Received: 15 April 2019 | Revised: 8 January 2020 | Accepted: 13 January 2020

DOI: 10.1111/mec.15358

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Biogeographic barriers, Pleistocene refugia, and climatic gradients in the southeastern Nearctic drive diversification in cornsnakes (*Pantherophis guttatus* complex)

Edward A. Myers[1,2] | **Alexander D. McKelvy**[3] | **Frank T. Burbrink**[2]

[1]Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA

[2]Department of Herpetology, The American Museum of Natural History, New York, NY, USA

[3]Department of Biology, The Graduate School and Center, City University of New York, New York, NY, USA

Correspondence
Edward A. Myers, Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA.
Email: eddie.a.myers@gmail.com

Abstract

The southeastern Nearctic is a biodiversity hotspot that is also rich in cryptic species. Numerous hypotheses (e.g., vicariance, local adaptation, and Pleistocene speciation in glacial refugia) have been tested in an attempt to explain diversification and the observed pattern of extant biodiversity. However, previous phylogeographic studies have both supported and refuted these hypotheses. Therefore, while data support one or more of these diversification hypotheses, it is likely that taxa are forming within this region in species-specific ways. Here, we generate a genomic data set for the cornsnakes (*Pantherophis guttatus* complex), which are widespread across this region, spanning both biogeographic barriers and climatic gradients. We use phylogeographic model selection combined with hindcast ecological niche models to determine regions of habitat stability through time. This combined approach suggests that numerous drivers of population differentiation explain the current diversity of this group of snakes. The Mississippi River caused initial speciation in this species complex, with more recent divergence events linked to adaptations to ecological heterogeneity and allopatric Pleistocene refugia. Lastly, we discuss the taxonomy of this group and suggest there may be additional cryptic species in need of formal recognition.

KEYWORDS

hindcast niche modelling, historical demography, Mississippi River, phylogeographic model selection, species delimitation

1 | INTRODUCTION

Genetic structure within a species forms due to reduced connectivity and limited gene flow between populations. The causes of reduced gene flow across the distribution of a species can be numerous. For example, disruption to continuous migration can be the result of local adaptation to contrasting environments, where ecologically mediated natural selection can result in divergence despite gene flow (Oliveira et al., 2015; Papadopulos et al., 2011; Rundle & Nosil, 2005). Historical changes in climate and shifts in species-specific suitable habitats may result in population divergence in historically allopatric refugia (Lumibao, Hoban, & McLachlan, 2017; Myers, Bryson, et al., 2019; Taberlet, Fumagalli, Wust-Saucy, & Cosson, 1998; Ursenbacher, Carlsson, Helfer, Tegelström, & Fumagalli, 2006). Alternatively, biogeographic barriers that bisect the geographic distribution of a taxon can hinder dispersal and gene flow, ultimately causing divergence and speciation (Hickerson, Stahl, & Lessios, 2006; Pyron & Burbrink, 2010). It is possible that multiple determinates of population structure can interact producing genetically discrete populations (e.g., Aleixo, 2004; Bradburd, Ralph, & Coop, 2013; Myers, Xue, et al., 2019). For example, populations may diverge in allopatric refugia and expand population sizes post

divergence. This population expansion may now result in their geographic distributions being determined by the borders of biogeographic barriers (e.g., major rivers) which currently limit gene flow between recently diverged sister taxa but that were not instrumental to their original divergence. Such scenarios are expected in regions that are geologically complex with diverse habitats, especially if these regions presented climatically stable refugia throughout the Pleistocene.

These causes of population differentiation may result in distinct demographic histories that can be evaluated using phylogeographic model selection methods, where unique models differ in the timing and directionality of gene flow, population size fluctuations, and timing of divergence (Carstens, Morales, Jackson, & O'Meara, 2017; Charles et al., 2018; Portik et al., 2017). If, for example, biogeographic barriers have driven diversification, the expected outcome is divergence with little or no gene flow across a barrier. Alternatively, population divergence resulting from natural selection across environmental gradients will result in a history of population divergence despite continuous gene flow between sister populations; under such a scenario population sizes might be expected to remain stable through time. Lastly, refugial isolation will leave a signature of allopatric divergence where sister pairs diverge in the absence of gene flow and reduced effective population sizes, where post-divergence population sizes expand with the potential for secondary contact and gene flow. While these demographic models are simplifications of the true underlying evolutionary histories, they are useful in testing major biogeographic patterns with population genomic data (Carstens et al., 2013).

These contrasting scenarios of population divergence and speciation can be better understood by incorporating both current and hindcast ecological niche models (ENMs) with coalescent analyses of genomic data (Carstens & Richards, 2007). By using climatic data to estimate potential geographic distributions, ecological niche models make it possible to explicitly test if discrete lineages have diverged into different climatic environments (Myers et al., 2013; Warren, Glor, & Turelli, 2010) or alternatively have retained a particular niche despite population divergence (Peterson, Soberón, & Sánchez-Cordero, 1999). For example, ecological divergence would show phylogeographic estimates supporting divergence with gene flow and ENM comparisons would suggest that sister lineages occur in ecologically distinct environments with no evidence of a biogeographic barrier. Alternatively, if biogeographic barriers are promoting divergence, phylogeographic models should show reduced gene flow while ENM demonstrate little or no ecological divergence. Furthermore, by hindcasting ENMs onto projected climates during the Quaternary, it is possible to identify regions that

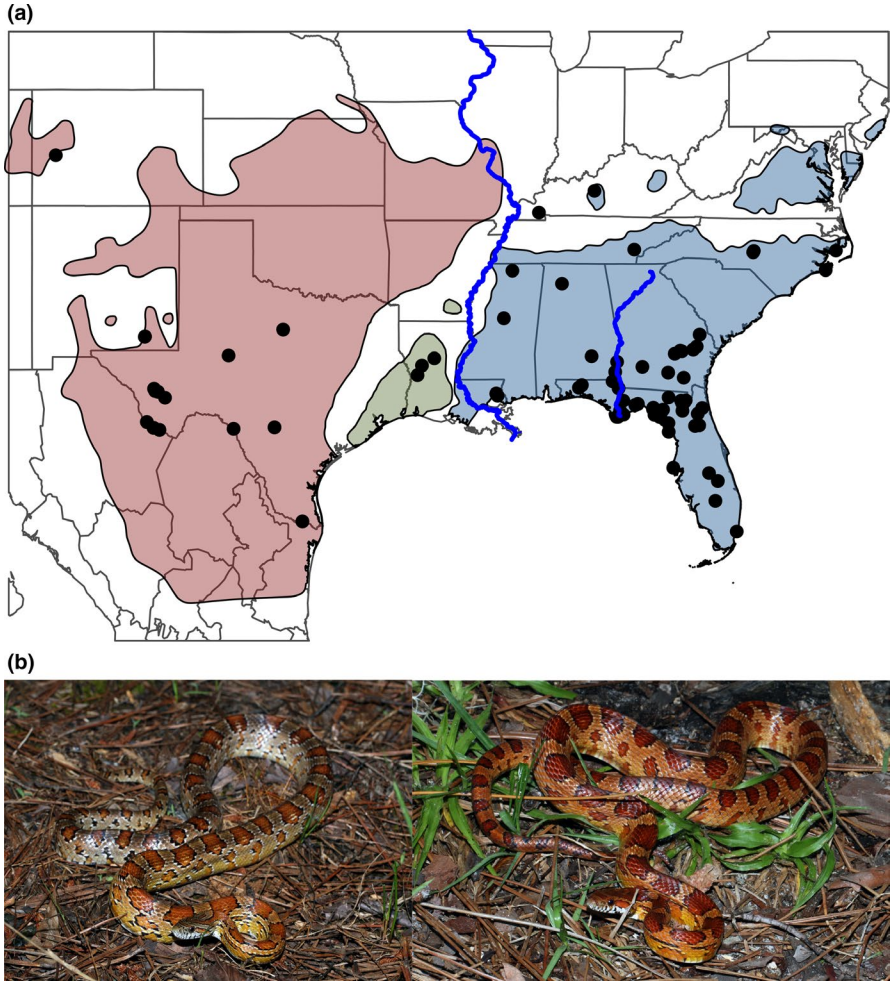


FIGURE 1 (a) Distribution of the *Pantherophis guttatus* species complex (*Pantherophis emoryi* in red, *Pantherophis slowinskii* in green, and *P. guttatus* in blue). Sampling localities of specimens used for genomic analyses are indicated by black circles, and the Mississippi and Apalachicola Rivers are highlighted in blue. Geographic distributions are based on IUCN assessments (Echternacht & Hammerson, 2016); (b) *P. guttatus* photographed from the southeastern United States (photo credits to ADM) [Colour figure can be viewed at wileyonlinelibrary.com]

have been climatically stable through time, potentially serving as refugia (Carnaval, Hickerson, Haddad, Rodrigues, & Moritz, 2009). Phylogeographic models may suggest divergence in allopatric refugia, which can be supported with hindcast ENMs that demonstrate isolated regions of climatically stable habitats (Ilves, Huang, Wares, & Hickerson, 2010; Waltari & Hickerson, 2012).

Within the southeastern Nearctic, many studies have attributed population divergence to a combination of local adaptation, physically isolating barriers, and historical refugia during the Pleistocene and Pliocene (Soltis, Morris, McLachlan, Manos, & Soltis, 2006). The southeastern Nearctic contains numerous diverse habitats that vary from temperate conifer and broad leaf forests, to grasslands that transition into the xeric shrubland of the Chihuahuan Desert (Bailey, 1995). These environmental transitions result in turnover of biological assemblages of vertebrates across the Nearctic (Burbrink & Myers, 2015) and has been implicated in driving speciation via ecological divergence (McKelvy & Burbrink, 2017; Soltis et al., 2006). This region is also bisected by numerous large rivers, most notably the Mississippi and Apalachicola Rivers (Figure 1), which are commonly cited as biogeographic barriers (Burbrink, Lawson, & Slowinski, 2000; Pauly, Piskurek, & Shaffer, 2007; Satler & Carstens, 2017). The southeastern Nearctic is a biodiversity hotspot and is thought to have acted as a climate refugium throughout the Quaternary for many endemic taxa (Noss et al., 2015; Waltari et al., 2007). It was further hypothesized that there have been multiple refugia within this region during the glacial periods of the Pleistocene leading to divergence and speciation (Waltari et al., 2007). While these sharp environmental transitions, bisecting rivers, and potential for habitat stability throughout the Quaternary have probably resulted in population divergence among many codistributed taxa, some authors have found little correspondence between the location of genetic breaks within species and biogeographic barriers (Barrow, Lemmon, & Lemmon, 2018) and lack of influence from historical climate change on population differentiation (Barrow, Soto-Centeno, Warwick, Lemmon, & Moriarty Lemmon, 2017). This suggests that particular species-specific demographic responses drive divergence and population demographic change among taxa in the southeastern Nearctic (Burbrink et al., 2016; Lumibao et al., 2017).

