(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 108220249 A
(43)申请公布日 2018.06.29

(21)申请号 201710328996.2

(22)申请日 2017.05.11

(66)本国优先权数据
201611139393.X 2016.12.12 CN

(71)申请人 上海交通大学医学院附属第九人民医院
地址 200011 上海市黄浦区制造局路639号

(72)发明人 张羽 单同领 冯希平 李飞
祝策 陈曦 叶玮 于阗

(74)专利代理机构 上海序伦律师事务所 31276
代理人 周东萍

(51)Int.Cl.
*C12N 7/00*(2006.01)
*C12Q 1/70*(2006.01)
*C12Q 1/686*(2018.01)
*C12R 1/92*(2006.01)

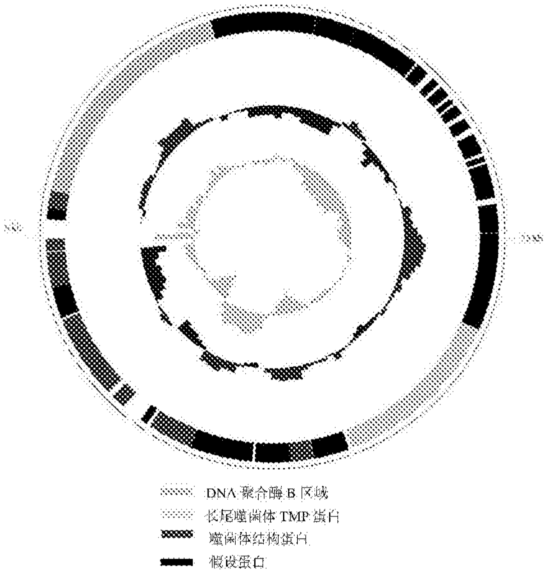
权利要求书1页 说明书29页 附图1页

(54)发明名称

# 长尾噬菌体及其获得方法和应用

(57)摘要

本发明公开了一种长尾噬菌体，该噬菌体包含：编码SEQ ID NO.1所示氨基酸序列的核苷酸序列。本发明还公开了上述长尾噬菌体的获得方法及应用。本发明的长尾噬菌体，是采用病毒宏基因组学方法从牙周炎患者口腔中分离获得的一株新病毒，为口腔疾病的病因、治疗及预防策略提供新的依据。



# 权 利 要 求 书

1.一种长尾噬菌体,其特征在于,其TMP蛋白的氨基酸序列如SEQ ID NO.3所示。

2.根据权利要求1所述的长尾噬菌体,其特征在于,其DNA聚合酶B区域蛋白的氨基酸序列如SEQ ID NO.4所示。

3.根据权利要求2所述的长尾噬菌体,其特征在于,其噬菌体结构蛋白的氨基酸序列如SEQ ID NO.5所示。

4.根据权利要求3所述的长尾噬菌体,其特征在于,该噬菌体包含:编码SEQ ID NO.1所示氨基酸序列的核苷酸序列。

5.根据权利要求4所述的长尾噬菌体,其特征在于,所述噬菌体具有SEQ ID NO.2所示的全长核苷酸序列。

6.一种长尾噬菌体,其特征在于,其DNA聚合酶B区域蛋白的氨基酸序列如SEQ ID NO.4所示。

7.根据权利要求6所述的长尾噬菌体,其特征在于,其噬菌体结构蛋白的氨基酸序列如SEQ ID NO.5所示。

8.一种长尾噬菌体,其特征在于,其噬菌体结构蛋白的氨基酸序列如SEQ ID NO.5所示。

9.根据权利要求8所述的长尾噬菌体,其特征在于,其TMP蛋白的氨基酸序列如SEQ ID NO.3所示。

10.一种包含权利要求1-9任一项所述长尾噬菌体的药物组合物或消毒液。

11.一种检测长尾噬菌体的试剂盒,其特征在于,包含特异性针对SEQ ID NO.2所示全部或部分核苷酸序列设计的引物对。

12.一种长尾噬菌体的获得方法,其特征在于,包括以下步骤:

取牙周炎患者牙龈组织样品,提取该样品中的病毒基因组;

将所述病毒基因组反转录成病毒cDNA,再用随机引物进行PCR扩增,扩增后的产物通过高通量测序,构建测序文库;

将高通量测序结果进行数据预处理后,把长度大于100bp的短序列拼接为重叠群;

将拼接的序列与数据库中的已知序列数据进行比对,找出序列之间的相似性;

分析比对结果,依据已知序列设计扩增引物,通过巢式PCR扩增,获得未知病毒的全长基因序列,该基因序列编码SEQ ID NO.1所示的氨基酸序列;

经同源性分析,对未知病毒进行物种分类,将该未知病毒归类为长尾噬菌体。

13.权利要求1-9任一项所述长尾噬菌体的应用,其特征在于,用于制备或筛选预防或治疗口腔疾病的药物。

# 长尾噬菌体及其获得方法和应用

## 技术领域

[0001] 本发明涉及口腔医学和病毒学技术领域，尤其涉及一种从牙周炎患者口腔中分离获得的长尾噬菌体及其获得方法和应用。

## 背景技术

[0002] 牙周病是一组复杂的感染性疾病，通常认为其由细菌刺激引起、由人体免疫等多因素参与。牙周病以牙槽骨吸收、牙齿松动及脱落为主要临床症状，已经成为成年人失牙的主要原因。重度牙周炎不仅会影响口腔健康，还与糖尿病、心血管疾病、低体重新生儿及早产等系统性疾病相关，对全身健康产生严重的负面影响。传统的牙周病病因学研究通常认为牙周病的发病与特定种类的细菌感染相关，然而细菌致病理论却并不能完整地解释牙周病的各种临床特征。例如，Umeda等研究发现，牙周致病菌广泛存在于唾液之中，接触所有牙齿，但是牙周病却常常不是全口发病，而仅影响有限数目的牙齿，具有位点特异性；尽管牙周环境中总有不同种类和数目的牙周致病菌持续感染和定植，但是牙周病却通常是短期发病，有自限性；病毒作为另一种潜在的牙周病致病因素逐渐受到关注，牙周病发病过程中，与病毒相关的牙周致病菌的粘附定植、病毒导致的宿主抗牙周致病菌的免疫功能抑制、与牙周病相关的病毒激活等机制可相对合理地解释菌斑量与牙周炎严重程度、进展速度的关系。病毒宏基因组学作为一项新兴技术能充分挖掘特定环境中的病毒群落和未知病毒。

[0003] 长尾病毒科是一种双股DNA病毒，只会感染细菌，外表形态具有长无收缩性的尾鞘。人类口腔中的噬菌体大多是属于有尾噬菌体目的长尾噬菌体科(通常都是溶原性噬菌体且有中等宽度的宿主范围)，肌尾噬菌体科(主要是裂解性噬菌体且宿主范围稍微宽一些)以及短尾噬菌体科(主要是裂解性噬菌体且宿主范围稍微窄一些)。关于噬菌体在牙周病的发生发展中的地位和作用，目前的相关研究并不很多，而且尚有争论。有研究者在比较了牙周炎患者和牙周健康人群的唾液及菌斑内的噬菌体数量后发现，牙周炎患者的菌斑内的噬菌体数量及种类与牙周健康者的菌斑相比有很大不同，且与牙周状态相关，而牙周健康人群内部的噬菌体结构则相对稳定(Ly,M.,et al.,Altered oral viral ecology in association with periodontal disease.MBio,2014.5(3):p.e01133-14)，由此认为噬菌体与牙周炎的发病可能密切相关。

## 发明内容

[0004] 本发明要解决的技术问题是提供一种长尾噬菌体，该病毒是从牙周炎患者牙周袋内壁牙龈上皮中分离获得的新病毒，为口腔疾病的治疗与预防提供新的解决方案。

[0005] 此外，还需要提供一种上述长尾噬菌体的获得方法和应用。

[0006] 为了解决上述技术问题，本发明通过如下技术方案实现：

[0007] 在本发明的一个方面，提供了一种长尾噬菌体，其TMP蛋白的氨基酸序列如SEQ ID NO.3所示。

[0008] 优选的，所述长尾噬菌体DNA聚合酶B区域蛋白的氨基酸序列如SEQ ID NO.4所示。

更优选的，其噬菌体结构蛋白的氨基酸序列如SEQ ID NO.5所示。进一步优选的，该噬菌体包含：编码SEQ ID NO.1所示氨基酸序列的核苷酸序列。最优选的，该噬菌体具有SEQ ID NO.2所示的全长核苷酸序列。

[0009] 在本发明的另一方面，还提供了一种长尾噬菌体，其DNA聚合酶B区域蛋白的氨基酸序列如SEQ ID NO.4所示。

[0010] 优选的，其噬菌体结构蛋白的氨基酸序列如SEQ ID NO.5所示。

[0011] 在本发明的另一方面，还提供了一种长尾噬菌体，其噬菌体结构蛋白的氨基酸序列如SEQ ID NO.5所示。

[0012] 优选的，其TMP蛋白的氨基酸序列如SEQ ID NO.3所示。

[0013] 在本发明的另一方面，还提供了一种包含上述长尾噬菌体的药物组合物。

[0014] 在本发明的另一方面，还提供了一种包含上述长尾噬菌体的消毒液。

[0015] 在本发明的另一方面，还提供了一种检测长尾噬菌体的试剂盒，包含特异性针对SEQ ID NO.2所示全部或部分核苷酸序列设计的引物对。利用该试剂盒，通过简单的PCR反应就可以检测出本发明的长尾噬菌体。

[0016] 在本发明的另一方面，还提供了一种长尾噬菌体的获得方法，包括以下步骤：

[0017] 取牙周炎患者牙龈组织样品，提取该样品中的病毒基因组；

[0018] 将所述病毒基因组反转录成病毒cDNA，再用随机引物进行PCR扩增，扩增后的产物通过高通量测序，构建测序文库；

[0019] 将高通量测序结果进行数据预处理后，把长度大于100bp的短序列拼接为重叠群；

[0020] 将拼接的序列与数据库中的已知序列数据进行比对，找出序列之间的相似性；

[0021] 分析比对结果，依据已知序列设计扩增引物，通过巢式PCR扩增，获得未知病毒的全长基因序列，该基因序列编码SEQ ID NO.1所示的氨基酸序列；

[0022] 经同源性分析，对未知病毒进行物种分类，将该未知病毒归类为长尾噬菌体。

[0023] 在本发明的另一方面，还提供了一种上述长尾噬菌体的应用，用于制备或筛选预防或治疗口腔疾病的药物。

[0024] 本发明的长尾噬菌体，是采用病毒宏基因组学方法从牙周炎患者牙周袋内壁牙龈上皮中分离获得的一株新病毒，为口腔疾病的病因、治疗及预防策略提供新的依据。

## 附图说明

[0025] 下面结合附图和具体实施方式对本发明作进一步详细的说明。

[0026] 图1是本发明实施例1长尾噬菌体Siphoviridae 29632的基因结构图。

## 具体实施方式

[0027] 下列实施例中，未注明具体条件的实验方法，通常按常规条件，如《精编分子生物学实验指南》(F.M.奥斯伯，R.E.金斯顿，J.G.塞德曼等主编，马学军，舒跃龙的译.北京：科学出版社，2004)中所述的方法进行。

[0028] 传统的牙周病病因学研究焦点多为牙周致病细菌，而细菌致病理论却不足以完整解释牙周病相关的临床症状，本发明将关注转移到牙周病发生、发展过程中的病毒因素，力图揭示牙周环境中的病毒群落，筛选出牙周病的可疑致病病毒，为后续治疗与预防牙周病

提供基础。由于病毒特殊的生物学特性，传统的病毒鉴定方法已经无法满足要求，本发明首次将病毒宏基因组学、高通量测序及生物信息学等技术应用于牙周病的病因学研究，建立了一个分析牙周环境中的病毒群落以及诊断牙周环境中未知噬菌体的病毒宏基因组学技术平台。

[0029] 本发明采用病毒宏基因组学方法从牙周炎患者牙周袋内壁牙龈上皮组织中钓取疑似新病毒的序列，经序列拼接、基因步移、PCR扩增，克隆出新病毒的全基因组序列，全长29632bp，经同源性分析，多个片段与Siphovirus contig89片段同源性较高，该新病毒属于长尾病毒科，将该新病毒命名为长尾噬菌体Siphoviridae 29632(GenBank:KY053532)。

[0030] 实施例1采用病毒宏基因组学方法从牙周炎患者口腔中分离获得长尾噬菌体的全长基因序列

[0031] 取牙龈组织约2g，放入2ml离心管中，加入1mlPBS及3粒直径0.3cm钢珠后封口膜封口，后进行匀浆(反复冻融三次并匀浆5min)。4℃条件下，13000g离心10分钟，抽取上清样品，再次4℃条件下，12000g离心5分钟，抽取上清样品，两次离心去除细胞碎片和其它微生物；通过0.45m滤器处理上清样品，7500rpm离心1min过滤，再次去除悬液中的非病毒壳粒；37℃下水浴90min，应用Turbo DNase、RNase、Baseline Zero Dnase和Benzonase去除病毒颗粒包裹外的遗传物质。目前已处理病例组66份样品，对照组32份样品。并且提取其遗传物质(包括DNA、RNA、dsDNA、ssDNA、dsRNA和ssRNA等)。用QIAamp Viral RNA extraction Kit提取样品中的病毒基因组，通过DNA酶去除DNA病毒的遗传物质及反转录和Klenow Fragment酶，提取完病毒基因组后立刻使用SuperScriptⅢReverse Transcriptase试剂盒进行反转录操作，反转录的引物为Fixed-Primer-8N，该引物的设计是在网站(http://www.changbioscience.com/primo/primor.html)上完成，在网站中产生随机引物，随后通过与NCBI的数据库进行BLAST比对，挑选其中不能够匹配到人或者细菌的引物作为Fixed-Primer；向提取的病毒基因组12μl中加入1μl 100pmol的Fixed-Primer-8N，上下颠倒5-6次混合均匀；70℃水浴5min；冰上静置2min；后进行反转录；反转录产物置于94℃下5min变性，冰上冷却2min；后向上述产物中加入1μl的Klenow酶进行第二链的合成反应，合成病毒双链cDNA，同时将引物标签分别加在不同样品的病毒双链cDNA上；通过单引物PCR扩增病毒遗传物质，PCR扩增的引物为Fixed-Primer-8N对应的Fixed-Primer引物，模板即为上一步中得到的双链cDNA模板；PCR产物经过1％的琼脂糖凝胶电泳，切取满足高通量测序的DNA；设置48个复孔，并设定电压为100伏特。经90分钟电泳后得到smear条带，分别切取48个复孔对应长度范围在500-1000bp的凝胶，置于2ml离心管，称重。产物经PCR Purification Kit纯化，电泳后切胶回收；根据Miseq高通量测序文库构建试剂盒的操作说明，构建Miseq测序文库，送至测序公司测序。测序结果由斯坦福大学基因组工程中心进行数据处理和序列分析，依据引物标签将各条序列信息分类至各个样品，随后去除引物及标签。将长度大于100bp的短序列经NEWBLER2.5软件拼接为重叠群(contigs)。下载NCBI数据库中的生物遗传信息，建立本地BLAST文库，并将拼接和翻译的核苷酸和氨基酸序列在本地文库和NCBI数据库中进行多重BLAST比对。设定E-value≤0.00001为临界参数，将比对结果依据E-Value值的大小分为真核生物、细菌、病毒和未知序列。比对结果上传至可视化网站，分析病毒序列信息，发现疑似全新噬菌体序列。

