Landscape genetics of Iowa's threatened black redhorse

Morganne C. Borsh

University of Northern Iowa
LANDSCAPE GENETICS OF
IOWA'S THREATENED BLACK REDHORSE

A Thesis Submitted
in Partial Fulfillment
of the Requirements for the Designation
University Honors

Morganne C. Borsh
University of Northern Iowa
May 2021
This Study by: Morganne C. Borsh

Entitled: Landscape Genetics of Iowa’s Threatened Black Redhorse

has been approved as meeting the thesis or project requirement for University Honors

________________________
Date

Peter Berendzen, Honors Thesis Advisor, Biology

________________________
Date

Dr. Jessica Moon, Director, University Honors Program
Introduction

Iowa has one of the most agriculturally altered landscapes in the United States, which has resulted in a decline of biodiversity. Currently, Iowa ranks last in the amount of presettlement vegetation remaining (Sampson and Knopf, 1994). Iowa’s rivers and streams are similarly impaired by high concentrations of nutrients, pesticides, bacteria, and soil sediment (IDNR, 2017; Iowa Association of Naturalists, 2018). Richter et al. (1997), highlighted habitat degradation associated with agriculture as one of the leading threats to aquatic biodiversity loss, including both altered sediment loads and excessive nutrient inputs. Overall ecological integrity of aquatic systems is determined by flow regime, habitat structure, water quality, and energy relationships; agricultural practices compromise these critical facets of stream ecology (Menzel, 1984).

Flow regime is a chief constituent in habitat requirements for aquatic life and is critical to various developmental stages across species (Menzel, 1984). Agricultural practices surrounding land drainage have significantly altered flow regime characteristics such as stream velocity, turbidity, sediment load, and flow rate, changing the stream environment. Many of these changes are derived from erosion which is exacerbated by unvegetated banks and drainage practices. These environmental changes result in reduced habitat for aquatic species not adapted to unstable environments (Menzel, 1984).

Changes in the physical structure of streams similarly affects aquatic species. Menzel (1984) notes the physical structure of a stream, including meandering, substrate, relief, and cover-providing features, determine the ecological relationships of stream biota. Specifically, channelization, or altering river flow into managed channels, is well-documented to decrease species diversity within affected streams. Channelization is
associated with decreased average water depth, wider channels, decrease in aquatic vegetation, and an increase in smaller sediment particles within the substrate (Trautman, 1939; Welker, 1967; Congdon, 1973; Hansen, 1972; Trautman and Gartman, 1974; Huggins & Moss, 1974; Headrick, 1976; Menzel & Fierstine, 1976; Gorman & Karr, 1978; Zimmer & Bachmann, 1978). Fish communities ultimately mirror these changes, seeing decreases in species adapted to deeper, cooler, clearer, and more vegetated waters; communities are unlikely to completely recover (Menzel, 1984).

Stream ecosystems see a decline in species richness with poorer water quality, a recognized concern in Iowa. In 2016, approximately 61% of Iowa’s streams and rivers were considered impaired, largely (84%) due to bacterial indicators, fish kills, and biological contaminants (IDNR, 2017). Additional pollutants include sediments, nutrients, chemical pollutants, organic pollutants, and pesticides. Sediment pollutants have proved particularly damaging to all levels of community organization, especially for native species evolutionarily adapted to clear water conditions (Menzel, 1984). This ultimately impairs reproduction, feeding, defense, and social interactions, among other life processes. Agriculture is a chief cause of several of these pollutants and has directly resulted in eutrophication through pesticides, fertilizers, and destruction of natural nutrient sinks (Menzel, 1984).

