A methylotrophic origin of methanogenesis and early divergence of anaerobic multicarbon alkane metabolism

Yinzhao Wang1*, Gunter Wegener2,3, Tom A. Williams4, Ruize Xie5, Jialin Hou1, Chen Tian5, Yu Zhang5, Fengping Wang1,5,6, Xiang Xiao1,7*

Methanogens are considered as one of the earliest life forms on Earth, and together with anaerobic methanoxidizing archaea, they have crucial effects on climate stability. However, the origin and evolution of anaerobic alkane metabolism in the domain Archaea remain controversial. Here, we present evidence that methylotrophic methanogenesis was the ancestral form of this metabolism. Carbon dioxide–reducing methanogenesis developed later through the evolution of tetrahydromethanopterin S-methyltransferase, which linked methanogenesis to the Wood-Ljungdahl pathway for energy conservation. Anaerobic multicarbon alkane metabolisms in Archaea also originated early, with genes coding for the activation of short-chain or even long-chain alkanes likely evolving from an ethane-metabolizing ancestor. These genes were likely horizontally transferred to multiple archaeal clades including Candidatus (Ca.) Bathymarcheia, Ca. Lokiarchaeia, Ca. Hadarchaeia, and the methanogenic Ca. Methanoliparia.

INTRODUCTION

Methanogenesis, one of the most ancient biochemical pathways on Earth, also plays a critical role in global climate change, as this process largely controls the formation of methane, a strong greenhouse gas. Methanogenesis is exclusively found in the domain Archaea, and changes in its activity may have caused severe fluctuations of Earth’s surface temperature and subsequent life mass extinction events, e.g., the mass extinction potentially caused by the methanogenic burst in the end-Permian (1). By contrast, microbial methanogenesis might have prevented early snowball Earth scenarios during the Hadean and Archean eons by producing methane to keep a warm atmosphere in the “faint young Sun” period of Earth history (2, 3). However, methane can be also consumed in anoxic habitats by anaerobic methanoxidizing archaea (ANME) using a reverse methanogenesis pathway (4, 5). It has been estimated that the anaerobic oxidation of methane removes ~80% of the methane produced by methanogenesis in the modern ocean, therefore keeping the atmospheric methane at a relatively low concentration, avoiding potential global warming effect (6).

Both geological evidence and molecular dating indicated that methanogenesis originated very early (2, 7), and it has been suggested that methanogens may represent one of the earliest life forms (8). The methyl–coenzyme M reductase (MCR) is the key enzyme of anaerobic methane metabolism, while the related alkyl–coenzyme M reductases (ACRs) catalyze the oxidation of multicarbon alkanes (4, 5, 9, 10). Most of the cultured methanogens reduce carbon dioxide with electron donors such as hydrogen. These organisms contain the enzymes of the Wood-Ljungdahl pathway to reduce carbon dioxide to tetrahydromethanopterin-bound methyl groups (11). Other methanogens thrive on the acetoclastic reaction or disproportionation of methylated substrates such as methanol to methane and carbon dioxide (11). ANME and the recently found anaerobic short-chain alkane-oxidizing archaea use the Wood-Ljungdahl pathway and MCR/ACR in an oxidative direction (9, 10, 12). The enzyme interconnecting the Wood-Ljungdahl pathway with MCR is the tetrahydromethanopterin S-methyltransferase (MTR) that transfers the methyl groups between the cofactors tetrahydromethanopterin and coenzyme M (11, 13). However, hydrogen-dependent methylotrophic archaea of the order Methanomassiliicoccales have been cultured, which lack both the Wood-Ljungdahl pathway and MTR and consequently require methylated compounds and hydrogen as electron acceptor and donor, respectively (14, 15). In recent years, environmental genomics has revealed many previously unknown lineages of potential methanogens across the archaeal species tree (16–22) [reviewed in (5)]. Phylogenomic and comparative genomic analyses of archaean diversity have supported the hypothesis that the last common ancestor of Euryarchaeota and TACK (an archaean superphylum including Thaumarcheota, Aigarchaeota, Crenarcheota, Korarchaeota, Geoarchaeota, and Bathyarchaeota, among other phyla) archaea might have been a methanogen (19–22). Some studies suggested that the first methanogens were carbon dioxide reducers using the Wood-Ljungdahl pathway (19). Others also pointed to the possibility that the ancestral methanogen was likely a hydrogen-dependent methylotroph (20). The finding of anaerobic multicarbon alkane-oxidizing archaea from different archaeal phyla suggested a more complex evolutionary history of ACR, potentially involving multiple horizontal gene transfers (HGTs) of ACR-encoding genes (20–22). Clearly elucidating the origin and evolution of methanogenesis and anaerobic alkane metabolisms requires better genomic sampling of early-diverging members of the Euryarchaeota and TACK archaea and mapping of mcr/acr and mtr gene family origins onto the archaeal species tree.
RESULTS AND DISCUSSION
Here, we studied the origin and diversification of methanogens and anaerobic multicarbon alkane-oxidizing archaea. To do so, we retrieved MCR/ACR-containing metagenome-assembled genomes (MAGs) from eight metagenome datasets collected from Tengchong hot spring in China and Captain Aryutinov mud volcano, two public metagenome datasets, and the Genomes from Earth’s Microbiomes (GEM) catalog (23). The data analyzed here [Table 1 and data file S1 (table S1)] comprise 24 MAGs binned from our own sequencing data, two MAGs binned from public metagenomic data, and 30 published MAGs from the GEM project (23). All of the data originally generated by other laboratories (including MAGs from GEM) were used with their permission, and we thank these colleagues for their generosity and willingness to share their data. The provenance of all data used is indicated in table S1 as well as the Acknowledgements for further details. The MAGs assigned to the class I (Methanobacteriales, Methanococcales, and Methanopyrales) and class II (Methanosarcinales, Methanomicrobiales, and Methanocellales) methanogens are well represented by genomes from cultured strains and, hence, were not included in subsequent analyses. A total of 56 MAGs containing MCR/ACR-encoding organisms mostly with completeness >80% (39 MAGs) and contamination below 5% (49 MAGs) were retained for further analyses [Table 1 and data file S1 (table S1)]. In the following discussion, we use the Genome Taxonomy Database (GTDB) to label and refer to MAGs (24), but in the interest of clarity, we retain the conventional names, i.e., Euryarchaeota, TACK, Asgard, and DPANN (an archaeal superphylum currently including Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota, Nanoarchaeota, and several other phyla), for higher taxonomic description. These MAGs and reference genomes of the domain Archaea from the National Center for Biotechnology Information (NCBI) prokaryotes database were used to construct a phylogenomic tree on the basis of a concatenated set of 37 marker genes [fig. S1 and data file S1 (tables S2 to S3)] (25). All MAGs with MCR/ACR affiliated with the Euryarchaeota, TACK, and Asgard archaea. Despite intensive search for additional MCR/ACR sequences on both NCBI prokaryotes and metagenomic databases (July 2019), no MCR/ACR-encoding gene was found in genomes assigned to the DPANN archaea or the domain Bacteria, indicating that our analysis covered the diversity of MCR/ACR-encoding archaeal lineages.

