Mortality and DNA metrics from minimally-lethal genotyping with eastern oysters (Crossosrea virginica) conducted in 2022 and 2023

Website: https://www.bco-dmo.org/dataset/1000264
Data Type: experimental
Version: 1
Version Date: 2026-06-04

Project
» CAREER: Evaluation of machine learning algorithms for understanding and predicting adaptation to multivariate environments with a Model Validation Program (MVP) (Model Validation Program)

Contributors	Affiliation	Role
Lotterhos, Katie	Northeastern University	Principal Investigator
Small, Jessica	Virginia Institute of Marine Science (VIMS)	Scientist
Leung, Elisabeth	Northeastern University	Student, Contact
Soenen, Karen	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

Abstract
Understanding how selection acts on individual genotypes often requires genotyping at different life stages and tracking their survival under experimental conditions. However, in mollusks, such as the Eastern oyster (Crassostrea virginica), collecting sufficient high-quality DNA without causing mortality remains challenging. While prior studies have explored different noninvasive genotyping methods, an optimal method hasn't been identified. Here, we tested different techniques by combining two minimally invasive accession techniques (relaxation, shell notching), four types of cell sampling (swab, mantle biopsy, hemolymph, and extrapallial fluid), and three preservation techniques (flash freezing, ethanol, and FTA card) using nine different treatment groups. We monitored mortality for 11 days after cell sampling and quantified the effects of each treatment on mortality rates, DNA quality, and DNA quantity. This dataset includes mortality results from the different treatments and DNA quality and quantification from the extracted samples.

Table of Contents

Coverage

Location: Virginia Institute of Marine Science (part of William & Mary) in Gloucester Point, Virginia, USA

Methods & Sampling

Juvenile oysters ranging from 30mm to 60mm were sourced from the Aquaculture Genetics and Breeding Technology Center (ABC) at the Virginia Institute of Marine Science (VIMS). Each oyster was cleaned using a 1% bleach solution, air-dried for an hour, and labeled with a shellfish tag (Hallprint FPN 8x4 mm) using coral epoxy (Instant Ocean HoldFast® Epoxy Stick) (Figure 1A). Shell width and length were measured for each oyster before the experiment (Kraeuter et al., 2007). To store the oysters before the experiment, the oysters were stored at 5°C for up to 48 hours before experimentation. After each trial, the oysters were transferred to sea tables (flow-through tanks filled with minimally filtered seawater from the adjacent York River), and

mortality was monitored daily for 11 days. Daily mortality checks were conducted where each oyster was removed from the sea table briefly to check for shell gapes, which indicated loss of adductor muscle function and death.

Treatment groups

We planned nine treatments to determine which protocol would have the best DNA yield with the lowest mortality (Figure 1 of Leung et al. 2026 (in press)). Each treatment had at least 20 individual oysters. Groups of 10 oysters per each treatment were placed into separate trays in the sea tables. The treatments and short names are described in Table 1 of Leung et al. 2026 (in press).

Control Groups: We established two control groups in our experiment: one to monitor mortality and the other to quantify the amount of DNA extracted.

Mortality Control Groups: We had two groups to control for mortality. *Mortality_Control_1:* At the beginning of the experiment, we labeled 30 individuals using coral epoxy and placed them into the sea tables. *Mortality_Control_2:* At the end of our trials, we added another 30 individuals to the sea tables that were stored in the refrigerator for two days. This was to monitor any effects of being in the refrigerator, controlling for the longest time of oyster refrigeration. *DNA_Control:* To establish a baseline for DNA quantification and quality, we sampled mantle, adductor, and gill tissues from 30 individuals and placed them into 2 mL O-ring vials with 95% molecular-grade ethanol. We stored the samples temporarily in a -20℃ freezer before transferring them to a -80℃ freezer.

