

# Genetic diversity metrics for oyster genetic clusters from an experimental oyster reef restoration in Ninigret Pond, Rhode Island (USA) in 2018-2020

**Website:** <https://www.bco-dmo.org/dataset/1004745>

**Data Type:** Other Field Results

**Version:** 1

**Version Date:** 2026-08-12

## Project

» [CAREER: Linking genetic diversity, population density, and disease prevalence in seagrass and oyster ecosystems](#) (Seagrass and Oyster Ecosystems)

| Contributors                       | Affiliation                                         | Role                      |
|------------------------------------|-----------------------------------------------------|---------------------------|
| <a href="#">Hughes, A. Randall</a> | Northeastern University                             | Principal Investigator    |
| <a href="#">Truskey, Sarit</a>     | Northeastern University                             | Co-Principal Investigator |
| <a href="#">Mickle, Audrey</a>     | Woods Hole Oceanographic Institution (WHOI BCO-DMO) | BCO-DMO Data Manager      |

## Abstract

This dataset contains genetic cluster-level estimates of genetic diversity for eastern oysters (*Crassostrea virginica*, urn:lsid:marinespecies.org:taxname:140657) sampled from each of the 12 experimental restored reefs in Ninigret Pond, Rhode Island, USA in fall 2018 and fall 2020, as part of a multi-year oyster reef restoration experiment using oysters sourced from four commercial hatcheries along the U.S. Atlantic coast. Genetic diversity metrics - observed heterozygosity ( $H_o$ ), expected heterozygosity ( $H_e$ ), allelic richness ( $A_r$ ), and inbreeding coefficient ( $F_{IS}$ ) - were estimated for each genetic cluster represented on each reef at the two sampling time points using SNP genotype data (4679 SNPs). Bootstrapped confidence intervals for  $F_{IS}$  are also provided. Genetic clusters broadly correspond to the four commercial hatchery sources (Maine, Massachusetts, New York, and Virginia) as described in the related genetic assignments dataset.

## Table of Contents

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

## Coverage

**Location:** Ninigret Pond, Charlestown, Rhode Island

**Spatial Extent:** Lat:41.3549 Lon:-71.6929

**Temporal Extent:** 2018-11 - 2020-10

## Methods & Sampling

These data were published in Truskey et al. 2025 (Evolutionary Applications).

This dataset contains genetic cluster-level estimates of genetic diversity for eastern oysters (*Crassostrea virginica*, urn:lsid:marinespecies.org:taxname:140657) sampled from each of the 12 experimental restored reefs in Ninigret Pond, Rhode Island, USA in fall 2018 and fall 2020, as part of a multi-year oyster reef restoration experiment using oysters sourced from four commercial hatcheries along the U.S. Atlantic coast.

We sampled live oysters from each reef on scuba or snorkel. Oysters were put on ice and transported to the Northeastern University Marine Science Center where they were held at  $-80^{\circ}\text{C}$  until DNA extraction.

Raw sequence reads for this study are deposited in the NCBI SRA (BioProject ID PRJNA1280068) and linked in the Related Datasets section. Also see Related Datasets for sequence metadata for all sampled oysters and individual genetic assignments source data.

### Estimating genetic diversity by oyster genetic cluster

We used the primary SNP dataset (4679 SNPs; filtered for  $\text{MAF} > 0.01$  and fall 2020 batch missing data, and pruned for linkage disequilibrium) to estimate genetic diversity metrics within assigned genetic clusters separately for each reef in fall 2018 and 2020. Using the hierfstat R package, we estimated observed ( $H_o$ ) and expected ( $H_e$ ) heterozygosity,  $F_{IS}$  inbreeding coefficients, and allelic richness ( $A_r$ ). Bootstrapped confidence intervals for  $F_{IS}$  were generated using the boot.ppfis function applying 1000 bootstrap replicates.

