Metadata for longread sequencing of Carcinus maenas collected from Buzzards Bay, Massachusetts from May 2022 to Aug 2022

Website: https://www.bco-dmo.org/dataset/949666
Data Type: Other Field Results
Version: 1
Version Date: 2025-01-29

Project

» Collaborative Research: Tracking fine-scale selection to temperature at the invasion front of a highly dispersive marine predator (West Coast Carcinus)

Contributors	Affiliation	Role
Tepolt, Carolyn	Woods Hole Oceanographic Institution (WHOI)	Principal Investigator
Mickle, Audrey	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

Abstract

This project explores genomic changes in the invasive European green crab (Carcinus maenas), including at a putative inversion polymorphism. To begin to explore structural variation without a reference genome, we conducted semi-targeted longread sequencing of the C. maenas genome using MinION sequencing. This dataset includes individual metadata for 6 raw MinION reads, archived at GenBank's SRA under BioProject PRJNA1171011. This sequencing was conducted using crabs from Massachusetts waters.

Table of Contents

Coverage

Location: Buzzards Bay, Massachusetts, USA
Spatial Extent: **N**:41.775408 **E**:-70.589905 **S**:41.30566 **W**:-71.201019
Temporal Extent: 2022-05-18 - 2022-08-08

Methods & Sampling

Samples of Carcinus maenas (urn:lsid:marinespecies.org:taxname:107381) were collected between May 2022 and Aug 2022 from Massachusetts waters.

Extractions were performed with an NEB Monarch HMW DNA Extraction Kit for Cells & Blood (May runs; low success) or a Circulomics Nanobind Tissue Big DNA kit (June and August runs; good success). Library prep was performed with a Oxford Nanopore Cas9 sequencing kit and custom Cas9 probes targeting putative inversion regions. May and June runs used the same first-round set of probes, while August runs used an updated second-round set of probes. Probe sequences are included in the supplemental file and can be cross-referenced using the probe_set value, though targeting is imperfect so much of the data simply reflect non-

targeted genome sequencing. Libraries were sequenced on a single flowcell of an Oxford Nanopore MinION mk1c. For each round of sequencing (May, June, and August), the same library was run twice for 24 hours each time, with a flowcell flush at 24 hours. The run_day value captures this, listing this pre- and post-flush runs as run_days 1 and 2, respectively, for each single sample.

Data Processing Description

Minimal initial processing with standard Oxford Nanopore MinKNOW software (v22.03.06) was used to convert native fast5 files into fastq format for downstream processing. Only the fastq files in the "pass" bin were shared to the SRA. The MinKNOW software includes the following subpackage versions: MinKNOW core v5.0.0, Guppy v6.0.7, Bream v7.0.9, and Script Configuration v5.0.8.

BCO-DMO Processing Description

- Imported "BCODMO_MinION_metadata_NSF-1850996.csv" into BCO-DMO system
- Converted "Collection_date" and "sequencing_date" to ISO 8601 format
- Exported resource as "949666_v1_longread_seq_metadata.csv"

Accepted species identifier confirmed on 2025-01-29.

Problem Description

First round of sequencing (May) resulted in few usable sequences on account of low HMW DNA yield.

Samples were collected in Buzzards Bay, MA; precise sample collection coordinates are not known.

[table of contents | back to top]

Data Files

File
949666_v1_longread_seq_metadata.csv(Comma Separated Values (.csv), 796 bytes) MD5:d3b60996bc3078cfa12587bda10556af Primary data file for dataset ID 949666, version 1

[table of contents | back to top]

Supplemental Files

File
CRISPR-Cas9 crRNA probes to target longread sequencing filename: BCO-DMO_Cas9-probe-seqs_NSF-1850996.csv   (Comma Separated Values (.csv), 2.05 KB) MD5:d4b438e23efff3eec0a3330eb172f2ff Round 1 samples were targeted with probes in round 1, while Round 2 samples (August) were targeted with probes in both rounds 1 and 2. probe_name: unique probe name probe_set: round 1 or 2 target_sequence: DNA sequence to be targeted for CRISPR-Cas9 cutting Cas9_crRNA_probe_sequence: complete CRISPR-Cas9 crRNA sequence to guide cutting

[table of contents | back to top]

Related Publications

MinKNOW Technical Document. Document version: MITD_5000_v1_revAJ_16May2016 (2016). Oxford Nanopore Technologies. https://nanoporetech.com/document/minknow-tech-doc
Methods

[table of contents | back to top]

Related Datasets

References

Woods Hole Oceanographic Institution. Carcinus maenas longread sequencing. 2024/10. In: BioProject [Internet]. Bethesda, MD: National Library of Medicine (US), National Center for Biotechnology Information; 2011-. Available from: http://www.ncbi.nlm.nih.gov/bioproject/PRJNA1171011. NCBI:BioProject: PRJNA1171011.

