

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

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 ddH ₂ O	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

48

49 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

50 另需：工作站（型号：DELL Precision T7920；系统：Ubuntu 20.04；CPU：Intel(R)
 51 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
 52 调用个人终端、以及访问网页版软件和下载序列。

53

54 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

55

56 实验步骤

57 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。

二、物种鉴定

1. 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:

1.1 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。

1.2 组织彻底消化后，加入 10 ul RNase Solution 至消化液中，37 °C 水浴 60 min。

1.3 室温，12000 x g 离心 3 min，转移上清液至新的 1.5 ml EP 管内。

1.4 加入 250 ul Buffer GB 至消化液中，涡旋混匀 20 s，70 °C 水浴 10 min。

1.5 加入 250 ul 无水乙醇至消化液中，涡旋混匀 20 s。

1.6 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。

1.7 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。

1.8 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。

1.9 弃滤液，将吸附柱置于收集管中，室温，12000 x g 离心 2 min。

1.10 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。

1.11 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。

2. 测量 DNA 浓度并评价其质量，保存于-20 °C 冰箱中备用。

3. 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):

1.1 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。

1.2 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。

1.3 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。

1.4 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。

1.5 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。

1.6 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。

- 125 1.7 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
126 离心 1 min。
- 127 1.8 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，
128 13400 x g 离心 1 min。此步骤重复一次。
- 129 1.9 弃滤液，将吸附柱置于收集管中，室温，13400 x g 离心 2 min。
- 130 1.10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
131 膜中央，室温放置 2 min，室温，13400 x g 离心 1 min。
- 132 1.11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
133 x g 离心 1 min。
- 134 1.12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
135 胶电泳检测 RNA 的完整性。
- 136 1.13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
137 ng/ul 之间。使用 Qubit 试剂（Equalbit RNA HS Assay Kit, Vazyme 公司，EQ211）
138 进行精确的浓度测定。Qubit Buffer 和 Qubit Reagent 按照 200:1 的比例配置成
139 Qubit Working Solution。取 199 ul 的 Qubit Working Solution 和 1 ul 稀释好的 RNA
140 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
141 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，
142 使用 Qubit 仪器（Qubit 3 或 4）进行检测。
- 143 1.14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
144 不低于 10 ng，OD₂₆₀/OD₂₈₀ 比值介于 1.8-2.1 之间，OD₂₆₀/OD₂₃₀ 比值介于 2-
145 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
146 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
147 建实验。
- 148 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
149 Universal V6 RNA-seq Library Prep Kit for Illumina, Vazyme 公司，NR604)、rRNA 去
150 除试剂盒 (Ribo-off® rRNA Depletion Kit (Human/Mouse/Rat), Vazyme 公司，N406)、
151 rRNA 去除试剂盒 (Ribo-off® rRNA Depletion Kit (Bacteria), Vazyme 公司，N407)、
152 高通量测序接头试剂盒 (VAHTS RNA Multiplex Oligos Set1- Set2 for Illumina,
153 Vazyme 公司，N323/N324) 进行建库实验 (Shi *et al.*, 2016; Zhang *et al.*, 2018):

2.1 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。

2.2 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)。

2.3 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。

2.4 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。

2.5 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。

2.6 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。

2.7 保持样品始终处于磁力架中，在室温下开盖干燥磁珠 10 min。

2.8 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。

2.9 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。

2.10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。

2.11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。

2.12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。

2.13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。

2.14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。

2.15保持样品始终处于磁力架中, 在室温下开盖干燥磁珠 10 min。

2.16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。

2.17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。

2.18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。

2.19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。

2.20保持样品始终处于磁力架中，在室温下开盖干燥磁珠 10 min。

2.21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。

2.22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。

2.23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。

2.24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。

2.25保持样品始终处于磁力架中，在室温下开盖干燥磁珠 10 min。

2.26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：

3.1 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。

3.2 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。

3.3 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:

