



## 核糖体 RNA 测序技术在海洋微生物多样性研究中的应用

刘宇航<sup>1,2</sup>, 陈松泽<sup>3</sup>, 张传伦<sup>1</sup>, 范陆<sup>1,4\*</sup>

<sup>1</sup>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

<sup>2</sup>哈尔滨工业大学环境学院, 黑龙江 哈尔滨 150001

<sup>3</sup>同济大学海洋地质国家重点实验室, 上海 200092

<sup>4</sup>南方科技大学前沿与交叉科学研究院, 广东 深圳 518055

**摘要:** 海洋微生物总数高达  $10^{30}$  个, 在海洋生态系统中发挥着驱动能量物质循环和维持生态平衡的重要作用。深入研究海洋微生物多样性有助于理解海洋生态系统运行机理、应对全球海洋生态危机, 以及开发海洋微生物资源。由于目前可培养分离的海洋微生物种类极少, 极大限制了对海洋微生物群落、丰度和生理生态特征的研究。核糖体 RNA 测序技术以较低的成本, 根据遗传信息差异对微生物快速准确地进行分类鉴定, 揭示群落结构、评估进化和生态学关系。近年来基于核糖体 RNA 序列测序分析的海洋微生物学研究随着测序技术本身的快速优化, 在发现海洋微生物新类群, 揭示海洋微生物生态规律, 分析海洋微生物进化关系, 以及与海洋微生物相关的代谢产物开发和海洋生态治理等研究中取得显著进展。本综述详细介绍了核糖体 RNA 序列测序技术的原理, 各代测序技术在海洋微生物多样性研究中的应用, 以及多种建库和测序技术的有机结合。最后对研究不同海洋微生物多样性问题提供了测序方案的建议和展望, 以期核糖体 RNA 测序技术在海洋微生物多样性研究中的应用提供参考。

**关键词:** 海洋微生物多样性, 核糖体 RNA, 测序技术

海洋微生物(包括细菌、古菌、原生动物、真菌、单细胞藻类和病毒等)是海洋生态系统中的主要生产者以及分解者。海洋中约含  $3.6 \times 10^{29}$  个原核生物细胞<sup>[1]</sup>。它们以不同功能的群落形式促进海

洋生态系统结构和功能的多样性与稳定性<sup>[2]</sup>。海洋微生物密切参与光合作用、养分吸收和有机质转移等海洋生态循环, 贡献地球上的一半的初级生产力<sup>[3]</sup>。在碳、氮、磷、硫等元素的代谢利用以及营

基金项目: 国家自然科学基金(91951120, 91851210); 广东省自然科学基金(2018B030311016); 深圳市科技创新委员会(JCYJ20180305123458107, ZDSYS201802081843490)

\*通信作者。Tel/Fax: +86-755-88018620; E-mail: fanl@sustech.edu.cn

收稿日期: 2020-03-28; 修回日期: 2020-05-10; 网络出版日期: 2020-07-11

养循环中发挥着不可替代的作用<sup>[4-5]</sup>,并影响着其他海洋大型生物的生存代谢<sup>[6]</sup>。部分海洋微生物对海洋环境具有潜在的保护作用<sup>[7]</sup>,能够高效降解石油等海洋有机污染物<sup>[8]</sup>,调控赤潮等海洋生物灾害<sup>[9-10]</sup>,也是环境监测的重要工具<sup>[11-13]</sup>。此外,海洋微生物是海洋资源开发与利用的关键,是工业和医疗所需的药物和酶开发的宝库<sup>[6]</sup>。

海洋微生物的多样性是海洋微生物生态研究的前提和重点<sup>[14]</sup>。微生物多样性的含义非常丰富,通常包括了基于研究尺度的 $\alpha$ 多样性(样本内多样性)、 $\beta$ 多样性(样本间多样性)和 $\gamma$ 多样性(地理分布多样性),以及基于研究单位类型的物种多样性、系统发生学多样性和功能多样性等。而 $\alpha$ 多样性中则又包括了丰富度(Richness)和均匀度(Evenness)等指标。本文着重讨论环境中微生物物种丰富度(Species richness),也就是特定环境中物种的数量。

微生物物种定义标准存在诸多争议<sup>[15-16]</sup>。传统物种鉴定及分类的方法是选择性培养微生物,并根据形态、生理和生化特征、营养代谢条件、系统发育和基因型等遗传特征进行区分<sup>[17-19]</sup>。虽然培养技术不断进步<sup>[20]</sup>,但绝大多数海洋微生物的培养仍面临挑战<sup>[21]</sup>。因此,仅依靠传统的培养分离技术鉴定物种,会极大地限制对环境微生物多样性的认知<sup>[15]</sup>。

现代分子生物学技术主要通过标记基因或全基因组测序研究海洋微生物多样性。目前较为广泛接受的微生物物种多样性分析是基于核糖体核糖核酸(核糖体 RNA,即 Ribosomal RNA, rRNA)序列分析<sup>[22]</sup>,根据序列一致性聚类形成“可操作分类单元”(Operational taxonomic unit, OTU),将 16S rRNA 基因序列一致性大于 97%的微生物粗略估

计为同一物种<sup>[23]</sup>。rRNA 序列分析技术可检测出许多尚未培养的微生物<sup>[24]</sup>,已广泛应用于海洋微生物系统多样性研究<sup>[25-26]</sup>。

当前基于 rRNA 分析技术对海洋微生物多样性的研究结果存在诸多争议。据预测,每毫升海水中有数百至数十万种不同的微生物物种,而海洋微生物物种多样性范围在十万至百万<sup>[27]</sup>。最近几年海洋微生物研究仍然不断发现大量新类群,总类群数应已远超这些先前的估计。虽然有科学家认为之前微生物多样性被远远高估,全球范围内仅存在 80-160 万个原核 OTU<sup>[28]</sup>,但依旧存在大量不可培养的微生物类群。因此,要系统准确地推测海洋微生物整体多样性仍然是一个非常困难的问题<sup>[29]</sup>。

本综述着重介绍了近年来测序平台技术优化对 rRNA 序列测序的变革,及其在海洋微生物多样性研究中的应用,特别是海洋微生物新类群的发现。同时讨论了不同测序技术存在的问题和未来可能的解决方案,以期开展海洋微生物多样性研究提供方法选择的参考。

## 1 rRNA 序列在微生物多样性研究中的重要意义

rRNA 是一种非编码 RNA,是组成核糖体的结构骨架分子,在蛋白质合成中发挥关键作用,为所有细胞生物所共有<sup>[30]</sup>。rRNA 基因在物种进化中具有高度的保守性,且几乎不发生水平基因转移<sup>[31]</sup>,是推断原核生物和真核生物的系统发育关系的重要分子标记<sup>[32]</sup>。rRNA 由大核糖体亚基(Large subunit, LSU)和小核糖体亚基(Small subunit, SSU)组成,包含大量遗传信息<sup>[33]</sup>。它们

在基因组中常常成串分布(图 1)。原核生物(细菌和古菌)的 SSU 为 16S rRNA (大肠杆菌中长度 1542 nt), LSU 为 5S rRNA (大肠杆菌中长度 120 nt) 和 23S rRNA (大肠杆菌中长度 2906 nt)。在这两者的基因之间通常存在一个内部转录间隔区 (Internal transcribed spacer, ITS)。真核生物的 SSU 为 18S rRNA (酵母中长度 1869 nt), LSU 为 5S rRNA (酵母中长度 121 nt)、5.8S rRNA (酵母中长度 156 nt) 和 28S rRNA (酵母中长度 5070 nt)。在 18S rRNA 基因和 5.8S rRNA 基因以及 5.8S rRNA 基因和 28S rRNA 基因之间各有一个 ITS<sup>[34]</sup>。此外, 植物和真核藻类中的叶绿体基因组和线粒体基因组通常含有 16S rRNA 基因和 23S rRNA 基因。在上述 rRNA 分子中, 16S rRNA 基因序列中存在 9 个可变区和 10 个高度保守区域, 保守区与可变区以不同的速率进化(图 1)。缓慢进化的保守区可用于设计通用引物来扩增跨不同分类单元的微生物的 rRNA 基因<sup>[35]</sup>, 而快速进化的可变区则因反映物种之间的差异而用于分类学鉴定和多样性分析<sup>[36]</sup>。当前, rRNA 基因测序已经成为实验室进行微生物鉴定和多样性分析最常用方法<sup>[37]</sup>。

其中, 16S rRNA 基因序列被广泛用于研究海洋原核微生物的类群鉴定<sup>[38]</sup>。

## 2 基于 rRNA 的海洋微生物群落多样性主要检测分析技术

### 2.1 指纹图谱技术

浓度梯度电泳 (Denaturing gradient gel electrophoresis, DGGE) 和温度梯度电泳 (Temperature gradient gel electrophoresis, TGGE) 是分子指纹图谱技术的代表<sup>[39]</sup>。通过聚合酶链式反应 (Polymerase chain reaction, PCR) 扩增复杂环境中的 rRNA 基因, 基于这些 DNA 分子不同的 GC 含量和二级结构, 使用含梯度变性剂或温度梯度电泳环境对 rRNA 基因片段进行分级分离<sup>[40]</sup>。这种梯度电泳技术和 16S rRNA 基因克隆文库测序<sup>[41]</sup> 联用, 并促进了海洋微生物多样性的研究<sup>[42]</sup>; 而对海洋沉积物古菌群落的研究发现群落多样性随深度而升高<sup>[43]</sup>。因为梯度电泳存在操作繁琐、分析通量较低、对低丰度微生物的分辨率有限、不能准确分类复杂环境的物种的缺点, 现已被逐渐淘汰。

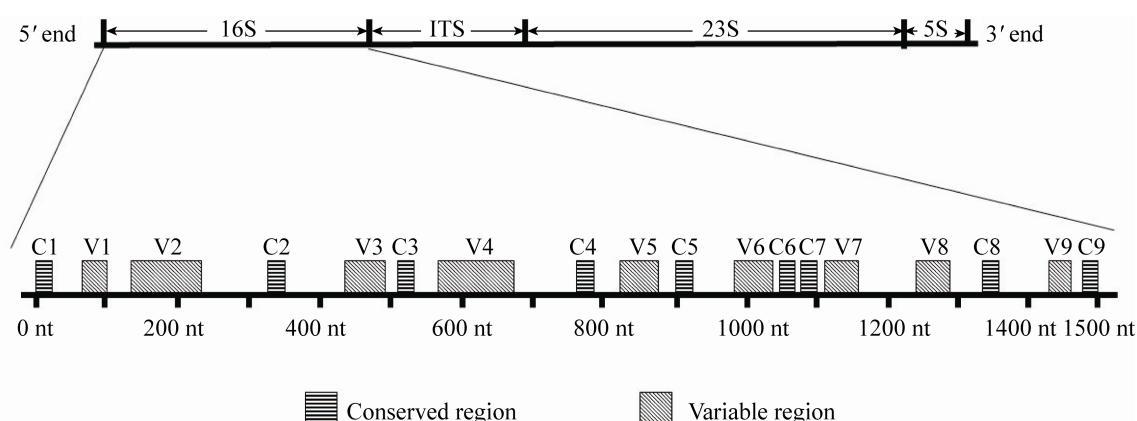


图 1. rRNA 操纵子及 16S 基因保守区和可变区示意图

Figure 1. Schematic diagram of the rRNA gene operon and conserved and variable regions of the 16S rRNA gene.

