Zoological Journal of the Linnean Society, 2024, **XX**, 1–17
https://doi.org/10.1093/zoolinnean/zlae018
Advance access publication 20 March 2024
Original Article

Original Article

Contrasting the depths of divergence between gene-tree and coalescent estimates in the North American racers (Colubridae: *Coluber constrictor*)

Edward A. Myers[1,2,*], Marcelo Gehara[3,4], Jamie L. Burgoon[2], Alexander D. McKelvy[5,6], Lauren Vonnahme[2], Frank T. Burbrink[2]

[1]Department of Herpetology, California Academy of Sciences, 55 Music Concourse Drive, San Francisco, California 94118, USA
[2]Department of Herpetology, The American Museum of Natural History, Central Park West and 79th Street, New York, NY 10024, USA
[3]Sackler Institute for Comparative Genomics, American Museum of Natural History, NY 10024, USA
[4]Department of Biological Sciences, Rutgers University-Newark, 195 University Ave, Newark, NJ 07102, USA
[5]Department of Biology, The Graduate School and Center, City University of New York, New York, New York 10016, USA
[6]Snake Evolution and Biogeography, Wacissa, FL 32361, USA

[*]Corresponding author. Department of Herpetology, California Academy of Sciences, 55 Music Concourse Drive, San Francisco, California 94118, USA.
E-mail: eddie.a.myers@gmail.com

ABSTRACT

The North American racers (*Coluber constrictor*) are widely distributed across the Nearctic and numerous studies have demonstrated extensive variation in morphology, ecology, and population genetic structure. Here we take an integrative approach to understand lineage diversification within this taxon by combining genomic sequence capture data, mtDNA sequence data, morphometrics, and ecological niche models. Both the genomic data and mtDNA phylogeographic analyses support five lineages distributed across the range of this species. However, demographic model selection based on these two datasets strongly conflict in both the model of divergence and estimates of timing of lineage divergence. While mtDNA and concatenated genomic data suggest a Miocene origin of these distinct groups, coalescent-based demographic models with the sequence capture data suggest lineage diversification occurred at ~33 kya in allopatry without gene flow. Using linear morphological measurements of head shape we demonstrate that lineages distributed largely east and west of the Mississippi River are distinguishable. Furthermore, ecological niche models demonstrate that lineages distributed in subtropical habitats have environmental niche space that is significantly differentiated from lineages distributed across the continent. Taken together, these results suggest that ecology is an important axis of lineage divergence within this group and that more fine-scale analyses may find even greater differentiation between the populations identified here. This abstract translated to Spanish is avaliable in the Supporting Infromation section (Este resumen traducido al español está disponible en la sección, Supporting Infromation).

Keywords: phylogeography; genomic discordance; Colubridae; ecological niche models; linear morphometrics

INTRODUCTION

Understanding the process of lineage formation and speciation is central to evolutionary biology. These processes can be studied using various phylogeographic methods, and when coupled with high-throughput sequencing data have the potential to uncover the timing and mode of lineage divergence with high accuracy and precision (Pinho and Hey 2010, Carstens *et al.* 2012, Schrider and Kern 2018). Central to these advances are model-based approaches to infer the evolutionary history of taxa in which competing demographic histories can be ranked in a statistically comparable manner (Beaumont 2010). Such demographic modelling provides a framework to understand the relative importance of how processes like gene flow and allopatric isolation affect speciation (Portik *et al.* 2017, Myers *et al.* 2020). These approaches for testing competing demographic models with genomic data in non-model taxa have become commonplace in evolutionary biology.

Demographic models, while providing important insights into the diversification and evolutionary dynamics of species, are limited in that they are simplified approximations of

Received 21 August 2023; revised 28 December 2023; accepted 26 January 2024

complex historical processes. If the tested models are not a close fit to the true underlying demographic histories, it is likely that both model selection and parameter estimation will be biased (Hickerson 2014). Additionally, demographic events, like linked selection, migration rates, or changes in ancestral population sizes, when unaccounted for can bias the outcome of these analyses (Ewing and Jensen 2016, Roux *et al.* 2016, Momigliano *et al.* 2021). Lastly, it is also possible that there may not be enough power to discriminate among similar models; for instance, it may be difficult to discern between isolation-only and secondary-contact models when time periods between isolation and secondary contact are very short (Roux *et al.* 2016, Provost *et al.* 2021).

Therefore, it is beneficial to have additional lines of evidence that can be useful to understand the evolutionary history of a taxon. For example, in addition to demographic analyses with genomic data, studies of lineage diversification can be enhanced by incorporating ecological niche models (ENMs) and analyses of morphological data. These additional sources of information can be informative about both the processes by which lineages have formed and how they are maintained. For example, ENMs can be used in testing the role of ecological divergence or niche conservatism in driving population differentiation (Pyron and Burbrink 2009), to help distinguish among geographic modes of speciation when coupled with demographic model selection (Myers *et al.* 2020, Jaynes *et al.* 2022), and to assess whether contact zones between lineages consists of habitat that may impede or facilitate gene flow (Costa *et al.* 2008). Ecological divergence is often an important step in the process of speciation (Nosil 2012), particularly where lineages potentially overlap in space, and this can, in part, be assessed using ENM methods (Rissler and Apodaca 2007).

