

DOI: 10.13930/j.cnki.cjea.180695

王新珍, 王凤花, 孙瑞波, 刘彬彬. 高通量测序技术在微生物分子生态学中的应用[J]. 中国生态农业学报, 2018, 26(10): 1593–1600

WANG X Z, WANG F H, SUN R B, LIU B B. Application of high-throughput DNA sequencing in microbial ecology[J]. Chinese Journal of Eco-Agriculture, 2018, 26(10): 1593–1600

## 高通量测序技术在微生物分子生态学中的应用\*

王新珍, 王凤花, 孙瑞波, 刘彬彬\*\*

(中国科学院遗传与发育生物学研究所农业资源研究中心 石家庄 050022)

**摘要:** 微生物在众多的自然和人工生态系统中发挥着核心的作用, 但能够被培养分离的微生物在大部分生态系统中只占极少一部分, 极大地限制了人们对微生物组成、功能及其潜在应用的认识。分子生物学方法, 尤其是高通量测序技术应用到微生物生态学研究, 为认识微生物多样性、群落结构组成及其生态功能提供了有利手段。高通量测序作为一种新兴的免培养分子生物学技术, 具备检测快速、准确、信息全面丰富等特点。随着高通量测序技术的不断升级换代, 测序通量、读长和准确度的不断提升以及成本的大幅下降, 该技术在过去十几年间被迅速应用于土壤、水体和肠道等微生物区系的研究中。本文简述了基于高通量测序技术的 PCR 产物测序技术和宏基因组学测序技术的原理、发展历程、数据分析方法与应用, 以及宏基因组学测序技术在病毒学领域的应用, 以期对微生物分子生态学研究提供参考。

**关键词:** 高通量测序; 微生物分子生态学; PCR 产物测序; 宏基因组学; 病毒

**中图分类号:** Q93-31; Q938.1 **文献标识码:** A **文章编号:** 1671-3990(2018)10-1593-08

## Application of high-throughput DNA sequencing in microbial ecology\*

WANG Xinzhen, WANG Fenghua, SUN Ruibo, LIU Binbin\*\*

(Center for Agricultural Resources Research, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Shijiazhuang 050022, China)

**Abstract:** Microorganisms play essential roles in natural and artificial ecosystems. However, only a small portion of microorganisms can be cultured and isolated in most ecosystems, which greatly limits our understanding of microbial community composition, ecological function and their potential applications. The novel molecular techniques, especially high-throughput sequencing, have provided advantages for further exploring the diversity, composition and ecological functions of microbial community. High-throughput sequencing is a culture-independent technology capable of deep, rapid and accurate detection of genetic information. With the update of the high-throughput sequencing technology, the sequencing throughput, read length and accuracy have been dramatically improved, and the cost has greatly declined. During the past decade, high-throughput sequencing technology has been rapidly applied in the microbial ecological studies with various types of samples such as soil, water and gut microbial communities. Amplicon and metagenomic sequencing are most widely used strategies for environmental samples. Amplicon sequencing refers to sequencing the PCR products of target gene fragments amplified with specifically designed primers. The 16S rRNA gene of prokaryotes and the 18S rRNA and ITS genes of eukaryotes are most commonly used marker genes to conduct microbial taxonomic analysis. In addition, functional genes such as *nirK*, *nirS* and *nosZ* genes of denitrifying bacteria, *amoA* gene of ammonia-oxidizing

\* 国家重点研发计划项目(2016YFD0800100)、国家自然科学基金青年基金项目(41601511)和河北省优秀青年基金项目(D2017503022)资助

\*\* 通信作者: 刘彬彬, 主要研究领域为微生物分子生态学和生物信息学。E-mail: binbinliu@sjziam.ac.cn

王新珍, 主要研究领域为环境微生物学。E-mail: xzwwang@sjziam.ac.cn

收稿日期: 2018-07-15 接受日期: 2018-08-05

\* Supported by the National Key Research and Development Project of China (2016YFD0800100), the National Natural Science Foundation of China (41601511) and Foundation for Distinguished Young Scholars of Hebei Province (D2017503022)

\*\* Corresponding author, E-mail: binbinliu@sjziam.ac.cn

Received Jul. 15, 2018; accepted Aug. 5, 2018

bacteria, and *ureC* gene of urea hydrolytic bacteria are frequently adopted to study the diversity of functional microorganisms. In metagenomics, sequencing is performed on the genomic DNA directly extracted from environmental samples therefore avoiding the bias from PCR. Theoretically this method can provide a representation of all genomes in the sample and can be used for fully exploring the genetic diversity, functional potentials and metabolic pathways of both cultured and uncultured microorganisms. Metagenomic sequencing technology has been applied in the field of medical diagnosis, human health, biological energy, environmental restoration and agricultural ecology, etc. and has provided us new insights into taxonomic diversity, ecological function, evolutionary succession and interaction in the complex microflora. The application of metagenomic sequencing technology to the field of virology is referred as viral metagenomics. Viruses are not only related to various diseases of crop, animal and human, but also indispensable in the natural ecosystems, and play an important role in regulating host diversity and community succession, mediating gene transfer between microbes, and promoting global biogeochemical cycles. Viral metagenomics has recently gained momentum in application in the field of environmental science to reveal the genetic diversity, explore the novel species of viruses, and investigate their interactions with environmental factors.

