

宏基因组学分析揭示深古菌 Bathyarchaeota B242 的代谢特征

冯晓远^{1,2} Venki Perumal^{1,2} 王风平^{1,2,3*}

(1. 上海交通大学生命科学技术学院 上海 200240)

(2. 上海交通大学微生物代谢国家重点实验室 上海 200240)

(3. 上海交通大学海洋工程国家重点实验室 上海 200240)

摘要:【背景】海洋沉积物中蕴含着丰富的微生物资源,估算约 2.9×10^{29} 个细胞,与海水中的微生物总量相当。但是由于缺少可培养物,大部分的微生物缺乏生理特征、代谢方式以及生态功能的相关研究。深古菌(Bathyarchaeota)是一类典型的未培养微生物,在全球海洋沉积物中普遍存在,并且具有很高的丰度。【目的】对深古菌代谢潜能及其在海洋沉积物中发挥的生态功能进行更加深入的研究。【方法】应用宏基因组学的技术手段,对采集自瓜伊马斯盆地的深海热液沉积物样本进行了分析,获得了一个接近完整的深古菌基因组 Bathyarchaeota B242。【结果】对 Bathyarchaeota B242 基因组的分析发现,其具有以降解蛋白质和多种碳水化合物为主的异养代谢途径,同时还具有通过还原型乙酰辅酶 A 途径实现的自养途径。【结论】同时具有自养和异养代谢途径对 Bathyarchaeota B242 适应低物质能量供给环境下的生存起到重要作用。

关键词: 海洋沉积物, 深古菌, 宏基因组, 蛋白质降解, WL 自养途径

Metagenomic analysis revealing metabolic characteristics of Bathyarchaeota B242

FENG Xiao-Yuan^{1,2} Venki Perumal^{1,2} WANG Feng-Ping^{1,2,3*}

(1. School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China)

(2. State Key Laboratory of Microbial Metabolism, Shanghai Jiao Tong University, Shanghai 200240, China)

(3. State Key Laboratory of Ocean Engineering, Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai 200240, China)

Abstract: [Background] Approximately 2.9×10^{29} cells reside in marine sediments, roughly equal to the estimates of cell abundance in seawater. Due to the lack of cultivated representative species of these microbes in laboratories, their physiology and metabolic functions remain largely unknown. The members of Bathyarchaeota are typical sedimentary microorganisms and widely distributed in global marine sediments. [Objective] For further understanding of their metabolic potential and ecological roles that remain largely elusive. [Methods] Samples from Guaymas Basin hydrothermal sediment

Foundation items: National Natural Science Foundation of China (41525011, 91228201, 91428308)

*Corresponding author: Tel: 86-21-34207205; E-mail: fengpingw@sjtu.edu.cn

Received: September 13, 2017; **Accepted:** November 02, 2017; **Published online** (www.cnki.net): November 20, 2017

基金项目: 国家自然科学基金项目(41525011, 91228201, 91428308)

*通信作者: Tel: 86-21-34207205; E-mail: fengpingw@sjtu.edu.cn

收稿日期: 2017-09-13; 接受日期: 2017-11-02; 网络首发日期(www.cnki.net): 2017-11-20

were analyzed by metagenomic technology, and a near-complete genome belonging to Bathyarchaeota B242 was obtained. **[Results]** Through comprehensive genomic analysis, Bathyarchaeota B242 was found having heterotrophic mode of life style mainly through protein fermentation, carbohydrate utilization as well as possess autotrophic acetogenic pathway that helps survive under energy deficient environments. **[Conclusion]** The heterotrophic and autotrophic metabolic mode together contributed to the adaptation in energy deficient marine sediments for Bathyarchaeota.