[0032] 分析比对结果，运用geneious 8.1软件(美国)对文库中的噬菌体片段进行拼接，

得到噬菌体序列后，设计34对验证引物扩增(见下表1)。PCR扩增后，产物经凝胶电泳、用AxyPrep DNAGel Extraction Kit(Axygene,Silicon Valley,USA)切胶回收、克隆后测序。测序后的序列经过载体的剪切和拼接，找出插入碱基和突变位点，并进行纠错。克隆出的基因序列通过blast比对分析其编码区(Coding Sequence,CDS)，结果如图1所示，新噬菌体的全长基因序列为29632bp(SEQ ID NO.2)，其编码的氨基酸序列如SEQ ID NO.1所示(2499AA)，其中，第6085-7005位核苷酸序列编码的306AA的氨基酸序列(SEQ ID NO.3)为长尾噬菌体TMP蛋白；第20599-20970位核苷酸序列编码的123AA的氨基酸序列(SEQ ID NO.4)为DNA聚合酶B区域；第21842-22369位核苷酸序列编码的175AA的氨基酸序列(SEQ ID NO.5)为噬菌体结构蛋白，这几个基因及蛋白在噬菌体中发挥着重要作用。

[0033] 克隆出的基因序列通过比对分析编码区后，再经同源性分析，多个片段与Siphovirus contig89片段同源性较高，该新噬菌体属于长尾病毒科。将该新噬菌体命名为Siphoviridae 29632(GenBank:KY053532)，全长29632bp，属于长尾噬菌体。

[0034] 表1 34对引物序列

[0035]

| | 上游 | 下游 |
|---|---|---|
| 1 | GAGGACACTCCGAACGAGAC（SEQ ID NO.6） | ATCCCCGAAAGTCAACCGAC（SEQ ID NO.7） |
| 2 | ATAATGTTGCGGAAGCCGCT（SEQ ID NO.8） | TTTCGTGATACGGGCCAGA（SEQ ID NO.9） |
| 3 | ACGATGCGGCGACGAAGAT（SEQ ID NO.10） | CAGCGACAGCTTCGAGTCGTT（SEQ ID NO.11） |
| 4 | CCTTCCGGGACATGAAGGAC（SEQ ID NO.12） | CTGCGAGCACCAGTATAGCA（SEQ ID NO.13） |
| 5 | AAGAACGCCATCGACCAGTT（SEQ ID NO.14） | CGAACGTCGCAATGATCACC（SEQ ID NO.15） |
| 6 | ATCCTGCTGGACAACATCCC（SEQ ID NO.16） | TTCCGAATAGCGTCAGCGAA（SEQ ID NO.17） |
| 7 | ATCGCGCTTATTCAGGTGCT（SEQ ID NO.18） | TTCGGACGATCGCTGTCATT（SEQ ID NO.19） |
| 8 | GGGCCTCTCCGAATCTTACG（SEQ ID NO.20） | AGTCAGGTGCTGATTGCGAA（SEQ ID NO.21） |
| 9 | GAGTTCGACTGGGAGGACAC（SEQ ID NO.22） | TGTTGTCGCCGAAATCGTTG（SEQ ID NO.23） |
| 10 | AGCGATGACCCCTGGTTCTA（SEQ ID NO.24） | GAGCGAGTGTACTCGGTGAG（SEQ ID NO.25） |
| 11 | GACTTCTTCATGGGCGACCT（SEQ ID NO.26） | GATCTACCAGAACCTGCCCG（SEQ ID NO.27） |
| 12 | CGTTCCAGGAATGGGCGAAG（SEQ ID NO.28） | CCGCATCGTACATCAGGGAG（SEQ ID NO.29） |
| 13 | CAGGACCACAAGGATGCGAT（SEQ ID NO.30） | CGACGAACTTCAGACGGTCA（SEQ ID NO.31） |
| 14 | TACTACCATCTCACCCGCCA（SEQ ID NO.32） | TCAGTCACGCGCTTCTTTCT（SEQ ID NO.33） |
| 15 | ATGAGCCCACACTCGAAAGG（SEQ ID NO.34） | CAAGGCGCTCTTCTCTCGAA（SEQ ID NO.35） |
| 16 | CTTGTGATACTTGCGCTGGC（SEQ ID NO.36） | TGAGACCCCTCGAAAATGCC（SEQ ID NO.37） |
| 17 | CTGGTCAACGACCCCAAGAA（SEQ ID NO.38） | CCGTCATAGCACAGAACCGT（SEQ ID NO.39） |
| 18 | GTGTACGAGATGGTGGTCGG（SEQ ID NO.40） | GTGTCGAGCTCCTTCATCCC（SEQ ID NO.41） |
| 19 | CTGCTGAAGCTCTACGCTCC（SEQ ID NO.42） | TCGTACTCGTGCATGTCGTC（SEQ ID NO.43） |
| 20 | AGGATGACGACATGCACGAG（SEQ ID NO.44） | ATCATGAAGTCCGTGGGCTC（SEQ ID NO.45） |
| 21 | AAGAAGGATGGTTCGGCCTG（SEQ ID NO.46） | GTACTCGTCCCGGCAGTATG（SEQ ID NO.47） |
| 22 | ACACGTACGAGCGTCTGATG（SEQ ID NO.48） | TGGATCTCCTTCCGGAGGTT（SEQ ID NO.49） |
| 23 | AGTCTCCGCGACCTTATCCT（SEQ ID NO.50） | CGTACACGTAACCGCCTTCT（SEQ ID NO.51） |
| 24 | ACAAGTACGAGTTCGGCGTT（SEQ ID NO.52） | AGCGAGATATCGCCGTTCAG（SEQ ID NO.53） |

[0036]

| | | |
|---|---|---|
| 25 | CGTCCGGAAGCAAGGGATAC（SEQ ID NO.54） | GCGGAAGTGCATGAGGTAGA（SEQ ID NO.55） |
| 26 | TTCCGCGTCATCTGAAAGGA（SEQ ID NO.56） | ATATGACGTGCTGTCGTCCG（SEQ ID NO.57） |
| 27 | TTATTCACTGGGTCCGTGTCG（SEQ ID NO.58） | TCCAACCAGCAGAACCGATG（SEQ ID NO.59） |
| 28 | CGTCAGGGCAATTCCCAGTT（SEQ ID NO.60） | AACCGATATGCATCCTCCGC（SEQ ID NO.61） |
| 29 | CCTTCGAGATCGAGGGCTTC（SEQ ID NO.62） | TCCCGTAATATGTCGCGCTC（SEQ ID NO.63） |
| 30 | GCGGGCTATCGTGTTCTACA（SEQ ID NO.64） | CTGCTTGTAGTTCGGGTCGT（SEQ ID NO.65） |
| 31 | CTCATCAAGGAGTTGCACGC（SEQ ID NO.66） | GCGCAATATCAGTGTCGCTC（SEQ ID NO.67） |
| 32 | TTTCTCGTCCACTACGGCAC（SEQ ID NO.68） | ATCGCTGCTGAACAACCGTA（SEQ ID NO.69） |
| 33 | CGACCCGGAGAATCCTTTCG（SEQ ID NO.70） | CTTCGATGAAGACCGGACCC（SEQ ID NO.71） |
| 34 | CCAAGGGTCCGGTCTTCA（SEQ ID NO.72） | TCCGCATTAGCCACGATCTC（SEQ ID NO.73） |

[0037] 以上所述实施例仅表达了本发明的实施方式，其描述较为具体和详细，但并不能因此而理解为对本发明专利范围的限制。应当指出的是，对于本领域的普通技术人员来说，在不脱离本发明构思的前提下，还可以做出若干变形和改进，这些都属于本发明的保护范围。因此，本发明专利的保护范围应以所附权利要求为准。

## 序 列 表

```
<110>  上海交通大学医学院附属第九人民医院
<120>  长尾噬菌体及其获得方法和应用
<160>  73
<170>  PatentIn version 3.3

<210>  1
<211>  2499
<212>  PRT
<213>  Siphoviridae

<220>
<221>  misc_feature
<222>  (1689)..(1689)
<223>  Xaa can be any naturally occurring amino acid

<400>  1
Val Glu Glu Leu Phe Leu Ser Thr Glu Trp Glu Gly Ile Trp Phe Asp
1               5                   10                  15
Ala Ala Lys Leu Leu Cys Leu His Ser Lys Arg Val Glu Thr Ala Arg
            20                  25                  30
Ser Tyr Pro Ile Glu Gly Ile Arg Tyr Leu Leu Ile Ser Asp Ile Glu
        35                  40                  45
Ser Leu His Asp Thr Tyr Ala Val Phe Gly Glu Asp Gln Val Tyr Leu
    50                  55                  60
Leu Glu Asp Ser Phe Gly Ser Arg Met Lys Gly Tyr Tyr Thr Phe Thr
65                  70                  75                  80
Pro Cys Arg Glu Ser Arg Leu Ile His Trp Val Arg Val Glu Ala Ser
                85                  90                  95
Phe Val Asp Leu Lys Val Val His Arg Pro Asp Asp Ser Thr Ser Tyr
            100                 105                 110
Leu Tyr Val Met Lys Lys Glu Pro Glu Gln Thr Val Cys Gly Arg Ser  [0038]
        115                 120                 125
Ser Leu Thr Arg Ser Ser Thr Thr Thr Gly Trp Pro Arg Ser Pro Thr
    130                 135                 140
Gly Arg Ser Ala Ala Thr Ser Thr Arg Pro Ala Ser Thr Ala Asp Ser
145                 150                 155                 160
Ser Gly His Phe Val Ser Ser Ser Arg Thr Gly Trp Ser Ser Ser Pro
                165                 170                 175
Arg Gly Ser Ser Thr Arg Ser Ser Thr Ser Lys Arg Ala Arg Ser Ser
            180                 185                 190
Ser Pro Pro Thr Asp His Ile Arg Thr Gly Ala Trp Ser Asp Leu Gly
        195                 200                 205
Pro Arg Arg Arg Lys Gly Ala Pro Cys Tyr Pro Thr Thr Ala Leu Phe
    210                 215                 220
Leu Trp Thr Ser Ser Arg Gly Arg Arg Ser Gly Arg Thr Ala Ser Asp
225                 230                 235                 240
Arg Ser Ser Ile Ile Ser Ala Gly Thr Ala Arg Thr Phe Trp Trp Ser
                245                 250                 255
Thr Glu Ile Ser Arg Pro Ile Ser Arg Thr Ser Ser Cys Trp Ile Gln
            260                 265                 270
Ile Ala Ser Ala Ser Ser Arg Lys Ser Ile Arg Leu Gly Thr Ser Thr
        275                 280                 285
Ile Trp Ser Trp Pro Met Val Asn Pro Pro Pro Arg Pro Thr Tyr Pro
    290                 295                 300
Ile Thr Glu Trp Ile Arg Asp Glu Lys Tyr Leu Gly Lys Arg Glu Tyr
305                 310                 315                 320
Trp Ile Arg Ser Gly Leu Thr Lys Gln Arg Tyr Leu Ile Ser Ser Lys
                325                 330                 335
Gly Arg Val Lys Gly Pro Arg Gly His Ile Gly Cys Ala Arg Gln Gly
            340                 345                 350
Glu Gly Val Glu Ala Arg Asn Arg Gln Gly Asn Ser Gln Phe Phe Arg
        355                 360                 365
Leu Ala Arg Leu Met Cys Glu Ala Phe Thr Thr Phe Glu Tyr Asp Pro
```