4.1 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。

4.2 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。

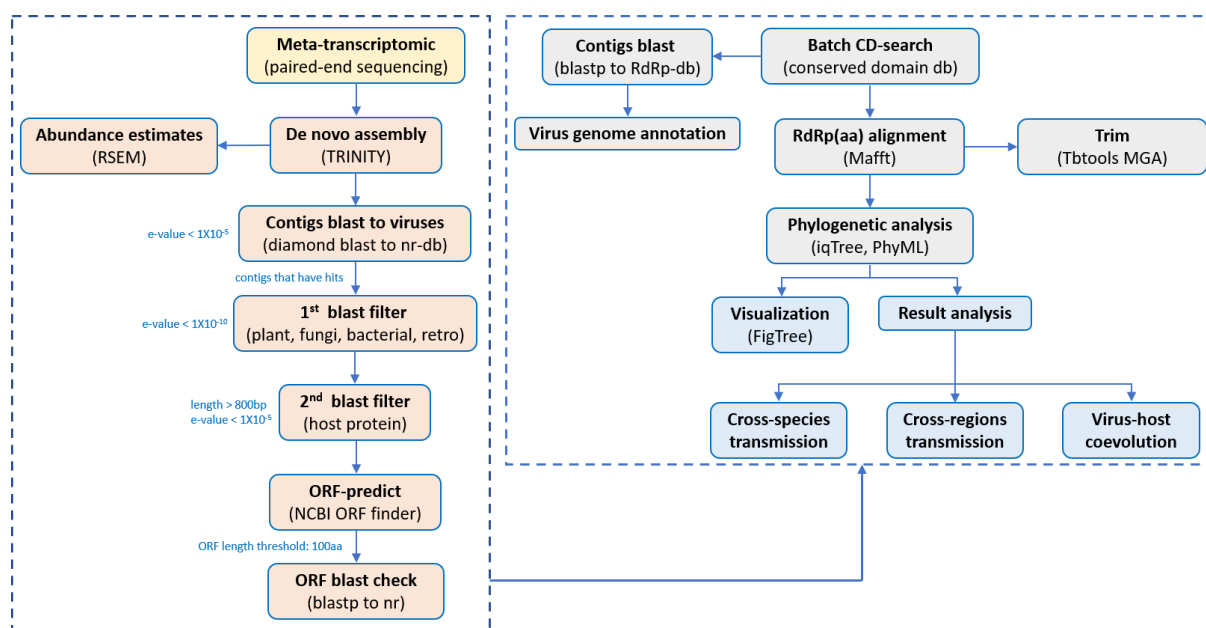
4.3 取 1 ul 的文库样品加入 5 ul 的 Marker, 加入制备好的芯片中 (高灵敏度 DNA 试剂盒, 安捷伦公司, 5067-4626), 使用安捷伦 2100 生物分析仪进行检测, 评估文库的质量。并用 Agilent 2100 进行文库质量的评价。质检合格的样品, 送公司进行高通量测序。

4.4 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：

- 2.1 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。

- 2.2 构建 nr 数据库：diamond makedb --in nr.fasta --db nr --taxonmap prot.accession2taxid.gz --taxonnodes nodes.dmp。

- 2.3 Blastx：diamond blastx --query [input.fasta] --out [output.txt] --db nr --taxonlist 10239 --evaluate 1E-5。

- 2.4 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：

- 3.1 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。

- 3.2 Blastx：diamond blastx --query [input.fasta] --out [output.txt] --db nr --taxonlist [filter_list] --evaluate 1E-10。

- 3.3 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：

- 4.1 从 NCBI 下载所有的 RdRp 的蛋白序列 (RefSeq)，RdRp_pr.fasta。

- 4.2 构建 RdRp 数据库：diamond makedb --in RdRp.fasta --db RdRp

- 4.3 Blastx：diamond blastx --query [input.fasta] --out [output.txt] --db RdRp --evaluate 1E-5。

