Microbiome data 16S rRNA results for sea urchin gut content, sediment, and surrounding seawater from sampled collected in 2023 from four locations along the coast of Puerto Rico

Website: https://www.bco-dmo.org/dataset/986537
Version: 1
Version Date: 2025-10-10

Project

» RAPID: The black urchin (Diadema antillarum) massive resurgent die-off: Causes, demographic, and community consequences (Diadema antillarum die-off associated microbiota in Puerto Rico)

Contributors	Affiliation	Role
Toledo-Hernandez, Carlos	Sociedad Ambiente Marino (SAM)	Principal Investigator
Godoy-Vitorino, Filipa	University of Puerto Rico School of Medicine (UPR-RCM)	Co-Principal Investigator
Ruiz-Diaz, Claudia Patricia	Sociedad Ambiente Marino (SAM)	Co-Principal Investigator
Chorna, Nataliya	University of Puerto Rico School of Medicine (UPR-RCM)	Scientist
Kardas, Elif	The University of Mons (Belgium) (UMons)	Scientist
Rodriguez-Barreras, Ruber	University of Puerto Rico - Mayaguez (UPRM)	Scientist
Gerlach, Dana Stuart	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

Abstract

May need to re-work This study investigated the microbiota associated with sea urchin gut content, sediment, and surrounding seawater using high-throughput sequencing of bacterial (16S rRNA), fungal (ITS), and eukaryotic (18S rRNA) gene markers. Data analyses is still ongoing. Sequences and associated metadata were deposited in QIITA (study ID 15616). Downstream analysis in QIIME2 included demultiplexing, quality trimming, DADA2 denoising, and taxonomic classification using the Greengenes reference database. Microbial community analyses assessed alpha- and beta-diversity, taxonomic composition, and biomarker discovery using MaAsLin and LEfSe. Comparative analyses were conducted across sample types, collection sites, and time periods, with additional ecological and functional inference (FAPROTAX) performed for gut samples. This integrative microbiome profiling provides critical insights into the taxonomic diversity, temporal dynamics, and potential functional roles of microbial communities in sea urchin-associated environments, offering a valuable perspective on host-microbe-environment interactions in marine ecosystems.

Table of Contents

Coverage

Location: Puerto Rican coast

Spatial Extent: **N**:18.485 **E**:-65.287 **S**:18.281 **W**:-66.298
Temporal Extent: 2023-03-23 - 2023-12-07

Dataset Description

One of ___ datasets investigating sea urchins and associated water and sediment from four locations in Puerto Rico.

- Capture data, sizes, and DNA yields
- Metabolites from GC-MS analysis
- Raw GC-MS data (metabolomics) --possibly combine with metabolites
- Microbiome data 16S rRNA results for sea urchin gut content, sediment, and surrounding seawater

Methods & Sampling

At the arrival of the sea urchin samples, seawater and sediment at the Microbiome lab (Laboratory of Dr. Filipa Godoy-Vitorino, FGV), Elif Kardas (EKA) proceeded to the anesthesia of the sea urchins, and to the filtration of the seawater sample, following the protocol. All sample types were then stored at -80°C until dissection. Before proceeding to the dissection, the size (equatorial diameter) and the presence of albinism on the sea urchin was noted. RRB and EKA then dissected all the collected sea urchins following the protocol in order to obtain the gut pellets (Suppl. File 1). At the end of the sampling period (December 2023), we obtained 45 sea urchin gut pellets, 14 sediment samples, and 16 seawater samples. All the metadata can be found in Supplementary File 2. Each sample of gut pellet and sea water was then divided into two sets, one for downstream microbial and one for metabolic processes (min 50 mg of gut pellets and 50 ml of seawater per sample).

