nature portfolio

Corresponding author(s): Alexander J. Probst

Last updated by author(s): Jun 19, 2023

Reporting Summary

Nature Portfolio 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 Portfolio 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	Datasets were created based on DNA and RNA sequencing as well as based on fluorescence in situ hybridization (FISH) as stated in the manuscript.
Data analysis	Software code and web links are available from online sources and appropriately referenced to in the method section. A description of how we performed the analysis and which software version we used can be found in this respective methods section in the manuscript. A list of software and versions used is provided below. metaSPAdes (version 3.10), bowtie2 (version 2.3.5.1), uBin (version 0.9.14), bbduk (version 37.09), sickle (version 1.33), HMMER 3.2.1, Phylosift, MUSCLE v3.8.31, BMGE (BLOSUM30), IQ-TREE v2.0.5, ModelFinder, PMSF model (LG+C60+F+G), Ultrafast bootstrap, SH-aLRT, aBayes, iTOL (version 5), MetaCRAST, CD-hit (cersion 4.8.1), CRISPRDirection 2.0, VirSorter (version 1), ISEScan, PSAMM (v1.0), cplex solver (v12.7.1.0) CRISPRTarget (June 2020), WebLogo (v2.8.2), BBmap (v38.92), VarScan (v2.4.3), CRISPRCasFinder (version 2.0.3), GTDB-Tk classify (v0.3.3 r89), Cytoscape (version 3.9.02), blastn (v 2.9.0+), CheckV v0.11.3, Refseq Database (release July 2022), vContact2 (v. 0.11.3), INPHARED (v1.2), VICTOR, VIRIDIC (v1.0 r3.6), Virus-Host DB:RefSeq release 217, ViPTree version 3.5, Zen 3.4 Pro (version 3.4.91.00000), BOOST design software (v1.3.9), DIAMOND (version 2.0.15.153), MAFFT FFT-NS-2 (v7.505), Seaview (v 5.0.4), SAMtools (version 1.10), BEDtools (v2.27.1), R basic, R studio (v.2023.03.0+386), checkM (v1.2.2), EggNOG (v5.0), Biorender, Prodigal (v2.6.3), Microsoft Office (v16.74), KEGG (v >94.0), Transporter Classification Database (2016)

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 Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Metagenomic datasets generated from the Crystal Geyser (CG) (Emerson et al. 2016, Probst et al. 2018) ecosystem (Utah, USA) in 2009, 2014, and 2015 (n = 66), and the Horonobe Underground Research Laboratory (HURL) (Hokkaido, Japan) (Hernsdorf et al. 2017) environment (n = 1) were downloaded from the NCBI' Sequence Read Archive (SRA) in April 2019 (Table S1). SAGs generated in a previous study (Probst et al. 2018) (n = 219) were retrieved from the JGI's Integrated Microbial Genomes and Microbiomes database (Chen et al. 2019) (Table S3). The metagenome-derived genomes of Ca. A. crystalense and Ca. H. crystalense from CG are publicly accessible from NCBI (accession numbers in Table S2). The genomes of Ca. A. horonobense and Ca. H. julieae from HURL were newly reconstructed in this investigation (Table 2). All previously unpublished genomes used in this study are available in a Figshare folder 10.6084/m9.figshare.22339555 and all viral genomes are available here: 10.6084/m9.figshare.22738568. All raw FISH images are deposited here: https://doi.org/10.6084/m9.figshare.22739849.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	This does not apply to the present study.
Population characteristics	This does not apply to the present study.
Recruitment	This does not apply to the present study.
Ethics oversight	This does not apply to the present study.

Note that full information on the approval of the study protocol must also be provided in 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

Ecological, evolutionary & environmental sciences study design

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

Study description	Our study deals with Altiarchaeota, a dominant archaeal primary producer of the deep subsurface and its interaction with viruses and the episymbiont of the uncultivated phylum Huberarchaeota. We investigated the host's CRISPR-Cas systems and associated spacer matches to the episymbiont genomes to investigate the nature of the relationship. We also transferred these findings to other archaeal symbiont-host associations.
Research sample	We used publicly available metagenomic/metatranscriptomic datasets of filtered groundwater samples from Crystal Geyser and Horonobe Underground Research Laboratory, in 350m and 250m depth, respectively. For FISH analyses we retrieved filter sets with microbes from the two sites as well.
Sampling strategy	The sampling strategies/procedures can be found in the method section in the manuscript. We analyzed 66 metagenomes from one ecosystem and compared the results to a distinct ecosystem with one publicly available metagenome, were the host-symbiont relationship was also found. Additionally, we also analyzed 3 metatranscriptomes of the respective site. Samples for FISH were additionally collected specifically in eruption periods of the geyser, which showed the respective symbiont-host association, and from the respective aquifer in HURL.
Data collection	Metagenomic data were downloaded from public datasets in April 2019 and June 2021.
Timing and spatial scale	Sampling of Crystal Geyser for metagenomes and metatranscriptomes was performed in 2009, 2014 and 2015. Sampling of the Horonobe Underground Research Laboratory was conducted between June 2012 and June 2013. Subsurface ecosystems rely on geological settings and geological time scales that do not affect the study. FISH samples of Crystal Geyser were collected in 2021 and of HURL in 2019. Dates of samplings are listed in the section "Field work, collection and transport". All further details are specified in the method section of our manuscript.

Data exclusions	We did not exclude any data.
Reproducibility	Detailed information can be found in the method section of the manuscript. Metagenomes had 66 replicates, transcriptomes had three biological replicates. We did not test for reproducibility of metagenomic sequencing with technical replicates but relied on biological replicates instead. Environmental parameters of the groundwater acquifer/sulfidic spring were extensively investigated previously and remained constant over years.
Randomization	This does not apply to the present study.
Blinding	This does not apply to the present study.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	Metatranscriptomics: Mai 25/26 2015 at Crystal Geyser (location, see below). Samples were retrieved from the subsurface and thus the weather conditions are not relevant. FISH: The sampling of Crystal Geyser was performed at the 13th of August 2021 within the minor eruption phase. Samples were retrieved from the subsurface and thus the weather conditions are not relevant. HURL/FISH: The sampling of HURL was performed at the 9th of July 2019. Samples were retrieved from the subsurface and thus the weather conditions are not relevant.
Location	Crystal Geyser (Utah, USA) longitude: -110° 08' 7.58" W; latitude: 38° 56' 17.71" N; Honorobe Underground Research Laboratory (Japan) longitude: 141° 51' 34.55" E; latitude: 45° 02' 43.41" N
Access & import/export	As the sampling site is in the middle of a desert, the sampling campaign was performed by car. The samples were directly frozen at site at -80 degrees celsius and transported frozen.
Disturbance	There was no disturbance of the ecosystem caused by this study.

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
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

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