2024

Range-wide Analysis of the Genetic Diversity and Genetic Structure of the Spotted Turtle (Clemmys guttata)

Madison S. Whitehurst
Virginia Commonwealth University

Range-wide Analysis of the Genetic Diversity and Genetic Structure of the Spotted Turtle (*Clemmys guttata*)

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University

Madison S. Whitehurst
B.S. Integrated Science and Technology – James Madison University, 2020

Advised by
Rodney Dyer, Ph.D.
Director of the Center for Environmental Studies
Virginia Commonwealth University

Virginia Commonwealth University
Richmond, VA
May 2024

Table of Contents

Acknowledgments

I want to express my utmost gratitude to my advisor, Dr. Rodney Dyer, and my committee members, Dr. Lesley Bulluck, John Kleopfer, and Dr. Lisabeth Willey. I was able to complete this project thanks to their help and support, especially from Dr. Dyer. This work could not have been initiated without the expertise, fieldwork, and ongoing research from the Eastern Spotted Turtle Working Group and the associated state agencies that collected samples. This thesis was funded by Wildlife International to Dr. R.J. Lastly, I would like to thank my mom and my partner for both their love and support while attending graduate school and working full time.

Abstract

Understanding the spatial distribution of genetic diversity and structure in endangered species is vital for effective conservation management. This research discusses the results of an initial analysis of the genetic structure and variation of the spotted turtle (*Clemmys guttata*), a species of conservation interest. The study sampled 913 individuals from 78 unique locales across 16 states in the eastern portion of the species' range. Utilizing ddRADSeq, a *de novo* genome assembly was developed to identify nearly 20,000 potential loci. A total of 926 were selected for this analysis. Results revealed no spatial clustering of genetic diversity or structure, which is unexpected for turtles. However, a gradual south-to-north reduction in heterozygosity, indicative of a relictual pattern left by post-Pleistocene range expansion, was found. Estimates of genetic differentiation were low, suggesting minimal population structuring. Although no structural clustering was observed, weak isolation by distance was noted, likely due to limitations in dispersal distance relative to the spatial extent of the sampling area. Future work would benefit from augmenting the dataset with samples from the rest of the species distribution and developing a whole-genome reconstruction to aid in population assignment, phylogeographic analysis, and investigation of adaptive genetic variance. These findings provide valuable insights for creating a comprehensive conservation strategy for spotted turtles.

Introduction

The distribution of spatial genetic structure within a species is a complex outcome of demographic, biogeographic, and ecological forces all acting upon its population structure (Walker & Avise, 1998). Landscape composition and configuration could impact a species population distribution and gene flow by isolating or restraining movement between habitat patches (Bouchard et al., 2019). Natural processes and conditions, such as climate change, glacial retreats, and river networks, can moderate the geographical distribution of a species (Bouchard et al., 2019). These forces and the confounding effects of anthropogenic activities can lead to genetic erosion by creating heterogeneous connectivity patterns. Consequently, identifying genetic subdivisions within existing populations is crucial for conservation strategies. Resulting changes in genetic diversity and structure—such as census size, allelic and genotypic diversity, and the amount of standing genetic structure — are of concern for threatened species and are essential factors to consider when developing effective management plans.

Genetic diversity plays a limiting role in how a species or population may react to environmental, ecological, or evolutionary selection. Genetic diversity is broadly defined as the amount of variation and can be measured at the level of the individual, the population, or any larger spatial extent (Balkenhol et al., 2015). It is estimated at the allelic, genotypic, or genomic scale. At an individual locus, the allelic level, the number of variant alleles in the total count, or as a frequency-based metric can provide localized estimates of the magnitude of raw diversity present. At the genetic locus, diversity quantifies how alleles are paired within genotypes. Both allelic and genotypic diversity define the raw material upon which selection pressures may operate and the extent to which the species can respond. At large spatial scales, diversity can also

be measured at sampling locales and populations and even across the entire species distribution, allowing for the determination of spatially relevant hotspots and areas of conservation concern.

When populations experience a reduction in diversity, their ability to respond to changing environmental conditions decreases. This reduction in diversity limits the range of traits available within the population, making it less likely that individuals possess the genetic variations necessary to survive and reproduce in new or changing conditions; the rate of evolutionary change is directly proportional to the amount of diversity present (Fisher, 1930). One consequence of low diversity is inbreeding, leading to an increase in homozygosity and a decrease in heterozygosity. Inbreeding can result in the expression of deleterious phenotypes, decreasing offspring's reproductive fitness (Buchanan et al., 2019). Genetic diversity provides valuable insight into different populations' genetic makeup and evolutionary/demographic history and conservation strategies, helping to inform efforts to preserve biodiversity and manage endangered species.

In addition to diversity, genetic structure quantifies among-site differentiation and is primarily shaped by historical demography and ongoing connectivity patterns. Measures of spatial genetic structure are defined as the non-random distribution of variation among individuals within populations (Rosetti & Remis, 2012). When populations mate entirely randomly, there will be no discernible differences in allele frequencies among strata (Balkenhol et al., 2015). Conversely, when populations have restricted connectivity, they may become increasingly differentiated due to genetic drift. The rate of gene flow can constrain differences between these temporal changes, and its magnitude will determine the strength of the co-variation. Essentially, disconnected populations become independent of each other. Through

drift, they follow individual evolutionary trajectories, which may have profound implications for population persistence and the probability of species extinction.

The most common metrics used to quantify population-level structure are derived from Wright's F_{st}, which partitions the total genetic variance into two additive components: variation within sampling locations and variation between them (Wright, 1931; Wright, 1949). Values of F_{st} close to zero indicate that populations may share genetic material through either high levels of ongoing gene flow or a history of recent divergence, meaning that populations have only recently separated. Populations can lose genetic diversity but still maintain their structure. With restrictions in genetic connectivity, within-site variation may remain the same, but drift would lead to an increase in among-site variation, thus increasing F_{st}. In extreme cases, where drift has led to the fixation of alleles, $F_{st}=1$, at different populations, it suggests complete isolation. As a result, F_{st} is often used as an inversely proportional indication of the magnitude of connectivity. At smaller spatial scales, one of the drivers of structure may be landscape heterogeneity and how it impacts individual movement patterns (Balkenhol et al., 2015). When landscapes change due to natural or anthropogenic forces, there may be disruptions in previously established connectivity patterns. We would expect organisms occupying these fragments to exhibit higher genetic divergence among subdivided populations, particularly if populations have a small effective population size (Nunney, 1991; Rosenbaum et al., 2007).

Species, such as turtles, whose life-history traits include small populations, longevity, and limited dispersal, often have genetic signatures of low at-site genetic diversity and limited interpopulation genetic exchange (Shaffer & Breden, 1989). Because of their slow reproductive rates and limited dispersal, genetic diversity can remain relatively stable within populations, making it possible for them to mask declines in genetic diversity for decades or even centuries

after severe population decline (Buchanan et al., 2019). For example, in Canadian populations of Blanding's turtles (*Emydoidea blandingii),* Mockford et al. (2005) revealed significant genetic structure (F_{st} = 0.042-0.124; $p < 0.05$) in pairwise comparisons between groups. Populations exhibited isolation by physical distance as the principal determinant of spatial structure, suggesting that the structure of this turtle species pre-dated human influence on the local landscape. Similarly, in a study done by Latch et al. (2011), they discovered that desert tortoises (*Gopherus agassizii*) exhibited weak but significant genetic structure (F_{st} = 0.0046, $p = 0.002$), hypothesizing that either the groups are the result of a recent separation or there are ongoing high levels of gene flow. The landscape genetic analysis identified that both natural (slope) and anthropogenic (roads) landscapes significantly influence the gene flow within these local populations, with gene flow predominantly influenced by the landscape's slope. The spatial genetic structure can be reduced with increased connectivity, whereas isolation, because of distance or properties of the landscape, will allow populations to diverge independently. In the face of rapid environmental change, understanding the spatial genetic structure of a species is essential, enabling the matching of management and ecological scales with the spatial and temporal dynamics of ecological processes, which is especially important in long-lived and late-maturing species.

The spotted turtle (*Clemmys guttata;* Schneider, 1792) is a species of conservation concern due to habitat fragmentation, degradation, and loss. It is also a species found in the illegal pet trade (Dijk, 2011). This species is a small (carapace length up to 14.3 cm) member of the North American Family of freshwater turtles, occupying a wide variety of shallow, unpolluted wetland habitats and the surrounding upland areas. They occur in disjunct populations ranging from the Atlantic coastal lowlands and foothills from southern Maine southwards to

northern Florida and from south Ontario westward along the southern shores of the Great Lakes (Figure 1; Dijk, 2011). Since this species occupies a specialized ecological niche in habitats that are comparatively rare and biologically diverse, coupled with their low vagility (individuals rarely travel > 2 km per year), longevity (30 years), and high fidelity to individual wetlands, they are more likely to experience low levels of gene flow and thus sensitive to localized extinction (Gibbons et al., 2000).