The DNA of the 3 sample types (gut pellets, sediments, sea water) was then extracted using DNeasy PowerSoil Pro kit (QIAGEN LLC, Germantown Road, Maryland, United States) and quantified using Qubit® dsDNA HS assay kit. We normalized the DNA to 4nM during the 16S rRNA gene library preparation process. Using region-specific primers that include sequencer adapter sequences used in the Illumina flowcell, we used the Earth Microbiome Project standard protocols1 to amplify the hypervariable region V4 of the 16S ribosomal RNA gene (~291 bp) using the universal bacterial primers 515F (5′GTGCCAGCMGCCGCGGTAA3′) and 806R (5′GGACTACHVGGGTWTCTAAT3′). Amplicons were measured with a plate reader (Infinite® 200 PRO, Tecan) and PicoGreen (Invitrogen). Volumes of all the products were combined into one tube after they were quantified, resulting in an equimolar representation of each amplicon. This pool was then cleaned up using AMPure XP Beads (Beckman Coulter), and finally quantified using a fluorometer (Qubit, Invitrogen). Customized sequencing was outsourced to Argonne National Laboratory in Illinois, USA, utilizing a 2 × 150 bp paired-end sequencing kit with an Illumina MiSeq.

The sequencer's reads and the **associated** metadata were uploaded to QIITA study ID 15616 (version 2024.02). This includes ITS, 18S and 16S. The downstream microbial analyses were performed locally in QIIME2 (version 2024.05) and include demultiplexing, trimming at 250 bp, denoising. Denoising method was done using DADA2 (Callahan et al 2016) and included quality filtering, removing chimeric sequences, combining paired-ends, and eliminating singleton reads in order to join, denoise, and duplicate sequences. The bacterial sequences were classified using the Greengenes reference database and taxonomy files (McDonald et al 2024), trained with classify-sklearn function in QIIME2. Microbiota analyses (diversity and composition) were divided into four sets: (1) sample types (gut digesta pellet, seawater, or sediment), (2) gut digesta pellet between the different collection locations, (3) gut digesta pellet between the different collection periods, and (4) gut digesta pellet between the different collection locations considering the location periods. For each set, beta- and alpha-diversity were calculated, and taxonomic compositions (taxa barplots) and putative biomarker taxa, e.g, MaAsLin (Mallick et al. 2021) and/or LEfSe (Segata et ql. 2011) are represented. To correct for difference in extraction dates, those were added from the metadata as a reference for the MaAsLin analysis. In addition, for collection locations (set of analyses#2), ecological and metabolic function inference for 16S (FAPROTAX, Louca et al. 2016) and putative biomarker data (LefSe) were computed. 18S and ITS data was not completed by time of submission.

~ ~ ~ ~ ~ ~

The study was conducted at 4 sites along the northeastern and eastern coasts of Puerto Rico. These sites were selected due to the robust ecological and microbial data on *Diadema antillarum* collected in previous

studies (Rodríguez-Barreras et al., 2018, 2021, and 2022). Furthermore, recent observations suggest contrasting levels of disease impact among these sites, ranging from deeply affected to unaffected by the disease. Punta Melones (**PME**, 18°16'51.40"N, 65°17'12.21"W), located in the Luis Peña Natural Reserve in Culebra had a high density of sea urchins (1.5 ind. per m^2) before the recent die-off. Nevertheless, Rodrigues-Barrera et al., (2023) observed over 90% mortality of sea urchin at the site. Punta Bandera (**PBA,** 18°23'18.46" N, 65°43'5.52"W), Cerro Gordo reef (**CGO** 18°29'05.54"N, 65°20'21.65"W), and Punta Sardinas (**PSA** 18°28'36.34"N, 66°17'52.48"W), are fringing reefs located on the northern coast of Puerto Rico. These sites exhibited densities around 1.26 ind. per m^2.

