Lower Colorado River Multi-Species Conservation Program

Balancing Resource Use and Conservation

Development of SNP Markers for Sex Determination, Parentage Assessment, and Population Genetics of Razorback Suckers

2018 Annual Report

U.S. DEPARTMENT OF THE INTERIOR
BUREAU OF RECLAMATION

May 2019

Work conducted under LCR MSCP Work Task C40

Lower Colorado River Multi-Species Conservation Program Steering Committee Members

Federal Participant Group

Bureau of Reclamation
U.S. Fish and Wildlife Service
National Park Service
Bureau of Land Management
Bureau of Indian Affairs
Western Area Power Administration

Arizona Participant Group

Arizona Department of Water Resources
Arizona Electric Power Cooperative, Inc.
Arizona Game and Fish Department
Arizona Power Authority
Central Arizona Water Conservation District
Cibola Valley Irrigation and Drainage District
City of Bullhead City
City of Lake Havasu City
City of Mesa
City of Somerton
City of Yuma
Electrical District No. 3, Pinal County, Arizona
Golden Shores Water Conservation District
Mohave County Water Authority
Mohave Valley Irrigation and Drainage District
Mohave Water Conservation District
North Gila Valley Irrigation and Drainage District
Town of Fredonia
Town of Thatcher
Town of Wickenburg
Salt River Project Agricultural Improvement and Power District
Unit "B" Irrigation and Drainage District
Wellton-Mohawk Irrigation and Drainage District
Yuma County Water Users' Association
Yuma Irrigation District
Yuma Mesa Irrigation and Drainage District

Other Interested Parties Participant Group

QuadState Local Governments Authority
Desert Wildlife Unlimited

California Participant Group

California Department of Fish and Wildlife
City of Needles
Coachella Valley Water District
Colorado River Board of California
Bard Water District
Imperial Irrigation District
Los Angeles Department of Water and Power
Palo Verde Irrigation District
San Diego County Water Authority
Southern California Edison Company
Southern California Public Power Authority
The Metropolitan Water District of Southern California

Nevada Participant Group

Colorado River Commission of Nevada
Nevada Department of Wildlife
Southern Nevada Water Authority
Colorado River Commission Power Users
Basic Water Company

Native American Participant Group

Hualapai Tribe
Colorado River Indian Tribes
Chemehuevi Indian Tribe

Conservation Participant Group

Ducks Unlimited
Lower Colorado River RC&D Area, Inc.
The Nature Conservancy

Lower Colorado River
Multi-Species Conservation Program

Development of SNP Markers for Sex Determination, Parentage Assessment, and Population Genetics of Razorback Suckers

2018 Annual Report

Prepared by:

Thomas E. Dowling,[1] Joel B. Corush,[1] and Trevor J. Krabbenhoft[2]

[1] Department of Biological Sciences
5047 Gullen Mall
Wayne State University
Detroit, MI 48202
313-577-3020
thomas.dowling@wayne.edu

[2] Department of Biological Sciences
Research and Education in Energy, Environment, and Water (RENEW) Institute
University at Buffalo
Buffalo, NY 14260-4610
716-645-4934
tkrabben@buffalo.edu

Lower Colorado River
Multi-Species Conservation Program
Bureau of Reclamation
Lower Colorado Region
Boulder City, Nevada
http://www.lcrmscp.gov

May 2019

Dowling, T.E., J.B. Corush, and T.J. Krabbenhoft. 2019. Development of SNP Markers for Sex Determination, Parentage Assessment, and Population Genetics of Razorback Suckers, 2018 Annual Report. Submitted to the Lower Colorado River Multi-Species Conservation Program, Bureau of Reclamation, Boulder City, Nevada, by Wayne State University, Detroit, Michigan, under Agreement No. R16AP00174.

ACRONYMS AND ABBREVIATIONS

bp	base pairs
contig	contiguous block of DNA
DNA	deoxyribonucleic acid
F_{IS}	variation among individuals within populations (standard inbreeding coefficient)
FLSU	flannelmouth sucker(s) (*Catostomus latipinnis*)
GT-seq	Genotyping in Thousands by sequencing
H_o	observed heterozygosity
max	maximum
min	minimum
MM	male-specific marker
opt	optimum
N	sample size
NGS	Next generation sequencing
PCR	polymerase chain reaction
PIT	passive integrated transponder
RAD-seq	restriction site associated DNA sequencing
RASU	razorback sucker(s) (*Xyrauchen texanus*)
SNP	single-nucleotide polymorphism

Symbols

≈	approximately
>	greater than
<	less than
%	percent

CONTENTS

Tables

Figures

INTRODUCTION

Management of razorback suckers (*Xyrauchen texanus*; RASU) in Lake Mohave is changing, fueled by technological advances. Advances in passive integrated transponder (PIT) tag scanning (Wisenall et al. 2015) have made it possible to remotely assess activity and abundance, potentially reducing the need for frequent monitoring of the population with more invasive methods (e.g., netting). Unfortunately, PIT tag scanning has some shortcomings. Individual fish are not handled after stocking; therefore, it is not possible to assess condition or identify the sex or genotypes of specific individuals.

