

Research
Deep Matter & Energy—Article

Core Metabolic Features and Hot Origin of Bathyarchaeota

Xiaoyuan Feng^a, Yinzhaoh Wang^a, Rahul Zubin^a, Fengping Wang^{a,b,*}

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

^bState 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

Keywords:

Bathyarchaeota

Metagenomics

Comparative genomics

Hyperthermophilic adaptation

ABSTRACT

The archaeal phylum Bathyarchaeota comprises highly diversified subgroups and is considered to be one of the most abundant microorganisms on earth. The metabolic features and evolution of this phylum still remain largely unknown. In this article, a comparative metabolic analysis of 15 newly reconstructed and 36 published metagenomic assembled genomes (MAGs) spanning 10 subgroups was performed, revealing the core metabolic features of Bathyarchaeota—namely, protein, lipid, and benzoate degradation; glycolysis; and the Wood–Ljungdahl (WL) pathway, indicating an acetyl–CoA–centralized metabolism within this phylum. Furthermore, a partial tricarboxylic acid (TCA) cycle, acetogenesis, and sulfur-related metabolic pathways were found in specific subgroups, suggesting versatile metabolic capabilities and ecological functions of different subgroups. Intriguingly, most of the MAGs from the Bathy-21 and -22 subgroups, which are placed at the phylogenetic root of all bathyarchaeotal lineages and likely represent the ancient Bathyarchaeota types, were found in hydrothermal environments and encoded reverse gyrase, suggesting a hyperthermophilic feature. This work reveals the core metabolic features of Bathyarchaeota, and indicates a hot origin of this archaeal phylum.

© 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. Introduction

The archaeal phylum Bathyarchaeota (formerly known as Miscellaneous Crenarchaeotal Group (MCG)) is one of the most abundant and ubiquitously distributed microorganisms living in diverse habitats such as marine/freshwater sediment, soil, bioreactor, animal-associated habitats, and the deep subsurface biosphere [1–5]. They are considered to actively participate in global geochemical cycling, particularly in the transformations of carbon [6,7]. Several studies have suggested an acetyl–CoA–centralized heterotrophic lifestyle in the Bathyarchaeota [4,8], and have found that they hold the potential to degrade proteins [2] and aromatic compounds [3,4], and to hydrolyze extracellular carbohydrates [9,10]. A recent breakthrough showed that Bathyarchaeota were able to utilize lignin while fixing carbon dioxide (CO₂); this finding was the first evidence for anaerobic lignin degradation in the domain Archaea [11]. Furthermore, several bathyarchaeotal lineages have been suggested to have potential for methane metabolism [12] and acetogenesis [4], two of the most fundamental and ancient microbial biochemical energy-conservation processes

[13]. The above mentioned metabolic pathways of Bathyarchaeota may actively interact with local microbes [4] and drive the cycling of other elements such as nitrogen and sulfur [9,10].

Phylogenetic analysis based on 16S rRNA gene sequences revealed a large intragroup diversity (Bathy-1 to -19), which was recently enlarged further to 25 subgroups [8]. These subgroups have versatile physiological capabilities and possess distinct ecological niches [1,14]. Previous research has demonstrated that the diversification of Bathyarchaeota subgroups may have been influenced by several saline–freshwater transition events [14]. However, the ecological role, origin, and evolution of Bathyarchaeota are still far from fully understood. It remains unclear whether this diverse archaeal phylum contains core metabolic features, and how this phylum originates and evolves. In this study, we analyzed bathyarchaeotal 16S rRNA genes from two newly sequenced Guaymas Basin sediment samples and from public databases. With new binned and downloaded bathyarchaeotal metagenomic assembled genomes (MAGs), a comprehensive comparative analysis of 51 out of 60 MAGs spanning 10 subgroups was performed. The results indicate that the phylum Bathyarchaeota shares a core set of metabolic pathways, including protein degradation, glycolysis, and the reductive acetyl–CoA (Wood–Ljungdahl, WL) pathway. Furthermore, our data suggests a hot origin for

* Corresponding author.

