# 环境微生物的宏基因组学实例与新发现

宏基因组学是直接研究环境样品中的遗传学材料，主体研究对象是微生物群落的基因组，即所有微生物个体基因组的总和。随着近15年以来DNA测序成本的大幅降低（> 105倍）和第二代高通量测序技术（NGS）大规模应用，宏基因组学（Metagenomics）正在快速发展成微生物学的一个新兴学科分支，其方法学正史无前例地转变着传统环境生物学与生态学的研究方式1,2。伴随着现代分子生物学方法与技术流程的不断自动化与商业化，当前微生物群落研究的时间与经济成本投入已从检测通量低、劳动力密集型的分子生物学“湿实验”转变为检测通量高、数据处理密集型的计算生物学“干实验”（例如生物信息和生物统计计算）2,3。

宏基因组学是打开通往未知微生物界大门的钥匙，广义上包括环境基因组、生态基因组学和群体基因组学的研究目标与内容。自然环境中超过99%以上的微生物在当前实验室条件下尚未被分离培养。相比于传统方法（如分离与培养）与现代分子生物学技术，宏基因组学的方法与技术优势体现在不依赖实验室分离培养来解码环境微生物组赋存的物种与基因多样性4。通过环境样品中所有遗传材料的分离与测序、生物信息学、生物统计学等多方法与技术手段可以直接研究微生物群落结构与功能，统称为“泛宏基因组学” (Bulk metagenomics)，以区分与带功能活性筛选（如功能宏基因组）或仅靶向部分微生物（如16S核糖体RNA扩增子测序）的组学策略与方法。泛宏基因组学无需对环境DNA样品进行PCR扩增，有效避免了引物设计与扩增偏好性问题对测序结果的干扰，更能真实的定性与定量研究微生物组的原位多样性与组成丰度的信息。

简言之，宏基因组学突破了环境微生物难培养的研究瓶颈，打开了微生物界物种和基因多样性的知识与资源宝库，促进了包括新物种、新酶基因（如CRISPR-Cas核酸酶）、新功能代谢过程（如完全氨氧化）与新活性物质（如抗生素）的发现。此外，针对自然或工程生态系统的大规模时间与空间尺度宏基因组学研究有助于探明微生物之间及微生物与环境之间的复杂相互作用，进而阐释群落构建机制，完善群落生态学理论4-8。过去10年基于高通量测序的宏基因组学技术被广泛应用于生物学与环境科学领域的前沿交叉学科研究，衍生了一系列新发现与重大成果；其中，部分代表性的研究成果可归纳为以下五类：1）新微生物物种的发现；2）新酶基因（簇）及功能的发现；3）新物质代谢途径的发现；4）微生物组-环境-效能新关联的发现；5）新生命进化树与物种分类新规则。

1. **新微生物物种的发现**

通过不同类型环境样品采集、核酸提取、泛宏基因组测序、基因组重构等技术手段有助于不断发现新的微生物物种。美国加州大学伯克利分校Jill Banfield教授是环境宏基因组学领域的先驱者与引领者9,10。她的团队通过利用泛宏基因组组装基因组（MAG）分析与转录组分析研究了科罗拉多河附近蓄水层的微生物组，发现了一类 “不寻常微生物”，命名为“候选门辐射群” (简称CPR) 。这类CPR微生物不同于传统观念中的原核生物：它们的基因组编码了外显子与内含子，还具有自我剪接的内含子及在其核糖体（rRNA）蛋白质编码基因。该发现更新了领域对原核生物组成的理解，相关成果于2015年在Nature上发表后引发了国际同行对包括CPR在内的未培养微生物暗物质的进一步关注和后续研究。