Realization that predation pressure will be difficult to overcome has led to the development of a plan (Minckley et al. 2003) in which predator-free environments (i.e., backwaters) are used to establish stable, recruiting populations of RASU. PIT tag scanning allows us to obtain regular estimates of abundance in such backwaters and to assess patterns of mortality over time, and molecular methods have been used to estimate the number of breeding individuals each year (Dowling et al. 2014). Molecular methods have been used to genotype stocked adults and their progeny in such backwaters, making it possible to characterize patterns of reproduction and to estimate reproductive success of specific individuals (Dowling et al. 2017). These environments are critical, as they hold refuge populations that will serve as a source for adults to be repatriated to natural settings (e.g., Lake Mohave) and they avoid many of the issues associated with selection in hatchery environments (Christie et al. 2012, 2016) while providing locations where stable, self-sustaining populations of RASU can contribute to conservation of this species.

A natural consequence of a successful backwater is recruitment of larvae into the reproductive population. Microsatellites have been a useful tool for identifying parentage in ephemeral backwaters; however, analysis of the permanent backwater at Yuma Cove indicates that there is insufficient statistical power to assign parentage when sex and genotypes of parents are unknown, as for newly recruited individuals (Dowling et al. 2017). In addition, the platform that has been used to generate microsatellite data (LiCor) has become obsolete and is no longer supported, requiring a change in approach.

Advances in molecular technology may provide a solution to these two issues. Next generation sequencing (NGS) approaches (Mardis 2008; Shendure and Ji 2008) allow researchers to survey the entire genome for variable markers (e.g., single nucleotide polymorphisms [SNPs]) that can be used to identify sex of unknown individuals (Fowler and Buonaccorsi 2016; Gamble 2016; Pan et al. 2019) and to characterize population genetic and reproductive characteristics (Andrews et al. 2018; Hauser et al. 2011).

This report covers the second year of a 3-year project in which NGS is used to identify deoxyribonucleic acid (DNA)-based markers to: (1) characterize population genetic parameters used in genetic monitoring of several large refuge populations (e.g., Reaches 1–3), (2) quantify reproductive success in backwater populations, and (3) identify the sex of individuals from small amounts of tissue (e.g., a fin clip).

These markers will replace previous methods for quantifying patterns of genetic variation, providing more statistical power than previously available. This is especially important in studies of backwater populations where complexities associated with incorporation of progeny of stocked adults and the presence of uncharacterized reproductive individuals requires greater statistical power for determining reproductive success of individuals (stocked adults and their progeny) reproducing in these ponds. Identification of individual sex from fin clips is also an essential need for monitoring efforts on Lake Mohave (and for any other RASU monitoring efforts) as management agencies move more toward hands-off approaches. This approach has been used successfully for development of sex-specific markers in other fishes (e.g., Fowler and Buonaccorsi 2016; Gamble 2016; Pan et al. 2019), including another cypriniform (zebrafish [*Danio rerio*]; Anderson et al. 2012).

Use of NGS to develop DNA-based markers is necessary for effective studies of RASU. Given issues with statistical power and the limited ability to generate microsatellite data in the future, NGS is an ideal approach, as it can be used to develop hundreds of genetic markers. This approach is routinely used during management of salmonid populations (Hauser et al. 2011, Larson et al. 2014; Storer et al. 2012) and should provide sufficient statistical power for questions associated with the genetic characterization of permanent backwater ponds.

METHODS

DNA samples from a total of 192 RASU and flannelmouth suckers (*Catostomus latipinnis*, FLSU) were used to construct a restriction site associated DNA sequencing (RAD-seq) library using NGS described in Dowling and Krabbenhoft 2017). The library consisted of 64 adult male and 64 adult female RASU that were selected from adults that were used in previous studies of reproductive success in backwaters ponds, larvae (N = 40) from 4 selected pairs of these adults, and 24 adult FLSU. These individuals were selected to develop SNP and sex-specific markers for: (1) monitoring population genetic parameters in populations from Reaches 1–3 and backwaters, (2) determining parentage to estimate reproductive success of individuals stocked and born into backwaters, and (3) assessing the impact of hybridization between RASU and FLSU, especially in Reach 1 (Lake Mead), where it has been previously documented.