### BCO-DMO Processing Description

- Loaded table Truskey\_EVA2025\_gdmetrics\_reefxcluster.csv
- Output file as 1004745\_v1\_gdmetrics\_reefxcluster.csv

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

---

### Related Publications

Truskey, S., Sotka, E., Grabowski, J., Kollars-Kjersten, N. M., Lotterhos, K. E., Schneider, E., & Hughes, A. R. (2025). Non-Random Mortality in an Experimental Oyster Restoration. *Evolutionary Applications*, 18(7). Portico. <https://doi.org/10.1111/eva.70128>  
*Results*

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

---

### Related Datasets

#### IsRelatedTo

Hughes, A. R., Truskey, S. (2026) **Genetic relatedness estimates calculated from SNP genotype data of oysters from an experimental oyster reef restoration in Ninigret Pond, Rhode Island (USA) in 2018-2020**. Biological and Chemical Oceanography Data Management Office (BCO-DMO). (Version 1) Version Date 2026-08-05 <http://lod.bco-dmo.org/id/dataset/1003777> [[view at BCO-DMO](#)]

*Relationship Description: Related dataset with pairwise individual genetic relatedness estimates from restored reefs*

Hughes, A. R., Truskey, S. (2026) **Oyster genetic assignment data from an experimental oyster reef restoration in Ninigret Pond, Rhode Island (USA) in 2017-2020**. Biological and Chemical Oceanography Data Management Office (BCO-DMO). (Version 1) Version Date 2026-08-07 <http://lod.bco-dmo.org/id/dataset/1004242> [[view at BCO-DMO](#)]

*Relationship Description: Companion dataset with individual genetic cluster assignments for oysters*

Hughes, A. R., Truskey, S. (2026) **Oyster morphometric, condition, and parasite infection data from an experimental oyster reef restoration in Ninigret Pond, Rhode Island (USA) in 2018-2020**. Biological and Chemical Oceanography Data Management Office (BCO-DMO). (Version 1) Version Date 2026-08-11 <http://lod.bco-dmo.org/id/dataset/1004478> [[view at BCO-DMO](#)]

*Relationship Description: Related dataset with trait measurements for oysters from restored reefs*

Hughes, A. R., Truskey, S. (2026) **Sequence metadata for all sampled oysters from an experimental oyster reef restoration in Ninigret Pond, Rhode Island (USA) in 2017-2020**. Biological and Chemical Oceanography Data Management Office (BCO-DMO). (Version 1) Version Date 2026-08-07 <http://lod.bco-dmo.org/id/dataset/1003894> [[view at BCO-DMO](#)]

*Relationship Description: Sequence metadata for all archived genetic samples associated with this project*

## Software

---

Goudet, J., & Jombart, T. (2004). hierfstat: Estimation and Tests of Hierarchical F-Statistics [Dataset]. In (Editor), CRAN: Contributed Packages. The R Foundation. <https://doi.org/10.32614/cran.package.hierfstat>  
<https://doi.org/10.32614/CRAN.package.hierfstat>

## References

---

Northeastern University. Non-random mortality in an experimental oyster restoration. 2025/06. 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/PRJNA1280068>. NCBI:BioProject: PRJNA1280068.

Truskey, S. B., & Hughes, A. R. (2025). Non-random mortality in an experimental oyster restoration [Data set]. Northeastern University. <https://hdl.handle.net/2047/D20775066>

