1 Fine scale population structure and extensive gene flow within an Eastern
2 Nearctic snake complex (*Pituophis melanoleucus*)

3 Zachary L. Nikolakis^{1,2,*§}, Richard W. Orton^{2*}, and Brian I. Crother¹,

4 ¹Department of Biology, Southeastern Louisiana University, Hammond, LA 70401, USA

5 ²Department of Biology, 501 S. Nedderman Drive, University of Texas at Arlington, Arlington,
 6 TX 76019, USA

7 [§] Corresponding author: Zachary L. Nikolakis

8 ^{*} Authors contributed equally

9 Department of Biology, 501 S. Nedderman Drive, University of Texas at Arlington, Arlington,
 10 TX 76019, USA

11 Email: Zachary.Nikolakis@uta.edu

Abstract

Understanding the processes and mechanisms that promote lineage divergence is a central goal in evolutionary biology. For instance, studies investigating the spatial distribution of genomic variation often highlight biogeographic barriers underpinning geographic isolation, as well as patterns of isolation by environment and isolation by distance that can also lead to lineage divergence. However, the patterns and processes that shape genomic variation and drive lineage divergence may be taxa-specific, even across closely related taxa co-occurring within the same biogeographic region. Here, we use molecular data in the form of ultra-conserved elements (UCEs) to infer the evolutionary relationships and population genomic structure of the Eastern Pinesnake complex (*Pituophis melanoleucus*) – a polytypic wide-ranging species that occupies much of the Eastern Nearctic. In addition to inferring evolutionary relationships, population genomic structure, and gene flow, we also test relationships between genomic diversity and putative barriers to dispersal, environmental variation, and geographic distance. We present results that reveal shallow population genomic structure and ongoing gene flow, despite an extensive geographic range that transcends geographic features found to reduce gene flow among many taxa, including other squamate reptiles within the Eastern Nearctic. Further, our results indicate that the observed genomic diversity is spatially distributed as a pattern of isolation by distance and suggest that the current subspecific taxonomy do not adhere to independent lineages, but rather, show a significant amount of admixture across the entire *P. melanoleucus* range.

Keywords: Phylogenomics; population genomics; ultra-conserved elements; subspecies

Introduction

Understanding the processes and underlying mechanisms that promote and maintain biological diversity (i.e., speciation) is a central goal in evolutionary biology (De Queiroz, 1998; Turelli, Barton, & Coyne, 2001; Wiley & Lieberman, 2011). For example, the most widely acknowledged mode of speciation is one in which populations diverge due to geographic isolation (allopatric speciation), allowing for the accumulation of fixed genetic differences over time (Hoskin, Higgie, McDonald, & Moritz, 2005; Mayr, 1963). Most often, geographic barriers to dispersal and environmental variation (i.e., isolation by environment [IBE]) underpin these reductions in gene flow, and are therefore important to identify when attempting to understand the evolutionary histories of diverged lineages. However, genomic variation may be distributed across geographic space as clines (i.e., isolation by distance [IBD]) when barriers to dispersal or environmental variation only minimally reduce gene flow (Slatkin, 1993; Wright, 1943). In instances where species have greater dispersal capabilities and encounter only weak barriers to gene flow, genomic variation can also be uniformly distributed across panmictic populations; this being most frequently observed in marine and/or migratory taxa (Coltman et al., 2007; Palm, Dannewitz, Prestegard, & Wickström, 2009; White, Fotherby, Stephens, & Hoelzel, 2011). Although examples of divergence with gene flow are becoming increasingly common (Feder, Egan, & Nosil, 2012), it may still be more likely that gene flow among demes would homogenize genetic differences and ultimately prevent lineage divergence (Bolnick & Fitzpatrick, 2007; Dobzhansky & Dobzhansky, 1971; Endler, 1973; Mayr, 1963). As such, it may be equally valuable to identify when genetic continuity has been maintained across a species' geographic distribution, especially when a species occupies an ecologically complex and expansive region with previously recognized biogeographic barriers.

Species with distributions that span large biogeographic regions are more likely to encounter barriers to gene flow and environmental variation that can lead to population genomic structuring compared to species inhabiting more homogenous landscapes (Avise et al., 1987; Soltis, Morris, McLachlan, Manos, & Soltis, 2006). For example, the Eastern Nearctic is a large biogeographic region that spans much of North America and harbors multiple biodiversity hotspots and several well-known biogeographic barriers that reduce gene flow among populations of different species in predictable ways (Burbrink & Guiher, 2015; Burbrink, Lawson, & Slowinski, 2000; McKelvy & Burbrink, 2017; Myers, McKelvy, & Burbrink, 2020). As such, previous studies have shown that the population genomic structure of squamate reptiles within this region tends to correlate with the Apalachicola and Mississippi river drainages (Alexander Pyron & Burbrink, 2009; Burbrink et al., 2000; Weinell & Austin, 2017). However, the impacts of biogeographic heterogeneity on population genetic structure may be taxon specific (as reviewed by Soltis et al., 2006), depending on species range distributions, genetic diversity, and overall vagility. Species with considerable dispersal abilities, such as large mammals (DeCesare et al., 2020; Latch, Heffelfinger, Fike, & Rhodes Jr, 2009) and some birds (Kekkonen et al., 2011), might be expected to exhibit more homogenized levels of genetic diversity than species with more limited vagility. Even within squamate reptiles, different mechanisms may underpin the population structure of species with different dispersal capabilities. Investigating links between landscape features and genomic diversity of widespread taxa within the Eastern Nearctic or similar regions may therefore refine our understanding of the processes and underlying mechanisms that promote and maintain biological diversity.

The Eastern Pinesnake complex (*Pituophis melanoleucus*) is a polytypic wide-ranging species complex occupying much of eastern North America with several disjunct populations. This

81 species typically inhabits upland pine ecosystems, has been categorized as a longleaf specialist in
 82 the most southern regions of its distribution (Baxley & Qualls, 2009), and has one of the largest
 83 home range sizes known across colubroid snakes (Miller, Smith, Johnson, & Franz, 2012). Color
 84 patterns vary across its range; from being uniformly black in areas of southern Mississippi and
 85 Alabama, to a combination of white and dark blotches of red and bronze in the south/central
 86 distribution and black and white in the New Jersey Pine Barrens (Guyer, Bailey, & Mount,
 87 2019). The current taxonomic classification of this complex – largely based on color pattern,
 88 recognizes three geographical sub-specific taxa: (1) the Northern Pinesnake (*P. m.*
 89 *melanoleucus*), (2) the Florida Pinesnake (*P. m. mugitus*), and (3) the Black Pinesnake (*P. m.*
 90 *lodingi*) (Crother, 2012). All three subspecies are of taxonomic interest because of conservation
 91 concerns stemming from habitat degradation/fragmentation and declining prey communities
 92 (e.g., southern pocket gopher decline, longleaf pine degradation; (Landers, Van Lear, & Boyer,
 93 1995; Van Lear, Carroll, Kapeluck, & Johnson, 2005). Previous molecular studies that included
 94 members within this complex have not incorporated population scale sampling and resulted in
 95 variable hypotheses regarding placement within the genus *Pituophis* (Bryson Jr, García-Vázquez,
 96 & Riddle, 2011; Pyron & Burbrink, 2009; Rodríguez-Robles & De Jesús-Escobar, 2000).
 97 Because of the current ambiguous relationships, lack of geographic structuring recovered from
 98 traditional molecular data, morphological variability, conservation concern, and vast geographic
 99 distribution, this group represents an ideal model to investigate phylogenetic and population
 100 genomic patterns. Moreover, this species complex provides additional opportunity to continue
 101 testing hypotheses regarding population differentiation and known geographic barriers.
 102 Here, we aim to quantify population genomic structure within the *P. melanoleucus* complex and
 103 determine if spatial patterns of genomic variation result from barriers to gene flow,

environmental variation, geographic distance, or continuous gene flow across the entire geographic range. To this end, we use nuclear genomic data sampled across the entire geographic range in the form of ultraconserved elements (UCEs) to (1) infer the evolutionary relationships and population structure within the *P. melanoleucus* complex, (2) test for barriers to gene flow, IBE, and IBD, and (3) address the taxonomic validity of the currently recognized sub-specific designations. Notably, this is the first population genomic study of this species, providing invaluable information regarding levels of genetic differentiation and insight for future conservation genomic efforts.