Marker Selection

RAD-seq loci generated last year (see Dowling and Krabbenhoft 2017 for details) were computationally filtered to provide a reduced subset of markers for Genotyping in Thousands by sequencing (GT-seq). First, the STACKS program "*populations*" (Catchen et al. 2013) was used to identify SNPs in RASU adults. A series of filtering steps were then conducted to identify a panel of 300 candidate SNPs for population genomics and parentage analysis. First, only RAD-seq tags present in > 85% of individuals were kept. We also only kept RAD-seq tags with 2–5 SNPs per tag. Next, SNPs with observed heterozygosity, H_o, > 0.45 were removed to reduce the chances of SNPs resulting from alignment of non-homologous duplicates of the same locus ("paralogs"). SNPs with H_o < 0.20 were also removed because the small amount of variation limits their utility for parentage analysis. SNPs with an inbreeding coefficient, F_{IS}, < -0.10 and > 0.40 were also removed due to their potential for including genotyping or assembly errors. Finally, the pipeline developed by McKinney et al. (2017) was used to identify loci likely to represent paralogs, based on a combination of high heterozygosity and distorted read-ratio counts. For heterozygotes, alleles should follow a 1:1 expected ratio in singletons (non-paralogs), while duplicated loci (paralogs) often deviate from this ratio. This approach was used to identify and remove potential paralogs following thresholds developed in McKinney et al. (2017). Loci with z-scores < -3 and > 3 were removed. Once a list of loci was generated, we reran *populations* and removed RAD-seq tags without at least 1 SNP with minor allele frequency > 0.2. Remaining loci were further filtered to remove any loci that contained (1) insertions or deletions of base pairs (bp), (2) ambiguous bp, and (3) stretches of the same bp repeated six times or more (e.g., AAAAAA, TTTTTT, CCCCCC, or GGGGGG). Finally, loci containing SNPs within the first or last 20 bp were eliminated, as those SNPs were too close to the primer and would not be included in resulting sequences. Of the loci that passed filtering, a set of 300 was randomly chosen for downstream population and parentage analysis.

Primer Development

To obtain forward and reverse amplification primers for these 300 loci, qualifying sequences were submitted to BatchPrimer3 v1.0 (https://probes.pw.usda.gov/batchprimer3/) (You et al. 2008). With the exception of minimum, optimum, and maximum product sizes (min = 100 bp, opt = 200 bp, and max = 400 bp, respectively), all default settings were used. Amplification is more complex than normal, with the first amplification reaction (polymerase chain reaction [PCR]) (PCR1) designed to produce sequences for each of the 300 loci from every individual. Each primer provided by BatchPrimer3 for PCR1 included an adapter sequence that is necessary for the next round of amplification (PCR2) (figure 1).

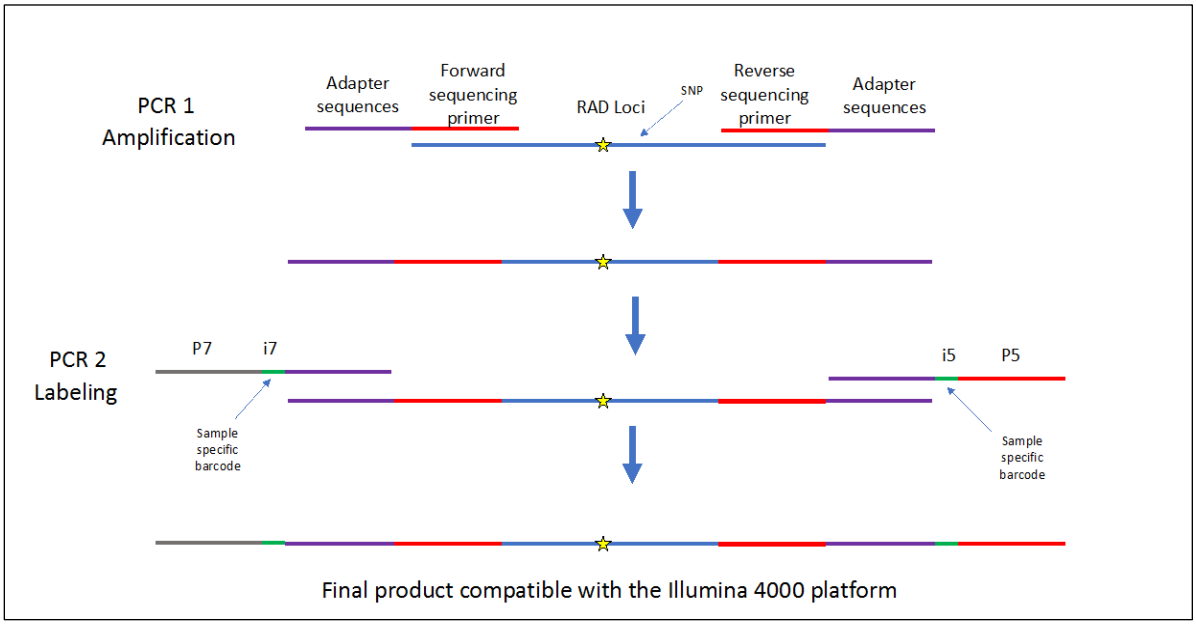

Figure 1.—Breakdown of amplification of DNA in GT-seq.
Primers used for PCR1 (amplification of targeted DNA) and PCR2 (labeling amplified DNA based on individual) are noted, including the Illumina specific adapters (P5 and P7) and the individual's specific barcodes (i5 and i7).

