

DOI: 10.13376/j.cblls/2023178

文章编号: 1004-0374(2023)12-1630-09



肖湘, 上海交通大学特聘教授, 国家杰出青年基金获得者, 深部生命国际研究中心主任, 主攻深海极端环境生命研究。2001年起, 肖湘教授在国内率先开展了高压下的深海微生物研究, 建设了具有自主知识产权的全海深深海微生物研究平台体系。曾担任中国大洋深海考察第一个生物基因航段首席科学家、中国载人深潜西南印度洋热液区航段指挥部科学家、国际首次环大洋洲载人深潜科考航次第一航段联合首席科学家, 迄今共完成五次载人深潜任务, 先后完成西南印度洋热液区、雅浦海沟、马里亚纳海沟及克马德克海沟底部调查取样工作, 抵达过马里亚纳海沟挑战者深渊、克马德克海沟底部及代表人类首次到达雅浦海沟最深点。2020年起承担国家自然科学基金委第一个深海生命的创新群体项目“深海及深部生物圈”; 2021年10月发起并领衔了深渊环境与生态研究 (MEER) 计划。

大数据时代的深渊微生物研究

王 景¹, 赵维姝^{2,3}, 肖 湘^{2,3*}

(1 上海交通大学海洋学院, 上海 200030; 2 上海交通大学生命科学技术学院, 微生物代谢国家重点实验室, 深部生命国际研究中心, 上海 200240; 3 上海交通大学三亚崖州湾深海科技研究院, 三亚 572024)

摘 要: 近年来, 我国载人、无人深潜技术和深海装备领域的重大突破, 使得大范围、系统性地获得深渊沉积物和水体样本成为了可能。以测序技术为代表的多组学技术的大力发展, 以及人工智能在生命科学领域的前沿探索, 使得深渊微生物的研究迎来了大样本量、大数据量、多数据类型和多学科交叉的大数据时代。充分利用大数据所提供的优势解析深渊微生物所涉及的生物学和生态学科学问题, 并进一步通过微生物大数据为海洋学和地球科学提供新的线索, 是新时代所面临的关键机遇与挑战。该文对深渊微生物研究的历程与现状进行了小结, 并通过对标国际人类微生物组、地球微生物组以及 Tara Oceans 计划等其他先行的微生物大数据研究范例, 对大数据时代开展深渊微生物研究的新思路和新范式进行了梳理和展望。

关键词: 深渊环境; 微生物组; 大数据; 新技术与新方法

中图分类号: P73; Q93; TP39 **文献标志码:** A

Hadal microbial research in the age of big data

WANG Jing¹, ZHAO Wei-Shu^{2,3}, XIAO Xiang^{2,3*}

(1 School of Oceanography, Shanghai Jiao Tong University, Shanghai 200030, China; 2 State Key Laboratory of Microbial Metabolism, International Center for Deep Life Investigation (IC-DLI), School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China; 3 SJTU Yazhou Bay Institute of Deepsea Sci-Tech, Sanya 572024, China)

Abstract: The recent advances in China's deep-submarine and deep-sea equipment technologies, both manned and unmanned, have made large-scale and systematic collection of hadal sediment and water samples possible. The

收稿日期: 2023-12-09

基金项目: 国家重点研发计划(2022YFC2805300, 2022YFC2805404); 国家自然科学基金创新研究群体项目(41921006)

*通信作者: E-mail: zjxiao2018@sjtu.edu.cn

rapid advancement of multi-omics technology, particularly sequencing technology, and the cutting-edge implementation of artificial intelligence in the life sciences field have inaugurated the big data era. This era encompasses extensive sample sizes, copious amounts of data, diverse data types, and cross-disciplinary research in the study of hadal microorganisms. Making full use of the benefits offered by big data to examine the scientific concerns of biology and ecology related to hadal microorganisms and generating fresh leads for oceanography and earth sciences via microbial big data present vital opportunities and challenges in this new era. Additionally, we analyze and anticipate new concepts and paradigms emerging in the era of big data by comparing them with other pioneering microbial big data research projects, such as the Human Microbiome Project, the Earth Microbiome Project, and the Tara Oceans Program.

Key words: hadal environment; microbiome; big data; new method and technology

深渊一般指水深超过 6 000 米的海洋环境, 主要由地质板块交界处的 37 条深渊海沟组成^[1]。其中, 马里亚纳海沟挑战者深渊水深接近 11 000 米, 静水压力接近 1 100 个大气压 (115 MPa), 是已知的地球表面最深点^[2-3]。深渊一度是人类的禁区, 至今仍是地球上少数未被人类完全探索的极端环境之一。除极端高压以外, 深渊环境平均水温仅约 2.5 °C, 同时由于有机质输入有限, 原位又缺乏光照驱动光合自养作用, 因此深渊环境过去长期被认为是海洋中的生命荒漠^[3-4]。然而近十年来的研究却发现, 相对于平均水深 6 000 米的深海平原, 深渊海沟中汇集了大量的新鲜有机质, 孕育了活跃的深渊微生物生态系统, 具有相对快速的有机碳周转^[4-7]。生存在深渊极端环境中的微生物群落不仅具有活跃的代谢功能^[6, 8-11], 而且具有非常高的生物多样性^[12-18]和物种新颖性^[16, 19]。对深渊微生物及其所处极端环境开展系统性的研究, 将有助于我们进一步认识和回答深渊独特微生物资源的挖掘潜力、微生物极端环境的适应机制、深渊微生物参与地球元素循环的新特征与新过程、类早期地球环境生命过程和生命起源等一系列重要科学问题^[1, 4, 20-21]。