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

[0039]

```
Phe Glu Ile Glu Gly Phe Trp Arg Thr Cys Cys Gly Ser Thr Ser Ala
                645                 650                 655
Arg Pro Cys Gly Gly Cys Ile Ser Val Ala Ser Ser Ala Gly Pro Arg
            660                 665                 670
Ala Pro Gly Arg Ala Trp Arg Ala Ser Arg Ser Pro Ser Arg Pro Ser
        675                 680                 685
Arg Val Ile Ile Leu Ser Arg Gly Arg Ala Pro Arg Ser Pro Arg Arg
    690                 695                 700
Ser Ile Ser Ser Arg Pro Arg Pro Ser Gly Thr Val Ser Thr Gly Arg
705                 710                 715                 720
Gly Ser Ser His Gly Ser Ala Ser Arg Pro Arg Asn Arg Arg Ala Ala
                725                 730                 735
Tyr Pro Ala Ser Ser Thr His Gly Thr Thr Ser Gly Ser Thr Arg Ser
            740                 745                 750
Arg Arg Ala Ser Ser Tyr Ser Thr Ser Ser Asp Ser Trp Gly Gly Ala
        755                 760                 765
Pro Gly Arg Arg His Ile Val Arg Ser Pro Pro Gly Thr Pro Gly Ser
    770                 775                 780
Cys Ser Arg Ala Arg Pro Glu Thr Pro Gly Trp Thr Thr Phe Arg Tyr
785                 790                 795                 800
Phe Trp Ala Thr Ala Ser Thr Ala Thr Arg Arg Ser Ser Trp Arg Ser
                805                 810                 815
Thr Trp Ser Thr Thr Ala Ser Arg Asp Ile Leu Arg Ser Ser Gly Thr
            820                 825                 830
Thr Thr Arg Arg Cys Cys Gly Ser Thr Gly Thr Pro Tyr Trp Cys Arg
        835                 840                 845
Cys Arg Trp Arg Gly Gly Pro Gly Val Thr Ser Ser Ala Tyr Pro Ala
    850                 855                 860
Ser Thr Arg Gln Arg Arg Thr Thr Arg Arg Cys Gly Asn Gly Gly Thr
865                 870                 875                 880
Phe Phe Gly Thr Val Arg Ser Arg Thr Leu Arg Val Leu His Gly Arg
                885                 890                 895
Cys Ala Ile Ser Trp Pro Gln Ala Thr Ala Glu Ala Pro Phe Gly Ser
```

```
                900                     905                     910
Trp Arg Pro Val Ser Gly Leu Ser Cys Ser Thr Cys Thr Thr Thr Ser
            915                     920                     925
Ala Thr Tyr Tyr Gly Thr Ser Ala Arg Leu Trp Arg Gly Arg Gly Gly
        930                     935                     940
Arg Thr Glu Pro Thr Thr Thr Gly Phe Leu Arg Ala Thr Arg Gly Arg
945                     950                     955                     960
Ile Trp Cys Ser Lys Pro Pro Gly Pro Arg Pro Gly Thr Ala Arg Pro
                    965                     970                     975
Ala Ile Pro Ser Tyr Ser Thr Ala Ser Pro Thr Arg Gly Ser Arg Trp
                980                     985                     990
Asn Arg Arg Arg Gly Ala Ser Thr  Gly Pro Thr Pro Arg  Thr Arg Thr
            995                     1000                    1005
Cys Thr  Ile Ser Leu Arg Ala  Ser Arg Ala Ser Thr  Ala Gln Ser
    1010                    1015                    1020
Arg Gly  Pro Arg Gly Arg Ser  Gly Ser Thr Lys Arg  Pro Ser Gly
    1025                    1030                    1035
Asp Asp  Asp Glu Val Pro Gly  Arg Ser Tyr Gly Arg  Leu Gly Phe
    1040                    1045                    1050
Ser Phe  Leu Ala Arg Ser Gln  Glu Trp Pro Phe Phe  Trp Pro Leu
    1055                    1060                    1065
Trp Pro  Asp Leu Ala Ser Arg  Glu Ser Leu Gln Arg  His Phe Leu
    1070                    1075                    1080
Tyr Leu  Met Ala Lys Arg Pro  Phe Phe Tyr Ser Tyr  Tyr Thr Ile
    1085                    1090                    1095
Ala Ser  Arg Ala Cys Lys Phe  Ser Arg Gly Val Asp  Lys Pro Val
    1100                    1105                    1110
Pro Phe  Leu Ala Ile Trp Pro  Phe Glu Asp Phe Ala  Ala Glu Ser
    1115                    1120                    1125
Arg Pro  Lys Arg Arg Gly His  Phe Trp Pro Arg Trp  Thr Arg Ile
    1130                    1135                    1140
Trp Pro  Leu Ala Ile Phe Gly  Gln Asp Pro Thr Pro  Ser Trp Val
    1145                    1150                    1155
[0040]
Gln Leu  Arg Thr Gln Asn Thr  Gly His Asn Glu Asp  Pro Asp Pro
    1160                    1165                    1170
Thr Pro  His Gly Ser Cys Ile  Phe Arg Arg Leu Asn  Ala Arg Thr
    1175                    1180                    1185
Arg Val  Pro Gly Ser Pro His  Gln Gly Val Ala Arg  His Phe Pro
    1190                    1195                    1200
Arg Leu  Tyr His Leu Glu Glu  Arg Pro Glu Leu Gln  Ala Gly Arg
    1205                    1210                    1215
Ala Gly  Pro Tyr Ser Pro Leu  Gly Gly Gln Val Gly  His Ala Gly
    1220                    1225                    1230
Gly Gln  Gly Leu Gln Ser Cys  Gln Ala Pro Pro Gln  Pro Gly Leu
    1235                    1240                    1245
Leu Arg  Ser Ser Arg Pro Asp  Val Leu Arg Arg Ile  Tyr Phe Ser
    1250                    1255                    1260
Gly Gln  Gln Gly Thr Arg Pro  Arg Cys Thr Ser Thr  Ile Thr His
    1265                    1270                    1275
Val Pro  Ala Pro Thr His Phe  Ser Ala Arg Val Ser  Thr Pro Gly
    1280                    1285                    1290
Ser Thr  Thr Thr Pro Ile Ala  Phe Glu Arg Ser Met  Ala Pro Ala
    1295                    1300                    1305
Arg Pro  Gln Leu Gly Ala His  Gly Phe Thr Ser Ser  Leu Pro Ser
    1310                    1315                    1320
Ile Phe  Ala Cys Glu Phe Glu  Cys Arg Arg Thr Arg  Ser Arg Ser
    1325                    1330                    1335
Thr Ala  Ile Ser Thr Thr Leu  Ser Ala Thr Val His  Arg Ser Arg
    1340                    1345                    1350
Ser Ser  Gly Thr Ala Thr Thr  Pro Ser Gly Arg Arg  Thr Pro Tyr
    1355                    1360                    1365
Pro Ser  Thr Arg Arg Arg Thr  Phe Ser Ala Ser Thr  Thr Ser Arg
    1370                    1375                    1380
Arg Ala  Leu Arg Arg Arg Pro  Arg Ser Ser Ile Ser  Thr Trp Arg
    1385                    1390                    1395
Cys Ser  Ala Ser Ser Thr Ala  Ile Ala Arg Ala Ile  Phe Met Arg
```

1400 1405 1410
Arg Pro Val Ser Thr Arg Thr Thr Arg Ser Leu Ser Thr Ser Pro
1415 1420 1425
Met Pro Arg Cys Ser Arg Ile Phe Arg Thr Asp Ser Ser Ala Ser
1430 1435 1440
Thr Arg Ser Ser Arg Asp Ser Ser Arg Arg Arg Asp Asp Arg Phe
1445 1450 1455
Ser Arg Pro Leu Arg His Ala Pro Pro Leu Arg Ala Leu Ser Leu
1460 1465 1470
Gly Val Gly Gln Gly Pro Val Pro Glu Arg Leu Val Ile Ser Gln
1475 1480 1485
Arg Ala Arg Ser Ser Pro Gln Gly Gly His Glu Arg His Tyr Cys
1490 1495 1500
Ala Cys Leu Gly Asn Asp His His Arg Val Pro Ser Ser Gly Gln
1505 1510 1515
His Arg Pro Arg Arg Glu Glu Gly Arg Arg Tyr Cys Ser Ser Arg
1520 1525 1530
Ser Ala Gln Gly Arg Arg Ser Pro Gln Tyr Arg His Arg Gly Glu
1535 1540 1545
Asp Gly Ala Arg Val Val Cys Ser Ser Ala Glu Ala Gly Arg Arg
1550 1555 1560
Ala Ser Pro Gly Arg His Gln Ala His Gln Arg Tyr Ser Gly Gly
1565 1570 1575
Val Gln Gly Gln Ala Val Tyr Val Arg His Arg Arg Gly Asp Glu
1580 1585 1590
Ser Pro Asp Leu Asp Arg His Ala Gln His Ser Arg Arg Val Leu
1595 1600 1605
Ala Pro Glu Gly Val Gln Asp Pro Arg Gly Leu Cys Gln Ala Ala
1610 1615 1620
Gly Thr Gly Arg Leu His Asp Pro Glu Gly Ala Arg Pro Ala Gly
1625 1630 1635
Arg Tyr Pro Arg Arg Ala Asp Gly Glu Ser Arg Gln Asp Pro His
[0041] 1640 1645 1650
Pro Gly Cys Cys Arg Gly Arg Glu Arg His Ile Asp Arg Asn Gln
1655 1660 1665
Ala Pro Asp Glu Gly Val Glu Gln Ala Ser Ala Tyr Ser Leu Arg
1670 1675 1680
Arg Gly Arg Trp His Xaa His Gly Arg Arg Tyr Ser Ala Pro Ser
1685 1690 1695
Arg Cys Pro Gly His Gln Ser Arg Leu Leu Ser Ile Cys Ser Gly
1700 1705 1710
Pro Tyr Cys Arg Arg His Ala Leu Pro Gln Gly Tyr Gly Asp Val
1715 1720 1725
His Arg Gln Ala Pro Gln Gly Arg His Asp Ile Gln His Gln Gln
1730 1735 1740
Ala Gln Arg Asp Thr Arg Ser Arg Pro Glu Gly Pro Leu Gly Ser
1745 1750 1755
Gln Ala Pro Gln Gly Arg Pro Gly Glu Ser Phe Arg Tyr Gly Cys
1760 1765 1770
Ser Ala Ala Ile Leu Gly Gln Glu Asp Gly Ala Glu Ala Val Gly
1775 1780 1785
Pro Gln Tyr Cys Glu Arg Gly Gly Val Leu Gly Pro Leu Val Cys
1790 1795 1800
Asn Ser Gly Val Pro Val Pro Val Lys Ala Val Ala Arg Ile Gly
1805 1810 1815
Glu Glu Ala Ala Gln Gly His Ser Arg Gly Glu Arg Ala Gly Val
1820 1825 1830
Gln Gly Asp His Glu Ser His Glu Ser Cys Tyr Ser Glu Glu Ala
1835 1840 1845
Ala Glu Val Leu Cys Arg Gly Leu Arg Leu Gly Ser Arg Thr Pro
1850 1855 1860
Gln Ser Gln Gly Pro Thr Arg Ser Gly Leu Ala Gly His Ile Ala
1865 1870 1875
Asp Ala Ala Ser Gln Glu Glu Pro Cys Val Cys Thr Lys Phe Pro
1880 1885 1890
Arg Arg Ser Arg His Glu Asp Asn Glu Gln Glu Thr Pro Gln Tyr

```
    1895                 1900                 1905
His Ser  Arg Ser Arg Ser Leu  Trp Leu Thr Thr Arg  Thr Arg Lys
    1910                 1915                 1920
Arg Val  Val Leu Trp Val Val  Leu Gly Met Leu Leu  Val Phe Ile
    1925                 1930                 1935
Leu Leu  Ser Pro Ser Asp Ser  Ala Ser Ala Asp Phe  Asp Gly Asp
    1940                 1945                 1950
Ser Val  Thr Ala Ile Pro His  Ile Gly Thr Thr Arg  Val Arg Ser
    1955                 1960                 1965
Thr Pro  Pro Leu Lys Gly Leu  Lys Gly Phe Glu Pro  Lys Arg Val
    1970                 1975                 1980
Tyr Pro  Ala Tyr Pro Gly Met  Lys Arg Met Gly Asp  Thr Gln Ala
    1985                 1990                 1995
Gln Met  Gly Lys Ile Ser Asn  Leu Ile Thr Gly His  Asp Pro Glu
    2000                 2005                 2010
Gly Gly  Phe Arg Val Arg Ala  Arg Ser Arg Cys Pro  Pro Leu Asp
    2015                 2020                 2025
Gly Cys  Tyr Arg Glu Ala Leu  Pro Gln Leu Gln Thr  Glu Ala Gly
    2030                 2035                 2040
Arg Tyr  Cys Cys Ala Ser Glu  Glu Val Ser Gly Arg  Lys Arg Ala
    2045                 2050                 2055
Arg Cys  Gly His Ala Tyr Phe  Gln Gly Gln Gly Ser  Gly Leu His
    2060                 2065                 2070
Arg Arg  Arg Thr Pro Arg Gly  Leu Gln Ser Pro Ser  Ser Leu Ala
    2075                 2080                 2085
Leu Ser  Ala Thr Arg Trp Leu  Leu Leu Thr Leu Arg  Ser Thr Thr
    2090                 2095                 2100
Ser Ile  Thr Asn Arg Val Ser  Gly Ile Thr Val Leu  Leu Arg Phe
    2105                 2110                 2115
Gly Arg  Ser Ile Arg Ala Lys  Thr Gly Thr Val Arg  Pro Arg Leu
    2120                 2125                 2130
Phe Pro  Gly Pro Thr Val Arg  Ser Ser Ser Glu Glu  Arg Lys Leu
    2135                 2140                 2145
[0042] Arg Lys  Ala Ala Lys Gly Gly  Pro Ile Asp Pro Lys  Thr Gly Lys
    2150                 2155                 2160
Leu Val  Tyr Glu Lys Thr Gly  Arg Gly Phe Tyr Asn  Arg Lys Gly
    2165                 2170                 2175
Lys Trp  Ile Pro Lys Val Thr  Lys Ala Arg Arg Met  Ala Leu Val
    2180                 2185                 2190
Asp Asp  Ala His Ile Leu Ser  Ser Gly Thr His Met  Glu Gly Ile
    2195                 2200                 2205
Tyr Ala  Asp His Ala Asn Glu  Leu Lys Ala Leu Ala  Asn Arg Ser
    2210                 2215                 2220
Arg Lys  Ala Ala Val Ser Ile  Pro Pro Leu Lys Arg  Asp Pro Arg
    2225                 2230                 2235
Met Ala  Arg Lys Tyr Ala Pro  Glu Val Ser Ser Leu  Arg Ala Ala
    2240                 2245                 2250
Leu Asn  Arg Ala Ile Lys Gln  Lys Pro Leu Glu Arg  Gln Ala Glu
    2255                 2260                 2265
Ile Ile  Ala Gln Gly Val Val  Ser Lys Lys Val Ala  Ala Asn Pro
    2270                 2275                 2280
Asp Met  Ser Lys Lys Glu Arg  Ala Lys Ile Glu Tyr  Met Ala Val
    2285                 2290                 2295
Glu Thr  Ala Arg Ala Arg Leu  Gly Thr Asp Arg Lys  Gly Thr Arg
    2300                 2305                 2310
Val Arg  Pro Thr Pro Arg Glu  Trp Glu Ala Ile Gln  Arg Gly Ala
    2315                 2320                 2325
Val Ser  Asn Ala Met Leu Glu  Glu Ile Val Ala Asn  Ala Asp Ser
    2330                 2335                 2340
Asp His  Ile Lys Lys Leu Ala  Met Pro Cys Glu Lys  Thr Arg Val
    2345                 2350                 2355
Thr Thr  Ala Gln Gln Ser Arg  Ile Met Thr Leu Arg  Ser Arg Gly
    2360                 2365                 2370
Ala Thr  Gln Ala Glu Ile Ala  Glu Ala Leu Gly Leu  Ser Val Ser
    2375                 2380                 2385
Gln Val  Lys Ser Val Ile Tyr  Asn Asp Glu Arg Gln  Leu His Tyr
```