**Keywords:** High-throughput sequencing; Microbial ecology; Amplicon sequencing; Metagenomics; Virus

20 世纪末和 21 世纪初, 随着微生物分子生态学的发展, 环境样品的微生物分析方法发生了快速的转变, 传统的微生物培养技术迅速地被新的分子技术取代。这一转变的主要原因是大多数生态系统中可培养的微生物只占一小部分<sup>[1]</sup>, 采用培养方法分析微生物群落结构会造成巨大的偏差。分子生物学方法直接从环境样品中提取核酸, 避开了培养过程造成的偏差, 能够快速分析环境中的微生物区系组成。由于大多数典型的环境样品中微生物数量大、种类多, 因此需要高通量的技术来进行研究。近年来, 高通量测序技术的发展满足了这一需求, 被广泛地应用到微生物分子生态学研究, 其中 PCR 产物测序和宏基因组学测序技术迅速发展起来。本文主要针对这些方法的原理、发展历程、数据分析方法以及应用情况进行介绍。

## 1 PCR 产物测序

PCR 产物测序是指通过设计特定的引物, 对目的基因片段的 PCR 扩增产物进行测序。原核生物 16S rRNA 基因和真核生物 18S rRNA 和 ITS 基因序列高度保守, 是微生物系统分类研究中最为常用的分子生物标记基因, 被广泛应用于微生物生态学研究。此外, 一些微生物的功能基因也常被用于功能微生物的多样性研究, 这些基因包括反硝化细菌的 *nirK*、*nirS* 和 *nosZ* 基因, 氨单加氧酶 *amoA* 基因(催化硝化作用的第一步), 以及尿素水解菌的尿素水解酶 *ureC* 基因等。通过对这些特定的序列区域进行测序和对比, 能够揭示环境微生物的物种组成及其特定生态功能<sup>[2]</sup>。

### 1.1 PCR 产物测序技术

传统的微生物研究需要依托分离培养方法获取环境样品中的微生物。然而, 环境中绝大部分的微

生物不能通过现有的培养条件获得, 从而限制了对环境微生物的分析。分子生物学技术可以直接从环境样品中提取微生物核酸(DNA), 采用 PCR 技术从环境样品 DNA 中扩增微生物分子生物标记基因和功能基因, 通过对这些 PCR 产物进行测序来揭示环境微生物的多样性、群落结构组成及其生态功能。早期的分子生物学技术, 如变性梯度凝胶电泳(denaturing gradient gel electrophoresis, DGGE)、温度梯度凝胶电泳(temperature gradient gel electrophoresis, TGGE)和限制性片段长度多态性分析(restriction fragment length polymorphism, RFLP)等因其特定的局限性, 导致检测到的微生物数量较少, 大约为微生物总数的 0.1%~10%<sup>[3]</sup>, 不能从整体上更为深入详细地揭示环境微生物的区系组成。随着测序技术的不断优化, 高通量测序技术是对传统测序技术的一次巨大改革, 能够提供丰富的、真实可靠的基因序列信息, 为微生物分子生态学研究提供了有利手段。环境样品 DNA 的提取是保障高通量测序顺利进行的关键步骤, 主要包括直接和间接提取法。直接提取法能够克服传统微生物培养方法的局限性, 并保证所得样品具备一定的代表性; 间接提取法则需要培养、分离纯化环境样品中的微生物, 然后提取 DNA。由于直接提取法获取 DNA 的效率高于间接提取法, 且更具有代表性, 因而得到更广泛的应用。需要指出的是, PCR 产物测序采用的样本是 DNA 经过 PCR 扩增后的产物, 而 PCR 扩增过程依托的通用引物并不能对所有微生物进行等效扩增, 因此扩增具有一定的选择性<sup>[4]</sup>。

PCR 产物测序技术以 454 测序技术和 Illumina 测序技术最为普遍。454 测序技术的优势在于单条序列读长较长, 其测序结果不经过拼接就能进行一定的信息分析。其利用焦磷酸测序的原理: 首先将

基因组 DNA 打碎成 300~800 bp 的片段, 在片段两端分别连接不同的接头; 将带有接头的单链 DNA 固定在 DNA 捕获磁珠上, 采用扩增试剂使之乳化, 构成仅含有 1 个磁珠和 1 条独特 DNA 片段的微反应区, 每条独特的片段在各自的微反应区里进行独立的扩增; 将携带有 PCR 产物的磁珠放入只能容纳单个磁珠的 Pico Titer Plate (PTP) 板中进行测序, T、A、C、G 碱基依次循环进入 PTP 板, 这些碱基伴随碱基配对释放焦磷酸, 焦磷酸在荧光素酶和 ATP 硫酸化酶的作用下释放出光信号, 依据光信号捕获、转化和记录的结果, 即可准确、快速地确定待测模板 DNA 的碱基序列<sup>[5]</sup>。Illumina 测序技术则是采用可逆性末端边合成边测序的原理: 将基因组 DNA 打碎成 100~200 bp 的小片段, 在片段两端加上序列已知的通用接头构建文库, 通过接头与测序芯片 Flowcell 基底上的 Oligo 序列互补, 每条文库片段都经过桥式 PCR 扩增形成 1 个簇, 然后测序时, 在合成反应中加入改造过的 DNA 聚合酶和带有 4 种荧光标记的 dNTP, 每个循环反应只能加入 1 个正确互补的碱基, 根据 4 种不同的荧光信号确认碱基种类, 最后依据每轮收集到的荧光信号读取核酸序列<sup>[6]</sup>。该技术单次测序获得的数据量大、覆盖率高, 可检测到更多低丰度的遗传信息, 并且成本相对较低<sup>[7]</sup>。

