

# Analysis protocols, sample information, code, and datasets associated with the manuscript Structure and long-term stability of the microbiome in diverse diatom cultures from samples collected in Guam, California, and Gulf of Mexico between 2008 and 2016

**Website:** <https://www.bco-dmo.org/dataset/855750>

**Data Type:** Other Field Results

**Version:** 1

**Version Date:** 2021-07-15

## Project

» [Collaborative Research: Ecology and Evolution of Microbial Interactions in a Changing Ocean](#) (LTPE)

Contributors	Affiliation	Role
<a href="#">Morris, James Jeffrey</a>	University of Alabama at Birmingham (UA/Birmingham)	Principal Investigator
<a href="#">Ashworth, Matt</a>	University of Texas at Austin (UT Austin)	Co-Principal Investigator
<a href="#">York, Amber D.</a>	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

## Abstract

Analysis protocols, sample information, code, and datasets associated with the manuscript Structure and long-term stability of the microbiome in diverse diatom cultures. Sequence data is available from the NCBI SRA archive, BioProject PRJNA706454.

---

## Table of Contents

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

## Coverage

**Spatial Extent:** N:33.71 E:-80.78 S:13.25 W:144.7

**Temporal Extent:** 2008-08-18 - 2015-11-19

---

## Dataset Description

See "Data Files" section for access to download the data and analysis code "Ashworth\_data\_and\_analysis.zip". The sample information and genetic accession identifiers are available as a data table from this page.

## Methods & Sampling

### Location:

Lab work performed at University of Alabama at Birmingham and University of Texas at Austin; Samples

collected from Guam, California, and Gulf of Mexico.

### **Sampling and analytical procedures:**

Cultures were obtained from sites in Guam, California, and the Gulf of Mexico near Florida. Cells were isolated from collections made primarily by plankton net with 20 m mesh or by collection of the top millimeters of sediment from the benthos. Individual cells were isolated by glass Pasteur pipet (Andersen) into 15x100 mm glass test tubes in approximately 12 mL of liquid f/2 medium with double the normal concentration of silica (Guillard, Andersen) at a salinity of 32-35 ppt. The f/2 base was seawater collected from the Texas coast of the Gulf of Mexico and passed through a 0.22 um filter. Isolates were maintained under natural light from a north facing window between 20-24oC, or in a Percival growth chamber on a 12:12 light:dark cycle at 27oC in the case of the isolates from Guam. Once unialgal growth was confirmed by microscopy, strains were maintained in triplicate with cells (approximately 0.25 mL) from one tube transferred to three fresh media tubes every 3-4 months.

During the strain transfer cycle, one replicate tube was harvested by Pasteur pipet into a 1.5 mL microcentrifuge tube and centrifuged in an Eppendorf 5414 C microcentrifuge at 8,000 rpm for 10 minutes. Liquid media was decanted off the pellet, and the pellet was stored at -80oC until DNA extraction. DNA was extracted from the pellets using a MoBio Powersoil DNA kit, following the manufacturers protocol.

The V4 region of the bacterial 16S rRNA gene was amplified using bar-coded PCR primers (Caporaso), purified by gel electrophoresis, and then sequenced using the Illumina MiSeq platform. Quality control and sequence analysis was performed in mothur according to the mothur MiSeq SOP protocol (Schloss, Kozich).

Sequence data is available from the NCBI SRA archive, BioProject PRJNA706454.

### **Species List (ScientificName,AphialD)**

Astrosyne radiata,837899  
Roundia cardiophora,627356  
Florella pascuensis,646734  
Triceratium dubium,418600  
Striatella unipunctata,149177  
Hanicella moenia,842549  
Paralia longispina,708079  
Astrosyne radiata,837899  
Neosynedra provincialis,175369  
Leptocylindrus danicus,149106

### **Data Processing Description**

R version 4.0.3

mothur version 1.42.3

BCO-DMO Data Manager processing notes:

\* Species names in file axes.csv run through the World Register of Marine Species (WoRMS) taxa match tool to check taxonomic names. All names matched accepted names exactly as of 2021-07-15. Unique species list with associated AphialDs added to the metadata.