Accessing the oyster tissue within the shell

Two methods were used to enter the oyster shell: creating a notch in the shell or relaxing the oyster. Creating a notch allowed for the extraction of fluid that contains genetic material (e.g., hemolymph or EPF), while relaxing the oyster allowed for tissue swabbing or direct sampling of the tissue. *Notch Protocol*: A variable-speed dremel was used to make a small, V-shaped notch in the oyster shell, just large enough for the needle (BD 23G PrecisionGlide Needle) to fit in at different angles. The size of the notch depended on how much new shell an oyster had recently produced, but was typically a few mm in size. The notch was created on the right valve of the oyster, approximately two-thirds to the top of the shell. The needle, attached to a syringe (BD 1mL Syringe), was used to extract fluid as described below for the *Hemolymph Protocol* and the *EPF Protocol*. *Relaxation Protocol*: For the relaxation protocol, 2 L of 1µm filtered seawater was mixed with 100g of Epsom salt until all the salt dissolved following protocol by Proestou et al. (2025). A maximum of ten oysters were placed in each replicate 2L container for 16 hours before the cell sampling (see *Treatment Groups* for experimental design). Oysters were fully submerged in the Epsom salt solution and held at room temperature. During relaxation, containers were aerated using air stones and airlines to maintain oxygen levels.

Collection of genetic material

Cell collection for notched oysters: Cells were collected from notched oysters by sampling hemolymph or EPF fluids. For the preservation of fluids, we split the samples and placed half in liquid nitrogen and the other half in FTA cards (see below). *Hemolymph Protocol*: We inserted the needle through the notch diagonally down into the adductor muscle, which we identified through increased resistance. We collected 125 to 250 µL of hemolymph. *EPF Protocol*: We inserted the needle into the notch until it reached the mantle cavity of the shell and collected 80 to 500 µL of EPF from each oyster. If necessary, we removed and reinserted the needle at multiple angles for small oysters to obtain a minimum fluid amount of 80 µL.

Cell collection for relaxed oysters: Cells were collected from relaxed oysters by taking a mantle biopsy or swabbing the mantle tissue. *Mantle Biopsy Protocol*: We used forceps and scissors to remove a small piece (~10-20mg) of the mantle wrapped around the adductor muscle. We sampled the distal edge of the mantle tissue, approximately 2-3mm away from the visible edge of the adductor muscle.

After each sample was acquired, we sterilized the forceps and scissors using an ethanol lamp. *Swab Protocol*: Using an OmniSwap (Qiagen; WB100035), we swabbed in a circular motion, starting at the top layer of the mantle and moving down to the mantle around the adductor, then to the mantle's bottom layer, and then reversing this path. The motion was repeated five times for each sampled individual.

Preservation of genetic material

We used three preservation protocols for the genetic samples: ethanol, flash-freezing, and FTA cards. *Ethanol Protocol*: The mantle sampled from the relaxed oysters was preserved with 95% molecular-grade ethanol. It was temporarily stored in a -20℃ freezer and then transferred to a -80℃ freezer. *Flash-Freeze Protocol*: Liquid nitrogen was used to flash-freeze half of each EPF or hemolymph sample from the notched oysters. After the

samples were flash-frozen, they were stored in a -80°C freezer. Additionally, the swabs used on the relaxed oysters were flash-frozen and then stored in a -80°C freezer. *FTA Cards Protocol*: The remaining half of the EPF and hemolymph samples were stored on QIAcard FTA Classic Cards (Qiagen; WB120305). Following the manufacturer's instructions, we put up to 125 µL of fluid on each circle in the FTA card. FTA cards were stored in WhirlPak bags with color-indicating silica beads (desiccants that absorb moisture and change color to indicate saturation) following the manufacturer's recommendations.

DNA extraction

A Qiagen DNeasy Blood & Tissue kit (Qiagen; 69504) was used for all DNA extractions, but with slight modifications for different cell types or preservation methods. The manufacturer's instructions were used for the control and the mantle biopsy groups with 20- 30mg of tissue, as well as the swabs (swabs were horizontally cut in half before extraction so that they could be submerged fully in the buffer). We modified the protocol for the EPF and Hemolymph samples by using 200 µL of the flash-frozen sample as input (based on preliminary trials of 20, 100, and 200 µL that showed the latter amount yielded the most DNA) but otherwise followed the manufacturer's instructions. For the Whatman FTA Cards, we used a 6.0mm biopsy punch (Cenmed™ PMD Sterile, Disposable Biopsy Punches) to cut a disk from the middle of the card. Per recommendation from Sigma Aldrich, 280 microliters of the ATL buffer was added to the disk for extraction, and the remainder of the manufacturer's instructions were followed (*Reliable DNA Extraction from WhatmanTM FTA Cards*, n.d.).

