CONSERVATION GENETIC ANALYSIS OF BLANDING'S TURTLES ACROSS OHIO, INDIANA, AND MICHIGAN

by

Dan Guinto

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

December 2021

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Mark A. Jordan, Chair
Department of Biology

Dr. Bruce A. Kingsbury
Department of Biology

Dr. Rebecca A.S. Palu
Department of Biology

Dr. Matthew D. Cross
Toledo Zoo

Gregory J. Lipps, Jr.
Ohio State University

Approved by:
Dr. Jordan M. Marshall

Dedicated to my friends, family, and many mentees who have guided and supported me along the way.

ACKNOWLEDGMENTS

I would like to thank Dr. Mark Jordan for his constant guidance and support through each aspect of my graduate school journey. I would also like to thank my committee members Dr. Bruce Kingsbury, Dr. Rebecca Palu, Dr. Matt Cross, and Greg Lipps wisdom, guidance, and support. For Funding support for tissue collection and genotyping I'd like to thank the Indiana Department of Natural Resources, Toledo Zoo, and the U.S. Fish and Wildlife Service. Next I would like to thank our partners: (Ohio) from The Toledo Zoo, Matt Cross, ZooTeens; Terry Breymaier, Chris Otto, and Felicity Galavan from The Ohio State University, Greg Lipps; from The Ohio Division of Wildlife, Kate Parsons and Jim Schott; the Ohio Division of Natural Areas and Preserves, the Ohio Division of Parks and Watercraft; from Winous Point Marsh Conservancy, John Simpson and Brendan Shirkey; from Ottawa Shooting Club, Pete Ochs; Old Woman Creek National Estuarine Research Reserve; (Michigan), from the Michigan Natural Features Inventory; Yu Man Lee and Helen Enander; Michigan State University Extension, From the Michigan Department of Natural Resources, Amy Derosier; Michigan Nature Association, Huron-Clinton Metroparks, Seven Ponds Nature Center, Addison Township , (Indiana) Purdue University at Fort Wayne, the Kingsbury Lab, from Indiana Department of Natural Resources, Linnea Petercheff, Teresa Clark; from the Nature Conservancy, Nathan Herbert. Next I'd like to thank Tyler Scoville and Connor Dempsey for aid and assistance in conducting lab work. Finally I'd like to thank the many field leads, field technicians, and volunteers who contributed data collection and field sampling: Kailyn Atkinson, Chloe Bates, Julia Boldrick, Katie Brandewie, Charlotte Brennan, Morgan Boyer, Ian Chick, Elizabeth Cubberly, Connor Dempsey, Diana Digges, Dan Earl, Stephanie Emerine, Molly Fava, Nick Friedeman, Talia Greenblatt, Gia Haddock, Jessica Hinson, Beckie Hippensteel, Trevor

Hoffman, Liz King, Minh Lee, John Lynch, Carly Martenson, Jen Mayer, Andrew Metz, Chrissey Miller, Sophie Mills, Sean Obrochta, Elspeth Pierce, Zack Pitman, Shelby Priester, Trevor Proctor, Joseph Redinger, Mic Rohde, Courtney Ross, Dale Shank, Frank Schroyer, Jacob Schott, Tyler Scoville, Ayley Shortridge, Nick Smeenk, Hunter Smith, Jesse Sockman, Reine Sovey, Bria Spalding, Courtney Thompson, Jenny Swonger, Megan Wise.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The Blanding's Turtle (*Emydoidea blandingii*) is considered a species of conservation need across much of its range. A key aspect to conserving a species is understanding the genetic diversity and population structure across the landscape. Several researchers have focused on *E. blandingii* genetic diversity in the northeastern United States, Canada, and the Midwest. However, little investigation has been done on localities within the Great Lakes region of Indiana, Michigan, and Ohio. Here 14 microsatellite loci are utilized to characterize the genetic diversity of *E. blandingii* in Indiana, Ohio, and Michigan. Understanding genetic trends within this region will allow for the defining of management units through genetic clustering, investigation of historic and recent migration between clusters, investigation of drivers of genetic differentiation, checks for bottlenecks, estimations of effective population size (N_e), and optimization of landscape resistance surfaces. Overall, little differentiation is observed between localities and within locality diversity tended to be high. A minimum of four clusters were identified and as many as seven clusters were detected in a hierarchical manner using three grouping methods (STRUCTURE, Tes3r, and DAPC). Historical migration between clusters was relatively low, and recent migration appears to be absent. Significant correlations between geographic distance and genetic differentiation (IBD), as well as watershed and genetic differentiation were observed. Optimized landscape resistance layers provided poor models and distance was maintained as the best driver of differentiation. No bottlenecking was detected, and *Ne* estimates were generally high, but likely biased by sample size. The long lifespan and delayed genetic differentiation of *E. blandingii* is likely responsible for the observed diversity and lack of differentiation between localities. This does not mean they are secure in the Great Lakes Region. Bottlesim analysis looking at the effects of population reduction and subsequent loss of genetic diversity indicates that many localities

within the study area are likely vulnerable to genetic loss in the next 200 years, which can be rapid and drastic in long-lived species.

CHAPTER 1. CONSERVATION GENETIC ANALYSIS

1.1 Introduction

Turtles (Testudines) in general are one of the most imperiled groups of vertebrates [1]. Blanding's Turtle (*Emydoidea blandingii*) is a species in decline that contributes to this trend, receiving protected status across much of its geographic range [2,3,9]. As a long-lived species with long generation times, large seasonal terrestrial movements, and low annual fecundity, the life history and spatial ecology of *E. blandingii* puts populations at a particular disadvantage in the face of habitat loss and degradation [2,4-6]. In addition, *E. blandingii* exhibit low haplotype and sequence diversity, which indicates potential for continued population decline due to lack of genetic adaptability [7].

E. blandingii populations in the Great Lakes region are reliant on shallow (~2.5 m deep) open marshes, ponds, and lakes with emergent herbaceous vegetation as well as a mix of forested ephemeral wetlands, prairies, and bare sands for upland movement/nesting [8]. A long-term (40+ years) study by Congdon and Gibbons (1996) demonstrates that *E. blandingii* populations can maintain stable population sizes with low recruitment in large (615 hectares), well protected preserves. Urban modification and conversion of such shallow wetland habitat complexes and uplands can lead to local extirpation and changes to the spatial ecology *E. blandingii [10]*. *E. blandingii* is reliant on large, contiguous wetlands and the destruction of this habitat is seen as a key threat to their persistence [6,8,11-16]. Of particular concern is the drastic reduction of wetlands in the Midwestern United States that have occurred since the 1700's [17]. Reduction of movement corridors and exposure to roadways has been shown to increase the incidence of road mortality of *E. blandingii* [15,18]. Increased urbanization can also cause wetland degradation due to introductions of pesticides, herbicide, and fertilizers which may affect *E. blandingii*, as has been

shown for the common snapping turtle (*Chelydra serpentina)* [15]. Bioaccumulation and morphological impacts have been recorded in *C. serpentina* in wetland systems with chemical contamination [19,20]. Beyond physical alteration of the landscape, urbanization can also increase human and predator interaction with *E. blandingii* exposing them to increased poaching and mortality [8,10]. Collection can be particularly detrimental to small or isolated populations which may already be at risk of extirpation [15]. Removal of individuals can be particularly detrimental to small or isolated populations, especially due to the long reproductive lives that are required for population viability [4,21].

An important aspect of conservation planning beyond habitat protection is to understand the genetic composition and diversity of a species within localized populations as well as across their range. Understanding the local and range-wide genetic diversity of a species can help conservationists determine if and when it is appropriate to reintroduce individuals to a landscape or supplement remnant populations using translocation and/or head-starting. By examining the genetic composition prior to conservation actions, researchers can reduce the likelihood of inbreeding and outbreeding depression which could otherwise further imperil an already vulnerable long lived species such as *E. blandingii* [3]. Assessing the effects of such conservation actions for *E. blandingii*, and turtles in general, can be difficult as the effects may take decades to manifest due to the low genetic variability and reduced micro-evolutionary rates present in the order Testudines [22,23].

Small, isolated populations are expected to have lost genetic variation due to bottlenecks caused by range reduction and expansion [7]. However, turtles appear to retain higher than expected levels of genetic variation despite population decline, possibly due to their long generation times [24-30]. Prior studies examining within population genetic diversity of *E.*

blandingii have reported relatively high levels of observed and expected heterozygosity (H_O & H_E) across the range from 0.30 to 0.79 [24-30]. Relatively high levels of allelic richness (AR) have been reported from Ontario and Illinois, ranging from 3.6 to 5.3 [27,30]. Finally, Anthonysamy et al., (2018) found no evidence of inbreeding in *E. blandingii* in northeast Illinois and even found potential evidence of outbreeding (F_{IS} ranging from -0.088-0.042).

Microsatellites or short tandem repeats (STR's) of base pairs in non-coding DNA are present in high frequencies in the eukaryotic genome [3,31]. Microsatellites have relatively high mutation rates caused by proof reading or polymerase slippage errors during DNA replication. As a result, they are prone to accumulate polymorphisms faster than other regions of the genome [32]. Microsatellite markers that are conserved across species but polymorphic between individuals can be particularly useful because they can be used to characterize genetic dissimilarity for cluster analysis. Microsatellite analysis uses the comparison of the length of DNA regions enriched with microsatellites between individuals to determine differences in the number of loci present at the region. Difference in copy numbers allows researchers to determine the presence or absence of polymorphisms within and among populations [31]. Microsatellites have proven to be an effective tool for determining genetic diversity of *E. blandingii* at regional scales through a number of studies [27,28,30,33].

E. blandingii originated between 5 and 19 million years ago (dating to the Hemiphillian or the Miocene), and have experienced slow but drastic changes in the landscape due to glacial expansion and reduction [34-36]. With these climactic changes the geographic range has also shifted likely causing bottlenecks and founder events, creating regional isolation and population structure [7,25,37]. Prior range-wide study using five microsatellite loci identified two evolutionary significant units being separated by the Appalachian Mountains, and found additional support for

recognizing Nova Scotia as its own evolutionary significant unit [25]. Recent range-wide genetic analysis using mitochondrial and nuclear loci supports the previously identified evolutionary units, and indicate that the glacial dynamics experienced by *E. blandingii* are responsible for the presence of at least two genetically distinct lineages: a Great Lakes/Midwest USA lineage to the west of the Appalachian Mountains and a Northeast USA/ Nova Scotia lineage to the east [7].

Region-wide microsatellite analyses find that *E. blandingii* seem to have relatively low levels of differentiation from locality to locality, with higher degrees of differentiation detected east of the Appalachian Mountains compared to west [24-30]. Within the Midwest and Great Lakes regions, these findings all support the south and west glacial retreat followed by north and east recolonization of *E. blandingii* resulting in a high degree of genetic mixing and ultimately creating a low degree of differentiation [27,28,30,38]. Despite the lack of differentiation in the Great Lakes and Midwest (low F_{ST}) there is still evidence of population structure through cluster analysis [27,28,30]. Davy et al. (2014) discovered four genetic clusters utilizing 12 microsatellite loci in southern Ontario, and Anthonysamy et al. (2018) found evidence of a hierarchical assortment of three clusters nested into two large broad clusters across six sites in northeast Illinois using 14 loci. Sethuraman et al. (2014) found four to 6 hierarchical clusters across Illinois, Iowa, Minnesota and Nebraska.

Although analyses of clustering are useful for identifying the distribution of populations, the connectivity of these populations is of conservation interest to identify habitat needs for connectivity and gene flow. The relatively slow rate of genetic differentiation and the long lifespan of turtles makes it difficult to differentiate the effects of contemporary and historical gene flow especially in the presence of a rapidly changing landscape [24,27,33]. Although the genetic differentiation may be limited in scope, it can still indicate historical trends in gene flow that likely

underlie modern population structure [7,25,27,28]. Comparing historical gene flow with modern landscape resistance can be useful in determining where gene flow likely no longer exists between localities. Understanding historic gene flow can be used to determine where corridors or translocation may be useful in promoting gene flow and maintaining genetic diversity by staving off founder's effects or genetic bottlenecks associated with the fragmentation of populations.

Little investigation into the population genetic structure and patterns of differentiation of *E. blandingii* have been conducted within Indiana, Ohio, and Michigan. Two prior studies from Osentoski (2001) and McGuire et al., (2013) within the E.S. George Reserve in Michigan found no evidence of genetic structure using eight microsatellite loci. My study looks to: (1) examine levels of genetic variation and population clustering in *E. blandingii* localities across Indiana, Michigan, and Ohio, and (2) assess the potential for landscape variables to correlate with genetic connectivity among populations. Results of this investigation can be used to provide more focus to ongoing species conservation in the Great Lakes region and further our understanding of historic colonization patterns of *E. blandingii*.

1.2 Methods

1.2.1 Ohio and Michigan Field Sampling

Field sampling was conducted in April to August of 2019 to 2021 in the Lake Erie Watershed in southeast Michigan and northern Ohio. For the first trapping season in Ohio and Michigan, sites were chosen based on current observed or historic presence, along with habitat suitability. For the 2020 and 2021 trapping seasons localities that fell within a circle with a 15 kilometer radius were grouped into a genetic neighborhood. These genetic neighborhoods were intended to encapsulate the home range, breeding dispersal distance, and hatchling dispersal of *E.*

blandingii at each locality [40-42]. We then focused on obtaining a minimum of 10 samples for at least one locality within an area presumed to be within the maximum movement distance of an individual Blanding's Turtle. Additional samples from northwest Michigan were obtained through a partnership with an ongoing study in the Kingsbury Lab (Purdue Fort Wayne). Trapping was conducted following the Northeast Blanding's Turtle Working Group trapping protocol, using a combination of Hoop traps (~0.8 m diameter) and Promar traps (~0.3 m diameter) [43]. Blood was drawn from the nuchal sinus using IACUC approved methods, preserved in 95% ethanol, and placed in a standard freezer until extraction.

1.2.2 Indiana Field Sampling

Field sampling in Indiana took place from March through July of 2017-2019 and a single locality in 2021. Indiana sites were chosen based on historical records [44]. Trapping was conducted using a combination of Hoop traps (~0.8 m diameter) and Promar traps (~0.5 m diameter).

1.2.3 Lab Protocol

DNA was extracted from 95 microliters of alcohol-preserved blood using the Qiagen, DNeasy Blood and Tissue extraction kit. Before starting extractions, stored blood was centrifuged and air-dried for ~15 minutes to separate and remove excess ethanol. A Nanodrop spectrophotometer was used to determine the concentration of extracted DNA. If sufficient DNA was not extracted and blood sample remained the sample was re-extracted.

Fifteen microsatellite markers and primers were chosen from prior studies to maximize genetic variation and facilitate cross-regional comparison for future study (Table 1.1). These loci were chosen for their number of alleles, ability to be multiplexed, and high degree of use across

regions. These microsatellite markers were developed for a variety of turtle species, and have been used on *E. blandingii* across the geographic range [26-30,33,45-49]. The 5' ends of the forward primers were all tagged with universal florescent tails following the methods of Blacket et al., (2012) (6-Fam, NED, PET, or VIC) so that markers could be multiplexed in 5 reactions rather than 15 (Table 1.1). The concentration of the forward primers and universal tails differed slightly to optimize allele calls (Table 1.1).

PCR reactions were performed using a Qiagen Multiplex PCR Kit following manufacturer protocol. Thermocycling included a denaturation step at 95° C for 15 minutes, 35 cycles of denaturation at 94° C for 30 seconds, annealing at 56° C for 90 seconds, and elongation at 72° C for 60 seconds, and final elongation at 72° C for 30 minutes. After thermocycling, completed samples were removed and stored at -80° C. PCR was performed using 2 microliters of DNA (5-50 nanograms per microliter) in a 10 microliter reaction.

Gel electrophoresis was performed for at least six samples from each round of PCR on a 2% agarose gel to ensure amplification took place at expected product lengths. PCR products were then sent to the Yale DNA Analysis Facility or the Yale Keck DNA Sequencing Lab for fragment analysis. All samples were run on an Applied Biosystems 3730xl 96-Capillary Genetic Analyzer using the GelCo. Liz 500 size standard. Electrographs were analyzed using Geneious v. 11.1.5; all loci were scored and binned based on the expected number of repeats. Each locus was rerun at least once with a replicate sample to establish confidence in allele scoring.

1.3 Statistical Analyses

1.3.1 Hardy Weinberg Equilibrium and Disequilibrium

Hardy Weinberg Equilibrium (HWE) is used as a null test for genetic forces contributing to a population. Populations exhibiting Hardy Weinberg Equilibrium are likely not experiencing the genetic effects of evolutionary forces outside of random mating in large populations [51]. PopGenReport version 3.0.4 was used to test for Hardy Weinberg Equilibrium [52]. Linkage disequilibrium can bias the analysis of genetic differentiation by creating false associations between loci within or among populations. Genepop version 1.1.7 was used to examine linkage disequilibrium by locus pair within each of the sample localities [53]. Tests for linkage disequilibrium were run in Poppr version 2.9.3 [54] using the Markov chain method with dememorization set at 10000 with 5000 iterations and 100 batches.

1.3.2 Descriptive Statistics

GenAlEx 6.5 was used to check raw data for missing values, and to export data into different formats [55]. PopGenReport version 3.0.4 was used to determine the relationship between allele numbers and sample size per location, screen for null alleles, identify private alleles, and determine allelic richness with rarefaction [52]. Checking for null alleles helps increase confidence in genetic distance measures, since high frequencies of null alleles can cause inaccurate estimates of F_{ST} and other measures of genetic distance [56]. Private alleles can be an effective way of determining population structure since they are unique to a locality, however the detection of null alleles can be heavily biased by the sample size [57]. Allelic richness by rarefaction accounts for differences in samples size allowing a more accurate assumption of the presence of private alleles [57]. Overall, 492 samples were collected from 49 localities. Initially, PopGenReport was run with

all sites regardless of sample size, to determine the allelic richness with rarefaction, which was then used to determine the minimum sample size needed to properly determine allelic richness and private alleles. Some localities had large samples sizes (>70), they were randomly sub-sampled to a maximum of thirty individuals to avoid bias in comparison to localities with smaller samples [58]. Descriptive statistics were then run for all sites with at least ten samples and a maximum of thirty samples. DiveRsity Version 1.9.90 was used to calculate observed heterozygosity (H_O), Nei's expected heterozygosity (H_E), F_{IS}, F_{ST}, and D [59]. Measures of heterozygosity, along with allelic richness, contribute to the assessment of the amount of genetic variation within sample localities. Heterozygosity measures are used to calculate F_{IS}, the inbreeding coefficient within sample localities, and F_{ST}, the traditional measure of differentiation among localities [60,61]. D (JostD) statistic is used for determining the relative differentiation of allele frequency among sub-divisions within localities [62]. Finally, a mixed linear model was used to determine the differences in allelic richness among sites and Tukey post-hoc test was used to assess pairwise differences.

1.3.3 Structure Analysis, Genetic Clustering

Population structure based off genetic clustering was assessed using three programs with slightly different approaches: STRUCTURE version 2.3.4 [63], TESS3r version 1.1.0 [64], and Adegenet version 2.1.4 [65]. Population structure analysis is used for identifying populations and management units for a species without relying on sample location *a priori*. By using multiple methods, concordance in inferred clustering gives higher confidence in attributing results to a biological process. Additionally, genetic clustering is often hierarchical and using multiple methods can help to determine finer scale patterns among broader clustering schemes. Areas of disagreement in clustering between methods can also highlight areas were clear population structure is difficult to discern and may require additional sampling and analysis.

STRUCTURE Version 2.3.4 uses a Bayesian cluster analysis method utilizing Markov Chain Monte Carlo (MCMC) to group individuals into clusters using unlinked genetic markers [63]. STRUCTURE is a model-based approach that uses the frequencies of each allele at each locus to probabilistically assign individuals to a given population/cluster (K) based on shared allelic frequency while also avoiding departure from HWE and LDE within assumed populations [63,66]. Assuming an admixture model within STRUCTURE allows individuals to be assigned to one or more populations by utilizing a Q matrix that compares the proportion of an individual's genome associated with a given assumed population [63,66]. Structure was run with a burn-in of 50,000 proceeded by 100,000 steps and was run 10 times for K values from 1 to 10. STRUCTURE was run with LOCPRIOR. LOCPRIOR is a function that incorporates the given sample site locations as *a priori* populations to determine if they influence the number of population clusters. Without LOCPRIOR sample location information is not considered. STRUCTURE results were viewed using STRUCTURE Selector [67]. STRUCTURE Selector allows the visualization of the STRUCTURE results in a bar graph format using the Puechmaille method (controls for uneven sampling size) and provides graphic representations of the Delta K and MedMeaK/MedMedK selection criteria [68]. The bar graph displays individual assignment to each given cluster represented by the different colors. Although Delta K methods are commonly presented, MedMeaK/MedMedK is more robust when sampling is uneven [68] and so only those values will be presented.

