



(12) 发明专利

(10) 授权公告号 CN 108513581 B

(45) 授权公告日 2022.02.08

(21) 申请号 201680060127.X
 (22) 申请日 2016.10.13
 (65) 同一申请的已公布的文献号
 申请公布号 CN 108513581 A
 (43) 申请公布日 2018.09.07
 (30) 优先权数据
 2015-202234 2015.10.13 JP
 (85) PCT国际申请进入国家阶段日
 2018.04.13
 (86) PCT国际申请的申请数据
 PCT/JP2016/080306 2016.10.13
 (87) PCT国际申请的公布数据
 W02017/065194 JA 2017.04.20
 (73) 专利权人 国立研究开发法人海洋研究开发
 机构
 地址 日本神奈川县

(72) 发明人 浦山俊一 布浦拓郎 出口茂
 (74) 专利代理机构 中国专利代理(香港)有限公
 司 72001
 代理人 张桂霞 鲁炜
 (51) Int.Cl.
 C12N 15/09 (2006.01)
 C12Q 1/686 (2018.01)
 G01N 33/53 (2006.01)
 G01N 33/569 (2006.01)
 C12N 7/00 (2006.01)
 (56) 对比文件
 US 2010279302 A1, 2010.11.04
 US 2010279302 A1, 2010.11.04
 US 2007031865 A1, 2007.02.08
 US 2014155274 A1, 2014.06.05
 审查员 王浩铭

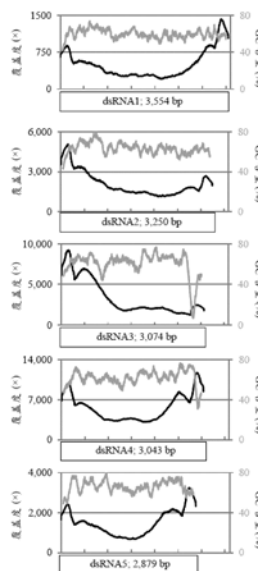
权利要求书2页 说明书20页
序列表(电子公布) 附图9页

(54) 发明名称

双链RNA的片段化方法及其利用

(57) 摘要

本发明的课题在于提供:可检测未知的病毒序列,且可有效地进行病毒的检测或探索的方法。所述方法包含:将对象双链(ds)RNA随机地片段化,获得dsRNA片段的步骤;将所得的dsRNA片段供于反转录反应及与其连续的聚合酶链反应(PCR),获得对应的DNA片段的步骤;和将所得的DNA片段供于序列分析操作,确定序列的步骤。反转录反应优选自dsRNA片段的3'末端开始。



1. RNA序列的确定方法,所述确定方法包含:
将试样中的dsRNA从DNA和单链RNA中分离,获得已纯化的双链RNA即dsRNA的步骤;
将已纯化的dsRNA随机地片段化,获得dsRNA片的步骤,此时所得的dsRNA片段为150~5000个碱基长;
将所得的dsRNA片段供于自dsRNA片段的3'末端开始的反转录反应及与其连续的聚合酶链反应即PCR,获得对应的DNA片段的步骤;
将所得的DNA片段供于序列分析操作,确定序列的步骤;和
将已确定的序列与已知的RNA病毒序列进行比较的步骤。
2. 权利要求1所述的方法,其中,已纯化的dsRNA的片段化是机械性、酶性、或化学性片段化。
3. 权利要求2所述的方法,其中,已纯化的dsRNA的片段化是基于超声波处理的机械性片段化。
4. 权利要求1~3的任一项中所述的方法,其中,已纯化的dsRNA的片段化是在所得的dsRNA片段的3'末端不残留磷酸基。
5. 权利要求4所述的方法,其进一步包含:将环引物与dsRNA片段的3'末端连接,获得连接引物的dsRNA片的步骤,
将所得的连接引物的dsRNA片段供于反转录反应及与其连续的聚合酶链反应即PCR。
6. 权利要求1~3和5的任一项中所述的方法,其中,已纯化的dsRNA来自RNA病毒。
7. 权利要求4所述的方法,已纯化的dsRNA来自RNA病毒。
8. 权利要求6或7所述的方法,其进一步包含:根据片段的序列确定RNA病毒的总长基因组序列。
9. 权利要求6或7所述的方法,其中,RNA病毒为未知的RNA病毒。
10. 权利要求1~3、5和7的任一项中所述的方法,其中,所得的dsRNA片段为1000~4000个碱基长。
11. DNA片段的制造方法,所述制造方法包含:
将试样中的dsRNA从DNA和单链RNA中分离,获得已纯化的双链RNA即dsRNA的步骤;
将已纯化的dsRNA随机地片段化,获得dsRNA片的步骤,此时所得的dsRNA片段为150~5000个碱基长;和
将所得的dsRNA片段供于自dsRNA片段的3'末端开始的反转录反应及与其连续的PCR,获得对应的DNA片段的步骤。
12. 权利要求11所述的方法,其中,已纯化的dsRNA来自RNA病毒。
13. 病毒的分析方法,所述分析方法包含:
将试样中的dsRNA从DNA和单链RNA中分离,获得已纯化的dsRNA的步骤;
将已纯化的 dsRNA随机地片段化,获得dsRNA片的步骤,此时所得的dsRNA片段为150~5000个碱基长;
将所得的dsRNA片段供于自dsRNA片段的3'末端开始的反转录反应及与其连续的PCR,获得对应的DNA片段的步骤;
将所得的DNA片段供于序列分析操作,确定序列的步骤;和
根据已确定的序列,分析试样中病毒的有无和/或特征的特征的步骤。

14. 权利要求13所述的方法,其中,试样来自生物或环境。

双链RNA的片段化方法及其利用

技术领域

[0001] 本发明涉及双链RNA的片段化方法和利用其的序列确定。本发明可以用于RNA病毒的全长基因组的序列确定、或者已知或未知的病毒的检测。本发明在生命科学和医疗等的领域是有用的。

背景技术

[0002] 病毒的检测和探索对理解病毒的生态学的作用和对控制病毒感染是重要的。作为以往的病毒检测方法,有基于利用抗原抗体反应来检测特定的病毒蛋白质的方法、和基于检测特定的病毒基因的碱基序列的方法。另外,RNA测序(RNA sequencing, RNA-Seq)是用于RNA病毒宏基因组学(metagenomics)的一般性方法,被广泛用于RNA病毒的鉴定或探索。该方法是下述的方法:全面性地解读试样中所含的核酸序列,在解读结果中,检测与已知的病毒序列或已知的病毒同源性高的序列(预测未知的病毒序列)。

[0003] 另一方面,DNA是担负几乎所有生物的遗传信息的分子,DNA测序成为用于分析(解析)遗传信息的基本手段。使用定序器的用于DNA测序的前处理主要包含以下的步骤。(1)由样品提取DNA、(2) DNA的片段化、(3) 向DNA末端附加接头序列(adapter sequence)、(4) DNA的扩增。在DNA的片段化中,采用超声波装置或使用喷雾器的物理破碎处理法。关于物理性的切断,在可以不依赖于碱基序列而将DNA无偏差地随机地片段化的角度,适于调制用于序列分析的DNA片段。关于基于超声波装置的片段化,非专利文献1中,对病毒的双链RNA适用。非专利文献1涉及基于双链RNA病毒的干扰素诱导的片段化效果。

[0004] 进行片段化时,切断末端磷酸基的有无对用于DNA扩增的接头序列的附加是重要的。专利文献1提案了:为了酶的检测方法,自核酸分子的3'末端去除磷酸基的切割方法,但不是以序列分析为目的。详细而言,专利文献1提案了用于检测试样中的酶的方法,该方法是对核酸分子或自核酸分子附加或去除化学性部分,由此,可以对核酸分子赋予延伸的能力以获得新型的可检测的核酸分子,该方法包含:使是否存在酶的试验的试样与核酸分子相互作用的步骤;及通过检测仪在酶的存在下所得的新型的核酸分子,试验酶是否与核酸分子相互作用的步骤。在此,优选为:酶是可以自核酸分子去除末端磷酸的磷酸酶,去除的末端磷酸存在于核酸分子的3'末端。

[0005] 现有技术文献

[0006] 专利文献

[0007] 专利文献1:特表2008-545384号公报

[0008] 非专利文献

[0009] 非专利文献1:J gen. Virol. (1974), 23, 191-195, Effect of fragmentation on interferon induction by double stranded virus RNA。

发明内容

[0010] 发明所要解决的课题

[0011] 为了实施以往的病毒检测方法,对病毒蛋白质特异性的抗体、或对目标病毒的基因序列特异性的核酸引物是必要的。因此所适用的病毒有限,另外,在这些方法中,无法检测事前没有信息的未知的病毒。另外,以往的RNA测序中所得的许多分析结果是来自非目标细胞的序列,被认为来自病毒的序列根据试样或分析条件有很大不同,但通常为1%以下。为了改善如此差的效率,研究了各种的改良,但原理上提高效率,会使操作烦杂,且发生偏差。因此,期望只要是可检测未知的病毒序列、且可有效地进行病毒的检测或探索的方法即可。

[0012] 用于解决课题的手段

[0013] 地球上存在单链DNA、双链DNA、单链RNA (ssRNA) 和双链RNA (dsRNA) 的4个核酸种。这些之中,细胞具有双链DNA和单链RNA,但病毒具有单链DNA、双链DNA、单链RNA和dsRNA。即,单链DNA及双链RNA是病毒特异性的核酸种。另外,许多RNA病毒在生命周期中有获取dsRNA状态的时期,但在健康体的植物、动物或真菌类的细胞中几乎不含dsRNA。因此,仅提取、纯化dsRNA,进行序列分析,便可以对RNA病毒进行分析。

[0014] 另外,在测序之前,如果可以将dsRNA随机地片段化至适当的尺寸,进行扩增,则对确定RNA病毒的全长基因组序列是有利的。

[0015] 本发明提供以下内容。

[0016] [1] RNA序列的确定方法,所述确定方法包含:

[0017] 将对象双链RNA (dsRNA) 随机地片段化,获得dsRNA片的步骤;

[0018] 将所得的dsRNA片供于反转录反应及与其连续的聚合酶链反应(PCR),获得对应的DNA片的步骤;和

[0019] 将所得的DNA片供于序列分析操作,确定序列的步骤。

[0020] [2] 1所述的方法,其中,反转录反应自dsRNA片的3'末端开始。

[0021] [3] 1或2所述的方法,其中,对象dsRNA的片段化是机械性、酶性、或化学性片段化。

[0022] [4] 3所述的方法,其中,对象dsRNA的片段化是基于超声波处理的机械性片段化。

[0023] [5] 1~4的任一项中所述的方法,其中,对象dsRNA的片段化是在所得的dsRNA片的3'末端不残留磷酸基。

[0024] [6] 5所述的方法,其中,进一步包含:将环引物与dsRNA片的3'末端连接,获得连接引物的dsRNA片的步骤,

[0025] 将所得的连接引物的dsRNA片供于反转录反应及与其连续的聚合酶链反应(PCR)。

[0026] [7] 1~6的任一项中所述的方法,其中,对象dsRNA来自RNA病毒。

[0027] [8] 7所述的方法,其用于确定RNA病毒的总长基因组序列。

[0028] [9] 7所述的方法,其用于确定未知的RNA病毒的序列。

[0029] [10] 1~9的任一项中所述的方法,其中,所得的dsRNA片为1000~4000个碱基长。

[0030] [11] DNA片的制造方法,所述制造方法包含:

[0031] 将对象dsRNA随机地片段化,获得dsRNA片的步骤;和

[0032] 将所得的dsRNA片供于反转录反应及与其连续的PCR,获得对应的DNA片的步

骤。

[0033] [12] 11所述的方法,其中,对象dsRNA来自RNA病毒。

[0034] [13] 病毒的分析方法,所述分析方法包含:

[0035] 将试样中的dsRNA从DNA和单链RNA中分离,获得已纯化的dsRNA的步骤;

[0036] 将所得的纯化dsRNA随机地片段化,获得dsRNA片段的步骤;

[0037] 将dsRNA片段供于反转录反应及与其连续的PCR,获得对应的DNA片段的步骤;

[0038] 将所得的DNA片段供于序列分析操作,确定序列的步骤;和

[0039] 根据已确定的序列,分析试样中病毒的有无和/或特征步骤。

[0040] [14] 13所述的方法,其中,试样来自生物或环境。

[0041] 发明效果

[0042] 根据本发明,可以有效地将从超过数万个碱基长(碱基长度)的长dsRNA分子至千个碱基左右的短dsRNA分子的序列相比以往进行均匀地确定。特别是,还可以确定在病毒的序列分析等中是重要的、且在以往技术中难以获得的、RNA基因组的两末端附近的序列。详细而言,本发明的手法由于以来自dsRNA病毒及ssRNA病毒的dsRNA为对象,因此通过确定存在于dsRNA的双链的3'末端序列,也可以确定原RNA病毒的5'末端附近的序列。

[0043] 另外,根据本发明,可以制造包含对应于RNA病毒的基因组的3'末端的序列的、RNA病毒基因组所对应的DNA片段。

[0044] 另外,根据本发明,可以有效地进行RNA病毒的探索。

[0045] 而且,根据本发明,可以确定已知或未知的RNA病毒的全长基因组序列。

附图说明

[0046] [图1] 将FLDS法用于已知dsRNA病毒(MoCV1-A)的基因组序列确定时的制图结果。作为试样,使用感染有由5条的dsRNA(3554、3250、3074、3043、2879 nt)构成的RNA病毒、即Magnaporthe oryzae chrysovirus 1 strain A(MoCV1-A)的稻瘟病菌(Magnaporthe oryzae),评价了FLDS的性能。通过将所得的序列数据接在一起,重构了MoCV1-A的全长基因组序列。另外,在各个分节基因组的末端区域,可见覆盖度(coverage)较中央区域变高的倾向。

