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

Yinzhao Wang 1*, Gunter Wegener 2,3, Tom A. Williams 4, Ruize Xie 5, Jialin Hou 1, Fengping Wang 1,5,6, Xiang Xiao 1,7*

Methanogens are considered as one of the earliest life forms on Earth, and together with anaerobic methane-oxidizing archaea, they have crucial effects on climate stability. Yet, the origin and evolution of anaerobic alkane metabolism in the domain Archaea remain controversial. Here, we show that methanogenesis was already present in the common ancestor of Euryarchaeota, TACK archaea, and Asgard archaea likely in the late Hadean or early Archean eon and that the ancestral methanogen was dependent on methylated compounds and hydrogen. Carbon dioxide-reducing methanogenesis developed later through the evolution of tetrahydromethanopterin S-methyltransferase, which linked methanogenesis to the Wood-Ljungdahl pathway for energy conservation. Multicarbon alkane metabolisms in Archaea also originated early, with genes coding for the activation of short- 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.) Bathyarchaeota, Ca. Helarchaeota, Ca. Hadesarchaeota, 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 methane-oxidizing 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 primary 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 aceticlastic 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 (CoM) (11, 13). However, hydrogen-dependent methylotrophic archaea of the order Methanomassiliicoccales have been cultured that 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 archaeal diversity have supported the hypothesis that the last common ancestor of Euryarchaeota and TACK archaea (Fig. 1A, fig. S1) 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 when methanogens and anaerobic multicarbon alkane-oxidizing archaea emerged and how they evolved on the early Earth. To do so, we retrieved MCR/ACR-containing metagenome-assembled genomes (MAGs) from 73 publicly available metagenome datasets in National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database. Many of the MAGs were assigned to the class I (Methanobacteriales, Methanococcales, and Methanopyrales) and class II (Methanosarcinales, Methanomicrobiales, and Methanocellales) methanogens. These groups are well represented by genomes from the cultured strains, and hence, our MAGs of these groups were removed. A selection of 127 MAGs of MCR/ACR-containing organisms mostly with a completeness of >80% and contamination below 5% were retained for further analyses [24 MAGs have less than 80% completeness, 11 have more than 5% contamination, and 3 have low completeness but are of importance in the current study given their gene contents and phylogenetic positions; Table 1 and data file S1 (tables S1 to S3)]. All MAGs with MCR/ACR affiliated nomic tree based on a concatenated set of 37 marker genes ([fig. 23](#fig23)), the NCBI prokaryotes database were used to construct a phylogenetic tree, which varied when using different phylogenetic algorithms. Consistent with previous reports ([15](#15)), topologies of trees for McrA sequences and species trees in the current study are highly congruent for the class I and class II methanogens in Euryarchaeota (Fig. 1, A and B, and fig. S1). In the TACK superphylum, four lineages ([Thaumarchaeota, *Candidatus* (Ca.) Verstraetearchaeota, Ca. Korarchaeota, and Ca. Nezharchaeota] with MCR-encoding genes on the phylogenomic tree also match well to their McrA phylogenetic tree (Fig. 1, A and B, and figs. S1 and S2), except for the branching position of Ca. Korarchaeota on species trees, which varied when using different phylogenetic algorithms.

In addition, we found that 15 MAGs assembled here and in previous studies branch near the base of Euryarchaeota and TACK in both the MCR and species trees, providing insight into the early evolution of methanogenesis (Fig. 1, A and B, and fig. S1). Two early diverging lineages from Euryarchaeota branch between the

