Sample information for the metagenomes from the Gulf of Mexico collected on R/V Endeavor cruise EN509 in May 2012

Website: https://www.bco-dmo.org/dataset/998957
Data Type: Cruise Results
Version: 1
Version Date: 2026-05-20

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 includes sample information for the metagenomes from the Gulf of Mexico used in the Conrad et al. AEM 2026 study. Water samples were collected during CTD casts from the May 2012 R/V Endeavour cruise EN509 and shotgun metagenomes were sequenced from the surface (~3 meters (m)), the mixed layer (ML; 15-25m), the deep chlorophyll maximum (DCM; 70-90m), below the DCM but above the oxygen minimum zone (OMZ; 100-150m), and the OMZ (200-400m) from three stations in the northwest Gulf of Mexico. Sequences are available in the National Center for Biotechnology Information (NCBI) under BioProject number PRJNA291283.

Table of Contents

Coverage

Location: Gulf of Mexico
Spatial Extent: **N**:28.8505 **E**:-88.4893 **S**:26.095 **W**:-92.701
Temporal Extent: 2012-05-26 - 2012-06-02

Methods & Sampling

Water column samples were collected during CTD casts from the May 2012 R/V Endeavour cruise EN509 and shotgun metagenomes were sequenced from the surface (~3 meters (m), the mixed layer (ML; 15-25m), the deep chlorophyll maximum (DCM; 70-90m), below the DCM but above the oxygen minimum zone (OMZ; 100-150m), and the OMZ (200-400m) from three stations (2, 5, and 8) in the northwest GoM (Fig. 1A of Conrad et al., 2026).

Four additional metagenomes were sequenced from below the OMZ (Deep; 600, 1000, 1470 and 2107m) from station 5. Station 2 was near the edge of the Texas-Louisiana Shelf about 50 miles southeast from the mouth of the Mississippi river. Station 8 and 5 were over the TX-LA Slope about 190 and 270 mi. southeast of

Galveston, TX with station 8 nearer the TX-LA Shelf and station 5 nearer the edge of the TX-LA slope. Fluorescence, oxygen, salinity, and temperature measurements were similar at all stations except salinity was lower (29.6 vs. 36 PSU) and fluorescence was increased (0.9 vs. 0.07 milligrams per cubic meter (mg/m³) at the surface of station 2 while the fluorescence peak of the DCM layer was reduced (0.3 vs. 0.8 mg/m³) compared to stations 5 and 8 (Fig. 1B, C, D, and E of Conrad et al., 2026). Temperature and salinity were most similar for all samples from the ML and DCM while salinity was variable across surface samples and temperature and salinity were variable across aOMZ and OMZ samples (Fig. 1E of Conrad et al., 2026).

Data Processing Description

DNA was purifiedusing the Ampure XP-Beads (Beckman Coulter). DNA sequencing was performed using a Nextera XT DNA Sample Prep kit, and a paired-end strategy on an Illumina MiSeq sequencer available at the Molecular Core of Georgia Tech. Sequences are available in the National Center for Biotechnology Information (NCBI) under BioProject number PRJNA291283.

BCO-DMO Processing Description

- Imported sheet 1 of the original file named, "Supplemental_File_01-metaGs-revised.xlsx into table "998957_v1_metagenome_gom" into the BCO-DMO system.
- Filtered rows to keep only the first 19 rows, excluding secondary tables located on the data sheet.
- Combined Year (format Y), Month (format %B), and Day (format %d) columns into a new Date column formatted as %Y-%m-%d with type date.
- Renamed all columns to comply with BCO-DMO naming conventions.
- Saved the final file as "998957_v1_metagenome_gom.csv".

[table of contents | back to top]

Related Publications

Conrad, R. E., Tsementzi, D., Meziti, A., Hatt, J. K., Montoya, J., & Konstantinidis, K. T. (2026). Metagenome-based vertical profiling of the Gulf of Mexico highlights its uniqueness and far-reaching effects of freshwater input. Applied and Environmental Microbiology, 92(2). https://doi.org/10.1128/aem.02589-25
Results

[table of contents | back to top]

Related Datasets

IsRelatedTo

Georgia Institute of Technology. marine metagenome, seawater sample Metagenome. 2016/08. 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/PRJNA291283. NCBI:BioProject: PRJNA291283. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA291283

[table of contents | back to top]