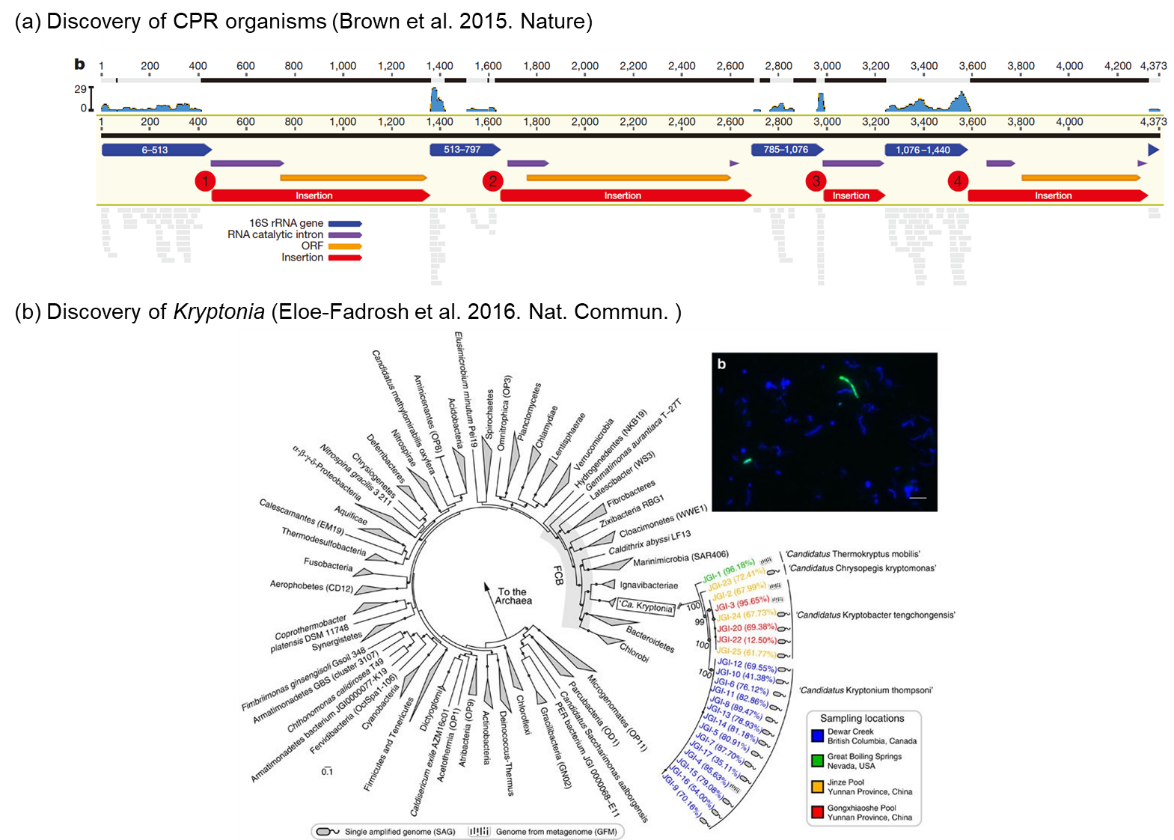


图 1. 候选门辐射群CPR微生物及*Kryptonia*候选门的发现。(a) CPR微生物基因组内内含子的16S rRNA编码区域。宏转录组序列覆盖了 16S编码区而非内含子插入片段区11。(b) “Candidatus *Kryptonia*”（绿色）的系统发育树和其在热泉沉积物中的荧光原位杂交（FISH）影像12。图片设计取材于文献。

宏基因组学研究产生了海量可开源获取的微生物组序列。相应地，针对全球主流公共数据库中宏基因组大数据的宏分析（meta-analysis）研究稳步成为新知识与新发现的重要途径。例如：来自美国能源部联合基因研究所JGI的研究团队通过JGI-IMG/M 数据库中5.2 TB宏基因组数据的重分析与深度挖掘，于2016率先发现了一种新的候选细菌门“Candidatus *Kryptonia*”。结果表明这类菌以异养方式生活，其仅在高温、 pH呈中性的地热泉中存在。该项研究是通过全球公共数据库中可开源获取宏基因组数据重分析与深度挖掘来发现新的微生物物种及其基因组编码的功能潜能的典型范例。

