POPULATION GENOMICS OF BLANDING'S TURTLE ON A REGIONAL SCALE IN THE MIDWEST

by

Connor Dempsey

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science

Department of Biology at Purdue Fort Wayne

Fort Wayne, Indiana

August 2021

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Mark Jordan, Chair
Department of Biology

Dr. Bruce Kingsbury
Department of Biology

Dr. Jordan Marshall
Department of Biology

Dr. Matt Cross
Toledo Zoo

Approved by:
Dr. Jordan M. Marshall

For my wife, Shannon

ACKNOWLEDGMENTS

I would like to thank my committee chair, Dr. Mark Jordan, for his guidance and assistance throughout this entire process. I would also like to thank my committee members Dr. Bruce Kingsbury, Dr. Jordan Marshall, and Dr. Matt Cross for their suggestions and for helping to review and refine my writing.

Projects such as mine depend on the sharing of tissue samples from collaborators conducting field work on wild populations. Tissue samples were contributed by Matt Cross (Toledo Zoo), Greg Lipps (Ohio State University), Yu Man Lee (Michigan Natural Features Inventory), Fred Janzen (Iowa State University), and Steve Mockford (Acadia University). Many thanks to Dan Guinto for all of his work and assistance in the laboratory and for providing valuable discourse on my analysis and results. I would also like to thank JJ Apodaca and Tangled Bank Conservation for their work in the library construction and initial bioinformatic filtering steps and for answering my many questions on sample data analysis and results. Funding support for tissue collection and genotyping was through grants from the Indiana Department of Natural Resources, Toledo Zoo, and the U.S. Fish and Wildlife Service to Mark Jordan and collaborators.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Maintaining high genetic diversity within and among wildlife populations is an important component to the management of threatened species. Population genomics utilizes recent advancements in high-throughput next-generation sequencing to obtain genome-wide data that can yield deeper perspectives on intraspecific genetic variation and elucidate evolutionary significant units that may require conservation management or augmentation. The semi-aquatic Blanding's Turtle (*Emydoidea blandingii*) has experienced drastic population declines in North America due in large part to anthropogenic activities. This species is listed as threatened or endangered across most of its range. A population genomic study can help to understand the status of this species and guide future management practices. Hence, a population genomic analysis was conducted using 3RAD to discover and analyze SNPs across the range using samples from Nebraska, Indiana, Michigan, Ohio, and Nova Scotia,. Range-wide analysis used 8,602 SNPs while analysis within the Great Lakes region used 7,893 SNPs. High amounts of missing data were found across all individuals and loci. Low levels of genetic variation relative to other turtle species were detected both across the range and within the Great Lakes region. Minimal population structure was detected range-wide via clustering and admixture analyses; however, a signal of population differentiation was detected among Nebraska, Nova Scotia, and the Great Lakes. Clustering and differentiation analyses focused on the Great Lakes region found a signal of population structure and differences between the Lake Michigan and Lake Erie watershed. These results may prove useful for conservation management of Blanding's Turtle populations, particularly related to efforts using translocation or head-starting practices.

CHAPTER 1. INTRODUCTION

1.1 Conservation Genetics

As the global loss of biodiversity has continued at an alarming rate, conservation genetics has become an important component in understanding the status of declining species and informing conservation management decisions (Frankham, 1995, 2003). Anthropogenic forces leading to habitat loss and population declines can have a direct impact on the genetics of wildlife populations, resulting in decreased genetic variation or inbreeding (Ashley et al., 1990; Gibbs, 1998; Templeton et al., 1990; Wayne et al., 1992). Reduced genetic diversity within and among populations is a cause for concern as low genetic variation can lead to reduced reproductive fitness and viability of individuals, i.e. inbreeding depression (Charlesworth & Willis, 2009; Lacy, 1987; Ralls et al., 1988). With the advent of new technologies, many conservation genetic studies routinely obtain genome-wide data that may supplement existing studies done with different molecular markers (Allendorf et al., 2010; De Cara et al., 2011; Luikart et al., 2003). Doing so can yield deeper perspectives on intraspecific genetic variation and elucidate evolutionary significant units that may require conservation management or augmentation (Meffe et al., 1995; Moritz, 1994).

1.2 Approaches to Population Genetic Analysis

Various genetic tools have been used in the past few decades to study genomic variation both within and between wildlife populations across a wide range of taxa (Fauvelot et al., 2003; García-Moreno et al., 1996; Jonker et al., 2012). One of the more popular markers for genetic studies have been microsatellites, which have been successful in numerous studies but require primers developed for the study species or a closely related species (Selkoe & Toonen, 2006; Wright & Bentzen, 1995). This can be an issue, as using primers for the study species that were developed for a different species, even if closely related, can lead to errors in the amplification of loci in the study species (Prosser et al., 1999). Significant loci that have occurred due to mutation may also be overlooked in the study species and repeat lengths of loci present in the study species may be misrepresented (Ellegren et al., 1995; Jarne & Lagoda, 1996). Another issue with microsatellites is the relatively low numbers of markers, which can be further reduced if issues such as those mentioned previously do occur (DeFaveri et al., 2013; Lemopoulos et al., 2019).

Using techniques that utilize higher numbers of markers may provide greater coverage and a better view of the level, distribution, and history of intraspecific genetic variation (Liu et al., 2005; Spinks et al., 2014).

Recent advancements in high-throughput next-generation sequencing have revolutionized how genetic studies are conducted and have facilitated the development and use of new genetic techniques (Emerson et al., 2010; Morin et al., 2004; Yang et al., 2009). One such technique that has gained popularity is Restriction site-Associated DNA Sequencing (RADseq). Its primary advantage over microsatellites is that there is potential for a high degree of resolution using large numbers of markers without the need for any prior genomic information (Andrews et al., 2016; Baird et al., 2008). The relatively lost-cost of discovery and genotyping of high numbers of genetic markers across the genome for non-model species has made RADseq an attractive method for conservation genetic projects working with a given budget (Andrews et al., 2016). This approach, while promising, has not been widely implemented for turtle populations at a regional scale.

RADseq uses a reduced-representation sequencing approach targeting a subset of the genome to identify hundreds or thousands of Single Nucleotide Polymorphisms (SNPs) (Luikart et al., 2003). In general, RADseq begins with multiple samples of high-quality DNA that are digested with one or more chosen restriction enzymes (Andrews et al., 2016). Next, specific sequencing adapters are constructed and ligated to target genomic DNA fragments adjacent to restriction cut sites. These adapters contain sample-specific barcodes that identify samples which are sequenced together (multiplexed) in a single library. The genomic DNA then includes a wide range of fragment lengths that are size-selected isolate fragments of ideal lengths for sequencing. Although many different approaches to RADSeq exist, one that has been developed more recently is a triple enzyme digestion approach (3RAD) (Bayona-Vásquez et al., 2019). 3RAD uses three restriction enzymes during the sequencing process which helps to accurately detect SNPs in the genome and overcomes some of the limitations of other RADSeq methods (Bayona-Vásquez et al., 2019). It does this by reducing the quantity of adapter-dimer formation during digestion and ligation, which can help the adapter sequence to better ligate to the DNA. 3RAD also reduces preparation steps through better design of the adapters and reagents while improving PCR yields of input product and allowing for the multiplexing and pooling of high quantities of individual samples (Bayona-Vásquez et al., 2019). RADseq can be used to gain an accurate picture of the landscape genetics of non-model organisms.

1.3 Blanding's Turtle

Blanding's Turtle (*Emydoidea blandingii*) is a semi-aquatic species that inhabits shallow, freshwater wetlands across the northern United States and southern Canada (Ernst et al., 2009; Standing et al., 1999). It is species of conservation concern whose life-history traits make management efforts difficult (Congdon et al., 2001; Congdon et al., 1993). The first of these is extended longevity, as this species is known to be long-lived with a maximum lifespan of 75 years or more (Brecke & Moriarty, 1989; Congdon et al., 2001). Its reproductive traits of delayed maturation and low annual fecundity are also of note, as females reach sexual maturity at 14-20 years old and lay an average of one clutch of 10 eggs per year (Congdon et al., 2001; Ernst et al., 2009). Furthermore, individuals have been shown to travel long distances over terrestrial landscapes, commonly moving anywhere from hundred meters to several kilometers over the duration of their active season (Standing et al., 1999). Blanding's Turtle is currently listed as endangered or threatened throughout most of its range due to habitat loss or degradation and road mortality coupled with its delayed maturation and low reproductive frequency (Congdon et al., 1993).

The center of the Blanding's Turtle's geographic range is the Great Lakes region with its entire range spanning from southwestern Quebec and southern Ontario westward to Nebraska along with a few disjunct populations in the northeastern United States and Nova Scotia (Ernst et al., 2009). Major geological changes over 18,000 years ago, namely cycles of glacial advance and retreat, resulted in the current distribution of many native North American species (Schmidt, 1938; Streicher et al., 2012). This is true as well for the Blanding's turtle with evidence suggesting that at least two evolutionarily distinct lineages occur across the range (Jordan et al., 2019; Mockford et al., 2007). The Great Lakes region is hypothesized to be one genetic lineage that is separated from populations in New York, New England, and Nova Scotia along a boundary in eastern Ontario (Jordan et al., 2019; Mockford et al., 2007). A possible boundary may also exist in eastern Nebraska, separating these populations from those in the Great Lakes region and creating a third genetic lineage (Jordan et al., 2019). Within the Great Lakes region, current populations in Indiana, Michigan, and Ohio are thought to have resulted from northern expansions of the species after the retreat of the Wisconsin glacier (Mockford et al., 2007).

Genetically this species has shown low variation, possibly as a result of a bottleneck event following the glacial retreats (Rodder et al., 2013). However, recent evidence from a series of

regional analyses using microsatellite loci suggests that there can be considerable differentiation among populations within lineages (Davy et al., 2014; Reid et al., 2017; Sethuraman et al., 2014). This differentiation has also been found to correspond with watersheds, possibly as a result of post-glacial colonization along watershed boundaries resulting from the recession of the Wisconsin Glacier (Sethuraman et al., 2014). Human activities have further segmented and isolated the populations of Blanding's Turtle, potentially reducing gene flow and contributing to inbreeding depression. The Blanding's Turtle is considered a non-model organism with no reference genome of a closely related species available. At present, there is no genome-wide regional analysis of genetic variation in Blanding's Turtle that includes Indiana, Michigan, and Ohio.

1.4 Study Objective

My purpose in this study was to use RADseq to analyze the genetic variation of Blanding's Turtle populations. I focused on the regional level within Indiana, Michigan, and Ohio but also compare results to those from samples within other hypothesized lineages. This will help identify the degree of population structure, levels of inbreeding within populations, and contribute to understanding the population history of Blanding's Turtles in the region. It will also contribute to plans for conservation management of this species by locating sites that can be used as potential source populations for head-starting or translocation to augment or establish populations. Blanding's Turtle's status as a species of conservation concern, along with its historical range patterns makes this a species of interest for a population genomics study.

CHAPTER 2. METHODS

2.1 DNA Extraction and Sequencing

I collected 96 Blanding's Turtle blood and tissue samples previously gathered from 5 localities in Indiana, 8 localities in Ohio, 7 localities in Michigan, 1 locality in Nova Scotia, and 1 locality in Nebraska (Appendix A). Blood and tissue samples were stored in 95% ethanol at -4˚C. I used the DNeasy Blood & Tissue Kit (Qiagen, MD, USA) to perform DNA extractions on the blood and tissue samples and I quantified DNA concentrations using a Nanodrop spectrophotometer.

