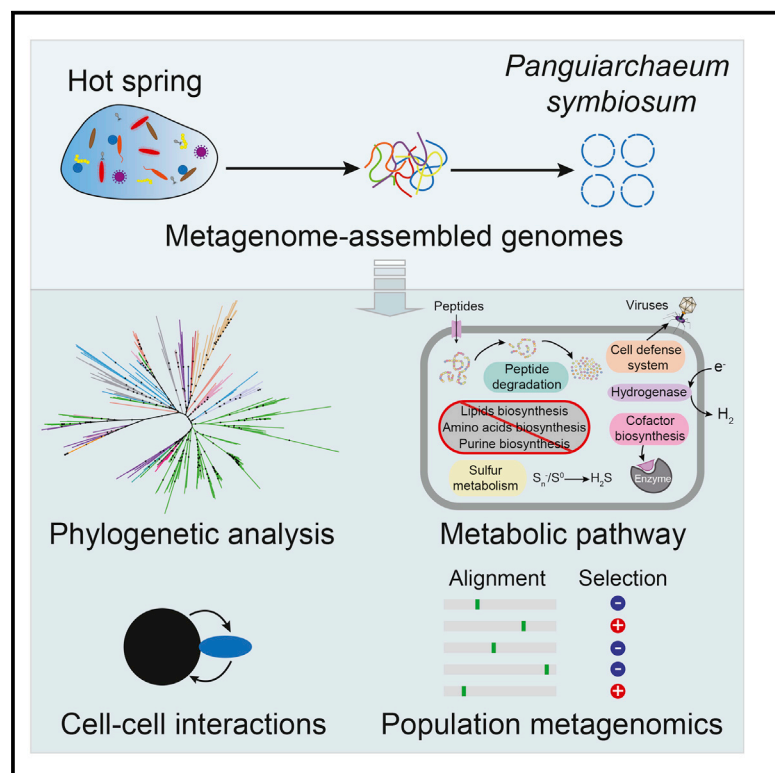


Panguiarchaeum symbiosum, a potential hyperthermophilic symbiont in the TACK superphylum

Graphical abstract



Authors

Yan-Ni Qu, Yang-Zhi Rao, Yan-Ling Qi, ..., Guo-Xing Nie, Zheng-Shuang Hua, Wen-Jun Li

Correspondence

hzhengsh@ustc.edu.cn (Z.-S.H.),
liwenjun3@mail.sysu.edu.cn (W.-J.L.)

In brief

Qu et al. report 10 closely related genomes from hot springs with the name *Panguiarchaeum symbiosum* proposed. By focusing on ecological niches and evolution, they reveal a symbiotic lifestyle of these microbes, where genes associated with nutrient acquisition and cell attachment are prone to be positively selected and highly transcribed.

Highlights

- *Panguiarchaeum symbiosum* belongs to a previously unknown order within the *Korarchaeia*
- *P. symbiosum* has a symbiotic lifestyle by fermenting peptides and reducing sulfur
- Genes related to proteolytic catabolism and cell attachment exhibit positive selection
- Genes related to proteolytic catabolism and cell attachment are highly expressed



Article

Panguiarchaeum symbiosum, a potential hyperthermophilic symbiont in the TACK superphylum

Yan-Ni Qu,¹ Yang-Zhi Rao,¹ Yan-Ling Qi,² Yu-Xian Li,² Andrew Li,² Marike Palmer,³ Brian P. Hedlund,^{3,4} Wen-Sheng Shu,⁵ Paul N. Evans,⁶ Guo-Xing Nie,⁷ Zheng-Shuang Hua,^{2,9,*} and Wen-Jun Li^{1,8,*}

¹State Key Laboratory of Biocontrol, Guangdong Provincial Key Laboratory of Plant Resources and Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), School of Life Sciences, Sun Yat-Sen University, Guangzhou 510275, PR China

²Chinese Academy of Sciences Key Laboratory of Urban Pollutant Conversion, Department of Environmental Science and Engineering, University of Science and Technology of China, Hefei 230026, China

³School of Life Sciences, University of Nevada Las Vegas, Las Vegas, NV 89154, USA

⁴Nevada Institute of Personalized Medicine, University of Nevada Las Vegas, Las Vegas, NV 89154, USA

⁵School of Life Sciences, South China Normal University, Guangzhou 510631, PR China

⁶The Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, University of Queensland, St Lucia, QLD 4072, Australia

⁷College of Fisheries, Henan Normal University, Xinxiang, China

⁸State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, PR China

⁹Lead contact

*Correspondence: hzhengsh@ustc.edu.cn (Z.-S.H.), liwenjun3@mail.sysu.edu.cn (W.-J.L.)

<https://doi.org/10.1016/j.celrep.2023.112158>

SUMMARY

The biology of *Korarchaeia* remains elusive due to the lack of genome representatives. Here, we reconstruct 10 closely related metagenome-assembled genomes from hot spring habitats and place them into a single species, proposed herein as *Panguiarchaeum symbiosum*. Functional investigation suggests that *Panguiarchaeum symbiosum* is strictly anaerobic and grows exclusively in thermal habitats by fermenting peptides coupled with sulfide and hydrogen production to dispose of electrons. Due to its inability to biosynthesize archaeal membranes, amino acids, and purines, this species likely exists in a symbiotic lifestyle similar to DPANN archaea. Population metagenomics and metatranscriptomic analyses demonstrated that genes associated with amino acid/peptide uptake and cell attachment exhibited positive selection and were highly expressed, supporting the proposed proteolytic catabolism and symbiotic lifestyle. Our study sheds light on the metabolism, evolution, and potential symbiotic lifestyle of *Panguiarchaeum symbiosum*, which may be a unique host-dependent archaeon within the TACK superphylum.

INTRODUCTION

The proliferation of DNA sequencing and bioinformatics workflows in recent decades has sparked a rapid expansion of recognized, but yet to be cultivated, archaeal diversity (e.g., Bathyarchaeota, Verstraetearchaeota, Marsarchaeota, Nezharchaeota, Brockarchaeota). The potential ecological roles and evolutionary histories of these lineages have been extensively studied.^{1–5} One major archaeal lineage, the “TACK superphylum,” is a phylogenetically and functionally diverse group that was initially described based on a monophyletic group composed of Thaumarchaeota, Aigarchaeota, Crenarchaeota, and Korarchaeota,⁶ although alternative names and taxonomies for this group are now widely used.⁷ Within the TACK superphylum, the class *Korarchaeia* is the most basal branch, implying an important role in the early evolution and diversification of the group.^{6,8} Unexpectedly, McKay and colleagues described metagenome-assembled genomes (MAGs) belonging to *Korarchaeia* with genomic capacity for dissimilatory reduction of sulfite and

production and/or oxidation of methane.⁹ Combined with the deep phylogenetic position of the *Korarchaeia*, this finding supported the hypothesis that life on Earth may have originated in thermal habitats and conserved energy via methanogenesis and/or dissimilatory sulfite reduction.^{10,11} However, most *Korarchaeia* remain uncultured and unclassified and are only based on single marker gene analyses.

Many thermophilic archaea, including members of the TACK superphylum, are known or proposed to be either heterotrophs or chemolithoautotrophs with sulfur-dependent metabolisms and a free-living lifestyle.^{12–15} Very few TACK lineages are reported to be symbiotic,¹⁶ and virtually all known or presumed symbiotic archaea belong to the distantly related DPANN superphylum. DPANN archaea have greatly reduced cell and genome sizes,^{17,18} and most lack complete biosynthetic pathways for nucleotides, lipids, and amino acids, which has been widely used to infer direct interactions with and dependence on other microbes. To date, no archaea within the TACK superphylum are known to have such features.



Table 1. Summary statistics of the *Panguiarchaeum symbiosum* MAGs reconstructed from hot spring samples

Bins	JZ-1 bin_83 ^T	SRBZ-1 bin_73	SRBZ-1 bin_65	JZ-2 bin_52	JZ-1 bin_98	JZ-1 bin_103	JZ-1 bin_107	ZZQ bin_139	SZTDM-1b bin_1	BF2 bin_16
No. of scaffolds	29	34	70	52	54	59	65	62	67	200
Sampling time	May, 2017	Aug, 2018	Jan, 2019	Jan, 2016	Sep, 2017	Dec, 2019	Mar, 2018	Mar, 2018	Sep, 2017	Aug, 2018
Genome size (bp)	1,356,921	1,425,407	1,425,832	1,256,628	1,257,246	1,248,635	1,267,925	1,228,896	1,317,496	1,114,919
GC content (%)	40.21	40.27	40.45	40.17	40.20	40.22	40.26	40.06	40.24	40.17
N50 value (bp)	149,542	106,043	50,799	38,034	38,904	47,189	35,341	31,118	40,452	6,331
No. of CDSs	1,297	1,346	1,379	1,239	1,235	1,209	1,246	1,228	1,244	1,194
Coding density (%)	90.2	87.9	89.2	88.5	90.0	90.3	89.9	90.3	89.6	90.3
No. of rRNAs	2	2	2	0	2	2	2	0	1	0
No. of tRNAs	45	46	49	43	46	40	45	44	43	41
No. of arCOGs	812 (62.6%)	822 (61.1%)	851 (61.7%)	765 (61.7%)	772 (62.5%)	752 (62.2%)	776 (62.3%)	756 (61.6%)	781 (62.8%)	725 (60.7%)
No. of KOs	578 (44.6%)	584 (43.4%)	585 (42.4%)	545 (44.0%)	537 (43.5%)	539 (44.6%)	554 (44.5%)	532 (43.3%)	544 (43.7%)	487 (40.8%)
Completeness (%) ^a	100 (85.2)	97.9 (83.8)	97.9 (83.8)	96.8 (82.9)	96.8 (82.9)	95.7 (81.9)	95.7 (81.9)	95.3 (81.1)	93.5 (80.5)	71.0 (61.4)
Contamination (%) ^a	1.61 (0.93)	2.69 (1.87)	1.61 (0.93)	1.08 (0.47)	1.08 (0.47)	1.08 (0.47)	0 (0)	1.08 (0.47)	2.15 (1.4)	0 (0)
Rel. abund. (%) ^b	0.08	0.14	0.06	0.28	0.04	0.06	0.06	0.06	0.04	0.03
SNPs	2,377	13,748	13,360	3,396	3,833	3,742	3,652	3,619	13,036	10,780

Abbreviations: CDSs, protein-coding genes; Rel. abund., Relative abundance.

^aGenome completeness and contamination were estimated using CheckM.²¹ Twenty marker proteins were excluded from the estimation due to the inherent absence in all *Panguiarchaeum symbiosum* MAGs. The results shown in parentheses were generated using the "checkm lineage_wf" procedure in CheckM.

^bThe relative abundance of each bin was computed as: total reads mapped to bins/total reads in the corresponding sample.

In this study, we applied a metagenomic approach to reconstruct 10 MAGs from terrestrial geothermal systems in southwest China and show that they comprise a single species within a previously undescribed order, herein named *Panguiarchaeales*, which is phylogenetically positioned as a sister lineage of *Korarchaeales*. Analysis of the gene repertoire of these *Panguiarchaeales* MAGs revealed limited biosynthetic potential with apparent dependencies similar to those in DPANN, suggesting a symbiotic lifestyle. By implementing population metagenomics and metatranscriptomic analyses, we provide evidence that nutrient uptake and cell attachment are under positive selection and are highly expressed and therefore of great importance to the ecology and evolution of *Panguiarchaeales*.

RESULTS

New MAGs from terrestrial geothermal systems represent the novel species *Panguiarchaeum symbiosum* within the novel order *Panguiarchaeales*

Ten MAGs were obtained from five high-temperature (>75°C) hot spring sediments located in Tengchong County, Yunnan Province, China (see [additional resources](#) for detailed physicochem-

ical and geolocation information of the sampling sites). All MAGs except one were of high quality, with genome sizes ranging from 1.23 to 1.43 Mbp, completeness >90%, and contamination <3% (Table 1).¹⁹ Most MAGs included two rRNA operons (16S rRNA and 23S rRNA) and >40 tRNAs. On average, a total of 1,262 protein-coding genes were predicted, with an average length of 916 base pairs (bp). To resolve the phylogenetic affiliations of the new MAGs, we constructed four phylogenetic trees based on different marker gene sets: (1) 122 archaeal marker proteins (Figure 1A); (2) 16S rRNA genes (Figure S1A); (3) 16 ribosomal proteins (Figure S1B); and (4) 56 universal marker proteins (Figure S1C). The topologies of the four taxonomic trees are generally consistent with each other and reveal a novel lineage branching within *Thermoproteota* (Figures 1A and S1). All 10 MAGs are closely related with short branch lengths (Figure 1B). Combined with pairwise average nucleotide identities >99%, we infer that they belong to the same species (Figure 1C). The name *P. symbiosum* within the novel order *Panguiarchaeales* is proposed under the nascent SeqCode,²⁰ with an average relative evolutionary divergence (RED) value of 0.32 (see [STAR Methods](#)) and the genome JZ-1.bin_83^T proposed as the nomenclatural type for the species. Pangu (盘古) is a god in

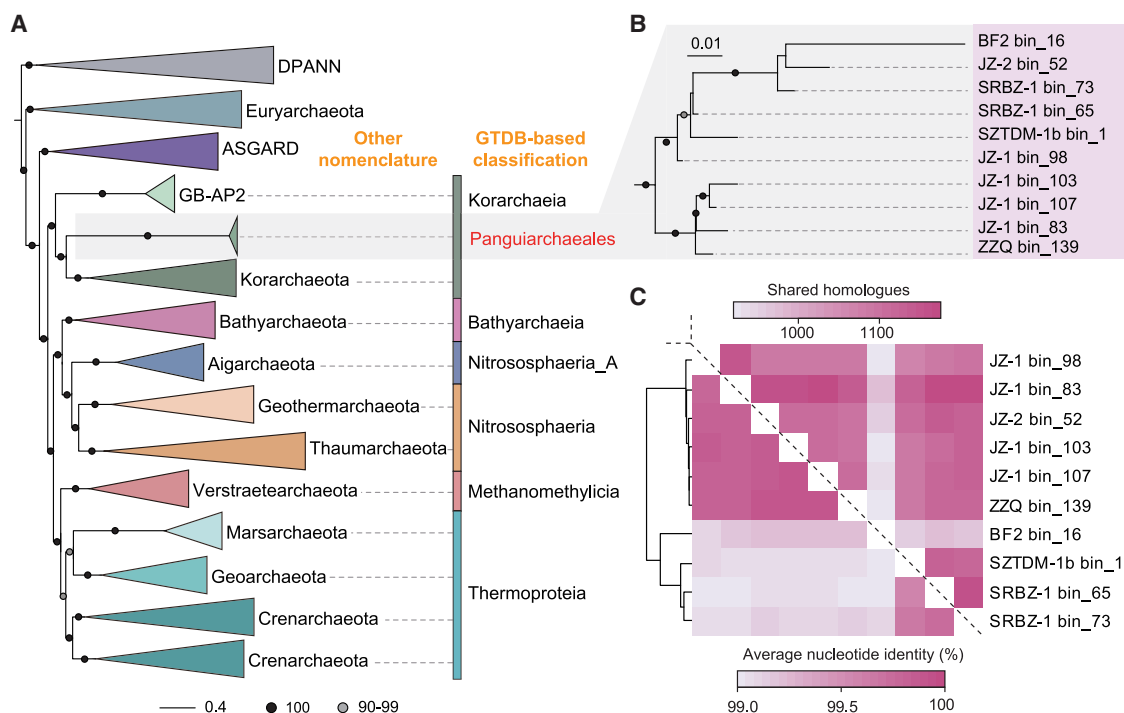


Figure 1. Phylogenetic placements of *Pangiarchaeum symbiosum* MAGs

(A) Maximum-likelihood tree based on 474 archaeal genomes (464 from public databases and 10 from the current study) was inferred from a concatenation of 122 marker proteins using IQ-TREE²² with 1,000 ultrafast bootstrap iterations. Genomes from DPANN superphylum are used as outgroups. Additional phylogenetic trees are shown in Figure S1.

