微生物学报
Acta Microbiologica Sinica
2022, 62(12): 4663–4683
http://journals.im.ac.cn/actamicrocn
DOI: 10.13343/j.cnki.wsxb.20220661

环境病毒的宿主鉴定技术进展

徐步[1,2]，邹雪蓉[2]，朱元清[2,4]，范陆[2,3*]，张传伦[2,3,4]

1 哈尔滨工业大学环境学院，黑龙江 哈尔滨 150001
2 南方科技大学海洋科学与工程系，深圳海洋地球古菌组学重点实验室，广东 深圳 518055
3 南方海洋科学与工程广东省实验室(广州)，广东 广州 511458
4 上海佘山地球物理国家科学野外观测研究站，上海 200062

徐步, 邹雪蓉, 朱元清, 范陆, 张传伦. 环境病毒的宿主鉴定技术进展. 微生物学报, 2022, 62(12): 4663–4683.
Xu Bu, Zou Xuerong, Zhu Yuanqing, Fan Lu, Zhang Chuanlun. Advances in host prediction approaches for environmental phages. *Acta Microbiologica Sinica*, 2022, 62(12): 4663–4683.

摘　要：病毒是地球上丰度最高的微小生命粒子，通过调控宿主的群落结构、介导宿主死亡和参与水平基因转移等方式影响着生物地球化学循环和地球生命演化。近年来，宏基因组学的发展实现了在全球尺度上对环境病毒的大规模探索和研究，大量新的病毒基因组被发掘，病毒在全球生态过程和生物地球化学循环中的角色和贡献也得到进一步认知。病毒在环境中的重要作用是通过感染宿主实现的。然而，环境病毒的宿主鉴定工作远落后于环境病毒基因组测序研究。本文综述了目前病毒宿主鉴定的主要技术及其优缺点和应用场景，总结了病毒的宿主鉴定在病毒生态学研究和生物工程领域的重要价值，并初步展望了未来病毒宿主鉴定技术的发展方向。

关键词：病毒；宿主鉴定；感染关系

基金项目：国家自然科学基金(91951120，91851210，42141003)；国家重点研发计划(2018YFA0605802)；深圳市科技创新委员会(ZDSYS201802081843490)；南方海洋科学与工程广东省实验室(广州) (K19313901)；上海佘山地球科学国家野外科学观测研究站(2020Z01)

Supported by the National Natural Science Foundation of China (91951120, 91851210, 42141003), by the National Key Research and Development Program of China (2018YFA0605802), by the Shenzhen Science and Technology Innovation Commission (ZDSYS201802081843490), by the Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) (K19313901) and by the Shanghai Sheshan National Geophysical Observatory (2020Z01)

***Corresponding author.** Tel/Fax: +86-755-88011400; E-mail: fanl@sustech.edu.cn

Received: 4 September 2022; **Revised:** 30 October 2022; **Published online:** 08 November 2022

Advances in host prediction approaches for environmental phages

XU Bu[1,2], ZOU Xuerong[2], ZHU Yuanqing[2,4], FAN Lu[2,3*], ZHANG Chuanlun[2,3,4]

1 School of Environment, Harbin Institute of Technology, Harbin 150001, Heilongjiang, China
2 Shenzhen Key Laboratory of Marine Archaea Geo-Omics, Department of Ocean Science and Engineering, Southern University of Science and Technology (SUSTech), Shenzhen 518055, Guangdong, China
3 Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou 511458, Guangdong, China
4 Shanghai Sheshan National Geophysical Observatory, Shanghai 200062, China

Abstract: Viruses are the most abundant biological entities on Earth. They play an important role in biogeochemical cycles and the evolution of life by regulating host community structure, causing mortality, and mediating genetic exchange. In recent years, the development of metagenomics has enabled the global-scale study of environmental viruses. A large number of novel viral genomes have been uncovered, and the roles and contributions of viruses in ecological processes and biogeochemical cycles have been recognized. The important role of viruses in ecosystem is mainly dependent on the infecting host. However, host prediction of environmental viruses has lagged far behind the studies of environmental viral genomics. This review aims to provide the latest knowledge of the main approaches for virus-host prediction, their pros and cons, and application, to reveal the importance of host prediction in the ecological research and bioengineering of viruses, and to present an outlook on the future development of host prediction approaches.

Keywords: virus; host prediction; infectious relationship

病毒是地球生物圈中丰度最高的生命形式，总数量可达 10^{31} 个，超过宿主微生物丰度的十倍，是地球上最丰富的微生物资源之一[1–2]。病毒在环境中不能单独存活，需要严格寄生在活的宿主细胞内进行增殖，其生活方式可大致分为裂解性感染(lytic)、溶源性感染(lysogenic)和慢性感染(chronic)[3]。病毒因其极高的丰度、丰富的多样性和独特的生活方式在地球圈的生态过程和生物地球化学循环中发挥着重要的作用[4–7]，主要表现在：(1) 病毒是环境微生物的主要致死因子，海洋中 40%–60%的微生物死亡是由病毒裂解导致的，是环境微生物生物量和群落的两大下行控制因素之一[4]；(2) 病毒在感染过程中能够重构宿主的代谢方式，改变宿主在生物地球化学循环中的角色；(3) 病毒参与遗传物质在微生物群落中水平传递，对宿主的环境适应和演化产生重大的影响[8]；(4) 病毒裂解微生物以及病毒自身降解所释放的有机物质重新进入环境，是生态系统的初级生产力和次级生产力的重要上行控制因素[9–10]，影响着海洋生态系统碳循环总量的 6%–26%[11]。

近年来，宏基因组技术为环境病毒研究带来了革命性的突破。宏基因组技术不依赖于微生物培养，直接采集环境样品，进行基因组提取和高通量测序，最后通过生物信息学方法重构环境病毒基因组[12]。这一方法极大地拓展了环境病毒学研究的广度和深度。随着环境病毒宏基因组研究如雨后春笋般的涌现，目前已发掘的环境病毒基因组的数目已非常可观，人们对环境病毒的分布、遗传多样性和生态功能等

也有了更深入的了解[13−14]。

宏基因组技术帮助发现海量病毒基因组的同时，也带来了病毒宿主无法确定的难题。由于环境病毒宏基因组技术获得的结果缺乏与宿主关联的直接证据，绝大部分病毒的宿主无法确定，导致未培养病毒的宿主鉴定工作远滞后于病毒多样性研究。目前公共数据库中 95%以上的病毒属于未培养病毒，其中大部分病毒的形态特征、宿主及与宿主的相互作用仍是一个“黑箱”[15]。随着病毒和微生物数据库的日益丰富、生物信息学分析方法的开发，以及先进仪器设备逐步应用到环境病毒学的研究中，病毒宿主鉴定也取得了一些突破性的进展[16]。本文系统梳理了目前主要的环境病毒宿主鉴定技术，以时间为主线概述了这些技术的发展历程(图 1)。基于技术的原理，将环境病毒宿主鉴定技术划分为 5 大类：序列相似性原则、共演化历史信息、病毒和宿主的丰度分布模式、机器学习方法和病毒宿主物理接触信息(表 1)。此外，本文还详细地描述了环境病毒鉴定技术的原理、主要优缺点和应用场景，讨论了这些技术在生态学研究和环境工程领域的重大意义，并展望了未来病毒宿主鉴定技术的发展趋势。

1 序列相似性原则

序列相似性原则是病毒宿主鉴定的重要依据之一。一方面，病毒和宿主之间的同源基因可以指征病毒对宿主的感染关系。另一方面，与已知宿主的病毒或前噬菌体(prophage)的基因组相似性常被用于指导未培养病毒的宿主预测。BLAST[17]和 Diamond[18]是进行相似性计算的常用软件。

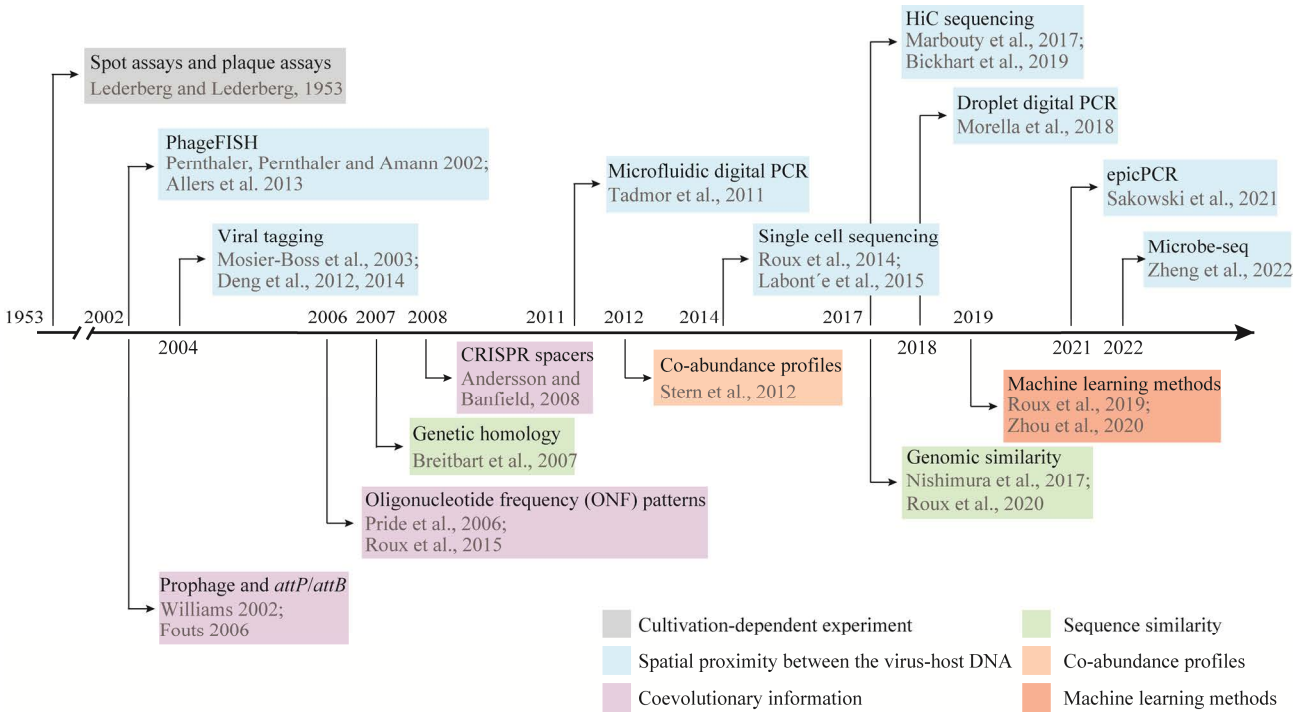

图 1 病毒宿主鉴定技术发展历程

Figure 1 The timeline of the current major approaches for virus-host prediction.