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

| Classification | MAGs | Metabolic potentials |
|----------------|------|----------------------|
| *Ca. Nuwarchaeia* | 5 | CH₃-reducing methanogen |
| *Ca. Methanofastidiosa* | 1 | CH₃-reducing methanogen or C₂H₆ metabolism |
| Thermoplasmata | 33 | CO₂-reducing methanogen with C₂H₆ oxidation |
| *Ca. Methanoliparia* | 2 | CO₂-reducing methanogen or CH₄ oxidation |
| Archaeoglobi | 1 | CO₂-reducing methanogen |
| Methanomicrobia | 2 | CH₄ oxidation |
| *Ca. Methanoflorens* | 6 | C₂H₂ oxidation |
| *Ca. Argoarchaeum* | 3 | C₂H₂ oxidation |
| *Ca. Ethanoperedens* | 1 | C₂H₂ oxidation |
| *Ca. Methanophagaes* | 22 | CH₄ oxidation |
| *Ca. santabarbaracales* | 2 | CH₂ = CH₂ oxidation (n ≥ 3) |
| *Ca. Syntropharchaeas* | 4 | C₂H₂ oxidation |
| *Ca. Hadesarchaeota* | 5 | C₂H₂ oxidation |
| *Ca. Nezharchaeota* | 7 | CO₂-reducing methanogen |
| *Ca. Verstraetearchaeota* | 17 | CH₃-reducing methanogen or C₂H₂ metabolism |
| *Ca. Korarchaeota* | 1 | CH₃-reducing methanogen |
| Thaumarchaeota | 2 | CH₃-reducing methanogen |
| *Ca. Helarchaeota* | 7 | C₂H₂ oxidation |
TACK and Euryarchaeota class I methanogens: the class Ca. Methanofastidiosa and one group branching next to Ca. Methanofastidiosa (including a previously reported MAG NM3) (20) belonging to an additional class-level lineage, which we here propose as Ca. Nuwarchaeia (fig. S3; Nuwa, the mother of earth in the Chinese myth), a potential hydrogen-dependent methylotrophic methanogen without the Wood-Ljungdahl pathway as also described in (20). Ca. Methanofastidiosa and Ca. Nuwarchaeia branch together with Thermococci and
The ancestral methanogen likely reduced methylated compounds

The high congruence between the topologies of the McrA phylogenetic tree and the methanogen species tree suggests that the common ancestor of Euryarchaeota and TACK already contained methanogenic McrA-encoding genes (Fig. 1, A and B). To evaluate the statistical support for this hypothesis, we used probabilistic gene tree–species tree reconciliation [the amalgamated likelihood estimation or ALE method; (26)] to estimate the number of copies of the McrA-encoding gene family 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 (27), particularly in the presence of HGT (28). The result indicated that McrA was already present in the common ancestor of the Euryarchaeota, TACK, and Asgard archaea (mean of 1.07 gene copies averaged across 100 sampled reconciliations). Previous studies pointed out that both Euryarchaeota and TACK combine the McrA and the Wood-Ljungdahl pathway via the MTR complex and therefore that the ancestral methanogen should be capable to reduce carbon dioxide to methane (19). However, on the basis of our current knowledge, the combination of Mcr, MTR, and Wood-Ljungdahl pathway is restricted to the class I and II methanogens in Euryarchaeota, as well as Ca. Nezhararchaeota in TACK. All other basal lineages of methanogens from TACK and Euryarchaeota lack the MTR complex (Fig. 1, A to C, and fig. S3). To resolve the evolutionary history of MTR-encoding genes, phylogenetic trees of MTR subunits were analyzed (Fig. 1C). For the Euryarchaeota class I and II methanogens, the MTR-encoding genes have very similar topologies as the corresponding phylogenomic trees. However, MTR-encoding genes from Ca. Nezhararchaeota clusters with those of Ca. Methanophagales and Ca. Methanoliparia, and an approximately unbiased test (29) for the catalysis subunit MtrA rejects the monophyly of Ca. Nezhararchaeota 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. Nezhararchaeota and, consequently, an origin of the MTR within the Euryarchaeota phylum, likely in the class I methanogens (see MTR phylogenetic tree; Fig. 1C and fig. S5). Some archaea from the TACK and Asgard superphyla 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). As above, we used ALE to evaluate when MTR originated in archaeal evolution. We found no support for an early origin of MTR (0.01 mean gene copies at the root, 0.16 in the common ancestor of Euryarchaeota, and 0.01 in the common ancestor of TACK/Asgard and TACK alone). Instead, the MTR appears to have originated within Euryarchaeota after the divergence of Ca. Hadesarchaeota, Ca. Nuwararchaeia, and Ca. Methanofastidiosa, with the gene appearing in the common ancestor of Methanobacteria, Methanococci, Methanopyri, and Methanomicrobia. 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 (Mtbb), and methylated-thiol–corrinoid
protein:coenzyme M methyltransferase (Mts) (fig. S3). 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 mostly vertically inherited and that they were likely present in the common ancestor of Euryarchaeota and TACK (fig. 1D and fig. S6).

