Dataset: Metabarcoding zooplankton at station ALOHA: Operational taxonomic unit (OTU) tables and fasta files for representative sequences from each OTU (Plankton Population Genetics project)

Project(s): Basin-scale genetics of marine zooplankton (Plankton Population Genetics)

Abstract: This data consists of metabarcoding data for the zooplankton community in the epipelagic, mesopelagic and upper bathypelagic zones (0-1500m) of the North Pacific Subtropical Gyre. The goal of this study was to assess the hidden diversity present in zooplankton assemblages in midwaters, and detect vertical gradients in species richness, depth distributions, and community composition of the full zooplankton assemblage. Samples were collected in June 2014 from Station ALOHA (22.75,-158.00) using a 1-meter square Multiple Opening and Closing Nets and Environmental Sampling System (MOCNESS, 200um mesh). Next generation sequence data (Illumina MiSeq, V3 chemistry, 300-bp paired-end) of the zooplankton assemblage derive from amplicons of the V1-V2 region of 18S rRNA (primers described in Fonseca et al. 2010). These data include read count abundance information for molecular OTUs from both holoplanktonic and meroplanktonic taxa. For a complete list of measurements, refer to the supplemental document 'Field_names.pdf', and a full dataset description is included in the supplemental file 'Dataset_description.pdf'. The most current version of this dataset is available at: http://www.bco-dmo.org/dataset/700279

Description: Operational taxonomic unit (OTU) tables and fasta files for representative sequences from each OTU

This data submission consists of metabarcoding data for the zooplankton community in the epipelagic, mesopelagic and upper bathypelagic zones (0-1500m) of the North Pacific Subtropical Gyre. The goal of this study was to assess the hidden diversity present in zooplankton assemblages in midwaters, and detect vertical gradients in species richness, depth distributions, and community composition of the full zooplankton assemblage. Samples were collected in June 2014 from Station ALOHA (22.75, -158) using a 1-meter square Multiple Opening and Closing Nets and Environmental Sampling System (MOCNESS, 200um mesh), on R/V Falkor cruise FK140613. Next generation sequence data (Illumina MiSeq, V3 chemistry, 300-bp paired-end) of the zooplankton assemblage derive from amplicons of the V1-V2 region of 18S rRNA (primers described in Fonseca et al. 2010). The data includes sequences and read count abundance information for molecular OTUs from both holoplanktonic and meroplanktonic taxa. All results derive from analyses in mothur v1.36.1 (Schloss et al. 2009, Kozich et al. 2013).

Tables from four analyses are included in this submission:

1) 97_OTUtable: read counts for each OTU (clustered at 97% similarity) across 54

samples (depth, size fractionated), with NCBI BLAST results for the top representative sequence from each OTU.

2) 97_OTU_subsampled: read counts for each OTU (clustered at 97% similarity), subsampled for even sequencing coverage across 54 samples (depth, size fractionated, with NCBI BLAST results for the top representative sequence from each OTU.

3) 99_OTU: read counts for each OTU (clustered at 99% similarity) across 54 samples (depth, size fractionated), with NCBI BLAST results for the top representative sequence from each OTU.

4) 99_OTU_subsampled: read counts for each OTU (clustered at 99% similarity), subsampled for even sequencing coverage across 54 samples (depth, size fractionated), with NCBI BLAST results for the top representative sequence from each OTU.

fasta files:
RepSeqs97.fasta: representative sequences for each OTU identified by abundance (top read) and with alignment gaps removed. Clustered at 97% similarity.

RepSeqs99.fasta: representative sequences for each OTU identified by abundance (top read) and with alignment gaps removed. Clustered at 99% similarity.

Related dataset containing NCBI accession numbers for sequence data: **Metabarcoding zooplankton at station ALOHA: NCBI SRA accession numbers**

Acquisition Description:

Sample Codes:

Read count parameter names (e.g. FA3_N7_SF1) include the following three codes:

MOCNESS tow
FA3: Night sampling
FA4: Day sampling

Depth range:
N1: 1500-1000m
N2: 1000-700m
N3: 700-500m
N4: 500-300m

N5: 300-200m
N6: 200-150m
N7: 150-100m
N8: 100-50m
N9: 50m-0m

Wet-sieved zooplankton size fractions
SF1: 0.2-0.5 mm
SF2: 0.5-1.0 mm
SF3: 1.0-2.0 mm

Data parameters with the prefix (NCBI_) are from the National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/)

Data parameters with the prefix (silva_) are from the Silva database (https://www.arb-silva.de/).

Processing Description: BCO-DMO processing notes:
* commas in the data were replaced with semicolons to support export as csv format.

Deployment Information

Deployment description for R/V Falkor FK140613

Student Cruise #3 More about this cruise from the Schmidt Ocean Institute page:https://schmidtocean.org/cruise/net-gains-at-station-aloha/

Instrument Information

Instrument	Illumina MiSeq
Description	Illumina MiSeq using V3 chemistry (300-bp, paired-end)
Generic Instrument Name	Automated DNA Sequencer
Generic Instrument Description	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.

Instrument	quantitative PCR by the Evolutionary Genetics Core Facility (Hawaii Institute of Marine Biology)
Description	*local description not specified*
Generic Instrument Name	PCR Thermal Cycler
Generic Instrument Description	General term for a laboratory apparatus commonly used for performing polymerase chain reaction (PCR). The device has a thermal block with holes where tubes with the PCR reaction mixtures can be inserted. The cycler then raises and lowers the temperature of the block in discrete, pre-programmed steps. (adapted from http://serc.carleton.edu/microbelife/research_methods/genomics/pcr.html)

Instrument	Agilent 2100 Bioanalyzer
Description	*local description not specified*
Generic Instrument Name	Bioanalyzer
Generic Instrument Description	A Bioanalyzer is a laboratory instrument that provides the sizing and quantification of DNA, RNA, and proteins. One example is the Agilent Bioanalyzer 2100.