

Metagenomic sample information, genetic accession identifiers (NCBI SRA, JGI IMG), and estimated gene copies from Orcas Island coastal waters (2 m depth) from 2021-05-27 to 2021-06-18

Website: <https://www.bco-dmo.org/dataset/984169>

Data Type: Other Field Results

Version: 1

Version Date: 2025-09-10

Project

» [Collaborative Research: Rhythm and Blooms: Deciphering metabolic, functional and taxonomic interactions over the life cycle of a phytoplankton bloom](#) (Rhythm and Blooms)

Contributors	Affiliation	Role
Kubaneck, Julia	Georgia Institute of Technology (GA Tech)	Principal Investigator
Nunn, Brook L.	University of Washington (UW)	Principal Investigator
Rynearson, Tatiana A.	University of Rhode Island (URI)	Principal Investigator
Mudge, Miranda	University of Washington (UW)	Scientist
Timmins-Schiffman, Emma	University of Washington (UW)	Scientist, Data Manager
Bartlett, Evelyn	University of Washington (UW)	Student
York, Amber D.	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

Abstract

This dataset contains NCBI Sequence Read Archive (SRA) accession numbers, DOE JGI Integrated Microbial Genomes & Microbiome (IMG/M) IDs, and estimated gene copies for metagenomic samples collected at Orcas Island, WA, USA Coastal Ocean (2m depth) from 5/27/21 to 6/18/21 collected as part of the following study. Study abstract: Floating, single-celled algae, or phytoplankton, form the base of marine food webs. When phytoplankton have sufficient nutrients to grow quickly and generate dense populations, known as blooms, they influence productivity of the entire food web, including rich coastal fisheries. The present research explores how the environment (nutrients) as well as physical and chemical interactions between individual cells in a phytoplankton community and their associated bacteria act to control the timing of bloom events in a dynamic coastal ecosystem. The work reveals key biomolecules within the base of the food web that can inform food web functioning (including fisheries) and be used in global computational models that forecast the impacts of phytoplankton activities on global carbon cycling. A unique set of samples and data collected in 2021 and 2022 that captured phytoplankton and bacterial communities before, during, and after phytoplankton blooms, is analyzed using genomic methods and the results are used to interrogate these communities for biomolecules associated with blooms stages. The team mentors undergraduates, graduate students, and postdoctoral researchers in the fields of biochemical oceanography, genome sciences, and time-series multivariate statistics. University of Washington organized hackathons to develop publicly accessible portals for the simplified interrogation and visualization of 'omics data, accessible to high schoolers and undergraduates. These portals are implemented in investigator-led undergraduate teaching modules in the University of Rhode Island Ocean Classroom. The research team also returns to Orcas Island, WA, where the field sampling takes place, to host a series of annual Science Weekends to foster scientific engagement with the local community. Phytoplankton blooms, from initiation to decline, play vital roles in biogeochemical cycling by fueling primary production, influencing nutrient availability, impacting carbon sequestration in aquatic ecosystems, and supporting secondary production. In addition to influences from environmental conditions, the physical and chemical interactions among planktonic microbes can significantly modulate blooms, influencing the growth, maintenance, and senescence of phytoplankton. Recent work in steady-state open ocean ecosystems has shown that important chemicals are transferred amongst plankton on time-dependent metabolic schedules that are related to diel cycles. It is unknown how these metabolic schedules operate in dynamic coastal environments that experience perturbations, such as phytoplankton blooms. Here, the investigators are examining metabolic scheduling using long-term, diel sample sets to reveal how chemical and biological signals associated with the initiation, maintenance, and cessation of phytoplankton blooms are modulated on both short (hrs) and long (days-weeks) time scales. Findings are advancing the ability to predict and manage phytoplankton dynamics, providing crucial insights into ecological stability and future oceanographic sampling strategies. Additionally, outcomes of this study are providing a new foundational understanding of the succession of microbial communities and their chemical interactions across a range of timescales. In the long term, this research has the potential to identify predictors of the timing of phytoplankton blooms, optimize fisheries management, and guide future research on carbon

