCSDFGYPVWDSGG
HLFRCPHCGYRVQADLNAAANIAQRFFGSLMEWKCKDYRYSNAMHPSFDARKDFDKWASDVQRRKSMPDAPF (SEQ ID NO:
356)
>SRR8554643_megahit_k177_4328113_1|P
[土壤宏基因组]
MCESPQQRAFLELTHRLFNESLRPLVPLLYAARKGRHGPDYQQILSTIKSGQGASAQVEAVSSLAARPGSRGGKEWTDLAKELV
ARGQILFDRHALMPNVHKEFRRKVFEMAFQIILSHDAKMKQWRTDHAKWLREKAEWERSYPEYLAVRPVIQEFEAREGQAGKRR
GRWHRWLDFLVLYPQLAAWRGGEPTVRQLTQQELQEARRSRRRAVPNTIDAFFAKNPELKELHGLHGEYQARFVRPWAKRRHQD
GFFLPPTFTLPSADRHPAWYSFKRGDTYRNLSLEQRTIELRVLKSADPNESGRHAYVQYQFVGDNRLGKLSPASKIVRSGRFPC
DLLVWDEGSDHARPARVQGAKLVIRKGQAYLYFSVYIEDAPSRLPVKQAQIDKYSPGWSINRIRGEPYDPPLRTMAIDLGIRHI
AAATVMENGNVLATRFLHSRPVSRVSGHVIEGIPTLPQISEMKQRLRRARRKFGKPLKGTMSCRRIQRHVTSMSEDRCKKSASA
IVDFGRAHRVDIILMERLEGLVPDAARERRINCGLINWNRGHLASWIRIIAEPYGIRVLELPPRWTSQVCSHCGSIGARFTCTR
GEMKISAGGKLFGCPTCAFEANADFNASLNLQRVFWGLFPQVSKAGRGRVQCNDRVIELASVNQAWENHWLRKHNSANDVF

[0397]

(SEQ ID NO: 355)
>SRR8554643_megahit_k177_1281129_3|P
[土壤宏基因组]
MATKTYYARFVNTTPRLRAALERSHYVYVECLRQMIERYIAMRKGKLGEECRQLAKIILTTGNTEAHGIMNQITRPVVTSGKDY
EWVKLAKQVHKLRGPLLLQDEGFARVDGVTIRTKHHGKVEPTRGELAVTAKFWMQICDMASSFLKSNAELMDDWREERTKWLRE
KAEWEERNSDFLKFWNGPYTEFENLYERMRLEAQTAAGQTPTVRKRQSRERGKRIDRWHLWYEWVISHPEIMAWRNQARASDFK
VMPIELQNQIKRKHPRQNKYIPEFLKWLRENNAEFDRLDSLRRVYVNNYCRFKRPPTLTLPSPKKHPYWFTFERDEFYKDADFE
TGAIRLLLIDQKDDGAWFFEWFDAQMKCDPRLKPSVRAEHFNKEGRYPPYIDGKPARKLNRSAASPEQRRAGYTGAKLVLKGAR
QELLFTVIEQDCPPKVKWQKVKMRACRADNAFSAEGERIPLKIMAIDLGIRHTAAYVIAEGTLQDGQWHLSWQKKGIVRDIHIS
PLMNIRQHDWELREGRSRQGKAAKGERTFVDLQDHRTDMADDRFKKAANTIVETAREHDVRLILFEQLKNWFPKAFDERWMNRQ
LRDMNRRKIVEMVEGQSKEFGIICADNISPWLTSRICSRCYKPGWRFSIKGKDPYKEKLSRRHCQDFGYPIWDRGGGHLFRCSH
CGYRANADINAAGNVAAKFFGAWPVLSCKKGVYTWQQDGQKHTFDARETFESWAQDVRRRKQIAESPF (SEQ ID NO: 357)
>SRR8554643_megahit_k177_1281129_3|M
[土壤宏基因组]
MIERYIAMRKGKLGEECRQLAKIILTTGNTEAHGIMNQITRPVVTSGKDYEWVKLAKQVHKLRGPLLLQDEGFARVDGVTIRTK
HHGKVEPTRGELAVTAKFWMQICDMASSFLKSNAELMDDWREERTKWLREKAEWEERNSDFLKFWNGPYTEFENLYERMRLEAQ
TAAGQTPTVRKRQSRERGKRIDRWHLWYEWVISHPEIMAWRNQARASDFKVMPIELQNQIKRKHPRQNKYIPEFLKWLRENNAE
FDRLDSLRRVYVNNYCRFKRPPTLTLPSPKKHPYWFTFERDEFYKDADFETGAIRLLLIDQKDDGAWFFEWFDAQMKCDPRLKP
SVRAEHFNKEGRYPPYIDGKPARKLNRSAASPEQRRAGYTGAKLVLKGARQELLFTVIEQDCPPKVKWQKVKMRACRADNAFSA
EGERIPLKIMAIDLGIRHTAAYVIAEGTLQDGQWHLSWQKKGIVRDIHISPLMNIRQHDWELREGRSRQGKAAKGERTFVDLQD
HRTDMADDRFKKAANTIVETAREHDVRLILFEQLKNWFPKAFDERWMNRQLRDMNRRKIVEMVEGQSKEFGIICADNISPWLTS
RICSRCYKPGWRFSIKGKDPYKEKLSRRHCQDFGYPIWDRGGGHLFRCSHCGYRANADINAAGNVAAKFFGAWPVLSCKKGVYT
WQQDGQKHTFDARETFESWAQDVRRRKQIAESPF (SEQ ID NO: 358)

[0398]

>SRR6056591_megahit_k177_2733919_1|P
[地下宏基因组]
MATKSYRSRIVCTKAELDDLVRTHRVFNAGVRAMVDLYISMKKGTYGETAKEIARYLLDHSGTKGHDLMNPLTDRGGKDTRDDE
GMKLVRRHRQEHGILFPRHERICEIDGRKVRAGSSAKKEKSRDTLTISSIFWHFICNQAIEYLRSNNELMAAWREERGKWLRDK
EKFFAENPGFREFVEGYLAAFDAHAEKARQEFQTKAGQRAHARKRKKSGSGKRFSRWHLWCEWICDHPEIIEWRGKAKGEGFAF
LTAEEVRAIEEEFPKREDRRAGRMLDMLTRKNPELAEVEKLRKKYNNDYARFRRPPTLTLPSAKKHPCWYRLEKGKLYRNEDFE
KGTIEILVICHDEKEDAYYMDWRPFSIKPEPRLVPSHRARAFAAEGRYPPYKEGKVGNTLNRPAESAEERLAGLKGAKLILKPK
GNAQLVFTVVEQNEPLKAKFRKKAGATSSADRLDILDGNKATRVMGIDLGARHVGAYVITEGKKKGDGWKLAPLRKGFIDSPDL
PRLRSITRHENQLKKGRRKRGKPVKNERSFISLQGHRTKMSTDRFKKAANLIVEKAREHDVHIIVFEDLRRLKPTAYDERWLNR
QVRSMNHRHIFDLAKEMAPEFGIQVTDINPYNTSRLCSKCGFPGYRFSMKRKNPYREQAPRRECADFGYPVWDSGGHLFRCPHC
GYRVQADLNAAGNIARKFFGHFMDWECKEYTYRNAVQPSFDARKDFDQWASGVLKRRTMPDAPF (SEQ ID NO: 324)
>SRR6263259_megahit_k177_232832_6|M
[陆生宏基因组]
MAVKTLQSRLHCKPKKLEALWRTHEIFNERLKDVLSLLFRMRRGECGRSPDERELYREIALFITACPANNAPYLLNSVCIQNWV
PTTARKIKAIVSRSSGELIKGTEESWTTRAAQLSAKGHLLFDKREVLGDLTPAMAQVIVREAAAYLSNYESHVRRWEREHEQWL
GEKAAWEAEPEHQLYLASRPRFTAFEAGIGGRLTERRHRWAQYLTWLSQNPELAAWRGGLATINPIGEEGRKRIAHAPRWRKAQ
VEAEEFFKLNPELKDLDALHGYYERKFVRRRKTKRHIDGFDHKPTFTQPDALVHPHWYLFNGPQTKPAGYRKLILPKSSKQAGS
IELSLMTCDKNDPNDTAWVPFRFHGDCRMSNFRSVVVTSKIKSGGKRGQTKEKEGYEMYDPALKTWRPAEIKGARLIFTVKGKR
PRAAYLYFTIKTKNTPFSEAARNIEWSETGETSVFGPPRKTKLVPEGLVTCAAHIGARDCVYATLAVGDGSTPHIIRQRNLWIE
HIEATGQHVGRRHRGPSLAHLASHKMELSRRRSQMGRIPRGETAHEGLQSHITNMAKDRYRQQARRIFDFALNVEGVGDPRAGK
SYPRADVLILESMENLIPNVERERGVNKALVAFNRAHLVSHLKEIVAEAGVMVFEISPVGTSQVCSRCGALGRPYSVRRRSKDH
PVAITFGPAEPLFACPSCHYRANSSHNASVNLHHRFYSDQALESFITYLRQPVERKLENLRRIEAELISPIAGPLSLYEMHGIK
IG (SEQ ID NO: 359)
>3300031539|Ga0307380_10018846_3|P

[陆生-土壤]
MATKTYYARFVNTTPRLRAALERSHYVYVECLRQMIERYIAMRKGKLGEECRQLAKIILTTGNTEAHGIMNQITRPVVTSGKDY
EWVKLAKQVHKLRGPLLLQDEGFARVDGVTIRTKHHGKVEPTRGELAVTAKFWMQICDMASSFLKSNAELMDDWREERTKWLRE
KAEWEERNSDFLKFWNGPYTEFENLYERMRLEAQTAAGQTPTVRKRQSRERGKRIDRWHLWYEWVISHPEIMAWRNQARASDFK
VMPIELQNQIKRKHPRQNKYIPEFLKWLRENNAEFDRLDSLRRVYVNNYCRFKRPPTLTLPSPKKHPYWFTFERDEFYKDADFE
TGAIRLLLIDQKDDGAWFFEWFDAQMKCDPRLKPSVRAEHFNKEGRYPPYIDGKPARKLNRSAASPEQRRAGYTGAKLVLKGAR
QELLFTVIEQDCPPKVKWQKVKMRACRADNAFSAEGERIPLKIMAIDLGIRHTAAYVIAEGTLQDGQWHLSWQKKGIVRDIHIS
PLMNIRQHDWELREGRSRQGKAAKGERTFVDLQDHRTDMADDRFKKAANTIVETAREHDVRLILFEQLKNWFPKAFDERWMNRQ
LRDMNRRKIVEMVEGQSKEFGIICADNISPWLTSRICSRCYKPGWRFSIKGKDPYKEKLSRRHCQDFGYPIWDRGGGHLFRCSH
CGYRANADINAAGNVAAKFFGAWPVLSCKKGVYTWQQDGQKHTFDARETFESWAQDVRRRKQIAESPF (SEQ ID NO: 357)
>3300031539|Ga0307380_10018846_4|M
[陆生-土壤]
MIERYIAMRKGKLGEECRQLAKIILTTGNTEAHGIMNQITRPVVTSGKDYEWVKLAKQVHKLRGPLLLQDEGFARVDGVTIRTK
HHGKVEPTRGELAVTAKFWMQICDMASSFLKSNAELMDDWREERTKWLREKAEWEERNSDFLKFWNGPYTEFENLYERMRLEAQ
TAAGQTPTVRKRQSRERGKRIDRWHLWYEWVISHPEIMAWRNQARASDFKVMPIELQNQIKRKHPRQNKYIPEFLKWLRENNAE
FDRLDSLRRVYVNNYCRFKRPPTLTLPSPKKHPYWFTFERDEFYKDADFETGAIRLLLIDQKDDGAWFFEWFDAQMKCDPRLKP
SVRAEHFNKEGRYPPYIDGKPARKLNRSAASPEQRRAGYTGAKLVLKGARQELLFTVIEQDCPPKVKWQKVKMRACRADNAFSA
EGERIPLKIMAIDLGIRHTAAYVIAEGTLQDGQWHLSWQKKGIVRDIHISPLMNIRQHDWELREGRSRQGKAAKGERTFVDLQD
HRTDMADDRFKKAANTIVETAREHDVRLILFEQLKNWFPKAFDERWMNRQLRDMNRRKIVEMVEGQSKEFGIICADNISPWLTS
RICSRCYKPGWRFSIKGKDPYKEKLSRRHCQDFGYPIWDRGGGHLFRCSHCGYRANADINAAGNVAAKFFGAWPVLSCKKGVYT
WQQDGQKHTFDARETFESWAQDVRRRKQIAESPF (SEQ ID NO: 358)
>3300031539|Ga0307380_10072822_2|M
[陆生-土壤]

[0399]

MCESPQQRAFLELTHRLFNESLRPLVPLLYAARKGRHGPDYQQILSTIKSGQGASAQVEAVSSLAARPGSRGGKEWTDLAKELV
ARGQILFDRHALMPNVHKEFRRKVFEMAFQIILSHDAKMKQWRTDHAKWLREKAEWERSYPEYLAVRPVIQEFEAREGQAGKRR
GRWHRWLDFLVLYPQLAAWRGGEPTVRQLTQQELQEARRSRRRAVPNTIDAFFAKNPELKELHGLHGEYQARFVRPWAKRRHQD
GFFLPPTFTLPSADRHPAWYSFKRGDTYRNLSLEQRTIELRVLKSADPNESGRHAYVQYQFVGDNRLGKLSPASKIVRSGRFPC
DLLVWDEGSDHARPARVQGAKLVIRKGQAYLYFSVYIEDAPSRLPVKQAQIDKYSPGWSINRIRGEPYDPPLRTMAIDLGIRHI
AAATVMENGNVLATRFLHSRPVSRVSGHVIEGIPTLPQISEMKQRLRRARRKFGKPLKGTMSCRRIQRHVTSMSEDRCKKSASA
IVDFGRAHRVDIILMERLEGLVPDAARERRINCGLINWNRGHLASWIRIIAEPYGIRVLELPPRWTSQVCSHCGSIGARFTCTR
GEMKISAGGKLFGCPTCAFEANADFNASLNLQRVFWGLFPQVSKAGRGRVQCNDRVIELASVNQAWENHWLRKHNSANDVF
(SEQ ID NO: 355)
>3300031539|Ga0307380_10072822_2|P
[陆生-土壤]
MIQGTTSCASPTIAVKCFQAKLVCESPQQRAFLELTHRLFNESLRPLVPLLYAARKGRHGPDYQQILSTIKSGQGASAQVEAVS
SLAARPGSRGGKEWTDLAKELVARGQILFDRHALMPNVHKEFRRKVFEMAFQIILSHDAKMKQWRTDHAKWLREKAEWERSYPE
YLAVRPVIQEFEAREGQAGKRRGRWHRWLDFLVLYPQLAAWRGGEPTVRQLTQQELQEARRSRRRAVPNTIDAFFAKNPELKEL
HGLHGEYQARFVRPWAKRRHQDGFFLPPTFTLPSADRHPAWYSFKRGDTYRNLSLEQRTIELRVLKSADPNESGRHAYVQYQFV
GDNRLGKLSPASKIVRSGRFPCDLLVWDEGSDHARPARVQGAKLVIRKGQAYLYFSVYIEDAPSRLPVKQAQIDKYSPGWSINR
IRGEPYDPPLRTMAIDLGIRHIAAATVMENGNVLATRFLHSRPVSRVSGHVIEGIPTLPQISEMKQRLRRARRKFGKPLKGTMS
CRRIQRHVTSMSEDRCKKSASAIVDFGRAHRVDIILMERLEGLVPDAARERRINCGLINWNRGHLASWIRIIAEPYGIRVLELP
PRWTSQVCSHCGSIGARFTCTRGEMKISAGGKLFGCPTCAFEANADFNASLNLQRVFWGLFPQVSKAGRGRVQCNDRVIELASV
NQAWENHWLRKHNSANDVF (SEQ ID NO: 360)
>3300031566|Ga0307378_10072489_1|M
[陆生-土壤]
MCESPQQRAFLELTHRLFNESLRPLVPLLYAARKGRHGPDYQQILSTIKSGQGASAQVEAVSSLAARPGSRGGKEWTDLAKELV
ARGQILFDRHALMPNVHKEFRRKVFEMAFQIILSHDAKMKQWRTDHAKWLREKAEWERSYPEYLAVRPVIQEFEAREGQAGKRR

GRWHRWLDFLVLYPQLAAWRGGEPTVRQLTQQELQEARRSRRRAVPNTIDAFFAKNPELKELHGLHGEYQARFVRPWAKRRHQD
GFFLPPTFTLPSADRHPAWYSFKRGDTYRNLSLEQRTIELRVLKSADPNESGRHAYVQYQFVGDNRLGKLSPASKIVRSGRFPC
DLLVWDEGSDHARPARVQGAKLVIRKGQAYLYFSVYIEDAPSRLPVKQAQIDKYSPGWSINRIRGEPYDPPLRTMAIDLGIRHI
AAATVMENGNVLATRFLHSRPVSRVSGHVIEGIPTLPQISEMKQRLRRARRKFGKPLKGTMSCRRIQRHVTSMSEDRCKKSASA
IVDFGRAHRVDIILMERLEGLVPDAARERRINCGLINWNRGHLASWIRIIAEPYGIRVLELPPRWTSQVCSHCGSIGARFTCTR
GEMKISAGGKLFGCPTCAFEANADFNASLNLQRVFWGLFPQVSKAGRGRVQCNDRVIELASVNQAWENHWLRKHNSANDVF
（SEQ ID NO: 355）

>3300031566|Ga0307378_10072489_2|P

[陆生-土壤]

MIQGTTSCASPTIAVKCFQAKLVCESPQQRAFLELTHRLFNESLRPLVPLLYAARKGRHGPDYQQILSTIKSGQGASAQVEAVS
SLAARPGSRGGKEWTDLAKELVARGQILFDRHALMPNVHKEFRRKVFEMAFQIILSHDAKMKQWRTDHAKWLREKAEWERSYPE
YLAVRPVIQEFEAREGQAGKRRGRWHRWLDFLVLYPQLAAWRGGEPTVRQLTQQELQEARRSRRRAVPNTIDAFFAKNPELKEL
HGLHGEYQARFVRPWAKRRHQDGFFLPPTFTLPSADRHPAWYSFKRGDTYRNLSLEQRTIELRVLKSADPNESGRHAYVQYQFV
GDNRLGKLSPASKIVRSGRFPCDLLVWDEGSDHARPARVQGAKLVIRKGQAYLYFSVYIEDAPSRLPVKQAQIDKYSPGWSINR
IRGEPYDPPLRTMAIDLGIRHIAAATVMENGNVLATRFLHSRPVSRVSGHVIEGIPTLPQISEMKQRLRRARRKFGKPLKGTMS
CRRIQRHVTSMSEDRCKKSASAIVDFGRAHRVDIILMERLEGLVPDAARERRINCGLINWNRGHLASWIRIIAEPYGIRVLELP
PRWTSQVCSHCGSIGARFTCTRGEMKISAGGKLFGCPTCAFEANADFNASLNLQRVFWGLFPQVSKAGRGRVQCNDRVIELASV
NQAWENHWLRKHNSANDVF（SEQ ID NO: 360）

>3300031670|Ga0307374_10070813_2|P

[陆生-土壤]

MAVRAIQVRLESDKTLLEHLWRTHKVFNERLPMIISILFKMRRGECGTTDEERTLYKNIARFVLARSSKDAPYLLNSISIKDWT
PSTARKMKATITIPDGSEQEVCGDTWADKVAELSAAGRLLYSKQEVLGDLPDSLRQMIVRECVAIISGHDELVRNWDKSHTEWL
KKREEWEAKEEHKKYLAVRSRFDEFEKSVGGKAGKRRARWHKYLQWLRESPDLAGWRGSLAVVNELSREAQERVRKAKPWRQRG
VEAEEFWKANPELQALDRLHGQYERDFVRRGKTKRNPDGFEHRPTFTLPHFLRHPRWLVFNAPQTSPPGYRNLKLPGKAGQYGE
VELRLLTNEIRGAEFPIDWTTLRFLGDPRLADFKPVKIQKAATKGKVKGELRGRPAYRFFDRELQMERPAQISGAKLIFKNVRL
DDGPLQSATPYLYFACEAKSLRWTENARKVKWNATGEVTKKGKKRGYRTLPDGLIACAVDLGIRNLGFATLARYENGVPRLLRS
RNLWIGHAETRGRHPGRWCEGPELAHLALHKRELRRLRSLRGDPIAGEESHIELQQHITHFAEDRFKKAARAIVNFALNVKEFA
DKNTGEIYPRADVLVLENLGSLLPDAEREKGINRSLIEFNRGHLVDRVKEVAKDVGLRVFEVSPVGTSQVCSRCGTLGRRYSIR
RSAETHRPEIQFGFVEKLFACPECGYRANADHNASVNLHRRLAQVPGAFAQWEELGKDQQRKRKRWEEIEAKLLPSLQKMHGFD
GAGKPHPFLSA（SEQ ID NO: 354）

>3300031965|Ga0326597_10006968_2|P

[陆生-土壤]

MGVKAWKVKLFVDAPEKQAVLDDTLQQWNLYAAHVCKVLLAFRRNDYGDVGLEVWRQLAKTPQNYHVLYCITPRSDAAPDQGPR
VTREKDYHRAARAFVQRHGYEPELAAVRRDAGPLWKQIVKAARERVLSYEQKVADWREAHEQWIVQRREWEGDNPPYMAVRPVI
EAFQRDHGATRGRRVRWRKWQAFLQSTPELAAWRGGAACVEPLSAQELRHAARNRRKAARRERELFLDKNPELAELDRVHFQYE
RLFARTAAKRRNPDGFRHRPTFTLPSLPNRPDWPRFQAMEGYKNLDITSRSVQLRLPTSDTRHWLSLSFVPDPRLARLVPLERS
VKEGRTRFDYRFPNRPGEFVLAQPQEIRLIRRGSSYFATIRLTLQPTPSPLPLTQTAAGKYSPQWVIRKIRDDAPDLSLVTCAV
DLAVRHLGAASVAREGRVFARRILHNRFFRPNDLRERALNIPSLPEIAAVRRRIRRAQRRSGRMSPAKDTCRRLWTHYRNLCED
HYKKAVAAIFAYALHHGAHLVIFEDLKMLNPDSANERGVNNALQNWNRGNIVRFAKMAAEDHGLRVVTVGAWWTSRLCARCGGM
GVRFNQGKRRPWGQRRHTAPRPVPIDRLTCNIAPLGHWFLCPGCRSCLHADINASENLHRQFLGSLPTVKRLGSSRPPIYEVNG
QRVEQARLEAAANELLLARTAPATPF（SEQ ID NO: 361）

>3300017540|Ga0180217_1000254_36|M

[陆生-土壤-堆肥]

MGTKAWKAKLILDSEDARAALLHTFSVWNAACAEVCEALLAFRRGEFGEEGRRVWTYLRDAGTQGYYILWCITKAPRPREVKEL
HRIAADFVARHGFHQGLLAVRERLPSALWKQVIHAARERVASYEEKVAIWNEDHAQWLADRAEWEAEHASYMAVRPLIEQFQSQ
FGATRGRRARWHAWLTFLGSTPALQSWRGGPPGVVSLTSGERQAARRNPRKAVAREAELFYAKNPELAELNRIHGEYERRFARS
WAKRKNPDGFRHRPTFTLPNEKSHPDWPRLQANEGWKALDPAARTIRLQLIAPDGTARLYPFRFVPDPRLSRLTPLAEPLKEGR

[0400]

[0401]

ITYAYSVPGPQGRPTPAQPQGIKVILRGDSAFISISLSLLPPPSAIPVTQAAMTKYPPTWALKQIREKAPELELVTCAVDLGVR
HIGAATIAREGEVIARRILHHRFFPPGGDPARPINIPSLAEIAEVKRQLRRARRRSGKPAPGKESCRRLQAHYRGLCEDRYKKV
VAAIFEFAAAHGAHIVIFEDLRMLNPDSANERGVNAALMNWNRGSIVRFARATAEDRGLRVATVGAWWTSRICCRCDALGVRFE
QGRRREWARAPVASPAPPDRTACTISRLGQWFFCPACRRRVHADINASENLHRVFLGTFPKVERVGREPPAYEVGGQHVDFREI
QRAAAAILSAPATPF (SEQ ID NO: 362)
>3300014494|Ga0182017_10000840_9|P
[陆生-土壤-沼地]
MGVRALRARIECDRETLLHLWRTHCAFNQRLLSILTKLFGMRRGELGETAARRALYQRVARFLLARDSKDAPYLLNSVSIQGWK
PATALKMKVRMPAPDGTLAEVSADTWAPEAAKLSAHGTLAYDKEAVRDGLPDSLFQPMLRDAVAYISGHIELARNWEAEHADWL
KAKAEWEEKDEHKKYLALRPRFDAFEQSVGGKAGKRRERWHLYLTWLRENPALAAWRGGPPAVHPLDAKATARVAKAKPWKKRS
VEADEFWKANPELRELDKLHGYYEREFVRRRKTKKNPDGFDHRPTFTLPHAVLHPRWTLFNAPQTSPQGYADLRLPQRPGTLGS
VKLRLLTGEKAPDGCPNEWVGVQFRADPRLSQFAERAEAQEIRKGKKKGTQKTVRVFRFVDQQLNCERPAQISGAKLIFKNVKL
NDDGSLKSATPYLVFACTIDNLPLTEHARAIQWSETGEVTKTGKKRKAKKLPTGLVTCAVDLGLRNLAFLTLATSGEGPQNNTH
DGIRVLRSRNQWLALEETAGTHPGRWSAGPDLAHIAAHKREIRRLRRLRGKPVKGQQSHVELQGHIDHMGQDRFKKAARAIVNF
ALNVDRAAGQSGQPYPRADLLIVEKLEGFIPDAARERGINRALVAWNRGQLVTRIKEVAADAGLKLFPVHPAGTSQCCSRCGAL
GRRYSIVRDQQTGVPAIRFGWVEKLFACPNRECPGRAPDKPDQPFTCNADHNASLNLHRRFVLDDRAVAAFFAWKNKTNAERER
ELTQMDQALRPELEKLHRLPLAELATPF (SEQ ID NO: 343)
>3300014839|Ga0182027_10004505_3|M
[陆生-土壤-沼地]
MFNRRLPDLISLLFRMRRGECGETEQASRLYSQVASFILARGSKDAPYLLNSVSVRGWKPHTAKKMKARALGPDGQQIELAGED
WADPAAALSAEGRLLYDKTELLGDVPDTLRQMAARECVAIISGHEELVRAWEKAHEEWLASKAKWESDPENQFYLGLRARFDEF
EQSVGGKVGKRRGRWQRYLDWLRANPDLAAWRGGPAAVVELSAGAKERIRKAKPRKVRSVEAEEFWKVNSELAALDRLHGYYER
EFVRRRKTKKHPDGFDHRPTFTLPHPVRHPRWFVFNAPQTSPQGYRKLCLPSSGGHTGTVELLLLTGEKADNKYPSGWETVRFA
ADPRLSDFKAITKRRQVLKGKTKGQETDTKAFTFLDRQINLERPAQISGVKLIFKGVRQNHDGSLLSAHPYLVFTCNVADVPLT
DGAKKIKWEDTGERRKRVVVPAGLIACAVDLGIRHVGFATVATQDAESGGLHVLRSRNIWIGQEEEKGRNPGRWSPGPDLAHLS
AHKRSLRHLRGLRGKPVKGENSHIELQDHITNMASDRFKKAAREIVNFALNADSKVDPRTGEPYPRADVLVLENLANLLPDSER
ERGINRALIEFNRGHLADRIVEVAKDCGLKVMMVSPVGTSQVCCRCGALGRRYSLRRHQDGGYTDIHFGFVEPLFACPSCGYRA
NSDHNASVNLHRRFRLGEAAIAKFKEWQKGTKPERDEVLRRIEQSLVGPLREMHRLAVADTPF (SEQ ID NO: 346)
>3300014839|Ga0182027_10004505_2|P
[陆生-土壤-沼地]
MSTRALQSRIQCDHQTLEHLWRTHLVFNRRLPDLISLLFRMRRGECGETEQASRLYSQVASFILARGSKDAPYLLNSVSVRGWK
PHTAKKMKARALGPDGQQIELAGEDWADPAAALSAEGRLLYDKTELLGDVPDTLRQMAARECVAIISGHEELVRAWEKAHEEWL
ASKAKWESDPENQFYLGLRARFDEFEQSVGGKVGKRRGRWQRYLDWLRANPDLAAWRGGPAAVVELSAGAKERIRKAKPRKVRS
VEAEEFWKVNSELAALDRLHGYYEREFVRRRKTKKHPDGFDHRPTFTLPHPVRHPRWFVFNAPQTSPQGYRKLCLPSSGGHTGT
VELLLLTGEKADNKYPSGWETVRFAADPRLSDFKAITKRRQVLKGKTKGQETDTKAFTFLDRQINLERPAQISGVKLIFKGVRQ
NHDGSLLSAHPYLVFTCNVADVPLTDGAKKIKWEDTGERRKRVVVPAGLIACAVDLGIRHVGFATVATQDAESGGLHVLRSRNI
WIGQEEEKGRNPGRWSPGPDLAHLSAHKRSLRHLRGLRGKPVKGENSHIELQDHITNMASDRFKKAAREIVNFALNADSKVDPR
TGEPYPRADVLVLENLANLLPDSERERGINRALIEFNRGHLADRIVEVAKDCGLKVMMVSPVGTSQVCCRCGALGRRYSLRRHQ
DGGYTDIHFGFVEPLFACPSCGYRANSDHNASVNLHRRFRLGEAAIAKFKEWQKGTKPERDEVLRRIEQSLVGPLREMHRLAVA
DTPF (SEQ ID NO: 347)
>3300031715|Ga0307476_10010105_2|P
[陆生-土壤-硬木森林土壤]
MSVRAIQARINCDRQTLLDLWRTHCVFNQRLPSILSVLFKMRRGECGDSPELQDLYQRISRFIIASSSQLADALLNSVSIRNWK
PDSARKRKIKVVSETGEIVEVTGETWADAAAAISKQGQLLYDKRALLFDLPNSMRQIVSREAVTIISGHDELTAIWHKEHAAWS
KQKAEWQNGAENQRYLALRPKFAAFEEEVGGAITKRRGRWHLYLQWLPEHPELAAWRGGPAIVRKLDHAALLRVQRSRPSKQRT
VEAEEFWKINPELKVLDRLHGDYESRFVRRRKTKKNSDGFDHRPTFTLPHPVNHPRWFVFNAPQTSPQGYGDLKLPAEPNGFGT
ISLRLLTGEKSEGGYPSNWRTVRFKADPRLADFRQVSVKYRTNKGASKGTEREKRGYIFVDRQLKLERKATISGAKLVFRDIRL

NPDGSLSSATPYLIFTCNIDDELMTERAKAIKWSDTGEITKKGKPRKKKTLPSSLVSCAIDLGHRHLGFATIAQAEETGLRILR
SRNLWIGHKETTGHHPGRWCEGPDLAHLGQHKRKLRALRRLRGKPVKDERTHVELQGHIDDMAEDRFKKGARAIINFALNTDNR
IDEKTGKPIPRADVLLLENLENLIPDAERERGINRMLVQFNRGNLVKRVKEMAADAGLKVFEVSPFGTSQVCSKCGELGRRYSI
RRQQVPEDPTQTRPDIHFGLVEKLFACPKCNYRANADHNASVNLHRRFHNDGLALKSFAEFRALSDQDRLRSIEQIDAKLLEPL
RELHNVGPVTPF (SEQ ID NO: 363)

>3300031715|Ga0307476_10010105_1|M

[陆生-土壤-硬木森林土壤]

[0402]

MADALLNSVSIRNWKPDSARKRKIKVVSETGEIVEVTGETWADAAAAISKQGQLLYDKRALLFDLPNSMRQIVSREAVTIISGH
DELTAIWHKEHAAWSKQKAEWQNGAENQRYLALRPKFAAFEEEVGGAITKRRGRWHLYLQWLPEHPELAAWRGGPAIVRKLDHA
ALLRVQRSRPSKQRTVEAEEFWKINPELKVLDRLHGDYESRFVRRRKTKKNSDGFDHRPTFTLPHPVNHPRWFVFNAPQTSPQG
YGDLKLPAEPNGFGTISLRLLTGEKSEGGYPSNWRTVRFKADPRLADFRQVSVKYRTNKGASKGTEREKRGYIFVDRQLKLERK
ATISGAKLVFRDIRLNPDGSLSSATPYLIFTCNIDDELMTERAKAIKWSDTGEITKKGKPRKKKTLPSSLVSCAIDLGHRHLGF
ATIAQAEETGLRILRSRNLWIGHKETTGHHPGRWCEGPDLAHLGQHKRKLRALRRLRGKPVKDERTHVELQGHIDDMAEDRFKK
GARAIINFALNTDNRIDEKTGKPIPRADVLLLENLENLIPDAERERGINRMLVQFNRGNLVKRVKEMAADAGLKVFEVSPFGTS
QVCSKCGELGRRYSIRRQQVPEDPTQTRPDIHFGLVEKLFACPKCNYRANADHNASVNLHRRFHNDGLALKSFAEFRALSDQDR
LRSIEQIDAKLLEPLRELHNVGPVTPF (SEQ ID NO: 364)

[0403] 表13.代表性CLUST.089537直接重复的核苷酸序列

效应子蛋白登录号	直接重复核苷酸序列
CAAACX010000652_28|M (SEQ ID NO: 301)	CGTGCATCGGCCGACGAAGAGCGAAGGTCAGTGACAG (SEQ ID NO: 401)
CAAACH010011569_2|M (SEQ ID NO: 302)	CGCCGCAGCGGCCGACGCGGCCCTGATCGATGGACAC (SEQ ID NO: 402)
CAAACI010012331_3|M (SEQ ID NO: 303)	GCCGCAGCGGCCGACGCGGCCCTGATCGATGGACAC (SEQ ID NO: 403)
UZXW01002767_11|M (SEQ ID NO: 304)	GGAGCGCCCCCCCGACGCGGCAATGATCGGCTGACAC (SEQ ID NO: 404)
3300009161|Ga0114966_10008307_8|M (SEQ ID NO: 305)	CTCTCAATGGCTAAAGATCTAGTTCTAATTGGACAC (SEQ ID NO: 405)
3300009161|Ga0114966_10008307_7|P (SEQ ID NO: 306)	CTCTCAATGGCTAAAGATCTAGTTCTAATTGGACAC (SEQ ID NO: 405)
3300010885|Ga0133913_10083255_2|P (SEQ ID NO: 306)	CTCTCAATGGCTAAAGATCTAGTTCTAATTGGACAC (SEQ ID NO: 405)
3300020109|Ga0194112_10008260_1|P (SEQ ID NO: 307)	CCCCCCCCCGCCATTCTCGGCTGTGTACAAC (SEQ ID NO: 406)
3300020109|Ga0194112_10008260_1|M (SEQ ID NO: 308)	CCCCCCCCCGCCATTCTCGGCTGTGTACAAC (SEQ ID NO: 406)
3300020222|Ga0194125_10021011_3|M (SEQ ID NO: 309)	CGGGTGTACCCCCCCCCGCCATTCTCGGCTGTGTACAAC (SEQ ID NO: 407)
3300027770|Ga0209086_10006474_1|P (SEQ ID NO: 306)	TCTCTCAATGGCTAAAGATCTAGTTCTAATTGGACACC (SEQ ID NO: 408)
3300015360|Ga0163144_10184575_1|M (SEQ ID NO: 310)	GCGGCGCTGGCCAGCGCCGCGCTGAA (SEQ ID NO: 409)
3300013130|Ga0172363_10020541_1|M (SEQ ID NO: 311)	GTTGCACCTCGCGCGCACCTGCGGGTTGACAC (SEQ ID NO: 410)

[0404]

[0405]

3300013133|Ga0172362_10044703_1|M (SEQ ID NO: 312)	GGCGCACCGCGTGGCGTGGAACTGACCGGCTGACAC (SEQ ID NO: 411)
3300031862|Ga0315280_10005430_2|M (SEQ ID NO: 313)	CCTGTATCCGCGCGCGCGTTTGCGGCTGATGAACAC (SEQ ID NO: 412)
3300031952|Ga0315294_10035890_2|P (SEQ ID NO: 314)	GTATTCATCCCCTTCAGAAAAGGAACTCTCTGGACAC (SEQ ID NO: 413)
3300031952|Ga0315294_10041214_2|M (SEQ ID NO: 315)	GGCGCACCCGCTGGCGCGGAACAGACCGACTGACAC (SEQ ID NO: 414)
3300031952|Ga0315294_10041214_2|P (SEQ ID NO: 316)	GGCGCACCCGCTGGCGCGGAACAGACCGACTGACAC (SEQ ID NO: 414)
3300031952|Ga0315294_10058002_1|M (SEQ ID NO: 317)	CTCTCAACCGCTTGGGCAGAACCGACCGACTGACAC (SEQ ID NO: 415)
3300032046|Ga0315289_10071835_2|M (SEQ ID NO: 318)	CTGTCCCCCCCGGACCTGACCAATAAGGGATTGACAC (SEQ ID NO: 416)
3300032053|Ga0315284_10046226_2|P (SEQ ID NO: 319)	GTGACATCTTATGCGTGTGTACAGGTGGACAC (SEQ ID NO: 417)
3300032069|Ga0315282_10002412_8|P (SEQ ID NO: 320)	CCTGTATCCGCGCGCGCGTTTGCGGCTGATGAACAC (SEQ ID NO: 412)
3300032118|Ga0315277_10001015_12|P (SEQ ID NO: 321)	GAAGAAAAGACCTCTCAAAAAAAATTCTCCTGGACAC (SEQ ID NO: 418)
3300032118|Ga0315277_10012802_3|M (SEQ ID NO: 322)	GTCGCGGCGGCCGATGCGATCCTGATCGATGGACAC (SEQ ID NO: 419)
3300009499|Ga0114930_10003500_21|M (SEQ ID NO: 323)	GTGTGTTTCCTCTTAGAAAAACAAATGCTCTGGACAC (SEQ ID NO: 420)
3300011118|Ga0114922_10000510_27|M (SEQ ID NO: 324)	GGGGGGCAAGCTCGCGCGTATGCACGGATTGGCAG (SEQ ID NO: 421)
3300017960|Ga0180429_10052609_1|M (SEQ ID NO: 325)	GGTGGACGTCCCCCGCGATTCCCGGGTGTGTACCACGA (SEQ ID NO: 422)
3300017963|Ga0180437_10045654_1|M (SEQ ID NO: 326)	GTTGCAATCTGTGCGTATGTATGGGTGGACAC (SEQ ID NO: 423)
3300017971|Ga0180438_10003700_2|P (SEQ ID NO: 326)	GTTGCAATCTGTGCGTATGTATGGGTGGACAC (SEQ ID NO: 423)
3300017991|Ga0180434_10065880_1|P (SEQ ID NO: 327)	ACTATAGGTACCCTAAAAATTCCGGGTAACTACACC (SEQ ID NO: 424)
3300017992|Ga0180435_10008274_7|P (SEQ ID NO: 328)	GATGTAGCCGCACGTTGATTCACCGGCGATTACCAC (SEQ ID NO: 425)
3300018080|Ga0180433_10080815_1|M (SEQ ID NO: 329)	ACTATAGGTACCCTAAAAATTCCGGGTAACTACACC (SEQ ID NO: 424)
3300009503|Ga0123519_10002165_1|M (SEQ ID NO: 330)	GAAGAAAAAACCTCTTAAAAAAATTCTCCTGGACAC (SEQ ID NO: 426)
CAAACB010000011_94|M (SEQ ID NO: 331)	GTTGCAGTTCGCGCGTAGGCGTGAGCGGACAC (SEQ ID NO: 427)
SRR5251193_megahit_k177_201695_120|P (SEQ ID NO: 332)	GTCGCGATCCGCACACCCGAGCGAGTCGACAG (SEQ ID NO: 428)
SRR5251487_megahit_k177_199995_1|P (SEQ ID NO: 332)	GTCGCGATCCGCACACCCGAGCGAGTCGACAG (SEQ ID NO: 428)

[0406]

SRR5576005_megahit_k177_1292294_1|M（SEQ ID NO: 331）	GTTGCAGTTCGCGCGTAGGCGTGAGCGGACAC（SEQ ID NO: 427）
UOUK01008448_2|M（SEQ ID NO: 333）	GCCGCCGCAGCGGCCGACGCGGCCCTGATCGATGGACAC（SEQ ID NO: 429）
UOUL01003058_1|M（SEQ ID NO: 334）	GCCGCAGCGGCCGACGCGGCCCTGA（SEQ ID NO: 430）
3300009652|Ga0123330_1010394_3|M（SEQ ID NO: 335）	GCCGCAGCGGCCGACGCGGCCCTGA（SEQ ID NO: 430）
3300012949|Ga0153798_10025413_2|P（SEQ ID NO: 336）	GTCGCGATCCGCACACCCGAGCGAGTCGACAG（SEQ ID NO: 428）
SRR7811943_megahit_k177_406423_2|M（SEQ ID NO: 337）	GCCGCGGAGGCCGACGCGGCCCTGATGGATGGACAC（SEQ ID NO: 431）
SRR5818194_megahit_k177_802550_7|M（SEQ ID NO: 305）	CTCTCAATGGCTAAAGATCTAGTTCTAATTGGACAC（SEQ ID NO: 405）
SRR5818194_megahit_k177_802550_6|P（SEQ ID NO: 306）	CTCTCAATGGCTAAAGATCTAGTTCTAATTGGACAC（SEQ ID NO: 405）
SRR6481359_megahit_k177_1860565_4|M（SEQ ID NO: 338）	ATAGCGCCCCCCGACGCAAATCTGATTAATGGACAC（SEQ ID NO: 432）
SRR6049666_megahit_k177_256604_1|P（SEQ ID NO: 330）	GAAGAAAAAACCTCTTAAAAAAATTCTCCTGGACAC（SEQ ID NO: 426）
SRR2080392_megahit_k177_207267_1|M（SEQ ID NO: 339）	GTCGCAATGAGATAAAACATAGATATGTGTTGCAAC（SEQ ID NO: 433）
SRR2080392_megahit_k177_207267_2|P（SEQ ID NO: 340）	GTCGCAATGAGATAAAACATAGATATGTGTTGCAAC（SEQ ID NO: 433）
3300025517|Ga0207869_1001046_11|M（SEQ ID NO: 301）	GTGCATCGGCCGACGAAGAGCGAAGGTCAGTGACAG（SEQ ID NO: 434）
3300025572|Ga0207864_1000213_47|M（SEQ ID NO: 341）	CGTGCATCGGCCGACGAAGAGCGAAGGTCAGTGACAG（SEQ ID NO: 401）
3300025572|Ga0207864_1014515_2|P（SEQ ID NO: 332）	GGCGTCGCGATCCGCACACCCGAGCGAGTCGACAGCA（SEQ ID NO: 435）
3300025572|Ga0207864_1014515_3|M（SEQ ID NO: 342）	GGCGTCGCGATCCGCACACCCGAGCGAGTCGACAGCA（SEQ ID NO: 435）
SRR6505097_megahit_k177_377355_35|M（SEQ ID NO: 343）	GCCGCGTCCCCCGACGCGACCCTGATCGATGGACAC（SEQ ID NO: 436）
SRR6962291_megahit_k177_1754533_5|M（SEQ ID NO: 344）	GGCGCAGCGGCCATCAGGATGGTGATCGGCTGACAC（SEQ ID NO: 437）
SRR6962291_megahit_k177_1754533_5|P（SEQ ID NO: 345）	GGCGCAGCGGCCATCAGGATGGTGATCGGCTGACAC（SEQ ID NO: 437）
SRR6962291_megahit_k177_3367137_15|M（SEQ ID NO: 346）	GGTGAAGCGGCCATCGGGATCGTGATCGGCTGACAC（SEQ ID NO: 438）
SRR6962291_megahit_k177_3367137_14|P（SEQ ID NO: 347）	GGTGAAGCGGCCATCGGGATCGTGATCGGCTGACAC（SEQ ID NO: 438）
AQT69685.1（SEQ ID NO: 348）	GGCGCAACCCGCACACAACCGCGAATGGACAC（SEQ ID NO: 439）
PMBN01000097_4|M（SEQ ID NO: 343）	GCCGCGTCCCCCGACGCGCCCCTGATCGAT（SEQ ID NO: 440）
PMDC01000375_4|M（SEQ ID NO: 343）	GCCGCGTCCCCCGACGCGACCCTGATCGATGGACAC（SEQ ID NO: 436）
PMMM01000072_9|M（SEQ ID NO: 343）	GCCGCGTCCCCCGACGCGACCCTGATCGATGGACAC（SEQ ID NO: 436）
QBM02859.1（SEQ ID NO: 349）	ATTGCACCGGCCAACGCAAATCTGATTGATGGACAC（SEQ ID NO: 441）
SRR6963308_megahit_k177_1876783_1|M	GGTGGACGTCCCCCGCGATTCCCGGGTGTGTACCAC（SEQ ID NO: 442）

[0407]

(SEQ ID NO: 325)	
SRR6963352_megahit_k177_1225001_5|P (SEQ ID NO: 329)	TATAGGTACCCTAAAAATTCCGGGTAACTACACC (SEQ ID NO: 443)
SRR6963480_megahit_k177_1813084_1|M (SEQ ID NO: 329)	ACTATAGGTACCCTAAAAATTCCGGGTAACTACACC (SEQ ID NO: 424)
SRR6963491_megahit_k177_4304123_3|P (SEQ ID NO: 328)	GATGTAGCCGCACGTTGATTCACCGGCGATTACCAC (SEQ ID NO: 425)
SRR6963587_megahit_k177_2272343_29|M (SEQ ID NO: 326)	GTTGCAATCTGTGCGTATGTATGGGTGGACAC (SEQ ID NO: 423)
SRR6963591_megahit_k177_1621237_2|P (SEQ ID NO: 326)	GTTGCAATCTGTGCGTATGTATGGGTGGACAC (SEQ ID NO: 423)
SRR8554444_megahit_k177_608234_1|M (SEQ ID NO: 350)	GAAGAAAAAACCTCTTAAAAAAATTCTCCTGGACACT (SEQ ID NO: 444)
SRR4030068_megahit_k177_2163203_10|M (SEQ ID NO: 351)	CTCTCACCCCAGTTCGATCGAGGAAGGCGACTGACAC (SEQ ID NO: 445)
SRR5487326_megahit_k177_254484_5|M (SEQ ID NO: 352)	GGCGTACCGCCTGACGGGTTTCTCATCGATGGACAC (SEQ ID NO: 446)
SRR5487761_megahit_k177_404752_1|M (SEQ ID NO: 353)	GGCGTACCGCCTGACGGGTTTCTCATCGATGGACAC (SEQ ID NO: 446)
SRR8554485_megahit_k177_439309_3|P (SEQ ID NO: 354)	GGCGCACCGCCTGACGGGTTTCTCATCGATGGACAC (SEQ ID NO: 447)
SRR8554511_megahit_k177_4512460_1|M (SEQ ID NO: 355)	ACCGCGGCGATACCGGCTTTCCCGGGTGACTGACAC (SEQ ID NO: 448)
SRR8554511_megahit_k177_1767604_11|P (SEQ ID NO: 356)	GTCGCACCTCGCGCGTATGCACGGATTGGCAG (SEQ ID NO: 449)
SRR8554643_megahit_k177_4328113_1|P (SEQ ID NO: 355)	ACCGCGGCGATACCGGCTTTCCCGGGTGACTGACAC (SEQ ID NO: 448)
SRR8554643_megahit_k177_1281129_3|P (SEQ ID NO: 357)	GTTGCAACCTGTGCGTATGTGCGGATGGACAC (SEQ ID NO: 450)
SRR8554643_megahit_k177_1281129_3|M (SEQ ID NO: 358)	GTTGCAACCTGTGCGTATGTGCGGATGGACAC (SEQ ID NO: 450)
SRR6056591_megahit_k177_2733919_1|P (SEQ ID NO: 324)	GGGGGGCAAGCTCGCGCGTATGCACGGATTGGCAG (SEQ ID NO: 421)
SRR6263259_megahit_k177_232832_6|M (SEQ ID NO: 359)	GGAGAGCGAAAAAAAGCGGGGGAGAT (SEQ ID NO: 451)
3300031539|Ga0307380_10018846_3|P (SEQ ID NO: 357)	GTTGCAACCTGTGCGTATGTGCGGATGGACAC (SEQ ID NO: 450)
3300031539|Ga0307380_10018846_4|M (SEQ ID NO: 358)	GTTGCAACCTGTGCGTATGTGCGGATGGACAC (SEQ ID NO: 450)
3300031539|Ga0307380_10072822_2|M (SEQ ID NO: 355)	GGTACCGCGGCGATACCGGCTTTCCCGGGTGACTGACAC (SEQ ID NO: 452)
3300031539|Ga0307380_10072822_2|P (SEQ ID NO: 360)	GGTACCGCGGCGATACCGGCTTTCCCGGGTGACTGACAC (SEQ ID NO: 452)
3300031566|Ga0307378_10072489_1|M (SEQ ID NO: 355)	GGTACCGCGGCGATACCGGCTTTCCCGGGTGACTGACAC (SEQ ID NO: 452)
3300031566|Ga0307378_10072489_2|P	GGTACCGCGGCGATACCGGCTTTCCCGGGTGACTGACAC (SEQ ID NO: 452)

[0408]

(SEQ ID NO: 360)	
3300031670|Ga0307374_10070813_2|P (SEQ ID NO: 354)	GGCGCACCGCCTGACGGGTTTCTCATCGATGGACAC (SEQ ID NO: 447)
3300031965|Ga0326597_10006968_2|P (SEQ ID NO: 361)	GGCGCAACCCCGCACATGCGGGCGGGATGACAG (SEQ ID NO: 453)
3300017540|Ga0180217_1000254_36|M (SEQ ID NO: 362)	GTCGCACCCCGCGCGTGCGAGCGTGTTGACAG (SEQ ID NO: 454)
3300014494|Ga0182017_10000840_9|P (SEQ ID NO: 343)	GCCGCGTCCCCCGACGCGACCCTGATCGATGGACAC (SEQ ID NO: 436)
3300014839|Ga0182027_10004505_3|M (SEQ ID NO: 346)	GGTGAAGCGGCCATCGGGATCGTGATCGGCTGACAC (SEQ ID NO: 438)
3300014839|Ga0182027_10004505_2|P (SEQ ID NO: 347)	GGTGAAGCGGCCATCGGGATCGTGATCGGCTGACAC (SEQ ID NO: 438)
3300031715|Ga0307476_10010105_2|P (SEQ ID NO: 363)	GGAGCGCCCCCCCGACGCGGCCATGATCGACTGACGG (SEQ ID NO: 455)
3300031715|Ga0307476_10010105_1|M (SEQ ID NO: 364)	GGAGCGCCCCCCCGACGCGGCCATGATCGACTGACGG (SEQ ID NO: 455)

[0409] 表14.代表性CLUST.089537系统的非编码序列

[0410]

>3300009161|Ga0114966_10008307_7|P
ACTTTTTCTACCATGAAAAAAAACCCATCTGATAAAACAATAATGAACCTAGAATTTATGTTTTTTTATATAGAATGGATGAAA
AGATTAGAAACTAGTCCTGATTATAATAAACCTTTTATTGAAGAACTGTTTGGGATTACTATTTAGATTATCGTTAGCTAAAGA
CAATTGAAAAAATGACTTGAAGATTCTGAGCAAAAGATTAGGATTTTAAAAGGAGAAATTCCGTGGGCACAAGAGCCATAAAAG
CAAAGATTATTTGTGACAAAACAACACTTGGTTATTTGTGGCGATCTCACTGTATGTACAATCAGAAACTCTCTGAGTATATGA
AATTACTTTTTAAACTTCGACGTGGTAAGATTGGTAAAAATAAAGATTATGCGAATAAAAATGAGGAAGAAAAGCAGAAAATAA
TTGACAATCTTGAAGCAATACTTAGACCTCATCTGCAAAAAATGCACTTTACCACCAAAAAAGAACGCTCACTTGACGTTCCAG
AACACCCACTTGACGTTCCTTTTTGAGCTTGTTAAGATATGGTCATTGAGGAGTTTTGGTGGAGTTACGACCATCCCACAGACA
TGGCAAGCCATAGCTGAAAGCCTTGATAGCCTTCGTTTTGTGTCCAAAGTCTGCTGATCTGACTGCTTTCGGGCATTTAGGATA
TTACAAAAAAGGAAACTTTTTTGGCGGCGGGCAAGGCGAGGGCAATTAGAAGTAGATCTTTTGCTGTTTGACAAATTTAGTCAA
TTTTCGTCGTATTCTATTGATGCGAGTGACGAGGTATGATCGTTTTTGCCGTAGCTGCTCGCGAGCTAGGTGTAGTAACGACTT
AAGCACAGCTAAGCATTGTAAAAATTTGAATTTAAACCCTTATTTTTCAAGGTGCTCGTCAAGTGTGTTCTAAGTTGTTGTAGA
GACTTCGTTAAGGACTAGGGCCCCGTTGGAAAAAGAAAGAGTTAGATCATCTAATCCATAAGCATTGGAATATTTCCAAAGATC
AAAGATAGGAATCCAAGTGGGAAAAAAGAAAGTAGGGGGAAAAGGAAAAATAAAACAAAAATAATTTTCAAAAGAGATGTTCTG
GATTCAGAAAAAAACCAGAAGAATTTAAAATTGTGAAGGGTAAAATCTGTTTGGGGAGGAGGGTATTTCTCCACCGAACTACTC
AAGATTCTCTAAAATTAGGAAAAGTTTATGAAGACATCTAAGGCAGAATTTAAAATTATTAAATTAACATTTGCTGATAAAAAA
TGGGGACTAGTATTGATTTTTGTTGGTATTTGTTTGGGTCCTGTGATTGTAGGTATTCCATTAATTATACTTGGTGTGGGGATG
TTATGCATA (SEQ ID NO: 500)
>SRR6481359_megahit_k177_1860565_4|M
CTGCAATCGGACCTCGAAGCTGAAATGAACCTCAAAGACGACCAGACAATGATTATCGATATGGGTCCGAACGAAAAAGAGGCC
AGAGACGCAGCCGCTGTAATCGGCCAGTCCCTGCCGAAATTCGAAACCAGAATGATCGTCGTATAAAATTGACATCCCGGTTAT
GTTTTATATGATTAGGCATAATAGAGGATTACGAAAGGAGCTAATATGTCGGTGAGGTCTTTTCAAGCAAAAGTGGAATGCGAC
AAACAAATAATGGAACATCTTTGGCGGACGCACACGGTTTTCAATGAACGGCTTCCGGTAATTATCAAAATTTTGTTCCAGATG
AAACGGGGCGAATGCGGCCAAAATGATTCGGTCAATTTGCACAGGCGATTTCTGATAGATGATGCGTTAAAGAATTATTATGGA
TTCAAAAAGCTGCCTGAGAAAAAGCAAAAAGAGGCAATAGCAGCTATTGAGAACTCGCTGAAAGAAAAATTATCTACAATTCAC
AAAATATAAAATATTTTCCTTGTGTCTTAACATAGTTGAACATTTTTACTTGTCTTTGCCGATATTCTGAAAAAGGGACAAATT

[0411]

GTTGTCCCATAAAGGATAACGTTCGGGAGGCAAGACAATGCAAGCTCTTTCAAAATTCTGTTGTCAAAATAAAAGATGCACGGA
TTACGGAAAACGTGGTGCAAATAATTTATCGGTGTGCGGCAGATTCGGCAAAAATGACCATATCCGGCTGCTATACTGCCGAAC
ATGCAAGAAGCGATTTTCCTATAATTTTTGCTGGCCGGTACGGACACTTCGTTTACAAGGCCAAGATGAGCGGTGGCACGCAAG
GACACCTGCGATGGCCGCGGATTTGACAGATCATGTCTGGTCTTTGCATGAATGGTTGAGTTATCCGGCTTTTCAATTAACATA
AGACACAAGGAATATTTTTCTTGCAGCATAAATATGCAAAAAGTACAATAATCTCGTGGAGTAAGGTAGGAGTGACGCCCGCCC
GTTTAGCGTGCTATAGGCCGCTGAATGCCACGTAGCGATATGTTTTGAGTGTCAAGAGCTGCTGACCCAAAGGCCAAAAGCTGA
GTAGGGCTTGATAGTCGTGGTTATTACCATGCAATGCGGCAGCGACACATATCGCGTTAATCAAATTTGCGTTGTCGTCAAGAT
GGATTCAGGCCATTGCGCAAGTGTCCTCTTTGACAAGTTCAGACGCGAGTGCGACGGTGGTGAAGGGGGGCACTGGGAATGCTC
GCAGCGGGAATTTGTGCAACTGCTTCCAGGATTTGAGGAACTATTGACAATATGGTTTGGCGCGAATTTTTACAGTTTTTTTAC
CTTGCTCGTAAACCAAACGCTAACCGCAGATGCAAAAAGAACTTGCAGAAGAAGCCATCAAGCCACAGAGAACAGAGGAATTTT
CATAATTACGCCACTGCATATAAATCCTTATGTGCCAAATACTGATGAATAAAAAATATCATAACTTTCAAATCCATCCAAGGG
CGAACCCTTTTCAGCCGCCTAACATGATAAGGCAGCATACTAACACGACTACGACCTTTGAACAAGCTATACCAACGCTCCACC
ACACTACGTCTTTCCAGCAGCCGTTGTCCCTCAACACTCTGCAACATACAATACGAAGCATTACGACCGGGTTGT(SEQ ID NO:
501)
>UOUK01008448_2|M
CTGGAGAATGCATTGCGGGATGTGGTTAACCTCAAGCAGGATCAGGTGCTAATTGTCGATTTGGGGTCGAATGAGGAGGCGGCG
CGCGAGTCGGCGACGACGCTCGGGCCGGCGTTGCCTTCCCAAGACGAGGGGATCGTGGTAATCTAATCTGAGCGGTTCAGCACC
GGTCCCGCACCCGCGGCGATAGCTATTCTTGGAGGTGTAGACGTGGGAGTCCGGGCACTTCGAGCGCGAATCGAGTGCGATCGA
GAAACGCTTATGCACCTTTGGCGCACGCACTGCGTCTTCAATCAACGCCTACCCGCGCTGCTGAAACGCTTGTTCGCAATGCGG
CGAGGGGAAGTGGGGGGGAACGAGGCGCAGCGGCAGGTTTACCAACGTGTTGCGCAGTTTGTTCTTGCGCGCGACGGCGATCAA
GCCGTCGCCGCCTTCCGTGCTCTCCCACCGCGGGATTCCCCCGCCAGGACCTTAGCCGTTAAAAGGGTCGAAGATACCCTTCGC
CCACAGTTGATGCGCGTGCACAAACTTGCGGATGCAGGGGTTGACAGCCCCTTCTGAGGCTCTGATATACTGGGTTTGTGGGGC
GAGCCTTCGGGCGGTATATCCGAGCCGAAGTCCACGTAGTCGTCGCCCCTGTGTGTCAACACGTGCTGACTCCAAGGCGAACAG
CCGAAGGAGGCCCCGCAAGGGTGACATCGTGGCGGGCTCGCACGCCCCGGCGGCACGGACAGCGACGGCATTGATCAGGTCCGC
GTCGGCGCGGGCGGCGCTTCGCGGCGGTGTTCTTTGAAAAGTCGGTCGTCGCGGCCTTGGCCGAGGATAAAGTCGGTGCGACGC
AGCCAACTGCGACCAACGCCGTGGGCGCGTCAGGGCGCCCGCGAGTGCGACGGTGGTGAAGGGGGGGCCTGGGCACGCTCGCGA
GGGCGTAAGTCTTTGAGAATTGAGGACTTGCAGCTTGCTGGTTTTCGGATGCGACCGAACAGTTGCCGTACGTGCGGGTCACGC
TCGCAAACGCACGCGCAAGTTCGGTATTTACAGCAACTTGCGAGTGCG (SEQ ID NO: 502)
>3300012949|Ga0153798_10025413_2|P
CAAGCCGGCGACGTCGAGCGCACCTATGCCGACCTGACGCGGTCGAAGGCAGAACTCGGCTACGACCCGCAGGTGGAGATTACG
GAAGGCCTGCGGCGGTTCGTGGCGTGGTATCGGGAGAATCGGGAGAGGTATCGGCTGGAGGTGTGAGGGGGATTCGGGAGCGAC
GAAGCGACGGAGCGACGAAGGGGAAGGCGTATGATGCATAGGGGCAGGCAACTAGCCCGCCCTACTGCCTGATCTTCTCGACCG
GCCACATAATAGTGGGCGCATCTCCACGGGCACAACAGTGAGCGAGTCAGGTCGGCATGACTTCCACAGATCTGATTGTCTGTT
CTTGCATGGGAAATTGCACAGTCAACTGAATCCACGTTATCCTTGATCTTGCTGCCAAAGGAGGACCAGGTGGGGATCAAATCT
TGGCGCGTTAAATTATGGGTGGAATCGCCTGAGATACGCGCGACTCTGGAATATTGTCTCAGCCTGTGGAACACTTCGGCTGCT
TTACTTTGTCGGCATATACTAGCGATGCGTCGTGGCGACTTTGGCCCGGAAGGGCAAGCGATCTGGGAGCACCTAAAGAAAACC
CCACATAATTTTCATGTCCTTTATCCTATCTCCAATGGCAGCGGCATAGAAAAGACTACCGGATATTACCCCGATGCACGAGCA
TTCGTTGTCTTCCTTGGAACATTTCCCACTGTGCAGACCATCCAGCGCCAACCGCGGATCTACGAGATCAATGGCCAGCGCGTT
GCTCTGGATGAGATACAGCAGCGCGCTGTTGAAATCCTTGCCAGTCGCCGCGCAGGTGATACACCCTTTTAAGATTTGTTTGCG
GTATCAGACAGGAATAGTTAACATTTGCGGTGGATCAGGTCCGAAAGGGCCGGCCCGAGGTTCGGCTATAGGCCCAACCTCGCA
CCAGCGGGCGCTGTCTAGCCGAAATTGGCGAGCGAGGCTCTTCTCAGGGCCCGCTCGAAACCTGTGGCTACGATTACGTGGCTG
CGTGGGCTTTTCGGCACCCCGTTGGACACGCTTCGGTATGCGGAACGCGAACCACCAGCGCGCACCAGTTGCATAGCGGATTCG
CGTTTAGCGCAAGTTCTTTCTCAAGCGAACTGTGCAAACAGTGGGTCAGCAAATCGCTTCTGTTGCATTTGCCTGGACGAGACG
ATTCGCGAAAGTGGACTGTAGCTCTTGGTCCTGCCTTTAACTAGAGCAAGGCCGCCGATACCACTGGCGACGGCTTGCACGCCG
ACAGGGTCGCGATCGGTAACGTCCTGAATTGTGCGCACATTTGAAAAGGCAGCTCGTCCGCTGGAAGGTGTTGGTGTGTAGACA
ACCCTTACCAGACAGGAGAGCTGCATGGACAGCGTAACAGAAGCGGCGCAGG (SEQ ID NO: 503)
>CAAACB010000011_94|M

[0412]

CAACACGGCCTGGGCTACGCCCCGTTCCCCGCCTCCGTGATTTCCCCCAGGAACTGGACCGACCACTGCCACCTGAACCCCAAC
GGCGAGCGCGAAAAAGCCACGCACATCGCCGACCATGTCCGCCCCTTCATCCGCGCTCCGCGATAAGCATTTTCGGACGGAATA
TGTCTCATCGGATGCACTTGACGGTGGCTGCGCCGCCGAGCGGAGTGCCCGTGCTTGCCCATCCATCGACCCGTTGGCATCCTG
TGAACAGGAACCAGGCACATGTCCACACGCGCTATTCAGGCCAGAGTTGTCGCCGACAAGGCACTGCTGCAAAAGCTCTGGCTG
ACCCACAGGGCGTTCAACGAACATTTGCCCCATTTTTCGCAAACCCTGTTCCGCATGATGCGGGGCGAGTTGGGCGACTCCGAT
CGCTTCGCCCTTGGTGAAACCGCCATCGCTGCCGCCATGGACTTGTGGCATCGCCCCAAGCCCGAAAAAGAGGCTATGCGAAAA
TCAATCGAAGCCCAACTCCGATCAAAACTGGAAATCGAACACAAGCTCGCCGAACCGCTGCCCTAAGGCTCTTCAAGGAGCTTG
CGCTGGGCACCCAGGTTTTCATACGTTGATCGCGTGGAGCAAGGGACCGGGTCGCGTCGGTTTCCACCATTTGGCTATAGGCCA
GGCGGTTGCCACATGGCGCCATGTTTTTGAGTGTCAAGATGTGCTGGTCCTAAGGCGACACGCCGAAGGGACTTTACCCTCCGG
GCAACTGGAAGGGGCGGCACAGACCCATGGCGCGTTCACTCCTACGATCTTTGAAGCGCGAGCTGCGACGTTCGCGAACCTATA
GCGGACAACCATCTACAGAGAGGTCCGCGGGCGGCGCAACCACCGCCCCCACGAGGACTTGTCACGCTCGTTCTCCGCTCGCCT
CTTGCGGGATCGGGAATCGCGTTTCGCTTCGCGAACTGGCGATGTGCGCACTTGTGCTTCAATCCTTTGCACTGCAAGG (SEQ
ID NO: 504)
>3300020109|Ga0194112_10008260_1|M
GCCGGCTGGACCTCCGAGTTCCCCGGCCAGCCGCCCCCCACCCGCATCGCCCTGGCCGGGCCGGACGCCCTCCTGTTCTTCTAC
CTCACGCTCCTCCAGGACATCAACACCAAGACCCTCCTCCCGCCCGACCTCGAATTTTACTGGTAGCCGGGCTAGATAGTCTCT
CTTGCGGAAAGTGCGATCACAGGCGGGACGCCTGTGCCACCGCTGAGACAAGGGGCTTGGTGGCACAGGCCTCTGGCCTGTGAT
TGGGGCACCGAGGGCGTGGATACGATGGACATGGTGGACCCCGAGGCCCGCGCCGCCCTCTGGCGCACCCATGGGGTCTACAAC
GAGTGCCTCCGGTTCGCCCTGCGCCACATGCACAAGATGAAGCGCGGGGAGGCCGACCCGCGCTACGCGGAGATATTCGGCGCG
ATCCGCGTGGAGAGGCCCAACAAGGACAAGCGGGTCTTCCGCCTGACTCCCAGGGGCCAGCCAACCGTGGAAGTGGACATGGCC
GACGTGGAGAAGCGCATCCTCGCGCGGGGAAGGGCCGAAGACCTCTGCCGCGGCAAGCCCACGCCATTTTGAGCTTGACTTTCC
CGGCCAAACTGTCATGAATGGTGTTGGCCCTGCGTGGAATAGCGCGGGCCGCCGGCGGGGCTGCTATAGGCCGCGTCGGACTAC
GTCGGGCATCCAGTTGTTACGGCCGCGTGGGCCTACGGGTTTCGGCCCAAAGGGCACAAGACCGGATGGGTCACTGGACCTGTC
AGGGCGGCGGCCAGGCCCGGCGACACGACCGAGAATGTTGGGGAAAAAGGAGGCCCGGACCGCCAAAGAATGGCCCGGATGGGG
CGCTTTTTTCAGCCGAAATGCGGGCTAACTCCTTGCGCCCTCTCGACTTCGGCGGTATCTGGTGTAGGCGCGGGTAGACTTTTC
GGACGATGTCAGGGCAAGGGCGAAGGGGTCAGACCTCAATTCTCTACTTTCCTGCCCCGGGTGGTGTGCCGACAGCGGCGTCGA
GCCGCCGCAGTCCGACGGGCTGACGCACGCTCGGGTGCCCGTGACGGTGCGAGAGGCATGGGGTACGGCGGGTGGCTGGGAAAG
GACAACAGGGACCACAGGGACAAC (SEQ ID NO: 505)
>UOUL01003058_1|M
CTGGAGAATGCATTGCGGGATGTGGTTAACCTCAAGCAGGATCAGGTGCTAATTGTCGATTTGGGGTCGAATGAGGAGGCGGCG
CGCGAGTCGGCGACGACGCTCGGGCCGGCGTTGCCTTCCCAAGACGAGGGGATCGTGGTAATCTAATCTGAGCGGTTCAGCACC
GGTCCCGCACCCGCGGCGATAGCTATTCTTGGAGGTGTAGACGTGGGAGTCCGGGCACTTCGAGCGCGAATCGAGTGCGATCGA
GAAACGCTTATGCACCTTTGGCGCACGCACTGCGTCTTCAATCAACGCCTACCCGCGCTGCTGAAACGCTTGTTCGCAATGCGG
CGAGGGGAAGTGGGGGGGAACGAGGCGCAGCGGCAGGTTTACCAACGTGTTGCGCAGTTTGTTCTTGCGCGCGACGGCGATCAA
GCCGTCGCCGCCTTCCGTGCTCTCGCACCGCGGGATTCGCCCGCCAGGACCTTAGCCGTTAAAAGGGTCGAAGATACCCTTCGC
CCACAGTTGATGCGCGTGCACAAACTTGCGGATGCAGGGGTTGACAGCCCCTTCTGAGGCTCTGATATACTGGGTTTGTGGGGC
GAGCCTTCGGGCGGTATATCCGAGCCGAAGTCCACGTAGTCGTCGCCCCTGTGTGTCAACACGTGCTGACTCCAAGGCGAACAG
CCGAAGGAGGCCCCGCAAGGGTGACATCGTGGCGGGCTCGCACGCCCCGGCGGCACGGACAGCGACGGCATTGATCAGGTCCGC
GTCGGCGCGGGCGGCGCTTCGCGGCGGTGTTCTTTGAAAAGTCGGTCGTCGCGGCCTTGGCCGAGGATAAAGTCGGTGCGACGC
AGCCAACTGCGACCAACGCCGTGGGCGCGTCAGGGCGCCCGCGAGTGCGACGGTGGTGAAGGGGGGGCCTGGGCACGCTCGCGA
GGGCGTAAGTCTTTGGGAATTGCGGACTTACAGCTTGCTGGTTTTGGGATGCGACCGAACAGTTGCCGTACGTGCGGGTCACGC
TCGCAAACGCACGCGCAAGTTCGGTATTTACAGCAACTTGCGAGTGCGGCA (SEQ ID NO: 506)
>3300025572|Ga0207864_1014515_3|M
GCGTCTACGGCAATGCCAGCAAGGTCCCCTTCAGCGAAGAAGACCCCGTCGATCAGCCCATCAGCCCCTACGCAGCCACCAAAC
GCGCCGGGGAGCTGATCTGCCATACGTTCTGGCATCTGTACAAAATGCCGATCTTCTGCCTGCGGTTTTTCACCGTCTACGGCC
CCCGCCAGCGGCCGGACCTGGCCATTCACAAGTTCACGCGGCTGATCGACGCCGGCAAGCCGATCCCCGTCTTCGGCGACGGCA
GCATGTCGCGCGATTACACGTATGTGTCGGACACGGTGGCTGGGATCATGAGCGCCCTGGACCGATGCGACCGCTACCGCATCT

[0413]

ACAACCTCGGCGGCTCGAATCCGGTGACGCTCACGCAGTTGATCCAGGAGATCGGGACCGCCCTCGGCAAGCAGGCCATCATCG
ACCGCAAACCCACCCAAGCCGGCGACGTCGAGCGCACCTATGCCGACCTGACGCGGTCGAAGGCAGAACTCGGCTACGACCCGC
AGGTGGAGATTACGGAAGGCCTGCGGCGGTTCGTGGCGTGGTATCGGGAGAATCGGGAGAGGTATCGGCTGGAGGTGTGAGGGG
GATTCGGGAGCGACGAAGCGACGGAGCGACGAAGGGGAAGGCGTATGATGCATAGGGGCAGGCAACTAGCCCGCCCTACTGCCT
GATCTTCTCGACCGGCCACATAATAGTGGGCGCATCTCCACGGGCACAACAGTGAGCGAGTCAGGTCGGCATGACTTCCACAGA
TCTGATTGTCTGTTCTTGCATGGGAAATTGCACAGTCAACTGAATCCACGTTATCCTTGATCTTGCTGCCAAAGGAGGACCAGG
TGGGGATCAAATCTTGGCGCGTTAAATTATGGGTGGAATCGCCTGAGATACGCGCGACTCTGGAATATTGTCTCAGCCTGTGGA
ACACTTCGGCTGCTTTACTTTGTCGACATGTACTGGCGATGCGTCGTGGCGACTTTGGCCCGGAAGGGCAAGCGATCTGGGAGC
ACCTAAAGAAAACCCCACATAATTTTCATGTCCTTTATCCTATCTCCAATGGCAGCGGCATTGAAAAGACTACCCGGTATTATC
CCGATGCACGAGCATTCGTTGTCTTCCTTGGAACATTTCCCACTGTGCAGACCATCCAGCGCCAACCGCGGATCTACGAGATCA
ATGGCCAGCGCGTTGCTCTGGATGAGATACAGCAGCGCGCTGTTGAAATCCTTGCCAGTCGCCGCGCAGGTGATACACCCTTTT
AAGATTTGTTTGCGGTATCAGACAGGAATAGTTAACATTTGTGGTGGATCAGGTCCGAAAGGGCCGGCCCGAGGTTCGGCTATA
GGCCCAACCTCGCACCAGCGGGCGCTGTCTAGCCGAAATTGGCGAGCGAGGCTCTTCTCAGGGCCCGCTCGAAACCTGTGGCTA
CGATTACGTGGCTGCGTGGGCTTTTCGGCACCCCGTTGGACACGCTTCGGTATGCGGAACGCGAACCACCAGCGCGCACCAGTT
GCATAGCGGATTCGCGTTTAGCGCAAGTTCTTTCTCAAGCGAACTGTGCAAACAGTGGGTCAGCAAATCGCTTCTGTTGCATTT
GCCTGGACGAGACGATTCGCGAAAGTGGACTGTAGCTCTTGGTCCTGCCTTTAACTAGAGCAAATGTCTCGGCTGGCGATATCG
TCGTCGTCGGCGTCGCGATCCG (SEQ ID NO: 507)
>CAAACX010000652_28|M
GAGGATGGACCAGACGGCGCCGCCGCGCTCCTTGAGCATCGCGGCAAGAACGTGCAGGCCGGTGACCTCCGGGTTCTGACGCGC
CGCGGCGTCGGACTGGGCGTCGGCAAGGGTCTGCTGGGCGAGCGTGGTGAAGCGGTCGGTGCGCATGGGTCGTGCTCCCGAGGG
AGTCTGGCTGACTCGGTCTGCCGGATTGTTGGGCAAAACGGGGGCCAGTCGTGAGAGGCTTTGGAGGGGGGTTTTACGGAGCGC
GGACTGCCGGAATGGCAGGGGGACGCCGGCCGTCGGCCAACCGCCCCCCGGGCGCCTCCCGGCAAGTTTCGTTAGGATGGTCGC
CGTGGAGGCCGCCTCAGATGGGTGTTAAAGCCTGGAAGGCGAAGATCATCGTGGCCGACCCGGACCTGGCGGCCGCTCTGGACC
ACACCGTCCGGCAATGGAACCTCTGGGCGGAACAGGCCTGCATCGCCATCATGCGCCTGCGCCGCGGCGACTACGGGGACGAGG
TCCTGATCGGGACGTTCCCGCAGGTCCGGCGCATCTCGCGCGACCCGCCGGTGCTGTCGATCGCCGATGTCCGGGTGGACCAGC
GGGAGGCGGATCGCCTGGCCGCGGCGTGGCTGAGCGAACTGTCGGTCGGCCCGACGCCCTTCTGAGGTGGCGTCGGATGGGGGC
TTCCGGTATACTGCCGTCGCTACCGCCTGCACAGGGCGGCGCGGCGGACCGGCTATAGGCCCGTCCGTCGTACGTCGGGAACTG
TCGAGCCGGGAGTGACGTGGCGGGGCCTCGGCCCTGTTCAAGACCTTCACCCCTCTCGTGGGGGTGGTGGGCTTCCCGGCAACC
CGGCGGCTGATCCATCACGCGGCCGCCGCGAATATCGAGCGCCCACGCCGACGCCGGGAGGTTCGCGGAAGTCGCAACGCTCTT
TGACAGCCCAAGTTGCGCTCAGGAGGCCCGATGACCGGCTTGGCTCTCCGTTCAAAAACAGCCGATTCGCGGAGGAACGCTCCG
AAGTGCCGACCGGTCAACCTCTTACGGGCGGGAGCTACTACGAGGCGATGCGGATCGGCGGGAACGCCCGCTGCATCGGCCG
(SEQ ID NO: 508)
>3300018080|Ga0180433_10080815_1|M
TTCGAGATCGACACCGAATCACGCATAGGGTGTCAGGAGATAGTACCCTAAAAATTCTGAGTAATCTACACCTGTCGCGTTCTC
CTTGTTGACCATGATTTTATTAGAGATTGGTAAAAATATTAGATTGAAGATCGAATATATTAGATATTAGAGAAAAGTTATTGG
AGATAGATAATGTCAGTGAAAACAATTCAAGCAAAATTAGAATGTGACAGAGAGACTTTAGAGACTCTATGGAGAACTCATACG
GTTTTTAATGAAGGTCTTTTACCTATTTTAGCTTTATTTTTTAAGATGAAGCGTAGAGAAACAGGATCAACAGAATTACACAGA
AAGTTCTATGATTGGTCAGTAGTAGACAAGTACTTTACTTTGCTTAGAAGTGATAATTATAAGGAACAGAAAAAAGTGATTGAA
GAAGAACTTCGAGAGTTTTTGAGGAAAGAGCATTTTCCTCAAATGGCTCTGAATTAAAAAGTGGAGAATTTATCCAATGCCTAT
AGGGAAATTGGATAACCACACAGGCCTTGGTTTTGATGGTGAATATCTGTTGATTATTTGGTCTGAGGGCCTTAAATAAATTAC
ATGGATAGTA (SEQ ID NO: 509)
>3300025572|Ga0207864_1000213_47|M
GAGGATGGACCAGACGGCGCCGCCGCGCTCCTTGAGCATCGCGGCAAGAACGTGCAGGCCGGTGACCTCCGGGTTCTGACGCGC
CGCGGCGTCGGACTGGGCGTCGGCAAGGGTCTGCTGGGCGAGCGTGGTGAAGCGGTCGGTGCGCATGGGTCGTGCTCCCGAGGG
AGTCTGGCTGACTCGGTCTGCCGGATTGTTGGGCAAAACGGGGGCCAGTCGTGAGAGGCTTTGGAGGGGGGTTTTACGGAGCGC
GGACTGCCGGAATGGCAGGGGGACGCCGGCCGTCGGCCAACCGCCCCCCGGGCGCCTCCCGGCAAGTTTCGTTAGGATGGTCGC
CGTGGAGGCCGCCTCAGATGGGTGTTAAAGCCTGGAAGGCGAAGATCATCGTGGCCGACCCGGACCTGGCGGCCGCTCTGGACC

[0414]

ACACCGTCCGGCAATGGAACCTCTGGGCGGAACAGGCCTGCATCGCCATCATGCGCCTGCGCCGCGGCGACTACGGGGACGAGG
TCCTGATCGGGACGTTCCCGCAGGTCCGGCGCATCTCGCGCGACCCGCCGGTGCTGTCGATCGCCGATGTCCGGGTGGACCAGC
GGGAGGCGGATCGCCTGGCCGCGGCGTGGCTGAGCGAACTGTCGGTCGGCCCGACGCCCTTCTGAGGTGGCGTCGGATGGGGGC
TTCCGGTATACTGCCGTCGCTACCGCCTGCACAGGGCGGCGCGGCGGACCGGCTATAGGCCCGTCCGTCGTACGTCGGGAACTG
TCGAGCCGGGAGTGACGTGGCGGGGCCTCGGCCCTGTTCGAAACCTTCACCCCTCTCGTGGGGGTGGTGGGCTTCCCGGCAACC
CGGCGGCTGATCCATCACGCGGCCGCCGCGAATATCGAGCGCCCACGCCGACGCCGGGAGGTTCGCGGAAGTCGCAACGCTCTT
TGACAGCCCAAGTTGCGCTCAGGAGGCCCGATGACCGGCTTGGCTCTCCGTTCAAAAACAGCCGATTCGCGGAGGAACGCTCCG
AAGTGCCGACCGGTCAACCTCTTACGGGCGGGAGCAGATGCAGCCAAGGGGCGGACTCGAACCGCCCCTCGCACGTATCCGCCG
GAAAAGGGCGAATGCCAGTGACCGAAGAACGGGTCCGGCCGTTCCGCAACACTCAATCGCCGCCGCGGCAATCAGGCAGAGCCC
GCCTCAGCAACATCATCCCCCGCGCCCCCTGCTCCGCCCTTCTGAGATCGCGCAGGCCCTCAACCCGTCTACGCCACCACGTCC
AGCCGCTCCCCCACTCCCGGCGGGCTCGCGCCCTGGGCGACCTTCGCCGCGGCCTGGATCATCTGCACCGCGGCCTGGCCCTGC
TGCTCGGCGTGGTCGAGGACCTTCCGGGCGACGCGGATGCTCGCGGCGGA (SEQ ID NO: 510)
>3300009652|Ga0123330_1010394_3|M
CTGGAGAATGCATTGCGGGATGTGGTTAACCTCAAGCAGGATCAGGTGCTAATTGTCGATTTGGGGTCGAATGAGGAGGCGGCG
CGCGAGTCGGCGACGACGCTCGGGCCGGCGTTGCCTTCCCAAGACGAGGGGATCGTGGTAATCTAATCTGAGCGGTTCAGCACC
GGTCCCGCACCCGCGGCGATAGCTATTCTTGGAGGTGTAGACGTGGGAGTCCGGGCACTTCGAGCGCGAATCGAGTGCGATCGA
GAAACGCTTATGCACCTTTGGCGCACGCACTGCGTCTTCAATCAACGCCTACCCGCGCTGCTGAAACGCTTGTTCGCAATGCGG
CGAGGGGAAGTGGGGGGGAACGAGGCGCAGCGGCAGGTTTACCAACGTGTTGCGCAGTTTGTTCTTGCGCGCGACGGCGATCAA
GCCGTCGCCGCCTTCCGTGCTCTCGCACCGCGGGATTCGCCCGCCAGGACCTTAGCCGTTAAAAGGGTCGAAGATACCCTTCGC
CCACAGTTGATGCGCGTGCACAAACTTGCGGATGCAGGGGTTGACAGCCCCTTCTGAGGCTCTGATATACTGGGTTTGTGGGGC
GAGCCTTCGGGCGGTATATCCGAGCCGAAGTCCACGTAGTCGTCGCCCCTGTGTGTCAACACGTGCTGACTCCAAGGCGAACAG
CCGAAGGAGGCCCCGCAAGGGTGACATCGTGGCGGGCTCGCACGCCCCGGCGGCACGGACAGCGACGGCATTGATCAGGTCCGC
GTCGGCGCGGGCGGCGCTTCGCGGCGGTGTTCTTTGAAAAGTCGGTCGTCGCGGCCTTGGCCGAGGATAAAGTCGGTGCGACGC
AGCCAACTGCGACCAACGCCGTGGGCGCGTCAGGGCGCCCGCGAGTGCGACGGTGGTGAAGGGGGGGCCTGGGCACGCTCGCGA
GGGCGTAAGTCTTTGGGAATTGCGGACTTACAGCTTGCTGGTTTTGGGATGCGACCGAACAGTTGCCGTACGTGCGGGTCACGC
TCGCAAACGCACGCGCAAGTTCGGTATTTACAGCAACTTGCGAGTGCGGCA (SEQ ID NO: 506)
>3300015360|Ga0163144_10184575_1|M
GAATTGAATGCAACCGCGAAGAATTACAACATCTTTGGCGGACTCACTGCGCCTTCAACGAAAGGTTGCCATCAATCCTTTCGG
TCCTGTTCAAAATGCGTCGTGGAGAATGGGGTGACACCGAACAGATAAAGAAAAAAATGCAGGCTTTCTTCTTTCTCGATGGGC
AGGCCGTCCAGGCATTCTACGATTGGAAGGGCAAGCCGGAGCCGGAAAGAAAAACGTTTTTCGCGGCAATGGACGGCAAGCTGC
AGCCAGCGCTGAGAGTGGAACATAAATTGGATGGTCCGACACCGTTTTAAAATTCGAAGACTTGCGGCAGGATGCAATCAGGTT
ACGTTTACGTTTGTGGGTTGTACCCGCTCGGCGTGCTATAGGCAGGCCGGGACCACAAAAGCACTGCGTTTGAGTGTCAAGACC
TGCTGACCGAACGGCCAAAAGCCTAGTAGGGCCAGACTGAGGAGGGGCAACCCACCAAGGGCGGCAGGAACACGCAGTGAGTTG
TTCAGCTTGCGTTGTGAGGACTGTTGGTGGGTCTTTGACAACTCATTTTGCGCGGCTTTTTGCCGCGAGCGCGGAGGTGGTGGT
GCGGGTGCGTCGTTGCGCTCGCACCTACTTAACCTCTTACGGTGCAAGGGCCTCGTCAACGGAAATGGTGAGTCGCATGTGTCT
CGATCAGTTGATTCCACGCAGGTGCTCGTTATGGGACTGCCAAGTCCCGGTCAAATCATGGTGTGCGAGCGCGACA (SEQ ID
NO: 511)
>3300017960|Ga0180429_10052609_1|M
TTCTACAGCAAGACTTAGCCGAAATGCCCCCAAGATGTTGCGGTGTCAGCAAAGCAGGCAATACAATATCCGGCACCAAATCGC
CCGAAGTCTTGCTGTAGAGAAAATGTACGCCTGACCCCGTTGTCCCCCCTTGCCCGCCCCAAGCGATGACGGCAAGCGGAGTAC
GTCTTCTACTGAATCGAAAGGACCCGAAATGTCACAAAGGGCCCTTGCCGTACGACTTCTGGTTGACTCTGAAGACGATCGCGC
GGCGCTGTGGCGAACGCACCGCATATTCAATGAGCGCCTCGGATGGGTGCTGCGTCAGATGCATCGGATGAAACGGGGCGAACC
GGACGCGCGCGAACTCCCGAAGGTCAAGAAGCTGCGCGAGGGCGTGTATGAGCTAGCGAAGCCTGATCACGCGCCCAAAGAAGT
GAAGCCTGCCGAGCTTAGGAGTGCGGTTAGCGTCGCGGCGGCCGCGATGTGCCGCGGGCAGGCGAGCCCCTTCTAGGCATTGAC
TTCCCTTCTGTACCTGATACGATATGGATCAGAGTTGCGAGATTACCGTAACCTGACCGGCAGGGCTATAGGCCGTGTCGGTGA
GCGCCGGGAACCAGTGGTTATGGCCGCATGGGCCTACGGGTTACCCCAATGGGCACGAAACCGACCCGGTCGGCTTCGCCGATC
CAGCGAGGGCGGCGGCCTTTGCCCGGCGACACAATCGGGAATTTCGGGGAAAACTGCCGCTTCGCACCCCGGGAAAGGGGCTGG

[0415]

AAAGAGTGTTTTTTCGAGGCGTTTTCGTTCTAAGTCCTTTTGCAGCGAGGAGTTCGGTCCGGAACTTATTGTCCAA (SEQ ID NO: 512)

[0416] 实例8-工程化CLUST.089537 CRISPR-Cas系统的功能验证(图19-21)

[0417] 鉴定了CLUST.089537 CRISPR-Cas系统的最小组分，我们从命名为CAAACX010000652的宏基因组源选择一个基因座用于功能验证。

[0418] DNA合成和效应子文库克隆

[0419] 为了测试示例性CLUST.089537 CRISPR-Cas系统的活性，我们设计并合成含有pET28a(+)载体的系统。简而言之，合成编码CLUST.089537 CAAACX010000652效应子(表12中提供的氨基酸序列)的大肠杆菌密码子优化的核酸序列(金斯瑞公司)并且将其克隆至来源于pET-28a(+)(EMD-密理博公司)的定制表达系统中。载体包含在lac启动子控制下编码CLUST.089537 CAAACX010000652效应子蛋白的核酸和大肠杆菌核糖体结合序列。载体还包含在CLUST.089537 CAAACX010000652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。

[0420] 我们计算设计了含有“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质粒以及大肠杆菌必需基因的序列。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。

[0421] 我们接下来使用金门组装方法将重复-间隔子-重复文库克隆至质粒中。简而言之，我们首先使用独特的PCR引物将每种重复-间隔子-重复从OLS池(安捷伦基因组学)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)纯化，之后添加至金门组装主混合物(新英格兰实验室)并且根据制造商的说明书孵育。我们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。

[0422] 按照卢西根公司推荐的方案使用基因Pulser Xcell®(伯乐公司)将含有不同的重复-间隔子-重复元件和Cas蛋白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)中的含有氯霉素(飞世尔公司)、四环素(阿法埃莎公司)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)提取质粒DNA，以产生“输出”文库。通过使用含有与依诺米那测序化学兼容的条形码和位点的定制引物执行PCR，我们从转化前“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Nextseq550(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。

[0423] 细菌筛选测序分析

[0424] 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。

[0425] 为了鉴定导致酶活性和细菌细胞死亡的特定参数,我们使用下一代测序(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。我们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0.1(超过10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,我们取给定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)。我们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。我们研究了此矩阵中不同特征解释CLUST.089537系统靶标耗减的程度,从而得到了单一筛选内功能参数的广泛调查。

[0426] 图19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。在每种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。

[0427] 图20A-B描绘了靶向pACYC184和大肠杆菌E.Cloni必需基因的CLUST.089537 CAAACX010000652效应子的强耗减靶标的位置。

