Bioproject accession information on tag-sequence data for Pisaster ochraceus samples collected from Bodega Bay, CA, in July 2019

Website: <https://www.bco-dmo.org/dataset/934800> **Data Type**: Other Field Results **Version**: 1 **Version Date**: 2024-08-13

Project

» [Collaborative](https://www.bco-dmo.org/project/704716) Proposal: Selection and Genetic Succession in the Intertidal -- Population Genomics of Pisaster ochraceus During a Wasting Disease Outbreak and its Aftermath (PoGOMO)

Abstract

Outbreaks of sea star wasting (SSW) have killed millions of sea stars across over 20 taxa in the last decade alone, threatening the health and stability of coastal communities around the world. While the causative agent remains unknown, it has recently been postulated that hypoxia exposure may play a dominant role in the onset of SSW. We leveraged a study that subjected ochre sea stars to organic matter amendment in a controlled laboratory setting to induce hypoxia and used a repeated sampling design to collect non-invasive tissue samples from both healthy and wasting individuals. Following tag-based RNAseq (TagSeq), we analyzed differential gene expression (DGE) patterns among and within these individuals sampled strategically throughout the 15-day experiment. Transcriptional profiles reveal a progressive change in gene expression accompanying the advancement of SSW, reflecting a transition from asymptomatic stars to the onset of characteristic SSW lesions that progressively worsen until, in some cases, the star dies of their symptoms. Included in this dataset is the accession information for 89 individual TagSeq samples across 20 individual Pisaster ochraceus sea stars at multiple time points during the study to make them available for subsequent re-evaluation. The sequence data have been deposited into the NCBI archive under BioProject PRJNA1116313 and will be publicly available on 2025-08-01.

Table of Contents

- [Coverage](#page-0-0)
- Dataset Description
	- o Methods & [Sampling](#page-0-0)
	- Data Processing [Description](#page-0-0)
	- o **BCO-DMO Processing [Description](#page-0-0)**
	- o Problem [Description](#page-0-0)
- [Data](#page-0-0) Files
- Related [Publications](#page-0-0)
- Related [Datasets](#page-0-0)
- [Parameters](#page-0-0)
- [Instruments](#page-0-0)
- Project [Information](#page-0-0)
- [Funding](#page-0-0)

Coverage

Location: Bodega Bay, CA, USA **Spatial Extent**: **Lat**:38.3181 **Lon**:-123.0171 **Temporal Extent**: 2019-07-01

Methods & Sampling

Specimens of Pisaster ochraceus were collected in July 2019 from the Bodega Marine Lab intertidal area. They were placed into four treatment groups with five replicates per treatment: a control group, with no dissolved organic matter (DOM) enrichment, and three experimental groups amended with different forms of DOM:

- 1. Peptone (organic nitrogen source usable by most marine bacteria; "Peptone")
- 2. Cultured phytoplankton (Dunaliela tertiolecta; "Dunaliela")
- 3. Coastal particulate matter (containing natural phytoplankton, collected from Bodega Bay daily; "Coastal POM")

All groups were placed on flow-through, large volume sea tables for initial acclimation for seven days. Treatment stars were then exposed to assigned conditions for 15 days. Tube feet were collected at the start of the experiment (day 0), every 48 hours thereafter, and upon termination of the experiment (day 15), and preserved in RNAlater™ (ThermoFisher Scientific; cat. AM7020). Observations were recorded daily on the presence/absence of sea star wasting symptoms, and individuals which succumbed to the disease were subsequently removed from the experiment. We strategically subsampled a collection of tissues from the DOM-enrichment experiment published in Aquino et al. (2021).

