NCBI accession numbers from a study of two tintinnid ciliate species, Schmidingerella arcuata and Schmidingerella meunieri

Website: https://www.bco-dmo.org/dataset/924260
Data Type: experimental
Version: 1
Version Date: 2024-06-13

Project
» Collaborative Research: Combining single-cell and community 'omics' to test hypotheses about diversity and function of planktonic ciliates (Ciliate Omics)

Contributors	Affiliation	Role
McManus, George	University of Connecticut (UConn)	Principal Investigator
Katz, Laura A.	Smith College	Co-Principal Investigator
Santoferrara, Luciana	Hofstra University	Co-Principal Investigator
Rauch, Shannon	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

Abstract
Many ciliates express their genes only after "unscrambling" them from the genome (portions of the gene may be reversed in the genome or expressed in a different order from what is in the genome). Because the scrambling process would have to be unique to a given species, we can use genome/transcriptome comparisons of single cells to define species boundaries. We showed that two tintinnid ciliate species that look almost identical, Schmidingerella arcuata and S. meunieri, scramble and unscramble their genes differently in the step before transcription, verifying that they represent two distinct biological species. We are also using the single-cell -omics data to better understand the evolutionary relationships (phylogenomics) among tintinnids and related planktonic ciliates. The information in this dataset indicates how to access the sequence data from NCBI.

Table of Contents

Coverage

Location: Northwest Atlantic continental shelf (New England coast) and Puget Sound
Spatial Extent: **N**:48.5 **E**:-72.06 **S**:41.31 **W**:-122.7
Temporal Extent: 2019 - 2019

Methods & Sampling

This project required sequencing whole genomes and transcriptomes from single cells picked from two cultures. One was *Schmidingerella arcuata* (urn:lsid:marinespecies.org:taxname:732664), isolated from the East Coast of the United States in Long Island Sound; the other was *Schmidingerella meunieri* (urn:lsid:marinespecies.org:taxname:732864), isolated by S. Strom from Puget Sound on the West Coast of the United States. *S. arcuata* was collected from the surface waters of northeastern Long Island Sound, CT

(41.31°N, 72.06°W), using a 20-micrometer (μm) mesh plankton net. Single cells were isolated with drawn capillaries and moved to six-well culture plates with 0.2-μm-filtered sample water. The goal was to quantify the degree of difference in genome architecture for these two close congeners.

The cultures were grown in filtered seawater at 18 degrees Celsius (°C) on a 12:12 light cycle and fed the dinoflagellate *Heterocapsa triquetra* and the prymnesiophyte *Isochrysis galbana*. Individual cells were picked with a drawn capillary pipette and processed for sequencing as detailed below (from Smith et al 2020).

The SMART-Seq2 v4 Ultra Low input RNA kit (Cat: 634889; Takara, Mountain View, CA) was used for whole transcriptome amplification (WTA) following the manufacturer's protocols, with the exception that we quartered the reaction volumes. For whole-genome amplification (WGA), the Repli-g single-cell kit (Cat: 150343; Qiagen, Hilden, Germany) was used following the manufacturer's protocols. The products (cDNA for WTA, gDNA for WGA) were quantified with the dsDNA Qubit assay (Invitrogen, Waltham, MA) and polymerase chain reaction-checked with eukaryotic 18S rDNA and genus-specific ITS primers. Minimum bacterial contamination was confirmed by polymerase chain reaction with 16S rDNA primers Sequencing libraries were prepared with the Illumina Nextera XT kit (Cat: FC1311096; Illumina, San Diego, CA), then processed with Illumina HiSeq 2500 at Macrogen Sequencing (Geumcheon-gu, Seoul, South Korea).

The isolation/cultivation was done in 2014-2015. The sequencing work was all completed and analyzed by the suummer of 2020.

Data Processing Description

Raw reads from WTA and WGA sequencing were trimmed for quality and size (Q28 and minimum length of 200 and 1,500 bp, respectively) using BBMap (V38.39). After trimming, two single-cell WTAs were assembled together using rnaSPAdes (V3.13.1), and seven singe-cell WGAs were assembled together using both SPAdes (V3.13.1) and MEGAHIT (V1.2.9). The MEGAHIT genome assembly was used for the final analysis.

Assemblies were processed through custom python scripts (https://github.com/maurerax/KatzLab/tree/HTS-Processing-PhyloGenPipeline) for the removal of rDNA and prokaryotic transcripts, and for the identification of orthologous gene families using USEARCH (V9.2) with OrthoMCLdatabases (V2.0.9).