Lastly, agricultural practices have negatively influenced energy relationships and community dynamics. Headwater streams are closely related to the surrounding land. As agriculture results in fewer vegetated banks, organic detritus decreases and primary production of algae significantly increases (Menzel, 1984; Kilkus et al., 1975; Schlosser

Factors compromising flow regime, habitat structure, water quality, and energy relationships can lead to population fragmentation, local extirpations, and have ultimately resulted in the decline of freshwater organisms (Page et al., 1997; Menzel, 1984). Areas with extensive agriculture that interrupt these integral aquatic ecological aspects, such as Iowa, are at risk for further declines in native stream species. Understanding threats to aquatic biodiversity is imperative for the preservation of Iowa’s native species and the aquatic ecosystem.

These concerns are reflected by the Black Redhorse (*Moxostoma duquesnei*), an Iowa native freshwater fish in the family Catostomidae, which is listed as threatened and a species of greatest conservation need (Zohrer, 2006). The Iowa Department of Natural Resources (IDNR) describes the Black Redhorse as a small, freshwater fish most common to clear, cool, and swift streams, small to medium in size. Streams with shallow riffles (< 2ft deep) and cobble substrate are particularly important for spawning, making increased stream sediment loads highly problematic to species reproduction. The Black Redhorse can reach 2 pounds in weight and a length of 17 inches and is slender in appearance (Figure 1). While they have a widespread distribution across eastern North America, they have a patchy disjunct distribution in northeast Iowa and are found in the Upper Iowa, Turkey-Volga, Maquoketa, Cedar, and Wapsipinicon river drainages (Figure 2) (Zohrer, 2006). This limited distribution, combined with their reduction in statewide population extent, and specialized ecological requirements, justify their threatened species designation (Kurtz, 2014).
Their limited distribution throughout Iowa is principally a result of glaciation. Multiple glacial cycles, with the last glacial advance approximately 10,500 years ago, altered the landscape and drainage patterns within the region (Anderson, 1998; Mayden, 1988; Prior, 1991). In Iowa, glacial maxiums extended as far south as central Iowa (Anderson, 1998; Prior, 1991). Aquatic species persisted in southern refugia, then migrated north through new drainages, establishing populations in areas with suitable habitat (Berendzen et al., 2003; Bernatchez & Wilson, 1998; Hewitt, 2000). This has shaped the primary distribution of Black Redhorse populations today.

The objectives of this study are to:

a. Characterize the genetic diversity of populations of Black Redhorse in Iowa
b. Determine the level of connectivity between populations in Iowa
Overall, this project aims to determine the degree of fragmentation of Black Redhorse populations to better direct conservation strategies. This study utilizes genome-wide Single Nucleotide Polymorphisms (SNPs) to assess the population genetic structure of the Black Redhorse across its range in northeast Iowa.

Figure 2: Black Redhorse range and predicted distribution in Iowa. Green areas represent predicted distribution while blue points indicate historical presence (IDNR, 2004).

A previous study on Black Redhorse population genetics was done utilizing microsatellites. In this study, Kurtz (2014) found unique colonization of the Maquoketa River, but failed to detect contemporary changes in genetic diversity and population substructure. To better assess fine-scale population structure and contemporary genetic
diversity, genome wide SNPs were chosen for their advantages derived from SNP loci abundance (Lemopoulos, 2019). Lemopoulos et al. (2019), recommended a SNP approach over microsatellites when seeking fine-scale population divergence and detailed demographic questions, finding SNPs to be more informative than microsatellites when describing relatedness from small sample sizes and isolated populations. Previous genetic studies employing microsatellites indicated that historical processes could have a stronger influence on genetic signatures that is not representative of contemporary gene flow (Davis et al., 2015). High-throughput sequencing utilizing SNPs have captured even tenuous relationships with more accuracy than microsatellites (Laoun et al., 2020), indicating the ability for SNPs to overcome such discrepancies. This study will further compare SNP-generated population structure with microsatellite-derived structure from Kurtz (2014).