[0428] 图21描绘了侧接耗减靶标的序列的weblogo,其指示出TTC的5'PAM。

[0429] 其他实施例

[0430] 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> A2186-7021WO
<140>
<141>
<150> 62/873,602
<151> 2019-07-12
<150> 62/873,606
<151> 2019-07-12
<150> 62/872,157
<151> 2019-07-09
<150> 62/861,582
<151> 2019-06-14
<160> 834
<170> PatentIn 3.5版
<210> 1
<211> 878
<212> PRT
<213> 未知
<220>
[0001] <223> 未知的描述：
水生-淡水序列
<400> 1

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

[0002]

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

[0003]

Leu Lys Lys Ser Ser Asp Pro Trp Met Leu Thr Asn Gly Arg Leu Leu
515 520 525
Arg Ile Tyr Pro Pro Asn Arg Pro Gln Phe Ser Asp Tyr Phe Leu Thr
530 535 540
Ile Thr Leu Gln Lys Asp Thr Pro Pro Leu Ala Ile Ser Ser Gly Glu
545 550 555 560
Lys Leu Val Gly Val Asp Arg Gly Glu Gln Asn Pro Ala Ser Phe Ala
565 570 575
Val Val Asp Ala Lys Gly Arg Leu Leu Glu Arg Gly Asn Val Gly Glu
580 585 590
Lys Tyr Thr Ala Lys Ile Arg Gln Phe Ala Glu Met Lys Ser Asp Leu
595 600 605
Gln Arg Thr Ala Gly Gly Tyr Thr Arg Lys Leu Arg Ala Lys Glu Arg
610 615 620
Asn Thr Ala Ser Ala Leu Gly Gly Leu Val Met Arg Glu Leu Leu Asp
625 630 635 640
Leu Leu Val Arg His Pro Gly Ala Leu Val Phe Glu Arg Leu Ser Ser
645 650 655
Gly Ile Thr Thr Arg Gly Gly Arg Gly Thr Leu Met Gly Asn Met His
660 665 670
Tyr Glu Arg Ile Leu Ser Gly Val Glu Gly Lys Leu Ala Glu Ala Gly
675 680 685
Cys Tyr Lys Ala Pro Ser Ala Ala Lys Tyr Arg Lys Gly Ile Thr Pro
690 695 700
Phe Leu Arg Phe Val Asn Pro Ala Tyr Thr Ser Ala Thr Cys Ser Ala
705 710 715 720
Cys Gly Glu Val Phe Ser Ser Ser Trp Tyr Lys Lys Thr Ala Ala Thr
725 730 735
Ile Ser Lys Glu Ser Pro Gly Ser Trp Ser Val Thr Leu Pro Asn Gly
740 745 750
Lys Lys Leu Gln Leu Pro Cys Ser Tyr Thr Ser Tyr Met Arg Gly Arg
755 760 765
Gly Glu Lys Thr Trp Asn Thr Ala Glu Arg Leu Glu Glu Leu Phe Lys
770 775 780
Asp Arg Thr Trp Ser Asp Leu Pro Pro Ser Thr Arg Lys Thr Val Glu
785 790 795 800
Gly Leu Val Ala Arg Leu Thr His His Arg Pro Lys Arg Gly Glu Phe
805 810 815
Val Cys Pro Ser Cys Gly His Lys Asp Asp Ala Asp Leu Gln Ala Ala
820 825 830
Leu Asn Ile Ala Arg Lys Ser Ile Trp Val Ser Ser Leu Pro Pro Arg
835 840 845
Gln Ala Lys Gln Ser Glu Gly Glu Asp Arg Arg Ser Thr Glu Glu Pro
850 855 860

Trp Arg Val Trp Tyr Lys Lys Gln Ile Met Glu Lys Trp Ser
865 870 875

<210> 2
<211> 877
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水序列
<400> 2

[0004]

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

[0005]

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

[0006]

Thr Ala Ser Ala Leu Gly Gly Leu Val Ile Arg Glu Leu Leu Asp Leu
625 630 635 640
Leu Val Arg His Pro Gly Asp Leu Val Phe Glu Arg Leu Thr Ser Gly
645 650 655
Ile Thr Thr Arg Gly Gly Arg Gly Thr Leu Met Gly Asn Met His Tyr
660 665 670
Glu Arg Ile Met Ser Gly Val Glu Gly Lys Leu Ala Glu Ala Gly Cys
675 680 685
Tyr Lys Val Pro Arg Ala Thr Lys Phe Arg Ser Lys Asp Thr Arg Phe
690 695 700
Val Arg Phe Val His Pro Ala Tyr Thr Ser Ala Thr Cys Ser Ala Cys
705 710 715 720
Gly Asp Val Phe Ser Ser Gly Trp Tyr Lys Lys Thr Ala Ala Thr Ile
725 730 735
Ala Lys Glu Ser Pro Gly Ser Trp Ser Val Thr Leu Pro Ser Gly Lys
740 745 750
Lys Ile Gln Leu Pro Ser Ser Tyr Thr Tyr Tyr Met Arg Gly Arg Gly
755 760 765
Glu Lys Thr Trp Asn Thr Ala Glu Arg Leu Glu Glu Leu Phe Lys Gly
770 775 780
Arg Lys Trp Ser Asp Leu Pro Pro Ser Thr Arg Lys Thr Val Glu Gly
785 790 795 800
Leu Val Ala Arg Leu Ile His His Arg Pro Lys Arg Gly Glu Phe Val
805 810 815
Cys Pro Ser Cys Gly His Lys Asp Asp Ala Asp Leu Gln Ala Ala Leu
820 825 830
Asn Ile Ala Arg Lys Ser Ile Trp Val Arg Ser Leu Pro Pro Arg Gln
835 840 845
Ala Lys Gln Ser Glu Gly Glu Asp Arg Arg Ser Thr Glu Glu Pro Trp
850 855 860
Arg Glu Trp Tyr Lys Lys Gln Ile Arg Glu Lys Trp Ala
865 870 875

<210> 3
<211> 877
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水序列
<400> 3

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

[0007]

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

[0008]

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

Ser Arg Glu Ser Ser Gly Ser Trp Ser Val Thr Leu Pro Ser Gly Lys
740 745 750
Lys Val Gln Leu Pro Ser Ser Tyr Thr Ser Tyr Met Arg Gly Arg Gly
755 760 765
Glu Lys Thr Trp Asn Thr Ala Glu Arg Leu Glu Glu Leu Phe Lys Gly
770 775 780
Arg Lys Trp Ser Asp Leu Pro Pro Ser Thr Arg Lys Thr Val Asp Gly
785 790 795 800
Leu Val Ala Arg Leu Ile His His Arg Pro Lys Arg Gly Glu Phe Val
805 810 815
Cys Pro Ser Cys Ala His Lys Asp Asp Ala Asp Leu Gln Ala Ala Leu
820 825 830
Asn Ile Ala Arg Lys Ser Ile Trp Val Ser Ser Leu Pro His Arg Gln
835 840 845
Ala Lys Gln Ser Glu Gly Glu Asp Arg Arg Ser Thr Glu Glu Pro Trp
850 855 860
Arg Glu Trp Tyr Lys Lys Arg Ile Arg Glu Lys Trp Ala
865 870 875

<210> 4
<211> 867
<212> PRT
<213> 未知

[0009]

<220>
<223> 未知的描述:
水生-淡水序列
<400> 4
Met Ile His Lys Ser Thr Val Lys Ala Ser Val Arg His Leu Thr Arg
1 5 10 15
His Ala Leu Asp Lys Pro Gly Glu Gly Asp Pro Gly Lys Pro Thr Thr
20 25 30
Lys Thr Leu Lys Phe Arg Ile Asp Ala Lys Gly His Thr Arg Thr Arg
35 40 45
Leu Asp Gln Ile Phe Thr Ala Gln Asn His Phe Arg Arg Asp Leu Tyr
50 55 60
Ile Lys Leu Glu Gly Met Trp Asp Lys Asp Pro Glu Ser Phe Ile Ser
65 70 75 80
Met Val Asn Ser Ser Pro Ser Lys Pro Phe Gln Lys Lys Thr Ser Cys
85 90 95
Ser Gly Trp Ile Glu Thr Lys Phe Val Thr Gly Val Ala Leu Pro Glu
100 105 110
His Leu Ser Arg Glu Thr Ala Ala Ala Thr Ala Val Ala Val Gly Asn
115 120 125
Ser Leu Thr Ala Thr Phe Glu Arg Arg Lys Asn Val Ile Ser Gln Leu
130 135 140

[0010]

```
Ala Lys Asn Ile Glu Glu Asn Arg Ser Val Trp Arg Gly Asn Leu Lys
145                 150                 155                 160
Glu Leu Cys Ser Glu Leu Lys Ile Thr Pro Phe Lys Ala Pro Pro Ser
                165                 170                 175
Ile His Pro Ala Lys Ile Asn Ala Lys Thr Val Glu Lys Tyr Asn Glu
            180                 185                 190
Trp Val Ala Ala Ala Arg Ile Trp Cys Asn Leu Ile Leu Val Gln Lys
        195                 200                 205
His Arg Val Pro Arg Asn Ser Ala Cys Leu Pro Val Arg Leu Lys His
    210                 215                 220
Tyr Pro Gly Phe Pro Gly Ser Val Ile Phe Lys Lys Gly Ala Ser Leu
225                 230                 235                 240
Ala Thr Ala Ile Lys Glu Phe Arg Lys Cys Leu Ala Gly Leu Lys Arg
                245                 250                 255
Ser Gln Leu Lys Leu Phe Asp Gly Leu Thr Ala Glu Glu Trp Gly Leu
            260                 265                 270
Val Ile Asp Arg Phe Pro Ala Pro Glu Gly Glu Gly Gly Lys Gly Met
        275                 280                 285
Arg His Gln Leu Gly Lys Thr Leu Ala Val Leu Gly Lys Lys His Pro
    290                 295                 300
Asp Trp Lys Pro Arg Gln Leu Val Glu Glu Leu Val Ala Gly Leu Glu
305                 310                 315                 320
Arg Gly Ala Asp Ser Leu Val Asn His Leu Lys Asp Arg Asp Phe Ser
                325                 330                 335
Asp Arg Pro Ala Ala Ile Lys Leu Ile Asn Leu Leu Asn Asn Ala Cys
            340                 345                 350
Val Val Leu Met Glu Pro Leu Arg Val Lys Asn Leu Tyr Arg Glu Leu
        355                 360                 365
Phe Glu Val Gly Thr Pro Arg Arg Asp Ala Phe Gly Gln Ala Arg Gly
    370                 375                 380
Ala Leu Ala Glu Pro Thr Asp Gln Ser Glu Gly Leu Glu Ile Arg Gly
385                 390                 395                 400
Phe Ser Ile Arg Thr Asn Arg Pro Pro Ile Gly His Ser Ser Phe Asn
                405                 410                 415
Gly Phe Leu Ala Val Arg Gln Thr Glu Gly Lys Asp Glu Trp Ser Phe
            420                 425                 430
Phe Phe Asp Pro Asp Gly Pro Gly Gly Met Gly Leu Ser Gly Leu Pro
        435                 440                 445
Pro Ala Gly Ser Ile Pro Thr Asp Tyr Ile Gly Phe Gly Ser Thr Gly
    450                 455                 460
Gly Ser Arg Lys Lys Ala Gln Ala Ser Glu Lys Lys Leu Val Arg Glu
465                 470                 475                 480
Arg Leu Trp Ile Arg Arg Asp Arg Asn Pro Leu Gln Leu Pro Leu Gln
                485                 490                 495
```

[0011]

Phe Gly Lys Arg Gln Gly Arg Glu Tyr Leu Trp Asn Phe Asp His Gly
500 505 510
Leu Lys Lys Ser Ser Asp Pro Trp Met Leu Ala Asn Gly Arg Leu Leu
515 520 525
Arg Val Tyr Pro Pro Asn Gln Pro Gln Phe Ser Asp Tyr Phe Leu Thr
530 535 540
Ile Thr Leu Gln Lys Asp Thr Pro Pro Leu Ala Leu Ser Ala Gly Glu
545 550 555 560
Met Leu Ile Gly Val Asp Gln Gly Glu Val Asn Pro Ala Ser Tyr Ala
565 570 575
Val Thr Asp Ala Lys Gly Lys Leu Leu Thr Arg Gly Asn Val Gly Glu
580 585 590
Ser Tyr Ser Ala Lys Ile Arg Gln Phe Ala Glu Met Lys Ser Ala Glu
595 600 605
Gln Arg Thr Ala Gly Gly Tyr Thr Arg Lys Leu Arg Ser Lys Glu Arg
610 615 620
Asn Thr Ala Ser Ala Cys Gly Gly Leu Val Ile Arg Glu Leu Leu Asp
625 630 635 640
Leu Leu Val Arg Tyr Pro Gly Ala Leu Val Phe Glu Arg Leu Ser Ser
645 650 655
Gly Ile Ala Thr Arg Gly Gly Arg Gly Thr Leu Met Ser Asn Met His
660 665 670
Tyr Gln Arg Ile Thr Gly Gly Ile Glu Lys Lys Leu Ser Glu Ala Gly
675 680 685
Cys His Arg Ile Thr Asp Phe Lys Ser Asn Phe Ser Pro Phe Leu Arg
690 695 700
Met Val Gln Pro His Tyr Thr Ser Ala Thr Cys Ser Asn Cys Gly Glu
705 710 715 720
Val Phe Asp Ser Asp Trp Tyr Glu Lys Leu Thr Ala Thr Ile Arg Lys
725 730 735
Ser Val Ser Gly Ser Trp Val Val Thr Leu Pro Asp Gly Lys Glu Leu
740 745 750
Trp Leu Gln Gly Arg Glu Glu Ile Pro Trp Asn Pro Thr Asp Arg Leu
755 760 765
Glu Glu Ile Phe Lys Gly Arg Thr Phe Pro Asp Leu Thr Ser Arg Ser
770 775 780
Gln Lys Ile Val Gln Gly Ile Val Gly Arg Leu Val Ala Phe Arg Pro
785 790 795 800
Lys Arg Gly Glu Phe Val Cys Pro Cys Cys Gly His Thr Asp Asp Ala
805 810 815
Asp Leu Gln Ala Ala Leu Asn Ile Ala Arg Lys Ser Ile Trp Leu Ser
820 825 830
Thr Leu Pro Lys Arg Ser Lys Lys Glu Thr Leu Ala Glu Ala Arg Gly
835 840 845

Ser Thr Lys Pro Ala Trp Lys Lys Trp Tyr Gln Asn Arg Ile Ser Glu
850 855 860
Asn Trp Cys
865

<210> 5
<211> 878
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水序列
<400> 5

[0012]

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

[0013]

```
Ser His Gln Lys Leu Phe Gly Ala Leu Ser Asn Glu Glu Trp Ser Leu
            260                 265                 270
Gly Leu Glu Arg Phe Pro Ala Pro Glu Gly Thr Gly Gly Lys Gly Met
        275                 280                 285
Arg His Gln Leu Gly Lys Thr Leu Ala Val Leu Gly Lys Lys His Pro
    290                 295                 300
Glu Trp Lys Pro Arg Gln Tyr Ile Glu Glu Leu Val Ala Gly Leu Glu
305                 310                 315                 320
Arg Gly Ala Asp Ser Met Ala Lys His Leu Glu Asn Arg Gln Phe Ser
                325                 330                 335
Asp Arg Pro Ala Ala Ile Lys Leu Ile Asn Leu Leu Asn Asn Ala Cys
            340                 345                 350
Val Val Leu Met Glu Pro Leu Arg Val Lys Ser Leu Tyr Lys Glu Tyr
        355                 360                 365
Phe Glu Gln Asp Thr Pro Arg Arg Asn Ala Phe Gly Gln Ala Arg Gly
    370                 375                 380
Ser Leu Ala Glu Pro Thr Asp Gln Ala Asp Ser Leu Gln Ile Leu Gly
385                 390                 395                 400
Phe Ser Leu Met Asp Glu Val Arg Gly Ile Cys Arg Tyr Asp Gly Phe
                405                 410                 415
Leu Ala Phe Ala Gly Asp Lys Glu Lys Asp Gln Trp Ala Phe Phe Phe
            420                 425                 430
Asp Pro Asp Gly Pro Gly Gly Met Gly Ile Ala Asp Pro Val Lys His
        435                 440                 445
Pro Lys Gly Ser Thr Asp Ser Gly Tyr Ile Gly Phe Gly Thr Met Gly
    450                 455                 460
Gly Ser Arg Lys Arg Ala Gln Ala Ala Glu Lys Arg Leu Val Arg Glu
465                 470                 475                 480
Arg Leu Trp Ile Arg Arg Asp Arg Lys Pro Leu Gln Leu Pro Leu His
                485                 490                 495
Phe Gly Thr Arg Gln Gly Arg Glu Tyr Leu Trp Asn Phe Asp His Gly
            500                 505                 510
Leu Lys Lys Ser Ser Asp Pro Trp Met Leu Thr Asn Gly Arg Leu Leu
        515                 520                 525
Arg Ile Tyr Pro Pro Asn Arg Pro Gln Phe Ser Asp Tyr Phe Leu Thr
    530                 535                 540
Ile Thr Leu Gln Lys Asp Thr Pro Pro Leu Ala Ile Ser Ser Gly Glu
545                 550                 555                 560
Lys Leu Val Gly Val Asp Arg Gly Glu Gln Asn Pro Ala Ser Phe Ala
                565                 570                 575
Val Val Asp Ala Lys Gly Arg Leu Leu Glu Arg Gly Asn Val Gly Glu
            580                 585                 590
Lys Tyr Thr Ala Lys Ile Arg Gln Phe Ala Glu Met Lys Ser Asp Leu
        595                 600                 605
```

[0014]

Gln Arg Thr Ala Gly Gly Tyr Thr Arg Lys Leu Arg Ala Lys Glu Arg
610 615 620
Asn Thr Ala Ser Ala Leu Gly Gly Leu Val Met Arg Glu Leu Leu Asp
625 630 635 640
Leu Leu Val Arg His Pro Gly Ala Leu Val Phe Glu Arg Leu Ser Ser
645 650 655
Gly Ile Thr Thr Arg Gly Gly Arg Gly Thr Leu Met Gly Asn Met His
660 665 670
Tyr Glu Arg Ile Leu Ser Gly Val Glu Gly Lys Leu Ala Glu Ala Gly
675 680 685
Cys Tyr Lys Ala Pro Ser Ala Ala Lys Tyr Arg Lys Gly Ile Thr Pro
690 695 700
Phe Leu Arg Phe Val Asn Pro Ala Tyr Thr Ser Ala Thr Cys Ser Ala
705 710 715 720
Cys Gly Glu Val Phe Ser Ser Ser Trp Tyr Lys Lys Thr Ala Ala Thr
725 730 735
Ile Ser Lys Glu Ser Pro Gly Ser Trp Ser Val Thr Leu Pro Asn Gly
740 745 750
Lys Lys Leu Gln Leu Pro Cys Ser Tyr Thr Ser Tyr Met Arg Gly Arg
755 760 765
Gly Glu Lys Thr Trp Asn Thr Ala Glu Arg Leu Glu Glu Leu Phe Lys
770 775 780
Asp Arg Thr Trp Ser Asp Leu Pro Pro Ser Thr Arg Lys Thr Val Glu
785 790 795 800
Gly Leu Val Ala Arg Leu Thr His His Arg Pro Lys Arg Gly Glu Phe
805 810 815
Val Cys Pro Ser Cys Gly His Lys Asp Asp Ala Asp Leu Gln Ala Ala
820 825 830
Leu Asn Ile Ala Arg Lys Ser Ile Trp Val Ser Ser Leu Pro Pro Arg
835 840 845
Gln Ala Lys Gln Ser Glu Gly Glu Asp Arg Arg Ser Thr Glu Glu Pro
850 855 860
Trp Arg Val Trp Tyr Lys Lys Gln Ile Met Glu Lys Trp Ser
865 870 875

<210> 6
<211> 878
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水序列
<400> 6
Met Asn His Lys Ser Thr Val Lys Ala Ser Val Arg His Leu Thr Arg
1 5 10 15

[0015]

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

Asp Arg Pro Ala Ala Ile Lys Leu Ile Asn Leu Leu Asn Asn Ala Cys
            340                 345                 350
Val Val Leu Met Glu Pro Leu Arg Val Lys Ser Leu Tyr Lys Glu Tyr
```

[0016]

```
          355                 360                 365
Phe Glu Gln Asp Thr Pro Arg Arg Asn Ala Phe Gly Gln Ala Arg Gly
      370                 375                 380
Ser Leu Ala Glu Pro Thr Asp Gln Ala Asp Ser Leu Gln Ile Leu Gly
385                 390                 395                 400
Phe Ser Leu Met Asp Glu Val Arg Gly Ile Cys Arg Tyr Asp Gly Phe
                  405                 410                 415
Leu Ala Phe Ala Gly Asp Lys Glu Lys Asp Gln Trp Ala Phe Phe Phe
              420                 425                 430
Asp Pro Asp Gly Pro Gly Gly Met Gly Ile Ala Asp Pro Val Lys His
          435                 440                 445
Pro Lys Gly Ser Thr Asp Ser Gly Tyr Ile Gly Phe Gly Thr Met Gly
      450                 455                 460
Gly Ser Arg Lys Arg Ala Gln Ala Ala Glu Lys Arg Leu Val Arg Glu
465                 470                 475                 480
Arg Leu Trp Ile Arg Arg Asp Arg Lys Pro Leu Gln Leu Pro Leu His
                  485                 490                 495
Phe Gly Thr Arg Gln Gly Arg Glu Tyr Leu Trp Asn Phe Asp His Gly
              500                 505                 510
Leu Lys Lys Ser Ser Asp Pro Trp Met Leu Thr Asn Gly Arg Leu Pro
          515                 520                 525
Arg Ile Tyr Pro Pro Asn Arg Pro Gln Phe Ser Asp Tyr Phe Leu Thr
      530                 535                 540
Ile Thr Leu Gln Lys Asp Thr Pro Pro Leu Ala Ile Ser Ser Gly Glu
545                 550                 555                 560
Lys Leu Val Gly Val Asp Arg Gly Glu Gln Asn Pro Ala Ser Phe Ala
                  565                 570                 575
Val Val Asp Ala Lys Gly Arg Leu Leu Glu Arg Gly Asn Val Gly Glu
              580                 585                 590
Lys Tyr Thr Ala Lys Ile Arg Gln Phe Ala Glu Met Lys Ser Asp Leu
          595                 600                 605
Gln Arg Thr Ala Gly Gly Tyr Thr Arg Lys Leu Arg Ala Lys Glu Arg
      610                 615                 620
Asn Thr Ala Ser Ala Leu Gly Gly Leu Val Met Arg Glu Leu Leu Asp
625                 630                 635                 640
Leu Leu Val Arg His Pro Gly Ala Leu Val Phe Glu Arg Leu Ser Ser
                  645                 650                 655
Gly Ile Thr Thr Arg Gly Gly Arg Gly Thr Leu Met Gly Asn Met His
              660                 665                 670
Tyr Glu Arg Ile Leu Ser Gly Val Glu Gly Lys Leu Ala Glu Ala Gly
          675                 680                 685
Cys Tyr Lys Ala Pro Ser Ala Ala Lys Tyr Arg Lys Gly Ile Thr Pro
      690                 695                 700
Phe Leu Arg Phe Val Asn Pro Ala Tyr Thr Ser Ala Thr Cys Ser Ala
```

705 710 715 720
Cys Gly Glu Val Phe Ser Ser Ser Trp Tyr Lys Lys Thr Ala Ala Thr
725 730 735
Ile Ser Lys Glu Ser Pro Gly Ser Trp Ser Val Thr Leu Pro Asn Gly
740 745 750
Lys Lys Leu Gln Leu Pro Cys Ser Tyr Thr Ser Tyr Met Arg Gly Arg
755 760 765
Gly Glu Lys Thr Trp Asn Thr Ala Glu Arg Leu Glu Glu Leu Phe Lys
770 775 780
Asp Arg Thr Trp Ser Asp Leu Pro Pro Ser Thr Arg Lys Thr Val Glu
785 790 795 800
Gly Leu Val Ala Arg Leu Thr His His Arg Pro Lys Arg Gly Glu Phe
805 810 815
Val Cys Pro Ser Cys Gly His Lys Asp Asp Ala Asp Leu Gln Ala Ala
820 825 830
Leu Asn Ile Ala Arg Lys Ser Ile Trp Val Ser Ser Leu Pro Pro Arg
835 840 845

Gln Ala Lys Gln Ser Glu Gly Glu Asp Arg Arg Ser Thr Glu Glu Pro
850 855 860
Trp Arg Val Trp Tyr Lys Lys Gln Ile Met Glu Lys Trp Ser
865 870 875

[0017]

<210> 7
<211> 878
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水序列
<400> 7
Met Asn His Lys Ser Thr Val Lys Ala Ser Val Arg His Leu Thr Arg
1 5 10 15
His Ala Leu Gln Lys Pro Gly Glu Glu Ser Asp Pro Gly Glu Thr Thr
20 25 30
Lys Thr Leu Lys Phe Arg Val Asp Ala Lys Gly His Thr Arg Thr Arg
35 40 45
Leu Asp Gln Ile Phe Thr Ala Gln Arg Gln Phe Val Arg Asp Leu Leu
50 55 60
Asn Lys Leu Glu Gly Met Trp Asp Lys Asp Pro Gln Arg Phe Ile Arg
65 70 75 80
Met Val Glu Ser Ser Ala Ser Asp Pro Phe Glu Ser Lys Thr Ser Cys
85 90 95
Tyr Gly Trp Met Met Thr Lys Phe Leu Thr Gly Ser Asp Leu Pro Glu

[0018]

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

[0019]

```
    450                 455                 460
Gly Ser Arg Lys Arg Ala Gln Ala Ala Glu Lys Arg Leu Val Arg Glu
465                 470                 475                 480
Arg Leu Trp Ile Arg Arg Asp Arg Lys Pro Leu Gln Leu Pro Leu His
                485                 490                 495
Phe Gly Thr Arg Gln Gly Arg Glu Tyr Leu Trp Asn Phe Asp His Gly
            500                 505                 510
Leu Lys Lys Ser Ser Asp Pro Trp Met Leu Thr Asn Gly Arg Leu Leu
        515                 520                 525
Arg Ile Tyr Pro Pro Asn Arg Pro Gln Phe Ser Asp Tyr Phe Leu Thr
    530                 535                 540
Ile Thr Leu Gln Lys Asp Thr Pro Pro Leu Ala Ile Ser Ser Gly Glu
545                 550                 555                 560
Lys Leu Val Gly Val Asp Arg Gly Glu Gln Asn Pro Ala Ser Phe Ala
                565                 570                 575
Val Val Asp Ala Lys Gly Arg Leu Leu Glu Arg Gly Asn Val Gly Glu
            580                 585                 590
Lys Tyr Thr Ala Lys Ile Arg Gln Phe Ala Glu Met Lys Ser Asp Leu
        595                 600                 605
Gln Arg Thr Ala Gly Gly Tyr Thr Arg Lys Leu Arg Ala Lys Glu Arg
    610                 615                 620
Asn Thr Ala Ser Ala Leu Gly Gly Leu Val Met Arg Glu Leu Leu Asp
625                 630                 635                 640
Leu Leu Val Arg His Pro Gly Ala Leu Val Phe Glu Arg Leu Ser Ser
                645                 650                 655
Gly Ile Thr Thr Arg Gly Gly Arg Gly Thr Leu Met Gly Asn Met His
            660                 665                 670
Tyr Glu Arg Ile Leu Ser Gly Val Glu Gly Lys Leu Ala Glu Ala Gly
        675                 680                 685
Cys Tyr Lys Ala Pro Ser Ala Ala Lys Tyr Arg Lys Gly Ile Thr Pro
    690                 695                 700
Phe Leu Arg Phe Val Asn Pro Ala Tyr Thr Ser Ala Thr Cys Ser Ala
705                 710                 715                 720
Cys Gly Glu Val Phe Ser Ser Ser Trp Tyr Lys Lys Thr Ala Ala Thr
                725                 730                 735
Ile Ser Lys Glu Ser Pro Gly Ser Trp Ser Val Thr Leu Pro Asn Gly
            740                 745                 750
Lys Lys Leu Gln Leu Pro Cys Ser Tyr Thr Ser Tyr Met Arg Gly Arg
        755                 760                 765
Gly Glu Lys Thr Trp Asn Thr Ala Glu Arg Leu Glu Glu Leu Phe Lys
    770                 775                 780
Asp Arg Thr Trp Ser Asp Leu Pro Pro Ser Thr Arg Lys Thr Val Glu
785                 790                 795                 800
Gly Leu Val Ala Arg Leu Thr His His Arg Pro Lys Arg Gly Glu Phe
```

805 810 815
Val Cys Pro Ser Cys Gly His Lys Asp Asp Ala Asp Leu Gln Ala Ala
820 825 830
Leu Asn Ile Ala Arg Lys Ser Ile Trp Val Ser Ser Leu Pro Pro Arg
835 840 845
Gln Ala Lys Gln Ser Glu Gly Glu Asp Arg Arg Ser Thr Glu Glu Pro
850 855 860
Trp Arg Val Trp Tyr Lys Lys Gln Ile Met Glu Lys Trp Ser
865 870 875

<210> 8
<211> 708
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水序列
<400> 8

[0020]

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

[0021]

210 215 220

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

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

[0022]

<210> 9
<211> 885
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水序列
<400> 9
Met Leu Asn Pro Ala Glu Met Gln Met Pro Ser Asn Ser Ile Lys Ser
1 5 10 15
Ile Lys Ala Ser Val Arg Glu Leu Tyr Asn His Ala Leu Asn Lys Gln
20 25 30
Gly Asp Gly Glu Pro Gly Glu Thr Thr Lys Thr Leu Lys Phe Lys Ile
35 40 45
Asp Pro Ser Pro Glu Asn Gln Ile Gln Leu Thr Lys Ile Phe Asp Ala
50 55 60
Gln Arg Ala Phe Leu Arg Asp Val Leu Gln Lys Leu Glu Gly Met Trp
65 70 75 80
Asp Lys His Pro Glu Lys Ile Glu Glu Met Ala Asn Cys Ser Gly Ser
85 90 95
Lys Pro Phe Glu Lys Lys Thr Ser Cys Tyr Gly Trp Leu Leu Thr Lys
100 105 110
Phe Leu Thr Gly Ser Asp Leu Pro Glu Asn Leu Ser Arg Ser Ala Ala
115 120 125
Thr Ala Ser Ile Gly Thr Leu Ala Gly Asn Leu Lys Ser Phe Leu Thr

[0023]

130 135 140

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

[0024]

485 490 495
Pro Met Ala Leu Pro Leu Leu Phe Gly Ala Arg Gln Gly Arg Glu Tyr
500 505 510
Leu Trp Asn Phe Asp Arg Gly Leu Lys Ser Ser Ser Thr Pro Trp Met
515 520 525
Leu Thr Asn Gly Arg Ile Leu Arg Ile Phe Pro Tyr Gln Lys Pro Glu
530 535 540
Leu Ala Glu Phe Tyr Leu Thr Leu Thr Leu Gln Lys Asp Asn Pro Pro
545 550 555 560
Leu Ala Leu Ser Leu Gly Glu His Leu Ile Gly Val Asp Arg Gly Glu
565 570 575
Glu Asn Pro Ala Ser Tyr Ala Leu Val Asp Ala Lys Gly Arg Leu Val
580 585 590
Glu Arg Gly His Ile Gly Gln Glu Tyr Ser Gln Arg Ile Lys Thr Phe
595 600 605
Ala Lys Glu Lys Ala Glu Leu Gln Arg Ser Leu Gly Gly Tyr Thr Arg
610 615 620
Arg Leu Arg Ala Lys Glu Arg Asn Ile Ala Ser Ala Leu Gly Gly Gln
625 630 635 640
Val Ile Arg Glu Leu Met Asn Leu Leu Val Ala Lys Lys Gly Ala Leu
645 650 655
Val Phe Glu Asn Leu Ser Ser Gly Ile Val Thr Arg Gly Gly Arg Gly
660 665 670
Thr Leu Met Gly Asn Met His Tyr Glu Arg Ile Leu Ser Gly Ile Glu
675 680 685
Gln Lys Leu Ala Glu Ala Gly Cys Tyr Glu Val Pro Asn Ala Lys Lys
690 695 700
Tyr Arg Lys Gly Val Ser Pro Phe Val Arg Phe Val Asn Pro Ser Tyr
705 710 715 720
Thr Ser Ser Thr Cys Ser Lys Cys Gly Glu Val Tyr Ser Arg Ala Trp
725 730 735
Tyr Glu Lys Ile Ala Pro Ser Leu Glu Arg Lys Gly Ala Gly Ile Trp
740 745 750
Ala Val Lys Leu Gly Glu Lys His Phe Thr Leu Pro Leu Ser Tyr Ser
755 760 765
Tyr Phe Leu Arg Gly Arg Gly Glu Gln His Lys Lys Thr Asp Glu Arg
770 775 780
Leu Thr Glu Ile Leu Lys Gly Lys Asp Trp Ala Lys Leu Ser Ala Thr
785 790 795 800
Ser Lys Lys Glu Ile Val Gly Leu Leu Gln Ser Lys Leu Val Ala Tyr
805 810 815
Arg Pro Lys Arg Gly Lys Phe Asn Cys Pro Ser Cys Glu His Glu Asp
820 825 830
Asp Ala Asp Leu Gln Ala Ala Leu Asn Ile Ala Arg Lys Thr Ile Trp

835 840 845

Val Ser Thr Leu Pro Pro Lys Lys Ala Lys Ser Gly Asp Gln Asp Phe
850 855 860

Arg Gly Val Gly Gln Glu Leu Trp Ala Lys Trp Tyr Lys Asp Arg Val
865 870 875 880

His Lys Asn Trp Ala
885

<210> 10
<211> 878
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水序列
<400> 10

Met Asn His Lys Asn Thr Val Lys Ala Ser Val Arg Glu Leu Thr Arg
1 5 10 15

His Ala Leu Gln Lys Pro Gly Glu Glu Ser Glu Pro Gly Glu Thr Thr
20 25 30

Lys Thr Leu Lys Phe Arg Val Asp Ala Lys Gly His Thr Arg Thr Arg
35 40 45

[0025]

Leu Asp Gln Ile Phe Thr Ala Gln Arg Gln Phe Val Arg Asp Leu Leu
50 55 60

Thr Lys Leu Glu Gly Met Trp Asp Lys Asp Pro Gln Arg Phe Ile Arg
65 70 75 80

Met Val Glu Ser Ser Ala Ser Asp Pro Phe Glu Ser Lys Thr Ser Cys
85 90 95

Tyr Gly Trp Met Met Thr Lys Phe Leu Thr Gly Ser Asp Leu Pro Glu
100 105 110

His Leu Ser Arg Ser Ala Ala Thr Ala Ser Val Ala Thr Leu Ser Gly
115 120 125

Asn Met Lys Ser Phe Val Thr Arg Arg Glu Asn Val Leu Lys Asp Met
130 135 140

Ala Val Val Ile Glu Arg Asn Met Glu Asn Cys Glu Ala Gly Leu Lys
145 150 155 160

Asp Leu Cys Ala Glu Leu Lys Ile Asp Arg Phe Lys Ser Pro Pro Ser
165 170 175

Val Asp Leu Glu Ser Ile Asn Ala Lys Thr Val Glu Lys Tyr Asn Glu
180 185 190

Trp Val Ala Ala Ala Arg Val Trp Cys Asn Leu Ile Leu Ala Gln Lys
195 200 205

His Lys Ile Pro Arg Gly Asp Ala Phe Phe Pro Thr Arg Leu Lys Ala
210 215 220

Tyr Pro Gly Phe Pro Gly Ser Val Ile His Lys Glu Ala Thr Ser Phe

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

[0026]

[0027]

580 585 590

Lys Tyr Thr Ala Lys Ile Arg Gln Phe Ala Glu Met Lys Ser Asp Leu
595 600 605

Gln Arg Thr Ala Gly Gly Tyr Thr Arg Lys Leu Arg Ala Lys Glu Arg
610 615 620

Asn Thr Ala Ser Ala Leu Gly Gly Leu Val Met Arg Glu Leu Leu Asp
625 630 635 640

Leu Leu Val Arg His Pro Gly Ala Leu Val Phe Glu Arg Leu Ser Ser
645 650 655

Gly Ile Thr Thr Arg Gly Gly Arg Gly Thr Leu Met Gly Asn Met His
660 665 670

Tyr Glu Arg Ile Leu Ser Gly Val Glu Gly Lys Leu Ala Glu Ala Gly
675 680 685

Cys Tyr Lys Ala Pro Ser Ala Ala Lys Tyr Arg Lys Gly Ile Thr Pro
690 695 700

Phe Leu Arg Phe Val Asn Pro Ala Tyr Thr Ser Ala Thr Cys Ser Ala
705 710 715 720

Cys Gly Glu Val Phe Ser Ser Gly Trp Tyr Lys Lys Thr Ala Ala Thr
725 730 735

Ile Ser Lys Glu Ser Leu Gly Ser Trp Ser Val Thr Leu Pro Asn Gly
740 745 750

Lys Lys Leu Gln Leu Pro Ser Ser Tyr Thr Ser Tyr Met Arg Gly Arg
755 760 765

Gly Glu Lys Thr Trp Asn Thr Ala Glu Arg Leu Glu Glu Leu Phe Gln
770 775 780

Asp Arg Thr Trp Ser Asp Leu Pro Pro Ser Thr Arg Lys Thr Val Glu
785 790 795 800

Gly Leu Val Ala Arg Leu Thr His His Arg Pro Lys Arg Gly Glu Phe
805 810 815

Val Cys Ser Ser Cys Gly His Lys Asp Asp Ala Asp Leu Gln Ala Ala
820 825 830

Leu Asn Ile Ala Arg Lys Ser Ile Trp Val Ser Ser Leu Pro Pro Arg
835 840 845

Gln Ala Lys Gln Ser Glu Gly Glu Asp Arg Arg Ser Thr Glu Glu Pro
850 855 860

Trp Arg Val Trp Tyr Lys Lys Gln Ile Met Gly Lys Trp Ser
865 870 875

<210> 11
<211> 878
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水序列

<400> 11

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

[0028]

[0029]

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

690 695 700

Phe Leu Arg Phe Val Asn Pro Ala Tyr Thr Ser Ala Thr Cys Ser Ala
705 710 715 720
Cys Gly Glu Val Phe Ser Ser Gly Trp Tyr Lys Lys Thr Ala Ala Thr
725 730 735
Ile Ser Lys Glu Ser Leu Gly Ser Trp Ser Val Thr Leu Pro Asn Gly
740 745 750
Lys Lys Leu Gln Leu Pro Ser Ser Tyr Thr Ser Tyr Met Arg Gly Arg
755 760 765
Gly Glu Lys Thr Trp Asn Thr Ala Glu Arg Leu Glu Glu Leu Phe Gln
770 775 780
Asp Arg Thr Trp Ser Asp Leu Pro Pro Ser Thr Arg Lys Thr Val Glu
785 790 795 800
Gly Leu Val Ala Arg Leu Thr His His Arg Pro Lys Arg Gly Glu Phe
805 810 815
Val Cys Ser Ser Cys Gly His Lys Asp Asp Ala Asp Leu Gln Ala Ala
820 825 830
Leu Asn Ile Ala Arg Lys Ser Ile Trp Val Ser Ser Leu Pro Pro Arg
835 840 845
Gln Ala Lys Gln Ser Glu Gly Glu Asp Arg Arg Ser Thr Glu Glu Pro
850 855 860
Trp Arg Val Trp Tyr Lys Lys Gln Ile Met Gly Lys Trp Ser
865 870 875

[0030]

<210> 12
<211> 951
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水-沼泽序列
<400> 12
Met Pro Gln Thr Ile Glu Leu Lys Leu Phe Met Asn Ser Ile His Lys
1 5 10 15
Ser Val Arg Ser Leu Leu Ser Leu Gly Leu Ser Arg Ser Ala Ser Ala
20 25 30
Ser Pro Gln Thr Ile Thr Arg Thr Val Lys Phe Lys Ile Asn Thr Asp
35 40 45
Ile Arg Pro Glu Leu Ile Pro Val Leu Asn Arg His Phe Asp Ala Phe
50 55 60
Glu Lys Phe Arg Arg Lys Val Leu Asp Glu Leu Glu Ala Trp Trp Asn
65 70 75 80
Lys Lys Pro Asp Asp Phe Thr Ser Met Val Arg Ala Thr Ala Lys Glu
85 90 95
Arg Phe Gln Asp Lys Ser Ser Cys Tyr Gly Trp Leu Tyr Arg His Phe

[0031]

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

[0032]

```
    450                 455                 460
Gly Pro Arg Pro Pro Arg Asp Phe Gln Pro Thr Asn Trp His Gly Phe
465                 470                 475                 480
Ala Thr Lys Gly Ser Asp Asp Glu Lys Glu Ile Val Arg Gly Thr Val
                485                 490                 495
Trp Leu Pro Thr Glu Ile Ser Lys Thr Lys Pro Ala Leu Ala Leu Pro
            500                 505                 510
Leu Ala Phe Gly Ala Arg Gln Gly Arg Glu Tyr Leu Trp His Phe Asp
        515                 520                 525
Arg Asp Leu Arg Thr Arg Asn Glu Trp Thr Pro Leu Asn Ala Arg Leu
    530                 535                 540
Leu Arg Val Met Pro Ala Ala Gln Glu Ser Glu Ser Glu Lys Glu Arg
545                 550                 555                 560
Ala Leu Arg Ala Asn Arg Ala Asp Phe Tyr Val Thr Ile Thr Phe Leu
                565                 570                 575
Arg Leu Ala Pro Pro Ile Ala Ser Val Lys Ser Lys Glu Leu Ile Gly
            580                 585                 590
Val Asp Arg Gly Glu Val Val Pro Ala Ala Phe Ala Val Val Gly Ala
        595                 600                 605
Thr Gly Arg Phe Leu Ala Lys Pro Ser Arg Leu Arg Glu Phe Met Glu
    610                 615                 620
Tyr Leu Arg Gln Leu Asp Asn Trp Glu Lys Leu Asn Leu Ala His Pro
625                 630                 635                 640
Glu Asn Ala Arg Glu Asn Glu Arg Lys Pro Arg Trp Glu Ala Arg Pro
                645                 650                 655
Asp Asp Gly Phe Gly Phe Ile Ala Pro Glu Tyr Arg Asp Gln Gln Arg
            660                 665                 670
Glu Phe Asn Lys Leu Lys Arg Glu Leu Gln Gln Thr Lys Gly Gly Tyr
        675                 680                 685
Thr Arg Trp Leu Arg Ala Lys Glu Arg Asn Arg Ala Arg Ala Leu Gly
    690                 695                 700
Gly Gln Val Met Arg Glu Leu Leu Met Leu Cys Ala Thr His Lys Ala
705                 710                 715                 720
Pro Leu Val Phe Glu Ser Leu Gly Ser Gly Leu Ala Thr Arg Gly Gly
                725                 730                 735
Lys Gly Thr Met Met Ser Gln Met Gln Tyr Ser Arg Val Phe Thr Ser
            740                 745                 750
Val Glu Gln Lys Leu Ala Glu Ala Gly Leu Cys Ala Ile Ala Arg Gln
        755                 760                 765
Pro Lys Phe Arg Lys Trp Asp Asn Gly Phe Ile Lys Leu Val Gly Pro
    770                 775                 780
Ala Cys Thr Ser Ala Thr Cys Ser Val Cys Gly Gln Val His Ser Thr
785                 790                 795                 800
Ser Phe Tyr Ala Thr Ile Ala Ala Thr Leu Lys Glu Ser Ser Lys Gly
```

805 810 815
Asn Trp Gln Val Thr Leu Pro Asn Gly Lys Val Arg Arg Leu Pro Lys
820 825 830
Glu Tyr Gly Tyr Trp Leu Arg Gly Lys Gly Glu Gln Thr Ala Asn Thr
835 840 845
Asp Glu Arg Ile His Gln Leu Leu Asn Lys Pro Phe Ala Glu Leu Ser
850 855 860
Thr Thr Asn His Asn Ala Leu Ile Ser Leu Leu Lys Ser Arg Trp Leu
865 870 875 880
Pro Phe Arg Pro Asp Gln Ala Thr Phe Gln Cys Leu Ala Cys Gly Tyr
885 890 895
Ser Val Asn Ala Asp Val Gln Ala Ala Leu Asn Ile Ala Arg Lys His
900 905 910
Leu Phe Arg Ile Gly Arg Glu Lys Lys Ser Asp Glu Glu Asn Glu Ala
915 920 925
Ala Arg Arg Lys Val Met Gly Glu Trp Glu Thr Trp Tyr Arg Gln Lys
930 935 940
Leu Ala Ser Val Trp Asn Lys
945 950

<210> 13
<211> 1000
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水-沼泽序列
<400> 13

[0033]

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

130 135 140
Glu Arg Phe Gln Asp Lys Ser Ser Cys Tyr Gly Trp Leu Tyr Arg His
145 150 155 160
Phe Leu Ser Gly Lys Asp Leu Pro Ala Gly Leu Thr Arg Glu Ala Ala
165 170 175
Asn Ser Val Leu Trp Ser Leu Lys Gly Gly Leu Lys Ser Phe Leu Thr
180 185 190
Arg Arg Glu Asn Val Ala Lys Glu Leu Glu Glu Asn Tyr Thr Thr Asn
195 200 205
Ala Asn Leu Trp Asn Glu Ser Ile Ile Arg Ile Ile Arg Asp His Lys
210 215 220
Leu Asp Lys Pro Thr Pro Pro Thr Ala Pro Pro Thr Val Asn Phe Lys
225 230 235 240
Lys Leu Thr Asp Asp Gln Ile Glu Lys Tyr Asn Gly Trp Val Gly Ile
245 250 255
Val Arg Met Trp Gly Asn Leu Ile Leu His Gln Gln Leu Val Gln Lys
260 265 270
Gly Leu Val Asp Arg Glu Asp Ala Lys Leu Pro Arg Arg Leu Lys Gly
275 280 285
Tyr Pro Gly Phe Pro Gly Ser Gln Arg Tyr Leu Gln Gly Ala Ser Leu
290 295 300
[0034] Ala Asp Ser Leu Lys Arg Leu Asp Asp Ile Thr Gly Lys Phe Ala Lys
305 310 315 320
Asp Ser Glu Lys Arg Phe Ala Lys Ile Thr Ser Glu His Trp Gly Ala
325 330 335
Val Leu Glu Arg Phe Phe Pro Pro Glu Ala Lys Glu Asp Glu Thr Arg
340 345 350
Arg Arg Pro Arg Thr Val Lys Gln Met Ala Ala Glu Arg Leu His Tyr
355 360 365
Leu Cys Lys His Asn Pro Asp Trp Lys Pro Glu Gln Thr Ala Lys Glu
370 375 380
Val Leu Ala Gly Leu Phe Arg Leu Ala Asp Lys Leu His Ala His Leu
385 390 395 400
Lys Asn Asn Ser Pro His Asp Arg Arg Ala Val Ile Lys Leu Ala Asn
405 410 415
Val Leu Asn Val Ala Ala Val Phe Ser Leu Glu Pro Leu Arg Ala Ser
420 425 430
Gly Asp Tyr Ile Ala Cys Phe Glu Ala Asp Glu Pro Arg Arg Gln Ala
435 440 445
Phe Ala Asp Leu Arg Gly Ala Leu Ala Gln Pro Gly Asp Asp Thr Val
450 455 460
Ala Ile Glu Val Thr Gly Phe Ser Leu Ser Ser Asn Ser Pro Gln Tyr
465 470 475 480
Cys Gly Leu Leu Ala Cys Arg Glu Ile Lys Pro Asp Cys Asp Glu Trp

[0035]

485 490 495
Ala Phe Phe Tyr Thr Leu Asp Asn Gln Asp Leu Lys Ile Phe Ala Lys
500 505 510
Pro Gly Pro Arg Pro Pro Arg Asp Phe Gln Pro Thr Asn Trp His Gly
515 520 525
Phe Ala Thr Lys Gly Ser Asp Asp Glu Lys Glu Ile Val Arg Gly Thr
530 535 540
Val Trp Leu Pro Thr Glu Ile Ser Lys Thr Lys Pro Ala Leu Ala Leu
545 550 555 560
Pro Leu Ala Phe Gly Ala Arg Gln Gly Arg Glu Tyr Leu Trp His Phe
565 570 575
Asp Arg Asp Leu Arg Thr Arg Asn Glu Trp Thr Pro Leu Asn Ala Arg
580 585 590
Leu Leu Arg Val Met Pro Ala Ala Gln Glu Ser Glu Ser Glu Lys Glu
595 600 605
Arg Ala Leu Arg Ala Asn Arg Ala Asp Phe Tyr Val Thr Ile Thr Phe
610 615 620
Leu Arg Leu Ala Pro Pro Ile Ala Ser Val Lys Ser Lys Glu Leu Ile
625 630 635 640
Gly Val Asp Arg Gly Glu Val Val Pro Ala Ala Phe Ala Val Val Gly
645 650 655
Ala Thr Gly Arg Phe Leu Ala Lys Pro Ser Arg Leu Arg Glu Phe Met
660 665 670
Glu Tyr Leu Arg Gln Leu Asp Asn Trp Glu Lys Leu Asn Leu Ala His
675 680 685
Pro Glu Asn Ala Arg Glu Asn Glu Arg Lys Pro Arg Trp Glu Ala Arg
690 695 700
Pro Asp Asp Gly Phe Gly Phe Ile Ala Pro Glu Tyr Arg Asp Gln Gln
705 710 715 720
Arg Glu Phe Asn Lys Leu Lys Arg Glu Leu Gln Gln Thr Lys Gly Gly
725 730 735
Tyr Thr Arg Trp Leu Arg Ala Lys Glu Arg Asn Arg Ala Arg Ala Leu
740 745 750
Gly Gly Gln Val Met Arg Glu Leu Leu Met Leu Cys Ala Thr His Lys
755 760 765
Ala Pro Leu Val Phe Glu Ser Leu Gly Ser Gly Leu Ala Thr Arg Gly
770 775 780
Gly Lys Gly Thr Met Met Ser Gln Met Gln Tyr Ser Arg Val Phe Thr
785 790 795 800
Ser Val Glu Gln Lys Leu Ala Glu Ala Gly Leu Cys Ala Ile Ala Arg
805 810 815
Gln Pro Lys Phe Arg Lys Trp Asp Asn Gly Phe Ile Lys Leu Val Gly
820 825 830
Pro Ala Cys Thr Ser Ala Thr Cys Ser Val Cys Gly Gln Val His Ser

835 840 845

Thr Ser Phe Tyr Ala Thr Ile Ala Ala Thr Leu Lys Glu Ser Ser Lys
850 855 860

Gly Asn Trp Gln Val Thr Leu Pro Asn Gly Lys Val Arg Arg Leu Pro
865 870 875 880

Lys Glu Tyr Gly Tyr Trp Leu Arg Gly Lys Gly Glu Gln Thr Ala Asn
885 890 895

Thr Asp Glu Arg Ile His Gln Leu Leu Asn Lys Pro Phe Ala Glu Leu
900 905 910

Ser Thr Thr Asn His Asn Ala Leu Ile Ser Leu Leu Lys Ser Arg Trp
915 920 925

Leu Pro Phe Arg Pro Asp Gln Ala Thr Phe Gln Cys Leu Ala Cys Gly
930 935 940

Tyr Ser Val Asn Ala Asp Val Gln Ala Ala Leu Asn Ile Ala Arg Lys
945 950 955 960

His Leu Phe Arg Ile Gly Arg Glu Lys Lys Ser Asp Glu Glu Asn Glu
965 970 975

Ala Ala Arg Arg Lys Val Met Gly Glu Trp Glu Thr Trp Tyr Arg Gln
980 985 990

Lys Leu Ala Ser Val Trp Asn Lys
995 1000

[0036]

<210> 14
<211> 878
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水-淡水和沉积物序列
<400> 14

Met Asn His Lys Ser Thr Val Lys Ala Ser Val Arg His Leu Thr Arg
1 5 10 15

His Ala Leu Gln Lys Pro Gly Glu Glu Ser Asp Pro Gly Glu Thr Thr
20 25 30

Lys Thr Leu Lys Phe Arg Val Asp Ala Lys Ala His Thr Arg Thr Arg
35 40 45

Leu Asp Gln Ile Phe Thr Ala Gln Arg Gln Phe Val Arg Asp Leu Leu
50 55 60

Asn Lys Leu Glu Gly Met Trp Asp Lys Asp Pro Gln Arg Phe Ile Arg
65 70 75 80

Met Val Glu Ser Ser Ala Ser Asp Pro Phe Glu Ser Lys Thr Ser Cys
85 90 95

Tyr Gly Trp Met Met Thr Lys Phe Leu Thr Gly Ser Asp Leu Pro Glu
100 105 110

His Leu Ser Arg Ser Ala Ala Thr Ala Ser Val Ala Thr Leu Ser Gly

[0037]

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

[0038]

```
465                 470                 475                 480
Arg Leu Trp Ile Arg Arg Asp Arg Lys Pro Leu Gln Leu Pro Leu His
                485                 490                 495
Phe Gly Thr Arg Gln Gly Arg Glu Tyr Leu Trp Asn Phe Asp His Gly
            500                 505                 510
Leu Lys Lys Ser Ser Asp Pro Trp Met Leu Thr Asn Gly Arg Leu Leu
        515                 520                 525
Arg Ile Tyr Pro Pro Asn Arg Pro Gln Phe Ser Asp Tyr Phe Leu Thr
    530                 535                 540
Ile Thr Leu Gln Lys Asp Thr Pro Pro Leu Ala Ile Ser Ser Gly Glu
545                 550                 555                 560
Lys Leu Val Gly Val Asp Arg Gly Glu Gln Asn Pro Ala Ser Phe Ala
                565                 570                 575
Val Val Asp Ala Lys Gly Arg Leu Leu Glu Arg Gly Asn Val Gly Glu
            580                 585                 590
Lys Tyr Thr Ala Lys Ile Arg Gln Phe Ala Glu Met Lys Ser Asp Leu
        595                 600                 605
Gln Arg Thr Ala Gly Gly Tyr Thr Arg Lys Leu Arg Ala Lys Glu Arg
    610                 615                 620
Asn Thr Ala Ser Ala Leu Gly Gly Leu Val Met Arg Glu Leu Leu Asp
625                 630                 635                 640
Leu Leu Val Arg His Pro Gly Ala Leu Val Phe Glu Arg Leu Ser Ser
                645                 650                 655
Gly Ile Thr Thr Arg Gly Gly Arg Gly Thr Leu Met Gly Asn Met His
            660                 665                 670
Tyr Glu Arg Ile Leu Ser Gly Val Glu Gly Lys Leu Ala Glu Ala Gly
        675                 680                 685
Cys Tyr Lys Ala Pro Ser Ala Ala Lys Tyr Arg Lys Gly Ile Thr Pro
    690                 695                 700
Phe Leu Arg Phe Val Asn Pro Ala Tyr Thr Ser Ala Thr Cys Ser Ala
705                 710                 715                 720
Cys Gly Glu Val Phe Ser Ser Ser Trp Tyr Lys Lys Thr Ala Ala Thr
                725                 730                 735
Ile Ser Lys Glu Ser Pro Gly Ser Trp Ser Val Thr Leu Pro Asn Gly
            740                 745                 750
Lys Lys Leu Gln Leu Pro Cys Ser Tyr Thr Ser Tyr Met Arg Gly Arg
        755                 760                 765
Gly Glu Lys Thr Trp Asn Thr Ala Glu Arg Leu Glu Glu Leu Phe Lys
    770                 775                 780
Asp Arg Thr Trp Ser Asp Leu Pro Pro Ser Thr Arg Lys Thr Val Glu
785                 790                 795                 800
Gly Leu Val Ala Arg Leu Thr His His Arg Pro Lys Arg Gly Glu Phe
                805                 810                 815
Val Cys Pro Ser Cys Gly His Lys Asp Asp Ala Asp Leu Gln Ala Ala
```

820 825 830

Leu Asn Ile Ala Arg Lys Ser Ile Trp Val Ser Ser Leu Pro Pro Arg
835 840 845

Gln Ala Lys Gln Ser Glu Gly Glu Asp Arg Arg Ser Thr Glu Glu Pro
850 855 860

Trp Arg Val Trp Tyr Lys Lys Gln Ile Met Glu Lys Trp Ser
865 870 875

<210> 15

<211> 941

<212> PRT

<213> 未知

<220>

<223> 未知的描述:
水生-淡水-泥炭地序列

<400> 15

[0039]

Met Lys Ala Lys Tyr Arg Ser Asn Ser Ile His Ala Ser Leu Arg Gln
1 5 10 15

Leu Leu Glu Leu Gly Leu Ser Lys Ser Ser Ser Ala Glu Pro Gln Lys
20 25 30

Ile Thr Arg Thr Ala Lys Phe Lys Ile Asn Thr Asp Ile Arg Pro Asp
35 40 45

Leu Val Pro Val Leu Asn Arg His Phe Asp Ala Phe Glu Lys Phe Arg
50 55 60

Arg Lys Val Leu Asp Glu Leu Glu Ala Arg Trp Asn Lys Asp Gln Lys
65 70 75 80

Ser Phe Gln Ala Met Val Gln Cys Ser Ala Lys Glu Pro Tyr Gln Lys
85 90 95

Lys Ser Ser Cys Tyr Ala Trp Leu Asp Thr His Phe Ile Thr Glu Ala
100 105 110

Lys Ala Val Leu Asp Leu Pro Arg Lys Pro Ala Thr Ser Leu Leu Tyr
115 120 125

Asn Leu Ser Gly Gly Leu Lys Gly Phe Leu Thr Arg Arg Glu Thr Val
130 135 140

Thr Glu Asp Ile Gln Lys Arg Phe Asn Asp Asn Leu Arg Glu Trp Asn
145 150 155 160

Gly Asp Leu Ser Gln Leu Ala Ser Asp Leu Lys Ala Pro Leu Pro Pro
165 170 175

Ala Pro Pro Asn Leu Asp Phe Glu Asn Leu Thr Glu Arg Ala Ile Glu
180 185 190

Arg Tyr Asn Asp Trp Val Gly Arg Thr Arg Ala Trp Cys Asn Leu Ile
195 200 205

Leu Val Gln Gln Lys Lys Val Glu Arg Arg Asp Ala Cys Leu Pro Arg
210 215 220

Tyr Leu Lys Gly Tyr Pro Gly Phe Phe Gly Ser Gln Arg Tyr Ala Thr

225 230 235 240
Thr Ala Ser Leu Ala Glu Asn Leu Lys Lys Leu Glu Gln Glu Ala Arg
245 250 255
Glu Gln Ser Lys Lys Ala Pro Thr Arg Phe Ala Lys Leu Ser Pro Glu
260 265 270
Ile Trp Thr Ala Ile Gln Glu Arg Phe Ser Pro Pro Glu Gly His Glu
275 280 285
Ala Gly Glu Lys Arg Arg Pro Arg Thr Ala His Gln Thr Val Cys Leu
290 295 300
Arg Phe Ala Ala Leu Arg Ala Ala His Pro Glu Trp Thr Pro Ala Gln
305 310 315 320
Leu Ala Gly Glu Ile Leu Ala Gly Ile Phe His Gly Thr Glu Lys Leu
325 330 335
Lys Lys His Leu Ala Ala Asn Gly Phe Glu Asp Arg Arg Ala Val Ile
340 345 350
Lys Leu Ala Asn Leu Tyr Asn Val Val Val Ala Phe Ser Leu Asp Pro
355 360 365
Ile Arg Ala Ala Gly Asn Tyr Ile Ser Phe Tyr Lys Glu Glu Thr Pro
370 375 380
Lys Arg Asn Ala Phe Gly Asp Val Arg Gly Gly Leu His Gln Pro Ser
385 390 395 400
[0040] Asp Glu Ser Ala Ala Ile Glu Ile Met Gly Phe Gly Leu Gln Lys Glu
405 410 415
Ser Gly Lys Pro Leu Tyr Asn Gly Leu Leu Val Cys Lys Lys Ser Glu
420 425 430
Lys Glu His Asp Asp Ser Trp Ala Phe Leu Tyr Cys His Thr Glu Gly
435 440 445
Gln Met Phe Glu Leu Ala Asn Glu Lys Ala Lys Leu Arg Gly Lys Leu
450 455 460
Leu Thr Asp Trp Thr Gly Phe Ala Ser Arg Gly Gly Ser Arg Lys Lys
465 470 475 480
Ala Glu Ala Ser Ala Lys Gln Leu Val Arg Gly Arg Val Trp Ile Ser
485 490 495
Glu Lys Thr Pro Pro Thr Val Met Pro Leu Ala Phe Gly Ser Arg Gln
500 505 510
Gly Arg Glu Tyr Leu Trp His Phe Asp Arg Asp Leu Arg Glu Lys Asn
515 520 525
Glu Trp Val Leu Gly Asn Gly Arg Leu Leu Arg Ile Met Pro Pro Gly
530 535 540
Arg Pro Asn Ala Ala Asp Phe Tyr Leu Thr Ile Thr Leu Glu Arg Gln
545 550 555 560
Ala Pro Pro Val Ala Glu Phe Ile Ala Ser Lys Leu Ile Gly Ile Asp
565 570 575
Arg Gly Glu Ala Val Pro Ala Ala Tyr Ala Leu Ile Asp Arg Glu Gly

[0041]

580 585 590
Arg Trp Leu Val Asn Gln Lys Arg Phe Gln Glu Phe Leu Lys Ala Leu
595 600 605
Gly Thr Trp His Asp Glu Leu Asp Glu Trp Trp Lys Lys Asn Lys Asn
610 615 620
Thr Pro Ile Asn Asp Arg Val Lys Arg Pro Arg Leu Gln Trp Thr Ser
625 630 635 640
Glu Leu Glu Asp Gly Phe Gly Phe Val Ala Ser Glu Tyr Arg Glu Gln
645 650 655
Gln Ala Asp Phe Asn Asp Gln Lys Arg Glu Leu Gln Arg Thr Gln Gly
660 665 670
Gly Tyr Thr Arg Trp Leu Arg Ser Lys Glu Arg Asn Arg Ala Arg Ala
675 680 685
Leu Gly Gly Glu Val Thr Arg Ala Val Leu Ser Leu Ala Ser Glu His
690 695 700
Arg Ser Pro Leu Val Leu Glu Lys Leu Gly Ser Ala Leu Ala Thr Arg
705 710 715 720
Gly Gly Lys Gly Thr Met Met Ser Gln Met Gln Tyr Glu Arg Met Leu
725 730 735
Thr Thr Leu Glu Gln Lys Leu Ala Glu Val Gly Leu Tyr Ala Met Pro
740 745 750
Ser Ala Pro Lys Tyr Arg Lys Leu Thr Asn Gly Phe Ile Asn Phe Ala
755 760 765
Ala Pro His Tyr Thr Ser Ser Thr Cys Ser Ser Cys Gly Gln Val His
770 775 780
Asp Ser Ile Phe Tyr Asn Glu Leu Ala Lys Lys Ile Val Arg Pro Asn
785 790 795 800
Asp Ser Lys Trp Gln Val Thr Leu Pro Ser Gly Gln Thr Arg Thr Leu
805 810 815
Pro Gln Thr Tyr Ser Tyr Trp Leu Lys Gly Lys Gly Glu Gln Thr Lys
820 825 830
Gln Thr Asn Glu Arg Leu Asn Glu Leu Leu Gly Glu Arg Thr Val Thr
835 840 845
Gln Leu Ser Lys Ser Ser Tyr Met Thr Leu Val Ser Leu Leu Lys Asn
850 855 860
Ser Trp Leu Pro Tyr Arg Pro Gln Gln Ala Asp Phe Cys Cys Leu Asn
865 870 875 880
Cys Gly Tyr Thr Thr Asn Ala Asp Val Gln Gly Ala Leu Asn Ile Ala
885 890 895
Arg Lys Phe Leu Phe Arg Ser Glu Arg Gly Lys Lys Thr Asp Asp Ala
900 905 910
Gly Glu Glu Gly Glu Ala Lys Arg Gly Lys Tyr Lys Asp Asp Trp Gln
915 920 925
Thr Trp Tyr Arg Gln Lys Leu Glu Lys Val Trp Cys Lys

930 935 940

<210> 16
<211> 887
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
土壤宏基因组序列
<400> 16

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

[0042]

[0043]

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

625 630 635 640
Glu Val Thr Arg Lys Leu Leu Ser Leu Ala Ala Glu His Cys Ala Pro
645 650 655
Leu Val Leu Glu Asn Leu Asn Ser Ser Leu Ala Thr Arg Gly Gly Lys
660 665 670
Asn Thr Met Met Ser Gln Met Gln Tyr Glu Arg Val Phe Thr Ala Leu
675 680 685
Glu Gln Lys Leu Ala Glu Ala Gly Leu Tyr Ser Leu Pro Ser Ala Pro
690 695 700
Lys Phe Arg Lys Gly Asp Asn Ala Phe Ile Lys Leu Val Gly Pro Ala
705 710 715 720
Tyr Thr Ser Gln Thr Cys Ser Ala Cys Gly His Val His Trp Ala Glu
725 730 735
Phe Tyr Glu Lys Ile Leu Asp Thr Leu Ala Pro Cys Gly Glu Lys Gln
740 745 750
Trp Gln Val Thr Val Glu Gly Lys Ala Arg Thr Leu Pro Glu Ser Tyr
755 760 765
Thr Tyr Trp Leu Arg Gly Lys Gly Glu Gln Thr Lys Asn Thr His Glu
770 775 780
Arg Ile Ile Glu Leu Leu Lys Asp Arg Pro Val Ser Lys Leu Asn Lys
785 790 795 800
[0044] Thr Asp Arg Lys Ser Leu Val Ser Leu Phe Lys Asn Arg Trp Met Pro
805 810 815
Tyr Arg Pro Lys Gln Ala Asp Phe Gln Cys Val Val Cys Gly His Lys
820 825 830
Met Asn Ala Asp Glu Gln Gly Ala Leu Asn Ile Ala Arg Lys Phe Leu
835 840 845
Phe Ala Ser Gly Arg Gly Lys Lys Ser Gly Glu Met Thr Glu Ala Glu
850 855 860
Arg Arg Lys Val Arg Gly Glu Trp Glu Lys Trp Tyr Lys Glu Lys Leu
865 870 875 880
Ala Thr Val Trp Lys Lys Gln
885

<210> 17
<211> 888
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
陆生-土壤-沼地序列
<400> 17
Met Ser Lys Glu Tyr Arg Pro Asp Ser Ile His Ala Ser Leu Arg His
1 5 10 15
Leu Leu Glu Leu Gly Leu Ser Lys Thr Gly Ser Ser Glu Pro Gln Lys

[0045]

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

[0046]

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

725 730 735
Ser Glu Phe Tyr Asp Gln Leu Thr Asp Thr Leu Ala Pro Cys Gly Asp
740 745 750
Lys Gln Trp Arg Val Thr Val Glu Gly Lys Pro Arg Ile Leu Pro Glu
755 760 765
Ser Tyr Thr Tyr Trp Leu Arg Gly Lys Gly Asp Gln Thr Arg Asn Thr
770 775 780
Asn Glu Arg Ile Val Glu Leu Leu Lys Asp Arg Pro Ile Ser Lys Leu
785 790 795 800
Thr Lys Thr Asp Arg Lys Ser Leu Val Ser Leu Val Lys Asn Arg Trp
805 810 815
Met Gln Tyr Arg Pro Thr Gln Ala Asp Phe Gln Cys Val Val Cys Gly
820 825 830
His Lys Met Asn Ala Asp Glu Gln Gly Ala Leu Asn Ile Ala Arg Lys
835 840 845
Phe Leu Phe Ala Leu Gly Arg Gly Lys Lys Ser Gly Glu Met Thr Glu
850 855 860
Ala Glu Arg Arg Lys Val Arg Gly Glu Trp Glu Lys Trp Tyr Lys Glu
865 870 875 880
Lys Leu Gly Thr Val Trp Arg Lys
885

[0047] <210> 18
<400> 18
000
<210> 19
<400> 19
000
<210> 20
<400> 20
000
<210> 21
<400> 21
000
<210> 22
<400> 22
000
<210> 23
<400> 23
000
<210> 24
<400> 24
000
<210> 25
<400> 25

000

<210> 26

<400> 26

000

<210> 27

<400> 27

000

<210> 28

<400> 28

000

<210> 29

<400> 29

000

<210> 30

<400> 30

000

<210> 31

<400> 31

000

<210> 32

<400> 32

000

[0048]

<210> 33

<400> 33

000

<210> 34

<400> 34

000

<210> 35

<400> 35

000

<210> 36

<400> 36

000

<210> 37

<400> 37

000

<210> 38

<400> 38

000

<210> 39

<400> 39

000

<210> 40

[0049]

<400> 40
000
<210> 41
<400> 41
000
<210> 42
<400> 42
000
<210> 43
<400> 43
000
<210> 44
<400> 44
000
<210> 45
<400> 45
000
<210> 46
<400> 46
000
<210> 47
<400> 47
000
<210> 48
<400> 48
000
<210> 49
<400> 49
000
<210> 50
<400> 50
000
<210> 51
<400> 51
000
<210> 52
<400> 52
000
<210> 53
<400> 53
000
<210> 54
<400> 54
000

<210> 55
<400> 55
000
<210> 56
<400> 56
000
<210> 57
<400> 57
000
<210> 58
<400> 58
000
<210> 59
<400> 59
000
<210> 60
<400> 60
000
<210> 61
<400> 61
000
[0050] <210> 62
<400> 62
000
<210> 63
<400> 63
000
<210> 64
<400> 64
000
<210> 65
<400> 65
000
<210> 66
<400> 66
000
<210> 67
<400> 67
000
<210> 68
<400> 68
000
<210> 69
<400> 69

000

<210> 70

<400> 70

000

<210> 71

<400> 71

000

<210> 72

<400> 72

000

<210> 73

<400> 73

000

<210> 74

<400> 74

000

<210> 75

<400> 75

000

<210> 76

<400> 76

000

[0051]

<210> 77

<400> 77

000

<210> 78

<400> 78

000

<210> 79

<400> 79

000

<210> 80

<400> 80

000

<210> 81

<400> 81

000

<210> 82

<400> 82

000

<210> 83

<400> 83

000

<210> 84

[0052]

<400> 84
000
<210> 85
<400> 85
000
<210> 86
<400> 86
000
<210> 87
<400> 87
000
<210> 88
<400> 88
000
<210> 89
<400> 89
000
<210> 90
<400> 90
000
<210> 91
<400> 91
000
<210> 92
<400> 92
000
<210> 93
<400> 93
000
<210> 94
<400> 94
000
<210> 95
<400> 95
000
<210> 96
<400> 96
000
<210> 97
<400> 97
000
<210> 98
<400> 98
000

[0053]

<210> 99
<400> 99
000
<210> 100
<400> 100
000
<210> 101
<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> 101
ggtttgctga cctcgattgt tggaggagtg caaaag 36
<210> 102
<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> 102
ggtttcgcac tccctcattg ttggtggtgt gcaaaag 37
<210> 103
<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> 103
ggtttgctcc ggtccttgtt ggtaggtgtg caaaag 36
<210> 104
<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> 104
ggttagctcc catccttgtt gggtggtttg caaaag 36
<210> 105
<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> 105
ttgcctcctg tggggttgta tgtgaag 27
<210> 106
<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> 106
ggtttgctga cctcgattgt tggaggtgtg caaaag 36
<210> 107
<211> 37
<212> DNA
<213> 人工序列

[0054]

<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 107
gtgagaaccg ctgcggttcg cggggatggt ttcaaag 37
<210> 108
<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> 108
ttcggaatcg ctgacgctcg cggggctggt ttcaaag 37
<210> 109
<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> 109
gtttgcatga tgacgtttgc ggggttggtt tcaaag 36

<210> 110
<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> 110
ctctgacgcg ctgacgttcg cggggttggt ttcaaag 37
<210> 111
<400> 111
000
<210> 112
<400> 112
000
<210> 113
<400> 113
000
<210> 114
<400> 114
000
[0055] <210> 115
<400> 115
000
<210> 116
<400> 116
000
<210> 117
<400> 117
000
<210> 118
<400> 118
000
<210> 119
<400> 119
000
<210> 120
<400> 120
000
<210> 121
<400> 121
000
<210> 122
<400> 122

000

<210> 123

<400> 123

000

<210> 124

<400> 124

000

<210> 125

<400> 125

000

<210> 126

<400> 126

000

<210> 127

<400> 127

000

<210> 128

<400> 128

000

<210> 129

<400> 129

000

[0056]

<210> 130

<400> 130

000

<210> 131

<400> 131

000

<210> 132

<400> 132

000

<210> 133

<400> 133

000

<210> 134

<400> 134

000

<210> 135

<400> 135

000

<210> 136

<400> 136

000

<210> 137

[0057]

<400> 137
000
<210> 138
<400> 138
000
<210> 139
<400> 139
000
<210> 140
<400> 140
000
<210> 141
<400> 141
000
<210> 142
<400> 142
000
<210> 143
<400> 143
000
<210> 144
<400> 144
000
<210> 145
<400> 145
000
<210> 146
<400> 146
000
<210> 147
<400> 147
000
<210> 148
<400> 148
000
<210> 149
<400> 149
000
<210> 150
<400> 150
000
<210> 151
<400> 151
000

<210> 152
<400> 152
000
<210> 153
<400> 153
000
<210> 154
<400> 154
000
<210> 155
<400> 155
000
<210> 156
<400> 156
000
<210> 157
<400> 157
000
<210> 158
<400> 158
000
[0058] <210> 159
<400> 159
000
<210> 160
<400> 160
000

<210> 161
<400> 161
000
<210> 162
<400> 162
000
<210> 163
<400> 163
000
<210> 164
<400> 164
000
<210> 165
<400> 165
000
<210> 166

<400> 166
000
<210> 167
<400> 167
000
<210> 168
<400> 168
000
<210> 169
<400> 169
000
<210> 170
<400> 170
000
<210> 171
<400> 171
000
<210> 172
<400> 172
000
<210> 173
<400> 173
000
<210> 174
<400> 174
000
<210> 175
<400> 175
000
<210> 176
<400> 176
000
<210> 177
<400> 177
000
<210> 178
<400> 178
000
<210> 179
<400> 179
000
<210> 180
<400> 180
000

[0059]

<210> 181
<400> 181
000
<210> 182
<400> 182
000
<210> 183
<400> 183
000
<210> 184
<400> 184
000
<210> 185
<400> 185
000
<210> 186
<400> 186
000
<210> 187
<400> 187
000
[0060] <210> 188
<400> 188
000
<210> 189
<400> 189
000
<210> 190
<400> 190
000
<210> 191
<400> 191
000
<210> 192
<400> 192
000
<210> 193
<400> 193
000
<210> 194
<400> 194
000
<210> 195
<400> 195

000
<210> 196
<400> 196
000
<210> 197
<400> 197
000
<210> 198
<400> 198
000
<210> 199
<400> 199
000
<210> 200
<211> 7
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 200
ccgcgag 7

[0061]

<210> 201
<211> 307
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
多核苷酸
<400> 201
tgttccgttc cgaacgaggc aaaaagactg acgacgcagg cgaagaaggt gaggcaaagc 60
gtggtaaata caaagacgac tggcagactt ggtatcgcca aaagttggaa aaggtctggt 120
gcaagtaatc tgacatcgcc cattcgagac ccgcgagcat gacatggaga ccgacgattt 180

ccttcctatc tactcgctcg cggagtggca ttattgcccg cgctcgtctt ttctgtcgtg 240
gtttggtgca gagcgtgagg atcgcgtcac cccggcttac cagaaaatgc gcgaagtcca 300
cgcggtg 307
<210> 202
<211> 792
<212> DNA
<213> 未知
<220>
<223> 未知的描述：
水生-淡水序列

<400> 202

cacaatattt cttaaatgtc gagaaaataa tattttacac ctatccatat gccaaattac 60
aaaaatgaat gttttacctc ataaagagta aaaagtgctc ccccgtcaca aaacattagt 120
ttggtgactg gttgttttta accaccaact tctctgtcca atttattggg gttttgtgaa 180
tccctattcc aaggaccgtt tttgtactac cccctaattt acagttatcc ttattatgaa 240
ctggaatagc tcccttaaaa gtccgattat aactgcttgc ttaatacgat gaatcacaaa 300
aataccgtta aagctcctgt tcgcgaactt acccggcatg cattgcagaa acccggggag 360
gagagtgaac cgggagaaac aacaaagact cttaagttcc gtgtggatgc caagggacat 420
acccgaaccc gactcgacca ggcagcactg aacattgctc ggaaaagcat ctgggtcagt 480
tcccttccac cccgtcaggc aaagcaatcg gaaggtgaag accggcgttc caccgaggaa 540
ccatggcgcg tttggtacaa gaagcagatc atgggaaagt ggtcttaaaa aatcccttgc 600
cacccggcta tcataggatc aaattaaccc cttagtattt tctttgcaca acgtcgactg 660
ccctgccaag gagcggctga taaccaacaa gttcttttac atttgaggtg caaaccgttc 720
ttgtgatttt tttataaaag gatgcacccg aaagggttgg ctatacccac atactgcccg 780
cttgcattat gt 792

<210> 203
<211> 944
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
水生-淡水序列

[0062]

<400> 203

gtctactttc acggcaacgt gaaagacttg gaacccaaac ccggttcggt cacaccgatc 60
cagaaaacgg ccgcggggct gatgtttctt gtcgatgcga tcatgctggg actcgctctt 120
tatttgatcc ttcgctaacg cgaaccgatc gataaggatt aagtcgcact acccctgcat 180
cgattcccgt tttttcgatt cccttccagg tatttaattt ttacttctca acgaggtatt 240
tggcttggca tagtctcctc tgcgcccccc gccccctgtc ccctgccccc tttcacttcc 300
agcctaccaa ttcccagtgt tcaaaacgtc ggtgttttgg cgtttatcct ttcctgagat 360
aacgcccagg actgtttttg tcccacctcc taatttacac ttatccttat tatgaactgg 420
aatagctccg ttaaaagtcc gattatgact gcttgcttaa tacgatgaat cacaaaagca 480
ctgtcaaagc ctctgttcgc catcttactc ggcatgcatt gcagaaaccc ggggaggaga 540
gtgatccggg agaaacgaca aagactttaa agttccgggt agatgccaaa ggacataccc 600
gaacccgact cgaccaggca gcactgaaca ttgcccgcaa aagcatctgg gtcagctccc 660
ttccaccccg tcaggcaaag caatcggaag gtgaagaccg gcgttccacc gaggaaccat 720
ggcgcgtttg gtacaaaaag cagatcatgg aaaagtggtc ttaaaaaatc ccttgccacc 780
cggctatcat aggatcattt tagtctctta gtattttctt tgcacaacgt cgactgccct 840
gccaaggagc ggctgacaac caacaagttc ttttacattt gaggtgcaaa ccgttcttgt 900
gatttttta caaaaggatg cacccgaaag ggttggccat accc 944

<210> 204
<211> 934
<212> DNA
<213> 未知
<220>

<223> 未知的描述:
水生-淡水序列
<400> 204
acggcaacgt gaaagacttg gaacccaaac ccggttcggt cacaccgatc cagaaaacgg 60
ccgcggggct gatgtttctt gtcgatgcga tcatgctggg actcgctctt tatttgatcc 120
ttcgctaacg cgaaccgatc gataaggatt aagtcgcact acccctgcat cgattcccgt 180
tttttcgatt cccttccagg tatttaattt ttacttctca acgaggtatt tggcttggca 240
tggtcttctc tgcgcccccc gccccctgtc ccctgccccc tttcacttcc agcctaccaa 300
ttcccagtgt tcaaaacgtc ggtgttttgg cgtttatcct ttcctgagat aacgcccagg 360
actgtttttg tcccacctcc taatttacac ttatccttat tatgaactgg aatagctccg 420
ttaaaagtcc gattatgact gcttgcttaa tacgatgaat cacaaaagca ctgtcaaagc 480
ctctgttcgc catcttactc ggcatgcatt gcagaaaccc ggggaggaga gtgatccggg 540
agaaacgaca aagactttaa agttccgggt agatgccaaa ggacataccc gaacccgact 600
cgaccaggca gcactgaaca ttgcccgcaa aagcatctgg gtcagctccc ttccaccccg 660
tcaggcaaag caatcggaag gtgaagaccg gcgttccacc gaggaaccat ggcgcgtttg 720
gtacaaaaag cagatcatgg aaaagtggtc ttaaaaaatc ccttgccacc cggctatcat 780
aggatcattt tagtctctta gtattttctt tgcacaacgt cgactgccct gccaaggagc 840
ggctgacaac caacaagttc ttttacattt gaggtgcaaa ccgttcttgt gatttttta 900
caaaaggatg cacccgaaag ggttggccat accc 934

[0063]

<210> 205
<211> 2207
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
水生-淡水序列
<400> 205
tcaaaaccag ccgtagctgg cggaaaatat gcttgggatt gtcattggta tgtggccaat 60
tcctttccca cacggcgtgc ttgtccacag gattttaacc ctgcctgttt ttttgatatc 120
aaacctgata tcggtcattt ctcctcgata ctctgaaacc aatcagaaac acccttccca 180
ggatgctgag ttaagaacca cacgctaact tatcgtttgg ctgaaaggga tttgaaccct 240
tttcttatta tattgctgaa tacagagaaa ttctaagaaa tcttaaaata tattttacct 300
gcaaacagca caaagcacta atagccagtt atatgcacac ctagtaggat tgaatcttat 360
taatacaacc taggacagca aatgacccac catatgagtt attattgtgg aatgaagaat 420
taccaaactt gtggaaagtc gatatctggg gatacgatca atgcttaatc cagctgaaat 480
gcaaatgccc tcaaatagta taaaatcgat caaaacttca gtacgagagc tgtacaacca 540
cgccttgaac aaacagggcg atggtgagcc aggagagaca accaagacac ttaagttcaa 600
gatcgatcca caagcggcgc tgaacattgc caggaagacg atctgggtca gtaccctacc 660
tcctaaaaaa gccaagagcg gtgatcagga ctttcgtggt gtaggacaag agctgtgggc 720
gaaatggtac aaggataggg ttcacaaaaa ctgggcataa aatagcactt gtctgaattg 780
ggatgggtta atacattgaa atcttagtat ttactttgca tagcgttgcc aaccttgccc 840
aaggatgagt caacgcccca tagaaggcag ttcattaaca cgcaaaaggc acaccgccgt 900
agtaagtagg gggcagcaac cacagatggt gagatgcctc ctatggggtc atatgtgaag 960
ttcatgttgt ctcccatctt ggacccatag ccgttggtga gttgcctcct gtggggccat 1020

atgtgaagga ggccaattct gctcgattgt gatccaatgt gatggtgagt tgcctcctgt 1080
ggggttgtat gtgaaggtca atgtggagcg attgctgggg gtattcagtt ggtgatttgc 1140
ctcctatggg gctatgtgtg aagggatgcg atcctgcaac attcaagcct agtcagttgc 1200
aagcgttttc tgcaaatgat agagatctgt tggtggtttg ccttctgtgg ggctatgtgt 1260
gaaggatgtc gaagcgccat tgctcggcat tgatgcgtgt gtgatctggt ttctgtggag 1320
ttatgtgtga aggctcgctc tcatagttct ttggcttggc actcatagca cttttccgtt 1380
taaaccagca aaaaatcgcc gcactagata atgcaacata atttgtataa attatccaaa 1440
taaatgcctg ttttctcaac acaagaaaaa cagcctatcc tttcatactt atctgtttct 1500
ctctgtgcat caattaaata tataatagga tcttaatcgt ataatagctt aatccagagt 1560
aacctcctat tagaaatccg ctgctctagt cccctgagct acgggcgcag cacgtagttt 1620
tgtagcagaa aattgtcaga gaggcaggga ttatggtgga gtaagaaggg ggttaaggga 1680
tgatgggatt aaggggaaat ctgaaaaagg ggttaaggga ttcaggggtt aaggaattaa 1740
ggaagagaca aatcgatgac tgacaagacg tttaaggttt tttaggacgg aggaggaggc 1800
cgaggtcggg gagcttgagc agtttggcga caaggaagta ggcgagtgac ccgagaagaa 1860
cgagcgcaaa ggttccaccg atacgaacag ggatgggttg gcccatgggt tgaatgaatt 1920
ggaggcctat aagtagggct ccaattaagg tgacgcagca gagaagtatc tttccgactc 1980
cttggaggaa gtgaggttcg aagtagctgc ctacgtgccg ttgggaggcc cacaccagct 2040
gcagaatatt cagcgtcgcc acagaggctg tggtgagagt gagtcctaaa atgccgaggt 2100
tgtagacccg aataagaatg tagttgaagg tgaggttgag aaaaattccc gtaaggctaa 2160
cgcgaagcgg aatccccggg cgatcgatag ccgagaaggc gggggcg 2207

<210> 206
<211> 737
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
水生-淡水序列
<400> 206

[0064]

cctcataaag agtaaaaagt gctcccccgt cacaaaacat tagtttggtg actggttgtt 60
tttaaccacc aacttctctg tccaatttat tggggttttg tgaatcccta ttccaaggac 120
cgtttttgta ctacccccta atttacagtt atccttatta tgaactggaa tagctccctt 180
aaaagtccga ttataactgc ttgcttaata cgatgaatca caaaaatacc gttagagctt 240
ctgttcgcga acttacccgg catgcattgc agaaacccgg ggaggagagt gaaccgggag 300
aaacaacaaa gactcttaag ttccgtgtgg atgccaaggg acatacccga acccgactcg 360
accaggcagc actgaacatt gctcggaaaa gcatctgggt cagttccctt ccaccccgtc 420
aggcaaagca atcggaaggt gaagaccggc gttccaccga ggaaccatgg cgcgtttggt 480
acaagaagca gatcatggga aagtggtctt aaaaaatccc ttgccacccg gctatcatag 540
gatcaaatta acccccttagt attttctttg cacaacgtcg actgccctgc caaggagcgg 600
ctgataacca acaagttctt ttacatttga ggtgcaaacc gttcttgtga ttttttttata 660
aaaggatgca cccgaaaggg ttggctatac ccaaggattc ccagccaccc atctgacgga 720
ctggtttgct gacctcg 737

<210> 207
<211> 1191
<212> DNA

<213> 未知
<220>
<223> 未知的描述:
陆生-土壤-沼地序列
<400> 207

[0065]

tgccgggcgg cggggaaagc gtatgatttg cgccgtcacc agaaaacaaa cgaaggcaca 60
aaatggggcg gaaaagaaaa ggcatcgggc agaaagagca agcccagccg gatttgggtc 120
cggccgggct tgtccaagaa acgcttttga acaggtgtat ccgcatggag gataagtcca 180
tattcagctt cccttgcttg accactttac tttacaacac ttcgaattta tgagcaaaga 240
ataccgaccg gattcgattc acgcctcgct tcgccatctg cttgaacttg gcttgagcaa 300
aaccggttcg agtgaaccgc aaaagatcac gcgcacagtt aagttcaagg tacacacgct 360
tgcccgaccg gaactggtgc aaggcgcatt gaacatcgcc cgcaagttcc ttttcgcctt 420
gggacgcggg aagaaatccg gcgagatgac cgaagcggag cggcgcaaag tcaggggcga 480
gtgggaaaaa tggtataagg aaaaactcgg aaccgtttgg agaaagtgat aagtctggat 540
tttcgcccat tgcgggcctg atatggaaaa cgatgatttt ctgcccatct actcgctggc 600
ggaatggcat tattgcccgc gctcggcgtt tctctcctgg ttcggcgctg aacgcgaaga 660
ccgcgtgacg ccggcttacc agaagatgcg tgaagtccat gccgtgagca accatgtggc 720
cttccgcggc gaactcgaac ggggcggttg gctgggcgga aaaaacgcgc tgctcattta 780
cccgcttcac gatgacgacc ttgaaaatgt ggagcggttc ggggcggtgc gcgtctggca 840
ggccacattt gaacttcttt aagaaaacgc ttgcgtctgt cctgctaatg gcgtaaaaag 900
gccccgttag tatttacttt gatacgatct ggcgtgcctt tccgcccaac ggattcgaca 960
cgacccacga acttcagcgt gtctgaggaa acctcaagcg gtggatgagc cgctgcttgt 1020
cgcgatcctt gataagtcaa acggagcgag tcggttaggt ttggaaaaag ccttgaaccg 1080
ttctcagcgg tttcctatgc gtaacttgcc taaaatcaac acgagtttgc gcgttagccg 1140
ctgcgacagc agcggctcac gcgcaaactc gtgtgatcca agggccgaag a 1191

<210> 208
<211> 947
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
水生-淡水-泥炭地序列
<400> 208

gcgcatcgcc gagttcggct tcttgggcgt gcgcgtcatg acctgaatgc acaccccgcg 60
tcgaaacggc gagccttcca aggctggcga cttggacttg tatttaactt tgcgtcggcc 120
ctggcggacc agttgattga tagtcggcat tgcgtcttaa aaatggaacg gggaatataa 180
caaaccaaat ccggcttgca acaactattt cgcccgattt gcggtcttaa ctcatcgtag 240
cggcggtcgg cagaccgccg catttcttct ggaaattgcg ctgtccggca ctttgccggg 300
agccgctgcg cccgttttgc agaagtctcg ttttaatcct tttggaaagc accgcgaagc 360
cagcttcaac taccgcttgc cgccgccggg acacgggcgt agcatcactc cacaatcagc 420
catcaggaaa atgggtatga aaatctcttt ggggatttca tcatgaaagc aaagtatcgg 480
tctaactcaa ttcatgcctc gctccggcag cttttggagc ttggcttaag caagagcagt 540
tcggcagaac cacagaaaat cacgcgcact gccaagttca agattaatac cgacatccgg 600
ccggatttgg ttttaaacat tgcccgaaag tttttgttcc gttccgaacg aggcaaaaag 660

actgacgacg caggcgaaga aggtgaggca aagcgtggta aatacaaaga cgactggcag 720
acttggtatc gccaaaagtt ggaaaaggtc tggtgcaagt aatctgacat cgcccattcg 780
ggacccgcga gcatgacatg gagactgacg atttccttcc tatctactcg ctcgcggagt 840
ggcattattg cccgcgctcg tcttttctgt cgtggtttgg tgcagagcgt gaggatcgcg 900
tcaccccggc ttaccagaaa atgcgcgaag tccacgcggt gagcaat 947

<210> 209
<211> 856
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
水生-淡水-淡水和沉积物序列
<400> 209

aaacccaacg gcataaagag gatcaggagt tttgggtgcc cttattatga gataaagaat 60
aagaataccc catacgatat tcatccattt aagcagacgt tgcacttggg atattctgtc 120
acctgcgttg accatttggc aacgtgccat tatcgatgac aagggagagg tataccccca 180
aatggatatg cctgttataa aacattagtt tggtgacttg gcatttttag ccacttttc 240
cctgtccaat tttaggggt tttgtgaatc cctattccaa ggactgtttt tgtcccacct 300
cctaatttac agttatcctt attatgaact ggaatagctc cgttaaaggt ccgattatga 360
cttcttgctt aatacgatga atcacaaaag caccgtcaag gcttctgttc gccatcttac 420
ccggcatgca ttgcagaaac ccggggagga gagtgatccg ggagaaacga caaagacttt 480
aaagttccgg gtagatgcca aagcacatac ccgaacccga ctcgaccagg cagcactgaa 540
[0066]
cattgcccgc aaaagcatct gggtcagctc ccttccaccc cgtcaggcaa agcaatcgga 600
aggtgaagac cggcgttcca ccgaggaacc atggcgcgtt tggtacaaaa agcagatcat 660
ggaaaagtgg tcttaaaaaa tcccttgcca cccggctatc ataggatcat tttagtctct 720
tagtattttc tttgcacaac gtcgactgcc ctgccaagga gcggctgaca accaacaagt 780
tcttttacat ttgaggtgca aaccgttctt gtgatttttt tacaaaagga tgcacccgaa 840
agggttggcc ataccc 856

<210> 210
<211> 823
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
土壤宏基因组序列
<400> 210

ctgaaattcc cggccatcaa tcccgccaac gccgccgagc gcgatgaatt tcgtctgaat 60
accttcccac cggacaacca caccgcgccc gcgccgccga ctccgcccgc gacaccgaag 120
ccgtaaaatc aaatttggaa aacgcaaagc ggcctttggg aatggaggcc gctttttgtt 180
ttggattgtt cctgttcacg aaatccgcta ttcaggctac aaatctaaaa gggtaaatat 240
gagcaaagag cctcggaaaa attcgactca cgcctcgctt cgccagttgc ttgaacttgg 300
tttgagcaaa agcagttcgg cagagccaca gaagatcacg cgcacggtga agttcaagat 360
tcacacgctt gcacgaccgg aactcgttgg cgcgttgaac atcgcccgca aattcctgtt 420
cgcctcggga cgcggaaaaa agtccggtga aatgaccgaa gccgagcggc gcaaagtgcg 480

gggcgagtgg gaaaagtggt ataaggaaaa acttgcgacc gtctggaaaa aacaatgaaa 540
ggcttgcgtc agttctgcga aaggcgtaaa caggcggcgt tagtatttac tttgatacga 600
tctagcgtgc ctttccgtcc aacggattcg acacgacccg tgaacttcag cgtgcaaaag 660
gaaatcacaa gtggtggttg accgctgcct gtagcgatcc ttgaaaactt aaaccgagcg 720
aatcagttag gttcggaaag aagcccgaac tggtttcaaa gtcaatcgcg tcaatgattt 780
cctgcaaacc ataaacgggt ctaaacgggt cagtcctaac ttt 823

<210> 211
<211> 1177
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
水生-淡水序列
<400> 211

[0067]

tgccccagcc cctaactcaa gttctgtaaa aacattattc gaaagaataa atccccgatc 60
agaggattgc ccaacaatag tgccttggtt gtttatatcg taacaggcgc tccacgtatg 120
ggcggaacta agcaccccca cggaactcat gcttttattt ctgtacacga acccccctgtc 180
accaagcgaa cttgaggagg caccaaccac aacacccgag tcgttaatgc tgtagcttcg 240
actggaatta ttcgtgtagg tgtactcgcc tccgaaccac gatgtccaag taacaacctg 300
aggaccaaga gtcctaagat ttaccagctt gtaagtcggg gcttgcgcat ttccgttttc 360
gcaaattacg agcgaagcaa ttaaaacgca gagccaccga accatatgat tattcaacac 420
ctattttact ttttttcaac tagggaaaac cggcttttta actactatcc cttttgtcct 480
aagcattaac ttttgcctca gaaactgtcc tgatcactaa actaatggtt tgtgaccggc 540
tgtgttttgg gcgttccgac tctgtgggtt tcattggggt tttgtgactg ccccctccta 600
ataccgtatt tgtaccaccc cctaccttag ggttggccct gttttgaact ggaatacgta 660
ccttaatcgt ctgattataa cccttcggtt aaaacgatga ttcataagag taccgttaaa 720
gcatcggttc gccacctgac ccgccatgcg ttggacaaac ccggcgaggg cgacccgggt 780
gaaacaacaa aaaccctaaa gttccggatt gatgccaaag atcacacccg cacccggctc 840
gaccagcagg cagccctaaa cattgcccga aaaagcatct gggtcagttc cttgccgcac 900
cggcaggcaa aacagtcgga gggagaagac cggcgttcca ctgaggagcc gtggcgggag 960
tggtacaaga agcggatctg ggagaagtgg gcttaaaaaa aagcccttgc cacccggtcc 1020
ccgcaggatc attttagtct cttagtattt tctttgcaca gcgttgactc ccctgccaag 1080
ggtgcggcca acaatccaac aagttctttc gcaacagggg gcaaaccccg ttcaattttg 1140
ctttaaggaa ctaaagccca tggcgttagg tatagct 1177

<210> 212
<211> 954
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
水生-淡水序列
<400> 212

aaatacggtt ccgtcaaata agttgcctga agagtcactg gcgttccctg aaaattcgta 60
ttttccaacc aagccataat tgaggttaac cacggcagac ttcgctgttg aacccatcaa 120

caagatggct aatgcgacca cccgattcat atattaatta aaccatgttc ttttgttttt 180
tcaatgtccc acaggctctt tttttgacta aataaaatat atccctttg gtatatcctt 240
ggttggtttc gtattatcgc ataaaacaaa tggtttgtga ccggtcatgt tttgagcgtt 300
ccatctctgt aggtttcatt ggggttttgt gattaccccc tcctaatacc gtatttgtac 360
caccccctac cttagggtta gccctgtttt gaactggaat acgtacctta atcgtctgat 420
tataaccctt cggttaaaac gatgattcat aagagtactg ttaaagcatc cgttcgccac 480
ctgacccgcc atgctctgga caaacccggc gagggcgatc cgggtaaacc aacaacaaaa 540
accctgaagt tccggattga tgccaagggg cacacccgca cccggctcga ccaagccgct 600
ttaaatatag ctcgcaagag tatctggctg agtactttgc ccaaaagatc caaaaaggaa 660
acactggctg aagcacgtgg gtccacaaaa ccagcgtgga aaaagtggta tcaaaatcgt 720
atcagcgaga attggtgtta aaaatacttg tcttacactc gaaaatatta taaaaagcag 780
tcttagtatt ttcttagcaa cacgttgtct gtcctcgcca aggaacggtc aacaatccaa 840
caagttcttt cgcaacgggg gggcaaacca gaaaattttc ccttaaattt gaatcttagg 900
actttagata ctgccatgcg cggatactac aaaaccaaat gcatattttt cgta 954

<210> 213
<211> 995
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
水生-淡水序列
<400> 213

[0068]

ggcagggtca aagtcatcat ggcttggaag ttggacagaa tcggccgctc gacccagcac 60
ttggccggaa tcatcaatcg actccttaag gcggatgttg gcctcgtggt tccctctcag 120
ggtattgata cctcggccgg cagcatgaat cctgccagta agctccagtt gaacgttctg 180
gccgcagtcg ctgagtttga gagggacctt atccgcgaga ggacgaaggc cggcatgaag 240
gcagctaagg ccagaggcgt taaattcggc cgtgaaacgg tggtaaccga agcccagcga 300
atcctagcca aagagattct tcagcttaat ccggatactt cggtgaccag cctgtcaaag 360
cggttgggct tgtctttggg aacggcgagc agacttaaga cgattttggg aggtagctgc 420
tagtaccgcc aatgtcctat ccacccgctt acgattgacc cagttttgaa ctggaatacg 480
caccttaatc gtctgattat aacccttcgg ttaacacgat gattcataag agtaccgtta 540
aagcatcggt tcgccacctg acccgccatg ctctggacaa acccggcgag ggcgatccgg 600
gtgaaacaac aaaaaccctg aagttccgca ttgatgccaa aggtcacacc cgcacccggc 660
tcgatcagca ggccgccctg aacatcgctc gtaaaagcat ctgggtcaga tcattgccgc 720
cccgacaggc aaagcagtcg gagggcgagg atcggcgttc cactgaggag ccatggaggg 780
agtggtataa gaagcagatc cgggagaagt gggcttaaaa aaacccttgc cacccggtcc 840
ccctaggatc atcttggtcc cttagtattt tctttgtaca gcgttgactc ccctgccaag 900
ggtgcggcca acaaccaaca agtttttttct caacgagggg tgttataacc ctgaaatttc 960
gacagggtgc tgaactgtag ggttttggcg ataac 995

<210> 214
<211> 1776
<212> DNA
<213> 未知
<220>

<223> 未知的描述:
水生-淡水-沼泽序列
<400> 214
caccacgccg tcttcgatgg cattgaacaa cgtctccaga aacttccgtt cccgggccaa 60
tcgttggaca accgtttgca acccttcggc atccaggcgt ccgatcctgc ccaggacttt 120
atcgagaaaa gtcgatatta ccgcagccat aatgatgata cccgccaatg cgcaaccacc 180
gcgccaattt gacgcaccca aagtattcaa gcaattctga ttaggcaagc attgctttca 240
atttgaaagg ccggaaaaaa ccgtgaaaaa ttcggggttt ttcaattaaa caagaaagtt 300
gtcaagggga aaattgcgca aaaaaaacag tcaaaatttc ggcatttttg acctaagcaa 360
ttgcaaacca aacacttacg aatttccacc tgcttgctgc ccatgactgc aaaatggtag 420
cttgggactg atggacgaat gcgactgatg gtctggcaac cagccagtgg tcaaggcggg 480
tttctgccag atggacagat aggattgtaa tcataagacg attcgggcaa aaccagggcg 540
gacagggaaa gcaggaattg gacggtgggg cgagattcca tcgagccgta aaatcgagtc 600
ggagcacagc ccgtcacgtt gcaggattaa tggtgaaaat gggtcttggc aggacgttct 660
gtttgggcca taatcttgcc cgctccacaa aacgaaggga aaaagaagat gagtgcggaa 720
acttcagagg ctggcaaaac catgtgggct ttcgacttgg gcacagactg cgaattgcgc 780
gtcgaatgcc gcaaaccatc gaattgaaac tatttatgga ttcaatccac aaatccgttc 840
gttcgctgct atcgcttggc ctgtcgcgta gcgcttcggc aagtccgcaa acaattacgc 900
gcacagtaaa attcaagatc aataccgaca ttcggcaagc cgcgctcaat atcgccagaa 960
agcatctgtt tcgcattggc cgcgaaaaaa aatcagatga ggagaatgaa gccgcacgac 1020
gcaaagttat gggtgagtgg gaaacttggt atcggcagaa actcgccagc gtttggaata 1080
aataattcac ccttagccat tttgactgtg agcgagatat ggagatcgac gatttccttc 1140
caatctattc gctcgcggag tggcattatt gcccacgctc gtcttttctg tcgtggttcg 1200
gcgcggaacg tgaggatcgc gtcaccccgg cttaccagaa gatgcgtgaa gtccatgccg 1260
tgagcaatca tgtggcgttt cggggcgcat tggaacgcgg cggctggctg gatggcaaac 1320
acgcgctgct ggtttatccc ttgcatgatg acgaccttga aaatgtggaa cgttacggcg 1380
cggtacgggt gtggcaggca acgtttgaat ttctttgaat ttaggcttgc agttggatgg 1440
gagcaggcat aatacttctc ccgttagctg tctcctcgga ggcagattta ctttgattct 1500
gtgcgccttt acgccccacg tatttgacgc acttcgcgaa cttcagcggt taacggaaac 1560
cgccagcggt ggatggacga ttcacccggc gttccttgac aacacagtcc gtagcagttg 1620
gttatatacg aaacggacta aaagcggttt gccatgcgta acttgttgcc agacaatgga 1680
gtcatcatct cggcagaaat tttgcattta ggtcttgatg caaatatttg gacggcaggg 1740
cgtgaagaca gatcagcatc gcgcttggcc aggccc 1776
<210> 215
<211> 1440
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
水生-淡水序列
<400> 215
ctctctggat tgggttggtc ttcgccacga tggtctactt tcacggcaac gtgaaagact 60
tggaacccaa acccggttcg gtcacaccga tccagaaaac ggccgcgggg ctgatgtttc 120
ttgtcgatgc gatcatgctg ggactcgctc tttatttgat ccttcgctaa cgcgaaccga 180

[0069]

tcgataagga ttaagtcgca ctacccctgc atcgattccc gttttttcga ttcccttcca 240
ggtatttaat ttttacttct caacgaggta tttggcttgg catggtcttc tctgcgcccc 300
ctgccccctg tcccctgccc cctttcactt ccagcctacc aattcccagt gttcaaaacg 360
tcggtgtttt ggcgtttatc ctttcctgag ataacgccca ggactgtttt tgtcccacct 420
cctaatttac acttatcctt attatgaact ggaatagctc cgttaaaagt ccgattatga 480
ctgcttgctt aatacgatga atcacaaaag cactgtcaaa gcctctgttc gccatcttac 540
tcggcatgca ttgcagaaac ccggggagga gagtgatccg ggagaaacga caaagacttt 600
aaagttccgg gtagatgcca aaggacatac ccgaacccga ctcgacaccc ggggtggtcg 660
gggaacattg atgggcaaca tgcactatga gcgaattttg tctgggttga gggcaagttg 720
gccgaggcag gttgctacaa ggcacccagt gccgccaagt accgcaaagg aatcacgccg 780
ttcttgcgat ttgtaaatcc agcatatacc agtgccacct gttcagcttg tggtgaggtg 840
ttcagttcta gttggtataa gaaaaccgca gccacaatca gtaaggaatc acccggttca 900
tggtcagtga cgcttcccaa cggaaaaaaa cttcagttgc catgctccta cacctcatac 960
atgcgtgggc gcggtgaaaa gacatggaac accgctgaac ggctggagga acttttcaag 1020
gatcgaacat ggtcagatct tccccccagc acccgaaaaa cagtagaggg attggttgct 1080
cgcctgactc atcatcgccc caaacggggt caggcagcac tgaacattgc ccgcaaaagc 1140
atctgggtca gctcccttcc accccgtcag gcaaagcaat cggaaggtga agaccggcgt 1200
tccaccgagg aaccatggcg cgtttggtac aaaaagcaga tcatggaaaa gtggtcttaa 1260
aaaatccctt gccacccggc tatcatagga tcattttagt ctcttagtat tttctttgca 1320
caacgtcgac tgccctgcca aggagcggct gacaaccaac aagttctttt acatttgagg 1380
tgcaaaccgt tcttgtgatt tttttacaaa aggatgcacc cgaaagggtt ggccataccc 1440

[0070] <210> 216
<400> 216
000
<210> 217
<400> 217
000
<210> 218
<400> 218
000
<210> 219
<400> 219
000
<210> 220
<400> 220
000
<210> 221
<400> 221
000
<210> 222
<400> 222
000
<210> 223
<400> 223

000
<210> 224
<400> 224
000
<210> 225
<400> 225
000
<210> 226
<400> 226
000
<210> 227
<400> 227
000
<210> 228
<400> 228
000
<210> 229
<400> 229
000
<210> 230
<400> 230
000

[0071]

<210> 231
<400> 231
000
<210> 232
<400> 232
000
<210> 233
<400> 233
000
<210> 234
<400> 234
000
<210> 235
<400> 235
000
<210> 236
<400> 236
000
<210> 237
<400> 237
000
<210> 238

[0072]

<210> 238
<400> 238
000
<210> 239
<400> 239
000
<210> 240
<400> 240
000
<210> 241
<400> 241
000
<210> 242
<400> 242
000
<210> 243
<400> 243
000
<210> 244
<400> 244
000
<210> 245
<400> 245
000
<210> 246
<400> 246
000
<210> 247
<400> 247
000
<210> 248
<400> 248
000
<210> 249
<400> 249
000
<210> 250
<400> 250
000
<210> 251
<400> 251
000
<210> 252
<400> 252
000

<210> 253
<400> 253
000
<210> 254
<400> 254
000

<210> 255
<400> 255
000
<210> 256
<400> 256
000
<210> 257
<400> 257
000
<210> 258
<400> 258
000
<210> 259
<400> 259
000

[0073]

<210> 260
<400> 260
000
<210> 261
<400> 261
000
<210> 262
<400> 262
000
<210> 263
<400> 263
000
<210> 264
<400> 264
000
<210> 265
<400> 265
000
<210> 266
<400> 266
000
<210> 267

[0074]

<400> 267
000
<210> 268
<400> 268
000
<210> 269
<400> 269
000
<210> 270
<400> 270
000
<210> 271
<400> 271
000
<210> 272
<400> 272
000
<210> 273
<400> 273
000
<210> 274
<400> 274
000
<210> 275
<400> 275
000
<210> 276
<400> 276
000
<210> 277
<400> 277
000
<210> 278
<400> 278
000
<210> 279
<400> 279
000
<210> 280
<400> 280
000
<210> 281
<400> 281
000

<210> 282
<400> 282
000
<210> 283
<400> 283
000
<210> 284
<400> 284
000
<210> 285
<400> 285
000
<210> 286
<400> 286
000
<210> 287
<400> 287
000
<210> 288
<400> 288
000
[0075] <210> 289
<400> 289
000
<210> 290
<400> 290
000
<210> 291
<400> 291
000
<210> 292
<400> 292
000
<210> 293
<400> 293
000
<210> 294
<400> 294
000
<210> 295
<400> 295
000
<210> 296
<400> 296

000
<210> 297
<400> 297
000
<210> 298
<400> 298
000
<210> 299
<400> 299
000
<210> 300
<400> 300
000
<210> 301
<211> 674
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
生物反应器宏基因组或实验室富集限定的培养基-离子液体和高固体富集的序列
<400> 301

[0076]

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

Arg Gly Arg Trp Arg Arg Trp Leu Asp Phe Leu Ser Ser His Pro Glu
180 185 190
Leu Ala Ala Trp Arg Gly Gly Pro Ala Thr Ile Val Pro Leu Ser Lys
195 200 205
Glu Glu Leu Ala Ala Cys Arg Arg Arg Arg Arg Arg Ala Val Ala Glu
210 215 220
Gln Ala Arg Leu Phe Phe Glu Lys Asn Pro Glu Leu Arg Ala Leu Asn
225 230 235 240
Arg Glu His Asp Glu Tyr Gln Glu Lys Phe Ala Arg Pro Trp Ala Lys
245 250 255
Arg Arg Asn Pro Asp Gly Phe Arg His Arg Pro Thr Phe Thr Leu Pro
260 265 270
His Pro Asp Ser His Pro Asp Trp Pro Arg Phe Lys Gly Asn Ala Gly
275 280 285

Trp Arg Asp Leu Ser Val Pro Asp Arg Thr Val Ser Leu Glu Leu Leu
290 295 300
Cys Pro Asp Gly Ser Arg Arg Trp Phe Asn Val Arg Phe Val Pro Asp
305 310 315 320
Pro Arg Leu Ala Ser Leu Arg Pro Ala Thr Gln Pro Val Arg Gln Gly
325 330 335
Asn Thr Thr Tyr His Tyr Val Ser Asp Ala Leu Gly Val Ala Glu Gln
340 345 350
Pro Val Gln Pro Gln Gly Ile Lys Leu Ile Leu Arg Pro Asp Gly Ala
355 360 365
Tyr Leu Ser Val Ser Val Ser Val Leu Ala Pro Pro Cys Gly Phe Asp
370 375 380
Val Arg Gln Ser Ala Ile Glu Lys Tyr Gly Ser Ile Trp Ala Cys Arg
385 390 395 400
Lys Tyr Lys Gln Asp His Pro Asp Arg Pro Leu Val Thr Cys Ala Val
405 410 415
Asp Leu Ala Val Arg His Asn Gly Ala Leu Thr Val Ala Arg Asp Gly
420 425 430
Glu Val Phe Ala Arg Arg Leu Leu His Asn Arg Phe Phe Leu Pro Gly
435 440 445
Asp Asp Arg Glu Arg Ala Met Asn Ile Pro Ser Leu Leu Glu Ile Ala
450 455 460
Ala Val Lys Arg Gln Leu Arg Arg Ala Arg Arg Gln Thr Gly Arg Ile
465 470 475 480
Ala Pro Gly Lys Pro Ser Cys Ala Arg Leu Gln Ala His Tyr Arg Ala
485 490 495
Leu Cys Glu Asp Arg Tyr Lys Lys Leu Val Ala Ala Ile Phe Ala Tyr
500 505 510
Ala Arg His His Gly Ala Glu Val Val Ile Phe Glu Asp Leu Lys Asn

[0077]

515 520 525
Leu Leu Pro Asn Ala Val Asn Glu Arg Gly Val Asn Arg Ala Leu Gln
530 535 540
Ser Trp Asn Arg Gly Ala Ile Val Arg Phe Ala Arg Gln Ser Ala Glu
545 550 555 560
Asp Val Gly Leu Arg Val Ala Thr Val Pro Ala Phe His Thr Ser Ile
565 570 575
Val Cys Ala Arg Cys Gly Ala Leu Gly Cys Arg Phe Glu Gln Thr Arg
580 585 590
Glu Gly Val Arg Ile Ala Lys Leu Gly Pro Trp Phe Gly Cys Arg Ala
595 600 605
Cys Gly Arg Arg Ile His Ala Asp Leu Asn Ala Ser Glu Asn Leu His
610 615 620
Lys Val Leu Ile Gly Thr Phe Pro Gln Val Arg Arg Ile Ser Arg Asp
625 630 635 640
Pro Pro Val Leu Ser Ile Ala Asp Val Arg Val Asp Gln Arg Glu Ala
645 650 655
Asp Arg Leu Ala Ala Ala Trp Leu Ser Glu Leu Ser Val Gly Pro Thr
660 665 670
Pro Phe

[0078]

<210> 302
<211> 759
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
激活的污泥宏基因组序列
<400> 302
Met Gln Leu Trp Arg Thr His Cys Val Phe Asn Gln Arg Leu Pro Ala
1 5 10 15
Leu Leu Lys Arg Leu Phe Ala Met Arg Arg Gly Glu Val Gly Gly Asn
20 25 30
Glu Ala Gln Arg Gln Val Tyr Gln Arg Val Ala Gln Phe Val Leu Ala
35 40 45
Arg Asp Ala Lys Asp Ser Val Asp Leu Leu Asn Ala Val Ser Leu Arg
50 55 60
Lys Arg Ser Ala Asn Ser Ala Phe Lys Lys Lys Ala Thr Ile Ser Cys
65 70 75 80
Asn Gly Gln Ala Arg Glu Val Thr Gly Glu Glu Val Phe Ala Glu Ala
85 90 95
Val Ala Leu Ala Ser Lys Gly Val Phe Ala Tyr Asp Lys Asp Asp Met
100 105 110
Arg Ala Gly Leu Pro Asp Ser Leu Phe Gln Pro Leu Thr Arg Asp Ala

[0079]

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

[0080]

