Insights into the ecological roles and evolution of methyl-coenzyme M reductase-containing hot spring Archaea

Zheng-Shuang Hua1,2,14, Yu-Lin Wang3,14, Paul N. Evans4,14, Yan-Ni Qu1, Kian Mau Goh5, Yang-Zhi Rao1, Yan-Ling Qi1, Yu-Xian Li1, Min-Jun Huang2, Jian-Yu Jiao1, Ya-Ting Chen1, Yan-Ping Mao3,6, Wen-Sheng Shu7, Wael Hozzein8,9, Brian P. Hedlund10,11, Gene W. Tyson4,12*, Tong Zhang3* & Wen-Jun Li1,13*

Several recent studies have shown the presence of genes for the key enzyme associated with archaeal methane/alkane metabolism, methyl-coenzyme M reductase (Mcr), in metagenome-assembled genomes (MAGs) divergent to existing archaeal lineages. Here, we study the mcr-containing archaeal MAGs from several hot springs, which reveal further expansion in the diversity of archaeal organisms performing methane/alkane metabolism. Significantly, an MAG basal to organisms from the phylum Thaumarchaeota that contains mcr genes, but not those for ammonia oxidation or aerobic metabolism, is identified. Together, our phylogenetic analyses and ancestral state reconstructions suggest a mostly vertical evolution of mcrABG genes among methanogens and methanotrophs, along with frequent horizontal gene transfer of mcr genes between alkanotrophs. Analysis of all mcr-containing archaeal MAGs/genomes suggests a hydrothermal origin for these microorganisms based on optimal growth temperature predictions. These results also suggest methane/alkane oxidation or methanogenesis at high temperature likely existed in a common archaeal ancestor.

https://doi.org/10.1038/s41467-019-12574-y
Methane is an important compound in both the abiotic and biotic cycling of carbon on Earth. While the majority of biological methane formation is mediated by certain archaea, it has also been suggested that aerobic bacteria are responsible for a limited amount of methane biosynthesis. Likewise, methane can be oxidized by certain lineages of Archaea in anaerobic environments, with bacterial methanotrophs also responsible for methane oxidation in aerobic and some anaerobic environments. The archaea that mediate methane metabolism in anaerobic environments are dominated by organisms from a small number of lineages within the phylum Euryarchaeota that share a core metabolism centered around the methyl-coenzyme M reductase complex (Mcr). This complex catalyzes the terminal step in methanogenesis and initial step in methane oxidation. While these mcr-containing organisms have been shown to produce methane are restricted to the Euryarchaeota, there is much debate on the evolutionary origin of this complex. More recently, genes encoding for the predicted metabolism of methane, including divergent mcr genes, were identified in two metagenome-assembled genomes (MAGs) from the archaeal phylum Bathyrarchaeota. Further studies have suggested that these divergent mcr genes oxidize alkanes other than methane based on the detection of metabolic intermediates predicted to be formed from propane and butane oxidation by the Mcr complex. Subsequently, other archaeal MAGs from non-euryarchaeal lineages have been discovered that were found to contain mcr genes similar to cultivated methanogens. These Verstreetaarchaeota MAGs have been previously observed in a cultivation-independent survey of mcrA genes in several environments, before they were taxonomically linked to MAGs. Recently other MAGs that harbor mcr genes have been identified from archaeal lineages not previously associated with methane or alkane metabolism. Together these results further consolidate the wide distribution of mcr-mediated methane metabolizing archaea beyond our current knowledge, suggesting that the mcr catalysis should be expanded to include alkane oxidation, and underscoring the need for the axenic cultivation of these organisms.

Here, we successfully reconstruct MAGs containing mcr gene clusters using bioinformatic methods from metagenomes of high spring microbial communities, which represent potential methanogens, methanotrophs, or alkanotrophs. Phylogenomic analysis shows that these genomes belong within the TACK superphylum and Euryarchaeota phylum, but often form distinct clades divergent from cultivated lineages. Similarly, metabolic reconstructions of these organisms inferred from high-quality MAGs suggest unreported methane and alkane metabolism pathways not seen in previous genomes. Also, evolutionary analyses of these and database MAGs revealed methanogenesis or methane oxidation may be a primordial form of energy metabolism in early free-living archaea. The ancestral state of optimal growth temperature indicated these lineages represent an ancient metabolic capacity that originated from thermal habitats. Overall, this study significantly expands our current knowledge about the diversity of methanogens, methanotrophs and alkanotrophs, summarizes recently identified lineages, and sheds further light on the evolution of Mcr-mediated alkane metabolism within the Archaea.

**Results and discussion**

**Discovery of mcr-containing archaeal MAGs in hot springs.** A total of 14 mcr-containing archaeal MAGs were successfully reconstructed from six metagenomes that were sequenced from hot spring sediments located in Tengchong County of Yunnan, China (Table 1). Nine mcr-containing MAGs were assembled...
from JinZe (JZ) hot springs, which were collected from three different sites (named JZ-1, -2, -3). An additional three mcr-containing MAGs from GuMingQuan (GMQ), one from DiReTiYanQu-6 (DRTY-6) and one from ZiMeiQuan (ZMQ) hot spring sediments were also reconstructed. These hot springs are high-temperature (60–98 °C) environments that range from near-neutral to alkaline pH values (6–9.6) (Supplementary Table 1). Most MAGs were classified as high quality with sizes ranging from 0.95 to 1.73 Mbp, with completeness estimates of greater than 90% (Supplementary Fig. 1; Table 1). Both rRNAs and tRNAs (>29) are detectable in nearly all MAGs and is consistent with these MAGs being of high quality. The key taxonomic marker genes for archaeal methane and alkane metabolism, methyl-coenzyme reductase (mcrABG), were identified in each of these 14 MAGs on scaffolds with even sequence depth at medium to high coverage (Supplementary Fig. 2). In most cases, each mcr complex is located on a long scaffold (>50 Kbp) with several nearby genes annotated as being associated with methane metabolism, as well as tRNAs, chaperon proteins and/or ribosomal proteins interspersed within the corresponding scaffolds (Supplementary Fig. 3). In combination, these results strongly indicate the high accuracy of assembled mcr-containing scaffolds and the identified mcr genes do belong to the reconstructed MAGs. Phylogenomic analysis showed that these MAGs are widely distributed across both the TACK superphylum (10 MAGs) and the Euryarchaeota phylum (four MAGs) (Fig 1; Supplementary Data 1). Taxonomic placement revealed that four of the 10 TACK-associated MAGs form a clade that is sister to the Nezhaarchaeota representing a new order within this phylum (JZ bin_38, JZ-1 bin_66, GMQP bin_37 and ZMQR bin_18); five belong to Verstraetearchaeota (JZ-2 bin_200, JZ-3 bin_106, JZ-3 bin_107, GMQP bin_44 and DRTY-6 bin_144); and the last one branches deeply within the Thaumarchaeota (JZ-2 bin_220). The four euryarchaeal MAGs were found to belong to the Hadesarchaeota (JZ-1 bin_103 and JZ-2 bin_199), Methanomassiliicoccales (JZ-2 bin_168), and Archaeoglobales (GMQP bin_32) lineages (Fig. 1). The placement of these and other mcr-containing MAGs reveals deep branching within the respective lineages, and in the case of Thaumarchaeota, Nezhaarchaeota, Hadesarchaeota, Verstraetearchaeota, and Archaeoglobi MAGs, they are basal to the last common ancestor of their respective phyla (Fig. 1). This result also indicates that the metabolic features of these organisms may predate the metabolism of the more derived, well studied methane metabolizing organisms from the Euryarchaeota.

