

基于元基因组学数据集的海洋微生物多样性分析

鲁水龙, 俞 晓, 全哲学*

(复旦大学 生命科学学院, 上海 200433)

摘 要 通过对海洋 13 个样品元基因组数据的 BLAST 搜索, 筛选到了 16S rRNA 基因序列 1 600 条, 18S rRNA 基因序列 61 条。分类结果显示, 细菌在海岸、公开海域深层海水和表层海水 3 种海洋环境类型中都占优势, 其相对百分比分别为 98%、59% 和 91%。相比于海岸和公开海域表层海水, 公开海域深层海水中古生菌和 Deltaproteobacteria 所占的相对含量较高, 各 31% 和 27%。海岸检测到的古生菌主要为 Euryarcheata, 公开海域深层海水检测到的古生菌主要为 Crenarchaeota (93% 为与氨氧化相关的 MGI 纲)。结果表明, 氨氧化相关古生菌在深海生态系统中的作用可能较大。

关键词 元基因组; 海洋; 细菌; 古生菌; 真核生物

中图分类号 Q938

文献标识码 A

文章编号 1005-7021(2014)05-0008-05

doi:10.3969/j.issn.1005-7021.2014.05.002

Diversity Analysis of Marine Microorganisms Based on Metagenomic Datasets

LU Shui-long, YU Xiao, QUAN Zhe-xue

(Fudan Uni., Schl. of Life Sci., Shanghai 200433)

Abstract Based on BLAST searching of metagenomic data of 13 marine samples, 1 600 pieces of 16S rRNA gene sequences and 61 pieces of 18S rRNA gene sequences were screened. Classification results showed that the bacteria was dominant in coast, open sea waters, deep sea water and surface sea waters, the relative percentage was 98, 59 and 91 respectively. As compared with coast and surface of open sea waters, the contents of Archaea and Deltaproteobacteria were relatively high, they accounted for 31% and 27% respectively. Archaea determined and tested in coast mainly was *Euryarchaea*, while Archaea determined and tested in open deep sea waters mainly was *Crenarchaeota*, and 93% of which was correlated to ammonia oxidizing related class of MGI. The results of this study showed that ammonia oxidation related Archaea may play a larger role in eco-system in deep open seas.

Keywords metagenome; ocean; bacteria; archaea; eukaryote

海洋是地球上最大的生态系统, 占地球总表面积的 71%。海洋微生物主要包括细菌、古生菌和真核微生物, 它们中的一些类群对全球的物质和能量循环起着重要的推动作用^[1-2]。关于海洋微生物的多样性研究一直是环境微生物学的研究热点, 然而至今为止, 关于海洋微生物的多样性研究都是通过基于“通用”引物的 PCR 反应^[3-4], 而

PCR 反应的引物偏向性会遗漏许多重要的微生物^[5]。要解决 PCR 偏向性扩增对微生物群落结构分析的影响就必须绕过 PCR 方法, 即选择不经过 PCR 方法扩增得到的序列进行分析。近些年发展起来的元基因组学(或叫作宏基因组学)为研究人员提供了分析此问题的方法, 元基因组学方法不通过 PCR 扩增, 而是将环境样品中的总

基金项目: 国家自然科学基金面上项目(31170114)

作者简介: 鲁水龙 男, 硕士研究生。研究方向为微生物分子生态学。Tel: 021-65643334, E-mail: Alanlu86@163.com

* 通讯作者。男, 博士, 教授。研究方向为微生物分子生态学。Tel: 021-65643334, E-mail: quanzx@fudan.edu.cn

收稿日期: 2014-03-11; 修回日期: 2014-04-16

DNA 提取后随机打断然后测序,有效避免了引物的偏向性扩增,可以比较真实地反映环境样品中微生物的组成。随着高通量测序技术的发展,许多环境样品进行了元基因组测序,同时这些序列数据也被提交到了公共数据库中。CAMERA(The Community Cyberinfrastructure for Advanced Microbial Ecology Research and Analysis, <http://www.camera.calit2.net/>) 数据库^[6] 由于覆盖了大部分海洋样品的元基因组数据,拥有比较完整的 meta-data 环境信息并提供在线 BLAST 搜索功能而作为本次研究的数据库来源。本研究以 16S rRNA 和 18S rRNA 基因作为分子标记,分别从域、门和纲 3 个分类水平分析海洋微生物在不同海洋环境中的多样性。