Table of Contents

- [Coverage](#)
 - [Dataset Description](#)
 - [Methods & Sampling](#)
 - [Data Processing Description](#)
 - [BCO-DMO Processing Description](#)
 - [Problem Description](#)
 - [Data Files](#)
 - [Supplemental Files](#)
 - [Related Publications](#)
 - [Related Datasets](#)
 - [Parameters](#)
 - [Instruments](#)
 - [Project Information](#)
 - [Funding](#)
-

Coverage

Location: Coastal waters, East Sound, WA, USA, depth 2 m

Spatial Extent: **Lat:**48.6765 **Lon:**-122.883233

Temporal Extent: 2021-05-27 - 2021-06-18

Dataset Description

Additional funding description:

This dataset was supported by NSF OCE-2401646, OCE-2401645, OCE-2401644, University of Washington Royalty Research Fund, NIH NIEHS grant R21ES034337-01, NSF IOS-2041497, NIH fellowship F31 ES032733-01A1

See the "Related Datasets" section for descriptions of related holdings at NCBI (umbrella BioProject PRJNA1093221), JGI (GOLD study Gs0160720), Zenodo (raw data), PanoramaWeb, PRIDE/ProteomeXchange (PXD052873), etc.

DOE = U.S. Department of Energy

JGI = Joint Genome Institute (DOE)

IMG/M = Integrated Microbial Genomes & Microbiomes (DOE JGI)

GOLD = Genomes Online Database (DOE JGI)

NCBI = National Center for Biotechnology Information

SRA = Sequence Read Archive (NCBI)

NSF = U.S. National Science Foundation

OCE = Division of Ocean Sciences (NSF)

IOS - Division of Integrative Organismal Systems (NSF)

NIH = National Institutes of Health

NIEHS = National Institute of Environmental Health Sciences (NIH)

PRIDE = PRoteomics IDentifications Database (EMBL-EBI)

Methods & Sampling

The 1–2 L of pre-filtered water was collected and then passed through a 47 mm × 0.22 µm polyethersulfone filter (Tisch, part number SF15021) using the Alexis peristaltic field pump (Proactive Environmental) at a flow rate of <0.1 L min⁻¹ to isolate the bacterial fraction. Filters were immediately placed in plastic resealable baggies (5 × 4 cm), then 200 µl of DNA/RNA Shield (Zymo Research) was added to the whole filter within 1 minute of collection and filter baggies were transferred to liquid nitrogen while in the field. Samples were transferred to the University of Washington and placed in a –80 °C freezer prior to extraction.

Estimated Gene copies (Supplemental File: 984169_v1_estimated_genome_copies.csv):

Reads were then quality controlled for length (> 50 bp), quality (> Q30), and any adapter contamination was removed along with mateless pairs. Filtered reads were error-corrected using `bbcms` v.38.90 (Bushnell, 2014; <https://bbmap.org>). This was run with the following command line options: `bbcms.sh -Xmx100g metadatafile = counts.metadata.json mincount = 2 highcountfraction = 0.6 in = bbcms.input.fastq.gz out1 = input.corr.left.fastq.gz out2 = input.corr.right.fastq.gz`. The readset was assembled with `metaSPAdes` v. 3.15.2 (Nurk, 2017). This was run using the following command line options: `spades.py -m2000-tmp-dir cromwell_root -o spades3-only-assembler -k 33,55,77,99,127-meta -t 16-1 input.corr.left.fastq.gz -2 input.corr.right.fastq.gz`. The input read set was mapped to the final assembly and coverage information generated with `bbmap` v. BMap:38.86 (Bushnell, 2014; <https://bbmap.org>). This was run using the following command line options: `bbmap.sh build = 1 overwrite = true fastreadlen = 500 -Xmx100g threads = 16 nodisk = true interleaved = true ambiguous = random rgid = filename in = reads.fastq.gz ref = reference.fastaout = pairedMapped.bam`. Following assembly, the metagenomes underwent processing through the DOE JGI Metagenome Annotation Pipeline (MAP version 5.1.13) and were subsequently loaded into the Integrated Microbial Genomes and Microbiomes (IMG/M) platform (Clum, 2021). Complete metagenomic datasets for each time point are listed by individual IMG Genome IDs (under the Study Name "Seawater microbial communities from East Sound, Orcas Island, WA, USA" (<https://gold.jgi.doe.gov/study?id=G0160720>)). For each time point, estimated gene copies for all classes in the domain "Bacteria" were downloaded (3.14.2024; Nunn et al. (2024) Dataset 5). At each time point, the cumulative count of gene copies was calculated, and subsequently, each taxonomic class was divided by the total count to determine the class distribution (Nunn et al. (2024) Fig. 4, Dataset 5).