1. **新酶基因（簇）及功能的发现**

通过环境样品的大规模泛宏基因组学测序分析还有助于发现新的酶基因系统以及生物合成基因簇。美国UC Berkeley 地球与环境科学系Jill Banfield教授和生物化学与分子生物学系Jennifer Doudna教授团队开展交叉学科合作，解析了来自地下水、底泥、生物膜、土壤以及婴儿肠道等环境微生物样品中TB级别的泛宏基因组数据；通过组装、分箱等深度挖掘手段从得到了1.55亿个蛋白编码基因和大量MAGs序列13。该研究通过未培养微生物基于组信息的深度挖掘，鉴定第一个来自古菌域的Cas9蛋白，还在未培养细菌中发现了两种新的CRISPR-Cas系统，并将其命名为CRISPR-CasX和CRISPR-CasY系统，该成果于2017年发表在《自然》13。Doudna教授及其合作者率先在基于CRISPR-Cas系统的基因编辑领域取得了重大原创性突破，被授予2020年授予诺贝尔化学奖。目前，CRISPR-Cas9已发展成真核生物基因编辑“神器”。宏基因组学为将来从未培养环境微生物中发现了更多新CRISPR-Cas系统提供了强有力的工具，进一步拓宽了识别DNA序列的范围，为发展CRISPR-Cas系统以更好应用于生物技术领域和疾病治疗等提供了更多的可能和选择。

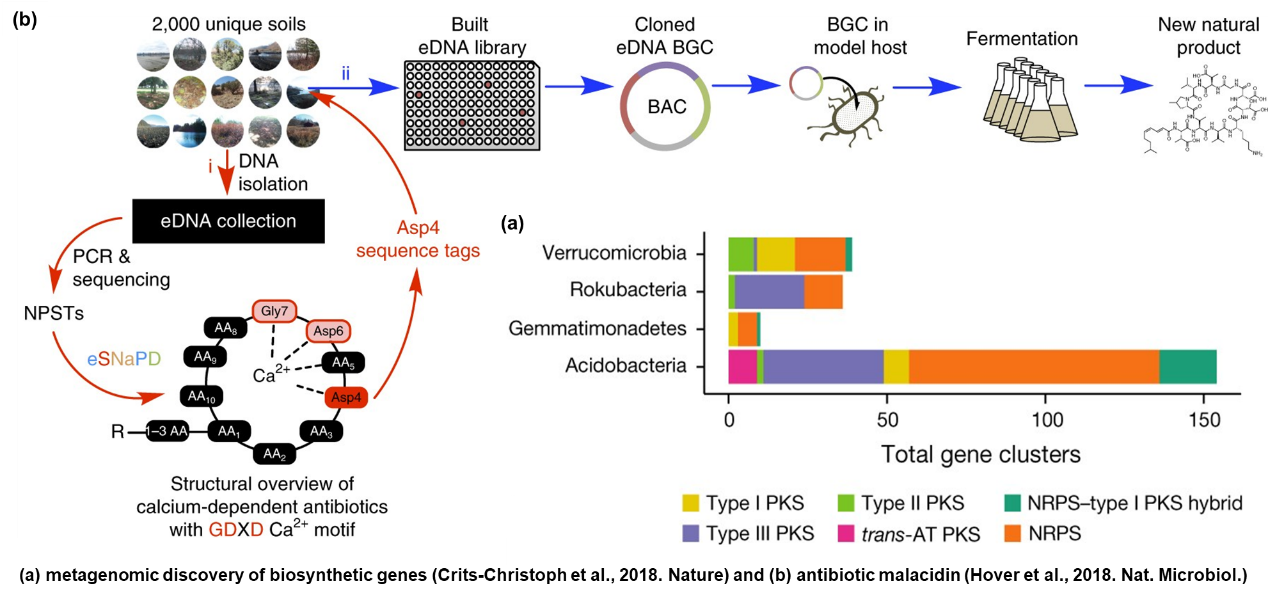


图2. 新型次级代谢产物生物合成基因簇及新型抗生素的发现。(a) 基于基因组重建从大于10 kb重叠群发现的NRPS 和 PKS合成基因簇。网络图的边连接基因簇之间共享的基因14。(b) 基于扩增子测序和克隆文库筛选从土壤微生物群落中筛选并发现一种全新的新型钙依赖环肽抗生素Malacidin 15。图片设计取材于文献。