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

Bathyarchaeota based on the metabolic potentials of ancestral genomes.

2. Materials and methods

2.1. Sample description

The sediment sample core-10 was collected by means of push core at a deep-sea vent site at the Guaymas Basin in the Gulf of California in October 2008, during Alvin dive 4460, which was part of cruise AT15-25. The sediment used in this study was oily and covered with a white microbial mat. The temperature, as measured by the Alvin temperature probe, increased from 7 to 10 °C at the surface to 31 °C at 6 cm below the surface (cmbsf). Two sediment layers, at 4–6 and 8–11 cmbsf, were designated as M8 and M10, respectively, and used for DNA isolation. The methods used for DNA extraction, metagenome library construction, and sequencing are described in Ref. [4].

2.2. De novo assembly and binning of metagenomic sequences

The raw reads from metagenomic sequencing were trimmed with Sickle v1.33 (<https://github.com/najoshi/sickle>) and assembled using IDBA-UD v1.1.1 [15] with the following parameters: “mink 52, maxk 92, step 8.” The coverage was determined by recruiting raw reads to assembled contigs using Bowtie v2.2.8 [16] with the parameter “–very-sensitive,” and was calculated using SAMtools v1.3.1 [17]. Binning of assembled sequences longer than 5 kb was performed using tetra-nucleotide frequencies, guanine-cytosine (GC) content, and sequencing coverages using MetaBAT v0.32.4 with the parameter “–specific” [18]. The retrieved MAGs were further checked manually using ESOM v1.1 [19] and mmgenome package v0.6.3 [20]. The completeness and contamination of each MAG were estimated using CheckM v1.0.7 with the “lineage_wf” function [21].

In addition to the MAGs retrieved from the present study, 44 reference bathyarchaeotal genomes were downloaded and used in the current analysis. These MAGs were retrieved from samples from Aarhus Bay sediment [2], Surat Basin coal-bed methane wells [12], Colorado soil samples [22], Guaymas Basin hydrothermal sediments [4,23], Northern California grassland [24], White Oak River estuarine sediments [9], Juan de Fuca basalt-hosted fluids [25], and other globally distributed sample sites [26].

This Whole Genome Shotgun project has been deposited at GenBank under the accession number PRJNA418890.

2.3. 16S rRNA gene sequences analysis

A total of 3738 reference bathyarchaeotal sequences longer than 900 nt were collected from the SILVA 132 database [27] and clustered into 1547 operational taxonomic units (OTUs) using CD-HIT [28] with the parameter “–c 0.97–G 1.” Another 23 16S rRNA gene sequences were predicted from the MAGs mentioned above using CheckM [21]. All of these 16S rRNA gene sequences, together with reference sequences from Thaumarchaeota and Aigarchaeota which were used as outgroups, were aligned using MAFFT v7.313 [29] with the parameters “–localpair” and “–maxiterate 100.” The phylogenetic tree was generated using maximum likelihood analysis in RAxML v8.2.8 [30] with the GTRGAMMA model, and visualized using Interactive Tree Of Life (iTOL) [31]. Affiliations of bathyarchaeotal sequences to each subgroup were carried out using reference sequences [1,8] as phylogenetic anchors.

2.4. Metagenome annotation

Protein coding sequences (CDSs) were determined using Prodigal v2.6.3 [32] with the parameter “–meta.” Functional information for each CDS was collected by comparing amino acid sequences against the National Center for Biotechnology Information (NCBI) “non-redundant” (nr), Kyoto Encyclopedia of Genes and Genomes (KEGG) [33], and Clusters of Orthologous Group (COG) databases [34] with E -values $< 10^{-5}$, and was inspected manually. Those predicted CDSs were further searched against the MEROPS [35] and CAZy [36] databases with E -values $< 10^{-5}$ to find genes related to the degradation of peptides and carbohydrates. The extracellular peptidases were further confirmed based on the identification of transport signals using PRED-SIGNAL [37].