表 1　病毒宿主预测方法概述

Table 1　A summary of representative technologies for virus-host prediction

Categories	Methods	Principle	Main advantages	Main limitations	Tools/Pipeline
Sequence similarity	Gene homology	During historical infection, the horizon gene transform (HGT) between viral and host genomes. Conserved signature genes between viruses infecting the same host	High accuracy	This method depends on a reference database and frequently HGT may result in genetic homology signals useless	BLAST, Diamond BLAST, phylogenetic tree
	Genomic similarity	The genomic similarity of viruses infecting similar hosts or genomic similarity of viruses and their hosts	High accuracy	It is database-dependent. The complement of the viral genome significantly impacts its performance	BLASTN, tBLASTx
Co-existence	Co-abundance profiles	Viruses lyse host cells to produce viral progeny. The abundances of viruses and their hosts in environment often corelate	Database-independent. Increasing number of metagenomics sequencing projects may improve the performance	Complex virus-host interactions and ecological processes have a significant impact on this approach	Bowtie, bwa
Coevolutionary information	CRISPR spacer matching	An acquired immune mechanism evolved by prokaryotic microorganisms to defend against exogenous genetic invasion represented by viruses	High accuracy, high specificity, up to the strain level	High false negatives because only ~40% of bacteria and ~70% of archaea encode a CRISPR system	BLASTN, SpacePHARER, and CRISPRDetect3
	Temperate phages	The full length of prophage can be identified from their host genome or the exact match of the homologous recombination sites	The longest exact match (prophage) with high accuracy	The short exact match (homologous recombination sites) may lead to false positives	Prophage identification tools *attB* and *attP* sites identification and exact match by the BLAST
	Oligonucleotide frequency (ONF) patterns	Adaptation of viruses to host replication, transcription and translation machinery, and evolutionary pressure to avoid recognition by host restriction endonucleases	Flexible and convenient, database-independent, good performance in host prediction of the fragmented viral genome	High rate of false positives	VirHostMatcher, HostPhinder

(待续)

(续表 1)

Categories	Methods	Principle	Main advantages	Main limitations	Tools/pipeline
Machine learning methods	Random forest classifiers; Gaussian model; neighborhood regularized; logistic matrix factorization; deep convolutional neural network; graph convolutional network (GCN)	Learning from features of existing virus-host pairs and building a virus-host prediction classifier	Good scalability, flexibility, and convenience	The challenge of machine learning lies in building a robust classifier that covers as many diverse viruses, host taxa, and virus-host interactions as possible to avoid over-prediction Need further revaluation	RaFAH, HostG
Spatial proximity	Single-cell sequencing	Based on the spatial proximity of the viruses and the host during infection	Can identify infection strategies of viruses and obtain viral and host genome sequences	High cost	Flow cytometry and multiple displacement amplification (MDA)
	Viral tagging		High specificity than single-cell sequencing	Adherent rather than infection may cause false positives	Flow cytometry and multiple displacement amplification (MDA)
	PhageFISH		With a certain specificity, and can be used to study viral infection strategies	The limitations of probes and can't obtain genomic information	FISH with phage probes
	Microfluidic digital PCR and Droplet digital PCR		With a certain specificity, and can be used to study viral infection strategies	The limitations of primers and can't obtain genomic information	Single-cell separates and PCR
	epicPCR		Higher throughput than microfluidic digital PCR and droplet digital PCR	The limitations of primers and can't obtain genomic information	Single-cell separates and PCR
	HiC sequencing		High throughput, ability to identify viral infection strategies, and cross-species infections	High demand for experimental operation and bioinformatic skills	HiC protocol and bioinformatics analysis
	Microbe-Seq		High throughput, ability to identify viral infection strategies, and cross-species infections	High demand for experimental operation and bioinformatic skills	Microbe-Seq protocol and bioinformatics analysis

1.1 同源基因

同源基因是一类具有共同祖先来源的基因，现代分子生物学将同源基因分为直向同源基因(orthologous gene)、横向同源基因(paralogous gene)和异源同源基因(xenologous gene)[19–20]。其中部分病毒和宿主之间的异源同源基因和病毒与病毒之间的直向同源基因具有被用于病毒宿主鉴定的潜力。病毒感染宿主的过程可能会导致基因元件在病毒和宿主基因组之间水平转移，其中部分基因被长期稳定地保存下来。辅助代谢基因(auxiliary metabolic genes，AMGs)就是一类常见的异源同源基因，它们能在感染期间表达，以重新引导宿主细胞的能量和资源用于病毒生产，是病毒参与物质代谢和元素循环的重要方式之一[6]。同时，部分具有相同宿主类群的病毒之间也保有直向同源基因，可用于确定系统发育关系。通过匹配未培养病毒和宿主或已知宿主的病毒之间同源基因可以用来

重建病毒与宿主之间的感染关系。

Philosof 等基于主要衣壳蛋白(major capsidprotein，MCP)和 DNA 聚合酶 B 家族蛋白(DNA polymerase B，DNApolB)的序列相似性，从 *Tara* Oceans 数据集中获得了 26 个感染海洋浮游古菌 MGII 的病毒[21]。Nishimura 等的研究结果表明，MGII 病毒与宿主 MGII 的 DNApolB 的平均氨基酸相似度为 36.1%[22]。本文作者正在进行的研究对全球河口环境的 MGII 病毒进行了调查，利用 MGII 病毒的 MCP 基因和 DNApolB 基因系统发育分析从全球河口环境中获得了 359 个非冗余 MGII 病毒基因组。系统发育结果显示，新发现的 MGII 病毒与已发表 MGII 病毒参考基因组聚在同一进化支上，验证了本研究的结果的可靠性，也表明利用 MCP 和 DNApolB 基因相似性可以用于环境未培养 MGII 病毒的挖掘(数据未发表)。Ahlgren 等在海洋奇古菌病毒基因组上发现了与宿主同源的参与氮代谢的氨单加氧酶基因 *amoC*，系统发育分析揭示这些病毒 *amoC* 基因与奇古菌 *amoC* 基因在演化树上形成紧密相邻的姊妹演化支，这暗示 *amoC* 基因可能可以用于鉴定环境未培养奇古菌病毒[23]。此外，光合系统相关的基因也被广泛用于鉴定未培养蓝细菌病毒，如 *psbA* 和 *psbD*[24–25]。Dutilh 等报道了一系列的保守基因可用于 CrAssphage 的鉴定和宿主预测[26]。

序列同源比对一直是非常有效的病毒宿主预测方法之一。Edwards 等利用测试数据对此方法进行了验证，研究表明同源比对方法能够在物种水平准确地鉴定出超过 30%的宿主[27]。值得注意的是，并不是所有的同源基因都能够用来进行宿主预测，如磷酸盐调节基因 *phoH*。这可能与基因的演化速率或者在物种之间频繁地水平转移有关。此外，基于核苷酸序列的同源比对较基于氨基酸序列具有更高的准确度和更低的假阳性率，这可能是因为同源的氨基酸序列覆盖了更加久远的演化历史，而病毒与宿主的感染关系可能并不能维持那么久[27]。病毒和宿主的同源基因的系统发育关系是判断同源基因是否可以被用于病毒宿主鉴定的重要参考依据。

1.2 长片段核苷酸序列相似性

与已知宿主的病毒的基因组相似性常被用于未培养病毒的物种分类和宿主预测，一般认为基因组相似性高的病毒之间往往亲缘关系更近，也更可能感染同一类群的宿主。Nishmura 等使用 tBLASTx 计算了不同病毒之间基因组相似性，发现感染相同宿主的病毒之间的基因组相似性要显著高于感染不同宿主的病毒[22]。由此得出，与已知宿主的病毒的基因组相似性可以用来指导未培养病毒的宿主预测，且当相似性得分超过一定阈值时，在属水平上进行宿主预测的准确率可达到 90%。他们利用该方法从 *Tara* Oceans 数据集中成功鉴定出 18 个蓝细菌病毒、8 个远洋杆菌病毒和 1 个假单胞菌病毒[22]。

Roux 等也利用病毒和宿主之间长片段序列(≥2 kb)的相似性(≥90%)特征来鉴定未培养病毒的宿主，并用 NCBI 病毒参考基因组数据库对方法进行了验证。结果表明该方法在目、科和属水平的宿主预测准确率分别为 96.2%、95.3%和 91.2%[28]。随后，Roux 等将这种方法与 CRISPR spacers 序列精确匹配方法联用对来自 IMG/VR 数据库中的未培养病毒进行了宿主预测，成功地对 15.3%的未培养病毒进行了准确的宿主预测，其中 85.8%可以预测到属水平[28]。值得注意的是，这类预测信号可能主要来源于宿主基因组上的前噬菌体。此外，病毒基因组的完整度对宿主预测的准确性具有很大的影响，完整度大于 50%的病毒基因组能够更准确

地预测到宿主[28]。最后，由于已分离病毒参考基因组库的极不完善，研究发现环境未培养病毒基因组与参考病毒基因组之间存在很大的差异。因此，基于与参考病毒基因组的相似性比较，只有少部分的未培养病毒能够成功地鉴定出宿主[15,22]。

2 共演化历史信息

病毒的生活方式决定着病毒及其宿主的主要相互作用为病毒对宿主的感染和宿主对病毒防御，两者的相互作用呈现为“军备竞赛式”的持续共演化现象。在长期的共演化历史中，部分表征病毒和宿主感染关系的证据被“写”进各自的基因组。主要表现为：(1) 宿主的规律成簇间隔的短回文重复序列(clustered regularly interspaced short palindromic repeats，CRISPR)中记录着感染它的病毒的遗传信息；(2) 病毒将自己的 DNA 整合到宿主的基因组上；(3) 病毒对宿主遗传、代谢和防御机制的演化性适应在其基因组上的体现。这些病毒–宿主相互作用的信息可以用于病毒的宿主鉴定。