[0047] [图2] 比较各自的每个病毒的序列出现频率。比较了来自各自的病毒序列的读出(read)在RNA-seq和FLDS数据中所出现的频率。将从RNA-seq数据检测到1读出以上的37个病毒序列进行绘图的结果,大部分的病毒的出现频率上升至100倍以上。显示出:与■所示的双链RNA病毒同样,△中所示的单链RNA病毒也在5例中的4例基于FLDS的浓缩是有效的。

[0048] [图3] RNA-seq和FLDS的比较。(a) RNA-seq中也对获得充分的读出数的3病毒序列比较了覆盖度变动的大小,因此算出了变动系数(表示标准偏差被平均而得的值的百分率)。其结果,可见FLDS中变动系数变低的倾向、即获得更均匀的覆盖度的倾向。■: 双链RNA病毒、△: 单链RNA病毒。(b~c) 比较了RNA-seq和FLDS中所得的病毒读出的覆盖度。

[0049] [图4] 由海水所得的RNA病毒候选序列(地点名:Jam)。在使用本发明方法的分析例中,获得了被认为705条的来自dsRNA的全长序列。半数以上的全长序列与已知RNA病毒基因不具有显著的序列同源性,除了新型RNA病毒序列之外,还获得了多数的完全新型的RNA

病毒候选序列。本发明的方法,无论所得的重叠群(コンテグ,contig)与已知RNA病毒基因是否具有显著的序列同源性,都可以完全地检测新型的RNA病毒的存在。以下的图中相同。

[0050] [图5] 由海水所得的RNA病毒候选序列(地点名:Jam)。

[0051] [图6] 由海水所得的RNA病毒候选序列(地点名:St.73)。

[0052] [图7] 由海水所得的RNA病毒候选序列(地点名:St.79)。

[0053] [图8] 由海水所得的RNA病毒候选序列(地点名:St.97)。

[0054] [图9] 由海水所得的RNA病毒候选序列(地点名:St.122)。

具体实施方式

[0055] 关于本发明和本说明书中的氨基酸、肽、碱基序列、核酸等的简略符号的表示,除了特别记载的情形,按照IUPAC-IUB的规定[IUPAC-IUB Communication on Biological Nomenclature、Eur. J. Biochem.、138: 9 (1984)]、“用于制作包含碱基序列或氨基酸序列的说明书等的指导原则”(专利局编)、和本技术领域常用的表现。

[0056] 将数值范围用“X~Y”表示时,除了特别记载的情形,包含两端的值X和Y。

[0057] 本发明提供RNA序列的确定方法。本发明的方法包含下述的步骤。

[0058] (1) 将对象dsRNA随机地片段化,获得dsRNA片的步骤;

[0059] (2) 将dsRNA片供于反转录反应及与其连续的聚合酶链反应(PCR),获得对应的DNA片的步骤;和

[0060] (3) 将所得的DNA片供于序列分析操作,确定序列的步骤。

[0061] [步骤1:dsRNA的片段化步骤]

[0062] <dsRNA分子的纯化>

[0063] 在步骤1中,对象dsRNA被片段化,但对象dsRNA在片段化前或后,可以进行纯化。纯化可以如下进行:从包含对象dsRNA的试样中首先提取总核酸种的步骤;及从所得的提取物中分离其他核酸种的步骤。从dsRNA的其他核酸种中的分离可以通过各种的方法进行,例如可以利用在羟磷灰石或纤维素等上的特异性吸附、使用苯酚的微分萃取、或对各种核酸酶的敏感性的差异等来进行。

[0064] 例如,纯化可以如下进行。从包含对象dsRNA的试样中通过以往法提取总核酸种之后,将所得的提取液流过纤维素柱,使dsRNA吸附于纤维素上。根据需要经过洗涤操作,进行使用适当的洗脱用缓冲液的洗脱操作。根据需要,为了分解dsRNA以外的其他核酸种可以在有效的条件下使核酸酶作用于所得的洗脱液。

[0065] 纯化可使用将4个核酸种分离而获取的串联色谱法(Syun-ichi Urayama et. al., A New Fractionation and Recovery Method of Viral Genomes Based on Nucleic Acid Composition and Structure Using Tandem Column Chromatography, Microbes and Environments Vol. 30 (2015) No.2, 199-203页)来进行。

[0066] <随机片段化>

[0067] 核酸的“随机(的)片段化”是指,除了特别记载的情形,以不依赖于碱基序列或位置的、不确定的样式的片段化。对象dsRNA的随机片段化可以机械性、酶性、或化学性地进行。片段化的程度(所得的dsRNA片的尺寸)只要可以实施持续的步骤,可以达成本发明的目的即可,没有特别限定,可以是任一的片段化手段,所得的dsRNA片的平均尺寸例如为

150~5000bp、优选为300~3000bp、更优选为1000~2000bp。

[0068] (机械性片段化)

[0069] 关于机械性的片段化,可以使用用于DNA片段化的现有的装置,适宜应用用于DNA片段化的条件来进行。基于机械化的片段化的例子,可举出:基于声波(例如,超声波)的处理、基于将对象dsRNA流过细的毛细管或孔的流体力学的剪切处理、基于dsRNA的喷雾化的处理。将片的平均大小进行一定地调节是比较容易的,如后所述,在可以制作无磷酸基的3'末端的角度,优选基于超声波的处理。

[0070] 基于超声波的处理例如可以如下进行:首先,将用于片段化的对象dsRNA溶解于适当的缓冲液中。dsRNA的量(或浓度)没有特别限定,确认到:即使是比较少的量、例如利用电泳无法检测的量的dsRNA也可以适用本发明。在此,将dsRNA溶液用现有的超声波破碎机处理,片段化至适当的尺寸为止。片段化可以在150bp~20000bp的范围下高度地控制来进行,为了本发明,例如片段化至150~5000bp、优选300~3000bp、更优选1000~2000bp为止。超声波处理的条件可以通过预实验来预先决定。关于超声波处理,为了防止温度上升,根据需要,可以边冷却边进行。片段化处理后,根据需要,使用电泳等的手段,可以确认已片段化的dsRNA的尺寸。关于所得的dsRNA片段,根据需要,在使用乙醇沉淀等的手段进行浓缩之后再溶解于适当的溶剂中。确认所得的dsRNA片段溶液的浓度,根据需要,调整为适当的浓度。

[0071] (酶性片段化)

[0072] 在此言及的酶性片段化是指,不依赖于碱基序列、可以将dsRNA无偏差地随机地、片段化至目标尺寸为止的方法。关于酶性片段化,在不需要超声波破碎装置等的特殊装置的角度,是优选的方法。切断被认为:现有的酶、例如RNase III系(Dicer等) (<http://www.amsbio.com/Recombinant-Human-Turbo-Dicer-Enzyme-Kit--siRNA.aspx>、http://catalog.takara-bio.co.jp/PDFFiles/6147_j.pdf)、RNase V1系(“RNase V1 preferentially cleaves phosphodiester bonds 3' of double-stranded RNA” Kertesz, Michael, et al. “Genome-wide measurement of RNA secondary structure in yeast.” *Nature* 467.7311 (2010): 103-107.)可将dsRNA随机地切断,可以用于本发明。如果通过酶性片段化所得的dsRNA片段的末端为突出末端,根据需要,也可以进行平滑化。需要说明的是,RNase III的切断产物的3'末端为羟基(Meister, Gunter, and Thomas Tuschl. “Mechanisms of gene silencing by double-stranded RNA.” *Nature* 431.7006 (2004): 343-349.)。

[0073] 基于酶的片段化处理例如可以如下进行:准备溶解于适当的溶剂中的dsRNA,加入已准备的酶混合液进行充分混合之后,以片段化上有效的温度(例如,30~45℃)开始反应。关于反应时间,例如在37℃下培育的情形、且想要获得100~1000bp的片段的情形,可以为15~25分钟。通过延长反应时间,可以获得更小的尺寸的dsRNA片段,通过缩短反应时间,可以获得更大的尺寸的dsRNA片段。供于反应的dsRNA少的情形等,优选使反应时间更长。通常,通过添加EDTA,可以停在反应。片段化处理后,根据需要,使用电泳等的手段,可以确认已片段化的dsRNA的尺寸。

[0074] (化学性片段化)

[0075] DNA的化学性片段化可以通过现有的手段进行实施。现有的手段的例子,可举出:酸或碱催化性水解、基于金属离子和复合物的水解、羟自由基处理或放射线处理。化学性片

段化的例子中也包含基于热的片段化。热片段化例如在约40℃以上进行。本领域技术人员可以理解pH和/或盐浓度之类的温度以外的参数对切断产生影响,并设计各种的条件。例如,热片段化的条件包含:在95℃(约80℃~100℃的温度范围)下、在低盐缓冲液(L-TE缓冲液)中、且中性pH(pH6.0~9.0)。

[0076] (在3'末端不残留磷酸基的片段)

[0077] 在优选的方案之一中,片段化以在3'末端不残留磷酸基的手段来进行。原因在于,无磷酸基的3'末端、即3'-羟基,通过一般的RNA连接酶(参照http://catalog.takara-bio.co.jp/product/basic_info.php?unitid=U100003136、http://www.promega.co.jp/Cre_Html.php?pGMPID=0206003),与具有5'-磷酸基的核酸以磷酸二酯键可以连接,因此容易与后述的引物连接。一般的RNA连接酶与3'-羟基及5'-磷酸基以外的组合、例如3'-磷酸基及5'-羟基、3'-羟基及5'-羟基、3'-双脱氧核苷酸及5'-磷酸基、3'-羟基及5'-三磷酸等不反应、不连接。

[0078] 通过步骤1所得的dsRNA片段供于接下来的步骤2。

[0079] [步骤2:反转录和PCR]

[0080] <自3'末端的反转录>

[0081] 在步骤2中,将步骤1中所得的dsRNA片段供于反转录反应,其次通过PCR,获得步骤1中所得的dsRNA片段所对应的DNA片段。关于本步骤,更详细而言,包含dsRNA的变性(单链化)、使用引物的反转录、RNA链的分解、基于DNA聚合酶的DNA扩增。

[0082] 关于反转录反应中所使用的引物,只要是在所使用的酶反应条件下成为自模板RNA反转录的起点的引物即可没有特别限定,但优选使用可以自模板RNA的3'末端开始反转录的引物。原因在于,可以获得在病毒的分析等中是重要的、且在以往技术中难以获得的、包含RNA 3'末端序列的反转录产物。可以自3'末端开始反转录的引物的例子中,包含在模板RNA的3'末端附加而使用的环引物;向模板RNA的3'末端附加多聚(A)核苷酸的同时所使用的寡dT引物;基于末端脱氧核苷酸转移酶的d(G)、d(T)或d(C)的附加的同时所使用的对应的引物;和向模板RNA的3'末端附加单纯的单链DNA接头的同时所使用的对应的引物。在优选的方案之一中,使用环引物。向模板RNA的3'末端附加引物或特定序列,可以通过本领域技术人员所公知的任意方法来实施。

[0083] 一般来说,引物的设计,通常留意于T_m值、各引物区域的末端稳定性、GC含量、二次结构这4个来进行。另外,为了防止引物二聚物的生成,使3'末端设计成非互补性。本发明所使用的引物中也同样。本发明中,也可利用现有的引物。

[0084] 使用环引物的情形,形成环的部分可以设计成20~80个碱基长,根据需要,也可以扩展至~100个、或~120个为止。整体的长度例如为40~200个碱基长、优选为50~100个碱基长。对多聚A部分或接头所使用的引物的情形,从进行特异性退火的观点考虑,长度优选为6个碱基长以上、更优选为9个碱基长以上。从DNA合成的观点考虑,优选为100个碱基长以下、更优选为30个碱基长以下。

[0085] 需要说明的是,以下,以使用环引物的情形为例来说明本发明,但本领域技术人员可以理解:其说明适用于适宜使用其他引物的情形。另外,有将包含dsRNA的随机片段化和环引物的使用的dsRNA序列的确定方法特别地称为FLDS(fragmented and loop primer ligated dsRNA sequencing)。

[0086] 关于环引物的附加、或多聚(A)核苷酸或单链DNA接头的附加,通常对dsRNA来进行,其次进行用于dsRNA变性(单链化)的反应。用于dsRNA变性的条件本领域技术人员可以适宜设定。只要是片段化至数百~5000bp左右的dsRNA即可,通常,可以通过在90~98℃下处理数秒~数分钟之后,进行骤冷来进行。

[0087] 关于反转录反应(cDNA链的合成),利用所附加的环引物或附加序列与引物的组合来进行。所使用的酶只要是以RNA为模板的具有DNA合成活性的酶即可,没有特别限定,例如可举出:来自禽类成髓细胞症病毒的反转录酶(AMV RTase,禽类成髓细胞性白血病病毒反转录酶)、来自莫罗尼鼠白血病病毒的反转录酶(MMLV RTase)、劳斯相关病毒2反转录酶(RAV-2 RTase)等各种起源的反转录酶。此外,还可以使用兼具反转录活性的DNA聚合酶。另外,为了本发明的目的,在高温下具有反转录活性的酶是适合的,例如可以使用来自栖热菌属细菌的DNA聚合酶、Tth (Thermus thermophilus,嗜热栖热菌) DNA聚合酶等或来自嗜热性杆菌属细菌的DNA聚合酶等。没有特别限定,例如优选来自嗜热性杆菌属细菌的DNA聚合酶,优选来自B.st (Bacillus stearothermophilus,嗜热脂肪芽孢杆菌)的DNA聚合酶(Bst DNA聚合酶)、以及来自Bca (Bacillus caldotenax,热坚芽孢杆菌)的DNA聚合酶。关于具有反转录酶活性的酶,在具有目标活性的范围,可以使用天然体、变异体的任一者。

