

DOI: 10.11766/trxb202104230215

阮楚晋, 熊广州, 牛欣尧, 陈国炜, 吴汉卿, 马泽超, 朱堃, 刘莹, 王钢. 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.

RUAN Chujun, XIONG Guangzhou, NIU Xinyao, CHEN Guowei, WU Hanqing, MA Zechao, ZHU Kun, LIU Ying, WANG Gang. Preliminary Analysis of the Virus Metagenome in a Farmland Black Soil of Northeast China[J]. Acta Pedologica Sinica, 2022, 59 (5): 1447–1456.

一个东北农田黑土样品宏病毒组的初步分析*

阮楚晋^{1, 5}, 熊广州², 牛欣尧¹, 陈国炜³, 吴汉卿¹, 马泽超¹, 朱堃¹,
刘莹⁴, 王钢^{1†}

(1. 中国农业大学土地科学与技术学院, 北京 100083; 2. 华中科技大学生命科学与技术学院, 武汉 430074; 3. 合肥工业大学土木与水利工程学院, 合肥 230009; 4. 北京大学地球与空间科学学院, 北京 100871; 5. 瑞士联邦水科学与技术研究所, 杜本多夫 8600)

摘要: 病毒在地球上无处不在, 几乎能感染任何生物, 包括动物、植物、真菌以及细菌等, 因而在生物地球化学元素和能量的循环过程中发挥重要作用。了解土壤病毒基因信息有助于深入理解病毒在生态系统中所扮演的角色。选取一个东北农田黑土样品中的病毒为研究对象, 基于病毒宏基因组测序技术获得土壤病毒基因序列, 并利用生物信息分析方法揭示土壤病毒多样性。同时, 结合个性化分析宏病毒组基因序列、进行功能基因分析、宿主预测和单病毒基因组组装和注释。研究发现该农田土壤检测到的病毒主要归属于有尾噬菌体目 (Caudovirales, 59.38%) 和疱疹病毒目 (Herpesvirales, 2.56%) 等 2 个病毒目中的 29 个病毒科, 其中以长尾噬菌体科 (Siphoviridae)、微小噬菌体科 (Microviridae) 的病毒数量最多, 分别占 44.48% 和 20.53%。基因功能分析表明土壤病毒可能参与土壤中的酶催化、生物代谢 (如氮化合物代谢、分解代谢、多生物代谢、细胞代谢、初级代谢、含碱基小分子代谢以及有机物代谢等) 等过程。宿主预测分析揭示检测到的病毒宿主分属 5 个菌门中的 35 个菌属。研究结果丰富了土壤病毒的基因数据库, 为土壤病毒分离提供参考, 并为进一步理解土壤病毒生态学意义提供数据支撑。

关键词: 黑土; 病毒; 宏病毒组; 病毒群落结构; 生态功能

中图分类号: S154.3 **文献标志码:** A

Preliminary Analysis of the Virus Metagenome in a Farmland Black Soil of Northeast China

RUAN Chujin^{1, 5}, XIONG Guangzhou², NIU Xinyao¹, CHEN Guowei³, WU Hanqing¹, MA Zechao¹, ZHU Kun¹, LIU Ying⁴, WANG Gang^{1†}

(1. College of Land Science and Technology, China Agricultural University, Beijing 100083, China; 2. College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China; 3. College of Civil Engineering, Hefei University of

* 国家自然科学基金项目 (41877412)、中国农业大学 2115 人才培养发展支持计划 (1191-00109012) 和中国农业大学研究生自主创新研究基金项目 (2020XYZC24A) 资助 Supported by the National Natural Science Foundation of China (41877412), the 2115 Talent Development Program of China Agricultural University (1191-00109012), and the Research Innovation Fund for Graduate Students of CAU (2020XYZC24A)

† 通讯作者 Corresponding author, E-mail: gangwang@cau.edu.cn

作者简介: 阮楚晋 (1993—), 男, 江西九江人, 博士研究生, 主要从事微生物学方面研究。E-mail: chujin_ruan@aliyun.com

收稿日期: 2021-04-23; 收到修改稿日期: 2021-06-07; 网络首发日期 (www.cnki.net): 2021-07-25

Technology, Hefei 230009; 4. School of Earth and Space Sciences, Peking University, Beijing 100871, China; 5. Swiss Federal Institute of Aquatic Science and Technology, Dübendorf 8600, Switzerland)

Abstract: 【Objective】 Viruses exist almost everywhere on Earth. They can infect humans, animals, plants, fungi, and even bacteria, and thus play critical roles in the biogeochemical cycling of elements and energy. Knowledge of the viruses' community patterns and genomic information is key for proper understanding of their crucial ecological functions. This study aims to quantify the viral community patterns and functions in a farmland black soil of Northeast China and analyze their potential roles in soil ecosystems. 【Method】 The metagenome sequences of the viruses were obtained based on metagenomics sequencing technology and were analyzed individually for functional gene analysis, host prediction, and genome assembly of viral Contigs. A bioinformatics software was employed to analyze the diversity patterns of the viruses. 【Result】 The results showed that the majority of the viruses in the farmland black soil belonged to Caudovirales(59.38%)and Herpesvirales(2.56%)orders. This consisted of 29 families, with Siphoviridae and Microviridae representing the top two abundant families and accounting for about 44.48% and 20.53%, respectively. Gene annotation analysis revealed that soil viruses may actively contribute to numerous biogeochemical cycling processes, including nitrogen compounds metabolism, catabolism, biosynthesis, methylation, multi-organism metabolism, cellular metabolism, primary metabolism, nucleobase-containing small molecules metabolism, and organic substance metabolism, etc. Additional analysis showed that the virus hosts could be classified into 35 genera that belong to 5 phyla. 【Conclusion】 With these quantitative measurements, this study provides necessary information towards advancing the virus gene database of the black soil and its ecological functions.