2.1 宿主的免疫防御机制

CRISPR 系统是细菌和古菌为了防御以病毒为代表的外源可移动基因元件入侵而演化出的一类获得性免疫机制。CRISPR 系统由规律成簇的间隔短回文重复序列 CRISPR、一段或若干段长 25–75 个碱基的外源特异性间隔序列(spacers)和一组具有剪切功能的蛋白组成[27,29]。CRISPR 系统对病毒的防御机制如下：初次感染时细菌和古菌的 CRISPR 系统通过识别病毒的特异性序列，作为 spacers 序列将其记录到 CRISPR 中，当再次被同类型的病毒感染时，细菌和古菌通过匹配 spacers 序列来识别病毒，并激发核酸酶等防御工具对病毒核酸进行剪切，使外源病毒核酸断裂，从而保护细胞抵御病毒的侵害[30]。基于这一原理，在环境病毒组学研究中，通过识别宿主的 CRISPR 系统和精确匹配病毒和宿主的 spacers 序列可以重建病毒和宿主之间的感染关系[27]。SpacePHARER[31]和 BLASTN[17]、CRISPRDetect[32]等软件被用来搜索和精确匹配病毒和宿主之间的 spacers 序列[16]。

此方法已经被广泛应用于海洋[33]、土壤[34]、淡水湖泊[35]、极地冰川[36]、动物[37]和人体[26]微生物组等环境病毒的宿主鉴定。本文作者利用此方法从珠江口淡咸水环境中鉴定出多对病毒和宿主的感染关系，其宿主涵盖了变形菌门、拟杆菌门、蛭弧菌门和浮霉菌门微生物。研究还发现了多种病毒对同一宿主的共感染现象[38]。一般来说，与其他方法相比，基于 CRISPR spacers 序列匹配的方法具有更高的特异性，能够在菌株水平鉴定病毒的宿主。当病毒和宿主之间有着多个匹配的 spacers 区域时，鉴定结果更可信[27]。然而，研究表明仅有约 40%的细菌和 70%的古菌具有 CRISPR 系统，且不同类群差异很大，导致此方法的假阴性率相对较高[27]。在实际应用中，由于 CRISPR 间隔序列比较短，为了降低假阳性概率，一般仅允许 1–2 个碱基错配[16]。此外，许多 CRISPR 间隔序列在病毒基因组上找不到同源序列，这表明仍有大部分环境病毒待挖掘。

2.2 溶源性感染特征

病毒的溶源性感染也为病毒的宿主鉴定提供了非常重要的信息。溶源性感染是指病毒感染宿主之后，暂时不进行子代病毒的繁殖和裂解宿主，而是将其基因组整合到宿主基因组上，随着宿主的繁殖而被保存下来，这类病毒的存在形式也被称为前噬菌体。病毒的溶源性感染在环境中普遍存在，研究表明，近一半的原核微生物基因组含有前噬菌体[39–40]。宿主基因组上的前噬菌体基因组可能是完整的，也可能随

着漫长的演化部分丢失成为基因片段。这些留存的证据可以很好地反映病毒对宿主的历史感染事件和潜在感染能力，因此也被用于病毒的宿主鉴定。通过对原核微生物基因组进行前噬菌体预测一方面可以重建病毒对该原核微生物宿主的感染关系，同时这些前噬菌体基因组也可以作为数据库，用于环境未培养病毒的宿主鉴定。

此外，溶源性病毒一般带有整合酶、切除酶和阻遏蛋白，其整合酶基因附近有整合位点 *attP*，而对应其宿主基因组的 tRNA 基因内部或者附近区域则有与之精确匹配的整合位点 *attB*[27]。因此，通过搜索和匹配病毒和原核微生物基因组上的整合位点，能够进行未培养病毒的宿主鉴定。

Mizuno 等利用这种方法对 2 株海洋未培养蓝细菌病毒进行了宿主鉴定，它们的 *attP* 区域分别与原绿球藻 (*Prochlorococcus marinus*) MED4 和聚球藻(*Synechoccus*) CC9605 的 *attB* 区域高度匹配，同源基因 *psb*A 基因的系统发育分析也支持预测的结果[41]。需要强调的是，tRNA 基因在亲缘关系近的原核生物中具有高度保守性，基于这种方法的宿主预测可能只能预测到一个较高的分类单元，即门或类[27]。此外，由于 tRNA 和整合位点序列较短，因此 *attP* 和 *attB* 可能因随机匹配而造成假阳性预测[27]。与 CRISPR 精确匹配方法一样，此方法也是基于短序列片段的精确匹配，因此应当用严格的过滤标准，整个序列最多只允许 1–2 个碱基错配[16]。结合长片段前噬菌体序列的比对，CRISPR spacers 序列和同源基因比对等能够提高宿主预测的准确度和获得更加精确的预测宿主。

2.3 病毒和宿主之间的寡核苷酸频率使用相似性

与其他方法不同，基于寡核苷酸频率相似性进行病毒宿主预测方法不依赖于序列比对，而是根据病毒和原核微生物基因组的序列特征来预测病毒的宿主。寡核苷酸序列是指微生物基因组上连续的核苷酸短序列，又被称为 kmer，其中 k 表示短序列的长度。在演化过程中，如果序列越短，那么被改变的几率也就越小，因此寡核苷酸序列频率被认为是一种保守的遗传标记。寡核苷酸序列频率被广泛应用于原核微生物基因组序列拼接和分箱、物种注释和环境病毒基因组鉴定等。

直到近年来，寡核苷酸使用频率才被发现可应用于未培养病毒的宿主预测。Roux 等首次系统地探究了寡核苷酸序列频率相似性在病毒宿主预测方面的应用潜力。研究比较了病毒与宿主和非宿主之间不同长度的寡核苷酸频率和密码子使用情况的差异，结果表明病毒与其宿主之间的差异小于与非宿主的差异，其中四核苷酸频率使得宿主和非宿主之间的差异最大[42]。关于病毒与其宿主之间相似的寡核苷酸利用频率的原因尚未有定论，推测可能与病毒对宿主复制、转录和翻译机器的适应以及避免被宿主限制性内切酶识别的演化压力有关[43–44]。

基于病毒和宿主之间寡核苷酸频率相似性的特点，许多病毒宿主预测软件被开发出来，PHP[45]和 VirHostMatcher[46]通过识别病毒和原核微生物之间 4-mer 和 6-mer 频率的最大相似性来预测病毒的潜在宿主。ILMF-VH[47]和 HostPhinder[48]通过比较未培养病毒与已知宿主的病毒之间 6-mer 和 16-mer 最大相似性来进行宿主预测。WISH 软件[49]则将原核微生物基因组 8-mer 的序列构建成同源的隐马尔可夫模型，通过搜索隐马尔可夫模型来进行宿主预测。Ahlgren 等用 VirHostMatcher 从海洋宏基因组数据中发掘出 15 个奇古菌病毒基因组，其基因组上编码了与宿主同源的 *amoC* 基因，证明了宿

主预测的准确性[23]。

寡核苷酸频率相似性的使用限制较少，且不需要考虑基因组序列的编码区和非编码区，因此具有灵活方便的优点，尤其是在预测不完整病毒基因组时有着不俗的表现。不可忽视的是，研究者经常发现以 ssDNA 为代表的部分病毒采取和宿主不同的寡核苷酸使用模式，这可能与病毒所采取的感染策略有关，如慢性感染。此外，一些噬菌体可能通过使用自身携带的 tRNA 基因来改变宿主的密码子使用，从而不需要演化出与宿主相同寡核苷酸使用模式[42]。这些因素可能使得此方法在实际应用中产生假阴性结果。最后，不同长度 kmer 在宿主预测方面的表现也存在差异，较短的 kmer 可能造成随机的相似性，导致假阳性预测；而 kmer 越长其在病毒和宿主基因组上出现的频率也越低，造成假阴性预测[27]。因此，一些软件结合寡核苷酸使用频率相似性和其他宿主预测方法一起使用以最大程度地提高病毒宿主预测的召回率和准确性，如 PHISDetector[50]和 VirHostMatcher-Net[51]等。

3 病毒和宿主在环境中的丰度分布模式

病毒不具备完整的细胞结构，必须依赖活性宿主细胞才能完成生活史，因此病毒的丰度在一定程度上受宿主丰度的影响。研究表明，环境病毒及其宿主在丰度分布模式上往往呈正相关，感染相同宿主的病毒也往往因为对宿主的竞争关系而具有相同的丰度分布模式[52]。因此，病毒和宿主在时间和空间上的丰度分布模式能够用来预测病毒和宿主之间的感染关系。

得益于宏基因组学的发展，测序数据可以反映出病毒和宿主的相对丰度，结合统计学方法来能够重建病毒和宿主之间的感染关系。其分析思路一般为：通过 BLAST[17]或者序列回帖软件[53−54]将宏基因组测序的短序列回帖到病毒和宿主基因组上以计算基因组相对丰度，最后通过计算病毒和宿主或竞争者在时间序列或空间序列宏基因组样本中的丰度相关性和共现性来确定潜在的感染关系。一般认为丰度分布模式呈正相关的病毒和宿主具有感染关系，呈正相关的不同病毒之间具有属分类水平上的共同宿主。

目前病毒和宿主的丰度相关性已被应用于海洋水体环境[52,55]和肠道环境[26,56]中的病毒宿主预测。Lima-Mendez 等对 *Tara* Oceans 宏基因组样本中病毒和宿主的丰度分布模式进行了研究，结果表明未培养病毒与环境中优势的 7 个细菌门类和 1 个广古菌门类存在 1 869 对正相关关系。其中 8 对预测的病毒与宿主的感染关系被公共数据库确认。共现性网络结果表明，约 43%的病毒仅有单一的宿主，其余病毒的宿主范围也相对较窄[55]。Dutilh 等也利用这种方法从 151 个人体微生物宏基因组样本中预测出 crAssphage 类噬菌体的潜在宿主为拟杆菌属细菌[26]。Coutinho 等通过计算来自于 *Tara* Oceans 宏基因组样本中未培养病毒基因组和已知宿主的参考病毒基因组的丰度相关性来鉴定未培养病毒的潜在宿主，对 1 279 个未培养病毒进行了宿主预测，结果表明这些病毒主要感染海洋蓝细菌和远洋杆菌[52]。

基于丰度相关性的环境病毒宿主预测不依赖于数据库，且避免了病毒和宿主基因组上与感染关系相关的遗传信号缺失的问题。此外，大量基于宏基因组测序的环境微生物调查为此方法的准确率和灵敏度提供了重要的保障，尤其是日益丰富的基于时间序列和空间序列的宏基因组学研究[27]。

然而由于多方面的因素，使得此方法存在争议[57]。首先，环境中病毒和宿主的丰度是动

态变化的。病毒与宿主的丰度比值(virus to microbe ratio，VMR)常用来表征病毒对宿主的感染率。VMR 受多种因素的影响，如宿主的活性、病毒的降解率、病毒的生存策略和环境因素等等。这些因素可能因时空条件的不同而发生变化，从而导致病毒及其宿主的丰度相关性往往呈非线性关系，造成假阳性预测。