TESS3r also implements admixture models and MCMC but incorporates spatial trends and special autocorrelation into the prior distribution of the Q-matrix by allowing admixture of each individual to change across geographic space (individual ancestry)[66]. Individual variation in admixture also decreases at a regional and local level to allow for clines in all directions [66]. In

addition to the Q-matrix Tess3r also uses a G-matrix that includes the ancestral genotypic frequencies to conduct a combination of matrix factorization and quadratic programing to determine the number of clusters present in the given data set [64]. Tess3r also allows results to be represented over geographical space as well as in a traditional bar graph format [64]. This allows the incorporation of raster files and presents a unique visualization of the genetic clustering presented over geographical space [64]. Tess3r was run for K = 1 through 10, and the number of clusters (ancestral populations) was determined using cross-validation criteria.

Adegenet was used to perform a discriminant analysis of principal components (DAPC). Adegenet uses a multivariate analysis approach rather than a model-based approach like STRUCTURE [69]. This allows Adegenet to avoid reliance on assumptions about population structure implicit in model-based approaches. Additionally, Adegenet can handle large amounts of data very quickly. Like a traditional principal component analysis (PCA), a (DAPC) provides a useful tool for examining clustering without relying on a Bayesian framework. DAPC does not require assumptions about populations like PCA [69]. The DAPC takes the benefits of both the traditional PCA and combines them with the genetic application of a Discernment Analysis (DA) while also overcoming the limitation of the traditional DA [69]. While a traditional PCA can be appropriate for examining genetic variation of individuals as well as variation among clusters, it lacks the ability to examine difference between clusters while ignoring the variability of individuals within those clusters, which is where the DA is effective [69]. DA however cannot handle the effects of linkage disequilibrium, and is not compatible with multi-allelic data [69]. DAPC uses a traditional PCA to transform the data to be compatible with DA to allow for between cluster analyses that ignore individual variability. Adegenet performs DAPC by using sequential K-means method and model selection to determine the appropriate number of clusters detected.

1.3.4 Migration

An important aspect of understanding observed genetic composition is determining which populations have interacted with each other, and the extent and direction in which gene flow has occurred. Understanding migration patterns can help inform historic and current source-sink dynamics, which may make some populations less stable than others. Since we were primarily interested in movement between populations/management units due to their potential application in conservation, site localities were grouped into the four clusters identified by TESS3r (Table 1.5). TESS3r was chosen, because it produced conservative population clusters, and has been indicated to be most robust for detecting effects of multidirectional clines into clustering compared to STRUCTURE or DAPC [66]. Migrate version 4.4.3 was used to examine rates of historic migration between clusters to determine the degree to which clusters have historically interacted and to examine historic source-sink dynamics between clusters [70]. Migrate uses a Brownian motion approximation stepwise mutation model as well as Bayesian inference to determine effective population size and past migration rates [70]. Migrate assumes a migration matrix model that uses asymmetric migration rates and assumes different sub-population sizes with population divergence and admixture present. Migrate was run using a Brownian motion model with priors for theta (Θ) set from 0-1000. Simulations used one long chain with sample increments of 200 and recorded 5000 steps per chain after burn in of 1000 steps. To extend the length of the run and allow for greater convergence a multiple Markov chain statistical heating scheme was used with four chains with temperatures of 1.00, 1.50, 3.00, and 1,000,000.00 respectively with the swapping interval set to 1.

BayesAss edition 3 (BA3; [71]) was used to examine more recent (past few generations) levels of migration between the same four TESS3r clusters used in Migrate. BayesAss takes a Bayesian approach to estimating recent migration using MCMC to estimate posterior probabilities

[71]. BayesAss allows for within population frequencies to deviate from HWE and uses the temporary states of disequilibrium to make inferences about recent population gene flow [71]. Input files for BayesAss were formatted using Formatomatic [72]. BayesAss was initially run with 10,000,000 iteration of the MCMC chain utilizing a burn in of 1,000,000, and a sampling frequency of 1,000. These initial parameters did not allow for proper convergence of the Markov Chain so the number of iterations was increased to 1,000,000,000. Tracer v1.7.2 was used to calculate 95% confidence intervals and to view trace files from BayesAss [73].

Population sources and sinks were identified following the methods of Ishiyama et al. (2015), where the net immigration (immigration-emigration) was calculated for each pair of clusters and then averaged. A cluster with a negative net immigration is considered a source population whereas a cluster with a positive net immigration is considered a sink.

1.3.5 Isolation by Distance and Geographic Isolation

Geographic distance and watershed have both been identified as potential drivers of population structuring in *E. blandingii* [26,28]. A Mantel test was run using Adegenet in R to test for isolation by distance between individuals using 9999 permutation and pairwise F_{ST}. To test for effects of watershed on clustering, sites were grouped based on shared watersheds at the HUC-8 (cataloguing unit) and HUC-6 (accounting unit) levels (Figure 1.12). HUC-6 was the largest watershed level used because the groupings did not differ at the HUC-4 or HUC-2 levels. An AMOVA was then used to assess variation within and between sites at different watershed levels. Arlequin v. 3.5.2.2 was used to run the AMOVAs following the approach of Sethuraman et al. (2014), in which 16,000 permutations were used [75].

1.3.6 Bottlenecking and Effective Population Size

Population bottlenecks tend to occur when populations experience a large reduction in effective population size, and results in a reduction of the number of alleles present among polymorphic loci [76]. A loss of alleles leads to a direct loss in genetic diversity which can make a population more vulnerable to environmental change and stochasticity by constraining the available genetic plasticity [77]. When bottlenecks occur the number of alleles present in a given population tend to drop more quickly than the expected heterozygosity causing the expected heterozygosity to be greater than the observed heterozygosity (heterozygosity excess) [76]. BOTTLENECK version 2.2.02 uses each population and loci to examine the expected vs observed heterozygosity relative to the number of individuals and alleles used in each population [76,78,79]. Since Davy and Murphy ([80]) had similar sample sizes and numbers of locations and loci for a similar long lived species of turtle, the same BOTTLENECK parameters were used. BOTTLENECK was run using a two-phase model replicated 1000 times to check for evidence of population bottlenecking. Variance in the model was set at 12%, and single step mutation rate was set at 95% whereas multistep mutation rate was set to 5%.

Since *E. blandingii* can be difficult to capture and have differing reproductive output based on age, it is important to understand the number of individuals in a population that are actually contributing to the next generation (effective population size, N_E) [81,82]. NeEstimator Version 2.1 was used to assess the effective population size of each site. NeEstimator uses linkage disequilibrium under a molecular co-ancestry method to determine the N_E and also provides jackknifed confidence intervals [83,84]. NeEstimator was run with the Linkage Disequilibrium random mating model. To evaluate the impact of sample size on N_E estimation, NeEstimator was run using three different randomly generated sub-samples from the largest site (OH08): one with

the 10 individuals (our N cut off), one with 30 individuals (or N max), and one with the entire 77 individuals sampled.

Bottlesim v. 2.6 was used to explore the impacts of potential future bottlenecks in the absence of gene flow [24]. Bottlesim was run for our largest population OH08 with the following assumptions: initial population size of 200 for all scenarios, a 1:1 sex ratio, dioecy with random mating, and diploid multilocus individuals. Longevity was set to 65, and age of maturity was set to 14 following Anthonysamy et al. (2018). In the first scenario, the population declines 50% over 200 years, and the second scenario saw a 90% population decline. Both scenarios were run with 1000 replicates.

1.4 Results

1.4.1 Hardy Weinberg Equilibrium and Disequilibrium

Out of 224 tests of HWE only GmuD40 showed deviation from HWE at one locality (OH-17). Since only one site exhibited deviation from HWE all loci were retained. GmuD28 and GmuD107 (p= <1.02e-19) showed significant linkage disequilibrium after Bonferroni correction. GmuD28 was removed from further analysis, and GmuD107 was retained. Descriptive and Frequency Based Statistics

One hundred and sixty-nine alleles were detected across 14 loci representing 16 sample localities and a total of 313 individuals (Table 1.2). Of the 14 loci used for analysis, GmuD79 was monomorphic and uninformative leading it to be ignored for analysis. The number of alleles per locality ranged from 68 to 97 and increased according to the number of individuals sampled (Table 1.2). The number of alleles by sample size at a location appears to have reached or come close to an asymptote, with sites having around 20 individuals having nearly the same or more alleles than

the largest samples (n = 30) (Figure 1.3). MI-7 was an exception having around 15 more alleles than any other site. Of the 315 individuals and the 14 microsatellites used 1.51% of the genotype data was missing. The number of private alleles ranged from 0 to 7 for the sites with at least 10 samples (Table 1.2).

Mean allelic richness ranged from 4.01 to 4.84 but no statistical difference (P>0.05) was observed between localities (Table 1.2). Observed heterozygosity across sites ranged from 0.58 to 0.65 (Table 1.2). F_{IS} values indicated no evidence of inbreeding among sites, however there is potential outbreeding in MI-6 and MI-10 (F_{IS} values below negative for the bootstrapped 95% confidence interval). Overall F_{ST} was 0.05 (pairwise F_{ST} = -0.01-0.15) and overall D was 0.08 (pairwise D = 0.00-0.17). The F_{ST} and D values showed a general pattern of increased differentiation by geographical distance from site to site Table (1.3).

1.4.2 Structure Analysis and Genetic Clustering

Using LOCPRIOR, K values of 6 and 7 were identified for the MedMed/MedMeaK and MeaxMed/MaxMeaK methods respectively, at the 0.5 assignment threshold (Figure 1.5). The K=6 cluster scheme maintained the IN01, IN06, and IN07 cluster, the MI05, MI06, MI07, MI10, MI15, OH01, and OH18 cluster, the OH06, OH08, and OH09 cluster, and the OH16 and OH17 clusters (Figure 1.6). However K=6 increased the amount of admixture in OH01 and OH16 but reduced the admixture in OH13 grouping primarily with the OH06, OH08, and OH09 cluster. The K=6 scheme also introduced a new cluster (pink Figure 1.6) that gradually increases in degree of admixture west to east from Indiana and Michigan and reaching the largest degree of assignment in OH01, before gradually decreasing across the rest of Ohio (Figure 1.6). The K=7 cluster maintained the same overall clusters as K=6, with the exception of IN07 which clustered on its own (Figure 1.6). To visualize sub-clusters by localities, the STRUCUTRE results for K=6 were

displayed over geographic space since it produced more conservative estimates than *K*=7 (Figure 1.7). Only the results using a 0.5 threshold level are displayed since this is the standard, however Puechmaille (2016) recommends varying different threshold levels to make more stringent clusters with less extraneous groupings. We viewed the STRUCTURE results under a 0.8 threshold level which uncovered a *K*=4 for the MedMedK and MaxMedK and a *K*=3 for the MeMeaK and MaxMeaK. The *K*=4 clustering scheme produced by STRUCTURE was very similar to TESS3r but clustered OH17 on its own and OH16 with OH06-OH13. The *K*=3 maintained OH17 as its own cluster and grouped Indian, Michigan and the Ohio localities east of OH06 together.

TESS3r identified four clusters using cross validation scores (Figure 1.8). TESS3r maintained a similar clustering pattern as STRUCTURE, but created a more gradual west-east cluster pattern. TESS3r clustered IN01, IN06, IN07, MI05, MI06, MI07, MI10, MI15, OH01, and OH18 as a cluster, OH06, OH08, OH09, and OH13 as a cluster, and OH16 and OH17 as a cluster refining the east to west trend indicated by STRUCTURE (Figures 1.6, 1.7, 1.8, & 1.9).

The DAPC run through Adegenet identified five clusters (using BIC vs # of Clusters) (Figure 1.10). Cluster 1 contained 2.5-12.5 individuals from each locality in Ohio and Michigan, except OH17 which had fewer than 2.5 individuals assigned to cluster 1. Less than two individuals from IN07 were also assigned to cluster 1(Figure 1.10). Cluster 2 contained the majority of OH17 (Figure 1.10). Cluster 3 contained the majority of the individuals from Indiana localities and 7.5 a few individuals from each Michigan localities (Figure 1.10). Cluster 4 and 5 were very similar except cluster 4 contained individuals from IN07, MI05, MI06, MI15 and no individuals from OH17 (Figure 1.10).

1.4.3 Migration

Mean historic mutation scaled migration rates ranged from 6.90 to 20.20 (or 23.86 and 102.56 individuals per 4 generations) with the highest rate of average migration being from cluster 2 to cluster 4, and the lowest average migration rate being from cluster 1 to cluster 2 (Table 1.6 and Figure 1.13). Migrate results indicate near equal amounts of migration between clusters 1 and 3, and clusters 1 and 4, (Table 1.6 and Figure 1.13). Cluster 4 historically contributed about half of the amount of mean migration to cluster 2 as cluster 2 had to cluster 4 (Table 1.6 and Figure 1.13). Cluster 2 had nearly three times the historic genetic contribution to cluster 1 as cluster 1 had to cluster 2 (Table 1.6 and Figure 1.13). Mean recent migration rates ranged from 0.08 to 0.09 however all values crossed zero at the 95% confidence interval indicating little to no recent migration (Table 1.6 and Figure 1.13). The same general trend was seen in the mean recent migration rates as was observed in the historic analysis with the exception of migration between cluster 1 and 3 (Table 1.6 and Figure 1.13).

Based on mean historic net immigration rates from Migrate, cluster two saw a net negative immigration rate implicating it as a source population at least in a historic sense to the other three clusters. Clusters 1, 3, and 4 all had positive net mean immigration rates (Figure 1.14). Looking at recent net immigration clusters 2 and 3 both had negative net immigration of about three and two individuals per generation, respectively (Figure 1.14). Whereas cluster 1 and 4 still saw a net increase in immigration of about three individuals per generation (Figure 1.14).

1.4.4 Isolation By Distance and Watershed Isolation

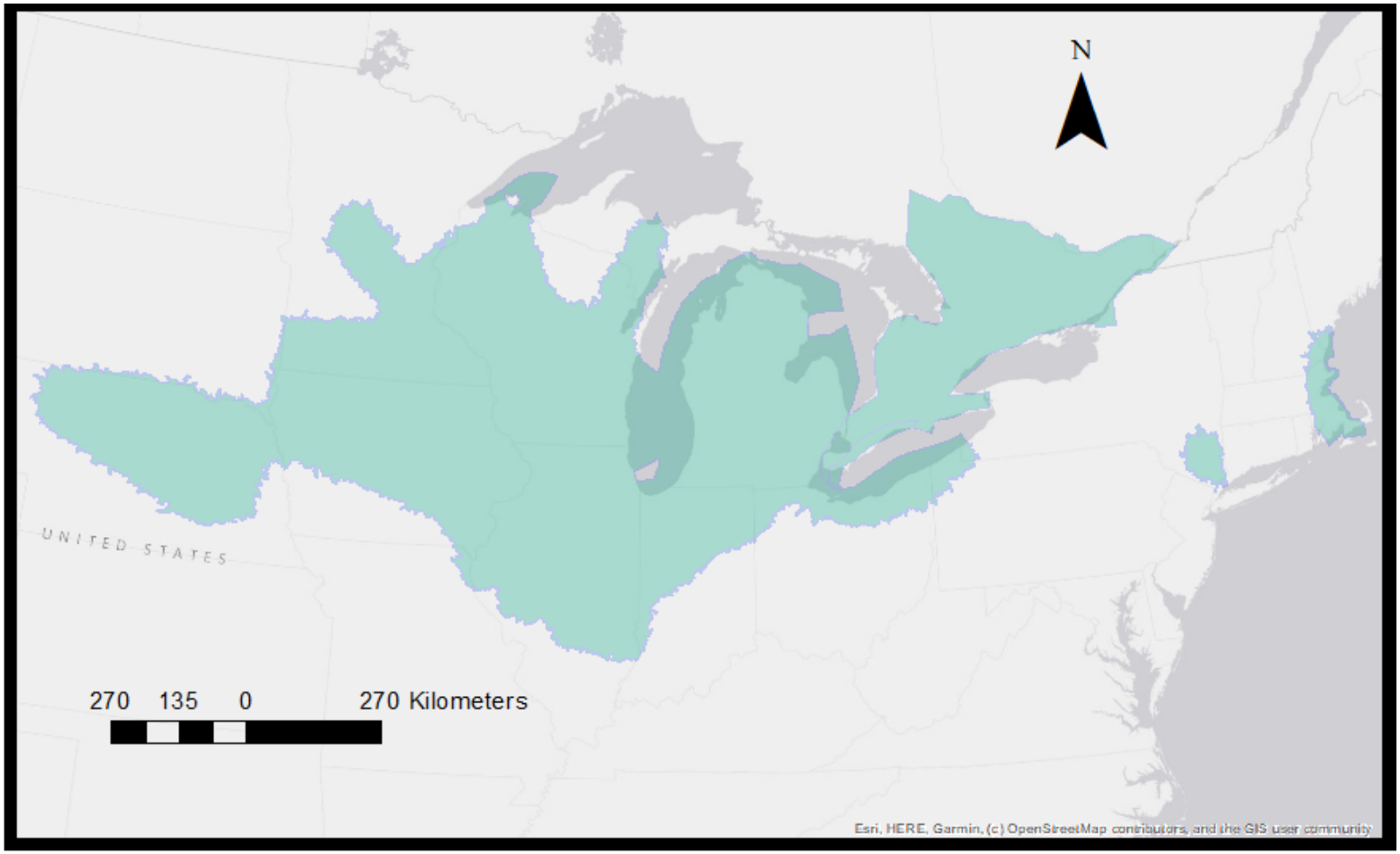
The observed variance for the Mantel test was 0.69, and fell on the high end of the normal distribution of similarity indicating significant IBD in the dataset (Figure 1.11, $r^2 = 0.31$, $P < 0.0001$). AMOVA results showed statistically significant ($p<0.05$) differentiation by watershed

grouping at both the HUC6 and HUC8 levels (Table 1.4). As expected, the highest level of variance was explained at the within population level for both the HUC6 and HUC8 watershed clustering schemes (92.62% and 94.70%) (Table 1.4). Among groups had relatively low levels of variance explained (HUC6 = 4.10% and HUC8 = 3.4%), but both explained more variance than between populations within groups (3.28% and 1.87% respectively) (Table 1.4).

1.4.5 Bottlenecking, Effective Population Size

No bottlenecks were detected in most localities, however a Wilcoxon test suggested heterozygosity excess in OH09. Conversely, sign tests indicated heterozygosity deficiency in IN01, MI06, OH01, and OH16 (Table 1.7). Effective population size estimates ranged from 7.9 to ∞ (Pcrit=0.05) and 8.5 to ∞ (Pcrit=0.02) (Table 1.7). However, effective population size estimates appear to be highly variable based on the sample size used to determine the estimate (Table 1.8). Sub-sampling from 77 individuals to 30 individuals seemed to have a much more marginal impact on estimations of effective populations size with the estimates varying by 30 to 40 individuals (for the Pcrit= 0.05 and 0.02 respectively) (Table 1.8).

Over a 200 year period Bottlesim models indicated a ~9% loss of alleles under a constant population size, a ~17% loss of alleles under a 50% population reduction, and a >50% reduction of alleles in a 90% population reduction (Figure 1.4). The bottleneck models indicated a ~3% loss of alleles under a constant population size, a ~5% loss of H_O under a 50% population reduction and a ~14% reduction of alleles in a 90% population reduction over 200 years (Figure 1.4).