Hydrogen-dependent methyl group reduction as the basal metabolic mode of the first methanogens is well supported by the conditions of the early Earth. Our planet was anoxic and likely rich in methylated compounds, hydrogen and simple organic acids (30, 31). These environments should have been suitable for hydrogen-dependent methyl-reducing methanogenesis. This finding does not contradict the hypothesis that both the MCR and the Wood–Ljungdahl pathway were early metabolic traits of archaea (4, 8, 15, 25, 28, 32) because methane production via the methylo trophic methanogenesis pathway and carbon dioxide fixation through the Wood–Ljungdahl pathway can work separately. Ancestral class I methanogens likely developed 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, 33). The class II methanogens are the most successful on modern Earth and members of this group thrive on carbon dioxide reduction, disproportional methylated substrates, and acetate (11). With the accumulation of electron acceptors such as sulfate in the late Archean (34), several methanogenic lineages started to reverse the methanogenesis pathway and turned into anaerobic methane oxidizers coupling with sulfate-reducing bacteria, 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 (35, 36).

**Multicarbon alkane metabolisms in Ca. Verstraetearchaeota 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) and also in Ca. Bathyarchaeota from TACK (16) and Ca. Helarchaeota from the Asgard superphylum (37). Among the ACR-containing MAGs assembled in the present study (Table 1), we identified two Ca. Verstraetearchaeota MAGs from hot spring sediments that code both an MCR and an ACR (Fig. 2, A to C, and fig. S7). The contig coding for these genes appears to be endogenous to Ca. Verstraetearchaeota because 176 of its 188 genes (93.6% of the encoded genes) show highest database amino acid identities (~65.8% on average) to sequences of other Ca. Verstraetearchaeota MAGs (fig. S8, A and B). One mcr gene clusters with those of other Ca. Verstraetearchaeota, and this gene is surrounded by other genes of the methanogenesis pathway (fig. S8C). On the MCR/ACR phylogenetic tree, the ACR sequences of Ca. Verstraetearchaeota cluster closely but as a basal lineage with the ACR from Ca. Argoarchaeum (10) and Ca. Ethanoperedens (38). 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. Verstraetearchaeota ECR-encoding genes are surrounded by genes coding for coenzyme A (CoA) ligase, cobalamin-binding protein, carboxylesterase, and fatty acid–CoA ligase (fig. S8C), enzymes that might help to metabolize reactants or products of ethane metabolism (Fig. 2C). In addition to the ECR-containing Ca. Verstraetearchaeota, we assembled three MAGs from Methanomassiliicoccales in Euryarchaeota from peatland samples that also contain both MCR- and/or ECR-encoding genes (Fig. 2, A, B, and D). However, ECR-encoding genes are next to corrinoid protein, cobalamin-binding protein, and methyltransferase-like enzyme–encoding genes (Fig. 2D). Both MCR/ECR-containing Ca. Verstraetearchaeota and Methanomassiliicoccales have several methyltransferases such as Mta, Mtb, and Mts that enable an 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 (energy-converting hydrogenases) for CoBS-SCoM cycling and energy conservation. However, in contrast to Ca. Argoarchaeum and Ca. Ethanoperedens, ECR-containing Ca. Verstraetearchaeota and Methanomassiliicoccales do not encode a Wood–Ljungdahl pathway; it therefore lacks an obvious pathway to oxidize CoM-bound ethyl groups to carbon dioxide. In contrast, Ca. Verstraetearchaeota and Methanomassiliicoccales 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 subsurface responsible for ethanogenesis in anoxic sediments (39). Ethanogenic organisms are not yet cultured, but based on their coded pathways, ECR-containing Ca. Verstraetearchaeota and Methanomassiliicoccales might perform these reactions. Alternatively these organisms may oxidize ethane, propane, or other alkanes, using thus far 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 duplication that took place before the radiation of the Euryarchaeota, TACK, and Asgard 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). The consensus-rooted MCR/ECR/ACR-encoding gene tree (fig. S9) excludes the root from the three clades with high bootstrap support: the TACK MCRs (100% bootstrap), Eur yarchaeota MCRs (87% bootstrap), and a clade comprising all ECRs and ACRs that metabolize longer carbon chains (100% bootstrap). The exclusion of the root from the ACR clade (fig. S9) indicates that the ACR sequences that bind longer-chain alkanes evolved from within a paraphyletic grade of ECR sequences, including those found from Ca. Verstraetearchaeota and Methanomassiliicoccales. This suggests that the enzymes that potentially metabolize longer-chain alkanes, such as the ACRs in Ca. Helarchaeota in Asgard and Ca. Bathyarchaeota in TACK, as well as Ca. Hadesarchaeota and Ca. Methanoliparia, evolved or transferred from those involved in potential ethane metabolism.
By extension, it is tempting to speculate 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). Alternatively, they might have evolved from an ancestral gene coding for a methyl-binding enzyme at the root of the MCR/ACR family. However, since there are still too few ECR/ACR-containing MAGs available, our analysis only showed a highly significant basal node within the Euryarchaeota (with 4.62 mean gene copies), 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, to 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 additional sequences from Ca. Verstraetearchaeota. Trees of both the alpha and gamma subunits of the MCR/ECR/ACR enzymes suggest that the Ca. Verstraetearchaeota and Methanomassiliicoccales 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 extinct, or unsampled, ACR lineage (Fig. 2, A and B). Exploration of MAGs from other extreme environments such as hot springs or hydrothermal vents will help to determine whether other TACK archaea encode relatives of this deep-branching ACR lineage. Alternatively, the Ca. Verstraetearchaeota 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 the deep-branching organisms

