

Research
Deep Matter & Energy—Article

深古菌门的核心代谢功能和热环境起源

冯晓远^a, 王寅焯^a, Rahul Zubin^a, 王风平^{a,b,*}

^a State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China

^b State Key Laboratory of Ocean Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

ARTICLE INFO

Article history:

Received 1 July 2018

Revised 17 October 2018

Accepted 16 January 2019

Available online 22 April 2019

关键词

深古菌

宏基因组

比较基因组

超嗜热适应性

摘要

深古菌门具有复杂的亚群分类,是地球上丰度最高的微生物之一。然而,深古菌的代谢特征和演化历史仍然有待进一步研究。本研究对来自10个不同亚群的15个新获得和36个已经发表的深古菌基因组进行了比较基因组分析,揭示了深古菌门的核心代谢特征,即蛋白质、脂质、芳香族化合物降解,糖酵解途径和Wood-Ljungdahl (WL) 途径。上述核心代谢特征表明深古菌使用乙酰辅酶A作为重要的代谢中间物。此外,部分深古菌亚群还具有不完整的柠檬酸循环、产乙酸途径和硫化物相关的代谢潜能,表明不同亚群具有多样的代谢能力和生态功能。亚群Bathy-21和Bathy-22位于深古菌系统发育树根部,是目前已知最古老的两个深古菌亚群。它们广泛分布于热液(泉)环境中,并编码超嗜热适应性特征的标记基因逆促旋酶(reverse gyrase, *rgy*)。综上,本研究对深古菌门的核心代谢能力进行了系统性研究和总结,并揭示了深古菌热环境起源的演化历史。

© 2019 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. 背景介绍

深古菌曾被称为Miscellaneous Crenarchaeotal Group (MCG),是地球上丰度最高的微生物之一,广泛分布于多种环境中,如海洋、淡水、沉积物、土壤、生物反应器、动物共生环境和深部生物圈中[1–5]。深古菌在地球元素化学循环过程,尤其是碳循环过程中起到重要的作用[6,7]。有研究表明深古菌可能使用乙酰辅酶A作为重要的代谢中间物[4,8],并具有降解胞外蛋白质[2]、芳香族化合物[3,4]和碳水化合物[9,10]的代谢潜能。近期的一项研究表明,深古菌可以在碳固定的同时进行木质素降解,这也是首次在古菌中发现的木质素厌氧降解途径

[11]。此外,某些深古菌亚群还具有烷烃代谢[12]和产乙酸[4]的代谢潜能,而上述两个代谢途径被认为是地球上最古老的生物化学能量产生过程[13]。综上,深古菌可能在不同的微生物群落中具有很高的活性[4],并驱动包含碳、氮和硫在内的多种元素循环过程[9,10]。

基于16S rRNA基因构建的系统发育分析揭示了深古菌复杂的亚群分类(Bathy-1到Bathy-19),而近期的一项研究工作则将其扩展到25个亚群[8]。不同的深古菌亚群具有多样的生理功能和生态位。一项研究表明,某些深古菌亚群受到多次咸水–淡水转换事件的影响[14]。然而,深古菌的生态地位、核心代谢特征、起源与演化历史仍然有待进一步研究。在本研究中,我们首

* Corresponding author.

E-mail address: fengpingw@sjtu.edu.cn (F. Wang).

先对瓜伊马斯盆地的两个样本测序得到的16S rRNA基因序列和公开数据库中的参考序列进行了系统性分析。随后,我们使用这两个样本中得到的深古菌基因组以及公共数据库中共计51个深古菌基因组进行了比较基因组分析。其中,深古菌的核心代谢功能包括蛋白质降解、糖酵解和Wood-Ljungdahl (WL) 途径等。最后,我们通过祖先重建分析,揭示了深古菌热环境起源的演化历史。

2. 材料和方法

2.1. 样本描述

本研究中使用编号为core-10的沉积物样品。此样品采集自2008年10月,位于加利福尼亚湾的瓜伊马斯盆地,由AT15-25航次Alvin dive 4460的插管采样。该沉积物表面为油性,并有白色微生物垫覆盖。样品的温度使用Alvin温度探针进行测量,表层温度为7~10 °C,在沉积物6 cm处达到31 °C。本研究使用深度为4~6 cm (编号M8)和8~11 cm (编号M10)沉积物柱的两个剖面进行DNA提取和测序。DNA提取、宏基因组建库和测序的详细方法请参见文献[4]。

2.2. 宏基因组测序数据的从头组装和分装

宏基因组的原始测序数据使用Sickle v1.33 (<https://github.com/najoshi/sickle>)进行低测序质量碱基的去除,并使用IDBA-UD v1.1.1 [15]进行组装,参数为“mink 52, maxk 92, step 8”。高质量测序数据使用Bowtie v2.2.8 [16]和“-very-sensitive”参数比对到组装片段上,随后使用SAMtools v1.3.1 [17]进行测序深度计算。长度大于5 kb的拼接片段使用MetaBAT v0.32.4 [18]和“-specific”参数,基于其四核苷酸多态性、GC含量和测序深度进行分装,以获得基因组数据。上述获得的基因组使用ESOM v1.1 [19]和mmgenome v0.6.3 [20]进行手动检查。基因组的完整度和污染度使用CheckM v1.0.7 [21]的“lineage_wf”功能进行计算。

