Wide diversity of methane and short-chain alkane metabolisms in uncultured archaea

Guillaume Borrel*, Panagiotis S. Adam, Luke J. McKay, Lin-Xing Chen, Isabel Natalia Sierra-García, Christian M. K. Sieber, Quentin Letourneur, Amine Ghozlane, Gary L. Andersen, Wen-Jun Li, Steven J. Hallam, Gerard Muyzer, Valéria Maia de Oliveira, William P. Inskeep, Jillian F. Banfield and Simonetta Gribaldo*

Methanogenesis is an ancient metabolism of key ecological relevance, with direct impact on the evolution of Earth’s climate. Recent results suggest that the diversity of methane metabolisms and their derivations have probably been vastly underestimated. Here, by probing thousands of publicly available metagenomes for homologues of methyl-coenzyme M reductase complex (MCR), we have obtained ten metagenome-assembled genomes (MAGs) belonging to potential methanogenic, anaerobic methanotrophic and short-chain alkane-oxidizing archaea. Five of these MAGs represent under-sampled (Verstraetearchaeota, Methanonatronarchaeia, ANME-1 and GoM-Arc1) or previously genomically undescribed (ANME-2c) archaeal lineages. The remaining five MAGs correspond to lineages that are only distantly related to previously known methanogens and span the entire archaeal phylogeny. Comprehensive comparative annotation substantially expands the metabolic diversity and energy conservation systems of MCR-bearing archaea. It also suggests the potential existence of a yet uncharacterized type of methanogenesis linked to short-chain alkane/fatty acid oxidation in a previously undescribed class of archaea (’Candidatus Methanoliparia’). We redefine a common core of marker genes specific to methanogenic, anaerobic methanotrophic and short-chain alkane-oxidizing archaea, and propose a possible scenario for the evolutionary and functional transitions that led to the emergence of such metabolic diversity.

Methanogenesis is an archaeal-specific metabolism of key relevance in the anaerobic degradation of organic matter and biogas production. It is considered one of the most ancient energetic metabolisms with direct impact on the evolution of the Earth’s climate system. Methanogens have been detected in virtually all types of anaerobic environment. Until recently, all methanogens were thought to belong to two meuryarchaeal clades, named Class I and Class II methanogens. The majority of Class I/II methanogens can grow by reducing CO₂ into methane using H₂ as an electron donor. Several representatives of the Methanosarcinales (Class II methanogens) use additional energetic substrates, including acetate and methylated compounds. Methanosphaera spp. (Class I methanogens) are restricted to the reduction of methanol with H₂ (ref. 1). Regardless of the encoded methanogenic pathway, all members of Class I/II methanogens possess the H₄MPT methyl-branch of the Wood–Ljungdahl pathway (m-WL), the N⁵-methyltetrahydromethanopterin–coenzyme M–methyltransferase complex (MtrABCDEFGH or MTR) and the methylcoenzyme M–reductase complex (McrABG or MCR)². The same enzymes are present in anaerobic methanotrophic archaea (ANME) and are used in reverse to oxidize methane³⁻⁴.

Our understanding of the diversity and metabolic versatility of methanogenic archaea is undergoing a rapid transformation with the availability of additional isolates and MAGs⁵⁻¹⁰. This has revealed additional lineages only distantly related to Class I/II methanogens, including Methanomassiliicoccales⁶, Methanofastidiosales⁷, Methanotransarchaeia⁸, Methanomonadarchaeota⁹ and Methanoproteobacteria. A striking characteristic of these recently described methanogens is the absence of the MTR complex, a partial or missing m-WL pathway, and the presence of specific methyltransferases for the use of methylated compounds. Accordingly, they are predicted to be limited to the reduction of methylated compounds with H₂ for methanogenesis, which was experimentally validated for Methanomassiliicoccales⁶ and Methanotransarchaeia⁸. More recently, the implication of a divergent McrABG-like complex in the oxidation of short-chain alkanes (butane, propane) has been demonstrated in two representatives of a recently described euryarchaeal order, the ‘Ca. Syntrophoarchaeales’¹². Divergent MCR sequences were also found in two members of the Batharchaeota (TACK superphylum), in the GoM-Arc1 (a lineage in the Methanosarcinales)¹³ and in environmental samples¹⁴. Altogether, this suggests that methanogens, anaerobic methanotrophs and short-chain alkane oxidizers may have an even wider phylogenetic and environmental distribution.

¹Department of Microbiology, Unit Evolutionary Biology of the Microbial Cell, Institut Pasteur, Paris, France. ²Université Paris Diderot, Sorbonne Paris Cité, Paris, France. ³Center for Biofilm Engineering, Montana State University, Bozeman, MT, USA. ⁴Department of Earth and Planetary Sciences, University of California, Berkeley, CA, USA. ⁵Microbial Resources Division, Research Center for Chemistry, Biology and Agriculture, Campinas University—UNICAMP, Campinas, Sao Paolo, Brazil. ⁶Department of Energy Joint Genome Institute, Walnut Creek, CA, USA. ⁷Bioinformatics and Biostatistics Hub, C3BI, Institut Pasteur, Paris, France. ⁸Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA, USA. ⁹State Key Laboratory of Biocontrol and Guangdong Provincial Key Laboratory of Plant Resources, School of Life Sciences, Sun Yat-Sen University, Guangzhou, China. ¹⁰Department of Microbiology & Immunology, ECOSCOPE Training Program, Graduate Program in Bioinformatics, and Genome Sciences and Technology Training Program, University of British Columbia, Vancouver, British Columbia, Canada. ¹¹Microbial Systems Ecology, Department of Freshwater and Marine Ecology, Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, The Netherlands. ¹²Department of Land Resources and Environmental Sciences, Montana State University, Bozeman, MT, USA. *e-mail: guillaume.borrel@pasteur.fr; simonetta.gribaldo@pasteur.fr
than previously anticipated, provoking new questions on the diversity and evolution of these metabolisms.

**Results and discussion**

**Additional lineages of archaea with an MCR or MCR-like complex.** To identify previously undescribed lineages of potential methanogens, anaerobic methanotrophs and short-chain alkane oxidizers, we probed available metagenomes from the JGI/IMG database for McrA homologues and identified sequences distantly related to well characterized lineages (Methods). Ten MAGs were reconstructed from the corresponding metagenomes sourced from a wide range of anoxic environments including an inland petroleum reservoir from Brazil, oil seeps from the United States, soda lake sediments from Russia and hot-springs from China and the United States (Table 1). Nine of the ten MAGs had an estimated completeness ranging from 78.4 to 94.4%, and one was only 51.5% complete. Estimated contamination (without strain heterogeneity) ranged from 0 to 3.3%.