2.5. Phylogenetic analysis based on conserved genes

To identify relevant phylogenetic homologs of bathyarchaeotal MAGs, a total of 60 bathyarchaeotal genomes (16 from the present study and 44 from public databases) were included in the phylogenetic analysis, along with other archaeal reference genomes. Following a previous study [9], 16 conserved ribosomal genes (ribosomal protein L14, L15, L16/L10AE, L18, L2, L22, L24, L3, L4, L5, L6P/L9E, S10, S17, S19, S3, and S8) were considered to undergo no horizontal gene transferring [38], and were used to infer the phylogenetic affiliation of the inspected bathyarchaeotal MAGs. Homologous sequences were individually aligned using MAFFT v7.313 [29] with the parameters “–localpair” and “–maxiterate 100,” and then concatenated in series. The phylogenetic trees were generated using maximum likelihood analysis in RAxML v8.2.8 [30] with the PROTGAMMA model, and visualized using iTOL [31].

2.6. Reconstruction of ancestral genomes

Ancestral genome reconstruction was implemented in the Count software [39]. The phyletic patterns (gene family presence/absence and gene copy number) of hydrothermal-adaptation-related genes were mapped to the bathyarchaeotal phylogeny based on the 16 concatenated conserved ribosomal protein sequences mentioned above. The phyletic pattern was visualized using iTOL [31] and plotted manually.

3. Results and discussion

3.1. Bathyarchaeotal subgroup classification

In total, 16 bathyarchaeotal MAGs were recovered from the Guaymas Basin M8 and M10 metagenomic datasets (Table 1). Together with 44 published bathyarchaeotal genomes, 23 16S rRNA gene sequences were predicted (Appendix A. Table S1) using CheckM [21]. Furthermore, an additional 1547 representative bathyarchaeotal 16S rRNA gene sequences with a maximum 97% identity cutoff were downloaded from the SILVA databases and used as references. A 16S rRNA gene phylogenetic tree was constructed (Fig. 1(a) and Appendix A. Fig. S1), whose topology was highly similar to and coherent with results from previous studies [1,4,8,14,40]. To be specific, Bathy-21 and -22, which were located at the base of the tree, branched monophyletically from the other subgroups. These two subgroups, together with subgroup Bathy-23, were recently named by Zhou et al. [8]. The related 16S rRNA gene sequences are listed in Appendix A. Table S2. The phylogeny for most bathyarchaeotal subgroups are relatively stable, with variations in only a few branches' orders still remaining due to the use of different reference sequences, methods, and parameters in different studies [1,4,8,14,40]. Instability in the tree topologies is

Table 1
General characteristics of the MAGs retrieved in the present study from Guaymas Basin hydrothermal sediments.

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	1 605	High
Bathy-22	M10_bin185	1 310 867	33.4	89.49	0	1 462	Medium
Bathy-1	M8_bin1642	1 094 533	40.8	89.25	0.93	1 171	Medium
Unclassified	M10_bin241	1 132 007	47.1	88.35	2.02	1 214	Medium
Bathy-23	M10_bin26	1 331 265	51.9	88.32	2.80	1 390	Medium
Bathy-23	M8_bin162	1 251 171	46.8	85.12	4.67	1 223	Medium
Bathy-23	M10_bin1901	1 202 259	44.8	83.96	4.67	1 188	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	1 224	Medium
Bathy-1	M10_bin208	907 528	41.1	74.37	1.71	1 030	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	1 833	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%.