Because all samples will ultimately be pooled into one tube for sequencing, barcodes must be added to allow for sorting of sequences by individuals as well as sample plate. This was achieved by performing a second amplification (PCR2) that utilized primers with unique individual specific DNA barcodes (i5 and i7; see figure 1) that were concatenated to the Illumina platform adapters (P5 and P7) as well as the adapter sequences from PCR1. In PCR2, the Illumina platform adapters are required for the sequencer to read the sequences, the specific DNA barcodes allow for assignment of sequences to specific individuals for post sequencing analysis, and the adapter sequences are required to link the barcodes to the amplified loci from PCR1. The final product contains an Illumina specific sequence, individual specific barcodes, an adapter region, and a segment of DNA containing two to five SNPs (see figure 1).

Genome Sequence

Sequence generation and assembly of the RASU genome has been initiated. An adult male RASU from the Lake Mead hatchery was selected for DNA extraction (due to expected lower levels of variation) and an initial sequencing run was conducted using Illumina short-read sequencing (150 base pair-long fragments, obtained from both directions). A total of 100 billion base pairs of high-quality DNA sequence was generated in this preliminary dataset. This sequence data was used to computationally predict the previously unknown genome size of RASU to

be 1,782,172,447 bp using KmerGenie software version 1.7051 (Chikhi and Medvedev 2014). This estimate was used to select correct parameter settings needed for generating a final assembly. Long-read Oxford Nanopore sequences were generated on the MinION and GridION platforms using a total of nine flow cells (version R9.4.1) as of December 2018. Long reads are being used for the primary RASU genome assembly using the same adult male RASU used in Illumina sequencing. In genome assembly, these long fragments of DNA sequence are computationally aligned to one another to create longer contiguous blocks of DNA ("contigs"). Long-read sequences are necessary for correctly assembling complex regions of the genome, such as long sequence repeat regions like microsatellites. Genome assembly was conducted using minimap2 and miniasm software (Li 2016, 2018) and visualized using Bandage (Wick et al. 2015). In the initial genome assembly presented here, only raw sequences > 10,000 bp were used (due to computational time required). Future assembly will include all raw sequences > 1,000 bp and additional sequencing depth (e.g., ≈60X target).

Male-Specific Markers

Methods for identification and analysis of three male-specific markers reported here were described in Dowling and Krabbenhoft (2017). These three loci were obvious outliers in the RAD-seq analysis, found in a large number of adult RASU males and completely absent in the RASU female adult RAD-seq data. Because RAD-seq often results in missing data, especially in low coverage individuals, results were validated, and primers were developed and optimized for three of these apparent male-specific markers (MM1, MM2, and MM3). Samples from the RAD-seq library were characterized by PCR with these three sets of primers, yielding results that were similar to each other (table 1) and identical to expectations from RAD-seq library sequences.

Table 1.—Results from assays of all individuals with the three male-specific markers

Sample	Male marker 1			Male marker 2			Male marker 3		
	Present	Absent	Total	Present	Absent	Total	Present	Absent	Total
RASU males	50	12	62	52	12	64	51	12	63
RASU females	0	63	63	0	64	64	0	64	64
RASU larvae	11	16	27	16	19	35	8	17	25
FLSU	0	22	22	11	7	18	0	20	20
Totals	**61**	**113**	**174**	**79**	**102**	**181**	**59**	**113**	**172**

RESULTS AND DISCUSSION

SNP Marker Development

NGS resulted in a total of 323,277,973 DNA sequence reads across the 3 RAD-seq libraries, of which 90.6% were very high quality (i.e., > Q30). Individual RAD-seq libraries ranged from ≈43 million reads in library 1 to ≈2 million reads in library 2. To avoid issues with low coverage (a common problem with RAD-seq), 20 of the 128 adult RASU with the lowest coverage were excluded, leaving 56 females and 52 males for downstream analyses. Coverage in these remaining samples was high (e.g., 14X median coverage), allowing for generation of SNP and sex-specific markers.

Using the optimal parameters, a raw catalog containing 147,650 RAD-seq loci was recovered, of which 84,991 RAD-seq loci were polymorphic and contained 316,639 SNPs (3.64 SNPs per polymorphic RAD-seq locus). Of these 316,639 SNPs, a total of 29,362 SNPs across 18,501 different RAD-seq loci passed initial filters for use as population genetic markers, with mean observed heterozygosity = 0.26. After additional filtering based on H_o, F_{IS}, and the McKinney paralog pipeline, we identified 7,315 markers that met predetermined criteria. Additional filtering of these markers yielded 3,784 loci, from which 300 loci were randomly selected for use in genotyping.

Four sets of primers were produced and tested through the amplification and labeling step. While only primer sets 2, 3, and 4 show bands after amplification (figure 2A), the appearance of bands after labeling indicates that amplification did work in all four primers sets (figure 2B). These results suggest that the primer design was successful in both amplifying the targeted loci and uniquely labeling each individual with a tag that could also be sequenced on the Illumina machine.