- 4.4 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：

- 313 5.1 Blastx: `diamond blastx --query [input.fasta] --out [output.txt] --db nr --evaluate 1E-`
 314 5。
- 315 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
 316 top hit 不为病毒蛋白的假阳性序列。
- 317 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
 318 酸序列保存至新的 fasta 文件中：`ORFfinder -in [input_nu.fasta] -out [output_pr.fasta] -`
 319 `s 0 -ml 150 -n true`。
- 320 7. 使用 DIAMOND 软件，对上一步的 ORF 序列进行过滤：
- 321 7.1 Blastp: `diamond blastp --query [input.fasta] --out [output.txt] --db nr --evaluate 1E-`
 322 5。
- 323 7.2 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
 324 top hit 不为病毒蛋白的假阳性序列。
- 325 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
 326 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
 327 序列，即 RNA 病毒序列，进行后续的分析。
- 328 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
 329 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
 330 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
 331 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
 332 续的分析。
- 333 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
 334 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
 335 AI 软件进行进一步的调整 and 美化。
- 336 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：
- 337 11.1 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，
 338 `ICTV_virus.fasta`。
- 339 11.2 构建病毒数据库：`makeblastdb -in ICTV_virus.fasta -out ICTV_virus -dbtype`
 340 `nucl`。

- 341 11.3tblastn: tblastn -query [input.fasta (RNA 病毒的 ORF 序列)] -out [output.txt] -db
342 ICTV_virus -evalue 1E-5 -qcov_hsp_perc 40。
- 343 11.4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
344 种信息，作为该 RNA 病毒的物种信息。
- 345 12. 使用 DIAMOND、TBtools、usearch、Mafft、trimAL、iqtree 等软件，对 RNA 病毒
346 按照病毒种类，进行进化学地位的分析：
- 347 12.1Blastp: diamond blastp --query [input.fasta] --out [output.txt] --db nr --evalue 1E-
348 5。提取每个病毒前 50 个 hit 的 accession 号至 ref_list。
- 349 12.2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
350 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
351 并存至 ref_list 中。
- 352 12.3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
353 有的参考序列。
- 354 12.4使用 usearch 软件对参考序列进行 85 %一致性的去重：usearch -cluster_fast
355 [input.fasta] -id 0.85 -centroids [output.fasta]。
- 356 12.5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
357 行比对：mafft --auto --reorder [input.fasta] > [output.fasta]。
- 358 12.6使用 trimAL 软件对序列进行修建：trimal -in [input.fasta] -out [output.fasta] -
359 gappyout。
- 360 12.7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
361 TEST -alrt 1000。
- 362 13. 使用 Mafft、Mega、Cytoscape 软件，对 RNA 病毒进行跨物种、跨地域传播的分
363 析：
- 364 13.1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
365 采集地点信息的 RNA 病毒序列。
- 366 13.2整理本课题分析中新发现的 RNA 病毒的核酸序列，对样品采样表整理病毒
367 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
368 类。将所有的 RNA 病毒的核酸序列整合到一个 fasta 文件内。

13.3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]。

13.4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。

13.5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 and 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>

-
- 401 4. Simmonds, P., Adams, M. J., Benko, M., Breitbart, M., Brister, J. R., Carstens, E. B.,
402 Davison, A. J., Delwart, E., Gorbalenya, A. E., Harrach, B., Hull, R., King, A. M.,
403 Koonin, E. V., Krupovic, M., Kuhn, J. H., Lefkowitz, E. J., Nibert, M. L., Orton, R.,
404 Roossinck, M. J., Sabanadzovic, S., Sullivan, M. B., Suttle, C. A., Tesh, R. B., van der
405 Vlugt, R. A., Varsani, A. and Zerbini, F. M. (2017). Consensus statement: Virus taxonomy
406 in the age of metagenomics. *Nat Rev Microbiol* 15(3): 161-168.
407 <https://www.ncbi.nlm.nih.gov/pubmed/28134265>
- 408
- 409 5. Zhang, Y. Z., Chen, Y. M., Wang, W., Qin, X. C. and Holmes, E. C. (2019). Expanding
410 the RNA Viroisphere by Unbiased Metagenomics. *Annu Rev Virol* 6(1): 119-139.
411 <https://www.ncbi.nlm.nih.gov/pubmed/31100994>
- 412
- 413 6. Zhang, Y. Z., Shi, M. and Holmes, E. C. (2018). Using Metagenomics to Characterize an
414 Expanding Viroisphere. *Cell* 172(6): 1168-1172.