**Fig. 1** The phylogeny of reconstructed methanogenic and methanotrophic MAGs. Maximum-likelihood tree based on 892 archaeal genomes including 14 MAGs in this study was inferred from a concatenated set of 122 proteins using IQ-TREE with 1000 ultrafast bootstrapping iterations. Support values >70% are shown as black circles. Stars in circles represent the MAGs reconstructed in this study. The MAG OPBin_054 of Berghuis et al. was suggested to belong to the clade with JZ bin_38, JZ-1 bin_66, GMQP bin_37, and ZMQR bin_18 has been excluded from this analysis because it is only 17% complete and is taxonomically difficult to place. 

**Methane or alkane metabolism in hot spring Archaea.** The five mcr-containing MAGs identified as *Verstraetearchaeota* were each approximately ~1.5 Mb and are similar in size to previously described *Verstraetearchaeota* MAGs. These MAGs are also predicted to be H2-dependent methylo trophic methanogens based on the absence of a complete archaeal Wood-Ljungdahl carbon dioxide reduction pathway and the presence of methyltransferase/coronoid proteins (Fig. 2). Furthermore, the mcr genes derived from these hot spring organisms form a clade with mcr genes from other characterized H2-dependent methanogens from the clades...
Methanomassiliicoccales, Methanotronarchaeia, Methanofastidiosales, and existing Verstraetearchaeota genomes, (Fig. 3a; Supplementary Data 2). Based on the presence of genes for methyltransferases and corrinoid proteins, the hot spring Verstraetearchaeota may form methane from methylated amines, methanol, methylsulfides, methanethiol, or other unknown methylated compounds (Fig. 2). Although, the absence of pyrrolysine biosynthesis and methylmethylene transferase (mtbA) genes suggest that methane formation from methylated amines is unlikely to occur. It is likely that methanogenesis from methylsulfides, methanethiol, or methanol occurs based on the presence of mtsA and mtaA genes in these Verstraetearchaeota MAGs (Fig. 2). Methylsulfides and methanethiol are common sulfur species observed in thermal springs. Congruent with other H2-dependent methylotrophic methanogens is the absence of genes for a carbon dioxide reduction pathway and tetrahydromethanopterin S-methyltransferase (mtrADBCDEFGH) that would allow hydrogenotrophic methanogenesis to proceed. While the complete mtr operon is absent, the presence of predicted corrinoid containing mtrH subunit homologs with adjacent methyltransferases may allow this organism to form methane from unknown methylated compounds. The absence of heterodisulfide reductase (hdrABC) subunits in Verstraetearchaeota MAGs (Supplementary Data 3) suggests that mechanisms for the regeneration of coenzyme M and B may be carried out by a proposed hdrD-glcD complex (Fig. 2). However, the presence of fpo genes (archaeal complex I) suggests energy is likely conserved in a similar manner to H2-dependent methanogens from the Methanomassiliicoccales, Methanotronarchaeia, Methanofastidiosales, and Verstraetearchaeota. In these organisms, energy could be conserved by an hdrD-like protein that regenerates the coenzyme M and B cofactors, with the concomitant translocation of protons across the membrane (Fig. 2). Also, the presence of the key energy-conserving complex ehhb seen in the coal seam and bioreactor Verstraetearchaeota genomes suggests a similar methane-forming metabolism. Core methanogenesis genes (mcrBCGCD, hdrD, glcD, mtsA, mtaA, mttBC) are also present in these organisms, and in previously published Verstraetearchaeota genomes, suggesting a similar metabolism type between all these organisms. Also, mcr-containing H2-dependent methanogens identified in other thermal springs have been suggested to possess a methane-forming metabolism similar to Methanomassiliicoccales methanogens. The inferred metabolism of these thermophilic Verstraetearchaeota organisms is not surprising as methylated compounds and hydrogen are key metabolic intermediates in thermal spring environments. While H2-dependent methylotrophic Verstraetearchaeota and Methanomassiliicoccales methanogens appear to be important in these hot spring environments, predicted hydrogenotrophic methanogens that form a clade sister to the Crenarchaeota are also present (Fig. 1). These organisms have been observed in several hot springs and appear to be the first hydrogenotrophic mcr-containing organisms outside the Euryarchaeota based on the presence of genes for the archaeal Wood-Ljungdahl pathway,
methyl-coenzyme M reductase complex (mcr; mcrABGCD), all subunits of tetrahydrodemitroanopterin-S-methyltransferase (mtr; mtrABCDEFGH), and heterodisulfide reductase (hdh; hdtABC) (Supplementary Data 3). We also show there is a greater diversity of these organisms from the same samples compared to Wang et al.17 (Fig. 1); however, there is little difference in the metabolic capabilities of these organisms based on gene complements (Fig. 2; Supplementary Data 3).

Another mcr-containing MAG, GMQP_bin_32, with a similar metabolic potential to the Nezhaarchaeota-like MAGs presented here and in Wang et al.17 was found to belong to the order Archaeoglobales (Fig. 1). However, the MAG only shared ~60% amino acid identity to the closest subgroup of organisms from the genus Archaeoglobus (Supplementary Fig. 4a). The similarity of the GMQP_bin_32 MAG to the Nezhaarchaeota-like MAGs suggests it is also a hydrogenotrophic methanogen consistent with the results of Wang et al.17. This would also be consistent with speculation that most Archaeoglobales lost the metabolic capability for methane metabolism19 but is intriguing as recent analysis of another Archaeoglobales genome (IdFR-42) that harbored mcr genes was predicted to perform alkane oxidation14,17. Therefore, the ability for methanotrophy in GMQP_bin_32 and the Nezhaarchaeota-like MAGs cannot be ruled out (Fig. 2). Together these results also suggest that the last common ancestor of these Archaeoglobales organisms possibly possessed the Mcr complex.

A further MAG from outside the Euryarchaeota was found to cluster with the Thaumarcheota and had a combination of methane metabolism genes that were distinct from other mcr-containing hot spring organisms (Fig. 2). Based on the presence of genes for methyltransferases and corrinoid proteins (mmtBC, mtaA, and mtaA), along with the absence of a complete mtrABCDEFGH operon, it is suggested that this organism is a H2-dependent methylotrophic methanogen. Consistent with its predicted methylotrophic lifestyle, the Thaumarcheota organism does possess mtrAH genes suggesting metabolism of unknown methylated compounds could occur. Furthermore, genes that encode for an incomplete archaeal complex I suggests an electron carrier other than F420, such as ferredoxin seen in Methanothrix organisms28.