Methods and Materials

Sample Collection and DNA Extraction

Forty-three tissue samples were collected from individual specimens of *Pituophis melanoleucus*, including all of the recognized subspecies across its geographic distribution with five additional samples collected from specimens of *P. ruthveni*, *P. catenifer*, and *Pantherophis obsoletus* to serve as outgroup taxa. Tissue samples were taken either from liver, muscle, shed skins or ventral scale clips and preserved in 95-100% ethanol until DNA extraction. Genomic DNA was isolated by using Qiagen's DNeasy kit following the manufacturers protocols and quantified using a Qubit 2.0 (Thermofisher Scientific, Waltham, Massachusetts, USA). DNA samples were sent to the University of Georgia's Department of Genetics for library preparation and sequencing of (UCEs).

UCE assembly and dataset generation

Samples were barcoded using Illumina TruSeq adapters with unique 8bp sequence tags for each individual and UCEs were targeted by using a tetrapod 5060-locus probe set (open source availability from ultraconserved.org). Demultiplexed samples were filtered and processed by removing adapter sequences and ambiguous bases using the program Illumiprocessor which is incorporated in the open source software package Phyluce v.1.5 (Faircloth, 2016). Reads were assembled *de novo* using standard settings in Velvet v.1.1 (Zerbino & Birney, 2008) and assembled contigs were matched against the 5k UCE tetrapod probe kit to identify and extract UCEs within Phyluce. Sequence alignments for each individual UCE locus were generated using MAFFT v.7.397 (Katoh & Standley, 2013) and concatenated data alignment matrices of 50% and 75% were created in order to examine the effects of missing data on resulting topologies. The data matrix percentages represent the amount of alignments that include every individual (e.g., 50% data matrix would indicate that all alignments will contain at least half of the individuals). Python scripts within Phyluce were used to generate summary statistics of alignment matrices, total reads, contigs, and UCE counts.

To call variants from our assembled sequence capture data, we used the pipeline outlined in Harvey, Smith, Glenn, Faircloth, & Brumfield, 2016, which uses samtools (Li et al., 2009) and Picard (<http://picard.sourceforge.net>) to convert, merge, and index BAM files and remove PCR duplicates. We then used the best practices workflow from GATK (McKenna et al., 2010) to call and hard filter biallelic SNPs that had low quality scores and read depth (i.e., GQ<30 and DP<5). To avoid effects of linkage and violating underlying assumptions for the population genetic clustering software used, we sampled only the first SNP from each UCE loci.

Phylogenetic Analysis and Population genomic analyses

We inferred phylogenetic hypotheses in a maximum-likelihood (ML) framework using the program RAxML v.8.2.3 (Stamatakis, 2014). Twenty ML searches were performed on each dataset (50%, 75%, and first SNP dataset) by partitioning each individual UCE locus under the GTR+ Γ model with the exception of the SNP dataset. This partitioning scheme was chosen over a single partition to estimate individual locus specific parameters (e.g., base frequencies, transition/transversion ratios). Node support was assessed by performing 100 BS replicates for each dataset and visualized using the program FigTree v1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>).

To understand and infer overall population structure and genetic clustering we first performed a principal component analyses (PCA) in the R package adegenet (Jombart & Ahmed, 2011) to visualize the spatial distribution of genetic variance. We then used the program STRUCTURE v.2.3.4 (Falush, Stephens, & Pritchard, 2003; Hubisz, Falush, Stephens, & Pritchard, 2009; Pritchard, Stephens, & Donnelly, 2000) to infer levels of gene flow using the admixture model and correlated allele frequencies with no prior assumption of individual or population assignments. The program Strauto (Chhatre & Emerson, 2017) was used to run multiple iterations of K values (2-7) by utilizing the parallelization function. For each K value, 20 iterations of 10,000 burn-in generations were conducted followed by 50,000 logged generations. To determine the optimal value of K , the ΔK (Evanno, Regnaut, & Goudet, 2005) values were evaluated by using the program STRUCTURE HARVESTER (Earl, 2012). The program CLUMPP was used to combine all iterations for each K value using the greedy algorithm to estimate the most likely membership coefficients. The results were visualized using the package pophelper in R (Francis, 2017).

168 In addition to STRUCTURE, we also used R package conStruct (Bradburd et al., 2018) to assess
 169 population genomic structure with regard to geographic distance. The program conStruct
 170 simultaneously infers discrete (non-spatial model) and continuous (spatial model) patterns of
 171 population genomic structure by estimating ancestry coefficients for each sampled individual
 172 from two-dimensional population layers. Using a cross-validation procedure, conStruct is then
 173 able to assess differential support for both non-spatial and spatial models. Spatial models are
 174 given greater support by conStruct when the decay of allelic covariance is strongly associated
 175 with increasing geographic distance. In such cases, conStruct can then assign all individuals to a
 176 single population layer, where other clustering algorithms cannot, overestimating the number of
 177 discrete clusters (Bradburd et al., 2018; Frantz, Cellina, Krier, Schley, & Burke, 2009;
 178 Meirmans, 2012). Given that conStruct does not perform well with sparse data matrices
 179 (Bradburd et al., 2018), we further filtered our original SNP dataset using VCFtools (Danecek et
 180 al., 2011), to include only sites that were present in at least 75% of samples and excluded 3
 181 samples that had high degrees of missing data, resulting in 1876 SNPs. We then used the cross-
 182 validation method to determine the best fit model between both spatial and non-spatial models
 183 for K values of two through six with ten replicates per K and 25,000 iterations per replicate with
 184 a training proportion of 0.75. We determined the optimal value of K by comparing the resulting
 185 predictive accuracy of each model with the individual layer contributions, testing for differences
 186 with a standard t-test in R base package. We then ran five independent conStruct analyses for
 187 both non-spatial and spatial models at each K value with 25,000 iterations per and confirmed
 188 convergence of analyses by analyzing trace plots in conStruct. Last, we tested correlations
 189 between ancestry coefficients estimated between STRUCTURE and a non-spatial model in
 190 conStruct for $K=2$ using a Pearson's correlation. Outputs from both STRUCTURE and conStruct

were then plotted in R using a combination of the conStruct package and custom scripts (Bradburd et al., 2018).

Estimation of Effective Migration Surface

To visualize patterns of gene flow across the *P. melanoleucus* range, we used the program EEMS – Estimation of Effective Migration Surface (Petkova, Novembre, & Stephens, 2016), which estimates effective migration and provides visualization of migration surfaces that highlight geographic regions of relatively increased and reduced gene flow. EEMS uses a stepping-stone model to relate effective migration rates to expected genetic dissimilarities of geo-referenced samples. We estimated gene flow using three independent chains, each with 750 demes, for 1,000,000 MCMC iterations, with 50,000 iterations of burn-in and a thinning interval of 1,000. We plotted results to check for agreement across all three independent chains and checked for MCMC convergence by visualizing trace plots using the R package, rEEMSplots (Petkova et al., 2016).

Generalized Dissimilarity Modeling

In order to test the roles of biogeographic barriers, environmental variation, and geographic distance in shaping genomic variation across the *P. melanoleucus* range, we used generalized dissimilarity modeling (GDM) on a reduced set of BioClim variables, elevation, geographic distance, and several river drainages using R package gdm (Manion et al., 2017). In particular, we included the Alabama river system, the Apalachicola river, and the Suwannee river, as these river systems have been previously shown to reduce gene flow between demes of several small vertebrate taxa (Avice, Giblin-Davidson, Laerm, Patton, & Lansman, 1979; Burbrink et al., 2000; Church, Kraus, Mitchell, Church, & Taylor, 2003; Jackson & Austin, 2010; Walker &

Awise, 1998). Generalized dissimilarity modeling is a permuted matrix regression technique that has gained recent popularity in population genomic studies for parsing relationships between geographic features, geographic distances, environmental distances, and genetic distances. Briefly, GDM uses a permuted nonlinear matrix regression through the use of I-spline basis functions to accommodate nonlinear statistical relationships (Manion et al., 2017) and incorporates backwards model selection to assess relationships between the response and individual predictor variables.