2390 2395 2400
Tyr Arg Arg Gln Trp Cys Leu Pro Leu Asp Ser Gln Gln Arg Thr
2405 2410 2415
Ile Arg Ser Ile Leu Ser Thr Ser Ser Thr Asn Gly Thr Ser Ser
2420 2425 2430
Met Cys Leu Met Ala Thr Ala Pro Leu Pro Thr Trp Val Gly His
2435 2440 2445
Thr Val Ala Thr Asn Cys Leu Leu Gln Ile Arg Leu Lys Leu Leu
2450 2455 2460
Met Lys Leu Leu Arg Leu Leu Arg Ser Ile Leu Ala Thr Thr Gly
2465 2470 2475
Leu Trp Asn Glu Met Ser Lys Ser Asp Glu Met Lys Leu Asn Asp
2480 2485 2490
His Pro Ser Pro Gly Gly
2495

<210> 2
<211> 29632
<212> DNA
<213> Siphoviridae

<400> 2
agggtcctcc gggagacaag gggaaacctg gcgaacaggg tcctcccgga gacaagggcg 60
cacctggcga ccctggcaaa cagggtcctc ccggagacaa gggcgcacct ggcgggaccc 120
tggcaaacag ggtcctcccg gagacaaggg gaaacctggc gataagggtc ctcccggaga 180
caagggcccg agcaagccgg tcacgctgtc ttcagaggca tcgggttacc cgggttacta 240
caaactggaa ctcgaggaca ctccgaacga gaccaagctc aagtggttct cctcgccaaa 300
cttcgctccc gcatatgtca caggtccgcg tgatacaggc aacaggcgcc tctgggcggt 360
tttcagcctc accgaggtga ctgttatcgg cggcacgaat ggtggtgaga ttcaggtcga 420
gcgcggttgg aactggttcg gcaagctccc acagaatatt accgtttatg agaacgccta 480
cgctccgttg ttcgacaaca actggaacat ctgcggatac acgacgattc gcgccgttga 540
ttccagtgta cgagtgttct cgcaggcgga aggcggagta ttcatgaccg cgtgcacggc 600
[0043] tccgcggaag acctgacatg ccctccatgt ccttgtcttg gcagggcgac taccgcgcta 660
ccaaacggtg gatggacaag atggaggccg gcagcgtcgc cgatgttctt cgctcgtgtg 720
gcgagaaggg catcgaggcg ctgtcggccg ccaccccgt tcgcacaggc aagacggcgg 780
cgtcctgggg attcgaggcg aagacctcgt ccaaggggt cacgctcacg tggacgaact 840
ccaatgttgt gaacggcgtt ccgatcgcta taattcttca gtacgggcac ggcaccggta 900
cgggggata cgttcacggc agggactata ttaatccggc tatgcggccg gtgttcgatg 960
agatcgagag tcgcatatct agaatgctga ggggctaatg gctaccattg acgagcgagt 1020
cgtaagtctc aagttcaaca acgcccagtt catggacggt gtccgagagt ccaaggccgg 1080
tattcagtcc cttgactcct cgcttcgtct tgagggcgct acgcagggca ttgataatgt 1140
tgcggaagcc gctcgtcggt tgactttcgg ggatgttatc tccggcgcgg ccaatgttat 1200
ctctaacatg ggactgatgg aggtcgcggg cgtcgcgtcg ctgggcggta tcgccgccaa 1260
ggctgttagc gtcggcggcg agctcgtcaa agccctcact attgacgccg ccaaggcggg 1320
tttcgaggag tacgagcttc agctcaactc cgttcagact atcttggcga acaccgcgtc 1380
caagggcgag gacatcaaca ccgtcaacgc cgccctggac gagctcaacg tctacgccga 1440
tcagactatc tacaacttcg gcgagatgac gcgcaacatc ggcacattca cggccgccgg 1500
tgtcgggctc aaggagtcgg tcagcgccat caagggtatt tccaacctgg cggcggcgtc 1560
cggctcctcg gcggcccagg cctccacagc catgtatcag ctttcgcagg ccattgcctc 1620
cggtaccgtg cgactgatgg actggaactc cgtgcagaac gccggtatgg gcggcgagca 1680
gatgcaggag gcgctcaagc ggaccgctcg agttcacggc caggcggttg acgaggccat 1740
cgcaaagcag ggctcgttcc gagagtctct ccaggagggc tggctcactt ccgaggtcat 1800
gctcgagacc ctgactctca tgaccggcga tctctctcgc gagcagatca tcgagatggg 1860
gtacacggag gagcagaccg acgagatcat gaagttcgcc gagacggcga acgatgcggc 1920
gacgaagatc aagaccttct cgcagcttat cgatacgctg aaggaggagc tcggctcggg 1980
ctgggcgcag acctggcgca ttatcttcgg cgacttcgag caggccaagg agctttggtc 2040
ggggatcggg aactacctca ccggagttct cagcaacgcg tcgaaggctc ggaatgagct 2100
tctccaggca tgggtcgatc ttggcggtcg cgaggacgtc cttcgcggac tcaagaacct 2160
gttcgacgcg atcatgatgc ctctgcgagc gatcaaggag gcctgggacg cgacattctc 2220
ggggccatcc gctgagggtc tggcccgtat cacgaaagcg ttcggcgact tcacggcgta 2280
cctcatcatg ggttacccgc ggttcgtcga tcttcggcgg acgttccagg gcatcttcac 2340
gatcttcggg atcggcatcg acatcattaa ggatgtcggg caggccctcg ggccgattct 2400
cgctgttgcg gcgggtttgg cgggcaaagc tatatcctac ctcggcttcg tgctgttcaa 2460
catcacaagc tacatcggat acctgcttat acgcttccgc gagtggtacc agggactcga 2520
cctcggtacc aagggcgcgg agaaagtctc tgccgccctt gctgtgctgt ataagtacct 2580
ccaggctgcg gcaagcatcg ccggcaagtt caaagagggc ttctacgtcg ggatcatcta 2640

```
ccccatgggt gttctcaagc gcagcgccga gcatctctgg gagacgattc acgacacgtt    2700
ctctcgtatc taccacgcca tcgtcgaccc gttcaagaag acggacgagg ccgagggatc    2760
cctcaacgtt ctcaccggcg atgtcgtgga cctcggagac aaggtcgagg aggccgctca    2820
gccagtcatc accttcgctc aggtcgtctc ggccatgggc aggaaggcct gggaggctgg    2880
caacaagctc gcgcagaagc tggctccgta tattctcgag ctggcgagct ggctcgacgg    2940
caccaaccac aagatcgagg agttcggcga caagaccgag gcgaacatca cagagaagat    3000
gcctacggtc atcgagtgga tcgagaagac caagcaggcc ttccgggaca tgaaggacgc    3060
cttcacgaat ggcctcatgg actccctcac cggcaacgac tcgaagctgt cgctgccgga    3120
cgtcgacacc agcaccttcc gcggtcagat cacgcttgct gcgtcatcac tcaagtcgat    3180
cgacacgtcg tcgttcacgg cgatgctggc cagcgcgggc gctgctctgt cgagcttcgg    3240
gacacacctc tggaacgccg gtaaggcgct gaaggagttc tggggcaaca tgagcatcgg    3300
cgagcatatc gccaacggct gggacaactt caaggaccgt tggggaggcc ttccgggtct    3360
tatcgggaag ggcgttaccg cggtcggcaa ggcgctcggt ttcctggcgg accaggcggc    3420
tcgtttcgcc ggatcggcgc tgaaggggct cggcgagttc atcaagaact cgatcacgtg    3480
gttcacgggg catcatcagg tcgcggccat gacgctcgcc atcggtggtg ttatcagcgg    3540
cctgcttcgc ttcgtcaccg cagggcgcgc catttccggc acgttcatct cgatcaagga    3600
ccttatcggc tcggtcaaga acgccatcga ccagttcagc ggcgtattca cctccttctc    3660
caagaagctg gaggccgagg cgaagaccga gacggcgaag cagatcctga tgtacgccgg    3720
tgctatactg gtgctcgcag ccgcgatgtg ggtgctcgcc cagatcccgg cggacaagct    3780
cctcggcgta tccgtggcca ttatcgcggt attcgccggc cttacggcgt ctgcggagtc    3840
gatcggcaac tcgctcaagg acaacgcgag catgctcgtc ggggctctcg ccatggtggc    3900
gatctgtatc gcggtcgcaa ctgcggcggg ggcgcttcgc gtcgtcgccg gcatggagtg    3960
gcccgagatc ctggcgggga cggccgcgct ctccgctata ttcgggatgc tcatcgtgac    4020
gtccaagaac gtcggcaaga acaagggctc actcatggcg ctcggtgtga gcgctatact    4080
gctgggcgcc ggcgtacgcc tcatggtcaa cagcgtcgcc aagctcggcg agatgtctct    4140
cgctcagctg ctgcaaggca tggcggcact cacgatcatg atcgcaattc tgatgatatt    4200
caaccgtgtg cctacgccgc ggaagaacgc catcgggaac gcgatccagt tcctcatcct    4260
gtcgtatgcc ctgaagaacg tcgcctcagt catcaaagag ttcggcgaga tggactggaa    4320
gacgtatctc gtggggctcg gcgagatgac cggcgcgctg gcgattctca tggggggtcgc    4380
tatcctgctg gacaacatcc cactctctac cgtgggcgcg gttggccttc ttgtcgtggc    4440
catggccgtg ggcaaggtcg cggtgatcat tgcgacgttc gccgcgatgc cgtggcagca    4500
gtatctgatc ggcctcggcg agatgacagg ggcgttggcg attatcgtcg ccgcgagcta    4560
[0044] tctggcctcg gggagcatcg gcggcgcaat agccctggca gccctggcgg cggctatatc    4620
cctgcttgtt ccggcactgg tcatcctcag tcagattccg ctcaagaagg ccgctagcgg    4680
tctgcttatc ctggcggggg cgctggctat tatcatcgcc gctggctatc ttgctatagg    4740
cgcggctgtc gggctcgggg cgcttgctct ggtgcttatc ggtcttggcg ccggattcgc    4800
cctcgctggc gccggtgtgt ggctgttcgc aacaggtctt acgctgctga cagtcaacgg    4860
cgcggccggg gtagccgtcc ttgccaagat gatcgagctc ctgccgattc tgggtacgaa    4920
cctggccgga ctccttgtga atttcctggc agttctcggc gagaacatgc aggcgattcg    4980
cgatggcctc ggggcgctct tgatcggcat attcggagcc atcaccgatt cgctccccgc    5040
gttcggtgag atgtgtatcc agcttatcgg aacgcttctc acagtcatcg tcacgatgac    5100
gccgcagatc atccaggccg cgttcgacct gcttatcgcg cttattcagg tgctgattga    5160
caagatgcct gagttcgccg agaagggcgc cgaactcatt atcgcctggt tgcagggcat    5220
cgaggaccac gctcctgatg ttatagagca ggccgtgtcg accatctcga cgctcgccac    5280
cgcgttcagt gagcagcttc cggtcatcat ccagacggcc ttcgatattg cgatcaactt    5340
cctcaacggg ttcgctgacg ctattcggaa taactccagc accgttggag acgcctgcgg    5400
gaacattgtt gacgcgcttg tcacgggctt caaggactac tcgtcctcct tcgggagtcg    5460
gatctgggat actctgactc gactcggccg cgatattatc aacggcatta agcagggggt    5520
cagggacaat gcgcacaaga tcggggacgc gcttatctcc ggtgcgaagg ccggcatcgg    5580
ccggctcaag tcttacctgg gcattcactc gccctctcgg cgcttcgccg agctcggcaa    5640
gtggaccatg atgggattcg ccaagggtat ctccgacaac tcctccgaga cggttgctgc    5700
tgcacagaac tcggcgaagg ccgtttccga ggagttcgag aagaacctgg acattgatcc    5760
ggagaacatc aatctcttcg gcaaccagga gcctccggtg atcaagccgg tagtggacct    5820
gtccgaggtc gacaaggcca aggaggcgat caacggtctc ggggcctctc cgaatcttac    5880
ggcggaggcc gccagggtcg aggcacttcg ctctgcgtac agtgttacgg ctcccgagcc    5940
gcagcgtcaa aatgacagcg atcgtccgaa gggcgacgtt atcttcaacc agtacaacac    6000
ctcgccaaga gcgctgtcgg aggcggagat ctaccgtcag acgcggagcc agctggttgg    6060
tcttcgagag gagatattca agctgtgatt cgatctatcg agcttattcc cgcagaccga    6120
gaggggctac tgctggagat tgacaagccg gaggacaccg gcatcctggt caagtccatc    6180
gacggccttg gtccggcgaa ggctacgatc aacaccacag cgctgtctct gaccgactcg    6240
gcgatgttca acggcagcag ggtgggtatg cgaactatca cgctcaccct gctgccgctc    6300
gaggttccta cggtcgagcg tgcacggcac cggatctaca accttctgcc gatcaagcag    6360
ccggtaacga ttatcgtgcg gacggacact cgggtcgtca agacactcgg gtacgtcgag    6420
tccagcgagc cggacatatt ctccaaggag gaggccataa aggtcgtcct catatgtccc    6480
gacggctact ggtcggacgg cgctgacgac actcagaact atctgccctt cgtgcaggag    6540
cacgcggcgt tcgagttcga ctgggaggac acgcctctct cggagagccc cacgcttata    6600
```

```
        ttttcgaaga ccgttggtct cccctccgtt attctggaga acggcggaga cgttccggcg    6660
        gggttcacga tcatgatcga catcctcaag gacaacgcca cgccgatcac gatctatgac    6720
        gatgttcgca atcagcacct gactctgact accaagtggc acccggacgc ggctaccacg    6780
        cagccggcca aggcgggcga tcgcttttac atcaacactc gtgtgggcca caaggcagtc    6840
        acgcgtattc gaaacggcaa gtccgagaag gcgctgcatc ttctcgatat cgactcggac    6900
        tggctcatgc tgtatccggg agagaaccgt ctctactaca cgatcaatgt cggcggcgga    6960
        gccaaggtgt ccctgtcccg cgatattctc taccaggggg tatagtgtat ctagccgtcc    7020
        tcgacaaagg acttaacctg gcgtacgtga ttgacgacta caagtcggtt atctggacgg    7080
        agcgctacaa caagtacggc gacttcgagc tggtggtccc gggcacttac gagaactacg    7140
        agaagttccg cctcgataca tatctgttca cccctgagtc gcagaaactc atgatcatcg    7200
        agcagctcga gtggacagag gagtacaaca aagccagcga gatcaagatc acagggcgat    7260
        ccctggagtc cattctcgac cgacgtatcg tagggccgat ccgctccagc gatgacccct    7320
        ggttctacta cgtcagaggc ttcaacgatt tcggcgacaa catcattacg tccctggtcg    7380
        cggatgcgtt tgcgtttcgt aagccggcac ttcctcgcca cgtggggtgg ctgagatgga    7440
        tggcgaaccc gaaggatctc ccgacggaag ttcttcgtgg aagagacggc gccgagatga    7500
        agcttccgca gcgccgtttc acaacggttg ctgagcagca cagcaaaccc gctggcccgg    7560
        gctcagacct tcggctcagt cagtactcga ttgacctctc gtcgacgctt gagggcagct    7620
        tgtacgagat cgtgaccact attctcgggt actacggcgg cggattcgca atcgacctgg    7680
        agctcggaaa ccccttccgc tggtggggat atgtgtacaa agggctcgac cgcacgcggt    7740
        cgcagacgaa gcggtcgccc gtcatcttct cgcccaagat ggacaatctg tcgaaggcaa    7800
        cctacatcga gtccgttgcg gattacaagt ccgttatcta ctcgggcctt gtcagcagcc    7860
        agagcgttgg cgaagcgcac gagtacggcg acacaaagcc gaagacgcag ctcgccaata    7920
        actcgaagac aatcggcatc tcgggcctgg agcgccgaga gggatatctc cagaacatcc    7980
        agatcaacga gatgaccggc gttcgctctg acaggaacgg atcaaaatac tcggttctgg    8040
        gtgagcagac gatgtccaag gagcaggttc tgcagaagat cgaggacgcc tgcaacgacg    8100
        agctgtacaa gcacggcaag aagacggtat acaccggcga ggcggacatg acgacgatgt    8160
        accgttacgg cgaggacttc ttcatgggcg acctggttca gctcgagaac ggtcacggtc    8220
        tgtctgagaa ggcgattctc accgagtaca ctcgctcctc ctccgagtcc gagggcgaca    8280
        agttctatcc cacatttacc aagccggacg agtcctcggc ctggcaccac tgaggaggga    8340
        gaatggcact tacctctggt ttctacccgt ccaaggacgg agaccgaaag tactcggcgc    8400
        tggacttcgg gcgaatgttc gacggtctta ttcacgacgg catctacgcg actgctctga    8460
        atggttttcg tccgcgaatg aaggacgcga acagtctcac cattcaggtg gacagcggcc    8520
[0045]  gcgcatggtt caaccacacg tgggtcgtca acgacgctat tatcgagatc gacggttccc    8580
        cggcacaccc gacatttcct cgatgggatg ctctggttct cactgtggat cggtcggacg    8640
        accagcgctc ggcatatctt gagattgtca acgggacgcc caaggatggg gcaacgcgac    8700
        cggacgtcac cagcacgaac aacgagatca agtcgcgtta cccgctcttc tacctgttca    8760
        tggacagcaa gtacaaggcc gggactgctc ccaagcaggt aagcgacaac cgaggccagt    8820
        cgacgtgtcc gttcgtgact ggtatcgtcg atcacctcga tggcgacacc ctgttcaagc    8880
        gctgggacga tgcgttccag gaatgggcga agaactacag cgccgaggtc aagaccgacc    8940
        tggatgagtg gaaggccgag cgactggccg actacaacac ctggaaagac accctgatca    9000
        acacgatgaa tggtgacacc aacaagaagg tcatcaacga gatcgcggcg atcaagtctc    9060
        agctccaggg gatgaccgcg ggttttagga tctaccagga catagtatgg tccaacgggc    9120
        aggttctggt agatcgtact acgcccgttc aaggaaaact cgtctacgag cttaaggtgt    9180
        aactcatgcg tatctccgat cttccgatcg tccgctacaa cgaggagtcg gactatatcg    9240
        tcctcgacaa tcccggaaag agcaccacga aacggatatc ctgctccgac tttcgcttct    9300
        ccgtgttctc gaactaccag ttcacccaca accagatcgt tcgcggagca tctctcggca    9360
        acgagttcag acaggaccac aaggatgcga tccggaccgg atcgttccag gacttgtggc    9420
        tcggggacta ctggctgtat aacaacgtcc gttggatgat cgtcgacttc aactacttca    9480
        ggggtacgga agagggcgtc aagaaccata ttgtggtcat gcccgaccga agccttaccg    9540
        ccacttttgc cactaaggca gaagacccgc tgaagaacta ctgcgactcc ctgatgtacg    9600
        atgcggcagc gaagctcaaa ccgcagttcg cgtcgctgtt cggcgacgcg tatatcatgg    9660
        ggcacaagga cgttcttgcg aacacctacg ccggcagcgc aacgttcccg tacacatcgg    9720
        accaggtact cgttcgcggc ggtatctatt ccacgattcc ggacgaggtc atgatgttcg    9780
        gcgcccgact catggctccg gttcaggcgg gacgagacgc ggccagccat attacgggaa    9840
        agcagttctc atactaccga cagggcatgg gtatcccgaa tccccaccag atcttctggc    9900
        tgcgggacaa gtgctggtac aactacttca cctgctgggc tgactaccgt ctgagcagtc    9960
        gcatctggaa cactaccgct ggtctacgtc ccttcgtctg catctcgggg gacgccaact   10020
        gatgtctcac acgctcgaac tggctatcac tgttgctaca tcagtgcttg gctcctcggg   10080
        cttgtgggca ttcattcagg cgaagaacgc gaaaaacgcc gcgcagtccc agctcatagt   10140
        tgggcttgcg catatccagc tcgtcgccat cgccgagggg tatatcgcgc gcggctggat   10200
        cacccacgcc gagtacgatg acattcgcgt gtatctgtat gacccgtaca gagcgctggg   10260
        cggcaacggg tccgcgcagc gtcttgtgca ggaactgtcc ctactaccat ctcacccgcc   10320
        atcagaggag catcatgaca aacaagacgt atgaccgtct gaagttcgtc gcgcaggtcg   10380
        tccttcctgc gctcgccacc ctgtacaccg ctctcgcggc cgcctggggc ttcggccatg   10440
        tcgagggcgt cgtcggcacc atcaccgccg tcgacctctt cctgggttcc ctcctgggca   10500
        tcacagccaa gcagtacacc cctcccgcgg acggtgtact gcacgtggac cacgagaaca   10560
```

```
aggaggtcgt ggccgcactg gagaagccgg cggtcgacct gcgcaggggc gactcggtca 10620
ctctgaaggt ctccgaggcc tgacgcgagg ggaacaggcc tcataatgag acctagagaa 10680
aggaaccaca tgtcccccga cacagagaag cccgaccagg acgacctcct ggaggacgcc 10740
tacgccttcg cgtacggtct ggaccccgac tccgaggcct accgtcggac gctggacagc 10800
atcgacaagc tcgaggccat ctcccgctcc aagcggaaga gctggaagcc gagccccgac 10860