Key words: Black soil; Virus; Viral metagenome; Viral community; Ecological function

病毒在土壤元素生物地球化学循环中发挥重要作用, 土壤生态系统中每克土含有的病毒颗粒 (Virus-like particles, VLPs) 丰度可达 10^9 VLPs^[1-3]。病毒基因结构组成非常复杂, 大量未知的遗传基因被认为是基因库中的“暗物质”, 对这些未知基因的研究, 有助于揭开生命的起源、演替和更迭规律, 帮助人们理解和确定病毒在生态系统中的地位和作用^[4]。

自然界大量的病毒是构成生态系统的重要基础, 而分离技术的缺乏是目前限制病毒研究发展的主要原因之一, 近年来发展起来的基因组学技术为探索病毒的奥秘提供了有效途径, 尤其是高通量测序法已经成为研究环境中微生物群落结构和功能分析的重要工具。高通量测序法具有信息量大、对样本所含信息无偏倚、操作较简单等优点, 避免了传统的微生物分离培养步骤, 可以直接从环境样品中提取总 DNA, 是有效提升微生物生态学研究的重要手段^[5]。由于土壤具有复杂的空间结构 (比表面积达 $1.1 \times 10^6 \sim 8 \times 10^6 \text{ cm}^2 \cdot \text{g}^{-1}$), 并与土壤各要素 (水分、养分、空气和温度等) 共同形成了特有的微观异质性, 为土壤病毒和其他微生物的生命活动创造了多重生态位, 这些因素给土壤微生物学相关研究带来

巨大的挑战^[6-7]。目前仅有少数宏病毒组学研究揭示了土壤中病毒群落结构^[8-11]。Yu 等^[12]收集了新疆独山子区火山土壤的病毒, 并通过宏基因组数据分析病毒群落结构和功能, 鉴定出了 8 个完整的新病毒基因组。Jin 等^[13]利用病毒宏基因组测序研究了从中国南方不同红树林生境采集的 6 个红树林土壤样本中的病毒群落, 并从中鉴定出了丰富的辅助代谢基因。

黑土是一种结构好、肥力高的土壤, 尤其适合种植玉米、小麦、大豆等农作物, 被誉为耕地中的“大熊猫”。在我国, 黑土主要分布于黑龙江、吉林两省, 是我国重要的粮食生产基地。解析黑土微生物群落结构和功能, 尤其是病毒的研究对我国的作物生产和安全有重要意义。当前, 关于黑土地土壤病毒的研究还十分匮乏, Wang 等^[14]利用分子生物学技术, 以 T4 型噬菌体主要衣壳蛋白基因 (*g23*) 作为目标基因, 揭示了东北黑土中不同施肥处理条件下 T4 型噬菌体的多样性。但目前对于黑土病毒多样性特征及其生态功能的认识仍十分有限。

基于目前发展的宏病毒组测序技术以及生物信息学的数据挖掘能力, 本研究拟以一个东北农田黑土样品中的病毒作为研究对象, 通过宏基因组测序

技术获得土壤宏病毒基因序列，利用生物信息学分析揭示土壤病毒多样性及其生态功能。

1 材料与方法

1.1 土壤样品采集

于 2019 年 11 月从位于吉林省梨树县大房身乡高家村的中国科学院保护性耕作实验基地 (43°19' N, 124°14' E) 采集耕层土壤样品 (0~20 cm)。该处土壤以玉米连作为主，经历多年传统耕作，质地为壤质黏土。利用直径为 2 cm 的土钻采用网格抽样法随机选取每个样地内采样点，剔除杂物和 2 mm 以上的沙砾，以四分法将 5 份样品等量混匀得到一份土样；采样过程使用无菌手套操作，并提前去除样品外围可能污染的部分；采集好的土壤样品放置无菌袋中并放置于冰盒中暂时保存，当日运回实验室后 4°C 保存备用。

1.2 病毒的收集

利用切向流膜过滤系统切向流过滤法收集病毒^[15]。具体步骤：称取 500 g 土壤样品，加入 800 mL 浸提液 (1% 柠檬酸钾、250 mmol·L⁻¹ 甘氨酸) 中。将土壤悬液置于 180 r·min⁻¹ 摇床震荡 30 min 后，转移至离心机 3 500×g 离心 20 min 收集上层溶液备用。离心管底的土壤样品加入 1.2 L 浸提液重悬并搅匀，180 r·min⁻¹ 摇床震荡 30 min 后，再次将土壤悬液转移至离心机 3500 g 离心并收集上层溶液备用。该浸提液洗脱过程，重复 3 次。所收集的上层溶液用 10 层纱布过滤去除漂浮物后转移溶液至切向流膜过滤系统 (AKTA Flux S) 中，分别用 0.45 μm、0.22 μm 和 30 kDa 滤管进行过滤和浓缩，得到 200 mL 病毒浓缩液。将病毒浓缩液转移至 50 mL 30 kDa 超滤管，3500×g 离心 8 min，移除滤液。每管超滤管加入 1 mL PBS 缓冲液充分重悬滤膜中富集的病毒颗粒，置入 -80°C 冰箱保存。

1.3 基因测序

为确保收集的病毒不被环境中游离的 DNA 干扰，使用 DNase I 降解土壤溶液中游离的 DNA，并使用 16S rRNA 引物 E9F GAGTTTGATCCTGGCT CAG / U1510r GGTTACCTTGTTACGACTT^[16] 通过 PCR 扩增检测样品无细菌 DNA 污染。病毒 DNA 提取采用改良 CTAB 法提取^[17]，消化体系为：200 μL

病毒溶液、140 μL 5×TE、50 μL 蛋白酶 K (2 mg·mL⁻¹)、40 μL 20% SDS。所提取的病毒 DNA 用全基因扩增试剂盒 (illustra™ Ready-To-Go™ GenomiPhi™ V3 DNA Amplification Kit) 进行扩增，扩增后浓度为 284.63 ng·μL⁻¹，并利用超声破碎仪进行随机打断得到 0~23 kb 的短片段 DNA，用于测序文库构建，质检合格的文库将采用 Illumina NovaSeq 6000 (Illumina, San Diego, CA, USA) 平台进行测序。将测序所得宏病毒组原始序列提交至 NCBI Sequence Read Archive (SRA) 数据库，获得数据收录号。