\* Sample collection information and accession numbers at NCBI extracted using the NCBI Run Selector. Sample information imported into the BCO-DMO data system.

\* Latitude and longitude split into individual columns and converted to decimal degrees. e.g. lat\_lon "24.84 N 80.78 W" -> lat: 24.84, lon: -80.78

\* After email correspondence about a coordinate outlier, the latitude for "USA: San Pedro California, kelp bed" was changed from 38.71 to 33.71 after the submitters double-checked their notes.

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

---

### **Data Files**

## File

### File Ashworth\_data\_and\_analysis

filename: Ashworth\_data\_and\_analysis.zip

(ZIP Archive (ZIP), 18.51 KB)  
MD5:9bcfe24aa36cdc39ec9fc7d1f88a477a

READ ME for Filho et al. (2021) data and data analysis package. Code and data files described below are packaged within Ashworth\_data\_and\_analysis.zip.

References to figures and tables below refer to the results paper "Structure and Long-Term Stability of the Microbiome in Diverse Diatom Cultures," Filho et al. (2021).

This data archival package contains 11 files:

1. This readme.txt file
2. Ashworth.mothurcode.txt which contains the mothur code used to analyze 16S sequencing data
3. Ashworth.rcode.txt which contains the R code used for some of the statistical analysis of the mothur output
- 4-11. Comma-separated values spreadsheets containing the raw data used in the R analyses.

Ashworth.mothurcode.txt:

To execute this code, first download the fastq sequence files from the NCBI SRA archive, BioProject PRJNA706454. Execute mothur from the directory containing these files, and then run each line in order. You will need to adjust the paths to the Silva 16S databases based on your own system; instructions on where to find these databases can be found on the mothur wiki page.

NOTE: there are a few lines of R code commented into the mothur code. These require a package called SRS that executes the ranked subsampling algorithm described in the manuscript text. The comments explain how to transfer your mothur data to R, execute the SRS code, and then transfer the output back into mothur.

Throughout the mothur code there are commented lines showing output relevant to our data analysis. These correspond to results reported in the manuscript and can be helpful guideposts if you are trying to replicate our results.

Ashworth.rcode.txt

To execute this code, set your R working directory to the location of the .csv files contained in this data archive. You should then be able to run all of the code at once, replicating our statistical analyses and re-creating our figures. Key results are included as commented lines.

NOTE: at the end of this file we have included, as commented lines, the results of our online blastn analyses with full details on the best hits for our unidentified bacteria.

Ashworth\_Culture1.csv

This file shows the relative abundances of the 10 most common OTUs in Culture 1 at each of the 4 sampled time points. The top row indicates the elapsed time since cultivation for each sample. This file is one of the inputs used to create the Muller plot in Figure 2.

Ashworth.Muller.csv

This file is the other input necessary for creating Figure 2. All it shows is that none of the OTUs are lineal descendants of any of the others.

axes.csv

**File** is the main data file containing the mothur output regarding diversity, culture identity, and ordination results. Columns are as follows:

Culture: Unique diatom cultures as described in Table 1  
Group: Code signifying which fastq files correspond to each sample  
Species: Diatom species  
Site: Which of the specific sampling locations the culture was collected at  
Locale: More coarse-grained region where the culture was collected  
Class: Diatom class  
Order: Diatom order  
Time: Time between culture isolation and DNA extraction  
MostAbundOTU: OTU number that was most abundant in each sample  
MostAbundTax: Taxonomy of the most abundant OTU  
PropMostAbund: Relative abundance of the most abundant OTU  
PropUbiquitous: Proportion of each sample comprised of the 32 ubiquitous OTUs  
NumberUbiquitous: Number of the 32 ubiquitous OTUs detected in the sample  
axis1, axis2, axis3: coordinates for each sample from NMDS analysis  
chao, chao\_lci, chao\_hci: chao index with low and high confidence intervals  
coverage: estimated sequencing coverage of the sample  
sobs: number of OTUs in the sample  
shannon, shannon\_lci, shannon\_hci: Shannon diversity index with low and high confidence intervals  
invsimpson, invsimpson\_lci, invsimpson\_hci: Inverse Simpson diversity index with low and high confidence intervals