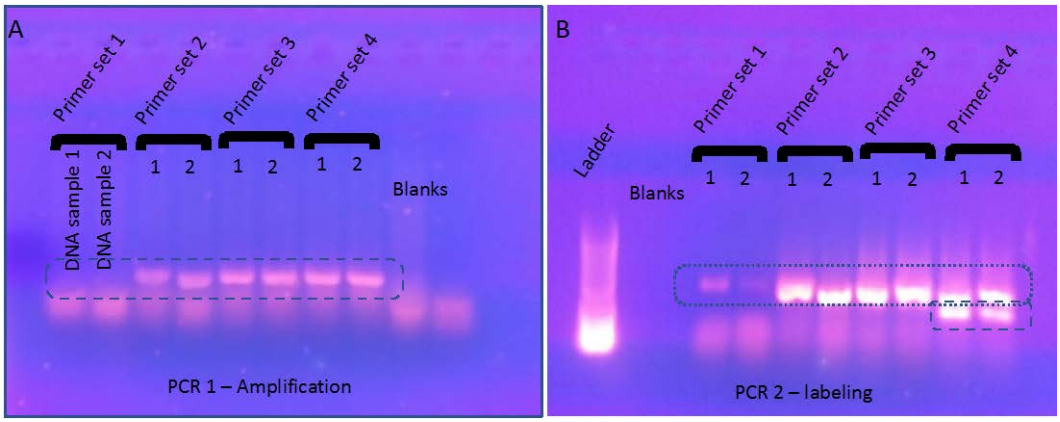

Figure 2.—Visualization of (A) amplification product and (B) labeling product. The dashed line box indicates amplification product and the dotted line box indicates amplified product with species barcode and Illumina adapters.

Based on the original GT-seq protocol (Campbell et al. 2015), final concentration of post-barcoding (PCR2, see figure 1) DNA from seven test samples was observed using the Qubit® 2.0 Fluorometer. Concentrations were then sent to Novogene (company that will perform sequencing) to verify that the final product would be of sufficient quantity for sequencing. Based on test sample quantities, our test protocol produced an adequate amount of DNA for sequencing.

Genome Sequencing

Long-read Nanopore sequencing resulted in 69,833,188,877 bp of sequence across 24,161,066 reads. This equates to an estimated 39X coverage of the genome. Individual raw DNA sequences ranged from < 1,000 bp to over 300,000 bp in length. A total of 12,821,383 sequences were > 1,000 bp in length. A preliminary genome assembly using only raw sequences > 10,000 bp in length resulted in a total assembly length of 2,034,512,007 bp, or approximately 115% of the predicted genome size. At present, 50% of the assembly is in contigs > 965,196 bp or larger ("N50"; a common benchmark for genome assembly quality). The longest assembled fragment is currently 6,061,287 bp, representing a significant piece of a typical chromosome (e.g., chromosomes range from 37–78 million bp in length in zebrafish). The current assembly has not yet been filtered for non-target DNA (e.g., bacterial contamination, microbiome sequences, etc.), and future addition of remaining long-read and short Illumina sequences is expected to improve the quality and contiguity of this assembly beyond this first pass assembly (i.e., final target of 100 assembled chromosomes based on typical karyotypes of sucker species).

Once a more complete genome assembly is completed, GT-seq and sex-specific loci will be mapped to the genome to refine marker design (e.g., improve specificity of PCR primers) and to ensure that markers are independent (not located on the same stretch of chromosome) and adequately cover the RASU genome.

The genome assembly will also facilitate the search for more informative sex-specific markers (e.g., those that are present in 100% of males and 0% of females). If present, these loci should allow for unambiguous identification of the sex of specific individuals based on genetic material. Genes responsible for sex determination are often physically proximate, occurring on the same chromosome (Andrews et al. 2016). Location of the three male-specific markers was mapped to the preliminary assembly (figure 3). This preliminary result places these markers on decoupled fragments of DNA, possibly indicating that genes involved in sex determination are not found on sex chromosomes but are scattered across the genome. Alternatively, future refinements of the assembly may result in these three DNA fragments being assembled onto the same chromosome (i.e., a sex chromosome).

