RADseq data from Atlantic silversides used for linkage and QTL mapping.

Website: https://www.bco-dmo.org/dataset/924886
Data Type: experimental
Version: 1
Version Date: 2024-04-24

Project
» Collaborative research: The genomic underpinnings of local adaptation despite gene flow along a coastal environmental cline (GenomAdapt)

Contributors	Affiliation	Role
Therkildsen, Nina Overgaard	Cornell University (Cornell)	Principal Investigator, Contact
Baumann, Hannes	University of Connecticut (UConn)	Co-Principal Investigator
Akopyan, Maria	Cornell University (Cornell)	Student
Soenen, Karen	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

Abstract
ddRADseq data from 568 Atlantic silversides (Menidia menidia) that are either F1 or F2 offspring to wild-caught parents from Georgia and New York used in a controlled breeding experiment. The data were used to build linkage maps for each of the separate populations and their inter-population cross, and to perform quantitative trait locus mapping.

Table of Contents

- Coverage
- Dataset Description
  - Methods & Sampling
  - Data Processing Description
  - BCO-DMO Curation Notes
- Data Files
- Supplemental Files
- Related Publications
- Related Datasets
- Parameters
- Instruments
- Project Information
- Funding

Coverage

Location: Jekyll Island, Georgia and Patchogue, NewYork USA
Spatial Extent: **N**:40.75 **E**:-73 **S**:31.02 **W**:-81.43
Temporal Extent: 2017-05-01 - 2018-05-09

Dataset Description

* Raw data from the RADseq libraries are available under NCBI BioProject accession number PRJNA771889 (see related dataset section).
* SNP genotype call files (VCF format) are available at doi:10.6084/m9.figshare.19521955.v1 (see related dataset section) and as supplemental files to this dataset.

Methods & Sampling

We generated three crosses for linkage mapping, including two F1 families resulting from reciprocal crossing of wild-caught silversides from two adaptively divergent parts of the distribution range (Georgia and New York), and one F2 family from intercrossing laboratory-reared progeny from one of the F1 families. Because linkage mapping measures recombination during gamete production in the parents, the F1 families give us separate information about the wild-caught male and female founder fish from each separate population (the F0 progenitors), and the F2 map reflects recombination in the hybrid F1 progeny.

In the spring of 2017, spawning ripe founders were caught by beach seine from Jekyll Island, Georgia (31°03’N, 81°26’W) and Patchogue, New York (40°45’N, 73°00’W) and transported live to the Rankin Seawater Facility at University of Connecticut's Avery Point campus. For each family, we strip-spawned a single male and a single female onto mesh screens submerged in seawater-filled plastic dishes, then transferred the fertilized embryos to rearing containers (20 L) placed in large temperature-controlled water baths with salinity (30 psu) and photoperiod held constant (15 L:9 D). Water baths were kept at 20°C for the New York mother and at 26°C for Georgia mother families, which increased hatching success by mimicking the ambient spawning temperatures at the two different latitudes. Post hatch, larvae were provided ad libitum rations of newly hatched brine shrimp nauplii (Artemia salina, brineshrimpdirect.com). At 22 days post hatch (dph), we sampled 138 full-sib progeny from each of the two F1 families to be genotyped. The remaining offspring from the Georgia-mother F1 family were reared to maturity in groups of equal density (40–50 individuals) in 24°C water baths. In spring 2018, one pair of adult F1 siblings from the Georgia family were intercrossed to generate the F2 mapping population. At 70 dph, we sampled 221 full-sib F2 progeny for genotyping. In total, we analyzed 503 individuals: the two founders (male and female) and 138 offspring from each of the two F1 families, plus two additional F1 siblings from the Georgia mother F1 family and their 221 F2 offspring. All animal care and euthanasia protocols were carried out in accordance with the University of Connecticut's Institutional Animal Care and Use Committee (A17-043).

We extracted DNA from each individual with a Qiagen DNeasy tissue kit following the manufacturer's instructions and used double-digest restriction-site associated DNA (ddRAD) sequencing (Peterson et al., 2012) to identify and genotype single nucleotide polymorphisms (SNPs) for linkage map construction. We created two ddRAD libraries, each with a random subset of ~250 barcoded individuals, using restriction enzymes MspI and PstI (New England BioLabs cat. R0106S and R3140S, respectively), following library construction steps as in Peterson et al. (2012). We size-selected libraries for 400– 650 bp fragments with a Pippin Prep instrument (Sage Science) and sequenced the libraries across six Illumina NextSeq500 lanes (75 bp single- end reads) at the Cornell Biotechnology Resource Centre. Raw reads were processed in Stacks v2.53 (Catchen et al., 2013) with the module process_radtags to discard low-quality reads and reads with ambiguous barcodes or RAD cut sites. The reads that passed the quality filters were demultiplexed to individual fastq files. To capture genomic regions potentially not included in the current reference genome assembly, we ran the ustacks module to assemble RAD loci de novo (rather than mapping to the reference genome). We required a minimum of three raw reads to form a stack (i.e., minimum read depth, default -m option) and allowed a maximum of four mismatches between stacks to merge them into a putative locus (-M option).