The cornsnakes (*Pantherophis guttatus* complex, including *P. guttatus*, *Pantherophis emoryi*, and *Pantherophis slowinskii*) are medium-sized snakes that are widely distributed across the eastern to central United States into northern Mexico (Figure 1; Schultz, 1996). These snakes inhabit a variety of habitats across this broad range, from tropical hammocks in southern peninsular Florida to dry Chihuahuan Desert scrub (Schultz, 1996). Previous phylogeographic research on this taxon has demonstrated significant lineage divergence within the mtDNA genome and suggested that these divergent lineages warrant species status recognition (Burbrink, 2002). The mtDNA phylogeographic structure within this group occurs at the Mississippi River and between populations within longleaf pine habitats of Louisiana and eastern Texas compared to those more widespread in the central prairies and scrub habitat (Burbrink, 2002). Based on these results, the mtDNA lineage east of the Mississippi River remained *P. guttatus*, the lineage west of the Mississippi River and restricted to longleaf pine habitat was named *P. slowinskii*, and the lineage broadly distributed in the prairies and Chihuahuan desert was elevated to species status, *P. emoryi* (Burbrink, 2002). This probably indicates that divergence occurred due to both biogeographic barriers and habitat turnover. However, it is also possible that isolation in allopatric refugia has been an important driver of diversification within this species complex. Here, we generate reduced-representation genomic data and use coalescent-based demographic models coupled with current and hindcast ecological niche models to assess the drivers of diversification within this taxon across the southeastern Nearctic. Additionally, we reassess species limits within this group in order to better quantify biodiversity, as this is essential information for conservation management (Folt et al., 2019; McKelvy & Burbrink, 2017).

2 | MATERIALS AND METHODS

2.1 | Data collection

We acquired a total of 80 tissue samples via fieldwork and museum sample loans; these samples largely span the geographic distribution of the *Pantherophis guttatus* species complex and include samples from all currently recognized species, as well as *Pantherophis bairdi* for use as an outgroup (see Figure 1). We extracted DNA using either Qiagen DNeasy or MagAttract HMW DNA kits, following manufacture protocols. We submitted DNA to the University of Wisconsin-Madison Biotechnology Center. DNA concentration was verified using the Quant-iT PicoGreen dsDNA kit (Life Technologies). Libraries were prepared as in Elshire et al. (2011) with minimal modification; in short, 150 ng of DNA was digested using PstI and MspI (New England Biolabs) after which barcoded adapters amenable to Illumina sequencing were added by ligation with T4 ligase (New England Biolabs). Ninety-six adapter-ligated samples were pooled and amplified to provide library quantities amenable for sequencing, and adapter dimers were removed by SPRI bead purification. Quality and quantity of the finished libraries were assessed using the Agilent Bioanalyzer High Sensitivity Chip (Agilent Technologies, Inc.) and Qubit dsDNA HS Assay Kit (Life Technologies), respectively. Libraries were standardized to 2 nM. Sequencing was performed using single read, 100 bp sequencing and HISEQ SBS Kit version 4 (Illumina) on a HiSeq2500 sequencer, samples were multiplexed to a total of 144 samples per lane.

2.2 | Bioinformatics and SNP calling

We used ipyrad (Eaton & Overcast, 2020) for de novo assembly of the genomic data set with many of the default settings (e.g., a clustering threshold of 0.85, no barcode mismatches, and a minimum sequencing depth of six reads for base calling). However, we changed the following parameters from default setting: permitted up to four

low quality bases per sequence read, trimmed reads for adapters and primers, only kept reads with a minimum length of 75 base pairs after trimming, and required a minimum of 67 samples per locus for outputs, which is equivalent to retaining no more than 17% missing data per locus.

Since selection can result in similar patterns of population genomic variation as historical demographic events and if not accounted for can bias demographic estimates (Hahn, 2008), we implemented the F_{ST} outlier method in BAYESCAN version 2.1 (Foll & Gaggiotti, 2008) to identify and remove loci putatively under selection for demographic analyses that assume neutrality. We used VCFTOOLS (Danecek et al., 2011) to thin the data set to one SNP per locus and remove the outgroup *P. bairdi* sample; and converted the resulting vcf file to BayeScan input format using the R package RADIATOR (Gosselin, 2017). We ran BayeScan for 50,000 iterations following a burnin length of 80,000 iterations and 50 pilot runs. We used a false discovery rate of 0.05 in identifying selected loci and assessed convergence of BayeScan by examining trace plots of the MCMC chain for stationarity in R.

2.3 | Population structure

To assess the number of genetic groups within the *P. guttatus* complex, we used both multivariate and model-based methods. For these analyses, we used only a single SNP per radseq locus. First, we implemented principle component analysis (PCA), a fast and effective approach for analysis of large SNP data sets (Patterson, Price, & Reich, 2006), in the *R* package ADEGENET (Jombart, 2008). Secondly, we implemented discriminant analysis of principal components (DAPC; Jombart, Devillard, & Balloux, 2010), a *K*-means clustering method also implemented in the *R* package ADEGENET (Jombart, 2008). We selected the number of genetic clusters based on the number of PC axes that minimized the Bayesian information criterion score. Third, we used sparse nonnegative matrix factorization (sNMF) in the R package LEA (Frichot & François, 2015), which implements sparse non-negative matrix factorization algorithms and computes least-squares estimates of ancestry coefficients (Frichot, Mathieu, Trouillon, Bouchard, & François, 2014) to estimate population structure. We allowed *K* to vary between 1 and 6, with 100 replicate runs for each value of *K*, and an alpha value set to 5,000. The cross-entropy calculation was used to assess support for *K* values; this generates masked genotypes to predict ancestry assignment error where lower values indicate a better prediction of the true *K* ancestral population value (Frichot et al., 2014). We compared the best cross-entropy score for each replicate and the *K* value that did not decrease in cross-entropy score for the next *K* iteration to infer the best supported number of ancestral populations. Because missing data can influence population assignment, we removed seven samples missing greater than 50% data and used this data set for both DAPC and sNMF (removed EAM191, FTB1243, FTB3293, FTB3429, H15968, H7478, H8423). We assessed phylogenetic relationships among sampled individuals by using the full concatenated GBS data set in RAxML-8.2.11 (Stamatakis, 2006). We applied the rapid bootstrapping approach in RAxML, and rooted the tree with *P. bairdi* using the GTR GAMMA model of sequence substitution.

2.4 | Ecological niche modelling

We downloaded *P. guttatus* and *P. emoryi* occurrence records from VertNet that contained georeferenced localities and all *P. slowinskii* occurrence records. *Pantherophis slowinskii* occurrences lacking latitude and longitude records were georeferenced in Google Earth. We removed any localities that fell outside the known distribution of these taxa and divided *P. guttatus* samples into two populations based on the population structure analyses. We spatially thinned occurrences using the R package spThin (Aiello-Lammens, Boria, Radosavljevic, Vilela, & Anderson, 2015), resulting in a data set of occurrences where nearest neighbors were no less than 50 km apart. We downloaded bioclimatic variables from WORLDCLIM version 1.4, which represent monthly averages of current climatic conditions (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005), at 2.5 arcmin resolution to be used in constructing ENMs. We removed all but one of all variables that had a Pearson's correlation >0.70, resulting in four bioclimatic variables: Annual Mean Temperature, Annual Precipitation, Mean Diurnal Range, Mean Temperature of the Wettest Quarter. To test if the lineages identified above have nonequivalent environmental niches, in which case ENMs should be conducted separately on each population, we used the enmtools.ecospat.id function of the *ENMTOOLS R* package (Warren et al., 2010). This test, based on the method developed in Broennimann et al. (2012), does not rely on ENMs, but instead uses kernel density smoothing to estimate the density of environmental niche space of a species, and corrects for the availability of environments when measuring overlap. We conducted this test on all adjacent population pairs using the uncorrelated bioclimatic variables.

To identify optimal model parameters, we tested all combinations of six feature classes (Linear; Linear Quadratic; Hinge; Linear Quadratic Hinge; Linear Quadratic Hinge Product; and Linear Quadratic Hinge Product Threshold) and six regularization multipliers (1, 2, 5, 10, 15, and 20) using the *ENMEVAL* package in R (Muscarella et al., 2014). To construct ENMs for each population independently we used BIOMOD2 (Thuiller, Georges, & Engler, 2013) and implemented the best fit model from ENMEVAL. We sampled 10,000 pseudoabsence points within the potential range of the *P. guttatus* complex (circumscribed within: −110, −74, 18, 50) and MAXENT version 3.4.1 (Phillips, Anderson, & Schapire, 2006) was used to construct ENMs using the four uncorrelated bioclimatic variables. We conducted each analysis with 25 evaluation runs and replicated for 5,000 iterations, reserving 25% of samples as a training data set for model evaluation; we created response curves and jackknifed our data to measure variable importance. We projected averages of these ENMs and saved as ascii files, hindcasting these models to both Last Glacial Maximum (21 kya) and the Mid-Holocene (6 kya) conditions using the CCSM4 general

circulation model projections for these time periods based on the same four noncorrelated variables identified in the current climate data set. To identify regions of habitat stability through time for each population, we stacked and averaged the current and two projected-paleo climate models. Regions highlighted in these stacked projections demonstrate the potential geographic extent of refugia through time (e.g., Portik et al., 2017).

2.5 | Species delimitation

We conducted species delimitation analyses using the Bayesian model comparison method of BFD* (Leaché, Fujita, Minin, & Bouckaert, 2014). The competing delimitation models were derived from current taxonomy based on single-locus analyses (Burbrink, 2002) and different combinations of the exploratory analyses of population structure of the genomic data collected here. Ultimately we tested six models: (a) four species based on sNMF and DAPC results, (b) three species based on current taxonomy, (c) three species with *P. guttatus* east of the Mississippi River split into two taxa and all individuals west of the river lumped (as suggested from sNMF results), (d) two species, separated by the Mississippi River (*K* = 2 results from sNMF), (e) two species, combining *P. slowinskii* with *P. guttatus* and testing that group against *P. emoryi*, and (f) the *P. guttatus* complex represents a single taxon.

