AMERICAN SOCIETY FOR MICROBIOLOGY | mSystems®

RESEARCH ARTICLE

Deciphering Symbiotic Interactions of "*Candidatus* Aenigmarchaeota" with Inferred Horizontal Gene Transfers and Co-occurrence Networks

Yu-Xian Li,[a] Yang-Zhi Rao,[a] Yan-Ling Qi,[a] Yan-Ni Qu,[a] Ya-Ting Chen,[a] Jian-Yu Jiao,[a] Wen-Sheng Shu,[b] Hongchen Jiang,[c] Brian P. Hedlund,[d,e] Zheng-Shuang Hua,[f] Wen-Jun Li[a,g]

[a]State 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, People's Republic of China

[b]School of Life Sciences, South China Normal University, Guangzhou, People's Republic of China

[c]State Key Laboratory of Biogeology and Environmental Geology, China University of Geosciences, Wuhan, People's Republic of China

[d]School of Life Sciences, University of Nevada Las Vegas, Las Vegas, Nevada, USA

[e]Nevada Institute of Personalized Medicine, University of Nevada Las Vegas, Las Vegas, Nevada, USA

[f]Department of Environmental Science and Engineering, University of Science and Technology of China, Hefei, People's Republic of China

[g]State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi, People's Republic of China

Yu-Xian Li and Yang-Zhi Rao contributed equally to this work. Author order was determined based on alphabetical order of authors' last names.

ABSTRACT "*Candidatus* Aenigmarchaeota" ("*Ca.* Aenigmarchaeota") represents one of the earliest proposed evolutionary branches within the *Diapherotrites*, *Parvarchaeota*, *Aenigmarchaeota*, *Nanoarchaeota*, and *Nanohaloarchaeota* (DPANN) superphylum. However, their ecological roles and potential host-symbiont interactions are still poorly understood. Here, eight metagenome-assembled genomes (MAGs) were reconstructed from hot spring ecosystems, and further in-depth comparative and evolutionary genomic analyses were conducted on these MAGs and other genomes downloaded from public databases. Although with limited metabolic capacities, we reported that "*Ca.* Aenigmarchaeota" in thermal environments harbor more genes related to carbohydrate metabolism than "*Ca.* Aenigmarchaeota" in nonthermal environments. Evolutionary analyses suggested that members from the *Thaumarchaeota*, *Aigarchaeota*, *Crenarchaeota*, and *Korarchaeota* (TACK) superphylum and *Euryarchaeota* contribute substantially to the niche expansion of "*Ca.* Aenigmarchaeota" via horizontal gene transfer (HGT), especially genes related to virus defense and stress responses. Based on co-occurrence network results and recent genetic exchanges among community members, we conjectured that "*Ca.* Aenigmarchaeota" may be symbionts associated with one MAG affiliated with the genus *Pyrobaculum*, though host specificity might be wide and variable across different "*Ca.* Aenigmarchaeota" organisms. This study provides significant insight into possible DPANN-host interactions and ecological roles of "*Ca.* Aenigmarchaeota."

IMPORTANCE Recent advances in sequencing technology promoted the blowout discovery of super tiny microbes in the *Diapherotrites*, *Parvarchaeota*, *Aenigmarchaeota*, *Nanoarchaeota*, and *Nanohaloarchaeota* (DPANN) superphylum. However, the unculturable properties of the majority of microbes impeded our investigation of their behavior and symbiotic lifestyle in the corresponding community. By integrating horizontal gene transfer (HGT) detection and co-occurrence network analysis on "*Candidatus* Aenigmarchaeota" ("*Ca.* Aenigmarchaeota"), we made one of the first attempts to infer their putative interaction partners and further decipher the potential functional and genetic interactions between the symbionts. We revealed that HGTs contributed by members from the *Thaumarchaeota*, *Aigarchaeota*, *Crenarchaeota*, and *Korarchaeota* (TACK) superphylum and *Euryarchaeota* conferred "*Ca.* Aenigmarchaeota" with the

Citation Li Y-X, Rao Y-Z, Qi Y-L, Qu Y-N, Chen Y-T, Jiao J-Y, Shu W-S, Jiang H, Hedlund BP, Hua Z-S, Li W-J. 2021. Deciphering symbiotic interactions of "*Candidatus* Aenigmarchaeota" with inferred horizontal gene transfers and co-occurrence networks. mSystems 6:e00606-21. https://doi.org/10.1128/mSystems.00606-21.

Editor Thulani P. Makhalanyane, University of Pretoria

Copyright © 2021 Li et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Zheng-Shuang Hua, hzhengsh@foxmail.com, or Wen-Jun Li, liwenjun3@mail.sysu.edu.cn.

Received 14 May 2021
Accepted 9 July 2021
Published 27 July 2021

Downloaded from https://journals.asm.org/journal/msystems on 27 August 2021 by 71.38.0.28.

ability to survive under different environmental stresses, such as virus infection, high temperature, and oxidative stress. This study demonstrates that the interaction partners might be inferable by applying informatics analyses on metagenomic sequencing data.

KEYWORDS "*Ca.* Aenigmarchaeota", DPANN, symbiont, horizontal gene transfer, co-occurrence network, coevolution network

With advances in sequencing technologies and bioinformatic approaches, insight into the "unseen majority" prokaryotes has become possible, even when they inhabit complex microbial communities, leading to a tremendous expansion of known archaeal diversity (1–7). Among recently proposed major archaeal lineages, the *Diapherotrites*, *Parvarchaeota*, *Aenigmarchaeota*, *Nanoarchaeota*, and *Nanohaloarchaeota* (DPANN) superphylum has inspired considerable research attention, which has uncovered their surprisingly small genome sizes, lack of genes associated with core biosynthetic pathways (3, 8, 9), and extensive phylogenetic and functional diversity (10–12). "*Candidatus* Aenigmarchaeota" ("*Ca.* Aenigmarchaeota"), which represent the "A" of the DPANN superphylum, were first uncovered and named as the "Deep Sea Euryarchaeotic Group (DSEG)" (13). Later, based on single-amplified genomes (SAGs), this lineage was defined and proposed as a novel phylum (2). Other studies integrating metagenomic and metatranscriptomic sequencing revealed that this phylum lacks many essential metabolic pathways and may possess fermentative and symbiotic lifestyles (3, 8). However, our understanding of the metabolic characteristics, functional diversity, and potential host-symbiont interactions of "*Ca.* Aenigmarchaeota" is far from sufficient.

Here, we apply comparative and evolutionary genomics analyses on eight new metagenome-assembled genomes (MAGs) along with 15 publicly available genomes to fill these gaps. Our study reveals a symbiotic lifestyle for "*Ca.* Aenigmarchaeota" based on the absence of many genes involved in core metabolic pathways. Further analyses suggest that the occurrence of horizontal gene transfer (HGT) improves the competitiveness of "*Ca.* Aenigmarchaeota" by expanding their gene repertoires relevant to stress response and virus defense. We also integrate the HGT inference and co-occurrence network construction to reveal potential functional and genetic interactions between "*Ca.* Aenigmarchaeota" and other microbes.

RESULTS AND DISCUSSION

Phylogeny and distribution of "*Ca.* Aenigmarchaeota." Eight MAGs of "*Ca.* Aenigmarchaeota" were successfully reconstructed from five hot spring sediment samples collected in Tengchong county in Yunnan, China (Fig. 1a), including four from a single sample from Diretiyanqu-6 (DRTY-6) and one for each of the other springs Diretiyanqu-7 (DRTY-7), Gumingquan (GMQ), Qiaoquan (QQ), and Jinze (JZ-2) (Table 1) (14). "*Ca.* Aenigmarchaeota" represents a rare group in hot spring ecosystems with relative abundances of all MAGs of $<0.4\%$ (Fig. 1b). Most MAGs are of high quality, with completeness of $>90\%$, nearly no contamination, and detectable 16S rRNAs and tRNAs (>21) (Table 1; Data Set S1 in the supplemental material) (15). Compared to MAGs of "*Ca.* Aenigmarchaeota" from other studies (Table S1), they have smaller genome sizes (0.64 versus 0.86 mega base pairs [Mbp]; Mann-Whitney *U* test, $P = 0.0003$; Fig. S1) and a lower range of GC content (average 31.74% versus 38.59%; Mann-Whitney *U* test, $P = 0.012$) (2, 3, 8); they also harbor a smaller number of genes (752 versus 1,070; Mann-Whitney *U* test, $P = 0.002$), shorter average gene length (771 versus 683 bp; Mann-Whitney *U* test, $P = 0.0005$), and remarkably high coding density (88 to 94.6%) and percentage of overlapping genes ($\sim$20.6%) (Table 1). This is consistent with previous studies suggesting that thermophiles harbor small genome sizes as a result of genomic streamlining due to high fitness costs of life at high temperatures (16). Both whole-genome-based phylogenomic and 16S rRNA gene-based phylogenetic analyses revealed that the eight MAGs from this study branched within the phylum "*Ca.* Aenigmarchaeota" with high bootstrap confidences (Fig. 1c; Fig. S2). 16S rRNA sequences of "*Ca.* Aenigmarchaeota" were retrieved from the NCBI-nr database and used to

Downloaded from https://journals.asm.org/journal/msystems on 27 August 2021 by 71.38.0.28.

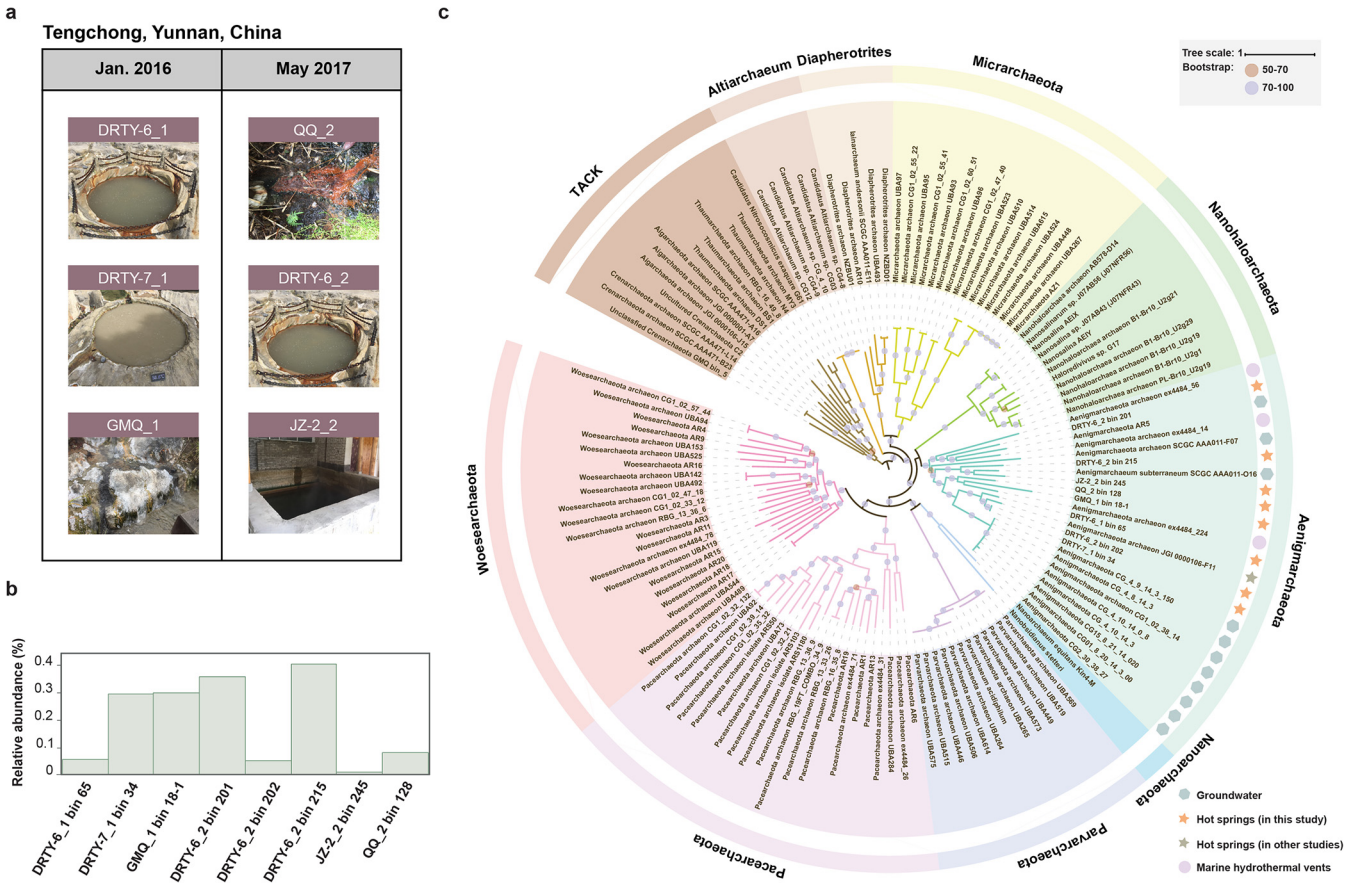


FIG 1 Phylogenetic analysis of reconstructed genomes of "*Ca.* Aenigmarchaeota" sampled from hot spring sediments in Yunnan, China. (a) Sampling sites of "*Ca.* Aenigmarchaeota" in Tengchong, Yunnan, China. Hot spring sediments from a total of five sites were collected in January 2016 and May 2017. (b) Relative abundances of "*Ca.* Aenigmarchaeota" MAGs were calculated by the coverage of the scaffolds of each MAG over the coverage of all the scaffolds of the corresponding metagenome. (c) Phylogenetic placement of the reconstructed MAGs. Maximum likelihood phylogeny of eight "*Ca.* Aenigmarchaeota" MAGs in this study and reference genomes from TACK and DPANN superphyla. Phylogeny was constructed based on a concatenated set of 16 ribosomal proteins with 1,000 bootstrap iterations.

illustrate the geographical distribution (see Materials and Methods). The results demonstrated that "*Ca.* Aenigmarchaeota" represented an evolutionarily diverse group that inhabits a broad range of ecosystems (Data Set S2), including freshwater (40.96%), marine water (27.71%), hot springs (7.63%), hydrothermal vents (6.83%), and groundwater (7.63%). A minor portion of the 16S rRNA gene sequences was retrieved from hypersaline lakes and soils (<5%).