Fig. 3. Evolutionary history of MCR/ECR/ACR-based anaerobic alkane metabolisms. (A) A scenario for the evolutionary history of MCR/ECR/ACR-encoding genes. After the emergence of an ancestral MCR that reduced methylated compounds, in the common ancestor of Euryarchaeota, TACK, and Asgard, MCR was inherited by each of the two lineages. ECR/ACR evolved before the radiation of extant Euryarchaeota, on the euryarchaeotal or TACK stem, or potentially earlier. MTR originated in Euryarchaeota and was subsequently horizontally transferred to Ca. Nezhararchaeota in TACK. Before the radiation of Euryarchaeota, a putative duplication of MCR-encoding genes created the ECR-encoding genes, which later evolved to metabolize longer-chain alkanes and spread by both vertical descent and horizontal transfer in the domain Archaea. "l." refers to long chain, and "s." refers to short chain. (B) A cartoon phylogenetic tree for archaeal evolution and MCR/ECR/ACR divergence. MCR-encoding gene might have originated by the time of the last archaea common ancestor (LACA) assuming that DPANN had lost their MCR-encoding gene during evolution. ECR/ACR might have been present in LACA or the stem lineage of Euryarchaeota, TACK, and Asgard but underwent multiple HGT processes. (C) Predicted ancestral MCR/ECR/ACR-based anaerobic alkane metabolism evolution in Archaea. First, a methyl-reducing methanogen or ethane-metabolizing archaea; second, the origin of MTR complex created the carbon dioxide–reducing methanogen; then, the introduction of β-oxidation pathway enabled the anaerobic multicarbon alkane-metabolizing archaea. MT, methyl transferase.
with ACR such as Ca. Bathyarchaeota in TACK, Ca. Helarchaeota in Asgard, and Ca. Hadesarchaeota in Euryarchaeota encode the Wood-Ljungdahl and β-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 (40). Considering the low sulfate levels in the Archean ocean (34), at that time, the partners of the alkane oxidizers might have been rather methanogens than sulfate reducers. Or, they might use a methanogenic alkane degradation pathway, which requires substrates with chain length of at least six carbons (hexane) (41). In contrast, the more modern lineages of Ca. Syntrophoarchaeia have cytochromes that allow direct electron transfer with cytochrome-containing partners such as Ca. Desulfurovibrio (42). The phylogenetic positions of these ACR sequences, from one Archaeoglobi lineage and members in the Euryarchaeota class Ca. Syntrophoaarchaeia, match well with their species tree (fig. S11). This suggests that these ACR-encoding genes were vertically inherited from their last common ancestor, which might be the ACR-containing deep-branching Archaeoglobi.