We used SNAPP version 1.3.0 (Bryant, Bouckaert, Felsenstein, Rosenberg, & RoyChoudhury, 2012) in BEAST version 2.5.1 (Bouckaert et al., 2014) to estimate species trees and calculate the marginal likelihoods of each species delimitation model. We set mutation rates u and v to 1.0, gave the lambda prior a gamma distribution with alpha = 2 and beta = 200, set the 'snapprior' to alpha = 1, beta = 100, and lambda = 20. We used a stepping-stone analysis PathSampleAnalyser

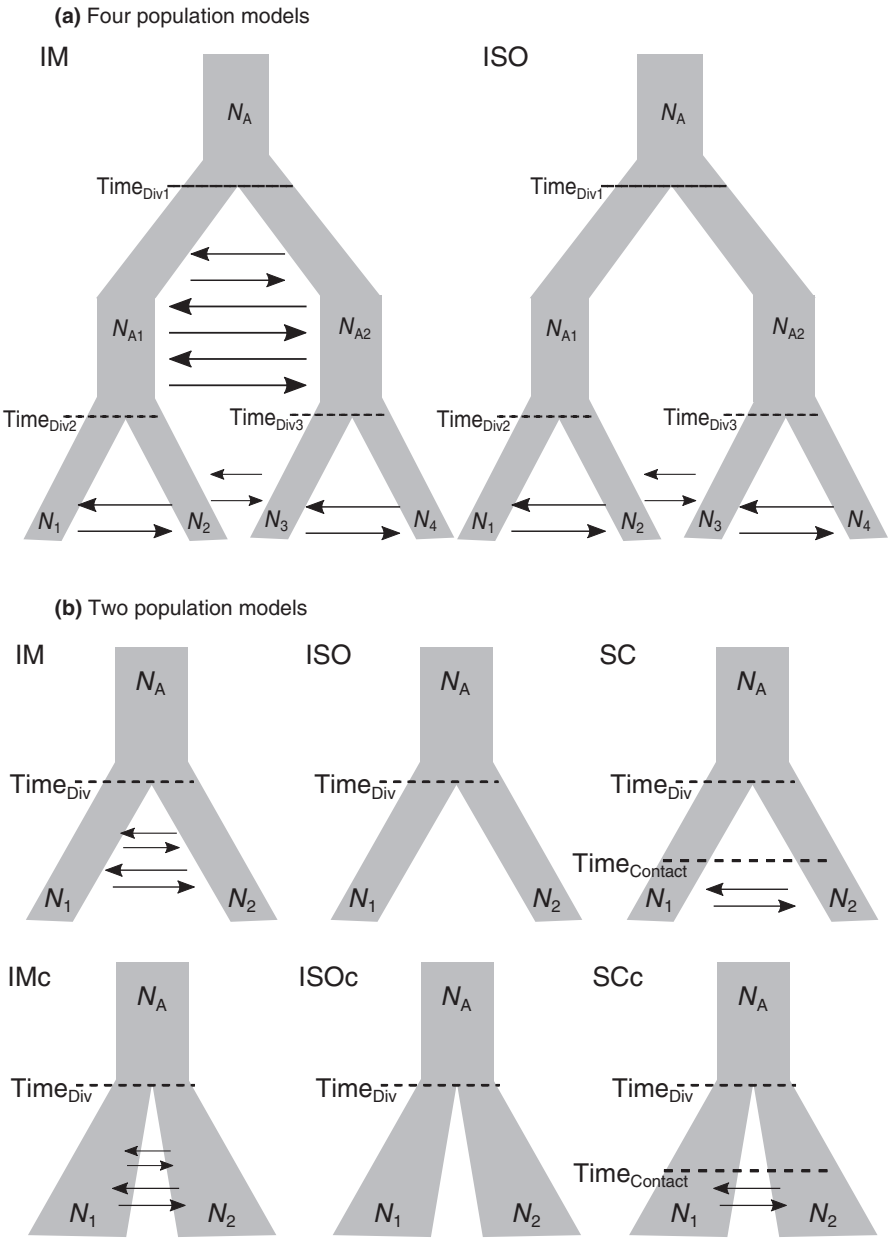


FIGURE 2 Competing phylogeographic models tested for both (a) the four populations models to understand diversification across the Mississippi River and (b) two population models tested for both sister lineage pairs (*N* = effective population sizes, $Time_{DIV}$ = divergence times, $Time_{CONTACT}$ = time of secondary contact, arrows represent gene flow)

with 70 steps, a chain length of 50,000 with burnin percent set to 50%, and a preburnin run of 12,000 iterations. We included three individuals per potential species in all models (for a total of 12 in-group samples; see Supporting Information for individual samples used in this analysis) plus the out-group taxon *P. bairdi* so that the single *P. guttatus* species model could be evaluated. Convergence was assessed by examining log plots in TRACER version 1.7.1 (Rambaut, Suchard, Xie, & Drummond, 2014) ensuring that all ESS values >200.

2.6 | Phylogeographic model selection and historical demography

We tested two four-population and six two-population demographic models (for both species pairs) using FASTSIMCOAL2 (fsc2; Excoffier, Dupanloup, Huerta-Sánchez, Sousa, & Foll, 2013). This method uses coalescent simulations to approximate an expected site frequency spectrum (SFS) and a composite likelihood approach for parameter optimizations. We constructed the two four-population models to test how divergence had occurred across the Mississippi River (Figure 2), and therefore only differed in whether gene flow occurred between the two ancestral lineages. In these models, we fixed the topology of the four populations based on the results from SNAPP and we fixed the initial divergence time within this group and the divergence between *P. emoryi* and *P. slowinskii* at 2.97 mya and 590 kya, respectively (Chen, Lemmon, Lemmon, Pyron, & Burbrink, 2017). These models also allowed for recent gene flow between adjacent populations and we used a mutation rate of 7.26×10^{-9} (Harrington, Hollingsworth, Higham, & Reeder, 2017) assuming a 3-year generation time (estimate for *P. emoryi*; Ernst & Ernst, 2003). The two-population models differ only in the parameterization of gene flow between the ancestral lineages (see Figure 2a), such that these models are simplifications of isolation with migration (Figure 2a—IM) or, alternatively, in strict allopatry across the Mississippi River (Figure 2a—ISO). We also implemented a total of six two population models, which differed in the parameterization of gene flow, the timing of gene flow, and population size changes in daughter lineages (Figure 2b). These models are simplifications of the actual evolutionary histories of these groups, but were useful for distinguishing between divergence across ecological gradients (divergence with continual gene flow; Figure 2b—IM and IMc), divergence in allopatry across biogeographic barriers (Figure 2b—ISO and ISOc), and divergence in Pleistocene refugia (Figure 2b—SC and SCc). We used easySFS (https://github.com/isaacovercast/easySFS), a wrapper around dadi (Gutenkunst, Hernandez, Williamson, & Bustamante, 2009), to generate the SFS for each analyses. We chose to project the full data set down to a smaller number of samples per population to average over missing data thereby maximizing the number of segregating sites in the SFS, while randomly selecting one SNP per locus to reduce the possibility of linkage. In the four population models we used 12 *P. emoryi*, 17 *P. guttatus* west, 26 *P. guttatus* east, and three *P. slowinskii* diploid samples, in the *P. emoryi—slowinskii* demographic models, we used 11 and two diploid samples, respectively, and in the *P. guttatus* east—west models we used 24 and 14 diploid samples, respectively. While the number of individuals per lineage included in some of these analysis is small, theoretical work has demonstrated the accuracy of reconstructing demographic size change using the SFS is based largely on the number of sampled segregating sites and is not dependent on the dimension of the SFS which is associated with the number of sampled individuals (Terhorst & Song, 2015). Therefore, sampling more individuals may increase the number of segregating sites in phylogeographic studies, yet at a fixed number of segregating sites increasing the number of individuals does not improve demographic reconstructions. Input files for all models are available on Dryad (https://doi.org/10.5061/dryad.k0p2ngf4m).

