RNA病毒组与生物信息学分析

**RNA Virome and Bioinformatics Analysis**

张誉译1, 3，陈毅聪1，魏小曼1, 3, 4，崔杰1, 2, \*

1中国科学院上海巴斯德研究所，中国科学院分子病毒与免疫重点实验室，中国科学院生物安全大科学研究中心，上海市，上海市；2青岛海洋科学与技术试点国家实验室，海洋生物与生物技术实验室，青岛市，山东省；3中国科学院大学，北京市，北京市；4中国科学院武汉病毒研究所，中国科学院生物安全大科学研究中心，武汉市，湖北省

\*通讯作者邮箱：[jcui@ips.ac.cn](mailto:jcui@ips.ac.cn)

**摘要：**病毒组学是一门借助Meta-genomics或Meta-transcriptomics的手段，对样品中所携带的所有病毒进行系统性分析的学科。随着高通量测序技术的兴起，以及其在病原生物学领域的广泛应用，人们对于病毒组的认识有了极大的拓展，其研究可以辅助预测病毒性传染病、了解病毒的进化和基因组的多样性、探究病毒在宿主或地域间的跨物种传播现象，是描绘病毒生态圈的理论基础。本实验方法旨在从目标样品中提取核酸并进行文库构建和高通量测序，从测序结果中挖掘新病毒信息，并做病毒组分析。

**关键词：**病毒组，宏转录组，病毒进化，跨物种传播

**研究背景**

致病性病毒的感染，是公众健康的巨大威胁，同时会造成严重的经济损失。然而，传统的病毒学研究具有明显的不足。一方面，从局部部位采集的样品，不足以反映整体的病毒情况，我们仅能从单一的点出发，对有限种类的病毒开展研究。另一方面，因为缺少合适的培养体系或动物模型，病毒学的基础研究困难重重。相比于其他疾病，以上这些不足极大地限制了我们对于病毒的研究，也使得病毒成为了当下研究最为不足的生物实体。

“病毒组（Virome）”是近年来兴起的一个新的组学概念，是指特定的生物个体、生物群体或生态环境中携带的所有病毒的集合，其包括DNA或RNA作为遗传物质的、已知或未知的、致病或非致病的、内源或外源的全部病毒。得益于高通量测序技术（HTS）的蓬勃发展，及其在病原生物学领域的广泛应用，病毒组的研究日益成熟并且一改传统的病毒学研究模式。人们不需要再颇费周章地寻找合适的病毒培养体系或模式动物，直接将感兴趣的样品构建成测序文库（Library），通过宏基因组（Meta-genomics）或宏转录组（Meta-transcriptomics）的测序方法便可以检测样品中的绝大部分病毒，从最宏观的视角剖析最微观的世界(Shi *et al.,* 2018; Zhang *et al.,* 2018; Zhang *et al.,* 2019)。

高通量测序技术，又称“下一代测序技术（NGS）”，是对传统的测序方法的革命性的变革，可以同时对高达数百万条的核酸分子进行序列测定，使得测序通量提高了成千万倍，也使得单一物种的基因组或转录组测序变得简单、快捷和便宜。高通量测序技术引领了多个生命科学领域的技术突破，有着广泛的应用。在病毒学研究当中，基于高通量测序技术的病毒组研究成为了人们认识病毒世界的强大工具。Zhang等人在他们的一篇评论中详细地介绍了利用宏基因组学、宏转录组学手段来拓展病毒圈（Virosphere），相比于其他传统方法（病毒培养、共有引物PCR）具有通量高、视角广等不可比拟的优势(Zhang *et al.,* 2018)。

2017年，ICTV宣布将通过宏基因组或宏转录组的方法发现的病毒加入正式的病毒学分类体系(Simmonds *et al.,* 2017)。高通量技术赋予了病毒学研究新的生机，自此人们发现和认识新病毒的过程进入了“快车道”。

**材料与试剂**

|  |  |  |
| --- | --- | --- |
| **名称** | **公司** | **货号** |
| RNAlater | 赛默飞世尔科技（中国）有限公司 | AM7021 |
| β-巯基乙醇（1000 X） | 赛默飞世尔科技（中国）有限公司 | 21985023 |
| 3 mm研磨珠（氧化锆） | 武汉赛维尔生物科技有限公司 | G0203 |
| 无水乙醇 | 上海盛稀化工有限公司 | 64-17-5 |
| RNase-free ddH2O | 宝日医生物技术（北京）有限公司 | 9012 |
| DL2000 DNA Marker | 宝日医生物技术（北京）有限公司 | 3427 |
| DL15000 DNA Marker | 宝日医生物技术（北京）有限公司 | 3582 |
| 50 X TAE Buffer | 翌圣生物科技（上海）有限公司 | 60116ES76 |
| 10 X Loading Buffer | 南京诺唯赞生物科技股份有限公司 | P022 |
| 高保真DNA聚合酶 | 南京诺唯赞生物科技股份有限公司 | P505 |
| 逆转录试剂盒 | 南京诺唯赞生物科技股份有限公司 | R323 |
| DNA提取试剂盒 | 南京诺唯赞生物科技股份有限公司 | DC102 |
| RNA提取试剂盒 | 南京诺唯赞生物科技股份有限公司 | RC101 |
| RNA文库构建试剂盒 | 南京诺唯赞生物科技股份有限公司 | NR604 |
| 核糖体去除试剂盒 | 南京诺唯赞生物科技股份有限公司 | N406/N407 |
| DNA纯化与分选磁珠 | 南京诺唯赞生物科技股份有限公司 | N411 |
| RNA纯化磁珠 | 南京诺唯赞生物科技股份有限公司 | N412 |
| 双端测序接头试剂盒 | 南京诺唯赞生物科技股份有限公司 | N323/N324 |
| DNA定量试剂盒 | 南京诺唯赞生物科技股份有限公司 | EQ121 |
| RNA定量试剂盒 | 南京诺唯赞生物科技股份有限公司 | EQ211 |
| 高灵敏度DNA试剂盒 | 安捷伦科技（中国）有限公司 | 5067-4626 |
| 高灵敏度DNA试剂 | 安捷伦科技（中国）有限公司 | 5067-4627 |