Methods

Study Area and Sampling

Twenty-eight individuals were sub-sampled from Kurtz (2014) for collection and analyses of SNP data. These individuals originated from four drainages across their range in the upper Mississippi River Basin in Iowa. The sample includes: twelve individuals from the Upper Iowa River, five from the Turkey River, five from the Volga River, and six from the Maquoketa River (Figure 3; Table 1). Fin clips were taken from each sample and stored in a 95% EtOH for later DNA extraction.
Figure 3: Kurtz’s (2014) tissue collection sites within the Upper Mississippi River Basin in Iowa. Subsamples were taken from the Upper Iowa (12), Turkey (5), Volga (5), and Maquoketa (6) rivers.

**Library Preparation**

Restriction-Site Associated DNA Sequencing (RADSeq) was used to collect SNPs to genotype individuals. Genomic DNA was first extracted from the fin clips and quantified using a nano-drop spectrophotometer. Approximately 1 μg of DNA from each sample was then digested at 37°C for 2 hours using the high-fidelity SbfI restriction enzyme (Figure 4A). The resulting overhang was then ligated to an adapter (P1) containing a matching DNA sequence and a molecular identifier (MID) (Figure 4B). The
samples were pooled and sheared using a sequence of cyclic sonication (Figure 4C). Resulting DNA segments were then selected for 300-400 base pair (bp) lengths using magnetic beads in a PEG/NaCl buffer. Libraries were then blunt-end repaired and dA-tailed (Rohland & Reich, 2012). The library identifying P2 adapter was then ligated to the fragments and amplified via PCR (Figure 4D & E) (Andrews et al., 2016). Samples were cleaned and again size selected using a GeneRead Size Selection Kit (Qiagen). Genomic libraries were then quantified and pooled using a Qubit Fluorometer. A Bioanalyzer assay (Agilent) was used to verify DNA fragment sizes. Illumina NovaSeq 6000 and was used to sequence 100 paired-end reads of the pooled libraries at the University of Iowa’s Institute of Human Genetics, Genomics Division (Figure 4F).
Figure 4: RADseq library preparation process. (A) SbfI shears DNA. (B) P1 adapter is ligated to SbfI-cut fragments. (C) Samples from multiple individuals are pooled together and sheared. (D) P2 adapter is ligated to all fragments. (E) PCR amplification with P1 and P2 primers. (F) Pooled samples with different MIDs are separated bioinformatically (Davey & Blaxter, 2010).

Data Preparation

Raw Illumina sequence reads were processed using STACKS v.2.54 utilizing the process_radtags pipeline (Catchen et al., 2013). Sequenced data was de-multiplexed and filtered by their adapter sequences. Two mismatches were allowed within the adapter
sequences ‘--adapter_mm 2’, and any reads marked as failures by the Illumina purity filter were discarded by using the ‘--filter_illumina’ options. Data were removed if they exhibited uncalled bases (-c) or low phred33 quality scores (-q), and recover and repair barcode segments (-r) (Rochette et al., 2019). Samples were trimmed to 95 bp length to account for low-quality bases from the 5’ end of reads (Catchen et al., 2011). The output was then concatenated into a single file for each respective sample (Rochette & Catchen, 2017).

As no reference genome is available, the denovo_map.pl script was used to execute the STACKS pipeline. The pipeline (Catchen et al.’s 2011), produces loci for each sample (ustacks); creates a catalog of all loci across the population (cstacks); organizes samples by RAD locus (tsv2bam); assembles a contig, calls SNPs and phases SNPs into haplotypes (gstacks); and generates de novo summary statistics (populations). The minimum depth of coverage required to create a stack (-m) and the maximum distance allowed between stacks (-M) were both set to 3. Mismatches between sample loci during assembly was set to 2 (-n).

Once the de novo genome was created, another ‘populations’ pipeline was used to call the SNP dataset and generate population genetics statistics (STACKS). Due to small sample size, SNPs were called using a population map, which grouped individuals by river. The minimum percentage of individuals required to process a locus for that population was set to 50% (-r 0.5) and a minimum of 34% of individuals across populations were required to process a locus (-R 0.34). Lastly, the analysis was further restricted to the first SNP per locus, reducing potential linkage disequilibrium error (‘--write-single-snp’). The minimum number of populations a locus must be present in to
process a locus was set to 1 (\( p 1 \)). This generated observed heterozygosity, observed homozygosity, and nucleotide diversity and the inbreeding coefficient.

