Metabarcoding data from samples collected at shore-based tide pools and ocean samples in New England waters in 2019

Website: https://www.bco-dmo.org/dataset/988266
Data Type: Other Field Results
Version: 1
Version Date: 2025-11-04

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

Contributors	Affiliation	Role
Katz, Laura A.	Smith College	Principal Investigator
York, Amber D.	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

Abstract
This dataset contains sampling and genetic accession information for metabarcoding data from bottle samples collected from shore-based tide pool and ocean sample collections in Acadia National Park, Maine in 2019. Samples were filtered for metabarcoding. Metabarcoding relied on taxon-specific primers to characterize ciliates and close relatives. Marine metagenome RNA and DNA sequences are available in the Sequence Read Archive (SRA) at the National Center for Biotechnology Information (NCBI) under BioProject PRJNA952215.

Table of Contents

Coverage

Location: Acadia National Park, ME tide pool and ocean samples
Spatial Extent: **Lat**:44.2283 **Lon**:-68.3125
Temporal Extent: 2019-05-20 - 2019-10-14

Dataset Description

See the "Related Datasets" section for methods and data from additional datasets from this study (CTD profile data and single-cell transcriptomic data).

Methods & Sampling

Sample type: Amplicon = communities collected on filters at varying sizes and amplified using primers as described in Grow et al. (2023, doi:10.3354/ame02003).

Metabarcoding relied on taxon-specifc primers to characters ciliates and close relatives.

Raw sequences are available in the Sequence Read Archive (SRA) at the National Center for Biotechnology

Information (NCBI) under BioProject https://www.ncbi.nlm.nih.gov/bioproject/**PRJNA952215**
*** BioProject Title: "**Exploring the diversity of microeukaryotic communities in New England tide pools"
*** BioProject Description:** "Extraction of community DNA & RNA from Acadia National Park, ME tide pool and ocean samples, amplification of SSU rRNA using SAR (Stramenopila, Alveolata, Rhizaria) primers, metabarcoding sequencing of amplified DNA."

Data Processing Description

Scripts developed under this study to analyze the NCBI data can be found at the GitHub (see Wiki https://github.com/Katzlab/EukPhylo/wiki), with versions of scripts developed through this project from 2020-2024. A full version of EukPhylo version 1 is published on Zenodo (doi: 10.5281/ZENODO.15866075). The parent DOI for all versions of the github repository, including past releases of PhyloToI is doi:10.5281/ZENODO.13323347.

BCO-DMO Processing Description

* Sheet 1 of submitted file "Katz_metabarcoding.xlsx" was imported into the BCO-DMO data system for this dataset. Values "NA" imported as missing data values. Table will appear as Data File: 988266_v1_marine-metabarcoding.csv (along with other download format options).

Missing Data Identifiers:
* In the BCO-DMO data system missing data identifiers are displayed according to the format of data you access. For example, in csv files it will be blank (null) values. In Matlab .mat files it will be NaN values. When viewing data online at BCO-DMO, the missing value will be shown as blank (null) values.

* Column names adjusted to conform to BCO-DMO naming conventions designed to support broad re-use by a variety of research tools and scripting languages. [Only numbers, letters, and underscores. Can not start with a number]

* NCBI Run Selector data were used to extract data from BioProject PRJNA952215. The Date, Instrument, and Depth columns were added from the NCBI metadata as the dates matched the provided excel data but were in a consistent (ISO) format.

* Time column renamed Time_local

* Additional column added: ISO_DateTime_UTC (ISO datetime with timezone) in UTC. Converted from Date and Time_Local (US Eastern EST/EDT).

* Column lat_lon separated into Lat,Lon columns in decimal degrees.

* Second "Sample_Type" column renamed Sample_Site_Type. The other Sample_Type contained RNA or DNA indication.

