

Whole genome sequence data for *Pisaster ochraceus* samples collected from the Pacific coast of North America from July 2004 to May 2018

Website: <https://www.bco-dmo.org/dataset/934772>

Data Type: Other Field Results

Version: 1

Version Date: 2024-08-07

Project

» [Collaborative Proposal: Selection and Genetic Succession in the Intertidal -- Population Genomics of *Pisaster ochraceus* During a Wasting Disease Outbreak and its Aftermath](#) (PoGOMO)

Contributors	Affiliation	Role
Wares, John P.	University of Georgia (UGA)	Principal Investigator
Duffin, Paige Joy	University of Georgia (UGA)	Scientist
Mickle, Audrey	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

Abstract

This dataset includes collection and accession information for whole genome sequence (WGS) data from 65 *Pisaster ochraceus* (ochre sea star) collected across latitudes ranging from SE Alaska to southern California. The sequence data have been deposited into NCBI SRA archive under BioProject PRJNA1117092 and will be publicly available on 2025-08-01. These data are used to evaluate the population genomic diversity and divergence of spatially and environmentally separated populations of *Pisaster ochraceus*.

Table of Contents

- [Coverage](#)
- [Dataset Description](#)
 - [Methods & Sampling](#)
 - [Data Processing Description](#)
 - [BCO-DMO Processing Description](#)
- [Data Files](#)
- [Related Publications](#)
- [Parameters](#)
- [Instruments](#)
- [Project Information](#)
- [Funding](#)

Coverage

Location: Pacific Coast of North America

Spatial Extent: N:57.075 E:-118.86 S:34.034 W:-135.372

Temporal Extent: 2004-07 - 2018-05

Methods & Sampling

Sea stars (*Pisaster ochraceus*) were collected from Santa Barbara, central California, Dabob Bay, Friday Harbor, Sokol Point, and Sitka. In all cases, tube feet were removed from *Pisaster ochraceus* with a razor blade and preserved using 95% undenatured ethanol.

The tube feet from Dabob Bay were preserved in dimethyl sulfoxide (DMSO) buffer per Wares (2000). DNA was subsequently isolated using a Puregene protocol as in Wares (2023) at our lab in Athens, Georgia, USA within the Department of Genetics at the University of Georgia. DNA concentration was quantified using Qubit

fluorometry (Invitrogen, Waltham, MA, USA). DNA processing for Illumina sequencing was performed as follows: DNA samples were prepped for blunt-end ligation using End-It™ DNA End-Repair Kit (Epicentre Biotechnologies) according to the manufacturer's instructions. Next, DNA was cleaned using AMPure beads (Beckman Coulter) to select fragments of 100 bp or higher, and prepped for Illumina sequencing as in Ji et al (2018). Following elution into 43 µl of Tris-HCl, samples were combined with 50 µl A-tailing reaction in NEBNext dA-tailing buffer with Klenow fragment (3'→5' exo-) and incubated at 37 °C for 30 min. Following addition of A-tails, the DNA fragments were ligated to Illumina Truseq adaptors and again purified with AMPure beads. Amplification was achieved in a 50 µl reaction using Phusion™ High-Fidelity polymerase (Thermo Scientific) according to the manufacturer's protocol and the following PCR thermocycler (Bio-RAD T100) specifications: an initial 95 °C for 2 min, then 98 °C for 30 s, followed by 15 cycles of 98 °C for 15 s, 60 °C for 30 s, 72 °C for 4 min and a final 10 min incubation at 72 °C. Primers were removed from the genomic DNA product with a third round of AMPure bead purification.

