Decorator worm Diopatra cuprea mitochondrial locus (COI) sequences from collections in the Gulf of Mexico and eastern United States estuaries from 2009 to 2022

Website: https://www.bco-dmo.org/dataset/999053
Version: 1
Version Date: 2026-05-19

Project
» The genetic legacy of an Asian oyster introduction and its disease-causing parasite (Oyster historical genetics)

Contributors	Affiliation	Role
Sotka, Erik	College of Charleston (CofC)	Principal Investigator
York, Amber D.	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

Abstract

The decorator worm Diopatra cuprea Bosc, 1802 (Annelid; Polychaete; Onuphidae) is an ecosystem engineer within high-salinity estuaries of the southern and eastern United States. A previous study revealed five morphologically cryptic mitochondrial lineages across its broad geographic distribution. We Sanger-sequenced a mitochondrial locus (COI). This dataset includes metadata, methods, links to published code and processed data used to generate figures for results publication Sotka et al. (2023, doi:10.3390/biology12040521) titled "Cryptic mtDNA Diversity of Diopatra cuprea (Onuphidae, Annelida) in the Northwestern Atlantic Ocean." This dataset also includes genetic accession identifiers for sequence data contributed to the National Center for Biotechnology Information (NCBI)'s GenBank database.

Table of Contents

Coverage

Location: Eastern United States and Gulf of Mexico estuaries, soft sediment intertidal zone
Spatial Extent: **N**:42.04647 **E**:-70.073089 **S**:27.4575 **W**:-84.474212
Temporal Extent: 2009-01-01 - 2022-12-31

Dataset Description

See the "Related Datasets" section for genotyped single nucleotide polymorphisms (SNPs) with RADseq that were also generated as part of this study.

Methods & Sampling

Note: This metadata section includes description of a related dataset that was generated as part of this study (see "Related Datasets" section) for that dataset which includes metadata and accessions for genotyped single nucleotide polymorphisms (SNPs) with RADseq.

Diopatra cuprea were sampled from 19 locations on the United States east coast, from Duxbury MA (the genus' northern limit) to St. Teresa Beach FL. Tubes were excavated with a shovel, the worm removed, and antennae clipped off and preserved in 95% ethanol for later DNA extraction. or 2022 samples, twenty-five randomly selected individuals from each population were extracted for DNA, with the exception of Broad River Estuary populations from which 20 and 22 samples were collected. Approximately 25 mg of tissue wet weight was rinsed of ethanol with deionized water and extracted with the Qiagen DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA, USA) following the manufacturer's extraction protocol. Extractions were screened with 1.5% agarose gels and DNA was quantified using a NanoDrop 200 spectrophotometer (Thermo Scientific, Waltham, MA, USA) to obtain concentration values and purity.

A portion of the COI gene was then PCR amplified using the protocol in Berke, et al. (2010, doi:10.1111/j.1466-8238.2009.00509.x). These were cleaned with an EXO-SAP-IT protocol and sent for Sanger sequencing with these same primers at a private company.

Double digest restriction-associated DNA sequencing (or ddRADSeq) library was prepared on 312 samples following the protocol in Parchman et al. (2012,doi: 10.1111/j.1365-294X.2012.05513.x). Briefly, we digested gDNA with two restriction enzymes, EcoRI and MseI, and ligated adaptors containing unique 8 to 10 bp barcodes to the digested DNA of each individual. The products were then PCR amplified in two independent reactions with standard Illumina primers. All amplicons were pooled and shipped to the University of Texas Genomic Sequencing and Analysis Facility or the Tufts University Core Facility, which used Pippin Prep® to isolate the 300–450 bp fraction. This fraction was then single-read sequenced (100 basepairs) with Illumina HiSeq 4000 machine. FASTQ sequences were demultiplexed using custom Unix code. 510 million reads contained barcode sequence (range = 39 to 6.8 million (M) reads per sample; mean = 1.6 M reads), and 234 individuals had at least 200 K reads that were analyzed.

The data table in this dataset 999053_v1_diopatra-coi-haplotype.csv provides the COI sequence accessions in NCBI's GenBank database.