**仪器设备**

|  |  |
| --- | --- |
| **名称** | **公司** |
| 移液器 | 艾本德（上海）国际贸易有限公司 |
| 低温台式离心机 | 赛默飞世尔科技（中国）有限公司 |
| 超低温冰箱 | 赛默飞世尔科技（中国）有限公司 |
| 微量分光光度计 | 赛默飞世尔科技（中国）有限公司 |
| Qubit荧光计 | 赛默飞世尔科技（中国）有限公司 |
| 超净台 | 北京东联哈尔仪器制造有限公司 |
| 研磨仪 | 宁波新芝生物科技股份有限公司 |
| 琼脂糖凝胶电泳 | 上海天能科技有限公司 |
| 凝胶成像仪 | 上海天能科技有限公司 |
| 2100生物分析仪 | 安捷伦科技（中国）有限公司 |
| PCR仪 | 伯乐生命医学产品（上海）有限公司 |

另需：工作站 (型号：DELL Precision T7920；系统：Ubuntu 20.04；CPU：Intel(R) Xeon(R) Gold 6140 CPU @ 2.30GHz，36核)，用于运行相关命令；普通个人电脑，用于调用个人终端、以及访问网页版软件和下载序列。

**软件和数据库**

|  |  |
| --- | --- |
| **名称** | **版本/网址** |
| Python | v3.8.5 |
| Miniconda | v4.9.2 |
| FastQC | v0.11.9 |
| Trimmomatic | v0.39 |
| Trinity | v2.1.1 |
| DIAMOND | v2.0.4 |
| ORFfinder | v0.4.3 |
| IBS | v1.0.3 |
| MAFFT | v7.453 |
| trimAL | v1.2.rev59 |
| MEGA 7 | v7.0.26 |
| IQ-TREE 2 | v2.1.1 |
| FigTree | v1.4.4 |
| Cytoscape | v3.8.1 |

**实验步骤**

**一、样品采集**

1. 依据课题需要收集合适的动物样品，样品应尽量保证鲜活，或在低温 (4 °C) 环境下妥善保存。样品在进行解剖前可以在适宜的人造环境中 (人造海水、生理盐水等) 平衡24 h，以减少采样环境中的杂质污染对后续实验的影响。对采集的样品进行真实、详细的文字记录，并配有图片、视频等。文字记录内容包括但不局限于物种的名称 (物种的鉴定由生态学专业相关人员进行)、采样数量、采样时间、采样地点、形态描述等。

2. 在干净的操作台上对样品进行解剖，应使用75 %的酒精棉球仔细擦拭操作台面、解剖工具、容器等 (在处理不同物种的样品间隙，也应该重复此步骤)。在冰上对样品进行解剖，剖取脏器等可能富含病毒的组织，转移至新的、装有RNA稳定和储存溶液 (RNAlater™ Stabilization Solution，Invitrogen公司，AM7020) 的容器内 (试剂应至少没过组织块)，并做好样品的标注。同一物种的不同个体，以及同个个体的不同脏器都应该分管储存。每个个体另取少量的肌肉组织于干净的EP管内。所有的样品冻存于-20 °C或-80 °C冰箱保存，按照RNA稳定和储存溶液的官方说明，样品RNA在37 °C条件下可以稳定保存1天，25 °C条件下1周，4 °C条件下1月，-20 °C条件下无期限。