CorrAxes.csv

These coordinates describe the vectors of the 10 most abundant OTUs with a significant impact on the position of samples in the NMDS plot. The "Cultures" column indicates how many different cultures each OTU was detected in.

SharedOTUs.csv, SharedOTUs.astrosyne.csv, SharedOTUs.gabgab.csv

These files show the number of OTUs shared between pairs of samples. Files show either all samples, only the samples from *Astrosyne radiata* cultures, or only the samples from culture originally collected at Gab Gab Beach in Guam.

ubiquitous.csv

This file shows the relative abundances of the 32 ubiquitous OTUs in each of the 15 samples. It was used to create the hierarchical clustering plot in Figure 3.

**longterm\_microbe\_sample\_info.csv**

(Comma Separated Values (.csv), 3.45 KB)  
MD5:28901e47972aacc30099bd4ee90277ba

Primary data file for dataset ID 855750

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

## Related Publications

Barreto Filho, M. M., Walker, M., Ashworth, M. P., & Morris, J. J. (2021). Structure and Long-Term Stability of the Microbiome in Diverse Diatom Cultures. *Microbiology Spectrum*. doi:10.1128/spectrum.00269-21

<https://doi.org/10.1128/Spectrum.00269-21>

### Results

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., ... Knight, R. (2010). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences*, 108(Supplement\_1), 4516–4522. doi:[10.1073/pnas.1000080107](https://doi.org/10.1073/pnas.1000080107)

### Methods

Guillard, R. R. L. (1975). Culture of Phytoplankton for Feeding Marine Invertebrates. *Culture of Marine Invertebrate Animals*, 29–60. doi:[10.1007/978-1-4615-8714-9\\_3](https://doi.org/10.1007/978-1-4615-8714-9_3)

### Methods

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., & Schloss, P. D. (2013). Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. *Applied and Environmental Microbiology*, 79(17), 5112–5120. doi:[10.1128/aem.01043-13](https://doi.org/10.1128/aem.01043-13)

### Methods

R Core Team (2020). R: A language and environment for statistical computing. R v4.0.3. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>

### Software

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., ... Weber, C. F. (2009). Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Applied and Environmental Microbiology*, 75(23), 7537–7541. doi:[10.1128/aem.01541-09](https://doi.org/10.1128/aem.01541-09) <https://doi.org/10.1128/AEM.01541-09>

### Software

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

---

## Related Datasets

### IsRelatedTo

Morris, J. J., University of Alabama at Birmingham (2021). Structure and Long-Term Stability of the Microbiome in Diverse Diatom Cultures. 2021/03. In NCBI:BioProject: PRJNA706454. Bethesda, MD: National Library of Medicine (US), National Center for Biotechnology Information; Available from: <http://www.ncbi.nlm.nih.gov/bioproject/PRJNA706454>.

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

---

## Parameters

Parameter	Description	Units
BioProject	NCBI BioProject	unitless
BioSample	NCBI BioSample	unitless
SRA_Study	NCBI Sequence Read Archive (SRA) Study identifier	unitless
Run	NCBI Sequence Read Archive (SRA) Run identifier	unitless
collection_date	Sample collection date in format yyyy-mm-dd	unitless
lat	Sample collection latitude	decimal degrees
lon	Sample collection longitude	decimal degrees
env_biome	Collection location environmental biome (e.g. kelp bed)	unitless
env_feature	Collection location environment feature (e.g. open water)	unitless
env_material	Collection location material (e.g. water)	unitless
geo_loc_name_country	Collection location country	unitless
geo_loc_name_country_continent	Collection location continent	unitless
geo_loc_name	Collection geolocation name (e.g. "Guam: Achang Reef")	unitless
Organism	Organism sampled (e.g. algae metagenome)	unitless
Sample_Name	Sample name	unitless
sample_title	Sample title	unitless
Description	Sample description	unitless