Samples collected from the remaining five sites (Santa Barbara, central California, Friday Harbor, Sokol Point, and Sitka) were processed as follows: DNA samples were prepped for blunt-end ligation using End-It™ DNA End-Repair Kit (Epicentre Biotechnologies) according to the manufacturer's instructions. Next, DNA was cleaned using AMPure beads (Beckman Coulter) to select fragments of 100 bp or higher, and prepped for Illumina sequencing as in Ji et al (2018). Following elution into 43 µl of Tris-HCl, samples were combined with 50 µl A-tailing reaction in NEBNext dA-tailing buffer with Klenow fragment (3'→5' exo-) and incubated at 37 °C for 30 min. Following addition of A-tails, the DNA fragments were ligated to Illumina Truseq adaptors and again purified with AMPure beads. Amplification was achieved in a 50 µl reaction using Phusion™ High-Fidelity polymerase (Thermo Scientific) according to the manufacturer's protocol and the following PCR thermocycler specifications: an initial 95 °C for 2 min, then 98 °C for 30 s, followed by 15 cycles of 98 °C for 15 s, 60 °C for 30 s, 72 °C for 4 min and a final 10 min incubation at 72 °C. Primers were removed from the genomic DNA product with a third round of AMPure bead purification.

Sequencing of WGS data was performed on Illumina NovaSeq sequencing machines. Some samples were sequenced at the UC Davis Sequencing Core (those from central California), the rest were sequenced at Novogene.

Data Processing Description

First, paired-end reads were trimmed using Trim Galore! (Krueger, 2015) using the paired-read setting (`trim_galore --paired {readA} {readB}`). FastQC reports (Andrews, 2010; module version: FastQC/0.11.9-Java-11) were generated for each trimmed read file (`fastqc {read}`). The *P. ochraceus* reference genome (version 2, DeBiasse et al. publication in prep) was obtained from collaborators and indexed with BWA (Li & Durbin, 2009; module version: BWA/0.7.17-GCC-8.3.0; command: `bwa index p.ochraceus_genome.fna.gz`). Paired reads were mapped to the reference genome using BWA MEM and stored as a .sam file (`bwa mem -t 8 p.ochraceus_genome.fna.gz {readA}.fq.gz {readB}.fq.gz > {sample_ID}.sam`). Next, sam (and later, bam) files were processed through a series of steps recommended in the Genome Analysis Toolkit (GATK) workflow (Broad Institute, DePristo et al., 2011): (1) duplicates were removed using the Picard ("Picard Toolkit," 2019) command `AddOrReplaceReadGroups`; (2) sequences within each file were sorted with `samtools sort` (Li et al., 2009; module version: SAMtools/0.1.19-foss-2019b); (3) paired-end information was verified/corrected with Picard's `FixMateInformation` command and output as bam files; finally, (4) duplicate reads were marked with Picard's `MarkDuplicates` command.

BCO-DMO Processing Description

- Imported original file "BCO-DMO Pisaster WGS.xlsx" into the BCO-DMO system
- Split `lat_lon` column
- Renamed fields to comply with BCO-DMO naming conventions
- Add column with specified dates in YYYY-MM-DD format

Data Files

File
934772_v1_wgs_pisaster_ochraceus.csv (Comma Separated Values (.csv), 18.85 KB) MD5:baa9fe1c3115a0587edce7837a4788a3
Primary data file for dataset ID 934772, version 1

Related Publications

Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at:

<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>

Software

,

Methods

Chandler, V. K., & Wares, J. P. (2017). RNA expression and disease tolerance are associated with a “keystone mutation” in the ochre sea star *Pisaster ochraceus*. *PeerJ*, 5, e3696. Portico.

<https://doi.org/10.7717/peerj.3696>

Methods

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43(5), 491–498. <https://doi.org/10.1038/ng.806>

Methods

,

Software

Harley, C. D. G., Pankey, M. S., Wares, J. P., Grosberg, R. K., & Wonham, M. J. (2006). Color Polymorphism and Genetic Structure in the Sea Star *Pisaster ochraceus*. *The Biological Bulletin*, 211(3), 248–262.