其次，最近提出来的病毒和宿主群落相互作用理论——“搭乘胜利者”假说(piggyback-the-winner，PtW)有力地冲击了传统“杀死胜利者”假说(kill-the-winner，KtW)，认为当环境宿主丰度越高的时候，病毒的生活策略往往由裂解性感染转变为溶源性感染[10,58–59]。这导致环境中病毒的丰度并不会随着宿主丰度的增加而增加，相反往往出现减少的现象，从而导致病毒丰度和宿主丰度出现负相关[60]。在环境中病毒和宿主的相互作用主要受 PtW 假说影响还是由 KtW 假说影响及其驱动因素等尚未有研究结果，由此也对此方法的实际应用提出了极大的挑战。

最后，环境样本的采集和处理方法、环境 DNA 测序是否有经过多重置换扩增(multiple displacement amplification，MDA)，以及病毒和宿主的样品是分开采集测序还是混合采集测序等等，都对病毒和宿主的丰度计算产生很大的影响。相关性和共现性关系的统计学计算方法是否能够可靠地反映病毒和宿主之间的相互作用也备受质疑。Coenen 等对几种相关性方法进行了测试，发现目前所有的方法都不能完全准确地预测环境微生物之间真实的相互作用，即使在改变网络结构、网络大小、初始条件扰动程度和采样频率等条件下表现仍然不佳[61]。

综合为数不多的研究表明，虽然目前基于丰度分布模式的方法对未培养病毒的宿主预测能力有限，然而随着环境微生物学研究的推进，对环境病毒和宿主相互作用的生态效应的理解加深以及丰度分布模式的生物信息学方法的加强，或可很大程度地提升丰度分布模式方法的预测能力，对发掘环境中更多病毒和宿主之间的相互作用也具有巨大的潜力。

4 机器学习在病毒宿主预测领域的应用

基于计算方法的宿主预测在实际应用中仍存在一些问题。首先，已知的病毒和宿主关系仍然非常有限。其次，无论是基于保守基因还是基于 CRISPR spacers 等特征序列，病毒和宿主之间序列比对往往存在结果缺失或者模糊的情况，影响宿主预测的准确性。近年来，机器学习的方法逐渐被应用于病毒宿主预测，以弥补传统计算方法的不足。

严格意义上的机器学习方法并不是一种单独的方法，它依赖于已知的病毒宿主之间的感染关系，并通过机器学习工具对知识进行分类学习和模型训练。这种模型一旦训练完成，可以直接用于环境未培养病毒的宿主鉴定，省去大量的计算成本。用户也可以基于最新的研究成果自主地对模型进行补充训练，或者训练特异性的模型用于对特定病毒物种进行宿主预测，是一种非常灵活方便的方法。目前主要的机器学习工具主要包括，包括随机森林分类器[62–63]、高斯模型[45]、邻域正则化逻辑矩阵分解模型[47]和深度卷积神经网络[64]和图卷积网络[65]等。

Roux 等利用丝状噬菌体科(*Inoviridae*)病毒的基因组特征和 ATP 酶基因构建了随机森林分类器。该研究从各类环境微生物基因组和宏基因组样本的拼接基因组中识别出 10 295 个丝状病毒基因组，其宿主包括细菌和古菌[62]。Coutinho 等开发了机器学习软件 RaFAH 来进行病毒宿主预测，他们首先将所有已知宿主的病

毒的蛋白序列聚类成蛋白簇，然后构建具有最佳皮尔逊相关系数的病毒基因组和蛋白簇之间的矩阵，这个矩阵被用来构建随机森林分类器，并通过测试数据对分类器进行调整和测试[63]。RaFAH在环境宏基因组的病毒宿主预测有着良好的表现，尤其是在古菌病毒的预测中表现不俗，Coutinho等利用RaFAH从环境宏基因组拼接序列中鉴定出537个古菌病毒基因组，其中仅有97个在公共数据库中有记录[63]。Shang等提出了一种半监督学习模型的病毒宿主预测方法HostG，利用病毒氨基酸序列相似性和病毒与宿主DNA序列相似性构建知识图谱，并通过图卷积网络(graph convolutional network，GCN)同时对已知宿主和未知宿主的病毒进行训练。此外，分类器的图卷积网络还能够拓展学习用户提供的知识图谱，对新的宿主类群进行病毒预测。这种方法在应对日益丰富的原核微生物基因组库和分类框架有着很强的适应性[65]。

机器学习方法在病毒预测领域具有不错的表现，极大地提高了未知病毒基因组的发掘和宿主预测。然而，机器学习的核心在于建立一个稳健的训练器，需要尽可能地覆盖多样化的病毒和宿主类群和病毒宿主感染关系，以免高估方法的性能或者造成过度预测[16]。随着对病毒和原核微生物基因组的挖掘以及病毒和宿主之间相互作用认知的加深，更多有用的信息能够被用于机器学习分类器的训练，也提高了方法预测的准确率。

5 病毒宿主物理接触信息

生物信息学方法的病毒宿主预测是一种基于经验的间接预测方法。这些方法仍存在一定的不足，主要表现在以下几个方面。首先基于经验的方法可能存在假阳性，虽然多方法联用能够降低假阳性概率，但这也意味着病毒及其宿主基因组上需要有多种能够反映感染关系的遗传信号或者序列特征，然而在实际应用中往往很难实现。其次，由于环境病毒多样性极高，大量未知病毒仍有待挖掘，关于病毒和宿主之间的相互作用的研究仍非常少，因此生物信息学方法的应用也极大地受到了限制，目前公共数据库中84.7%的病毒未能匹配到任何可能的宿主信息[28]。最后，病毒和宿主之间的感染关系往往因长时间的演化或者环境因素影响而发生改变，而记录在基因组上的遗传信号很可能反应了两者的感染历史，但并不一定能反映当前生态系统中病毒和宿主真实的相互作用。

近年来，一类基于病毒宿主物理接触信息的方法逐渐被应用于环境中未培养病毒的宿主预测。这类方法利用病毒感染宿主过程中两者核酸在空间距离上接近的原理，通过特定技术实时捕捉病毒与宿主的物理接触，并还原病毒和宿主感染关系。这类方法最大的优势在于不依赖病毒和宿主数据库以及病毒和宿主的序列同源性等特征，具有独立鉴定环境大量病毒宿主的巨大潜力。目前基于病毒宿主物理接触信息的方法主要有单细胞测序技术、病毒标签、病毒荧光原位杂交、微流控数字PCR、液滴数字PCR、细胞内融合基因技术、Microbe-Seq技术和HiC技术等。

5.1 单细胞测序技术(single cell sequencing)

单细胞测序技术是指在单细胞水平上，对细胞内基因组或转录组进行扩增和测序的技术。与宏基因组技术相比，单细胞测序技术在研究种群内部泛基因组学和物种间相互作用等方面具有非常突出的优势。由于单细胞测序技术是对单个细胞内所有核酸进行总体研究，因此感染该细胞的病毒也能够被检测到。近年来，单细胞技术被广泛应用于环境微生物学研究中，病毒及其宿主的相互作用是其中的研究热点之一，在海洋、热泉和肠道等环境的病毒生

态学研究中取得了一系列进展。

Roux 等利用单细胞测序技术从海洋低氧区获得了 127 个未培养硫氧化细菌 SUP05 基因组和 69 个病毒基因组，其中三分之一的宿主细胞被病毒感染。研究还率先报道了双链 DNA 病毒和单链 DNA 病毒对同一宿主的共感染现象[66]。Labonté 等的研究发现在表层海洋中，被病毒感染的细胞约占总分选细胞的 34.5%，研究还首次报道了感染海洋奇古菌门、海微菌门、疣微菌门和变形菌门等微生物的病毒[67]。Jarett 等的研究也发现病毒感染广泛发生在热泉环境中，他们从热泉环境中分选出 130 个微生物单细胞，测序发现其中四分之三的细胞至少被一种病毒感染。结合宏基因组学分析发现这些病毒复制活动并不活跃，表明这些病毒可能采用溶源性感染的生活方式，采用 PtW 的群落生态策略[68]。

单细胞测序技术也存在着一定的缺点。单细胞分选技术能够从极少样品中分离出单个微生物细胞，这也决定了若要尽可能全面地对环境中病毒和宿主包括低丰度类群进行调查，需要大量地进行分选、核酸提取、建库和测序，这无疑是一项高成本的工作。

5.2 病毒标签(viral tagging)

病毒标签也是一类基于病毒宿主物理接触信息来进行病毒宿主鉴定的方法。它采用不影响病毒活性的方式对病毒核酸进行荧光标记，并与环境中未知的微生物细胞进行共培养，当病毒感染潜在宿主时，受感染微生物细胞被标记上荧光信号，然后通过荧光显微镜确定感染关系或者利用流式细胞分选技术将目标细胞分选出来。早期的病毒标签主要结合荧光显微镜的检测手段用于搜索环境中特定病毒的宿主[69]。随着流式细胞分选技术的应用，被感染细胞能够被单独分选出来并测序，病毒及其宿主的分类和基因组信息得到确认。目前病毒标签技术已成为研究病毒宿主范围、环境中感染特定宿主的病毒以及环境中病毒和宿主群落感染关系的有力手段[70]。

Deng 等通过模式体系对病毒标签进行了验证，结果表明病毒标签方法得到的结果与空斑实验方法一致[71]。在之后的研究中，Deng 等从太平洋海水环境中鉴定出 107 株感染海洋聚球藻的未培养病毒，其中包括部分已分离的株系。基因组学分析揭示了海洋聚球藻病毒丰富的遗传多样性[72]。Džunková 等利用病毒标签技术对不同人粪便样本中的病毒进行了标记和交叉感染实验，研究不仅发掘出 363 对病毒和宿主感染关系，还揭示了病毒可能不是人类肠道环境中微生物水平基因转移的主要载体[73]。

病毒标签方法在一定程度上改进了单细胞测序技术，提高了方法的特异性和应用场景，但仍然依赖于环境中病毒对宿主细胞的感染行为，因此对环境中高丰度的病毒宿主鉴定效果要优于低丰度病毒。此外，收集和标记病毒的操作可能会对病毒活性造成影响，导致培养过程中感染事件可能低于实际情况。最后，病毒和细胞的物理吸附作用可能导致假阳性，因而产生错误的宿主范围，这一点对高丰度病毒类群的影响尤为大。

5.3 病毒荧光原位杂交(phageFISH)