Each site was surveyed on months 0, 3, 6, and 9 to collect samples from March 2023 to December 2023, between the hours of 10:00 and 13:00. We set up eight belt transects of 20 m² (10 m x 2 m) parallel to the coast to estimate density. Transects were positioned at least 5 meters apart from each other, at depths ranging from 1 to 3 meters; as sea urchin abundance tends to be higher at these depths. All individuals within each transect were counted. These data were used to estimate sea urchin density (i.e., the number of urchins per transect per site). At each site, between 1 and 3 healthy sea urchin specimens were collected, for a total of 45 individuals. Additionally, we measured the horizontal test diameter (*td*) of individuals collected from the transects to assess the size distribution at each reef. A total of 50 individuals per reef were measured using a caliper. If necessary, sea urchins found outside of the transects were also measured until reaching 50 individuals per reef, following Mercado-Molina et al., (2015) and Rodríguez-Barreras et al., (2018, 2023). We also measured the tests of dead and diseased sea urchins when possible. Based on test diameter, sea urchins were categorized into three size classes: small or juvenile (*td* ≤ 4.0 cm), medium or young adult (4.01 < *td* ≤ 6.0 cm), and large or adult (*td* ≥ 6.01 cm). These data were used to construct a size-frequency distribution (Miller et al., 2003; Lugo-Ascorbe, 2004; Rodríguez-Barreras et al., 2014; Rodríguez-Barreras et al., 2023).

Additionally, 1 liter of seawater (in a sterile glass bottle) and 50 mL of sediment (in a sterile plastic tube) were collected during each survey. Sampling was approved by the Department of Natural and Environmental Resources of Puerto Rico permit #DRNA-2022-IC-026 (O-VS-PVSlS-SJ-01291-06052022). The IACUC permit previously approved for Rodríguez-Barreras used in a former sea urchin-microbiota project was renewed until September 29, 2026 (UPRRCM # A530118).

Sample individual processing:

Upon arrival at the Microbiome lab (Laboratory of Dr. Filipa Godoy-Vitorino University of Puerto Rico at Medical School), sea urchin, water, and sediment samples were proceeded (***do you mean processed?***). Sea urchins were anesthetized, and seawater samples were filtrated. All sample types were then stored at -80°C until dissection. Before dissection, the horizontal test diameter was measured, along with any evidence of albinism. Sea urchins were dissected following the standard protocol to obtain fecal pellets. At the end of the sampling period (December 2023), a total of 45 fecal pellet samples, 14 sediment samples, and 16 seawater samples were collected. Each fecal pellet and seawater sample was divided into two aliquots: one for downstream microbial and one for metabolomic analyses

Data Processing Description

Info DM notes: Some of this information might be better in the regular METHODS section

~ ~~ ~ ~

At the arrival of the sea urchin samples, seawater and sediment at the Microbiome lab (Laboratory of Dr. Filipa Godoy-Vitorino, FGV), Elif Kardas (EKA) proceeded to the anesthesia of the sea urchins, and to the filtration of the seawater sample, following the protocol (***what protocol? is there a publication?***). All sample types were then stored at -80°C until dissection. Before proceeding to the dissection, the size (equatorial diameter) and the presence of albinism on the sea urchin was noted. RRB and EKA then dissected all the collected sea urchins following the protocol in order to obtain the gut pellets (Suppl. File 1). At the end of the sampling period (December 2023), we obtained 45 sea urchin gut pellets, 14 sediment samples, and 16 seawater samples. All the metadata can be found in Supplementary File 2. Each sample of gut pellet and sea water was then divided into two sets, one for downstream microbial and one for metabolic processes (min 50 mg of gut pellets and 50 ml of seawater per sample).