除上述基因组外,我们还在公开数据库中收集和下载了44个深古菌参考基因组。这些基因组分别采集于奥胡斯湾沉积物[2]、苏拉特盆地煤层气井[12]、科罗拉多土壤[22]、瓜伊马斯盆地热液沉积物[4,23]、北加州草原[24]、白橡树河口沉积物[9]、胡安德富卡玄武岩层流体[25]以及其他采样点[26]。

该研究的相关序列已经上传到NCBI数据库,编号

为PRJNA418890。

2.3. 16S rRNA 基因序列分析

我们从SILVA 132数据库[27]中收集了3738条长度大于900 nt的深古菌16S rRNA基因序列,并使用CD-HIT [28]和参数“-c 0.97-G 1”将其划分为1547个OTU (operational taxonomic unit)。此外,我们从所有深古菌基因组中使用CheckM [21]预测得到了23条16S rRNA基因序列。我们对上述两种序列,以及作为外源的奇古菌和曙古菌参考序列,使用MAFFT v7.313 [29]进行序列联配 (alignment),使用参数为“-localpair”和“-maxiterate-100”。随后,使用RAxML v8.2.8 [30]的GTRGAMMA模型进行最大似然估计方法的系统发育树构建,bootstrap值为1000次重复计算获得,并使用iTOL [31]进行可视化。其他研究工作中使用的参考序列[1,8]将作为深古菌亚群分类的参考。

2.4. 宏基因组注释

我们使用Prodigal v2.6.3 [32]和“-meta”参数对编码基因进行预测。编码基因的功能注释通过比对NCBI (National Center for Biotechnology Information) NR (non-redundant) 库、KEGG (Kyoto Encyclopedia of Genes and Genomes) 数据库[33]和COG (Clusters of Orthologous Group) 数据库[34]获得,使用参数“-E-values < 10⁻⁵”。此外,这些编码基因还使用同样的参数比对到MEROPS数据库[35]和CAZy数据库[36],以寻找其中肽类和碳水化合物降解的相关基因。胞外蛋白酶使用PRED-SIGNAL [37]进行进一步预测。

2.5. 基于保守基因的系统发育分析

为了确定深古菌基因组的系统发育地位,我们收集了60个深古菌基因组(16个为本研究中新得到的基因组,44个为公开数据库中的参考基因组)和其他古菌的参考基因组。参照之前的研究工作[9],我们使用了16个极少受到水平基因转移影响的保守核糖体蛋白基因序列(L14、L15、L16/L10AE、L18、L2、L22、L24、L3、L4、L5、L6P/L9E、S10、S17、S19、S3、S8) [38],进行系统发育树的构建。每个基因序列使用MAFFT v7.313 [29]进行序列联配,使用参数为“-localpair”和“-maxiterate 100”。上述序列按顺序连接,使用RAxML v8.2.8 [30]的PROTGAMMA模型进行最大似然估计方法的系统发育树构建,bootstrap值为1000次重复计算获

得，并使用iTOL [31]进行可视化。

2.6. 祖先基因组重建

祖先基因组特征使用Count [39]进行重建。我们将Count得到的热适应相关基因的预测结果，如基因功能存在与否、基因拷贝数目等，比对到基于16个保守基因的深古菌系统发育树上，并使用iTOL [31]进行可视化。

3. 结果和讨论

3.1. 深古菌门亚群分类

本研究在瓜伊马斯盆地编号为M8和M10的宏基因组测序数据中获得了16个深古菌基因组（表1）。此外，我们还收集了44个已经公开发表的深古菌参考基因组。通过使用CheckM [21]对上述共计60个基因组中进行预测，共计得到23条16S rRNA基因序列（Appendix A. Table S1）。此外，以97%为相似度阈值，在SILVA数据库中得到1547条16S rRNA基因参考序列。随后，我们对上述16S rRNA基因序列进行系统发育树的构建（图1、Appendix A. Fig. S1），该发育树的拓扑结构和之前的研究工作高度相似[1,4,8,14,40]。此外，两个深古菌亚群Bathy-21和Bathy-22位于深古菌发育树的最根部，和其他亚群形成单系同源。这两个亚群（以及亚群Bathy-23）在近期的一项研究工作中得以命名[8]。相关

的16S rRNA基因序列请参见Appendix A. Table S2。多数深古菌亚群的系统发育树位置相对稳定，但有少数进化支由于参考序列、分析方法和参数的不同，在不同研究中的演化地位仍有差异[1,4,8,14,40]。上述拓扑结构的不稳定性可能是由目前参考基因组的缺乏造成的，这也进一步说明深古菌门内部极其复杂的多样性。