**Population Structure and Genetic Diversity**

To identify and characterize independent genetic clusters, a Discriminant Analysis of Principal Components (DAPC) was conducted. A DAPC is a multivariate method that assigns individuals to groups based on identified genetic relatedness (Jombart et al., 2010). The genepop *populations* output file was imported into the DAPC web server via the *adegenet* package in R (Jombart, 2008; Jombart et al., 2010). To identify the correct number of PCA axes to retain (1-18), 10,000 cross-validation iterations were performed, signifying the retention of 12 PCAs. These are then used to generate uncorrelated input variables for the Discriminant Analysis (DA), which seeks to maximize between-group variation and minimize within-group variation (Jombart et al., 2010; Viengkone et al., 2016). The suggested 3 DA axes were then retained.

**Results**

Population filtering removed 6,978,688 loci from 7,394,147 loci. Of these, 415,459 loci were kept from the 156,955 remaining variant sites among all individuals, after loci filtering. Observed and expected heterozygosity is low across populations (Table 1). On average, each population has a moderate degree of inbreeding (\( F_{IS} = 0.15875 \)), ranging from 0.11817 to 0.25802 (Table 1). The Upper Iowa River exhibits the highest proportion of inbreeding, while the Turkey, Volga, and Maquoketa rivers all share similar inbreeding levels between approximately 0.12 and 0.13. The
Upper Iowa River also had the highest number of private alleles, totaling 70,242, compared to the average of 24,458.5. The Turkey River has the fewest number of private alleles, totaling 4,648 (Table 1). The Turkey River additionally maintains the lowest nucleotide diversity, at approximately 0.1697. Nucleotide diversity ranges from 0.16964 to 0.21815, with the Upper Iowa again exhibiting the highest degree of differentiation (Table 1).

Table 1: Summary of genetic diversity by river drainage. Private Alleles, observed heterozygosity (H_{ob}), expected heterozygosity (H_{ex}), nucleotide diversity (\pi), and inbreeding coefficient (F_{is}) estimates by river in northeast Iowa.

<table>
<thead>
<tr>
<th>River</th>
<th>n</th>
<th>Private Alleles</th>
<th>H_{ob}</th>
<th>H_{ex}</th>
<th>\pi</th>
<th>F_{is}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turkey</td>
<td>5</td>
<td>4648</td>
<td>0.10549</td>
<td>0.14311</td>
<td>0.16964</td>
<td>0.12407</td>
</tr>
<tr>
<td>Volga</td>
<td>5</td>
<td>11290</td>
<td>0.11653</td>
<td>0.15086</td>
<td>0.17585</td>
<td>0.11817</td>
</tr>
<tr>
<td>Maquoketa</td>
<td>6</td>
<td>11654</td>
<td>0.10351</td>
<td>0.14702</td>
<td>0.17086</td>
<td>0.13474</td>
</tr>
<tr>
<td>Upper Iowa</td>
<td>12</td>
<td>70242</td>
<td>0.13684</td>
<td>0.20378</td>
<td>0.21815</td>
<td>0.25802</td>
</tr>
</tbody>
</table>

Pairwise F_{st} values range from 0.055203 to 0.11716, with the Volga and Upper Iowa rivers being the most genetically similar, and the Turkey-Maquoketa rivers sharing the least amount of genetic diversity, respectively (Table 2). The average F_{st} between populations is 0.084191, indicating low population subdivision. Overall, the pairwise analyses indicate that all populations are relatively similar, with a low likelihood of diversity between populations (D_{xy} = 0.0027) (Table 2).
Table 2: Pairwise analysis of the mean fixation index ($F_{ST}$) and absolute nucleotide divergence ($D_{XY}$)

