Population genomics study on the planktonic copepod Pleuromamma xiphias: RADSeq data and metadata (Plankton Population Genetics project)

Website: https://www.bco-dmo.org/dataset/684156
Data Type: Cruise Results
Version: 0
Version Date: 2017-03-08

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

Contributors	Affiliation	Role
Goetze, Erica	University of Hawaii at Manoa (SOEST)	Principal Investigator
Copley, Nancy	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

Abstract
Population genomics study on the planktonic copepod Pleuromamma xiphias: RADSeq data and metadata

Table of Contents

Coverage

Spatial Extent: **N**:39.647 **E**:-22.4662 **S**:-39.88 **W**:-42.0517
Temporal Extent: 2012-10-16 - 2012-11-17

Dataset Description

This dataset includes RADSeq data as well as NCBI Short Read Archive (SRA) BioProject and BioSample accessions and collection metadata from animals collected on Atlantic Meridional Transect 22 (AMT22) in Oct. - Nov. 2012. Field work was conducted on the RRS James Cook cruise JC079. See NCBI GenBank Bioproject PRJNA368728 [https://www.ncbi.nlm.nih.gov/bioproject/PRJNA368728]

The sequences are embargoed until 2019-12-01. Please check back after that date.

README for processed data files associated with the article: “Genetic isolation between populations in distinct pelagic habitats of the oceanic copepod Pleuromamma xiphias” – Authors: Lauren Van Woudenberg, Matthew Iacchei, Jonathon Whitney, Katja T. C. A. Peijnenburg, Erica Goetze (2017?). in preparation for submission to Molecular Ecology.

Data files included in this archive:
(1) Supplementary Table 1. Overview of RADSeq data for all animals included in the study. VanWoudenberg_et_al_PLXI_RADSeq.xlsx

(2) VCF file used in downstream analyses, including mitochondrial clade 3 animals only. 289 total animals included. M4n3_5X_60%indiv_40%miss_final.vcf.

(3) VCF file used in downstream analyses, for analyses regarding mitochondrial clades 2 & 3 and SNP clusters 1 & 2. 112 total animals included. MTclades_M4n3_5X_60%indiv_40%miss.vcf

(4) Summary table and metadata of the sequence files submitted to the NCBI Sequence Read Archive (SRA), with BioProject and BioSample numbers. VanWoudenberg_et_al_2017_SRA_metadata.xlsx

Methods & Sampling

Please refer to the paper for methodological details. If you have further questions, please contact the corresponding author (Dr. Erica Goetze): egoetze[at]hawaii[dot]edu.

From the cruise report:
Sample collection. Plankton samples were collected with 0.71m diameter bongo nets (200, 333 µm), and with an RMT1 midwater trawl (333 µm) that has a nominal mouth area of 1m2. A total of 50 plankton tows were conducted along the cruise leg (Table 1), with 35 tows conducted using the bongo and 14 samples collected with the RMT net. The bongo tows were oblique tows that sampled from between 211 to 488 m depth and the surface (324m average maximum depth of tow). The bongo samples will be used for quantitative estimates of animal abundance along the cruise leg (target species only, tows conducted with timedepth-recorder and flowmeter). The RMT tows were also oblique tows that sampled between 62 to 216 m depth and the surface (153 m average maximum depth of tow). All tows except one (station 42) were conducted at night, in order to efficiently sample the migratory community.

Sample handling and preservation. All plankton from the 200 µm mesh bongo net was preserved immediately in 100% ethyl alcohol for use in molecular studies, including DNA sequencing and microsatellite genotyping (and possibly RAD tag sequencing), in addition to estimates of abundance of target species. Plankton material from the 333 µm mesh bongo net and the RMT net was sorted live immediately following collection, and animals were individually identified, and preserved in acetone, RNALater, cryopreserved, and in some cases used for live imaging prior to preservation. These animals will be used for molecular, genomic and transcriptomic analyses. Both RNA/DNA ratios and prosome length - dry weight relationships will be used as measures of animal condition in copepods. In total, over 17,000 animals from 40 target species were individually sorted and preserved for this panel of measurements. Following live sorting and imaging of the 333 µm samples, the remaining plankton was preserved either in 4% buffered formalin or 100% ethyl alcohol for morphological studies.

Data Processing Description

SEQUENCE DATA FILES
Illumina HiSeq reads are available NCBI Sequence Read Archive (SRA). Libraries were prepared following the ezRAD protocol (Toonen et al. 2013). Sequences from Illumina HiSeq 2500, with quality trimming and adaptor removal using TrimGalore (as follows).

#ADAPTERS
#Illumina TruSeq HT dual-indexed Adapters (96 barcode combinations)
GATCGGAAGAGCACACGTCTGAACTCCAGTCACNNNNNNNNATCTCGTATGCCGTCTTCTGCTTG #Read1 w/ 8 digit wildcard i7 #barcode
GATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTNNNNNNNNGTGTAGATCTCGGTGGTCGCCGTATCATT #Read2 w/ wildcard i5 barcode (reverse complemented)

#TrimGalore Command
#first make directory for cleaned files
mkdir cleaned_for_stacks

##FOR R1 loop for trim_galore##
declare -a TEST=(site09_12 site09_15 site09_18 site09_21 site09_24 site09_13 site09_16 site09_19 site09_22 site09_14 site09_17 site09_20 site09_23)

for i in "${TEST[@]}"; do perl ~/ddocent/trim_galore —-phred33 —-dont_gzip -a gatcggaagagcacacgtctgaactccagtcacnnnnnnnnatctcgtatgccgtcttctgcttg --stringency 5 -e 0.1 -r1 100 --output_dir ./cleaned_for_stacks $i.R1.fq; done

##FOR R2 loop for trim_galore##
declare -a TEST=(site09_12 site09_15 site09_18 site09_21 site09_24 site09_13 site09_16 site09_19 site09_22 site09_14 site09_17 site09_20 site09_23)

for i in "${TEST[@]}"; do perl ~/ddocent/trim_galore —-phred33 —-dont_gzip -a gatcggaagagcgtcgtgtagggaaagagtgtnnnnnnnngtgtagatctcggtggtcgccgtatcatt --stringency 5 -e 0.1 -r1 100 --output_dir ./cleaned_for_stacks $i.R2.fq; done

Contact: Erica Goetze for any questions, or for subsequent use of these data.

