

## 研究报告

基于 *amoA* 基因扩增子高通量测序的氨氧化古菌多样性  
分析方法何翔<sup>1</sup> 吴佳鹏<sup>2</sup> 焦黎静<sup>2</sup> 温晓梅<sup>3</sup> 王岩<sup>2</sup> 欧林坚<sup>1</sup> 洪义国<sup>3\*</sup>

(1. 暨南大学赤潮与海洋生物研究中心 广东 广州 510632)

(2. 中国科学院南海海洋研究所 广东 广州 510301)

(3. 广州大学环境科学与工程学院 广东 广州 510006)

**摘要:**【背景】对于环境样品中氨氧化古菌(Ammonia-oxidizing archaea, AOA)多样性的研究, 利用 *amoA* 功能基因作为分子标记会比 16S rRNA 基因有更强的特异性和更高的分辨率, 能更准确地反映环境样品中氨氧化古菌的种群结构和分布特征。然而, 目前对 *amoA* 基因扩增子高通量测序的分析存在两大限制因素: 一是缺乏相应的 *amoA* 基因参考数据库; 二是 AOA *amoA* 基因在种水平上的相似性阈值未知, 分析过程中没有明确的划分种水平操作分类单元(Operational taxonomic unit, OTU)的阈值。【目的】构建基于 *amoA* 功能基因序列分析氨氧化古菌多样性的方法, 为基于高通量测序的功能微生物多样性分析提供参考。【方法】基于目前已通过分离纯化或富集培养获得的 34 株氨氧化古菌及功能基因数据库中收录的环境样品 *amoA* 基因序列, 构建氨氧化古菌 *amoA* 基因参考数据库。通过菌株间两两比对获得的 *amoA* 基因相似度与 16S rRNA 基因相似度的相关性分析, 确定 *amoA* 基因在种水平上的相似性阈值。基于 MOTHUR 软件平台, 利用建立的参考数据库和确定的阈值对南海一个垂直水体剖面样品的 *amoA* 基因序列进行多样性分析。【结果】构建了含有 26 091 条序列信息的古菌 *amoA* 基因参考数据库, 确定了 89%作为分析过程中古菌 *amoA* 基因划分种水平 OTU 的阈值, 对南海水体样品氨氧化古菌的多样性分析结果很好地显示了南海不同深度水层水体中氨氧化古菌的种群结构和系统发育关系, 有效揭示了南海氨氧化古菌的垂直分布差异。【结论】建立了基于 *amoA* 基因高通量测序的氨氧化古菌多样性分析方法, 此方法可以有效分析环境样品中氨氧化古菌的多样性。

**关键词:** 功能基因, 参考数据库, 相似性阈值, *amoA* 基因, 氨氧化古菌

**Foundation items:** National Natural Science Foundation of China (41576123); National Key Basic Research Program of China (2015CB452902); Scientific Research Project of Guangzhou City (201504010005); Research Foundation for Talented Scholars of Guangzhou University (GZHU-B-05)

**\*Corresponding author:** Tel: 86-20-39158007; E-mail: yghong@gzhu.edu.cn

**Received:** March 09, 2018; **Accepted:** June 14, 2018; **Published online** (www.cnki.net): June 20, 2018

**基金项目:** 国家自然科学基金(41576123); 国家重点基础研究发展计划(2015CB452902); 广州市科学研究专项重点项目(201504010005); 广州大学百人计划人才引进启动基金(GZHU-B-05)

**\*通信作者:** Tel: 86-20-39158007; E-mail: yghong@gzhu.edu.cn

**收稿日期:** 2018-03-09; **接受日期:** 2018-06-14; **网络首发日期**(www.cnki.net): 2018-06-20

## Development of a method for ammonia-oxidizing archaea diversity analysis based on *amoA* gene amplicons with high-throughput sequencing

HE Xiang<sup>1</sup> WU Jia-Peng<sup>2</sup> JIAO Li-Jing<sup>2</sup> WEN Xiao-Mei<sup>3</sup> WANG Yan<sup>2</sup>  
OU Lin-Jian<sup>1</sup> HONG Yi-Guo<sup>3\*</sup>

(1. Research Center for Harmful Algae and Marine Biology, Jinan University, Guangzhou, Guangdong 510632, China)

(2. South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou, Guangdong 510301, China)

(3. School of Environmental Science and Engineering, Guangzhou University, Guangzhou, Guangdong 510006, China)

**Abstract:** [Background] Using *amoA* gene as marker for analyzing the diversity of ammonia-oxidizing archaea (AOA) has stronger specificity and higher resolution than 16S rRNA gene, it can more accurately reflect the community structure and distribution pattern of AOA in environmental samples. However, there are two limitations in the analysis of *amoA* gene sequence with high-throughput sequencing: one is the lack of corresponding *amoA* gene reference taxonomic database; and the other is no determined species-level cut-off value for operational taxonomic units (OTUs) clustering. [Objective] The aim of this article was to develop a method for ammonia-oxidizing archaea diversity analysis based on *amoA* gene sequence with high-throughput sequencing, providing a reference method to analyze the functional microbial diversity based on high-throughput sequencing. [Methods] *amoA* gene sequences of 34 AOA strains from pure or enrichment culture were used as seed sequences. Uncultured *amoA* gene sequences from environmental samples were downloaded from functional gene database. All sequences were used to construct taxonomic database for *amoA* gene sequences. By pairwise comparison of 16S rRNA gene and archaeal *amoA* identities of all recognized species of AOA, we determined the cut-off value for OTU clustering at species-level. We used the established reference database and determined cut-off value to analyze the diversity for water samples in a vertical profile of the South China Sea with the MOTHUR software. [Results] We constructed a reference taxonomic database containing 26 091 *amoA* gene sequences, and determined 89% sequence identity as cut-off value for OTU clustering at species-level. The diversity analysis of AOA sufficiently showed the community structure and phylogenetic relationship, and effectively revealed the AOA vertical distribution differences in the South China Sea. [Conclusion] We developed a method for ammonia-oxidizing archaea diversity analysis based on *amoA* gene sequence with high-throughput sequencing. This method can effectively analyze the diversity of AOA in environmental samples.