[0088] 关于用于反转录反应的条件,本领域技术人员可以根据所使用的酶来适宜设定。如果是数百~5000bp左右的长度的RNA为模板的情形,则反转录反应例如可以在30~50℃下、实施数十分钟~数小时,接着在60~80℃下处理数分钟~数十分钟,由此可以将酶失活(不活性化)。通过反转录反应,形成RNA-DNA杂化物。

[0089] 所得的RNA-DNA杂化物的RNA链也可以通过化学性手法或适当的酶等进行分解。基于化学性手法的情形,例如可以使溶液变成碱性之后,进行加热,由此仅将RNA分解。使用酶的情形,关于所使用的酶,代表性的是核糖核酸酶H (Ribonuclease H、RNase H)。该酶为非特异性的核酸内切酶,通过水解而催化RNA切断。RNase H产生5'末端磷酸化物。关于用于RNA分解的条件,本领域技术人员可以根据所使用的酶来适宜设定。例如可以在30~50℃下实施数十分钟~数小时,接着在60~80℃下处理数分钟~数十分钟,由此可以将酶失活。在反转录反应之后或RNase H处理之后,可以进行去除剩余的模板RNA链的操作。通过分解RNA-DNA杂化物的RNA链,可以获得单链DNA。

[0090] RNA-DNA杂化物的RNA链分解的效果,有可以通过PCR的初期变性进行代替的情形。因此,代替使用RNase H来进行分解RNA的反应,接着,考虑使DNA与RNA的杂化物为牢固,可以将PCR反应的初期变性较长地设定反应时间。

[0091] 通过RNA-DNA杂化物的RNA链分解所得的单链DNA可以根据需要供于脱盐和浓缩步骤。关于单链DNA,可以通过在一旦达到高温之后缓慢地降低温度而退火,根据情形,用DNA聚合酶处理之后,可以获得双链的cDNA。双链cDNA通过接下来的PCR反应进行扩增。

[0092] <PCR反应>

[0093] 本发明所适用的PCR反应可以适用在同种目的中所利用的各种反应。关于PCR反应,通常,可以通过双链模板DNA向单链的变性、引物的退火、和由引物的互补链合成(延伸)的步骤来实施。另外,关于适用于本发明的用于PCR的条件,本领域技术人员可以适宜设计,例如将由变性(80~98℃、数分钟)、退火和基于聚合酶的延伸反应(65~75℃、数分钟)构成的处理进行5~50循环。该处理之后,可以进行用于去除小分子的cDNA或引物二聚物的处

理。通过PCR反应, dsRNA片段所对应的、已扩增的双链DNA片段, 供于接下来的序列确定步骤。PCR产物可以供于适当的纯化方法、例如供于Microcon-100之类的PCR产物纯化用的分子筛, 在纯化之后供于序列确定步骤。

[0094] [步骤3: 确定序列的步骤]

[0095] <序列数据的获取>

[0096] 序列确定步骤包含: 从作为PCR产物的各DNA片段获取碱基序列数据, 及通过这些序列数据进行全长序列的分析。

[0097] 在从DNA片段获取碱基序列数据中, 可以使用现有的定序器。另外, 也可以使用市售的测序试剂盒。在优选的方案之一中, 在碱基序列的确定中, 使用被称为下一代定序器的、用于边合成互补链DNA边进行读取的(Sequencing By Synthesis, SBS. Bentley et al. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456: 53-59)的手段。

[0098] 基于下一代定序器的序列确定步骤通常包含: 样品调制、簇(cluster)形成、测序、和数据分析的步骤。样品调制是制作用于通过定序器分析序列的文库的步骤, 通常包含: 片段化为200~600bp的DNA、附加接头序列及引物序列、根据需要附加用于识别DNA所来自的试样的标记(index, 标志)。为了样品调制, 对应于所用的定序器的各种的试剂盒被市售, 本发明中也可以利用。簇形成是为了将已调制的DNA文库在序列时增加至可检测的足够量而进行的扩增步骤。代表性的是, 簇制作在被称为流动池的玻璃基板上进行。基于SBS法的测序是用于读取依赖于碱基序列的荧光的步骤。SBS法是桑格法的改良法, 使用经荧光标记、且在3'末端具有保护基而使链延伸终止的4种碱基来进行。4种碱基通过荧光的种类进行识别。数据分析是综合数据以获得碱基序列的步骤。

[0099] 用于基于下一代定序器的序列确定的信息例如可从Illumina Technologies (Illumina的技术) (www.illumina.com) 获取, 而且可以参照W02004/018497、W02004/018493、W02004/050915、W02004/076692、W02005/021786、W02005/047301、W02005/065814、W02005/068656、W02005/068089、W02005/078130。

[0100] 利用下一代定序器, 可以生成每1运行(run)至少1000 读出(读取的片段)、每1运行至少10,000 读出、每1运行至少100,000 读出、每1运行至少500,000 读出、或每1运行至少1,000,000 读出。而且, 每1 读出可以生成约30个碱基长的序列数据、约40个碱基长的序列数据、约50个碱基长的序列数据、约60个碱基长的序列数据、约70个碱基长的序列数据、约80个碱基长的序列数据、约90个碱基长的序列数据、约100个碱基长的序列数据、约110个碱基长的序列数据、约120个碱基长的序列数据、约150个碱基长的序列数据、约200个碱基长的序列数据、约250个碱基长的序列数据、约300个碱基长的序列数据、约350个碱基长的序列数据、约400个碱基长的序列数据、约450个碱基长的序列数据、约500个碱基长的序列数据、约550个碱基长的序列数据、约600个碱基长的序列数据、约650个碱基长的序列数据、或约700个碱基长的序列数据、或其以上。

[0101] <序列数据处理>

[0102] 关于序列数据处理, 根据来自数百个碱基长的DNA的片段的序列数据, 为了重建更长的、优选全长的基因或全长基因组的序列而进行。为了序列数据处理, 开发了各种的程序, 也可以用于本发明。序列数据处理通常包含下述的各步骤: 来自定序器的数据组的输

入、序列数据接头和引物等的多余的序列数据的修剪、低质量的序列区域的去除、读出序列数据的装配(assembly)。序列的装配可以大致区分为2个(de-novo(从头):将读取的片段进行装配,重建目前为止未知的基因组序列;和制图(mapping):将现有的基因组序列作为参照序列,对其将读出进行制图)的方法,但本发明可以适合地实施任一者。

[0103] <作用、效果、优点等>

[0104] 通过上述的本发明的RNA序列的确定方法,可以有效地将从超过数万个碱基长的长dsRNA分子至千个碱基左右的短dsRNA分子的序列较以往均匀地确定序列。特别是,也可以确定在病毒的序列分析等中是重要的、且以往技术中难以获得的、RNA基因组的两末端附近的序列。详细而言,本发明的手法由于以来自dsRNA病毒及ssRNA病毒的dsRNA为对象,因此通过确定存在于dsRNA的双链的3'末端序列,也可以确定原RNA病毒的5'末端附近的序列。病毒基因组的末端序列包含重要的信息,能够确实地获取该信息是有利的。

[0105] 另外,病毒所具有的分节基因组的末端序列,多数的情形,在分节间保存。分节基因组之内,一个分节用同源性检索推定为病毒样序列、而其它分节与已知的病毒序列不显示同源性的情形,若末端序列共通则这些分节可以推定为来自相同的病毒。即,与已知的病毒序列不显示同源性的未知的分节也有通过本发明可以鉴定为病毒的可能性。

[0106] [本发明的应用、其他方式等]

[0107] <病毒的分析>

[0108] 本发明的方法可以用于确定RNA病毒的总长基因组序列。另外可以用于确定未知的RNA病毒的序列。

[0109] 通过本发明的方法确定dsRNA的序列,许多RNA病毒在生命周期之中具有保持dsRNA状态的时期,而健康体的植物、动物或真菌类的细胞中几乎不含dsRNA。因此dsRNA的存在表示病毒的存在,所确定的dsRNA的序列可以用于病毒分析。病毒的分析包含病毒的检测、和基于与已知病毒的序列比较的特征等。

[0110] 本发明的病毒的分析方法至少包含以下的步骤:

[0111] •将试样中的dsRNA从DNA和单链(ssRNA)中分离,获得已纯化的dsRNA的步骤;

[0112] •将所得的纯化dsRNA随机地片段化,获得dsRNA片段的步骤;

[0113] •将dsRNA片段供于反转录聚合酶链反应(RT-PCR),获得对应的DNA片段的步骤;

[0114] •将所得的DNA片段供于序列分析操作,确定序列的步骤;和

[0115] •根据已确定的序列,分析试样中病毒的有无和/或特征的步骤。

[0116] 对上述的RNA序列的确定方法进行的说明在病毒的分析方法中也适用。

[0117] <新型RNA病毒、其全长序列>

[0118] 本发明还提供:通过本发明的方法所发现的31条的新型的多核苷酸(全长序列)、由该多核苷酸的任一者构成的22个的病毒基因组、和包含该病毒基因组而成的病毒。本发明还提供:通过本发明的方法所发现的718条的来自新型病毒的多核苷酸、由该多核苷酸的任一者构成的病毒基因组、和包含该病毒基因组而成的病毒。具体而言,本发明提供以下的内容。

[0119] 下述的(A)、(B)或(C)的多核苷酸、由该多核苷酸构成的病毒基因组、或包含该病毒基因组而成的病毒:

[0120] (A) 由SEQ ID NO: 1~31、和SEQ ID NO: 34~751的任一项中记载的碱基序列组

成的多核苷酸；

[0121] (B) 与由与(A)中记载的多核苷酸互补的碱基序列组成的多核苷酸在严格的条件下杂交,且可构成与包含由(A)中记载的多核苷酸的任一者构成的病毒基因组而成的病毒在分类学上相同的变异病毒的基因组的多核苷酸；

[0122] (C) 与(A)中记载的多核苷酸具有高的同源性,且可构成与包含由(A)中记载的多核苷酸的任一者构成的病毒基因组而成的病毒在分类学上相同的变异病毒的基因组的多核苷酸。需要说明的是,在序列表中,<212>中所记载的序列的DNA型可以换读作RNA,序列中的t可以换读作u。

[0123] 对于SEQ ID NO: 1~31的多核苷酸,将病毒名、碱基长和分类汇总于下表中。表中,关于SEQ ID NO: 1、2、23、27和29,基因组结构与新型的病毒或动物病原病毒为相同科的病毒。

[0124] [表1]

病毒名称	图例	碱基长	分类	SEQ ID NO
歧蘼菌落伴随 dsRNA 病毒 1	DCADSRV-1 片段 1	1734	Viruses: dsRNA viruses: Potôviridae: unclassified	1
	DCADSRV-1 片段 2	1562		2
歧蘼菌落伴随 dsRNA 病毒 2	DCADSRV-2	4326	Viruses: dsRNA viruses	3
歧蘼菌落伴随 dsRNA 病毒 3	DCADSRV-3	4911	Viruses: dsRNA viruses: Totôviridae	4
歧蘼菌落伴随 dsRNA 病毒 4 基因组A型	DCADSRV-4 基因组A型	4982	Viruses: dsRNA viruses: Totôviridae	5
歧蘼菌落伴随 dsRNA 病毒 4 基因组B型	DCADSRV-4 基因组B型	4979	Viruses: dsRNA viruses: Totôviridae	6
歧蘼菌落伴随 dsRNA 病毒 5	DCADSRV-5	5252	Viruses: dsRNA viruses: Totôviridae	7
歧蘼菌落伴随 dsRNA 病毒 6	DCADSRV-6	4928	Viruses: dsRNA viruses: Totôviridae	8
歧蘼菌落伴随 dsRNA 病毒 7	DCADSRV-7	5327	Viruses: dsRNA viruses: Totôviridae	9
歧蘼菌落伴随 dsRNA 病毒 8	DCADSRV-8	4600	Viruses: dsRNA viruses: Totôviridae	10
歧蘼菌落伴随 dsRNA 病毒 9 基因组A型	DCADSRV-9 基因组A型	4844	Viruses: dsRNA viruses: Totôviridae	11
歧蘼菌落伴随 dsRNA 病毒 9 基因组B型	DCADSRV-9 基因组B型	4845	Viruses: dsRNA viruses: Totôviridae	12
歧蘼菌落伴随 dsRNA 病毒 10	DCADSRV-10	5082	Viruses: dsRNA viruses: Totôviridae	13
歧蘼菌落伴随 dsRNA 病毒 11	DCADSRV-11	5160	Viruses: dsRNA viruses: Totôviridae	14
歧蘼菌落伴随 dsRNA 病毒 12	DCADSRV-12	5941	Viruses: dsRNA viruses: Totôviridae	15
歧蘼菌落伴随 dsRNA 病毒 13	DCADSRV-13	4671	Viruses: dsRNA viruses: Totôviridae	16
歧蘼菌落伴随 dsRNA 病毒 14	DCADSRV-14 片段 1	1576	Viruses: dsRNA viruses: Partôviridae: unclassified	17
	DCADSRV-14 片段 2	1490		18
歧蘼菌落伴随 dsRNA 病毒 15	DCADSRV-15	12172	Viruses: dsRNA viruses: Endonôviridae: unclassified	19
歧蘼菌落伴随 dsRNA 病毒 16	DCADSRV-16	6635	Viruses: unclassified viruses	20
歧蘼菌落伴随 dsRNA 病毒 17 基因组A型	DCADSRV-17 基因组A型	5907	Viruses: dsRNA viruses	21
歧蘼菌落伴随 dsRNA 病毒 17 基因组B型	DCADSRV-17 基因组B型	5909	Viruses: dsRNA viruses	22
歧蘼菌落伴随 ssRNA 病毒 1	DCASSRV-1	11413	Viruses: ssRNA viruses: ssRNA positive-strand viruses, no DNA stage: Flavôviridae	23
歧蘼菌落伴随 ssRNA 病毒 2	DCASSRV-2	4506	Viruses: ssRNA viruses: ssRNA positive-strand viruses, no DNA stage: Namôviridae: unclassified Namôviridae	24
歧蘼菌落伴随病毒样 RNA 片段 1	病毒样 dsRNA-1	4567	Viruses: dsRNA viruses	25
歧蘼菌落伴随病毒样 RNA 片段 2	病毒样 dsRNA-2	4788	Viruses: dsRNA viruses	26
歧蘼菌落伴随病毒样 RNA 片段 3	病毒样 dsRNA-3	3458	Viruses: dsRNA viruses: Totôviridae	27
歧蘼菌落伴随病毒样 RNA 片段 4	病毒样 dsRNA-4	3190	Viruses: dsRNA viruses: Totôviridae	28
歧蘼菌落伴随病毒样 RNA 片段 5	病毒样 dsRNA-5	3262	Viruses: dsRNA viruses: Totôviridae	29
歧蘼菌落伴随病毒样 RNA 片段 6	病毒样 dsRNA-6	3325	Viruses: dsRNA viruses: Totôviridae	30
歧蘼菌落伴随病毒样 RNA 片段 7	病毒样 dsRNA-7	1988	Viruses: dsRNA viruses: Partôviridae: unclassified Partôviridae	31

