

Expanding anaerobic alkane metabolism in the domain of Archaea

Yinzhao Wang¹, Gunter Wegener^{2,3}, Jialin Hou¹, Fengping Wang^{1*} and Xiang Xiao^{1,4*}

Methanogenesis and anaerobic methane oxidation through methyl-coenzyme M reductase (MCR) as a key enzyme have been suggested to be basal pathways of archaea¹. How widespread MCR-based alkane metabolism is among archaea, where it occurs and how it evolved remain elusive. Here, we performed a global survey of MCR-encoding genomes based on metagenomic data from various environments. Eleven high-quality *mcr*-containing metagenomic-assembled genomes were obtained belonging to the Archaeoglobi in the Euryarchaeota, Hadesarchaeota and different TACK superphylum archaea, including the Nezaarchaeota, Korarchaeota and Verstraetearchaeota. Archaeoglobi WYZ-LMO1 and WYZ-LMO3 and Korarchaeota WYZ-LMO9 encode both the (reverse) methanogenesis and the dissimilatory sulfate reduction pathway, suggesting that they have the genomic potential to couple both pathways in individual organisms. The Hadesarchaeota WYZ-LMO4-6 and Archaeoglobi JdFR-42 encode highly divergent MCRs, enzymes that may enable them to thrive on non-methane alkanes. The occurrence of *mcr* genes in different archaeal phyla indicates that MCR-based alkane metabolism is common in the domain of Archaea.

Since the emergence of life on our planet, anaerobic methane metabolism, including methanogenesis and methane oxidation, has been a crucial element in the Earth's carbon cycle¹, and both processes are key to the global methane budget. Methanogenic archaea produce ~500–600 Tg of methane per year², whereas anaerobic methane-oxidizing archaea (ANME) oxidize a large portion of methane within the seafloor before it reaches the water column^{3,4}. The metabolic pathways of methane formation and anaerobic oxidation of methane are largely identical, as both contain exclusively C₁-compound-transforming enzymes that were described originally in the methanogenesis pathway^{5–7}.

Among these enzymes, methyl-coenzyme M (CoM) reductase (MCR) plays the crucial role⁸. In methanogens, MCR reduces CH₃-CoM to CH₄ (refs. ^{9,10}), whereas in ANMEs, this enzyme activates CH₄ to form CH₃-CoM as the primary intermediate⁵. This canonical MCR type is highly conserved, and the gene encoding the α -subunit (*mcrA*) of the enzyme complex has been used as a diagnostic marker for the detection and phylogenetic classification of methanogens and ANMEs¹¹. The presence of MCR was thought to be limited to the phylum Euryarchaeota, yet the occurrence of *mcr* genes in metagenome-assembled genomes (MAGs) of the Bathyarchaeota¹² and Verstraetearchaeota¹³ has indicated a much wider distribution of this gene within archaea. In addition, highly divergent MCR types have been identified in two strains of the thermophilic *Ca. Syntrophoarchaeum* spp. that use the encoded enzymes to activate

short-chain alkanes, such as propane and *n*-butane, therewith documenting a first case of a physiological activation of non-methane alkanes using MCR and the production of the corresponding alkyl-CoM as a primary intermediate¹⁴. None of the known MCR-based alkane-oxidizing archaea growing under sulfate-reducing conditions possess genes encoding enzymes for dissimilatory sulfate reduction; hence, all of them rely on syntrophic interactions with sulfate-reducing partner bacteria^{14–17}. By contrast, some ANME-2d contain nitrite or nitrate reductases, or may transfer electrons to metal oxides, and hence may not require partner bacteria when growing under these conditions^{18,19}.

In this study, we aimed to provide a larger overview of the distribution of MCR-based alkane metabolism in archaea from different environments. Thus, we established a database of known representative McrA protein sequences as a reference ($n = 153$; Supplementary Table 1), and with this list as a basis, the Joint Genome Institute (JGI) metagenomic protein databases (January 2017) were used as queries for screening homologous sequences (identity >30%, $e < 10^{-20}$, coverage >75%). Sixty-four metagenomic data sets produced on Illumina HiSeq 2000 or 2500 platforms containing McrA sequences were further analysed (Supplementary Table 2). These data sets were then downloaded from the NCBI Sequence Read Archive public database, reassembled, binned and annotated as described in the Methods section. A phylogenetic analysis of the McrA subunits retrieved from these metagenomes shows that they cluster with those of the Euryarchaeota phylum, including ANME-1, ANME-2a/b, ANME-2c, ANME-2d, ANME-3, *Ca. Syntrophoarchaeum* spp. and most of the known methanogens, and with those of non-Euryarchaeota phyla, including the Verstraetearchaeota and Bathyarchaeota. The majority of the McrA sequences that affiliate with ANME homologues show high identities to the known McrA sequences, whereas nearly half of the McrA sequences affiliating with the Verstraetearchaeota and *Ca. Syntrophoarchaeum* spp./Bathyarchaeota have less than 80% and 60% identities to the known sequences, respectively (Supplementary Table 3). This suggests that many of these sequences come from unknown microorganisms.

From the metagenomic sequences of 3 geothermal environments, we recovered 11 high-quality *mcr*-containing MAGs that contained divergent *mcr* genes (Table 1 and Supplementary Table 2). All MAGs were checked for lineage-specific (149–228) and Archaea (149) marker proteins to evaluate their completeness and contamination (Supplementary Table 4). The completeness of all reported MAGs was higher than 80%, whereas their contamination was below 2%. The GC contents, tetranucleotide frequencies and codon usage patterns of the *mcr* gene containing contigs show high congruence with their respective MAGs (Supplementary Figs. 1–3). This strongly

¹State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China.

²Max Planck Institute for Marine Microbiology, Bremen, Germany. ³MARUM, Center for Marine Environmental Sciences, University of Bremen, Bremen, Germany. ⁴State Key Laboratory of Ocean Engineering, Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai, China.

*e-mail: fengpingw@sjtu.edu.cn; zxiao2018@sjtu.edu.cn

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

MAG name	Sampling location	MCR group	Potential alkane metabolic pathways
Archaeoglobi WYZ-LMO1	Washburn Spring, WY, USA	II	MAO coupled to DSR in one cell or methanogenesis
Archaeoglobi WYZ-LMO2	Obsidian Pool, WY, USA	II	Methanogenesis from CO ₂ and H ₂ or MAO coupled to DSR
Archaeoglobi WYZ-LMO3	Obsidian Pool, WY, USA	II	MAO coupled to DSR in one cell or methanogenesis
Archaeoglobi JdFR-42 ^a	Juan de Fuca Ridge, Pacific	III	MAO coupled to nitrate reduction or external e ⁻ sinks
Hadesarchaeota WYZ-LMO4	Jinze Hot Spring, Yunnan, China	III	MAO with unknown or external e ⁻ sink or alkane production
Hadesarchaeota WYZ-LMO5	Jinze Hot Spring, Yunnan, China	III	MAO with unknown or external e ⁻ sink or alkane production
Hadesarchaeota WYZ-LMO6	Washburn Spring, WY, USA	III	MAO with unknown or external e ⁻ sink or alkane production
Nezhaarchaeota WYZ-LMO7	Jinze Hot Spring, Yunnan, China	II	Methanogenesis from CO ₂ and H ₂
Nezhaarchaeota WYZ-LMO8	Jinze Hot Spring, Yunnan, China	II	Methanogenesis from CO ₂ and H ₂
Korarchaeota WYZ-LMO9	Washburn Spring, WY, USA	II	MAO coupled to sulfite reduction; methanogenesis from methyl groups and H ₂ or reduced sulfur species as an e ⁻ source
Verstraetearchaeota WYZ-LMO10	Washburn Spring, WY, USA	II	Methanogenesis from methyl groups and H ₂
Verstraetearchaeota WYZ-LMO11	Washburn Spring, WY, USA	II	Methanogenesis from methyl groups and H ₂

DSR, dissimilatory sulfate/sulfite reduction; MAO, MCR-based alkane oxidation. ^aOrigin of this bin from ref. 53.

indicates that these contigs were correctly assigned into MAGs. By contrast, these measures show clear differences when comparing them with genomes of other closely related archaea. The 11 MAGs with divergent *mcr* genes and reference archaea genomes from the NCBI genome database were used to construct phylogenetic trees based on a concatenated set of 37 marker genes (Supplementary Table 5) and 16S rRNA gene (Fig. 1a,b). According to these analyses, the MAGs belong to the Euryarchaeota, Hadesarchaeota²⁰, Korarchaeota²¹, Verstraetearchaeota¹³ and an additional clade in the TACK superphylum (Table 1).