**Keywords:** Functional gene, Taxonomic database, Cut-off value, *amoA* gene, AOA

氨氧化古菌 (Ammonia-oxidizing archaea, AOA) 是自然环境中一类具有将氨氧化成亚硝酸盐功能的古菌, 在氮素的生物地球化学循环中起着重要的作用<sup>[1]</sup>。2005 年, Könnike 等从西雅图热带海洋水族馆的海水中成功分离培养到第一株具有氨氧化功能的古菌 *Nitrosopumilus maritimus* SCM1, 开辟了微生物氨氧化作用研究的新领域, 自此氨氧化古菌作为新的氮循环微生物成为研究热点<sup>[2]</sup>。2008 年, 科学家们通过系统发生学和酶学分析,

将包括氨氧化古菌在内的部分泉古菌类群重新划分成一个新门——奇古菌门 (Thaumarchaeota)<sup>[3-4]</sup>。*amoA* 基因是编码氨氧化过程中的关键酶——氨单加氧酶 (Ammonia monooxygenase, AMO)  $\alpha$  亚基的基因, 对 AOA 的系统发育分析与 16S rRNA 基因有着高度的一致性<sup>[5]</sup>, 与氨氧化细菌 (Ammonia-oxidizing bacteria, AOB) 基因组中 *amoA* 基因存在多拷贝不同<sup>[6]</sup>, 目前已发现的 AOA 基因组中 *amoA* 基因均为单拷贝。目前, *amoA* 基因作为分子标记

物广泛地用于海洋、河口、沉积物、土壤等生境中氨氧化微生物的生态学研究<sup>[7-9]</sup>。相较于保守的 16S rRNA 基因, *amoA* 基因序列差异程度更大, 作为分子标记分析氨氧化古菌多样性分辨率更高, 可以更好地区分 16S rRNA 基因相似性较高的不同种。此外, 由于 AOA *amoA* 基因为单拷贝, 并不会出现类似 16S rRNA 基因由于多拷贝和基因组内部异质性导致的对多样性分析的偏差<sup>[10]</sup>。但 *amoA* 基因作为分子标记分析氨氧化古菌多样性的过程中存在两大难点: (1) 目前没有基于 *amoA* 基因序列构建的参考数据库(Taxonomic database); (2) 进行操作分类单元(Operational taxonomic unit, OTU)聚类分析时, 由于 AOA *amoA* 基因与 16S rRNA 基因在进化速率上存在差异, AOA *amoA* 基因在种水平上的相似性阈值并不清楚, *amoA* 基因并不能类似于 16S rRNA 基因, 将 97% 的序列相似度作为划分种水平 OTU (Species-level OTU) 的阈值<sup>[11]</sup>, 即对 *amoA* 基因测序结果聚类分析过程中没有明确地划分种水平 OTU 的阈值。

本文以南海氨氧化古菌的多样性分析为例, 以 *amoA* 功能基因作为分子标记, 基于高通量测序技术, 利用 MOTHUR 软件<sup>[12]</sup> 进行数据分析, 建立基于 *amoA* 基因高通量测序分析环境中氨氧化古菌多样性的研究方法。

## 1 材料与方法

### 1.1 样品采集

选取南海 S28 站位(北纬 18°00'00", 东经 113°30'00"), 通过船载 CTD (Conductivity-Temperature-Depth sensors) 上配备的尼斯金采样瓶(Niskin bottle)采集不同深度水层水体(75、100、150、200、500、800、1 000 m), 量取 2 L 海水样品经 0.22  $\mu\text{m}$  聚碳酸酯滤膜(Millipore, USA)进行过滤, 收集海洋水体中的微生物。将过滤后的滤膜装入冻存管中, 置于液氮中冷冻保存。

### 1.2 主要试剂和仪器

PowerWater<sup>®</sup> DNA Isolation Kit, Qiagen 公司; GoTaq<sup>®</sup> Colorless Master Mix, Promega 公司;

TaKaRa MiniBEST Agarose Gel DNA Extraction Kit, 宝生物工程(大连)有限公司; PCR 引物由 Invitrogen 公司合成。PCR 仪, ABI 公司; NanoDrop LITE 核酸定量仪, Thermo 公司; Mupid<sup>®</sup>-One 电泳仪, Advance 公司; 凝胶成像系统, Bio-Rad 公司。

### 1.3 样品总 DNA 提取及 *amoA* 基因扩增

按照强力水样 DNA 提取试剂盒操作说明书提取不同深度水层样本滤膜上的微生物宏基因组 DNA, 使用特异性引物 Arch-*amoA*F (5'-STAATGG TCTGGCTTAGACG-3') 和 Arch-*amoA*R (5'-GCGG CCATCCATCTGTATGT-3') 扩增 *amoA* 基因, 其中上游引物前端通过加 8 碱基的 Barcode 标记进行修饰, 使用带有不同 Barcode 标记的上游引物扩增不同样品, PCR 反应体系及反应条件参照 Francis 等<sup>[13]</sup>。利用凝胶成像系统的定量工具初步测定 PCR 产物的相对丰度, 根据各样本目标测序深度对各样本 PCR 产物进行相应比例混合, 利用胶回收试剂盒对混合后的 PCR 产物进行切胶纯化回收。最后, 将纯化后的 PCR 产物-80 °C 保存, 委托苏州金唯智生物科技有限公司进行高通量测序。

