Metagenome from the Pacific Ocean OMZ

Website: https://www.bco-dmo.org/dataset/998771
Version: 1
Version Date: 2026-05-18

Project
» Collaborative Research: Microdiversity drives ecosystem function: SAR11 bacteria as models for oceanic nitrogen loss (SAR11 in OMZs)

Contributors	Affiliation	Role
Konstantinidis, Kostas	Georgia Institute of Technology (GA Tech)	Principal Investigator
Rauch, Shannon	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

Abstract

This dataset represents the shotgun metagenome samples used in used in Zhao et al., ISME J 2025 and described in the following abstract: Surveys of microbial communities (metagenomics) or isolate genomes have revealed sequence-discrete species. That is, members of the same species show >95% average nucleotide identity (ANI) of shared genes among themselves vs. <83% ANI to members of other species while genome pairs showing between 83% and 95% ANI are comparatively rare. In these surveys, aquatic bacteria of the ubiquitous SAR11 clade (Class Alphaproteobacteria) are an outlier and often do not exhibit discrete species boundaries, suggesting the potential for alternate modes of genetic differentiation. To explore evolution in SAR11, we analyzed high-quality, single-cell amplified genomes, and companion metagenomes from an oxygen minimum zone (OMZ) in the Eastern Tropical Pacific Ocean, where the SAR11 make up ~20% of the total microbial community. Our results show that SAR11 do form several sequence-discrete species, but their ANI range of discreteness is shifted to lower identities between 86% and 91%, with intra-species ANI ranging between 91% and 100%. Measuring recent gene exchange among these genomes based on a recently developed methodology revealed higher frequency of homologous recombination within compared to between species that affects sequence evolution at least twice as much as diversifying point mutation across the genome. Recombination in SAR11 appears to be more promiscuous compared to other prokaryotic species, likely due to the deletion of universal genes involved in the mismatch repair, and has facilitated the spread of adaptive mutations within the species (gene sweeps), further promoting the high intraspecies diversity observed. Collectively, these results implicate rampant, genome-wide homologous recombination as the mechanism of cohesion for distinct SAR11 species.

Table of Contents

Coverage

Location: Eastern Tropical Pacific Ocean
Spatial Extent: **N**:16.9798 **E**:-90.4497 **S**:11.8122 **W**:-107.7
Temporal Extent: 2021-12-30 - 2022-10-16

Methods & Sampling

Samples for metagenomic sequencing were collected from the Eastern Tropical North Pacific (ETNP) Oxygen Minimum Zone (OMZ) on R/V Sally Ride (SR2114) from December 2021 to January 2022. Sea water for metagenomes was collected from all nine depths for all five sampling stations. Collections were made using Niskin bottles on a rosette containing a conductivity–temperature–depth profiler (Sea-Bird SBE 911plus), as described in Tsementzi et al., 2016.

DNA was extracted from biomass on collecting filters using the MoBio Power Soil kit (MoBio Inc. Carlsbad, CA, USA) and libraries were prepared for metagenomic sequencing using the Illumina DNA library prep kit with unique dual indexing according to manufacturer's instructions, except that the protocol was terminated after isolation of cleaned double stranded libraries. An equimolar mixture of the libraries was 2 sequenced on an Illumina NovaSeq 6000 instrument at the Molecular Evolution Core, Georgia Institute of Technology.

Data Processing Description

The raw shotgun metagenomes are provided, and are available in NCBI under BioProject number PRJNA1124864.

BCO-DMO Processing Description

- Imported original Excel file "Table S2. Sample information for the metagenomes from the ETNP OMZ.xlsx" (sheet 1) into the BCO-DMO data processing system.
- Converted "Date Sampled" from format "%m-%d-%y" to "%Y-%m-%d".
- Converted "Date Extracted" from format "%m.%d.%y" to "%Y-%m-%d".
- Converted "Lat N" from degrees-decimal_minutes format to decimal degrees using directional N, and renamed the new column "Latitude".
- Converted "Long W" from degrees-decimal_minutes format to decimal degrees using directional W, and renamed the new column "Longitude".
- Rounded Latitude and Longitude to maximum 4 decimal places.
- Renamed columns to comply with BCO-DMO naming conventions.
- Saved the final file as "998771_v1_etnp_metagenomes_zhao.csv".