Three MAGs (WYZ-LMO1–3) clustered within the class Archaeoglobi, neighbouring with *Archaeoglobus fulgidus* (Fig. 1a,b). They showed whole genome average amino acid identities (AAIs) of 65% and 16S rRNA gene identities of ~85–94% to their most closely related cultured Archaeoglobaceae strains (Supplementary Tables 6 and 7). Three MAGs (WYZ-LMO4–6) clustered within the Hadesarchaeota and neighboured with MSBL-1 archaea (Fig. 1a,b). Whereas Hadesarchaeota have been previously suggested to cluster within the Euryarchaeota²⁰, recent analyses suggested the classification of Hadesarchaeota as a separate phylum Hadesarchaeota, probably branching at the base of the Asgard/TACK superphyla^{22,23}. Our results suggest that the Hadesarchaeota branch with Euryarchaeota. Yet, the bootstrap support for this is low, hence further analyses with additional archaeal lineages, including a congruent placement of the root of archaea^{24–26}, may be required to finally resolve the phylogenetic placement of the Hadesarchaeota. Two MAGs (WYZ-LMO7 and WYZ-LMO8) form a deep-branching cluster within the TACK superphylum. Although the partial 16S rRNA gene sequences (~900bp) from MAGs WYZ-LMO7 and WYZ-LMO8 shared ~89% identities with the Thermoprotei strains in Crenarchaeota (Fig. 1b), the AAIs with the closest reference genomes of the Verstraetearchaeota, Geothermarchaeota and Crenarchaeota were only ~48–50%. Moreover, both genomic protein sequences and MCR coding sequences have very low average identities (on average <50%) to the known protein sequences in the NCBI nr database (Supplementary Fig. 4), indicating that these MAGs may represent an additional lineage of alkane-metabolizing archaea. Although, more genomes from representatives of this lineage are required to determine its taxonomic level, here, we temporarily named it Nezhaarchaeota, after Nezha, an immortal in Chinese mythology. One MAG (WYZ-LMO9) clustered closely with the Korarchaeota (Fig. 1a,b) but shared only ~58% AAI and ~92% 16S rRNA gene identity with the *Ca. Korarchaeum cryptofilum* OPF8, suggesting

that MAG WYZ-LMO9 might represent an additional family in the phylum Korarchaeota. Two MAGs (WYZ-LMO10 and WYZ-LMO11) are affiliated with the Verstraetearchaeota, but showed <65% AAI with all published Verstraetearchaeota genomes and also displayed large distances in the phylogenomic tree. The MAGs WYZ-LMO10 and WYZ-LMO11 also share low AAI (~65%) between each other and show large differences in their average genomic GC content (WYZ-LMO10: 50.3%; WYZ-LMO11: 28.5%).

Our study extends the known distribution of *mcr* genes in the Archaea domain and identifies additional lineages of potential alkane-metabolizing archaea within the Archaeoglobi in the Euryarchaeota, Hadesarchaeota and TACK superphylum. The highly similar tree topologies for all three subunits clearly indicate a coevolution of the encoded protein subunits (Fig. 2). Based on phylogenetic analyses, the MCR sequences were categorized into three groups (I, II and III). Group I contains sequences of the canonical MCRs including those originating from the Euryarchaeota (Fig. 2). Group II is between the canonical and divergent MCRs and is dominated by sequences assigned to archaea of the TACK superphylum, yet it also includes the MCRs of Archaeoglobi WYZ-LMO1–3 (Fig. 2). Group III represents the putative multi-carbon alkane metabolizing MCRs, and includes not only divergent MCR sequences of *Ca. Syntrophoarchaeum* spp.¹⁴ and the Bathyarchaeota¹² but also those of the Hadesarchaeota and one Archaeoglobi MAG (Archaeoglobi JdFR-42, NCBI genomic database; Fig. 2).

The Archaeoglobi MAGs WYZ-LMO1–3 contain *mcr* genes of group II MCR and complete methanogenesis pathways (Fig. 3a,b and Supplementary Tables 8 and 9), whereas two MAGs (WYZ-LMO1 and WYZ-LMO3) additionally possess genes encoding the complete sulfate reduction pathway, including genes for sulfate adenylyltransferase, adenylylsulfate reductase (*AprAB*) and dissimilatory sulfite reductase (*DsrAB*). We also identified genes encoding the *DsrC* protein and the *DsrMKJOP* complex, which are regularly found in sulfate reducers. The *DsrA* sequences from WYZ-LMO1 and WYZ-LMO3 are closely related to those of the reducing-type enzymes in the Firmicutes (Supplementary Fig. 5). Many Archaeoglobi probably acquired the *dsrAB* and *aprAB* genes from a bacterial donor early during the evolution of this lineage and thus obtained the ability to perform dissimilatory sulfate reduction^{27,28}. For decades, it was believed that sulfate-dependent methane oxidation requires the interaction of anaerobic methane-oxidizing archaea and sulfate-reducing partner bacteria²⁹. However, based on the genomic potential described here, some Archaeoglobi might be