Methodology is from the results paper Nunn et al. (2024, doi: 10.1038/s41597-024-04013-5).

Data Processing Description

DNA extraction was performed using Qiagen's DNeasy PowerWater kit following the manufacturer's instructions. Quant-iT DNA Assay Kits (Invitrogen) were used to generate accurate DNA quantitation (at the recommended molecular weight) in conjunction with a fluorescent plate reader (Varioskan Lux, SkanIt Software 7.0.2). On average, 1.3 µg of high-quality DNA was isolated from each time point collected and 300 ng of isolated DNA was aliquoted into DNA-free tubes (Matrix, catalog #3743) and sent to the Department of Energy (DOE) Joint Genome Institute (JGI) for sequencing.

Of the 133 time points collected between May 27 (13:00) to June 18 (13:00) of 2021, a total of 128 DNA samples were successfully sequenced at the Department of Energy (DOE) Joint Genome Institute (JGI) using Illumina technology.

Data processing is from the results paper Nunn et al. (2024, doi: 10.1038/s41597-024-04013-5).

These tools (from M. Riffle, JGI) describe the data processing done by JGI to produce the metagenomics dataset. These are good references for understanding the data workflow for the metagenomics data.

* <https://github.com/mriffle/nf-filter-fasta>

* <https://www.yeastrc.org/metagomics>

BCO-DMO Processing Description

All sheets in submitted excel files were exported to csv in preparation for import into BCO-DMO.

Data file "984169_v1_metagenome.csv" history:

The following sheets were combined into the resulting table attached to this dataset

984169_v1_metagenome.csv:

* NCBI and JGI Linking.xlsx (Sheet1).

* Nunn_OrcasIsland_Data_JGI_metadata.xlsx (Sheets 1 and 2). These data were joined using the identifiers supplied in lookup table "NCBI and JGI Linking.xlsx"

* ISO DateTime with timezone (UTC) column added in ISO 8601 format, converted from DateTime_PT (time in US/Pacific time zone PST/PDT).

Supplemental file history:

The csv export of the lookup table "NCBI and JGI Linking.xlsx" was attached to this dataset as a supplemental file "NCBI_and_JGI_Linking.csv"

* Sheet 3 of Nunn_OrcasIsland_Data_JGI_metadata.xlsx was exported as csv and attached as a supplemental file:

"984169_v1_estimated_genome_copies.csv" with an additional column added with DateTime with timezone in UTC. Column names were not edited in this table except spaces were converted to underscores.

* The data submitter indicated the names in the dataset should match the names as published in the corresponding literature (results publications). Taxonomic names with current standing as synonyms or misspellings at LPSN were not corrected to be the names as shown at NCBI (or correct name shown at LPSN as of 2025-09-10). A supplemental file "name_matches_and_ids.csv" was added with more information about the names used in this dataset as well as information in the Problems/Issues section.

The originally submitted file "Nunn_OrcasIsland_Data_JGI_metadata.xlsx" was attached as a supplemental file with no additions or changes.

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]

Problem Description

Bacterial names and standing change frequently due to reclassification, synonymy, or spelling corrections, which makes consistency across literature, registries, and datasets difficult.