此外， 新功能基因簇的发现也推动了新型抗生素的研发，当今大多数已知的细菌天然产物（包括抗生素）是从能被成功分离培养的放线菌、变形杆菌、 芽孢杆菌分离物中获取的，但它们仅占自然界中微生物群落的极少数，还有更多可以合成新抗生素的功能基因簇等待被发现和使用。例如，研究者通过大规模重建土壤中未培养微生物的基因组草图，发现了大量可合成如抗生素、抗真菌剂、免疫抑制剂等化合物的新型生物合成基因簇。来自美国洛克菲勒大学的研究团队利用扩增子测序结合BAC 文库筛选的方法从 2000 个土壤样本中发现了一个未知的抗生素合成基因簇并利用该基因簇合成了一种可杀灭多重耐药的革兰氏阳性致病菌的新型抗生素，将其命名为malacidins。因此，通过基于宏基因组学研究可以进一步拓展对于自然界中微生物生产抗菌类次级代谢产物合成能力的认知，进一步结合功能基因组学方法验证合成基因簇的功能活性为新型抗生素的发现提供了途径。

1. **新物质代谢途径的发现**

除了有助于新功能基因与酶系统的发现，宏基因组学还推动了领域对于未培养微生物的物质与能量代谢以及在全球生物地球化学循环中作用的全新认知的认知。比如，2015年奥地利Wagner Michael教授和荷兰Sebastian Lücker教授分别领导的独立研究团队Nature同期发表论文，声称发现了完全氨氧化微生物（Comammox）的存在。这两项独立研究均基于实验室反应器分别富集了来自水产养殖系统生物过滤器的生物膜和被热水覆盖金属管道表面的微生物生物膜样品，再利用泛宏基因组测序，同位素示踪技术结合MAG分析，最终证实了完全氨氧化*Nitrospira*具备独立将氨氮氧化为硝酸盐的能力。这一新发现解开了困扰微生物学界100多年的谜题，打破了“硝化作用过程需要两种不同微生物分两步完成”的传统观念，为领域探索Comammox在氮元素的生物地球化学循环过程中的作用奠定了基础。此外，基于泛宏基因组测序和MAG分析的方法也推动了硝酸亚还原厌氧甲烷化过程 、深古菌门等未培养微生物参与甲烷代谢过程的发现【插入文献】。简言之，将宏基因组学技术与生物反应器富集、微生物原位表征技术（例如稳定同位素示踪）等有机结合，是发现新功能微生物和揭示其代谢途径和生态功能的强有力手段与重要途径，可有效推动对未培养微生物的功能与代谢潜力的认知。

