**Interim Progress Report for PRRIP project "Resolving Pallid Sturgeon Species Identification, Demographics and Hybridization using GT-Seq"**

Edward J. Heist
Center for Fisheries, Aquaculture & Aquatic Sciences
Southern Illinois University Carbondale
December 30, 2021

This report details progress made during the 2021 calendar year with expected progress for 2022.

**Student Recruiting and Training** – Following an open search for an appropriate Ph.D. student to Dr. Heist identified Mr. Junman Huang as the best candidate among more than 20 who applied for the position. Mr. Huang has a BS in Aquaculture from Shanghai Ocean University in China and is expected to complete his MS from the same institution in June 2022. The title of Mr. Huang's MS thesis is "High-level interrelationships of the chondrichthyans revealed by exon capture data." Mr. Huang has experience with preparing DNA libraries for Illumina sequencing and with the use of various computer languages (R, Perl) for analyzing sequence data from high-throughput DNA sequencing. He has published four peer-reviewed manuscripts related to genomics in fish. Mr. Huang applied to the Ph.D. program in Zoology and there do not appear to be any impediments for Mr. Huang to begin the Ph.D. program in August 2022. When Mr. Huang arrives, will begin the process of streamlining the GTseq bioinformatics pipeline for genotyping sturgeon. In 2022, Mr. Huang will begin writing a Ph.D. dissertation proposal that includes the objectives of the funded project.

**Equipment and Supplies** – An Illumina MiSeq instrument was delivered to Dr. Heist's and installed following site-preparation guidelines provided by Illumina. Illumina also provided a list of necessary equipment and reagents for running the instrument, and everything needed for running and maintaining the instrument is currently in place.

**Mi-Seq install and training** – Following the delivery of the MiSeq instrument, a technician from Illumina performed an initial sequencing run to ensure that the instrument was functioning properly, and it passed all checks. Illumina has scheduled two days of on-site training on the MiSeq for January 13-14, 2022. Attending that training will be PI Ed Heist, graduate student Richard Flamio, and lab manager Amy Buhman. During the training we will discuss the prices and capabilities of available sequencing kits and flow cells to help optimize the numbers of individuals that can be run at once for either the quickest turnaround or the most efficient runs.

**SNP study and linkage map development** – Graduate student Richard Flamio developed a panel of Single Nucleotide Polymorphism (SNP) markers as part of his dissertation research. This project, which was funded by the US Geological Survey, has already produces one peer-reviewed manuscript describing the production of haploid sturgeon for developing disomic (i.e. two alleles per locus) markers in sturgeon with tetraploid ancestry. A second manuscript describing SNP variation within and between both species from two management units is nearly ready for submission. A third objective of that study is the development of a linkage map that describes the relative locations of those markers in the

genome. The linkage map will be used to select unlinked markers, which will provide more powerful and less confounded estimates of species identity and hybrid status. Mr. Flamio presented a draft linkage map at the American Fisheries Society annual meeting in Baltimore in November 2021. The linkage map requires some additional bioinformatic manipulation of parameters to produce a more comprehensive and accurate map. Mr. Flamio anticipates completing a draft manuscript on the final linkage map in time for graduating with his Ph.D. in May 2022. As soon as the linkage map is complete (likely February 2022) we will select 200-300 loci for SNP panels to be scored using the Genotyping-in-Thousands (GTseq) method.

**GT-Seq primer design and validation** – Matthew Campbell developed the GTseq method and formed a company called "GTseek" for consulting on the method. A Zoom meeting was held on November 16, 2021 with Matthew Campbell, GTseek technician Cheyenne Lobato, PI Ed Heist and Graduate Student Richard Flamio. During that meeting, Matthew Campbell noted that in addition to primer design, GTseek now offers "full-service panel development" which includes a validation of the developed primers based on a run of 96 samples and preparation of a "primer cocktail" which will make future more time-efficient. This service was not available when the proposal for this project was submitted.

The original budget for the project included $3000 for GTseek consulting (primer design) and $5750 for GTseq primers. The quote provided from GTseek for developing and validating 200 loci was $11,235.60 which includes genotyping of the first 96 sturgeon samples. Optimization of GT-seq reactions can be an expensive and time-consuming process, and for some researchers this leads to time delays and cost overruns. Dr. Heist believes that the additional expense up-front for the full-service panel development performed by the experts is justified. Dr. Heist informally discussed using some of the commodity budget from the first year of the project to make up for the extra expenses associated with the full-service panel development.

During the Zoom call we also discussed the sturgeon data file formats needed for GTseq panel development. Once candidate loci are chosen based on current SNP data and a draft linkage map, we will submit the appropriate data files to GTseek for primer development along with 96 DNA samples from pallid and shovelnose sturgeon from the SNP study for genotyping and validation of results. The characteristics of the panel developed (e.g. length of amplicons, number of reads requires for adequate coverage) will be used to guide choices of sequencing kits and flow cells for either the quickest turnaround or the most efficient runs.

**Field sampling** – Dr. Heist will supply sampling materials for young of the year (YOY) to collaborators at the University of Nebraska at Lincoln (UNL). UNL researchers should request sampling kits for juvenile and adult pallid sturgeon from the US Fish and Wildlife Service which include two tubes for collecting duplicate samples from each fish. One sample will be sent to Dr. Heist and the other will be sent to the repository at Lamar Fish Technology Center in Pennsylvania following the Range-wide Pallid Sturgeon Handling Protocols and Procedures.

**Anticipated timeline for 2022**

March 1 – Completion of linkage map and submission of data files to GTseek for primer development and validation.

May-June – Receipt of first young of the year (YOY) sturgeon from UNL. Ms. Buhman will use existing SNP and microsatellite technology to quickly identify any pallid sturgeon so that field sampling can be modified as warranted.

June 1 – Receipt of GTseek primer cocktails and protocols. Dr. Heist and Ms. Buhman will perform Initial runs of GTseq on sturgeon samples already provided by ACOE and begin troubleshooting any issues that arise with GTseq through GTseek and with the MiSeq instrument with Illumina.

August 1 – Receipt of first batch of adult PRRIP samples from UNL. Samples received from UNL will be genotyped using GTseq. Additional samples already in hand will be used to "fill" the run so that samples are run most efficiently.

August 16 – Huang Junman begins Ph.D. program.

December 31 – Completion of first 1000 individual GTseq runs and initial data analysis. This number includes the original 96 samples used for the SNP panel development by GTseq, all samples provided by UNL by October 1, and additional samples already in hand from the Missouri River.