## 2.2 分子标记技术

分子标记技术是通过标记核苷酸序列反映特异性 DNA 片段的方法。限制性片段长度多态性分析 (Restriction fragment length polymorphism, RFLP) 是分子标记方法的代表性技术。该技术通过标记 rRNA 基因限制性片段实现快速精准识别微生物序列间的差异<sup>[44]</sup>。但该技术操作要求高, 获得信息量有限。末端限制性片段长度多态性分析 (Terminal restriction fragment length polymorphism, TRFLP) 与 PCR 联用, 比 RFLP 更为快捷灵敏<sup>[45]</sup>。但该技术无法区分共享末端限制位点的不同序列。代表性研究包括借助 TRFLP 技术筛选新泽西州沿海的海洋微生物的 16S rRNA 基因<sup>[46]</sup>和通过 TRFLP 对海底真菌种群与环境条件进行了相关性研究<sup>[47]</sup>。但 RFLP 与 TRFLP 均受限于复杂微生物群落分析, 现已逐渐被淘汰。

## 2.3 荧光原位杂交技术(FISH)

荧光原位杂交技术 (Fluorescent *in situ* hybridization, FISH) 通过荧光标记与目标序列杂交互补的探针序列(16–24 bp), 实现复杂微生物群落的物种鉴定、物种丰度及功能基因检测<sup>[48–49]</sup>。该技术可实现定量分析海洋微生物的优势群落、分布规律及生态功能<sup>[26]</sup>。例如, 通过借助 FISH 对 Sargasso Sea 海域的细胞丰度定量测量, 其中一株命名为 SAR11 的类群被认定是全球海洋中最丰富的细菌<sup>[50]</sup>。Karner 等通过 FISH 技术评估了海水中古菌的细胞数约为  $1.3 \times 10^{28}$ , 总量和细菌在一个数量级<sup>[51]</sup>。但 FISH 技术可以同时检测到的微生物种类数量有限, 探针制备成本较高, 杂交过程耗时较长<sup>[52]</sup>, 限制了该技术的普及及应用。

## 2.4 rRNA 序列测序和分析技术

rRNA 序列测序技术是采用测序仪, 对 rRNA 或 rRNA 基因的序列进行读取, 通过序列一致性比对、相似性聚类、物种注释、丰度计算、进化树构建、多样性统计和可视化等生物信息学方法, 对环境微生物多样性进行综合分析<sup>[53]</sup>。rRNA 核酸序列(包括 rRNA 和 rRNA 基因)测序技术可以鉴定和分类更广泛的微生物类群, 解析度可以达到单核苷酸解析度, 从而整体上更为深入详细地揭示环境微生物的区系组成<sup>[54]</sup>, 已成为研究海洋微生物生态学的一种标准方法<sup>[55]</sup>。

其中, 16S rRNA 基因扩增子测序技术(通过 PCR 引物扩增 16S rRNA 基因后再测序)在环境微生物生态应用中最为广泛。但该技术对微生物多样性分析可靠性会受到 PCR 扩增效率的影响。包括: (1) 简并引物对不同微生物的扩增效率存在偏好性, 导致部分不与引物互补的 DNA 模板将不会被扩增, 其相应的微生物在群落多样性分析中将成为假阴性。特别是环境中微生物组成未知, 在引物设计时很难考虑到所有未知物种, 因此往往在软件预测中表现良好的引物在实验室条件下不一定有较好表现<sup>[56]</sup>。(2) 尽管引物序列中的错配是扩增偏向的主要原因, 但在简并位置含有 C/G 而非 A/T 的序列的优先结合也是一个重要因素<sup>[57]</sup>。(3) 部分扩增子在 PCR 中可能会断裂, 并与其他序列相似谱系的模板发生非特异性退火。这一过程可形成包含来自两个或多个微生物序列的 DNA 分子, 即嵌合体。这一现象与 PCR 循环数<sup>[58]</sup>和样品多样性密切相关<sup>[59]</sup>。嵌合体不来自于自然界存在的微生物, 从而导致最终对微生物多样性的高估。(4) PCR 扩增过程本身会引入突变, 因此需要使用高保真酶以及对测序结果进行降噪处理。

与另一种目前广泛采用的微生物多样性研究手段——宏基因组测序相比, rRNA 测序技术的优势在于能在较少成本下获取更多更详细的微生物物种多样性信息, 分析过程相对简单<sup>[60]</sup>。特别是在对生物量较少的样本以及与真核宿主共生的微生物样本中有突出优势。其缺点在于受引物偏好性影响, 难以进行微生物功能多样性分析<sup>[61]</sup>。因两者对比已有大量前期文献进行了讨论, 在此不再赘述。

### 3 测序技术的进步给 rRNA 序列分析带来的变革

DNA 测序技术起源于 1977 年。第一个被测定的基因组序列是噬菌体 X174<sup>[62]</sup>。近年来随着技

术进步, 测序平台在通量、操作过程、成本、精确度、读长、测序周期和测序仪器尺寸等方面不断优化, 主要经历了三代变革(表 1)。而测序技术的进步也同时极大促进海洋微生物多样性研究。

#### 3.1 第一代测序技术及其应用

1977 年, Sanger 发明了双脱氧核苷酸链终止法(Dideoxy chain termination method, Sanger 测序法)并开发了第一代测序技术<sup>[62]</sup>。而后该技术快速发展, 将基因测序带入高度自动化时代<sup>[63]</sup>。基于克隆的 Sanger 测序可提供准确的近全长 16S rRNA 序列(图 2-①), 是获得准确序列信息的有效方法<sup>[64]</sup>。借助该技术对 Sargasso Sea 中微生物的 16S rRNA 基因序列进行扩增测序, 首次发现了原本认为仅栖息在极端环境中的古菌, 在海洋中也

表 1. 代表性测序平台的比较

Table 1. Comparison of representative DNA sequencing platforms

Sequencing technology	Years to market	Features	Limitation
Sanger (1 <sup>st</sup> generation)	1977	1. Long reads (1000 nt) 2. High accuracy, reading repeated regions and homopolymers well	1. Low throughput (6 Mb/d) 2. High cost for sample preparation
Roche/454 (2 <sup>nd</sup> generation)	2005	1. Short reads (~500 nt) 2. High throughput (750 Mb/d) 3. Low cost	1. Error-prone in reading single-base continuous regions 2. Higher error rate than Sanger sequencing
Illumina (2 <sup>nd</sup> generation)	2006	1. Short reads (~500 nt) 2. High throughput (5000 Mb/d) 3. Low cost	1. GC biased 2. Higher error rate than Sanger sequencing
PacBio SMRT (3 <sup>rd</sup> generation)	2009	1. Long reads (1–10 kb) 2. No need for PCR amplification, single molecule direct sequencing	1. Lower throughput than the 2 <sup>nd</sup> generation technologies (10 M/run) 2. Higher cost than 2 <sup>nd</sup> gen 3. High original error rate and error-prone in reading homopolymers but can be corrected by CCS
Oxford Nanopore (3 <sup>rd</sup> generation)	2010	1. Long reads 2. No need for PCR amplification, single molecule direct sequencing 3. Direct sequencing of RNA molecules 4. Easy operation and portable equipment	1. Low throughput (1 M/run) 2. High error rate but in continuous optimization

广泛存在<sup>[65-66]</sup>。在随后的环球大洋采样项目 (Global ocean sampling expedition) 中, 测序发现 85% 以上的基因是前所未有的<sup>[67]</sup>; 并发现一些物种和基因变异有较为明确的空间分布<sup>[68]</sup>。虽然一代测序能得到较长序列, 但研究环境微生物多样性需要通过大肠杆菌构建克隆文库, 存在一定的克隆偏差<sup>[69]</sup>。且由于无法对混合 DNA 进行并行测序, 不适合用于大规模的基因测序研究。第一代测序技术在通量、成本、测序速度等方面的局限性, 无法满足对各类环境微生物的基因序列的大量测序需求, 在海洋微生物多样性研究中, 已经被第二代测序技术所取代。

### 3.2 第二代测序技术及其应用

第二代测序技术(又称高通量测序技术)的发展, 解除了 Sanger 测序一个反应只能测定一条序列的限制, 成本也出现了断崖式下降。实现了准确、快速和低成本地研究各类微生物群落(图 2-②)。第二代测序技术的主要代表公司有

454/Roche<sup>[70]</sup>、Solexa/Illumina<sup>[71]</sup>、BGISeq/华大基因(基于 Complete Genomics 公司技术改造)<sup>[72]</sup>、ion Torrent/Life<sup>[73]</sup>和 SOLiD/ABI<sup>[12]</sup>等。

2005 年, Roche 公司基于 454 焦磷酸测序原理推出的 Genome Sequencer FLX 测序平台是首个商业化的第二代测序技术平台。该技术兼具并行通量较高且读长相对较长(500 nt)的优势, 较为适合对 16S rRNA 序列进行测序<sup>[74]</sup>。例如, Klier 等对波罗的海低盐-中盐过渡带沉积物的底栖细菌群落 16S rRNA 基因序列进行了测序。首次揭示了波罗的海盐度梯度下沉积物中细菌的多样性模式和组成。研究表明, 基于其广泛的盐度耐受性, 几种典型的海洋和寡盐沉积细菌能够在中等盐度环境中分布<sup>[75]</sup>。该测序方法也推动了深海<sup>[76]</sup>和热液喷口<sup>[77-78]</sup>等海洋环境附近的微生物多样性研究。454 焦磷酸测序的缺点在于通量与成本比仍然较低, 对连续单核苷酸重复的读取错误率较高, 因此在随后的竞争中逐步被 Illumina 公司的测序仪所取代<sup>[79]</sup>。

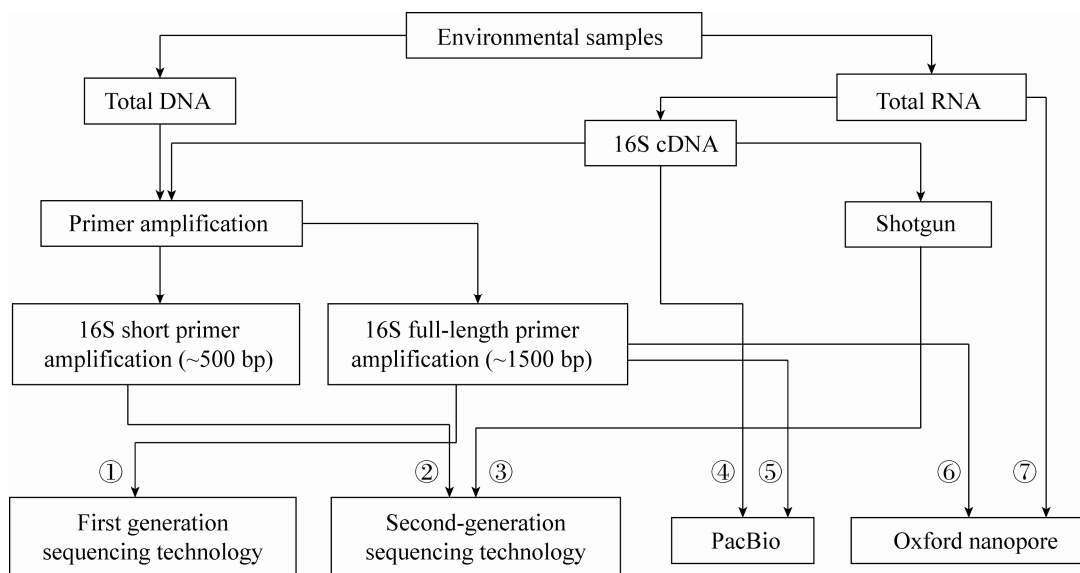


图 2. 16S rRNA 序列分析策略

Figure 2. Technical strategies in analyzing 16S rRNA sequences.