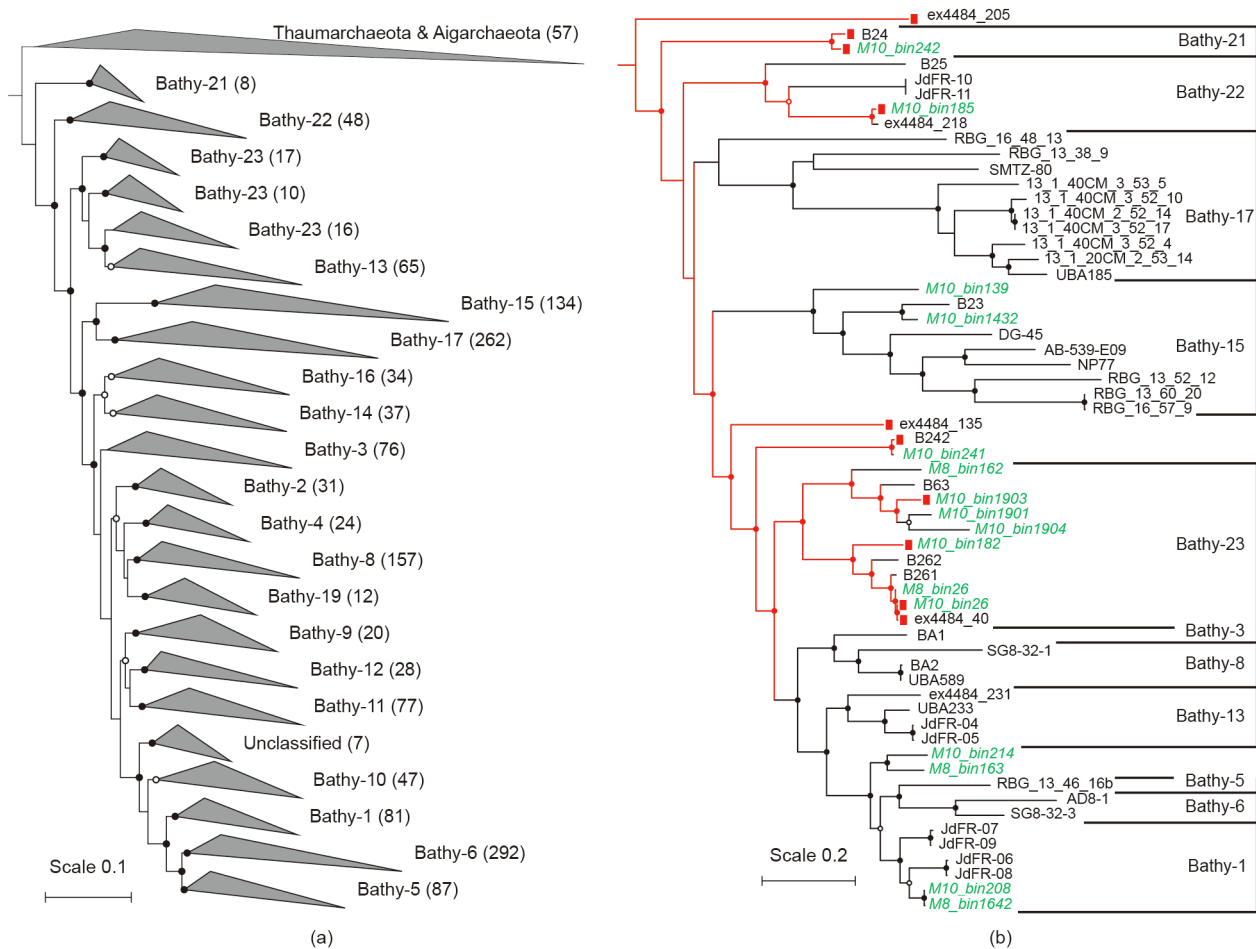


Fig. 1. (a) Phylogenetic analysis based on 16S rRNA gene sequences. Each bathyarchaeotal subgroup was cladded, followed by the total 16S rRNA gene sequences number of SILVA references and those predicted from the MAGs. For a full version of the uncollapsed phylogeny, please refer to Fig. S1. (b) Phylogenomic analysis based on 16 concatenated ribosomal proteins. This phylogeny is rooted by the phyla Thaumarchaeota and Aigarchaeota (not shown). MAGs retrieved from the present study are colored in green with italics. MAGs containing hyperthermophilic adaptation marker gene reverse gyrase are marked with a red rectangle, and ancestor nodes with hyperthermophilic adaptation ability, based on ancestral genome analysis, are colored in red. Bootstrap values for these phylogenies are shown with open ($\geq 80\%$) and filled ($\geq 90\%$) circles.

probably due to insufficient sequences for phylogenetic analyses, indicating the extremely high diversity of the phylum Bathyarchaeota.