1.4 序列数据预处理

使用 Trimmomatic^[18]，对原始数据进行质量控制，去除低质量读段得到高质量的过滤后读段条数 (Clean reads)。随后使用 BWA v0.7.17^[19] 将 Clean reads 与 NT 数据库分离的病毒参考序列进行比对，初步获得样本中的病毒分类信息。最后使用 Megahit v1.1.2^[20] 高质量 Clean reads 进行组装获得不短于 300 bp 的 Contigs。

1.5 群落结构分析

使用两种方法对病毒 Contigs 进行注释以提高注释的广度和精度。方法一，从 NT 和 NR 数据库中分离出相应的病毒序列，使用 BLAST 将组装得到的 Contigs 与构建的病毒数据库进行比对，基于参考序列进行物种注释。得到可信度较高 (Identity>=80%，length>=500bp，E-value<=1e⁻⁵) 的确认序列。方法二，使用 BLAST 将 Contigs 分别与从 NT 和 NR 分离得到的病毒数据库进行比对，同时将 MetaGeneMark v3.38^[21] 预测得到的基因使用 Hmsearch^[22] 在病毒的 HMM 数据库 (VPFs^[23] 和 vFam^[24]) 进行搜索比对得到候选的病毒 Contigs。使用 Blastn 将候选 Contigs 与 NT 数据库进行比对，随后将未比对上的 Contigs 使用 Diamond v0.9.10^[25] 与 NR 数据库进行比对。将比对结果使用 NCBI Taxonomy 数据库进行注释，若比对前 50 个结果中注释结果超过 80% 为病毒，则认为该 Contig 为病毒序列。方法二得到的新病毒序列和方法一得到的高可信度序列则认为是病毒 Contigs，进行后续分析。

使用 BWA v0.7.17^[19] 将 Clean reads 分别与病毒 Contigs 比对，过滤比对长度低于 80% Reads 总长的比对结果，统计病毒 Reads 比例。随后计算每条病

毒 Contig 的 RPKM 值和样本 Alpha 多样性的 Shannon 指数。

1.6 基因注释以及功能分析

采用 Prokka v1.13^[26]来预测病毒 Contigs 上长度大于 200 bp 的基因序列。将预测得到的基因蛋白序列与 UniProtKB/Swiss-Prot 数据库的病毒序列 (ViralZone^[27], reviewed proteins, <https://viralzone.expasy.org/>) 进行比对, 过滤 E-value>1e⁻³ 的比对结果获得病毒基因功能注释信息^[28]。

1.7 宿主预测

利用 CRISPR-Cas 系统, 研究噬菌体与细菌的关系, 推测噬菌体可能的宿主。使用 CRISPR Recognition Tool^[28] (CRT, <http://www.room220.com/crt/>) 从 RefSeq 数据库的细菌基因组中构建 CRISPR-Cas spacer 数据库, 然后使用 Blastn-short v2.9.0 与前述鉴定的病毒 Contigs 比对, 获取噬菌体可能的宿主信息。由于目前已知的 Spacer 序列有限, 因此只有一部分病毒 Contigs 能预测出可能的宿主。

1.8 特征基因组分析

使用 Prodigal v2.6.3^[29]对两条最长的 Contigs 进行基因预测, 使用 Blastp 将预测得到的基因蛋白序列与 UniProtKB/Swiss-Prot 数据库的病毒序列 (ViralZone^[27], reviewed proteins, <https://viralzone.expasy.org/>) 进行比对, 过滤 E-value>1e⁻³ 的比对结果获得病毒基因功能注释信息。病毒基因组使用 SnapGene 进行可视化。

2 结果与讨论

2.1 病毒丰度及群落结构

通过高通量测序, 从东北农田黑土样品中获得 35 611 102 条原始下机序列 (Raw reads), 测序数据量为 12.2 Gbp, 原始数据在 NCBI 数据库的收录号为 SRR14209300。经过质控后 Clean reads 数占比为 78.85% (28 078 029 条)。对 Clean reads 使用 Megahit 进行拼接, 获得 229 988 条 Contigs (>300 bp)。

通过两种方法对病毒 Contigs 进行物种注释, 注释得到 9 851 条可能的病毒 Contigs (>300 bp)。由于土壤环境的复杂性导致病毒收集过程中其他微生物碎片所携带基因组的干扰, 以及目前土壤病毒基因组库还尚未完善, 大量未知的基因无法被注释,

因此仅有 4% 组装的 Contigs 被注释为病毒。此外, 虽然采用全基因组扩增后测定序列有偏好性, 偏向于单链病毒, 而本文得到最多的是双链病毒, 可能由于样品中 dsDNA 病毒占比较大, 因此能够被注释的 Contigs 也多属于 dsDNA 病毒。病毒群落主要归属于有尾噬菌体目 (Caudovirales) 和疱疹病毒目 (Herpesvirales) 2 个病毒目, 以及 29 个病毒科和一些未分类的病毒 (图 1)。其中长尾噬菌体科 (Siphoviridae)、微小噬菌体科 (Microviridae)、肌病毒科 (Myoviridae) 相对丰度最高, 分别为 44.48%、20.53% 和 8.38%。噬菌体占 55.42%, 而其他的病毒占 44.58%。基于病毒序列的 RPKM 值计算该东北农田土壤病毒的香农指数为 8.93。研究发现, 噬菌体是该农田土壤的主要类群, 其中长尾噬菌体科 (Siphoviridae) 是一类 dsDNA 病毒, 可以感染细菌和古菌, 不仅在深海沉积物^[30-31]、沙漠土壤^[32]、农田土壤^[33]以及森林土壤^[11]广泛分布, 在生活污水^[34]、人体肠道^[35]等与人类息息相关的环境中也均被报道, 是一类重要的环境病毒。微小噬菌体科则是一类 ssDNA 病毒, 有研究报道在部分农田土壤和森林土壤中微小噬菌体科病毒数量占比最大^[36]。其他以微生物为宿主的病毒科肌病毒科 (Myoviridae)、短尾噬菌体科 (Podoviridae)、代列尔噬菌体科 (Herelleviridae)、Ackermannviridae 科、三层病毒科 (Tristromaviridae, 侵染古菌)、丝杆病毒科 (Inoviridae)、Lavidaviridae 科。除噬菌体外, 该样品中还有以原生动物为宿主的拟菌病毒科 (Mimiviridae)、阔口罐病毒科 (Pithoviridae)、马赛病毒科 (Marseilleviridae), 它们属于核质巨型 DNA 病毒, 拥有较大的衣壳直径和庞大的基因组。与植物相关的病毒则有藻类 DNA 病毒科 (Phycodnaviridae)、Genomoviridae 科、双生病毒科 (Geminiviridae)、矮化病毒科 (Nanoviridae)、Alphasatellitidae 科。此外, 还发现了多种与动物致病相关的病毒, 包含杆状病毒科 (Baculoviridae)、虹彩病毒科 (Iridoviridae)、Polydnaviridae 科、Nudiviridae 科、细小病毒科 (Parvoviridae)、疱疹病毒科 (Herpesviridae)、圆环病毒科 (Circoviridae)、痘病毒科 (Poxviridae)、异疱疹病毒科 (Alloherpesviridae)、腺病毒科 (Adenoviridae)、乳头瘤病毒科 (Papillomaviridae)。而其中人类健康息息相关的病毒, 有疱疹病毒科 (Herpesviridae)、痘病毒科 (Poxviridae)、