### 1.4 *amoA* 基因参考数据库的构建

从 GenBank 数据库(<https://www.ncbi.nlm.nih.gov/genbank/>)中下载目前已通过分离或富集培养获得菌株的 *amoA* 基因序列作为种子序列, 同时下载 FunGene 数据库<sup>[14]</sup> (<http://fungene.cme.msu.edu/>) 中收集的环境样品 *amoA* 基因序列, 将获得的所有序列进行整合, 用 MOTHUR 软件([https://mothur.org/wiki/MiSeq\\_SOP](https://mothur.org/wiki/MiSeq_SOP)) 对序列进行筛选: (1) 以种子序列作为参比序列进行序列对齐; (2) 对环境样品序列进行去冗余, 提取非冗余序列; (3) 对序列进行长度控制, 删除长度小于 555 个碱基的序列; (4) 对剩余序列进行嵌合体检测, 去除含有嵌合体片段的序列。利用筛选处理后的序列生成基因序列文件(Template file), 基于 *amoA* 基因的序列号从 NCBI 数据库中搜集各序列对应的种属信息, 生成种属信息文件(Taxonomic

file)。基因序列文件和对应的种属信息文件构成功能基因的参考数据库<sup>[15]</sup>。

### 1.5 *amoA* 基因种水平相似性阈值的确定

收集 NCBI 数据库 (<https://www.ncbi.nlm.nih.gov/>) 中 *amoA* 基因以及 16S rRNA 基因序列信息均完整的氨氧化古菌菌株, 利用 MEGA 7.0 分别计算菌株间 *amoA* 基因以及 16S rRNA 基因的相似度。以 16S rRNA 基因相似度作为 *X* 轴, 功能基因 *amoA* 基因相似度作为 *Y* 轴, 用 GraphPad Prism7.00 进行散点图绘制, 通过回归分析计算菌株间 2 个基因相似度的相关关系。基于回归方程得出 97% 的 16S rRNA 基因相似度所对应的 *amoA* 基因相似度, 从而得到 *amoA* 基因在种水平上的相似性阈值 (Species-level cut-off value)。

### 1.6 高通量测序数据分析

利用 MOTHUR (V.1.39.5) 软件对原始序列进行分析处理, 包括序列质量控制和相关生物信息学分析。对初始序列进行质控过滤, 通过长度筛选以及嵌合体检测, 去掉过短序列和嵌合体序列 (Chimera sequence), 序列对齐后, 去除 Barcode 和引物序列, 得到有效序列用于进一步分析。通过与参考数据库进行比对, 对有效序列进行物种注释, 绘制种群结构组分图 (属水平), 分析南海不同深度水层中氨氧化古菌的种属组成; 以 2.5 中确定的相似性阈值进行 OTU 聚类分析, 将相似度处于阈值范围内的序列归为同一个 OTU; 去除稀有 OTU (序列数 < 总序列数的 0.1%), 绘制稀释曲线 (Rarefaction curve) 进行取样深度验证, 评价测序量是否足以覆盖样品中的所有类群; 选取主要 OTU (序列数 > 总序列数的 1%) 的代表序列构建系统发育树, 分析样品中优势物种进化关系及分类地位, 同时基于主要 OTU 的丰度信息绘制热图 (Heatmap), 反映不同样本间物种分布的相似性与差异性; 基于非稀有 OTU (序列数 > 总序列数的 0.1%) 在每个样品中丰度分布状况 (格式为 Shared 的文件), 计算样品中的 Chao1 指数<sup>[16]</sup>、Shannon 指数<sup>[17]</sup>等多样性指数, 从而分析样品中的  $\alpha$  多样

性 (Alpha-diversity); 基于主坐标分析 (Principal coordinates analysis, PCoA) 计算样品间的多样性差异, 进行  $\beta$  多样性 (Beta-diversity) 分析。

### 1.7 获取序列登录号

将得到的 AOA *amoA* 基因高通量测序结果提交至 NCBI SRA (Sequence read archive) 数据库获取登录号。

## 2 结果与分析

### 2.1 *amoA* 基因参考数据库

构建的序列种子库中, 共收集了迄今通过分离或富集培养获得的 34 株氨氧化古菌的 *amoA* 基因序列信息, 基于种子序列构建的系统发育树见图 1。可以看出, 目前已知的氨氧化古菌可分为 4 大簇, 包括海洋簇 (Group I.1a)、土壤簇 (Group I.1b)、嗜酸簇 (Group I.1a-associated)、嗜热簇 (THAOA)。从 FunGene 数据库中共下载到 47 792 条环境样品序列。经筛选处理并整合后, 得到 AOA *amoA* 基因序列文件 (Template file)。表 1 简要总结了参考数据库中序列的质量信息, 基因序列文件中共收集了 26 091 条 *amoA* 基因序列, 所有序列中均不含模糊碱基, 但含有 3–5 个碱基长度的同聚物, 序列长度控制在 555–635 个碱基的范围。基因序列文件与种属信息文件均由 2 个纵列组成, 基因序列文件第 1 列为序列号, 第 2 列为碱基序列; 种属信息文件第 1 列与基因序列文件的第 1 列一致, 第 2 列为对应的分类信息, 其中不同分类级别用分号分隔。基因序列文件和种属信息文件构成 *amoA* 基因参考数据库。

### 2.2 *amoA* 基因在种水平上的相似性阈值

通过菌株间的两两比对, 得到成对菌株间 16S rRNA 基因相似度和 *amoA* 基因相似度 (Pairwise distances), 从回归分析结果 (图 2) 可以看出, 这两者之间有着很好的相关性, 线性拟合  $R^2$  达到 0.84, 而多项式拟合  $R^2$  达到 0.91。表明 16S rRNA 基因与 *amoA* 基因在进化过程中有很高的相关性, 但 16S rRNA 基因比 *amoA* 基因在进化速率上表现得更保守。选择  $R^2$  值更高的多项式拟合曲线作为 16S rRNA

基因与 *amoA* 基因相关关系的回归方程可以得到, 16S rRNA 基因 97% 的序列相似度对应的 *amoA* 功能基因的相似度为 89.4%。因此, 基于功能基因 *amoA* 序列对氨氧化古菌的多样性分析, 选取以 89% 的序列相似度作为在种水平上划分 OTU 的阈值。