Genomic DNA samples were sent to Tangled Bank Conservation (Asheville, North Carolina) for PCR amplification and construction of a genomic library. A 3RAD protocol from the University of Georgia EHS DNA laboratory (Athens, GA) was followed using the restriction enzymes BamHI, MspI, and ClaI. Samples were first normalized to 20ng/µL and quantified on a Qubit fluorometer before preparation for ligation and PCR, with lower quality samples being purified with SpeedBead and checked on an agarose gel. PCR products were then sent to Genewiz (South Plainfield, NJ) for sequencing using an Illumina flow cell Novaseq S4 PE150 on a NovaSeq 6000 platform.

2.2 SNP Identification and Filtering

Initial identification, analysis, and filtering of the raw Illumina reads of 96 Blanding's Turtle samples was done using iPyRAD (Eaton 2020). The original pipeline pyRAD (Eaton, 2014) was created to provide a user friendly and flexible program for accurate assembly, analysis, and quality filtering of *de novo* locus identification applicable to larger scale studies. iPyRAD continues to provide many of the same benefits, including the ability to handle insertion-deletions among sequences, while greatly increasing performance and scalability (Eaton & Overcast, 2020).

iPyRAD includes a 7 step assembly workflow for processing raw Illumina data that results in multiple data analysis and statistics output files (for overview see, https://ipyrad.readthedocs.io/en/latest/7-outline.html). My samples were run through this workflow beginning with step 1, which loaded in the sequence data and demultiplexed the raw reads by barcodes. Demultiplexed reads were run through step 2, which performed quality control by filtering out base calls with a quality of less than 5 before moving on to step 3 where *de novo*

clustering was performed at a threshold of 85%. Steps 4-7 performed more filtering and clustering including removing duplicates and a maximum of 8 indels and 20 SNPs allowed per locus, a minimum of 4 samples per locus needed for output, a maximum of 2 alleles allowed at each site in the consensus sequences per individual, and sequences with more than 50% heterozygous sites being excluded. Initial filtering revealed one individual did not sequence well and was removed from the data.

The programs VCFtools (Danecek et al. 2011) and the POPULATIONS module in STACKS v2.55 (Catchen, Hohenlohe, et al., 2013) were used for further filtering of the output files from iPyRAD. Two different filtering schemes were conducted, one including samples from Nova Scotia and Nebraska with 3 populations assignments used based on geographic distance (FS-1) (Figure 1). The second filtering scheme excluded the Nova Scotia and Nebraska samples with two approaches to population assignments taken: one using 15 populations assignment based on locality and one using 2 population assignments based on the Lake Michigan and Lake Erie watersheds at the hydrologic unit 4 level (FS-2) (Figure 2). Both filtering schemes included filtering sites by a minimum genotype read depth of 2 which improves confidence that false homozygotes are excluded from further analysis. A minimum minor allele frequency of 5% was also applied to both filtering schemes to lessen the chance of including incorrectly called bases and to reduce the number of rare alleles in the data which can be uninformative (Roesti et al., 2012). Finally, sites with greater than 99% missing data were also filtered out. In order to reduce the chance of retaining linked loci, a random SNP from each locus was written to the output files for both filtering schemes from STACKS (Andrews et al. 2018).

2.3 Genetic Diversity and Population Structure

STRUCTURE v2.3.4 (Pritchard et al., 2000) a model-based approach which assumes Hardy-Weinberg equilibrium and linkage equilibrium among clustered samples, was used first to examine population structure. Runs in STRUCTURE were setup using values for K from 1 to 10 with both burn-ins and MCMC replications set to 10,000 repeated over 10 iterations for each K value. STRUCTURE runs were performed both with and without the LOCPRIOR parameter for FS-1 and FS-2. When using LOCPRIOR, *a priori* localities were assigned using sample location or watershed. LOCPRIOR builds the *a priori* localities into the model, which can be informative in situations with weak population structure signals (Porras-Hurtado et al., 2013). Output files from

STRUCTURE were run through StructureSelector (Li & Liu, 2018) with the Puechmaille method (Puechmaille, 2016) being used to determine the best-fit K value. The Puechmaille method is useful for judging the number of clusters in situations with uneven sampling across localities (Puechmaille, 2016). Clustering assignments from STRUCTURE results were visualized using Clumpak (Kopelman et al., 2015). Data files were then converted to PED format using PGDspider v2.1.1.5 (Lischer & Excoffier, 2012). PED files were converted into LFMM format for further analysis of population structure using the R package tess3r (Caye et al., 2016). R package tess3r differs from STRUCTURE by using ancestry coefficients stored in a *Q*-matrix and by incorporating geographic sample coordinates to visualize results on a geographical map.

The genetics tool PLINK was then used to convert PED files to a PLINK format for use in the R package SambaR for further analysis of population structure (de Jong et al., 2021; Purcell et al., 2007). In SambaR I ran discriminant analysis of principal components (DAPC) and principal coordinate analysis (PCoA) (Gower, 1966; Jombart et al., 2010). DAPC takes a model-free approach and can incorporate prior population assignments while PCoA is conceptually similar to principal component analysis (PCA) but uses the genetic data as an input in the form of a dissimilarity matrix. SambaR was also used to measure population differentiation (*F*st) between Nebraska, the Great Lakes, and Nova Scotia and between the Lake Michigan and Lake Erie watersheds., with low *F*st values indicating a low level of genetic differentiation between populations.

The POPULATIONS module in STACKS was used to gain measures of within site diversity including the number of private alleles, the number of polymorphic loci, observed heterozygosity (H_O), expected heterozygosity (H_E), and inbreeding coefficient (F_{is}). Analyses were done for both FS-1 and FS-2, with low levels of *Ho* indicating low genetic variability within a population and high levels of *Fis* indicating higher levels of inbreeding.

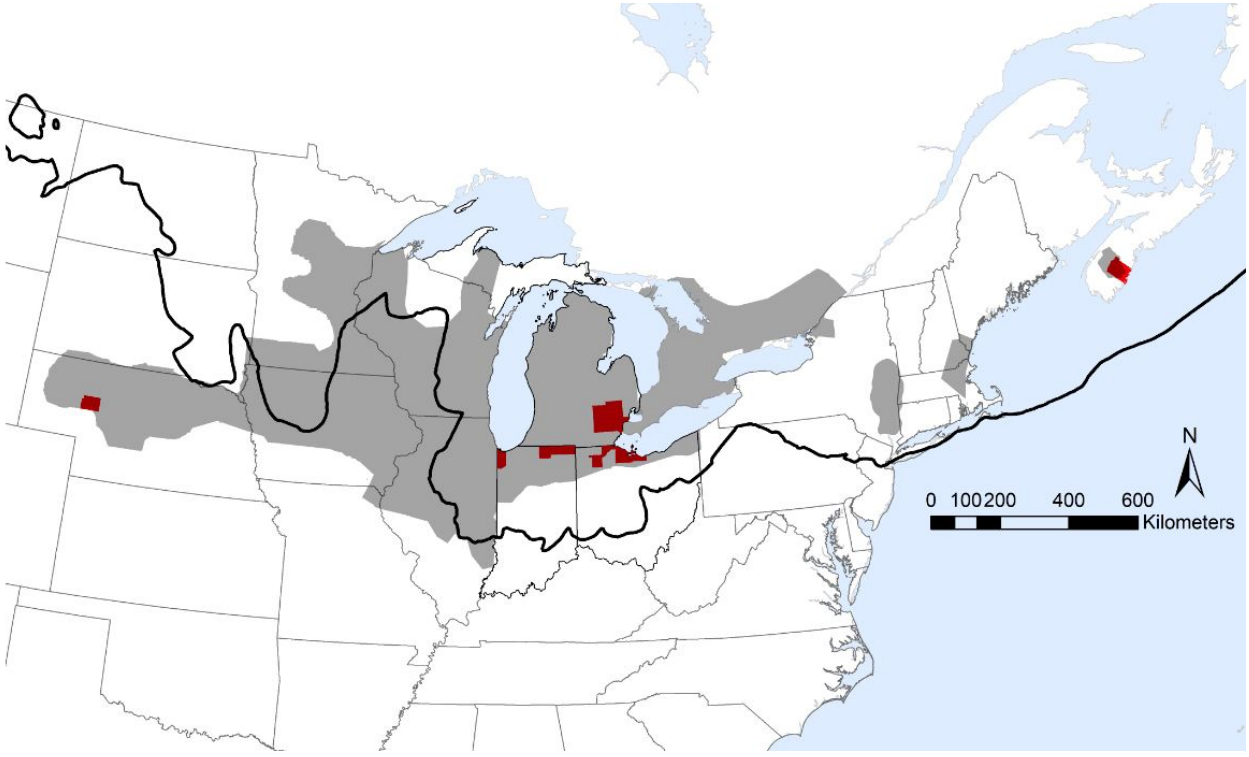


Figure 1: Distribution of Blanding's Turtle populations. Red layer indicates county locations of samples used in the analysis. Dark grey layer shows the range-wide geographic distribution of Blanding's Turtle. Black line shows last glacial maximum of the Wisconsin Glacier.

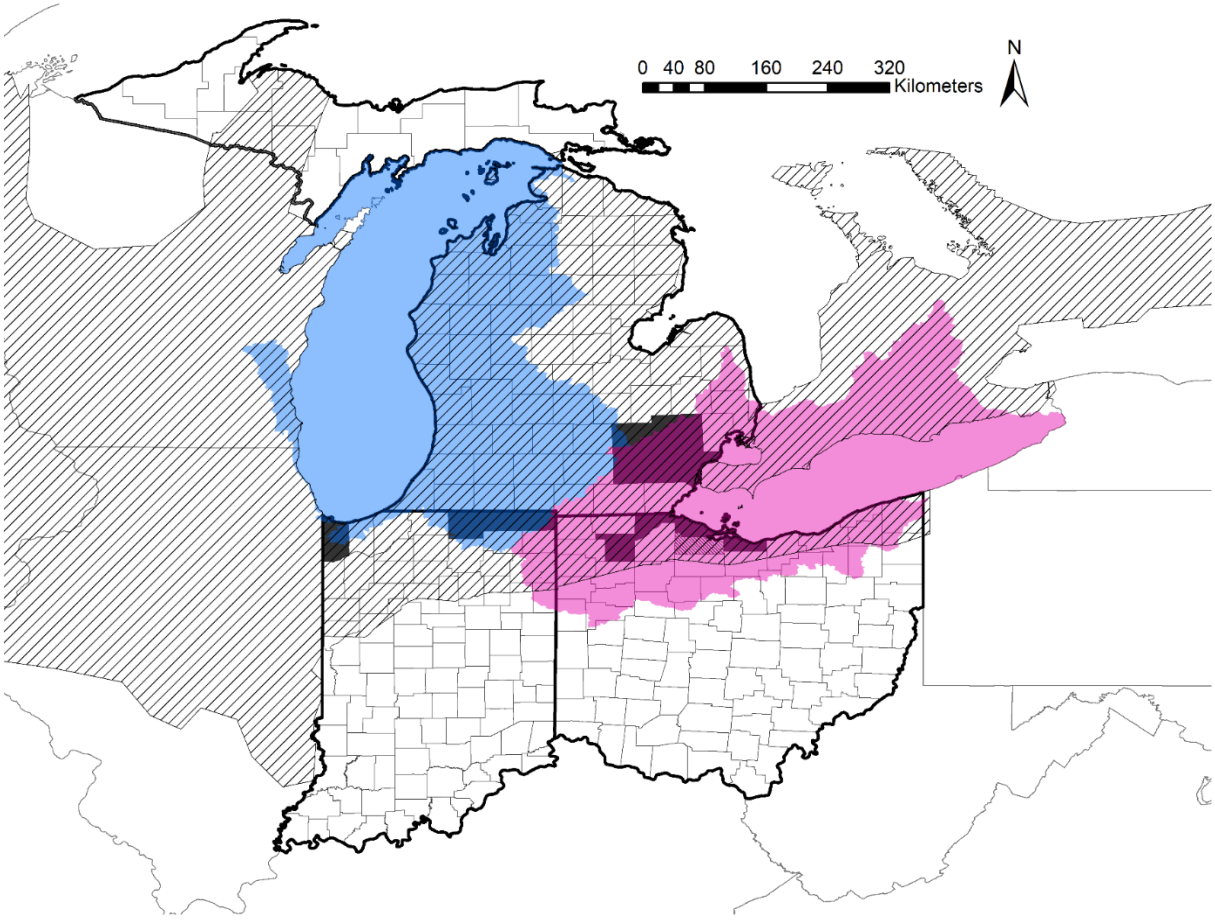


Figure 2: Sample and watershed distribution for the Midwest Blanding's Turtle populations. Dark counties show locations of samples used in the analysis. Cross-hatched layer shows regional geographic distribution of Blanding's Turtle. Blue layer shows the Lake Michigan Watershed and pink layer shows the Lake Erie Watershed

CHAPTER 3. RESULTS

3.1 Sequencing and Filtering

Initial *de novo* identification of loci using iPyRad resulted in the discovery of 25,778 loci. Primary filtering through iPyRad to remove duplicates and loci with more than 8 indels, more than 20 SNPs, more than 50% heterozygous sites, less than 4 samples, and more than 2 alleles per locus resulted in 10,070 SNPs retained across 95 individuals. Missing data per individual was 92.23% while the mean read depth per individual was 0.88 and the maximum read depth was 9.74. Further filtering through VCFtools of individuals based on a missing data threshold of 99% and outlier individuals discovered through PCoA resulted in 91 individuals retained in FS-1 and 78 individuals retained in FS-2. Final filtering of the remaining individuals in each filtering scheme by minimum depth per genotype of 2, minimum minor allele frequency of 0.05, and max missing data per site of 99% resulted in 8,602 unlinked SNPs retained in FS-1 and 7,893 unlinked SNPs retained in FS-2.