Metabolic features of "*Ca.* Aenigmarchaeota." Based on 8 MAGs from this study and 3 SAGs and 12 MAGs from previous studies, we constructed the metabolic pathways of "*Ca.* Aenigmarchaeota" (Fig. 2). Consistent with previous studies on the DPANN superphylum, all 23 "*Ca.* Aenigmarchaeota" MAGs have limited metabolic capacities. Pathways including the tricarboxylic acid cycle (TCA), fatty acid metabolism, and dissimilatory/assimilatory sulfur and nitrogen metabolism were missing (8, 17, 18). "*Ca.* Aenigmarchaeota" MAGs from hot springs and hydrothermal vents possess an incomplete glycolytic pathway (Fig. 2). All MAGs except DRTY-6_2 bin 201 lack the key enzyme phosphofructokinase (PFK), which impedes the formation of fructose-1,6-bisphosphate (fructose-1,6P) from fructose-6-phosphate (fructose-6P) (Data Set S1). DRTY-6_2 bin 201 seems to have a rather complete glycolysis pathway. However, the lack of glucokinase indicates the incapacity in the production of glucose-6-phosphate (glucose-6P) from glucose. Interestingly, the solely detected glycogen phosphorylase in this MAG and widely distributed phosphoglucomutase suggest that DRTY-6_2 bin 201 can phosphorylate glycogen into glucose-1-phosphate (glucose-1P) and further enter the glycolysis pathway by converting glucose-1P into glucose-6P, which could

Downloaded from https://journals.asm.org/journal/msystems on 27 August 2021 by 71.38.0.28.

TABLE 1 General genomic features of the "*Ca.* Aenigmarchaeota" MAGs reconstructed from hot spring metagenome sequencing

Bins	DRTY-6_1 bin 65	DRTY-6_2 bin 201	DRTY-6_2 bin 202	DRTY-6_2 bin 215	DRTY-7_1 bin 34	GMQ_1 bin 18-1	JZ-2_2 bin 245	QQ_2 bin 128
Genome size (Mbp)	0.75	0.71	0.69	0.55	0.82	0.55	0.52	0.64
GC content (%)	31.6	30.9	28.6	25.7	29.1	25.5	47.9	34.2
N_{50}	394,136	119,657	15,097	31,689	45,374	62,116	24,348	157,870
No. of scaffolds	19	9	69	25	27	19	25	8
Completeness (%)[a]	96.3	92.6	98.1	94.4	90.7	94.4	96.3	94.4
Contamination[b]	0	0	0	0	0	0	0	0.93
Strain heterogeneity[b]	0	0	0	0	0	0	0	0
No. of RNAs	42	20	35	22	43	47	28	34
5S rRNAs	2	1	1	0	1	1	1	0
16S rRNAs	0	1	1	1	1	1	1	1
23S rRNAs	1	1	0	0	1	1	1	1
tRNAs	39	17	33	21	40	44	25	32
No. of protein-coding genes	865	729	840	662	923	662	637	757
Avg length (bp)	817	857	754	789	820	790	747	747
Coding density (%)	93.9	88.0	92.1	94.5	92.8	94.4	92.3	88.0
Overlapped genes	223 (25.8%)	50 (6.9%)	193 (23%)	113 (17.1%)	229 (24.8%)	202 (30.5%)	105 (16.5%)	154 (20.3%)
No. of genes annotated by COG[c]	392 (43.2%)	422 (56.3%)	382 (43.7%)	367 (53.6%)	451 (46.6%)	336 (47.3%)	350 (52.6%)	383 (48.4%)
No. of genes annotated by KO[c]	313 (34.5%)	348 (46.4%)	324 (37%)	321 (46.9%)	353 (36.5%)	279 (39.3%)	287 (43.2%)	321 (40.5%)

[a]Genome completeness was calculated as the percentage of detected marker genes among 54 conserved single-copy genes as listed in Data Set S1 in the supplementary material.
[b]Genome quality, including contamination and heterogeneity, were estimated by CheckM (16).
[c]Functional annotation was conducted by uploading genomes to the IMG database.

Downloaded from https://journals.asm.org/journal/msystems on 27 August 2021 by 71.38.0.28.

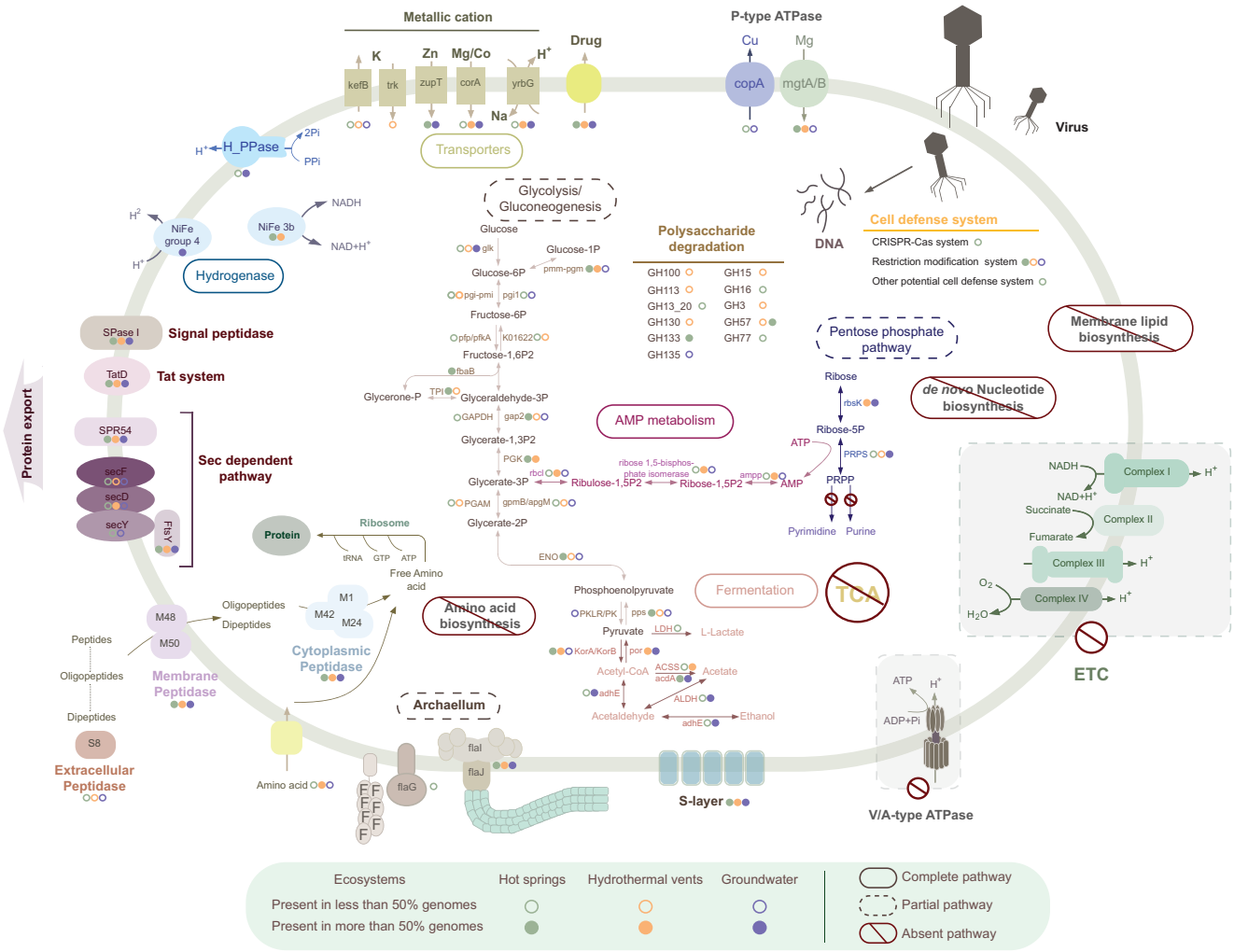


FIG 2 Reconstructed metabolic pathway of "*Ca.* Aenigmarchaeota." Key genes involved in glycolysis, gluconeogenesis, pentose phosphate pathway, pyruvate metabolism, fermentation, AMP metabolism, protein biosynthesis and exportation, transporters, and flagellum are shown. Solid circles indicate that certain genes are present in more than or equal to 50% of the genomes, while hollow circles are genes presented in less than 50% of the genomes. Abbreviations: TCA, tricarboxylic acid; ETC, electron transport chain; PRPP, phosphoribosyl pyrophosphate.

subsequently enter the rest of glycolysis pathway. The absence of pyruvate kinase (PK) and pyruvate kinase isozymes R/L (PKLR) prohibits the conversion of phosphoenolpyruvate (PEP) to pyruvate during the last step of glycolysis in DRTY-6_2 bin 201. However, phosphoenolpyruvate synthase (*pps*), which might perform the same function as PK in thermophiles (19, 20), was detected in most MAGs in this study, including DRTY-6_2 bin 201. As a result, these genes may provide an alternative glycolysis pathway to DRTY-6_2 bin 201. Subsequently, genes encoding 2-oxoglutarate/2-oxoacid ferredoxin oxidoreductase (*korAB*) in most of the MAGs suggest that "*Ca.* Aenigmarchaeota" is able to catalyze the reaction from pyruvate to acetyl-CoA. The reverse reaction could be performed by pyruvate ferredoxin oxidoreductase (*por*), which is widely detected in MAGs of hydrothermal vents and groundwater. Three MAGs can generate membrane proton motive force (PMF) via a hydrolysis process encoded by the membrane-bound H^+-phosphatase (H^+-PPase) (21). Alternatively, PMF could be generated via reactions involved in the degradation of amino acids or Na^+/H^+ antiporters (22). However, the absence of the electron transport chain (ETC), especially V/A-type ATPase, may suggest that "*Ca.* Aenigmarchaeota" could not produce ATP via PMF. Each "*Ca.* Aenigmarchaeota" contains at least one type of fermentation pathway. All MAGs from hot springs and groundwater except DRTY-6_2 bin 201 and GMQ_1 bin 18-1 harbor acetate-coenzyme A (acetate-CoA) ligase, which performs the conversion of acetate and ATP from acetyl-CoA, ADP, and phosphate. DRTY-6_2 bin

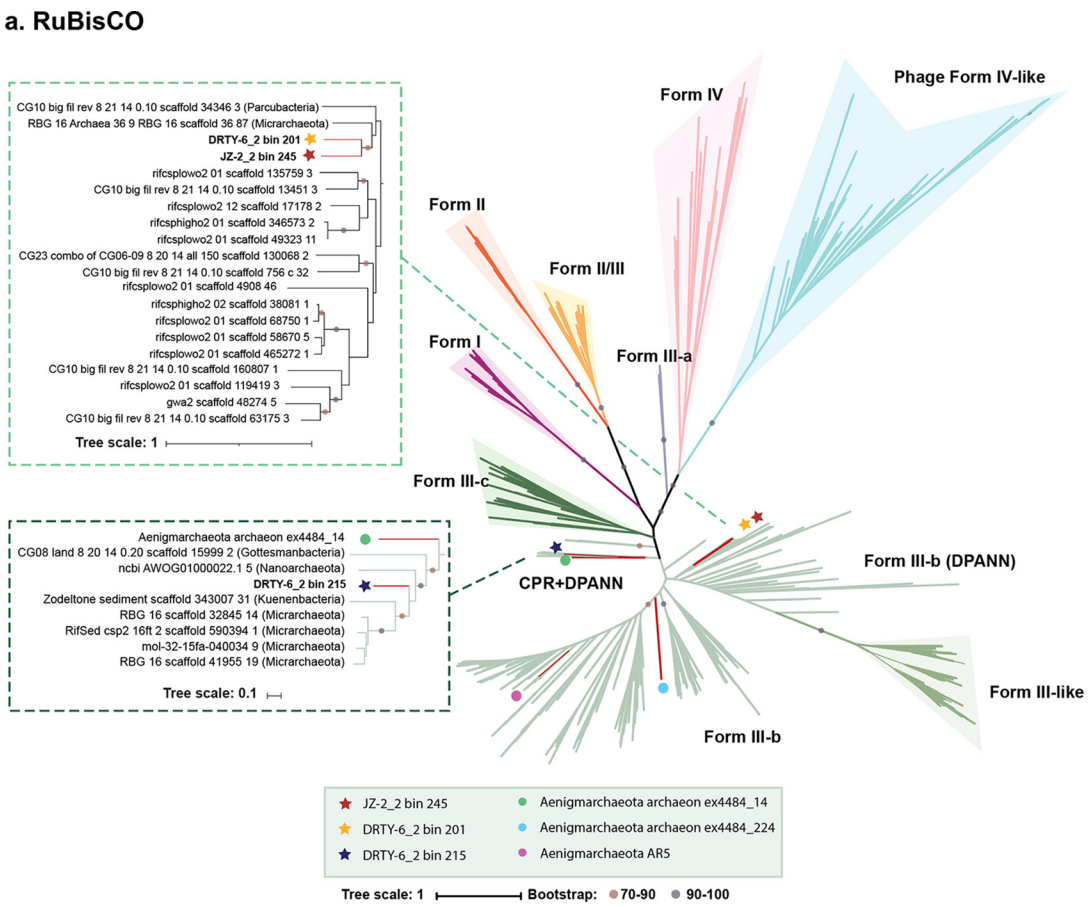
Downloaded from https://journals.asm.org/journal/msystems on 27 August 2021 by 71.38.0.28.