gccgtcctgt cgtccgcgac cagcctcctc ggaatcctga tcatcgtgaa ggcggagagc 10920
atcttcccga tcgcctcgaa ggcgctctct atcgcgacca agatccgact ctgagatcta 10980
accccatgcc atgggcccct gtcataaccg actcgggccc atggcatggt ttatcaaaat 11040
cgggttctaa aaattgcccg ggagggattt tcggaaacgc gatttcaaca gggctcataa 11100
tgagcccaca ctcgaaagga accgcaatgc tgctcttcat caccttcgtg ctcgctctca 11160
tctcgctcgt cgagttcctg ctgctgacct gcgcgctcat cgcgtacagg gacgccgcca 11220
ggagaaagaa gcgcgtgact gagttcgaca aggagttgat gcgcctctac aagtcgcttt 11280
ccgagaactg acctctcgcc tatagcccca acccgggcta taggtctcgc ggaaccatca 11340
tgcctcataa tgagaaccaa gaaaggaacc gcaatgaacg ctcgtaagat cctcggtttc 11400
accgctgaca tctccgcaag ctacgccgtc ggcgtcgcca tggttgccgt catgcccgag 11460
ctctcgatcc ctgcaatgat cgggtaccgt ttgggtacga tggcgatcgg cgccgcagtc 11520
ggacgcttcg cgcggaagaa cgcggaggag acctacgacg aagttacgaa ggctgacaac 11580
gtccacttct actccacccg ccgctgaagg cagattcatg acccgcatgg gtcatgggtc 11640
tcgcggaacc gtcatggccc ataatgagaa ccaagaaagg aaccgcaatg aacttcttca 11700
gctggatcat caagaacatc gtcttcggtg ttggtctctg gaccatcatc aagtggatgc 11760
tctcgaagaa cggctacgag tccgcccccc ggaacttcta ctactgaccc tccgtctata 11820
gcccccgacc cggggctata ggtttctcga aaggaaccac atgtactccc aggtattctg 11880
caacgtcgtc acaatcatcg ccattatcgg gatcgccgat atcttccacc gccagctcga 11940
ggagaaggag gccatgcgca agcagatggc cgagctccag aaggagaaga acgagacgga 12000
gaagcagctt tccgacctca caatgaccat cctcgatctg cgcgcccagc ttcgagagaa 12060
gagcgccttg tgatacttgc gctggccgct gcgaccatga tcatgcttgt cgtgttcgcc 12120
gtgatgggct tctggtgact cgcgaaaaac acatctccta taatgagacc ccatagaaag 12180
gaatcgcaat gaacaacacc gacctcactc ccgccgagaa cgacgagaag aagtccccgt 12240
tcaacctccc cccgctctcg gaccagacca agtccgagct caaggccctc ggctgggccg 12300
tgctcaaggg cgtcgccgtt gccggcgccg tcgcgggcac cctggccatc gccgctgcgc 12360
tcagcaagtc tgacgagacg gacgaggacg aggacgactt cgacaccgtt gacgcggacg 12420
aggacgagtg accatcgtct ccgtctaact ctcgcctata gcccctaacc cgggctatag 12480
[0046] gttttcgcca tgtgcttctc agtcgctgtt gcgctcatat tcatagcgct cgctcttgtc 12540
ctgttcaact gaagggaata ccatgatcaa gaacaccatt ctcgccgaga acttcgacgg 12600
cgagtccgtc gcccagacct gctggttcaa cctcacccgc aaggaggtcg tcgacttcct 12660
cattgaggag tcgcagcagg acagcgtctt cgcccgcagc cttctctcgg gcgagtccac 12720
gccggagaac ctgcgggtcg tcgacggtat caccgcgtac cagttcttcg tccgtctcgt 12780
cgaccgctcg ttcggcgaga gggaccagga cggaatccac ttccgcaagt ccgagaaggc 12840
gctcaaggac ttcaaggagt ccgtgttcta cgaggagttc ctctttgacc tggtcaacga 12900
ccccaagaag gcggtcaagt tcttcaacgg agtcatgcct aagaagatga tcgaggacgc 12960
tcgtagggct aacccggagg cgttcgaggc tctgggtgtc tgacggcatt ttcgaggggt 13020
ctcactagcg agttctacat ggctcatagt gagacccctc gaaaggaaca cgcaatgaac 13080
aacaccgact cggtcaactc ccacgacatc gttctcttcg gatccgcgat ctggatcgcg 13140
ctctggaccc gaacgattct ctggtacacc gacggaactg acatcgccaa ctgacgaacc 13200
tcgccataac cctctgcaag ggttatggtt ttctcgaaag gacacacgtg aatggacttg 13260
tccgcggact tatatccgct acgaacagga acgcccccggc gatcctctcc gccatcggag 13320
tcgccggaac tgtactcacg gcggttgccg ctgttgcggg agaccgacag gccgcgctcc 13380
gctgcgagaa cgcaggaatc gacgccccca cgaaccgaga gcgtctcgag cacggatgga 13440
cgtgctatat tcctgcactc cttgcggcgg cgggaactgt ttcagccgtt attgccggtt 13500
atcgagtcgg cgctgcacgg caggccgcag cggtagccgc ttgcaccatg acgtccgagg 13560
cgttcgatcg ttatcgttcg gccgccaagg aggttctgcc cgaggagcag aggaagaagg 13620
tcgagtcggc ctccgccgcc aagatccgtc ttcccgaaaa ccgacagaac gtcatcttcg 13680
tcgagggcgg aacggttctg tgctatgacg ggcacagtgg gcggtacttc cgctccagca 13740
tgaatcacct ccgcaaggtg cagaacgacc tgaaccggat cctcatctcc gagtcgtcgg 13800
tctccctcaa cgagttctac gagcacgtcg gtcttccgac gacggccatg ggcgatcagc 13860
tcggatggcg cgtcggagcg gagatcgagc tccatttctc caccagcatg ggcgacaacg 13920
aggagccctg cctcgtggtg gacttcatca ccgaaccgat ccccgactgg ttcaaactcg 13980
gctgagccta tctgaaagga cacaatgtca ttctgggagt acgccgttct cacggcagtc 14040
gatatcgtat ggcgaccgat caacgcgacg ctcggtttcc ttgagaagca cacgaactgg 14100
tgggagtgat agacatggac gtacccgttc gaggcagcct tcccggcaac agcgacaagg 14160
acaaggagcg cgtcgccaag cagatcacga gcaagcccgg acgccatatc gcgccgagct 14220
tccgcgaccg cgctatgaac gcgttcttcg gagacgacct gaagtccgtc gggtcatatc 14280
tgttctggga catcgcgatc ccggcgttca agaacatggt gtacgagatg gtggtcgggg 14340
gcaccgagcg gacgctgttc tctcagggcg gcgctcctgc atatcagcgc cccacgacgc 14400
ctcgctacac ctcctacaac cgcgtctacc agagcaggac gtacggaacg acgccgggca 14460
gcgaacgaac gctcactaag agcgtctacc agcccgtcgt gctcgactcg aagagcgagg 14520
```

```
       cggagtccgt cgtcactgag ctcaccaaca tcgtcgacac gtacgggtcc gcctccgtgg 14580
       gcgacctgta cggcatggtc ggcatctccg cggactaccc cgtcgaggac cgcggctgga 14640
       cgaacctgac atctgctcag gtcacacgcg tccgagacgg atatctgctc gatcttccgt 14700
       cccccgtcaa gatcaactga aaggcatatt ctcaaatgaa cctcggagtc attactcgcg 14760
       ctgctgggaa ggtcggtctg gtgctgtcca agcactcccc ggccatgctc agcgttatcg 14820
       gcggcgtcgg tgttgtcgcc accgcagttc tcgcgtcgca ggagaccctg cgcgtcaagg 14880
       aggtcgtcga gccccacgtc gagaacctgg aactcatttc cgccacgctg gaggacgagt 14940
       ccaagcagta cagcgaccag gacgccctgc acgaccggac ggttatctac acccgtctcg 15000
       gtcgggacct gctgaagctc tacgctccgg ctctcgttgt cggcgtgctc accatcgctt 15060
       ccatcgcggc gtcccaccgc atctctgcca agcgcatcgc cggtctgacg gccgcttacg 15120
       gcgcgcttga ccagtcctac cgccgctaca ggggtcgtgt cgagcaggcc ctcggagagg 15180
       aggggatgaa ggagctcgac accaagatcc gcgagcaggc gaagaaggat attgccgagc 15240
       gtcgcaagcc cgacgccgat atctccgaga tcggggacag catcttcgac ctggccggcg 15300
       cctcccagta cgcggtgctc tacgacgaga acgccgccac ctggaacaag aaccggaacc 15360
       tctccacatc gattctccgt gcgcaggaga actacgccaa cgacctgctg aactgccgcg 15420
       ggtacgtgat gctgaacgag gtctacgccg gcctcggtct accccagaca tctgccgggg 15480
       cggtcgtcgg ctggatccgc aaggacgacg gcggtgccga cggctacatc accttcggcg 15540
       actgggacgc gaactacttc gacgacatct acaaggacgt cgacggcgtg tgcgaggcga 15600
       ggcgctggat cctcgacttc aacgtcgacg gcgttatctg ggaccgcatc gacgaggtgt 15660
       cggtccgttg atgttctcgc cgcgtgacgc ggtgctcgtg gcggccgggg cgctttcggg 15720
       tgccctggcc gccgtcgggg cgaccgccta tattctgaac cgcaagcact tccacgagct 15780
       cgaggacatg aaggacgcca tcgcccgaga ggcgatgatc cgcaaggagg cgaccgacac 15840
       ctcgaagatg ctcaaggccg ctatctccga gcaggagaag cagctcgagt tcttcctgaa 15900
       gatgcccgca gaggtgcttg caccgtatct accatcggag gatgacgaca tgcacgagta 15960
       cgaccctgac gtggaggcgt acgacagcga gatgcttccg gagggaactg actccctcgg 16020
       gtgccaggtg ctgtccgaaa aggagtgggc ggagcgcgag ggtcgaggcg atttcgacga 16080
       tctcactctg cgctactacg tgaaggacga cacgctggtc gggccgaaca accgaccggt 16140
       tgagttcccg caggaccttg tcggggaagc gctcagcgtt atctccgtgc cctcgagcgg 16200
       caccactacc gcctatattc gcaactacct gatctgggcg gacatcacgc tcgagatcta 16260
       cgcggacgcg ttcttcgaga aggaccccga tgcctgacac gcagcgttat ttcgagtggc 16320
       tgtacggacg agcgatctcc ttcgtggagc actaccgcat gctctgctcc gtcctgcaca 16380
       ctattatgtt cttctcgaac gtgccgatgg acgacaaccg cgtcgcggac gccgtctctc 16440
[0047] tccggaccat gttcgagaca gaggctcagg cggatctgtc cgagcttcgg gagccctcgg 16500
       tgctggagat gcttgtcgtc ctggcgcagg acctgtcatt ccagacgatg gacgaggagg 16560
       atattcagat cgcgttcgcc gagatcctcg gcaacctcgg cctcagtaag tggatcgacg 16620
       agcgctgggc ggcagaaccg ttcgacgcgg agcacgacac gcgcctcacg atccaccgat 16680
       ggctcgatcg tgaggaactc acccccccgca cgccatttcc ggttttcgag aacggcgagc 16740
       ttgtagacac tcgccgcatc gagctctggt accagatgca tctgtacaac agaagcatcg 16800
       accaattcta ggagcacgct atggatttct tcgaagtggc caggaccgag aagaaggatg 16860
       gttcggcctg cctcgccccg gagttcctcg tgggggagcc cacggacttc atgatccggg 16920
       gaggagcgtt ctacgccatc tgggacgacg agaagggcaa gtggtccact cgcatggatg 16980
       acgtcgcccg gctcgtcgac cgcgaggtcc tggccactgc ggagaagtcg gatgacaatc 17040
       cctcggtacg tctcatgcgg aggttcaagt ccggtatctg gcgcgagttc cgttcgtggg 17100
       agtcgagcat gttcgaccat tttcacccgc tggacgacac gctggtgttc gcagacgatc 17160
       cggtgcgcaa ggaggactac gcctccaagc gtctgccata ttcgctcaaa gcgggagacc 17220
       actcggcctt cgaccgtctc ctcggggtcc tgtacgacga ggaggagcgt cggaagatca 17280
       tctgggcgat cggctctgtc atctccggag acagcgctcg gatccagaag ttcttcgtac 17340
       tgtacgggaa gccgggatcg ggcaagtcca cgatcctgaa catcatatcc tggctcttcg 17400
       agggctacag caagccgttc gactcgcggg ctcttgtgcg gtcgaacaac cagttcgctc 17460
       tcgaggcgtt caagacgaac cccctcatcg ccattgacca cgatggcgac atgagcaggc 17520
       tcagcgacaa cacggttctc aactcgctcg tctctcacga ggagatgctg gtgaacgaga 17580
       agaacaagtc gctgtacggg atggcctttc ggacgaccct gttcgtgggc acgaactcgc 17640
       ctgttcgcat cacggacagc aagtcgggca tcatccgcag gctgatcgat attcacccga 17700
       cgggcagcac ggtccccaag gacacgtacg agcgtctgat gaccgagatc agtcagacac 17760
       gcggcgagat cgcggcatac tgccgggacg agtacaaccg tatgggcatc tcctactacg 17820
       acgagtaccg cccgctgcgg atgatggggg acacggacgc gctgtacaac ttcgtggagt 17880
       acaaccagga ggagctcgag gccgagtcga gcatatccct gcgtcgtgcc tacacgatgt 17940
       acaaggagta ctgcgacatg gcgagcgttt cctcgccccct gcccatgtac gtgttccgtt 18000
       cgcagttccg gacgtatttc gaggactaca aggaccgtgc ggttctgtcg gacgggtccc 18060
       gcggacggag catctacatc ggcttccgaa gcgatatcct gacgcagaag cagctgaagg 18120
       agaagcccaa cgagacgacc tggttgcacc tgtctgagca gccctcgctc cttgacaagc 18180
       gatattccga ccgcccccgcg cagtactcgt cggacgaggg caccccgctg agtcggtggg 18240
       acgacgtcac gaccacgctg agcgatattg acactcgaaa ggagcactat gtacgacccc 18300
       ccgagaatga gattgtcatc gacttcgata tctgcgagaa cggagtcaag tcgccttctc 18360
       gaaaccttga cgctgcttct cggtggcctc gaacctatgc tgagaggtct cgaggaggtg 18420
       gaggtctcca tctccattac gtctaccgag gagaccccga gcgcctcgcc catatttacg 18480
```

```
agccgggaat cgaagtcaag tggtatcccg gacgttcagc tctacggagg cgactgagtg 18540
tatgcaatga tattcctccg gctactcttg aggaaggctc gctccctctc aaggaggaga 18600
agatgctcaa ccgcacgacg atgaaggacg agaagagtct ccgcgacctt atcctgcgga 18660
acctccggaa ggagatccat ccgggcacca agccatcgat cgacttcatc gagaagatcc 18720
tgaacgacgc gtactcgcag gggatggagt acgacctgag cgacatgcgt caggcggtca 18780
tgacgttcgc cctccggtcc acgcaccact cctcgtactg cctctcgcgc gttgcccaga 18840
tgcatttcca ctcggagatg gatgtcgaca agaccgacgg agtggagaag ggcgatatcg 18900
tcttcttcga ctgcgaggtc ttcccgaacc tgttgctcgt gaactggaag gttcagggta 18960
acgaccatat tcaccgcatg atcaaccctg cgccggcaga gatcgaggcg ctcgtcgagc 19020
accgtctgat cggcttcaac aaccgacggt acgacaacca tatcctgtac gggcggatga 19080
tggggtactc gaacgagcag ctgttccacc tgagtcagaa gatcatcggg aacgtgctgg 19140
acgcggggtt ccgcaacgcg tacgacctct catattccga catctacgac ttctcctcga 19200
agaagcagag cctcaagaag tgggagatcg agctcgggat acaccacaag gagctcggcc 19260
tgccctggga catgcctgtt ccggaggagc gctggcccga ggtctccgca tactgcgaca 19320
acgacgtacg ggccactgag gccgtgttcg acgcacggca cgaggactgg acggcgcgtc 19380
gggttcttgc ggacgtcgcc ggcatgactc cgaacgccac gaaccgtatc ctggcgacca 19440
agatcatttt cggggacaac cgccacccgc agaaggacct ggtctacaca gatctgagca 19500
aggagttccc gggctacaag tacgagttcg gcgttagctc atatcgcggg gaggtcaccg 19560
gagaaggcgg ttacgtgtac gccgagccgg gaatccaccg caacgtggct ctgctcgata 19620
tcgcttcgat gcaccccacc tcgatcgagg tcctgaatat gttcgggcct tacacgaagc 19680
gcttctccga gatcaaggcc gctcgcatcg ccatcaagca cggcgatatt cagcaggcca 19740
agtcgatgct cggcggtgcg ctcggtccat atctgacgga cgagtcgcag gcgaaggacc 19800
tggcgtacgc tctgaagatc gtgatcaaca gcgtgtacgg ccttaccgcg gcgtccttcg 19860
acaacgcatt tcgtgatccg cgaaacgtgg acaacatctg cgccaagcgg ggcgcgctgt 19920
tcatgatcga tctgaagcac ttcgtccagg agcagggctt cacggtcgca catatcaaga 19980
cggactcgat caagatcccg gacgcgacgc ccgagatcat atccgcggtc accgagttcg 20040
gcaagaagta cggctacacc ttcgagcacg aggccacgta cgaccgcatg tgcctcctga 20100
acgacgccgt ctatatcgct cacgacgaga agggctggca cgccacaggc gcccagttcc 20160
tgcaccccgta cgtttttcaag tcactgttca gccatgagga ggtgacactc gatgatttct 20220
gtgagacgaa ggccgtcacg accagcctgg ttctgggggga ccaggagact gcaccggagg 20280
gcatccgctt tgtgggcaga gtcggtcgct tcgtacctgt tcgcgaaggc ggcttcgata 20340
ttctgcgcga gaaggaggga aagtactcct tcgcgtccgg aagcaaggga tacaagtggc 20400
[0048] tcgaggcgga ggacctgaac ggcgatatct cgctggtgga cgaccgctac gccaactcgc 20460
tcgccaacaa ggcgcgcgac gccatatccg cattcggcga tgcggacgag ttcctcgcgc 20520
agaactaaca aagctcctaa cgagacccat agcccctaac ccgggctatg gctttcgttg 20580
gatcagaaag gaaacacaat ggagatcaag ttcgtaccca gaatcatcgt caccaaccac 20640
aacgcatatc gcgaggacac gcccttcgag acgcttgtgc acgcggcgat gaacgcagcg 20700
gtcaagaagt actgcggagt cgacaccccg atggagtgct gccccattcc gactcgagtg 20760
gtcgtctcgg ccatgggtaa tgtccgtctc gagtgggacg aggacgcaga gggcccgatg 20820
gtcatcgccg ggcgccacac ggacggacac tacggcctca tcgccgaggt ccgtccctcc 20880
aatgagaccg cccatattct cgacggctac cagttcgtcg gttacgccgt caaccccctc 20940
cgctccgtct acatgaggaa ggcatcatga gcagccacat cgtctccgtc cgttccgacg 21000
atccctgcat cgacgggatc aagatccttc tgtcggacgc cgtccgcaac gacgccgaca 21060
agtcgggata ccaccatttc ccggtcgtct caccgaagac cttcgacggg atcaagtcgt 21120
acatcatcgc cggctcgctc gatatctcgg aggactggta cgacgtccag aggaacggcc 21180
tctgcgtcaa tgtcatttct cgcaagagga ccgtcgagtg atctggatcc tctggacgat 21240
cctcgccgtg atgtccctct actatgcggg cacacggtgg ggatccgagt acgcatgggc 21300
cttgttctgc gccctgctcg tcatctacct catgcacttc cgcgtcatct gaaaggaaca 21360
cattcatgcc gtcagcacag atcaccctca aggacgtccg cctcctcttc cgcaacttca 21420
gcggtgcccg cgatcgtttc ggcaacgagg gcaagcggac gttctgcgtc gtcctgcccg 21480
aggacgctgc gaacgacctc cgtgcacagg gagtcaacgt cagggttctc aagccccgcg 21540
acgaggacga ccagccggag cccttcgtca aggtcaacgt gaagttcggg acccgcccgc 21600
cgaagatcac ggtcatctcc gggaagcacc ggacgatcct gaaggacgaa gcgtccctcg 21660
ggatgctgga ctgggcggtt attcagcgcg ccgatatcca gatccgcgtc tggcacaacg 21720
ccgagcgcgg gttcaccacc tgctggctga acaaggcctt catcaccgtc caggaggacg 21780
agctcgacgc gatgtacgcc gagcctccga gcgacgagga cgccccagag gagtggcggg 21840
catgaggatc acgcctgaca ccatttccaa gctgccgcag ctcgaggacg ttcgtgttcg 21900
tatcgtgcgg ccggacggct cctcggtcag tgcccgaggg gatatctacg ccgctctgcc 21960
gggcgacgac aagtggtccg tccttattcg gagcaacggc gtcgaagaac tcttcctctc 22020
gaccgagtgg gagggcatct ggttcgacgc cgcgaagctc ctgtgcctcc acagcaagcg 22080
cgtcgagacc gcccggtcat atccgatcga ggggatcagg tacctcctga tcagcgacat 22140
cgagtcgctg cacgacacct acgccgtatt cggcgaggac caggtctacc tgctcgagga 22200
ctccttcggg agccgcatga agggatacta caccttcacc ccctgccggg agtcccgtct 22260
tattcactgg gtccgtgtcg aagcgagctt cgtggacctc aaggtcgtcc accgaccgga 22320
cgacagcacg tcatatctct acgtcatgaa gaaggagccc gagcagtgaa cggtatgtgg 22380
aagatcatct ctcacgagga gttctacaac gacaggctgg ccacgatcgc cgacgggcag 22440
```