我们还使用16个串联的保守核糖体蛋白序列对本研究得到的16个深古菌基因组和44个参考基因组进行了系统发育树的构建[图1 (b)]。共计54个基因组归类于11个不同的亚群（Table S1），另外6个基因组（ex4484_135、ex4484_205、M10_bin214、M10_bin241、M8_bin163、JdFR-01）由于缺乏16S rRNA基因序列，暂时无法获得亚群分类。和基于16S rRNA基因序列构建的系统发育树相似，Bathy-21和Bathy-22两个亚群位于深古菌的最根部，代表最古老的深古菌类型。亚群Bathy-1, Bathy-3, Bathy-5, Bathy-6, Bathy-8, Bathy-15, Bathy-17在不同方法构建的发育树中位置保持稳定，而亚群Bathy-13和Bathy-23的位置并不一致（Table S1）。考虑到基于保守核糖体蛋白构建的系统发育树具有更高的分辨率和稳定度，我们使用此发育树进行后续分析。

3.2. 深古菌门核心代谢基因组

在60个深古菌基因组中，6个基因组（ex4484_135、ex4484_205、M10_bin214、M10_bin241、M8_bin163、

表1 本研究从瓜伊马斯盆地热液沉积物中得到的基因组基本统计信息表

Lineage	Bin	Assembly size	GC content (%)	Completeness (%)	Contamination (%)	Gene number	Quality ^a
Bathy-15	M10_bin139	1 527 575	58.4	90.42	1.87	1605	High
Bathy-22	M10_bin185	1 310 867	33.4	89.49	0	1462	Medium
Bathy-1	M8_bin1642	1 094 533	40.8	89.25	0.93	1171	Medium
Unclassified	M10_bin241	1 132 007	47.1	88.35	2.02	1214	Medium
Bathy-23	M10_bin26	1 331 265	51.9	88.32	2.80	1390	Medium
Bathy-23	M8_bin162	1 251 171	46.8	85.12	4.67	1223	Medium
Bathy-23	M10_bin1901	1 202 259	44.8	83.96	4.67	1188	Medium
Unclassified	M8_bin163	816 136	43.4	79.44	0.93	897	Medium
Bathy-23	M10_bin1904	1 257 418	42.7	77.57	1.87	1224	Medium
Bathy-1	M10_bin208	907 528	41.1	74.37	1.71	1030	Medium
Unclassified	M10_bin214	597 499	37.2	65.87	0	695	Medium
Bathy-23	M10_bin1903	1 879 918	43.1	92.60	15.26	1833	Low
Bathy-23	M8_bin26	503 563	53.6	46.96	2.34	548	Low
Bathy-23	M10_bin182	597 004	44.3	37.70	0	604	Low
Bathy-21	M10_bin242	497 722	48.8	33.35	0.97	543	Low
Bathy-15	M10_bin1432	474 510	56.3	32.97	0	605	Low

^a High quality: completeness > 90%, contamination < 5%; medium quality: completeness > 50%, contamination < 10%; low quality: completeness < 50% or contamination > 10%.