[0125] [0126] 本发明中,关于多核苷酸,称之为“在严格的条件下杂交”时,除外特别记载的情形,任一种的多核苷酸,杂交的条件均可以按照Molecular Cloning. A Laboratory Manual. 2nd ed. (Sambrook et al., Cold Spring Harbor Laboratory Press)和Hybridization of Nucleic Acid Immobilization on Solid Supports (ANALYTICAL BIOCHEMISTRY 138, 267-284 (1984))的记载,使与要获取的多核苷酸一致进行适宜选定。例如获取具有85%以上同源性的DNA的情形,只要使用下述的条件即可:2倍浓度的SSC溶液和50%甲酰胺的存在下,在45℃下进行杂交之后,使用0.1倍浓度的SSC溶液(1倍浓度的SSC溶液的组成:由150mM氯化钠、15mM枸橼酸钠构成),在60℃下洗涤过滤器。另外,获取具有90%以上同源性的DNA的情形,只要使用下述的条件即可:2倍浓度的SSC溶液和50%甲酰胺的存在下,在50℃下进行杂交之后,使用0.1倍浓度的SSC溶液,在65℃下洗涤过滤器。

[0127] 本发明中,关于碱基序列(也称为核苷酸序列),称之为同源性时,除外特别记载的情形,是指在将2个序列以最佳方式排列的情形下,在2个序列间所共有的一致的核苷酸的个数的百分率。即,可以通过同源性=(一致的位置的数/位置的总数)×100算出,可以使用市售的算法进行计算。另外,这样的算法可编入Altschul et al., J. Mol. Biol. 215 (1990) 403-410中记载的NBLAST和XBLAST程序中。更详细而言,与碱基序列的同源性相关的检索、分析可以通过本领域技术人员周知的算法或程序(例如,BLASTN、BLASTP、BLASTX、ClustalW)来进行。关于使用程序的情形的参数,本领域技术人员可以适当地设定,另外,也可使用各程序的缺省参数。这些分析方法的具体手法也是本领域技术人员周知的。

[0128] 本说明书中,碱基序列的同源性“高”,除外特别记载的情形,是指具有至少70%、优选80%以上、更优选85%以上、进一步优选90%以上、进一步优选95%以上、进一步优选97.5%以上、进一步优选99%以上的序列同源性。

[0129] <DNA片段的制造方法>

[0130] 本发明还提供DNA片段的制造方法。该方法至少包含下述的步骤:

[0131] ·将对象dsRNA随机地片段化,获得dsRNA片的步骤;和

[0132] ·将所得的dsRNA片供于反转录反应及与其连续的PCR,获得对应的DNA片的步骤。

[0133] 通过该方法所得的DNA片段,

[0134] 对上述的RNA序列的确定方法进行的说明,在病毒的分析方法中也适用。

[0135] <试样>

[0136] 本发明中,使用包含对象dsRNA的试样,但试样可由来自所有的生物或环境的物质调制。关于试样,例如可以为:病毒、微生物、植物、或动物、它们的一部分(例如,器官或脏器、细胞等)、由它们所得的物质(例如,提取液、体液、排泄物等,它们的生活圈的一部分(例如,培养液、水系环境、土壤、空气等)。关于生物,可以有处于健康状态的情形,及处于疾患或任何的病态状态的情形。

[0137] 作为试样的具体例,可举出:细胞培养物(例如、哺乳动物细胞培养物和细菌培养物等)这样的来自生物体的试样,类病毒、病毒、细菌、霉、酵母、植物和动物这样的含核酸试样,有可能混入或感染病毒或细菌这样的微生物的试样(食品、生物学制剂等),或有可能含有土壤、废水(排水)、海水、和温泉水这样的生物的试样。另外,也可以是将前述试样等通过公知的方法进行处理所得的含核酸调制物。作为这样的调制物,例如细胞破碎物或将其分馏(分画)所得的试样、该试样中的核酸、或特定的核酸分子组例如将mRNA多量化(富化)的试样等可以用于本发明。而且,也可以适合地使用通过公知方法扩增上述试样中所含的核酸而得的DNA或RNA等的核酸等。

实施例

[0138] 以下所公开的技术是将本发明的方案之一以通过实验得到支持的目的表示,这是本领域技术人员可以理解的。本申请的保护范围依据权利要求书进行解释,不限于该实施例之项所记载的方式。

[0139] [材料和方法]

[0140] <Magnaporthe oryzae chrysovirus 1 strain A>

[0141] 将感染有Magnaporthe oryzae chrysovirus 1 strain A (MoCV1-A)的稻瘟病菌S-0412-II 1a菌株于YG液体培养基(0.5%酵母提取物、2%葡萄糖),在25℃下、以60r.p.m往返振荡培养了2周。(本菌是以越南产的稻瘟病菌进行操作,由于需要许可,因此在具有许可的东京农工大学的寺冈教授之下进行。另外,关于MoCV1-A,有相关的专利或专利申请(例如,EP2679675、US2011-0020289等)。

[0142] <硅藻试样>

[0143] 硅藻菌落是2014年4月从东京湾的退潮时海水滞留的地方(35.3405° N, 139.6396° E)取样。菌落用蒸馏水洗涤后,在-80℃下保管。

[0144] <dsRNA的纯化及片段化>

[0145] dsRNA纯化是将Okada et al.一部分改变来进行。将试样在液氮存在下,用乳钵进行破碎,提取了总核酸。使用填充有纤维素粉末(Cellulose D; ADVANTEC, Tokyo, Japan)的Micro Spin Column (empty Bio-spin column; Bio-Rad Laboratories, Inc., Hercules, CA, USA),将dsRNA进行了2次纯化。将已(关于MoCV1-A试样,到此为止在东京农工大学进行。)洗脱的核酸溶液进行调整(终浓度:57mM CH₃COONa、9.5mM MgCl₂、1.9mM ZnSO₄、189mM NaCl),将DNaseI (amplification grade, Invitrogen, Carlsbad, CA, USA)及S1核酸酶(Invitrogen)在37℃下处理了2小时。将所得的dsRNA溶液进行调制(终浓度:90mM CH₃COONa、15mM MgCl₂、3mM ZnSO₄、300mM NaCl),通过RNeasy Mini Kit (Qiagen, Valencia, CA)回收了dsRNA。

[0146] 向已回收的dsRNA中加入添加有ShortCut RNase III (NEB Japan, Tokyo, Japan)的10×ShortCut buffer及10×MnCl使达到1×。将溶液装入Snap-Cap microTUBE,通过Covaris S220 (Woburn, MA, USA) (条件:35秒运行(sec run),峰值功率140.0W,占空因数2.0%,和200循环/爆发(cycles/burst))将dsRNA进行了片段化。将其分为2等分,制作添加有ShotCut RNase III (NEB)的样品和未添加ShotCut RNase III (NEB)的样品,在37℃下保持了40分钟。其后,使用ZymoClean Gel RNA Recovery Kit (ZymoResearch, Orange, CA)将dsRNA进行了回收。需要说明的是,该实验中使用100ng左右的dsRNA,在预定为用于切断成约1500bp的条件下进行片段化。

[0147] <cDNA合成与扩增>

[0148] 按照Potgieter, et al. (“Improved strategies for sequence-independent amplification and sequencing of viral double-stranded RNA genomes.” Journal of General Virology 90.6 (2009): 1423-1432.)中记载的方法,将PC3-T7环引物(5' -p-GGA TCC CGG GAA TTC GGT AAT ACG ACT CAC TAT ATT TTT ATA GTG AGT CGT ATT A-OH-3' (SEQ ID NO:1))与片段化dsRNA连接。连接后的dsRNA通过MinElute Gel Extraction Kit (Qiagen)进行了浓缩、纯化。加入DMSO使终浓度达到15%(v/v)之后,在95℃下热处理3分钟,在冰上进行骤冷。将已附加的环引物部分作为引物替代,使用Superscript III First-Strand Synthesis System (Invitrogen)进行了反转录反应。去除DNA-RNA杂化物的RNA之后,将对应的DNA通过MinElute PCR cleanup kit (Qiagen)进行了浓缩、纯化。所得的DNA缓慢地由95℃降至50度,由此与具有互补序列的DNA进行退火。在KOD-plus Neo (Toyobo, Osaka, Japan) PCR溶液的热活化后,加入经退火的DNA,在68℃下维持,由此制成完全的双链DNA。使用以下的PC2引物(5' - CCGAATCCCGGGATCC-3')在下述的条件下进行

了基于PCR反应的DNA扩增。96℃ 2分钟; [98℃ 10秒, 68℃ 2分钟] 18循环。扩增产物所含的引物等的低分子产物使用SPRIselect reagent kit (Beckman Coulter, Brea, CA, USA) 来去除。

[0149] <总RNA提取、cDNA合成、文库构建>

[0150] 使用TRIzol Plus RNA Purification Kit (Invitrogen)由硅藻菌落提取了总RNA。将所得的RNA组分(画分)用DNaseI (Takara, Otsu, Japan)处理,由此去除了残存的DNA。具有对应于RNA的序列的dsDNA通过PrimeScript Double Strand cDNA Synthesis Kit (Takara)使用随机引物(9碱基)进行了合成。所得的dsDNA使用Qubit dsDNA HS Kit进行了定量。

[0151] <使用Illumina的序列分析>

[0152] 将cDNA使用Covaris S220 (Woburn, MA, USA) (4℃、Snap-Cap microTUBE、55秒,峰值功率175.0W、占空因数5.0%、循环/爆发(cycles/burst) 200循环)通过超声波进行了片段化。Illumina序列文库使用KAPA Hyper Prep Kit Illumina platforms (Kapa Biosystems, Woburn, MA, USA),按照附带的手法进行了构建。文库使用KAPA library quantification kit (Kapa Biosystems)进行了定量。将所得的文库使用Illumina MiSeq platforms (San Diego, CA, USA)进行了配对(paried-end)序列分析。

[0153] <序列数据处理>

[0154] 通过测序分析所得的原始序列(生配列)使用CLC Genomics Workbench (CLC Bio, Aarhus, Denmark)进行了操作。去除了低质量序列、和序列接头或PC2引物序列、作为对照而添加的PhiX序列、实验上的混入序列(0.05%以下)。将残留的序列进行从头装配,根据该序列使用Tablet viewer按照手册进行了重叠群的确认和延伸。最终,所得的重叠群中,将平均覆盖度(average coverage)为10以上、最低覆盖度为3以上、1000个碱基长以上的物质供于以后的分析。特别优先的序列在10 读出以上一定的位置结束的情形,将其部位认定为重叠群的末端。该认定正确性也得到如下的支持:修剪前的这些序列中在末端序列直后存在PC2序列。通过(其中,除外在碱基长具有变异的保有polyA序列的病毒)重叠群间的比较分析可见具有70-9-%的序列同源性的重叠群的情形,这些序列作为相同病毒种的不同基因组型。可见具有90%以上序列同源性的重叠群的情形,仅将这些序列作为相同基因组型的序列优先的重叠群序列用于分析。装配序列的操作使用了Genetyx-MAC software version 17.0.0 (Genetyx Corp., Tokyo, Japan)。small subunit rRNA的获取使用了EMIRGE。

[0155] <系统分析>

[0156] 根据在RNA病毒中共通地保存的RNA依赖RNA合成酶RNA-dependent RNA polymerase (RdRp)的氨基酸序列推定了各病毒的系统关系。经de novo(从头)装配的重叠群中的推定氨基酸序列和已知RNA病毒的RdRp序列的多重对比使用ClustalX 2.0和MEGA5软件进行了制作。根据多重对比的系统推定使用MrBayes 3.2.3,氨基酸取代模型使用了RtREV+I+G+F。

