United States Department of the Interior
Fish and Wildlife Service
Southwestern Native Aquatic Resources and Recovery Center
P.O. Box 219, 7116 Hatchery Road
Dexter, New Mexico 88230
575-734-5910, 575-734-6130 fax
March 2019

Title: Using Molecular Techniques to Determine Effective Number of Breeders (N_b) for Razorback Sucker and Colorado Pikeminnow in the San Juan River

Principal Investigator: Tracy Diver and Steve Mussmann, Southwestern Native Aquatic Resources and Recovery Center, Dexter, NM

Introduction & Justification

Recent advances in population-level genetic analyses are increasingly helping managers monitor and adaptively manage the recovery of endangered populations (Hartl and Clark 2007). Recovery plans for many endangered fishes include the production and release of hatchery-reared individuals to augment populations with two major objectives: increase population sizes and promote genetic diversity (Miller and Kapuscinski, 2003). In order to meet these recovery criteria, survival, reproduction, and recruitment of wild and augmented individuals must occur at a sustainable scale; however, understanding factors that limit success in achieving a self-sustaining population can be difficult to identify. Long-term datasets can provide insight into population responses that might hinder recovery, and genetic monitoring can be an additional tool for providing insight into complex ecological, demographic, and genetic factors that can impede the establishment of self-sustaining populations.

Augmentation of endangered fish populations in the San Juan River using captive-reared Razorback Sucker (*Xyrauchen texanus*) and Colorado Pikeminnow (*Ptychocheilus lucius*) began in the mid-1990s and continues as a recovery action (USFWS 2005; USFWS 2015). Annual monitoring of survival, reproduction, and recruitment of these populations has been supported through the San Juan River Basin Recovery Implementation Program (SJRRIP). Mark-recapture data on PIT tagged individuals have enabled estimation of survival of both stocked species (Franssen and Durst *unpublished data*; Clark et al. 2018), which in turn has prompted additional research investigating ways to increase survival. Larval fish surveys have documented successful reproduction of both species in the river (Farrington et al. 2015), but recruitment to the adult stage is extremely limited. Successful recovery likely requires a significant portion of the reestablished population to reproduce annually to both increase population sizes and ensure maintenance of genetic diversity. Therefore, quantifying the number of individuals that reproduce annually can provide data to make informed management decisions to aid in the reestablishment of self-sustaining populations.

Population-level spawning success (i.e., number of reproducing adults) is innately difficult to quantify from field studies, especially for highly fecund species where few individuals can produce a large number of offspring. Furthermore, there is substantial evidence that shows reproductive output can depend on environmental conditions and age- or size-related factors (Lauer et al. 2005; Lambert 2008). Both temporal and spatial variation in spawning effort has also been observed for Razorback Sucker with adults vacating a spawning area early in the season and later returning to spawn again that year (Marsh et al. 2015) and with individuals visiting multiple sites during the same spawning period (Modde and Irving 1998). Such reproductive strategies further compound the difficulty in determining individual contribution to cohorts over a reproductive season; however, this question can be addressed by

estimating the effective number of breeders (N_b) using genetic analyses. For long-lived, highly fecund, iteroparous species with overlapping generations, such as Colorado Pikeminnow and Razorback Sucker, N_b is an extremely useful metric for understanding population-level spawning success due to its defined seasonal reproductive bouts (Waples et al. 2013; Waples et al. 2014). Single cohort N_b estimates can reliably quantify the number of individuals that contributed to a given cohort (Waples et al 2014). Obtaining annual N_b estimates for the endangered fishes of the San Juan River may provide insight into how management activities (i.e., increasing passage, managing flows) effect population-level reproduction, thus, improving the opportunity to manage the San Juan River to favor spawning success.

Objectives:

1. Continue N_b monitoring for both Razorback Sucker (N = 120) and Colorado Pikeminnow (N = 120) collected in the San Juan River in 2020.
2. Expand upon N_b estimates for Razorback Sucker (N = 120) and Colorado Pikeminnow (N = 120) larvae captured below the waterfall for 2020.