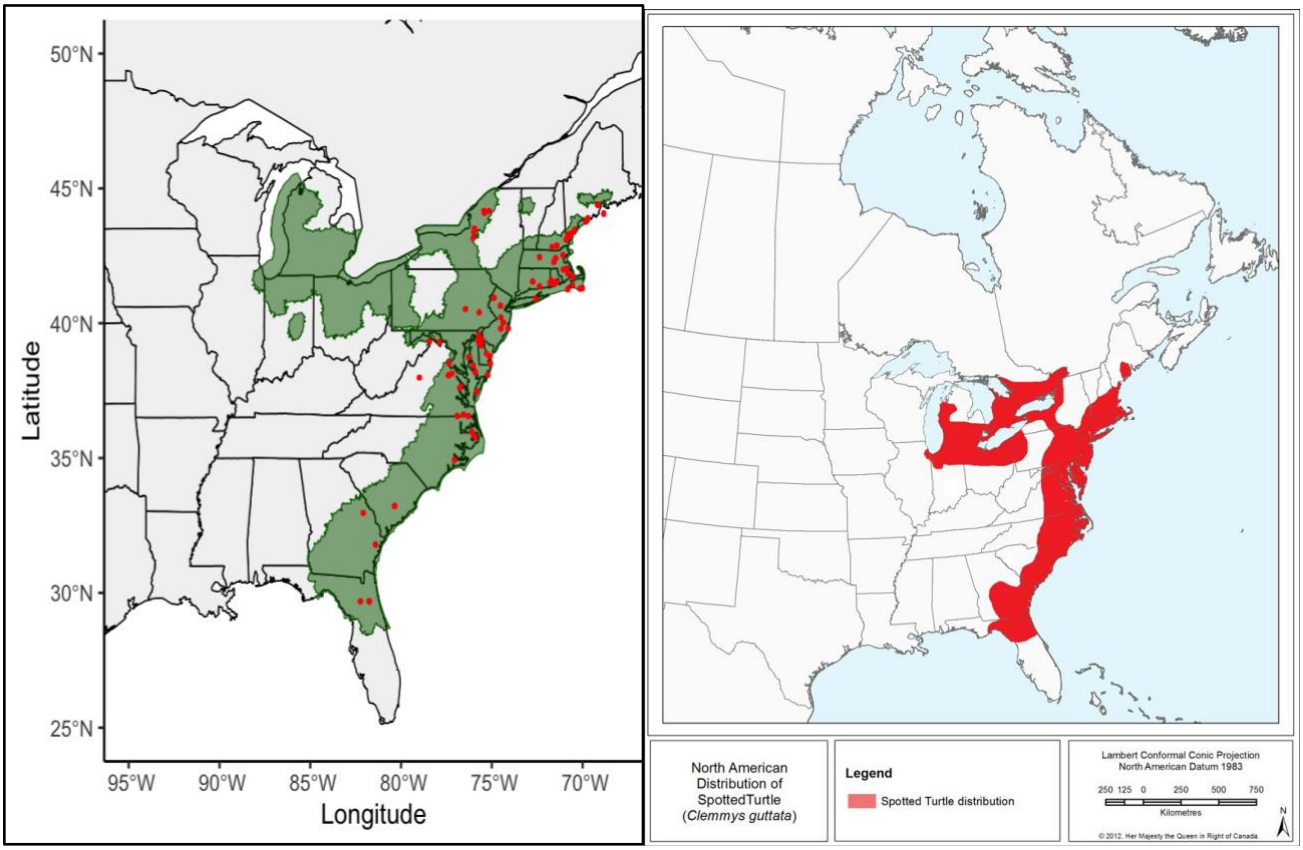

Figure 1: *Species distribution of the Spotted Turtle. Left: The red dots identify sampling locations for broad-scale genetic analyses. Right: Spotted Turtle distribution in North America. The range includes portions of southern Ontario and Quebec and areas south of the Great Lakes in Michigan, Ohio, Indiana, and New York. The range extends down the eastern seaboard from Maine to Florida, extending approximately 200 km inland. From Environment and Climate Change Canada (2018).*

For the Spotted Turtle, characterization of both the spatial distribution of genetic variation and the underlying connectivity network that maintains genetic diversity and structure can be invaluable information for supporting ongoing conservation efforts. There has been limited analysis of the spatial distribution of genetic variation within this species. In a study by Davy & Murphy (2014), the heterozygosity and allelic richness of spotted turtle populations in Canada (Southwest Ontario: $H_O = 0.679$, $H_E = 0.728$, $A = 6.75$; Southeastern Ontario: $H_E = 0.718$, $H_) = 0.707$, $A = 6.11$) were comparable to those observed in larger, sympatric populations. They also found that variations within populations accounted for 91.79% of the variation in the data (AMOVA: $\Phi_{ST} = 0.082$, $p < 0.0001$; n.b., Φ_{ST} is a multilocus corollary of F_{st}). Weak but significant variation occurred among the different populations of turtles ($\Phi_{CT} = 0.038$, $p < 0.0001$) and among sites ($\Phi_{SC} = 0.046$, $p < 0.0001$). Additionally, populations showed isolation by distance, consistent with ongoing genetic connectivity, although limited by spatial dispersal. Similarly, Buchanan et al. (2019) found low estimates of single-locus ($F_{st} = 0.014$) and multi-locus ($\Phi_{ST} = 0.021$) among-population structures and ascribed it to population decline.

Conservation genetics provides fundamental information for effective planning of conservation management through the integration of genetic analyses that quantify genetic diversity, uncover evidence of fragmentation, determine the spatial genetic structure, and estimate rates of gene flow. This study aimed to examine the spatial distribution of the diversity and structure of the spotted turtle, *Clemmys guttata*, from locales throughout the eastern portion of their range. To do this, we worked to develop a comprehensive library of single nucleotide polymorphisms (SNPs) to be used as genome-wide markers. These markers allow for estimating the amount of allelic and genotypic diversity within and among populations and identifying unique subpopulations and the spatial connectivity network of these populations. These estimates

will aid in determining if there are spatial components to how genetic variance is positioned across the landscape and if any diversity hotspots would be of higher conservation value, guiding proposed conservation area networks. The results of this work can be used to provide valuable insights for future conservation planning.

Methods

DNA Collection and Extraction

Individual state wildlife agencies from the eastern seaboard states and commonwealths collected and sampled *C. guttata*. The samples were then sent to the Dyer Laboratory at Virginia Commonwealth University as tissue or blood samples. The sampling organization handled all collecting permits and IACUC protocols. In total, 1643 samples were collected.

Genomic DNA was extracted from scale tips and blood stored in ethanol using the DNeasy Blood and Tissue kit (Qiagen, Boston, MA). A modified protocol was followed for extracting DNA from both blood and tissue. For blood stored in ethanol, the ethanol was spun, pipetted, and dried out of the blood. The blood was then incubated in a 180 µl Buffer ATL and 20 µl proteinase K solution overnight at 56℃ until complete lysis. For tissue samples, the standard protocol was followed except for extending the lysis step to a minimum of 6 hours to increase yield. Following lysis, the DNA extraction procedure adhered to the manufacturer's protocol. All samples were eluted with ultrapure water and stored until further analysis. The concentrations of all samples were standardized before the creation of the sequencing libraries.

Library Preparation and Sequencing

To produce the genetic libraries of *C. guttata*, a modified version of the ddRADseq protocol from Parchman et al. (2012) was used to generate anonymous single nucleotide polymorphic loci for sequencing on the Illumina NovoSeq platform. Our genomic DNA was digested with two restriction enzymes, *EcoRI* and *MseI*, resulting in a large set of random genomic fragments with sticky ends bearing unique DNA sequences. Adapter oligonucleotides containing custom ten base pair (bp) barcodes, which allow individuals to be uniquely identified, were ligated onto the digested fragment ends using T4-ligase (Figure 2). Once barcoded, the

ligated fragments of DNA were amplified using normal polymerase chain reaction (PCR). Individuals with unique barcodes were pooled together in sets of 96 samples. A total of 10 plates were created. Individual locales were randomized among the three libraries to minimize plate-level biases. Pooled libraries were then size-selected using a Bluepippen 300-400bp 2% DF cassette to contain fragments in the approximate range of 300-500bp.