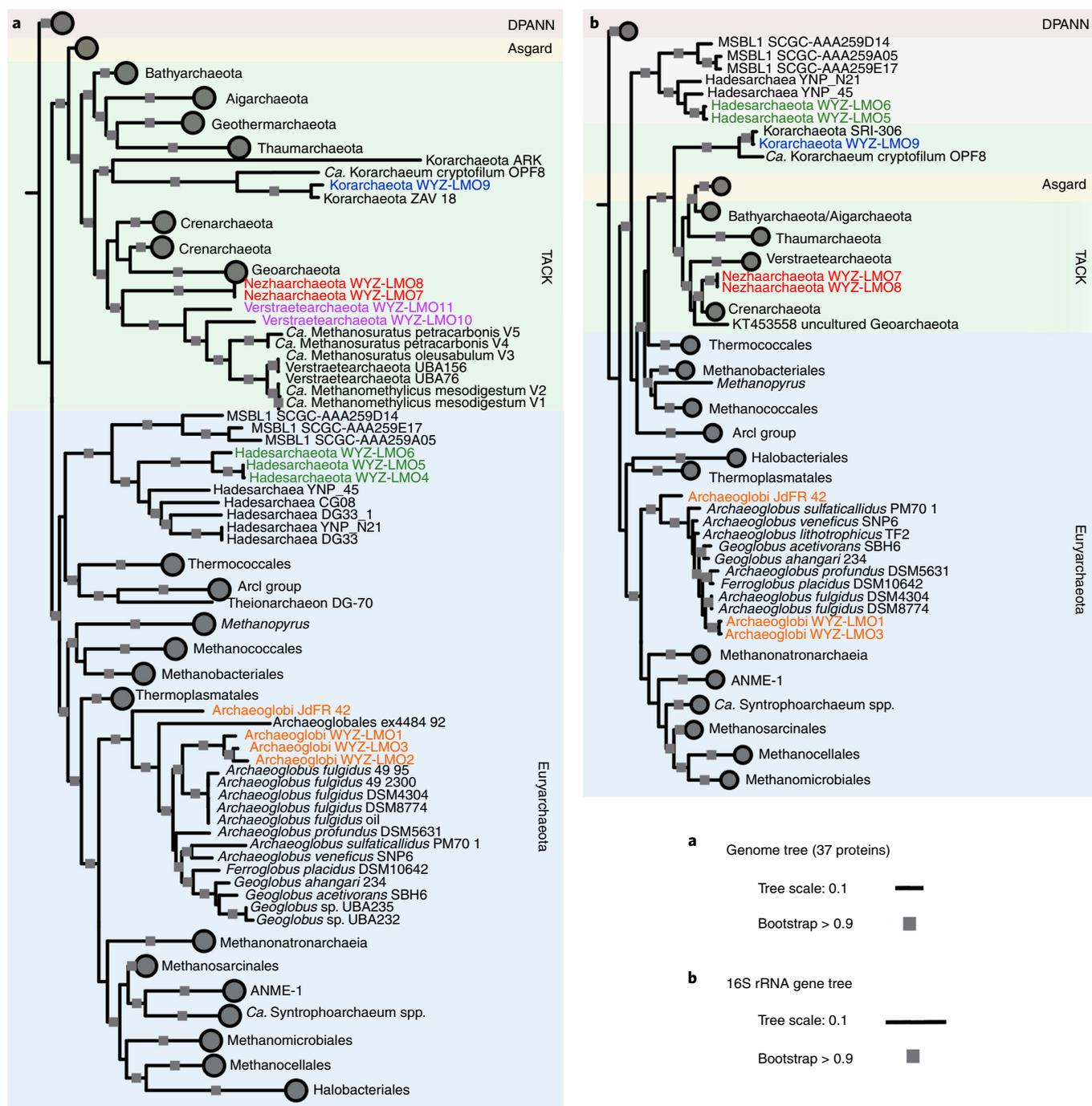


Fig. 1 | Classification of the 12 described MAGs. a, Phylogenomic affiliation of the MAGs based on 37 conserved protein sequences and using 233 representative archaeal genomes from 4 superphyla. **b**, Phylogenetic affiliation of the MAGs and representative archaea based on their 16S rRNA genes. Both alignments were based on MAFFT and then filtered with trimAl, and the trees were built by the IQ-Tree method with the model LG + C60 + F + G or HKY + F + G with 1,000 bootstrap replicates. The phylogenetic trees were rooted at the DPANN superphylum and all four superphyla were assigned different background colours. Distinct *mcr* genes containing archaea are shown with coloured text (the Archaeoglobi WYZ-LMO1–3/JdFR-42, orange; Hadesarchaeota WYZ4–6, green; Nezharchaeota WYZ-LMO7 and WYZ-LMO8, red; Korarchaeota WYZ-LMO9, blue; Verstraetearchaeota WYZ-LMO10 and WYZ-LMO11, purple).

able to perform sulfate-dependent methane oxidation in single cells. Yet, we cannot rule out the possibility that these Archaeoglobi are also methanogens, although no complete hydrogenase was found in their genomes. The Archaeoglobi MAG WYZ-LMO2 does not have the *dsrAB* and *aprAB* genes. Hence, as alkane oxidizer, it would require an electron-accepting partner organism. However, the Archaeoglobi MAG WYZ-LMO2 contains a complete Ni-Fe

hydrogenase gene set that resembles F_{420} -non-reducing hydrogenase (MvhADG). This enzyme is usually found in methanogens and can produce reduced ferredoxin and reduced coenzyme M (CoMSH) by consuming hydrogen³⁰. Hence, the Archaeoglobi WYZ-LMO2 might be instead a hydrogenotrophic methanogen. Only cultivation of representative organisms will be able to verify the genome-based hypotheses on these Archaeoglobi.

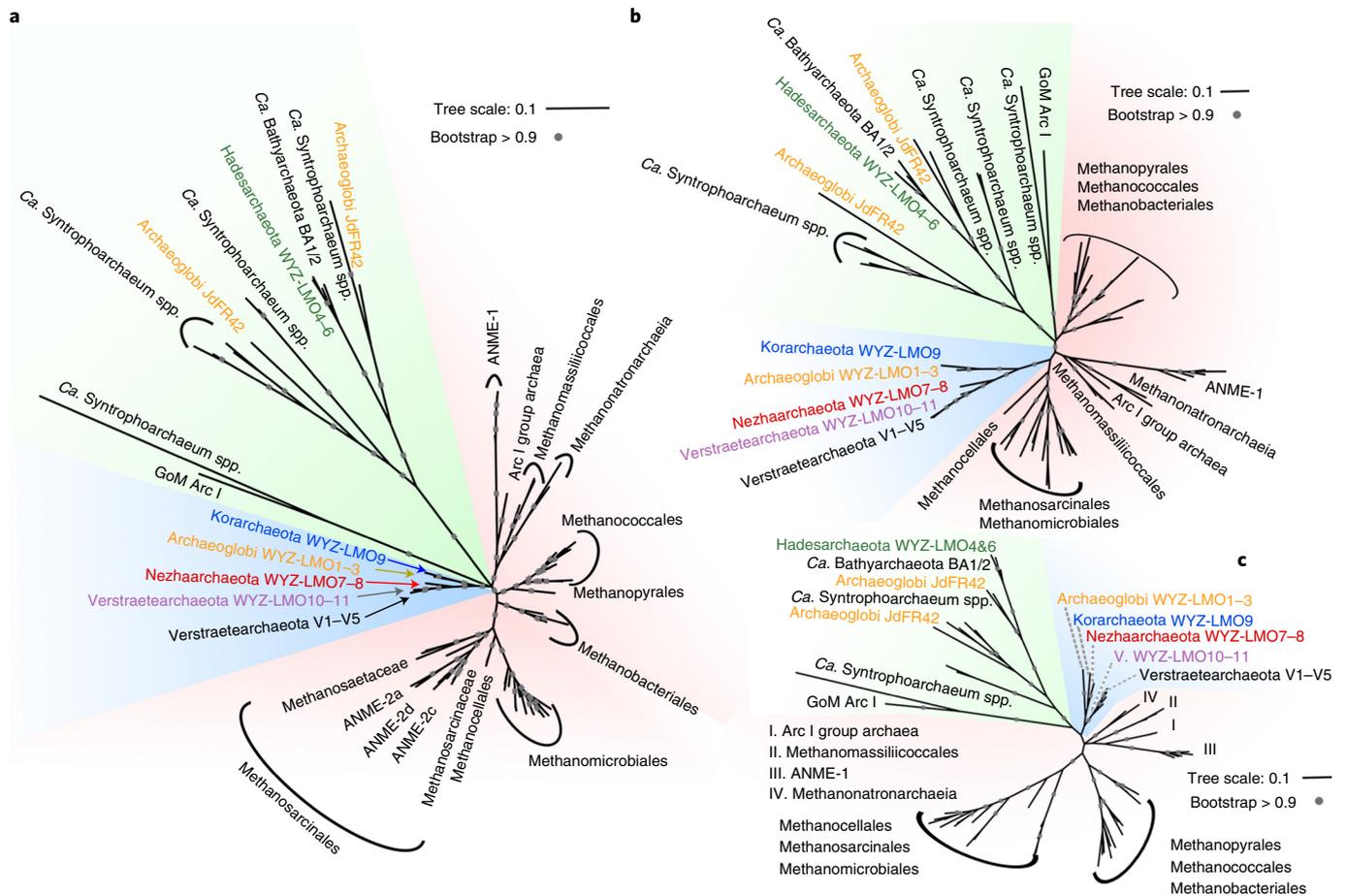


Fig. 2 | Phylogenetic affiliations of the McrA, McrB and McrG protein sequences of the 12 studied archaeal MAGs. a–c. The phylogenetic tree was constructed based on the alignments of McrA with 461 aligned positions (**a**), McrB with 351 aligned positions (**b**) and McrG with 221 aligned positions (**c**). Alignments were generated using MAFFT and then filtered with trimAl, and the trees were built by the IQ-TREE method with the model LG + C60 + F + G with 1,000 bootstrap replicates. The tree branches were classified into three distinct MCR groups: canonical euryarchaeotal MCRs (group I, red shaded background), the MCRs of group II dominated by sequences of the TACK superphylum archaea but including the Archaeoglobi WYZ-LMO1–3 Mcr sequences (group II, blue shaded background) and the putative multi-carbon alkane oxidization MCRs (group III, green shaded background) that seem to be laterally transferred between different Archaeoglobi, Hadesarchaeota, Bathyarchaeota and *Ca. Syntrophoarchaeum* spp. Previously unidentified MCR protein sequences are also coloured (the Archaeoglobi WYZ-LMO1–3/JdFR-42, orange; Hadesarchaeota WYZ4–6, green; Nezhaarchaeota WYZ-LMO7 and WYZ-LMO8, red; Korarchaeota WYZ-LMO9, blue; Verstraetearchaeota WYZ-LMO10 and WYZ-LMO11, purple).