2006年, Illumina 公司开始陆续提供 HiSeq、MiSeq 等多个高通量测序系统, 相较于 454 测序平台, 其成本更低、通量更高<sup>[80]</sup>。通过该技术可研究环境中的极低丰度物种<sup>[81]</sup>, 从而引起了微生物研究方法的变革, 并长期以来主导着 rRNA 扩增子测序研究。例如, Tara Oceans 项目<sup>[82]</sup>使用 Illumina 测序技术对全球海洋浮游微生物样本进行测序, 分析了其中细菌、古菌和真核微生物的 16S rRNA 基因多样性, 研究其相对丰度、进化关系及对环境和生物地球化学循环的影响<sup>[83]</sup>。

第二代测序技术已被普遍用于探索不同生境中微生物的物种组成和多样性。目前, 数据分析流程也已标准化(如 Mothur、Qiime2、USEARCH 等软件<sup>[84-85]</sup>)。但二代测序技术依旧存在读长较短(约 500 bp)的缺点。再者, 其测序数据准确率显著低于第一代测序, 一定程度上限制了微生物多样性分析的精确性<sup>[86]</sup>(详细讨论见下文)。

为了克服读长短的缺点, 部分公司和实验室开发了基于二代测序的长读长分析技术。主要思路是将全长 rRNA 扩增子随机打断后进行短读长测序, 而后通过混合从头组装或者标签指导下单个序列组装的方式来获得完整的全长 rRNA 序列。而后者又被称为合成长读长(Synthetic long reads)方法。该方法的关键是在制备短读文库之前, 将样本中的每个模板核苷酸序列分子都标记一个由 10-20 个随机碱基组成的独特的分子标识符(Unique molecular identifiers, UMIs)<sup>[87]</sup>, 使得所有目标片段在整个加工和测序过程中都包含 UMI 标签。在测序完成后, 即可根据读长中的 UMI 序列, 分析其在原始模板分子中的排序, 从而组装生成共识序列<sup>[88]</sup>。该技术的主要不足在于, 受长 DNA 片段的生成效率所限, UMI 合成片段长度上限约

为 2000 bp (但也有报道可以合成达到 11.6 kb 长度的序列<sup>[89]</sup>), 无法对 LSU rRNA 序列进行测序分析。此外, 该技术操作繁琐, 限制了其更广泛应用。同时由于合成长读长依赖于短读长的从头组装, 也使得这种方法依然无法解析大于短读长的内部序列重复。

### 3.3 第三代测序技术及其应用

Pacific Biosciences (PacBio)公司开发的单分子实时测序技术(Single molecule and real-time, SMRT)和 Oxford Nanopore (ONT)公司开发的纳米孔测序技术<sup>[90-91]</sup>是第三代测序技术的代表。第三代测序技术的最大技术特点是实现了单分子高通量测序, 并显著提高了测序读长(>1500 bp)。

PacBio 通过检测 DNA 分子上每个核苷酸与纳米井(Nanowell)底部聚合酶结合时发出的荧光实现序列读取。为了克服其原始测序精度较低的缺点, 采用了环形共识序列(Circular consensus sequence, CCS)策略对同一分子循环测序和多重比对。PacBio SMRT 测序技术已经被应用于分析 16S rRNA 基因(图 2-⑤)<sup>[92]</sup>和 23S rRNA 基因<sup>[93]</sup>。该技术可提高海洋微生物种水平的注释及物种丰度预测的准确性<sup>[94]</sup>。比如, Pootakham 等利用 PacBio CCS 技术的高通量、长读长优势, 对印度西太平洋地区主要造礁珊瑚 *Porites lutea* 的相关细菌群落多样性及组成进行分析<sup>[95]</sup>, 首次展示了 PacBio 技术进行海洋环境样品全长 16S rRNA 基因测序分析的可行性和优越性。

当前 PacBio 测序平台用于 rRNA 基因测序仍有改进的空间。首先, 虽然 CCS 策略能将碱基平均错误率降低到 0.2%<sup>[96]</sup>, 但仍高于 Illumina MiSeq (0.1%)<sup>[97]</sup>平台。其次, 尽管 PacBio 平台捕获全长 16S rRNA 序列比 Sanger 测序更具成本效益, 但目

前依然高于二代测序<sup>[98]</sup>。此外, 尽管存在一些生物信息学工具和软件包支持处理长读长数据<sup>[99-100]</sup>, 专门用于分析 PacBio 测序所得的 CCS 序列的一体化软件还有待开发。

ONT 推出的便携式测序设备 MinION<sup>TM</sup> 通过流动室(Flowcell)的马达蛋白控制核酸分子通过纳米孔, 进而检测纳米孔中电流的改变来实现序列读取。该仪器具有体积小、成本低和读长长(1-10 kb)的特点<sup>[101]</sup>, 为 16S rRNA 基因测序提供了一种新的选择<sup>[102]</sup>。更提供了 DNA 或 RNA 分子的实时和现场分析的可能性, 在临床应用和野外环境监测中具有巨大应用前景<sup>[103]</sup>。如 Curren 等使用 ONT MinION<sup>TM</sup> 测序平台对东南亚热带海洋生态系统中蓝细菌群落的 16S rRNA 全长基因扩增子进行测序<sup>[104]</sup>, 发现沿海地区蓝藻等物种丰富度的增加与富营养化和中等盐度的海水有关, 螺旋藻 *Spirulina* 等特定蓝细菌物种的存在可作为水质参数的指标。这是首次将该技术用于对浮游植物样品进行物种分辨率的研究。

ONT 平台最初受到较高的原始误差率(13%)的限制<sup>[105]</sup>。直到 R9.0 版本的出现将错误率降低到 7.5%, 才使其成为与其他测序平台的有力竞争者<sup>[106]</sup>。目前, 该平台的精确度仍在开发提高过程中<sup>[107]</sup>。例如, 使用可变驱动电压代替恒定驱动电压通过纳米孔, 可有效提高单次通过读取精度, 进而提高纳米孔测序的最终准确性<sup>[108]</sup>。最新推出的 R10.3 版本, 在结合 UMI 技术的基础上, 已经能将扩增子序列单核苷酸测序精度达到 99.995% (nanoporetech.com)。从而完全能满足 rRNA 序列高精度测序的要求。Shinichi 直接通过 PCR 扩增细胞悬液中近全长 16S rRNA 基因, 并建立了 ONT 测序的生物信息学流水线。这一方法加快了样品

制备和微生物鉴定的过程, 具有较高的准确性和便捷性<sup>[109]</sup>。但细胞壁裂解效率差异会显著影响扩增效率和测序结果。

## 4 rRNA 测序分析的不同策略

### 4.1 16S rRNA 基因部分长度测序和全长测序技术比较

受二代测序读长限制, 对 16S rRNA 基因的不同区域进行测序直接影响着多样性分析和进化分析结果<sup>[110-111]</sup>。不同高变区的微生物物种分辨率差异明显<sup>[112]</sup>, 且对不同类群有偏好性<sup>[113]</sup>。比如在不同序列一致性阈值下, 每个子区域产生的 OTU 的相对数量也不一致, 因此不同子区域的测序结果不能直接比较<sup>[114]</sup>。研究表明 16S rRNA 基因单个可变区的分类分辨率较低<sup>[115]</sup>, 而跨越越多可变区域的扩增子序列分类精度越高<sup>[116]</sup>。目前最佳可变区域组合的标准尚有争议<sup>[117-118]</sup>。

总的来说, 所测基因区域越长, 获得的分类和遗传信息量也更大, 可以以更高分辨率鉴定物种和构建进化树<sup>[92,119]</sup>。相对于短序列分析, 全长 16S rRNA 基因序列得到的 OTU 数量显著增加, 所构建提供的系统发育树分辨率更高<sup>[120]</sup>, 可进行种水平的分析。

目前可以实现 16S rRNA 基因全长测序的高通量测序技术包括 PacBio CCS 测序技术(图 2-④⑤)、ONT 测序技术(图 2-⑥)和基于 UMI 长读长建库的 Illumina 测序(图 2-③), 而后者相对于前两者更为繁琐。随着 PacBio 测序成本的下降和 ONT 测序精确度的不断提高, 未来基于三代测序技术的 rRNA 全长测序必将成为海洋微生物多样性分析的主流。



## 4.2 SSU 测序、LSU 测序和 ITS 测序技术的比较

如前文所述, rRNA 基因操纵子包括 SSU、ITS 和 LSU 等结构。SSU rRNA 基因因其长度较短, 易于测通全长, 从第一代测序技术开始就成为了主要研究对象。虽然 16S rRNA 基因测序是研究微生物组成和发现新类群的有效方法, 但是在很多情况下, 其长度范围内提供的序列信息仍不足以可靠地区分菌株和构建足够分辨率的系统发育树。例如, 研究表明 SAR11 细菌基因组中 16S rRNA 基因两侧存在高度可变的区域, 仅对 16S rRNA 基因进行分析将低估不同 SAR11 基因组间实际的多样性<sup>[121]</sup>。此外, 18S rRNA 基因在真核生物中进化得不够快, 也无法用于高分辨率的分类学研究<sup>[122]</sup>。与 SSU rRNA 相比, LSU rRNA 更长, 因此能为微生物多样性分析提供更多信息<sup>[123]</sup>。28S rRNA 基因适合对真菌进行属水平的解析<sup>[124]</sup>。但由于 LSU 序列太长, 目前测序技术很难对其进行全长的精确测序分析, 因此相关研究较少, 相关的引物序列和数据库支持也非常不足, 影响了其在微生物多样性分析中的应用。ITS 位于 rRNA 编码基因之间, 位点通常有较高的进化速率, 被认为是真菌鉴定的条形码基因<sup>[125-126]</sup>

