

# Atlantic silverside (*Menidia menidia*) cDNA transcriptome and TSA accessions from specimens collected at Poquott Beach, New York in June of 2013 (Fishery Genome Changes project)

**Website:** <https://www.bco-dmo.org/dataset/686981>

**Data Type:** experimental

**Version:**

**Version Date:** 2017-04-06

## Project

» [High resolution genome changes during evolution in a classic fisheries experiment](#) (Fishery Genome Changes)

Contributors	Affiliation	Role
<a href="#">Palumbi, Stephen R.</a>	Stanford University	Principal Investigator, Contact
<a href="#">York, Amber D.</a>	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

## Table of Contents

- [Coverage](#)
- [Dataset Description](#)
  - [Methods & Sampling](#)
  - [Data Processing Description](#)
- [Data Files](#)
- [Parameters](#)
- [Instruments](#)
- [Deployments](#)
- [Project Information](#)
- [Funding](#)

## Coverage

**Spatial Extent:** Lat:40.9475 Lon:-73.1025

## Dataset Description

The data include Atlantic silverside (*Menidia menidia*) genetic accession information at the National Center for Biotechnology Information (NCBI). Accessions in this dataset are for cDNA transcriptome or a Transcriptome Shotgun Assembly (TSA). Links to the accession at NCBI are provided along with accession number and accession type. Atlantic silversides were collected at Poquott Beach, New York on June 20th and 21st, 2013.

## Methods & Sampling

### Raw cDNA transcriptome sequence reads:

The raw sequence data are deposited in the NCBI Sequence Read Archive (SRA) with accession numbers SRR3990241- SRR3990248 associated with BioProject PRJNA330848 and BioSamples SAMN05427525 - SAMN05427532

Species: Atlantic silverside (*Menidia menidia*)

Sample type: mRNA from mix of tissues

RNA extraction method: Qiagen RNeasy Plus Universal Tissue Mini Kit

Library preparation: Illumina's TruSeq RNA sample prep kit v2

Sequencing instrument: Illumina HiSeq 2000

### **Assembled Atlantic silverside transcriptome:**

An assembled Atlantic silverside transcriptome is deposited in the NCBI GenBank Transcriptome Shotgun Assembly Sequence Database (TSA). This version of the project (01) has the accession number GEVY01000000, and consists of sequences GEVY01000001-GEVY01020998. The cleaned RNA-seq reads from all samples were de novo assembled with two different programs: CLC Genomic Workbench v6.0.2 (both with an automatically optimized word size of 25 and a longer word size of 40) and Trinity v. r20131110 (with default settings, but retaining only the isoform with the highest mapped read depth within each subcomponent). We saw that each assembly contained a substantial set of unique transcripts not present in the other assemblies and therefore merged all three to maximize the gene space coverage in our final contig set. To reduce redundancy, we used cd-hit-est v4.5.4 to collapse the contig set into the longest representative for each unique sequence, and CAP3 v12/21/07 to meta-assemble partial assemblies of the same transcript. Following these procedures, we broke up likely chimeric contigs with the method by Yang and Smith (2013, BMC Genomics 14:328). Because we wanted to reduce our contig set to only include a single representative transcript for each silverside gene, we used a reciprocal best hit blast approach to extract non-redundant putative orthologs to the gene sets in three related species: platyfish (*Xiphophorus maculatus*), medaka (*Oryzias latipes*), and Nile tilapia (*Oreochromis niloticus*). We compared our contig set against the full peptide set for each reference species (downloaded from Ensemble release 75) with blastx, and then compared the peptide sequences for each species to our contig set with tblastn, in both cases using soft masking and an e-value cut-off of  $10e-4$ . For each reference species, we recorded reciprocal best hits (RBHs) when a contig and a protein had a best match to each other. We used a sequential approach to select putative orthologs. We first extracted the contigs that were RBHs to platyfish proteins (since this species yielded the highest number of RBHs). We also added additional contigs that had a best hit to a portion of an RBH protein not covered by the RBH contig (secondary hits (maximum overlap of 10 amino acids allowed)), under the assumption that these contigs represented transcript fragments. We then added contigs that were RBHs (and the associated secondary non-overlapping hits to the same proteins) to medaka proteins that were non-redundant to the platyfish proteins. Medaka proteins were considered non-redundant if they did not have a RBH to the previously extracted RBH platyfish protein set (in a direct blastp comparison of the two protein set) or was annotated to the same zebrafish gene (ZFIN ID) as an RBH platyfish protein. We similarly added contigs that were RBH or associated secondary hits to tilapia proteins that were non-redundant to the proteins included from the other species. To recover additional high quality non-redundant transcripts, we used TransDecoder to predict coding regions in our redundancy-reduced contig set on the basis of nucleotide composition, open reading frame (ORF) length and Pfam domain content. Of the contigs predicted to contain a complete ORF, we retained the subset which did not have a significant ( $e\text{-value} < 10e-2$ ) blastn hit to the RBH contig set (and therefore are non-redundant).