荧光原位杂交(fluorescence *in situ* hybridization，FISH)是一类利用特定的荧光分子探针与细胞核酸片段杂交并通过荧光显微镜显影的技术。Allers 等在 geneFISH 的基础上开发出 phageFISH，不仅将细胞的检测效率由原来的 40%提高到 92%以上，且能够同时检测胞外游离的病毒和被感染细胞内的病毒基因以及宿主 rRNA[74]。此外，phageFISH 通过量化每个宿主细胞的相对病毒的标记基因拷贝数，从而

推测宿主所处的感染状态，亦可用于推断病毒所采取的感染策略[74]。

phageFISH 能够在原位环境对病毒和宿主之间的感染关系进行群落水平的研究，对环境病毒和宿主种群生态学研究有着重大的推动作用。然而 phageFISH 也存在一定的局限性。首先，phageFISH 依赖引物探针来对病毒和宿主进行杂交，由于病毒并不像原核微生物具有保守的基因片段可用于探针设计，因此在实际应用中受到很大的限制。利用宏基因组学挖掘未培养病毒的可利用引物在一定程度上可以弥补此方法应用上的局限。其次，由于无法在基因组层面对 phageFISH 所鉴定的病毒和宿主进行物种鉴定，因此 phageFISH 对病毒宿主鉴定的分辨率主要取决于病毒和宿主的引物所覆盖的物种类群。

5.4 微流控数字 PCR (microfluidic digital PCR)和液滴数字 PCR (droplet digital PCR)

聚合酶链式反应(polymerase chain reaction，PCR)也被用于病毒宿主鉴定，其中微流控数字 PCR[75]和液滴数字 PCR[76]是 2 类常用的方法。微流控数字 PCR 利用微流控芯片将细胞群分散到微孔板小室中，确保一个小室最多只有一个细胞。对单个小室中的细胞进行裂解后，通过 PCR 技术即可对小室中的 DNA 序列进行扩增和检测。Tadmor 等用不同噬菌体亚类群引物和宿主 16S rRNA 基因引物对微流控小室中的 DNA 模板进行 PCR 扩增，当同时检测到病毒和宿主的信号时，即可确认病毒和宿主的感染关系，而病毒和宿主的物种信息也可用通过测序得到验证[75]。液滴数字 PCR 则将样品中的细胞分散到单独的油包水液滴中，替代了微流控数字 PCR 的微孔板，而细胞裂解和后续的 PCR 反应也都在液滴中进行，在一定程度上提高了检测的通量和降低了检测的成本[76]。

微流控数字 PCR 和液滴数字 PCR 方法是一类快速且高通量的病毒宿主鉴定方法，结合测序结果可以对病毒和宿主进行物种鉴定，提高了病毒宿主鉴定的分辨率。此外，这 2 种方法还能用于测量病毒的暴发量和裂解周期，以及进行病毒和宿主的种群动态研究。不可忽视的是，与 phageFISH 类似，微流控数字 PCR 和液滴数字 PCR 方法同样受限于不同亚类群的病毒是否有可用的引物。此外，细胞的裂解方法和效率也是这 2 种方法不可忽视的关键问题。

5.5 细胞内融合基因技术(emulsion paired isolation-concatenation PCR，epicPCR)

细胞内融合基因技术最早由 Spencer 等提出[77]，用于建立微生物细胞物种分类和功能之间的联系。其原理是利用液滴微流控进行单细胞分离，在液滴内利用功能基因引物和 16S rRNA 基因引物将微生物特定的功能基因和 16S rRNA 基因连接，通过 PCR 和高通量测序确定微生物的物种分类及其潜在代谢功能。细胞内融合基因技术是解决生态系统中“谁做了什么？”——微生物物种分类和生态功能的重大生态问题的重要技术手段之一[77]。2021 年，Sakowski 等用噬菌体核糖核酸还原酶基因(ribonucleotide reductase gene，*rnr*)引物替换功能基因引物，将细胞内融合基因技术应用于病毒宿主鉴定领域，以期解决“谁感染了谁？”的科学问题。其研究结果表明，在切萨皮克湾西岸的一个潮汐河口，拟杆菌门微生物是具有 *rnr* 基因的病毒的主要宿主，其相互作用与环境因素相关[78]。细胞内融合基因技术不仅能够快速且高通量地对原位环境进行病毒宿主鉴定，还在研究病毒宿主范围、揭示影响病毒–宿主相互作用的环境和生态驱动因素等方面有着重要的价值。然而，病毒类群是否具有可用的引物仍是此方法最大的限制因素，此外，缺乏病毒和

宿主的全基因组信息也是所有基于引物PCR或探针 FISH 手段的病毒宿主鉴定方法普遍存在的问题。

5.6 Microbe-Seq 技术

2022 年，Zheng 等提出了一种高通量单细胞基因组测序技术 Microbe-Seq，用于在菌株水平解析微生物基因组[79]。首先，它通过液滴微流控技术将成千上万的微生物单独地包裹在液滴中，实现了高通量的检测。其次，通过特异性标签对液滴所有 DNA 进行标记并混合测序，最后通过特异性标签和优化的基因组组装方法还原菌株水平的物种基因组信息或者微生物之间的相互作用。它同时结合了单细胞测序技术的全基因组测序和微流控技术的高通量的优点，在病毒宿主鉴定方面也有着巨大的潜力。一方面，利用特异性标签同时对液滴内病毒和宿主 DNA 进行标记，避免了病毒引物的限制；另一方面，混合的高通量测序和优化的基因组组装能够获得病毒及其宿主的基因组信息，能够从基因组水平研究病毒和宿主的相互作用。因此，Microbe-Seq 在环境病毒学研究方面具有广阔的前景。

5.7 HiC 技术

HiC (high-through chromosome conformation capture)技术是一种由基因组捕获技术发展而来的高通量三维基因组技术，最早 Lieberman-Aiden 等开发出来[80]，用于研究真核生物核内基因组位点之间的相互作用。其流程主要是利用低浓度的甲醛溶液对空间距离相近的 DNA 序列进行交联固定，再通过限制性内切酶对不同来源的DNA进行酶切，然后利用DNA连接酶对不同来源的 DNA 片段的剪切末端进行联结，在末端补平和连接的步骤中加入生物素标记，后续筛选出有生物素标记的片段直接进行建库测序，最终通过生物信息学计算还原基因组内的空间邻近关系和相互作用。Lieberman-Aiden 等用 HiC 构建了分辨率为 1 兆字节的人类基因组的空间三维图谱，这表明 HiC 技术对于空间距离较近的 DNA 具有很好的识别能力[80]。

近年来，研究者逐渐开始将 HiC 技术应用于研究病毒或质粒与原核微生物宿主的相互作用。其原理是病毒在感染宿主细胞的过程中，会将其基因组注入到宿主细胞内，造成与宿主基因组在空间上的接近，因此可以利用 HiC 技术将病毒和宿主基因组交联在一起，最终通过测序和生物信息学方法还原病毒和宿主之间的感染关系。

目前仅有为数不多的研究利用 HiC 方法对环境病毒进行了宿主鉴定和病毒宿主相互作用研究。Marbouty 等首次利用 HiC 技术建立了小鼠肠道中病毒和微生物的感染关系[81]。Bickhart 等结合三代测序和 HiC 技术从牛瘤胃中鉴定出 188 对病毒和宿主之间的联系，研究还发现病毒和宿主之间有较高的 HiC 连接，暗示这些病毒可能主要选择裂解性感染的生活方式[82]。Marbouty 等利用 HiC 技术从人类肠道样品中发掘出约 6 000 对病毒和宿主对应关系[83]。本课题组也利用 HiC 方法对珠江口水样和海绵共生系统中的病毒进行了宿主鉴定，初步的结果均验证了环境中病毒和宿主复杂的感染关系，而其中的部分感染关系通过同源基因和 CRISPR spacers 比对等方法得到验证。大部分的感染关系未能通过已有的数据库或者生物信息学方法得到验证，这也暗示 HiC 方法在环境未培养病毒的宿主鉴定方面有着不可替代的作用。仅有的研究证明 HiC 技术在环境病毒宿主鉴定和病毒宿主相互作用等研究中具有重大的应用潜力。首先，相较于大部分生物信息学方法，HiC 技术不仅能够在物种水平构建病毒和宿主之间

的对应关系，还能够识别环境中病毒的跨物种感染事件。其次，与基于单细胞分选的单细胞测序技术和病毒标签等方法相比，HiC 技术有着更强的鉴定能力和更高的灵敏度。此外，基于矫正后的 HiC 连接数在预测病毒生活方式和计算裂解性病毒的子代病毒释放量等方面可能有着重要的指示作用，但这还有待于严谨的验证。然而，基于空间距离相近的原则，HiC 方法获得的宿主细胞自身 DNA 的 HiC 连接往往要多于宿主 DNA 与病毒 DNA 的连接，对测序数据的利用率产生影响。

6 病毒宿主预测技术在病毒生态学、演化学以及在生物工程技术中的应用展望

6.1 生命起源和演化

病毒和细胞生命的起源问题一直是生命起源和演化领域悬而未决的重大问题。目前关于病毒的起源主要有 3 种假说[84]，(1) 最早起源假说：病毒起源于细胞形成前早期的遗传物质——RNA；(2) 退化假说：病毒由细胞生命退化而来；(3) 逃逸假说：病毒起源于脱离细胞控制的细胞遗传物质物质，获得部分自主复制能力并成为营寄生生活的生命形式。病毒的起源和演化与它的宿主都息息相关，而且越来越多的证据证明病毒在宿主细胞的遗传和演化过程中起到了非常重要的作用，病毒和宿主之间共演化模式逐渐被广泛接受。虽然病毒学家们试图通过各种方法来阐释病毒和宿主之间的共演化模式和演化历史，但是由于大部分病毒缺乏宿主信息或对感染原核生物的病毒认知的局限性，病毒和宿主之间复杂的相互作用和历史感染信息仍得不到很好的挖掘。因此，加速病毒的宿主鉴定方法和工具的开发以及推进病毒宿主的鉴定工作，有助于帮助我们构建复杂而庞大的病毒和宿主的感染网络，进而为厘清病毒和宿主之间复杂的相互作用和病毒宿主共演化提供重要的数据支撑。