Because the founders contain all the possible alleles that can occur in the progeny (except from any new mutations), we assembled a catalogue of loci with cstacks using only the four wild-caught F0 progenitors. We built the catalogue with both sets of founders to allow cross-referencing of common loci across the resulting F1 maps and we allowed for a maximum of four mismatches between loci (-n option). We matched loci from all progeny against the catalogue with sstacks, transposed the data with tsv2bam to be organized by sample rather than locus, called variable sites across all individuals, and genotyped each individual at those sites with gstacks using the default SNP model (marukilow) with a genotype likelihood ratio test critical value (α) of 0.05. Finally, we ran the populations module three times to generate a genotype output file for each mapping cross. For each run of populations, we specified the type of test cross (-- map- type option cp or F2), pruned unshared SNPs to reduce haplotype-wise missing data (-H option), and exported loci present in at least 80% of individuals in that cross (-r option) to a VCF file, without restricting the number of SNPs retained per locus.

Data Processing Description

The NCBI accessions refer to raw sequencing files.

The .vcf files contain SNP genotype data processed as described under methods

Raw data from the RADseq libraries are available under NCBI BioProject accession number PRJNA771889 (see related datasets)

SNP genotype call files (VCF format) are available atfigshare (see related publications) and as supplemental files to this dataset.

BCO-DMO Curation Notes

* Added lat/lon of sampling location for the mother and father
* Added SRA information: SRA study, experiment, run & sample name
* Adjusted field names to database requirements

[table of contents | back to top]

Data Files

File
924886_v1_seq.csv (Comma Separated Values (.csv), 162.49 KB) MD5:d2a5c032593fce2be4a3e36753cb9725 Primary data file for dataset ID 924886, version 1

[table of contents | back to top]

Supplemental Files

File
Genotypes for F2 offspring filename: F2.vcf.gz (GZIP (.gz), 95.23 MB) MD5:beb8dbe941d6e916ffc105e104d32534 This file contains called genotypes (in vcf format) for F2 offspring generated by an intercross among F1 individuals of Atlantic silversides from Jekyll Island GA and Patchogue, NY. The file was generated with the procedures described under methods.
Genotypes for Patchogue mother F1 linkage map filename: PJ.vcf.gz (GZIP (.gz), 57.70 MB) MD5:7a81f146baca4b49721009b3e41740aa This file contains called genotypes (in vcf format) for individuals used to generate the linkage map of F1 individuals with a mother from Patchogue, NY and a father from Jekyll Island, GA. The file was generated with the procedures described under methods. A sample ID including F1 indicates that the sample is an F1 offspring. Samples IDs including F0 indicates that the sample was a parent of the cross.
Genotypes filename: JP.vcf.gz (GZIP (.gz), 53.74 MB) MD5:7465e6593eebe1fcbf3648fbff267a67 This file contains called genotypes (in vcf format) for individuals used to generate the linkage map of F1 individuals with a mother from Jekyll Island, GA and a father from Patchogue, NY. The file was generated with the procedures described under methods. A sample ID including F1 indicates that the sample is an F1 offspring. Samples IDs including F0 indicates that the sample was a parent of the cross.

[table of contents | back to top]

Related Publications

Akopyan, M., Tigano, A., Jacobs, A., Wilder, A. P., Baumann, H., & Therkildsen, N. O. (2022). Comparative linkage mapping uncovers recombination suppression across massive chromosomal inversions associated with local adaptation in Atlantic silversides. Molecular Ecology, 31(12), 3323–3341. Portico. https://doi.org/10.1111/mec.16472
Results

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. Molecular Ecology, 22(11), 3124–3140. Portico. https://doi.org/10.1111/mec.12354

Methods

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. PLoS ONE, 7(5), e37135. https://doi.org/10.1371/journal.pone.0037135
Methods

[table of contents | back to top]

Related Datasets

References

Cornell University. Menidia menidia Linkage Map. 2021/10. 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/PRJNA771889. NCBI:BioProject: PRJNA771889. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA771889

IsSupplementedBy

Akopyan, M. (2022). *SNPS for linkage mapping* [Data set]. figshare. https://doi.org/10.6084/M9.FIGSHARE.19521955.V1 https://doi.org/10.6084/m9.figshare.19521955.v1

[table of contents | back to top]