201, QQ_2 bin 128, and most MAGs from groundwater could utilize *adhE* to produce acetaldehyde and utilize aldehyde dehydrogenase to produce acetate. In addition, QQ_2 bin 128 is predicted to produce lactate and ethanol, based on the presence of L-lactate dehydrogenase, and acetaldehyde dehydrogenase/alcohol dehydrogenase. Moreover, MAGs from hydrothermal vents and DRTY-6_2 bin 201 could utilize acetyl-CoA synthetase (ACSS) to produce acetate as previously described for other DPANN genomes (3). The production of acetate is predicted to support the growth of aerobic/anaerobic respiratory organisms, indicating the role of "*Ca.* Aenigmarchaeota" in the energy cycle of the microbial community (8). These results show that fermentation pathways could be the main sources of ATP generation for "*Ca.* Aenigmarchaeota" (23).

Two genes relevant to polysaccharide degradation, including α-amylase (for starch and glycogen) and α-1,6-glucosidase (for starch and disaccharides), were identified in many of the hot spring MAGs (Data Set S1). DRTY-6_1 bin 65 might also degrade and utilize pullulan (GH13_20) and xyloglucan (GH16) (24, 25). Aside from α-amylase, MAGs from hydrothermal vents harbored different and more glycoside hydrolases, including β-glucosidases (for disaccharides), glucoamylases (for starch), β-1,2-mannosidases (for β-1,2-mannotriose and β-1,2-mannobiose), and endo-1,4-β-mannanase (for β-1,4-mannans, β-1,4-galactomannans, and β-1,4-glucomannans). However, MAGs from groundwater only had α-1,4-galactosaminogalactan hydrolase (for galactosaminogalactan). This might reflect a more active carbohydrate metabolism in thermal environments. None of the known carbon fixation pathways were detected in these MAGs, though three MAGs contain archaeal ribulose-bisphosphate carboxylase (RuBisCO). As previously described in other archaea, RuBisCO genes may function in the CO_2-incorporating AMP pathway, together with genes encoding AMP phosphorylase and ribose-1,5-biphosphate isomerase (26, 27). This pathway could produce glycerate-3-phosphate that enters the glycolysis pathway. Phylogenetic analysis suggests that the six RuBisCO genes recovered from "*Ca.* Aenigmarchaeota" belong to the form III group, of which five are from thermal environments (Fig. 3a). Four of them belong to form III-b, and the remaining two could be classified as a novel lineage that clustered with RuBisCO genes from candidate phyla radiation (CPR) genomes, which were previously suggested to have been passed by HGT from "*Ca.* Aenigmarchaeota" to CPR (26). Additionally, five MAGs distributed in both thermal and nonthermal ecosystems were identified to encompass all three genes involved in the AMP pathway (Data Set S1). We also identified different types of hydrogenases in "*Ca.* Aenigmarchaeota." Eight MAGs from thermal environments harbor NiFe 3b-type hydrogenases, which are clustered into one clade (Fig. 3b). This type of hydrogenase is functionally reversible and is capable of catalyzing the oxidation for anabolic metabolism or evolution of H_2 during fermentation (28). Unlike the MAGs from thermal environments, NiFe 3b-type of hydrogenases are absent in nonthermal environments. Instead, membrane-bound hydrogenases (group 4) are identified, illustrating that different strategies are used by nonthermophiles in producing PMF and H_2.

Despite the possession of genes involved in glycolysis and the fermentation pathway, the absence of many pivotal pathways strongly suggests a symbiotic lifestyle for "*Ca.* Aenigmarchaeota." First, this archaeal phylum is devoid of *de novo* amino acid biosynthetic pathways. Although we found a variety of extracellular peptidases, membrane peptidases, and cytoplasmic peptidases that can degrade extracellular and intracellular proteins and peptides (Data Set S1) (3), only a few amino acid transporters were detected. The only identified one is an uncharacterized amino acid transporter (arCOG00009), suggesting a poor ability in the transport of peptides/amino acids extracellularly. Therefore, "*Ca.* Aenigmarchaeota" presumably obtain amino acids from hosts by physical contact, similar to *Nanohaloarchaeota* and *Nanoarchaeota* (29–31), and rely on a great number of peptidases to recycle their amino acids. Second, *de novo* nucleotide biosynthetic pathways are absent in most of the genomes of this phylum (32). Moreover, genes for purine and pyrimidine salvage pathways are rarely detected in most MAGs from hot springs (Data Set S1), illustrating the further reliance on a host

Downloaded from https://journals.asm.org/journal/msystems on 27 August 2021 by 71.38.0.28.



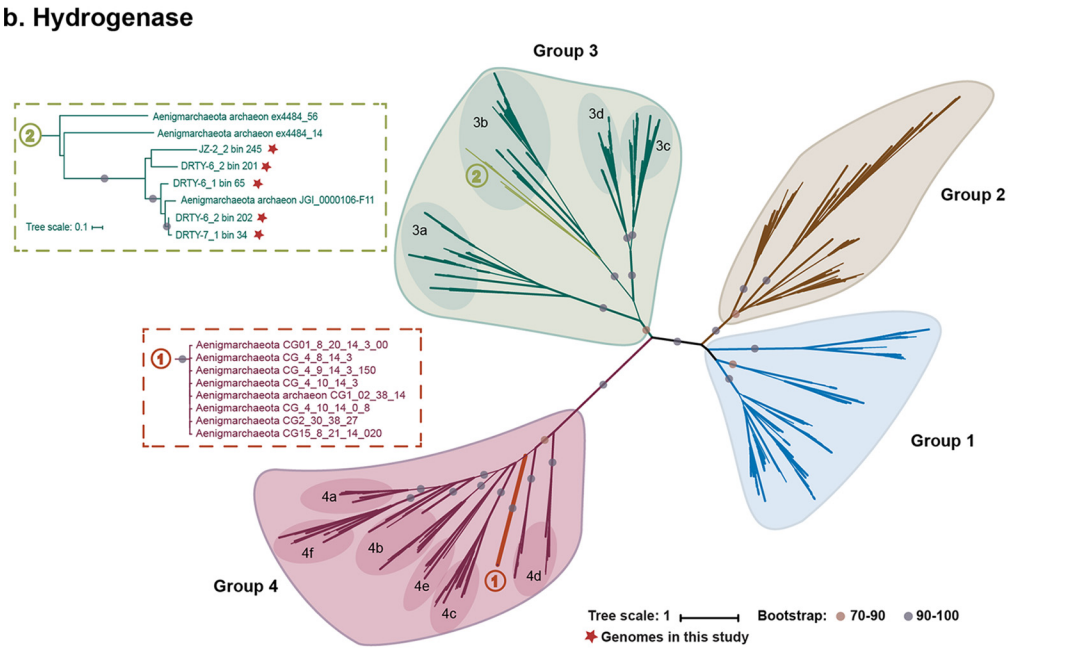


FIG 3 Phylogenetic tree of RuBisCO and hydrogenase. (a) Unrooted maximum likelihood tree of RuBisCO large subunit. RuBisCO genes from the present study are shown in stars. The bootstrap values of main clades in this phylogenetic tree are shown with dots in different colors. (b) Unrooted maximum likelihood tree of hydrogenases. Red stars represent genomes in this study. Dots on the branches indicate the bootstrap values of ≥70. Bootstrap confidences are shown only for the ancestral nodes of each clade.

Downloaded from https://journals.asm.org/journal/msystems on 27 August 2021 by 71.38.0.28.

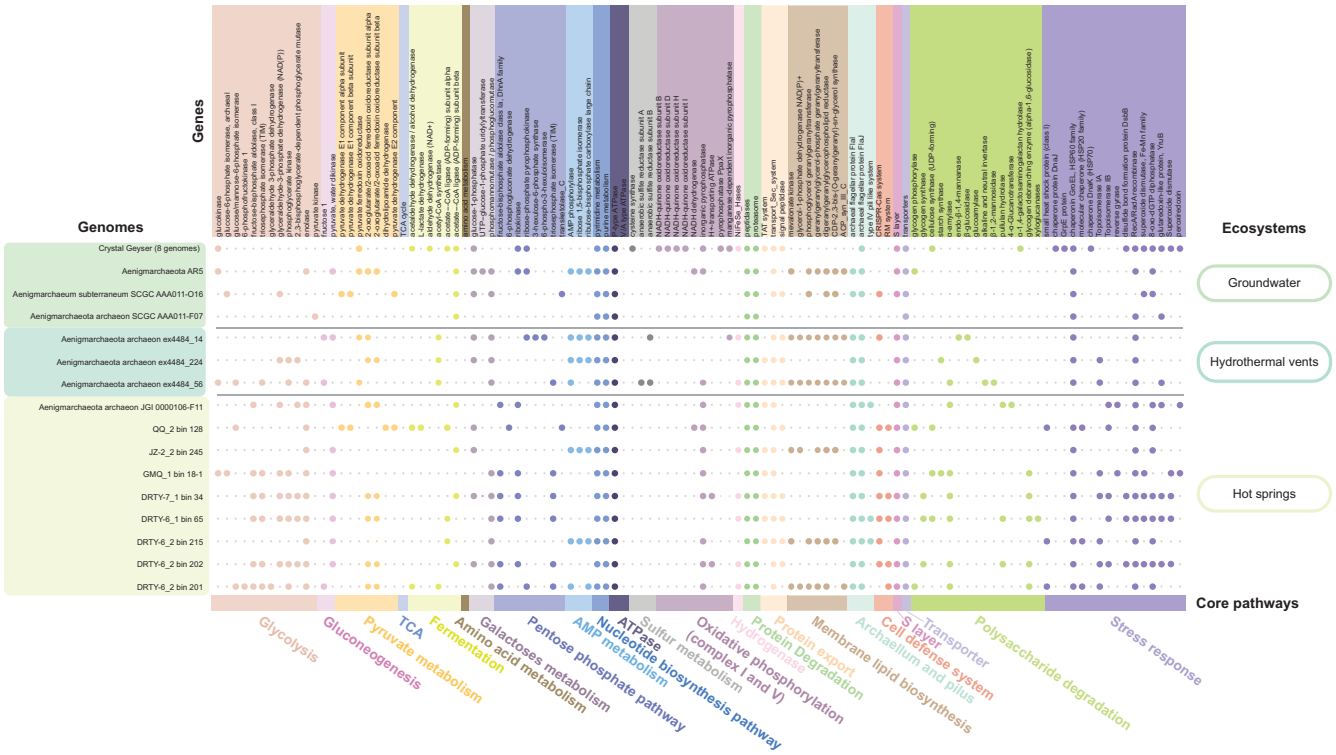


FIG 4 The core metabolic pathways with the presence/absence of genes of all "*Ca.* Aenigmarchaeota" genomes. The metabolic potential (columns) is mainly generated based on KEGG orthologs (KOs), clusters of orthologous groups (COGs), and archaeal clusters of orthologous groups (arCOGs). Genomes (rows) of "*Ca.* Aenigmarchaeota" were clustered by ecosystems. Eight genomes obtained from groundwater in Crestal Geyser were shown as one row because of the high metabolic similarity of the eight genomes. Small gray dots represent genes or metabolic pathways that were absent, and big dots represent genes or metabolic pathways that were present. Distinct metabolic pathways were distinguished by different colored dots. See Data Set S1 in the supplemental material for details.

to provide requisite nutrients. Third, "*Ca.* Aenigmarchaeota" genomes are unable to synthesize cell membranes *de novo* due to the lack of genes for synthesis of sterol isoprenoids involved in the mevalonate pathway (MVA), although genes encoding mevalonate kinase, glycerol-1-phosphate dehydrogenase, and associated enzymes for phospholipid biosynthesis have been detected (33, 34).

Cell-surface structures might enable the interactions between DPANN archaea with their hosts (30). Genes encoding S-layers, a subset of confirmed archaellum proteins (FlaG, FlaI, and FlaJ), and several adjacent archaellum homologs are identified in most "*Ca.* Aenigmarchaeota" MAGs. Notably, type-IV pili in "*Ca.* Aenigmarchaeota" are solely found from MAGs inhabiting thermal environments (35–37). To a certain extent, these genes endow "*Ca.* Aenigmarchaeota" with protection, motility, and cell-to-cell attachment abilities, which might consequently facilitate host-symbiont interactions.

Stress responses used by "*Ca.* Aenigmarchaeota." Comparative genomics showed that "*Ca.* Aenigmarchaeota" inhabiting thermal environments harbor higher abundances of genes involved in genetic information processing, including "transcription," "translation," "replication and repair," and "folding, sorting, and degradation" (Fig. S3a). In addition, genes involved in "cell motility" and "posttranslational modification, protein turnover, and chaperones" predominantly were enriched in thermophiles (Fig. S3b). This reflects the fact that cells at high temperatures have to combat constant thermal denaturation of both macromolecules and monomers (38–40). "*Ca.* Aenigmarchaeota" MAGs from thermal environments have evolved multiple strategies to overcome this stress. Chaperonin GroEL, associated with the repair of DNA and protein damage caused by high temperature, was present in all "*Ca.* Aenigmarchaeota" genomes (Fig. 4) (8, 41). The prevalence of DNA repair protein RadA could be used for homologous recombination and as an alternative strategy for DNA repair (42), indicating the pervasiveness of genome reduction among these genomes. Type I

Downloaded from https://journals.asm.org/journal/msystems on 27 August 2021 by 71.38.0.28.