We downloaded the complete set of BioClim variables at a resolution of 2.5 minutes using the R raster package and removed variables with correlations of 0.75 or greater with the *raster.cor.matrix* function in ENMtools R package (Warren, Glor, & Turelli, 2010), as in Myers et al. (2019). Elevation for each sampling point was also acquired with the raster package using the function *get_elev_points*. Variable reduction retained seven BioClim variables (mean diurnal temperature range, maximum temperature of the warmest month, mean temperature of the wettest quarter, mean temperature of the driest quarter, annual precipitation, precipitation of the driest month, and precipitation of the warmest quarter), which were set as predictor variables in our models. River drainages were determined using custom R scripts which assigned samples to different “basins” according to their orientation to different river systems. For example, samples located west of the Alabama river were assigned a value of four while samples west of the Suwanee River but east of the Apalachicola River were assigned a value of two. We then calculated the ‘distances’ between samples where a value of one would represent samples being separated by a single river system, a value of two represented samples separated by two river systems, and so forth. Our response variable, Euclidean genetic distance, was estimated from our

original VCF using the *dist* function in R package adegenet (Jombart & Ahmed, 2011), using the ‘Euclidean’ flag.

We used GDM to assess the relationships between all predictor variables and the response variable with three I-spline knots and 100 jackknife replicates as in Fitzpatrick & Keller (2015). For each replicate, we removed all samples from a random 10% of sampling locations and model fit was assessed according to the percent deviance explained between permuted and un-permuted models. In order to assess the importance of each individual variable (each of the seven BioClim variables, geographic distance, elevation, and the river basins), we used backwards elimination by running the *gdm.varimp* function in the *gdm* package (Ferrier, Manion, Elith, & Richardson, 2007; Manion et al., 2017) with 100 permutations for each step. This function determines significance for each predictor variable by permuting predictor variables individually and assessing the deviance explained between the un-permuted model and the different permuted models. The relationships between genetic distance and the predictor variables that were determined significant through backwards elimination of the full model were then plotted with the R package *gdm* (Ferrier et al., 2007) using the *plotUncertainty* function after running a second, refined model using only these significant predictor variables. Moreover, we once again ran the *gdm.varimp* function in the *gdm* packaged to assess the relative contribution of each of the predictor variables included in our refined model (geographic distance, mean diurnal range, precipitation of the warmest quarter, and river basin).

Results

Sequence data and data matrices

The average number of reads for each individual was 1,886,033 with a contig range of 7,136 to 19,784 (Table S1). The number of recovered UCEs from all individuals ranged from 3,383 to 4,156 with an average length of 596 bp. The 50% data matrix contained 4,332 UCEs with a total of 2,768,605 bp and the 75% alignment had 3,996 UCEs and 2,623,726 bp. From our variant calling pipeline we called a total of 2910 SNPs that were used in all subsequent population clustering analyses with the exception of conStruct in which we filtered down to include 1876 SNPs. For phylogenetic analyses using only SNP data, which included outgroup taxa, we retained 2715 SNPs that passed all hard filtering parameters.

Phylogenomic analyses

Our phylogenetic results across all datasets (50%, 75%, and SNP) recovered *P. melanoleucus* as a monophyletic group and sister to *P. ruthveni* (Bootstrap value of 100; Figure 1a and Supplemental Figure 1). All inferred topologies did not recover the currently recognized subspecific taxa as monophyletic and generally held the same geographic pattern with samples from the extreme edges of the species distribution tend to group together with high node support (Bootstrap value > 70; e.g., North Carolina/New Jersey, Mississippi, and Peninsula Florida; Figure 1a). Samples that were located in the central region showed the greatest shift in subclade membership as the topology placement was the most variable for locations in central Alabama, Georgia, and panhandle of Florida (e.g., Covington, AL, Marion, GA, and Hernando, FL; Supplemental Figure 1). Node support varied across topologies with most receiving low support values (<70), with the exception of the subclade of all North Carolina and New Jersey individuals and the two Tennessee individuals within the 50 and 75% topologies.

Population Cluster analysis

The final dataset used in all population clustering analyses (with the exception of the conStruct analysis) contained a total of 2,910 SNPs with the amount of individual missing data ranging from 1-58% (average:13%). Principal component analysis revealed a pattern of continuous variation with PC1 describing 6.55% of the variation in the data, and PC2 describing 5.43%. Notably, PCA revealed overlap between all three recognized subspecies classifications (Figure 1b). Results from our STRUCTURE runs showed that the best supported model using the ΔK method was a K value of 2 (Supplemental Figure 2) with ancestry coefficients/admixture levels exhibiting a gradient distributed from east to west across the Florida panhandle and Apalachicola River drainage (Figure 2a). This clinal gradient was further distributed from North Florida through the Piedmont and into central Tennessee. Samples along the East coast in North Carolina and New Jersey were maintained within the cluster assigned to samples throughout the Florida peninsula. Shifts between majority proportions of ancestry occurred at the Alabama river drainage, discretely clustering samples east and west of this drainage. Like STRUCTURE, conStruct also revealed a gradient of ancestry coefficients and gave greatest support to a model with two population layers ($t = 10.46$; $p < 0.001$) (Figure 2b). Notably, for $K = 2$, ancestry coefficients estimated with a non-spatial model in conStruct were highly correlated with the ancestry coefficients estimated by STRUCTURE (correlation = 0.98; $p < 0.001$). Observations of clinal genomic variation were further supported by the cross validation of non-spatial and spatial models in conStruct, which gave greater support for spatial models than for non-spatial models at all values of K (Supplemental Figure 3). The preference of a spatial model indicates an underlying pattern of isolation by distance (IBD), where allelic covariation is associated with geographic distance. The analysis of layer contributions then suggested that little resolution in the genomic data was described by adding even a second layer (Figure 2c, inset).

This inference is apparent in our visualization of ancestry coefficients generated through a spatial model in conStruct plotted atop a geographic map, which showed a majority population layer that persisted across the entire *P. melanoleucus* range (Figure 2c). Unlike STRUCTURE and non-spatial conStruct models, visualization of ancestry coefficients generated by the spatial model of $K = 2$ by conStruct showed very little population genomic structure occurring around any river system. Indeed, the greatest population genomic structure according to this model appears to have occurred throughout the core of the *P. melanoleucus* range, which roughly corresponds to the Piedmont region.

Last, a conStruct non-spatial model of $K = 3$ showed population genomic structuring that occurred at the Alabama river drainage, along with a third population layer that became evident at higher latitudes (e.g., Tennessee, North Carolina, and New Jersey) (Supplemental figure 4a). However, adding a third population layer in the spatial conStruct model did not show any population genomic structuring that occurred around any river system, and only began to add this third layer into samples from the Piedmont region (Supplemental figure 4b).

Estimation of Effective Migration Surface

Visualization of Estimation of Effective Migration Surface (EEMS) is an additional method used to assess gene flow and population genomic structure with respect to geographic features. Our results from EEMS (Figure 3a) revealed patterns of gene flow that were consistent with other population genomic analyses presented here (Figures 1 and 2), where no clear barrier to effective migration was highlighted. Indeed, there appears to be relatively consistent effective migration across the entire *P. melanoleucus* range, which would reduce any population genomic stratification. Notably, the only region of reduced effective migration highlighted by EEMS

occurs throughout much of the Piedmont, where conStruct models indicate increased population genomic structure.