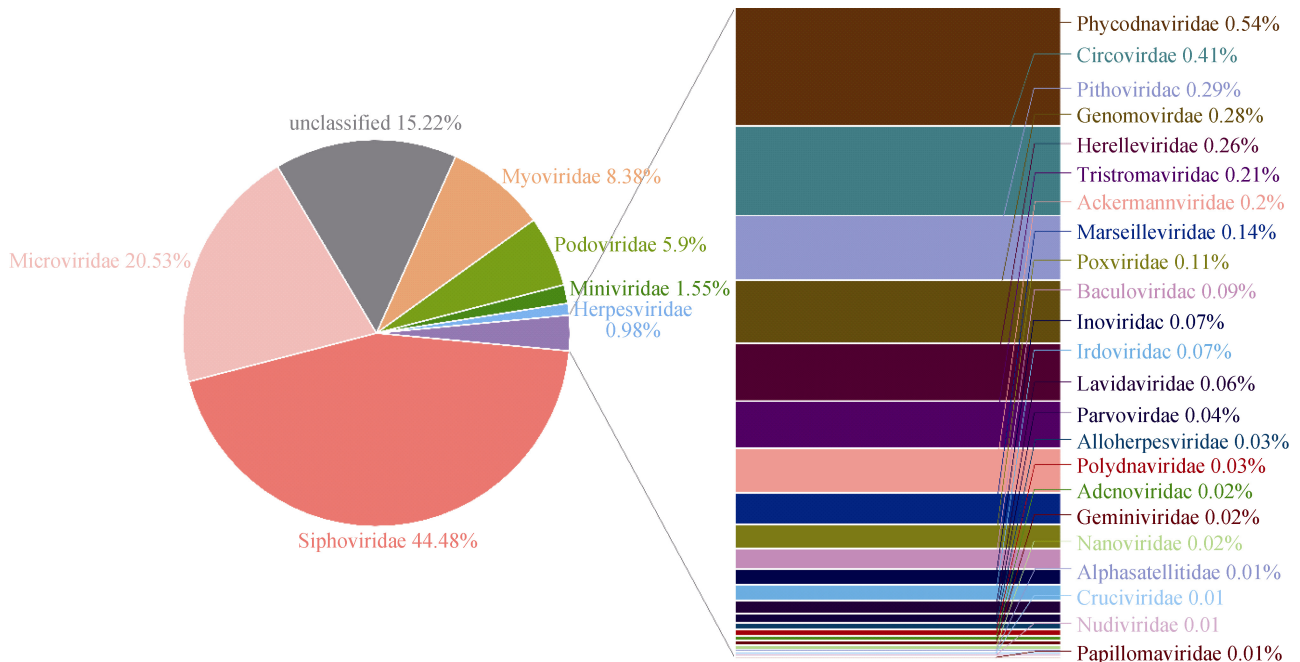


图 1 病毒物种注释结果 (科水平)

Fig. 1 Taxonomic composition of viromes (family level)

异疱疹病毒科 (Alloherpesviridae)、腺病毒科 (Adenoviridae)、乳头瘤病毒科 (Papillomaviridae) 和圆环病毒科 (Circoviridae) 等, 这也暗示土壤成为了人类病毒的储存库, 对人类的健康有着潜在的影响。

2.2 土壤病毒基因预测及功能分析

病毒在生态系统中发挥着关键作用。病毒可以通过辅助代谢基因 (Auxiliary metabolic genes, AMGs) 重新编码宿主基因而影响其新陈代谢^[37-38]。同时, 病毒还可以通过介导水平基因转移驱动宿主进化^[39-40]。本研究中, 使用 Prokka v1.13 对病毒 Contigs 进行基因预测, 一共预测到 808 个 tRNA 和 12 058 个 CDS (Coding sequence)。在 UniprotKB/Swiss-Prot 数据库中有效注释病毒序列 1 867 个基因。

注释结果中, 分子功能相关基因共有 179 个, 其中占比较高的是催化活性基因 (Catalytic activity), 共有 146 个, 包含氧化还原酶活性、转移酶活性、水解酶活性、裂解酶活性、肽聚糖的遗传活性、催化蛋白质活性、催化 DNA 活性、催化 RNA 活性。这些微生物催化酶活性方面的相关基因在微生物细胞生长和遗传发育中起着不可替代的调控作用。维持病毒繁殖和生理结构的相关基因共有 124 个, 包含病毒衣壳蛋白、病毒膜蛋白、尾部和末端

酶等。生物过程代谢相关基因共有 105 个, 包含氮化合物代谢过程、分解代谢过程、生物合成过程、甲基化、多生物代谢过程、细胞代谢过程、初级代谢过程、含碱基小分子代谢过程以及有机物代谢过程, 相关基因参与了微生物介导的诸多物质循环过程, 如代谢过程。这些过程在不同程度下均影响了微生物参与的地球元素化学循环。