3 | RESULTS

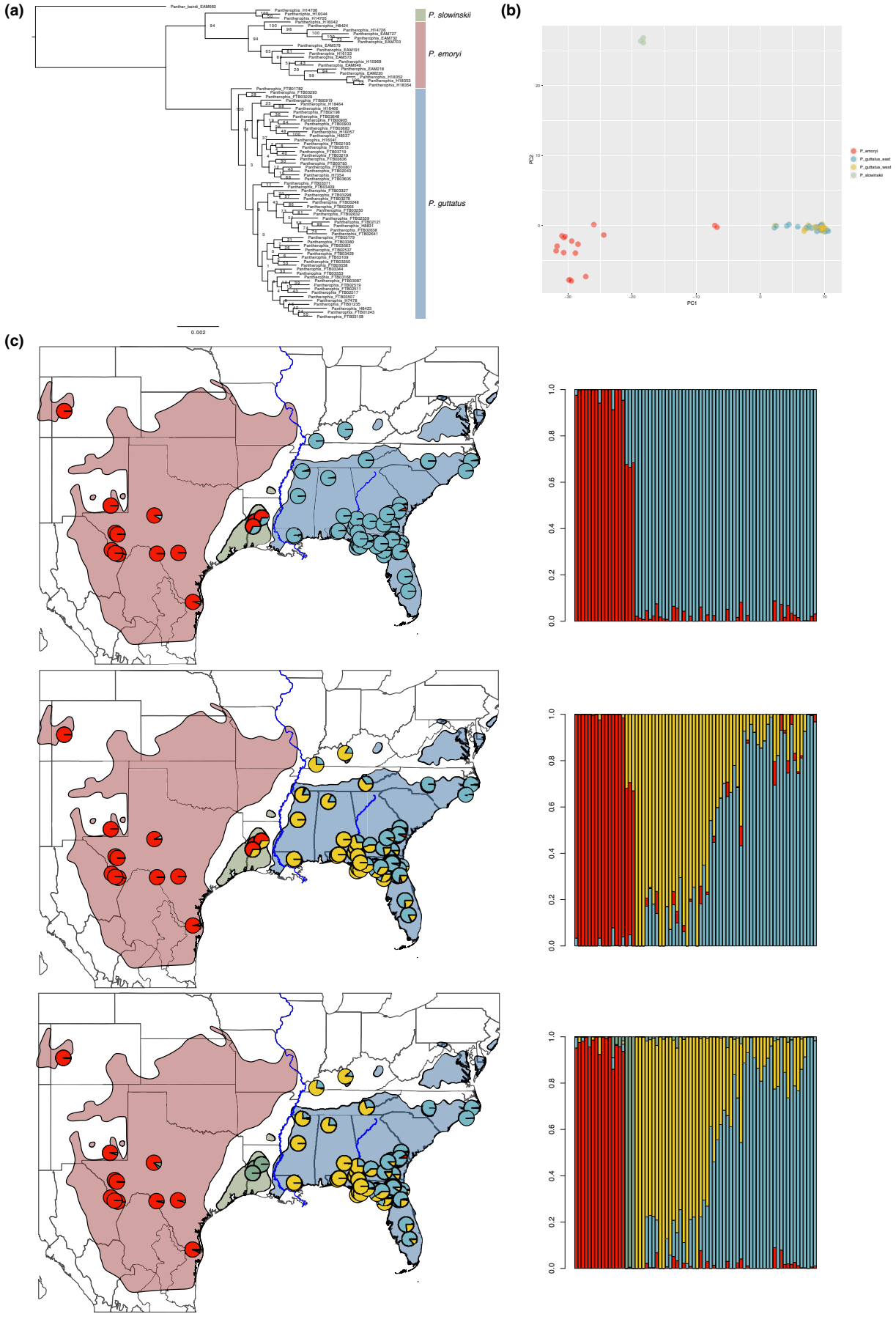
3.1 | Sequencing and selection analyses

We sequenced 80 individuals resulting in 159 million sequence reads, with 1.99 million reads per sample on average (± 1.26 million reads). The final data set contained 4,256 loci and 4,010 unlinked SNPs. Our BayeScan analyses reached stationarity after 80,000 iterations and indicated that no SNPs had significantly elevated F_{ST} values, suggesting that the sequenced loci are not under strong selection, we therefore retained all loci for further analyses.

3.2 | Population structure

The concatenated maximum likelihood gene tree inferred a topology that is in agreement with the currently recognized species within the *Pantherophis guttatus* complex, showing three major lineages corresponding to the three species (Figure 3). Both DAPC and sNMF find similar support for between two and four genetic clusters, largely in agreement on the individual sample membership to these populations. In these analyses, at $K = 2$ the populations were divided by the Mississippi River, at $K = 3$ DAPC corresponded to the current taxonomy (*P. guttatus*, *P. slowinskii*, and *P. emoryi*) whereas sNMF found two populations within *P. guttatus* and a single cluster west of the Mississippi River, lastly at $K = 4$ these analyses found *P. slowinskii*, *P. emoryi*, and two populations within

FIGURE 3 Population structure and historical relationships within the *Pantherophis guttatus* complex. (a) Concatenated gene-tree inferred in Raxml, with bootstrap support values indicated at each node; (b) Principle components analysis of rad-seq data (note that the two *Pantherophis emoryi* samples in the middle of the plot are missing a large amount of data); (c) population assignments based on sNMF analysis with *K* values ranging from 2 to 4. These inferred clusters are plotted on the geographic distributions of the currently defined species (*P. emoryi* in red, *Pantherophis slowinskii* in green, and *P. guttatus* in blue) [Colour figure can be viewed at wileyonlinelibrary.com]


(a)
P. slowinskii
P. emoryi
P. guttatus
0.002
(b)
PC1
PC2
P_emoryi
P_guttatus_east
P_guttatus_west
P_slowinskii
(c)

P. guttatus (Figure 3). The PCA plot shows three distinct clusters, where the two populations within *P. guttatus* are largely overlapping in PC space (Figure 3).

3.3 | Ecological niche models

After thinning the locality records from VertNet, a total of 174 *P. emoryi*, eight *P. slowinskii*, 81 eastern *P. guttatus*, and 24 western *P. guttatus* were retained for ENMs (locality data and museum specimen numbers can be found in Dryad: https://doi.org/10.5061/dryad.k0p2ngf4m). Comparisons of niche equivalency demonstrate that sister population pairs are occupying distinct environments (*P. emoryi*—*P. slowinskii* comparison: *D* statistic = 0.02, *I* statistic = 0.07, *p*-values = 0; *P. guttatus* eastern–western comparison: *D* statistic = 0.26, *I* statistic = 0.39, *p*-values = 0). However, the nonsister comparison between *P. slowinskii* and western *P. guttatus* that are separated by the Mississippi River, suggest that the two are distributed in equivalent habitats (*D* statistic = 0.41, *p*-value = .30; *I* statistic = 0.55, *p*-value = .23).

Best-fit combinations of feature class and regularization multipliers differed among the four lineages (see Supporting Information for AIC scores). In several cases AIC values could not distinguish between multiple combinations of feature classes, in which case we generated ENMs using all best fit models (e.g., for *P. slowinskii* both combinations of LQHP and LQHPT feature classes had equal support). Ecological niche models performed well, with AUC values equal to or above 0.9 (*P. emoryi* = 0.9, *P. guttatus* east = 0.96, *P. guttatus* west = 0.97, *P. slowinskii* = 0.94). Current ENMs predicted the geographic distributions of each taxon (Supporting Information). These current habitat suitability maps also suggest that *P. slowinskii* had a much larger region of suitable habitat than its current realized niche; the Mississippi River and/or competition from *P. guttatus* probably restrict its geographic distribution. This is not surprising as niche comparisons between the western population of *P. guttatus* and *P. slowinskii* suggest that the two taxa are occupying similar ecological niches. Stacked suitability maps suggest that core regions of stability for *P. emoryi* and *P. slowinskii* have largely been allopatric (Figure 4), although with some limited regions of low suitability for both taxa in current day eastern Texas. Stability models for eastern and western populations of *P. guttatus* suggest that both taxa had large regions of suitability from the late Pleistocene to present day along the Gulf Coast (Figure 4), indicating that the two may have been in contact since the LGM. We present suitability models generated under all best fit combinations of feature classes in the Supporting Information; core regions of stability do not differ greatly among these models.

3.4 | Species delimitation and historical demography

Species delimitation analyses using BFD* overwhelmingly supported a model in which all four lineages are distinct taxa (Table 1). The second-best model, although with considerably less support (BF = 82.6), was represented by the currently recognized taxonomy of three species. SNAPP analyses demonstrate that the oldest divergence within this complex is across the Mississippi River, and that the populations on either side of this barrier are sister to each other (Figure 3).

Demographic model selection with all four populations favored an initial population divergence across the Mississippi River with no gene flow between these two ancestral lineages (Table 2), demonstrating the importance of this river in promoting divergence. The best-supported model for divergence between eastern and western populations of *P. guttatus* was an isolation-with-migration model without population size changes through time, followed by a model of secondary contact. This suggested continual or repeated bouts of migration between these lineages post-divergence (Table 2), which could be attributed to divergence across environmental gradients or in allopatric refugia. There is considerable uncertainty in which demographic model best fits the evolutionary history between *P. emoryi* and *P. slowinskii*; however, all three of the best-fit models included gene flow (Table 2). These FSC2 analyses demonstrated divergence across environmental gradients or in allopatric refugia, possibly with population expansion. Estimated mean divergence times under the best-fit model between eastern and western populations of *P. guttatus* was 448 kya (92–912 kya), and 289 kya (range 17–763 kya) between *P. emoryi* and *P. slowinskii*. Rates of migration between sister populations were low and rates were probably asymmetric between pairs (Table 3).

4 | DISCUSSION

The southeastern Nearctic is a biodiversity hotspot characterized by rapid diversification and high concentration of endemic lineages (Noss et al., 2015). This diversity could have resulted from one of several evolutionary or ecological drivers of diversification, including environmental transitions, historical climate change, and the potential for allopatric divergence across numerous biogeographic barriers (Soltis et al., 2006). We explicitly tested how these hypothesized factors influenced population connectivity and lineage divergence among the widespread cornsnake complex (Figure 1). By combining phylogeographic model selection with environmental niche comparisons and ENMs, we demonstrate the importance of all of these drivers of lineage-formation in producing cryptic diversity within a species complex. Speciation within the *Pantherophis guttatus* complex initially occurred via vicariance with a lack of gene flow across the Mississippi River (Table 2). More recent divergence events, between *P. emoryi* and *P. slowinskii* and between eastern and western populations of *P. guttatus*, occurred in the late Pleistocene probably with continuous or frequent episodes of migration after initial divergence (Table 2). These sister lineage pairs also diversified into significantly different environmental niches that occurred in mostly allopatric regions of habitat stability from the LGM to present day (Figure 4).