Vertical evolution of MCR-encoding genes in Euryarchaeota and TACK
This and other recent metagenome-based studies (5, 16–22, 26) remarkably expanded the known distribution of MCR/ACR-encoding archaea.

### Table 1. MAGs described in this study and their main potential metabolic features.

<table>
<thead>
<tr>
<th>Superphylum</th>
<th>Phylum</th>
<th>Class</th>
<th>Order or family</th>
<th>MAGs</th>
<th>Metabolic potentials</th>
<th>Database source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Euryarchaeota</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methanobacterota</td>
<td>Thermococi</td>
<td>Ca. Nuwarchaeales</td>
<td></td>
<td>6</td>
<td>CH$_3$-reducing methanogen</td>
<td>Present study, PRJNA3573577 and Nayfach et al. (23)</td>
</tr>
<tr>
<td>Thermoplasmatota</td>
<td>Thermoplasmata</td>
<td>Methanomassilicoccales</td>
<td></td>
<td>1</td>
<td>CH$_3$-reducing methanogen or C$_2$H$_6$ metabolism</td>
<td>PRJNA4453587 and Reiss et al. (75)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ca. Methanoliparia</td>
<td>Ca. Methanoliparales</td>
<td>1</td>
<td>CO$_2$-reducing methanogen with C$_2$H$_6$ metabolism</td>
<td>Nayfach et al. (23)</td>
</tr>
<tr>
<td></td>
<td>Archaeglobi</td>
<td>Archaeglobales</td>
<td></td>
<td>4</td>
<td>CO$_2$-reducing methanogen or CH$_4$ oxidation</td>
<td>Present study &amp; Nayfach et al. (23)</td>
</tr>
<tr>
<td></td>
<td>Halobacterota</td>
<td>Ca. Polytropaceales</td>
<td></td>
<td>1</td>
<td>C$<em>2$H$</em>{2n+2}$ oxidation</td>
<td>Nayfach et al. (23)</td>
</tr>
<tr>
<td></td>
<td>Methanosarcina</td>
<td>Ca. Ethanoperedenaceae</td>
<td></td>
<td>1</td>
<td>C$<em>2$H$</em>{4}$ oxidation</td>
<td>Nayfach et al. (23)</td>
</tr>
<tr>
<td></td>
<td>Ca. Syntropharchaeia</td>
<td>Ca. Santabarbaraeales</td>
<td></td>
<td>1</td>
<td>C$<em>2$H$</em>{2n+2}$ oxidation (n &gt; 2)</td>
<td>Hawley et al. (31, 32) and Nayfach et al. (23)</td>
</tr>
<tr>
<td></td>
<td>Ca. Holophaga</td>
<td>Ca. Alkanophaga</td>
<td></td>
<td>1</td>
<td></td>
<td>Dombrowski et al. (30)</td>
</tr>
<tr>
<td></td>
<td>Ca. Hadarchaeota</td>
<td>Ca. Hadarchaeas</td>
<td></td>
<td>3</td>
<td>C$<em>2$H$</em>{2n+2}$ oxidation</td>
<td>Nayfach et al. (23)</td>
</tr>
<tr>
<td></td>
<td>Ca. Nezharchaeia</td>
<td>Ca. Nezharchaeas</td>
<td></td>
<td>5</td>
<td>CO$_2$-reducing methanogen</td>
<td>Present study, Nayfach et al. (23), and Dombrowski et al. (30)</td>
</tr>
<tr>
<td></td>
<td>TACK</td>
<td>Thermoproteota</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ca. Methanomethylia</td>
<td>Ca. Methanomethylicales</td>
<td></td>
<td>30</td>
<td>CH$_3$-reducing methanogen or C$_2$H$_6$ metabolism</td>
<td>Present study and Nayfach et al. (23)</td>
</tr>
<tr>
<td></td>
<td>Ca. Korarchaeia</td>
<td>Ca. Korarchaeas</td>
<td></td>
<td>1</td>
<td>CH$_3$-reducing methanogen</td>
<td>Nayfach et al. (23)</td>
</tr>
<tr>
<td><strong>Asgard</strong></td>
<td>Asgardarchaeota</td>
<td>Ca. Lokiarchaeia</td>
<td>Ca. Helarchaeales</td>
<td>4</td>
<td>C$<em>2$H$</em>{2n+2}$ oxidation</td>
<td>Present study</td>
</tr>
</tbody>
</table>
genes in the domain Archaea. In most of these archaea, the MCR alpha subunit (McrA)—encoding genes are well conserved and can be used as a phylogenetic marker, in the sense that their evolution follows the species tree (27). Consistent with previous reports (15), topologies of trees for McrA sequences and species trees in the current study are highly congruent for the class I and II methanogens in Eurysyarchaeota (Fig. 1, A and B, and fig. S1). McrA-encoding genes are found less frequently in TACK archaea, but the McrA-encoding genes that have been identified [in Nitrososphaeria, Candidatus (Ca.) Methanomethylcyxia, Ca. Korarchaeia, and Ca. Nezhaarchaeia] form a monophyletic group whose internal branching order is congruent with the inferred TACK species tree (Fig. 1, A and B, and figs. S1 and S2), potentially consistent with vertical inheritance of McrA from their common ancestor. However, the branching position of Ca. Korarchaeia varied when using different combinations of conserved genes and phylogenetic algorithm, placing Ca. Korarchaeia either close to Ca. Bathysarchaeia and Nitrososphaeria or to Thermoprotea and Ca. Methanomethylxyxia (20–22, 26). Nevertheless, with substitution model (LG+C60+G) of a better Bayesian information criterion (BIC) score, the phylogenomic tree of anaerobic alkane-metabolizing archaea with 37 concatenated conserved protein sequences shows identical tree topology with their McrA phylogenetic tree (Fig. 1, A and B, and fig. S2).