gtccgcggcc acatcgacga gaccggcttc cacggcggat tcgtcggggc atttcgtctc 22500
gtcctcgcga acggggtgga gctgatcatc cccaaggggc agttctactc gatcatcgac 22560
ttcgaagagg gcacgatcga gctgatcgcc gccgactgac catattcgga cgggggcctg 22620
gagcgatctg ggcccccgtc gacgcaaagg agcaccatgt taccctacaa ccgccttatt 22680
cttgtggact tcctcgaggg gccgaagatc tgggcgtact gcatcagaca gatcatcgat 22740
tatcagtgcg ggtaccgcaa ggacattctg gtgatggagc accgagatct cgagaccaat 22800
cagccgtacg agttcgtgct ggatacagat cgcttcagca tcgtccagga agtcgatccg 22860
tctcggaacg agtaccatct ggtcgtgatg gcctatggtt aacccgcccc ctcggccgac 22920
atatccgatc acggagtgga tcagggacga gaagtacctc ggaaagaggg agtactggat 22980
ccgatccggg ctcacgaaac agaggtactt gatatccagt aagggccgcg tgaagggtcc 23040
tcgcggccat atcgggtgcg ccaggcaggg cgagggcgtc gaggcccgaa atcgtcaggg 23100
caattcccag ttcttccgac tcgctcgtct gatgtgcgag gccttcacga ccttcgagta 23160
cgatccgaag aagcatcggt tctgctggtt ggacggcgac tcgaataacc gagccctctt 23220
gaacctcggc gtgcagatga tgacggggga ggggcggaaa agacaggaga ggcgcaggga 23280
ccctggctcg gcgcatatca ttcgtatcaa gatcgagggg gaggacgcga tgatcgaacg 23340
gcgcacccat atcccccgcg tcgttctgcc gaacatagtc aagaaggaga aactcaccga 23400
tatctggctc gtgtcgcacg gcatggtgga cgacgagcac aggtggacat actcagcaag 23460
gaggaagaca gatggggaac aagatcgttg acgcattcgt ccgggaggcc gttctggcag 23520
gcgcggatga tgtgatcatt tccgaggacg acgtggcccc cgagaccaac gccatcaagt 23580
tcgtggcggt gaagcgttga gcgaggagac ctggcggacc ctgtacgcat attacgatct 23640
tgaggtcagc agctacggcc gcttccgcaa gatcagcacg ggaacatatc gcaagccgac 23700
cctgaacaag gggtcgctcg tcatgagcgt cctggacgcc gagacgatgc agcagcagca 23760
gatggtcgcc caccggctcg tgtgggaggc atttcgcggg cccgtaccgc cgatgcactg 23820
catcaggcac cggatggagg acccgcgcga tattcgtctc gacaacctct actgccagtc 23880
gtactcggac ctcgtgtctg agaactggca ggggaagaag gaggagtggc gtcagacggc 23940
cttcgagatc gagggcttct gatggcggac ctgctgtgga agcaccagcg cgaggccctg 24000
cggaggatgc atatcggttg cgtcctctgc ggggcccggg gctccgggaa gagcatggcg 24060
ggcatcgcgt tcgccatcga gaccatcacg ggtcataatt ttgagtcggg gaagagcgcc 24120
gagaagccca cgacgctcta tatcatcacg accgcgacca agcgggacag tctcgactgg 24180
caggggggagc tcgcacggtt cggcctcacg cccaaggaac cgacgcgcgg catatcctgc 24240
gtcgtcgact catggaacaa catccggaag tacgagaagt cgaagggctt cttcatattc 24300
gacgagcagc gactcgtggg gaggggcgcc tgggcgaagg catatcgtaa gatcgccgcc 24360
[0049] cggaacccct ggatcctgct ctcgggcacg cccggagaca cctggatgga ctaccttcag 24420
gtatttctgg gcaacggctt ctaccgcaac cagacggagt tcgtggagaa gcacgtggag 24480
tacgaccgct tcgcgcgata tcctaaggtc aagcggtacc acaacgagaa ggtgctgcgg 24540
aagcaccggg acgccgtact ggtgcagatg ccgatggaga ggaggacccg gcgtcacgag 24600
cagcgcatat cctgcaagta cgagacagag acgtacaacg agacgctgcg gcaacggtgg 24660
gacgttcttc ggaaccgtcc gatcgaggac gctgcgggtc ttgcacgggc gctgcgctat 24720
atcgtagtgg cctcaggcga ccgcagaggc acccttctga ggctcctggc ggcccgtaag 24780
cgggctatcg tgttctacat gtacaactac gagcgcgaca tattacggga cgtctgcgag 24840
gctctggcga ggccgtgggg ggagatgaac ggaacctacc acgacggggt tcctcagagc 24900
gacgaggggg cgtatctggt gcagtaaacc tccggggccg aggcctggaa ctgcacgacc 24960
tgcgatacca tcatattcta cagcgagtcc tactcgtgga agcaggtgga acaggcgaag 25020
gggcgcatcg accgggccaa caccccgtac aaggacctgt actatatcac tctgacgagc 25080
gagtcgggct tcgaccgcgc agtcgagagg gccttgaaga ggaaggagcg gttcaacgaa 25140
gcggccttcg ggcgacgatg acgaggtccc ggggcggtct tacggccgtc tcgggttctc 25200
gtttttggcc agatcccagg agtggccatt tttttggcca ctgtggccag atttggccag 25260
tcgcgaatcg ttgcaatgac gccactttct ctatcttatg gccaaaaggc catttttcta 25320
ttaataaagt tactattaat aatagtaaac tatagccagt agagcctgta agttctcccg 25380
tggcgtggac aaacccgtcc cgtttttggc catttggcca ttcgaggact tcgcggcgga 25440
atcacgccca aaacgccgcg gccatttttg gccacgctgg acacggattt ggccactggc 25500
cattttcggc caagacccga ccccgtcttg ggtgcaatta cgcacacaga acacgggcca 25560
taatgaagat cctgacccga ccccgcacgg gtcatgcatt ttcaggaggc tcaatgctcg 25620
aacgagagtt ccaggctcgc ctcatcaagg agttgcacgc cattttcccc ggctgtatca 25680
tcttgaagaa cgacccgaac tacaagcagg gcgtgccgga ccttattctc ctctgggagg 25740
acaggtgggc catgctggag gtcaaggcct ccagtcgtgc caggcaccgc cccaaccagg 25800
actactacgt tgatcttctc gaccggatgt ccttcgccgc atttatttct ccggacaaca 25860
aggaacacgt cctagatgca cttcaacgat cactcacgtc tgaccggcgc ccacgcattt 25920
ctcagcgcga gtaagcacgc ctgggtcaac tacgacaccg atcgccttcg agaggtctat 25980
ggcaccagcc aggccgcagc tcggggcaca cggcttcacg agctcgctgc cgagcatatt 26040
cgcctgcgaa ttcgaatgcc gaagaaccag gtcacgctca accgctatat caacgacgct 26100
atcggctacc gtatgacacc ggagcagatc ctctggtaca gcgacaacgc cttcgggacg 26160
gcggacgcca tatccttcga cgagaagaag aaccttctcc gcgtccacga cctcaagacg 26220
ggcgttacgc cggcgtccat gacgcagctc catatctacg tggcgctgtt ctgcctcgag 26280
tacggctatc gcccgggcga tattcatgcg gagacccgta tctaccagaa caacgaggtc 26340
tttgtcgacg agcccgatgc cgaggtgctc tcgcatattc aggaccgact cgtccgcttc 26400