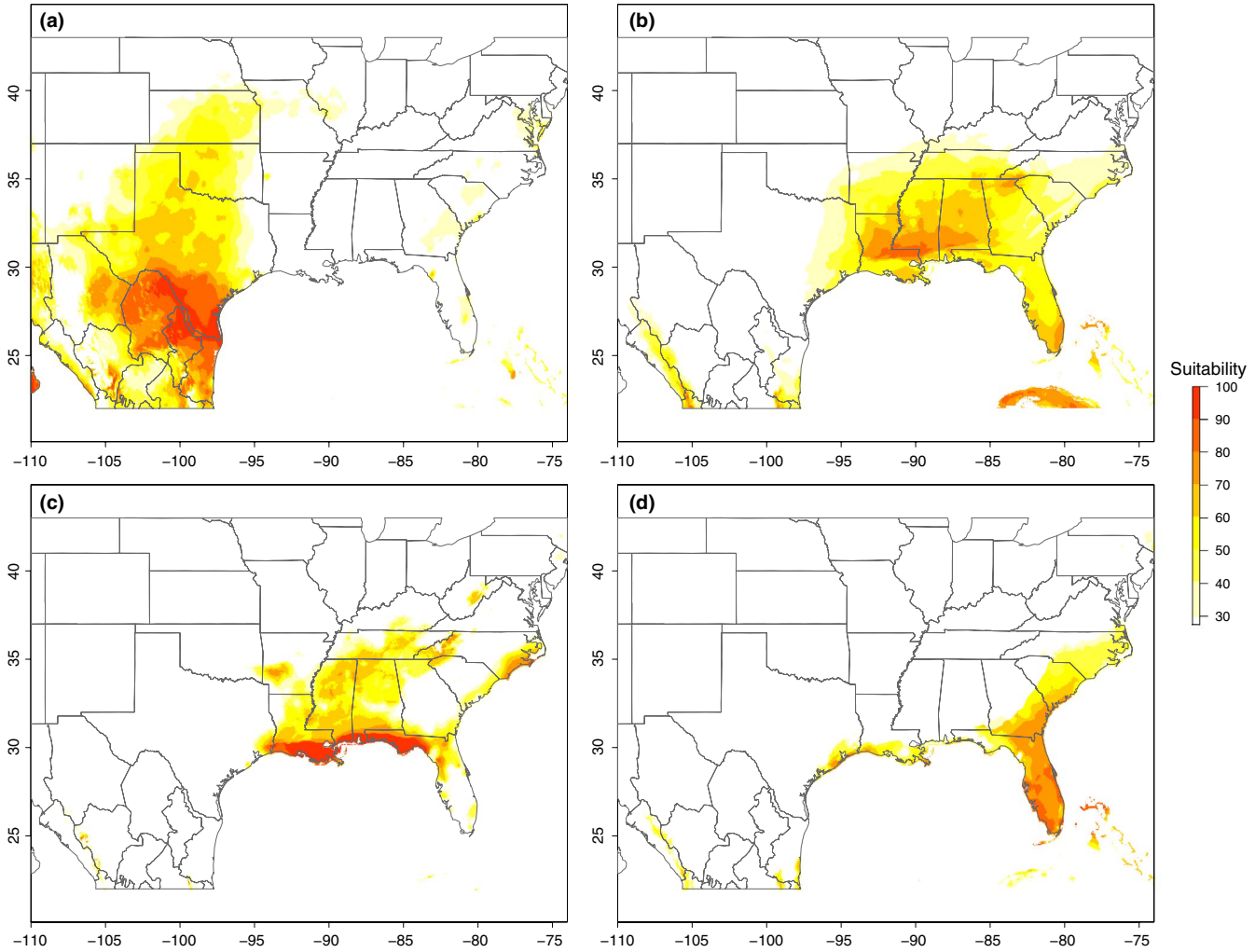


FIGURE 4 Regions of habitat suitability through time. These projections are the result of stacked and averaged ENMs from the Last Glacial Maximum, the Mid-Holocene, and current climatic conditions. Warmer colours indicate greater suitability scores. (a) *Pantherophis emoryi*; (b) *Pantherophis slowinskii*; (c) western *Pantherophis guttatus*; (d) eastern *P. guttatus*. These projections used LQHPT feature classes (with the exception of *P. emoryi* where the bet fit model was LQ) [Colour figure can be viewed at wileyonlinelibrary.com]

4.1 | Diversification in the southeastern Nearctic

Proposed biogeographic barriers in the southeastern Nearctic include major rivers, for example the Mississippi and Apalachicola rivers (Soltis et al., 2006). Within the North American cornsnakes, initial divergences occurred across the Mississippi River (Figure 3) approximately ~2.97 mya (Chen et al., 2017), a time when the Mississippi River would have had a discharge 6–8 times greater than current rates (Cox, Lumsden, & Van Arsdale, 2014). The demographic models tested here suggest that this initial speciation event occurred without gene flow across the Mississippi River. No gene flow, despite suitable habitat on either side of the barrier, suggests that the Mississippi River has been an important feature of the landscape in promoting divergence, potentially via allopatry. Numerous studies have identified lineage formation across this barrier in taxonomic groups as varied as terrestrial (Brant & Ortí, 2003; Burbrink, Fontanella, Pyron, Guiher, & Jimenez, 2008; Burbrink et al., 2000; Leaché & Reeder, 2002; Myers et al., 2017) and aquatic vertebrates (Brandley, Guiher, Pyron, Winne, & Burbrink, 2010; Near, Page, & Mayden, 2001), plants (Al-Rabab'ah & Williams, 2002; Zellmer, Hanes, Hird, & Carstens, 2012), and invertebrates (Katz, Taylor, & Davis, 2018; Satler & Carstens, 2017). The timing of lineage divergence across this important barrier is unknown for many taxa, however estimates from other studies range from the late Miocene (e.g., Lemmon, Lemmon, Collins, Lee-Yaw, & Cannatella, 2007) to the Pleistocene (e.g., Howes, Lindsay, & Lougheed, 2006). This suggests that the Mississippi River is a long-standing, important biogeographic barrier that has promoted the origins and maintenance of biodiversity for extant taxa across the southeastern Nearctic. To better understand community-wide patterns of dispersal, divergence, and gene flow in relation to changes in the hydrology of the Mississippi River (e.g., Cox et al., 2014), future comparative analyses should focus on both the number and timing of population divergence events and shared historical demography across entire assemblages of taxa (Myers et al., 2017; Xue & Hickerson, 2015).

TABLE 1 Results from species delimitation models compared using BFD* in SNAPP (Bryant et al., 2012)

Model	Marginal likelihood	Bayes factor
All four lineages ([*P. emoryi*, *Pantherophis slowinskii*], [*Pantherophis guttatus* east, *P. guttatus* west])	−17,450.3	0
Current taxonomy (*P. emoryi*, *P. slowinskii*, [*P. guttatus*])	−17,491.6	82.6
Three species (*P. emoryi* + *P. slowinskii*, [*P. guttatus* east, *P. guttatus* west])	−18,450.6	2,000.6
Two species, split across the Mississippi River (*P. emoryi* + *P. slowinskii*, *P. guttatus*)	−18,500.5	2,100.4
Two species, relationships based on mtDNA gene tree (*P. emoryi*, *P. guttatus* + *P. slowinskii*)	−20,217.2	5,533.8
Single species (*P. emoryi* + *P. guttatus* + *P. slowinskii*)	−22,013	9,125.4

Note: Models are ranked in descending order by Bayes factor support; the model with the lowest Bayes factor is the best fit model suggesting that all four lineages represent distinct species.

TABLE 2 Results from demographic model selection in fastsimcoal2

	Number of parameters	Log-likelihood	AIC	ΔAIC	AIC weight
Four lineage models					
ISO	**14**	**−6,206.2**	**12,440.4**	**0**	**1**
IM	16	−6,226.6	12,485.3	44.8	1.81457E−10
Eastern and Western *Pantherophis guttatus* lineages models					
IM	**6**	**−1,744.0**	**3,500.1**	**0**	**0.605123276**
IMc	8	−1,744.0	3,504.1	4.072	0.078998763
ISO	4	−1,758.6	3,525.2	25.122	2.12164E−06
ISOc	6	−1,753.3	3,518.5	18.486	5.85679E−05
SC	7	−1,743.9	3,501.9	1.844	0.240671323
SCc	9	−1,743.1	3,504.2	4.172	0.075145948
Pantherophis emoryi and *Pantherophis slowinskii* models					
IM	**6**	**−1,764.094**	**3,540.188**	**0**	**0.461148985**
IMc	8	−1,762.796	3,541.592	1.404	0.228542267
ISO	4	−1,779.46	3,566.92	26.732	7.22872E−07
ISOc	6	−1,772.356	3,556.712	16.524	0.000119042
SC	7	−1,763.749	3,541.498	1.31	0.23954018
SCc	9	−1,762.97	3,543.94	3.752	0.070648803

Note: Best-fit models were selected based on ΔAIC scores and AIC weights, models best supported by the data are highlighted in bold.

Model abbreviations are as follows: IM, isolation with migration; ISO, isolation without migration; IMc, isolation with migration and demographic size change; ISOc, isolation without migration and demographic size change; SC, secondary contact; SCc, secondary contact with demographic size change

TABLE 3 Bootstrapped parameter estimates for each of the two lineages modeled in fastsimcoal2

	Divergence time	Ancestral N_e	Lineage 1 N_e	Lineage 2 N_e	Migration 1 → e2	Migration 2 → 1
(1) *Pantherophis emoryi* and (2) *Pantherophis slowinskii*	289 kya (17–763 kya)	89,000 (7,200–309,000)	103,000 (7300– 279,000)	270,000 (46,000–532,000)	2.5×10^{-4} (4.25×10^{-5} – 2.2×10^{-3})	0
(1) Eastern and (2) Western *P. guttatus* lineages	448 kya (92–912 kya)	79,000 (13,000–219,000)	65,000 (16,000–127,000)	283,000 (77,000–507,000)	7.2×10^{-6} (3.39×10^{-6} – 2.32×10^{-5})	1×10^{-6} (4.37×10^{-7} – 4.12×10^{-6})

Note: Parameters are listed as mean and range of each estimate. Note that the linage number is listed in the column of taxa being compared (e.g., *P. emoryi* is Lineage 1)

The turnover in habitats and environments across the southeastern Nearctic is an important determinant of population genetic structure (Burbrink & Guiher, 2015; McKelvy & Burbrink, 2017) and influences the assemblage of taxa into communities (Burbrink & Myers, 2015). Here we find significant ecological differentiation between both sister lineage pairs, which likely lead to ecologically-mediated population differentiation and speciation within this complex. This is an important distinction to make; by simply visualizing lineage membership on a map it may appear that divergence between eastern and western populations of *P. guttatus* occurred across or within the vicinity of the Apalachicola River (Figure 3). However, by combining niche comparisons with demographic modeling, we can reject a scenario of divergence across this river in favour of continual gene flow with divergence across climatic gradients or, albeit with lesser support, refugial isolation with secondary contact (Figure 4 and Table 2). Similarly, the *P. emoryi* and *P. slowinskii* species pair, which are distributed in different habitats (e.g., desert scrub and prairie vs. longleaf pine), have probably diverged due to adaptation to local environmental conditions and do not share regions of habitat stability through time (Figure 4). These niche comparisons and hindcast ENMs supported by our estimates that gene flow occurred throughout the history of divergence between *P. emoryi* and *P. slowinskii*, possibly with population size changes (Table 2). Although we find a strong signal of environmental niche differentiation, we find no SNPs that deviate from expectations of neutral population divergence. Future analyses could elucidate the role of selection in maintaining population differentiation by using targeted sequencing of functional loci (Christmas, Biffin, Breed, & Lowe, 2016; Namroud, Bealieu, Juge, Laroche, & Bousquet, 2008) or whole genome sequence data to explore the role of structural genomic variation in relation to local adaptation to different climatic regimes (Lucek, Gompert, & Nosil, 2019; Rinker, Specian, Zhao, & Gibbons, 2019).