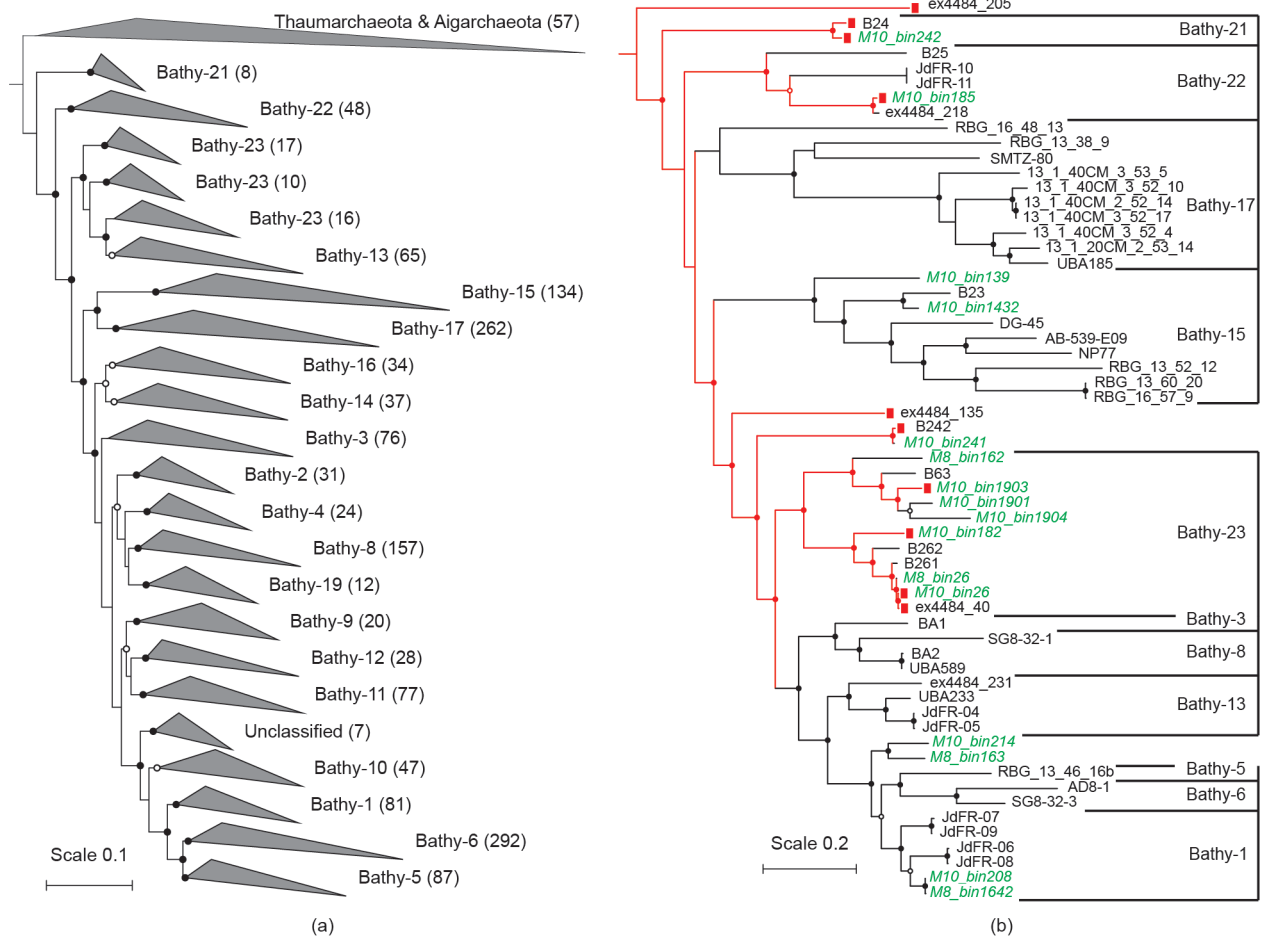


图1. (a) 基于16S rRNA基因构建的系统发育树。每个深古菌亚群以折叠的方式呈现，后面括号中的数字代表该亚群在SILVA数据库中的16S rRNA参考基因和基因组中预测得到的16S rRNA基因总数量。该发育树的完整版请参见Fig. S1。(b) 基于16个核糖体蛋白构建的系统发育树。该发育树使用奇古菌和嗜古菌作为外群（未展示）。本研究得到的基因组以绿色斜体字表示。包含超嗜热适应性标志基因促旋酶的基因组标注为红色矩形，且基于祖先重建分析预测具有超嗜热适应性的祖先节点标记为红色。高于80%和90%的bootstrap值分别使用空心圆和实心圆表示。

JdFR-01) 暂时没有亚群分类。此外，还有3个基因组 (M10_bin1903、DG-45、RBG_13_46_16b) 具有较高的污染度 (10%)。在去除上述9个基因组后，我们使用分属于10个亚群的51个基因组进行深古菌核心和其他代谢功能的分析 (图2、表2)。除亚群Bathy-3外，其余亚群都包含多个基因组。深古菌的核心代谢能力主要包括碎屑蛋白、脂质和芳香族化合物的降解。碎屑蛋白被认为是海洋中丰度最高的有机物之一[41]。所有深古菌亚群都包含蛋白质降解的代谢途径 (Appendix A. Table S3)，其编码的胞外蛋白酶主要属于丝氨酸、半胱氨酸和金属辅酶蛋白酶家族 (Table S3)，其中，枯草杆菌蛋白酶 (S08A)、木瓜蛋白酶 (C01A) 和谷氨酰水解酶 (C26) 在多数深古菌亚群中广泛出现。这和之前的研究结果类似[2,4]，进一步揭示了深古菌可能利用碎屑蛋白和氨基酸作为主要碳源和能源物质。脂质降解和芳香族化合物降解是深古菌的另一种核心代谢特征。除亚群Bathy-21

未发现脂质降解的相关基因，亚群Bathy-3未发现芳香族化合物降解相关的基因之外，其他亚群都编码脂质降解关键基因长链酰基辅酶A合成酶 (*fadD*) 和芳香族化合物降解关键基因酰基磷酸酶 (*acyP*, 表2、Appendix A. Table S4)。这些代谢能力表明深古菌具有降解多种复杂有机化合物的营养策略[11]。

深古菌的能量产生过程主要包括Embden–Meyerhof–Parnas (EMP) 糖酵解途径和Wood–Ljungdahl (WL) 途径 (图2、表2)。多数深古菌基因组都包含几乎完整的从葡萄糖生成乙酰辅酶A的EMP糖酵解途径 (表2、Appendix A. Table S4)。虽然EMP糖酵解途径的最后一步，即催化磷酸烯醇式丙酮酸转化为丙酮酸的丙酮酸激酶 (*pyk*) 仅在6个亚群中出现，其他亚群可能通过磷酸烯醇式丙酮酸羧化酶 (*ppc*) 催化磷酸烯醇式丙酮酸转化为草酰乙酸。