2.3 噬菌体宿主预测

目前土壤病毒资源的获取在微生物学中仍然是个巨大的挑战, 病毒宿主的预测可以有效地指导单病毒的分离工作。本研究从 Refseq 数据库中的细菌基因组挖掘 CRISPR-Cas 序列构建 CRISPR-Cas spacer 数据库, 对病毒 Contigs 比对搜索获取可能的宿主信息。但由于已知的 Spacer 序列有限, 因此样品中仅有 66 条病毒 Contigs 能预测出可能的宿主, 具体结果详见表 1。其中有 30 条 (45.45%) Contigs 被预测宿主属变形菌门, 占能被预测宿主最多比例, 宿主分属于无色杆菌属 (*Achromobacter*)、不动杆菌属 (*Acinetobacter*)、气单胞菌属 (*Aeromonas*)、弓形杆菌属 (*Arcobacter*)、伯克氏菌属 (*Burkholderia*)、*Candidimonas* 属、柠檬酸杆菌属 (*Citrobacter*)、克洛诺菌属 (*Cronobacter*)、克雷伯氏菌属 (*Klebsiella*)、军团菌属 (*Legionella*)、

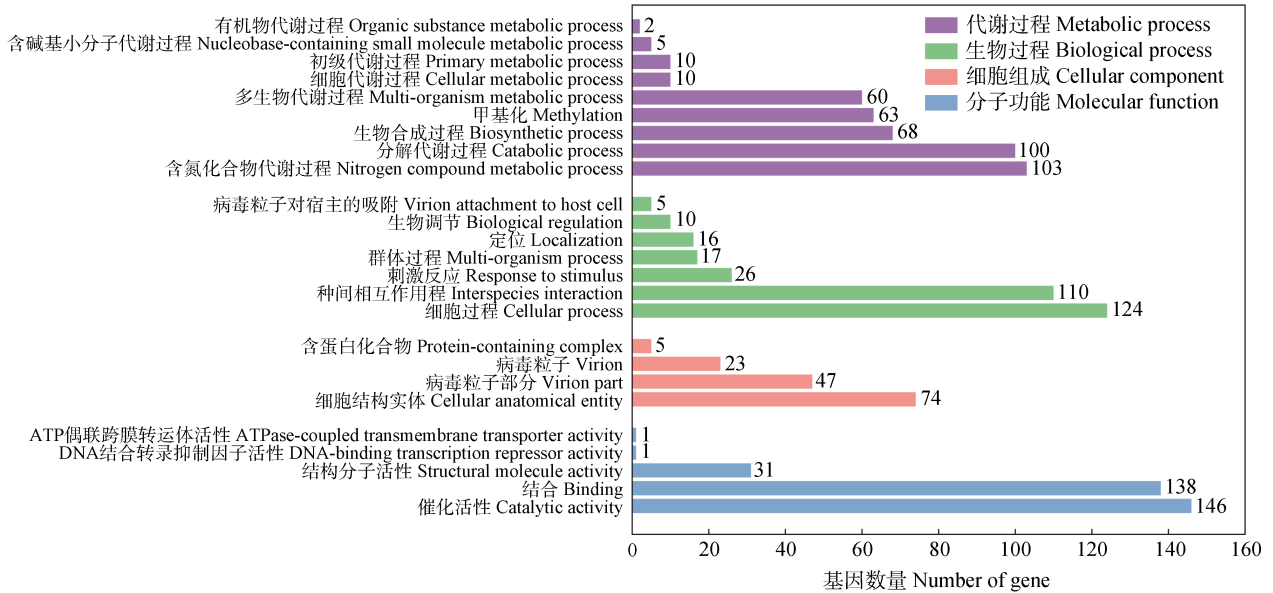


图 2 病毒基因组在 UniProt 数据库中的注释结果

Fig. 2 Annotation of viral genes against UniProt database

表 1 土壤噬菌体宿主预测结果

Table 1 Prediction results of soil phage host

门 Phylum	预测宿主-菌属 Predicted host-Genus	预测宿主-菌种 Predicted host-Species	数量 Number
Actinobacteri	<i>Actinomyces</i>	<i>Actinomyces provencensis/Actinomyces viscosus</i>	2
	<i>Bifidobacterium</i>	<i>Bifidobacterium asteroides</i>	2
	<i>Cellulomonas</i>	<i>Cellulomonas iranensis</i>	1
	<i>Micromonospora</i>	<i>Micromonospora chokoriensis/Micromonospora parva</i>	2
	<i>Nocardia</i>	<i>Nocardia lijiangensis/Nocardia sp. Y48</i>	2
	<i>Plantactinospira</i>	<i>Plantactinospira sp. KBS50</i>	1
	<i>Quadrisphaera</i>	<i>Quadrisphaera sp. DSM 44207</i>	1
	<i>Salinispora</i>	<i>Salinispora arenicola</i>	1
	<i>Streptomyces</i>	<i>Streptomyces autolyticus/Streptomyces xiamenensis/Streptomyces thermovulgaris</i>	7
	<i>Thermobifida</i>	<i>Thermobifida halotolerans</i>	1
Bacteroidetes	<i>Niastella</i>	<i>Niastella koreensis</i>	1
	<i>Parapedobacter</i>	<i>Parapedobacter composti</i>	3
Chlorobi	<i>Pedobacter</i>	<i>Pedobacter tournemirensis</i>	1
	<i>Prosthecochloris</i>	<i>Prosthecochloris sp. ZM</i>	2
Firmicutes	<i>Bacillus</i>	<i>Bacillus amyloliquefaciens/Bacillus cereus/Bacillus licheniformis</i>	4
	<i>Clostridioides</i>	<i>Clostridioides difficile</i>	1
	<i>Cohnella</i>	<i>Cohnella sp. CIP</i>	1
	<i>Desulfitobacterium</i>	<i>Desulfitobacterium hafniense</i>	1