3.2 Filtering Scheme 1 (FS-1)

3.2.1 Population Structure

Analyses of population structure for FS-1 revealed variable best-fit K values (Table 1). STRUCTURE results of FS-1 based on the Puechmaille method both without LOCPRIOR and with LOCPRIOR based on geographic region revealed K = 3 but with high admixture and no obvious pattern of cluster assignment (Figure 3). PCoA results for FS-1 from SambaR showed minimal separation of Nova Scotia and Nebraska from the Great Lakes region (Figure 4). However, DAPC analysis run through SambaR using prior population assignment based on region showed a clear separation of most individuals into the three regions respectively (Figure 5).

Genetic differentiation analyses in SambaR showed significant ($P<0.001$) differences between among Nebraska, the Great Lakes, and Nova Scotia (Table 2). The lowest differentiation was between Nebraska and the Great Lakes (Fst = 0.174). The highest genetic differentiation was between Nebraska and Nova Scotia (Fst = 0.490).

3.2.2 Genetic Diversity

Genetic diversity results for each sampling locality from STACKS showed private alleles ranged from 41 to 1188 (Table 3). Samples IN06 and OH13 had the lowest number of private alleles while IN07 and MI10 had the highest. Number of polymorphic loci ranged from 179 to 2921 with OH13 and IN06 having the lowest and IN07 and MI10 having the highest (Table 3). Mean observed heterozygosity, $H_{O,}$ ranged from 0.2246 to 0.3259 (Table 3). Samples NS01 and NE01 had the lowest H_O values while IN06 and OH16 had the highest. Mean expected heterozygosity, H_E, ranged from 0.1180 to 0.1895 (Table 3). Samples NS01 and NE01 were the sites with the lowest H_E while IN07, MI10, and MI06 had the highest H_E values. Inbreeding coefficient values, F_{is}, ranged from -0.0075 to 0.0194 with IN07 having the highest and OH08 having the lowest (Table 3).

Comparing the Great Lakes region to Nebraska and Nova Scotia, the Great Lakes region had the highest values for each with 1929 private alleles, 8125 polymorphic loci, 0.3523 H_O, 0.3026 H_E, and 0.0510 F_{is} (Table 3).

3.3 Filtering Scheme 2 (FS-2)

3.3.1 Population Structure

Analyses of population structure for FS-2 revealed variable best-fit K values (Table 1). STRUCTURE results based on the Puechmaille method identified K = 2 for FS-2 both without LOCPRIOR and using LOCPRIOR with watersheds but found K = 3 for LOCPRIOR using site localities (Figure 6). Cluster assignments closely resembled watershed designation for all three approaches in FS-2 but was much more obvious when using watershed designation (Figure 6). Much more admixture (blue and red bars) was exhibited when comparing localities than when comparing watersheds (Figure 6). K = 2 was chosen for visualization of all Great Lakes region STRUCTURE results based on evidence of fewer clusters. Results from tess3r for FS-2 using the geographic coordinates of individuals identified K = 4 with the majority of individuals assigned to clusters roughly based on watershed designations (Figure 7).

PCoA results for FS-2 displayed no clear separation of individuals into clusters and lacked any pattern associated with watershed designation (Figure 8). However, DAPC analysis for FS-2

using a priori site localities based on watershed showed a clear separation of most individuals into the two watershed clusters (Figure 9).

Genetic differentiation analyses in SambaR showed significant ($P<0.001$) between the Lake Michigan and Lake Erie watershed clusters (Fst=0.028; Table 4).

3.3.2 Genetic Diversity

Genetic diversity results from STACKS comparing clusters based on watersheds showed the Lake Erie watershed cluster had a higher number of private alleles and polymorphic loci with 51 and 2764 respectively while the Lake Michigan watershed had 27 private alleles and 1582 polymorphic loci (Table 3). The Lake Erie Watershed also had higher values for H_O, H_E, and F_{is}, with 0.3632, 0.2768 and 0.0293 respectively (Table 3). The Lake Michigan watershed had an H_O value of 0.0343, am H_E value of 0.2324, and a F_{is} value of 0.0257 (Table 3).

Table 1: Summary of best-fit K values for each filtering scheme based on the program and model used.

Program/Model	Best-fit K (FS-1)	Best-fit K (FS-2)
STRUCTURE (no LOCPRIOR)	3	2
STRUCTURE (LOCPRIOR using site localities)	3	3
STRUCTURE (LOCPRIOR using watersheds)	-	2
SambaR (PCoA)	1*	1*
SambaR (DAPC)	3	2**

*little to no structure seen

**using location information based on watershed

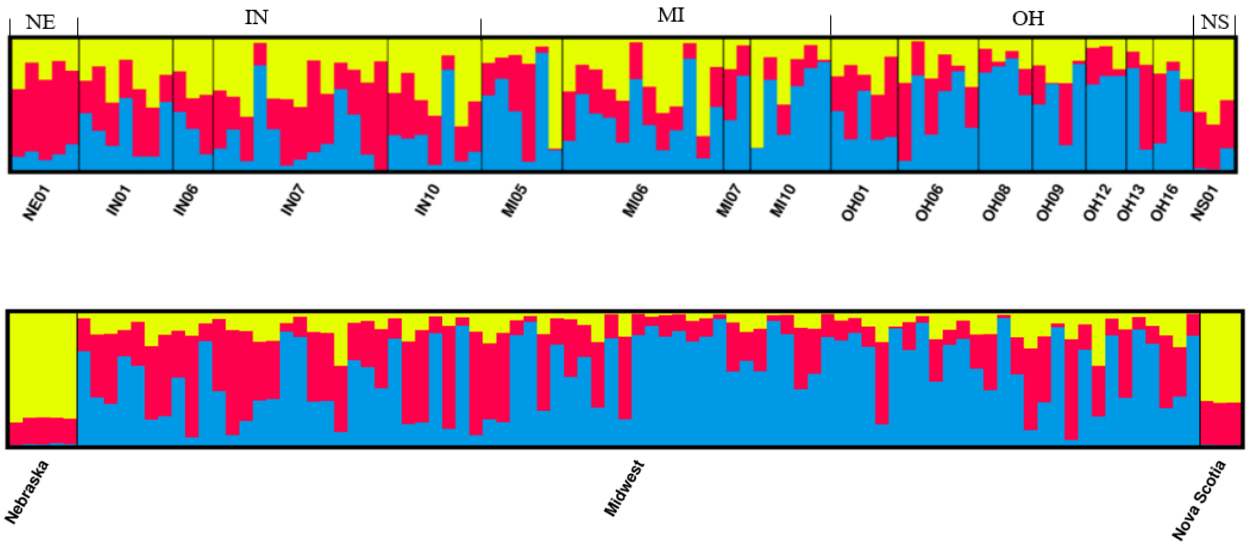


Figure 3: Cluster results for FS-1 from STRUCTURE both with and without LOCPRIOR across all 17 site localities based on K = 3. Blue bars represent cluster 1, red bars represent cluster 2, and yellow bars represent cluster 3.

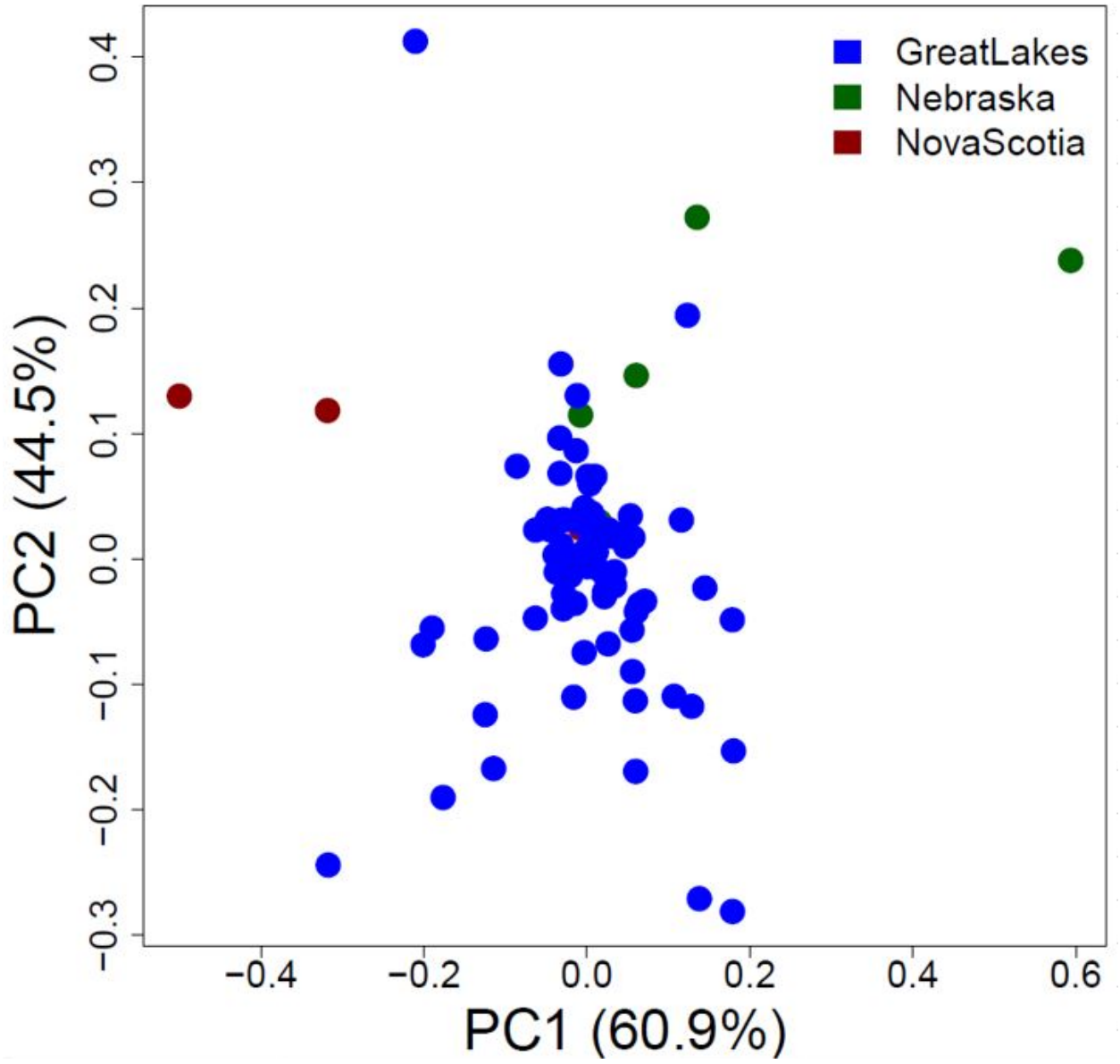


Figure 4: PCoA results from SambaR for FS-1 comparing Nebraska, the Great Lakes, and Nova Scotia.