(B) Enlarged phylogenetic tree of the *Pangiarchaeum symbiosum* MAGs assembled from the present study.

(C) Hierarchical clustering heatmap based on average nucleotide identities (ANIs, lower left of the squares beneath the dashed line) and shared homologs (upper right of the squares above the dashed line) between each pair of *Pangiarchaeum symbiosum* MAGs. The dendrogram was clustered according to the Euclidean distance of ANIs, which was implemented in the *heatmap* v1.0.12 package in R v4.1.1.

Chinese mythology who separated heaven and Earth from chaos and created the world. A protologue for each proposed taxonomic name is provided (see [experimental model and subject details](#)). While we present 10 MAGs here, as yet, only 20 near-full-length 16S rRNA sequences publicly available from the NCBI-nt database are affiliated with this novel lineage (Figure S1A). None of them have genomic information, underscoring the importance of the MAGs described in this study. All of these 16S rRNA genes are derived from ecosystems with high temperatures, with deeper branching taxa from marine hydrothermal vents and sediments and a single monophyletic group from terrestrial hot springs, including the MAGs from this study, suggesting a marine hydrothermal origin of this lineage.

P. symbiosum is an anaerobic amino acid fermenter with a symbiotic lifestyle

P. symbiosum MAGs lack known carbon fixation pathways and are thus likely to be heterotrophs. The absence of several genes involved in glycolysis; the tricarboxylic acid (TCA) cycle; respiratory complexes I, III, and IV; the high-affinity cytochrome *bd* ubiquinol oxidase; and other terminal oxidase complexes suggest an anaerobic and non-respiratory mode of energy conservation (Figure 2 and Table S1). Similar to most DPANN archaea, *P. symbiosum* MAGs lack a complete pathway for *de novo*

biosynthesis of most amino acids, except for alanine, glutamate, and citrulline (Figures 2 and 3). Given the considerable number of peptide/amino acid transporters and peptidases/aminotransferases, *P. symbiosum* may use extracellular peptides and amino acids to support cell growth (Tables S2 and S3). Most of the peptidases/aminotransferases likely function intracellularly, except four peptidases (C11A, C14B, C25 and S08A) with N-terminal signal peptides identified. The former three are cysteine peptidases that strictly cleave arginine residues (or sometimes lysine, for C14B and C25), and the latter is a nonspecific serine peptidase with a preference to cleave peptide bonds adjacent to hydrophobic residues.²³ The detected aminotransferases lead to the deamination of amino acids to 2-keto acids (Table S1). The generated 2-keto acids, including pyruvate, oxoglutarate, and indolepyruvate, could be further decarboxylated via the putative enzymes pyruvate:ferredoxin oxidoreductase (POR), oxoglutarate:ferredoxin oxidoreductase (KOR), and indolepyruvate:ferredoxin oxidoreductase (IOR). The *P. symbiosum* MAGs also harbor a homolog encoding aldehyde:ferredoxin oxidoreductase (AOR) that could catalyze the oxidation of aldehydes to corresponding organic acids. In addition, the CoA derivatives may be fermented to acetate for energy conservation with an ATP-yielding acetate-CoA ligase (ACD) (Figure 2). Given the identification of an apparently complete protein degradation

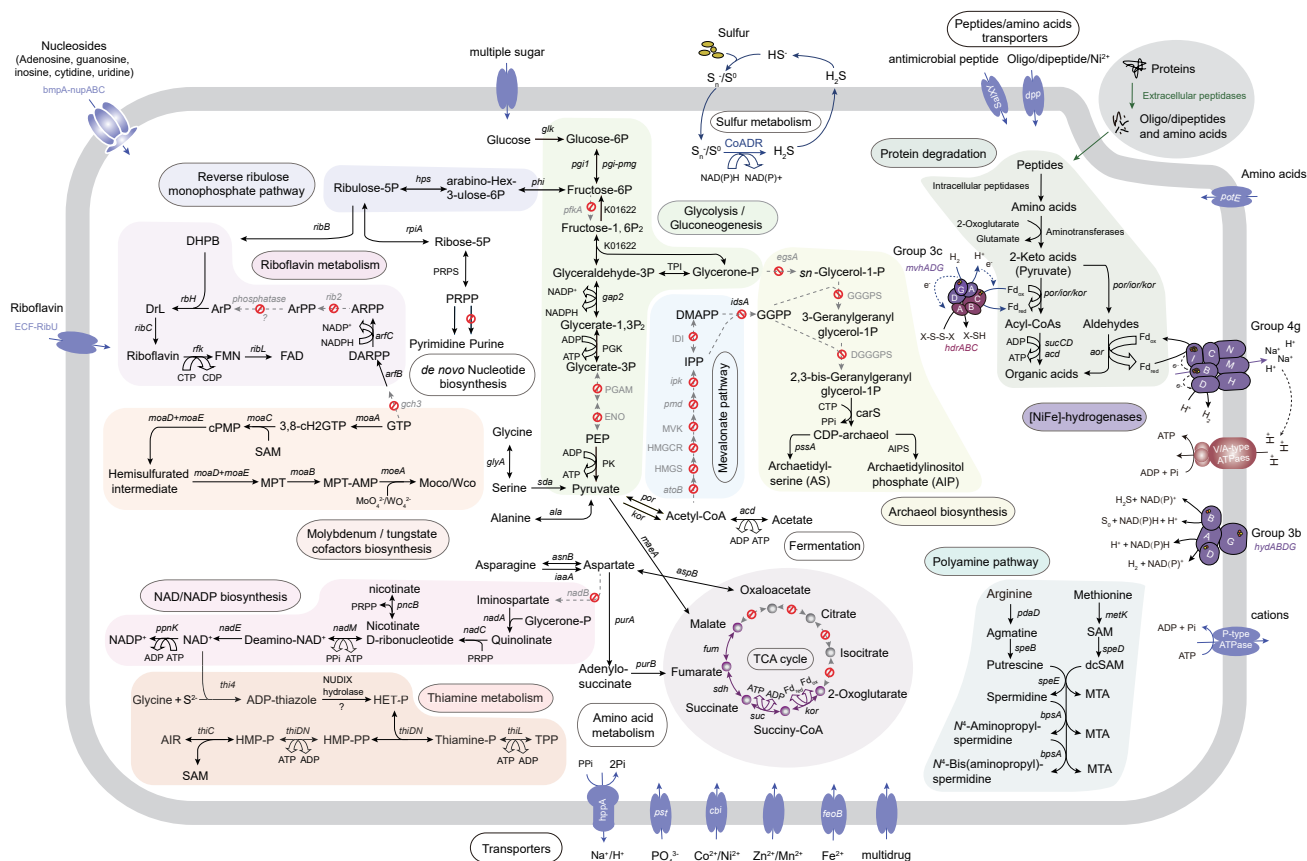


Figure 2. Overview of the metabolic potential of *Pangiarchaeum symbiosum*

Metabolic pathways were constructed based on annotation of predicted putative protein-coding genes against different databases. Pathways related to glycolysis, fermentation, nucleotide and amino acid biosynthesis and salvage, cofactor (e.g., Moco, TPP, FAD, FMN, NAD, and NADP) biosynthesis, and other significant metabolisms are shown (see Table S1 for full names and copy numbers of the genes). Abbreviations: ADT, ADP-thiazole; AIR, aminoimidazole ribotide; ArP, 5-Amino-6-(1-D-ribitylamino)uracil; ArPP, 5-Amino-2,6-dioxy-4-(5'-phospho-D-ribitylamino)pyrimidine; ARPP, 5-Amino-6-(ribosylamino)-2,4-(1H,3H)-pyrimidinedione 5'-phosphate; cPMP, cyclic pyranopterin monophosphate; DARPP, 2,5-Diamino-6-(1-D-ribosylamino)pyrimidin-4(3H)-one 5'-phosphate; dcSAM, decarboxylated S-Adenosylmethionine; DHPB, L-3,4-Dihydroxybutan-2-one 4-phosphate; DrL, 6,7-Dimethyl-8-(D-ribityl)lumazine; FAD, flavin adenine dinucleotide; FMN, flavin mononucleotide; GTP, guanosine 5'-triphosphate; HET, 4-methyl-5-β-hydroxyethylthiazole; HMP, 4-amino-5-hydroxy-2-methylpyrimidine; Moco, molybdenum cofactor; MPT, molybdopterin; MTA, Methyl-5'-thioadenosine; NAD, nicotinamide adenine dinucleotide; NADP, nicotinamide adenine dinucleotide phosphate; SAM, S-adenosylmethionine; TCA, tricarboxylic acid cycle; TPP, thiamine pyrophosphate; IPP, isopentenyl pyrophosphate; DMAPP, dimethylallyl pyrophosphate; GGPP, geranylgeranyl pyrophosphate.

pathway, including genes for the oxidation of amino acids to organic acids and their subsequent fermentation, we therefore propose that *P. symbiosum* strains likely use exogenous proteins/peptides/amino acids as carbon, nitrogen, and energy sources to support their growth.

Both the oxidative and non-oxidative pentose phosphate pathways (PPP) are absent in *P. symbiosum* due to the lack of transketolase and transaldolase. Putative 3-hexulose-6-phosphate synthase and 6-phospho-3-hexuloisomerase encoded by the *hps* and *phi* genes suggest that the reverse ribulose monophosphate pathway (rRuMP) may function in the biosynthesis of ribulose 5-phosphate (Ru5P) in place of the conventional PPP.^{24,25} Instead of forming a fused gene as proposed in most hyperthermophilic Archaea,^{25,26} *hps* and *phi* form two separate genes in *P. symbiosum* that are not adjacent to each other. Notably, the *phi* gene encodes two equivalent domains

that possibly perform the same function. Phylogenetic analysis of key sugar isomerase (SIS) domains identified in *phi* genes reveals that this domain structure exclusively occurs within Archaea (Figure S2A). Phylogenetic rooting based on the minimal ancestor deviation (MAD) method separates Archaea and Bacteria into two distinct branches,²⁷ suggesting an independent evolutionary trajectory. The similar topology between the two SIS domains in these archaea reveals that an ancient duplication event may have occurred, leading to the divergence of the two SIS copies (Figure S2B). Each SIS domain is well conserved and distinct from the other (Figure S2C), suggesting that the domains may confer different catalytic activities under different environmental conditions. All *P. symbiosum* MAGs contain genes encoding 5-phosphoribosyl-α-1-pyrophosphate (PRPP) synthase, which leads to the formation of PRPP from Ru5P, a key precursor for purine and pyrimidine biosynthesis.

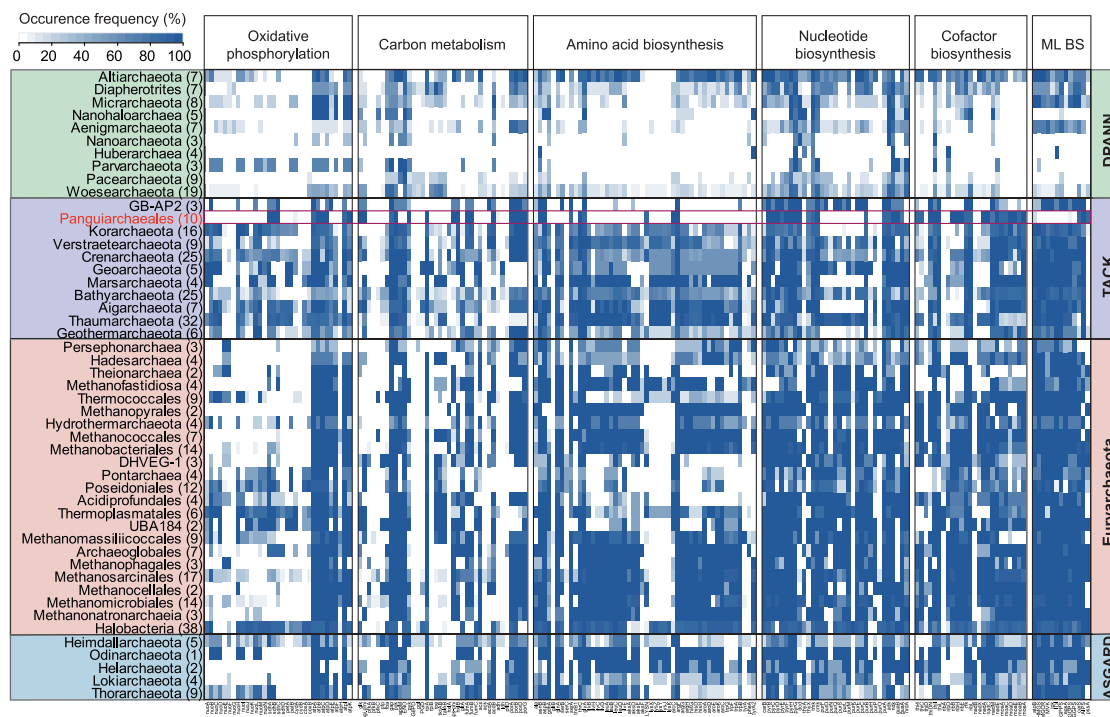


Figure 3. General functional profiles of major archaeal lineages regarding basic metabolism

Pathways including oxidative phosphorylation, carbon/nitrogen/sulfur/hydrogen cycling, and biosynthesis of amino acids, nucleotides, lipids and cofactors are shown. Each column in the heatmap represents one gene in a specific pathway. Characters in the color-coded boxes on the left represent the taxonomic information, and numbers in the brackets indicate the number of genomes in that lineage. A total of 408 archaeal genomes were taken into consideration with 398 from the public database and 10 from the present study (see [additional resources](#) for the detailed genomic information and gene occurrences in each archaeal lineage). Each square in the heatmap describes the occurrence frequency of a gene in each phylogenetic cluster, which is calculated as the percentage of genomes containing the gene across the total number of genomes of the corresponding taxonomic group. ML BS, membrane lipid biosynthesis.