## 2.3 基于 MOTHUR 平台分析南海深海水体样品中的氨氧化古菌多样性

### 2.3.1 序列登录号

AOA *amoA* 基因高通量测序结果在 NCBI SRA 数据库获取的登录号为 SRP131845。

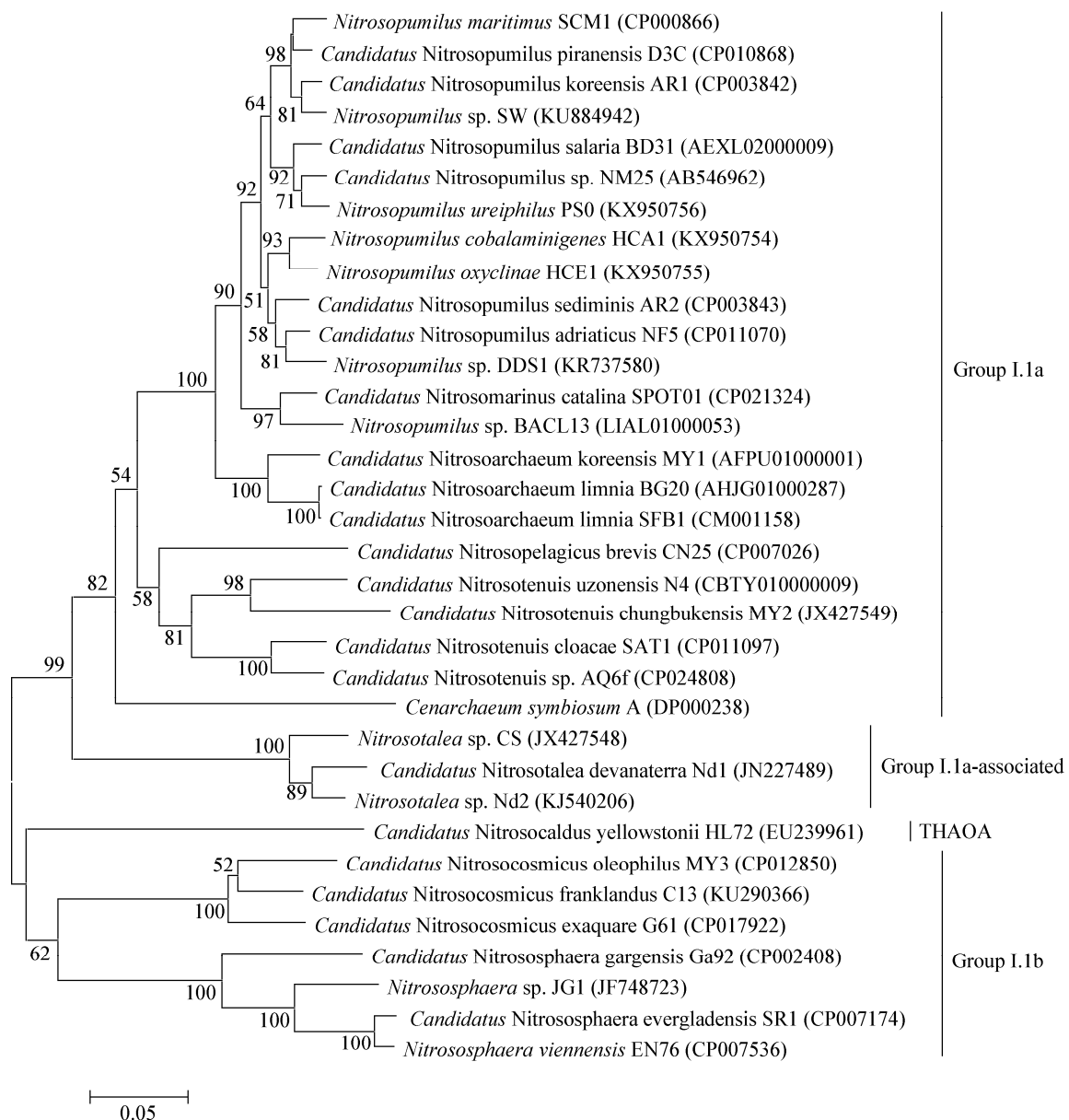


图 1 已知的氨氧化古菌 *amoA* 基因系统发育分析

Figure 1 Phylogenetic analysis of recognized species of archaeal *amoA* genes

注: 基于邻接法, 节点上的数值为自展值(Bootstrap), 低于 50% 的数值(1 000 次重复)没有显示, 序列选自 NCBI 数据库, 括号内为含有 *amoA* 基因的序列登录号。

Note: Based on neighbor-joining method, numbers at the node are the bootstrap values (%), bootstrap values (1 000 replicates) less than 50% are not shown. Sequences are obtained from NCBI database, accession numbers of sequences of *amoA* gene are shown in parentheses.

表 1 AOA *amoA* 基因参考数据库序列基本信息

Table 1 Summary of the quality of sequences in the reference template database

项目 Items	起始 Start	末端 End	碱基数 NBases	模糊碱基 Ambigs	同聚物 Polymer	序列数 Sequences number
Minimum	1	564	555	0	3	1
2.5%-tile	4	615	594	0	3	653
25%-tile	4	615	597	0	4	6 523
Median	4	633	630	0	4	13 046
75%-tile	22	633	630	0	4	19 569
97.5%-tile	22	635	632	0	5	25 439
Maximum	37	635	635	0	9	26 091
Mean	9.288 38	627.591	619.302	0	4.013 91	

### 2.3.2 南海水体样品中的氨氧化古菌属水平分布分析

运行 MOTHUR 软件中的 `classify.seqs` 语句, 基于构建的参考数据库对质量控制后的有效序列进行分类学注释。利用运行结束后生成的分类结果统计文件, 计算氨氧化古菌属水平下的相对丰度分布情况, 图 3 为基于 OriginPro 2017 绘制的不同深度水层 AOA 在属水平上的相对丰度分布图。