RNA was isolated from the tube feet samples using a standard lab TRIzol RNA isolation protocol. Tube feet were removed from RNAlater™ and transferred to new 1.5 ml microcentrifuge tubes, followed by tissue lysis with 1 ml TRIzol reagent (Invitrogen, Carlsbad, CA, USA). To prevent overflow, TRIzol was added in two steps: an initial 500 µl was added to each tube and the tissue was manually homogenized with a plastic pestle, followed by another 500 µl TRIzol, after which samples were vortexed and incubated 5 minutes at room temperature (RT) using a Scientific Industries Vortex Genie 2 and a WVR incubator. Next, to separate the RNAcontaining aqueous phase, 200 µl of chloroform was added to the tubes, and the mixture was vortexed for 15 seconds, then incubated for 3 minutes at RT. Following centrifugation at 11.8 rcf for 15 minutes at 4°C, the upper clear, aqueous layer was carefully transferred to a new tube. RNA was precipitated with the addition of 500 µl isopropyl alcohol, then incubated for 10 minutes at RT and centrifuged at 11.8 rcf for 10 minutes at 4°C. After removing the supernatant, pelleted RNA was then washed by adding 1 ml 75% ethanol, followed by a final centrifugation at 7.5 rcf for 5 minutes 4°C. The supernatant was decanted, and the pellet was air-dried for 5 minutes. The RNA product was resuspended in 25 µl H2O and stored at -80°C following quantification with a Qubit® 2.0 Fluorometer and corresponding RNA High Sensitivity (HS) Assay Kit (ThermoFisher Scientific, Waltham, MA, USA). If the concentration exceeded the maximum capacity of the HS kit, quantification was repeated with the Qubit Broad Range (BR) kit.

The isolated RNA was treated with the TURBO DNA-free™ Kit (ThermoFisher Scientific; cat. AM1907) to minimize genomic DNA contamination in our samples using an adapted protocol for routine DNase treatment. First, samples with over 200 ng/µl starting RNA were diluted to meet kit recommendations. Then, 0.1 volume 10X TURBO DNase Buffer and 1 μl TURBO DNase Enzyme were added to the RNA and mixed gently, followed by incubation at 37°C for 30 minutes. Resuspended DNase Inactivation Reagent was then added (2 μl or 0.1 volume, whichever is greater) and the tube was mixed well, then incubated for 5 minutes at RT during which the tube was flicked 2-3 times to redisperse the Inactivation Reagent. Finally, the solution was centrifuged at $10,000 \times g$ for 1.5 minutes, and the supernatant containing the RNA product was transferred to a fresh tube without disturbing the DNase Inactivation Reagent pellet. Samples were again quantified using the same Qubit® procedure and stored at -80°C.

The next step used RNA Clean & Concentrator Kits (Zymo Research, Irvine, CA, USA) to remove inhibitors and other contaminants, including those introduced during the previous step, and to concentrate the RNA product to an ideal volume in preparation for sequencing. Zymo Research offers several different versions of this kit, catered to different ranges of input RNA and desired volume eluent; the Zymo-5 kit (Zymo Research, cat. R1015) has an RNA binding capacity and delivers a more concentrated eluent, while the Zymo-25 kit (Zymo Research, cat. R1018) handles larger amounts of RNA but is limited in its ability to deliver concentrated eluents when input RNA is low. At this stage, 8 samples had undetectable levels of total RNA as measured by Qubit® and the remaining 86 samples had values ranging from 51.51 ng to 13.22 mg. To maximize retention of the isolate in samples with low RNA concentrations while also properly handling samples with high RNA concentrations, the Zymo clean up step into two groups: the first group used the Zymo-5 kit on the 49 samples with less than 600 total ng of RNA and the second used the Zymo-25 kit on the remaining 45 samples with over 600 total ng. Samples were cleaned and concentrated according to the manufacturer's protocols for either kit (Zymo Research, Irvine, CA, USA).

We quantified RNA one final time in preparation for submission. A total of 12 samples had low total RNA recovered (<250 ng) (from lowest to highest: 13-12, 19-15, 2-0, 17-8, 16-10, 4-8, 18-0, 8-8, 4-4, 8-4, 12-8, 5-12) and 5 had levels of RNA so low they were undetectable (4-0, 13-0, 16-4, 16-8, 16-12). We relayed this information to the sequencing facility but ultimately requested they do their best to amplify sequences regardless. We conducted sequencing through the Genomic Sequencing and Analysis Facility at UT Austin, following submission of 94 samples, we received a set of two read files for each of the 89 successfully sequenced samples (179 read files total).