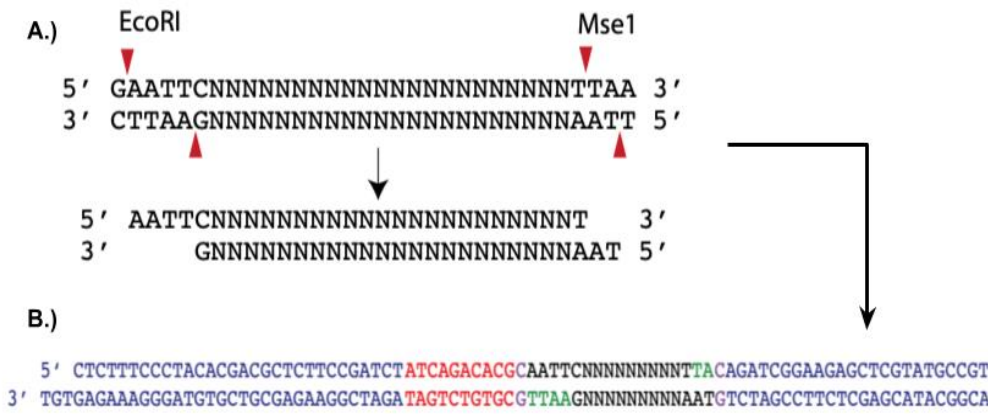

Figure 2: *A.) Double digestion of genomic C. guttata DNA using the restriction enzymes EcoRI and MseI. B.) Ligation of adaptors to genomic fragments created from digestion. Adapter primers for EcoRI and MseI (in blue) with unique sequence barcodes for each individual (in red), and restriction enzyme cut sites (in green). Image from Parchman et al. (2012).*

Single-end sequencing with one multiplexed library was performed by Novogene Corporation using the Illumina NovaSeq platform. All bioinformatic processes were conducted on the Huff cluster at Virginia Commonwealth University's High-Performance Computing Core. After sequencing, reads were de-multiplexed and partitioned using STACKS (Catchen et al., 2013). The resulting de-multiplexed FASTQ files were processed in the *dDocent* bioinformatics pipeline (Puritz et al., 2014). No reference genome exists for *C. guttata*, so a *de novo* reference assembly was developed to identify variable genetic markers. A random subset of reads per individual was selected based on a set of criteria: the absence of insertion or deletion patterns, strictly 2-allele sites, individual quality scores exceeding 20, minor allele frequencies exceeding

5% across individuals, and sequencing depth per individual. The number of reads per individual was too large for analysis. A random subset of sequencing reads was determined and extracted for each sample that could be mapped onto the reference assembly using `bwa` Li & Durbin (2009). Individual SNPs were called using `FreeBayes` (Garrison & Marth (2012)) and all variable sites were converted to minor allele counts for the subsequent genetic analysis. Overall, 913 individuals were retained for the analysis of the eastern portion of the spotted turtle's range, which were sampled from 16 states and 78 locales (Table 1).

Table 1: *Sample sizes and locales for individual turtles yielding SNP genotypes by state. A total of 960 samples from 16 states were used for the spatial genetic analysis for the Spotted Turtle eastern seaboard range.*

State	Sites	Samples
CT	2	23
DE	8	97
FL	2	36
GA	2	36
MA	12	118
MD	4	54
ME	10	134
NC	3	51
NH	3	31
NJ	6	69
NY	8	64
PA	2	12
RI	4	40
SC	1	12
VA	10	112
WV	2	24

Genetic Diversity

Several parameters were employed to assess the genetic diversity of this species. Estimates of allelic diversity, genotypic diversity, and the average heterozygosity per individual were utilized to quantify measures of genetic diversity. At the allelic level, the effective allelic diversity (A_e) was chosen to refer to the number of different alleles present at a particular locus within a population, weighted by their frequencies. It considers both the number of alleles (allele

richness) and their relative abundance in the population. Genotypic diversity was used to measure how alleles are arranged within genotypes within populations and across the entire species. To do this, we looked at both the observed heterozygosity (H_o), the proportion of individuals in a population that are heterozygous at a particular locus, and the expected heterozygosity (H_e), the probability that two randomly chosen alleles from the population will be different at a given locus. The average heterozygosity (H_I), the fraction of all loci in an individual that are heterozygous, was used to provide a summary statistic of genetic diversity that reflects the overall proportion of heterozygous individuals in the population. These analyses were conducted on the R statistical analysis platform using the package `gstudio` (Dyer 2022). Analyzing these parameters both at specific sites and across broader spatial scales aids conservation management efforts. This approach helps identify specific areas with heightened genetic diversity, enabling targeted conservation actions to preserve and enhance genetic health within populations as they exhibit greater resilience to environmental changes.

Population Structure

To quantify and describe the extent of the genetic structure of the spotted turtle, multiple complementary methods were used to provide insights into the evolutionary processes that influence the structure of genetic variation within and among populations. Genetic structure was estimated using Weir & Cockerham's (1984) estimator θ (for F_{ST}) for single locus estimates and extended to multi-locus estimates using Φ_{ST} (Excoffier et al., 1992). A principal component analysis (PCA) of multi-locus genotypes following the methods described by Patterson et al. (2006) was conducted to further assess the spatial distribution of diversity and structure.

The presence of natural clusters of populations due to historical vicariance within our data was identified through a historical admixture model using the clustering program

STRUCTURE (Pritchard et al. (2000), which identifies natural partitions in the data. The number of relevant partitions (K) was assessed through simulations using 50,000 burn-in generations and 100,000 Markov Chain Monte Carlo (MCMC) steps per iteration. Ten replicates were rerun for each value of K, and the most likely number of clusters (K) was determined by following the guidelines of Evanno et al. (2005). We then quantified genetic divergence from the natural clusters identified, providing estimates of inter-population connectivity. In addition to characterizing natural clustering, the overall structure (F_{st}) was described using multi-locus F-statistics using the `gstudio` package (Dyer, 2017). We then conducted a hierarchical analysis of molecular variance (AMOVA) to quantify the partitioning of genetic variance within and among sampled sites and genetic clusters in the `vegan` package (Oksanen et al., 2022.). Lastly, we examined the extent to which genetic differentiation was spatially arranged under a model of Isolation by Distance. For this, we converted the measure of among-population structure using F_{ST} /(1 - F_{ST}) and then fit it to a log-transformed model of inter-population physical distance. The magnitude of the interaction and significance were evaluated using a standard Mantel test (1,000 permutations) in the R package `adegenet` (Jombart, 2008).

Results

Overall Genetic Diversity and Structure

The genetic diversity and structure analysis encompassed 913 sampling locales (sites), utilizing 926 SNP loci. Regardless of sampling locale, allelic and genotypic diversity both exhibit left-skewed distributions, indicative of low genetic variation (mean H_e = 0.088, median H_e = 0.034; Figure 3A & 3B). The mean expected heterozygosity of 0.088 indicates the average level of genetic diversity across the population, with higher values suggesting more genetic variability and lower values suggesting lower genetic variability. The median expected heterozygosity of 0.03 represents the middle value in the dataset, signifying that a substantial portion of the population exhibits lower genetic diversity. The distribution of single-locus genetic structure, measured as θ, for each locus, reveals low levels of inter-local differentiation (mean =0.0294, median =0.0290; Figure 3C). These distributions align with previous observations in spotted turtle populations, highlighting low genetic diversity and structure.