N
UNITED STATES
270 135 0 270 Kilometers
Esri, HERE, Garmin, (c) OpenStreetMap contributors, and the GIS user community

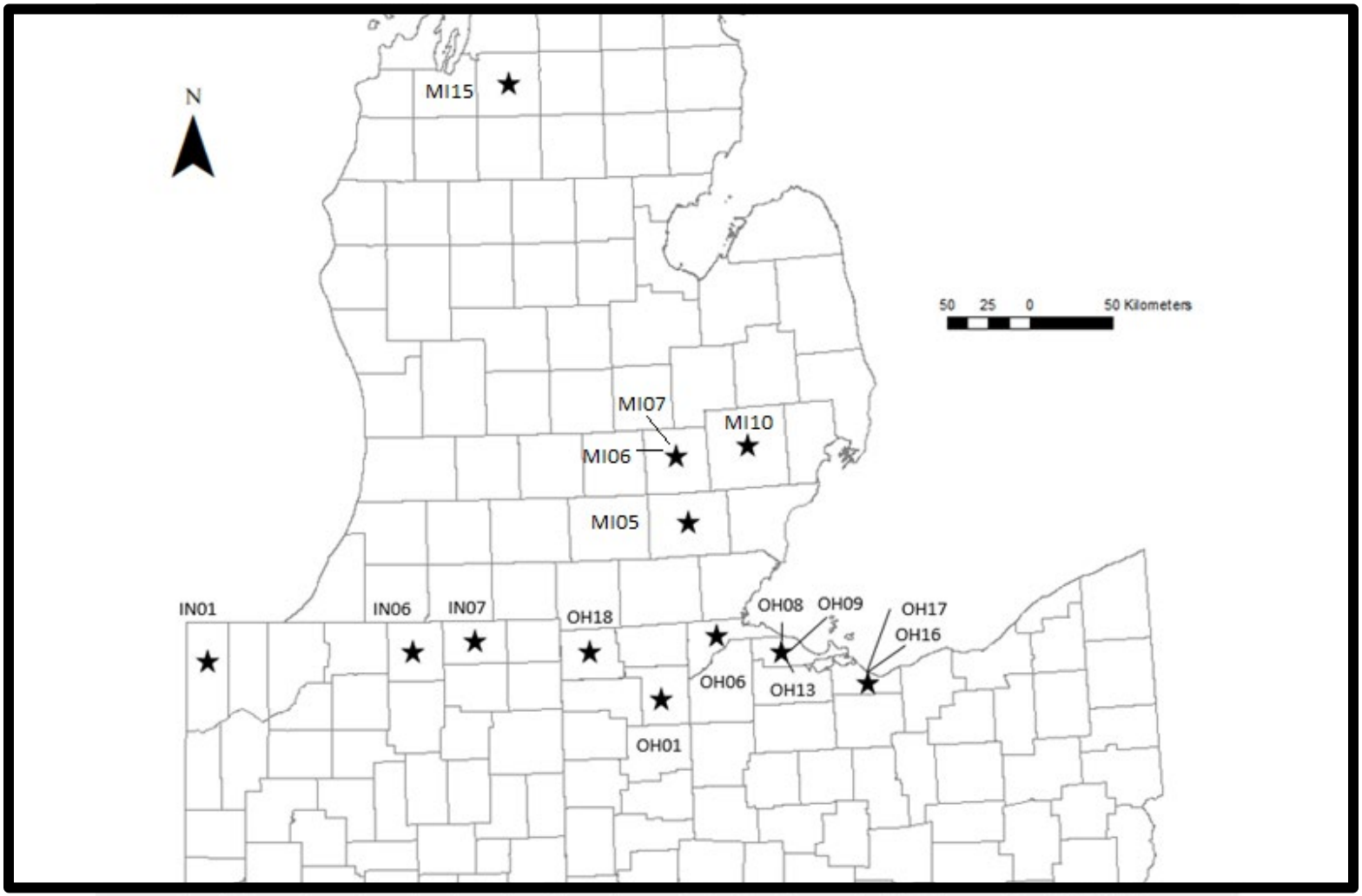


Figure 1.2 Site Localities with greater than 10 turtles.

Table 1.1 Summary of fifteen microsatellite loci run for all samples. Universal tail sequences were bound to the 5' tail of the forward primer following Blacket et al., (2014). Microsatellites were grouped run together in a multiplex PCR reaction.

Locus	Size range (bp)	Forward primer tail/label	Forward primer sequence (5'-3')	Reverse primer sequence (5'-3')	Multiplex group	Number of alleles	Label Volume (ul)	Forward Primer Volume (ul)	Reverse Primer Volume(ul)
GmuD55	175-220	Universal NED	GTG ATA CTC TGC AAC CCA TCC	TTG CAT TCA GAA TAT CCA TCA G	1	12	1.0	9.0	10.0
GmuD90	122-134	Universal FAM	ATA GCA GGA CAA TTA CCA CCA G	CCT AGT TGC TGC TGA CTC CAC	1	3	0.5	9.5	10.0
GmuD87	190-255	Universal VIC	AAA CCC TAA GAC ATC AGA CAG G	CAA ATC CAG TAC CCA GAA AGT C	1	10	1.0	9.0	10.0
GmuD88	115-171	Universal VIC	AAC AAT GCC TGA AAA TGC A C	AGG CTA CCT CTG AAA ATG CTG	2	12	0.5	9.5	10.0
Cp2	187-229	Universal PET	C TCT AAG GGT TGC ACT TCT CAA A	GAG GTG GCA TCA AAA CAT CAT	2	9	1.0	9.0	10.0
GmuD28	180-230	Universal NED	AGC TGT TTG TCA TCA TAC ACT CTC	TGG CCC TCA TGT TTT ATA AGT G	2	14	2.0	8.0	10.0
BTCA9	147-188	Universal FAM	TAC TCA AGA TTT GAA GCA GAT ACA	GGC TTG ATT CTA CTG TCA CTT AC	2	11	1.0	9.0	10.0
Eb19	97-110	Universal NED	AGG GCT CTG AAG CAC TAA AGT AA	GGC ACT GAA ATA AGA GAA AGT A	3	3	1.0	9.0	10.0
GmuD93	185-389	Universal VIC	AGA CTC TCT TGA CCA GAT TTT CTC	TCT GCC TTC TAT CAC TCT CCT G	3	2	1.0	9.0	10.0
GmuD107	189-209	Universal FAM	GAC AAA CAT GAA CAG GAG AAG AG	ATT AGA GAG ACA GAT AGA TAG GAC TTG	3	10	1.0	9.0	10.0
Eb17	94-117	Universal VIC	CCC ACA AAA GTA GAC ACC TAT	GGC ACT GAA ATA AGA GAA AGT A	4	6	1.0	9.0	10.0
GmuD121	138-178	Universal FAM	GGCAAA TAT CCA ATA GAA ATC C	CAA CTT CCT CGT GGG TTC AG	4	7	1.0	9.0	10.0
GmuD79	164-192	Universal PET	GCC CTG TTC CAT TCT TAT TCT G	ATC CCC TTA GTC GTC TCT TTT C	4	1	1.0	9.0	10.0
GmuD16	149-210	Universal NED	ATC CCT GAA ATT TTG TGT GTT C	TTT ACT CTA GAA GGG GCA ATC C	5	15	0.5	9.5	10.0
GmuD40	182-285	Universal PET	T TTG TCA TAT CAT CCA CTC ACC	TTT GTC ACA GAT GGG AAT TAG C	5	25	2.0	8.0	10.0

Table 1.2 Summary of descriptive statistics by locality. Listed statistics include number of turtles sampled (N), total number of alleles (A), number of private alleles (PA), mean allelic richness (AR with SE), mean observed Heterozygosity (HO with SE) expected heterozygosity (HE with SE), and inbreeding coefficient (FIS includes 95% confidence interval).

Site	State	County	HUC6	HUC8	N	A	PA	AR	H_O	H_E	F_{IS}
IN-1	Indiana	Lake	1	1	12	70	4	4.08	0.59	0.56	-.06
IN-6	Indiana	Elkhart	2	2	15	75	3	4.27	0.63	0.60	-0.05
IN-7	Indiana	LaGrange	2	2	22	92	7	4.77	0.64	0.64	0.00
MI-5	Michigan	Washtenaw	4	6	20	84	1	4.42	0.61	0.61	-0.01
MI-6	Michigan	Livingston	4	6	16	86	3	4.60	0.65	0.63	-0.03
MI-7	Michigan	Livingston	4	6	30	97	1	4.84	0.64	0.66	0.03
MI-10	Michigan	Oakland	4	6	11	74	0	4.40	0.62	0.61	-0.01
MI-15	Michigan	Crawford	3	3	10	71	1	4.32	0.59	0.60	0.03
OH-1	Ohio	Henry	4	5	11	81	6	4.75	0.58	0.62	0.06
OH-6	Ohio	Lucas	4	7	30	87	1	4.44	0.61	0.62	0.03
OH-8	Ohio	Ottawa	4	7	30	81	2	4.22	0.61	0.61	-0.01
OH-9	Ohio	Ottawa	4	7	14	68	0	4.06	0.64	0.62	-0.04
OH-	Ohio	Ottawa	4	8	21	88	0	4.58	0.62	0.63	.01
OH-	Ohio	Erie	4	8	30	79	1	4.12	0.63	0.61	-0.04
OH-	Ohio	Erie	4	8	29	76	2	4.01	0.65	0.62	-0.04
OH-	Ohio	Williams	4	4	12	68	2	4.07	0.58	0.58	-0.00

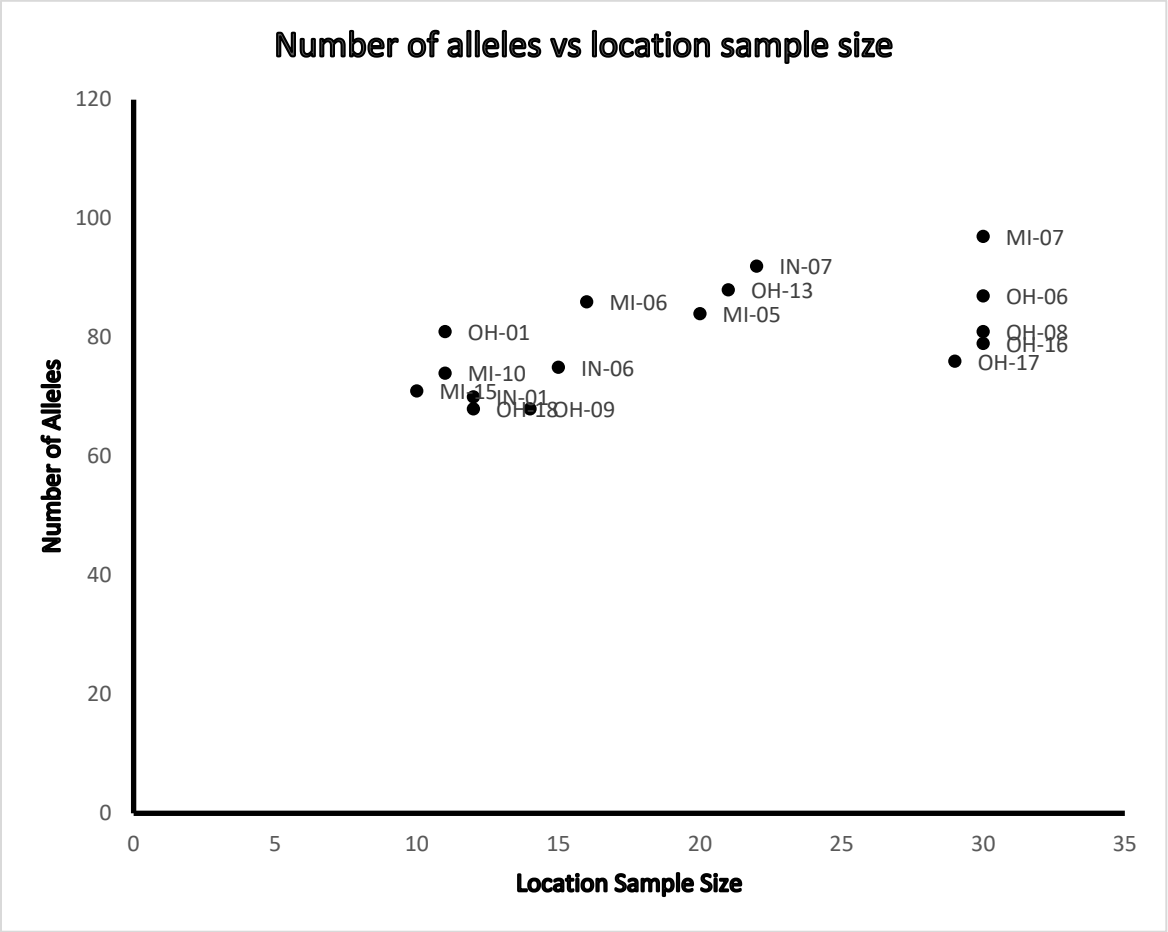


Figure 1.3 Number of alleles versus sample size by locality.

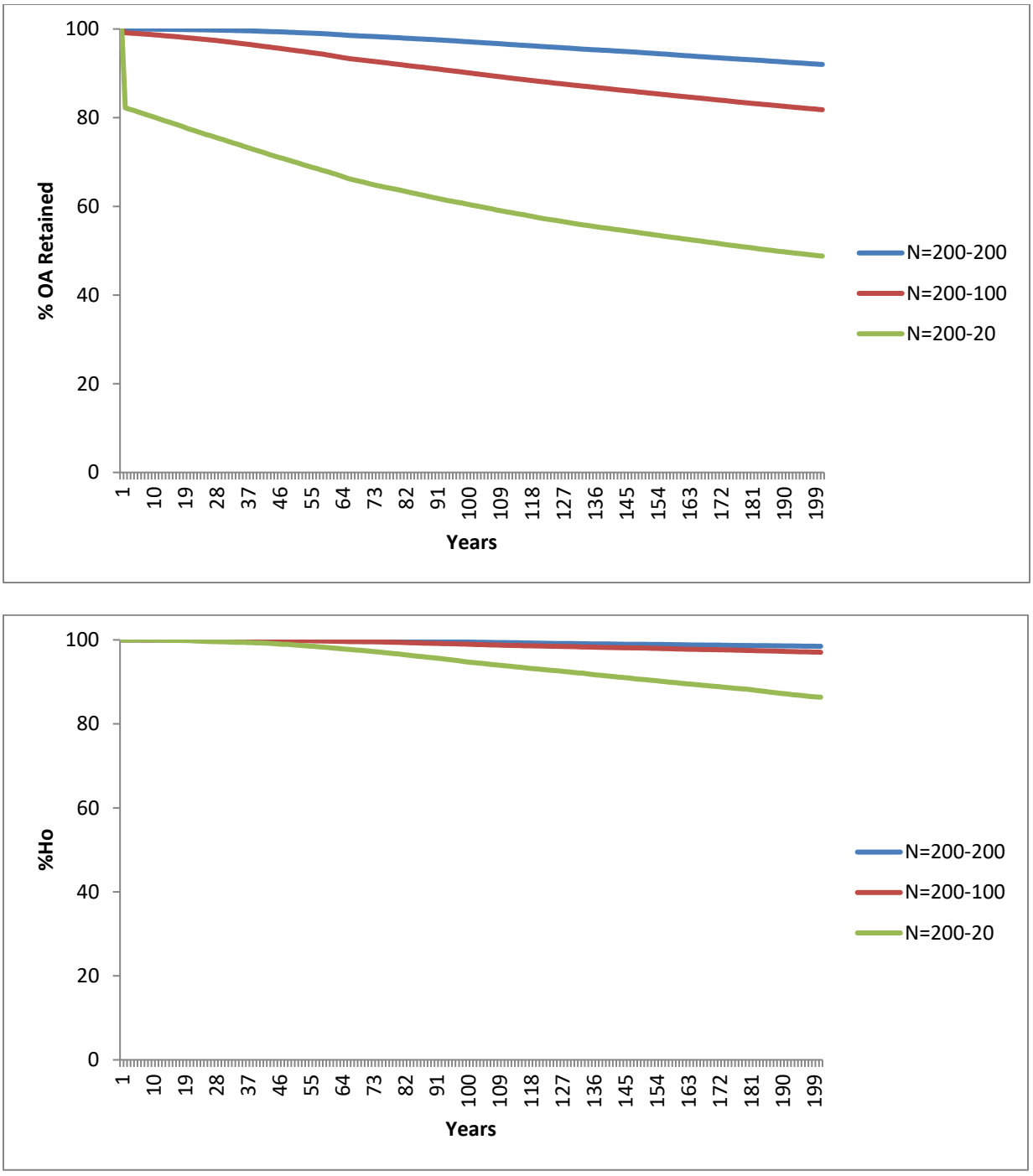


Figure 1.4 Observed alleles and observed heterozygosity over 200 years for three different bottleneck scenarios. N = 200-100 Bottleneck resulting in 50% population decline. N = 200-20 Bottleneck resulting in a 90% population decline. N = 200-200 Bottleneck in population with stable population size.

Table 1.3 Summary of pairwise FST and D (Jost D) differentiation scores. FST below and D above. Bold values do not include 0 in the 95% confidence intervals.

	IN01	*IN06*	*IN07*	*MI05*	*MI06*	*MI07*	*MI10*	*MI15*	*OH01*	*OH06*	*OH08*	*OH09*	*OH13*	*OH16*	*OH17*	*OH18*
IN01	0	0.0284	0.0441	**0.08**	**0.0945**	**0.0681**	0.0541	0.0695	**0.0683**	**0.1465**	**0.1727**	**0.1739**	**0.1418**	**0.1349**	**0.2093**	**0.0758**
IN06	**0.0448**	0	0.0283	0.0556	**0.0706**	**0.0548**	0.0491	0.0507	0.0652	**0.1395**	**0.1314**	**0.1446**	**0.0957**	**0.1112**	**0.1188**	**0.0543**
IN07	**0.0618**	**0.0451**	0	0.0422	**0.0683**	0.0332	0.036	0.0281	**0.0528**	**0.0972**	**0.1133**	**0.1243**	**0.0928**	**0.1049**	**0.1494**	0.0297
MI05	**0.0786**	**0.07**	**0.062**	0	-1.00E-04	0.0052	2.00E-04	-4.00E-04	0.0117	**0.052**	**0.0492**	**0.0549**	**0.042**	**0.041**	**0.0914**	3.00E-04
MI06	**0.0848**	**0.0844**	**0.0646**	0	0	1.00E-04	-0.0035	0	2.00E-04	**0.0497**	**0.0338**	**0.0322**	0.0195	**0.0441**	**0.1013**	5.00E-04
MI07	**0.0676**	**0.065**	**0.0418**	0.0084	0.0021	0	-0.0051	0	0.0031	**0.0422**	**0.0361**	**0.0258**	0.0147	**0.0284**	**0.093**	0.0027
MI10	**0.0643**	**0.0708**	**0.0573**	0.0035	-0.0123	-0.0056	0	-0.0079	0.0038	**0.0499**	**0.0495**	0.0132	0.0112	**0.0267**	**0.0995**	0
MI15	**0.0812**	**0.0708**	**0.0534**	0.0023	0.0068	-0.001	-0.0126	0	0.0012	0.0308	0.0304	0.0187	0.0043	0.008	**0.0595**	0
OH01	**0.0868**	**0.0689**	**0.0498**	**0.0206**	0.0013	0.0075	0.0052	0.0025	0	0.0048	0.015	0.0185	0.0089	0.0175	**0.0597**	0.0106
OH06	**0.1266**	**0.1103**	**0.0714**	**0.0467**	**0.0344**	**0.0304**	**0.0367**	**0.0324**	0.0164	0	9.00E-04	0	2.00E-04	0.0089	**0.086**	**0.038**
OH08	**0.1257**	**0.1062**	**0.0804**	**0.0421**	**0.0251**	**0.0311**	**0.0368**	**0.0341**	**0.0177**	0.0054	0	0.0021	0.0087	0.0114	**0.0929**	0.0259
OH09	**0.1236**	**0.1121**	**0.0862**	**0.0485**	**0.0305**	**0.0304**	**0.0355**	**0.0298**	**0.0274**	-0.0011	0.0103	0	0	0.0109	**0.0638**	0.0243
OH13	**0.1048**	**0.0884**	**0.0643**	**0.0325**	**0.0275**	**0.0142**	**0.0226**	0.0096	0.017	0.0069	0.0148	0.0076	0	0.0092	**0.0583**	0.0085
OH16	**0.1106**	**0.1034**	**0.0834**	**0.0435**	**0.0342**	**0.0246**	**0.0268**	0.0178	**0.0216**	**0.017**	**0.0256**	**0.0191**	**0.0114**	0	**0.068**	**0.0262**
OH17	**0.1488**	**0.1071**	**0.1055**	**0.0907**	**0.0798**	**0.0678**	**0.0842**	**0.0626**	**0.0615**	**0.0656**	**0.0715**	**0.0582**	**0.0548**	**0.0473**	0	**0.0516**
OH18	**0.0896**	**0.0661**	**0.0509**	0.0015	0.0062	0.0111	-0.0014	0.0042	0.0253	**0.0449**	**0.0367**	**0.0461**	**0.0293**	**0.0373**	**0.0721**	0

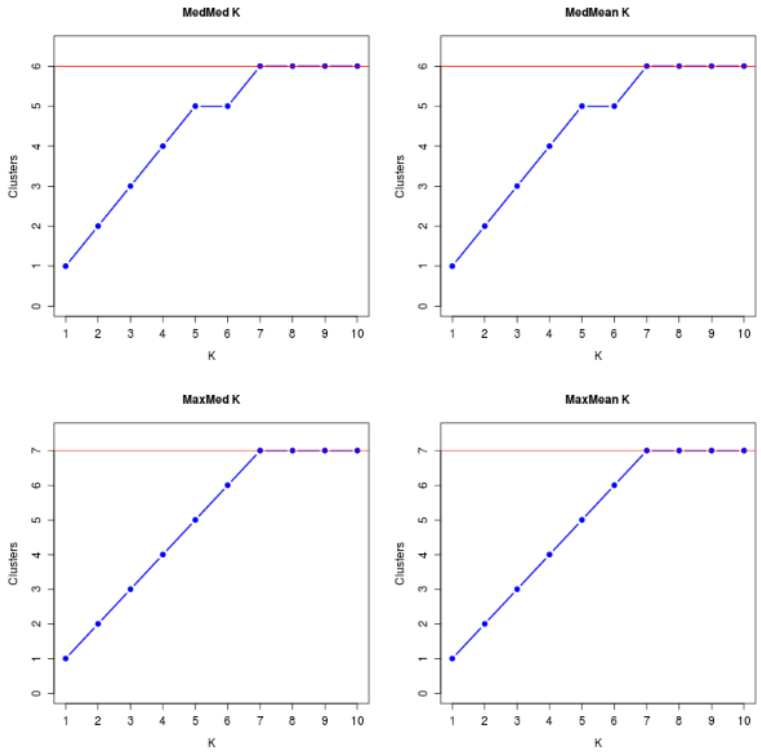


Figure 1.5 MedMeaK and MedMedK values reducing inclusion of extra clusters, estimated using the methods of Puechmaille (2016). Models are based on STRUCTURE using LOCPRIOR.