(IA- and IB-type) topoisomerases and reverse gyrases, the latter considered a hallmark of hyperthermophily, were solely detected in MAGs from thermal environments, which could stabilize DNA and modulate DNA topology to maintain the structure and integrity of chromosomes (40, 43, 44). HSP70 (DnaK), DnaJ, and GrpE were absent in MAGs from thermal environments but were commonly detected in groundwater MAGs (Fig. 4), which is consistent with the potential role of this system in the adaptation to mesophily (45).

The presence of cell defense systems protects prokaryotic cells from virus infection. Based on the results of functional annotation, we investigated cell defense systems in "*Ca.* Aenigmarchaeota." DRTY-7_1 bin 34 was detected to contain a CRISPR-Cas system (Class III-A; Fig. S4a). All three types of restriction modification (RM) systems were found in "*Ca.* Aenigmarchaeota" (46) (Fig. S4b). The most frequently detected type III RM system was found in 15 (65%) of the total MAGs. Five of the eight MAGs from this study harbor a type II RM system, indicating an alternative common strategy for terrestrial thermal microbes to resist virus infection. Additionally, the recently discovered Hachiman system was detected in DRTY-6 bin 65 and the Gabija system was detected in DRTY-6 bin 215, providing broad protection against viruses (47) (Fig. S4c). The wide distribution of defense systems in thermophilic "*Ca.* Aenigmarchaeota" suggests that viruses could be an important threat to the survival of microbes in hot spring ecosystems (48–50). Viruses with different morphologies have been detected in hot springs and are highly active *in situ* (14, 47). Hence, it seems plausible that "*Ca.* Aenigmarchaeota" may confer their hosts with immunity to viruses by serving as "viral decoys" (51). The attracted virus could be degraded, and the released DNA could be recycled as a nucleotide source (51). Therefore, the host-symbiont interaction between "*Ca.* Aenigmarchaeota" and its potential hosts appear to be mutually beneficial.

Horizontal gene transfer in "*Ca.* Aenigmarchaeota." Horizontal gene transfer has been recognized as a substantial force in shaping the genetic diversity of prokaryotes (52–55). To ensure a high-quality detection of HGTs, we removed three low-quality genomes for HGT analysis (see details in Materials and Methods). Surprisingly, results uncovered a lower proportion of HGT-derived genes in "*Ca.* Aenigmarchaeota" than in thermophilic, free-living archaea, for example, *Aigarchaeota* (mean 14.2% versus 22.9%; Mann-Whitney *U* test, $P = 8.687E-06$) (7). Intriguingly, "*Ca.* Aenigmarchaeota" MAGs from different ecosystems possessed comparable percentages of HGT-derived genes (Fig. 5a), in which significant positive correlation (Pearson's $R^2 = 0.57$, $P = 7.213E-05$) was observed between detected HGTs and predicted gene totals in corresponding genomes regardless of ecosystems and genome sizes (Fig. 5b). By looking into the potential donors, we found that members from *Euryarchaeota* (998, 38.7%), "*Ca.* Bathyarchaeota" (193, 7.49%), and *Firmicutes* (147, 5.70%) are the top three contributors to the genetic innovations of "*Ca.* Aenigmarchaeota" MAGs (Fig. 5c and Data Set S3). Among the HGTs from *Euryarchaeota*, 186 (7.2% among all HGTs) are derived from *Methanomicrobia* and 178 (6.9%) are from *Methanococcus*. The high percentage of HGTs derived from *Euryarchaeota*, *Crenarchaeota*, and "*Ca.* Bathyarchaeota" is consistent with previous studies reporting of symbioses between DPANN and TACK and *Euryarchaeota* (10, 29, 30, 56).

To reveal potential factors that facilitate the adaptation of "*Ca.* Aenigmarchaeota" across groundwater, hot springs, and hydrothermal vents, we compared the acquired genes between these groups. Results indicated very little overlap between HGT-derived genes among these three ecosystems. Only 54 KEGG orthologs (KOs) and 67 archaeal clusters of orthologous groups (arCOGs) (11.04% and 12.27% among all KO and arCOG assignable HGTs) are shared by all the three groups (Data Set S3), suggesting that they have distinct adaptation strategies.

Through the BLAST-based analysis, as well as phylogenetic analysis, we can verify that HGT plays a crucial role during the adaptation to different stresses. For instance, the detected reverse gyrases in "*Ca.* Aenigmarchaeota" genomes seem to derive from "*Ca*. Bathyarchaeota," which may have improved the fitness of these organisms to inhabit high-temperature environments (Fig. S5a). Also, several genes encoding

Downloaded from https://journals.asm.org/journal/msystems on 27 August 2021 by 71.38.0.28.

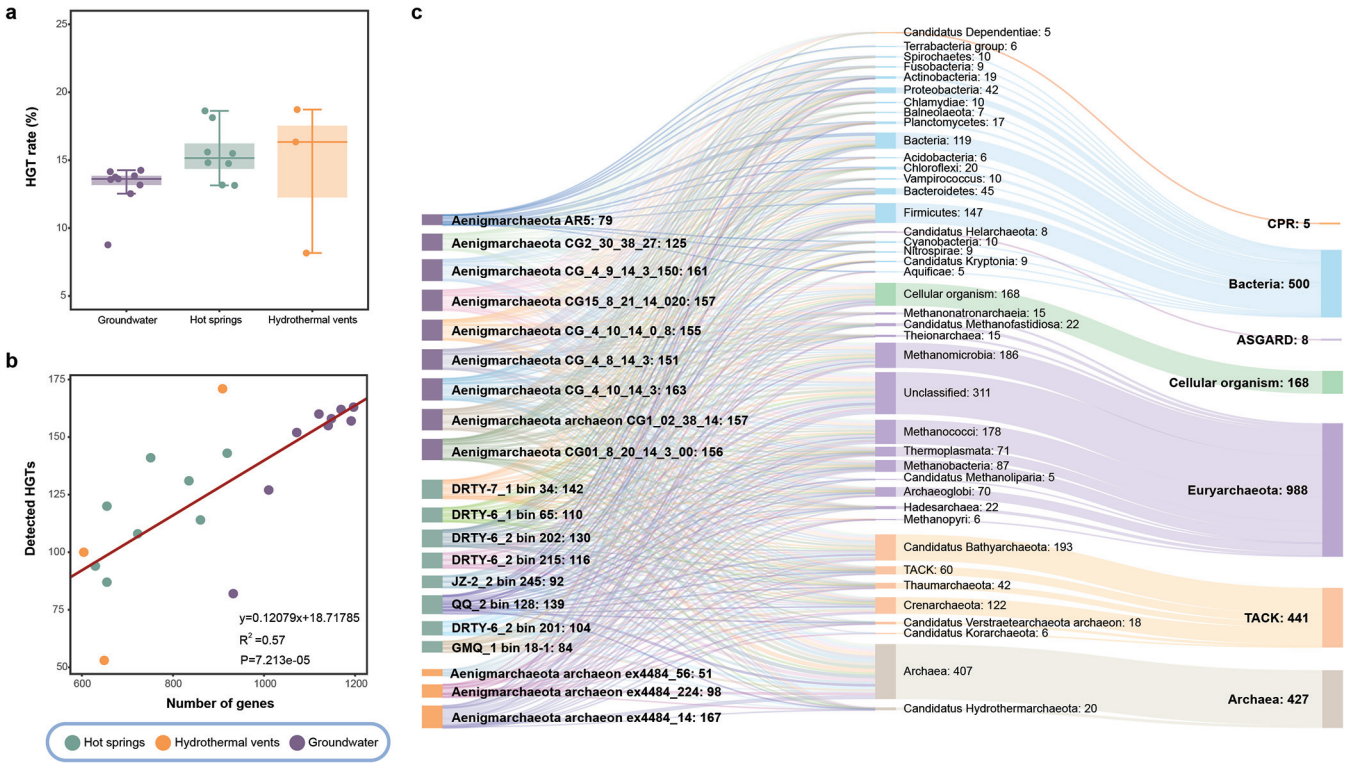


FIG 5 The identified horizontally gene transfers (HGTs) in genomes from this study, hydrothermal vents, and groundwater. (a) Comparisons of HGT rates between these three groups. (b) A linear regression analysis was conducted to model the relationship between genomes size and detected interphylum HGTs. The fitted formula was shown on the plot with R^2 = 0.57 and $P = 7.213E-05$. (c) The Sankey plot shows the detected interphylum HGTs and potential donors for each genome. Number of donors that were below five were removed. Interphylum HGTs were detected using HGTector as described in Materials and Methods.

superoxide dismutase (SOD2) (Fig. S5b) and 8-oxo-dGTP diphosphatase (*mutT*) (Fig. S5c) were identified as HGTs, which could be used to resist oxidative damage and to generate PMF (22). Two genes, including transcription initiation factor IIB (TFIIB) and phage integrase, in these two MAGs DRTY-6_1 bin 65 and DRTY-7_1 bin 34 were identified as belonging to phyla outside DPANN, both of which show high identity to *Euryarchaeota*. However, neither of them was identified as HGTs by HGTector. From the constructed phylogeny, the TFIIB gene in *Theionarchaea* archaeon DG-70 is surrounded by genes from "*Ca.* Aenigmarchaeota," suggesting that members of "*Ca.* Aenigmarchaeota" are possible gene donors rather than recipients (Fig. S5d). In DRTY-6 bin 65, both genes are in the same scaffold. The taxonomic information of genes close to TFIIB is mostly affiliated with "*Ca.* Aenigmarchaeota," consolidating the inference of vertical inheritance of this gene. The phage integrase may be horizontally transferred from *Euryarchaeota*, since 6 of the 10 downstream genes are close relatives to *Theionarchaea* (51.3 to 77.4%) and three are close relatives to *Thermoplasmata* (48.3 to 61.1%). Among them, three genes exhibit homologies to methyltransferase, DNA-binding protein, and restriction endonucleases associated with type II restriction modification (RM) systems. This suggests that integrase-mediated HGT may confer "*Ca.* Aenigmarchaeota" the special niche to resist virus infection. Overall, the above observations further support that HGT plays a substantial role in shaping the genetic diversity of "*Ca.* Aenigmarchaeota" for stress response.

Putative functional and genetic interaction partners of "*Ca.* Aenigmarchaeota" inferred from *in situ* communities. Cell-to-cell contact possibly leads to an opportunity for extensive HGTs (57–59). The inferred HGTs of "*Ca.* Aenigmarchaeota" may facilitate us to infer their potential hosts. However, only xenologous sequences with high identities, the so-called "recent HGTs," could be used for the inference of current symbiotic relationships between the associated donor and recipient (59). Therefore, we

Downloaded from https://journals.asm.org/journal/msystems on 27 August 2021 by 71.38.0.28.

ruled out the possibilities of Bacteria as potential hosts since most of the transferred genes represent ancient events with high divergence, though some Bacteria may contribute a lot to the genomic innovation of "*Ca.* Aenigmarchaeota." For instance, those genes transferred from *Firmicutes* only shared ~42% of the identities to the recipient genes in "*Ca.* Aenigmarchaeota." Additionally, most DPANN, such as *Nanoarchaeota*, *Nanohaloarchaeota*, and "*Ca.* Micrarchaeota," are incapable of synthesizing membrane spontaneously and must rely on assistance from their hosts (8, 30, 60, 61). Therefore, it is more likely that the putative hosts are restricted to Archaea due to the similar membrane structure shared by DPANN and most other archaea (8). Additionally, all previous studies support our conjecture that DPANN archaea were exclusively associated with archaea to form the symbiotic relationship (e.g., *Nanoarchaeum equitans* and *Ignicoccus hospitalis*, "*Ca.* Micrarchaeota acidiphilum" Mia14 and *Cuniculiplasma divulgatum* PM4, "*Ca.* Nanohaloarchaeota antarcticus" and *Halorubrum lacusprofundi*, "*Ca.* Huberiarchaeum crystalense," and "*Ca.* Altiarchaeum hamiconexum") (17, 29–31, 62, 63).

To further explore the potential host-symbiont relationships between "*Ca.* Aenigmarchaeota" and other archaea, a network interface of microbial communities in hot springs was constructed (see Materials and Methods), aiming to reveal microbial co-occurrence patterns and possible ecological interactions (59, 62–65). The network encompasses 97 nodes with 257 edges, and only one of them is identified as "*Ca.* Aenigmarchaeota" (Fig. 6a). After extracting the module that contains an "*Ca.* Aenigmarchaeota" operational taxonomic unit (OTU) (GMQ-1 bin_18-1), we observed tight connections ($|\text{rho}| \geq 0.6$ and pseudo-P values of <0.05) between GMQ-1 bin_18-1 and 11 OTUs from *Crenarchaeota* (rho values are listed in Data Set S4; Fig. 6b).

Notably, we identified a recent HGT event between "*Ca.* Aenigmarchaeota" and one of the *Crenarchaeota* OTUs, which strengthened the putative symbiont-host relationship between them. The SOD2 gene in GMQ_1 bin 18-1 shows 89.1% identity and 100% of coverage to the gene in *Pyrobaculum* sp. WP30. By looking into the belonging sample, we identified a SOD2 gene from the GMQ_1 bin 7 that shares a high identity (89%) to the one in GMQ_1 bin 18-1. Remarkably, the GMQ_1 bin 7 MAG could be assigned to the genus *Pyrobaculum* as well. BLAST searches suggest that all SOD2 genes except the one in GMQ_1 bin 18-1 show identities of $<30\%$, indicating the recent HGT event specifically occurs in GMQ_1 bin 18-1 rather than in all "*Ca.* Aenigmarchaeota" members. In addition, the OTUs that GMQ_1 bin 7 and GMQ_1 bin 18-1 belonged to have a statistically significant positive correlation (Pearson's $R^2 = 0.85$, $P < 1.68\text{E}-14$; Fig. 6c). These two bins are the first and fourth most abundant organisms in one sample by taking up greater than 50% of the cells in total. Based on the metabolic features of "*Ca.* Aenigmarchaeota" and *Pyrobaculum*, we proposed the potential interaction scenario between them (Fig. 7) (66). Specifically, for the MAGs of this study, *Pyrobaculum* GMQ_1 bin 7 could provide amino acids, nucleotides, membrane lipids, ATP, and active sugars to support growth of GMQ_1 bin 18-1, and GMQ_1 bin 18-1 possessed RM systems to protect the host for cell defense, which were absent in GMQ_1 bin 7. Additionally, the recent HGT of the SOD2 gene from GMQ_1 bin 7 provides GMQ_1 bin 18-1 with the capacity of oxidative stress resistance. However, the *Pyrobaculum* OTU is completely absent in the five communities where other "*Ca.* Aenigmarchaeota" MAGs in this study came from (Data Set S4). These results indicate that "*Ca.* Aenigmarchaeota" microbes in this study are likely to be associated with different interaction partners.