1 材料与方法

1.1 Reference 16S rRNA 和 18S rRNA 基因代表序列获取

从 Mothur^[7] (http://www.mothur.org/wiki/Main_Page) 网站下载 3 个含有分类信息的标准小亚基核糖体 RNA (SSU rRNA) 基因序列的 alignment 文件,分别为细菌 16S、古生菌 16S 和真核生物 18S rRNA 基因序列,将其 SSU rRNA 基因序列的文件合并成一个大的文件,命名为“silva.SSU.fasta”;同时从 Silva 数据库中下载大亚基核糖体 rRNA 基因的序列(包括 23S 和 28S rRNA 基因)并命名为“silva.LSU.fasta”。用 Usearch^[8] 软件以 75% 的相似度对“silva.SSU.fasta”分出 71 组 OTU (operational taxonomic unit),再从每组抽取代表序列,作为 BLAST 的待查 reference 序列。

1.2 元基因组数据集的选取

本实验所用的元基因组数据集全部来自 CAMERA 网站(Camera Portal version 2.0.9)^[9]。选择与海洋相关的数据集,分别为 Marine Metagenome from Coastal Waters project、North Pacific metagenomes from Monterey Bay to Open Ocean、Sargasso Sea Bacterioplankton Community 以及 Surface Water Marine Microbial Community Gene Expression。

1.3 16S rRNA 和 18S rRNA 基因元基因组序列的获取及分类

以 71 条代表序列作为 reference,分别对 4 个元基因组数据集进行本地 BLAST 搜索。将筛选出的序列从原序列组中抽出,除去重复的、长度小于 80 bp 的以及含有模糊碱基“N”的序列后,得到 4 个潜在 16S rRNA 基因序列的 FASTA 文件,进一步通过序列比对(Alignment)和分类(classification)对这些潜在序列进行二次筛选和分类。二次筛选和分类采用实验室之前在对通用引物的覆盖率评估时所使用的方法^[5],对该方法的流程进行了自动化并开发 perl 脚本。

2 结果与分析

2.1 筛选到的 13 个海洋样品描述

本研究共筛选到 13 个不同海洋样品,根据环境类型区分为海岸(coast),公开海域深层海水(deep open ocean)和公开海域表层海水(surface open ocean)3 种类型;海岸指深度小于 200 m,公开海域深层海水指深度和采样深度同时大于 200 m,公开海域表层海水指深度大于 200 m,但采样深度小于 200 m。样品的详细环境参数描述见表 1。

表 1 13 个海洋样品环境参数描述

Table 1 Description of environmental parameters of 13 marine samples

样品名	样品类型	采样位置	采样深度/m	水深/m
PML_SMPL_SRS000294	Coa(C1)	60°27'N/5°22'W	-	-
PML_SMPL_SRS000295	Coa(C2)	60°27'N/5°22'W	-	-
PML_SMPL_SRS000296	Coa(C3)	60°27'N/5°22'W	-	-
PML_SMPL_SRS000297	Coa(C4)	60°27'N/5°22'W	-	-
HF_SMPL_HOT179_500M_GDNA	Doo(D1)	22°75'N/158°E	500	4 750
HF_SMPL_HOT186_500M_GDNA	Doo(D2)	22°73'N/158°03'E	500	4 750
HF_SMPL_BATS216_20M_SG	Soo(S1)	31°67'N/64°17'E	20	4 200
HF_SMPL_BATS216_50M_SG	Soo(S2)	31°67'N/64°17'E	50	4 200
HF_SMPL_HOT179_25M_SG	Soo(S3)	22°73'N/158°03'E	25	4 750

续表

HF_SMPL_HOT179_75M_SG	Soo(S4)	22°73'N/158°03'E	75	4 750
HF_SMPL_HOT186_110M_GDNA	Soo(S5)	22°73'N/158°03'E	110	4 750
HF_SMPL_HOT186_25M_GDNA	Soo(S6)	22°73'N/158°03'E	25	4 750
HF_SMPL_HOT186_75M_GDNA	Soo(S7)	22°73'N/158°03'E	75	4 750

注: Coa 代表海岸样品, Doo 代表公开海域深层海水样品, Soo 代表公开海域表层海水样品, 下表同

2.2 域分类水平海洋微生物多样性研究

用 71 条代表序列对 4 个海洋元基因组数据集进行 BLAST 搜索和筛选后, 得到 1 661 条元基因组序列。分类结果表明古生菌共有 119 条, 细菌 1 481 条, 真核生物 61 条。古生菌、细菌和真核生物在每个样品中的详细分布见表 2。同时计算了古生菌、细菌和真核生物在海岸、公开海域深层海水和表层海水 3 种海洋环境类型中的相对百分

