Corresponding author(s): Rohit Ghai

Last updated by author(s): Feb 1,2019

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.

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

The following data sources were used in this study.
Database versions
SILVA 16S rRNA database version SSURef_NR99_132 database38
NCBI NR protein database: Downloaded on March 30, 2018
Clusters of Orthologous Groups: http://www.ncbi.nlm.nih.gov/COG/
Archaeal Cluster of Orthologous Groups: ftp://ftp.ncbi.nih.gov/pub/wolf/COGs/arCOG
InterPro databases as available in version InterProScan Version 5.25:
TIGRFAM,SMART,Pfam,Hamap,ProDom,PRINTS,CDD,SUPERFAMILY,Gene3D,PIRSF

Data analysis

The following programs were used to analyze the data in this study.
Software versions:
hmmsearch version 3.1b2
Kalign version 2.04
FastTree version 2.1.7
MMseqs version: d96d24698a58604a3daaf2c18fda6cd19afcf53e
MAFFT version 7.055b
USEARCH version 6
BBMap version 36.x
PRANK version 150803
T-COFFEE version_11.00.8cbe486
MEGAHIT version 1.1.2
MetaBAT version 2.12.1
reformat.sh version 36.19
uclust version 1.2.22q

ublast version 6
SSU-ALIGN version 0.1.1
blastall version 2.2.21
PRODIGAL version 2.6.2
bbwrap.sh version 36.x
CheckM version 1.0.11
Prokka version 1.11
BlastKOALA version 2.1
barrnap version 0.8
tRNAscan-SE version 1.4
InterProScan version 5.25-64.0
HmmerWeb version 2.26.0
Phyre2 version 2.0
Phobius version 1.01
BMGE version 1.12
IQ-TREE version 1.6.2
PhyloBayes MPI version 1.8
MUSCLE version 3.8.31
SINA version 1.2.11
RAxML version 7.2.8
ARB version 6.0.6
LUCIA version 1.0
catfasta2phyml (https://github.com/nylander/catfasta2phyml)
Phylogears (https://github.com/astanabe/Phylogears)
A complete description of the parameters used to run the Bayesian phylogenomic analyses is provided in the materials and methods section of the manuscript. No custom code that is central to the conclusions of this study was generated.

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. 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

All sequence data produced within the study is deposited in the Sequence Read Archive (SRA) database of the National Center for Biotechnology Information (NCBI) and can be found linked to the Bioproject PRJNA483005. Raw figure data files are available from FigShare (LINK). There are no restrictions on data availability.

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

Ecological, evolutionary & environmental sciences study design

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

Study description	This is an exploratory metagenomic study in which we recover and analyze a high number of Asgardaeota-related genomes. The nature of the study does not necessitate any treatment factors, interactions, design structure (factorial, nested, hierarchical) or replicates.
Research sample	In this study we generated metagenomic datasets that targeted the microbial communities present in two saline sediment samples. The sample collection sites were thought to contain a high number of novel prokaryotic lineages. The genomes described in this study are representative for the saline sediment habitats. Accession numbers are provided in the manuscript. In addition, we also used several previously published available datasets and genomes for which details are provided in the Supplementary Information.
Sampling strategy	Two sediment environmental samples were collected from two sites. Sediment sampling was performed on 10 October 2017 at 12:00 in Tekirghiol Lake, Romania, (44°03.19017 N, 28°36.19083 E) and on 11 October 2017 at 15:00 in Amara Lake, Romania, (44°36.30650 N, 27°19.52950 E). Two plunger cores of 0.3 m each were collected from a water depth of 0.8 m in Tekirghiol Lake and 4 m in Amara Lake. No sample-size determination was used in the experimental design. Replicate DNA extractions were performed from the same environmental sample in order to minimize the recovery biases.
Data collection	Sediment sampling was performed on 10 October 2017 at 12:00 in Tekirghiol Lake, Romania, (44°03.19017 N, 28°36.19083 E) and on 11 October 2017 at 15:00 in Amara Lake, Romania, (44°36.30650 N, 27°19.52950 E). Two plunger cores of 0.3 m each were collected from a water depth of 0.8 m in Tekirghiol Lake and 4 m in Amara Lake. Sediment samples were stored in the dark at 4 °C and processed within 24 hours after collection

DNA was extracted from approximately 10 g of wet sediment from each mixed core sample using the DNeasy PowerMax Soil Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. Extracted DNA was further purified by passing it through humic acid removal columns (type IV-HRC) provided in the ZR Soil Microbe DNA MiniPrep kit (Zymo Research, Irvine, CA, USA). Purified DNA was quality checked and quantified using a ND-1000 NanoDrop spectrophotometer (Thermo Scientific, Waltham, MA, USA). From each sample, 4 µg of pure DNA were vacuum dried in a SpeedVac concentrator (Thermo Scientific, Waltham, MA, USA) and shipped for library construction and NGS sequencing to Macrogen (Seoul, South Korea).