系统研究深渊微生物可以回答生命科学和生态学的一些基本问题, 如深渊是否存在大量未知的生物类群, 深渊生物细胞如何抵抗极端高压, 其中哪些类群和代谢功能具有活性, 哪些类群和代谢功能被高压抑制, 深渊微生物生态系统如何发端, 多样性和新颖性的成因是什么, 深渊环境是否蕴含陆地和非深渊海洋所没有的潜在微生物资源等; 同时, 也需要回答深渊生态系统与地球和海洋环境的关系, 如深渊微生物圈的营养和能量来源于何处, 深渊微生物的代谢活动如何参与深渊环境甚至海洋环境中的碳、氮、硫、磷等元素的循环活动, 与全球气候的变化有怎样的关系等; 此外, 研究者们更期

待通过对深渊微生物的研究, 在全新的角度上获得关于生命科学与地球科学的新认知。

1 深渊微生物研究发展历程与现状

1.1 发展历程和现状

深渊环境的现代科学研究肇始于 20 世纪 50 年代丹麦组织的“Galathea II”和苏联组织的“Vitjaz”全球调查航次^[22]。但对深渊生态系统的探索长期受到设备的限制, 主要使用拖网和抓斗等方式捕获深渊鱼类、钩虾等宏生物, 针对其共生微生物开展了部分研究, 深渊微生物分布特征和环境意义长期以来并不明确^[3]。到本世纪初, 随着深渊采样、检测技术的发展, 研究者们对于深渊生态系统特别是微生物生态系统开始有了基本认知。

深渊生物地球化学研究首先通过对微生物群落代谢活动的定量检测颠覆了深渊环境是生命荒漠的传统认知^[3-4, 20]。2013 年, Glud 等^[5]发现相对于平均水深 6 000 米的深海平原, 马里亚纳海沟挑战者深渊底部沉积物具有更高的有机质含量、更新鲜的有机质组分、更高的微生物含量, 以及显著更快的氧消耗速率, 表明海沟深渊底部孕育了活跃的微生物生态系统。利用类似的原位检测和采样设备, 研究者们先后分别在伊豆-小笠原海沟、汤加海沟、马里亚纳海沟、穆绍海沟、新不列颠海沟、克马克海沟、阿塔卡马海沟等开展了系统性的沉积物氧气消耗实验, 结果显示海沟深渊中的微生物代谢速率不仅普遍高于普通深海平原 2~3 倍, 在不同海沟之间和同一海沟的不同区域间也呈现出高达 10 倍的差异, 且可能受到有机质输入量和有机质组分类型的调控^[6-7, 23]。研究者已初步观察到在水深超过 6 000 米的深渊区域, 沉积物微生物群落活跃程度随水深增加而提高, 这一趋势与表层到深海区域微生物活性逐渐降低的大趋势相反, 但该反转趋势的

原因目前仍未得到证实。

此外,综合多种生物地球化学指标,我们对支撑活跃的深渊微生物生态系统的有机质来源及分布特征有了初步的认识。通过测定深渊环境样品中的总有机碳、氮含量及其同位素^[24-26],有机质组分如蛋白质、色素等的含量和比例^[5-6, 27],长链烷烃、支链或不饱和脂肪酸的类型和含量^[25, 28],甘油二烷基甘油四醚(GDGT)的组成^[28-31],以及沉积物中铅-210同位素、稀有金属含量等信息^[32-35],研究者推测深渊中富集的有机质可能主要包括四个来源:表层海洋生物圈中产生的海源有机质沉降入深渊、来自陆地生态系统的陆源有机质沿洋底的陆坡受侧向搬运作用汇集到深渊、深渊生态系统原位化能自养过程以及大型海洋生物死亡后形成的鲸落^[4, 36]。目前推测深渊微生物代谢活动的强弱和异质性可能受到海沟所处海区初级生产力水平、海沟距大陆或大型岛屿的距离,以及海沟内部不同区域地形特征及水岩反应等潜在能量供应因素的共同影响^[6, 37]。

对于深渊微生物生态系统的研究,早期主要基于分离培养实验和16S rRNA基因扩增子测序两种技术手段。深渊微生物中潜在的异养微生物类群如Actinobacteria、Bacteroidetes、Chloroflexi、Alphaproteobacteria、Gammaproteobacteria、Gemmatimonadetes、Planctomycetes、SAR406、Woesearchaeota等占主导^[9, 12, 15, 17, 38-41];深渊异养菌株降解明胶、淀粉、DNA等多种大分子有机质及降解难利用的D型氨基酸的功能也得到了基于纯菌的实验验证^[42-43];少数基于宏基因组的研究也显示出深渊微生物中还广泛分布有降解碳水化合物的基因(如糖基转移酶和糖苷酶等)、降解烷烃的基因(如*alkB*和*almA*等)、降解芳香族化合物的基因(如*catE*和*hcaC*等)^[41, 44-45]。在深渊微生物中同时也发现了具有较高丰度的氮代谢微生物类群,包括氨氧化古菌Nitrososphaeria、亚硝化细菌Nitrospira和Nitrospina、厌氧氨氧化菌Brocadiales,以及分布在Gammaproteobacteria中的多种潜在反硝化类群等^[9, 17-18, 46]。研究推测深渊微生物生态系统可能具有活跃的碳、氮周转潜力^[8, 10-11, 15, 18]。

