

DOI: 10.11766/trxb202306130230

刘泽锟, 黄士轩, 杨美艳. 稻田土壤中病毒多样性研究[J]. 土壤学报, 2024, 61(6): 1703–1713.

LIU Zekun, HUANG Shixuan, YANG Meiyuan. Virus Diversity Studies in Paddy Soils[J]. Acta Pedologica Sinica, 2024, 61(6): 1703–1713.

稻田土壤中病毒多样性研究*

刘泽锟^{1, 2}, 黄士轩^{1, 2}, 杨美艳^{1†}

(1. 华南农业大学食品学院, 广东省食品质量安全重点实验室, 广州 510642; 2. 华南农业大学农学院, 广州 510642)

摘要: 病毒是地球上最丰富多样的微生物类群, 广泛存在于各种生物体及环境中, 但关于稻田土壤的病毒组研究较少。使用公共数据库中的稻田土壤宏基因组测序数据, 构建由 8 791 个病毒操作分类单元 (viral operational taxonomic units, vOTU) 和 168 940 个蛋白序列组成的稻田土壤病毒数据库 (Paddy soil virus database, PSVD), 该数据库中 31.41% 的序列能够预测出物种分类地位信息, 这些病毒序列分属于 76 个已知的病毒科, 以有尾病毒纲 (Caudoviricetes) 病毒为主, 宿主分布在 11 个细菌门中。同时还对稻田土壤病毒中所携带的辅助代谢基因 (Auxiliary Metabolic Genes, AMGs) 进行分析, 发现了 39 个可能广泛存在与稻田系统的碳水化合物活性酶 (Carbohydrate active enzymes, CAZymes), 该类基因的存在, 可能对生物地球化学循环产生潜在的影响。综上, 本研究创新性构建了 PSVD 数据库 (<https://github.com/lzk98/PSVD>), 并利用 DRAM-v 挖掘了稻田土壤系统中的 AMGs 基因, 为后期稻田土壤病毒组的研究奠定了研究基础, 并为稻田病毒群落的解析提供新的见解。

关键词: 病毒组学; 土壤; 稻田; 碳水化合物活性酶

中图分类号: S154.4 **文献标志码:** A

Virus Diversity Studies in Paddy Soils

LIU Zekun^{1,2}, HUANG Shixuan^{1,2}, YANG Meiyuan^{1†}

(1. Key Laboratory of Food Quality and Safety of Guangdong Province, College of Food Science, South China Agricultural University, Guangzhou 510642, China; 2. College of Agriculture, South China Agricultural University, Guangzhou 510642, China)

Abstract: **【Objective】** Viruses are the most abundant and diverse group of microorganisms on Earth as they can be found in a wide range of organisms and environments. Despite their ubiquity, there are relatively few studies on the virome in paddy soil systems, which is an area that warrants investigation. **【Method】** In this study, virus sequences were mined from rice soil macrogenome sequencing data in public databases and characterized. **【Result】** A database of paddy soil viruses (PSVD) was constructed, which contained 8 791 viral operational taxonomic units (vOTUs) and 168 940 protein sequences. In PSVD, 31.41% of the sequences could predict the taxonomic status of the species. These viral sequences belonged to 76 known virus families,

* 国家自然科学基金项目 (32202194)、广东省微生物安全与健康重点实验室开放课题 (2020B121201009-PKLMSh003-2021) 和广州市重点研发计划项目 (SL2022B03J01243) 资助 Supported by the National Natural Science Foundation of China (No. 32202194), the Guangdong Provincial Key Laboratory of Microbial Safety and Health Open Topic (No.2020B121201009-PKLMSh003-2021), and the Key R & D Projects of Guangzhou (No. SL2022B03J01243)

† 通讯作者 Corresponding author, E-mail: ymy@scau.edu.cn

作者简介: 刘泽锟 (1998—), 男, 广东饶平人, 硕士研究生, 主要从事病毒组学研究。E-mail: lzk_98@foxmail.com

收稿日期: 2023-06-13; 收到修改稿日期: 2023-09-09; 网络首发日期 (www.cnki.net): 2024-07-19

with the majority of viruses coming from the *Caudoviricetes*, and their hosts were mainly distributed across 11 bacterial phyla. Also, Auxiliary Metabolic Genes (AMGs) carried by rice field soil viruses were analyzed, and 39 carbohydrate-active enzymes (CAZymes) genes, which may have been widely distributed in rice soil systems, were identified, and the presence of these genes contributed to biogeochemical cycling. 【Conclusion】 In summary, this study innovatively constructed a PSVD database (<https://github.com/lzk98/PSVD>) and used DRAM-v to mine AMG genes in paddy soil systems, thus, laying a research foundation for later studies of paddy soil virus groups and providing new insights for the resolution of paddy virus communities.

Key words: Viromic; Soil; Rice field; Carbohydrate-active enzymes

土壤微生态中的微生物群落（细菌、真菌、古菌和病毒）在生物地球化学循环，农业可持续化和植物生产等方面发挥着重要作用。以往研究多集中在不同生态系统中细菌和真菌的群落组成和多样性上^[1]，而土壤中病毒对生态系统影响的研究较少，尤其是对深层陆地生物圈的病毒知之甚少。土壤中含有极高丰度的病毒，预计其数量约为 4.88×10^{30} 个 VLPs，超过海洋中病毒数量（ 1.29×10^{30} 个 VLPs）^[2]，土壤病毒组可能是地球上最大的遗传多样性库，并且在土壤生态系统中具有不可替代的作用^[3-4]。土壤病毒中，绝大部分由噬菌体组成，噬菌体与细菌之间的互作关系，是土壤病毒研究的重点^[5]。

土壤中的病毒，尤其是噬菌体可以携带辅助代谢基因（Auxiliary Metabolic Genes, AMGs），进而影响宿主生物被膜的形成^[6-7]或参与宿主碳、硫、磷代谢^[4, 8-10]。其中碳循环相关基因^[11]是目前的研究热点，韩丽丽课题组^[12]在土壤病毒组中挖掘出 137 个碳水化合物活性酶（Carbohydrate active enzymes, CAZymes）基因，发现病毒能够通过携带 CAZymes 来影响生态系统的碳利用。目前 AMGs 已在多个土壤生态系统中被鉴定出^[13]，并认为在生物地球化学循环中起到重要的作用，但大多数 AMGs 在塑造土壤细菌的生态和进化轨迹方面的作用仍不清楚。

目前多数土壤病毒生态学的研究侧重于在自然生态系统中进行，如沙漠和永冻土等，而对农田等受到人工干预的土壤生态系统研究较少^[14]。农业生态系统不仅具有经济效益，为人类提供了大量食物，而且还通过无机氮固定和甲烷合成等过程影响全球生态系统。其中，稻田系统是最重要的农业系统，但对稻田系统中病毒群及病毒-细菌互作的相关研究仍然较少^[15-16]。与水生和其他陆地生态系统不同，稻田生态系统处于一般干湿循环的过程中^[17]。在这个干湿循环的过程中，水稻土具有特殊的氧化还原