```
465                 470                 475                 480
Gln Thr His His Gly Val Arg Leu Leu Gln Ser Arg Tyr Val Ala Val
                485                 490                 495
Gly Gln Val Glu Ala Arg Ala Ser Gly Glu Ala Glu Trp Ser Pro Gly
            500                 505                 510
Pro Asp Leu Ala His Ile Ala Arg His Lys Arg Glu Ile Arg Arg Leu
        515                 520                 525
Arg Gln Leu Arg Gly Lys Pro Val Lys Gly Glu Arg Ser His Val Arg
    530                 535                 540
Leu Gln Ala His Ile Asp Arg Met Gly Glu Asp Arg Phe Lys Lys Ala
545                 550                 555                 560
Ala Arg Lys Ile Val Asn Glu Ala Leu Arg Gly Ser Asn Pro Ala Ala
                565                 570                 575
Gly Asp Pro Tyr Thr Arg Ala Asp Val Leu Leu Tyr Glu Ser Leu Glu
            580                 585                 590
Thr Leu Leu Pro Asp Ala Glu Arg Glu Arg Gly Ile Asn Arg Ala Leu
        595                 600                 605
Leu Arg Trp Asn Arg Ala Lys Leu Ile Glu His Leu Lys Arg Met Cys
    610                 615                 620
Asp Asp Ala Gly Ile Arg His Phe Pro Val Ser Pro Phe Gly Thr Ser
625                 630                 635                 640
Gln Val Cys Ser Lys Cys Gly Ala Leu Gly Arg Arg Tyr Ser Leu Ala
                645                 650                 655
Arg Glu Asn Gly Arg Ala Val Ile Arg Phe Gly Trp Val Glu Arg Leu
            660                 665                 670
Phe Ala Cys Pro Asn Pro Glu Cys Pro Gly Arg Arg Pro Asp Arg Pro
        675                 680                 685
Asp Arg Pro Phe Thr Cys Asn Ser Asp His Asn Ala Ser Val Asn Leu
    690                 695                 700
His Arg Val Phe Ala Leu Gly Asp Gln Ala Val Ala Ala Phe Arg Ala
705                 710                 715                 720
Leu Pro Pro Arg Asp Ser Thr Ala Arg Thr Leu Ala Val Lys Arg Val
                725                 730                 735
Glu Asp Thr Leu Arg Pro Gln Leu Met Arg Leu His Lys Leu Ala Asp
            740                 745                 750
Ala Gly Val Asp Ser Pro Phe
        755
<210> 303
<211> 759
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
      激活的污泥宏基因组序列
```

<400> 303

Met Gln Leu Trp Arg Thr His Cys Val Phe Asn Gln Arg Leu Pro Ala
1 5 10 15
Leu Leu Lys Arg Leu Phe Ala Met Arg Arg Gly Glu Val Gly Gly Asn
20 25 30
Glu Ala Gln Arg Gln Val Tyr Gln Arg Val Ala Gln Phe Val Leu Ala
35 40 45
Arg Asp Ala Lys Asp Ser Val Asp Leu Leu Asn Ala Val Ser Leu Arg
50 55 60
Lys Arg Ser Ala Asn Ser Ala Phe Lys Lys Lys Ala Thr Ile Ser Cys
65 70 75 80
Asn Gly Gln Ala Arg Glu Val Thr Gly Glu Glu Val Phe Ala Glu Ala
85 90 95
Val Ala Leu Ala Ser Lys Gly Val Phe Ala Tyr Asp Lys Asp Asp Met
100 105 110
Arg Ala Gly Leu Pro Asp Ser Leu Phe Gln Pro Leu Thr Arg Asp Ala
115 120 125
Val Ala Cys Met Arg Ser His Glu Glu Leu Val Ala Thr Trp Lys Lys
130 135 140
Glu Tyr Gln Glu Trp Leu Asp Arg Lys Ser Glu Trp Glu Ala Glu Pro
145 150 155 160
[0081] Glu His Ala Leu Tyr Leu Asn Leu Arg Pro Lys Phe Glu Glu Gly Glu
165 170 175
Ala Ala Arg Gly Gly Arg Phe Arg Lys Arg Ala Glu Arg Asp His Ala
180 185 190
Tyr Leu Asp Trp Leu Glu Ala Asn Pro Gln Leu Ala Ala Trp Arg Gly
195 200 205
Lys Ala Pro Pro Ala Val Val Pro Ile Asp Glu Ala Gly Lys Arg Arg
210 215 220
Ile Ala Arg Ala Lys Ala Trp Lys Gln Ala Ser Val Arg Ala Glu Glu
225 230 235 240
Phe Trp Lys Arg Asn Pro Glu Leu His Ala Leu His Lys Ile His Val
245 250 255
Gln Tyr Leu Arg Glu Phe Val Arg Phe Lys Gln Arg Pro Thr Phe Thr
260 265 270
Met Pro Asp Pro Val His His Pro Arg Trp Cys Leu Phe Asn Ala Pro
275 280 285
Gln Thr Ser Pro Gln Gly Tyr Arg Leu Leu Arg Leu Pro Gln Ser Arg
290 295 300
Arg Thr Val Gly Ser Val Glu Leu Arg Leu Leu Thr Gly Pro Ser Asp
305 310 315 320
Gly Ala Gly Phe Pro Asp Ala Trp Val Asn Val Arg Phe Lys Ala Asp
325 330 335
Pro Arg Leu Ala Gln Leu Arg Pro Val Lys Val Pro Arg Thr Val Thr

[0082]

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

690 695 700

His Arg Val Phe Ala Leu Gly Asp Gln Ala Val Ala Ala Phe Arg Ala
705 710 715 720

Leu Pro Pro Arg Asp Ser Thr Ala Arg Thr Leu Ala Gly Lys Arg Val
725 730 735

Glu Asp Thr Leu Arg Pro Gln Leu Met Arg Leu His Lys Leu Ala Asp
740 745 750

Ala Gly Val Asp Ser Pro Phe
755

<210> 304
<211> 758
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
激活的污泥宏基因组序列
<400> 304

[0083]

Met Ser Val Arg Ala Ile Gln Ala Lys Val Met Cys Asp Arg Ala Thr
1 5 10 15

Leu Glu Asn Leu Trp Arg Thr His Arg Val Phe Asn Glu Arg Val Pro
20 25 30

Arg Ile Leu Ser Leu Leu Phe Arg Met Arg Arg Gly Glu Cys Gly Lys
35 40 45

Thr Pro Ala Glu Arg Gln Leu Tyr Gln Arg Ile Gly Gln Phe Ile Thr
50 55 60

Asn Cys Ser Ala Gln Asn Ala Asp Tyr Leu Met Asn Ala Val Thr Met
65 70 75 80

Arg Gly Trp Lys Pro Gly Ser Ala Lys Lys Tyr Lys Ile Arg Val Lys
85 90 95

Asp Glu Asn Gly Glu Thr Val Glu Ile Ser Gly Glu Ser Trp Ala Asp
100 105 110

Glu Ala Ala Ala Leu Ser Ala Lys Gly Ser Ile Leu Cys Asp Lys Glu
115 120 125

Asp Leu Ala Gly Asp Leu Pro Gly Cys Met Arg Gln Met Leu Asn Arg
130 135 140

Glu Ala Val Ala Ile Ile Ser Gly His Asp Glu Leu Val Ala Asn Trp
145 150 155 160

Ser Thr Asp His Val Gln Trp Leu Ala Glu Lys Ala Lys Trp Glu Arg
165 170 175

Glu Asp Glu His Lys Gln Tyr Leu Ala Leu Arg Ser Lys Phe Asp Glu
180 185 190

Phe Glu Lys Ser Ile Gly Gly Lys Val Gly Lys Arg Arg Gly Arg Trp
195 200 205

His Leu Tyr Ile Asp Trp Leu Arg Ala Asn Pro Gln Leu Ala Ala Trp

[0084]

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

565 570 575
Asp Gly Thr Val Ala Thr Lys Gly Ala Asp Pro Phe Pro Arg Ala Asp
580 585 590
Val Ile Leu Ile Glu Lys Leu Glu Gly Phe Ile Pro Asp Ala Glu Lys
595 600 605
Glu Arg Gly Val Asn Arg Ala Leu Ala Ala Trp Asn Arg Gly Gln Leu
610 615 620
Val Asn Arg Leu Lys Glu Leu Ala Ile Asp Ala Gly Tyr Lys Ala Arg
625 630 635 640
Ile Phe Glu Ile Leu Pro His Gly Thr Ser Gln Val Cys Ser Arg Cys
645 650 655
Gly Ala Met Gly Arg Arg Tyr Ser Ile Ala Arg Asp Glu Glu Ser Gly
660 665 670
Arg Pro Glu Ile Arg Phe Gly Trp Val Glu Lys Leu Phe Ala Cys Pro
675 680 685
Lys Cys His Phe Ser Ala Asn Ser Asp His Asn Ala Ser Val Asn Leu
690 695 700
His Arg Lys Phe Leu Leu Gly Asp Ala Ala Val Ala Ala Phe Phe Asp
705 710 715 720
Trp Lys Asn Leu Pro Glu Glu Lys Arg Arg Gly Glu Leu Asp Ala Ile
725 730 735
[0085] Asp Glu Ala Leu Arg Gly Tyr Leu Arg Lys Met His Lys Leu Gln Ala
740 745 750
Glu Ile Glu Thr Pro Phe
755

<210> 305
<211> 728
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水-淡水湖或淡水宏基因组
序列
<400> 305
Met Lys Leu Leu Phe Lys Leu Arg Arg Gly Lys Ile Gly Lys Asn Lys
1 5 10 15
Gln Gln Arg Asp Asp Tyr Lys Arg Trp Phe Glu Tyr Val Ile Tyr Asn
20 25 30
Gly Ala Lys Asn Ala Asp Tyr Asn Leu Asn Ala Leu Ser Ile Asn Asn
35 40 45
Trp Thr Pro Asn Thr Ala Lys Lys Tyr Ile Thr Lys Leu Met Ala Gly
50 55 60
Lys Asn Lys Ile His Ala Lys Gly Leu Asp Leu Val Val Ser Arg Ile
65 70 75 80

[0086]

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

[0087]

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

<210> 306
<211> 762
<212> PRT
<213> 未知
<220>
<223> 未知的描述:

水生-淡水-淡水湖或淡水宏基因组

序列

<400> 306

[0088]

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

[0089]

325 330 335
Ala Gly Ser Phe Ala Leu Asn Leu Leu Thr Gly Asp Ile Ile Asp Gly
340 345 350
Lys Tyr Pro Lys Glu Trp Val Pro Phe Lys Phe Ile Ala Asp Lys Arg
355 360 365
Leu Ser Cys Phe Lys Gly Lys Ile Lys Val Val Asn Asn Lys Ile Arg
370 375 380
Thr Glu Tyr Ile Phe Glu Asp Pro His Leu Ser Lys Lys Arg Gln Ala
385 390 395 400
Thr Ile Gly Gly Ile Lys Ile Ile Leu Lys Asp Ile Lys Ile Asp Asn
405 410 415
Leu Gly His Leu Leu Ser Ala Thr Pro Tyr Val Val Phe Ala Cys Asp
420 425 430
Val Asp Asn Glu Pro Ile Ser Glu Gln Ala Lys Lys Leu Lys Ser Glu
435 440 445
Asp Thr Gly Glu Thr Gly Asn Ser Gly Leu Pro Ile Arg Lys Arg Ile
450 455 460
Leu Pro Glu Gly Cys Leu Ile Ser Cys Ala Val Asp Leu Gly Ile Arg
465 470 475 480
Asn Leu Gly Phe Ala Thr Leu Ala Glu Ser Gly Gly Ile Pro Ser Glu
485 490 495
Cys Thr His Gly Asp Ile Lys Leu Leu Arg Ser Tyr Asn Leu Leu Leu
500 505 510
Glu Asp Gly Pro Ser Leu Glu Lys Ile Ala Lys Gln Lys Lys Leu Leu
515 520 525
Gly Lys Phe Arg Arg Leu Arg Gly Lys Pro Ile Lys Gly Glu Lys Ser
530 535 540
His Val Phe Leu Gln Asn His Ile Thr Lys Met Gly Val Asp Arg Phe
545 550 555 560
Asn Lys Ala Ala Arg Leu Ile Val Asn Phe Ala Leu Asn Thr Glu Lys
565 570 575
Lys Lys Asn Val Gln Gly Ile Glu Tyr Pro Arg Ala Asp Val Ile Val
580 585 590
Leu Glu Asn Leu Ala Gly Leu Ile Pro Asp Ala Ala Lys Arg Lys Gly
595 600 605
Ile Asn Lys Ala Leu Met Asn Trp Asn Arg Gly His Leu Val Lys Arg
610 615 620
Ile Gln Glu Ile Ala His Asp Ser Gly Tyr Leu Tyr Arg Val Tyr Ile
625 630 635 640
Met Ser Pro Trp Gly Ser Ser Gln Val Cys Cys Lys Cys Lys Gly Leu
645 650 655
Gly Arg Arg Tyr Ser Ile Ile Lys Asp Glu Asn Gly Arg Gln Ile Gln
660 665 670
Phe Gly Asn Gly Glu Lys Leu Phe Ala Cys Ala Asn Cys Arg Tyr Met

675 680 685

Ala Asn Ala Asp His Asn Ala Ser Ile Asn Leu His His Ile Phe Lys
690 695 700

Asp Lys Asp Ala Val Gly Phe Tyr Lys Asp Tyr Ala Asn Lys Asn Glu
705 710 715 720

Glu Glu Lys Gln Lys Ile Ile Asp Asn Leu Glu Ala Ile Leu Arg Pro
725 730 735

His Leu Gln Lys Met His Phe Thr Thr Lys Lys Glu Arg Ser Leu Asp
740 745 750

Val Pro Glu His Pro Leu Asp Val Pro Phe
755 760

<210> 307
<211> 676
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水-淡水湖序列
<400> 307

[0090]

Met Val Asp Pro Glu Ala Arg Ala Ala Leu Trp Arg Thr His Gly Val
1 5 10 15

Tyr Asn Glu Cys Leu Arg Phe Ala Leu Arg His Met His Lys Met Lys
20 25 30

Arg Gly Glu Ala Asp Pro Arg Tyr Ala Glu Ile Phe Gly Ala Ile Arg
35 40 45

Asn Ser Gln His Ala Thr Ala Val Met Glu Ala Val Thr Asp Pro Arg
50 55 60

Trp Lys Pro Lys Lys Asp Pro Lys Pro Trp His Ala Pro Ala Ala Ala
65 70 75 80

Leu Ile Ala Glu Gly Lys Leu Leu Phe Asp Arg Trp Lys Glu Leu Pro
85 90 95

Gly Leu Pro Gly Arg Phe His Arg Lys Leu Trp Glu Gly Ala Tyr Gln
100 105 110

Ala Met Asp Ser His Lys Glu Leu Met Ala Ile Trp Ala Ser Ala His
115 120 125

Glu Glu Trp Gln Lys Ala Arg Ala Glu Trp Glu Ala Glu Asn Pro Glu
130 135 140

Tyr Met Lys Leu Arg Pro Ala Leu Glu Ala Phe Glu Ala Glu His Gly
145 150 155 160

Gln Ala Ala Lys Arg Arg Gln Arg Trp His Lys Trp Leu Ala Phe Leu
165 170 175

Gln Ala Asn Pro Gly Leu Ala Ala Trp Arg Gly Gly Ala Asp Val Val
180 185 190

Asn Pro Ile Ser Asp Asp Gly Lys Glu Arg Leu Arg Arg Ala Lys Arg

[0091]

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

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

<210> 308
<211> 678
<212> PRT
<213> 未知
<220>

[0092]

<223> 未知的描述:
水生-淡水-淡水湖序列
<400> 308
Met Asp Met Val Asp Pro Glu Ala Arg Ala Ala Leu Trp Arg Thr His
1 5 10 15
Gly Val Tyr Asn Glu Cys Leu Arg Phe Ala Leu Arg His Met His Lys
20 25 30
Met Lys Arg Gly Glu Ala Asp Pro Arg Tyr Ala Glu Ile Phe Gly Ala
35 40 45
Ile Arg Asn Ser Gln His Ala Thr Ala Val Met Glu Ala Val Thr Asp
50 55 60
Pro Arg Trp Lys Pro Lys Lys Asp Pro Lys Pro Trp His Ala Pro Ala
65 70 75 80
Ala Ala Leu Ile Ala Glu Gly Lys Leu Leu Phe Asp Arg Trp Lys Glu
85 90 95
Leu Pro Gly Leu Pro Gly Arg Phe His Arg Lys Leu Trp Glu Gly Ala
100 105 110
Tyr Gln Ala Met Asp Ser His Lys Glu Leu Met Ala Ile Trp Ala Ser
115 120 125
Ala His Glu Glu Trp Gln Lys Ala Arg Ala Glu Trp Glu Ala Glu Asn
130 135 140
Pro Glu Tyr Met Lys Leu Arg Pro Ala Leu Glu Ala Phe Glu Ala Glu

145 150 155 160
His Gly Gln Ala Ala Lys Arg Arg Gln Arg Trp His Lys Trp Leu Ala
165 170 175
Phe Leu Gln Ala Asn Pro Gly Leu Ala Ala Trp Arg Gly Gly Ala Asp
180 185 190
Val Val Asn Pro Ile Ser Asp Asp Gly Lys Glu Arg Leu Arg Arg Ala
195 200 205
Lys Arg Ser Lys Arg Asn Lys Val Glu Ser Glu Glu Phe Phe Asn Ala
210 215 220
Asn Pro Glu Leu Lys Glu Leu Asp Lys Leu His Gly Tyr Tyr Gln Arg
225 230 235 240
Glu Phe Val Arg Pro Trp Ala Lys Lys Arg Asn Pro Asp Gly Phe Arg
245 250 255
His Arg Pro Thr Phe Thr Glu Pro Ser Pro Glu Lys His Pro Phe Trp
260 265 270
Tyr Gln Phe Lys Lys Gly Glu Thr Tyr Lys Ala Leu Asp Leu Gln Ala
275 280 285
Gly Thr Ile Ser Leu Gln Leu Leu Ala Pro Glu Gly Ala Lys Arg Ala
290 295 300
Ala Leu Trp Thr Thr Phe Ala Phe Arg Pro Asp Pro Arg Leu Arg Leu
305 310 315 320

[0093]

Ile Gln Lys Val Asn Pro Pro Leu Asn Glu Lys Phe Thr Tyr Ala Phe
325 330 335
Ala Asp Pro Ala Tyr Arg Gly Gln Pro Ala Arg Pro Ala Glu Ile Arg
340 345 350
Gly Ala Lys Leu Ile Phe Arg Pro Ala Gln Pro Asp Gly Ala Val Tyr
355 360 365
Leu Ser Phe Thr Cys Asp Met Pro Asp Leu His Cys Arg Leu Ser Val
370 375 380
Gly Gln Asp Ser Ile Asp Arg Tyr Thr Thr Lys Gly Thr Lys Trp Leu
385 390 395 400
Val Gly Arg Leu Arg Glu Glu Pro Glu Leu Ala Gly His Glu Pro Val
405 410 415
Thr Cys Ala Val Asp Leu Gly Leu Arg His Leu Gly Ala Ala Thr Val
420 425 430
Arg Arg Asp Gly Lys Ile Val Arg Ala Arg Leu Leu Arg Glu Gln Gly
435 440 445
Asp Pro Gly Lys Gly Pro Gly Leu Pro Ala Ile Ala Gly His Lys Arg
450 455 460
Ala Leu Ala Lys Gly Arg Arg Lys Arg Gly Lys Pro Ile Ala Gly Glu
465 470 475 480
Glu Ser Phe Val Glu Leu Gln Arg His Thr Asp Lys Met Gly Glu Asp
485 490 495
Arg Phe Lys Lys Gly Ala Arg Arg Ile Val Ala Phe Ala His Glu Asn

500 505 510
Gly Ala Asp Leu Ile Ile Leu Glu Lys Leu Ala Gly Leu Val Pro Asp
515 520 525
Ala Glu Lys Glu Arg Gly Ile Asn Arg Ala Leu Val Ser Trp Asn Arg
530 535 540
Ala Asn Leu Glu Lys Trp Val Lys Gln Leu Ala Asn Asp Ala Gly Met
545 550 555 560
Arg Val Val Glu Val Phe Pro His Trp Thr Ser Gln Leu Cys Ser Arg
565 570 575
Cys Gly Ser Met Gly Ala Arg Phe Ser Ala Lys Glu Gly Leu Pro Ala
580 585 590
Phe Glu Pro Val Gly Lys Leu Phe Ala Cys Pro Glu Cys Gly Tyr Ile
595 600 605
Ala Asn Ala Asp His Asn Ala Ser Val Asn Leu His Lys Arg Phe Tyr
610 615 620
Gly Glu Leu Pro Glu Val Glu Arg Pro Asn Lys Asp Lys Arg Val Phe
625 630 635 640
Arg Leu Thr Pro Arg Gly Gln Pro Thr Val Glu Val Asp Met Ala Asp
645 650 655
Val Glu Lys Arg Ile Leu Ala Arg Gly Arg Ala Glu Asp Leu Cys Arg
660 665 670
Gly Lys Pro Thr Pro Phe
675

[0094]

<210> 309
<211> 687
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水-淡水湖序列
<400> 309
Met Ser Thr Lys Ala Phe Thr Ala Arg Leu Ile Val Glu Ser Pro Glu
1 5 10 15
Ala Arg Asp Ala Leu Trp Arg Thr His Arg Leu Tyr Asn Glu Cys Leu
20 25 30
Arg Phe Val Leu Arg His Met His Lys Met Lys Arg Gly Glu Ala Asp
35 40 45
Pro Arg Tyr Gly Glu Ile Phe Gly Ala Ile Arg Asn Ser Gln His Ala
50 55 60
Thr Ala Val Met Glu Ala Val Thr Asp Leu Arg Trp Lys Pro Lys Lys
65 70 75 80
Asp Pro Lys Pro Trp His Thr Pro Ala Ala Ala Leu Ile Ala Glu Gly
85 90 95
Lys Leu Leu Phe Asp Arg Trp Lys Glu Leu Pro Gly Leu Pro Gly Arg

[0095]

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

450 455 460

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

[0096]

<210> 310
<211> 779
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水-淡水微生物垫序列
<400> 310
Met Glu Cys Asn Arg Glu Glu Leu Gln His Leu Trp Arg Thr His Cys
1 5 10 15
Ala Phe Asn Glu Arg Leu Pro Ser Ile Leu Ser Val Leu Phe Lys Met
20 25 30
Arg Arg Gly Glu Trp Gly Asp Thr Glu Gln Ile Lys Lys Lys Met Gln
35 40 45
Ala Phe Ala Trp Gly Val Leu Gln Leu Glu Ala Gln Asn Ser Tyr Asp

[0097]

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

[0098]

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

755 760 765
Val Glu His Lys Leu Asp Gly Pro Thr Pro Phe
770 775
<210> 311
<211> 761
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水-沉积物序列
<400> 311
Met Ser Ile Arg Ala Ile Gln Ala Arg Val Glu Cys Asp Arg Pro Thr
1 5 10 15
Leu Glu Arg Leu Trp Leu Thr His Arg Ile Phe Asn Glu Arg Leu Pro
20 25 30
Val Ile Leu Ser Phe Leu Phe Arg Met Arg Arg Gly Glu Cys Gly Asp
35 40 45
Thr Val Val Glu Arg Lys Met Tyr His Glu Ile Ala Asp Phe Ile Leu
50 55 60
Thr His Ser Gln Glu Ala Pro Tyr Leu Phe Asn Ala Val Ser Ile Lys
65 70 75 80
Gly Trp Lys Pro Gly Thr Ala Val Lys Gln Ser Pro Gly Ser Glu Trp
85 90 95
Pro Ala Gln Ala Ala Gln Leu Ser Lys Glu Gly Lys Leu Leu Tyr Asp
100 105 110
Lys Ser Asn Val Leu Gly Asp Leu Pro Gly Thr Leu Gln Gln Met Val
115 120 125
Cys Arg Glu Ser Ala Ala Ile Ile Ser Gly His His Glu Leu Val Arg
130 135 140
Met Trp Glu Ala Glu His Ala Gln Trp Leu Glu Asp Lys Ser Ala Trp
145 150 155 160
Glu Ala Asp Ala Glu His Gln Lys Tyr Leu Ala Leu Arg Pro Arg Phe
165 170 175
Asp Ala Phe Glu Gln Glu Ala Gly Gly Pro Ala Gly Lys Arg Arg Gly
180 185 190
Arg Trp His Leu Tyr Leu Asn Trp Leu Lys Gln Asn Pro Asp Leu Ala
195 200 205
Ala Trp Arg Gly Gly Lys Pro Glu Val Ile Glu Leu Ser Pro Ala Ala
210 215 220
Arg Arg Arg Val Glu Arg Ala Lys Pro Asn Arg Gln Arg Ser Val Glu
225 230 235 240
Ala Glu Glu Phe Trp Lys Gly Asn Pro Glu Leu Ser Ala Leu Asp Arg
245 250 255
Leu His Gly Tyr Tyr Glu Arg Glu Phe Val Arg Arg Arg Lys Thr Lys

[0099]

[0100]

```
                260                 265                 270
Lys Asn Ala Asp Gly Phe Lys His Arg Pro Thr Phe Thr Leu Pro His
        275                 280                 285
Pro Arg Ile His Pro Arg Trp Phe Val Phe Asn Ala Pro Gln Thr Gln
    290                 295                 300
Pro Ser Gly Tyr Ala Asp Leu Val Leu Pro Thr Lys Pro Gly Glu Glu
305                 310                 315                 320
Gly Ser Val Arg Met Leu Val Ile Asn Ala Glu Lys Ala Asn Gly Lys
                325                 330                 335
Tyr Pro Ser Glu Trp Val Arg Ile Arg Phe His Gly Asp Pro Arg Leu
            340                 345                 350
Cys Asn Phe Arg Ala Val Ser Ser Thr Arg Thr Ile Ser Arg Gly Lys
        355                 360                 365
Asp Lys Gly Lys Thr Lys Glu Ser Pro Ser Tyr Val Phe Arg Asp Arg
    370                 375                 380
His Leu Gly Gln Asp Arg Pro Ala Thr Ile Ser Gly Ile Lys Leu Ile
385                 390                 395                 400
Phe Arg Thr Asn Thr Asp Arg Thr Pro Lys Ala Ala Tyr Leu Tyr Phe
                405                 410                 415
Ala Cys Thr Ile Asp Asp Leu Pro Gln Ser Asp Ala Ala Lys Ser Ile
            420                 425                 430
Gln Trp Thr Glu Thr Gly Glu Ile Thr Ser Lys Gly Lys Lys Arg Lys
        435                 440                 445
Asn Met Lys Val Pro Asp Gly Leu Val Gly Cys Ala Val Asp Leu Gly
    450                 455                 460
Ile Arg Asn Leu Gly Phe Ala Thr Leu Ala Arg Tyr Ala Asp Gly Gln
465                 470                 475                 480
Pro Gln Ile Leu Arg Ser Arg Asn Ile Trp Val Arg Arg Gln Glu Ala
                485                 490                 495
Lys Gly Ala His Thr Gly Arg Trp Ser Ala Gly Ala Asp Leu Ala His
            500                 505                 510
Ile Gly Glu His Lys Arg Thr Ile Arg Arg Leu Arg Arg Gln Arg Gly
        515                 520                 525
Lys Pro Val Lys Asp Glu Lys Ser His Val Gly Leu Gln Leu His Ile
    530                 535                 540
Asp His Met Gly Glu Asp Arg Phe Lys Gln Ala Ala Arg Ala Ile Ile
545                 550                 555                 560
Asn Phe Ala Leu Asn Ala Asp Gly Asp Cys Asn Arg Lys Thr Gly Glu
                565                 570                 575
Val His Pro Arg Ala Asp Val Leu Val Leu Glu Asn Leu Gly Gly Leu
            580                 585                 590
Ile Pro Asp Ala Glu Lys Glu His Gly Ile Asn Arg Ala Leu Val Asn
        595                 600                 605
Trp Asn Arg Gly Gln Leu Val Glu Arg Ile Lys Glu Met Ala Gln Asp
```

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

[0101]

<210> 312
<211> 654
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水-沉积物序列
<400> 312
Met Asp Gly Asp Thr Trp Ala Asn Glu Ala Ala Ala Leu Ser Ala Gln
1 5 10 15
Gly Val Leu Leu Tyr Asp Lys His Ala Val Leu Gly Asp Leu Pro Ala
20 25 30
His Leu Arg Gln Met Val Cys Arg Glu Ser Val Ala Ile Leu Ser Gly
35 40 45
His Asp Glu Leu Val Ala Arg Trp Asn Lys Glu His Gln Glu Trp Leu
50 55 60
Thr Ala Lys Ala Lys Trp Glu Asp Glu Glu Glu His Lys Lys Tyr Leu
65 70 75 80
Ala Leu Arg Pro Arg Phe Glu Ala Phe Glu Lys Glu Ala Gly Gly Lys
85 90 95
Ala Gly Lys Arg Arg Gly Arg Trp His Leu Tyr Phe Ser Trp Leu Lys
100 105 110
Ala Asn Pro Asp Leu Ala Ala Trp Arg Gly Lys Ala Asp Ser Val Ala
115 120 125
Leu Leu Ser Asp Lys Ala Lys Glu Arg Ile Arg Lys Ala Arg Pro Trp

[0102]

```
    130                 135                 140
Lys Lys Arg Ser Val Glu Ala Glu Glu Phe Trp Lys Val Asn Pro Glu
145                 150                 155                 160
Leu Gln Ala Leu Asp Lys Leu His Gly Tyr Tyr Glu Arg Glu Phe Val
                165                 170                 175
Arg Arg Arg Lys Thr Gln Arg Asn Pro Asp Gly Phe His His Arg Pro
            180                 185                 190
Thr Phe Thr Leu Pro His Phe Thr Arg His Pro Arg Trp Phe Val Phe
        195                 200                 205
Asn Ala Pro Gln Thr Ser Pro Gln Gly Tyr Asp Asn Leu Gln Leu Pro
    210                 215                 220
Ala Lys Pro His Ala Gly Gly Ser Val Glu Leu Leu Leu Leu Thr Gly
225                 230                 235                 240
Asp Lys Thr Asn Gly Glu Tyr Pro Ser Gln Trp Ile Ser Ile Gly Phe
                245                 250                 255
Gln Gly Asp Pro Arg Leu Ser Leu Phe Arg Lys Val Lys Lys Gln Arg
            260                 265                 270
Thr Val Asn Arg Gly Lys Asp Lys Gly Thr Gln Lys Glu Ser Asp Ala
        275                 280                 285
Tyr Glu Phe Arg Asp Leu His Leu Asn Arg Trp Arg Lys Ala Glu Ile
    290                 295                 300
Ser Gly Ile Lys Leu Leu Phe Arg Leu Asn Pro Asn Gly Ser Pro Lys
305                 310                 315                 320
Ala Ala Tyr Leu Cys Phe Thr Cys Ser Ile Thr Asp Glu Pro Leu Ser
                325                 330                 335
Ser Asn Ala Arg Lys Ile Lys Trp Ser Glu Thr Gly Glu Thr Thr Lys
            340                 345                 350

Lys Gly Lys Lys Arg Lys Arg Lys Thr Leu Pro Asp Gly Leu Val Thr
        355                 360                 365
Cys Ala Val Asp Leu Gly Ile Arg Asn Thr Gly Phe Ala Thr Leu Ala
    370                 375                 380
Val His Asp Gln Gly Asn Val Leu Ile Arg Arg Ser Arg Asn Leu Trp
385                 390                 395                 400
Ala Gly Trp Leu Gly Glu Asp Ser Lys Asp Arg Asp Arg Trp Met Asp
                405                 410                 415
Gly Pro Val Leu Ala His Ile Ala Ala His Lys Arg Lys Ile Arg Phe
            420                 425                 430
Leu Arg Arg Lys Arg Gly Lys Pro Val Lys Gly Glu Thr Ser His Val
        435                 440                 445
Glu Leu Gln Thr His Ile Asp His Met Gly Glu Asp Arg Phe Lys Arg
    450                 455                 460
Ala Ala Arg Ala Ile Ile Asn Phe Ala Leu Asn Val Gly Ser Glu Cys
465                 470                 475                 480
```

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

[0103]

<210> 313
<211> 710
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水-沉积物序列
<400> 313
Met Val Glu Leu Tyr Leu Arg Met Arg Lys Gly Lys His Arg Glu Thr
1 5 10 15
Ala Arg Lys Ile Ala Glu Ile Ile Leu Lys Ser Pro Lys Asn Thr Ala
20 25 30
Ala Ala Ile Met Glu Gly Leu Thr Ala Lys Glu Trp Thr Ser Thr Arg
35 40 45
Thr Thr Glu Trp Ala Asn Leu Val Arg Glu Arg Arg Asn Ala Glu Gly
50 55 60
Leu Leu Phe Asp Arg His Glu Ser Val Gly Ser Ala Asp Gly Arg Asn
65 70 75 80
Ala Cys Cys Gln Arg Gln Gln Gly Pro Arg Lys Lys Pro Thr Asp Pro
85 90 95
Ser Arg Leu Pro Val Ser Ser Lys Phe Trp Arg Trp Leu Cys Asp Asp
100 105 110

[0104]

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

[0105]

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

<210> 314
<211> 691
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水-沉积物序列
<400> 314
Met Ser Val Lys Cys Phe Gln Ala Lys Val Val Cys Asp Thr Pro Glu
1 5 10 15
Lys Arg Glu Tyr Leu Trp Leu Thr His Thr Leu Phe His Glu Ala Leu
20 25 30

[0106]

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

[0107]

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

<210> 315
<211> 741
<212> PRT
<213> 未知

<220>

<223> 未知的描述:
水生-淡水-沉积物序列

<400> 315

[0108]

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

[0109]

Asp Leu Ala Phe Pro Glu Lys Arg Gly Ala Thr Gly Ser Ile Arg Leu
325 330 335
Cys Leu Leu Thr Gly Asp Lys Ala Ala Asp Gly Lys Tyr Pro His Glu
340 345 350
Trp Val Pro Val Arg Phe Leu Ala Asp Pro Arg Leu Ser Ala Phe Arg
355 360 365
Pro Val Lys Arg Gln Arg Val Ile Gly Arg Gly Lys Arg Lys Gly Glu
370 375 380
Ala Ala Glu Ala Lys Ser Ala Tyr Glu Tyr Phe Asp Arg His Leu Gly
385 390 395 400
Lys Thr Arg Glu Ala Arg Phe Ser Gly Val Lys Leu Val Phe Arg Leu
405 410 415
Tyr Pro Asp Gly Lys Pro Lys Ala Ala Tyr Leu Tyr Phe Ala Cys Glu
420 425 430
Leu Thr Asp Glu Pro Leu Ser Glu Arg Ala Lys Ala Ile Gln Trp Thr
435 440 445
Glu Thr Gly Glu Val Thr Arg Lys Gly Lys Lys Arg Lys Ser Lys Arg
450 455 460
Leu Pro Asp Gly Leu Val Ser Ala Ala Val Asp Leu Gly Ile Arg Asn
465 470 475 480
Leu Gly Phe Ala Thr Leu Ala Arg Lys Glu Asp Gly Ala Val Gln Val
485 490 495
Leu Arg Ser Arg Asn Leu Trp Thr Gly Pro Asp Leu His His Ile Thr
500 505 510
Arg His Lys Arg Arg Leu Arg Ala Leu Arg Arg Asp Arg Gly Lys Pro
515 520 525
Val Ala Gly Glu Ser Ser His Val Arg Leu Gln Gly His Ile Asp His
530 535 540
Met Gly Glu Asp Arg Phe Lys Arg Ala Ala Arg Ala Ile Val Asn Phe
545 550 555 560
Ala Leu Asn Thr Asp Gly Ala Arg Asn Gly Arg Gly Gln Ala His Pro
565 570 575
Arg Ala Asp Val Leu Val Leu Glu Ser Met Glu Gly Leu Ile Pro Asp
580 585 590
Ala Glu Lys Glu Arg Gly Ile Asn Arg Ala Leu Val Ala Trp Asn Arg
595 600 605
Gly His Leu Val Glu Arg Ile Arg Asp Met Ala Leu Asp Ala Gly Leu
610 615 620
Lys Val Phe Glu Val Ser Pro Phe Gly Thr Ser Gln Val Cys Ser Arg
625 630 635 640
Cys Gly Ala Leu Gly Arg Arg Tyr Ser Leu Val Arg Glu Ala Glu Thr
645 650 655
Gly Glu Gln Ala Val Arg Phe Gly Phe Val Glu Lys Leu Phe Ala Cys
660 665 670

Pro Ala Cys Gly Tyr Arg Ala Asn Ala Asp His Asn Ala Ser Val Asn
675 680 685
Leu His Arg Arg Phe Ile Leu Glu Asp Ala Ala Val Lys Ala Tyr Gln
690 695 700
Asp Leu Tyr Asn Arg Pro Glu Ser Glu Arg Lys Val Ala Leu Ala Asp
705 710 715 720
Met Glu Ser Arg Leu Leu Ser Arg Leu Gly Val Glu His Gly Leu Ala
725 730 735
Lys Pro Pro Pro Phe
740

<210> 316
<211> 745
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水-沉积物序列
<400> 316

[0110]

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

[0111]

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

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

[0112]

<210> 317
<211> 794
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水-沉积物序列
<400> 317
Met Ser Val Arg Ser Ile Gln Ala Arg Leu His Cys Asp Ala Lys Thr
1 5 10 15
Leu Asp Asp Leu Trp Arg Thr His Arg Val Phe Asn Glu Arg Leu Pro
20 25 30
Gly Ile Asn Ala Ile Leu Phe Lys Met Arg Arg Gly Glu Phe Gly Gly
35 40 45
Thr Pro Lys Glu Arg Glu Met Met Gln Ala Phe Ala Trp Cys Val Leu
50 55 60
Arg Leu Asp Ala Lys Asn Ser Tyr Asp Pro Leu Asn Lys Ala Ala Ile
65 70 75 80
Gly Ser Ala Arg Ser Ala Ser Ala Val Lys Ala Val Asn Ala Leu Ile
85 90 95

[0113]

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

[0114]

```
Phe Arg Phe Asn His Asp Gly Thr Pro Lys Ala Ala Tyr Leu Tyr Phe
    450                 455                 460
Thr Cys Asn Ile Lys Asp Val Pro Leu Thr Ala Asn Ala Lys Ala Ile
465                 470                 475                 480
Gln Trp Thr Glu Thr Gly Glu Val Thr Lys Lys Gly Lys Lys Arg Lys
                485                 490                 495
Arg Lys Thr Leu Pro Asp Gly Ile Leu Thr Cys Ala Val Asp Leu Gly
            500                 505                 510
Ile Arg Asn Leu Ala Phe Ala Thr Leu Ala Val His His Glu Lys Gly
        515                 520                 525
Asn Leu Val Ile Lys Arg Ser Arg Asn Leu Trp Val Gly Trp Gln Glu
    530                 535                 540
Gly Gly Asp Ala Lys Gly Glu Arg Trp Leu Asp Gly Pro Asp Leu Ala
545                 550                 555                 560
His Ile Ala Lys His Lys Arg Lys Ile Arg Trp Leu Arg Gly Lys Arg
                565                 570                 575
Gly Lys Pro Val Lys Gly Glu Ser Ser His Val Glu Leu Gln Thr His
            580                 585                 590
Ile Asp His Met Gly Glu Asp Arg Phe Lys Arg Ala Ala Arg Ala Ile
        595                 600                 605
Ile Asn Phe Ala Leu Asn Ala Gln Ala Ala Cys Asp Arg His Gly Glu
    610                 615                 620
Val Tyr Pro Arg Ala Asp Val Leu Ile Leu Glu Asn Met Asp Gly Leu
625                 630                 635                 640
Ile Pro Asp Ala Glu Lys Glu His Gly Ile Asn Arg Ala Leu Ile Ala
                645                 650                 655
Trp Asn Arg Gly Gln Leu Thr Lys Arg Ile Glu Glu Met Ala Lys Asp
            660                 665                 670
Val Gly Leu Lys Ile Ile Pro Val Ser Pro Met Gly Thr Ser Gln Val
        675                 680                 685
Cys Ser Arg Cys Gly Ser Leu Gly Arg Arg Tyr Ser Ile Gly Arg Asp
    690                 695                 700
Glu Lys Thr Gly Gln Ala Glu Ile Arg Phe Gly Phe Val Glu Lys Leu
705                 710                 715                 720
Phe Ala Cys Pro Asn Ser Asp Cys Gly Tyr Leu Ala Asn Ala Asp His
                725                 730                 735
Asn Ala Ser Val Asn Leu His Arg Arg Phe Ala Leu Asp Asp Lys Ala
            740                 745                 750
Val Ala Gly Tyr Ala Glu His Arg Val Lys Gly Asp Asp Glu Arg Thr
        755                 760                 765
Lys Glu Ile Ala Asp Ile Glu Ala Arg Leu Arg Pro Arg Leu Arg Leu
    770                 775                 780
Lys His Ser Leu Asp Gly Pro Val Pro Phe
785                 790
```

<210> 318
<211> 727
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水-沉积物序列
<400> 318

[0115]

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

[0116]

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

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

<210> 319
<211> 740
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水-沉积物序列
<400> 319

[0117]

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

[0118]

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

[0119]

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

<210> 320
<211> 744
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水-沉积物序列
<400> 320
Met Val Thr Arg Ala Leu Arg Ala Arg Val Ile His Pro Asp Ala Thr
1 5 10 15
Leu Leu Ala Ala Leu Lys Arg Thr His Arg Val Tyr Asn Val Leu Leu
20 25 30
Arg Gln Met Val Glu Leu Tyr Leu Arg Met Arg Lys Gly Lys His Arg
35 40 45
Glu Thr Ala Arg Lys Ile Ala Glu Ile Ile Leu Lys Ser Pro Lys Asn
50 55 60
Thr Ala Ala Ala Ile Met Glu Gly Leu Thr Ala Lys Glu Trp Thr Ser
65 70 75 80

[0120]

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

Leu Ile Phe Lys Gln Gly Ala Ala His Leu Thr Phe Thr Ile Ile Asp
```

[0121]

420 425 430

Gln Asp Arg Ala Pro Arg Val Lys Pro Arg Lys Val Ala Asp Arg Arg
435 440 445

Cys Pro Ala Asp Asn Ile Phe Asp Pro Gln Gly Asp Arg Ile Pro Leu
450 455 460

Arg Val Val Ala Val Asp Leu Gly Val Arg His Val Gly Gly Tyr Ala
465 470 475 480

Val Ala Glu Gly Arg Phe Glu Asn Gly Ala Trp Ser Val Gly Tyr Ile
485 490 495

Thr Lys Gly Ile Val Ser Gly Pro Gln Val Pro Glu Arg Phe Ser Ile
500 505 510

Asp Ala His His Arg Gln Leu Lys Trp Ala Arg Arg Leu Arg Gly Lys
515 520 525

Pro Val Phe Asp Glu Glu Ser Cys Ala Glu Leu Gln Asp His Val Thr
530 535 540

His Met Gly Glu Asp Arg Phe Lys Lys Ala Ala His Ala Ile Val Glu
545 550 555 560

Leu Ala Arg Arg His Asp Val His Val Ile Leu Phe Glu Asn Leu Lys
565 570 575

Gly Phe Val Pro Ser Ala Arg Asp Glu Arg Trp Ala Asn Arg Thr Arg
580 585 590

Thr Lys Met Leu Arg Arg Lys Thr Val Glu Met Val Lys Ala Gln Ala
595 600 605

Gly Glu Phe Gly Phe Leu Cys Asn Glu Lys Ile Ser Pro Lys Leu Thr
610 615 620

Ser His Ile Cys Ser Lys Cys Phe Arg Pro Gly Trp Arg Phe Ser Met
625 630 635 640

Lys Ala Lys Gln Pro Trp Ser Glu Lys Gln Pro Arg Ser Lys Cys Arg
645 650 655

Asp Phe Gly Tyr Pro Ala Trp Asp Ala Gly Gly His Leu Phe Arg Cys
660 665 670

Pro His Cys Gly Tyr Arg Val Asn Ala Asp Ile Asn Ala Ala Gly Asn
675 680 685

Leu Val Ala Lys Phe Phe Gly Leu Trp Pro Asp Glu Lys Leu Glu Arg
690 695 700

Lys Asp Trp Val Tyr Arg Trp Arg Glu Asn Gly Glu Ser Arg Gln Phe
705 710 715 720

Asp Ala Lys Ala Glu Phe Glu Lys Trp Ser Asn Asp Val Arg Ser Arg
725 730 735

Lys Ala Ala Lys Ser Phe Pro Tyr
740

<210> 321
<211> 796
<212> PRT

<213> 未知
<220>
<223> 未知的描述:
水生-淡水-沉积物序列
<400> 321

[0122]

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

[0123]

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

660 665 670
Leu Leu Ile Tyr Asp Arg Ile Asn Pro Ala Tyr Thr Ser Ser Met Cys
675 680 685
Ser Val Cys Gly Asn Pro Gly Ala Arg Ile Ser Leu Pro Ser Lys Arg
690 695 700
Asn Trp Glu Lys Tyr Tyr Ala Glu Tyr Tyr Gly Lys Glu Arg Arg Met
705 710 715 720
Ile Ala Val Val Gly Gly Gln Phe Phe Cys Cys Pro Val Cys Lys Lys
725 730 735
Ile Ile Asn Ala Asp Ile Asn Ala Ser Gln Asn Met His Lys Val Phe
740 745 750
Tyr Asn Thr Phe Asn Trp Pro Pro Lys Met Glu Lys Lys Glu Lys Lys
755 760 765
Ile Phe Trp Glu Ser Val Arg Glu Gln Ala Gln Asn Ile Leu Asp Gln
770 775 780
Lys Ala Glu Pro Lys Arg Glu Glu Asp Ile Ser Tyr
785 790 795

<210> 322
<211> 759
<212> PRT
<213> 未知
<220>

[0124]

<223> 未知的描述:
水生-淡水-沉积物序列
<400> 322
Met Ser Val Arg Ala Ile Gln Ala Arg Val Glu Cys Asp Arg Ala Thr
1 5 10 15
Leu Glu His Leu Trp Arg Thr His Arg Val Phe Asn Glu Arg Val Pro
20 25 30
Val Val Leu Ser Trp Leu Phe Lys Met Arg Arg Gly Glu Cys Gly Arg
35 40 45
Thr Glu Glu Glu Lys Arg Leu Tyr Arg Thr Ile Ala Asp Phe Val Leu
50 55 60
Ala Asn Ser Gln Asn Ala Glu Ala Ile Leu Asn Ala Val Ser Ile Pro
65 70 75 80
Asn Trp Thr Pro Thr Ser Ala Arg Lys Leu Lys Ala Lys Thr Val Leu
85 90 95
Ser Asp Gly Thr Glu Lys Val Ser Thr Gly Met Asp Trp Ala Asp Lys
100 105 110
Ala Ala Glu Leu Ser Ala Ala Gly Arg Leu Leu Phe Asp Lys Arg Glu
115 120 125
Met Leu Phe Asp Leu Pro Gly Thr Leu Arg Gln Met Val Cys Arg Asp
130 135 140
Ala Gly Ala Ile Ile Ser Gly His Glu Glu Leu Val Ala Val Trp Lys

145 150 155 160
Lys Ala His Ala Asp Trp Leu Lys Ala Lys Ala Asp Trp Glu Ser Gln
165 170 175
Glu Glu Asn Lys Leu Tyr Leu Asn Leu Arg Pro Glu Phe Glu Ala Phe
180 185 190
Glu Gln Ala Ala Gly Gly Arg Ala Ala Lys Arg Arg Gly Arg Trp His
195 200 205
Leu Tyr Leu Lys Trp Leu Lys Asp His Pro Arg Leu Gly Gly Trp Arg
210 215 220
Gly Gly Glu Pro Ile Val Asn Glu Leu Ser Glu Gln Ala Arg Arg Arg
225 230 235 240
Ile Gln Lys Ala Lys Pro Trp Lys Gln Arg Ser Val Glu Gly Glu Glu
245 250 255
Phe Trp Arg Ala Asn Pro Glu Leu Ala Ala Leu Asp Arg Leu His Gly
260 265 270
Tyr Tyr Glu Arg Glu Phe Ile Arg Arg Arg Lys Thr Lys Lys Asn Pro
275 280 285
Asp Gly Phe Thr His Pro Pro Thr Phe Thr Val Pro His Ala Val Arg
290 295 300
His Pro Arg Trp Phe Val Phe Asn Ala Pro Gln Thr Ser Pro Ile Gly
305 310 315 320
Tyr Arg Asn Leu Val Leu Pro Pro His Ala Gly Pro Glu Gly Thr Val
[0125] 325 330 335
Glu Leu Arg Leu Leu Thr Gly Asp Lys Thr Asp Gly Lys His Pro Gly
340 345 350
Asn Trp Val Lys Val Arg Phe Arg Ala Asp Pro Arg Leu Ser Asp Phe
355 360 365
Arg Pro Ala Lys Arg Thr Arg Ser Phe Asn Arg Gly Arg Asn Lys Gly
370 375 380
Gln Thr Thr Gln Ser Asp Ala Tyr Glu Tyr Phe Asp His His Leu Gly
385 390 395 400
Leu Met Arg Ser Ala Glu Ile Lys Gly Leu Lys Leu Ile Phe Arg Phe
405 410 415
Asn Pro Asp Asp Ser Leu Arg Ala Ala Tyr Leu Tyr Phe Thr Cys Asp
420 425 430
Val Gly Ser Ile Pro Phe Thr Glu Arg Thr Arg Ala Ile Gln Trp Thr
435 440 445
Glu Thr Gly Glu Val Thr Lys Lys Gly Lys Lys Arg Lys Thr Lys Lys
450 455 460
Leu Pro Asp Gly Leu Ile Ala Cys Ala Val Asp Leu Gly Leu Arg His
465 470 475 480
Leu Gly Phe Ala Thr Arg Ala Arg Tyr Ala Gly Ala Asn Pro Glu Ile
485 490 495
Leu Arg Ser Arg Asn Leu Trp Leu Gly Arg Glu Glu Gln Lys Gly Arg

[0126]

```
            500                 505                 510
His Pro Gly Arg Leu Ser Pro Gly Ala Asp Leu Arg His Ile Ala His
        515                 520                 525
His Lys Arg Asn Ile Arg Arg Leu Arg Arg Leu Arg Gly Lys Pro Val
    530                 535                 540
Lys Ser Glu Ser Ser His Val Glu Leu Gln Thr His Ile Asp His Met
545                 550                 555                 560
Gly Glu Asp Arg Phe Lys Gln Gly Ala Arg Ala Ile Ile Asn Phe Ala
                565                 570                 575
Leu Asn Val Lys Lys Gly Pro Asp Thr Arg Thr Gly Glu Val His Pro
            580                 585                 590
Arg Ala Asp Val Leu Ile Leu Glu Lys Leu Ala Gly Leu Leu Pro Asp
        595                 600                 605
Ala Glu Lys Glu Arg Gly Ile Asn Arg Ala Leu Val Ala Trp Asn Arg
    610                 615                 620
Gly Gln Leu Val Lys Arg Val Lys Glu Met Ala Gln Asp Val Gly Leu
625                 630                 635                 640
Lys Val Phe Glu Val Ser Pro Val Gly Thr Ser Gln Val Cys Ser Arg
                645                 650                 655
Cys Gly Ala Leu Gly Arg Arg Tyr Ser Ile Val Pro Asp Asp Arg Thr
            660                 665                 670
Gly Arg Pro Gly Phe Arg Phe Gly Trp Val Glu Lys Leu Phe Ala Cys
        675                 680                 685
Pro Asn Pro Asp Cys Arg Tyr Arg Gly Asn Ala Asp His Asn Ala Ser
    690                 695                 700
Val Asn Leu His Arg Arg Phe Met Leu Glu Asp Ala Ala Val Lys Ala
705                 710                 715                 720
Phe Leu Asp Trp Arg Ala Lys Thr Lys Lys Glu Gln Glu Glu Val Leu
                725                 730                 735
Arg Arg Ile Glu Lys Ala Leu Cys Ala Asp Leu Leu Arg Met His Gly
            740                 745                 750
Leu Glu Glu Pro Thr Pro Phe
        755
<210> 323
<211> 702
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
      水生-海生-深部地下序列
<400> 323
Met Ser Thr Lys Ala Ile Gln Ala Lys Val Ile Cys Asp Thr Trp Glu
1               5                   10                  15
His Arg Tyr Phe Leu Trp Ile Thr His Lys Leu Phe Asn Asp Ser Val
```

[0127]

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

370 375 380
Gln Trp Ile Asp Lys Leu Arg Pro Tyr Phe Val Phe Thr Leu Thr Leu
385 390 395 400
Ala Asp Glu Ala Lys Pro Val Lys Leu Glu Trp Asn Lys Lys Glu Lys
405 410 415
His Ser Val Ile Pro Lys Glu Leu Val Thr Met Ala Ile Asp Val Gly
420 425 430
Gln Arg Asp Leu Gly Ala Val Thr Ile Ala Arg Glu His Asn Gly Val
435 440 445
Pro Val Gln Ile Ala Phe Leu Pro Gln Tyr Trp Lys Pro Leu Lys Asn
450 455 460
Gly Asn Ser Ala Lys Pro Val Lys Ala Trp Leu Val Asp Ile Pro Gly
465 470 475 480
Ile Ser Phe Glu Ala Leu Lys Lys His Glu Lys Glu Ala Arg Gly Arg
485 490 495
Glu Arg Gly Ala Lys Ala Arg Arg Val Lys Gly Glu Asn Tyr Ala Ile
500 505 510
Gly Leu Arg Gln His Ile Leu His Met Lys Glu Asp Arg Tyr Lys Lys
515 520 525
Ala Gly Ser Ile Leu Leu Arg Thr Ala Ile Lys Asn Lys Val His Val
530 535 540
[0128] Ile Ile Leu Glu Asp Leu Lys Lys Tyr Lys Met Asn Leu Glu Arg Ser
545 550 555 560
Lys Gln Glu Asn Arg Gln Leu Met Val Trp Ala Ile Arg Lys Ile Gln
565 570 575
Asp Phe Leu Glu Leu Gln Ala Lys Pro Leu Gly Val Arg Val Arg Ser
580 585 590
Val Trp Pro His Tyr Thr Ser Gln Val Cys Ser Ala Cys Gly Glu Pro
595 600 605
Gly Lys Arg Ala Ser Leu Pro Ser Lys Ser Ser Trp Glu Lys Phe Tyr
610 615 620
Gln Lys Lys Tyr Gly Glu Lys Arg Lys Trp Ile Val Glu Asn Gly Gly
625 630 635 640
Lys Met Phe Val Cys Ser Asn Pro Asn Cys Val Arg Lys Ser Arg Met
645 650 655
Ile Asn Ala Asp Ile Asn Ala Ser Leu Asn Leu His Lys Lys Phe Tyr
660 665 670
Gly Ile Leu Thr Tyr Pro Pro Lys Thr Lys Glu Phe Lys Gln Gln Leu
675 680 685
Thr Leu Asp Ile Gln Asn Tyr Leu Asp Thr Ile Leu Lys Phe
690 695 700
<210> 324
<211> 736
<212> PRT

<213> 未知
<220>
<223> 未知的描述:
水生-海生-深部地下或地下宏基因组
序列
<400> 324

[0129]

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

[0130]

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

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

<210> 325
<211> 679
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-非海生盐水和碱性-超盐水湖
沉积物或沉积物宏基因组序列
<400> 325

[0131]

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

[0132]

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

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

<210> 326
<211> 741
<212> PRT
<213> 未知
<220>

[0133]

<223> 未知的描述:
水生-非海生盐水和碱性-超盐水湖
沉积物或沉积物宏基因组序列

<400> 326
Met Ala Thr Lys Thr Tyr Arg Ala Arg Ile Val Gln Ile Ser Pro Arg
1 5 10 15
Leu Gly Ala Ala Leu Glu Arg Thr His Tyr Leu Phe Ala Glu Ala Leu
20 25 30
Asn Gln Met Val Arg Arg Tyr Ile Asp Met Arg Lys Gly Lys Tyr Gly
35 40 45
Gln Glu Cys Lys Lys Leu Ala Asp Ile Met Leu Thr Arg Ser Asn Thr
50 55 60
Phe Ala His Gly Val Met Asp Gln Leu Thr Arg Glu Ile Ser Thr Ser
65 70 75 80
Gly Leu Thr Asp Glu Trp Val Asp Leu Ala Arg Ala Val His Lys Ser
85 90 95
His Gly Pro Leu Phe Leu Gln His Asp Gly Phe Ala Glu Val Asp Gly
100 105 110
Val Met Leu His Thr Lys Arg His Gly Lys Val Ser Pro Met Pro Gly
115 120 125
Arg Leu Pro Val Asn Ala Lys Phe Trp His Gln Val Cys Asp Ser Ala
130 135 140

[0134]

```
Ser Ala Phe Leu Lys Ser Asn Asn Glu Leu Leu Glu Asn Trp Arg Gln
145                 150                 155                 160
Gly Arg Lys Glu Trp Leu Gln Glu Lys Ala Asp Trp Glu Ser Ala Asn
                165                 170                 175
Ala Asp Phe Met Arg Phe Trp Asn Gly Ser Tyr Gln Glu Phe Glu Glu
            180                 185                 190
Thr Cys Glu Ile Lys Arg Lys Glu Ser Gln Ala Leu Glu Gly Gln Lys
        195                 200                 205
Val Val Glu Lys Lys Arg His Arg Arg Asp Arg Gly Lys Arg Met Ser
    210                 215                 220
Arg Trp His Leu Trp Tyr Asn Trp Val Ile Thr His Pro Glu Ile Ile
225                 230                 235                 240
Glu Trp Arg Asn Arg Ala Lys Ala Ala Glu Phe Arg Pro Val Pro Pro
                245                 250                 255
Thr Val His Lys Lys Ile Lys Asn Lys Tyr Pro Arg Gln Asp Lys His
            260                 265                 270
Ile Thr Lys Phe Leu Asp Trp Leu Lys Glu Asn Asn Pro Glu Leu Lys
        275                 280                 285
Ala Leu Asp Ser Arg Arg Arg Phe Tyr Val Arg Asn Tyr Leu Arg Phe
    290                 295                 300
Arg Arg Pro Pro Thr Leu Thr Leu Pro Ser Ala Asn Lys His Pro Tyr
305                 310                 315                 320
Trp Phe Thr Phe Glu Leu Asn Gln Phe Tyr Lys Asn Val Asp Phe Asp
                325                 330                 335
Lys Gly Ser Ile Gln Leu Leu Leu Val Asp Glu Thr Glu Glu Gly Ser
            340                 345                 350
Trp Phe Phe Arg Trp Phe Asp Ala Lys Ile Gln Cys Asp Pro Arg Leu
        355                 360                 365
Arg Pro Cys Val Arg Ala Gly Val Phe Ala Thr Lys Gly Arg Tyr Pro
    370                 375                 380
Pro Tyr Ile Gly Gly Lys Val Gly Asn Lys Leu Asn Arg Pro Ala Glu
385                 390                 395                 400
Arg Pro Glu Asn Arg Lys Ala Gly Tyr Ala Gly Ala Lys Leu Val Leu
                405                 410                 415
Asn Lys Arg Arg Lys Glu Leu Phe Phe Thr Val Ile Glu Gln Asp Cys
            420                 425                 430
Pro Arg Lys Val Ser Trp Lys Lys Thr Lys Gln Arg Thr Cys Arg Gly
        435                 440                 445
Asp Asn Val Phe Ala Lys Asp Gly Ser Gln Ile Pro Leu Lys Val Met
    450                 455                 460
Ala Ile Asp Leu Gly Ile Arg His Ile Gly Ala Tyr Val Ile Ala Glu
465                 470                 475                 480
Ala Cys Arg Asn Arg Gly Val Trp Asp Ile Ser Trp Gln His Lys Gly
                485                 490                 495
```

Ile Thr Gly Arg His Ala Val Pro Gly Leu Lys Glu Ile Arg Gln His
500 505 510
Asp Arg Gln Leu Arg Lys Ala Arg Ser Gln Arg Gly Lys Pro Val Gln
515 520 525
Gly Glu Glu Ser Phe Ile Glu Leu Gln Asp His Arg Thr Asn Met Ala
530 535 540
Asp Asp Arg Phe Lys Lys Ala Ala Asn Thr Ile Val Glu Leu Ala Arg
545 550 555 560
Gln Tyr Asp Val His Leu Ile Leu Phe Glu Arg Leu Gly Asn Leu Ser
565 570 575
Pro Thr Ala Phe Asp Glu Arg Trp Met Asn Arg Gln Leu Arg Asp Met
580 585 590
Asn Arg Arg Gln Ile Val Asp Thr Val Lys Leu Gln Ala Arg Glu Phe
595 600 605
Gly Ile Glu Cys Leu Glu Asp Ile Ser Pro Trp Met Thr Ser His Ile
610 615 620
Cys Ser Arg Cys Phe Lys Pro Gly Trp Arg Phe Ser Ile Lys Arg Lys
625 630 635 640
Asn Pro Tyr Gly Glu Arg Phe Thr Arg Lys Asn Cys Arg Asp Tyr Gly
645 650 655
Tyr Pro Val Trp Asp Thr Gly Gly His Leu Phe Arg Cys Pro His Cys
660 665 670
[0135] Glu Tyr Lys Cys Asn Ala Asp Ile Asn Ala Ala Cys Asn Val Ala Ala
675 680 685
Lys Phe Phe Gly Leu Trp Ser Gly Leu Asp Gly Leu Lys Arg Asp Gln
690 695 700
Trp Val Tyr Thr Trp Gln Glu Asn Gln Pro Lys Thr Phe Asp Ala Lys
705 710 715 720
Lys Thr Phe Leu Glu Trp Ala Leu Asp Val Arg Asn Arg Lys Leu Met
725 730 735
Gly Asp Ser Pro Phe
740

<210> 327
<211> 762
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-非海生盐水和碱性-超盐水湖
沉积物序列
<400> 327
Met Ser Val Lys Thr Ile Gln Ala Lys Leu Glu Cys Asp Arg Glu Thr
1 5 10 15
Leu Glu Thr Leu Trp Arg Thr His Thr Val Phe Asn Glu Gly Leu Leu

[0136]

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

[0137]

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

725 730 735
Lys Glu Gln Lys Lys Val Ile Glu Glu Glu Leu Arg Glu Phe Leu Arg
740 745 750
Lys Glu His Phe Pro Gln Met Ala Leu Asn
755 760

<210> 328
<211> 670
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-非海生盐水和碱性-超盐水湖
沉积物或沉积物宏基因组序列
<400> 328

[0138]

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

Lys Leu Asp Ala Lys Tyr Gly Arg Asn Met Val Arg Arg Ser Ala Arg
245 250 255
His Arg Asn Ile Asp Gly Phe Arg His Pro Pro Thr Met Thr Leu Pro
260 265 270
His Ser Arg Leu His Pro Glu Trp Ile Thr Phe Gln Ala Gly Glu Thr
275 280 285
Tyr Lys Asp Leu Asp Leu Asn Gly Gly Thr Val Arg Leu Asn Ile Leu
290 295 300
Pro Pro Asp Ala Glu Thr Ser Lys Leu His Ser Val Glu Phe Arg Ala
305 310 315 320
Glu Lys Arg Leu Arg Arg Ile Thr Pro Arg Glu Thr Pro Leu Arg Lys
325 330 335
Gly Arg Asn Glu Phe Asn Lys Asn Tyr Arg Asp Ala Ala Asp Arg Val
340 345 350

His Pro Ala Arg Val Lys Gly Ile Arg Leu Ile Leu Arg Pro Arg Gly
355 360 365
Ala Tyr Leu Lys Ile Thr Val Glu Leu Asp Arg Pro Glu Ser Arg Thr
370 375 380
Gly Leu Lys Gln Gln Thr Cys Asp Lys Tyr Ser Ser Asn Trp Cys Tyr
385 390 395 400
Gln Glu Ile Val Lys Arg Arg Glu Gly Lys Pro Leu Lys Cys Leu Ser
405 410 415
Leu Asp Leu Gly Val Arg His Leu Ala Val Gly Val Val Val Gln Glu
420 425 430
Thr Arg Ala Asn Gly Asp Arg Glu Thr Val Ile Glu Gly Val Arg Phe
435 440 445
Phe Lys Glu Val Gly Gly Arg Pro Phe Pro Ser Leu Asp Glu Ile Arg
450 455 460
Arg His Lys Asp Arg Leu Arg Glu Lys Arg Ser Gln Thr Arg Arg Ala
465 470 475 480
Pro Lys Gly Arg Arg Thr Cys Ile Ala Leu Gln Asp His Val Thr Lys
485 490 495
Met Gly Gln Asp Arg Tyr Arg Lys Gly Ala Ala Ala Ile Val Arg Tyr
500 505 510
Ala Val Glu Arg Asp Val Asp Val Ile Leu Leu Glu Lys Leu Ala Gly
515 520 525
Phe Ile Pro Asp Ala Ala Lys Gln Arg Gly Ile Asn Lys Ala Leu Met
530 535 540
Glu Trp Asn Arg Arg Lys Ile Ala Asp Phe Val Lys Leu Leu Ala Ala
545 550 555 560
Asp His Gly Ile Ser Val Val Glu Val Ala Pro His His Leu Ser Gln
565 570 575
Leu Cys Ser Lys Cys Gly Glu Leu Gly Arg Arg Phe Ser Trp Lys Asp

[0139]

580 585 590

Gly Leu Pro Lys Phe Gly Thr Val Glu Lys Leu Phe Val Cys Pro Ala
595 600 605

Cys Gly Tyr Gln Ala Asn Ser Asp Tyr Asn Ala Ala Val Asn Leu Tyr
610 615 620

Lys Ala Phe Ser Gly Ala Tyr Glu Glu Arg Pro Lys Lys Ala Lys Lys
625 630 635 640

Gly Val Tyr Lys Tyr Arg Asp Arg Thr Ile Cys Leu Lys Lys Leu Asp
645 650 655

Gly Gln Leu Ala Glu Arg Leu Thr Gln Glu Gln Thr Pro Phe
660 665 670

<210> 329
<211> 762
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-非海生盐水和碱性-超盐水湖
沉积物或沉积物宏基因组序列