The discovery of phylogeographic lineages within wide-ranging taxa may suggest that hidden cryptic species are present, where there is little to no obvious morphological differentiation. However, this assumption is often not immediately quantified by assessing morphological differences. Given that phylogeographic lineages are independently evolving, it might be expected that phenotypic variation is associated with divergence in ecology due to interspecific competition for resources and niche partitioning and/or reproductive strategies (Fabre *et al.* 2016). Therefore, combining independent datasets (e.g. genomic, morphological, and ecological) can provide a more robust view of the process of population divergence and lineage formation, particularly in areas that have extensive heterogeneity in environments and complex geological histories.

The demographic histories of Nearctic species have been influenced by biogeographic barriers, dynamic glacial cycles of the Pleistocene, and climatic gradients (O'Connell *et al.* 2017, Myers *et al.* 2019b). There are several well-recognized biogeographic barriers to gene flow within North America that have contributed to lineage divergence. This includes hard-barriers like the Mississippi River and the Appalachian Mountains (Burbrink *et al.* 2000, Martin *et al.* 2016), as well as soft-barriers like the discontinuity seen at the Florida peninsula in terrestrial taxa that is thought to be the result of past sea-level changes (Fetter and Weakley 2019). Historical changes in climate throughout the Pleistocene have also resulted in population divergence and speciation in many taxa within allopatric refugia (Walker *et al.* 2009, Wilson and Pitts 2012). Additionally, there are climatic gradients across these regions that influence species' distributions and community composition (Costa *et al.* 2008), and that have probably had an impact on population genetic structure between species (Burbrink *et al.* 2022). Therefore, several factors in this region may be responsible for lineage formation, resulting in many wide-ranging species being composed of multiple distinct lineages (Hamilton *et al.* 2014, O'Connell *et al.* 2017, Morales and Carstens 2018, Kim *et al.* 2022).

Study system

The North American racers (*Coluber constrictor*) are moderately-sized snakes with adult snout-to-vent lengths reaching up to 190 cm and are widely distributed across much of the Nearctic (Fig. 1). Across this wide distribution they can be found in numerous environments, from subtropical habitats to eastern deciduous forests to California chaparral (Ernst and Ernst 2003). Given the wide geographic distribution of this species, ecological tolerances, and variability in coloration and pattern, some biologists recognize 11 subspecies within *C. constrictor*. Phylogeographic analyses, however, have demonstrated that these subspecific groups do not correspond to population genetic structure (Burbrink *et al.* 2008, Myers *et al.* 2017). Instead, there are five mtDNA lineages distributed across the Nearctic, which were estimated to have begun diversifying as early as the Late Miocene (Burbrink *et al.* 2008).

We revisit the diversification of the North American racers, combining genomic sequence capture data, by-catch mtDNA data, morphometric measurements, and ecological niche models. We examine population genetic structure to identify mechanisms that have contributed to intraspecific diversification. Specifically, we characterize divergence in genomes, ecology, and morphology to address hypotheses about potential biogeographic barriers across North America and to differentiate among mechanisms proposed for the accumulation of biodiversity in the Nearctic. We expect to find population genetic structure across previously identified barriers in North America (e.g. the Mississippi River and the Florida peninsula discontinuity) and that much of this population divergence predates the Pleistocene glacial cycles, as demonstrated previously with mtDNA analyses. Lastly, we hypothesize that phylogeographic lineages within this species have also differentiated in morphology and ecology. We synthesize these findings in an integrative framework to address how biogeographic barriers and ecological gradients have resulted in species diversification.

METHODS AND MATERIALS

Sampling design

We sampled 91 individuals across the geographic distribution of *Coluber constrictor*, sampling all previously defined mtDNA phylogeographic lineages and all 11 putative subspecies (Fig. 1; Supporting Information, Table S1). We also included three *Masticophis flagellum* and one *M. fuliginosus* as outgroups (Myers *et al.* 2017). DNA was extracted using Qiagen DNeasy DNA extraction kits following manufacturer protocols, and DNA concentrations were quantified on a Qubit assay; 100 µL of DNA extract, with average DNA concentration of 34 ng/µL (2.3–151 ng/µL), was submitted to Arbor Biosciences for a