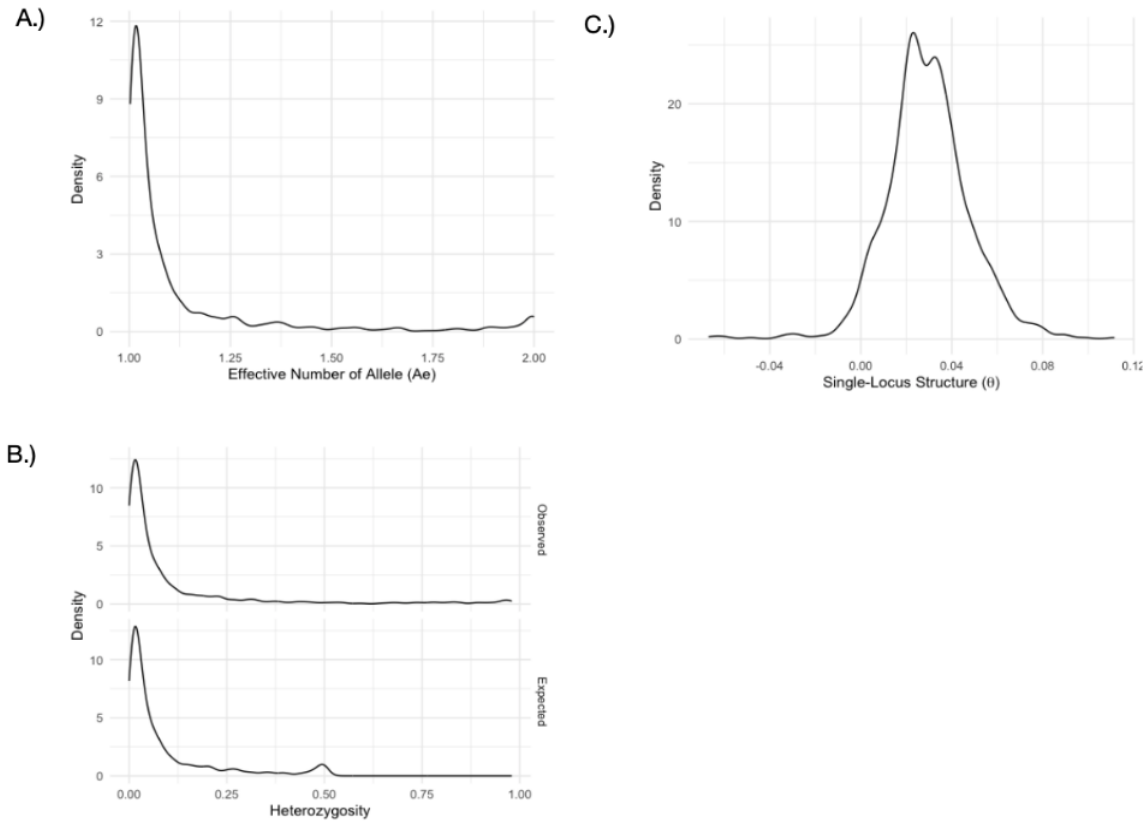

Figure 3: *Genetic diversity and structure analysis for the eastern range of the spotted turtle. Analyses were done for (A) the distribution of effective allelic diversity, (B) the distribution of genotypic diversity as measured by observed (H_o) and expected (H_e) heterozygosity, and (C) the distribution of single locus genetic structure (θ).*

Informative Loci of Genetic Diversity and Structure

To explore loci with spatial signals, we selected a subset from the initial set of 926 loci that have some spatial structuring across the landscape. Employing a conservative approach, loci exhibiting significant levels of inter-local differentiation, having θ => 0.05, were explicitly chosen (Figure 4C; Table S.1). The top 50 most informative loci from this selection were then separated for subsequent analysis. The distribution of allelic diversity and genotypic diversity were calculated. The results revealed a higher level of genetic variation (mean H_e = 0.110,

median H_e =0.050; Figure 4A & 4B) compared to the assessment involving all 926 loci. This suggests that regions with greater genetic structure also exhibit higher genetic variation.

Further analysis of the spotted turtle's genetic structure was done by applying an evolutionary model of historical admixture to determine if there is a natural clustering of individuals. The appropriate number of clusters was determined using the ∂K method proposed by Evanno et al. (2005). Notably, no better representation of the data was found than that of no sub-clustering (K = 1; Figure S.1). Posterior log-likelihood for K=2 and beyond sharply declined. These results do not indicate the presence of any distinct genetic clusters in this species.

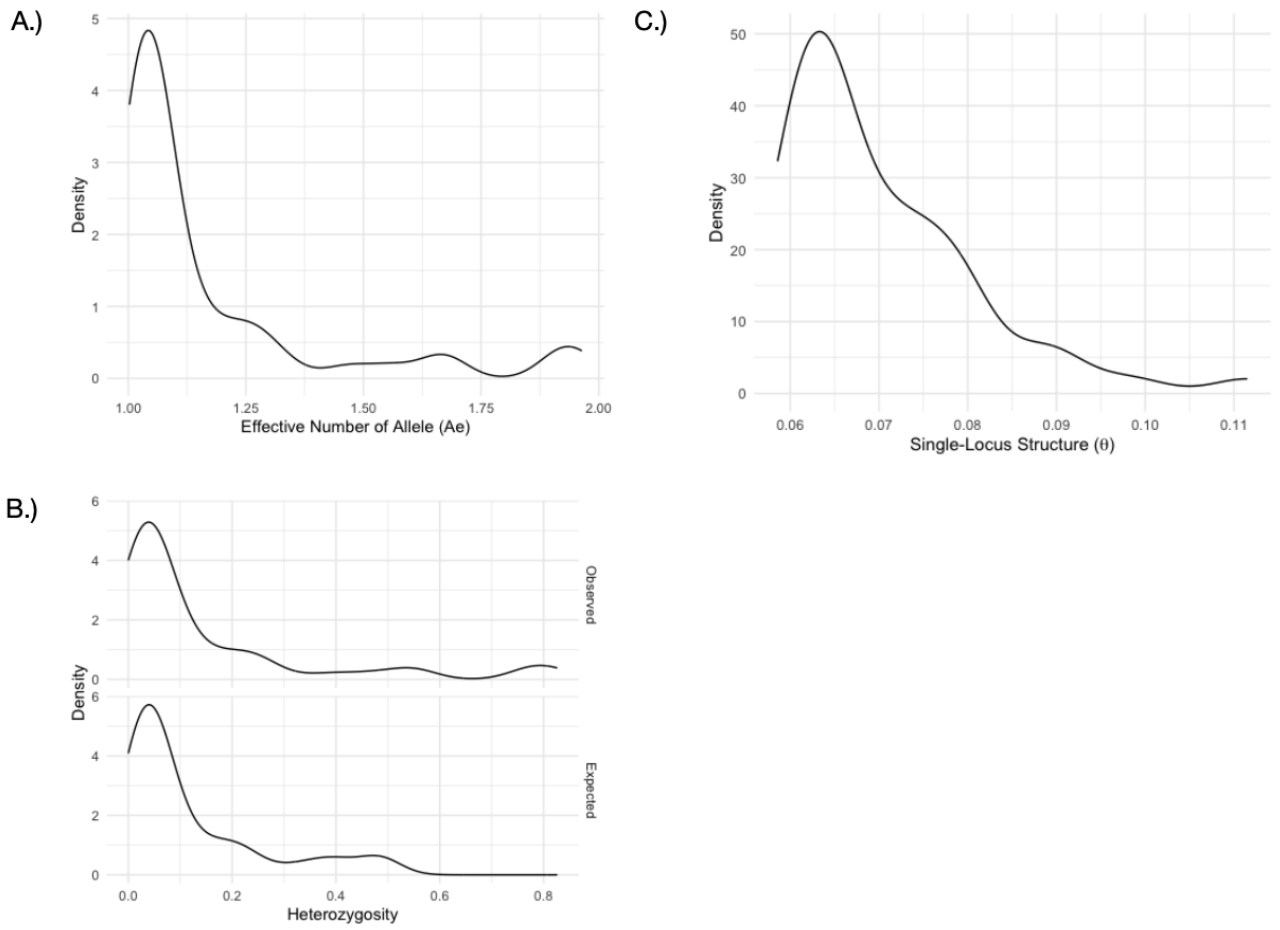

Figure 4: *Genetic diversity and structure analysis for the 50 most informative loci. Analyses were done for (A) the distribution of allele-level genetic effective allelic diversity across the 50 most informative loci, (B) the distribution of genotypic diversity as measured by observed (H_o) and expected (H_e) heterozygosity across the 50 most informative loci, and (C) the distribution of single locus genetic structure (θ) across the 50 most informative loci.*

Spatial Distribution of Genetic Diversity and Structure

By partitioning the data based on individual locales, one can assess the distribution of diversity and structure patterns across the landscape. Within the sampled locales, estimates for locale-specific parameters for allelic and genotypic diversity measures were calculated for the subset of informative loci. The results revealed that two particular loci exhibit both longitudinal and latitudinal gradients in diversity (Table 2), indicating variations in diversity as one traverses the landscape. Both allelic and genotypic diversity have significant negative correlations for both

latitude and longitude, suggesting that there are discernible changes in the genetic makeup at these specific loci as you move across different geographical coordinates.

The analysis of additive multilocus genetic structure via an analysis of Molecular Variance for our informative loci revealed a statistically significant result ($p < 0.05$) of similarly low levels of differentiation (Table 3). Merely 1% of the observed genetic variation is attributed to individuals sampled from alternative populations. This suggests that, while there is a significant difference, the overall genetic variation among individuals from different populations is relatively modest, with most of the genetic variation residing within populations rather than between them.