此外,也有部分研究对微生物适应深渊环境的机制进行了初步探索,其技术手段主要依赖纯菌的高压培养实验和比较基因组学。已经发现深渊微生物在高静水压环境下表现出细胞水平的功能变化,涉及到DNA的损伤修复、细胞壁与细胞膜的组成、抗氧化、压力渗透因子合成、重金属耐受等多种多

样的机制^[21, 47-50];并且发现深渊微生物群落的组成和功能可能受到有机质输入量和氧化还原环境的调控^[17-18]。尽管如此,深渊如何塑造独特的微生物生态系统,微生物如何适应深渊区域60~115 MPa的极高静水压力,哪些类群和功能在深渊环境中保持活跃,深渊微生物生态系统如何记录和参与全球变化,这一系列问题至今仍未得到系统的回答。

1.2 我国深渊工程技术发展带来的新机遇

近十年来我国深渊工程技术取得了长足的发展,形成了以全海深级载人潜水器(human-occupied vehicles, HOV)“奋斗者”号、7 000米级“蛟龙”号、4 500米级“深海勇士”号为代表的有人潜水器和无人遥控潜水器(remotely operated vehicles, ROV)、自主潜水器(autonomous underwater vehicles, AUV)、着陆器等相互配合的深海装备集群^[1, 51]:HOV能够帮助下潜科学家实际体验深渊环境的各种特征,可以根据下潜科学家的现场决策实现精确到厘米级别的环境采样,样品携带能力和环境探测设备搭载能力也超过其他设备;ROV运行和维护成本更低,科研人员同样可以通过线缆获取实时的观测数据并进行远程决策,适合在高价值区域进行频繁作业;AUV可以在更大范围进行自主巡弋,快速扩充深渊探测的覆盖面;着陆器则可以搭载各型探测和实验设备,实现深渊原位环境的长时间尺度观测和实验验证工作。

相较于传统的抓斗、箱式采样器、多管采样器、重力采样柱、采水器等船基采样设备,深渊装备的发展升级带来的是科学范式的改变:以“奋斗者”号载人深潜器为例,采样精度提升四个数量级(百米级到厘米级),实现对深渊特殊位点样本的精确采集;采样规模提升两个数量级(个位数到数百份),实现对深渊样本的大范围覆盖;设备迭代提升两个数量级(年级到天级),实现装备根据科考需求进行现场调试和更新;同时实现了现场决策的从无到有,研究对象更具有针对性与代表性。与此同时,原位技术和模拟技术的发展使得关键科学问题和假设得以在可控条件下加以验证。新型的调查、采样方式配合实验室模拟技术和高通量组学检测技术,有望在不久的将来获得更多的深渊样本,实现系统的深渊微生物大数据研究。

1.3 组学研究手段在深渊微生物中的初步应用

宏基因组、宏转录组、宏病毒组等分析手段对微生物研究产生了深远影响,相较于经典的基于16S rRNA基因序列差异的微生物多样性分析手段,

这些组学新技术不仅使得微生物研究的分辨率由属水平提升到种,甚至更精细的水平,而且能够建立微生物物种和功能之间的直接对应关系。对这些组学新技术的应用探索是深渊微生物研究的发展趋势。Zhou等^[19]通过着陆器和箱式采样等手段采集了马里亚纳海沟挑战者深渊及其周边5 000~1 000米水深13个站点共计37份沉积物样本,通过高通量测序共获取了586个高质量的宏基因组组装基因组(metagenome-assembled genomes, MAGs),是目前已发表的最大规模的深渊微生物数据集。该工作不仅揭示了深渊微生物群落具有较高的物种新颖性,并且结合宏转录组数据展示了深渊微生物代谢重金属的潜力和整体代谢功能在不同地理位置的异质性分布。Chen等^[44]综合使用地球化学检测技术、宏基因组测序和培养实验技术,对从挑战者深渊底部获取的沉积物样品开展了综合分析。此项工作从宏基因组数据中获取了覆盖26个门的178个深渊微生物MAGs,其中16个门是首次在深渊沉积物中发现;获取了6个可能参与深渊氮、硫元素循环的真菌类群;同时还发现了深渊中独特的病毒类群和特征性的病毒-宿主互作关系;除此之外,还分离得到了1 070株细菌和19株真菌,扩充了深渊来源微生物物种资源库。Jian等^[16]通过对已发表的马里亚纳海沟、雅浦海沟、克马德克海沟深渊微生物宏基因组数据重分析,构建了海洋深渊病毒基因组数据集(oceanic trench viral genome dataset, OTVGD),其中超过99%的病毒可操作分类单元(viral operational taxonomic unit, vOTU)和超过50%的病毒蛋白簇(protein cluster, PC)为其他病毒数据库中未包含的新结果。此项工作不仅展示出深渊病毒的新颖性、迁移动态以及随生境变化的多样性特征,同时还显示出深渊病毒有潜力通过对宿主代谢过程的重编辑来影响深渊生态系统和元素循环过程。

现有数据向我们初步展示了独特的、动态的、多样化的深渊微生物生态系统的冰山一角。但现有深渊微生物组学研究的样本量多为几个到十几个位点^[19, 44],数据量多在300 GB以内,采样位点在深渊的覆盖度、精细度及采样量等均比较有限。同时,深渊环境样本具有异质性高、前处理困难等问题,因此相较于其他环境的微生物研究,多组学的协同研究在深渊微生物中进展缓慢。为了系统地描述深渊微生物群落的整体物种和功能特征,阐释海沟与非海沟区域、不同海沟区域之间,以及海沟内不同