Data Processing Description

Raw reads were trimmed and filtered using the iRNAseq pipeline (Dixon, et al. 2015; Matz, M. V. (2022)), which is intended for use with TagSeq technologies. Briefly, each trimming step was achieved with the following line of code with perl (module version: Perl/5.36.1-GCCcore-12.3.0), cutadapt (Martin, 2011; module version: cutadapt/4.5-GCCcore-11.3.0), and a perl script retrieved from the aforementioned Github repository: perl tagseg_clipper.pl {sampleID}.fastg | cutadapt - -a AAAAAAAA -a AGATCGG -q 15 -m 25 -o {sampleID}.filt.trim. This step removes the first 20 bases, specified adaptor sequences, and 3' end bases of low quality (Phred score < 15), as well as duplicated reads and reads shorter than 25 bases after trimming.

The P. ochraceus reference genome (Ruiz-Ramos et al. 2020; GenBank accession JAAFGO000000000) was used to build the transcriptome for read alignment with STAR (Spliced Transcripts Alignment to a Reference, transcriptome build version: STAR/2.7.10a-GCC-8.3.0) by first converting the publicly available annotation file (Ruiz-Ramos et al. 2020; file name: "augustus.hints.gff3") from '.gff3' to '.gtf' format using the module gffread (version: gffread/0.11.6-GCCcore-8.3.0) and the code gffread -E augustus.hints.gff3 -T -o P.ochraceus annotation.gtf. Next, fasta headers in the reference genome file were modified so that they did not contain spaces. Finally, the STAR-compatible transcriptome files were built using the following code: STAR - runThreadN 16 --genomeSAindexNbases 11 --runMode genomeGenerate --genomeDir /path/to/prepped_genome --genomeFastaFiles P.ochraceus_ref_modified_headers.fa --sjdbGTFfile P.ochraceus_annotation.gtf --sjdbGTFfeatureExon CDS; exit 0.

Trimmed reads were then aligned to the transcriptome using STAR (alignment version: STAR/2.7.10b-GCC-11.3.0), aided by Bowtie2 (module version: Bowtie2/2.5.2-GCC-11.3.0), and the following code: STAR - genomeDir /path/to/prepped_genome --readFilesIn {sampleID}_trimmed.fastq --runThreadN 16 -outFileNamePrefix /path/to/STAR_align_output/{sampleID}_output --outSAMtype BAM SortedByCoordinate -outReadsUnmapped Fastx --outFilterScoreMinOverLread 0.66 --outFilterMatchNminOverLread 0.66 - outFilterMismatchNmax 5 --sjdbGTFfile p.ochraceus genome annotation.gtf --sjdbGTFfeatureExon CDS -quantMode GeneCounts. Read counts per gene were contained in the * ReadsPerGene.out.tab files, which were processed using a custom script to retain counts within the first column of each sample's output file, sum counts across each of the two replicates sequenced for each sample and generate a final gene counts matrix for use in our DGE analyses, below.

DGE Analyses

DGE analyses were achieved using a combination of edgeR (Robinson et al., 2010), PERMANOVA and DESeq2 (Love et al., 2014)-based approaches in R v4.02 (R Core Team, 2020) using RStudio (RStudio Team, 2020). Representative code for core steps using either approach will be made available on the Paige Joy Duffin's Github (Duffin, P. (n.d.)). All analyses employ a FDR/adjusted p-value threshold of 0.05 (or lower, when specified).

BCO-DMO Processing Description

- Imported original file "RNAseq BCO-DMO.xlsx" into the BCO-DMO system
- Remove extraneous columns
- Split lat lon column
- Convert collection date from Jul-2019 to 2019-07-01
- Renamed fields to comply with BCO-DMO naming conventions

Problem Description

A small number of samples failed to generate sufficient reads for analysis, thus we only present data from 89 total RNAseq libraries.