Table 2: *Spatial distribution of allelic and genotypic measures of diversity. Two loci were found to reduce genetic diversity across the landscape at both the latitudinal and longitudinal levels, as seen by their significant negative values.*

Spatial	Locus	Ho	He	Ae	HoP	HeP	AeP
Latitude	Loc.1064.105	-0.36994	-0.36672	-0.37099	0.00086	0.00096	0.00083
	Loc.5995.25	-0.33702	-0.33971	-0.33460	0.00255	0.00234	0.00275
Longitude	Loc.1064.105	-0.26695	-0.26658	-0.26612	0.01815	0.01831	0.01852
	Loc.5995.25	-0.35765	-0.35663	-0.35834	0.00131	0.00135	0.00128

Table 3: *Analysis of Molecular Variance (AMOVA) table for the multi-locus genetic differences among populations for the top 50 most informative loci. Findings show significant results ($p < 0.05$) of low levels of differentiation.*

	df	SSD	MSD	Phi	Prob
Pop	77	19.347690	0.251269	0.012	0.048
Error	631	142.774450	0.226267		
Total	708	162.122140	0.228986		

Isolation by Distance

To detect the occurrence of isolation by distance, a pairwise Φ_{ST} from Excoffier et al. (1992) was computed and transformed as $\Phi / (1 - \Phi)$. These transformed values were compared to pairwise physical separation (great circle distance). A significant positive correlation was identified (Pearson $\rho = 0.117$, $P = 0.001$; note the significance tested via permutation; Figure 6). Upon closer examination of the data within the 0-20 km range, there appears to be a notable and statistically significant positive pattern of isolation by distance (Pearson $\rho = 0.311$; Figure S.2). However, beyond distances of 20 km, the observed isolation by distance seems to occur randomly.

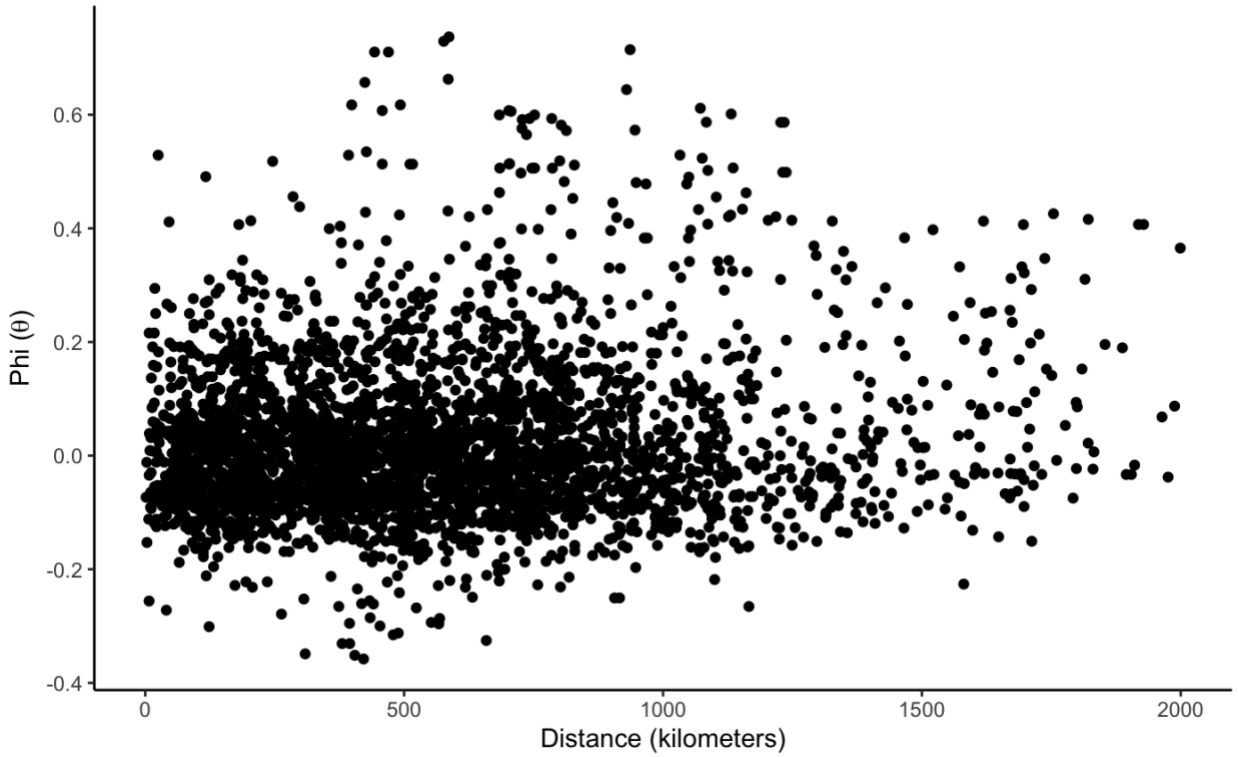

Figure 6: *Isolation by distance for the entire eastern seaboard range of the spotted turtle. To detect the occurrence of isolation by distance, a pairwise* Φ_{ST} *was calculated. A significant positive correlation was identified (Pearson* $\rho = 0.117$*,* $P = 0.001$*, note that the significance tested via permutation).*

Discussion

Conservation Genetics

This research aimed to analyze the spatial genetic variation among spotted turtles across the eastern portion of their current population distribution. Despite facing significant challenges such as habitat fragmentation and loss, populations of *C. guttata* have preserved both their demographic and genetic "integrity" across the entire range. This resilience is particularly evident in the retention of heterozygosity and genetic structure. Previous studies conducted by Davey & Murphy (2014) and Buchanan et al. (2019) have highlighted similarly low levels of genetic structure (F_{ST} = 0.038 and F_{ST} = 0.014, respectively) in smaller, more localized populations of spotted turtles with much more confined spatial distributions. Building upon these findings, our study extends this understanding to a broader context of the entire eastern seaboard range. Consistently, we found that across this expansive geographical area, genetic structure remains relatively low (F_{ST} = 0.04 across the entire genome; Table S.2, and F_{ST} = 0.012 for loci with spatial signal; Table 3).

The observed low genetic structure indicates some combination of an active exchange of genetic material among populations through ongoing levels of gene flow or relatively short times since divergence. Factors such as migration, dispersal, and occasional interbreeding between neighboring populations likely facilitate this gene flow. These processes allow genetic material to move across different sites within the species' range, promoting a relatively even distribution of genetic variation across the region. There could also be historical trends in gene flow, as seen in *E. blandingii*, that likely underlie modern population structure due to a degree of connectivity from the past that would act to minimize differentiation across the landscape (Guinto et al., 2023). This limited yet continuous exchange of individuals over multiple generations or

historical gene flow would result in this genetic homogeneity, as suggested in studies of other freshwater turtles (Kiester et al., 1982; Scribner et al., 1986). Ultimately, this contributes to the spotted turtle species' overall genetic resilience and adaptability.

In addition, our investigation revealed a weak association of inter-locale differentiation, primarily attributed to spatial isolation, with no partitioning between populations due in large part to admixture. The observed association of inter-locale differentiation being weak suggests that while spatial isolation influences genetic diversity, it may not be the sole determinant of population differentiation. Other factors, such as historical events and contemporary ecological processes, such as glaciation, likely interact with spatial isolation to shape genetic patterns across populations. The weak partitioning across populations suggests that while there may have been instances of introgression or gene flow between populations in the past, the extent of genetic mixing has not been sufficient to remove population-specific genetic signatures completely. This means that gene flow has not been strong or frequent enough to erase the genetic differences accumulated between populations over time. These observations underscore the intricate interplay of spatial dynamics and historical factors shaping this species' genetic makeup and population dynamics.

The majority of genetic variation observed in the spotted turtle is due to variation within populations, with little contribution to variation between populations (Table 3). The low inter-population variation may stem from various, sometimes mutually exclusive, mechanisms. One possibility is that the effective population sizes of existing populations are large enough to compensate for the diversifying influence of genetic drift. Additionally, while isolation by distance implies some limitations on dispersal potential, it may not be sufficient to result in spatial segregation across the landscape. This study reveals a complete absence of spatial

partition among the eastern extent of the species' range, indicating sufficient connectivity among sampled populations to resist genetic drift. Whether analyzed through historical admixture models or multi-locus frequency spectra, the absence of nested structure further supports this notion. Notably, no apparent sources of vicariance in the landscape have created sharp discontinuities in the genetic structure. Another factor contributing to low genetic variation could be the slow rates of molecular evolution associated with the species' long generation time (Shaffer et al., 2013), estimated at around 25 years. The divergence time hypothesis suggests that evolutionary rates may not be slow but that recently evolved species have not had adequate time to accumulate polymorphisms (Jordan et al., 2019).