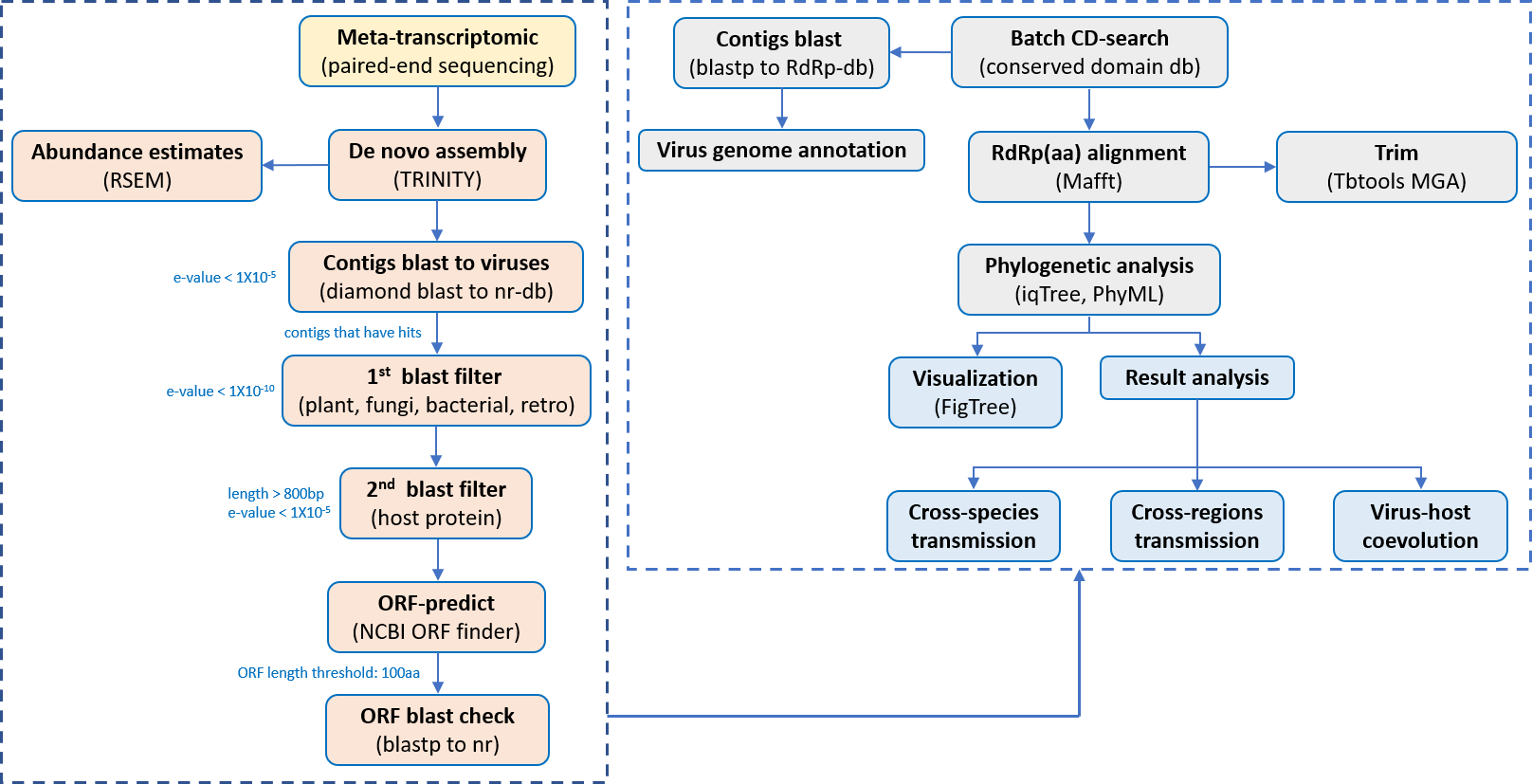
1. **物种鉴定**
2. 所有样品的肌肉组织从冰箱取出于室温解冻，吸水纸上吸干多余水分，切取约20 mg的肌肉组织于新的EP管中。使用DNA提取试剂盒 (FastPure Cell/Tissue DNA Isolation Mini Kit，Vazyme公司，DC102) 提取肌肉组织gDNA：
   1. 对于1-20 mg的组织样品，加入230 ul Buffer GA和20 ul Proteinase K试剂 (若样品质量大于20 mg，按比例扩大试剂加入量) 于55 °C水浴进行消化裂解。为了促进消化，组织块应尽量剪碎，或对组织进行匀浆。
   2. 组织彻底消化后，加入10 ul RNase Solution至消化液中，37 °C水浴60 min。
   3. 室温，12000 x g离心3 min，转移上清液至新的1.5 ml EP管内。
   4. 加入250 ul Buffer GB至消化液中，涡旋混匀20 s，70 °C水浴10 min。
   5. 加入250 ul无水乙醇至消化液中，涡旋混匀20 s。
   6. 将gDNA Columns吸附柱置于2 ml的收集管中，将上一步所得的混合液转移至吸附柱中，室温，12000 x g离心1 min。
   7. 弃滤液，将吸附柱置于收集管中，加入500 ul Washing Buffer A至吸附柱中，轻柔颠倒吸附柱3次，室温，12000 x g离心1 min。
   8. 弃滤液，将吸附柱置于收集管中，加入650 ul Washing Buffer B至吸附柱中，轻柔颠倒吸附柱3次，室温，12000 x g离心1 min。此步骤重复一次。
   9. 弃滤液，将吸附柱置于收集管中，室温，12000 x g离心2 min。
   10. 将吸附柱置于新的1.5 ml离心管，加入30 ul预热至70 °C的Elution Buffer至吸附柱的膜中央，室温放置3 min，室温，12000 x g离心1 min。
   11. 将得到的滤液重新吸取并加入到吸附柱的膜中央，室温放置3 min，室温，12000 x g离心1 min。
3. 测量DNA浓度并评价其质量，保存于-20 °C冰箱中备用。
4. 对样品的细胞色素c氧化酶I基因 (COI基因) 进行PCR扩增和鉴定，使用引物LCO1490：5’-GGTCAACAAATCATAAAGATATTGG-3’和HCO2198：5’-TAAACTTCAGGGTGACCAAAAAATCA-3’ (Folmer *et al.,* 1994)。PCR体系使用高保真酶 (Phanta Max Super-Fidelity DNA Polymerase，Vazyme公司，P505)。25 ul的反应体系中含有2X Phanta Max Buffer 12.5 ul，dNTP Mix (10 mM each) 0.5 ul，上游引物LCO1490和下游引物HCO2198各1 ul，Phanta Max Super-Fidelity DNA Polymerase 0.5 ul，上一步中抽提的gDNA 100 ng，用ddH2O补足体积。PCR反应条件：95 °C预变形3 min，35个循环的反应 (95 °C 15 s，54 °C 15 s，72 °C 1 min)，72 °C彻底延伸5 min。PCR产物取5 ul，使用2 %琼脂糖凝胶电泳进行检测，应在700 bp处有清晰可见的单一条带。对此条带进行回收，送测序公司进行Sanger测序。返回的结果使用NCBI Blastn进行验证，检查物种信息是否正确。