A phylogenomic tree of the MAGs from the current study (16) and public database (44) was further constructed using 16 concatenated conserved ribosomal protein sequences (see Section 2.5

and Fig. 1(b)). A total of 54 MAGs were classified into 11 different subgroups (Table S1), whereas six MAGs (ex4484_135, ex4484_205, M10_bin214, M10_bin241, M8_bin163, and JdFR-01) were categorized as unclassified due to the lack of 16S rRNA gene sequences. Similar to the phylogenetic analysis based on 16S rRNA gene sequences, the subgroups Bathy-21 and Bathy-22 are still located at the root of the phylogenomic tree, which suggests that these two lineages may represent the ancient Bathyarchaeota types. Subgroups 1, 3, 5, 6, 8, 15, and 17 have corresponding genomes and show similar topology with the 16S rRNA gene phylogenetic tree. Nevertheless, the phylogeny for subgroups 13 and 23 was not matched between the 16S rRNA-based and concatenated ribosomal protein-based phylogenomic analyses (Table S1). Considering the higher resolution and stability of the phylogenetic analysis based on concatenated conserved ribosomal protein sequences compared with 16S rRNA genes, the phylogenomic tree was used for further analysis.

3.2. Core metabolic properties of Bathyarchaeota

Among 60 bathyarchaeotal MAGs, six MAGs (ex4484_135, ex4484_205, M10_bin214, M10_bin241, M8_bin163, and JdFR-01) remained in an unclassified subgroup, while three MAGs (M10_bin1903, DG-45, and RBG_13_46_16b) contained potentially high contamination > 10%; these nine MAGs were not used for the core and specific metabolic analysis of different subgroups. The remaining 51 phylogenetically affiliated MAGs formed 10 bathyarchaeotal subgroups, and their metabolic potentials were systematically compared (Fig. 2 and Table 2). While Bathy-3 had only one representative MAG, other subgroups contained two or more MAGs. Shared core metabolic features among different lineages were discovered, including the pathways for the utilization of detrital proteins, lipids, and benzoates. Detrital proteins are considered to be one of the most abundant components of marine organic matter [41]. Pathways involved in the degradation of proteins were present in all bathyarchaeotal subgroups (Appendix A.

Table S3). The detected extracellular peptidases mainly belonged to the serine, cysteine, and metallo peptidase families (Table S3), while subtilisin (S08A), papain (C01A), and gamma-glutamyl hydrolase (C26) existed in most bathyarchaeotal subgroups. These observations further agreed with previous studies [2,4], suggesting that detrital proteins and/or amino acids are used as major carbon and energy sources for Bathyarchaeota. The degradation of lipids and benzoates was another core genomic feature. Except for one subgroup (Bathy-21 for lipids and Bathy-3 for benzoates), the majority of the bathyarchaeotal subgroups contained long-chain acyl-CoA synthetase (*fadD*) and acylphosphatase (*acyP*, Table 2 and Appendix A. Table S4), the key genes for lipid and benzoate degradation, respectively. In general, these metabolic potentials indicate the capability of Bathyarchaeota to degrade diverse complex organic compounds for an organoautotrophic life strategy, as was recently suggested [11].