Conclusions. The enigmatic "*Ca.* Aenigmarchaeota" is still underexplored due to the insufficient cultured representatives or assembled genomes available. Here, we expanded the phylogenetic diversity of "*Ca.* Aenigmarchaeota" in hot spring environments and showed that "*Ca.* Aenigmarchaeota" can be found in diverse ecosystems on earth. They harbor limited metabolic capacities by missing several pivotal biosynthetic pathways, such as nucleotide, amino acid, and cell membrane biosynthesis, suggesting that such molecules need to be obtained from the environment or from the host as symbionts. Comparative genomics analysis reveals that genomes from thermal environments possess smaller genome sizes but stronger capacities in metabolizing carbohydrates. HGT

Downloaded from https://journals.asm.org/journal/msystems on 27 August 2021 by 71.38.0.28.

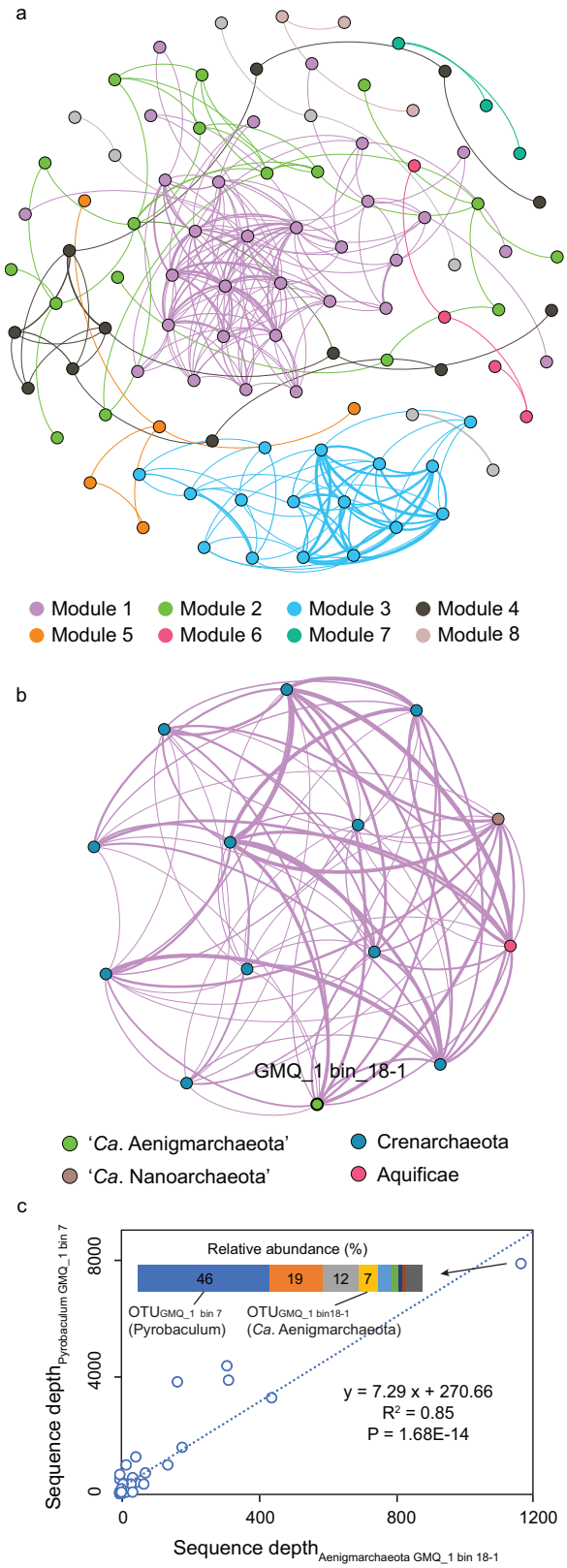


FIG 6 The co-occurrence network between "*Ca.* Aenigmarchaeota" and other community members. (a) The co-occurrence network was constructed based on the depth of the representative sequences of all OTUs in all samples using SparCC (95) with the default parameters, and 100 bootstrap samples were used to infer pseudo-*P* values. The nodes represent OTUs at a 95% cutoff, and these edges denote significant ($P < 0.05$, two-sided) and robust correlations (rho > 0.6) between pairwise OTUs.

(Continued on next page)

Downloaded from https://journals.asm.org/journal/msystems on 27 August 2021 by 71.38.0.28.

identifies a salient number of gene flows from TACK and *Euryarchaeota* to "*Ca.* Aenigmarchaeota," especially the genes related to stress responses. By conducting co-occurrence network and recent HGT detection analyses, we highlight the power of informatics analysis in identifying putative interaction partners. However, even though significant correlations and genetic interactions are observed, further analyses such as fluorescence *in situ* hybridization, should be integrated to confirm the inference of host-symbiont relationship. Overall, this study enables us to better understand the metabolic potentials and possible interactions between "*Ca.* Aenigmarchaeota" and their putative hosts, shedding light on the understanding of coevolution between hosts and symbionts.

MATERIALS AND METHODS

Site description, sampling, DNA extraction, and sequencing. All five hot spring sediment samples were collected from Tengchong County, Yunnan, China (24.95, 98.44) in January 2016 and May 2017. DRTY-6_1, DRTY-7_1, and GMQ_1 samples were collected from DiReTiYanQu (DRTY), GuMingQuan (GMQ), and QiaoQuan (QQ) hot springs in Rehai National Park in January 2016. DRTY-6_2 and QQ_2 samples were collected from DiReTiYanQu and QiaoQuan (QQ) hot springs in May 2017. JZ-2_2 was collected from the JinZe Hot Spring Resort in May 2017. DiReTiYanQu is an artificial concrete hot spring landscape experiencing area. DRTY-6 and DRTY-7 are two medium-size pools that have mixtures of spall, sand, and soil on the bottom. The pH and temperatures of DRTY-6 were 6.1/50°C in January 2016 and were 6.0/60°C in May 2017. The pH and temperature of DRTY-7 were 6.0 and 56°C, respectively, in January 2016. GuMingQuan (GMQ) is a pool with a hot spring fall above and with a length, width, and depth of around 98, 79, and 9.5 cm, respectively. Leaf litter and other debris on the bottom could be seen clearly. The sampling site is located upstream of the GMQ pool, which was named GMQS. The pH and temperature of GMQ were 9.0 and 89°C, respectively, in January 2016. QiaoQuan (QQ) is a hot spring stream flowing out from a soil slope with a rust-color trace. This spring is surrounded by bush and grass. The pH and temperature of QQ were 7.2 and 77°C, respectively, in May 2017. The JinZe-2 (JZ-2) pool is an artificial concrete cubic hot spring water reservoir with a ceiling covering the top in JinZe Hot Spring Resort. The JZ-2 pool contains turbid water and sediments on the bottom of the pool. The pH and temperature of JZ-2 were 7.6 and 63°C, respectively, in May 2017. The top 1 cm of sediment of each site was collected with a sterile iron spoon and transferred to a 50-ml centrifuge tube. All sediment samples were then stored in liquid nitrogen before transporting to the lab and were stored at −20°C in the lab until DNA extraction.

Community DNA was extracted from approximately 20 g of sediment for each sample using the PowerSoil DNA isolation kit (MoBio). Libraries with an insert size of 350 bp were constructed by using an M220 Focused-ultrasonicator NEBNext and an Ultra II DNA library prep kit. The concentration of genomic DNA was measured with a Qubit fluorometer. The total genomic DNA was sequenced using an Illumina Hiseq 4000 instrument at Beijing Novogene Bioinformatics Technology Co., Ltd. (Beijing, China). On average, 30 giga base pairs (Gbp) (2×150 bp) of raw sequencing data for each sample were generated.

Metagenomic assembly and genome binning. Raw sequencing data were preprocessed to eliminate replicated reads and trim bases with low qualities, following workflows that were described previously (67). All quality reads of each data set were *de novo* assembled using SPAdes v3.9.0 (68) with the following parameters: -k 33,55,77,99,111 –meta. Scaffolds with a length of <2,500 bp in each assembly were removed. BBMap v38.85 (http://sourceforge.net/projects/bbmap/) with the parameters k = 15 minid = 0.97 build = 1 was used to compute the coverage information by mapping clean reads to the corresponding assembled scaffolds without cross-mapping. Genome binning was performed based on the calculated sequence depth and tetranucleotide frequency (TNF) of each scaffold using MetaBAT v2.12.1 (69). Marker genes that occurred more than once in each bin were treated as contaminations, and associated contigs were removed manually. Specifically, genome bins were visualized by fragmenting each scaffold into pieces with the length ranging from 5 to 10 kb using ESOM v1.1 (70) for further curation, in which the discordant points were removed manually from the clusters. Scaffolds with similar TNFs but abnormal sequence depth (the abnormal sequences depth were examined manually, and the difference was mostly over 10-fold) compared to other scaffolds in the corresponding bins were also discarded. Subsequently, quality reads of the associated samples of each optimized genome bins were recruited by mapping onto all optimized genome bins using BBMap and were reassembled using SPAdes under the "–careful" mode with the parameter "-k 21,33,55,77,99,127." Contaminations and strain heteroge-

FIG 6 Legend (Continued)

OTUs are colored by modularity classes. Nodes have the same size, and edges have the same thickness. (b) The subnetworks that contain "*Ca.* Aenigmarchaeota" *rpS3* genes. Only nodes and edges that have connections with "*Ca.* Aenigmarchaeota" in the corresponding modules are shown. Modules 1 in (a) were detected to contain "*Ca.* Aenigmarchaeota" *rpS3* genes. The size of each node is proportional to the number of connections (i.e., degree). OTUs are colored by the phylum-level taxonomy. The thickness of edges denotes the Spearman rank correlation coefficients (rho values). Edges in purple show the connections between "*Ca.* Aenigmarchaeota" (green circles deviated from core networks) and other members. (c) Correlation between *Pyrobaculum* GMQ_1 bin 7- and GMQ_1 bin 18-1-associated OTUs. The two OTUs are observed in 33 out of 88 metagenomic samples. Results show a statistically significant positive correlation (Pearson's R^2 = 0.85, $P < 1.68E-14$) between the sequence depths of these two OTUs.

Downloaded from https://journals.asm.org/journal/msystems on 27 August 2021 by 71.38.0.28.

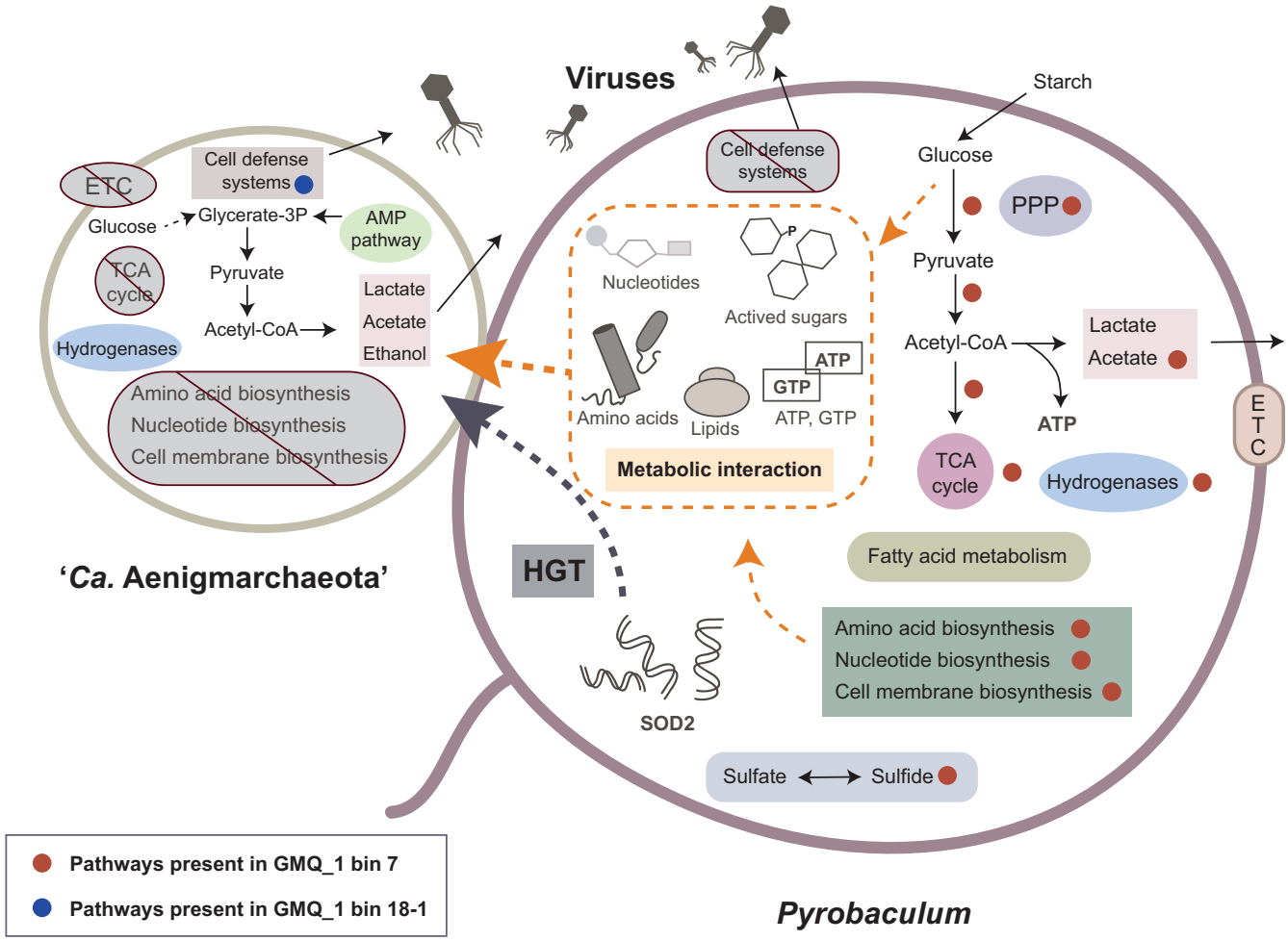


FIG 7 The potential interaction between "*Ca.* Aenigmarchaeota" and *Pyrobaculum*. Based on the annotation results of eight MAGs of "*Ca.* Aenigmarchaeota" in this study and the metabolic features of *Pyrobaculum* GMQ_1 bin 7 and other *Pyrobaculum* genomes, the schematic metabolic capacities and possible interaction scenario of "*Ca.* Aenigmarchaeota" and *Pyrobaculum* were shown. Metabolic pathways that presented in GMQ_1 bin 18-1 and GMQ_1 bin 7 were marked with blue and red circles, respectively. Abbreviations: TCA cycle, tricarboxylic acid cycle; ETC, electron transport chain; PPP, pentose phosphate pathway.