比(图 1)。古生菌在海岸、公开海域深层海水和表层海水的相对百分比分别为 0、31% 和 3%; 真核生物在海岸、公开海域深层海水和表层海水 3 种海洋环境类型中均可检测到, 但其相对百分比比较低, 分别为 3%、11% 和 6%。细菌在海岸、公开海域深层海水和表层海水 3 种海洋环境类型中都占主导优势, 其相对百分比分别为 98%、59% 和 91%。

表 2 古生菌、细菌和真核生物在 13 个海洋样品中的分布

Table 2 The distribution of Archaea, Bacteria and Eukaryote in 13 marine samples

样品名	样品编号	古生菌	细菌	真核生物	总计
PML_SMPL_SRS000294	C1	0	140	1	141
PML_SMPL_SRS000295	C2	0	100	3	103
PML_SMPL_SRS000296	C3	0	123	3	126
PML_SMPL_SRS000297	C4	0	146	6	152
HF_SMPL_HOT179_500M_GDNA	D1	11	13	6	30
HF_SMPL_HOT186_500M_GDNA	D2	90	265	5	360
HF_SMPL_BATS216_20M_SG	S1	0	112	1	113
HF_SMPL_BATS216_50M_SG	S2	0	103	0	103
HF_SMPL_HOT179_25M_SG	S3	1	48	11	60
HF_SMPL_HOT179_75M_SG	S4	2	120	3	125
HF_SMPL_HOT186_110M_GDNA	S5	6	37	7	50
HF_SMPL_HOT186_25M_GDNA	S6	6	166	11	183
HF_SMPL_HOT186_75M_GDNA	S7	3	108	4	115
总计		119	1 481	61	1 661

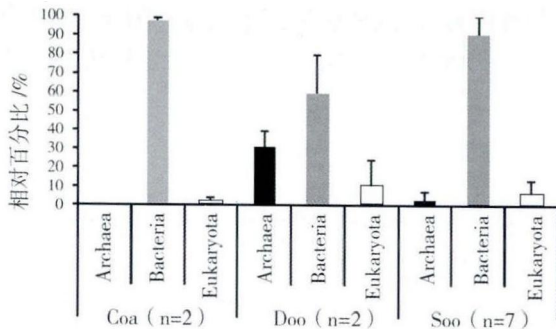


图 1 古生菌、细菌和真核生物在 3 种海洋环境类型中的分布

Fig. 1 The distribution of Archaea, Bacteria and Eukaryote in three marine habitats
n 表示样品数量

n means the number of samples included in this analysis

2.3 门分类水平海洋微生物多样性研究

门分类结果表明古生菌、细菌和真核生物可分为 30 个门, 按照平均百分比进行排序, 选择了平均百分比大于 1% 的 13 个门(图 2), 其中古生菌包括 Crenarchaeota(3%) 和 Euryarchaeota(3%) 2 个门; 真核生物包括 Viridiplantae(2%)、Alveolata(2%)、Stramenopiles(1%) 和 Metazoa(1%) 4 个门; 细菌包括 Proteobacteria(52%)、Cyanobacteria(16%)、Bacteroidetes(8%)、Deferribacteres(4%)、Actinobacteria(3%)、Verrucomicrobia(1%) 和 Chloroflexi(1%) 7 个门。Crenarchaeota 门只在深海样品 D1 和 D2 中检测到, 相对百分比分别为 27% 和 18%。真核生物中 Viridiplantae 门主要分布在 D1 样品中(13%), Metazoa 门主要分