<table>
<thead>
<tr>
<th>River Comparisons</th>
<th>Mean $F_{ST}$</th>
<th>Mean $D_{XY}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turkey-Volga</td>
<td>0.10488</td>
<td>0.0032075</td>
</tr>
<tr>
<td>Turkey-Maquoketa</td>
<td>0.11716</td>
<td>0.0032647</td>
</tr>
<tr>
<td>Turkey-Upper Iowa</td>
<td>0.055842</td>
<td>0.0021607</td>
</tr>
<tr>
<td>Volga-Maquoketa</td>
<td>0.11046</td>
<td>0.003131</td>
</tr>
<tr>
<td>Volga-Upper Iowa</td>
<td>0.055203</td>
<td>0.0021559</td>
</tr>
<tr>
<td>Maquoketa-Upper Iowa</td>
<td>0.061599</td>
<td>0.0021892</td>
</tr>
</tbody>
</table>

Despite high population similarities, the DAPC retained 3 DAPC axes for the 4 population clusters (Figure 5). Individual river systems were distinguished, denoting a high amount of variation within systems (Figure 5). The Turkey River is not as varied and has evidence of shared ancestry across most individuals. Approximately 78.6% of the individuals had the overall correct assignment (Table 3). Of the Turkey River individuals, 100% were reassigned to their original cluster, as the population cluster is not unique and was not retained as a DAPC axis. Individuals from the Volga and Upper Iowa rivers were reassigned to their original groups at 80% and 75%, respectively. Maquoketa River individuals were reassigned at 67%. From prior to posterior assignment, the Turkey River increased from 5 to 11 individuals. The Volga and Maquoketa both decreased from 5 to 4 and 6 to 4, respectively. The Upper Iowa River posterior assignment population consisted of 9 individuals, opposed to the 12 prior individual assignments. Figure 6 displays river membership probability for each individual that is the basis for posterior assignment.
**Figure 5**: DAPC scatter plot with inertia ellipses displaying a spatial degree of variance between Black Redhorse river populations. Individuals are represented by dots and predefined rivers are differentiated by color. Distance between points represents relatedness, with shorter distances being more related. The inertia ellipses further group similar individuals (Jombart et al., 2010). Eigenvalues are displayed in the insets.

**Figure 6**: Comoplot showing the admixture or probability of group membership for each individual. Bars with >1 color indicate the sharing of genetic ancestry (admixture) with the associated probability of belonging to a specified population (y-axis).
Table 3: DAPC summary statistics for group assignment. 3 axes retained, 4 populations (78.57% average retention of individuals across populations).

<table>
<thead>
<tr>
<th>River</th>
<th>Percentage of Individuals Retained</th>
<th>Prior Individual Assignment</th>
<th>Posterior Individual Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turkey</td>
<td>100.00</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>Volga</td>
<td>80.00</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Maquoketa</td>
<td>66.67</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Upper Iowa</td>
<td>75.00</td>
<td>12</td>
<td>9</td>
</tr>
</tbody>
</table>

Discussion

The results of the analyses of the SNP dataset reveal fine scale population subdivision, moderate levels of connectivity and an even distribution of genetic diversity. The DAPC identified genetic clusters corresponding with each river drainage, as displayed in Figure 5. The three prominent clusters (the Maquoketa, Upper Iowa, and Volga) are likely a function of geographic distance, as spatial isolation frequently influences genetic variation (Bradburd, Ralph & Coop, 2013; Wright, 1943). Although the Turkey was highly similar to the other drainages (indicated by the high overlap and lack of DA axis), it still maintained a unique genetic signature and was loosely clustered in the DAPC. All individuals originally assigned to the Turkey River drainage were reassigned, and several were added in the posterior DAPC, again indicating high connectivity between the Turkey River and other localities (Table 3). This is again seen in the comoplot, where most individuals are admixed with the Turkey (Figure 6).