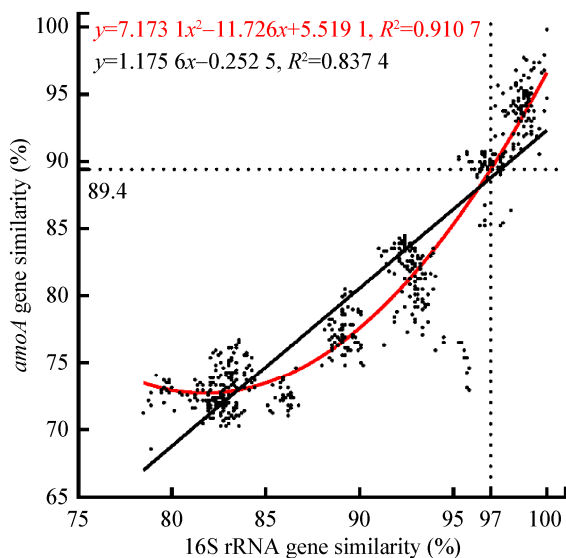


图 2 *amoA* 基因相似度与 16S rRNA 基因序列相似度相关性分析

Figure 2 Correlation of *amoA* gene similarity versus 16S rRNA gene similarity

注: 黑色直线为线性拟合, 红色曲线为二项式拟合。

Note: Black line represents linear fitting, red curve represents binomial fitting.

从图 3 可以看出, 随着深度变化, 南海水体中的 AOA 菌群分布有明显的变化趋势: 南海 75 m 和 100 m 的浅水层中的 AOA 主要属于 *Nitrosopelagicus* 属; 500、800 和 1 000 m 的深水层样品序列匹配到的参考序列主要为未培养(Uncultured)的环境样品序列, 未明确分类地位菌属主要分布在深层水体; 而 150 m 和 200 m 中层水体的氨氧化古菌种属分布表现出一个明显的过渡段。此外, 测序产生的所有序列均来自于奇古菌门的 *amoA* 基因, 说明本研究选择的扩增引物对 AOA 的 *amoA* 基因具有很强的特异性。而分析结果中仅有很小一部分序列(3.7%)被划分到了 Unclassified, 即仅有 3.7% 的序列无法从参考数据库中找到匹配序列, 说明构建的参考数据库收集的序列信息较为完整, 库容量足够支撑环境中氨氧化古菌的多样性分析。

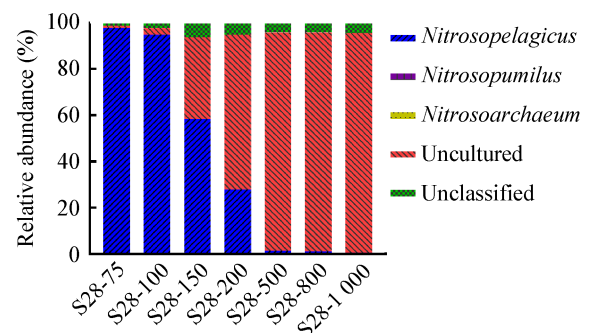


图 3 南海不同深度水层氨氧化古菌的属水平分布

Figure 3 Relative abundance of the AOA at genus level in different depth in the South China Sea

### 2.3.3 南海水体样品中氨氧化古菌的系统发育分析

利用 MOTHUR 平台上的 *dist.seqs* 命令语句计算有效序列间的相似度, 生成距离矩阵, 将相似度 $\geq 89\%$ 的序列归为一个 OTU。挑取主要 OTU 的代表序列, 得到了 8 个主要 OTU 的代表序列, 利用 MEGA 7.0 进行系统进化树的构建(图 4), 分析主要 OTU 间的系统进化关系; 基于 OTU 在每个样品的丰度分布及进化关系进行 OTU 热图(Heatmap)绘制(图 5), 通过颜色梯度变化及相似程度来反映不同深度水层中 AOA 种属组成的相似性和差异性。系统发育分析结果表明, 所有样品序列均属于奇古菌中的 Group I.1a 部分, 其中丰度最高的 OTU01 与从海洋中通过富集培养得到的 *Candidatus Nitrosopelagicus brevis* CN25<sup>[18]</sup>有很近的亲缘关系, 序列经两两比对(BLAST)后相似度为 99%。结合图 4 和图 5 来看, 样品序列被划分为两大簇,

其中包含有 OTU01、OTU04、OTU07 的一簇主要分布在表层及中层(75–200 m), 在底层水体中丰度极低; 而另一簇(OTU02、OTU03、OTU05、OTU06、OTU08)主要分布在中层和底层(150–1 000 m), 在表层水体中几乎没有分布。氨氧化古菌在南海海洋水体中表现出很高的垂直分布差异。

### 2.3.4 取样深度验证

通过随机抽样的方法提取数据, 以抽取的序列数作为横轴, 以这些序列所包含的 OTU 数目作为纵轴, 绘制稀释曲线(Rarefaction curve)。稀释曲线反映了样品的取样深度, 可以用来评价测序量是否足以覆盖所有类群。南海不同深度水层样品的稀释曲线如图 6 所示, 可以看出稀释曲线已趋于平缓, 表明当前的测序量已经能够很好地反映出样品中 AOA 的群落结构和多样性, 测序数据量合理, 再增大数据量对发现新 OTU 的影响很小。