Keywords: Marine sediments, Bathyarchaeota, Metagenomics, Protein degradation, Autotrophic WL pathway

海洋占据了全球 71% 的地表面积, 其中超过半数的微生物生活在能量供应稀缺的海底沉积物中^[1]。这些海底沉积物中的微生物生存并繁衍于高度变化的环境之中, 主要依赖于上层海水有机质和营养物质的沉降获取能量。然而由于现有实验培养条件和方法的限制, 多数此类微生物缺乏可培养代表, 仅有基于 16S rRNA 基因的相关研究。因此, 不依赖于培养、基于组学数据的分析手段为研究此类微生物的进化地位和代谢潜能提供了新的可行途径, 并为进一步研究其在全球元素循环过程中的重要地位提供了有效帮助。

深古菌(Bathyarchaeota), 旧称 MCG (Miscellaneous Crenarchaeota Group), 是一类新近发现的古菌, 为门(Phylum)水平分类单元, 在世界范围内具有广泛多样的生态分布^[2-3], 在全球碳循环过程中也发挥着至关重要的作用^[4], 据推测可能是全球丰度最高的微生物之一。基于 16S rRNA 基因的系统发育学研究揭示出其复杂的内部亚群多样性, 目前被分为 MCG1-19 等多个亚群^[2,5-6], 具有多样的生理生化代谢特征。然而, 截至目前, 仅有数篇直接在基因组层面揭示其代谢潜能和生态功能的相关研究报道。其中, 深古菌被报道具有降解蛋白质的能力^[3,7], 并且在某些亚群中还发现了与产甲烷途径相关的代谢途径^[8]。此外, 我们之前的研究还证实其作为产乙酸菌的特性, 并对相关基因进行了生理功能验证^[9]。但考虑到其内部复杂的亚群分类和代谢特征, 对其内部的进化分类地位和生态功能仍然亟待研究。

本研究使用宏基因组学分析手段, 对采集自瓜伊马斯盆地深海热液沉积物的样本进行了精细的拼接和分装, 获得了一个编号为 Bathyarchaeota

B242 的接近完整基因组, 并对其代谢潜能和进化地位进行了全面研究, 为后续深古菌的比较基因组分析提供了更加充实的基因组信息。

1 材料与方法

1.1 样品信息及 DNA 提取测序

样本为 2007 年 10 月在 Alvin 4358 次潜航(北纬 27°0'43", 西经 111°24'19", 深度 2 016 m), 采集自瓜伊马斯盆地(加利福尼亚墨西哥湾)的深海热液沉积物(深度为 0-10 cm), 样品采集后储存于 4 °C 丁基胶塞玻璃瓶中保存。该沉积物样本为油性且有白色生物层覆盖。其温度由 Alvin 温度探针检测, 样本表面温度为 5 °C, 且温度随着沉积物深度的增加而增加, 在 10 cm 处达到 44 °C。后续的 DNA 提取和测序分析同文献[9]。

1.2 测序数据组装和基因组分装

原始宏基因组测序数据由测序公司返回后, 使用 Sickle 软件(<https://github.com/najoshi/sickle>)对其进行质量控制, 对测序质量低的碱基进行剪切和删除, 使用参数为--se。随后使用 IDBA-UD^[10]进行拼接, 使用参数为--mink 52、--maxk 92 和--step 8。对于拼接后得到长度大于 5 kb 的拼接片段, 通过比较四核苷酸频率特征, 使用 ESOM^[11]进行分装, 并使用 Bowtie 2^[12]将原始的 DNA 短读段比对到拼接片段上, 计算每条拼接片段的测序深度。随后使用 mmgenome^[13]对这些初步分装获得的基因组进行基于四核苷酸频率特征分析、G+C%含量和测序深度的整合分析, 手动检查并进行更加精细的基因组片段分装, 最后使用 CheckM^[14]对其进行初步的物种鉴定以及完整度和污染度的评估。

1.3 功能基因的预测和注释

对于上述拼接和分装得到的高质量基因组, 使用 Prodigal^[15]进行蛋白编码序列的预测, 使用参数-meta; 同时分别使用 RNAmmer^[16]和 CheckM^[14]进行 16S rRNA 基因的预测。对于蛋白质编码序列, 将其比对已知的注释信息数据库 NCBI NR、KEGG^[17]和 COG^[18]得到功能注释, 并手动检测注释的准确性, 使用参数为 E-values<10⁻⁵。此外, 还将这些蛋白质编码序列比对到肽酶数据库 MEROPS^[19], 以寻找编码肽酶的基因并进行分类, 并使用 PRED-SIGNAL^[20]进行细胞外转运信号的检测。