[table of contents | back to top]

Related Publications

Tsementzi, D., Wu, J., Deutsch, S., Nath, S., Rodriguez-R, L. M., Burns, A. S., Ranjan, P., Sarode, N., Malmstrom, R. R., Padilla, C. C., Stone, B. K., Bristow, L. A., Larsen, M., Glass, J. B., Thamdrup, B., Woyke, T., Konstantinidis, K. T., & Stewart, F. J. (2016). SAR11 bacteria linked to ocean anoxia and nitrogen loss. Nature, 536(7615), 179–183. https://doi.org/10.1038/nature19068
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

[table of contents | back to top]

Parameters

Parameter	Description	Units
Sample_ID	Sample ID number	unitless
DNA_Concentration	DNA concentration	micromoles per liter
Date_Sampled	Date sampled	unitless
Date_Extracted	Date of DNA extraction	unitless
Latitude	Latitude where sample was collected	decimal degrees
Longitude	Longitude where sample was collected	decimal degrees
Depth	Depth of sample collection	meters (m)
Volume_Filtered	Volume of water filtered	milliliters (mL)
Station	Station number	unitless
Oxygen_Concentration	Oxygen concentration	micromoles per kilogram
Experiment_Accession	NCBI experiment accession number	unitless
Instrument	Instrument used in sequencing	unitless
Study_Accession	NCBI study accession number	unitless
Sample_Accession	NCBI sample accession number	unitless
Bioproejct_number	NCBI BioProject number	unitless
Total_Bases	Total number of bases	number of bases
Library_Name	ID of the sequenced library	Matches Sample ID above plus another unique identifier.

[table of contents | back to top]

Instruments

Dataset-specific Instrument Name	Illumina NovaSeq 6000
Generic Instrument Name	Automated DNA Sequencer
Dataset-specific Description	Libraries were sequenced on an Illumina NovaSeq 6000 instrument at the Molecular Evolution Core, Georgia Institute of Technology.
Generic Instrument Description	A DNA sequencer is an instrument that determines the order of deoxynucleotides in deoxyribonucleic acid sequences.

Dataset-specific Instrument Name	Niskin bottles
Generic Instrument Name	Niskin bottle
Dataset-specific Description	Niskin bottles on a rosette were used to collect the water samples.
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

SR2114

Website	https://www.bco-dmo.org/deployment/931391
Platform	R/V Sally Ride
Start Date	2021-12-23
End Date	2022-01-21
Description	Additional cruise information is available from R2R: https://www.rvdata.us/search/cruise/SR2114

[table of contents | back to top]

Project Information

Collaborative Research: Microdiversity drives ecosystem function: SAR11 bacteria as models for oceanic nitrogen loss (SAR11 in OMZs)

Coverage: Eastern Tropical North Pacific, off Colima, Mexico

NSF Award Abstract:
This project studies how low oxygen availability influences the biodiversity and ecological role of SAR11

bacteria, one of the most abundant microbial groups in the ocean. The work involves oceanographic sampling across a range of oxygen and nutrient levels in the Eastern Tropical North Pacific Ocean. Using a combination of genomic, microbiological, and biogeochemical methods, the study identifies the mechanisms by which SAR11 strains diversify into separate niches and species and contribute biochemically to the ecosystem, likely through removing nitrogen from seawater. The project equips the next generation of researchers and educators, notably those from underrepresented minority groups, to use oceanographic, genomic, and microbiological concepts to meet contemporary scientific challenges. This goal is met through a combination of bioinformatic workshops that target undergraduate students from the University System of Puerto Rico, middle school teacher-training workshops, and middle or high school teacher internships in the investigator’s labs. This multifaceted research and educational agenda fills a gap in our understanding of marine biological diversity, identifies the contribution of SAR11 bacteria to nutrient and carbon cycles in low oxygen oceans, and provides lessons and analytical tools to study microbial processes in other ecosystems.

This project has two aims. Aim 1 employs comparative metagenomic and single-cell genomic analyses to identify metabolic properties that distinguish SAR11 clades from low oxygen regions and processes of selection or gene flow operating across the clades. Aim 2 combines microbial transcriptomics, incubation experiments with isotope tracers, and culturing to delimit the oxygen and nutrient conditions that define the niche space of each SAR11 clade and to correlate SAR11 gene transcription with community biochemical outcomes, including nitrogen loss through denitrification. The results of these aims and the informatic methods used to probe microbial microdiversity are disseminated through genomics-focused undergraduate workshops, and new teacher-training educational modules, including lab-based modules focused on the importance of microorganisms under environmental change in the oceans. Data, manuscripts, and informatics workflows from this project are made publicly available. The results are critical for resolving the processes that create and sustain microbial diversity in the oceans and informing biogeochemical models that predict how diversity influences ecosystem processes.

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-2129823

[table of contents | back to top]