In addition, we found that 13 MAGs assembled here (Table 1) and in previous studies (18, 20–22, 26) branch near the base of Euryarchaeota and TACK in both the MCR and species trees, helping provide insight into the early evolution of methanogenesis (Fig. 1, A and B, and fig. S1). Two early diverging lineages from Eurysyarchaeota branch between the TACK and Eurysyarchaeota class I methanogens: Ca. Methanofastidiosales and one group branching next to Ca. Methanofastidiosales (including a previously reported MAG, NM3) (20) belonging to a new order-level lineage, which we here propose as Ca. Nuwarchaeales (Nuwa, the mother of earth in the Chinese myth; fig. S3), a potential hydrogen-dependent methylotrophic methanogen without the Wood-Ljungdahl pathway as also described in (20). Ca. Methanofastidiosales and Ca. Nuwarchaeales branch together with Thermococci and Ca. Theionarchaea and are close to the divergence of Eurysyarchaeota and TACK on the species tree, which is known as a superclass Ca. Acherontia and has been reported previously (28) (fig. S1), such that comparisons of their gene content with other Eurysyarchaeota may elucidate early metabolic transitions within the clade. Their MCR sequences also form a monophyletic cluster on the MCR/ACR phylogenetic tree, and therefore we here classified these MAGs as the class III methanogens (Fig. 1, A and B). Within the class I and II methanogens, the previously reported Ca. Methanopiliparia (20, 29) branches between these two groups of methanogens. This position is supported by both MCR and genome trees, and the branching position of Ca. Methanopiliparia in both the MCR and species trees strengthens the existing case (19–22) for vertical evolution of MCR-based metabolism within the class I and II methanogen clades, with lineage-specific losses giving rise to the metabolisms of methanogenesis and carbon fixation (Fig. 1, A and B, and fig. S1). Together, the results here suggest that the MCR-encoding genes of the TACK and Eurysyarchaeota methanogens have descended vertically from the common ancestor of the two superphyla, as suggested previously (19–22).

Although most methanogen lineages showed vertical evolution, we identified and confirmed some incongruences with the species tree, in accordance with a previous report (20) (Fig. 1, A and B, and fig. S1) (20). Within Eurysyarchaeota, McrA-encoding genes belonging to the class I/II (carbon dioxide–reducing) and class III (methyl-reducing) methanogens form reciprocally monophyletic lineages with high bootstrap support (100 and 93%, respectively); however, some methyl-reducing methanogens such as Methanomassiliicoccales and Methanomassiliicoccales have the class III MCR-encoding genes but branch within the class I/II methanogens on species trees (fig. S1). This pattern might reflect the sorting out of a putatively duplicated ancestral euryarchaeotal MCR-encoding gene through parallel gene losses or, alternatively, gene transfer of a single ancestral MCR-encoding gene during the diversification of Eurysyarchaeota (Fig. 1, A and B, and figs. S1 and S4). In the first scenario, ancestral Eurysyarchaeota should have both the class I and III MCR but gradually lost one during their evolution, and only one type of MCR-encoding gene was kept within their genomes. In the latter scenario, the MCR-encoding genes of the Methanomassiliicoccales/Methanomassiliicoccales clades would have been acquired from an ancient class III methanogen ancestral to Ca. Methanofastidiosales and Ca. Nuwarchaeales, and the deeper MCR lineage branching outside Methanomassiliicoccales on this cluster is still somewhere that was not sampled in this study. The phylogeny indicates that Ca. Methanophagales (previously called ANME-1), a lineage of methane oxidizers, may also have acquired MCR-encoding genes from a relative of Ca. Nuwarchaeales by HGT, as suggested by Borrel et al. (20). Ca. Methanophagales also contains genes for beta-oxidation, which would not be required for methane metabolism; however, this organism emerges from within a clade of archaea capable of multivcarbon alkane metabolism and may have retained these beta-oxidation genes by vertical inheritance (fig. S3). All the neighboring branches of Ca. Methanophagales, including Ca. Syntrophoarchaeales, the Ca. Alkanophagales proposed here (previously called ANME-1 B39_G2) (30), and Ca. Santabarbaracales (from Santa Barbara Coal Oil Point) (31, 32), contain ACR and beta-oxidation enzyme-encoding genes for anaerobic multicarbon alkane metabolisms (fig. S3). Therefore, the Ca. Methanophagales ancestor, such as those archaea described above, may originally metabolize multicarbon alkanes rather than methane.