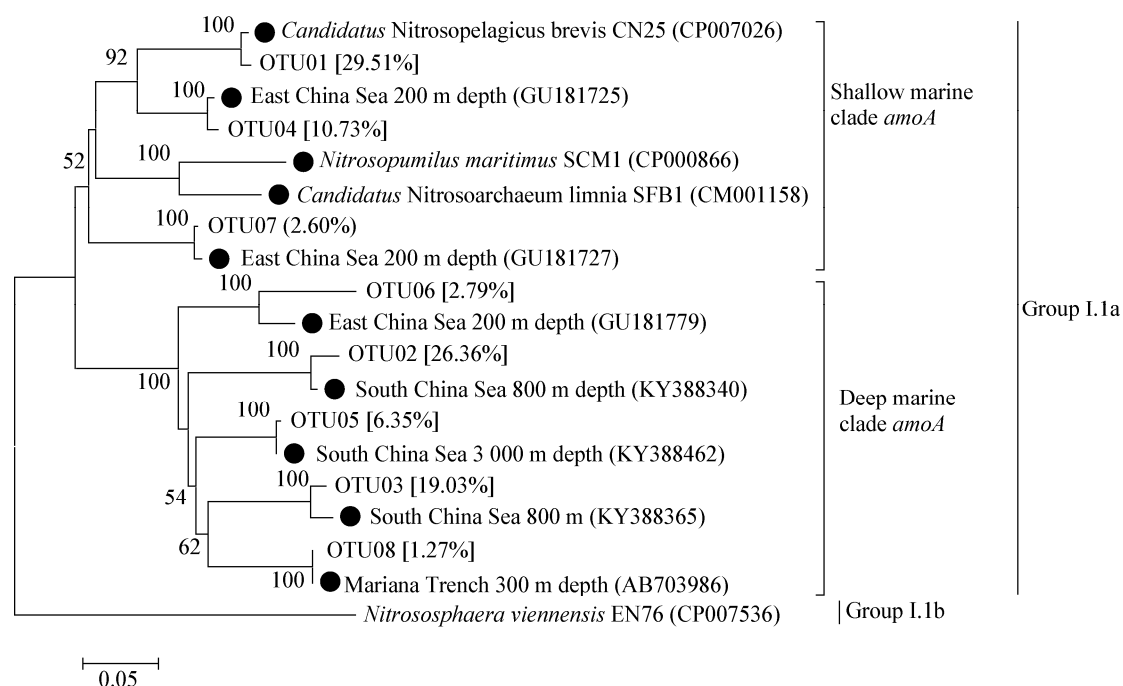


图 4 基于主要 OTU 构建的系统进化树

Figure 4 Phylogenetic tree of dominant OTUs in the South China Sea

注: 基于邻接法, Bootstrap 值选自 1 000 重复, 黑点标志为参考序列, 参考序列选自 NCBI 数据库, 圆括号里为参考序列的登录号, 方括号内为各主要 OTU 序列数占总序列数的比例。

Note: Based on neighbor-joining method, numbers on branches are the supporting percentage by 1 000 replicates. Reference sequences obtained from NCBI database are marked by black circle, accession number of reference sequences are shown in parenthesis, the percentage of each dominant OTU in total sequences are shown in bracket.

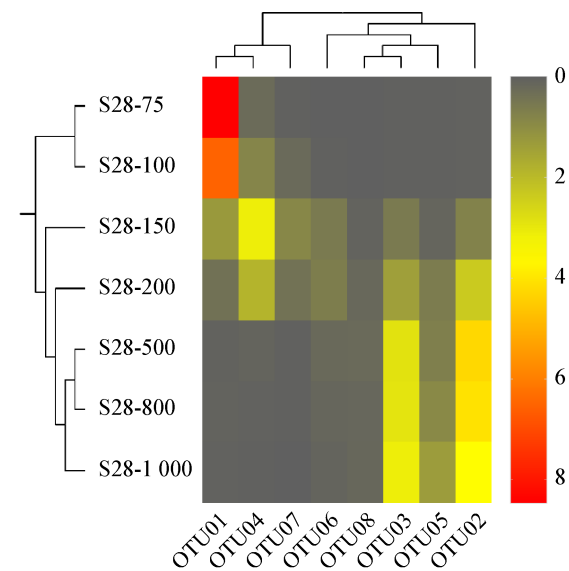


图 5 基于主要 OTU 绘制的热图  
**Figure 5 Heat map of dominant OTUs in the South China Sea**  
注：热图左侧为样品名称和样本聚类树，上面和下面分别为物种聚类树和主要 OTU 名称。  
Note: The left side of the heat map are the sample name and the sample clustering tree, the upper and the bottom side are the species clustering tree and dominant OTU name, respectively.

2.3.5  $\alpha$  多样性分析

利用 Mothur 软件对南海不同深度水层样品进行多样性指数(Alpha-diversity)统计分析(表 2)，测序结果覆盖率均在 99.55%以上，表明数据有效可靠。从整体上来看，150 m 和 200 m 的样品中 OTU

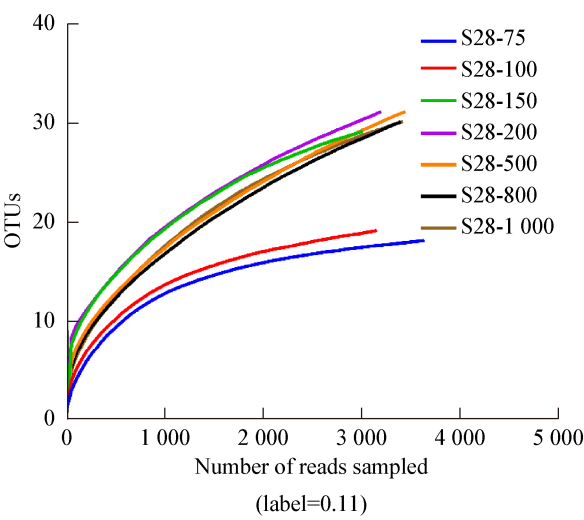


图 6 操作分类单元稀释曲线  
**Figure 6 Rarefaction curves for OTUs**

数目最多，Shannon 指数最高，Simpson 指数最低；而 75 m 和 100 m 的样品中 OTU 数目最少，Shannon 指数最低，Simpson 指数最高。这表明南海氨氧化古菌在表层水体中的物种丰富度、种群多样性以及物种组成均匀度最低，而在中层水体中表现最高，深层水体次之。

2.3.6  $\beta$  多样性分析

基于 OTU 聚类结果，利用 MOTHUR 软件中的 thetayc 算法计算样品间的差异，运行 pcoa 语句进行主坐标分析(Principal coordinates analysis)，结果见图 7。由图 7 可知，主坐标分析其第 1 主成

表 2 不同深度水层样品的 Alpha 多样性指数  
**Table 2 The alpha diversity index of samples in different depth**

样本 Sample	覆盖率 Coverage (%)	OTUs	Chao1 指数 Chao1 index	Shannon 指数 Shannon index	Simpson 指数 Simpson index
S28-75	99.88	18	19.828	0.277	0.908
S28-100	99.83	19	23.562	0.611	0.730
S28-150	99.70	29	35.000	1.742	0.245
S28-200	99.58	31	53.536	1.894	0.192
S28-500	99.55	31	46.013	1.257	0.378
S28-800	99.56	30	45.751	1.232	0.373
S28-1 000	99.64	30	38.287	1.266	0.344

注：序列相似性在 89%以上为 1 个 OTU。  
Note: The operational taxonomic unit (OTU) was defined with 89% similarity.