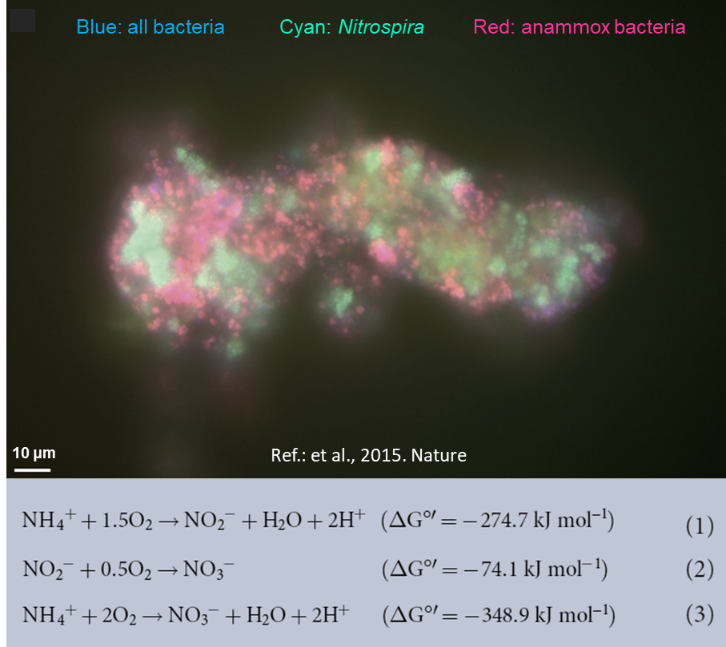


图3. 完全氨氧化菌*Nitrospira*与厌氧氨氧化菌*Brocadia*富集物的荧光原位杂交影像16。图片设计取材于文献。

1. 微生物组-环境-系统效能新关联的发现

宏基因组学揭开了自然界中微生物群落组成和功能多样性面纱，有望推动微生物组科学向生物工程技术应用的转化。例如通过土壤微生物群落组成定向提高作物的生产力，通过刺激自然产生或者人工投加微生物菌剂修复受污染的水体，通过富集培养特定功能的微生物实现从废水中回收资源等过程【插入文献】。探索微生物组、环境因素和系统功能之间的内部关联可以进一步将关于微生物结构与功能的微生物组科学知识更好的应用于改善社会生产力，以提高人类生活水平。例如，2015年来自香港大学研究人员利用细菌16S 核糖体RNA扩增子测序结合共现网络模型分析、多参数统计学分析等手段解析了城镇污水处理厂活性污泥工艺中的细菌群落、环境因素（如污水水质、运行参数等）以及处理效能三者之间的内在关联并提出了通过微生物组调控来减轻活性污泥起泡问题的工程措施与系统设计策略【插入文献】。近期，来自美国长期从事工程微生物组研究的科研团队提出了基于“设计-建造-测试-学习（DBTL）”的研究和技术开发体系，为微生物组在医药、农业、能源和环境方面的广泛应用提供了理论指导【插入文献】。其中，宏基因组学技术在捕获必要的生态学原理，把控微生物组群落组成，预测的群落功能等相关的设计、建造和测试环节中都发挥了重要作用。

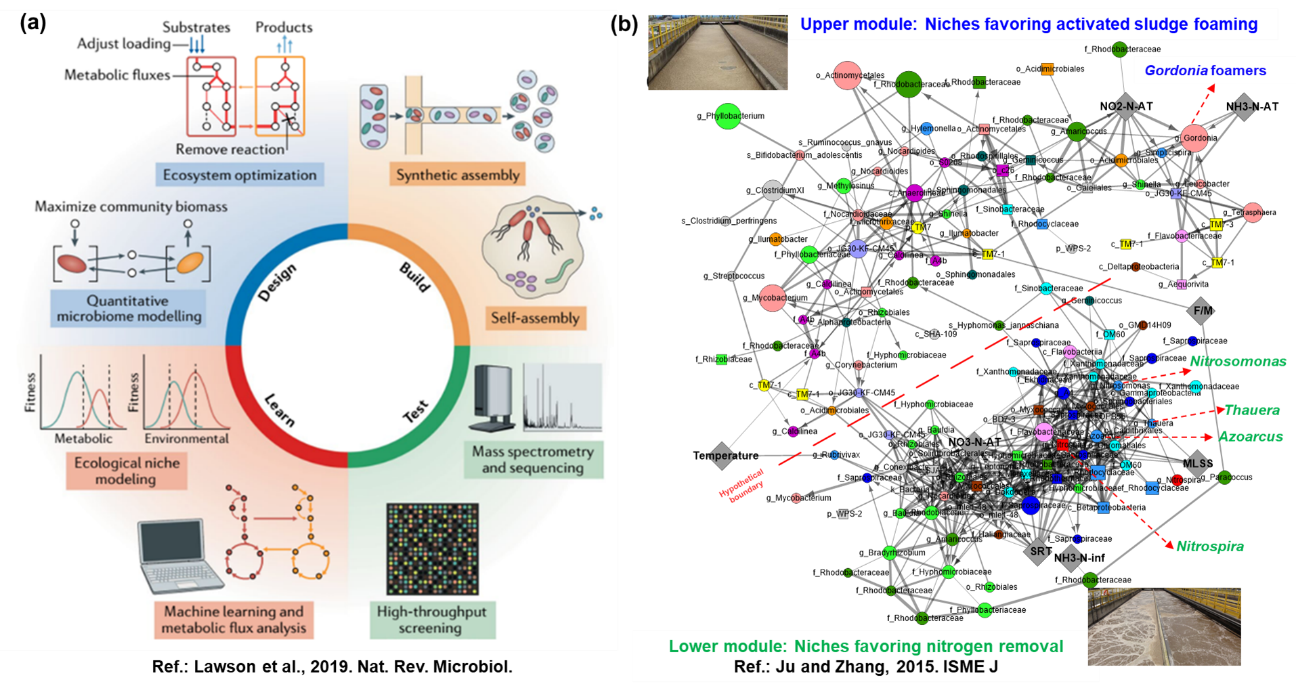


图4. 基于“设计-建造-测试-学习（DBTL）”的研究思路 (Lawson et al., 2019) 及“微生物组（M）-环境参数（E）-系统效能（P）”关联解析（简称“MEP 关联”）【插入文献】。图片设计取材于文献。