6.2 微生物的环境生态效应

病毒是地球上丰度最高的生命粒子，而原核生物在地球上有着巨大的生物量，两者的相互作用对地球生态系统产生了重大的影响。虽然关于病毒裂解宿主和病毒的辅助代谢所造成的生态环境效应已有许多突破性的研究。然而关于如何量化这一生态效应，尚未有清晰的认知。而厘清病毒和宿主之间的感染关系是阐述其中生态学机制和量化生态效应的重要前提。以海洋中丰度最高的自养古菌类群(奇古菌)及其病毒为例。在深部生物圈，病毒引起的奇古菌裂解占被杀死的微生物总生物量的三分之一，导致全球每年释放约 0.3–0.5 GtC (1 GtC= 10^{15} g 碳)[85]。近年从海洋宏基因组数据中鉴定出的奇古菌病毒基因组的证据表明，奇古菌病毒可能通过其基因组上编码的 *amoC* 基因参与宿主的氮代谢，从而影响全球的氮元素循环[23]。最新关于奇古菌病毒分离株的研究结果表明，采用慢性感染策略的病毒虽然不会裂解宿主细胞，但是会显著抑制宿主的氨氧化功能[86]。总而言之，病毒和宿主之间存在着复杂而多样的相互作用，对生态系统和全球元素循环产生的影响也不同，厘清病毒和宿主之间的感染关系是量化这些生态过程的重要前提，这对病毒的宿主预测技术提出了非常大的需求和挑战。

6.3 环境风险评价和管理

病毒的宿主鉴定在环境风险评价和监控方面也发挥着重要的作用。病毒是地球上最大的基因库之一和作为重要的可移动基因元件，在参与微生物基因水平转移中发挥着重要的作用，尤其是对人类和水产养殖动物影响较大的有害基因，如抗生素抗性基因和毒力基因等。

近年来，越来越多的研究表明病毒参与了抗生素抗性基因在不同微生物类群之间的流转，是抗生素抗性基因传播的关键载体之一，且病毒携带的抗生素抗性基因类型与人类活动密切相关[87–89]。此外，病毒不仅是环境微生物的重要致死因子，也是人和其他高等生物疾病和死亡的重要始作俑者。由于对这些病毒及其宿主的未知性，往往疾病暴发或者导致海洋动物大量死亡，造成巨量经济损失时，才能够确定致病病毒[90]。因此迫切地需要建立健全的环境生态监测体系，对环境病毒和宿主的感染关系有着更加清楚的认知，从而为揭示与环境风险相关的基因在环境中的流转机制提供帮助，也为防范生物灾害和减少经济损失提供重要的数据支撑。

6.4 生物工程技术

随着抗生素导致的环境和健康问题的日益暴露，以及抗生素抗性基因在病原微生物间的流行和抗生素的效用逐步降低。人们开始逐渐寻求新的替代物取代抗生素来进行人和动物病原性疾病的治疗。噬菌体治疗利用了病毒感染和裂解致病细菌来清除感染源的原理，被认为是一种可靠而具有广阔前景的替代品[91–92]。相较于抗生素，噬菌体治疗有着安全、特异性高和获取来源广等优势。由于病原微生物对噬菌体抗性的演化，多噬菌体混合使用的“鸡尾酒疗法”是目前噬菌体治疗的主要策略，病毒的宿主预测和鉴定技术有助于扩充可用于噬菌体治疗的病毒库。基于事实感染的病毒宿主鉴定技术(如病毒标签)可用于病毒的宿主范围和病毒感染裂解效应等方面的研究，为将病毒应用于噬菌体治疗提供重要的参考作用。

7 总结和展望

测序技术的飞速发展揭示了大量环境病毒基因组序列，如何确定这些病毒的宿主是当前病毒学研究的难点，也是今后病毒学研究的重要方向。本文系统综述了目前主要的环境病毒的宿主鉴定方法，包括生物信息学方法(同源比对、共演化历史信息、病毒和宿主的丰度分布模式、机器学习)和基于病毒和宿主物理接触信息的实验方法。总体而言，这些方法各有优势和缺陷，有着不同的应用场景。一般来说，基于同源比对方法适用范围更广，能够鉴定出更多、更准确和更精细的病毒宿主感染关系。其他的生物信息学方法也有着不俗的预测表现，是对同源比对方法的重要补充。基于病毒和宿主物理接触信息的实验方法则是近年迅速发展的一类方法，由低通量鉴定(单细胞测序和病毒标签)发展到高通量鉴定(微流控 PCR 等)，由仅鉴定病毒和宿主物种信息(微流控 PCR 等)到获得病毒宿主的相对完整的基因组(HiC 和 Microbe-Seq 技术)。实时揭示环境中病毒和宿主感染关系是此类方法相较于其他方法最大的优势，这也使得这类方法成为今后研究环境病毒相互作用及生态效应重要的方法之一。

虽然环境病毒的宿主鉴定技术已经取得蓬勃的发展，不可忽视的是环境病毒的宿主鉴定技术仍存在很多不足，在此我们也提出今后环境病毒的宿主鉴定技术发展的一些建议：(1) 传统的分离培养技术是目前全面研究病毒生理生化特性和研究病毒宿主相互作用的主要手段，在环境病毒宿主鉴定方面仍有强大的生命力。利用现今先进的技术手段(高通量培养、基因组学指导等)加强或改进传统分离培养手段以提高病毒宿主鉴定效率，是今后环境病毒的宿主预测技术发展的重要目标之一。(2) 目前几乎所有的病毒宿主预测方法都无法避免假阳性预测。多方法联用或补充验证能够有效降低假阳性率，提高预测的准确率，也能得到更多的预测结果。不可忽视的是，对预测出来的病毒宿

主进行更严谨验证也是非常必要的。(3) 以 HiC 技术和 Microbe-Seq 技术为代表的物理接触信息方法，是实时鉴定环境中病毒感染关系和研究病毒宿主相互作用的有力手段，具有很大的发展潜力和广阔的应用场景。优化这类方法的实验操作和生物信息学计算，可以提高这类方法的应用场景和鉴定能力以及降低减低成本。(4) 随着研究的推进，被预测出来的病毒和宿主的感染关系日益增加，亟需建立系统的病毒宿主感染关系库以指导今后环境未培养病毒的宿主预测，以及为改进和开发病毒宿主预测方法提供数据支撑。

致谢

感谢上海交通大学刘浩东、香港科技大学蔡兰兰、南方科技大学陈雨霏和郑峰峰对本论文写作提出宝贵的修改意见。

参考文献

[1] Mushegian AR. Are there 10^{31} virus particles on earth, or more, or fewer? *Journal of Bacteriology*, 2020, 202(9): e00052–e00020.

[2] Breitbart M, Rohwer F. Here a virus, there a virus, everywhere the same virus? *Trends in Microbiology*, 2005, 13(6): 278–284.

[3] Weinbauer MG. Ecology of prokaryotic viruses. *FEMS Microbiology Reviews*, 2004, 28(2): 127–181.

[4] Suttle CA. Marine viruses—major players in the global ecosystem. *Nature Reviews Microbiology*, 2007, 5(10): 801–812.

[5] Breitbart M, Bonnain C, Malki K, Sawaya NA. Phage puppet masters of the marine microbial realm. *Nature Microbiology*, 2018, 3(7): 754–766.

[6] Warwick-Dugdale J, Buchholz HH, Allen MJ, Temperton B. Host-hijacking and planktonic piracy: how phages command the microbial high seas. *Virology Journal*, 2019, 16(1): 15.

[7] Kuzyakov Y, Mason-Jones K. Viruses in soil: nano-scale undead drivers of microbial life, biogeochemical turnover and ecosystem functions. *Soil Biology and Biochemistry*, 2018, 127: 305–317.

[8] Paul JH. Microbial gene transfer: an ecological perspective. *Journal of Molecular Microbiology and Biotechnology*, 1999, 1(1): 45–50.

[9] Zhang R, Wei W, Cai LL. The fate and biogeochemical cycling of viral elements. *Nature Reviews Microbiology*, 2014, 12(12): 850–851.

[10] Chen XW, Weinbauer MG, Jiao NZ, Zhang R. Revisiting marine lytic and lysogenic virus-host interactions: kill-the-winner and piggyback-the-winner. *Science Bulletin*, 2021, 66(9): 871–874.

[11] Wilhelm S, Suttle C. Viruses and nutrient cycles in the sea viruses play critical roles in the structure and function of aquatic food webs. *BioScience*, 1999, 49: 781–788.

[12] Coutinho FH, Gregoracci GB, Walter JM, Thompson CC, Thompson FL. Metagenomics sheds light on the ecology of marine microbes and their viruses. *Trends in Microbiology*, 2018, 26(11): 955–965.

[13] Gregory AC, Zayed AA, Conceição-Neto N, Temperton B, Bolduc B, Alberti A, Ardyna M, Arkhipova K, Carmichael M, Cruaud C, Dimier C, Domínguez-Huerta G, Ferland J, Kandels S, Liu YX, Marec C, Pesant S, Picheral M, Pisarev S, Poulain J, Tremblay JÉ, Vik D, Babin M, Bowler C, Culley AI, De Vargas C, Dutilh BE, Iudicone D, Karp-Boss L, Roux S, Sunagawa S, Wincker P, Sullivan MB, Acinas SG, Babin M, Bork P, Boss E, Bowler C, Cochrane G, De Vargas C, Follows M, Gorsky G, Grimsley N, Guidi L, Hingamp P, Iudicone D, Jaillon O, Kandels-Lewis S, Karp-Boss L, Karsenti E, Not F, Ogata H, Pesant S, Poulton N, Raes J, Sardet C, Speich S, Stemmann L, Sullivan MB, Sunagawa S, Wincker P. Marine DNA viral macro- and microdiversity from pole to pole. *Cell*, 2019, 177(5): 1109–1123.e14.

[14] Breitbart M, Salamon P, Andresen B, Mahaffy JM, Segall AM, Mead D, Azam F, Rohwer F. Genomic analysis of uncultured marine viral communities. *PNAS*, 2002, 99(22): 14250–14255.

[15] Andrade-Martínez JS, Camelo Valera LC, Chica Cárdenas LA, Forero-Junco L, López-Leal G, Moreno-Gallego JL, Rangel-Pineros G, Reyes A. Computational tools for the analysis of uncultivated phage genomes. *Microbiology and Molecular Biology Reviews*: *MMBR*, 2022, 86(2): e0000421.

[16] Coclet C, Roux S. Global overview and major challenges of host prediction methods for uncultivated phages. *Current Opinion in Virology*, 2021, 49: 117–126.

[17] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *Journal of Molecular Biology*, 1990, 215(3): 403–410.

[18] Buchfink B, Reuter K, Drost HG. Sensitive protein alignments at tree-of-life scale using diamond. *Nature Methods*, 2021, 18(4): 366–368.

[19] Serres MH, Goswami S, Riley M. GenProtEC: an updated and improved analysis of functions of *Escherichia coli* K-12 proteins. *Nucleic Acids Research*, 2004, 32(suppl_1): D300–D302.