In 984169_v1_estimated_genome_copies.csv, bacterial class names are given as published in the corresponding results papers. A supplemental file, name_matches_and_ids.csv, provides additional details on the names used. See that file description for more information.

The List of Prokaryotic names with Standing in Nomenclature (LPSN, <https://lpsn.dsmz.de/>) can be consulted for the current standing of names used at NCBI, literature, and in this dataset.


[[table of contents](#) | [back to top](#)]

Data Files

File	
984169_v1_metagenome.csv	(Comma Separated Values (.csv), 80.17 KB) MD5:53e77f79a7b49cc284db86c5e161ae48
Primary data file for dataset ID 984169, version 1. This is a combined table that contains columns with information about the metagenomic samples, NCBI Sequence Read Archive accession information, and JGI IMG Genome IDs.	

[[table of contents](#) | [back to top](#)]

Supplemental Files

File	
984169_v1_estimated_genome_copies.csv	(Comma Separated Values (.csv), 67.67 KB) MD5:ae32d17c6bcb0cc45d748a6108650ead
<p>Supplemental table containing the estimated genome copies with columns by taxon (bacteria class). This table contains columns: _Genome_ID = Contains the JGI sample ID of the specified timepoint DateID_PT = Contains the DateID of the sample (time zone: US/Pacific (PST/PDT)) ISO_DateTime_UTC = Contains the Datetime with timezone of the sample (time zone: UTC) Followed by an estimated gene copy column per bacteria class. See name_matches_and_ids.csv for more information about class names along with identifiers.</p>	
name_matches_and_ids.csv	(Comma Separated Values (.csv), 10.25 KB) MD5:d30c0e01def39c42f5ea77ae250c633b
<p>Information about the bacteria class names used in the estimated Gene Copy table: 984169_v1_estimated_genome_copies.csv. These names were matched to NCBI:txid identifiers on 2025-09-10 using the Global Names Verifier (GNV) . Additional links to the List of Prokaryotic names with Standing in Nomenclature (LPSN, https://lpsn.dsmz.de/) were provided when the NCBI differed from the name as provided in this dataset.</p> <p>This supplemental name table includes columns: Name_in_dataset, The bacteria class name as used in this dataset Exact_Match, Did the name match the exact name at NCBI? (Yes, or No with more details). MatchedName_at_NCBI, The matched name as it appears at NCBI. NCBI_Taxonid_for_MatchedName, The identifier (NCBI:txid) for the MatchedName_at_NCBI ClassificationPath_at_NCBI, The current classification shown at NCBI (as of 2025-09-18) LPSN_page_for_name_in_dataset, The page containing information about the Name_in_dataset at LPSN (https://lpsn.dsmz.de/) which may clarify if it is a misspelling, or a synonym of some kind.</p> <p>Note: LPSN can be consulted for information about the standing of names used at NCBI and literature. NCBI provides this disclaimer on taxonomy pages: "The NCBI taxonomy database is not an authoritative source for nomenclature or classification - please consult the relevant scientific literature for the most reliable information."</p>	
NCBI_and_JGI_Linking.csv	(Comma Separated Values (.csv), 4.93 KB) MD5:6f508b462c783a7971cc581e0b10facd
<p>Supplemental file containing NCBI SRA accession identifiers, JGI IMG accession identifiers, and the DateID. This table can be used as a lookup table to connect the NCBI and JGI holdings.</p>	
Nunn_OrcasIsland_Data_JGI_metadata.xlsx	(Microsoft Excel, 124.89 KB) MD5:03eea3de7d936eedb8c51933b7ac4945
<p>This file contains the same data as the combined table 984169_v1_metagenome.csv and the supplemental file 984169_v1_metagenome-estimated-genome-copies.csv. In this Excel file, the data are separated into three tables:</p> <p>Orcas Island, WA, USA 2021 Coastal Ocean (2m depth) Time Series Sheets:</p> <p>Sheet1 - NCBI Sequence Read Archive Accession Numbers for Metagenomic Samples Sheet2 - IMG Genome IDs for Metagenomic Samples Sheet3 - Estimated Gene Copy for Metagenomic Samples</p>	