<400> 329

[0140]

Met Ser Val Lys Thr Ile Gln Ala Lys Leu Glu Cys Asp Arg Glu Thr
1 5 10 15

Leu Glu Thr Leu Trp Arg Thr His Thr Val Phe Asn Glu Gly Leu Leu
20 25 30

Pro Ile Leu Ala Leu Phe Phe Lys Met Lys Arg Arg Glu Thr Gly Ser
35 40 45

Thr Glu Glu Glu Lys Asp Leu Tyr Lys Arg Ile Cys Glu Phe Ile Leu
50 55 60

Arg Thr Asn Ala Ser Asn Pro Gly Tyr Leu Leu Asn Ser Val Cys Leu
65 70 75 80

Glu Asn Trp Lys Pro Asp Ser Ala Arg Lys Thr Lys Ala Glu Val Val
85 90 95

Arg Asp Gly Lys Lys Glu Lys Ile Ser Ala Glu Glu Trp Ile Asp Asp
100 105 110

Leu Leu Asn Met Lys Glu Glu Ser Gly Leu Leu Phe Leu Lys Gln Glu
115 120 125

Leu Leu Asn Gly Leu Pro Ile Met Phe Phe Glu Lys Leu Val Ala Glu
130 135 140

Ser Lys Ala Ile Leu Ser Ala Tyr Glu Glu Leu Gln Tyr Gln Trp Glu
145 150 155 160

Val Asp His Asn Lys Trp Leu Lys Glu Lys Glu Gly Trp Glu Lys Lys
165 170 175

His Lys Glu Tyr Leu Ala Val Arg Ser Leu Phe Gln Lys Phe Glu Gln

[0141]

```
                180                 185                 190
Glu Val Lys Gly Lys Ile Thr Lys Lys Arg Arg Arg Trp His Leu Tyr
        195                 200                 205
Lys Asp Trp Met Lys Arg Thr Pro Glu Leu Val Ala Trp Arg Gly Lys
    210                 215                 220
Glu Thr Lys Val Tyr Gly Leu Thr Glu Glu Asp Arg Glu Ala Ile Lys
225                 230                 235                 240
Lys Ala Pro Leu Lys Lys Lys Ile Gly Val Glu Val Gly Leu Phe Phe
                245                 250                 255
Lys Arg Asn Pro Glu Leu Ser Ala Leu Asp Lys Leu His Ser Tyr Tyr
            260                 265                 270
Glu Glu Lys Phe Val Arg Arg Gly Lys Lys Lys Arg Asn Cys Asp Gly
        275                 280                 285
Tyr Glu His Arg Pro Thr Phe Thr Leu Pro His Ser Val Lys His Pro
    290                 295                 300
Arg Trp Tyr Val Phe Ser Gly Pro Gln Thr Gly Asn Gly Tyr Lys Asn
305                 310                 315                 320
Leu Asn Ile Pro Glu Arg Lys Lys Asp Phe Gly Gln Leu Asp Leu Leu
                325                 330                 335
Val Leu Thr Asp Asp Lys Val Asp Gly Glu Tyr Pro Arg Glu Tyr Arg
            340                 345                 350
Thr Val Lys Phe Arg Gly Asp His Arg Phe Leu Asp Ile Arg Gly Ala
        355                 360                 365
Thr Arg Thr Arg Leu Val Thr Lys Gly Lys Asn Lys Gly Gln Glu Val
    370                 375                 380
Glu Glu Lys Phe Ala Tyr Glu Met Tyr Asp Arg Tyr Leu Lys Ile Trp
385                 390                 395                 400
Arg Pro Leu Lys Ile Ser Gly Ala Lys Leu Ala Phe Arg Asp Ile Lys
                405                 410                 415
Ile Asn Gln Asp Gly Ser Leu Lys Ser Ala Val Pro Tyr Leu Trp Phe
            420                 425                 430
Thr Leu Asn Leu Glu Asp Leu Pro Ile Ser Glu Lys Ala Lys Glu Ile
        435                 440                 445
Lys Trp Val Glu Thr Asp Glu Val Lys Asp Asn Gly Asp Lys Arg Lys
    450                 455                 460
Lys Ile Gln Ile Pro Glu Gly Leu Ile Thr Cys Ala Val Asn Ile Gly
465                 470                 475                 480
Ile Arg Asn Leu Gly Phe Gly Thr Leu Ala Val Gly Glu Lys Gly Lys
                485                 490                 495
Ala Arg Ile Leu Arg Ser Arg Asn Ile Trp Ile Gly Glu Glu Glu Lys
            500                 505                 510
Thr Gly Lys His Lys Gly Arg Trp Gln Lys Gly Pro Thr Leu Arg Glu
        515                 520                 525
Leu Arg Asp His Lys Gln Glu Ile Arg Arg Leu Arg Lys Leu Arg Gly
```

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

[0142]

<210> 330
<211> 792
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-温泉-热泉或热泉宏基因组
序列
<400> 330
Met Ala Thr Lys Ser Phe Glu Ala Lys Ile Val Cys Lys Pro Asp Glu
1 5 10 15
Lys Tyr Thr Ala Glu Gln Lys Lys Gln Phe Leu Trp Phe Thr His Gln
20 25 30
Val Phe Asn Asp Gly Val Arg Lys Val Ile Pro Tyr Val Phe Lys Met
35 40 45

[0143]

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

[0144]

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

Lys Asp Thr Lys Asn Phe Ile Trp Gln Gly Lys Asn Tyr Asn Trp Asp
755 760 765
Gln Ile Ala Asp Asp Val Gln Ser Phe Leu Asp Gln Lys Ala Gly Ile
770 775 780
Lys Lys Glu Asp Asp Ile Pro Tyr
785 790

<210> 331
<211> 789
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
生物反应器宏基因组序列
<400> 331

[0145]

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

[0146]

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

[0147]

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

<210> 332
<211> 679
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
生物反应器宏基因组 或实验室富集限定的培养基-离子液体和高固体富集的序列

<400> 332
Met Gly Ile Lys Ser Trp Arg Val Lys Leu Trp Val Glu Ser Pro Glu
1 5 10 15
Ile Arg Ala Thr Leu Glu Tyr Cys Leu Ser Leu Trp Asn Thr Ser Ala
20 25 30
Ala Leu Leu Cys Arg His Val Leu Ala Met Arg Arg Gly Asp Phe Gly
35 40 45
Pro Glu Gly Gln Ala Ile Trp Glu His Leu Lys Lys Thr Pro His Asn
50 55 60
Phe His Val Leu Tyr Pro Ile Ser Asn Gly Ser Gly Ile Glu Lys Thr

65 70 75 80
Thr Arg Tyr Tyr Pro Asp Ala Arg Ala Phe Val Asn Lys Tyr Gly Tyr
85 90 95
Gly Ala Leu Gly Pro Val Arg Ser Ala Ser Gly Cys Leu Trp Lys Gln
100 105 110
Leu Thr Lys Ala Val Arg Glu Arg Thr His Ser Tyr Asp Lys Thr Cys
115 120 125
Glu Lys Trp Ser Glu Glu His Ser Lys Trp Leu Glu Glu Arg Ala Lys
130 135 140

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

[0148]

[0149]

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

Ala Leu Asp Glu Ile Gln Gln Arg Ala Val Glu Ile Leu Ala Ser Arg
            660                 665                 670
Arg Ala Gly Asp Thr Pro Phe
        675
<210> 333
<211> 757
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
    生物反应器宏基因组序列
<400> 333
Met His Leu Trp Arg Thr His Cys Val Phe Asn Gln Arg Leu Pro Ala
```

[0150]

```
1               5                   10                  15
Leu Leu Lys Arg Leu Phe Ala Met Arg Arg Gly Glu Val Gly Gly Asn
            20                  25                  30
Glu Ala Gln Arg Gln Val Tyr Gln Arg Val Ala Gln Phe Val Leu Ala
        35                  40                  45
Arg Asp Ala Lys Asp Ser Val Asp Leu Leu Asn Ala Val Ser Leu Arg
    50                  55                  60
Lys Arg Ser Ala Asn Ser Ala Phe Lys Lys Lys Ala Thr Ile Leu Arg
65                  70                  75                  80
Asn Gly Gln Arg Arg Glu Val Thr Gly Glu Glu Val Phe Ala Glu Ala
                85                  90                  95
Val Ala Leu Ala Ser Lys Gly Val Phe Ala Tyr Asp Lys Ala Asp Met
            100                 105                 110
Arg Ala Gly Leu Pro Asp Ser Leu Phe Gln Pro Leu Thr Arg Asp Ala
        115                 120                 125
Val Ala Cys Met Arg Ser His Glu Glu Leu Val Glu Thr Trp Lys Lys
    130                 135                 140
Glu Cys Arg Glu Trp Leu Asp Arg Lys Ser Lys Trp Glu Ala Glu Pro
145                 150                 155                 160
Glu His Ala Leu Tyr Leu Lys Leu Arg Pro Lys Phe Glu Lys Gly Glu
                165                 170                 175
Ala Glu Arg Gly Gly Arg Phe Arg Lys Arg Ala Glu Arg Asp Arg Ala
            180                 185                 190
Tyr Leu Asp Trp Leu Glu Ala Asn Pro Gln Leu Ala Ala Trp Arg Gly
        195                 200                 205
Lys Ala Pro Pro Ala Val Val Pro Ile Asp Glu Ala Gly Lys Arg Arg
    210                 215                 220
Ile Ala Arg Ala Lys Ala Trp Lys Gln Ala Ser Val Arg Ala Glu Glu
225                 230                 235                 240
Phe Trp Lys Arg Asn Pro Glu Leu His Ala Leu His Lys Ile His Val
                245                 250                 255
Gln Tyr Leu Arg Glu Phe Val Arg Phe Lys Gln Arg Pro Thr Phe Thr
            260                 265                 270
Met Pro Asp Pro Val Arg His Pro Arg Trp Cys Leu Phe Asn Ala Pro
        275                 280                 285
Gln Thr Ser Pro Gln Gly Tyr Arg Leu Leu Arg Leu Pro Gln Ser Arg
    290                 295                 300
Arg Thr Val Gly Ser Val Glu Leu Arg Leu Leu Thr Gly Pro Ser Asp
305                 310                 315                 320
Gly Ala Gly Phe Pro Asp Ala Trp Val Asn Val Arg Phe Lys Ala Asp
                325                 330                 335
Pro Arg Leu Ala Gln Leu Arg Pro Val Lys Val Ala Arg Thr Val Thr
            340                 345                 350
Arg Gly Lys Asn Lys Gly Ala Lys Val Glu Ala Asp Gly Phe Arg Tyr
```

[0151]

355 360 365
Tyr Asp Asp Gln Leu Leu Ile Glu Arg Asp Ala Gln Val Ser Gly Val
370 375 380
Lys Leu Leu Phe Arg Asp Ile Arg Met Ala Pro Phe Ala Asp Lys Pro
385 390 395 400
Ile Glu Asp Arg Leu Leu Ser Ala Thr Pro Tyr Leu Val Phe Ala Val
405 410 415
Glu Ile Lys Asp Glu Ala Arg Thr Glu Arg Ala Lys Ala Ile Arg Phe
420 425 430
Asp Glu Thr Ser Glu Leu Thr Lys Ser Gly Lys Lys Arg Lys Thr Leu
435 440 445
Pro Ala Gly Leu Val Ser Val Ala Val Asp Leu Asp Thr Arg Gly Val
450 455 460
Gly Phe Leu Thr Arg Ala Val Ile Gly Val Pro Glu Ile Gln Gln Thr
465 470 475 480
His His Gly Val Arg Leu Leu Gln Ser Arg Tyr Val Ala Val Gly Gln
485 490 495
Val Glu Ala Arg Ala Ser Gly Glu Ala Glu Trp Ser Pro Gly Pro Asp
500 505 510
Leu Ala His Ile Ala Arg His Lys Arg Glu Ile Arg Arg Leu Arg Gln
515 520 525
Leu Arg Gly Lys Pro Val Lys Gly Glu Arg Ser His Val Arg Leu Gln
530 535 540
Ala His Ile Asp Arg Met Gly Glu Asp Arg Phe Lys Lys Ala Ala Arg
545 550 555 560
Lys Ile Val Asn Glu Ala Leu Arg Gly Ser Asn Pro Ala Ala Gly Asp
565 570 575
Arg Tyr Thr Arg Ala Asp Val Leu Leu Tyr Glu Ser Leu Glu Thr Leu
580 585 590
Leu Pro Asp Ala Glu Arg Glu Arg Gly Ile Asn Arg Ala Leu Leu Arg
595 600 605
Trp Asn Arg Ala Lys Leu Ile Glu His Leu Lys Arg Met Cys Asp Asp
610 615 620
Ala Gly Ile Arg His Phe Pro Val Ser Pro Phe Gly Thr Ser Gln Val
625 630 635 640
Cys Ser Lys Cys Gly Ala Leu Gly Arg Arg Tyr Ser Leu Ala Arg Glu
645 650 655
Asn Gly Arg Ala Val Ile Arg Phe Gly Trp Val Glu Arg Leu Phe Ala
660 665 670
Cys Pro Asn Pro Glu Cys Pro Gly Arg Arg Pro Asp Arg Pro Asp Arg
675 680 685
Pro Phe Thr Cys Asn Ser Asp His Asn Ala Ser Val Asn Leu His Arg
690 695 700
Val Phe Ala Leu Gly Asp Gln Ala Val Ala Ala Phe Arg Ala Leu Pro

705 710 715 720
Pro Arg Asp Ser Pro Ala Arg Thr Leu Ala Val Lys Arg Val Glu Asp
725 730 735
Thr Leu Arg Pro Gln Leu Met Arg Val His Lys Leu Ala Asp Ala Gly
740 745 750
Val Asp Ser Pro Phe
755

<210> 334
<211> 769
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
生物反应器宏基因组序列

[0152]

<400> 334
Met His Leu Trp Arg Thr His Cys Val Phe Asn Gln Arg Leu Pro Ala
1 5 10 15
Leu Leu Lys Arg Leu Phe Ala Met Arg Arg Gly Glu Val Gly Gly Asn
20 25 30
Glu Ala Gln Arg Gln Val Tyr Gln Arg Val Ala Gln Phe Val Leu Ala
35 40 45
Arg Asp Ala Lys Asp Ser Val Asp Leu Leu Asn Ala Val Ser Leu Arg
50 55 60
Lys Arg Ser Ala Asn Ser Ala Phe Lys Lys Lys Ala Thr Ile Ser Cys
65 70 75 80
Asn Gly Gln Ala Arg Glu Val Thr Gly Glu Glu Val Phe Ala Glu Ala
85 90 95
Val Ala Leu Ala Ser Lys Gly Val Phe Ala Tyr Asp Lys Asp Asp Met
100 105 110
Arg Ala Gly Leu Pro Asp Ser Leu Phe Gln Pro Leu Thr Arg Asp Ala
115 120 125
Val Ala Cys Met Arg Ser His Glu Glu Leu Val Ala Thr Trp Lys Lys
130 135 140
Glu Tyr Arg Glu Trp Arg Asp Arg Lys Ser Glu Trp Glu Ala Glu Pro
145 150 155 160
Glu His Ala Leu Tyr Leu Asn Leu Arg Pro Lys Phe Glu Glu Gly Glu
165 170 175
Ala Ala Arg Gly Gly Arg Phe Arg Lys Arg Ala Glu Arg Asp His Ala
180 185 190
Tyr Leu Asp Trp Leu Glu Ala Asn Pro Gln Leu Ala Ala Trp Arg Gly
195 200 205
Lys Ala Pro Pro Ala Val Val Pro Ile Asp Glu Ala Gly Lys Arg Arg
210 215 220
Ile Ala Arg Ala Lys Ala Trp Lys Gln Ala Ser Val Arg Ala Glu Glu

225 230 235 240
Phe Trp Lys Arg Asn Pro Glu Leu His Ala Leu His Lys Ile His Val
245 250 255
Gln Tyr Leu Arg Glu Phe Val Arg Pro Arg Arg Thr Arg Arg Asn Lys
260 265 270
Arg Arg Glu Gly Phe Lys Gln Arg Pro Thr Phe Thr Met Pro Asp Pro
275 280 285
Val Arg His Pro Arg Trp Cys Leu Phe Asn Ala Pro Gln Thr Ser Pro
290 295 300
Gln Gly Tyr Arg Leu Leu Arg Leu Pro Gln Ser Arg Arg Thr Val Gly
305 310 315 320
Ser Val Glu Leu Arg Leu Leu Thr Gly Pro Ser Asp Gly Ala Gly Phe
325 330 335
Pro Asp Ala Trp Val Asn Val Arg Phe Lys Ala Asp Pro Arg Leu Ala
340 345 350
Gln Leu Arg Pro Val Lys Val Pro Arg Thr Val Thr Arg Gly Lys Asn
355 360 365
Lys Gly Ala Lys Val Glu Ala Asp Gly Phe Arg Tyr Tyr Asp Asp Gln
370 375 380
Leu Leu Ile Glu Arg Asp Ala Gln Val Ser Gly Val Lys Leu Leu Phe
385 390 395 400
[0153] Arg Asp Ile Arg Met Ala Pro Phe Ala Asp Lys Pro Ile Glu Asp Arg
405 410 415
Leu Leu Ser Ala Thr Pro Tyr Leu Val Phe Ala Val Glu Ile Lys Asp
420 425 430
Glu Ala Arg Thr Glu Arg Ala Lys Ala Ile Arg Phe Asp Glu Thr Ser
435 440 445
Glu Leu Thr Lys Ser Gly Lys Lys Arg Lys Thr Leu Pro Ala Gly Leu
450 455 460
Val Ser Val Ala Val Asp Leu Asp Thr Arg Gly Val Gly Phe Leu Thr
465 470 475 480
Arg Ala Val Ile Gly Val Pro Glu Ile Gln Gln Thr His His Gly Val
485 490 495
Arg Leu Leu Gln Ser Arg Tyr Val Ala Val Gly Gln Val Glu Ala Arg
500 505 510
Ala Ser Gly Glu Ala Glu Trp Ser Pro Gly Pro Asp Leu Ala His Ile
515 520 525
Ala Arg His Lys Arg Glu Ile Arg Arg Leu Arg Gln Leu Arg Gly Lys
530 535 540
Leu Val Lys Gly Glu Arg Ser His Val Arg Leu Gln Ala His Ile Asp
545 550 555 560
Arg Met Gly Glu Asp Arg Phe Lys Lys Ala Ala Arg Lys Ile Val Asn
565 570 575
Glu Ala Leu Arg Gly Ser Asn Pro Ala Ala Gly Asp Pro Tyr Thr Arg

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

[0154]

<210> 335
<211> 769
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
生物反应器-厌氧-厌氧沼气反应器序列
<400> 335
Met His Leu Trp Arg Thr His Cys Val Phe Asn Gln Arg Leu Pro Ala
1 5 10 15
Leu Leu Lys Arg Leu Phe Ala Met Arg Arg Gly Glu Val Gly Gly Asn
20 25 30
Glu Ala Gln Arg Gln Val Tyr Gln Arg Val Ala Gln Phe Val Leu Ala
35 40 45
Arg Asp Ala Lys Asp Ser Val Asp Leu Leu Asn Ala Val Ser Leu Arg
50 55 60
Lys Arg Ser Ala Asn Ser Ala Phe Lys Lys Lys Ala Thr Ile Ser Cys
65 70 75 80
Asn Gly Gln Ala Arg Glu Val Thr Gly Glu Glu Val Phe Ala Glu Ala

[0155]

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

435 440 445

Glu Leu Thr Lys Ser Gly Lys Lys Arg Lys Thr Leu Pro Ala Gly Leu
450 455 460

Val Ser Val Ala Val Asp Leu Asp Thr Arg Gly Val Gly Phe Leu Thr
465 470 475 480

Arg Ala Val Ile Gly Val Pro Glu Ile Gln Gln Thr His His Gly Val
485 490 495

Arg Leu Leu Gln Ser Arg Tyr Val Ala Val Gly Gln Val Glu Ala Arg
500 505 510

Ala Ser Gly Glu Ala Glu Trp Ser Pro Gly Pro Asp Leu Ala His Ile
515 520 525

Ala Arg His Lys Arg Glu Ile Arg Arg Leu Arg Gln Leu Arg Gly Lys
530 535 540

Pro Val Lys Gly Glu Arg Ser His Val Arg Leu Gln Ala His Ile Asp
545 550 555 560

Arg Met Gly Glu Asp Arg Phe Lys Lys Ala Ala Arg Lys Ile Val Asn
565 570 575

Glu Ala Leu Arg Gly Ser Asn Pro Ala Ala Gly Asp Pro Tyr Thr Arg
580 585 590

Ala Asp Val Leu Leu Tyr Glu Ser Leu Glu Thr Leu Leu Pro Asp Ala
595 600 605

[0156] Glu Arg Glu Arg Gly Ile Asn Arg Ala Leu Leu Arg Trp Asn Arg Ala
610 615 620

Lys Leu Ile Glu His Leu Lys Arg Met Cys Asp Asp Ala Gly Ile Arg
625 630 635 640

His Phe Pro Val Ser Pro Phe Gly Thr Ser Gln Val Cys Ser Lys Cys
645 650 655

Gly Ala Leu Gly Arg Arg Tyr Ser Leu Ala Arg Glu Asn Gly Arg Ala
660 665 670

Val Ile Arg Phe Gly Trp Val Glu Arg Leu Phe Ala Cys Pro Asn Pro
675 680 685

Glu Cys Pro Gly Arg Arg Pro Asp Arg Pro Asp Arg Pro Phe Thr Cys
690 695 700

Asn Ser Asp His Asn Ala Ser Val Asn Leu His Arg Val Phe Ala Leu
705 710 715 720

Gly Asp Gln Ala Val Ala Ala Phe Arg Ala Leu Ala Pro Arg Asp Ser
725 730 735

Pro Ala Arg Thr Leu Ala Val Lys Arg Val Glu Asp Thr Leu Arg Pro
740 745 750

Gln Leu Met Arg Val His Lys Leu Ala Asp Ala Gly Val Asp Ser Pro
755 760 765

Phe

<210> 336

<211> 638
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
生物反应器-柳枝稷降解序列
<400> 336

[0157]

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

290 295 300

His Thr Val Leu Ala Gln Pro Gln Glu Val Arg Leu Ile Arg Arg Gly
305 310 315 320
Ser Ser Tyr Phe Ala Thr Ile Arg Leu Ser Leu His Ser Ala Pro Ser
325 330 335
Ala Ile Pro Val Pro Gln Thr Ala Ala Ser Ser Pro Ala Gly Met Leu
340 345 350
Arg Lys Ile Arg Thr Asp Ala Pro Glu Leu Lys Leu Ile Thr Cys Ala
355 360 365
Val Asp Leu Gly Val Arg His Leu Gly Ala Ala Thr Ile Ala Arg Glu
370 375 380
Gly Glu Ile Ile Gly Arg Arg Ile Leu His Asn Arg Phe Leu Ala Pro
385 390 395 400
Gly Ala Ala Gly Gly Val Ile Asn Ile Pro Ser Leu Pro Gln Ile Ala
405 410 415
Gln Val Arg Arg Gln Ile Arg Cys Ala Gln Arg Arg Ser Gly Arg Ile
420 425 430
Ser Pro Lys Val Ala Gly Cys Arg His Leu Trp Thr His Tyr Arg His
435 440 445
Leu Cys Glu Asp Arg Tyr Lys Lys Ala Val Ala Ala Ile Phe Gly Phe
450 455 460

[0158]

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

<210> 337

<211> 778
<212> PRT
<213> 未知

<220>
<223> 未知的描述:
洞穴宏基因组序列
<400> 337

[0159]

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

[0160]

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

Ala Lys Asp Ala Gly Leu Lys Leu Phe Glu Val His Pro Ala Gly Thr
645 650 655
Ser Gln Cys Cys Ser Lys Cys Gly Ala Leu Gly Arg Arg Tyr Ser Ile
660 665 670
Val Arg Asp Asp Glu Thr Arg Ala Pro Val Ile Arg Phe Gly Trp Val
675 680 685
Glu Lys Leu Phe Ala Cys Pro Asn Leu Asn Cys Ala Gly Arg Lys Ser
690 695 700
Met His Cys Asp Gln Pro Phe Val Cys Asn Ala Asp His Asn Ala Ser
705 710 715 720
Val Asn Leu His Arg Arg Phe Val Met Asp Asp Lys Ala Val Ala Ala
725 730 735
Phe Phe Ala Trp Lys Ala Arg Pro Asp Gly Glu Arg Glu Gln Ser Val
740 745 750
Arg Ser Ile Asp Asp Ala Leu Arg Pro Gln Leu Glu Arg Ile His Gly
755 760 765
Leu Ser Ala Ser Pro Leu Glu Thr Pro Phe
770 775

<210> 338
<211> 743
<212> PRT
<213> 未知

[0161]

<220>
<223> 未知的描述:
淡水沉积物宏基因组序列
<400> 338
Met Ser Val Arg Ser Phe Gln Ala Lys Val Glu Cys Asp Lys Gln Ile
1 5 10 15
Met Glu His Leu Trp Arg Thr His Thr Val Phe Asn Glu Arg Leu Pro
20 25 30
Val Ile Ile Lys Ile Leu Phe Gln Met Lys Arg Gly Glu Cys Gly Gln
35 40 45
Asn Asp Lys Gln Lys Ala Leu Tyr Gln Asn Ile Ala Lys Ser Ile Leu
50 55 60
Asp Thr Asn Ala Gln Asn Ala Asp Tyr Leu Leu Asn Ser Val Ser Ile
65 70 75 80
Lys Gly Trp Lys Pro Gly Ala Ala Arg Lys Tyr Asn Lys Gly Ser Phe
85 90 95
Thr Trp Leu Asp Glu Ala Ala Asn Leu Ser Ser Lys Gly Ile Leu Val
100 105 110
Tyr Asp Lys Asp Gln Val Leu Gly Asp Ile Pro Leu Met Met Arg Gln
115 120 125
Met Ile Cys Arg Gln Ser Val Glu Thr Ile Lys Ser Gln Gln Glu Ile
130 135 140

[0162]

```
Leu Lys Asn Trp Glu Asn Glu Tyr Lys Glu Trp Leu Lys Lys Lys Glu
145                 150                 155                 160
Glu Trp Glu Ser Gln Glu Asp Asn Lys Lys Tyr Leu Ala Leu Arg Asp
                165                 170                 175
Lys Phe Glu Gln Phe Arg Lys Ser Tyr Gly Gly Lys Lys Ser Lys Lys
            180                 185                 190
Arg Gly Arg Trp His Leu Tyr Leu Lys Trp Leu Arg Asp Asn Pro Asp
        195                 200                 205
Leu Ala Ala Trp Arg Gly Gly Lys Ala Glu Ile Asn Pro Ile Ser Glu
    210                 215                 220
Lys Ala Gln Ser Arg Ile Asn Lys Ala Lys Pro Asn Lys Lys Asn Ala
225                 230                 235                 240
Ile Glu Leu Glu Glu Phe Leu Glu Ala Asn Pro Glu Ile Lys Lys Leu
                245                 250                 255
Asp Glu Leu His Gly Tyr Tyr Glu Arg Thr Phe Ile Arg Arg Arg Glu
            260                 265                 270
Thr Arg Lys Gly Lys Glu Leu Thr Gly Phe Lys Gln Lys Pro Thr Phe
        275                 280                 285
Thr Leu Pro His Ser Thr Thr His Pro Arg Trp Phe Val Phe Asn Ala
    290                 295                 300
Pro Gln Thr Asn Pro Ala Gly Tyr Arg Asn Leu Ile Leu Pro Gln Lys
305                 310                 315                 320
Thr Gly Ala Leu Gly Ser Leu Glu Met Arg Leu Leu Asn Gly Asp Lys
                325                 330                 335
Asn Asn Gly Asp Tyr Pro Asp Asp Trp Ile Asn Val Arg Phe Lys Ala
            340                 345                 350
Asp Pro Arg Leu Ala Leu Met Arg Gly Ile Lys Ile Gln Arg Ile Val
        355                 360                 365
Arg Arg Gly Lys Asp Ala Gly Gln Thr Lys Glu Ser Asp Ser Tyr Lys
    370                 375                 380
Tyr Phe Asp Lys Phe Leu Asn Lys Trp Arg Pro Ala Thr Leu Ser Gly
385                 390                 395                 400
Val Lys Leu Ile Phe Lys Phe Asn Pro Asp Lys Ser Val Asn Thr Ala
                405                 410                 415
Tyr Leu Tyr Phe Thr Cys Asp Ile Thr Asn Glu Pro Leu Thr Glu Thr
            420                 425                 430
Ala Lys Lys Ile Gln Trp Ile Glu Thr Gly Glu Val Thr Lys Lys Gly
        435                 440                 445
Lys Ile Arg Lys Lys Lys Val Leu Pro Glu Gly Leu Ile Ser Cys Ala
    450                 455                 460
Val Asp Leu Gly Met Arg Lys Gly Thr Thr Gly Phe Gly Thr Leu Cys
465                 470                 475                 480
Lys Tyr Lys Asn Gly Lys Pro Tyr Ile Leu Arg Ser Arg Asn Leu Trp
                485                 490                 495
```

[0163]

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

<210> 339
<211> 785
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
热液喷口宏基因组序列
<400> 339
Met Leu Val Leu Pro His Gly Ala His Gly Ala Ala Ser Ser Ile Ile
1 5 10 15
Cys Pro Phe Gly Ile Lys Asn Ile Ser Ala Lys Trp Asp Ser Pro Glu
20 25 30

[0164]

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

[0165]

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

Asn Leu Leu Asn Ile Val Leu Leu Asn Lys Arg Met Thr Pro Glu Asp
740 745 750
Lys Thr Asn Leu Ser Pro Ala Glu Lys Arg Asn Tyr Tyr Thr Gln Leu
755 760 765
Asp Asp Leu Ala Met Lys Lys Leu Ser Pro Gln Lys Ser Pro Val Lys
770 775 780
Val
785

<210> 340
<211> 742
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
热液喷口宏基因组序列
<400> 340

[0166]

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

[0167]

Glu Arg Lys Arg Ser Gln Arg Lys Arg Ser Leu Arg Trp Glu Arg Trp
225 230 235 240
Asp Lys Tyr Val Asn Leu Leu Ile Asp Asn Ser Glu Leu Met Gln Trp
245 250 255
Arg Glu Lys Asn Thr Thr Leu Thr Ile Pro Pro Ala Leu Asn Asp Pro
260 265 270

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

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

[0168]

<210> 341
<211> 674
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
实验室富集限定的培养基-离子液体和高固体
富集的序列
<400> 341
Met Gly Val Lys Ala Trp Lys Ala Lys Ile Ile Val Ala Asp Pro Asp
1 5 10 15
Leu Ala Ala Ala Leu Asp His Thr Val Arg Gln Trp Asn Leu Trp Ala
20 25 30
Glu Gln Ala Cys Ile Ala Ile Met Arg Leu Arg Arg Gly Asp Tyr Gly
35 40 45
Asp Glu Gly Gln Ala Val Trp Arg Phe Leu Ser Glu Lys Gly Thr Phe
50 55 60
Gly Thr Ser Ile Leu Asp Lys Ile Thr Arg Val Thr Gly Arg Gly Ile
65 70 75 80
Lys Asn Asp Glu Glu Arg Thr Ile Ala Ser Asn Phe Leu Ser Arg His
85 90 95

Gly Pro Phe Glu Pro Leu Arg Ala Ala Arg Asp Ala Leu Gly Pro Asn
100 105 110
Phe Trp Tyr Gln Phe Gln Arg Ala Val Arg Glu Arg Ile Glu Ser His
115 120 125
Asp Ala Lys Leu Arg Gln Trp Arg Glu Asp His Ala Glu Trp Ile Glu
130 135 140
Thr Arg Arg Thr Trp Glu Glu Glu His Arg Ala Tyr Met Asn Val Arg
145 150 155 160
Pro Val Ile Glu Ala Phe Glu Ala Ala Glu Gly Gln Ser Gly Arg Arg
165 170 175
Arg Gly Arg Trp Arg Arg Trp Leu Asp Phe Leu Ser Ser His Pro Glu
180 185 190
Leu Ala Ala Trp Arg Gly Gly Pro Ala Thr Ile Val Pro Leu Ser Lys
195 200 205

Glu Glu Leu Ala Ala Cys Arg Arg Arg Arg Arg Arg Ala Val Ala Glu
210 215 220
Gln Ala Arg Leu Phe Phe Glu Lys Asn Pro Glu Leu Arg Ala Leu Asn
225 230 235 240
Arg Glu His Asp Glu Tyr Gln Glu Lys Phe Ala Arg Pro Trp Ala Lys
245 250 255
[0169] Arg Arg Asn Pro Asp Gly Phe Arg His Arg Pro Thr Phe Thr Leu Pro
260 265 270
His Pro Asp Ser His Pro Asp Trp Pro Arg Phe Lys Gly Asn Ala Gly
275 280 285
Trp Arg Asp Leu Ser Val Pro Asp Arg Thr Val Ser Leu Glu Leu Leu
290 295 300
Cys Pro Asp Gly Ser Arg Arg Trp Phe Asn Val Arg Phe Val Pro Asp
305 310 315 320
Pro Arg Leu Ala Ser Leu Arg Pro Ala Thr Gln Pro Val Arg Gln Gly
325 330 335
Asn Thr Thr Tyr His Tyr Val Ser Asp Ala Ile Gly Val Ala Glu Gln
340 345 350
Pro Val Gln Pro Gln Gly Ile Lys Leu Ile Leu Arg Pro Asp Gly Ala
355 360 365
Tyr Leu Ser Val Ser Val Ser Val Leu Ala Pro Pro Cys Gly Phe Asp
370 375 380
Val Arg Gln Ser Ala Ile Glu Lys Tyr Gly Ser Ile Trp Ala Cys Arg
385 390 395 400
Lys Tyr Lys Gln Asp His Pro Asp Arg Pro Leu Val Thr Cys Ala Val
405 410 415
Asp Leu Ala Val Arg His Asn Gly Ala Leu Thr Val Ala Arg Asp Gly
420 425 430
Glu Val Phe Ala Arg Arg Leu Leu His Asn Arg Phe Phe Leu Pro Gly

435 440 445

Asp Asp Arg Glu Arg Ala Met Asn Ile Pro Ser Leu Leu Glu Ile Ala
450 455 460
Ala Val Lys Arg Gln Leu Arg Arg Ala Arg Arg Gln Thr Gly Arg Ile
465 470 475 480
Ala Pro Gly Lys Pro Ser Cys Ala Arg Leu Gln Ala His Tyr Arg Ala
485 490 495
Leu Cys Glu Asp Arg Tyr Lys Lys Leu Val Ala Ala Ile Phe Ala Tyr
500 505 510
Ala Arg His His Gly Ala Glu Val Val Ile Phe Glu Asp Leu Lys Asn
515 520 525
Leu Leu Pro Asn Ala Val Asn Glu Arg Gly Val Asn Arg Ala Leu Gln
530 535 540
Ser Trp Asn Arg Gly Ala Ile Val Arg Phe Ala Arg Gln Ser Ala Glu
545 550 555 560
Asp Val Gly Leu Arg Val Ala Thr Val Pro Ala Phe His Thr Ser Ile
565 570 575
Val Cys Ala Arg Cys Gly Ala Leu Gly Cys Arg Phe Glu Gln Thr Arg
580 585 590
Glu Gly Val Arg Ile Ala Lys Leu Gly Pro Trp Phe Gly Cys Arg Ala
595 600 605
[0170] Cys Gly Arg Arg Ile His Ala Asp Leu Asn Ala Ser Glu Asn Leu His
610 615 620
Lys Val Leu Ile Gly Thr Phe Pro Gln Val Arg Arg Ile Ser Arg Asp
625 630 635 640
Pro Pro Val Leu Ser Ile Ala Asp Val Arg Val Asp Gln Arg Glu Ala
645 650 655
Asp Arg Leu Ala Ala Ala Trp Leu Ser Glu Leu Ser Val Gly Pro Thr
660 665 670
Pro Phe

<210> 342
<211> 638
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
实验室富集限定的培养基-离子液体和高固体
富集的序列
<400> 342
Met Arg Arg Gly Asp Phe Gly Pro Glu Gly Gln Ala Ile Trp Glu His
1 5 10 15
Leu Lys Lys Thr Pro His Asn Phe His Val Leu Tyr Pro Ile Ser Asn
20 25 30

[0171]

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

Val Pro Asp Pro Arg Leu Gln Arg Leu Ala Arg Leu Lys Gln Pro Val
        275                 280                 285
Lys Glu Gly Ala Thr Arg Tyr Asp Tyr Ser Leu Asp Met Pro Ser Gly
    290                 295                 300
His Thr Val Leu Ala Gln Pro Gln Glu Val Arg Leu Ile Arg Arg Gly
305                 310                 315                 320
Ser Ser Tyr Phe Ala Thr Ile Arg Leu Ser Leu His Ser Ala Pro Ser
                325                 330                 335
Ala Ile Pro Val Pro Gln Thr Ala Ala Ser Ser Pro Ala Gly Met Leu
            340                 345                 350
Arg Lys Ile Arg Thr Asp Ala Pro Glu Leu Lys Leu Ile Thr Cys Ala
        355                 360                 365
Val Asp Leu Gly Val Arg His Leu Gly Ala Ala Thr Ile Ala Arg Glu
```

370 375 380
Gly Glu Ile Ile Gly Arg Arg Ile Leu His Asn Arg Phe Leu Ala Pro
385 390 395 400
Gly Ala Asp Gly Gly Val Ile Asn Ile Pro Ser Leu Pro Gln Ile Ala
405 410 415
Gln Val Arg Arg Gln Ile Arg Cys Ala Gln Arg Arg Ser Gly Arg Ile
420 425 430
Ser Pro Lys Val Ala Gly Cys Arg His Leu Trp Thr His Tyr Arg His
435 440 445
Leu Cys Glu Asp Arg Tyr Lys Lys Ala Val Ala Ala Ile Phe Gly Phe
450 455 460
Ala Arg Ala His Gly Ala His Ile Val Leu Phe Glu Asp Leu Lys Met
465 470 475 480
Leu Asn Pro Asp Ser Ala Asn Glu Arg Gly Val Asn Arg Ala Leu Gln
485 490 495
Asn Trp Asn Arg Gly Arg Ile Val Glu Phe Ala Lys Asn Thr Ala Glu
500 505 510
Asp Ala Gly Leu Cys Val Val Thr Val Pro Ala Tyr Trp Thr Ser Arg
515 520 525
Ile Cys Ala Arg Cys Asn Ala Phe Gly Val Arg Phe Asn Gln Gly Gln
530 535 540
[0172] Arg Arg Lys Pro Gly Asp Asp Gly His Val Ser Arg Val Ala Cys Glu
545 550 555 560
Leu Asp Phe Leu Gly His Trp Phe Leu Cys Pro Ser Cys Gly Arg Ser
565 570 575
Val His Ala Asp Leu Asn Ala Ser Glu Asn Leu His Arg Val Phe Leu
580 585 590
Gly Thr Phe Pro Thr Val Gln Thr Ile Gln Arg Gln Pro Arg Ile Tyr
595 600 605
Glu Ile Asn Gly Gln Arg Val Ala Leu Asp Glu Ile Gln Gln Arg Ala
610 615 620
Val Glu Ile Leu Ala Ser Arg Arg Ala Gly Asp Thr Pro Phe
625 630 635

<210> 343
<211> 784
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
永久冻土宏基因组或菲西芬氏菌目（Phycisphaerales）细菌或陆生-土壤-沼地序列
<400> 343
Met Gly Val Arg Ala Leu Arg Ala Arg Ile Glu Cys Asp Arg Glu Thr
1 5 10 15

[0173]

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

[0174]

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

Ala Asp His Asn Ala Ser Leu Asn Leu His Arg Arg Phe Val Leu Asp
725 730 735
Asp Arg Ala Val Ala Ala Phe Phe Ala Trp Lys Asn Lys Thr Asn Ala
740 745 750
Glu Arg Glu Arg Glu Leu Thr Gln Met Asp Gln Ala Leu Arg Pro Glu
755 760 765
Leu Glu Lys Leu His Arg Leu Pro Leu Ala Glu Leu Ala Thr Pro Phe
770 775 780

<210> 344
<211> 722
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
永久冻土宏基因组序列

[0175]

<400> 344
Met Lys Arg Gly Asp Ser Gly Thr Thr Asp Ala Gln Lys His Leu Tyr
1 5 10 15
Ser Gln Ile Ala Arg Phe Val Leu Ala Arg Ser Ser Arg Asp Ala Pro
20 25 30
Tyr Leu Met Asn Ser Ile Ser Ile Arg Gly Trp Thr Pro Asn Thr Ala
35 40 45
Lys Lys Lys Met Lys Ala Thr Leu Met Gly Pro Asp Gly Gln Glu Val
50 55 60
Glu Thr Ser Gly Asp Asp Trp Ala Asp Ala Ala Ala Ala Leu Ser Ala
65 70 75 80
Glu Gly Thr Leu Leu Tyr Asp Lys Ser Glu Leu Leu Gly Asp Val Pro
85 90 95
Asp Thr Leu Arg Gln Val Val Val Arg Glu Cys Ala Ala Ile Ile Ser
100 105 110
Gly His Glu Glu Leu Val Arg Ala Trp Glu Leu Ser His Glu Glu Trp
115 120 125
Leu Glu Gln Lys Ala Lys Trp Glu Ser Asp Pro Glu His Glu Arg Tyr
130 135 140
Met Val Leu Arg Pro Gln Phe Glu Thr Phe Glu Gln Ser Val Gly Gly
145 150 155 160
Lys Ala Gly Lys Arg Arg Gly Arg Trp Gln Arg Tyr Leu Asp Trp Leu
165 170 175
Arg Ser Asn Pro Glu Leu Ala Ala Trp Arg Gly Gly Pro Ala Thr Val
180 185 190
Val Glu Leu Pro Ala Ala Ala Arg Glu Arg Val Arg Lys Ala Arg Pro
195 200 205
Ser Lys Val Arg Ser Val Glu Ala Glu Glu Phe Trp Lys Val Asn Pro
210 215 220

[0176]

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

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

[0177]

<210> 345
<211> 762
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
永久冻土宏基因组序列
<400> 345
Met Ser Thr Arg Ala Ile Gln Ser Arg Ile Gln Cys Asp His Gln Thr
1 5 10 15
Leu Glu His Leu Trp Arg Thr His Val Val Phe Asn Arg Arg Leu Pro
20 25 30
Pro Leu Ile Ser Leu Leu Phe Arg Met Lys Arg Gly Asp Ser Gly Thr
35 40 45
Thr Asp Ala Gln Lys His Leu Tyr Ser Gln Ile Ala Arg Phe Val Leu
50 55 60
Ala Arg Ser Ser Arg Asp Ala Pro Tyr Leu Met Asn Ser Ile Ser Ile
65 70 75 80
Arg Gly Trp Thr Pro Asn Thr Ala Lys Lys Lys Met Lys Ala Thr Leu
85 90 95
Met Gly Pro Asp Gly Gln Glu Val Glu Thr Ser Gly Asp Asp Trp Ala
100 105 110
Asp Ala Ala Ala Ala Leu Ser Ala Glu Gly Thr Leu Leu Tyr Asp Lys
115 120 125

[0178]

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

[0179]

```
His Val Gly Phe Ala Thr Ile Ala Ala Lys Asp Glu Ala His Pro Glu
                485                 490                 495
Gly Leu Arg Val Leu Arg Ser Arg Asn Val Trp Met Gly His Glu Glu
            500                 505                 510
Gln Lys Gly Arg His Pro Gly Arg Trp Ser Pro Gly Pro Glu Leu Pro
        515                 520                 525
His Leu Ala Ser His Lys Arg Ser Leu Arg Arg Leu Arg Arg Leu Arg
    530                 535                 540
Gly Lys Pro Val Lys Asp Glu Ser Ser His Ala Glu Leu Gln Gly His
545                 550                 555                 560
Ile Thr Asn Met Ala Asn Asp Arg Phe Lys Lys Ala Ala Arg Glu Ile
                565                 570                 575
Val Asn Phe Ala Leu Asn Val Glu Arg Lys Val Asp Pro Arg Thr Gly
            580                 585                 590
Glu Pro Tyr Ser Arg Ala Asp Leu Leu Val Val Glu Asn Leu Ala Asn
        595                 600                 605
Leu Ile Pro Asp Ser Glu Arg Glu Arg Gly Ile Asn Arg Ala Leu Ile
    610                 615                 620
Glu Phe Asn Arg Gly His Leu Val Asp Arg Ile Ala Glu Val Ala Lys
625                 630                 635                 640
Asp Cys Gly Leu Lys Leu Met Lys Val Ser Pro Val Gly Thr Ser Gln
                645                 650                 655
Val Cys Cys Arg Cys Gly Ala Leu Gly Arg Arg Tyr Ser Ile Gln Arg
            660                 665                 670
Asp Arg Asp Ser Gly Tyr Ala Asp Ile Arg Phe Gly Phe Val Glu Ala
        675                 680                 685
Leu Phe Ala Cys Pro His Cys Gly Tyr Arg Ala Asn Ala Asp His Asn
    690                 695                 700
Ala Ser Ile Asn Leu His Arg Arg Phe Leu Leu Gly Glu Ala Ala Ile
705                 710                 715                 720
Arg Gly Phe Lys Glu Trp Gln Gln Gly Ser Arg Gln Glu Arg Ala Asp
                725                 730                 735
Thr Val Arg Arg Ile Glu Glu Ser Leu Val Ala Pro Leu Arg Leu Met
            740                 745                 750
His Arg Leu Glu Val Ala Asp Met Pro Phe
        755                 760
<210> 346
<211> 735
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
      永久冻土宏基因组或陆生-土壤-沼地序列
<400> 346
```

[0180]

```
Met Phe Asn Arg Arg Leu Pro Asp Leu Ile Ser Leu Leu Phe Arg Met
1               5                  10                  15
Arg Arg Gly Glu Cys Gly Glu Thr Glu Gln Ala Ser Arg Leu Tyr Ser
            20                  25                  30
Gln Val Ala Ser Phe Ile Leu Ala Arg Gly Ser Lys Asp Ala Pro Tyr
        35                  40                  45
Leu Leu Asn Ser Val Ser Val Arg Gly Trp Lys Pro His Thr Ala Lys
    50                  55                  60
Lys Met Lys Ala Arg Ala Leu Gly Pro Asp Gly Gln Gln Ile Glu Leu
65                  70                  75                  80
Ala Gly Glu Asp Trp Ala Asp Pro Ala Ala Ala Leu Ser Ala Glu Gly
                85                  90                  95
Arg Leu Leu Tyr Asp Lys Thr Glu Leu Leu Gly Asp Val Pro Asp Thr
            100                 105                 110
Leu Arg Gln Met Ala Ala Arg Glu Cys Val Ala Ile Ile Ser Gly His
        115                 120                 125
Glu Glu Leu Val Arg Ala Trp Glu Lys Ala His Glu Glu Trp Leu Ala
    130                 135                 140
Ser Lys Ala Lys Trp Glu Ser Asp Pro Glu Asn Gln Phe Tyr Leu Gly
145                 150                 155                 160
Leu Arg Ala Arg Phe Asp Glu Phe Glu Gln Ser Val Gly Gly Lys Val
                165                 170                 175
Gly Lys Arg Arg Gly Arg Trp Gln Arg Tyr Leu Asp Trp Leu Arg Ala
            180                 185                 190
Asn Pro Asp Leu Ala Ala Trp Arg Gly Gly Pro Ala Ala Val Val Glu
        195                 200                 205
Leu Ser Ala Gly Ala Lys Glu Arg Ile Arg Lys Ala Lys Pro Arg Lys
    210                 215                 220
Val Arg Ser Val Glu Ala Glu Glu Phe Trp Lys Val Asn Ser Glu Leu
225                 230                 235                 240
Ala Ala Leu Asp Arg Leu His Gly Tyr Tyr Glu Arg Glu Phe Val Arg
                245                 250                 255
Arg Arg Lys Thr Lys Lys His Pro Asp Gly Phe Asp His Arg Pro Thr
            260                 265                 270
Phe Thr Leu Pro His Pro Val Arg His Pro Arg Trp Phe Val Phe Asn
        275                 280                 285
Ala Pro Gln Thr Ser Pro Gln Gly Tyr Arg Lys Leu Cys Leu Pro Ser
    290                 295                 300
Ser Gly Gly His Thr Gly Thr Val Glu Leu Leu Leu Leu Thr Gly Glu
305                 310                 315                 320
Lys Ala Asp Asn Lys Tyr Pro Ser Gly Trp Glu Thr Val Arg Phe Ala
                325                 330                 335
Ala Asp Pro Arg Leu Ser Asp Phe Lys Ala Ile Thr Lys Arg Arg Gln
            340                 345                 350
```

[0181]

Val Leu Lys Gly Lys Thr Lys Gly Gln Glu Thr Asp Thr Lys Ala Phe
355 360 365
Thr Phe Leu Asp Arg Gln Ile Asn Leu Glu Arg Pro Ala Gln Ile Ser
370 375 380
Gly Val Lys Leu Ile Phe Lys Gly Val Arg Gln Asn His Asp Gly Ser
385 390 395 400
Leu Leu Ser Ala His Pro Tyr Leu Val Phe Thr Cys Asn Val Ala Asp
405 410 415
Val Pro Leu Thr Asp Gly Ala Lys Lys Ile Lys Trp Glu Asp Thr Gly
420 425 430
Glu Arg Arg Lys Arg Val Val Val Pro Ala Gly Leu Ile Ala Cys Ala
435 440 445
Val Asp Leu Gly Ile Arg His Val Gly Phe Ala Thr Val Ala Thr Gln
450 455 460
Asp Ala Glu Ser Gly Gly Leu His Val Leu Arg Ser Arg Asn Ile Trp
465 470 475 480
Ile Gly Gln Glu Glu Glu Lys Gly Arg Asn Pro Gly Arg Trp Ser Pro
485 490 495
Gly Pro Asp Leu Ala His Leu Ser Ala His Lys Arg Ser Leu Arg His
500 505 510
Leu Arg Gly Leu Arg Gly Lys Pro Val Lys Gly Glu Asn Ser His Ile
515 520 525
Glu Leu Gln Asp His Ile Thr Asn Met Ala Ser Asp Arg Phe Lys Lys
530 535 540
Ala Ala Arg Glu Ile Val Asn Phe Ala Leu Asn Ala Asp Ser Lys Val
545 550 555 560
Asp Pro Arg Thr Gly Glu Pro Tyr Pro Arg Ala Asp Val Leu Val Leu
565 570 575
Glu Asn Leu Ala Asn Leu Leu Pro Asp Ser Glu Arg Glu Arg Gly Ile
580 585 590
Asn Arg Ala Leu Ile Glu Phe Asn Arg Gly His Leu Ala Asp Arg Ile
595 600 605
Val Glu Val Ala Lys Asp Cys Gly Leu Lys Val Met Met Val Ser Pro
610 615 620
Val Gly Thr Ser Gln Val Cys Cys Arg Cys Gly Ala Leu Gly Arg Arg
625 630 635 640
Tyr Ser Leu Arg Arg His Gln Asp Gly Gly Tyr Thr Asp Ile His Phe
645 650 655
Gly Phe Val Glu Pro Leu Phe Ala Cys Pro Ser Cys Gly Tyr Arg Ala
660 665 670
Asn Ser Asp His Asn Ala Ser Val Asn Leu His Arg Arg Phe Arg Leu
675 680 685
Gly Glu Ala Ala Ile Ala Lys Phe Lys Glu Trp Gln Lys Gly Thr Lys
690 695 700

Pro Glu Arg Asp Glu Val Leu Arg Arg Ile Glu Gln Ser Leu Val Gly
705 710 715 720
Pro Leu Arg Glu Met His Arg Leu Ala Val Ala Asp Thr Pro Phe
725 730 735

<210> 347
<211> 760
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
永久冻土宏基因组或陆生-土壤-沼地序列

<400> 347

[0182]

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

[0183]

```
Glu Phe Trp Lys Val Asn Ser Glu Leu Ala Ala Leu Asp Arg Leu His
            260                 265                 270
Gly Tyr Tyr Glu Arg Glu Phe Val Arg Arg Arg Lys Thr Lys Lys His
        275                 280                 285
Pro Asp Gly Phe Asp His Arg Pro Thr Phe Thr Leu Pro His Pro Val
    290                 295                 300
Arg His Pro Arg Trp Phe Val Phe Asn Ala Pro Gln Thr Ser Pro Gln
305                 310                 315                 320
Gly Tyr Arg Lys Leu Cys Leu Pro Ser Ser Gly Gly His Thr Gly Thr
                325                 330                 335
Val Glu Leu Leu Leu Leu Thr Gly Glu Lys Ala Asp Asn Lys Tyr Pro
            340                 345                 350
Ser Gly Trp Glu Thr Val Arg Phe Ala Ala Asp Pro Arg Leu Ser Asp
        355                 360                 365
Phe Lys Ala Ile Thr Lys Arg Arg Gln Val Leu Lys Gly Lys Thr Lys
    370                 375                 380
Gly Gln Glu Thr Asp Thr Lys Ala Phe Thr Phe Leu Asp Arg Gln Ile
385                 390                 395                 400
Asn Leu Glu Arg Pro Ala Gln Ile Ser Gly Val Lys Leu Ile Phe Lys
                405                 410                 415
Gly Val Arg Gln Asn His Asp Gly Ser Leu Leu Ser Ala His Pro Tyr
            420                 425                 430
Leu Val Phe Thr Cys Asn Val Ala Asp Val Pro Leu Thr Asp Gly Ala
        435                 440                 445
Lys Lys Ile Lys Trp Glu Asp Thr Gly Glu Arg Arg Lys Arg Val Val
    450                 455                 460
Val Pro Ala Gly Leu Ile Ala Cys Ala Val Asp Leu Gly Ile Arg His
465                 470                 475                 480
Val Gly Phe Ala Thr Val Ala Thr Gln Asp Ala Glu Ser Gly Gly Leu
                485                 490                 495
His Val Leu Arg Ser Arg Asn Ile Trp Ile Gly Gln Glu Glu Glu Lys
            500                 505                 510
Gly Arg Asn Pro Gly Arg Trp Ser Pro Gly Pro Asp Leu Ala His Leu
        515                 520                 525
Ser Ala His Lys Arg Ser Leu Arg His Leu Arg Gly Leu Arg Gly Lys
    530                 535                 540
Pro Val Lys Gly Glu Asn Ser His Ile Glu Leu Gln Asp His Ile Thr
545                 550                 555                 560
Asn Met Ala Ser Asp Arg Phe Lys Lys Ala Ala Arg Glu Ile Val Asn
                565                 570                 575
Phe Ala Leu Asn Ala Asp Ser Lys Val Asp Pro Arg Thr Gly Glu Pro
            580                 585                 590
Tyr Pro Arg Ala Asp Val Leu Val Leu Glu Asn Leu Ala Asn Leu Leu
        595                 600                 605
```

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

[0184]

<210> 348
<211> 747
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
菲西芬氏菌纲（Phycisphaerae）细菌 ST-NAGAB-D1 序列
<400> 348
Met Ala Thr Lys Ser Tyr Arg Ala Arg Ile Leu Thr Asp Ser Arg Leu
1 5 10 15
Ala Ala Ala Leu Asp Arg Thr His Val Val Phe Val Glu Ser Leu Lys
20 25 30
Gln Met Ile Asn Thr Tyr Leu Arg Met Gln Asn Gly Lys Phe Gly Pro
35 40 45
Asp His Lys Lys Leu Ala Gln Ile Met Leu Ser Arg Ser Asn Thr Phe
50 55 60
Ala His Gly Val Met Asp Gln Ile Thr Arg Asp Gln Pro Thr Ser Thr
65 70 75 80
Leu Asp Glu Glu Trp Thr Asp Leu Ala Arg Arg Ile His Lys Thr Thr
85 90 95
Gly Pro Leu Phe Leu Gln Ala Glu Arg Phe Ala Thr Val Lys Asn Arg
100 105 110
Ala Ile His Thr Lys Ser Arg Gly Lys Val Ile Pro Ser Pro Glu Thr
115 120 125

[0185]

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

Gln Gly Thr Arg Asn Asp Ser Ala Trp Gln Thr Glu Ser Leu Lys Lys
485 490 495
Gly Ile Ile Asn Ser Pro Ser Ile Pro Pro Leu Arg Gln Val Arg Arg
500 505 510
His Asp Tyr Asp Leu Lys Arg Lys Arg Arg Arg His Gly Lys Pro Val
515 520 525
Lys Gly Gln Arg Ser Asn Ala Asn Leu Gln Ala His Arg Thr Asn Met
530 535 540
Ala Gln Asp Arg Phe Lys Lys Gly Ala Ser Ala Ile Val Ser Leu Ala
545 550 555 560
Arg Glu His Ser Ala Asp Leu Ile Leu Phe Glu Asn Leu His Ser Leu
565 570 575
Lys Phe Ser Ala Phe Asp Glu Arg Trp Met Asn Arg Gln Leu Arg Asp
580 585 590
Met Asn Arg Arg His Ile Val Glu Leu Val Ser Glu Gln Ala Pro Glu
595 600 605
Phe Gly Ile Thr Val Lys Asp Asp Ile Asn Pro Trp Met Thr Ser Arg
610 615 620
Ile Cys Ser Asn Cys Asn Leu Pro Gly Phe Arg Phe Ser Met Lys Lys
625 630 635 640
Lys Asn Pro Tyr Arg Glu Lys Leu Pro Arg Glu Lys Cys Thr Asp Phe
645 650 655
[0186]
Gly Tyr Pro Val Trp Glu Pro Gly Gly His Leu Phe Arg Cys Pro His
660 665 670
Cys Asp His Arg Val Asn Ala Asp Ile Asn Ala Ala Ala Asn Leu Ala
675 680 685
Asn Lys Phe Phe Gly Leu Gly Tyr Trp Asn Asn Gly Leu Lys Tyr Asp
690 695 700
Ala Glu Thr Lys Thr Phe Thr Val His Thr Asp Lys Lys Thr Pro Pro
705 710 715 720
Leu Ile Phe Lys Pro Arg Pro Gln Phe Asp Leu Trp Ala Asp Ser Val
725 730 735
Lys Thr Arg Lys Gln Leu Gly Pro Asp Pro Phe
740 745

<210> 349
<211> 743
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
浮霉菌门（Planctomycetes）细菌 RBG_13_46_10 序列
<400> 349
Met Ser Val Arg Ser Phe Gln Ala Arg Val Glu Cys Asp Lys Gln Thr
1 5 10 15

[0187]

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

Gly Lys Glu Gln Gly Gln Thr Lys Glu Thr Asp Ser Tyr Glu Phe Phe
370 375 380
Asp Lys His Leu Lys Lys Trp Arg Pro Ala Lys Leu Ser Gly Val Lys
385 390 395 400
Leu Ile Phe Pro Asp Lys Thr Pro Lys Ala Ala Tyr Leu Tyr Phe Thr
405 410 415
Cys Asp Ile Pro Asp Glu Pro Leu Thr Glu Thr Ala Lys Lys Ile Gln
420 425 430
Trp Leu Glu Thr Gly Asp Val Thr Lys Lys Gly Lys Lys Arg Lys Lys
435 440 445
Lys Val Leu Pro His Gly Leu Val Ser Cys Ala Val Asp Leu Ser Met
450 455 460
Arg Arg Gly Thr Thr Gly Phe Ala Thr Leu Cys Arg Tyr Glu Asn Gly
465 470 475 480
Lys Ile His Ile Leu Arg Ser Arg Asn Leu Trp Val Gly Tyr Lys Glu
485 490 495
Gly Lys Gly Cys His Pro Tyr Arg Trp Thr Glu Gly Pro Asp Leu Gly
500 505 510
His Ile Ala Lys His Lys Arg Glu Ile Arg Ile Leu Arg Ser Lys Arg
515 520 525
Gly Lys Pro Val Lys Gly Glu Glu Ser His Ile Asp Leu Gln Lys His
530 535 540

[0188]

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

Thr Ile Glu Ser Lys Leu Met Asp Lys Leu Cys Ala Met His Lys Ile
725 730 735
Ser Arg Gly Ser Ile Ser Lys
740

<210> 350
<211> 792
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
沉积物宏基因组序列
<400> 350

[0189]

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

[0190]

```
Gly Lys Ala Lys Glu Thr Ile Pro Leu Thr Lys Glu Glu Arg Ala Gly
            260                 265                 270
Phe Arg Lys Pro Gly Gln His Phe Ala Ala Phe Phe Asn Lys Asn Pro
        275                 280                 285
Glu Leu Gln Glu Leu Asp Arg Leu His Lys Glu Tyr Gln Glu Lys Phe
    290                 295                 300
Ala Arg Thr Gln Ser Lys Arg Thr Pro His Pro Asp Gly Phe Lys His
305                 310                 315                 320
Arg Pro Thr Phe Thr Leu Pro Asp Ala Met Arg His Pro Val Trp Tyr
                325                 330                 335
Ser Phe Lys Gly Ala Thr Asp Pro Thr Lys Gly Ser Thr Tyr Arg Asn
            340                 345                 350
Leu Asp Leu Glu Asn Cys Thr Leu Asp Leu Lys Val Leu Thr Ala Met
        355                 360                 365
Glu Gly Glu Gly Arg Asn Pro Gly Gly Met Ile Gln Tyr Ala Phe Glu
    370                 375                 380
Pro Asp Glu Arg Ile Lys Gly Phe Arg Tyr Val Gly Thr Thr Glu Lys
385                 390                 395                 400
Gly Lys Arg Ala Lys Gly Tyr Ile Tyr Tyr Asp Pro Ile Leu Glu Lys
                405                 410                 415
Glu Arg Pro Ala Lys Ile Gln Gly Ile Lys Leu Val Phe Arg Pro Pro
            420                 425                 430
Arg Pro Asp Gly Thr Ala Tyr Leu Ile Phe Ser Cys Gln Ile Glu Asp
        435                 440                 445
Glu Lys Pro Lys Ile Lys Ile Trp Lys Asp Lys Glu Glu Glu Ser Pro
    450                 455                 460
Gly Glu Ile Thr Lys Arg Lys Lys Thr Glu Val Tyr Pro Pro Glu Leu
465                 470                 475                 480
Ile Thr Leu Ala Ile Asp Phe Gly Gln Arg His Leu Gly Ala Ile Thr
                485                 490                 495
Ile Cys Lys Asn Asn Asn Gly Arg Pro Glu Pro Ile Arg Phe Ile Pro
            500                 505                 510
Ala Tyr Pro Lys Arg Arg Lys Asp Arg Glu Ser Lys Pro Val Ser Ala
        515                 520                 525
Trp Leu Ala Lys Ile Pro Gly Leu Thr Phe Asn Ala Val Gly Met His
    530                 535                 540
Glu Lys Glu Ile Ser Ala Gly Met Ser Arg Arg Phe Gln Asp Pro Lys
545                 550                 555                 560
Ser Ile Arg Gln Ala Gly Glu Lys Glu Gly Arg Lys Ser Lys Gly Gln
                565                 570                 575
His Ile Pro Glu Thr Glu Thr Pro Trp Ala His Leu Arg Glu His Ile
            580                 585                 590
Ala Asn Met Lys Glu Asp His Tyr Lys Lys Ala Ala Asn Leu Ile Ile
        595                 600                 605
```

[0191]

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

<210> 351
<211> 754
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
土壤宏基因组序列
<400> 351
Met Gly Val Arg Ala Ile Lys Ser Lys Ile Ile Cys Asp Lys Val Val
1 5 10 15
Leu Thr His Leu Trp Arg Thr His Lys Val Phe Asn Glu Arg Leu Pro
20 25 30
Leu Ile Ile Ser Leu Leu Phe Lys Met Arg Arg Gly Glu Ala Gly Glu
35 40 45
Thr Pro Glu Arg Gln Asp Leu Tyr Arg Arg Ile Ala Glu Phe Ile Leu
50 55 60
Met Arg Ser Ser Lys Asp Ala Pro Tyr Leu Phe Asn Ser Ile Cys Ile
65 70 75 80
Arg Asn Trp Lys Pro Asn Ser Ala Lys Gly Tyr Lys Val Arg Leu Ile
85 90 95

[0192]

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

[0193]

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

<210> 352
<211> 767
<212> PRT
<213> 未知

<220>

<223> 未知的描述:
土壤宏基因组序列

<400> 352

[0194]

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

[0195]

Gly Tyr Arg Asn Leu Lys Leu Pro Gly Lys Val Gly Gln Tyr Gly Glu
325 330 335
Val Asp Leu Arg Leu Leu Thr Asp Glu Ile Arg Gly Ala Glu Phe Pro
340 345 350
Val Asp Trp Thr Arg Leu Arg Phe Leu Gly Asp Pro Arg Leu Ala Asp
355 360 365
Phe Lys Pro Val Lys Ile Gln Lys Ala Ala Thr Lys Gly Lys Val Lys
370 375 380
Gly Glu Leu Arg Glu Arg Pro Ala Tyr Arg Phe Phe Asp Gly Glu Leu
385 390 395 400
Gln Met Glu Arg Pro Ala Gln Ile Ser Gly Ala Lys Leu Ile Phe Lys
405 410 415
Asn Val Arg Leu Asp Asp Gly Ser Leu Gln Ser Ala Thr Pro Tyr Leu
420 425 430
Tyr Phe Thr Cys Glu Ala Glu Ser Leu Gln Trp Thr Glu Asn Ala Arg
435 440 445
Lys Val Lys Trp Asn Ala Thr Gly Glu Val Met Lys Lys Gly Lys Lys
450 455 460
Arg Gly Tyr Arg Thr Leu Pro Asp Gly Leu Ile Ala Cys Ala Val Asp
465 470 475 480
Leu Gly Ile Arg Asn Leu Gly Phe Ala Thr Leu Ala Arg Tyr Glu Asn
485 490 495
Gly Val Pro Arg Leu Leu Arg Ser Arg Asn Leu Trp Ile Gly His Ala
500 505 510
Glu Ala Arg Gly Ile His Pro Gly Arg Trp Ser Glu Gly Pro Glu Leu
515 520 525
Ala His Leu Ala Leu His Lys Arg Glu Leu Arg Arg Leu Arg Ser Leu
530 535 540
Arg Gly Asp Pro Ile Ala Gly Glu Glu Ser His Ile Lys Leu Gln Gln
545 550 555 560
His Ile Thr Tyr Phe Ala Glu Asp Arg Phe Lys Lys Ala Ala Arg Ala
565 570 575
Ile Val Asn Phe Ala Leu Asn Val Lys Glu Phe Ala Asp Lys Asn Thr
580 585 590
Gly Glu Ile Tyr Pro Arg Ala Asp Val Leu Val Leu Glu Asn Leu Gly
595 600 605
Thr Leu Leu Pro Asn Ala Glu Arg Glu Lys Gly Ile Asn Arg Ser Leu
610 615 620
Ile Glu Phe Asn Arg Gly His Leu Val Asp Arg Val Lys Glu Leu Ala
625 630 635 640
Lys Asp Val Gly Leu Arg Val Phe Glu Val Ser Pro Val Gly Thr Ser
645 650 655
Gln Val Cys Ser Arg Cys Gly Thr Leu Gly Arg Arg Tyr Ser Ile Arg
660 665 670

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

<210> 353
<211> 767
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
土壤宏基因组序列

<400> 353

[0196]

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

[0197]

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

Arg Gly Asp Pro Ile Ala Gly Glu Glu Ser His Ile Lys Leu Gln Gln
545 550 555 560
His Ile Thr Tyr Phe Ala Glu Asp Arg Phe Lys Lys Ala Ala Arg Ala
565 570 575
Ile Val Asn Phe Ala Leu Asn Val Lys Glu Phe Ala Asp Lys Asn Thr
580 585 590
Gly Glu Ile Tyr Pro Arg Ala Asp Val Leu Val Leu Glu Asn Leu Gly
595 600 605
Thr Leu Leu Pro Asn Ala Glu Arg Glu Lys Gly Ile Asn Arg Ser Leu
610 615 620
Ile Glu Phe Asn Arg Gly His Leu Val Asp Arg Val Lys Glu Leu Ala
625 630 635 640
Lys Asp Val Gly Leu Arg Val Phe Glu Val Ser Pro Val Gly Thr Ser
645 650 655
Gln Val Cys Ser Arg Cys Gly Thr Leu Gly Arg Arg Tyr Ser Ile Arg
660 665 670
Arg Ser Ala Glu Thr His Arg Pro Glu Ile Gln Phe Gly Phe Val Glu
675 680 685
Lys Leu Phe Ala Cys Pro Glu Cys Gly Tyr Arg Ala Asn Ala Asp His
690 695 700
Asn Ala Ser Val Asn Leu His Arg Arg Leu Ala Gln Val Pro Gly Ala
705 710 715 720
[0198] Phe Ala Gln Trp Glu Glu Leu Gly Lys Asp Gln Gln Ser Lys Arg Lys
725 730 735
Arg Trp Glu Glu Ile Glu Ala Lys Leu Leu Pro Ser Leu Gln Lys Met
740 745 750
His Gly Phe Asp Gly Ala Gly Lys Pro His Pro Phe Leu Ser Ala
755 760 765

<210> 354
<211> 767
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
土壤宏基因组或陆生-土壤序列
<400> 354
Met Ala Val Arg Ala Ile Gln Val Arg Leu Glu Ser Asp Lys Thr Leu
1 5 10 15
Leu Glu His Leu Trp Arg Thr His Lys Val Phe Asn Glu Arg Leu Pro
20 25 30
Met Ile Ile Ser Ile Leu Phe Lys Met Arg Arg Gly Glu Cys Gly Thr
35 40 45
Thr Asp Glu Glu Arg Thr Leu Tyr Lys Asn Ile Ala Arg Phe Val Leu
50 55 60

[0199]

```
Ala Arg Ser Ser Lys Asp Ala Pro Tyr Leu Leu Asn Ser Ile Ser Ile
65                  70                  75                  80
Lys Asp Trp Thr Pro Ser Thr Ala Arg Lys Met Lys Ala Thr Ile Thr
                85                  90                  95
Ile Pro Asp Gly Ser Glu Gln Glu Val Cys Gly Asp Thr Trp Ala Asp
            100                 105                 110
Lys Val Ala Glu Leu Ser Ala Ala Gly Arg Leu Leu Tyr Ser Lys Gln
        115                 120                 125
Glu Val Leu Gly Asp Leu Pro Asp Ser Leu Arg Gln Met Ile Val Arg
    130                 135                 140
Glu Cys Val Ala Ile Ile Ser Gly His Asp Glu Leu Val Arg Asn Trp
145                 150                 155                 160
Asp Lys Ser His Thr Glu Trp Leu Lys Lys Arg Glu Glu Trp Glu Ala
                165                 170                 175
Lys Glu Glu His Lys Lys Tyr Leu Ala Val Arg Ser Arg Phe Asp Glu
            180                 185                 190
Phe Glu Lys Ser Val Gly Gly Lys Ala Gly Lys Arg Arg Ala Arg Trp
        195                 200                 205
His Lys Tyr Leu Gln Trp Leu Arg Glu Ser Pro Asp Leu Ala Gly Trp
    210                 215                 220
Arg Gly Ser Leu Ala Val Val Asn Glu Leu Ser Arg Glu Ala Gln Glu
225                 230                 235                 240
Arg Val Arg Lys Ala Lys Pro Trp Arg Gln Arg Gly Val Glu Ala Glu
                245                 250                 255
Glu Phe Trp Lys Ala Asn Pro Glu Leu Gln Ala Leu Asp Arg Leu His
            260                 265                 270
Gly Gln Tyr Glu Arg Asp Phe Val Arg Arg Gly Lys Thr Lys Arg Asn
        275                 280                 285
Pro Asp Gly Phe Glu His Arg Pro Thr Phe Thr Leu Pro His Phe Leu
    290                 295                 300
Arg His Pro Arg Trp Leu Val Phe Asn Ala Pro Gln Thr Ser Pro Pro
305                 310                 315                 320
Gly Tyr Arg Asn Leu Lys Leu Pro Gly Lys Ala Gly Gln Tyr Gly Glu
                325                 330                 335
Val Glu Leu Arg Leu Leu Thr Asn Glu Ile Arg Gly Ala Glu Phe Pro
            340                 345                 350
Ile Asp Trp Thr Thr Leu Arg Phe Leu Gly Asp Pro Arg Leu Ala Asp
        355                 360                 365
Phe Lys Pro Val Lys Ile Gln Lys Ala Ala Thr Lys Gly Lys Val Lys
    370                 375                 380
Gly Glu Leu Arg Gly Arg Pro Ala Tyr Arg Phe Phe Asp Arg Glu Leu
385                 390                 395                 400
Gln Met Glu Arg Pro Ala Gln Ile Ser Gly Ala Lys Leu Ile Phe Lys
                405                 410                 415
```

[0200]

Asn Val Arg Leu Asp Asp Gly Pro Leu Gln Ser Ala Thr Pro Tyr Leu
420 425 430
Tyr Phe Ala Cys Glu Ala Lys Ser Leu Arg Trp Thr Glu Asn Ala Arg
435 440 445
Lys Val Lys Trp Asn Ala Thr Gly Glu Val Thr Lys Lys Gly Lys Lys
450 455 460
Arg Gly Tyr Arg Thr Leu Pro Asp Gly Leu Ile Ala Cys Ala Val Asp
465 470 475 480
Leu Gly Ile Arg Asn Leu Gly Phe Ala Thr Leu Ala Arg Tyr Glu Asn
485 490 495
Gly Val Pro Arg Leu Leu Arg Ser Arg Asn Leu Trp Ile Gly His Ala
500 505 510
Glu Thr Arg Gly Arg His Pro Gly Arg Trp Cys Glu Gly Pro Glu Leu
515 520 525
Ala His Leu Ala Leu His Lys Arg Glu Leu Arg Arg Leu Arg Ser Leu
530 535 540
Arg Gly Asp Pro Ile Ala Gly Glu Glu Ser His Ile Glu Leu Gln Gln
545 550 555 560
His Ile Thr His Phe Ala Glu Asp Arg Phe Lys Lys Ala Ala Arg Ala
565 570 575
Ile Val Asn Phe Ala Leu Asn Val Lys Glu Phe Ala Asp Lys Asn Thr
580 585 590
Gly Glu Ile Tyr Pro Arg Ala Asp Val Leu Val Leu Glu Asn Leu Gly
595 600 605
Ser Leu Leu Pro Asp Ala Glu Arg Glu Lys Gly Ile Asn Arg Ser Leu
610 615 620
Ile Glu Phe Asn Arg Gly His Leu Val Asp Arg Val Lys Glu Val Ala
625 630 635 640
Lys Asp Val Gly Leu Arg Val Phe Glu Val Ser Pro Val Gly Thr Ser
645 650 655
Gln Val Cys Ser Arg Cys Gly Thr Leu Gly Arg Arg Tyr Ser Ile Arg
660 665 670
Arg Ser Ala Glu Thr His Arg Pro Glu Ile Gln Phe Gly Phe Val Glu
675 680 685
Lys Leu Phe Ala Cys Pro Glu Cys Gly Tyr Arg Ala Asn Ala Asp His
690 695 700
Asn Ala Ser Val Asn Leu His Arg Arg Leu Ala Gln Val Pro Gly Ala
705 710 715 720
Phe Ala Gln Trp Glu Glu Leu Gly Lys Asp Gln Gln Arg Lys Arg Lys
725 730 735
Arg Trp Glu Glu Ile Glu Ala Lys Leu Leu Pro Ser Leu Gln Lys Met
740 745 750
His Gly Phe Asp Gly Ala Gly Lys Pro His Pro Phe Leu Ser Ala
755 760 765

<210> 355
<211> 669
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
土壤宏基因组或陆生-土壤序列
<400> 355

Met Cys Glu Ser Pro Gln Gln Arg Ala Phe Leu Glu Leu Thr His Arg
1 5 10 15
Leu Phe Asn Glu Ser Leu Arg Pro Leu Val Pro Leu Leu Tyr Ala Ala
20 25 30
Arg Lys Gly Arg His Gly Pro Asp Tyr Gln Gln Ile Leu Ser Thr Ile
35 40 45
Lys Ser Gly Gln Gly Ala Ser Ala Gln Val Glu Ala Val Ser Ser Leu
50 55 60
Ala Ala Arg Pro Gly Ser Arg Gly Gly Lys Glu Trp Thr Asp Leu Ala
65 70 75 80
Lys Glu Leu Val Ala Arg Gly Gln Ile Leu Phe Asp Arg His Ala Leu
85 90 95
Met Pro Asn Val His Lys Glu Phe Arg Arg Lys Val Phe Glu Met Ala
100 105 110

[0201]

Phe Gln Ile Ile Leu Ser His Asp Ala Lys Met Lys Gln Trp Arg Thr
115 120 125
Asp His Ala Lys Trp Leu Arg Glu Lys Ala Glu Trp Glu Arg Ser Tyr
130 135 140
Pro Glu Tyr Leu Ala Val Arg Pro Val Ile Gln Glu Phe Glu Ala Arg
145 150 155 160
Glu Gly Gln Ala Gly Lys Arg Arg Gly Arg Trp His Arg Trp Leu Asp
165 170 175
Phe Leu Val Leu Tyr Pro Gln Leu Ala Ala Trp Arg Gly Gly Glu Pro
180 185 190
Thr Val Arg Gln Leu Thr Gln Gln Glu Leu Gln Glu Ala Arg Arg Ser
195 200 205
Arg Arg Arg Ala Val Pro Asn Thr Ile Asp Ala Phe Phe Ala Lys Asn
210 215 220
Pro Glu Leu Lys Glu Leu His Gly Leu His Gly Glu Tyr Gln Ala Arg
225 230 235 240
Phe Val Arg Pro Trp Ala Lys Arg Arg His Gln Asp Gly Phe Phe Leu
245 250 255
Pro Pro Thr Phe Thr Leu Pro Ser Ala Asp Arg His Pro Ala Trp Tyr
260 265 270
Ser Phe Lys Arg Gly Asp Thr Tyr Arg Asn Leu Ser Leu Glu Gln Arg
275 280 285

[0202]

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

Asn Asp Arg Val Ile Glu Leu Ala Ser Val Asn Gln Ala Trp Glu Asn
645 650 655
His Trp Leu Arg Lys His Asn Ser Ala Asn Asp Val Phe
660 665

<210> 356
<211> 744
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
土壤宏基因组序列
<400> 356

[0203]

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

[0204]

```
Asp Phe Val Leu Leu Thr Pro Asp Glu Val His Glu Val Glu Lys Gln
            260                 265                 270
Phe Pro Lys Arg Gln Asp Arg Arg Thr Gly Lys Met Leu Asp Met Leu
        275                 280                 285
Arg Arg Lys Asn Pro Glu Leu Asp Glu Ile Glu Lys Leu Arg Lys Thr
    290                 295                 300
Tyr Thr Asn Asp Phe Ala Arg Phe Arg Arg Pro Pro Thr Leu Thr Leu
305                 310                 315                 320
Pro Ser Ala Asp Lys His Pro Cys Trp Cys Arg Leu Glu Lys Asp Lys
                325                 330                 335
Leu Tyr Arg Lys Ala Asp Phe Glu Lys Gly Thr Ile Glu Leu Cys Val
            340                 345                 350
Val Cys Lys Asp Glu Thr Asp Gly Glu Tyr Asp Met Lys Trp Arg Thr
        355                 360                 365
Phe Arg Leu Lys Pro Asp Pro Arg Leu Gln Pro Ser His Arg Ala Lys
    370                 375                 380
Ala Phe Ala Ala Glu Gly Arg Tyr Pro Pro Tyr Lys Glu Gly Lys Val
385                 390                 395                 400
Gly Asn Thr Leu Asn Arg Pro Ala Glu Ser Pro Asp Lys Arg Leu Ala
                405                 410                 415
Gly Leu Lys Gly Ala Lys Leu Ile Leu Lys Arg Asn Asp Asn Ala Gln
            420                 425                 430
Ile Val Phe Thr Val Val Glu Gln Asn Glu Pro Leu Lys Ala Lys Phe
        435                 440                 445
Ser Lys Lys Glu Gly Ala Lys Ser Ser Ala Asp Arg Leu Glu Ile Leu
    450                 455                 460
Asp Gly Lys Glu Ala Thr Arg Val Met Ala Val Asp Leu Gly Ala Arg
465                 470                 475                 480
His Val Gly Ala Tyr Val Ile Thr Glu Gly Lys Arg Ser Ser Asp Val
                485                 490                 495
Trp Arg Leu Glu His Leu Lys Lys Ala Phe Val Asp Ser Pro Asp Leu
            500                 505                 510
Pro Asp Leu Arg Ser Ile Thr Gln His Asp Trp Gln Leu Arg Lys Ala
        515                 520                 525
Arg Arg Lys Arg Gly Lys Pro Val Lys Asn Gln Arg Ser Phe Ile Ser
    530                 535                 540
Leu Gln Asp His Arg Thr Lys Met Ser Asp Asp Arg Phe Lys Lys Ala
545                 550                 555                 560
Ala Asn Val Ile Val Glu Asn Ala Arg Arg His Ser Val His Ile Ile
                565                 570                 575
Val Phe Glu Asp Leu Arg Asp Leu Lys Pro Thr Ala Tyr Asn Glu Arg
            580                 585                 590
Trp Leu Asn Arg Gln Val Arg Gly Met Asn His Arg His Ile Phe Asp
        595                 600                 605
```

Phe Thr Lys Lys Met Ala Ala Glu Phe Gly Ile Asn Val Thr Asp Val
610 615 620
Arg Ala Tyr Asn Thr Ser Arg Leu Cys Ser Lys Cys Gly Phe Pro Gly
625 630 635 640
Tyr Arg Phe Ser Met Lys Gln Lys Ser Pro Tyr Arg Glu Lys Ala Pro
645 650 655
Arg Lys Gly Cys Ser Asp Phe Gly Tyr Pro Val Trp Asp Ser Gly Gly
660 665 670
His Leu Phe Arg Cys Pro His Cys Gly Tyr Arg Val Gln Ala Asp Leu
675 680 685
Asn Ala Ala Ala Asn Ile Ala Gln Arg Phe Phe Gly Ser Leu Met Glu
690 695 700
Trp Lys Cys Lys Asp Tyr Arg Tyr Ser Asn Ala Met His Pro Ser Phe
705 710 715 720
Asp Ala Arg Lys Asp Phe Asp Lys Trp Ala Ser Asp Val Gln Arg Arg
725 730 735
Lys Ser Met Pro Asp Ala Pro Phe
740

[0205]

<210> 357
<211> 740
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
土壤宏基因组或陆生-土壤序列
<400> 357
Met Ala Thr Lys Thr Tyr Tyr Ala Arg Phe Val Asn Thr Thr Pro Arg
1 5 10 15
Leu Arg Ala Ala Leu Glu Arg Ser His Tyr Val Tyr Val Glu Cys Leu
20 25 30
Arg Gln Met Ile Glu Arg Tyr Ile Ala Met Arg Lys Gly Lys Leu Gly
35 40 45
Glu Glu Cys Arg Gln Leu Ala Lys Ile Ile Leu Thr Thr Gly Asn Thr
50 55 60
Glu Ala His Gly Ile Met Asn Gln Ile Thr Arg Pro Val Val Thr Ser
65 70 75 80
Gly Lys Asp Tyr Glu Trp Val Lys Leu Ala Lys Gln Val His Lys Leu
85 90 95
Arg Gly Pro Leu Leu Leu Gln Asp Glu Gly Phe Ala Arg Val Asp Gly
100 105 110
Val Thr Ile Arg Thr Lys His His Gly Lys Val Glu Pro Thr Arg Gly
115 120 125
Glu Leu Ala Val Thr Ala Lys Phe Trp Met Gln Ile Cys Asp Met Ala
130 135 140

Ser Ser Phe Leu Lys Ser Asn Ala Glu Leu Met Asp Asp Trp Arg Glu
145 150 155 160
Glu Arg Thr Lys Trp Leu Arg Glu Lys Ala Glu Trp Glu Glu Arg Asn
165 170 175
Ser Asp Phe Leu Lys Phe Trp Asn Gly Pro Tyr Thr Glu Phe Glu Asn
180 185 190
Leu Tyr Glu Arg Met Arg Leu Glu Ala Gln Thr Ala Ala Gly Gln Thr
195 200 205
Pro Thr Val Arg Lys Arg Gln Ser Arg Glu Arg Gly Lys Arg Ile Asp
210 215 220

Arg Trp His Leu Trp Tyr Glu Trp Val Ile Ser His Pro Glu Ile Met
225 230 235 240
Ala Trp Arg Asn Gln Ala Arg Ala Ser Asp Phe Lys Val Met Pro Ile
245 250 255
Glu Leu Gln Asn Gln Ile Lys Arg Lys His Pro Arg Gln Asn Lys Tyr
260 265 270
Ile Pro Glu Phe Leu Lys Trp Leu Arg Glu Asn Asn Ala Glu Phe Asp
275 280 285
Arg Leu Asp Ser Leu Arg Arg Val Tyr Val Asn Asn Tyr Cys Arg Phe
290 295 300
[0206] Lys Arg Pro Pro Thr Leu Thr Leu Pro Ser Pro Lys Lys His Pro Tyr
305 310 315 320
Trp Phe Thr Phe Glu Arg Asp Glu Phe Tyr Lys Asp Ala Asp Phe Glu
325 330 335
Thr Gly Ala Ile Arg Leu Leu Leu Ile Asp Gln Lys Asp Asp Gly Ala
340 345 350
Trp Phe Phe Glu Trp Phe Asp Ala Gln Met Lys Cys Asp Pro Arg Leu
355 360 365
Lys Pro Ser Val Arg Ala Glu His Phe Asn Lys Glu Gly Arg Tyr Pro
370 375 380
Pro Tyr Ile Asp Gly Lys Pro Ala Arg Lys Leu Asn Arg Ser Ala Ala
385 390 395 400
Ser Pro Glu Gln Arg Arg Ala Gly Tyr Thr Gly Ala Lys Leu Val Leu
405 410 415
Lys Gly Ala Arg Gln Glu Leu Leu Phe Thr Val Ile Glu Gln Asp Cys
420 425 430
Pro Pro Lys Val Lys Trp Gln Lys Val Lys Met Arg Ala Cys Arg Ala
435 440 445
Asp Asn Ala Phe Ser Ala Glu Gly Glu Arg Ile Pro Leu Lys Ile Met
450 455 460
Ala Ile Asp Leu Gly Ile Arg His Thr Ala Ala Tyr Val Ile Ala Glu
465 470 475 480
Gly Thr Leu Gln Asp Gly Gln Trp His Leu Ser Trp Gln Lys Lys Gly

[0207]

