Genome-wide population structure at microgeographic scales in the endangered Bluemask Darter (*Etheostoma akatulo*) from the Caney Fork River System

Benavides E.[1], Taylor L.U.[1], J. Simmons[2], D. Macguigan[1], C. Parker[1], D. Kim[1], and T. J. Near[1,3]

[1]Department of Ecology and Evolutionary Biology, Yale University; [2]Tennessee Valley Authority, Chattanooga, Tennessee; [3]Yale Peabody Museum of Natural History, New Haven, Connecticut

Introduction

The Bluemask Darter (Percidae: *Etheostoma akatulo*) is a federally endangered darter species with a tightly restricted distribution along four waterways in the Caney Fork river system of Tennessee (Layman and Mayden, 2009). A population once occupying a fifth waterway (the Calfkiller River) is now extirpated.

We normally expect genetic differentiation to follow patterns of geographic distance, physical boundaries, or ecological differentiation across heterogeneous habitats. None of these potential drivers are obvious in the narrow range of *E. akatulo* upstream of the Great Falls Dam. Despite this fact, previous work with microsatellites has uncovered unexpected signals of differentiation across three populations (Collins River//Rocky River//Cane Creek+Caney Fork) for these darters (Robinson et al., 2013).

In this project, we use genome-scale reduced representation sequencing (double-digest RAD-seq) data to assess the phylogenetic relationships, population structure, and genetic connectivity of *E. akatulo* individuals from across their range, with additional context from eight closely related darter species within the *Doration* clade.

Our phylogeny strongly supports every species within the *Doration* clade except for E. *clinton and E. teddyroosevelt.* Within *E. akatulo,* we find strong support for a split between western individuals (from Collins River and Scott Creek) and eastern individuals (from Rocky River, Caney Fork, and Cane Creek). Although Rocky River individuals show some signs of admixture, we find the strongest support for east/west population differentiation and little support for patterns of genetic structure within those clades. Our results highlight hidden population structure in a species with a narrow range and no clear geographical boundaries.

Sample Collection and ddRAD sequencing

E. akatulo tissue samples were collected by the Tennessee Valley Authority in 2005, 2015, and 2017. Samples consisted of pelvic or anal fin clips and individuals were released back into the sampling area. We isolated DNA with the Qiagen DNeasy extraction protocol (Qiagen, Valencia, CA) and determined extraction concentrations with gel electrophoresis and a Qubit v. 3.0 fluorometer (ThermoFisher, Philadelphia, PA). We prepared 96-sample, multiplexed RAD-seq libraries using double digestion with PstI and MspI restriction enzymes and sample-specific barcodes (Floragenex Inc, Portland, OR, USA). Each library included a replicate to minimize the influence of PCR duplication bias and other technical errors. We sequenced libraries on an Illumina HiSeq 2000 using single-end 100 base pair sequencing at the University of Oregon GC3F facility (https://gc3f.uoregon.edu/).

Data Wrangling

We used Ipyrad (http://github.com/dereneaton/ipyrad) to demultiplex sequence data, filter reads, and generate de novo assemblies. Reads with Phred scores <20 for more than five bases were excluded, and a minimum depth of six reads was required for statistical base calling. The clustering threshold of sequence similarity for building loci within individuals was set to 0.96 (McCartney-Melstad et al., 2019).

Three *E. akatulo* individuals were removed from downstream analysis due to low loci counts or signs of contamination. We performed downstream analyses and visualization in R (R Core Team, 2018) with the tidyverse package (Wickham, 2017). All computation was done on the Farnam cluster at the Yale Center for Research Computing.

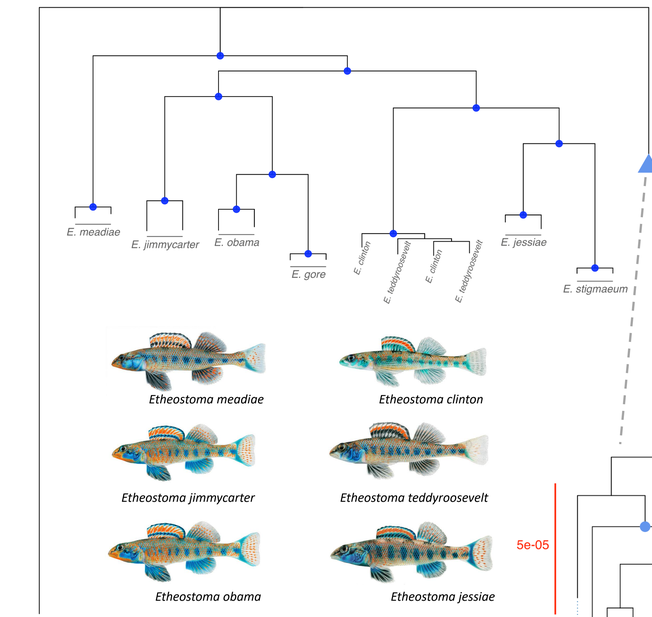
Phylogeny

We assembled a maximum likelihood consensus tree with IQTree (Nguyen et al, 2015) with a GTR substitution model and 1000 ultrafast bootstraps. The phylogenetic dataset used loci present in at least 20 of 93 individuals (76 *E. akatulo,* 17 outgroup). Data ranged from 14,451 to 86,946 loci per individual with an average of 49,029 loci. We used the ggtree package for visualization (Yu et al., 2017). Blue nodes indicate ≥95% bootstrap support.

Population Structure

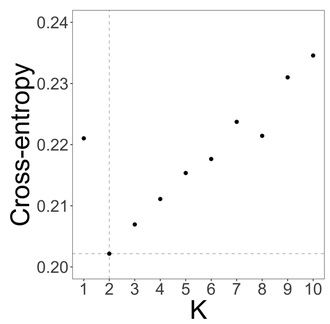
We assessed underlying population structure with SNMF in the LEA R package (Frichot et al., 2014; Frichot and François, 2015). We tested cross-entropy values for K=1:10 with 1000 replicate runs for each K. The population structure dataset used loci present in at least 60 of the 76 *E. akatulo* individuals. Data ranged from 2,445 to 8,857 loci per individual, with an average of 6,806 loci. The height of the bars in the SNMF plot indicate the posterior probability of population assignment for K=2. We used the Threepop program in TreeMix (Pickrell and Pritchard, 2012) to calculate f_3 statistics for *E. akatulo* individuals assigned to West, East, and Rocky River populations.

For acknowledgements, references, and further information, visit:
https://ltaylor2.github.io/akatulo



Bluemask Darter breeding male (*Etheostoma akatulo*)

Credit all drawings except *E. stigmaeum*: Joseph Tomelleri (www. americanfishes.com)



Strongest support for K=2