[0157] [使用已知病毒的FLDS的性能评价]

[0158] 作为试样,使用感染有Magnaporthe oryzae chrysovirus 1 strain A (MoCV1-A)的稻瘟病菌(Magnaportheoryzae),评价了FLDS的性能。需要说明的是,MoCV1-A是由5条

的dsRNA (3554, 3250, 3074, 3043, 2879 nt)构成的RNA病毒。

[0159] 通过将所得的序列数据连接在一起,重建了MoCV1-A的全长基因组序列。经重建的MoCV1-A序列与在数据库(DNA Data Bank of Japan、<http://www.ddbj.nig.ac.jp/index-j.html>)上公开的MoCV1-A序列(AB560761~AB560764和AB700631)几乎一致(>99.9%)。另外,在分节基因组各自的末端区域可见覆盖度较中央区域变高的倾向(参照图1)。

[0160] 通过FLDS,确定了dsRNA分子的全长序列。另外,根据FLDS,则不需要通常在RNA分子的末端序列确定所使用的RACE法等。通过该方法明确了:可以期待包含重要信息的末端序列的确实获取。另外,如果末端序列共通,则可以推定它们的分节为来自相同的病毒,因此与已知的病毒序列不显示同源性的未知分节也有可以鉴定为病毒的可能性。

[0161] [从环境试样中探索RNA病毒]

[0162] (1) 探索了硅藻的群体1g的试样中所含的RNA病毒。通过由所得的序列数据重建的序列,将42条认定为病毒序列。其中31条获得了全长序列,由此再构成了22个的病毒基因组(所有新种)(参照表2)。●标记的序列在数据库上具有同源性的已知序列不存在,但根据基因组末端序列或近缘病毒的基因组结构,确定了是病毒基因组的一部分。

[0163] [表2]

由 31 条的全长序列重构的 22 个病毒基因组及 7 个病毒样序列

RNA 病毒种	类型	尺寸 (nt)	制图的读出的数量	平均覆盖率	BlastX 分析
DCADSRV-1	●	1,734	1,301,278	191,942	-
		1,562	1,717,396	279,580	Fox picobirnavirus Picovirnaviridae
DCADSRV-2		4,026	1,337,570	83,876	Ustilagoidea virens nonsegmented 病毒 1 Not assigned
DCADSRV-3		4,911	14,544	703	Ustilagoidea virens RNA 病毒 1 Totiviridae
DCADSRV-4	基因组 A 型	4,982	12,325	591	Aspergillus mycovirus 178 Totiviridae
DCADSRV-4	基因组 B 型	4,979	1,074	52	Ustilagoidea virens RNA 病毒 1 Totiviridae
DCADSRV-5		5,252	7,863	359	Aspergillus foetidus slow 病毒 1 Totiviridae
DCADSRV-6		4,939	2,720	131	Aspergillus mycovirus 178 Totiviridae
DCADSRV-7		5,327	1,957	87	Gremmeniella abietina RNA 病毒 L1 Totiviridae Ustilagoidea virens RNA 病毒 3 Totiviridae
DCADSRV-8		4,660	1,163	60	Aspergillus foetidus slow 病毒 1 Totiviridae
DCADSRV-9	基因组 A 型	4,844	1,198	60	Magnaporthe oryzae 病毒 2 Totiviridae
DCADSRV-9	基因组 B 型	4,845	364	18	Aspergillus foetidus slow 病毒 1 Totiviridae
DCADSRV-10		5,082	1,244	59	Rosellinia necatrix victorivirus 1 Totiviridae Ustilagoidea virens RNA 病毒 1 Totiviridae
DCADSRV-11		5,160	1,173	55	Ustilagoidea virens RNA 病毒 1 Totiviridae Ustilagoidea virens RNA 病毒 1 Totiviridae
DCADSRV-12		5,941	1,219	49	Beauveria bassiana RNA 病毒 1 Totiviridae
DCADSRV-13		4,671	820	42	Aspergillus foetidus slow 病毒 1 Totiviridae
DCADSRV-14		1,576	438	67	Persimmon cryptic 病毒 Partitiviridae
●		1,490	274	43	-
	DCADSRV-15		12,172	1,482	29
DCASSRV-1		11,413	1,011	21	Border disease 病毒 - BD31 Flaviviridae
DCASSRV-2		4,586	4,153	224	Tuber excavatum mitovirus NaRNAviridae
DCADSRV-16		6,635	8,735	310	Rhizoctonia fumigata mycovirus Not assigned
DCADSRV-17	基因组 A 型	5,907	5,325	218	dsRNA virus environmental sample Not assigned
DCADSRV-17	基因组 B 型	5,909	1,564	63	Botrytis porri RNA 病毒 1 Not assigned
DCAVLS-1	断续的 RdRp	4,567	57,802	3,039	Ustilagoidea virens nonsegmented 病毒 1 Not assigned
DCAVLS-2	断续的 RdRp	4,786	41,181	2,100	Ustilagoidea virens nonsegmented 病毒 1 Not assigned
DCAVLS-3	仅 CP	3,458	13,140	876	Ustilagoidea virens RNA 病毒 1 Totiviridae
DCAVLS-4	仅 RdRp	3,190	3,995	294	Magnaporthe oryzae 病毒 2 Totiviridae
DCAVLS-5	仅 CP	3,262	1,331	96	Phomopsis vexans RNA 病毒 Totiviridae
DCAVLS-6	仅 RdRp	3,325	891	65	Ustilagoidea virens RNA 病毒 3 Totiviridae
DCAVLS-7	断续的 RdRp	1,986	164	20	Flammulina velutipes browning 病毒 Partitiviridae

基准：平均覆盖度为 10 以上、最低覆盖度为 3 以上、末端序列为 10 读出以上

[0165] (2-1) 其次,比较了通常RNA病毒探索所利用的RNA-seq法及FLDS。对于已重建的42条的病毒序列,将用各自的分析所得的读出进行制图的结果,相对于FLDS中98.2%的读出所制图的结果,RNA-seq停留于0.3% (参照表3)。另外,仅使用RNA-seq数据重建病毒基因组的情形,仅获得了6条的部分序列。不存在仅用RNA-seq数据所检测的病毒。

[0166] [表3]

[0167] 所得的读出的内译

	FLDS		总 RNA-seq	
	读出的数量	比率(%)	读出的数量	比率(%)
[0168] 已修剪	4,631,738	100.0	6,979,561	100.0
主要的病毒的读出	4,549,629	98.2	24,036	0.3
未制图的读出 (包括次要的病毒的读出)	82,109	1.7	6,955,525	99.6

[0169] (2-2) 比较了来自各自的病毒序列的读出在RNA-seq和FLDS数据中所出现的频率(参照图2)。将由RNA-seq数据检测出1 读出以上的37个病毒序列进行绘图的结果,大部分的病毒的出现频率上升了100倍以上。关于△所示的单链RNA病毒,显示出:5例中有4例基于FLDS的浓缩是有效的。

[0170] (2-3) 比较了RNA-seq和FLDS中所得的病毒读出的覆盖度(参照图3)。RNA-seq中也对获得充分的读出数的3个病毒序列比较了覆盖度的变动大小,因此算出了变动系数(表示标准偏差被平均而得的值的百分率)。其结果,可见FLDS中变动系数变低的倾向、即获得更均匀的覆盖度的倾向。

[0171] [从海水中探索RNA病毒]

[0172] •实验方法

[0173] <海水试样采取及病毒粒子纯化>

[0174] 由计5地点[Jam, St73, St79, St97, St122] (参照下表)对海洋表层水取样。将各地点2L的海水用孔径0.2 μ m的醋酸纤维素膜过滤器进行过滤,过滤器在-80℃下保管。滤液中所含的病毒粒子按照John et al. (“A simple and efficient method for concentration of ocean viruses by chemical flocculation.” Environmental microbiology reports 3.2 (2011): 195-202.)中记载的方法进行了浓缩、保管。将病毒浓缩试样溶解后,进一步供于氯化铯密度梯度离心(274,000g、48小时),将密度1.30-1.48 (g/cm³)的区域回收、浓缩,由此获得了纯化病毒粒子。

[0175] [表4]

地点名	Jam	St. 73	St. 79	St. 97	St. 122
时间(mm/dd/yyyy)	12/21/2015	07/27/2014	07/30/2014	08/04/2014	08/13/2014
坐标 (纬度, 经度)	35.199, 139.392	47.0122, 160.0192	46.9718, 166.7472	46.9859, -179.4263	47.0035, -151.4048
深度(m)	<1	<1	<1	<1	<1
盐分浓度(ppt)	-	-	32.9282	32.8586	32.4992
叶绿素 a (mgL ⁻¹)	-	-	-	0.29	0.61

[0177] <核酸提取及RNA纯化>

[0178] 将醋酸纤维素膜过滤器上的细胞以每过滤器在液氮存在下,用乳钵进行破碎,按

照Urayama et al. (“FLDS: a comprehensive dsRNA sequencing method for intracellular RNA virus surveillance.” *Microbes and Environments* 31.1 (2016): 33.)提取了总核酸。纯化病毒粒子也溶解于同样的核酸提取液,提取了总核酸。从这些总核酸溶液中将dsRNA和ssRNA按照Urayama et al. (“A new fractionation and recovery method of viral genomes based on nucleic acid composition and structure using tandem column chromatography.” *Microbes and Environments* 30.2 (2015): 199.)进行了分馏。

[0179] 所得的dsRNA依据硅藻的情形使用填充有纤维素粉末(Cellulose D; ADVANTEC, Tokyo, Japan)的Micro Spin Column (empty Bio-spin column; Bio-Rad Laboratories, Inc., Hercules, CA, USA)进行了2次纯化。将已洗脱的核酸溶液进行调整(终浓度:57mM CH₃COONa、9.5mM MgCl₂、1.9mM ZnSO₄、189mM NaCl),将DNase I (amplification grade, Invitrogen, Carlsbad, CA, USA)及S1核酸酶(Invitrogen)在37°C下处理了2小时。将所得的dsRNA溶液进行调制(终浓度:90mM CH₃COONa、15mM MgCl₂、3mM ZnSO₄、300mM NaCl),通过RNeasy Mini Kit (Qiagen, Valencia, CA)回收了dsRNA。而且,使用无核酸酶的水,将已洗脱的dsRNA溶液进行调整(终浓度:200mM NaCl、20mM Tri-HCl pH 8.0、2mM EDTA pH 8.0),依据硅藻的情形将dsRNA进行了片段化。

[0180] <cDNA合成及扩增>

[0181] 按照Potgieter, et al. (“Improved strategies for sequence-independent amplification and sequencing of viral double-stranded RNA genomes.” *Journal of General Virology* 90.6 (2009): 1423-1432.)中记载的方法,将U2引物(5'-p-GAC GTA AGA ACG TCG CAC CA-p-3', SEQ ID NO: 32)与片段化dsRNA进行连接。连接后的dsRNA通过MinElute Gel Extraction Kit (Qiagen)进行了浓缩、纯化。使用U2-comp引物(5'-OH-TGG TGC GAC GTT CTT ACG TC-OH-3', SEQ ID NO: 33),通过SMARTer RACE 5' /3' Kit (TaKaRa, Japan)进行了反转录反应。将DNA-RNA杂化物的RNA去除后,使用U2-comp引物及UPM引物(SMARTer RACE 5' /3' Kit添加),将cDNA进行了PCR扩增。PCR使用KOD-plus Neo (Toyobo, Osaka, Japan),在以下的条件下进行。96°C 2分钟; [98°C 10秒、60°C 15秒、68°C 2分钟] 30-35循环。扩增产物中所含的引物等的低分子产物使用SPRIselect reagent kit (Beckman Coulter, Brea, CA, USA)进行了去除。

[0182] <使用Illumina的序列分析>

[0183] 依据硅藻的情形进行。

[0184] <序列数据处理>

[0185] 由通过测序分析所得的原始序列(生配列),使用Trimmomatic version 0.32 (Bolger et al. “Trimmomatic: a flexible trimmer for Illumina sequence data.” *Bioinformatics* (2014): btu170)去除了接头序列及低质量序列。使用Cutadapt version 1.9.1 (Martin. “Cutadapt removes adapter sequences from high-throughput sequencing 读出.” *EMBnet. journal* 17.1 (2011): pp-10)去除了cDNA合成和扩增中所使用的引物序列。使用Bowtie2 version 2.2.5 (Langmead & Salzberg. “Fast gapped-read alignment with Bowtie 2.” *Nature methods* 9.4 (2012): 357-359)去除了作为对照而添加的PhiX序列、实验上的混入序列。将50个碱基长以下的序列通过Trimmomatic

version 0.32去除,用于以下的分析。

[0186] 将所得的序列使用CLC Genomics Workbench (CLC Bio, Aarhus, Denmark) 进行了从头装配。根据平均覆盖度为3以上的装配序列,使用Tablet viewer version 1.14.10.20 (Milne et al. “Tablet-next generation sequence assembly visualization.” *Bioinformatics* 26.3 (2010): 401-402) 及PRICE version 1.2 (Ruby et al. “PRICE: software for the targeted assembly of components of (Meta) genomic sequence data.” *G3 (Bethesda)*. (2013): 20; 3(5): 865-80),按照手册进行了重叠群的确认和延伸。最终,在所得的重叠群中,将平均覆盖度(average coverage,平均盖度)为10以上、最低覆盖度为3以上、500个碱基长以上的物质供于以下的分析。特别优先的序列在10 读出以上一定位置结束的情形,与硅藻的情形同样,将其部位认定为重叠群的末端。