深古菌也可以利用WL途径产生能量。多数深古菌

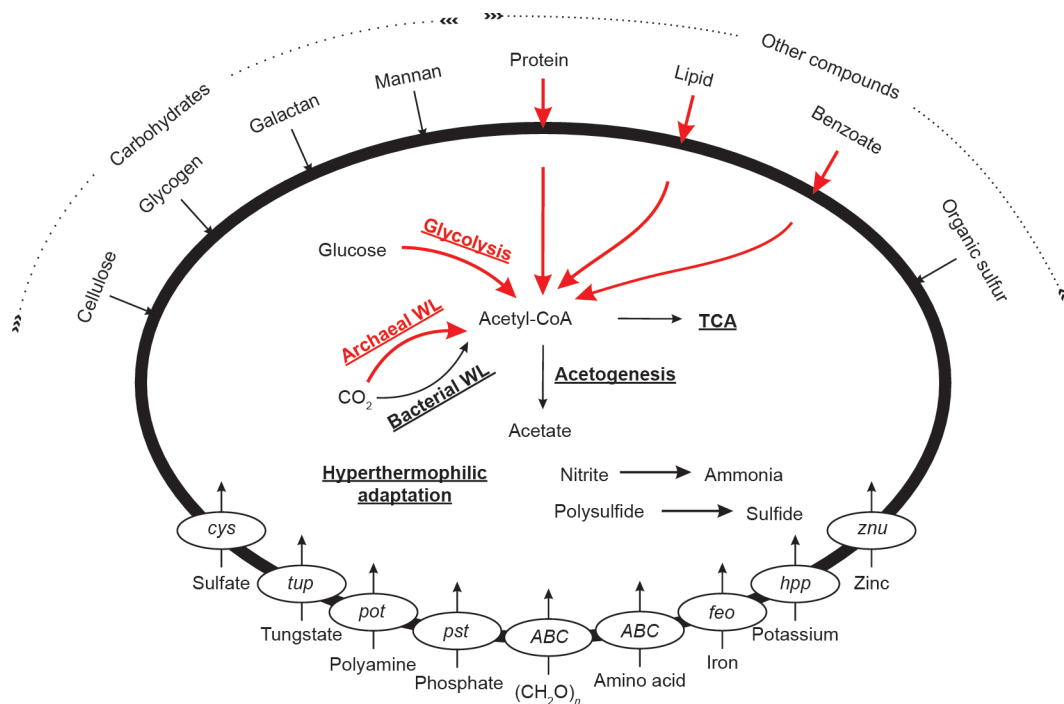


图2. 深古菌代谢特征示意图。核心代谢能力使用红色粗线表示，其他代谢能力使用黑色细线表示。代谢途径使用粗体下划线表示，基因名使用斜体表示，化合物使用正常字体表示。

表2 不同深古菌亚群代谢特征的比较基因组分析

Genomic potential	Subgroups									
	1	3	6	8	13	15	17	21	22	23
Number of MAGs used for metabolism comparative analysis	6	1	2	3	4	8	10	2	5	10
Protein degradation	+	+	+	+	+	+	+	+	+	+
EMP glycolysis pathway	+++	+	+	+	+++	+	+++	+	+	+++
WL pathway	+++	+	+	+	+++	+++	+++	+++	+++	+++
WL pathway (bacterial)	+++					+++	+++			
Lipid degradation	+	+	+	+	+	+	+		+	+
Benzoate degradation	+		+	+	+	+	+	+	+	+
Cellulose degradation	+			+	+	+	+		+	+
Glycogen degradation	+++		+++		+	+	+++			+++
Galactan degradation	+++	+	+	+	+	+	+	+++	+	+++
Mannan degradation	+						+	+		+
Organic sulfur compound degradation	+							+		+++
TCA cycle	+++		+	+		+++	+++		+	
Acetogenesis (acd)	+					+	+	+	+	+
Acetogenesis (ackpta)	+++							+++	+++	+++
Methanogenesis		+		+						
Dissimilatory nitrite reduction	+++			+		+	+		+	+
Sulfidogenesis	+							+		+
Hyperthermophilic adaptation								+++	+	+++

EMP: Embden–Meyerhof–Parnas glycolysis pathway.
+++ : pathway complete; + : pathway near-complete.