neity were estimated by CheckM v1.0.12 (71); genome completeness was estimated by calculating the proportion of detected marker genes among 54 conserved archaeal single-copy genes (SCGs) (Data Set S1) (3). Finally, eight MAGs identified as "*Ca.* Aenigmarchaeota" were kept for the later analysis.

Functional annotation of "*Ca.* Aenigmarchaeota" genomes. The eight MAGs were submitted to the Integrated Microbial Genomes & Microbiomes (IMG-M) (https://img.jgi.doe.gov/cgi-bin/m/main.cgi) database for gene prediction and functional annotation. For comparative genomics analysis, the annotation pipeline was also conducted locally. In brief, putative protein-coding sequences (CDS) of all MAGs, including eight MAGs from the present study and 15 genomes downloaded from public databases, were determined using Prodigal v2.6.3 (72) under the "-p single" model. Functional annotations were performed by comparing predicted CDSs against KEGG (73), evolutionary genealogy of genes: nonsupervised orthologous groups (eggNOG) (74), Pfam (75), and arCOG (76) databases using DIAMOND v0.7.9 (77) with a cutoff E value of $<1E-5$. rRNA-coding regions were identified using RNAmmer v1.2 (78). All MAGs were uploaded to the web server of tRNAscan-SE v2.0 (79) to identify the tRNA. The dbCAN2 webserver (80) was used to identify carbohydrate-active enzymes based on the carbohydrate-active enzymes (CAZy) database. The localization signals of genes annotated as peptidases were determined using the online tool PSORTb v3.0 (81). To detect the putative CRISPR-Cas systems in "*Ca.* Aenigmarchaeota" MAGs, tandem repeats, and spacers were identified using the online tool CRISPRFinder (82). Then, the genes nearby the region (both the upstream and downstream) were investigated manually to decide the type. The restriction modification (RM) systems and other cell defense systems, including Hachiman and Gabija reported in a previous study (44), were identified according to the KEGG and arCOG annotation results.

Phylogenetic and phylogenomic analysis. Sixteen ribosomal protein sequences (L2, L3, L4, L5, L6, L14, L15, L16, L18, L22, L24, S3, S8, S10, S17, and S19) were selected to reconstruct the phylogenomic tree (83). These sequences were identified by AMPHORA2 (84) and aligned using MUSCLE v3.8.31 with 100 iterations (85). The poorly aligned regions were eliminated using TrimAl v1.4.rev22 (86) with the

Downloaded from https://journals.asm.org/journal/msystems on 27 August 2021 by 71.38.0.28.

parameters set as -gt 0.95 -cons 50. Then, multiple alignments were concatenated and applied to reconstruct a maximal likelihood phylogenetic tree using IQ-TREE v1.6.11 with the following parameters: iqtree -s a -alrt 1000 -bb 1000 -nt AUTO (87, 88).

16S rRNA genes were predicted for each MAG using RNAmmer. MAGs without 16S rRNA genes were further searched against the Ribosomal Database Project (RDP) database (downloaded on 18 October 2018) (89) using the BLASTn v2.8.1+ program, and sequences with a length of >300 bp were selected and combined with those retrieved by RNAmmer. All 16S rRNA sequences were aligned using the online tool SINA v1.2.11 (90). Columns containing more than 95% gaps were removed, after which a maximum likelihood phylogenetic tree was constructed using IQ-TREE with the parameters iqtree -s a -alrt 1000 -bb 1000 -nt AUTO.

Reference sequences of RuBisCo large subunit were obtained from a previous study (26). All sequences were aligned using MAFFT v6.864b (91). Poorly aligned regions were removed using TrimAl v1.4.rev22 (86). The unrooted phylogeny was generated using RAxML v7.2.7 (86) with the following parameters: -f a -n boot -m PROTGAMMAIJTT -c 4 -e 0.001 -p 13452 -x 1165 -# 1000.

Reference sequences of hydrogenases were selected from a previous study (28). Alignments were generated using MUSCLE v3.8.31 (85), and divergent regions were filtered using TrimAl (86). The same model and detailed parameters of RAxML as RuBisCo large subunit phylogeny were used to construct the phylogeny of hydrogenase sequences.

Recruiting 16S rRNA from NCBI. Seventeen 16S rRNA gene sequences recovered from currently available "*Ca.* Aenigmarchaeota" genomes were used as input to search against the NCBI-nt (https://www.ncbi.nlm.nih.gov/) database via the BLASTn program with the default parameters. BLAST hits with coverage of >95% and identity of >85% were kept for the phylogenetic tree construction. Sequences were aligned, while poor alignment regions were eliminated. A maximum likelihood-based phylogeny was generated using RAxML v7.2.7 (86). Finally, we identified 236 16S rRNA gene sequences of "*Ca.* Aenigmarchaeota" by the result of phylogenetic analysis.

Detection of horizontally transferred genes. Twenty genomes with completeness of >75% and contamination of <5% evaluated by CheckM were taken into consideration for the inferences of horizontal gene transfers (HGTs). As a result, three SAGs were removed because of the low quality. Putative HGTs were identified for each genome using HGTector v2 (92). To determine the interaction partners or possible hosts of "*Ca.* Aenigmarchaeota," recent horizontal gene transfers were identified by applying BLAST searches against the genome bins reconstructed from the corresponding communities. Only genes from outside DPANN are considered the so-called interphylum HGTs. Due to a lack of representative archaeal genomes in the prebuilt default database, especially the genomes from DPANN and TACK superphyla, 3,358 genomes were downloaded from the RefSeq database on 14 May 2019. Genome quality was evaluated using CheckM except microbes from the DPANN superphylum. Genome quality of the DPANN superphylum was performed using the same procedure as mentioned above. Genomes with completeness of <80% and contamination of >10% were discarded. The remaining high-quality genomes were dereplicated at the phylum level using dRep v2.3.2 (93). Finally, 1,133 genomes were picked out, and 689 of those genomes complementary to the default database were appended. The combined sequences were compiled into a database using DIAMOND (77), and the relevant taxonomy files were changed correspondingly. Then, the "search" step was performed for the 20 high-quality "*Ca.* Aenigmarchaeota" genomes with default parameters. During the "analyze" step, genomes belonging to the DPANN superphylum (taxonomy ID, 1783276) were treated as "closeTax," but different species were used as "selfTax." Specifically, *Aenigmarchaeum subterraneum* SCGC AAA011-O16 (taxonomy ID, 743730) was set as the "selfTax" for all MAGs for DRTY-6_2 bin 215 and JZ-2_2 bin 245; *Aenigmarchaeota archaeon* ex4484_224 (taxonomy ID, 2012503) was used as "selfTax" for GMQ_1 bin 18-1 and QQ_2 bin 128; *Aenigmarchaeota archaeon* ex4484_14 (taxonomy ID, 2012502) was used as "selfTax" for DRTY-6_2 bin 201; and *Aenigmarchaeota archaeon* JGI 0000106-F11 (taxonomy ID, 1130284) was used for DRTY-6_1 bin 65, DRTY-7_1 bin 34, and DRTY-6_2 bin 202. The identified interphylum HGTs for each genome were visualized using SankeyMATIC (http://sankeymatic.com/).

In more detail, preliminary genome binning was conducted as described above for each "*Ca.* Aenigmarchaeota"-containing microbial community. The taxonomic information was determined using GTDB-tk v.0.2.2 (94). Then, gene calling was performed for all genome bins. The predicted putative coding sequences were formatted into a BLAST database. A BLAST search against this database was conducted with genes identified from "*Ca.* Aenigmarchaeota" as input. Only BLAST hits from outside DPANN with a sequence identity of ≥70% and aligned region of ≥100 amino acids were retained. The target genes and BLAST hits that represented the first or last genes of the belonging scaffolds (≥5,000 bp) were discarded. The remaining genes were used for inferring the potential functional interactions between putative hosts and symbionts.

Phylogenetic analysis of four horizontally transferred genes. (i) Reverse gyrase (*rgy*). Scaffolds identified as *rgy* in "*Ca.* Aenigmarchaeota" were collected and used as input to blast against the NCBI-nr database (E value of <1E−5). The top 50 hits for each query were kept. The obtained *rgy* genes for all queries were combined and clustered using CD-HIT v4.6 (95), with a sequence identity cutoff of 90%. The identified representative sequences were used to build the phylogenetic tree. Sequences were aligned using MUSCLE, and poorly aligned regions were eliminated using TrimAl. Phylogeny was generated using IQ-TREE v1.6.11 (87) by integrating 1,000 times with the best model "LG+F+R9." The generated phylogenetic tree was rooted using MAD v2.2 (96) with the default parameters.

Superoxide dismutase (SOD2). The same procedures as the *rgy* phylogeny construction were applied, including sequence recruitment, clustering, alignment, and poorly aligned region elimination, except the sequence identity cutoff was set to 0.8 during the clustering using CD-HIT. The phylogenetic tree was generated using IQ-TREE with the best model "WAG+R10."

Downloaded from https://journals.asm.org/journal/msystems on 27 August 2021 by 71.38.0.28.

(ii) 8-oxo-dGTP diphosphatase (*mutT*). The same procedures as the *rgy* phylogeny construction were applied, including sequence recruitment, clustering, alignment, and poorly aligned region elimination, except the sequence identity cutoff was set to 0.65 during the clustering using CD-HIT. The phylogenetic tree was generated using IQ-TREE with the best model "VT+F+R10."

(iii) Transcription initiation factor IIB (*TFIIB*). The same procedures as the *rgy* phylogeny construction were applied, including sequence recruitment, clustering, alignment, and poorly aligned region elimination. The phylogenetic tree was generated using IQ-TREE with the best model "LG+R10."

Network-based co-occurrence analysis. A total of 88 hot spring samples across the time and spatial scale were used to reveal the co-occurrence pattern of "*Ca.* Aenigmarchaeota" with other community members (Data Set S4). Metagenomic sequencing was conducted for all samples. Detailed quality control and assembly steps were done as described above. Genome binning was conducted for each community (as described above), and only genome bins with completeness of >50%, contamination of <5%, and *rpS3* gene called by AMPHORA2 (84) occurring exactly once in one genome were taken into consideration. The corresponding nucleotide sequences were extracted for the later analysis. Taxonomic information of each bin was obtained using GTDBtk v.0.2.2 (94). All predicted *rpS3* gene sequences from different data sets were combined and clustered into OTUs using USEARCH 9.2.84 with the following parameters: -cluster_smallmem -id 0.95. Taxonomy of the representative sequences was assigned according to the taxonomic information of belonging bins. The sequence depth of each *rpS3* gene sequence was used to build the OTU table and was calculated by mapping clean reads in each sample to the dereplicated *rpS3* gene sequences. Specifically, clean reads from each sample were mapped to *rpS3* gene sequences using BBmap with the same parameter settings as described above. The generated .bam files were sorted using SAMtools v.1.3.1 (97). Sequence depth was subsequently calculated using the "jgi_summarize_bam_contig_depths" program in MetaBAT. OTUs that occurred in less than six samples were filtered out to reduce the complexity, and 844 OTUs were kept for the subsequent network construction. The co-occurrence network was constructed using SparCC (98) with the default parameters, and 100 bootstrap samples were used to infer pseudo-*P* values. Those significant ($P < 0.05$, two-sided) and robust correlations (rho > 0.6) between pairwise OTUs were used to infer a reliable network. Network visualization and relevant parameter calculations regarding modularity, betweenness, closeness, average clustering coefficient, average weighted degree, and average shortest path length were conducted in Gephi v.0.9.2 (99).