4.2 | Habitat stability and Pleistocene refugia

Habitat stability models illustrate largely nonoverlapping regions of core suitable habitat since the LGM for *P. emoryi* and *P. slowinskii*. These regions have likely served as allopatric refugia during the late Pleistocene, further reinforcing population differentiation and divergence. A potential caveat here is the smaller sample size used for *P. slowinskii*, which may influence the ability to predict suitable habitat. However, previous studies focused on the effects of small sample size use in Maxent have demonstrated that such models can identify regions with similar environmental conditions to where a taxon is known to occur, but should not be interpreted as absolutely predicting species ranges (Pearson, Raxworthy, Nakamura, & Townsend Peterson, 2007). We similarly interpret our resulting ENMs as regions of similar, suitable habitats (i.e., potential refugia) and not absolute range limits. Additionally, many of the same regions of stability that we have identified for the *P. guttatus* species complex has been found in other, largely codistributed taxa across the eastern Nearctic (Barrow et al., 2017; Walker, Stockman, Marek, & Bond, 2009; Waltari et al., 2007).

The comparison of stable regions shared by the eastern and western populations of *P. guttatus* suggests potential overlap in suitable climate since the LGM within the panhandle region of Florida (Figure 4). Given the potential stablility within this region for the eastern and western *P. guttatus* lineages and an extensive zone of contact in population clustering analyses in present day Florida and Georgia (Figure 3), it is possible that these lineages have maintained continuous distributions through time, a scenario supported by the best-fit demographic model of isolation-with-migration. The future generation of whole genome sequence data will allow for tests of multiple bouts of isolation and migration throughout the Pleistocene versus parapatric divergence by taking advantage of linkage disequilibrium block structure (e.g., Sousa & Hey, 2013).

Because each of the sister lineages have different environmental niches, we chose to generate ENMs and hindcast models separately for each taxon. This is an important, but perhaps overlooked consideration when inferring suitable refugia during the LGM. Unique responses to climate change have been shown to occur between intraspecific groups, which can result in differences in adaptation or dispersal between populations within a species (Maguire, Shinneman, Potter, & Hipkins, 2018; Pearman, D'Amen, Graham, Thuiller, & Zimmermann, 2010). As expected, accounting for phylogeographic structure when forecasting species' distributions onto models of future climate change has demonstrated that predicted distributions are smaller than when not considering this structure (Valladares et al., 2014). Given that differences exist in the environmental space across the *P. guttatus* species complex, lumping occurrences from sister pairs to make predictions about current distributions or regions of stability may return unreliable results (e.g., Smith, Godsoe, Rodríguez-Sánchez, Wang, & Warren, 2018). For example, areas of stability would probably be overestimated, resulting in more widespread and continuous regions of stability. It is likely an unjustifiable assumption to group cryptic lineages into single ENMs without first testing for niche differentiation, an important consideration for future studies.

4.3 | Taxonomic considerations

The study group has historically been treated as a single species with five subspecies largely differentiated by color and pattern (*P. guttatus guttatus*, *P. guttatus rosaceus*, *P. guttatus emoryi*, *P. guttatus intermontanus*, and *P. guttatus meahllmorum*; Schultz, 1996). However, these subspecies were subjectively defined (e.g., defining subspecies as having >44.5 or <44.5 body blotches; Smith, Chiszar, Staley, & Tepedelen, 1994) and not supported by gene tree analyses of mtDNA data (Burbrink, 2002), nor detectable with genomic data (this study). The most recent taxonomic treatments recognize three allopatrically distributed cryptic species that form monophyletic lineages; *P. guttatus*, *P. emoryi*, and *P. slowinskii* (Figure 1; Burbrink, 2002); note that

P. slowinskii had previously been classified as being part of the wide ranging *P. guttatus guttatus* (Schultz, 1996). Our analyses, given our sampling of individuals and genomic data, support the recognition of these three species and further suggest that species diversity is underestimated within *P. guttatus* (Table 1; Figure 3). These distinctions are supported by multiple lines of evidence: these lineages are genetically distinct (Figure 3), supported as separate species using BFD* (Table 1), have low rates of migration between sister taxa (Table 3), and each sister-pair has evolved distinct ecological niches.

Using coalescent species delimitation methods, we support the current taxonomy within the *P. guttatus* complex, and suggest that while there are potentially further cryptic species within this group, we refrain from formally recognizing additional species pending integrative taxonomic analyses. A number of recent papers have suggested that multi-species coalescent (MSC) models used for species delimitation may instead be delimiting population substructure, (e.g., Jackson, Carstens, Morales, & O'Meara, 2016; Leaché, Zhu, Rannala, & Yang, 2018; Sukumaran & Knowles, 2017), though how populations differ from species using the MSC is often not defined a priori. Future species delimitation analyses could focus on the zone of contact between the eastern and western *P. guttatus* lineages, assessing the frequency of F_1 hybrids (Leaché et al., 2018), the width of hybrid clines across this contact zone, and implement additional species delimitation models that incorporate gene flow and spatial data. Previous analyses of morphological characters have shown that scale counts are not discreet but instead continuous between the currently named taxa (Burbrink, 2002), whether this is the case for the eastern and western populations of *P. guttatus* is not yet known. Morphological variation among these recently evolved taxa might be subtle (e.g., differences in head shape variation; Ruane, 2015) and may be detected using high-resolution morphological data from specimen-based studies (Chaplin, Sumner, Hipsley, & Melville, 2019).

In conclusion, here we have demonstrated that biogeographic barriers, Pleistocene refugia, and niche divergence have interacted to drive population differentiation and produce cryptic diversity within the *P. guttatus* species complex. These results support the importance of the Mississippi River as an essential biogeographic barrier in North America that has generated biodiversity via strict allopatric divergence. Furthermore, our analyses suggest that more recent divergence events in the southeastern Nearctic have probably occurred due to niche divergence and vicariance between allopatric refugia during the Pleistocene. This study illustrates the significance of environmental gradients in driving ecologically mediated divergence across much of eastern North America. Lastly, our analyses support the recognition of three cryptic species within this complex and suggest there may also be additional taxa in need of description.

ACKNOWLEDGEMENTS

We thank Louisiana State University Museum of Natural Sciences (J. Boundy, D. Dittman, R. Brumfeld, F. Sheldon) for donating tissues samples for this work. S. Ruane, R. A. Pyron, and T. Guiher also collected samples used in this study. E. Chen was a huge help in the laboratory and is a DNA extracting machine. Additionally, we would like to acknowledge the following institutions for which locality data from their specimens were used in generating ecological niche models: Cornell University Museum of Vertebrates, Fort Hays Sternberg Museum of Natural History, Georgia Southern University, Natural History Museum of Los Angeles County, Museum of Southwestern Biology, Sam Noble Oklahoma Museum of Natural History, Ohio State University, San Diego Natural History Museum, Texas A&M University Biodiversity Research and Teaching Collections, University of British Columbia Beaty Biodiversity Museum, University of Colorado Museum of Natural History, Florida Museum of Natural History, National Museum of Natural History – Smithsonian Institution, University of Texas at El Paso Biodiversity Collections, Yale Peabody Museum, and Louisiana State University Museum of Natural Science. We also would like to thank K. de Queiroz and K. O'Connell for comments on this manuscript, as well as D. Rivera for thoughtful discussion on ENMs. Lastly, we thank the University of Wisconsin Biotechnology Center DNA Sequencing Facility for providing GBS services. EAM was funded by the Gerstner Scholar/Theodore Roosevelt and Peter Buck/Walter Rathbone Bacon fellowships.

AUTHOR CONTRIBUTIONS

E.A.M., and F.T.B. conceived the study. E.A.M. collected samples, performed laboratory work and analyses. A.D.M. collected samples and helped design figures. F.T.B. provided funding. All authors contributed to the writing of the manuscript.

DATA AVAILABILITY STATEMENT

The genetic data generated during the current study are available from the NCBI Sequence Read Archive (BioProject ID: PRJNA596278); the assembled rad data and fsc2 models used in this study are available on Dryad (https://doi.org/10.5061/dryad.k0p2ngf4m). All sampling localities used for ENMs as well as all output ascii files are available on Dryad (https://doi.org/10.5061/dryad.k0p2ngf4m).

ORCID

Edward A. Myers https://orcid.org/0000-0001-5369-6329
Alexander D. McKelvy https://orcid.org/0000-0002-3042-7465
Frank T. Burbrink https://orcid.org/0000-0001-6687-8332

REFERENCES

Aiello-Lammens, M. E., Boria, R. A., Radosavljevic, A., Vilela, B., & Anderson, R. P. (2015). spThin: An R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography*, *38*(5), 541-545. https://doi.org/10.1111/ecog.01132

Aleixo, A. (2004). Historical diversification of a terra-firme forest bird superspecies: A phylogeographic perspective on the role of different hypotheses of Amazonian diversification. *Evolution*, *58*(6), 1303–1317.

Al-Rabab'ah, M. A., & Williams, C. G. (2002). Population dynamics of *Pinus taeda* L. based on nuclear microsatellites. *Forest Ecology and Management*, *163*(1–3), 263–271.

Bailey, R. G. (1995). *Description of the ecoregions of the United States* (2nd ed). Fort Collins, CO: Rocky Mountain Research Station.

Barrow, L. N., Lemmon, A. R., & Lemmon, E. M. (2018). Targeted sampling and target capture: Assessing phylogeographic concordance with genome-wide data. *Systematic Biology*, *67*(6), 979–996.

Barrow, L. N., Soto-Centeno, J. A., Warwick, A. R., Lemmon, A. R., & Moriarty Lemmon, E. (2017). Evaluating hypotheses of expansion from refugia through comparative phylogeography of south-eastern Coastal Plain amphibians. *Journal of Biogeography*, *44*(12), 2692–2705.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., … Drummond, A. J. (2014). BEAST 2: A software platform for Bayesian evolutionary analysis. *PLOS Computational Biology*, *10*(4), 1-6. https://doi.org/10.1371/journal.pcbi.1003537

Bradburd, G. S., Ralph, P. L., & Coop, G. M. (2013). Disentangling the effects of geographic and ecological isolation on genetic differentiation. *Evolution*, *67*(11), 3258–3273.