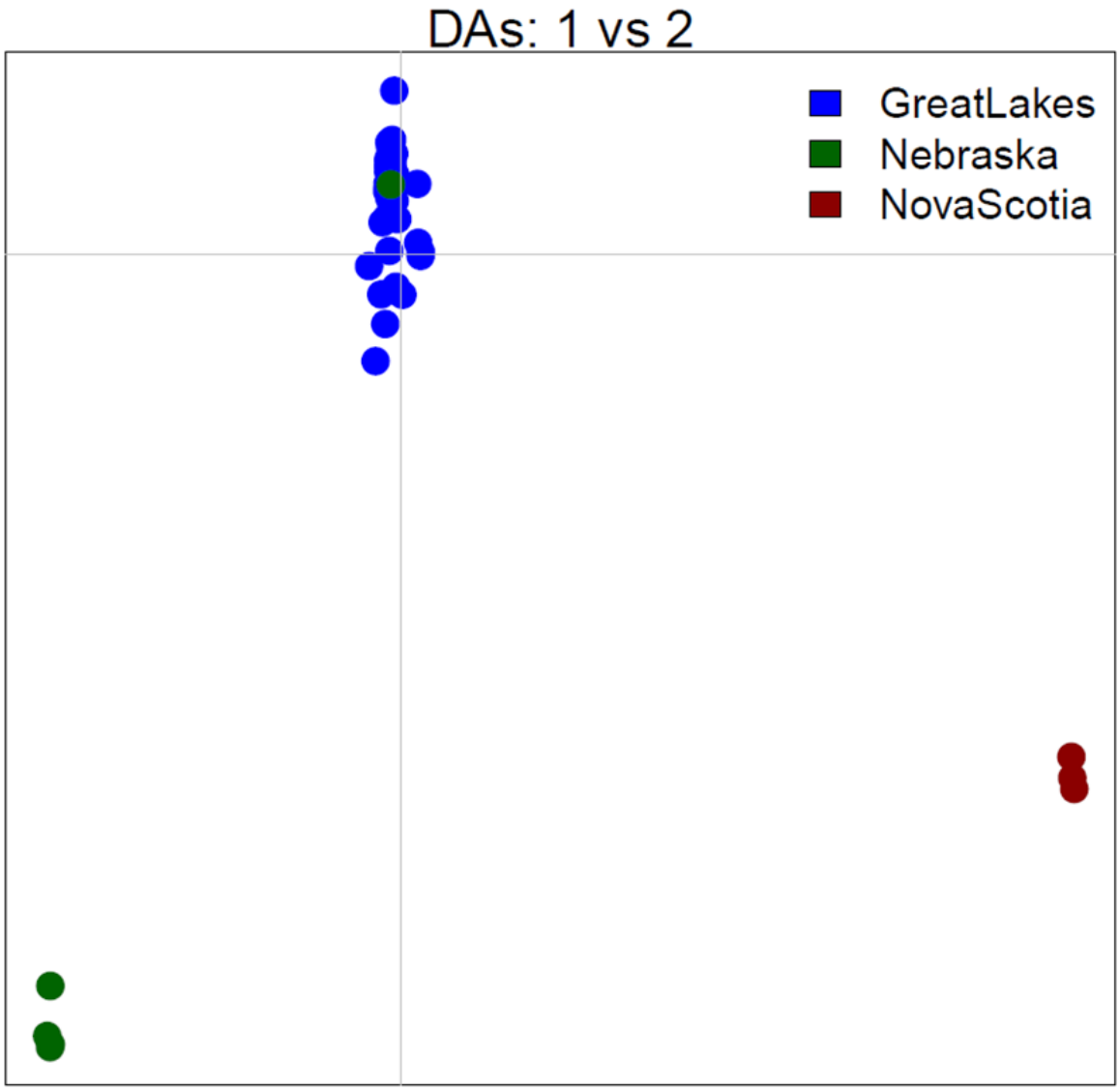


Figure 5: DAPC results from SambaR for FS-1 using Nebraska, the Great Lakes, and Nova Scotia as prior population assignments.

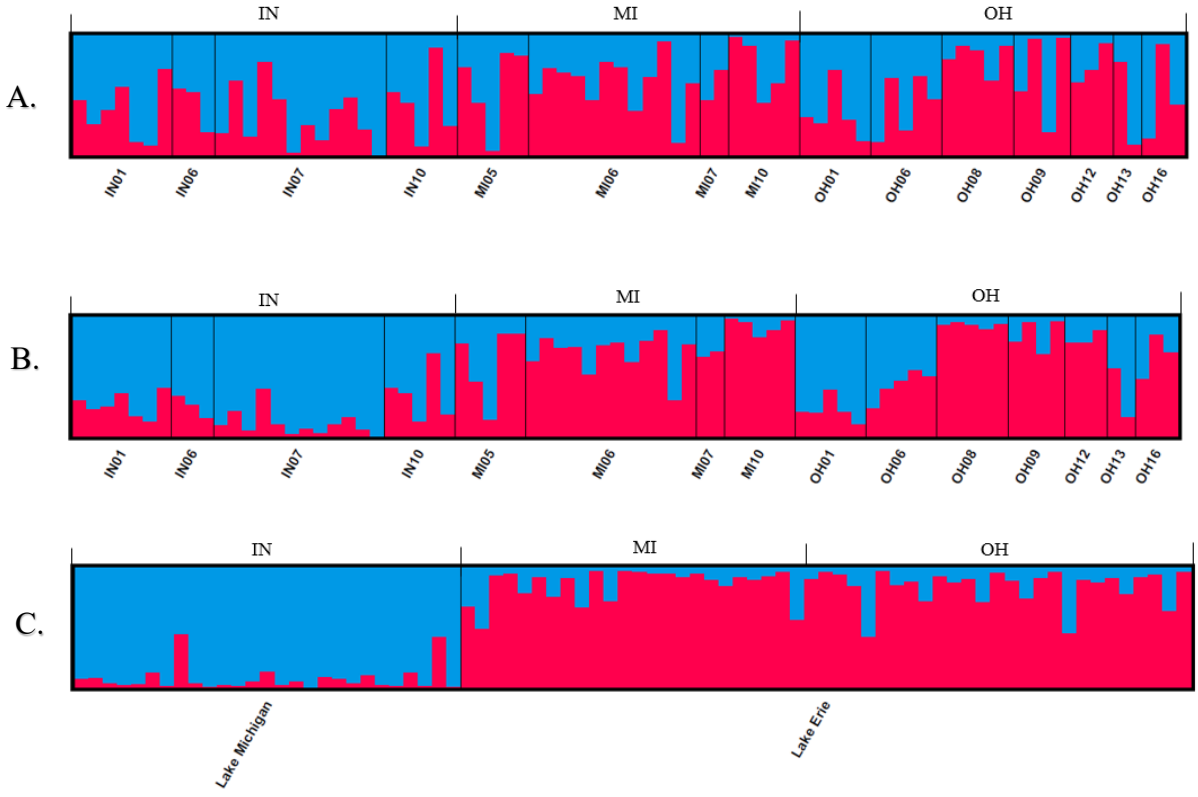


Figure 6: Cluster results from STRUCTURE. Results without LOCPRIOR across the 15 Midwest site localities based on K = 2 (A). Results with LOCPRIOR using the 15 Midwest site localities based on K = 2 (B). Results with LOCPRIOR using watershed information based on K = 2 (C). Blue bars represent cluster 1 and red bars represent cluster 2.

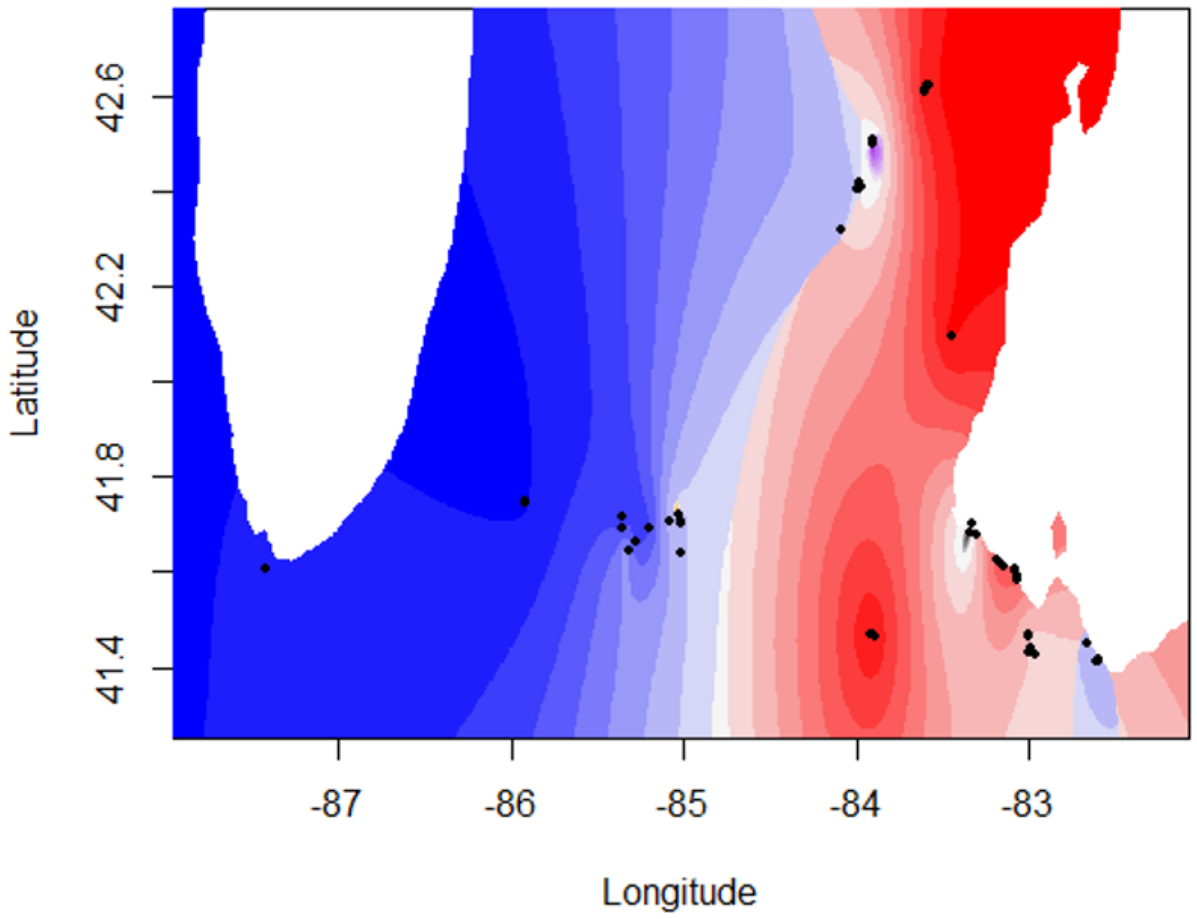


Figure 7: Results for FS-2 from tess3r incorporating geographic coordinates of individual samples. Blue layer represents cluster 1, red layer represents cluster 2, purple layer represents cluster 3, and the grey layer represents cluster 4.