Strikingly, we found an Archaeoglobi MAG that contains two highly divergent group III *mcr* gene sets in the NCBI genome database (Archaeoglobi JdFR-42; Fig. 3c and Supplementary Table 8). Both *mcr* gene sets are similar to those found in *Ca. Syntrophoarchaeum* spp. In *Ca. Syntrophoarchaeum* spp., the encoded MCR complexes have been shown to activate the multi-carbon alkanes propane and *n*-butane¹⁴. Moreover, MAG JdFR-42 encodes a Wood–Ljungdahl pathway and a pathway for fatty acid oxidation, including acyl-CoA dehydrogenase (FadE), crotonase (FadB1), 3-hydroxyacyl-CoA dehydrogenase (FadB2) and acetyl-CoA acetyltransferase (AtoB). These pathways are similar to those found in *Ca. Syntrophoarchaeum* spp. Hence, we suggest that Archaeoglobi JdFR-42 is another multi-carbon alkane degrader. Similar to *Ca. Syntrophoarchaeum* spp., MAG JdFR-42 does not carry genes encoding enzymes for sulfate reduction, such as *dsrAB* and *aprAB*. By contrast, its genome contains genes encoding nitrate reductase/nitrite oxidoreductase (NarGHI), periplasmic nitrate reductase (NapA), nitrate/nitrite transporter (NrtP) and the nitrogenase iron protein NifH similar to ANME-2d, thus might use nitrate/nitrite as electron sinks. Alternatively, Archaeoglobi JdFR-42 may have electron-accepting partner bacteria as shown for *Ca. Syntrophoarchaeum* spp.

The Hadesarchaeota MAGs WYZ-LMO4, WYZ-LMO5 and WYZ-LMO6 (Fig. 3d and Supplementary Tables 8 and 9) contain group III *mcr* genes that cluster with those of the *Ca. Syntrophoarchaeum* spp./Bathyarchaeota group. All three MAGs have β -oxidation pathways and they all contain the Wood–Ljungdahl pathway, while lacking genes encoding sulfate or nitrate reduction. The published MAGs of Hadesarchaeota (YNP_N21, YNP_45, DG-33 and DG-33-1) do not contain *mcr* genes and thus have been suggested to thrive as carbon monoxide and hydrogen oxidizers that can fix CO₂ in subsurface environments²⁰. However, the gene repertoire of the Hadesarchaeota obtained in this study suggests that they may oxidize short-chain alkanes similar to *Ca. Syntrophoarchaeum* spp.¹⁴.

The Nezhaarchaeota MAGs WYZ-LMO7 and WYZ-LMO8 (Fig. 3e and Supplementary Tables 8 and 9) contain *mcr* genes of group II MCR. These archaea also encode the complete methanogenesis pathway and two hydrogenases, coenzyme F₄₂₀ hydrogenase (FrhABDG) and ferredoxin-dependent hydrogenase (ECH), which are considered to act in coenzyme F₄₂₀ and ferredoxin recycling, respectively. The MAGs WYZ-LMO7 and WYZ-LMO8 contain formate dehydrogenase-like genes next to a heterodisulfide reductase