随着三代长读长测序技术的不断优化, rRNA 全操纵子测序成为未来不可避免的选择。近期, Martijn 等开发了对原核生物 rRNA 全操纵子 (16S-ITS-5.8S-23S) 区域进行扩增测序的方法<sup>[93]</sup>。经过 PacBio 的 CCS 策略提高精确率, 整体读取错误率低至 0.18%。在几种不同的环境样本中捕获到了大量近乎全长的全操纵子序列。将其与 SILVA 数据库相比, 发现了大量新的古菌 23S rRNA 基因序列, 并实现了更高分辨率的物种鉴定<sup>[93]</sup>。在另一项研究中, Karst 等将 UMI 技术与 PacBio 和 Oxford

Nanopore MinION 测序结合, 通过构建共识序列的方式获得大于 10 kb、平均错误率低于 0.0049% 的长读长扩增子<sup>[127]</sup>。但目前用于全操纵子扩增的引物还十分有限, 这些研究所使用引物的扩增效率和偏差还有待更多研究的检验。此外, 由于并不是所有微生物的 rRNA 操纵子各个组分在基因组内都严格按照 16S-ITS-5.8S-23S 的顺序排列, 因此该方法难以对不按照此规律排列的微生物进行有效扩增, 从而可能严重低估微生物多样性。

## 4.3 rRNA 和 rRNA 基因测序的区别

rRNA 是 rRNA 基因(DNA)的转录产物。因为 DNA 提取和随后 PCR 引物扩增的便捷性, 目前大多数研究都是基于 rRNA 基因的扩增子测序。然而, 由于 DNA 相对于 RNA 的耐降解性较高, 环境样品中存在大量死细胞中未降解残留的 rRNA 基因。这部分序列也会被引物扩增, 从而干扰对于微生物群落多样性的分析。与此相对, RNA 分子比 DNA 分子更不稳定, 会因细胞内和细胞外普遍存在核糖核酸酶(RNase)而迅速降解<sup>[128]</sup>。同时, 每个细胞的 rRNA 拷贝数与细胞活性相关<sup>[129]</sup>。因此, 环境样品中的 rRNA 丰度能直接反映其所对应微生物物种的转录活性<sup>[130]</sup>, 从而可以研究环境微生物群落的代谢活跃程度——也就是发现直接发挥生态功能的类群<sup>[131]</sup>。对于环境中丰度低但活性高的微生物, 通过 16S rRNA 与 16S rRNA 基因拷贝的比例可以衡量其在群落中的活跃程度<sup>[132]</sup>。此外, 基于 rRNA 基因得到的物种相对丰度分析还受到微生物基因组中 rRNA 重复基因数量的干扰。而基于 rRNA 的分析则不受这种干扰影响。比如, 由于浮游植物物种之间存在大量的 rRNA 基因拷贝数变异, 使用基于 rRNA 测序的物种组成分析结果比基于 rRNA 基因的更加符合生物量

指标<sup>[133]</sup>。然而对 rRNA 进行测序需要从各类环境中提取高质量 RNA, 该操作经常存在较大挑战。而后续将纯化后的 rRNA 反转录成 cDNA, 并进行引物扩增后测序, 同样会引入因扩增效率引发的偏差。

#### 4.4 无引物扩增的 rRNA 测序技术

为了避免前文所述的引物扩增效率对微生物群落多样性分析的影响, 可以将 rRNA 反转录后的 cDNA 直接进行测序。然而, 细菌与古菌的 rRNA 因为不存在聚合 A 尾(PolyA tail), 无法通过 dT 引物对其进行直接反转录并添加标签用于建库测序。Botero 等开发了一种将聚合 A 尾加至环境样品中 RNA 的方法<sup>[134]</sup>。而 Karst 等利用这项技术, 对各类环境样品中的 rRNA 加聚合 A 尾后进行反转录, 并用长片段建库方法, 实现全长 SSU rRNA 的 Illumina 测序分析<sup>[135]</sup> (图 2-③)。通过研究包括海洋沉积物在内的 7 类环境的微生物组样本, 捕获了原始错误率为 0.17% 的高质量全长 SSU rRNA 序列。因为这一方法避免了引物扩增的偏好性, 并第一次发现大量的新物种 rRNA 序列, 以至于能将 SILVA 数据库中的非冗余 SSU 序列量提高一倍, 对完善微生物的分类和对系统发育树的构建具有重要意义, 特别是在海洋沉积物中发现了大量古菌新类群。但该方法的操作步骤繁琐, 特别是 16S rRNA 的胶纯化后经 UMI 长读长建库过程繁琐, 是否普遍适用于微生物多样性分析仍需要论证。此外, 如前所述, UMI 技术只能构建小于 2 kb 的长读长, 因此无法对 LSU rRNA 进行建库测序。除了使用聚合 A 尾以外, Maki 等将 RNA 寡核苷酸直接连接到 rRNA 的 5'端并构建反转录文库来实现对 SSU rRNA 免引物扩增测序<sup>[136]</sup>。

以上两项研究都采用电泳切胶的方法从总

rRNA 样品中分离到 SSU rRNA 进行建库测序。而用三代长读长测序技术理论上可以不用切胶分离 SSU rRNA 而直接对全部 rRNA 反转录后的产物进行建库测序, 从而显著简化建库步骤(图 2-④)。但目前尚无相关研究报道。另一方面, rRNA 反转录成 cDNA 也会引入一定的技术误差, 而且逆转录酶对 ONT 平台的读长会产生不良影响<sup>[137]</sup>。ONT 技术理论上可以直接对 rRNA 分子进行读取测序<sup>[138]</sup>(图 2-⑦), 但目前尚不具备成熟应用。

#### 4.5 宏基因组 SSU 片段重构法

宏基因组测序结果数据中的部分序列为 rRNA 基因序列, 因此可以利用这部分序列进行微生物群落多样性分析。该方法首先从宏基因组原始测序数据中提取 rRNA 相关片段, 随后或对这些序列直接进行物种注释<sup>[139]</sup>, 或通过从头组装或参考模板序列组装的方式<sup>[140]</sup>获取全长 rRNA 片段。该方法能避免引物扩增带来的偏好性影响, 因此能更真实地反映微生物群落结构特征, 也能发现引物扩增无法发现的海洋微生物新物种。例如, 在海绵共生微生物的多样性研究中, rRNA 基因扩增中采用的简并引物都是基于研究较为透彻的海洋浮游微生物序列设计的, 因此相关研究结果很可能会低估共生微生物的多样性。我们的前期工作采用从宏基因组片段中重构 SSU rRNA 序列的方法, 分析了两种海绵动物的共生微生物群落, 发现了大量引物扩增所没有发现的微生物类群<sup>[141]</sup>, 且这些物种与海水中的浮游微生物种类截然不同, 有重要的生态和进化意义<sup>[142]</sup>。

但此类方法也存在局限性。首先, rRNA 基因序列只占宏基因组测序总数据中的较少一部分, 通常这些 rRNA 相关序列只能覆盖群落中丰度较高的物种, 而无法对低丰度物种多样性进行分析。

其次, 这类方法为获取全长 rRNA 片段, 需对短读长片段进行混合从头组装。该过程不可避免会引入组装错误产生的嵌合体序列, 从而影响多样性分析的可靠性。

#### 4.6 rRNA 序列与微生物群落功能多样性

虽然基于 rRNA 序列本身只能获取微生物物种多样性信息, 但由于进化关系相近的微生物其基因组功能组成也具有一定的相似性<sup>[143]</sup>, 所以可以基于 rRNA 序列获得微生物分类学信息, 推测其对应微生物的功能基因组成。相关软件包括 PICRUSt<sup>[144]</sup>和 Tax4Fun<sup>[145]</sup>等。然而, 这项技术的可预测性和预测准确性都严重依赖于参考基因组数据库的完整度<sup>[146]</sup>。由于目前公共基因组数据库中主要包含人体(特别是肠道)微生物, 而来自环境中的微生物相对稀少。因此该技术在预测环境微生物群落, 特别是分析海洋沉积物、海洋极端环境等生境时, 其可用性较差<sup>[147]</sup>。此外, PICRUSt 和 Tax4fun 等软件基于 SILVA 数据库的版本更新不及时也限制两个软件的广泛使用。

除此之外, 细胞内融合基因技术(epicPCR)将 16S rRNA 基因与所要研究的功能基因在细胞内进行融合, 从而可以同时研究微生物群落中的物种多样性和功能多样性<sup>[148]</sup>。

## 5 综合运用测序技术研究海洋微生物多样性的具体问题

### 5.1 估算全球海洋微生物多样性

目前用 rRNA 序列估算全球原核微生物多样性的方法是将全长 SSU rRNA 序列以 97%序列一致性为阈值构建 OTU, 以一个 OTU 为物种单位进行计算。而计算模型的估算准确度依然受采集

自不同地域和环境的参考数据集的影响<sup>[149]</sup>。然而, 由于存在严重的采样偏好性和引物特异性偏差, 目前公共数据库中海洋微生物, 特别是海洋沉积物等环境中的全长 SSU rRNA 序列仍然不足<sup>[135]</sup>。因此, 为了更准确地对全球海洋微生物多样性进行估算, 首先要做的是对全球不同海域内更多类型的海洋环境如海洋极端环境<sup>[76]</sup>、海洋沉积物<sup>[135]</sup>、海洋生物被膜<sup>[150]</sup>和海洋共生微生物<sup>[151]</sup>等进行 SSU rRNA 全长测序。随着 PacBio CCS 技术的成熟商业化和 ONT R10 高精度版本的推出, 接下来几年必将有大量海洋微生物研究会采用全长 SSU rRNA 扩增子测序, 从而快速补充公共数据库。此外, 随着免引物的 rRNA 直接反转录后测序或 rRNA 直接测序技术的开发, 相关研究也能补充扩增子测序中由于引物偏好性而忽略的微生物类群, 从而进一步完善公共数据库中海洋微生物序列的代表性。随着公共数据库数据的量变积累, 我们将很快能够较为准确地预测出全球海洋微生物的物种多样性范围<sup>[152]</sup>。

### 5.2 应用于不同时空尺度范围的海洋微生物生态学研究

测序技术的快速和多样性发展为海洋微生物生态学研究提供了丰富选择, 以满足不同类型研究的需求。对于预算有限、样本量需求较少的微生物多样性研究, 每个样本的测序深度、每条序列的精确度和信息量无疑是最重要的。因此, 虽然基于 Illumina 的 SSU rRNA 部分区域扩增子测序仍然是当前研究的主流, 但随着 PacBio CCS 的价格快速降低, 三代全长 SSU rRNA 测序很快对前者的统治地位产生挑战。与之相比, 长片段合成技术因其操作更为复杂, 目前并没有太多测序服务公司开展相关服务, 限制了其成本降低和应

用范围扩展。

对于需要在一定预算范围内尽可能对更多环境进行采样的大时空尺度研究, 分配到每个样本的预算可能更为有限。因此基于 Illumina 或 BGISEq 测序平台的 SSU rRNA 部分区域扩增子测序因其较为突出的测序深度和价格比而成为目前的必然选择。虽然读长较短限制了物种分析的准确性, 但尚能满足在大尺度范围内揭示生态学规律。