亚群都可以在代谢过程中使用自养WL途径，使用二氧化碳作为碳源产生乙酰辅酶A。本研究中分析的10

类亚群都编码古菌类型WL途径的相关基因（表2、Table S4），即通过四氢蝶呤作为一碳载体。此外，亚群

Bathy-1, Bathy-15, Bathy-17还编码细菌类型WL途径的相关基因, 此类途径使用四氢叶酸作为一碳载体。虽然这三类亚群中并未发现催化二氧化碳生成甲酸的甲酸脱氢酶 (*fdh*), 但都编码其他细菌类型WL途径的相关基因, 其中包括: 一氧化碳脱氢酶 (*coo*)、甲酸四氢叶酸连接酶 (*fhs*)、亚甲基四氢叶酸脱氢酶 (*fol*)、亚甲基四氢叶酸还原酶 (*met*) 以及乙酰辅酶A脱羧酶 (*cdh*)。综上, 这三个亚群可能同时使用四氢蝶呤和四氢叶酸作为一碳载体, 用两种类型的WL途径进行二氧化碳固定。

3.3. 深古菌门亚群的其他代谢功能

基于不同亚群基因组信息的代谢预测揭示了每个亚群特有的代谢能力 (图2、表2、Appendix A. Fig. S4), 其中包括碳水化合物利用、三羧酸循环、产乙酸途径和硫化物相关的代谢途径。我们发现不同深古菌亚群对不同碳水化合物具有吸收和降解的偏好 (表2、Table S4), 如纤维素、糖原、半乳糖和甘露聚糖等。完整的三羧酸循环仅在亚群Bathy-1, Bathy-15, Bathy-17中检测到, 其他亚群如Bathy-6, Bathy-8, Bathy-22等则具有不完整的三羧酸循环, 这可能是由于基因组不完整造成的。亚群Bathy-23的所有基因组 (包括污染度高的M10_Bin1903) 都缺乏用于三羧酸循环第一步的柠檬酸合成酶 (*cs*)、第二步的乌头酸水合酶 (*acn*) 以及琥珀酰辅酶A合成酶 (*suc*)。上述结果表明, 深古菌可以利用多种有机物进行异养代谢, 这也是碳限制的深海沉积物微生物常见的生存策略。该发现与之前基于¹³C有机物的稳定同位素探测实验结果相符, 也与不同生态环境中深古菌丰度与总有机碳 (TOC) 含量的相关性结果一致[5,42–44]。

产乙酸代谢是海洋沉积物中有机碳循环相关的重要生物过程[45], 并且在多个深古菌亚群中出现 (表2、Table S4)。亚群Bathy-15, Bathy-17可以通过依赖于ADP的乙酰辅酶A合成酶 (*acd*) 进行磷酸化, 而Bathy-1, Bathy-21, Bathy-22, Bathy-23则可以通过磷酸乙酰转移酶 (*pta*) 和乙酸激酶 (*ack*) 进行产乙酸过程。后两个基因广泛分布在使用氢气和二氧化碳为底物的产乙酸细菌中。然而, 其他亚群 (Bathy-3, Bathy-6, Bathy-8, Bathy-13) 是否具有产乙酸能力仍然需要更多基因组证据进行确认。最近的研究结果表明, 产乙酸途径在瓜伊马盆地沉积物的微生物碳水化合物的利用、发酵和呼吸途径中起到关键作用[23]。在本研究中, 深古菌门的10个亚群中有不少于6个亚群具有潜在的产乙酸能力, 进一步说明深古菌可能在海洋沉积物的碳循环过程中发挥

重要作用。

基因组证据还表明, 不同亚群中具有硫化物相关的特有代谢潜能 (表2、Table S4)。亚群Bathy-1, Bathy-21, Bathy-23含有硫酸盐水解有关的基因, 具有降解有机硫化物的功能。此外, 上述亚群 (以及亚群Bathy-17和Bathy-21) 还含有编码硫化物形成相关的硫氢化酶基因 (*hyd*) 和硫酸盐转运基因 (*cysU*)。上述结果表明, 深古菌在全球硫循环过程中的重要作用是被低估的。

3.4. 深古菌的热环境起源特征

大多数深古菌亚群都编码多种在热休克过程中表达的分子伴侣蛋白 (如热休克蛋白基因 *htpX*、*sHSP*, 表4)[46]。我们发现, 采集自瓜伊马斯盆地热液沉积物的深古菌亚群Bathy-21和Bathy-23并不编码多数亚群所具有的细菌类型DnaK-DnaJ-GrpE伴侣蛋白系统 (Table S4)。上述简化的热休克系统被认为是超嗜热古菌的特征之一[46], 表明这两类深古菌亚群可能具有超嗜热适应性特征。此外, 亚群Bathy-21, Bathy-22, Bathy-23和其他三个未分类的深古菌基因组 (*ex4484_205*, *ex4484_135*, *B242*) 还编码超嗜热适应性的标记基因逆促旋酶 (Table S4)。祖先重建分析[图1 (b)]显示, 深古菌的祖先可能具有逆促旋酶。此外, 这些具有超嗜热性特征的深古菌基因组全部采集自瓜伊马斯盆地热液沉积物。其他深古菌亚群可能在演化过程中丢失了超嗜热适应性特征。基于已有的基因组数据, 我们推测深古菌的祖先起源于热环境中, 具有超嗜热性特征, 并逐渐演化以适应不同环境。然而, 此推论还需要更多的基因组证据进行验证。深古菌门属于TACK超古菌门 (或称“Proteoarchaeota”)。虽然TACK超古菌门的其他谱系, 如初古菌、泉古菌和曙古菌等, 都具有超嗜热特征[47], 但迄今为止还没有发现深古菌的基因组或亚群具有超嗜热特征。综上, 采集自高温沉积物环境的亚群Bathy-21, Bathy-22, Bathy-23可能具有超嗜热特征。此外, 对这些亚群的全球分布调查也表明, 这些亚群大多分布于高温环境, 为上述推论进一步提供了地理证据。