<https://doi.org/10.2307/4134547>

Methods

Ji, L., Jordan, W. T., Shi, X., Hu, L., He, C., & Schmitz, R. J. (2018). TET-mediated epimutagenesis of the *Arabidopsis thaliana* methylome. *Nature Communications*, 9(1). <https://doi.org/10.1038/s41467-018-03289-7>

Methods

Krueger, F. (2015). Trim Galore!: A wrapper around Cutadapt and FastQC to consistently apply adapter and quality trimming to FastQ files, with extra functionality for RRBS data. Babraham bioinformatics - trim galore!

https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/

Software

,

Methods

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform.

Bioinformatics, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>

Methods

,

Software

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., ... Homer, N. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. doi:[10.1093/bioinformatics/btp352](https://doi.org/10.1093/bioinformatics/btp352)

Methods

,

Software

Wares, J. P. (2023). The Genomic Ghosts of *Geukensia granosissima*. *Estuaries and Coasts*, 47(2), 494–503.

<https://doi.org/10.1007/s12237-023-01296-6>

Methods

Wares, J. P., Duke University. Zoology Department, & Duke University. (2000). Abiotic influences on the population dynamics of marine invertebrates

Methods

“Picard Toolkit.” 2019. Broad Institute, GitHub Repository. <https://broadinstitute.github.io/picard/>; Broad Institute <https://github.com/broadinstitute/picard>

Software

[[table of contents](#) | [back to top](#)]

Parameters

Parameter	Description	Units
sample_name	Unique identifier in genomic samples	unitless
bioproject_accession	Accession number for short read archive data maintained at NCBI	unitless
bioproject_ncbi	Name of the NCBI data collection	unitless
assay_type	Assay type, whole genome sequencing	unitless
organism	Host species identity	unitless
isolation_source	Location of organism sample collection and how DNA was isolated	unitless
collection_date	Date organism sample was collected, local	unitless
iso_collection_date	Date organism sample was collected, local, YYYY-mm-dd	unitless
geo_loc_name	NCBI browser form for geographic location of sample	unitless
lat	Latitude of sampling site in degrees North	decimal degrees
lon	Longitude of sampling site in degrees East (negative values are West)	decimal degrees
tissue	Tissue of organism used in genome sample	unitless
biomaterial_provider	Lab that provided the genomic material	unitless
collected_by	Description of collection party	unitless
host_tissue_sampled	Tissue of organism used in genome sample	unitless

Instruments

Dataset-specific Instrument Name	Illumina NovaSeq sequencing machines
Generic Instrument Name	Automated DNA Sequencer
Dataset-specific Description	Sequencing of WGS data was performed on Illumina NovaSeq sequencing machines.
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	Novogene
Generic Instrument Name	Automated DNA Sequencer
Dataset-specific Description	Sequencing of WGS data was performed on Illumina NovaSeq sequencing machines. Some samples were sequenced at the UC Davis Sequencing Core (those from central California), the rest were sequenced at Novogene.
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	Qubit fluorometer
Generic Instrument Name	Fluorometer
Dataset-specific Description	DNA concentration was quantified using Qubit fluorometry (Invitrogen, Waltham, MA, USA).
Generic Instrument Description	A fluorometer or fluorimeter is a device used to measure parameters of fluorescence: its intensity and wavelength distribution of emission spectrum after excitation by a certain spectrum of light. The instrument is designed to measure the amount of stimulated electromagnetic radiation produced by pulses of electromagnetic radiation emitted into a water sample or in situ.

Dataset-specific Instrument Name	PCR thermocycler, Bio-RAD T100
Generic Instrument Name	Thermal Cycler
Dataset-specific Description	Following addition of A-tails, the DNA fragments were ligated to Illumina Truseq adaptors and again purified with AMPure beads. Amplification was achieved in a 50 µl reaction using Phusion™ High-Fidelity polymerase (Thermo Scientific) according to the manufacturer's protocol and the following PCR thermocycler specifications: an initial 95 °C for 2 min, then 98 °C for 30 s, followed by 15 cycles of 98 °C for 15 s, 60 °C for 30 s, 72 °C for 4 min and a final 10 min incubation at 72 °C.
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](#)]

