# POPULATION GENOMICS OF KIRTLAND'S SNAKE (CLONOPHIS KIRTLANDII)

by

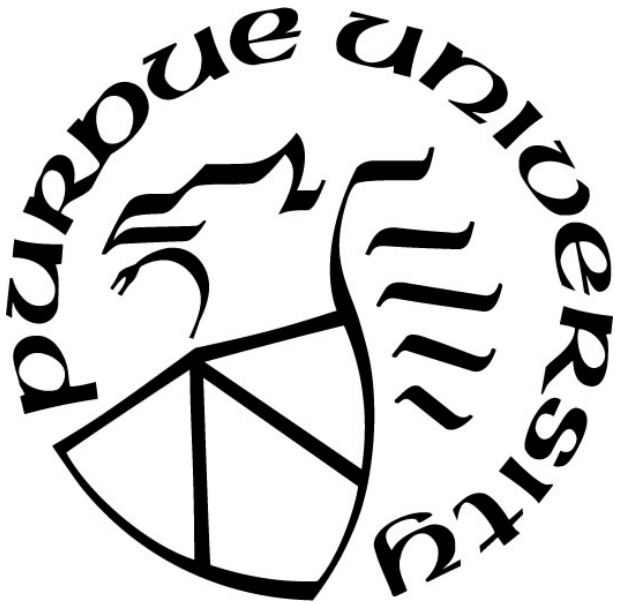
**Alanna D Noland**

**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

Fort Wayne, Indiana

August 2019

# 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

**Approved by:**

Dr. Jordan M. Marshall
Head of the Graduate Program

*For my dad and mom, Don and Kimberly*

# ACKNOWLEDGMENTS

I would like to thank my committee chair, Dr. Jordan for assistance and suggestions. I would also like to thank my committee members Dr. Jordan Marshall and Dr. Bruce Kingsbury for reviewing my writing. Thank you, Dar and Colleen for being very helpful and answering all of my questions. I would especially like to thank Purdue Fort Wayne Institute for Research, Scholarship, and Creative Endeavor for funding this study.

Many thanks to the following people and organizations for providing samples: Illinois Natural History Survey (Chris Phillips and Tristan Schramer), Jeff Johnson, Dave Frymire, Rich King, Toledo Zoo (Matt Cross), Ohio State University (Greg Lipps), U.S. Fish and Wildlife Service (Megan Seymour), Indiana Department of Natural Resources, Jeff Davis, and Brad Shoger for providing information on a sample site and allowing me to assist on his surveys. Additional thanks to Wade Lahue and Bryan Eads for providing information on sample sites.

I would also like to thank Rikki Ratsch for providing suggestions on survey methods and for assisting on some surveys for samples, and Tangled Bank Conservation for answering questions concerning sample data analysis and their involvement in library construction and initial bioinformatic filtering steps. Lastly, I would like to thank friends and family who have provided support, laughs, and a lot of good food.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# ABSTRACT

Author: Noland, Alanna, D. MS
Institution: Purdue University
Degree Received: August 2019
Title: Population Genomics of Kirtland's Snake (*Clonophis kirtlandii*).
Committee Chair: Mark Jordan

The semi-fossorial Kirtland's Snake *Clonophis kirtlandii* is a relatively small species in the subfamily Natricinae that is endemic to the Midwest and prefers moist habitats such as open meadows and wetlands. Threatened or endangered over most of its range, a population genomic study for this species is necessary to better understand the status of its populations and provide a basis for developing management plans to further its protection. Here, a population genomics analysis using 3RAD was used to discover 166 SNPs within this species from across its geographic range. Multiple cluster analyses revealed populations formed three clusters, which were not clearly based on geographic location. Samples from Indiana and Ohio showed admixture and were assigned to two different clusters, while Illinois samples were all located in a single cluster. Michigan samples and some individuals collected from single site in southeast Indiana formed a third distinct cluster. Pairwise comparison of DAPC results revealed that the three identified clusters were distinct from each other. Individuals from non-urban habitats trended toward higher levels of nucleotide diversity and expected heterozygosity compared to individuals from urban habitats but without statistical support. This study successfully identified loci that were used to describe genetic variation between *C. kirtlandii* populations.

# INTRODUCTION

Changes in the amount of gene flow occurring between populations is a potential consequence of fragmentation of suitable habitat for snakes (Clark et al. 2010; Jansen, Mushinsky, & Karl 2008). Investigating genetics of populations that have reduced connectivity can provide insight to the level of genetic diversity within fragmented populations, and whether there is indication of reduced diversity or inbreeding (Wood et al. 2018). Furthermore, examining genetics of populations can be useful when determining whether a species requires management (Dann et al. 2013). Genome-wide data can be examined for genetic variation and may potentially uncover differences that suggest that populations require separate management or augmentation (Pan et al. 2016). In particular, low genetic variation within and among populations may suggest that populations are at risk, as low variation has been attributed to individuals having reduced reproductive fitness and viability (Madsen, Stille, & Shine 1996; Vrijenhoek 1994). Differences in levels of allelic richness and genetic diversity among populations may be used to identify genetic units and populations that may require conservation management (Gautschi et al. 2002; Prior, Gibbs, & Weatherhead 1997). In addition to informing management decisions, exploring population genetics has the potential to elucidate historical movement and distributions of a species (Richmond et al. 2016).

The semi-fossorial Kirtland's Snake *Clonophis kirtlandii* is a relatively small species that is found in wet open areas such as meadows and wetlands, or other areas that contain a body of water (Bavetz 1994; Gibson & Kingsbury 2004). Existing information on Kirtland's Snake is very limited due to its secretive behavior, use of subterranean habitat, and small populations (U.S. Fish and Wildlife Service 2017). The geographic range of Kirtland's Snake includes the Midwestern states of Illinois, Indiana, Kentucky, Michigan, Missouri, Ohio, and Pennsylvania

(Bavetz 1994), as well as in northwestern Tennessee (Frymire & Scott 2012). The most recent status report of Kirtland's Snake has determined it extant in seven states including Illinois, Indiana, Kentucky, Michigan, Missouri, Ohio, and Tennessee (U.S. Fish and Wildlife Service 2017). It is listed as threatened in Illinois, Kentucky, Ohio; endangered in Indiana, Michigan, and Pennsylvania; and critically imperiled in Missouri and Tennessee (Gibson & Kingsbury 2004; U.S. Fish and Wildlife Service 2017).

Due to its preference to moist, open habitats such as wetlands, loss of habitat is a likely cause of Kirtland's Snake decline. Additionally, gene flow may be an issue for fragmented populations, particularly for populations present in urbanized areas. Studies that have investigated the genetics of snake populations belonging to other closely related species located in or near urban areas have found reduced genetic variation (Gangloff et al. 2017; Richmond et al. 2016). This has been attributed to population isolation and fragmentation from barriers such as roads, buildings, and other anthropogenic disturbances. Another problem is that Kirtland's Snake may exhibit low dispersal ability due to small population sizes, thus exacerbating the effect of fragmented populations (Capparella, Springer, & Brown 2012; Gibson & Kingsbury 2004). The ability to examine the genetics of a population is often an important step to making population management decisions, particularly of species that are already suggested to be threatened or endangered. Currently, the only genetic population study involving the Kirtland's Snake examined the mitochondrial ND2 and cytochrome b genes in a sample of 24 individuals (collected from Illinois, Indiana, Michigan, Missouri, and Ohio) and found no population differentiation (Ray 2009).

## Historic Distribution

Most commonly found in formerly glaciated areas, Kirtland's snakes are thought to be associated primarily with the Prairie Peninsula region and are proposed to be relics of the Xerothermic period which occurred after the post-glacial maximum (Smith 1957). Kirtland's populations probably migrated in an eastward manner which followed the former prairie area that expanded from west to east during this time (Conant 1943; Smith 1957). Some populations that exist south of the former glaciated areas could be relicts of the glacial period that survived at the southern edge of the glacial extent (Conant 1943), and then spread northward following retreat of the glaciers. Small populations in some habitats have been discovered far apart from other more well-known populations (Frymire & Scott 2012); these may be part of a refugia that survived at the southern extent of the glacial edge. Kirtland's snakes most likely spread outward from their most heavily populated ranges in the former wet-prairie and swampy regions to new areas (Conant 1943).

It was previously suggested that the range of Kirtland's occurred in Illinois, Indiana, and Ohio (Conant 1943; Smith 1957), but there is evidence of other possible relict populations far removed such as southern Kentucky and northwestern Tennessee (Frymire & Scott 2012). Additionally, small populations have also been found in Missouri (U.S. Fish and Wildlife Service 2017).

Conant (1943) proposed that because this species is found in a variety of habitats that Kirtland's was once very widespread, but that the distribution became highly fragmented and populations were most likely exterminated from many regions due to post- glacial forest expansion from the south, followed by drainage of wetlands and development. The species may have also survived in wet forested areas, and then dispersed more recently to new areas that have

been cleared and maintained as wet-prairie habitats or other wet open areas (Smith 1957). For some populations, such as those that occur in urban areas, this may be important, as many urban areas where Kirtland's are present were formerly forested (Conant 1943).

Thus, the main range most likely spread from Illinois and Indiana, into Kentucky, Michigan, Missouri, Ohio, and Pennsylvania, after which the range of Kirtland's was fragmented after forest expansion from the south. Small populations probably survived in wet prairies, wetlands, and open areas, or wet open forested areas prone to flooding, such as bottom river valleys. Fragmentation and reduction of suitable habitat may have further compounded the issue of dispersal of small populations. For urban areas, small populations may have persisted in wet areas and waterways. Kirtland's populations at the southern edge of distribution may be more diverse because they likely experienced more gene flow as a result of post-glacial range expansion compared to populations in the main range.

