Project Report for USDA-ARS Project Entitled

Genetic Variation Related to Growth Performance in Domestic and Wild Striped Bass

Submitted on September 28, 2018
by Dr. Adrienne Kovach, University of New Hampshire

with project collaborators:
Dr. Benjamin Reading, North Carolina State University
&
Dr. David Berlinsky, University of New Hampshire

Project Goal and Rationale

The overarching goal of this project was to explore genetic variation related to growth performance of domestic striped bass (*Morone saxatilis*) and those of wild origin derived from wild spawning populations from six geographic locations along the Atlantic coast. The project leveraged ongoing striped bass growth studies of domestic and wild-origin striped bass reared in competitive growth trials at the University of New Hampshire and North Carolina State University. Integrating genotyping from next generation sequencing with these growth trial studies will provide insight into genetic markers associated with growth performance and ultimately inform mass selection practices.

Approach

Fingerling striped bass (0.5-4.0 g), bred in captivity during two consecutive natural spawning seasons (spring 2013 and 2014) at state, federal, private, and university hatcheries in Texas (TX), Florida (FL), South Carolina (SC), Virginia (VA), Delaware (DE), North Carolina (NC) and Nova Scotia (NS) were transported to the Ritzman Aquaculture Laboratory (University of New Hampshire, UNH; Durham, NH, USA) for two growth trials. Striped bass broodstock were wild-caught from the spawning grounds except the NS broodstock that were wild-captured as juveniles and maintained in captivity. The striped bass strains used in these trials were selected because of their geographic distribution, hatchery availability, and genetic contributions to the USDA national breeding program housed at North Carolina State University.

The domesticated striped bass line (DOM) was originally created from hundreds of outcrosses of six distinct striped bass strains (Canada, Pacific Ocean, Roanoke River, Chesapeake Bay, Santee-Cooper Reservoir, and Florida-Gulf of Mexico) in the *National Program for Genetic Improvement and Selective Breeding for the Hybrid Striped Bass Industry* at the North Carolina State University (NCSU) Pamlico Aquaculture Field Laboratory (Aurora, NC). The DOM striped bass were mass selected for performance and body conformation in fresh and brackish water aquaculture systems for ~8-9 generations and

primarily serve as broodstock for producing the hybrid striped bass and supplying select aquaculture producers. Juvenile DOM striped bass were produced at NCSU and transported to UNH.

Comparative growth trials were conducted from 2013-2016 in recirculating aquaculture systems maintained at different salinities through a collaboration at UNH and University of Maryland. Performance was closely monitored until strains reached two years of at which point final observations were made (finish weights and specific growth rates, normalized for 750 days post spawn). Fin clips were collected from all individuals and stored in 95% ethanol for family identification by microsatellite genotyping upon conclusion of the growth trials. The full methodological details and results of the growth studies are found in Kenter et al. (2018).

For the genotyping studies to identify genetic markers associated with growth performance, we selected a subset of fish representing the fastest and slowest growing fish in each strain. To make this selection, we identified fish that were one standard deviation above or below the mean finish weight of their respective strain. This resulted in the selection of the 30 largest and 30 smallest fish from each of seven wild strains and one domestic strain, totaling 480 individuals. In some strains, the size selection bins were relaxed to 0.8 standard deviations above and below the mean, to obtain consistent sample sizes of 30 for each comparative grouping.

From the 480 selected samples, we designed an experiment for three ddRAD Sequencing libraries following the Peterson et al. (2012) protocol. Our experimental design employed barcode sequences in conjunction with Illumina indices, so that we could uniquely barcode 160 samples for multiplexing in a single lane. We targeted a DNA concentration of 5-25 ng/ul. Samples below 10ng/ul after initial extraction were vacuum centrifuged to concentrate to within the target range. Samples that were above 25 ng/ul were diluted down to 25 ng/ul. DNA was digested with the enzymes *SbfI* and *MspI*, and ligated to P1 and P2 adapters using T4 DNA ligase (30 min at 37 ˚C and 60min at 20 ˚C, held at 10 ˚C). Samples were pooled into index groups by their unique P1 adapter and cleaned using 1.5x Agen- court AMPure XP beads. Using BluePippin (Sage Science, MA, USA), fragments were size selected between 400-700 bp in length. Low cycle PCR reactions were then preformed to incorporate the Illumina TruSeq primer sequences into the library, as well as a final clean up using AMPure XP beads. Libraries were visualized on a fragment Bioanalyzer to ensure desired fragment size/distribution and index groups pooled.