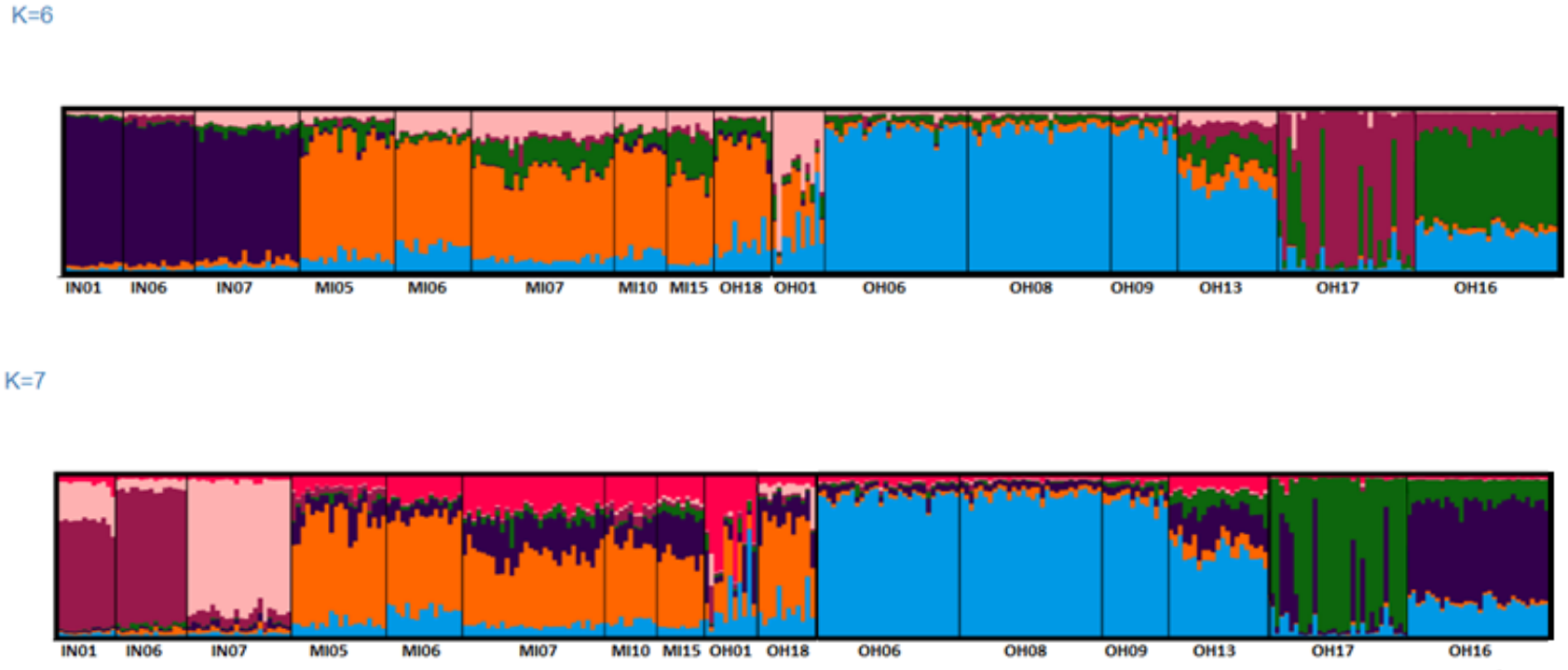


Figure 1.6 Bar graphs showing inferred clusters by individuals by site. K = 6 inferred from MedMeaK (top); K = 7 inferred from MedMedK (bottom). Both models derived from STRUCTURE using LOCPRIOR. Localities listed from west to east.

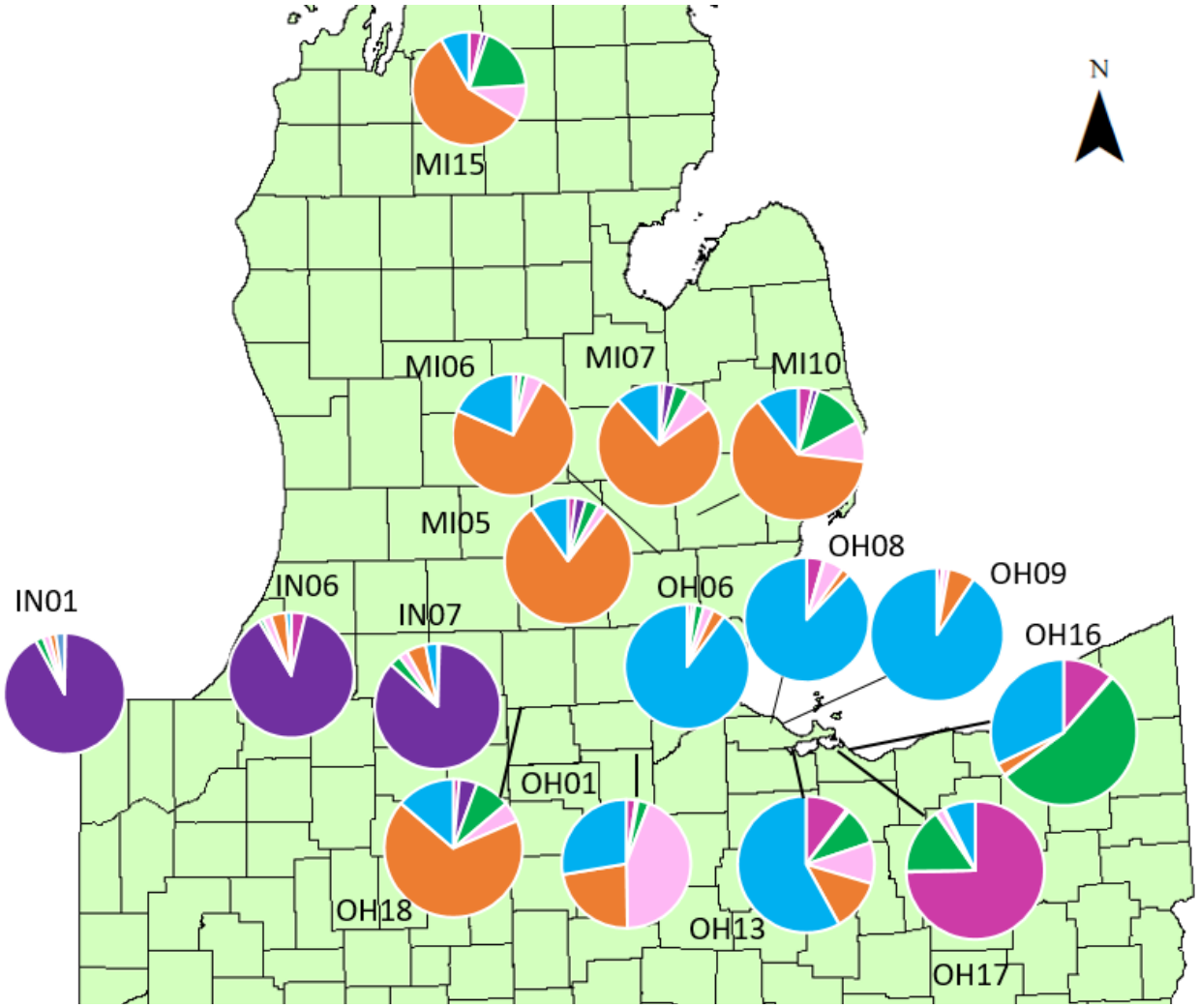


Figure 1.7 Map displaying STRUCTURE results for K = 6 clusters over geographic space.

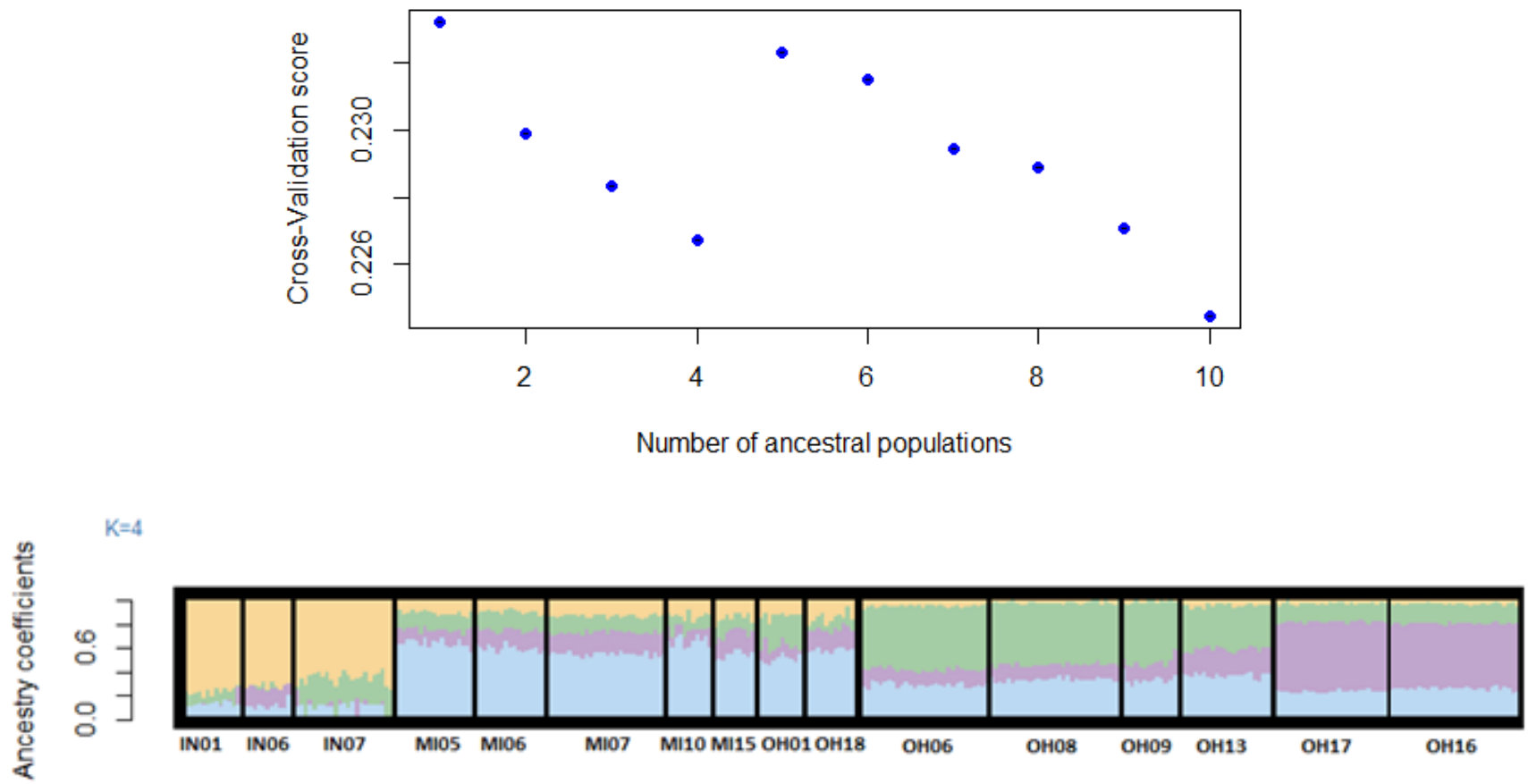


Figure 1.8 Cross validation score for inferring number of clusters (top). TESS3r bar graphs showing inferred clusters by individuals by site for K=5 localities listed from west to east (bottom).

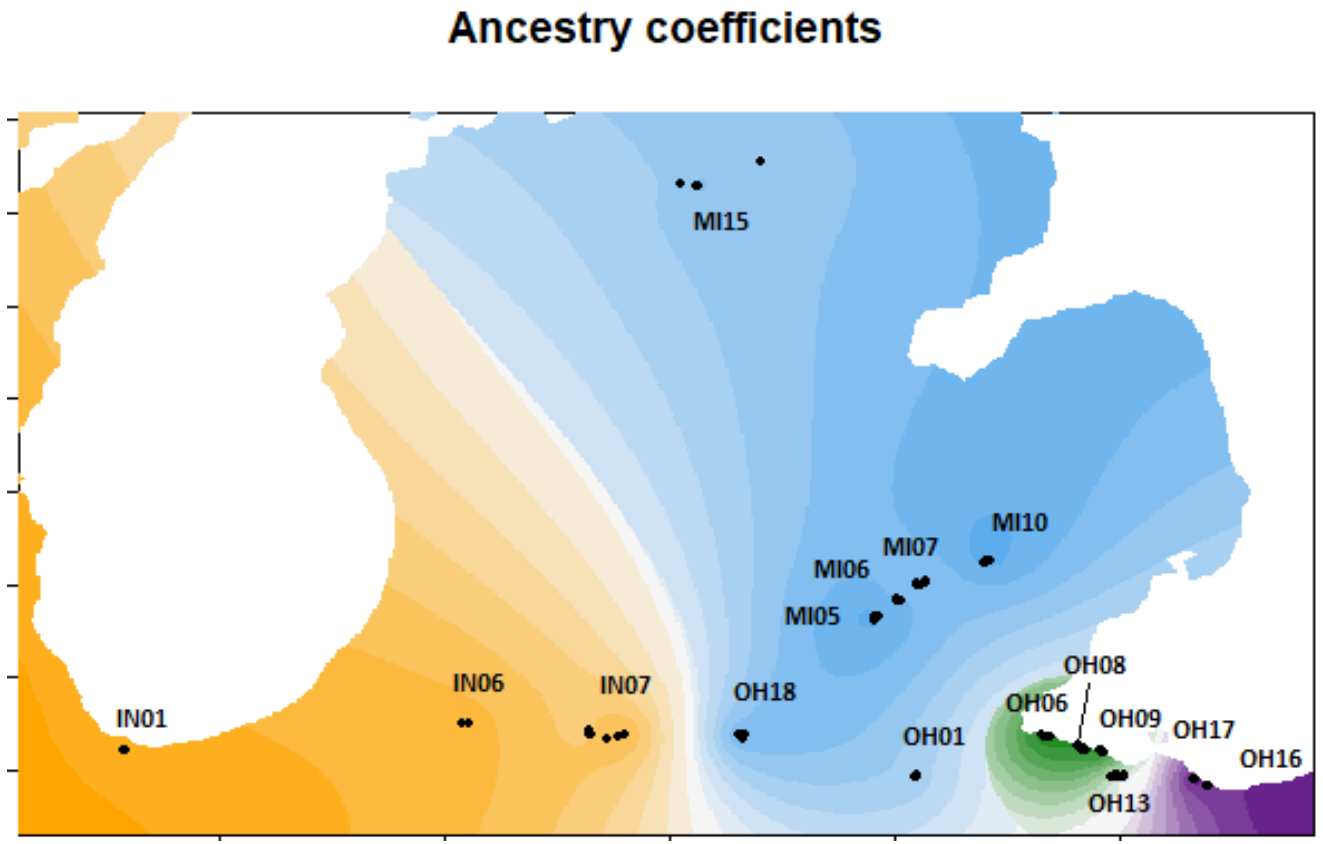


Figure 1.9 Inferred clusters displayed over geographic space for K=5 . Note that large portions of Michigan were not sampled and this geographic interpretation does not show admixture.

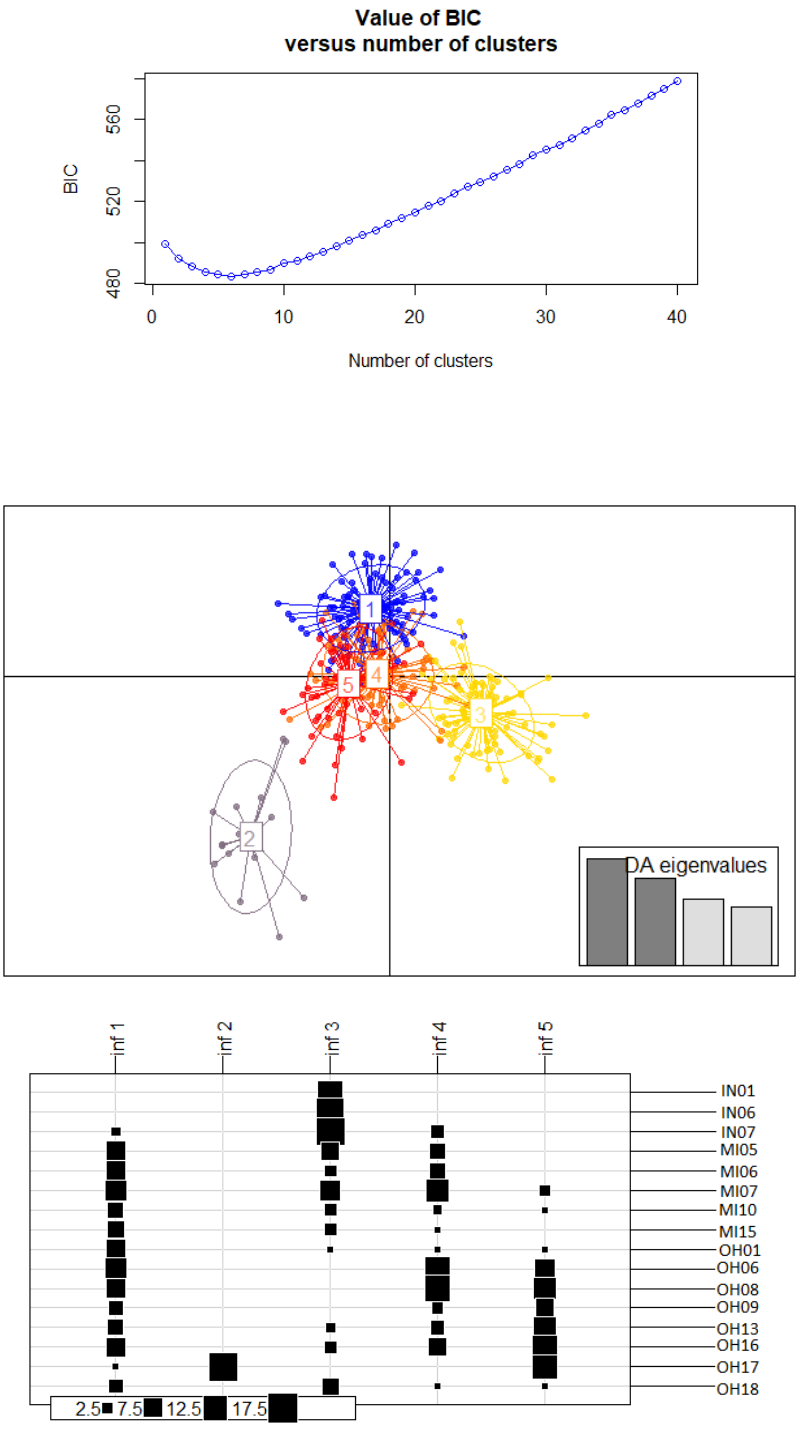


Figure 1.9 Values of BIC for inferring number of clusters (top). DAPC scatter chart of five inferred clusters (middle). Individual assignment from given populations to inferred clusters (bottom).