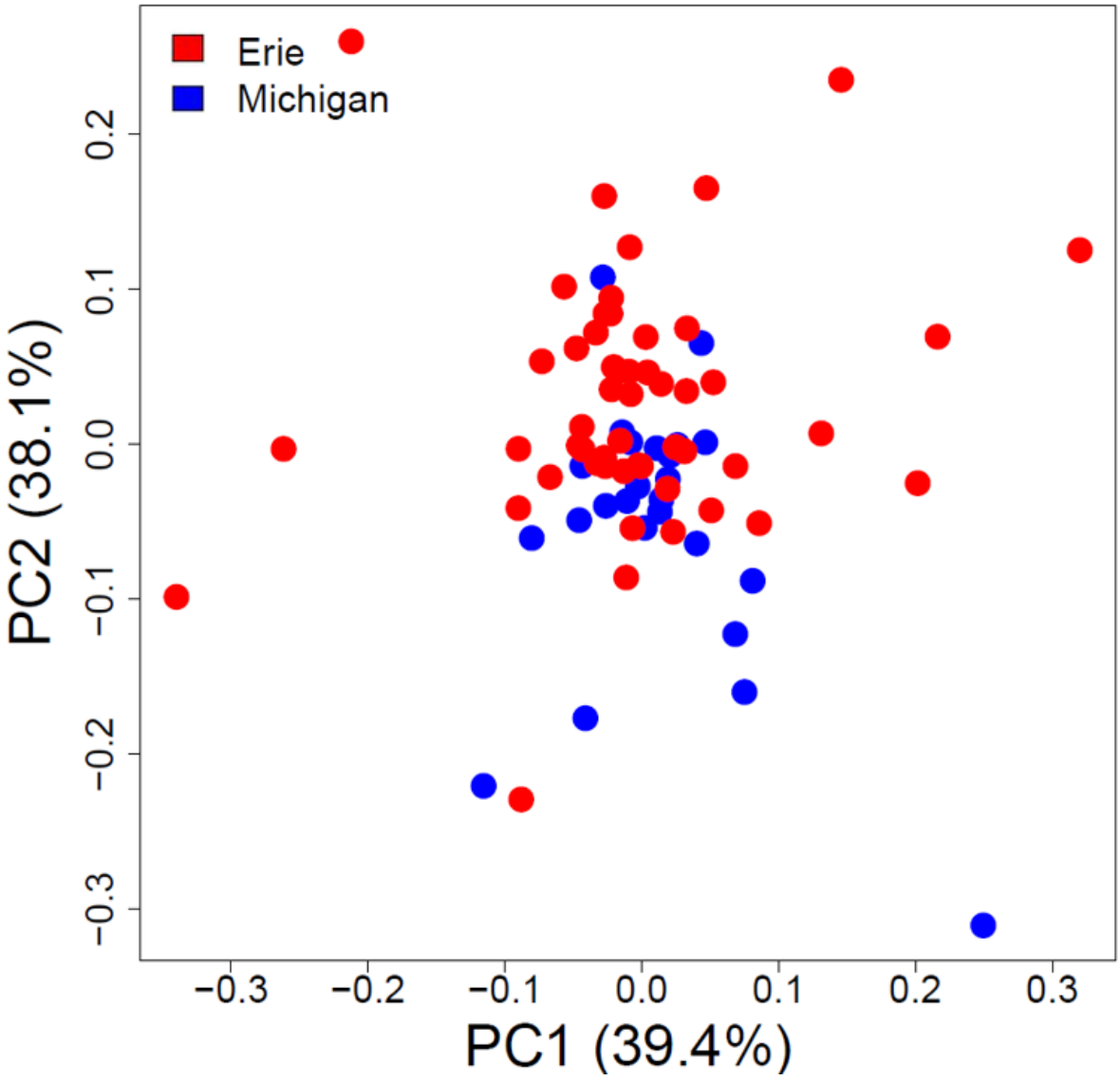


Figure 8: PCoA results from SambaR for FS-2 comparing Lake Michigan and Lake Erie watersheds.

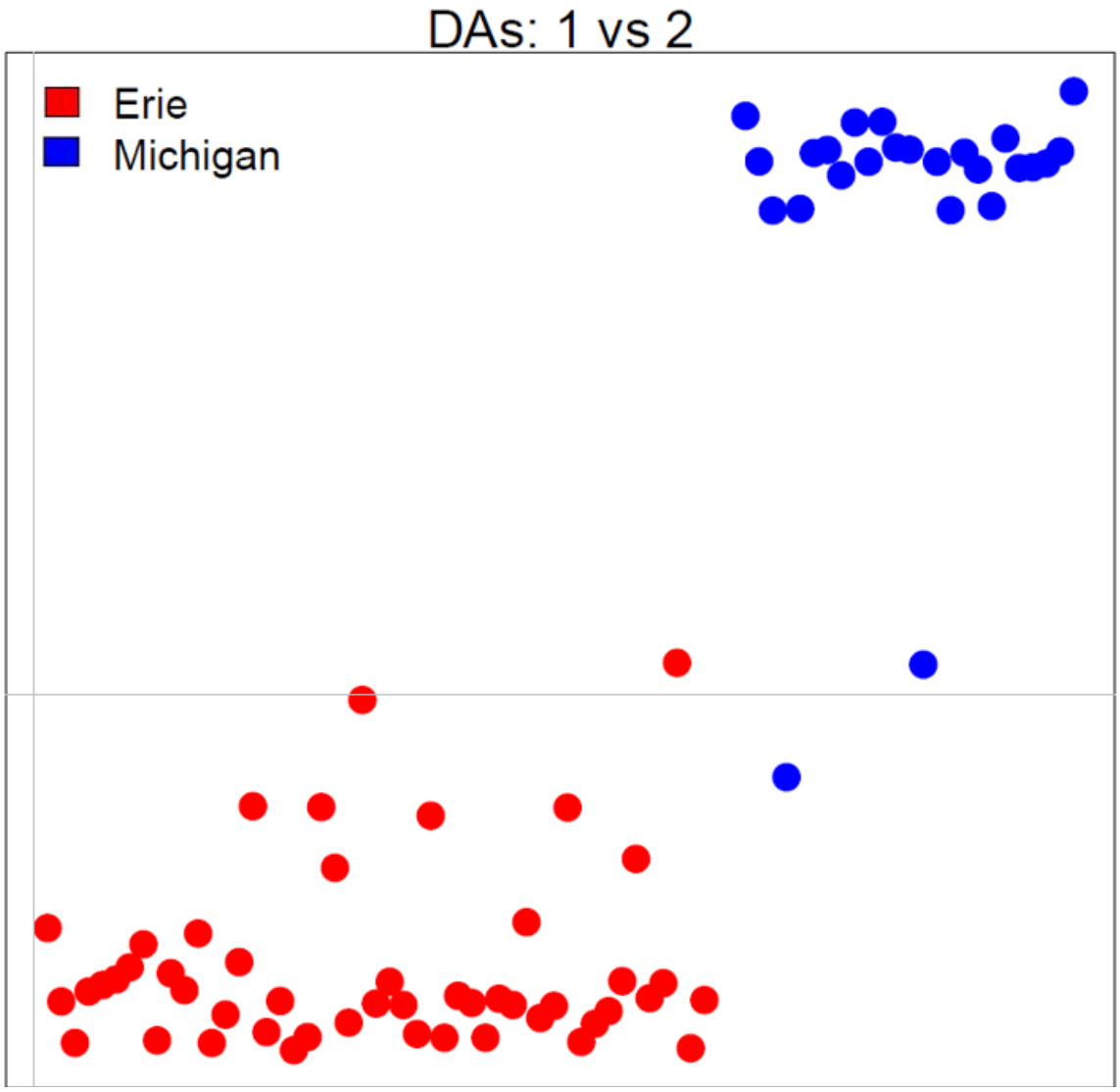


Figure 9: DAPC from SambaR for FS-2 using Lake Erie and Lake Michigan watershed designations as prior population assignments.

Table 2: Pairwise F_{ST} genetic differentiation results between clusters based on region. The F_{ST} value is shown below the diagonal and *P* value is shown above the diagonal.

Region	Nebraska	Great Lakes	Nova Scotia
Nebraska	-	< 0.001	< 0.001
Great Lakes	0.174	-	< 0.001
Nova Scotia	0.490	0.210	-

Table 3: Blanding's Turtle genetic diversity results based on all 17 sampling localities, three regions, and two watersheds. (N), total samples size; H_E, mean expected heterozygosity; H_O, mean observed heterozygosity; *Fis*, mean inbreeding coefficient; *SE*, standard error

Locality	N	Private Alleles	Polymorphic Loci	HO (SE)	HE (SE)	Fis (SE)
NE01	5	215	458	0.2470 (0.0052)	0.1318 (0.0053)	-0.0001 (0.0107)
IN01	7	217	742	0.2708 (0.0045)	0.1488 (0.0046)	0.0025 (0.0119)
IN06	3	41	222	0.3259 (0.0090)	0.1647 (0.0091)	-0.0060 (0.0093)
IN07	13	1188	2921	0.2989 (0.0026)	0.1895 (0.0027)	0.0194 (0.0118)
IN10	7	111	423	0.2822 (0.0060)	0.1496 (0.0061)	0.0011 (0.0119)
MI05	6	192	657	0.2772 (0.0048)	0.1515 (0.0049)	0.0026 (0.0107)
MI06	12	490	1528	0.2993 (0.0035)	0.1774 (0.0036)	0.0115 (0.0132)
MI07	2	77	260	0.2958 (0.0077)	0.1497 (0.0078)	0.0000 (0.0075)
MI10	6	892	2308	0.3118 (0.0029)	0.1834 (0.0030)	0.0070 (0.0089)
OH01	5	183	607	0.2689 (0.0048)	0.1423 (0.0049)	0.0005 (0.0098)
OH06	6	260	885	0.2793 (0.0041)	0.1501 (0.0042)	-0.0013 (0.0095)
OH08	4	76	386	0.2985 (0.0065)	0.1545 (0.0065)	-0.0075 (0.0103)
OH09	4	144	466	0.2678 (0.0055)	0.1425 (0.0056)	-0.0032 (0.0100)
OH12	3	106	392	0.2694 (0.0059)	0.1390 (0.0060)	-0.0002 (0.0087)
OH13	2	43	179	0.2739 (0.0089)	0.1409 (0.0089)	0.0032 (0.0075)
OH16	3	68	291	0.3174 (0.0078)	0.1627 (0.0078)	-0.0034 (0.0081)
NS01	3	231	396	0.2246 (0.0051)	0.1180 (0.0052)	-0.0004 (0.0085)
Nebraska	5	215	458	0.2470 (0.0052)	0.1318 (0.0053)	-0.0001 (0.0107)
Great Lakes	83	1929	8125	0.3523 (0.0030)	0.3026 (0.0016)	0.0510 (0.0422)

Table 3 continued.

Nova Scotia	3	231	396	0.2246 (0.0051)	0.1180 (0.0052)	-0.0004 (0.0085)
Lake Michigan	27	1582	3776	0.34278 (00483)	0.2324 (0.0027)	0.0257 (0.0186)
Lake Erie	51	2764	5758	0.3632 (0.0039)	0.2768 (0.0021)	0.0293 (0.0310)

Table 4: Pairwise F_{ST} genetic differentiation results between clusters based on watershed (K = 2). The F_{ST} value is shown below the diagonal and P value is shown above the diagonal.

Watershed	Lake Michigan	Lake Erie
Lake Michigan	-	< 0.001
Lake Erie	0.028	-

CHAPTER 4. DISCUSSION

4.1 Overview

This study used a RAD-seq method, 3RAD, to discover and genotype loci in Blanding's Turtle. Genetic analysis of identified SNPs found little to no structure across all 17 site but with analyses using prior population assignments of Nebraska, the Great Lakes, and Nova Scotia showing some genetic signal. Little to no structure was found within the Great Lakes region without using geographic coordinates or prior population localities. However, incorporating geographic coordinates and prior population information identified two genetic cluster that aligned with the watershed of origin. Preliminary analysis found very high missing data both per individual and per locus.