## Approaches to Population Genetic Analysis

Microsatellites are commonly used in genetic analysis studies that investigate genetic differentiation and are often used to examine population genetics of Natricine snakes (Jansen, Mushinsky, & Karl 2008; King et al. 2001). However, an issue with microsatellites is that they require the development of species-specific primers or the use of primers developed for closely related species. In many cases, using primers to study the genetics of a species of interest for which those primers have not been specially designed, results in reduced or failed amplification of loci in the species of interest (Prosser, Gibbs, & Weatherhead 1999). This is problematic, as a larger number of loci will result in a more accurate genetic analysis. Furthermore, only using primers that have been developed for certain species poses the risk of overlooking other significant loci that may be present through mutation, or misrepresenting repeat lengths of

present loci in the species of interest (Ellegren, Primmer, & Sheldon 1995; Jarne & Lagoda 1996).

There are currently no available primers designed specifically for the Kirtland's Snake and only a few primers that have been developed in Natricines show cross-amplification in related species, *Storeria dekayi* and *S. occipitomaculata* (Prosser, Gibbs, & Weatherhead 1999). For species that do not have genomic information already available, loci must be isolated *de novo*, which requires a thorough library screening with probes. This is suitable for species with high numbers of loci or when only a small number of loci are needed, but screening can be tedious if the species has low frequencies of microsatellite loci. Also, the probes required as well as other necessary reagents can be quite expensive (Zane, Bargelloni, & Patarnello 2002). There is limited genomic information available for the Kirtland's Snake, therefore, isolating microsatellites for this species may not be cost effective.

RADseq (restriction site-associated DNA sequencing) is an alternative approach that provides a high degree of resolution and is particularly useful for species that have not been extensively studied and have poorly characterized genomes (Davey & Blaxter 2010). There are several modern RADseq methods that use restriction enzymes to digest the DNA, such as genotyping by sequencing (GBS), Double-digest RAD (ddRAD), and ezRAD (Andrews et al. 2016; Çilingir et al. 2017). The use of restriction enzymes often produces fragments of variable lengths, which require other reduction or size selection steps in the protocols. Additionally, larger amounts of DNA are usually required for these other methods (Andrews et al. 2016).

3RAD, a RADseq method that incorporates three restriction enzymes, can be used to accurately detect and characterize single nucleotide polymorphisms (SNPs) in the genome and overcomes the limitations of other methods (Bayona-Vasquez et al. 2019). In particular, the third

restriction enzyme reduces the number of adapter-dimers that form during the digestion and ligation step of the protocol. Reducing adapter-dimer formation is important because it increases the ability of the adapter sequence to ligate to the input DNA. Therefore, it is helpful when starting with samples that have lower or less than ideal starting DNA concentrations. In addition, design of the adapters and specific reagents that are used permit fewer preparation steps, produce high yields of input product for the PCR step, and give the ability to multiplex and pool a large number of individual samples (Bayona-Vasquez et al. 2019).

## Study Objective

The overall objective of this study was to use genomic information produced from the 3RAD method to describe genetic variation within and among populations to inform conservation of Kirtland's Snake. This information will be used to determine whether this species exhibits population structure across its range. In addition, genetic and nucleotide diversity measures between urban and non-urban populations will be examined to determine whether urban populations show lower genetic diversity compared to populations located in non-urban habitats. Lastly, hypotheses on the distribution of this species will be considered regarding post-glacial range expansion.

# METHODS

## Tissue Samples and DNA Extraction

I compiled tissue samples from across the geographic range of Kirtland's Snake. The majority of extant records for Kirtland's Snake populations are in south central Indiana, mid-state Illinois, and the western half of Ohio and most tissues were collected from these regions, representing both urban (Cincinnati, Indianapolis, Louisville), and more natural habitats (U.S. Fish and Wildlife Service 2017; Appendix A). A small number of samples were obtained from Michigan and northern Missouri. However, samples from Missouri were of poor quality and did not pass the PCR step in preparation for sequencing. Tissue samples from 34 other localities were acquired with 1-12 snakes per locality. These were provided from partners in other states that had taken samples during prior surveys of this species including: Illinois and Kentucky (Illinois Natural History Survey), Indiana (Indiana Department of Natural Resources, Jeff Johnson), and Ohio (Ohio State University, Toledo Zoo, and U.S. Fish and Wildlife Service). These samples were added to those previously collected (Northern Illinois University, natural history museum collections, and samples available in the research laboratory). I also collected samples from Indiana in Summer 2018.

Tissue samples comprised of tail clippings were placed in 95% ethanol in the field and then stored at -4°C. DNA extractions from both tail clippings and blood were performed using the DNeasy Blood & Tissue Kit (Qiagen, MD, USA). Quantification of DNA was done using a Nanodrop spectrophotometer.

## Preparation of Samples and Library Construction

Tangled Bank Conservation (North Carolina) prepared the genomic DNA prior to construction of the genomic library. The protocol followed the high-throughput 3RAD Protocol from the University of Georgia EHS DNA laboratory (Athens, GA). The restriction enzymes used were BamHI-HF, MspI, and ClaI. Briefly, DNA samples were normalized to 20ng/μL and quantified on a Qubit fluorometer, then samples were prepared individually for ligation. Following ligation and PCR, samples were purified with SpeedBead, then checked on an agarose gel and normalized. Six samples did not present satisfactory PCR products following the agarose gel check, thus they were excluded from further preparation and the subsequent sequencing. Samples were size selected between ~472 and ~577 base pairs using Pippin Prep. Lastly, PCR products were again quantified by fluorometry and sent to the University of Georgia laboratory (Athens, GA) for sequencing on the Illumina NovaSeq S1 6000 platform.

## Sequence Filtering and SNP Identification

Individual sequences were sorted and filtered using a range of statistical tools. Primary analysis and identification of loci was done using iPyRAD (Eaton 2014). Like the original pipeline PyRAD, iPyRAD is designed to work for analyses encompassing a larger scale but, it can deal with insertions-deletion among the sequences, unlike some other programs. Additionally, it can provide accurate analyses on *de novo* locus identification and genotyping, as well as quality filtering of sequencing reads (Andrews et al. 2016).

Raw Illumina reads of 90 Kirtland's Snake samples were processed with iPyRAD v0.7.28 (https://github.com/dereneaton/ipyrad). The raw reads per sample were first re-demultiplexed with step 1 of ipyrad workflow to remove residual internal tags. Then the post step 1 reads were demultiplexed, quality filtered, and had loci assembled *de novo* with Steps 1-7, generating

multiple output files for data analysis and a statistics file (for overview see, https://ipyrad.readthedocs.io/outline.html). Reads with a depth lower than 6 or more than 10,000 for each individual were discarded and a clustering threshold for *de novo* assembly was set at 85%. To reduce chance of paralogous regions, a maximum of two alleles at each site were allowed in consensus sequences of each individual, and consensus sequences with more than 50% heterozygous sites were excluded.

Output files from iPyRAD were further filtered using VCFtools (Danecek et al. 2011) and the POPULATIONS module in STACKS v2.4 (Catchen et al. 2013). Initial analysis revealed that nine individuals had a large amount of missing data. These nine individuals were removed from the data set prior to any further filtering. Variant sites with more than 50% missing data were also excluded. The vcf file output from VCFtools was imported into STACKS, where data were further filtered based on a minimum minor allele frequency of 5%. Filtering by lower minor allele frequency (MAF) thresholds has the potential to retain incorrectly called bases (or errors) that occur during the sequencing process (Roesti, Salzburger, & Berner 2012). Furthermore, filtering by a MAF < 0.05 could increase the number of rare alleles that are kept in the data set, which are often not as informative and may introduce bias in calculated population differentiation values (Roesti, Salzburger, & Berner 2012). Data was also filtered by an 80% threshold for loci among individuals at each sample locality, to further reduce the occurrence of missing data (Arnold et al. 2013). A locus also had to be present across at least 15 sample localities to be retained, and a random SNP from each locus was written to the output files from STACKS. Keeping only one SNP per locus decreases the chance of linked loci being included further analysis (Andrews et al. 2018).

## Population Structure and Genetic Diversity

Population structure was first examined using snapclust (Beugin et al. 2018) in adegenet v2.1.1 (Jombart 2008). Snapclust implements two steps; first, a suitable starting point is identified which is distance-based. Then, the expectation-maximization (EM) algorithm is used. The implementation of a distance-based step, followed by the EM algorithm step, greatly reduces computation time and yields quick output of visual results. Similar to the STRUCTURE software, snapclust examines the probability of genotype assignment to clusters based on allele frequencies, is model-based, and follows Hardy-Weinberg equilibrium assumptions (Beugin et al. 2018). The number of putative populations (K) was varied from 1 to 10, and the best clustering solution was identified through Bayesian information criterion (BIC).