[0187] •结果

[0188] <RNA病毒探索>

[0189] 由5地点、各2L的计10L的海水试样探索了RNA病毒。由根据所得的序列数据重建的序列明确了:656个的新型RdRp基因得以检测,至少存在656个的RNA病毒。另外,明确了:这些RNA病毒的一部分与已知的非反转录RNA病毒44家族中27家族的病毒显示同源性,对于已知的非反转录RNA病毒的半数以上,其近缘种在海水中生息(参照下表b)。目前为止以175L的海水为对象进行的RNA病毒研究中,仅检测到7家族的RNA病毒(参照下表a),显示出:该手法是非常有效的RNA病毒探索手法。

[0190] [表5]

a

175L → 7 家族 (*)

<i>Alphaflexiviridae</i>	<i>Coronaviridae</i>	<i>Ophioviridae</i>
<i>Alphatetraviridae</i>	<i>Cystoviridae</i>	<i>Orthomyxoviridae</i>
<i>Alvernnaviridae</i>	<i>Dicistroviridae*</i>	<i>Paramyxoviridae</i>
<i>Amalgaviridae</i>	<i>Endornaviridae</i>	<i>Partitiviridae</i>
<i>Arenaviridae</i>	<i>Filoviridae</i>	<i>Permutotetraviridae</i>
<i>Arteriviridae</i>	<i>Flaviviridae</i>	<i>Picobimaviridae</i>
<i>Astroviridae</i>	<i>Gammaflexiviridae</i>	<i>Picornaviridae*</i>
<i>Barnaviridae</i>	<i>Hepeviridae</i>	<i>Potyviridae</i>
<i>Benyviridae</i>	<i>Hypoviridae</i>	<i>Quadriviridae</i>
<i>Betaflexiviridae</i>	<i>Iflaviridae</i>	<i>Reoviridae*</i>
<i>Birnaviridae</i>	<i>Leviviridae</i>	<i>Rhabdoviridae</i>
<i>Bornaviridae</i>	<i>Luteoviridae</i>	<i>Roniviridae</i>
<i>Bromoviridae</i>	<i>Marnaviridae*</i>	<i>Secoviridae*</i>
<i>Bunyaviridae</i>	<i>Megabirnaviridae</i>	<i>Togaviridae</i>
<i>Caliciviridae</i>	<i>Mesoniviridae</i>	<i>Tombusviridae*</i>
<i>Carmotetraviridae</i>	<i>Narnaviridae</i>	<i>Totiviridae</i>
<i>Chrysoviridae</i>	<i>Nodaviridae*</i>	<i>Tymoviridae</i>
<i>Closteroviridae</i>	<i>Nyamaviridae</i>	<i>Virgaviridae</i>

[0191]

(Culley, A. I. et. al., 2014, *Mbio*)
(Steward, G. F. et. al., 2013, *ISME*)
(Culley, A. I. et. al., 2006, *Science*)

b

10L → 26 家族 (*)

<i>Alphaflexiviridae</i>	<i>Coronaviridae</i>	<i>Ophioviridae</i>
<i>Alphatetraviridae*</i>	<i>Cystoviridae*</i>	<i>Orthomyxoviridae</i>
<i>Alvernnaviridae*</i>	<i>Dicistroviridae*</i>	<i>Paramyxoviridae</i>
<i>Amalgaviridae*</i>	<i>Endornaviridae*</i>	<i>Partitiviridae*</i>
<i>Arenaviridae</i>	<i>Filoviridae</i>	<i>Permutotetraviridae</i>
<i>Arteriviridae</i>	<i>Flaviviridae*</i>	<i>Picobirnaviridae*</i>
<i>Astroviridae</i>	<i>Gammaflexiviridae</i>	<i>Picornaviridae*</i>
<i>Barnaviridae*</i>	<i>Hepeviridae</i>	<i>Potyviridae*</i>
<i>Benyviridae</i>	<i>Hypoviridae*</i>	<i>Quadriviridae*</i>
<i>Betaflexiviridae</i>	<i>Iflaviridae*</i>	<i>Reoviridae*</i>
<i>Bimaviridae</i>	<i>Leviviridae*</i>	<i>Rhabdoviridae</i>
<i>Bornaviridae</i>	<i>Luteoviridae*</i>	<i>Roniviridae</i>
<i>Bromoviridae</i>	<i>Marnaviridae</i>	<i>Secoviridae</i>
<i>Bunyaviridae</i>	<i>Megabimaviridae*</i>	<i>Togaviridae</i>
<i>Caliciviridae</i>	<i>Mesoniviridae</i>	<i>Tombusviridae*</i>
<i>Carmotetraviridae</i>	<i>Narnaviridae*</i>	<i>Totiviridae*</i>
<i>Chrysoviridae*</i>	<i>Nodaviridae*</i>	<i>Tymoviridae</i>
<i>Closteroviridae*</i>	<i>Nyamaviridae</i>	<i>Virgaviridae*</i>

[0192]

[0193] <使用全长序列的完全未知RNA病毒的推定>

[0194] 关于该手法,无论所得的重叠群与已知RNA病毒基因是否具有显著的序列同源性,均可以完全地检测新型的RNA病毒的存在。本分析中获得被认为来自705条的dsRNA的全长序列(SEQ ID NO: 34~738)。半数以上的全长序列与已知RNA病毒基因不具有显著的序列同源性,获得了多数的RNA病毒候选序列(图4~9)。

[0195] [高温酸性温泉的环境微生物]

[0196] •实验方法

[0197] <温泉试样采取>

[0198] 将目前为止未发现RNA病毒的被推定为archaea优先的九州云仙的高温酸性温泉水用孔径0.2 μ m的醋酸纤维素膜过滤器进行过滤,将过滤器在-80 $^{\circ}$ C下保管。

[0199] <核酸提取及dsRNA纯化>

[0200] 依据硅藻试样来进行。

[0201] <cDNA合成及扩增>

[0202] 依据海水试样来进行。

[0203] <使用Illumina的序列分析>

[0204] 依据海水试样来进行。

[0205] <序列数据处理>

[0206] 依据海水试样来进行。

[0207] •结果

[0208] <RNA病毒探索>

[0209] 在所得的重叠群中,未发现与已知RNA病毒序列显示显著的同源性的重叠群。

[0210] <使用全长序列的完全未知RNA病毒的推定>

[0211] 获得被认为来自13条的dsRNA的全长序列,全读出的48%制图成它们的序列。虽然在使用Blastx的同源性检索中未显示显著的值,但在一部分的重叠群中检测到RdRp中高度地保存的GDD基序。另外,13条中有8条的两末端的序列在重叠群间保存,若参考碱基长信

息,则预测到这8条的重叠群来自由2分节构成的4种RNA病毒(SEQ ID NO: 739~751)。根据以上的结果,强烈暗示到:archaea中也存在RNA病毒。