地形地域之间的物种和功能异质性^[12, 17, 19, 52-53]的成因,探究特殊功能类群在深渊环境中的具体代谢特征和环境作用^[54-56],需要从大数据产出的根本入手,借鉴已成功开展大数据研究的其他环境微生物计划蓝本,探索以微生物大数据为基础的深渊研究创新体系和理论框架。

2 深渊微生物研究面临的挑战

2.1 科学认知的挑战

在全球海洋环境中,海水表层初级生产力产生的有机质碎屑在向海底沉降的过程中逐级消耗,最终平均只有不到1%的有机质能够输入到平均水深6 000米的深海沉积物中。海底沉积物中的有机质含量随所处水深的增加呈递减趋势,伴随着微生物总量和代谢活动的减弱,微生物活动也从近岸的以厌氧和还原性过程为主,过渡到深海平原的以好氧过程为主^[57-58]。而在水深超过6 000米的深渊区域中,有机质输入量随水深而升高^[5, 32, 36],微生物的代谢活动特征有可能发生逆转^[8, 10]。此外,对氧气消耗速率的测量忽略了厌氧呼吸的贡献,有可能明显低估了深渊微生物的代谢活性,但目前对于深渊微生物过程特别是厌氧过程的定量认识仍然非常有限^[10-11]。不仅如此,通常认为浅层海水微生物以压力敏感型为主,沉降到深渊高压环境下后生长和代谢活动将受到抑制^[21, 59-60],但最新的证据显示在深渊沉积物中对氮转化有主要贡献的厌氧氨氧化细菌类群却与浅水类群高度同源^[10]。此外,我们发现深渊的极端静水压力导致了微生物细胞内部的过氧化应激,但这一过程驱动深渊微生物适应性演化的机制尚不清楚^[21]。种种迹象表明,深渊环境与普通深海环境之间存在不可忽视的差异,且现有的理论很难解释,暗示在海洋非深渊环境中总结出的推论在深渊环境中可能并不适用,需要以大数据为基础,深入探索深渊生命过程及微生物生态系统的特征,重新获得具有统计学意义的规律。

2.2 技术方法的挑战

实现深渊到达以后,深渊微生物大数据研究同样还面临着一系列共性和特性问题。(1)采样精度的挑战:船基采样技术在万米深渊的精度误差往往在百米级,难以满足对深渊精确采样和分析的需求,需要依托潜水器实现深渊环境与生态的精细刻画。(2)大样本量处理和检测通量的挑战:传统依赖手工的处理方式不能满足深渊研究的特殊需求(低温、高压、厌氧、洁净度等),需要提高标准化和自动

化水平,以满足深渊微生物大数据研究的通量和时效要求。(3) 大数据存储和计算的挑战:目前的深渊总数据规模已经从吉字节(giga byte, GB)级上升到了万亿字节(tera byte, TB)级(表1),并随着高通量测序技术的快速发展将进入千万亿字节(peta byte, PB)级。如此大的数据规模对于高性能计算资源和存储资源的占用是非常可观的。(4) 多数据类型对分析工具和数据集成的挑战:深渊微生物数据的分析不仅需要整合基因组、转录组、蛋白质组等生物数据,还需要紧密结合地质、地球物理、地球化学等多学科环境理化数据。这些数据在存储格式、空间尺度、时间跨度等多个层面存在巨大差异。需要开发适应不同数据类型和时空跨度的数据分析集成新工具、新流程和新算法。

3 国际微生物大数据研究计划给深渊微生物研究的启示

以人类微生物组计划(Human Microbiome Project, HMP)^[61-62]、地球微生物组计划(Earth Microbiome Project, EMP)^[63-64]、Tara Oceans 计划^[65]为范例,大数据在微生物学研究中已经有一系列的成功应用。我们对其中的共性特征进行了系统性梳理,为深渊微生物大数据研究提供参考(表1)。

3.1 微生物大数据研究的三个阶段

第一阶段:大范围的基本微生物图谱绘制,获取“静态快照”。无论是HMP的健康人群及其共生微生物研究队列^[66],EMP的全球微生物“种子库”数据集^[67-68],还是Tara Oceans的全球海洋微生物调查^[65],首要目标都是通过大样本量、大数据量的积累,获取微生物组的全貌信息,并通过相关分析和差异分析,构建目标研究环境中微生物组成特征和形成机制的初步理论模型。

第二阶段:关键研究对象的长时序跟踪,重构“因果性”。以HMP第二阶段为例,研究计划聚焦于早产儿、炎症肠病、糖尿病前期三个队列进行持续的多组学检测^[69],以时序信息过滤大数据中的噪声^[70],将单时间点相关性扩展为时序因果性,进一步构建微生物与环境的互作机制和因果关系。

第三阶段:多学科交叉,验证数据爆炸引发的新理论、新猜想。大数据催生的新理论框架最终应该通过实验手段,在原位环境或模拟体系里加以验证和迭代。最终根据更新的理论框架,实现以微生物组为靶点的宿主(环境)改造^[71-72],完成微生物理论在健康、环境等领域的应用转化。

3.2 研究的可重复性依赖于标准化建设

可重复性低是目前微生物组学研究中最为突出的问题之一。这由几个原因共同导致,主要包括(1)环境异质性和个体差异;(2)数据产出平台的差异;(3)数据分析流程的差异;(4)统计方法使用不当产生的假阳性。以大数据分析需求作为驱动,催生新技术和新方法,并推动建立标准化的数据产出和分析流程,是提升微生物组学研究可重复性的基础保障。比如,从HMP和EMP中孵化出的16S rRNA扩增子测序数据的标准化综合分析平台QIIME2就是一个成功的例子^[73]。而目前尚未出现统一的基于鸟枪法测序的宏基因组和宏转录组分析流程和平台,蛋白质组和代谢组更是往往停留在小样本手工处理的阶段。未来在这个方向的建设,将有助于提升微生物研究工作的可重复性,形成可靠的理论体系框架^[63, 74-75]。