The ancestral methanogen likely reduced methylated compounds

The congruence between the topologies of the McrA phylogenetic tree and the methanogen species tree suggests that the common ancestor of Eurysyarchaeota and TACK archaea already contained methanogenic MCR-encoding genes (Fig. 1, A and B). Previous studies pointed out the possibility that both Eurysyarchaeota and TACK archaea combine the MCR and Wood-Ljungdahl pathways via the MTR complex and, therefore, that the ancestral methanogen might have been able to reduce carbon dioxide to methane (19). However, on the basis of our current knowledge, the combination of MCR, MTR, and the Wood-Ljungdahl pathway is restricted to the class I and II methanogens in Eurysyarchaeota, as well as Ca. Nezhaarchaeia in TACK. The other early-diverging lineages of methanogens within Eurysyarchaeota and TACK lack the MTR complex (Fig. 1, A to C, and fig. S3). To resolve the evolutionary history of MTR-encoding genes, we inferred phylogenetic trees of MTR subunits (Fig. 1C). For the Eurysyarchaeota class I and II methanogens, the MTR-encoding genes have very similar topologies as the corresponding phylogenomic trees. However, MTR-encoding genes from Ca. Nezhaarchaeia clusters with those of Ca. Methanophagales and
Fig. 1. Phylogenetic analyses of enzymes in archaeal methane metabolism. (A) The McrA/AcrA phylogenetic tree is constructed on the basis of the alignments of 259 McrA/AcrA sequences with 472 aligned positions. Only the McrA branches are showed here. (B) Phylogenomic affiliation of 177 MAGs is based on 37 conserved protein sequences using representative methane metabolism archaea. Background colors: TACK, red shaded; class I methanogen, blue shaded; class II methanogen, green shaded; class III methanogen, pink shaded. (C) MtrA phylogenetic tree shows the classification of these sequences from different archaeal lineages. The phylogenetic tree is constructed on the basis of the alignments of 152 MtrA sequences with 148 aligned positions. Other phylogenetic trees are presented in fig. S5, which are in agreement with Liu et al. (73). The MtrA and H-like subunits are possibly either HGT from methanogens or originated before class I methanogens, as also described in Wang et al. (74). (D) One type of methyltransferase (MtaA, alpha subunit of methanol-corrinoid protein:coenzyme M methyltransferase) phylogenetic tree (134 sequences with 309 aligned positions) shows that methyltransferases are likely vertically transferred and might originate before the divergence of TACK and Euryarchaeota. Some class I methanogens that contain MtaA might be HGT from the TACK methanogens. Other methyltransferase phylogenetic trees are displayed in fig. S7. All conserved protein phylogenomic and McrA, MtrA, and MtaA phylogenetic alignments are based on MAFFT and then filtered with trimAl, and the trees that were built by the IQ-Tree method with model LG+C60+F+G using SH approximate likelihood ratio test implemented with 1000 bootstrap replicates with bootstrap higher than 0.8 are shown with gray squares on tree branches. (E) Evolutionary history of MCR-based methane metabolism.

Ca. Methanoliparia and an approximately unbiased (AU) test (33) for the catalytic subunit MtrA rejected the monophyly of Ca. Nezhaarchaeia MtrA cluster with other TACK archaea. This suggests a horizontal transfer of MTR-encoding genes from an ancestral lineage of Ca. Methanophagales/Ca. Methanoliparia to Ca. Nezhaarchaeia and, consequently, an origin of the MTR well within Euryarchaeota, likely in the class I methanogens (see MTR phylogenetic tree; Fig. 1C and fig. S5). Some archaea from TACK and Asgard contain MtrA or H-like subunits; however, these appear to be only distantly related to the enzymes of the MTR complex in methanogens and thus may have alternative functions (16, 22). To evaluate the origination points of the McrA- and MtrA-encoding gene families statistically, we used probabilistic gene tree–species tree reconciliation [the amalgamated likelihood estimation (ALE) method (34)] to estimate the probability that these gene families were present at each internal node of their rooted species tree. ALE reconciliations account for gene duplication, transfer, and loss by estimating the rates of these events for the data, and analysis of empirical data and simulations suggest that ALE is an accurate method for rooting gene trees (35), particularly in the presence of HGT (36). The results suggested that MtrA evolved later than McrA during archaeal evolution. The nodes in the species tree to which McrA could be mapped with high confidence [presence probability (PP) > 0.95] are the common ancestor of class I and II methanogens, Ca. Methanofastidiosales/Ca. Nwuarchaeales in Euryarchaeota, as well as Ca. Methanomethylcica, Ca. Korarchaeia, and Ca. Nezhaarchaeia in TACK (fig. S6). By contrast, the earliest node at which MtrA is present with high confidence (PP = 0.94 and PP > 0.95 on descendant nodes) is only within the class I and II methanogens following the divergence of Archaeoglobales (fig. S6). This is consistent with the hypothesis that MtrA evolved later than McrA in archaeal evolution. This late origin of MTR is consistent with a methylotrophic origin of methanogenesis and the subsequent evolution of carbon dioxide–reducing methanogenesis within Euryarchaeota. By contrast, in addition to MCR, both the Euryarchaeota and TACK methanogens contain a variety of methyltransferase genes such as methanol-corrinoid protein:coenzyme M methyltransferase (Mta), methylamine-corrinoid protein:coenzyme M methyltransferase (Mtb), and methylated-thiol-corrinoid protein:coenzyme M methyltransferase (Mts). These methyltransferases enable methanogens to use methyl compounds such as methanol, methylamine, and methanethiol. The phylogenetic analyses of these methyltransferases also indicate that they were likely vertically inherited and that they might be present in the common ancestor of Euryarchaeota and TACK archaea (Fig. 1D and fig. S7).

Hydrogen-dependent methyl-group reduction as the original metabolic mode of methanogens is consistent with the conditions of the early Earth. Our planet was anoxic but likely rich in methylated compounds and hydrogen, as well as other simple organic compounds (37, 38). These environments should have been suitable for hydrogen-dependent methyl-reducing methanogenesis. This finding does not contradict the hypothesis that both MCR and the Wood-Ljungdahl pathway were early metabolic traits of archaea (4, 8, 15, 28, 36, 39) because methane production via the methylotrophic methanogenesis pathway and carbon dioxide fixation through Wood-Ljungdahl pathway can work separately. Ancestral class I methanogens likely developed the MTR complex shortly after the divergence of the TACK and Euryarchaeota methanogens. The MTR connected the MCR with the Wood-Ljungdahl pathway and enabled energy generation by carbon dioxide reduction to methane. After that, some class II methanogens acquired the capability for cytochrome c synthesis that allows for more efficient methanogenic growth (11, 40). The class II methanogens are now the most successful on the modern Earth environments; members of this group thrive on carbon dioxide reduction, disproportion of methylated substrates, and acetate (11). With the accumulation of electron acceptors such as sulfate on Earth in the late Archean (41), several methanogenic lineages started to reverse the methanogenesis pathway and turned into anaerobic methane oxidizer coupling with sulfate-reducing bacteria or nitrate or metal oxides (4, 5). All known members of this group have cytochromes and are suggested to have crucial roles in direct interspecies electron transfer (42, 43).