## 1.2 PCR 产物测序技术的应用

21 世纪初, 454 高通量测序技术的诞生为揭示环境微生物基因组组成提供了新手段, 极大地丰富了人们对海洋、污水、活性污泥和土壤等各种环境中微生物多样性的认识。Chen 等<sup>[8]</sup>通过大规模地焦磷酸测序获得 20 万条细菌 V6 区基因序列, 为揭示珊瑚细菌和微藻共生体内部的相互关系提供了数据参考。Ye 等<sup>[9]</sup>采用 454 测序技术揭示了城市污水处理厂不同处理区域的细菌群落组成, 发现活性污泥样品中细菌多样性高于进水区、出水区和消化泥样品, 这些样品主要的微生物类群包括  $\alpha$ -变形杆菌门 (Alphaproteobacteria)、热袍菌门 (Thermotogae)、 $\delta$ -变形菌门 (Deltaproteobacteria) 和  $\gamma$ -变形菌门 (Gammaproteobacteria)。此外, 454 测序技术可用于一些功能微生物相关研究如农田土壤中氨单加氧酶基因 (*amoA*) 多样性<sup>[10]</sup>, 以及高含量 CO<sub>2</sub> 土壤中微生物群落结构与功能<sup>[11]</sup>等。近年来, Illumina 测序技术以其在数据通量和运行时间等方面的显著优势, 发展成为微生物分子生态学研究中主流的高通量测序技术。赵柏霞等<sup>[12]</sup>研究了樱桃 (*Prunus pseudocerasus*) 不同生长发育期根际土壤的细菌多样

性, 发现不同生长发育期能够对根际土壤细菌群落结构产生一定的影响。Sun 等<sup>[13]</sup>采用高通量测序技术分析了中国科学院栾城长期定位试验田翻耕、旋耕和免耕 3 个处理不同层次土壤细菌和真菌的群落组成, 从微生物方面评价了不同耕作方式对土壤生态的影响, 指出不同耕作方式对真菌群落结构的影响大于细菌, 但细菌群落结构在不同层次土壤中的变化较真菌更为明显。Chen 等<sup>[14]</sup>采集华北平原典型厚包气带 0~10 m 不同层次土壤, 通过调控碳源和氧气含量的培养试验, 测定土壤反硝化速率、反硝化功能基因以及土壤微生物组成区系的变化, 解析有机碳含量和氧气浓度对土壤反硝化作用的影响, 指出添加有机碳可显著影响土壤微生物的区系组成, 而调控氧气浓度几乎不影响土壤微生物的区系组成, 并推测假单胞菌 (*Pseudomonas*) 和芽孢杆菌 (*Bacillus*) 在土壤反硝化过程中起到重要的作用。

## 2 宏基因组学测序分析技术及其应用

宏基因组学 (metagenomics) 一词最早由 Handelsman 等<sup>[15]</sup>提出, 它直接对环境样品中微生物群落的遗传物质进行研究, 可充分发掘未培养和未鉴定微生物的基因多样性与生化反应<sup>[16]</sup>。由于不依赖于可培养技术, 宏基因组学技术的出现大大扩展了我们对复杂的微生物群落结构和功能的认知, 极大地推动了微生物学的发展。目前宏基因组学技术正日臻成熟, 也形成了比较完善的研究流程 (图 1), 其中 DNA 测序和数据分析是宏基因组学研究的关键部分。

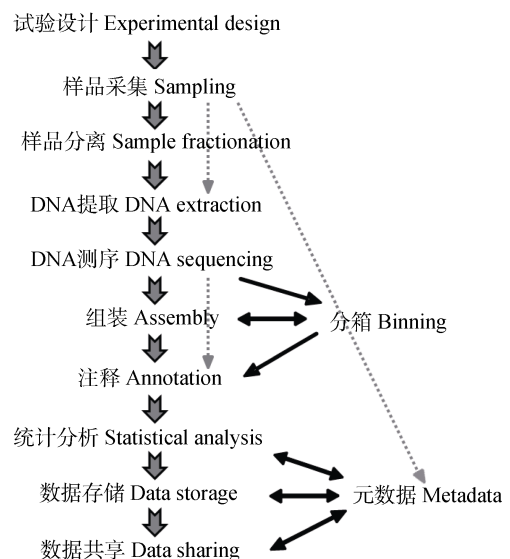


图 1 典型的宏基因组研究流程图(虚线表示可省略的步骤)<sup>[17]</sup>  
Fig. 1 Flow diagram of a typical metagenome project (dashed arrows indicate steps that can be omitted)<sup>[17]</sup>