Project Information

Collaborative Proposal: Selection and Genetic Succession in the Intertidal -- Population Genomics of *Pisaster ochraceus* During a Wasting Disease Outbreak and its Aftermath (PoGOMO)

Coverage: Northeastern Pacific (32–60 °N), particularly northern Central California (35–40 °N)

NSF abstract:

This project seeks to understand the outcomes of predator-disease dynamics by exploring a recent pandemic that decimated 90% of ochre sea stars (*Pisaster ochraceus*) in the eastern North Pacific in 2013. The research team will explore how recovery may depend upon often difficult-to-see processes such as the interplay of migration and natural selection in marine species. While the population of sea stars is currently rebounding due to several years of unusually high recruitment, the sea star wasting disease continues to persist at low levels. This project aims to determine the genetic consequences of the pandemic and subsequent recovery. The team will determine whether the majority of susceptible sea stars have died and identify possible refuges where susceptible sea stars survived. They will examine the potential for heritable variation in resistance to this disease in order to assess whether the new recruits are tolerant or susceptible to wasting. Resolving these issues will enable predictions about the trajectory of their recovery and the potential responses to future large scale disease outbreaks. Research findings will be shared with resource managers and scientists at a collaborative workshop that will focus on state-of-the-art methods to advance research on marine diseases. The public will have the opportunity to learn more about sea star wasting disease through a partnership with the UCSC Seymour Marine Discovery Center and can track the incidence of disease using an online interactive map available at www.seastarwasting.org. Results will be incorporated into professional development for teachers with CalTeach and adapted for teaching materials up to college-level. This project will train diverse early career scientists - undergraduates, graduates, and a postdoctoral scholar - in integration of ecological and genomic methods.

Understanding the consequences of large-scale pandemics in the broader contexts of geographic heterogeneity and chronic changes in ocean pH and temperature is an emerging contemporary issue. This project employs long-term characterization of population dynamics and genetic consequences of a sea star wasting disease (SSWD) outbreak, which caused median 90% mortality in *Pisaster ochraceus* populations in the northeastern Pacific, to estimate potential long-term consequences for the species. While the largest

recorded influx of new recruits occurred in 2014-2016, it is unknown where they originated from, whether recruits and surviving adults remained susceptible to the disease, which persisted at low levels, and for how long these dynamics might continue. This long-term dataset provides a unique opportunity for exploring the short and long term repercussions of such large-scale disease outbreaks and the population dynamics that they precipitate. This project builds on long-term field studies of wild populations to describe host population dynamics, the disease, and genomic diversity. The goal is to discover genetic variation associated with SSWD and to dissociate that variation from population genomic effects attributable to abiotic environmental variation. Objectives are: (1) Census *P. ochraceus* at 24 sites throughout its range to describe population dynamics, the prevalence of SSWD, and measure abiotic variables. (2) Conduct laboratory experiments coupled with RNAseq analyses to determine loci differentially regulated during exposure to SSWD, temperature, salinity, and pCO₂ anomalies. (3) Map ddRAD, RNAseq, and candidate loci under selection to a *P. ochraceus* genome. (4) Conduct range-wide population genomic analyses for 3 years to assess genetic (SNP) variation among wild-caught specimens with, versus without, SSWD across a geographic mosaic of abiotic variation. (5) Explore links between SSWD and candidate loci, such as EF1A. These analyses will describe the immediate genomic consequences of the disease outbreak, the population dynamics that the outbreak set in motion, and the interplay of factors and mechanisms - such as disease, temperature, migration, selection - that affected these changes. The results will advance understanding of general processes and interactions that shape population genomic structure in coastal ecosystems, providing resources to inform future research and applications in design of management strategies for coastal living resources.