续表

门	预测宿主-菌属 Predicted	预测宿主-菌种	数量
Phylum	host-Genus	Predicted host-Species	Number
Firmicutes	<i>Lactobacillus</i>	<i>Lactobacillus delbrueckii</i>	1
	<i>Megasphaera</i>	<i>Megasphaera stantonii</i>	1
	<i>Achromobacter</i>	<i>Achromobacter</i> sp. AONIH1	1
	<i>Acinetobacter</i>	<i>Acinetobacter</i> sp. CIP 102143	1
	<i>Aeromonas</i>	<i>Aeromonas veronii</i>	1
	<i>Arcobacter</i>	<i>Arcobacter defluvii</i>	1
	<i>Burkholderia</i>	<i>Burkholderia pseudomallei</i>	2
	<i>Candidimonas</i>	<i>Candidimonas nitroreducens</i>	1
	<i>Citrobacter</i>	<i>Citrobacter</i> sp. TSA-1	1
	<i>Cronobacter</i>	<i>Cronobacter sakazakii</i>	1
	<i>Klebsiella</i>	<i>Klebsiella variicola</i>	1
	<i>Legionella</i>	<i>Legionella pneumophila</i>	1
	<i>Methylocaldum</i>	<i>Methylocaldum</i> sp. 14B	1
	<i>Methylomonas</i>	<i>Methylomonas methanica</i>	1
Proteobacteri	<i>Novosphingobium</i>	<i>Novosphingobium</i> sp. AAP83	1
	<i>Oligella</i>	<i>Oligella urethralis</i>	1
	<i>Oxalobacter</i>	<i>Oxalobacter formigenes</i>	1
	<i>Pannonibacter</i>	<i>Pannonibacter phragmitetus</i>	1
	<i>Pseudomonas</i>	<i>Pseudomonas aeruginosa</i> / <i>Pseudomonas</i> sp. TTU2014-080ASC	3
	<i>Pseudospirillum</i>	<i>Pseudospirillum japonicum</i>	1
	<i>Pseudoxanthomonas</i>	<i>Pseudoxanthomonas spadix</i>	1
	<i>Rhodopseudomonas</i>	<i>Rhodopseudomonas faecalis</i>	1
	<i>Salinivibrio</i>	<i>Salinivibrio kushneri</i>	1
	<i>Salmonella</i>	<i>Salmonella enterica</i>	3
	<i>Sutterella</i>	<i>Sutterella parvirubra</i>	1
<i>Vibrio</i>	<i>Vibrio parahaemolyticus</i>	1	
<i>Yersinia</i>	<i>Yersinia pekkaneii</i>	1	

Methylocaldum 属、甲基单胞菌属 (*Methylomonas*)、新鞘氨醇杆菌属 (*Novosphingobium*)、寡养杆菌属 (*Oligella*)、草酸杆菌属 (*Oxalobacter*)、潘隆尼亚碱湖杆菌 (*Pannonibacter*)、假单胞菌属 (*Pseudomonas*)、假螺菌属 (*Pseudospirillum*)、假黄色单胞菌属 (*Pseudoxanthomonas*)、红假单胞菌属 (*Rhodopseudomonas*)、盐弧菌属 (*Salinivibrio*)、沙门氏菌属 (*Salmonella*)、*Sutterella* 属、弧菌属 (*Vibrio*)、耶尔森氏菌属 (*Yersinia*) 15 个菌属中。有 20 条 (30.30%) Contigs 被预测宿主属放线菌门，

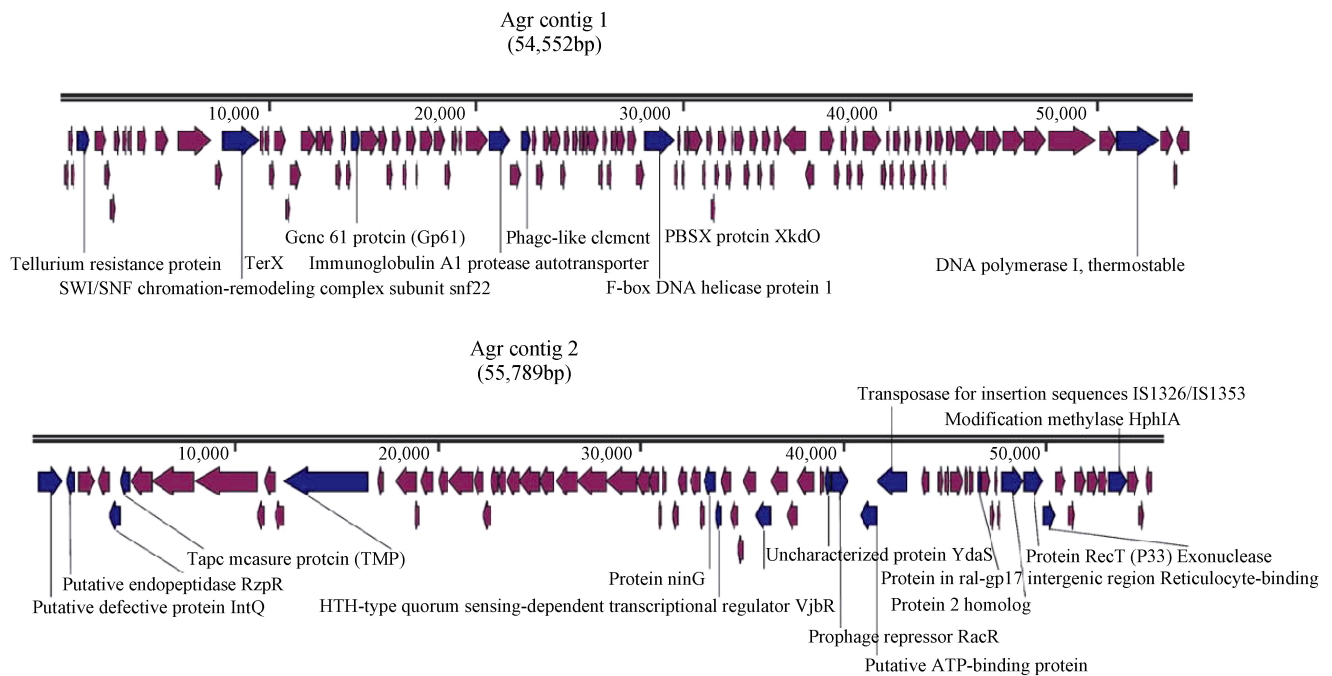
宿主分属于放线菌属 (*Actinomyces*)、双歧杆菌属 (*Bifidobacterium*)、纤维菌属 (*Cellulomonas*)、小单胞菌属 (*Micromonospora*)、诺卡式菌属 (*Nocardia*)、*Plantactinospira* 属、四折叠球菌属 (*Quadrisphaera*)、盐胞菌属 (*Salinispora*)、链霉菌属 (*Streptomyces*)、*Thermobifida* 属 10 个菌属中。有 9 条 (13.63%) Contigs 被预测宿主属厚壁菌门，宿主分属于芽孢杆菌属 (*Bacillus*)、*Clostridioides* 属、*Cohnella* 属、*Desulfitobacterium* 属、乳酸菌属 (*Lactobacillus*)、巨球菌属 (*Megasphaera*) 6 个菌