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

Glu Ser Pro Phe
            740
<210> 358
<211> 706
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
      土壤宏基因组或陆生-土壤序列
<400> 358
Met Ile Glu Arg Tyr Ile Ala Met Arg Lys Gly Lys Leu Gly Glu Glu
1               5                   10                  15
```

[0208]

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

[0209]

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

<210> 359
<211> 758
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
陆生宏基因组序列
<400> 359

[0210]

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

[0211]

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

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

<210> 360
<211> 691
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
陆生-土壤序列

[0212]

<400> 360
Met Ile Gln Gly Thr Thr Ser Cys Ala Ser Pro Thr Ile Ala Val Lys
1 5 10 15
Cys Phe Gln Ala Lys Leu Val Cys Glu Ser Pro Gln Gln Arg Ala Phe
20 25 30
Leu Glu Leu Thr His Arg Leu Phe Asn Glu Ser Leu Arg Pro Leu Val
35 40 45
Pro Leu Leu Tyr Ala Ala Arg Lys Gly Arg His Gly Pro Asp Tyr Gln
50 55 60
Gln Ile Leu Ser Thr Ile Lys Ser Gly Gln Gly Ala Ser Ala Gln Val
65 70 75 80
Glu Ala Val Ser Ser Leu Ala Ala Arg Pro Gly Ser Arg Gly Gly Lys
85 90 95
Glu Trp Thr Asp Leu Ala Lys Glu Leu Val Ala Arg Gly Gln Ile Leu
100 105 110
Phe Asp Arg His Ala Leu Met Pro Asn Val His Lys Glu Phe Arg Arg
115 120 125
Lys Val Phe Glu Met Ala Phe Gln Ile Ile Leu Ser His Asp Ala Lys
130 135 140
Met Lys Gln Trp Arg Thr Asp His Ala Lys Trp Leu Arg Glu Lys Ala
145 150 155 160

[0213]

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

Thr Ser Met Ser Glu Asp Arg Cys Lys Lys Ser Ala Ser Ala Ile Val
515 520 525
Asp Phe Gly Arg Ala His Arg Val Asp Ile Ile Leu Met Glu Arg Leu
530 535 540
Glu Gly Leu Val Pro Asp Ala Ala Arg Glu Arg Arg Ile Asn Cys Gly
545 550 555 560
Leu Ile Asn Trp Asn Arg Gly His Leu Ala Ser Trp Ile Arg Ile Ile
565 570 575
Ala Glu Pro Tyr Gly Ile Arg Val Leu Glu Leu Pro Pro Arg Trp Thr
580 585 590
Ser Gln Val Cys Ser His Cys Gly Ser Ile Gly Ala Arg Phe Thr Cys
595 600 605
Thr Arg Gly Glu Met Lys Ile Ser Ala Gly Gly Lys Leu Phe Gly Cys
610 615 620
Pro Thr Cys Ala Phe Glu Ala Asn Ala Asp Phe Asn Ala Ser Leu Asn
625 630 635 640
Leu Gln Arg Val Phe Trp Gly Leu Phe Pro Gln Val Ser Lys Ala Gly
645 650 655
Arg Gly Arg Val Gln Cys Asn Asp Arg Val Ile Glu Leu Ala Ser Val
660 665 670
Asn Gln Ala Trp Glu Asn His Trp Leu Arg Lys His Asn Ser Ala Asn
675 680 685

[0214]

Asp Val Phe
690

<210> 361
<211> 698
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
陆生-土壤序列
<400> 361
Met Gly Val Lys Ala Trp Lys Val Lys Leu Phe Val Asp Ala Pro Glu
1 5 10 15
Lys Gln Ala Val Leu Asp Asp Thr Leu Gln Gln Trp Asn Leu Tyr Ala
20 25 30
Ala His Val Cys Lys Val Leu Leu Ala Phe Arg Arg Asn Asp Tyr Gly
35 40 45
Asp Val Gly Leu Glu Val Trp Arg Gln Leu Ala Lys Thr Pro Gln Asn
50 55 60
Tyr His Val Leu Tyr Cys Ile Thr Pro Arg Ser Asp Ala Ala Pro Asp
65 70 75 80
Gln Gly Pro Arg Val Thr Arg Glu Lys Asp Tyr His Arg Ala Ala Arg
85 90 95

[0215]

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

[0216]

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

<210> 362
<211> 687
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
陆生-土壤-堆肥序列
<400> 362
Met Gly Thr Lys Ala Trp Lys Ala Lys Leu Ile Leu Asp Ser Glu Asp
1 5 10 15
Ala Arg Ala Ala Leu Leu His Thr Phe Ser Val Trp Asn Ala Ala Cys
20 25 30

[0217]

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

[0218]

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

<210> 363
<211> 768
<212> PRT
<213> 未知
<220>
<223> 未知的描述:

陆生-土壤-硬木森林土壤序列

<400> 363

[0219]

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

[0220]

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

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

<210> 364
<211> 699
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
陆生-土壤-硬木森林土壤序列
<400> 364

[0221]

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

[0222]

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

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

<210> 365
<400> 365
000

[0223]

<210> 366
<400> 366
000
<210> 367
<400> 367
000
<210> 368
<400> 368
000
<210> 369
<400> 369
000
<210> 370
<400> 370
000
<210> 371
<400> 371
000
<210> 372
<400> 372
000
<210> 373
<400> 373

000

<210> 374

<400> 374

000

<210> 375

<400> 375

000

<210> 376

<400> 376

000

<210> 377

<400> 377

000

<210> 378

<400> 378

000

[0224]

<210> 379

<400> 379

000

<210> 380

<400> 380

000

<210> 381

<400> 381

000

<210> 382

<400> 382

000

<210> 383

<400> 383

000

<210> 384

<400> 384

000

<210> 385

<400> 385

000

<210> 386

<400> 386

000

<210> 387

<400> 387

000

[0225]

<210> 388
<400> 388
000
<210> 389
<400> 389
000
<210> 390
<400> 390
000
<210> 391
<400> 391
000
<210> 392
<400> 392
000
<210> 393
<400> 393
000
<210> 394
<400> 394
000
<210> 395
<400> 395
000
<210> 396
<400> 396
000
<210> 397
<400> 397
000
<210> 398
<400> 398
000
<210> 399
<400> 399
000
<210> 400
<400> 400
000
<210> 401
<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> 401
cgtgcatcgg ccgacgaaga gcgaaggtca gtgacag 37
<210> 402
<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> 402
cgccgcagcg gccgacgcgg ccctgatcga tggacac 37
<210> 403
<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> 403
[0226] gccgcagcgg ccgacgcggc cctgatcgat ggacac 36
<210> 404
<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> 404
ggagcgcccc ccgacgcggc aatgatcggc tgacac 36
<210> 405
<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> 405
ctctcaatgg ctaaagatct agttctaatt ggacac 36
<210> 406
<211> 30
<212> DNA
<213> 人工序列

<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 406
ccccccccgc cattctcggc tgtgtacaac 30
<210> 407
<211> 38
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 407
cgggtgtacc ccccccgcca ttctcggctg tgtacaac 38
<210> 408
<211> 38
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 408

[0227]

tctctcaatg gctaaagatc tagttctaat tggacacc 38
<210> 409
<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> 409
gcggcgctgg ccagcgccgc gctgaa 26
<210> 410
<211> 32
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 410
gttgcacctc gcgcgcacct gcgggttgac ac 32
<210> 411
<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> 411
ggcgcaccgc gtggcgtgga actgaccggc tgacac 36
<210> 412
<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> 412
cctgtatccg cgcgcgcgtt tgcggctgat gaacac 36
<210> 413
<211> 37
<212> DNA
<213> 人工序列
<220>
[0228] <223>人工序列的描述：合成
寡核苷酸
<400> 413
gtattcatcc ccttcagaaa aggaactctc tggacac 37
<210> 414
<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> 414
ggcgcacccg ctggcgcgga acagaccgac tgacac 36
<210> 415
<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> 415
ctctcaaccg cttgggcaga accgaccgac tgacac 36
<210> 416

<210> 37
<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> 416
ctgtcccccc cggacctgac caataaggga ttgacac 37
<210> 417
<211> 32
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 417
gtgacatctt atgcgtgtgt acaggtggac ac 32
<210> 418
<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> 418
gaagaaaaga cctctcaaaa aaaattctcc tggacac 37
<210> 419
<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> 419
gtcgcggcgg ccgatgcgat cctgatcgat ggacac 36
<210> 420
<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> 420
gtgtgtttcc tcttagaaaa acaaatgctc tggacac 37

[0229]

[0230]

<210> 421
<211> 35
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 421
gggggcaag ctcgcgcgta tgcacggatt ggcag 35
<210> 422
<211> 38
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 422
ggtggacgtc ccccgcgatt cccgggtgtg taccacga 38
<210> 423
<211> 32
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 423
gttgcaatct gtgcgtatgt atgggtggac ac 32
<210> 424
<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> 424
actataggta ccctaaaaat tccgggtaac tacacc 36
<210> 425
<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> 425

gatgtagccg cacgttgatt caccggcgat taccac 36

<210> 426

<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> 426

gaagaaaaaa cctcttaaaa aaattctcct ggacac 36

<210> 427

<211> 32

<212> DNA

<213> 人工序列

<220>

<223>人工序列的描述：合成

寡核苷酸

<400> 427

gttgcagttc gcgcgtaggc gtgagcggac ac 32

<210> 428

<211> 32

<212> DNA

[0231]

<213> 人工序列

<220>

<223>人工序列的描述：合成

寡核苷酸

<400> 428

gtcgcgatcc gcacacccga gcgagtcgac ag 32

<210> 429

<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> 429

gccgccgcag cggccgacgc ggccctgatc gatggacac 39

<210> 430

<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> 430
gccgcagcgg ccgacgcggc cctga 25
<210> 431
<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> 431
gccgcggagg ccgacgcggc cctgatggat ggacac 36
<210> 432
<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> 432
atagcgcccc ccgacgcaaa tctgattaat ggacac 36
<210> 433
<211> 36
[0232]
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 433
gtcgcaatga gataaaacat agatatgtgt tgcaac 36
<210> 434
<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> 434
gtgcatcggc cgacgaagag cgaaggtcag tgacag 36
<210> 435
<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> 435

ggcgtcgcga tccgcacacc cgagcgagtc gacagca 37

<210> 436

<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> 436

gccgcgtccc ccgacgcgac cctgatcgat ggacac 36

<210> 437

<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> 437

ggcgcagcgg ccatcaggat ggtgatcggc tgacac 36

[0233]

<210> 438

<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> 438

ggtgaagcgg ccatcgggat cgtgatcggc tgacac 36

<210> 439

<211> 32

<212> DNA

<213> 人工序列

<220>

<223>人工序列的描述：合成

寡核苷酸

<400> 439

ggcgcaaccc gcacacaacc gcgaatggac ac 32

<210> 440

<211> 30

<212> DNA

<213> 人工序列

<220>

<223>人工序列的描述：合成
寡核苷酸
<400> 440
gccgcgtccc ccgacgcgcc cctgatcgat 30
<210> 441
<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> 441
attgcaccgg ccaacgcaaa tctgattgat ggacac 36
<210> 442
<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> 442
[0234] ggtggacgtc ccccgcgatt cccgggtgtg taccac 36
<210> 443
<211> 34
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 443
tataggtacc ctaaaaattc cgggtaacta cacc 34
<210> 444
<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> 444
gaagaaaaaa cctcttaaaa aaattctcct ggacact 37
<210> 445
<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> 445
ctctcacccc agttcgatcg aggaaggcga ctgacac 37
<210> 446
<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> 446
ggcgtaccgc ctgacgggtt tctcatcgat ggacac 36
<210> 447
<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> 447
[0235]
ggcgcaccgc ctgacgggtt tctcatcgat ggacac 36
<210> 448
<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> 448
accgcggcga taccggcttt cccgggtgac tgacac 36
<210> 449
<211> 32
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 449
gtcgcacctc gcgcgtatgc acggattggc ag 32
<210> 450
<211> 32
<212> DNA

<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 450
gttgcaacct gtgcgtatgt gcggatggac ac 32
<210> 451
<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> 451
ggagagcgaa aaaaagcggg ggagat 26
<210> 452
<211> 39
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
[0236]
<400> 452
ggtaccgcgg cgataccggc tttcccgggt gactgacac 39
<210> 453
<211> 33
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 453
ggcgcaaccc cgcacatgcg ggcgggatga cag 33
<210> 454
<211> 32
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 454
gtcgcacccc gcgcgtgcga gcgtgttgac ag 32
<210> 455
<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> 455
ggagcgcccc cccgacgcgg ccatgatcga ctgacgg 37
<210> 456
<400> 456
000
<210> 457
<400> 457
000
<210> 458
<400> 458
000
<210> 459
<400> 459
000
<210> 460

[0237]

<400> 460
000
<210> 461
<400> 461
000
<210> 462
<400> 462
000
<210> 463
<400> 463
000
<210> 464
<400> 464
000
<210> 465
<400> 465
000
<210> 466
<400> 466
000
<210> 467
<400> 467
000

<210> 468
<400> 468
000
<210> 469
<400> 469
000
<210> 470
<400> 470
000
<210> 471
<400> 471
000
<210> 472
<400> 472
000
<210> 473
<400> 473
000
<210> 474
<400> 474
000
[0238] <210> 475
<400> 475
000
<210> 476
<400> 476
000
<210> 477
<400> 477
000
<210> 478
<400> 478
000
<210> 479
<400> 479
000
<210> 480
<400> 480
000
<210> 481
<400> 481
000
<210> 482
<400> 482

000

<210> 483

<400> 483

000

<210> 484

<400> 484

000

<210> 485

<400> 485

000

<210> 486

<400> 486

000

<210> 487

<400> 487

000

<210> 488

<400> 488

000

<210> 489

<400> 489

000

[0239]

<210> 490

<400> 490

000

<210> 491

<400> 491

000

<210> 492

<400> 492

000

<210> 493

<400> 493

000

<210> 494

<400> 494

000

<210> 495

<400> 495

000

<210> 496

<400> 496

000

<210> 497

<400> 497
000
<210> 498
<400> 498
000
<210> 499
<400> 499
000
<210> 500
<211> 1353
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
水生-淡水-淡水湖序列
<400> 500

[0240]

actttttcta ccatgaaaaa aaacccatct gataaaacaa taatgaacct agaatttatg 60
ttttttata tagaatggat gaaaagatta gaaactagtc ctgattataa taaacctttt 120
attgaagaac tgtttgggat tactatttag attatcgtta gctaaagaca attgaaaaaa 180
tgacttgaag attctgagca aaagattagg attttaaaag gagaaattcc gtgggcacaa 240
gagccataaa agcaaagatt atttgtgaca aaacaacact tggttatttg tggcgatctc 300
actgtatgta caatcagaaa ctctctgagt atatgaaatt acttttaaa cttcgacgtg 360
gtaagattgg taaaaataaa gattatgcga ataaaaatga ggaagaaaag cagaaaataa 420
ttgacaatct tgaagcaata cttagacctc atctgcaaaa aatgcacttt accaccaaaa 480
aagaacgctc acttgacgtt ccagaacacc cacttgacgt tccttttga gcttgttaag 540
atatggtcat tgaggagttt tggtggagtt acgaccatcc cacagacatg gcaagccata 600
gctgaaagcc ttgatagcct tcgttttgtg tccaaagtct gctgatctga ctgctttcgg 660
gcatttagga tattacaaaa aaggaaactt ttttggcggc gggcaaggcg agggcaatta 720
gaagtagatc ttttgctgtt tgacaaattt agtcaatttt cgtcgtattc tattgatgcg 780
agtgacgagg tatgatcgtt tttgccgtag ctgctcgcga gctaggtgta gtaacgactt 840
aagcacagct aagcattgta aaaatttgaa tttaaaccct tatttttcaa ggtgctcgtc 900
aagtgtgttc taagttgttg tagagacttc gttaaggact agggccccgt tggaaaaaga 960
aagagttaga tcatctaatc cataagcatt ggaatatttc caaagatcaa agataggaat 1020
ccaagtggga aaaaagaaag taggggaaa aggaaaaata aaacaaaaat aattttcaaa 1080
agagatgttc tggattcaga aaaaaaccag aagaatttaa aattgtgaag ggtaaaatct 1140
gtttggggag gagggtattt ctccaccgaa ctactcaaga ttctctaaaa ttaggaaaag 1200
tttatgaaga catctaaggc agaatttaaa attattaaat taacatttgc tgataaaaaa 1260
tggggactag tattgatttt tgttggtatt tgtttgggtc ctgtgattgt aggtattcca 1320
ttaattatac ttggtgtggg gatgttatgc ata 1353

<210> 501
<211> 1671
<212> DNA
<213> 未知
<220>

<223> 未知的描述:
淡水沉积物宏基因组序列
<400> 501

[0241]

ctgcaatcgg acctcgaagc tgaaatgaac ctcaaagacg accagacaat gattatcgat 60
atgggtccga acgaaaaaga ggccagagac gcagccgctg taatcggcca gtccctgccg 120
aaattcgaaa ccagaatgat cgtcgtataa aattgacatc ccggttatgt tttatatgat 180
taggcataat agaggattac gaaaggagct aatatgtcgg tgaggtcttt tcaagcaaaa 240
gtggaatgcg acaaacaaat aatggaacat ctttggcgga cgcacacggt tttcaatgaa 300
cggcttccgg taattatcaa aattttgttc cagatgaaac ggggcgaatg cggccaaaat 360
gattcggtca atttgcacag gcgatttctg atagatgatg cgttaaagaa ttattatgga 420
ttcaaaaagc tgcctgagaa aaagcaaaaa gaggcaatag cagctattga gaactcgctg 480
aaagaaaaat tatctacaat tcacaaaata taaaatattt tccttgtgtc ttaacatagt 540
tgaacatttt tacttgtctt tgccgatatt ctgaaaaagg gacaaattgt tgtcccataa 600
aggataacgt tcgggaggca agacaatgca agctctttca aaattctgtt gtcaaaataa 660
aagatgcacg gattacggaa aacgtggtgc aaataattta tcggtgtgcg gcagattcgg 720
caaaaatgac catatccggc tgctatactg ccgaacatgc aagaagcgat tttcctataa 780
tttttgctgg ccggtacgga cacttcgttt acaaggccaa gatgagcggt ggcacgcaag 840
gacacctgcg atggccgcgg atttgacaga tcatgtctgg tctttgcatg aatggttgag 900
ttatccggct tttcaattaa cataagacac aaggaatatt tttcttgcag cataaatatg 960
caaaaagtac aataatctcg tggagtaagg taggagtgac gcccgcccgt ttagcgtgct 1020
ataggccgct gaatgccacg tagcgatatg ttttgagtgt caagagctgc tgacccaaag 1080
gccaaaagct gagtagggct tgatagtcgt ggttattacc atgcaatgcg gcagcgacac 1140
atatcgcgtt aatcaaattt gcgttgtcgt caagatggat tcaggccatt gcgcaagtgt 1200
cctctttgac aagttcagac gcgagtgcga cggtggtgaa ggggggcact gggaatgctc 1260
gcagcgggaa tttgtgcaac tgcttccagg atttgaggaa ctattgacaa tatggtttgg 1320
cgcgaatttt tacagttttt ttaccttgct cgtaaaccaa acgctaaccg cagatgcaaa 1380
aagaacttgc agaagaagcc atcaagccac agagaacaga ggaatttca taattacgcc 1440
actgcatata aatccttatg tgccaaatac tgatgaataa aaaatatcat aactttcaaa 1500
tccatccaag ggcgaaccct tttcagccgc ctaacatgat aaggcagcat actaacacga 1560
ctacgacctt tgaacaagct ataccaacgc tccaccacac tacgtctttc cagcagccgt 1620
tgtccctcaa cactctgcaa catacaatac gaagcattac gaccgggttg t 1671

<210> 502
<211> 1056
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
生物反应器宏基因组序列
<400> 502

ctggagaatg cattgcggga tgtggttaac ctcaagcagg atcaggtgct aattgtcgat 60
ttggggtcga atgaggaggc ggcgcgcgag tcggcgacga cgctcgggcc ggcgttgcct 120
tcccaagacg aggggatcgt ggtaatctaa tctgagcggt tcagcaccgg tcccgcaccc 180
gcggcgatag ctattcttgg aggtgtagac gtgggagtcc gggcacttcg agcgcgaatc 240
gagtgcgatc gagaaacgct tatgcacctt tggcgcacgc actgcgtctt caatcaacgc 300

ctacccgcgc tgctgaaacg cttgttcgca atgcggcgag gggaagtggg ggggaacgag 360
gcgcagcggc aggtttacca acgtgttgcg cagtttgttc ttgcgcgcga cggcgatcaa 420
gccgtcgccg ccttccgtgc tctcccaccg cgggattccc ccgccaggac cttagccgtt 480
aaaagggtcg aagataccct tcgcccacag ttgatgcgcg tgcacaaact tgcggatgca 540
ggggttgaca gccccttctg aggctctgat atactgggtt tgtggggcga gccttcgggc 600
ggtatatccg agccgaagtc cacgtagtcg tcgcccctgt gtgtcaacac gtgctgactc 660
caaggcgaac agccgaagga ggccccgcaa gggtgacatc gtggcgggct cgcacgcccc 720
ggcggcacgg acagcgacgg cattgatcag gtccgcgtcg gcgcgggcgg cgcttcgcgg 780
cggtgttctt tgaaaagtcg gtcgtcgcgg ccttggccga ggataaagtc ggtgcgacgc 840
agccaactgc gaccaacgcc gtgggcgcgt cagggcgccc gcgagtgcga cggtggtgaa 900
ggggggcct gggcacgctc gcgagggcgt aagtctttga gaattgagga cttgcagctt 960
gctggttttc ggatgcgacc gaacagttgc cgtacgtgcg ggtcacgctc gcaaacgcac 1020
gcgcaagttc ggtatttaca gcaacttgcg agtgcg 1056

<210> 503
<211> 1396
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
生物反应器-柳枝稷降解序列
<400> 503

[0242]
caagccggcg acgtcgagcg cacctatgcc gacctgacgc ggtcgaaggc agaactcggc 60
tacgacccgc aggtggagat tacggaaggc ctgcggcggt tcgtggcgtg gtatcgggag 120
aatcgggaga ggtatcggct ggaggtgtga gggggattcg ggagcgacga agcgacggag 180
cgacgaaggg gaaggcgtat gatgcatagg ggcaggcaac tagcccgccc tactgcctga 240
tcttctcgac cggccacata atagtgggcg catctccacg ggcacaacag tgagcgagtc 300
aggtcggcat gacttccaca gatctgattg tctgttcttg catgggaaat tgcacagtca 360
actgaatcca cgttatcctt gatcttgctg ccaaaggagg accaggtggg gatcaaatct 420
tggcgcgtta aattatgggt ggaatcgcct gagatacgcg cgactctgga atattgtctc 480
agcctgtgga acacttcggc tgctttactt tgtcggcata tactagcgat gcgtcgtggc 540
gactttggcc cggaagggca agcgatctgg gagcacctaa agaaaacccc acataatttt 600
catgtccttt atcctatctc caatggcagc ggcatagaaa agactaccgg atattacccc 660
gatgcacgag cattcgttgt cttccttgga acatttccca ctgtgcagac catccagcgc 720
caaccgcgga tctacgagat caatggccag cgcgttgctc tggatgagat acagcagcgc 780
gctgttgaaa tccttgccag tcgccgcgca ggtgatacac ccttttaaga tttgtttgcg 840
gtatcagaca ggaatagtta acatttgcgg tggatcaggt ccgaaagggc cggcccgagg 900
ttcggctata ggcccaacct cgcaccagcg ggcgctgtct agccgaaatt ggcgagcgag 960
gctcttctca gggcccgctc gaaacctgtg gctacgatta cgtggctgcg tgggcttttc 1020
ggcaccccgt tggacacgct tcggtatgcg gaacgcgaac caccagcgcg caccagttgc 1080
atagcggatt cgcgtttagc gcaagttctt tctcaagcga actgtgcaaa cagtgggtca 1140
gcaaatcgct tctgttgcat ttgcctggac gagacgattc gcgaaagtgg actgtagctc 1200
ttggtcctgc ctttaactag agcaaggccg ccgataccac tggcgacggc ttgcacgccg 1260
acagggtcgc gatcggtaac gtcctgaatt gtgcgcacat ttgaaaaggc agctcgtccg 1320
ctggaaggtg ttggtgtgta gacaaccctt accagacagg agagctgcat ggacagcgta 1380

acagaagcgg cgcagg 1396

<210> 504

<211> 1003

<212> DNA

<213> 未知

<220>

<223> 未知的描述:
生物反应器宏基因组序列

<400> 504

[0243]

caacacggcc tgggctacgc cccgttcccc gcctccgtga tttcccccag gaactggacc 60
gaccactgcc acctgaaccc caacggcgag cgcgaaaaag ccacgcacat cgccgaccat 120
gtccgccccc tcatccgcgc tccgcgataa gcattttcgg acggaatatg tctcatcgga 180
tgcacttgac ggtggctgcg ccgccgagcg gagtgcccgt gcttgcccat ccatcgaccc 240
gttggcatcc tgtgaacagg aaccaggcac atgtccacac gcgctattca ggccagagtt 300
gtcgccgaca aggcactgct gcaaaagctc tggctgaccc acagggcgtt caacgaacat 360
ttgccccatt tttcgcaaac cctgttccgc atgatgcggg gcgagttggg cgactccgat 420
cgcttcgccc ttggtgaaac cgccatcgct gccgccatgg acttgtggca tcgccccaag 480
cccgaaaaag aggctatgcg aaaatcaatc gaagcccaac tccgatcaaa actggaaatc 540
gaacacaagc tcgccgaacc gctgccctaa ggctcttcaa ggagcttgcg ctgggcaccc 600
aggttttcat acgttgatcg cgtggagcaa gggaccgggt cgcgtcggtt tccaccattt 660
ggctataggc caggcggttg ccacatggcg ccatgttttt gagtgtcaag atgtgctggt 720
cctaaggcga cacgccgaag ggactttacc ctccgggcaa ctggaagggg cggcacagac 780
ccatggcgcg ttcactccta cgatctttga agcgcgagct gcgacgttcg cgaacctata 840
gcggacaacc atctacagag aggtccgcgg gcggcgcaac caccgccccc acgaggactt 900
gtcacgctcg ttctccgctc gcctcttgcg ggatcgggaa tcgcgtttcg cttcgcgaac 960
tggcgatgtg cgcacttgtg cttcaatcct ttgcactgca agg 1003

<210> 505

<211> 1116

<212> DNA

<213> 未知

<220>

<223> 未知的描述:
水生-淡水-淡水湖序列

<400> 505

gccggctgga cctccgagtt ccccggccag ccgcccccca cccgcatcgc cctggccggg 60
ccggacgccc tcctgttctt ctacctcacg ctcctccagg acatcaacac caagaccctc 120
ctcccgcccg acctcgaatt ttactggtag ccgggctaga tagtctctct tgcggaaagt 180
gcgatcacag gcgggacgcc tgtgccaccg ctgagacaag gggcttggtg gcacaggcct 240
ctggcctgtg attggggcac cgagggcgtg gatacgatgg acatggtgga ccccgaggcc 300
cgcgccgccc tctggcgcac ccatggggtc tacaacgagt gcctccggtt cgccctgcgc 360
cacatgcaca agatgaagcg cggggaggcc gacccgcgct acgcggagat attcggcgcg 420
atccgcgtgg agaggcccaa caaggacaag cgggtcttcc gcctgactcc caggggccag 480
ccaaccgtgg aagtggacat ggccgacgtg gagaagcgca tcctcgcgcg gggaagggcc 540
gaagacctct gccgcggcaa gcccacgcca ttttgagctt gactttcccg gccaaactgt 600

catgaatggt gttggccctg cgtggaatag cgcgggccgc cggcggggct gctataggcc 660
gcgtcggact acgtcgggca tccagttgtt acggccgcgt gggcctacgg gtttcggccc 720
aaagggcaca agaccggatg ggtcactgga cctgtcaggg cggcggccag gcccggcgac 780
acgaccgaga atgttgggga aaaaggaggc ccggaccgcc aaagaatggc ccggatgggg 840
cgctttttc agccgaaatg cgggctaact ccttgcgccc tctcgacttc ggcggtatct 900
ggtgtaggcg cgggtagact tttcggacga tgtcagggca agggcgaagg ggtcagacct 960
caattctcta ctttcctgcc ccgggtggtg tgccgacagc ggcgtcgagc cgccgcagtc 1020
cgacgggctg acgcacgctc gggtgcccgt gacggtgcga gaggcatggg gtacggcggg 1080
tggctgggaa aggacaacag ggaccacagg gacaac 1116

<210> 506
<211> 1059
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
生物反应器宏基因组或生物反应器-厌氧-厌氧沼气
反应器序列

<400> 506

[0244]

ctggagaatg cattgcggga tgtggttaac ctcaagcagg atcaggtgct aattgtcgat 60
ttggggtcga atgaggaggc ggcgcgcgag tcggcgacga cgctcgggcc ggcgttgcct 120
tcccaagacg aggggatcgt ggtaatctaa tctgagcggt tcagcaccgg tcccgcaccc 180
gcggcgatag ctattcttgg aggtgtagac gtgggagtcc gggcacttcg agcgcgaatc 240
gagtgcgatc gagaaacgct tatgcacctt tggcgcacgc actgcgtctt caatcaacgc 300
ctacccgcgc tgctgaaacg cttgttcgca atgcggcgag gggaagtggg ggggaacgag 360
gcgcagcggc aggtttacca acgtgttgcg cagtttgttc ttgcgcgcga cggcgatcaa 420
gccgtcgccg ccttccgtgc tctcgcaccg cgggattcgc ccgccaggac cttagccgtt 480
aaaagggtcg aagataccct tcgcccacag ttgatgcgcg tgcacaaact tgcggatgca 540
ggggttgaca gccccttctg aggctctgat atactgggtt tgtggggcga gccttcgggc 600
ggtatatccg agccgaagtc cacgtagtcg tcgcccctgt gtgtcaacac gtgctgactc 660
caaggcgaac agccgaagga ggccccgcaa gggtgacatc gtggcgggct cgcacgcccc 720
ggcggcacgg acagcgacgg cattgatcag gtccgcgtcg gcgcgggcgg cgcttcgcgg 780
cggtgttctt tgaaaagtcg gtcgtcgcgg ccttggccga ggataaagtc ggtgcgacgc 840
agccaactgc gaccaacgcc gtgggcgcgt cagggcgccc gcgagtgcga cggtggtgaa 900
ggggggggcct gggcacgctc gcgagggcgt aagtctttgg gaattgcgga cttacagctt 960
gctggttttg ggatgcgacc gaacagttgc cgtacgtgcg ggtcacgctc gcaaacgcac 1020
gcgcaagttc ggtatttaca gcaacttgcg agtgcggca 1059

<210> 507
<211> 1702
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
实验室富集限定的培养基-离子液体和高固体
富集的序列

<400> 507

gcgtctacgg caatgccagc aaggtcccct tcagcgaaga agaccccgtc gatcagccca 60
tcagccccta cgcagccacc aaacgcgccg gggagctgat ctgccatacg ttctggcatc 120
tgtacaaaat gccgatcttc tgcctgcggt ttttcaccgt ctacggcccc cgccagcggc 180
cggacctggc cattcacaag ttcacgcggc tgatcgacgc cggcaagccg atccccgtct 240
tcggcgacgg cagcatgtcg cgcgattaca cgtatgtgtc ggacacggtg gctgggatca 300
tgagcgccct ggaccgatgc gaccgctacc gcatctacaa cctcggcggc tcgaatccgg 360
tgacgctcac gcagttgatc caggagatcg ggaccgccct cggcaagcag gccatcatcg 420
accgcaaacc cacccaagcc ggcgacgtcg agcgcaccta tgccgacctg acgcggtcga 480
aggcagaact cggctacgac ccgcaggtgg agattacgga aggcctgcgg cggttcgtgg 540
cgtggtatcg ggagaatcgg gagaggtatc ggctggaggt gtgaggggga ttcgggagcg 600
acgaagcgac ggagcgacga aggggaaggc gtatgatgca taggggcagg caactagccc 660
gccctactgc ctgatcttct cgaccggcca cataatagtg ggcgcatctc cacgggcaca 720
acagtgagcg agtcaggtcg gcatgacttc cacagatctg attgtctgtt cttgcatggg 780
aaattgcaca gtcaactgaa tccacgttat ccttgatctt gctgccaaag gaggaccagg 840
tggggatcaa atcttggcgc gttaaattat gggtggaatc gcctgagata cgcgcgactc 900
tggaatattg tctcagcctg tggaacactt cggctgcttt actttgtcga catgtactgg 960
cgatgcgtcg tggcgacttt ggcccggaag ggcaagcgat ctgggagcac ctaaagaaaa 1020
ccccacataa ttttcatgtc ctttatccta tctccaatgg cagcggcatt gaaaagacta 1080
cccggtatta tcccgatgca cgagcattcg ttgtcttcct tggaacattt cccactgtgc 1140
agaccatcca gcgccaaccg cggatctacg agatcaatgg ccagcgcgtt gctctggatg 1200
agatacagca gcgcgctgtt gaaatccttg ccagtcgccg cgcaggtgat acacccttt 1260
aagatttgtt tgcggtatca gacaggaata gttaacattt gtggtggatc aggtccgaaa 1320
gggccggccc gaggttcggc tataggccca acctcgcacc agcgggcgct gtctagccga 1380
aattggcgag cgaggctctt ctcagggccc gctcgaaacc tgtggctacg attacgtggc 1440
tgcgtgggct tttcggcacc ccgttggaca cgcttcggta tgcggaacgc gaaccaccag 1500
cgcgcaccag ttgcatagcg gattcgcgtt tagcgcaagt tctttctcaa gcgaactgtg 1560
caaacagtgg gtcagcaaat cgcttctgtt gcatttgcct ggacgagacg attcgcgaaa 1620
gtggactgta gctcttggtc ctgcctttaa ctagagcaaa tgtctcggct ggcgatatcg 1680
tcgtcgtcgg cgtcgcgatc cg 1702

[0245]

<210> 508
<211> 1090
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
生物反应器宏基因组序列

<400> 508

gaggatggac cagacggcgc cgccgcgctc cttgagcatc gcggcaagaa cgtgcaggcc 60
ggtgacctcc gggttctgac gcgccgcggc gtcggactgg gcgtcggcaa gggtctgctg 120
ggcgagcgtg gtgaagcggt cggtgcgcat gggtcgtgct cccgagggag tctggctgac 180
tcggtctgcc ggattgttgg gcaaaacggg ggccagtcgt gagaggcttt ggaggggggt 240
tttacggagc gcggactgcc ggaatggcag ggggacgccg gccgtcggcc aaccgccccc 300
cgggcgcctc ccggcaagtt tcgttaggat ggtcgccgtg gaggccgcct cagatgggtg 360

ttaaagcctg gaaggcgaag atcatcgtgg ccgacccgga cctggcggcc gctctggacc 420
acaccgtccg gcaatggaac ctctgggcgg aacaggcctg catcgccatc atgcgcctgc 480
gccgcggcga ctacggggac gaggtcctga tcgggacgtt cccgcaggtc cggcgcatct 540
cgcgcgaccc gccggtgctg tcgatcgccg atgtccgggt ggaccagcgg gaggcggatc 600
gcctggccgc ggcgtggctg agcgaactgt cggtcggccc gacgcccttc tgaggtggcg 660
tcggatgggg gcttccggta tactgccgtc gctaccgcct gcacagggcg gcgcggcgga 720
ccggctatag gcccgtccgt cgtacgtcgg gaactgtcga gccgggagtg acgtggcggg 780
gcctcggccc tgttcaagac cttcacccct ctcgtggggg tggtgggctt cccggcaacc 840
cggcggctga tccatcacgc ggccgccgcg aatatcgagc gcccacgccg acgccgggag 900
gttcgcggaa gtcgcaacgc tctttgacag cccaagttgc gctcaggagg cccgatgacc 960
ggcttggctc tccgttcaaa aacagccgat tcgcggagga acgctccgaa gtgccgaccg 1020
gtcaacctct tacgggcggg agctactacg aggcgatgcg gatcggcggg aacgcccgct 1080
gcatcggccg 1090

<210> 509
<211> 598
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
水生-非海生盐水和碱性-超盐水湖沉积物
序列
<400> 509

[0246]

ttcgagatcg acaccgaatc acgcataggg tgtcaggaga tagtacccta aaaattctga 60
gtaatctaca cctgtcgcgt tctccttgtt gaccatgatt ttattagaga ttggtaaaaa 120
tattagattg aagatcgaat atattagata ttagagaaaa gttattggag atagataatg 180
tcagtgaaaa caattcaagc aaaattagaa tgtgacagag agactttaga gactctatgg 240
agaactcata cggtttttaa tgaaggtctt ttacctattt tagctttatt ttttaagatg 300
aagcgtagag aaacaggatc aacagaatta cacagaaagt tctatgattg gtcagtagta 360
gacaagtact ttactttgct tagaagtgat aattataagg aacagaaaaa agtgattgaa 420
gaagaacttc gagagttttt gaggaaagag cattttcctc aaatggctct gaattaaaaa 480
gtggagaatt tatccaatgc ctatagggaa attggataac cacacaggcc ttggttttga 540
tggtgaatat ctgttgatta tttggtctga gggccttaaa taaattacat ggatagta 598

<210> 510
<211> 1394
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
实验室富集限定的培养基-离子液体和高固体
富集的序列
<400> 510

gaggatggac cagacggcgc cgccgcgctc cttgagcatc gcggcaagaa cgtgcaggcc 60
ggtgacctcc gggttctgac gcgccgcggc gtcggactgg gcgtcggcaa gggtctgctg 120
ggcgagcgtg gtgaagcggt cggtgcgcat gggtcgtgct cccgagggag tctggctgac 180

tcggtctgcc ggattgttgg gcaaaacggg ggccagtcgt gagaggcttt ggaggggggt 240
tttacggagc gcggactgcc ggaatggcag ggggacgccg gccgtcggcc aaccgccccc 300
cgggcgcctc ccggcaagtt tcgttaggat ggtcgccgtg gaggccgcct cagatgggtg 360
ttaaagcctg gaaggcgaag atcatcgtgg ccgacccgga cctggcggcc gctctggacc 420
acaccgtccg gcaatggaac ctctgggcgg aacaggcctg catcgccatc atgcgcctgc 480
gccgcggcga ctacggggac gaggtcctga tcgggacgtt cccgcaggtc cggcgcatct 540

cgcgcgaccc gccggtgctg tcgatcgccg atgtccgggt ggaccagcgg gaggcggatc 600
gcctggccgc ggcgtggctg agcgaactgt cggtcggccc gacgcccttc tgaggtggcg 660
tcggatgggg gcttccggta tactgccgtc gctaccgcct gcacagggcg gcgcggcgga 720
ccggctatag gcccgtccgt cgtacgtcgg gaactgtcga gccgggagtg acgtggcggg 780
gcctcggccc tgttcgaaac cttcacccct ctcgtggggg tggtgggctt cccggcaacc 840
cggcggctga tccatcacgc ggccgccgcg aatatcgagc gcccacgccg acgccgggag 900
gttcgcggaa gtcgcaacgc tctttgacag cccaagttgc gctcaggagg cccgatgacc 960
ggcttggctc tccgttcaaa aacagccgat tcgcggagga acgctccgaa gtgccgaccg 1020
gtcaacctct tacgggcggg agcagatgca gccaaggggc ggactcgaac cgcccctcgc 1080
acgtatccgc cggaaaaggg cgaatgccag tgaccgaaga acgggtccgg ccgttccgca 1140
acactcaatc gccgccgcgg caatcaggca gagcccgcct cagcaacatc atcccccgcg 1200
ccccctgctc cgcccttctg agatcgcgca ggccctcaac ccgtctacgc caccacgtcc 1260
agccgctccc ccactcccgg cgggctcgcg ccctgggcga ccttcgccgc ggcctggatc 1320
atctgcaccg cggcctggcc ctgctgctcg gcgtggtcga ggaccttccg ggcgacgcgg 1380
atgctcgcgg cgga 1394

[0247]

<210> 511
<211> 748
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
水生-淡水-淡水微生物垫序列
<400> 511
gaattgaatg caaccgcgaa gaattacaac atctttggcg gactcactgc gccttcaacg 60
aaaggttgcc atcaatcctt tcggtcctgt tcaaaatgcg tcgtggagaa tggggtgaca 120
ccgaacagat aaagaaaaaa atgcaggctt tcttctttct cgatgggcag gccgtccagg 180
cattctacga ttggaagggc aagccggagc cggaaagaaa aacgtttttc gcggcaatgg 240
acggcaagct gcagccagcg ctgagagtgg aacataaatt ggatggtccg acaccgtttt 300
aaaattcgaa gacttgcggc aggatgcaat caggttacgt ttacgtttgt gggttgtacc 360
cgctcggcgt gctataggca ggccgggacc acaaaagcac tgcgtttgag tgtcaagacc 420
tgctgaccga acggccaaaa gcctagtagg gccagactga ggaggggcaa cccaccaagg 480
gcggcaggaa cacgcagtga gttgttcagc ttgcgttgtg aggactgttg gtgggtcttt 540
gacaactcat tttgcgcggc tttttgccgc gagcgcggag gtggtggtgc gggtgcgtcg 600
ttgcgctcgc acctacttaa cctcttacgg tgcaagggcc tcgtcaacgg aaatggtgag 660
tcgcatgtgt ctcgatcagt tgattccacg caggtgctcg ttatgggact gccaagtccc 720
ggtcaaatca tggtgtgcga gcgcgaca 748
<210> 512

<211> 832
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
水生-非海生盐水和碱性-超盐水湖沉积物
序列
<400> 512

ttctacagca agacttagcc gaaatgcccc caagatgttg cggtgtcagc aaagcaggca 60
atacaatatc cggcaccaaa tcgcccgaag tcttgctgta gagaaaatgt acgcctgacc 120
ccgttgtccc cccttgcccg ccccaagcga tgacggcaag cggagtacgt cttctactga 180
atcgaaagga cccgaaatgt cacaaagggc ccttgccgta cgacttctgg ttgactctga 240
agacgatcgc gcggcgctgt ggcgaacgca ccgcatattc aatgagcgcc tcggatgggt 300
gctgcgtcag atgcatcgga tgaaacgggg cgaaccggac gcgcgcgaac tcccgaaggt 360
caagaagctg cgcgagggcg tgtatgagct agcgaagcct gatcacgcgc ccaaagaagt 420
gaagcctgcc gagcttagga gtgcggttag cgtcgcggcg gccgcgatgt gccgcgggca 480
ggcgagcccc ttctaggcat tgacttccct tctgtacctg atacgatatg gatcagagtt 540
gcgagattac cgtaacctga ccggcagggc tataggccgt gtcggtgagc gccgggaacc 600
agtggttatg gccgcatggg cctacgggtt accccaatgg gcacgaaacc gacccggtcg 660
gcttcgccga tccagcgagg gcggcggcct ttgcccggcg acacaatcgg gaatttcggg 720
gaaaactgcc gcttcgcacc ccgggaaagg ggctggaaag agtgtttttt cgaggcgttt 780
tcgttctaag tccttttgca gcgaggagtt cggtccggaa cttattgtcc aa 832

[0248]

<210> 513
<400> 513
000
<210> 514
<400> 514
000
<210> 515
<400> 515
000
<210> 516
<400> 516
000
<210> 517
<400> 517
000
<210> 518
<400> 518
000
<210> 519
<400> 519
000
<210> 520

[0249]

<400> 520
000
<210> 521
<400> 521
000
<210> 522
<400> 522
000
<210> 523
<400> 523
000
<210> 524
<400> 524
000
<210> 525
<400> 525
000
<210> 526
<400> 526
000
<210> 527
<400> 527
000
<210> 528
<400> 528
000
<210> 529
<400> 529
000
<210> 530
<400> 530
000
<210> 531
<400> 531
000
<210> 532
<400> 532
000
<210> 533
<400> 533
000
<210> 534
<400> 534
000

<210> 535
<400> 535
000
<210> 536
<400> 536
000
<210> 537
<400> 537
000
<210> 538
<400> 538
000
<210> 539
<400> 539
000
<210> 540
<400> 540
000
<210> 541
<400> 541
000
[0250] <210> 542
<400> 542
000
<210> 543
<400> 543
000
<210> 544
<400> 544
000
<210> 545
<400> 545
000
<210> 546
<400> 546
000
<210> 547
<400> 547
000
<210> 548
<400> 548
000
<210> 549
<400> 549

000

<210> 550

<400> 550

000

<210> 551

<400> 551

000

<210> 552

<400> 552

000

<210> 553

<400> 553

000

<210> 554

<400> 554

000

<210> 555

<400> 555

000

<210> 556

<400> 556

000

[0251]

<210> 557

<400> 557

000

<210> 558

<400> 558

000

<210> 559

<400> 559

000

<210> 560

<400> 560

000

<210> 561

<400> 561

000

<210> 562

<400> 562

000

<210> 563

<400> 563

000

<210> 564

[0252]

<400> 564
000
<210> 565
<400> 565
000
<210> 566
<400> 566
000
<210> 567
<400> 567
000
<210> 568
<400> 568
000
<210> 569
<400> 569
000
<210> 570
<400> 570
000
<210> 571
<400> 571
000
<210> 572
<400> 572
000
<210> 573
<400> 573
000
<210> 574
<400> 574
000
<210> 575
<400> 575
000
<210> 576
<400> 576
000
<210> 577
<400> 577
000
<210> 578
<400> 578
000

<210> 579
<400> 579
000
<210> 580
<400> 580
000
<210> 581
<400> 581
000
<210> 582
<400> 582
000
<210> 583
<400> 583
000
<210> 584
<400> 584
000
<210> 585
<400> 585
000
[0253] <210> 586
<400> 586
000
<210> 587
<400> 587
000
<210> 588
<400> 588
000
<210> 589
<400> 589
000
<210> 590
<400> 590
000
<210> 591
<400> 591
000
<210> 592
<400> 592
000
<210> 593
<400> 593

000
<210> 594
<400> 594
000
<210> 595
<400> 595
000
<210> 596
<400> 596
000
<210> 597
<400> 597
000
<210> 598
<400> 598
000
<210> 599
<400> 599
000
<210> 600
<400> 600
000

[0254]

<210> 601
<211> 481
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
废水-厌氧消化器-厌氧消化器污泥
序列
<400> 601

```
Met Lys Ala Thr Thr Ile Thr Lys Ser Leu Arg Phe Trp Ile Asp Glu
1               5                  10                  15
Pro Pro Glu Arg Cys Arg Leu Leu Tyr Gly Val Ser Asp Glu Leu Thr
            20                  25                  30
Asp Ala Tyr Asn Ala Ile Leu Glu Tyr Trp Glu Thr Asp Val Arg Lys
        35                  40                  45
Val Ala Val Glu Ala Thr Glu Ala Ala Ile Glu Ala Tyr His Val Trp
    50                  55                  60
Lys Asp Ala Pro Lys Glu Thr Arg Gly Asp Lys Pro Lys Trp Trp Ser
65                  70                  75                  80
Ala Asp Thr Ala Thr Arg Ala Ala Phe Leu Arg Val Gly Val Ala Ala
                85                  90                  95
Thr Ser Thr Ile Arg Thr Arg Leu Val Asp Asn Met Val Gln Gly Glu
```

[0255]

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

450 455 460

Gln Pro Ala Arg Thr Pro Ala Ala Cys Gly Gly Asp Pro Tyr Ala Gly
465 470 475 480

Glu

<210> 602
<211> 512
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-温泉-热泉序列
<400> 602

[0256]

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

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

[0257]

<210> 603
<211> 481
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
激活的污泥宏基因组序列
<400> 603
Met Lys Ala Thr Thr Ile Thr Lys Ser Leu Arg Phe Trp Ile Asp Glu
1 5 10 15
Pro Pro Glu Arg Cys Arg Leu Leu Tyr Gly Val Ser Asp Glu Leu Thr

[0258]

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

370 375 380
Thr Ser Gln Thr Cys Ser Arg Cys Ala Val Arg Asp Pro Glu Ser Arg
385 390 395 400
Asp Gly Lys Arg Phe Glu Cys Arg His Cys Gly Tyr Lys Gly Asp Ala
405 410 415
Asp Leu Asn Ala Ala Asn Asn Ile Ala Ser Gly Arg Phe Arg Arg Ser
420 425 430
Ile Arg Pro Ser Thr Gln Pro Ala Gly Arg Ile Pro Ser Ala Ala Pro
435 440 445
Asp Cys Thr Gly Glu Ala Val Lys Ala Ser Asp Val Asp Ala Glu Leu
450 455 460
Gln Pro Ala Arg Thr Pro Ala Ala Cys Gly Gly Asp Pro Tyr Ala Gly
465 470 475 480
Glu

<210> 604
<211> 481
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
激活的污泥宏基因组序列

[0259]

<400> 604
Met Lys Ala Thr Thr Thr Thr Lys Ser Leu Arg Phe Trp Ile Asp Glu
1 5 10 15
Pro Pro Glu Arg Cys Arg Leu Leu Tyr Gly Val Ser Asp Glu Leu Thr
20 25 30
Ala Ala Tyr Asn Ala Ile Leu Glu Tyr Trp Glu Thr Asp Val Arg Ser
35 40 45
Val Ala Val Glu Val Thr Glu Ala Ala Gln Ala Ala Tyr His Val Trp
50 55 60
Lys Asp Ala Pro Lys Glu Thr Arg Gly Asp Lys Pro Lys Trp Trp Ser
65 70 75 80
Ala Asp Thr Ala Thr Arg Ala Ala Phe Leu Arg Val Gly Val Ala Ala
85 90 95
Thr Ser Thr Ile Arg Thr Arg Leu Val Asp Asn Met Val Gln Gly Glu
100 105 110
Trp Leu Arg Asp Val Lys Thr Tyr Met Ser Arg Arg Phe Lys Asp His
115 120 125
Ile Ala Ala Gly Ile Arg His Arg Ala Lys Gly Val Val Ile Ser Leu
130 135 140
Asp Gln Ser Thr Met Lys Pro Asp Val Gln Leu Gly Asp Asp Gly Arg
145 150 155 160
Trp Thr Ile Arg Cys Ala Leu Trp Ala Gln Gly Thr Glu Glu Ser Ala

[0260]

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

<210> 605
<211> 481
<212> PRT

<213> 未知
<220>
<223> 未知的描述:
激活的污泥宏基因组序列
<400> 605

[0261]

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

305 310 315 320
Asp Cys Cys Arg Arg Asp Arg Val Gly Thr Leu Arg Val Glu Asp Leu
325 330 335
Thr Gly Ile Arg Asp Gln Gly Ser Asp Asp Ala Asp Arg Asn Phe Ala
340 345 350
Leu Gly Ala Arg Phe Pro Tyr Tyr Lys Leu Gln Thr Tyr Leu Glu Gln
355 360 365
Ala Ala Ala Ser Ala Gly Val Arg Leu Glu Lys Val Gln Pro Ala Gly
370 375 380
Thr Ser Gln Thr Cys Ser Arg Cys Ala Val Arg Asp Pro Glu Ser Arg
385 390 395 400
Asp Gly Lys Arg Phe Glu Cys Arg His Cys Gly Tyr Lys Gly Asp Ala
405 410 415
Asp Leu Asn Ala Ala Asn Asn Ile Ala Ser Gly Arg Phe Arg Arg Ser
420 425 430
Ile Arg Pro Ser Thr Gln Pro Ala Gly Arg Ile Pro Ser Ala Ala Pro
435 440 445
Asp Cys Thr Gly Glu Ala Val Lys Ala Ser Asp Val Asp Ala Glu Leu
450 455 460
Gln Pro Ala Arg Thr Pro Ala Ala Cys Gly Gly Asp Pro Tyr Ala Gly
465 470 475 480
Glu

[0262]

<210> 606
<211> 419
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
激活的污泥宏基因组或废水-厌氧
消化器-厌氧消化器污泥序列
<400> 606
Met Ile Lys Thr Val Thr Phe Tyr Ile Ala Glu Met Pro Arg Glu Asp
1 5 10 15
Lys Arg Leu Cys Asp Arg Leu Cys Ile Asp Leu Gly Ala Ala Gly Ala
20 25 30
Arg Met Val Arg Glu Trp Phe Arg Ala Cys Asp Glu Thr Pro Pro Gly
35 40 45
Ala Ala Ile Asp Trp Pro Glu Val Gln Arg Arg Gly Val Ala Val Gly
50 55 60
Lys Ala Met Val Glu Ala Ala Gly Tyr Arg Thr Glu Leu Gly Asn Asp
65 70 75 80
Leu Ala Arg Ala Glu Phe Phe Ala Arg Phe Lys Met Glu Phe Lys Lys

85 90 95
Val Leu Lys Gly Glu Arg Ala Ser Leu Pro Phe Pro Gly Arg Ala Pro
100 105 110
Phe Ile Phe Leu Arg Met Asp Pro Gly His Glu Thr Ala Ile Ile Glu
115 120 125
Gln Arg Asp Asp Ala Trp Tyr Met Arg Asn Pro Arg Phe Tyr Pro Gly
130 135 140
Gly Asp Ser Asn Glu Ala Asn Arg Arg Glu Trp Arg Leu Lys Pro Val
145 150 155 160
Asp Arg Lys Gly Trp Ala Glu Asn Ile Leu Glu Ala Ala Glu Lys Phe
165 170 175
Ala Cys Val Arg Ile Met Pro Asp Arg Lys Lys Pro Gly Ser Trp Val
180 185 190
Val Lys Ile Thr Val His Leu Pro Asp Ala Leu Glu Arg Lys Leu Val
195 200 205
Glu Glu His Glu Cys Trp Ala Gly Ile Asp Leu Gly Ile Asn Cys Pro
210 215 220
Ala Val Leu Ser Ile Pro Asp Arg Asp Lys Gly Gly Phe Val Arg Phe
225 230 235 240
Phe Gly Met Asp Lys Tyr Thr Ala Leu Trp Ala Lys Leu Asn Arg Tyr
245 250 255
[0263] Glu Glu Arg Lys Arg Ala Leu Asn Arg Val Gly Gln Gln His Ala Ala
260 265 270
Arg Asp Leu Arg Ala Lys Ile Thr Gly Leu Arg Gln Tyr Ile Asn Glu
275 280 285
Gln Ile Ser Arg Glu Ile Ile Asp Leu Cys Leu Arg Met His Val Thr
290 295 300
Gly Ile Arg Met Glu Asp Leu Lys Gly Leu Arg Thr Glu Gly Thr Gly
305 310 315 320
Lys Leu Lys Tyr Trp Pro Arg Phe His Leu Thr Thr Arg Ile Glu Gln
325 330 335
Lys Ala Arg Glu Val Gly Leu Ala Phe Glu Gln Ile Arg Ala Ala Gly
340 345 350
Thr Ser Gln Thr Cys Ser Val Cys Gly His Arg His Arg Asp Asn Arg
355 360 365
His Gly Ser Asp Phe Leu Cys Leu Arg Cys Gly Phe Ala Arg His Ala
370 375 380
Asp Val Asn Ala Ala Asn Asn Ile Ala Arg Gly Met Gln Asp Phe Asp
385 390 395 400
Leu Pro Ala Gly Ser Ala Ala Pro Gly Arg Ala Leu Arg Leu Glu Leu
405 410 415
Ser Ala Gln

<210> 607

<211> 341
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水-淡水湖沉积物或淡水
宏基因组序列
<400> 607

Met Thr Ala Thr Lys Gln Gln Ile Leu Gln Thr Glu Tyr Asp Asn Leu
1 5 10 15
Gln His Phe Leu Gln Thr Gly Glu Asp Leu Gly Val Tyr Ser Ala Asn
20 25 30
Lys Gln Gln Ala Lys Arg Tyr Tyr Lys Thr Ile Lys Pro Glu His Gln
35 40 45
Tyr Pro Leu Ser Ile Arg Asn Asp Leu Ile Gln Val Gln His Lys Pro
50 55 60

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

[0264]

275 280 285

Thr Cys His Arg Thr Ser Phe Asn Ala Asp Tyr Asn Gly Ala Leu Asn
290 295 300
Ile Leu Lys Arg Ala Thr Ser Tyr Met Asp Ile Ala Gly Ala Val Ser
305 310 315 320
Glu Pro Ala Arg Asp Trp Gly Asn Asp Asp Gln Thr Ser Glu Ala Thr
325 330 335
Gln Leu Phe Glu Trp
340

<210> 608
<211> 353
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水-淡水湖沉积物或淡水
宏基因组序列

[0265]

<400> 608
Met Ala Gln Lys Thr Ile Thr Ala Lys Ile Val Ala Met Thr Ala Thr
1 5 10 15
Lys Gln Gln Ile Leu Gln Thr Glu Tyr Asp Asn Leu Gln His Phe Leu
20 25 30
Gln Thr Gly Glu Asp Leu Gly Val Tyr Ser Ala Asn Lys Gln Gln Ala
35 40 45
Lys Arg Tyr Tyr Lys Thr Ile Lys Pro Glu His Gln Tyr Pro Leu Ser
50 55 60
Ile Arg Asn Asp Leu Ile Gln Val Gln His Lys Pro Asn Ser Ile Ala
65 70 75 80
Glu Tyr Trp Leu Lys Ile Pro Val Lys Gly Arg Arg Gly Gly Leu Trp
85 90 95
Val Ala Ile Arg Pro His Gln Gln Ile Pro Asp Asn Ala Glu Phe Leu
100 105 110
Glu Ser Lys Ile Arg Arg Lys Arg Lys Gly Trp Phe Ala Asn Ile Val
115 120 125
Ile Gln Lys Gln Ile Glu Pro Ile Arg Thr Lys Asn Ile Leu Ala Val
130 135 140
Asp Leu Gly Glu His Val Ile Ala Thr Val Cys Gly Ser Phe Asp Asn
145 150 155 160
Gln Arg Pro Arg Phe Tyr Gly Arg Asp Val Arg Gly Ile Arg Arg His
165 170 175
Tyr Ala Trp Leu Arg Lys Arg Leu Gly Glu Lys Lys Leu Leu Gly Val
180 185 190
Ile Arg Arg Ile Lys Asp Thr Glu Lys Arg Lys Val Asn Asp Thr Leu
195 200 205

```
His Lys Ile Ser Arg Ser Ile Val Asp Leu Ala Gln Glu His Asn Ala
    210                 215                 220
Thr Ile Val Leu Gly Glu Leu Gln Gly Ile Arg Asn His Ala Lys Gly
225                 230                 235                 240
Arg Arg Met Asn Arg Ile Val Ser Asn Met Pro Tyr Leu Lys Leu Ser
                245                 250                 255
Gln Met Ile Glu Tyr Lys Ala Glu Trp Leu Gly Ile Pro Val Ile Lys
            260                 265                 270
Met Gln Glu His Asp Thr Ser Lys Thr Cys Ser Arg Cys Asn Glu Tyr
        275                 280                 285
Gly Lys Arg Pro Thr Gln Gly Gln Phe Arg Cys Asp Thr Cys His Arg
    290                 295                 300
Thr Ser Phe Asn Ala Asp Tyr Asn Gly Ala Leu Asn Ile Leu Lys Arg
305                 310                 315                 320
Ala Thr Ser Tyr Met Asp Ile Ala Gly Ala Val Ser Glu Pro Ala Arg
                325                 330                 335
Asp Trp Gly Asn Asp Asp Gln Thr Ser Glu Ala Thr Gln Leu Phe Glu
            340                 345                 350
Trp
```

[0266]

<210> 609
<211> 373
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水-沉积物或淡水沉积物宏基因组
序列

```
<400> 609
Met Ala Asn Arg Glu Leu Val Val Val Met Gln Val Gln Lys Val Val
1               5                   10                  15
Lys Gly Lys Val Val Gly Leu Thr His Ala Lys Arg Gln Ile Leu Asp
            20                  25                  30
His Glu Tyr Glu Gly Leu Gln Arg Phe Leu Gln Gly Asp Ser Asn Val
        35                  40                  45
Glu Leu Tyr Ser Ala Asn Lys Gln Gln Ala Lys Arg Phe Tyr Lys Thr
    50                  55                  60

Val Lys Pro Glu Arg Glu Tyr Pro Leu Ser Ile Arg Lys Asp Leu Ile
65                  70                  75                  80
Arg Val Glu Arg Arg Asp Thr Lys Ile Ala Arg Tyr Trp Ala Arg Ile
                85                  90                  95
Pro Ala Lys Leu Arg Arg Gly Gly Val Trp Val Ala Val Lys Pro His
            100                 105                 110
```

[0267]

Cys Glu Ile Thr Asp Asp Leu Glu Ile Cys Glu Ser Lys Leu Leu Arg
115 120 125
Arg Asn Gly Glu Tyr Phe Leu His Leu Ala Val Lys Arg Glu Val Glu
130 135 140
Leu Asn Thr Ser Tyr Ser Ser Val Leu Ala Ile Asp Leu Gly Ile Arg
145 150 155 160
His Val Ala Cys Ser Val Asp Thr Arg Ser Gly Glu Thr His Phe His
165 170 175
Gly Arg Glu Leu Arg Arg Val Arg Gly His Tyr Phe His Leu Arg Arg
180 185 190
Arg Leu Gly Glu Lys Lys Leu Leu Lys Val Ile Lys Arg Ile Gly Ala
195 200 205
Arg Glu Ser Arg Ile Thr Asn Asp Gln Leu His Lys Ile Ala Lys Ala
210 215 220
Ile Val Asp Glu Ala Glu Arg Leu Asn Ala Leu Ile Ala Ile Gly Asn
225 230 235 240
Pro Lys Gly Ile Arg Arg Asn Gly Lys Gly Arg Arg Phe Asn Arg Lys
245 250 255
Leu Asn Ser Trp Pro Phe Trp Lys Leu Arg Gln Tyr Ile Glu Tyr Lys
260 265 270
Ala Asn Trp Arg Gly Ile Ala Val Val Glu Val Asn Glu Ala Tyr Thr
275 280 285
Ser Gln Arg Cys Trp Arg Cys Gly Thr Ile Gly Thr Arg Ser Gly Thr
290 295 300
His His Gly Leu Phe Glu Cys Pro Arg Cys Gly Leu Arg Glu Asn Ala
305 310 315 320
Asp Arg Asn Gly Ala Phe Asn Ile Gly Arg Arg Ala Leu Gly Gln Val
325 330 335
Ser Lys Val Gly Ala Val Val Ser Gln Pro Val Thr Gly Ala Val Pro
340 345 350
Phe Ser Glu Leu Gly Ser Val Val Gln Cys Pro Thr Pro Glu Ala Thr
355 360 365
Leu Phe Lys Gly Trp
370

<210> 610
<211> 373
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水-沉积物序列
<400> 610
Met Ala Asn Arg Glu Leu Val Val Val Met Gln Val Gln Lys Val Val
1 5 10 15

[0268]

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

Leu Phe Lys Gly Trp
370

<210> 611
<211> 342
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水-沉积物序列
<400> 611

[0269]

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

Ile Leu Val Val Asn Ala Asn Glu Glu Trp Thr Ser Lys Thr Cys Ser
275 280 285
Phe Cys Gly Glu Ile Ser Asp Arg Pro Arg Gln Asp Leu Ile Val Cys
290 295 300
Pro His Gly His Arg Leu Asn Ala Asp Ile Asn Ala Cys Arg Asn Leu
305 310 315 320
Ile Lys Arg Phe Ile Asp His Trp Leu Ile Asp Gly Ala Ile Leu Ala
325 330 335
Ser Pro Lys Leu Pro His
340

<210> 612
<211> 368
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-淡水-沉积物序列
<400> 612

[0270]

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

Lys Gln Leu Ser Asp Lys Glu His Arg Ile Val Asn Asp Leu Leu His
210 215 220
Lys Ile Ser Lys Gln Ile Val Glu Lys Ala Glu Gln Thr Asp Ser Tyr
225 230 235 240
Ile Val Leu Gly Asp Leu Lys Gly Ile Arg Asn Ser Thr Lys Asn Lys
245 250 255
Gly Arg Asn Phe Arg Arg Ile Val Ser Asn Met Pro Tyr Leu Lys Leu
260 265 270
Thr Gln Phe Ile Glu Tyr Lys Ala Lys Glu Lys Gly Ile Lys Ile Ile
275 280 285
Arg Ile Ser Glu Lys Gly Thr Ser Lys Ile Cys Ser Lys Cys Gly Tyr
290 295 300
Glu Asp Lys Gly Asn Arg Lys Ser Gln Gly Leu Phe Lys Cys Lys Gln
305 310 315 320
Cys Gly Phe Glu Leu Asn Ala Asp Ala Asn Gly Ala Arg Asn Ile Leu
325 330 335
Lys Phe Ser Asp Gly Tyr Thr Leu Pro Glu Arg Ala Met Ser Glu Leu
340 345 350
Ala Leu Asn Ser Asn Gln Leu Glu Lys Pro Thr Cys Phe Ser Gly Gly
355 360 365

[0271]

<210> 613
<211> 343
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-海生-深部地下序列
<400> 613
Met Leu Val Glu Lys Thr Val Lys Ala Lys Val Ile Ala Pro Thr Asn
1 5 10 15
Arg Lys Gln Glu Leu Leu Asp Arg Glu Tyr Gln Ser Phe Gln Tyr Thr
20 25 30
Leu Tyr Gly Gly Glu Ala Asn Leu Tyr Ser Ala Thr Lys Gln Gln Ala
35 40 45
Glu Arg Phe Leu Arg Lys Ile Lys Thr Pro Lys His Arg His Tyr Pro
50 55 60
Leu Ile Leu Arg Arg Asp Val Ile Arg Leu Glu Arg Gln Glu Thr Lys
65 70 75 80
Val Ala Pro Tyr Trp Phe Lys Phe Pro Val Phe Gly Val Arg Gly Gly
85 90 95
Ile Trp Leu Gly Leu Lys Pro His Cys Asp Ile Pro Val Asp Cys Ser
100 105 110
Leu Arg Glu Ala Lys Leu Ile Arg Arg Lys Asn Ala Trp Phe Ile His
115 120 125

Leu Thr Ile Gln Lys Glu Val Pro Glu Ser Ile Ile Asp Glu His His
130 135 140
Ile Leu Ala Val Asp Leu Gly Glu Arg Tyr Leu Ala Thr Val Cys Gly
145 150 155 160
Asn Asn Gly Thr Asn Pro Gln Phe Leu Gly Lys Glu Val Arg Gly Ile
165 170 175
Arg Arg His Tyr Ala Trp Ile Arg Lys Arg Leu Gly Glu Arg Lys Leu
180 185 190
Leu Asn Thr Thr Arg Lys Val Gly His Thr Glu Gln Lys Lys Val Asn
195 200 205
Asp Ile Cys His Gln Ile Ser Arg Arg Ile Val Asn Glu Ala Arg Glu
210 215 220
Thr Asp Ser Val Ile Ile Leu Gly Asp Leu Lys Gly Ile Arg Lys Asn
225 230 235 240
Ala Lys Gly Lys Arg Met Asn Arg Ile Val Ser Asn Met Pro Tyr Tyr
245 250 255
Lys Leu Thr Gln Phe Ile Lys Tyr Lys Ala Ala Trp Asp Gly Val Pro
260 265 270
Val Leu Val Ile Pro Glu Ala Tyr Thr Ser Lys Gly Cys His Arg Cys
275 280 285
Gly Ser Leu Gly Ser Arg Pro Tyr Arg Gly Leu Phe Leu Cys His Ser
290 295 300

[0272]

Cys Lys Leu His Tyr Asn Ala Asp Leu Asn Gly Ala Arg Asn Ile Leu
305 310 315 320
Lys Arg Ala Leu Asp Tyr Met Ser Ser Ala Gly Ala Ala Leu Ala Gln
325 330 335
Pro Val Thr Gln Ser Leu Asn
340

<210> 614
<211> 365
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-海生-海生沉积物序列
<400> 614
Met Lys Val Lys Lys Thr Ile Lys Ala Lys Ile Leu Glu Leu Arg Lys
1 5 10 15
Gly Lys Glu Glu Leu Leu Lys Arg Glu Tyr Glu Asn Trp Gln Arg Tyr
20 25 30
Leu Arg Gly Asp Lys Thr Val Pro Leu Tyr Ser Ala Thr Lys Gln Gln
35 40 45
Ala Asp Arg Leu Leu Lys Arg Leu Lys Gly Arg Leu Lys Pro Asn Lys
50 55 60

Glu Tyr Pro Met Ile Leu Arg Arg Asp Val Tyr Arg Ala Asn Thr Lys
65 70 75 80
Leu Thr Pro Tyr Trp Leu Lys Ile Ser Ile Tyr Gly Val Arg Gly Gly
85 90 95
Ile Asn Val Pro Ile Lys Thr His Glu Pro Ile Thr Glu Asp Met Val
100 105 110
Cys Arg Glu Ala Lys Ile Ile Arg Lys Gly Asp Glu Trp Phe Val Tyr
115 120 125
Ile Thr Val Glu Lys Glu Val Glu Glu Arg Asn Pro Lys Ser Ile Leu
130 135 140
Ala Ile Asp Leu Gly Ile Arg Trp Met Ala Thr Thr Val Asn Ser Asn
145 150 155 160
Asn Pro Lys Pro Asn Phe Tyr Gly Lys Glu Leu Arg Lys Val Lys Gly
165 170 175
His Tyr Phe Tyr Leu Arg Arg Ser Leu Thr Leu Lys Lys Ala Tyr Lys
180 185 190
Thr Ile Lys Lys Ile Gly Leu Lys Glu Arg Arg Val Val Asn Asp Ile
195 200 205
Leu His Lys Ile Ser Arg Ala Ile Val Asn Glu Ala Leu Glu Asn Asp
210 215 220
Ser Met Ile Val Leu Gly Asn Leu Lys Gly Ile Arg Arg Asn Gly Lys
225 230 235 240
[0273] Gly Arg Arg Phe Asn Arg Lys Leu Asn Asn Gly Phe Pro Tyr Tyr Arg
245 250 255
Leu Ser Gln Phe Ile Glu Tyr Lys Ala Lys Trp Leu Gly Ile Lys Val
260 265 270
Ile Lys Ile Ser Glu Lys Asn Thr Ser Lys Leu Cys His Lys Cys Gly
275 280 285
Glu Glu Gly Val Arg Val Gly Ser Ser Phe Lys Cys Pro Asn Cys Gly
290 295 300
Tyr Ser Cys His Ala Asp Tyr Asn Gly Ala Met Asn Ile Leu Lys Arg
305 310 315 320
Ala Met Gly Leu Val Ser Val Ala Gly Ala Gly Leu Thr Gln Pro Arg
325 330 335
Thr Arg Tyr Asp Glu Ala Arg Arg Pro Lys Asn Arg Glu Ser Pro His
340 345 350
Leu Ser Gly Lys Ser Val Lys Ser Tyr Ser Phe Lys Phe
355 360 365

<210> 615
<211> 382
<212> PRT
<213> 未知
<220>
<223> 未知的描述:

水生-非海生盐水和碱性-超盐水或
利氏盐环境杆菌（Halohasta litchfieldiae）tADL 序列

<400> 615

Met Lys Gln Thr Ala Glu Lys Thr Val Val Ala Lys Ile Phe Thr Asn
1 5 10 15
Asp Leu Thr Gln Ser Lys Leu Asp Ala Ile Asn His Glu Tyr Glu Ala
20 25 30
Phe Gln Arg Tyr Ile Arg Gly Asp Glu Glu Ala Asp Leu Tyr Ser Ala
35 40 45
Thr Lys Gln Cys Ala Asp Ala Tyr Val Asp Thr Asp Asn Leu Arg Asp
50 55 60
Asp His Asp Tyr Pro Trp Phe Ile Arg Asn Asp Val Phe Asp Val Gly
65 70 75 80

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

[0274]

Ala Asp Lys Asn Gly Ala Thr Asn Ile Ala Lys Lys Gly Leu Gly Lys
325 330 335
Asp Ile Ala Cys Pro Leu Ser Ser Leu Gly Ala Val Cys Glu Pro Val
340 345 350
Leu Gly Pro Ser Gly Ala Asp Gln Ala Ser Ala Thr Val Cys Leu Arg
355 360 365
Ala Asp Leu Glu Ala Pro Ala Ser Thr Lys Arg Thr Val Arg
370 375 380

<210> 616
<211> 432
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-非海生盐水和碱性-盐水湖或盐
湖宏基因组序列

<400> 616

[0275]

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

210 215 220
Leu Ser Ile Pro Asp Lys Gly Phe Val Gln Phe Met Gly Arg Glu Arg
225 230 235 240
Asp Glu Tyr Pro Arg Leu Trp Glu Lys Leu Arg Glu Tyr Glu Asp Arg
245 250 255
Lys Arg Arg Leu Asn Arg Ala Gly Lys Arg Arg Ala Ala Arg Asp Leu
260 265 270
Arg Pro Arg Ile Thr Asn Val Arg Lys Tyr Ile Asn Glu Ala Ile Ser
275 280 285
Lys Asp Val Ala Asp Thr Cys Arg Ala Leu Gly Val Thr His Leu Arg
290 295 300
Met Glu Asp Leu Lys Gly Leu Ser Arg His Lys Ser Met Thr Gly Lys
305 310 315 320
Leu Arg Tyr Trp Pro Arg Tyr His Leu Gln Glu Arg Ile Glu Arg Lys
325 330 335
Cys Cys Glu Val Gly Ile Glu Val Glu Arg Ile Lys Pro Ala Gly Thr
340 345 350
Ser Gln Thr Cys Ser Val Cys Gly His Arg Glu Lys Ala Asn Arg Asn
355 360 365
Gly Ala Gln Phe Val Cys Ile Glu Cys Gly Tyr Gln Gln His Ala Asp
370 375 380
[0276] Leu Asn Ala Ala Asn Asn Ile Ala Arg Gly Met Gln Thr Phe Gly Pro
385 390 395 400
Cys Ala Lys Ser Ser Ala Ser Asp Asp Ala Ile Gly Leu Arg Phe Asp
405 410 415
Ser Glu Gly Gln Glu Ser Ala Leu Gly Gln Ala Lys Arg Leu Phe Ala
420 425 430

<210> 617
<211> 521
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-温泉序列
<400> 617
Met Lys Lys Val Val Ile Thr Val Thr Thr Gly Val Glu Phe Leu Gln
1 5 10 15
Pro Gln Asp Glu Gln Ile Val Arg Lys Ala Ala Glu Trp Tyr Thr Lys
20 25 30
Ala Arg Glu Ala Leu Phe Ala His Ala Leu Gln Gln Gly Lys Val Lys
35 40 45
Glu Ala Ala Ser Thr Leu Arg Thr Leu Asn Phe Pro Met Ala Pro His
50 55 60
Ser Arg Leu Pro Asp Leu Leu Thr Arg Asp Val Gly Gly Ile Leu Lys

65 70 75 80
Ala Leu Lys Glu Lys Arg Glu Gly Tyr Leu Gly Glu Phe Lys Arg Trp
85 90 95
Trp Glu Glu Asp Cys Arg Val Ala Asn Glu Lys Gly Ser Gln Ala Val
100 105 110
Ala Asp Leu Lys Thr His Val Gly Lys Leu Leu Asn Leu Val His Gln
115 120 125
Leu Pro Trp Asn Ser Pro Val Trp Asp Gly Glu Ala Gln Glu Ala Phe
130 135 140
Lys Asn Leu Lys Glu Ala Trp Ala Val Val Arg Gln Lys Asp His Glu
145 150 155 160
Gln Leu Lys Ile Arg Trp Arg Ala Glu Val Ser Pro Leu Arg Arg Pro
165 170 175
Pro Arg Thr Pro Lys Phe Gln Arg Leu Val Phe Asn Ala Asp Gly Arg
180 185 190
Val Ser Gly Gly Ser Lys Ser Glu Ala Leu Gln Ile Glu Glu Arg Glu
195 200 205
Asp Gly Asp Leu Tyr Leu Arg Leu Gln Leu Arg Asp Pro Glu Gly Val
210 215 220
Val Gly Gly Gly Ala Tyr Gln Glu Val Trp Leu Arg Leu Val Val Glu
225 230 235 240
[0277] Asp Glu Glu Arg Lys Gln Met Leu Arg Asp Phe Ala Pro Lys Tyr Leu
245 250 255
Gln Leu Arg Gln Asp Gly Lys Arg Trp Ile Ala His Ile Val Ile Glu
260 265 270
Arg Glu Ile Lys Phe Gly Leu Lys Gly Ile Met Trp Thr Ala Gly Asp
275 280 285
Tyr Leu Tyr Gly Val Asp Ile Gly Ala Arg Lys Tyr Leu Leu Ala Ala
290 295 300
Val Ala Val Pro Leu Lys Pro Gly Leu Pro Lys Leu Pro Asn Ile Ala
305 310 315 320
Ile Pro Ser Gly Lys Phe Trp Gln Ala Val Glu Arg Val Glu Ala Glu
325 330 335
Trp Arg Lys Trp Gln Arg Leu Arg Thr Gly Lys Lys Asp Lys Glu Gly
340 345 350
Lys Val Val Leu Lys Lys Arg Thr Ala Trp Arg Val Leu Lys Arg Lys
355 360 365
Arg Arg Met Arg Arg Glu Val Ala Lys Asn Met Ile Ala Glu Ala Val
370 375 380
Asn Thr Phe Ile Asn Phe Ile Lys Gly Asp Asn Val Asp Ile Arg Leu
385 390 395 400
Trp Leu Ala Ile Glu Gln Ile Lys Asn Leu Asn Leu Ser Pro Ala Phe
405 410 415
Gln Asp Arg Lys Ser Arg Trp Arg Gln Arg Arg Trp Ala Tyr Arg Leu

420 425 430

Leu Leu Glu Met Leu Arg Tyr Lys Ala Glu Trp Glu Gly Ile Arg Val
435 440 445

Ile Glu Val Asn Pro Ala His Thr Ser Gln Ile Cys Pro Arg Cys Gly
450 455 460

Lys Lys Asp Gly Lys Ala Arg Gln Gly Asn Arg Phe Cys Cys Thr Lys
465 470 475 480

Cys Lys Phe Gln His Asn Ala Asp Phe Val Ala Ala Tyr Asn Ile Ala
485 490 495

Val Arg Ala Leu Lys Lys Leu Gly Met Arg Glu Glu Ile Gly Trp Lys
500 505 510

Val Leu Thr Gly Lys Pro Arg Val Leu
515 520

<210> 618

<211> 546

<212> PRT

<213> 未知

<220>

<223> 未知的描述:
水生-温泉-热泉序列

<400> 618

[0278]

Met Lys His Leu Val Ser Asn Ala Gln Val Asn Glu Lys Asn Glu Ala
1 5 10 15

Arg Leu Thr Leu Lys Cys Gly Leu Glu Phe Glu Asn Ala Glu Gly Glu
20 25 30

Glu Leu Val Lys Arg Ala Ala Glu Trp Tyr Cys Lys Ala Arg Ile Phe
35 40 45

Leu Leu Ser Tyr Ala Leu Lys Asn Asn Lys Thr Arg Val Gln Lys Ile
50 55 60

Ala Ser Leu Leu Lys Ser Glu Gly Phe Pro Ser Pro Pro His Pro Arg
65 70 75 80

Leu Ile Asp Leu Leu Gln Arg Thr Ile Thr Pro Ala Ile Lys Asn Met
85 90 95

Arg Arg Leu Arg Arg Arg Tyr Pro Lys Thr Leu Arg Ala Trp Phe Asp
100 105 110

Arg Val Lys Lys Phe Ala Glu Glu Lys Gly Ser Asp Ala Ala Glu Arg
115 120 125

Leu Gln Pro Tyr Val Glu Asp Leu Leu Gln Leu Val Lys Asn Leu Pro
130 135 140

Ile Asn Ser Pro Ala Trp Asp Thr Asp Ala Gln Glu Lys Trp Lys Glu
145 150 155 160

Leu Lys Lys Leu Trp Gly Lys Val Arg Glu Ser Asp Trp Lys Lys Gly
165 170 175

Ile Val Trp Lys Phe Gln Ile Pro Asn Pro Lys Lys Pro Pro Ser Thr

180 185 190

Pro Ala Phe Arg Cys Ala Asp Ser Ser Asn Ile Phe Val Thr Ala Asp
195 200 205

Tyr Arg Met Ser Gly Lys Asn Ala Arg Asn Ser Tyr Asp Met Lys Thr
210 215 220

Asp Ser Glu Glu Asn Leu Trp Val Leu Val Arg Leu Asp Asp Val Asp
225 230 235 240

Gly Asp Val Lys Val Arg Lys Asp Val Trp Ile Lys Leu Cys Val Asp
245 250 255

Lys Ser Arg Tyr Asp Leu Leu Arg Ile Gly Lys Ile Thr Glu Ile Lys
260 265 270

Ile Leu Arg Asn Arg Asp Gly Arg Trp Glu Gly Arg Ile Val Val Lys
275 280 285

Met Asn Ala Gly Tyr Asn Ser Pro Gln Val Val Val Gly Val Asp Val
290 295 300

Gly Ala Ser Glu His Leu Ile Ala Ala Val Gly Ile Pro Leu Asp Gly
305 310 315 320

Ser Asn Asn Asn Phe Leu Leu Ser Val Pro Ala Ser Tyr Phe Trp Lys
325 330 335

Arg Val Glu Gln Ala Glu Arg Leu Trp Arg Arg Trp Gln Arg Leu Tyr
340 345 350

[0279]

Thr Gly Arg Thr Ser Glu Lys Lys Lys Ala Pro Ile Leu Pro Arg Arg
355 360 365

Arg Lys Ala Tyr Gln Ile Leu Lys Arg Asn Arg Asn Trp Arg Arg Glu
370 375 380

Val Thr Lys His Cys Ile Phe His Ala Val Lys Lys Phe Val Gln Leu
385 390 395 400

Leu Pro Pro Asn Ser Val Val Val Val Glu Asn Leu Lys Asp Ile Pro
405 410 415

Pro Asn Val Trp Lys Glu Asp Lys Asp Lys Arg Phe Arg Arg Trp Val
420 425 430

Tyr Ala Phe Leu Leu Phe Thr Leu Lys Tyr Lys Cys Gln Gln Arg Lys
435 440 445

Ile Ala Val Ile Ala Val Ser Pro His Asn Thr Ser Gln Thr Cys Pro
450 455 460

Arg Cys Gly Val Val Asp Pro Lys Ser Arg Gly Leu Trp Lys Asp Gly
465 470 475 480

Glu Leu Ile Ser Arg Lys His Phe Cys Cys Thr Gln Cys Lys Phe Thr
485 490 495

His Asn Ala Asp Phe Val Ala Ala His Asn Ile Ala Val Leu Gly Ala
500 505 510

Lys Ala Leu Gly Ile Thr Leu Pro Glu Ser Trp Arg Ala Leu Arg Gly
515 520 525

Glu Val Gln Tyr Val Ile Pro Tyr Gln Asn Met Pro Lys Val Arg Lys
530 535 540
Ser Trp
545
<210> 619
<211> 543
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-温泉-热泉序列
<400> 619
Met Asn Asn Lys Gln Val Thr Phe Gly Gly Leu Phe Leu Met Ala Leu
1 5 10 15
Gln Lys Lys Arg Val Arg Arg Thr Ile Val Ala Gly Val Glu Ile Glu
20 25 30
Ser Ala Glu Glu Glu Asn Leu Val Thr Glu Ala Ala Lys Trp Tyr Leu
35 40 45
Gln Ala Lys Gln Lys Leu Leu Gln Tyr Ala Ile Glu Lys Arg Lys Glu
50 55 60
Arg Ile Thr Ala Pro Ala Arg Ile Leu Lys Asp Glu Leu Lys Leu Pro
65 70 75 80
[0280]
Glu Pro Pro His Ser Arg Leu Pro Asp Glu Leu Ala Gln Val Ile Asn
85 90 95
Gly Val Leu Lys Ala Leu Lys Lys Arg Arg Thr Glu Tyr Leu Arg Arg
100 105 110
Phe Gln Asp Trp Trp Asn Lys Val Val Glu Val Gly Asn Gln His Gly
115 120 125
Leu Glu Ser Tyr Arg Ala Met Lys Pro Lys Val Glu Ala Leu Ile Ser
130 135 140
Ser Val Pro Asn Asn Pro Phe Asp Pro Thr Trp Asp Ser Asp Ile Gln
145 150 155 160
Lys Arg Phe Ala Glu Val Lys Glu Ile Trp Ser Lys Ile Arg Leu Lys
165 170 175
Asp Ser Asn Glu Arg Asn Phe Arg Trp Arg Tyr His Val Pro Pro Leu
180 185 190
Val Lys Pro Pro Lys Thr Pro Glu Phe Lys Leu Leu Gly Phe Asp Ala
195 200 205
Gly Ala Arg Val Ser Gly Gly Gly Lys Ala Lys Thr Leu Thr Ile Lys
210 215 220
Ser Gln Asp Gly Glu Ile Phe Leu Arg Leu Leu Leu Arg Gln Pro Gly
225 230 235 240
Val Ser Asn His Gly Arg Gly Ser Tyr Arg Glu Ile Trp Leu His Ile
245 250 255

[0281]

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

<210> 620
<211> 530
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-温泉-热泉序列
<400> 620

[0282]

Met Ala Leu Gln Lys Lys Arg Val Arg Arg Thr Ile Val Ala Gly Val
1 5 10 15
Glu Ile Glu Ser Ala Glu Glu Glu Asn Leu Val Thr Glu Ala Ala Lys
20 25 30
Trp Tyr Leu Gln Ala Lys Gln Lys Leu Leu Gln Tyr Ala Ile Glu Lys
35 40 45
Arg Lys Glu Arg Ile Thr Ala Pro Ala Arg Ile Leu Lys Asp Glu Leu
50 55 60
Lys Leu Pro Glu Pro Pro His Ser Arg Leu Pro Asp Glu Leu Ala Gln
65 70 75 80
Val Ile Asn Gly Val Leu Lys Ala Leu Lys Lys Arg Arg Thr Glu Tyr
85 90 95
Leu Arg Arg Phe Gln Asp Trp Trp Asn Lys Val Val Glu Val Gly Asn
100 105 110
Gln His Gly Leu Glu Ser Tyr Arg Ala Met Lys Pro Lys Val Glu Ala
115 120 125
Leu Ile Ser Ser Val Pro Asn Asn Pro Phe Asp Pro Thr Trp Asp Ser
130 135 140
Asp Ile Gln Lys Arg Phe Ala Glu Val Lys Glu Ile Trp Ser Lys Ile
145 150 155 160
Arg Leu Lys Asp Ser Asn Glu Arg Asn Phe Arg Trp Arg Tyr His Val
165 170 175
Pro Pro Leu Val Lys Pro Pro Lys Thr Pro Glu Phe Lys Leu Leu Gly
180 185 190
Phe Asp Ala Gly Ala Arg Val Ser Gly Gly Gly Lys Ala Lys Thr Leu
195 200 205
Thr Ile Lys Ser Gln Asp Gly Glu Ile Phe Leu Arg Leu Leu Leu Arg
210 215 220
Gln Pro Gly Val Ser Asn His Gly Arg Gly Ser Tyr Arg Glu Ile Trp
225 230 235 240
Leu His Ile Ile Leu Glu Asp Glu Ser Arg Lys Ala Leu Leu Leu Ser
245 250 255
Asp Lys Val Gln Ala Lys Phe Ile Arg Leu Arg Arg Arg Thr Asp Lys
260 265 270
Trp Glu Ala His Ile Thr Val Glu Lys Glu Val Ser Phe Glu Lys Pro
275 280 285
Arg Val Ile Tyr Gly Ile Asp Ile Gly Met Lys Asp Ala Leu Leu Phe
290 295 300
Cys Val Pro Ile Phe Ile Asp Ser Gln Glu Thr Thr Glu Thr Ala Ser
305 310 315 320
Glu Gly Lys Lys Leu Ile Val Val Glu Ala Gly Gln Phe Trp His Arg
325 330 335
Leu Glu Val Glu Asp Ala Gln Trp Arg Lys Trp Gln Phe Arg Arg Ser
340 345 350

Gly Lys Val Lys Glu Gly Ile Lys Gly Lys Arg Met Val Trp Arg Ile
355 360 365
Leu Lys Arg Arg Arg Arg Arg Arg Arg Glu Val Thr Lys His Leu Ile
370 375 380
His Gln Ile Val Asn Ala Leu Ile Ser Gln Leu Gln Arg Pro Cys Ala
385 390 395 400
Val Ala Ile Glu Arg Leu Lys Gly Leu Arg Asn Lys Pro His Leu Thr
405 410 415
Asp Lys Arg Ser Gln Trp Lys Phe His Arg Trp Ala Tyr Arg Phe Leu
420 425 430
Leu Glu Thr Leu Arg Tyr Lys Leu Glu Trp Glu Gly Ile Met Ile Ile
435 440 445
Glu Val Pro Pro Tyr Gly Thr Ser Gln Thr Cys Pro Arg Cys Ser His
450 455 460
Arg Asp Lys Arg Ser Arg Asp Gly Arg Arg Phe Arg Cys Thr Lys Cys
465 470 475 480
Lys Phe Gln His Asn Ala Asp Ala Ala Gly Ala Phe Asn Ile Ala Met
485 490 495
Arg Ala Val Glu Gln Leu Gln Ser Ser Arg Asp Leu Gln Glu Val Ser
500 505 510
Ile Ala Glu Ser Ala Trp Lys Ile Leu Thr Gly Lys Pro Met Val Leu
515 520 525

[0283]

Lys Tyr
530

<210> 621
<211> 512
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-温泉-热泉序列
<400> 621
Met Glu Lys Val Ile Leu Thr Ile Thr Ser Gly Val Glu Phe Leu Gln
1 5 10 15
Pro Gln Asp Glu Gln Leu Val Ile Lys Ala Ala Lys Trp Tyr Thr Glu
20 25 30
Ala Arg Lys Ala Leu Leu Tyr Asn Ala Leu Gln Lys Gly Glu Glu Gly
35 40 45
Val Lys Gln Ala Ala Ser Thr Leu Ser Thr Leu Asn Phe Pro Pro Ala
50 55 60
Pro His Ser Arg Leu Pro Asp Leu Leu Thr Arg Asp Val Gly Gly Thr
65 70 75 80
Leu Gln Ala Leu Lys Glu Arg Arg Glu Gly Tyr Ala Gln Asp Phe Ser
85 90 95

[0284]

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

Gln Thr Cys Pro Arg Cys Gly Glu Arg Asp Arg Asn Ala Arg Gln Gly
450 455 460
Asn Leu Phe Arg Cys Thr Lys Cys Arg Phe Gln His Asn Ala Asp Phe
465 470 475 480
Val Ala Ala Tyr Asn Ile Thr Val Arg Gly Leu Met Lys Leu Arg Thr
485 490 495
Arg Gly Arg Ile Asp Trp Lys Ile Leu Thr Gly Lys Pro Arg Val Leu
500 505 510

<210> 622
<211> 490
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-温泉-热泉序列

[0285]

<400> 622
Met Leu Gln Tyr Ala Ile Glu Lys Arg Lys Glu Arg Ile Thr Ala Pro
1 5 10 15
Ala Arg Ile Leu Lys Asp Glu Leu Lys Leu Pro Glu Pro Pro His Ser
20 25 30
Arg Leu Pro Asp Glu Leu Ala Gln Val Ile Asn Gly Val Leu Lys Ala
35 40 45
Leu Lys Lys Arg Arg Thr Glu Tyr Leu Arg Arg Phe Gln Asp Trp Trp
50 55 60
Asn Lys Val Val Glu Val Gly Asn Gln His Gly Leu Glu Ser Tyr Arg
65 70 75 80
Ala Met Lys Pro Lys Val Glu Ala Leu Ile Ser Ser Val Pro Asn Asn
85 90 95
Pro Phe Asp Pro Thr Trp Asp Ser Asp Ile Gln Lys Arg Phe Ala Glu
100 105 110
Val Lys Glu Ile Trp Ser Lys Ile Arg Leu Lys Asp Ser Asn Glu Arg
115 120 125
Asn Phe Arg Trp Arg Tyr His Val Pro Pro Leu Val Lys Pro Pro Lys
130 135 140
Thr Pro Glu Phe Lys Leu Leu Gly Phe Asp Ala Gly Ala Arg Val Ser
145 150 155 160
Arg Gly Gly Lys Ala Lys Thr Leu Thr Ile Lys Ser Gln Asp Gly Glu
165 170 175
Ile Phe Leu Arg Leu Leu Leu Arg Gln Pro Gly Val Ser Asn His Gly
180 185 190
Arg Gly Ser Tyr Arg Glu Ile Trp Leu His Ile Ile Leu Glu Asp Glu
195 200 205
Ser Arg Lys Ala Leu Leu Leu Ser Asp Lys Val Gln Ala Lys Phe Ile
210 215 220

Arg Leu Arg Arg Arg Thr Asp Lys Trp Glu Ala His Ile Thr Val Glu
225 230 235 240
Lys Glu Val Ser Phe Glu Lys Pro Arg Val Ile Tyr Gly Ile Asp Ile
245 250 255
Gly Met Lys Asp Ala Leu Leu Phe Cys Val Pro Ile Phe Ile Asp Ser
260 265 270
Gln Glu Thr Thr Glu Thr Ala Ser Glu Gly Lys Lys Leu Ile Val Val
275 280 285
Glu Ala Gly Gln Phe Trp His Arg Leu Glu Val Glu Asp Ala Gln Trp
290 295 300
Arg Lys Trp Gln Phe Arg Arg Ser Gly Lys Val Lys Glu Gly Ile Lys
305 310 315 320
Gly Lys Arg Met Val Trp Arg Ile Leu Lys Arg Arg Arg Arg Arg Arg
325 330 335
Arg Glu Val Thr Lys His Leu Ile His Gln Ile Val Asn Ala Leu Ile
340 345 350
Ser Gln Leu Gln Arg Pro Cys Ala Val Ala Ile Glu Arg Leu Lys Gly
355 360 365
Leu Arg Asn Lys Pro His Leu Thr Asp Lys Arg Ser Gln Trp Lys Phe
370 375 380

[0286]

His Arg Trp Ala Tyr Arg Phe Leu Leu Glu Thr Leu Arg Tyr Lys Leu
385 390 395 400
Glu Trp Glu Gly Ile Met Ile Ile Glu Val Pro Pro Tyr Gly Thr Ser
405 410 415
Gln Thr Cys Pro Arg Cys Ser His Arg Asp Lys Arg Ser Arg Asp Gly
420 425 430
Arg Arg Phe Arg Cys Thr Lys Cys Lys Phe Gln His Asn Ala Asp Ala
435 440 445
Ala Gly Ala Phe Asn Ile Ala Met Arg Ala Val Glu Gln Leu Gln Ser
450 455 460
Ser Arg Asp Leu Gln Glu Val Ser Ile Ala Glu Ser Ala Trp Lys Ile
465 470 475 480
Leu Thr Gly Lys Pro Met Val Leu Lys Tyr
485 490

<210> 623
<211> 521
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-温泉-热泉和微生物垫飘带
序列
<400> 623

[0287]

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

Arg Glu Ile Lys Phe Gly Leu Lys Gly Ile Met Trp Thr Ala Gly Asp
        275                 280                 285
Tyr Leu Tyr Gly Val Asp Ile Gly Ala Arg Lys Tyr Leu Leu Ala Ala
    290                 295                 300
Val Ala Val Pro Leu Lys Pro Gly Leu Pro Lys Leu Pro Asn Ile Ala
305                 310                 315                 320
Ile Pro Ser Gly Lys Phe Trp Gln Ala Val Glu Arg Val Glu Ala Glu
                325                 330                 335
Trp Arg Lys Trp Gln Arg Leu Arg Thr Gly Lys Lys Asp Lys Glu Gly
```