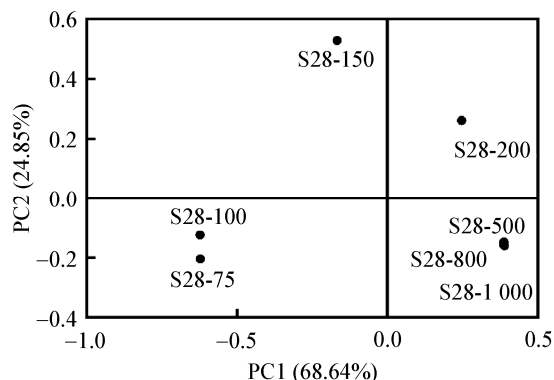


图 7 南海不同深度水层样品中氨氧化古菌种群结构主坐标分析

Figure 7 Principal coordinates analysis for AOA community in different depth in the South China Sea

分和第 2 主成分的贡献率分别为 68.64% 和 24.85%，其中，75 m 和 100 m 距离较近，而代表 500、800 和 1 000 m 样品的点很好地聚集在一起，代表样品的点按深度聚集成 3 簇。由此可以得出，南海氨氧化古菌在浅层(75、100 m)、中层(150、200 m)、深层(500、800、1 000 m)的海洋水体中分布差异明显，氨氧化古菌在南海水体的分布呈现垂直分层现象。

### 3 讨论与结论

基于 MOTHUR 软件平台，利用构建的 *amoA* 基因参考数据库以及将 89% 的相似度作为在种水平上划分 OTU 的阈值，可以有效分析环境样品中的氨氧化古菌多样性。对南海不同深度水层水体样品中 *amoA* 基因高通量测序结果的分析表明，本文建立的分析方法可以有效揭示南海水体中氨氧化古菌的垂直分布差异。

本研究构建了含有 26 091 条序列的参考数据库，其中大部分 *amoA* 基因序列来源于未培养微生物，仅是通过分子生物学的手段从环境样品中获得，其对应的种属信息和分类地位都不明确。因此，已鉴定物种过少以及分类信息不完整仍是目前针对环境样品中氨氧化古菌多样性进行分析的制约因素。2012 年 Biller 等分析了 NCBI 中来

自不同生境的 8 296 条 AOA *amoA* 基因序列，按照 85% 的相似度在种水平上进行物种划分时自然环境中 AOA 的物种数在 138 个左右，当按照 90% 的相似度进行物种划分时 AOA 的物种数为 315 个<sup>[19]</sup>。而由于氨氧化古菌生长缓慢，代时长，目前通过分离或富集培养得到并鉴定的 AOA 仅有二十几个物种，这就意味着自然界中将近 85%–95% 的 AOA 还未被人类所认识，因此对氨氧化古菌新物种的分离与鉴定仍然是目前微生物氨氧化研究的重点。

参照 Purkhold 等<sup>[20]</sup>确定的氨氧化细菌(AOB) *amoA* 基因在种水平上相似性阈值的方法，Pester 等<sup>[21]</sup>基于当时 16S rRNA 基因与 *amoA* 基因序列信息均完整的 10 株氨氧化古菌菌株，提出将 *amoA* 基因序列相似度大于 85% 的 AOA 定义为同一个物种。本研究中增加了近年新鉴定的氨氧化古菌菌株，利用两两比对后的 561 组数据重新计算了 AOA *amoA* 基因在种水平上的相似性阈值，将该值从 85% 提升至 89%。而随着分离鉴定的氨氧化古菌菌株的增加，数据量更加丰富，阈值的选取也将更加真实准确。基于目前已发现的氨氧化古菌的全基因组，并未发现 AOA 基因组中 16S rRNA 基因与 *amoA* 基因存在多拷贝的现象，因此本文对阈值的选取过程均是基于基因组中单拷贝的 16S rRNA 基因与单拷贝的 *amoA* 基因比对得出。但有研究发现，古菌基因组中 16S rRNA 基因存在多拷贝现象，而且多拷贝的基因序列并不完全一致，即 16S rRNA 基因在基因组内部存在异质性。16S rRNA 基因的多拷贝及异质性对基于 16S rRNA 的多样性分析会造成分析结果的偏差，会引起一定程度的高估<sup>[10]</sup>。但对于功能基因相似性阈值的选取，由于绝大多数原核微生物基因组中多拷贝的 16S rRNA 基因序列通常是相同或是相似度非常高的(序列差异小于 1%)<sup>[22]</sup>，而对 AOA *amoA* 基因的相似性阈值的确定是基于 16S rRNA 基因相似度与 *amoA* 基因相似度在统计学上表现出的相关关系得到的，因此 16S rRNA 基因的多拷贝及异质性对相