[20] Riley M. Functions of the gene products of *Escherichia coli*. *Microbiological Reviews*, 1993, 57(4): 862–952.

[21] Philosof A, Yutin N, Flores-Uribe J, Sharon I, Koonin EV, Béjà O. Novel abundant oceanic viruses of uncultured marine group II euryarchaeota. *Current Biology*: *CB*, 2017, 27(9): 1362–1368.

[22] Nishimura Y, Watai H, Honda T, Mihara T, Omae K, Roux S, Blanc-Mathieu R, Yamamoto K, Hingamp P, Sako Y, Sullivan MB, Goto S, Ogata H, Yoshida T. Environmental viral genomes shed new light on virus-host interactions in the ocean. *mSphere*, 2017, 2(2): e00359–e00316.

[23] Ahlgren NA, Fuchsman CA, Rocap G, Fuhrman JA. Discovery of several novel, widespread, and ecologically distinct marine Thaumarchaeota viruses that encode *amoC* nitrification genes. *The ISME Journal*, 2019, 13(3): 618–631.

[24] Sullivan MB, Waterbury JB, Chisholm SW. Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature*, 2003, 424(6952): 1047–1051.

[25] Sharon I, Battchikova N, Aro EM, Giglione C, Meinnel T, Glaser F, Pinter RY, Breitbart M, Rohwer F, Béjà O. Comparative metagenomics of microbial traits within oceanic viral communities. *The ISME Journal*, 2011, 5(7): 1178–1190.

[26] Dutilh BE, Cassman N, McNair K, Sanchez SE, Silva GGZ, Boling LC, Barr JJ, Speth DR, Seguritan V, Aziz RK, Felts B, Dinsdale EA, Mokili JL, Edwards RA. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nature Communications*, 2014, 5: 4498.

[27] Edwards RA, McNair K, Faust K, Raes J, Dutilh BE. Computational approaches to predict bacteriophage-host relationships. *FEMS Microbiology Reviews*, 2016, 40(2): 258–272.

[28] Roux S, Páez-Espino D, Chen IMA, Palaniappan K, Ratner A, Chu K, Reddy TBK, Nayfach S, Schulz F, Call L, Neches RY, Woyke T, Ivanova NN, Eloe-Fadrosh EA, Kyrpides NC. IMG/VR v3: an integrated ecological and evolutionary framework for interrogating genomes of uncultivated viruses. *Nucleic Acids Research*, 2021, 49(D1): D764–D775.

[29] Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. *Science*, 2010, 327(5962): 167–170.

[30] Pougach K, Semenova E, Bogdanova E, Datsenko KA, Djordjevic M, Wanner BL, Severinov K. Transcription, processing and function of CRISPR cassettes in *Escherichia coli*. *Molecular Microbiology*, 2010, 77(6): 1367–1379.

[31] Zhang RS, Mirdita M, Levy Karin E, Norroy C, Galiez C, Söding J. SpacePHARER: sensitive identification of phages from CRISPR spacers in prokaryotic hosts. *Bioinformatics*, 2021, 37(19): 3364–3366.

[32] Biswas A, Staals RHJ, Morales SE, Fineran PC, Brown CM. CRISPRDetect: a flexible algorithm to define CRISPR arrays. *BMC Genomics*, 2016, 17: 356.

[33] Cassman N, Prieto-Davó A, Walsh K, Silva GGZ, Angly F, Akhter S, Barott K, Busch J, McDole T, Haggerty JM, Willner D, Alarcón G, Ulloa O, DeLong EF, Dutilh BE, Rohwer F, Dinsdale EA. Oxygen minimum zones harbour novel viral communities with low diversity. *Environmental Microbiology*, 2012, 14(11): 3043–3065.

[34] Trubl G, Jang HB, Roux S, Emerson JB, Solonenko N, Vik DR, Solden L, Ellenbogen J, Runyon AT, Bolduc B, Woodcroft BJ, Saleska SR, Tyson GW, Wrighton KC, Sullivan MB, Rich VI. Soil viruses are underexplored players in ecosystem carbon processing. *mSystems*, 2018, 3(5): e00076–e00018.

[35] Coutinho FH, Cabello-Yeves PJ, Gonzalez-Serrano R, Rosselli R, López-Pérez M, Zemskaya TI, Zakharenko AS, Ivanov VG, Rodriguez-Valera F. New viral biogeochemical roles revealed through metagenomic analysis of Lake Baikal. *Microbiome*, 2020, 8(1): 163.

[36] Sanguino L, Franqueville L, Vogel TM, Larose C. Linking environmental prokaryotic viruses and their host through CRISPRs. *FEMS Microbiology Ecology*, 2015, 91(5): fiv046.

[37] Berg Miller ME, Yeoman CJ, Chia N, Tringe SG, Angly FE, Edwards RA, Flint HJ, Lamed R, Bayer EA, White BA. Phage-bacteria relationships and CRISPR elements revealed by a metagenomic survey of the rumen microbiome. *Environmental Microbiology*, 2012, 14(1): 207–227.

[38] Xu B, Li FY, Cai LL, Zhang R, Fan L, Zhang CL. A holistic genome dataset of bacteria, archaea and viruses of the Pearl River Estuary. *Scientific Data*, 2022, 9: 49.

[39] Correa AMS, Howard-Varona C, Coy SR, Buchan A, Sullivan MB, Weitz JS. Revisiting the rules of life for

viruses of microorganisms. *Nature Reviews Microbiology*, 2021, 19(8): 501–513.

[40] Paul JH. Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas? *The ISME Journal*, 2008, 2(6): 579–589.

[41] Mizuno CM, Rodriguez-Valera F, Kimes NE, Ghai R. Expanding the marine virosphere using metagenomics. *PLoS Genetics*, 2013, 9(12): e1003987.

[42] Roux S, Hallam SJ, Woyke T, Sullivan MB. Viral dark matter and virus-host interactions resolved from publicly available microbial genomes. *eLife*, 2015, 4: e08490.

[43] Pride DT, Wassenaar TM, Ghose C, Blaser MJ. Evidence of host-virus co-evolution in tetranucleotide usage patterns of bacteriophages and eukaryotic viruses. *BMC Genomics*, 2006, 7: 8.

[44] Gouy M, Gautier C. Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Research*, 1982, 10(22): 7055–7074.

[45] Lu CY, Zhang Z, Cai ZN, Zhu ZZ, Qiu Y, Wu AP, Jiang TJ, Zheng HP, Peng YS. Prokaryotic virus host predictor: a Gaussian model for host prediction of prokaryotic viruses in metagenomics. *BMC Biology*, 2021, 19(1): 5.

[46] Ahlgren NA, Ren J, Lu YY, Fuhrman JA, Sun FZ. Alignment-free d_2^* oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences. *Nucleic Acids Research*, 2016, 45(1): 39–53.

[47] Liu D, Ma YJ, Jiang XP, He TT. Predicting virus-host association by kernelized logistic matrix factorization and similarity network fusion. *BMC Bioinformatics*, 2019, 20(suppl 16): 594.

[48] Villarroel J, Kleinheinz KA, Jurtz VI, Zschach H, Lund O, Nielsen M, Larsen MV. HostPhinder: a phage host prediction tool. *Viruses*, 2016, 8(5): 116.

[49] Galiez C, Siebert M, Enault F, Vincent J, Söding J. WIsH: who is the host? Predicting prokaryotic hosts from metagenomic phage contigs. *Bioinformatics*, 2017, 33(19): 3113–3114.

[50] Zhou F, Gan R, Zhang F, Ren C, Yu L, Si Y, Huang Z. PHISDetector: a tool to detect diverse *in silico* phage-host interaction signals for virome studies. *Genomics*, *Proteomics & Bioinformatics*, 2022. Doi: https://doi.org/10.1016/j.gpb.2022.02.003.

[51] Wang WL, Ren J, Tang KJ, Dart E, Ignacio-Espinoza JC, Fuhrman JA, Braun J, Sun FZ, Ahlgren NA. A network-based integrated framework for predicting virus-prokaryote interactions. *NAR Genomics and Bioinformatics*, 2020, 2(2): lqaa044.

[52] Coutinho FH, Silveira CB, Gregoracci GB, Thompson CC, Edwards RA, Brussaard CPD, Dutilh BE, Thompson FL. Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans. *Nature Communications*, 2017, 8: 15955.

[53] Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nature Methods*, 2012, 9(4): 357–359.

[54] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*: *Oxford*, *England*, 2009, 25(14): 1754–1760.

[55] Lima-Mendez G, Faust K, Henry N, Decelle J, Colin S, Carcillo F, Chaffron S, Ignacio-Espinosa JC, Roux S, Vincent F, Bittner L, Darzi Y, Wang J, Audic S, Berline L, Bontempi G, Cabello AM, Coppola L, Cornejo-Castillo FM, D'Ovidio F, De Meester L, Ferrera I, Garet-Delmas MJ, Guidi L, Lara E, Pesant S, Royo-Llonch M, Salazar G, Sánchez P, Sebastian M, Souffreau C, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Gorsky G, Not F, Ogata H, Speich S, Stemmann L, Weissenbach J, Wincker P, Acinas SG, Sunagawa S, Bork P, Sullivan MB, Karsenti E, Bowler C, De Vargas C, Raes J, Coordinators TO. Determinants of community structure in the global plankton interactome. *Science*, 2015, 348(6237): 1262073.

[56] Nielsen HB, Almeida M, Juncker AS, Rasmussen S, Li JH, Sunagawa S, Plichta DR, Gautier L, Pedersen AG, Le Chatelier E, Pelletier E, Bonde I, Nielsen T, Manichanh C, Arumugam M, Batto JM, Quintanilha Dos Santos MB, Blom N, Borruel N, Burgdorf KS, Boumezbeur F, Casellas F, Doré J, Dworzynski P, Guarner F, Hansen T, Hildebrand F, Kaas RS, Kennedy S, Kristiansen K, Kultima JR, Léonard P, Levenez F, Lund O, Moumen B, Le Paslier D, Pons N, Pedersen O, Prifti E, Qin JJ, Raes J, Sørensen S, Tap J, Tims S, Ussery DW, Yamada T, Renault P, Sicheritz-Ponten T, Bork P, Wang J, Brunak S, Ehrlich SD. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nature Biotechnology*, 2014, 32(8): 822–828.

[57] Alrasheed H, Jin R, Weitz JS. Caution in inferring viral strategies from abundance correlations in marine metagenomes. *Nature Communications*, 2019, 10: 501.