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

---

## Parameters

| Parameter                  | Description                                                                                                                                                                                                                                                                                                                                       | Units                            |
|----------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------|
| yearcoll                   | Year and season of sample collection (e.g., F2018 = Fall 2018, F2020 = Fall 2020)                                                                                                                                                                                                                                                                 | unitless                         |
| block                      | Experimental block in which a restored reef was situated (A - D)                                                                                                                                                                                                                                                                                  | unitless                         |
| reef                       | Restored reef from which an individual was collected. Letter refers to experimental block and number refers to the reef number within a block (e.g., D3)                                                                                                                                                                                          | unitless                         |
| primary_genetic_assignment | Assigned genetic cluster for an individual using the primary genetic assignment set applied to all main text analyses (DAPC approach, SNP set filtered for MAF > 0.01, LD pruning, and Fall 2020 missing data). Label names correspond to the state of origin of the hatchery source associated with a given genetic cluster (gME, gMA, gNY, gVA) | unitless                         |
| n_inds                     | Number of individuals sampled for this reef x genetic cluster combination                                                                                                                                                                                                                                                                         | count                            |
| Ho                         | Observed heterozygosity: mean proportion of heterozygous loci across individuals. Value: 0 to 1                                                                                                                                                                                                                                                   | unitless (proportion)            |
| He                         | Expected heterozygosity under Hardy-Weinberg equilibrium. Value: 0 to 1                                                                                                                                                                                                                                                                           | unitless (proportion)            |
| Ar                         | Allelic richness: rarefied mean number of alleles per locus, standardized to account for differences in sample size. Derived from rarefaction; Value: 1 to 2                                                                                                                                                                                      | alleles per locus (standardized) |
| FIS                        | Inbreeding coefficient (FIS) estimated using hierfstat. Value: -1 to 1; Negative = excess heterozygosity                                                                                                                                                                                                                                          | unitless                         |
| FIS_ll                     | Lower limit of 95% confidence interval for FIS via bootstrap                                                                                                                                                                                                                                                                                      | unitless                         |
| FIS_hl                     | Upper limit of 95% confidence interval for FIS via bootstrap                                                                                                                                                                                                                                                                                      | unitless                         |

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

## Instruments

|                                         |                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|-----------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Dataset-specific Instrument Name</b> | Illumina HiSeq 2500                                                                                                                                                                                                                                                                                                                                                                                                                          |
| <b>Generic Instrument Name</b>          | Automated DNA Sequencer                                                                                                                                                                                                                                                                                                                                                                                                                      |
| <b>Dataset-specific Description</b>     | See Related Datasets for individual genetic assignments from hatchery samples and restored reefs and sequence metadata for NCBI dataset. The initial and fall 2020 batches were sequenced on a single lane of the Illumina HiSeq 2500 at Tufts University Core Facility Genomics; the fall 2018 batch was sequenced on two lanes of the Illumina NovaSeq 6000 at the University of Texas at Austin Genomic Sequencing and Analysis Facility. |
| <b>Generic Instrument Description</b>   | A DNA sequencer is an instrument that determines the order of deoxynucleotides in deoxyribonucleic acid sequences.                                                                                                                                                                                                                                                                                                                           |

|                                         |                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|-----------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Dataset-specific Instrument Name</b> | Illumina NovaSeq 6000                                                                                                                                                                                                                                                                                                                                                                                                                        |
| <b>Generic Instrument Name</b>          | Automated DNA Sequencer                                                                                                                                                                                                                                                                                                                                                                                                                      |
| <b>Dataset-specific Description</b>     | See Related Datasets for individual genetic assignments from hatchery samples and restored reefs and sequence metadata for NCBI dataset. The initial and fall 2020 batches were sequenced on a single lane of the Illumina HiSeq 2500 at Tufts University Core Facility Genomics; the fall 2018 batch was sequenced on two lanes of the Illumina NovaSeq 6000 at the University of Texas at Austin Genomic Sequencing and Analysis Facility. |
| <b>Generic Instrument Description</b>   | A DNA sequencer is an instrument that determines the order of deoxynucleotides in deoxyribonucleic acid sequences.                                                                                                                                                                                                                                                                                                                           |

|                                         |                                                                                                                                                                                                                                                               |
|-----------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Dataset-specific Instrument Name</b> |                                                                                                                                                                                                                                                               |
| <b>Generic Instrument Name</b>          | Manual Biota Sampler                                                                                                                                                                                                                                          |
| <b>Dataset-specific Description</b>     | We sampled live oysters from each reef on scuba or snorkel. Oysters were put on ice and transported to the Northeastern University Marine Science Center where they were held at $-80^{\circ}\text{C}$ until DNA extraction.                                  |
| <b>Generic Instrument Description</b>   | "Manual Biota Sampler" indicates that a sample was collected in situ by a person, possibly using a hand-held collection device such as a jar, a net, or their hands. This term could also refer to a simple tool like a hammer, saw, or other hand-held tool. |

