Population Genomic Calls for green ash, Fraxinus pennsylvanica, collected in eastern North America 2015

Website: https://www.bco-dmo.org/dataset/1005382
Data Type: Other Field Results
Version: 1
Version Date: 2026-08-17

Project
» The genetic legacy of an Asian oyster introduction and its disease-causing parasite (Oyster historical genetics)

Contributors	Affiliation	Role
Strand, Allan	College of Charleston (CofC)	Principal Investigator
York, Amber D.	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

Abstract
This dataset contains field sampling metadata (individual tree IDs, population names, and sampling coordinates) and a corresponding SNP genotype matrix (361 individuals x 1,306 loci) for Fraxinus pennsylvanica (green ash) sampled from 26 localities across eastern North America in 2015. These data accompany the results publication Castilla et al. (2023, doi: 10.1111/jbi.14754) and the archived HoloSimCell software package Strand et al. (2026, doi: 10.5281/ZENODO.18673622). See the below study description for larger context of how these data were used. Study description: Biogeographers have used three primary data types to examine shifts in tree ranges in response to past climate change: fossil pollen, genetic data and contemporary occurrences. Although recent efforts have explored formal integration of these types of data, we have limited understanding of how integration affects estimates of range shift rates and their uncertainty. We compared estimates of biotic velocity (i.e. rate of species; range shifts) using each data type independently to estimates obtained using integrated models, focused on Eastern North America populations of Fraxinus pennsylvanica Marshall (green ash). Using fossil pollen, genomic data and modern occurrence data, we estimated biotic velocities directly from 24 species distribution models (SDMs) and 200 pollen surfaces created with a novel Bayesian spatio-temporal model. We compared biotic velocity from these analyses to estimates based on coupled demographic-coalescent simulations and Approximate Bayesian Computation that combined fossil pollen and SDMs with population genomic data collected across the F. pennsylvanica range.

Table of Contents

Coverage

Location: Eastern half of North America
Spatial Extent: **N**:50.56 **E**:-66.88 **S**:29.17 **W**:-110.75
Temporal Extent: 2015

Methods & Sampling

Fossil pollen records from eastern North America were downloaded from the Neotoma Paleoecology Database.

There were 465 pollen records within our study region with at least one pollen count sample in the considered time period (LGM to present). Each pollen record includes pollen taxon counts at a set of depths. Most records include reliable sample age estimates derived from an age-depth model. Occurrence data were downloaded from the Botantical Information and ecology network (BIEN version 4.1.1). Using specimen and plot data for *F. pennsylvanica* from across its native range, we calibrated 24 different species distribution models (SDM) using present-day climate reconstructions from two paleoclimate models (CCSM, the Community Climate System Model, and ECBilt).

We collected leaf or twig samples from adult trees (> 10 cm DBH) in 48 naturally regenerated sites. Leaf DNA was extracted using a modified CTAB protocol. For each individual tree sampled, 100ng of genomic DNA was cut using two restriction enzymes, EcoRI and MseI. Subsequently, adaptors were ligated on each end of fragments that resulted from restriction enzyme digestion. Adaptors complementary to the EcoRI cut site also included a 8-base sequence unique to each individual included in the library to serve as a barcode to identify sequences from that individual during bioinformatic processing. Two rounds of PCR followed the ligations in parallel (each reaction started with the same product from ligation reactions). Finally all reaction products were pooled. This pooled library was size-selected to include fragments from 300-500bp in size and then sequenced on two lanes of the Illumina HiSeq 2000 platform.

Organism identifiers:
green ash, Fraxinus pennsylvanica, Fraxinus pennsylvanica Marshall, urn:lsid:dyntaxa.se:Taxon:221644, https://www.checklistbank.org/dataset/315834/taxon/6JMHT

Data Processing Description

DNA sequence reads were demultiplexed, assembled against the F. pennsylvanica reference genome (PE_00248), and reads mapping to more than one genomic location were removed to reduce paralogy. Genotype likelihoods were estimated with ANGSD, retaining SNPs with range-wide minor allele frequency >= 0.01 and Phred-scaled base quality >= 20. Explicit genotype calls required the most likely genotype to be at least 15 times more likely than the next-best genotype; otherwise the genotype was coded as missing. For the ABC analyses reported in the paper, SNPs and individuals were further filtered to support balanced, no-missing-data summary statistics, with the final analysis set reported as 294 individuals from 21 populations genotyped at 1,000 SNP loci. This BCO-DMO upload contains the broader genotype matrix exported from the HoloSimCell data package: 361 individuals from 26 named sampling localities and 1,306 SNP rows.

Encoding:
Genotypes are diploid nucleotide calls (AA, AC, ..., TT); no missing values are present in the provided matrix in this dataset.