似性阈值的选取影响较小。

目前已发展了多种技术手段用于环境微生物组研究,包括基因芯片、高通量测序、宏基因组测序等。高通量测序技术以其低成本、高通量、流程自动化的优点备受青睐,功能基因的高通量测序是继 16S rDNA/18S rDNA/ITS 等特定区域片段 PCR 产物测序后的又一项高靶向性测序技术。选择对功能基因区域覆盖度和特异性最佳的扩增引物,通过 PCR 扩增技术获得功能基因扩增产物,利用功能基因的碱基差异来反映环境中功能微生物物种之间的差异。功能基因具有的高特异性与高分辨率,使其作为分子标记不仅能更加准确地靶定环境微生物群中具有特定功能的类群微生物,也能更加清晰地反映菌群中特定类群微生物的物种组成及多样性;不仅可以区分生态功能不同但亲缘相近的微生物种群,也可以鉴定功能相同但亲缘关系较远的微生物种群。此外,随着测序技术的迅猛发展,测序成本的不断降低,尤其第三代测序技术的推广,让更长的功能基因用于功能微生物多样性分析成为了现实。因此,采用功能基因的高通量测序技术,利用本文中提出的分析方法,可以针对自然环境中的特定功能微生物进行有效的多样性分析。

## REFERENCES

- [1] Hatzenpichler R. Diversity, physiology, and niche differentiation of ammonia-oxidizing archaea[J]. *Applied and Environmental Microbiology*, 2012, 78(21): 7501-7510
- [2] Könneke M, Bernhard AE, de la Torre JR, et al. Isolation of an autotrophic ammonia-oxidizing marine archaeon[J]. *Nature*, 2005, 437(7058): 543-546
- [3] Zhang LM, He JZ. A novel archaeal phylum: Thaumarchaeota—a review[J]. *Acta Microbiologica Sinica*, 2012, 52(4): 411-421 (in Chinese)  
张丽梅, 贺纪正. 一个新的古菌类群——奇古菌门 (Thaumarchaeota)[J]. *微生物学报*, 2012, 52(4): 411-421
- [4] Brochier-Armanet C, Boussau B, Gribaldo S, et al. Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota[J]. *Nature Reviews Microbiology*, 2008, 6(3): 245-252
- [5] Jung MY, Well R, Min D, et al. Isotopic signatures of N<sub>2</sub>O produced by ammonia-oxidizing archaea from soils[J]. *The ISME Journal*, 2013, 8(5): 1115-1125
- [6] Arp DJ, Chain PSG, Klotz MG. The impact of genome analyses on our understanding of ammonia-oxidizing bacteria[J]. *Annual Review of Microbiology*, 2007, 61: 503-528
- [7] Pouliot J, Galand PE, Lovejoy C, et al. Vertical structure of archaeal communities and the distribution of ammonia monooxygenase A gene variants in two meromictic High Arctic lakes[J]. *Environmental Microbiology*, 2009, 11(3): 687-699
- [8] Dang HY, Luan XW, Chen RP, et al. Diversity, abundance and distribution of *amoA*-encoding archaea in deep-sea methane seep sediments of the Okhotsk Sea[J]. *FEMS Microbiology Ecology*, 2010, 72(3): 370-385
- [9] Mosier AC, Francis CA. Relative abundance and diversity of ammonia-oxidizing archaea and bacteria in the San Francisco Bay estuary[J]. *Environmental Microbiology*, 2008, 10(11): 3002-3016
- [10] Sun DL, Jiang X, Wu QL, et al. Intragenomic heterogeneity of 16S rRNA genes causes overestimation of prokaryotic diversity[J]. *Applied and Environmental Microbiology*, 2013, 79(19): 5962-5969
- [11] Tindall BJ, Rosselló-Móra R, Busse HJ, et al. Notes on the characterization of prokaryote strains for taxonomic purposes[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2010, 60: 249-266
- [12] Schloss PD, Westcott SL, Ryabin T, et al. Introducing Mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities[J]. *Applied and Environmental Microbiology*, 2009, 75(23): 7537-7541
- [13] Francis CA, Roberts KJ, Beman JM, et al. Ubiquity and diversity of ammonia-oxidizing archaea in water columns and sediments of the ocean[J]. *Proceedings of the National Academy of Sciences*, 2005, 102(41): 14683-14688
- [14] Fish JA, Chai B, Wang Q, et al. FunGene: the functional gene pipeline and repository[J]. *Frontiers in Microbiology*, 2013, 4: 291
- [15] Yang SZ, Liebner S, Alawi M, et al. Taxonomic database and cut-off value for processing *mcrA* gene 454 pyrosequencing data by MOTHUR[J]. *Journal of Microbiological Methods*, 2014, 103: 3-5
- [16] Chao A. Non-parametric estimation of the number of classes in a population[J]. *Scandinavian Journal of Statistics*, 1984, 11: 265-270
- [17] Hill MO. Diversity and evenness: a unifying notation and its consequences[J]. *Ecology*, 1973, 54(2): 427-432
- [18] Santoro AE, Dupont CL, Richter RA, et al. Genomic and proteomic characterization of “*Candidatus Nitrosopelagicus brevis*”: an ammonia-oxidizing archaeon from the open ocean[J]. *Proceedings of the National Academy of Sciences*, 2015, 112(4): 1173-1178
- [19] Biller SJ, Mosier AC, Wells GF, et al. Global biodiversity of aquatic ammonia-oxidizing archaea is partitioned by habitat[J]. *Frontiers in Microbiology*, 2012, 3: 252
- [20] Purkhold U, Pommerening-Röser A, Juretschko S, et al. Phylogeny of all recognized species of ammonia oxidizers based on comparative 16S rRNA and *amoA* sequence analysis: implications for molecular diversity surveys[J]. *Applied and Environmental Microbiology*, 2000, 66(12): 5368-5382
- [21] Pester M, Rattei T, Flechl S, et al. *amoA*-based consensus phylogeny of ammonia-oxidizing archaea and deep sequencing of *amoA* genes from soils of four different geographic regions[J]. *Environmental Microbiology*, 2012, 14(2): 525-539
- [22] Yan YW, Zhang DM. Multiplicity and heterogeneity of 16S rRNA genes in prokaryotes—a mini review[J]. *Journal of Biology*, 2013, 30(4): 63-66 (in Chinese)  
阎永伟, 张德民. 原核生物 16S rRNA 基因多重拷贝及其序列异化[J]. *生物学杂志*, 2013, 30(4): 63-66