The processes for energy conservation were mostly identified as the Embden–Meyerhof–Parnas (EMP) glycolysis pathway and the reductive acetyl-CoA (Wood–Ljungdahl, WL) pathway (Fig. 2 and Table 2). Most of the Bathyarchaeota genomes contain near-complete enzymes for the utilization of glucose to produce acetyl-CoA by the glycolysis process (Table 2 and Table S4). Although the pyruvate kinase gene (*pyk*), which converts phosphoenolpyruvate (PEP) into pyruvate in the last step of EMP, was found in only six out of 10 subgroups, another gene encoding phosphoenolpyruvate carboxylase (*ppc*), which converts PEP into oxaloacetate, was present and could be used for an alternative pathway.

Bathyarchaeota can also use the WL pathway to generate energy. In nearly all subgroups, most of the carbon metabolic pathways lead to acetyl-CoA production—including the autotrophic WL pathway which uses CO₂ as the carbon source. All the analyzed subgroups possessed an archaeal-type WL pathway for autotrophic inorganic carbon fixation using tetrahydromethanopterin (H₄MPT) as the C₁-carrier, which is prevalently found in Archaea (Table 2 and Table S4). Interestingly, a bacterial type of the WL pathway that uses tetrahydrofolate as an intermediate was also found in

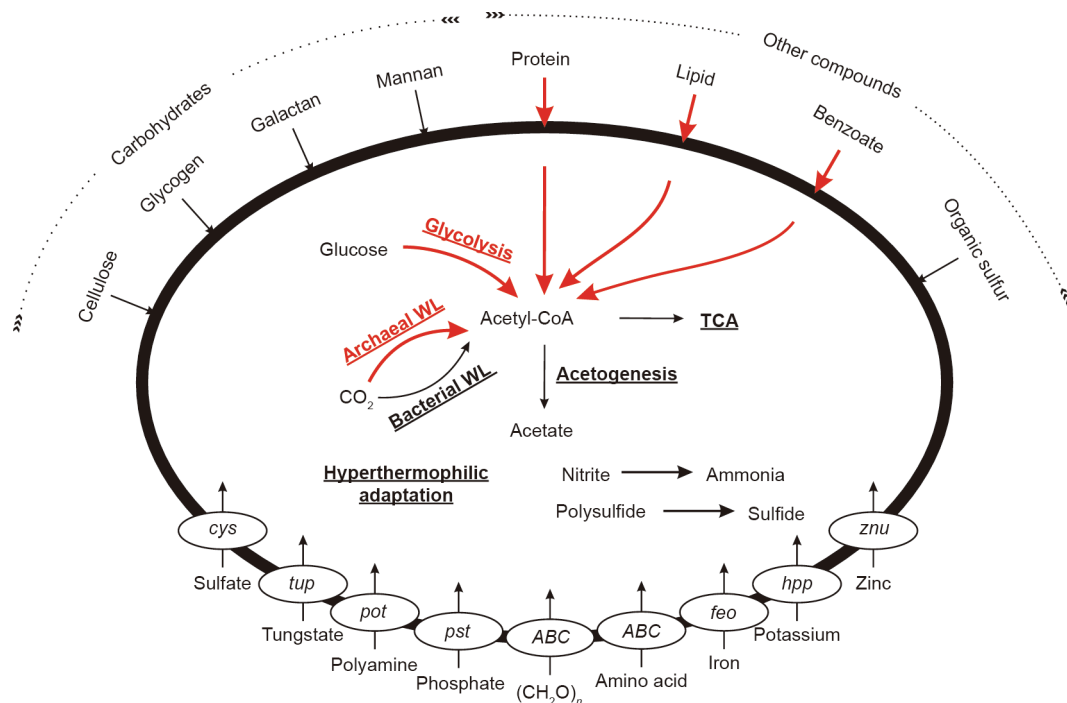


Fig. 2. Schematic representation of the metabolic potential of the phylum Bathyarchaeota. The core metabolic features among the lineages are colored in red with bold lines, while accessory metabolic features are shown in black lines. Metabolic and adaptive pathways are shown in bold and underlined, gene names are shown in italic, and compounds are shown in normal font. TCA: tricarboxylic acid.

