nature research

Corresponding author(s): *Double-blind peer review submissions: write DBPR and your manuscript number here instead of author names.*

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 ; Sreening 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 completness/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 ☒ 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; 3300000557). 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 downloded 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 ☒ 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