While many of these hot spring microbial community MAGs appear to have conventional mcr genes associated with methanogenesis or methanotrophy (Fig. 2). A group of archaea associated with Hadesarchaeota MAGs contain mcr-like genes similar to Batharchaeota and Ca. Syntrophoarchaeum organisms that have been associated with Mcr-mediated alkane oxidation9,29 (Fig. 3a). These Hadesarchaeota MAGs also possess genes associated with the archaeal Wood-Ljungdahl, β-oxidation, and
other pathways similar to those found in other mcr-containing Hadesarchaeota genomes that have been recently identified. Previously identified non-mcr-containing Hadesarchaeota genomes were reported to have the ability to metabolize sugars and amino acids in a heterotrophic lifestyle, oxidize carbon monoxide (COxMLSIP), and perform dissimilatory nitrite reduction to ammonium. These current mcr-containing Hadesarchaeota MAGs in this study do not contain these pathways. While Wang et al. identified only a single copy of a mcr gene in each of their Hadesarchaeota MAGs, a second copy of mcrA in a single mcr operon was identified in the analysis of our MAGs (Supplementary Figs. 2 and 3). This operon structure may be the result of gene duplication and may lead to the neofunctionalisation of the Mcr complex to metabolize short chain alkanes of different lengths suggested recently by Evans et al.

The expansion of mcr gene diversity from the MAGs presented here and in other studies reveals organisms with relatively similar mcr complexes from across the archaeal species tree (Fig. 3a, d). These organisms have characteristics of either H2-dependent methanotrophic or hydrogenotrophic methanogens, although given the similarity to pathways from Anaerobic Methanotrophs (ANME), the ability of these organisms to oxidize methane cannot be ruled out. Overall, the discovery of these many novel MAGs suggests that yet more novel lineages of mcr-containing Archaea will be identified in the future, with mcr-containing Asgardarchaeota being a recent example of this.

Carbon fixation. Four of the six mcr-containing MAG clades (Nezhaarchaeota, Thaumarchaeota, Hadesarchaeota, and Archaeoglobales) appear to encode a conventional archaeal Wood-Ljungdahl pathway carbon dioxide fixation pathway, that includes genes for acetyl-CoA synthase/carbon monoxide dehydrogenase (acs/codh), and pyruvate oxidoreductase complexes (porABDG) (Fig. 2; Supplementary Data 3). This result is consistent with many archaea, including methanogens, which use this pathway to fix organic carbon, which has been suggested in previous studies.

Ribulose 1,5- biphosphate carboxylases/oxygenases (RuBisCOs) were identified in Archaeoglobobales, Methanomassiliicoccales and four of the five Verstraetearchaeota MAGs (Supplementary Fig. 5). Among them, the Archaeoglobales and one of the Verstraetearchaeota MAGs harbored form III-b RuBisCOs, suggesting the potential function in a nucleoside-salvaging pathway instead of carbon dioxide fixation via the Calvin–Benson–Bassham (CBB) cycle. The remaining MAGs have RuBisCOs categorized as the form III-a group that is specific to methanogenic archaea. No complete CBB cycle was identified in these MAGs. It was suggested that form III-a RuBisCOs may be involved in the reductive hexulose-phosphate pathway to fix carbon dioxide, with energy and reducing power supported by methanogenic pathways.

However, no phosphoribulokinase was detected in any of the MAGs. This may be the result of the high-temperature environment driving genome streamlining, leading to a deficiency in this autotrophic pathway.

Overall, there appears to be a diverse range of mechanisms for fixing carbon in these mcr-containing MAGs, reflecting the wide diversity of these lineages across the archaeal tree. This result is congruent with the multiple mechanisms that organisms across the archaean possess ability to fix carbon. Four of the six genome lineages also appear to generate energy from ß-oxidation of fatty acids or other linear chain hydrocarbons (Fig. 2). In the case of the Hadesarchaeota, the presence of this pathway is not surprising because a modified ß-oxidation pathway has been proposed previously in this lineage for Mcr-mediated alkane oxidation. Alkane oxidation has been seen in archaea previously including those belonging to the order Archaeoglobales, but is mediated by the addition of fumarate rather than ß-oxidation.

Evolution of methane and alkane metabolism. Previous studies have shown that Mcr-mediated methane and alkane metabolism is widespread across the archaeal tree based on the expanding taxonomy of archaea with these metabolism types. However, it is unclear if this wide diversity of organisms carrying these genes is the result of vertical descent or horizontal gene transfer (HGT), based on the paucity of mcr gene-containing lineages within the TACK superphylum. Here, the recovery and analysis of MAGs containing mcr genes from hot springs provides stronger evidence that mcr within TACK superphylum genomes are the result of vertical descent (Fig. 3a). These lineages include those that are notionally within the Nezhaarchaeota (four MAGs), Verstraetearchaeota (five MAGs), Korarchaeota (three MAGs), and Thaumarchaeota (one MAG) (Fig. 1). The mcr-containing scaffolds from these MAGs always harbor genes related to methane metabolism including several putative methanogenesis marker proteins, genes related to energy conservation with the exception of Hadesarchaeota (Supplementary Fig. 3). In these cases, HGT is unlikely based on the absence of differences in DNA sequence composition as there is little variation in the mcr-containing sequence composition above the natural variation in their respective MAGs (Supplementary Table 2).

For the mcr-containing TACK lineages that fall within or sister to Verstraetearchaeota MAGs, the genome and mcr gene taxonomies appear to be congruent (Figs. 1 and 3a) and strengthens the case for the mcr complex being present in the common ancestor of the TACK and Euryarchaeota lineages. The presence of the euryarchaeal Archaeoglobobus GMQP bin_32 MAG Mcr sequences sister to the Nezhaarchaeota and H2-dependent methanogens (Figs. 1 and 3a) could be explained by HGT. It has been suggested that this mechanism is common in Archaeoglobi and has been suggested as a driver of gaining sulfate reduction at the expense of methane or alkane metabolism in these organisms. Ribulose 1,5- biphosphate carboxylases/oxygenases (RuBisCOs) were identified in Archaeoglobobales, Methanomassiliicoccales and four of the five Verstraetearchaeota MAGs (Supplementary Fig. 5). Among them, the Archaeoglobobales and one of the Verstraetearchaeota MAGs harbored form III-b RuBisCOs, suggesting the potential function in a nucleoside-salvaging pathway instead of carbon dioxide fixation via the Calvin–Benson–Bassham (CBB) cycle. The remaining MAGs have RuBisCOs categorized as the form III-a group that is specific to methanogenic archaea. No complete CBB cycle was identified in these MAGs. It was suggested that form III-a RuBisCOs may be involved in the reductive hexulose-phosphate pathway to fix carbon dioxide, with energy and reducing power supported by methanogenic pathways. However, no phosphoribulokinase was detected in any of the MAGs. This may be the result of the high-temperature environment driving genome streamlining, leading to a deficiency in this autotrophic pathway.

Overall, there appears to be a diverse range of mechanisms for fixing carbon in these mcr-containing MAGs, reflecting the wide diversity of these lineages across the archaeal tree. This result is congruent with the multiple mechanisms that organisms across the archaean possess ability to fix carbon. Four of the six genome lineages also appear to generate energy from ß-oxidation of fatty acids or other linear chain hydrocarbons (Fig. 2). In the case of the Hadesarchaeota, the presence of this pathway is not surprising because a modified ß-oxidation pathway has been proposed previously in this lineage for Mcr-mediated alkane oxidation. Alkane oxidation has been seen in archaea previously including those belonging to the order Archaeoglobales, but is mediated by the addition of fumarate rather than ß-oxidation.