A complete *de novo* pyrimidine synthesis pathway was detected within all *Panguiarchaeales* MAGs (Figure S3A). However, they are unable to synthesize purines *de novo* because they lack genes to convert PRPP to the key intermediate inosine monophosphate. Consequently, they likely rely on the salvage system to recycle purines based on the activities of purine nucleoside phosphorylase, adenosine kinase, and hypoxanthine phosphoribosyltransferase (Figure S3B and Table S1). Furthermore, the detected nucleoside ABC uptake system encoded by *bmpA* and *nupABC* ensures the transport of all common nucleosides across the membrane. Collectively, the presence of rRuMP provides *P. symbiosum* the precursor to synthesize pyrimidines *de novo*, whereas purines have to be salvaged.

Similar to DPANN archaea, most genes encoding enzymes in the archaeal mevalonate pathway and phospholipid biosynthesis are absent in the *P. symbiosum* MAGs (Figure 2). In particular, none of them harbor mevalonate kinase encoded by *mvk*, a key gene involved in the isoprenoid biosynthesis pathway. Despite homologs of *mvk* being annotated, phylogenetic analyses place these amino acid sequences distant from all characterized *mvk* genes (Figure S4). Further identification of functional domains demonstrates that they are putative galactokinases, differing from *mvk* but belonging to the same superfamily. Hence, we speculate that *P. symbiosum* is unable to synthesize archaeal membranes independently, which is uncharacteristic of

microbes within the TACK superphylum, including their closest relatives in *Korarchaeales* and GB-AP2 (Figure 3). Given the lack of genes involved in lipid transporters, they presumably rely on partner organisms to provide lipids or intermediates to assist the biosynthesis of their own membrane lipids.

Based on the presence of proteins containing pectin lyase, lectin, fibronectin type III, immunoglobulin-like, beta propeller, WD40 (beta-transducin repeat), PDK (phosphoinositide-dependent protein kinase), adhesion, and von Willebrand factor domains, which are commonly involved in cell-cell interactions in DPANN,^{18,28} we predict that 32 dereplicated proteins (average 16 in each MAG) may have a role in cell adhesion (Table S4) and in turn may contribute to the attachment of *P. symbiosum* to partner organisms. We also identified several hypothetical proteins containing thrombospondin type-3 repeat domains (IPR028974) in *P. symbiosum* MAGs, which may participate in cell-extracellular matrix interactions. Combined with the lack of most genes related to the biosynthesis of amino acids, purines, and lipids, the presence of abundant cell-cell and cell-matrix proteins suggests that *P. symbiosum* may represent a symbiotic archaeal lineage that is outside of the DPANN superphylum.

Metabolism of cofactors

Despite their possible symbiotic lifestyle, genes encoding cofactors are widely detected among the *P. symbiosum*

MAGs. Cofactors are necessary to maintain the proper functioning of enzymes and promote the growth of microorganisms. Almost all *P. symbiosum* MAGs harbor the key genes *moaABCDE* and *moeAB*, which are essential for molybdenum or tungstate cofactor biosynthesis (Figure 2).^{29,30} Consistent with most thermophilic archaea, *mogA* is lacking in all *P. symbiosum* MAGs. Instead, they possibly adopt *moaB*, a *mogA*-homolog that is commonly detected in thermophiles, to catalyze the adenylation of metal-binding pterin.²⁹ Two copies of MoeA-like proteins in most *P. symbiosum* MAGs suggest either metal-selective activities or the potential formation of two different cofactors.²⁹ These genes, combined with the presence of TupA (TC# 3.A.1.6.2; Table S3), an ABC-type transporter with extremely high affinity for tungstate, suggest *P. symbiosum* may have the capacity to use tungsten-containing ferredoxin oxidoreductases to oxidize various aldehydes derived from amino acid degradation.³¹ The absence of the *mobA* gene in *P. symbiosum* indicates the inability to synthesize enzymes belonging to the DMSO reductase family. Given the presence of the *mobB* gene, which is normally co-transcribed with *mobA* and facilitates the cofactor transfer to target enzymes, the possibility that *P. symbiosum* might use a novel enzyme to perform the function cannot be ruled out. *P. symbiosum* is able to synthesize thiamine *de novo* based on genes encoding sulfide-dependent adenosine diphosphate thiazole synthase (Thi4), phosphomethylpyrimidine synthase (ThiC), thiamine monophosphate kinase (ThiL), and hydroxymethylpyrimidine kinase/phosphomethylpyrimidine kinase/thiamine-phosphate diphosphorylase (ThiDN) (Figure 2). Similar to some methanogens and other hyperthermophiles, *P. symbiosum* might use sulfide as the sulfur source for thiamine biosynthesis due to the presence of a conserved His residue in Thi4.³² Although the gene coding thiamine phosphate synthase is absent, an annotated ThiN domain with a length of ~180 amino acids adjacent to an annotated ThiD domain within the ThiDN gene possibly performs the same function.³³

Riboflavin (vitamin B2) is the precursor of the coenzymes flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN), which act as coenzymes in numerous redox reactions.^{34,35} All *P. symbiosum* MAGs have an incomplete riboflavin biosynthesis pathway, lacking the key enzyme GTP cyclohydrolase II, which hydrolyzes the imidazole ring of GTP. However, *P. symbiosum* may be able to transport riboflavin directly via the ECF riboflavin uptake complex (EcfA₁A₂T-RibU) and subsequently catalyze the phosphorylation of riboflavin with the production of FMN performed by archaeal-specific CTP-dependent riboflavin kinase.³⁶ FMN can then be further converted into FAD via an archaeal-specific FAD synthetase (Figure 2). Nicotinamide adenine dinucleotide (NAD), another essential cofactor involved in numerous redox reactions, can likely be synthesized *de novo* by *P. symbiosum* starting from quinolinate.

Energy conservation in *P. symbiosum*

Two homologs sharing 27% to 43% amino acid identities with Coenzyme A disulfide reductases (CoADRs) in *Thermococcus kodakarensis*, *Pyrococcus furiosus*, and *Pyrococcus horikoshii* (see details in additional resources), were identified in the *P. symbiosum* MAGs. CoADRs are a class of single-cysteine-

containing FAD- and NAD(P)H-dependent enzymes that use a tightly bound CoA to assist the reduction of elemental sulfur, di-, per-, and/or polysulfide substrates to hydrogen sulfide.³⁷ Phylogenetic analysis revealed that two copies of CoADRs are divergent from each other at the amino acid sequence level (Figures 4A and 4B). Though one copy and their homologs from archaea form a separate clade, frequent inter-phylum HGTs occurred since CoADRs from Euryarchaeota and TACK superphylum are interwoven across the tree (Figure 4A). Both genes exhibit conservation at coenzyme- and substrate-binding regions and the catalytic center. In particular, the indispensable active-site cysteine within the FAD-binding domain is present in all CoADR homologs with the capacity of S⁰ reduction and/or dissimilatory sulfur respiration, as well as all *P. symbiosum* homologs (Figure 4C). The predicted crystal structures of two copies of CoADRs within *P. symbiosum* are very similar to those encoded by *Pyrococcus horikoshii* CoADR (PDB ID: 4FX9), with RMSD values of 0.35 Å and 1.88 Å, respectively (Figures 4D and 4E), which have been experimentally confirmed to reduce persulfide or polysulfide substrates.³⁷ In addition, bifunctional sulfhydrogenases (*hydABCD*) identified as group 3b [NiFe]-hydrogenases may confer *P. symbiosum* the ability to oxidize H₂ with NAD(P)⁺ as the electron acceptor (Figure S5A). Alternatively, it may catalyze the production of H₂ with NAD(P)H as an electron donor when sulfur is absent or reduce elemental sulfur to H₂S when S⁰ is available.³⁸ The predicted four subunits (*hydABDG*) have comparable molecular masses to the *Thermococcus litoralis* homologs, and are organized in an operon, suggesting they may form a single transcriptional unit and work in concert to perform the function (Figures S5B and S5C). Genes encoding for all subunits of a heterodisulfide reductase HdrABC and an F420-non-reducing hydrogenase MvhADG were detected, revealing the capacity for electron bifurcation. Unlike the associations with CoM-S-S-CoB or sulfate reduction in methanogens or sulfate-reducing microorganisms, this complex could reduce ferredoxin and an unknown disulfide reducer coupling with hydrogen oxidation.³⁹ All *P. symbiosum* MAGs encode a gene cluster with nine subunits showing homology to subunits of the NADH:quinone oxidoreductase (*nuoBCD-HIKLMN*) (Figure S5B). The lack of NADH binding module (*nuoEFG*) and the phylogenetic placement of the *nuoD* gene suggest that the Nuo-like complex of the *P. symbiosum* MAGs is likely to be a membrane-bound and energy-conserving group 4g [NiFe]-hydrogenase (Figure S5A). This type of hydrogenase can couple the formation of H₂ to the regeneration of oxidized ferredoxin to conserve free energy, while pumping protons across the membrane to produce the proton motive force and ultimately leading to the formation of ATP by archaeal V-type ATP synthase. Collectively, these observations suggest that *P. symbiosum* might have the ability to use hydrogenases and sulfur-based respiration for energy conservation, which is common in hyperthermophilic archaea.¹³

Determinants of thermophily and evasion of viral infection

P. symbiosum is very likely thermophilic, since all 16S rRNA genes and genomes belonging to this order derive from environments with high temperatures (Figure S1A). The high fractions of

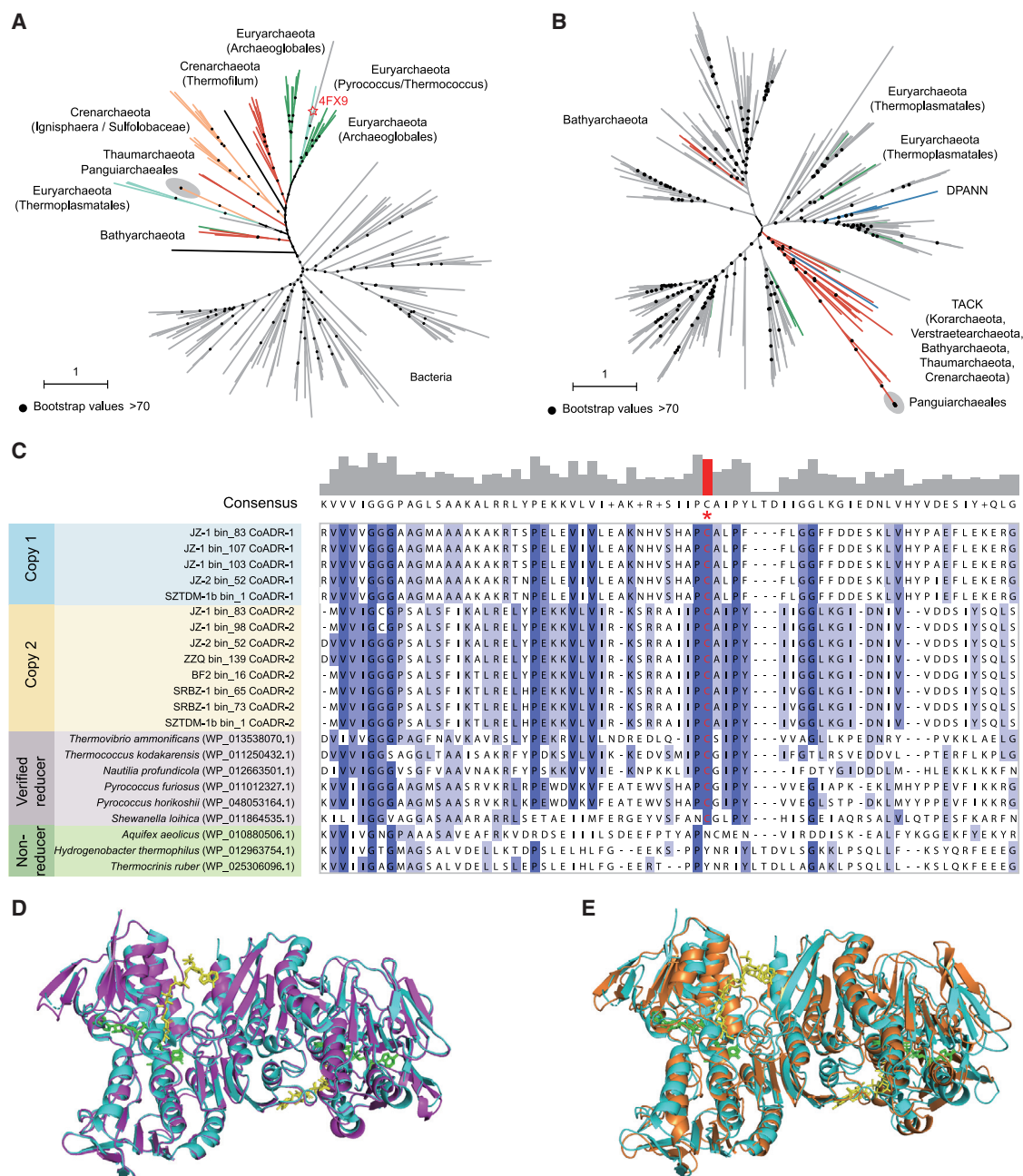


Figure 4. CoA-dependent NAD(P)H sulfur oxidoreductases in Pangiarchaeum symbiosum MAGs

(A and B) Phylogenetic analyses of two copies of CoADRs in *Pangiarchaeum symbiosum* MAGs.