Parameters

Parameter	Description	Units
bioproject_accession	NCBI BioProject accession number	units
biosample_accession	NCBI BioSample accession number	units
taxonomic_name	Taxonomic name of specimen	units
mother_f0_sampling_location	Sampling location ()Jekyll Island, GA for the wild-caught mother of the F0 cross used to generate F1 and F2 offspring (the generation of each sample is listed in the filename column).	units
lat_mother	Latitude of sampling location of wild-caught mother	units
lon_mother	Longitude of sampling location of wild-caught mother	units
father_f0_sampling_location	Sampling location (Patchogue, NY) for the wild-caught father of the F0 cross used to generate F1 and F2 offspring (the generation of each sample is listed in the filename column).	units
lat_father	Latitude of sampling location of wild-caught father	units

lon_father	Longitude of sampling location of wild-caught father	units
SRA_study_accession	NCBI SRA study accession number	units
SRA_experiment_accession	NCBI SRA experiment accession number	units
SRA_run_accession	NCBI SRA run accession number	units
library_ID	Short, unique identifier for the sequencing library	units
title	Short description that identifies the dataset in NCBI	units
library_strategy	The library preparation type used for the sample (details in Akopyan et al. 2022	units
library_source	The type of DNA used to prepare the sequencing library	units
library_selection	NCBI Controlled vocabulary of terms describing selection or reduction method use in library construction	units
library_layout	Either paired-end or single-end reads	units
platform	Manufacturer of the sequencing instrument	units
instrument_model	Model of the sequencing instrument	units
design_description	Type of experimental design for original study	units
filetype	Type of file wiith raw sequencing data	units
sample_name	Name of the sample. Sample names containing F0 were wild-caught founders of the cross. Sample names including F1 are from F1 offspring. File names containing F1_x_ are F1 offspring intercrossed with other F1 to generate F2. Sample names containing F2 are for F2 offspring.	units
filename	Name of the fastq file. File names containing F0 were wild-caught founders of the cross. File names including F1 are from F1 offspring. File names containing F1_x_ are F1 offspring intercrossed with other F1 to generate F2. File names containing F2 are for F2 offspring.	units

[table of contents | back to top]

Instruments

Dataset-specific Instrument Name	Illumina NextSeq500
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.

[table of contents | back to top]

Project Information

Collaborative research: The genomic underpinnings of local adaptation despite gene flow along a coastal environmental cline (GenomAdapt)

Website: https://befel.marinesciences.uconn.edu/2018/03/07/research-news-new-nsf-grant-to-study-silverside-genes/

Coverage: Eastern coastline of North America

NSF Abstract:

Oceans are large, open habitats, and it was previously believed that their lack of obvious barriers to dispersal would result in extensive mixing, preventing organisms from adapting genetically to particular habitats. It has recently become clear, however, that many marine species are subdivided into multiple populations that have evolved to thrive best under contrasting local environmental conditions. Nevertheless, we still know very little about the genomic mechanisms that enable divergent adaptations in the face of ongoing intermixing. This project focuses on the Atlantic silverside (Menidia menidia), a small estuarine fish that exhibits a remarkable degree of local adaptation in growth rates and a suite of other traits tightly associated with a climatic gradient across latitudes. Decades of prior lab and field studies have made Atlantic silverside one of the marine species for which we have the best understanding of evolutionary tradeoffs among traits and drivers of selection causing adaptive divergence. Yet, the underlying genomic basis is so far completely unknown. The investigators will integrate whole genome sequencing data from wild fish sampled across the distribution range with breeding experiments in the laboratory to decipher these genomic underpinnings. This will provide one of the most comprehensive assessments of the genomic basis for local adaptation in the oceans to date, thereby generating insights that are urgently needed for better predictions about how species can respond to rapid environmental change. The project will provide interdisciplinary training for a postdoc as well as two graduate and several undergraduate students from underrepresented minorities. The findings will also be leveraged to develop engaging teaching and outreach materials (e.g. a video documentary and popular science articles) to promote a better understanding of ecology, evolution, and local adaptation among science students and the general public.

The goal of the project is to characterize the genomic basis and architecture underlying local adaptation in M. menidia and examine how the adaptive divergence is shaped by varying levels of gene flow and maintained over ecological time scales. The project is organized into four interconnected components. Part 1 examines fine-scale spatial patterns of genomic differentiation along the adaptive cline to a) characterize the connectivity landscape, b) identify genomic regions under divergent selection, and c) deduce potential drivers and targets of selection by examining how allele frequencies vary in relation to environmental factors and biogeographic features. Part 2 maps key locally adapted traits to the genome to dissect their underlying genomic basis. Part 3 integrates patterns of variation in the wild (part 1) and the mapping of traits under controlled conditions (part 2) to a) examine how genomic architectures underlying local adaptation vary across gene flow regimes and b) elucidating the potential role of chromosomal rearrangements and other tight linkage among adaptive alleles in facilitating adaptation. Finally, part 4 examines dispersal - selection dynamics over seasonal time scales to a) infer how selection against migrants and their offspring maintains local adaptation despite homogenizing

connectivity and b) validate candidate loci for local adaptation. Varying levels of gene flow across the species range create a natural experiment for testing general predictions about the genomic mechanisms that enable adaptive divergence in the face of gene flow. The findings will therefore have broad implications and will significantly advance our understanding of the role genomic architecture plays in modifying the gene flow - selection balance within coastal environments.

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-1756316

[table of contents | back to top]