BCO-DMO Processing Description

- Imported original file "NCBI_data_Schmidingerella.csv" into the BCO-DMO system.
- Removed "N" and "W" from the latitude and longitude columns.
- Made longitude negative to indicate the west direction.
- Renamed fields to comply with BCO-DMO naming conventions.
- Saved the final file as "924260_v1_ncbi_accessions.csv"

[ ]

Related Publications

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., … Pevzner, P. A. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. Journal of Computational Biology, 19(5), 455–477. doi:10.1089/cmb.2012.0021
Software

Bushnell, B. (2014). BBMap: A Fast, Accurate, Splice-Aware Aligner. Lawrence Berkeley National Laboratory. LBNL Report #: LBNL-7065E. Retrieved from https://escholarship.org/uc/item/1h3515gn
Software

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics, 26(19), 2460–2461. https://doi.org/10.1093/bioinformatics/btq461
Software

Fischer, S., Brunk, B. P., Chen, F., Gao, X., Harb, O. S., Iodice, J. B., Shanmugam, D., Roos, D. S., & Stoeckert, C. J. (2011). Using OrthoMCL to Assign Proteins to OrthoMCL-DB Groups or to Cluster Proteomes Into New

Ortholog Groups. Current Protocols in Bioinformatics, 35(1). Portico.
https://doi.org/10.1002/0471250953.bi0612s35
Methods

Li, D., Liu, C.-M., Luo, R., Sadakane, K., & Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics, 31(10), 1674–1676.
https://doi.org/10.1093/bioinformatics/btv033
Software

Smith, S. A., Maurer-Alcalá, X. X., Yan, Y., Katz, L. A., Santoferrara, L. F., & McManus, G. B. (2020). Combined Genome and Transcriptome Analyses of the Ciliate Schmidingerella arcuata (Spirotrichea) Reveal Patterns of DNA Elimination, Scrambling, and Inversion. Genome Biology and Evolution, 12(9), 1616–1622.
https://doi.org/10.1093/gbe/evaa185
Results

Smith, S. A., Santoferrara, L. F., Katz, L. A., & McManus, G. B. (2022). Genome architecture used to supplement species delineation in two cryptic marine ciliates. Molecular Ecology Resources, 22(8), 2880–2896.
Portico. https://doi.org/10.1111/1755-0998.13664
Results

[table of contents | back to top]

Related Datasets

IsRelatedTo

Smith, S., Santoferrara, L. F., Katz, L., & McManus, G. B. (2022). *Genome architecture used to supplement species delineation in two cryptic marine ciliates (study on congeneric tintinnid ciliates Schmidingerella arcuata and Schmidingerella meunieri (published as Smith et al. 2022 in Molecular Ecology Resources)* [Data set]. figshare. https://doi.org/10.6084/M9.FIGSHARE.16892893 https://doi.org/10.6084/m9.figshare.16892893

University of Connecticut. Schmidingerella arcuata isolate:LIS Genome sequencing and assembly. 2022/06. In: BioProject [Internet]. Bethesda, MD: National Library of Medicine (US), National Center for Biotechnology Information; 2011-. Available from: http://www.ncbi.nlm.nih.gov/bioproject/PRJNA829814. NCBI:BioProject: PRJNA829814.

University of Connecticut. Schmidingerella arcuata isolate:SAS-2020, Genome and transcriptome of marine ciliate Schmidingerella arcuata. 2020/09. In: BioProject [Internet]. Bethesda, MD: National Library of Medicine (US), National Center for Biotechnology Information; 2011-. Available from: http://www.ncbi.nlm.nih.gov/bioproject/PRJNA626068. NCBI:BioProject: PRJNA626068.

University of Connecticut. Schmidingerella meunieri isolate:PAC Genome sequencing and assembly. 2022/06. In: BioProject [Internet]. Bethesda, MD: National Library of Medicine (US), National Center for Biotechnology Information; 2011-. Available from: http://www.ncbi.nlm.nih.gov/bioproject/PRJNA828278. NCBI:BioProject: PRJNA828278.