(C) Structure-based sequence alignment of FAD-binding domain segments in selected CoADRs. The active-site cysteine residue is marked with red “*” on the top. The alignment was performed using MUSCLE⁴⁰ v3.8.31.

(D) Superposition in stereo of JZ-1 bin_83 CoADR-1 (magenta) and *Pyrococcus horikoshii* CoADR (4FX9, cyan) with the rmsd value of 0.35 Å.

(E) Superposition in stereo of JZ-1 bin_83 CoADR-2 (orange) and *Pyrococcus horikoshii* CoADR (4FX9, cyan) with the RMSD value of 1.88 Å. CoA (yellow) and FAD (green) in (D) and (E) are represented as sticks. The star in (A) represents CoADR gene (4FX9) from *Pyrococcus horikoshii*.

the amino acids IVYWREL (average 45.5%) in the proteome, which is correlated with optimal growth temperature (OGT),⁴¹ further support this notion. By applying the previously fitted formula, the predicted OGT for *P. symbiosum* can reach up to 92.2°C with 91.2°C on average, which is similar to the measured

temperatures of sediments from which the metagenomes were derived (75–91°C). Besides, each *P. symbiosum* MAG encodes a reverse gyrase, a DNA topoisomerase that introduces positive supercoils into DNA to maintain genomic stability in thermophiles.⁴² In addition, almost all *P. symbiosum* MAGs possess

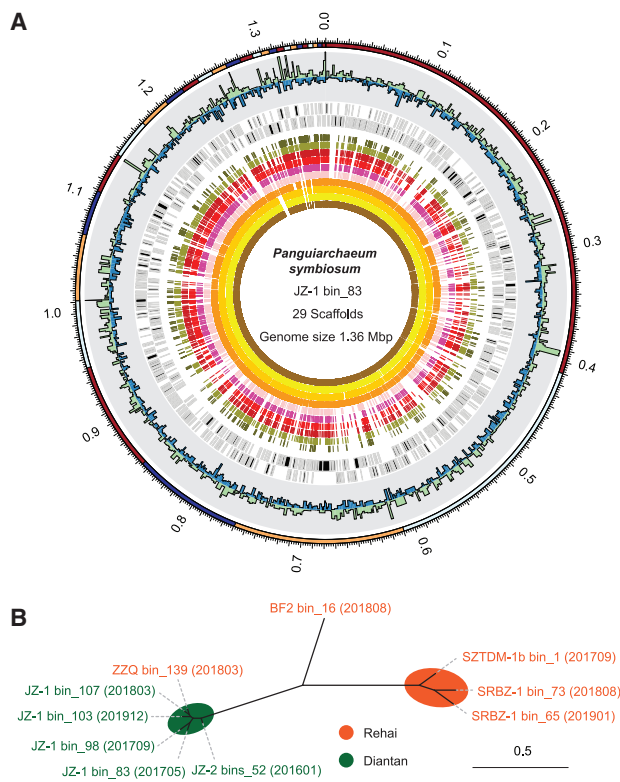


Figure 5. Single-nucleotide polymorphisms (SNPs) in *Pan-guarchaeum symbiosum* MAGs

(A) Identified SNPs in each assembled *Pan-guarchaeum symbiosum* MAG. JZ-1 bin_83 was selected as the template to show the position of SNPs. Circles from the outermost to the innermost depict the following: (1) Genome fragments distinguished by colors. (2) GC content, with green and blue bars representing values higher and lower than the average GC rate of the genome. GC was calculated with a window size of 1,000 and a step size of 1,000. (3) Putative protein-coding regions (see [STAR Methods](#) for detailed prediction procedures), with black bars on the outer and inner sides representing positive and negative orientations of genes. The width of the bars indicates the gene length proportionally. The 10 inner circles represent SNPs scanned in JZ-1 bin_83, JZ-1 bin_98, JZ-1 bin_107, JZ-1 bin_103, JZ-2 bin_52, ZZQ bin_139, BF2 bin_16, SRBZ-1 bin_73, SRBZ-1 bin_65 and SZTDM-1 bin_1. This plot was visualized using Circos⁴⁶ v0.69-9.

(B) Phylogenetic tree based on identified SNPs (see details for SNP detection and phylogenetic tree reconstruction in [STAR Methods](#)).

the polyamine biosynthesis pathway, which confers the ability to biosynthesize polyamines, putrescine, and spermidine from arginine and methionine that may stabilize the cytoplasm ([Figure 2](#)). Moreover, a novel aminopropyltransferase, branched-chain polyamine synthase A, that is responsible for the biosynthesis of branched-chain polyamines, such as *N*⁴-aminopropylspermidine and *N*⁴-bis(aminopropyl)spermidine, was detected in all *P. symbiosum* MAGs. These positively charged polyamines might be synthesized by *P. symbiosum* to maintain the stability of DNA and RNA structure under high-temperature conditions.⁴³ In addition, a previous study revealed that the hyperthermophilic archaeon *Pyrococcus furiosus* relied on protein refolding (Hsp60-like chaperonin, Hsp20-like small heat shock protein and two VAT-related chaperones), proteolytic degrada-

tion (proteasome), and stabilization (di-myo-inositol phosphate) to cope with thermal stress.⁴⁴ All related genes without exception were found in the *P. symbiosum* MAGs ([Table S1](#)). Overall, these observations reveal that *P. symbiosum* is likely to be hyperthermophilic and may have adopted diverse strategies to protect itself against thermal stress.

The MAG SRBZ-1 bin_73 harbors a complete type I CRISPR-Cas system, suggesting the potential capacity to resist phage infection ([Figure S6A](#)). Four of the 10 MAGs contain CRISPR arrays. By taking a deeper look into the array sequences possessed by the MAGs, we found that only a minority of the spacers were shared even between MAGs from the same sampling site. Specifically, two MAGs with the best genome quality, SRBZ-1 bin_73 and JZ-1 bin_83, possess 31 and 17 spacers, respectively. By applying BLAST searches, only two of these spacers shared significant homology ([Figure S6B](#)), suggesting a high diversity of viruses infecting these *P. symbiosum* populations and that host and virus populations may vary across temporal and/or spatial scales. Functional annotation by applying BLAST search against the NCBI-nr database suggests that four genes in some of the MAGs derived from SRBZ encode proteins involved in either restriction-modification (RM) or bacteriophage exclusion (BREX) systems, including *brxL*, *hsdM*, *pglX*, and *Hinp1I*, which could also resist infection by viruses ([Figure S6C](#)). Interestingly, the putative BREX system encoded by *brxL* and *pglX*, albeit incomplete, might allow phage adsorption but prevent DNA replication.⁴⁵ The entered virus could be further degraded and recycled as a nucleotide source to support the growth of *P. symbiosum*.

Geographic distance contributed more to genetic differentiation than temporal variation

In addition to differences at the whole-genome level, variations at the nucleotide level were also observed, enabling analysis of the population genetics of *P. symbiosum*. With high accuracy, we detected single-nucleotide polymorphisms (SNPs) by mapping clean reads from different samples to the highest quality reference genome JZ-1 bin_83. A total of 71,543 SNPs (average 7,154 SNPs per MAG) covering 24,216 loci of the reference genome were identified among the 10 MAGs ([Figure 5A](#) and [Table 1](#)). With JZ-1 bin_83 as the reference, fewer SNPs were observed in *P. symbiosum* MAGs derived from JZ than MAGs from SRBZ, demonstrating higher similarities between MAGs from the same location. These SNPs traced the distinct topology structure as the phylogenomic tree ([Figures 1B](#) and [5B](#)). Unlike the phylogeny reconstructed based on conserved markers ([Figure 1B](#)), a phylogeny based on the SNP alignment formed three groups roughly according to geographical distance ([Figure 5B](#)). MAGs reconstructed from Diantan and Rehai form two separate groups, with BF2 bin_16 representing the third cluster. An exception is that ZZQ bin_139 was sampled from Rehai but clustered into the Diantan group, suggesting that selection based on physicochemical or ecological niches, rather than dispersal limitation, drives these geographic clusters. Despite the wide range of temperatures of the sampling sites from which the *P. symbiosum* MAGs were derived, temperature plays a limited role in the observed genetic variations (Spearman correlation, $\rho = 0.28$ and $p = 0.43$). Collectively, the SNP tree indicated

that geographic distance contributes to the genomic variation of *P. symbiosum*, with dispersal contributing little to its evolution.

Positively selected genes contribute to amino acid degradation and host-symbiont interactions

To determine the selection pressure on loci with variations among 10 *P. symbiosum* MAGs, gene families were clustered, and phylogenies were inferred, which were further used for the calculation of ratios of nonsynonymous to synonymous SNPs, dN/dS, a common feature used to describe the selection pressure (see details in [STAR Methods](#)). A total of 373 of 1,000 gene families (>5 genes in each) were detected with one or more SNP, with a median dN/dS of 0.31, indicating a general stabilizing selection trend rather than adaptive selection. However, positive selection was detected at 249 loci within 52 gene families ([Table S5](#)). This includes one of the CoADR genes with one locus showing positive selection, albeit not significant after FDR correction, suggesting the importance of sulfur reduction to *P. symbiosum*. FDR correction based on chi-square tests resulted in only 36 genes exhibiting significant levels of positive selection, mostly distributed across pathways associated with cell-cell interactions and amino acid or peptide uptake ([Table 2](#)). Fourteen and 16 genes were annotated to contain signal peptides and transmembrane regions respectively. Most positively selected loci (211 of the 221 sites, 95.5%) were located within either transmembrane or non-cytoplasmic regions. Interestingly, half of the genes identified with signal peptides encode one or more domains implicated in cell-cell attachment, including immunoglobulin-like, invasion, fibronectin, adhesin, von Willebrand factor, and intimin domains, which might be involved in the perception and activation of host-symbiont interactions.^{18,47} In addition, five genes involved in the transport and degradation of peptidases/proteins, including ABC transporters (K01990 and K02035), 4-aminobutyrate aminotransferase, Peptidase_C14B (MER1151619), and formiminoglutamase exhibited positive selection signals, supporting the inference that *P. symbiosum* ferments amino acids and peptides to support its growth ([Tables 2 and S5](#)).

To further explore the biology of *P. symbiosum*, metatranscriptomic sequencing was conducted to determine which genes are expressed *in situ*. Due to the extremely low biomass, sufficient RNA was only obtained from JZ-2 and JZ-3 collected in January 2021 and July 2022, respectively. After removal of rRNAs, 267 and 2,134 reads were mapped onto 325 genes in the *P. symbiosum* genome ([Table S6](#)), including 14 of the positively selected genes predicted to be important for cell-cell interactions and amino acid/peptide metabolism ([Table S6](#)). Notably, four genes with positive selection, including fibronectin-binding autotransporter adhesin, peptide/nickel transport system substrate-binding protein, peptidase C14B, and one signal peptide-containing hypothetical gene were identified as the top 10 highest expressed genes in both samples. Another positively selected gene predicted to include a bacterial immunoglobulin domain and inferred to be involved in cell attachment was also expressed, albeit only in one sample. These results further demonstrate the investment of resources *in situ* to express proteins that may mediate cell-cell interactions to promote symbiosis and amino acid/peptide metabolism, suggesting these functions are critical for the biology of *P. symbiosum*.

DISCUSSION

Phylogenomic trees are crucial for a meaningful interpretation of the biological and evolutionary relationships of all organisms. Yet, the position of the lineage herein named *Panguiarchaeales* has been unstable. For example, an older version of GTDB (release 202) based on 122 archaeal marker proteins placed *Panguiarchaeales* within the TACK superphylum, whereas the latest version (release 207v2) considering only 53 marker proteins assigned them into the ASgard superphylum.⁴⁸ The latter resulted in a less robust phylogenetic tree with low bootstrap support for the node connecting *P. symbiosum* to the ASgard superphylum and aberrant long branches within the *P. symbiosum* MAGs, contrary to the close relationship between the *P. symbiosum* MAGs (>99% average nucleotide identities). Save for this case, *P. symbiosum* was consistently placed as a sister lineage of the *Korarchaeales* by applying four different marker sets for phylogenetic reconstruction: 16S rRNA ([Figure S1A](#)), 16 universal ribosomal proteins ([Figure S1B](#)), 122 archaeal-specific marker proteins ([Figure 1A](#)), and 56 conserved archaeal-specific marker proteins without horizontal gene transfer (HGT) ([Figure S1C](#)). Therefore, there is support for the position of *P. symbiosum* diverging deep within the TACK superphylum, which raises the question of how they diverged from the common ancestor of the TACK superphylum.

