nature portfolio

Corresponding author(s): Denis Tikhonenkov

Last updated by author(s): Oct 25, 2022

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	Sequence quality: FastQC v0.10.1 Transcriptome assembly: Trinity v2.4.0; TransDecoder v5.5.0; Genome assembly: SPAdes v3.14.1, NOVOPlasty v4.3 Sequencing read processing: Trimmomatic v0.36; PEAR v0.9.6; BBMap v37.36
Data analysis	Transcriptome and genome analysis: BLAST v2.2.30+; CD-HIT v4.6; HMMER3.1 (hmmer.org); OGDRAW v1.3.1; RNA2Drawer; BUSCO v3.0.0; PfamScan v1.6 (http://ftp.ebi.ac.uk/pub/databases/Pfam/Tools); SMART v9.0; SignalP v5.0; Trophic Mode Prediction Tool v1.0.0; MFannot (https://megasun.bch.umontreal.ca/apps/mfannot/); NCBI Genome Workbench v3.6.0 Transcriptomic data filtering: BlobTools; DIAMOND v0.9.24; TaxonKit v0.3.0 Phylogenomic and phylogenetic analysis: MAFFT v7.222; MAFFT v7.313; trimAL v1.2; trimAl v1.4; BMGE v1.1.2; SCaFoS v1.2.5; PhyloSuite v1.2.2; IQ-TREE v1.6.8; IQ-TREE v2.0.7; IQ-TREE v1.6.10, IQ-TREE v1.6.12; PhyloBayes MPI v1.8c; PREQUAL v1.02; BaCoCa v1.105.r Comparative genomics: OrthoFinder v2.5.4; KEGG Automatic Annotation Server (KAAS) v2.1 Visualization: Seaborn v0.8.1; MEGA v7.0.21; BioEdit v7.2.5

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

Raw transcriptome reads from Provora are deposited in GenBank (PRJNA866092), along with the SSU rRNA gene sequences of species (OP101998-OP102010). Assembled transcriptomes, mitochondrial genomes, materials of orthogroup and phylogenetic analyses, along with individual gene alignments, concatenated and trimmed alignments, and maximum-likelihood and Bayesian tree files for the phylogenomic dataset are available at figshare with the identifier doi.org/10.6084/m9.figshare.20497143. The following databases were used in this study: NCBI nt database (https://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/nt.gz), NCBI non-redundant database (https://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/nr.gz), Swiss-Prot database (https://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/complete/uniprot_sprot.fasta.gz), EukProt database (https://figshare.com/articles/dataset/EukProt_a_database_of_genome-scale_predicted_proteins_across_the_diversity_of_eukaryotic_life/12417881/2), KEGG database (https://www.genome.jp/kegg/), Pfam database (http://ftp.ebi.ac.uk/pub/databases/Pfam/releases/Pfam32.0/). Environmental sequencing datasets were used for 18S rRNA gene analysis: Tara Oceans (https://zenodo.org/record/3768510#.Y1ZtKuzMI1I), Protists in European coastal waters and sediments (https://doi.org/10.1111/1462-2920.12955), Autonomous Reef Monitoring Structures (ARMS) in Red Sea (https://doi.org/10.1038/s41598-018-26332-5), Stream biofilm eukaryotic assemblages (https://doi.org/10.1016/j.ecolind.2020.106225), Deep sea basin sediments (https://doi.org/10.1038/s42003-021-02012-5), Eukaryotic plankton in reef environments in Panama (https://doi.org/10.1007/s00338-020-01979-7), Eukaryote communities in a high-alpine lake (https://doi.org/10.1007/s12275-019-8668-8), Mountain lake microbial communities (https://doi.org/10.1111/mec.15469), Microbial eukaryotes in lake Baikal (https://doi.org/10.1093/femsec/fix073); 320-gene dataset was used for constructing alignments for phylogenomic analyses (https://static-content.springer.com/esm/art%3A10.1038%2Fs41467-021-22044-z/MediaObjects/41467_2021_22044_MOESM5_ESM.zip). The novel taxa have been registered with the Zoobank database (http://zoobank.org/) urn:lsid:zoobank.org:act: 9EE01A01-E294-415B-A36F-0FB4373183D0, urn:lsid:zoobank.org:act:A54BD0FB-7FA3-42CB-9D3D-2211FA657DC0, urn:lsid:zoobank.org:act: F6395E20-7BDF-4CBE-95FB-E4CE1E7B8185, urn:lsid:zoobank.org:act:F1E8545D-BAC1-44FF-9B6B-8FEE4AC028BB, urn:lsid:zoobank.org:act:66A5C066-890F-4F25-AAB6-5CDCE2028034, urn:lsid:zoobank.org:act:830A4372-62D9-4CE1-BFD8-9FE9EED67FED, urn:lsid:zoobank.org:act:DFE7080B-6201-455A-99CE-903103CBB049, urn:lsid:zoobank.org:act:A230EC14-DC4B-4F05-8D69-8FE0BAB3DE09, urn:lsid:zoobank.org:act:B8894608-40D4-4D16-A4D9-6F448614F22C, urn:lsid:zoobank.org:act:97B89F6F-72D6-482A-9EA7-88E5C63E6EB6.