[58] Knowles B, Silveira CB, Bailey BA, Barott K, Cantu VA, Cobián-Güemes AG, Coutinho FH, Dinsdale EA, Felts B, Furby KA, George EE, Green KT, Gregoracci GB, Haas AF, Haggerty JM, Hester ER, Hisakawa N, Kelly LW, Lim YW, Little M, Luque A, McDole-Somera T, McNair K, De Oliveira LS, Quistad

SD, Robinett NL, Sala E, Salamon P, Sanchez SE, Sandin S, Silva GGZ, Smith J, Sullivan C, Thompson C, Vermeij MJA, Youle M, Young C, Zgliczynski B, Brainard R, Edwards RA, Nulton J, Thompson F, Rohwer F. Lytic to temperate switching of viral communities. *Nature*, 2016, 531(7595): 466–470.

[59] Silveira CB, Rohwer FL. Piggyback-the-Winner in host-associated microbial communities. *npj Biofilms and Microbiomes*, 2016, 2: 16010.

[60] Wigington CH, Sonderegger D, Brussaard CPD, Buchan A, Finke JF, Fuhrman JA, Lennon JT, Middelboe M, Suttle CA, Stock C, Wilson WH, Wommack KE, Wilhelm SW, Weitz JS. Re-examination of the relationship between marine virus and microbial cell abundances. *Nature Microbiology*, 2016, 1: 15024.

[61] Coenen AR, Weitz JS. Limitations of correlation-based inference in complex virus-microbe communities. *mSystems*, 2018, 3(4): e00084–e00018.

[62] Roux S, Krupovic M, Daly RA, Borges AL, Nayfach S, Schulz F, Sharrar A, Matheus Carnevali PB, Cheng JF, Ivanova NN, Bondy-Denomy J, Wrighton KC, Woyke T, Visel A, Kyrpides NC, Eloe-Fadrosh EA. Cryptic inoviruses revealed as pervasive in bacteria and archaea across Earth's biomes. *Nature Microbiology*, 2019, 4(11): 1895–1906.

[63] Coutinho FH, Zaragoza-Solas A, López-Pérez M, Barylski J, Zielezinski A, Dutilh BE, Edwards R, Rodriguez-Valera F. RaFAH: host prediction for viruses of bacteria and archaea based on protein content. *Patterns*: *New York*, *N Y* , 2021, 2(7): 100274.

[64] Li ML, Wang YN, Li FY, Zhao Y, Liu MY, Zhang SJ, Bin YN, Smith AI, Webb GI, Li J, Song JN, Xia JF. A deep learning-based method for identification of bacteriophage-host interaction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2021, 18(5): 1801–1810.

[65] Shang JY, Sun YN. Predicting the hosts of prokaryotic viruses using GCN-based semi-supervised learning. *BMC Biology*, 2021, 19(1): 250.

[66] Roux S, Hawley AK, Torres Beltran M, Scofield M, Schwientek P, Stepanauskas R, Woyke T, Hallam SJ, Sullivan MB. Ecology and evolution of viruses infecting uncultivated SUP05 bacteria as revealed by single-cell- and meta-genomics. *eLife*, 2014, 3: e03125.

[67] Labonté JM, Swan BK, Poulos B, Luo HW, Koren S, Hallam SJ, Sullivan MB, Woyke T, Eric Wommack K, Stepanauskas R. Single-cell genomics-based analysis of virus-host interactions in marine surface bacterioplankton. *The ISME Journal*, 2015, 9(11): 2386–2399.

[68] Jarett JK, Džunková M, Schulz F, Roux S, Paez-Espino D, Eloe-Fadrosh E, Jungbluth SP, Ivanova N, Spear JR, Carr SA, Trivedi CB, Corsetti FA, Johnson HA, Becraft E, Kyrpides N, Stepanauskas R, Woyke T. Insights into the dynamics between viruses and their hosts in a hot spring microbial mat. *The ISME Journal*, 2020, 14(10): 2527–2541.

[69] Mosier-Boss PA, Lieberman SH, Andrews JM, Rohwer FL, Wegley LE, Breitbart M. Use of fluorescently labeled phage in the detection and identification of bacterial species. *Applied Spectroscopy*, 2003, 57(9): 1138–1144.

[70] Dang VT, Sullivan MB. Emerging methods to study bacteriophage infection at the single-cell level. *Frontiers in Microbiology*, 2014, 5: 724.

[71] Deng L, Gregory A, Yilmaz S, Poulos BT, Hugenholtz P, Sullivan MB. Contrasting life strategies of viruses that infect photo- and heterotrophic bacteria, as revealed by viral tagging. *mBio*, 2012, 3(6): e00373–e00312.

[72] Deng L, Ignacio-Espinoza JC, Gregory AC, Poulos BT, Weitz JS, Hugenholtz P, Sullivan MB. Viral tagging reveals discrete populations in *Synechococcus* viral genome sequence space. *Nature*, 2014, 513(7517): 242–245.

[73] Džunková M, Low SJ, Daly JN, Deng L, Rinke C, Hugenholtz P. Defining the human gut host-phage network through single-cell viral tagging. *Nature Microbiology*, 2019, 4(12): 2192–2203.

[74] Allers E, Moraru C, Duhaime MB, Beneze E, Solonenko N, Barrero-Canosa J, Amann R, Sullivan MB. Single-cell and population level viral infection dynamics revealed by phageFISH, a method to visualize intracellular and free viruses. *Environmental Microbiology*, 2013, 15(8): 2306–2318.

[75] Tadmor AD, Ottesen EA, Leadbetter JR, Phillips R. Probing individual environmental bacteria for viruses by using microfluidic digital PCR. *Science*, 2011, 333(6038): 58–62.

[76] Morella NM, Yang SC, Hernandez CA, Koskella B. Rapid quantification of bacteriophages and their bacterial hosts *in vitro* and *in vivo* using droplet digital PCR. *Journal of Virological Methods*, 2018, 259: 18–24.

[77] Spencer SJ, Tamminen MV, Preheim SP, Guo MT, Briggs AW, Brito IL, A Weitz D, Pitkänen LK, Vigneault F, Virta MP, Alm EJ. Massively parallel sequencing of single cells by epicPCR links functional

genes with phylogenetic markers. *The ISME Journal*, 2016, 10(2): 427–436.

[78] Sakowski EG, Arora-Williams K, Tian FN, Zayed AA, Zablocki O, Sullivan MB, Preheim SP. Interaction dynamics and virus-host range for estuarine actinophages captured by epicPCR. *Nature Microbiology*, 2021, 6(5): 630–642.

[79] Zheng WS, Zhao SJ, Yin YH, Zhang HD, Needham DM, Evans ED, Dai CL, Lu PJ, Alm EJ, Weitz DA. High-throughput, single-microbe genomics with strain resolution, applied to a human gut microbiome. *Science*, 2022, 376(6597): eabm1483.

[80] Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 2009, 326(5950): 289–293.

[81] Marbouty M, Baudry L, Cournac A, Koszul R. Scaffolding bacterial genomes and probing host-virus interactions in gut microbiome by proximity ligation (chromosome capture) assay. *Science Advances*, 2017, 3(2): e1602105.

[82] Bickhart DM, Watson M, Koren S, Panke-Buisse K, Cersosimo LM, Press MO, Van Tassell CP, Van Kessel JAS, Haley BJ, Kim SW, Heiner C, Suen G, Bakshy K, Liachko I, Sullivan ST, Myer PR, Ghurye J, Pop M, Weimer PJ, Phillippy AM, Smith TPL. Assignment of virus and antimicrobial resistance genes to microbial hosts in a complex microbial community by combined long-read assembly and proximity ligation. *Genome Biology*, 2019, 20(1): 153.

[83] Marbouty M, Thierry A, Millot GA, Koszul R. MetaHiC phage-bacteria infection network reveals active cycling phages of the healthy human gut. *eLife*, 2021, 10: e60608.

[84] Krupovic M, Dolja VV, Koonin EV. Origin of viruses: primordial replicators recruiting capsids from hosts. *Nature Reviews Microbiology*, 2019, 17(7): 449–458.

[85] Danovaro R, Dell'Anno A, Corinaldesi C, Rastelli E, Cavicchioli R, Krupovic M, Noble RT, Nunoura T, Prangishvili D. Virus-mediated archaeal hecatomb in the deep seafloor. *Science Advances*, 2016, 2(10): e1600492.

[86] Kim JG, Kim SJ, Cvirkaite-Krupovic V, Yu WJ, Gwak JH, López-Pérez M, Rodriguez-Valera F, Krupovic M, Cho JC, Rhee SK. Spindle-shaped viruses infect marine ammonia-oxidizing thaumarchaea. *PNAS*, 2019, 116(31): 15645–15650.

[87] Haaber J, Leisner JJ, Cohn MT, Catalan-Moreno A, Nielsen JB, Westh H, Penadés JR, Ingmer H. Bacterial viruses enable their host to acquire antibiotic resistance genes from neighbouring cells. *Nature Communications*, 2016, 7: 13333.

[88] Hussain FA, Dubert J, Elsherbini J, Murphy M, VanInsberghe D, Arevalo P, Kauffman K, Rodino-Janeiro BK, Gavin H, Gomez A, Lopatina A, Le Roux F, Polz MF. Rapid evolutionary turnover of mobile genetic elements drives bacterial resistance to phages. *Science*, 2021, 374(6566): 488–492.

[89] Debroas D, Siguret C. Viruses as key reservoirs of antibiotic resistance genes in the environment. *The ISME Journal*, 2019, 13(11): 2856–2867.

[90] Suttle CA. Viruses in the sea. *Nature*, 2005, 437(7057): 356–361.

[91] Toporek A, Lechtzin N. Viruses to the rescue-use of bacteriophage to treat resistant pulmonary infections. *Cell*, 2022, 185(11): 1807–1808.

[92] Dedrick RM, Guerrero-Bustamante CA, Garlena RA, Russell DA, Ford K, Harris K, Gilmour KC, Soothill J, Jacobs-Sera D, Schooley RT, Hatfull GF, Spencer H. Engineered bacteriophages for treatment of a patient with a disseminated drug-resistant *Mycobacterium abscessus*. *Nature Medicine*, 2019, 25(5): 730–733.

范陆，南方科技大学海洋与工程系助理教授，博士生导师，独立 PI。毕业于浙江大学和澳大利亚新南威尔士大学，曾在澳大利亚昆士兰大学和深圳华大基因工作。主要研究领域为微生物与病毒的生态和演化。曾以第一或通讯作者，在 *Nature Ecology & Evolution*、*PNAS*、*ISME J* 等国际一流杂志上发表研究成果。曾获欧洲微生物学会(FEMS)青年科学家奖，澳大利亚新南威尔士州海洋科学最佳研究奖。