Proposal abstract

Extreme disturbances are expected to increase in frequency and intensity with climate change; their consequences for marine species will depend upon the often enigmatic interplay of dispersal and selection (and drift). This project seeks to understand the population and genomic consequences of a decimating epizootic of the sea star *Pisaster ochraceus*. Existing collections, which immediately preceded and followed the outbreak and documented >90% mortality of adults and massive subsequent recruitment, will be coupled with continuing annual surveys and population genomic, transcriptomic, and candidate locus analyses. The project aims to determine the extent to which this disease outbreak may (or may not) lead to long-term changes in the frequencies of alleles associated with survival of SSWD.

Understanding the consequences of large perturbations set against a backdrop of geographic heterogeneity and gradual environmental change is an emerging contemporary issue. It requires long-term characterization of population dynamics, genetic consequences, and future implications. In 2013, sea star wasting disease (SSWD) swept through *P. ochraceus* populations in the northeastern Pacific. We captured this epizootic in long-term ecological-genetic studies, which documented median 90% mortality coast-wide (site-specific rates 51-96%). In the aftermath of the initial outbreak, we quantified the largest influx of new recruitments on record. The disease currently persists at low-levels among surviving populations, and recruitment continues to be above average. Given heterogeneity in the environment and in mortality rates, and because 2013 recruits may have been spawned by adults pre-outbreak, but 2014-to-current recruits are progeny of adults that survived, the genomic consequences of the outbreak and the implications for future population and disease dynamics are uncertain.

This project builds on long-term field studies of wild populations of *P. ochraceus* to describe population dynamics, the disease, and genomic diversity. Goals are to discover genetic variation associated with SSWD and to dissociate that variation from population genomic effects attributable to abiotic environmental variation. Objectives are: (1) Census *P. ochraceus* at 24 sites throughout its range to describe population dynamics, the prevalence of SSWD, and measure abiotic variables. (2) Conduct laboratory experiments coupled with RNAseq analyses to determine loci differentially regulated during exposure to SSWD, temperature, salinity, and pCO₂ anomalies. (3) Map ddRAD, RNAseq, and candidate loci under selection to a *P. ochraceus* genome. (4) Conduct range-wide population genomic analyses for 3 years, including intensive study of a focal region, in which we will assess genetic (RAD) variation among wild-caught specimens with versus without SSWD and experiencing the geographic mosaic of abiotic variation. (5) Explore links between SSWD and candidate loci, such as EF1A.

Preliminary results are consistent with an association between SSWD, very high mortality (90%), and differential susceptibility of *P. ochraceus* linked to variation in ddRAD markers, expression of RNAseq loci, and overdominance at a candidate locus (EF1A). RAD analyses show site-specific differences between *P. ochraceus* adults despite high gene flow, and while intertidal juveniles and adults were selected by SSWD in 2013, the subsequent pulse of new recruits was most genetically similar to the pre-outbreak population. The consequences of the SSWD outbreak are still unfurling in a dynamic eco-evolutionary landscape.

Research mentoring

This project trains a postdoc, 3 graduate students, >= 6 undergraduates. The postdoc and graduates will be

cross-trained in field, lab, and genomics. Undergraduates will participate in field and laboratory research, have opportunities for internships, and be involved in outreach activities addressing environmental change. UCM and UCSC are designated Hispanic Serving Institutions. Teaching. Collaboration with the CalTeach program at UC Merced, lab research experience for high school students at Cornell, and an interactive web-based instructional exercise at UCSC will draw upon the iconic natures of *Pisaster*, the rocky intertidal, and keystone predators in education, policy, and management, interweaving project outcomes. Public understanding. Outreach efforts will target general public and marine resource managers using a website and interactive map for tracking sea star wasting (www.seastarwasting.org). Scientific understanding. An end-of-project workshop will host several groups working on different aspects of SSWD and its consequences.

[[table of contents](#) | [back to top](#)]

Funding

Funding Source	Award
NSF Division of Ocean Sciences (NSF OCE)	OCE-1737091

[[table of contents](#) | [back to top](#)]