Methods

Larval fish surveys are conducted annually along a 140 mile section of the San Juan River between Shiprock, NM, and Clay Hills, UT. Approximately 240 larval samples representing the spatial and temporal distribution of sampling efforts will be examined for N_b estimates. Tissue subsamples from the posterior portion of each specimen will be collected from 120 Razorback Sucker and Colorado Pikeminnow from 2020. The anterior portion of all specimens will be saved for otolith studies. In order to ensure N_b estimates are not artificially lowered due to limited spatial representation of samples, rare collections were targeted while sites with high larval densities were proportionally reflected in samples. Larval Razorback Sucker are collected during much of the sampling season. Early larval stages (e.g., protolarvae to mesolarvae) will be targeted throughout larval collections under the assumption that these individuals were from recent spawning events; thus, ensuring sampling was representative of the temporal spawning season. Conversely, Colorado Pikeminnow are collected later in the sampling season, making it relatively easy to have those captured individuals reflect the entire seasonal spawning period of adults. Finally, larval sampling downstream of the waterfall located in the San Juan arm of Lake Powell began in 2018. Up to 120 samples for both endangered species will be included to evaluate larval emigration or adult contribution downstream of the waterfall.

Previous N_b estimates have been obtained using multi-locus microsatellite markers; however, the field of population genetics has been shifting to a more recent technology termed “next-generation sequencing” (NGS). This technology has provided a cost-effective means of quantifying massive amounts of genetic data from individuals through the identification of thousands of single nucleotide polymorphism (SNPs). These SNPs are analogous to microsatellite markers, however, SNPs can quantify an order of magnitude more loci compared to microsatellites (i.e., SNP = thousands of loci, microsatellites = 10 – 30 loci), functionally increasing our resolution of genetic variation among individual genomes. This increase in genomic markers not only improves confidence assignments for parental reconstruction (Thrasher et al. 2018), but SNPs also offer benefits over traditional microsatellites methods due to lower error rates and broader genome coverage (Smouse 2010; Hauser et al. 2011). Therefore, N_b estimates for 2020 will be collected using this method along with the methods used in previous years to ensure these results are comparative. The cost of this comparison will include an in-kind contribution for the NGS data collection to: not increase the cost to the Program, ensure 2020

data are comparable to previous years, and provide an avenue to improve data collection while lowering Program costs in the future.

Genomic DNA will be extracted from tissues following standard protocols used at Southwestern ARRC. Microsatellite genotyping will follow the same methods used in previous reports (Diver and Wilson 2018). NGS data will be prepared using double digest Restriction-Site Associated DNA (ddRAD) libraries (Peterson et al. 2012). Restriction digest of 1µg genomic DNA/sample will be performed in 50µl reactions containing 5µl New England BioLabs CutSmart Buffer and 20 units each PstI and MspI. Samples will be digested at 37°C for 18 hours then purified using Agencourt AMPure XP beads (Beckman Coulter, Inc.). Barcoded samples (100 ng DNA each) will be pooled in sets of 48 following Illumina adapter ligation, then size-selected using the Pippin Prep System (Sage Science) to retrieve DNA fragments between 350 and 400 bp in length for Razorback Sucker and 325 to 425 for Colorado Pikeminnow (Bangs et al. 2018; Chafin et al. 2018). Size-selected DNA will be subjected to 10 cycles of PCR amplification using Phusion high-fidelity DNA polymerase (New England Bioscience), according to manufacturer protocols. Four indexed libraries (192 samples) will be pooled per lane for 100bp single-end sequencing on an Illumina HiSeq 4000 (University of Oregon Genomics & Cell Characterization Core Facility). Data will be de-multiplexed and filtered in STACKS (Catchen et al. 2013) to discard reads with uncalled bases or low Phred quality scores (<10), while simultaneously attempting to recover those reads with ambiguous barcodes (=1 mismatched nucleotide). The *de novo* assembly of ddRAD loci will be accomplished in STACKS (Catchen et al. 2013) with clustering parameters being determined by the methods of Rochette & Catchen (2017). Only loci appearing in 95% of individuals will be retained for analysis. SNPs will be filtered to retain one per ddRAD locus. A second researcher will perform a 10% quality assurance/quality control of samples to ensure accuracy.

Both microsatellites and SNPs will be evaluated for linkage disequilibrium and Hardy-Weinberg Equilibrium (Raymond and Rousset 1995; Purcell et al. 2007). For both datasets, N_b will be estimated using the sibship-assignment (SA) method in COLONY version 2.0.4.0 (Jones and Wang 2010). This software uses a maximum likelihood method to estimate relationships among offspring belonging to a single cohort by identifying full and half-sibling families while incorporating genotyping errors (i.e., E_1 and E_2) and allowing for inferences related to the mating strategy of the organism. Analyses will be conducted separately for each year to estimate N_b, the number of adults that contributed at least one offspring, number of sampled offspring produced by each parent, and the number of parental pairs. For both species, male and female polygamy will be assumed and parameter settings (i.e., dioecious, diploid, inbreeding, medium run length, full-likelihood with medium likelihood precision, no sibship prior, and updated allele frequencies) will be maintained across years.