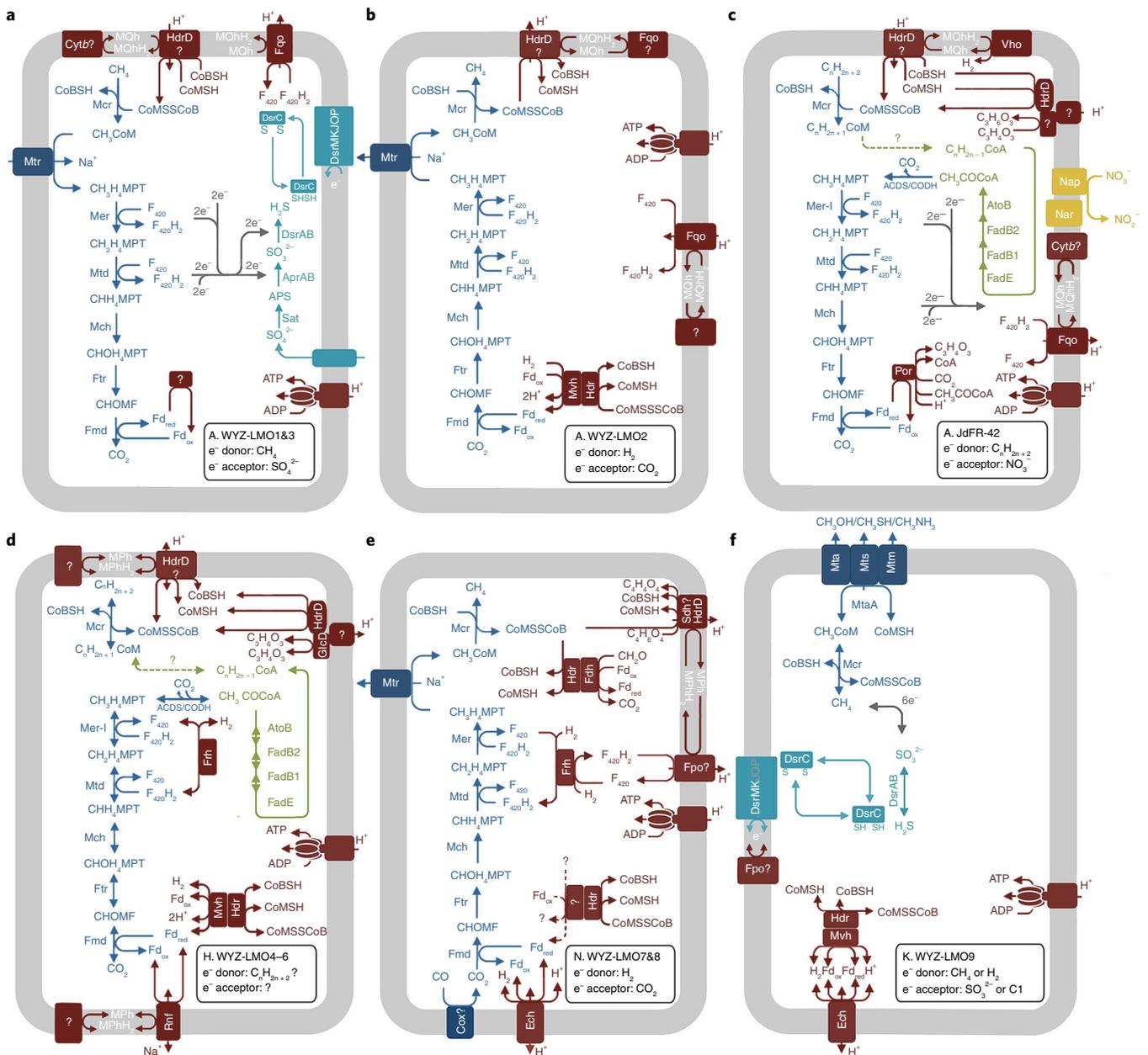


Fig. 3 | Alkane metabolic schemes of the studied MAGs with previously unidentified *mcr* genes. a–c, The MAGs demonstrate the large metabolic diversity of *mcr*-containing Archaeoglobi. WYZ-LMO1 (88.89% genome completeness) and WYZ-LMO3 (87.40% completeness) are probably methane oxidizers capable of sulfate reduction (a). WYZ-LMO2 (97.60% completeness), with a hydrogenase *mvh* gene cluster but no sulfate-reducing genes, is probably a methanogen or partner-bacterium-dependent methane oxidizer (b). JdFR-42 (99.84% completeness) contains the multi-carbon alkane oxidation group III *mcr* gene cluster, genes for fatty acid degradation and genes for nitrate/nitrite reduction; hence, it is probably a nitrate/nitrite- or partner-bacterium-dependent alkane oxidizer (c). **d,** The MAGs of Hadesarchaeota WYZ-LMO4–6 (88.89%, 83.26% and 92.37% completeness, respectively) encode sets of enzymes for multi-carbon alkane metabolism; the absence of reductive pathways suggests a need for partner bacteria. **e,** The Nezaarchaeota WYZ-LMO7 (93.90% completeness) and WYZ-LMO8 (97.20% completeness) contain the genes for hydrogenotrophic methanogenesis. **f,** The Korarchaeota WYZ-LMO9 has both *mcr* and *dsr* genes but lacks the Wood–Ljungdahl pathway, it might also oxidize methane with sulfite reduction; alternatively, it is probably a methylotrophic methanogen that uses electrons from sulfide or hydrogen oxidation. Question marks indicate unknown enzymes and substrates; dashed arrows indicate unknown reactions. Dark blue reactions are those involved in hydrocarbon metabolism; light blue reactions are those involved in dissimilatory sulfate reduction; red reactions are those involved in energy conservation; green reactions are those involved in β -oxidation; and yellow reactions are those involved in dissimilatory nitrate reduction. ACDS, acetyl-CoA decarboxylase/synthase; CODH, carbon-monoxide dehydrogenase; Cox, carbon monoxide dehydrogenase; Cytb, cytochrome b reductase; CHOMF, formylmethanofuran; Fmd, formylmethanofuran dehydrogenase; Fpo, $F_{420}H_2$ dehydrogenase; Fqo, $F_{420}H_2$ quinone oxidoreductase; Ftr, formylmethanofuran-tetrahydromethanopterin N-formyltransferase; GlcD, glycolate oxidase; Mch, methenyltetrahydromethanopterin cyclohydrolase; Mer, 5,10-methylenetetrahydromethanopterin reductase; Mer-I, 5,10-methylenetetrahydromethanopterin reductase-like enzyme; Mtd, methylenetetrahydromethanopterin dehydrogenase; NarDHI, nitrate reductase/nitrite oxidoreductase; Por, pyruvate ferredoxin oxidoreductase; Rnf, Na^+ -translocating ferredoxin:NAD⁺ oxidoreductase; Sat, sulfate adenylyltransferase; Vho, methanophenazine-dependent hydrogenase; APS, adenosine phosphosulfate; H_4 MPT, tetrahydromethanopterin; CoMSSCoB, heterodisulfide coenzyme B coenzyme M; Fd_{ox} , oxidized ferredoxin; Fd_{red} , reduced type ferredoxin; MPH, methanophenazine; MQH, menaquinone; MPT, methanopterin.

(HdrABC) with Fe-S-cluster-containing components. Moreover, two *hdrD* genes are located near succinate dehydrogenase-like genes and FAD/FMN-containing dehydrogenase genes; together, these enzymes could be used for the reduction of the heterodisulfide coenzyme B coenzyme M (CoBS–SCoM) to reduced coenzyme M and reduced coenzyme B. Thus, these Nezaarchaeota are most likely hydrogenotrophic methanogens.

The Korarchaeota MAG WYZ-LMO9 (Fig. 3f and Supplementary Tables 8 and 9) also contains group II *mcr* genes, as well as the sulfite reductase (*dsrAB*) genes that are related to those of the Firmicutes (Supplementary Fig. 5). In case the encoded DSR enzyme complex is used in an oxidative direction, electrons liberated in the transformation of sulfide to sulfite could be used for the production of methane from methyl groups. Conversely, a reductive use of the DSR might also be likely, as it would accept the electron from methane consumption. MAG WYZ-LMO9 lacks the genes for the upstream part of the methanogenesis pathway, but encodes methyltransferases, such as (methyl-Co(III) methanol-specific corrinoid protein):CoM methyltransferase (MtaA), methylthiol:CoM methyltransferase (MtsA) and methylamine-corrinoid protein co-methyltransferase (MtmB), suggesting that it may use methylated compounds for methane production or vice versa. MAG WYZ-LMO9 also contains genes encoding the HDR complex for the regeneration of free CoMSH and CoBSH, and encodes four potential sets of hydrogenases (MvhADG, FrhDG and two putative Ni-Fe hydrogenase: NiFe-group-1g and NiFe-group-4b).

The Verstraetearchaeota MAGs WYZ-LMO10 and WYZ-LMO11 (Supplementary Tables 8 and 9) also contain *mcr* genes of the group II MCR. They miss the upstream part of the methanogenesis pathway, but contain genes encoding MtaA, MtsA and MtmB, indicating that they are likewise methylotrophic methanogens. MAGs WYZ-LMO10 and WYZ-LMO11 also contain genes encoding the HdrDE complex for the regeneration of free CoMSH and CoBSH, as well as the ECH complex and energy-converting hydrogenase (EHB) for ferredoxin recycling. The combination of these pathways may allow the Verstraetearchaeota to thrive as methylotrophic methanogens, yet they would require an electron donor, presumably molecular hydrogen as suggested before¹³.

The genes encoding MCR-based alkane metabolism have now been identified in the Euryarchaeota, Hadesarchaeota and different clades of the TACK superphylum. Our phylogenetic analysis of all three subunits of the MCR complex showed that the canonical Euryarchaeota Mcr sequences (group I MCR) form at least three distinct monophyletic clusters (euryarchaeotal-cluster I includes the Methanobacteriales, Methanococcales and Methanopyrales; euryarchaeotal-cluster II includes the Methanocellales, Methanomicrobiales and Methanosarcinales; euryarchaeotal-cluster III includes the Methanomassiliococcales, Arc I group (Methanofastidiosales), Methanonatronarchaeia and ANME-1). The MCR in group II forms a well-supported monophyletic cluster next to group I and contains MCR sequences from the TACK superphylum archaea, yet includes the MCRs of the Archaeoglobi MAGs WYZ-LMO1–3. Besides this exception, the phylogeny of both the group I and group II MCRs are consistent with the topologies from phylogenomic and 16S rRNA gene phylogenetic analyses (Figs. 1 and 2). Two scenarios could explain the presence of the euryarchaeotal Archaeoglobi Mcr sequences in group II MCR. One possibility is that a common ancestor of the Euryarchaeota and TACK archaea already contained a *mcr* gene cluster that was vertically transferred into the two emerging superphyla, which since then would have evolved independently. In this scenario, archaea from the Archaeoglobi might have obtained the group II *mcr* genes by horizontal gene transfer from a TACK archaeon. In a second scenario, one ancestor of the TACK archaea might also have acquired its MCR after the divergence of the two superphyla, possibly from the *mcr*-containing Archaeoglobi lineage.