1. **新生命进化树与物种分类新规则**

真核生物的起源是进化生物学界研究的谜题与热点17。近年来瑞典乌普萨拉大学Thijs Ettema教授团队18,19和美国加州大学伯克利分校 Jill Banfield教授团队10,20基于环境样品采集、泛宏基因组学测序、大规模基因组构建与比较基因组学宏分析，发现了大量未培养环境微生物类群，提出了真核生物起源于古菌的新证据。例如：2015年，Spang等在《自然》发表文章，他们采集了北冰洋大洋中脊受到热液作用的深海沉积物样品，对其进行16S 核糖体RNA基因序列的宏基因组学分析，发现了一个全新的微生物门类名为*Lokiarchaeota*，系统发育树分析表明这是当时已知与真核生物亲缘关系最近的古菌生物，该发现填补了原核生物向真核生物过渡的进化学研究空白18。在后续的研究中，*Odinarchaeota*、*Heimdallarchaeota*和 *Thorarchaeota*也陆续被发现，它们与真核生物一同构成了Asgard 古菌门，并被认为是真核生物起源的谱系19。结合之前对TACK古菌的研究，新的真核生物起源的系统发生树被勾勒出来。这些发现表明真核生物起源于古菌，改写了从1977 年以来学界普遍认为的生命体划分为“真核生物、细菌和古菌”三大界的历史21。

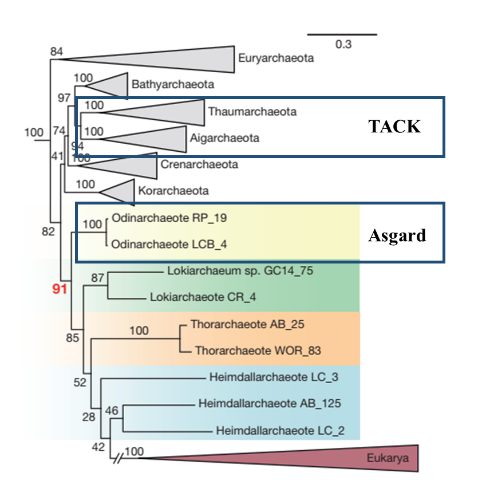


图5. Asgard 古菌的系统发育关系。图片设计取材于文献（18）

近年来，全球公共数据库中储存的宏基因组数据量在不断扩大，促使领域学者可以更好的利用来自不同环境的微生物样品的基因组序列开展宏分析研究，并基于此提出关于生命进化树的新见解，开发新工具。微生物的生命进化树与物种分类信息相互关联。昆士兰大学Hugenholtz 教授团队通过使用公认的平均核苷酸同源性（ANI）标准来设置物种界限并提出包含所有公共可获得的细菌和古细菌基因组的物种簇, 提出了一个可以为细菌和古细菌基因组提供完整的从域到种的分类学新框架。在这个新的分类系统下， NCBI 数据库中 58% 的物种分类信息将被更改，这也预示着将会有更多的对物种基因水平界限和相关生态学以及进化研究的新认知产生【插入文献】。目前，宏基因组数据分析中常用于基因组分类的GTDB数据库是基于单系和亲缘系进化的差异加以限定的细菌和古细菌类群基因组的广泛分类法，其中大约40%的基因组在NCBI数据库中缺乏种名【插入文献】。