**三、文库构建**

1. 用于建库的组织样品 (浸泡于RNA稳定和储存溶液中)，从冰箱取出，于冰上融化，将组织取出并用吸水纸吸干表面液体。使用RNA提取试剂盒 (FastPure Cell/Tissue Total RNA Isolation Kit，Vazyme公司，RC101) 提取组织总RNA (提取过程注意防范RNase污染)：
   1. 对于10-20 mg的组织样品，加入500 ul Buffer RL1 (已加入β-巯基乙醇或4 M DTT至终浓度为1 %，现配现用) (若样品质量大于20 mg，按比例扩大试剂加入量)，使用电动匀浆器将组织彻底研磨均匀。
   2. 匀浆转移到gDNA-Filter Columns中 (放入收集管内)，室温，13400 x g离心2 min。离心后弃掉gDNA-Filter Columns，保留收集管内上清液。
   3. 向收集管内加入1.6倍体积的Buffer RL2，并轻柔混匀。混合液转移至RNAPure Columns中 (放入收集管内)，室温，13400 x g离心1 min。
   4. 弃滤液，将吸附柱置于收集管中，加入500 ul Buffer RW1至吸附柱中，室温，13400 x g离心1 min。
   5. 弃滤液，将吸附柱置于收集管中，加入700 ul Buffer RW2至吸附柱中，室温，13400 x g离心1 min。
   6. 弃滤液，将吸附柱置于收集管中，向膜中央加入70 ul的DNase I反应工作液(65 ul Buffer RDD和5 ul DNase I)，室温静置30 min。
   7. 将吸附柱置于收集管中，加入500 ul Buffer RW1至吸附柱中，室温，13400 x g离心1 min。
   8. 弃滤液，将吸附柱置于收集管中，加入700 ul Buffer RW2至吸附柱中，室温，13400 x g离心1 min。此步骤重复一次。
   9. 弃滤液，将吸附柱置于收集管中，室温，13400 x g离心2 min。
   10. 将吸附柱置于新的1.5 ml离心管，加入50 ul的RNase-free ddH2O至吸附柱的膜中央，室温放置2 min，室温，13400 x g离心1 min。
   11. 将得到的滤液重新吸取并加入到吸附柱的膜中央，室温放置2 min，室温，13400 x g离心1 min。
   12. 使用微量紫外分光光度计（Nanodrop）检测RNA的浓度和纯度，使用琼脂糖凝胶电泳检测RNA的完整性。
   13. 依据初步检测的浓度，使用RNase-free ddH2O将RNA溶液调整至10 pg/ul - 100 ng/ul之间。使用Qubit试剂（Equalbit RNA HS Assay Kit，Vazyme公司，EQ211）进行精确的浓度测定。Qubit Buffer和Qubit Reagent按照200:1的比例配置成Qubit Working Solution。取199 ul的Qubit Working Solution和1 ul稀释好的RNA样品于干净的0.5 ml薄壁离心管中，另取190 ul的Qubit Working Solution和10 ul的标准品于干净的0.5 ml的薄壁离心管中。所有样品于室温避光孵育2 min，使用Qubit仪器（Qubit 3或4）进行检测。
   14. 下游的建库实验要求较高质量的RNA，应保证RNA的浓度高于10 ng/ul且总量不低于10 ng，OD260/OD280比值介于1.8-2.1之间，OD260/OD230比值介于2-2.5之间，电泳条带可以看到清晰、明亮的两条条带。若RNA质量不合格，应重新提取RNA。将RNA溶液保存于-80 °C冰箱中备用，并尽快开展下游的文库构建实验。
2. 用于测序的总RNA样品，从冰箱取出，于冰上融化。使用RNA建库试剂盒(VAHTS Universal V6 RNA-seq Library Prep Kit for Illumina，Vazyme公司，NR604)、rRNA去除试剂盒 (Ribo-off ® rRNA Depletion Kit (Human/Mouse/Rat)，Vazyme公司，N406)、rRNA去除试剂盒 (Ribo-off ® rRNA Depletion Kit (Bacteria)，Vazyme公司，N407)、高通量测序接头试剂盒 (VAHTS RNA Multiplex Oligos Set1- Set2 for Illumina，Vazyme公司，N323/N324) 进行建库实验 (Shi *et al.,* 2016; Zhang *et al.,* 2018)：
   1. 取1 ug总RNA于微量离心管中，并用Nuclease-free H2O稀释至10 ul，冰上放置备用。
   2. 向管中加入1 ul rRNA Probe (Human/Mouse/Rat)，1 ul rRNA Probe (Bacteria)，3 ul Probe Buffer至总体积15 ul，用移液枪吹打混匀。于PCR仪中进行探针杂交反应，反应条件：105 °C热盖，95 °C 2 min，95-22 °C 0.1 °C/s，22 °C 5 min。注：因为市面在售的rRNA去除试剂盒仅有针对宿主为人/小鼠/大鼠、细菌和植物三种。在病毒组研究的实践过程中，联合使用人/小鼠/大鼠以及细菌来源的rRNA去除试剂盒，可以取得较为理想的清楚效果。然而在用于非哺乳或非细菌宿主类型的样品处理时，rRNA的清除效率可能会有所降低。若想检测rRNA的去除效率，请设置对照组，并在此处使用2 ul的RNase-free ddH2O替代rRNA Probe (Human/Mouse/Rat)和rRNA Probe (Bacteria)。
   3. 向管中加入4 ul RNase H Buffer，1 ul RNase H至总体积20 ul，用移液枪吹打混匀。于PCR仪中进行RNase H消化反应，反应条件：37 °C 30 min，4 °C Hold。
   4. 向管中加入29 ul DNase I Buffer，1 ul DNase I至总体积50 ul，用移液枪吹打混匀。于PCR仪中进行DNase I消化反应，反应条件：37 °C 30 min，4 °C Hold。
   5. 涡旋振荡混匀VAHTS RNA Clean Beads，吸取110 ul至上一步的样品中，用移液枪吹打混匀，冰上静置15 min，使RNA结合到磁珠上。
   6. 将样品置于磁力架上5 min，待溶液澄清后，小心移除上清液，随即加入200 ul 用Nuclease-free H2O配置的80 %乙醇 (现用现配)，漂洗磁珠 (此步骤保持样品在磁力架上)。此步骤重复一次。
   7. 保持样品始终处于磁力架中，在室温下开盖干燥磁珠10 min。
   8. 将样品从磁力架上取出，加入18.5 ul Frag/Prime Buffer，用移液枪吹打混匀，室温静置2 min，在磁力架上静置5 min，待溶液澄清后，小心吸取16 ul上清至新的微量离心管中。于PCR仪中进行片段化反应，反应条件：94 °C 5 min，4 °C Hold。注：若想检测rRNA的去除效率，请在此处用21 ul的RNase-free ddH2O替代Frag/Prime Buffer进行洗脱，并吸取18ul的上清至新的微量离心管中。后续步骤见步骤3。
   9. 将Actinomycin D溶液从5 mg/ml稀释至0.12 mg/ml：向48.8 ul的Nuclease-free H2O中加入1.2 ul Actinomycin D (5 mg/ml) 至总体积50 ul，混匀备用。