Initially, we submitted for sequencing a single ddRAD library consisting of 160 samples selected across four of the strains – FL, TX, SC and VA. This library was sequenced across on an Illumina HiSeq 2500 platform by Genewiz Next Generation Sequencing Services (South Plainfield, NJ).

Sequences were initially evaluated for overall quality using FastQC, then trimmed and filtered using FASTX-Toolkit. Specifically, reads were trimmed on the 3' end to 97 bp and eliminated if the Phred quality scores were below 10 or if 95% of the bases had Phred quality scores below 20. Using STACKS (version 1.48; Catchen et al. 2013), we

demultiplexed the remaining sequences. We used the process_radtags command with the following conditions: any reads not meeting Illumina's chastity/purity filter and of low quality were discarded, data were cleaned such that any read with an uncalled base was removed, reads with mismatches in the adapter sequence >1 were removed, and reads were only processed if the sequence had an intact SbfI RAD site and one of the unique barcodes. Subsequently, fastx_trimmer was used to trim all sequences to the length of the shortest sequences. Reads were aligned to the striped bass reference genome (Reading et al. in prep; https://cals.ncsu.edu/applied-ecology/striped-bass-genome-project/) using STACKS (version 1.48). Minimum stack depth for a read to be assembled into a catalog was 6. The number of mismatches allowed between sample loci was set at 5. The Populations program within STACKS was used to identify a subset of SNPs across the different size groups of this sample.

Downstream analyses of the SNP loci were performed to identify FSTs for the different sizes of fish (pooling across strain), and outlier methods (Bayescan, Foll and Gaggiotti 2008) were used to identify markers with elevated divergence between the two datasets, indicating a genetic variant that differs between the two size groups of fish.

Results and Next Steps

The sequencing produced a total of 627,131,910 raw reads. Upon demultiplexing, <25% of these reads were retained, resulting in 139,712,647 for downstream analysis. Upon further troubleshooting, we determined that the likely cause of this was a problem with the ligation of the barcoded adapters during the library preparation. The performance of this step will need to be improved in future work (see below).

As a result of the low numbers of reads retained, 40 of the 160 samples were dropped from the analysis due to low sequencing sample coverage, leaving 120 samples for this initial proof-of-concept analysis. We completed the SNP detection pipeline using this reduced dataset, which resulted in the recovery of 370 SNP loci in total, and 280 SNPs once we selected only 1 SNP per locus. This reduced dataset had low power to identify markers that diverged between the large and small weight groups of fish, and as a result, no outliers were identified in this initial analysis.

In ongoing work, we have developed an approach to improve the recovery of barcoded sequences. We have also identified another combination of enzymes *SphI* and *MluCI* that will yield a much higher number of SNPs. A virtual digest of these enzymes projected the recovery of 12,683 SNPs. In ongoing work, we are completing library preparation for the remaining 360 samples and building new libraries for the original 160 samples using these modifications to our ddRAD protocols. We anticipate that with orders of magnitude more SNPs and a more robust sample size (the full 480 sample dataset), we will have power to identify genetic variation associated with growth performance. Results of the future application of this improved methodology to the full dataset, representing all wild and domestic strains, will be disseminated in the form of a peer-reviewed application upon completion.

Literature Cited

Catchen, J., P. Hohenlohe, S. Bassham, A. Amores, and W. Cresko. 2013. Stacks: an analysis tool set for population genomics. Molecular Ecology 22: 3124-3140.

Foll M and O. Gaggiotti. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180(2):977–93.

Kenter, L., A.I. Kovach, C.L. Woods, B.J. Reading, and D.L. Berlinsky. 2018. Strain evaluation of striped bass (Morone saxatilis) cultured at different salinities. *Aquaculture* https://doi.org/10.1016/j.aquaculture.2018.04.017

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) https://doi.org/10.1371/journal.pone.0037135

Reading*, B.J., Baltzegar*, D.A., Dashiell, C., Chapman, R.W., Schilling, J., Williams, V.N., Ring, B.D., Hopper, M., McGinty, A.S., Borski, R.J., McCoy, S., Hiramatsu, N., Matsubara, T., Schaff, J.E., Opperman, C.H., and Sullivan, C.V. The striped bass (Morone saxatilis) genome sequence reveals a complex evolutionary history for teleost vitellogenins. *These authors contributed equally to this manuscript. (Manuscript in preparation).