4.2 Dataset Characteristics

All RAD-seq datasets suffer from missing data to some extent due to allelic dropout and null alleles, errors arising from library preparation and sequencing, and variation in the depth of coverage among loci (Andrews et al., 2016; Arnold et al., 2013; Eaton et al., 2017; Gautier et al., 2013). Most studies choose to filter out the loci and/or individuals with the highest amounts of missing data, typically requiring around 50-80% of the individuals to have data for a given locus, otherwise that locus is discarded from analysis, or discarding individuals that lack data at more than 90% of the loci (Bernardi et al., 2016; Blanco-Bercial & Bucklin, 2016; Catchen, Bassham, et al., 2013; Jackson et al., 2014; Rodríguez-Ezpeleta et al., 2016; Van Wyngaarden et al., 2017).

Although these cutoff values are common, the choice is arbitrary and many studies have been successful in revealing significant genetic signals even with very high levels of missing data (Eaton et al., 2017; Hodel et al., 2017). For example, Tripp et al. (2017) found that the run that resulted in the best-supported and most fully resolved phylogenetic trees for *Petalidium* (Acanthaceae) used a dataset in which allowed for 90% missing data across 176,198 SNPs. The authors of these studies argue that allowing for high amounts of missing data can actually benefit RAD-seq studies by retaining a much larger number of SNPs, essentially overcoming the high missing data through a much greater number of markers (Eaton et al., 2017; Hodel et al., 2017;

Tripp et al., 2017). The numbers of SNPs used in this study (8,602 SNPs retained in FS-1 and 7,893 SNPs retained in FS-2) were relatively high compared to other freshwater turtle studies (Dresser, 2017; Gallego-García et al., 2019; Spinks et al., 2014; Xie et al., 2021). However, these numbers are rather low relative to other studies with a high amount of missing data which typically retained anywhere from around 25,000 to 180,000 SNPs (Eaton et al., 2017; Hodel et al., 2017; Tripp et al., 2017). This may have an impact on the results of this study, as the data gained from the retained number of SNPs may not be enough to overcome the high amounts of missing data.

The differences in the number of SNPs discovered in this study compared with other freshwater turtle studies is a result of the difference in both the study species and the number of samples being used. For example, although Dresser (2017) and Gallego-García et al. (2019) similiary utilized 3RAD and iPyRad for SNP discovery and bioinformatic filtering, they discovered 2,658 and 3,211 SNPs using 171 *Glyptemys muhlenbergii* samples and 175 *Mesoclemmys dahli* samples respectively. Differences in dataset preparation methods such as bioinformatic filtering and choice of parameters such as MAF can also impact the number of SNPs found and affect downstream analysis (Shafer et al., 2017). Another study used RAD-seq with 35 *Pelodiscus sinensis* samples but used SAMtools and bcftools for SNP discovery and filtering, finding just 105 high-quality SNPs (Xie et al., 2021).

To date, this is the first genomic study of Blanding's Turtle using a RADseq method. Locus identification and SNP calling was done *de novo* as no prior genomic information, or a closely related reference genome, were available to use. This also may have impacted the number of markers used in this study as having the ability to align sequences with a reference genome will usually result in the discovery of a larger number of SNPs (Catchen et al., 2013). Additionally, low and uneven samples sizes among many of the sites could have impacted genetic analysis and clustering results (Kalinowski, 2011; Shafer et al., 2017).

4.3 Population Structure and Genetic Diversity

Analysis comparing samples for FS-1 from Nebraska, Nova Scotia, and the Great Lakes region using both model based and model-free based methods revealed very little to no population structure in Blanding's Turtle across the range. PCoA of range-wide samples showed a very minor pattern, with samples from all three regions grouped close together with minimal separation of regions across the plot while DAPC results using prior population assignment based on region

found most individuals clustering within their respective geographic regions. STRUCTURE results of these samples found three clusters, however the high amount of admixture of the three clusters across all samples precluded assignment of any region to any one cluster.

Similarly, initial analysis focusing in on the Great Lakes region and including samples from Indiana, Michigan, and Ohio also showed little to no population structure using model based and model-free methods. Just as seen in the range-wide analysis, PCoA results of the Great Lakes samples did no exhibit enough of a pattern to assign samples or localities to any cluster. Similarly, basic STRUCTURE results revealed two clusters but the admixture exhibited across the samples made clustering assignment difficult.

These results disagree with recent work done using microsatellites and mitochondrial and nuclear loci that found a clear separation between the Great Lakes region and eastern North America, with a possible third separation between the Great Lakes Region and Nebraska (Jordan et al., 2019; Mockford et al., 2007). Although my results were somewhat surprising, there have been previous genetic studies conducted on a smaller scale on populations of Blanding's Turtle in areas such as Chicago, Illinois, Ann Arbor, Michigan, and Nova Scotia to New York that also found very little to no genetic structures among localities (Howes et al., 2009; McGuire et al., 2013; Mockford et al., 2005; Rubin et al., 2001).

However, using the LOCPRIOR parameter for FS-2 in STRUCTURE incorporates sampling locations for each individual as prior population information (Pritchard et al., 2010). This can be useful in analysis of datasets with a very weak structure signal that standard structure models may fail to detect (Pritchard et al., 2010). Employing the LOCPRIOR parameter for the Great Lakes dataset displayed less admixture across all samples and allowed for stronger inferences on clustering assignment. Using sample localities for prior population assignment revealed two clusters following a pattern somewhat consistent with watershed designation of sample localities. Hence, another STRUCTURE run was conducted using the Lake Michigan and Lake Erie watersheds as prior population assignments, revealing two clusters with much less admixture allowing for a much stronger inference of cluster assignment. DAPC analysis affirmed these results by incorporating watershed as prior population assignment, showing strong clustering of Indiana samples into the Lake Michigan watershed and Michigan and Ohio samples into the Lake Erie watershed. Furthermore, integrating exact geographic coordinates of individual samples into clustering analysis using tess3r revealed four clusters. Despite the larger amount of clusters, the

majority of samples concurred with the LOCPRIOR and DAPC analyses with most samples following the pattern of clustering based on watershed.

Analyses focused on genetic differentiation of Nebraska, the Great Lakes, and Nova Scotia based on *F*st reported vastly different results from clustering and admixture analyses. These results were closer in agreement with previous work done by Jordan et al. (2019) in showing that the Great Lakes region is distinct from both Nebraska and Nova Scotia but more similar to Nebraska. Unsurprisingly, Nebraska and Nova Scotia were much more highly differentiated from each other than from the Great Lakes. Expected heterozygosity estimates were low across all localities supporting previous work showing low genetic variation in Blanding's Turtle using mitochondrial and nuclear loci (Jordan et al., 2019).

Similar analyses conducted within the Great Lakes region showed a significant, but rather small, differentiation between the Lake Michigan watershed and the Lake Erie watershed. These results support previous findings by Sethuraman et al. (2014) that watersheds may be an important factor to consider when trying to detect genetic signals between populations. This pattern and the low genetic variation detected may be a product of population bottlenecks following the displacement of *E. blandingii* populations during the Pleistocene resulting from glacial advance during glaciation events, with the most recent being the Wisconsin Glaciation period (Hewitt, 2000). Studies have shown *E. blandingii* may have had smaller sources of refugia than other turtle species during these glacial events, which may have led to reduced genetic variation across the range (Rödder et al., 2013). The subsequent northward recolonization events may have then aligned with the Lake Michigan and Lake Erie watersheds created from the retreat of the Wisconsin Glacier, with more limited gene flow occurring across watershed boundaries (Larson & Schaetzl, 2001).

Another possible explanation for the limited genetic variations and structure seen in this study may be the evolution rate of the chosen marker. SNPs tend to have a slower rate of evolution and a lower degree of polymorphism than other markers such as microsatellites, which are known for their fast mutation rate and high degree of polymorphism (Hess et al., 2011; Narum et al., 2008; Vignal et al., 2002). This coupled with the long estimated generation time of 37 years for Blanding's Turtle (Congdon et al., 1993) as well as small population sizes may lead to stronger genetic drift occurring for more quickly evolving markers than for SNPs.

4.4 Management Implications and Future Directions

Translocation efforts of reptiles can be a useful tool when motivated by conservation, and supplementation of Blanding's Turtle populations in northwestern Illinois via head-starting programs have proven to be successful (Germano & Bishop, 2009; Thompson et al., 2020). However, localities for potential source populations must be carefully considered, as movement of individuals between populations with a deeper, historical divergence carries the risk of decreased fitness per increased outbreeding depression.

Although minimal population structure was revealed across the range in this study, a signal of differentiation among the easternmost site of Nova Scotia, the Great Lakes region, and the western-most site of Nebraska was detected and should be considered when designing management plans for Blanding's Turtle. This is particularly relevant for management at sites near the confluence of these major regions, where adaptive difference in populations across the boundary may preclude supplementation via translocation. Similar considerations should be given to watershed boundaries within the Great Lakes region, as genetic differentiation between localities within the Lake Michigan watershed and the Lake Erie watershed was detected.

Further studies using RADseq to discover genome-wide SNPs in Blanding's Turtle are needed to gain a clearer picture on the effectiveness of this technique for this species. Priority should be given to reducing the amount of missing data in the dataset and increasing the number of SNPs discovered, as this may result in greater coverage and stronger genetic signals. This may be overcome by using high quality DNA samples, as RADseq approaches tend to perform poorly with highly degraded genomic DNA (Graham et al., 2015). Another approach may be to design baits using a method called RADcap (Hoffberg et al., 2016a, 2016b). RADcap works by first running 3RAD on a small number of samples to identify variable loci (Hoffberg et al., 2016a, 2016b). Then, custom sequence baits are designed to enrich the candidate SNPs found among these loci across many individuals (Hoffberg et al., 2016a, 2016b). 3RAD is then rerun with all of the samples, resulting in lower rates of PCR duplicates and missing data (Hoffberg et al., 2016a, 2016b). The creation and use of a reference genome of either this species or of a closely related species, such as Western Pond Turtle (*Actinemys marmorata*), may also prove useful, as being able to align sequences with a reference genome tends to result in the discovery of a larger number of SNPs (Catchen et al., 2013).

4.5 Conclusion

To date, this is the first study on Blanding's Turtle using genome-wide SNPs discovered using a RAD-sequencing method. Low genetic variation was detected across the range and within the Great Lakes regions. Weak signals of population structure were revealed among the localities in Nebraska, the Great Lakes, and Nova Scotia and between the Lake Michigan and Lake Erie watersheds in the Great Lakes region. These results should be considered in management plans for this species, especially in areas where the regions or watersheds meet.

High amounts of missing data were present across all samples and loci and a relatively low number of SNPs were discovered which may be having a negative impact on the results of this study. Futures genomic studies of Blanding's Turtle should prioritize obtaining a larger number of SNPs and decreasing the amount of missing data in the dataset.