DNA quantification and quality assessment

We used a Qubit 3.0 Fluorometer to quantify DNA yield from the extractions. We used the High Sensitivity Kit (ThermoFisher; Q32851) for samples with a yield between 0 and 20 ng/ul and the Broad Range Kit (ThermoFisher; Q32850) for all samples with a yield higher than 20 ng/ul.

To test the quality of the DNA, we ran 3 µL of each DNA extract on a 1% agarose gel (0.6g of agarose with 60 mL of 1x TAE buffer), stained with GelRed at 130V for 30 minutes. Gels were visualized, and images were captured on a BioRad Gel Doc XR+ Molecular Imager machine. We categorized each sample on the gel into one of four levels: a bright band greater than 3000 bp with a slight smear (level 1), a completely smeared sample (level 2), a barely visible gel (level 3), or no band (level 4).

Statistical Analyses

Oyster size distribution across treatments

Although we randomly assigned oysters to treatments, we conducted a one-way ANOVA of oyster length and treatment to assess potential sampling bias. We used the "aov" function in R to perform an analysis of variance with the treatment groups as the explanatory variable and the length of the oysters as the response variable, and a Tukey HSD test to examine pairwise differences among treatments ("TukeyHSD" in R).

Comparison of Survival between Control Groups

First, we tested the null hypothesis that the probability of survival was equal between the two control groups (*Mortality_Control_1* and *Mortality_Control_2*) with a binomial test (function `binom.test()` in R with the number of successes set to the number of surviving oysters after 11 days and the sample size the total number of oysters in each control group). The alternate hypothesis was that the survival of *Mortality_Control_1* and *Mortality_Control_2* was not equal.

Survival through time

To ensure the length of the experiment was sufficient, we conducted a survival analysis. Survival data were converted into a numerical binary variable, with 1 representing alive and 0 representing dead after 11 days. Using all samples, we created a survival object with the binary survival data and the lifespan of each oyster, with no grouping term, using the `survfit` function from the `survival` package in R. The survival object was plotted in a Kaplan-Meier survival curve.

Effect of treatment and size on survival

To test the effects of treatment and oyster size on mortality, we created a logistic survival model with survival (binary) as the response variable, treatment as an explanatory factor, and length as an explanatory covariate. We implemented the model using the `glm` function in R and specified family = binomial(link = "logit").

DNA Quantity Analysis

To test if there was a statistical difference in the total amount of DNA among treatment groups (DNA_Control, Relax_Swab_Flash, Relax_Mantle_EtOH, Notch_Hemo_Flash, Notch_EPF_Flash), we conducted an ANOVA test using the `aov` function in R. The explanatory variable was the different treatment groups, and the response variable was the DNA concentration. The null hypothesis was that the treatment groups had no difference in DNA concentration. After a significant ANOVA result, we conducted post-hoc comparisons using Tukey's HSD test to identify which treatment groups differed significantly in DNA concentration while controlling for multiple comparisons.

DNA Quality Analysis

To address the question of which method had the highest DNA quality, we scored the gels on a range from 1 to 4, with 1 being the highest quality. All scoring was done by the same person for consistency and to minimize variability. For statistical analysis, we performed a chi-square test to test the null hypothesis of no association (i.e., independence) between the gel categorization scores and treatment groups using the `chisq.test()` in R. The alternative hypothesis was that the different treatment groups and the gel categorization scores were not independent.

[table of contents | back to top]

Related Publications

KRAEUTER, J. N., FORD, S., & CUMMINGS, M. (2007). OYSTER GROWTH ANALYSIS: A COMPARISON OF METHODS. Journal of Shellfish Research, 26(2), 479–491. https://doi.org/10.2983/0730-8000(2007)26[479:ogaaco]2.0.co;2 https://doi.org/10.2983/0730-8000(2007)26[479:OGAACO]2.0.CO;2
Methods

Leung, E., Small, J., & Lotterhos, K. (in press). A comparison of non-lethal genotyping methods for juvenile Eastern oysters. Ecology and Evolution.
Results