[table of [contents](#page-0-0) | [back](#page-0-0) to top]

Data Files

[table of [contents](#page-0-0) | [back](#page-0-0) to top]

Related Publications

Aquino, C. A., Besemer, R. M., DeRito, C. M., Kocian, J., Porter, I. R., Raimondi, P. T., Rede, J. E., Schiebelhut, L. M., Sparks, J. P., Wares, J. P., & Hewson, I. (2021). Evidence That Microorganisms at the Animal-Water Interface Drive Sea Star Wasting Disease. Frontiers in Microbiology, 11. doi[:10.3389/fmicb.2020.610009](https://doi.org/10.3389/fmicb.2020.610009) **Methods**

Dixon, E. E., Wu, H., Muto, Y., Wilson, P. C., & Humphreys, B. D. (2022). Spatially Resolved Transcriptomic Analysis of Acute Kidney Injury in a Female Murine Model. Journal of the American Society of Nephrology, 33(2), 279–289. doi:10.1681/asn.2021081150 <https://doi.org/10.1681/ASN.2021081150> **Methods**

Dixon, G. B., Davies, S. W., Aglyamova, G. V., Meyer, E., Bay, L. K., & Matz, M. V. (2015). Genomic determinants of coral heat tolerance across latitudes. Science, 348(6242), 1460–1462. doi[:10.1126/science.1261224](https://doi.org/10.1126/science.1261224) Methods

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2012). STAR: ultrafast universal RNA-seq aligner. Bioinformatics, 29(1), 15–21. https://doi.org[/10.1093/bioinformatics/bts635](https://doi.org/10.1093/bioinformatics/bts635) **Software**

, **Methods**

Duffin, P. (n.d.). Paige Duffin's GitHub Profile. GitHub. <https://github.com/paigeduffin> General

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 15(12). doi[:10.1186/s13059-014-0550-8](https://doi.org/10.1186/s13059-014-0550-8) **Methods**

, **Software**

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal, 17(1), 10. doi[:10.14806/ej.17.1.200](https://doi.org/10.14806/ej.17.1.200) **Software**

, **Methods**

Matz, M. V. (2022). z0on/tag-based_RNAseq: Tag-Seq: low-cost alternative to RNAseq (Version v1) [Computer software]. Zenodo. https://doi.org/10.5281/ZENODO.7392165 <https://doi.org/10.5281/zenodo.7392165> **Software**

Pertea, G., & Pertea, M. (2020). GFF Utilities: GffRead and GffCompare. F1000Research, 9, 304. https://doi.org[/10.12688/f1000research.23297.2](https://doi.org/10.12688/f1000research.23297.2)

Software

, **Methods**

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2009). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics, 26(1), 139–140. https://doi.org[/10.1093/bioinformatics/btp616](https://doi.org/10.1093/bioinformatics/btp616) **Software**

, **Methods**

[table of [contents](#page-0-0) | [back](#page-0-0) to top]

Related Datasets

IsRelatedTo

Ruiz-Ramos, D. V., Schiebelhut, L. M., Hoff, K. J., Wares, J. P., & Dawson, M. N. (2020). Data from: An initial comparative genomic autopsy of wasting disease in sea stars (Version 4) [Data set]. Dryad. https://doi.org[/10.6071/M3ND50](https://doi.org/10.6071/M3ND50)

References

Cornell University. Microbial Ecology of Wasting Disease in Pisaster ochraceus. 2020/06. 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/PRJNA637333>. NCBI:BioProject: PRJNA637333.

University of California Conservation Genomics Consortium. Pisaster ochraceus isolate:M0D055189C Genome sequencing and assembly. 2020/02. 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/PRJNA532896. NCBI:BioProject: PRJNA532896. <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA532896>

[table of [contents](#page-0-0) | [back](#page-0-0) to top]

Parameters

[table of [contents](#page-0-0) | [back](#page-0-0) to top]

Instruments

[table of [contents](#page-0-0) | [back](#page-0-0) 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 ºN), particularly northern Central California (35–40 ºN)

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](http://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](#page-0-0) | [back](#page-0-0) to top]

Funding

[table of [contents](#page-0-0) | [back](#page-0-0) to top]