Four MAGs represent three previously undescribed lineages only distantly related to known methanogenic/methanotrophic archaea (Fig. 1): NM1 (NM1a and NM1b MAGs) branches in the Methanotecta superclass, between Archaeoglobales and the clade formed by ‘Ca. Syntrophoarchaeum’ and ‘Ca. Methanophagales’ (ANME-1); NM3 branches in the Acherontia superclass, at the base of the clade formed by the non-methanogenic ‘Ca. Theionarchaeum’ and ‘Ca. Methanofastidiosa’ (former WSA2/Arc1); NM4 branches in the TACK superphylum and is related to *Korarchaeum cryptofilum* (Fig. 1).

Table 1 | General information on the ten novel MAGs

| Genome   | Origin                       | Scaffolds (number) | Size (Mb) | Genes (number) | GC (%) | Compl. (%) | Cont. (%) | Strain hetero.* | Cont. excl. strain hetero. (%) |
|----------|------------------------------|--------------------|-----------|----------------|--------|------------|-----------|-----------------|-------------------------------|
| NM1a     | Enrichment culture (50 °C) from petroleum sample, Brazil | 12                 | 1.26      | 1,388          | 35.7   | 92.5       | 1.3       | 0               | 1.3                           |
| NM1b     | Santa Barbara Channel oil seeps, USA | 183                | 1.66      | 1,860          | 43.8   | 90.2       | 3.6       | 66.7            | 1.2                           |
| NM2      | Santa Barbara sediments, USA | 210                | 1.03      | 1,254          | 41.8   | 51.5       | 2         | 0               | 2                             |
| NM3      | Enrichment culture (40 °C) from petroleum sample, Brazil | 26                 | 1.49      | 1,578          | 55.2   | 85.5       | 2.9       | 0               | 2.9                           |
| NM4      | Yellowstone sulfidic hot spring, USA | 122                | 1.42      | 1,603          | 43.4   | 85.5       | 1.8       | 66.7            | 0.6                           |
| Verst-YHS | Yellowstone sulfidic hot spring, USA | 46                 | 1.05      | 1,220          | 28.4   | 94.4       | 0         | 0               | 0                             |
| Mnatro-ASL | Altai Soda Lake sediments, Russia | 73                 | 1.34      | 1,451          | 42.2   | 93.3       | 1.3       | 0               | 1.3                           |
| ANME-1-THS | Tibetan Hot Spring sediment, China | 181                | 2.03      | 2,155          | 48.7   | 78.4       | 4.9       | 33              | 3.3                           |
| GoM-Arc1-GOS | Gulf of Mexico natural oil seep, USA | 119                | 1.46      | 1,623          | 41.0   | 91.2       | 0         | 0               | 0                             |
| ANME-2c  | Gulf of Mexico natural oil seep, USA | 249                | 2.66      | 2,867          | 48.5   | 92         | 2         | 0               | 0                             |

Genome ID, origin, number of scaffolds, number of protein-coding genes, guanine-cytosine (GC) content, estimated completeness (Compl.), estimated contamination (Cont.), strain heterogeneity (Strain hetero.) and contamination excluding strain heterogeneity (Cont. excl. strain hetero.) are shown. *Percentage of contamination that can be due to binning of contigs from closely related strains.

The remaining MAGs harbour canonical MCR complexes (Fig. 2a), branching next to their closest MCR-bearing neighbours in the reference archaeal phylogeny (Fig. 1), suggesting no recent horizontal gene transfers (HGTs). The clustering of NM3 with Methanofastidiosa supports an early presence of methanogenesis in the Acherontia. The clustering of NM4 with *Verstraetearchaeota*, support that it is a genuine methanogenic/methanotrophic representative of the TACK. The separate branching of the ANME-2c MAG from the other ANME-2 lineages suggests that anaerobic methane oxidation in Methanosarcinales emerged multiple times independently from methanogenic ancestors. Strikingly, both NM1a and NM1b encode, in
addition to the McrABG-like complex, a canonical MCR complex branching at the base of Class II methanogens, consistent with the reference phylogeny. The co-existence of MCR and MCR-like complexes in the same archaeon has never been observed before and brings into question the metabolism of this lineage (see below).

Fig. 1 | Placement of nine MAGs described in this study in the reference phylogeny of Archaea. NM2 was not included due to low completeness. Bayesian phylogeny (PhyloBayes, CAT+GTR+Γ4) based on concatenation of 40 conserved phylogenetic markers (8,564 amino acid positions) and 156 genomes/MAGs (see Supplementary Tables 5 and 6 for details). Node supports refer to posterior probabilities and for reasons of readability, only values above 0.8 are shown. The tree is rooted according to Raymann et al. The scale bar represents the average number of substitutions per site. Scale bar, 0.1. Black arrows point to the nine obtained MAGs and accolated pie charts indicate their estimated completeness. Colours indicate that genomes of these lineages encode an MCR/MCR-like complex, Class I/II methanogens are in green, methyl-dependent hydrogenotrophic lineages are in red, methanotrophs are in orange (some being in Class II), potential or validated short-chain alkane users are in in blue. NM1 could also have a methane metabolism (see text for discussion).
Fig. 2 | Phylogeny of the MCR/MCr-like complex and conservation of important positions in the catalytic site. a. Unrooted Bayesian phylogeny (CAT+GTR+I+4) based on a concatenation of McrABG/McrABG-like subunits (1,187 amino acid positions) from 109 genomes/MAGs (see Supplementary Table 6 for details). Node supports refer to posterior probabilities and for reasons of readability only values above 0.8 are shown. The scale bar represents the average number of substitutions per site, 0.1. The colour code is similar to that in Fig. 1 with the exception of NM1 that have both an MCR-like (in blue) and a canonical MCR (in purple) (see text for discussion). b. Conservation of 17 residues previously described to interact with CoM, CoB, F430 cofactors, forming the catalytic site cavity wall or being post-translationally modified, are not conserved in ‘Ca. S. butanivorans’28,48,76. Replacement of conserved aromatic residues (Fig. 2b). The replacement of large aromatic residues (for example, Phe330, Tyr333, Tyr444, Tyr446; Fig. 2b) preserving an important role in canonical MCR, either by interacting with CoM, CoB, F443 cofactors, or having post-translational modifications (those marked with an asterisk)28,48,76. Replacement of conserved amino acids associated to a negative value in the Blossom45 matrix are indicated by white on black background, those with a null or positive value in the Blossom45 matrix are in bold, ‘-’ indicate conserved positions and ‘-’ indicate missing positions in the sequence due to sequencing incompleteness.