2.1.Microbial characterization

The DNA of the 3 sample types (gut pellets, sediments, sea water) was then extracted using DNeasy PowerSoil Pro kit (QIAGEN LLC, Germantown Road, Maryland, United States) and quantified using Qubit® dsDNA HS assay kit. We normalized the DNA to 4nM during the 16S rRNA gene library preparation process. Using region-specific primers that include sequencer adapter sequences used in the Illumina flowcell, we used the Earth Microbiome Project standard protocols (Caporaso et al., 2023) <-- [**added this. Please confirm**] to amplify the hypervariable region V4 of the 16S ribosomal RNA gene (~291 bp) using the universal bacterial primers 515F (5′GTGCCAGCMGCCGCGGTAA3′) and 806R (5′GGACTACHVGGGTWTCTAAT3′). Amplicons were measured with a plate reader (Infinite® 200 PRO, Tecan) and PicoGreen (Invitrogen). Volumes of all the products were combined into one tube after they were quantified, resulting in an equimolar representation of each amplicon. This pool was then cleaned up using AMPure XP Beads (Beckman Coulter), and finally quantified using a fluorometer (Qubit, Invitrogen). Customized sequencing was outsourced to Argonne National Laboratory in Illinois, USA, utilizing a 2 × 150 bp paired-end sequencing kit with an Illumina MiSeq. The sequencer's reads and the metadata that went along with them were uploaded to QIITA study ID 15616 (version 2024.02).

The downstream microbial analyses were performed locally in QIIME2 (version 2024.05) and include demultiplexing, trimming at 250 bp, denoising. Denoising method was done using DADA2 (Callahan et al 2016) and included quality filtering, removing chimeric sequences, combining paired-ends, and eliminating singleton reads in order to join, denoise, and duplicate sequences. The bacterial sequences were classified using the Greengenes reference database and taxonomy files (McDonald et al 2024), trained with classify-sklearn function in QIIME2. Microbiota analyses (diversity and composition) were divided into four sets: (1) sample types (gut digesta pellet, seawater, or sediment), (2) gut digesta pellet between the different collection locations, (3) gut digesta pellet between the different collection periods, and (4) gut digesta pellet between the different collection locations considering the location periods. For each set, beta- and alpha-diversity were calculated, and taxonomic compositions (taxa barplots) and putative biomarker taxa, e.g, MaAsLin (Mallick et al. 2021) and/or LEfSe (Segata et al. 2011) are represented. To correct for difference in extraction dates, those were added from the metadata as a reference for the MaAsLin analysis. In addition, for collection locations (set of analyses#2), ecological and metabolic function inference (FAPROTAX, Louca et al. 2016) and putative biomarker data (LefSe) were computed.

[<u>table of contents</u> | <u>back to top</u>]

Related Publications

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., … Asnicar, F. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature Biotechnology, 37(8), 852–857. doi:<u>10.1038/s41587-019-0209-9</u>
Software

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods, 13(7), 581–583. doi:<u>10.1038/nmeth.3869</u>
Software

Caporaso, J.G, Ackermann, G., Apprill, A., Bauer, M., Berg-Lyons, D., Betley, J., Fierer, N., Fraser, L., A. Fuhrman, J., A. Gilbert, J., Gormley, N., Humphrey, G., Huntley, J., K. Jansson, J., Knight, R., L. Lauber, C., A. Lozupone, C., McNally, S., M. Needham, D., … Weber, L. (2023). Earth Microbiome Project (EMP) 16S Illumina Amplicon Protocol v2. https://doi.org/<u>10.17504/protocols.io.kqdg3dzzl25z/v2</u>
Methods

Gonzalez, A., Navas-Molina, J. A., Kosciolek, T., McDonald, D., Vázquez-Baeza, Y., Ackermann, G., DeReus, J., Janssen, S., Swafford, A. D., Orchanian, S. B., Sanders, J. G., Shorenstein, J., Holste, H., Petrus, S., Robbins-Pianka, A., Brislawn, C. J., Wang, M., Rideout, J. R., Bolyen, E., … Knight, R. (2018). Qiita: rapid, web-enabled microbiome meta-analysis. Nature Methods, 15(10), 796–798. https://doi.org/<u>10.1038/s41592-018-0141-9</u>
Software

Louca, S., Parfrey, L. W., & Doebeli, M. (2016). Decoupling function and taxonomy in the global ocean microbiome. Science, 353(6305), 1272–1277. https://doi.org/<u>10.1126/science.aaf4507</u>
Software