   10. 向片段化反应后的样品中加入1 ul Actinomycin D (0.12 mg/ml)，6 ul 1st Strand Buffer 2，1st Strand Enzyme Mix 2至总体积25 ul，用移液枪吹打混匀。于PCR仪中进行第一链cDNA合成反应，反应条件：105 °C热盖，25 °C 10 min，42 °C 15 min，70 °C 15 min，4 °C Hold。
   11. 向上一步的反应液中加入25 ul 2nd Strand Buffer 2 (with dUTP)，15 ul 2nd Strand Enzyme Super Mix至总体积65 ul，用移液枪吹打混匀。于PCR仪中进行第二链cDNA合成反应，反应条件：105 °C热盖，16 °C 30 min，65 °C 15 min，4 °C Hold。
   12. 向上一步的反应液中加入5 ul RNA Adapter，25 ul Rapid Ligation Buffer 3，5 ul Rapid DNA Ligase 2，加入Nuclease-free H2O将体积补至100 ul，用移液枪吹打混匀。于PCR仪中进行接头连接反应，反应条件：105 °C热盖，20 °C 15 min，4 °C Hold。
   13. 涡旋振荡混匀VAHTS DNA Clean Beads (磁珠及所用Buffer都应提前从冰箱取出，静置使其平衡至室温)，吸取45 ul至上一步的样品中，用移液枪吹打混匀，室温静置10 min，使DNA结合到磁珠上。
   14. 将样品置于磁力架上5 min，待溶液澄清后，小心移除上清液，随即加入200 ul 用80 %乙醇 (现用现配)，漂洗磁珠 (此步骤保持样品在磁力架上)。此步骤重复一次。
   15. 保持样品始终处于磁力架中，在室温下开盖干燥磁珠10 min。
   16. 将样品从磁力架上取出，加入102.5 ul Nuclease-free H2O，用移液枪吹打混匀，室温静置2 min，在磁力架上静置5 min，待溶液澄清后，小心吸取100 ul上清至新的微量离心管中。
   17. 涡旋振荡混匀VAHTS DNA Clean Beads (磁珠及所用Buffer都应提前从冰箱取出，静置使其平衡至室温)，吸取70 ul至上一步的样品中，用移液枪吹打混匀，室温静置10 min，使DNA结合到磁珠上。
   18. 将样品置于磁力架上5 min，待溶液澄清后，吸取155 ul上清液于新的微量离心管中，再次加入10 ul VAHTS DNA Clean Beads，用移液枪吹打混匀，室温静置10 min，使DNA结合到磁珠上。
   19. 将样品置于磁力架上5 min，待溶液澄清后，小心移除上清液，随即加入200 ul 用80 %乙醇 (现用现配)，室温孵育30 s漂洗磁珠 (此步骤保持样品在磁力架上)，小心移除上清。漂洗的步骤重复一次。
   20. 保持样品始终处于磁力架中，在室温下开盖干燥磁珠10 min。
   21. 将样品从磁力架上取出，加入21.5 ul Nuclease-free H2O，用移液枪吹打混匀，室温静置2 min，置于磁力架上5 min，待溶液澄清后，小心吸取19 ul上清至新的微量离心管中。
   22. 向上一步的样品中加入2.5 ul i5 PCR Primer，2.5 ul i7 PCR Primer，25 ul VAHTS HiFi Amplification Mix，1 ul Heat-labile UDG，至总体积50 ul，用移液枪吹打混匀。于PCR仪中进行文库扩增反应，反应条件：105 °C热盖，37 °C 10 min，98 °C预变形30 s，13个循环的反应 (98 °C 10 s，60 °C 30 s，72 °C 30 s），72 °C彻底延伸5 min，4 °C Hold。
   23. 涡旋振荡混匀VAHTS DNA Clean Beads (磁珠及所用Buffer都应提前从冰箱取出，静置使其平衡至室温)，吸取45 ul至上一步的样品中，用移液枪吹打混匀，室温静置10 min，使DNA结合到磁珠上。
   24. 将样品置于磁力架上5 min，待溶液澄清后，小心移除上清液，随即加入200 ul 用80 %乙醇 (现用现配)，室温孵育30 s漂洗磁珠 (此步骤保持样品在磁力架上)，小心移除上清。漂洗的步骤重复一次。
   25. 保持样品始终处于磁力架中，在室温下开盖干燥磁珠10 min。
   26. 将样品从磁力架上取出，加入25 ul Nuclease-free H2O，用移液枪吹打混匀，室温静置2 min，置于磁力架上5 min，待溶液澄清后，小心吸取22.5 ul上清至新的离心管中。
3. rRNA残留量检测：
   1. 18 ul的RNA洗脱液（实验组为进行了rRNA去除的，对照组为未进行rRNA去除的）使用反转录试剂盒进行反转录（HiScript III RT SuperMix for qPCR (+gDNA wiper)，Vazyme公司，R323）。向RNA洗脱液中加入6 ul的4 X gDNA wiper Mix，于PCR仪内42°C孵育2min。加入6ul的5 X HiScript III qRT SuperMix，于PCR仪内进行反应，反应条件：37°C 15 min，85°C 5 s。
   2. 使用荧光定量PCR（ChamQ Universal SYBR qPCR Master Mix，Vazyme公司，Q711）进行检测。取2.5 ul的cDNA，加入12.5 ul的2 X ChamQ Universal SYBR qPCR Master Mix，加入2.5 ul的上游引物和2.5 ul的下游引物（引物依据具体的宿主类型进行设计，并针对持家基因进行引物设计作为对照），用ddH2O补足25 ul。使用荧光定量PCR仪进行检测，反应条件：95°C 30 s，40循环的95°C 10s，60°C 30s。
   3. 对PCR结果进行分析，计算rRNA的残留量或清除效率。
4. 文库质检：
   1. 使用微量紫外分光光度计（Nanodrop）检测文库的浓度和纯度检测，使用琼脂糖凝胶电泳检测文库的片段大小分布。
   2. 依据初步检测的浓度，使用RNase-free ddH2O将文库浓度调整至10 pg/ul - 100 ng/ul之间。使用Qubit试剂（Equalbit 1 × dsDNA HS Assay Kit，Vazyme公司，EQ121）进行精确的浓度测定。取199 ul的Qubit Working Solution和1 ul稀释好的文库样品于干净的0.5 ml的薄壁离心管中，另取190 ul的Qubit Working Solution和10 ul的标准品于干净的0.5 ml的薄壁离心管中。所有样品于室温避光孵育2 min，使用Qubit仪器（Qubit 3或4）进行检测。
   3. 取1 ul的文库样品加入5 ul的Marker，加入制备好的芯片中（高灵敏度 DNA 试剂盒，安捷伦公司，5067-4626），使用安捷伦2100生物分析仪进行检测，评估文库的质量。并用Agilent 2100进行文库质量的评价。质检合格的样品，送公司进行高通量测序。
   4. 为保证测序质量，文库浓度应大于10 ng/ul，电泳条带应为350-650 bp间较窄的弥散条带，2100的分析结果应显示基线平整，单一窄峰，无杂峰，无接头或引物二聚体。
5. 高通量测序：样品送测序公司，进行高通量测序。