## 2.1 宏基因组学测序技术

宏基因组学的发展离不开测序技术的进步,环境微生物数量大、多样性高,其遗传物质(DNA)包含海量的信息,要解读这浩瀚的遗传信息,获取其基因序列是第 1 步。自经典的 Sanger 测序出现以来,测序技术主要经历了 3 个重要的发展阶段:以荧光标记测序技术和毛细管阵列电泳测序技术为代表的第 1 代测序技术,以循环阵列合成测序为基础的第 2 代测序技术和第 3 代的单分子直接测序技术。第 2 代测序技术实现了高通量的大规模测序,并大大降低了测序成本,被认为是测序技术划时代的成就<sup>[18]</sup>。第 2 代测序技术的出现为宏基因组学的快速发展奠定了技术基石,是目前最为主流的测序技术,主要有 Apply Biosystems(ABI)公司的 SOLiD sequencing 平台,罗氏公司(Roche)的 454 测序平台以及 Illumina 公司的 Miseq 和 Hiseq 测序平台。454 测序平台在一段时期内占据着第 2 代测序的主要市场,但由于其数据通量较低,并且随着 Miseq 和 Hiseq 测序技术的不断进步,其片段读长大的优势也逐渐失去,目前已经淡出历史舞台,现在应用最为广泛的是 Hiseq 测序平台,但由于其测序原理的限制,片段读长低是其最大的缺陷。采用单分子测序的第 3 代测序技术,成功地摆脱了 DNA 扩增过程,极大地提高了测序长度,其测序长度可达 10~25 kb,目前主要有 Pacific Biosciences 和 Oxford Nanopore Technologies<sup>[19]</sup>,但其错误率较高,常用的做法是进行多次重复测序进行修正,或者以第 3 代测序长片段作为“骨架”,用第 2 代测序数据进行修正。相信随着技术的不断进步,第 3 代测序技术会不断完善,更新的技术也会涌现,为微生物学研究提供强有力的支持。

## 2.2 宏基因组学测序数据分析方法

由于宏基因组数据包含了整体微生物群落的大量 DNA 信息,而且片段长度短、重复片段多,如何进行序列拼接和注释是宏基因组技术面临的重大挑战。目前宏基因组拼接主要基于图论的算法,包括 Overlap-layout-consensus (OLC)和 de Bruijn Graph (DBG)两种。前者直接利用序列间的重叠序列建立公共路径以解决拼装问题,主要用于较大基因组的组装;后者则将短的测序片段转化成定长的 k-mer,然后寻找 k-mer 之间的重叠关系,通过建立 de Bruijn Graph,把碱基序列拼接问题转为图论问题,DBG 算法主要用于第 2 代高通量测序数据的拼接<sup>[20]</sup>。拼接后的功能注释是宏基因组技术的另一大难题,目前用于功能注释

的数据库主要有 KEGG(Kyoto Encyclopedia of Genes and Genomes)数据库、COG(Clusters of Orthologous Groups of proteins)数据库和 GO(Gene Ontology)数据库。随着宏基因组技术产生的大量数据需要分析,很多研究组发展了宏基因组学数据分析方法和工具。这些工具在用户使用上可以分为两类:一类是针对于具有较强生物信息学基础的用户,可以独立下载安装的宏基因组分析软件包;另一类则是面向生物信息基础薄弱的研究者开发的在线的宏基因组数据分析平台。目前常用的宏基因组数据分析平台包括:在线分析平台 MG-RAST 和 CoMet,以及本地分析平台 EBI Metagenomics 和 MEGAN 等。这些工具针对不同的研究者,为宏基因组数据分析提供了各具特色的研究工具<sup>[21]</sup>。但是,它们都针对全部功能进行拼接和注释,而对于特定功能基因来说,尤其是丰度在整体基因组中占比较小的基因,从庞大的宏基因组数据中找到这些基因犹如大海捞针,所以很难开展有针对性的研究,而基因靶向装配(gene-targeted assembly)工具的出现,为我们实现对特定功能基因的“大海捞针”提供了有力的工具。Xander 便是其中的代表之一<sup>[22]</sup>,其由密歇根州立大学开发,可以针对特定基因进行组装与注释。但是其分析依赖于高覆盖度、高准确度的参考数据库和准确的分析模型,目前软件开发者只提供了一些氮循环过程的参考基因数据库,而且已有的数据库质量有待提高,所以想要开展其他功能基因的研究,必须构建准确度高、覆盖度广的相应的参考数据库和分析模型,这在一定程度上限制了它的应用。

## 2.3 宏基因组学测序技术的应用

宏基因组测序和数据分析技术的蓬勃发展,逐渐冲破了微生物学研究的技术屏障,基于宏基因组技术的研究在医疗诊断与人类健康、生物能源、生物技术、环境修复、农业和环境生态等方面取得了丰硕的成果,对于微生物的物种和多样性组成、生态功能、进化演替、相互作用等也有了新的突破。研究发现人体微生物与人体健康密不可分,例如肠道微生物对人体营养代谢和维持肠道功能具有重要作用,并且与肥胖症、糖尿病、衰老等的发生、发展关系密切<sup>[23-25]</sup>,人为调控肠道微生物群落被看作是治疗相关疾病的新途径。例如,Zhao 等<sup>[26]</sup>通过宏基因组研究发现增加大量多样化的膳食纤维,可通过改变肠道菌群结构而显著改善 2 型糖尿病人的胰岛素分泌和胰岛素敏感性,而这与有利于增加胰岛素分泌和提高胰岛素敏感性的“短链脂肪酸”产生菌