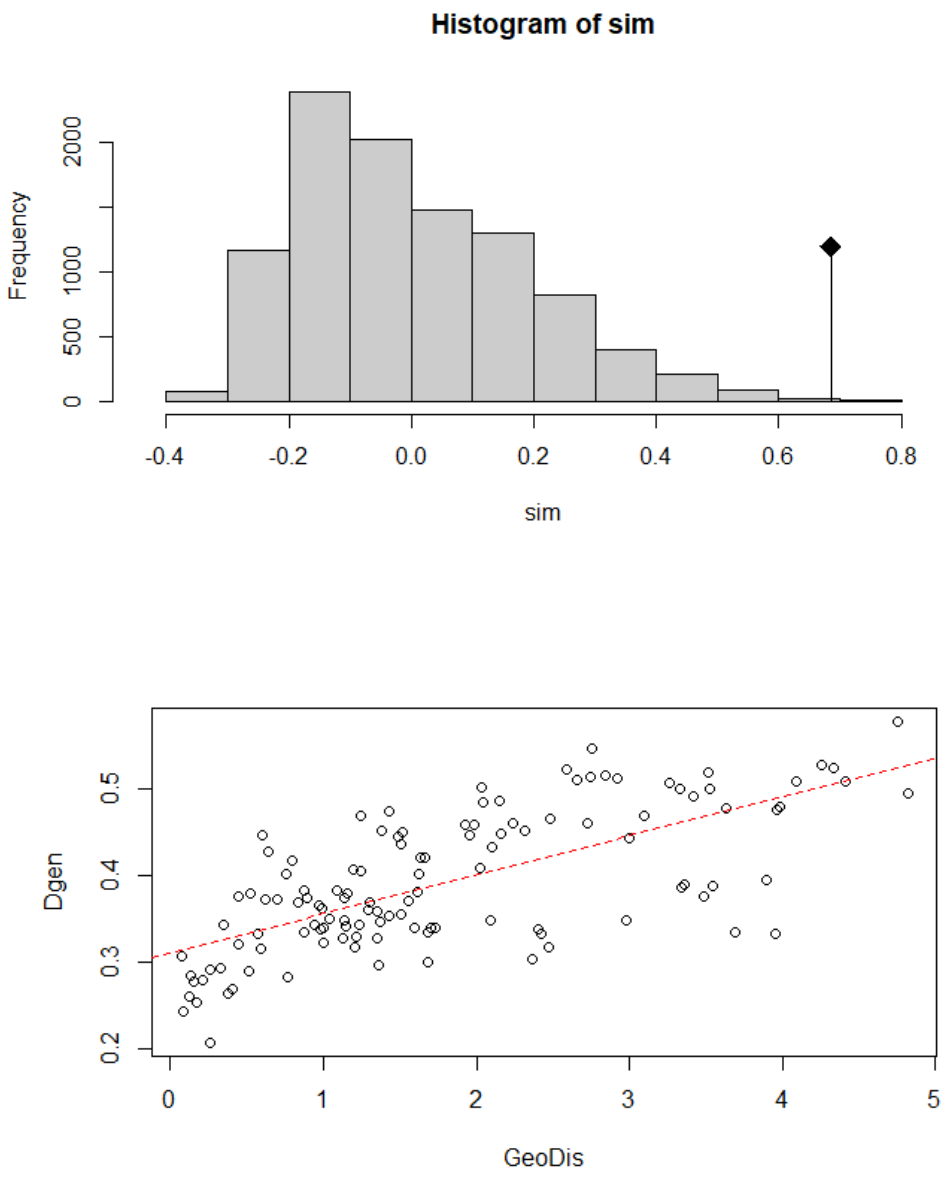


Figure 1.10 Mantel test for isolation by distance (IBD) (top) and regression line for individual genetic distance (Dgen) vs geographic distance (GeoDis), r2 = 0.31, P <0.0001 (bottom).

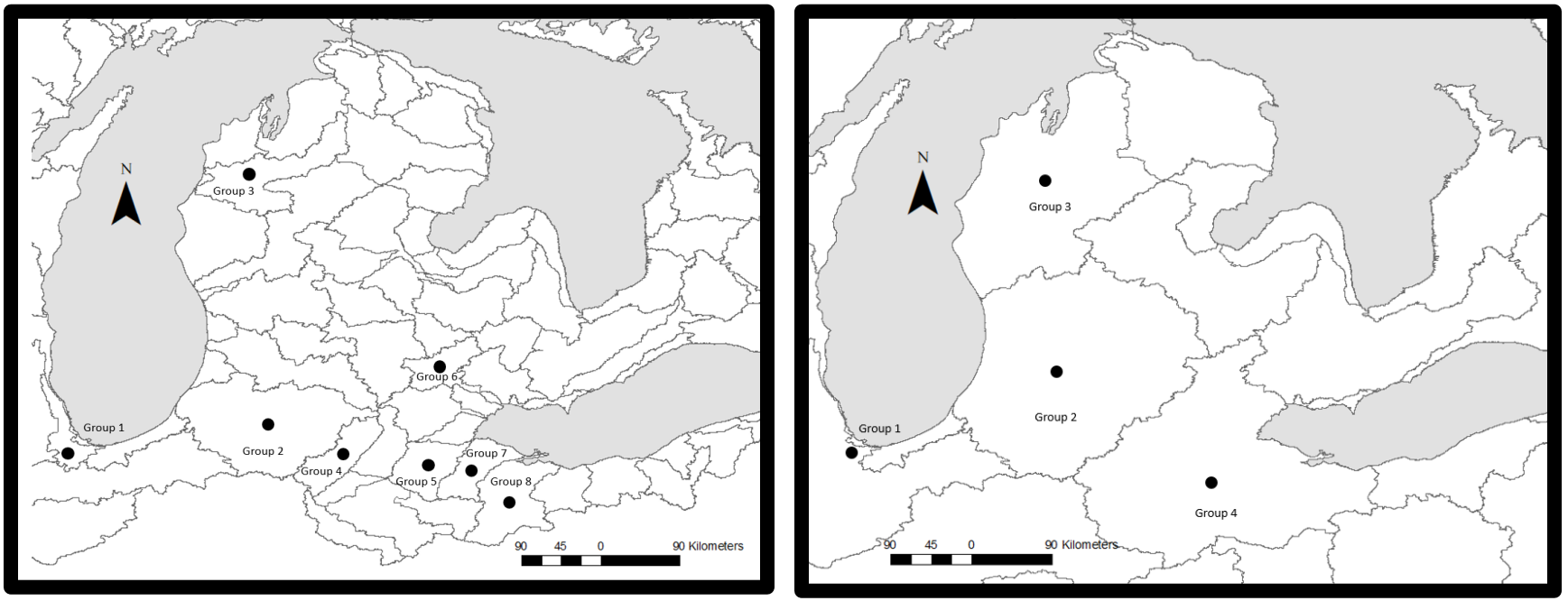


Figure 1.11 HUC 8 (left) and HUC 6 (right) watershed levels.

Table 1.4 AMOVA results displaying variance for the HUC6 (bottom) and HUC8 (top) watershed levels.

HUC Level	Source of Variation	d.f.	Sum of Squares	Variance Components	P-value	%of Variation	Fixation Indices
HUC8	Among Groups	7	127.71	0.16 Va	0.00	3.42	0.03 (F_{CT})
	Among Pop within Group	8	63.72	0.08 Vb	0.00	1.87	0.02 (F_{SC})
	Within Pop	610	2620.34	4.30 Vc	0.00	94.70	0.05 (F_{ST})
HUC6	Among Groups	3	64.54	0.19 Va	0.00	4.10	0.04 (F_{CT})
	Among Pop within Group	12	126.89	0.15 Vb	0.00	3.28	0.03 (F_{SC})
	Within Pop	610	2620.34	4.30 Vc	0.00	92.62	0.07 (F_{ST})

Table 1.5 Locality grouping based on TESS3r cluster assignment.

TESS3r Cluster	**N**	**Sites**
Cluster 01	49	IN01, IN06, IN07
Cluster 02	108	MI05, MI06, MI07, MI10, MI15, OH01, OH18
Cluster 03	97	OH06, OH08, OH09, OH13
Cluster 04	59	OH16, OH17

Table 1.6 Mean historic migration calculated through Migrate, recent migration rate from BayesAss. Migrate indicates the mean mutation-scaled migration rate with 95% CI. BayesAss displays mean migration rate with 95% CI derived from migration rate by population size.

TESS3r Cluster	**Mean Historic Migration Rate**	**Mean Recent Migration Rate**
Cluster 01-Cluster 02	6.90 (±6.56)	0.08 (±0.12)
Cluster 01-Cluster 03	16.09 (±8.25)	0.08 (±0.12)
Cluster 01-Cluster 04	8.51(±6.81)	0.09 (±0.13)
Cluster 02-Cluster 01	16.76 (±8.40)	0.08 (±0.12)
Cluster 02-Cluster 03	16.81 (±8.75)	0.08 (±0.13)
Cluster 02-Cluster 04	20.20 (±8.89)	0.09 (±0.13)
Cluster 03-Cluster 01	15.73 (±8.12)	0.08 (±0.13)
Cluster 03-Cluster 02	10.37 (±9.68)	0.08 (±0.12)
Cluster 03-Cluster 04	17.45 (±8.54)	0.09 (±0.13)
Cluster 04-Cluster 01	10.55 (±7.75)	0.08 (±0.12)
Cluster 04-Cluster 02	10.06 (±7.41)	0.08 (±0.13)
Cluster 04-Cluster 03	12.37 (±7.31)	0.08 (±0.13)

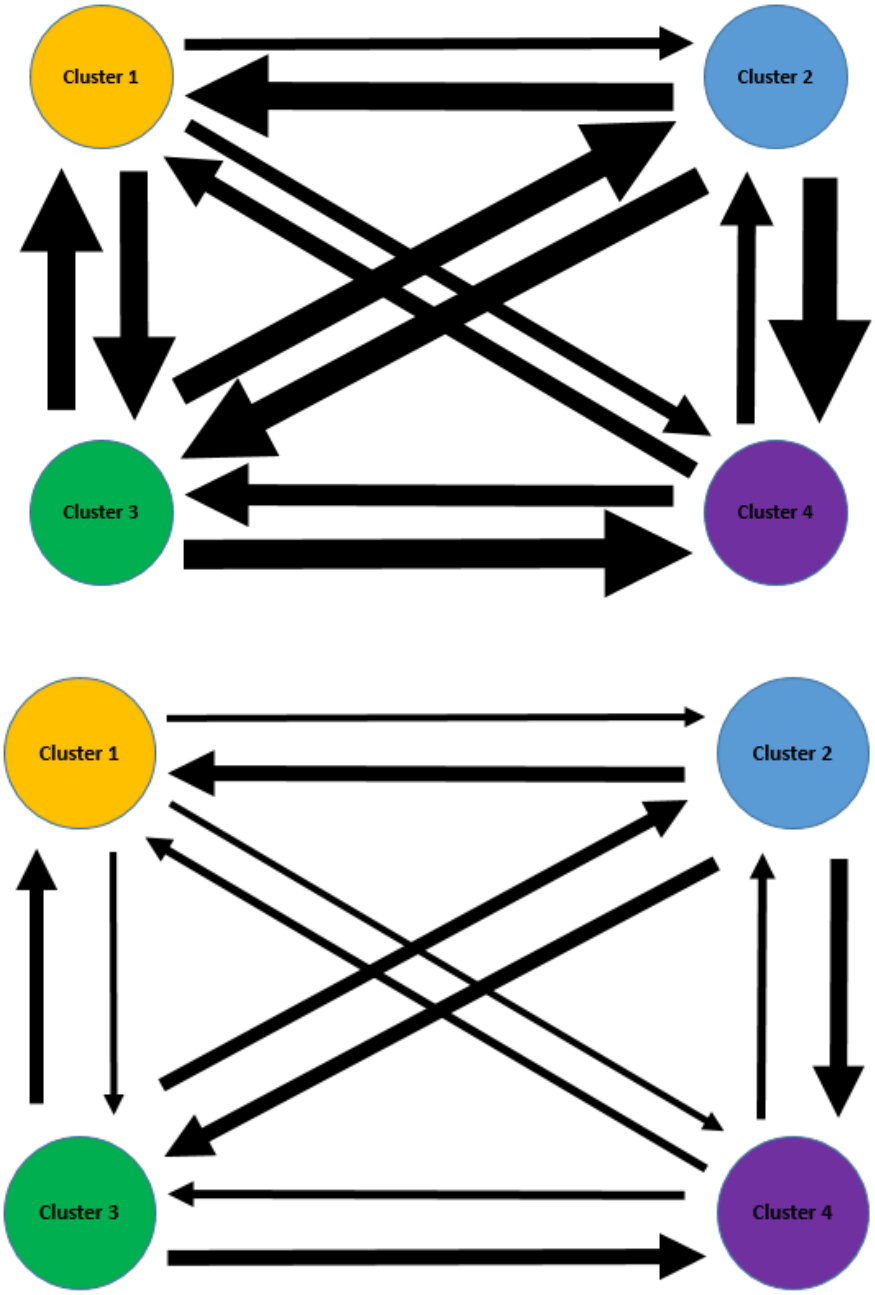


Figure 1.12 Mean historic migration rate (top), mean recent migration rate (bottom). Scaled arrows indicate direction and magnitude of migration.

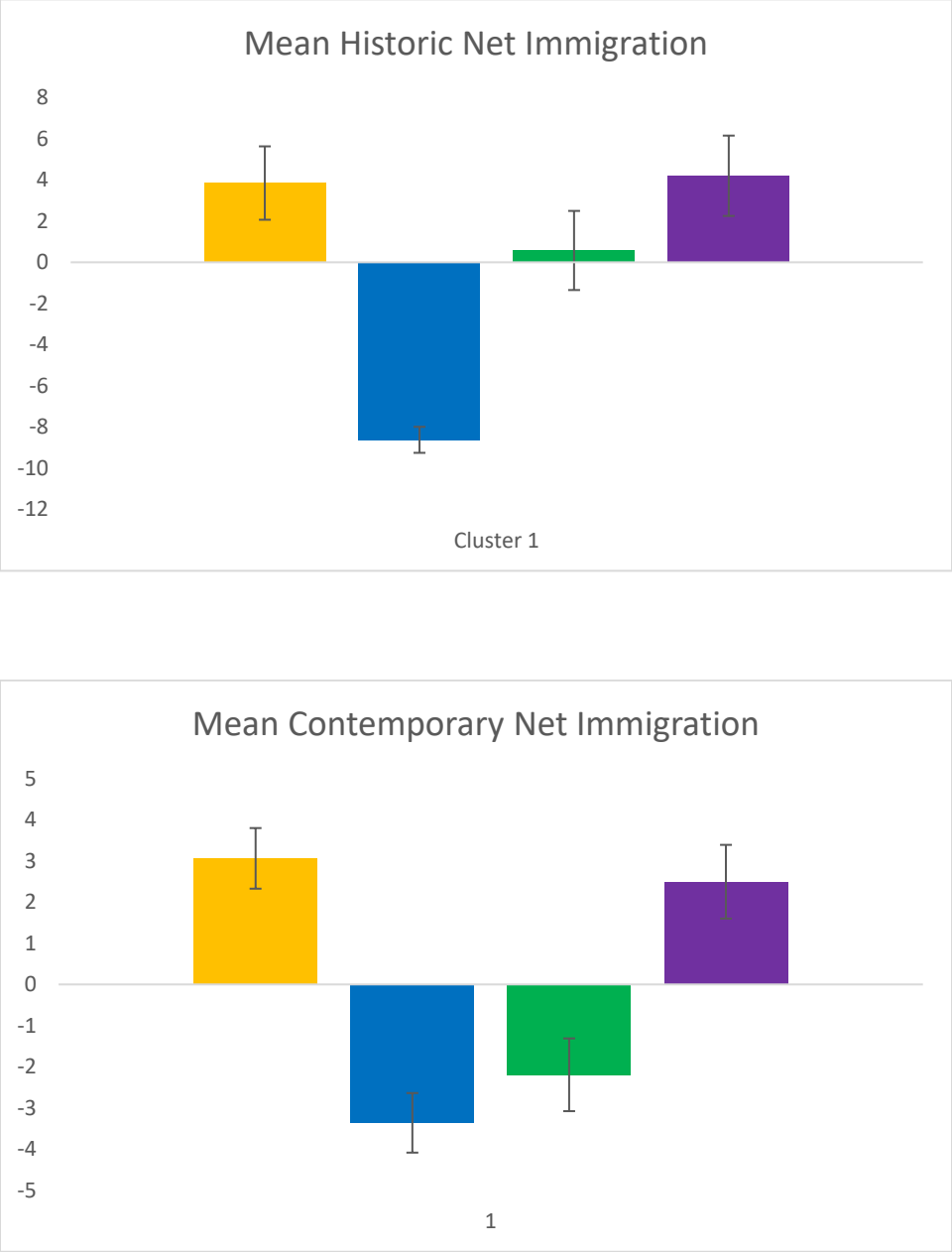


Figure 1.13 Mean historic and contemporary net immigration. Contemporary uses individuals per generation, while historic uses mutation-scaled migration rate.

Table 1.7 Summary of Bottleneck tests and effective population size estimates with 95% confidence intervals.

Site	**BOTTLENECK**			**Effective Population Size**			
	Wilcoxon test	***Sign test***	***Mode shift***	***N*e 0.05**	**95% CI**	***N*e 0.02**	**95% CI**
IN-1	*P* = 0.892	*P* = 0.049	none	53.3	15.3 - ∞	97.5	23.8 - ∞
IN-6	*P* = 0.812	*P* = 0.271	none	450.0	33.1 - ∞	∞	69.0 - ∞
IN-7	*P* = 0.729	*P* = 0.473	none	323.9	55.5 - ∞	632.7	83.4 - ∞
MI-5	*P = 0.945*	*P* = 0.069	none	71.6	31.4 - ∞	730.4	77.3 - ∞
MI-6	*P* = 0.996	*P* = 0.002	none	∞	53.5 - ∞	∞	68.4 - ∞
MI-7	*P* = 0.596	*P* = 0.157	none	∞	448.0 - ∞	∞	242.3 - ∞
MI-10	*P* = 0.607	*P* = 0.452	none	∞	44.3 - ∞	∞	70.9 - ∞
MI-15	*P* = 0.793	*P* = 0.224	none	1770.8	27.4 - ∞	1770.8	27.4 - ∞
OH-1	*P* = 0.996	*P* = 0.034	none	7.9	4.4 – 13.8	8.5	5.6 – 13.3
OH-6	*P* = 0.446	*P* = 0.498	none	∞	132.9 - ∞	350.4	84.6 - ∞
OH-8	*P* = 0.607	*P* = 0.477	none	195.5	68.1 – ∞	262.4	80.3 – ∞
OH-9	*P* = 0.040	*P* = 0.339	none	106.2	25.0 - ∞	∞	41.6 - ∞
OH-13	*P* = 0.473	*P* = 0.521	none	54.2	27.9 – 270.7	202.8	59.0 – ∞
OH-16	*P* = 0.905	*P* = 0.042	none	14.6	11.1 – 19.8	22.1	16.8 – 30.5
OH-17	*P* = 0.729	*P* = 0.259	none	10.9	8.4 – 14.1	14.0	10.9 – 18.4
OH-18	*P* = 0.661	*P* = 0.156	none	42.1	14.5 – ∞	50.1	18.2 - ∞

Table 1.8 Comparison of effective population size estimates (Ne) for one site subsampled to different population sizes with 95% confidence intervals.

Site	**N**	***N*e 0.05**	**95% CI**	***N*e 0.02**	**95% CI**
OH-8	10	772.0	22.7 - ∞	772.0	22.7 - ∞
OH-8	30	195.5	68.1 – ∞	262.4	80.3 - ∞
OH-8	77	137.1	87.5 – 271.8	189.1	121.4 – 380.9

Table 1.9 Comparison of descriptive statistics for *E. blandingii*. * H_O reported rather than H_E.

Species	Location	Pairwise F_{ST}	AR	H_E	Sites	Loci	N	Reference
E. blandingii	Nova Scotia	0.04 – 0.12	-	0.45 – 0.54	3	5	110	Mockford et al. (2005)
	Rangewide	0.00 – 0.47	-	0.45 – 0.71	12	5	200	Mockford et al. (2007)
	NE, IA, MN, IL	0.01 – 0.47	-	0.49 - 0.79	12	8	202	Sethuraman et al. (2014)
	Ontario	0.04 – 0.10	4.8 – 5.3	0.59 – 0.66	12	4	97	Davy et al. (2014)
	NY and southeast Ontario	0.01 – 0.38		0.31 – 0.63	5	7	115	McCluskey et al. (2016)
	WI	0.00 – 0.18		0.59 – 0.70	18	14	389	Reid et al. (2017)
	northeast IL	0.02 – 0.10	3.6 – 3.9	0.51 – 0.64*	6	14	186	Anthonysamy et al. (2018)
	IN, OH, MI	0.00 – 0.15	4.01 – 4.84	0.56 – 0.66	16	14	313	This study

1.5 Discussion

In order to provide further insight into *E. blandingii* conservation, levels of genetic variation and population clustering in localities across Indiana, Michigan, and Ohio were examined. Additionally distance and watershed were correlated with genetic connectivity among populations. Microsatellite analysis of *E. blandingii* from 13 loci were examined across Indiana and the Lake Erie marshes of Ohio and Michigan. Overall genetic differentiation between localities was low across the study region whereas within locality diversity was relatively high (Tables 1.2 & 1.3). Despite low differentiation between localities, population structure was observed (Figures 1.6, 1.8, & 1.9). Pairwise F_{ST} scores as well as AR and H_E observed were very similar to other studies looking at *E. blandingii* (Table 1.9). The relatively high genetic diversity observed across Ohio, Indiana, and Michigan is comparable to range-wide diversity (Tables 1.9). An AMOVA and a Mantel test were also used to assess the influence of geographic distance and watershed boundaries on genetic differentiation (Figures 1.10 & 1.11). Significant isolation by distance and isolation between watersheds at both the HUC6 and HUC8 levels were detected through a Mantel test and an AMOVA respectively (Figure 1.11 & 1.11). Despite the lack of observable bottlenecking, and the relatively high degree of observed genetic diversity, the predicted loss of alleles in the next 200 years by Bottlesim would imply that most localities within this region are likely less secure than then may be assumed (Tables 1.2, 1.6, and 1.9 and Figure 1.4). The inability of BOTTLENECK to detect bottlenecks in a long lived species, and the lack of recent gene flow would also indicate more genetic vulnerability than would otherwise be expected [24,80].