Schedule:

Completion of genetic analysis	March 30, 2021
Final Report	June 30, 2021

Intended Method of Information Dissemination:

Dissemination of the results will include a draft and final report and presentation of project results at the San Juan Researcher's meeting. Data will be submitted per SJRRIP timelines.

FY20 - Detailed Spending Plan

1. PERSONNEL		
A. Laboratory Work		
1 Bio/Geneticist (GS-11; 400 hours -10 pay periods) @ $32.27/hr		$12,908
B. Report Writing		
1 Bio/Geneticist (GS-11; 150 hours -3.75 pay periods) @ $32.27/hr		$4,841
	Subtotal Personnel	$17,749
1. MATERIALS/SUPPLIES		
A. Extractions		$1,711
B. PCR Reactions		$12,232
C. Genetic Analyzer		$4,158
D. Other (tubes, tips, etc.)		$6,513
	Subtotal Supplies	$24,614
	TOTAL	$42,363
Southwestern ARRC Utilities		
-Electrical, (approx. 4,259 KW/h @ 0.34569 per KW/h) =		$1,000
Administrative and Overhead Costs Regional Office @ 3%		$1,301
Project Total FY2020		**$44,664**

Literature Cited

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

Clark, S.R., M.M. Conner, S.L. Durst and N.R. Franssen. 2018. Age-Specific Estimates Indicate Potential Deleterious Capture Effects and Low survival of Stocked Juvenile Colorado Pikeminnow. North American Journal of Fisheries Management. 38(5): 1059-1074.

Diver, T.A. and W.D. Wilson. 2018. Using Molecular Techniques to Determine Effective Number of Breeders (N_b) for Razorback Sucker and Colorado Pikeminnow in the San Juan River. San Juan River Basin Recovery Implementation Program, Albuquerque, New Mexico.

Farrington, M.A., R.K. Dudley, J.L. Kennedy, S.P. Platania, and G.C. White. 2015. Colorado Pikeminnow and Razorback Sucker larval fish survey in the San Juan River during 2014.

Hartl, D.L. and A.G. Clark. 2007. Principles of Population Genetics. Sinauer Associates, Inc., Sunderland, MA.

Hauser, L,, M. Baird, R.A. Hilborn, L.W. Seeb and J.E. Seeb. 2011. An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (*Oncorhynchus nerka*) population. Molecular Ecology Resources. 11: 150-161.

Jones, O.R. and J. Wang. 2010. COLONY: a program for parentage and sibship interference from multilocus genotype data. Molecular Ecology Resources. 10(3): 551-555.

Lambert, Y. 2008. Why should we closely monitor fecundity in marine fish populations. J. Northwest Atlantic Fisheries Science. 41:93-106.

Lauer, T.E., Shroyer, S.M., Kilpatrick, J.M., McComish, T.S., Allen, P.J. 2005. Yellow perch length–fecundity and length–egg size relationships in Indiana waters of Lake Michigan. North American Journal of Fisheries Management. 25:791-796.

Marsh, P.C., T.E. Dowling, B.R. Kesner, T.F. Turner, and W.L. Minckley. 2015. Conservation to Stem Imminent Extinction: The Fight to Save Razorback Sucker *Xyrauchen texanus* in Lake Mohave and Its Implications for Species Recovery. Copeia. 103: 141-156.

Miller, L.M. and A.R. Kapuscinski. 2003. Genetic Guidelines for Hatchery Supplementation Programs. Pages 329-355 *in* E.M. Hallerman, editor. Population genetics: principles and applications for fisheries scientists. American Fisheries Society, Bethesda, Maryland.

Modde, T. and D.B. Irving. Use of multiple spawning sites and season movement by Razorback Sucker in the Middle Green River, Utah. North American Journal of Fisheries Management. 18:318-326.

Peterson, B.K., J.N. Weber, E.H. Kay, H.S. Fisher and H.E. Hoekstra. 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.

Purcell, S., B. Neale, K. Todd-Brown, M.A. Ferreira, D. Bender, J. Maller, P. Sklar, P.I. de Bakker, M.J. Daly, and P.C. Sham. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics. 81(3): 559-575.

Raymond, M., and F. Rousset. 1995. GENEPOP (version 1.2): population genetics software for exact test and ecumenicism. Journal of Heredity 86:248-249. Available: http://genepop.curtin.edu.au/. (April 2012).

Rochette, N.C and J.M. Catchen. 2017. Deriving genotypes from RAD-seq short-read data using Stacks. Nature Protocols. 12: 2640.