It is notable that most of the predicted or experimentally proven methyl-dependent hydrogenotrophic methanogens are closely related in the MCR tree (Fig. 2a, in red), irrespective of their placement in the reference phylogeny (Fig. 1). This might be the consequence of ancient exchanges of the MCR complex among these lineages, whose direction is hard to define. Nevertheless, some more recent transfers may be identified. For example, ‘Ca. Methanophagothrix’ (ANME-1) MCRs branch far from their Methanococcales relatives, and might have acquired their MCR complex from a methanogenic member of the Achromatium.

The clustering of MCR-like homologues belonging to distantly related lineages (Fig. 2a, in blue) is also puzzling. This might be due to HGTs and/or tree reconstruction artefacts linked to their high sequence divergence with respect to canonical MCRs, exemplified by their longer-than-average branches. Such divergence is probably related to a change in function, as MCR-like complexes are involved in activating short-chain alkanes (butane and propane) in ‘Ca. Syntrophoarchaeum’11. Accordingly, several residues playing an important role in canonical MCR, either by interacting with cofactors, forming the catalytic site cavity wall or being post-translationally modified, are not conserved in ‘Ca. Syntrophoarchaeum’ sequences (Fig. 2b). The replacement of large aromatic residues (for example, Phe330, Tyr333, Tyr444, Tyr446; Fig. 2b) present in the cavity wall of canonical MCR28 by smaller ones in ‘Ca. Syntrophoarchaeum’ MCR-like complexes could have occurred to accommodate butane/propane (a larger substrate than methane).
in the catalytic site. The presence of smaller amino acids at these positions in both NM1 and Bathyaarchaeota MCR-like complex suggests a similar function in short-chain alkane oxidation. Finally, the MCR-like sequences of GoM-Arc1 show fewer modifications at these sites, suggesting the use of a smaller alkane, possibly ethane or methane.

**Expanded diversity of methyl-dependent hydrogenotrophic methanogens.** The NM3 and NM4 MAGs share several similarities with the recently discovered order-level lineages of methanogens that were proposed or experimentally proven to perform methyl-dependent hydrogenotrophic methanogenesis\(^\text{17-19}\) (Fig. 3, Supplementary Fig. 1 and Supplementary Table 1). First, these relatively complete MAGs (85.5% completeness) lack at least 24 genes coding for the MTR complex, H\(_{\text{MPT}}\) biosynthesis and the H\(_{\text{MPT}}\) methyl-branch of the Wood–Ljungdahl pathway, otherwise present in all Class I/II methanogens (Supplementary Table 1). Second, they encode [Ni-Fe] hydrogenases and methyltransferases with the potential to support methanogenesis from methanol (MtaABC in NM3 and NM4) and methanethiol (MtsAB in NM3) (Fig. 3 and Supplementary Table 1). Energy conservation complexes of NM3 are mostly similar to Methanostefidiosiales\(^\text{17}\) (Supplementary Fig. 1), their closest related methanogens in the reference phylogeny (Fig. 1). Altogether, these data suggest that NM3 and NM4 rely on methyl-dependent hydrogenotrophic methanogenesis (Fig. 3 and Supplementary Discussion for details on energy conservation in NM3 and NM4).

The predicted methanogenesis pathway in Verst-YHS (Verstraetearchaeota) and Mnatro-ASL (Methanonatronarchaeia) MAGs also supports methyl-dependent hydrogenotrophic methanogenesis (Fig. 3 and Supplementary Table 1), as described in the first genomic assembly for these lineages\(^\text{18,19}\). However, comparison of the energy conservation enzymes in the seven currently available Verstraetearchaeota (order Methanomethyliales) suggests an alternative model than previously described\(^\text{19}\) (Fig. 3 and Supplementary Table 1). Indeed, we found that all Methanomethyliales MAGs (95% average completeness) lack the HdrA/MvhD and possibly MvhAG subunits of the electron-bifurcating complex HdrABC/MvhADG, suggesting that this complex is absent in these archaea. In contrast, we identified in these genomes a gene cluster encoding a potential complex composed of a membrane-bound hydrogenase and of HdrBC (tentatively named Energy-converting Hydrogenase D complex or Ehd, see Supplementary Fig. 2). We propose that this complex could be involved in a previously unreported mode of energy conservation associated with methanogenesis (Fig. 3 and Supplementary Discussion).

**Insights into methane and short-chain alkane oxidizers.** GoM-Arc1-GOS, ANME-1-THS and ANME-2c MAGs possess a Wood–Ljungdahl pathway and lack the methyltransferases and [Ni-Fe] hydrogenases required for methylotrophic and hydrogenotrophic methanogenesis, respectively (Fig. 3 and Supplementary Table 1), similar to all available MAGs of methanotrophs and short-chain
alkane oxidizer (Supplementary Fig. 3). Although they encode an AMP-producing acetyl-CoA synthetase (Acs) that is used for acetoclastic methanogenesis in \textit{Methanosaeta} spp., they could rather use it for acetate assimilation\textsuperscript{11}. Comparison with methanotrophs and short-chain alkane oxidizers also reveals a common core of enzymes for energy conservation, comprising the F420H2:quinone (or methanophenazine) oxidoreductase (Fqo/Fpo) and a potential electron confurcating complex (HdrABC/MvhD/FdhB\textsuperscript{29}) coded by a conserved gene cluster (Supplementary Fig. 4). ANME-2c and GoM-Arc1-GOS encode 17 and 10 multihaeme \textit{c}-type cytochromes, respectively, supporting the importance of direct electron transfer to syntrophic partners in anaerobic methane\textsuperscript{30,31} and short-chain alkane oxidation\textsuperscript{21} metabolisms (Supplementary Fig. 3 and Supplementary Table 1).