1.5.1 Population Structure

Across localities, little differentiation among populations was found (F_{ST}= 0.05, D=0.08), though there was an observed trend of increasing differentiation by geographic distance (Table 1.3), with the greatest pairwise differentiation occurring between IN01 (western most locality) and OH17 (second most eastern site) (F_{ST} = 0.15, D= 0.21, Table 1.3).

The number of genetic clusters identified through each method ranged from K=4 to K=6 (STRUCTURE, TESS3r, and DAPC). Most comparable studies examining *E. blandingii* structure that utilized STRUCTURE employed the delta k method for determining the number of clusters [27-29]. Only Reid et al., (2017) reported both the MedMeaK and delta K clusters. Puechmaille (2016) argues that the MedMeaK and MedMedK interpretation of determining clusters is likely more accurate for detecting genetic clusters when sampling effort is not equal across localities. Since differentiation across sites was generally low, fewer inferred clusters likely captures the overall trend of population deviation. Higher K values from STRUCTURE and DAPC are representative of hierarchical clustering or sub-structure that fall within the broader clusters identified by TESS3r. OH17 was consistently identified as its own cluster by STRUCTURE (even using K=4 with increased assignment threshold) and DAPC, suggesting that this grouping represents some ecological significance.

Some populations in New York and Illinois showed greater differentiation and additional population structure than would be expected based on the geographic distance (or lack thereof) between them, which is thought to be attributed to glacial retreat and expansion as well as impacts of the watershed on the dispersal of the turtles [7,25,28]. Due to the similar geographic and glacial history, I expected to find similar degrees of differentiation and structure across Indiana, Michigan and Ohio. Although the differentiation based on F_{ST} and D values did not necessarily support this

hypothesis, the *K* values of the STRUCTURE run, TESS3r and DAPC show evidence of this degree of population differentiation. Francois and Durand (2010) have demonstrated that Bayesian approaches that incorporate geographic priors (TESS3r and STRUCTURE LOCPRIOR) are more advantageous at detecting accurate population structure. Additionally, Francois and Durand (2010) demonstrated that TESS3r does a better job at capturing the geographic driven clustering than STRUCTURE. TESS3r and STRUCTURE with LOCPRIOR have been shown to effectively predict population structure in organisms that have experience high degrees of glacial driven expansion and contraction [65]. Since Tess3r produced the most conservative estimates (with the exception of increasing assignment threshold) and seems to be the most robust method for estimates of population clustering I suggest that the four identified clusters from TESS3r be considered management units for the purpose of conservation though it is worth noting that differentiation was generally low. It should also be noted that northwest Michigan was heavily under sampled with only one locality being included and Indiana may share population structure with un-sampled localities from Illinois. That being said, the identified management units should apply to the Lake Erie Marshes of Ohio and southeast Michigan.

All observed *K* values, regardless of the method employed, provide support to prior findings that *E. blandingii* has higher levels of between locality differentiation east of the Appalachians compared to west of the Appalachians, (F_{ST}, *D*) but are not in panmixia across the Midwest and Great Lake Regions (Table 1.9) [25-30].

Similar to other studies, isolation by distance and watershed were observed [26,28]. Large historical migration driven by glacial dynamics are thought to have increased genetic admixture across the Midwest and Great Lakes region which may also dampen detectable distance-driven isolation [28,38]. Similar to Sethuraman et al. (2014) the majority of variance was observed within

localities, but a significant signature of watershed was detected by AMOVA analysis. The watershed signature may be partially explained by effect of distance especially within large watershed basins.

Utilizing the clusters identified by TESS3r, Migrate helped to further explain likely historic migration patterns. Average historic migration rates tended to show a general trend of movement out of cluster 2 and into the other three clusters (Table 1.6 and Figure 1.13). The movement out of cluster 2 could in part be explained by the settlement pattern and subsequent draining of wetland in northwest Ohio and southeast Michigan [85]. As the Great Black Swamp (which stretched across northwest Ohio into northeastern Indiana) was divided by roads and railroads, and eventually drained in the mid to late 1800's it opened up northeast Ohio, southeast Michigan, and northeast Indiana to increased settlement and land modification [85]. It is possible that as this region opened up and was converted to productive agricultural land that the *E. blandingii* in that region (Cluster 2) would be driven to remnant wetlands to the east and west which would explain the observed pattern of movement out of cluster 2 [85]. Further investigation into migration patterns and population cluster at a range-wide scale would help to further explore the validity of this trend.

Mean recent migration estimates from BayesAss showed the same general trend in migration rates between clusters as was seen in historic migration patterns, although recent migration crossed zero at the 95% confidence interval indicating little to no actual immigration. Although the observed genetic differentiation between localities does not support a finding of recent loss of gene flow, the long generation time of *E. blandingii* will likely make it difficult to observe the reduction of gene flow for some time (1 to 2 generations or longer Figure 1.4). The reduction to little or no migration in the past few generations (150 to 250 years) compared to historic rates may implicate

the well documented effects of urbanization and habitat fragmentation as the culprit [10,15,18]. However differences in migration estimation methods between the coalescent-based Migrate and the disequilibrium-based BayesAss makes it difficult to interpret the validity of such a comparison [86].

The largest cluster, cluster 2, was identified as a source by both the historic and recent migration rates, and the second largest cluster, cluster 3 was identified as a source population by the recent migration rates (Figure 1.14). Large reductions of average migration rates from clusters 2 and 3 to cluster 1 and 4 could potentially set up conditions for bottlenecking within those clusters. Since large portions of Michigan were not sampled and the clustering scheme covered large geographical areas, the mean migration estimate should be interpreted with caution.

1.5.2 Genetic Status Within Populations

The lack of observed bottlenecking does not necessarily mean bottlenecking has not occurred (or is not occurring) since BOTTLENECK poorly detects population bottlenecks in long lived species [24,80]. Studies on the similarly long-lived ornate box turtle (*Terrapene ornata*) [24] and spotted turtle (*Clemmys guttata*) [80] had similar difficulty detecting population bottlenecks using traditional methods. Kuo and Janzen (2004) also found that once bottlenecks occur they may result in rapid genetic decline.

Effective population sizes for *E. blandingii* were as high as ∞ for four populations, but these estimates are likely biased by the sample sizes of the localities (Tables 1.7 and 1.8). Most populations of *E. blandingii* are thought to be small and disjunct which contradicts our estimates [2]. Citizen science based population estimates for this same locality found an estimated female population size of 87 individual *E. blandingii* or a population size of 174 adult individuals

assuming a 1:1 sex ratio [87]. Comparison of different samples size effects on N_e estimates indicates that localities with less than 30 individuals likely do not allow for accurate estimations (Table 1.8). As our sample size increase our estimates and confidence interval became closer and closer to the population estimates of Cross et al., (2021). Additional replicates with other large populations would help determine the sample size necessary for more accurate estimates of effective population size.

The projection of genetic variation into the future using Bottlesim suggests that large reductions in population size can result in drastic loss of genetic diversity in isolated populations of *E. blandingii* (Figure 1.4). It should be noted that this model does not account for differential reproductive success or survival based on the age of the individuals, environmental stochasticity, or low recruitment, features seen in *E. blandingii* [2,4,81,88]. Excluding life stage dynamics in Bottlesim models likely leads to under-estimation of the genetic impacts resulting from population reductions in the absence of gene flow. Population Viability Analysis (PVA) in Illinois by King et al. (2021) included environmental stochasticity and showed that inclusion of environmental catastrophes increased the population size threshold to maintain genetic diversity (50 to 110 without, and 110 to 200 with) and reduce extinction risk (20 to 50 without, and 50 to 200 with). Furthermore, the Bottlesim results indicate that populations around 200 individuals will show a steady loss of genetic diversity over 200 years even with a constant population size. Since most localities in our study appear to be small and isolated, and likely have *E. blandingii* estimates far below 200 individuals, it is possible they have begun to lose a great degree of genetic diversity which is not yet observable due to long lifespan [24].

The PVA models used by King et al. (2021) corroborate the loss of genetic diversity and a high risk of extinction in scenarios where starting population size was small (less than 50 adults).

In that study, fifty individuals was used as a cut off for the minimum number of adult breeding *E. blandingii* necessary (assuming a 1:1 sex ratio) to reduce extinction risk and maintain a high degree of genetic diversity in the presence of environmental stochasticity. N_e tends to be much lower than the population census size for wild populations (~10% across taxonomic groups) since N_e accounts for sex ratio, number of adults, genetic makeup and population size [90]. However our best estimate of N_e (137.1 for OH08) was very close to the adult census size (174 assuming a 1:1 sex ratio) from that same locality as estimated by Cross et al. (2021). Assuming a similar relationship between adult census size and N_e as was observed for OH08 only four localities of the 16 included in this study fall below this threshold for the number of adults to prevent extinction set by King et al. (2021) for estimated N_e and two were at this or near this threshold (Table 1.7). Using the most conservative estimates of N_e (the lower 95% confidence interval) 11 of the 16 sites included fall below the 50 individual cut off (Table 1.7). Additionally, the best *Ne* estimates for our largest locality within the study area had an effective population size below 200 individuals indicating most actual effective population sizes are likely far below this. It is also worth noting that based on our estimates from OH08, small sample size seems to inflate estimates of N_e and only five of the sample localities included had enough individuals to provide reasonable estimates (Tables 1.2, 1.7, and 1.8).

1.5.3 Management Implications

Although within population diversity appears to be high, this does not mean populations are ecologically secure. Based on the Bottlesim result, many localities within this region have likely seen substantial population declines and are vulnerable to lose up to 50% of the observed genetic diversity in the next 200 years (Figure 1.4). Effective population size estimates are heavily

influenced by sample size tending to overestimate when less than thirty samples are included, for this reason N_e estimates should be viewed conservatively focusing on the lower end of the 95% confidence interval (pcrit <0.05) for populations that had less than thirty samples (Tables 1.7 & 1.8).

King et al., (2021) estimated adult population sizes necessary to minimize extinction probability and maintain 95% of the genetic diversity in *E. blandingii* using genetically informed population viability analysis. Based on the estimates of King et al., (2021) an adult population size of at least 20 individuals is required to minimize (<5%) extinction risk and at least 50 individuals are needed to maintain 95% of the genetic diversity of the next 100 years in the absence of stochastic events. In the presence of stochastic events at least 50 individuals are required to minimize extinction risk and 110 are needed to maintain 95% of the genetic diversity [89]. Assuming that most localities within our study have little to no gene flow due to fragmentation and urbanization and maintain a similar relationship between N_e and adult census size as was observed in OH08 (Table 1,8 and [87]) only 9 of the 16 localities included produced conservative N_e estimates > 20 threshold and only 5 meet the > 50 individual threshold (Table 1.7). Only two localities have conservative N_e estimates above the 110 threshold indicating that in the presence of stochastic events that could increase mortality, only 2 of the 16 localities produced conservative estimates minimizing loss of genetic diversity and only 5 are large enough to minimize extinction risk (Table 1.7). Population and habitat suitability modeling for *E. blandingii* under different climate change scenarios indicates that the amount of available suitable habitat is likely to shift north with a decline in habitat connectivity, which argues for targeting a threshold of adult individuals > 110 for each locality to prevent genetic loss and extirpation [91,92]. Since 110

individuals per locality may not be a realistic goal, a minimum of 50 individuals per locality could help to reduce extinction risk, particularly if gene flow is maximized through translocation [89].

Additional population viability analysis conducted by Ross et al., (2020) indicated that head-starting individuals could help to stave off extinction and prevent local extirpations. Head-starting a minimum of 50 eggs per year (under an assumed 0 to 1 age mortality rate of 92.8% [94]) can help to prevent extinction, whereas 100 eggs per year can increase population growth [93]. Head-starting can also help to increase juvenile survival and lead to increased recruitment [95,96]. Carstairs et al., (2019) found that although head started individuals had an acclimatization period of 1 to 2 years that resulted in reduced growth, movement, and survival they had nearly three times the probability of surviving to age 10 compared to wild hatched turtles. Thompson et al. (2020), found that in the long term head-starting can increase the body size distribution of individuals within a population, increasing the number of juveniles and ultimately leading to recruitment of reproductive adults. Although head-starting would need to be maintained until populations could naturally reach sustainable recruitment rates of at least 50 individuals a year, it provides a viable strategy for minimizing extinction risk, especially in small, isolated localities [93].

The observed genetic clusters provide a basis for population management units. These management units should be used as a guideline for determining which localities should be used for translocations or head-starting at other localities. When possible, limiting translocations to localities that share a cluster under the K=7 scheme should be maintained, however since differentiation is generally low using the K=4 cluster scheme allows for greater leeway and should be maintained when there are not an ample number of individuals from localities sharing a K=7 cluster. Additionally, the presence of IBD and the isolation between watersheds implies that when possible individuals should be translocated from nearby localities within the same HUC6 or HUC8

(when staying within HUC8 is not feasible) watershed. Aiming to maintaining connectivity between localities within the life time home range size and maximum movement range of *E. blandingii* could also act to maintain genetic diversity and increase effective population size by increasing the number of breeding individuals able to freely breed [41]. Since roads have been found to be a significant source of mortality in *E. blandingii*, causing as many as 61 mortalities over 4 years at a single locality in Ontario, it would make sense to target high traffic roads within a 2 kilometer distance (based on length of breeding and nesting movements) from known *E. blandingii* localities for reducing mortality and allowing gene flow [18,41,97]. Reducing the effect of roads and urbanization to promote gene flow could be accomplished in a number of ways, from creating buffers of natural landscapes around localities, utilizing artificial crossing structures such as well-maintained exclusion fences and culverts interspersed based on home range size, or providing supplemental artificial nesting habitat to provide an alternative to crossing roads [41,98].

1.6 Conclusions

Observed levels of within locality genetic diversity and the lack of bottlenecking imply a level of genetic security within the Great Lakes Region. However, the observed diversity may be representative of remnant population structure and historic gene flow masking ongoing or developing bottlenecks and a potential forthcoming decline in genetic diversity [24].

Despite the lack of differentiation between localities within each state, as well as between localities across the study area, a minimum of four genetic clusters were identified. Although these clusters do not necessarily represent localities connected by current gene flow, they provide guidance for genetically significant management units. As many as seven genetic clusters were observed depending on method, but a hierarchical clustering scheme was observed across methods

(except for OH17). Increasing the number of clusters followed the same general assignment scheme when using fewer cluster and provides insight on differentiation within those larger groupings. These higher levels of clustering help to inform within unit diversification, but in light of the low levels of differentiation observed across the study (Pairwise F_{ST} between 0 and 0.15) are likely less genetically relevant. A Mantel test and AMOVA's indicate that part of the detected genetic clustering is explained by a combination of watershed boundaries and distance between localities.

Although average migration rates lack resolution without increased sampling across the region (particularly northwest Michigan) the detected reduction in mean historic vs mean recent migration rate between genetic clusters would imply urbanization has had an effect on gene flow between identified clusters.

My findings support the currently reported genetic trends observed in *E. blandingii* across their range; although genetic differentiation is low between localities, there is genetic structure [25-30]. Further sampling and analysis and a range wide scale using microsatellite markers would help to further understand the genetic trends and historic drivers of population structure within *E. blandingii*.

CHAPTER 2. LANDCSAPE RESISTANCE MODELING

2.1 Introduction

In order to further understand the complex patterns that have driven the observed genetic composition and population structure, it is important to understand how a species moves across the landscape in reference to different geographic and geomorphic features. The practice of correlating genetic differentiation with traditional landscape ecology to look for genetic discontinuities caused by environmental features has come to be known as landscape genetics [99]. Landscape genetics takes advantage of advances in molecular genetics, statistical modeling, and computing to help further the exploration of the effects of landscape features and geographic barriers on functional gene flow and population dynamics that drive genetic clustering [99,100]. In particular, landscape genetics has provided a new approach to understand the effects of ecological succession and urban modification on both flora and fauna gene flow and clustering [101,102]. Landscape genetic approaches also prove effective in identifying specific landscape features that can act as barriers for gene flow [103]. More complex modeling utilizing high quality landscape data and fine scale spatial autocorrelation allows for the identification of movement corridors, as well as optimization of features that drive landscape resistance that can be utilized for targeted habitat conservation [104-108].

There is a large body of work focusing on the spatial ecology and landscape use of *E. blandingii* [6,8,11,14,16,109,110]. *E. blandingii* are known to use a variety of wetland habitats from season to season including open marshes, ponds, lakes, ephemeral wetlands, creeks, streams, and ditches [6,8,11,14,16,111]. Habitat use also tends to vary slightly among regions, complicating precise delineation of suitable habitat [2]. In general, *E. blandingii* tend to utilize shallow vegetated

wetlands, and at a macrohabitat level show selection of wetlands over other water bodies or uplands [2,8,16]. Despite selection of wetland habitat, Joyal et al. (2001) noted that *E. blandingii* will use upland habitat extensively for nesting, dormancy, and dispersal movements. Focusing on habitat selection in a mostly pristine location in Ontario, Canada, Edge et al. (2010) found no selection in microhabitat use indicating that *E. blandingii* was likely not overly selective of habitat when there is an abundance of open resources and low levels of landscape modification. Conversely, urbanization and habitat modification has well documented impacts on *E. blandingii* persistence and habitat availability [10,15,18]. For example, in two wetlands located within agriculture-dominated watersheds that are heavily invaded by European common reed (*Phragmites australis*), Markle and Chow Fraser (2018) found selection for aquatic and mixed organic marshes and locally avoidance of patches of European common reed.

Beyond movement-based landscape analysis, prior studies utilizing microsatellite markers in *E. blandingii* have inferred a mixture of historic landscapes and life history affecting gene flow and genetic clustering. Since *E. blandingii* has long lifespans (up to 83 years in the wild) accompanied with long generation times and increasing reproductive output with age, it is possible that genetic variation will not have had time to respond to contemporary landscape change which isolates populations [22,81,113]. Davy et al. (2014) indicated low levels of historic migration between clusters in Ontario indicating a pre-settlement driver of genetic units across the landscape other than the modern urban disconnect. Anthonysamy et al. (2018) found no significant relationship between *E. blandingii* population structure and geographic distance indicating that landscape and or habitat features likely drive gene flow. Conversely Reid et al. (2017) saw a significant relationship between roads and population differentiation in *E. blandingii* in Wisconsin using a transect-based landscape metric.

In order to account for the lag in genetic changes associated with the life history of *E. blandingii,* and to determine drivers of gene flow and differentiation within the Great Lakes region, it is useful to characterize landscapes using both historic and contemporary features. The comparison of historic and contemporary population dynamics has the potential to help determine where gene flow has been restricted and where populations may experience decreases in genetic diversity or bottlenecks in the coming years. Furthermore, such analyses provide opportunities for identifying corridors of habitat that connected populations in the past to identify areas that would benefit most from habitat restoration.

Publicly available GIS data can be used to examine elevation and pre-settlement land class features which can be used to infer the conditions that would have been encountered by *E. blandingii* as they moved across the landscape prior to contemporary landscape modification [114]. Although a transect-based landscape metric is simple to do and can quickly make inferences about movement across a landscape, it lacks the complexity of the actual landscape the turtles experience traversing the habitat. Peterman (2018) instead implemented least cost paths of random walk method using CIRCUITSCAPE to rank landscape resistance faced by individuals scored by using a genetic algorithm to optimize the resistance surface [108,115]. This approach removes the need for *a priori* inference or expert opinion which could potentially bias landscape models of gene flow [108].

Here, contemporary and historic landscape features within three Lake Erie watershed marshes in Michigan are analyzed to: (1) optimize landscape resistance surfaces to better understand the influence of landscape feature on movement, (2) determine the potential driving effects of landscape resistance on gene flow, and (3) compare the influence of contemporary

landscapes versus historic landscapes that are likely to have shaped standing genetic variation among populations of *E. blandingii.*

2.2 Methods

2.2.1 Site Description

Three localities within Livingston County in Michigan (MI6, MI7, and MI17) were chosen due to their close relative distance from each other as well as available sample size, and landscape heterogeneity (each about 10 kilometers from each other). All three of these localities are dedicate state recreation areas making their use primarily for outdoor recreation. MI06 encompasses an area of 11,000 acres, MI07 an area of 4,947 acres, and MI17 an area of 20,500 acres. Currently all three localities are dominated by deciduous forest, and woody wetlands with several kettle lakes and other bodies of open water interspersed. MI06 and MI17 as well as MI07 and MI17 are separated by mix of agricultural land and low intensity developed urban spaces. MI06 and MI07 are separated by higher density and more intensely developed urban area in addition to sparse agricultural land. The pre-settlement landscape of this area was primarily deciduous forest, scrub shrubs, and wood wetlands with interspersed grasslands and open water kettle lakes.