Smouse, P.E. 2010. How many SNPs are enough?. Molecular Ecology. (7):1265-6.

Thrasher, D.J., B.G. Butcher, L. Campagna, M.S. Webster, and I.J. Lovette. 2018. Double-digest RAD

sequencing outperforms microsatellite loci at assigning paternity and estimating relatedness: a proof of concept in a highly promiscuous bird. Molecular Ecology Resources. 18(5): 953-965.

U.S. Fish and Wildlife Service. 2005. An augmentation plan for Razorback Sucker in the San Juan River. U.S. Fish and Wildlife Service, Colorado River Fishery Project, Grand Junction, Colorado.

U.S. Fish and Wildlife Service. 2015. San Juan River Razorback Sucker *Xyrauchen texansu* & Colorado Pikeminnow *Ptychocheilus Lucius* population augmentation: 2014. U.S. Fish and Wildlife Service, New Mexico Fish and Wildlife Conservation Office, Albuquerque, New Mexico.

Waples, R.S., G. Luikart, J.R. Faulkner, and D.A. Tallmon. 2013. Simple life history trains explain key effective population size ratios across diverse taxa. Proceedings of the Royal Society B: Biological Sciences. 280: 20131339.

Waples, R.S., T. Antao, and G. Luikart. 2014. Effects of overlapping Generations on Linkage Disequilibrium Estimates of Effective Population Size. Genetics. 197: 769-780.

Response to Comments

SOW20-2 - Using molecular techniques to determine effective number of breeders (N_b) for Razorback Sucker and Colorado Pikeminnow in the San Juan River

Responses to reviewer comments are presented in bold text.

Wayne Hubert: "This is a novel approach to determine effective population size that is worthy of continuation. A concern is if the sample size for an individual species (n = 120) in a given year is sufficient and if the sample approximates a random sample of the population of larval fish in the SJR. Can a sample of 120 larvae effectively represent the spatial and temporal variation in the population of larvae in a given year? Estimates of numbers of breeders may be biased by insufficient sampling. It would be wise for the research from the Southwest Resources Aquatic Resources and Recovery Center to collaborate with ASIR (the organization sampling larval fish) and statisticians with expertise in sampling to assess sampling needs to obtain an accurate and relative precise estimates of the number of breeders of a species within a given year."

Thanks to Dr. Hubert for his comments. We hope our responses help alleviate some of his concerns. We agree appropriate sample sizes for these types of assessments will be extremely important for developing accurate estimates. For each year, we randomly selected 120 fish from all of ASIR's annual collections; this was weighted by densities at each site (i.e., backwater at river mile X) while specifically including rare samples. Our genetic sampling is therefore limited to the temporal and spatial sampling of larval fish collections, thus, we are estimating the effective number of breeders that contributed to the larval fish that were collected in the field. While, we agree that there remains the possibility that cohorts of Razorback Sucker and Colorado Pikeminnow that are not sampled in the field (spatially and/or temporally) would lower our estimate, we are unable to factor for that in our genetic sampling if those cohorts are missed in field. Nonetheless, because ASIR's annual sampling is relatively constant, we are intending these estimates to be used as an index of annual reproductive output rather than an estimate of the actual number of all the breeding adults in the San Juan Basin. Furthermore, this estimate factors for variance in reproductive success. Thus, if a few individuals contributed a majority of the larvae sampled in the field, then we would predict that biased contribution would lower our estimate; this is a genetic estimate for quantifying the number of adults that effectively contributed to a single cohort and is not a count of spawning adults.

Although we are taking a random sample from the annual larval fish collections, our data also suggest that in most years, a sample size of 120 is adequate for the endangered fishes we are targeting. Other studies have indicated that a sample size that is close to the true N_b is adequate (England et al. 2006; Wang 2016; Sánchez-Montes et al. 2017; Bacles et al. 2018). Our data also suggest that increasing our sample size from the larvae collected in the field is not necessarily going to change our estimates due to the high level of relatedness among the larval fish sampled. In other words, increasing our sample size generally increases the number of siblings (whose parents have already been included in the estimate), and therefore would not change the overall N_b estimate. However, we are aware that if the number of breeders from our samples start to increase substantially, then our sample sizes will need to be increased. But at this point, we think the minimum sample size of 120 should be adequate to estimate N_b for the larval fish that are collected annually in the field. If the Program would like to test if our samples size is sufficient, we would

recommend both increasing field sampling of larval fishes over space and time and increasing the number of fish included in genetic sampling; however, the cost of this increase must also be considered.