BCO-DMO Processing Description

No modifications made to the data files as provided to BCO-DMO. The main data file (holoSimCell_genotypes.csv) contains a matrix and thus was not imported into the BCO-DMO table data servers. Data were not transformed to tabular (unpivot operation) due to emphasis on data usage with related tools as a matrix not a table.
- Added holoSimCell_column_definitions.csv as supplemental file with "parameter" column descriptions in the file comment. At the time this dataset was drafted the dedicated "Parameter" section only supports column information for primary "main" data tables in our system.
- The original holoSimCell_README.txt was also added to the supplemental files section. The readme file contains similar but slightly differently worded metadata from what is included in the dataset Methods & Sampling and Data Processing sections.
* The related package was cited with 10.5281/zenodo.18673622 which is the umbrella DOI for all versions of the zenodo package (currently there is only one (Version v0.1.4)). 10.5281/zenodo.18673623 is the doi for the exact version (Version v0.1.4). The exact version was cited in the related publications section.

Problem Description

NA

[table of contents | back to top]

Related Publications

Castilla, A. R., Brown, A., Hoban, S., Abhainn, E. A., Robinson, J. D., Romero-Severson, J., Smith, A. B., Strand, A. E., Tipton, J. R., & Dawson, A. (2023). Integrative demographic modelling reduces uncertainty in estimated rates of species' historical range shifts. Journal of Biogeography, 51(2), 325–336. Portico. https://doi.org/10.1111/jbi.14754
Results

Strand, A., Robinson, J., Smith, A. B. (2026). stranda/holoSimCell: documented version for release (Version v0.1.4) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.18673623 https://doi.org/https://doi.org/10.5281/ZENODO.18673622
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	Illumina sequencing machine (Illumina HiSeq platform)
Generic Instrument Name	Automated DNA Sequencer
Generic Instrument Description	A DNA sequencer is an instrument that determines the order of deoxynucleotides in deoxyribonucleic acid sequences.

Dataset-specific Instrument Name	Thermocycler (BioRad S1000)
Generic Instrument Name	Thermal Cycler
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

The genetic legacy of an Asian oyster introduction and its disease-causing parasite (Oyster

historical genetics)

Coverage: Global

NSF abstract:

During the 20th century, the Pacific oyster *Crassostrea gigas* was deliberately introduced from its native range of coastal Asia to the estuaries of six continents. While the introduced Pacific oysters are widely aquacultured and thus can generate local economic wealth, they sometimes outcompete native oysters, and can carry microbial, animal and plant hitchhikers that negatively impact local economies and the ecological functioning of local estuaries. This study comprehensively assesses the pathways and sources of Pacific oyster introductions using a worldwide, population genetic survey. Simultaneously, the study also assesses the pathways and source of one hitchhiking protist (*Haplosporidium nelsoni*) that causes the disease MSX (multinucleated sphere X) in the Virginia oyster (*Crassostrea virginica*) along the eastern seaboard of the United States. One goal of this research is to generate management strategies that combat the negative impacts of the Pacific oyster and its associated invaders, and minimize future invasions. A second goal is to minimize some uncertainty about the population biology of the devastating Haplosporidium parasite, and thus, increase confidence of policy makers who are managing shellfish health, restoration and commerce. By quantifying the pathways and sources of *C. gigas*, this project may inform strategies to combat negative impacts of *C. gigas* and its associated invaders, as well as minimize future invasions. Moreover, quantifying dispersal within and among populations of *H. nelsoni* along the US East Coast will provide perspective on the effectiveness of regional biosecurity measures in preventing the ongoing dispersal of this destructive pathogen via aquaculture. In addition, the project lends itself well to programs that foster critical thinking and research experience among both undergraduate and K-12 students. The project provides opportunities for 6-9 undergraduates to perform research, includes a 2-day workshop on bioinformatics for the wider undergraduate community, and facilitates ongoing opportunities for K-12 students to participate in citizen-science research.

There is a wealth of information on the source, pathways and vectors of *C. gigas* based largely on historical documents but no study has comprehensively tested whether these historical accounts are correct using a worldwide, population genetic survey. Using >14K single-nucleotide polymorphisms (SNPs) from 41 populations across five continents a high level of spatial genetic differentiation was found within the native range and differences in source populations among non-native regions. Preliminary genetic data indicated that the parasitic protist, *Haplosporidium nelsoni* arrived with C. gigas imports to the US Atlantic coastline and then infected the native C. virginica, however the native source populations, the pathways and vector from which *H. nelsoni* arrived remain unknown. This project couples high-throughput sequencing technologies and Approximate Bayesian Computing (ABC)-based models to answer the following: What are the population genomic patterns among *C. gigas* from native and non-native regions? What are the population genomic patterns of *Haplosporidium nelsoni* among Asian and North American Crassostrea gigas and eastern North American *C. virginica*? What were the source populations and invasion pathways of *C. gigas* and *H. nelsoni*? Identifying source locations, pathways and vectors of introduction of C. gigas will provide researchers with a null-model of invasion history for dozens of other non-native species that were transported with *C. gigas*. Currently, there are no verified 'vector maps' for historical shipments of *C. gigas* that are similar to those generated from modern-day or historical shipping records.

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-1924599
NSF Division of Biological Infrastructure (NSF DBI)	DBI-1759759

[table of contents | back to top]