ANME-1-THS MAG is the first sequenced representative of a 'Land clade' in the '\textit{Ca}. Methanophagales' (Supplementary Fig. 5), suggesting different adaptations to environmental conditions than members of the ANME-1b clade, which are mainly from marine methane seeps. ANME-1-THS differs from the ANME-1b MAG\textsuperscript{32} by the presence of a bacterial-like Rnf complex that could couple the NAD–ferredoxin oxidoreduction with chemiosmotic gradient generation/use (Fig. 3, Supplementary Fig. 6 and Supplementary Discussion). If these genes are not in the missing region of this MAG, ANME-1-THS might also differ from the other ANMEs by the lack of multihaeme \textit{c}-type cytochromes to transfer electrons from methane oxidation to a syntrophic partner (Fig. 3 and Supplementary Fig. 3). Alternatively, two PsrABC-like complexes, including a molybdenum/selenocysteine-containing dehydrogenase subunit, could be involved in the reduction of inorganic compounds such as polysulfide/elemental sulfur\textsuperscript{33,34} (Fig. 3). This contrasts with ANME-1b MAG that misses the membrane integral (PsrC-like) subunit needed to transfer electrons from membrane-associated electron transporters (Supplementary Fig. 3). These characteristics might indicate growth of ANME-1-THS without bacterial syntrophs.
The gene content of GoM-Arc1-GOS is consistent with the recent description of the first member of the GoM-Arc1 lineage\(^2\). While GoM-Arc1 members encode an MCR-like complex possibly involved in short-chain alkane oxidation (Fig. 3), they lack the beta-oxidation pathway proposed to be involved in the use of butane/propane in ‘Ca. Syntrophoarchaeales’\(^2\) (Supplementary Fig. 3). If GoM-Arc1 members are capable of oxidizing ethane (\(\text{C}_2\text{H}_6\)), as suggested by the fewer modifications observed in the catalytic site of its MCR-like complex relative to canonical MCRs (Fig. 2), the oxidation of the ethyl group would lead to an acetyl group that could directly enter the oxidative Wood–Ljungdahl pathway, making the beta-oxidation pathway unnecessary (Fig. 3 and Supplementary Discussion). With the presence of Fqo, HdrABC/MvhD/FhdB, multiheme c-type cytochromes and HdrDE (Supplementary Table 1), the energy conservation system associated with this potential ethane-oxidation metabolism in GoM-Arc1 would mostly resemble that associated with methanotrophy in their closely related ANME-2 lineages (Supplementary Fig. 3). The question remains whether the MCR-like homologues of GoM-Arc1 could also be capable of methane oxidation.

A previously uncharacterized type of methanogenesis? The two NM1 MAGs represent the first archaea predicted to encode both an MCR and an MCR-like complex (Fig. 2), suggesting that they might be potentially capable of both methane and short-chain alkane metabolisms (Fig. 4 and Supplementary Table 1). While both NM1 MAGs encode the MTR and the m-WL pathway similarly to Class I/II methanogens, they lack the [Ni–Fe] hydrogenases (MvhA and FrhA) and methyltransferases needed for hydrogenotrophic and methylothrophic methanogenesis, respectively. They also diverge from Class I/II methanogens by the replacement of the \(F_{\text{430}}\) dependent methane-tetrahydromethanopterin dehydrogenase (Mtd) by MtdB, which relies on NAD(P) redox cofactor in *Methylobacterium extorquens*\(^3\).

Beyond the presence of an MCR-like complex, the potential ability of NM1 for short-chain alkane oxidation is also suggested by the presence of a complete beta-oxidation pathway with several gene copies per step, and a complete Wood–Ljungdahl pathway (including CODH/ACS) as in ‘Ca. Syntrophoarchaeales’\(^2\). In addition, NM1 encode multiple LCFA acyl-CoA synthases (FadD-like), not present in ‘Ca. Syntrophoarchaeales’. LCFAS activated with these enzymes can enter the beta-oxidation pathway. NM1a and NM1b also encode multiple AMP-forming acyl-CoA synthetase (Acd) to generate ATP from LCFA degradation. These enzymatic redundancies suggest a versatility towards substrates, as previously proposed for *Syntrophus aciditrophicus*\(^3\) and *Archaeoglobus fulgidus*\(^3\). Consistently, analysis of the environmental distribution of NM1 (Supplementary Fig. 7) reveals their common association with anoxic hydrocarbon-rich environments including methane seeps and oil-rich environments, where short-chain alkanes and long-chain carboxylic acids can be present in substantial concentrations\(^3\). In particular, NM1a and NM1b originate from an enrichment culture based on petroleum fluids and from a natural oil seep.

In addition to this potential wide substrate range, NM1 also contrast with ‘Ca. Syntrophoarchaeales’ in terms of energy conservation by lacking homologues of the NADH/F\(_{\text{430}}\)-H\(_2\)-quinone oxidoreductase (Nuo/Fqo) and multiheme c-type cytochromes (Fig. 4 and Supplementary Table 1). Also, NM1 contain an Rnf complex potentially using NAD instead of menaquinone for ferredoxin oxidoreduction, similarly to ANME-1-THS (Supplementary Fig. 6 and Supplementary Discussion). In the absence of membrane-bound enzymes involved in oxidoreduction of lipid-soluble electron carriers, of multiheme c-type cytochromes for direct interspecies electron transfer, of a confurcating Fe-hydrogenase for interspecies H\(_2\) transfer\(^3\), of a complete formate dehydrogenase, and of enzymes involved in dissimilatory reduction of inorganic compounds, the nature of the terminal electron acceptor coupled to alkane/LCFA oxidation remains elusive. Although both MAGs are mostly complete (~90%), it cannot be excluded that some of these enzymes are coded in their missing regions, or that an alternative way to transfer electrons to a terminal acceptor exists (for example, use of the assimilatory-type sulfite reductase present in both MAGs for dissimilatory reduction of sulfite, direct electron transfer not relying on cytochromes or use of cytochromes produced by a syntrophic partner). Alternatively, we speculate that in NM1, methanogenesis involving the canonical MCR complex could act as a sink for the electrons produced during alkane and LCFA oxidation. Several electron-bifurcating/confurcating complexes encoded in the two NM1 MAGs (Supplementary Fig. 8) together with the Rnf complex could be involved in this metabolism. The conversion of alkane and LCFA into \(\text{CH}_4\) and acetate is thermodynamically feasible but was only reported to occur through syntrophic partnerships between a bacterium (performing the beta-oxidation) and a \(\text{H}_2\)-consuming methanogen\(^3\), and it thus remains to be proven experimentally whether this can occur in a single organism.

On the basis of the presence of methane and short-chain alkane/fatty acid-related enzymes and the preferential association with hydrocarbon-rich environments, we propose the provisional class ‘Ca. Methanolipariia’, with ‘Ca. Methanoliparium thermophilum’ for NM1a and ‘Ca. Methanolliviera hydrocarbonicum for NM1b (see Supplementary Discussion for full taxonomy and nomenclature).

A core of markers related to methane and short-chain alkane metabolisms. A group of 38 genes present in most methanogens and absent from most other organisms, generally referred to as ‘methanogenesis core markers’, was previously defined from Class I/II methanogen genomes\(^3\) (Supplementary Table 2). Half of them have an unknown function. The others correspond to MCR and MTR subunits, enzymes for biosynthesis and activation of the \(F_{\text{430}}\) prosthetic group of MCR and post-translational modifications in the McrA catalytic site\(^3\). We reassessed the occurrence of these markers in the ten assembled MAGs as well as reference genomes covering all recently discovered lineages of methanogens, methanotrophs and short-chain alkane oxidizers (Table 2).