3.3 研究的系统性离不开科学共同体

伴随大数据而来的技术挑战决定了微生物领域的大数据研究越来越不可能依靠单一团队完成,而是逐渐演变成工程、技术、科学团队紧密结合的系统性研究工程,并在项目推进的过程中逐步实现跨学科、跨地域、多中心、分布式的科学共同体架构。HMP和EMP公开样本采集、处理、分析的标准化流程,鼓励领域内科学家遵循相同的技术规范产生数据并汇交到统一的平台进行比较研究,这一研究范式的转变大大提升了可比较、可重复的微生物组学数据的扩张速度,使得微生物领域共性法则的研究成为可能^[76]。未来受益于元宇宙、云计算、人工智能等新兴技术,微生物研究的范式将有望迎来新的蜕变。

4 大数据时代的深渊微生物研究展望

4.1 大数据量变推动深渊微生物知识质变

我们认为深渊微生物研究应当首先变单钩钓鱼为细网捞鱼,从以功能基因和已知功能类群为指标,到多学科多层次的全面覆盖。可以开展基于以下三方面的工作:(1)依赖于多组学手段对比的大数据研究,区分生命潜能与实际功能,系统性地揭示深渊生命过程与生态系统的分布结构与时空变化;(2)生态学理论与环境适应相结合,从简单的结构与功能描述,发展到对深渊微生物群落特征的解构与重塑;(3)研究对象贯通宏生物和微生物(原核、真核、病毒),研究层次贯穿生态系统、细胞与生物大分子,以深渊为切入点,完整解析地球环境对生命的塑造。

4.2 深渊大数据与高通量实验相结合

同时, 我们还认为深渊大数据的研究需要与高通量实验相结合, 从大数据获得的具有统计学意义的相关性规律, 需要实验验证解析其背后的生物学机制、建立因果关系。当前的结合方式可以围绕以下三方面: (1) 通过大数据指导获取深渊特有的高阶分类单元的微生物种群; (2) 通过实验室模拟验证生态学理论与深渊适应性的关系; (3) 深渊原位实验研究人类活动对深渊环境的影响。此外, 大数据还需要通过人工智能 (AI) 算法与模型提取深渊环境与生命的高维特征, 拓展对生命科学和地球科学的认知。

4.3 深渊大数据研究需要新工具、新方法、新技术

最后, 我们建议深渊微生物大数据研究应当抓住当前我国深渊采样装备领先的先机, 不断发展新技术、新工具、新方法, 带来深渊环境与生命的系统性认知。这需要着重以下三个方面: (1) 需要建立全流程贯穿的技术体系, 能够打通从原位采样、甲板实验、样本处理、实验室检测到大数据的信息学分析, 利用好深潜装备体系和实验流程与分析流程的搭配, 实现深渊生态环境从粗粒度描绘到精确刻画的跃升; (2) 在全流程贯穿的基础上, 向成熟的工程化体系取经, 建立标准化的深渊样本和数据的采集和处理流程, 并逐步实现自动化, 满足样本高通量处理的需求, 保证深渊样本从采集到数据分析的可重复性及不同批次样本的可重复性; (3) 需要针对深渊样本的特点和科学问题的需求不断发展和迭代新技术, 如深渊微生物长周期原位监测技术、多组学 (宏基因组、转录组、蛋白质组、代谢组) 共提取与共分析技术、聚焦稀有高阶分类单元的高通量深渊微生物分离培养技术、基于人工智能的深

渊微生物高维特征分析技术, 以及计算机模拟深渊生命动态过程及预测技术等。总之, 大数据时代的到来为深渊微生物研究提供了无限的可能, 深渊微生物大数据的获取与整理运用需要解决自动化处理分析流程, 而新的分析手段如人工智能的接入更将为我们带来深渊生命的高维特征。如何理解并利用这些高维认知, 需要原位调查及实验室实验联动, 计算预测与高通量实验室验证的交互是从相关性发现到因果关系建立的关键环节, 这必将催生技术方法的快速发展和迭代 (图 1)。

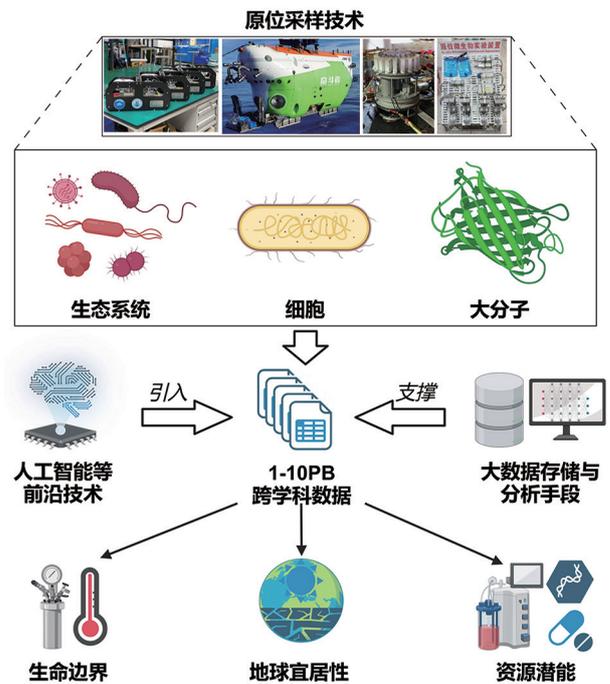


图1 大数据时代的深渊微生物研究展望 (使用BioRender.com制作)