Generalized Dissimilarity Modeling

Last, to test for biogeographic barriers to gene flow, as well as patterns of isolation by environment (IBE), and IBD, we used generalized dissimilarity modeling (GDM). Results of GDM indicated (Table S2: Full model and Table 2: refined model) that the percent deviance explained by the full model that included all seven BioClim variables, geographic distance, elevation, and the river basins was 46.7% ($p < 0.001$). However, a refined model incorporating only the predictor variables determined to be significant in the full model (geographic distance, mean diurnal temperature range, precipitation of the warmest quarter, and river basins) still accounted for 46.3% ($p < 0.001$) of the deviance explained.

Backwards elimination of predictor variables within the refined model further indicated that the river basins only account for 1.1% of the deviance explained from the refined model (Figure 3b) and were ultimately not significant predictors of variation in genetic distance ($p = 0.17$). Mean diurnal range and precipitation of the warmest quarter, meanwhile, explained 22.6% (Figures 4a and 4b) ($p = 0.03$) and 11.9% (Figures 4c and 4d) ($p = 0.02$) deviance from the full model, respectively. However, geographic distance accounted for 63.2% (Figure 4e) of the deviance explained from the full model ($p < 0.001$), lending statistical support for a model of IBD over geographic barriers, IBE, or panmixia.

Discussion

Evidence for a single lineage

344 Here, we present the first phylogenetic and population genomic-scale study of the Eastern
345 Pinesnake (*Pituophis melanoleucus*) species complex. Our results indicate that the currently
346 recognized subspecific nomenclature does not reflect evolutionary history as subspecies were not
347 recovered as independent lineages (i.e., monophyletic) (Figure 1 and Supplemental Figure 1).
348 Rather, this complex appears to be a single species composed of multiple subpopulations with
349 little genetic differentiation indicated by shallow population genomic structuring and moderate
350 levels of gene flow across the entire species distribution, which contrasts many of the previously
351 studied squamate taxa within the Eastern Nearctic (Soltis et al., 2006).
352 Results from our principal components analysis (PCA) revealed a pattern of genetic diversity
353 that anecdotally resembled our geographic sampling distribution and showed no obvious
354 clustering between sub-specific taxonomic designations (Figure 1b). Such patterns in PCA (and
355 lack of monophyly) tend to reflect genetic diversity that is continuously distributed across
356 species' ranges with no phylogenetic breaks and little population stratification (Novembre &
357 Stephens, 2008; Reich, Price, & Patterson, 2008). Results from our STRUCTURE and conStruct
358 analyses also show a clinal pattern of genetic diversity where ancestry coefficients appear to be
359 continuously distributed across geographic space (Figures 2a-c). Moreover, cross validation in
360 conStruct supports a spatial model over a non-spatial model at all values of K , indicating a strong
361 correlation between genetic and geographic distance. Combined with results from our tree-based
362 approach, our clustering analyses suggest a pattern of genetic variation would be consistent with
363 IBE or IBD (Wang & Bradburd, 2014), and not biogeographic barriers or panmixia, which
364 would likely lead to either monophyly and multiple discrete clusters or a single population layer.
365 Although STRUCTURE (Figure 2a) and non-spatial conStruct models (Figures 2b and
366 Supplemental Figure 4a), which are highly correlated, reveal population genomic structure ($K=2$)

that occurs near the Western edge of the *P. melanoleucus* range at the Alabama river drainage, this is likely an artefact of the observed clinal variation and not evidence of a biogeographic barrier. Spatial conStruct models (which are supported over non-spatial models) do not indicate population stratification that occurs near any river system, but rather show shallow structure that occurs throughout the central region of the *P. melanoleucus* range (i.e., the Piedmont) (Figure 2c and Supplemental Figure 4b), which is likely due to genetic drift (Bradburd et al., 2018), potentially brought on by habitat fragmentation (Maloney & Weller, 2011; Taverna, Peet, & Phillips, 2005). Indeed, differences between non-spatial and spatial clustering models presented here may even begin to highlight the importance of incorporating a spatial term in clustering algorithms, as phylogenomic breaks can be prematurely inferred from genomic gradients that transcend potential biogeographic barriers. Here, visualization of ancestry coefficients estimated with spatial conStruct models reveals a majority layer contribution that occurs across the entire *P. melanoleucus* range. Layer contribution analysis implemented in conStruct also shows that little additional resolution in the genomic data is gained by adding layers above $K=1$ (Figure 2c inset). Collectively, these results suggest that geographic barriers have had little to no impact on historical population genomic structure within this species complex and that gene flow across the *P. melanoleucus* range and current subspecific designations is likely ongoing.

Gene flow across a previously recognized genetic barrier

Like clustering analyses, visualization of effective migration across the *P. melanoleucus* range using EEMS shows a moderate reduction in gene flow within Piedmont region, but not adjacent to any river systems (Figure 3a). Rather, effective migration across the entire *P. melanoleucus* range appears to be moderate, with even slightly increased effective migration in parts of South Carolina and Florida. These results contrast previous squamate phylogeography studies

(Burbrink, Fontanella, Pyron, Guiher, & Jimenez, 2008; Burbrink et al., 2000; Weinell & Austin, 2017, Myers et al., 2020) which tend to show substantial genetic variation between populations and clade memberships occurring on opposite sides of different river drainages. Discord among results from our phylogenetic and population genomic structuring analyses compared to other squamate taxa within the Eastern Nearctic may feasibly be underpinned by the dispersal capabilities of these species. Previous studies of movement patterns and annual home range sizes within the genus *Pituophis* span from approximately 35 hectares in *P. catenifer* (Kapfer, Pekar, Reineke, Coggins, & Hay, 2010), to 70 hectares in *P. m. mugitus* (Miller et al., 2012), and over 105 hectares in *P. m. melanoleucus* (Zappalorti, Burger, & Peterson, 2015). Conversely, home range estimates for *Pantherophis* tend to converge on a mean of approximately 10 hectares in *P. alleghaniensis* (Carfagno & Weatherhead, 2008; Durner & Gates, 1993) and 50 hectares in *P. gloydi* (Row, Blouin-Demers, & Lougheed, 2012). Additionally, annual range estimates within *Lampropeltis* have been documented at approximately 15 hectares for *L. getula* (Wund, Torocco, Zappalorti, & Reinert, 2007) and eight hectares for *L. calligaster* (Richardson, Weatherhead, & Brawn, 2006). Because dispersal capability is often directly proportional to home range size (Choudoir, Barberán, Menninger, Dunn, & Fierer, 2018; Pérez-Pérez, López-Moreno, Suárez-Rodríguez, Rheubert, & Hernández-Gallegos, 2017) and inversely related to genetic structure (Brunke et al., 2020; Epperson & Alvarez-Buylla, 1997; Polato et al., 2018; Wright, 1943), it is possible that the larger home range size estimates in *P. melanoleucus* accounts for the lack of discrete phylogenetic breaks that are commonly observed in snake species within the Eastern Nearctic.

Isolation by Environment and Isolation by Distance

In addition to barriers to gene flow, environmental variation can also structure genomic diversity between populations and lead to discrete phylogenetic breaks (Coyne & Orr, 2004). However, environmental variation can also shape genetic diversity across species' ranges as clines, which may be difficult to distinguish from patterns of IBD (Lee and Mitchell-Olds, 2011). Using generalized dissimilarity modeling (GDM), which holds individual predictor variables constant, we simultaneously tested hypotheses of barriers to gene flow, IBE, IBD, and panmixia. Results from our GDM analysis support the hypothesis that continuous gene flow has not been impacted from river drainages within the *P. melanoleucus* complex as these features explain only 1.1% of the deviance and it is not significant (Figure 3b). Using backwards elimination implemented in GDM, we also found moderate but significant patterns of IBE (Figures 4a-4d). Additionally, we find that the range of diurnal temperatures across the year, as well as the amount of precipitation during the warmest quarter of the year, describe 22.6 and 11.9% of the deviance from our refined model with the associations between genetic distance and environmental variation to be the greatest when environmental differences were high. This may explain the variation in ancestry coefficients within our genomic dataset, such as those observed at the transition between the Piedmont and Coastal Plains regions (Figures 2a-2c and Supplemental Figures 4a and 4b). Although this environmental variation could be linked to local adaptation, potentially even for morphological traits such as color pattern (Rosenblum, 2006), it may also drive variation in movement patterns and thus, rates of gene flow between the demes of different regions. However, we still found greatest support for a pattern of IBD within *P. melanoleucus*, where geographic distance described over 62% of the variation in our genomic data (deviance from the refined model) (Figure 4e). This correlation indicates that *P. melanoleucus* is not a truly panmictic species, which would be represented by a weak association between genetic and

geographic distances. These results are also concordant with previous findings that ectothermic species are less likely to be panmictic than other terrestrial vertebrates due to metabolic and energetic requirements (Jenkins et al., 2010). However, among the squamate reptiles of the Eastern Nearctic, *P. melanoleucus* appears to be unique in that river systems have had little to do with shaping genomic diversity across its range.

