nature research

Corresponding author(s): Nikolay A. Chernyh, Filipa L. Sousa, Inês A. Cardoso Pereira

Last updated by author(s): Jul 8, 2020

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 our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the 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
☐	☒	A statement on 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 statistical parameters 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

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

Sulphate was analysed with a Stayer ion chromatograph (Aquilon) equipped with an IonPack AS4-ASC column (Dionex).
The sequencing of the libraries of 16S rRNA gene fragments was performed with the Illumina MiSeq platform.

All qPCR assays were performed using StepOnePlus™ Real-Time PCR System (Life Technologies, USA) and qPCRmix-HS ROX (Evrogen, Russia).
The 35S radioactivity was measured in an LKB wallac 1215 RackBeta liquid scintillation counter.

For proteome analysis we used a nano liquid chromatography system (Eksigent Technologies nanoLC Ultra 1D plus, AB SCIEX) coupled to a high-speed Triple TOF 5600 mass spectrometer (SCIEX) with a Nanospray III source.

All genomes used for bioinformatics analysis were downloaded manually from IMG or from the NCBI ftp site, using NCBI recommended strategy (see https://www.ncbi.nlm.nih.gov/genome/doc/ftpfaq/) and with the aid of a small script that we provide via Figshare (https://figshare.com/s/fb5a1541be68639f4026)

Data analysis

Analyses of data from high-throughput sequencing of 16S rRNA gene fragments, including quality trimming, demultiplexing, taxonomic assignments, and core diversity analyses, were performed with QIIME (version 1.9.162) and SILVA online data analysis service. The coverage of the microbial community was estimated by determining Chao1 richness.

In proteomic analysis, MS and MS–MS data obtained for individual samples were processed using Analyst TF 1.7 Software (SCIEX). The reconstituted chromosome sequence was used to generate the database for protein identification using Mascot Server v. 2.5.1 (Matrix Science).

Databases used to analyze data and or against which data was compared:
- nr (ftp://ftp.ncbi.nlm.nih.gov/) and https://blast.ncbi.nlm.nih.gov/Blast.cgi
- Uniprot/Swissprot Knowledgebase (https://www.uniprot.org/)

Data was analyzed using the following published software:
- rBBH approach - BLAST+ version 2.4.0+
- RStudio version 1.0.153
- R-package genoPloR70 (version 0.1)
- CDD (version 3.16 or higher)
- Clustalw2 (version 2.1)
- trimAl (v1.4.rev22)
- IQ-TREE version 1.5.2 or higher
- FigTree (version 1.4.3)
- Analyst® TF 1.7 Software (SCIEX)
- Proteome Discoverer v2.4 (Thermo Fisher Scientific)
- Mascot Server v. 2.5.1 and v2.7.0.1
- QIIME (version 1.9.1)
- SILVA online data analysis service (https://ngs.arb-silva.de/silvangs/)
- Minimal Ancestor Deviation method (MAD v. 2.22)

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 and reviewers. 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

Candidatus Vulcanisaeta mountnovskia 768-28 has been deposited into the All-Russian Collection of Microorganisms (VKM) under the number VKM B-3479. The mass spectrometry proteomics data have been deposited in the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD012750 and 10.6019/PXD012750. Pyrosequencing read data obtained for the 16S rRNA gene fragments were deposited in Sequence Read Archive (SRA) under the accession numbers PRJNA642261 .

Genomic data and code availability: all genomes analysed have been made publicly available on NCBI or IMG previously. Detailed accession codes, download date and location of the records in the original dataset are given in the Supplementary Table 4, accompanying this manuscript. The binary culture genomes used here have the accession codes (GenBank CP002529, CP002590) and have been previously sequenced and described in the following publications:

- Mardanov, A. V. et al. Complete genome sequence of the thermoacidophilic crenarchaeon Thermoproteus uzoniensis 768-20. J. Bacteriol. 193, 3156–3157 (2011).

- Gumerov, V. M. et al. Complete genome sequence of Vulcanisaeta moutnovskia strain 768-28, a novel member of the hyperthermophilic crenarchaeal genus Vulcanisaeta. J. Bacteriol. 193, 2355–6 (2011).

The alignments, phylogenetic reconstructions and small custom scripts are available via figshare under the link https://figshare.com/s/fb5a1541be68639f4026. The source raw data underlying the remaining figures is provided in the Supplementary Information accompanying the manuscript.

Field-specific reporting

Please select the one below that is 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/documents/nr-reporting-summary-flat.pdf

Life sciences study design

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

Sample size	As two to three biological replicates are common practice in prokaryotic in vivo research, sample size was determined accordingly. Experiments with the binary culture 768-28 and with the pure cultures of the putative sulphate reducers Thermoproteus tenax, Vulcanisaeta distributa, V. souniana, and Caldivirga maquilingensis were performed 2-3 times. Similar trends were observed for the biological replicates. Bioinformatics Analysis: We did not obtain any new genomic assemblies, as the genomes used for computational analyses have already been reconstructed in previous and published work and are publicly available. All complete archaeal (253) and bacterial (4856) genomes available at the download date were retrieved from NCBI. To increase the taxonomic diversity within the archaeal domain, 33 selected metagenomes were additionally downloaded. In addition, the entire NCBI nr database was used for comparative analysis.
Data exclusions	no data was excluded from the analysis
Replication	Growth and proteomic experiments had 2-3 independent biological replicates and each biological replicate had three technical replicates.

We hadataset, ve described our methodological approaches for phylogenetic analyses to ensure reproducibility. We have calculated bootstrap support values for the different protein families analysed in our phylogenies to provide an indication as to how stable the tree topology is. Using the same methodological approach, software and dataset it can be reproduced with aid of the small scripts provided in figshare (https://figshare.com/s/fb5a1541be68639f4026).

Randomization: Samples were independently and randomly analyzed.

Blinding: Samples were blindly analyzed.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

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