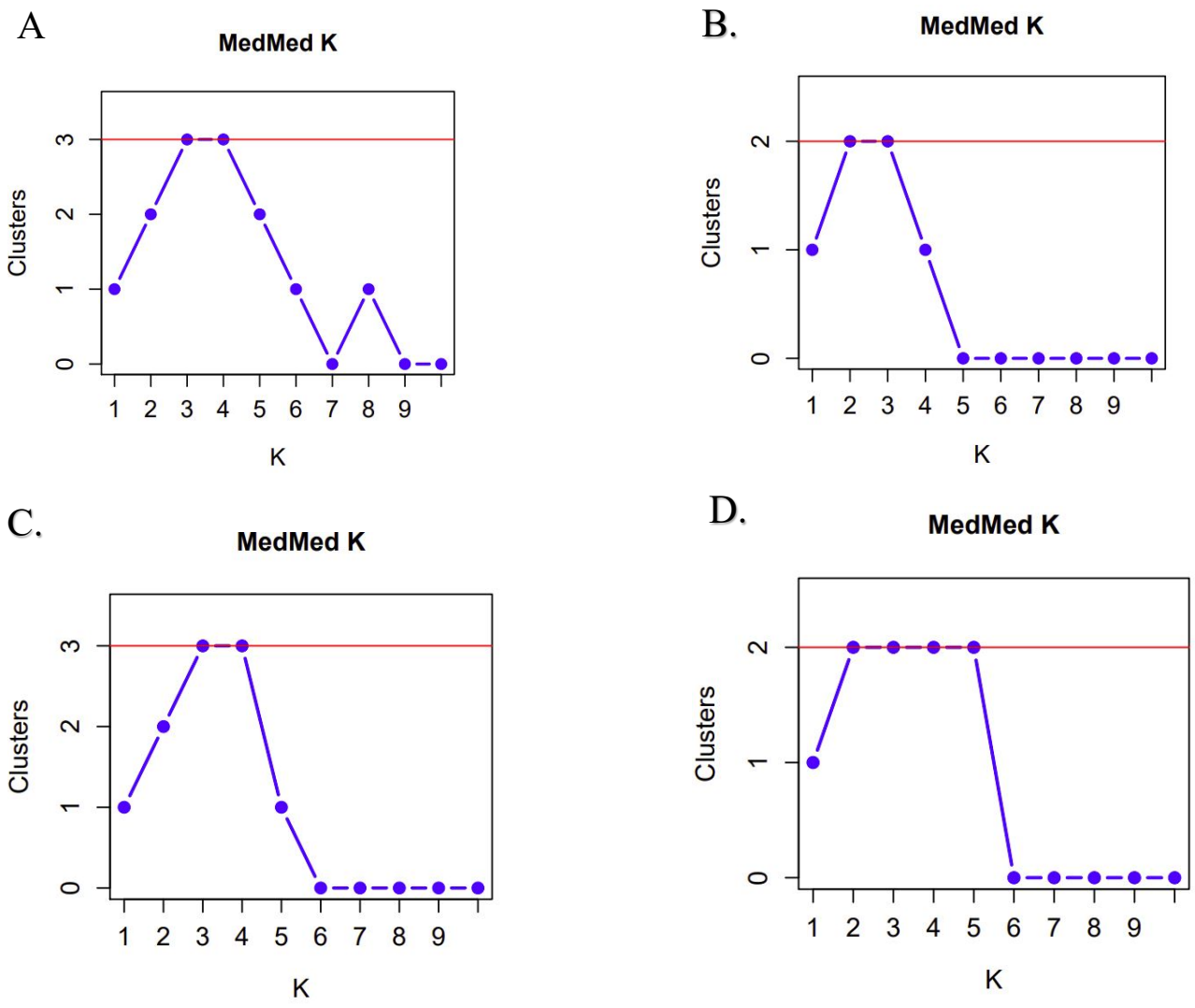
APPENDIX A. SAMPLES PER LOCALITY

Sample localities with location codes for pre-filtered Blanding's Turtle samples. N represents the total number of samples for each locality.

Sample Locality		N
Nebraska		
NE01		5
Indiana		
IN01		7
IN06		3
IN07		13
IN10		6
IN12		2
Michigan		
MI01		1
MI05		5
MI06		13
MI07		2
MI10		5
MI12		1
MI13		1
Ohio		
OH01		4
OH06		5
OH08		5
OH09		5
OH12		3
OH13		2
OH16		4
OH17		1
Nova Scotia		
NS01		3

APPENDIX B. BEST-FIT K VALUE

StructureSelector results using Puechmaille MedMedK plots for determining the best-fit K value through. Results for the FS-1 (K = 3) (A). Results for the Great Lakes region without LocPrior (K = 2) (B). Results for the Great Lakes region using LocPrior based on sample locality (K = 3) (C). Results for the Great Lakes region using LocPrior based on watershed location (K = 2) (D).



REFERENCES

Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews Genetics*, *11*(10), 697-709.

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, *17*(2), 81.

Arnold, B., Corbett-Detig, R. B., Hartl, D., & Bomblies, K. (2013). RAD seq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology*, *22*(11), 3179-3190.

Ashley, M. V., Melnick, D. J., & Western, D. (1990). Conservation genetics of the black rhinoceros (*Diceros bicornis*), I: evidence from the mitochondrial DNA of three populations. *Conservation Biology*, *4*(1), 71-77.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A., & Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*, *3*(10), e3376.

Bayona-Vásquez, N. J., Glenn, T. C., Kieran, T. J., Pierson, T. W., Hoffberg, S. L., Scott, P. A., Bentley, K. E., Finger, J. W., Louha, S., & Troendle, N. (2019). Adapterama III: Quadruple-indexed, double/triple-enzyme RADseq libraries (2RAD/3RAD). *Peerj,* 7:e7724

Bernardi, G., Azzurro, E., Golani, D., & Miller, M. R. (2016). Genomic signatures of rapid adaptive evolution in the bluespotted cornetfish, a Mediterranean Lessepsian invader. *Molecular Ecology*, *25*(14), 3384-3396.

Blanco-Bercial, L., & Bucklin, A. (2016). New view of population genetics of zooplankton: RAD-seq analysis reveals population structure of the North Atlantic planktonic copepod *Centropages typicus*. *Molecular Ecology*, *25*(7), 1566-1580.

Brecke, B., & Moriarty, J. (1989). *Emydoidea blandingii* (Blanding's Turtle) longevity. *Herpetological Review*, *20*(2), 53.

Catchen, J., Bassham, S., Wilson, T., Currey, M., O'Brien, C., Yeates, Q., & Cresko, W. A. (2013). The population structure and recent colonization history of O regon threespine stickleback determined using restriction-site associated DNA-sequencing. *Molecular Ecology*, *22*(11), 2864-2883.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124-3140.

Caye, K., Deist, T. M., Martins, H., Michel, O., & François, O. (2016). TESS3: fast inference of spatial population structure and genome scans for selection. *Molecular Ecology Resources*, *16*(2), 540-548.

Charlesworth, D., & Willis, J. H. (2009). The genetics of inbreeding depression. *Nature Reviews Genetics*, *10*(11), 783-796.

Congdon, J., Nagle, R., Kinney, O., & van Loben Sels, R. (2001). Hypotheses of aging in a long-lived vertebrate, Blanding's turtle (*Emydoidea blandingii*). *Experimental Gerontology*, *36*(4-6), 813-827.

Congdon, J. D., Dunham, A. E., & van Loben Sels, R. (1993). Delayed sexual maturity and demographics of Blanding's turtles (*Emydoidea blandingii*): implications for conservation and management of long-lived organisms. *Conservation Biology*, *7*(4), 826-833.

Davy, C. M., Bernardo, P. H., & Murphy, R. W. (2014). A Bayesian approach to conservation genetics of Blanding's turtle (*Emys blandingii*) in Ontario, Canada. *Conservation Genetics*, *15*(2), 319-330.

De Cara, M., Fernández, J., Toro, M., & Villanueva, B. (2011). Using genome-wide information to minimize the loss of diversity in conservation programmes. *Journal of Animal Breeding and Genetics*, *128*(6), 456-464.

de Jong, M. J., de Jong, J. F., Hoelzel, A. R., & Janke, A. (2021). SambaR: An R package for fast, easy and reproducible population-genetic analyses of biallelic SNP data sets. *Molecular Ecology Resources*, *21*(4), 1369-1379.

DeFaveri, J., Viitaniemi, H., Leder, E., & Merilä, J. (2013). Characterizing genic and nongenic molecular markers: Comparison of microsatellites and SNP s. *Molecular Ecology Resources*, *13*(3), 377-392.

Dresser, C. M. (2017). Assessment of genetic and education recovery plan objectives for the Bog Turtle (*Glyptemys muhlenbergii*). Ph.D. dissertation, The University of Tennessee, Knoxville.

Eaton, D. A. (2014). PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, *30*(13), 1844-1849.

Eaton, D. A., & Overcast, I. (2020). ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics*, *36*(8), 2592-2594.

Eaton, D. A., Spriggs, E. L., Park, B., & Donoghue, M. J. (2017). Misconceptions on missing data in RAD-seq phylogenetics with a deep-scale example from flowering plants. *Systematic Biology*, *66*(3), 399-412.

Ellegren, H., Primmer, C. R., & Sheldon, B. C. (1995). Microsatellite 'evolution': directionality or bias? *Nature Genetics*, *11*(4), 360-362.

Emerson, K. J., Merz, C. R., Catchen, J. M., Hohenlohe, P. A., Cresko, W. A., Bradshaw, W. E., & Holzapfel, C. M. (2010). Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences*, *107*(37), 16196-16200.

Ernst, C. H., Ernst, C. H., & Lovich, J. E. (2009). *Turtles of the United States and Canada*. JHU Press.

Fauvelot, C., Bernardi, G., & Planes, S. (2003). Reductions in the mitochondrial DNA diversity of coral reef fish provide evidence of population bottlenecks resulting from Holocene sea-level change. *Evolution*, *57*(7), 1571-1583.

Frankham, R. (1995). Conservation genetics. *Annual Review of Genetics*, *29*(1), 305-327.

Frankham, R. (2003). Genetics and conservation biology. *Comptes Rendus Biologies*, *326*, 22-29.

Gallego-García, N., Forero-Medina, G., Vargas-Ramírez, M., Caballero, S., & Shaffer, H. B. (2019). Landscape genomic signatures indicate reduced gene flow and forest-associated adaptive divergence in an endangered neotropical turtle. *Molecular Ecology*, *28*(11), 2757-2771.

García-Moreno, J., Matocq, M. D., Roy, M. S., Geffen, E., & Wayne, R. K. (1996). Relationships and genetic purity of the endangered Mexican wolf based on analysis of microsatellite loci. *Conservation Biology*, *10*(2), 376-389.

Gautier, M., Gharbi, K., Cezard, T., Foucaud, J., Kerdelhué, C., Pudlo, P., Cornuet, J. M., & Estoup, A. (2013). The effect of RAD allele dropout on the estimation of genetic variation within and between populations. *Molecular Ecology*, *22*(11), 3165-3178.

Germano, J. M., & Bishop, P. J. (2009). Suitability of amphibians and reptiles for translocation. *Conservation Biology*, *23*(1), 7-15.

Gibbs, J. P. (1998). Genetic structure of redback salamander *Plethodon cinereus* populations in continuous and fragmented forests. *Biological Conservation*, *86*(1), 77-81.

Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, *53*(3-4), 325-338.

Graham, C. F., Glenn, T. C., McArthur, A. G., Boreham, D. R., Kieran, T., Lance, S., Manzon, R. G., Martino, J. A., Pierson, T., & Rogers, S. M. (2015). Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADS eq). *Molecular Ecology Resources*, *15*(6), 1304-1315.

Hess, J., Matala, A., & Narum, S. (2011). Comparison of SNPs and microsatellites for fine-scale application of genetic stock identification of Chinook salmon in the Columbia River Basin. *Molecular Ecology Resources*, *11*, 137-149.

Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, *405*(6789), 907-913.

Hodel, R. G., Chen, S., Payton, A. C., McDaniel, S. F., Soltis, P., & Soltis, D. E. (2017). Adding loci improves phylogeographic resolution in red mangroves despite increased missing data: comparing microsatellites and RAD-Seq and investigating loci filtering. *Scientific Reports*, *7*(1), 1-14.

Hoffberg, S. L., Kieran, T. J., Catchen, J. M., Devault, A., Faircloth, B. C., Mauricio, R., & Glenn, T. C. (2016a). Adapterama IV: Sequence capture of dual-digest RADseq libraries with identifiable duplicates (RADcap). *bioRxiv*, 044651.