|                                         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|-----------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Dataset-specific Instrument Name</b> | scuba                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| <b>Generic Instrument Name</b>          | Self-Contained Underwater Breathing Apparatus                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| <b>Dataset-specific Description</b>     | We sampled live oysters from each reef on scuba or snorkel. Oysters were put on ice and transported to the Northeastern University Marine Science Center where they were held at $-80^{\circ}\text{C}$ until DNA extraction.                                                                                                                                                                                                                                                                                                                                                      |
| <b>Generic Instrument Description</b>   | The self-contained underwater breathing apparatus or scuba diving system is the result of technological developments and innovations that began almost 300 years ago. Scuba diving is the most extensively used system for breathing underwater by recreational divers throughout the world and in various forms is also widely used to perform underwater work for military, scientific, and commercial purposes. Reference: <a href="https://oceanexplorer.noaa.gov/technology/technical/technical.html">https://oceanexplorer.noaa.gov/technology/technical/technical.html</a> |

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

## Project Information

### **CAREER: Linking genetic diversity, population density, and disease prevalence in seagrass and oyster ecosystems (Seagrass and Oyster Ecosystems)**

**Coverage:** Coastal New England

#### *NSF Award Abstract:*

Disease outbreaks in the ocean are increasing, causing losses of ecologically important marine species, but the factors contributing to these outbreaks are not well understood. This 5-year CAREER project will study disease prevalence and intensity in two marine foundation species - the seagrass *Zostera marina* and the Eastern oyster *Crassostrea virginica*. More specifically, host-disease relationships will be explored to understand how genetic diversity and population density of the host species impacts disease transmission and risk. This work will pair large-scale experimental restorations and smaller-scale field experiments to examine disease-host relationships across multiple spatial scales. Comparisons of patterns and mechanisms across the two coastal systems will provide an important first step towards identifying generalities in the diversity-density-disease relationship. To enhance the broader impacts and utility of this work, the experiments will be conducted in collaboration with restoration practitioners and guided by knowledge ascertained from key stakeholder groups. The project will support the development of an early career female researcher and multiple graduate and undergraduate students. Students will be trained in state-of-the-art molecular techniques to quantify oyster and seagrass parasites. Key findings from the surveys and experimental work will be incorporated into undergraduate courses focused on Conservation Biology, Marine Biology, and Disease Ecology. Finally, students in these courses will help develop social-ecological surveys and mutual learning games to stimulate knowledge transfer with stakeholders through a series of workshops.

The relationship between host genetic diversity and disease dynamics is complex. In some cases, known as a dilution effect, diversity reduces disease transmission and risk. However, the opposite relationship, known as the amplification effect, can also occur when diversity increases the risk of infection. Even if diversity directly reduces disease risk, simultaneous positive effects of diversity on host density could lead to amplification by increasing disease transmission between infected and uninfected individuals. Large-scale field restorations of seagrasses (*Zostera marina*) and oysters (*Crassostrea virginica*) will be utilized to test the effects of host genetic diversity on host population density and disease prevalence/intensity. Additional field experiments independently manipulating host genetic diversity and density will examine the mechanisms leading to dilution or amplification. Conducting similar manipulations in two marine foundation species - one a clonal plant and the other a non-clonal animal - will help identify commonalities in the diversity-density-disease relationship. Further, collaborations among project scientists, students, and stakeholders will enhance interdisciplinary training and help facilitate the exchange of information to improve management and restoration efforts. As part of these efforts, targeted surveys will be used to document the perceptions and attitudes of managers and restoration

practitioners regarding genetic diversity and its role in ecological resilience and restoration.

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

---

## Funding

| Funding Source                                           | Award                       |
|----------------------------------------------------------|-----------------------------|
| <a href="#">NSF Division of Ocean Sciences (NSF OCE)</a> | <a href="#">OCE-1652320</a> |

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