BCO-DMO Processing Notes:
added conventional header with dataset name, PI name, version date
modified parameter names to conform with BCO-DMO naming conventions
combined SRA metadata with collection information
converted latitude and longitude to decimal degrees
added links to NCBI GenBank BioProject and BioSample pages

[table of contents | back to top]

Parameters

Parameter	Description	Units
bioproject_accession	NCBI BioProject accession number	unitless
biosample_accession	NCBI BioSample accession number	unitless
library_ID	NCBI Library identifier	unitless
title	NCBI project title	unitless
library_strategy	NCBI term meaning genomic method used for analysis	unitless
library_source	NCBI term meaning the type of genomic material that was analyzed	unitless
library_selection	NCBI term meaning method that the source material was selected	unitless
library_layout	NCBI library layout: Paired-end or Single	unitless
platform	platform used for sequencing	unitless
instrument_model	sequencing instrument model	unitless
design_description	NCBI	unitless
filetype	type of file	unitless
filename	file name	unitless
sample_id	sample identifier	unitless
cruise_id	cruise identifier	unitless
sta	station number	unitless
lat_collection	latitude; north is positive	decimal degrees
lon_collection	longitude; east is positive	decimal degrees
date_collection	collection date formatted as yyyy-mm-dd	unitless
sex	sex of specimens	unitless
mtCOI_Clade	mitochondrial COI clade	unitless
num_Seq_reads_initial	number of sequence reads pre-cleaning	reads
num_Seq_reads_post_cleaning	number of sequence reads post-cleaning	reads

[table of contents | back to top]

Instruments

Dataset-specific Instrument Name	Illumina HiSeq 2500
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.

Dataset-specific Instrument Name	
Generic Instrument Name	Thermal Cycler
Generic Instrument Description	A thermal cycler or "thermocycler" is a general term for a type of laboratory apparatus, commonly used for performing polymerase chain reaction (PCR), that is capable of repeatedly altering and maintaining specific temperatures for defined periods of time. 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. They can also be used to facilitate other temperature-sensitive reactions, including restriction enzyme digestion or rapid diagnostics. (adapted from http://serc.carleton.edu/microbelife/research_methods/genomics/pcr.html)

[table of contents | back to top]

Deployments

JC079

Website	https://www.bco-dmo.org/deployment/540458
Platform	RRS James Cook
Report	http://dmoserv3.whoi.edu/data_docs/Goetze/AMT22_cruise/jc079.pdf
Start Date	2012-10-10
End Date	2012-11-24
Description	The AMT22 cruise set sail from Southampton in the UK on 10 October 2012 and arrived in Punta Arenas, Chile on 24 November 2012. The final cruise report and other cruise information, including all science components, can be found online at the Atlantic Meridional Transect webpage (http://www.amt-uk.org/Cruises), or through the British Oceanographic Data Centre (BODC) (http://www.bodc.ac.uk/projects/uk/amt/). Zooplankton ecology data from the project "Does habitat specialization drive population genetic structure of oceanic zooplankton?" (NSF OCE-1029478) were collected on this cruise.

[table of contents | back to top]

Project Information

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

Coverage: Atlantic Ocean, 46 N - 46 S

Description from NSF award abstract:
Marine zooplankton show strong ecological responses to climate change, but little is known about their capacity for evolutionary response. Many authors have assumed that the evolutionary potential of zooplankton is limited. However, recent studies provide circumstantial evidence for the idea that selection is a dominant evolutionary force acting on these species, and that genetic isolation can be achieved at regional spatial scales in pelagic habitats. This RAPID project will take advantage of a unique opportunity for basin-scale transect sampling through participation in the Atlantic Meridional Transect (AMT) cruise in 2014. The cruise will traverse more than 90 degrees of latitude in the Atlantic Ocean and include boreal-temperate, subtropical and tropical waters. Zooplankton samples will be collected along the transect, and mitochondrial and microsatellite markers will be used to identify the geographic location of strong genetic breaks within three copepod species. Bayesian and coalescent analytical techniques will test if these regions act as dispersal barriers. The physiological condition of animals collected in distinct ocean habitats will be assessed by measurements of egg production (at sea) as well as body size (condition index), dry weight, and carbon and nitrogen content. The PI will test the prediction that ocean regions that serve as dispersal barriers for marine holoplankton are areas of poor-quality habitat for the target species, and that this is a dominant mechanism driving population genetic structure in oceanic zooplankton.

Note: This project is funded by an NSF RAPID award. This RAPID grant supported the shiptime costs, and all the sampling reported in the AMT24 zooplankton ecology cruise report (PDF).

Online science outreach blog at: https://atlanticplankton.wordpress.com

[table of contents | back to top]

Funding

Funding Source	Award
NSF Division of Ocean Sciences (NSF OCE)	OCE-1338959

[table of contents | back to top]