Climatic fluctuations, such as ice ages, interglacial periods, tectonic movements, and sea-level shifts, have led to profound changes in habitats and species distribution—the most recent of which is the Pleistocene, which ended roughly 11,000 years ago. Low genetic diversity is not uncommon in species inhabiting historically glaciated regions (Rosenbaum et al., 2007), which is hypothesized to be impacted by loss of overall diversity during range expansion. In our data, we observed a pattern of multilocus individual heterozygosity, which varies along a latitudinal gradient (Table 2). Specifically, we found that regions in the south exhibit higher genetic diversity than their north counterparts. This pattern aligns closely with post-pleistocence expansion models proposed by Hewitt (1999). It is comparable to patterns of genetic variation across other North American herpetofauna that depend on an aquatic environment (Weisrock, 2000). According to these models, populations in northern latitudes are expected to exhibit reduced genetic diversity and appear comparatively "younger" in evolutionary terms due to their recent expansion. As species undergo directional range expansion in response to changing environmental conditions, it is expected that measures of genetic diversity will decline alone the

spatial axis of expansion. This phenomenon arises because the leading edge of the expansion recruits individuals predominantly from that edge rather than from random locations across the species distribution. In the case of the spotted turtle, as populations expanded northward, they experienced a reduction in local genetic diversity. This provides a nice set of hypotheses for subsequent testing of phylogeographic patterns using a more complete data set that includes the western portion of the species range.

Regardless of the sampling locale, our analysis reveals consistent patterns of low but significant genetic variation within spotted turtle populations (mean $H_e = 0.088$, median $H_e = 0.034$; Figure 3A & 3B). Such a disparity between the values suggests there may be a skewness or asymmetry in the distribution of genetic variation within the population. Several factors could contribute to this observed pattern. Adding to the rapid expansion, historical events such as population bottlenecks, founder effects, or genetic drive may have led to the differential accumulation of genetic variation across different population segments. Additionally, environmental factors such as habitat fragmentation, limited dispersal abilities, or selective pressures may have also influenced the distribution of genetic diversity within the population. Given the rapid pace of human-induced landscape changes relative to post-Pleistocene range expansion, ongoing monitoring and conservation efforts are crucial to assess if further loss of diversity among populations is occurring.

Despite the observed low levels of genetic structure, our analysis reveals evidence of restricted dispersal potential among populations, as indicated by isolation by distance (IBD). This suggests that genetic differentiation between pairs of populations tends to increase with greater geographic separation (Wright, 1938). Implying that populations are effectively isolated from one another at specific spatial scales. In the case of the spotted turtle, our findings suggest

that between distances of 0 to 20 km, there is a clear indication of IBD (Figure 6). As the geographic distance between populations increases within this range, their genetic differences also tend to increase. Individuals within closer proximity exhibit more genetic similarity than those farther apart due to barriers to dispersal or other ecological factors that prevent genetic exchange. However, beyond distances of 20 km, the observed pattern of IBD becomes more random, and as a result, the relationship between genetic and physical separation reaches an asymptote. This implies that, within shorter distances, geographic barriers or localized factors play a role in shaping genetic differentiation. In comparison, at greater distances, other factors might contribute to a more random distribution of genetic variation.

Conservation Implications

Understanding the mechanisms that promote gene flow and maintain genetic variation is critical for conserving and managing threatened and endangered species, such as the spotted turtle. Genetic structure and diversity provide valuable insights into the connectivity and resilience of populations, guiding conservation efforts toward preserving habitat connectivity, managing movement corridors, and mitigating threats like habitat loss and fragmentation. By identifying isolated populations, conservation efforts can be targeted to promote connectivity and gene flow, thereby enhancing vulnerable populations' long-term viability and genetic health across their range.

The limited genetic variability observed within and among spotted turtle populations highlights critical considerations for conservation strategies. Populations exhibited relatively low levels of genetic structure and a high degree of genetic homogeneity across the eastern seaboard range. With no prominent conservation hotspots, these findings underscore the importance of

managing spotted turtle populations as interconnected entities across entire landscapes rather than focusing solely on individual populations. By adopting a landscape-scale approach, conservation efforts can better preserve the species' evolutionary potential in response to changing environmental conditions. Conservation initiatives should prioritize the conservation of representative populations across diverse geographic, ecoregional, and genetic contexts to safeguard the species' genetic diversity and resilience effectively. By conserving populations in various habitats and regions, we can ensure that the species retains the genetic variability needed to adapt to future challenges.

When this project started, only populations from the eastern seaboard were being collected. Based on the findings presented herein, the lack of differentiation and absence of spatial structure is not particularly informative for conservation and management strategies that target specific locales for differential conservation efforts. To gain a more comprehensive understanding of the large-scale, species-level conservation management needs, it is essential to conduct sampling across the entire range of the species. For instance, if populations from the Great Lakes region follow a migration path up the western side of the Appalachian Mountains, we might expect to observe more variations between populations from the Piedmont and Great Lakes regions (as is seen in Blanding’s turtles). On the contrary, if the Great Lakes populations are derived from the Piedmont populations, we might anticipate minimal genetic structure and continued gradients in diversity. Future research endeavors, particularly those utilizing the rest of the genome sequences initiated in this project, hold immense promise in elucidating the species' evolutionary history and adaptive potential. Whole-genome sequencing and establishment of a reference genome would enable the exploration of more intricate patterns of spatial structure and

local adaptation. These insights will build upon the foundation laid by this study and contribute significantly to advancing conservation efforts for this species.

Conclusion

The main goal of this study was to quantify the genetic variation in the spotted turtle throughout its eastern seaboard range. Our analysis provides insights into spotted turtle populations' genetic structure and connectivity across this extensive geographic area. Despite the vast expanse of the eastern seaboard, spotted turtle populations exhibit relatively low levels of genetic structure, indicating a high degree of genetic homogeneity across their range. Additionally, our study reveals evidence of restricted dispersal potential among populations, as indicated by isolation by distance. Suggesting that while gene flow occurs between neighboring populations, it is limited over larger spatial scales. Overall, our findings contribute to a better understanding of the spotted turtle's population dynamics and genetic diversity, which is crucial for informing conservation efforts to preserve this species in its natural habitat.

References

Balkenhol, N., Cushman, S., Storfer, A., & Waits, L. (2015). Introduction to landscape genetics–concepts, methods, applications. In *Landscape Genetics: Concepts, Methods, Applications* (pp. 1–8). https://doi.org/10.1002/9781118525258.ch01

Bouchard, C., Tessier, N., & Lapointe, F. J. (2019). Watersheds influence the wood turtle's (Glyptemys insculpta) genetic structure. *Conservation Genetics*. https://doi.org/10.1007/s10592-019-01169-1

Buchanan, S. W., Kolbe, J. J., Wegener, J. E., Atutubo, J. R., & Karraker, N. E. (2019). A comparison of the population genetic structure and diversity between a common (Chrysemys picta) and an endangered (Clemmys guttata) freshwater turtle. *Diversity*, *11*. https://doi.org/10.3390/d11070099

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*. https://doi.org/10.1111/mec.12354

Davy, C. M., & Murphy, R. W. (2014). Conservation genetics of the endangered spotted turtle (Clemmys guttata) illustrate the risks of "bottleneck tests". In *Canadian Journal of Zoology* (Vol. 92). https://doi.org/10.1139/cjz-2013-0188

Dyer, R. J. (2022). *Gstudio: A suite of functions for the spatial analysis of population and landscape genetic data*. https://github.com/dyerlab/gstudio

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, *14*. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*, *131*, 479–491.

Fisher, R. A., The Genetical Theory of Natural Selection (Clarendon Press, Oxford, 1930).

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing – free bayes – variant calling – longranger. *arXiv Preprint arXiv:1207.3907*.

Gibbons, J.W.; Scott, D.E.; Ryan, T.J.; Buhlmann, K.A.; Tuberville, T.D.; Metts, B.S.; Greene, J.L.; Mills, T.; Leiden, Y.; Poppy, S.; et al. The Global Decline of Reptiles, Déjà Vu Amphibians. BioScience 2000, 50, 653–666.

Government of Canada. (2018). Recovery Strategy for the Spotted Turtle (Clemmys guttata) in Canada [Proposed]. Ottawa, Environment and Climate Change Canada**:** ix-61.

Holman, J.A. (1995). "PleistoceneAmphibians and Reptiles in North America," Oxford Monographs on Geology and Geophysics, Oxford University Press, NewYork.