340 345 350
Lys Val Val Leu Lys Lys Arg Thr Ala Trp Arg Val Leu Lys Arg Lys
355 360 365
Arg His Met Arg Arg Glu Val Ala Lys Asn Met Ile Ala Glu Ala Val
370 375 380
Asn Thr Phe Ile Asn Phe Ile Lys Gly Asp Asn Val Asp Ile Arg Leu
385 390 395 400
Trp Leu Ala Ile Glu Gln Ile Lys Asn Leu Asn Leu Ser Pro Ala Phe
405 410 415
Gln Asp Arg Lys Ser Arg Trp Arg Gln Arg Arg Trp Ala Tyr Arg Leu
420 425 430
Leu Leu Glu Met Leu Arg Tyr Lys Ala Glu Trp Glu Gly Ile Arg Val
435 440 445
Ile Glu Val Asn Pro Ala His Thr Ser Gln Ile Cys Pro Arg Cys Gly
450 455 460
Lys Lys Asp Gly Lys Ala Arg Gln Gly Asn Arg Phe Cys Cys Thr Glu
465 470 475 480
Cys Lys Phe Gln His Asn Ala Asp Phe Val Ala Ala Tyr Asn Ile Ala
485 490 495
Val Arg Ala Leu Lys Lys Leu Gly Met Arg Glu Glu Ile Gly Trp Lys
500 505 510
Val Leu Thr Gly Lys Pro Arg Val Leu
[0288] 515 520

<210> 624
<211> 516
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
水生-温泉-沉积物序列
<400> 624
Met Lys Lys Val Val Ile Thr Ile Thr Thr Gly Val Glu Phe Leu Gln
1 5 10 15
Pro Gln Asp Glu Gln Ile Val Arg Lys Ala Ala Glu Trp Tyr Thr Lys
20 25 30
Ala Arg Glu Ala Leu Phe Ala His Ala Leu Gln Gln Gln Gln Gly Lys
35 40 45
Val Glu Lys Ala Ala Ser Thr Leu Arg Thr Leu Asn Phe Pro Met Ala
50 55 60
Pro His Ser Arg Leu Pro Asp Leu Leu Thr Arg Asp Val Gly Gly Ile
65 70 75 80
Leu Lys Ala Leu Lys Glu Lys Arg Glu Gly Tyr Leu Gly Glu Phe Lys
85 90 95
Arg Trp Trp Glu Glu Val Cys Arg Val Ala Asn Glu Lys Gly Ser Gln

[0289]

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

450 455 460

Ala Arg Gln Gly Asn Arg Phe Cys Cys Thr Lys Cys Lys Phe Gln His
465 470 475 480
Asn Ala Asp Phe Val Ala Ala Tyr Asn Ile Ala Val Arg Ala Leu Lys
485 490 495
Lys Leu Gly Met Arg Glu Glu Ile Gly Trp Lys Val Leu Thr Gly Lys
500 505 510
Pro Lys Val Leu
515

<210> 625
<211> 516
<212> PRT
<213> 未知

<220>
<223> 未知的描述:
水生-温泉-沉积物或沉积物宏基因组
序列

<400> 625

[0290]

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

[0291]

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

<210> 626
<211> 417
<212> PRT

<213> 未知
<220>
<223> 未知的描述:
发酵宏基因组序列
<400> 626

[0292]

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

305 310 315 320
Leu Lys Phe Trp Pro Arg Phe His Leu Val Thr Arg Ile Glu Gln Lys
325 330 335
Cys Gly Glu Val Gly Leu Thr Phe Glu Lys Ile Lys Pro Ala Gly Thr
340 345 350
Ser Gln Thr Cys Ser Ser Cys Gly His Arg Asp Lys Asp Asn Arg Asn
355 360 365
Gly Pro Asp Phe Val Cys Leu Arg Cys Gly Phe Ala Arg His Ala Asp
370 375 380
Val Asn Ala Ala Asn Asn Ile Ala Arg Gly Met Asp Asp Phe Asp Leu
385 390 395 400
Pro Thr Gly Ala Ser Ala Pro Thr Lys Ala Arg Arg Leu Glu Ile Ala
405 410 415
Ser

<210> 627
<211> 512
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
Fervidibacteria 细菌 JGI MDM2 序列

[0293]
<400> 627
Met Glu Lys Val Ile Leu Thr Ile Thr Ser Gly Val Glu Phe Leu Gln
1 5 10 15
Pro Gln Asp Glu Gln Leu Val Ile Lys Ala Ala Lys Trp Tyr Thr Glu
20 25 30
Ala Arg Lys Ala Leu Leu Tyr Asn Ala Leu Gln Lys Gly Glu Glu Gly
35 40 45
Val Lys Gln Ala Ala Ser Thr Leu Ser Thr Leu Asn Phe Pro Pro Ala
50 55 60
Pro His Ser Arg Leu Pro Asp Leu Leu Thr Arg Asp Val Gly Gly Thr
65 70 75 80
Leu Gln Ala Leu Lys Glu Arg Arg Glu Gly Tyr Ala Gln Asp Phe Ser
85 90 95
Arg Trp Trp Glu Glu Ile Cys Arg Val Ala Asn Glu Lys Gly Ser Gln
100 105 110
Ala Val Ser Val Leu Lys Pro His Val Asn Lys Leu Ser Ser Leu Ile
115 120 125
His Gln Leu Ser Trp Asn Ser Pro Glu Trp Asp Glu Glu Ala Gln Glu
130 135 140
Val Phe Lys Glu Leu Lys Gln Ile Trp Gly Val Val Lys Gln Lys Asp
145 150 155 160
His Glu Arg Leu Lys Ile Lys Trp Cys Ala Glu Val Pro Pro Val Pro

[0294]

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

Gln Thr Cys Pro Arg Cys Gly Glu Arg Asp Arg Asn Ala Arg Gln Gly
    450                 455                 460
Asn Leu Phe Arg Cys Thr Lys Cys Arg Phe Gln His Asn Ala Asp Phe
465                 470                 475                 480
Val Ala Ala Tyr Asn Ile Thr Val Arg Gly Leu Met Lys Leu Arg Thr
                485                 490                 495
Arg Gly Arg Ile Asp Trp Lys Ile Leu Thr Gly Lys Pro Arg Val Leu
            500                 505                 510
```

<210> 628
<211> 514
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
淡水宏基因组序列
<400> 628

[0295]

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

Gln Leu Arg Gln Asp Gly Lys Arg Trp Ile Ala His Ile Val Ile Glu
            260                 265                 270
Arg Glu Ile Lys Phe Gly Glu Pro Asp Tyr Leu Tyr Gly Val Asp Ile
```

[0296]

```
        275                 280                 285
Gly Ala Arg Lys Tyr Leu Leu Ala Ala Val Ala Val Pro Leu Lys Pro
    290                 295                 300
Asp Leu Pro Lys Leu Pro Asn Ile Ala Ile Pro Ser Gly Glu Phe Trp
305                 310                 315                 320
Gln Ala Val Glu Arg Val Glu Ala Glu Trp Arg Lys Trp Gln Arg Leu
                325                 330                 335
Arg Thr Gly Lys Lys Asp Lys Glu Gly Arg Val Val Leu Lys Lys Lys
            340                 345                 350
Arg Met Ala Trp Arg Val Leu Lys Arg Lys Arg Arg Met Arg Arg Glu
        355                 360                 365
Val Ala Lys Asn Met Ile Ala Glu Ala Val Asn Thr Phe Ile Asn Phe
    370                 375                 380
Ile Lys Arg Gly Asn Val Asn Ser Gly Leu Ala Ile Glu Gln Ile Lys
385                 390                 395                 400
Asn Leu Asn Leu Ser Pro Thr Leu Gln Asp Arg Lys Ser Arg Trp Arg
                405                 410                 415
Gln Arg Arg Trp Ala Tyr Arg Leu Leu Leu Glu Met Leu Arg Tyr Lys
            420                 425                 430
Ala Glu Trp Glu Gly Ile Arg Val Ile Glu Val Asn Pro Thr His Thr
        435                 440                 445
Ser Gln Ile Cys Pro Arg Cys Arg Lys Lys Asp Gly Lys Ala Arg Gln
    450                 455                 460
Gly Asn Arg Phe Cys Cys Thr Lys Cys His Phe Gln His Asn Ala Asp
465                 470                 475                 480
Phe Val Ala Ala Tyr Asn Ile Ala Met Arg Gly Leu Arg Lys Leu Gly
                485                 490                 495
Met Arg Gly Glu Ile Gly Trp Lys Val Leu Thr Gly Lys Pro Arg Val
            500                 505                 510
Leu Gln
```

<210> 629
<211> 454
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
淡水宏基因组序列
<400> 629

```
Met Ala Pro His Ser Arg Leu Pro Asp Leu Leu Thr Arg Asp Val Gly
1               5                   10                  15
Gly Ile Leu Lys Ala Leu Lys Glu Lys Arg Glu Gly Tyr Leu Arg Glu
            20                  25                  30
Phe Lys Arg Trp Trp Glu Glu Val Cys Arg Val Ala Asn Glu Lys Gly
```

[0297]

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

Trp Gln Arg Leu Arg Thr Gly Lys Lys Asp Lys Glu Gly Arg Val Val
          275                 280                 285
Leu Lys Lys Lys Arg Met Ala Trp Arg Val Leu Lys Arg Lys Arg Arg
      290                 295                 300
Met Arg Arg Glu Val Ala Lys Asn Met Ile Ala Glu Ala Val Asn Thr
305                 310                 315                 320
Phe Ile Asn Phe Ile Lys Arg Gly Asn Val Asn Ser Gly Leu Ala Ile
                325                 330                 335
Glu Gln Ile Lys Asn Leu Asn Leu Ser Pro Thr Leu Gln Asp Arg Lys
            340                 345                 350
Ser Arg Trp Arg Gln Arg Arg Trp Ala Tyr Arg Leu Leu Leu Glu Met
          355                 360                 365
Leu Arg Tyr Lys Ala Glu Trp Glu Gly Ile Arg Val Ile Glu Val Asn
      370                 375                 380
```

Pro Thr His Thr Ser Gln Ile Cys Pro Arg Cys Arg Lys Lys Asp Gly
385 390 395 400
Lys Ala Arg Gln Gly Asn Arg Phe Cys Cys Thr Lys Cys His Phe Gln
405 410 415
His Asn Ala Asp Phe Val Ala Ala Tyr Asn Ile Ala Met Arg Gly Leu
420 425 430
Arg Lys Leu Gly Met Arg Gly Glu Ile Gly Trp Lys Val Leu Thr Gly
435 440 445
Lys Pro Arg Val Leu Gln
450

<210> 630
<211> 332
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
真菌-菌丝-切叶（attine）蚂蚁真菌园序列

[0298]

<400> 630
Met Lys Ile Thr Val Lys Ala Lys Val Thr Asp Asp Ser Lys Asn Lys
1 5 10 15
Leu Gly Lys Ile Glu Ser Glu Phe Gln Asn Tyr Gln Val Tyr Leu Gln
20 25 30
Ser Gly Cys Thr Thr Arg Val Asn Leu Tyr Ser Ala Thr Lys Gln Gln
35 40 45
Ala Gln Arg Ala Leu Lys Lys Ile Lys Gln Pro Asn Pro Lys Lys Gln
50 55 60
Tyr Pro Val Ile Leu Arg Arg Asp Leu Val Asp Ile Gln Gln Asp Ile
65 70 75 80
Lys Phe Glu Gly Val Tyr Trp Ile Lys Ile Pro Val Tyr Pro Asn Ser
85 90 95
Ile Asn Ile Arg Ile Lys Thr Ser Cys Lys Tyr Asp Leu Thr Gln His
100 105 110
Ser Ile Arg Glu Ser Lys Ile Leu Lys Ile Gln Asn Val Trp Tyr Ile
115 120 125
Phe Ile Thr Ile Glu Lys Gln Asp Glu Arg Ile Ile Arg Ala Asp Asn
130 135 140
Val Leu Ala Val Asp Leu Gly Cys Lys Asn Ile Ala Val Thr Val Asn
145 150 155 160
Thr Ile Asn Thr Lys Pro Asn Phe Tyr Gly Ala Ala Ile Arg Gly Ile
165 170 175
Arg Gly Phe Tyr Phe Asn Leu Arg Arg Lys Leu Gly Glu Lys Lys Ala
180 185 190
Ser Asp Lys Ile Lys Leu Leu Lys Asn Ile Glu Phe Leu Gln Val Asn
195 200 205

His Glu Leu His Lys Ile Ser Lys Ala Ile Val Glu Glu Ala Lys Arg
210 215 220
Thr Asn Ala Val Ile Val Leu Gly Lys Leu Lys Gly Ile Arg Lys Asn
225 230 235 240
Ile Lys Gly Ser Arg Arg Ile Arg Arg Leu Ile Asn Asn Phe Pro Tyr
245 250 255
Tyr Arg Leu Val Gln Tyr Ile Lys Tyr Lys Ala Ala Trp His Ser Ile
260 265 270
Lys Thr Leu Glu Ile Ser Glu Ala Asp Thr Ser Asn Thr Cys Phe Asn
275 280 285
Cys Arg Thr Lys Asp Lys Lys Ala Arg Lys Thr Gln Gly Leu Phe Gln
290 295 300
Cys Gly Asn Cys Ser Ile Lys Thr Asn Ala Asp Tyr Asn Gly Ala Met
305 310 315 320
Asn Ile Leu Gln Arg Gly Val Gly Tyr Phe Gln Leu
325 330

<210> 631
<211> 384
<212> PRT
<213> 未知
<220>

[0299]

<223> 未知的描述:
方形盐盒菌（Haloarcula quadrata）序列
<400> 631
Met Lys Arg Asn Val Ser Thr Thr Val Arg Ala Lys Leu His Ser Leu
1 5 10 15
Thr Asn Arg Lys Thr Asp Leu Leu Ala Arg Glu Tyr Glu Ala Phe Gln
20 25 30
Thr Glu Val His Gly Gly Asp Ala Asp Leu Tyr Ser Ala Thr Asp Gln
35 40 45
Gln Ala Ser Lys Val Gln Arg Gln Lys Tyr Pro Asn Pro Gly Thr Glu
50 55 60
Gln Pro Val Val Leu Arg Asn Asp Val Phe Asp Val Ala Tyr Asp Glu
65 70 75 80
Asp Thr Val Leu Ser Ser Trp Trp Val Lys Val Pro Val Tyr Asp Pro
85 90 95
Glu Lys Gly Gln Gly Asn Ser Ile Trp Cys Pro Ala His Val Pro His
100 105 110
Lys Asp Glu Gln Leu Val Arg Glu Gly Asn Leu Arg Asp Ser Glu Leu
115 120 125
Val Arg Ser Asp Gly Asp Trp Tyr Val His Leu Val Val Lys Arg Ser
130 135 140
Val Thr Val Gln Asp Glu Tyr Asp Asp Val Leu Ala Ile Asp Met Gly
145 150 155 160

Ala Arg Trp Val Ala Thr Cys Ala Phe Leu Ser Asp Arg Lys Thr Thr
165 170 175
Phe Tyr Gly Glu Glu Val Arg Arg Ile Arg Glu His Tyr Lys Gln Leu
180 185 190
Arg Lys Ser Ile Gly Lys Ala Lys Pro Arg Gln Gly Gln Gln Val Val
195 200 205
Glu Arg Ile Gly Asp Ala Glu Ala Arg Lys Val Asp Asp Arg Leu His
210 215 220
Lys Ile Ala Arg Ser Ile Val Glu Asp Ala Glu Glu Arg Asn Ala Val
225 230 235 240
Ile Val Val Gly Asp Leu Gly Gly Ile Arg Lys Asp Asn Asp Lys Gly
245 250 255
Arg Tyr Val Asn Asp Lys Thr His Lys Met Pro Phe Ala Arg Leu Leu
260 265 270
Asn Tyr Ile Glu Tyr Lys Ala His Gly Ala Gly Ile Asp Val Gln Leu
275 280 285
Val Glu Glu Tyr Asp Thr Ser Lys Thr Cys Asn Arg Cys Ala Cys Glu
290 295 300
Gly Val Arg Thr Thr Gln Gly Arg Phe Glu Cys Pro Glu Cys Gly Leu
305 310 315 320
Asp Asp Asn Ala Asp Lys Asn Gly Ala Leu Asn Ile Gly Lys Arg Ala
325 330 335

[0300]

Leu Gly Lys Phe Leu Lys Pro Leu Ser Glu Ala Gly Ala Val Leu Ala
340 345 350
Gln Ser Glu Thr Gln Val Ile Val Gln Arg Asp Asp Glu Pro Ala Asn
355 360 365
Leu Ser Val Ser Val Asp Ser Thr Pro Ser Gly Gly Thr Pro Arg Leu
370 375 380

<210> 632
<211> 384
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
盐红菌属物种（Halorubrum sp.）阿特里特（Atlit）-26R 序列
<400> 632
Met Gln Arg Asn Val Ser Thr Thr Val Arg Val Lys Leu His Ser Leu
1 5 10 15
Thr Asn Arg Lys Ala Asp Leu Leu Ala Arg Glu Tyr Glu Ala Phe Gln
20 25 30
Thr Ala Val His Gly Gly Asp Ala Asp Leu Tyr Ser Ala Thr Asp Gln
35 40 45
Gln Ala Ser Lys Val Gln Arg Gln Lys Asp Pro Asn Pro Asp Thr Glu
50 55 60

[0301]

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

<210> 633
<211> 384
<212> PRT
<213> 未知

<220>

<223> 未知的描述:

特班齐赫盐红菌（Halorubrum tebenquichense）DSM 14210 序列

<400> 633

[0302]

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

Asp Asp Asn Ala Asp Lys Asn Gly Ala Leu Asn Ile Gly Lys Arg Ala
325 330 335
Ser Gly Lys Phe Ser Lys Pro Leu Ser Glu Ala Gly Ala Val Leu Ala
340 345 350
Gln Pro Lys Thr Gln Val Ile Val Glu Arg Asp Glu Glu Pro Ala Asn
355 360 365
Leu Ser Ile Ser Val Gly Ser Thr Pro Ser Gly Gly Thr Pro Arg Leu
370 375 380

<210> 634
<211> 535
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
热泉宏基因组序列

<400> 634

[0303]

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

[0304]

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

<210> 635
<211> 537
<212> PRT
<213> 未知

<220>

<223> 未知的描述:
热泉宏基因组序列

<400> 635

[0305]

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

[0306]

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

<210> 636
<211> 546
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
热泉宏基因组序列
<400> 636

Met Lys His Leu Val Ser Asn Ala Gln Val Asn Glu Lys Asn Glu Ala
1 5 10 15
Arg Leu Thr Leu Lys Cys Gly Leu Glu Phe Glu Asn Ala Glu Gly Glu
20 25 30
Glu Leu Val Lys Arg Ala Ala Glu Trp Tyr Cys Lys Ala Arg Ile Phe
35 40 45
Leu Leu Ser Tyr Ala Leu Lys Asn Asn Lys Thr Arg Val Gln Lys Ile
50 55 60

Ala Ser Leu Leu Lys Ser Glu Gly Phe Pro Ser Pro Pro His Pro Arg
65 70 75 80
Leu Ile Asp Leu Leu Gln Arg Thr Ile Thr Pro Ala Ile Lys Asn Met
85 90 95
Arg Arg Leu Arg Arg Arg Tyr Pro Lys Thr Leu Arg Ala Trp Phe Asp
100 105 110
Arg Val Lys Lys Phe Ala Glu Glu Lys Gly Ser Asp Ala Ala Glu Arg
115 120 125
Leu Gln Pro Tyr Val Glu Asp Leu Leu Gln Leu Val Lys Asn Leu Pro
130 135 140
Ile Asn Ser Pro Ala Trp Asp Thr Asp Ala Gln Glu Lys Trp Lys Glu
145 150 155 160
Leu Lys Lys Leu Trp Gly Lys Val Arg Glu Ser Asp Trp Lys Lys Gly
165 170 175
Ile Val Trp Lys Phe Gln Ile Pro Asn Pro Lys Lys Pro Pro Ser Thr
180 185 190
Pro Ala Phe Arg Cys Ala Asp Ser Ser Asn Ile Phe Val Thr Ala Asp
195 200 205
Tyr Arg Met Ser Gly Lys Asn Ala Arg Asn Ser Tyr Asp Met Lys Thr
210 215 220
Asp Ser Glu Glu Asn Leu Trp Val Leu Val Arg Leu Asp Asp Val Asp
225 230 235 240

[0307]

Gly Asp Val Lys Val Arg Lys Asp Val Trp Ile Lys Leu Cys Val Asp
245 250 255
Lys Ser Arg Tyr Asp Leu Leu Arg Ile Gly Lys Ile Thr Glu Ile Lys
260 265 270
Ile Leu Arg Asn Arg Asp Gly Arg Trp Glu Gly Arg Ile Val Val Lys
275 280 285
Met Asn Ala Gly Tyr Asn Ser Pro Gln Val Val Val Gly Val Asp Val
290 295 300
Gly Ala Ser Glu His Leu Ile Ala Ala Val Gly Ile Pro Leu Asp Gly
305 310 315 320
Ser Asn Asn Asn Phe Leu Leu Ser Val Pro Ala Ser Tyr Phe Trp Lys
325 330 335
Arg Val Glu Gln Ala Glu Arg Leu Trp Arg Arg Trp Gln Arg Leu Tyr
340 345 350
Thr Gly Arg Thr Ser Glu Lys Lys Lys Ala Pro Ile Leu Pro Arg Arg
355 360 365
Arg Lys Ala Tyr Gln Ile Leu Lys Arg Asn Arg Asn Trp Arg Arg Glu
370 375 380
Val Thr Lys His Cys Ile Phe His Ala Val Lys Lys Phe Val Gln Leu
385 390 395 400
Leu Pro Pro Asn Ser Val Val Val Val Glu Asn Leu Lys Asp Ile Pro
405 410 415

Pro Asn Val Trp Lys Glu Asp Lys Asp Lys Arg Phe Arg Arg Trp Val
420 425 430
Tyr Ala Phe Leu Leu Phe Thr Leu Lys Tyr Lys Cys Gln Gln Arg Lys
435 440 445
Ile Ala Val Ile Ala Val Ser Pro His Asn Thr Ser Gln Thr Cys Pro
450 455 460
Arg Cys Gly Val Val Asp Pro Lys Ser Arg Gly Leu Trp Lys Asp Gly
465 470 475 480
Glu Leu Ile Ser Arg Lys His Phe Cys Cys Thr Gln Cys Glu Phe Thr
485 490 495
His Asn Ala Asp Phe Val Ala Ala His Asn Ile Ala Val Leu Gly Ala
500 505 510
Lys Ala Leu Gly Ile Thr Leu Pro Glu Ser Trp Arg Ala Leu Arg Gly
515 520 525
Glu Val Gln Tyr Val Ile Pro Tyr Gln Asn Met Pro Lys Val Arg Lys
530 535 540
Ser Trp
545

<210> 637
<211> 512
<212> PRT
<213> 未知

[0308]

<220>
<223> 未知的描述:
热泉宏基因组序列
<400> 637
Met Glu Lys Val Ile Leu Thr Ile Thr Ser Gly Val Glu Phe Leu Gln
1 5 10 15
Pro Gln Asp Glu Gln Leu Val Ile Lys Ala Ala Lys Trp Tyr Thr Glu
20 25 30
Ala Arg Lys Ala Leu Leu Tyr Asn Ala Leu His Lys Gly Glu Glu Gly
35 40 45
Val Lys Gln Ala Ala Ser Thr Leu Ser Thr Leu Asn Phe Pro Pro Ala
50 55 60
Pro His Ser Arg Leu Pro Asp Leu Leu Thr Arg Asp Val Gly Gly Thr
65 70 75 80
Leu Gln Ala Leu Lys Glu Arg Arg Glu Gly Tyr Ala Gln Asp Phe Ser
85 90 95
Arg Trp Trp Glu Glu Ile Cys Arg Val Ala Asn Glu Lys Gly Ser Gln
100 105 110
Ala Val Ser Val Leu Lys Pro His Val Asn Lys Leu Ser Ser Leu Ile
115 120 125
His Gln Leu Ser Trp Asn Ser Pro Glu Trp Asp Glu Glu Ala Gln Glu
130 135 140

[0309]

```
Val Phe Lys Glu Leu Lys Gln Ile Trp Ala Val Val Arg Gln Lys Asp
145                 150                 155                 160
His Glu Arg Leu Lys Ile Lys Trp Arg Ala Glu Val Pro Pro Val Pro
                165                 170                 175
Arg Pro Pro Arg Thr Pro Lys Phe Gln Arg Leu Val Phe Asn Ala Asp
            180                 185                 190
Gly Arg Val Ser Gly Gly Ser Arg Ser Asn Ala Leu Gln Ile Glu Glu
        195                 200                 205
Arg Glu Asp Gly Asp Leu Tyr Leu Arg Leu Gln Leu Arg Asp Pro Glu
    210                 215                 220
Arg Val Ala Lys Lys Gly Ser Tyr Gln Asp Val Trp Leu Arg Leu Val
225                 230                 235                 240
Ile Glu Asp Glu Glu Arg Lys Gln Met Leu Lys Asn Phe Ala Pro Lys
                245                 250                 255
Tyr Leu Gln Leu Arg Gln Asp Gly Lys Arg Trp Ile Ala His Ile Val
            260                 265                 270
Ile Glu Arg Glu Asn Glu Phe Glu Val Pro Gln Tyr Val Tyr Gly Val
        275                 280                 285
Asp Ile Gly Ala Arg Lys Tyr Leu Leu Ala Ala Val Ala Ile Pro Leu
    290                 295                 300
Asn Pro Ala Leu Pro Lys Pro Pro Asn Ile Ala Val Pro Ser Ser Lys
305                 310                 315                 320
Phe Trp Gln Ile Val Glu Arg Val Glu Ala Glu Trp Arg Lys Trp Gln
                325                 330                 335
Gln Leu Arg Thr Gly Lys Arg Asp Lys Glu Gly Arg Val Val Leu Lys
            340                 345                 350
Lys Arg Lys Met Ala Trp Arg Val Leu Lys Arg Lys Arg Arg Met Arg
        355                 360                 365
Arg Glu Val Thr Lys Ser Met Ile Ala Glu Ala Val Asn Thr Phe Ile
    370                 375                 380
Lys Ser Leu Glu Arg Pro Cys Ala Ile Val Ile Glu Gln Ile Lys Asn
385                 390                 395                 400
Leu His Leu Ser Pro Thr Leu Gln Asp Arg Lys Ser Arg Trp Arg Gln
                405                 410                 415
Arg Arg Trp Ala Tyr Arg Phe Leu Leu Glu Met Leu Arg Tyr Lys Ala
            420                 425                 430
Glu Trp Glu Gly Ile Arg Val Ile Glu Val Asn Pro Ala Tyr Thr Ser
        435                 440                 445
Gln Thr Cys Pro Arg Cys Gly Lys Arg Asp Arg Asn Ala Arg Gln Gly
    450                 455                 460
Asn Leu Phe Gln Cys Thr Lys Cys Arg Phe Gln His Asn Ala Asp Phe
465                 470                 475                 480
Val Ala Ala Tyr Asn Ile Thr Val Arg Gly Leu Met Lys Leu Arg Thr
                485                 490                 495
```

Arg Gly Arg Ile Asp Trp Lys Ile Leu Thr Gly Lys Pro Arg Val Leu
500 505 510

<210> 638
<211> 532
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
热泉宏基因组序列
<400> 638

[0310]

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

[0311]

Arg Glu Ile Lys Phe Gly Leu Lys Gly Ile Met Trp Thr Ala Gly Asp
275 280 285
Tyr Leu Tyr Gly Val Asp Ile Gly Ala Arg Lys Tyr Leu Leu Ala Ala
290 295 300
Val Ala Val Pro Leu Lys Pro Gly Leu Pro Lys Leu Leu Glu Ser Gln
305 310 315 320
Gly Asp Asn Val Gly Ile Arg His Ile Ala Ile Pro Ser Gly Lys Phe
325 330 335
Trp Gln Ala Val Glu Arg Val Glu Ala Glu Trp Arg Lys Trp Gln Arg
340 345 350
Leu Arg Thr Gly Lys Lys Asp Lys Glu Gly Arg Val Val Leu Lys Lys
355 360 365
Lys Arg Met Ala Trp Arg Val Leu Lys Arg Lys Arg Arg Met Arg Arg
370 375 380
Glu Val Ala Lys Asn Met Leu Ala Lys Ala Val Asn Thr Phe Ile Leu
385 390 395 400
Asp Phe Ile Lys Arg Asn Asn Val Asp Ile Trp Leu Ala Ile Glu Gln
405 410 415
Ile Lys Asn Leu Asn Leu Ser Pro Thr Leu Gln Asp Arg Lys Ser Arg
420 425 430
Trp Arg Gln Arg Arg Trp Ala Tyr Arg Leu Leu Leu Glu Met Leu Arg
435 440 445
Tyr Lys Ala Glu Trp Glu Gly Ile Arg Val Ile Glu Val Asn Pro Ala
450 455 460
His Thr Ser Gln Ile Cys Pro Arg Cys Gly Lys Lys Asp Gly Lys Ala
465 470 475 480
Arg Gln Gly Asn Arg Phe Cys Cys Thr Lys Cys His Phe Gln His Asn
485 490 495
Ala Asp Phe Val Ala Ala Asp Asn Ile Ala Met Arg Gly Leu Arg Lys
500 505 510
Val Gly Met Arg Gly Glu Ile Gly Trp Lys Val Leu Thr Gly Lys Pro
515 520 525
Arg Val Leu Gln
530

<210> 639
<211> 384
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
实验室富集限定的培养基-有氧富集培养基序列
<400> 639
Met Gln Gln Thr Val Ser Thr Thr Val Arg Val Lys Leu Pro Ser Leu
1 5 10 15

[0312]

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

Leu Ser Val Ser Val Asp Ser Thr Pro Ser Gly Gly Thr Pro Arg Leu
370 375 380

<210> 640
<211> 376
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
实验室富集限定的培养基-有氧富集培养基序列
<400> 640

[0313]

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

Asp Ala Gly Ile Asp Leu Gln Leu Val Asp Glu Tyr Asp Thr Ser Gln
275 280 285
Thr Cys Asn Arg Cys Ala Cys Glu Gly Val Arg Ala Thr Gln Gly Arg
290 295 300
Phe Glu Cys Pro Gly Cys Gly Leu Asp Asp Asn Ala Asp Lys Asn Gly
305 310 315 320
Ala Leu Asn Ile Gly Lys Arg Ala Leu Gly Lys Phe Ser Arg Pro Leu
325 330 335
Ser Glu Ala Gly Ala Val Leu Ala Gln Ser Glu Thr Gln Val Ile Val
340 345 350
His Arg Asp Glu Glu Pro Ala Asn Leu Ser Val Ser Val Asp Ser Thr
355 360 365
Pro Ser Gly Gly Thr Pro Arg Leu
370 375

<210> 641
<211> 429
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
填埋物宏基因组或固体废物-填埋物-填埋物沥滤液
序列

[0314]

<400> 641
Met Ala Leu Lys Thr Leu Thr Phe Trp Pro Leu Glu Ile Pro Pro Glu
1 5 10 15
His Lys Arg Ile Phe Asp Asp Ile Arg Lys Gln Leu Gly Asn Ala Gln
20 25 30
Arg Glu Met Val Lys Ala Trp Leu Lys Leu Ala Leu Ala Ala Glu Pro
35 40 45
Tyr Gly Ile Thr Pro Gly Glu Gln Ala Gly Arg Val Leu Phe Gly Glu
50 55 60
Thr Asp Tyr Ser Asn Gly Leu Leu Lys Ser Leu Ala Tyr Thr Asp Ala
65 70 75 80
His Asp Ser Phe Lys Glu Gln Trp Lys Gly Val Leu Lys Gly Asp Ala
85 90 95
Thr Ser Val Ser Phe Pro Asp Glu Ala Ser Gly Val Val Leu Leu Val
100 105 110
Gly Ser Gly Gln Ala Thr Val Glu Lys Ile Asp Gly Glu Trp Trp Val
115 120 125
Thr Arg Leu Leu Val Phe Gly Gly Arg Thr Ala Glu Asp Arg Ser Arg
130 135 140
Asn Asp Gln Phe Arg Trp Gln Leu Lys Gly Val Ser Arg Phe Gly Gly
145 150 155 160
His Tyr Glu Lys Ser Val Met Asp Arg Ala Glu Arg Ile Cys Gln Val

[0315]

```
                165                 170                 175
Arg Val Leu Pro Ala Arg Lys Asp Arg Gly Lys Tyr Val Val Gln Val
            180                 185                 190
Thr Val Glu Val Pro Asp Glu Gln Ile Glu Arg Thr Ala Ile Pro Val
        195                 200                 205
Cys Ala Gly Ile Asp Leu Gly Ile Asn Cys Pro Val Val Leu Ser Ile
    210                 215                 220
Pro Asp Lys Gly Phe Val Arg Phe Met Gly Arg Glu Arg Asp Glu Tyr
225                 230                 235                 240
Pro Arg Leu Trp Glu Lys Leu Arg Glu Tyr Glu Asp Arg Lys Arg Thr
                245                 250                 255
Leu Asn Arg Ala Gly Lys Arg Arg Ala Ala Arg Asp Leu Arg Pro Arg
            260                 265                 270
Leu Thr Asn Ile Arg Lys Tyr Ile Asn Glu Ala Ile Ser Lys Asp Val
        275                 280                 285
Ala Asp Thr Cys Arg Ala Met Arg Val Thr Thr Leu Arg Met Glu Asn
    290                 295                 300
Leu Lys Gly Leu Ser Phe His Lys Ser Gly Thr Gly Lys Leu Arg His
305                 310                 315                 320
Trp Pro Arg Tyr His Leu Gln Glu Arg Ile Glu Gln Lys Cys Gly Glu
                325                 330                 335
Val Gly Ile Glu Val Glu Arg Ile Lys Pro Ala Gly Thr Ser Gln Thr
            340                 345                 350
Cys Ser Met Cys Gly His Arg Asp Lys Gly Asn Arg Asn Gly Ser Asp
        355                 360                 365
Phe Ile Cys Leu Ala Cys Gly Tyr Arg Gln His Ala Asp Leu Asn Ala
    370                 375                 380
Ala Asn Asn Ile Ala Arg Gly Met Gln Glu Phe Gly Pro Cys Val Gln
385                 390                 395                 400
Ser Ser Ala Ser Asp Asn Thr Ile Glu Leu Arg Phe Asp Gly Ala Gly
                405                 410                 415
Gln Glu Pro Ala Leu Gly Arg Ala Met Arg Leu Phe Ala
            420                 425
<210> 642
<211> 468
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
      填埋物宏基因组或固体废物-填埋物-填埋物沥滤液
      序列
<400> 642
Met Asp Thr Ala Ala Asn Glu Thr Thr Pro Lys Ala Leu Cys Phe Trp
1               5                   10                  15
```

[0316]

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

Arg Ile Glu Gln Lys Ala Ala Ala Val Asn Ile Ala Val Glu Lys Ile
370 375 380
Asn Pro His Leu Thr Ser Gln Thr Cys Ser Ala Cys Gly His Val Asp
385 390 395 400
Lys Gly Asn Arg Glu Gly Ala Gln Phe Arg Cys Leu Ser Cys Gly Phe
405 410 415
Glu Cys His Ala Asp Leu Asn Ala Ala Asn Asn Ile Ala Gly Arg Asp
420 425 430
Pro Ala Thr Asp Phe Asp Thr Pro Asp Arg Ala Ala Val Arg Ser Asp
435 440 445
Gly Pro Lys Ser Arg Pro Gln Ala Gly Arg Ala Ser Arg Leu Val Ala
450 455 460
Ala Ser Thr Pro
465

<210> 643
<211> 440
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
填埋物宏基因组序列

[0317] <400> 643

Met Leu Gly Gln Val Lys Thr Val Lys Phe His Ile Thr Gly Val Thr
1 5 10 15
Arg Asp Gln Gly Glu Ala Leu Thr Gly Ser Leu Arg Ala Leu Gly Ala
20 25 30
Ala Ala Arg Thr Ala Val Lys Thr Ala Leu Met Glu Val Ala Asp Lys
35 40 45
Asp Leu His Gly Val Lys Asp Ala Tyr Trp Gln Gly Val Gln Val Arg
50 55 60
Ala Glu Gln Ser Ala Arg Glu Ala Phe Lys Glu Thr Gly Tyr Ser Pro
65 70 75 80
Ala Ile Leu Asn Glu Tyr Trp Asn Ala Phe Gly Phe Thr Gly Phe Lys
85 90 95

Ile Arg Trp Ser Gln Val Arg Lys Gly Glu Arg Thr Ser Ile Pro Phe
100 105 110
Pro Gly Asp Ala Pro Phe Leu Gln Phe Thr Asn Gly Ala Thr Ile Gln
115 120 125
Val Ala Glu Thr Glu Asp Gly Tyr Cys Ile Arg Arg Ile Arg Ile Tyr
130 135 140
Gly Leu Val Gly Asn Lys Pro Asp Thr Arg Asn Glu Val Glu Trp Ala
145 150 155 160
Leu Lys Pro Val Gly Arg Gly Arg His Ala Ala Trp Gly Ile Glu Gln

[0318]

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

<210> 644
<211> 417
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
宏基因组序列
<400> 644
Met Ile Lys Thr Ala Thr Phe Tyr Ile Ala Asp Met Pro Gln Glu Asp

[0319]

```
1               5                   10                  15
Lys Arg Leu Cys Glu Arg Leu Cys Ala Glu Leu Gly Ala Ala Gly Ala
            20                  25                  30
Thr Met Val Arg Glu Trp Tyr Arg Ala Cys Asp Glu Val Thr Gly Glu
        35                  40                  45
Arg Ile Asn Trp Pro Glu Val Gln Lys Arg Gly Val Gln Val Gly Arg
    50                  55                  60
Gln Ala Phe Ala Glu Arg Gly Tyr Arg Ala Glu Leu Val Asn Asp Leu
65                  70                  75                  80
Gln Arg Ala Glu Phe Phe Ala Arg Phe Lys Met Glu Phe Lys Lys Ile
                85                  90                  95
Gln Ala Gly Glu Arg Ala Ser Leu Pro Phe Pro Gly Ser Ala Pro Phe
            100                 105                 110
Ile Phe Leu Arg Met Asp Pro Gly Lys Glu Lys Ala Thr Leu Val Gln
        115                 120                 125
Glu Asp Gly Thr Trp Tyr Leu Arg Asp Pro His Phe Tyr Pro Gly Gly
    130                 135                 140
Gly Arg Asn Glu Asp Asn Lys Arg Glu Trp Arg Leu Lys Asp Val Gly
145                 150                 155                 160
Arg Asn Gly Trp Gly Tyr Ala Ile Leu Ala Glu Ala Glu Lys Ala Ala
                165                 170                 175
Cys Ile Arg Ile Met Pro Asp Arg Lys Lys Arg Gly Ser Trp Val Val
            180                 185                 190
Lys Ile Thr Val His Leu Pro Asp Ala Ala Leu Arg Lys Glu Val Glu
        195                 200                 205
Arg Gln Glu Val Trp Ala Gly Ile Asp Leu Gly Leu Asn Cys Pro Ala
    210                 215                 220
Val Leu Ser Ile Pro Asp His Pro Asp Gly Gly Phe Val Arg Phe Phe
225                 230                 235                 240
Gly Asn Asp Glu Tyr Ala Arg Leu Trp Gln Arg Leu Lys Glu Tyr Glu
                245                 250                 255
Asp Arg Lys Arg Arg Leu Asn Arg Ala Gly Lys Arg Arg Ala Ala Arg
            260                 265                 270
Asp Leu Arg Ala Lys Ile Thr Gly Val Arg Gln His Ile Asn Glu Leu
        275                 280                 285
Ile Ser Arg Glu Ile Ile Asp Leu Cys Ile Ala Met Arg Val Thr Gly
    290                 295                 300
Leu Arg Met Glu Asp Leu Lys Gly Leu Gly Ala Asp Gly Ser Gly Lys
305                 310                 315                 320
Leu Lys Phe Trp Pro Arg Phe His Leu Val Thr Arg Ile Glu Gln Lys
                325                 330                 335
Ala His Glu Ala Gly Leu Glu Phe Val Lys Ile Lys Pro Ala Gly Thr
            340                 345                 350
Ser Gln Thr Cys Ser Ala Cys Gly Tyr Arg Asp Lys Asp Asn Arg Asn
```

355 360 365
Gly Ser Asp Phe Thr Cys Lys Arg Cys Gly Phe Ser Arg His Ala Asp
370 375 380
Val Asn Ala Ala Asn Asn Ile Ala Arg Gly Met Glu Glu Phe Ser Leu
385 390 395 400
Pro Thr Gly Ala Thr Ala Pro Asn Lys Ala Gln Arg Leu Val Ile Ala
405 410 415
Ser

<210> 645
<211> 389
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
解淀粉嗜盐碱球菌 (Natronococcus amylolyticus) DSM 10524 序列
<400> 645

[0320]

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

210 215 220

Gln Val Ala Asn Glu Leu Ile Ala Arg Ala Arg Glu Arg Asn Ala Val
225 230 235 240
Ile Val Phe Gly Asp Met Thr Gly Leu Arg Phe Asp Asn Asp Glu Gly
245 250 255
Arg Tyr Val Asn Asp Lys Thr His Lys Met Pro Tyr Ala Lys Leu Ala
260 265 270
Asn Ile Leu Thr Tyr Lys Ser His Leu Asp Gly Arg Glu Cys Ile Pro
275 280 285
Val Ser Glu Tyr Asp Thr Ser Val Thr Cys Trp Arg Cys Gly Ser Gln
290 295 300
Asn Thr Ser Arg Gly Ala Gln Gly Arg Val Glu Cys Cys Asp Cys Gly
305 310 315 320
Leu Asp Asp Asn Ala Asp Lys Asn Gly Ala Ser Asn Ile Gly Lys Arg
325 330 335
Ala Val Gly Lys Asp Ile Gln Ser Pro Leu Ser Thr Val Gly Ala Ala
340 345 350
Val Ala Gln Pro Glu Thr Gln Val Val Leu Asp Gly Thr Ser Gly Glu
355 360 365
Met Glu Pro Ala Asn Ser Pro Lys Asp Val Gly Leu Thr Leu Ser Glu
370 375 380
Gly Ser Pro Arg Leu
385

[0321]

<210> 646
<211> 363
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
泥炭宏基因组序列
<400> 646
Met Ser Leu Val Lys Val Lys Lys Cys Tyr Tyr Cys Lys Val Ile Glu
1 5 10 15
Asp Leu Thr Asn Thr Lys Leu Val Ala Leu Gln Gln Glu Tyr Asp Asn
20 25 30
Leu Gln His Tyr Leu Gln Thr Lys Glu Asp Leu Gly Leu Tyr Ser Ala
35 40 45
Asn Lys Gln Gln Ala Asn Arg Phe Tyr Lys Val Ile Lys Arg Gly Lys
50 55 60
Lys Tyr Pro Leu Ser Ile Arg Asn Asp Leu Ile Arg Ile Glu His Lys
65 70 75 80
Pro Ser Thr Ile Ala Glu Tyr Trp Ile Arg Ile Pro Val Lys Ser Val
85 90 95
Lys Gly Gly Leu Trp Ile Gly Leu Ile Lys Pro Tyr Glu Pro Ile Pro

[0322]

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

<210> 647
<211> 358
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
沉积物宏基因组序列
<400> 647
Met Leu Val Arg Lys Val Ile Lys Ala Lys Ile Met Asp Glu Ala Asn
1 5 10 15
Ser Arg Lys Leu Lys Ala Leu Glu Lys Glu Tyr Asn Asn Phe Gln Arg

[0323]