而在微生物多样性的实时监测应用中(如出海航次采样过程中及时评估微生物多样性, 以指导进一步采样计划), ONT 因其便携性和用户可操作性成为未来的必然选择。随着其测序精度进一步提升(如 R10.3 版本)和单次测序成本的进一步下降(如 Flongle 芯片), 其必将成为海洋微生物 rRNA 多样性研究的利器。

### 5.3 菌株水平的微生物多样性研究

通常序列一致性高于 97% 的微生物被认为属于同一物种的不同菌株。16S 序列也被用来区分菌株(有时称为亚种)的基因多态性, 其中单核苷酸多态性(Single-nucleotide polymorphisms)已被用于预测表型特征<sup>[153]</sup>。但是由于 SSU rRNA 基因在进化中较为保守, 所以即使具有完全相同的全长 SSU rRNA 基因序列的两个微生物也可能具有不同的基因组结构和生理生态功能。为了在更精细的水平上分析微生物多样性, 需要更多序列信息。因此, 除了在测序精度上要达到能区分单碱基水平外, LSU 甚至是全 rRNA 操纵子测序已成为这个领域的发展方向。因此基于单分子的高精度三代长读长测序技术将在此领域发挥更多作用。此外, SSU 序列相同的不同菌株往往在 rRNA 操纵子的拷贝数上也存在差异。通过 PacBio 对 SSU

rRNA 全长甚至全操纵子测序的方法, 也能基于单核苷酸多态性鉴别不同菌株之间的拷贝数差异, 从而改进对物种甚至菌株的区分<sup>[86]</sup>。

## 6 未来测序技术发展对海洋微生物 rRNA 多样性研究的影响

基于 rRNA 序列测序的分析技术因其便捷、成本低、灵敏度高而必将长期成为海洋微生物多样性研究的支柱技术之一。未来测序技术将继续向高通量、长读长、低误差、短时间、可便携和易分析等方向差异化快速发展。而这些指标的提升将直接促进基于 rRNA 测序技术的海洋微生物多样性研究。随着长读长高精度测序技术的普及应用, 将使海洋微生物全长 SSU 数据库更为完善, 并极大改善 LSU 乃至全操纵子数据严重缺乏的现状。而这些数据库的丰富, 反过来有利于我们设计更为丰富也更为合理的引物序列(特别是 LSU 和全操纵子扩增序列), 减少扩增偏好性, 增加特定类群引物的准确性, 从而普遍提升研究的可重复性。与此同时, 测序长度和精度的提高, 将大大促进海洋微生物生态学研究的解析度。基于菌株水平和基于 rRNA 的生态学模型, 在理论上能更好地与微生物的生理生态功能以及环境参数相关联, 从而对现有部分基于 OTU 和 DNA 数据的生态理论进行修正。

另一方面, 测序技术的不断发展, 也对分析软件和配套试剂的开发、人员的培训提出了更多要求。因此, 海洋微生物研究者将不可避免地需要及时了解和学习测序技术的发展, 并结合自己的研究需求合理规划测序方案。甚至基于已有技术, 开发新的建库和测序策略。从而能使海洋微生物多样性研究及时获益于测序技术的进步成果。

## 参考文献

- [1] Whitman WB, Coleman DC, Wiebe WJ. Prokaryotes: the unseen majority. *Proceedings of the National Academy of Sciences of the United States of America*, 1998, 95(12): 6578–6583.
- [2] Fuhrman JA, Cram JA, Needham DM. Marine microbial community dynamics and their ecological interpretation. *Nature Reviews Microbiology*, 2015, 13(3): 133–146.
- [3] Hansell DA, Carlson CA, Repeta DJ, Schlitzer R. Dissolved organic matter in the ocean: a controversy stimulates new insights. *Oceanography*, 2009, 22(4): 202–211.
- [4] Dyrman ST, Ammerman JW, Van Mooy BAS. Microbes and the marine phosphorus cycle. *Oceanography*, 2007, 20(2): 110–116.
- [5] Hutchins DA, Mulholland MR, Fu FX. Nutrient cycles and marine microbes in a CO<sub>2</sub>-enriched ocean. *Oceanography*, 2009, 22(4): 128–145.
- [6] Arrigo KR. Marine microorganisms and global nutrient cycles. *Nature*, 2005, 437(7057): 349–355.
- [7] Spackeen JL, Sipler RE, Xu K, Tatters AO, Walworth NG, Bertrand EM, McQuaid JB, Hutchins DA, Allen AE, Bronk DA. Interactive effects of elevated temperature and CO<sub>2</sub> on nitrate, urea, and dissolved inorganic carbon uptake by a coastal California, USA, microbial community. *Marine Ecology Progress Series*, 2017, 577: 49–65.
- [8] Lu ZM, Deng Y, Van Nostrand JD, He ZL, Voordeckers J, Zhou AF, Lee YJ, Mason OU, Dubinsky EA, Chavarria KL, Tom LM, Fortney JL, Lamendella R, Jansson JK, D'haeseleer P, Hazen TC, Zhou JZ. Microbial gene functions enriched in the deepwater horizon deep-sea oil plume. *The ISME Journal*, 2012, 6(2): 451–460.
- [9] Bai SJ, Huang LP, Su JQ, Tian Y, Zheng TL. Algicidal effects of a novel marine actinomycete on the toxic dinoflagellate *Alexandrium tamarense*. *Current Microbiology*, 2011, 62(6): 1774–1781.
- [10] Rodríguez GR, Villasante S, Do Carme García-Negro M. Are red tides affecting economically the commercialization of the Galician (NW Spain) mussel farming? *Marine Policy*, 2011, 35(2): 252–257.
- [11] Hering D, Borja A, Jones JI, Pont D, Boets P, Bouchez A, Bruce K, Drakare S, Hänfling B, Kahlert M, Leese F, Meissner K, Mergen P, Reyjol Y, Segurado P, Vogler A, Kelly M. Implementation options for DNA-based identification into ecological status assessment under the European water framework directive. *Water Research*, 2018, 138: 192–205.
- [12] Shokralla S, Spall JL, Gibson JF, Hajibabaei M. Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology*, 2012, 21(8): 1794–1805.
- [13] Valentini A, Taberlet P, Miaud C, Civade R, Herder J, Thomsen PF, Bellemain E, Besnard A, Coissac E, Boyer F, Gaboriaud C, Jean P, Poulet N, Roset N, Copp GH, Geniez P, Pont D, Argillier C, Baudoin JM, Peroux T, Crivelli AJ, Olivier A, Acqueberge M, Le Brun M, Møller PR, Willerslev E, Dejean T. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology*, 2016, 25(4): 929–942.
- [14] Arakaki A, Shibusawa M, Hosokawa M, Matsunaga T. Preparation of genomic DNA from a single species of uncultured magnetotactic bacterium by multiple-displacement amplification. *Applied and Environmental Microbiology*, 2010, 76(5): 1480–1485.
- [15] Rosselló-Mora R, Amann R. The species concept for prokaryotes. *FEMS Microbiology Reviews*, 2001, 25(1): 39–67.
- [16] Doolittle WF, Zhaxybayeva O. On the origin of prokaryotic species. *Genome Research*, 2009, 19(5): 744–756.
- [17] Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences of the United States of America*, 1990, 87(12): 4576–4579.
- [18] Davidson BS. New dimensions in natural products research: cultured marine microorganisms. *Current Opinion in Biotechnology*, 1995, 6(3): 284–291.
- [19] Vandamme P, Pot B, Gillis M, De Vos P, Kersters K, Swings J. Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiological Reviews*, 1996, 60(2): 407–438.
- [20] Stott MB, Crowe MA, Mountain BW, Smirnova AV, Hou SB, Alam M, Dunfield PF. Isolation of novel bacteria, including a candidate division, from geothermal soils in New Zealand. *Environmental Microbiology*, 2008, 10(8): 2030–2041.
- [21] Joint I, Mühling M, Querellou J. Culturing marine bacteria—an essential prerequisite for biodiscovery. *Microbial Biotechnology*, 2010, 3(5): 564–575.
- [22] Woese CR, Fox GE. Concept of cellular evolution. *Journal of Molecular Evolution*, 1977, 10(1): 1–6.
- [23] Glöckner FO, Yilmaz P, Quast C, Gerken J, Beccati A, Ciuprina A, Bruns G, Yarza P, Peplies J, Westram R, Ludwig

- W. 25 years of serving the community with ribosomal RNA gene reference databases and tools. *Journal of Biotechnology*, 2017, 261: 169–176.
- [24] Ward DM, Weller R, Bateson MM. 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature*, 1990, 345(6270): 63–65.
- [25] Knittel K, Lösekann T, Boetius A, Kort R, Amann R. Diversity and distribution of methanotrophic archaea at cold seeps. *Applied and Environmental Microbiology*, 2005, 71(1): 467–479.
- [26] Rappé MS, Giovannoni SJ. The uncultured microbial majority. *Annual Review of Microbiology*, 2003, 57: 369–394.
- [27] Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B. How many species are there on earth and in the ocean? *PLoS Biology*, 2011, 9(8): e1001127.
- [28] Louca S, Mazel F, Doebeli M, Parfrey LW. A census-based estimate of earth's bacterial and archaeal diversity. *PLoS Biology*, 2019, 17(2): e3000106.
- [29] Heidelberg KB, Gilbert JA, Joint I. Marine genomics: at the interface of marine microbial ecology and biodiscovery. *Microbial Biotechnology*, 2010, 3(5): 531–543.
- [30] Klappenbach JA, Saxman PR, Cole JR, Schmidt TM. Rnldb: the ribosomal RNA operon copy number database. *Nucleic Acids Research*, 2001, 29(1): 181–184.
- [31] Green R, Noller HF. Ribosomes and translation. *Annual Review of Biochemistry*, 1997, 66: 679–716.
- [32] Woese CR. Bacterial evolution. *Microbiological Reviews*, 1987, 51(2): 221–271.
- [33] Ludwig W, Schleifer KH. Bacterial phylogeny based on 16S and 23S rRNA sequence analysis. *FEMS Microbiology Reviews*, 1994, 15(2/3): 155–173.
- [34] Lafontaine DLJ, Tollervey D. The function and synthesis of ribosomes. *Nature Reviews Molecular Cell Biology*, 2001, 2(7): 514–520.
- [35] Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research*, 2013, 41(1): e1.
- [36] Clarridge III JE. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clinical Microbiology Reviews*, 2004, 17(4): 840–862.
- [37] Tringe SG, Hugenholtz P. A renaissance for the pioneering 16S rRNA gene. *Current Opinion in Microbiology*, 2008, 11(5): 442–446.
- [38] Orsini L, Procaccini G, Sarno D, Montresor M. Multiple rDNA ITS-types within the diatom *Pseudo-Nitzschia Delicatissima* (Bacillariophyceae) and their relative abundances across a spring bloom in the Gulf of Naples. *Marine Ecology Progress Series*, 2004, 271: 87–98.
- [39] Sheffield VC, Cox DR, Lerman LS, Myers RM. Attachment of a 40-base-pair G+C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. *Proceedings of the National Academy of Sciences of the United States of America*, 1989, 86(1): 232–236.
- [40] Holben WE, Harris D. DNA-based monitoring of total bacterial community structure in environmental samples. *Molecular Ecology*, 1995, 4(5): 627–631.
- [41] Apajalahti JHA, Kettunen H, Kettunen A, Holben WE, Nurminen PH, Rautonen N, Mutanen M. Culture-independent microbial community analysis reveals that inulin in the diet primarily affects previously unknown bacteria in the mouse cecum. *Applied and Environmental Microbiology*, 2002, 68(10): 4986–4995.
- [42] Gordon DP, Mawatari SF, Kajihara H. New taxa of Japanese and New Zealand eurytomellidae (phylum Bryozoa) and their phylogenetic relationships. *Zoological Journal of the Linnean Society*, 2002, 136(2): 199–216.
- [43] Vetriani C, Jannasch HW, MacGregor BJ, Stahl DA, Reysenbach AL. Population structure and phylogenetic characterization of marine benthic archaea in deep-sea sediments. *Applied and Environmental Microbiology*, 1999, 65(10): 4375–4384.
- [44] Osborn AM, Moore ERB, Timmis KN. An evaluation of terminal-restriction fragment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics. *Environmental Microbiology*, 2000, 2(1): 39–50.
- [45] Schütte UME, Abdo Z, Bent SJ, Shyu C, Williams CJ, Pierson JD, Forney LJ. Advances in the use of terminal restriction fragment length polymorphism (T-RFLP) analysis of 16S rRNA genes to characterize microbial communities. *Applied Microbiology and Biotechnology*, 2008, 80(3): 365–380.
- [46] Babcock DA, Wawrik B, Paul JH, McGuinness L, Kerkhof LJ. Rapid screening of a large insert BAC library for specific 16S