Conclusion

Our analysis of genomic diversity, population genomic structuring, and gene flow provides insight into how widespread, phenotypically distinct species complexes may have traditionally been regarded as separate lineages, yet these complexes exhibit significant levels of genetic homogenization. More specifically, our results highlight that geographic features typically considered to be biogeographic barriers, can have varying impacts on the population genomic structure of different taxa. It may therefore be naïve to accept certain taxa as paradigms for phylogeographic studies, even among closely related species. Here, genomic variation across the *P. melanoleucus* range is best describe by IBD, as individuals of proximate demes are more likely to exchange alleles than individuals of more geographically distant localities. It is this relationship between geographic distance and genetic decay that has ultimately shaped the shallow and continuous population genomic structure observed across the *P. melanoleucus* range. Although genomic patterns of IBD can lead to reproductively-isolated lineages through either stochastic “breaks” (Baptistini, de Aguiar, & Bar-Yam, 2013; Hoelzer, Drewes, Meier, & Doursat, 2008) or “speciation through forced distance” (i.e., ring species; Irwin, & Price, 2005), this is unlikely to be the case here provided the lack of reciprocal monophyly and shallow levels of population genomic stratification.

While future studies focusing on intraspecies level relationships within the genus *Pituophis* would benefit from increased sampling, the results observed here provide additional evidence for the current lack of understanding regarding the evolutionary relationships among these snakes and suggests that currently this species complex appears to be comprised of a single lineage. We acknowledge that molecular markers other than UCEs may be better suited for these analyses, however, given the inferred shallow nature of this species complex, UCEs are likely still viable for our data and even benefit because they do not suffer from lineage specific dropout as opposed to other reduced representation methods.

Acknowledgments

We would like to thank the following institutions and people for providing samples for this project: the Florida Museum of Natural History (K. Krysko, and K. Enge) Louisiana State University (D. Dittman), North Carolina Museum of Natural Sciences (J. Beane), the Nature Conservancy (J. Lee), American Museum of Natural History (F. Burbrink and A. Mckelvy), Auburn University (B. Folt and D. Laurencio), and the University of Southern Mississippi (B. Kreiser). This research was funded by the Schlieder Foundation.

Data accessibility

Raw data has been deposited under the BioProject number: PRJNAxxxxxx

References

- Alexander Pyron, R., & Burbrink, F. T. (2009). Lineage diversification in a widespread species: Roles for niche divergence and conservatism in the common kingsnake, *Lampropeltis getula*. *Molecular Ecology*, 18(16), 3443–3457.
- Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., ... Saunders, N. C. (1987). Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, 18(1), 489–522.
- Avise, J. C., Giblin-Davidson, C., Laerm, J., Patton, J. C., & Lansman, R. A. (1979). Mitochondrial DNA clones and matriarchal phylogeny within and among geographic populations of the pocket gopher, *Geomys pinetis*. *Proceedings of the National Academy of Sciences*, 76(12), 6694–6698.
- Baptestini, E. M., de Aguiar, M. A., & Bar-Yam, Y. (2013). Conditions for neutral speciation via isolation by distance. *Journal of Theoretical Biology*, 335, 51–56.
- Baxley, D. L., & Qualls, C. P. (2009). Black pine snake (*Pituophis melanoleucus lodingi*): Spatial ecology and associations between habitat use and prey dynamics. *Journal of Herpetology*, 43(2), 284–293.
- Bolnick, D. I., & Fitzpatrick, B. M. (2007). Sympatric speciation: Models and empirical evidence. *Annu. Rev. Ecol. Evol. Syst.*, 38, 459–487.
- Bradburd, G. S., Coop, G. M., & Ralph, P. L. (2018). Inferring continuous and discrete population genetic structure across space. *Genetics*, 210(1), 33–52.
- Brunke, J., Russo, I.-R. M., Orozco-terWengel, P., Zimmermann, E., Bruford, M. W., Goossens, B., & Radespiel, U. (2020). Dispersal and genetic structure in a tropical small mammal,

497 the Bornean tree shrew (*Tupaia longipes*), in a fragmented landscape along the
498 Kinabatangan River, Sabah, Malaysia. *BMC Genetics*, 21, 1–13.

499 Bryson Jr, R. W., García-Vázquez, U. O., & Riddle, B. R. (2011). Phylogeography of Middle
500 American gophersnakes: Mixed responses to biogeographical barriers across the Mexican
501 Transition Zone. *Journal of Biogeography*, 38(8), 1570–1584.

502 Burbrink, F. T., Fontanella, F., Pyron, R. A., Guiher, T. J., & Jimenez, C. (2008).
503 Phylogeography across a continent: The evolutionary and demographic history of the
504 North American racer (Serpentes: Colubridae: *Coluber constrictor*). *Molecular*
505 *Phylogenetics and Evolution*, 47(1), 274–288.

506 Burbrink, F. T., & Guiher, T. J. (2015). Considering gene flow when using coalescent methods
507 to delimit lineages of North American pitvipers of the genus *Agkistrodon*. *Zoological*
508 *Journal of the Linnean Society*, 173(2), 505–526.

509 Burbrink, F. T., Lawson, R., & Slowinski, J. B. (2000). Mitochondrial DNA phylogeography of
510 the polytypic North American rat snake (*Elaphe obsoleta*): A critique of the subspecies
511 concept. *Evolution*, 54(6), 2107–2118.

512 Carfagno, G. L., & Weatherhead, P. J. (2008). Energetics and space use: Intraspecific and
513 interspecific comparisons of movements and home ranges of two Colubrid snakes.
514 *Journal of Animal Ecology*, 416–424.

515 Chhatre, V. E., & Emerson, K. J. (2017). StrAuto: Automation and parallelization of
516 STRUCTURE analysis. *BMC Bioinformatics*, 18(1), 192.

517 Choudoir, M. J., Barberán, A., Menninger, H. L., Dunn, R. R., & Fierer, N. (2018). Variation in
518 range size and dispersal capabilities of microbial taxa. *Ecology*, 99(2), 322–334.

519 Church, S. A., Kraus, J. M., Mitchell, J. C., Church, D. R., & Taylor, D. R. (2003). Evidence for
520 multiple Pleistocene refugia in the postglacial expansion of the eastern tiger salamander,
521 *Ambystoma tigrinum tigrinum*. *Evolution*, 57(2), 372–383.

522 Clark, M. I., Bradburd, G. S., Akopyan, M., Vega, A., Rosenblum, E. B., & Robertson, J. M.
523 (2021). Genetic isolation by distance underlies color pattern divergence in red-eyed
524 treefrogs (*Agalychnis callidryas*). *BioRxiv*.

525 Coltman, D. W., Stenson, G., Hammill, M. O., Haug, T., Davis, C. S., & Fulton, T. L. (2007).
526 Panmictic population structure in the hooded seal (*Cystophora cristata*): HOODED
527 SEAL GENETIC STRUCTURE. *Molecular Ecology*, 16(8), 1639–1648. doi:
528 10.1111/j.1365-294X.2007.03229.x

529 Coyne, J. A., & Orr, H. A. (2004). *Speciation* (Vol. 37). Sinauer Associates Sunderland, MA.

530 Crother, B. (2012). Standard common and current scientific names for North American
531 amphibians, turtles, reptiles, and crocodilians. *Herpetological Circular*, 39, 1–92.