**Multicarbon alkane metabolism in Ca. Methanomethylcica and Methanomassiliicoccales**

Experimentally characterized ACRs catalyze the anaerobic oxidation of multicarbon alkanes including ethane, propane, and n-butane, similar to the biochemical process in anaerobic methane oxidation, i.e., activating alkanes to alkyl–coenzyme M (9, 10). ACR-encoding genes are found in at least seven euryarchaeotal lineages (5), as well as in Ca. Batharchaeia from the TACK superphylum (16) and Ca. Lokiarchaeia from the Asgard superphylum (44). Among the ACR-containing MAGs assembled in the present study (Table 1), we identified five Ca. Methanomethylcica MAGs from five independent hot spring sediment samples that code both an MCR and an ACR (Fig. 2, A to C, and figs. S4 and S8). The contigs coding for these genes appears to be endogenous to Ca. Methanomethylcica because these genes show the highest amino acid identities (~87.2% on average) to sequences of other Ca. Methanomethylcica MAGs (fig. S9, A and B). One mcr gene clusters with those of other Ca. Methanomethylcica, and this gene is arranged with other genes of the methanogenesis pathway (fig. S9C). On the MCR/ACR phylogenetic tree, the ACR sequences of Ca. Methanomethylcica cluster closely but as a basal lineage with the ACR from Ca. Argoarchaeum (10) and Ca. Ethanolopederens (45). These organisms activate ethane as ethyl–coenzyme M; hence, we propose to assign these ACR sequences to the subgroup of ethyl–coenzyme M reductases (ECRs). The Ca. Methanomethylcica ECR-encoding genes are surrounded by genes coding for cobalamin-binding protein and carboxylase (fig. S9C), enzymes that might help metabolize reactants or products of ethane metabolism (Fig. 2C). In addition to the ECR-containing Ca. Methanomethylcica, we assembled one MAG from Methanomassiliicoccales in Euryarchaeota from peatland samples that also contain both MCR- and ECR-encoding genes as well as bacterial-type Wood-Ljungdahl pathway with tetrahydrofolate as C1 carrier (Fig. 2, A, B, and D). Both MCR/ECR-containing Ca. Methanomethylcica and Methanomassiliicoccales have several methyltransferases such as Mta, Mtb, and Mts that enable activation of methylated compounds such as methanol, methylamines, and methyl sulfides (Fig. 2, C and D, and fig. S3). Both organisms also contain genes coding for heterodisulfide reductase/F420-nonreducing hydrogenase complex and EcH hydrogenase (ECH) for heterodisulfide (CoB-SS-CoM) cycling and energy conservation. However, in contrast to Ca. Argoarchaeum, Ca. Ethanolopederens, and Methanomassiliicoccales, the ECR-containing Ca. Methanomethylcica lack the gene coding for the Wood-Ljungdahl pathway (the completeness of one MAG approaches 100% with 0.93% contamination); it therefore lacks an obvious pathway to oxidize coenzyme M–bound ethyl groups to...
carbon dioxide. In contrast, Ca. Methanomethylicia may rather reduce C2 compounds such as acetate or ethanol to ethane using hydrogen as an electron donor (Fig. 2, C and D). Biogenic ethanogenesis is thermodynamically feasible even at low hydrogen and acetate or ethanol concentrations, and this process has been postulated to occur in the subsurface and may account for ethanogenesis in anoxic sediments (46). Ethanogenic organisms are not yet cultured, but on the basis of their coded pathways, ECR-containing Ca. Methanomethylicia might perform such reactions; however, at this stage, we cannot rule out the possibility of oxidation of ethane or propane, etc., using unknown pathways.

**Origin of anaerobic multicarbon alkane metabolism in Archaea**

Genes coding for putative anaerobic multicarbon alkane metabolism are widespread within three archaeal superphyla (5). MCR- and
ACR-encoding genes are part of the same broader gene family. On the basis of the congruence of the MCR tree with their species tree (Fig. 1, A and B), MCR- and ACR-encoding genes likely originated from an ancient gene divergence that took place before the radiation of the Euryarchaeota and TACK archaea. However, while the MCR subfamily has evolved largely vertically during archaeal evolution, ACR-encoding genes appear to have been subjected to frequent horizontal transfer (Fig. 2, A and B). Reconciliation of the MCR and ACR subtrees of the MCR/ACR gene family independently against the archaeal species tree suggested that the MCR subfamily is the older of the two, appearing coincident with the family as a whole in the common ancestor of class I and II methanogens at the latest (PP = 0.95). By contrast, other ACR genes can only be mapped confidently (PP > 0.95) to relatively shallow nodes within the Eu-
yarchaeota and TACK archaea. Together, these analyses are consist-
ent with a root for the ACR/MCR tree on, or within, the MCR subfamily (Fig. S6). The exclusion of the root from the ACR clade indicates that the ACR sequences that bind longer-chain alkanes evolved from within a paraphyletic grade of ECR sequences, includ-
ing those newly discovered from Ca. Methanomethylicia and Meth-
animassiliicoccales. This suggests that the enzymes that potentially metabo-
lize longer-chain alkanes, such as the ACRs in Ca. Iar-
chaeia in Asgard, Ca. Bathyarchaeia in TACK, and Ca. Hadarchaeia and Ca. Methanoliparia, evolved or transferred from those involved in
potential ethane metabolism. By extension, it is tempting to spec-
ulate that the genes of potential ethane metabolism might have evolved from within the MCR clade (i.e., that the root lies on either the TACK MCR or Euryarchaeota MCR stem branches, two of the
three root positions that are not rejected by the ALE analysis). Al-
ternatively, they might have evolved from an ancestral gene coding
for methyl-binding enzyme at the root of the MCR/ACR family.
However, since there are still not enough ECR/ACR-containing MAGs
that have been sampled up to now, our calculation only showed
a strong statistic earliest node within the Euryarchaeota, perhaps also
as a result of high rates of gene transfers and losses for this family
(Fig. 3, A and B). Comparison of sequence identity among the
MCR-, ECR-, and ACR-encoding genes (Fig. S10) indicates that
ECR-encoding genes show greater sequence identities to MCR-
encoding genes generally and, at least for EcrG, TACK MCR-
encoding genes specifically, than to other ACR-encoding genes.
Speculatively, this higher sequence identity might indicate a specific
evolutionary relationship between ECR-encoding genes and TACK
MCR-encoding genes, as might be expected if TACK MCR and all
ACRs form a clade on the rooted gene tree. However, this pattern
might also reflect lower rates of sequence evolution in TACK MCR-
and ECR-encoding genes compared to other ACR family members.