```
                20                  25                  30
Ala Leu Arg Gly Glu Asp Val Lys Leu Tyr Ser Ala Thr Arg Gln Gln
        35                  40                  45
Ala Glu Arg Leu Leu Arg Lys Ile Arg Lys Gln Arg Gly Lys Leu Lys
    50                  55                  60
Gln Lys Glu Tyr Pro Leu Ile Leu Arg Asn Asp Val Tyr Asp Val Arg
65                  70                  75                  80
Ile Thr Asn Asn Lys Leu Ser Lys Tyr Trp Ile Lys Ile Pro Val Tyr
                85                  90                  95
Gly Val Arg Gly Gly Ile Lys Leu Pro Ile Lys Pro His Cys Glu Ile
            100                 105                 110
Asp Ala Ser Trp Lys Phe Lys Glu Ala Lys Val Ile Arg Lys Glu Asn
        115                 120                 125
Gly Trp Phe Val His Ile Thr Val Glu Lys Glu Ile Glu Ile Lys Lys
    130                 135                 140
Glu Tyr Glu Asn Val Ile Ala Ile Asp Leu Gly Ile Arg Asn Ile Ala
145                 150                 155                 160
Thr Val Val Ser Thr Ala Asp Gly Lys Pro Arg Phe Tyr Gly Lys Glu
                165                 170                 175
Leu Arg Arg Ile Arg Gly His Tyr Phe His Leu Arg Arg Arg Leu Ala
            180                 185                 190
Lys Leu Lys Lys Phe Lys Ala Ile Lys Lys Ile Gly Asn Lys Glu Arg
        195                 200                 205
Arg Val Val Asp Asp Leu Leu His Lys Ile Ser Arg Glu Ile Val Asn
    210                 215                 220
Arg Ala Lys Glu Glu Asn Ala Val Ile Val Leu Gly Asp Leu Lys Arg
225                 230                 235                 240
Ile Arg Lys Gln Asn Arg Gly Arg Lys Phe Asn Arg Lys Leu Asn Thr
                245                 250                 255
Phe Pro Tyr Tyr Arg Leu Ser Lys Phe Ile Glu Tyr Lys Ala Leu Trp
            260                 265                 270
Glu Gly Ile Ala Val Ile Lys Val Asn Glu Ala Asn Thr Ser Lys Ile
        275                 280                 285
Cys Ser Arg Cys Gly Ser Gln Gly Leu Arg Arg Asn Gly Lys Phe Val
    290                 295                 300
Cys Ser Lys Cys Gly Val Glu Leu Asn Ala Asp Tyr Asn Gly Ala Leu
305                 310                 315                 320
Asn Ile Leu Lys Arg Ala Leu Gly Tyr Met Ser Lys Ala Gly Ala Asp
                325                 330                 335
Leu Thr Gln Pro Phe Gln Ser His Phe Gly Leu Ile Leu Thr Met Lys
            340                 345                 350
Lys Arg Tyr Gly Arg Asp
        355
<210> 648
```

<211> 512
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
沉积物宏基因组序列
<400> 648

[0324]

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

290 295 300

Asn Pro Ala Leu Pro Lys Pro Pro Asn Ile Ala Val Pro Ser Ser Lys
305 310 315 320
Phe Trp Gln Ile Val Glu Arg Val Glu Ala Glu Trp Arg Lys Trp Gln
325 330 335
Gln Leu Arg Thr Gly Lys Arg Asp Lys Glu Gly Arg Val Val Leu Lys
340 345 350
Lys Arg Lys Met Ala Trp Arg Val Leu Lys Arg Lys Arg Arg Met Arg
355 360 365
Arg Glu Val Thr Lys Ser Met Ile Ala Glu Ala Val Asn Thr Phe Ile
370 375 380
Lys Ser Leu Glu Arg Pro Cys Ala Ile Val Ile Glu Gln Ile Lys Asn
385 390 395 400
Leu His Leu Ser Pro Thr Leu Gln Asp Arg Lys Ser Arg Trp Arg Gln
405 410 415
Arg Arg Trp Ala Tyr Arg Phe Leu Leu Glu Met Leu Arg Tyr Lys Ala
420 425 430
Glu Trp Glu Gly Ile Arg Val Ile Glu Val Asn Pro Ala Tyr Thr Ser
435 440 445
Gln Thr Cys Pro Arg Cys Gly Glu Arg Asp Arg Asn Ala Arg Gln Trp
450 455 460

[0325]

Asn Leu Phe Arg Cys Thr Lys Cys Arg Phe Gln His Asn Ala Asp Phe
465 470 475 480
Val Ala Ala Tyr Asn Ile Thr Val Arg Gly Leu Met Lys Leu Arg Thr
485 490 495
Arg Gly Arg Ile Asp Trp Lys Ile Leu Thr Gly Lys Pro Arg Val Leu
500 505 510

<210> 649
<211> 512
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
沉积物宏基因组序列
<400> 649

Met Glu Lys Val Ile Leu Thr Val Thr Ser Gly Val Glu Phe Leu Gln
1 5 10 15
Pro Gln Asp Glu Gln Leu Val Ile Lys Ala Ala Lys Trp Tyr Thr Glu
20 25 30
Ala Arg Lys Ala Leu Leu Tyr Asn Ala Leu Gln Lys Gly Glu Glu Gly
35 40 45
Val Lys Gln Ala Ala Ser Thr Leu Ser Thr Leu Asn Phe Pro Pro Ala
50 55 60
Pro His Ser Arg Leu Pro Asp Leu Leu Thr Arg Asp Val Gly Gly Thr

65 70 75 80
Leu Gln Ala Leu Lys Glu Arg Arg Glu Gly Tyr Ala Gln Asp Phe Ser
85 90 95
Arg Trp Trp Glu Glu Ile Cys Arg Val Ala Asn Glu Lys Gly Ser Gln
100 105 110
Ala Val Ser Val Leu Lys Pro His Val Asn Lys Leu Ser Ser Leu Ile
115 120 125
His Arg Leu Ser Trp Asn Ser Pro Glu Trp Asp Glu Glu Ala Gln Glu
130 135 140
Val Phe Lys Glu Leu Lys Gln Ile Trp Gly Val Val Lys Gln Lys Asp
145 150 155 160
His Glu Arg Leu Lys Ile Lys Trp Cys Ala Glu Val Pro Pro Val Pro
165 170 175
Arg Pro Pro Arg Thr Pro Lys Phe Gln Arg Leu Val Phe Asn Ala Asp
180 185 190
Gly Arg Val Ser Gly Gly Ser Arg Ser Asn Ala Leu Gln Ile Glu Glu
195 200 205
Arg Glu Asp Gly Asp Leu Tyr Leu Arg Leu Gln Leu Arg Asp Pro Glu
210 215 220
Arg Val Ala Lys Lys Gly Ser Tyr Gln Asp Val Trp Leu Arg Leu Val
225 230 235 240
Ile Glu Asp Glu Glu Arg Lys Gln Met Leu Lys Asn Phe Ala Pro Lys
245 250 255
Tyr Leu Gln Leu Arg Gln Asp Gly Lys Arg Trp Ile Ala His Ile Val
260 265 270
Ile Glu Arg Glu Asn Glu Phe Glu Val Pro Gln Tyr Val Tyr Gly Val
275 280 285
Asp Ile Gly Ala Arg Lys Tyr Leu Leu Ala Ala Val Ala Ile Pro Leu
290 295 300
Asn Pro Ala Leu Pro Lys Pro Pro Asn Ile Ala Val Pro Ser Ser Lys
305 310 315 320
Phe Trp Gln Ile Val Glu Arg Val Glu Ala Glu Trp Arg Lys Trp Gln
325 330 335
Gln Leu Arg Thr Gly Lys Arg Asp Lys Glu Gly Arg Val Val Leu Lys
340 345 350
Lys Arg Lys Met Ala Trp Arg Val Leu Lys Arg Lys Arg Arg Met Arg
355 360 365
Arg Glu Val Thr Lys Ser Met Ile Ala Glu Ala Val Asn Thr Phe Ile
370 375 380
Lys Ser Leu Glu Arg Pro Cys Ala Ile Val Ile Glu Gln Ile Lys Asn
385 390 395 400
Leu His Leu Ser Pro Thr Leu Gln Asp Arg Lys Ser Arg Trp Arg Gln
405 410 415
Arg Arg Trp Ala Tyr Arg Phe Leu Leu Glu Met Leu Arg Tyr Lys Ala

[0326]

420 425 430

Glu Trp Glu Gly Ile Arg Val Ile Glu Val Asn Pro Ala Tyr Thr Ser
435 440 445

Gln Thr Cys Pro Arg Cys Gly Glu Arg Asp Arg Asn Ala Arg Gln Gly
450 455 460

Asn Leu Phe Arg Cys Thr Lys Cys Arg Phe Gln His Asn Ala Asp Phe
465 470 475 480

Val Ala Ala Tyr Asn Ile Thr Val Arg Gly Leu Met Lys Leu Arg Thr
485 490 495

Arg Gly Arg Ile Asp Trp Lys Ile Leu Thr Gly Lys Pro Arg Val Leu
500 505 510

<210> 650

<211> 351

<212> PRT

<213> 未知

<220>

<223> 未知的描述:
沉积物宏基因组序列

<400> 650

[0327]

Met Val Ala Val Lys Thr Ile Lys Ala Lys Ile Ile Ala Pro Thr Arg
1 5 10 15

Arg Lys Ala Glu Ile Leu Glu Lys Glu Tyr Gln Gly Phe Gln Asp Ile
20 25 30

Leu His Gly Gln Gln Ala Glu Ile Tyr Ser Arg Thr Arg Asp Gln Ala
35 40 45

Leu Arg Val Leu Arg Arg Ile Lys Ala Pro Lys Asp Arg His Tyr Pro
50 55 60

Leu Ile Leu His Arg Ala Ser Ile Arg Leu Glu Arg Arg Asp Thr Lys
65 70 75 80

Ile Ala Gln Tyr Trp Phe Arg Phe Pro Cys His Gly Val Arg Gly Gly
85 90 95

Ile Trp Leu Ala Ile Lys Pro His Cys Gln Ile Pro Glu Gly Cys Ser
100 105 110

Leu Arg Glu Thr Lys Leu Ile Lys Arg Asn Asn Ala Trp Phe Leu Tyr
115 120 125

Ile Ile Ile Lys Lys Glu Val Pro Glu Pro Val Ile Asn Pro Asp Arg
130 135 140

Ile Val Ala Val Asp Leu Gly Glu Arg Tyr Leu Ala Thr Val Cys Gly
145 150 155 160

Gly Glu Gly Ile Arg Pro Gln Phe Leu Gly Lys Glu Ala Arg Gly Ile
165 170 175

Arg Arg His Tyr Ala Trp Leu Arg Arg Arg Leu Gly Glu Leu Lys Leu
180 185 190

Leu Arg Thr Ile Lys Lys Val Gly Asn Thr Glu Arg Arg Lys Ile Asn

195 200 205

Thr Ile Cys His Gln Ile Ser Arg Lys Ile Val Asp Leu Ala Lys Gln
210 215 220
Thr Gly Ser Thr Ile Val Leu Gly Asp Leu Lys Gly Ile Arg Gln Arg
225 230 235 240
Ala Arg Gly Arg Arg Met Asn Arg Ile Val Ser Gly Met Pro Tyr Tyr
245 250 255
Lys Leu Thr Gln Phe Ile Lys Tyr Lys Ala Ala Trp Glu Ser Val Pro
260 265 270

Val Leu Ile Val Ser Glu Arg Asp Thr Ser Lys Thr Cys His Arg Cys
275 280 285
Gly Ala Ile Gly Thr Arg Pro Tyr Gln Gly Leu Phe Leu Cys Gln Ala
290 295 300
Cys Gly Leu Arg Tyr Ser Ala Asp Leu Asn Gly Ala Lys Asn Ile Met
305 310 315 320
Lys Arg Ala Arg Glu Tyr Ile Ser Leu Ala Gly Ala Ala Phe Ser Gln
325 330 335
Pro Met Thr Arg Ala Leu Ala Gly Asn Pro Leu Met Glu Thr Gly
340 345 350

[0328]

<210> 651
<211> 359
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
土壤宏基因组序列
<400> 651
Met Lys Ile Thr Val Lys Ala Lys Val Thr Asp Asp Ser Lys Asn Lys
1 5 10 15
Leu Glu Lys Ile Glu Ser Glu Phe Gln Asn Tyr Gln Val Tyr Leu Gln
20 25 30
Ser Gly Cys Thr Ala Ile Val Asn Leu Tyr Ser Ala Thr Lys Gln Gln
35 40 45
Thr Gln Arg Ala Leu Lys Lys Ile Lys Gln Pro Asn Gln Lys Lys Gln
50 55 60
Tyr Pro Val Ile Leu Arg Arg Asp Leu Ile Asp Ile Gln Lys Asp Ile
65 70 75 80
Lys Phe Lys Gly Val Tyr Trp Met Arg Ile Pro Val Tyr Pro Asn Ser
85 90 95
Ile Asn Ile Arg Ile Lys Thr Ser Cys Lys Tyr Asp Leu Thr Glu Tyr
100 105 110
Ser Ile Arg Glu Cys Lys Ile Leu Lys Ile Arg Asn Glu Trp Tyr Ile
115 120 125

Phe Ile Thr Ile Glu Lys Gln Asp Glu Pro Ile Ile Pro Pro Asp Asn
130 135 140
Val Leu Ala Val Asp Leu Gly Cys Lys Asn Ile Ala Val Thr Val Asn
145 150 155 160
Thr Ile Asn Thr Lys Pro Asn Phe Tyr Gly Ala Ala Leu Arg Gly Ile
165 170 175
Arg Gly Phe Tyr Phe Asn Leu Arg Arg Lys Leu Gly Glu Lys Lys Ala
180 185 190
Ser Asn Lys Ile Lys Ser Leu Lys Asn Ile Glu Phe Leu Gln Val Asn
195 200 205
His Glu Leu His Lys Ile Ser Lys Ala Ile Val Glu Glu Ala Lys Arg
210 215 220
Thr Asn Ala Val Ile Val Leu Gly Lys Leu Lys Gly Ile Arg Lys Asn
225 230 235 240
Ile Lys Gly Ser Arg Arg Met Arg Arg Leu Ile Asn Asn Phe Pro Tyr
245 250 255
Tyr Arg Leu Val Gln Tyr Ile Lys Tyr Lys Ala Ala Trp Arg Ser Ile
260 265 270
Arg Ala Leu Glu Ile Ser Glu Ala Asp Thr Ser Asn Thr Cys Phe Asn
275 280 285
Cys Arg Thr Lys Asp Lys Lys Ala Arg Lys Thr Gln Gly Leu Phe Gln
290 295 300

[0329]

Cys Ser Asn Cys Ser Ile Arg Thr Asn Ala Asp Tyr Asn Gly Ala Met
305 310 315 320
Asn Ile Leu Gln Arg Gly Val Gly Ile Leu Ser Thr Leu Gly Gly Phe
325 330 335
Leu Thr Tyr Pro Glu Pro Ser Val Ile Ile Asp Arg Asn Lys Met Ile
340 345 350
Thr Lys Glu Pro Gln Met Leu
355

<210> 652
<211> 457
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
土壤宏基因组序列
<400> 652
Met Ile Lys Thr Ala Thr Phe Trp Leu Cys Glu Val Thr Ser Asp Gln
1 5 10 15
Arg Ala Val Leu Val Asp Ile Ser Lys Gln Leu Gly Ala Ala His Arg
20 25 30
Ala Met Val Cys Glu Trp Leu Arg Leu Ala Met Ala Gly Glu Pro Tyr
35 40 45

[0330]

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

Ala Ala Gly Thr Ser Gln Thr Cys Ser Val Cys Gly His Arg His Arg
355 360 365
Asp Asn Arg Gln Arg Ser Gln Phe Glu Cys Leu Ala Cys Gly Tyr Thr
370 375 380
Ala His Ala Asp Val Asn Ala Ala Asn Asn Ile Ala Arg Asn Leu Glu

385 390 395 400
Thr Phe Gly Pro Val Thr Gly Ser Arg Ser Ala Gln Glu Ala Pro Val
405 410 415
Gly Leu Arg Phe Asp Trp Glu Gly Arg Glu Pro Gly Leu Arg Gly Lys
420 425 430
Ser Ser Arg Leu Gln Leu Asp Leu Gly Thr Gln Val Ser Arg Ser Ala
435 440 445
Glu Gly Gly Val Ile Asn Ala Ala Thr
450 455

<210> 653
<211> 437
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
土壤宏基因组序列

<400> 653

[0331]

Met Asp Ile Ser Lys Gln Leu Gly Ala Ala His Arg Ala Met Val Cys
1 5 10 15
Glu Trp Leu Arg Leu Ala Met Ala Gly Glu Pro Tyr Gly Glu Asn Ala
20 25 30
Gly Lys Gln Lys Ala Arg Glu Val Phe Ala Asp Thr Pro Tyr Ser Asn
35 40 45
Gly Leu Val Asn Asn Tyr Ser His Ala Asp Ala His Pro Arg Phe Lys
50 55 60

Lys Gln Trp Glu Ala Val Leu Lys Gly Glu Arg Ser Ser Val Ala Phe
65 70 75 80
Pro Gly His Ala Pro Phe Ile Tyr Leu Gly Ala Pro Glu Gln Val Ser
85 90 95
Val Glu Leu Leu Asp Gly Gly Trp His Ile Val Asp Pro Arg Leu Tyr
100 105 110
Ile Gly Asp Ser Lys Asp Asn Gly Arg Ser Ile Thr Trp Gln Glu Gly
115 120 125
Arg Ser Arg Asn Arg Ser Ile Ala Trp Arg Leu Asp Pro Val Asp Asn
130 135 140
Arg Arg Thr Gly Trp Cys Leu Gly Leu Leu Arg Asp Ala Glu Cys Val
145 150 155 160
Ser Cys Val Arg Leu Met Lys Ser Asp Arg Arg Pro Gly Arg Trp Leu
165 170 175
Ala Lys Ile Ser Val Glu Leu Ala Glu Ala Ala Ser Ala Arg Asp Lys
180 185 190
Glu Glu Val Glu Leu Cys Ala Gly Ile Asp Leu Gly Leu Asn Cys Pro
195 200 205

Val Val Leu Ser Val Pro Asp Lys Gly Leu Leu Arg Phe Phe Gly Arg
210 215 220
Glu Gly Asp Met Tyr Pro Arg Leu Trp Gln Lys Leu Arg Gln Tyr Glu
225 230 235 240
Asp Arg Lys Arg Ala Phe Asn Arg Ala Gly His Lys Arg Glu Ala Arg
245 250 255
Asp Leu Arg Met Asn Leu Thr Asn Val Arg Gln Tyr Ile Asn Glu Cys
260 265 270
Val Ser Lys Ala Val Val Asp Tyr Cys Leu Ala Asn Arg Val Thr Lys
275 280 285
Leu Arg Leu Glu Asn Leu Lys Gly Leu Gly Val Gly Ala Thr Gly Lys
290 295 300
Leu Lys His Trp Pro Arg Tyr His Leu Gln Gln Arg Ile Glu Gln Lys
305 310 315 320
Gly Ala Ala Ala Gly Leu Glu Val Glu Tyr Val Lys Ala Ala Gly Thr
325 330 335
Ser Gln Thr Cys Ser Val Cys Gly His Arg His Arg Asp Asn Arg Gln
340 345 350
Arg Ser Gln Phe Glu Cys Leu Ala Cys Gly Tyr Thr Ala His Ala Asp
355 360 365
Val Asn Ala Ala Asn Asn Ile Ala Arg Asn Leu Glu Thr Phe Gly Pro
370 375 380

[0332]

Val Thr Gly Ser Arg Ser Ala Gln Glu Ala Pro Val Gly Leu Arg Phe
385 390 395 400
Asp Trp Glu Gly Arg Glu Pro Gly Leu Arg Gly Lys Ser Ser Arg Leu
405 410 415
Gln Leu Asp Leu Gly Thr Gln Val Ser Arg Ser Ala Glu Gly Gly Val
420 425 430
Ile Asn Ala Ala Thr
435

<210> 654
<211> 419
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
陆生-土壤序列
<400> 654
Met Ser Leu Lys Thr Val Gly Leu Tyr Leu Ala Glu Met Pro Gln Glu
1 5 10 15
Asp Lys Val Val Cys Lys Ala Ile Cys Arg Asp Leu Gly Arg Ala Gly
20 25 30
Ala Arg Met Val Arg Glu Trp Tyr Arg Ala Cys Asp Glu Val Ala Pro
35 40 45

[0333]

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

Leu Pro Thr Gly Ala Gln Ala Met Arg Pro Ala Arg Arg Leu Glu Leu
405 410 415
Val Ala Gly

<210> 655
<211> 469
<212> PRT
<213> 未知
<220>
<223> 未知的描述:
废水宏基因组序列
<400> 655

[0334]

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

[0335]

Leu Trp His Glu Ile Glu Arg His Cys Glu Arg Arg Thr Ile Leu Asn
260 265 270
Lys Ala His Lys Arg His Ala Ala Ala Ala Gln Ser Asp Lys Ile Ser
275 280 285
Arg Ile Arg Lys His Gly Cys Glu Leu Ala Ser Arg Gln Leu Ile Asp
290 295 300
Trp Cys Met Tyr His His Val Thr Thr Ile Arg Leu Glu Asp Leu Ser
305 310 315 320
Gly Ile Arg Asp Arg Cys Ser Glu Asp Glu Asn Ser Glu Lys Ala Arg
325 330 335
Thr Trp Asn Arg Arg Leu Ser His Trp Pro Trp Tyr Gln Leu Gln Gln
340 345 350
Arg Ile Glu Gln Lys Ala Asp Ala Tyr Gly Ile Thr Val Glu Lys Val
355 360 365
Asn Pro His Leu Thr Ser Gln Thr Cys Ser Ser Cys Gly Tyr Ala Asn
370 375 380
Pro Ala Asn Arg Glu Gly Thr Gln Phe Lys Cys Gln Arg Cys Gly Phe
385 390 395 400
Glu Arg His Ala Asp Leu Asn Ala Ala Asn Asn Ile Ala Arg Arg Glu
405 410 415
Pro Gly Val Asp Phe Asp Ala Pro Asn Thr Pro Gly Arg Ala Ala Val
420 425 430
Leu Ser Asp Gly Pro Lys Ser Arg Pro Lys Ala Gly Arg Ser Gly Arg
435 440 445
Leu Val Leu Ser Asp Asn Pro Val Thr Pro Arg Pro Asp Thr Leu Gln
450 455 460
Leu Asp Met Val Leu
465

<210> 656
<400> 656
000
<210> 657
<400> 657
000
<210> 658
<400> 658
000
<210> 659
<400> 659
000
<210> 660
<400> 660
000
<210> 661

[0336]

<210> 661
<400> 661
000
<210> 662
<400> 662
000
<210> 663
<400> 663
000
<210> 664
<400> 664
000
<210> 665
<400> 665
000
<210> 666
<400> 666
000
<210> 667
<400> 667
000
<210> 668
<400> 668
000
<210> 669
<400> 669
000
<210> 670
<400> 670
000
<210> 671
<400> 671
000
<210> 672
<400> 672
000
<210> 673
<400> 673
000
<210> 674
<400> 674
000
<210> 675
<400> 675
000

<210> 676
<400> 676
000
<210> 677
<400> 677
000
<210> 678
<400> 678
000
<210> 679
<400> 679
000
<210> 680
<400> 680
000
<210> 681
<400> 681
000
<210> 682
<400> 682
000
[0337] <210> 683
<400> 683
000
<210> 684
<400> 684
000
<210> 685
<400> 685
000
<210> 686
<400> 686
000
<210> 687
<400> 687
000
<210> 688
<400> 688
000
<210> 689
<400> 689
000
<210> 690
<400> 690

000
<210> 691
<400> 691
000
<210> 692
<400> 692
000
<210> 693
<400> 693
000
<210> 694
<400> 694
000
<210> 695
<400> 695
000
<210> 696
<400> 696
000
<210> 697
<400> 697
000

[0338]

<210> 698
<400> 698
000
<210> 699
<400> 699
000
<210> 700
<400> 700
000
<210> 701
<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> 701
ctctccacgc gcgcgcggga tgcggg 26
<210> 702
<211> 29
<212> DNA
<213> 人工序列

<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 702
gcctgctgtc cacactatcc cccaaaaac 29
<210> 703
<211> 28
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 703
gtgctctcca cgcgcgcgcg ggatgcgg 28
<210> 704
<211> 29
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 704
[0339]
ggtctctcca cgcgcgcgcg ggatgcggg 29
<210> 705
<211> 29
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 705
gtgctctcca cgcgcgcgcg ggatgcggg 29
<210> 706
<211> 29
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 706
gtgcggccca cgcgcgcgcg gggcagagg 29
<210> 707
<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> 707
tggtgaccat gtttttgttg aa 22
<210> 708
<211> 20
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成寡核苷酸
<400> 708
ccagtccaag tttcgatgag 20
<210> 709
<211> 20
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成寡核苷酸
[0340]
<400> 709
ggaaacaaaa taacaaacaa 20
<210> 710
<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> 710
tataaggtgt tttgaaaatg a 21
<210> 711
<211> 20
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成寡核苷酸
<400> 711
gagaaggcaa aaagagaagc 20
<210> 712
<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> 712
tttcaattcc cattatatgg gatttctctt tgaaacc 37
<210> 713
<400> 713
000
<210> 714
<211> 29
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 714
gttgtctcca cgcgcgcgag ggcggtttg 29
<210> 715
<211> 31
[0341] <212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 715
ctttcaatcc cacaatggtg cgattgctgc g 31
<210> 716
<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> 716
agtgcggggg gaacctagga gggcagtggg c 31
<210> 717
<211> 29
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸

<400> 717
gttcgcgaga gactttgtga ataacaagc 29
<210> 718
<211> 29
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 718
gcctgatgtc cacattatcc ccttaagac 29
<210> 719
<211> 30
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 719
gcctgatgtc cacattatcc ccttaagacc 30
<210> 720
<211> 30
[0342]
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 720
cgtgcgatcc acgcgcgcgc ggggcagagg 30
<210> 721
<211> 29
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 721
gcctgctgtc cacattatcc ccttaagac 29
<210> 722
<211> 32
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成

寡核苷酸

<400> 722

atattatgat aaagcattgg ctatagatcc aa 32

<210> 723

<211> 19

<212> DNA

<213> 人工序列

<220>

<223>人工序列的描述：合成

寡核苷酸

<400> 723

caacagagtc accaacagc 19

<210> 724

<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> 724

agtaattact gagtcattac tc 22

[0343]

<210> 725

<211> 23

<212> DNA

<213> 人工序列

<220>

<223>人工序列的描述：合成

寡核苷酸

<400> 725

gcagctgggg ctttggcggt gtt 23

<210> 726

<211> 20

<212> DNA

<213> 人工序列

<220>

<223>人工序列的描述：合成

寡核苷酸

<400> 726

acacttcgat agaggggtgt 20

<210> 727

<211> 30

<212> DNA

<213> 人工序列

<220>

<223>人工序列的描述：合成
寡核苷酸
<400> 727
agcttactat ccacaaagtc ttccgcgaac 30
<210> 728
<211> 34
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 728
gaccctgagc ggtgaaagcc ctcgacccgc tcgc 34
<210> 729
<211> 35
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 729
[0344] cgtcgcacct cgcgcgcgcg taggagatgg tttgg 35
<210> 730
<211> 32
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 730
gtcgcacctc gcgcgcgcgt aggagatggt tt 32
<210> 731
<211> 29
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 731
gtgccgccca cgggcgcagg ggtggattg 29
<210> 732
<211> 33
<212> DNA
<213> 人工序列

<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 732
gtcgcacctc gcgcgcgcgt aggagatggt ttg 33
<210> 733
<211> 29
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 733
ggtcggcccg cgcgcgcgag gggggttgg 29
<210> 734
<211> 29
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 734
[0345]
gtgcgaccca cgcgcgcgcg gggcagagg 29
<210> 735
<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> 735
gtcgtcttcg tcgccgatat cg 22
<210> 736
<211> 19
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 736
taaccagaat agcggtaat 19
<210> 737
<211> 30
<212> DNA

<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 737
ctttcaatcc cattttggtc tgattttaac 30
<210> 738
<211> 29
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 738
gttgagtggc cacatagtgg gagctgaac 29
<210> 739
<211> 24
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
[0346]
<400> 739
gtattatgat aaagcattgg ctat 24
<210> 740
<211> 32
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 740
gccgtgctct ccgcgcgcgc gtgggcggtt cg 32
<210> 741
<211> 32
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 741
ggcgcaaccc gcacgcgcgc gcgagggcag ag 32
<210> 742
<211> 29

<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 742
gcagagccca cgggcgtggg ggtggattg 29
<210> 743
<211> 28
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<400> 743
tgcggcccac gcgcgcgcgg ggcagagg 28
<210> 744
<400> 744
000
<210> 745
<400> 745
000
[0347]
<210> 746
<400> 746
000
<210> 747
<400> 747
000
<210> 748
<400> 748
000
<210> 749
<400> 749
000
<210> 750
<400> 750
000
<210> 751
<400> 751
000
<210> 752
<400> 752
000
<210> 753

<400> 753
000
<210> 754
<400> 754
000
<210> 755
<400> 755
000
<210> 756
<400> 756
000

<210> 757
<400> 757
000
<210> 758
<400> 758
000
<210> 759
<400> 759
000
[0348] <210> 760
<400> 760
000
<210> 761
<400> 761
000
<210> 762
<400> 762
000
<210> 763
<400> 763
000
<210> 764
<400> 764
000
<210> 765
<400> 765
000
<210> 766
<400> 766
000
<210> 767
<400> 767

000

<210> 768

<400> 768

000

<210> 769

<400> 769

000

<210> 770

<400> 770

000

<210> 771

<400> 771

000

<210> 772

<400> 772

000

<210> 773

<400> 773

000

<210> 774

<400> 774

000

[0349]

<210> 775

<400> 775

000

<210> 776

<400> 776

000

<210> 777

<400> 777

000

<210> 778

<400> 778

000

<210> 779

<400> 779

000

<210> 780

<400> 780

000

<210> 781

<400> 781

000

<210> 782

[0350]

<400> 782
000
<210> 783
<400> 783
000
<210> 784
<400> 784
000
<210> 785
<400> 785
000
<210> 786
<400> 786
000
<210> 787
<400> 787
000
<210> 788
<400> 788
000
<210> 789
<400> 789
000
<210> 790
<400> 790
000
<210> 791
<400> 791
000
<210> 792
<400> 792
000
<210> 793
<400> 793
000
<210> 794
<400> 794
000
<210> 795
<400> 795
000
<210> 796
<400> 796
000

<210> 797
<400> 797
000
<210> 798
<400> 798
000
<210> 799
<400> 799
000
<210> 800
<211> 2142
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
废水-厌氧消化器-厌氧消化器污泥
序列
<400> 800

[0351]

gccttccgcg cgtggctgtc ctacgtgggc gcggaggtga cgccggagga gcggcgcaag 60
gcggagagcg cggcggtgaa cggggagtgc ccgccacaac gcagttacgg gacttcccaa 120
gtctttgatg gctttgagaa cccagcgggc gcagggatga ccaacccaga acacgccgcg 180
caacgtcacc gcgacggcgc gggtatcctc acccgcctgc tgaccgcgca ctggccgtcg 240
tcggcgcagg agcagcgcgc gcactgggag gcggtcatcg gcgaggtgcc ggagggctgg 300
gattgcacga ggccgccggg ggtgagggtt tcgcaggggc gatcccattg gggcgcgag 360
agaggccgca atggttctca tgctggcgtg ttggacaagt cctccacttt cgcgaacgag 420
gcatcgcacg acaccgcccc atcgactccc gacgccgcgc gcttcgacgg cctgcgcggc 480
tccgtcgccc cgctcaagcc cttcgcctac cccatcctgc acgcgcagaa gccgcgcgag 540
ggcagctacc tcgcgaacct gcgcgcgcat ggcgtcggcg cgtggaatgt ggaggaggcg 600
agggtgccgt tcagtgagga tggtgtcgag ggcttctggc cgggcgcccg aagtggacgc 660
ggcgacctgc agcccccgtc gtggggagcc acagaagggg ggcagcgcga tgacgcaggc 720
cgcacccccct ccaacctcct gtcctacggc gacctgctcg gcgacctgca acgctactgc 780
gacctcgacg cctggtgcga gcgcctcggc ctgcccgacg ccgccgccga cctgctcgaa 840
gcggggttgg tctacgcgcc gaagccgagc agggcggaga aagaggccgg gctggaggca 900
gacaagccgc ccgcgttcgg caacgataag ggcgatggcc ttgggcgggg catcagcaac 960
acgagacagg attgggcgcg gggcaacacg catcccaccg cgaagcccgc cgcgctctgc 1020
gcctacctcg tcgccctcgc cacccgcgag gggcagaccg tgcttgaccc gttctgcggg 1080
tcgggcagca cgggcgtcgc ctgcgcgcgc atcggcagga acttcgtcgg cgtcgaactc 1140
gacgccgact tcgccgccat cgcacgagcg cggatcgcac acgccaagac cacacgaccg 1200
gctcaggagg ccctcccgct atgaccgcgg acctccgcga ccggaagttg actcattccc 1260
tccattctgg tatactgtca gcgttccagg caagggactc agaggaggtc gtttcagtgt 1320
acgctcgaca gttttgcccc cgtgtctcca gcgacctcct gttcccttgc ctgggcagag 1380
acgctgtcac gggggctgtt cgtgtacgcg cttcgagaca agccgacccg gctgcggttg 1440
ggttggttgc ggctactctg ggccgcgagg ccgggcgatc agacccggca aacccgcgag 1500
tacaagtgtg ggcagcatgg ggcgggcaac tctgctcatc gtagtgggag ctggcaagca 1560

tgatcacagg aggcaagact gatgatcaag accgtcacct tctatatcgc tgaaatgccg 1620
agggaggaca agcgcctctg cgaccgcctc tgcatcgacc tcggggccgc cggggcgcgc 1680
atggtgcgcg agtggttccg cgcctgtgac gaaaccccgc ccggtgcggc ctcagacttc 1740
ctctgcttgc ggtgcggctt cgcgcgccat gccgacgtga acgccgcgaa caacatcgcg 1800
cgggggatgc aggacttcga tttgccagcc gggtccgccg cccctgggcg agcgcttcgg 1860
ctggagttga gcgcgcaatg agacaatgcg ggcgggctgc ggttcgcgcg tgtgcggcga 1920
ctgtgcgccg cgaggacgac ggtcaagttg gcgagtcatt gttatgcaat gaacacaggc 1980
ctgcccacgg gacggatgaa gggcgcggtg cgcctcaccg cctgcgcgac cgtcagcgac 2040
ctggtgcgcg cggggctggt cgtcgtcgcg tgggcctacg agaacggcga cctcgggctg 2100
cgggggacgg aggcggcggc gactttcttt cgcgagggtt ga 2142

<210> 801
<211> 1339
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
真菌-菌丝-切叶蚂蚁真菌园序列

<400> 801

[0352]

agaatgaaga tgcattaatt tggaaaggta tagctttctt tacaatgaag aaaacgttat 60
ttctgacaat cgcaataaca atcgctgtgg tggcggtagc gttacctttc gtagcaccaa 120
taatcgcaaa agtagctaat gctggcggta gcgttacctt tcgtagcacc aataatcgca 180
aaagtagcta atgctgacgg taattagttt tgataaattg tcggcctata aggtcatggg 240
cattagcaat ttacaatata tttgaagcct gaaagcccgc ctcctttagg agcgggatga 300
aatgctgacc gatatcttaa tagctccaaa gctctagtat ttgtatgagt cgaggactct 360
tcgaatgaaa atcacagtta aagcaaaagt cacagatgac agcaaaaaca aactaggaaa 420
aatagaatca gagtttcaaa actatcaggt atatttgcaa tcaggatgta ccacaagagt 480
taatctatat tctgcgacga aacaacaagc acaaaacaca tgctttaact gtcgcactaa 540
agataagaaa gcaagaaaaa cacaaggatt gttccaatgt ggtaattgca gcattaaaac 600
gaatgctgat tataatggag caatgaacat cttgcaacga ggggttggat actttcaact 660
ctagtgagct ttctgacata ccccgaaccg tcggtgatta tagacaggaa taaaatgatc 720
acgaaggaac cccacatgct ttatctctgg gaggatgtca gaggatcaaa cggctaacaa 780
taccatagat aggttctcct ccgaaatgat gtcctgtgtc tttatcgttt aggctaagag 840
tcattgtgtt tcccttggaa atcaatactg atatatgcgc attattccac atagttacac 900
catttatttt tacatcagca gtaccattca aagataagct attgtctgat gtcatttgaa 960
ctagttgtgt ttttcctcca atattatcaa ttttgaagtt tgttatttgg tgaatatgtg 1020
gtcttctacc atccatcggt gatgcaagga agtctacacc aaaattcgta acgttccctt 1080
tatgcacatt caagttccat tctccagata agataacctt aaaagcatca gttatattaa 1140
atccagattc gggaattgta ataattaaag aatttattgt tcctattgca gagaatgtat 1200
cctttgctgg gattggaaaa aatactggta gtgaagtcac cgctattttt ttagctgcat 1260
atgcctcata gtaataacta tgagtattgt tattactagc gttgatatct actacaacga 1320
cgatccctat tattatcca 1339

<210> 802
<211> 897
<212> DNA

<213> 未知
<220>
<223> 未知的描述:
废水宏基因组序列
<400> 802
ctgcggccac cctgccccgt ggactcccc gaggggcagc aggtgctgcg gggcctcttt 60
cggcaggcgt tcgcggcgtt gccgccggag gcgctggagc ggtgctgcga ggcgatcacg 120
gcggaggatg acgaggacgc tggtgggtga tgcccacacc cgccacgcgg caggcccccc 180
gtccggcaca tcctcgactg ggcaggggggc caggcacaca caggacgcac ccgccgccgg 240
cctccttccc ccgcggcggt agacggaccc ggtcttgctc ccgcgcggac cggcgacgac 300
aggggccggg gctcaactct cactcccggc cccggggccc gagcaacaga ctacagacag 360
gagtaacggc agtgaccacc gacaagcaga cggtaaccaa gacgctccgc ttctggctca 420
ccgacctgag caaggaggac ttcgacctcc tgcacaagct cagccgcgaa ctggcacgtt 480
gccaccgtct gtttgtccag gcgtcccttg agtacgcgtt ctttgacgca cccaacacac 540
ccggcagggc tgcggttctg tcggacgggc cgaaatcacg gcccaaggcg gggcggtctg 600
ggaggttggt tctgtcagac aaccccgtaa ccccgcgccc agacacactg caacttgaca 660
tggtgctctg acctgtacag gtttcgcgag attccttgta ggagcgtggt agacatgccg 720
acatgcagca actgcctggc cctcactgag cgcatccgtc gactcaaggc tgacgaggac 780
gcgcacctgg aacagatcgc cgccctggga gcccggcttg cgcgcagcga ggccaaactt 840
gcgcggtcgc aggcgctgtg ccggcggctg gcgagcaacg tcgccgcgct gggtggc 897
<210> 803
[0353] <211> 35
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
沉积物宏基因组序列
<400> 803
ctcgttgtgc ctatgacggc ctgaatgact tcctc 35
<210> 804
<211> 3563
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
水生-淡水-沉积物序列
<400> 804
tcgcttaacc aaagatggcg aatagagaac tggtggttgt gatgcaggtg cagaaagtcg 60
tcaaggcaaa agtcgtgggg ctgactcacg ctaaacgcca gatactcgac cacgaatacg 120
aagggctcca gcgattcttg caggggggatt ccaacgttga actcgccttc aacataggca 180
gacgagcctt ggggcaagtc tccaaggtag gggctgtcgt gagccagccc gtaactgggg 240
cagtgccgtt cagcgaactc ggtagcgtcg ttcaatgtcc gaccccagaa gccacccttt 300
ttaaagggtg gtagttcacg tgcgcgagcg cctgcggcgt ttagtccaat ctactaatag 360
gtgctctccc atctccctaa atttcctcgc cgctggagaa cgcggggcat agcgaagaac 420

[0354]

tgagctgcca gcggtcaagc tctcacgaac tttatcgtca tcgggttgcc actcaattgg 480
tgcttggaca gccccagtcc atgtttcaat cagattccac ccaattggag cttgcaccgt 540
gccaatccaa gtttcaatca gttgccatgc ggcagtgacc gtgactgtcc cagtccaggt 600
ttcgatgagt ttccattcac cagtggcagt gactgttcca gtccaagttt caatctcgta 660
ccactgagca gctgctggtg cttggaccgc cccggtccat gactcggcga gttgccattc 720
agccggagca cttactgtcc cagtccaaga atcgatctgc tgccactgtc ccgtggttgc 780
attaacagtc ccgctccaag actcgacctg acgccactga ccagtggtcg cgttgaccgt 840
tccagtccaa ctctcaatct gccgccactg tcccgtggtc gcgttcacag tcccagtcca 900
gctctcaatc tcgtaccact gaacagctgc tggtgcttgg accgtcccgg cccatgactc 960
gatgagttgc catgcagcag gggcggtcac cgtgcccgtc cagctctcga cgagctgcca 1020
cgcggccggg gcgctcacgg tgccagtcca tgtttcgatg atctgccaga ctgcggaagc 1080
attgaccgtc cccatccacg actccaccaa ttgccattca gctggcgcct gaacggttcc 1140
agtccacgtc tctatgagct gccacgctgc aggtgcatta atagtacccg tccatgtctc 1200
gatcacttgc cagacggcag gcgctgtcac ggtcccagtc caggtttcga ttgtttgcca 1260
agccgcgggt gcgctgaccg ttccagtcca ggcctcgatt agctgccatt gggcaggagg 1320
tggaaccacc gtgaactgcc aagtgttttc agaaacattc tcacccgcag cgttgaccgc 1380
ccacactctc cagtagtagg tccccggggt ggagaagttg cgctggtaga tgttgtcggt 1440
gttgtcttgg tagttgtcca cgagcggcgt ggagaaggtg ggctcgttgt caatctcgat 1500
cctatggttc tccccaaacc acccgttgtc ccactcgaag gtcacagttt ccccgacccc 1560
gaagacagag ttgttctctg gactgtagag cctgggcctc ttcggaaaga atgtgtactc 1620
gtatgtgaaa aagatagtac cggtaagagc ggcgcctgtt cccatgtcgt attcaaagtt 1680
gtaggtcagg tcggcaaggg tcctatctac aacaacgtcg atgttgtctt tgtttacaga 1740
gatggatgtc aaggcctccc cagagctgcc tgtgctcata ctgacggctg agccccagtt 1800
taccccatca gcttagatcc tggtctgtcg cgtttccgaa gcggttgcac ttcctatcca 1860
atatgccgtg aggccaacgc tcttcacagt ctttttgttt tccggggcat agaacgtcag 1920
ctccccgctg gtccaagggt caccggcagc cccgctgggt ttttggcccg cggtgaacat 1980
cgtggttttc agctttgtcg cggccgatgg gcggaagtga taggtgatgt gcaaaaggcc 2040
gcttacctca tagaggtcgg aggccgcagc gctctggatg ttgaggttca cattgtcccc 2100
gttcgaccag ttctccctca tgaaggcgct gatgtcgtgg aagtagtcga acccctctac 2160
atgctcgctg cctatgttgg tcgtgtaagc tcccgaggtc tgggttttcg ccccgacgct 2220
cgcgtacacg ctcgtgctgc cgtcactagt caagatggtg tggaaaattt caaagtacgc 2280
gttctcgatg ttgtccacgc gagagctgtc caggtccgtc agataaacgg gcgtcccgaa 2340
ggcgaacagg ttggccgagc cgttgacgtt ggcggccccc atggcaaacg tgagtgtttt 2400
tatctggtcc gcgctgttgt ctgaatactc atatgtgacg taagtttcac agctgagcgc 2460
atcgtttgtt tgtcctccgg tagtaattgt aatatcctga tcaacgttgt agttccagga 2520
agccgtcggc cgggaaatat accagaaatt gagatcgttg ccggtacggg cttgctctgt 2580
ccaagcattt gtggtggctg tatcgcttcc gatcttgatg ttaagagaaa agtcggttga 2640
agaggcttgg gaagagtggg catggatttc gtaccagttg tccctcagcg tgaccccgtc 2700
ttctgggatt tcgattctgt gtgtgtaggt cgccgtgccg ccagccccta tctcggtgtt 2760
cgtgctcccc acaaaatacc tgacagtctt gatcttcgtg tccgccagct cgttctcctc 2820
atatgtgatg tacatcttgg ccgtgtggct ctggataccc gcgtcagtga cggttaattt 2880
tgggtagacg gcggccccgt ttgtccagta tgttgtaaaa gcagaggtaa aatctgcgac 2940
cacccatgcc gccacgcttt ccccgctgtt ggcggtgtac tggccggtca ctgggtctgc 3000
gtctgggttc cggtcggtcc ccgcgatgtt cagggtgacg gtcatatctg tcacgttggc 3060

agcagtcgat gcccccgtct ggacgtgata catcacgaac accttcctga acggcccagc 3120
cttgtccggg atgtagatgg tttttgtcgg ggtgaacgag taggccgtgc cagtgcccaa 3180
agcggacgat gaggttcctt gcctcgtatc ccatgaatat tggattgtgt ttatggtcac 3240
tatattttcg ttgggtttgg tgggggaggg agtagccacg gctaggggct gcatcgagat 3300
aacgaaaacg ctggctaaca tcatcgctat caatacgaat acaaagattt gttttgcctc 3360
cttccgcccg aactcttccc gccccaacct aaacaagcta caacccatct ccctcacctg 3420
cttgcctcca cttggactat ggaatgggcg cggctcacgg cagagccatc tctcatctct 3480
ctcagcaggt tcgtgaagta ctcaatcccg ctcgcggtga gcttgcagag ctttcgaggc 3540
ctccccgggg ttaaacgggg ctc 3563

<210> 805
<211> 2255
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
发酵宏基因组序列

<400> 805
cgcgttcgtg cagggcttcc ttgcgatccg cgcagacgtg aagcacccgc aggcggagga 60
tggcgaggac tcgttcggcg atcagtagcg agtggtattg acaagccgcc ccatatcatg 120
taaactgaaa gtgtcccggc aaggacttga gaggaggtcg cgttcacaat gcacatcaat 180
aggattgccc ccgcgaaagc gcgacctccg tttccttgcc gggacgagag cgttgctttt 240
[0355] cgcgggggtt ttgcatgtac gcttacaaga cgcctcgaac cggctgcggt tggttcgact 300
gcggctactc tgagccgcga ggccgggcaa cgagtcgtcc ggcaaacctg cgcgcgcccc 360
aagcgtgcga tccacgcgcg cggggcagag gacaacggca cgtgacgacc ggttcgacgt 420
gcgatccacg cgcgcgcggg gcagaggtaa ccgggtcaga gtgtcgtaca tagggcaccg 480
tgcgatccac gcgcgcgcgg ggcagaggca aggcggcctg ggcggcgctg gagaaggcgg 540
agaggcaggc ctcgcagcgt gcgatccacg cgcgcgcggg gcagaggagg agagatagct 600
aaagcaatga taaagaccgc gacattctat ctggccgaca tgccgcagga ggacaagcgc 660
ctgtgtcaac tcctctgcga tgaactcggc gcggcgggct ccgcgatggt tcgcgagtgg 720
tatcgcgcct gtgacgaagt agatggcgag cgtatcggcc ccgactttgt ttgcctgcgg 780
tgcggtttcg cccgccatgc cgacgtgaac gcagcgaaca acattgcccg tggcatggat 840
gacttcgact tgccgacggg cgcaagcgcg ccgacgaagg cgcggcgatt ggagattgca 900
tcatgagaag tggctgtcgg gctgaggttt gacggcttgc ggtgattgcg caccgcgagg 960
atagccgatg gttatcaagt gctgatacag catcgcaagc gcagatgatg gtgcagagat 1020
cgtgcttgac gaagacgaag tgcagaagct cagggaggcg ctggaggagc acgcgaagag 1080
cagtccgcag gactgacgca cacgccgcac ctgctgacgc gcatggggac gcacggggat 1140
gcagagcgcg cagcgagagg ttgccgggat gcgcttgcaa tcgacgaacg ttcgtgctat 1200
aatggaggtg tctggagaga cacggcgcgg ctttgcgttc tatggccgcc caccacaatc 1260
tgaccgtgac catattgccc tcgtgcaaac ctccgcgcgt gtctctccag actgcgtagg 1320
cggccagcgc gagggctttt ctttgctccg gggtgatcca gttcgtggct aacccaaacc 1380
actacgtcaa actcttcctc gcaacgttcg acgggcagtc cgccgatatg caggcgctca 1440
ctatcggcca gaagtacgtg tggctggcgc tcctcgcatt tgcccgagcc tacagaccgt 1500
atggcattct gaacgcagga gagcacgagg ggcgtcctat gcccgcgagt gtccgccgcc 1560

tggctgactg ggcttccgtt ggagtcggca ctgtgagtga ggcgctggca atctatgaga 1620
agagaggaat gataaagcag gataaccggc gcagtgatcc gtgttgggag atcgtcaact 1680
gggagaagtg gcaggtgtct cacggcctcg aagagtgttc gggagccgaa caagatggtg 1740
ttcgggaggc gaacacacag actgtagcgc gttcgggaga tgaacagcac tgttcgccat 1800
ctgaacgcca gtgttcggca aatgaacagc agtgttcaga tggcgaacac ctatctccgg 1860
acgcagagtg ttcaggaggc gaacacaaac aggtagaaag cctaggggcg caagggctat 1920
tggccttagt atctgacgag caggggtgtt cgcctaccga acacgattcc ctagacgtta 1980
agagaacaac attaacagca gaagaaggca cggctacttc tgcttctcaa gtagatcagg 2040
cacgacaaat cctctgcacc cttcttcgct tgcccgatgg atcaatcgcg attgatgccg 2100
cgcttatggc ggcctctcgt gttcatccgc gcgccgccga tgtccttgtc caagagtgtg 2160
gagcacttgc ggcgcgggaa gctcaaggcg gcaagccggt tgatctacaa tacctcgcct 2220
cgaatgttcg ccaagcggcg gagcgacttg agcag 2255

<210> 806
<211> 482
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
废水-厌氧消化器-厌氧消化器污泥
序列
<400> 806

[0356]

gtgcgggtgt ggcgctcgca ccgggctgaa tatcaggagg cgagcggcat gcaggaggcg 60
ctgatggtcg tggctgggat cgtgatcgcg ctcgcggcga ccggatgggt ctggttcgcg 120
ctgtgtatgc gggagggcta gattgaaggc aacgaccatc acaaagagcc tccgcttctg 180
gatcgacgaa ccgcctgagc ggtgccgtct cctctatggc gtcagcgatg aactcaccga 240
tgcatacaac gctatccccg agtactggaa gaccgatgtc cgcaaggtag ccatccgacc 300
gtctacacag ccggcgggac gtatcccgtc ggccgcgccc gactgcacgg gcgaggcagt 360
gaaggcgtca gacgttgacg ctgagctgca acctgcgcgc acgccggcag cttgtggagg 420
tgatccatat gccggcgagt gagcgagtgg agaggtgtgg ggccctggtc tgtacggggt 480
ag 482

<210> 807
<211> 350
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
固体废物-填埋物-填埋物沥滤液序列
<400> 807

gaccttaatg ctgcgaataa catcgcgcgg gggatgcagg agtttgggcc ttgtgtccag 60
tcctccgcgt cagacaacac tatcgagctg cggttcgatg gtgcggggca agaacctgcc 120
ctggggcggg caatgcgatt attcgcatga gtccgcaagt ctctacaagc gagtgcgacc 180
ccgagatggt ttggaggcaa ccagtgacca tacaggagcg agtctaccag acattagccg 240
cacgcaacat ccgtccatca gcgcacctgc tgcacgatat gggccactac ggcgaggctc 300
aggcgctcag tcagaaccct gccgagaacc cgttgcagca ggcgctcaga 350

<210> 808
<211> 1032
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
固体废物-填埋物-填埋物沥滤液序列
<400> 808

[0357]

atctaccgta tgtgccagat agtccgcata aagcgcctgc agatggagca gtttcggaga 60
gcgcaggcca gggcggaagc ggagatggac tggcgccgtc aggcgcatcg caaatggcag 120
cggttgcgcg aagcgatgga cagccggtag tcgtcttgcc gtggacctct cccgccgcga 180
ctgacatgcc cggaggaaac tctggccccc cccacacccg gggctccgat cacagccggg 240
gccccggccc acaggccgca cccccgcgtt ggcggggaca acgacataag aggtaggccg 300
tagacgcaga ggcacagagg acgcgtgacg aggggcgctg cgacaactgc acgggaagca 360
gtgacgtagc gcccaaccca ccataatcga ggagcaagcg gaccgtggat actgcagcga 420
acgagacgac acccaaggct ctgtgcttct ggctcggcgc tctgagtgat gccggattca 480
agctcctaca tgagctgagt cgccaactgg cacggtgcca tgcgacgttc gtgcaggcgg 540
ccctggagta tcaggcagac ctgaacgctg cgaacaacat cgccgggcgt gacccggcga 600
cagactttga cacacccgat cgggctgcgg ttcggtcgga cgggccgaaa tcacggcccc 660
aggcggggcg ggcgtcgcga ctggtcgcgg cgtcaacccc gtaaccctgc gcacctagac 720
atcggtgttg gcacatttct ggtctgcgca gcttgcacgg aaagcagttg acgtggcgcc 780
ccacccaaca cacaggccat tccccgcgtt agcgggttgg caggcaaggc gtgcaggggc 840
cgggctggac acccggcccc actattccac aggaggcgga cgatgccagg cacggagacg 900
gtgacggtga caatccctgt ggactacgtg acggcccgcc ggatagccga gctggagagg 960
aaactctccc aggccagggg cgaggccaac gaccaggagt gcgacgccac cttctggcgg 1020
gagaggtgcg ag 1032

<210> 809
<211> 3022
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
解淀粉嗜盐碱球菌 DSM 10524 序列
<400> 809

gcggaaccgt gccgtttcat cgtcgacgcg aaacacgact aacccgcggc cgcctgcgtg 60
aagcgcggtt cgctcgggta atcccgacac cggctccgtc ggctcgatcg ccgccgggtc 120
aactacctga atctcgtcgc tcatgttcat tgtaccgatg gggtcgggtt acagaggttc 180
gaggagaaag cccacgactt cagtcgtggg atgaatccga tcaaatacga cacaatccat 240
cctcaatagc cgaccgaagg ttgatagaat cataacatct aatatagatt gacgtctata 300
taacagtatg ccacgggggt acagccgaga gcgaacgtcg gttcacaacc tccactacca 360
cttcgtgtgg tgtccgaagt atcgtaaacc ggtactgaca gacgaggttg ccgaccgcct 420
cgaacacctc atcgaggaga aagccgacga actcggacag gtcaagggct acacctcgcg 480
gaaccttcgc gacgagtttg atttcggcct cccgtcgctg tggacacgct cgtacttcgt 540
ctcctcggcg ggcgaagtgt cgagtcagac catcgaggag tacatcgaag cacagactgg 600

[0358]

gcaatagcga tgcaccgcga agtcactacc acagtacgtg tcaaactcca ctcgctcacc 660
gagcggaaag cccgcctcat cgaacgcgag tatacggcgt tccaagacgc cgttcacggt 720
ggcgacgcga gcctctactc cgccaccaaa caacaggcgg acattcagag cccgctatcc 780
acggtggggg ctgccgtggc tcagcccgaa acgcaggtcg tactcgacgg aaccagcggt 840
gagatggaac ctgcgaactc cccgaaagac gtgggcctaa ccctcagtga gggaagtcca 900

cgactttagt cgtgggagaa gtcacgtctg tcggatgacc cacgggctcg cggttcgagc 960
cgtcgctccg ccgctatacc tcccagggtt cgccggtcgt ctcgccgaac agctcccgca 1020
tgagcgtgac gatgctttcc ggcgggaact ggccccgttc ggtgacgatc gcgtcgacgt 1080
gacgcggcgg cgtgacgtcg aacgccgggt tctccaccgt gagtcctcct cggagtcgtc 1140
tatctcgtcg tcgccgtcgt catccgagtc gtactcctct ccgtcgtact gctcgtcgtc 1200
tccgttttcg ctctcgtctt cgtcgtcggc caccccttcg ttgtcatccg cttcgccctc 1260
ctcatctccg tcgtactgct cgtcgtggtc gtccctctgg tcttcgctct cctcctcacc 1320
gtcgtcgtca tcctcgtcgt ccccatctcc gcgatcatcg tccgccgttt cctcactttc 1380
gtcctcctct ccatcgtctt cggcatcgtc gccgttttct ccatcacctt cgtcaccctc 1440
gtcgtcctcg ccagtctctt cgtccccatc atcctcctcg cgaagcgcga cgagactgtc 1500
ctcgtgttcg acgtagagcg tttcgtcgct gatacccagg acgttgccga ctggcagctc 1560
cccggcttgg atcgcccagc gctcgcttcc gtcgtccttg tcgagggcga cgagttcgcc 1620
gtcgtattct ccgtacgtcg catccgtccc cctgaagtag acgtagacgg tgtcgccggc 1680
gatgatcgcg cggccgacgt tgcgagacac gtcgacaccc cagtcgtcct cgtcgccgtg 1740
ataggcgctt cccgtgagct ggtcggcctc gtacccgctg aggtagcgct ccttgccgag 1800
cgtcaccgag atcccctccg ccgacacctt cgcaatatgc tcgccggtct gtgcgtcgag 1860
gagcacccgg gcgtagaaat cgacccggct cacgacgacg gcggtcgacg tcgcccgtgc 1920
ctgggcgcct ccggtgtaaa aattcgaatc ccctcgccag acctcatcac ccgtatcggg 1980
gtcgagtgca accgttgcac cgccttccag atcccccgcg ccggcgtaga cgacgccgtt 2040
cgccgcggcg gtcgtatcga ggaagtcgta ctcctgttcg ccctcgtcat cgtgttgggt 2100
ttcagcggtg agcgtgtcga tctcccatcg caccgacccg tcgtcggtcc cgagggcgta 2160
cagcgtgccg ttgacgacga cgaagacgcc tccgtacgcg acggtctgac tctgaatcga 2220
atcctccggc tcgaactcgg tctgccaccg gacgcttccg tcggcgatgt cgagcgcgaa 2280
cagttcgtcg ccggtaacgt agacggtgtc gtcgacaacc gacggcgcgt tcgcgtcgac 2340
gtcggtgttc tcccagataa tcgaaccgtc ggcgacgtcg agcgcgacga ctccgttcgc 2400
attcgtcgcg tagacgacgt cgtcggcgac ggccacgccg gcgtctcgtc cgaagccgaa 2460
gtcgaccgac caggctgcct cgggcgcgtc gccgtcgagc gcggtcccgt cggggacgta 2520
cctcgcgtgc ccggcgtcgc cgtcgaacga cggccagtcc tcgtccatcc ctgggttcgg 2580
atcgcggtcc gaatccggta gatcgtcggt cgactcctcc ccggcaccga tcgacgcgag 2640
ggctcccccc gcagtgagcg ttgcaccggt cgccagcacc gaacgccgtc gccacttcgt 2700
catacgggtg tcgacgacag aatatcggat tactatccag ttgctacggc tgacttgtcc 2760
ttcctgcaac tgtttcagac ggacgagcac cgagcttact cgacggaacg gaaccgacgg 2820
catcgaccct cgaagacgga ctcgcgggct cgggcgaagc ggcttccggc cgttaccact 2880
cctggccggc gggttcggcc tgccgtggtt ccccctcgtc ctcggacgtc ttctccccgc 2940
ggacgacgaa gtacgcgatt gcggcgacca cgcccagcgg gaagaacagg aacaacagcc 3000
cggcggtgcc gagcgcccag ag 3022
<210> 810
<211> 3403

<212> DNA
<213> 未知
<220>
<223> 未知的描述:
水生-非海生盐水和碱性-超盐水序列
<400> 810

[0359]

gcagtcgttg catcgatctc ccgactcgca gctggcgaga tttcgtattc ggcccgtcca 60
tcctcggcag tgaggttctg tttgattagg tcagcccgct cgttgaactc gacggcggtg 120

ctggtgtcgg cgtcgtattc cggatcgcgg aagtattccc gtggcgtttc gggtttgcgg 180
taggcggcgt tggggttgtc gcggttccac tctctgtttt cagggatctc ggcgtcaagc 240
tgttggtgga agaactcgat agccgctgcc attgccgcct tgatccggtg ttcttcgggt 300
gggtcgacgg cgagttcggc ttcgagttcc gagtagtctt catcaagata gtcctgtttg 360
gtcttggtgg gagtcgtgtc tgtctcggtc ccttcccccg tctctgattc agtctccgtc 420
gactcgtcgg tgggtgttgc ggtgtcgtct tcgtcggcgt cggcgtcttg atcgtcgcta 480
gttgctgtgg tttcggcagt cccgtctgtg ttggccgctg tctgcggttc gtctgtgtcg 540
gtggtaggat gtatcggctt cggcgtccga ttgctcacca gccgattctg cggcgtctgt 600
gtcagcggca gtctgtgcgg gggcttcgtc tgacgtgttc gaggtgtcgg ccccatcact 660
tccattttca ctctcagcac tttccccacc ttcctcaccc gtggttgctg ggctgtcgtc 720
tgcagtctcg gtgtcgattg attgctcacc gtctccgttg tcgcctccat catcgtcatc 780
gtcgttgtcg tcggtggttt caccgccgag tttccgggtg atctctgggc cctcgtctgt 840
ggcggggtcg ttgatcacat actggccgtc tgactcaaac gccggagagt agatctcgcc 900