Guinto, D., Cross, M., Lipps, G., Lee, Y., Kingsbury, B., Earl, D., Dempsey, C., Hinson, J., & Jordan, M. (2023). Conservation Genetic Analysis of Blanding's Turtles across Ohio, Indiana, and Michigan. *Diversity*, *15*(5). https://doi.org/10.3390/d15050668

Hewitt GM (1999) Post-glacial colonization of European biota. Biol J Linn Soc 68:87–112

Jordan, M. A., Mumaw, V., Millspaw, N., Mockford, S. W., & Janzen, F. J. (2019). Range-wide phylogeography of blanding's turtle [Emys (= Emydoidea) blandingii]. *Conservation Genetics*, *20*. https://doi.org/10.1007/s10592-018-01140-6

Kiester, A. R., Schwartz, C. W., & Schwartz, E. R. (1982). Promotion of Gene Flow by Transient Individuals in an Otherwise Sedentary Population of Box Turtles (Terrapene carolina triunguis). *Evolution*, *36*(3). https://doi.org/10.2307/2408108

Latch, E. K., Boarman, W. I., Walde, A., & Fleischer, R. C. (2011). Fine-scale analysis reveals cryptic landscape genetic structure in desert tortoises. *PLoS ONE*, *6*. https://doi.org/10.1371/journal.pone.0027794

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, *25*. https://doi.org/10.1093/bioinformatics/btp324

Litzgus, J. D., & Mousseau, T. A. (2004). Demography of a southern population of the spotted turtle (clemmys guttata). *Southeastern Naturalist*, *3*. https://doi.org/10.1656/1528-7092(2004)003[0391:DOASPO]2.0.CO;2

Mockford, S. W., McEachern, L., Herman, T. B., Snyder, M., & Wright, J. M. (2005). Population genetic structure of a disjunct population of blanding's turtle (Emydoidea blandingii) in nova scotia, canada. *Biological Conservation*, *123*. https://doi.org/10.1016/j.biocon.2004.11.021

Nunney, L. (1991). The influence of age structure and fecundity on effective population size. *Proceedings: Biological Sciences*, *246*(1315), 71–76. http://www.jstor.org/stable/76636

Oksanen J, Simpson G, Blanchet F, Kindt R, Legendre P, Minchin P, O'Hara R, Solymos P, Stevens M, Szoecs E, Wagner H, Barbour M, Bedward M, Bolker B, Borcard D, Carvalho G, Chirico M, De Caceres M, Durand S, Evangelista H, FitzJohn R, Friendly M, Furneaux B, Hannigan G, Hill M, Lahti L, McGlinn D, Ouellette M, Ribeiro Cunha E, Smith T, Stier A, Ter Braak C, Weedon J (2022). vegan: Community Ecology Package. R package version 2.6-4, <https://CRAN.R-project.org/package=vegan>.

Parchman, T. L., Gompert, Z., Mudge, J., Schilkey, F. D., Benkman, C. W., & Buerkle, C. A. (2012). Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology*, *21*, 2991–3005. https://doi.org/10.1111/j.1365-294X.2012.05513.x

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

Puritz, J. B., Hollenbeck, C. M., & Gold, J. R. (2014). dDocent: A RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ*, *2014*. https://doi.org/10.7717/peerj.431

Rosenbaum, P. A., Robertson, J. M., & Zamudio, K. R. (2007). Unexpectedly low genetic divergences among populations of the threatened bog turtle (Glyptemys muhlenbergii). *Conservation Genetics*, *8*. https://doi.org/10.1007/s10592-006-9172-3

Rosetti, N., & Remis, M. I. (2012). Spatial genetic structure and mitochondrial DNA phylogeography of argentinean populations of the grasshopper Dichroplus elongatus. *PLoS ONE*, *7*. https://doi.org/10.1371/journal.pone.0040807

Scribner, K. T., Evans, J. E., Morreale, S. J., Smith, M. H., & Gibbons, J. W. (1986). Genetic Divergence among Populations of the Yellow-Bellied Slider Turtle (Pseudemys scripta) Separated by Aquatic and Terrestrial Habitats. *Copeia*, *1986*(3). https://doi.org/10.2307/1444951

Shaffer, H. B., & Breden, F. (1989). The relationship between allozyme variation and life history: Non-transforming salamanders are less variable. *Copeia*, *1989*. https://doi.org/10.2307/1445988

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., Borchert, G. M., Botka, C. W., Bowden, R. M., Braun, E. L., Bronikowski, A. M., Bruneau, B. G., Buck, L. T., Capel, B., Castoe, T. A., … Wilson, R. K. (2013). The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. *Genome Biology*, *14*. https://doi.org/10.1186/gb-2013-14-3-r28

van Dijk, P.P. 2011. Clemmys guttata. The IUCN Red List of Threatened Species 2011: e.T4968A97411228. http://dx.doi.org/10.2305/IUCN.UK.2011-1.RLTS.T4968A11103766.en

Walker, D. E., & Avise, J. C. (1998). Principles of phylogeography as illustrated by freshwater and terrestrial turtles in the southeastern united states. *Annual Review of Ecology and Systematics*, *29*. https://doi.org/10.1146/annurev.ecolsys.29.1.23

Weir, B. S., & Cockerham, C. C. (1984). Estimating f-statistics for the anlaysis of population structure. *Evolution*, *38*, 1358–1370.

Weisrock, D. W., & Janzen, F. J. (2000). Comparative molecular phylogeography of North American softshell turtles (Apalone): Implications for regional and wide-scale historical evolutionary forces. *Molecular Phylogenetics and Evolution*, *14*(1). https://doi.org/10.1006/mpev.1999.0689

Wright, S. (1931). Evolution In Mendelien Populations. *Genetics*, *16*(2), 97–159. https://doi.org/10.1093/genetics/16.2.97

Wright S. (1938). Size of population and breeding structure in relation to evolution. *Science.* 87:430–431.

Wright, S. (1949). The Genetical Structure Of Populations. *Annals of Eugenics*, *15*(1), 323–354. https://doi.org/https://doi.org/10.1111/j.1469-1809.1949.tb02451.x

Appendix - Supplementary Figures and Tables

Table S.1: *The allelic diversity, genotypic diversity, and the single-locus genetic structure for informative loci from the spotted turtle's eastern seaboard range.*

Locus	Ae	Ho	He	Theta
Loc.1064.100	1.07676	0.07403	0.07129	0.07910
Loc.1064.105	1.02670	0.02635	0.02600	0.07869
Loc.1064.23	1.25662	0.23087	0.20422	0.07805
Loc.1064.29	1.07407	0.06650	0.06896	0.08834
Loc.1064.51	1.14886	0.13927	0.12957	0.07924
Loc.1064.56	1.46278	0.38394	0.31637	0.07091
Loc.1064.70	1.27653	0.24216	0.21663	0.09009
Loc.1064.86	1.17621	0.15558	0.14981	0.07330
Loc.1064.92	1.03965	0.03890	0.03814	0.08076
Loc.107.61	1.93002	0.78993	0.48187	0.07658
Loc.1151.102	1.02153	0.02130	0.02107	0.06347
Loc.1151.110	1.02274	0.02249	0.02223	0.06338
Loc.1151.114	1.04598	0.04497	0.04396	0.06318
Loc.1151.24	1.91319	0.76805	0.47731	0.11150
Loc.1286.46	1.05203	0.04579	0.04946	0.06401
Loc.131.70	1.27377	0.23775	0.21493	0.06662
Loc.147.24	1.01487	0.00000	0.01465	0.07685
Loc.147.9	1.00741	0.00000	0.00735	0.08046
Loc.18.83	1.67622	0.55594	0.40342	0.06644
Loc.1836.14	1.19331	0.17780	0.16200	0.05876
Loc.1836.141	1.09837	0.09398	0.08956	0.05958
Loc.1836.89	1.04637	0.04535	0.04432	0.05858
Loc.1836.97	1.10261	0.09785	0.09306	0.06313
Loc.2033.11	1.66222	0.53706	0.39839	0.08912
Loc.2033.132	1.96427	0.82625	0.49090	0.06180
Loc.426.86	1.22435	0.20167	0.18324	0.05874
Loc.5995.110	1.12794	0.12071	0.11343	0.06198
Loc.5995.123	1.01928	0.01910	0.01892	0.06308
Loc.5995.126	1.05485	0.05068	0.05200	0.05910
Loc.5995.135	1.04779	0.04121	0.04561	0.05910
Loc.5995.18	1.06327	0.05321	0.05951	0.06567
Loc.5995.2	1.06908	0.06685	0.06461	0.06066
Loc.5995.20	1.05318	0.05184	0.05050	0.06376
Loc.5995.25	1.02906	0.02865	0.02824	0.06243
Loc.5995.3	1.02207	0.02183	0.02159	0.06322
Loc.5995.4	1.05318	0.05184	0.05050	0.06609
Loc.5995.55	1.01097	0.01091	0.01085	0.06313
Loc.5995.57	1.02346	0.02319	0.02292	0.06320
Loc.5995.61	1.06762	0.06548	0.06334	0.06385
Loc.5995.65	1.01097	0.01091	0.01085	0.06317
Loc.5995.67	1.05606	0.04366	0.05308	0.06648
Loc.5995.87	1.05318	0.05184	0.05050	0.06305
Loc.61.26	1.04100	0.04019	0.03938	0.07404
Loc.61.46	1.32919	0.28014	0.24766	0.07356
Loc.61.51	1.00474	0.00473	0.00472	0.07195
Loc.61.53	1.00474	0.00473	0.00472	0.07187
Loc.61.58	1.55740	0.46454	0.35790	0.09843
Loc.61.60	1.09000	0.08156	0.08257	0.06790
Loc.61.8	1.03609	0.03546	0.03483	0.07144
Loc.61.9	1.00237	0.00236	0.00236	0.07192