A root for the gene tree on the ACR stem would imply that in
addition to MCR, at least one ACR-encoding gene was present at
one time in a lineage ancestral to extant TACK and Asgard archaea,
either in the common ancestor of TACK and Euryarchaeota. However,
this ACR clade appears to be absent from the described genomes of
extant TACK, with the possible exception of the new sequences from
Ca. Methanomethylicia. Trees of Mcr/Ecr/AcrA and G sug-
gest that the new Ca. Methanomethylicia and Methanimassiliicoccales
sequences branch either within the ECR clade or between MCRs
and all ACRs. Thus, one possibility is that these sequences represent
a vestige of this otherwise extant, or unsampled, ACR lineage
(Fig. 2, A and B). Exploration of MAGs from other extreme envi-
ronments such as hot springs or hydrothermal vents will help determine
whether other TACK archaea encode relatives of this deep-branching
ACR lineage. Alternatively, the Ca. Methanomethylicia sequences
might have been acquired by gene transfer from an unsampled
deep-branching euryarchaeotal lineage.

The second scenario focuses on metabolic features of the ACR-
containing archaea. All those deep-branching organisms with ACR
such as Ca. Bathyarchaeia, Ca. Lokiararchaeia, and Ca. Hadarchaeia
encode the Wood-Ljungdahl and beta-oxidation pathways and, hence,
could potentially oxidize long-chain alkanes. Notably, they
do not contain cytochromes and hence likely interact with partner
organisms via interspecies hydrogen or formate transfer (47).
Considering the low sulfate levels in the Archean ocean (41), at that
time, the partners of the alkane oxidizers might have been rather metha-
nogens than sulfate reducers, or they might use a methanogenic al-
kan degradation pathway, which requires a chain length of at least six carbons (hexane) (48). In contrast, the more modern lineages of
Ca. Syntrophoarchaeia have cytochromes that allow direct electron trans-
fer with cytochrome-containing partners such as Ca. Desulfofer-
dus (49). The phylogenetic positions of these ACR sequences, from
one Archaeoglobi lineage and members in the Euryarchaeota class
Ca. Syntrophoarchaeia, match with their species tree (Fig. S11). This
suggests that these ACR-encoding genes might be vertically inherited
from their last common ancestor.

**Dating the origin and subsequent evolution of anaerobic
alkane metabolisms**

The acquisition of SMC (structural maintenance of chromosomes)
protein-encoding genes by the cyanobacterial ancestor from an ar-
cheal lineage in Euryarchaeota, most likely within Methanomicro-
bia or Halobacteria, has been recently used to date the origin of
euryarchaeotal methanogens before 3.51 billion years (Ga) ago (2).
We here combined the temporal constraint implied by this HGT
event to estimate the time at which methanogenesis first evolved and
the divergence times of carbon dioxide–reducing methanogens. Be-
cause the euryarchaeotal SMC donor lineage must be older than the
common ancestor of Cyanobacteria, we can import an absolute time
constraint from Cyanobacteria [which have fossil records (50, 51)] to
date events within Archaea, for which no unambiguous fossil re-
cord is available. The SMC gene partition provides a fossil calibra-
tion for the archaeal tree at the point at which the cyanobacterial clade
branches inside Euryarchaeota (fig. S12). The SMC sequences of
Cyanobacteria cluster with Halobacteria and Methanomicrobia in
Euryarchaeota, and the topology within cyanobacterial clade for the
SMC phylogenetic tree shows high congruence with the phyloge-
nomic tree of nonphotosynthetic and photosynthetic Cyanobacteria,
indicating that HGT of SMC-encoding genes occurred even long before
the origin of photosynthetic Cyanobacteria lineages
(fig. S12). On the basis of the timing constraints of Cyanobacteria
fossils (50, 51) and the predicated divergence of archaea and bacte-
ia (52), we predict that the common ancestor of methanogenic ar-
chea originated at ~3.8 to 4.1 Ga ago, at the border between
the Hadean and the Archean eon (Fig. 3, A and B, and figs. S12 and
S13). The age of the common ancestor of class I and II methano-
genics, that is, a conservative estimate for the origin of MCR genes
based on the earliest node to which MCR could be mapped with
high confidence in the ALE analysis, was 3.66 ± 0.20 Ga. The poste-
rior age estimates for the key nodes in the analysis of the tree built
by the conserved ribosome protein plus SMC sequence alignment
from Archaea and Cyanobacteria were similar even in an analysis
that did not include the sequence alignment, e.g., age of the most recent common ancestor of Euryarchaeota, TACK, and Asgard at 3.89 ± 0.21 Ga (fig. S13), suggesting that the tree topology and calibrations are providing much of the information for this analysis, as has been noted previously for estimates of ancestral methanogen age (2, 53, 54). This is earlier than the recent prediction for the origin of methanogenesis (55), and the authors also stated that the discovery of non-euryarchaeal methanogens would result in an even earlier origin of methanogenesis. Our analyses indicate that the first methanogen was a hydrogen-dependent methylotroph (Fig. 3, A to C), a possibility that has been raised previously (20). It is not clear whether methanogenesis is at the root of all archaea, as, so far, MCR-encoding genes have not been found in the DPANN superphylum. This might be because the DPANN genomes have undergone reductive evolution (56), as it remains difficult to determine whether the absence of MCR-encoding genes is ancestral to the DPANN clade or has resulted from gene loss. If so, the archaean ancestor would be capable of anaerobic alkane metabolism. Following
the divergence of Euryarchaeota and TACK, some Euryarchaeota
developed the MTR complex that provided a link between the MCR
and the Wood-Ljungdahl pathway. The latter pathway might have
first been developed or acquired to allow an autotrophic lifestyle.
With this, the class I carbon dioxide–reducing methanogens emerged,
possibly before ∼3.66 Ga ago (Fig. 3, A to C), supporting the geolog-
eical evidence that methane produced via carbon dioxide reduction
as characterized by highly depleted isotope signatures was dated
back to 3.46 Ga ago (7). In TACK, the ancestral methyl-reducing
methanogenesis pathway persisted, although some members such as
Ca. Nezhaarchaeia might have acquired the MTR complex from
Euryarchaeota (Fig. 1C).
Together, our results indicate that methanogenesis developed
soon after the divergence of Bacteria and Archaea, possibly in the
late Hadean period. The first methanogen was likely a hydrogen-depen-
dent methylotrophic archaeon. Life likely originated at hydro-
thermal vents or serpentization sites that provided ideal conditions
for such methanogens (fig. S14) (8, 57, 58). The temperatures were
elevated compared to the otherwise possibly cold planet owing to
the low illumination from the young Sun, and large amounts of mo-
olecular hydrogen and simple organic compounds such as methanol
and acetate would provide separated carbon and energy sources (59).
Meanwhile, microorganisms may have developed the Wood-Ljungdahl
pathway either for carbon fixation or energy conservation from ac-
etate metabolism. The later evolution of MTR could have allowed
methanogenesis from carbon dioxide reduction, which would
have strongly increased methane production on Earth. In the
anoxic Archean atmosphere, methane remained ∼1000 times
longer than today (60), and hence, methanogenesis may have had a
crucial impact on the climate of the early Earth. The accumulating
methane would enhance an early greenhouse effect and retain the
radiation from the young Sun, which would increase the surface
temperature on Earth, providing suitable habitats for other life
to evolve.