- rRNA genes using TRFLP. *Journal of Microbiological Methods*, 2007, 71(2): 156–161.
- [47] Orsi W, Biddle JF, Edgcomb V. Deep sequencing of subseafloor eukaryotic rRNA reveals active fungi across marine subsurface provinces. *PLoS One*, 2013, 8(2): e56335.
- [48] Wagner M, Nielsen PH, Loy A, Nielsen JL, Daims H. Linking microbial community structure with function: Fluorescence *in situ* hybridization-microautoradiography and isotope arrays. *Current Opinion in Biotechnology*, 2006, 17(1): 83–91.
- [49] Amann R, Fuchs BM. Single-cell identification in microbial communities by improved fluorescence *in situ* hybridization techniques. *Nature Reviews Microbiology*, 2008, 6(5): 339–348.
- [50] Morris RM, Rappé MS, Connon SA, Vergin KL, Siebold WA, Carlson CA, Giovannoni SJ. SAR11 clade dominates ocean surface bacterioplankton communities. *Nature*, 2002, 420(6917): 806–810.
- [51] Karner MB, DeLong EF, Karl DM. Archaeal dominance in the mesopelagic zone of the pacific ocean. *Nature*, 2001, 409(6819): 507–510.
- [52] Huber D, Kaigala GV. Rapid micro fluorescence *in situ* hybridization in tissue sections. *Biomicrofluidics*, 2018, 12(4): 042212.
- [53] Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. *Nature Reviews Genetics*, 2019, 20(11): 631–656.
- [54] Alegria Á, Szczesny P, Mayo B, Bardowski J, Kowalczyk M. Biodiversity in oscypek, a traditional polish cheese, determined by culture-dependent and -independent approaches. *Applied and Environmental Microbiology*, 2012, 78(6): 1890–1898.
- [55] Kirchman DL, Cottrell MT, Lovejoy C. The structure of bacterial communities in the western arctic ocean as revealed by pyrosequencing of 16S rRNA genes. *Environmental Microbiology*, 2010, 12(5): 1132–1143.
- [56] Claesson MJ, Wang Q, O’Sullivan O, Greene-Diniz R, Cole JR, Ross RP, O’Toole PW. Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Research*, 2010, 38(22): e200.
- [57] Lanzén A, Jørgensen SL, Bengtsson MM, Jonassen I, Øvreås L, Urich T. Exploring the composition and diversity of microbial communities at the jan mayen hydrothermal vent field using RNA and DNA. *FEMS Microbiology Ecology*, 2011, 77(3): 577–589.
- [58] Stevens JL, Jackson RL, Olson JB. Slowing PCR ramp speed reduces chimera formation from environmental samples. *Journal of Microbiological Methods*, 2013, 93(3): 203–205.
- [59] Fonseca VG, Nichols B, Lallias D, Quince C, Carvalho GR, Power DM, Creer S. Sample richness and genetic diversity as drivers of chimera formation in nSSU metagenetic analyses. *Nucleic Acids Research*, 2012, 40(9): e66.
- [60] Siegwald L, Touzet H, Lemoine Y, Hot D, Audebert C, Caboche S. Assessment of common and emerging bioinformatics pipelines for targeted metagenomics. *PLoS One*, 2017, 12(1): e0169563.
- [61] Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology*, 2017, 35(9): 833–844.
- [62] Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 1977, 74(12): 5463–5467.
- [63] Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang JH, Miklos GLG, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng ZM, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan WN, Ge WM, Gong FC, Gu ZP, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke ZX, Ketchum KA, Lai ZW, Lei YD, Li ZY, Li JY, Liang Y, Lin XY, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun JT, Wang ZY, Wang AH, Wang X, Wang J, Wei MH, Wides R, Xiao CL, Yan CH, Yao A, Ye J, Zhan M, Zhang WQ, Zhang HY, Zhao Q, Zheng LS, Zhong F, Zhong WY, Zhu SC, Zhao SY, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An HJ, Awe A, Baldwin D, Baden H, Barnstead M,

- Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferreira S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigó R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi HY, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays AD, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu XJ, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen MY, Wu D, Wu M, Xia A, Zandieh A, Zhu XH. The sequence of the human genome. *Science*, 2001, 291(5507): 1304–1351.
- [64] Bonetta L. Genome sequencing in the fast lane. *Nature Methods*, 2006, 3(2): 141–147.
- [65] DeLong EF. Archaea in coastal marine environments. *Proceedings of the National Academy of Sciences of the United States of America*, 1992, 89(12): 5685–5689.
- [66] Fuhrman JA, McCallum K, Davis AA. Novel major archaeobacterial group from marine plankton. *Nature*, 1992, 356(6365): 148–149.
- [67] Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S, Wu DY, Eisen JA, Hoffman JM, Remington K, Beeson K, Tran B, Smith H, Baden-Tillson H, Stewart C, Thorpe J, Freeman J, Andrews-Pfannkoch C, Venter JE, Li K, Kravitz S, Heidelberg JF, Utterback T, Rogers YH, Falcon LI, Souza V, Bonilla-Rosso G, Eguarte LE, Karl DM, Sathyendranath S, Platt T, Bermingham E, Gallardo V, Tamayo-Castillo G, Ferrari MR, Strausberg RL, Nealson K, Friedman R, Frazier M, Venter JC. The *Sorcerer II* global ocean sampling expedition: northwest atlantic through eastern tropical Pacific. *PLoS Biology*, 2007, 5(3): e77.
- [68] Yutin N, Suzuki MT, Teeling H, Weber M, Venter JC, Rusch DB, Bèjà O. Assessing diversity and biogeography of aerobic anoxygenic phototrophic bacteria in surface waters of the atlantic and pacific oceans using the global ocean sampling expedition metagenomes. *Environmental Microbiology*, 2007, 9(6): 1464–1475.
- [69] Temperton B, Field D, Oliver A, Tiwari B, Mühling M, Joint I, Gilbert JA. Bias in assessments of marine microbial biodiversity in fosmid libraries as evaluated by pyrosequencing. *The ISME Journal*, 2009, 3(7): 792–796.
- [70] Ronaghi M. Pyrosequencing sheds light on DNA sequencing. *Genome Research*, 2001, 11(1): 3–11.
- [71] Gloor GB, Hummelen R, Macklaim JM, Dickson RJ, Fernandes AD, MacPhee R, Reid G. Microbiome profiling by Illumina sequencing of combinatorial sequence-tagged PCR products. *PLoS One*, 2010, 5(10): e15406.
- [72] Callaway E. Genome giant offers data service. *Nature*, 2011, 475(7357): 435–437.
- [73] Salipante SJ, Kawashima T, Rosenthal C, Hoogstraat DR, Cummings LA, Sengupta DJ, Harkins TT, Cookson BT, Hoffman NG, Löffler FE. Performance comparison of Illumina and Ion Torrent next-generation sequencing platforms for 16S rRNA-based bacterial community profiling. *Applied and Environmental Microbiology*, 2014, 80(24): 7583–7591.
- [74] Schloss PD, Gevers D, Westcott SL. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS One*, 2011, 6(12): e27310.
- [75] Klier J, Dellwig O, Leipe T, Jürgens K, Herlemann DPR. Benthic bacterial community composition in the oligohaline-marine transition of surface sediments in the baltic sea based on rRNA analysis. *Frontiers in Microbiology*, 2018, 9: 236.
- [76] Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, Arrieta JM, Herndl GJ. Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proceedings of the National Academy of Sciences of the United States of*



