nature research

Corresponding author(s): Virginia Edgcomb 2019-08-12301

Last updated by author(s): 2019/12/28

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	For Raman analysis of thin section samples we used the LabSpec 6 ® Spectroscopy Suite Software (https://www.horiba.com/en_en/products/detail/action/show/Product/labspec-6-spectroscopy-suite-software-1843/) for data acquisition and background subtraction of acquired spectra. The processed spectra were then compared with the BIO-RAD KnowItAll® Raman Spectral Library databases for identification of organic substances and with the RRUFF (TM) Project Raman database for identification of minerals (Lafuente et al. (2015) (Lafuente B, Downs R T, Yang H, Stone N (2015) The power of databases: the RRUFF project. In: Highlights in Mineralogical Crystallography, T Armbruster and R M Danisi, eds. Berlin, Germany, W. De Gruyter, pp 1-30). For lipid analyses, Sciex Analyst 1.6.3 and Sciex MultiQuant 3.0.3 (AB Sciex LP, Concord, Canada) were used for triple quadrupole mass spectral data acquisition and data processing. Bruker Compass 1.9 and Bruker Data Analysis Version 4.4 (Bruker Daltonics, Bremen, Germany) were used for quadrupole time-of flight lipid data acquisition and processing.
Data analysis	The software R was used with the pvclust package for nMDS and clustering analysis of OTU data. Trimmomatic v. 0.32 was used to trim adapter sequences in transcript data, FastQC v. 0.11.7 and FASTX-toolkit v. 0.014 were used to quality check and trim paired reads in transcriptome libraries, Trinity v. 2.4.0 was used for assembly of transcriptomes, Bowtie v. 2.3.4.1 was used to align reads to our assembly and for de novo assembly of reads in our libraries, Bowtie aligner was used to align reads to assembled contigs, RSEM was used to estimate expression levels of reads in our libraries, TMM was used to perform cross-sample normalization and to generate a TMM-normalized expression matrix. The Trinotate suite (TransDecoder v. 3.0.1) was used to identify coding regions, BLASTx and BLASTp against UniProt, Swissprot (release 2018_02) and RefSeq nr were used to look for homologies between our reads and databases. HMMER v. 3.1b2 was used to identify conserved domains by searching against the Pfam v. 31.0 database. SignalP v. 4.1 and TMHMM 2.0c were used to predict signal peptides and transmembrane domains, RNAMMER v.1.2 was used to identify rRNA homologies. BLAST against the KEGG, COG, SEED and MetaCyc databases using MetaPathways v. 2.0 was used to interpret functions and pathways active in our transcriptome data. For our iTAG data Raw sequence reads were processed through Trim Galore [http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/], FLASH (ccb.jhu.edu/software/FLASH/) and FASTX Toolkit [http://hannonlab.cshl.edu/fastx_toolkit/] for trimming and removal of low quality/short reads. Quality filtering included requiring a minimum average quality of 25 and rejection of paired reads less

than 250 nucleotides. Operational Taxonomic Unit (OTU) clusters were constructed at 99% similarity with the script pick_otus.py within the Quantitative Insights Into Microbial Ecology (QIIME) v.1.9.1 software 45and 'uclust'. The OTU data table was transformed to a presence/absence table and the Jaccard method was used to generate a distance matrix using the dist.binary() function in the R package ade4. A hierarchical clustering dendrogram was created using hclust() and the stability of the clusters was evaluated using the clusterboot() function in the fpc package in R with 500 iterations.

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

iTAG data for the 11 samples as well as 13 different negative controls (drilling muds and fluids, seawater, kit controls) are deposited in GenBank under BioProject PRJNA497074. Results are presented at ~phylum level, however taxonomic assignments at finer resolution are available from the authors upon request. Raw reads for transcriptome data have also been doposited in GenBank SRA under BioProject PRJNA497074. Assemblies for curated portions of the data presented are available upon request to the authors.

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	Samples available for microbiological studies were selected by the Co-Chief Scientists of Expedition 360. They were limited to ~10-20cm long whole round rock sections from recovered cores. Replication was not possible because there was a single hole drilled for all studies conducted during Expedition 360, and only the 10-20 cm sections of 10m cores were allocated to microbiologists. Each of the individual 10-20cm core sections was processed to remove exterior rock and remaining material was subdivided for all analyses presented here.
Data exclusions	Outside of subtraction of molecular data matching any contamination controls (as noted in methods), no data were excluded. All data collected from analyzed samples is presented.
Replication	Cell counts were replicated by processing two replicate 1ml tubes of material from each core sample. Alkaline phosphatase activity measurements were conducted on triplicate technical replicates of material for those analyses and results were consistent. Methane production measurements are single measurements due to limited sample availability, except for the artificial seawater blank (n=3), for which the mean value is presented. All attempts at replication were successful. Sufficient material was not available for replicating other analyses. Lipid biomarker measurements were single measurements due to limited sample material. Raman spectral analysis presented in Figure 3a was made at two spots on this particular feature of interest, 10 independent times per spot, with similar results. In total, we performed >2000 analyses (with shorter integration times) on similar features in the same thin section sample with similar results.
Randomization	Eleven rock samples were selected from our collection for analyses presented here, purely on visual observations of the rock samples, prioritizing samples from the full depth of the hole drilled, and samples with evident veins and signs of alteration.
Blinding	Blinding was not applicable to this study of rock samples and associated control samples beyond the fact that all samples were number coded and treated identically throughout preparation and analyses.

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