**MATERIALS AND METHODS**

**Raw sequencing reads treatment and genome binning**

Six sediment samples from Tengchong hot spring in China and two
methane-enriched samples from Captain Aryutinov mud volcano
were collected, and DNA were extracted and sequenced on Illumina
NovaSeq platform (Novogene Co.). Two metagenomic datasets
from NCBI Sequence Read Archive (SRA) were downloaded [data
file S1 (table S1)], and GEM catalog was downloaded from the website
(https://genome.jgi.doe.gov/portal/GEMS/GEMS.home.html) (23).
The raw sequencing reads were trimmed using the Sickle algorithm
version 1.33 (https://github.com/najoshi/sickle). The trimmed reads
were assembled using MEGAHIT version 1.0.6-hotfix1 (61) with
step 4 and/or SPAdes-3.13.1 (62) with k-step 4. We then also download-
loaded the McrA/AcrA sequences from the NCBI protein nr data-
base and built a local McrA/AcrA database by DIAMOND version
0.8.28.90 (http://github.com/bbuchfink/diamond). A DIAMOND
search of potential McrA/AcrA sequences was carried out for
the assembled metagenomic datasets, and McrA/AcrA sequences with
best hits to the recently published alkane-metabolizing McrA/AcrA
sequences or McrA/AcrA with low identities or without the known
sequences were selected [data file S1 (table S1)]. Then, each assemb-
led contig coverage was determined by mapping the trimmed reads
back to the contigs using Bowtie version 2.2.8 (63) with parameter --
very-sensitive. The assembled metagenomic sequences were binned
on the basis of MaxBin version 2.2.4 (64) with the run_MaxBin.pl
script and on the basis of abundance and tetranucleotide frequency
using MetaBAT version 2.12.1 (65) with 1 kb (or 1.5 kb) and 3 kb
as contig length cutoffs. The 4-mer and 5-mer frequencies of each
contig were calculated, and dimensionality was reduced by t-sne
clustering algorithm (https://github.com/lejon/T-SNE-Java/tree/master/tsne-core) (66). The selected MAGs were then checked with
the mmgenome package version 0.6.3. Final completeness
and contamination of each MAG [data file S1 (table S1)] were as-
essed with CheckM version 1.0.7 (67) using lineage-specific (149 to
228) marker genes, and the MAGs with higher completeness and
lower contamination were selected from each dataset for further
analyses. Taxonomic classification was calculated using GTDB-Tk
(24) with slight modification based on the potential metabolic dif-
fferences. Open reading frames (ORFs) of these MAGs were predicted
with Prodigal version 2.6.3 (https://github.com/hyattpd/Prodigal).
The predicted ORFs were searched against the NCBI nr protein database (July 2019) and eggng (68) database with the
BLASTP algorithm (e values < 1 × 10−10) to check their protein
identities to the most closely related sequences using DIAMOND
data file S2). For metabolic pathway analyses, we used the web por-
tal GhostKOALA on the Kyoto Encyclopedia of Genes and Genomes
website (www. kegg.jp/ghostkoala/).

