nature portfolio

Corresponding author(s): David Walsh

Last updated by author(s): July, 8, 2026

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	NA
Data analysis	multiple sequence alignment using the MUSCLE algorithm in MEGA v.11, The maximum likelihood tree was constructed using the JTT substitution model, gamma distribution (4 categories), and the nearest neighbor heuristic search model in MEGA v.11 with 100 bootstrap replicates. To determine MAG gene expression abundance from, Canada Basin metatranscriptomes, we performed fragment recruitment of metatranscriptomic data previously described in19. Competitive fragment recruitment of the Canada Basin metatranscriptomes was carried out in the same manner as the metagenomic fragment recruitment. Metatranscriptomic RPKG values for the MAGs was calculated by dividing the total number of reads mapped to the MAG by the size of the MAG in Kbp and the size of the metatranscriptome in Gbp. For gene expression analysis in P. arcticus, metatranscriptomic RPKM values were calculated for each of the gene-coding sequences by dividing the total number of reads mapped to the gene by the size of the gene in Kbp and the total number of metatranscriptome reads in millions. A multiple-sequence alignment of near full-length 16S rRNA genes from Gammaproteobacteria was generated using the MUSCLE algorithm as implemented in MEGA v.1130. A maximum likelihood tree was constructed using the GTR + gamma distribution (4 categories) model of nucleotide substitution in MEGA v.11 with 100 bootstraps. Gene prediction and annotation was performed using Prokka v.1.1233, which implemented Prodigal v.2.6.3 for protein coding genes34, Barrnap for ribosomal RNA genes, and ARAGORN v.1.235 for transfer and transfer-messenger RNA genes. Functional annotation of proteins was performed using KofamKOALA with default settings and a threshold score of 0.7 or higher and an e-value of 1 x 1010 or lower (Aramaki et al., 2020). To explore differences in gene content between P. arcticus, Porticoccus sp. Uisw_50_02, and P. hydrocarbonoclasticus, the distributions of orthologous genes were analyzed by Proteinortho36. The KEGG Mapper "Reconstruct" tool was used to identify central metabolic pathways harboured by P. arcticus and P. hydrocarbonoclasticus37. To verify proteorhodopsin annotations, we performed a phylogenetic analysis to increase confidence in the biochemical function of the rhodopsin genes. Reference sequences for proteorhodopsin and sensory rhodopsin I and II were included in a multiple sequence alignment using the MUSCLE algorithm in MEGA v.1130. The maximum likelihood tree was constructed using the JTT substitution model, gamma distribution (4

categories), and the nearest neighbor heuristic search model in MEGA v.11 with 100 bootstrap replications30. Finally, to gain additional confidence in the biochemical function of the proteorhodopsin genes, we analysed the primary amino acid sequence for the presence of proton pumping and light absorbing motifs

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

Single sample metagenome and metatranscriptome dataset are available at the Joint Genome Institute Genomes OnLine Database under study ID Gs0134626 (see Table S1 for details). Two co-assemblies are also available under study ID Gs0134626: the CB co-assembly (Analysis ID Ga0485091) and the Upper water co-assembly (Analysis ID Ga0485101). The western Arctic Ocean MAG catalog is available at Dryad (https://doi.org/10.5061/dryad.g4f4qrfxk). The P. arcticus genome is available at NCBI (JAZHGD000000000; BioSample SAMN39732788).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	NA
Reporting on race, ethnicity, or other socially relevant groupings	NA
Population characteristics	NA
Recruitment	NA
Ethics oversight	NA

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	Metagenomic and 16S rRNA analysis of Arctic Ocean microbial communities
Research sample	3 um to 0.22 um size fraction of bacterial cells from different depths as described in Supplementary Table 1. These samples represent bacterioplankton populations.
Sampling strategy	Water samples (7-14L were collected by CTD rosette. Bacterioplankton were collected on a 0.22 um filter after passing through a 3 um prefilter.) No staistics were used to predetermine sample size.
Data collection	Data was collected on board the Louis St. Laurent ice breaker by a collection of scientists.
Timing and spatial scale	Summer sampling across the Canada Basin in the Arctic Ocean (72oN and 83oN) between 2017-2023.
Data exclusions	No data was excluded from the study.
Reproducibility	No measures were taken to verify reproducibility of the finding
Randomization	*Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.*
Blinding	Inviestigators were not blinded. Blinding is not relevant to an observational study on distributions of organisms in the environment.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	Ice covered and ice free regions of the Arctic Ocean
Location	Canad Basin Arctic Ocean (between 72oN and 83oN)
Access & import/export	Samples were collected in Canadian or international waters in the Arctic and no permits were required.
Disturbance	NA

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
☒	☐ Plants

Methods

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

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*