Mallick, H., Rahnavard, A., McIver, L. J., Ma, S., Zhang, Y., Nguyen, L. H., Tickle, T. L., Weingart, G., Ren, B., Schwager, E. H., Chatterjee, S., Thompson, K. N., Wilkinson, J. E., Subramanian, A., Lu, Y., Waldron, L., Paulson, J. N., Franzosa, E. A., Bravo, H. C., & Huttenhower, C. (2021). Multivariable association discovery in population-scale meta-omics studies. PLOS Computational Biology, 17(11), e1009442.

https://doi.org/[10.1371/journal.pcbi.1009442](https://doi.org/10.1371/journal.pcbi.1009442)
Software

McDonald, D., Jiang, Y., Balaban, M., Cantrell, K., Zhu, Q., Gonzalez, A., Morton, J. T., Nicolaou, G., Parks, D. H., Karst, S. M., Albertsen, M., Hugenholtz, P., DeSantis, T., Song, S. J., Bartko, A., Havulinna, A. S., Jousilahti, P., Cheng, S., Inouye, M., … Knight, R. (2023). Greengenes2 unifies microbial data in a single reference tree. Nature Biotechnology, 42(5), 715–718. https://doi.org/[10.1038/s41587-023-01845-1](https://doi.org/10.1038/s41587-023-01845-1)
Software

Ruiz-Barrionuevo, J. M., Kardas, E., Rodríguez-Barreras, R., Quiñones-Otero, M. A., Ruiz-Diaz, C. P., Toledo-Hernández, C., & Godoy-Vitorino, F. (2024). Shifts in the gut microbiota of sea urchin Diadema antillarum associated with the 2022 disease outbreak. Frontiers in Microbiology, 15. https://doi.org/[10.3389/fmicb.2024.1409729](https://doi.org/10.3389/fmicb.2024.1409729)
Results

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., & Huttenhower, C. (2011). Metagenomic biomarker discovery and explanation. Genome Biology, 12(6). https://doi.org/[10.1186/gb-2011-12-6-r60](https://doi.org/10.1186/gb-2011-12-6-r60)
Software

[table of contents | back to top]

Related Datasets

IsRelatedTo

Chorna, N., Toledo-Hernandez, C., Ruiz-Diaz, C. P., Kardas, E., Godoy-Vitorino, F. (2025) **Metabolite and microbiota data from sea urchin specimens and seawater samples collected in 2023 from four locations in coastal waters of Puerto Rico.** Biological and Chemical Oceanography Data Management Office (BCO-DMO). (Version 1) Version Date 2025-10-10 http://lod.bco-dmo.org/id/dataset/986552 [view at BCO-DMO]

Godoy-Vitorino, F., Toledo-Hernandez, C., Kardas, E., Chorna, N., Rodriguez-Barreras, R., Ruiz-Diaz, C. P. (2025) **Capture data and DNA yields from sea urchin specimens plus surrounding water and sediments collected in 2023 from four locations along the coast of Puerto Rico.** Biological and Chemical Oceanography Data Management Office (BCO-DMO). (Version 1) Version Date 2025-11-14 http://lod.bco-dmo.org/id/dataset/985892 [view at BCO-DMO]

Results

University of California San Diego Microbiome Initiative. uropean Nucleotide Archive (ENA) (2023-11-20). Microbiota associated with the 2022 Diadema antillarum die-off in Puerto Rico [Project: PRJEB70304]. Retrieved from https://www.ebi.ac.uk/ena/browser/view/PRJEB70304.

[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	Illumina MiSeq at Argonne National Lab
Generic Instrument Name	Automated DNA Sequencer
Dataset-specific Description	Customized sequencing was outsourced to Argonne National Laboratory in Illinois, USA, utilizing a 2 × 150 bp paired-end sequencing kit with an Illumina MiSeq.
Generic Instrument Description	A DNA sequencer is an instrument that determines the order of deoxynucleotides in deoxyribonucleic acid sequences.