[table of contents | back to top]

Related Publications

Auden Cote-L'Heureux, Godwin N. Ani, Katzlab, Adri K. Grow, MCLeleu, Elinor Sterner, GiuliaRibeiro, & rebeccagawron. (2025). *Katzlab/EukPhylo: EukPhylo version 1.0* (Version v1.0) [Computer software]. Zenodo. https://doi.org/10.5281/ZENODO.15866075 https://doi.org/10.5281/zenodo.15866075
Software

Grow, A., Sleith, R., Sehein, T., Labare, M., & Katz, L. (2023). Exploring the diversity of microeukaryotic communities in New England tide pools. Aquatic Microbial Ecology, 89, 143–155. https://doi.org/10.3354/ame02003
Results

Katzlab. (2025). EukPhylo version 1.0 wiki. GitHub. https://github.com/Katzlab/EukPhylo/wiki
Methods

Shazib, S. U. A., Ahsan, R., Leleu, M., McManus, G. B., Katz, L. A., & Santoferrara, L. F. (2025). Phylogenomic workflow for uncultivable microbial eukaryotes using single-cell RNA sequencing − A case study with planktonic ciliates (Ciliophora, Oligotrichea). Molecular Phylogenetics and Evolution, 204, 108239. https://doi.org/10.1016/j.ympev.2024.108239
Results

[table of contents | back to top]

Related Datasets

IsRelatedTo

Katz, L. A. (2025) **Single-cell transcriptomic data from ciliates isolated in New England waters between 2019 and 2023.** Biological and Chemical Oceanography Data Management Office (BCO-DMO). (Version 1) Version Date 2025-11-04 http://lod.bco-dmo.org/id/dataset/988253 [view at BCO-DMO]
Relationship Description: The Single-cell transcriptomics dataset (988253) and metabarcoding dataset (988266) are the omics data associated with the collections from the CTD dataset (879380). Both omics datasets used the same bottle samples collected with the CTD profiles.

McManus, G., Santoferrara, L., Katz, L. A. (2022) **CTD profiles collected with RV Connecticut on the Continental shelf and slope south of Montauk, NY on 14-15 June 2022.** Biological and Chemical Oceanography Data Management Office (BCO-DMO). (Version 1) Version Date 2022-09-16 doi:10.26008/1912/bco-dmo.879380.1 [view at BCO-DMO]
Relationship Description: The Single-cell transcriptomics dataset (988253) and metabarcoding dataset (988266) are the omics data associated with the collections from the CTD dataset (879380). Both omics datasets used the same bottle samples collected with the CTD profiles.

Smith College (2023). Exploring the diversity of microeukaryotic communities in New England tide pools. 2023/04. In: NCBI:BioProject: PRJNA952215 [Internet]. Bethesda, MD: National Library of Medicine (US), National Center for Biotechnology Information; Available from: http://www.ncbi.nlm.nih.gov/bioproject/PRJNA952215.

[table of contents | back to top]

Parameters

Parameter	Description	Units
SRA_Run	NCBI Sequence Read Archive (SRA) Run accession	unitless
SRA_Study	NCBI Sequence Read Archive (SRA) Study accession	unitless
bioproject_accession	NCBI BioProject accession	unitless
biosample_accession	NCBI BioSample accession	unitless
LKH_Number	LKH Number. In house tracking system sample identifier.	unitless
Sample_ID	Sample identifier	unitless
Year	Sampling year	unitless

Month	Sampling month (Full month name)	unitless
Sampling_Day	Sampling day	unitless
Time_local	Sampling time. Local time zone (EDT/EST)	unitless
ISO_DateTime_UTC	Sampling datetime with timezone (ISO 8601 format). UTC	unitless
Tide	Tide (e.g. Ebbing or Flooding)	unitless
Sample	Sample number	unitless
Filter_Size	Filter Size	unitless
Sample_Type	Sample type (RNA or DNA)	unitless
Sampling_Location	Sampling location description (e.g. ANP)	unitless
Sample_Site_Type	Sample site description (e.g. Ocean, middle pool)	unitless
Location_Details	Location description (geolocation)	unitless
GPS_Lat	Latitude	decimal degrees
GPS_Lon	Longitude	decimal degrees
Water_Temperature	Water temperature	degrees Celsius
Salinity	Salinity	parts per thousand (ppt)
pH	pH	pH scale
Depth	Depth	meters (m)
Instrument	Instrument name	unitless

[table of contents | back to top]

Instruments

Dataset-specific Instrument Name	Illumina NovaSeq 6000
Generic Instrument Name	Automated DNA Sequencer
Generic Instrument Description	A DNA sequencer is an instrument that determines the order of deoxynucleotides in deoxyribonucleic acid sequences.

[table of contents | back to top]

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 Environmental Biology (NSF DEB)	DEB-1651908
NSF Division of Environmental Biology (NSF DEB)	DEB-1541511

[table of contents | back to top]