Test dataset created for testing and troubleshooting purposes

Website: https://www.bco-dmo.org/dataset/885287
Version: 1
Version Date: 2024-04-03

Project
» BCO-DMO: Accelerating Scientific Discovery through Adaptive Data Management (BCO-DMO)

Contributors	Affiliation	Role
York, Amber D.	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	Principal Investigator, Co-Principal Investigator
Heyl, Taylor	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	Co-Principal Investigator
Newman, Sawyer	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	Co-Principal Investigator
Soenen, Karen	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	Co-Principal Investigator
Rauch, Shannon	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	Scientist
Gerlach, Dana Stuart	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

Abstract
Raw data and assembled scaffolds for the Atlantic silverside genome Test text: Ipsum lorep test text Another paragraph with stuff Some italics here

Table of Contents

Coverage

Location: A really long description of where the data was collected or focused on.
Spatial Extent: **N**:65 **E**:-30 **S**:60 **W**:-45
Temporal Extent: 2020-01-01 - 2024-04-03

Dataset Description

DNA extracts from the vicinity of Station ALOHA (22.75 N, 158.0 W) just north of Hawaii.

CMORE RNA and DNA Archive

Goal: The goal of this effort is to provide a time series collection of planktonic microbial DNA and RNA depth profiles from the HOT station, for CMORE research. DNA and RNA samples will be available for researchers to conduct and coordinate taxonomic and functional gene studies, for example using PCR and RT-PCR methods.

Where possible, metagenomic datasets will be generated by pyrosequencing for selected profiles, and posted for downloading the blast searching on the MIT-CMORE server (http://genesis2.mit.edu/).

Activity: Collect microbial cell fraction (1.6um prefiltered, >0.22um) from HOT hydrocasts, and extract nucleic acids, for RNA and DNA for downstream analyses by CMORE investigators. Collection depths are 25m, 45m, 75m, 125m, 200m, 500m, 770m and 1000m. Extracted DNA samples will be available for distribution.

Methods & Sampling

Sampling and analytical procedures:

The genomes of two different individuals were sequenced with different approaches:

1. An individual sampled at Jekyll Island, Georgia:

We built a reference genome for the Atlantic silverside through three steps. First, we created a draft assembly using 10x Genomics linked-reads technology (10x Genomics, Pleasanton, CA). Second, we used proximity-ligation data— ChicagoVR (Putnam et al. 2016) and Dovetail

Hi-C (Lieberman-Aiden et al. 2009)—from Dovetail Genomics (Santa Cruz, CA) to increase contiguity, break up mis-joins, and orient and join scaffolds into chromosomes. Finally, we used short-insert reads to close gaps in the scaffolded and error-corrected assembly. The data were generated from muscle tissue dissected from two lab-reared F1 offspring of Atlantic silversides collected from the wild on Jekyll Island, GA, USA (N31.02 ,W81.43 ; the southern end of the species distribution range) in May 2017. For 10x Genomics library preparation, we extracted DNA from fresh tissue from one individual using the MagAttract HMW DNA Kit (Qiagen). Prior to library preparation, we selected fragments longer than 30 kb using a BluePippin device (Sage Science). A 10x Genomics library was prepared following standard procedure and sequenced using two lanes of paired-end 150 bp reads on a HiSeq2500 (rapid run mode) at the Biotechnology Resource Center Genomics Facility at Cornell University. To assemble the linked reads, we ran the program Supernova 2.1.1 (Weisenfeld et al. 2017) from 10x Genomics with varying numbers of reads and compared assembly statistics to identify the number of reads that resulted in the most contiguous assembly. Tissue from the second individual was flash frozen in liquid nitrogen and shipped to Dovetail Genomics, where Chicago and Hi-C libraries were prepared for further scaffolding. These long-range libraries were sequenced on one lane of Illumina HiSeqX using paired-end 150 bp reads. Two rounds of scaffolding with HiRiseTM, a software pipeline developed specifically for genome scaffolding with Chicago and Hi-C data, were run to scaffold and error-correct the best 10x Genomics draft assembly using Dovetail long-range data. Finally, the barcode-trimmed 10x Genomics reads were used to close gaps between contigs as the final step of the HiRise pipeline.

2. An individual sampled in Mumford Cove, Connecticut

This assembly was a lower-quality draft assembly used to identify structural variants in comparison to the chromosome-level assembly from the individual sampled in Georgia

The individual sampled for this assembly was sampled from Mumford Cove, Connecticut (N 41.32 , W 72.02) in June 2016. Genomic DNA was extracted from muscle tissue using the DNeasy Blood and Tissue kit (Qiagen) and normalized to 40 ng/ul. We prepared a genomic DNA library using the TruSeq DNA PCR-free library kit (Illumina) following the manufacturer's protocol for 550 bp insert libraries. The shotgun library was sequenced using paired-end 150 bp reads on an Illumina HiSeq4000. Mate- pair libraries with insert sizes of 3, 5.3, and 8.2 kb were pre- pared at the Huntsman Cancer Institute at the University of Utah using the Nextera Mate Pair Library Prep Kit (Illumina) and sequenced using paired-end 125 bp reads on an Illumina HiSeq2500. We used Trimmomatic 0.36 (Bolger et al. 2014) to remove adapter contamination and low-quality data from both the shotgun and the mate pair libraries and used these filtered reads to assemble a draft assembly and fill assembly gaps with Platanus v.1.2.4 (Kajitani et al. 2014) with the commands assemble, scaffold, and gap_close. Finally,we filtered scaffolds shorter than 1 kb.