Our analysis shows that some markers are no longer universal in Class I/II methanogens (for example, m37, 38). Also, several marker genes shared by all or most Class I/II methanogens were predicted to be non-essential in *Methanococcus maripaludis* S2\(^2\). These non-universal and non-essential genes could possibly be involved in fine-tuning of methanogenesis (for example, post-translational modification of MCR\(^4\)) or in its regulation under specific environmental conditions that are not encountered by all methanogens. For example, m21 and m24 are missing in several methanogens from nutrient-rich environments, such as *Methanobrevibacter spp.*, *Methanosphaera spp.* and *Methanocorpusculum spp.*, and could be involved in regulatory processes related to changes in substrate/nutrient availability.

All the lineages of predicted and experimentally proven methyl-dependent hydrogenotrophic methanogens\(^17\)–\(^19\) lack numerous markers (Table 2), similar to what was previously noted in Methanomassiliicoccales\(^4\). These markers correspond to MTR complex subunits (m27–31), an MCR post-translational modification enzyme (m33)\(^6\) and several uncharacterized markers that are mostly non-essential in *M. maripaludis*\(^2\) (Table 2). The existence of the same pattern in NM3 and NM4 supports our inference of a potential methyl-dependent hydrogenotrophic methanogenesis. Finally, Bathyarchaeota BA1 and BA2\(^2\), which were described as methyl-dependent hydrogenotrophic methanogens\(^2\) but possess an MCR-like complex instead of the canonical MCR (Fig. 2), lack almost all methanogenesis markers (Table 2), questioning their actual metabolism.
Several homologues of the methanogenesis markers are also known to be present in non-methanogenic archaea. This is the case of the MCR/MCR-like (m1–3) and MTR (m27–31) complexes in archaeal methanotrophs\(^\text{12}\) and GoM-Arc1\(^\text{23}\), as well as the MCR-like complex in *Ca. Syntrophoarchaeales*\(^\text{21}\). Based on our analysis, archaeal methanotrophs and short-chain alkane oxidizers also appear to possess numerous markers previously exclusively associated with methanogenesis (Table 2), supporting the common origin and functional links of these metabolisms.

In addition to the MCR/MCR-like complex subunits, the most specific and conserved markers in all lineages of methanogens, methanotrophs and short-chain alkane oxidizers appear to be the genes involved in the biosynthesis (\(m\text{I}Dc\text{fB}\), \(m\text{urDcB}\) and possibly \(m\text{crD}\)) and activation (\(\text{atwA}\) and possibly \(m\text{crC}\)) of the \(F_{\text{ex}}\), prosthetic group of MCR, along with six genes encoding uncharacterized proteins (m4 to m9) (Table 2). These six genes are co-localized in most genomes (Supplementary Fig. 9) and are among those that were predicted to be co-transcribed in *Methanobacterium phaffii* R15\(^\text{52}\), suggesting that they operate in a common process. These six marker enzymes do not co-purify with MCR\(^\text{81}\). However, their phylogeny (Supplementary Fig. 10) and their restriction to archaea having MCR or MCR-like complexes strongly suggest they are involved in essential aspects of the regulation, folding and/or function of the respective holoenzymes (Supplementary Discussion).

Finally, several markers are present in archaeal lineages without MCR/MCR-like complexes (Supplementary Table 3) and are possibly remnants of an ancestral methane metabolism (Supplementary Figs. 11–13 and Supplementary Discussion).

Taken together, these observations indicate that none of the previously defined methanogenesis markers are unique to methanogens but are rather more generally indicative of metabolisms involving MCR or MRC-like complexes, including methanogenesis, methanotrophy and short-chain alkane oxidation. Elucidating the roles of these markers (MCR-associated markers) will be essential not only for understanding methanogenesis, but also anaerobic methanotrophy and short-chain alkane oxidation in archaea.

### Evolution of methane and short-chain alkane metabolisms

Our results substantially extend recent data by highlighting the overwhelming presence of lineages with an MCR or MCR-like complex in the Archaea (Fig. 1). This supports an early origin of methanogenesis in this domain of life, and multiple losses of this metabolism during archaeal diversification\(^\text{14,15,53}\).

The sharing of a common set of genes (Table 2) clearly indicates that methanogens, anaerobic methanotrophs and short-chain
alkane oxidizers are evolutionarily linked. However, it remains unclear which type of metabolism is the most ancient, and what evolutionary and functional transitions led to such diversity\(^\text{10}\). The antiquity of the Wood–Ljungdahl pathway\(^\text{4,5}\), and the recent proposal that the root of the archaeal tree might lie in between Class I and II methanogens\(^\text{13}\), would suggest that CO\(_2\)-dependent hydrogenotrophic methanogenesis is the ancestral type of methanogenesis. Nevertheless, the growing diversity of methyl-dependent hydrogenotrophic methanogens, including this work (Fig. 1 in red), indicates that this metabolism has been largely overlooked. Its origin and evolutionary relationship with CO\(_2\)-dependent hydrogenotrophic methanogenesis, however, remain unclear. The fact that it is a simpler metabolism, requiring fewer genes than CO\(_2\)-dependent hydrogenotrophic methanogenesis might suggest its earliest origin. However, it may also signify that it could have emerged later through loss of the Wood–Ljungdahl pathway and/or HGT, as suggested by the grouping of most archaea sharing this metabolism in the phylogenies of MCR (Fig. 2a) and of m4–m9 markers (Supplementary Fig. 10). Also, the clustering of NM4 with Verstraetearchaeota on a separate and well-supported clade in the MCR tree (Fig. 2a) is compatible with a possible inheritance of this metabolism from the last archaeal common ancestor, even under the classical root in between Euryarchaeota and the TACK. However, the possibility of an acquisition through ancient HGT cannot be excluded at present. More insights into the ancestral type of methanogenesis might also be gained from re-examination of the root of the archaeal tree\(^\text{18}\) including all recently discovered archaeal lineages.

The phylogenetic placement of the ANME lineages (Fig. 1), strongly suggests that the capabilities for anaerobic methanotroph emerged multiple times independently during archaeal diversification. In the Methanosarcinales this could have occurred relatively recently and repeatedly by reversal of methanogenesis, possibly through switch of function of a resident canonical MCR, leading to the different ANME-2 (Fig. 2a) and possibly ANME-3 lineages. The pool of genes associated with energy conservation in methanogenic and methanotrophic Methanosarcinales is in fact relatively similar\(^\text{16}\) (Fig. 3 and Supplementary Fig. 3) and some methanogenic Methanosarcinales encode multihaeme c-type cytochromes\(^\text{23}\) providing the necessary background for electron transfer in AOM archaea.