表1 已发表深渊微生物组学研究与国际微生物大数据研究计划的规模对比

	数据类型	样本量	数据量
Jian等 ^[16]	宏基因组	19	1.9 TB
Zhou等 ^[19]	宏基因组	37	0.5 TB
	宏转录组	3	0.1 TB
Chen等 ^[44]	宏基因组	11	0.2 TB
	分离培养菌株	-	1 070株细菌, 19株真菌
HMP ^[61]	宏基因组(16S rRNA扩增子)	13 572	38 TB
	分离培养菌株	-	> 1 000 代表菌株
iHMP ^[62]	宏基因组(16S rRNA扩增子 + 全基因组)	18 024	10.5 TB
EMP ^[63-64]	宏基因组(16S rRNA扩增子)	27 751	0.4 TB
Tara Oceans ^[65]	宏基因组	370	10.2 TB
	宏转录组	187	4.8 TB

[参 考 文 献]

- [1] Du M, Peng X, Zhang H, et al. Geology, environment, and life in the deepest part of the world's oceans. *Innovation (Camb)*, 2021, 2: 100109
- [2] Stewart HA, Jamieson AJ. Habitat heterogeneity of hadal trenches: considerations and implications for future studies. *Prog Oceanogr*, 2018, 161: 47-65
- [3] Jamieson AJ, Fujii T, Mayor DJ, et al. Hadal trenches: the ecology of the deepest places on Earth. *Trends Ecol Evol*, 2010, 25: 190-7
- [4] Jamieson A. *The hadal zone: life in the deepest oceans*[M]. Cambridge: Cambridge University Press, 2015
- [5] Glud RN, Wenzhöfer F, Middelboe M, et al. High rates of microbial carbon turnover in sediments in the deepest oceanic trench on Earth. *Nat Geosci*, 2013, 6: 284-8
- [6] Glud RN, Berg P, Thamdrup B, et al. Hadal trenches are dynamic hotspots for early diagenesis in the deep sea. *Commun Earth Environ*, 2021, 2: 21
- [7] Wenzhöfer F, Oguri K, Middelboe M, et al. Benthic carbon mineralization in hadal trenches: assessment by *in situ* O₂ microprofile measurements. *Deep Sea Res 1 Oceanogr Res Pap*, 2016, 116: 276-86
- [8] Nunoura T, Nishizawa M, Kikuchi T, et al. Molecular biological and isotopic biogeochemical prognoses of the nitrification-driven dynamic microbial nitrogen cycle in hadopelagic sediments. *Environ Microbiol*, 2013, 15: 3087-107
- [9] Nunoura T, Takaki Y, Hirai M, et al. Hadal biosphere: insight into the microbial ecosystem in the deepest ocean on Earth. *Proc Natl Acad Sci U S A*, 2015, 112: E1230-6
- [10] Thamdrup B, Schauburger C, Larsen M, et al. Anammox bacteria drive fixed nitrogen loss in hadal trench sediments. *Proc Natl Acad Sci U S A*, 2021, 118: e2104529118
- [11] Liu S, Peng X. Organic matter diagenesis in hadal setting: insights from the pore-water geochemistry of the Mariana Trench sediments. *Deep Sea Res 1 Oceanogr Res Pap*, 2019, 147: 22-31
- [12] Cui G, Li J, Gao Z, et al. Spatial variations of microbial communities in abyssal and hadal sediments across the Challenger Deep. *PeerJ*, 2019, 7: e6961
- [13] Peoples LM, Grammatopoulou E, Pombrol M, et al. Microbial community diversity within sediments from two geographically separated hadal trenches. *Front Microbiol*, 2019, 10: 347
- [14] Li Y, Cao W, Wang Y, et al. Microbial diversity in the sediments of the southern Mariana Trench. *J Oceanol Limnol*, 2019, 37: 1024-9
- [15] Nunoura T, Nishizawa M, Hirai M, et al. Microbial diversity in sediments from the bottom of the Challenger Deep, the Mariana Trench. *Microbes Environ*, 2018, 33: 186-94
- [16] Jian H, Yi Y, Wang J, et al. Diversity and distribution of viruses inhabiting the deepest ocean on Earth. *ISME J*, 2021, 15: 3094-110
- [17] Hiraoka S, Hirai M, Matsui Y, et al. Microbial community and geochemical analyses of trans-trench sediments for understanding the roles of hadal environments. *ISME J*, 2020, 14: 740-56
- [18] Schauburger C, Glud RN, Hausmann B, et al. Microbial community structure in hadal sediments: high similarity along trench axes and strong changes along redox gradients. *ISME J*, 2021, 15: 3455-67
- [19] Zhou YL, Mara P, Cui GJ, et al. Microbiomes in the Challenger Deep slope and bottom-axis sediments. *Nat Commun*, 2022, 13: 1515
- [20] Liu R, Wang L, Wei Y, et al. The hadal biosphere: recent insights and new directions. *Deep Sea Res 2 Top Stud Oceanogr*, 2018, 155: 11-8
- [21] Xiao X, Zhang Y, Wang F. Hydrostatic pressure is the universal key driver of microbial evolution in the deep ocean and beyond. *Environ Microbiol Rep*, 2021, 13: 68-72
- [22] Belyaev G. *Deep-sea ocean trenches and their fauna* [Z]. Moscow: Nauka, 1989
- [23] Luo M, Glud RN, Pan B, et al. Benthic carbon mineralization in hadal trenches: insights from *in situ* determination of benthic oxygen consumption. *Geophys Res Lett*, 2018, 45: 2752-60
- [24] Xiao W, Xu Y, Haghpour N, et al. Efficient sequestration of terrigenous organic carbon in the New Britain Trench. *Chem Geol*, 2020, 533: 119446
- [25] Xu Y, Li X, Luo M, et al. Distribution, source, and burial of sedimentary organic carbon in Kermadec and Atacama Trenches. *J Geophys Res Biogeosci*, 2021, 126: e2020JG006189
- [26] Wang N, Shen C, Sun W, et al. Penetration of bomb ¹⁴C into the deepest ocean trench. *Geophys Res Lett*, 2019, 46: 5413-9
- [27] Hu T, Luo M, Wunsch UJ, et al. Probing sedimentary DOM in the deepest sector of Earth's surface. *Mar Chem*, 2021, 237: 104033
- [28] Guan H, Chen L, Luo M, et al. Composition and origin of lipid biomarkers in the surface sediments from the southern Challenger Deep, Mariana Trench. *Geosci Front*, 2019, 10: 351-60
- [29] Xu Y, Jia Z, Xiao W, et al. Glycerol dialkyl glycerol tetraethers in surface sediments from three Pacific trenches: distribution, source and environmental implications. *Org Geochem*, 2020, 147: 104079
- [30] Xiao W, Wang Y, Liu Y, et al. Predominance of hexamethylated 6-methyl branched glycerol dialkyl glycerol tetraethers in the Mariana Trench: source and environmental implication. *Biogeosciences*, 2020, 17: 2135-48
- [31] Ta K, Peng X, Xu H, et al. Distributions and sources of glycerol dialkyl glycerol tetraethers in sediment cores from the Mariana subduction zone. *J Geophys Res Biogeosci*, 2019, 124: 857-69
- [32] Luo M, Gieskes J, Chen L, et al. Provenances, distribution, and accumulation of organic matter in the southern Mariana Trench rim and slope: implication for carbon cycle and burial in hadal trenches. *Mar Geol*, 2017, 386: 98-106
- [33] Luo M, Algeo TJ, Tong H, et al. More reducing bottom-