2.2.2 Creating Resistance Surfaces

Initial runs utilized the full dataset presented in Chapter 1, however this proved to be computationally prohibitive and likely minimally informative since isolation by distance was found across populations. To focus in on landscape influence and reduce the impacts of distance,

the three closely clustered localities in Michigan described above had a reasonable sample size (n = 59) and provided relatively high landscape heterozygosity compared to the rest of the study.

All resistance surfaces were generated at or resampled to a 30 meter by 30 meter cells to allow for timely computation. Resistance surfaces were created in ArcMap version 10.7.1 using publicly available data [116]. The 2016 national land cover database map (NLCD) was used for contemporary landscape [117]. A historic land class map was developed from pre-settlement vegetation maps from Michigan Natural Features Inventory (MNFI) [118,119]. The pre-settlement vegetation map was re-categorized to follow the same land class features as the NLCD map. The re-classified historic land class shapefile was then converted into a 30 meter by 30 meter raster file. A 1 meter by 1 meter digital elevation model (DEM) was accessed through the U.S. Geological Survey (USGS) [120] an used to develop a Topographic Position Index (TPI) which is a measure of relative elevation and better represents fine scale elevational changes that an organism would experience moving across the landscape. Landscape surfaces were selected based on prior literature reporting on *E. blandingii* habitat use, movement, and fragmentation [2,11,14]. Pre-settlement maps were reclassified to group similar habitats into the same land class categories used by the national land cover database map to make the historic land class layer comparable to the modern land class map.

Since *E. blandingii* is believed to utilize ephemeral wetlands, as well as ditches to aid in seasonal movement, a TPI was generated as a proxy layer for examining wetland and elevation resistance due to the species' ability to detect small ephemeral wetlands [121]. TPI calculates the average elevation value of each cell within a raster based on the surrounding cells within a given radius providing a finer scale picture of valleys (negative values), and ridges [122]. A TPI was generated from digital elevation models (DEM) using the model builder in ArcGIS. Each surface

layer was then exported from ArcGIS using the export to CIRCUITSCAPE plug-in to ensure that each layer was the proper dimensions and in the correct raster format to be processed in R [123] (Figure 2.1).

2.2.3 Genetic Distance

Individual based genetic distances can be calculated a number of different ways, however Principal Component Analysis (PCA) based distance matrices appear to be highly robust for inferring landscape resistance relationships [124,125]. PCA based distance metrics allow the model to focus on the loci that are driving the majority of the variance between individuals [125]. A PCA genetic distance matrix was developed following [124]. A matrix was developed comparing allele usage at each locus for every individual and then used to develop a PCA using the R package Adegenet Version 2.1.4 [65]. The Euclidean distances between the 90 principal components were then used to develop pairwise genetic distance between individuals occurring within the three different sites [124-126].

2.2.4 Landscape Resistance Optimization

Genetic resistance modeling was run in Resistance GA (an R package utilizing Julia and CIRCUITSCAPE; [108]) using the previously described landscape surfaces. Resistance GA utilizes a genetic algorithm developed by Scrucca (2013) to optimize landscape surface using iterative data transformation to test all possible resistance values of the landscapes [108]. Resistance GA uses pairwise genetic data and landscape distances determined by CIRCUITSCAPE to test the fit of each iterative data transformation using a mixed effect linear model [108]. In each iteration, the best model is retained and included in the proceeding iteration until the model no longer improves

[108]. Resistance GA tests eight data transformations for each continuous landscape layer and explores each possible resistance value for each categorical layer [108]. Once optimized, each resistance layer is assigned an AICc score and compared to null and geographic distance models to determine which model is the best fit for the true landscape resistance. Resistance GA also generates resistance layers for each landscape surface allowing one to visualize the optimized landscape resistance layer. Each surface was run independently utilizing individual based PCA axis for genetic distance.

Since individuals were sampled using traps during the mating season when individual *E. blandingii* are known to make large terrestrial movements we wanted to include all individuals regardless of proximity [6,41]. To control for the potential effects of shared locality due to trapping, and to maximize the exploration of the landscape, comparisons between individuals within the same locality were excluded. A total of 59 individuals from three localities were included (MI06=16, MI07=36, MI17= 7).

2.3 Results

2.3.1 Genetic Distance

A total of 90 PCA axis were generated for allele usage by individual. Euclidean distance between all 90 PC axes for each individual ranged from 3.63 to 8.23 with a mean of 6.04. The overall variance for the Euclidean distance between the 90 PC axes was 0.31.

2.3.2 Landscape Resistance Optimization

Optimized resistance surfaces were generated for each provided surface layer assigning values of the relative resistance an *E. blandingii* would face moving through each feature (Figure

2.3). The resistance values increase exponentially as the relative elevation increased, and neared an asymptote at the highest relative elevation observed on the landscape (Figure 2.2). For the TPI layer lower elevation was associated with low relative resistance, and higher elevation was associated with higher relative resistance (Figure 2.3). For the contemporary landscape map (LAN) woody wetlands, open water, and deciduous forests had the lowest relative resistance values were as, developed urban lands had the highest followed by agricultural lands (Figure 2.3). For the pre-settlement land class map (PRE) emergent and woody wetlands had the lowest relative resistance values followed by deciduous forests while open water had the highest relative resistance followed by scrub/shrub (Figure 2.3).

2.3.3 Model Fit

The distance model was by far the best explanation for the genetic differentiation observed across the landscape at these localities (Table 2.1). The TPI model was the second best model but was only as good as the null (AICc values of 1278.139 and 1280.037, respectively) (Table 2.1). Both the TPI model and the distance model had relatively weak fits to the data (Rm2= 0.014 and 0.015 respectively) (Table 2.1). The contemporary landscape had the highest regression value, (though still rather weak) but was also the worst model based on AICc (Rm2= 0.027) (Table 2.1). The optimized resistance surface for TPI utilized a monomolecular transformation in which the lowest relative elevation values produced the least landscape resistance (Figures 2.3 & 2.4).

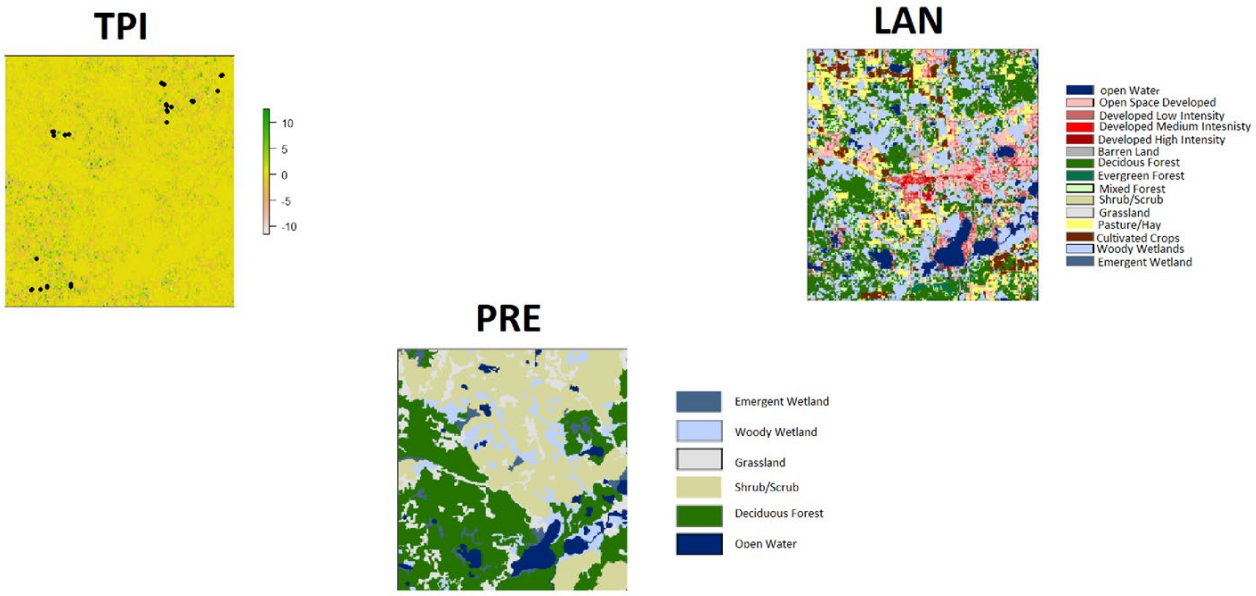


Figure 2.1 Resistance surfaces TPI (Topographic Position Index), LAN (National Land Class Map), PRE (Pre-Settlement Land Classes). TPI layer to show position of individuals. Locality information removed.

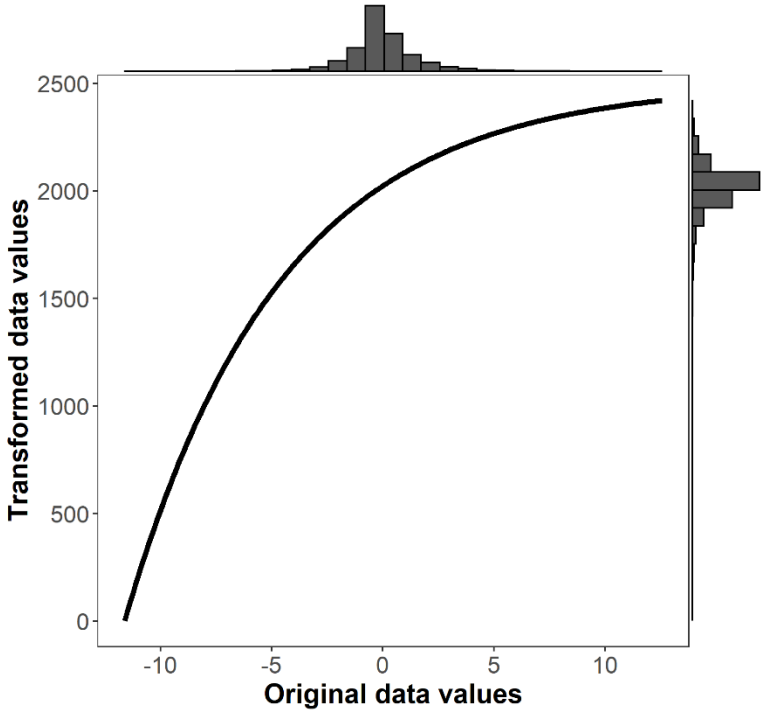


Figure 2.2 Monomolecular Transformation applied to TPI.

Table 2.1 Summary of resistance model optimization.

Surface	obj.func_LL	k	AIC	AICc	R^2m	R^2c	LL
Distance	-635.734	2	1275.468	1275.678	0.014486	0.257794	-635.734
Null	-638.035	1	1278.07	1278.139	0	0.247083	-638.035
TPI	-635.655	4	1279.31	1280.037	0.015069	0.25782	-635.655
PRE	-635.540	7	1285.081	1287.234	0.019872	0.252395	-635.54
LAN	-633.175	16	1298.351	1311.002	0.027415	0.250808	-633.175

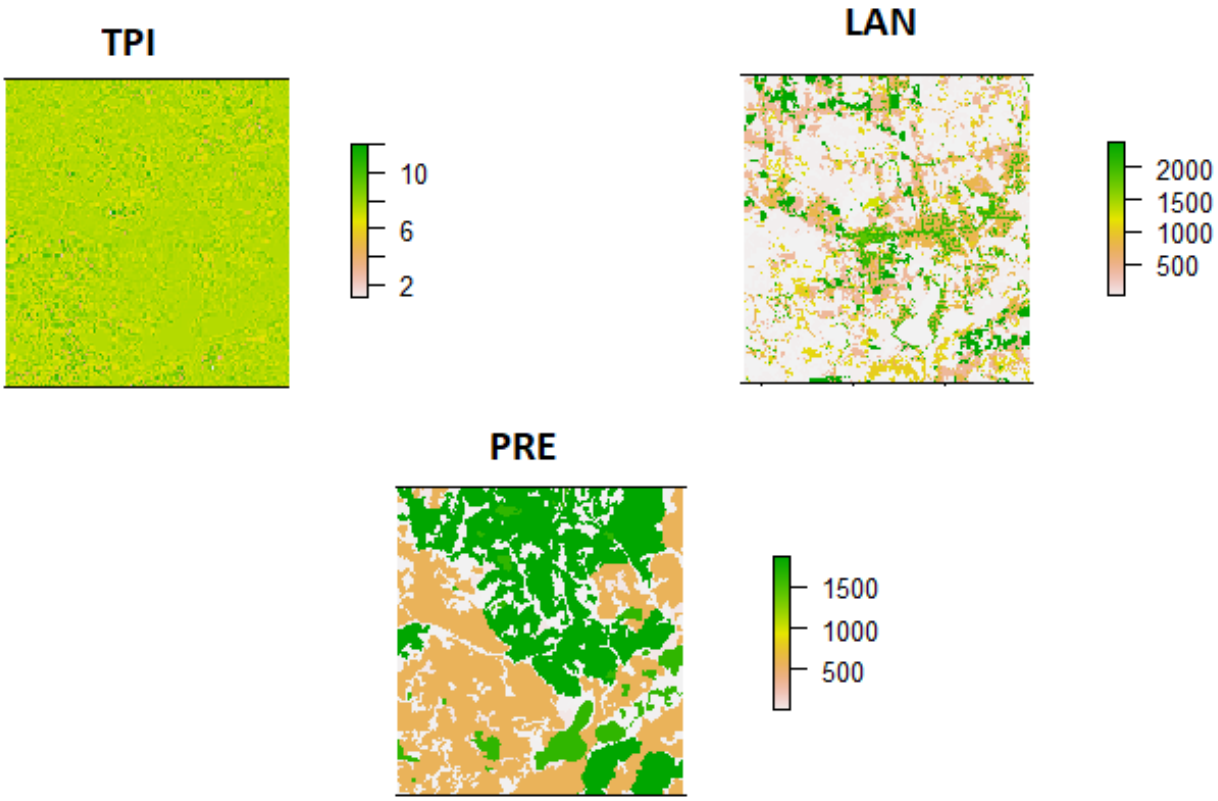


Figure 2.3 Optimized resistance surfaces, TPI (Topographic Position Index), LAN (National Land Class Map), PRE (Pre-Settlement Land Classes). TPI layer to show position of individuals. Locality information removed.

2.4 Discussion

The distance model as the best model (by AICc) with weak regression reinforced the observed weak IBD detected in chapter one (Figure 1.10). The optimized TPI model indicated that low elevation exerts the least landscape resistance on *E. blandingii* (Figures 2.2 &2.3). However, the weak relationship (R^2m 0.022) between relative elevation and genetic distance implies that the null model was just as good, indicating that TPI is not a strong driver of differentiation for *E. blandingii* and explains little to no differentiation.

The weak overall relationship between TPI resistance and genetic distance between localities is not surprising based on the previous research. Joyal et al., (2001) found extensive use of uplands for movement, nesting, and dormancy. *E. blandingii* are known to move extensively, use a wide variety of habitats, and have large variable home ranges [6,8,11,14,16]. Grgurovic and Sievert (2005) found an estimated home range size of 22 hectares in *E. blandingii* and saw little overlap in home range use from year to year indicating that *E. blandingii* likely roams across the landscape from year to year. It has been noted that *E. blandingii* will use uplands extensively for nesting and terrestrial dispersal and will move long distances to breed and nest (up to 2 km) which could aid in gene flow between localities even if individuals spend most of their time in a single wetland [6,41].

Edge et al. (2010) found selection at a macro scale for wetlands over other water bodies and uplands for habitat use (although that does not necessarily implicate them as barriers for gene flow). In the historic landscape layer it is possible that when landscapes have an abundance of available habitat, *E. blandingii* will show little to no microhabitat selection making it difficult to detect a correlation between differentiation and resistance distance [16]. The rather pristine landscape observed by Edge et al., (2010) is more likely be representative of pre-settlement conditions faced by *E. blandingii.*

Lack of correlation between genetic resistance and contemporary land cover parameters could be due to the long generation time of *E. blandingii* [22,81,113]. Despite substantial evidence for the effect of urbanization on *E. blandingii*, from direct mortality to habitat loss and extirpation, prior genetic assessments have observed a lack of differentiation between localities separated by urban development [6,8,10-17,27].

Although attempts were made to account for the effects of the traps on the model, it is still possible that luring individuals to one location may influence the observed relationship between genetic distance and landscape resistance distance. Increasing the extent of the model to include more sites and explore a greater diversity of landscapes could potentially allow for a more informative model, however doing so would also increase the effect of geographic distance on differentiation and require a reduction in resolution of the landscape characteristics because the raster data is large and memory intensive. Additionally, a larger extent is computationally and memory intensive, especially if attempting to utilize individual based genetic distance. Using site based genetic distance could reduce the number of parameters and reduce the time of computations but would also reduce sample size and constrain representative landscapes characteristics resulting in model bias.

2.5 Conclusion

Overall optimization of resistance surfaces at a localized region in Michigan found little to no influence of landscape features on the observed pattern of genetic differentiation in *E. blandingii* (Table 2.1). Consistent with chapter one, IBD was found to be the best explanation for the observed genetic differentiation between localities in Michigan. The degree of genetic differentiation and the observed pattern across the landscape does not appear to show any meaningful relationship with the historic or contemporary landscape, at least at this geographic extent (Table 2.1). The lack of correlation between resistance distances for land class features and genetic differentiation could be due to low levels of observed genetic differentiation across the region which is likely due to long generation time, high historic migration, and habitat flexibility in *E. blandingii* [22,81,113]. This would also explain why no significant pattern was observed in the

contemporary landscape despite the substantial evidence of the effect of urbanization on *E. blandingii* [6,8,10-17]. Lack of correlation in the pre-settlement model could be due to low levels of landscape resistance to gene flow in *E. blandingii* in unmodified landscapes [6,14,33].

Finally, although a strong relationship between genetic differentiation and landscape resistance was not seen in this region using these resistance surfaces, it does not mean there are no drivers of differentiation on the landscape but rather the resistance surface optimization that was implemented here may not be appropriate for examining the relationship between long-lived slowly differentiating organisms. Further investigation into other landscape features, or utilizing different extents or regions, may determine different relationships and shed light on the ability of this framework for assessing resistance surfaces for *E. blandingii*. Landscape resistance may be variable between regions and localities and should be considered when examining for drivers of differentiation.

REFERENCES

1. Lovich, J.E.; Ennen, J.R.; Agha, M.; Gibbons, J.W. Where have all the turtles gone, and why does it matter? *BioScience* **2018**, *68*, 771-781.
2. Congdon, J.; Graham, T.; Herman, T.; Lang, J.; Pappas, M.; Brecke, B. *Emydoidea blandingii* (Holbrook 1838)—Blanding's turtle. *Conservation biology of freshwater turtles and tortoises: a compilation project of the IUCN/SSC Tortoise and Freshwater Turtle Specialist Group. Edited by AG Rhodin, PCH Pritchard, PP van Dijk, RA Saumure, KA Buhlmann, and JB Iverson. Chelonian Research Monographs* **2008**, 015.011-015.012.
3. Alacs, E.A.; Janzen, F.J.; Scribner, K.T. Genetic issues in freshwater turtle and tortoise conservation. *Chelonian Research Monographs* **2007**, *4*, 107.
4. Congdon, J.D.; Dunham, A.; van Loben Sels, R. Implications for conservation and management of long-lived organisms. *Conservation Biology l* **1993**, *7*, 826-833.
5. Kinney, O.M. Movements and habitat use of Blanding's turtles in southeast Michigan: implications for conservation and management. University of Georgia, **1999**.
6. Joyal, L.A.; McCollough, M.; Hunter Jr, M.L. Landscape ecology approaches to wetland species conservation: a case study of two turtle species in southern Maine. *Conservation Biology* **2001**, *15*, 1755-1762.
7. Jordan, M.A.; Mumaw, V.; Millspaw, N.; Mockford, S.W.; Janzen, F.J. Range-wide phylogeography of Blanding's Turtle [*Emys* (= *Emydoidea*) *blandingii*]. *Conservation Genetics* **2019**, *20*, 419-430.

8. Hartwig, T.S. Habitat selection of Blanding's turtle (*Emydoidea blandingii*): A range-wide review and microhabitat study. Bard College--Annandale-on-Hudson, **2004**.

9. Congdon, J.D.; Gibbons, J.W. In *Long-term Studies of Vertebrate Communitie;* Cody, M.L.; Smallwood, J.A., Ed.; *Academic Press, USA* **1996**; Structure and dynamics of a Turtle community, 137-160.

10. Kiviat, E. Blanding's turtle habitat requirements and implications for conservation in Dutchess County, New York. In Proceedings of the Proceedings: Conservation, Restoration, and Management of Tortoises and Turtles-An International Conference, New York Turtle and Tortoise Society, New York, **1997**; pp. 377-382.

11. Ross, D.A.; Anderson, R.K. Habitat use, movements, and nesting of *Emydoidea blandingii* in central Wisconsin. *Journal of Herpetology* **1990**, 6-12.