Data availability. All genomes in our study are available at Joint Genome Institute (JGI) IMG-MER under the study ID Gs0127627 and whole-genome sequencing accession numbers Ga0181641 (unclassified Aenigmarchaeota DRTY7 bin_34), Ga0181640 (unclassified Aenigmarchaeota DRTY6 bin_65), Ga0181639 (unclassified Aenigmarchaeota GMQ bin_18-1), Ga0227293 (unclassified Aenigmarchaeota JZ-2 bin_245), Ga0227294 (unclassified Aenigmarchaeota DRTY-6 bin_215), Ga0261588 (unclassified Aenigmarchaeota DRTY-6 bin_201), Ga0261590 (unclassified Aenigmarchaeota DRTY-6 bin_202), and Ga0261591 (unclassified Aenigmarchaeota QQ bin_128). All genomes are also available in the NCBI database. The BioProject number is PRJNA544494. The genome accession numbers are JAHLMM000000000 (Aenigmarchaeota_DRTY-6_1_bin_65), JAHLMN000000000 (Aenigmarchaeota_DRTY-7_1_bin_34), JAHLMO000000000 (Aenigmarchaeota_GMQ_1_bin_18-1), JAHLMP000000000 (Aenigmarchaeota_DRTY-6_2_bin_201), JAHLMQ000000000 (Aenigmarchaeota_DRTY-6_2_bin_202), JAHLMR000000000 (Aenigmarchaeota_DRTY-6_2_bin_215), JAHLMS000000000 (Aenigmarchaeota_JZ-2_2_bin_245), and JAHLMT000000000 (Aenigmarchaeota_QQ_2_bin_128).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

DATA SET S1, XLSX file, 0.1 MB.
DATA SET S2, XLSX file, 0.02 MB.
DATA SET S3, XLSX file, 0.3 MB.
DATA SET S4, XLSX file, 0.4 MB.
FIG S1, PDF file, 0.8 MB.
FIG S2, PDF file, 0.3 MB.
FIG S3, PDF file, 0.9 MB.
FIG S4, PDF file, 0.8 MB.
FIG S5, PDF file, 1.3 MB.
TABLE S1, XLSX file, 0.01 MB.

ACKNOWLEDGMENTS

We thank Guangdong Magigene Biotechnology Co., Ltd., China, for the assistance in data analysis and the entire staff from Yunnan Tengchong Volcano and Spa Tourist Attraction Development Corporation for strong support.

We declare no competing interests.

This work was financially supported by the National Natural Science Foundation of China (no. 91951205 and 91751206), Natural Science Foundation of Guangdong

Downloaded from https://journals.asm.org/journal/msystems on 27 August 2021 by 71.38.0.28.

Province, China (no. 2016A030312003), and Guangzhou Municipal People's Livelihood Science and Technology Plan (no. 201803030030).

Y.X.L., Y.Z.R., Y.L.Q., Y.N.Q., Z.S.H., and W.J.L. conceived the study. Y.X.L., Y.Z.R., Y.L.Q., H.C.J., and Y.N.Q. performed the measurement of physiochemical parameters and DNA extraction. Y.X.L., Y.Z.R., Z.S.H., Y.L.Q., Y.N.Q., Y.T.C., J.Y.J., and B.P.H. performed the metagenomic analysis, genome binning, functional annotation, and evolutionary analysis. Y.X.L., Y.Z.R., Z.S.H., Y.L.Q., Y.N.Q., W.J.L., W.S.S., and B.P.H. wrote the manuscript. All authors discussed the results and commented on the manuscript.

REFERENCES

1. Whitman WB, Coleman DC, Wiebe WJ. 1998. Prokaryotes: the unseen majority. Proc Natl Acad Sci U S A 95:6578–6583. https://doi.org/10.1073/pnas.95.12.6578.
2. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng JF, Darling A, Malfatti S, Swan BK, Gies EA, Dodsworth JA, Hedlund BP, Tsiamis G, Sievert SM, Liu WT, Eisen JA, Hallam SJ, Kyrpides NC, Stepanauskas R, Rubin EM, Hugenholtz P, Woyke T. 2013. Insights into the phylogeny and coding potential of microbial dark matter. Nature 499:431–437. https://doi.org/10.1038/nature12352.
3. Castelle CJ, Wrighton KC, Thomas BC, Hug LA, Brown CT, Wilkins MJ, Frischkorn KR, Tringe SG, Singh A, Markillie LM, Taylor RC, Williams KH, Banfield JF. 2015. Genomic expansion of domain Archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. Curr Biol 25:690–701. https://doi.org/10.1016/j.cub.2015.01.014.
4. Adam PS, Borrel G, Brochier-Armanet C, Gribaldo S. 2017. The growing tree of Archaea: new perspectives on their diversity, evolution and ecology. ISME J 11:2407–2425. https://doi.org/10.1038/ismej.2017.122.
5. Spang A, Ettema T. 2016. Microbial diversity: the tree of life comes of age. Nat Microbiol 1:16056. https://doi.org/10.1038/nmicrobiol.2016.56.
6. Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, Bäckström D, Juzokaite L, Vancaester E, Seitz KW, Anantharaman K, Starnawski P, Kjeldsen KU, Stott MB, Nunoura T, Banfield JF, Schramm A, Baker BJ, Spang A, Ettema TJ. 2017. Asgard Archaea illuminate the origin of eukaryotic cellular complexity. Nature 541:353–358. https://doi.org/10.1038/nature21031.
7. Hua ZS, Qu YN, Zhu Q, Zhou EM, Qi YL, Yin YR, Rao YZ, Tian Y, Li YX, Liu L, Castelle CJ, Hedlund BP, Shu WS, Knight R, Li WJ. 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.
8. Castelle CJ, Brown CT, Anantharaman K, Probst AJ, Huang RH, Banfield JF. 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.
9. Kellner S, Spang A, Offre P, Szöllosi GJ, Petitjean C, Williams TA. 2018. Genome size evolution in the Archaea. Emerg Top Life Sci 2:595–605. https://doi.org/10.1042/ETLS20180021.
10. Wurch L, Giannone RJ, Belisle BS, Swift C, Utturkar S, Hettich RL, Reysenbach AL, Podar M. 2016. Genomics-informed isolation and characterization of a symbiotic *Nanoarchaeota* system from a terrestrial geothermal environment. Nat Commun 7:12115. https://doi.org/10.1038/ncomms12115.
11. Ortiz-Alvarez R, Casamayor EO. 2016. High occurrence of *Pacearchaeota* and *Woesearchaeota* (Archaea superphylum DPANN) in the surface waters of oligotrophic high-altitude lakes. Environ Microbiol Rep 8:210–217. https://doi.org/10.1111/1758-2229.12370.
12. Youssef NH, Rinke C, Stepanauskas R, Farag I, Woyke T, Elshahed MS. 2015. Insights into the metabolism, lifestyle and putative evolutionary history of the novel archaeal phylum '*Diapherotrites*'. ISME J 9:447–460. https://doi.org/10.1038/ismej.2014.141.
13. Takai K, Moser DP, DeFlaun M, Onstott TC, Fredrickson JK. 2001. Archaeal diversity in waters from deep south African gold mines. Appl Environ Microbiol 67:5750–5760. https://doi.org/10.1128/AEM.67.21.5750-5760.2001.
14. Hedlund BP, Cole JK, Williams AJ, Hou W, Zhou E, Li W, 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.
15. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA, Tringe SG, Ivanova NN, Copeland A, Clum A, Becraft ED, Malmstrom RR, Birren B, Podar M, Bork P, Weinstock GM, Garrity GM, Dodsworth JA, Yooseph S, Sutton G, Glöckner FO, Gilbert JA, Nelson WC, Hallam SJ, Jungbluth SP, Ettema TJG, Tighe S, Konstantinidis KT, Liu WT, Baker BJ, Rattei T, Eisen JA, Hedlund B, McMahon KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Tyson GW, Rinke C, Genome Standards Consortium, Lapidus A, Meyer F, Yilmaz P, Parks DH, Eren AM, 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.
16. Sabath N, Ferrada E, Barve A, Wagner A. 2013. Growth temperature and genome size in Bacteria are negatively correlated, suggesting genomic streamlining during thermal adaptation. Genome Biol Evol 5:966–977. https://doi.org/10.1093/gbe/evt050.
17. Golyshina OV, Toshchakov SV, Makarova KS, Gavrilov SN, Korzhenkov AA, La Cono V, Arcadi E, Nechitaylo TY, Ferrer M, Kublanov IV, Wolf YI, Yakimov MM, Golyshin PN. 2017. 'ARMAN'archaea depend on association with euryarchaeal host in culture and in situ. Nat Commun 8:60. https://doi.org/10.1038/s41467-017-00104-7.
18. Baker BJ, Comolli LR, Dick GJ, Hauser LJ, Hyatt D, Dill BD, Land ML, Verberkmoes NC, Hettich RL, Banfield JF. 2010. Enigmatic, ultrasmall, uncultivated Archaea. Proc Natl Acad Sci U S A 107:8806–8811. https://doi.org/10.1073/pnas.0914470107.
19. Sapra R, Bagramyan K, Adams MW. 2003. A simple energy-conserving system: proton reduction coupled to proton translocation. Proc Natl Acad Sci U S A 100:7545–7550. https://doi.org/10.1073/pnas.1331436100.
20. Imanaka H, Yamatsu A, Fukui T, Atomi H, Imanaka T. 2006. Phosphoenolpyruvate synthase plays an essential role for glycolysis in the modified Embden-Meyerhof pathway in *Thermococcus kodakarensis*. Mol Microbiol 61:898–909. https://doi.org/10.1111/j.1365-2958.2006.05287.x.
21. Baykov AA, Malinen AM, Luoto HH, Lahti R. 2013. Pyrophosphate-fueled Na^+ and H^+ transport in prokaryotes. Microbiol Mol Biol Rev 77:267–276. https://doi.org/10.1128/MMBR.00003-13.
22. Sieber CMK, Paul BG, Castelle CJ, Hu P, Tringe SG, Valentine DL, Andersen GL, Banfield JF. 2019. Unusual metabolism and hypervariation in the genome of a gracilibacterium (BD1-5) from an oil-degrading community. mBio 10:e02128-19. https://doi.org/10.1128/mBio.02128-19.
23. Wrighton KC, Thomas BC, Sharon I, Miller CS, Castelle CJ, VerBerkmoes NC, Wilkins MJ, Hettich RL, Lipton MS, Williams KH, Long PE, Banfield JF. 2012. Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. Science 337:1661–1665. https://doi.org/10.1126/science.1224041.
24. Blesák K, Janeček Š. 2012. Sequence fingerprints of enzyme specificities from the glycoside hydrolase family GH57. Extremophiles 16:497–506. https://doi.org/10.1007/s00792-012-0449-9.
25. Møller MS, Henriksen A, Svensson B. 2016. Structure and function of α-glucan debranching enzymes. Cell Mol Life Sci 73:2619–2641. https://doi.org/10.1007/s00018-016-2241-y.
26. Jaffe AL, Castelle CJ, Dupont CL, Banfield JF. 2019. Lateral gene transfer shapes the distribution of RuBisCO among candidate phyla radiation Bacteria and DPANN Archaea. Mol Biol Evol 36:435–446. https://doi.org/10.1093/molbev/msy234.
27. Sato T, Atomi H, Imanaka T. 2007. Archaeal type III RuBisCOs function in a pathway for AMP metabolism. Science 315:1003–1006. https://doi.org/10.1126/science.1135999.
28. Greening C, Biswas A, Carere CR, Jackson CJ, Taylor MC, Stott MB, Cook GM, Morales SE. 2016. Genomic and metagenomic surveys of hydrogenase distribution indicate H_2 is a widely utilised energy source for microbial growth and survival. ISME J 10:761–777. https://doi.org/10.1038/ismej.2015.153.

Downloaded from https://journals.asm.org/journal/msystems on 27 August 2021 by 71.38.0.28.