布在 S3 样品中(12%), Stramenopiles 门主要分布在 S5 样品中(10%)。

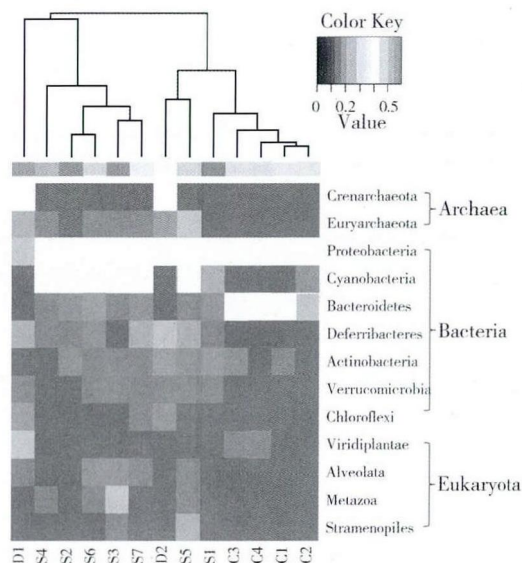


图2 13个门在13个海洋样品中的分布

Fig.2 The distribution of 13 phylums in 13 marine samples
颜色深浅表示该菌在样品中的相对百分比,这里只展示了平均相对百分比大于1%的门,13个不同样品进行了 hclust 聚类分析

The different colors mean the percent of each type of microorganisms in each sample. Only the phyla avarage of which content was larger than 1% was shown in this table. Thirteen samples were grouped based on hclust analysis

2.4 纲水平 Proteobacteria 门和 Crenarchaeota 门多样性分析

由于 Proteobacteria 门在 13 个海洋样品中的平均百分比达到了 52%,因此对其下面的主要纲在 3 种海洋环境类型中的分布进行了分析(图 3)。选取了平均百分比大于 1% 的 4 个纲,即 Alphaproteobacteria(55%)、Betaproteobacteria(2%)、Deltaproteobacteria(6%)和 Gammaproteobacteria(36%)。Alphaproteobacteria 纲在海岸、公开海域深层海水和表层海水的相对百分比分别为 55%、34%和 61%,说明其在海洋中分布广泛;Betaproteobacteria 纲在海岸、公开海域深层海水和表层海水的相对百分比分别为 6%、0.3%和 0.2%,这也与 Betaproteobacteria 纲本身在海洋中分布较少相关。Deltaproteobacteria 纲在海岸、公开海域深层海水和表层海水的相对百分比分别为 0、27%和 3%。Gammaproteobacteria 纲在海岸、公开海域深层海水和表层海水的相对百分比分别为 38%、

38%和 35%,与 Alphaproteobacteria 纲一样, Gammaproteobacteria 纲也广泛分布于海洋中。同时,又对 Crenarchaeota 门的主要纲进行了分析,结果表明 Marine Group I 占主导优势,相对百分比为 93%。Marine Benthic Group A 纲和 AK31 纲的相对百分比分别为 4%和 3%。

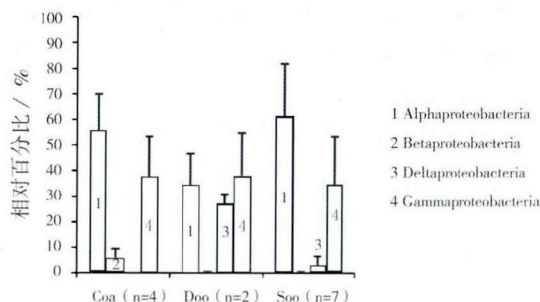


图3 Proteobacteria 门主要纲在 3 种海洋环境类型中的分布

Fig.3 The distribution of main class of Proteobacteria in three marine habitats

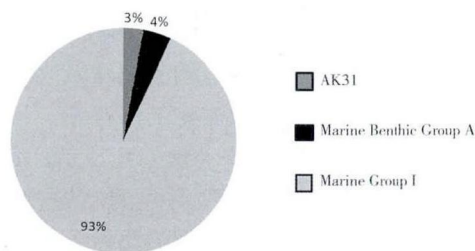


图4 Crenarchaeota 门主要纲多样性分析

Fig.4 Analysis of main class of Crenarchaeota

3 讨论

在海岸样品中,没有检测到古生菌,这也与实际相符合,因为古生菌一般分布于深海或极端环境,如碳循环中的产甲烷古生菌^[10]和氮循环中的氨氧化古生菌^[11]。通常认为真核微生物分布在海岸及海面表层,那里生存环境氧气充足,阳光充裕,符合大多真核微生物好氧的特点,但在本研究结果中,深海的相对百分比高,这也可能是因为只考虑 2 个深海样品所造成的误差。在海洋微生物群中,细菌占绝大多数,虽然样品采集地点较少,但也基本覆盖了海洋从表层到深层、从海岸到公开海域、从赤道附近到极地各种环境类型。