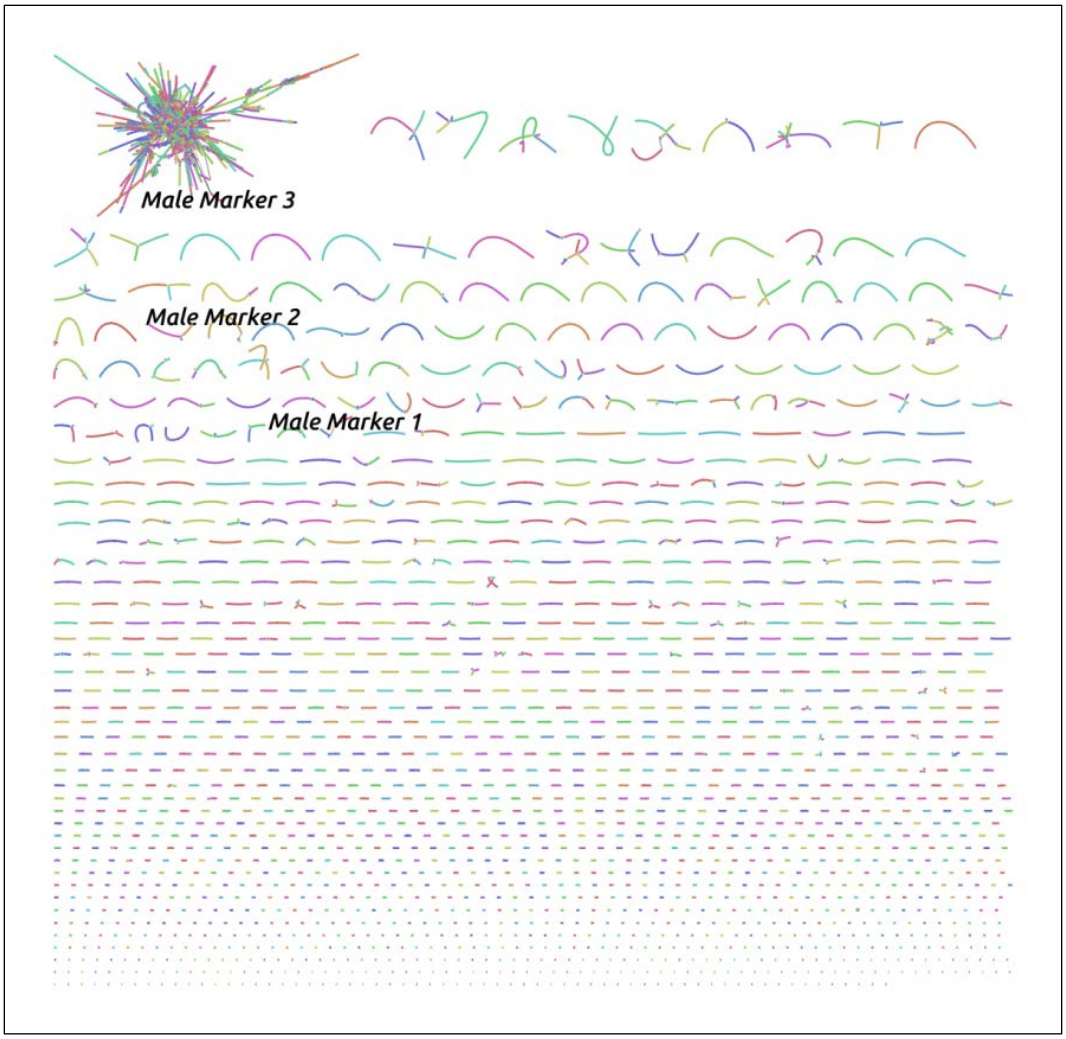

Figure 3.—Graph-based representation of a preliminary RASU genome assembly using long sequence reads.
Lines represent assembled DNA fragments (contigs). The tangled ball in the upper left may represent assembly ambiguities or complex repeat regions. Further sequencing depth is needed to disentangle this portion of the genome.

Male-Specific Markers

To assess the utility of the male-specific markers, 139 samples were drawn at random by B. Kesner (Marsh & Associates, LLC) from adults collected by trammel netting during the period of 1998–2017 and characterized with MM1-3. Of these 139 individuals, 4 DNA samples would not amplify; therefore, sex was obtained from 135 individuals. Once the sex of each individual was determined using these markers, it was compared to the sex identified at capture (68 males and 71 females), yielding 11 discrepancies of the 135 samples (table 2). Nine of these 11 discrepancies were males identified as females using the male-specific markers, with the last 2 females that were identified as males. It is important

Table 2.—Comparison of phenotypic and genetic identifications for 139 individuals randomly drawn from samples collected from Lake Mohave in years 1998–2017

(Samples are separated by year stocked. Sex was identified in the field using morphology and, in the lab, using male-specific markers, and are denoted as "Phenotype" or "P" and "Genotype" or "G", respectively.)