Hoffberg, S. L., Kieran, T. J., Catchen, J. M., Devault, A., Faircloth, B. C., Mauricio, R., & Glenn, T. C. (2016b). RAD cap: sequence capture of dual-digest RAD seq libraries with identifiable duplicates and reduced missing data. *Molecular Ecology Resources*, *16*(5), 1264-1278.

Howes, B. J., Brown, J. W., Gibbs, H. L., Herman, T. B., Mockford, S. W., Prior, K. A., & Weatherhead, P. J. (2009). Directional gene flow patterns in disjunct populations of the black ratsnake (*Pantheropis obsoletus*) and the Blanding's turtle (*Emydoidea blandingii*). *Conservation Genetics*, *10*(2), 407-417.

Jackson, A. M., Semmens, B. X., De Mitcheson, Y. S., Nemeth, R. S., Heppell, S. A., Bush, P. G., Aguilar-Perera, A., Claydon, J. A., Calosso, M. C., & Sealey, K. S. (2014). Population structure and phylogeography in Nassau grouper (*Epinephelus striatus*), a mass-aggregating marine fish. *PLoS One*, *9*(5), e97508.

Jarne, P., & Lagoda, P. J. (1996). Microsatellites, from molecules to populations and back. *Trends in Ecology & Evolution*, *11*(10), 424-429.

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, *11*(1), 1-15.

Jonker, R. M., Zhang, Q., Van Hooft, P., Loonen, M. J., Van der Jeugd, H. P., Crooijmans, R. P., Groenen, M. A., Prins, H. H., & Kraus, R. H. (2012). The development of a genome wide SNP set for the Barnacle goose *Branta leucopsis. PLoS One*, *7*(7), e38412.

Jordan, M. A., Mumaw, V., Millspaw, N., Mockford, S. W., & Janzen, F. J. (2019). Range-wide phylogeography of Blanding's Turtle [Emys (= *Emydoidea) blandingii*]. *Conservation Genetics*, *20*(3), 419-430.

Kalinowski, S. T. (2011). The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure. *Heredity*, *106*(4), 625-632.

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, *15*(5), 1179-1191.

Lacy, R. C. (1987). Loss of genetic diversity from managed populations: interacting effects of drift, mutation, immigration, selection, and population subdivision. *Conservation Biology*, *1*(2), 143-158.

Larson, G., & Schaetzl, R. (2001). Origin and evolution of the Great Lakes. *Journal of Great Lakes Research*, *27*(4), 518-546.

Lemopoulos, A., Prokkola, J. M., Uusi-Heikkilä, S., Vasemägi, A., Huusko, A., Hyvärinen, P., Koljonen, M. L., Koskiniemi, J., & Vainikka, A. (2019). Comparing RADseq and microsatellites for estimating genetic diversity and relatedness—Implications for brown trout conservation. *Ecology and Evolution*, *9*(4), 2106-2120.

Li, Y. L., & Liu, J. X. (2018). StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. *Molecular Ecology Resources*, *18*(1), 176-177.

Lischer, H. E., & Excoffier, L. (2012). PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, *28*(2), 298-299.

Liu, N., Chen, L., Wang, S., Oh, C., & Zhao, H. (2005). Comparison of single-nucleotide polymorphisms and microsatellites in inference of population structure. *BMC Genetics*

Luikart, G., England, P. R., Tallmon, D., Jordan, S., & Taberlet, P. (2003). The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, *4*(12), 981-994.

McGuire, J. M., Scribner, K. T., & Congdon, J. D. (2013). Spatial aspects of movements, mating patterns, and nest distributions influence gene flow among population subunits of Blanding's turtles (*Emydoidea blandingii*). *Conservation Genetics*, *14*(5), 1029-1042.

Meffe, G. K., Carroll, C. R., & Ralls, K. (1995). Principles of conservation biology. *Trends in Ecology and Evolution*, *10*(9), 387.

Mockford, S., Herman, T., Snyder, M., & Wright, J. M. (2007). Conservation genetics of Blanding's turtle and its application in the identification of evolutionarily significant units. *Conservation Genetics*, *8*(1), 209-219.

Mockford, S., McEachern, L., Herman, T., Snyder, M., & Wright, J. M. (2005). Population genetic structure of a disjunct population of Blanding's turtle (*Emydoidea blandingii*) in Nova Scotia, Canada. *Biological Conservation*, *123*(3), 373-380.

Morin, P. A., Luikart, G., & Wayne, R. K. (2004). SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution*, *19*(4), 208-216.

Moritz, C. (1994). Defining 'evolutionarily significant units' for conservation. *Trends in Ecology & Evolution*, *9*(10), 373-375.

Narum, S., Banks, M., Beacham, T., Bellinger, M., Campbell, M., Dekoning, J., Elz, A., Guthrie Iii, C., Kozfkay, C., & Miller, K. (2008). Differentiating salmon populations at broad and fine geographical scales with microsatellites and single nucleotide polymorphisms. *Molecular Ecology*, *17*(15), 3464-3477.

Porras-Hurtado, L., Ruiz, Y., Santos, C., Phillips, C., Carracedo, Á., & Lareu, M. (2013). An overview of STRUCTURE: applications, parameter settings, and supporting software. *Frontiers in Genetics*, *4*, 98.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945-959.

Pritchard, J. K., Wen, W., & Falush, D. (2010). Documentation for STRUCTURE software: Version 2. *University of Chicago, Chicago, IL*.

Prosser, M. R., Lisle Gibbs, H., & Weatherhead, P. J. (1999). Microgeographic population genetic structure in the northern water snake, *Nerodia sipedon sipedon* detected using microsatellite DNA loci. *Molecular Ecology*, *8*(2), 329-333.

Puechmaille, S. J. (2016). The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Molecular Ecology Resources*, *16*(3), 608-627.

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

Ralls, K., Ballou, J. D., & Templeton, A. (1988). Estimates of lethal equivalents and the cost of inbreeding in mammals. *Conservation Biology*, *2*(2), 185-193.

Reid, B. N., Mladenoff, D. J., & Peery, M. Z. (2017). Genetic effects of landscape, habitat preference and demography on three co-occurring turtle species. *Molecular Ecology*, *26*(3), 781-798.

Rödder, D., Lawing, A. M., Flecks, M., Ahmadzadeh, F., Dambach, J., Engler, J. O., Habel, J. C., Hartmann, T., Hörnes, D., & Ihlow, F. (2013). Evaluating the significance of paleophylogeographic species distribution models in reconstructing Quaternary range-shifts of Nearctic chelonians. *PLoS One*, *8*(10).

Rodder, D., Lawing, A. M., Flecks, M., Ahmadzadeh, F., Dambach, J., Engler, J. O., Habel, J. C., Hartmann, T., Hornes, D., Ihlow, F., Schidelko, K., Stiels, D., & Polly, P. D. (2013). Evaluating the significance of paleophylogeographic species distribution models in reconstructing quaternary range-shifts of nearctic chelonians. *PLoS One*, *8*(10), e72855. https://doi.org/10.1371/journal.pone.0072855

Rodríguez-Ezpeleta, N., Bradbury, I. R., Mendibil, I., Álvarez, P., Cotano, U., & Irigoien, X. (2016). Population structure of Atlantic mackerel inferred from RAD-seq-derived SNP markers: Effects of sequence clustering parameters and hierarchical SNP selection. *Molecular Ecology Resources*, *16*(4), 991-1001.

Roesti, M., Salzburger, W., & Berner, D. (2012). Uninformative polymorphisms bias genome scans for signatures of selection. *BMC Evolutionary Biology*, *12*(1), 1-7.

Rubin, C. S., Warner, R. E., Bouzat, J. L., & Paige, K. N. (2001). Population genetic structure of Blanding's turtles (*Emydoidea blandingii*) in an urban landscape. *Biological Conservation*, *99*(3), 323-330.

Schmidt, K. P. (1938). Herpetological evidence for the postglacial eastward extension of the steppe in North America. *Ecology*, *19*(3), 396-407.

Selkoe, K. A., & Toonen, R. J. (2006). Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters*, *9*(5), 615-629.

Sethuraman, A., McGaugh, S. E., Becker, M. L., Chandler, C. H., Christiansen, J. L., Hayden, S., LeClere, A., Monson-Miller, J., Myers, E. M., & Paitz, R. T. (2014). Population genetics of Blanding's turtle (*Emys blandingii*) in the midwestern United States. *Conservation Genetics*, *15*(1), 61-73.

Shafer, A. B., Peart, C. R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C. W., & Wolf, J. B. (2017). Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods in Ecology and Evolution*, *8*(8), 907-917.

Spinks, P. Q., Thomson, R. C., & Shaffer, H. B. (2014). The advantages of going large: genome-wide SNP s clarify the complex population history and systematics of the threatened western pond turtle. *Molecular Ecology*, *23*(9), 2228-2241.

Standing, K. L., Herman, T. B., & Morrison, I. P. (1999). Nesting ecology of Blanding's turtle (*Emydoidea blandingii*) in Nova Scotia, the northeastern limit of the species' range. *Canadian Journal of Zoology*, *77*(10), 1609-1614.

Streicher, J. W., Cox, C. L., Campbell, J. A., Smith, E. N., & de Sá, R. O. (2012). Rapid range expansion in the Great Plains narrow-mouthed toad (*Gastrophryne olivacea*) and a revised taxonomy for North American microhylids. *Molecular Phylogenetics and Evolution*, *64*(3), 645-653.

Templeton, A. R., Shaw, K., Routman, E., & Davis, S. K. (1990). The genetic consequences of habitat fragmentation. *Annals of the Missouri Botanical Garden*, 13-27.

Thompson, D., Glowacki, G., Ludwig, D., Reklau, R., Kuhns, A. R., Golba, C. K., & King, R. (2020). Benefits of Head-starting for Blanding's Turtle Size Distributions and Recruitment. *Wildlife Society Bulletin*, *44*(1), 57-67.

Tripp, E. A., Tsai, Y. H. E., Zhuang, Y., & Dexter, K. G. (2017). RAD seq dataset with 90% missing data fully resolves recent radiation of Petalidium (Acanthaceae) in the ultra-arid deserts of Namibia. *Ecology and Evolution*, *7*(19), 7920-7936.

Van Wyngaarden, M., Snelgrove, P. V., DiBacco, C., Hamilton, L. C., Rodríguez-Ezpeleta, N., Jeffery, N. W., Stanley, R. R., & Bradbury, I. R. (2017). Identifying patterns of dispersal, connectivity and selection in the sea scallop, *Placopecten magellanicus*, using RAD seq-derived SNP s. *Evolutionary Applications*, *10*(1), 102-117.

Vignal, A., Milan, D., SanCristobal, M., & Eggen, A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution*, *34*(3), 275-305.

Wayne, R. K., Lehman, N., Allard, M. W., & Honeycutt, R. L. (1992). Mitochondrial DNA variability of the gray wolf: genetic consequences of population decline and habitat fragmentation. *Conservation Biology*, *6*(4), 559-569.

Wright, J. M., & Bentzen, P. (1995). Microsatellites: genetic markers for the future. In *Molecular Genetics in Fisheries* (pp. 117-121). Springer.

Xie, Q., Liu, F., Zhang, J., Li, X., Chen, T., Fang, G., Ma, R., & Su, S.-P. (2021). Development of 105 SNP markers in endangered turtle species *Pelodiscus sinensis* using RAD-seq.

Yang, M. Q., Athey, B. D., Arabnia, H. R., Sung, A. H., Liu, Q., Yang, J. Y., Mao, J., & Deng, Y. (2009). High-throughput next-generation sequencing technologies foster new cutting-edge computing techniques in bioinformatics. *BMC Genomics*, *10 Suppl 1*, I1. https://doi.org/10.1186/1471-2164-10-S1-I1