532 Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... Sherry, S. T.
533 (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158.

534 De Queiroz, K. (1998). The general lineage concept of species, species criteria, and the process
535 of speciation. *Endless Forms: Species and Speciation*.

536 DeCesare, N. J., Weckworth, B. V., Pilgrim, K. L., Walker, A. B., Bergman, E. J., Colson, K. E.,
537 ... Jesmer, B. R. (2020). Phylogeography of moose in western North America. *Journal of*
538 *Mammalogy*, 101(1), 10–23.

539 Dobzhansky, T., & Dobzhansky, T. G. (1971). *Genetics of the evolutionary process* (Vol. 139).
540 Columbia University Press.

541 Durner, G. M., & Gates, J. E. (1993). Spatial ecology of black rat snakes on Remington Farms,
542 Maryland. *The Journal of Wildlife Management*, 812–826.

543 Earl, D. A. (2012). STRUCTURE HARVESTER: a website and program for visualizing
544 STRUCTURE output and implementing the Evanno method. *Conservation Genetics
545 Resources*, 4(2), 359–361.

546 Endler, J. A. (1973). Gene flow and population differentiation. *Science*, 179(4070), 243–250.

547 Epperson, B. K., & Alvarez-Buylla, E. R. (1997). Limited seed dispersal and genetic structure in
548 life stages of *Cecropia obtusifolia*. *Evolution*, 51(1), 275–282.

549 Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals
550 using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14(8), 2611–
551 2620.

552 Faircloth, B. C. (2016). PHYLUCE is a software package for the analysis of conserved genomic
553 loci. *Bioinformatics*, 32(5), 786–788.

554 Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using
555 multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, 164(4),
556 1567–1587.

557 Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends
558 in Genetics*, 28(7), 342–350.

559 Ferrier, S., Manion, G., Elith, J., & Richardson, K. (2007). Using generalized dissimilarity
560 modelling to analyse and predict patterns of beta diversity in regional biodiversity
561 assessment. *Diversity and Distributions*, 13(3), 252–264.

562 Fitzpatrick, M. C., & Keller, S. R. (2015). Ecological genomics meets community-level
563 modelling of biodiversity: Mapping the genomic landscape of current and future
564 environmental adaptation. *Ecology Letters*, 18(1), 1–16.

565 Francis, R. M. (2017). pophelper: An R package and web app to analyse and visualize population
566 structure. *Molecular Ecology Resources*, 17(1), 27–32.

567 Frantz, A., Cellina, S., Krier, A., Schley, L., & Burke, T. (2009). Using spatial Bayesian methods
568 to determine the genetic structure of a continuously distributed population: Clusters or
569 isolation by distance? *Journal of Applied Ecology*, 46(2), 493–505.

570 Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian
571 paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society of*
572 *London. Series B. Biological Sciences*, 205(1161), 581–598.

573 Guyer, C., Bailey, M. A., & Mount, R. H. (2019). *Lizards and Snakes of Alabama*. University
574 Alabama Press.

575 Harvey, M. G., Smith, B. T., Glenn, T. C., Faircloth, B. C., & Brumfield, R. T. (2016). Sequence
576 capture versus restriction site associated DNA sequencing for shallow systematics.
577 *Systematic Biology*, 65(5), 910–924.

578 Hoelzer, G. A., Drewes, R., Meier, J., & Doursat, R. (2008). Isolation-by-distance and
579 outbreeding depression are sufficient to drive parapatric speciation in the absence of
580 environmental influences. *PLoS Computational Biology*, 4(7).

581 Hoskin, C. J., Higgie, M., McDonald, K. R., & Moritz, C. (2005). Reinforcement drives rapid
582 allopatric speciation. *Nature*, 437(7063), 1353–1356.

583 Hubisz, M. J., Falush, D., Stephens, M., & Pritchard, J. K. (2009). Inferring weak population
584 structure with the assistance of sample group information. *Molecular Ecology Resources*,
585 9(5), 1322–1332.

586 Irwin, D. E., Bensch, S., Irwin, J. H., & Price, T. D. (2005). Speciation by distance in a ring
587 species. *Science*, 307(5708), 414–416.

588 Jackson, N. D., & Austin, C. C. (2010). The combined effects of rivers and refugia generate
589 extreme cryptic fragmentation within the common ground skink (*Scincella lateralis*).
590 *Evolution: International Journal of Organic Evolution*, 64(2), 409–428.

591 Jenkins, D. G., Carey, M., Czerniewska, J., Fletcher, J., Hether, T., Jones, A., ... Tursi, R.
592 (2010). A meta-analysis of isolation by distance: Relic or reference standard for
593 landscape genetics? *Ecography*, no-no. doi: 10.1111/j.1600-0587.2010.06285.x

594 Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide
595 SNP data. *Bioinformatics*, 27(21), 3070–3071.

596 Kapfer, J., Pekar, C., Reineke, D., Coggins, J., & Hay, R. (2010). Modeling the relationship
597 between habitat preferences and home-range size: A case study on a large mobile
598 colubrid snake from North America. *Journal of Zoology*, 282(1), 13–20.

599 Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7:
600 Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4),
601 772–780.

602 Kekkonen, J., Seppä, P., Hanski, I., Jensen, H., Väisänen, R., & Brommer, J. (2011). Low
603 genetic differentiation in a sedentary bird: House sparrow population genetics in a
604 contiguous landscape. *Heredity*, 106(1), 183–190.

Landers, J. L., Van Lear, D. H., & Boyer, W. D. (1995). The longleaf pine forests of the southeast: Requiem or renaissance? *Journal of Forestry* 93 (11): 39-44.

Latch, E. K., Heffelfinger, J. R., Fike, J. A., & Rhodes Jr, O. E. (2009). Species-wide phylogeography of North American mule deer (*Odocoileus hemionus*): Cryptic glacial refugia and postglacial recolonization. *Molecular Ecology*, 18(8), 1730–1745.

LEE, C.-R., & MITCHELL-OLDS, T. (2011). Quantifying effects of environmental and geographical factors on patterns of genetic differentiation. *Molecular Ecology*, 20(22), 4631–4642.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079.

Maloney, K. O., & Weller, D. E. (2011). Anthropogenic disturbance and streams: Land use and land-use change affect stream ecosystems via multiple pathways. *Freshwater Biology*, 56(3), 611–626.

Manion, G., Lisk, M., Ferrier, S., Nieto-Lugilde, D., Fitzpatrick, M. C., Fitzpatrick, M. M. C., ... Rcpp, I. (2017). Package ‘gdm.’ *A Toolkit with Functions to Fit, Plot, and Summarize Generalized Dissimilarity Models: CRAN Repository, R*.

Mayr, E. (1963). Animal species and evolution. *Animal Species and Evolution*.

McKelvy, A., & Burbrink, F. (2017). Ecological divergence in the yellow-bellied kingsnake (*Lampropeltis calligaster*) at two North American biodiversity hotspots. *Molecular Phylogenetics and Evolution*, 106, 61–72.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., ... Daly, M. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303.