Further details of the samples and methodology are available in the following publication:

Tigano, A., Jacobs, A., Wilder, A. P., Nand, A., Zhan, Ye, Dekker, J., and Therkildsen, N. O. 2021. Chromosome-level assembly of the Atlantic silverside genome reveals extreme levels of sequence diversity and structural genetic variation. **Genome Biology and Evolution** 13, evab163

Super and subscript display test

subscripts:

Our experiment assessed the CO_2 sensitivity of embryos and early life stages of BSB at a single static temperature (22°C) and three pCO_2 conditions (~400, ~2200, ~4200 µatm). On May 23rd, 2022, we strip-spawned wild, running-ripe BSB (($N_{female/male}$ = 4/3) to produce viable embryos. Upon water-hardening a 5 ml sample of eggs was randomly allocated to a replicate 19-l rearing container held within one of nine recirculating systems within the Baumann labs automated larval rearing system (ALFiRiS).

superscripts:

At the late-exponential phase, cultures were transferred in triplicate to one of two SN media: (1) +P*i* (45 µmol L^{-1} KH_2PO_4, following

Data Processing Description

All files except those listed with “genome assembly” in the date type column are raw sequence data files that have not undergone any processing.

The two genome assemblies were processed as described under “Sampling and analytical procedures” described above.

Problem Description

No collection issues reported.

No quality issues reported.

[table of contents | back to top]

Supplemental Files

File	
Filelist with species names filename: filename_list_with_species_ids.csv	(Comma Separated Values (.csv), 111.83 KB) MD5:334e664ff92878c63dca431edf70be56
Filelist with the associated species names and identifiers. Identifiers are the aphiaID from the World Register of Marine Species (WoRMS).	

[table of contents | back to top]

Related Publications

Goldstein, J. I., Newbury, D. E., Michael, J. R., Ritchie, N. W. M., Scott, J. H. J., & Joy, D. C. (2018). Scanning Electron Microscopy and X-Ray Microanalysis. Springer Science + Business Media, LLC, New York. (Third edition) https://doi.org/10.1007/978-1-4939-6676-9
Methods

[table of contents | back to top]

Parameters

Parameters for this dataset have not yet been identified

[table of contents | back to top]

Deployments

HOT_cruises

Website	https://www.bco-dmo.org/deployment/58879
Platform	Unknown Platform
Report	http://hahana.soest.hawaii.edu/hot/
Start Date	1988-10-31
Description	Since October 1988, the Hawaii Ocean Time-series (HOT) program has investigated temporal dynamics in biology, physics, and chemistry at Stn. ALOHA (22°45' N, 158°W), a deep ocean field site in the oligotrophic North Pacific Subtropical Gyre (NPSG). HOT conducts near monthly ship-based sampling and makes continuous observations from moored instruments to document and study NPSG climate and ecosystem variability over semi-diurnal to decadal time scales.

[table of contents | back to top]

Project Information

BCO-DMO: Accelerating Scientific Discovery through Adaptive Data Management (BCO-DMO)

NSF Award Abstract:

Scientific research is intrinsically reliant upon the creation, management, analysis, synthesis, and interpretation of data. Once generated, data are essential to demonstrating the veracity and reproducibility of scientific results, and existing data hold great potential to accelerate scientific discovery through reuse. The Biological and Chemical Oceanography and Data Management Office (BCO-DMO) was created in 2006 to assemble, curate, and publicly serve all data and related products resulting from grants funded by the NSF core programs for Biological and Chemical Oceanography, and Office of Polar Programs. BCO-DMO provides limnological and marine chemical, biological, and physical data inventories from several large and intermediate-sized programs, as well as single-investigator projects to support cross-disciplinary collaboration to address pressing environmental questions, problems, and challenges that are exacerbated with the increasing pace of climate change. BCO-DMO is committed to data management capacity building efforts, improving data literacy and increasing science engagement in data management topics through education, training, and outreach. The project collaborates with academic institutions and teachers, where the BCO-DMO database is leveraged for oceanographic curricula, and engages in targeted training of informatics students, cross-pollinating their knowledge with geoscience domain data management.

BCO-DMO's goal is to facilitate the integration of its diverse datasets to enable researchers to achieve a deeper understanding of ocean ecological and biogeochemical systems. As a domain repository, BCO-DMO adds value and improves interoperability of data to support activities such as synthesis and modeling, and the reuse of oceanographic data for new research. Open access to the BCO-DMO database lowers barriers to allow economically challenged countries to gain access to research quality data for field decision support, policy-relevant issues, and educational purposes. The project takes an active role in the exchange of knowledge at national and international geoscience and informatics meetings and workshops, where standards development and adoption occur. BCO-DMO also participates in the development and use of open-source, standards-based technologies that enable interoperable data systems to exchange data and information that will foster next-generation research in all disciplines. While continuing to perform its core mission of data management, BCO-DMO will reconstitute its data infrastructure to mobilize a new adaptive data management strategy for addressing the evolutionary change coinciding with the big data revolution. Leveraging data semantics BCO-DMO will construct a knowledge graph for sustainably operating an adaptive data repository. This infrastructure will support dataset-level and repository-level metrics, an improved data submission experience

and new data and metadata access capabilities. Through declarative workflows, the processing of contributed data will increase in efficiency, and result in actionable provenance records for complete transparency of data curation practices. Taking a holistic perspective on education, outreach and community engagement, formalized programs will be developed to promote data reuse and interest in oceanographic science.

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-1924618

[table of contents | back to top]