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

The acquisition of chromosome structural maintenance protein SMC-encoding genes by the cyanobacterial ancestor from an archaeal lineage in Euryarchaeota, most likely within the Methanomicrobia or Halobacteria, has been recently used to date the origin of euryarchaeal methanogens before 3.51 billion years (Ga) (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. Because the euryarchaeal 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; (43, 44)] to date events within Archaea, for which no unambiguous fossil record is available for other prokaryotes before Archean. The SMC gene partition provides a fossil calibration 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 the cyanobacterial clade for the SMC phylogenetic tree shows high congruence with the phylogenomic 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 (43, 44) and the predicted divergence of archaea and bacteria (45), we predict that the common ancestor of methanogenic archaea originated ~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 posterior age estimates for the key nodes in the analysis of the tree built by the conserved ribosome protein plus SMC sequences 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 (~3.94 Ga) and 95% highest probability density (3.50 to 4.35 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, 46, 47). This is earlier than a recent estimate of the ancestral methanogen age (48), and indeed the authors of that study also stated that the discovery of non-euryarchaeal methanogens would result in an even earlier origin of methanogenesis. Our analyses showed that the first methanogen was a hydrogen-dependent methylotroph (Fig. 3, A to C, and fig. S13), 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 (49); however, 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 archaeal 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 ~3.78 Ga (Fig. 3, A to C), supporting the geological evidence that methane produced via carbon dioxide reduction as characterized by highly depleted isotope signatures was dated back to 3.46 Ga (7). In TACK, the ancestral methyl-reducing methanogenesis pathway persisted, although some members such as Ca. Nezharchaeota 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-dependent methylotrophic archaea. Life likely originated at hydrothermal vents or serpentinitization sites that provided ideal conditions for these methanogens (fig. S14) (8, 50, 51). The temperatures were elevated compared to the otherwise rather cold planet owing to the low illumination from the young Sun, and large amounts of molecular hydrogen and simple organic compounds such as methanol and acetate would provide separated carbon and energy sources for methanogens (52). Meanwhile, with increasing population sizes, microorganisms would need to fix carbon themselves, which may have led to the development of the Wood-Ljungdahl pathway. Last, the evolution of MTR would 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 (53), and hence, methanogenesis may have had crucial impact on the climate of the early Earth. The accumulating methane would cause 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**