Table 2
Comparative analysis of genomic potential in different bathyarchaeotal subgroups.

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.

nearly all genomes from the Bathy-1, -15, and -17 subgroups. Although gene encoding formate dehydrogenase (*fdh*), which converts CO₂ to formate, was not found in these three subgroups, all other genes essential for the bacterial WL pathway were present, including carbon-monoxide dehydrogenase (*coo*), formate–tetrahydrofolate ligase (*fhs*), methylenetetrahydrofolate dehydrogenase (*fol*), methylenetetrahydrofolate reductase (*met*), and acetyl-CoA decarbonylase (*cdh*). Therefore, these three subgroups might contain unidentified enzymes for the conversion of CO₂, using both H₄MPT and tetrahydrofolate as the C₁-carrier.

3.3. Specific metabolic properties of bathyarchaeotal subgroups

Metabolic predictions based on the genomic information of different subgroups inferred group-specific metabolic properties that are not commonly shared in all lineages (Fig. 2, Table 2, and Table S4); these pathways include carbohydrates utilization, the tricarboxylic acid (TCA) cycle, acetogenesis, and sulfur-related metabolism. Different preferences of bathyarchaeotal subgroups for uptaking and degrading a wide variety of carbohydrates were found (Table 2 and Table S4), including cellulose, glycogen, galactan, and mannan. A complete TCA cycle was only detected in the subgroups Bathy-1, -15, and -17; other subgroups such as Bathy-6, -8, and -22 possessed incomplete TCA cycles, probably due to the incompleteness of the genomes. All 10 MAGs belonging to Bathy-23 (as well as the highly contaminated MAG M10_bin1903) lacked the citrate synthase (*cs*) and aconitate hydratase (*acn*) used in the first and second step of the TCA cycle, as well as succinyl-CoA synthetase (*suc*), which converts succinyl-CoA into succinate. The results above suggest that these widespread Archaea might be heterotrophs with extensive abilities to efficiently use organic matter—a survival strategy in carbon-starved deep marine sediments. This suggestion is consistent with the results of previous stable isotopic probing experiments on ¹³C-organic compounds, as well as with the correlation between bathyarchaeotal abundance and the content of total organic carbon (TOC) in diverse habitats [5,42–44].

Acetogenesis is an important feature shared by some bathyarchaeotal subgroups (Table 2 and Table S4): a biological process in organic carbon cycling in subsurface sediments [45]. Subgroups Bathy-15 and -17 have the potential to carry out phosphorylation catalyzed by adenosine diphosphate (ADP) forming acetyl-CoA

synthase (*acd*), while Bathy-1, -21, -22, and -23 can also perform acetogenesis via phosphate acetyltransferase (*pta*) and acetate kinase (*ack*). The latter two genes were found to be widely distributed in bona-fide bacterial homoacetogens, which produce acetate solely from hydrogen and CO₂. However, further genomic evidence is still required to answer the question of whether Bathy-3, -6, -8, and -13 have acetogenesis ability. Recent findings suggest that acetate is considered to play a crucial role in mediating the microbial carbohydrate utilization, fermentation and respiratory pathways in the Guaymas Basin sediments [23]. In the present study, no less than six out of 10 subgroups of Bathyarchaeota were considered to be potential acetogens, and may play a vital role in carbon cycling in marine sediments.

The genomic evidence also indicated diverse metabolic potentials in different subgroups regarding sulfur metabolic pathways (Table 2 and Table S4). Bathy-1, -21, and -23 contained genes related to the hydrolysis of sulfate esters, which is responsible for the decomposition of organic sulfur compounds. It is notable that these subgroups also contained genes encoding the sulfhydrogenase gene (*hyd*, also found in Bathy-17 and -21) involved in sulfide formation from polysulfide (sulfidogenesis), and the sulfate transport gene (*cysU*). These results suggest a formerly ignored role of Bathyarchaeota in global sulfur cycling.