The group III MCRs are present in different phyla, including the Euryarchaeota, Hadesarchaeota and Bathyarchaeota. The wide phylogenetic distribution and the low number of available MAGs with this MCR type may allow limited predictions on its origin. Nevertheless, the McrA and McrG proteins of the group III MCR also show close affiliations with the group II MCR, as well as with euryarchaeotal-cluster III MCR in group I (Fig. 2a,c). One possible explanation for the distribution of group III *mcr* genes in different archaeal phyla is horizontal gene transfer, which allowed several formerly most likely methanogenic or methane-oxidizing archaea to become multi-carbon alkane oxidizers. The high similarities between MCR and genome GC contents, tetranucleotide frequencies and codon usages (Supplementary Figs. 1–3) show that this horizontal gene transfer should have appeared early in the evolution of these organisms. Alternatively, owing to positive natural selection, several methanogens or methane oxidizers might have developed multi-carbon-type MCRs by convergent evolution of canonical MCRs. Genomic information from additional *mcr*-containing organisms should be required to further evaluate whether the wide distribution of *mcr* genes in the domain of Archaea is due to mechanisms of horizontal or vertical gene transfer and to identify the most ancestral *mcr*-containing archaeon.

Our study provides a global-scale metagenome-based analysis of the diversity of *mcr*-containing microorganisms. Particularly from geothermal environments, we retrieved *mcr*-containing MAGs of representatives affiliated with the Archaeoglobi in the Euryarchaeota, Hadesarchaeota, Korarchaeota, Nezaarchaeota and MAGs of recently described Verstraetearchaeota that probably perform (reverse) methanogenesis and multi-carbon alkane oxidation, and we provide genomic evidence for the coexistence of methane metabolism and sulfate reduction in single MAGs of Archaeoglobi and Korarchaeota. This suggests a greater importance of archaea in global carbon balancing. However, *in vitro* cultivation of representatives of the here described organisms and additional genomes of alkane-metabolizing archaea will be required to validate the metabolic hypotheses suggested in this study.

Methods

Data selection and treatments of raw sequencing reads. Predicted protein sequences from high-throughput metagenomes were downloaded from the JGI database (<https://genome.jgi.doe.gov>; January 2017). To identify metagenomic data sets containing MCR-based alkane metabolism-related genes, the metagenomic protein sequences were queried against a local McrA protein database ($n = 153$; Supplementary Table 1) containing sequences from the known methanogens, ANMEs and multi-carbon alkane-oxidizing archaea (sequences were retrieved from the NCBI database: <https://www.ncbi.nlm.nih.gov>), using DIAMOND³¹ version 0.8.28.90 (identity > 30%, coverage > 75%, $e < 1 \times 10^{-20}$). This screening identified 64 *mcrA* gene-containing data sets (Supplementary Table 2), for which the metagenomic raw sequencing reads data were downloaded from the NCBI Sequence Read Archive public database (<https://www.ncbi.nlm.nih.gov/sra/>). Raw sequencing reads were trimmed using the Sickle algorithm version 1.33 (<https://github.com/najoshi/sickle>), and trimmed reads were assembled using MEGAHIT³² version 1.0.6-hotfix1 with k-min 27 and k-max 127. The sequence coverage of each contig was determined by mapping the original trimmed reads to the contigs using Bowtie³³ version 2.2.8 with parameter --very-sensitive. The assembled data were subjected to open reading frame (ORF) prediction with Prodigal³⁴ version 2.6.3 with parameter -meta. The predicted ORFs were searched against the NCBI nr protein database (March 2017) and the web-based KEGG portal GhostKOALA (metagenome annotation) and KEGG³⁵ mapping for primary annotation.

Binning of genomes and evaluations. The assembled metagenomic sequences were binned based on the tetranucleotide frequency dimensionality reduction by ESOM³⁶, based on abundance information using MaxBin³⁷ version 2.2.4 with the run_MaxBin.pl script and based on abundance and tetranucleotide frequency using MetaBAT³⁸ version 2.12.1 with 1 kb (or 1.5 kb) and 3 kb as contig length cut-offs. The MAGs with the highest completeness and lowest contamination as evaluated using CheckM³⁹ version 1.0.7 using lineage-specific marker genes with parameter lineage_wf were considered for further analyses. The selected MAGs were then refined with the mmgenome⁴⁰ package version 0.6.3. Contigs that did not match the 95% confidence interval for depth and GC contents as calculated using the R function 'quantile' were removed. In addition, the annotation results

of each contig were checked, and potential contamination (defined as the entire contig was annotated as another microorganism with protein sequence identities higher than 80%) was manually removed as suggested by Bowers et al.⁴¹. The correct assignment of the *mcr* and *dsr* containing contigs to a certain MAG was additionally evaluated by comparing the GC contents, tetranucleotide frequencies and codon usages of the *mcr*- and *dsr*-containing contigs with those of their complete MAGs and genomes of methanogens, as well as closely related archaea. In detail, for selecting the most related methanogen genomes, we compared the McrA protein sequences from each of the MAGs to the NCBI refseq database to determine the highest-scoring McrA hits and downloaded the genome of the respective organisms, namely, the Archaeoglobi WYZ-LMO1–3 with *Methanocaldococcus bathoardescens* JH146, the Hadesarchaeota WYZ-LMO4–6 with Bathyarchaeota archaeon BA2, the Nezharchaeota WYZ-LMO7 and WYZ-LMO8 with *Methanopyrus* sp. KOL6, the Korarchaeota WYZ-LMO9 with *Methanocaldococcus vulcanius* M7 and the Verstraetearchaeota WYZ-LMO10 and WYZ-LMO11 with *Methanosuratus petracarbonis* V4. For selecting genomes of archaea related to the MAGs in the current study, the genomes that closely clustered to the MAG based on the phylogenomic analysis were considered, that is, the Archaeoglobi WYZ-LMO1–3 with *A. fulgidus* DSM 4304, the Hadesarchaeota WYZ-LMO4–6 with Hadesarchaeota archaeon YNP_45, the Nezharchaeota WYZ-LMO7 and WYZ-LMO8 with *M. petracarbonis* V4, the Korarchaeota WYZ-LMO9 with *Korarchaeum cryptofilum* OPF8, the Verstraetearchaeota WYZ-LMO10 and WYZ-LMO11 with *M. petracarbonis* V4. The average GC contents were calculated using the sum of cytosine and thymine divided by the sum of cytosine, thymine, adenine and guanine of contigs or genomes. Tetranucleotide frequencies were calculated by the Perl script tetramer_freqs_esom.pl³⁶. The tetranucleotide frequencies of all *mcr*- and *dsr*-containing contigs, the studied MAGs, their related archaea and their related methanogens were displayed by Principal Component Analysis (PCA) for dimensionality reduction as calculated by R function 'prcomp'. Codon usage frequencies were calculated at http://www.bioinformatics.org/sms2/codon_usage.html and the codon usage distances were shown as the difference of square values between each *mcr*- or *dsr*-containing contig and its host MAG, and also genomes of their related methanogens and archaea. Completeness and contamination of each MAG was assessed with CheckM³⁹ version 1.0.7 using lineage-specific (149–228) and archaea (149) marker genes. ORFs of these MAGs were predicted with Prodigal⁴¹ version 2.6.3. The predicted ORFs were searched against the NCBI nr protein database (March 2017) and eggNOG⁴² database with the BLASTP algorithm (coverage > 75% and $e < 1 \times 10^{-20}$) to check their protein identities to the most closely related sequences using DIAMOND version 0.8.28.90. Specifically, the NCBI nr protein database and eggNOG database were downloaded and local protein databases were constructed. 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>). The HydDB was first downloaded and a local database was established. Then, all predicted protein sequences from the MAGs were queried against the local HydDB with parameters coverage > 75% and $e < 1 \times 10^{-20}$. Next, the result sequences were searched again in the NCBI nr database for further confirmations. For metabolic pathway analyses, we used the web portal GhostKOALA on the KEGG⁴³ website. The AAI values were generated by compareM version 0.0.23 with parameters aai_wf and --proteins (<https://github.com/dparks1134/CompareM>). The taxonomic assignments of MAGs were based on criteria provided in references^{44–47}.