**四、RNA病毒的鉴定和分析流程图：**

本实验方案使用的RNA病毒的鉴定和分析流程，见下图：

****

**五、数据预处理**

1. 数据获取：从测序公司获取原始数据（文件格式为sample.fastq.gz），对照送样信息检查文件是否缺漏。将文件上传至服务器home目录下data文件夹内，使用gzip命令进行解压。
2. 数据质检：使用FastQC软件对每个文库的测序数据进行质检，使用命令“fastqc -o fastqc -f fastq \*.fastq”。执行此命令前切换至存放fastq文件的路径，并提前创建用于存放质检报告的文件夹fastqc。
3. 数据修剪：使用Trimmomatic软件对数据进行修剪和接头去除，使用命令“trimmomatic PE -threads 4 -phred33 sample\_R1.fastq sample\_R2.fastq sample-1P.fq sample-1U.fq sample-2P.fq sample-2U.fq ILLUMINACLIP:adapter.fa:2:30:10:1:true LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:50”。其中，测序模式选择双端测序模式，输入R1和R2两个测序结果文件，输出1P、1U、2P、2U四个结果文件，使用adapter.fa文件中记录的adapter信息进行接头去除，从reads的头部和末尾分别切除质量值低于3的碱基，从reads的头部开始进行长度为4的滑窗检查并去除平均质量低于20的滑窗内的所有碱基，去除修剪后长度低于50 nt的reads。
4. 从头拼接（如从公司接收的数据为clean data，可以直接进行此步骤）：使用Trinity软件对数据进行从头拼接，使用命令“Trinity --seqType fq --max\_memory 50G --left sample-1P.fq --right sample-2P.fq --CPU 12”。其中数据类型选择fastq格式，输入经过修剪的1P、2P文件，依据工作站性能选择合适的内存量和线程数。需要注意的是，Trinity的输出文件夹必须包含有“trinity”字段。

**六、RNA病毒鉴定与分析**

1. 序列准备：将样品名称添加到各条contig名称的前面，用“\_\_”分隔，并删除长度和位置信息，确保contig名称内没有空格和特殊字符。使用cat命令，将所有样品的contig文件合并成一个文件。
2. 使用DIAMOND软件，对来自于病毒的contigs序列进行富集：
   1. 从NCBI FTP服务器下载数据库序列以及物种对应信息：nr.fasta，prot. accession2taxid.gz，nodes.dmp。
   2. 构建nr数据库：diamond makedb --in nr.fasta --db nr --taxonmap prot. accession2taxid.gz --taxonnodes nodes.dmp。
   3. Blastx：diamond blastx --query [input.fasta] --out [output.txt] --db nr --taxonlist 10239 --ev alue 1E-5。
   4. 提取有hit的contigs序列至新的fasta文件中，进行后续分析。
3. 构建filter\_list，使用DIAMOND软件，对上一步的contigs序列进行过滤：
   1. 从Virus-Host database里，整理出宿主为植物、真菌、细菌、古菌的病毒以及逆转录病毒，收集taxonid，作为filter\_list。
   2. Blastx：diamond blastx --query [input.fasta] --out [output.txt] --db nr --taxonlist [filter\_list] --ev alue 1E-10。
   3. 提取没有hit的contigs序列至新的fasta文件中，进行后续分析。
4. 使用DIAMOND软件，对RNA病毒进行发掘：
   1. 从NCBI下载所有的RdRp的蛋白序列 (RefSeq)，RdRp\_pr.fasta。
   2. 构建RdRp数据库：diamond makedb --in RdRp.fasta --db RdRp
   3. Blastx：diamond blastx --query [input.fasta] --out [output.txt] --db RdRp --ev alue 1E-5。
   4. 提取有hit的contigs序列至新的fasta文件。
5. 使用DIAMOND软件，对上一步的contigs序列进行进一步的过滤：
   1. Blastx：diamond blastx --query [input.fasta] --out [output.txt] --db nr --ev alue 1E-5。
   2. 提取有hit且top hit为病毒蛋白的contigs序列至新的fasta文件，舍弃有hit但top hit不为病毒蛋白的假阳性序列。
6. 使用ORFfinder软件，对上一步的contigs序列进行ORF预测，并将ORF的氨基酸序列保存至新的fasta文件中：ORFfinder -in [input\_nu.fasta] -out [output\_pr.fasta] -s 0 -ml 150 -n true。