Evolution of methane and alkane metabolism. Previous studies have shown that Mcr-mediated methane and alkane metabolism is widespread across the archaeal tree based on the expanding taxonomy of archaea with these metabolism types. However, it is unclear if this wide diversity of organisms carrying these genes is the result of vertical descent or horizontal gene transfer (HGT), based on the paucity of mcr gene-containing lineages within the TACK superphylum. Here, the recovery and analysis of MAGs containing mcr genes from hot springs provides stronger evidence that mcr within TACK superphylum genomes are the result of vertical descent (Fig. 3a). These lineages include those that are notionally within the Nezhaarchaeota (four MAGs), Verstraetearchaeota (five MAGs), Korarchaeota (three MAGs), and Thaumarchaeota (one MAG) (Fig. 1). The mcr-containing scaffolds from these MAGs always harbor genes related to methane metabolism including several putative methanogenesis marker proteins, genes related to energy conservation with the exception of Hadesarchaeota (Supplementary Fig. 3). In these cases, HGT is unlikely based on the absence of differences in DNA sequence composition as there is little variation in the mcr-containing sequence composition above the natural variation in their respective MAGs (Supplementary Table 2). For the mcr-containing TACK lineages that fall within or sister to Verstraetearchaeota MAGs, the genome and mcr gene taxonomies appear to be congruent (Figs. 1 and 3a) and strengthens the case for the mcr complex being present in the common ancestor of the TACK and Euryarchaeota lineages. The presence of the euryarchaeal Archaeoglobobus GMQP bin_32 MAG Mcr sequences sister to the Nezhaarchaeota and H2-dependent methanogens (Figs. 1 and 3a) could be explained by HGT. It has been suggested that this mechanism is common in Archaeoglobi and has been suggested as a driver of gaining sulfate reduction at the expense of methane or alkane metabolism in these organisms. Ribulose 1,5- biphosphate carboxylases/oxygenases (RuBisCOs) were identified in Archaeoglobobales, Methanomassiliicoccales and four of the five Verstraetearchaeota MAGs (Supplementary Fig. 5). Among them, the Archaeoglobobales and one of the Verstraetearchaeota MAGs harbored form III-b RuBisCOs, suggesting the potential function in a nucleoside-salvaging pathway instead of carbon dioxide fixation via the Calvin–Benson–Bassham (CBB) cycle. The remaining MAGs have RuBisCOs categorized as the form III-a group that is specific to methanogenic archaea. No complete CBB cycle was identified in these MAGs. It was suggested that form III-a RuBisCOs may be involved in the reductive hexulose-phosphate pathway to fix carbon dioxide, with energy and reducing power supported by methanogenic pathways. However, no phosphoribulokinase was detected in any of the MAGs. This may be the result of the high-temperature environment driving genome streamlining, leading to a deficiency in this autotrophic pathway.

Overall, there appears to be a diverse range of mechanisms for fixing carbon in these mcr-containing MAGs, reflecting the wide diversity of these lineages across the archaeal tree. This result is congruent with the multiple mechanisms that organisms across the archaean possess ability to fix carbon. Four of the six genome lineages also appear to generate energy from ß-oxidation of fatty acids or other linear chain hydrocarbons (Fig. 2). In the case of the Hadesarchaeota, the presence of this pathway is not surprising because a modified ß-oxidation pathway has been proposed previously in this lineage for Mcr-mediated alkane oxidation. Alkane oxidation has been seen in archaea previously including those belonging to the order Archaeoglobales, but is mediated by the addition of fumarate rather than ß-oxidation.
branching order differences for the methane metabolizing Ca. Methanophages ANME-1-THS and Ca. Methanofastidiosa lineages also suggest a potential HGT event (Fig. 3a, b).

Further expansion of mcr-containing organisms to include the *Thaumarchaeota* associated JZ-2 bin_220 MAG suggests an interesting evolutionary story in the context of this metabolism type. Phylogenomic analysis places this MAG close to the pSL12 lineage (DRTY-7 bin_36) that does not contain mcr genes (Supplementary Fig. 6a). Unlike other *Thaumarchaeota* genomes and previously cultured isolates, JZ-2 bin_220 does not contain genes for ammonia oxidation (amoABC) or aerobic respiration (aa₅, bd, cbb₃, and bo₂-cytochromes) and suggests that the ancestor of *Thaumarchaeota* was an anaerobe with the ability for methane metabolism. Along with ancestral gene changes, JZ-2 bin_220 shows a significant alteration in gene complement with the acquisition of 457 and loss of 194, genes that likely led to the divergence of these aerobic and anaerobic lineages (Supplementary Fig. 6b; Supplementary Data 4). Consistent with the metabolism of the mcr-containing *Thaumarchaeota* MAG being the result of vertical descent, it is the basal nature of this genome within the *Thaumarchaeota* and the presence of many genes associated with methane metabolism (Supplementary Fig. 6c; Supplementary Data 4). Many of these genes, excluding mcr, are also found in the *Aigarchaeota* MAGs RBG_16_49_8, JGI MDM2 JNZ-1-N15, and JGI MDM2 JNZ-1-K18. For example, phylogeny of RNA polymerase sequences also places the *Aigarchaeota* into the *Thaumarchaeota* clade with a high bootstrap confidence.40,41. However, a definitive picture of the evolutionary history of these *Thaumarchaeota* organisms awaits additional genomic data from related organisms.

In the case of the mcr-containing *Hadesarchaeota* MAGs, they appear to have a high degree of similarity to *Hadesarchaeota* MAGs without mcr genes as 1125 gene families at the nucleotide (Supplementary Fig. 7a) and amino acid (Supplementary Fig. 7b) levels have high homology to each other. The presence of genes indicative of methane metabolism in JZ-1 bin_103 and JZ-2 bin_199 MAGs suggests there has either been a gain of genes by non-mcr-containing *Hadesarchaeota* MAGs or gene loss events resulting in the deficiency among non-mcr-containing *Hadesarchaeota*. The presence of genes, such as *mtd*, *mch*, and along with *hdr* genes suggests an ancestor of extant *Hadesarchaeota* organisms potentially, had the ability for methane or alkane oxidation (Supplementary Fig. 7c; Supplementary Data 4). Alternatively, these genes associated with methane metabolism were already present in the *Hadesarchaeota* MAGs and mcr genes and other associated genes were acquired in a manner that is similar to that predicted for *Archeoglobales* GMQP bin_32 (Supplementary Fig. 4) and *Bathyarchaeota* organisms.49,52.

Together these results reaffirm the results presented in Evans et al.32 and adds further to the complex interplay between methane- and alkane-metabolizing archaea due to vertical descent, gene duplication and neofunctionalisation along with some HGT of this metabolism type within the archaea. The now widespread nature of genes encoding the mcr complex across the archaea suggests this metabolism has a central role within archaeal metabolism alongside the similarly widespread and key acetyl-CoA synthase/carbon monoxide dehydrogenase complex.42-44. This result combined with the presence of archaeal Wood-Ljungdahl carbon dioxide fixation and β-oxidation pathways suggests a shared evolutionary history in these organisms of these metabolism types across the archaeal tree, despite being interspersed with non-mcr-containing lineages.42,43.