- America*, 2006, 103(32): 12115–12120.
- [77] Huber JA, Welch DB, Morrison HG, Huse SM, Neal PR, Butterfield DA, Sogin ML. Microbial population structures in the deep marine biosphere. *Science*, 2007, 318(5847): 97–100.
- [78] Kan JJ, Clingenpeel S, Macur RE, Inskeep WP, Lovalvo D, Varley J, Gorby Y, McDermott TR, Nealson K. *Archaea* in yellowstone lake. *The ISME Journal*, 2011, 5(11): 1784–1795.
- [79] Harismendy O, Ng PC, Strausberg RL, Wang XY, Stockwell TB, Beeson KY, Schork NJ, Murray SS, Topol EJ, Levy S, Frazer KA. Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biology*, 2009, 10(3): R32.
- [80] Glenn TC. Field guide to next-generation DNA sequencers. *Molecular Ecology Resources*, 2011, 11(5): 759–769.
- [81] Walters WA, Caporaso JG, Lauber CL, Berg-Lyons D, Fierer N, Knight R. Primerprospector: *de novo* design and taxonomic analysis of barcoded polymerase chain reaction primers. *Bioinformatics*, 2011, 27(8): 1159–1161.
- [82] Bork P, Bowler C, De Vargas C, Gorsky G, Karsenti E, Wincker P. *Tara* Oceans studies plankton at planetary scale. *Science*, 2015, 348(6237): 873.
- [83] Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A, Cornejo-Castillo FM, Costea PI, Cruaud C, d'Ovidio F, Engelen S, Ferrera I, Gasol JM, Guidi L, Hildebrand F, Kokoszka F, Lepoivre C, Lima-Mendez G, Poulain J, Poulos BT, Royo-Llonch M, Sarmiento H, Vieira-Silva S, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Coordinators TO, Bowler C, de Vargas C, Gorsky G, Grimsley N, Hingamp P, Iudicone D, Jaillon O, Not F, Ogata H, Pesant S, Speich S, Stemmann L, Sullivan MB, Weissenbach J, Wincker P, Karsenti E, Raes J, Acinas SG, Bork P. Structure and function of the global ocean microbiome. *Science*, 2015, 348(6237): 1261359.
- [84] Westcott SL, Schloss PD. *De novo* clustering methods outperform reference-based methods for assigning 16S rRNA gene sequences to operational taxonomic units. *PeerJ*, 2015, 3: e1487.
- [85] Puente-Sanchez F, Aguirre J, Parro V. A novel conceptual approach to read-filtering in high-throughput amplicon sequencing studies. *Nucleic Acids Research*, 2015, 44(4): e40.
- [86] Johnson JS, Spakowicz DJ, Hong BY, Petersen LM, Demkowicz P, Chen L, Leopold SR, Hanson BM, Agresta HO, Gerstein M, Sodergren E, Weinstock GM. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nature Communications*, 2019, 10(1): 5029.
- [87] Smith T, Heger A, Sudbery I. Umi-tools: modeling sequencing errors in unique molecular identifiers to improve quantification accuracy. *Genome Research*, 2017, 27(3): 491–499.
- [88] Hiatt JB, Patwardhan RP, Turner EH, Lee C, Shendure J. Parallel, tag-directed assembly of locally derived short sequence reads. *Nature Methods*, 2010, 7(2): 119–122.
- [89] Stapleton JA, Kim J, Hamilton JP, Wu M, Irber LC, Maddamsetti R, Briney B, Newton L, Burton DR, Brown CT, Chan C, Buell CR, Whitehead TA. Haplotype-phased synthetic long reads from short-read sequencing. *PLoS One*, 2016, 11(1): e0147229.
- [90] Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H. Continuous base identification for single-molecule nanopore DNA sequencing. *Nature Nanotechnology*, 2009, 4(4): 265–270.
- [91] Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 2016, 17(6): 333–351.
- [92] Singer E, Bushnell B, Coleman-Derr D, Bowman B, Bowers RM, Levy A, Gies EA, Cheng JF, Copeland A, Klenk HP, Hallam SJ, Hugenholtz P, Tringe SG, Woyke T. High-resolution phylogenetic microbial community profiling. *The ISME Journal*, 2016, 10(8): 2020–2032.
- [93] Martijn J, Lind AE, Schön ME, Spiertz I, Juzokaite L, Bunikis I, Pettersson OV, Ettema TJG. Confident phylogenetic identification of uncultured prokaryotes through long read amplicon sequencing of the 16S-ITS-23S rRNA operon. *Environmental Microbiology*, 2019, 21(7): 2485–2498.
- [94] Schloss PD, Jenior ML, Koumpouras CC, Westcott SL, Highlander SK. Sequencing 16S rRNA gene fragments using the PacBio SMRT DNA sequencing system. *PeerJ*, 2016, 4: e1869.
- [95] Pootakham W, Mhuantong W, Yoocha T, Putchim L, Sonthirod C, Naktang C, Thongtham N, Tangphatsornruang S. High resolution profiling of coral-associated bacterial communities using full-length 16S rRNA sequence data from PacBio smrt sequencing system. *Scientific Reports*, 2017, 7(1): 2774.

- [96] Bowden R, Davies RW, Heger A, Pagnamenta AT, De Cesare M, Oikkonen LE, Parkes D, Freeman C, Dhalla F, Patel SY, Popitsch N, Ip CLC, Roberts HE, Salatino S, Lockstone H, Lunter G, Taylor JC, Buck D, Simpson MA, Donnelly P. Sequencing of human genomes with nanopore technology. *Nature Communications*, 2019, 10(1): 1869.
- [97] Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, Pallen MJ. Performance comparison of benchtop high-throughput sequencing platforms. *Nature Biotechnology*, 2012, 30(5): 434–439.
- [98] Weirather JL, De Cesare M, Wang YH, Piazza P, Sebastiano V, Wang XJ, Buck D, Au KF. Comprehensive comparison of Pacific Biosciences and Oxford Nanopore technologies and their applications to transcriptome analysis. *F1000Research*, 2017, 6: 100.
- [99] Loman NJ, Quick J, Simpson JT. A complete bacterial genome assembled *de novo* using only nanopore sequencing data. *Nature Methods*, 2015, 12(8): 733–735.
- [100] Sović I, Šikić M, Wilm A, Fenlon SN, Chen S, Nagarajan N. Fast and sensitive mapping of nanopore sequencing reads with graphmap. *Nature Communications*, 2016, 7(1): 11307.
- [101] Leggett RM, Clark MD. A world of opportunities with nanopore sequencing. *Journal of Experimental Botany*, 2017, 68(20): 5419–5429.
- [102] Benítez-Páez A, Portune KJ, Sanz Y. Species-level resolution of 16S rRNA gene amplicons sequenced through the MinION™ portable nanopore sequencer. *Gigascience*, 2016, 5: 4.
- [103] Quick J, Ashton P, Calus S, Chatt C, Gossain S, Hawker J, Nair S, Neal K, Nye K, Peters T, de Pinna E, Robinson E, Struthers K, Webber M, Catto A, Dallman TJ, Hawkey P, Loman NJ. Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of *Salmonella*. *Genome Biology*, 2015, 16(1): 114.
- [104] Curren E, Yoshida T, Kuwahara VS, Leong SCY. Rapid profiling of tropical marine cyanobacterial communities. *Regional Studies in Marine Science*, 2019, 25: 100485.
- [105] Burke CM, Darling AE. A method for high precision sequencing of near full-length 16S rRNA genes on an Illumina MiSeq. *PeerJ*, 2016, 4(12): e2492.
- [106] Jain M, Tyson JR, Loose M, Ip CLC, Eccles DA, O'Grady J, Malla S, Leggett RM, Wallerman O, Jansen HJ, Zalunin V, Birney E, Brown BL, Snutch TP, Olsen HE. MinION analysis and reference consortium: phase 2 data release and analysis of R9.0 chemistry. *F1000Research*, 2017, 6: 760.
- [107] Tyler AD, Mataseje L, Urfano CJ, Schmidt L, Antonation KS, Mulvey MR, Corbett CR. Evaluation of Oxford Nanopore's MinION sequencing device for microbial whole genome sequencing applications. *Scientific Reports*, 2018, 8(1): 10931.
- [108] Noakes MT, Brinkerhoff H, Laszlo AH, Derrington IM, Langford KW, Mount JW, Bowman JL, Baker KS, Doering KM, Tickman BI, Gundlach JH. Increasing the accuracy of nanopore DNA sequencing using a time-varying cross membrane voltage. *Nature Biotechnology*, 2019, 37(6): 651–656.
- [109] Kai S, Matsuo Y, Nakagawa S, Kryukov K, Matsukawa S, Tanaka H, Iwai T, Imanishi T, Hirota K. Rapid bacterial identification by direct PCR amplification of 16S rRNA genes using the MinION™ nanopore sequencer. *FEBS Open Bio*, 2019, 9(3): 548–557.
- [110] Liu ZZ, DeSantis TZ, Andersen GL, Knight R. Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. *Nucleic Acids Research*, 2008, 36(18): e120.
- [111] Wang Q, Garrity GM, Tiedje JM, Cole JR. Naïve bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, 2007, 73(16): 5261–5267.
- [112] Chakravorty S, Helb D, Burday M, Connell N, Alland D. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *Journal of Microbiological Methods*, 2007, 69(2): 330–339.
- [113] Srinivasan R, Karaöz U, Volegova M, MacKichan J, Kato-Maeda M, Miller S, Nadarajan R, Brodie EL, Lynch SV. Use of 16S rRNA gene for identification of a broad range of clinically relevant bacterial pathogens. *PLoS One*, 2015, 10(2): e0117617.
- [114] Yang B, Wang Y, Qian PY. Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC Bioinformatics*, 2016, 17: 135.
- [115] Mignard S, Flandrois JP. 16S rRNA sequencing in routine bacterial identification: a 30-month experiment. *Journal of Microbiological Methods*, 2006, 67(3): 574–581.
- [116] Winand R, Bogaerts B, Hoffman S, Lefevre L, Delvoye M, van Braekel J, Fu Q, Roosens NH, de Keersmaecker SC,

- Vanneste K. Targeting the 16S rRNA gene for bacterial identification in complex mixed samples: comparative evaluation of second (Illumina) and third (Oxford Nanopore Technologies) generation sequencing technologies. *International Journal of Molecular Sciences*, 2020, 21(1): 298.
- [117] Cao Y, Fanning S, Proos S, Jordan K, Srikumar S. A review on the applications of next generation sequencing technologies as applied to food-related microbiome studies. *Frontiers in Microbiology*, 2017, 8: 1829.
- [118] Kim M, Morrison M, Yu ZT. Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes. *Journal of Microbiological Methods*, 2011, 84(1): 81–87.
- [119] Wagner J, Coupland P, Browne HP, Lawley TD, Francis SC, Parkhill J. Evaluation of PacBio sequencing for full-length bacterial 16S rRNA gene classification. *BMC Microbiology*, 2016, 16(1): 274.
- [120] Morales SE, Cosart TF, Johnson JV, Holben WE. Extensive phylogenetic analysis of a soil bacterial community illustrates extreme taxon evenness and the effects of amplicon length, degree of coverage, and DNA fractionation on classification and ecological parameters. *Applied and Environmental Microbiology*, 2009, 75(3): 668–675.
- [121] Gilbert JA, Muhling M, Joint I. A rare SAR11 fosmid clone confirming genetic variability in the ‘*Candidatus Pelagibacter ubique*’ genome. *The ISME Journal*, 2008, 2(7): 790–793.
- [122] Vandenkoornhuyse P, Baldauf SL, Leyval C, Straczek J, Young JPW. Extensive fungal diversity in plant roots. *Science*, 2002, 295(5562): 2051.
- [123] Ludwig W, Strunk O, Klugbauer S, Klugbauer N, Weizenegger M, Neumaier J, Bachleitner M, Schleifer KH. Bacterial phylogeny based on comparative sequence analysis (review). *Electrophoresis*, 1998, 19(4): 554–568.
- [124] Liu KL, Porras-Alfaro A, Kuske CR, Eichorst SA, Xie G. Accurate, rapid taxonomic classification of fungal large-subunit rRNA genes. *Applied and Environmental Microbiology*, 2012, 78(5): 1523–1533.
- [125] Thiéry O, Vasar M, Jairus T, Davison J, Roux C, Kivistik PA, Metspalu A, Milani L, Saks Ü, Moora M, Zobel M, Öpik M. Sequence variation in nuclear ribosomal small subunit, internal transcribed spacer and large subunit regions of *Rhizophagus irregularis* and *Gigaspora margarita* is high and isolate-dependent. *Molecular Ecology*, 2016, 25(12): 2816–2832.
- [126] Seifert KA. Progress towards DNA barcoding of fungi. *Molecular Ecology Resources*, 2009, 9(S1): 83–89.
- [127] Karst SM, Ziels RM, Kirkegaard RH, Sørensen EA, McDonald D, Zhu QY, Knight R, Albertsen M. Enabling high-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. *bioRxiv*, 2020: 645903.
- [128] Houseley J, Tollervey D. The many pathways of RNA degradation. *Cell*, 2009, 136(4): 763–776.
- [129] Jones SE, Lennon JT. Dormancy contributes to the maintenance of microbial diversity. *Proceedings of the National Academy of Sciences of the United States of America*, 2010, 107(13): 5881–5886.
- [130] Blazewicz SJ, Barnard RL, Daly RA, Firestone MK. Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. *The ISME Journal*, 2013, 7(11): 2061–2068.
- [131] Gremion F, Chatzinotas A, Harms H. Comparative 16S rDNA and 16S rRNA sequence analysis indicates that *Actinobacteria* might be a dominant part of the metabolically active bacteria in heavy metal-contaminated bulk and rhizosphere soil. *Environmental Microbiology*, 2003, 5(10): 896–907.
- [132] Campbell BJ, Yu LY, Heidelberg JF, Kirchman DL. Activity of abundant and rare bacteria in a coastal ocean. *Proceedings of the National Academy of Sciences of the United States of America*, 2011, 108(31): 12776–12781.
- [133] Mäki A, Salmi P, Mikkonen A, Kremp A, Tiirola M. Sample preservation, DNA or RNA extraction and data analysis for high-throughput phytoplankton community sequencing. *Frontiers in Microbiology*, 2017, 8: 1848.
- [134] Botero LM, D’Imperio S, Burr M, McDermott TR, Young M, Hassett DJ. Poly(A) polymerase modification and reverse transcriptase PCR amplification of environmental RNA. *Applied and Environmental Microbiology*, 2005, 71(3): 1267–1275.
- [135] Karst SM, Dueholm MS, McIlroy SJ, Kirkegaard RH, Nielsen PH, Albertsen M. Retrieval of a million high-quality, full-length microbial 16S and 18S rRNA gene sequences without primer bias. *Nature Biotechnology*, 2018, 36(2): 190–195.
- [136] Mäki A, Tiirola M. Directional high-throughput sequencing of