The NCBI SRA files are selected by text searching of potential environments that may harbor methanogens, such as sediment, hot spring, hydrothermal vent, peatland, bioreactor, etc. on the NCBI database (www.ncbi.nlm.nih.gov/sra/). Then, these SRA files were downloaded, and raw sequencing reads were trimmed by the Sickle algorithm version 1.33 (https://github.com/najoshi/sickle). The trimmed reads were assembled using MEGAHIT version 1.0.6-hotfix1 (54) with step 4 and/or SPAdes-3.13.1 (55) with k-step 4. We then also downloaded the McrA/AcrA sequences from NCBI protein nr
database and built a local McrA/AcrA diamond database. 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 with the known sequences were selected and resulted in 73 SRA datasets [data file S1 (table S1)]. Then, some of the datasets were combined and reasssembled, and each assembled contig coverage was determined by mapping the trimmed reads back to the contigs by Bowtie version 2.2.8 (56) with parameter --very-sensitive. The assembled metagenomic sequences were binned on the basis of MaxBin version 2.2.4 (57) with the run_MaxBin.pl script and on the basis of abundance and tetrancleotide frequency using MetaBAT version 2.12.1 (58) with 1 (or 1.5 kb) and 3 kb as contig length cutoffs. The 4-mer and 5-mer frequency of each contig was calculated, and dimensionality was reduced by t-SNE clustering algorithm (59) (https://github.com/lejon/T-SNE-Java/tree/master/tsne-core). The selected MAGs were then checked with the mngenome package version 0.6.3. Final completeness and contamination of each MAG [data file S1 (table S2)] were assessed with CheckM version 1.0.7 (60) using lineage-specific (149 to 228) marker genes. 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 (2019/07) and eggNOG (61) database with the BLASTP algorithm (coverage of >75%, e values of <1 × 10−20) to check their protein identities to the most closely related sequences using DIAMOND sequence aligner version 0.8.28.90 (http://github.com/bbuchfink/diamond). Specifically, the NCBI nr protein database and eggNOG database were downloaded, and local protein databases were constructed (data file S2). Then, the MAG protein sequences were searched against the established databases as query. The highest-scored sequences from the two databases were cross-checked, and sequences related to MCR-based alkane metabolism were also manually checked by BLASTP search on the NCBI website (https://blast.ncbi.nlm.nih.gov/Blast.cgi). For metabolic pathway analyses, we used the web portal GhostKOALA on the KEGG (Kyoto Encyclopedia of Genes and Genomes) website (www.kegg.jp/genes/koala/).

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

For phylogenetic analysis, a total of 337 representative archaea reference genomes [data file S1 (table S3)] 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 based on a concatenated alignment of a set of 37 marker genes as suggested in (23) [data file S1 (table S4)]. Specifically, each of the 37 marker protein sequences from the reference genomes and the MAGs was aligned using the MAFFT algorithm version 7.313 (https://mafft.cbrc.jp/alignment/software/) with parameters --ep 0 --globalpair or genapair --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 (62) with model LG+G+F+G with bootstrap value of 1000. The tree was rooted at DPANN as suggested by Williams et al. (28). For the phylogenetic analysis of functional marker proteins (Mcr/Acr/Acr, SMC, Mtr, Mts, MtA, and Mta; data file S3), the respective protein and nucleotide sequences were retrieved from the MAGs, and additional reference sequences were obtained from NCBI (www.ncbi.nlm.nih.gov/protein/). Alignment 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 best-fitting model chosen according to the Bayesian information criterion (LG+C+G+F+G) with 1000 ultrafast bootstraps. The ALE analyses were performed using the maximum likelihood implementation of the updated ALE algorithm using a sample of 1000 ultrafast bootstrap trees for each gene family (Mcr/Acr/Acr/Acr and Mtr) 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 (Bayesian Evolutionary Analysis Sampling Trees) version v1.10.4 (63) with Yela relax model; Bayesian estimation of species divergence times using soft fossil constraints MCMCtree version v4.9c (64) with JCE9 model; and treePL (65) with thorough option, which will make sure to 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; fossil resting cells similar to the Nostocales and Stigonematales from Francjeville Group of Gabon; (43, 44)]. The other is the predicted time of differentiation time of bacterial and archaeal lineages, i.e., 4.26 Ga (45). The SMC phylogenetic tree calculating with model LG+C+G+F+G by IQ-TREE was used for BEAST, MCMCtree, and treePL configuration.

**Supplementary materials**

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/7/7/eabd7180/DC1

View/request a protocol for this paper from Bio-protocol.

**References and notes**

1. D. H. Rothman, G. F. Fournier, K. L. French, E. J. Alm, E. A. Boyle, C. Cao, R. E. Summons, Methanogenic burst in the end-Permain carbon cycle. *Proc. Natl. Acad. Sci.* 111, 5462–5467 (2014).

2. J. M. Wolfe, G. F. Fournier, Horizontal gene transfer constrains the timing of methanogen evolution. *Nat. Ecol. Evol.* 2, 897–903 (2018).