Similar to DPANN MAGs, *P. symbiosum* lacks key pathways involved in purine, amino acid, and membrane lipid biosynthesis, suggesting that they may have a symbiotic lifestyle. Such reduced metabolic capacities are seldom detected in archaea, except in DPANN lineages. Although *Candidatus Cenarchaeum symbiosum* and several methanogens are reported to be involved in archaea-eukarya symbiotic associations, most of them encode near-complete amino acid, nucleotide, and cell membrane biosynthesis pathways.^{16,49} This suggests a greater degree of host dependency of *P. symbiosum* compared with these other examples. It could alternatively be argued that *P. symbiosum* may instead be free-living auxotrophs that are unable to synthesize many essential metabolites. We think this is unlikely because *P. symbiosum* lacks genes involved in cell membrane biosynthesis and would depend on other organisms to assist the assembly of their cell envelope. We treat this as the strongest indicator that they rely on close interactions with other organisms, as has been argued by others for both DPANN archaea and Patescibacteria.¹⁸ Direct evidence for sharing lipids between *Candidatus Nanoarchaeum equitans* and its host *Ignicoccus hospitalis* has been demonstrated by nearly identical lipid content and lipid ¹³C stable isotope analysis.⁵⁰ We further examined the co-occurrence between *P. symbiosum* and microbes from hot spring communities and identified recent HGT events to assist the inference of the potential host.⁵¹ Interestingly, we observed that *P. symbiosum* is positively correlated with one MAG from Aigarchaeota (Spearman correlation, rho = 0.64 and p = 5.4E-6, [Figure S7](#)). By conducting BLAST searches against 10 MAGs reconstructed from the same community where JZ-1 bin_83 derived, we identified 11 genes that exhibited high sequence identities to genes in other MAGs, suggesting recent HGTs ([Table S7](#)). Four of them were from one *Wolframiraptor* MAG in Aigarchaeota.⁵² Thus,

Table 2. Genes in *Panguiarchaeum symbiosum* JZ-1 bin_83^T exhibiting signatures of positive selection

Category	Index	dN/dS	P _{M1 vs M2}	P _{M7 vs M8}	dN	PS	Functional annotation	Location of PS ^a			
								S	T	C	N
Amino acid / peptide uptake and transport	1	13	0.043	0.043	7	4	ABC-2 type transport system ATP-binding protein				4
	2	4.1	0.033	0.032	18	1	Peptide transport system substrate-binding protein				1
	3	19	0.012	0.015	15	4	Peptidase_C14B				4
	4	46	0.001	0.002	12	3	Type I phosphodiesterase			3	
	5	145	0	0	5	3	Calcineurin-like phosphoesterase superfamily domain				3
	6	56	0	0	14	5	Formimidoylglutamase			5	
	7	54	0	0	7	5	4-aminobutyrate aminotransferase				5
Cell-cell attachment	8	3.4	0.088	0.036	23	2	Bacterial Ig domain		1	1	
	9	3.3	0.012	0.011	47	1	Invasin/intimin cell-adhesion fragments				1
	10	1.9	0.012	0.009	170	3	MSCRAMM family adhesin SdrC				3
	11	9.0	0.006	0.005	54	6	Bacterial Ig domain				6
	12	4.8	0	0	30	5	Ig-like domain				5
	13	22	0	0	95	32	Invasin: a bacterial integrin-binding protein				32
	14	58	0	0	77	14	Serine-rich adhesin, immunoglobulin-like receptor B				14
	15	548	0	0	34	8	Fibronectin-binding autotransporter adhesin ShdA				8
Others	16	11	0.051	0.040	4	1	2-oxoacid ferredoxin oxidoreductase subunit beta				1
	17	94	0.019	0.020	7	1	GTPase				1
	18	999	0.024	0.019	2	1	Hypothetical protein				1
	19	2.6	0.017	0.008	72	1	Hypothetical protein				1
	20	387	0.011	0.005	4	1	Hypothetical protein				1
	21	999	0.005	0.005	5	2	Magnesium chelatase subunit D				2
	22	35	0.005	0.005	11	3	HAD superfamily (subfamily IA) hydrolase				3
	23	48	0.006	0.004	5	2	Archaeosine tRNA-ribosyltransferase				2
	24	7.7	0.003	0.004	29	8	Mutual gliding-motility protein MglA				8
	25	50	0.001	0.001	4	3	Daunorubicin resistance ABC transporter ATPase				3
	26	28	0	0	15	5	Hypothetical protein				5
	27	204	0	0	6	3	Hypothetical protein				3
	28	17	0	0	27	8	Hypothetical protein		1	1	6
	29	14	0	0	22	12	Chromosome segregation ATPase				12
	30	25	0	0	33	7	Plasma-membrane choline transporter			7	
	31	20	0	0	36	10	Chromosome segregation ATPase				10
	32	3.0	0	0	127	16	Exported calcium-binding glycoprotein-related				16
	33	17	0	0	16	7	Malate dehydrogenase				7
	34	122	0	0	13	5	CoA-transferase family III			5	
	35	24	0	0	34	12	DEAD/DEAH box helicase domain-containing protein				12
	36	22	0	0	46	17	N-6 DNA Methylase				17

^aNumbers denote positive selections in different regions of the gene, including signal peptide (S), transmembrane (T), cytoplasmic (C) and non-cytoplasmic (N) region. Items shaded with black represent the presence of S and T in the corresponding gene. The prediction of locations was conducted by applying Phobius (see [STAR Methods](#)).

Abbreviations: PS, positive selection.

both co-occurrence and possible HGT suggest that Aigarchaeota may act as the potential hosts, although other possibilities cannot be ruled out. Collectively, we suggest that *P. symbiosum* may possess a symbiotic lifestyle rather than a free-living auxotroph with the reliance on metabolites that they cannot produce, most importantly lipids.

It is unclear whether gaps in the core metabolism of these MAGs represent an ancient feature or are due to genome reduction. Concordant with most DPANN genomes, the *P. symbiosum* MAGs contain genes encoding homologous recombination (e.g., RadAB, SbcCD, HerA, NurA and HelS) and DNA mismatch repair proteins (e.g., ERCC-2, ERCC-3, ERCC-4, PCNA, Nth, Nfo, and Lig1).^{18,53} Thus, no direct evidence supports that *P. symbiosum* is currently undergoing genome reduction owing to unrepaired DNA damage, and homologous recombination may promote the incorporation of external genetic material to better adapt to harsh conditions. Although the potential hosts are currently unknown, we speculate that *P. symbiosum* may form a mutualistic

relationship with its hosts. Similar to *Wolbachia*, an endosymbiont of the nematode *Brugia malayi*, *P. symbiosum* genomes encode biosynthetic pathways for vitamins and cofactors, such as NAD, FAD, FMN, thiamin, and molybdenum cofactor, which could be shared with host cells. In contrast, they require amino acids, purines, and perhaps other nutrients that may be supplied by hosts.⁵⁴ In contrast, genes relevant to *de novo* biosynthesis of vitamins and cofactors are rarely identified in DPANN archaea and the closest *P. symbiosum* neighbors in the *Korarchaeales*.^{8,18} This contributes to the larger genome sizes of *P. symbiosum* compared with DPANN and reflects that they may adopt different strategies to coexist with hosts and/or survive independent of their hosts temporarily. Another mechanism to form the mutual beneficial interaction is that *P. symbiosum* may be able to attract viruses because they would necessarily share membrane components with their host cells.⁵⁵ If so, that could diminish the host viral load and protect the host against viral infection. The presence of several virus defense

systems, such as the type I CRISPR-Cas, RM, and the putative BREX system, would provide a strong defense mechanism by the *Panguiarchaeales* to the host partner. Notably, the aforementioned DNA methylase encoded by *hsdM* involved in the Type I RM system is positively selected in the *Panguiarchaeales*, indicating the rapid evolution of antiviral systems (Table 2). Given the absence of enzymes required for *de novo* biosynthesis of purines, this may serve as one approach to salvage nucleotides.

Few population metagenomics studies have been applied to reveal intraspecies differences and to understand the evolution of natural populations.^{56–58} Comprehensive analysis of heterogeneity within the 10 *P. symbiosum* MAGs reveals fairly large genetic diversity within the species. The topology structure between the SNP tree and the phylogenomic tree differs markedly and reflects the different evolutionary trajectories between the core gene sets and landscapes of the whole genome. In comparison, the SNP tree exhibits better resolution than the latter and might be applicable to resolve the recent diversification at a shallower evolutionary time scale.⁵⁹ Several recent studies have explored how landscape genomics can provide greater power and resolution for examining spatial patterns and adaptive variation.^{60–62} The overall clustering pattern in the present study reveals that geographical separation and environmental heterogeneity may govern the population structure of *P. symbiosum* and have a profound influence on the observed population genetic differentiation, although the effect of dispersal cannot be ruled out on longer time scales. Consistent with a previous study, such extensive strain variations were maintained by an overall genomic pattern of purifying selection rather than positive selection.⁵⁸ The few positive selection events target very few functions in *P. symbiosum*, mostly involved in nutrient acquisition and cell attachment. Notably, most residues under positive selection occurred in the extracellular regions known to be targets of viruses but also cell-cell interactions.^{63,64} Besides the adaptive selection, these genes were also highly expressed. Therefore, both adaptive evolution and regulation of gene expression are effective strategies to increase the fitness of *P. symbiosum*, suggesting a remarkable flexibility to survive under extreme conditions. Previous studies reported that short-term disturbances are prone to induce physiological responses, while a longer duration of changes allows the adaptive evolution by accumulating mutations that increase fitness.⁶⁵ To unravel this, time-series genome and transcriptome records are needed. Our study contributes to the understanding of how microbes with functional defects survive under harsh environments and opens an avenue for revealing how microbes evolve on a shorter time scale by applying population metagenomics and technology.

Limitations of the study

There are several limitations of the study. To date, no studies have reported the cultivation of *P. symbiosum*. This precludes the accurate genome reconstruction and functional annotation of *P. symbiosum*. Genomes reconstructed from metagenomic analyses may inevitably introduce noises during the sequence assembly and genome binning steps. With the absence of additional experimental verification, it is often insufficient to assign

function unambiguously based on sequence similarity searches. Moreover, co-occurrence analyses and identified recent HGT events suggest that *P. symbiosum* and Aigarchaeota, may form symbionts; however, further direct experimental evidence is needed to support this inference.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Protologues
- METHOD DETAILS
 - Sample collection and metagenomic sequencing
 - Metagenome assembly and genome binning
 - Phylogenetic tree reconstruction
 - Functional annotation of genomes
 - Metatranscriptomic analysis
 - CoADR sequences and structure analysis
 - SNP calling and selection pressure calculation
- QUANTIFICATION AND STATISTICAL ANALYSIS
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.celrep.2023.112158>.

ACKNOWLEDGMENTS

This work was financially supported by the National Natural Science Foundation of China (32170014, 91951205, 92251302), Science and Technology Infrastructure work project of China Ministry of Science & Technology (No. 2015FY110100), Natural Science Foundation of Guangdong Province, China (No. 2016A030312003), the Guangdong Province Key Laboratory of Computational Science, the Guangdong Province Computational Science Innovative Research Team, the U.S. National Science Foundation (DEB 1557042), and the U.S. National Aeronautics and Space Administration (80NNSC17K0548).

AUTHOR CONTRIBUTIONS

Conceptualization, Y.N.Q., Z.S.H., and W.J.L.; Investigation, Y.Z.R., A.L., Y.X.L., G.X.N., and Y.L.Q.; Data Curation, Z.S.H., Y.N.Q., Y.Z.R., Y.X.L., A.L., Y.L.Q., and M.P.; Writing - Original Draft, Y.N.Q., B.P.H., and Z.S.H.; Writing - Review & Editing, Y.N.Q., B.P.H., W.S.S., P.N.E., W.J.L., and Z.S.H.; Funding Acquisition, Z.S.H., B.P.H., and W.J.L.; Supervision, Z.S.H. and W.J.L.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 8, 2022