属中。有 5 条 (7.58%) Contigs 被预测宿主属于拟杆菌门, 宿主分属于 *Niastella* 属、类土地杆菌属 (*Parapedobacter*)、地杆菌属 (*Pedobacter*) 3 个菌属中。有 2 条 (3.03%) Contigs 被预测宿主属于绿菌门, 宿主属于突柄绿菌属 (*Prosthecochloris*)。该农田土壤病毒可预测的宿主分属于放线菌门 (*Actinobacteria*)、拟杆菌门 (*Bacteroidetes*)、绿菌门 (*Chlorobi*)、厚壁菌门 (*Firmicutes*) 和变形菌门 (*Proteobacteria*) 5 个菌门中的 35 个菌属, 其中以变形菌门和放线菌门的病毒预测宿主居多, 分别占 45.45% 和 30.30%。变形菌门和放线菌门作为土壤中高丰度的菌门被广泛报道, 目前已得到分离的噬菌体也大多来自于这两个菌门当中。

2.4 代表性 Contigs 的基因组特征

在土壤的宏病毒组中组装注释得到 2 284 条病

毒 Contigs (>1 kb)。筛选出其中最长的两条 Contigs 分别为 Contig 1 (54 552 bp) 和 Contig 2 (55 789 bp), 分析其功能基因组组成(图 3)。使用 BLAST 基于 NCBI 的 Complete bacteriophages 数据库进行序列比对并进行物种注释, 比对发现 Contig 1 与 *Streptomyces* phage VWB 相似性最高, 相似度为 76.85%^[41]。Contig 2 与 *Pseudomonas* phage PS-1 相似性最高, 相似度为 81.13%^[42]。Contig 1 预测得到 115 个编码序列, Contig 2 预测得到 75 个编码序列。使用 Blastp 基于 Uniprot/Swiss-Prot 数据库进行基因注释, Contig 1 有 7 个基因能被注释, Contig 2 有 17 个基因能被注释。由于土壤病毒研究较少, 相关数据库并不完善, 因此只有少部分基因能够进行功能注释。虽然能够进行功能注释的基因不多, 但是大部分都与病毒的组装、转录翻译及裂解等重要的生理过程相关。



注: 箭头方向为转录方向, 紫色箭头为无法进行功能注释的假象蛋白, 蓝色箭头为能够进行功能注释的蛋白。Note: The arrow direction is the transcription direction, and the purple arrow is the hypothetical protein that cannot be annotated functionally, and the blue arrow is the protein that can be annotated functionally.

图 3 典型 contig 基因注释图 (>1 kb)

Fig. 3 Genomic features of representative viral contigs (>1 kb)

3 结论

本研究首次通过宏病毒组学方法报道了一个东北农田黑土样品中病毒的群落组成和功能特征。通

过数据库比对分析, 发现病毒主要归属于有尾噬菌体目 (*Caudovirales*, 59.38%) 和疱疹病毒目 (*Herpesvirales*, 2.56%) 2 个病毒目以及 29 个病毒科, 其中长尾噬菌体科 (*Siphoviridae*)、微小噬菌体科 (*Microviridae*) 相对丰度最高, 分别为 44.48%

和 20.53%。功能基因注释分析表明了土壤病毒在微生物酶催化和生物代谢等过程中发挥着潜在的作用。对病毒基因组进行了宿主预测，可预测的宿主归属于放线菌门、拟杆菌门、绿菌门、厚壁菌门和变形菌门等 5 个菌门的 35 个菌属中。

参考文献 (References)