Dataset-specific Instrument Name	caliper
Generic Instrument Name	calipers
Dataset-specific Description	Sea urchin tests were measured using a caliper
Generic Instrument Description	A caliper (or "pair of calipers") is a device used to measure the distance between two opposite sides of an object. Many types of calipers permit reading out a measurement on a ruled scale, a dial, or a digital display.

Dataset-specific Instrument Name	Infinite® 200 PRO, Tecan Plate Reader
Generic Instrument Name	plate reader
Dataset-specific Description	Amplicons were measured with a plate reader (Infinite® 200 PRO, Tecan) and PicoGreen (Invitrogen) dye.
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 uL 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 µL 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.

Dataset-specific Instrument Name	Invitrogen Qubit fluorometer
Generic Instrument Name	Qubit fluorometer
Dataset-specific Description	Volumes of all the products were combined into one tube after they were quantified, resulting in an equimolar representation of each amplicon which was then quantified using a fluorometer (Qubit, Invitrogen).
Generic Instrument Description	Benchtop fluorometer. The Invitrogen Qubit Fluorometer accurately and quickly measures the concentration of DNA, RNA, or protein in a single sample. It can also be used to assess RNA integrity and quality. Manufactured by Invitrogen, Carlsbad, CA, USA (Invitrogen is one of several brands under the Thermo Fisher Scientific corporation.)

Project Information

RAPID: The black urchin (Diadema antillarum) massive resurgent die-off: Causes, demographic, and community consequences (Diadema antillarum die-off associated microbiota in Puerto Rico)

Website: https://storymaps.arcgis.com/stories/40c072999aab4662b5fff6555a8c77b9

Coverage: Caribbean coral reefs. Fringing reefs located in the northern coast of Puerto Rico.

In recent decades, many marine species in Caribbean coral reef ecosystems have been impacted by disease. A well-documented mass mortality event affecting the long-spined black sea urchin Diadema antillarum in the early 1980s stands out because it had wide-ranging impacts on reef ecosystems. The urchins function as gatekeeper grazers, feeding mainly on macroalgae and preventing algae from overgrowing reefs. In the 1980s, an unknown disease killed over 90% of these urchins across the Caribbean, changing the reefscape from coral to algal dominated. Nearly 40 years later, black sea urchin populations have yet to recover. In early 2022, a new mortality event of D. antillarum was reported along the Caribbean, including Puerto Rico. This RAPID project is identifying the microbes involved in the current mortality event. The investigators are also assessing urchin populations under contrasting environmental conditions and disease incidences. Results are providing a better understanding of the causes and consequences of disease in Diadema, and insights learned may help prevent or mitigate future mortality events. The project is providing training for underrepresented undergraduate and graduate students in microbiology, bioinformatics, biochemistry, and ecology.

With use of multi-omics technology, this RAPID project is advancing our understanding of the current die-off of Diadema in the Caribbean, including host-pathogen interactions and how these are influenced by environmental factors. The investigators are pursuing the following questions: (1) Are the microbial and biochemical profiles of diseased urchins similar to healthy ones? (2) Do environmental factors, i.e., temperature, salinity, pH, and dissolved oxygen, influence disease incidence? (3) Are different size classes of urchins differentially being affected by the disease? This project is focusing on four study sites along the eastern and northern coast of Puerto Rico, where demographic and biochemical data have been previously collected. At three-month intervals, healthy and diseased D. antillarum urchins are being evaluated for changes in the microbiome using metagenomic and untargeted metabolomic strategies. In addition, urchins from eight transects per site are being counted and measured to determine how disease modulates life-history traits, i.e., population size-structure and density, and to assess disease incidence. The relationship between disease incidence and environmental measurements is being assessed. These multidisciplinary approaches and cutting-edge techniques, combined with the existing demographic and microbial data, make this a one-of-a-kind project to study the progression and effects of the disease at the metabolic, microbial, population, and ecosystem levels.

Funding

Funding Source	Award
NSF Division of Ocean Sciences (NSF OCE)	OCE-2243580

[table of contents | back to top]