- RNAs without gene-specific primers. *Biotechniques*, 2018, 65(4): 219–223.
- [137] Prazsák I, Moldován N, Balázs Z, Tombác D, Megyeri K, Szűcs A, Csabai Z, Boldogkői Z. Long-read sequencing uncovers a complex transcriptome topology in varicella zoster virus. *BMC Genomics*, 2018, 19(1): 873.
- [138] Smith AM, Jain M, Mulroney L, Garalde DR, Akeson M. Reading canonical and modified nucleobases in 16S ribosomal RNA using nanopore native RNA sequencing. *PLoS One*, 2019, 14(5): e0216709.
- [139] Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S, Wu DY, Eisen JA, Hoffman JM, Remington K, Beeson K, Tran B, Smith H, Baden-Tillson H, Stewart C, Thorpe J, Freeman J, Andrews-Pfannkoch C, Venter JE, Li K, Kravitz S, Heidelberg JF, Utterback T, Rogers YH, Falcón LI, Souza V, Bonilla-Rosso G, Eguiarte LE, Karl DM, Sathyendranath S, Platt T, Bermingham E, Gallardo V, Tamayo-Castillo G, Ferrari MR, Strausberg RL, Neelson K, Friedman R, Frazier M, Venter JC. The *Sorcerer II* global ocean sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biology*, 2007, 5(3): e77.
- [140] Gruber-Vodicka HR, Seah BKB, Pruesse E. PhyloFlash—rapid SSU rRNA profiling and targeted assembly from metagenomes. *bioRxiv*, 2019: 521922.
- [141] Fan L, McElroy K, Thomas T. Reconstruction of ribosomal RNA genes from metagenomic data. *PLoS One*, 2012, 7(6): e39948.
- [142] Thomas T, Moitinho-Silva L, Lurgi M, Björk JR, Easson C, Astudillo-García C, Olson JB, Erwin PM, López-Legentil S, Luter H, Chaves-Fonnegra A, Costa R, Schupp PJ, Steindler L, Erpenbeck D, Gilbert J, Knight R, Ackermann G, Lopez JV, Taylor MW, Thacker RW, Montoya JM, Hentschel U, Webster NS. Diversity, structure and convergent evolution of the global sponge microbiome. *Nature Communications*, 2016, 7(1): 11870.
- [143] Goberna M, Verdú M. Predicting microbial traits with phylogenies. *The ISME Journal*, 2016, 10(4): 959–967.
- [144] Langille MGI, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepille DE, Thurber RLV, Knight R, Beiko RG, Huttenhower C. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnology*, 2013, 31(9): 814–821.
- [145] Abhauer KP, Wemheuer B, Daniel R, Meinicke P. Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data. *Bioinformatics*, 2015, 31(17): 2882–2884.
- [146] Eng A, Borenstein E. Taxa-function robustness in microbial communities. *Microbiome*, 2018, 6(1): 45.
- [147] Louca S, Parfrey LW, Doebeli M. Decoupling function and taxonomy in the global ocean microbiome. *Science*, 2016, 353(6305): 1272–1277.
- [148] Qin HY, Wang S, Feng K, He ZL, Virta MPJ, Hou WG, Dong HL, Deng Y. Unraveling the diversity of sedimentary sulfate-reducing prokaryotes (SRP) across Tibetan saline lakes using epicPCR. *Microbiome*, 2019, 7(1): 71.
- [149] Louca S, Mazel F, Doebeli M, Parfrey LW. A census-based estimate of Earth's bacterial and archaeal diversity. *PLoS Biology*, 2019, 17(2): e3000106.
- [150] Zhang WP, Ding W, Li YX, Tam C, Bougouffa S, Wang RJ, Pei B, Chiang H, Leung P, Lu YH, Sun J, Fu H, Bajic VB, Liu HB, Webster NS, Qian PY. Marine biofilms constitute a bank of hidden microbial diversity and functional potential. *Nature Communications*, 2019, 10(1): 517.
- [151] Fan L, Reynolds D, Liu M, Stark M, Kjelleberg S, Webster NS, Thomas T. Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. *Proceedings of the National Academy of Sciences of the United States of America*, 2012, 109(27): 10765–10766.
- [152] Canonico G, Buttigieg PL, Montes E, Muller-Karger FE, Stepien C, Wright D, Benson A, Helmuth B, Costello M, Sousa-Pinto I, Saeedi H, Newton J, Appeltans W, Bednaršek N, Bodrossy L, Best BD, Brandt A, Goodwin KD, Iken K, Marques AC, Miloslavich P, Ostrowski M, Turner W, Achterberg EP, Barry T, Defeo O, Bigatti G, Henry LA, Ramiro-Sánchez B, Durán P, Morato T, Roberts JM, García-Alegre A, Cuadrado MS, Murton B. Global observational needs and resources for marine biodiversity. *Frontiers in Marine Science*, 2019, 6: 367.
- [153] Fitz-Gibbon S, Tomida S, Chiu BH, Nguyen L, Du C, Liu M, Elashoff D, Erfe MC, Loncaric A, Kim J, Modlin RL, Miller JF, Sodergren E, Craft N, Weinstock GM, Li HY. *Propionibacterium acnes* strain populations in the human skin microbiome associated with acne. *Journal of Investigative Dermatology*, 2013, 133(9): 2152–2160.

# Advances in ribosomal RNA sequencing technologies for studying marine microbial diversity

Yuhang Liu<sup>1,2</sup>, Songze Chen<sup>3</sup>, Chuanlun Zhang<sup>1</sup>, Lu Fan<sup>1,4\*</sup>

<sup>1</sup> Shenzhen Key Laboratory of Marine Archaea Geo-Omics, Department of Ocean Science and Engineering, Southern University of Science and Technology (SUSTech), Shenzhen 518055, Guangdong Province, China

<sup>2</sup> School of Environment, Harbin Institute of Technology, Harbin 150001, Heilongjiang Province, China

<sup>3</sup> State Key Laboratory of Marine Geology, Tongji University, Shanghai 200092, China

<sup>4</sup> Academy for Advanced Interdisciplinary Studies, Southern University of Science and Technology (SUSTech), Shenzhen 518055, Guangdong Province, China

**Abstract:** The total number of marine microorganisms is estimated to be  $10^{30}$ . They are the key components driving marine energy and material recycling and maintaining ecological balance. Research advances in marine microbial diversity contribute to our understanding of the operative mechanisms of marine ecosystems. These knowledges can guide our response to ecological crisis of global oceans and our exploration of marine microbial resources. As the number of culturable marine microorganisms is currently low, our knowledges to the abundance, physiological characteristics and ecological functions of marine microbial communities are greatly limited. Ribosomal RNA sequencing technologies can rapidly and accurately identify or classify microorganisms based on nucleic acid sequences at low-cost. These technologies have been used in studying community structure and evolutionary and ecological relationships between microorganisms. In recent years, with the benefit of rapid development in sequencing technologies, marine microbial studies based ribosomal RNA sequencing analyses have achieved remarkable progresses in discovering novel marine microbial lineages, revealing ecological mechanisms and evolutionary relationships between marine microorganisms, discovering novel metabolites and applying in marine bioremediation. In this review, we introduce the technical principles of ribosomal RNA sequencing, the application of three generations of sequencing technologies in studying marine microbial diversity, and variable combinations of library construction and sequencing technologies. Finally, we address prospectives in utilizing these sequencing strategies in studies with different purposes.

**Keywords:** marine microbial diversity, ribosomal RNA, sequencing technology

(本文责编: 李磊)

Supported by the National Natural Science Foundation of China (91951120, 91851210), by the Natural Scientific Foundation of Guangdong (2018B030311016) and by the Shenzhen Science and Technology Innovation Commission (JCYJ20180305123458107, ZDSYS201802081843490)

\*Corresponding author. Tel/Fax: +86-755-88018620; E-mail: fanl@sustech.edu.cn

Received: 28 March 2020; Revised: 10 May 2020; Published online: 11 July 2020

**范陆**, 博士, 南方科技大学海洋学院研究助理教授。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, 是微生物生态与进化理论研究领域的青年科学家。主要研究领域包括微生物生态与进化、微生物共生、微生物组与病毒组互作、计算微生物学等。曾担任广州医科大学和昆士兰大学客座研究员。曾以第一或通讯作者, 在 *Nature Ecology & Evolution*、*PNAS*、*ISME J*、*EST* 等国际一流杂志上发表研究成果。已主持国家自然科学基金重大研究计划培育项目 1 项, 深圳市基础研究 1 项。曾获 2011 年欧洲微生物学会 (FEMS) 青年科学家奖, 2014 年澳大利亚新南威尔士州海洋科学最佳研究奖。