tcgtgagact gcatcctgga gtatctcttc ggctttgccg agtggcttcg ggtctgctga 960
ataagattat ctcatgtatg gtggttgcga ctttgtatgg cttcgaacgc ctgagtatat 1020
tatttttatt tgttaatatt gcagttgtaa tataaataag accatgaaga atacgatact 1080
agagataatt acgacgcgtc gaaaccaacc aacacggaag caccagtatt gctctcgatc 1140
cgacggatca ggatactggc acctcatcta tgagtgggta gacggttcgt ggcagatcgt 1200
aagccgcgac ccactcgacg agatcgagat cgacatctcc aactcgcgta gtgattgcac 1260
accaacccca ctcgttgagt gtcctgtgtg tggtcggact ggacttcccg aacggatcgc 1320
ccagcacaac tgtacaccaa ggtcgacaca gtaagctagg agtcaaagat acaaccaata 1380
taaacacacc aagccattac caacaaaaat atatggttac gggtgttgta taactagtac 1440
tagttgttat tcacgcatct ccgctcgctc tcgtaggctg ttttcggtgc tatcatagag 1500
tataaattgg ctatcgactc atctggtgag ttcctcagtc gaccgcccga cgcggcgatc 1560
agtggtgaga ggaacagcaa gacagcagaa gagtgagatt gagaactaag atagagcacc 1620
tcgcacctcc aaacagatgg acgaccagtt ccgaacacgc tcgtgtttcg tcggaatcgg 1680
cgccaccggt gccgagcgca tcggtgacgg catcgaacgt cgtgacgcaa gtgccacgat 1740
ggttccggac gaagttccgg acccagccga cgccgtcgtc gactgcgttg ccgcaccgag 1800
ggcaggttgt actctcccag aactgcttga gtcgccgctc gatggccgca tcgagcggct 1860
ccgtgttgat tgtgatcacg cccagatcga caagctggcg aagcatccgg ccgaaggctg 1920
gatctgagga ggacatactc cctcatttga caccgtagac cgtagctata tggactttcc 1980
acgtttctac ataagagcgc aacaggttaa ccaacaaatc tatcttatac tcgttttgtt 2040
ggttaacaca aggcgctctt atgtgaacca cctcgcggtc aagccccgag gcactcgcct 2100
tgcttatctg tagacgtgga aatgttcaaa aagtggttcg ataccctcaa gagatcacca 2160
gccggccaac attcgacgtc ccatacaggt actaagtgac gctagcactg attgtcagaa 2220

aaaaatttat gaaacggtgg tgtggccctc cgatttaact cttgcaaggc taagtccccg 2280
tcctcaagga gcaacgcagg gagcgtagcg accgagtagc gcagtagggc ggggatacag 2340
ccgacacgaa attaggcacc ccacgaaatc gcacacccta accaaagaat tacatgattg 2400
tcacccgata gtggaagtga acccacaatg gaatacgacc tcgatacggg agcgcactca 2460
acgtactccc tgcactacca cctcgtcctc gtcgtgaaat atcgacgcga ggtgttcacc 2520
gaagaacttc atgttcaccg aagaacttca agggttcttc cgcgaggtcg tgagcgggtt 2580
cgctgacaac tacggcgtcg aaatcaaaaa cctcgaagcc gaccgaggcc acgttcacct 2640
cctgttcaaa gccacgccca caaccgacct caccaaattc atcaacgtcc tcaaaggcgc 2700
atcggctcgt cgcatccgca atgagtacgg cgagcaaatc ggggacaaac tgtggggtga 2760
ctcgttctgg gcagattcgt actgtctcat ctcgacggga caggtatcgc tggacatgct 2820
gaaagaatac gtcgagaacc agcgggagcg tgacgacgaa tgaaacagac tgccgagaaa 2880
acggtcgtgg caaaaatctt caccaacgac ctaacgcagt cgaaacttga cgccatcaac 2940
cacgagtatg aggcgttcca gcgatacatt cgtggtgacg aggaagcaga tctctactcg 3000
gccacgaaat tgggtaagga catcgcgtgc ccactatcca gtctcggcgc agtctgtgaa 3060
cctgtgctcg gaccgagcgg tgctgaccaa gcctctgcca cggtatgctt gcgagccgac 3120
ttggaagccc ctgcctcaac gaagcgaacc gttaggtgag cgaagtaggc aggggtagtt 3180
cacttgaaat tgctgaaatg gtgggggtg ttgaggagcc taatattggt tcaagacccg 3240
ctcaccaaca agactaggcg ttcggtcgat cagcactgat gcgctcaaaa tcgatatctg 3300
cgaatcgagt cgtatccttg ataacttcga ctgcgatctc ggtatcaact tgcagctggt 3360
actcccatgc tggcccaccg atcgaacctg ccgtcacctt ttc 3403

[0360]

<210> 811
<211> 1156
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
水生-海生-海生沉积物序列
<400> 811

ggcaggttgt attagacgga ttttatttaa aattgtttct ggtaggttgt tttctctggg 60
tcgttgtgaa ggttggaaaa tctgttaggg gtttataaag atcttcacgg tttggttggt 120
ggtttagaaa gtgttaaaaa tcttataggt ttagatttct ttactgggtg gtgtgtgttg 180
aagacttttg tgatctatga tatctccgat gatcggatgc gggaggggtt gagggagcat 240
attaggaact atggtcttcg gcgggttcag tatagtggtt tcctgggggа gatgagtccc 300
catgatcggt ttgtcctttc tcgtgaggtt gatcgtgagg acctgagggc actttacagc 360
gatgtgatga aaaccttgca tgatgaggcc ttttatgagg gcgtttataa cccagaggag 420
gtcgagtatg ctataaaaca ggttaagaag atgattgaag aatttaaagg aagttaaaac 480
cgcatattgt taggattctt ctttttacct attctcggcg agaaatcctc ccactttagt 540
gggggggatga atcgccgttt aagtttaaat atcagaaatc ttatatttgt ttcgggcgag 600
tcgcccatga aggttaagaa gacaattaag gctaagattc ttgagcttcg gaagggcaag 660
gaagagcttt tgaaacgtga gtatgagaat tggcaacgct acctaagagg cgataaaacg 720
gttccgcttt attctgcgac caagcaacag gcagatatac tcaagcgggc tatggggcta 780
gtctccgtag caggggctgg attgacacag cccagaactc ggtatgatga agccagaagg 840
ccgaaaaacc gagaatctcc ccacttaagt gggaagagtg tcaagagtta ttccttcaag 900
ttctaagatg agtttctcgc ataggttgta ggtcttcttg gctctgttgg tggcgtcctc 960

[0361]

tgcatcgttc ttgtccatga cctcgttggg tggtgtgagg gtcttttcat ccccgtatag 1020
gcttggctcc cttttcttga agaggagttt ggaggattcg cgtaggaagt ctatctcggt 1080
tttgaaccag tctgggaatc agtctctgat gtcggagaga aggtctgata cctcgtggat 1140
tttggggtac tcgatt 1156

<210> 812
<211> 1589
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
特班齐赫盐红菌 DSM 14210 序列
<400> 812

cccttccggt cagttgaacc gtagatcgaa gcggaaaatt aagtcaagtc caaatgttgg 60
tcaccgtcga tactggcagg tagaacaggt ttatagccgg atattgatgt ctacacaccc 120
cgttctagtt atggaagcga cagcactgaa cgctactttc cttcgaagaa atcatctatt 180
tgtgcgttga caacagattt gattagatcg ccttcattgt ctgcctcttc aatccggatg 240
tcagctcgaa gagtttcagc aaccacaccc ggatcatcaa caaacgtgta ctccttgtga 300
acgacctcgt cacggatctc ggcgagtgct tcgtcgtcga cgccgtggtc gtgtaatctc 360
gattcgaggc gatcgatcga atcccgcgcg agtgccgcct cctcgtcctc atcaggccga 420
tacgtctgtg gatcgcccat gaagtggccc atccgtcggt ggacttgaac ttcgaggagt 480
gtgggtccgt tcctgtcccg ggctctgccg attgcgtcag cggcagtttc gtagatggca 540
acggcgtcgt cgcggtcgac gcgaacacct ggcatatcaa acccggctgc tctgtctgac 600
ccgtcttcta cgtccgttac agcggtagca aggcgaaagc ccacccgttc acgggtggga 660
tgaagccgac aactgggaat gtatccacgt tcgtagccac ggcttggggt tgatatttgc 720
gtaacatctt tcgatgagtt acttctatat accgacgatg cctcgcgggt tcgaccgaga 780
acgtaccaac atccacaaac tccagtacca cttcgtctgg tgtccgaagt accgcaagtc 840
ggtactcaaa ggcgaggtac gccatcgact cgaagaactc atcgaggaca aagctaacga 900
actcgaccag gtcaagggct actcctcgca tcacctccgt gagaagtacg actttggctt 960
gccctcgctg tggacgcgct cgtacttcgt ctccagtggg ggtgacgtct ccagcgaggt 1020
catcgaggag tatatcgacg cacaagcggg tgaataacga tgcaacggaa cgtctcaacc 1080
acggtgcggg tcaagctcca ctcgctgacc aacagaaaat ccgatatgct ggcccgtgag 1140
tacgaggcgt tccaacgcga ggttcacggc ggggatgccg acctctactc cgcgacgaaa 1200
cagcaagcgg cctcgggcaa gttttcgaaa ccgctgtccg aggcgggggc tgtactggca 1260
cagcccaaaa cgcaggtcat cgtcgaacgt gacgaggaac ctgcgaacct ctcaatttct 1320
gtgggctcaa cccccagtgg aggaacccca cggctttagc cgtgggagga tgtcagacgg 1380
tctttgggca tgctgattgc ccagtcatta tcctcgacga cgaatacaac tggtaggttg 1440
tggacactcg cgaaattcag cgattcgagg aacgcaccct gatcgatcgc tccctcaccg 1500
aggaacgcga ctgccactgc gtcagtattg cgcttttgg ccgcaagccc tgccccggct 1560
gctggtggac atccctcagc gatgatgcc 1589

<210> 813
<211> 1180
<212> DNA
<213> 未知
<220>

<223> 未知的描述:
水生-海生-深部地下序列
<400> 813
cgagaaagct aaggaagcca gagaagcaga gaagctagct cagaaacaag ccgaactgga 60
ggctaaagaa aaggcaaaga gagaggccga ggaggcaaaa agagagactg aagaagctaa 120
gaagaccaaa gctgaaagga tgctagccga gaaggaagcc gagataagcc gtctggaagc 180
ggcattatca cgcggagata tacgactggg taaagttatc tatcattcct ggaagctggg 240
ctcagcacga gccactaagg aaaaggccag acgagacgcc gagaaagcca agcaggccaa 300
ggcggctcag cagtaggccg agcgggaagc ggagcggggc gctaaggaga aggcgaagcg 360
agaggctgaa gaggctaagc aggcactaca atcgctgata gattttgacc aagagaaaca 420
aattgaagaa tgtctaaaag aggtgaagag catttgtgaa gaactaaggg cagggaaagc 480
tgagacaaag gaagcaatac aaaggctttt agatatttca gaaaggattc ccgagtagat 540
attccccagc taatatcagt aggctgatgt gactgggctt catccgtggc aacctatttg 600
agtttgataa ttcactacct cagtgtacat tcaggcggac tcctaaaatt acagaatcag 660
gctgagaaag ccccctgctt cagcgggggg atgaatcagc ccttccctcc tttcccaaag 720
tataccaata tatgtttatg ggacaagaat atagacataa cagacattcg atctctttgg 780
tcaactatca tctcatctgg attccgaggc gtcggagaaa agtattggtg aatcaggtta 840
aggatagact tgataaactg cttaggcaaa aagctgaaga gcttggtata ggagtgatag 900
ctttggagat aaacccagac catgttcatc tttttgtgaa tgctatgccg gatttggcca 960
ttcatcaagt ggcctttcct tatcgcggtc tttttctttg ccattcttgc aaactccact 1020
ataatgccga ccttaatggt gcccgtaata ttctaaagag ggctttggac tatatgtcct 1080
[0362] cagccggggc tgctttggca cagcccgtaa ctcagtctct gaactgagaa tcctccgcct 1140
tcagacgggg gagtgtcaac tgcaggaaca ttactggcaa 1180
<210> 814
<211> 198
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
宏基因组序列
<400> 814
gggcagagaa ggtagatgca gtagaggcag gagaggattg gcctactcat gattaagaca 60
gcgacattct acatcgccga tatgccgcag gaggacaagc gcctctgcga acgcctttgc 120
gcagaactcg gagcggcggg cgcgacgatg gttcgcgagt ggtatcgcgc ctgtgacgaa 180
gtgaccggcg aacgtatc 198
<210> 815
<211> 3685
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
水生-淡水-沉积物序列
<400> 815
aagctgtttc acatggcaag agatttttga ttcgagtcgt gcggccaggt cgggctgatt 60

[0363]

gaaaagaagg agatgggttt ttatcgctaa cttttacctt aaaactgctt cacaaccgaa 120
aatcgcttaa ccaaagatgg cgaatagaga actggtggtc gtgatgcagg tgcagaaagt 180
cgtcaagggg aaggtcgtgg ggctgactca cgctaaacgc cagatactcg accacgaata 240
cgaagggctc cagcgattct tgcaggggga ttccaacgtt gaactcgcct tcaacatagg 300
cagacgagcc ttggggcaag tctccaaggt aggggctgtc gtgagccagc ccgtaactgg 360
ggcagtgccg ttcagcgaac tcggtagcgt cgttcaatgt ccgaccccag aagccaccct 420
ttttaaaggg tggtagttca cgtgcgcgag cgcctgcggc gtttagtcca atctactaat 480
aggtgctctc ccatctccct aaatttcctc gccgctggag aacgcggggc atagcgaaga 540
actgggctgc cagcggtcaa gctctcacga actttatcgt catcgggttg ccactcaatt 600
ggtgcttgga cagccccagt ccatgtttca atcagattcc acccaattgg agcttgcacc 660
gtgccaatcc aagtttcaat cagttgccat gcggcagtga ccgtgactgt cccagtccag 720
gtttcgatga gtttccattc accagtggca gtgactgttc cagtccaagt ttcaatctcg 780
taccactgag cagctgctgg tgcttggacc gccccggtcc atgactcggc gagttgccat 840
tcagccggag cacttactgt cccagtccaa gaatcgatct gctgccactg tcccgtggtt 900
gcattaacag tcccgctcca agactcgacc tgacgccact gaccagtggt cgcgttgacc 960
gttccagtcc aactctcaat ctgccgccac tgtcccgtgg tcgcgttcac agtcccagtc 1020
cagctctcaa tctcgtacca ctgaacagct gctggtgctt ggaccgtccc ggcccatgac 1080
tcgatgagtt gccatgcagc aggggcggtc accgtgcccg tccagctctc gacgagctgc 1140
cacgcggccg gggcgctcac ggtgccagtc catgtttcga tgatctgcca gactgcggaa 1200
gcattgaccg tccccatcca cgactccacc aattgccatt cagctggcgc ctgaacggtt 1260
ccagtccacg tctctatgag ctgccacgct gcaggtgcat taatagtacc cgtccatgtc 1320
tcgatcactt gccagacggc aggcgctgtc acggtcccag tccaggtttc gattgtttgc 1380
caagccgcgg gtgcgctgac cgttccagtc caagtctcga ttagctgcca ttgggcagga 1440
ggtggaacca ccgtgaactg ccaagtgttt tcagaaacat tctcacccgc agcgttgacc 1500
gcccacactc tccagtagta ggtccccggg gtggagaagt tgcgctggta gatgttgtcg 1560
gtgttgtctt ggtagttgtc cacgagcggc gtggagaagg tgggctcgtt gtcaatctcg 1620
atcctatggt tctccccaaa ccacccgttg tcccactcga aggtcacagt ttccccgacc 1680
ccgaagacag agttgttctc tggactgtag agcctgggcc tcttcggaaa gaatgtgtac 1740
tcgtatgtga aaaagatggc accggtaaga gcggcgcctg ttcccatgtc gtattcaaag 1800
ttgtaggtca ggtcggcaag ggtcctatct acaacaacgt cgatgttgtc tttgtttaca 1860
gagatggatg tcaaggcctc cccagagctg cctgtgctca tactgacggc tgagccccag 1920
tttaccccat cagcttgaat cctggtctgt cgcgtttccg aagcggttgc acttcctatc 1980
caatatgccg tgaggccaac gctcttcaca gtctttttgt tttccggggc atagaacgtc 2040
agctccccgc tggtccaagg gtcaccggca gccccgctgg gtttttggcc cgcggtgaac 2100
atcgtggttt tcagctttgt cgcggccgat gggcggaagt gataggtgat gtgcaaaagg 2160
ccgcttacct catagaggtc ggaggccgca gcgctctgga tgttgaggtt cacattgtcc 2220
ccgttcgacc agttctccct catgaaggcg ctgatgtcgt ggaagtagtc gaacccctct 2280
acatgctcgc tgcctatgtt ggtcgtgtaa gctctcaagg tctgggtttt cgccccgacg 2340
ctcgcgtaca cgctcgtgct gccgtcacta gtcaagatgg tgtggaaaat ttcaaagtac 2400
gcgttctcga tgttgtccac gcgagagctg tccaggtccg tcagataaac gggcgtcccg 2460
aaggcgaaca ggttggccga gccgttgacg ttggcggccc ccatggcaaa cgtgactgtt 2520
tttatctggt ccgcgctgtt gtctgaatac tcatatgtga cgtaagtttc acagctgagc 2580
gcatcgtttg tttgtcctcc ggtagtaatt gtaatatcct gatcaacgtt gtagttccag 2640
gaagccgtcg gccgggaaat ataccagaaa ttgagatcgt tgccggtacg ggcttgctct 2700

gtccaagcat ttgtggtggc tgtatcgctt ccgatcttga tgttaagaga aaagtcggtt 2760
gaagaggctt gggaagagtg ggcatggatt tcgtaccagt tgtccctcag cgtgaccccg 2820
tcttctggga tttcgattct gtgtgtgtag gtcgccgtgc cgccagcccc tatctcggtg 2880
ttcgtgctcc ccacaaaata cctgacagtc ttgatcttcg tgtccgccag ctcgttctcc 2940
tcatatgtga tgtacatctt ggccgtgtgg ctctgggtat ccgcgtcagt gacggttaat 3000
tttgggtaga cggcggcccc gtttgtccag tatgttgtaa aagcagaggt aaaatctgcg 3060
accacccatg ccgccacgct ttccccgctg ttggcggtgt actggccggt cactgggtct 3120
gcgtctgggt tccggtcggt ccccgcgatg ttcagggtga cggtcatatc tgtcacgttg 3180
gcagcagtcg atgcccccgt ctggacgtga tacatcacga acaccttcct gaacggccca 3240
gccttgtccg ggatgtagat ggtttttgtc ggggtgaacg agtaggccgt gccagtgccc 3300
aaagcggacg atgaggttcc ttgcctcgta tcccatgaat attggattgt gtttatggtc 3360
actatatttt cgttgggttt ggtgggggag ggagtagcca cggctagggg ctgcatcgag 3420
ataacgaaaa cgctggctaa catcatcgct atcaatacga atacaaagat ttgttttgcc 3480
tccttccgcc cgaactcttc ccgccccaac ctaaacaagc tacaacccat ctccctcacc 3540
tgcttgcctc cacttggact atggaatggg cgcggctcac ggcagagcca tctctcatct 3600
ctctcagcag gttcgtgaag tactcaatcc cgctcgcggt gagcttgcag agctttcgag 3660
gcctcccccg ggttaaacgg ggctc 3685

<210> 816
<211> 1615
<212> DNA
<213> 未知
<220>

[0364]

<223> 未知的描述：
盐红菌属物种阿特里特-26R 序列

<400> 816

gcgaagcgaa cgaacgagga gcgcagcgag tgtgcgccgc cctcgcggct ggggctttgg 60
cggtgttttc tgtcgatccg ttggcggcta tttataaatg agcagctggg gctttggcgg 120
tgttttctgt cgatccgttg gcggctattt ataaatgagc agctggggct ttggcggtgt 180
tccccgccga tccagaatca gcaatttata agcgagcaga tagggatata gcggtgctct 240
ctgccattcc gcagtcagcc atttataagc agtcggcggt agttcagtac gtctctttcc 300
cccaccgcag gtcgtcgtag gtggcctgct ggtcgctgcc gagccgctcc tcgaacttcc 360
ggcggagcgc ggccttctcg gtcgcggaga tccgcgccag ctcgtccgcc ttcgaaaccg 420
gctccggcgg gcccagcccg cgaccgacgg cgacgtcgac gaggcgcttc gcctggtcgg 480
gcgagaacca ctcccgctct aacgacagcg cgacgacgaa ctccccctcc ccgagccggt 540
cctgggcccg ggtcttgaac ggggcggcga cggcgacctc cagactcatg tgcgaatcgg 600
ggggcgccga cggcggtaaa cctaccggac cgcatcccgc ttcagggcgg ccgcgacgtc 660
cggcttatag cggtagcgag gcgaaagccc acccgtttac gggtgggatg aagtcgacaa 720
ctgggaatct atccacgctc gtagccacgg ctggggttgg atgcttgtgc aacatattta 780
ggtgaagata ctctatatac tgacgatgcc tcgtgggttc gacagggaac gtaccaacgt 840
ccacaaactc cagtaccact tcatctggtg ttcgaaatac cgtaaatcgg tgcttgaggg 900
cgaggtacgc gaccgtctcg aagaactcat cgaggagaaa gccgacgaac tcgaccaggt 960
caagggctac tcctcgcgcc gacttcggga cgagttcgac ttcggcttgc cgtcgctgtg 1020
gacgcgctcg tacttcgtct caagcgcggg cgacgtgtcc agcgaagtca tcgaggagta 1080
catcgacacc caagcaggtg aatagtcatg caacggaacg tctcgaccac ggtgcgagtc 1140

aaactccact cgctgaccaa caggaaagcc gacctactcg cccgcgagta tgaggcgttc 1200
cagaccgcag ttcacggcgg agacgctgac ctctactccg cgaccgacca gcaggcggcc 1260
ttgggcaagt tctcgaaacc gctgtccgag gcgggggctg tactggcaca gcccgaaacg 1320
caggtcatcg tcccacgtga cgacgaacct gcgaacctct ccgtgtccgt gggttcaacc 1380
cccagtgggg gaaccccacg gctttagccg tgggaggatg tcaggcccgg caaccgttat 1440
gtcgggccag cgagatggcg cacgtatgaa gacccgcccg ctcggcgaca ccggccacga 1500
cagcaccgtc atgacgttcg gcgcgatcgc gctcaactgg ctcgaacagg agggcgcgaa 1560
ccagctggtc gaacacgtcc tcgactacgg ggtcaatcac ttcgacgtgg ccccg 1615

<210> 817
<211> 2223
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
水生-淡水-淡水湖沉积物序列
<400> 817

[0365]

tccggggtgt atcttgcttc tggtttgtat gtgaatgcag ggaccatgtc tgcgaatgtg 60
tctgggaatt ggattgaggc gcatgtgatg tctggtgctt tgaccatcat ctctggtggc 120
gtgtttgtct ctggttctat cacagtgatt tctggtttga ttggtgcgta tgtttctggt 180
tttgtcactg tgcagagtgg tgtgattcag gttctatctg gtgtaggtgt tcttacttct 240
gtttctggga atattgtttc tattgcttct gggttgtatg tgactattca gtctggtggt 300
gctgtttctg ttcctatcac ttctggttct gtgattgtgt ctgggaatat tggtgtctct 360
ggctgggtga gtgtgcagag tggtgtgatt caggttctat ctggtgtagg tgttcttact 420
tctgtgtctg ggaatattgt ttctgtcgct tctgggttgt atcttgaatc tggtcaagag 480
attatctggc gtggtgtgca agtctctggc actgtcgctg tctctggttt gattggtgtg 540
caatctggtg aggttcatgt gatgtctggt gctttgacca tcatctctgg tggcgtgttt 600
gtctctggtt ctatcacagt gatttctggt ttgattggtg cgtatgtttc tggttttgtc 660
actgtgcaga gtggtgtgat tcaggttctg tctggtgcgg gtgtggttgc ttctgtttct 720
gggaatattg tttctgttgc ttctgggttg tatcttgctt ctgggttgta tgtgaatgcg 780
gggaccatgt ctgcgaatgt gtctgggaac tgggttgaga cgcatgtgat gtctgggaat 840
cttcaggttt tacagacatc tggttttgtt gctcttcaat ctggtgaaat tcatataatg 900
agtggttcag tgactatacc cgcattgaca acagttgttg ccaccccagc aacaaacccc 960
atgctgatta ccggtctttc aggaggacaa attttgactt caagtgtagt aataagcgct 1020
attgtgaaag cagtatcact aaacagtagt gatatgtata ttggtggagc atcgggggta 1080
aatagccctc cgtattctgg tggagggttt gtgcttgctc ctggtgaagc tattaatctt 1140
gatgtcgcaa atatcgggta tgtgcgggcc tgtgcagtga tgagtggtga taaagtatca 1200
tggattggga tagtataaga ggcaaatatg gttaaagaat ttctctttga acagttttac 1260
acagggacaa caggtgtcga tacagcgtat ggagtaaatt ggagggcaac gacattcacg 1320
atagggactg ttggcgcgaa agcgtctcat tatgcgacaa gcataggatg gcgagggcaa 1380
gattttccaa ccattcgttt cagcgataat gttgaaacag tccttgagga aaatttcaat 1440
atctccgacc aagacttgtt ttctatattt cgggcagaaa acaacatggt aatttagatt 1500
gtatactgcg gaacttgacc tgcccataaa tcttaaatac ctataggtgt atataagatt 1560
tgtggcacag aaaaccatca ccgcaaaaat cgtagcaatg accgctacga aacaacaaat 1620
actccaaacc gagtatgaca acctccagca ttttctacaa actggagaag accttggagt 1680

atactcagcg aacaaacaac aagccaaaag aacaagtttc aatgccgact acaatggtgc 1740
attaaatatc ttgaagaggg caacatccta catggatatt gctggggctg tgagtgaacc 1800
agcccgtgac tggggcaacg acgaccagac ctcagaagcc acccaacttt ttgagtggta 1860
gttcactaat gtaactggga ataagaaaat cattgaagta tttggtgaac gatggcacac 1920
tggggaaaag attccatata gacgaactga gattgggagg aaaaaaatat tcgaagaaaa 1980
tgggtatgct cttttgataa tatgggatta tgaattaatt gatgaacaaa aagttataga 2040
aaaaataaaa gattttgtct aaattagaca aaattaaatc cttcctgttt cgagtactat 2100
tgttgttgtg gttgtgattc ccgagtgagt atatgtttta gtaacagttc cattcacagt 2160
gacgctttga ctccaatatg aaccagcgag ttcgattcca cccacgagca gatgagcgta 2220
gag 2223

<210> 818
<211> 2586
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
方形盐盒菌序列
<400> 818

[0366]

taccccaaca ccggcgccaa cagaagcacc aactgcgacc ccgacgccga ctccaactcc 60
agatccagta atcgttgaag tcggcgaatg ggggaagata gagcagaacg gtacagtggt 120
tcggtttcga gtggttagct atgacatcac ggaatctgtt gaagaagccg atggagagac 180
agaagagcca tctgaagggt ataaatttgt agttgcagac gtacatatct ctactgatcc 240
cggcaatgaa gttacgattc gaaaagacca gtggtcgttt actgaactta ccgataagaa 300
gcgcggccct gatggcgtga ccaacggaat tagagagccg ttcccaaaca agactgcatt 360
acagggtgat tcggttcgag gaactattgt ttgggaagta aaatatccgt ctgatgggca 420
gttcgaggtg gagatacata aaaacaaaac aggcagggct ggtaaagtgt taatataaca 480
gagtagcttg ataagagttt tacatattat cgtggataat ttatgctcga aaggggttca 540
aaaacagttc tgctcggaga taatttttga gggctgactg tagttgtgga agtggtccat 600
cgtgccgggt gatgctctgt ataattatac catgtagcat cgctaggaga aattcagcgg 660
tttcttcggg ggagagagtc accgccgaag cgaccggagt tcccggcaga aggccggtga 720
cgatgactcg gctatggatg gggcctcttt gatcgggtca caagacagcc acgctgatgc 780
tgaaccccag cagacgacgc gtggaacgta tgcgtcttga aaccaacagg ccgctacact 840
cggcctgaaa tcccgtggtg ggactcctcc acgttcacgc ggaggaggag gtcaatttgc 900
cgatcttgtt tgcacgcatt cagtgagaca atgcgagaaa ccgcccggtt gcagtccaaa 960
ctacgagaag gtcgatacaa tagcgagcgt atccggcggt tgtaccggaa acagactcgt 1020
cgtcacgaca cacccgagga gcactgtgtc atgacctgct ggaacaacct gtacgccgag 1080
ggtatcaaca tcgtgtatat cggtgacttg accgacgtgc tggatacgca ctggtcggtc 1140
gaaagcaacg ccaagaccca gaacttttgg gcgctcaaga agtttcagtc gatcacatcg 1200
aaagctcaca aacccttcca gtagaagttg aatctgtgtc tacggtgcct ggccagttct 1260
cgaagtgaga cggacagtga gttttggcgg tgaggagtcg ctgtcggcgg cggatcgccg 1320
ccgacgcgac agtgtcgcca ctcacgcccg ctgtcctagc tgggacaact acctgaagaa 1380
ccggtcgtca agtggttggt tcgctggaac agacctgcgg acaccaagtg ccgtccaagc 1440
tgcgcgaagc gtcatttcgc agaactcggt gaacgaccac tgtaagaggc ggcgcccccc 1500
gcggcggggc gccgccagac tgcacatcct ctccaagcag acgcttactc taacccgatg 1560

tgatcgactg agtttacagc ggtagcgagg cgaaagccca tcggctttag ccgtgggatg 1620
aagccgacaa ctgggaatct atccacgctt gtagccacga ctggggttag atgtttgcgt 1680
aacatcttta gctgagtgat atctatatac tgacgatgcc tcgtgggttc gacagggaac 1740
gtaccagcgt ccacaaactc cagtaccact tcatctggtg tccgaaatac cgtaaatcgg 1800
tgcttgaggg cgaggtacgt acccgcctcg gggaactcat cgaagagaaa gccgacgagt 1860
tcgggcaggt caagggctac tcctcacgcc gacttcggga cgagttcgac ttcggtttgc 1920
cgtcgctgtg gacgcgctcg tacttcgtct caagtgcggg cgacgtatcc agcgaagtca 1980
tcgaggagta catcgacgcc caagcaggtg agtagccatg aaacggaacg tctcaaccac 2040
ggttcgggcc aaactccact cactgacaaa cagaaagact gatctgctcg cccgtgagta 2100
tgaggcgttc cagactgagg tacacggtgg agacgctgac ctgtactctg caaccgacca 2160
gcaggcggcc ttgggcaagt tcttgaaacc gctgtccgag gcgggggctg tactggcaca 2220
gtccgaaacg caggtcatcg tccaacgcga cgacgaacct gcgaacctct ctgtttccgt 2280
ggattcaacc cccagcgggg gaaccccacg gctttagccg tgtggaggat gtcagaagcg 2340
gtctaggaga atgccccggg gcttgacccc ggggttgaat ccgataagca ggagtacaaa 2400
tcattaagtg aacatatccc gaatcaagag acagcgatgg agtacagtca ccgctacccc 2460
gcgtacccaa cacaagaggt ggcggctgaa ctggaacgac acatcgacgc tcatcgccaa 2520
gcgtacaact acaccctgta caagtacgag aacgtggacg ccgacaatat cggttccgcg 2580
tacaaa 2586

<210> 819
<211> 1561
<212> DNA
<213> 未知

[0367]

<220>
<223> 未知的描述:
实验室富集限定的培养基-有氧富集培养基序列
<400> 819

accctgagcc taggcgtcgt cggcgtgatc gccggcccgc tcgccatcgc cctgctcgtc 60
gaagtggtcg aactgatcac cgacgaacgg gtcagcatcc agcagacctt ccaggagagc 120
gccaccgagg ccgccggcga cgacgactga gcggcggccg taccgacgcg agtgcgggcg 180
gcgttacagc ggtagcgagg cgaaagctca cccgtttacg ggtgggatga agccgacaat 240
tgggaatcta tccacgttcg tagccgcggc tggggttgga tacttgcata acatctttag 300
gtgaatagta tctatatact gacgatgcct cgtgggttcg acagggaacg taccagcgtc 360
cacaaactcc agtatcactt catttggtgt ccgaagtacc gcaaatcggt gcttgagggt 420
gaggtacgtg accgtctcga agaactcatt gaggagaaag gagacgaact cggccaggtc 480
aagggctact cctcgcgcca cctccgcgac gagttcgact tcggtcttcc ctcgctgtgg 540
acgcgctcgt acttcgtctc aactgcgggc gacgtatcca gcgaagtcat cgaggagtac 600
atcgacgccc aagcaggtga gtagtcttgc aacaaaccgt ctcgaccacg gtgcgggtca 660
aactcccctc gctgaccaac aggaaggcag acctgctcgc ccgtgagtac gaggcgttcc 720
agaccgaagt tcacggcggc gatgctgacc tctactccgc gactaagcaa caggcggcct 780
tgggcaagtt ctcgaaaccg ctgtccgagg cgggggctgt actggcacag tccgaaacgc 840
aggtcatcgt ccaccgtgac gaggaacctg cgaacctctc cgtatccgtg gattcaaccc 900
ccagtggggg aaccccacgg ctttagccgt gggaggatgt cagtcgcccg tctcgggttc 960
gcgccccgcg gccttccact cggtcagact cccctcgtag aaggcgaggt tctcgtaccc 1020
gaggtgtcgg aggacgacgt aggtgtgact gatccggcga gcggtgttgc agtacagcag 1080

aacgcgccgg gtcgctgagc gggatattct cactctccgc accgcccgcg agaatagggc 1140
cgcgctcagt cgaccacgca agcgcgagcg cgcctcgccc tttcagtcca ccaggaccgc 1200
accgctccgc cccgcaccgc caccgcagac agcctcacac ctccccagcc tcctcgcgtc 1260
tcgtcgctcg cgtgactcgc tcctcggtgc tcgtccctcg cacggtgcgt cgcggcacgg 1320
aggccgcgac ggcgcgcgcc accgcaatcg atcggtgtcg gaacggccgc gaacgcacct 1380
ctcaagagcc tgcgggccgt ggcgatcggt aatgagcgac accggcggcc cactctcacc 1440
ggacagaccg gacgccgaga gcgatttccg ggtggacgcc cccttcgacc ccgcgggcga 1500
ccagcccgag gccatcgagc aactggtcga cggcttcgaa cgccagggcc acgacaaaca 1560
g 1561

<210> 820
<211> 506
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
盐湖宏基因组序列
<400> 820

[0368]

atggtcacgg ttcggatcaa tgaagtcacc aagtcaaagc tcgacaagtt gtgcgaggac 60
aacggcctca agatgcggtg ggttgttgaa caggcgatcc ttcaatattg tgagatgcac 120
gagtcgaaac atcgtccaga ggatgggtaa cttatgtcac tcaaaacact cacattctgg 180
ccgattggta tgtctgcaaa atacaggcgg atatttgaca acatccgtaa gcaacttggg 240
gcggcgcagc gcgaaatggt gaaagcatgg ctacggctgg cgttggcaga ggatgcttat 300
ggtgacctta atgctgcgaa taacatcgcg cgggggatgc agacgtttgg gccttgtgcc 360
aagtcctctg cgtcagacga cgctatcggg ctgcggttcg atagcgaggg gcaagaatct 420
gccctggggc aggcgaagcg attgtttgct tgagcttgca agttcttgtg tgtgtgtggg 480
tatgaactgc tgagcgtttg tgtagt 506

<210> 821
<211> 1915
<212> DNA
<213> 未知
<220>
<223> 未知的描述:
水生-温泉-沉积物序列
<400> 821

gtcgtaacga tggctgatgt gttagaggga ggggctgtgg ggatcaatga caataaaatg 60
aacactctcg tattaatgca tgctaaccca ctatggggcg agattgagat aaaaactccc 120
gcagggcgaa cggtttggcg tgctccttaa ccgttggtga cctgaatgtg ggaaacaatc 180
ttgattggcg tttgtggggc gcactttccc gctaccgtag gggcgacaaa aagcagcact 240
attctggctt tccgtcgcca tcatcgctat catcgctctc atagccatca acactttgac 300
cgtccgctgg ctcggtcatc cgccacaaaa ctcatccaca acaatagcac tgtcagagtg 360
aaaggactta gcagcgttcg ttaaagcgcc gacctgcagc gccgaaagac acttgctaag 420
ccttaccgtc ggcttggtca cttcccgttt cccgtagcgc gtgaagccag aatggtaaag 480
cctaaaatgc cgtttttgcg ttgcgacaac gcgcaagtta gcgtcagcag gcgcacatag 540
gcacaggcgc acgggcgtag gcacgcaggc gcacgcacgc gggcacgggc gcgctttcaa 600

cggggatacc cgagcgcgca ggcgggtcag gctgcccgcc gaccttaaag tcaaccttaa 660
agactacaac ccccaaaccg acgctatgtg cgctatgctt tgggacgctt ataaagttca 720
ggtgcttaag gaacttgttg gcgaggatga gggcgactac attaagtaaa gccgtgagga 780
gttattagag gctttgcggg aacgggagca caagtatggt tataaagcgc cccttaatgc 840
gaacttaaca aaaccgtccc gaccagttat ttcgggtaaa attttaaccc atttgcccaa 900
attccgcaac ttcgttgcag caaacctcca gtcacccctt tcccctaagg ggtttggcgc 960
aagcaattct tgaaattcct gccaaaatcg gcaaaattca aaggtcccgt tctttgatgc 1020
ctatcacggc agaataaagg gggaggttga cggcaaatga aaaaagtggt cataaccatc 1080
actactggag tagagtttct acaacctcaa gatgagcaaa tcgtgaggaa ggcagcggaa 1140
tggtatacaa aggcacgaga agccctttt gcccatgccc tccaacaaca acaaggtaaa 1200
gttgagggaa accgattttg ttgcacaaag tgtaaatttc agcacaatgc ggactttgtg 1260
gcggcgtata acattgccgt aagggcattg aagaagttag ggatgaggga agagataggt 1320
tggaaggtat tgacaggcaa gccgaaggtg ttataatgac gatggctcat ggggagcagg 1380
ctccccatgg gcagcgtgcc cgccgaaagg tgggtggagg gcagagcctg tctgcctgta 1440
agcagcacct tcccacatgg ggaggttgtc tgcaagttca cattatcccc tgaccccagc 1500
ggtgaagagc cgctggggtt tcaattgtct tagcaccctg atgtccacat taacccctta 1560
agacagaaaa cttgcccaag cgaggattag cacgaactat tgccttgtaa cgctcccctt 1620
agcagttgca cattgcaaag caagtttctt ggcattaaga gactctttca gtcccacaat 1680
ggtttgattg ctacattgcc tgccaagata aagcccttcc tttccattat cacttgtaaa 1740
ggaccagtgt cataattcct caggggtgag ggagatgaag gtcattattg ggcgtgtaat 1800
ggaagacact tacacggtca aaagtcaaga acccttgcca ccaggacagg aagtcttcat 1860
tgtaactaag atggacgctg aggatgtcta ttaccttgca aaaactttgg cgtgg 1915

[0369]

<210> 822
<211> 13
<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> (6)..(6)
<223> a、c、t 或 g
<220>
<221>经修饰的碱基
<222> (10)..(11)
<223> a、c、t 或 g
<400> 822
gyyygnkrkn nac 13
<210> 823
<211> 16
<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> (13)..(13)
<223> a、c、t 或 g
<400> 823
yccrcgcgcg cgnggg 16
<210> 824
<211> 12
<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> (4)..(4)
<223> a、c、t 或 g
<400> 824
rbbnrbkgac as 12

[0370]

<210> 825
<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> (6)..(6)
<223> a、c、t 或 g
<400> 825
gkykknrcyg myswcgwtyg ygggrskkgt kymaaag 37
<210> 826
<211> 108
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
多核苷酸
<220>
<221> 经修饰的碱基
<222> (38)..(71)

<223> a、c、t、g、未知或其他
<400> 826
ctctgacgcg ctgacgttcg cggggttggt ttcaaagnnn nnnnnnnnnn nnnnnnnnnn 60
nnnnnnnnnn nctctgacgc gctgacgttc gcggggttgg tttcaaag 108
<210> 827
<211> 33
<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> (30)..(30)
<223> a、c、t 或 g
<400> 827
cgycgyrcbb yccrcgcgcg cgsgggvbgn dkg 33
<210> 828
<211> 34
<212> DNA
<213> 人工序列
<220>
[0371]
<223>人工序列的描述：合成
寡核苷酸
<400> 828
cgtcgcacct cgcgcgcgcg taggagatgg tttg 34
<210> 829
<211> 92
<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> (27)..(66)
<223> a、c、t、g、未知或其他
<400> 829
ctctccacgc gcgcgcggga tgcgggnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 60
nnnnnnctct ccacgcgcgc gcgggatgcg gg 92
<210> 830
<211> 40
<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> (10)..(13)
<223> a、c、t 或 g
<220>
<221> 经修饰的碱基
<222> (24)..(24)
<223> a、c、 t 或 g
<400> 830
gsygbygyrn nnnsybhsvv vdhnvhdrbb rrykgacaca 40
<210> 831
<211> 106
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
多核苷酸
<220>
[0372] <221> 经修饰的碱基
<222> (37)..(70)
<223> a、c、t、g、未知或其他
<400> 831
gtgcatcggc cgacgaagag cgaaggtcag tgacagnnnn nnnnnnnnnn nnnnnnnnnn 60
nnnnnnnnnn gtgcatcggc cgacgaagag cgaaggtcag tgacag 106
<210> 832
<211> 13
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<220>
<223> 组合的 DNA/RNA 分子的描述：合成
寡核苷酸
<220>
<221> 经修饰的碱基
<222> (6)..(6)
<223> a、c、 t、g 或 u
<220>
<221> 经修饰的碱基
<222> (10)..(11)

<223> a、c、 t、g或u
<400> 832
gyyygnkrkn nac 13
<210> 833
<211> 16
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<220>
<223> 组合的DNA/RNA分子的描述：合成
寡核苷酸
<220>
<221> 经修饰的碱基
<222> (13)..(13)
<223> a、c、 t、g或u
[0373] <400> 833
yccrcgcgcg cgnggg 16
<210> 834
<211> 12
<212> DNA
<213> 人工序列
<220>
<223>人工序列的描述：合成
寡核苷酸
<220>
<223> 组合的DNA/RNA分子的描述：合成
寡核苷酸
<220>
<221> 经修饰的碱基
<222> (4)..(4)
<223> a、c、 t、g或u
<400> 834
rbbnrbkgac as 12

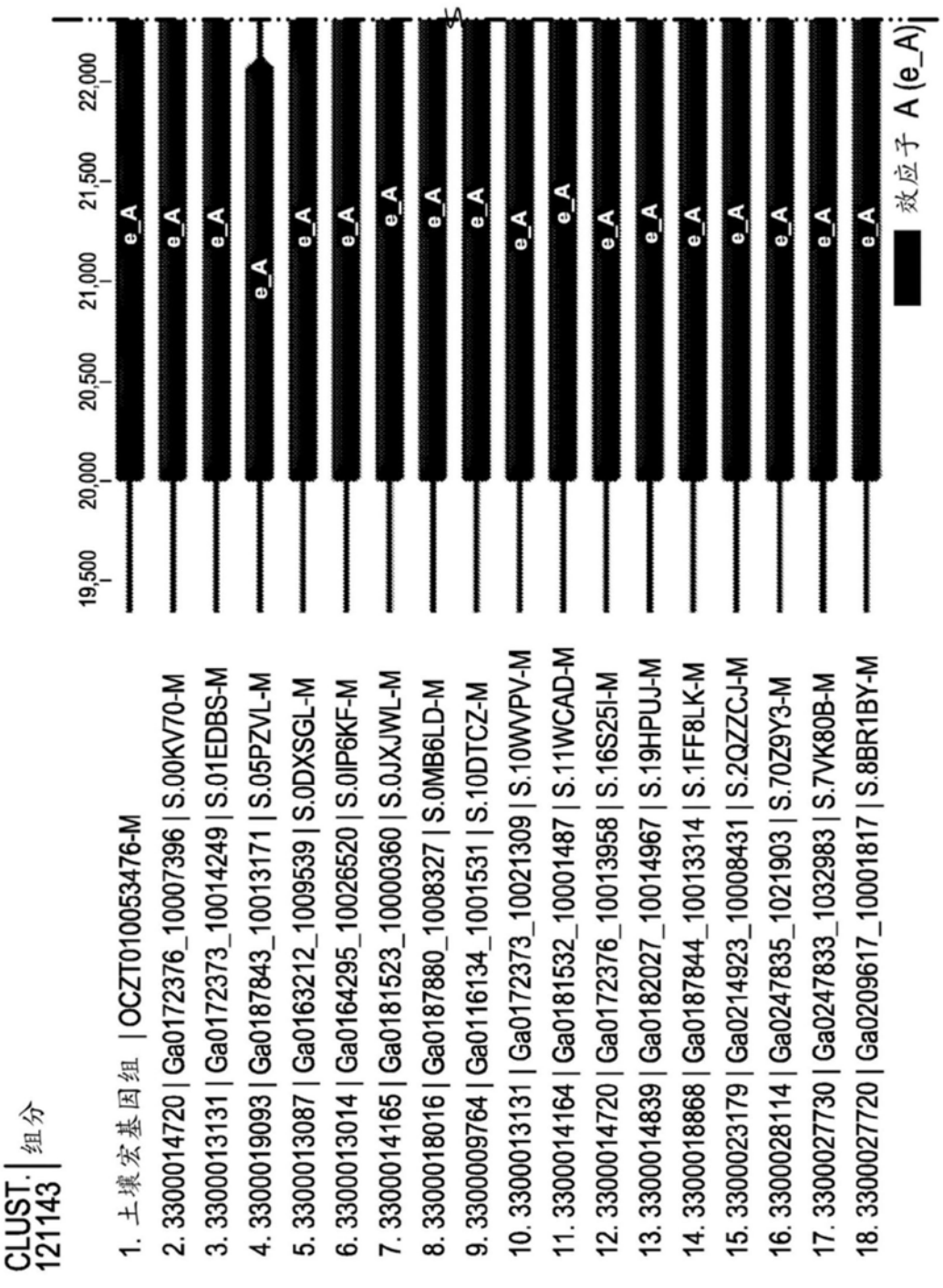

CLUST.
121143 | 组分
1. 土壤宏基因组 | OCZT010053476-M
2. 3300014720 | Ga0172376_10007396 | S.00KV70-M
3. 3300013131 | Ga0172373_10014249 | S.01EDBS-M
4. 3300019093 | Ga0187843_10013171 | S.05PZVL-M
5. 3300013087 | Ga0163212_1009539 | S.0DXSGL-M
6. 3300013014 | Ga0164295_10026520 | S.0IP6KF-M
7. 3300014165 | Ga0181523_10000360 | S.0JXJWL-M
8. 3300018016 | Ga0187880_1008327 | S.0MB6LD-M
9. 3300009764 | Ga0116134_1001531 | S.10DTCZ-M
10. 3300013131 | Ga0172373_10021309 | S.10WVPV-M
11. 3300014164 | Ga0181532_10001487 | S.11WCAD-M
12. 3300014720 | Ga0172376_10013958 | S.16S25I-M
13. 3300014839 | Ga0182027_10014967 | S.19HPUJ-M
14. 3300018868 | Ga0187844_10013314 | S.1FF8LK-M
15. 3300023179 | Ga0214923_10008431 | S.2QZZCJ-M
16. 3300028114 | Ga0247835_1021903 | S.70Z9Y3-M
17. 3300027730 | Ga0247833_1032983 | S.7VK80B-M
18. 3300027720 | Ga0209617_10001817 | S.8BR1BY-M
19,500
20,000
20,500
21,000
21,500
22,000
e_A
效应子 A (e_A)

图1(部分1)

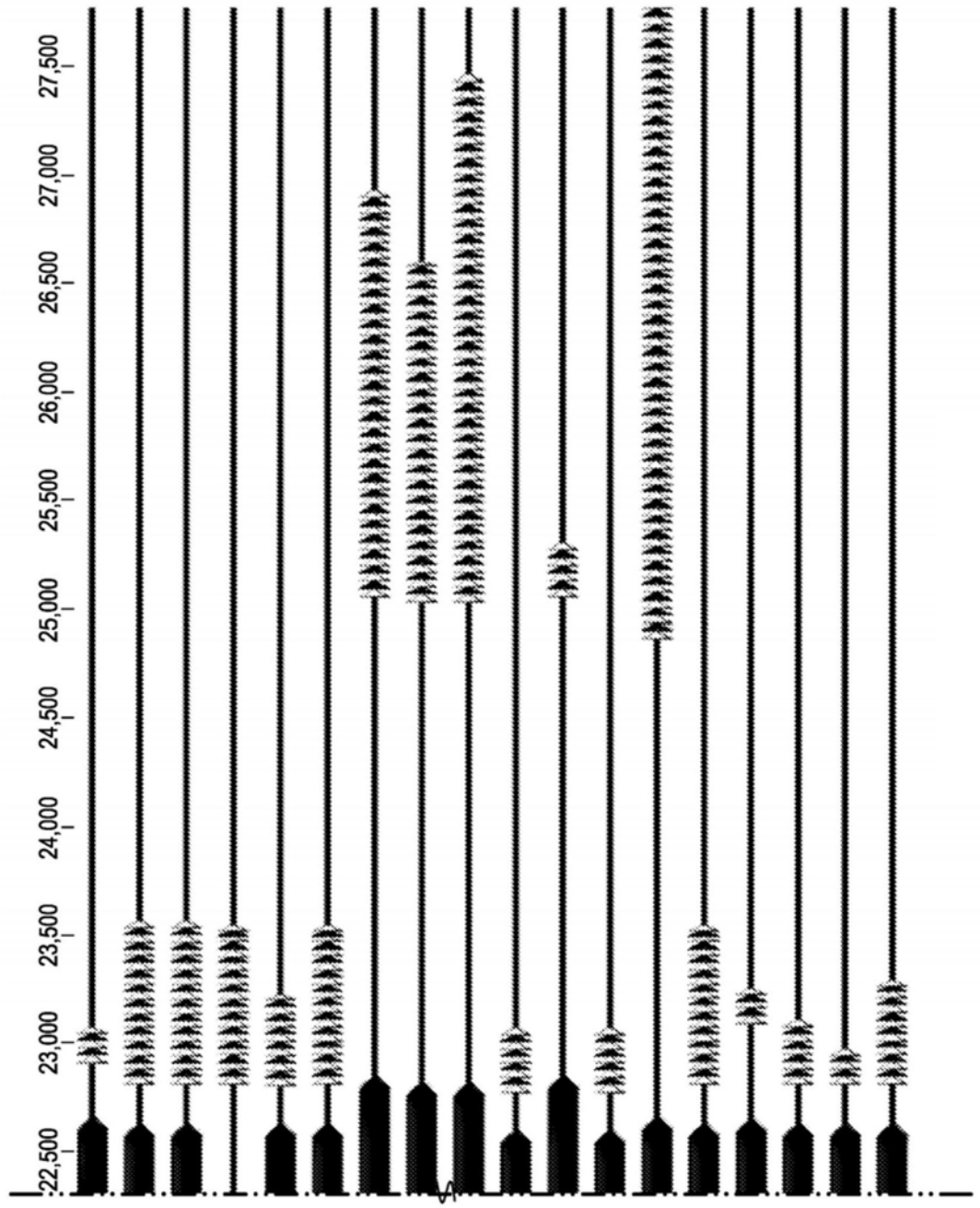

27,500
27,000
26,500
26,000
25,500
25,000
24,500
24,000
23,500
23,000
22,500

图1(部分2)

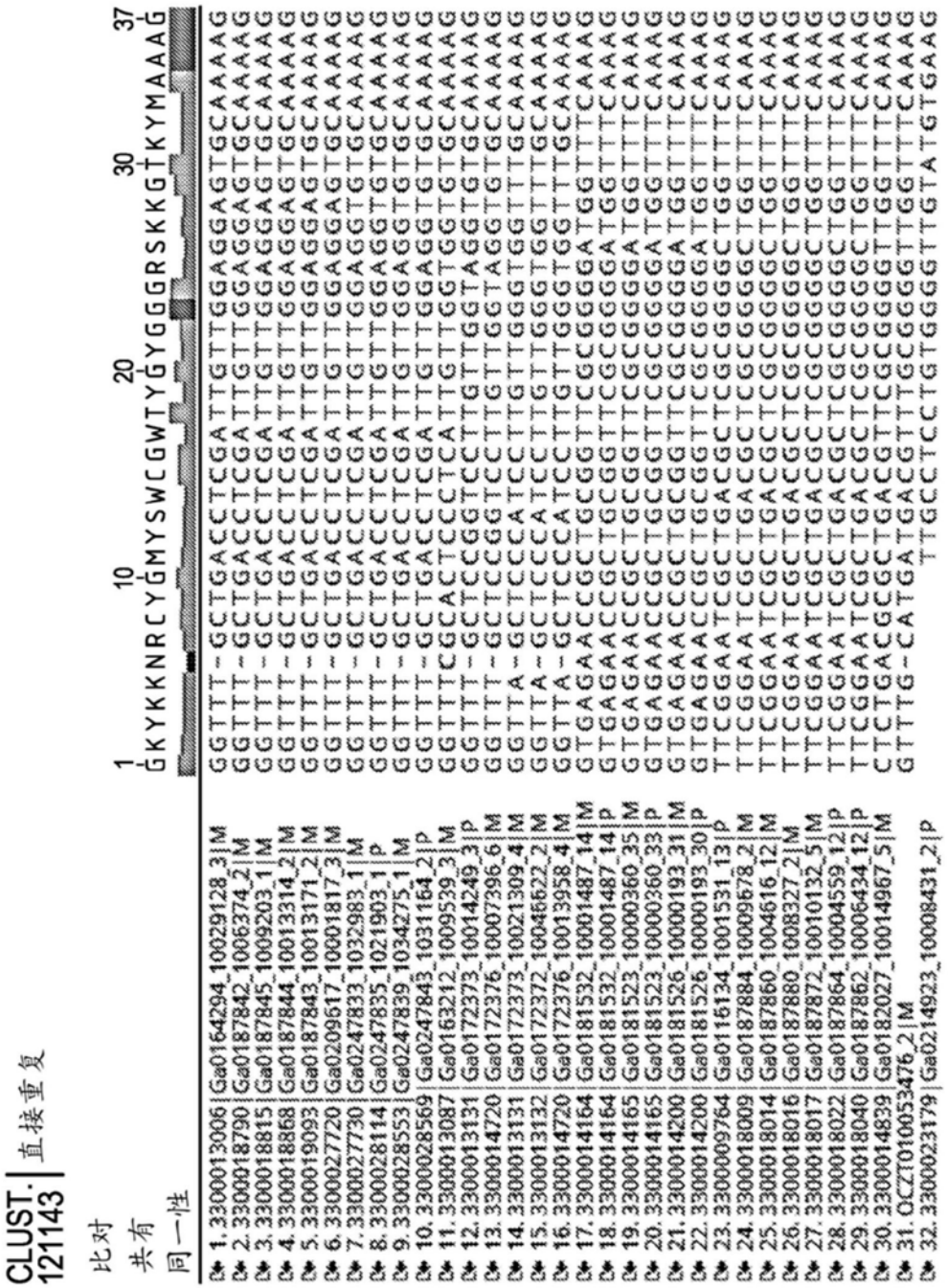

CLUST. 121143 直接重复
比对
共有
同一性
1 10 20 30 37
GKYKKNRCYGMYSWCGWTYGYGGGRSKKGTKYMAAAG
1.3300013006|Ga0164294_10029128_3|M
2.3300018790|Ga0187842_1006374_2|M
3.3300018815|Ga0187845_1009203_1|M
4.3300018868|Ga0187844_10013314_2|M
5.3300019093|Ga0187843_10013171_2|M
6.3300027720|Ga0209617_10001817_3|M
7.3300027730|Ga0247833_1032983_1|M
8.3300028114|Ga0247835_1021903_1|P
9.3300028553|Ga0247839_1034275_1|M
10.3300028569|Ga0247843_1031164_2|P
11.3300013087|Ga0163212_1009539_3|M
12.3300013131|Ga0172373_10014249_3|P
13.3300014720|Ga0172376_10007396_6|M
14.3300013131|Ga0172373_10021309_4|M
15.3300013132|Ga0172372_10046622_2|M
16.3300014720|Ga0172376_10013958_4|M
17.3300014164|Ga0181532_10001487_14|M
18.3300014164|Ga0181532_10001487_14|P
19.3300014165|Ga0181523_10000360_35|M
20.3300014165|Ga0181523_10000360_33|P
21.3300014200|Ga0181526_10000193_31|M
22.3300014200|Ga0181526_10000193_30|P
23.3300009764|Ga0116134_1001531_13|P
24.3300018009|Ga0187884_10009678_2|M
25.3300018014|Ga0187860_1004616_12|M
26.3300018016|Ga0187880_1008327_2|M
27.3300018017|Ga0187872_10010132_5|M
28.3300018022|Ga0187864_10004559_12|P
29.3300018040|Ga0187862_10006434_12|P
30.3300014839|Ga0182027_10014967_5|M
31.OCZT01003476_2|M
32.3300023179|Ga0214923_10008431_2|P

图2

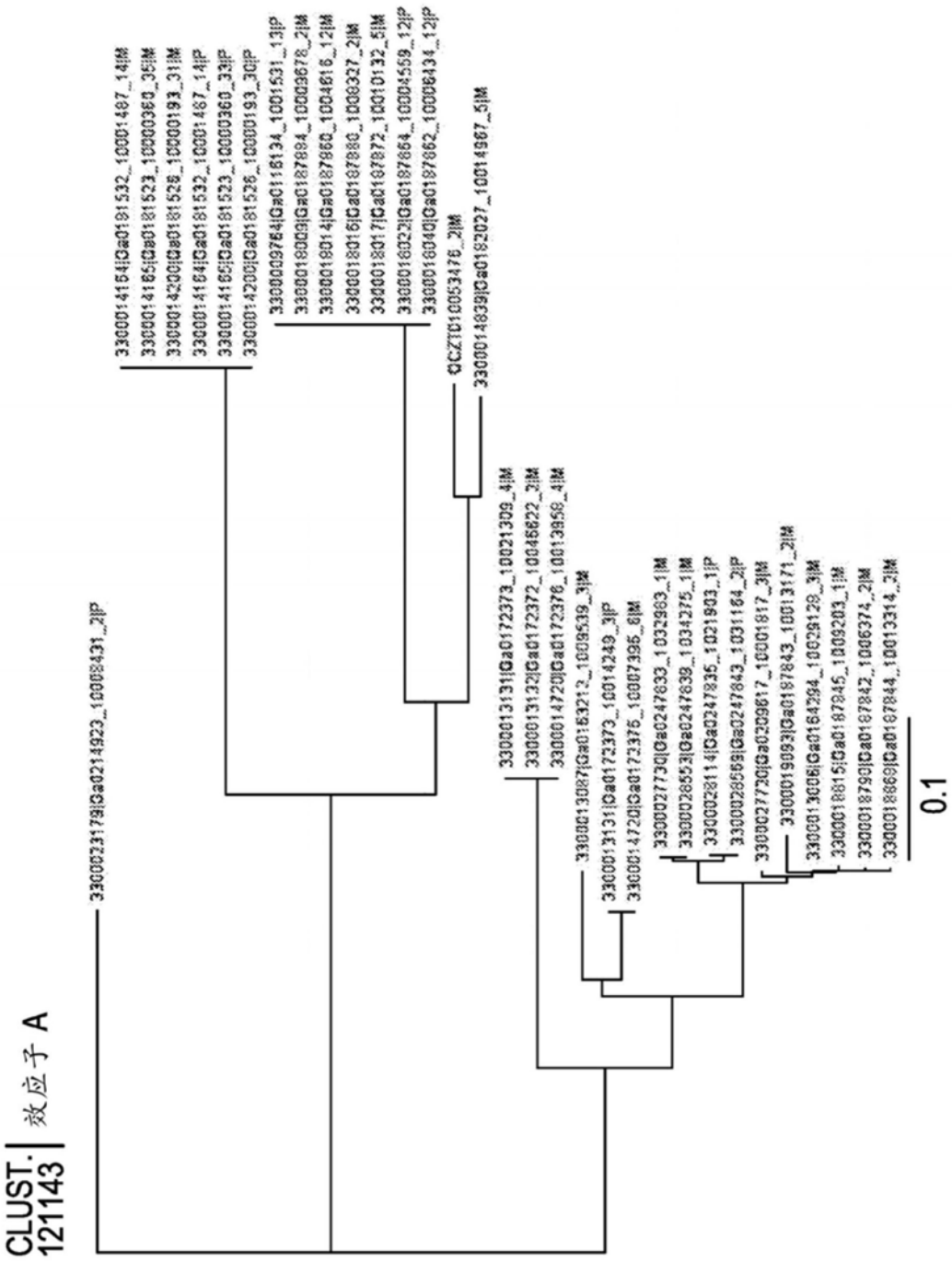


图3

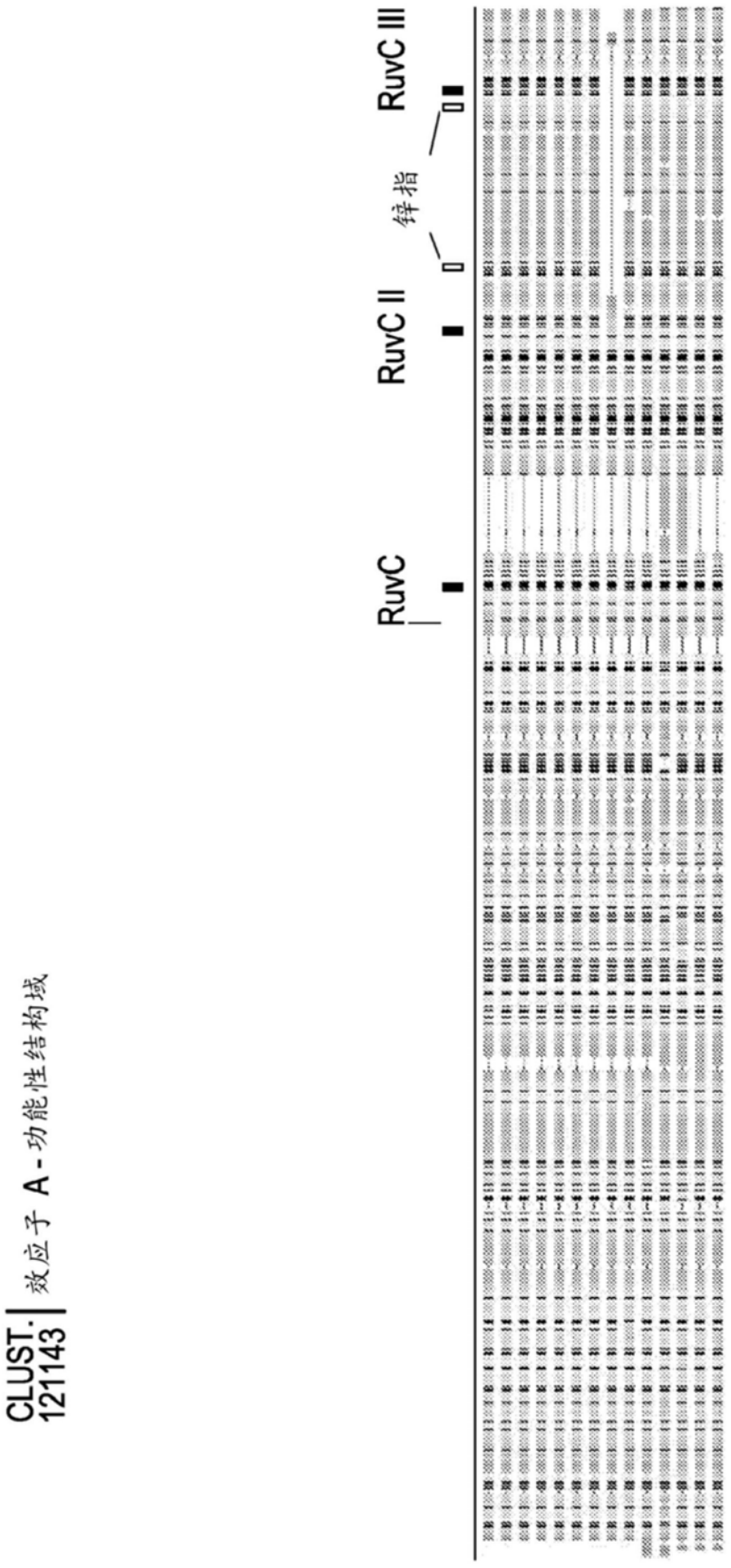


图4

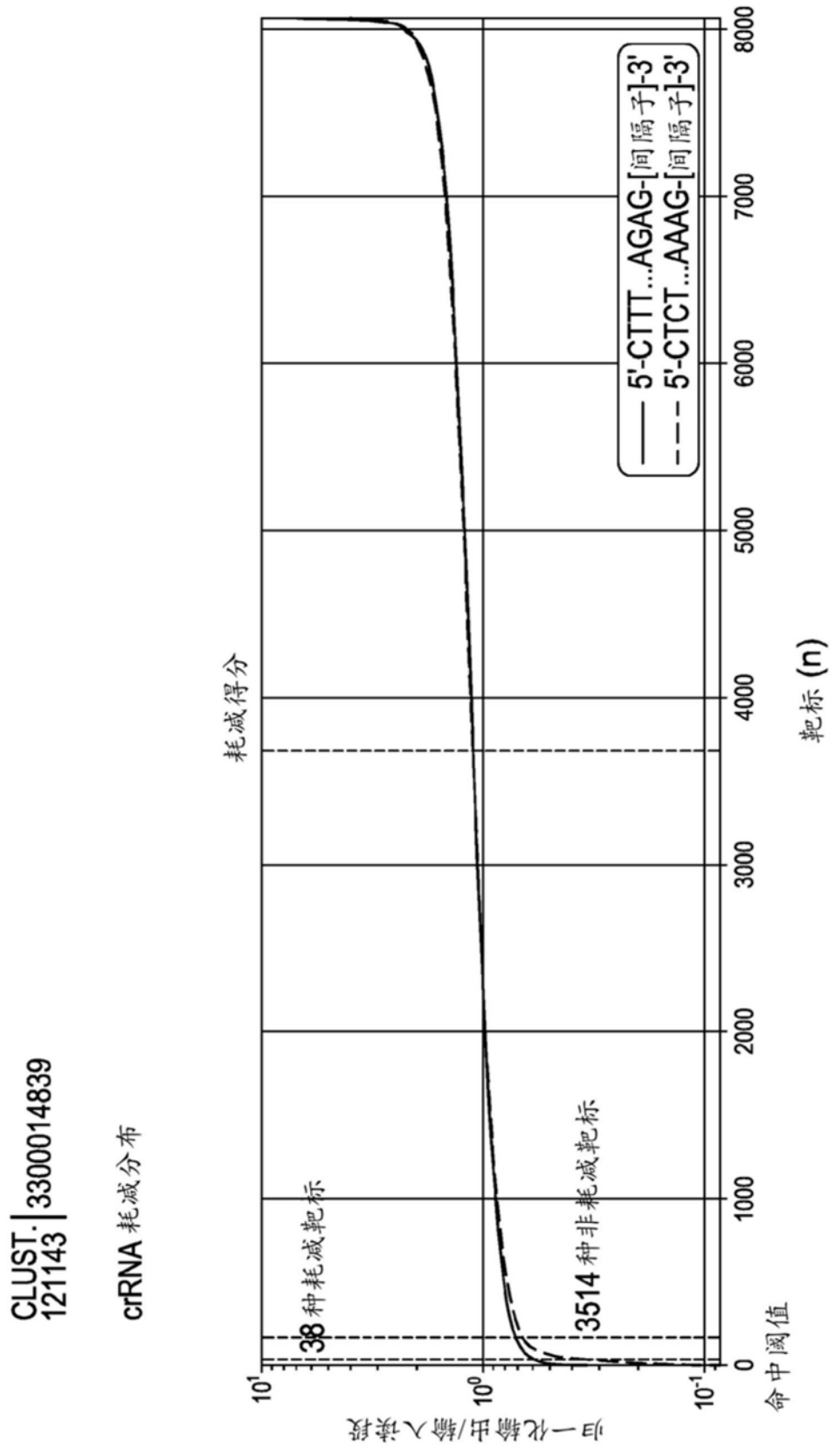

CLUST.
121143 3300014839
crRNA 耗减分布
耗减得分
归一化输出/输入读段
10^1
10^0
10^-1
38 种耗减靶标
3514 种非耗减靶标
0
1000
2000
3000
4000
5000
6000
7000
8000
命中阈值
靶标 (n)
5'-CTTT...AGAG-[间隔子]-3'
5'-CTCT...AAAG-[间隔子]-3'

图5

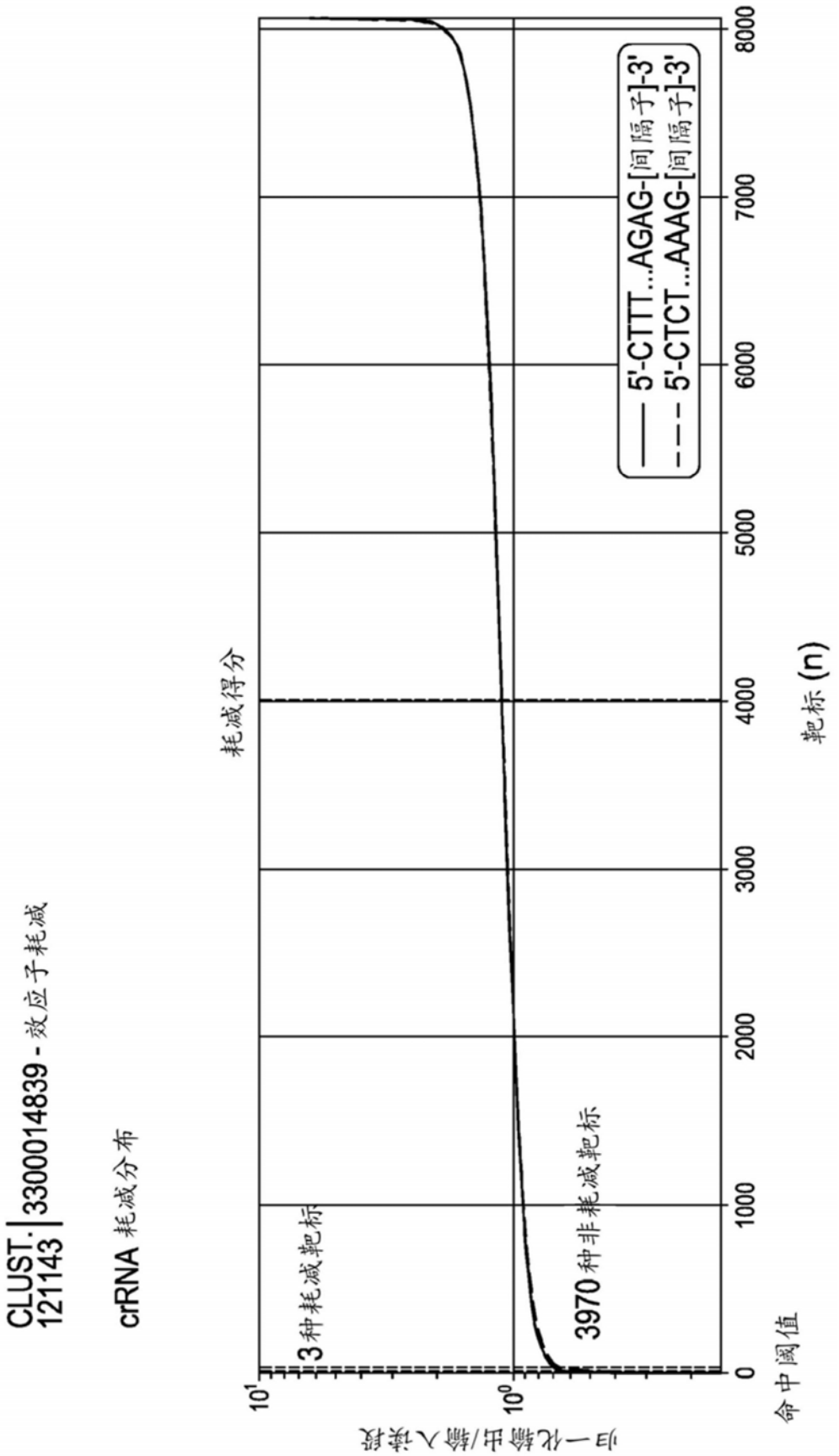
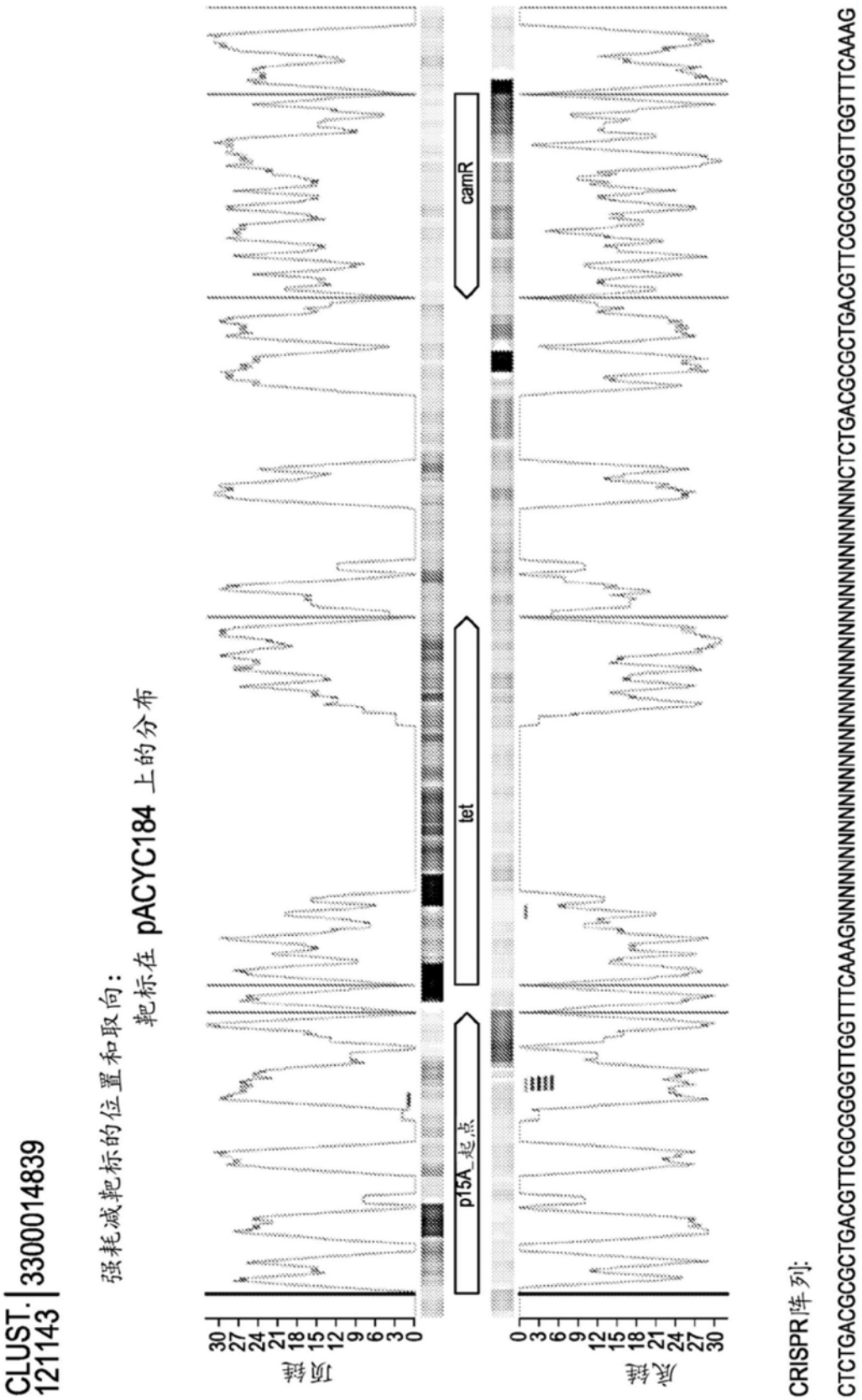
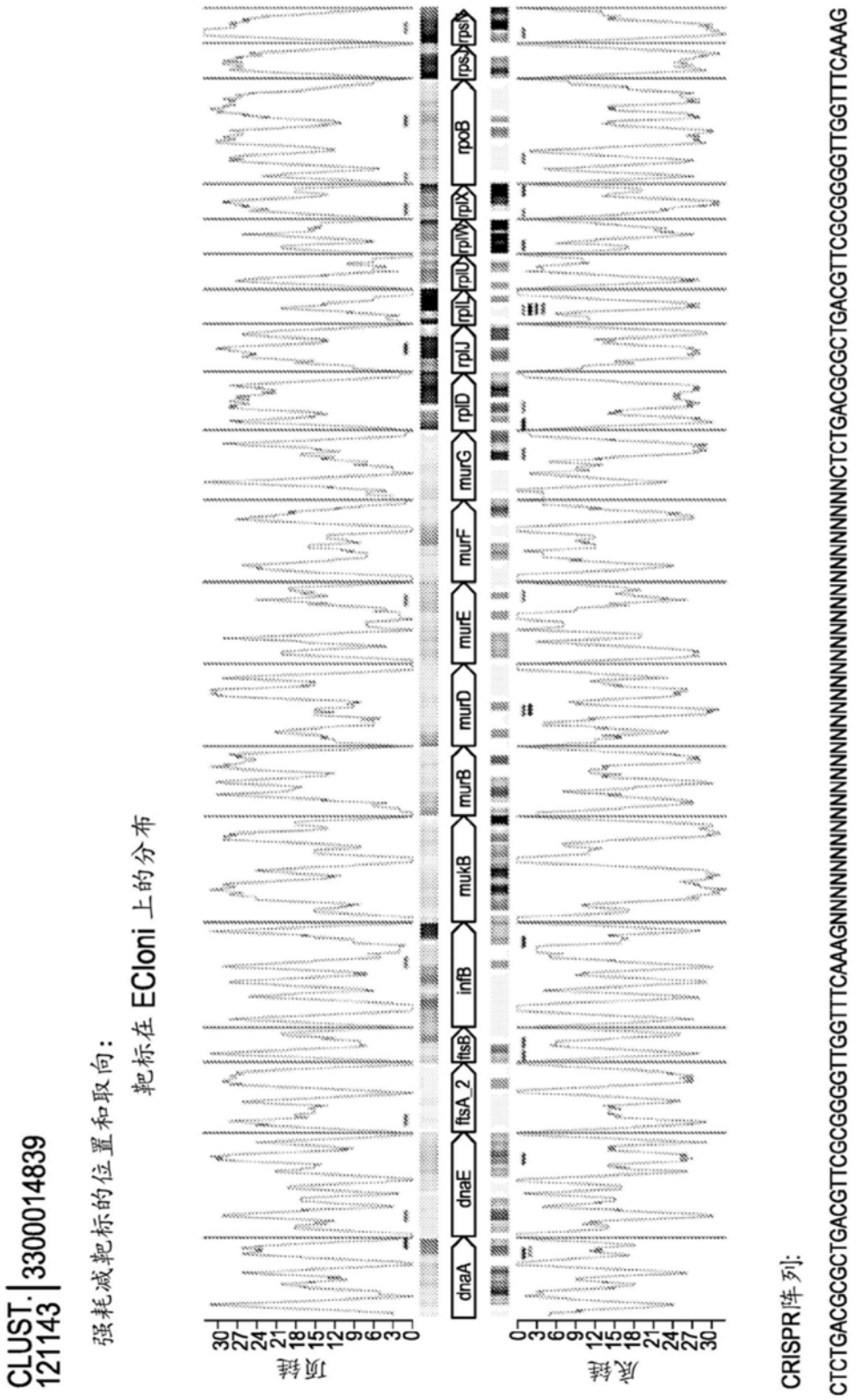

CLUST. 121143 | 3300014839 - 效应子耗减
crRNA 耗减分布
耗减得分
归一化输出/输入读段
命中阈值
靶标 (n)
3种耗减靶标
3970种非耗减靶标
5'-CTTT...AGAG-[间隔子]-3'
5'-CTCT...AAAG-[间隔子]-3'
10^1
10^0
0
1000
2000
3000
4000
5000
6000
7000
8000

图6


CLUST. 121143 | 3300014839
强耗减靶标的位置和取向：
靶标在 pACYC184 上的分布
顶链
反链
p15A_起点
tet
camR
CRISPR阵列:
CTCTGACGCGCTGACGTTCGCGGGGTTGGTTTCAAAGNNN


CLUST. 121143 | 3300014839
强耗减靶标的位置和取向：
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
rpsK
CRISPR阵列:
CTCTGACGCGCTGACGTTCGCGGGGTTGGTTTCAAAGNNN

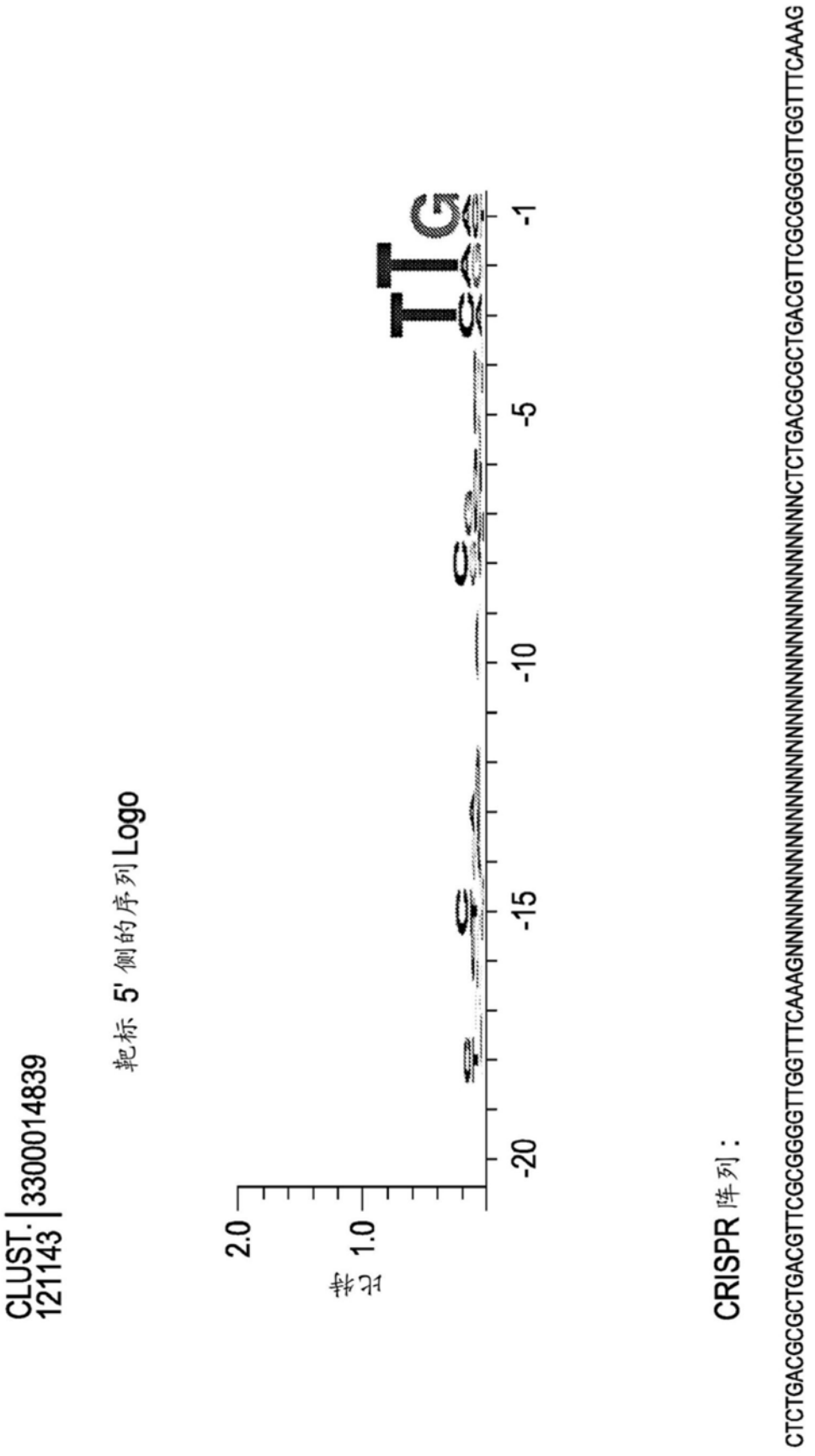

CLUST.
121143 | 3300014839
靶标 5' 侧的序列 Logo
比特
2.0
1.0
-20
-15
-10
-5
-1
CRISPR 阵列：
CTCTGACGCGCTGACGTTCGCGGGGTTGGTTTCAAAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCTCTGACGCGCTGACGTTCGCGGGGTTGGTTTCAAAG

图8

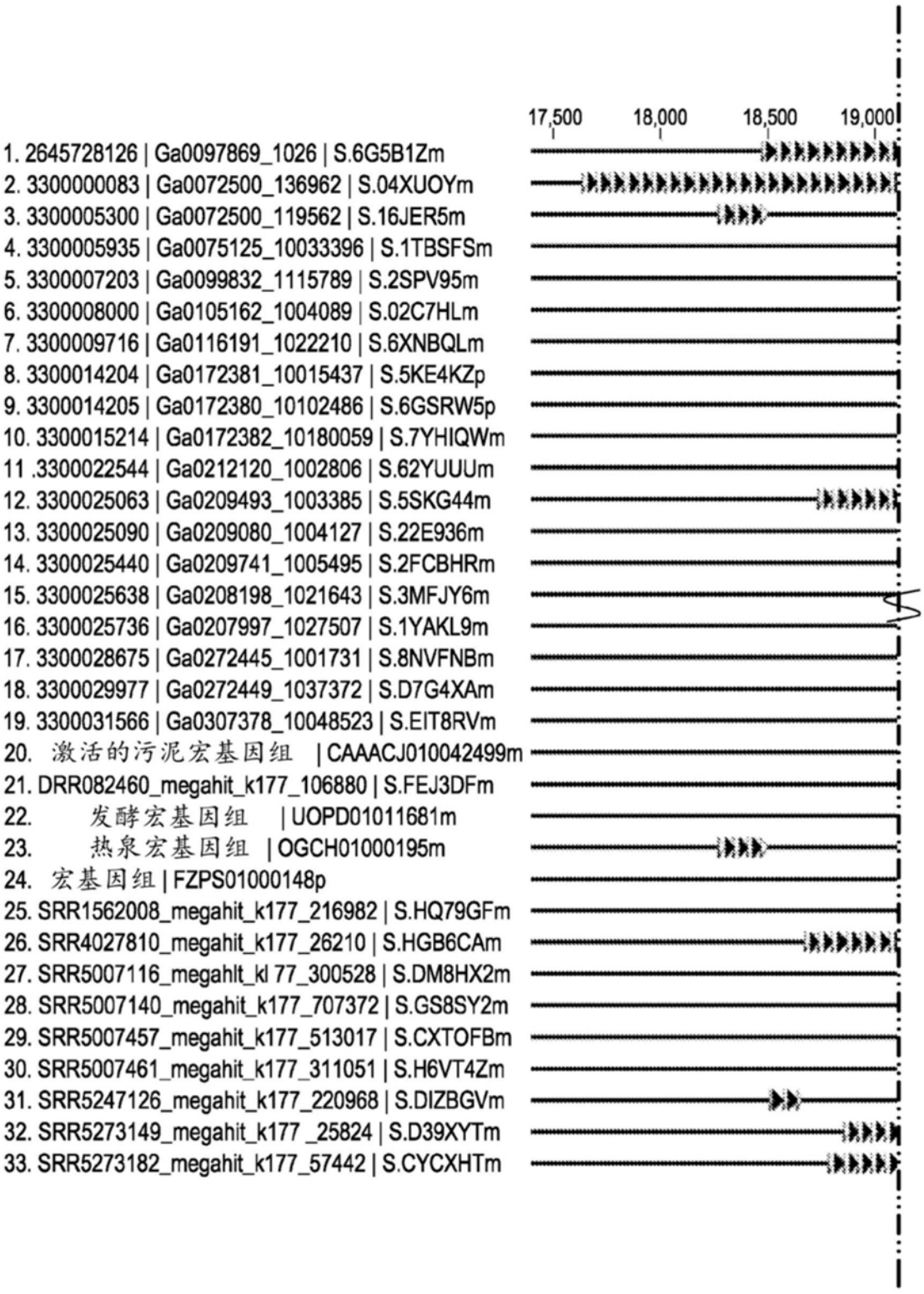

CLUST.
196682
组分
17,500
18,000
18,500
19,000
1. 2645728126 | Ga0097869_1026 | S.6G5B1Zm
2. 3300000083 | Ga0072500_136962 | S.04XUOYm
3. 3300005300 | Ga0072500_119562 | S.16JER5m
4. 3300005935 | Ga0075125_10033396 | S.1TBSFSm
5. 3300007203 | Ga0099832_1115789 | S.2SPV95m
6. 3300008000 | Ga0105162_1004089 | S.02C7HLm
7. 3300009716 | Ga0116191_1022210 | S.6XNBQLm
8. 3300014204 | Ga0172381_10015437 | S.5KE4KZp
9. 3300014205 | Ga0172380_10102486 | S.6GSRW5p
10. 3300015214 | Ga0172382_10180059 | S.7YHIQWm
11 .3300022544 | Ga0212120_1002806 | S.62YUUUm
12. 3300025063 | Ga0209493_1003385 | S.5SKG44m
13. 3300025090 | Ga0209080_1004127 | S.22E936m
14. 3300025440 | Ga0209741_1005495 | S.2FCBHRm
15. 3300025638 | Ga0208198_1021643 | S.3MFJY6m
16. 3300025736 | Ga0207997_1027507 | S.1YAKL9m
17. 3300028675 | Ga0272445_1001731 | S.8NVFNBm
18. 3300029977 | Ga0272449_1037372 | S.D7G4XAm
19. 3300031566 | Ga0307378_10048523 | S.EIT8RVm
20. 激活的污泥宏基因组 | CAAACJ010042499m
21. DRR082460_megahit_k177_106880 | S.FEJ3DFm
22. 发酵宏基因组 | UOPD01011681m
23. 热泉宏基因组 | OGCH01000195m
24. 宏基因组 | FZPS01000148p
25. SRR1562008_megahit_k177_216982 | S.HQ79GFm
26. SRR4027810_megahit_k177_26210 | S.HGB6CAm
27. SRR5007116_megahlt_kl 77_300528 | S.DM8HX2m
28. SRR5007140_megahit_k177_707372 | S.GS8SY2m
29. SRR5007457_megahit_k177_513017 | S.CXTOFBm
30. SRR5007461_megahit_k177_311051 | S.H6VT4Zm
31. SRR5247126_megahit_k177_220968 | S.DIZBGVm
32. SRR5273149_megahit_k177 _25824 | S.D39XYTm
33. SRR5273182_megahit_k177_57442 | S.CYCXHTm

图9(部分1)

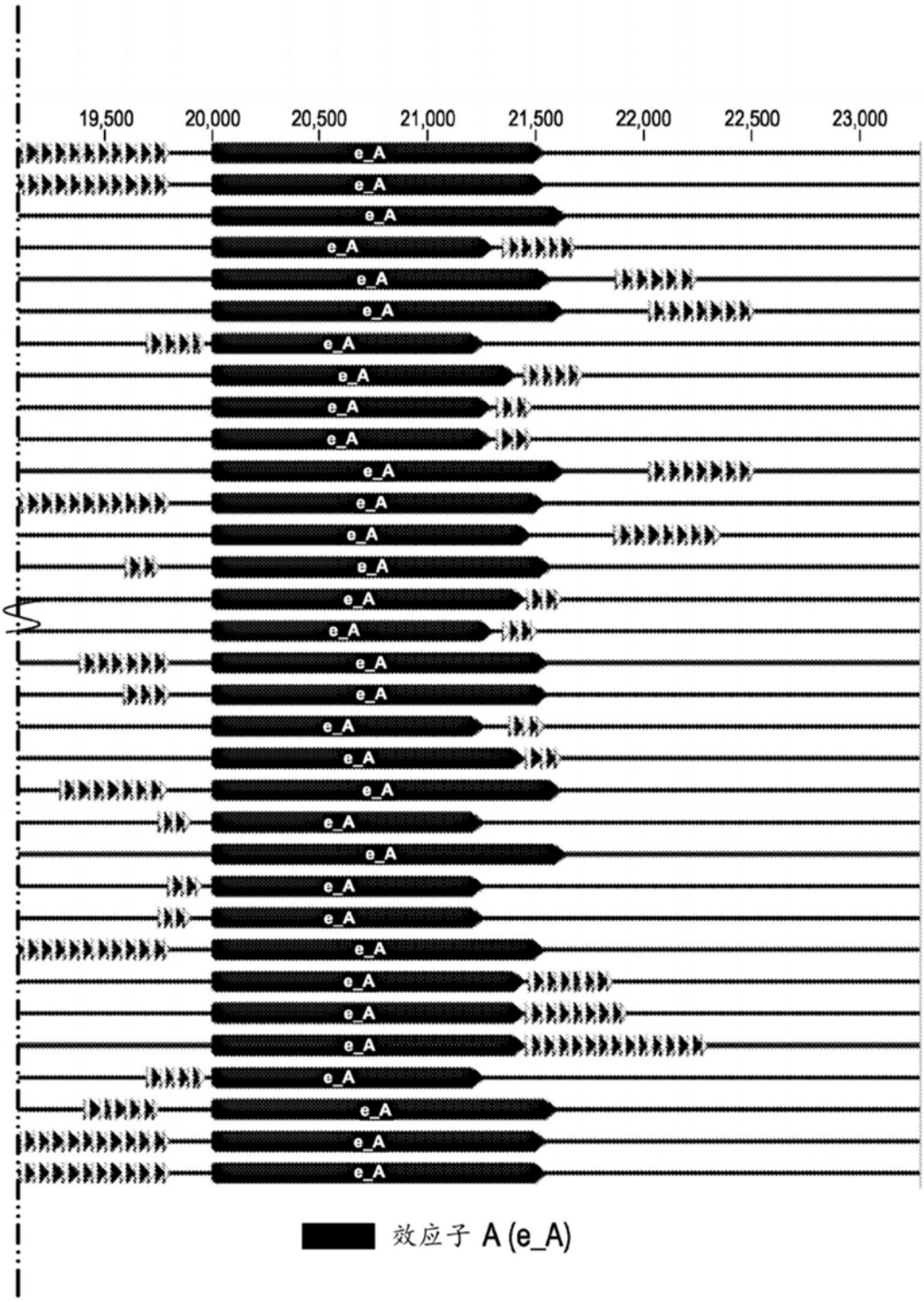


图9(部分2)

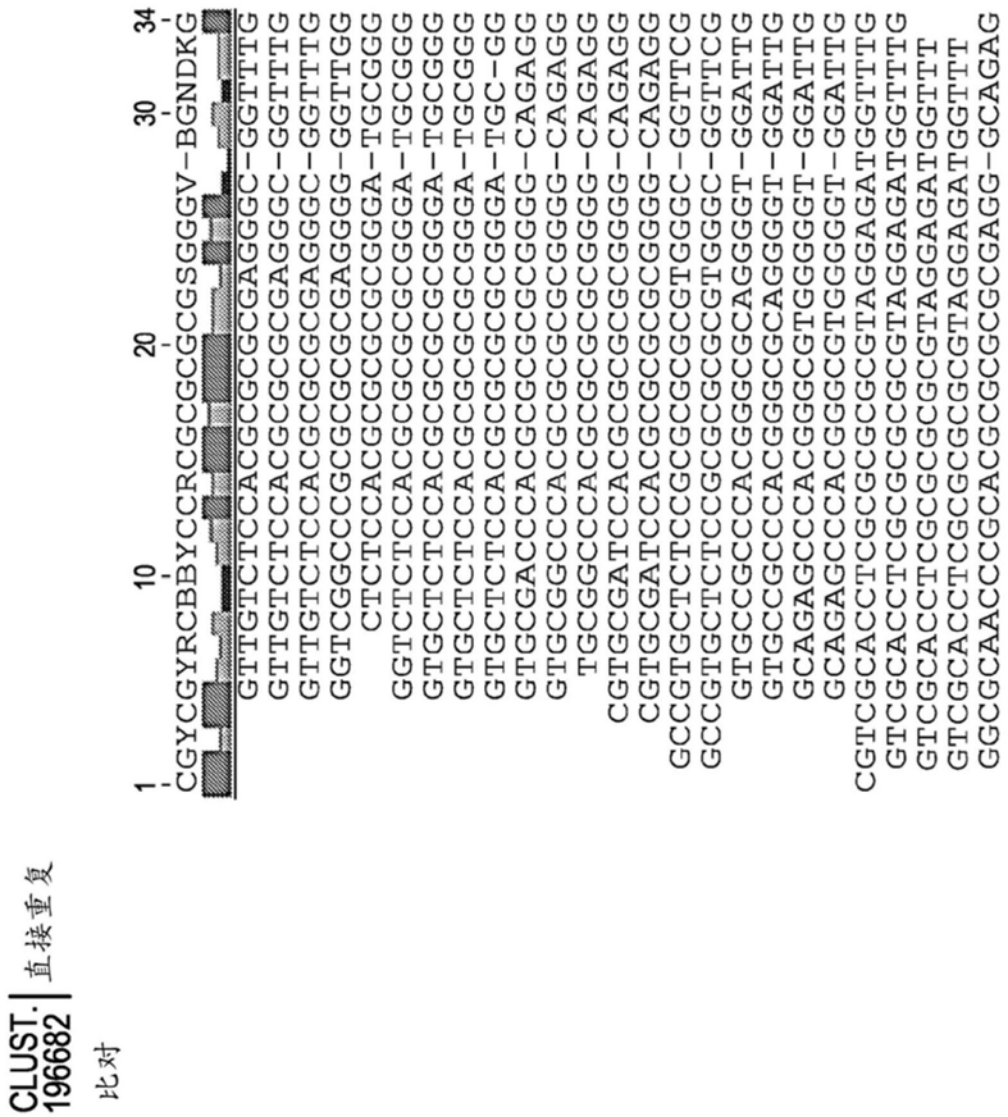

CLUST.
196682
直接重复
比对
1
10
20
30
34
CGYCGYRCBBYCCRCGCGCGCGSGGGV-BGNDKG
GTTGTCTCCACGCGCGCGAGGGC-GGTTTG
GTTGTCTCCACGCGCGCGAGGGC-GGTTTG
GTTGTCTCCACGCGCGCGAGGGC-GGTTTG
GGTCGGCCCGCGCGCGCGAGGGG-GGTTGG
CTCTCCACGCGCGCGCGGGA-TGCGGG
GGTCTCTCCACGCGCGCGCGGGA-TGCGGG
GTGCTCTCCACGCGCGCGCGGGA-TGCGGG
GTGCTCTCCACGCGCGCGCGGGA-TGCGGG
GTGCTCTCCACGCGCGCGCGGGA-TGC-GG
GTGCGACCCCACGCGCGCGCGGGG-CAGAGG
GTGCGGCCCCACGCGCGCGCGGGG-CAGAGG
TGCGGCCCCACGCGCGCGCGGGG-CAGAGG
CGTGCGATCCCACGCGCGCGCGGGG-CAGAGG
CGTGCGATCCCACGCGCGCGCGGGG-CAGAGG
GCCGTGCTCTCCGCGCGCGCGTGGGC-GGTTCG
GCCGTGCTCTCCGCGCGCGCGTGGGC-GGTTCG
GTGCCGCCCCACGGGGCGCAGGGGT-GGATTG
GTGCCGCCCCACGGGGCGCAGGGGT-GGATTG
GCAGAGCCCCACGGGCGTGGGGGGT-GGATTG
GCAGAGCCCCACGGGCGTGGGGGGT-GGATTG
CGTCGCACCCTCGCGCGCGCGTAGGAGATGGTTTG
GTCGCACCCTCGCGCGCGCGTAGGAGATGGTTTG
GTCGCACCCTCGCGCGCGCGTAGGAGATGGTTT
GTCGCACCCTCGCGCGCGCGTAGGAGATGGTTT
GGCGCAACCCGCACGCGCGCGCGAGG-GCAGAG

图10

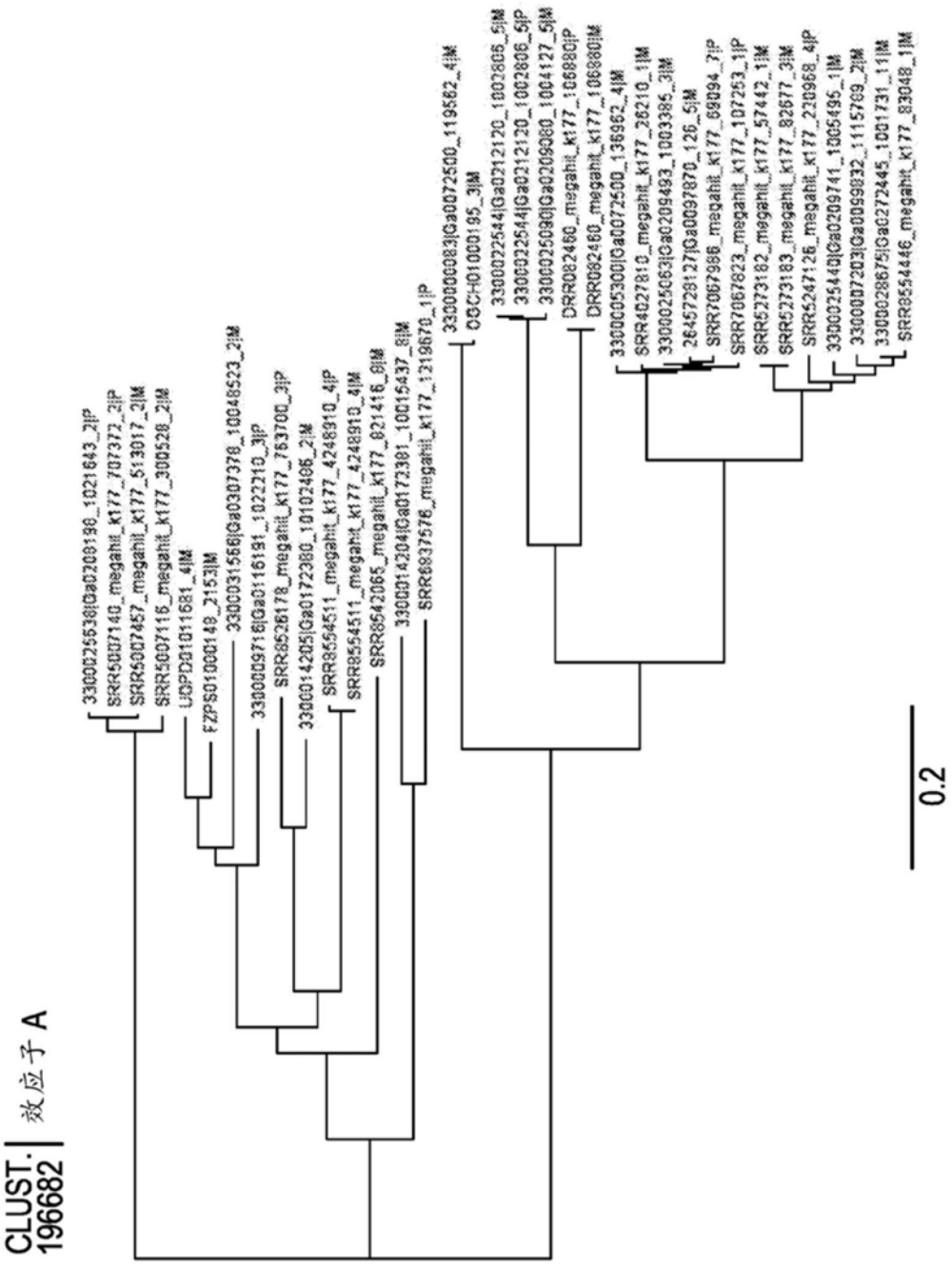


图11

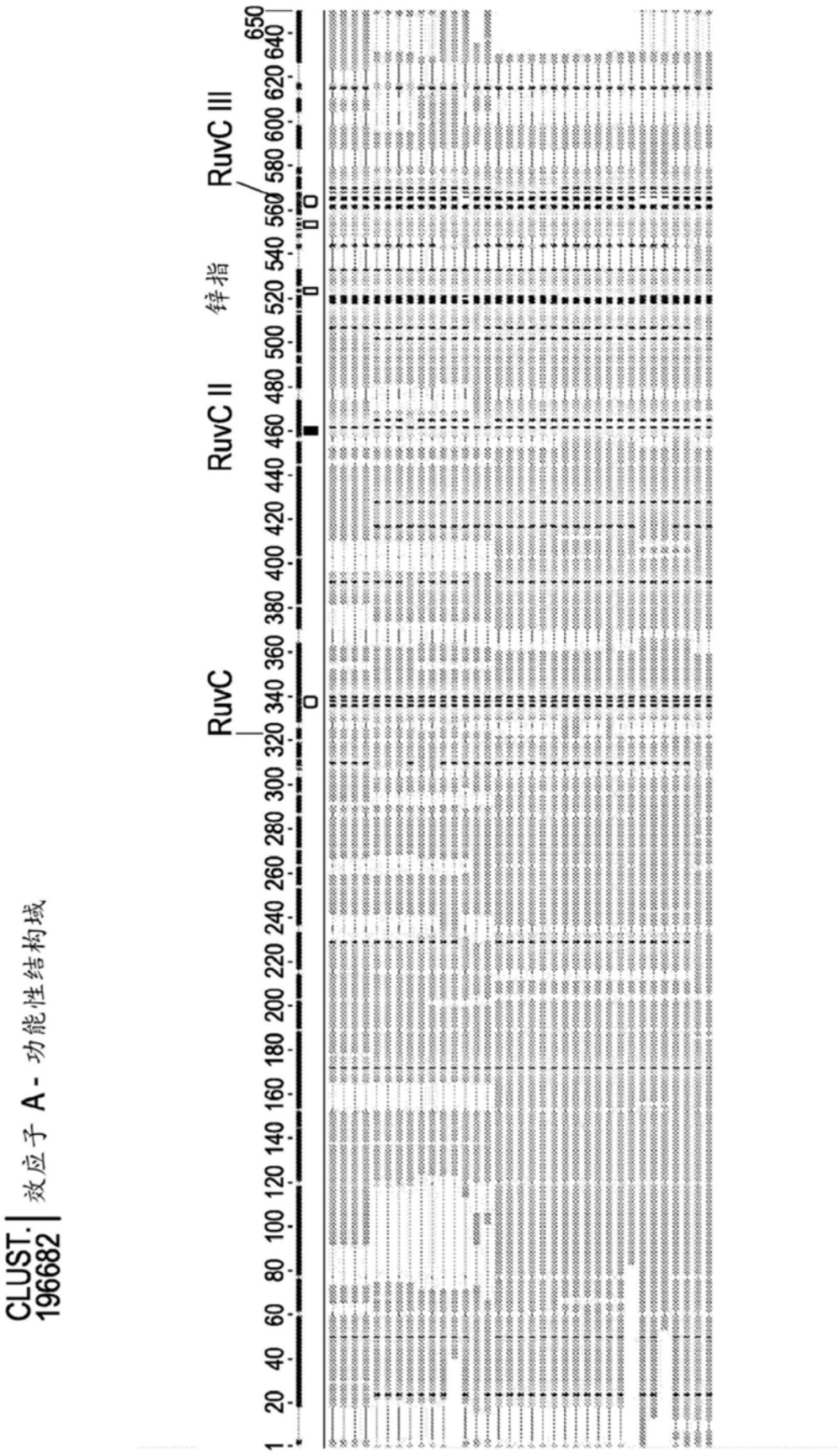

CLUST.
196682
效应子 A - 功能性结构域
RuvC
RuvC II
锌指
RuvC III
1 20 40 60 80 100 120 140 160 180 200 220 240 260 280 300 320 340 360 380 400 420 440 460 480 500 520 540 560 580 600 620 640 650

图12

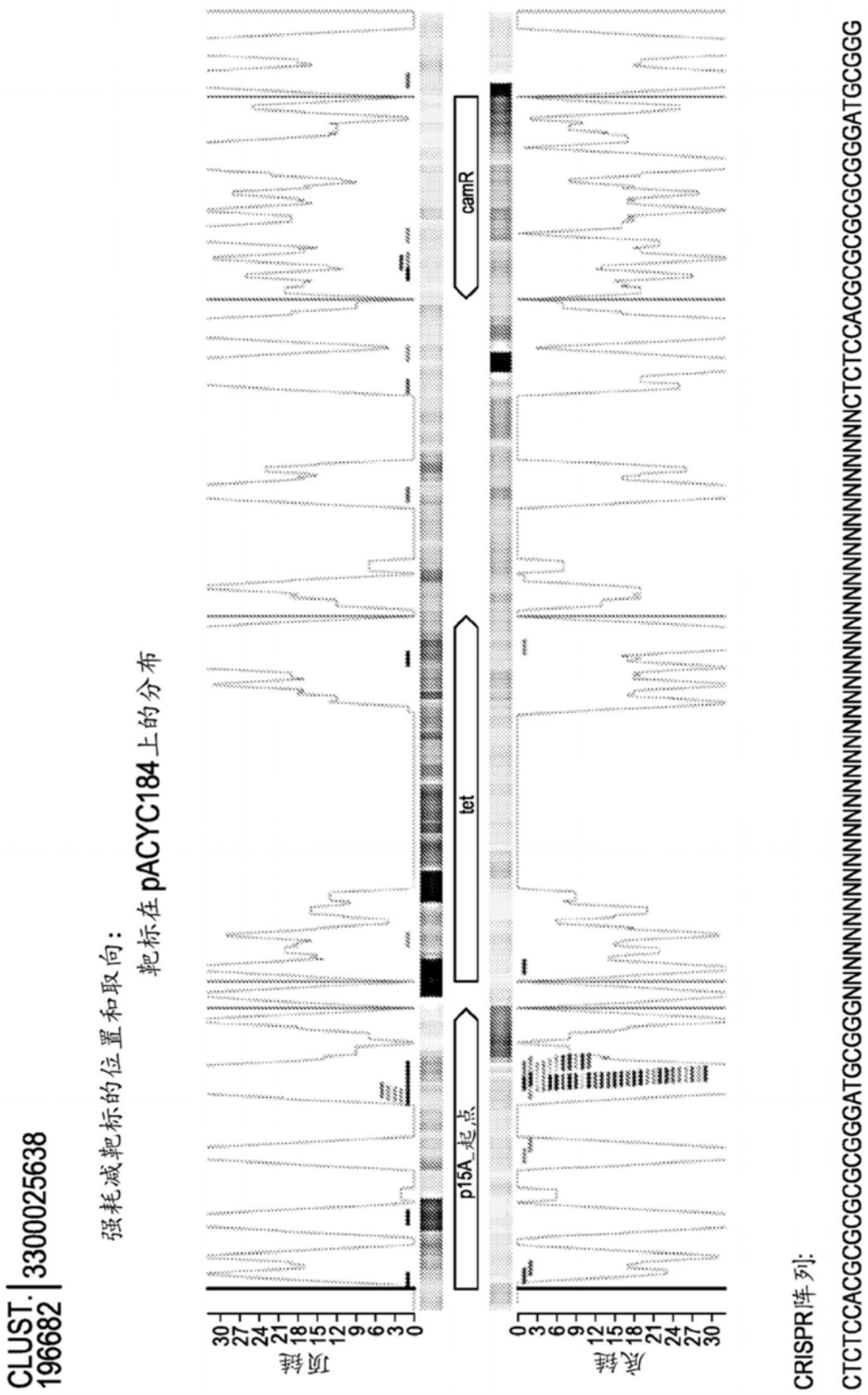
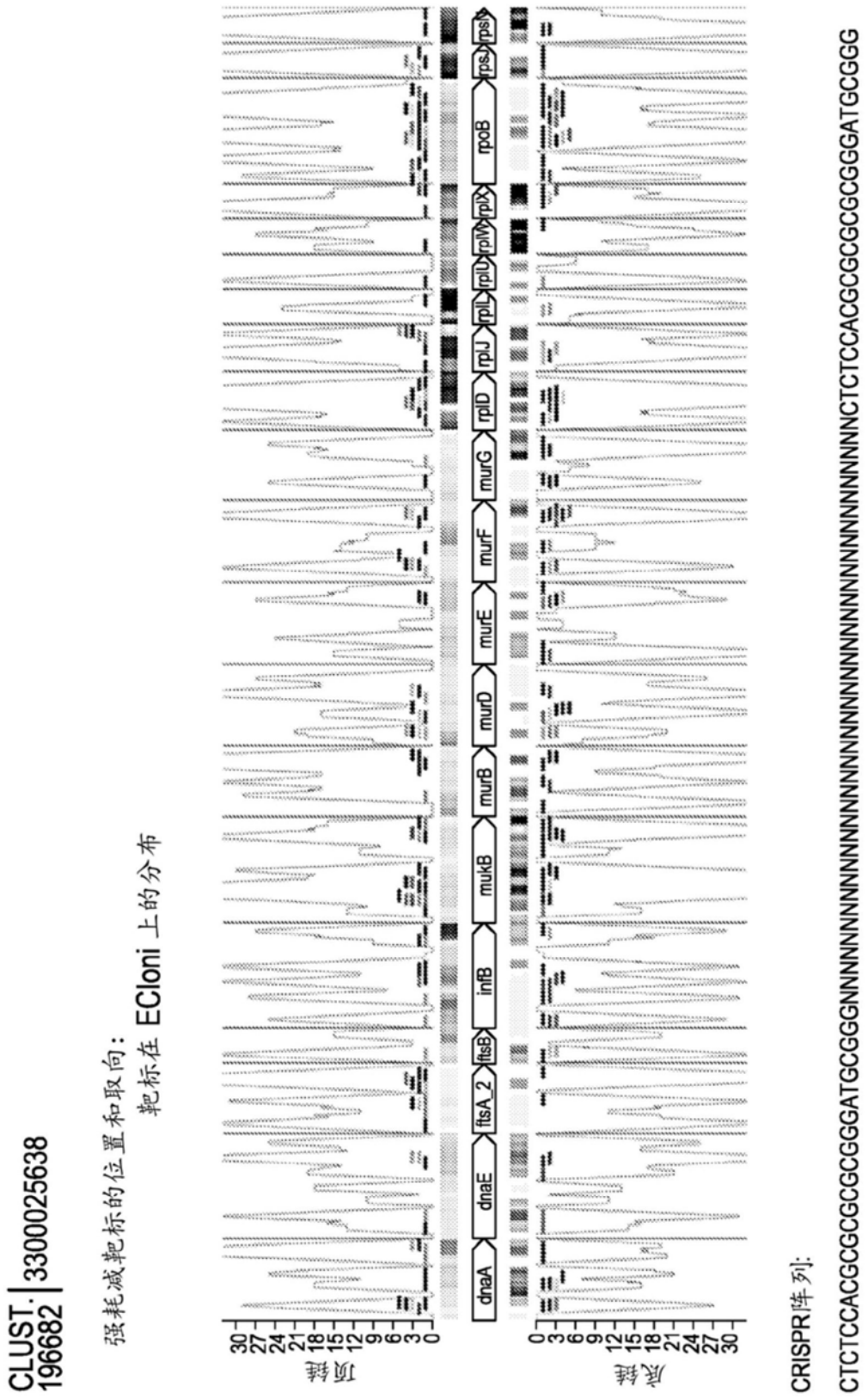

CLUST.
196682
3300025638
强耗减靶标的位置和取向：
靶标在 pACYC184 上的分布
顶链
30
27
24
21
18
15
12
9
6
3
0
p15A_起点
tet
camR
底链
0
3
6
9
12
15
18
21
24
27
30
CRISPR阵列:
CTCTCCACGCGCGCGGGATGCGGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCTCTCCACGCGCGCGGGATGCGGG

图13A


CLUST.
196682 | 3300025638
强耗减靶标的位置和取向：
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
rpsL
CRISPR阵列:
CTCTCCACGCGCGCGCGGGATGCGGGNNN

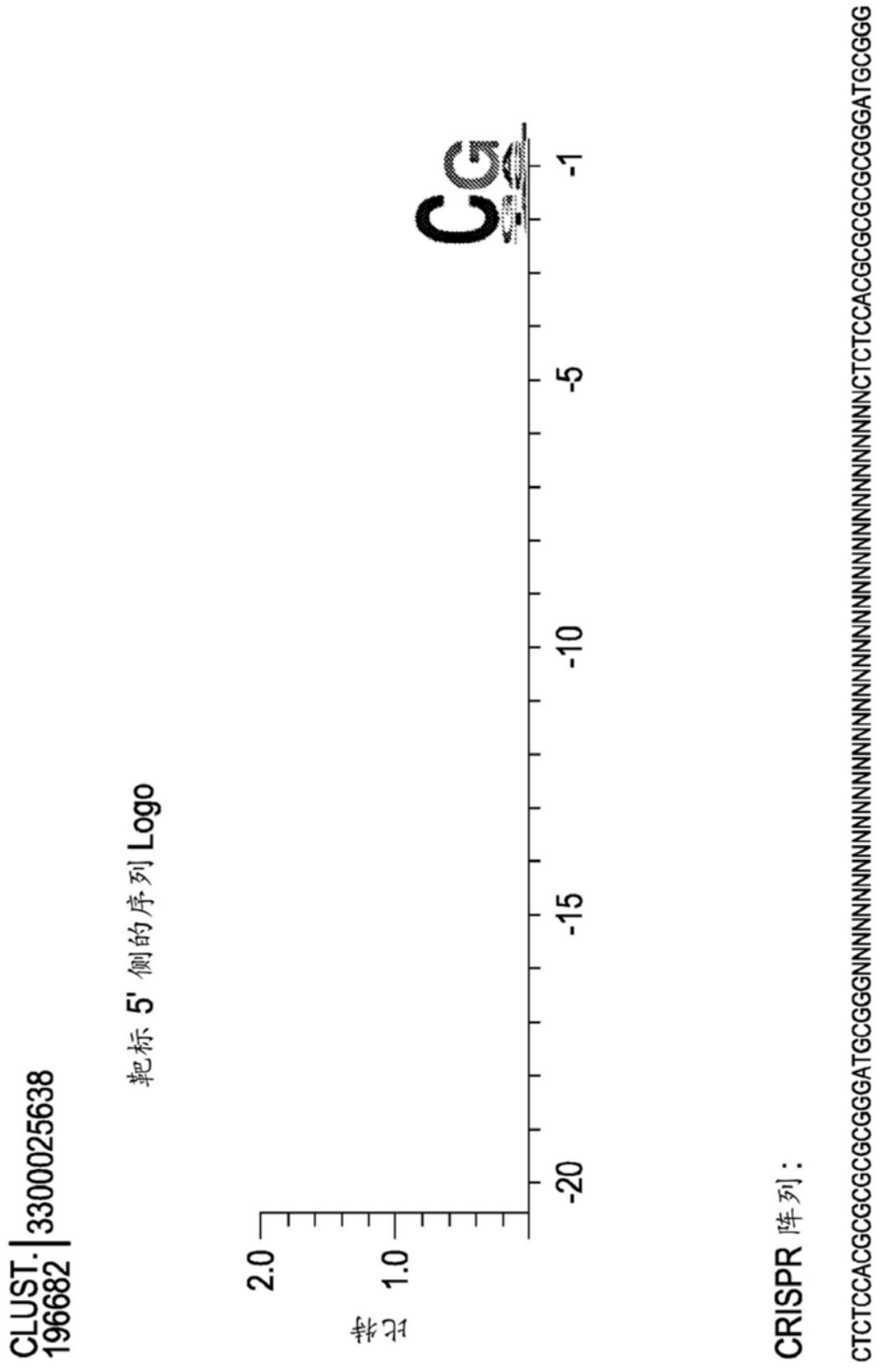

CLUST.
196682 | 3300025638
靶标 5' 侧的序列 Logo
比特
2.0
1.0
-20
-15
-10
-5
-1
CRISPR 阵列:
CTCTCCACGCGCGCGGGATGCGGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCTCTCCACGCGCGCGGGATGCGGG

图14


CLUST.
089537
组分
47,000
48,000
49,000
3300032069|Ga0315282_10002412|S.HR1Q9Lm
3300032118|Ga0315277_10001015|S.G1CNZTm
3300032118|Ga0315277_10012802|S.DWLF5Xm
激活的污泥宏基因组 |CAAACH010011569m
激活的污泥宏基因组 |CAAACI010012331m
激活的污泥宏基因组 |UZXW01002767m
生物反应器宏基因组|CAAACB010000011m
生物反应器宏基因组|CAAACX010000652m
生物反应器宏基因组|UOUK01008448m
生物反应器宏基因组|UOUL01003058m
菲西芬氏菌纲细菌 ST-NAGAB-D1|CP019791
菲西芬氏菌目细菌 |PMBN01000097m
菲西芬氏菌目细菌 |PMDC01000375m
菲西芬氏菌目细菌 |PMMM01000072m
浮霉菌门细菌 RBG_13_46_10|MHYD01...
SRR2080392_megahit_k177_207267|S.HVM1SJm
SRR4030068_megahit_k177_2163203|S.GMD1AFm
SRR5251193_megahit_k177_201695|S.DSZSF5p
SRR5251487_megahit_k177_199995|S.EREB5Kp
SRR5487326_megahit_k177_254484|S.ECZY6Um
SRR5487761_megahit_k177_404752|S.CXAVXXm
SRR5576005_megahit_k177_1292294|S.CDSB59m
SRR5818194_megahit_k177_802550|S.EWIWT1m
SRR6049666_megahit_k177_256604|S.CY1NZ4m
SRR6056591_megahit_k177_2733919|S.II9QB2m
SRR6263259_megahit_k177_232832|S.ICXWNNm
SRR6481359_megahit_k177_1860565|S.BVZN8Im
SRR6505097_megahit_k177_377355|S.FMM7ZXm
SRR6962291_megahit_k177_1754533|S.GKD7LLm
SRR6962291_megahit_k177_3367137|S.I1EHF4m
SRR6963308_megahit_k177_1876783|S.I1L86Pm
SRR6963352_megahit_k177_1225001|S.GMUXPKm
SRR6963480_megahit_k177_1813084|S.DW8OZ7m

图15(部分1)

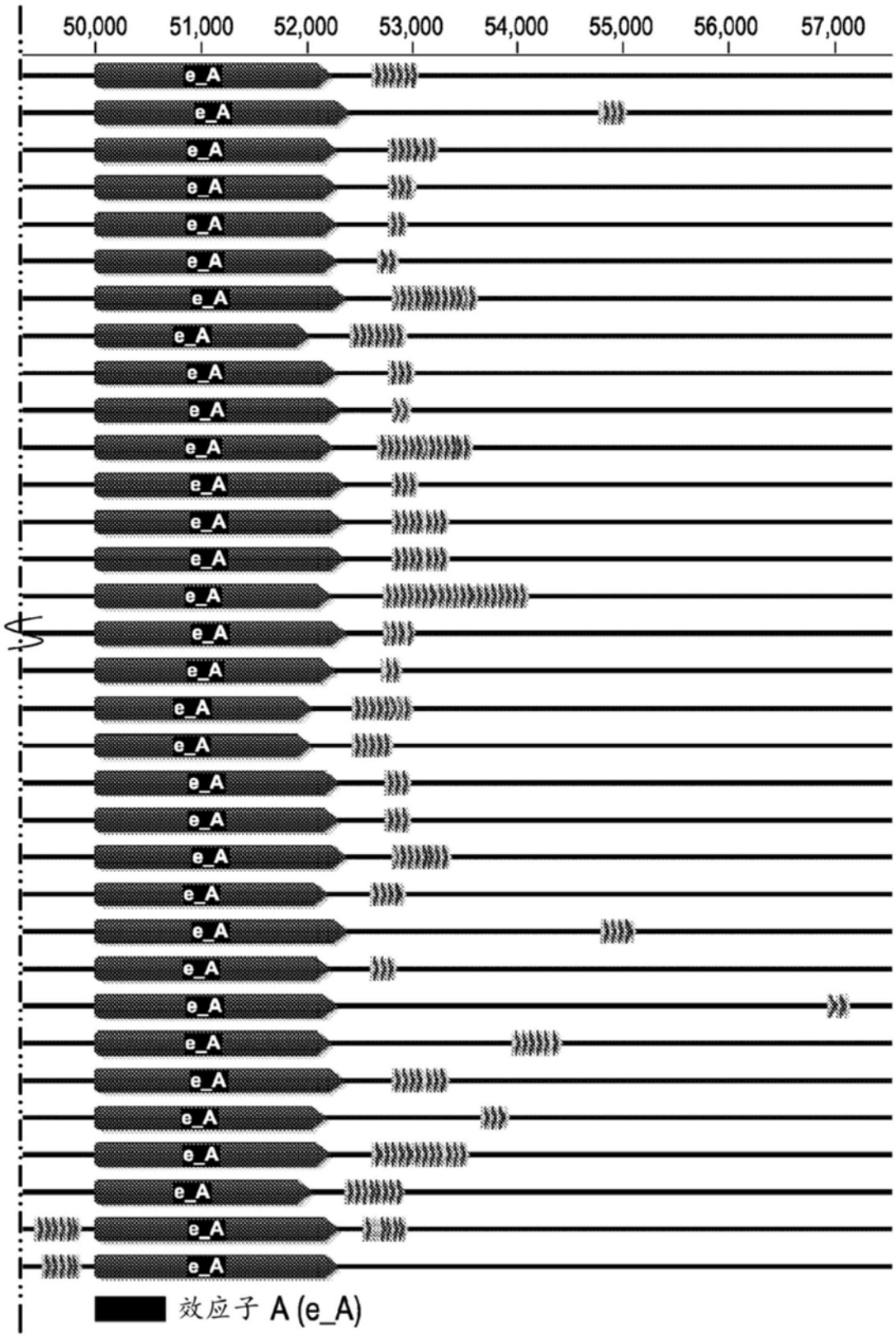


图15(部分2)


CLUST.
089537
直接重复
比对
1
10
20
30
40
43
GSYGBYGYR-NNNNSYBHSVVVDHNVHDRBBRRYKGACA-C-A
CGCCGCA-GCGGCCGACGCGGCCCTGATCGATGGACA-C
GCCGCCGCA-GCGGCCGACGCGGCCCTGATCGATGGACA-C
GCCGCA-GCGGCCGACGCGGCCCTGATCGATGGACA-C
GTCGCG-GCGGCCGATGCGATCCTGATCGATGGACA-C
GCCGCG-GAGGCCGACGCGGCCCTGATGGATGGACA-C
GCCGCG-TCCCCCGACGCGACCCTGATCGATGGACA-C
ATAGCG-CCCCCCGACGCAAATCTGATTAATGGACA-C
ATTGCA-CCGGCCAACGCAAATCTGATTGATGGACA-C
GGAGCG-CCCCCCGACGCGGCAATGATCGGCTGACA-C
GGAGCGCCCCCCCGACGCGGCCATGATCGACTGACG-G
GGCGTA-CCGCCTGACGGGTTTCTCATCGATGGACA-C
GGCGCA-CCGCCTGACGGGTTTCTCATCGATGGACA-C
GGCGCA-CCGCGTGGCGTGGAACTGACCGGCTGACA-C
GGCGCA-CCCGCTGGCGCGGAACAGACCGACTGACA-C
CTCTCA-ACCGCTTGGGCAGAACCGACCGACTGACA-C
GCCGCG-TCCCCCGACGCGCCCCTGATCGAT
GGCGCA-GCGGCCATCAGGATGGTGATCGGCTGACA-C
GGTGAA-GCGGCCATCGGGATCGTGATCGGCTGACA-C
GCCGCA-GCGGCCGACGCGGCCCTGA
CCTGTA-TCCGCGCGCGCGTTTGCGGCTGATGAACA-C
GATGTA-GCCGCACGTTGATTCACCGGCGATTACCA-C
ACCGCG-GCGATACCGGCTTTCCCGGGTGACTGACA-C
GGTACCGCG-GCGATACCGGCTTTCCCGGGTGACTGACA-C
CCCCCCCCGCCATTCTCGGCTGTGTACAA-C
CGGGTGTA-CCCCCCCCGCCATTCTCGGCTGTGTACAA-C
GGTGGA-CGTCCCCCGCGATTCCCGGGTGTGTACCA-CGA
GGTGGA-CGTCCCCCGCGATTCCCGGGTGTGTACCA-C
TATA-GGTACCCTAAAAATTCCGGGTAACTACAC-C
ACTATA-GGTACCCTAAAAATTCCGGGTAACTACAC-C
GCGGCG-CTGGCCAGCGCCGCGCTGA----------A
GTATTCATCCCCTTCAGAAAAGGAACTCTCTGGACA-C
GTGTGTTTCCTCTTAGAAAAACAAATGCTCTGGACA-C
GAAGAAAAGACCTCTCAAAAAAAATTCTCCTGGACA-C
GAAGAAAAAACCTCTT-AAAAAAATTCTCCTGGACA-C
GAAGAAAAAACCTCTT-AAAAAAATTCTCCTGGACA-C-T
CTGTCCCCCCCCGGACCTGACCAATAAGGGATTGACA-C
CTCTCACCCCAGTTCGATCGAGGAAGGCGACTGACA-C
GTGCATCGGCCGACGAAGAGCGAAGGTCAGTGACA-G
CGTGCATCGGCCGACGAAGAGCGAAGGTCAGTGACA-G
GTTGCA-----CCTCGCGCGCACCTGCGGGTTGACA-C
GTCGCA-----CCTCGCGCGTATGCACGGATTGGCA-G
GTCGCA-----CCCCGCGCGTGCGAGCGTGTTGACA-G
GTGACA-----TCTTATGCGTGTGTACAGGTGGACA-C
GTTGCA-----GTTCGCGCGTAGGCGTGAGCGGACA-C
GTTGCA-----ATCTGTGCGTATGTATGGGTGGACA-C
GTTGCA-----ACCTGTGCGTATGTGCGGATGGACA-C
GTCGCG-----ATCCGCACACCCGAGCGAGTCGACA-G
GGCGCA-----ACCCGCACACAACCGCGAATGGACA-C
GGCGTCGCG-----ATCCGCACACCCGAGCGAGTCGACAGC-A
GGCGCA----ACCCCGCACATGCGGGCGGGATGACA-G
GGGGGG--CAAGCTCGCGCGTATGCACGGATTGGCA-G
TCTCTCA-ATGGCTAAAGATCTAGTTCTAATTGGACACC
CTCTCA-ATGGCTAAAGATCTAGTTCTAATTGGACA-C
GTCGCA-ATGAGATAAAACATAGATATGTGTTGCAA-C
GGAGAG----------CGAAAAAAAGCGGGGGAGA-T

图16

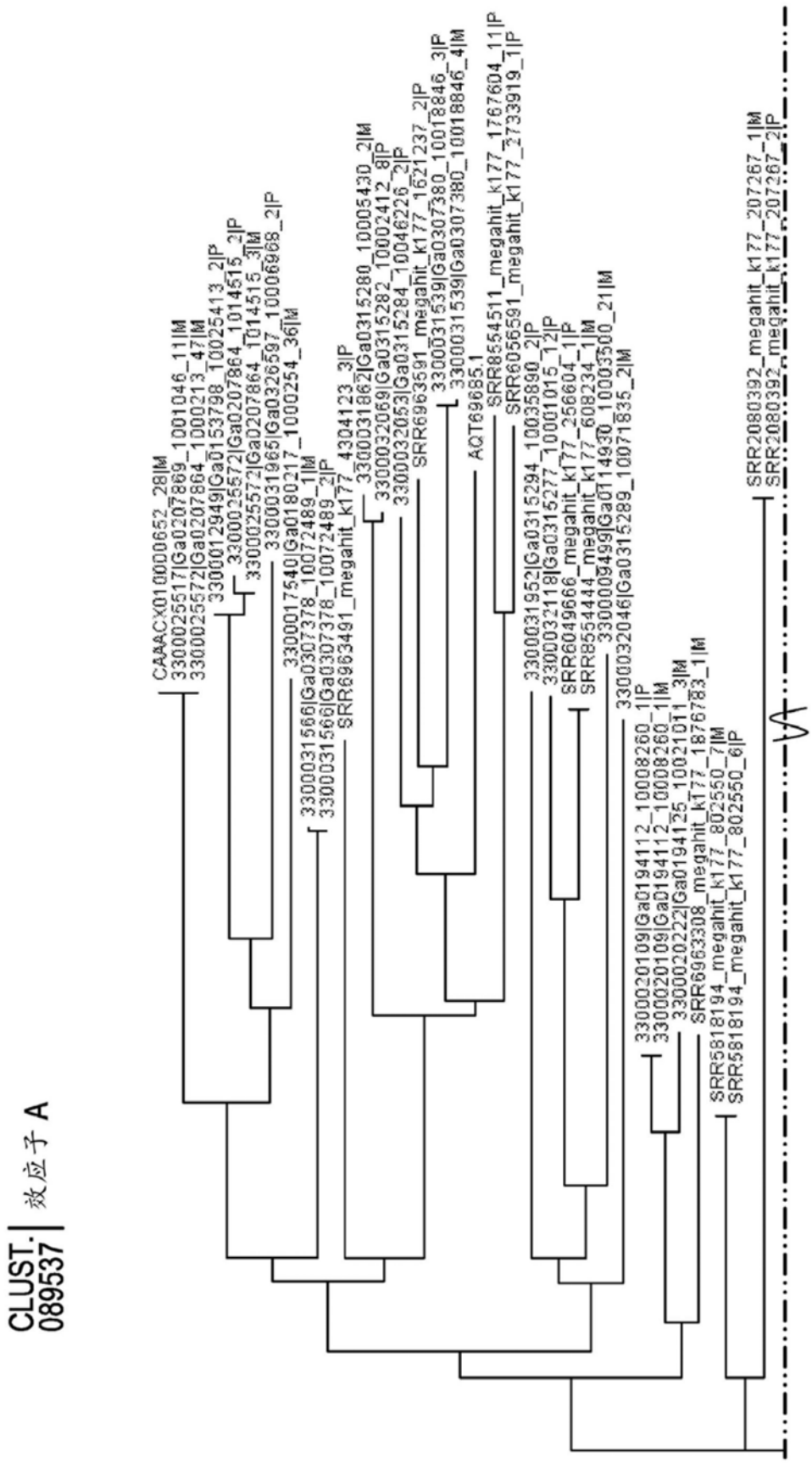

CLUST.
089537
效应子 A
CAAACX010000652_28|M
3300025517|Ga0207869_1001046_11|M
3300025572|Ga0207864_1000213_47|M
3300012949|Ga0153798_10025413_2|P
3300025572|Ga0207864_1014515_2|P
3300025572|Ga0207864_1014515_3|M
3300031965|Ga0326597_10006968_2|P
3300017540|Ga0180217_1000254_36|M
3300031566|Ga0307378_10072489_1|M
3300031566|Ga0307378_10072489_2|P
SRR6963491_megahit_k177_4304123_3|P
3300031862|Ga0315280_10005430_2|M
3300032069|Ga0315282_10002412_8|P
3300032053|Ga0315284_10046226_2|P
SRR6963691_megahit_k177_1621237_2|P
3300031539|Ga0307380_10018846_3|P
3300031539|Ga0307380_10018846_4|M
AQT69685.1
SRR8554511_megahit_k177_1767604_11|P
SRR6056591_megahit_k177_2733919_1|P
3300031952|Ga0315294_10035890_2|P
3300032118|Ga0315277_10001015_12|P
SRR6049666_megahit_k177_256604_1|P
SRR8554444_megahit_k177_608234_1|M
3300009499|Ga0114930_10003500_21|M
3300032046|Ga0315289_10071835_2|M
3300020109|Ga0194112_10008260_1|P
3300020109|Ga0194112_10008260_1|M
3300020222|Ga0194125_10021011_3|M
SRR6963308_megahit_k177_1876783_1|M
SRR5818194_megahit_k177_802550_7|M
SRR5818194_megahit_k177_802550_6|P
SRR2080392_megahit_k177_207267_1|M
SRR2080392_megahit_k177_207267_2|P

图17(部分1)

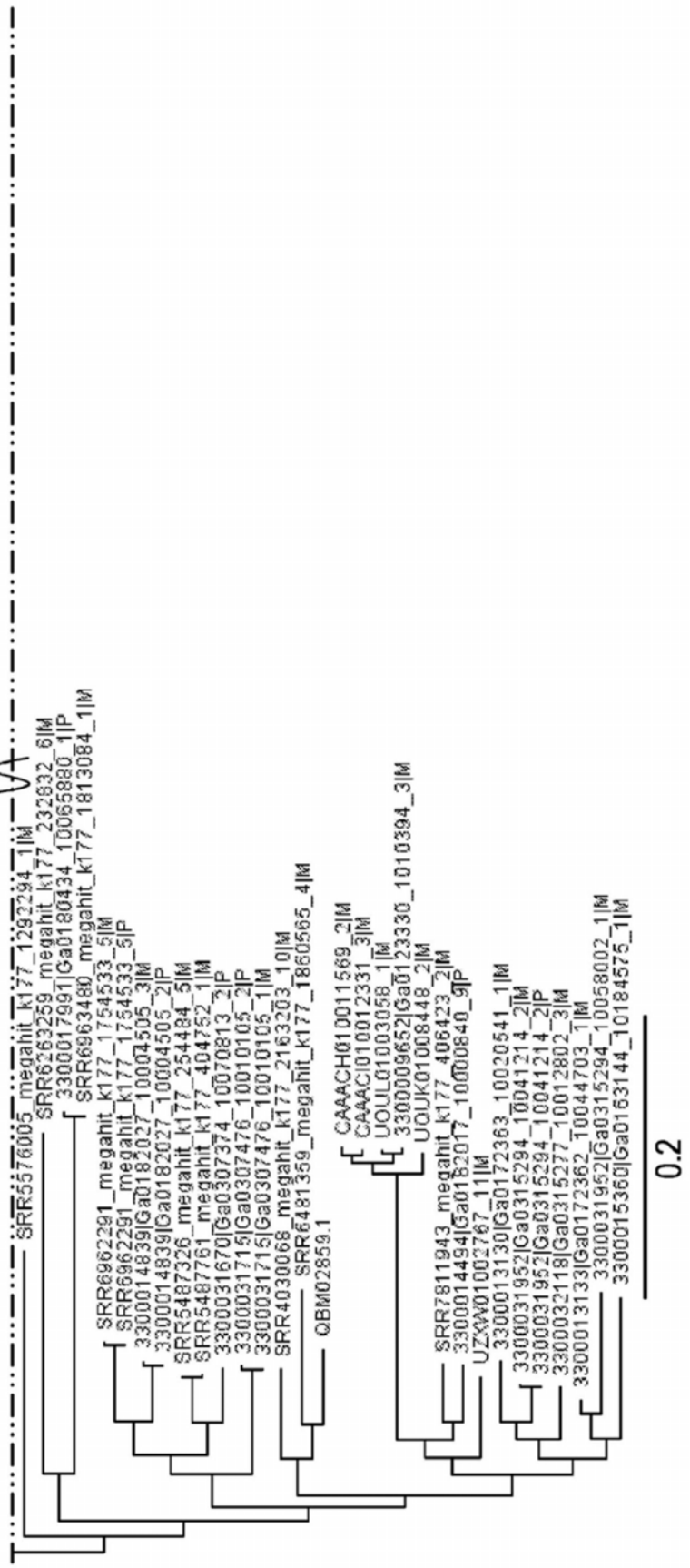

SRR5576005_megahit_k177_1292294_1|M
SRR6263259_megahit_k177_232832_6|M
3300017991|Ga0180434_10065880_1|P
SRR6963480_megahit_k177_1813084_1|M
SRR6962291_megahit_k177_1754533_5|M
SRR6962291_megahit_k177_1754533_3|P
3300014839|Ga0182027_10004505_3|M
3300014839|Ga0182027_10004505_2|P
SRR5487326_megahit_k177_254484_5|M
SRR5487761_megahit_k177_404752_1|M
3300031670|Ga0307374_10070813_2|P
3300031715|Ga0307476_10010105_2|P
3300031715|Ga0307476_10010105_1|M
SRR4030068_megahit_k177_2163203_10|M
SRR6481359_megahit_k177_1860565_4|M
QBM02859.1
CAAACH010011569_2|M
CAAACI010012331_3|M
UOUL01003058_1|M
3300009652|Ga0123330_1010394_3|M
UOUK01008448_2|M
SRR7811943_megahit_k177_406423_2|M
3300014494|Ga0182017_10000840_9|P
UZXW01002767_11|M
3300013130|Ga0172363_10020541_1|M
3300031952|Ga0315294_10041214_2|M
3300031952|Ga0315294_10041214_2|P
3300032118|Ga0315277_10012802_3|M
3300013133|Ga0172362_10044703_1|M
3300031952|Ga0315294_10058002_1|M
3300015360|Ga0163144_10184575_1|M
0.2

图17(部分2)

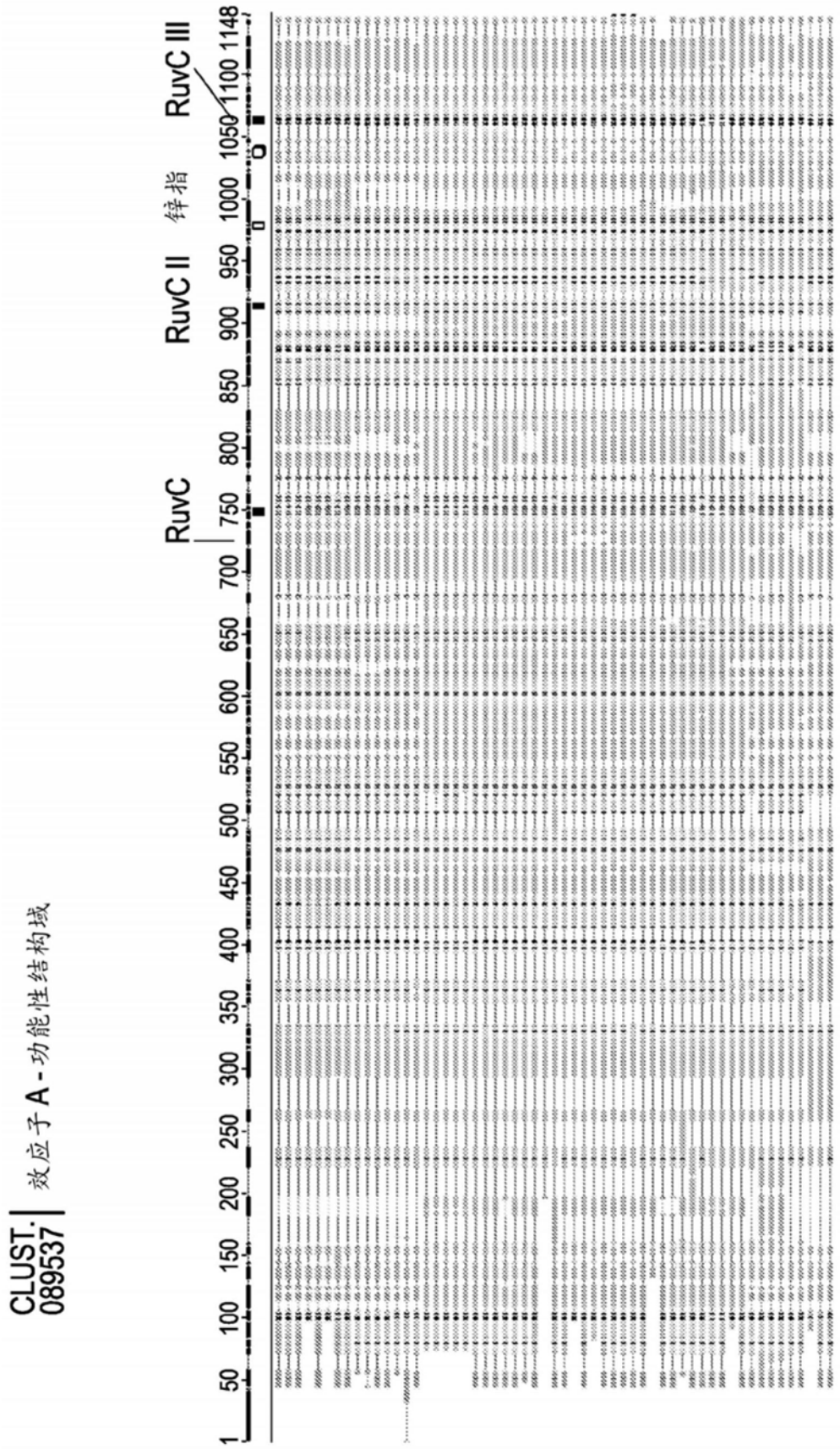

CLUST.
089537
效应子A-功能性结构域
RuvC
RuvC II
锌指
RuvC III
1
50
100
150
200
250
300
350
400
450
500
550
600
650
700
750
800
850
900
950
1000
1050
1100
1148

图18

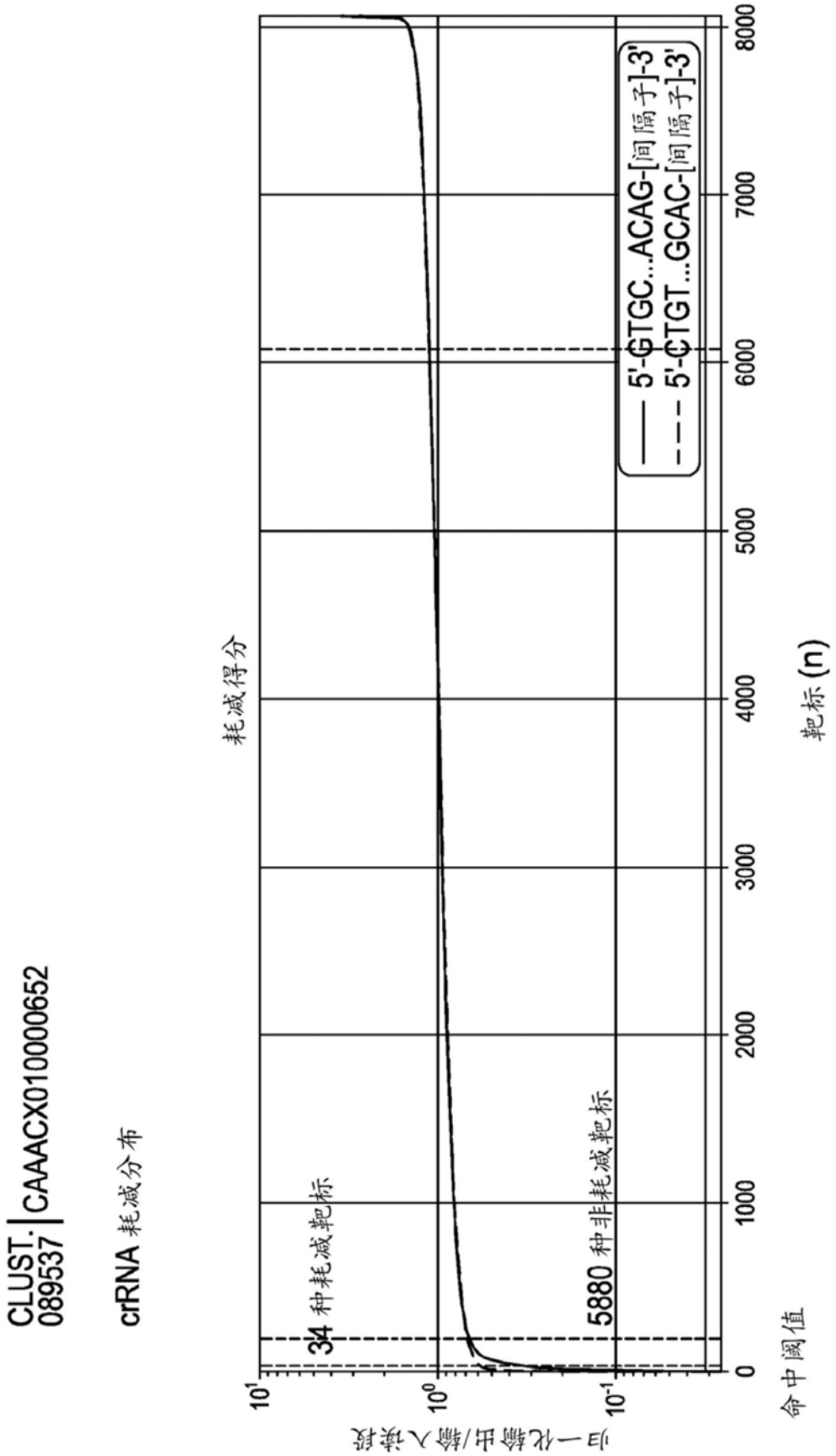

CLUST.
089537 | CAACX010000652
crRNA 耗减分布
34 种耗减靶标
5880 种非耗减靶标
归一化输出/输入读段
耗减得分
命中阈值
靶标 (n)
5'-GTGC...ACAG-[间隔子]-3'
5'-CTGT...GCAC-[间隔子]-3'

图19

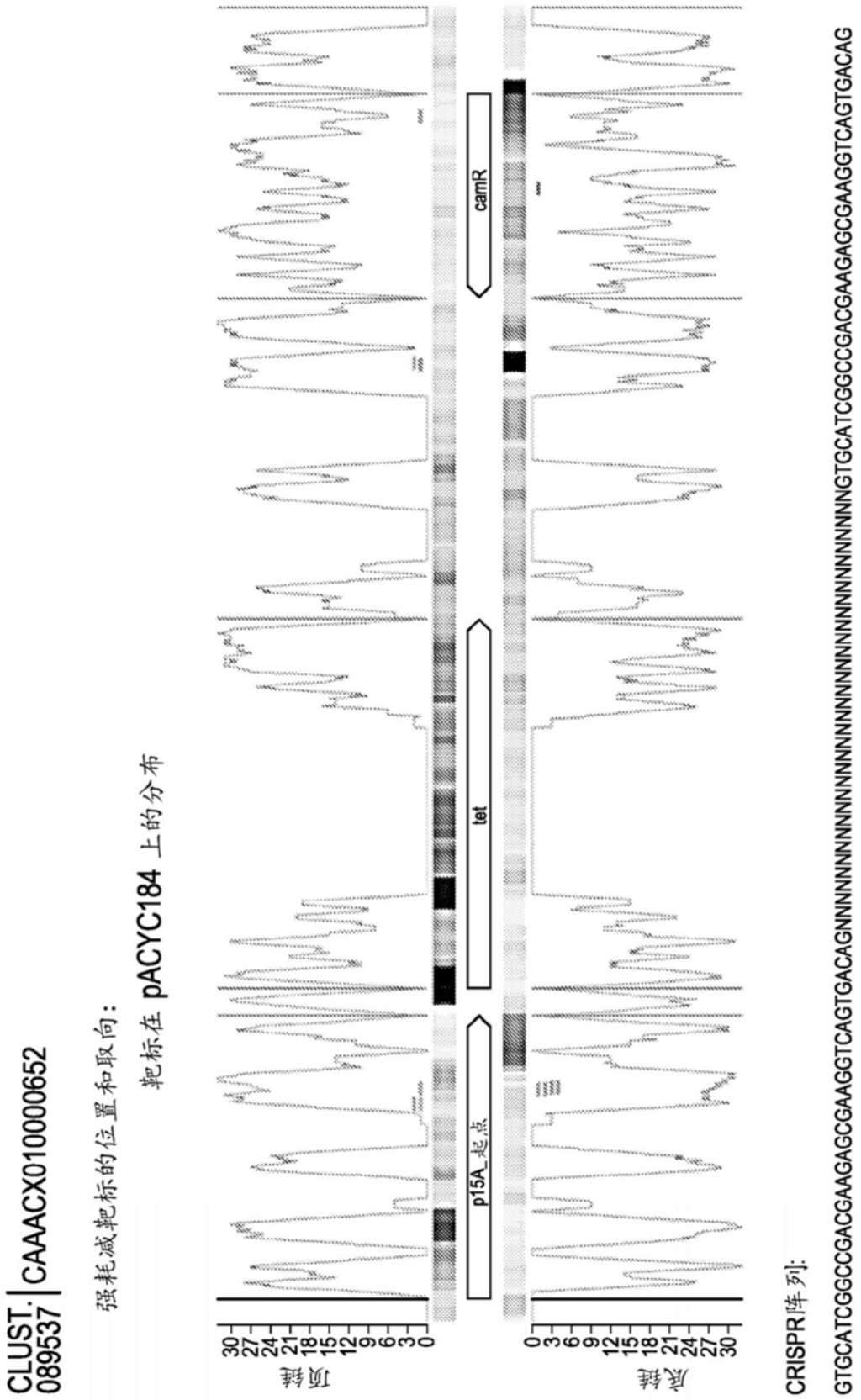
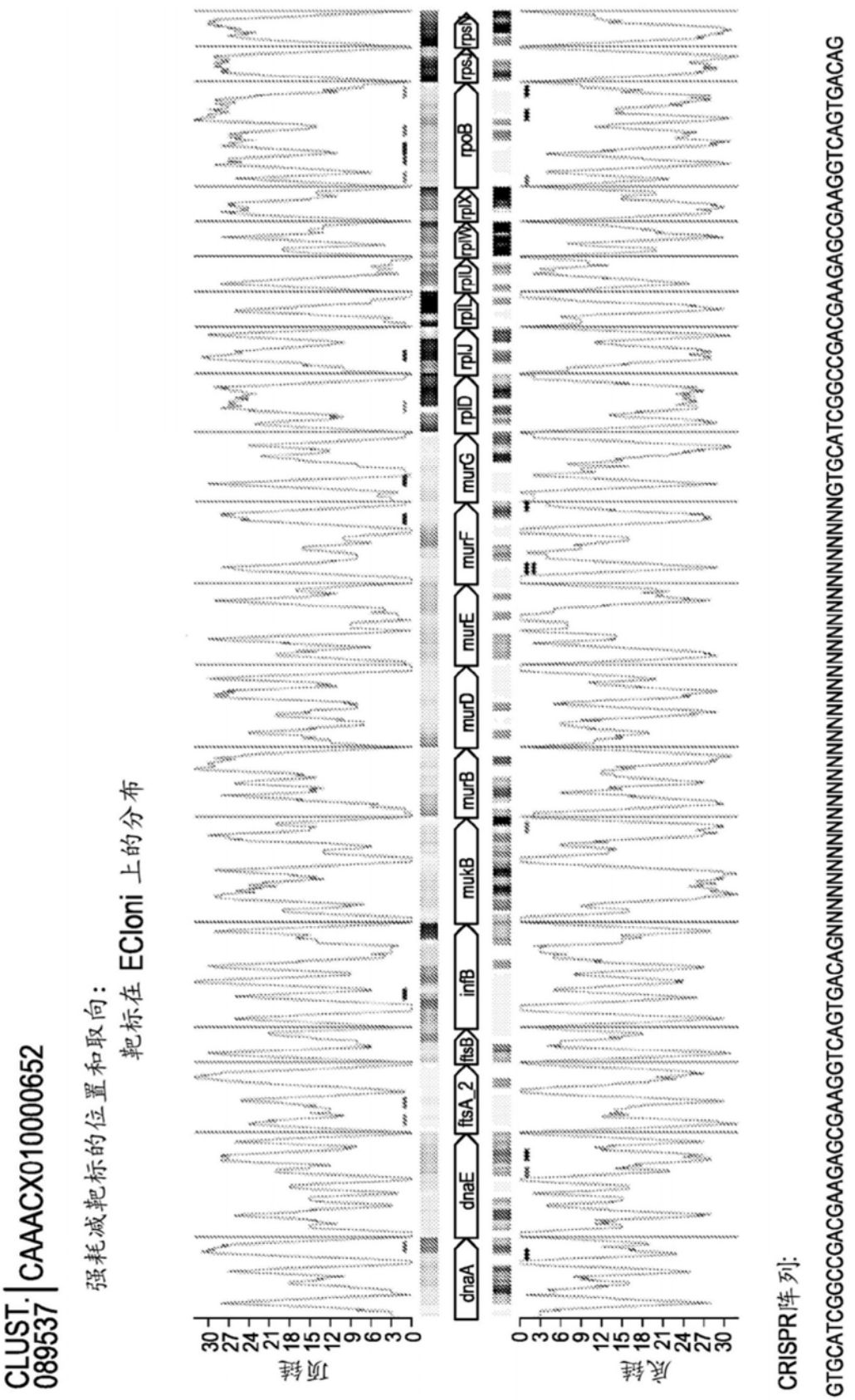

CLUST.
089537 | CAAACX010000652
强耗减靶标的位置和取向：
靶标在 pACYC184 上的分布
顶链
底链
p15A_起点
tet
camR
CRISPR阵列:
GTGCATCGGCCGACGAAGAGCGAAGGTCAGTGACAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGTGCATCGGCCGACGAAGAGCGAAGGTCAGTGACAG

图20A


CLUST.
089537 | CAAACX010000652
强耗减靶标的位置和取向：
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
rpsM
CRISPR阵列:
GTGCATCGGCCGACGAAGAGCGAAGGTCAGTGACAGNNN

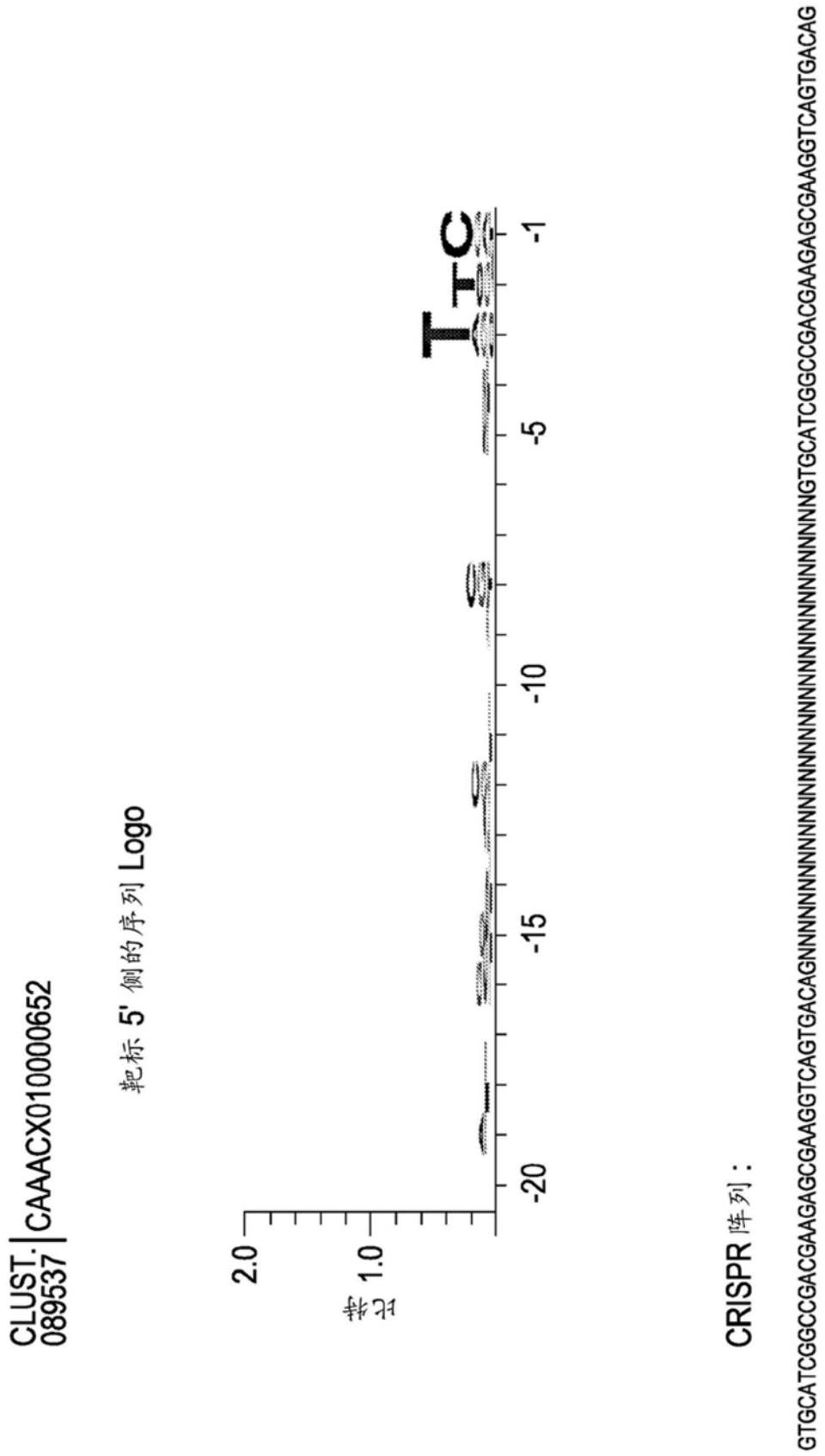

CLUST.
089537 | CAAACX010000652
靶标 5' 侧的序列 Logo
比特
2.0
1.0
-20
-15
-10
-5
-1
CRISPR 阵列：
GTGCATCGGCCGACGAAGAGCGAAGGTCAGTGACAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGTGCATCGGCCGACGAAGAGCGAAGGTCAGTGACAG

图21