密切相关。自然环境中,微生物对维持生态功能起着重要的作用,宏基因组技术为揭示环境生态过程提供了有力的工具。Liu 等<sup>[27]</sup>通过宏基因组测序和 Xander 解析了地表水-地下水的交错带氮循环功能微生物的群落,研究发现反硝化、异化还原成铵和硝化速率会对水体和沉积物改变做出快速响应,而且这与相应的功能微生物类群变化相关。Chu 等<sup>[28]</sup>通过宏基因组分析了污水处理厂排出物中的微生物、抗生素抗性基因和可移动基因原件在湖底沉积物中的扩散,研究发现污水处理厂排出物中的细菌携带的抗生素抗性基因可能会通过可移动基因原件转移到环境中。宏基因组为微生物研究提供了有力的工具,但也存在巨大的挑战,其中之一便是环境微生物具有极高的多样性,尤其对于复杂的生态系统,如土壤,目前的测序深度很难覆盖全部的微生物群落,而且对于复杂数据的解读依然是有待解决的主要问题。

### 3 病毒宏基因组学测序技术及其应用

将宏基因组学的研究方法应用于病毒学领域即为病毒宏基因组学(viral metagenomics)<sup>[29]</sup>。采用宏基因组学测序技术研究环境或生物体组织中的病毒遗传物质,能够克服大环境下病毒遗传物质浓度低的缺点,提高基因组信息的丰度。其在环境分子生态领域和人类健康与动植物检疫领域的应用,为揭示病毒基因多样性及其与环境之间的相互关系、发掘新病毒、鉴别病原体以及预警病毒感染提供了技术支持和参考依据。

#### 3.1 病毒宏基因组学测序技术在环境分子生态领域的应用

病毒是海洋中数量最多的生物类群,在海洋生态系统物质循环、能量流动、生物进化以及维持海洋生态系统功能等方面起着重要的作用<sup>[30]</sup>。2002 年, Breitbart 等<sup>[31]</sup>采用传统的鸟枪测序法(shortgun sequencing)揭示了美国加州海岸水体病毒的基因组组成,拉开了水体环境病毒宏基因组学研究的序幕。随着第 2 代测序技术的不断完善,宏基因组学测序技术在海洋病毒多样性及其群落结构组成研究中被广泛应用,极大地推动了病毒分子生态学的发展。Angly 等<sup>[32]</sup>从全球尺度对来自北冰洋、马尾藻海、不列颠哥伦比亚和墨西哥湾 4 个海域的 68 个样点的 184 个病毒集群进行了 10 年的追踪研究,发现了大量新的基因资源,并指出了病毒种类和丰度随纬度梯度变化的空间异质性。2008 年,国际期刊《Freshwater Biology》发表淡水浮游病毒专刊,综述

了噬藻体、藻病毒等病毒群落在沼泽、沉积物、沿岸带以及开放水域等淡水环境中的分布情况及其在淡水生态系统中的生态作用,表明淡水病毒与海水病毒一样,已经受到环境生态学家的高度关注,被认为是不可忽视的战略生物资源<sup>[33-34]</sup>。López-Bueno 等<sup>[35]</sup>报道了南极湖泊病毒群落与其他水体病毒基因组以细菌病毒序列为主不同,其大部分序列归属于真核生物病毒,包括许多未曾发现的 ssDNA 病毒和藻病毒;同时指出南极湖泊病毒群落春季以 ssDNA 病毒为主而夏季以 dsDNA 病毒为主的、可能与寄主密切关联的季节性动态变化过程。Ge 等<sup>[36]</sup>研究了东湖病毒基因组 4 个季节的多样性特征,发现病毒多样性在 8 月最高,12 月和 6 月居中,3 月最低。夏骏等<sup>[37]</sup>针对渤海秋冬季节的病毒宏基因组研究,指出其结构组成呈现季节性变化,与远洋病毒组相比具有独特性,并推测了未知 T4-like 病毒分支的存在。这些研究对于认识水体生态系统中病毒的分布规律,明确病毒在调控寄主微生物的丰度、群落结构组成以及它们之间的物种进化关系等具有重要意义。

相比之下,土壤环境病毒宏基因组学的研究进展比较缓慢。土壤种类的多样性和土壤环境的高度异质性,可能导致土壤环境病毒组成比海洋等水体环境病毒组成更复杂、更多样,而且病毒个体微小,容易被土壤颗粒吸附固定,从而给提取满足病毒宏基因组学测序条件的病毒遗传物质造成了一定的困难<sup>[38]</sup>。人们通常采用牛肉浸膏、甘油溶液、焦磷酸钠和柠檬酸钾等作为土壤病毒浸提液,同时结合震荡、超声波等物理方法从土壤中提取病毒<sup>[39]</sup>。需要注意的是,土壤颗粒对病毒的吸附与病毒种类、土壤 pH、土壤结构和离子浓度等密切相关,浸提剂的选择还需依据土壤样品性质等具体情况而定。近年来,病毒宏基因组学测序技术被应用于陆地生态系统如沙漠和耕地等环境。Adriaenssens 等<sup>[40]</sup>发现 Namib 沙漠土壤病毒宏基因组以有尾噬菌体为主,且细胞壁裂解酶基因、核苷酸还原酶基因以及噬菌体基因等功能基因丰度较高;此外,系统发育分析发现了许多新的生物标记基因。次年,Adriaenssens 等<sup>[41]</sup>又在 Namib 沙漠土壤病毒宏基因组中发现了大量新的基因资源。Reavy 等<sup>[42]</sup>研究了腐殖质铁灰壤土、高山灰壤土、耕作棕壤土和沿海土 4 种不同类型土壤中的病毒组,发现耕作棕壤土和沿海土病毒组以 ssDNA 病毒为主,并推测土壤性质影响了 ssDNA 病毒的分布。农业生产中,施肥方式、耕作措施、土壤类型、植被、温度、季节变化以及地域差异等对土壤细菌和真菌群落结构影响较大,但人