[[table of contents](#) | [back to top](#)]

Related Publications

Bushnell, B. (2014) BBMap: A Fast, Accurate, Splice-Aware Aligner. <https://bbmap.org>
Software

Clum, A., Huntemann, M., Bushnell, B., Foster, B., Foster, B., Roux, S., Hajek, P. P., Varghese, N., Mukherjee, S., Reddy, T. B. K., Daum, C., Yoshinaga, Y., O'Malley, R., Seshadri, R., Kyrpides, N. C., Elae-Fadrosch, E. A., Chen, I.-M. A., Copeland, A., & Ivanova, N. N. (2021). DOE JGI Metagenome Workflow. *MSystems*, 6(3).
<https://doi.org/10.1128/msystems.00804-20> <https://doi.org/10.1128/mSystems.00804-20>
Methods

Mudge, M. C., Riffle, M., Chebli, G., Plubell, D. L., Ryneerson, T. A., Noble, W. S., Timmins-Schiffman, E., Kubanek, J., & Nunn, B. L. (2025). Harmful algal blooms are preceded by a predictable and quantifiable shift in the oceanic microbiome. *Nature Communications*, 16(1). <https://doi.org/10.1038/s41467-025-59250-y>
Results

Nunn, B. L., Timmins-Schiffman, E., Mudge, M. C., Plubell, D. L., Chebli, G., Kubanek, J., Riffle, M., Noble, W. S., Harvey, E., Nunn, T. A., Ryneerson, T., Huntemann, M., LaButti, K., Foster, B., Roux, S., Palaniappan, K.,

Mukherjee, S., Reddy, T. B. K., ... Elie-Fadrosh, E. A. (2024). Microbial Metagenomes Across a Complete Phytoplankton Bloom Cycle: High-Resolution Sampling Every 4 Hours Over 22 Days. *Scientific Data*, 11(1). <https://doi.org/10.1038/s41597-024-04013-5>

Results

Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Research*, 27(5), 824-834. <https://doi.org/10.1101/gr.213959.116>

Methods

Riffle, M. (2022). metaGOMics (Version 0.2.0) [Web application]. University of Washington.

<https://meta.yeastrc.org/metagomics/>

Software

Riffle, M. (2025). nf-filter-fasta: Nextflow workflow for creating a small FASTA from a large FASTA based on hits found by comet [Computer software]. GitHub. <https://github.com/mriffle/nf-filter-fasta>

Software

Thermo Fisher Scientific. (n.d.). SkanIt Software for Microplate Readers (version 7.0.2) [Software]. Retrieved from <https://www.thermofisher.com/us/en/home/life-science/lab-equipment/microplate-instruments/plate-readers/software.html>

<https://www.thermofisher.com/us/en/home/life-science/lab-equipment/microplate-instruments/plate-readers/software.html>

Software

[[table of contents](#) | [back to top](#)]

Related Datasets

IsRelatedTo

Brook Nunn. (2023). Investigating the interactions of a phytoplankton community and its microbiome on a 4-hour timescale to reveal emerging and predictive properties across an algal bloom and bust cycle (Proposal ID: 509101) [Project information]. DOE Joint Genome Institute.

<https://genome.jgi.doe.gov/portal/Invtheabustcycle/Invtheabustcycle.info.html>

DOE Joint Genome Institute (2024). Orcas Island East Sound 22 day time series of the ocean microbiome Metagenome. 2024/03. NCBI:BioProject: PRJNA1093221. [Internet]. Bethesda, MD: National Library of Medicine (US), National Center for Biotechnology Information; Available from:

<http://www.ncbi.nlm.nih.gov/bioproject/PRJNA1093221>

Genomes OnLine Database (2024). GOLD study Gs0160720: 22 day time series of the seawater microbial communities from East Sound, Orcas Island, WA, USA. Genomes OnLine Database.

<https://gold.jgi.doe.gov/study?id=Gs0160720>

L. Nunn, B., Timmins-Schiffman, E., Mudge, M. C., Plubell, D. L., Chebli, G., Kubanek, J., Riffle, M., Noble, W. S., Harvey, E., Nunn, T., Huntemann, M., Clum, A., Foster, B., Foster, B., Roux, S., Palaniappan, K., Mukherjee, S., Reddy, T. B. K., Daum, C., ... Elie-Fadrosh, E. A. (2024). *Microbial Metagenomes across a Full Phytoplankton Bloom: High-Resolution Sampling Every 4 Hours for 22 Days* [Data set]. figshare.

<https://doi.org/10.6084/M9.FIGSHARE.26882737>

Mudge, Miranda, & Nunn, Brook L. (2024). High-resolution metaproteomic analysis of an ocean microbiome preceding a Chaetoceros bloom [Data set]. Panorama Public. <https://doi.org/10.6069/GN0Q-VH43>

<https://doi.org/10.6069/gn0q-vh43>

Nunn, B. & Timmins-Schiffman, E. (2025). Nunn-Lab/Publication-2021-Orcas-Island-Time-Series: Nat.Comm Harmful algal blooms are preceded by a predictable and quantifiable shift in the oceanic microbiome (Orcas_metaproteomics). Zenodo. <https://doi.org/10.5281/zenodo.14976385>

<https://doi.org/10.5281/zenodo.14976385>

Nunn, B. L. (2025). High-resolution metaproteomic analysis of an ocean microbiome six days preceding a Chaetoceros bloom (Dataset PXD052873) [Data set]. ProteomeXchange via PRIDE.

<https://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD052873>

Nunn, B. L., Kubanek, J., Rynearson, T. A., Timmins-Schiffman, E., Mudge, M., Bartlett, E. (2025) **Environmental YSI EX01 Sonde Probe data from Orcas Island, WA, USA Coastal Ocean (2m depth) from 2021-05-27 to 2021-06-18**. Biological and Chemical Oceanography Data Management Office (BCO-DMO). (Version 1) Version Date 2025-09-10 doi:10.26008/1912/bco-dmo.984153.1 [[view at BCO-DMO](#)]

Relationship Description: These data were used in results publications from this project Nunn et al (2024, doi: 10.1038/s41597-024-04013-5) and Mudge et al. (2025, doi: 10.1038/s41467-025-59250-y).

Nunn, B. L., Kubanek, J., Rynearson, T. A., Timmins-Schiffman, E., Mudge, M., Bartlett, E. (2025) **Flow**

cytometry data from samples collected from Orcas Island, WA, USA Coastal Ocean (2m depth) every four hours from 2021-05-28 to 2021-06-18. Biological and Chemical Oceanography Data Management Office (BCO-DMO). (Version 1) Version Date 2025-09-09 doi:10.26008/1912/bco-dmo.984014.1 [view at BCO-DMO]

Relationship Description: These data were used in results publications from this project Nunn et al (2024, doi: 10.1038/s41597-024-04013-5) and Mudge et al. (2025, doi: 10.1038/s41467-025-59250-y).

Nunn, B. L., Kubanek, J., Ryneerson, T. A., Timmins-Schiffman, E., Mudge, M., Bartlett, E. (2025) **Nutrient data for samples collected every 4 hours from Orcas Island, WA, USA Coastal Ocean (2m depth) during the period from 2021-05-27 to 2021-06-18.** Biological and Chemical Oceanography Data Management Office (BCO-DMO). (Version 1) Version Date 2025-09-09 doi:10.26008/1912/bco-dmo.984065.1 [view at BCO-DMO]

Relationship Description: These data were used in results publications from this project Nunn et al (2024, doi: 10.1038/s41597-024-04013-5) and Mudge et al. (2025, doi: 10.1038/s41467-025-59250-y).

University of Washington (2024). Pooled metagenome of seawater microbial communities from East Sound, Orcas Island, WA, USA. 2024/07. NCBI:BioProject: PRJNA1130528. [Internet]. Bethesda, MD: National Library of Medicine (US), National Center for Biotechnology Information; Available from: <http://www.ncbi.nlm.nih.gov/bioproject/PRJNA1130528>

[[table of contents](#) | [back to top](#)]

Parameters

Parameter	Description	Units
DateID_PT	Character value for the combined date and time of sample collection (time zone: US Pacific (PST/PDT))	unitless
ISO_DateTime_UTC	Datetime with timezone (ISO format) of sample collection (time zone: UTC)	unitless
Latitude	Latitudinal coordinate of where experiment occurred	unitless
Longitude	Longitudinal coordinate of where experiment occurred	unitless
Experiment_Accession	SRA Experiment Accession number	unitless
Experiment_Title	SRA Experiment Title (which indicates the location the sample was taken, the date the sample was taken, the day number the sample was taken out of 22 days, and the time the sample was taken in (US Pacific PST/PDT time zone))	unitless
Instrument	instruments sequencing was completed on	unitless
Submitter	who submitted the archive entry	unitless
Study_Accession	study accession number	unitless
Study_Title	contains the study title	unitless

Sample_Accession	contains the SRA sample accession number	unitless
Total_Size_Mb	the total size of the file	Megabytes (Mb)
Total_RUNs	the total number of runs	unitless
Total_Spots	total spots	unitless
Total_Bases	total bases	unitless
Library_Name	library name	unitless
Library_Strategy	library strategy	unitless
Library_Source	library source	unitless
Library_Selection	library selection	unitless
IMG_GenomeID	JGI Integrated Microbial Genomes & Microbiomes identification number	unitless
Domain	type of sample sequenced	unitless
Sequencing_Status	status of JGI sequencing effort for this ID	unitless
Study_Name	name of the umbrella project study	unitless
Genome_Name_Sample_Name	name of the specific sample includes standard annotation location_sample_datecollected_daycollected_timecollected	unitless
Sequencing_Center	location of sequencing	unitless
IMG_Submission_ID	Submission ID number	unitless
GOLD_Analysis_Project_ID	Joint Genome Institute Genome OnLine Database project ID	unitless
GOLD_Analysis_Project_Type	type of Joint Genome Institute Genome OnLine Database project	unitless

GOLD_Sequencing_Project_ID	Joint Genome Institute Genome OnLine Database project ID	unitless
Genome_Size_assembled	size of assembled genome	unitless
Gene_Count_assembled	number of genes in metagenome	unitless
Genome_MetaBAT_Bin_Count_assembled	number of metaBAT counts	unitless
Estimated_Number_of_Genomes_assembled	estimated number of genomes assembled	unitless
Estimated_Average_Genome_Size_assembled	estimated average genome size	unitless

[[table of contents](#) | [back to top](#)]

Instruments

Dataset-specific Instrument Name	NovaSeq 6000 with S4 flow cell
Generic Instrument Name	Automated DNA Sequencer
Dataset-specific Description	Illumina technology was used for DNA sequencing. NovaSeq 6000 with S4 flow cell.
Generic Instrument Description	A DNA sequencer is an instrument that determines the order of deoxynucleotides in deoxyribonucleic acid sequences.

Dataset-specific Instrument Name	Varioskan Lux (fluorescent plate reader)
Generic Instrument Name	plate reader
Dataset-specific Description	Skant Software 7.0.2 was used as the fluorescent plate reader (Varioskan Lux).
Generic Instrument Description	Plate readers (also known as microplate readers) are laboratory instruments designed to detect biological, chemical or physical events of samples in microtiter plates. They are widely used in research, drug discovery, bioassay validation, quality control and manufacturing processes in the pharmaceutical and biotechnological industry and academic organizations. Sample reactions can be assayed in 6-1536 well format microtiter plates. The most common microplate format used in academic research laboratories or clinical diagnostic laboratories is 96-well (8 by 12 matrix) with a typical reaction volume between 100 and 200 μ L per well. Higher density microplates (384- or 1536-well microplates) are typically used for screening applications, when throughput (number of samples per day processed) and assay cost per sample become critical parameters, with a typical assay volume between 5 and 50 μ L per well. Common detection modes for microplate assays are absorbance, fluorescence intensity, luminescence, time-resolved fluorescence, and fluorescence polarization. From: http://en.wikipedia.org/wiki/Plate_reader , 2014-09-0-23.

[[table of contents](#) | [back to top](#)]

Project Information

Collaborative Research: Rhythm and Blooms: Deciphering metabolic, functional and taxonomic interactions over the life cycle of a phytoplankton bloom (Rhythm and Blooms)

Coverage: Coastal waters East Sound WA

NSF Award Abstract:

Floating, single-celled algae, or phytoplankton, form the base of marine food webs. When phytoplankton have sufficient nutrients to grow quickly and generate dense populations, known as blooms, they influence productivity of the entire food web, including rich coastal fisheries. The present research explores how the environment (nutrients) as well as physical and chemical interactions between individual cells in a phytoplankton community and their associated bacteria act to control the timing of bloom events in a dynamic coastal ecosystem. The work reveals key biomolecules within the base of the food web that can inform food web functioning (including fisheries) and be used in global computational models that forecast the impacts of phytoplankton activities on global carbon cycling. A unique set of samples and data collected in 2021 and 2022 that captured phytoplankton and bacterial communities before, during, and after phytoplankton blooms, is analyzed using genomic methods and the results are used to interrogate these communities for biomolecules associated with blooms stages. The team mentors undergraduates, graduate students, and postdoctoral researchers in the fields of biochemical oceanography, genome sciences, and time-series multivariate statistics. University of Washington organized hackathons develop publicly accessible portals for the simplified interrogation and visualization of 'omics data by high schoolers and undergraduates and are implemented in investigator-led undergraduate teaching modules and the University of Rhode Island Ocean Classroom. The research team also returns to Orcas Island, WA, where the field sampling takes place, to host a series of annual Science Weekends to foster scientific engagement with the local community.

Phytoplankton blooms, from initiation to decline, play vital roles in biogeochemical cycling by fueling primary production, influencing nutrient availability, impacting carbon sequestration in aquatic ecosystems, and supporting secondary production. In addition to environmental conditions, the physical and chemical interactions between individual phytoplankton can significantly modulate blooms, influencing the growth, maintenance, and senescence of phytoplankton. Recent work in steady-state open ocean ecosystems has shown that important chemicals are transferred amongst plankton on time-dependent metabolic schedules that are related to diel cycles. It is unknown how these metabolic schedules operate in dynamic coastal environments that experience perturbations, such as phytoplankton blooms. Here, the investigators are examining metabolic scheduling using long-term, diel sample sets to reveal how chemical and biological signals associated with the initiation, maintenance, and cessation of phytoplankton blooms are modulated on both short (hrs) and long (days-weeks) time scales. Findings are advancing the ability to predict and manage phytoplankton dynamics, providing crucial insights into ecological stability and future oceanographic sampling strategies. Additionally, outcomes of this study are providing a new foundational understanding of the succession of microbial communities and their chemical interactions across a range of timescales. In the long term, this research has the potential to identify predictors of the timing of phytoplankton blooms, optimize fisheries management, and guide future research on carbon sequestration.

[[table of contents](#) | [back to top](#)]

Funding

Funding Source	Award
NSF Division of Ocean Sciences (NSF OCE)	OCE-2401646
NSF Division of Ocean Sciences (NSF OCE)	OCE-2401645
NSF Division of Ocean Sciences (NSF OCE)	OCE-2401644
National Institute of Environmental Health Sciences (NIEHS)	R21ES034337-01
NSF Division of Integrative Organismal Systems (NSF IOS)	IOS-2041497
National Institutes of Health (NIH)	ES032733-01A1

[[table of contents](#) | [back to top](#)]