Organism identifier (Life Science Identifier (LSID)):
Diopatra cuprea, Bosc, 1802, LSID(urn:lsid:marinespecies.org:taxname:157339)

BCO-DMO Processing Description

- Loaded tab-separated file "AccessionReport_wMeta.txt" into table "999053_v1_diopatra-coi-haplotype" with header row 1; treated empty strings and "nd" as missing values
- Renamed column "X.Accession" to "X_Accession"
- Set column types: SampleID, X_Accession, hap, pop1, pop2, state as string; lat and lon as number
- Updated field metadata for all columns: SampleID (individuals, standard_name_id 960), X_Accession (GenBank COI accession numbers/NCBI GenBank nucleotide database accession identifier, standard_name_id 873475), hap (unique haplotype, standard_name_id 1741), lat (Latitude, primary parameter, decimal degrees, standard_name_id 730), lon (Longitude, primary parameter, decimal degrees, standard_name_id 731), pop1 (abbreviation of site, standard_name_id 1080), pop2 (long description of site, standard_name_id 1124), state (State, standard_name_id 952)
- Output written to "999053_v1_diopatra-coi-haplotype.csv"

Problem Description

Note: The individual GenBank accessions are provided directly in the data table of this dataset (999053_v1_diopatra-coi-haplotype.csv). They were not associated with a BioProject and thus are not listed under the "Related Datasets" section of this dataset. Example of one of the 153 GenBank accessions: https://www.ncbi.nlm.nih.gov/nuccore/OQ700161.1

[table of contents | back to top]

Related Publications

Berke, S. K., Mahon, A. R., Lima, F. P., Halanych, K. M., Wethey, D. S., & Woodin, S. A. (2010). Range shifts and

species diversity in marine ecosystem engineers: patterns and predictions for European sedimentary habitats. Global Ecology and Biogeography, 19(2), 223–232. Portico. https://doi.org/10.1111/j.1466-8238.2009.00509.x
Methods

Parchman, T. L., Gompert, Z., Mudge, J., Schilkey, F. D., Benkman, C. W., & Buerkle, C. A. (2012). Genome-wide association genetics of an adaptive trait in lodgepole pine. Molecular Ecology, 21(12), 2991–3005. Portico. https://doi.org/10.1111/j.1365-294x.2012.05513.x https://doi.org/10.1111/j.1365-294X.2012.05513.x
Methods

Sotka, E. E., Bell, T., & Berke, S. (2023). Cryptic mtDNA Diversity of Diopatra cuprea (Onuphidae, Annelida) in the Northwestern Atlantic Ocean. Biology, 12(4), 521. https://doi.org/10.3390/biology12040521
Results

Ziegler, A. J., Bell, T. M., Berke, S. K., Strand, A. E., & Sotka, E. E. (2025). Multiple cryptic lineages and restricted gene flow in the decorator worm Diopatra Cuprea. Marine Biology, 172(3). https://doi.org/10.1007/s00227-025-04613-8
Results

[table of contents | back to top]

Related Datasets

IsRelatedTo

Sotka, E. (2026) **Decorator worm Diopatra cuprea single nucleotide polymorphism (SNP) genotypes from collections in the Gulf of Mexico and eastern United States estuaries from 2009 to 2022.** Biological and Chemical Oceanography Data Management Office (BCO-DMO). (Version 1) Version Date 2026-05-08 http://lod.bco-dmo.org/id/dataset/998297 [view at BCO-DMO]
Relationship Description: Data generated from the same study.