**Evolution of mcrABG genes and their hydrothermal origin.** The similar topologies for mcrA, mcrB, and mcrG genes show that these three genes have similar evolutionary histories (Supplementary Fig. 8a, b, c, d). This result supports the hypothesis that mcrABG has coevolved as an integral functional unit.45 While it is likely mcrABG has coevolved, discrepancies in the mcrG taxonomy were observed with the methanogenic Ca. Verstreeterarchaeota is showing higher divergence to most other TACK members instead of existing as a sister lineage to *Nezhaarchaeota* (Supplementary Fig. 8d). This may be an artefact due to the low confidence values at several parent nodes of the TACK lineage and the smaller length of mcrG compared to mcrA and mcrB gene subunits. It seems plausible that Mcr complexes are more conserved in TACK lineages compared to *Euryarchaeota* lineages based on the shorter branch lengths and the presence of only one copy per genome (Fig. 3d). In euryarchaeal methanogen lineages, two copies of the mcr operon were observed in *Methanobacteriales* and *Methanomicrobiales* and appear to have evolved in parallel (Fig. 3d; Supplementary Fig. 8). A likely duplication event has also occurred in the *Hadesarchaeota* MAGs presented here based on the organization of a second mcrA copy within an existing mcr gene operon and the phylogenetic distance between the two copies is small (Supplementary Fig. 8b). Also, duplications of mcr genes in the Ca. *Syntrophoarchaeum* and *Archeoglobi* MAGs appear to be more numerous and ancient.44

To unwind the evolutionary history of mcr-containing microorganisms, we used PAML.45 to computationally reconstruct the ancestral sequences of each internal node. By coupling this with an estimation of the optimal growth temperature (OGT), this gives us an opportunity to infer the environmental temperature of ancestral microbes and better understand their evolutionary history.46 For the traditional euryarchaeal methanogens, two major ancestral gene duplication events were observed, allowing them to survive in wide range of temperatures. Results show that CaP bias (difference between proportions of charged and polar non-charged amino acids) of concatenated mcrABG genes shows significant positive correlation with their OGTs (Person’s R² = 0.7288, P < 0.0000001; Fig. 3c). By reconstructing the ancestral amino acid sequences for each node, we observed a general pattern that the common ancestor of both methanogens and methanotrophs may have evolved from thermal habitats, as nodes near the root are predicted to be thermophilic. The predicted OGTs for the common ancestors of TACK-belonging methanogens/methanotrophs and alkanotrophs are ~93 and ~100°C, respectively, which are in line with a thermal origin for archaea.47. As expected, most extant genomes from hot spring habitats were predicted to be adapted to high temperatures ranging from 63–82°C, which are comparable to their habitats (Supplementary Table 1). To adapt to the thermal environments, thermophiles evolved both heat stable proteins and increased cell content of heat-shock proteins. Several MAGs contained genes encoding the heat-shock protein htpX, as well as high-temperature stress DNA repair systems encoded by radAB and RAD51 genes (Supplementary Data 3). All the genomic bins have the predicted capacity to synthesize archaeal membrane lipids. Unlike the bacterial lipid membranes composed of fatty acids linked to glycerol-3-phosphate via ester bonds, archaea possess isoprene-based alkyl chains linked by ether linkages to glycerol-1-phosphate, giving cell walls increased stability at high temperatures.48. Also, most bins harbored two copies of reverse DNA gyrase, which could stabilize DNA to cope with the increased temperature.49.

Fossil evidence suggests that geothermal habitats like submarine hydrothermal vents and terrestrial hot springs are a likely place where life on Earth first evolved,50,51 which is also supported by geochemical, biological, thermodynamic and phylogenetic inferences.43,44,52. On this basis and from the reconstructed archaean phylogenomic tree (Fig. 1) it is reasonable
to infer that thermophilic methanogenesis or methanotrophy may be a metabolism type associated with early life on earth. Consistent with this result we show the Mcr complex likely has a thermal origin based on ancestral state reconstruction from amino acid residues. Furthermore, it speculated that hydrogenotrophic methanogenesis via the Wood-Ljungdahl pathway might be the most primordial metabolisms for energy conservation and carbon dioxide fixation\textsuperscript{22-24}. Those thermophilic mcr-containing organisms that branch deeply in the concatenated McrABG tree (Fig. 3a, d) suggest a long evolutionary history of this metabolism type in archaea. This result would be consistent with some of the mcr-containing organisms from the TACK lineages harboring pathways for hydrogenotrophic methanogenesis could be primordial. Alternatively, by applying phylogenetic analysis of involved key enzymes in acetogenic, methanogenic, and methanotrophic pathways, the same view was proposed for methanotrophy being more likely than Wood-Ljungdahl pathway to date back to the emergence of life\textsuperscript{55}. Methane consumed by methanotrophs may come from the abiotic serpentinization, where seawater reacts with olivine at high temperatures to release iron oxide and various carbon compounds\textsuperscript{56}.

**Summary.** Together these results show that hot spring environments harbor many mcr-containing organisms from outside of the Euryarchaeota and further expand the diversity of Mcr mediated methane or alkane metabolism processes. The predicted wide range of metabolic mechanisms suggests that these organisms may utilize diverse and as yet unidentified substrates. Here, we propose a plausible evolutionary scenario where the common ancestor of archaea harbors the ability for methane metabolism that is mostly the result of vertical inheritance, with some HGT events. Frequent HGT events have also led to alkaneotrophy being found in several lineages that cannot be explained by vertical descent of mcr genes. Also, it is also likely that mcr gene duplication has led to changes in the substrate specificity to longer chain alkanes in several lineages within the Euryarchaeota. We also infer that these mcr-containing archaea may originate from thermal habitats such as hydrothermal vents or terrestrial hot springs predicted by a high ancestral optimal growth temperature. Overall, this study enables a better understanding of the origin of last common ancestor in Archaea using a combination of bioinformatic techniques.

**Methods**

Sample collection, DNA extraction, and sequencing. A total of six biomass samples were collected from four thermal spring sediments (IZ (JinZi), GMQ (Gu MingQuan), DRTY (DiTe YiYan), and ZMQ (Zi MeiQuan)) located at the collision boundary between the India and Eurasia plates near Tengchong city in Yunnan province (China). These hot springs span a wide range of physiochemical parameters with temperature ranging from 60 to 98 °C and pH ranging from 6.0 to 9.6 (see Supplementary Table 1 for the detailed geographical and physicochemical parameters). Three of the four springs are classiﬁed as containing bacteria and archaea habitats with temperatures greater than 80 °C. Following collection and DNA extraction, each of the six samples was subjected to metagenomic sequencing. Detailed method for sample collection, DNA extraction, and metagenomic sequencing is described in Hua et al.\textsuperscript{40}.

**Metagenomic assembly and genome binning.** Metagenomic assembly and genome binning were performed separately for each sample due to the significant differences in microbial community composition among the six samples (Supplementary Fig. 1). Briefly, the raw sequencing data were first preprocessed to remove adapters and duplicated sequences. Low quality reads with average phred value < 20 in a 50 bp sliding window were discarded. Remaining low quality sequences (phred value < 20) were trimmed at both ends. All quality control steps were conducted with custom Perl scripts\textsuperscript{57}. The quality reads were assembled using SPAdes\textsuperscript{48} (v3.9.0) with the parameters as--mapa-k 21.33,35.77,99.127. GapCloser (v1.12; http://soap.genomics.org.cn/) was used to eliminate the gaps within scaffolds. BBMap (v53.85; http://sourceforge.net/projects/bbmap/) was used to map all the quality reads onto assembled scaffolds with the parameters as k = 15 min = 0.9 build = 1. Then, the script “jgi_summaryize_bam_contig_depths” in MetaBAT\textsuperscript{59} (v2.12.1) was used to compute the coverage information of each scaffold. Genome binning for each sample was carried on scaffolds using MetaBAT with the coverage threshold set to 80% from each dataset. To further verify the taxonomically assembled genomes (MGAs), emergent self-organizing maps (ESOM)\textsuperscript{60} was performed to visualize the bins and scaffolds with abnormal coverage information or discordant positions were removed manually (Supplementary Fig. 1). Subsequently, reads mapped to the curated MGAs were re-clustered using SPAdes (v3.9.0) with the following options--careful -k 21.33,35.77,99.127. Further optimization of each MGAs was conducted as described above. The completeness, contamination and strain heterogeneity of each MAG was evaluated using CheckM\textsuperscript{61} (version 1.0.5). A total of 14 MAGs identified as containing mcrABG gene sequences from the assembled MGAs were identified using GrafM\textsuperscript{62} (v0.10.2).