gacaagatca tcgagggact caagcaggag gcgtgatgac tgacgatttt ctcgtccact 26460
acggcacgcc ccgccactca gggcgctatc cctgggggtc gggcaaggac ccgtaccaga 26520
gcgcctcgtc atttctcagc gagcgcgatc gtctccgcaa ggagggcatg agcgacactg 26580
atattgcgcg tgcctgggga atgaccacca ccgagttccg agctcagggc agcatcgccc 26640
gcgccgagaa gaaggccggc gatattgctc gagccgttcg gctcaaggac gccggtctcc 26700
ccaataccgc catcgcggag aagatggggc ttaacgagtc gtctgttcgt gatctgctga 26760
agccggacgc cgggcatcgc caggacgcca tcaagcgcac cagcgatatt cttgaggcgg 26820
agtgcaaggc caagcggtat attgagtacg gcaccggcgt ggagatgaat ctccagacct 26880
cgaccgccac gctcaacaca gccgtcgagt ccttgcgcca gaaggggtac aagacccacg 26940
aggtctatgt caagcagccg ggacgggacg acttcacgat cctgaaggtg ctcgtcccgc 27000
cgggcgttac ccgcgccgag ctgatggcga atcgcgacaa gatccgcacc ccgggtgttg 27060
ccgtggacga gaacggcata tcgacaggaa tcaagccccc gacgaaggtg tcgagcaagc 27120
gtctgcgtat tcgctacgcc gaggacggtg gcacakacat ggacggcgtt attcagctcc 27180
gtcgcggtgt ccaggacatc agtctcggct cctcagcata tgctcaggtc cgtattaatg 27240
tcgacggcac gcactacctc aagggtatgg cgatgtacac cgacaagctc cccaagggcg 27300
ttgacatgat attcaacacc aacaagccca aagggacacc cgctctcggc ccgaaggacc 27360
actcggttct caagcccctc aaggacgacc cggagaatcc tttcggtacg gttgttcagc 27420
agcgatatta taaggacaag aagacgggta agcggaagct gtcggccctc aatattgtga 27480
acgaggaggg gtcctgggac cgctggtctg caactctggc gtcccagttc ctgtcaaagc 27540
agtcgcccgt attggcgaag aagcagctca aggccactcg cgaggcgaaa gagcgggagt 27600
tcaaggagat catgagtctc acgaatcctg ttattcggaa gaagctgctg aagtcctttg 27660
ccgaggactg cgactcggca gccgtacacc tcaaagccaa ggccctaccc ggtcaggcct 27720
cgcaggtcat attgccgatg ccgcatctca agaagaacca tgtgtatgca ccaaatttcc 27780
gcgaaggtcc cgccatgagg ataacgaaca agagacacca cagtaccatt cgagatcccg 27840
cagcttgtgg ttaacaacaa ggacaagaaa gcgcgtcgta ttatgggtcg tgctcgggat 27900
gctattggta ttcatccttc tgtcgccgag cgactcagcg agcgctgact tcgacggcga 27960
ctccgtcaca gccatccccc atattggcac gactcgggtc cgctccaccc cgccactgaa 28020
ggggctcaaa ggcttcgaac ctaagcgagt ctatccggca tatcccggga tgaagaggat 28080
gggggacact caggctcaga tgggcaagat cagcaacctt attacaggac atgaccctga 28140
agggggcttc agagtccgag ctcgctcgcg ctgtccgcca ctcgatggtt gttattgacg 28200
ctgagaagca ctacctcaat tacaaacaga gtgagcggga taacggtatt gctgcgcttc 28260
ggaagaagta tcagggcgaa aacgggcacg gtgcggccac gcttatttcc agggccaagg 28320
[0050] gtccggtctt catcgaagaa ggacatgacc ctgaaggggg cttcagagtc cgagctcgct 28380
cgcgctgtcc gccactcgat ggttgttatt gacgctgaga agcactacct caattacaaa 28440
cagagtgagc gggataacgg tattgctgcg cttcggaaga agtatcaggg cgaaaacggg 28500
cacggtgcgg ccacgcttat ttccagggcc aacggtccgg tcctcctcgg aagagcgcaa 28560
attgcgcaaa gcagcgaaag gcggccccat cgatcccaag accggtaaac tcgtttacga 28620
gaagactggc cggggatttt acaaccgcaa gggcaagtgg attcccaagg tcaccaaggc 28680
ccgtcggatg gccttggtgg acgatgcgca tatcctcagt agtggtactc atatggaagg 28740
catatacgcc gaccatgcta atgagttgaa ggcactggca aatcgtagcc gcaaggctgc 28800
tgtttctatc cccccgctga aacgagaccc ccggatggca cgaaagtacg cccccgaggt 28860
ttcaagcctc cgggccgccc tcaaccgagc catcaagcag aagcctttgg aacgtcaagc 28920
ggaaattatt gcgcaaggcg ttgtgtccaa gaaggtcgct gcaaatccgg atatgtctaa 28980
gaaggagcgg gcgaagatcg agtacatggc tgtcgagacg gctcgtgctc gtctcggaac 29040
tgatcgcaag ggtactcgcg tccgccccac ccctcgcgag tgggaagcta ttcaacgcgg 29100
cgctgtgtcc aatgcgatgt tggaggagat cgtggctaat gcggactctg atcacatcaa 29160
gaagctcgct atgccatgtg agaagactcg tgtcactaca gcacagcagt ctcgcatcat 29220
gacactgcgc agcaggggcg ctacacaggc tgagattgca gaggctctcg gtctctcggt 29280
ctcacaagtc aagtcagtga tctacaacga cgagtagcga cagctacact actacaggag 29340
gcagtggtgc ttgcctttag actcacaaca gaggacaatc cgttcaatcc tttcgacgag 29400
ttcgacgaat ggtacaagtt cgatgtgtct aatggctacg gcaccactgc ctacctgggt 29460
agggtgacat acagtagcga cgaactgtct cttgcagatc agattgaagc ttctaatgaa 29520
gctgtaattg aggcttttgc gctcaatctg actggcaact acaggattgt ggaacgagat 29580
gtcgaagtct gatgaaatga aatgacttaa cgaccatccg tcccccgggg gg 29632