- 628 Meirmans, P. G. (2012). The trouble with isolation by distance. *Molecular Ecology*, 21(12),
629 2839–2846.
- 630 Miller, G. J., Smith, L. L., Johnson, S. A., & Franz, R. (2012). Home range size and habitat
631 selection in the Florida Pine Snake (*Pituophis melanoleucus mugitus*). *Copeia*, 2012(4),
632 706–713.
- 633 Myers, E. A., Bryson Jr, R. W., Hansen, R. W., Aardema, M. L., Lazcano, D., & Burbrink, F. T.
634 (2019). Exploring Chihuahuan Desert diversification in the gray-banded kingsnake,
635 *Lampropeltis alterna* (Serpentes: Colubridae). *Molecular Phylogenetics and Evolution*,
636 131, 211–218.
- 637 Myers, E. A., McKelvy, A. D., & Burbrink, F. T. (2020). Biogeographic barriers, Pleistocene
638 refugia, and climatic gradients in the southeastern Nearctic drive diversification in
639 cornsnakes (*Pantherophis guttatus* complex). *Molecular Ecology*, 29(4), 797–811.
- 640 Novembre, J., & Stephens, M. (2008). Interpreting principal component analyses of spatial
641 population genetic variation. *Nature Genetics*, 40(5), 646–649.
- 642 Palm, S., Dannewitz, J., Prestegard, T., & Wickström, H. (2009). Panmixia in European eel
643 revisited: No genetic difference between maturing adults from southern and northern
644 Europe. *Heredity*, 103(1), 82–89.
- 645 Pérez-Pérez, A., López-Moreno, A. E., Suárez-Rodríguez, O., Rheubert, J. L., & Hernández-
646 Gallegos, O. (2017). How far do adult turtles move? Home range and dispersal of
647 *Kinosternon integrum*. *Ecology and Evolution*, 7(20), 8220–8231.
- 648 Petkova, D., Novembre, J., & Stephens, M. (2016). Visualizing spatial population structure with
649 estimated effective migration surfaces. *Nature Genetics*, 48(1), 94.

650 Polato, N. R., Gill, B. A., Shah, A. A., Gray, M. M., Casner, K. L., Barthelet, A., ... Encalada, A.
651 C. (2018). Narrow thermal tolerance and low dispersal drive higher speciation in tropical
652 mountains. *Proceedings of the National Academy of Sciences*, 115(49), 12471–12476.

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

655 Pyron, R. A., & Burbrink, F. T. (2009). Neogene diversification and taxonomic stability in the
656 snake tribe Lampropeltini (Serpentes: Colubridae). *Molecular Phylogenetics and*
657 *Evolution*, 52(2), 524–529.

658 Reich, D., Price, A. L., & Patterson, N. (2008). Principal component analysis of genetic data.
659 *Nature Genetics*, 40(5), 491–492.

660 Richardson, M. L., Weatherhead, P. J., & Brawn, J. D. (2006). Habitat use and activity of prairie
661 kingsnakes (*Lampropeltis calligaster calligaster*) in Illinois. *Journal of Herpetology*, 423–
662 428.

663 Rodríguez-Robles, J. A., & De Jesús-Escobar, J. M. (2000). Molecular systematics of New
664 World gopher, bull, and pinesnakes (Pituophis: Colubridae), a transcontinental species
665 complex. *Molecular Phylogenetics and Evolution*, 14(1), 35–50.

666 Rosenblum, E. B. (2006). Convergent evolution and divergent selection: Lizards at the White
667 Sands ecotone. *The American Naturalist*, 167(1), 1–15.

668 Row, J. R., Blouin-Demers, G., & Lougheed, S. C. (2012). Movements and habitat use of eastern
669 foxsnakes (*Pantherophis gloydi*) in two areas varying in size and fragmentation. *Journal*
670 *of Herpetology*, 46(1), 94–99.

671 Slatkin, M. (1993). Isolation by distance in equilibrium and non-equilibrium populations.
672 *Evolution*, 47(1), 264–279.

673 Soltis, D. E., Morris, A. B., McLachlan, J. S., Manos, P. S., & Soltis, P. S. (2006). Comparative
674 phylogeography of unglaciated eastern North America. *Molecular Ecology*, 15(14),
675 4261–4293.

676 Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of
677 large phylogenies. *Bioinformatics*, 30(9), 1312–1313.

678 Taverna, K., Peet, R. K., & Phillips, L. C. (2005). Long-term change in ground-layer vegetation
679 of deciduous forests of the North Carolina Piedmont, USA. *Journal of Ecology*, 93(1),
680 202–213.

681 Turelli, M., Barton, N. H., & Coyne, J. A. (2001). Theory and speciation. *Trends in Ecology &*
682 *Evolution*, 16(7), 330–343.

683 Van Lear, D. H., Carroll, W., Kapeluck, P., & Johnson, R. (2005). History and restoration of the
684 longleaf pine-grassland ecosystem: Implications for species at risk. *Forest Ecology and*
685 *Management*, 211(1–2), 150–165.

686 Walker, D., & Avise, J. C. (1998). Principles of phylogeography as illustrated by freshwater and
687 terrestrial turtles in the southeastern United States. *Annual Review of Ecology and*
688 *Systematics*, 29(1), 23–58.

689 Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular Ecology*, 23(23),
690 5649–5662. doi: 10.1111/mec.12938

691 Warren, D. L., Glor, R. E., & Turelli, M. (2010). ENMTools: A toolbox for comparative studies
692 of environmental niche models. *Ecography*, 33(3), 607–611.

693 Weinell, J. L., & Austin, C. C. (2017). Refugia and speciation in North American scarlet snakes
694 (Cemophora). *Journal of Herpetology*, 51(1), 161–171.

695 White, T., Fotherby, H., Stephens, P., & Hoelzel, A. (2011). Genetic panmixia and demographic
696 dependence across the North Atlantic in the deep-sea fish, blue hake (*Antimora rostrata*).
697 *Heredity*, 106(4), 690–699.

698 Wiley, E. O., & Lieberman, B. S. (2011). *Phylogenetics: Theory and practice of phylogenetic*
699 *systematics*. John Wiley & Sons.

700 Wright, S. (1943). Isolation by distance. *Genetics*, 28(2), 114.

701 Wund, M. A., Torocco, M. E., Zappalorti, R. T., & Reinert, H. K. (2007). Activity ranges and
702 habitat use of *Lampropeltis getula getula* (Eastern Kingsnakes). *Northeastern Naturalist*,
703 14(3), 343–360.

704 Zappalorti, R., Burger, J., & Peterson, F. (2015). Home range size and distance traveled from
705 hibernacula in northern pinesnakes in the New Jersey pine barrens. *Herpetologica*, 71(1),
706 26–36.

707 Zerbino, D. R., & Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using
708 de Bruijn graphs. *Genome Research*, 18(5), 821–829.

709

710

Table 1. Results from generalized dissimilarity modeling (gdm) including backwards elimination of individual predictor variables for the refined model. The full refined model includes the four predictor variables determined to be significant in the original full model (Table 1). Each successive model includes all variables from the previous model barring the predictor variable with the least contribution to percent deviance explained. Significance was determined in gdm through 100 permutations. Note that we highlighted our final accepted model that included a model where all predictor variables were determined to be significant. This final model included geographic distance, precipitation of the warmest quarter, and mean diurnal temperature range.

Predictor variable	Full model		Model two		Model three	
	% deviance	P value	% deviance	P value	% deviance	P value
Geographic distance	24.31	0.00	63.22	0.00	63.25	0.00
Mean diurnal temperature range	21.40	0.00	22.63	0.00	22.70	0.03
Precipitation of the warmest quarter	12.44	0.01	11.93	0.02	NA	NA
River basin	1.05	0.17	NA	NA	NA	NA