**Phylogenetic and phylogenomic analysis.** To understand the phylogenetic relationship of the 14 genomes, a set of 122 conserved marker genes were retrieved from 1213 MAGs/genomes downloaded from public databases which covers nearly the whole archaeal diversity to date to build the phylogenomic tree\textsuperscript{41}. Where mcrABG genes were identified by the Relative Evolutionary Distant (RED) from the Genome Tree Database project were used for the assignment of taxonomic ranks\textsuperscript{41}. Only MGAs of medium to high quality (above a 70% CheckM quality metric) were used in analyses. The marker gene sets were extracted independently using AMPHORA\textsuperscript{28}. All identified marker genes were aligned individually using MUSCLE\textsuperscript{43} (v3.8.31). Poorly aligned regions were eliminated by TrimAl\textsuperscript{64} (v1.4.v22; -gt 0.95–cons 50). Individual alignments were concatenated and used as input to reconstruct the phylogenomic tree using IQ-TREE\textsuperscript{66} (v1.6.10) with the mixture model of LG + F + R10 and with ultrafast bootstrapping (-bb 1000), as well as Shimodaira–Hasegawa–like approximate likelihood-ratio test (SH-aLRT, -al 1000). The best model determined by ModelFinder\textsuperscript{67} is well supported by all criteria including Akaike Information Criterion (AIC), corrected AIC and Bayesian Information Criterion.

For functional genes, including concatenated mcrABG genes and individual subunits of these three genes, sequences were aligned using MUSCLE and IQ-tree was used to infer maximum-likelihood phylogenies with the same parameters as above. The best models for mcrA, mcrB, mcrG, and concatenated mcrABG genes were LG + P + I4 + G4, LG + F + R6, LG + I4 + G4 and LG + F + R6, respectively. We adopted two approaches to root the individual and concatenated mcrABG gene trees which place the root at the same place: phylogenetic rooting using the minimal ancestor deviation (MAD) method proposed by Tria et al.\textsuperscript{71} and Bayesian tree constructions using BEAST with a lognormal uncorrelated molecular clock\textsuperscript{72}. The generated trees in newick format were visualized by iTOL\textsuperscript{73}. v3.

**Optimal growth temperature (OGT) estimation.** It has been reported that OGT is significantly correlated to charged versus polar amino acid ratios (CvP bias)\textsuperscript{73,74}. Charged amino acids include arginine, lysine, aspartic acid, and glutamic acid, whereas the polar amino acids contain alanine, threonine, and serine. Here, 304 sequenced methanogens were downloaded from the National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/) genome repository database. Putative genes were predicted for all genomes using Prodigal\textsuperscript{78} with the “-p single -g 11” option. Orthologous genes were identified for each pairwise genome by extracting all reciprocal best BLAST hits (E-value < 1e-5). Average amino acid identity (AAI) was computed as the mean similarity of all orthologous genes. Single representatives were manually selected for subsequent analysis when two or more genomes showed an average AAI > 97%. A total of 117 genomes were used for OGT analyses including 25 from newly reported MAGs in recently published papers\textsuperscript{72-78}. Taxonomic information, sampling site and OGT of genomes were manually investigated from original publications and from the website of the German Collection of Microorganisms and Cell Cultures (http://www.dsmz.de), if available (Supplementary Data 5). Two psychrophiles were excluded since they evolved distinctly\textsuperscript{79}. The distribution of OGTs among these methanogens showed 98 species out of 102 OGTs of 37 °C. Linear regression was conducted based on the known OGTs and CvP biases of mcrABG genes. Species with two copies of mcr complex were computed twice with the same OGT. Significant correlation was observed between them by fitting the formula: y = 6.7254x + 13.858, with Pearson R = 0.738 and P < 0.00001 (Fig. 3c). Key genes including mcrA, mcrB and mcrG were searched against in-house database using AMPHORA2.
Ancient amino acid sequence reconstruction. Amino acid sequences at all internal nodes were computationally inferred based on the sequence alignment and tree topology of the reconstructed ARKO phylogenomic program in PAML 4.5 package (v4.9.4). An empirical Bayesian statistical framework incorporating Gamma distribution was employed for the inference of posterior amino acid probability per site and no molecular clock was set. The universal code (icode = 0) and fixed branch length option (Mgene = 0), as well as 10 categories in the distribution under LG model (suggested by ModelFinder as described above) were used throughout the PAML calculation. Both marginal and joint posterior probabilities were calculated in this program.

Comparative genomics. Comparative genomic analyses for Hadesarchaeota, Thaumarchaeota, and Archaeoglobales were conducted separately. Reference genomes for each group were downloaded from NCBI and IMG-M (https://img.jgi.doe.gov/m/m/mg1?source=ncbi) (Supplementary Data 6). Genome quality was estimated by CheckM and only genomes with completeness >80% and contamination <10% were taken into consideration for the later analysis. Due to the lack of representatives, all MAGs from Hadesarchaeota were included for the later analysis even their completeness <80%. Average amino acid identity of each pair of genomes of each group were inferred using COUNT78 (v9.1106) as described previously40. Briefly, all-against-all genomes BLAST for all the genomes in each group were performed to yield reciprocal best BLAST hits (rBBHs). MCL 76 algorithm (v14-137) was subsequently employed to cluster rBBHs into protein clusters. Bayesian trees were constructed using MrBayes77 (v3.2.6) by running four independent chains for 100 million generations for each run and sampling one tree every 1000 generations. These analysis starts with 0.1, 0.5, and 1 million generations for the deep-branching position in the Archaea. The type species is “Ca. Hadesarchaeum hydrogenovorans” (an.iti.qum) M. L. adj., also referring to Archaea. The type species is “Ca. Methanohydrogenoarchaeum antiquum”. “Ca. Methanohydrogenoarchaeum antiquum” (an.iti.qum) M. L. adj., also referring to Archaea. The type material is the metagenomic bin GMQP bin_44 (Ga0263253). Similarly, two new orders are proposed correspondingly, Methanomethylvoroarchaeota (Me.tha.no.me.thy.vo.ra) N.L. n. Methanomethylvoroarchaeum type genus of the order, and Methanomethylarchaeales (Me.tha.no.me.thy.ar.chae.ales) N.L. n. Methanomethylarchaeum type genus of the order; L. suff. -ales ending to denote an order.

Hadesarchaeum” (Ha.des.ar.chae.uch). M. G. n. Greek god of the underworld; Gr. adj. archeo ancient, also referring to Archaea. L. part. Hadesarchaeum, the archael god of the underworld. The type species is “Ca. Hadesarchaeum tengchongensis” “Ca. Hadesarchaeum tengchongensis” (teng.chong.en.siis) originating from Tengchong, a region of Yunnan Province, China. The type material is the metagenomic bin JZ-2 bin_199 (Ga0263248).

