

Red seaweed *Gracilaria vermiculophylla* population genomics (SNP genotype likelihoods) from collections of populations in Asia North America and Europe in 2014 and 2015

Website: <https://www.bco-dmo.org/dataset/998990>

Data Type: Other Field Results

Version: 1

Version Date: 2026-05-19

Project

- » [Detecting genetic adaptation during marine invasions](#) (Genetic Adaptation Marine Inv)
- » [The genetic legacy of an Asian oyster introduction and its disease-causing parasite](#) (Oyster historical genetics)

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

Abstract

We describe the relative influence of bottlenecks, clonality, and population expansion in determining genomic variability of the widespread red macroalga *Gracilaria vermiculophylla*. Its introduction from mainland Japan to the estuaries of North America and Europe coincided with shifts from predominantly sexual to partially clonal reproduction and rapid adaptive evolution. We surveyed 62,285 single nucleotide polymorphisms (SNPs) for 351 individuals from 35 populations, aligned to 24 chromosome-length scaffolds that we publish for the first time. This dataset includes metadata, methods, links to published code and processed data used to generate figures for the results publication Flanagan et al. (2021,doi:10.1111/mec.15854) titled "Founder effects shape linkage disequilibrium and genomic diversity of a partially clonal invader." This dataset also includes genetic accession identifiers for sequence data contributed to the National Center for Biotechnology Information (NCBI)'s Sequence Read Archive (SRA) and Genbank databases, available under BioProject PRJNA700770.

Table of Contents

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

Coverage

Location: intertidal estuaries worldwide

Spatial Extent: N:54.486315 E:144.879657 S:31.7321 W:-123.055447

Temporal Extent: 2014 - 2015

Dataset Description

The related BioProject PRJNA700770 aggregates accessions in the Sequence Read Archive (SRA), and related Genbank nucleotide accession JAHNZQ000000000.2, and assembly "ASM1915520v2" (GCA_019155205.2).

Methods & Sampling

We collected *A. vermiculophyllum* thalli at 35 intertidal sites across the Northern Hemisphere. This included 11 Japanese sites, five sites along the western coast of North America, 10 sites on the eastern coast of the United States, and nine sites along the European coast. At each site, we haphazardly collected 100 thalli separated by at least one meter and determined the reproductive state of each thallus (i.e., male gametophyte, female gametophyte, diploid tetrasporophyte, or nonreproductive) using a dissecting microscope. We then preserved 5–10 cm fragments of each thallus in silica gel as vouchers and for DNA extraction. Total gDNA was extracted from approximately 5–10 mg of dried tissue of 560 thalli ($n = 16$ per site) that were either phenotypically-diploid (i.e., bearing reproductive structures) or when not enough thalli were reproductive from a site, we used nonreproductive thalli. We used the Nucleospin Plant Kit (Macherey-Nagel) and followed the manufacturer's protocol with two exceptions: we performed the lysis step for 1 h at room temperature and eluted in 100 μ l of molecular grade water. Briefly, we digested gDNA with two restriction enzymes, EcoRI and MseI, and ligated adaptors containing unique 8–10 bp barcodes to the digested DNA of each individual. The products were then PCR amplified in two independent reactions with standard Illumina primers. All amplicons were pooled and shipped to the University of Texas Genomic Sequencing and Analysis Facility, which used BluePippin Prep to isolate the 300–500 bp fraction. This fraction was then single-read sequenced with one lane each on Illumina HiSeq 2500 and HiSeq 4000 platforms.

Name synonyms for "*Gracilaria (Agarophyton) vermiculophyllum*"

Accepted name (as of 2026-05-21):

Gracilaria vermiculophylla, LSID(urn:lsid:marinespecies.org:taxname:236157)

NCBI Taxonomy ID:2608709

Unaccepted synonym (as of 2026-05-21):

Agarophyton vermiculophyllum, LSID(urn:lsid:marinespecies.org:taxname:1327786)

Data Processing Description

The two Illumina sequencing lanes yielded 4.4×10^8 short read fragments after removal of PhiX sequences. Fragments that did not have adapter nor barcode sequences were removed using a custom perl script. All reads were then aligned to the 24 scaffolds to identify variant, multiallelic SNPs and generate genotype likelihoods. We then removed loci with a minor allele frequency less than or equal to 2%. We filtered out poorly sequenced individuals (i.e., reads at <20% of all loci) and loci (i.e., 1+ reads at 50% of individuals). We were left with a set of 62,285 loci and 351 individuals, sequenced to a depth of 19.4 reads per locus

BCO-DMO Processing Description

- Loaded data from "meta_forBCODMO.csv" as table "998990_v1_g-verm-pop-genomics" (CSV format, header row 1, missing values: "", "nd")
- Split column "lat_lon" into "lat", "lon" and converted to decimal degrees. lat_lon column removed after verifying correct parsing.
- Renamed column "Sample Name" to "Sample_Name"
- Set types for all retained columns: "BioProject", "BioSample", "Run", "Sample_Name", "SiteName_long", "SiteName_short", "geo_loc_name_country", "geo_loc_name_country_continent", "lat_lon" as string; "Collection_Date" as integer; "lat" and "lon" as number
- Updated field metadata (descriptions, standard name IDs, units) for all columns including BioProject, BioSample, Run, Sample_Name, SiteName_long, SiteName_short, geo_loc_name_country, geo_loc_name_country_continent, Collection_Date, lat, and lon
- Replaced escape character sequences (backslash followed by spaces) with a single space in "SiteName_long" values
- Output final table as "998990_v1_g-verm-pop-genomics.csv"

Organism name note: the name as provided in the metadata "*Gracilaria (Agarophyton) vermiculophyllum*" combines genera and species of both the currently accepted and unaccepted synonyms. Both names and life science identifiers were included for discoverability purposes.