Possible groups were further assessed with the find.clusters option in the adegenet package. Find.clusters transforms the data by principal component analysis (PCA), and then runs successive K-means on the data, each time with more clusters. K-means is an algorithm that examines a number of groups that are specified in the find.clusters function, and while calculating, maximizes dissimilarities among the different groups (Jombart 2010). The most suitable K value can then be visualized using the computed BIC scores. Unlike snapclust, find.clusters does not require Hardy-Weinberg assumptions among the samples. Possible clusters were examined for K values from 1 to 10, and the best-fit K was chosen based on the point at which an elbow in the curve was observed. Identified clusters were then visualized by discriminant analysis of principle components (DAPC) in adegenet (Jombart, Devillard, & Balloux 2010) using the K value chosen from the observed elbow. With some datasets, there may not be a clear decrease to the lowest BIC value, which is then followed by a sharp increase

(as seen in Figure 2B for snapclust). In this case, a sudden decrease, followed by another gradual decrease, may indicate a possible cluster solution (Figure 3B).

Because DAPC examines genetic differences based on model-free clustering, STRUCTURE v2.3.4 (Pritchard, Stephens, & Donnelly 2000), a model-based method, was also used to explore possible genetic units and clusters. STRUCTURE assumes Hardy-Weinberg equilibrium and linkage equilibrium among the samples. STRUCTURE runs were performed at K values 1 to 10; with 10,000 burn-ins and 10,000 MCMC replications repeated 10 iterations for each K value. The ΔK method (Evanno, Regnaut, & Goudet 2005) was used through Structure Harvester v0.6.94 (Earl & vonHoldt 2012) to determine the best-fit K value. Clumpak (Kopelman et al. 2015), was used to visualize individual assignment to clusters from STRUCTURE results.

Data files were converted to ARLEQUIN format using GenAlex v6.5 (Peakall & Smouse 2006, 2012). Population level measures of genetic differentiation comparing the variability of loci among individuals such as observed heterozygosity ($H_O$), expected heterozygosity ($H_E$), nucleotide diversity ($\pi$) and population differentiation ($F_{ST}$), were calculated using the program ARLEQUIN v3.5 (Excoffier & Lischer 2010). Low $H_O$ values within a population indicates low genetic variability, and greater $F_{ST}$ values indicate a larger degree of genetic differentiation among populations. Analyses were performed with overall sample locality sites and clusters identified through DAPC.

A two-way ANOVA was also performed in SPSS v25 (IMB Corp. 2017) to determine if there were differences in genetic diversity measures between urban and non-urban populations by cluster. For this analysis, 33 sample localities were examined. Sample localities where

individuals were assigned to two separate clusters were assigned to a fourth (mixed) cluster for the urban and non-urban comparison.

## Identifying Outlier Loci

The program BAYESCAN v2.1 (Foll & Gaggiotti 2008) was used to determine if there were any outlier loci. BAYESCAN uses Bayesian approaches for analyzing the genomic information and is capable of detecting possible outlier loci (Andrews et al. 2016). Outlier loci are identified by having fixed allelic differences between localities and indicate that natural selection is driving the differentiation beyond that expected due to population isolation. Data files were converted to BAYESCAN format with PGDSpider v2.1.1.5 (Lischer & Excoffier 2012). Clusters identified through DAPC were examined for outlier loci. BAYESCAN was run with 50,000 burn-ins and 100,000 iterations, with a default prior odds value of 10.

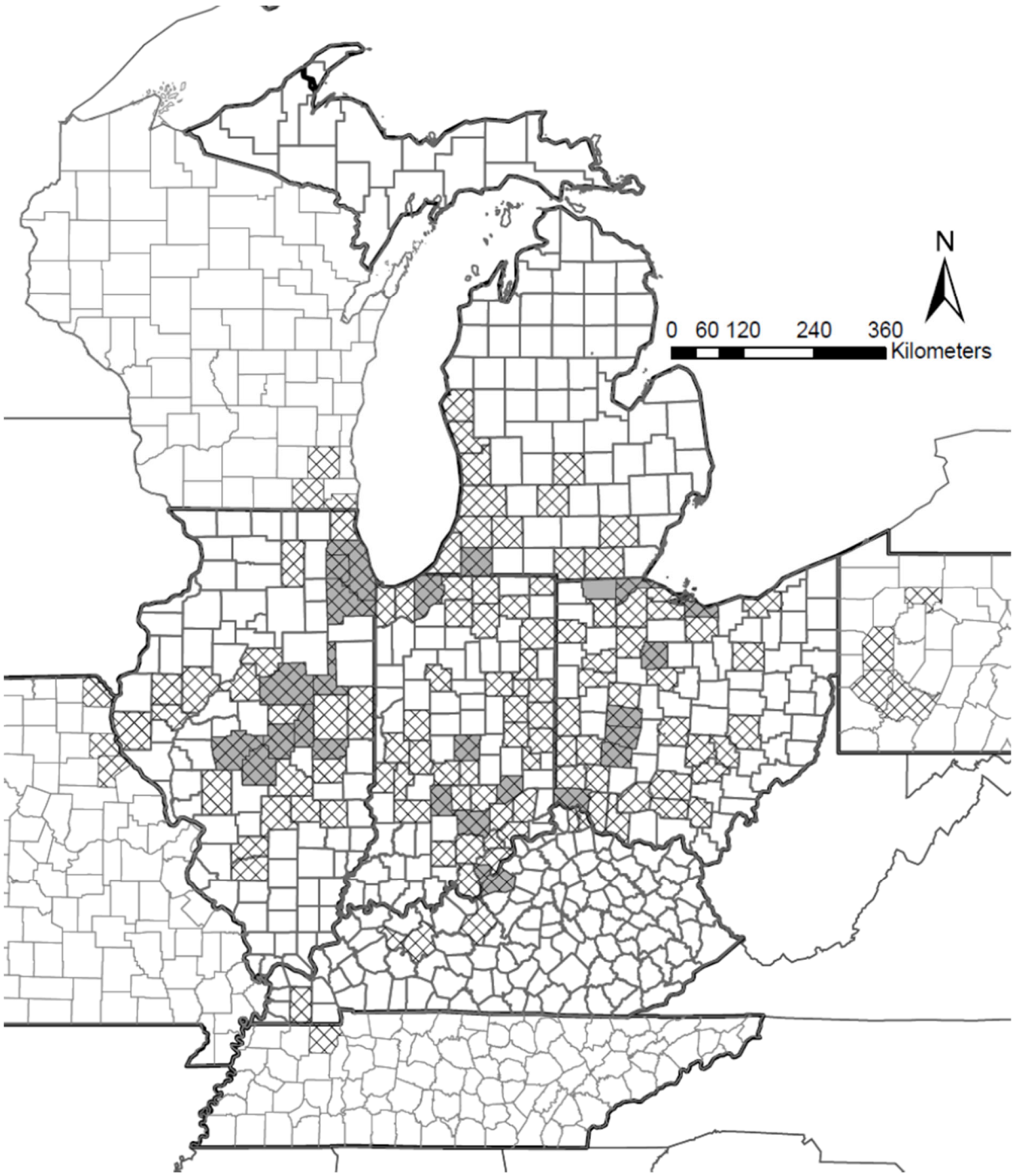



Figure 1: County historic distributions and collection location of tissue samples from Kirtland's Snakes. Grey counties are location of samples used in the analysis. Cross-hatched counties show historic distribution. Counties that did not yield viable genomic DNA products are excluded from sample locations.

# RESULTS

## Sequencing Results and SNP Recovery

*De novo* locus discovery through iPyRad yielded 68,492 loci. After filtering duplicates, heterozygous sites, and a maximum of two alleles per consensus sequence, there were 10,565 loci. After filtering by proportion of missing data at each locus, there were 1038 loci across 81 individuals. Mean read depth across all individuals was 35, max read depth was 99, and median read depth was 33 (Appendix B). Final filtering through stacks by minimum allele frequency, variable sites that occurred in at least 80% of the individuals at each sampling site and across 15 sample localities, returned 166 unlinked SNPs that were used in clustering, genetic diversity, and further analyses.

## Population Structure and Genetic Diversity

Grouping among Kirtland's Snake populations revealed three clusters from the different analyses performed. Snapclust results from the lowest BIC score assigned individuals to three clusters (Figure 2): Cluster results were not completely based on geographic location (Figure 3), as some samples from Indiana and Ohio showed mixture and were not assigned to the same clusters. All samples from Illinois were assigned to cluster one. Samples from northern Indiana, southeast Indiana, and south-central Indiana were also assigned to cluster one along with northern Ohio, western Ohio, southern Ohio, and Kentucky samples. Cluster two was formed by samples from Michigan and from some individuals collected from single site in southeast Indiana. Cluster three included individuals from central Indiana and Ohio.

The find.clusters function and DAPC produced similar results between sites (Figure 4A). The find.clusters function revealed a lowest BIC score of 10 however, there was a slight elbow in

the curve at K=3 (Figure 4B). When cluster solutions 3-10 were run for DAPC, clusters were still placed in the same space along the axes. Therefore, it was determined that three clusters were also the best fit solution for the DAPC. All samples were assigned to the same clusters in both snapclust and DAPC, except for one individual: CK49, which was assigned to cluster three for snapclust and cluster two for the DAPC.

STRUCTURE also identified K=3 based on the ΔK method. Clusters identified through STRUCTURE had similar grouping to both snapclust and DAPC (Figure 5). Indiana and Ohio populations exhibited more admixture (blue and orange bars) compared to populations from Illinois.