1.4 系统发育树的构建

为了确定样本中分装得到的基因组物种分类信息, 收集了 350 个已知的古菌基因组和 20 个已经发表的深古菌基因组^[3,7-9,21-22]作为参考基因组, 用于构建后续的系统发育树。通过文献调研, 使用 16 个经证实认为没有水平基因转移的古菌单拷贝保守基因(核糖体蛋白 L14、L15、L16/L10AE、L18、L2、L22、L24、L3、L4、L5、L6P/L9E、S10、S17、S19、S3、S8)^[23]构建系统发育树以确定各基因组的进化地位。对于上述的每一个保守单拷贝基因的蛋白序列进行相同的操作: 合并目标基因组和参考基因组中的对应蛋白序列, 使用 MAFFT^[24]进行比对, 参数使用默认。随后根据其所属的基因组, 将比对好的每一组序列按照相同顺序重新首位连接, 用空位补齐缺失的蛋白。将上述连接后得到的序列使用 SeaView^[25]进行可视化编辑, 手动检查并删除高度可变区和比对时引入的大量空位。系统发育树通过使用 FastTree^[26]、Jones-Taylor-Thornton (JTT)模型和 CAT 预测法构建最大似然树。自展值通过 1 000 次重复计算得到。

2 结果与分析

2.1 B242 基本信息

通过对来自瓜伊马斯盆地深海热液沉积物样本宏基因组测序数据的重新拼接和分装, 获得了一个接近完整的深古菌基因组 B242 (表 1)。其基因组大小为 1.34 Mb, 在已知的深古菌基因组中相对较小,

表 1 B242 基因组基本信息

Table 1 Summary information for Bathyarchaeota B242

基因组信息 Genomic information	B242
基因组大小 Genome size	1 347 170 bp
G+C%含量 G+C% content	47.36%
完整度 Completeness	96.88%
污染度 Contamination	1.01%
N50	19 870 bp
测序深度 Sequencing depth	16.92
基因数 Gene count	1 365

G+C%含量为 47.36%, 平均测序深度为 16.92。根据 CheckM 评估, 其完整度为 96.88%, 污染度为 1.01%, 属于高质量的基因组, 适合后续的相关分析研究。

相关的基因组序列已提交至公共数据库 DDBJ/ENA/GenBank, 其 BioProject 编号为 PRJNA312413, BioSample 编号为 SAMN07644978。

然而, 在深古菌 B242 基因组中并没有检测到 16S rRNA 编码基因。因此, 仅能根据基于串联保守基因蛋白质序列进行进化树的构建(图 1)确定其进化地位。结果显示, B242 的基因组独立于已知的其他深古菌基因组分支之外, 进化地位尚不明确。对其进化地位的进一步确定有待于更多深古菌相关基因组的发现。

2.2 B242 异养代谢潜能分析

基因组分析结果表明, 深古菌 B242 具有异养代谢潜能(图 2, 表 2), 可以吸收和降解蛋白质、淀粉和聚乳糖等复杂化合物, 并通过糖酵解、磷酸戊糖等代谢途径对其降解产物进行进一步利用。

在蛋白质降解方面, B242 编码多种胞外蛋白酶、转运蛋白、胞内蛋白酶、氨基转移酶和铁氧还原蛋白酶, 具有完整的蛋白质降解途径。B242 共编码 1 365 个基因, 其中被蛋白酶数据库 MEROPS^[19]注释为具有蛋白酶降解功能的基因共有 151 个, 占总体基因组的 11.06%, 显示出蛋白质降解对 B242 的生存代谢具有重要作用。此外, 蛋白质代谢的相关基因也在其他已经公开发表的深古菌基因组中

Note: The phylogeny is based on conserved single copy genes; The black and white dots represent bootstrap value larger than 90 and 80, respectively.