<210> 3
<211> 306
<212> PRT
<213> Siphoviridae

<400> 3

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

```
        35                  40                  45
Leu Thr Asp Ser Ala Met Phe Asn Gly Ser Arg Val Gly Met Arg Thr
    50                  55                  60
Ile Thr Leu Thr Leu Leu Pro Leu Glu Val Pro Thr Val Glu Arg Ala
65                  70                  75                  80
Arg His Arg Ile Tyr Asn Leu Leu Pro Ile Lys Gln Pro Val Thr Ile
                85                  90                  95
Ile Val Arg Thr Asp Thr Arg Val Val Lys Thr Leu Gly Tyr Val Glu
            100                 105                 110
Ser Ser Glu Pro Asp Ile Phe Ser Lys Glu Glu Ala Ile Lys Val Val
        115                 120                 125
Leu Ile Cys Pro Asp Gly Tyr Trp Ser Asp Gly Ala Asp Asp Thr Gln
    130                 135                 140
Asn Tyr Leu Pro Phe Val Gln Glu His Ala Ala Phe Glu Phe Asp Trp
145                 150                 155                 160
Glu Asp Thr Pro Leu Ser Glu Ser Pro Thr Leu Ile Phe Ser Lys Thr
                165                 170                 175
Val Gly Leu Pro Ser Val Ile Leu Glu Asn Gly Gly Asp Val Pro Ala
            180                 185                 190
Gly Phe Thr Ile Met Ile Asp Ile Leu Lys Asp Asn Ala Thr Pro Ile
        195                 200                 205
Thr Ile Tyr Asp Asp Val Arg Asn Gln His Leu Thr Leu Thr Thr Lys
    210                 215                 220
Trp His Pro Asp Ala Ala Thr Thr Gln Pro Ala Lys Ala Gly Asp Arg
225                 230                 235                 240
Phe Tyr Ile Asn Thr Arg Val Gly His Lys Ala Val Thr Arg Ile Arg
                245                 250                 255
Asn Gly Lys Ser Glu Lys Ala Leu His Leu Leu Asp Ile Asp Ser Asp
            260                 265                 270
Trp Leu Met Leu Tyr Pro Gly Glu Asn Arg Leu Tyr Tyr Thr Ile Asn
        275                 280                 285
Val Gly Gly Gly Ala Lys Val Ser Leu Ser Arg Asp Ile Leu Tyr Gln
    290                 295                 300
[0051]
Gly Val
305

<210>  4
<211>  123
<212>  PRT
<213>  Siphoviridae

<400>  4
Met Glu Ile Lys Phe Val Pro Arg Ile Ile Val Thr Asn His Asn Ala
1               5                   10                  15
Tyr Arg Glu Asp Thr Pro Phe Glu Thr Leu Val His Ala Ala Met Asn
            20                  25                  30
Ala Ala Val Lys Lys Tyr Cys Gly Val Asp Thr Pro Met Glu Cys Cys
        35                  40                  45
Pro Ile Pro Thr Arg Val Val Val Ser Ala Met Gly Asn Val Arg Leu
    50                  55                  60
Glu Trp Asp Glu Asp Ala Glu Gly Pro Met Val Ile Ala Gly Arg His
65                  70                  75                  80
Thr Asp Gly His Tyr Gly Leu Ile Ala Glu Val Arg Pro Ser Asn Glu
                85                  90                  95

Thr Ala His Ile Leu Asp Gly Tyr Gln Phe Val Gly Tyr Ala Val Asn
            100                 105                 110
Pro Leu Arg Ser Val Tyr Met Arg Lys Ala Ser
        115                 120

<210>  5
<211>  175
<212>  PRT
<213>  Siphoviridae

<400>  5
```

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

<210> 6
<211> 20
<212> DNA
<213> 人工序列

<400> 6
gaggacactc cgaacgagac 20

<210> 7
[0052] <211> 20
<212> DNA
<213> 人工序列

<400> 7
atccccgaaa gtcaaccgac 20

<210> 8
<211> 20
<212> DNA
<213> 人工序列

<400> 8
ataatgttgc ggaagccgct 20

<210> 9
<211> 19
<212> DNA
<213> 人工序列

<400> 9
tttcgtgata cgggccaga 19

<210> 10
<211> 19
<212> DNA
<213> 人工序列

<400> 10
acgatgcggc gacgaagat 19

<210> 11
<211> 21
<212> DNA

<213>　人工序列

<400>　11
cagcgacagc ttcgagtcgt t　21

<210>　12
<211>　20
<212>　DNA
<213>　人工序列

<400>　12
ccttccggga catgaaggac　20

<210>　13
<211>　20
<212>　DNA
<213>　人工序列

<400>　13
ctgcgagcac cagtatagca　20

<210>　14
<211>　20
<212>　DNA
<213>　人工序列

<400>　14
aagaacgcca tcgaccagtt　20

<210>　15
<211>　20
<212>　DNA
[0053] <213>　人工序列

<400>　15
cgaacgtcgc aatgatcacc　20

<210>　16
<211>　20
<212>　DNA
<213>　人工序列

<400>　16
atcctgctgg acaacatccc　20

<210>　17
<211>　20
<212>　DNA
<213>　人工序列

<400>　17
ttccgaatag cgtcagcgaa　20

<210>　18
<211>　20
<212>　DNA
<213>　人工序列

<400>　18
atcgcgctta ttcaggtgct　20

<210>　19
<211>　20
<212>　DNA
<213>　人工序列

<400> 19
ttcggacgat cgctgtcatt 20

<210> 20
<211> 20
<212> DNA
<213> 人工序列

<400> 20
gggcctctcc gaatcttacg 20

<210> 21
<211> 20
<212> DNA
<213> 人工序列

<400> 21
agtcaggtgc tgattgcgaa 20

<210> 22
<211> 20
<212> DNA
<213> 人工序列

<400> 22
gagttcgact gggaggacac 20

<210> 23
<211> 20
<212> DNA
<213> 人工序列

[0054] <400> 23
tgttgtcgcc gaaatcgttg 20

<210> 24
<211> 20
<212> DNA
<213> 人工序列

<400> 24
agcgatgacc cctggttcta 20

<210> 25
<211> 20
<212> DNA
<213> 人工序列

<400> 25
gagcgagtgt actcggtgag 20

<210> 26
<211> 20
<212> DNA
<213> 人工序列

<400> 26
gacttcttca tgggcgacct 20

<210> 27
<211> 20
<212> DNA
<213> 人工序列

<400> 27
gatctaccag aacctgcccg 20

<210> 28
<211> 20
<212> DNA
<213> 人工序列

<400> 28
cgttccagga atgggcgaag 20

<210> 29
<211> 20
<212> DNA
<213> 人工序列

<400> 29
ccgcatcgta catcagggag 20

<210> 30
<211> 20
<212> DNA
<213> 人工序列

<400> 30
caggaccaca aggatgcgat 20

<210> 31
<211> 20
<212> DNA
<213> 人工序列

<400> 31
cgacgaactt cagacggtca 20

[0055] <210> 32
<211> 20
<212> DNA
<213> 人工序列

<400> 32
tactaccatc tcacccgcca 20

<210> 33
<211> 20
<212> DNA
<213> 人工序列

<400> 33
tcagtcacgc gcttctttct 20

<210> 34
<211> 20
<212> DNA
<213> 人工序列

<400> 34
atgagcccac actcgaaagg 20

<210> 35
<211> 20
<212> DNA
<213> 人工序列

<400> 35
caaggcgctc ttctctcgaa 20

<210> 36

<211> 20
<212> DNA
<213> 人工序列

<400> 36
cttgtgatac ttgcgctggc 20

<210> 37
<211> 20
<212> DNA
<213> 人工序列

<400> 37
tgagacccct cgaaaatgcc 20

<210> 38
<211> 20
<212> DNA
<213> 人工序列

<400> 38
ctggtcaacg accccaagaa 20

<210> 39
<211> 20
<212> DNA
<213> 人工序列

<400> 39
ccgtcatagc acagaaccgt 20

<210> 40
[0056] <211> 20
<212> DNA
<213> 人工序列

<400> 40
gtgtacgaga tggtggtcgg 20

<210> 41
<211> 20
<212> DNA
<213> 人工序列

<400> 41
gtgtcgagct ccttcatccc 20

<210> 42
<211> 20
<212> DNA
<213> 人工序列

<400> 42
ctgctgaagc tctacgctcc 20

<210> 43
<211> 20
<212> DNA
<213> 人工序列

<400> 43
tcgtactcgt gcatgtcgtc 20

<210> 44
<211> 20
<212> DNA

<213> 人工序列

<400> 44
aggatgacga catgcacgag 20

<210> 45
<211> 20
<212> DNA
<213> 人工序列

<400> 45
atcatgaagt ccgtgggctc 20

<210> 46
<211> 20
<212> DNA
<213> 人工序列

<400> 46
aagaaggatg gttcggcctg 20

<210> 47
<211> 20
<212> DNA
<213> 人工序列

<400> 47
gtactcgtcc cggcagtatg 20

<210> 48
<211> 20
<212> DNA
[0057] <213> 人工序列

<400> 48
acacgtacga gcgtctgatg 20

<210> 49
<211> 20
<212> DNA
<213> 人工序列

<400> 49
tggatctcct tccggaggtt 20

<210> 50
<211> 20
<212> DNA
<213> 人工序列

<400> 50
agtctccgcg accttatcct 20

<210> 51
<211> 20
<212> DNA
<213> 人工序列

<400> 51
cgtacacgta accgccttct 20

<210> 52
<211> 20
<212> DNA
<213> 人工序列

<400> 52
acaagtacga gttcggcgtt 20

<210> 53
<211> 20
<212> DNA
<213> 人工序列

<400> 53
agcgagatat cgccgttcag 20

<210> 54
<211> 20
<212> DNA
<213> 人工序列

<400> 54
cgtccggaag caagggatac 20

<210> 55
<211> 20
<212> DNA
<213> 人工序列

<400> 55
gcggaagtgc atgaggtaga 20

<210> 56
<211> 20
<212> DNA
<213> 人工序列

[0058] <400> 56
ttccgcgtca tctgaaagga 20

<210> 57
<211> 20
<212> DNA
<213> 人工序列

<400> 57
atatgacgtg ctgtcgtccg 20

<210> 58
<211> 21
<212> DNA
<213> 人工序列

<400> 58
ttattcactg ggtccgtgtc g 21

<210> 59
<211> 20
<212> DNA
<213> 人工序列

<400> 59
tccaaccagc agaaccgatg 20

<210> 60
<211> 20
<212> DNA
<213> 人工序列

<400> 60
cgtcagggca attcccagtt 20

<210> 61
<211> 20
<212> DNA
<213> 人工序列

<400> 61
aaccgatatg catcctccgc 20

<210> 62
<211> 20
<212> DNA
<213> 人工序列

<400> 62
ccttcgagat cgagggcttc 20

<210> 63
<211> 20
<212> DNA
<213> 人工序列

<400> 63
tcccgtaata tgtcgcgctc 20

<210> 64
<211> 20
<212> DNA
<213> 人工序列

<400> 64
gcgggctatc gtgttctaca 20

[0059] <210> 65
<211> 20
<212> DNA
<213> 人工序列

<400> 65
ctgcttgtag ttcgggtcgt 20

<210> 66
<211> 20
<212> DNA
<213> 人工序列

<400> 66
ctcatcaagg agttgcacgc 20

<210> 67
<211> 20
<212> DNA
<213> 人工序列

<400> 67
gcgcaatatc agtgtcgctc 20

<210> 68
<211> 20
<212> DNA
<213> 人工序列

<400> 68
tttctcgtcc actacggcac 20

<210> 69

<211> 20
<212> DNA
<213> 人工序列

<400> 69
atcgctgctg aacaaccgta 20

<210> 70
<211> 20
<212> DNA
<213> 人工序列

<400> 70
cgacccggag aatcctttcg 20

<210> 71
<211> 20
<212> DNA
<213> 人工序列

[0060]

<400> 71
cttcgatgaa gaccggaccc 20

<210> 72
<211> 18
<212> DNA
<213> 人工序列

<400> 72
ccaagggtcc ggtcttca 18

<210> 73
<211> 20
<212> DNA
<213> 人工序列

<400> 73
tccgcattag ccacgatctc 20

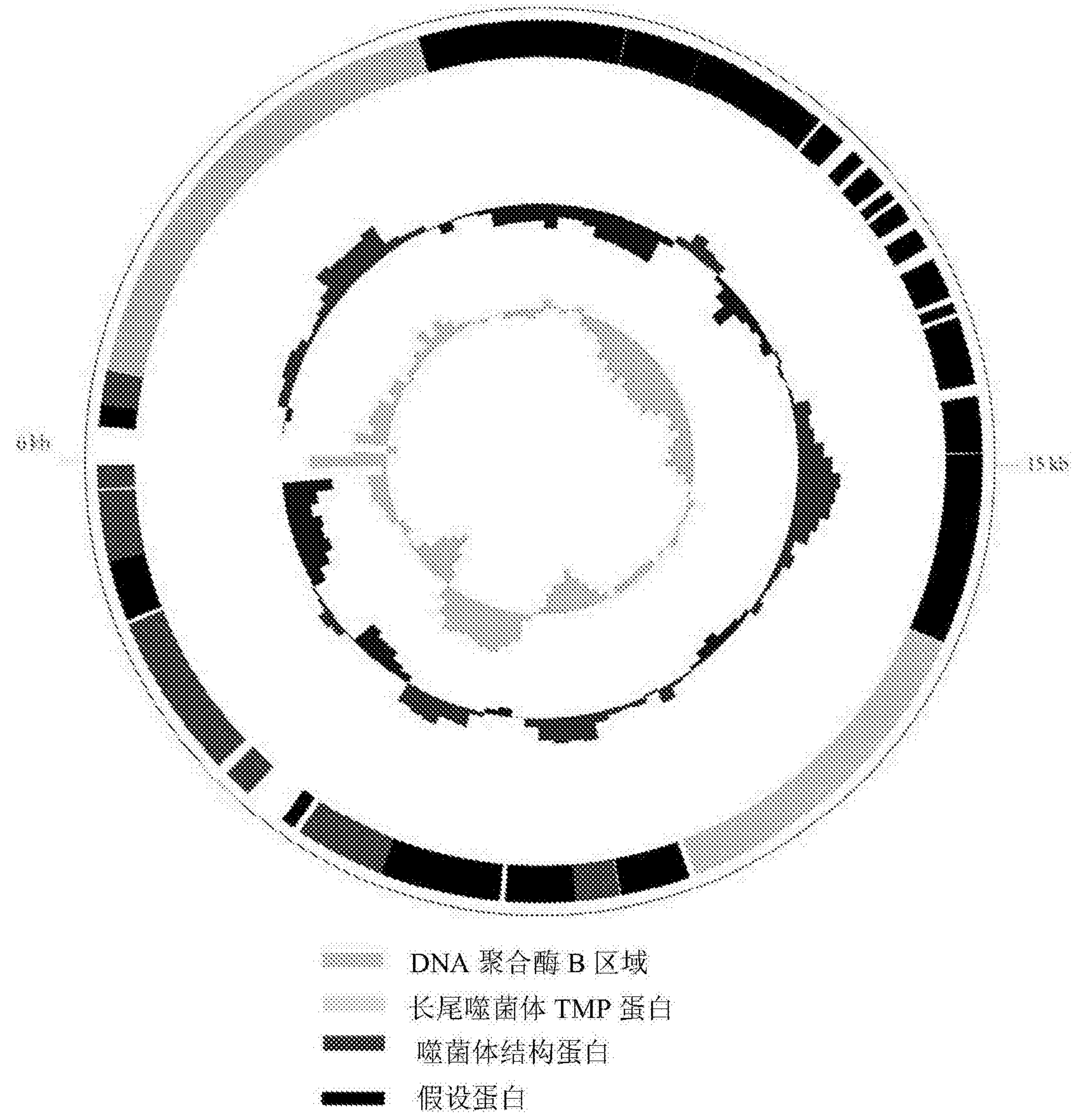


图1