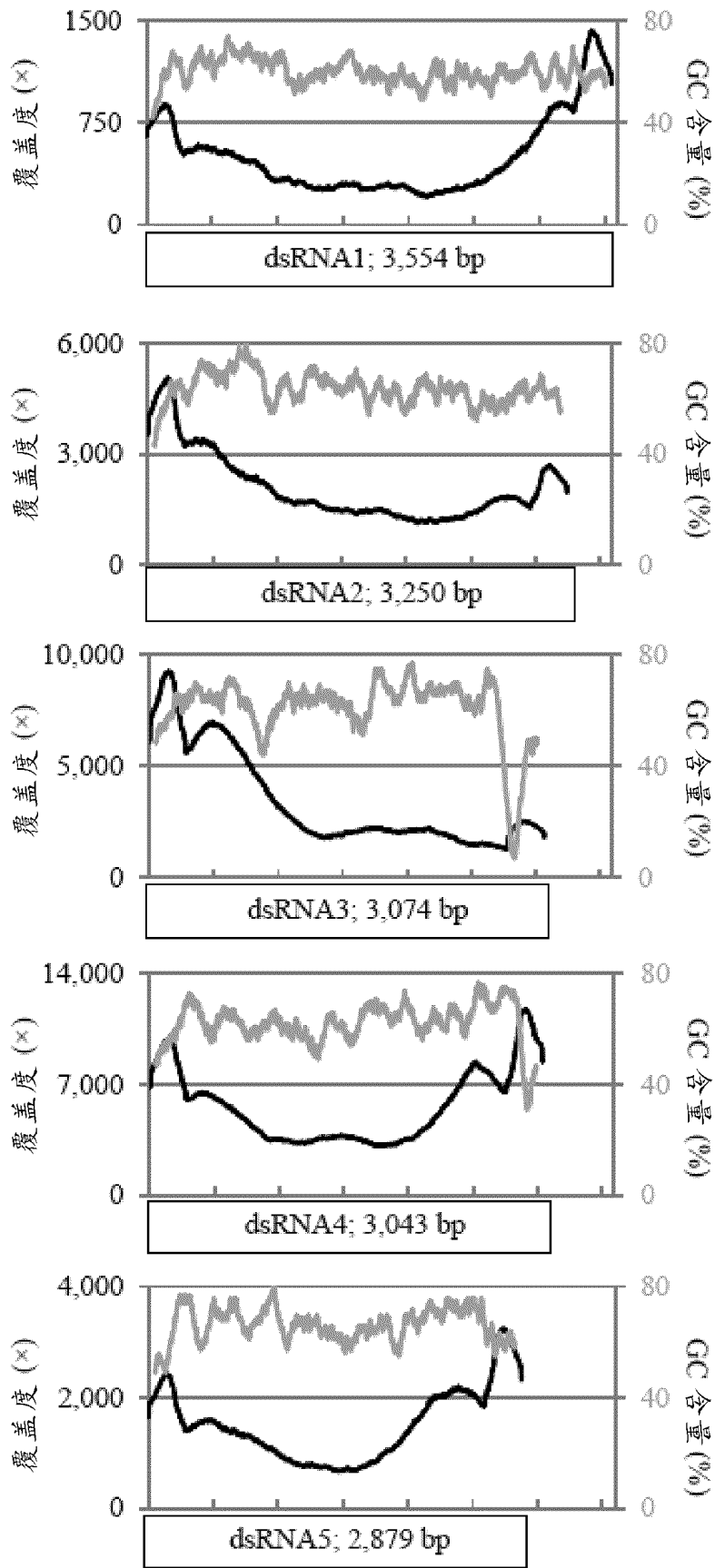


图 1

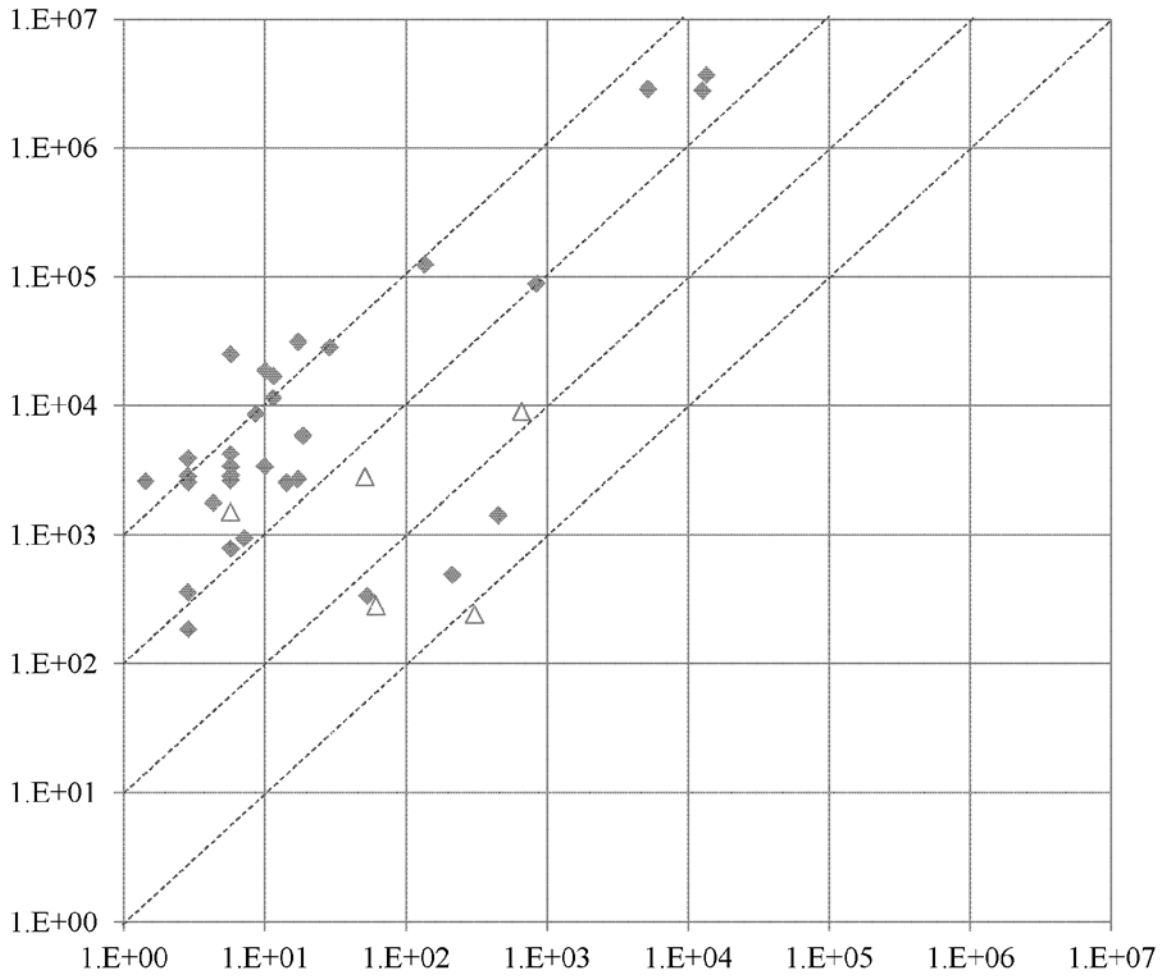


图 2

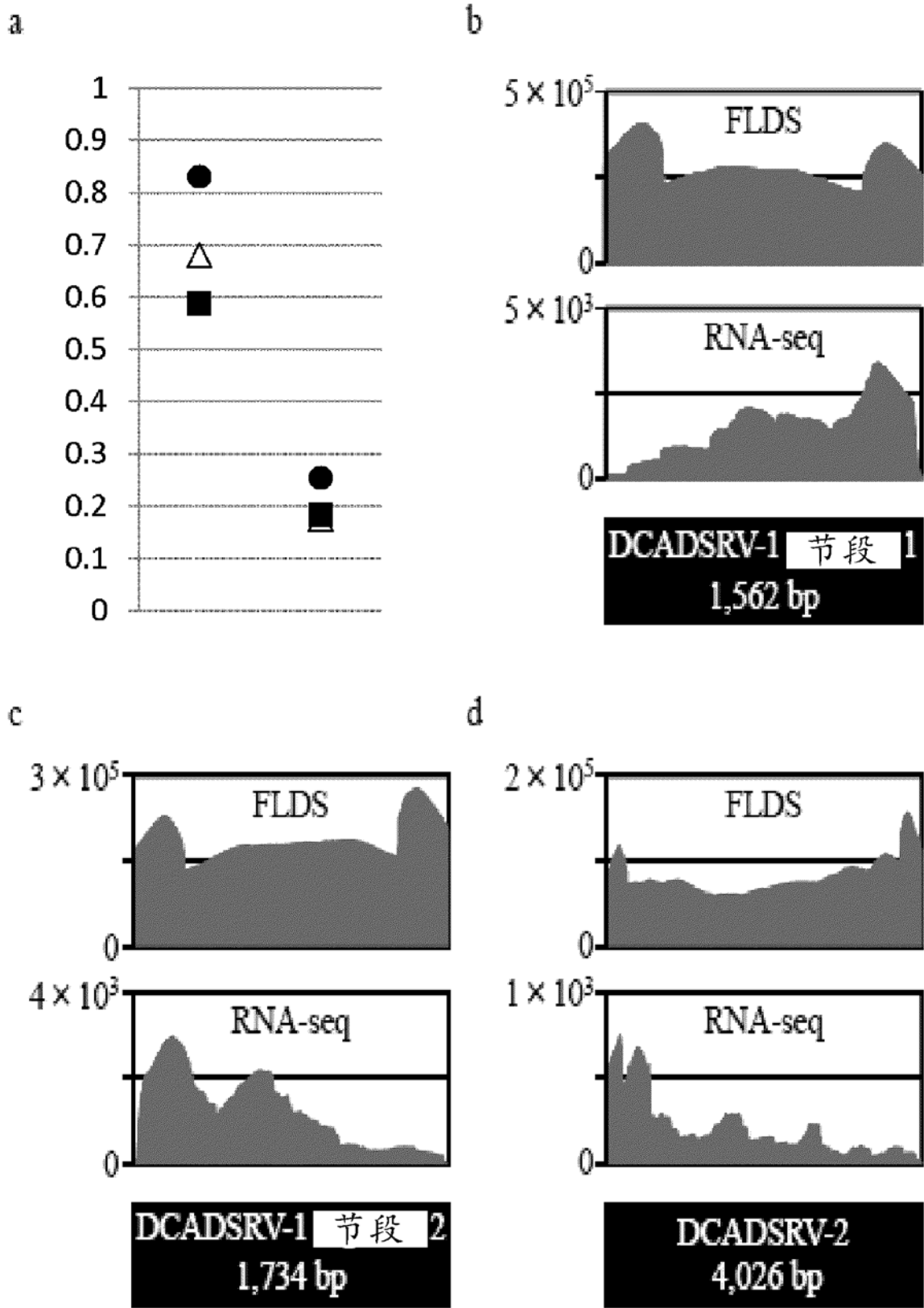


图 3

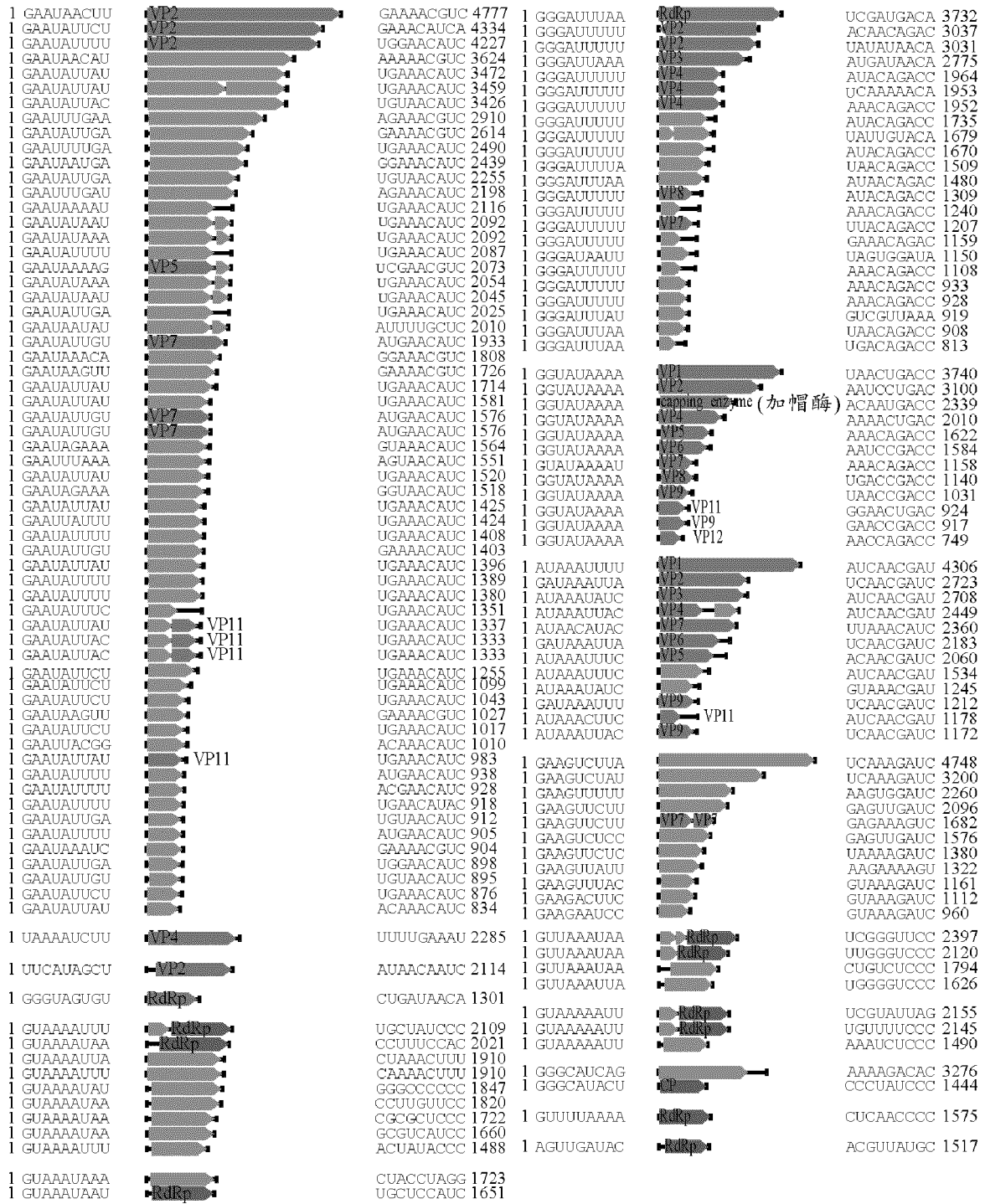


图 4

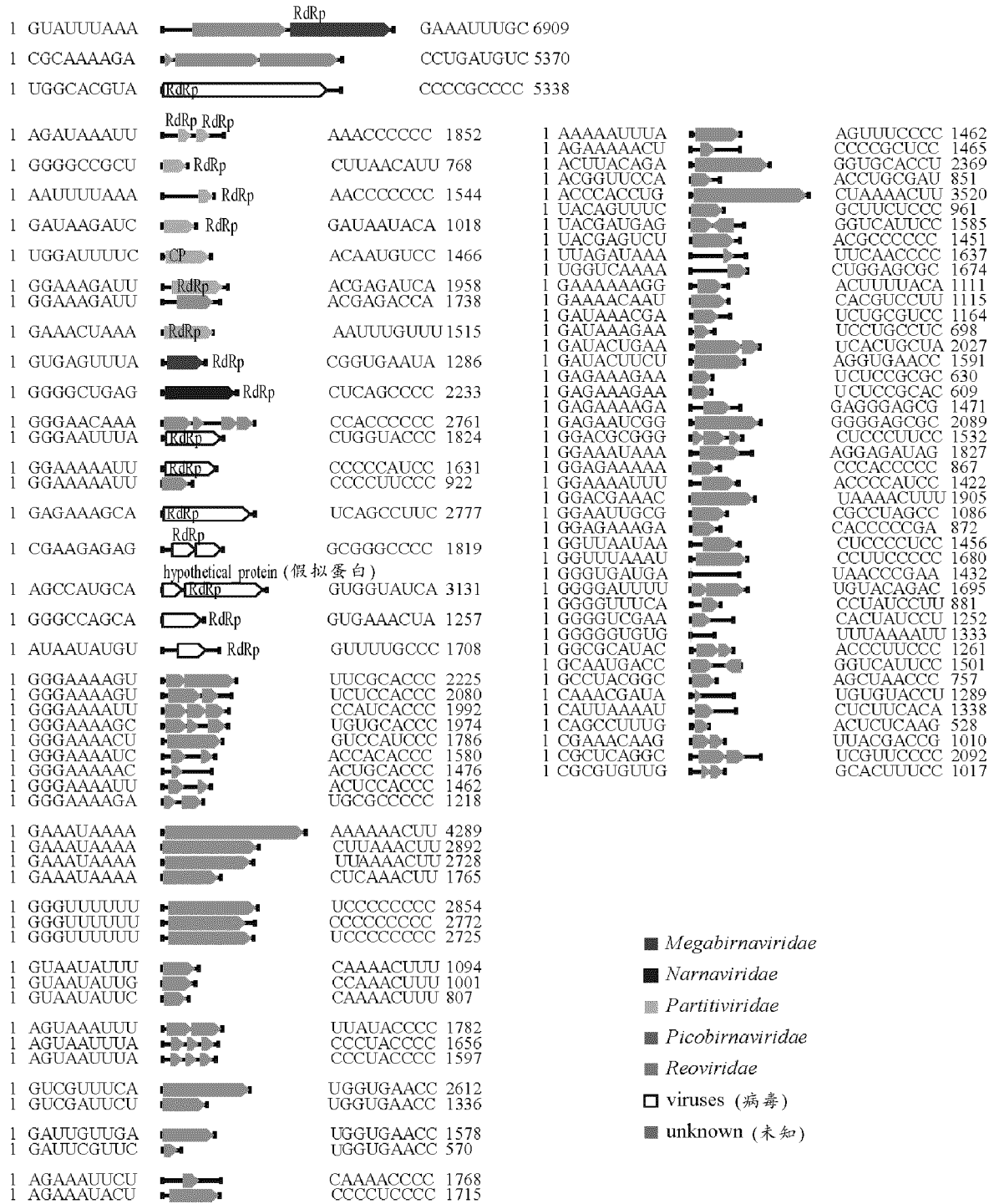