29. Hamm JN, Erdmann S, Eloe-Fadrosh EA, Angeloni A, Zhong L, Brownlee C, Williams TJ, Barton K, Carswell S, Smith MA, Brazendale S, Hancock AM, Allen MA, Raftery MJ, Cavicchioli R. 2019. Unexpected host dependency of Antarctic *Nanohaloarchaeota*. Proc Natl Acad Sci U S A 116:14661–-14670. https://doi.org/10.1073/pnas.1905179116.
30. La Cono V, Messina E, Rohde M, Arcadi E, Ciordia S, Crisafi F, Denaro R, Ferrer M, Giuliano L, Golyshin PN, Golyshina OV, Hallsworth JE, La Spada G, Mena MC, Merkel AY, Shevchenko MA, Smedile F, Sorokin DY, Toshchakov SV, Yakimov MM. 2020. Symbiosis between nanohaloarchaeon and haloarchaeon is based on utilization of different polysaccharides. Proc Natl Acad Sci U S A 117:20223–20234. https://doi.org/10.1073/pnas.2007232117.
31. Jahn U, Gallenberger M, Paper W, Junglas B, Eisenreich W, Stetter KO, Rachel R, Huber H. 2008. *Nanoarchaeum equitans* and *Ignicoccus hospitalis*: new insights into a unique, intimate association of two archaea. J Bacteriol 190:1743–1750. https://doi.org/10.1128/JB.01731-07.
32. Chen LX, Méndez-García C, Dombrowski N, Servín-Garcidueñas LE, Eloe-Fadrosh EA, Fang BZ, Luo ZH, Tan S, Zhi XY, Hua ZS, Martinez-Romero E, Woyke T, Huang LN, Sánchez J, Peláez AI, Ferrer M, Baker BJ, Shu WS. 2018. Metabolic versatility of small archaea *Micrarchaeota* and *Parvarchaeota*. ISME J 12:756–775. https://doi.org/10.1038/s41396-017-0002-z.
33. Villanueva L, von Meijenfeldt FB, Westbye AB, Hopmans EC, Dutilh BE, Damste JS. 2018. Bridging the divide: bacteria synthesizing archaeal membrane lipids. bioRxiv https://doi.org/10.1101/448035.
34. Coleman GA, Pancost RD, Williams TA. 2018. Extensive transfer of membrane lipid biosynthetic genes between Archaea and Bacteria. bioRxiv https://doi.org/10.1101/312991.
35. Albers SV, Meyer BH. 2011. The archaeal cell envelope. Nat Rev Microbiol 9:414–427. https://doi.org/10.1038/nrmicro2576.
36. Pohlschroder M, Pfeiffer F, Schulze S, Halim MFA. 2018. Archaeal cell surface biogenesis. FEMS Microbiol Rev 42:694–717. https://doi.org/10.1093/femsre/fuy027.
37. Albers SV, Jarrell KF. 2015. The archaellum: how Archaea swim. Front Microbiol 6:23. https://doi.org/10.3389/fmicb.2015.00023.
38. Hedlund BP, Thomas SC, Dodsworth JA, Zhang CL. 2016. Life in high-temperature environments, p 863–877. *In* Yates M, Nakatsu CH, Miller RV, Pillai SD (ed), Manual of environmental microbiology, 4th ed. ASM Press, Washington, DC.
39. Bains W, Xiao Y, Yu C. 2015. Prediction of the maximum temperature for life based on the stability of metabolites to decomposition in water. Life (Basel) 5:1054–1100. https://doi.org/10.3390/life5021054.
40. López-García P, Zivanovic Y, Deschamps P, Moreira D. 2015. Bacterial gene import and mesophilic adaptation in archaea. Nat Rev Microbiol 13:447–456. https://doi.org/10.1038/nrmicro3485.
41. Zeilstra-Ryalls J, Fayet O, Georgopoulos C. 1991. The universally conserved GroE (Hsp60) chaperonins. Annu Rev Microbiol 45:301–325. https://doi.org/10.1146/annurev.mi.45.100191.001505.
42. Nakai R, Abe T, Takeyama H, Naganuma T. 2011. Metagenomic analysis of 0.2-μm-passable microorganisms in deep-sea hydrothermal fluid. Mar Biotechnol (NY) 13:900–908. https://doi.org/10.1007/s10126-010-9351-6.
43. 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.
44. Kampmann M, Stock D. 2004. Reverse gyrase has heat-protective DNA chaperone activity independent of supercoiling. Nucleic Acids Res 32:3537–3545. https://doi.org/10.1093/nar/gkh683.
45. Petitjean C, Moreira D, López-García P, Brochier-Armanet C. 2012. Horizontal gene transfer of a chloroplast DnaJ-Fer protein to *Thaumarchaeota* and the evolutionary history of the DnaK chaperone system in Archaea. BMC Evol Biol 12:226. https://doi.org/10.1186/1471-2148-12-226.
46. Dupuis MÈ, Villion M, Magadán AH, Moineau S. 2013. CRISPR-Cas and restriction–modification systems are compatible and increase phage resistance. Nat Commun 4:2087. https://doi.org/10.1038/ncomms3087.
47. Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, Amitai G, Sorek R. 2018. Systematic discovery of antiphage defense systems in the microbial pangenome. Science 359:eaar4120. https://doi.org/10.1126/science.aar4120.
48. Rachel R, Bettstetter M, Hedlund BP, Häring M, Kessler A, Stetter KO, Prangishvili D. 2002. Remarkable morphological diversity of viruses and virus-like particles in hot terrestrial environments. Arch Virol 147:2419–2429. https://doi.org/10.1007/s00705-002-0895-2.
49. Breitbart M, Wegley L, Leeds S, Schoenfeld T, Rohwer F. 2004. Phage community dynamics in hot springs. Appl Environ Microbiol 70:1633–1640. https://doi.org/10.1128/AEM.70.3.1633-1640.2004.
50. Ortmann AC, Wiedenheft B, Douglas T, Young M. 2006. Hot crenarchaeal viruses reveal deep evolutionary connections. Nat Rev Microbiol 4:520–528. https://doi.org/10.1038/nrmicro1444.
51. Burstein D, Sun CL, Brown CT, Sharon I, Anantharaman K, Probst AJ, Thomas BC, Banfield JF. 2016. Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. Nat Commun 7:10613. https://doi.org/10.1038/ncomms10613.
52. Zhaxybayeva O, Swithers KS, Lapierre P, Fournier GP, Bickhart DM, DeBoy RT, Nelson KE, Nesbø CL, Doolittle WF, Gogarten JP, Noll KM. 2009. On the chimeric nature, thermophilic origin, and phylogenetic placement of the *Thermotogales*. Proc Natl Acad Sci U S A 106:5865–5870. https://doi.org/10.1073/pnas.0901260106.
53. Dombrowski N, Williams TA, Sun J, Woodcroft BJ, Lee JH, Minh BQ, Rinke C, 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.
54. Nelson-Sathi S, Dagan T, Landan G, Janssen A, Steel M, McInerney JO, Deppenmeier U, Martin WF. 2012. Acquisition of 1,000 eubacterial genes physiologically transformed a methanogen at the origin of *Haloarchaea*. Proc Natl Acad Sci U S A 109:20537–20542. https://doi.org/10.1073/pnas.1209119109.
55. Soucy SM, Huang JL, Gogarten JP. 2015. Horizontal gene transfer: building the web of life. Nat Rev Genet 16:472–482. https://doi.org/10.1038/nrg3962.
56. Giannone RJ, Wurch LL, Heimerl T, Martin S, Yang Z, Huber H, Rachel R, Hettich RL, Podar M. 2015. Life on the edge: functional genomic response of *Ignicoccus hospitalis* to the presence of *Nanoarchaeum equitans*. ISME J 9:101–114. https://doi.org/10.1038/ismej.2014.112.
57. Hotopp JCD, Clark ME, Oliveira DCSG, Foster JM, Fischer P, Torres MCM, Giebel JD, Kumar N, Ishmael N, Wang S, Ingram J, Nene RV, Shepard J, Tomkins J, Richards S, Spiro DJ, Ghedin E, Slatko BE, Tettelin H, Werren JH. 2007. Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. Science 317:1753–1756. https://doi.org/10.1126/science.1142490.
58. Podar M, Anderson I, Makarova KS, Elkins JG, Ivanova N, Wall MA, Lykidis A, Mavromatis K, Sun H, Hudson ME, Chen W, Deciu C, Hutchison D, Eads JR, Anderson A, Fernandes F, Szeto E, Lapidus A, Kyrpides NC, Saier MH, Jr, Richardson PM, Rachel R, Huber H, Eisen JA, Koonin EV, Keller M, Stetter KO. 2008. A genomic analysis of the archaeal system *Ignicoccus hospitalis-Nanoarchaeum equitans*. Genome Biol 9:R158. https://doi.org/10.1186/gb-2008-9-11-r158.
59. Jarett JK, Nayfach S, Podar M, Inskeep W, Ivanova NN, Munson-McGee J, Schulz F, Young M, Jay ZJ, Beam JP, Kyrpides NC, Malmstrom RR, Stepanauskas R, Woyke T. 2018. Single-cell genomics of co-sorted *Nanoarchaeota* suggests novel putative host associations and diversification of proteins involved in symbiosis. Microbiome 6:161. https://doi.org/10.1186/s40168-018-0539-8.
60. Castelle CJ, Banfield JF. 2018. Major new microbial groups expand diversity and alter our understanding of the tree of life. Cell 172:1181–1197. https://doi.org/10.1016/j.cell.2018.02.016.
61. Waters E, Hohn MJ, Ahel I, Graham DE, Adams MD, Barnstead M, Beeson KY, Bibbs L, Bolanos R, Keller M, Kretz K, Lin X, Mathur E, Ni J, Podar M, Richardson T, Sutton GG, Simon M, Soll D, Stetter KO, Short JM, Noordewier M. 2003. The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism. Proc Natl Acad Sci U S A 100:12984–12988. https://doi.org/10.1073/pnas.1735403100.
62. Schwank K, Bornemann TL, Dombrowski N, Spang A, Banfield JF, Probst AJ. 2019. An archaeal symbiont-host association from the deep terrestrial subsurface. ISME J 13:2135–2139. https://doi.org/10.1038/s41396-019-0421-0.
63. Probst AJ, Ladd B, Jarett JK, Geller-McGrath DE, Sieber CMK, Emerson JB, Anantharaman K, Thomas BC, Malmstrom RR, Stieglmeier M, Klingl A, Woyke T, Ryan MC, Banfield JF. 2018. Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. Nat Microbiol 3:328–336. https://doi.org/10.1038/s41564-017-0098-y.
64. Merkel AY, Pimenov NV, Rusanov II, Slobodkin AI, Slobodkina GB, Tarnovetckii IY, Frolov EN, Dubin AV, Perevalova AA, Bonch-Osmolovskaya EA. 2017. Microbial diversity and autotrophic activity in Kamchatka hot springs. Extremophiles 21:307–317. https://doi.org/10.1007/s00792-016-0903-1.
65. Liu X, Li M, Castelle CJ, Probst AJ, Zhou Z, Pan J, Liu Y, Banfield JF, Gu JD. 2018. Insights into the ecology, evolution, and metabolism of the

Downloaded from https://journals.asm.org/journal/msystems on 27 August 2021 by 71.38.0.28.

widespread Woesearchaeotal lineages. Microbiome 6:102. https://doi.org/10.1186/s40168-018-0488-2.
66. Jay ZJ, Beam JP, Dohnalkova A, Lohmayer R, Bodle B, Planer-Friedrich B, Romine M, Inskeep WP. 2015. *Pyrobaculum yellowstonensis* strain WP30 respires on elemental sulfur and/or arsenate in circumneutral sulfidic geothermal sediments of yellowstone national park. Appl Environ Microbiol 81:5907–5916. https://doi.org/10.1128/AEM.01095-15.
67. Hua ZS, Han YJ, Chen LX, Liu J, Hu M, Li SJ, Kuang JL, Chain PSG, Huang LN, Shu WS. 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.
68. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 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.
69. Kang DD, Froula J, Egan R, 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.
70. Dick GJ, Andersson AF, Baker BJ, Simmons SL, Thomas BC, Yelton AP, Banfield JF. 2009. Community-wide analysis of microbial genome sequence signatures. Genome Biol 10:R85. https://doi.org/10.1186/gb-2009-10-8-r85.
71. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 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.
72. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. https://doi.org/10.1186/1471-2105-11-119.
73. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 45:D353–D361. https://doi.org/10.1093/nar/gkw1092.
74. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa S, Kuhn M, Jensen LJ, von Mering C, Bork P. 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res 44:D286–D293. https://doi.org/10.1093/nar/gkv1248.
75. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44:D279–D285. https://doi.org/10.1093/nar/gkv1344.
76. Makarova KS, Wolf YI, Koonin EV. 2015. Archaeal clusters of orthologous genes (arCOGs): an update and application for analysis of shared features between *Thermococcales*, *Methanococcales*, and *Methanobacteriales*. Life (Basel) 5:818–840. https://doi.org/10.3390/life5010818.
77. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59–60. https://doi.org/10.1038/nmeth.3176.
78. Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 35:3100–3108. https://doi.org/10.1093/nar/gkm160.
79. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964. https://doi.org/10.1093/nar/25.5.955.
80. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490–D495. https://doi.org/10.1093/nar/gkt1178.
81. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ, Brinkman FS. 2010. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 26:1608–1615. https://doi.org/10.1093/bioinformatics/btq249.
82. Grissa I, Vergnaud G, Pourcel C. 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res 35:W52–W57. https://doi.org/10.1093/nar/gkm360.
83. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hernsdorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM, Amundson R, Thomas BC, Banfield JF. 2016. A new view of the tree of life. Nat Microbiol 1:16048. https://doi.org/10.1038/nmicrobiol.2016.48.
84. Wu M, Scott AJ. 2012. Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. Bioinformatics 28:1033–1034. https://doi.org/10.1093/bioinformatics/bts079.
85. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340.
86. Capella-Gutiérrez S, Silla-Martínez JM, 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.
87. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 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.
88. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14:587–589. https://doi.org/10.1038/nmeth.4285.
89. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, Brown CT, Porras-Alfaro A, Kuske CR, Tiedje JM. 2014. Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Res 42:D633–D642. https://doi.org/10.1093/nar/gkt1244.
90. Pruesse E, Peplies J, Glöckner FO. 2012. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics 28:1823–1829. https://doi.org/10.1093/bioinformatics/bts252.
91. Katoh K, Standley DM. 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. Zhu Q, Kosoy M, Dittmar K. 2014. HGTector: an automated method facilitating genome-wide discovery of putative horizontal gene transfers. BMC Genomics 15:717. https://doi.org/10.1186/1471-2164-15-717.
93. Olm MR, Brown CT, Brooks B, Banfield JF. 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.
94. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. 2020. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 36:1925–1927. https://doi.org/10.1093/bioinformatics/btz848.
95. Fu LM, Niu BF, Zhu ZW, Wu ST, Li WZ. 2012. CD-HIT: accelerated for clustering the next generation sequencing data. Bioinformatics 28:3150–3152. https://doi.org/10.1093/bioinformatics/bts565.
96. Tria FDK, Landan G, Dagan T. 2017. Phylogenetic rooting using minimal ancestor deviation. Nat Ecol Evol 1:0193. https://doi.org/10.1038/s41559-017-0193.
97. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, 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.
98. Friedman J, Alm EJ. 2012. Inferring correlation networks from genomic survey data. PLoS Comput Biol 8:e1002687. https://doi.org/10.1371/journal.pcbi.1002687.
99. Bastian M, Heymann S, Jacomy M. 2009. Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media.

Downloaded from https://journals.asm.org/journal/msystems on 27 August 2021 by 71.38.0.28.