Figure 2 Key metabolic pathways in Bathyarchaeota B242

表 2 深古菌 B242 的基本代谢模型相关基因

Table 2 Genes involved in key metabolic pathways in Bathyarchaeota B242

代谢过程 Metabolic pathways	功能注释 Annotation	缩写 Abbreviation	BLAST 物种 BLAST taxonomy	BLAST 相似度 BLAST identity (%)
蛋白质降解 Protein degradation	Extracellular serine protease	—	Bathyarchaeota B23	54
		—	Bathyarchaeota B26-2	46
淀粉降解 Starch degradation	Alpha-amylase	<i>amy</i>	Bathyarchaeota BA2	53
	Glycogen phosphorylase	<i>pyg</i>	Bathyarchaeota SG8-32-3	67
	Glycogen debranching enzyme	<i>agl</i>	Bathyarchaeota BA2	46
聚乳糖降解 Galactan degradation	Galactosidase	<i>lacZ</i>	Bathyarchaeota B26-2	72
		<i>lacZ</i>	Bathyarchaeota B26-2	69
糖酵解 Glycolysis	Glucokinase	<i>glk</i>	Bathyarchaeota B63	52
	Glucose-6-phosphate isomerase	<i>pgi</i>	Bathyarchaeota BA2	46
	6-Phosphofructokinase	<i>pfkA</i>	Bathyarchaeota B26-2	64
	Fructose 1,6-bisphosphate aldolase	—	Bathyarchaeota BA1	76
	Fructose-bisphosphate aldolase	—	Bathyarchaeota B63	76
	Triosephosphate isomerase	<i>tpi</i>	Bathyarchaeota B26-2	63
	Glyceraldehyde-3-phosphate dehydrogenase	<i>gapA</i>	Bathyarchaeota B26-2	81
		<i>gapA</i>	Bathyarchaeota B26-2	86
	3-Phosphoglycerate kinase	<i>pgk</i>	Bathyarchaeota B26-1	71
	Enolase	<i>eno</i>	Bathyarchaeota SMTZ-80	63
		<i>eno</i>	Bathyarchaeota B26-2	62
	Pyruvate kinase	<i>pyk</i>	Bathyarchaeota B26-2	62
磷酸戊糖途径 Pentose phosphate pathway	6-Phosphogluconolactonase	<i>pgl</i>	Bathyarchaeota B63	73
	6-Phosphogluconate dehydrogenase	<i>pgd</i>	Bathyarchaeota RBG_13_38_9	70
	Ribose-phosphate pyrophosphokinase	<i>prpS</i>	Bathyarchaeota BA2	61
	Ribokinase	<i>rbsK</i>	Bathyarchaeota B63	60
		<i>rbsK</i>	Bathyarchaeota B26-2	61
	Xylulokinase	<i>xylB</i>	Bathyarchaeota SMTZ1-55	49
		<i>xylB</i>	Bathyarchaeota SMTZ1-55	49
		<i>xylB</i>	Bathyarchaeota B26-1	52
WL 途径 WL pathway	Formylmethanofuran dehydrogenase	<i>fwdA</i>	Bathyarchaeota B26-2	70
		<i>fwdB</i>	Bathyarchaeota BA1	73
		<i>fwdC</i>	Bathyarchaeota BA1	60
		<i>fwdD</i>	Bathyarchaeota BA1	61
		<i>fwdF</i>	Bathyarchaeota BA2	61
	Formylmethanofuran tetrahydromethanopterin N-Formyltransferase	<i>ftr</i>	Bathyarchaeota B26-1	69
	Methenyltetrahydromethanopterin cyclohydrolase	<i>mch</i>	Bathyarchaeota BA1	63
	Methylenetetrahydromethanopterin dehydrogenase	<i>mtd</i>	Bathyarchaeota B63	60
	Acetyl-CoA synthase	<i>cdhA</i>	Bathyarchaeota B63	80
		<i>cdhB</i>	Bathyarchaeota BA1	67
		<i>cdhC</i>	Bathyarchaeota SG8-32-3	78

被检测到,说明蛋白质降解这一代谢过程可能作为一种通用的生存策略在深古菌中广泛存在。

在碳水化合物降解方面,B242 编码蛋白淀粉酶(Amy)、糖原磷酸化酶(Pyg)和糖原脱支酶(Agl),具有将淀粉降解为葡萄糖或磷酸葡萄糖的功能。此外,B242 还编码半乳糖苷酶(LacZ),具有降解聚乳糖的能力。

对于其他主要的复杂化合物,如纤维素、脂肪酸、芳香族化合物等,由于缺乏一系列关键代谢基因,因此 B242 可能并不具有相关的代谢功能。虽然 B242 并非完整基因组,其代谢相关的关键基因可能并没有得以分装。但是考虑到上述代谢过程所包含的一系列编码基因都没有被检测到,因此可以初步认为,深古菌 B242 很可能通过降解蛋白质、淀粉和聚乳糖的方式进行异养代谢。