过程，影响养分循环利用以及微生物介导的土壤有机质（Soil organic matter, SOM）周转^[18]，同时也会影响土壤团聚体之间的连通性，提高土壤中微生物群的接触机会。最近的研究表明，土壤含水量和 SOM 的变化会影响病毒-细菌的相互作用，进而影响土壤中病毒和细菌群落的组成^[19-20]。

鉴于稻田系统的复杂性和重要性，亟需开发出用于该系统中病毒-细菌互作的研究方法和相关数据库。本研究从公共数据库中收集稻田土壤相关宏基因组数据，挖掘并注释其中的病毒相关序列和基因，构建稻田土壤病毒数据库，为后续水稻-微生物互作等研究奠定基础。

1 材料与方法

1.1 稻田土壤宏基因组数据下载及整理

在 PubMed 上使用关键词“rice soils metagenomic”检索近 10 年内的文献，并且在欧洲核苷酸档案数据库（European Nucleotide Archive, ENA）中以分类号为 1672641 rice paddy metagenome、939928 rhizosphere metagenome 和 410658 soil metagenome 下载所有项目的信息表，通过逐一手动筛选，获得 53 个水稻宏基因组项目，并收集项目内所有样品的 SRA 号（合计 344 个），之后使用 prefetch 进行下载，使用 fasterq-dump 进行文件解析，获得 fastq 文件。最终经过数据质控和病毒序列筛选后，仅 51 个项目、总计 335 个样本符合要求。

1.2 病毒序列识别

使用 MEGAHIT^[21]进行序列组装，参数设置为 --presets meta-large --k-min 27。对获得的样品 contig 序列采用两种病毒序列识别软件进行同步筛选，分别为 DeepVirFinder^[22]（参数为 p 值小于 0.05）和 VirSorter2^[23]（参数为 --include-groups

dsDNAPhage, RNA, ssDNA--min-length 1000--min-score 0.5)。之后汇总两个软件的病毒序列预测结果,使用 CheckV^[24]对预测结果进行剪切去除细菌污染。之后参考 Sullivan Lab 的病毒筛选条件 (doi: dx.doi.org/10.17504/protocols.io.bwm5pc86) 并进行调整,具体筛选条件如下:

保留条件 1: viral_genes 大于 0 and (host_r 小于 0.6 or max_score 大于 0.8)

保留条件 2: viral_genes 等于 0 and (host_genes 等于 0 or max_score 大于或等于 0.95 or hallmark 大于 2)

丢弃条件 1: viral_genes 等于 0 and host_genes 大于 1

丢弃条件 2: viral_genes 等于 0 and host_genes 等于 1 and contig_length 小于 10 000

丢弃条件 3: viral_genes 大于 0 and (host_r 大于 0.6 or max_score 小于 0.8)

其中, viral_genes, host_genes 和 host_r 由 CheckV 提供; max_score 为 DeepVirFinder 或 VirSorter2 的病毒性得分; hallmark 为 VirSorter2 判定序列所含有病毒标记的基因数量。

保留所有满足保留条件的序列和不满足丢弃条件的序列。最后对序列集合进行去重,使用 CheckV 自带的 aniclust.py 脚本进行去 (参数为--min_ani 95--min_tcov 85), 获得非冗余病毒数据集 (即, vOTUs)。

使用 Prodigal 从非冗余病毒数据集中注释开放阅读框 (Open Reading Frames, ORFs), 参数为: -p mata -g 11 -c, 并对 ORFs 基于 eggNOG-mapper 进行基因注释, 具体参数为: --go_evidence non-electronic--cpu 20 --override。最后使用 R 进行可视化。

病毒序列在 eggNOG 注释结果中存在溶原相关基因, 如: 整合酶 (integrase)、转座酶 (transposase)、重组酶 (recombinase)、细胞溶素原 (lysogen) 和切除酶 (excisionase), 则被认定为溶原病毒。

数据库存储在 <https://github.com/lzk98/PSVD> 上。

1.3 病毒序列分类和宿主预测

病毒聚类分析使用 vConTACT2^[25]对非冗余病毒序列和 GenBank 上所有病毒序列 (2023 年 4 月) 进行聚类, 获得聚类病毒簇 (viral cluster, VC) 信息。病毒分类使用 PhaGCN2^[26]进行, 使用其默认

数据库 (2022 年 8 月)。病毒序列比对使用 BLASTn 进行, 对比 NCBI 的核酸数据库 (Nucleotide Sequence Database, NT) 中的病毒子数据库 (txid: 10239, 下载时间为 2022 年 11 月)。末端酶大亚基 (large terminase subunit, TerL) 系统发育树的构建分别基于 Terminase_3 (PF04466)、GpA_ATPase (PF05876)、Terminase_6N (PF03237) 和 TerL_ATPase (PF03354) 四个结构域进行构建。简而言之, 经过 MAFFT 多序列比对^[27], trimAl 序列对齐剪切^[28], FastTree 构建系统发育树^[29], 最后使用 TVBOT^[30]进行美化。

宿主信息使用 RefSeq-spacers 数据库与病毒序列进行匹配预测, 其中 RefSeq-spacers 数据库构建参考前人的构建方法^[31], 简而言之, 下载 RefSeq 中细菌和古菌的基因组数据集合 (2022 年 6 月), 使用 CRT v1.2 和 PILER v1.06 识别 CRISPR 阵列, 再汇总二者的结果。使用 BLASTn 进行比对, 参数为 blastn-short -word_size 18 -dust no-evalue 1e-5, 过滤 “bit score 大于 45” 的比对结果^[31-32]。

1.4 辅助代谢基因的鉴定

辅助代谢基因注释主要基于 DRAM^[33]软件进行。简而言之, 首先通过 VirSorter2, 参数为--prep-for-dramv, 运行所有已识别的 vOTU 序列以生成 affi-contigs.tab 文件, 然后使用默认数据库进行注释。最后根据 DRAM-v 注释提取出的结果, 选择具有辅助评分小于 4 和基因描述的推定 AMGs。遵循 ter Horst 等^[34]报告的方法, 保守起见, 对注释输出进行了过滤, 以提高 AMG 鉴定的可信度, 主要是去除与核苷酸代谢、有机氮、糖基转移酶和核糖体蛋白相关的基因类别。此外, 如果推定的 AMG 位于两个病毒标志基因或病毒样基因之间或附近, 则该基因在后续分析中被认为是高置信度的病毒 AMG^[35]。每个编码 AMG 的重叠群的基因组图使用 chipLOT (<https://www.chipLOT.online/>) 基于 DRAM-v 的注释结果完成可视化。

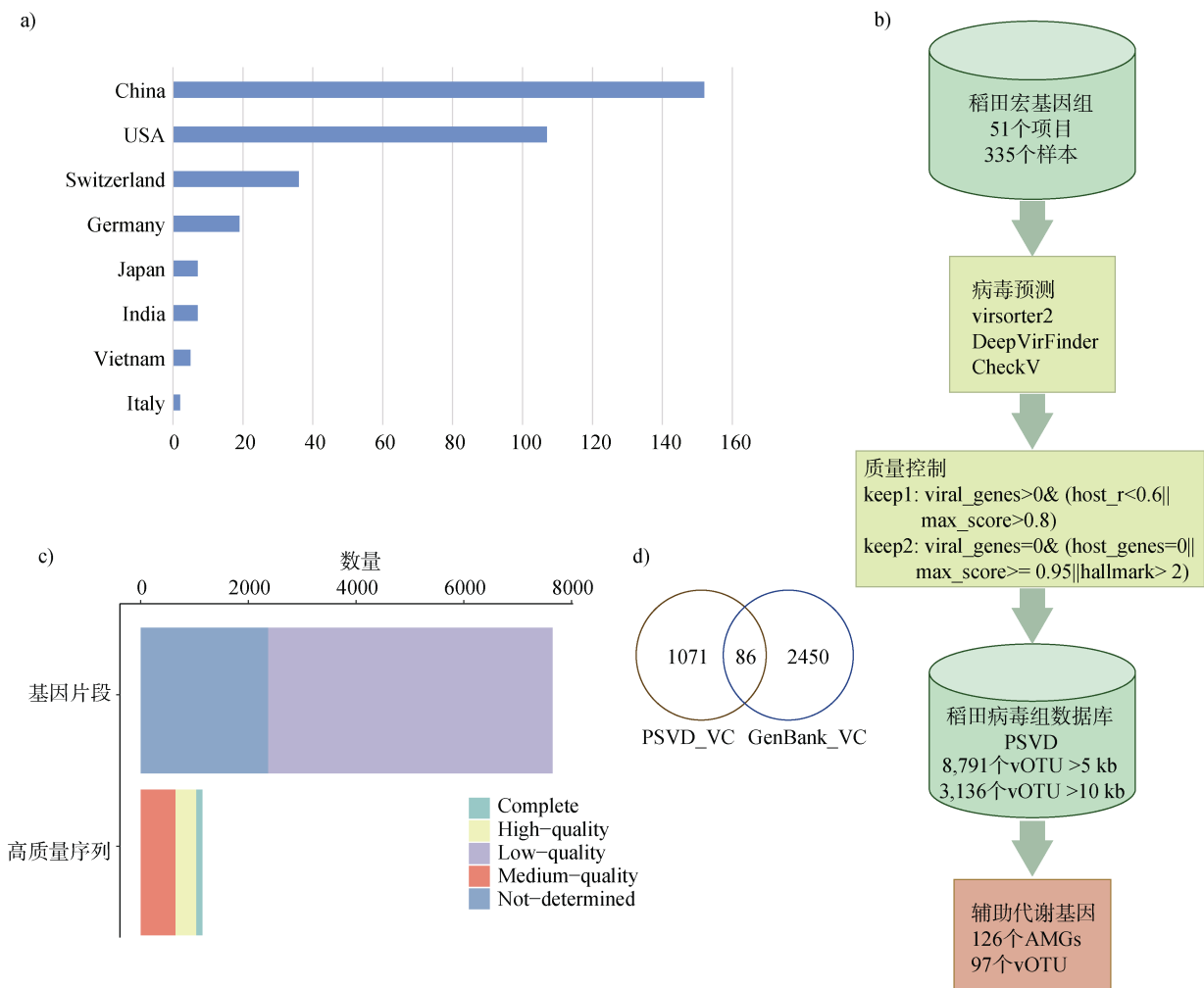
AMG 序列的进化溯源使用 BLASTp 检索 NR 数据库, 使用 MAFFT 进行多序列比对^[27], FastTree 构建系统发育树^[29], 最后使用 TVBOT^[30]进行美化, 通过系统发育树了解该基因的进化情况。AMG 三维结构预测使用 Phyre2 在线网站进行。

2 结果与讨论

2.1 稻田土壤病毒数据库构建

收集到的 51 个稻田宏基因组相关项目，合计 335 个样本，其中来自中国的测序数据最多，占 45.37%，其次为美国，占 31.94% (图 1a)。按图 1b 所示的病毒组分析流程进行稻田土壤病毒数据库的构建与分析。具体而言，通过 VirSorter2、DeepVirFinder 和 CheckV 进行病毒序列挖掘，获得 54 万条病毒序列，再通过自定义的筛选标准过滤后，获得约 37 万条病毒序列，去冗余处理，余下约 29 万条非冗余的病毒序列，其中序列长度大于等于 5 kb 和 10 kb 的数量分别为 8 791 和 3 136 条。

在大于 5 kb 序列的基础上构建稻田土壤病毒数据库 (Paddy Soil Viruses Database, PSVD)，该数据库由 8 791 个病毒 vOTU 构成 (合计 138 Mbp)，并且由 168 940 个 ORFs 组成。vOTU 序列长度主要分布在 5 kb~50 kb 之间 (96.05%)。进一步统计稻田土壤病毒数据库中病毒序列完整性情况，被认为具有中等质量完整、高等质量完整和完全完整的分别有 642、390 和 113 条序列的 (图 1c)，合计占比 13.02% (1 145/8 791)，在这些具有较高序列完整性 (中等质量以上) 的序列中，多数长度依旧分布在 5 kb~50 kb 之间 (80.70%)，其中有 43 条 vOTU 可能为巨噬菌体 (序列长度大于 200 kb)。通过 vConTACT2 对病毒群聚类发现，稻田病毒中存在大



a) 样本来源的地理分布; b) 数据库构建流程; c) 序列完整性分布情况; d) 稻田土壤病毒数据库与已知病毒的 vConTACT2 聚类情况。a) geographical distribution of sample sources; b) database construction process; c) sequence integrity distribution; d) vConTACT2 clustering of the PSVD with known viruses.

图 1 稻田土壤病毒数据库构建

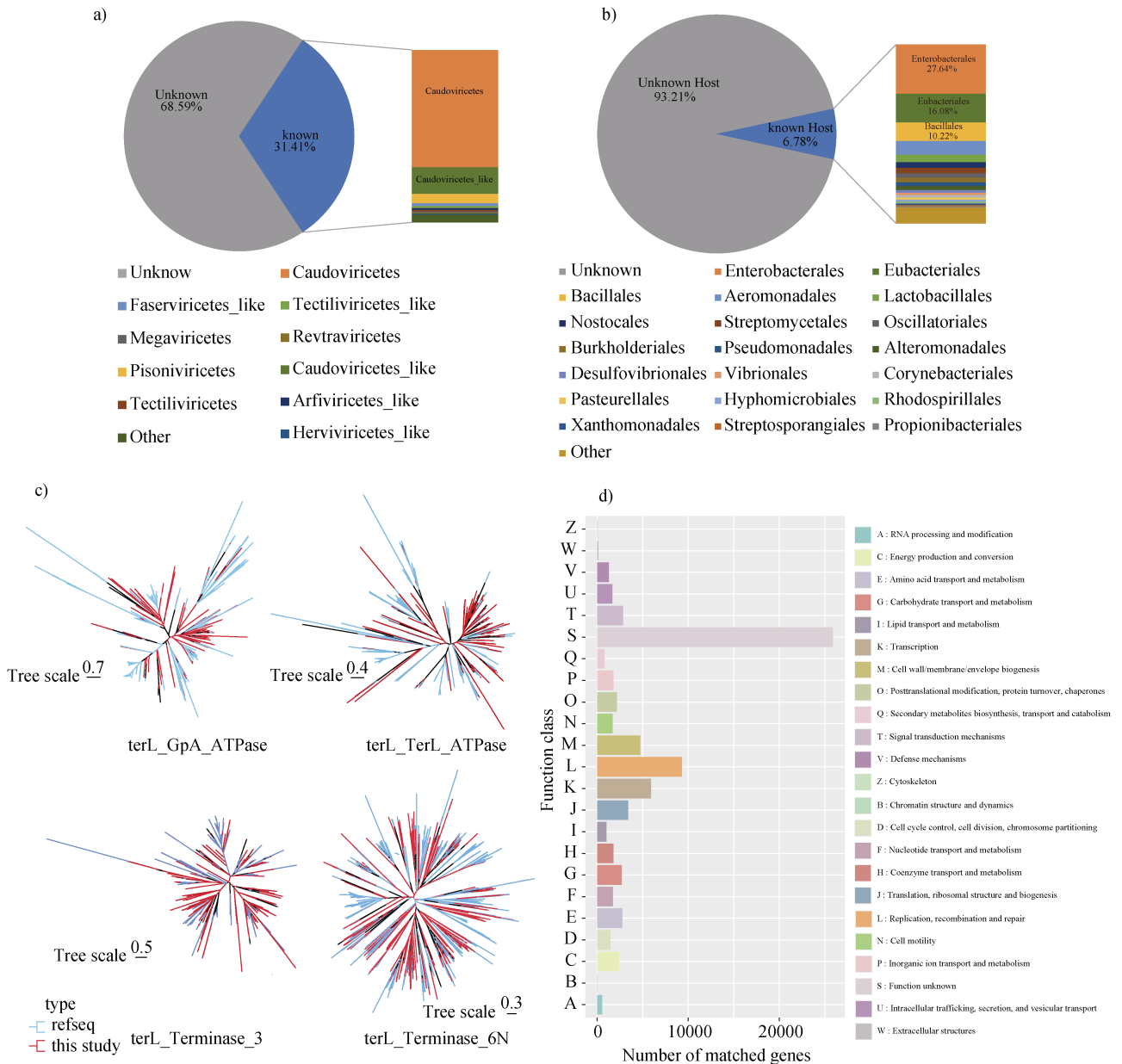
Fig. 1 Construction of the Paddy Soil Viruses Database

量与已知病毒不同的 VC 簇 (图 1d), 仅 86 个 VC 簇与已知的病毒相同, 92.57% (1 071/1 157) 的稻田土壤病毒 VC 簇是独有的。因此, PSVB 很好弥补了当前土壤病毒库的空缺。

2.2 稻田病毒的未知性和多样性

在对稻田土壤病毒数据库中的序列进行分类地位预测的过程中, 发现只有部分病毒序列 (31.41%) 能够预测出物种分类地位信息, 包括了 76 个已知的

病毒科, 分属于 20 个病毒纲, 其中纲水平分类主要为有尾病毒纲 (Caudoviricetes) 占全部序列的 83.45% (2 304/2 761), 而科水平分类又以 Peduoviridae 科和螺旋病毒科 (Herelleviridae) 为主 (图 2a)。Caudoviricetes 基本由噬菌体组成, 说明噬菌体是稻田土壤病毒圈中重要的成员。同时, 还预测出 56 个与已知病毒群具有一定关系的新病毒类群, 包括 615 个病毒 vOTU, 该类群使用 _like 后缀标识 (图 2a)。接着,



a) 分类预测; b) 宿主预测; c) 末端酶结构域系统发育树; d) 基因注释。a) classification prediction; b) host prediction; c) terminase phylogenetic tree; d) gene annotation.

图 2 PSVD 中病毒序列

Fig. 2 Virus sequences in PSVD

将序列与 NCBI 的 NT 数据库进行比对, 只有部分序列 (255/8 791) 能够比对上, 并且所比对的序列多数是未经验证的宏基因组组装病毒序列。最后, 基于 TerL 的系统发育树 (图 2c) 和大部分预测的病毒 ORFs 未能得到注释的结果 (图 2d), 进一步说明稻田土壤病毒群在很大程度上是未知的。

进一步研究稻田土壤病毒群是否与其他环境病毒一样, 可能通过感染微生物来驱动生物地球化学循环, 进而影响生态系统。通过自主构建的 RefSeq-spacers 数据库进行宿主预测分析表明, 8 791 个 vOTU 中的 597 个病毒序列预测出假定的宿主 (图 2b)。预测出的宿主主要分布在 11 个细菌门, 其中变形菌门 (Proteobacteria) 的肠杆菌目 (Enterobacterales) 和厚壁菌门 (Bacillota) 的真杆菌目 (Eubacteriales) 与芽孢杆菌目 (Bacillales) 的数量最多, 该结论与稻田土壤系统中细菌物种组成相类似^[36-37]。同时说明, 稻田土壤中存在许多未被表征的病毒, 其基因组和侵染特性均有待进一步挖掘。如此高比例的未知病毒序列和基因, 进一步阐明了稻田土壤病毒组的多样性和未知性。

2.3 稻田病毒组中的辅助代谢基因

土壤中的许多病毒编码辅助代谢基因 (AMGs), 参与到宿主代谢过程中, 从而影响生物地球化学循环并促进宿主对环境的适应。本研究评估了稻田土

壤病毒 AMGs 对生物地球化学循环的影响, 以加深对稻田土壤病毒的生态位特化和生态功能的了解。

在手动筛选前, DRAM-v 预测了 561 个假定的 AMGs, 随后参考 Pratama 等^[35]的方法和标准, 手动筛选了 DRAM-v 软件的注释结果, 共获得 126 个 AMGs, 分布在 97 个 vOTU 序列上, 参与多种代谢途径, 如碳代谢、氮代谢、光合作用和生物素合成等。其中碳源利用模块数量最多, 达到 73.02% (92/126)。根据预测的 AMGs 周围是否具有病毒基因再进行筛选, 共获得 72 个高置信度的 AMGs, 仅占全部 AMGs 的半数 (57.14%, 72/126), 其代谢模块的数量分布情况与前者相似 (图 3a)。高置信度的 AMGs 分布在 57 个 vOTU 上, 其中有尾噬菌体纲的病毒占据多数, 其余为不具有尾部结构的 Tectiviricetes 纲 dsDNA 噬菌体和具有较大基因组的 Megaviricetes 纲巨型病毒 (图 3b)。对病毒的宿主进行分析, 发现其属于 *Shewanella*、*Methylomonas*、*Klebsiella*、*Geobacillus*、*Enterobacter*、*Clostridium*、*Citrobacter* 和 *Bacillus* 属。在所有高置信度的 AMGs 中, CAZymes 模块数量最多, 为 39 个 (表 1), 其中又以糖苷水解酶 (Glycoside Hydrolases, GH) 和碳水化合物结合模块 (Carbohydrate-Binding Modules, CBM) 模块的数量最多, 分别为 19 和 12 个, 合计占比 79.45%。

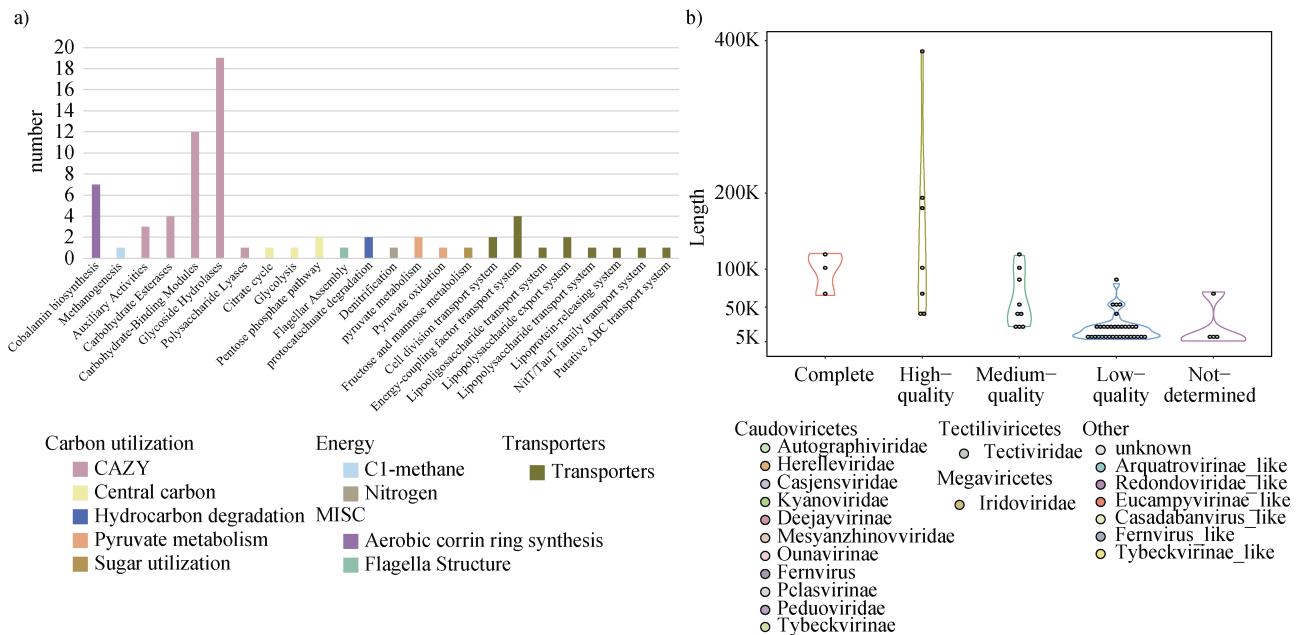


图 3 稻田土壤病毒中辅助代谢基因的分布情况及其序列表征

Fig. 3 Distribution of auxiliary metabolic genes and their sequence characterization in viruses of paddy soils

表 1 碳水化合物活性酶基因统计表

Table 1 Statistics of carbohydrate-active enzyme genes

碳水化合物活性酶类型	基因	描述	酶编号	数量
CAZymes type	Genes	Description	Accession	Number
辅助活性类	AA4	香草醇氧化酶	EC 1.1.3.38	1
Auxiliary activities families	AA7	低聚糖氧化酶	EC 1.1.3.-	2
碳水化合物酯酶	CE1	乙酰木聚糖酯酶	EC 3.1.1.72	1
Carbohydrate esterases	CE12	果胶乙酰酯酶	EC 3.1.1.-	1
	CE4	乙酰木聚糖酯酶	EC 3.1.1.72	2
碳水化合物结合模块	CBM16	能与纤维素和葡甘露聚糖结合	—	2
Carbohydrate-binding modules	CBM32	能与半乳糖和乳糖的结合	—	8
	CBM44	能与纤维素和木葡聚糖的结合	—	1
	CBM56	能与 β -1, 3-葡聚糖结合	—	1
糖苷水解酶	GH113	β -甘露聚糖酶	EC 3.2.1.78	1
Glycoside hydrolases	GH13	α -淀粉酶	EC 3.2.1.1	1
	GH130	β -1, 4-甘露糖基葡萄糖磷酸化酶	EC 2.4.1.281	2
	GH136	低聚糖乳糖-N-生物磺酰水解酶	EC 3.2.1.140	1
	GH145	鼠李糖水解酶	EC 3.2.1.-	1
	GH2	β -半乳糖苷酶	EC 3.2.1.23	1
	GH26	β -甘露聚糖酶	EC 3.2.1.78	1
	GH29	α -L-岩藻糖苷酶	EC 3.2.1.51	1
	GH3	β -葡萄糖苷酶	EC 3.2.1.21	1
	GH33	唾液酸酶或神经氨酸酶	EC 3.2.1.18	1
	GH39	alpha-L-iduronidase	EC 3.2.1.76	1
	GH43	β -木糖苷酶	EC 3.2.1.37	1
	GH5	纤维素酶	EC 3.2.1.4	1
	GH74	内切葡聚糖酶	EC 3.2.1.4	3
	GH87	mycodextranase	EC 3.2.1.61	1
	GH94	纤维二糖磷酸化酶	EC 2.4.1.20	1
多糖裂解酶	PL22	寡半乳糖醛酸裂解酶	EC 4.2.2.6	1
Polysaccharide lyases				

病毒 AMGs 还可能参与了核心碳循环, 如糖酵解、丙酮酸代谢、戊糖途径和三羧酸循环, 能够影响稻田系统中细菌的能量代谢。值得注意的是, 有两个 AMGs(vOTU303-cat_1_21 和 vOTU590-cat_2_12) 注释为 4-羧基粘内酯脱羧酶(4-carboxymuconolactone decarboxylase), 能使苯甲酸盐降解, 参与芳香族化合物的降解, 该基因的存在能够提高宿主细菌对被石油污染土壤的耐受能力。同时 vOTU551-cat_2_17

编码的 AMGs 被注释为砷酸还原酶 (arsenate reductase), 能够提高稻田系统对重金属污染的修复能力, 虽然该 AMGs 并非高置信度, 但可以说明其具有一定潜力。还有些稻田土壤的病毒 AMGs (vOTU325-cat_1_25) 通过编码乙酰辅酶 A 合成酶 (acetyl-CoA synthetase), 参与包括甲烷合成、糖酵解和丙酮酸代谢在内的多种代谢途径。此外, 还有一些 AMGs 编码亚硝酸还原酶和钴硫合酶, 分别参

与了亚硝酸盐降解和钴胺素（维生素 B12）生物合成。

病毒携带的 AMGs 说明了稻田土壤病毒是参与生物地球化学循环的重要组成部分，并且对环境土壤改善也起到一定的调节作用。

2.4 碳水化合物活性酶基因多样性

稻田病毒群中含有丰富的 CAZymes 基因，而病毒 AMGs 通常通过水平基因转移从宿主中获得^[38]。在本次预测出的 CAZymes 基因中有 13 个来源于具有较高完整性病毒序列（病毒序列完整性大于 50%，图 4a），分别属于糖苷水解酶、多糖裂解酶和碳水化合物结合模块。其中 PL22 蛋白与果胶酶同源，能使果胶骨架裂解，而 GH74、GH43、GH2、GH26 和 GH130 蛋白与纤维素酶同源，能够水解半纤维。这些病毒碳水化合物代谢酶可能有助于稻田系统中多糖的降解。此外还有一种病毒 AMGs 编码 GH13，参与淀粉水解过程。

为预测这些 CAZymes 基因（表 1）的潜在来源，对上述 CAZymes 基因进行 BLASTp 比对和系统发育分析，其中有 24 个高置信度 CAZymes 基因能够预测出假定的基因来源宿主（图 4b），其余 AMGs 基因在系统发育树构建过程中，与多种宿主来源的基因混杂在一起，不能明确基因的宿主来源情况。

稻田病毒群中的 CAZymes 基因主要来源于厚壁菌门的芽孢杆菌目（*Bacillales*）、拟杆菌门（*Bacteroidota*）、假单胞菌门的肠杆菌目（*Enterobacterales*）和放线菌门的放线菌纲（*Actinomycetes*）中，该结果与前人在稻田中细菌群落丰度描述的结果一致^[16]，该研究发现稻田系统中，变形杆菌门（*Proteobacteria*）、放线菌门（*Actinobacteria*）、绿弯菌门（*Chloroflexi*）、酸杆菌门（*Acidobacteria*）和拟杆菌门（*Bacteroidetes*）是五个丰度最高的门，这意味着病毒 CAZymes 基因可能在稻田群落中广泛存在。

在能够预测出宿主的 CAZymes 基因中，CBM32 表现出较高的多样性，其分布在伯克氏菌目的 *Ramlibacter* 和芽孢杆菌目的 *Paenibacillus*，*Lysinibacillus* 和 *Clostridium*。值得注意的是，有 5 个 AMGs 基因（包括 2 个 CBM32）来源于假定的溶原病毒（噬菌体）序列（图 4b 中椭圆形节点），推测该类基因的存在可能参与到细菌代谢过程中，能够增强宿主细菌的碳水化合物合成功能。

在具有较高序列完整性的病毒中，序列 vOTU206 含有两个高置信度的糖苷水解酶基因（图 4a），分别是糖苷水解酶家族 43（GH43）和 β -半乳糖苷酶（GH2）。GH43 中的大多数酶利用芳基糖苷作为底物，同时显示 α -L-阿拉伯呋喃糖苷酶和 β -D-木糖苷酶活性，而 GH2 的主要天然底物是乳糖，可将乳糖转化为半乳糖和葡萄糖（水解）或异乳糖（半乳糖基转移）。通过系统发育分析发现（图 4c），这两个糖苷水解酶基因均来自拟杆菌门（*Bacteroidota*），而根据蛋白质三维结构模型，预测其存在催化复合糖的水解的能力。

此外，稻田病毒群中还含有水解碳水化合物的酯类（CE）相关的酶（表 1），如多糖脱乙酰基酶（*polysaccharide deacetylase*, CE4）和 α/β 水解酶（CE1）等。其中，携带多糖脱乙酰基酶的病毒具有两种不同的宿主来源，分别是放线菌和芽孢杆菌，进一步说明了稻田病毒群中 CAZymes 的来源多样性。

目前研究发现，病毒 AMGs 与土壤中核心碳代谢和土壤有机质降解有密切的关系，其具有与“自上而下”的病毒裂解作用相反的调控土壤系统碳循环的能力，即病毒通过 AMGs 基因加强或赋予宿主菌的碳代谢功能，也被称为“自下而上”的调控方式^[39]。AMGs 参与代谢多种具有复杂特性的碳水化合物^[20]，其中检测最广泛的是 CAZymes 基因。目前对红树林系统中病毒组的研究中发现，编码 CAZymes 基因的病毒可能有助于红树林沉积物中多糖的降解^[40]。而 CAZymes 基因不仅仅在宏基因组数据中被表征，也出现在噬菌体分离株中，如枯草芽孢噬菌体 SP32、 ϕ NIT10 和 SPG1 中均含有 GH32，该基因是由水平基因转移从乳酸芽孢杆菌（*Sporolactobacillus laevolacticus*）原始祖先宿主中获得，其具有左旋果糖转移酶的功能^[41]。

本研究在稻田土壤中发现多种 CAZymes 基因，包括了低聚糖氧化酶、乙酰木聚糖酯酶、果胶乙酰酯酶、 β -甘露聚糖酶、 α -淀粉酶、鼠李糖水解酶、 β -半乳糖苷酶、纤维素酶和 β -葡萄糖苷酶等，可能参与到纤维素、乳糖、鼠李糖、淀粉和葡萄糖等多种糖类物质的代谢过程中。这意味着含有 CAZymes 基因的病毒群体可能具有在稻田土壤进行有机碳分解的潜在作用，值得注意的是，尽管病毒 AMGs 在不同的土壤环境中共享，但 CAZymes 基因会在局部尺度上表现出生态位划分。例如，因土壤中碳源底

物构成的差异，导致了红树林土壤中具有参与海藻糖降解的 GH37 和 GT20，而在稻田土壤系统中并不存在^[40, 42]，说明不同土壤系统中病毒参与全球碳循环过程也可能存在差异。虽然稻田土壤病毒，特别

是噬菌体，携带大量碳水化合物的催化基因，但目前仅有少部分的病毒 CAZymes 基因被验证参与碳代谢^[41, 43]。总之，关于稻田土壤病毒 AMGs 如何参与到土壤碳代谢过程还需要进一步的研究。

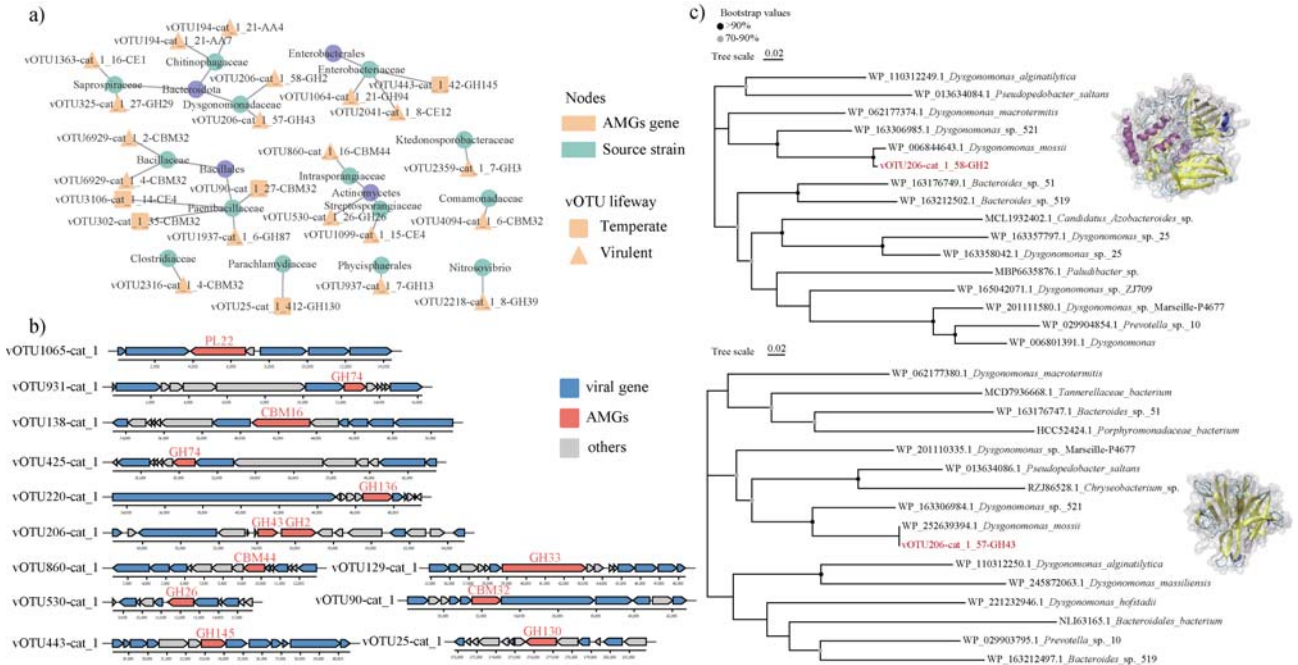


图 4 稻田土壤病毒中碳水化合物活性酶基因中部分高置信基因簇 (a)、基因来源情况 (b) 及 vOTU206 序列中 GH43 和 GH2 系统发育树 (c)

Fig. 4 Partial high-confidence gene clusters of carbohydrate-active enzyme genes in paddy soil virus (a), gene origin status (b), and phylogenetic tree of GH43 and GH2 in vOTU206 sequences (c)

3 结论

本研究使用 ENA 公开数据，构建了迄今为止最大的稻田土壤病毒数据库 (PSVD)，该数据库由 8 791 个病毒 vOTU 组成，有 31.41% 的序列能够预测出物种分类地位信息，这些病毒序列分属于 76 个已知的病毒科，其中 1 145 个 vOTU 是具有较高完整性的病毒序列。与已知的病毒序列比对发现，稻田土壤中存在很多未被表征的病毒群系，该数据库的发布将极大地填补稻田病毒组研究的不足，并为之后稻田病毒组的相关研究奠定基础。根据对稻田土壤病毒中所携带的 AMGs，尤其是 CAZymes 进行分析，该基因主要来源于稻田土壤中常见的细菌病毒类型，说明含有该类型基因的病毒（噬菌体）可能广泛存在于稻田群落中，可能对该系统中的碳循环起到潜在的作用。

参考文献 (References)

- [1] Sun A Q, Jiao X Y, Chen Q L, et al. Microbial communities in crop phyllosphere and root endosphere are more resistant than soil microbiota to fertilization[J]. *Soil Biology & Biochemistry*, 2021, 153: 108113.
- [2] Cobián Güemes A G, Youle M, Cantú V A, et al. Viruses as winners in the game of life[J]. *Annual Review of Virology*, 2016, 3 (1): 197—214.
- [3] Han L L, Cao M M, Bi L, et al. Research advances and application prospect of soil viruses [J]. *Science & Technology Review*, 2022, 40 (3): 75—86. [韩丽丽, 曹苗苗, 毕丽, 等. 土壤病毒的研究进展与应用前景[J]. *科技导报*, 2022, 40 (3): 75—86.]
- [4] Wang G H, Liu J J, Zhu D, et al. A review of researches on viruses in soil – Advancement and challenges[J]. *Acta Pedologica Sinica*, 2020, 57 (6): 1319—1332. [王光华, 刘俊杰, 朱冬, 等. 土壤病毒的研究进展与挑战[J]. *土壤学报*, 2020, 57 (6): 1319—1332.]
- [5] Gao R C, Hu M, Li F B, et al. Research progress and ecological function of phages in soil[J]. *Journal of South*

- China Agricultural University, 2022, 43 (5): 1—11. [高瑞川, 胡敏, 李芳柏, 等. 土壤噬菌体的研究进展及生态功能解析[J]. 华南农业大学学报, 2022, 43 (5): 1—11.]
- [6] Erez Z , Steinberger-Levy I , Shamir M , et al. Communication between viruses guides lysis-lysogeny decisions[J]. *Nature*, 2017, 541 (7638): 488—493.
- [7] Yang M Y , Chen H F , Huang Q L , et al. Characterization of the novel phage vB_VpaP_FE11 and its potential role in controlling *Vibrio parahaemolyticus* biofilms[J]. *Viruses*, 2022, 14 (2): 264.
- [8] Roux S , Brum J R , Dutilh B E , et al. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses[J]. *Nature*, 2016, 537 (7622): 689—693.
- [9] Zhan Y C , Chen F. Bacteriophages that infect marine roseobacters: Genomics and ecology[J]. *Environmental Microbiology*, 2019, 21 (6): 1885—1895.
- [10] Han L , Yu D , Bi L , et al. Distribution of soil viruses across China and their potential role in phosphorous metabolism[J]. *Environmental Microbiome*, 2022, 17. <https://doi.org/10.1186/s40793-022-00401-9>.
- [11] Bi L , Yu D T , Du S , et al. Diversity and potential biogeochemical impacts of viruses in bulk and rhizosphere soils[J]. *Environmental Microbiology*, 2021, 23 (2): 588—599.
- [12] Cao M M , Liu S Y , Bi L , et al. Distribution characteristics of soil viruses under different precipitation gradients on the Qinghai-Tibet plateau[J]. *Frontiers in Microbiology*, 2022, 13: 848305.
- [13] Santos-Medellin C , Zinke L A , ter Horst A M , et al. Viromes outperform total metagenomes in revealing the spatiotemporal patterns of agricultural soil viral communities[J]. *The ISME Journal*, 2021, 15 (7): 1956—1970.
- [14] Ruan C J , Xiong G Z , Niu X Y , et al. Preliminary analysis of the virus metagenome in a farmland black soil of northeast China[J]. *Acta Pedologica Sinica*, 2022, 59 (5): 1447—1456. [阮楚晋, 熊广州, 牛欣尧, 等. 一个东北农田黑土样品宏病毒组的初步分析[J]. 土壤学报, 2022, 59 (5): 1447—1456.]
- [15] Li Y T , Sun H , Yang W C , et al. Dynamics of bacterial and viral communities in paddy soil with irrigation and urea application[J]. *Viruses*, 2019, 11 (4): 347.
- [16] Wang Y , Liu Y , Wu Y , et al. Heterogeneity of soil bacterial and bacteriophage communities in three rice agroecosystems and potential impacts of bacteriophage on nutrient cycling[J]. *Environmental Microbiome*, 2022, 17. <https://doi.org/10.1186/s40793-022-00410-8>.
- [17] Yu G Y. Discussion on water-saving irrigation cultivation technology of rice in Northern China[J]. *Agricultural Machinery and Agronomy*, 2023, 54 (6): 75—77. [于广洋. 浅谈北方水稻节水灌溉栽培技术[J]. 南方农机, 2023, 54 (6): 75—77.]
- [18] Wei L , Ge T D , Zhu Z K , et al. Comparing carbon and nitrogen stocks in paddy and upland soils: Accumulation, stabilization mechanisms, and environmental drivers[J]. *Geoderma*, 2021, 398: 115121.
- [19] Williamson K E , Radosevich M , Wommack K E. Abundance and diversity of viruses in six Delaware soils[J]. *Applied and Environmental Microbiology*, 2005, 71 (6): 3119—3125.
- [20] Wu R , Davison M R , Nelson W C , et al. DNA viral diversity, abundance, and functional potential vary across grassland soils with a range of historical moisture regimes[J]. *mBio*, 2021, 12 (6): e259521.
- [21] Li D H , Liu C M , Luo R B , et al. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph[J]. *Bioinformatics*, 2015, 31 (10): 1674—1676.
- [22] Ren J , Song K , Deng C , et al. Identifying viruses from metagenomic data using deep learning[J]. *Quantitative Biology*, 2020, 8 (1): 64—77.
- [23] Guo J R , Bolduc B , Zayed A A , et al. VirSorter2: A multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses[J]. *Microbiome*, 2021, 9 (1): 37.
- [24] Nayfach S , Camargo A P , Schulz F , et al. CheckV assesses the quality and completeness of metagenome-assembled viral genomes[J]. *Nature Biotechnology*, 2021, 39 (5): 578—585.
- [25] Bin Jang H , Bolduc B , Zablocki O , et al. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks[J]. *Nature Biotechnology*, 2019, 37 (6): 632—639.
- [26] Jiang J Z , Yuan W G , Shang J Y , et al. Virus classification for viral genomic fragments using PhaGCN2[J]. *Briefings in Bioinformatics*, 2023, 24 (1): bbac505.
- [27] Nakamura T , Yamada K D , Tomii K , et al. Parallelization of MAFFT for large-scale multiple sequence alignments[J]. *Bioinformatics*, 2018, 34 (14): 2490—2492.
- [28] Capella-Gutiérrez S , Silla-Martínez J M , Gabaldón T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses[J]. *Bioinformatics*, 2009, 25 (15): 1972—1973.
- [29] Price M N , Dehal P S , Arkin A P. FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix[J]. *Molecular Biology and Evolution*, 2009, 26 (7): 1641—1650.
- [30] Xie J M , Chen Y R , Cai G J , et al. Tree Visualization By One Table (tvBOT): A web application for visualizing, modifying and annotating phylogenetic trees[J]. *Nucleic Acids Research*, 2023, 51 (W1): W587—W592.
- [31] Nayfach S , Pérez-Espino D , Call L , et al. Metagenomic

- compendium of 189, 680 DNA viruses from the human gut microbiome[J]. *Nature Microbiology*, 2021, 6 (7): 960—970.
- [32] Shkoporov A N, Clooney A G, Sutton T D S, et al. The human gut virome is highly diverse, stable, and individual specific [J]. *Cell Host & Microbe*, 2019, 26 (4): 527—541.e5.
- [33] Shaffer M, Borton M A, McGivern B B, et al. DRAM for distilling microbial metabolism to automate the curation of microbiome function[J]. *Nucleic Acids Research*, 2020, 48 (16): 8883—8900.
- [34] ter Horst A M, Santos-Medellin C, Sorensen J W, et al. Minnesota peat viromes reveal terrestrial and aquatic niche partitioning for local and global viral populations[J]. *Microbiome*, 2021, 9: 242.
- [35] Pratama A A, Bolduc B, Zayed A A, et al. Expanding standards in viromics: In silico evaluation of dsDNA viral genome identification, classification, and auxiliary metabolic gene curation[J]. *Peer J*, 2021, 9: e11447.
- [36] Tang Z Q, Zhang L Y, He N, et al. Soil bacterial community as impacted by addition of rice straw and biochar[J]. *Scientific Reports*, 2021, 11 (1): 22185.
- [37] Tang Z, Zhang L, He N, et al. Influence of planting methods and organic amendments on rice yield and bacterial communities in the rhizosphere soil[J]. *Frontiers in Microbiology*, 2022, 13: 918986.
- [38] Sullivan M B, Lindell D, Lee J A, et al. Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts[J]. *PLoS Biology*, 2006, 4 (8): e234.
- [39] Trubl G, Jang H B, Roux S, et al. Soil viruses are underexplored players in ecosystem carbon processing[J]. *mSystems*, 2018, 3 (5): e00076-18.
- [40] Jin M, Guo X, Zhang R, et al. Diversities and potential biogeochemical impacts of mangrove soil viruses[J]. *Microbiome*, 2019, 7 (1): 58.
- [41] Maaroufi H, Levesque R C. Glycoside hydrolase family 32 is present in *Bacillus subtilis* phages[J]. *Virology Journal*, 2015, 12: 157.
- [42] Sun M M, Yuan S J, Xia R, et al. Underexplored viral auxiliary metabolic genes in soil: Diversity and eco-evolutionary significance[J]. *Environmental Microbiology*, 2023, 25 (4): 800—810.
- [43] Emerson J B, Roux S, Brum J R, et al. Host-linked soil viral ecology along a permafrost thaw gradient[J]. *Nature Microbiology*, 2018, 3 (8): 870—880.

(责任编辑：卢萍)