Mean nucleotide diversity values ranged from 0.045 to 0.314; Michigan (CASS), a locality from south-central Indiana (JAC2), and three localities from west-central Ohio (CLAR, WYAN, GREE), had the highest values. Four localities from Illinois (PIATT, SANG, DOUG, WILL), and the locality from northern Indiana (LAPO) had the lowest values. $H_E$, mean expected heterozygosity ranged from 0.255 to 0.625. Three localities from Indiana (MON2, MAR1, JAC2), a locality from Ohio (WYAN), and a locality from Illinois (FORD) had the highest values. A locality from southern Indiana (JAC3), two localities from Ohio (HAM1, CLAR), and a locality from Illinois (PIATT) had the lowest values (Table 1).

Cluster two which grouped individuals from MI and SE IN showed the highest mean nucleotide diversity value of 0.321, while cluster one represents individuals from IL, IN, OH, and KY had the lowest mean nucleotide diversity value of 0.072. Cluster three with individuals from IN and OH had a mean nucleotide diversity value of 0.160. Mean expected heterozygosity $H_E$ ranged from 0.160 to 0.455; MI and SE IN samples exhibited the highest value, while individuals from IL, IN, OH, and KY cluster exhibited the lowest value (Table 2).

The lowest genetic differentiation was between cluster 2 (MI and SE IN samples), and cluster 3 (IN and OH samples), ($F_{ST}$ = 0.239; Table 3). The highest differentiation was between cluster 1 and 2 ($F_{ST}$ = 0.444) and 1 and 3 ($F_{ST}$ = 0.309). All comparisons were statistically significant ($P < 0.05$).

Comparisons of urban and non-urban populations showed similar or somewhat reduced genetic diversity but not at a level of statistical significance (Fig. 7-9). Urban populations trended toward lower nucleotide diversity compared to non-urban populations (Figure 8), but these values were also not statistically different for urban ($F_{1,29} = 2.260$, $P = 0.144$), or cluster comparisons ($F_{2,29} = 1.781$, $P = 0.186$). Mean expected (urban: $F_{1,29} = 0.225$, $P = 0.639$; cluster: $F_{2,29} = 2.906$, $P = 0.071$) and observed heterozygosity (urban: $F_{1,29} = 0.000$, $P = 0.985$; cluster: $F_{2,29} = 2.942$, $P = 0.069$) appeared to be less different between landscape types than nucleotide diversity. Note that cluster 2 was not included in the analysis due to its presence at only two localities, both of which were non-urban.

Outlier analysis with BAYESCAN did not detect any outlier loci using a False Discovery Rate (FDR) < 0.05.

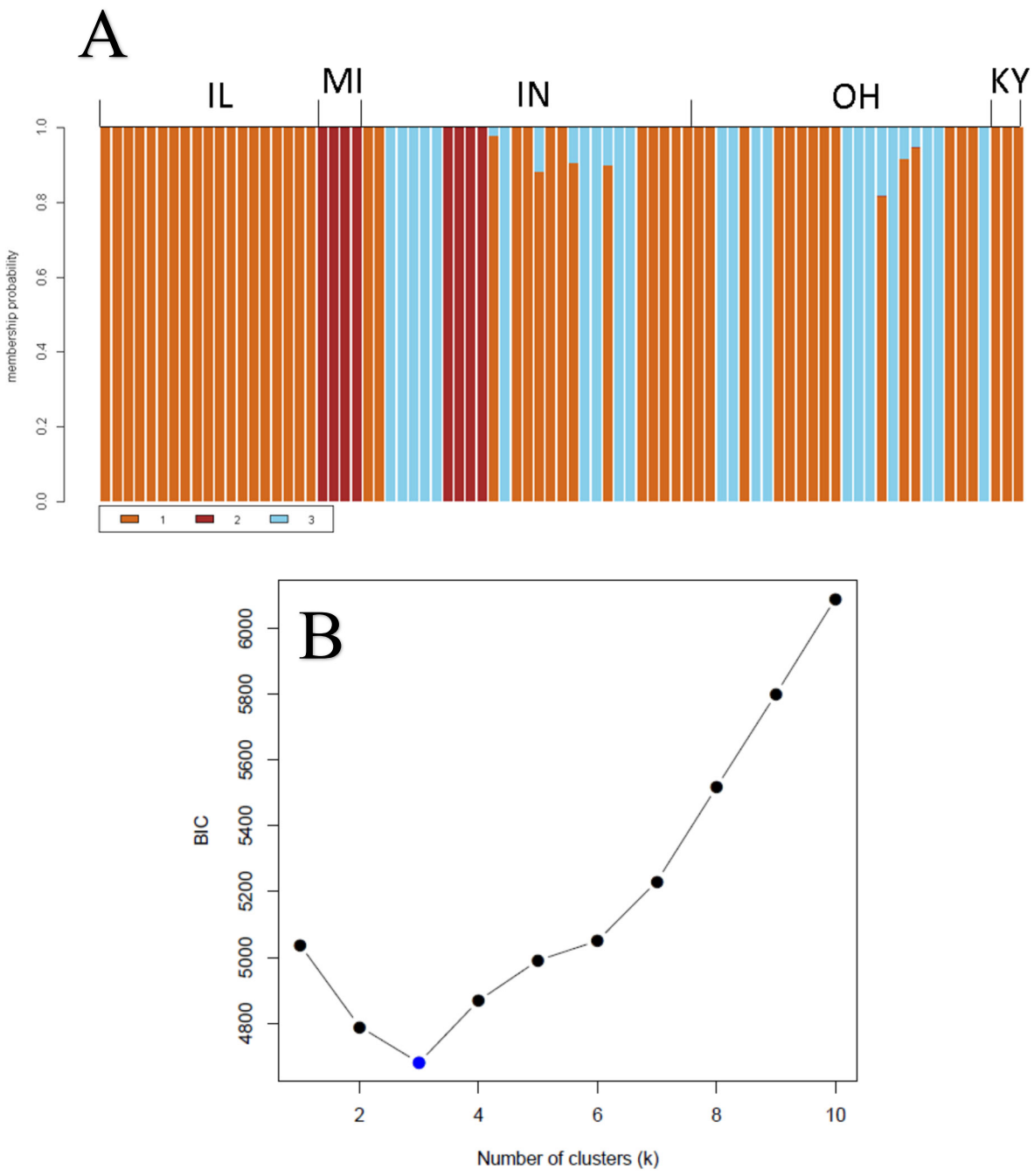



Figure 2: Clustering results for Kirtland's Snake based on (A) snapclust, with cluster one is colored orange, cluster two purple, and cluster three blue, and (B) inference of the cluster number through Bayesian information criterion (BIC). The lowest BIC value corresponding with K=3 is colored blue.

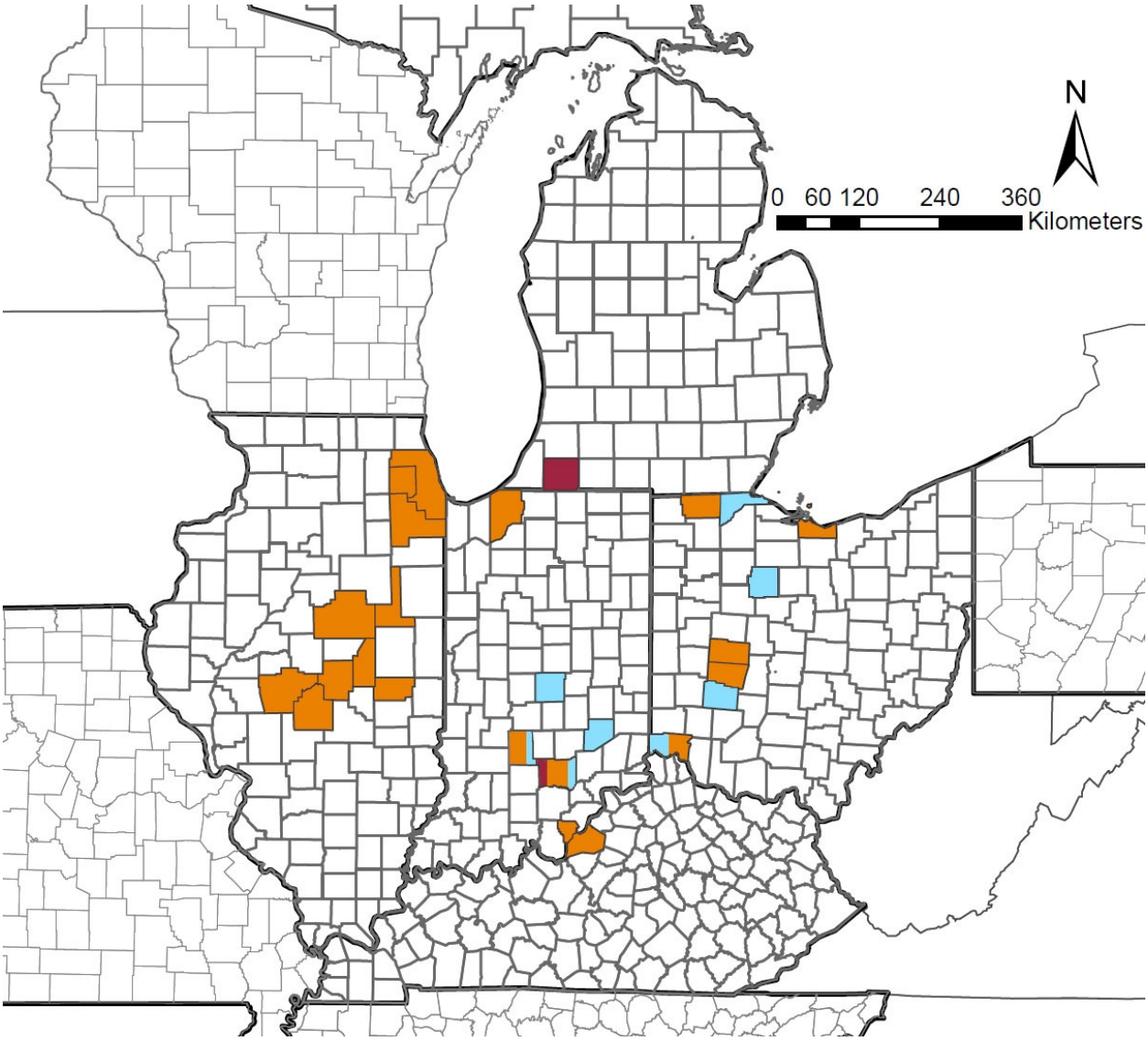



Figure 3: Cluster results based on snapclust showing the range-wide distribution of cluster assignments. Sample localities (counties) are colored based on proportion of individuals assigned to the respective clusters.