此外,深古菌 B242 还具有糖酵解途径和磷酸戊糖途径,可以对上述复杂化合物,尤其是碳水化合物降解得到的产物进行进一步利用。B242 编码完整的糖酵解途径,能够将葡萄糖转化为乙酰辅酶 A 并从中获得能量,这一代谢途径更加证明其具有异养代谢的特征。此外,B242 还编码不完整的磷酸戊糖途径,包含其中的磷酸葡萄糖酸内酯酶(Pgl)、磷酸葡萄糖酸脱氢酶(Pgd),以及下游代谢相关的核糖磷酸焦磷酸激酶(Prp)、核糖激酶(Rbs)和木酮糖激酶(Xyl)等基因,具有合成生命必需的核苷酸和芳香酸前体的潜能。

2.3 B242 自养代谢潜能分析

虽然大量的复杂化合物降解相关基因显示出深古菌 B242 具有异养代谢特征,B242 同时还编码还原型乙酰辅酶 A 途径(Wood-Ljungdahl, WL pathway)过程中的大多数基因,其中包括 Formylmethanofuran dehydrogenase (Fwd)、Formylmethanofuran tetrahydromethanopterin N-formyltransferase (Ftr)、Methenyltetrahydromethanopterin cyclohydrolase (Mch)、Methylenetetrahydromethanopterin dehydrogenase (Mtd)以及 Acetyl-CoA decarbonylase (Acs/codh)。然而,编码 5,10-Methylenetetrahydromethanopterin reductase

(Mer)的基因并没有在 B242 中被检测到,该现象可能是由 B242 并非完整基因组的原因造成。

总之,深古菌 B242 可以通过使用四氢叶酸作为一碳载体,以典型的古菌类型^[27]还原型乙酰辅酶 A 途径进行自养代谢。此外,这一代谢途径还在其他深古菌基因组中被发现,是深古菌共有的核心代谢特征之一。

2.4 B242 其他代谢潜能分析

虽然之前的数篇研究结果表明,某些亚群的深古菌可以进行乙酸代谢途径^[9]或甲烷代谢途径^[8],但是通过对深古菌 B242 的基因组研究表明,B242 中并没有发现相关的基因,如乙酸代谢途径的基因磷酸乙酰转移酶(Pta)、乙酸激酶(Ack)或乙酰辅酶 A 合成酶(Acd)和甲烷代谢途径的甲基辅酶 M 还原酶(Mcr)等。类似的,之前报道的亚硝酸盐还原途径^[7]也没有在深古菌 B242 基因组中被检测到。考虑到深古菌 B242 并非完整基因组,其是否具有相应的产乙酸、甲烷或者亚硝酸盐还原代谢特征仍然有待进一步研究证实。

深古菌 B242 还编码多种电子传递体,其中包括二硫化物还原酶(Hdr)甲基紫精还原酶(Mvh)等还原铁氧蛋白的复合物。此外,大量氢化酶和辅酶 F₄₂₀ 相关的基因也在 B242 中被发现,其中包括 F₄₂₀H₂ 脱氢酶(Fpo)、辅酶 F₄₂₀ 氢化酶(Frh)、电子转移黄素蛋白(Etf)和氢化酶成熟酶(Hyp)等,显示出深古菌 B242 可能使用氢气和辅酶 F₄₂₀ 作为电子传递链中的主要供体和受体。

深古菌 B242 还编码多种跨膜运输蛋白,可以进行亚精胺(Pot)、糖类(ABC.MS)、支链氨基酸(Liv)、锌离子(Znu)、二价铁离子(Feo)等物质的跨膜运输,说明 B242 具有复杂多样的代谢功能。

深古菌 B242 还编码一系列氧化应激反应相关的酶,其中包括硫氧还蛋白还原酶(TrxB)、核苷三磷酸还原酶(Rtp)、硫氧还蛋白(TrxA)等。此外,B242 还编码仅在嗜热菌中存在的 DNA 修复蛋白逆促旋酶(Rgy)。参考先前的研究结果^[9],B242 同样具有潜在的兼性厌氧代谢特征和热适应特征。

3 讨论

之前的研究表明, 本实验采样的瓜伊马斯盆地具有大量的热液羽流区域, 其沉积物中包含丰富的硫化物、甲烷以及溶解的无机碳。本实验中使用的样品来自于有机物和硫化物富集环境, 其温度随沉积物深度增加而迅速增加。