Crenarchaeota 门是古生菌的一大分支,最初

是在海岸样品中通过 16S rRNA 系统发育分析得到的^[12],在深海样品检测到 Crenarchaeota 门,说明其在海洋分布的广泛性。细菌中含量最高的门是 Proteobacteria 门,特别是在海岸样品 C1 (82%)、C2 (80%)、C3 (69%)、C4 (76%) 和公开海域表层海水样品 S1 (74%),表明其偏好阳光和氧气充足的海水表层。同时观察发现 Bacteroidetes 门也主要分布于海岸样品 C1 (14%)、C2 (14%)、C3 (22%)、C4 (18%)。

Deltaproteobacteria 纲的微生物通常参与硫还原反应,因此往往分布于深海或深海沉积物^[13],这也与本研究得出的结果相一致。Marine Group I (MGI) 是古生菌中多样性很高的一个类群,并且在海洋环境中广泛分布,对全球海洋的氨氧化有显著作用^[14]。MGI 起初被分类到 Crenarchaeota 门,不过最近的比较基因组学和进化分析结果表明 MGI 是属于 Thaumarchaeota 门^[15]。尽管 MGI 在深海有非常高的丰度(图 2 和图 4),但与浅海的 MGI 类群相比研究的还很少,未来还有很多关于 MGI 的生理和代谢问题需要回答。

本研究基于元基因组的方法从域、门、纲 3 种分类水平分析了海洋微生物的多样性。由于元基因组只能反映环境中微生物的组成,而元转录组可以反映环境中具有生物活性的微生物的组成,因此未来还可以从元转录组的角度分析海洋环境微生物的活性组成。同时,由于元基因组和元转录组都可以避免 PCR 的偏向性扩增,因此还可以对常用“通用引物”的覆盖率进行评估。

参考文献:

- [1] Boetius A, Ravenschlag K, Schubert CJ, et al. A marine microbial consortium apparently mediating anaerobic oxidation of methane [J]. *Nature*, 2000, 407(6804): 623-626.
- [2] Wuchter C, Abbas B, Coolen M J, et al. Archaeal nitrification in the ocean [J]. *Proceedings of the National Academy of Sciences*, 2006, 103(33): 12317-12322.
- [3] Pommier T, Canback B, Riemann L, et al. Global patterns of diversity and community structure in marine bacterioplankton [J]. *Molecular Ecology*, 2007, 16(4): 867-880.
- [4] Zinger L, Amaral-Zettler L A, Fuhrman J A, et al. Global Patterns of Bacterial Beta-Diversity in Seafloor and Seawater Ecosystems [J]. *Plos One*, 2011, 6(9): e24570.
- [5] Mao DP, Zhou Q, Chen CY, et al. Coverage evaluation of universal bacterial primers using the metagenomic datasets [J]. *BMC Microbiology*, 2012, 12(1): 66.
- [6] Sun SL, Chen J, Li WZ, et al. Community cyberinfrastructure for Advanced Microbial Ecology Research and Analysis: the CAMERA resource [J]. *Nucleic Acids Research*, 2011, 39: D546-D551.
- [7] 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.
- [8] Edgar, Robert C. Search and clustering orders of magnitude faster than BLAST [J]. *Bioinformatics*, 2010, 26(19): 2460-2461.
- [9] Seshadri R, Kravitz SA, Smarr L, et al. CAMERA: a community resource for metagenomics [J]. *Plos Biology*, 2007, 5(3): e75.
- [10] Hinrichs KU, Hayes JM, Sylva SP, et al. Methane-consuming archaeobacteria in marine sediments [J]. *Nature*, 1999, 398(6730): 802-805.
- [11] Park SJ, Park BJ, Rhee SK. Comparative analysis of archaeal 16S rRNA and amoA genes to estimate the abundance and diversity of ammonia-oxidizing archaea in marine sediments [J]. *Extremophiles*, 2008, 12(4): 605-615.
- [12] DeLong EF. Archaea in coastal marine environments [J]. *Proceedings of the National Academy of Sciences*, 1992, 89(12): 5685-5689.
- [13] Miyatake T, MacGregor BJ, Boshker HT. Linking microbial community function to phylogeny of sulfate-reducing Deltaproteobacteria in marine sediments by combining stable isotope probing with magnetic-bead capture hybridization of 16S rRNA [J]. *Applied and Environmental Microbiology*, 2009, 75(15): 4927-4935.
- [14] 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 of the United States of America*, 2005, 102(41): 14683-14688.
- [15] 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.