们对于农田土壤病毒组的认识局限于 Han 等<sup>[43]</sup>报道的我国 3 种不同土壤类型农田土壤中病毒物种和功能基因的组成。农业生态系统中的病毒是否同来自水体生态系统的病毒一样呈现空间异质性、时间异质性和寄主相关性的分布特征, 以及这些病毒如何响应各种农业生产活动还有待研究。

### 3.2 病毒宏基因组学测序技术在人类健康与动植物检疫领域的应用

病毒宏基因组学测序技术通过探索病毒组结构组成及其动态变化过程为多种人类临床疾病的诊断与治疗提供了重要理论依据。Willner 等<sup>[44]</sup>通过对比研究肺囊性纤维化患者和非患者痰液中的病毒宏基因组, 发现两组痰液样品中病毒群体存在明显差异, 并且患者有机体内芳香族氨基酸代谢方式在不同患病阶段有所不同, 因而推测可以通过调节患者有机体内微生物群体结构达到治疗该病的目的。Svraka 等<sup>[45]</sup>采用病毒宏基因组学测序技术对通过肠病毒检测项目收集的 1 834 份病因不详的临床样品细胞培养上清液进行了病原筛查, 发现了多种病毒和新型 Saffold 病毒, 为临床诊断提供了重要信息。Law 等<sup>[46]</sup>通过对慢性乙肝、丙肝以及非酒精脂肪肝等患者血浆的临床研究, 认为病毒宏基因组学测序技术不但检测过程快速、检测结果可靠, 而且检测范围广泛, 检测样品可涉及胆汁、尿液和唾液等。与此同时, 世界第二大哺乳动物蝙蝠作为许多人畜共患病毒如埃博拉病毒、尼帕病毒、狂犬病毒和严重急性呼吸系统综合症病毒(SARS)等的携带者和传播者, 其粪便、肠道以及肺脏组织等样品均被用作病毒宏基因组学研究的对象, 为蝙蝠病毒相关传染病的预警提供了数据支持<sup>[47-48]</sup>。值得关注的是, 全球病毒组项目(Global Virome Project)拟于 2018 年启动, 该项目将通过调查鸟类和哺乳动物携带的病毒组, 重点发掘威胁人类健康的病毒, 以期将流行性病毒疾病防患于未然<sup>[49]</sup>。此外, 病毒宏基因组学测序技术也因其快速、灵敏和适用范围广等特点成为动物病毒学和植物病毒学研究的强大工具, 这些研究结果对动植物检疫以及抗病毒治疗具有重要指导意义。Day 等<sup>[50]</sup>采用病毒宏基因组学方法对患有肠道病毒性肠炎的火鸡群体开展研究, 发现了大量小 RNA 病毒、星状病毒以及呼肠孤病毒等, 推测火鸡生产性能降低可能是由多种病毒共同感染造成的。Shan 等<sup>[51]</sup>以猪粪便为研究对象, 获得了猪肠道病毒组的群落结构组成, 并发现了多种新病毒, 为新发病毒性疾病的诊断提供了参考。Yu 等<sup>[52]</sup>对发生病害的槟榔和

烟草植株样品的小 RNA 病毒进行宏基因组学分析, 其中 *Velarivirus* 属新病毒 APV1 的发现, 为揭示病毒与槟榔黄化病之间的关系提供了基础依据。

### 参考文献 References

- [1] STALEY J T, KONOPKA A. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats[J]. *Annual Review of Microbiology*, 2003, 39(1): 321-346
- [2] ALEGRÍA A, SZCZESNY P, MAYO B, et al. Biodiversity in oscypek, a traditional Polish cheese, determined by culture-dependent and -independent approaches[J]. *Applied and Environmental Microbiology*, 2012, 78: 1890-1898
- [3] AMANN R I, LUDWIG W, SCHLEIGER K H. Phylogenetic identification and in-situ detection of individual microbial-cells without cultivation[J]. *Microbiological Reviews*, 1995, 59(1): 143-169
- [4] BAKER G C, SMITH J J, COWAN D A. Review and re-analysis of domain-specific 16S primers[J]. *Journal of Microbiological Methods*, 2003, 55(3): 541-555
- [5] MARGULIES M, EGHOLM M, ALTMAN W E, et al. Genome sequencing in microfabricated high-density picolitre reactors[J]. *Nature*, 2005, 437(7057): 376
- [6] QAIL M A, KOZARWA I, SMITH F, et al. A large genome center's improvements to the Illumina sequencing system[J]. *Nature Methods*, 2008, 5(12): 1005-1010
- [7] 楼骏, 柳永, 李延. 高通量测序技术在土壤微生物多样性研究中的研究进展[J]. *中国农学通报*, 2014, 30(15): 256-260
- [8] LOU J, LIU Y, LI Y. Review of high-throughput sequencing techniques in studies of soil microbial diversity[J]. *Chinese Agricultural Science Bulletin*, 2014, 30(15): 256-260
- [9] CHEN C P, TSENG C H, CHEN C A, et al. The dynamics of microbial partnerships in the coral *Isopora palifera*[J]. *ISME Journal*, 2011, 5(4): 728-740
- [10] YE L, ZHANG T. Bacterial communities in different sections of a municipal wastewater treatment plant revealed by 16S rDNA 454 pyrosequencing[J]. *Applied Microbiology & Biotechnology*, 2013, 97(6): 2681-2690
- [11] LEININGER S, URICH T, SCHOLTER M, et al. Archaea predominate among ammonia-oxidizing prokaryotes in soils[J]. *Nature*, 2006, 442(7104): 806-809
- [12] HE Z L, XU M Y, DENG Y, et al. Metagenomic analysis reveals a marked divergence in the structure of belowground microbial communities at elevated CO<sub>2</sub>[J]. *Ecology Letters*, 2010, 13(5): 564-575
- [13] 赵柏霞, 潘凤荣, 韩晓日. 基于高通量测序技术的樱桃根际细菌群落研究[J]. *土壤通报*, 2018, 49(3): 596-601
- [14] ZHAO B X, PAN F R, HAN X R. Bacterial community development based on Illumina amplicon sequencing of 16S rDNA tag in the cherry rhizosphere[J]. *Chinese Journal of Soil Science*, 2018, 49(3): 596-601
- [15] SUN R B, LI W Y, DONG W X, et al. Tillage changes vertical distribution of soil bacterial and fungal communities[J].

