Data Management Plan

Data Policy Compliance

This project will comply with the NSF OCE Data and Sample Policy. We will also follow policies and recommendations of the Environmental Data Initiative (environmentaldatainitiative.org/).

Description of Data Types

- Simulations: This research will generate a large number of simulations. Outputs of the simulations including tree sequencing files and VCF files.
- Machine learning outputs: Machine learning outputs will be generated for the simulated and empirical data. Data types include importance values of individual SNPs and a machine learning prediction of the multivariate environment.
- Genomics: This project will generate ~ 2TB of SNP data from the 200K Axiom SNP array and 1K AgriSeq GBS panel.
- Phenotyping: This research will generate phenotypic data for a few thousand juvenile oysters.
- Environmental data: This research will compile available environmental data for oyster habitats and will also collect environmental data during the experiment (temperature salinity, and disease).
- Genetics samples: Genetic samples (Preserved tissues and DNA extractions) will be deposited into a -80^{o}C freezer at NU and will be made available to researchers upon request.

Data and Metadata Formats and Standards

Metadata associated with this proposed research, including information on sites, experiments, and data collected (e.g., date, time, location, experimental treatments and maintenance, and environmental variables measured) will be documented for all data.

Data Storage and Access During the Project

All data will be stored permanently and backed up on Discovery Cluster at Northeastern University.

Mechanisms and Policies for Access, Sharing, Re-Use, and Re-Distribution

Data will be released once the results are published or no later than one year after the project end date. The PI is dedicated to conducting tractible and reproducible research. Custom scripts and pipelines will be developed collaboratively on GitHub, thus allowing transparency during the code development phase as well as access of completed code to other researchers.

Plans for Archiving

For each publication, raw data and reproducible pipelines (including scripts used to create publication figures) will be compressed and archived permanently on a public repository such as Dryad.

Raw data will also be contributed in tab-delimited format to the BCO-DMO system.

Roles and Responsibilities

Lotterhos will take the lead on ensuring that all project personnel comply with the data management plan.