尽管近10年宏基因组学在生物学与环境生态学研究中的应用催生了一系列重大的新发现，但是我们对环境赋存海量未培养微生物的认识依旧有限，仍然需要进行不断的方法学改进与理论探究。显而易见，微生物样品的收集、DNA提取以及宏基因组测序等过程会对组学研究结果产生影响，因此在制定标准规范化的操作流程的同时，针对不同特点的微生物样品的DNA提取等分子生物学流程进行优化以避免研究结果的偏差是必要的1。目前领域学者可较为容易的通过宏基因组组装重构大量基因组草图（MAGs）。尽管通过未培养微生物基因组学分析可有效预测其功能与代谢潜能以及微生物群落内物种相互作用，但是DNA水平的潜能不等同于功能与活性。因此，迫切需要新的微生物组学研究手段实现微生物组成与功能的耦合研究。鉴于此，需要进行更多的研究以提供原位生理功能测定，将微生物群落的物种与基因多样性信息与真实功能进行关联。目前已有多种方法可以用来量化微生物群落的活性，例如稳定性同位素探针技术、转录组学、培养组学、代谢组学和蛋白质组学。随着更多的表征手段和新方法的有机结合，宏基因组学必将为环境微生物学领域研究带来更多的新发现和新惊喜。

参考文献：

1 Ju, F. & Zhang, T. Experimental design and bioinformatics analysis for the application of metagenomics in environmental sciences and biotechnology. *Environ. Sci. Technol.* **49**, 12628-12640 (2015).

2 Liu, Y.-X. *et al.* A practical guide to amplicon and metagenomic analysis of microbiome data. *Protein Cell* (2020).

3 Ju, F. & Zhang, T. 16S rRNA gene high-throughput sequencing data mining of microbial diversity and interactions. *Appl. Microbiol. Biotechnol.* **99**, 4119-4129 (2015).

4 刘永鑫, 秦媛, 郭晓璇 & 白洋. 微生物组数据分析方法与应用. *遗传* **41**, 845-862 (2019).

5 鞠峰 & 张彤. 活性污泥微生物群落宏组学研究进展. *微生物学通报* **46**, 2038-2052 (2019).

6 Ju, F. & Zhang, T. Bacterial assembly and temporal dynamics in activated sludge of a full-scale municipal wastewater treatment plant. *The ISME journal* **9**, 683 (2015).

7 Ju, F., Xia, Y., Guo, F., Wang, Z. & Zhang, T. Taxonomic relatedness shapes bacterial assembly in activated sludge of globally distributed wastewater treatment plants. *Environ. Microbiol.* **16**, 2421-2432 (2014).

8 Hu, A. *et al.* Strong impact of anthropogenic contamination on the co‐occurrence patterns of a riverine microbial community. *Environ. Microbiol.* **19**, 4993-5009 (2017).

9 Tyson, G. W. *et al.* Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**, 37-43 (2004).

10 Castelle, C. J. & Banfield, J. F. Major new microbial groups expand diversity and alter our understanding of the tree of life. *Cell* **172**, 1181-1197 (2018).

11 Brown, C. T. *et al.* Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* **523**, 208-211 (2015).

12 Eloe-Fadrosh, E. A. *et al.* Global metagenomic survey reveals a new bacterial candidate phylum in geothermal springs. *Nat. Commun.* **7**, 1-10 (2016).

13 Burstein, D. *et al.* New CRISPR–Cas systems from uncultivated microbes. *Nature* **542**, 237-241 (2017).

14 Crits-Christoph, A., Diamond, S., Butterfield, C. N., Thomas, B. C. & Banfield, J. F. Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis. *Nature* **558**, 440-444 (2018).

15 Hover, B. M. *et al.* Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens. *Nature microbiology* **3**, 415-422 (2018).

16 van Kessel, M. A. *et al.* Complete nitrification by a single microorganism. *Nature* **528**, 555-559 (2015).

17 Williams, T. A., Foster, P. G., Cox, C. J. & Embley, T. M. An archaeal origin of eukaryotes supports only two primary domains of life. *Nature* **504**, 231-236 (2013).

18 Spang, A. *et al.* Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**, 173-179 (2015).

19 Zaremba-Niedzwiedzka, K. *et al.* Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353-358 (2017).

20 Hug, L. A. *et al.* A new view of the tree of life. *Nature microbiology* **1**, 1-6 (2016).

21 Eme, L., Spang, A., Lombard, J., Stairs, C. W. & Ettema, T. J. Archaea and the origin of eukaryotes. *Nat. Rev. Microbiol.* **15**, 711 (2017).