Revised: December 27, 2022

Accepted: February 9, 2023

Published: February 22, 2023

REFERENCES

- De Anda, V., Chen, L.-X., Dombrowski, N., Hua, Z.-S., Jiang, H.-C., Banfield, J.F., Li, W.-J., and Baker, B.J. (2021). Brockarchaeota, a novel archaeal phylum with unique and versatile carbon cycling pathways. *Nat. Commun.* 12, 2404. <https://doi.org/10.1038/s41467-021-22736-6>.
- Evans, P.N., Parks, D.H., Chadwick, G.L., Robbins, S.J., Orphan, V.J., Golding, S.D., and Tyson, G.W. (2015). Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. *Science* 350, 434–438. <https://doi.org/10.1126/science.aac7745>.
- Hua, Z.-S., Wang, Y.-L., Evans, P.N., Qu, Y.-N., Goh, K.M., Rao, Y.-Z., Qi, Y.-L., Li, Y.-X., Huang, M.-J., Jiao, J.-Y., et al. (2019). Insights into the ecological roles and evolution of methyl-coenzyme M reductase-containing hot spring Archaea. *Nat. Commun.* 10, 4574. <https://doi.org/10.1038/s41467-019-12574-y>.
- Jay, Z.J., Beam, J.P., Dlakić, M., Rusch, D.B., Kozubal, M.A., and Inskeep, W.P. (2018). Marsarchaeota are an aerobic archaeal lineage abundant in geothermal iron oxide microbial mats. *Nat. Microbiol.* 3, 732–740. <https://doi.org/10.1038/s41564-018-0163-1>.
- Vanwonterghem, I., Evans, P.N., Parks, D.H., Jensen, P.D., Woodcroft, B.J., Hugenholtz, P., and Tyson, G.W. (2016). Methylophilic methanogenesis discovered in the archaeal phylum Verstraetearchaeota. *Nat. Microbiol.* 1, 16170. <https://doi.org/10.1038/nmicrobiol.2016.170>.
- Guy, L., and Ettema, T.J.G. (2011). The archaeal 'TACK' superphylum and the origin of eukaryotes. *Trends Microbiol.* 19, 580–587. <https://doi.org/10.1016/j.tim.2011.09.002>.
- Rinke, C., Chuvpochina, M., Mussig, A.J., Chaumeil, P.-A., Davin, A.A., Waite, D.W., Whitman, W.B., Parks, D.H., and Hugenholtz, P. (2021). A standardized archaeal taxonomy for the Genome Taxonomy Database. *Nat. Microbiol.* 6, 946–959. <https://doi.org/10.1038/s41564-021-00918-8>.
- Elkins, J.G., Podar, M., Graham, D.E., Makarova, K.S., Wolf, Y., Randau, L., Hedlund, B.P., Brochier-Armanet, C., Kunin, V., Anderson, I., et al. (2008). A korarchaeal genome reveals insights into the evolution of the Archaea. *Proc. Natl. Acad. Sci. USA* 105, 8102–8107. <https://doi.org/10.1073/pnas.0801980105>.
- McKay, L.J., Dlakić, M., Fields, M.W., Delmont, T.O., Eren, A.M., Jay, Z.J., Klingel-Smith, K.B., Rusch, D.B., and Inskeep, W.P. (2019). Co-occurring genomic capacity for anaerobic methane and dissimilatory sulfur metabolisms discovered in the Korarchaeota. *Nat. Microbiol.* 4, 614–622. <https://doi.org/10.1038/s41564-019-0362-4>.
- Shen, Y., Buick, R., and Canfield, D.E. (2001). Isotopic evidence for microbial sulphate reduction in the early Archaean era. *Nature* 410, 77–81. <https://doi.org/10.1038/35065071>.
- Ueno, Y., Yamada, K., Yoshida, N., Maruyama, S., and Isozaki, Y. (2006). Evidence from fluid inclusions for microbial methanogenesis in the early Archaean era. *Nature* 440, 516–519. <https://doi.org/10.1038/nature04584>.
- Amend, J.P., and Shock, E.L. (2001). Energetics of overall metabolic reactions of thermophilic and hyperthermophilic Archaea and Bacteria. *FEMS Microbiol. Rev.* 25, 175–243. <https://doi.org/10.1111/j.1574-6976.2001.tb00576.x>.
- Jenney, F.E., Jr., and Adams, M.W.W. (2008). Hydrogenases of the model hyperthermophiles. *Ann. N. Y. Acad. Sci.* 1125, 252–266. <https://doi.org/10.1196/annals.1419.013>.
- Hedlund, B.P., Thomas, S.C., Dodsworth, J.A., and Zhang, C.L. (2016). Life in high-temperature environments. In *Manual of Environmental Microbiology*, 4th Edn, M.V. Yates, C.H. Nakatsu, R.V. Miller, and S.D. Pillai, eds. (Washington, DC: ASM Press), pp. 863–887. <https://doi.org/10.1128/9781555818821.ch4.3.4>.
- Hua, Z.-S., Qu, Y.-N., Zhu, Q., Zhou, E.-M., Qi, Y.-L., Yin, Y.-R., Rao, Y.-Z., Tian, Y., Li, Y.-X., Liu, L., et al. (2018). Genomic inference of the metabolism and evolution of the archaeal phylum Aigarchaeota. *Nat. Commun.* 9, 2832. <https://doi.org/10.1038/s41467-018-05284-4>.
- Hallam, S.J., Konstantinidis, K.T., Putnam, N., Schleper, C., Watanabe, Y.I., Sugahara, J., Preston, C., de la Torre, J., Richardson, P.M., and DeLong, E.F. (2006). Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum*. *Proc. Natl. Acad. Sci. USA* 103, 18296–18301. <https://doi.org/10.1073/pnas.0608549103>.
- Baker, B.J., De Anda, V., Seitz, K.W., Dombrowski, N., Santoro, A.E., and Lloyd, K.G. (2020). Diversity, ecology and evolution of Archaea. *Nat. Microbiol.* 5, 887–900. <https://doi.org/10.1038/s41564-020-0715-z>.
- Castelle, C.J., Brown, C.T., Anantharaman, K., Probst, A.J., Huang, R.H., and Banfield, J.F. (2018). Biosynthetic capacity, metabolic variety and unusual biology in the CPR and DPANN radiations. *Nat. Rev. Microbiol.* 16, 629–645. <https://doi.org/10.1038/s41579-018-0076-2>.
- Bowers, R.M., Kyrpides, N.C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T.B.K., Schulz, F., Jarett, J., Rivers, A.R., Elie-Fadrosh, E.A., et al. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* 35, 725–731. <https://doi.org/10.1038/nbt.3893>.
- Hedlund, B.P., Chuvpochina, M., Hugenholtz, P., Konstantinidis, K.T., Murray, A.E., Palmer, M., Parks, D.H., Probst, A.J., Reysenbach, A.-L., Rodriguez-R, L.M., et al. (2022). SeqCode: a nomenclatural code for prokaryotes described from sequence data. *Nat. Microbiol.* 7, 1702–1708. <https://doi.org/10.1038/s41564-022-01214-9>.
- Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. <https://doi.org/10.1101/gr.186072.114>.
- Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. <https://doi.org/10.1093/molbev/msu300>.
- Rawlings, N.D., Barrett, A.J., Thomas, P.D., Huang, X., Bateman, A., and Finn, R.D. (2018). The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* 46, D624–D632. <https://doi.org/10.1093/nar/gkx1134>.
- Grochowski, L.L., Xu, H., and White, R.H. (2005). Ribose-5-phosphate biosynthesis in *Methanocaldococcus jannaschii* occurs in the absence of a pentose-phosphate pathway. *J. Bacteriol.* 187, 7382–7389. <https://doi.org/10.1128/JB.187.21.7382-7389.2005>.
- Orita, I., Sato, T., Yurimoto, H., Kato, N., Atomi, H., Imanaka, T., and Sakai, Y. (2006). The ribulose monophosphate pathway substitutes for the missing pentose phosphate pathway in the archaeon *Thermococcus kodakaraensis*. *J. Bacteriol.* 188, 4698–4704. <https://doi.org/10.1128/JB.00492-06>.
- Orita, I., Yurimoto, H., Hirai, R., Kawarabayashi, Y., Sakai, Y., and Kato, N. (2005). The archaeon *Pyrococcus horikoshii* possesses a bifunctional enzyme for formaldehyde fixation via the ribulose monophosphate pathway. *J. Bacteriol.* 187, 3636–3642. <https://doi.org/10.1128/JB.187.11.3636-3642.2005>.
- Tria, F.D.K., Landan, G., and Dagan, T. (2017). Phylogenetic rooting using minimal ancestor deviation. *Nat. Ecol. Evol.* 1, 193. <https://doi.org/10.1038/s41559-017-0193>.
- Dombrowski, N., Williams, T.A., Sun, J., Woodcroft, B.J., Lee, J.-H., Minh, B.Q., Rinke, C., and Spang, A. (2020). Undinarchaeota illuminate DPANN phylogeny and the impact of gene transfer on archaeal evolution. *Nat. Commun.* 11, 3939. <https://doi.org/10.1038/s41467-020-17408-w>.
- Bever, L.E., Hagedoorn, P.-L., Santamaria-Araujo, J.A., Magalon, A., Hagen, W.R., and Schwarz, G. (2008). Function of MoaB proteins in the biosynthesis of the molybdenum and tungsten cofactors. *Biochemistry* 47, 949–956. <https://doi.org/10.1021/bi7020487>.
- Bever, L.E., Hagedoorn, P.-L., and Hagen, W.R. (2009). The bio-inorganic chemistry of tungsten. *Coord. Chem. Rev.* 253, 269–290. <https://doi.org/10.1016/j.ccr.2008.01.017>.

31. Debnar-Daumler, C., Seubert, A., Schmitt, G., and Heider, J. (2014). Simultaneous involvement of a tungsten-containing aldehyde:ferredoxin oxidoreductase and a phenylacetaldehyde dehydrogenase in anaerobic phenylalanine metabolism. *J. Bacteriol.* 196, 483–492. <https://doi.org/10.1128/JB.00980-13>.
32. Eser, B.E., Zhang, X., Chanani, P.K., Begley, T.P., and Ealick, S.E. (2016). From suicide enzyme to catalyst: the iron-dependent sulfide transfer in *Methanococcus jannaschii* thiamin thiazole biosynthesis. *J. Am. Chem. Soc.* 138, 3639–3642. <https://doi.org/10.1021/jacs.6b00445>.
33. Hayashi, M., Kobayashi, K., Esaki, H., Konno, H., Akaji, K., Tazuya, K., Yamada, K., Nakabayashi, T., and Nosaka, K. (2014). Enzymatic and structural characterization of an archaeal thiamin phosphate synthase. *Biochim. Biophys. Acta* 1844, 803–809. <https://doi.org/10.1016/j.bba-pap.2014.02.017>.
34. Rodionova, I.A., Vetting, M.W., Li, X., Almo, S.C., Osterman, A.L., and Rodionov, D.A. (2017). A novel bifunctional transcriptional regulator of riboflavin metabolism in Archaea. *Nucleic Acids Res.* 45, 3785–3799. <https://doi.org/10.1093/nar/gkw1331>.
35. Tsyrlunyk, A.O., Andreieva, Y.A., Ruchala, J., Fayura, L.R., Dmytruk, K.V., Fedorovich, D.V., and Sibirny, A.A. (2020). Expression of yeast homolog of the mammal BCRP gene coding for riboflavin efflux protein activates vitamin B2 production in the flavinogenic yeast *Candida famata*. *Yeast* 37, 467–473. <https://doi.org/10.1002/yea.3470>.
36. Ammelburg, M., Hartmann, M.D., Djuranovic, S., Alva, V., Koretke, K.K., Martin, J., Sauer, G., Truffault, V., Zeth, K., Lupas, A.N., et al. (2007). A CTP-dependent archaeal riboflavin kinase forms a bridge in the evolution of cradle-loop barrels. *Structure* 15, 1577–1590. <https://doi.org/10.1016/j.str.2007.09.027>.
37. Herwald, S., Liu, A.Y., Zhu, B.E., Sea, K.W., Lopez, K.M., Sazinsky, M.H., and Crane, E.J., 3rd. (2013). Structure and substrate specificity of the pyrococcal coenzyme A disulfide reductases/polysulfide reductases (CoADR/Psr): implications for S⁰-based respiration and a sulfur-dependent antioxidant system in *Pyrococcus*. *Biochemistry* 52, 2764–2773. <https://doi.org/10.1021/bi3014399>.
38. Ma, K., Schicho, R.N., Kelly, R.M., and Adams, M.W. (1993). Hydrogenase of the hyperthermophile *Pyrococcus furiosus* is an elemental sulfur reductase or sulfhydrogenase: evidence for a sulfur-reducing hydrogenase ancestor. *Proc. Natl. Acad. Sci. USA* 90, 5341–5344. <https://doi.org/10.1073/pnas.90.11.5341>.
39. Mander, G.J., Pierik, A.J., Huber, H., and Hedderich, R. (2004). Two distinct heterodisulfide reductase-like enzymes in the sulfate-reducing archaeon *Archaeoglobus profundus*. *Eur. J. Biochem.* 271, 1106–1116. <https://doi.org/10.1111/j.1432-1033.2004.04013.x>.
40. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. <https://doi.org/10.1093/nar/gkh340>.
41. Zeldovich, K.B., Berezovsky, I.N., and Shakhnovich, E.I. (2007). Protein and DNA sequence determinants of thermophilic adaptation. *PLoS Comput. Biol.* 3, e5. <https://doi.org/10.1371/journal.pcbi.0030005>.
42. Forterre, P. (2002). A hot story from comparative genomics: reverse gyrase is the only hyperthermophile-specific protein. *Trends Genet.* 18, 236–237. [https://doi.org/10.1016/S0168-9525\(02\)02650-1](https://doi.org/10.1016/S0168-9525(02)02650-1).
43. Okada, K., Hidese, R., Fukuda, W., Niitsu, M., Takao, K., Horai, Y., Umezawa, N., Higuchi, T., Oshima, T., Yoshikawa, Y., et al. (2014). Identification of a novel aminopropyltransferase involved in the synthesis of branched-chain polyamines in hyperthermophiles. *J. Bacteriol.* 196, 1866–1876. <https://doi.org/10.1128/JB.01515-14>.
44. Shockley, K.R., Ward, D.E., Chhabra, S.R., Connors, S.B., Montero, C.I., and Kelly, R.M. (2003). Heat shock response by the hyperthermophilic archaeon *Pyrococcus furiosus*. *Appl. Environ. Microbiol.* 69, 2365–2371. <https://doi.org/10.1128/AEM.69.4.2365-2371.2003>.
45. Goldfarb, T., Sberro, H., Weinstock, E., Cohen, O., Doron, S., Charpak-Amikam, Y., Afik, S., Ofir, G., and Sorek, R. (2015). BREX is a novel phage resistance system widespread in microbial genomes. *EMBO J.* 34, 169–183. <https://doi.org/10.15252/embj.201489455>.
46. Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. <https://doi.org/10.1101/gr.092759.109>.
47. Barr, J.J., Auro, R., Furlan, M., Whiteson, K.L., Erb, M.L., Pogliano, J., Stotland, A., Wolkowicz, R., Cutting, A.S., Doran, K.S., et al. (2013). Bacteriophage adhering to mucus provide a non-host-derived immunity. *Proc. Natl. Acad. Sci. USA* 110, 10771–10776. <https://doi.org/10.1073/pnas.1305923110>.
48. Parks, D.H., Chuvochina, M., Rinke, C., Mussig, A.J., Chaumeil, P.-A., and Hugenholtz, P. (2022). GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res.* 50, D785–D794. <https://doi.org/10.1093/nar/gkab776>.
49. Pester, M., and Brune, A. (2007). Hydrogen is the central free intermediate during lignocellulose degradation by termite gut symbionts. *ISME J.* 1, 551–565. <https://doi.org/10.1038/ismej.2007.62>.
50. Jahn, U., Summons, R., Sturt, H., Grosjean, E., and Huber, H. (2004). Composition of the lipids of *Nanoarchaeum equitans* and their origin from its host *Ignicoccus* sp. strain KIN4/I. *Arch. Microbiol.* 182, 404–413. <https://doi.org/10.1007/s00203-004-0725-x>.
51. Li, Y.-X., Rao, Y.-Z., Qi, Y.-L., Qu, Y.-N., Chen, Y.-T., Jiao, J.-Y., Shu, W.-S., Jiang, H., Hedlund, B.P., Hua, Z.-S., et al. (2021). Deciphering symbiotic interactions of “*Candidatus Aenigmarchaeota*” with inferred horizontal gene transfers and co-occurrence networks. *mSystems* 6, e0060621. <https://doi.org/10.1128/mSystems.00606-21>.
52. Buessecker, S., Palmer, M., Lai, D., Dimapilis, J., Mayali, X., Mosier, D., Jiao, J.-Y., Colman, D.R., Keller, L.M., St. John, E., et al. (2022). An essential role for tungsten in the ecology and evolution of a previously uncultivated lineage of anaerobic, thermophilic Archaea. *Nat. Commun.* 13, 3773. <https://doi.org/10.1038/s41467-022-31452-8>.
53. Moran, N.A., and Bennett, G.M. (2014). The tiniest tiny genomes. *Annu. Rev. Microbiol.* 68, 195–215. <https://doi.org/10.1146/annurev-micro-091213-112901>.
54. Foster, J., Ganatra, M., Kamal, I., Ware, J., Makarova, K., Ivanova, N., Bhattacharya, A., Kapatral, V., Kumar, S., Posfai, J., et al. (2005). The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. *PLoS Biol.* 3, e121. <https://doi.org/10.1371/journal.pbio.0030121>.
55. Burstein, D., Sun, C.L., Brown, C.T., Sharon, I., Anantharaman, K., Probst, A.J., Thomas, B.C., and Banfield, J.F. (2016). Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. *Nat. Commun.* 7, 10613. <https://doi.org/10.1038/ncomms10613>.
56. Denef, V.J., and Banfield, J.F. (2012). In situ evolutionary rate measurements show ecological success of recently emerged bacterial hybrids. *Science* 336, 462–466. <https://doi.org/10.1126/science.1218389>.
57. Johnson, P.L.F., and Slatkin, M. (2006). Inference of population genetic parameters in metagenomics: a clean look at messy data. *Genome Res.* 16, 1320–1327. <https://doi.org/10.1101/gr.5431206>.
58. Simmons, S.L., DiBartolo, G., Denef, V.J., Goltsman, D.S.A., Thelen, M.P., and Banfield, J.F. (2008). Population genomic analysis of strain variation in *Leptospirillum* Group II Bacteria involved in acid mine drainage formation. *PLoS Biol.* 6, e177. <https://doi.org/10.1371/journal.pbio.0060177>.
59. Leaché, A.D., and Oaks, J.R. (2017). The utility of single nucleotide polymorphism (SNP) data in phylogenetics. *Annu. Rev. Ecol. Syst.* 48, 69–84. <https://doi.org/10.1146/annurev-ecolsys-110316-022645>.
60. Lasky, J.R., Des Marais, D.L., McKay, J.K., Richards, J.H., Juenger, T.E., and Keitt, T.H. (2012). Characterizing genomic variation of *Arabidopsis thaliana*: the roles of geography and climate. *Mol. Ecol.* 21, 5512–5529. <https://doi.org/10.1111/j.1365-294X.2012.05709.x>.