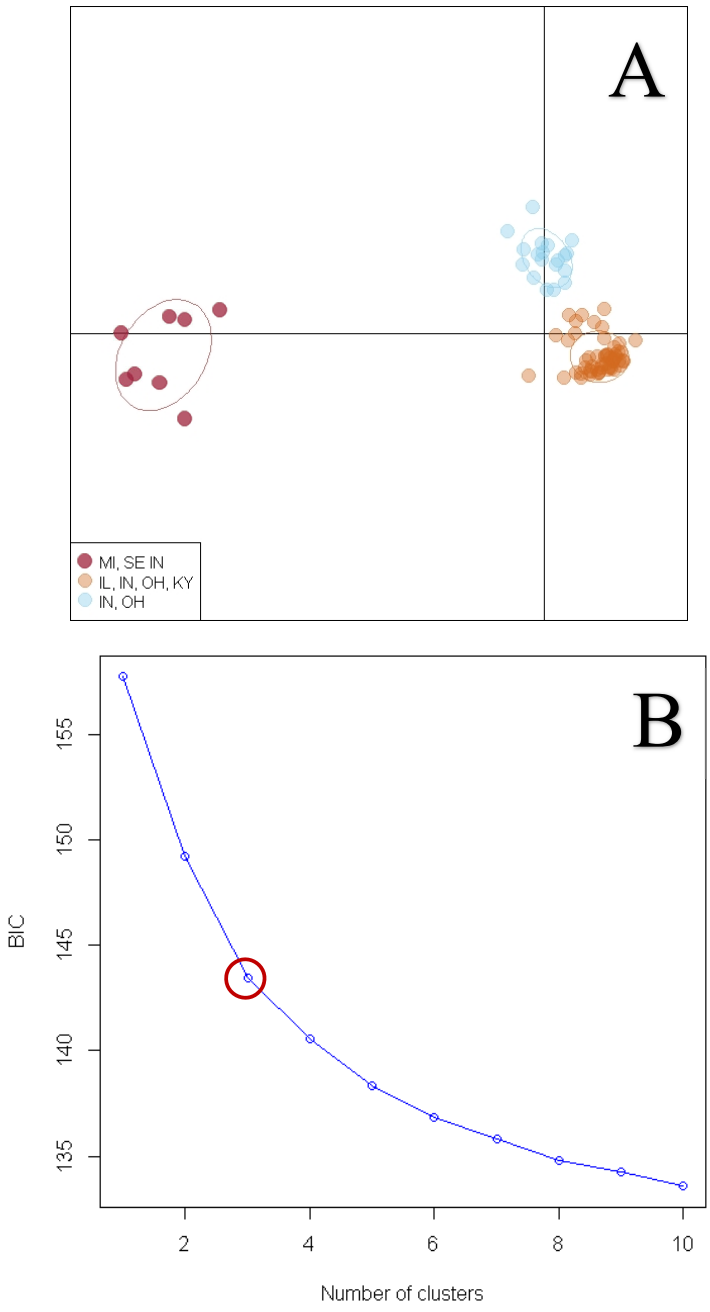



Figure 4: Discriminate analysis of principal component (DAPC) results of SNP data for all 81 samples (A) and (B) value of Bayesian information criterion (BIC) versus number of clusters, inference of the cluster number through BIC value. The slight elbow at K=3 is circled.

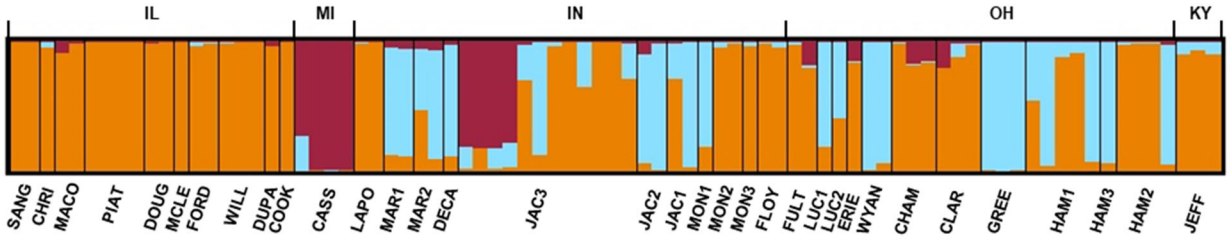



Figure 5: Cluster results from STRUCTURE across all 34 populations based on an optimal value of K=3. Orange bars represent cluster one, purple bars represent cluster two, and blue bars represent cluster three.

Table 1: Kirtland's Snake genetic diversity results based on all 34 sampling localities. (N), total sample size; π, mean nucleotide diversity; $H_E$, mean expected heterozygosity; $H_O$, mean observed heterozygosity.

| **Sample Locality** | **N** | **π** | $H_E$ | $H_O$ |
|---|---|---|---|---|
| SANG | 2 | 0.059 | 0.542 | 0.625 |
| CHRI | 1 | - | - | - |
| MACO | 2 | 0.103 | 0.536 | 0.607 |
| PIAT | 4 | 0.045 | 0.378 | 0.286 |
| DOUG | 2 | 0.061 | 0.542 | 0.500 |
| MCLE | 1 | - | - | - |
| FORD | 2 | 0.086 | 0.567 | 0.500 |
| WILL | 3 | 0.062 | 0.429 | 0.370 |
| DUPA | 1 | - | - | - |
| COOK | 1 | - | - | - |
| CASS | 4 | 0.314 | 0.505 | 0.732 |
| LAPO | 2 | 0.048 | 0.500 | 0.500 |
| MAR1 | 2 | 0.120 | 0.569 | 0.458 |
| MAR2 | 2 | 0.101 | 0.533 | 0.600 |
| DECA | 1 | - | - | - |
| JAC3 | 12 | 0.162 | 0.255 | 0.236 |
| JAC2 | 2 | 0.202 | 0.556 | 0.333 |
| JAC1 | 2 | 0.170 | 0.511 | 0.467 |
| MON1 | 1 | - | - | - |
| MON2 | 2 | 0.066 | 0.625 | 0.125 |
| MON3 | 1 | - | - | - |
| FLOY | 2 | 0.106 | 0.550 | 0.450 |
| FULT | 2 | 0.169 | 0.525 | 0.500 |
| LUC1 | 1 | - | - | - |
| LUC2 | 1 | - | - | - |
| ERIE | 1 | - | - | - |
| WYAN | 2 | 0.178 | 0.574 | 0.556 |
| CHAM | 3 | 0.144 | 0.438 | 0.476 |
| CLAR | 3 | 0.180 | 0.367 | 0.295 |
| GREE | 3 | 0.176 | 0.430 | 0.222 |
| HAM1 | 5 | 0.095 | 0.339 | 0.217 |
| HAM3 | 1 | - | - | - |
| HAM2 | 4 | 0.098 | 0.425 | 0.194 |
| JEFF | 3 | 0.104 | 0.436 | 0.545 |

Table 2: Genetic diversity of clusters based on DAPC for K=3. Cluster one represents individuals from IL, IN, OH, and KY; cluster two represents individuals from MI and SE IN; cluster three represents individuals from IN and OH.

| **Cluster** | **N** | $\boldsymbol{\pi}$ | $\boldsymbol{H_E}$ | $\boldsymbol{H_O}$ |
|---|---|---|---|---|
| 1 | 53 | 0.072 | 0.160 | 0.114 |
| 2 | 8 | 0.321 | 0.455 | 0.587 |
| 3 | 20 | 0.160 | 0.298 | 0.204 |

Table 3: Pairwise $F_{ST}$ genetic differentiation results among clusters identified through DAPC. Cluster one represents individuals from IL, IN, OH, and KY; cluster two represents individuals from MI and SE IN; cluster three represents individuals from IN and OH. $F_{ST}$ values are shown below the diagonal and P values are shown above the diagonal. All comparisons were significantly different from each other at a significance level of 0.05.

| Cluster | 1 | 2 | 3 |
|---|---|---|---|
| 1 | - | <0.001 | <0.001 |
| 2 | **0.444** | - | <0.001 |
| 3 | **0.309** | **0.239** | - |