图 5

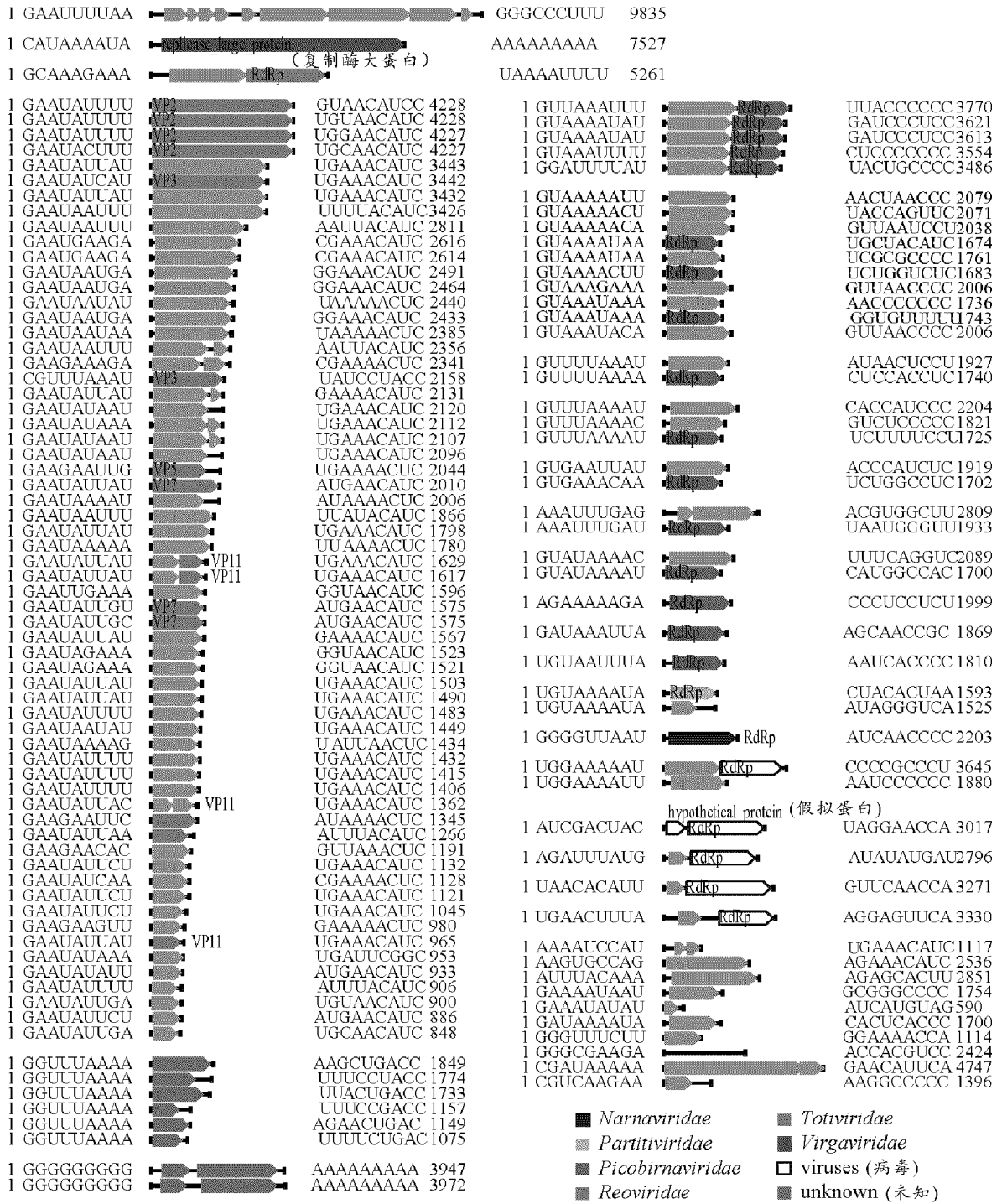


图 6

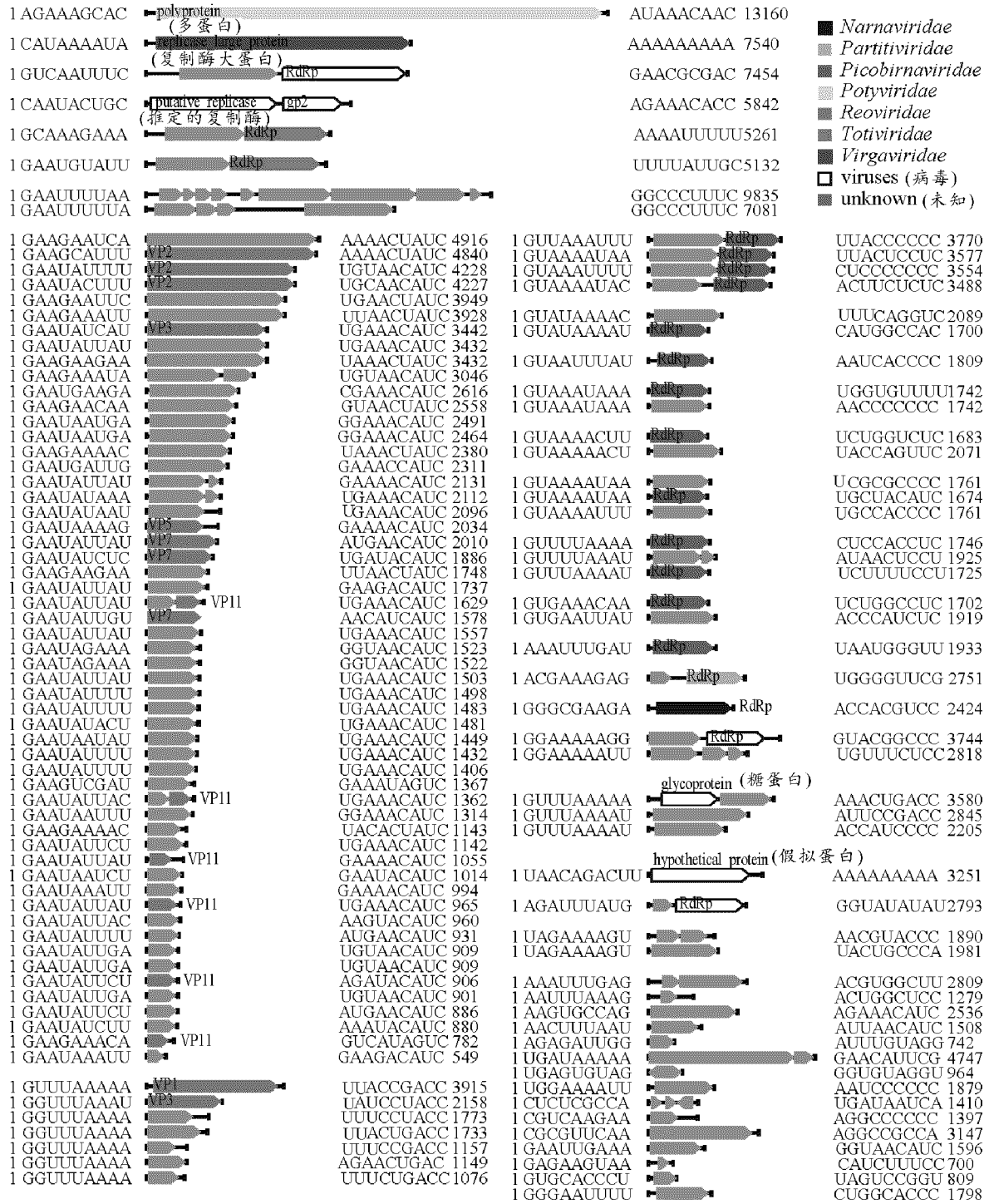


图 7

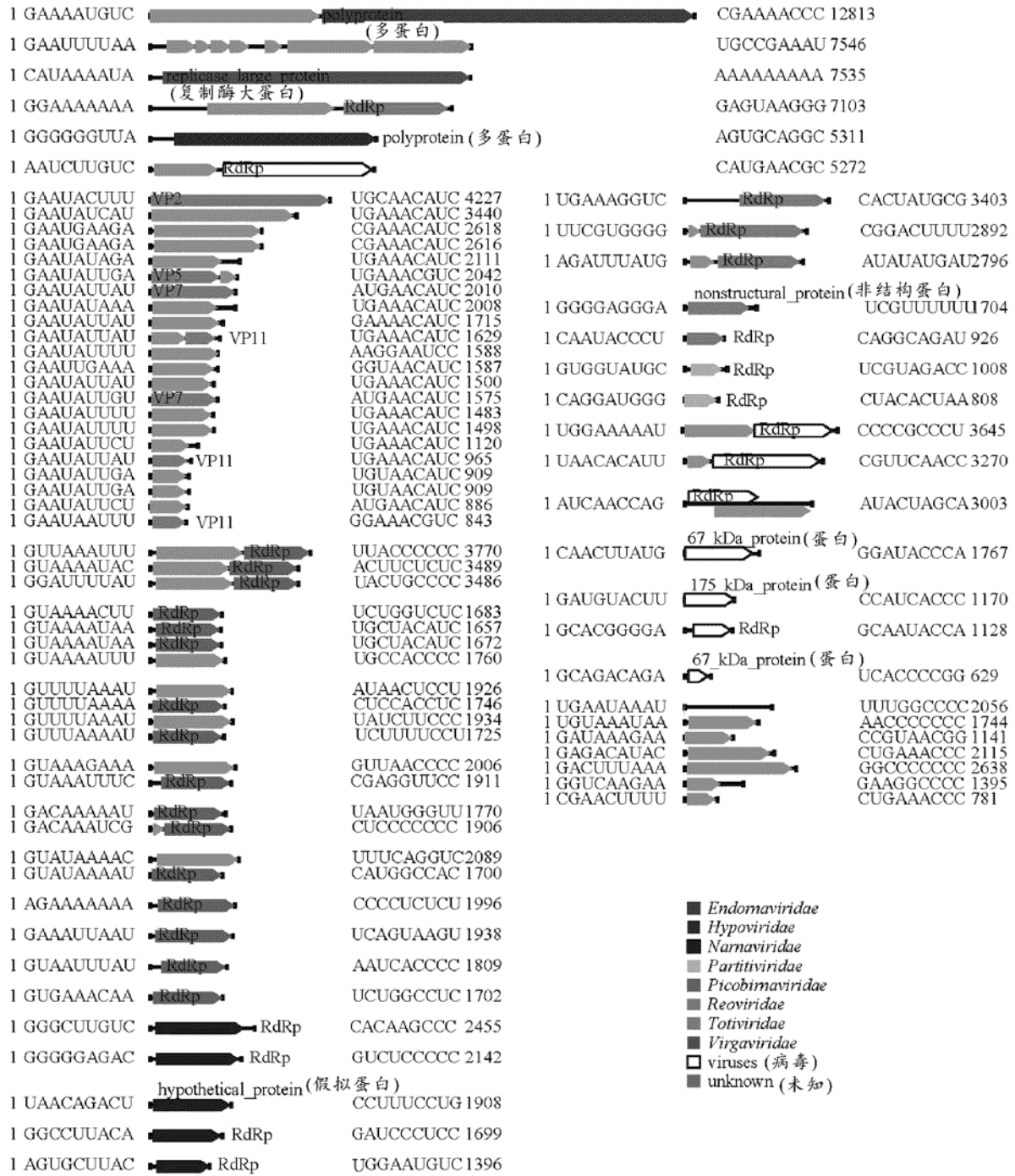


图 8

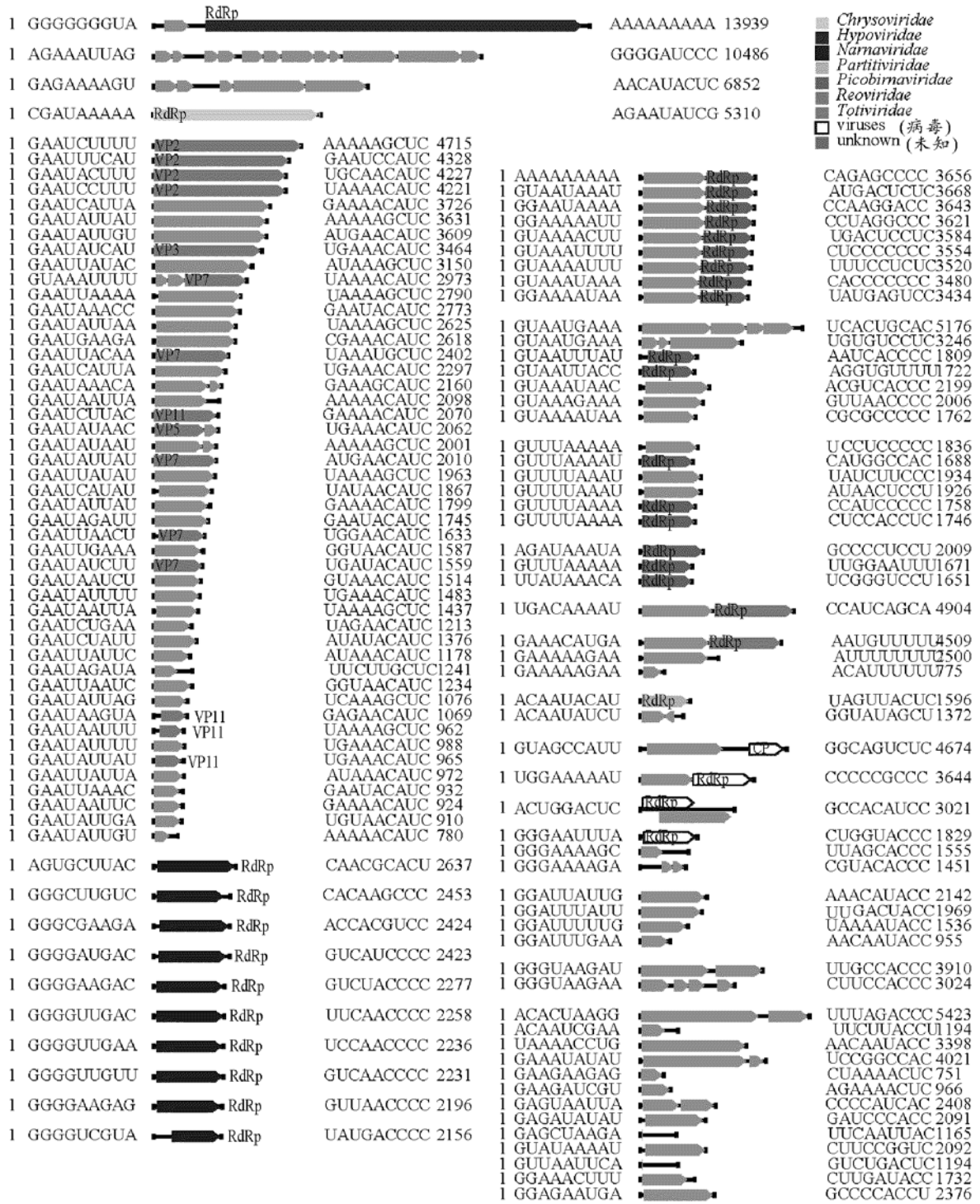


图 9