Methanourarchaeum” (Me.tha.no.ur.ar.chae.uch). M. L. n. methanum methane; ur, primitive, original, referring to the deeply branching position in the organism in phylogenetic analyses; Gr. adj. archeo ancient, also referring to Archaea. L. part. Methanourarchaeum, the primitive methanogen. The type species is “Ca. Methanourarchaeum thermodiluricium”. “Ca. Methanourarchaeum thermodiluricium” (ther.mo.tel.lu.r'i,ciu). Gr. n. therme heat; L. suff. -luricium ending to denote a family. Methanourarchaeum type genus of the family; L. suff. -aceae ending to denote a family; N.L. fem. pl. n. Hadesarchaeaceae the family of the genus Hadesarchaeum. Similarly, a new order is proposed, Hadesarchaeales (Ha.des.ar.ca.les) N.L. n. Hadesarchaeaceae type genus of the order; L. suff. -ales ending to denote an order; N.L. fem. pl. n. Hadesarchaeales the order of the genus Hadesarchaeum. A new class is proposed, Hadesarchaeota (Ha.des.ar.ca.ta) N.L. n. Hadesarchaeaceae type genus of the class; Hadesarchaeota the class of the genus Hadesarchaeum. A new phylum is proposed, Hadesarchaeota (Ha.des.ar.ca.o) N.L. n. Hadesarchaeaceae type genus of the phylum; L. suff. -ota ending to denote an phylum; N.L. fem. pl. n. Hadesarchaeota the phylum of the genus Hadesarchaeum.

Methanomethylarchaeum” (Me.tha.no.mix.to phil.pi) cu. M. L. n. methanum methane; M. L. adj. mixed, referring to H₂-dependent methanogenesis; N.L. neut. adj. trophicum metabolism; L. part. Methanomethylarchaeum, the methane-consuming archaeon. The type species is “Ca. Methanomethylarchaeum sicinus”. “Ca. Methanomethylarchaeum sicinus” (si.eni.sis). L. adj. originating from China. The type material is the metagenomic bin JZ-2 bin_199 (Ga0263247). Phylogenetic analyses support the proposal for a new family, Methanomethylarchaeaceae (Me.tha.no.me.thy.ar.chae.ac. ae) N.L. n. Methanomethylarchaeum type genus of the family; L. suff. -aceae ending to denote a family; N.L. fem. pl. n. Methanomethylarchaeaceae the family of the genus Methanomethylarchaeum.

Methanoproducendum” (Me.tha.pro.du.cen.dum) cu. M. L. n. methanum methane; M. L. adj. producendo, producing, referring to Methanoproducendum, the methane producer. The type species is “Ca. Methanoproducendum senex”. “Ca. Methanoproducendum senex” (senex). L. n. old man, referring to the deeply branching position of the organism. The type material is the metagenomic bin GMQP bin_32 (Ga0263258). Phylogenetic analyses support the proposal for a new family, Methanoproducendaceae (Me.tha.pro.du.cen.dae) N.L. n. Methanoproducendaceae type genus of the family; L. suff. -aceae ending to denote a family; N.L. fem. pl. n. Methanoproducendaceae the family of the genus Methanoproducendum.

Methylarchaeum” (Me.thyl.ar.chae.uch) cu. M. L. n. methanum methane; methyl, referring to methylation metabolism; Gr. adj. archeo ancient, also referring to Archaea. L. part. Methylarchaeum, the methane-consuming archaeon. The type species is “Ca. Methylarchaeum tengchongensis”. “Ca. Methylarchaeum tengchongensis” (teng.chong.en.siis) originating from Tengchong, a region of Yunnan Province, China. The type material is the metagenomic bin JZ-2 bin_220 (Ga0263250).

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
The 14 complete near-archaeal genomes are publicly available in the JGI IMG-MER under the Study ID Ga021767 and WGS accessions Ga0180368 (Unclassified Nezhaarchaeota JZ-38 bin), Ga0263245 (Unclassified Nezhaarchaeota JZ-1 bin_66), Ga0263246 (Unclassified Nezhaarchaeota GMQP bin_37), Ga0263254 (Unclassified Nezhaarchaeota GMQP bin_18), Ga0263254 (Unclassified Nezhaarchaeota DRTY-6 bin_144), Ga0263253 (Unclassified Nezhaarchaeota GMQP bin_44), Ga0263249 (Unclassified Nezhaarchaeota JZ-2 bin_200), Ga0263252 (Unclassified Nezhaarchaeota JZ-1 bin_66), Ga0263245 (Unclassified Nezhaarchaeota GMQP bin_37), Ga0263249 (Unclassified Nezhaarchaeota JZ-2 bin_200), Ga0263252 (Unclassified Nezhaarchaeota JZ-1 bin_66).
Verstraeten, A. et al. \( \text{JZ-3 bin}_3 \) (2018). Methyl-coenzyme M reductase genes of a deep-subsea iron-only methanotroph. Nature. Microbiol. \textbf{3}, 281 (2018).

Liu, Y. C. & Whitman, W. B. Metabolic, phylogenetic, and ecological diversity of the methanogenic archaea. \textit{Incred. Ann. N. Y. Acad. Sci.} \textbf{1125}, 171–189 (2008).

Evans, P. et al. Methane metabolism in the archaean phylum Bathyaarchaeota revealed by genome-centric metagenomics. \textit{Science} \textbf{350}, 434–438 (2015).

Luo, R. et al. Thermophilic archaea activate butane via alkyl-coenzyme M formation. Nature. \textit{Microbiol.} \textbf{3}, 1269–1279 (2019).

Vanderwerken, I. et al. Methylophilic methanogenesis discovered in the archaean phylum \textit{Verstraetenarchaeota}. \textit{Nature. Microbiol.} \textbf{1}, 16170 (2019).

Steinberg, L. M. & Regan, J. M. mcrA-targeted real-time quantitative PCR method to examine methanogen communities. \textit{Appl. Environ. Microbiol.} \textbf{79}, 4435–4442 (2009).

Berghuis, B. A. et al. A novel denitrifying methanotroph of the NC10 phylum and its role in global \( {\text{CH}}_3 {\text{COOH}} \) cycling. \textit{Nat. Commun.} \textbf{10}, 16002 (2016).

Evans, P. N. et al. An evolving view of methane metabolism in the Archaea. \textit{Nat. Rev. Microbiol.} \textbf{17}, 219–232 (2019).

Thauer, R. K. The Wolfe cycle comes full circle. \textit{Proc. Natl Acad. Sci. USA} \textbf{109}, 15084–15085 (2012).

Dridi, S., Fardeau, M. L., Ollivier, B., Raoult, D. & Drancourt, M. Methanomassiliicoccus luminyensis gen. nov., sp. nov., a methanogenic archaea isolated from human faeces. \textit{Int. J. Syst. Evol. Microbiol.} \textbf{62}, 1902–1907 (2012).

Castelle, C. et al. Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. \textit{Carr. Biol.} \textbf{25}, 690–701 (2015).

Kono, T. et al. A RubisCO-mediated carbon metabolic pathway in methanogenic archaea. \textit{Nat. Commun.} \textbf{8}, 14007 (2017).