The identification and experimental demonstration of the capacity for oxidation of short-chain alkanes (butane, propane) by a divergent MCR-like complex in the Synthrophoarchaeales\(^\text{41}\) is among the most interesting findings of the recent years in the field of environmental microbiology. Our results extend the distribution of these MCR-like complexes in the archaea (Fig. 2a) and therefore, of potential short-chain alkane oxidation capabilities (Figs. 3 and 4). The rapid evolutionary rates of MCR-like homologues coupled to the change of key residues (Fig. 2b) suggest that these complexes might have arisen from canonical MCRs through modifications in the catalytic site to accommodate larger hydrocarbons than methane. Transitions between anaerobic methanotrophy and short-chain alkane use could have occurred in both directions as suggested by: (1) the close phylogenetic relationships between ‘Ca. Methanophagae’ and ‘Ca. Synthrophoarchaeae’ and the position of GoM-Arc1 in a clade comprising ANME-2a/ANME-2d (Fig. 1), (2) the proposed mechanism of alkane activation in their MCR/ MCR-like complexes\(^\text{41}\), (3) their very similar modes of energy conservation (Supplementary Fig. 3) and (4) their numerous shared markers (Table 2). If GoM-Arc1 is a short-chain alkane oxidizer, as suggested by its MCR-like complex, this capacity could have emerged from methanotrophy. Conversely, ‘Ca. Methanophagae’ (ANME-1) might have shifted from short-chain alkane oxidation to methanotrophy after acquisition of their MCR through HGT (Fig. 2). Finally, the first report of co-existence of an MCR and an MCR-like complex in members of the ‘Ca. Methanopila’ class opens up the possibility of an additional type of methanogenesis associated with alkane and/or LCFA oxidation. Further exploration of archaeal lineages with an MCR/MCR-like complex and their experimental characterization will lead to a more complete understanding of methane metabolisms and their derivations, as well as their environmental impact.

Methods

Metagenomic database probing and contig binning. Contigs of 6,108 metagenomes publicly available on the IMG/GI database in April 2017 were screened for the presence of COG4058, corresponding to McrA, using search tools of the database. Of these, 819 contigs containing an McrA sequence with a minimal length of 750 base pairs were downloaded. The McrA sequences present on these contigs were aligned on those of 188 published genomes by using Mafft\(^\text{8}\) (mafft-lins) and were trimmed with BMGE\(^\text{80}\) (BLOSUM30). A maximum likelihood phylogeny was calculated in IQTree\(^\text{59}\) with the TEST option for best model selection and 100 bootstrap replicates. Metagenomes containing one or several contigs coding for an McrA homologue that was only distantly related to known lineages or belonged to under-sampled lineages were downloaded from the IMG database. These metagenomes were assembled with MetaSPades\(^\text{49}\) and IDBA-UD\(^\text{60}\) and Newbler (Roche) (see Supplementary Table 4 for details). The contigs were screened for the presence of McrABG using MetaBAT\(^\text{47}\), MetaXOM, MetaBAT+, ANME-2a/ANME-2d (http://ggkbase.berkeley.edu), MaxBin 2.0 (ref. \(^\text{1}\)) and CONCOCT\(^\text{11}\) (Supplementary Table 4). An in-house pipeline (Let-it-bin, https://github.com/QuentinLetourneur/Let-it-bin) was used for read trimming, assembly and contig binning. Two of the MAGs were refined using the DAS Tool\(^\text{80}\). Completeness and contamination of the assembled MAGs were estimated with CheckM\(^\text{47}\).

Phylogenetic analyses. A reference archaeal phylogeny was built from a concatenation of 40 phylogenetic markers corresponding to the 36 proteins of the Phylosoft database, plus the alpha and beta subunits of the RNA polymerase and two universal ribosomal proteins (L30, S4) (Supplementary Table 5). We used a subset of the genomes available for each order/class/phyllum level lineages (Supplementary Table 6) to minimize biased associated with uneven distribution of taxa among them (for example, >100 taxa in Halobacteriales versus three taxa in Methanocellales). The 147 genomes were chosen because they were the most complete and the most distant to each other in each lineage. Two phylogenies were constructed: (1) a concatenation of McrABG, and of six of the most localized markers (m4 to m9) specific to genomes encoding an MCR/MCR-like complex. Sequences used for these trees were searched by hidden Markov models (HMM) in the ten MAGs obtained in this study and in genomes present in the NCBI or IMG-databases, aligned with Mafft (mafft-lins), trimmed with BMGE\(^\text{80}\) (BLOSUM30) and concatenated with an in-house script. Before concatenation, the genomes of the two datasets (McrABG and m4 to m9) were tested for congruence using the internode certainty\(^\text{75}\) test in RaxML\(^\text{69}\). Maximum likelihood phylogenies for each gene and blind concatenations were calculated in IQTree\(^\text{59}\) with the TEST option for best model selection and 100 bootstrap replicates. Sequences causing strong incongruences (with a bootstrap ≥80%) at high taxonomic ranks (order to phylum applicable) were removed, and the alignment was repeated until no further incongruence was found. Bayesian phylogenies were constructed in PhyloBayes\(^\text{81}\) under the CAT+GTR+I model. Four independent Markov chain Monte Carlo chains were run until convergence and checked by sampling over two cycles with a 25% burn-in. Support at nodes was evaluated by posterior probability values.

Maximum likelihood phylogenies were constructed in IQTree\(^\text{59}\) under the LG+G60 model.

Metabolic prediction. Gene prediction was performed using Prodigal\(^\text{47}\). All metabolic genes were identified using HHMMs searches with PFAM, TIGR and custom HMM profiles. Annotation of proteins displayed in Supplementary Table 1 was improved by inspecting the genomic context of the metabolic genes using RAST\(^\text{76}\) and SynTax\(^\text{77}\) (http://archaea.e-psud.fr/syntax/), by phylogenetic analyses including the sequences of characterized enzymes and by identifying their conserved domains using CD-search database\(^\text{78}\) (https://www.ncbi.nlm.nih.gov/ Structure/bwrpsb/bwrpsb.cgi). Genomic context was also inspected to identify conserved patterns among multiple genomes. Identification of energy conservation systems was performed with MacSyFinder\(^\text{75}\), by defining specific and sensitive models for each system followed by manual curation.