Despite individuals clustering by drainage, diversity was fairly uniform across drainages. All populations had relatively low ranges of variation, observed
heterozygosity, and observed homozygosity, highlighting moderate to high levels of connectivity between populations. The Upper Iowa River recorded the highest number of private alleles and inbreeding coefficient (Table 1), indicating it is slightly more isolated, but still showed relatively low amounts of genetic variation (\( \pi = 0.21 \)). This is again displayed in the comoplot, where individuals corresponding with the Upper Iowa River generally show less admixture (Figure 6).

Admixture and uniform diversity were again seen in pairwise \( F_{ST} \) values. The Maquoketa and Upper Iowa river systems are separated by the furthest geographic distance, but showed little divergence from one another. River drainages closer in distance (such as the Maquoketa and Volga) had the most amount of divergence. While other geologic aspects should be considered, the initial lack of pattern likely reflects the species life history, and relative recent expansion from glacial refugia.

These results are consistent with previous studies that similarly find unique population clusters with low genetic distinction (Kurtz, 2014; Davis et al., 2015). These findings likely result from Iowa’s unique geologic history more than contemporary influences. During the quaternary, Iowa experienced several glacial cycles that have shaped recurring recolonization patterns and resulting population structure (Anderson, 1998; Prior, 1991; Davis et al., 2015; Hewitt, 2000). As glaciers advanced and retreated, fish populations recolonized as habitat became available. Davis et al. (2015) found that these historical processes often had lasting effects on population structure, obscuring contemporary influences in Iowa fish populations. Black Redhorse populations likely reflect this as well, and a species’ life history should be considered as recent geologic processes may obscure genetic representation.

Commented [MCB4]: Then summarize the analyses. Here is a bit of a guide:

1. The DAPC did identify genetic clusters that correspond to each river drainage – explain
   a. Can talk about the assignment tests with this
2. Interestingly individuals from the Turkey river did not form a distinct cluster, but overlapped the other clusters
   a. This is seen in the complot where a lot of individuals are admixed with the Turkey
3. Even though they clustered by drainage the diversity was fairly uniform across drainages
   a. Discuss diversity measures
4. The only exception was the upper Iowa
   a. Explain how it has more private alleles
   b. The complot shows the individuals have less admixture

Then discuss \( F_{ST} \) measures.
Microsatellite Comparison

The SNP dataset revealed a more fine-scaled population structure than that indicated by microsatellites (Kurtz, 2014). As SNP datasets have previously been acclaimed for providing better inferences to population structure than microsatellites, we compared Kurtz’s microsatellite study of Black Redhorse population structure to our findings. The largest advantage of SNPs was the identification of each genetically unique river system, compared to Kurtz’s (2014) differentiation between two clusters (the Maquoketa and all other river systems). Microsatellites indicated genetic distinction between the Maquoketa and all other drainages, while SNPs separated the Maquoketa, Upper Iowa, and Volga, finding all with a high degree of admixture. Microsatellites indicated the Maquoketa River is of most conservation concern; however, SNPs revealed it is less isolated than anticipated, instead denoting the Upper Iowa as having the least amount of genetic variation and connectivity. Our SNP dataset does indicate lower variation within the Maquoketa that should be further explored.

Conclusion

Iowa’s aquatic species are growing increasingly imperiled through a series of anthropogenic land-use changes. Species adapted for cool, deep, clean waters, such as the Black Redhorse, are of particular concern. Our results indicate the Upper Iowa River populations should be of primary conservation concern; however, more analyses should be explored to better understand current population structure.

Corresponding with Liu et al.’s (2005) findings, our SNP dataset appeared more sensitive than the microsatellites, despite a smaller sample size. The low amount of data
points associated with microsatellites may not be able to overcome Iowa’s recent historical events, while the increased resolution of data derived from SNPs can reveal more contemporary processes. This should be considered for future genetic studies where historical processes may be heavily influencing species’ genetic structure.
References


Trautman, M. B. (1939). The Effects of Man-Made Modifications on the Fish Fauna in Lost and Gordon Creeks, Ohio, Between 1887-1938.