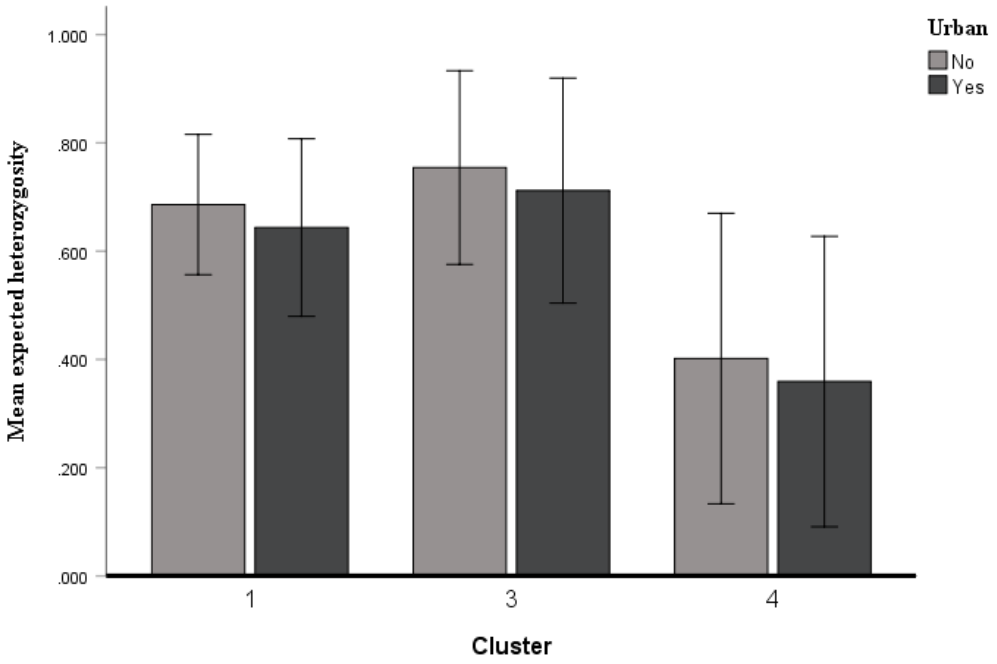



Figure 6: Mean expected heterozygosity ($H_E$), for comparison of populations from urban and non-urban habitats. Individuals assigned to cluster two are excluded. Cluster 4 represents mixed locality sites (with individuals assigned to both cluster one and three). Error bars show 95% confidence interval.

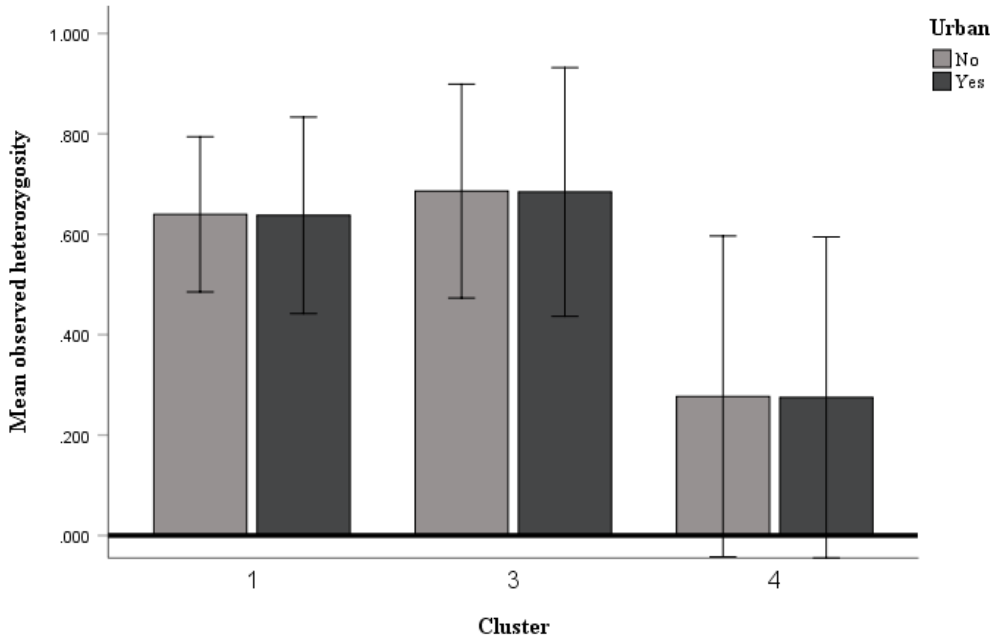



Figure 7: Mean observed heterozygosity (*Ho*), for comparison of populations from urban and non-urban habitats. Individuals assigned to cluster two are excluded. Cluster 4 represents mixed locality sites (with individuals assigned to both cluster one and three). Error bars show 95% confidence interval.

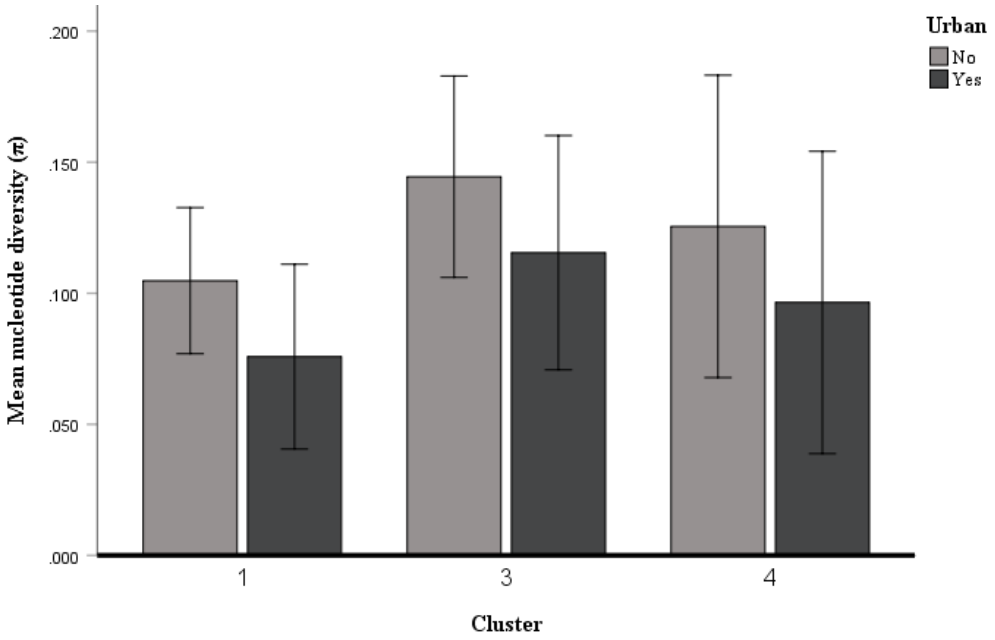



Figure 8: Mean nucleotide diversity ($\pi$), for comparison of populations from urban and non-urban habitats. Individuals assigned to cluster two are excluded. Cluster 4 represents mixed locality sites (with individuals assigned to both cluster one and three). Error bars show 95% confidence interval.

# DISCUSSION

In this study, 3RAD, a RAD-seq method was used to discover and genotype loci in Kirtland's Snake, *C. kirtlandii*. Analysis of identified SNPs revealed that Kirtland's Snake formed three distinct genetic clusters across its geographic range but that these clusters were not completely based on geographic location. The pattern of genetic differentiation revealed suggests that its current distribution may have resulted from range expansion from two separate refugia following glacial retreat. Comparison of genetic diversity measures between urban and non-urban populations showed that urban populations trend toward reduced genetic variation but not at a statistically supported level.

## Dataset Characteristics

After filtering, sequence data for 81 individual samples and 166 SNPs were retained for downstream analyses. The number of loci identified in this study is lower compared to other snake studies. For example, Wood et al. (2018) examined population genetics and structure in two species of garter snakes using both the PyRAD and STACKS pipelines for bioinformatic filtering. For *Thamnophis rufipunctatus* the study yielded 2505-5606 loci through STACKS and PyRAD and ended up with 253 SNPs from STACKS. For *Thamnophis eques megalops*, it yielded 2278-2948 loci through PyRAD and STACKS, respectively, and 537 SNPs from STACKS. In addition to the study by Wood et al. (2018) yielding a higher number of loci (and subsequently, SNPs), the authors also used stricter filters for missing data (30%-40% thresholds). It should also be noted that the study used more individuals for each species (108 and 132). The study presented here included 81 individuals which may have resulted in fewer loci being called. Furthermore, the data produced here may have had higher initial levels of missing data; thus,

after the filtering steps, there were lower numbers of SNPs retained. Another study using ddRAD with 85 *Sistrurus catenatus* samples identified 2073 polymorphic loci (1019 when filtered for occurring in 95% of individuals) *de novo* using the AftrRAD pipeline (Sovic, Fries, & Gibbs 2016).

Thus, depending on the species being studied and methods utilized for bioinformatic filtering, variation in the number of discovered loci and SNPs does occur. Although the number of loci produced in this study is lower compared to some other studies using RADseq methods, *de novo* based sequencing methods do produce smaller numbers of SNPs compared to referenced-aligned sequencing methods (Shafer et al. 2017).

This was the first genomic study of Kirtland's Snake using a RADseq method. Low and uneven sample sizes among some of the sites could have impacted clustering results (Kalinowski 2011; Shafer et al. 2017). Additionally, having no prior genomic information or a closely related reference genome meant that locus identification and SNP calls had to be done *de novo*. The ability to align sequences with a reference genome usually produces a larger number of SNPs (Catchen et al. 2013). Furthermore, dataset preparation methods such as filtering and choice of parameters such as MAF has the potential to impact downstream analyses (Shafer et al. 2017). However, variable MAF and missingness thresholds were applied to this dataset and there was no indication of differences in the number of clusters identified during preliminary analyses.