61. Parchman, T.L., Gompert, Z., Mudge, J., Schilkey, F.D., Benkman, C.W., and Buerkle, C.A. (2012). Genome-wide association genetics of an adaptive trait in lodgepole pine. *Mol. Ecol.* 21, 2991–3005. <https://doi.org/10.1111/j.1365-294X.2012.05513.x>.
62. Vincent, B., Dionne, M., Kent, M.P., Lien, S., and Bernatchez, L. (2013). Landscape genomics in atlantic salmon (*salmo Salar*): searching for gene-environment interactions driving local adaptation. *Evolution* 67, 3469–3487. <https://doi.org/10.1111/evo.12139>.
63. Blanc, G., Ngwamidiba, M., Ogata, H., Fournier, P.-E., Claverie, J.-M., and Raoult, D. (2005). Molecular evolution of Rickettsia surface antigens: evidence of positive selection. *Mol. Biol. Evol.* 22, 2073–2083. <https://doi.org/10.1093/molbev/msi199>.
64. Petersen, L., Bollback, J.P., Dimmic, M., Hubisz, M., and Nielsen, R. (2007). Genes under positive selection in *Escherichia coli*. *Genome Res.* 17, 1336–1343. <https://doi.org/10.1101/gr.6254707>.
65. Walworth, N.G., Zakem, E.J., Dunne, J.P., Collins, S., and Levine, N.M. (2020). Microbial evolutionary strategies in a dynamic ocean. *Proc. Natl. Acad. Sci. USA* 117, 5943–5948. <https://doi.org/10.1073/pnas.1919332117>.
66. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Pribelski, A.D., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. <https://doi.org/10.1089/cmb.2012.0021>.
67. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
68. Kang, D.D., Froula, J., Egan, R., and Wang, Z. (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 3, e1165. <https://doi.org/10.7717/peerj.1165>.
69. Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11, 119. <https://doi.org/10.1186/1471-2105-11-119>.
70. Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., and Kanehisa, M. (2007). KAAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35, W182–W185. <https://doi.org/10.1093/nar/gkm321>.
71. Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. <https://doi.org/10.1038/nmeth.3176>.
72. Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>.
73. Käll, L., Krogh, A., and Sonnhammer, E.L.L. (2007). Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res.* 35, W429–W432. <https://doi.org/10.1093/nar/gkm256>.
74. Bagos, P.G., Tsigirgos, K.D., Plessas, S.K., Liakopoulos, T.D., and Hamodrakas, S.J. (2009). Prediction of signal peptides in archaea. *Protein Eng. Des. Sel.* 22, 27–35. <https://doi.org/10.1093/protein/gzn064>.
75. Yu, N.Y., Wagner, J.R., Laird, M.R., Melli, G., Rey, S., Lo, R., Dao, P., Sahinalp, S.C., Ester, M., Foster, L.J., et al. (2010). PSORTb 3.0: improved protein subcellular localization prediction with refined localization sub-categories and predictive capabilities for all prokaryotes. *Bioinformatics* 26, 1608–1615. <https://doi.org/10.1093/bioinformatics/btq249>.
76. Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. <https://doi.org/10.1093/bioinformatics/btl158>.
77. Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858. <https://doi.org/10.1038/nprot.2015.053>.
78. Eddy, S.R. (2011). Accelerated profile HMM searches. *PLoS Comput. Biol.* 7, e1002195. <https://doi.org/10.1371/journal.pcbi.1002195>.
79. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>.
80. Lagesen, K., Hallin, P., Rødland, E.A., Stærfeldt, H.-H., Rognes, T., and Ussery, D.W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35, 3100–3108. <https://doi.org/10.1093/nar/gkm160>.
81. Pruesse, E., Peplies, J., and Glöckner, F.O. (2012). SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* 28, 1823–1829. <https://doi.org/10.1093/bioinformatics/bts252>.
82. Olm, M.R., Brown, C.T., Brooks, B., and Banfield, J.F. (2017). dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* 11, 2864–2868. <https://doi.org/10.1038/ismej.2017.126>.
83. Wu, M., and Scott, A.J. (2012). Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. *Bioinformatics* 28, 1033–1034. <https://doi.org/10.1093/bioinformatics/bts079>.
84. Thompson, J.D., Gibson, T.J., and Higgins, D.G. (2002). Multiple sequence alignment using ClustalW and ClustalX. *Curr. Protoc. Bioinformatics Chapter 2, Unit 2.3*. <https://doi.org/10.1002/0471250953.bi0203s00>.
85. Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44, W242–W245. <https://doi.org/10.1093/nar/gkw290>.
86. Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191. <https://doi.org/10.1093/bioinformatics/btp033>.
87. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., et al. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46, W296–W303. <https://doi.org/10.1093/nar/gky427>.
88. Lilkova, E., Petkov, P., Ilieva, N., and Litov, L. (2015). The PyMOL Molecular Graphics System, vol 273 (Gaussian, Inc), pp. 10919–10925.
89. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498. <https://doi.org/10.1038/ng.806>.
90. Enright, A.J., Van Dongen, S., and Ouzounis, C.A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* 30, 1575–1584. <https://doi.org/10.1093/nar/30.7.1575>.
91. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. <https://doi.org/10.1093/molbev/mst010>.
92. Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34, W609–W612. <https://doi.org/10.1093/nar/gkl315>.
93. Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. <https://doi.org/10.1093/molbev/msm088>.
94. Makarova, K.S., Wolf, Y.I., and Koonin, E.V. (2015). Archaeal clusters of orthologous genes (arCOGs): an update and application for analysis of shared features between *Thermococcales*, *Methanococcales*, and *Methanobacteriales*. *Life* 5, 818–840. <https://doi.org/10.3390/life5010818>.
95. Saier, M.H., Jr., Reddy, V.S., Tsu, B.V., Ahmed, M.S., Li, C., and Moreno-Hagelsieb, G. (2016). The transporter classification database

- (TCDB): recent advances. *Nucleic Acids Res.* 44, D372–D379. <https://doi.org/10.1093/nar/gkv1103>.
96. Søndergaard, D., Pedersen, C.N.S., and Greening, C. (2016). HydDB: a web tool for hydrogenase classification and analysis. *Sci. Rep.* 6, 34212. <https://doi.org/10.1038/srep34212>.
 97. Hedlund, B.P., Cole, J.K., Williams, A.J., Hou, W., Zhou, E., Li, W., and Dong, H. (2012). A review of the microbiology of the Rehai geothermal field in Tengchong, Yunnan Province, China. *Geosci. Front.* 3, 273–288. <https://doi.org/10.1016/j.gsf.2011.12.006>.
 98. Hua, Z.-S., Han, Y.-J., Chen, L.-X., Liu, J., Hu, M., Li, S.-J., Kuang, J.-L., Chain, P.S.G., Huang, L.-N., and Shu, W.-S. (2015). Ecological roles of dominant and rare prokaryotes in acid mine drainage revealed by metagenomics and metatranscriptomics. *ISME J.* 9, 1280–1294. <https://doi.org/10.1038/ismej.2014.212>.
 99. Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P., and Parks, D.H. (2019). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36, 1925–1927. <https://doi.org/10.1093/bioinformatics/btz848>.
 100. Parks, D.H., Rinke, C., Chuvochina, M., Chaumeil, P.-A., Woodcroft, B.J., Evans, P.N., Hugenholtz, P., and Tyson, G.W. (2017). Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat. Microbiol.* 2, 1533–1542. <https://doi.org/10.1038/s41564-017-0012-7>.
 101. Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J., Butterfield, C.N., Hemsdorf, A.W., Amano, Y., Ise, K., et al. (2016). A new view of the tree of life. *Nat. Microbiol.* 1, 16048. <https://doi.org/10.1038/nmicrobiol.2016.48>.
 102. Hoshino, Y., and Gaucher, E.A. (2018). On the origin of isoprenoid biosynthesis. *Mol. Biol. Evol.* 35, 2185–2197. <https://doi.org/10.1093/molbev/msy120>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Metagenomic and Metatranscriptomic sequencing data	This paper	BioProject: PRJNA544494
Metagenome-assembled genomes	This paper	Accession numbers: GCA_024720835.1, GCA_024720815.1, GCA_024720855.1, GCA_024720875.1, GCA_024720895.1, GCA_024720915.1, GCA_024720935.1, GCA_024720955.1, GCA_024720975.1, GCA_024720985.1
Software and algorithms		
SPAdes v3.9.0	Bankevich et al. ⁶⁶	https://github.com/ablab/spades
Samtools v1.3.1	Li et al. ⁶⁷	http://www.htslib.org/
MetaBAT v2.12.1	Kang et al. ⁶⁸	https://bitbucket.org/berkeleylab/metabat/src/master/
CheckM v1.0.12	Parks et al. ²¹	https://ecogenomics.github.io/CheckM/
Prodigal v2.6.3	Hyatt et al. ⁶⁹	https://github.com/hyatt/Prodigal
KAAS v2.1	Moriya et al. ⁷⁰	https://www.genome.jp/kegg/kaas/
DIAMOND v0.8.22	Buchfink et al. ⁷¹	https://www.crystalimpact.com/diamond/
BBmap v35.85		http://sourceforge.net/projects/bbmap/
InterProScan v5.36–75.0	Jones et al. ⁷²	https://www.ebi.ac.uk/interpro/
Phobius	Käll et al. ⁷³	https://phobius.sbc.su.se/
PRED-SIGNAL	Bagos et al. ⁷⁴	http://bioinformatics.biol.uoa.gr/PRED-SIGNAL/
PSORTb v3.0.2	Yu et al. ⁷⁵	https://www.psorth.org/psorth/
CD-HIT v4.6	Li and Godzik ⁷⁶	http://weizhong-lab.ucsd.edu/cd-hit/
Phyre2	Kelley et al. ⁷⁷	http://www.sbg.bio.ic.ac.uk/phyre2
Sickle v1.33		https://github.com/najoshi/sickle
HMMER v3.1b2	Eddy ⁷⁸	http://hmmer.org/
MUSCLE v3.8.31	Edgar ⁴⁰	https://www.ebi.ac.uk/Tools/msa/muscle/
TrimAL v1.4.rev22	Capella-Gutiérrez et al. ⁷⁹	https://vicfero.github.io/trimal/
IQ-tree v1.6.10	Nguyen et al. ²²	http://www.iqtree.org/
RNAmmmer v1.2	Lagesen et al. ⁸⁰	https://services.healthtech.dtu.dk/service.php?RNAmmmer-1.2
SINA v1.2.11	Pruesse et al. ⁸¹	https://www.arb-silva.de/aligner
dRep v2.3.2	Olm et al. ⁸²	https://github.com/MrOlm/drep
AMPHORA2	Wu and Scott ⁸³	https://github.com/wu-lab-uva/AMPHORA2
ClustalW v1.2.4	Thompson et al. ⁸⁴	http://www.clustal.org/
ITOL v3	Letunic and Bork ⁸⁵	https://itol.embl.de/
FigTree v1.4.4		https://github.com/rambaut/figtree
Jalview v2	Waterhouse et al. ⁸⁶	https://www.jalview.org/
SWISS-MODEL	Waterhouse et al. ⁸⁷	https://swissmodel.expasy.org/
PyMOL v2	Lilko et al. ⁸⁸	https://pymol.org/2/
GATK v4.1.4.1	DePristo et al. ⁸⁹	https://gatk.broadinstitute.org/hc/en-us
Picard v2.21.4		http://broadinstitute.github.io/picard/
MCL v14-137	Enright et al. ⁹⁰	https://github.com/micans/mcl
MAFFT v7.487	Katoh and Standley ⁹¹	https://mafft.cbrc.jp/alignment/software/
PAL2NAL	Suyama et al. ⁹²	http://www.bork.embl.de/pal2nal/
PAML v4.9	Yang ⁹³	http://abacus.gene.ucl.ac.uk/software/paml.html
R v4.1.1	The R Foundation	https://www.r-project.org/

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Circos v0.69-9	Krzywinski et al. ⁴⁶	http://circos.ca/
Other		
arCOG database	Makarova et al. ⁹⁴	ftp://ftp.ncbi.nih.gov/pub/wolf/COGs/arCOG/
Merops database (Release 12.1)	Rawlings et al. ²³	https://www.ebi.ac.uk/merops/download_list.shtml
Transporter Classification Database	Saier et al. ⁹⁵	https://www.tcdb.org/download.php
Hydrogenase database	Sondergaard et al. ⁹⁶	https://services.birc.au.dk/hyddb/browser/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Zheng-Shuang Hua (hzhengsh@ustc.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Data have been deposited at NCBI GenBank and are publicly available as of the data of publication. Accession numbers are list in the [key resources table](#).

This paper does not report original code.

Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Protologues

Description of *Panguiarchaeum* gen. Nov

Panguiarchaeum (Pan.gu.ar.chae'um N.L. n. Pangu, creation figure in Chinese mythology who separated heaven from Earth; N.L. neut. n. archaeum [from Gr. adj. archaios, -e, -on] ancient archaeon; N.L. neut. n. *Panguiarchaeum* an archaeon created by Pangu).

The genus description is the same as that of the nomenclatural type *Panguiarchaeum symbiosum*.

Description of *Panguiarchaeum symbiosum* sp. nov

Panguiarchaeum symbiosum [sym.bi.o'sum. Gr. masc./fem. adj. symbios, companion; N.L. neut. adj. symbiosum, living together with, symbiotic (referring to the proposed symbiotic lifestyle)]

Archaea that inhabit circumneutral, terrestrial geothermal systems and are likely hyperthermophiles. Genomes belonging to this species are predicted to range between 1.25 Mb and 1.43 Mb. The GC content for the species range between 40.06 and 40.45%. Genomes of this species lack annotated pathways for synthesis of membrane lipids, many amino acids, and purines, suggesting they are likely to be symbiotic. Predicted to be obligately anaerobic and fermentative. Genomic assemblies for this species originated from sediments from geothermal springs in both the Rehai and Diantan geothermal areas in Tengchong, Yunnan Province, China. The nomenclatural type for the species is the genomic assembly JZ-1.bin_83^T (GCA_024720975.1).