3. B. Sauterey, B. Charnay, A. Affholder, S. Mazevet, R. Ferrière, Co-evolution of primitive methane-cycling ecosystems and early Earth’s atmosphere and climate. *Nat. Commun.* 11, 2705 (2020).

4. P. N. Evans, J. A. Boyd, A. O. Leu, B. J. Woodcroft, D. H. Parks, P. Hugenholtz, G. W. Tyson, An evolving view of methane metabolism in the Archaea. *Nat. Rev. Microbiol.* 17, 219–232 (2019).

5. Y. Wang, G. Wegener, S. E. Ruff, F. Wang, Methyl/alkyl-coenzyme M reductase-based anaerobic alkane oxidation in archaea. *Environ. Microbiol.* 10.1111/1462-2920.15057 (2020).

6. K. Knittel, A. Boetius, Anaerobic oxidation of methane: Progress with an unknown process. *Annu. Rev. Microbiol.* 63, 311–334 (2009).

7. Y. Jeno, K. Yamada, N. Yoshida, S. Maruyama, Y. Isozaki, Evidence from fluid inclusions for microbial methanogenesis in the early Archaean era. *Nature* 440, 516–519 (2006).

8. W. F. Martin, F. L. Sousa, Early microbial evolution: the age of anaerobes. *Cold Spring Harb. Perspect. Biol.* 8, a018127 (2016).

9. R. Laso-Pérez, G. Wegener, K. Knittel, F. Widdel, K. J. Harding, V. Krukenberg, D. V. Meier, M. Richter, H. E. Tegetmeyer, D. Riedel, H.-H. Richnow, L. Adrian, T. Reemtsma, O. J. Lichtenfeld, F. Musat, Thermophilic archaea activate butane via alkyl-coenzyme M formation. *Nature* 539, 396–401 (2016).

10. S. C. Chen, N. Musat, O. J. Lichtenfeld, H. Paschke, M. Schmidt, N. Said, D. Popp, F. Calabrese, H. Styrhanyuk, U. Jaekel, Y. G. Zhu, S. B. Joye, H.-H. Richnow, F. Widdel, F. Musat, Anaerobic oxidation of ethane by archaea from a marine hydrocarbon seep. *Nature* 568, 108–111 (2019).
64. M. dos Reis, Notes on the birth–death prior with fossil calibrations for Bayesian estimation of species divergence times. *Philos. Trans. R. Soc. B Biol. Sci.* **371**, 20150128 (2016).

65. S. A. Smith, B. C. O’Meara, treePL: Divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* **28**, 2689–2690 (2012).

Acknowledgments: We thank J. Wang and X. Feng for valuable discussion and suggestions. We are grateful to the researchers who published their sequence data on NCBI (www.ncbi.nlm.nih.gov/). We also thank the scientists who analyzed the public datasets and assembled MAGs that were reassembled in the current study. The computations in this paper were run on the x2.0 cluster supported by the Center for High Performance Computing at Shanghai Jiao Tong University. Funding: We thank the following sources for funding: the National Key Research and Development Program of China (grant nos. 2018YFC0309800 and 2016YFA0601102), COMRA project (DY135-82-12), the National Natural Science Foundation of China (grant nos. 41525011, 41902313, 91751205, and 92051116), the Natural Science Foundation of Shanghai (20ZR1428000), the Senior User Project of RV KEXUE (KEXUE2019GZ06), and Shanghai Jiao Tong University interdisciplinary grant (20CX-01). G.W. was supported by the DFG Cluster of Excellence, The Ocean Floor–Earth’s Uncharted Interface (EXC-2077 – 390741603) at MARUM, University of Bremen. T.A.W. is supported by a Royal Society University Research Fellowship (UF140626). Author contributions: Y.W. designed research, performed the analyses, and wrote the paper. G.W. provided knowledge on metabolism of archaea and wrote the paper. T.A.W. provided guidance on inferring and interpreting phylogenies, performed gene tree–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. 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: The datasets generated and/or analyzed during the current study are available in the NCBI repository at www.ncbi.nlm.nih.gov/.