Phylogenetic analyses based on conserved proteins, 16S rRNA genes and McrABG/DsrA/AprA subunits. For phylogenomic analysis, 233 representative archaea reference genomes from the superphyla Euryarchaeota, TACK, Asgard and DPANN were downloaded from the NCBI prokaryote genome database (<https://www.ncbi.nlm.nih.gov/assembly/>); the genome list can be found in <https://figshare.com/>, <https://doi.org/10.6084/m9.figshare.7149623>. 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 38 marker genes as suggested by Hug et al.⁴⁸ (Supplementary Table 5). Nevertheless, the removal of the protein sequence translation elongation factor GTPase affects the placement of the Hadesarchaeota in the phylogenomic tree; thus, the sequence of the translation elongation factor GTPase was not considered in further phylogenomic analysis. Specifically, each of the 37 marker protein sequences from the reference genomes and the MAGs was aligned using the MAFFT⁴⁹ algorithm version 7.313 with parameters --ep 0 --genafpair --maxiterate 1,000 and filtered with trimAl⁵⁰ version 1.4.rev2 with parameter -automated1. Then, all 37 marker genes were concatenated into a single alignment and phylogenetic trees were built using both IQ-Tree⁵¹ version 1.6.6 with the model LG + C60 + F + G and RAXML⁵² version 8.0 with the model PROTGAMMAAUTO with a bootstrap value of 1,000. For the phylogenetic analysis of functional marker proteins (McrA, McrB, McrG, DsrA and AprA) and the 16S rRNA gene, the respective protein and nucleotide sequences were retrieved from the MAGs, and additional reference sequences were obtained from the NCBI (<https://www.ncbi.nlm.nih.gov/protein/>). Alignment and filtering was carried out with the same programs described above. For 16S rRNA gene sequences, phylogenetic trees were built using both IQ-Tree⁵¹ version 1.6.6 with the model

HKY + F + G and RAXML⁵² version 8.0 with the model GTRGAMMA with a bootstrap value of 1,000.

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

Code availability

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 1.33: <https://github.com/najoshi/sickle>, MEGAHIT version 1.0.6-hotfix1: <https://hku-bal.github.io/megabox/>, Bowtie version 2.2.8: <http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>, Prodigal version 2.6.3: <http://compbio.ornl.gov/prodigal/>, MaxBin version 2.2.4: <http://sourceforge.net/projects/maxbin/>, MetaBAT version 2.12.1: <https://bitbucket.org/berkeleylab/metabat>, CheckM version 1.0.7: <http://ecogenomics.github.io/CheckM>, compareM version 0.0.23: <https://github.com/dparks1134/CompareM>, MAFFT version 7.313: <https://mafft.cbrc.jp/alignment/software/>, trimAl version 1.4.rev2: <http://trimal.cgenomics.org>, IQ-Tree version 1.6.6: <http://www.cibiv.at/software/iqtree>, and RAXML version 8.0: <https://github.com/stamatak/standard-RAXML>.

Data availability

The data sets generated and/or analysed during the current study are available in the NCBI repository at <https://www.ncbi.nlm.nih.gov/>. The MAGs from the current study have been deposited in the NCBI GenBank under the project ID PRJNA475886.

Received: 19 June 2018; Accepted: 7 January 2019;

Published online: 04 March 2019

References

- Ueno, Y., Yamada, K., Yoshida, N., Maruyama, S. & Isozaki, Y. Evidence from fluid inclusions for microbial methanogenesis in the early Archaean era. *Nature* **440**, 516–519 (2006).
- Conrad, R. The global methane cycle: recent advances in understanding the microbial processes involved. *Environ. Microbiol. Rep.* **1**, 285–292 (2009).
- Reeburgh, W. Oceanic methane biogeochemistry. *Chem. Rev.* **107**, 486–513 (2007).
- Knittel, K. & Boetius, A. Anaerobic oxidation of methane: progress with an unknown process. *Annu. Rev. Microbiol.* **63**, 311–334 (2009).
- Krüger, M. et al. A conspicuous nickel protein in microbial mats that oxidize methane anaerobically. *Nature* **426**, 878–881 (2003).
- Hallam, S. J. et al. Reverse methanogenesis: testing the hypothesis with environmental genomics. *Science* **305**, 1457–1462 (2004).
- Scheller, S., Goenrich, M., Boecher, R., Thauer, R. K. & Jaun, B. The key nickel enzyme of methanogenesis catalyses the anaerobic oxidation of methane. *Nature* **465**, 606–608 (2010).
- Gunsalus, R. & Wolfe, R. Methyl coenzyme M reductase from *Methanobacterium thermoautotrophicum*. Resolution and properties of the components. *J. Biol. Chem.* **255**, 1891–1895 (1980).
- Ermler, U., Grabarse, W., Shima, S., Goubeaud, M. & Thauer, R. K. Crystal structure of methyl-coenzyme M reductase: the key enzyme of biological methane formation. *Science* **278**, 1457–1462 (1997).
- Thauer, R. K. Anaerobic oxidation of methane with sulfate: on the reversibility of the reactions that are catalyzed by enzymes also involved in methanogenesis from CO₂. *Curr. Opin. Microbiol.* **14**, 292–299 (2011).
- Lever, M. A. & Teske, A. P. Diversity of methane-cycling archaea in hydrothermal sediment investigated by general and group-specific PCR primers. *Appl. Environ. Microbiol.* **81**, 1426–1441 (2015).
- Evans, P. N. et al. Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. *Science* **350**, 434–438 (2015).
- Vanwongterghem, I. et al. Methylothermophilic methanogenesis discovered in the archaeal phylum Verstraetearchaeota. *Nat. Microbiol.* **1**, 16170 (2016).
- Laso-Pérez, R. et al. Thermophilic archaea activate butane via alkyl-coenzyme M formation. *Nature* **539**, 396–401 (2016).
- Boetius, A. et al. A marine microbial consortium apparently mediating anaerobic oxidation of methane. *Nature* **407**, 623–626 (2000).
- McGlynn, S. E., Chadwick, G. L., Kempes, C. P. & Orphan, V. J. Single cell activity reveals direct electron transfer in methanotrophic consortia. *Nature* **526**, 531–535 (2015).
- Wegener, G., Krukenberg, V., Riedel, D., Tegetmeyer, H. E. & Boetius, A. Intercellular wiring enables electron transfer between methanotrophic archaea and bacteria. *Nature* **526**, 587–590 (2015).
- Haroon, M. F. et al. Anaerobic oxidation of methane coupled to nitrate reduction in a novel archaeal lineage. *Nature* **500**, 567–570 (2013).