## Population Structure and Genetic Diversity

Analysis of population structure using both non-model based and model based methods revealed *C. kirtlandii* populations formed three clusters. These results disagree with the previous study that examined this species that found no population differentiation in mitochondrial sequences (Ray 2009). Clusters were not clearly based on geographic location since several

sample localities within the same regions were assigned to a different cluster. Interestingly, all samples from Illinois were only present in one cluster, whereas samples from Indiana and Ohio were placed into different clusters. All samples from Michigan were in a single cluster along with four samples from Indiana. Samples from the single site in Indiana clustering with Michigan samples may be due to translocation by humans because clustering through snapclust they were assigned to the cluster with over 99% probability, although STRUCTURE results indicated low amounts of admixture in the Indiana locality. Most Indiana and Ohio populations exhibited some admixture; samples from northern Indiana, south-central Indiana, and the most southern Indiana sample locality (Floyd) showed the least amount of admixture. Two populations, one urban and one from a more natural area (HAM2 and MON2) exhibited some of the lowest amounts of admixture but potential of inbreeding (Table 1) due to their very low observed to expected heterozygosity (Arnaud-Haond et al. 2006).

Cluster and admixture results may support the idea that Kirtland's Snake moved in a north-eastward manner following the former prairie area that expanded after the post-glacial maximum. This is because all populations from Illinois were assigned to the same cluster, while populations from Indiana and Ohio were not and displayed greater levels of admixture. Increased levels of admixture in Indiana and Ohio populations may indicate that population expansion occurred from two separate refugia, and that these separate distributions could have come into contact at the distribution edge in some areas prior to becoming fragmented from forest expansion as the prairie peninsula retracted in the east. The fact that the Michigan population is assigned to a single cluster may be further indication of movement into a new area following expansion, after which the population became isolated due to development and reduction in suitable habitat.

A study that examined population structure in the copperbelly water snake across Ohio, SE Michigan, Indiana, and Kentucky identified 2-3 clusters as optimal (Marshall, Kingsbury, & Minchella 2009). However, like this study, the cluster results were not completely based on geographic location. Samples from Muscatatuck National Wildlife Refuge formed a separate cluster when either two or three clusters were chosen. Additionally, the other Indiana sample site did not cluster with Muscatatuck, but with Ohio and Kentucky. The authors of this study suggested their clustering results may be due to habitat fragmentation or because of post-glacial range expansion.

Pairwise $F_{ST}$ measures between clusters identified through snapclust revealed significant differences. Cluster one with Illinois, Indiana, OH, and KY samples was more genetically distinct from cluster two that had samples from Michigan, and cluster three that had other samples from Indiana and Ohio. The cluster with Michigan samples and the cluster with samples from Indiana and Ohio were the least divergent, with an $F_{ST}$ value of 0.239.

Studies investigating genetics and gene flow within *Sistrurus catenatus* found that this species, which occupies a similar range to Kirtland's Snake, exhibited high levels of range-wide population differentiation between the identified clusters (Chiucchi & Gibbs 2010; Sovic et al. 2019). Another study that examined phylogenetic lineages within *S. catenatus* found that populations at the most western edge of their current distribution located in Iowa formed a distinct cluster and was separated from populations that occur in Wisconsin and Illinois, as well as other eastern states (Michigan, Ohio, and Pennsylvania). It was suggested that the Iowa lineage may either be an artifact of a refugial population that formed in an unglaciated area, or that the Mississippi River acts as a barrier to gene flow between western and eastern populations (Sovic, Fries, & Gibbs 2016).

Pairwise comparison results also suggest that distribution of Kirtland's Snake followed a north-eastward movement, as the cluster including individuals from Illinois showed the highest divergence when compared with the other two clusters. All $F_{ST}$ measures were moderate and represented a low amount of gene flow between clusters (Lowe & Allendorf 2010). Measures of population differentiation are often influenced by diversity measures within subpopulations (Charlesworth 1998). Clusters did exhibit differences in nucleotide diversity, with the cluster which had the fewest samples (n=8, Michigan site CASS and some samples from Indiana site JAC3) showing the highest nucleotide diversity value, while the cluster with samples from Illinois, Indiana, and Ohio showed the lowest. Lower nucleotide diversity between populations of other species has been suggested to be attributed to past population expansion following glacial retreat (Milá et al. 2000). On the other hand, greater nucleotide diversities may indicate that populations could have potentially occupied refuges or had slower distribution rates (Runck & Cook 2005).

Measures of nucleotide diversity between the separate sample localities also varied; most populations from Illinois exhibited some of the lowest nucleotide diversity values, along with populations collected from urban areas. The idea that populations located at the most southern range may be more diverse was not supported by the results, as populations from Kentucky (JEFF) and Indiana (FLOY) had similar gene diversity values compared to some of the other sites.

Comparisons of urban and non-urban populations revealed that populations in urban habitats had reduced mean nucleotide diversity and mean expected heterozygosity compared to populations in non-urban habitats. However, values were not significantly different. These

results could suggest that Kirtland's snakes in urban areas have reduced gene flow occurring with other populations.

## Management Implications

Because inbreeding has the potential to reduce the amount of genetic variation in future generations, isolated populations such as those occurring in urban areas may require management plans such as introducing additional snakes into these populations, before doing the same for populations in more natural habitats. Another study that has investigated the impacts of urbanization on garter snakes (Gangloff et al. 2017) found low levels of genetic variation and a high indication of inbreeding. Snakes in urban areas most likely have reduced movement and migration due to obstacles such as roads (Shine et al. 2004). Thus, urban populations may benefit from having other snakes introduced in order to alleviate the possibility of inbreeding. Since this study indicated that some urban populations may be distinct from each other, it may be more helpful to introduce snakes that originate outside the urban area from a more natural habitat. However, STRUCTURE results did indicate that there was admixture occurring in some urban populations; Kirtland's Snake may be capable of dispersal and migration via waterways that are present at urban sites. Management of populations located in non-urban areas may also be necessary, particularly populations that exhibit signs of inbreeding such as MON2. In this case, enabling gene flow using populations in closest proximity may benefit these natural areas. However, due to the signs of admixture in Indiana and Ohio populations, it is recommended that the cluster assignment of these populations be examined to make certain distinctive populations are not mixed.

Further studies on movement of Kirtland's Snake in both natural and urban areas are required to better understand home range size and dispersal patterns. After more information is

gathered on dispersal capabilities, other management decisions such as extension of suitable habitat could potentially permit gene flow between populations located in more fragmented natural habitats.

## Conclusion

To date, this is the first study on Kirtland's Snake using a RAD-sequencing method. Using genome-wide SNPs, I discovered population structure and found high differentiation between identified clusters. Comparison between urban and non-urban populations revealed that the snakes from urban areas had reduced nucleotide diversity and expected heterozygosity; however, these values were not significantly different. Analyses also revealed that some populations, particularly those in Indiana and Ohio, have varying levels of admixture.

Other studies investigating the demographics and spatial ecology of Kirtland's Snake are warranted in order to gain insight on population sizes, movement, and dispersal. Some populations, particularly those in urban areas that exhibit genetic signatures of inbreeding may require management before other populations located in more natural areas. Lastly, further studies that utilize RAD-sequencing methods for this species may benefit from processing data through a different pipeline for comparison of the number of SNPs that are called.

# REFERENCES

Andrews, K., Adams, J., Cassirer, E., Plowright, R., Gardner, C., Dwire, M., Hohenlohe, P.A., & Waits, L.P. 2018. A bioinformatic pipeline for identifying informative SNP panels for parentage assignment from RADseq data. Molecular Ecology Resources 18:1263-281.

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):81-92.

Arnaud-Haond, S., S. Teixeira, S. I. Massa, C. Billot, P. Saenger, G. Coupland, C. M. Duarte, & Serrão E.A. 2006. Genetic structure at range edge: low diversity and high inbreeding in Southeast Asian mangrove (*Avicennia marina*) populations. Molecular Ecology 15(12):3515-3525.

Arnold, B., Corbett-Detig, R. B., Hartl, D. & Bomblies, K. 2013. RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. Molecular Ecology 22:3179-3190.

Bavetz, M. 1994. Geographic variation, distribution, and status of Kirtland's snake, *Clonophis kirtlandii* (Kennicott) in Illinois. Transactions of the Illinois State Academy of Science 87(3-4):151-163.

Bayona-Vásquez, N.T, Glenn, T.C., Kieran, T.J., Pierson, T.W., Hoffberg, S.L., Scott, P.A., Bentley, K.E., Finger Jr. J.W., Louha, S., Troendle, N., Díaz-Jaimes, P., Mauricio, R., & Faircloth, B.C. 2019. Adapterama III: Quadruple-indexed, double/triple-enzyme RADseq libraries (2RAD/3RAD). bioRxiv 205799.

Beugin, M-P., Gayet, T., Pontier, D., Devillard, S., & Jombart, T. 2018. A fast likelihood solution to the genetic clustering problem. Methods in Ecology and Evolution 9:1006–1016.

Capparella, A. P., Springer, T., & Brown, L. E. 2012. Discovery of new localities for the threatened Kirtland's snake (*Clonophis kirtlandii*) in Central Illinois. Transactions of the Illinois State Academy of Science 105(1/2):101-105.

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.

Charlesworth, B. 1998. Measures of divergence between populations and the effect of forces that reduce variability. Molecular Biology and Evolution 15(5):538-543.