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

Data availability

MG sequences are available in the BioProject PRJNA472146 and Biosamples SANN10387997, SANN10390724, SANN10390732, SANN10390733, SANN10390735, SANN10390737, SANN10390739, SANN10390730. NM2 sequences corresponding to markers reported in Table 2 are deposited under MK202738 to MK202758.
References

1. Demirel, B. & Scherer, P. The roles of aceticotrophic and hydrogenotrophic methanogens during anaerobic conversion of biomass to methane: a review. Rev. Environ. Sci. Biotechnol. 7, 173–190 (2008).
2. Schink, B. Energetics of syntrophic cooperation in methanogenic degradation. Microbiol. Mol. Biol. Rev. 61, 262–280 (1997).
3. 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).
4. Sousa, F. L. et al. Early bioenergetic evolution. Philos. Trans. R. Soc. Lond. B 368, 20130081 (2013).
5. Kasting, J. F. & Siefer, J. L. Life and the evolution of Earth’s atmosphere. Science 296, 1066–1068 (2002).
6. Baptiste, E., Brochier, C. & Boucher, Y. Higher-level classification of the Archaea: evolution of methanogenesis and methanogens. Archaea. 1, 353–363 (2005).
7. Whitman, W. B., Bowen, T. L. & Boone, D. R. The Methanogenic Bacteria. Prokaryotes Vol. 3 (Springer, New York, 2006); https://doi.org/10.1007/3-5704-30743-5_9.
8. Kendall, M. M. & Boone, D. R. in The Prokaryotes 244–256 (Springer, New York, 2001); https://doi.org/10.1007/3-5704-30743-5_9.
9. Oren, A. in The Prokaryotes: Other Major Lineages of Bacteria and The Archaea (eds Rosenberg, E. et al.) 165–193 (Springer, New York, 2014).
10. Borrel, G., Adam, P. S. & Gribaldo, S. Methanogenesis and the Wood-Ljungdahl pathway: an ancient, versatile, and fragile association. Genome Biol. Evol. 8, 1706–1711 (2016).
11. Timmers, P. H. A. & Borrel, G. Reverse methanogenesis and respiration in Methanotrophic Archaea. Archaea. 2017, 1654237 (2017).
12. Scheller, S., Ermler, U. & Shima, S. in Anaerobic Utilization of Hydrocarbons, Oils, and Lipids (ed. Bool, M.) I–29 (Springer, Chamixon, 2017); https://doi.org/10.1007/978-3-319-35589-8_3.
13. Hallam, S. J. et al. Reverse methanogenesis: testing the hypothsis with environmental genomics. Science 305, 1457–1462 (2004).
14. 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).
15. Spang, A., Caceres, E. F. & Ettema, T. J. G. Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. Science 357, pii: eaad3883 (2017).
16. Borrel, G. et al. Phylogenomic data support a seventh order of methylothrophic methanogens and provide insights into the evolution of methanogenesis. Genome Biol. Evol. 5, 1769–1780 (2013).
17. Nobu, M. K., Nazhath, T., Kuroda, K., Mei, R. & Liu, W. T. Chasing the elusive Euryarchaeota class WAS2a genomes reveal a uniquely fastidious methyl-reducing methanogen. ISME. J. 10, 2478–2487 (2016).
18. Sorokin, D. Y. et al. Discovery of extremely halophilic, methyl-reducing methanogen. ISME. J. 2, 1783–1792 (2008).
19. Vanwonterghem, I. et al. Methylotrophic methanogenesis discovered in the novel archaeal phylum Verstatrixarchaeota. Nat. Microbiol. 1, 16170 (2016).
20. Brugière, J. F. et al. Archaeobiotics: proposed therapeutic use of archaea to prevent trimethylaminuria and cardiovascular disease. Gut Microbes 5, 5–10 (2014).
21. Laso-Pérez, R. et al. Thermophilic Archaea activate butane via alkyl-coenzyme M formation. Nature 539, 396–401 (2016).
22. Evans, P. N. et al. Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. Science 350, 434–438 (2015).
23. Dombrowski, N., Seitz, K. W., Teske, A. P. & Baker, B. J. Genomic insights into potential interspecies dependencies in microbial hydrocarbon and nutrient cycling in hydrothermal sediments. Microbiobone 5, 106 (2017).
24. McKay, L. J., & Hatzenpichler, R., Inskeep, W. P. & Fields, M. W. Occurrence and expression of novel methyl-coenzyme M reductase gene (mcrA) variants in hot spring sediments. Science. 357, 7252 (2017).
25. Hawley, E. R. et al. Metagenomes from two microbial consortia associated with Santa Barbara seep oil. Mar. Genom 18, 97–99 (2014).
26. Seitz, K. W., Lazar, C. S., Hinrichs, K.-U., Teske, A. P. & Baker, B. J. Genomic reconstruction of a novel, deeply branched sediment archaeal phylum with pathways for acetogenesis and sulfur reduction. ISME. J. 10, 1696–1705 (2016).
27. McKay, L. J. et al. Co-occurring genomic capacity for anaerobic methane and dissimilatory sulfur metabolisms discovered in the Korarchaeota. Nat. Microbiol. https://doi.org/10.1038/s41556-019-0363-4 (2019).
28. Ermler, U., Gruber, 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).
55. Adam, P. S., Borrel, G. & Gribaldo, S. Evolutionary history of carbon monoxide dehydrogenase/acetyl-CoA synthase, one of the oldest enzymatic complexes. *Proc. Natl Acad. Sci. USA* **115**, E5837 (2018).

56. McGlynn, S. E. Energy metabolism during anaerobic methane oxidation in ANME Archaea. *Microbes Environ.* **32**, 5–13 (2017).

57. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

58. Criscuolo, A. & Gribaldo, S. BMGE (block mapping and gathering with entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* **10**, 210 (2010).

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

60. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. MetaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).

61. Peng, Y., Leung, H. C. M., Yiu, S.-M. & Chin, F. Y. L. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**, 1420–1428 (2012).

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

63. Wu, Y. W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated binning technology. *BMC Genomics* **9**, 75 (2008).

64. Criscuolo, A. & Gribaldo, S. BMGE (block mapping and gathering with entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* **10**, 210 (2010).

65. Sieber, C. M. K. et al. Recovery of genomes from metagenomes via a dereplication, aggregation, and scoring strategy. *Nat. Microbiol.* **4**, D222–D226 (2015).

66. 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).

67. Lamb, A. E. & Lefebvre, N. PhyloSift: phylogenetic analysis of genomes and metagenomes. *Peer J.* **2**, e243 (2014).

68. Kubert, K., Salichos, L., Rokas, A. & Stamatakis, A. Computing the internode certainty and related measures from partial gene trees. *Mol. Biol. Evol.* **33**, 1606–1617 (2016).

69. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).

70. Cartil, N., Lepage, T. & Blanquart, S. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**, 2286–2288 (2009).

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

72. Aziz, R. K. et al. The RAST server: rapid annotations using subsystems technology. *BMC Genomics* **9**, 75 (2008).

73. Oberto, J. SyntTax: a web server linking synteny to prokaryotic taxonomy. *BMC Bioinformatics* **14**, 4 (2013).

74. Marchler-Bauer, A. et al. CDD: NCBI’s conserved domain database. *Nucleic Acids Res.* **43**, D222–D226 (2015).

75. Abby, S. S., Néron, B., Ménager, H., Touchon, M. & Rocha, E. P. A. MacSyFinder: a program to mine genomes for molecular systems with an application to CRISPR-Cas systems. *PLoS One* **9**, e110726 (2014).