[table of contents | back to top]

Parameters

Parameter	Description	Units
sample_name	Individual sample ID	unitless
Collection_date	Date of collection	unitless
SRA_accession	SRA accession number for individual sequence files	unitless
biosample_accession	Individual NCBI BioSample code	unitless
embayment	General coastal water body from which sample was collected	unitless
state	US state or Canadian province where samples was collected	unitless
sex	Sex (M=Male, F=Female, or U=Unknown)	unitless
life_stage	Stage of life of sampe (Adult)	unitless
run_ID	Unique identifier for MinION run	unitless
sequencing_date	Date that first sequencing run was started	unitless
run_day	For a single sample, whether the data were collected on day 1 or 2 of sequencing, where a flow-cell flush was performed in between	days
probe_set	Set of Cas9 probes used to target sequencing (see additional file in this dataset)	unitless

Instruments

Dataset-specific Instrument Name	Oxford Nanopore MinION mk1c.
Generic Instrument Name	Nanopore Sequencer
Dataset-specific Description	Sequencing was done with an Oxford Nanopore MinION mk1c.
Generic Instrument Description	A proprietary high-throughput DNA sequencing technology from Oxford Nanopore Technologies that can directly identify and sequence DNA molecules as they pass through nanopores, driven by electrophoresis.

Project Information

Collaborative Research: Tracking fine-scale selection to temperature at the invasion front of a highly dispersive marine predator (West Coast Carcinus)

Coverage: North American west coast: 36 N to 51 N. Emphasis on the Salish Sea

NSF Award Abstract:
Marine invasive species pose a serious and ongoing risk to ocean ecosystems and the economies that rely on them. Understanding how such species adapt rapidly to new environments is key to preventing and managing invasions. Traditionally, the focus has been on inherent traits and flexibility of an invasive species, ignoring the potential for evolutionary change after introduction. However, recent research has shown that some marine species may evolve specific genomic features which allow highly efficient selection over as little as a single generation. This project tests the importance of genomic traits in allowing marine invasive species to survive and thrive on new shores. Its focus is on the high-impact invasive European green crab, which has spread over 1,500 km of the West Coast of North America since 1989 and has very recently begun expanding into the Salish Sea. This project tracks the earliest stages of green crab invasion into a new environment where the species is predicted to have substantial ecological and economic impacts. Genetic differences are followed over time and space across the entire West Coast, with a focus on crabs found in the Salish Sea where the species is currently expanding. Genetic data is complemented by oceanographic modeling to predict the spread of green crabs into the Salish Sea and across the West Coast. Finally, targeted sequencing and prior sampling are used to probe the genomic traits underlying these changes and determine if the same traits have played a role in the species' invasive success on other shores. Sampling for this project is conducted by Washington Sea Grant's Crab Team, an expansive outreach and monitoring program powered largely by hundreds of volunteers who monitor green crabs across 3,000 miles of coastline in the Salish Sea. The results of this project are shared with these volunteers and other stakeholders and is used to inform trans-boundary green crab management and spread prediction on the West Coast.

Recent work has hypothesized that genomic architecture, which has been increasingly discovered to play a role in local adaptation, may also be key to a species' ability to adapt quickly when gene flow is high. This project integrates multiple approaches to track the speed and dynamics of adaptation-with-gene flow across a thermal gradient in an explicit oceanographic context using the invasive European green crab (Carcinus maenas). Prior work in this system identified a suite of genes that appear to constitute balanced polymorphisms whose allele frequencies correlate strongly with site temperature against a homogeneous neutral genetic background. This project has three main goals: 1) To examine fine-scale selection to temperature over a comprehensive spatial and temporal data set comprising most of the species' history on the West Coast, 2) To track the expanding range front in the Salish Sea, comparing the genetic trajectory of individuals at the range edge with oceanographic modeling of dispersal, and 3) To characterize the genomic regions surrounding putative balanced polymorphisms and examine the ubiquity of their association with temperature across globally replicated populations. This coupled evolutionary oceanography approach represents an unprecedented test of the speed and nature of rapid adaptation in a highly dynamic natural marine environment.

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.

Funding

Funding Source	Award
NSF Division of Ocean Sciences (NSF OCE)	OCE-1850996
NSF Division of Ocean Sciences (NSF OCE)	OCE-1850945