- [1] Breitbart M, Rohwer F. Here a virus, there a virus, everywhere the same virus?[J]. *Trends in Microbiology*, 2005, 13 (6): 278—284.
- [2] Reavy B, Swanson M M, Taliany M. Viruses in soil[M]//Dighton J, Krumins J A. Interactions in soil: promoting plant growth. Netherlands: Springer, 2014: 163—180.
- [3] Chen L, Xun W B, Sun L, et al. Effect of different long-term fertilization regimes on the viral community in an agricultural soil of Southern China[J]. *European Journal of Soil Biology*, 2014, 62: 121—126.
- [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] Weinbauer M G. Ecology of prokaryotic viruses[J]. *FEMS Microbiology Reviews*, 2004, 28 (2): 127—181.
- [6] Wilpiszeski R L, Aufrecht J A, Retterer S T, et al. Soil aggregate microbial communities: Towards understanding microbiome interactions at biologically relevant scales[J]. *Applied and Environmental Microbiology*, 2019, 85(14). <https://doi.org/10.1128/aem.00324—19>.
- [7] Fierer N, Breitbart M, Nulton J, et al. Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, Archaea, fungi, and viruses in soil[J]. *Applied and Environmental Microbiology*, 2007, 73 (21): 7059—7066.
- [8] Diemer G S, Stedman K M. A novel virus genome discovered in an extreme environment suggests recombination between unrelated groups of RNA and DNA viruses[J]. *Biology Direct*, 2012, 7: 13.
- [9] Emerson J B, Thomas B C, Andrade K, et al. Dynamic viral populations in hypersaline systems as revealed by metagenomic assembly[J]. *Applied and Environmental Microbiology*, 2012, 78 (17): 6309—6320.
- [10] Yoshida M, Takaki Y, Eitoku M, et al. Metagenomic analysis of viral communities in (hado) pelagic sediments[J]. *PLoS One*, 2013, 8 (2): e57271.
- [11] Kuzyakov Y, Blagodatskaya E. Microbial hotspots and hot moments in soil: Concept & review[J]. *Soil Biology & Biochemistry*, 2015, 83: 184—199.
- [12] Yu D T, He J Z, Zhang L M, et al. Viral metagenomics analysis and eight novel viral genomes identified from the Dushanzi mud volcanic soil in Xinjiang, China[J]. *Journal of Soils and Sediments*, 2019, 19 (1): 81—90.
- [13] Jin M, Guo X, Zhang R, et al. Diversities and potential biogeochemical impacts of mangrove soil viruses[J]. *Microbiome*, 2019, 7 (1): 58.
- [14] Wang G H, Yu Z H, Liu J J, et al. Molecular analysis of the major capsid genes(g23)of T4-type bacteriophages in an upland black soil in Northeast China[J]. *Biology and Fertility of Soils*, 2011, 47 (3): 273—282.
- [15] Thurber R V, Haynes M, Breitbart M, et al. Laboratory procedures to generate viral metagenomes[J]. *Nature Protocols*, 2009, 4 (4): 470—483.
- [16] Ramond J B, Welz P J, Tuffin M I, et al. Assessment of temporal and spatial evolution of bacterial communities in a biological sand filter mesocosm treating winery wastewater[J]. *Journal of Applied Microbiology*, 2013, 115 (1): 91—101.
- [17] Wang H N, Guo Z Q, Feng H L, et al. Novel sulfobolus virus with an exceptional capsid architecture[J]. *Journal of Virology*, 2018, 92 (5): e01727—e01717.
- [18] Bolger A M, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data[J]. *Bioinformatics*: Oxford, England, 2014, 30 (15): 2114—2120.
- [19] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform[J]. *Bioinformatics*, 2009, 25 (14): 1754—1760.
- [20] Li D H, Luo R B, Liu C M, et al. MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices[J]. *Methods*, 2016, 102: 3—11.
- [21] Zhu W H, Lomsadze A, Borodovsky M. Ab initio gene identification in metagenomic sequences[J]. *Nucleic Acids Research*, 2010, 38 (12): e132.
- [22] Eddy S R. A new generation of homology search tools based on probabilistic inference[J]. *Genome Informatics International Conference on Genome Informatics*, 2009, 23 (1): 205—211.
- [23] Paez-Espino D, Eloie-Fadrosh E A, Pavlopoulos G A, et al. Uncovering Earth's virome[J]. *Nature*, 2016, 536 (7617): 425—430.
- [24] Skewes-Cox P, Sharpton T J, Pollard K S, et al. Profile hidden Markov models for the detection of viruses within metagenomic sequence data[J]. *PLoS One*, 2014, 9 (8): e105067.
- [25] Buchfink B, Xie C, Huson D H. Fast and sensitive protein alignment using DIAMOND[J]. *Nature Methods*, 2015, 12 (1): 59—60.
- [26] Seemann T. Prokka : rapid prokaryotic genome annotation[J]. *Bioinformatics*, 2014, 30 (14): 2068—2069.
- [27] Hulo C, de Castro E, Masson P, et al. ViralZone: A knowledge resource to understand virus diversity[J].

- Nucleic Acids Research, 2011, 39 (suppl_1): D576—D582.
- [28] Bland C, Ramsey T L, Sabree F, et al. CRISPR recognition tool(CRT): A tool for automatic detection of clustered regularly interspaced palindromic repeats[J]. BMC Bioinformatics, 2007, 8: 209.
- [29] Hyatt D, Chen G L, Locascio P F, et al. Prodigal: Prokaryotic gene recognition and translation initiation site identification[J]. BMC Bioinformatics, 2010, 11: 119.
- [30] Zheng X W, Liu W, Dai X, et al. Extraordinary diversity of viruses in deep-sea sediments as revealed by metagenomics without prior virion separation[J]. Environmental Microbiology, 2021, 23 (2): 728—743.
- [31] Cai L L, Jørgensen B B, Suttle C A, et al. Active and diverse viruses persist in the deep sub-seafloor sediments over thousands of years[J]. The ISME Journal, 2019, 13 (7): 1857—1864.
- [32] Adriaenssens E M, Van Zyl L, De Maayer P, et al. Metagenomic analysis of the viral community in Namib Desert hypoliths[J]. Environmental Microbiology, 2015, 17 (2): 480—495.
- [33] 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.
- [34] Guerrero L D, Pérez M V, Orellana E, et al. Long-run bacteria-phage coexistence dynamics under natural habitat conditions in an environmental biotechnology system[J]. The ISME Journal, 2021, 15 (3): 636—648.
- [35] Mirzaei M K, Maurice C F. Ménage à trois in the human gut: Interactions between host, bacteria and phages[J]. Nature Reviews Microbiology, 2017, 15 (7): 397—408.
- [36] Bi L, Du S, Yu D T, et al. Compositional and functional characteristics of viruses in soil under two types of land-use in Xinjiang, China[J]. Acta Ecologica Sinica, 2021, 41 (7): 2728—2737. [毕丽, 杜帅, 于丹婷, 等. 新疆两种土地利用方式下土壤病毒的群落组成与功能特征[J]. 生态学报, 2021, 41 (7): 2728—2737.]
- [37] 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.
- [38] 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.
- [39] Mann N H, Cook A, Millard A, et al. Bacterial photosynthesis genes in a virus[J]. Nature, 2003, 424 (6950): 741.
- [40] Suttle C A. Marine viruses—major players in the global ecosystem[J]. Nature Reviews Microbiology, 2007, 5 (10): 801—812.
- [41] Anné J, Van Mellaert L, Decock B, et al. Further biological and molecular characterization of actinophage VWB[J]. Journal of General Microbiology, 1990, 136 (7): 1365—1372.
- [42] Yoshida M, Yoshida-Takashima Y, Nunoura T, et al. Identification and genomic analysis of temperate *Pseudomonas* bacteriophage PstS-1 from the Japan trench at a depth of 7000 M[J]. Research in Microbiology, 2015, 166 (9): 668—676.

(责任编辑: 卢 萍)