深古菌是一类在全球范围内广泛分布并大量存在的微生物, 具有多样的亚群分类。然而, B242 的基因组中并没有包含 16S rRNA 编码基因, 这为其亚群分类的定位带来了一定困难。此外, 在基于单拷贝保守核糖体蛋白的系统进化树上, B242 并没有与之相近的参考基因组, 因此, 对深古菌 B242 的亚群分类定义仍需要未来更多基因组数据的支持。

蛋白质被认为是海洋有机物中最大的组分之一^[28]。与之前的研究结果类似, 我们发现 B242 中包含了蛋白质降解的完整途径。此外, B242 还具有淀粉和聚乳糖的代谢潜能, 以及相关的糖酵解能力, 也证明 B242 具有异养代谢的能力。B242 还编码不完整的磷酸戊糖途径, 可以完成细胞内必需分子的转化、合成和利用。最后, 通过对电子传递体, 尤其是多种氢酶的分析, B242 可能使用氢气和还原型 F_{420} 作为电子传递链中的供体和受体。

产乙酸途径和产甲烷途径被认为是微生物能量代谢最基本和最古老的生化代谢途径之一^[29]。虽然在其他深古菌的研究工作中发现了对应的基因和代谢途径, 但是在 B242 中, 相关的基因并没有被检测到。考虑到深古菌 B242 的完整度为 96.88%, 其相关的基因仍可能存在于未分装得到的完整基因组中, 因此, B242 是否具有相应的产乙酸或产甲烷能力仍然需要更多基因组数据的支持。

对已有的深古菌基因组^[3,7-9]进行研究发现, 深古菌可能具有兼性厌氧代谢潜能。B242 中也编码多种氧压力适应相关的基因, 可能具有类似的兼性厌氧的代谢潜能。此外, B242 还编码仅存在于超嗜热菌中存在的 DNA 修复蛋白逆促旋酶, 然而其是否具有相关的热适应能力仍然有待进一步研究。

4 结论

深古菌 *Bathyarchaeota* 是一个新近发现的古菌门, 其广泛的地理分布和多样的代谢潜能使其成为目前微生物生态学研究的一个热点领域。考虑到其在海洋沉积物中的广泛分布及其内部复杂的亚群分类和代谢潜能(如复杂化合物降解、产乙酸过程、产甲烷过程等), 深古菌在全球碳元素的循环过程中可能发挥着重要作用。本研究通过对瓜伊马斯盆地深海热液沉积物样本的宏基因组分析, 获得了一个接近完整的深古菌基因组 *Bathyarchaeota* B242, 为深古菌进一步研究提供了新的基因组数据。B242 具有降解蛋白质、淀粉和聚乳糖等多种异养代谢途径和还原型乙酰辅酶 A 的自养代谢途径, 而且具有潜在的兼性厌氧和热适应能力。本研究为深古菌类群复杂代谢潜能的进一步探索, 及其在全球海洋沉积物的碳循环过程中所发挥的重要作用提供了更加丰富的数据支持和帮助。

REFERENCES

- [1] Kallmeyer J, Pockalny R, Adhikari RR, et al. Global distribution of microbial abundance and biomass in subseafloor sediment[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2012, 109(40): 16213-16216
- [2] Kubo K, Lloyd KG, Biddle JF, et al. Archaea of the Miscellaneous Crenarchaeotal Group are abundant, diverse and widespread in marine sediments[J]. *The ISME Journal*, 2012, 6(10): 1949-1965
- [3] Lloyd KG, Schreiber L, Petersen DG, et al. Predominant archaea in marine sediments degrade detrital proteins[J]. *Nature*, 2013, 496(7444): 215-218
- [4] Fry JC, Parkes RJ, Cragg BA, et al. Prokaryotic biodiversity and activity in the deep subseafloor biosphere[J]. *FEMS Microbiology Ecology*, 2008, 66(2): 181-196
- [5] Lazar CS, Biddle JF, Meador TB, et al. Environmental controls on intragroup diversity of the uncultured benthic archaea of the miscellaneous Crenarchaeotal group lineage naturally enriched in anoxic sediments of the White Oak River estuary (North Carolina, USA)[J]. *Environmental Microbiology*, 2015, 17(7): 2228-2238
- [6] Fillol M, Auguet JC, Casamayor EO, et al. Insights in the ecology and evolutionary history of the Miscellaneous Crenarchaeotic Group lineage[J]. *The ISME Journal*, 2016, 10(3): 665-677
- [7] Lazar CS, Baker BJ, Seitz K, et al. Genomic evidence for distinct carbon substrate preferences and ecological niches of *Bathyarchaeota* in estuarine sediments[J]. *Environmental Microbiology*, 2016, 18(4): 1200-1211
- [8] Evans PN, Parks DH, Chadwick GL, et al. Methane metabolism in the archaeal phylum *Bathyarchaeota* revealed by