- Frontiers in Microbiology, 2018, doi: 10.3389/fmicb.2018.00699
- [14] CHEN S, WANG F, ZHANG Y, et al. Organic carbon availability limiting microbial denitrification in the deep vadose zone[J]. Environmental Microbiology, 2018, 20(3): 980–992
- [15] HANDELSMAN J, RONDON M R, BRADY S F, et al. Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products[J]. Chemistry & Biology, 1998, 5(10): 245–249
- [16] BLAMEY J M, FISCHER F, MEYER H P, et al. Chapter 14 — Enzymatic biocatalysis in chemical transformations: A promising and emerging field in green chemistry practice[M]// BRAHMACHARI G. Biotechnology of Microbial Enzymes. Salt Lake City: Academic Press, 2017: 347–403
- [17] THOMAS T, GILBERT J, MEYER F. Metagenomics, a guide from sampling to data analysis[J]. Microbial Informatics and Experimentation, 2012, 2(1): 3
- [18] SCHUSTER S C. Next-generation sequencing transforms today's biology[J]. Nature Methods, 2008, 5(1): 16–18
- [19] GOODWIN S, MCPHERSON J D, MCCOMBIE W R, et al. Coming of age: Ten years of next-generation sequencing technologies[J]. Nature Reviews Genetics, 2016, 17(6): 333–351
- [20] 叶丹丹, 樊萌萌, 关琼, 等. 宏基因组研究的生物信息学平台现状[J]. 动物学研究, 2012, 33(6): 574–585  
YE D D, FAN M M, GUAN Q, et al. A review on the bioinformatics pipelines for metagenomic research[J]. Zoological Research, 2012, 33(6): 574–585
- [21] 程福东, 丁啸, 李晟, 等. 宏基因组样本数据的分析比较与分类[J]. 生物技术通报, 2016, 32(5): 1–10  
CHENG F D, DING X, LI S, et al. Analysis, comparison and classification of metagenomic samples[J]. Biotechnology Bulletin, 2016, 32(5): 1–10
- [22] WANG Q, FISH J A, GILMAN M, et al. Xander: Employing a novel method for efficient gene-targeted metagenomic assembly[J]. Microbiome, 2015, 3(1): 32
- [23] LOOMBA R, SEGURITAN V, LI W Z, et al. Gut microbiome-based metagenomic signature for non-invasive detection of advanced fibrosis in human nonalcoholic fatty liver disease[J]. Cell Metabolism, 2017, 25(5): 1054–1065
- [24] 朱真, 朱嗣博, 张铁军, 等. 宏基因组学与人类健康关系的研究进展[J]. 中国公共卫生, 2018, DOI: 10.11847/zgggws1118997  
ZHU Z, ZHU S B, ZHANG T J, et al. Association between metagenome and human health: A research progress review[J]. Chinese Journal of Public Health, 2018, DOI: 10.11847/zgggws1118997
- [25] 吴宇佳, 迟晓培, 陈峰, 等. 肥胖者唾液微生物宏基因组学特点[J]. 北京大学学报: 医学版, 2018, 50(1): 5–12  
WU Y J, CHI X P, CHEN F, et al. Salivary microbiome in people with obesity: A pilot study[J]. Journal of Peking University: Health Sciences, 2018, 50(1): 5–12
- [26] ZHAO L P, ZHANG F, DING X Y, et al. Gut bacteria selectively promoted by dietary fibers alleviate type 2 diabetes[J]. Science, 2018, 359(6380): 1151–1156
- [27] LIU Y, LIU C, NELSON W C, et al. Effect of water chemistry and hydrodynamics on nitrogen transformation activity and microbial community functional potential in hyporheic zone sediment columns[J]. Environmental Science & Technology, 2017, 51: 4877–4886
- [28] CHU B T T, PETROVICH M L, CHAUDHARY A, et al. Metagenomics reveals the impact of wastewater treatment plants on the dispersal of microorganisms and genes in aquatic sediments[J]. Applied and Environmental Microbiology, 2018, 84(5): e02168–17
- [29] 李晶, 杨洪一, 刘崇. 病毒宏基因组学的研究概况及应用[J]. 黑龙江农业科学, 2015, (12): 182–185  
LI J, YANG H Y, LIU C. Research and application of viral metagenomics[J]. Heilongjiang Agricultural Sciences, 2015, (12): 182–185
- [30] MANN N H. The third age of phage[J]. PLoS Biology, 2005, 3: e182
- [31] BREITBART M, SALAMON P, ANDRESEN B, et al. Genomic analysis of uncultured marine viral communities[J]. Proceedings of the National Academy of Sciences of the United States of America, 2002, 99(22): 14250–14255
- [32] ANGLY F E, Ben FELTS B, BREITBART M, et al. The marine viromes of four oceanic regions[J]. Plos Biology, 2006, 4(11): e386
- [33] MIDDELBOE M, JACQUET S, WEINBAUER M. Viruses in freshwater ecosystems: An introduction to the exploration of viruses in new aquatic habitats[J]. Freshwater Biology, 2008, 53(6): 1069–1075
- [34] 张奇亚, 桂建芳. 一类不可忽视的战略生物资源——淡水与海水中的病毒及其在生态系统中的作用[J]. 中国科学院院刊, 2009, 24(4): 414–420  
ZHANG Q Y, GUI J F. A kind of strategic bio-resource not to be neglected — Freshwater and marine viruses and their roles in the global ecosystem[J]. Bulletin of Chinese Academy of Sciences, 2009, 24(4): 414–420
- [35] LÓPEZ-BUENO A, TAMAMES J, VELÁZQUEZ D, et al. High diversity of the viral community from an Antarctic Lake[J]. Science, 2009, 326: 858–861
- [36] GE X Y, WU Y Q, WANG M N, et al. Viral metagenomics analysis of planktonic viruses in East Lake, Wuhan, China[J]. Virologica Sinica, 2013, 28(5): 280–290
- [37] 夏骏, 汪岷, 宫政, 等. 以宏基因组技术探讨渤海秋冬季病毒多样性[J]. 海洋与湖泊, 2016, 47(3): 572–580  
XIA J, WANG M, GONG Z, et al. Metagenomic study on viral diversity in autumn and winter in Bohai Sea[J]. Oceanologia et Limnologia Sinica, 2016, 47(3): 572–580
- [38] KIMURA M, JIA Z J, NAKAYAMA N, et al. Ecology of viruses in soils: Past, present and future perspectives[J]. Soil Science and Plant Nutrition, 2008, 54: 1–32
- [39] 韩丽丽, 于丹婷, 贺纪正. 土壤病毒生态学研究方法[J]. 生态学报, 2017, 37(6): 1749–1756  
HAN L L, YU D T, HE J Z. Research methods for soil viral ecology[J]. Acta Ecologica Sinica, 2017, 37(6): 1749–1756
- [40] ADRIAENSSENS E M, COWAN D A. Using signature genes