- water redox conditions during the Last Glacial Maximum in the southern Challenger Deep (Mariana Trench, western Pacific) driven by enhanced productivity. *Deep Sea Res 2 Top Stud Oceanogr*, 2018, 155: 70-82
- [34] Jiang Z, Sun Z, Liu Z, et al. Rare-earth element geochemistry reveals the provenance of sediments on the southwestern margin of the Challenger Deep. *J Oceanol Limnol*, 2019, 37: 998-1009
- [35] Oguri K, Masqué P, Zabel M, et al. Sediment accumulation and carbon burial in four hadal trench systems. *J Geophys Res Biogeosci*, 2022, 127: e2022JG006814
- [36] Xu Y, Ge H, Fang J. Biogeochemistry of hadal trenches: recent developments and future perspectives. *Deep Sea Res 2 Top Stud Oceanogr*, 2018, 155: 19-26
- [37] Ichino MC, Clark MR, Drazen JC, et al. The distribution of benthic biomass in hadal trenches: a modelling approach to investigate the effect of vertical and lateral organic matter transport to the seafloor. *Deep Sea Res 1 Oceanogr Res Pap*, 2015, 100: 21-33
- [38] Liu R, Wang Z, Wang L, et al. Bulk and active sediment prokaryotic communities in the Mariana and Mussau Trenches. *Front Microbiol*, 2020, 11: 1521
- [39] Tarn J, Peoples LM, Hardy K, et al. Identification of free-living and particle-associated microbial communities present in hadal regions of the Mariana Trench. *Front Microbiol*, 2016, 7: 665
- [40] Tian J, Fan L, Liu H, et al. A nearly uniform distributional pattern of heterotrophic bacteria in the Mariana Trench interior. *Deep Sea Res 1 Oceanogr Res Pap*, 2018, 142: 116-26
- [41] Liu J, Zheng Y, Lin H, et al. Proliferation of hydrocarbon-degrading microbes at the bottom of the Mariana Trench. *Microbiome*, 2019, 7: 47
- [42] Wang X, Yang Y, Lv Y, et al. The capability of utilizing abiotic enantiomers of amino acids by *Halomonas* sp. LMO_D1 derived from the Mariana Trench. *Front Astron Space Sci*, 2021, 8: 741053
- [43] Zhao X, Liu J, Zhou S, et al. Diversity of culturable heterotrophic bacteria from the Mariana Trench and their ability to degrade macromolecules. *Mar Life Sci Technol*, 2020, 2: 181-93
- [44] Chen P, Zhou H, Huang Y, et al. Revealing the full biosphere structure and versatile metabolic functions in the deepest ocean sediment of the Challenger Deep. *Genome Biol*, 2021, 22: 207
- [45] Eloë EA, Fadrosch DW, Novotny M, et al. Going deeper: metagenome of a hadopelagic microbial community. *PLoS One*, 2011, 6: e20388
- [46] Nunoura T, Hirai M, Yoshida-Takashima Y, et al. Distribution and niche separation of planktonic microbial communities in the water columns from the surface to the hadal waters of the Japan Trench under the eutrophic ocean. *Front Microbiol*, 2016, 7: 1261
- [47] Peoples LM, Kyaw TS, Ugalde JA, et al. Distinctive gene and protein characteristics of extremely piezophilic *Colwellia*. *BMC Genomics*, 2020, 21: 692
- [48] Lauro FM, Chastain RA, Blankenship LE, et al. The unique 16S rRNA genes of piezophiles reflect both phylogeny and adaptation. *Appl Environ Microbiol*, 2007, 73: 838-45
- [49] Yayanos AA, Dietz AS, van Boxtel R. Obligately barophilic bacterium from the Mariana trench. *Proc Natl Acad Sci U S A*, 1981, 78: 5212-5
- [50] Yayanos AA. Microbiology to 10,500 meters in the deep sea. *Annu Rev Microbiol*, 1995, 49: 777-805
- [51] Feng JC, Liang J, Cai Y, et al. Deep-sea organisms research oriented by deep-sea technologies development. *Sci Bull (Beijing)*, 2022, 67: 1802-16
- [52] Gao ZM, Huang JM, Cui GJ, et al. *In situ* meta-omic insights into the community compositions and ecological roles of hadal microbes in the Mariana Trench. *Environ Microbiol*, 2019, 21: 4092-108
- [53] Zhang C, Liu Q, Li X, et al. Spatial patterns and co-occurrence networks of microbial communities related to environmental heterogeneity in deep-sea surface sediments around Yap Trench, Western Pacific Ocean. *Sci Total Environ*, 2021, 759: 143799
- [54] Liu R, Wei X, Song W, et al. Novel *Chloroflexi* genomes from the deepest ocean reveal metabolic strategies for the adaptation to deep-sea habitats. *Microbiome*, 2022, 10: 75
- [55] Zheng PF, Wei Z, Zhou Y, et al. Genomic evidence for the recycling of complex organic carbon by novel *Thermoplasmata* clades in deep-sea sediments. *mSystems*, 2022, 7: e00077-22
- [56] Cui G, Zhou Y, Li W, et al. A novel bacterial phylum that participates in carbon and osmolyte cycling in the Challenger Deep sediments. *Environ Microbiol*, 2021, 23: 3758-72
- [57] Glud RN. Oxygen dynamics of marine sediments. *Mar Biol Res*, 2008, 4: 243-89
- [58] Jørgensen BB, Wenzhöfer F, Egger M, et al. Sediment oxygen consumption: role in the global marine carbon cycle. *Earth Sci Rev*, 2022, 228: 103987
- [59] Tamburini C, Boutrif M, Garel M, et al. Prokaryotic responses to hydrostatic pressure in the ocean -- a review. *Environ Microbiol*, 2013, 15: 1262-74
- [60] Amano C, Zhao Z, Sintès E, et al. Limited carbon cycling due to high-pressure effects on the deep-sea microbiome. *Nat Geosci*, 2022, 15: 1041-7
- [61] Human Microbiome Project Consortium. A framework for human microbiome research. *Nature*, 2012, 486: 215-21
- [62] Integrative HMP (iHMP) Research Network Consortium. The Integrative Human Microbiome Project: dynamic analysis of microbiome-host omics profiles during periods of human health and disease. *Cell Host Microbe*, 2014, 16: 276-89
- [63] Gilbert JA, Jansson JK, Knight R. Earth microbiome project and global systems biology. *mSystems*, 2018, 3: e00217-17
- [64] Gilbert JA, Jansson JK, Knight R. The Earth Microbiome Project: successes and aspirations. *BMC Biol*, 2014, 12: 69
- [65] Tara Oceans Coordinators, Sunagawa S, Acinas SG, et al. Tara Oceans: towards global ocean ecosystems biology.