Sotka,E., Bell,T. and Berke,S. (2023). Diopatra cuprea isolate MA.BA.15 cytochrome c oxidase subunit I (COX1) gene, partial cds; mitochondrial. NCBI:Genbank: OQ700161.1 [Internet] https://www.ncbi.nlm.nih.gov/nuccore/OQ700161.1

Software

Erik Sotka. (2026). *esotka/DiopatraCOI: DiopatraCOI* (Version v1.0) [Computer software]. Zenodo. https://doi.org/10.5281/ZENODO.18672658 https://doi.org/10.5281/zenodo.18672658

[table of contents | back to top]

Parameters

Parameter	Description	Units
X_Accession	GenBank COI accession numbers. NCBI Genbank nucleotide database accession identifier.	unitless
SampleID	individuals	unitless
state	State	unitless
pop1	abbreviation of site	unitless
pop2	long description of site	unitless
hap	unique haplotype	unitless
lat	Latitude	decimal degrees
lon	Longitude	decimal degrees

[table of contents | back to top]

Instruments

Dataset-specific Instrument Name	Illumina sequencing machine
Generic Instrument Name	Automated DNA Sequencer
Dataset-specific Description	Illumina HiSeq 4000 or Illumina NovaSeq 6000
Generic Instrument Description	A DNA sequencer is an instrument that determines the order of deoxynucleotides in deoxyribonucleic acid sequences.

[table of contents | back to top]

Project Information

The genetic legacy of an Asian oyster introduction and its disease-causing parasite (Oyster historical genetics)

Coverage: Global

NSF abstract:

During the 20th century, the Pacific oyster *Crassostrea gigas* was deliberately introduced from its native range of coastal Asia to the estuaries of six continents. While the introduced Pacific oysters are widely aquacultured and thus can generate local economic wealth, they sometimes outcompete native oysters, and can carry microbial, animal and plant hitchhikers that negatively impact local economies and the ecological functioning of local estuaries. This study comprehensively assesses the pathways and sources of Pacific oyster introductions using a worldwide, population genetic survey. Simultaneously, the study also assesses the pathways and source of one hitchhiking protist (*Haplosporidium nelsoni*) that causes the disease MSX (multinucleated sphere X) in the Virginia oyster (*Crassostrea virginica*) along the eastern seaboard of the United States. One goal of this research is to generate management strategies that combat the negative impacts of the Pacific oyster and its associated invaders, and minimize future invasions. A second goal is to minimize some uncertainty about the population biology of the devastating Haplosporidium parasite, and thus, increase confidence of policy makers who are managing shellfish health, restoration and commerce. By quantifying the pathways and sources of *C. gigas*, this project may inform strategies to combat negative impacts of *C. gigas* and its associated invaders, as well as minimize future invasions. Moreover, quantifying dispersal within and among populations of *H. nelsoni* along the US East Coast will provide perspective on the effectiveness of regional biosecurity measures in preventing the ongoing dispersal of this destructive pathogen via aquaculture. In addition, the project lends itself well to programs that foster critical thinking and research experience among both undergraduate and K-12 students. The project provides opportunities for 6-9 undergraduates to perform research, includes a 2-day workshop on bioinformatics for the wider undergraduate community, and facilitates ongoing opportunities for K-12 students to participate in citizen-science research.

There is a wealth of information on the source, pathways and vectors of *C. gigas* based largely on historical documents but no study has comprehensively tested whether these historical accounts are correct using a worldwide, population genetic survey. Using >14K single-nucleotide polymorphisms (SNPs) from 41 populations across five continents a high level of spatial genetic differentiation was found within the native range and differences in source populations among non-native regions. Preliminary genetic data indicated that the parasitic protist, *Haplosporidium nelsoni* arrived with C. gigas imports to the US Atlantic coastline and then infected the native C. virginica, however the native source populations, the pathways and vector from which *H. nelsoni* arrived remain unknown. This project couples high-throughput sequencing technologies and Approximate Bayesian Computing (ABC)-based models to answer the following: What are the population genomic patterns among *C. gigas* from native and non-native regions? What are the population genomic patterns of *Haplosporidium nelsoni* among Asian and North American Crassostrea gigas and eastern North American *C. virginica*? What were the source populations and invasion pathways of *C. gigas* and *H. nelsoni*? Identifying source locations, pathways and vectors of introduction of C. gigas will provide researchers with a null-model of invasion history for dozens of other non-native species that were transported with *C. gigas*. Currently, there are no verified 'vector maps' for historical shipments of *C. gigas* that are similar to those generated from modern-day or historical shipping records.

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-1924599

[table of contents | back to top]