19. Ettwig, K. F. et al. Archaea catalyze iron-dependent anaerobic oxidation of methane. *Proc. Natl Acad. Sci. USA* **113**, 12792–12796 (2016).
20. Baker, B. J. et al. Genomic inference of the metabolism of cosmopolitan subsurface Archaea, Hadesarchaea. *Nat. Microbiol.* **1**, 16002 (2016).
21. Elkins, J. G. et al. A korarchaeal genome reveals insights into the evolution of the Archaea. *Proc. Natl Acad. Sci. USA* **105**, 8102–8107 (2008).
22. Adam, P. S., Borrel, G., Brochier-Armanet, C. & Gribaldo, S. The growing tree of Archaea: new perspectives on their diversity, evolution and ecology. *ISME J.* **11**, 2407–2425 (2017).
23. Parks, D. H. et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* **36**, 996–1004 (2018).
24. Raymann, K., Brochier-Armanet, C. & Gribaldo, S. The two-domain tree of life is linked to a new root for the Archaea. *Proc. Natl Acad. Sci. USA* **112**, 6670–6675 (2015).
25. Petitjean, C., Deschamps, P., López-García, P., Moreira, D. & Brochier-Armanet, C. Extending the conserved phylogenetic core of archaea disentangles the evolution of the third domain of life. *Mol. Biol. Evol.* **32**, 1242–1254 (2015).
26. Williams, T. A. et al. Integrative modeling of gene and genome evolution roots the archaeal tree of life. *Proc. Natl Acad. Sci. USA* **114**, E4602–E4611 (2017).
27. Stahl, D. A., Fishbain, S., Klein, M., Baker, B. J. & Wagner, M. Origins and diversification of sulfate-respiring microorganisms. *Antonie Van Leeuwenhoek* **81**, 189–195 (2002).
28. Müller, A. L., Kjeldsen, K. U., Rattai, T., Pester, M. & Loy, A. Phylogenetic and environmental diversity of DsrAB-type dissimilatory (bi) sulfite reductases. *ISME J.* **9**, 1152–1165 (2015).
29. Timmers, P. H. et al. Reverse methanogenesis and respiration in methanotrophic archaea. *Archaea* **2017**, 1654237 (2017).
30. Stojanovic, A., Mander, G. J., Duin, E. C. & Hedderich, R. Physiological role of the F420-non-reducing hydrogenase (Mvh) from *Methanothermobacter marburgensis*. *Arch. Microbiol.* **180**, 194–203 (2003).
31. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2014).
32. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
33. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
34. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
35. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361 (2016).
36. Dick, G. J. et al. Community-wide analysis of microbial genome sequence signatures. *Genome Biol.* **10**, R85 (2009).
37. Wu, Y.-W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605–607 (2015).
38. 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).
39. 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).
40. Karst, S. M., Kirkegaard, R. H. & Albertsen, M. mmgenome: a toolbox for reproducible genome extraction from metagenomes. Preprint at <https://doi.org/10.1101/059121> (2016).
41. Bowers, R. M. et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* **35**, 725–731 (2017).
42. Huerta-Cepas, J. et al. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* **44**, D286–D293 (2016).
43. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **428**, 726–731 (2016).
44. Yarza, P. et al. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat. Rev. Microbiol.* **12**, 635–645 (2014).
45. Hugenholtz, P., Skarshewski, A. & Parks, D. H. Genome-based microbial taxonomy coming of age. *Cold Spring Harb. Perspect. Biol.* **8**, a018085 (2016).
46. Konstantinidis, K. T., Rosselló-Móra, R. & Amann, R. Uncultivated microbes in need of their own taxonomy. *ISME J.* **11**, 2399–2406 (2017).
47. Chuvpochina, M. et al. The importance of designating type material for uncultured taxa. *Syst. Appl. Microbiol.* <https://doi.org/10.1016/j.syapm.2018.07.003> (2018).
48. Hug, L. A. et al. A new view of the tree of life. *Nat. Microbiol.* **1**, 16048 (2016).
49. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
50. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
51. Stamatakis, A. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
52. 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 (2014).
53. Jungbluth, S. P., Amend, J. P. & Rappé, M. S. Metagenome sequencing and 98 microbial genomes from Juan de Fuca Ridge flank subsurface fluids. *Sci. Data* **4**, 170037 (2017).

Acknowledgements

We thank R. K. Thauer for his advice on the discussed metabolic pathways, and V. Krukenberg and J. Wang for valuable discussion and suggestions for the manuscript. We are grateful to the researchers who published their sequence data on the NCBI (<https://www.ncbi.nlm.nih.gov/>), and to the US Department of Energy Joint Genome Institute (<http://www.jgi.doe.gov/>) for providing protein sequence files in collaboration with the user community. We thank the following sources for funding: the Natural Science Foundation of China (grant numbers 91751205, 41525011 and 91428308), the National Key R&D project of China (grant number 2018YFC0309800) and China Postdoctoral Science Foundation Grant (grant number 2018T110390). This study is also a contribution to the Deep Carbon Observatory.

Author contributions

Y.W. and F.W. designed the research, performed the analyses, developed the metabolic models and wrote the paper. G.W. developed the metabolic models and wrote the paper. J.H. provided useful discussion and helped with the double-blind assessments of the MAGs. F.W. and X.X. provided guidance and useful suggestion.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41564-019-0364-2>.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to F.W. or X.X.

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

© The Author(s), under exclusive licence to Springer Nature Limited 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated
- Clearly defined error bars
State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on [statistics for biologists](#) may be useful.

Software and code

Policy information about [availability of computer code](#)

Data collection

No software was used to collect data.

Data analysis

All scripts and softwares to perform metagenome assembly, annotation and binning can be accessed from github (<https://github.com/>) or are listed in methods with versions and can be downloaded from the websites provided in the publications in the references part in the manuscript. 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 1.33: <https://github.com/najoshi/sickle>, MEGAHIT version 1.0.6-hotfix1: <https://hku-bal.github.io/megabox/>, Bowtie version 2.2.8: <http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>, Prodigal version 2.6.3: <http://compbio.ornl.gov/prodigal/>, MaxBin version 2.2.4: <http://sourceforge.net/projects/maxbin/>, MetaBAT version 2.12.1: <https://bitbucket.org/berkeleylab/metabat>, CheckM version 1.0.7: <http://ecogenomics.github.io/CheckM>, compareM version 0.0.23: <https://github.com/dparks1134/CompareM>, MAFFT version 7.313: <https://mafft.cbrc.jp/alignment/software/>, trimAl version 1.4.rev2: <http://trimal.cgenomics.org>, IQ-Tree version 1.6.6: <http://www.cibiv.at/software/iqtree>, RAxML version 8.0: <https://github.com/stamatak/standard-RAxML>.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The datasets generated and/or analysed during the current study are available in the NCBI repository at <https://www.ncbi.nlm.nih.gov/>. The accession numbers are listed in Supplementary table 2. The genomic bins from the current study have been deposited in NCBI GenBank under the project ID PRJNA475886.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/authors/policies/ReportingSummary-flat.pdf](https://www.nature.com/authors/policies/ReportingSummary-flat.pdf)

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We show that methyl-CoM reductase (MCR) based alkane metabolism is present in a wide array of phyla of the domain Archaea through a global survey of MCR-encoding genes in public metagenomic data from various environments. Hydrothermally heated environments were found to contain mostly novel types of mcr genes, and eleven high-quality mcr-containing genomic bins were obtained belonging to a variety of archaeal phyla, including the Archaeoglobi in the Euryarchaeota, the Hadesarchaeota, and different TACK superphylum archaea, including the novel Nezaarchaeota, Korarchaeota and recently described Verstraetearchaeota. The Archaeoglobi here and a novel strain of Korarchaeota combine the genes for (reverse) methanogenesis and reductive sulfur metabolism, suggesting a coupling of methane oxidation with sulfate reduction in individual organisms.
Research sample	The data used in the current study are existing shotgun metagenome sequence data from natural environments in public databases, including marine sediment, hot springs, freshwater niches. All metagenomes raw data analyzed here are available in the NCBI repository at https://www.ncbi.nlm.nih.gov/ . The accession numbers of the datasets are listed in Supplementary table 2. Specifically, the geothermal environment metagenomic data sets were assembled and binned. Previously unidentified alkane metabolism archaea were found within these data sets.
Sampling strategy	In the current study, sampling strategy is not applicable because we mainly focused on discovery of potential alkane metabolism archaea in natural environments and we screened all JGI metagenomic protein database while only took the samples with McrA protein sequences. We then selected the mcrA gene containing metagenomic datasets from public available NCBI SRA database for further analyses.
Data collection	The first author Yinzhao Wang collected all datasets from the NCBI SRA repository and did the analyses.
Timing and spatial scale	The data collection started/ended at Jan. 2017.
Data exclusions	No data were excluded from the analyses.
Reproducibility	No experiment was carried out in the current study. The reproducibility can be conducted by following the methods for shotgun metagenome sequence data as mentioned in the main text.
Randomization	No randomization was made because no statistical analyses were performed here. We mainly focused on discovery of potential alkane metabolism archaea in natural environments.
Blinding	Blinding was not relevant to this study, given that it primarily involved the analyses of metagenomic datasets.
Did the study involve field work?	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No

Reporting for specific materials, systems and methods

Materials & experimental systems

- | n/a | Involvement in the study |
|-------------------------------------|--|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Unique biological materials |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Human research participants |

Methods

- | n/a | Involvement in the study |
|-------------------------------------|---|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |