

Genomic data collected from R/V Atlantis cruise AT50-08 in the Eastern Tropical North Pacific during February-March 2023

Website: <https://www.bco-dmo.org/dataset/999321>

Data Type: Cruise Results

Version: 1

Version Date: 2026-05-28

Project

» [Collaborative Research: Key Microbial Processes in Oxygen Minimum Zones: From In Situ Community Rate Measurements to Single Cells](#) (MicroPro)

| Contributors | Affiliation | Role |
|--------------------------------------|---|------------------------|
| Pachiadaki, Maria G. | Woods Hole Oceanographic Institution (WHOI) | Principal Investigator |
| Rauch, Shannon | Woods Hole Oceanographic Institution (WHOI BCO-DMO) | BCO-DMO Data Manager |

Abstract

This dataset corresponds to the raw sequences of the single-cell amplified genomes (SAGs) collected for the MicroPro project in the Eastern Tropical North Pacific oxygen minimum zone in February and March of 2023 aboard R/V Atlantis on cruise AT50-08B. It includes genome identifiers and associated metadata (e.g., sample IDs, collection context) and links each genome to its corresponding accession in the National Center for Biotechnology Information's Short Read Archive (NCBI SRA).

Table of Contents

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

Coverage

Location: Eastern Tropical North Pacific

Methods & Sampling

Water samples (see data file for sampling depths and coordinates) were prepared by cryopreservation according to the protocol recommended by the Bigelow Single Cell Genomics Center (https://scgc.bigelow.org/wp-content/uploads/2018/06/Sample_cryopreservation_glyTE.pdf).

Briefly, small volumes of whole water (1 milliliter (ml)) were aliquoted under Helium (He) atmosphere using a glove box. Redox-Sensor Green (1 microliter (uL)) was added to the aliquots and left to incubate in the dark at in situ temperatures for 30 minutes. At the end of the incubation, glyTE (100 uL) was added and they were subsequently frozen at -80 degrees Celsius.

Sorting was performed on 4 April 2023 (within <2 months from the date the first sample was collected) and SAGs were generated with the modified genomic DNA amplification technique, WGA-Y, which enables a substantially improved average genome recovery from single cells (service S-202).

Data Processing Description

The obtained sequence reads were quality-trimmed with Trimmomatic v0.32 using the following settings: -phred33 LEADING:0 TRAILING:5 SLIDINGWINDOW:4:15 MINLEN:36. Reads matching the *H. sapiens* reference assembly GRCh38 and a local database of WGA-X reagent contaminants ($\geq 95\%$ identity of ≥ 100 bp alignments), as well as low complexity reads (containing $< 5\%$ of any nucleotide) were removed. The remaining reads were digitally normalized with kmernorm 1.05 (<http://sourceforge.net/projects/kmernorm>) using settings -k 21 -t 30 -c 3 and then assembled with SPAdes v.3.15.2 using the following settings: --careful --sc --phred-offset 33. Each end of the obtained contigs was trimmed by 100 bp, and then only contigs longer than 2,000 bp were retained. Contigs matching the *H. sapiens* reference assembly GRCh38 and a local database of WGA-X reagent contaminants ($\geq 95\%$ identity of ≥ 100 bp alignments) were removed. This workflow was evaluated for assembly errors using three bacterial benchmark cultures with diverse genome complexity and %GC, indicating no non-target and undefined bases in the assemblies and average frequencies of mis-assemblies, indels, and mismatches per 100 kbp: 1.5, 3.0 and 5.0.

SAG taxonomic assignments were obtained with GTDB-Tk v2.2.6. In addition, 16S rRNA gene regions longer than 500 bp were identified using local alignments provided by BLAST against CREST's curated SILVA reference database SILVAMod v128 and classified using a reimplement of CREST's last common ancestor algorithm. Genome functional annotation was first performed using Prokka with default Swiss-Prot databases supplied by the software. Prokka was run a second time with a custom protein annotation database built from compiling Swiss-Prot entries for Archaea and Bacteria.

BCO-DMO Processing Description

currently being processed

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

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](https://doi.org/10.1089/cmb.2012.0021)
Methods

Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Adesina, A., Ahmad, S., Bowler-Barnett, E. H., Bye-A-Jee, H., Carpentier, D., Denny, P., Fan, J., Garmiri, P., Gonzales, L. J. d. C., Hussein, A., Ignatchenko, A., Insana, G., Ishtiaq, R., Joshi, V., et al. (2024). UniProt: the Universal Protein Knowledgebase in 2025. *Nucleic Acids Research*, 53(D1), D609–D617. <https://doi.org/10.1093/nar/gkae1010>
Methods

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. doi:[10.1093/bioinformatics/btu170](https://doi.org/10.1093/bioinformatics/btu170)
Methods

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10(1). doi:[10.1186/1471-2105-10-421](https://doi.org/10.1186/1471-2105-10-421)
Methods

Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2022). GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics*, 38(23), 5315–5316. <https://doi.org/10.1093/bioinformatics/btac672>
Methods

Lanzén, A., Jørgensen, S. L., Huson, D. H., Gorfer, M., Grindhaug, S. H., Jonassen, I., Øvreås, L., & Urich, T. (2012). CREST - Classification Resources for Environmental Sequence Tags. *PLoS ONE*, 7(11), e49334. <https://doi.org/10.1371/journal.pone.0049334>
Methods

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F. O. (2012). The

SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, 41(D1), D590–D596. doi:[10.1093/nar/gks1219](https://doi.org/10.1093/nar/gks1219)

Methods

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30(14), 2068–2069.

<https://doi.org/10.1093/bioinformatics/btu153>

Methods

Stepanauskas, R., Fergusson, E. A., Brown, J., Poulton, N. J., Tupper, B., Labonté, J. M., Becraft, E. D., Brown, J. M., Pachiadaki, M. G., Povilaitis, T., Thompson, B. P., Mascena, C. J., Bellows, W. K., & Lubys, A. (2017).

Improved genome recovery and integrated cell-size analyses of individual uncultured microbial cells and viral particles. *Nature Communications*, 8(1). <https://doi.org/10.1038/s41467-017-00128-z>

Methods

Zhao, J., Pachiadaki, M., Conrad, R. E., Hatt, J. K., Bristow, L. A., Rodriguez-R, L. M., Rossello-Mora, R., Stewart, F. J., & Konstantinidis, K. T. (2025). Promiscuous and genome-wide recombination underlies the sequence-discrete species of the SAR11 lineage in the deep ocean. *The ISME Journal*, 19(1).

<https://doi.org/10.1093/ismejo/wraf072>

Results

limingkun. (2015). kmernorm (Version 1.05) [Software]. SourceForge. Retrieved date you accessed it, from <http://sourceforge.net/projects/kmernorm>

Methods

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

Related Datasets

IsRelatedTo

Georgia Institute of Technology. Candidatus Pelagibacterales Raw sequence reads. 2024/06. 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/PRJNA1124867>. NCBI:BioProject: PRJNA1124867. <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1124867>

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

Parameters

Parameters for this dataset have not yet been identified

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

Instruments

| | |
|---|--|
| Dataset-specific Instrument Name | NextSeq 2000 DNA sequencer |
| Generic Instrument Name | Automated DNA Sequencer |
| Dataset-specific Description | The genomic material was sequenced using NextSeq 2000 DNA sequencer. |
| Generic Instrument Description | A DNA sequencer is an instrument that determines the order of deoxynucleotides in deoxyribonucleic acid sequences. |

| | |
|---|--|
| Dataset-specific Instrument Name | BD InFlux Mariner flow cytometer |
| Generic Instrument Name | Flow Cytometer |
| Dataset-specific Description | Cells were sorted using an BD InFlux Mariner flow cytometer |
| Generic Instrument Description | Flow cytometers (FC or FCM) are automated instruments that quantitate properties of single cells, one cell at a time. They can measure cell size, cell granularity, the amounts of cell components such as total DNA, newly synthesized DNA, gene expression as the amount messenger RNA for a particular gene, amounts of specific surface receptors, amounts of intracellular proteins, or transient signalling events in living cells. (from: http://www.bio.umass.edu/micro/immunology/facs542/facswhat.htm) |

| | |
|---|---|
| Dataset-specific Instrument Name | 12L Niskin bottles |
| Generic Instrument Name | Niskin bottle |
| Dataset-specific Description | Seawater samples were collected using a 24-bottle rosette equipped with 12 L Niskin bottles. |
| Generic Instrument Description | A Niskin bottle (a next generation water sampler based on the Nansen bottle) is a cylindrical, non-metallic water collection device with stoppers at both ends. The bottles can be attached individually on a hydrowire or deployed in 12, 24, or 36 bottle Rosette systems mounted on a frame and combined with a CTD. Niskin bottles are used to collect discrete water samples for a range of measurements including pigments, nutrients, plankton, etc. |

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

Deployments

AT50-08B

| | |
|--------------------|---|
| Website | https://www.bco-dmo.org/deployment/995472 |
| Platform | R/V Atlantis |
| Start Date | 2023-02-10 |
| End Date | 2023-03-16 |
| Description | Project: Collaborative Research: Key Microbial Processes in Oxygen Minimum Zones: From In Situ Community Rate Measurements to Single Cells Chief: Pachiadaki, Maria G Start port: Putarenas, Costa Rica End port: Puntarenas, Costa Rica See additional information at R2R: https://www.rvdata.us/search/cruise/AT50-08B |

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

Project Information

Collaborative Research: Key Microbial Processes in Oxygen Minimum Zones: From In Situ Community Rate Measurements to Single Cells (MicroPro)

Coverage: East Tropical North Pacific Ocean

NSF Award Abstract:

Oxygen availability shapes the distributions and activities of marine organisms. Ongoing human activities and climate change are expected to lead to expansion and intensification of already large oxygen-stressed areas of the coastal and open ocean. Decreases in ocean oxygen have significant ecological consequences, including habitat loss for migratory and bottom-dwelling organisms, modification of the marine food web, and production of trace gases with pronounced feedbacks on climate, such as methane and nitrous oxide. Intense chemical cycling by microorganisms occurs in oxygen-depleted marine habitats. However, a full understanding of the consequences for marine ecosystems is hampered by limited knowledge of actual rates of key microbiological processes and dynamics of the microorganisms mediating them. This study combines novel methods and sampling techniques to understand how these processes are influenced by changes in oxygen concentration to inform predictions of important chemical exchanges within a changing ocean and its production of climate-active gases. This deeply collaborative project trains undergraduates (four of whom participate on the cruise), a graduate student and a postdoctoral fellow. Outreach takes place in middle and high schools and through social media. Data and samples from the cruise are integrated in coursework.

Oxygen depletion alters cycling of major elements (especially carbon, nitrogen, and sulfur) as well as food web functionality. This project addresses major gaps in our knowledge of oxygen minimum zone (OMZ) processes by applying in situ approaches to more accurately measure rates of several key microbial processes (chemoautotrophy, denitrification, anammox, sulfate reduction and sulfide oxidation) central to marine biogeochemical cycling. This work studies the Eastern Tropical North Pacific OMZ, the largest open ocean oxygen-depleted system, to 1) determine the in situ rates of microbial processes involved in carbon, nitrogen, and sulfur cycling, 2) reveal the genomic blueprint of active single cells involved in these processes, and 3) obtain estimates of the relative contributions of the dominant chemoautotrophic and heterotrophic groups to the measured rates. This work includes cutting-edge equipment for in situ sampling and incubations that minimize artifacts associated with traditional water sampling approaches, allowing more accurate estimates of rates of important biogeochemical processes. Additionally, rate measurements of relatively undisturbed bulk and fractionated water samples make it easier to distinguish the potential role of particle-associated microorganisms in these OMZ processes. Single cell sorting of microorganisms using a fluorescent dye indicative of cell activity together with metatranscriptomics informs on metabolic pathways used for key processes by active microbial community members, as well as the potential coupling of chemoautotrophy and nitrogen or/and sulfur cycling. By combining stable isotope probing, fluorescence in situ hybridization and single cell Raman microspectrometry the relative activity levels of different microbial phylotypes involved in chemoautotrophic and heterotrophic elemental cycling are assessed.

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-1924492 |
| NSF Division of Ocean Sciences (NSF OCE) | OCE-1924424 |

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