7. 使用DIAMOND软件，对上一步的ORF序列进行过滤：
   1. Blastp：diamond blastp --query [input.fasta] --out [output.txt] --db nr --ev alue 1E-5。
   2. 提取有hit且top hit为病毒蛋白的ORF序列至新的fasta文件，舍弃有hit但top hit不为病毒蛋白的假阳性序列。
8. 使用网页版Batch CD-search，对上一步的ORF序列进行保守结构域的预测，将具有RdRp结构域的ORF序列提取至新的fasta文件里。这些包含有RdRp结构域的序列，即RNA病毒序列，进行后续的分析。
9. 从原始的contigs文件中提取RNA病毒对应的核酸序列，针对这些序列分别设计引物。将抽提的总进行反转录，然后使用高保真酶对其进行PCR扩增。琼脂糖凝胶电泳检测条带，并切取对应的条带送公司进行一代测序，用于对高通量测序和从头拼接的验证。对于扩增失败或者测序结果不符合的序列，应予以剔除，不再进行后续的分析。
10. 基因组可视化：依据每条contig的长度、开放阅读框、保守结构域，使用IBS软件的Nucleotide模式进行绘图。作图完成后保存工程文件，输出PDF文件。使用AI软件进行进一步的调整和美化。
11. 使用NCBI Blast+软件，利用ICTV的数据，对RNA病毒进行物种信息的确定：
    1. 从ICTV网站下载最新公平的病毒信息表，下载所有的病毒核酸序列，ICTV\_virus.fasta。
    2. 构建病毒数据库：makeblastdb -in ICTV\_virus.fasta -out ICTV\_virus -dbtype nucl。
    3. tblastn：tblastn -query [input.fasta (RNA病毒的ORF序列)] -out [output.txt] -db ICTV\_virus -evalue 1E-5 -qcov\_hsp\_perc 40。
    4. 提取每条RNA病毒序列的top hit的病毒信息，并在ICTV信息表中查找其物种信息，作为该RNA病毒的物种信息。
12. 使用DIAMOND、TBtools、usearch、Mafft、trimAL、iqtree等软件，对RNA病毒按照病毒种类，进行进化学地位的分析：
    1. Blastp：diamond blastp --query [input.fasta] --out [output.txt] --db nr --ev alue 1E-5。提取每个病毒前50个hit的accession号至ref\_list。
    2. 从ICTV的病毒信息表中提取该种类病毒的所有参考序列的accession号 (核酸序列)，依据其注释信息提取其RdRp区域的蛋白序列，并将其accession号合并存至ref\_list中。
    3. 对于ref\_list当中记录的所有accession号进行去重，使用TBtools软件下载所有的参考序列。
    4. 使用usearch软件对参考序列进行85 %一致性的去重：usearch -cluster\_fast [input.fasta] -id 0.85 -centroids [output.fasta]。
    5. 将该种类新发现的RNA病毒的ORF序列和参考序列合并，使用Mafft软件进行比对：mafft --auto --reorder [input.fasta] > [output.fasta]。
    6. 使用trimAL软件对序列进行修建：trimal -in [input.fasta] -out [output.fasta] -gappyout。
    7. 使用iqtree对修建后的序列进行系统发生树的构建：iqtree -s [input.fasta] -m TEST -alrt 1000。
13. 使用Mafft、Mega、Cytoscape软件，对RNA病毒进行跨物种、跨地域传播的分析：
    1. 序列准备：广泛阅读文献，收集文献中报道的，记录有详细的病毒宿主、样品采集地点信息的RNA病毒序列。
    2. 整理本课题分析中新发现的RNA病毒的核酸序列，对照样品采样表整理病毒宿主、样品采集地点信息。依据病毒宿主和样品采集信息对RNA病毒进行分类。将所有的RNA病毒的核酸序列整合到一个fasta文件内。
    3. 序列比对：将RNA病毒序列和参考序列合并至一个fasta文件内，使用Mafft软件进行比对：mafft --auto --reorder [input.fasta] > [output.fasta]。
    4. 比对后的序列文件用Mega打开并计算遗传距离 (Model/Method使用p-distance，Gaps/Missing Data Treatment使用Pairwise deletion)，将结果导出至Excel文件。依据分组情况，进行组间比较，分别提取两个组内病毒之间的遗传距离，将遗传距离小于0.3的数据输出至新的Excel表格中。
    5. 网络作图：使用Cytoscape软件作图，excel表格中的第一列和第二列作为节点信息，第三列作为节点间的连线信息。使用两个地点间的跨物种传播事件数为每条连线的宽度进行加权；每个地点发现的RNA病毒数目，取10为底的对数，为每个节点的大小进行加权。调整网络图的形状、颜色等，完成后保存工程文件，输出PDF文件。使用AI软件进行进一步的调整和美化。