Description of *Panguiarchaeaceae* fam. Nov

Panguiarchaeaceae (Pan.gu.ar.chae.a.ce'ae. N.L. neut. n. *Panguiarchaeum* the type genus of the family; -aceae ending to denote a family; N.L. fem. pl. n. *Panguiarchaeaceae* the *Panguiarchaeum* family).

The family description is the same as that of the only species assigned to the family, *Panguiarchaeum symbiosum*.

Description of *Panguiarchaeales* ord. Nov

Panguiarchaeales (Pan.gu.ar.chae.a'les. N.L. neut. n. *Panguiarchaeum* the type genus of the order; -ales ending to denote an order; N.L. fem. pl. n. *Panguiarchaeales* the *Panguiarchaeum* order).

The order description is the same as that of the only species assigned to the order, *Panguiarchaeum symbiosum*.

METHOD DETAILS

Sample collection and metagenomic sequencing

A total of ten samples were obtained in five thermal spring sediments (JZ (JinZe) from Diantan Town, BF (BianFu), SRBZ (ShuiReBaoZhaQu), SZTDM (ShiZiTuDuiMian) and ZZQ (ZhenZhuQuan) from Rehai Town located at Tengchong County, Yunnan, China.⁹⁷ The same procedures for sample collection, DNA extraction and metagenomic sequencing were conducted as previously

described.¹⁵ The sampling date, geographic locations (latitude and longitude) and *in situ* physiochemical parameters are recorded in [additional resources](#).

Metagenome assembly and genome binning

Raw sequencing data generated from each sediment sample were processed to filter out poor quality reads as described in a previous study.⁹⁸ The remaining high-quality reads from each samples were assembled separately using SPAdes⁶⁶ v3.9.0 with the following parameters: `-meta -k 21,33,55,77,99,127`. Genomic binning of the individual assemblies was performed with the consideration of both tetranucleotide frequency and sequence depth. First, high-quality reads were mapped to metagenomic assemblies using BBmap v35.85 (<http://sourceforge.net/projects/bbmap/>). Then, the generated bam files were sorted using SAMtools⁶⁷ v1.3.1. Sequence depth was calculated using the `"jgi_summarize_bam_contig_depths"` program in MetaBAT⁶⁸ v2.12.1. Preliminary bins were generated using the `"metabat"` program in MetaBAT. Genome quality including completeness, contamination and heterogeneity was evaluated using CheckM²¹ v1.0.12. Scaffolds with abnormal coverage information were manually eliminated. To improve MAG quality, the high-quality reads mapped to each optimized MAG were reassembled using SPAdes with the same K-mers as above and allowing read correction (careful mode). The taxonomy of assembled MAGs were obtained using GTDB-TK^{48,99} v1.6.0 and those MAGs identified as members of the *Panguiarchaeales* were further analyzed.

Phylogenetic tree reconstruction

Phylogenomic trees

A total of 464 archaeal reference genomes were downloaded from NCBI database (<https://www.ncbi.nlm.nih.gov/>) and utilized to build phylogenomic trees (see details in [additional resources](#)). We applied different marker sets including a set of 122 archaeal-specific marker proteins,¹⁰⁰ 16 universal ribosomal proteins¹⁰¹ and 56 conserved archaeal-specific marker proteins²⁸ to determine the phylogenetic placement of *Panguiarchaeales* MAGs. Marker proteins for each genome were identified using `'hmmsearch'` command in HMMER⁷⁸ v3.1b2 (settings: `-E 1e-5`). Sequences from individual marker sets were aligned using MUSCLE⁴⁰ v3.8.31 with 100 iterations. Poorly aligned regions were filtered using TrimAL⁷⁹ v1.4.rev22 with the parameters `-gt 0.9 -cons 50`. Phylogenomic trees were constructed from the concatenated alignments using IQ-TREE²² v1.6.10 with LG + F + R10 as the best model.

16S rRNA gene tree

To discover the distribution pattern of *Panguiarchaeales*, all 16S rRNA gene sequences publicly available in the NCBI-nt database related to this lineage were downloaded for phylogenetic inference. Specifically, only sequences with similarity >75% and covering >90% of the whole region were considered. 16S rRNA sequences from MAGs in the present study were called using RNAmmer⁸⁰ v1.2. Representatives from Euryarchaeota were utilized as outgroups. Sequences from TACK superphylum were also included to better describe the phylogenetic placement of *Panguiarchaeales*. All collected sequences were aligned using the online tool SINA.⁸¹ Poorly aligned regions where columns exclusively constituted by gaps were eliminated using TrimAL v1.4.rev22 (`-automated1`). Phylogeny was inferred using IQ-TREE v1.6.10 with GTR + F + R6 as the best model.

Phi gene tree

Reference datasets of the *phi* gene were derived from the NCBI RefSeq database downloaded on May 21st, 2019. Briefly, the genome quality was evaluated via CheckM v1.0.12 for all publicly available complete and draft genomes. Genomes with completeness <80% or contamination <5% were discarded. Then, high quality genomes were dereplicated using dRep⁸² v2.3.2. Representative genomes were selected from genome groups that had an average nucleotide identity >90%. The *phi* gene sequences from this study were used to construct a local database using the `"hmmbuild"` program in HMMER v3.1b2. Subsequently, a set of *phi* genes was identified from dereplicated genomes using AMPHORA2.⁸³ The predicted genes were further searched against the KEGG database using DIAMOND (E-values < 1e⁻⁵). Only hits matching K08094 (*hxlB*) and K13831 (*hps-phi*) were kept for further analysis. To reduce the complexity of the phylogenetic tree, all genes verified to be *phi* homologues were clustered using CD-HIT at a sequence identity of 0.9. Finally, 1464 genes that satisfied these criteria were retained and used for further analyses. Among them, 24 were identified as genes containing two sugar isomerase (SIS) domains. Individual domains from those genes were selected and combined with the remaining gene sequences to construct a phylogenetic tree and resulted in a collection of 1488 gene fragments. Sequences were aligned and trimmed using MUSCLE v3.8.31 and TrimAL v1.4.rev22 (`-gt 0.05 -cons 50`). Phylogeny was generated using IQ-TREE v1.6.10 with LG + R10 as the best model.

Hydrogenase gene tree

Reference sequences of Groups I, II, III, and IV NiFe hydrogenases were downloaded from the Hydrogenase database.⁹⁶ The obtained protein sequences belonging to different groups were aligned using MUSCLE v3.8.31 separately, and individual alignments were merged using ClustalW⁸⁴ v1.2.4. The phylogeny was generated using IQ-TREE v1.6.10 with the best parameters determined by internal model selection (LG + R10).

Mvk gene tree

Mevalonate kinase protein sequences encoded by *mvk* were extracted from the 408 selected archaeal genomes using hmmsearch v3.1b2 (settings: `-E 1e-5`). Bacterial mevalonate kinase sequences were downloaded from NCBI database according to the accession number extracted from a previous study.¹⁰² The sequences were aligned using MUSCLE and trimmed using TrimAL (`-gt 0.9 -cons 50`). A phylogenetic tree was generated with IQ-TREE (`-m LG + R10 -bb 1000`).

All generated trees in Newick format were visualized and decorated using iTOL⁸⁵ v3 and FigTree v1.4.4 (<https://github.com/rambaut/figtree>).

Functional annotation of genomes

For metabolic function analyses, both genomes from the present study and public databases were used for functional annotation. A total of 398 genomes publicly available in the NCBI database were manually selected and downloaded that cover the current archaeal diversity at the phylum level (see details in [additional resources](#)). The protein-coding genes (CDSs) were predicted using Prodigal⁶⁹ v2.6.3 and functional annotations were conducted by searching against several public databases. All CDSs were assigned to KOs using KAAS⁷⁰ v2.1 with default settings. ArCOGs were searched against the arCOG database⁹⁴ using DIAMOND⁷¹ v0.8.22 with E-values < 1e-5. PFAMs and TIGRFAMs for all CDS were determined using InterProScan⁷² v5.36–75.0 package with the following parameters: -dp -f tsv -goterms -pa. Peptidases/aminotransferases and transporters in *Panguiarchaeales* MAGs were identified by searching against the Merops database²³ (Release 12.1) and Transporter Classification Database⁹⁵ using the DIAMOND program with an E-value < 1e-5. Phobius,⁷³ PRED-SIGNAL⁷⁴ and PSORTb⁷⁵ v3.0.2 software were used to predict the locations of peptidases.

Candidate proteins potentially involved in cell-cell interactions were filtered according to IPR descriptions and amino acid length as previously described.²⁸ The selected proteins were clustered using CD-HIT⁷⁶ v4.6 at a sequence identity of 90%. Then, the resulting representative sequences were searched using Phyre2⁷⁷ web version (<http://www.sbg.bio.ic.ac.uk/phyre2>), and results are recorded in [Table S4](#).

Metatranscriptomic analysis

Hot spring sediment samples collected in Jan 2021 and Jul 2022 were conducted for metatranscriptomic sequencing. All samples were failed to obtain sufficient total RNA except JZ-2 and JZ-3 collected in Jan 2021 and Jul 2022 respectively. Total RNA was extracted using an RNeasy PowerSoil Total RNA kit (QIAGEN) and sent for metatranscriptomic sequencing without removal of rRNA. The generated paired-end reads (~16 Gbp) were quality-controlled using Sickle v1.33 (<https://github.com/najoshi/sickle>) with the following parameters: -t sanger -quiet -l 50. Trimmed reads were mapped to the representative *Panguiarchaeales* MAG JZ-1 bin_83 using BBmap v35.85 with the same parameters as above and RPKM values were also generated in this program.

CoADR sequences and structure analysis

Two copies of CoADR genes that are divergent from each other at the sequence level were detected in *Panguiarchaeales* MAGs. To determine their putative functions, phylogenetic trees of CoADR homologs were built, and their 3-D structures were examined. The CoADR homologs in *Panguiarchaeales* were searched against the NCBI-nr database and the top 500 BLAST hits of each were collected. Redundant sequences were eliminated using CD-hit with the following parameters: -c 0.9 -as 1. Sequence alignments were performed using MUSCLE and visualized using Jalview⁸⁶ v2. Poorly aligned regions were discarded, and phylogeny was generated with the same procedures as described above (see [phi gene tree](#)). The 3-D structures of the two genes were predicted using the online tool SWISS-MODEL,⁸⁷ and further used for 3-D structure visualization, manipulation and comparison with experimentally verified CoADRs using PyMOL⁸⁸ v2 software.

SNP calling and selection pressure calculation

SNP calling was conducted by mapping sequenced reads to the reference genome. The MAG JZ-1 bin_83 with the highest genome quality was selected as the reference. The quality reads were mapped to the reference genome using BBMap v35.85 with the same parameters as above. The generated bam files were sorted using samtools v1.3.1. SNPs and indels were identified and split into individual files using the “HaplotypeCaller” algorithm in GATK⁸⁹ v4.1.4.1 package and the “SplitVcfs” program in Picard v2.21.4 (<http://broadinstitute.github.io/picard/>). Only loci with Phred quality ≥ 20 were counted as effective variants.

All-vs-all BLAST searches were conducted to obtain orthologues among the ten *Panguiarchaeales* MAGs. Only BLAST hits with identity ≥ 90% and aligned length ≥ 100 bp were kept for later analyses. Genes with gaps were excluded due to the high AALs between pairwise *Panguiarchaeales* MAGs. All MAGs except BF2 bin_16 were detected to contain at most one gene with gaps. This results in the elimination of 32 (0.25%) gene sequences with gaps and therefore, the low proportion of discarded genes will not affect the general pattern of later analysis. All BLAST hits were clustered into gene families using MCL⁹⁰ v14-137 with the following parameters: -l 1.4. Gene families with individuals < 5 were discarded. This resulted in the detection of 1,000 gene families covering a total of 9,371 genes (74.3%). Among them, 230 gene families with identical sequences were eliminated from later analyses. The remaining gene families were aligned using MAFFT⁹¹ v7.487. The codon-based alignments were constructed using the PAL2NAL⁹² script with nucleotide sequences and the aligned amino acid sequences as inputs. The amino acid-based phylogenies were built for each gene family using the IQ-TREE v1.6.10.

Natural selection was detected among the identified gene families using the “codeml” program in the PAML⁹³ v4.9 application suite. Nssites = 0 was performed to determine the fixed dN/dS, though a single ω for all sites is probably wrong in every functional protein. Positive selection on codons was examined by applying random-site models M1a (neutral evolution), M2a (positive selection), M7 (neutral with beta distribution) and M8 (positive selection with beta distribution) which allow the calculation of dN/dS (ω , nonsynonymous/synonymous ratio) with varied values among sites (model = 0, Nssites = 1, 2, 7, 8). Bayes Empirical Bayes

estimates from the model M8 giving a posterior probability >0.95 were used to identify sites under positive selection. A likelihood ratio test was conducted for each candidate family to fit the positive selection model over the neutral and negative selection model (M1a vs. M2a and M7 vs. M8) if the chi-square p value <0.05 . We applied the Benjamini-Hochberg approach to correct p values with the “*p.adjust*” function in the “states” package in R 4.0.3. One gene was considered as evidence for positive selection if the FDR-corrected p-value from either nested model (M1a vs. M2 or M7 vs. M8) < 0.05 .

QUANTIFICATION AND STATISTICAL ANALYSIS

In Figure S7, the Spearman correlation test was performed in R 4.0.3. The figure legend states the sample size ($n = 43$). Positively selected genes were determined using Chi-square test performed by “chi2” program in PAML v4.9 and p values were adjusted using “*p.adjust*” function in the “states” package in R 4.0.3. All p values <0.05 were considered statistically significant.

ADDITIONAL RESOURCES

Additional supplementary files including physicochemical characteristics of sampling sites, archaeal reference genomes used for phylogenetic inference and comparative genomic analyses, metabolic comparisons of archaeal reference genomes are available in FigShare repository at: [10.6084/m9.figshare.21782240](https://figshare.com/10.6084/m9.figshare.21782240).