Proestou, D. A., Delomas, T. A., & Small, J. M. (2025). Genetic parameters for dermo challenge survival in a Virginia, U.S. eastern oyster breeding population. Aquaculture, 609, 742816. https://doi.org/10.1016/j.aquaculture.2025.742816
Methods

elisabethleung. (2026). *elisabethleung/non_lethal_oysters: 02/15/2026* (Version oysters) [Computer software]. Zenodo. https://doi.org/10.5281/ZENODO.18654417 https://doi.org/10.5281/zenodo.18654417
Software

[table of contents | back to top]

Parameters

Parameters for this dataset have not yet been identified

[table of contents | back to top]

Instruments

Dataset-specific Instrument Name	Qubit & Qiagen
Generic Instrument Name	DNA Analysis Kit
Dataset-specific Description	Qiagen DNeasy Blood & Tissue kit (Qiagen; 69504) Qubit High Sensitivity Kit (ThermoFisher; Q32851) Qubit Broad Range Kit (ThermoFisher; Q32850)
Generic Instrument Description	A laboratory kit containing reagents and materials used to extract, purify, detect, measure, or otherwise evaluate DNA from a sample.

[table of contents | back to top]

Project Information

CAREER: Evaluation of machine learning algorithms for understanding and predicting adaptation to multivariate environments with a Model Validation Program (MVP) (Model Validation Program)

Coverage: East coast of North America

NSF Award Abstract:
Environmental change can be rapid and involve multiple aspects of the environment changing at the same time, such as warming and increased disease pressure. Rapid environmental change threatens the productivity of aquaculture and crops on which humans depend. Predicting organisms' vulnerabilities to rapid and multifactor environmental change, however, is a major scientific challenge. A hurdle to addressing this challenge arises from the complex and non-intuitive ways that organisms adapt, through changes at the level of the DNA sequence, to many environmental stresses at the same time. Thus, there is a need for new approaches to understand and predict adaptation in multivariate environments. To address this need, this project integrates research and education with a Model Validation Program (MVP). The research is developing and evaluating Machine Learning Algorithms (MLAs) for understanding and predicting adaptation of organisms to multivariate environments from their DNA sequences. To evaluate MLAs, this research combines both data simulation and an empirical test in the field with the Eastern Oyster, which provide important ecosystem services and support a multi-million dollar industry. For oysters, this research is studying how temperature, disease pressure, and salinity interact with evolutionary history to determine fitness in the field. This research advances efforts toward addressing the major scientific challenge of predicting adaptation in complex environments by integrating concepts across the frontiers of marine, evolutionary, and statistical sciences in a new way. Machine learning and model validation are not traditionally taught in the marine and environmental sciences, but are becoming increasingly relevant to these fields. As part of a broader education program, this research is developing MVP Learning Modules for high school students and undergraduates, which help students build the foundational knowledge they need to critically evaluate and apply models. Modules are being disseminated to hundreds of students in the greater Boston area and are being made available online for widespread use. The MVP mentoring program is training graduate students, undergraduates, and high school students in marine evolutionary ecology, statistical genomics, and machine learning. This research addresses a pressing societal need to more informatively match genotypes to environments for restoration, farming, and assisted gene flow efforts. Results are being disseminated to stakeholders in the oyster industry.

The goal of this research is to evaluate if MLAs, which can model non-linearities, can be used to understand and predict adaptation to multivariate environments under a wide range of scenarios. In Objective 1, the Principal Investigator (PI) is creating simulated datasets with different aspects of realism, and using them to evaluate and refine the MLAs. This novel set of simulations is studying genome evolution under high gene flow in complex, multivariate environments. In Objective 2, the PI is building on their expertise with the Eastern oyster to evaluate the MLAs in a field setting. The PI is first developing a comprehensive seascape genomic dataset and using it to train MLAs to predict an individual's multivariate environment based on a single nucleotide polymorphism genotype. Then, the PI is testing if the MLA prediction can predict the fitness of different genotypes from across the species range when raised in common garden field conditions. In Objective 3, the PI is integrating research and education by using the data obtained from Objs. 1 and 2 to develop a series of original "MVP Learning Modules" with interactive web apps for persons at different levels of understanding, using the relatable example of an oyster restoration project. This research lays the foundation for future studies by producing datasets that could become classical examples for developing and benchmarking innovative modeling approaches.

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.

[table of contents | back to top]

Funding

Funding Source	Award
NSF Division of Ocean Sciences (NSF OCE)	OCE-2043905

[table of contents | back to top]