Problem Description

NA

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

Related Publications

Flanagan, B. A., Krueger-Hadfield, S. A., Murren, C. J., Nice, C. C., Strand, A. E., & Sotka, E. E. (2021). Founder effects shape linkage disequilibrium and genomic diversity of a partially clonal invader. *Molecular Ecology*, 30(9), 1962–1978. Portico. <https://doi.org/10.1111/mec.15854>
Results

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

Related Datasets

Software

Erik Sotka. (2026). *esotka/gvermSNPs: GracilariaVermiculophyllaSNPs* (Version v1.0) [Computer software]. Zenodo. <https://doi.org/10.5281/ZENODO.18672830>

IsRelatedTo

College of Charleston (2021). Agarophyton vermiculophyllum population genomics. 2021/02.NCBI:BioProject: PRJNA700770. In: BioProject [Internet]. Bethesda, MD: National Library of Medicine (US), National Center for Biotechnology Information; Available from: <http://www.ncbi.nlm.nih.gov/bioproject/PRJNA700770>

Flanagan, B., Krueger-Hadfield, S., Murren, C., Nice, C., Strand, A., & Sotka, E. (2021). *Founder effects shape linkage disequilibrium and genomic diversity of a partially clonal invader* (Version 4) [Dataset]. Dryad. <https://doi.org/10.5061/DRYAD.DV41NS1XC>

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

Parameters

Parameter	Description	Units
Sample_Name	Individuals	unitless
SiteName_short	Site name identifier (3 characters)	unitless
Run	The National Center for Biotechnology Information (NCBI) Run accession in the Sequence Read Archive (SRA)	unitless
BioProject	National Center for Biotechnology Information (NCBI) BioProject identifier.	unitless
BioSample	The National Center for Biotechnology Information (NCBI) BioSample accession number.	unitless
Collection_Date	Collection year (yyyy)	Year
SiteName_long	Full site name (not truncated)	unitless
geo_loc_name_country	Country	unitless
geo_loc_name_country_continent	Continent	unitless
lat	latitude	decimal degrees
lon	longitude	decimal degrees

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

Instruments

Dataset-specific Instrument Name	Illumina sequencing machine (HiSeq 2500 and 4000)
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	384-well Thermocycler (Eppendorf)
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

Detecting genetic adaptation during marine invasions (Genetic Adaptation Marine Inv)

Coverage: Estuaries of NW and NE Pacific; estuaries of NW and NE Atlantic

Description from NSF award abstract:

Biological introductions, defined as the establishment of species in geographic regions outside the reach of their natural dispersal mechanisms, have dramatically increased in frequency during the 20th century and are now altering community structure and ecosystem function of virtually all marine habitats. To date, studies on marine invasions focus principally on demographic and ecological processes, and the importance of evolutionary processes has been rarely tested. This knowledge gap has implications for management policies, which attempt to prevent biological introductions and mitigate their impacts. The Asian seaweed *Gracilaria vermiculophylla* has been introduced to every continental margin in the Northern Hemisphere, and preliminary data indicate that non-native populations are both more resistant to heat stress and resistant to snail herbivory. The project will integrate population genetics, field survey and common-garden laboratory experiments to comprehensively address the role of rapid evolutionary adaptation in the invasion success of this seaweed. Specifically, the PIs will answer the following. What is the consequence of introductions on seaweed demography and mating systems? How many successful introductions have occurred in North America and Europe? Where did introduced propagules originate? Do native, native-source and non-native locations differ in environmental conditions? Do native, native-source and non-native populations differ in phenotype?

The intellectual merit of this project is based on three gaps in the literature. First, while biological invasions are widely recognized as a major component of global change, there are surprisingly few studies that compare native and non-native populations in their biology or ecology. Native and non-native populations will be surveyed in a similar manner, allowing assessment of differences in population dynamics, mating system, epifaunal and epiphytic communities, and the surrounding abiotic and biotic environment. Second, *G. vermiculophylla* exhibits a life cycle typical of other invasive species (including some benthic invertebrates), yet we still lack data on the effects of decoupling the haploid and diploid stages on genetic structure, and in turn, on the evolvability of their populations. Finally, this project will provide unequivocal evidence of an adaptive shift in a marine invasive. To our knowledge, such evolutionary change has been described previously for only a complex of marine copepod species. *G. vermiculophylla* will serve as a model for understanding evolution in other nuisance invasions, and perhaps lead to novel methods to counter future invasions or their spread.

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

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-1357386
NSF Division of Ocean Sciences (NSF OCE)	OCE-1924599

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