Brandley, M. C., Guiher, T. J., Pyron, R. A., Winne, C. T., & Burbrink, F. T. (2010). Does dispersal across an aquatic geographic barrier obscure phylogeographic structure in the diamond-backed watersnake (*Nerodia rhombifer*)? *Molecular Phylogenetics and Evolution*, *57*(2), 552–560.

Brant, S. V., & Ortí, G. (2003). Phylogeography of the Northern short-tailed shrew, *Blarina brevicauda* (Insectivora: Soricidae): Past fragmentation and postglacial recolonization. *Molecular Ecology*, *12*(6), 1435–1449. https://doi.org/10.1046/j.1365-294X.2003.01789.x

Broennimann, O., Fitzpatrick, M. C., Pearman, P. B., Petitpierre, B., Pellissier, L., Yoccoz, N. G., … Guisan, A. (2012). Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecology and Biogeography*, *21*(4), 481–497. https://doi.org/10.1111/j.1466-8238.2011.00698.x

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., & RoyChoudhury, A. (2012). Inferring species trees directly from Biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, *9*(8), 1917–1932. https://doi.org/10.1093/molbev/mss086

Burbrink, F. T. (2002). Phylogeographic analysis of the cornsnake (*Elaphe guttata*) complex as inferred from maximum likelihood and Bayesian analyses. *Molecular Phylogenetics and Evolution*, *25*(3), 465–476. https://doi.org/10.1016/S1055-7903(02)00306-8

Burbrink, F. T., Chan, Y. L., Myers, E. A., Ruane, S., Smith, B. T., & Hickerson, M. J. (2016). Asynchronous demographic responses to Pleistocene climate change in Eastern Nearctic vertebrates. *Ecology Letters*, *19*(12), 1457–1467. https://doi.org/10.1111/ele.12695

Burbrink, F. T., Fontanella, F., Pyron, R. A., Guiher, T. J., & Jimenez, C. (2008). Phylogeography across a continent: The evolutionary and demographic history of the North American racer (Serpentes : Colubridae : Coluber constrictor). *Molecular Phylogenetics and Evolution*, *47*(1), 274–288. https://doi.org/10.1016/j.ympev.2007.10.020

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

Burbrink, F. T., Lawson, R., & Slowinski, J. B. (2000). Mitochondrial DNA phylogeography of the polytypic North American rat snake (*Elaphe obsoleta*): A critique of the subspecies concept. *Evolution; International Journal of Organic Evolution*, *54*(6), 2107–2118. https://doi.org/10.1554/0014-3820(2000)054

Burbrink, F. T., & Myers, E. A. (2015). Both traits and phylogenetic history influence community structure in snakes over steep environmental gradients. *Ecography*, 1036-1048.

Carnaval, A. C., Hickerson, M. J., Haddad, C. F. B., Rodrigues, M. T., & Moritz, C. (2009). Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Science*, *323*(5915), 785–789. https://doi.org/10.1126/science.1166955

Carstens, B. C., Brennan, R. S., Chua, V., Duffie, C. V., Harvey, M. G., Koch, R. A., … Satler, J. D. (2013). Model selection as a tool for phylogeographic inference: An example from the willow *Salix melanopsis*. *Molecular Ecology*, *22*(15), 4014–4028.

Carstens, B. C., Morales, A. E., Jackson, N. D., & O'Meara, B. C. (2017). Objective choice of phylogeographic models. *Molecular Phylogenetics and Evolution*, *116*, 136-140. https://doi.org/10.1016/j.ympev.2017.08.018

Carstens, B. C., & Richards, C. L. (2007). Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution*, *61*(6), 1439–1454. https://doi.org/10.1111/j.1558-5646.2007.00117.x

Chaplin, K., Sumner, J., Hipsley, C., & Melville, J. (2019). An integrative approach using phylogenomics and high-resolution X-ray computed tomography (CT) for species delimitation in cryptic taxa. *Systematic Biology*. syz048. https://doi.org/10.1093/sysbio/syz048

Charles, K. L., Bell, R. C., Blackburn, D. C., Burger, M., Fujita, M. K., Gvoždík, V., … Portik, D. M. (2018). Sky, sea, and forest islands: Diversification in the African leaf-folding frog *Afrixalus paradorsalis* (Anura: Hyperoliidae) of the Lower Guineo-Congolian rain forest. *Journal of Biogeography*, *45*(8), 1781–1794. https://doi.org/10.1111/jbi.13365

Chen, X., Lemmon, A. R., Lemmon, E. M., Pyron, R. A., & Burbrink, F. T. (2017). Using phylogenomics to understand the link between biogeographic origins and regional diversification in ratsnakes. *Molecular Phylogenetics and Evolution*, *111*, 206–218.

Christmas, M. J., Biffin, E., Breed, M. F., & Lowe, A. J. (2016). Finding needles in a genomic haystack: Targeted capture identifies clear signatures of selection in a nonmodel plant species. *Molecular Ecology*, *25*(17), 4216–4233.

Cox, R. T., Lumsden, D. N., & Van Arsdale, R. B. (2014). Possible relict meanders of the Pliocene Mississippi River and their implications. *The Journal of Geology*, *122*(5), 609–622.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Eaton, D. A. R., & Overcast, I. (2020). ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics*, btz966. https://doi.org/10.1093/bioinformatics/btz966

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, *6*(5), 1-10. https://doi.org/10.1371/journal.pone.0019379

Ernst, C. H., & Ernst, E. M. (2003). *Snakes of the United States and Canada*. Washington, DC: Smithsonian Books.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLOS Genetics*, *9*(10), 1-17. https://doi.org/10.1371/journal.pgen.1003905

Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, *180*(2), 977–993. https://doi.org/10.1534/genetics.108.092221

Folt, B., Bauder, J., Spear, S., Stevenson, D., Hoffman, M., Oaks, J. R., … Guyer, C. (2019). Taxonomic and conservation implications of population genetic admixture, mito-nuclear discordance, and male-biased dispersal of a large endangered snake, *Drymarchon couperi*. *PLoS ONE*, *14*(3), e0214439.

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, *6*(8), 925–929. https://doi.org/10.1111/2041-210X.12382

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics*, *196*(4), 973–983. https://doi.org/10.1534/genetics.113.160572

Gosselin, T. (2017). *radiator: RADseq data exploration. Manipulation and Visualization Using R*. doi: https://doi.org/10.5281/zenodo.154432

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, *5*(10), e1000695. https://doi.org/10.1371/journal.pgen.1000695

Harrington, S. M., Hollingsworth, B. D., Higham, T. E., & Reeder, T. W. (2017). Pleistocene climatic fluctuations drive isolation and secondary contact in the red diamond rattlesnake (*Crotalus ruber*) in Baja California. *Journal of Biogeography*, *45*(1), 64-75. https://doi.org/10.1111/jbi.13114

Hickerson, M. J., Stahl, E. A., & Lessios, H. A. (2006). Test for simultaneous divergence using approximate Bayesian computation. *Evolution*, *60*(12), 2435–2453.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*(15), 1965–1978. https://doi.org/10.1002/joc.1276

Howes, B. J., Lindsay, B., & Lougheed, S. C. (2006). Range-wide phylogeography of a temperate lizard, the five-lined skink (*Eumeces fasciatus*). *Molecular Phylogenetics and Evolution*, *40*(1), 183–194.

Ilves, K. L., Huang, W. E. N., Wares, J. P., & Hickerson, M. J. (2010). Colonization and/or mitochondrial selective sweeps across the North Atlantic intertidal assemblage revealed by multi-taxa approximate Bayesian computation. *Molecular Ecology*, *19*(20), 4505–4519. https://doi.org/10.1111/j.1365-294X.2010.04790.x

Jackson, N. D., Carstens, B. C., Morales, A. E., & O'Meara, B. C. (2016). Species delimitation with gene flow. *Systematic Biology*, *66*(5), 799-812. https://doi.org/10.1093/sysbio/syw117

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics*, *11*(1), 94. https://doi.org/10.1186/1471-2156-11-94

Katz, A. D., Taylor, S. J., & Davis, M. A. (2018). At the confluence of vicariance and dispersal: Phylogeography of cavernicolous springtails (Collembola: Arrhopalitidae, Tomoceridae) codistributed across a geologically complex karst landscape in Illinois and Missouri. *Ecology and Evolution*, *8*(20), 10306–10325.

Leaché, A. D., Fujita, M. K., Minin, V. N., & Bouckaert, R. R. (2014). Species delimitation using genome-wide SNP data. *Systematic Biology*, *63*(4), 534–542.

Leaché, A. D., & Reeder, T. W. (2002). Molecular systematics of the Eastern Fence Lizard (*Sceloporus undulatus*): A comparison of Parsimony, Likelihood, and Bayesian approaches. *Systematic Biology*, *51*(1), 44–68. https://doi.org/10.1080/106351502753475871

Leaché, A. D., Zhu, T., Rannala, B., & Yang, Z. (2018). The spectre of too many species. *Systematic Biology*, *68*(1), 168–181.

Lemmon, E. M., Lemmon, A. R., Collins, J. T., Lee-Yaw, J. A., & Cannatella, D. C. (2007). Phylogeny-based delimitation of species boundaries and contact zones in the trilling chorus frogs (Pseudacris). *Molecular Phylogenetics and Evolution*, *44*(3), 1068–1082.

Lucek, K., Gompert, Z., & Nosil, P. (2019). The role of structural genomic variants in population differentiation and ecotype formation in *Timema cristinae* walking sticks. *Molecular Ecology*, *28*(6), 1224-1237.

Lumibao, C. Y., Hoban, S. M., & McLachlan, J. (2017). Ice ages leave genetic diversity 'hotspots' in Europe but not in Eastern North America. *Ecology Letters*, *20*(11), 1459–1468. https://doi.org/10.1111/ele.12853

Maguire, K. C., Shinneman, D. J., Potter, K. M., & Hipkins, V. D. (2018). Intraspecific niche models for ponderosa pine (*Pinus ponderosa*) suggest potential variability in population-level response to climate change. *Systematic Biology*, *67*(6), 965–978.