Parameters

Parameter	Description	Units
Acession	NCBI experiment accession number	unitless

Sample_Name	Sample ID number	unitless
Vessel	Ship name	unitless
Cruise	Cruise ID	unitless
Ocean	Ocean where samples were collected	unitless
Date	Date	unitless
Year	4-digit year	unitless
Month	Month	unitless
Day	Day of month	unitless
Station	Station number	unitless
Lat	Latitude of sample collection	decimal degrees
Lon	Longitude of sample collection	decimal degrees
Description	Sample description: SURFACE =Surface (3m); ML = mixed layer (15–25 m); DCM = Deep Chlorophyll Maximum (70–90 m); aOMZ = below the DCM but above the oxygen minimum (OM) depth (100–150 m); OMZ = oxygen minimum zone depth (100–150 m); DEEP = below the OM depth (600, 1,000, 1,470, or 2,107 m)	unitless
Depth	Sample depth	meters (m)
Temp	Water temperature	degrees Celsius
Salinity	Water sailinity	PSU
Density	Density	sigma theta
Fluoresence	Fluorescence	milligrams per cubic meter (mg/m^3)
Oxygen	Oxygen concentration	micromolar (uM)

Xmiss	percent of total	ranging 0-100%.
Conductivity	Conductivity	Siemens per meter (S/m)
Par	Photosynthetically active radiation	micromoles of photons per square meter per second (µmol*m^-2*s^-1)
Mean_PO4	Mean phosphate concentration	micromolar (uM)
Mean_Si	Mean silicon concetration	micromolar (uM)
Mean_NO3_NO2	Mean nitrate + nitrite	micromolar (uM)
Mean_N_star	(description needed)	unknown
Sequencing_Effort_Gb	Total sequencing performed	Giga-basepairs (Gb)
Sequencing_Effort_Bp	Total sequencing performed	Basepairs (Bp)
Metagenome_Reads_GC	average G+C% of metagenomics reads	ranging 25-75%, typically
Average_Genome_Size	Average Genome Size of sequences community	Mega-basepairs (Mb)
Nonpareil_Diversity	Community diversity as estimated by Nonpareil v3.401	ranging 1-30, typically
Diversity_Covered	% of community diversity recovered by sequencing based on Nonpareil v3.401	ranging 0-100%.
Contigs_Assembled	Number of large contigs assembled	integer, number
Assembled_Length	total length of assembled contigs	Basepairs (Bp)
Assembled_N50	The length of the shortest contig that covers at least 50% of the total assembly	Basepairs (Bp)
Assembled_GC	average G+C% of contigs	ranging 25-75%, typically

Predicted_CDS	number of predicted proteins-coding genes of the assembly	integer, number
Reads_Mapping_to_Assembly	percent of total reads mapping back to the assembly.	ranging 0-100%.
Reads_Mapping_to_rMAGs	percent of total reads mapping back to the representative MAGs	ranging 0-100%.

[table of contents | back to top]

Instruments

Dataset-specific Instrument Name	Sea-Bird SBE 911plus
Generic Instrument Name	CTD Sea-Bird SBE 911plus
Dataset-specific Description	Collections were made using Niskin bottles on a rosette containing a conductivity–temperature–depth profiler (Sea-Bird SBE 911plus).
Generic Instrument Description	The Sea-Bird SBE 911 plus is a type of CTD instrument package for continuous measurement of conductivity, temperature and pressure. The SBE 911 plus includes the SBE 9plus Underwater Unit and the SBE 11plus Deck Unit (for real-time readout using conductive wire) for deployment from a vessel. The combination of the SBE 9 plus and SBE 11 plus is called a SBE 911 plus. The SBE 9 plus uses Sea-Bird's standard modular temperature and conductivity sensors (SBE 3 plus and SBE 4). The SBE 9 plus CTD can be configured with up to eight auxiliary sensors to measure other parameters including dissolved oxygen, pH, turbidity, fluorescence, light (PAR), light transmission, etc.). more information from Sea-Bird Electronics

Dataset-specific Instrument Name	Niskin bottles
Generic Instrument Name	Niskin bottle
Dataset-specific Description	Collections were made using Niskin bottles on a rosette.
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

EN509

Website	https://www.bco-dmo.org/deployment/58933
Platform	R/V Endeavor
Start Date	2012-05-25
End Date	2012-06-20
Description	See additional information in R2R: https://www.rvdata.us/search/cruise/EN509

[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]