[table of contents | back to top]

Parameters

Parameter	Description	Units
species	Scientific name (Genus species) or "primer"	unitless
latitude	latitude of location from which species was isolated; positive values = North	decimal degrees
longitude	longitude of location from which species was isolated; negative values = West	decimal degrees
depth	depth from which ciliate was isolated	meters
date	year of ciliate isolation	unitless
NCBI_BioprojectProject_ID	NCBI BioProject identifier	unitless
Sample_Accession	NCBI BioSample identifier	unitless
Sequence_Accession	Accession numbers at NCBI (sample or sequence) (Genbank or SRA database)	unitless
sample	sequence locus or library type	unitless
other_links	web links to data illustrations on figshare website	unitless

Instruments

Dataset-specific Instrument Name	Illumina HiSeq 2500
Generic Instrument Name	Automated DNA Sequencer
Generic Instrument Description	General term for a laboratory instrument used for deciphering the order of bases in a strand of DNA. Sanger sequencers detect fluorescence from different dyes that are used to identify the A, C, G, and T extension reactions. Contemporary or Pyrosequencer methods are based on detecting the activity of DNA polymerase (a DNA synthesizing enzyme) with another chemoluminescent enzyme. Essentially, the method allows sequencing of a single strand of DNA by synthesizing the complementary strand along it, one base pair at a time, and detecting which base was actually added at each step.

Dataset-specific Instrument Name	20-μm-mesh plankton net
Generic Instrument Name	Phytoplankton Net
Generic Instrument Description	A Phytoplankton Net is a generic term for a sampling net having mesh size of 150 microns or less that is used to collect phytoplankton. It is used only when detailed instrument documentation is not available.

Project Information

Collaborative Research: Combining single-cell and community 'omics' to test hypotheses about diversity and function of planktonic ciliates (Ciliate Omics)

Website: http://microzooplankton.uconn.edu

Coverage: New England continental shelf

NSF Award Abstract:
Planktonic ciliates are key members of marine food webs where they serve diverse roles, including as food chain links between smaller microbes and larger plankton. Due to their small size and difficulties in identifying and cultivating them, we know less about ciliate diversity and distributions in the ocean than we do about larger organisms such as fish and invertebrates. Previous work from this team measured ciliate diversity in coastal waters and found that distinct genetic variants were separated in time and space in a way that could be related to factors such as ocean temperature, salinity, and depth gradients. Many questions remained unanswered, and it is important to understand the environmental factors that control the diversity and distribution of plankton such as ciliates to predict how these organisms may respond to a changing enviroment in the coming decades. This project focuses on: 1) how ciliate species are delineated using single-cell genomics and transcriptomics; 2) DNA-based studies of all ciliates and other planktonic members of the SAR clade (Stramenopila, Alveolata, Rhizaria), which will provide ecological context; 3) in situ gene expression by single-cell and meta- transcriptomics; and 4) laboratory studies of gene expression in cultivated ciliate species. This project involves training of postdoctoral scholars, graduate students, and undergraduates. The researchers are committed to creating diverse and inclusive research labs; recruitment of participants will be done through partnership with appropriate groups on our campuses. The project integrates with summer Research Experiences for Undergraduates (REU) activities at both Smith College and UCONN (including the UCONN/Mystic Aquarium joint REU), which are especially focused on underrepresented students. This project also enhances efforts to broaden understanding of biodiversity in partnership with the UCONN Noyce Scholars Program, which facilitates career-changing STEM professionals to become teachers in underserved secondary schools.

This project will assess distributions of reproductively-isolated species, determined using a new method to characterize regions of the ciliate germline genome. Furthermore, it will use phylogenomic methods to identify clade-specific transcripts (e.g. those of spirotrich ciliates) within metatranscriptomes from the shelf environment and to expand knowledge of ciliate function with single-cell transcriptomics of field-collected cells. These approaches will be a substantial improvement over the culture-based methods that are potentially biased towards "weedy" species in the ocean. The combination of definitive species identification with assessment of function via single-cell and meta- transcriptomics promises to provide significant advances in marine plankton ecology. The research focuses on two broad questions: 1) does the observed high diversity in phylogenetically-informative genes reflect reproductive isolation and functional differentiation in planktonic ciliates? and 2) do different co-occurring species of planktonic ciliates show substantial functional differences that correspond to different niches in the ocean? The project assesses species boundaries (i.e. reproductive isolation) through analyses of patterns in the germline micronuclei of planktonic ciliate morphospecies; characterizes transitions of closely-related ciliates across ecological gradients in the ocean; and examines functional differences within and between species, and in communities, through analyses of transcriptomics.

This award reflects NSF's statutory mission and has been deemed worthy of support through evaluation using the Foundation's intellectual merit and broader impacts review criteria.

[table of contents | back to top]

Funding

Funding Source	Award
NSF Division of Ocean Sciences (NSF OCE)	OCE-1924570
NSF Division of Ocean Sciences (NSF OCE)	OCE-1924527

[table of contents | back to top]