McKelvy, A. D., & Burbrink, F. T. (2017). Ecological divergence in the yellow-bellied kingsnake (*Lampropeltis calligaster*) at two North American biodiversity hotspots. *Molecular Phylogenetics and Evolution*, *106*, 61–72. https://doi.org/10.1016/j.ympev.2016.09.006

Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J. M., Uriarte, M., & Anderson, R. P. (2014). ENM eval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods in Ecology and Evolution*, *5*(11), 1198–1205.

Myers, E. A., Bryson, R. W. Jr, Hansen, R. W., Aardema, M. L., Lazcano, D., & Burbrink, F. T. (2019). Exploring Chihuahuan Desert diversification in the gray-banded kingsnake, *Lampropeltis alterna* (Serpentes: Colubridae). *Molecular Phylogenetics and Evolution*, *131*, 211–218.

Myers, E. A., Burgoon, J. L., Ray, J. M., Martínez-Gómez, J. E., Matías-Ferrer, N., Mulcahy, D. G., & Burbrink, F. T. (2017). Coalescent species tree inference of *Coluber* and *Masticophis*. *Copeia*, *105*(4), 642–650. https://doi.org/10.1643/CH-16-552

Myers, E. A., Rodríguez-Robles, J. A., Denardo, D. F., Staub, R. E., Stropoli, A., Ruane, S., & Burbrink, F. T. (2013). Multilocus phylogeographic assessment of the California Mountain Kingsnake (Lampropeltis zonata) suggests alternative patterns of diversification for the California Floristic Province. *Molecular Ecology*, *22*(21), 5418–5429. https://doi.org/10.1111/mec.12478

Myers, E. A., Xue, A. T., Gehara, M., Cox, C., Davis Rabosky, A. R., Lemos-Espinal, J., … Burbrink, F. T. (2019). Environmental heterogeneity and not vicariant biogeographic barriers generate community wide population structure in desert adapted snakes. *Molecular Ecology*, *28*, 4535–4548.

Namroud, M.-C., Bealieu, J., Juge, N., Laroche, J., & Bousquet, J. (2008). Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Molecular Ecology*, *17*(16), 3599–3613. https://doi.org/10.1111/j.1365-294X.2008.03840.x

Near, T. J., Page, L. M., & Mayden, R. L. (2001). Intraspecific phylogeography of *Percina evides* (Percidae: Etheostomatinae): An additional test of the Central Highlands pre-Pleistocene vicariance hypothesis. *Molecular Ecology*, *10*(9), 2235–2240.

Noss, R. F., Platt, W. J., Sorrie, B. A., Weakley, A. S., Means, D. B., Costanza, J., & Peet, R. K. (2015). How global biodiversity hotspots may go unrecognized: Lessons from the North American Coastal Plain. *Diversity and Distributions*, *21*(2), 236–244.

Oliveira, E. F., Gehara, M., São-Pedro, V. A., Chen, X., Myers, E. A., Burbrink, F. T., … Rodrigues, M. T. (2015). Speciation with gene flow in whiptail lizards from a Neotropical xeric biome. *Molecular Ecology*, *24*(23), 5957–5975.

Papadopulos, A. S. T., Baker, W. J., Crayn, D., Butlin, R. K., Kynast, R. G., Hutton, I., & Savolainen, V. (2011). Speciation with gene flow on Lord Howe Island. *Proceedings of the National Academy of Sciences*, *108*(32), 13188–13193.

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLOS Genetics*, *2*(12), e190.

Pauly, G. B., Piskurek, O., & Shaffer, H. B. (2007). Phylogeographic concordance in the southeastern United States: The flatwoods salamander, *Ambystoma cingulatum*, as a test case. *Molecular Ecology*, *16*(2), 415–429.

Pearman, P. B., D'Amen, M., Graham, C. H., Thuiller, W., & Zimmermann, N. E. (2010). Within-taxon niche structure: Niche conservatism, divergence and predicted effects of climate change. *Ecography*, *33*(6), 990–1003.

Pearson, R. G., Raxworthy, C. J., Nakamura, M., & Townsend Peterson, A. (2007). Predicting species distributions from small numbers of occurrence records: A test case using cryptic geckos in Madagascar. *Journal of Biogeography*, *34*(1), 102–117.

Peterson, A. T., Soberón, J., & Sánchez-Cordero, V. (1999). Conservatism of ecological niches in evolutionary time. *Science*, *285*(5431), 1265–1267. https://doi.org/10.1126/science.285.5431.1265

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*(3–4), 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Portik, D. M., Leaché, A. D., Rivera, D., Barej, M. F., Burger, M., Hirschfeld, M., … Fujita, M. K. (2017). Evaluating mechanisms of diversification in a Guineo-Congolian tropical forest frog using demographic model selection. *Molecular Ecology*, *26*(19), 5245–5263. https://doi.org/10.1111/mec.14266

Pyron, R. A., & Burbrink, F. T. (2010). Hard and soft allopatry: Physically and ecologically mediated modes of geographic speciation. *Journal of Biogeography*, *37*(10), 2005–2015. https://doi.org/10.1111/j.1365-2699.2010.02336.x

Rambaut, A., Suchard, M. A., Xie, D., & Drummond, A. J. (2014). *Tracer v1. 6. Computer Program and Documentation Distributed by the Author*. http://Beast.Bio.Ed.Ac.Uk/Tracer (Accessed 27 July 2014).

Rinker, D. C., Specian, N. K., Zhao, S., & Gibbons, J. G. (2019). Polar bear evolution is marked by rapid changes in gene copy number in response to dietary shift. *Proceedings of the National Academy of Sciences*, *16*(27), 13446–13451.

Ruane, S. (2015). Using geometric morphometrics for integrative taxonomy: An examination of head shapes of milksnakes (genus Lampropeltis). *Zoological Journal of the Linnean Society*, *174*(2), 394–413.

Rundle, H. D., & Nosil, P. (2005). Ecological speciation. *Ecology Letters*, *8*(3), 336–352. https://doi.org/10.1111/j.1461-0248.2004.00715.x

Satler, J. D., & Carstens, B. C. (2017). Do ecological communities disperse across biogeographic barriers as a unit? *Molecular Ecology*, *26*, 3533–3545. https://doi.org/10.1111/mec.14137

Schultz, K. D. (1996). In K. S. Books (Ed.), *A Monograph of the colubrid snakes of the genus* Elaphe Fitzinger. Czech Republic: Havlickuv Brod.

Smith, A. B., Godsoe, W., Rodríguez-Sánchez, F., Wang, H.-H., & Warren, D. (2018). Niche estimation above and below the species level. *Trends in Ecology & Evolution*, *34*, 260–273.

Smith, H. M., Chiszar, D., Staley, J. R., & Tepedelen, K. (1994). Populational relationships in the corn snake *Elaphe-Guttata* (Reptilia, Serpentes). *Texas Journal of Science*, *46*(3), 259–292.

Soltis, D. E., Morris, A. B., McLachlan, J. S., Manos, P. S., & Soltis, P. S. (2006). Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology*, *15*(14), 4261–4293. https://doi.org/10.1111/j.1365-294X.2006.03061.x

Sousa, V., & Hey, J. (2013). Understanding the origin of species with genome-scale data: Modelling gene flow. *Nature Reviews Genetics*, *14*(6), 404–414.

Stamatakis, A. (2006). RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, *22*(21), 2688–2690.

Sukumaran, J., & Knowles, L. L. (2017). Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences*, *114*(7), 1607–1612. https://doi.org/10.1073/pnas.1607921114

Taberlet, P., Fumagalli, L., Wust-Saucy, A.-G., & Cosson, J.-F. (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, *7*(4), 453–464.

Terhorst, J., & Song, Y. S. (2015). Fundamental limits on the accuracy of demographic inference based on the sample frequency spectrum. *Proceedings of the National Academy of Sciences*, *112*(25), 7677–7682.

Thuiller, W., Georges, D., & Engler, R. (2013). *biomod2: Ensemble platform for species distribution modeling. R Package Version*. https://doi.org/10.1017/CBO9781107415324.004

Ursenbacher, S., Carlsson, M., Helfer, V., Tegelström, H., & Fumagalli, L. (2006). Phylogeography and Pleistocene refugia of the adder (Vipera berus) as inferred from mitochondrial DNA sequence data. *Molecular Ecology*, *15*(11), 3425–3437.

Valladares, F., Matesanz, S., Guilhaumon, F., Araújo, M. B., Balaguer, L., Benito-Garzón, M., … Naya, D. E. (2014). The effects of phenotypic plasticity and local adaptation on forecasts of species range shifts under climate change. *Ecology Letters*, *17*(11), 1351–1364.

Walker, M. J., Stockman, A. K., Marek, P. E., & Bond, J. E. (2009). Pleistocene glacial refugia across the Appalachian Mountains and coastal plain in the millipede genus Narceus: Evidence from population genetic, phylogeographic, and paleoclimatic data. *BMC Evolutionary Biology*, *9*(1), 25.

Waltari, E., & Hickerson, M. J. (2012). Late Pleistocene species distribution modelling of North Atlantic intertidal invertebrates. *Journal of Biogeography*, *40*, 249–260. https://doi.org/10.1111/j.1365-2699.2012.02782.x

Waltari, E., Hijmans, R. J., Peterson, A. T., Nyári, Á. S., Perkins, S. L., & Guralnick, R. P. (2007). Locating Pleistocene Refugia: Comparing phylogeographic and ecological niche model predictions. *PLoS ONE*, *2*(7), e563. https://doi.org/10.1371/journal.pone.0000563

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

Xue, A. T., & Hickerson, M. J. (2015). The aggregate site frequency spectrum (aSFS) for comparative population genomic inference. *Molecular Ecology*, *24*(24), 6223-6240.

Zellmer, A. J., Hanes, M. M., Hird, S. M., & Carstens, B. C. (2012). Deep phylogeographic structure and environmental differentiation in the carnivorous plant *Sarracenia alata*. *Systematic Biology*, *61*(5), 763–777. https://doi.org/10.1093/sysbio/sys048

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Myers EA, McKelvy AD, Burbrink FT. Biogeographic barriers, Pleistocene refugia, and climatic gradients in the southeastern Nearctic drive diversification in cornsnakes (*Pantherophis guttatus* complex). *Mol Ecol*. 2020;29:797–811. https://doi.org/10.1111/mec.15358