Table S.2: *Analysis of Molecular Variance (AMOVA) table for the multi-locus genetic differences among populations for the entire eastern seaboard range. Findings show significant results ($p < 0.005$) of low levels of differentiation.*

	df	SSD	MSD	Phi	Prob
Pop	77	1421999	18467.52	0.0401	0.004
Error	835	10363310	12411.15		
Total	912	11785309	12922.49		

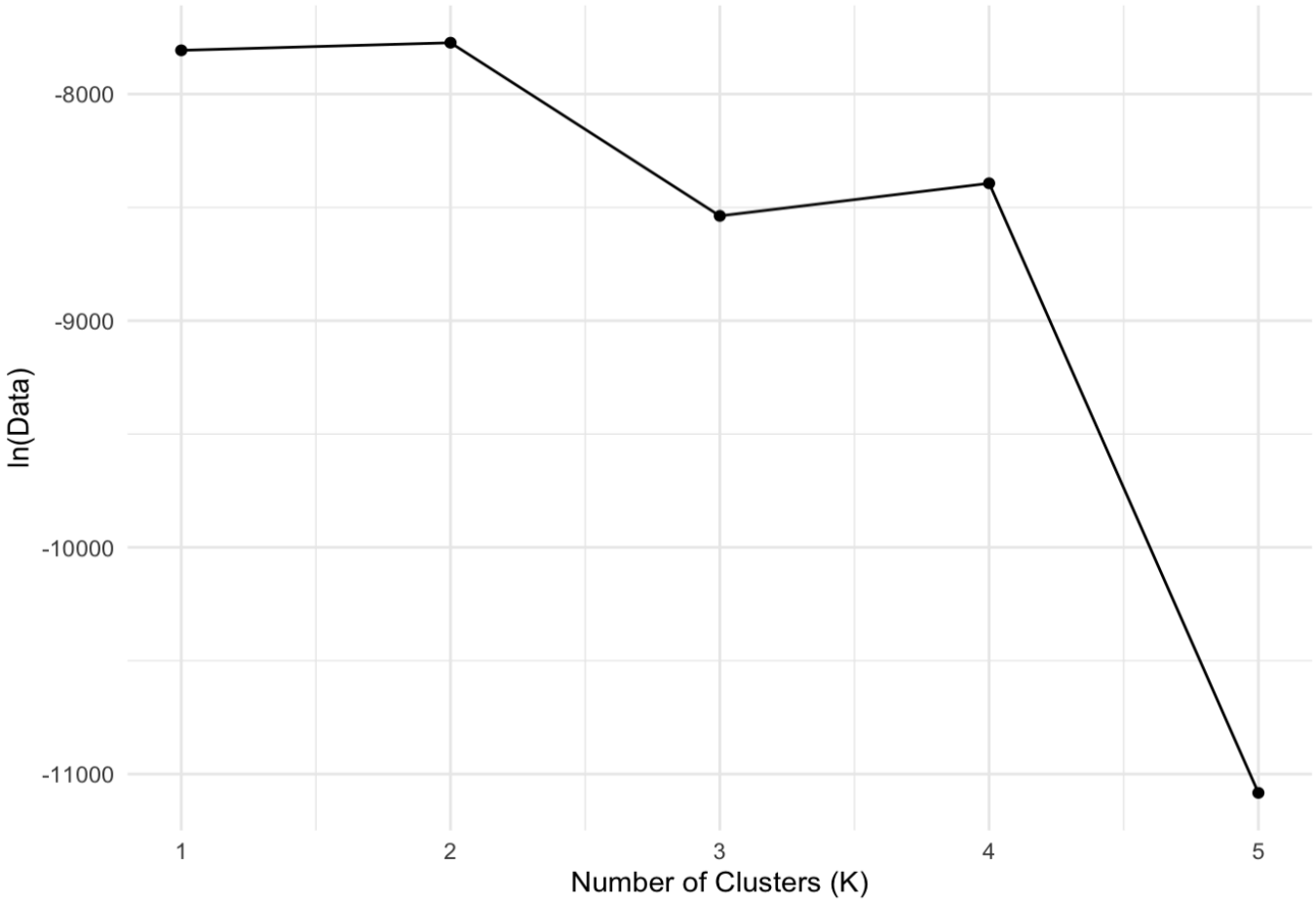

Figure S.1: *STRUCTURE output for further analysis of genetic structure. The likelihood of the data as a function of the number of partitions. For each level, we reported the median of 10 replicate runs. It was found that there is no genetic clustering (K = 1).*

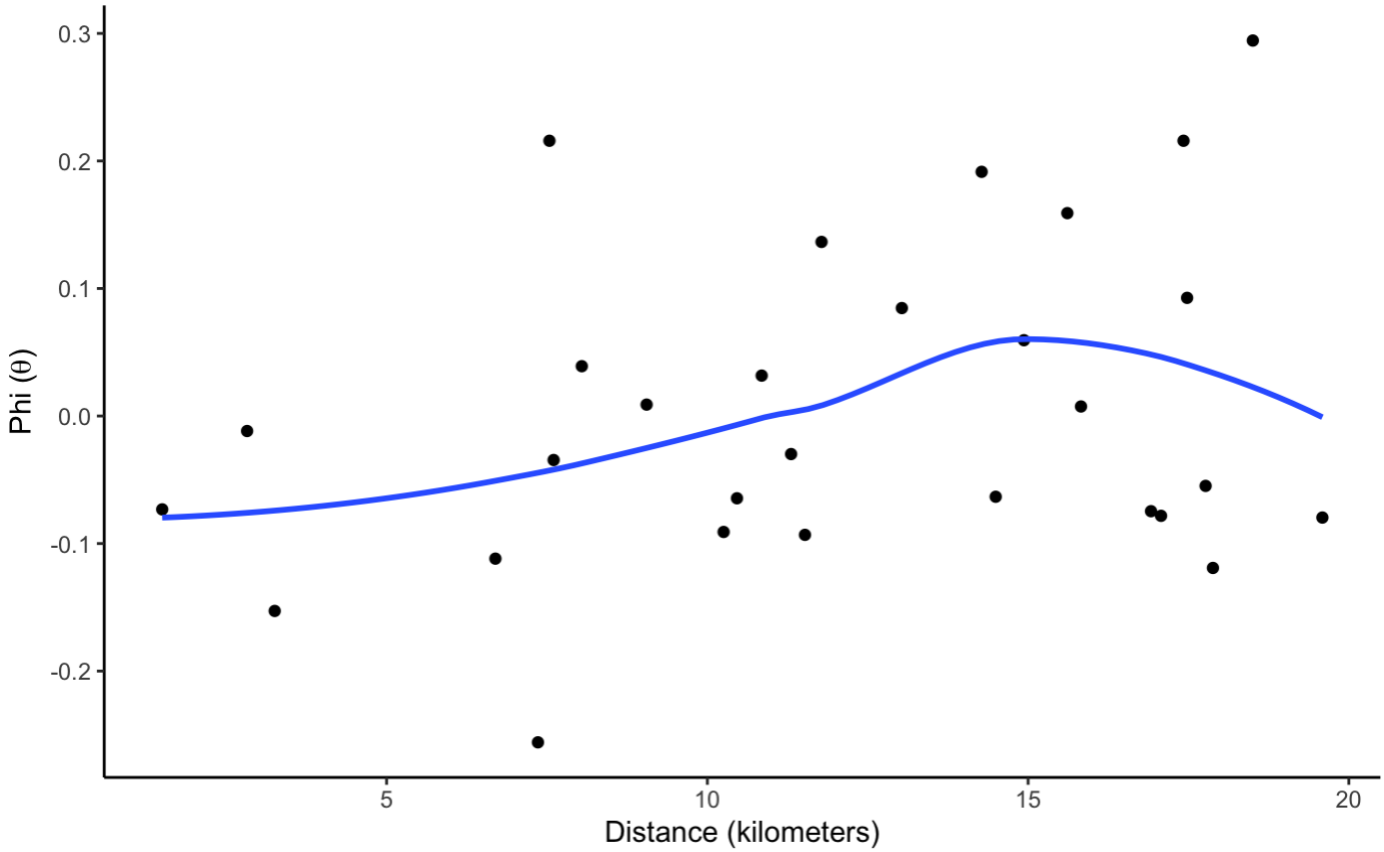

Figure S.2: *Isolation by distance from 0 -20 km. To detect the occurrence of isolation by distance, a pairwise Φ_{ST} was calculated. A significant positive correlation was identified (Pearson ρ = 0.311, P= 0.001, note that the significance tested via permutation)*