Chiucchi, J. E. & Gibbs, H. L. 2010. Similarity of contemporary and historical gene flow among highly fragmented populations of an endangered rattlesnake. Molecular Ecology 19(24):5345-5358.

Çilingir, F. G., Rheindt, F. E., Garg, K. M., Platt, K., Platt, S. G. & Bickford, D. P. 2017. Conservation genomics of the endangered Burmese roofed turtle. Conservation Biology 31:1469-1476.

Clark, R. W., Brown, W. S., Stechert, R., & Zamudio, K. R. 2010. Roads, interrupted dispersal, and genetic diversity in timber rattlesnakes. Conservation Biology: The Journal of the Society for Conservation Biology 24(4):1059-1069.

Conant, R. 1943. Studies on North American water snakes-I *Natrix kirtlandii* (Kennicott). The American Midland Naturalist 29(2):313-341.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. Bioinformatics (Oxford, England) 27(15):2156–2158.

Dann, T.H., Habicht, C., Baker, T.T., & Seeb, J.E. 2013. Exploiting genetic diversity to balance conservation and harvest of migratory salmon. Canadian Journal of Fisheries and Aquatic Sciences 70:785-793.

Davey, J. W. & Blaxter, M. L. 2010. RADSeq: next-generation population genetics. Briefings in Functional Genomics 9(5-6):416–423.

Earl, D. A. & vonHoldt, B. M. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4(2):359-361.

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

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

Excoffier, L. & Lischer, H. E. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10(3):564-567.

Evanno, G., Regnaut, S., & Goudet, J. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. Molecular Ecology 14:2611-2620.

Foll, M. & Gaggiotti, O. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A bayesian perspective. Genetics 180(2):977-993.

Frymire, D. & Scott, A. F. 2012. *Clonophis kirtlandii* (Kirtland's Snake). Herpetological Review (43):443.

Gangloff, E. J., Reding, D. M., Bertolatus, D., Reigel, C. J., Gagliardi-Seeley, J. L., & Bronikowski, A. M. 2017. Snakes in the city: Population structure of sympatric gartersnakes (*Thamnophis spp*.) in an urban landscape. Herpetological Conservation and Biology 12(2)509-521.

Gautschi, B., Widmer, A., Joshi, J., & Koella, J.C. 2002. Increased frequency of scale anomalies and loss of genetic variation in serially bottlenecked populations of the dice snake, *Natrix tessellata*. Conservation Genetics 3(3):235-245.

Gibson, J. & Kingsbury, B. 2004. Conservation assessment for Kirtland's snake *(Clonophis kirtlandii)*. USDA Forest Service, Eastern Region.

IBM Corp. Released 2017. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp.

Jansen, K. P., Mushinsky, H. R., & Karl, S. A. 2008. Population genetics of the mangrove salt marsh snake, *Nerodia clarkii compressicauda*, in a linear, fragmented habitat. Conservation Genetics 9(2):401-410.

Jarne, P. & Lagoda, P. J. L. 1996. Microsatellites, from molecules to populations and back. Trends in Ecology and 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 11:94.

Jombart, T. 2008. *adegenet*: a R package for the multivariate analysis of genetic markers. Bioinformatics 24(11):1403-1405.

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.

King, R. B., Milstead, W. B., Gibbs, H. L., Prosser, M. R., Burghardt, G. M., & McCracken, G. F. 2001. Application of microsatellite DNA markers to discriminate between maternal and genetic effects on scalation and behavior in multiply-sired garter snake litters. Canadian Journal of Zoology 79(1):121.

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.

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

Lowe, W. H. & Allendorf, F. W. 2010. What can genetics tell us about population connectivity? Molecular Ecology 19(15):3038-3051.

Madsen, T., Stille, B., & Shine, R. 1996. Inbreeding depression in an isolated population of adders *Vipera berus*. Biological Conservation 75(2):113-118.

Marshall, J. C., Kingsbury, B. A., & Minchella, D. J. 2009. Microsatellite variation, population structure, and bottlenecks in the threatened copperbelly water snake. Conservation Genetics 10(2):465-476.

Milá, B., Girman, D., Kimura, M., & Smith, T. 2000. Genetic evidence for the effect of a postglacial population expansion on the phylogeography of a North American songbird. Proceedings: Biological Sciences 267(1447):1033-1040.

Pan, Y., Wang, X., Sun, G., Li, F., & Gong, X. 2016. Application of RAD sequencing for evaluating the genetic diversity of domesticated Panax notoginseng (Araliaceae). PLoS ONE 11(11):1

Peakall, R. & Smouse, P.E. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. Bioinformatics 28:2537-2539.

Peakall, R. & Smouse, P.E. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6:288-295.

Prior, K.A., Gibbs, H.L., & Weatherhead, P.J. 1997. Population genetic structure in the black rat snake: Implications for management. Conservation Biology 11(5):1147-1158.

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

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

Ray, J. W. 2009. Conservation genetics and ecological niche modeling of Kirtland's snake, *Clonophis kirtlandii,* and the Eastern Massasauga rattlesnake, *Sistrurus catenatus catenatus*. MS Thesis. Department of Biological Sciences, Northern Illinois University.

Richmond, J. Q., Wood, D. A., Swaim, K. E., Fisher, R. N., & Vandergast, A. G. 2016. Historical habitat barriers prevent ring-like genetic continuity throughout the distribution of threatened Alameda Striped Racers (*Coluber lateralis euryxanthus*). Herpetologica 72(3):202-213.

Roesti, M., Salzburger, W., & Berner, D. 2012. Uninformative polymorphisms bias genome scans for signatures of selection. BMC Evolutionary Biology 12:94.

Runck, A. M. & Cook, J. A. 2005. Postglacial expansion of the southern red-backed vole (*Clethrionomys gapperi*) in North America. Molecular Ecology 14(5):1445-1456.

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 8:907-917.

Shine, R., M. Lemaster, M., Wall, T. Langkilde, & Mason, R. 2004. "Why did the snake cross the road? Effects of roads on movement and location of mates by garter snakes (*Thamnophis sirtalis parietalis*). Ecology and Society 9(1):9.

Smith, P. W. 1957. An analysis of post-Wisconsin biogeography of the prairie peninsula region based on distributional phenomena among terrestrial vertebrate populations. Ecology 38:205.

Sovic, M. G., Fries, A. C., & Gibbs, H. L. 2016. Origin of a cryptic lineage in a threatened reptile through isolation and historical hybridization. Heredity 117(5):358-366.

Sovic, M., Fries, A., Martin, S. A., & Gibbs, H.L. 2019. Genetic signatures of small effective population sizes and demographic declines in an endangered rattlesnake, *Sistrurus catenatus*. Evolutionary Applications 12(4):664-678.

U.S. Fish and Wildlife Service. 2017. Species status assessment report for Kirtland's snake (*Clonophis kirtlandii*). 43 pp.

Vrijenhoek, R.C. 1994. Genetic diversity and fitness in small populations. In: Loeschcke V., Jain S.K., Tomiuk J. (eds) Conservation Genetics. EXS, vol 68. Birkhäuser, Basel.

Wood, D.A., Emmons, I.D., Nowak, E.M., Christman, B.L., Holycross, A.T., Jennings, R.D., & Vandergast, A.G. 2018. Conservation genomics of the Mogollon Narrow-headed gartersnake (*Thamnophis rufipunctatus*) and Northern Mexican gartersnake (*Thamnophis eques megalops*): U.S. Geological Survey Open-File Report 2018 –1141, 47 p.

Zane, L., Bargelloni, L., & Patarnello, T. 2002. Strategies for microsatellite isolation: a review. Molecular Ecology 11(1):1-16.

# APPENDIX A

Sample localities with location codes for Kirtland's Snake samples used in the analysis. N represents the total number of samples for each locality.

| Sample location | N |
|---|---|
| *Illinois* | |
| Christian (CHRI) | 1 |
| Cook (COOK) | 1 |
| Douglas (DOUG) | 2 |
| DuPage (DUPA) | 1 |
| Ford (FORD) | 2 |
| Macon (MACO) | 2 |
| McLean (MCLE) | 1 |
| Piatt (PIAT) | 4 |
| Sangamon (SANG) | 2 |
| Will (WILL) | 3 |
| *Indiana* | |
| Decatur (DECA) | 1 |
| Floyd (FLOY) | 2 |
| Jackson site 1 (JAC1) | 2 |
| Jackson site 2 (JAC2) | 2 |
| Jackson site 3 (JAC3) | 12 |
| LaPorte (LAPO) | 2 |
| Marion site 1 (MAR1) | 2 |
| Marion site 2 (MAR2) | 2 |
| Monroe site 1 (MON1) | 1 |
| Monroe site 2 (MON2) | 2 |
| Monroe site 3 (MON3) | 1 |
| *Kentucky* | |
| Jefferson (JEFF) | 3 |
| *Michigan* | |
| Cass (CASS) | 4 |
| *Ohio* | |
| Champaign (CHAM) | 3 |
| Clark (CLAR) | 3 |
| Erie (ERIE) | 1 |

| | |
|---|---|
| Fulton (FULT) | 2 |
| Greene (GREE) | 3 |
| Hamilton site 1 (HAM1) | 5 |
| Hamilton site 2 (HAM2) | 4 |
| Hamilton site 3 (HAM3) | 1 |
| Lucas site 1 (LUC1) | 1 |
| Lucas site 2 (LUC2) | 1 |
| Wyandot (WYAN) | 2 |

# APPENDIX B

Violin plot showing mean read depth across retained samples. Identification is based on a unique code for each sample.