	Phenotype			Phenotype-genotype comparison				
Year stocked	M[1]	F[2]	Total	P = G[3]	Number of M as F[4]	Number of F as M[5]	?[6]	Total
1992	3	0	3	3	0	0	0	3
1993	3	0	3	1	1	0	1	3
1994	1	1	2	2	0	0	0	2
1995	0	1	1	1	0	0	0	1
1996	3	2	5	5	0	0	0	5
1997	3	0	3	3	0	0	0	3
1998	5	2	7	6	1	0	0	7
1999	9	2	11	8	3	0	0	11
2000	6	0	6	6	0	0	0	6
2001	1	3	4	4	0	0	0	4
2002	2	0	2	2	0	0	0	2
2003	2	1	3	3	0	0	0	3
2004	10	2	12	11	1	0	0	12
2005	3	1	4	4	0	0	0	4
2006	3	4	7	7	0	0	0	7
2007	2	2	4	2	1	1	0	4
2008	1	1	2	2	0	0	0	2
2009	2	14	16	14	0	1	1	16
2010	3	9	12	12	0	0	0	12
2011	2	6	8	7	0	0	1	8
2012	1	2	3	3	0	0	0	3
2013	0	5	5	4	0	0	1	5
2014	1	6	7	7	0	0	0	7
2015	1	1	2	1	1	0	0	2
2016	0	2	2	2	0	0	0	2
2017	1	4	5	4	1	0	0	5
Total	68	71	139	124	9	2	4	139

[1] M = Number of males.
[2] F = number of females.
[3] P = G = Samples with consistent sex identification.
[4] Number of M as F = Number of phenotypic males confirmed as misidentified using male markers.
[5] Number of F as M = Number of phenotypic females confirmed as misidentified using male markers.
[6] ? = Samples with poor DNA (would not amplify).

to note that the only two phenotypic females that were both identified as males by these markers had both lost their PIT tags in the field and had to be retagged, raising the possibility that there was some mixup at some time in their history. The proportion of males misidentified as females (9 of 68 = 13.2%) from Lake Mohave is similar to numbers identified from the RAD-seq library (11 of 64 = 17.2%; Dowling and Krabbenhoft 2017), as was the total proportion of misidentified individuals here (11 of 135 = 8.2%) as compared to results from the RAD-seq library (12 of 128 = 9.4%; Dowling and Krabbenhoft 2017). Therefore, these markers performed comparably for backwater and lake sampled individuals whose sexes were known or not known prior to analysis, respectively.

Inspection of the data identified a significant trend in sex ratios (table 2), with males stocked more frequently prior to 2008 (54 of 73 individuals) and females more common in 2008 and later (52 of 66 individuals) (G-test, $G = 40.3$, $df = 1$, $P < 0.0001$). It is unclear why this pattern exists, warranting further investigation.

Because of the unusual variation in sex ratio observed in samples from Lake Mohave, sex was determined for 96 randomly sampled individuals from 617 RASU reared at Achii Hanyo and stocked in 2016. This sample was biased, as 55 of the 96 individuals were females, yielding a sex ratio of 1.34 females for each male. As expected, females were significantly larger than males (436 millimeters versus 412 millimeters total length, T-test, $T = 6.7$, $P < 0.0001$). The biased sex ratio is also not surprising, as larger individuals were selected for stocking; therefore, one would expect the observed bias.

To further assess the sex ratio in Lake Mohave, 48 larvae were randomly sampled from 5 different locations and 4 months, covering the major spawning aggregations and spawning period (table 3). Using the three male-specific markers, sex was identified for 47 of 48 individuals. Sex ratios were skewed in two samples; however, sample sizes were too small to perform statistical analyses on individual temporal samples. Overall, this approach identified 23 males and 24 females (one sample did not work), yielding a 1:1 sex ratio.

Table 3.—Identifications of RASU larvae with male-specific markers from samples collected in 2018

Location	Date	Male	Female	Total
Above Willow Beach	1/22/2018	4	4	8
Yuma Cove	1/25/2018	4	4	8
Tequila Cove	1/30/2018	1	6	7
Nine Mile	2/12/2018	6	2	8
Yuma Cove	3/19/2018	4	4	8
Red Tail Cove	4/10/2018	4	4	8
Total		**23**	**24**	**47**

LITERATURE CITED

Anderson, J.L., A. Rodríguez Marí, I. Braasch, A. Amores, and P. Hohenlohe. 2012. Multiple sex-associated regions and a putative sex chromosome in zebrafish revealed by RAD mapping and population genomics. PLoS ONE 7(7): =e40701. doi: 10.1371/journal.pone.0040701.

Andrews, K.R., J.M. Good, M.R. Miller, G. Luikart, and P.A. Hohenlohe. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nature Reviews Genetics 17:81–92.

Andrews, K.R., J.R. Adams, E.F. Cassirer, R.K. Plowright, C. Gardner, M. Dwire, and L.P. Waits. 2018. A bioinformatic pipeline for identifying informative SNP panels for parentage assignment from RAD seq data. Molecular Ecology Resources 18:1263–1281.

Campbell, N.R., S.A. Harmon, and S.R. Narum. 2015. Genotyping-in-thousands by sequencing (GT-seq): a cost effective SNP genotyping method based on custom amplicon sequencing. Molecular Ecology Resources 15:855–867.

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.

Chikhi, R. and P. Medvedev. 2014. Informed and automated *k*-mer size selection for genome assembly. Bioinformatics 30:31–37.

Christie, M.R., M.L. Marine, R.A. French, and M.S. Blouin. 2012. Genetic adaptation to captivity can occur in a single generation. Proceedings of the National Academy of Sciences 109:238–242.