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

## Instruments

<b>Dataset-specific Instrument Name</b>	Illumina MiSeq platform
<b>Generic Instrument Name</b>	Automated DNA Sequencer
<b>Generic Instrument Description</b>	General term for a laboratory instrument used for deciphering the order of bases in a strand of DNA. Sanger sequencers detect fluorescence from different dyes that are used to identify the A, C, G, and T extension reactions. Contemporary or Pyrosequencer methods are based on detecting the activity of DNA polymerase (a DNA synthesizing enzyme) with another chemoluminescent enzyme. Essentially, the method allows sequencing of a single strand of DNA by synthesizing the complementary strand along it, one base pair at a time, and detecting which base was actually added at each step.

<b>Dataset-specific Instrument Name</b>	Eppendorf 5414 C microcentrifuge
<b>Generic Instrument Name</b>	Centrifuge
<b>Generic Instrument Description</b>	A machine with a rapidly rotating container that applies centrifugal force to its contents, typically to separate fluids of different densities (e.g., cream from milk) or liquids from solids.

## Project Information

### Collaborative Research: Ecology and Evolution of Microbial Interactions in a Changing Ocean (LTPE)

**Coverage:** Lab work: Birmingham, Alabama and New York, New York. Field Work: Bermuda Atlantic Time Series.

#### *NSF Award Abstract:*

Carbon dioxide released from fossil fuels is causing the ocean to become more acidic. Much attention has been given to how this will affect shelled animals like corals, but acidification also affects the algae that form the base of the ocean food chain. It is possible that future algal communities will look very different than they do today, with potentially negative consequences for fisheries, recreation, and climate. Alternatively, it is possible that these algae will be able to adapt rapidly enough to avoid the worst of it. This study looks at algae adapting to acidification in real time in the lab, focusing on "marketplace" interactions between the algae and the bacteria they live alongside. The researchers also go to sea to learn whether adaptations from the lab experiments are beneficial under real-world conditions. Ultimately, this project is helping scientists better understand how the ocean's most important and most overlooked organisms will respond to the changes humans are causing in their habitat. The researchers also use their scientific work to create fun educational opportunities from grade school to college, including agar art classes where students learn about microbial ecology by "painting" with freshly-isolated ocean bacteria.

The effect of ocean acidification on calcifying organisms has been well-studied, but less is known about how changing pH will affect phytoplankton. Previous work showed that the mutualistic interaction between the globally abundant cyanobacterium *Prochlorococcus* and its "helper" bacterium *Alteromonas* broke down under projected future CO<sub>2</sub> conditions, leading to a strong decrease in the fitness of *Prochlorococcus*. It is possible that such interspecies interactions between microbes are important for many ecological processes, but a lack of understanding of how these interactions evolve makes it difficult to predict how important they are. This project is using laboratory evolution experiments to discover how evolution shapes the interactions between bacteria and algae like *Prochlorococcus*, and how these co-evolutionary dynamics might influence the biogeochemical processes that shape Earth's climate. Four research cruises to the Bermuda Atlantic Time Series are also planned to study how natural algal/bacterial communities respond to acidification, and whether evolved microbes from laboratory experiments have a competitive advantage in complex, natural communities exposed to elevated CO<sub>2</sub>. The ultimate goal of this project is to gain a mechanistic understanding of microbial interactions that can be used to inform models of Earth's oceans and biological feedbacks on global climate.

## Funding

Funding Source	Award
<a href="#">NSF Division of Ocean Sciences (NSF OCE)</a>	<a href="#">OCE-1851085</a>