Paul Bulzu performed the sampling and DNA extraction.

Library preparation was performed by the commercial company by using the TruSeq DNA PCR Free Library prep kit (Illumina). Whole-genome shotgun sequencing of the 150 paired-end libraries (350bp insert size) was done using a HiSeq X (Illumina) platform. The amount of total raw sequence data generated for each metagenome was: 64.5Gbp for Amara and 57.6 Gbp for Tekirghiol.

The leachable major ions were water-extracted using a sediment-to- (milli-Q) water ratio of 1:10 at room temperature. The suspension was centrifuged and the supernatant was filtered through 0.22 µm-pore sized membranes. The obtained filtrate was further analyzed for ion content (Supplementary Table S9). Cations (Na+, K+, Mg2+) were measured by inductively coupled plasma atomic emission spectrometry (ICP-AES) using Optima 5300DV spectrometer (Perkin Elmer, USA). Chloride (Cl-) ions were measured by titrimetric method. Sulfate (SO4 2-) was assessed by ion chromatography on ICS-1500 (Dionex, Sunnyvale, CA, USA). The analysis of salt contents of Tekirghiol and Amara sediments indicated that dominant cations and anions were (g ·Kg-1): Na+ (16.5 and 7.0), K* (1.0 and 0.22), Mg2+ (1.1 and 4.0), Cl- (27.7 and 11.2) and SO4 2- (0.25 and 13.2). All chemical analyses were performed by E.A. Levei and M. Șenilă at INCDO-INOE 2000 - Research Institute for Analytical Instrumentation (Cluj-Napoca, Romania).

Timing and spatial scale

Sediment sampling was performed on 10 October 2017 at 12:00 in Tekirghiol Lake, Romania, (44°03.19017 N, 28°36.19083 E) and on 11 October 2017 at 15:00 in Amara Lake, Romania, (44°36.30650 N, 27°19.52950 E). Two plunger cores of 0.3 m each were collected from a water depth of 0.8 m in Tekirghiol Lake and 4 m in Amara Lake.

Data exclusions

No data were excluded from the analyses.

Reproducibility

The main findings of the manuscript are reproducible at several environmental sites already. e.g. assembly and recovery of genomes of Asgardaeota from varied sites e.g. Loki's castle, mangrove sediment, and the Red Sea etc. All methods employed are referenced extensively and the software parameters used are also provided.

Randomization

This is an exploratory study and randomization is not relevant to the study design.

Blinding

Blinding was not performed because it was not relevant to this study. This study was an exploratory survey of microbial diversity without a priori expectations that would influence the analyses.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions

Sediment sampling was performed on 10 October 2017 at 12:00 in Tekirghiol Lake, Romania, (44°03.19017 N, 28°36.19083 E) and on 11 October 2017 at 15:00 in Amara Lake, Romania, (44°36.30650 N, 27°19.52950 E). Two plunger cores of 0.3 m each were collected from a water depth of 0.8 m in Tekirghiol Lake and 4 m in Amara Lake. Sediment samples were stored in the dark at 4 °C and processed within 24 hours after collection. Chemical analysis results for the samples are provided in the Supplementary Tables. The temperature for the Tekirghiol samples was 15°C and for Amara 15.7°C.

Location

Sediment sampling was performed on 10 October 2017 at 12:00 in Tekirghiol Lake, Romania, (44°03.19017 N, 28°36.19083 E) and on 11 October 2017 at 15:00 in Amara Lake, Romania, (44°36.30650 N, 27°19.52950 E). Two plunger cores of 0.3 m each were collected from a water depth of 0.8 m in Tekirghiol Lake and 4 m in Amara Lake.

Access and import/export

All samples were collected in compliance with local, national and international laws and all required permits were obtained from the counties administration offices and the Romanian Ornithological Society (No.: 11526 for Amara Lake and 1669 for Tekirghiol Lake).

Disturbance

No disturbances were caused by the sampling procedures.

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
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

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