- Nat Rev Microbiol, 2020, 18: 428-45
- [66] Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature*, 2012, 486: 207-14
- [67] Gibbons SM, Caporaso JG, Pirrung M, et al. Evidence for a persistent microbial seed bank throughout the global ocean. *Proc Natl Acad Sci U S A*, 2013, 110: 4651-5
- [68] Lennon JT, Jones SE. Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nat Rev Microbiol*, 2011, 9: 119-30
- [69] Proctor LM, Creasy HH, Fettweis JM, et al. The Integrative Human Microbiome Project. *Nature*, 2019, 569: 641-8
- [70] Lloyd-Price J, Mahurkar A, Rahnavard G, et al. Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature*, 2017, 550: 61-6
- [71] Bucci V, Tzen B, Li N, et al. MDSINE: Microbial Dynamical Systems Inference Engine for microbiome time-series analyses. *Genome Biol*, 2016, 17: 121
- [72] Angulo MT, Moog CH, Liu YY. A theoretical framework for controlling complex microbial communities. *Nat Commun*, 2019, 10: 1045
- [73] Bolyen E, Rideout JR, Dillon MR, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol*, 2019, 37: 852-7
- [74] Poussin C, Sierro N, Boué S, et al. Interrogating the microbiome: experimental and computational considerations in support of study reproducibility. *Drug Discov Today*, 2018, 23: 1644-57
- [75] Liu YX, Qin Y, Chen T, et al. A practical guide to amplicon and metagenomic analysis of microbiome data. *Protein Cell*, 2021, 12: 315-30
- [76] Gonzalez A, Navas-Molina JA, Kosciolk T, et al. Qiita: rapid, web-enabled microbiome meta-analysis. *Nat Methods*, 2018, 15: 796-8