12. Rowe, J.W.; Moll, E.O. A radiotelemetric study of activity and movements of the Blanding's turtle (*Emydoidea blandingii*) in northeastern Illinois. *Journal of Herpetology* **1991**, 178-185.

13. Pappas, M.J.; Brecke, B.J.; Congdon, J.D. The Blanding's turtles (*Emydoidea blandingii*) of Weaver Dunes, Minnesota. *Chelonian Conservation and Biology* **2000**, *3*, 557-568.

14. Grgurovic, M.; Sievert, P.R. Movement patterns of Blanding's turtles (*Emydoidea blandingii*) in the suburban landscape of eastern Massachusetts. *Urban Ecosystems* **2005**, *8*, 203-213.

15. Congdon, J.; Keinath, D. Blanding's turtle (*Emydoidea blandingii*): a technical conservation assessment. USDA Forest Service, Rocky Mountain Region. **2006**.

16. Edge, C.B.; Steinberg, B.D.; Brooks, R.J.; Litzgus, J.D. Habitat selection by Blanding's turtles (*Emydoidea blandingii*) in a relatively pristine landscape. *Ecoscience* **2010**, *17*, 90-99.

17. Dahl, T.E. *Wetlands losses in the United States, 1780's to 1980's*; US Department of the Interior, Fish and Wildlife Service **1990**.

18. Ashley, E.P.; Robinson, J.T. Road mortality of amphibians, reptiles and other wildlife on the Long Point Causeway, Lake Erie, Ontario. *Canadian Field Naturalist* **1996**, *110*, 403-412.

19. Overmann, S.R.; Krajicek, J.J. Snapping turtles (*Chelydra serpentina*) as biomonitors of lead contamination of the Big River in Missouri's old lead belt. *Environmental Toxicology and Chemistry: An International Journal* **1995**, *14*, 689-695.

20. de Solla, S.R.; Bishop, C.A.; Van Der Kraak, G.; Brooks, R.J. Impact of organochlorine contamination on levels of sex hormones and external morphology of common snapping turtles (*Chelydra serpentina serpentina*) in Ontario, Canada. *Environmental Health Perspectives* **1998**, *106*, 253-260.

21. Congdon, J.D.; Dunham, A.E.; Sels, R.V.L. Demographics of common snapping turtles (*Chelydra serpentina*): implications for conservation and management of long-lived organisms. *American Zoologist* **1994**, *34*, 397-408.

22. Avise, J.C.; Bowen, B.W.; Lamb, T.; Meylan, A.B.; Bermingham, E. Mitochondrial DNA evolution at a turtle's pace: evidence for low genetic variability and reduced microevolutionary rate in the Testudines. *Molecular Biology and Evolution* **1992**, *9*, 457-473.

23. Shaffer, H.B.; Minx, P.; Warren, D.E.; Shedlock, A.M.; Thomson, R.C.; Valenzuela, N.; Abramyan, J.; Amemiya, C.T.; Badenhorst, D.; Biggar, K.K. The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. *Genome Biology* **2013**, *14*, 1-23.

24. Kuo, C.-H.; Janzen, F.J. Genetic effects of a persistent bottleneck on a natural population of ornate box turtles (*Terrapene ornata*). *Conservation Genetics* **2004**, *5*, 425-437.

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

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

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

28. 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. Population genetics of Blanding's turtle (*Emys blandingii*) in the midwestern United States. *Conservation Genetics* **2014**, *15*, 61-73.

29. McCluskey, E.M.; Mockford, S.W.; Sands, K.; Herman, T.B.; Johnson, G.; Gonser, R.A. Population Genetic Structure of Blanding's Turtles (*Emydoidea blandingii*) in New York. *Journal of Herpetology* **2016**, *50*, 70-76.

30. Anthonysamy, W.; Dreslik, M.; Douglas, M.; Thompson, D.; Klut, G.; Kuhns, A.; Mauger, D.; Kirk, D.; Glowacki, G.; Douglas, M. Population genetic evaluations within a co-distributed taxonomic group: a multi-species approach to conservation planning. *Animal Conservation* **2018**, *21*, 137-147.

31. Ellegren, H. Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics* **2004**, *5*, 435-445.

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

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

34. Parmley, D. Turtles from the late Hemphillian (latest Miocene) of Knox County, Nebraska. *Texas Journal of Science* **1992**, *44*, 339-348.

35. Holman, J.; Parmley, D. Noteworthy turtle remains from the Late Miocene (Late Hemphillian) of northeastern Nebraska. *Texas Journal of Science* **2005**, *57*, 307-316.

36. Spinks, P.Q.; Thomson, R.C.; McCartney-Melstad, E.; Shaffer, H.B. Phylogeny and temporal diversification of the New World pond turtles (Emydidae). *Molecular Phylogenetics and Evolution* **2016**, *103*, 85-97.

37. Rödder, D.; Lawing, A.M.; Flecks, M.; Ahmadzadeh, F.; Dambach, J.; Engler, J.O.; Habel, J.C.; Hartmann, T.; Hörnes, D.; Ihlow, F. Evaluating the significance of paleophylogeographic species distribution models in reconstructing Quaternary range-shifts of Nearctic chelonians. *PLoS One* **2013**, *8*, e72855.

38. Smith, P.W. An analysis of post-Wisconsin biogeography of the Prairie Peninsula region based on distributional phenomena among terrestrial vertebrate populations. *Ecology* **1957**, *38*, 205-218.

39. Osentoski, M.F. *Population genetic structure and male reproductive success of a Blanding's turtle (Emydoidea blandingii) population in southeastern Michigan*; Ph.D dissertation, University of Miami: **2001**.

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

41. Congdon, J.; Kinney, O.; Nagle, R. Spatial ecology and core-area protection of Blanding's Turtle (*Emydoidea blandingii*). *Canadian Journal of Zoology* **2011**, *89*, 1098-1106.

42. Starking-Szymanski, M.D.; Yoder-Nowak, T.; Rybarczyk, G.; Dawson, H.A. Movement and habitat use of headstarted Blanding's turtles in Michigan. *The Journal of Wildlife Management* **2018**, *82*, 1516-1527.

43. Willey, L.L.; Jones, M.T. Conservation Plan for the Blanding's Turtle and associated Species of Conservation Need in the Northeastern United States. Northeast Blanding's Working Group.; New Hampshire Fish and Game Department.; U.S. Fish and Wildlife Service **2014**.

44. Hinson, J. Distribution of Populations and Suitable Habitat for Spotted Turtles (*Clemmys guttata*) and Blanding's Turtles (*Emydoidea blandingii*) in Indiana. M.S thesis, Purdue University Fort Wayne **2019**.

45. Pearse, D.E.; Janzen, F.J.; Avise, J.C. Genetic markers substantiate long-term storage and utilization of sperm by female painted turtles. *Heredity* **2001**, *86*, 378-384.

46. Osentoski, M.; Mockford, S.; Wright, J.M.; Snyder, M.; Herman, T.; Hughes, C. Isolation and characterization of microsatellite loci from the Blanding's turtle, *Emydoidea blandingii*. *Molecular Ecology Notes* **2002**, *2*, 147-149.

47. King, T.L.; Julian, S. Conservation of microsatellite DNA flanking sequence across 13 Emydid genera assayed with novel bog turtle (*Glyptemys muhlenbergii*) loci. *Conservation Genetics* **2004**, *5*, 719-725.

48. Libants, S.; Kamarainen, A.; Scribner, K.; Congdon, J. Isolation and cross-species amplification of seven microsatellite loci from Emydoidea blandingii. *Molecular Ecology Notes* **2004**, *4*, 300-302.

49. Reid, B.N.; Peery, M.Z. Land use patterns skew sex ratios, decrease genetic diversity and trump the effects of recent climate change in an endangered turtle. *Diversity and Distributions* **2014**, *20*, 1425-1437.

50. Blacket, M.; Robin, C.; Good, R.; Lee, S.; Miller, A. Universal primers for fluorescent labelling of PCR fragments—an efficient and cost-effective approach to genotyping by fluorescence. *Molecular Ecology Resources* **2012**, *12*, 456-463.

51. Mayo, O. A century of Hardy–Weinberg equilibrium. *Twin Research and Human Genetics* **2008**, *11*, 249-256.

52. Adamack, A.T.; Gruber, B. PopGenReport: simplifying basic population genetic analyses in R. *Methods in Ecology and Evolution* **2014**, *5*, 384-387.

53. Rousset, F. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources* **2008**, *8*, 103-106.

54. Kamvar, Z.N.; Tabima, J.F.; Grünwald, N.J. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* **2014**, *2*, e281.

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

56. Chapuis, M.-P.; Estoup, A. Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* **2007**, *24*, 621-631.

57. Kalinowski, S.T. Counting alleles with rarefaction: private alleles and hierarchical sampling designs. *Conservation Genetics* **2004**, *5*, 539-543.

58. Hale, M.L.; Burg, T.M.; Steeves, T.E. Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population is enough to accurately estimate allele frequencies. *PLoS ONE* **2012, 7,** e45170.

59. Keenan, K.; McGinnity, P.; Cross, T.F.; Crozier, W.W.; Prodöhl, P.A. diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution* **2013**, *4*, 782-788.

60. Wright, S. The genetical structure of populations. *Annals of Eugenics* **1949**, *15*, 323-354.

61. Weir, B.S.; Cockerham, C.C. Estimating F-statistics for the analysis of population structure. *Evolution* **1984**, 1358-1370.

62. Jost, L. D vs. GST: response to Heller and Siegismund (2009) and Ryman and Leimar (2009). *Molecular Ecology* **2009**, *18*, 2088-2091.

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

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

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

66. François, O.; Durand, E. Spatially explicit Bayesian clustering models in population genetics. *Molecular Ecology Resources* **2010**, *10*, 773-784.

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

68. Puechmaille, S.J. 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* **2016**, *16*, 608-627.

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

70. Beerli, P. How to use MIGRATE or why are Markov chain Monte Carlo programs difficult to use. *Population Genetics for Animal Conservation* **2009**, *17*, 42-79.

71. Wilson, G.A.; Rannala, B. Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* **2003**, *163*, 1177-1191.

72. Manoukis, N.C. FORMATOMATIC: a program for converting diploid allelic data between common formats for population genetic analysis. *Molecular Ecology Notes* **2007**, *7*, 592-593.

73. Rambaut, A.; Drummond, A.J.; Xie, D.; Baele, G.; Suchard, M.A. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology* **2018**, *67*, 901.

74. Ishiyama, N.; Sueyoshi, M.; Nakamura, F. To what extent do human-altered landscapes retain population connectivity? Historical changes in gene flow of wetland fish Pungitius pungitius. *Royal Society Open Science* **2015**, *2*, 150033.

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

76. Piry, S.; Luikart, G.; Cornuet, J.M. Computer note. BOTTLENECK: a computer program for detecting recent reductions in the effective size using allele frequency data. *Journal of Heredity* **1999**, *90*, 502-503.

77. Frankel, O.H.; Frankel, O.; Soulé, M.E. *Conservation and Evolution*; CUP Archive: 1981.

78. Cornuet, J.M.; Luikart, G. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **1996**, *144*, 2001-2014.

79. Luikart, G.; Cornuet, J.-M. Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conservation Biology* **1998**, *12*, 228-237.

80. Davy, C.M.; Murphy, R.W. Conservation genetics of the endangered Spotted Turtle (*Clemmys guttata*) illustrate the risks of "bottleneck tests". *Canadian Journal of Zoology* **2014**, *92*, 149-162.

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

82. Cosentino, B.J.; Phillips, C.A.; Schooley, R.L. *Wetland occupancy and landscape connectivity for Blanding's and Western painted turtles in the Green River Valley*; Illinois Natural History Survey: **2008**.

83. Nomura, T. Estimation of effective number of breeders from molecular coancestry of single cohort sample. *Evolutionary Applications* **2008**, *1*, 462-474.

84. Do, C.; Waples, R.S.; Peel, D.; Macbeth, G.; Tillett, B.J.; Ovenden, J.R. NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources* **2014**, *14*, 209-214.

85. Kaatz, M.R. The Black Swamp: a study in historical geography. *Annals of the Association of American Geographers* **1955**, *45*, 1-35.

86. Samarasin, P.; Shuter, B.J.; Wright, S.I.; Rodd, F.H. The problem of estimating recent genetic connectivity in a changing world. *Conservation Biology* **2017**, *31*, 126-135.

87. Cross, M.D.; Mayer, J.; Breymaier, T.; Chiotti, J.A.; Bekker, K. Estimating Population Size of a Threatened Turtle Using Community and Citizen Science. *Chelonian Conservation and Biology* **2021**.

88. Gutzke, W.H.; Packard, G.C. The influence of temperature on eggs and hatchlings of Blanding's Turtles, *Emydoidea blandingii*. *Journal of Herpetology* **1987**, 161-163.

89. King, R.B.; Golba, C.K.; Glowacki, G.A.; Kuhns, A.R. Blanding's turtle demography and population viability. *Journal of Fish and Wildlife Management* **2021**, *12*, 112-138.

90. Frankham, R. Effective population size/adult population size ratios in wildlife: a review. *Genetics Research* **1995**, *66*, 95-107.

91. Hamilton, C.M.; Bateman, B.L.; Gorzo, J.M.; Reid, B.; Thogmartin, W.E.; Peery, M.Z.; Heglund, P.J.; Radeloff, V.C.; Pidgeon, A.M. Slow and steady wins the race? Future climate and land use change leaves the imperiled Blanding's turtle (*Emydoidea blandingii*) behind. *Biological Conservation* **2018**, *222*, 75-85.

92. Byer, N.W.; Reid, B.N.; Thiel, R.P.; Peery, M.Z. Strong Climate Associations but No Temporal Trends in Nesting Phenology of Blanding's Turtles (*Emydoidea blandingii*). *Herpetologica* **2020**, *76*, 396-402.

93. Ross, J.P.; Thompson, D.; Dreslik, M.J. Population Viability Analysis and The Role of Head-Starting for a Northern Illinois Blanding's Turtle Population. In Proceedings of the Midwest Fish and Wildlife Conference 2020, **2020**.

94. District, L.C.F.P.; Glowacki, G.; Kuhns, A.R. Recovery Of The Blanding's Turtle (*Emydoidea Blandingii*) At Spring Bluff Nature Preserve, Lake County Forest Preserves. **2010**.

95. Carstairs, S.; Paterson, J.E.; Jager, K.L.; Gasbarrini, D.; Mui, A.; Davy, C. Population reinforcement accelerates subadult recruitment rates in an endangered freshwater turtle. *Animal Conservation* **2019**, *22*, 589-599.

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

97. Paterson, J.E.; Steinberg, B.D.; Litzgus, J.D. Not just any old pile of dirt: evaluating the use of artificial nesting mounds as conservation tools for freshwater turtles. *Oryx* **2013**, *47*, 607-615.

98. Huijser, M.P.; Gunson, K.E.; Fairbank, E.R. Effectiveness of chain link turtle fence and culverts in reducing turtle mortality and providing connectivity along US Hwy 83, Valentine National Wildlife Refuge, Nebraska, USA. Nebraska Department of Transportation Research Report 4W6072 **2017**.

99. Manel, S.; Schwartz, M.K.; Luikart, G.; Taberlet, P. Landscape genetics: combining landscape ecology and population genetics. *Trends in ecology & evolution* **2003**, *18*, 189-197.

100. Storfer, A.; Murphy, M.; Evans, J.; Goldberg, C.; Robinson, S.; Spear, S.; Dezzani, R.; Delmelle, E.; Vierling, L.; Waits, L. Putting the 'landscape'in landscape genetics. *Heredity* **2007**, *98*, 128-142.

101. Cegelski, C.; Waits, L.; Anderson, N. Assessing population structure and gene flow in Montana wolverines (*Gulo gulo*) using assignment-based approaches. *Molecular Ecology* **2003**, *12*, 2907-2918.

102. Sezen, U.U.; Chazdon, R.L.; Holsinger, K.E. Genetic consequences of tropical second-growth forest regeneration. *Science* **2005**, *307*, 891-891.

103. Epps, C.W.; Palsbøll, P.J.; Wehausen, J.D.; Roderick, G.K.; Ramey, R.R.; McCullough, D.R. Highways block gene flow and cause a rapid decline in genetic diversity of desert bighorn sheep. *Ecology Letters* **2005**, *8*, 1029-1038.

104. Vignieri, S.N. Streams over mountains: influence of riparian connectivity on gene flow in the Pacific jumping mouse (*Zapus trinotatus*). *Molecular Ecology* **2005**, *14*, 1925-1937.

105. Peterman, W.E.; Connette, G.M.; Semlitsch, R.D.; Eggert, L.S. Ecological resistance surfaces predict fine-scale genetic differentiation in a terrestrial woodland salamander. *Molecular Ecology* **2014**, *23*, 2402-2413.

106. Ruiz-Lopez, M.; Barelli, C.; Rovero, F.; Hodges, K.; Roos, C.; Peterman, W.E.; Ting, N. A novel landscape genetic approach demonstrates the effects of human disturbance on the Udzungwa red colobus monkey (*Procolobus gordonorum*). *Heredity* **2016**, *116*, 167-176.

107. Khimoun, A.; Peterman, W.; Eraud, C.; Faivre, B.; Navarro, N.; Garnier, S. Landscape genetic analyses reveal fine-scale effects of forest fragmentation in an insular tropical bird. *Molecular Ecology* **2017**, *26*, 4906-4919.

108. Peterman, W.E. ResistanceGA: An R package for the optimization of resistance surfaces using genetic algorithms. *Methods in Ecology and Evolution* **2018**, *9*, 1638-1647.

109. Fortin, G.; Blouin-Demers, G.; Dubois, Y. Landscape composition weakly affects home range size in Blanding's turtles (*Emydoidea blandingii*). *Ecoscience* **2012**, *19*, 191-197.

110. Vinella-Brusher, E.K.; Haase, J.P. Assessing Urban Habitat Connectivity: Using Circuit Theory to Model Blanding's Turtle Movement. Senior Comprehensive Exercise, Carelton College, **2016**.

111. Spetz, J.C.; Sheil, C.A.; Robison, T.L. *Blanding's Turtle.*; Ohio Biological Survey Bulletin New Series: 2021; Volume 20, p. xiv + 402 p.

112. Markle, C.E.; Chow-Fraser, P. Effects of European common reed on Blanding's turtle spatial ecology. *The Journal of Wildlife Management* **2018**, *82*, 857-864.

113. Erickson, J. Oldest well-documented Blanding’s turtle recaptured at U-M reserve at age 83. *Michigan News.* 2016.

114. McCauley, L.A.; Jenkins, D.G. GIS-based estimates of former and current depressional wetlands in an agricultural landscape. *Ecological Applications* **2005**, *15*, 1199-1208.

115. Anantharaman, R.; Hall, K.; Shah, V.; Edelman, A. Circuitscape in Julia: High performance connectivity modelling to support conservation decisions. *arXiv preprint arXiv:1906.03542* **2019**.

116. ESRI. ArcGIS Desktop: Release 10. **2011**.

117. Dewitz, J. National Land Cover Database (NLCD) 2016 Products. **2019**.

118. Comer, P.J.; Albert, D.A.; Corner, R.; Hart, B.; Kashian, D.; Price, D.; Raab, J.; Schuen, D.; Wells, H. Vegetation of Michigan circa 1800. *Michigan Natural Features Inventory, Lansing* **1998**.

119. (MNFI), M.N.F.I. Michigan Circa 1800 Presettlement Vegetation Cover. **2021**.

120. (USGS), U.S.G.S. 1 meter Digital Elevation Models (DEMs). **2017**.

121. Riley, J.W.; Calhoun, D.L.; Barichivich, W.J.; Walls, S.C. Identifying small depressional wetlands and using a topographic position index to infer hydroperiod regimes for pond-breeding amphibians. *Wetlands* **2017**, *37*, 325-338.

122. Weiss, A. Topographic position and landforms analysis. In Proceedings of the Poster presentation, ESRI user conference, San Diego, CA, 2001.

123. Jenness, J. Export to Circuitscape Tool for ArcGIS 9.x. **2010**.

124. Shirk, A.; Landguth, E.; Cushman, S. A comparison of individual-based genetic distance metrics for landscape genetics. *Molecular Ecology Resources* **2017**, *17*, 1308-1317.

125. Shirk, A.; Wallin, D.; Cushman, S.; Rice, C.; Warheit, K. Inferring landscape effects on gene flow: a new model selection framework. *Molecular Ecology* **2010**, *19*, 3603-3619.

126. Winiarski, K.J.; Peterman, W.E.; McGarigal, K. Evaluation of the R package 'resistancega': A promising approach towards the accurate optimization of landscape resistance surfaces. *Molecular Ecology Resources* **2020**, *20*, 1583-1596.

127. Scrucca, L. GA: a package for genetic algorithms in R. *Journal of Statistical Software* **2013**, *53*, 1-37.