711

712 Table S1. Taxon, sample locality, number of reads, number of contigs, and number of UCEs for
713 each individual.

Taxon	Sample Locality	Reads	Contigs	UCEs
<i>Pantherophis obsoletus</i>	Jefferson LA	2192806	13226	4050
<i>Pituophis catenifer</i>	Sioux NE	2138167	9525	3937
<i>Pituophis catenifer</i>	Wheeler TX	2789601	16526	3705
<i>Pituophis ruthveni</i>	Bienville LA_1	1834169	8905	4000
<i>Pituophis ruthveni</i>	Bienville LA_2	1940223	10197	4082
<i>Pituophis m. lodingi</i>	Jackson MS_1	2476764	11792	3830
<i>Pituophis m. lodingi</i>	Jackson MS_2	1595826	8416	4076
<i>Pituophis m. lodingi</i>	Jackson MS_3	1591374	8514	4077
<i>Pituophis m. lodingi</i>	Mobile AL	2100669	10270	4105
<i>Pituophis m. lodingi</i>	Perry MS_1	1414642	8744	3867
<i>Pituophis m. lodingi</i>	Perry MS_2	1356673	7619	3889
<i>Pituophis m. lodingi</i>	Perry MS_3	1097785	7715	3897
<i>Pituophis m. melanoleucus</i>	Autauga AL	2116151	9591	3909
<i>Pituophis m. melanoleucus</i>	Burlington NJ_1	1061120	8046	3858
<i>Pituophis m. melanoleucus</i>	Burlington NJ_2	1336710	9439	3868
<i>Pituophis m. melanoleucus</i>	Humphreys TN	1854737	12198	3913
<i>Pituophis m. melanoleucus</i>	Huntingdon TN	1630352	8578	3992
<i>Pituophis m. melanoleucus</i>	Marion GA_1	1853974	12478	4011
<i>Pituophis m. melanoleucus</i>	Marion GA_2	2613221	19784	3436
<i>Pituophis m. melanoleucus</i>	Ocean NJ	1950980	10749	3828
<i>Pituophis m. melanoleucus</i>	Richmond NC	1671822	9973	3969
<i>Pituophis m. melanoleucus</i>	Rockingham NC	1822674	9829	3919
<i>Pituophis m. melanoleucus</i>	Scotland NC	1824724	7731	4156
<i>Pituophis m. melanoleucus</i>	Winston AL	2356280	14303	3810
<i>Pituophis m. melanoleucus</i>	Hoke NC	1335112	7503	4033
<i>Pituophis m. mugitus</i>	Baker GA	2038991	11954	4041

<i>Pituophis m. mugitus</i>	Covington AL_1	2787593	13899	3910
<i>Pituophis m. mugitus</i>	Covington AL_2	2628895	14689	3832
<i>Pituophis m. mugitus</i>	Pierce GA	2422072	10629	4115
<i>Pituophis m. mugitus</i>	Turner GA	1580711	8190	4039
<i>Pituophis m. mugitus</i>	Alachua FL	1830880	10553	4080
<i>Pituophis m. mugitus</i>	Bristol FL	2882788	15246	3622
<i>Pituophis m. mugitus</i>	Brooksville FL	1618837	7840	4089
<i>Pituophis m. mugitus</i>	Cantonment FL	3491931	14175	3832
<i>Pituophis m. mugitus</i>	Hamilton FL	1845695	12142	3895
<i>Pituophis m. mugitus</i>	Hernando FL	1255625	7583	3848
<i>Pituophis m. mugitus</i>	Highlands FL	2214725	12322	4032
<i>Pituophis m. mugitus</i>	Lafayette FL	2314768	10400	3903
<i>Pituophis m. mugitus</i>	Liberty FL	1526608	9735	3860
<i>Pituophis m. mugitus</i>	Marion FL	1308565	7136	4056
<i>Pituophis m. mugitus</i>	Okaloosa FL	1353711	8932	3709
<i>Pituophis m. mugitus</i>	Okeechobee FL_1	1722263	11758	3383
<i>Pituophis m. mugitus</i>	Okeechobee FL_2	1759949	11434	3970
<i>Pituophis m. mugitus</i>	PonceDe Leon FL	1202307	8690	3902
<i>Pituophis m. mugitus</i>	Santa Rosa FL	1128023	7291	3832

Table S2. Results from generalized dissimilarity modeling (gdm) including backwards elimination of individual predictor variables. The full model includes all seven bioclim variables determined to lack autocorrelation with other bioclim variables, elevation, and river drainage. Each successive model includes all variables from the previous model barring the predictor variable with the least contribution to percent deviance explained. Significance was determined in gdm through 100 permutations. Note that we highlighted the independent predictor variables determined to be significant in our original full model.

Predictor variable	Full model		Model one		Model two		Model three		Model four		Model five		Model six		Model seven		Model eight	
	% Deviance	P value	% Deviance	P value	% Deviance	P value	% Deviance	P value	% Deviance	P value	% Deviance	P value	% Deviance	P value	% Deviance	P value	% Deviance	P value
Geographic distance	6.18	0	6.18	0	6.18	0	7.01	0	7.24	0	8.72	0	25.85	0	63.2	0	63.47	0
Elevation	0.14	0.64	0.14	0.62	0.14	0.65	0.12	0.63	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Mean diurnal range	21.42	0.01	21.42	0	21.42	0.02	21.47	0.01	21.45	0.02	21.41	0.02	21.99	0.01	23.59	0.01	24.32	0.02
Max temp warmest month	0.92	0.45	0.92	0.48	0.92	0.44	0.9	0.35	0.84	0.4	0.8	0.44	NA	NA	NA	NA	NA	NA
Mean temp wettest month	0	0.99	0	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Mean temp driest quarter	0.03	0.66	0.03	0.67	0.03	0.7	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Annual precipitation	0.08	0.61	0.08	0.62	0.08	0.69	0.06	0.61	0.05	0.71	NA	NA	NA	NA	NA	NA	NA	NA
Precipitation of the driest month	0	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Precipitation of the warmest quarter	10.08	0.01	10.08	0.02	10.08	0.02	10.05	0.01	11.39	0.03	11.34	0.03	11.01	0.01	10.29	0.08	NA	NA
River drainage	2.09	0.04	2.09	0.05	2.09	0.04	2.06	0.02	2.01	0.02	2.26	0.04	1.53	0.08	NA	NA	NA	NA

Figure 1. Panel highlighting the genetic structuring and patterns across the Eastern Pinesnake complex. A. Results of the maximum likelihood SNP tree with tips and shapes corresponding to the subspecific taxonomic classification and state locality which is also highlighted in the subsequent panel. B. Principle coordinate analysis (PCA) plot showing the distribution of genetic variance of all individuals for the first two loadings.

Figure 2. Geographic patterns of ancestry coefficients. A., B., and C. Map of samples with individuals shown as pie charts for $K=2$ model. Colors represent cluster membership across STRUcTURE (A) and conStruCT (both non-spatial and spatial models; B. & C) analyses, which are indicated in the bottom left corner of each panel. Panel C inset shows the dissection of layer contributions for conStruCT spatial models sampled across K values ranging from one to four.

/44

Figure 3. Estimates of effective migration surfaces (EEMS) and correlations of genetic distance across major river basins. A. EEMS plot showing the migration surface for all sampled individuals. Reduced rates of effective migration are represented by red and yellow shades while blue and purple shades reflect increased rates of effective migration. B. Output from generalized dissimilarity modeling showing the relationship between pairwise genetic (y-axis) distances and the predicted number of river basins separating samples (x-axis). Note that the percent deviance explained and the p-value for the model are in the top right corner and show the variation in genetic distance explained by the predictor variable does not significantly deviate between unpermuted and permuted models.

Figure 4. Results from generalized dissimilarity modeling and maps of environmental variation across the *Pituophis melanoleucus* range. A. The relationship between pairwise genetic distance and variation in mean diurnal monthly temperature range (degrees Celsius x 10 [max – min]) estimated from generalized dissimilarity modeling. B. Mean diurnal monthly temperature range (degrees Celsius x 10 [max – min]) across the *P. melanoleucus* range. C. The relationship between pairwise genetic distance and variation in precipitation of the warmest quarter (mm) from generalized dissimilarity modeling. D. Precipitation of the warmest quarter (mm) across the *P. melanoleucus* range. E. The relationship between pairwise genetic distance and transformed geographic distance estimated with generalized dissimilarity modeling. The percent deviance explained by the permuted model from the unpermuted model and the associated p-value are in the top right or bottom right of each plot from our generalized dissimilarity modeling output.

Supplemental Figure 1. Output from STRUCTURE Harvester. Line plot showing support for a $K=2$ model using the Delta K method for the STRUCTURE analysis.

791
792 **Supplemental Figure 3.** Output from conStruct showing the predictive accuracy for both non-
793 spatial and spatial models from values of $K=1$ to $K=6$.

Supplemental Figure 4. Results from conStruct spatial and non-spatial models. Map of samples with individuals shown as pie charts for $K=3$ model for both spatial (A) and non-spatial (B) models.