Human research participants

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

Reporting on sex and gender	This study did not involve human research
Population characteristics	This study did not involve human research
Recruitment	This study did not involve human research
Ethics oversight	This study did not involve human research

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	In this study, we describe ten new strains of microbial predators, which collectively form a diverse new supergroup of eukaryotes Provora. We performed detailed ultrastructural, transcriptomic/genomic, and phylogenomic analyses, and showed that Provora is genetically and morphologically distinct from all other eukaryotes
Research sample	This research describes three new genera and five new species from a new phylum of predatory eukaryotic microbes that is the sister lineage of the Haptista+TSAR assemblage, possibly also including Hemimastigophora. The organisms were collected from marine habitats, including coral reefs, nearshore sediments, and the water column.
Sampling strategy	Sample size is not relevant to the present study.
Data collection	Samples were collected from marine sediments, water column, and corals, and the new organisms were subsequently grown in the

Data collection	laboratory. Microscopic data were recorded by Denis Tikhonenkov. Sequencing data were generated by D. Tikhonenkov. Transcriptome and genome data were assembled by K. Mikhailov and R. Gawryluk.
Timing and spatial scale	Sampling relevant to the present study was carried out seven times: in the Strait of Georgia, British Columbia, June 13, 2017; Arctic waters of the Kara Sea, September 19, 2015; Arctic waters of the East Siberian Sea, September 5, 2017; shoreland of Quarantine Bay, Black Sea, May 13, 2017; sea waters of the Curaçao island, April 24, 2018; Red Sea, Sharm El Sheikh, April 2015; Kazachya Bay, Black Sea, September 1, 2018. We had no reason to expect to find the organisms that we did, so there is no specific rationale to sampling sites.
Data exclusions	Sequencing data from prey organisms were excluded from the analyses for studied predatory protists. To do this, we subtracted transcripts derived from prey (kinetoplastids) and any non-eukaryotic transcripts from the total datasets. The raw data associated with this will be accessible in the raw read files deposited in the NCBI SRA database.
Reproducibility	Microscopic analyses were conducted several times. Phylogenomic analyses were carried out with a number of different approaches (maximum likelihood, Bayesian etc.) and all associated datasets have been made available.
Randomization	Randomization is not relevant to the present study because organisms were not allocated into groups.
Blinding	Blinding was not relevant to the present study.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	Climatic conditions in the field were not recorded and are not relevant to the study.
Location	1) Strait of Georgia, British Columbia, Canada (49°10'366'' N, 123°28'50'' W) 2) Arctic waters of the Kara Sea (75°53'16.8'' N, 89°30'28.8'' E) 3) Arctic waters of the East Siberian Sea (71°27'59.8'' N, 152°53'59.3'' E) 4) Shoreland of Quarantine Bay, Black Sea (44°36'41.4'' N, 33°30'6.2'' E) 5) Eastern point of the Curaçao island (12°12'32.3'' N, 68°48'58.8'' W) 6) Red Sea, Sharm El Sheikh (27°50'50.5'' N, 34°18'59.4'' E) 7) Kazachya Bay, Black Sea (44°34'18.8"N 33°24'40.2"E) 8) Curaçao island (12°12'32.3'' N, 68°48'58.8'' W)
Access & import/export	Habitats were accessed via a research vessel (locations 1, 2, 3), a car (locations 4-7), and a diving boat (lication 8). No permissions were required for sampling in the selected sampling sites.
Disturbance	No disturbances to the sites were caused; we sampled a small amount of water and surface sediment from marine habitats.

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

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	Ten clonal cultures of protists were isolated from marine habitats
Authentication	Phase and DIC contrast light microscopy and 18S rRNA gene sequencing was used for authentication.

Mycoplasma contamination	This is not relevant to protist cell culture.
Commonly misidentified lines (See ICLAC register)	This is not relevant to protist cell culture.

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	The study did not involve laboratory animals.
Wild animals	The study did not involve wild animals (or any animals).
Reporting on sex	This is not relevant to protist cell culture.
Field-collected samples	Cultures of predatiry protists were established by isolating cells with a glass micropipette. Cultures were maintained at room temperature and at +4C. Cultures were propagateed using the kinetoplastid protist Procryptobia sorokini B-69 as prey. The kinetoplastid was grown in marine Schmalz-Pratt's medium or artificial marine medium (RS-R11040, Red Sea) and preyed upon Pseudomonas fluorescens.
Ethics oversight	No ethical approval was required. The organisms described here are novel eukaryotic microbes (protists) that feed on other protists and pose no risk.

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