- genome-centric metagenomics[J]. *Science*, 2015, 350(6259): 434-438
- [9] He Y, Li M, Perumal V, et al. Genomic and enzymatic evidence for acetogenesis among multiple lineages of the archaeal phylum Bathyarchaeota widespread in marine sediments[J]. *Nature Microbiology*, 2016, 1(6): 16035
- [10] Peng Y, Leung HCM, Yiu SM, et al. IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth[J]. *Bioinformatics*, 2012, 28(11): 1420-1428
- [11] Dick GJ, Andersson AF, Baker BJ, et al. Community-wide analysis of microbial genome sequence signatures[J]. *Genome Biology*, 2009, 10(8): R85
- [12] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2[J]. *Nature Methods*, 2012, 9(4): 357-359
- [13] Albertsen M, Hugenholtz P, Skarshewski A, et al. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes[J]. *Nature Biotechnology*, 2013, 31(6): 533-538
- [14] Parks DH, Imelfort M, Skennerton CT, et al. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes[J]. *Genome Research*, 2015, 25(7): 1043-1055
- [15] Hyatt D, Chen GL, LoCascio PF, et al. Prodigal: prokaryotic gene recognition and translation initiation site identification[J]. *BMC Bioinformatics*, 2010, 11: 119
- [16] Lagesen K, Hallin P, Rødland EA, et al. RNAmmer: consistent and rapid annotation of ribosomal RNA genes[J]. *Nucleic Acids Research*, 2007, 35(9): 3100-3108
- [17] Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes[J]. *Nucleic Acids Research*, 2000, 28(1): 27-30
- [18] Tatusov RL, Fedorova ND, Jackson JD, et al. The COG database: an updated version includes eukaryotes[J]. *BMC Bioinformatics*, 2003, 4: 41
- [19] Rawlings ND, Waller M, Barrett AJ, et al. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors[J]. *Nucleic Acids Research*, 2014, 42(D1): D503-D509
- [20] Bagos PG, Tsirigos KD, Plessas SK, et al. Prediction of signal peptides in archaea[J]. *Protein Engineering, Design and Selection*, 2009, 22(1): 27-35
- [21] Zhang WP, Ding W, Yang B, et al. Genomic and transcriptomic evidence for carbohydrate consumption among microorganisms in a cold seep brine pool[J]. *Frontiers in Microbiology*, 2016, 7: 1825
- [22] Anantharaman K, Brown CT, Hug LA, et al. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system[J]. *Nature Communications*, 2016, 7: 13219
- [23] Puigbò P, Wolf YI, Koonin EV. Search for a 'Tree of Life' in the thicket of the phylogenetic forest[J]. *Journal of Biology*, 2009, 8(6): 59
- [24] Katoh K, Misawa K, Kuma KI, et al. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform[J]. *Nucleic Acids Research*, 2002, 30(14): 3059-3066
- [25] Gouy M, Guindon S, Gascuel O. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building[J]. *Molecular Biology and Evolution*, 2010, 27(2): 221-224
- [26] Price MN, Dehal PS, Arkin AP. FastTree 2-approximately maximum-likelihood trees for large alignments[J]. *PLoS One*, 2010, 5(3): e9490
- [27] Fuchs G. Alternative pathways of carbon dioxide fixation: insights into the early evolution of life?[J]. *Annual Review of Microbiology*, 2011, 65: 631-658
- [28] Wakeham SG, Lee C, Hedges JI, et al. Molecular indicators of diagenetic status in marine organic matter[J]. *Geochimica et Cosmochimica Acta*, 1997, 61(24): 5363-5369
- [29] Ueno Y, Yamada K, Yoshida N, et al. Evidence from fluid inclusions for microbial methanogenesis in the early Archaean era[J]. *Nature*, 2006, 440(7083): 516-519