**致谢**

1. 本实验的经费来源于国家自然科学基金 (31970176)，中国科学院“百人计划”以及中科院分子病毒学与免疫学重点实验室基金（KLMVI-OP-201901），广东省渔业生态环境重点实验室基金（FEEL-2019-6）。
2. 本实验方法改编自Shi等人于2016和2018年发表在Nature上的两篇文章。

**参考文献**

1. Folmer, O., Black, M., Hoeh, W., Lutz, R. and Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol Mar Biol Biotechnol 3(5): 294-299. <https://www.ncbi.nlm.nih.gov/pubmed/7881515>
2. Shi, M., Lin, X. D., Tian, J. H., Chen, L. J., Chen, X., Li, C. X., Qin, X. C., Li, J., Cao, J. P., Eden, J. S., Buchmann, J., Wang, W., Xu, J., Holmes, E. C. and Zhang, Y. Z. (2016). Redefining the invertebrate RNA virosphere. Nature 540(7634): 539-543.
3. Shi, M., Zhang, Y. Z. and Holmes, E. C. (2018). Meta-transcriptomics and the evolutionary biology of RNA viruses. Virus Res 243: 83-90. <https://www.ncbi.nlm.nih.gov/pubmed/29111455>
4. Simmonds, P., Adams, M. J., Benko, M., Breitbart, M., Brister, J. R., Carstens, E. B., Davison, A. J., Delwart, E., Gorbalenya, A. E., Harrach, B., Hull, R., King, A. M., Koonin, E. V., Krupovic, M., Kuhn, J. H., Lefkowitz, E. J., Nibert, M. L., Orton, R., Roossinck, M. J., Sabanadzovic, S., Sullivan, M. B., Suttle, C. A., Tesh, R. B., van der Vlugt, R. A., Varsani, A. and Zerbini, F. M. (2017). Consensus statement: Virus taxonomy in the age of metagenomics. Nat Rev Microbiol 15(3): 161-168. <https://www.ncbi.nlm.nih.gov/pubmed/28134265>
5. Zhang, Y. Z., Chen, Y. M., Wang, W., Qin, X. C. and Holmes, E. C. (2019). Expanding the RNA Virosphere by Unbiased Metagenomics. Annu Rev Virol 6(1): 119-139. <https://www.ncbi.nlm.nih.gov/pubmed/31100994>
6. Zhang, Y. Z., Shi, M. and Holmes, E. C. (2018). Using Metagenomics to Characterize an Expanding Virosphere. Cell 172(6): 1168-1172.