Christie, M.R., M.L. Marine, S.E. Fox, R.A. French, and M.S. Blouin. 2016. A single generation of domestication heritably alters the expression of hundreds of genes. Nature communications 7, Online Early, doi: 10.1038/ncomms10676.

Dowling, T.E. and T.J. Krabbenhoft. 2017. Development of SNP Markers for Sex Determination, Parentage Assessment, and Population Genetics of Razorback Sucker, 2017 Annual Report. Submitted to the Lower Colorado River Multi-Species Conservation Program, Bureau of Reclamation, Boulder City, Nevada, by Wayne State University, Detroit, Michigan, under Agreement No. R16AP00174.

Dowling, T.E., T.F. Turner, E.W. Carson, M.J. Saltzgiver, D. Adams, B. Kesner, and P.C. Marsh. 2014. Genetic and demographic responses to intensive management of razorback sucker in Lake Mohave, Arizona-Nevada. Evolutionary Applications 7:339–354.

Dowling, T.E., B.R. Kesner, P.C. Marsh, and T.F. Turner. 2017. Genetic and Demographic Studies to Guide Conservation Management of Razorback Suckers in Off-Channel Habitats, 2010–2015. Submitted to the Lower Colorado River Multi-Species Conservation Program, Bureau of Reclamation, Boulder City, Nevada, by Wayne State University, Detroit, Michigan; Marsh & Associates, LLC, Tempe, Arizona; and University of New Mexico, Albuquerque, New Mexico, under Agreement No. R13AC300016.

Fowler, B.L. and V.P. Buonaccorsi. 2016. Genomic characterization of sex-identification markers in *Sebastes carnatus* and *S. chrysomelas* rockfishes. Molecular Ecology 25:2165–2175.

Gamble, T. 2016. Using RAD-seq to recognize sex-specific markers and sex chromosome systems. Molecular Ecology 25:2114–2116.

Hauser, L., M. Baird, R.A.Y. 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(s1):150–161.

Larson, W.A., J.E. Seeb, C.E. Pascal, W.D. Templin, and L.W. Seeb. 2014. Single-nucleotide polymorphisms (SNPs) identified through genotyping-by-sequencing improve genetic stock identification of Chinook salmon (*Oncorhynchus tshawytscha*) from western Alaska. Canadian Journal of Fisheries and Aquatic Sciences 71:698–708.

Li, H. 2016. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. Bioinformatics 32:2103–2110. https://doi.org/10.1093/bioinformatics/btw152

_____. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094-3100. doi: 10.1093/bioinformatics/bty191.

Mardis, E.R. 2008. Next-generation DNA sequencing methods. Annual Review of Genomics and Human Genetics 9:387–402.

McKinney, G.J., R.K. Waples, L.W. Seeb, and J.E. Seeb. 2017. Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. Molecular Ecology Resources 17:656–669.

Minckley, W.L., P.C. Marsh, J.E. Deacon, T.E. Dowling, P.W. Hedrick, W.J. Matthews, and G. Mueller. 2003. A conservation plan for lower Colorado River native fishes. BioScience 53:219–234.

Pan, Q., R. Feron, A. Yano, R. Guyomard, E. Jouanno, E. Vigouroux, M. Wen, J.-M. Busnel, J. Bobe, J.-P. Concordet, H. Parrinello, L. Journot, C. Klopp, J. Lluch, C. Roques, J. Postlethwait, M. Schartl, A. Herpin, and Y. Guiguen. 2019. Identification of the master sex determining gene in Northern pike (*Esox lucius*) reveals restricted sex chromosome differentiation. bioRxiv 549527; doi: https://doi.org/10.1101/549527.

Shendure, J. and H. Ji. 2008. Next-generation DNA sequencing. Nature Biotechnology 26:1135–1145.

Storer, C.G., C.E. Pascal, S.B. Roberts, W.D. Templin, L.W. Seeb, and J.E. Seeb. 2012. Rank and order: evaluating the performance of SNPs for individual assignment in a non-model organism. PloS ONE 7:e49018.

Wick, R.R., M.B. Schultz, J. Zobel, and K.E. Holt. 2015. Bandage: interactive visualisation of *de novo* genome assemblies. Bioinformatics 31:3350–3352.

Wisenall, J.B., B.R. Kesner, C.A. Pacey, and P.C. Marsh. 2015. Demographics and Monitoring of Repatriated Razorback Sucker in Lake Mohave, 2011–2014. Final report submitted to Bureau of Reclamation, Boulder City, Nevada, by Marsh & Associates, LLC, Tempe, Arizona. 54 p.

You, F.M., N. Huo, Y.Q. Gu, M.C. Luo, Y. Ma, D. Hane, G.R. Lazo, J. Dvorak, and O.D. Anderson. 2008. BatchPrimer3: a high throughput web application for PCR and sequencing primer design. BMC Bioinformatics 9:253.