- as tools to assess environmental viral ecology and diversity[J]. *Applied and Environmental Microbiology*, 2014, 80: 4470–4480
- [41] 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
- [42] REAVY B, SWANSON M M, COCK P J, et al. Distinct circular single-stranded DNA viruses exist in different soil types[J]. *Applied and Environmental Microbiology*, 2015, 81(12): 3934–3945
- [43] HAN L L, YU D T, ZHANG L M, et al. Genetic and functional diversity of ubiquitous DNA viruses in selected Chinese agricultural soils[J]. *Scientific Reports*, 2017, 7: 45142
- [44] WILLNER D, FURLAN M, HAYNES M, et al. Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals[J]. *PLoS One*, 2009, 4(10): e7370
- [45] SVRAKA S, ROASRIO K, DUIZER E, et al. Metagenomic sequencing for virus identification in a public-health setting[J]. *Journal of General Virology*, 2010, 91: 2846–2856
- [46] LAW J, JOVEL J, PATTERSON J, et al. Identification of hepatotropic viruses from plasma using deep sequencing: A next generation diagnostic tool[J]. *PLoS One*, 2013, 8(4): e60595
- [47] DONALDSON E F, HASKEW A N, GATES J E, et al. Metagenomic analysis of the viromes of three North American bat species: Viral diversity among different bat species that share a common habitat[J]. *Journal of Virology*, 2010, 84(24): 13004–13018
- [48] 杨凡力, 王意银, 郑文成, 等. 中国部分地区蝙蝠携带病毒的宏基因组学分析[J]. *生物工程学报*, 2013, 29(5): 586–600
- YANG F L, WANG Y Y, ZHENG W C, et al. Metagenomic analysis of bat virome in several regions of China[J]. *Chinese Journal of Biotechnology*, 2013, 29(5): 586–600
- [49] CARROLL D, DASZAK P, WOLGE N D, et al. The global virome project[J]. *Science*, 2018, 359(6378): 872–874
- [50] DAY J M, BALLARD L L, DUKE M V, et al. Metagenomic analysis of the turkey gut RNA virus community[J]. *Virology Journal*, 2010, 7: 313
- [51] SHAN T L, LAN D L, LI L L, et al. Genomic characterization and high prevalence of bocaviruses in swine[J]. *PLoS One*, 2011, 6: e17292
- [52] YU H M, OI S S, CHANG Z X, et al. Complete genome sequence of a novel *Velarivirus* infecting areca palm in China[J]. *Archives of Virology*, 2015, 160(9): 2367–2370