**Phylogenetic analyses based on conserved proteins and
Mcr/Ecr/Acr, SMC, Mtr, Mts, Mtb, and Mta protein sequences**

For phylogenetic analysis, a total of 337 representative archaea
reference genomes [data files S1 (table S2)] from the superphyla
Euryarchaeota, TACK, Asgard, and DPANN were downloaded from
the NCBI prokaryote genome database (www.ncbi.nlm.nih.gov/
assembly/). These reference genomes and the MAGs from this study
were used to construct a phylogenomic tree on the basis of a concate-
inated alignment of a set of 37 marker genes as suggested by (25)
data file S1 (table S3)]. Specifically, each of the 37 marker protein
sequences from the reference genomes and the MAGs were aligned
using the MAFFT algorithm version 7.313 (https://mafft.cbrc.jp/
alignment/software/) with parameters --ep 0 –genafpair --maxiterate
1000 and filtered with trimAl version 1.4.rev2 (https://sourceforge.
net/projects/trimal) with parameter -automated1. Then, all 37 marker
genes were concatenated into a single alignment, and phylogenetic
trees were built using IQ-TREE version 1.6.6 (69) with model
LG+C60+F+G with bootstrap value of 1000. The tree was rooted at
DPANN, as suggested by Williams et al. (36). For the phylogeneti-
cal analysis of functional marker proteins (Mcr/Ecr/Acr, SMC,
Mtr, Mts, Mtb, and Mta; data file S3), the protein sequences were
retrieved from the MAGs and the reference genomes in the pre-
sent study by a DIAMOND search of reference protein sequences
with parameters of coverage >75% (e values < 1 × 10−20). Align-
ment and filtering were carried out with the same programs
described above, and phylogenetic trees were built using IQ-TREE
version 1.6.6 with the fitting model chosen according to the BIC
criterion (LG+C60+F+G) with 1000 ultrafast bootstraps. The
ALE analyses were performed using the maximum likelihood
implementation of the undated ALE algorithm using a sample of
1000 ultrafast bootstrap trees for each gene family (Mcr/Ecr/Acr/
MtrA and MtaA) and their species trees to estimate conditional clade
probabilities.
Divergence time estimation

Molecular timing analyses were conducted by three methods, i.e., Bayesian analysis of molecular sequences using Markov chain Monte Carlo (MCMC) BEAST version 1.10.4 (70) with Yule relaxed clock, Bayesian estimation of species divergence times using soft fossil constraints MCMECTree version 4.9c (71) with JC69 model, and treePL (72) with thorough option, which will continue iterating until convergence. Two different age constraints were considered for the divergence time estimation. One is the fossil evidence of the cyanobacterial clades Nostocales and Stigonematales (1.2 to 2.0 Ga) and the other is the predicted age of all methanogens. Proc. Natl. Acad. Sci. U.S.A. 116, 5037–5044 (2019).


Acknowledgments: We wish to clarify an unusual situation regarding the work presented here. An earlier version of this paper unintentionally included data that had been deposited in the NCBI SRA database but was under embargo in another database; to protect the interests of the original data generators, we retracted that version of the paper. In this updated and reworked paper, all of the data generated by other laboratories that we analyzed here were used with the explicit permission of the original data generators whom we thank for generously sharing the data with us. In particular, we thank Z. Du, R. Reiss, E. Lilleskov, C. Schadt, B. Hedlund, S. Caffrey, K. McMahon, R. Stepanauskas, R. Kelly, M. Hess, I. N. Sierra Garcia, and B. J. Baker for help, as well as the GEM project. We are grateful to the researchers who published their sequence data on NCBI (www.ncbi.nlm.nih.gov/) and the U.S. Department of Energy (DOE) Joint Genome Institute (JGI) (https://jgi doe.gov/), a DOE Office of Science User Facility, supported under contract no. DE-AC02-05CH11231. We also thank the scientists who analyzed the public datasets and assembled MAGs that were reassembled in the current study. The computations here were run on the x.2.0 cluster supported by the Center for High Performance Computing at Shanghai Jiao Tong University and with help from C. Li and J. Wang, and we thank W. Liu for the MAGs submission to eLMSG (eLibrary of Microbial Systematics and Genomics; www.biosino.org/elmgen/index). We also thank E. Lilleskov for valuable suggestions for this manuscript and J. Wang and X. Feng for useful discussion.

Funding: We thank the following sources for funding: the National Key Research and Development Program of China (grant nos. 2018YFC0309800 and 2016YFA060102), COMRA project (DY135-B2-12), the National Natural Science Foundation of China (grant nos. 141525011, 41902313, 917511150, and 90251116), the National Natural Science Foundation of Shanghai (20ZR1428000), the Senior User Project of RV KEXUE (KEXUE2019GZ06), and Shanghai Jiao Tong University interdisciplinary grant (20XK-01). G.W. was supported by the DFG Cluster of Excellence, The Ocean Floor-Earth’s Uncharted Interface (EXC-2007-39071643) at MARUM, University of Bremen. T.A.W. is supported by a Royal Society University Research Fellowship (UF146026). Author contributions: Y.W. designed the research, performed the analyses, and wrote the paper. T.A.W. provided guidance on inferring and interpreting phylogenies, performed gene–species tree analyses, and wrote the paper. R.X. performed the analyses of evolutionary time calculation. J.H. performed double-blind assessments of the MAGs. C.T. and Y.Z. provide important MAGs in the present study. F.W. and X.X. performed the analyses and wrote the paper. Competing interests: The authors declare that they have no competing interests.

Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. All scripts and analyses necessary to perform metagenome processing can be accessed from GitHub (https://github.com/) or the websites provided in the original research articles. The specific links to the custom software are listed below: DIAMOND version 0.8.28.90, http://ab.inf.uni-tuebingen.de/software/diamond/; Sickle version 1.3.3, https://github.com/najoshi/sickle; MEGAHIT version 1.3.0, https://github.com/krishnanlab/megahit; Bowtie version 2.2.8, https://bowtie-bio.sourceforge.net/bowtie2/index.shtml; Prodigal version 2.6.3, https://github.com/hyatt/pdProdigal;

Submitted 23 April 2021
Accepted 21 May 2021
Published 2 July 2021
10.1126/sciadv.abj1453

A methylotrophic origin of methanogenesis and early divergence of anaerobic multicarbon alkane metabolism

Yinzhao Wang, Gunter Wegener, Tom A. Williams, Ruize Xie, Jialin Hou, Chen Tian, Yu Zhang, Fengping Wang and Xiang Xiao

Sci Adv 7 (27), eabj1453.
DOI: 10.1126/sciadv.abj1453