76. Kahn, J. et al. Post-translational modifications in the active site region of methyl-coenzyme M reductase from methanogenic and methanotrophic archaea. *FEBS J.* **274**, 4913–4921 (2007).

Acknowledgements

We thank R. Thauer for feedback on an earlier version of the manuscript. G.B. acknowledges support from the Institut Pasteur through a Roux-Cantarini fellowship. P.S.A. is supported by a PhD fellowship from Paris-Saclay University. This work was supported by funds from the PhD Programme ‘Frontières du Vivant (Fdv)’, the Natural Sciences and Engineering Research Council (NSERC) of Canada, Genome British Columbia, Genome Canada, Canada Foundation for Innovation (CFI) and the Tula Foundation. L.N.S.-G. and V.M.d.O. are grateful to São Paulo Research Foundation—FAPESP (process Nos. 2011/14501-6 and 2013/20436-8) and Petrobras for financial support and to N. Gray and J. Head from the School of Civil Engineering and Geosciences at Newcastle University for lab facilities. W-J.L. was supported by Key Projects of Ministry of Science and Technology (MOST) (Nos. 2013DFA31880 and 2015FY110100). G.M. was supported by the ERC Advanced Grant PARASOL (No. 322551). L.M. appreciates funding from the NASA Postdoctoral Programme through the NASA Astrobiology Institute and W.P.I. was supported by the Montana Agricultural Experiment Station (Project No. 911300).

Author contributions

G.B. and S.G. conceived the study. L.J.M., L-X.C., I.N.S.-G., C.M.K.S., G.L.A., W.-J.L., S.J. H., G.M., V.M.d.O., W.-J.L., S.J. H., G.M., V.M.d.O., W.P.I. and J.F.B. sequenced and assembled the metagenomes. G.B. screened the IMG database for McrA and identified these metagenomes. Q.L., A.G., C.F.M. and G.B. developed the pipeline Let-it-bin. G.B. screened the IMG database for McrA and identified these metagenomes. Q.L., A.G., C.F.M. and G.B. developed the pipeline Let-it-bin. G.B. performed the contig binning of ANME-1-THS MAG and C.M.K.S. those of GoM-Arc1-GOS and NM1b, NM2, NM3, NM4, Verst-YHS, and Mnatro-ASL MAGs. L-X.C. carried out the dereplication, aggregation, and scoring strategy. G.B. and S.G. wrote the manuscript. All authors read and commented on the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41564-019-0363-3.

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

Correspondence and requests for materials should be addressed to G.B. or S.G.

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
- Local genome databank assembly: NCBI e-utilities ; Screening of IMG database : IMG tools

Data analysis
- Assembly: MetaSPAdes 3.11.1, IDBA_UD v1, Newbler 2.9, SOAPdenovo v1.05; Binning: ESOM, DAS_Tool, Metabat, ABAWACA 1.07, MaxBin 2.0, CONCOCT, Let-It-Bin; Bin completeness/contamination estimation: CheckM; homology searches: HMMer, BLAST; alignments: Mafft; Sequence trimming: BMGE; phylogenies: IQTree, Phylobayes; congruence tests: IC test in RaxML; annotation: RAST server, KOALA (KEGG), CD-Search Batch, BLAST, Prodigal, Macsyfinder;

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

Gene accession numbers are provided in the Supplementary Tables. Raw phylogenetic trees (newick files) and alignment files are available upon request.

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
- [X] Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see 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 | Phylogenomic study of methanogens, methanotrophs and potential short-chain alkane oxidizer, based on all archaeal genomes available in November 2017 and including MAGs (metagenome assembled genomes) of yet unreported lineages of archaea. |
| Research sample | 6108 metagenomes available on IMG/JGI; 694 archaeal genomes (from IMG/JGI and NCBI databases), 10 novel archaeal MAGs obtained by contig binning on several metagenomes from the IMG/JGI database (3300001749; 3300013881; 2077657018; 2077657019; 2077657014; 2084038021; 3300013883; 3300001446; 3300001446; 330000557). These metagenomes were obtained from environmental samples corresponding to enrichment culture (40°C and 50°C) from petroleum samples (Brazil), Santa Barbara Channel and Gulf of Mexico oil seeps (USA), Yellowstone (USA) and Tibetan (China) hot spring sediments, Altai hypersaline soda lake sediments (Russia). |
| Sampling strategy | All the metagenomes available on IMG in April 2017 (6108) were screened for the presence of COG4058 (McrA). All metagenomes presenting McrA sequences corresponding to poorly characterized or yet unreported lineages were downloaded for binning. 10 MAGs were obtained by following this procedure. |
| Data collection | Guillaume Borrel downloaded the metagenomes of interest (containing the mcrA sequence of interest) with IMG/JGI online tools. Protein datasets of each genome were downloaded through NCBI's e-utilities, or the download services of each database by Guillaume Borrel and Panagiotis Adam. If no protein dataset was available (only contigs), proteins were predicted with Prodigal. |
| Timing and spatial scale | Screening of metagenomes in IMG/JGI database was first performed in November 2015 and repeated in May 2016 and April 2017. Genomes of our local database are gathered every two month from the IMG/JGI and NCBI databases. Genomes included in the current study are part of our local database updated in November 2017. Metagenomes screened in IMG/JGI databases and genomes gathered from both IMG/JGI and NCBI databases originate from a wide range of environments from all continents and many oceanic locations. |
| Data exclusions | No metagenome was excluded for the screening of McrA. During binning with ESOM, contigs that had less than 75% of their sub-part (3-5kb or 5-10kb fragments) in the bins of interest were excluded from these bins. The exclusion criteria was not pre-established. |
| Reproducibility | Bootstrap replicates for Maximum Likelihood, posterior probabilities for Bayesian phologenies. All attempts to repeat the experiment were successful. |
| Randomization | Randomization was not relevant to our study. We analysed all metagenomes present in IMG/JGI databases in April 2017 and all MAGs corresponding to novel lineages according to their position in phylogenies. |
| Blinding | Blinding was not relevant to this study. We performed phylogenetic analyses and metabolic predictions. |

Did the study involve field work?  [ ] Yes  [X] No

Reporting for specific materials, systems and methods
**Materials & experimental systems**

| n/a | Involved in the study |
|-----|-----------------------|
| ☒   | Unique biological materials |
| ☒   | Antibodies |
| ☒   | Eukaryotic cell lines |
| ☒   | Palaeontology |
| ☒   | Animals and other organisms |
| ☒   | Human research participants |

**Methods**

| n/a | Involved in the study |
|-----|-----------------------|
| ☒   | ChIP-seq |
| ☒   | Flow cytometry |
| ☒   | MRI-based neuroimaging |