3.4. Hyperthermophilic origin of Bathyarchaeota

The majority of the bathyarchaeotal subgroups encoded diverse molecular chaperones, such as heat-shock protein (*hspX*) and small heat-shock proteins (*shsp*, Table S4), which were previously considered to express during the heat-shock process [46]. Interestingly, Bathy-21 and -23, which were retrieved from Guaymas Basin hydrothermal sediments, were devoid of the bacterial-type DnaK-DnaJ-GrpE chaperone system that is prevalent in other subgroups (Table S4). This simplified heat-shock system was previously reported to be one of the characteristics of hyperthermophilic Archaea [46], and thus indicates a potential hyperthermophilic feature of these bathyarchaeotal subgroups. Genes encoding reverse gyrase, a marker gene for hyperthermophiles, were detected in Bathy-21, -22, and -23 (Table S4), and in three unclassified bathyarchaeotal MAGs (ex4484_205, ex4484_135, and B242). The ancestral genome analysis (Fig. 1(b)) showed that the putative Bathyarchaeota ancestral lineage

possessed the reverse gyrase gene. However, the hyperthermophilic bathyarchaeotal MAGs were exclusively retrieved from Guaymas Basin hydrothermal sediment. Other bathyarchaeotal subgroups may have lost the hyperthermophilic ability during evolutionary adaptation to mesothermal environments. Based on existing bathyarchaeotal MAGs, we hypothesize that the ancient Bathyarchaeota type might be hyperthermophiles, which gradually adapted and evolved into ambient environments. However, more bathyarchaeotal genomic data are needed to further confirm this assumption. The archaeal phylum Bathyarchaeota belongs to the TACK superphylum (or “Proteoarchaeota”). Although many lineages from the TACK superphylum (e.g., Korarchaeota, Crenarchaeota, and Aigarchaeota) have been reported to be hyperthermophilic [47], no bathyarchaeotal MAGs or subgroups have been found to be potentially hyperthermophilic to date. The above results indicated that the subgroups Bathy-21, -22, and -23, which were retrieved from sediments with a high temperature, may be hyperthermophiles. Furthermore, a global survey of the distribution of these subgroups indicated that the majority of these subgroups are found in high-temperature environments (Table S2), which provides geographical evidence of the hyperthermophilic nature of these bathyarchaeotal subgroups.

4. Conclusion

Bathyarchaeota is of great interest to microbial ecologists for its wide distribution, high abundance, and diversity, as well as its potential ability to degrade detrital organic matter in aquatic environments and drive global elements cycling [8]. In this study, a comparative genomic analysis was performed using 51 phylogenetically affiliated MAGs spanning 10 bathyarchaeotal subgroups. Degradation of detrital proteins, lipids, and benzoates, as well as the EMP glycolysis pathway and the WL pathway, were identified in most bathyarchaeotal subgroups, suggesting a metabolic potential relying on both heterotrophic degradation and autotrophic carbon fixation. Other pathways, including carbohydrates utilization, the TCA cycle, acetogenesis, and sulfur-related metabolism, were also detected, indicating diverse and flexible metabolic potentials among different bathyarchaeotal subgroups. Intriguingly, we also found some bathyarchaeotal subgroups that encoded reverse gyrase and/or simplified chaperones, demonstrating putative hyperthermophilic features. Based on the ancestral analysis, the phylum Bathyarchaeota is suggested to have a hot origin. Taken together, these findings are further steps toward the elucidation of the origin, evolution, and roles of Bathyarchaeota, a globally important archaeal phylum.

Acknowledgements

This work was financially supported by the National Natural Science Foundation of China (41525011, 91751205, and 31661143022), and the Deep Carbon Observatory project.

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 single-cell 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:e1165.
- [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 next-generation 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 post-analysis 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, Hamodrakas 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 JL, 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 seafloor 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 seafloor 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.