### **Methods are also published in:**

Therkildsen, N. O., and S. R. Palumbi. 2016. Practical low-coverage genomewide sequencing of hundreds of individually barcoded samples for population and evolutionary genomics in nonmodel species. *Molecular Ecology Resources*. doi: [10.1111/1755-0998.12593](https://doi.org/10.1111/1755-0998.12593)

### **Data Processing Description**

Assembly Method :: CLC Genomics Workbench 6.0.2; Trinity r20131110; CAP3 v12/21/07

BCO-DMO Data Manager Processing Notes:

- \* added a conventional header with dataset name, PI name, version date
- \* modified parameter names to conform with BCO-DMO naming conventions
- \* broke sampling location description into sample\_location and sample\_state
- \* added lat/lon of sample site to dataset

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

---

### **Data Files**

File
<b>accessions.csv</b> (Comma Separated Values (.csv), 1.63 KB) MD5:389670823074b9182bc115ad5bee27d7
Primary data file for dataset ID 686981

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

## Parameters

Parameter	Description	Units
accession_id	Accession identifier at National Center for Biotechnology Information (NCBI)	unitless
accession_link	Accession link at NCBI	unitless
sample_type	Sample description	unitless
BioProject	BioProject identifier at NCBI	unitless
species	Scientific name of specimen	unitless
sample_site	Location specimens were sampled	unitless
sample_state	state specimens were sampled	unitless
lat_approx	Approximate latitude of sampling site; South is negative	decimal degrees
lon_approx	Approximate longitude of sampling site; West is negative	decimal degrees

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

## Instruments

<b>Dataset-specific Instrument Name</b>	Illumina HiSeq 2000
<b>Generic Instrument Name</b>	Automated DNA Sequencer
<b>Generic Instrument Description</b>	General term for a laboratory instrument used for deciphering the order of bases in a strand of DNA. Sanger sequencers detect fluorescence from different dyes that are used to identify the A, C, G, and T extension reactions. Contemporary or Pyrosequencer methods are based on detecting the activity of DNA polymerase (a DNA synthesizing enzyme) with another chemoluminescent enzyme. Essentially, the method allows sequencing of a single strand of DNA by synthesizing the complementary strand along it, one base pair at a time, and detecting which base was actually added at each step.

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

## Deployments

### Palumbi silversides

<b>Website</b>	<a href="https://www.bco-dmo.org/deployment/687817">https://www.bco-dmo.org/deployment/687817</a>
<b>Platform</b>	Poquott Beach
<b>Description</b>	Atlantic silverside ( <i>Menidia menidia</i> ) specimens collected from Poquott Beach (NY), Jekyll Island (GA), Patchogue (NY), Minas Basin (Nova Scotia), and Madalen Islands (Quebec).

## Project Information

### High resolution genome changes during evolution in a classic fisheries experiment (Fishery Genome Changes)

**Coverage:** The East coast of North America

*Description from NSF award abstract:*

One of the strongest impacts there is on ocean species is reduction in population size due to fishing. Even for species that are not overfished, harvest takes a large fraction of the biggest and fastest growing individuals. As a result, fishing exerts strong natural selection on fish population, selecting for slow growing, small individuals. Experiments in artificial fishing have shown rapid evolution of growth rates, maturation size and other traits for lab populations. The classic Conover-Munsch experiments ten years ago showed the power of fishing to generate rapid evolution. However, no analysis of the genetic impact of fishing under such controlled conditions has been done, and no investigation of the way whole genomes respond to strong fisheries evolution has been attempted. Luckily, the Conover-Munsch samples - from the fish used in their classic experiment - have been preserved, and modern genomic techniques are now available that can analyze the way fisheries-induced evolution shaped the genetic diversity and genome architecture of these populations. Strong selection is known in other systems to leave a legacy of deleterious changes in the genome. The results of this study will be important components of understanding the long-term effects of fishing because they will for the first time allow a mechanistic understanding of how natural selection works on fish populations. The project will also compare the changes that occur after the relaxation of fishing pressure to estimate if there is a legacy of deleterious genome changes in fished species that impedes their recovery. The data from this study will show how fishing creates change at loci under selection, and also how this strong selection generates other, non-adaptive shifts because of genetic hitchhiking and inbreeding.

The investigators will use next generation DNA sequencing to sequence the protein coding regions of the Conover-Munsch fish samples. They will discover, document and compare genetic variants across the genome in lines selected for large size, in lines selected for small size, and in the original populations in order to chart evolutionary changes at the genomic level imposed by fishing. They will use outlier analyses to pinpoint loci at which strong selection has acted to change allele frequencies. They will compare these changes to changes in fish populations after the relaxation of fishing in order to distinguish evolutionary changes that are easily reversible from those that are not. In addition, they will also compare genetic changes induced by fishing to those that occur naturally along an environmental gradient on the US east coast. Preliminary data indicate that many of the genetic variants selected for in the fisheries experiment are old variants - estimated by patterns of linkage disequilibrium - already present in the original population. Working on the genetics of the natural gradient will show which of these variants have been selected by evolutionary forces in the native environment, and subsequently been favored by the novel evolutionary pressure of fishing.

## Funding

Funding Source	Award
<a href="#">NSF Division of Ocean Sciences (NSF OCE)</a>	<a href="#">OCE-1434325</a>