4. 结论

深古菌具有分布广泛、高丰度和复杂多样性的特征, 并通过降解碎屑蛋白和其他有机物等途径在全球地球化学元素循环过程中扮演重要作用[8]。本研究对10

个深古菌门亚群共计51个基因组进行比较基因组分析,发现深古菌的核心代谢特征包括碎屑蛋白、脂质、芳香族化合物降解途径、糖酵解途径和WL碳固定途径,表明深古菌同时具有复杂有机质降解和自养碳固定的代谢潜能。此外,不同深古菌亚群还具有碳水化合物利用、三羧酸循环、产乙酸途径和硫化物相关的代谢功能,表明深古菌还具有多样的代谢能力。此外,我们还发现了编码逆促旋酶和简化的伴侣蛋白系统的深古菌亚群,说明这些亚群具有潜在超高温适应性特征。基于祖先重建分析,深古菌门可能具有热环境起源的演化历史。综上,本研究的结果进一步揭示了深古菌重要的生态特征及其演化起源。

致谢

本工作得到了国家自然科学基金(41525011, 91751205和31661143022)以及地球深部碳观测项目的支持。

Compliance with ethics guidelines

Xiaoyuan Feng, Yinzhao Wang, Rahul Zubin, and Fengping Wang declare that they have no conflict of interest or financial conflicts to disclose.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.eng.2019.01.011>.

References

- [1] Kubo K, Lloyd KG, Biddle JF, Amann R, Teske A, Knittel K. Archaea of the Miscellaneous Crenarchaeotal Group are abundant, diverse and widespread in marine sediments. *ISME J* 2012;6(10):1949–65.
- [2] Lloyd KG, Schreiber L, Petersen DG, Kjeldsen KU, Lever MA, Steen AD, et al. Predominant Archaea in marine sediments degrade detrital proteins. *Nature* 2013;496(7444):215–8.
- [3] Meng J, Xu J, Qin D, He Y, Xiao X, Wang F. Genetic and functional properties of uncultivated MCG Archaea assessed by metagenome and gene expression analyses. *ISME J* 2014;8(3):650–9.
- [4] He Y, Li M, Perumal V, Feng X, Fang J, Xie J, et al. Genomic and enzymatic evidence for acetogenesis among multiple lineages of the archaeal phylum Bathyarchaeota widespread in marine sediments. *Nat Microbiol* 2016;1(6):16035.
- [5] Xiang X, Wang R, Wang H, Gong L, Man B, Xu Y. Distribution of Bathyarchaeota communities across different terrestrial settings and their potential ecological functions. *Sci Rep* 2017;7(1):45028.
- [6] Fry JC, Parkes RJ, Cragg BA, Weightman AJ, Webster G. Prokaryotic biodiversity and activity in the deep seafloor biosphere. *FEMS Microbiol Ecol* 2008;66(2):181–96.
- [7] Li Q, Wang F, Chen Z, Yin X, Xiao X. Stratified active archaeal communities in the sediments of Jiulong River estuary, China. *Front Microbiol* 2012;3:311.
- [8] Zhou Z, Pan J, Wang F, Gu JD, Li M. Bathyarchaeota: globally distributed metabolic generalists in anoxic environments. *FEMS Microbiol Rev* 2018;42(5):639–55.
- [9] Lazar CS, Baker BJ, Seitz K, Hyde AS, Dick GJ, Hinrichs KU, et al. Genomic evidence for distinct carbon substrate preferences and ecological niches of Bathyarchaeota in estuarine sediments. *Environ Microbiol* 2016;18(4):1200–11.
- [10] Zhang W, Ding W, Yang B, Tian R, Gu S, Luo H, et al. Genomic and transcriptomic evidence for carbohydrate consumption among microorganisms in a cold seep brine pool. *Front Microbiol* 2016;7:1825.
- [11] Yu T, Wu W, Liang W, Lever MA, Hinrichs KU, Wang F. Growth of sedimentary Bathyarchaeota on lignin as an energy source. *Proc Natl Acad Sci USA* 2018;115(23):6022–7.
- [12] Evans PN, Parks DH, Chadwick GL, Robbins SJ, Orphan VJ, Golding SD, et al. Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. *Science* 2015;350(6259):434–8.
- [13] Martin WF, Sousa FL, Lane N. Energy at life's origin. *Science* 2014;344(6188):1092–3.
- [14] Fillol M, Auguet JC, Casamayor EO, Borrego CM. Insights in the ecology and evolutionary history of the Miscellaneous Crenarchaeotic Group lineage. *ISME J* 2016;10(3):665–77.
- [15] Peng Y, Leung HC, Yiu SM, Chin FY. IDBA-UD: a *de novo* assembler for singlecell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 2012;28(11):1420–8.
- [16] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9(4):357–9.
- [17] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009;25(16):2078–9.
- [18] Kang DD, Froula J, Egan R, Wang Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 2015;3:e11665.
- [19] Dick GJ, Andersson AF, Baker BJ, Simmons SL, Thomas BC, Yelton AP, et al. Community-wide analysis of microbial genome sequence signatures. *Genome Biol* 2009;10(8):R85.
- [20] Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat Biotechnol* 2013;31(6):533–8.
- [21] Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2015;25(7):1043–55.
- [22] Anantharaman K, Brown CT, Hug LA, Sharon I, Castelle CJ, Probst AJ, et al. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat Commun* 2016;7(1):13219.
- [23] Dombrowski N, Seitz KW, Teske AP, Baker BJ. Genomic insights into potential interdependencies in microbial hydrocarbon and nutrient cycling in hydrothermal sediments. *Microbiome* 2017;5(1):106.
- [24] Butterfield CN, Li Z, Andeer PF, Spaulding S, Thomas BC, Singh A, et al. Proteogenomic analyses indicate bacterial methylotrophy and archaeal heterotrophy are prevalent below the grass root zone. *PeerJ* 2016;4:e2687.
- [25] Jungbluth SP, Glavina del Rio T, Tringe SG, Stepanauskas R, Rappé MS. Genomic comparisons of a bacterial lineage that inhabits both marine and terrestrial deep subsurface systems. *PeerJ* 2017;5:e3134.
- [26] Parks DH, Rinke C, Chuvochina M, Chaumeil PA, Woodcroft BJ, Evans PN, et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* 2017;2(11):1533–42.
- [27] Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* 2007;35(21):7188–96.
- [28] Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the nextgeneration sequencing data. *Bioinformatics* 2012;28(23):3150–2.
- [29] Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 2002;30(14):3059–66.
- [30] Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and postanalysis of large phylogenies. *Bioinformatics* 2014;30(9):1312–3.
- [31] Letunic I, Bork P. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res* 2011;39(Suppl 2):W475–8.
- [32] Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinf* 2010;11(1):119.
- [33] Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 2000;28(1):27–30.
- [34] Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, et al. The COG database: an updated version includes eukaryotes. *BMC Bioinf* 2003;4(1):41.
- [35] Rawlings ND, Waller M, Barrett AJ, Bateman A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* 2014;42(D1):D503–9.
- [36] Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res*

- 2014;42(D1):D490–5.
- [37] Bagos PG, Tsirigos KD, Plessas SK, Liakopoulos TD, Hamdrakas SJ. Prediction of signal peptides in Archaea. *Protein Eng Des Sel* 2009;22(1):27–35.
- [38] Sorek R, Zhu Y, Creevey CJ, Francino MP, Bork P, Rubin EM. Genome-wide experimental determination of barriers to horizontal gene transfer. *Science* 2007;318(5855):1449–52.
- [39] Csurös M. Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics* 2010;26(15):1910–2.
- [40] Lazar CS, Biddle JF, Meador TB, Blair N, Hinrichs KU, Teske AP. 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). *Environ Microbiol* 2015;17(7):2228–38.
- [41] Wakeham SG, Lee C, Hedges JI, Hernes PJ, Peterson MJ. Molecular indicators of diagenetic status in marine organic matter. *Geochim Cosmochim Acta* 1997;61 (24):5363–9.
- [42] Heuer VB, Pohlman JW, Torres ME, Elvert M, Hinrichs KU. The stable carbon isotope biogeochemistry of acetate and other dissolved carbon species in deep subseafloor sediments at the northern Cascadia Margin. *Geochim Cosmochim Acta* 2009;73(11):3323–36.
- [43] Seyler LM, McGuinness LM, Kerkhof LJ. Crenarchaeal heterotrophy in salt marsh sediments. *ISME J* 2014;8(7):1534–43.
- [44] Yu T, Liang Q, Niu M, Wang F. High occurrence of Bathyarchaeota (MCG) in the deep-sea sediments of South China Sea quantified using newly designed PCR primers. *Environ Microbiol Rep* 2017;9(4):374–82.
- [45] Lever MA, Heuer VB, Morono Y, Masui N, Schmidt F, Alperin MJ, et al. Acetogenesis in deep subseafloor sediments of the Juan de Fuca Ridge Flank: a synthesis of geochemical, thermodynamic, and gene-based evidence. *Geomicrobiol J* 2010;27(2):183–211.
- [46] Richter K, Haslbeck M, Buchner J. The heat shock response: life on the verge of death. *Mol Cell* 2010;40(2):253–66.
- [47] Spang A, Caceres EF, Ettema TJG. Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. *Science* 2017;357(6351):eaaf3883.