Berg, I. et al. Autotrophic carbon fixation in archaea. \textit{Nat. Rev. Microbiol.} \textbf{8}, 447–460 (2010).

Khelifi, N. et al. Anaerobic oxidation of long-chain n-alkanes by the hyperthermophilic sulfate-reducing archaeon, \textit{Archaeoglobus fulgidus}. \textit{ISME J.} \textbf{8}, 2153–2166 (2014).

Borrel, G., Adam, P. & Gribaldo, S. Methanogenesis and the Wood-Ljungdahl pathway: an ancient, versatile, and fragile association. \textit{Genome Biol. Evol.} \textbf{8}, 1706–1716 (2016).

Hua, Z. S. et al. Genomic inference of the metabolism and evolution of the archaean phylum \textit{Aigarchaeota}. \textit{Nat. Commun.} \textbf{9}, 2832 (2018).

Parks, D. H. et al. Discovery of nearly 8000 metagenome-assembled genomes substantially expands the tree of life. \textit{Nat. Microbiol.} \textbf{2}, 1533–1542 (2018).

Adam, P., Borrel, G. & Gribaldo, S. Evolutionary history of carbon monoxide dehydrogenase/acetyl-CoA synthase, one of the oldest enzymatic complexes. \textit{Proc. Natl Acad. Sci. USA} \textbf{115}, E1166–E1173 (2018).

Weiss, M. C. et al. The physiology and habitat of the last universal common ancestor. \textit{Nat. Microbiol.} \textbf{1}, 16116 (2016).

Williams, T. A. et al. Integrative modeling of gene and genome evolution roots the archaean tree of life. \textit{Proc. Natl Acad. Sci. USA} \textbf{114}, E4602–E4611 (2017).

Yang, Z. \textit{PAML 4}: phylogenetic analysis by maximum likelihood. \textit{Mol. Biol. Evol.} \textbf{24}, 1586–1591 (2007).

Boussau, B., Blanchart, S., Neczula, A., Lartillot, N. & Gouy, M. Parallel adaptations to high temperatures in the Archean eon. \textit{Nature} \textbf{456}, 942–945 (2008).

Raymann, K., Bröchner-Armanet, C. & Gribaldo, S. The two-domain tree of life is linked to a new root for the Archaea. \textit{Proc. Natl Acad. Sci. USA} \textbf{112}, 6670–6675 (2015).

Koga, Y. Thermal adaptation of the archaean and bacterial lipid membranes. \textit{Archaea} \textbf{2012}, 789652 (2012).

Kampmann, M. & Stock, D. Reverse gyrase has heat-protective DNA chaperone activity independent of supercoiling. \textit{Nucleic Acids Res.} \textbf{32}, 3357–3365 (2004).

Djokic, T., Van Kranendonk, M. J., Campbell, K. A., Walter, M. R. & Ward, C. R. Earliest signs of life on land preserved in ca. 3.5 Ga hot spring deposits. \textit{Nat. Commun.} \textbf{8}, 15263 (2017).

Dodds, M. S. et al. Evidence for early life in Earth’s oldest hydrothermal vent precipitates. \textit{Nature} \textbf{543}, 60–64 (2017).

Martin, W. & Russel, M. J. On the origin of biochemistry at an alkaline hydrothermal vent. \textit{Philos. Trans. R. Soc. B} \textbf{362}, 1887–1926 (2007).

Ferry, J. G. & House, C. H. The stepwise evolution of early life driven by energy conservation. \textit{Mol. Biol. Evol.} \textbf{23}, 1286–1292 (2005).

Martin, W., Baross, J., Kelley, D. & Russel, M. J. Hydrothermal vents and the origins of life. \textit{Nat. Rev. Microbiol.} \textbf{6}, 805–814 (2008).

Nitschke, W. & Russel, M. J. Beating the alkyl coenzyme A-pathway to the origin of life. \textit{Philos. Trans. R. Soc. B} \textbf{368}, 20120258 (2013).
56. Schulte, M., Blake, D., Hoehler, T. & McCollom, T. H. O. M. A. S. Serpentinization and its implications for life on the early Earth and Mars. Astrobiology 6, 364–376 (2006).

57. Hua, Z. S. et al. Ecological roles of dominant and rare prokaryotes in acid mine drainage revealed by metagenomics and metatranscriptomics. ISME J. 9, 1280–1294 (2015).

58. Bankevich, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477 (2012).

59. Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ. 3, e1165 (2015).

60. Dick, G. J. et al. Community-wide analysis of microbial genome sequence signatures. Gerton J. Biol. Chem. 288, W159–W168 (2013).

61. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25, 1043–1055 (2015).

62. Boyd, J. A., Woodcroft, B. J. & Tyson, G. W. GraFFiM: a tool for scalable, phylogenetically informed classification of genes within metagenomes. Nucleic Acids Res. 46, 659–659 (2018).

63. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11, 119 (2010).

64. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12, 59–60 (2015).

65. Yin, Y. et al. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 40, W445–W451 (2012).

66. Wu, M. & Scott, A. J. Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. Bioinformatics 28, 1033–1034 (2012).

67. Edgar, R. C. Muscle: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1799 (2004).

68. Capella-Gutierrez, S., Silla-Martinez, J. M. & Gabaldon, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973 (2009).

69. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274 (2015).

70. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. X., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587 (2017).

71. Tria, F. D. K., Landan, G. & Dagan, T. Phylogenetic rooting using minimal ancestor deviation. Nat. Ecol. Evol. 1, 0193 (2017).

72. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–1973 (2012).

73. Letunic, I. & Bork, P. Interactive tree of life (ITOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 42, 242–245 (2016).

74. Sauer, D. B. & Wang, D. N. Prediction of optimal growth temperature using only genome derived features. Bioinformatics 35, 3224–3231 (2019).

75. Suhr, K. & Claverie, J. M. Genomic correlates of hyperthermostability, an update. J. Biol. Chem. 278, 17198–17202 (2003).

76. Enright, A. J., Van Dongen, S. & Ouzounis, C. A. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 30, 1575–1584 (2002).

77. Ronquist, F. et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542 (2012).

78. Csíkos, M. C.ount: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. Bioinformatics 26, 1910–1912 (2010).

Acknowledgements
We thank Guangdong Magigene Biotechnology Co., Ltd. China for the assistance in data analysis. This work was financially supported by Science and Technology Infrastructure work project of China Ministry of Science & Technology (No. 2015FY110100), National Natural Science Foundation of China (19151005, 31600298, U1201233 and 31370154), National Science Foundation of Guangdong Province, China (No. 2016A030312003), Guangdong Province Science and Technology Innovation Strategy Special Fund (No. 2018B020206001), Guangzhou Municipal People’s Livelihood Science and Technology Plan (No. 201803000300), China Postdoctoral Science Foundation (2016M602567), State Key Laboratory of Marine Pollution (SKLMP) Seed Collaborative Research Fund (2019), W.J.L was supported by Guangdong Province Higher Vocational Colleges & Schools Pearl River Scholar Funded Scheme (2014). P.N.E. is supported by an Australian Research Council Discovery Early Career Researcher Award (170110428). The authors are grateful to the Researchers Supporting Project number (RSP-2019/53), King Saud University, Riyadh, Saudi Arabia.

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

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41467-019-12574-y.

Correspondence and requests for materials should be addressed to G.W.T., T.Z. or W.-J.L.

Peer review information Nature Communications thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information are available at http://www.nature.com/reprints-

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.