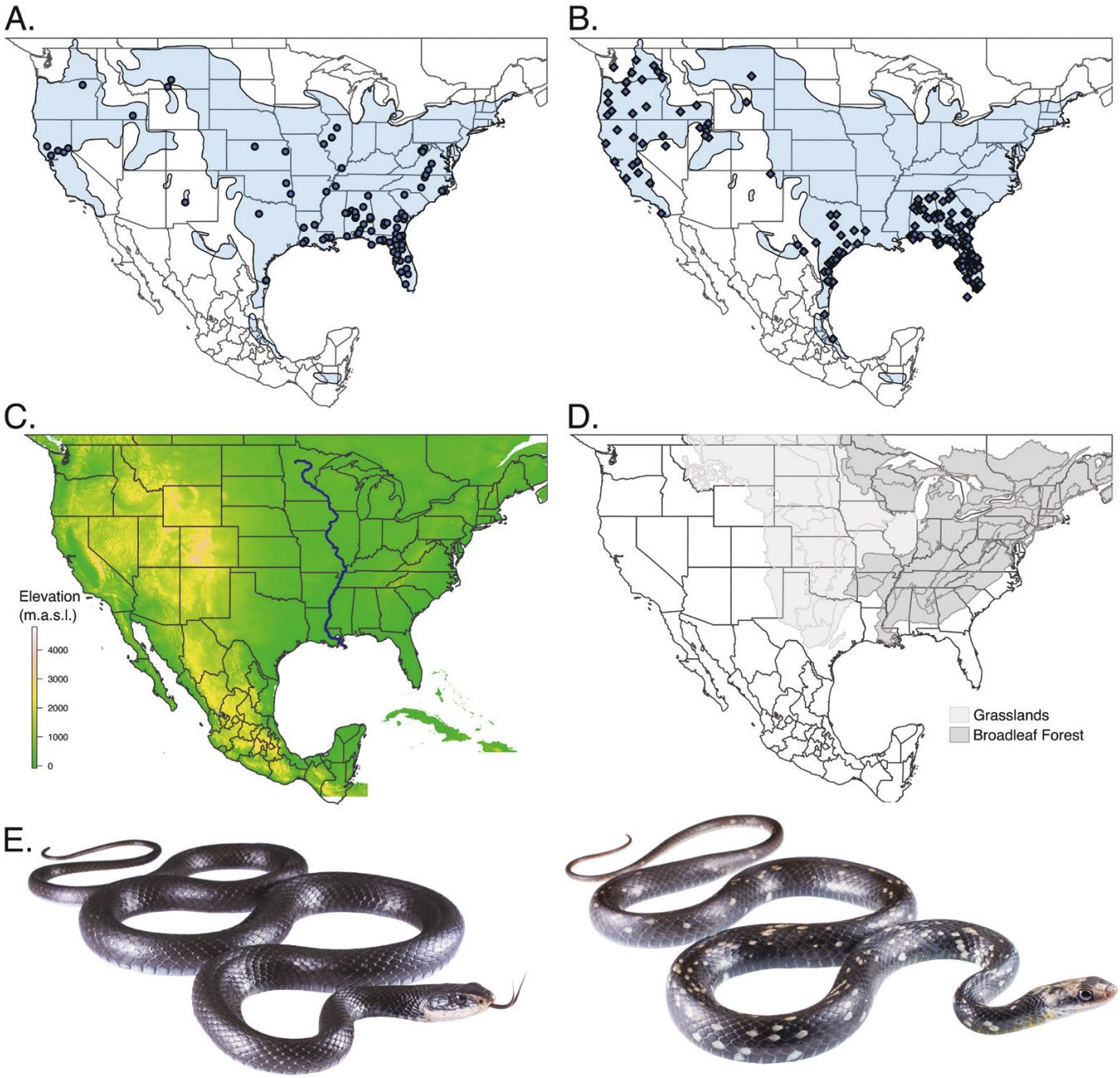

Figure 1. A, geographic distribution of sampling localities used to generate genomic sequence data. B, distribution of museum specimens measured for morphological analyses. In both (A) and (B) the shaded regions represent the approximate distribution of *Coluber constrictor*. C, elevation map of North America with the Mississippi River highlighted. The high elevation regions of central/western North America correspond to the Rocky Mountains. D, the distribution of grasslands and broadleaf forests in central and eastern North America; these biomes are defined by the World Wildlife Fund Terrestrial Ecoregions of the World. E, representative *C. constrictor* inidividuals from the eastern and southern United States (photo credits: T. Schramer); note that the snakes are not to scale with the maps.

myReads targeted sequencing project, where libraries were prepared and enriched with the myBaits SqCL probe set (Singhal *et al.* 2017). This sequence capture kit includes 5061 ultraconserved element loci (Faircloth *et al.* 2012), 394 anchored-hybrid enrichment loci (Lemmon *et al.* 2012), and 47 of the commonly sequenced Sanger loci for squamate systematics, and has been optimized for target capture in squamate reptiles (Singhal *et al.* 2017). Libraries were sequenced on a single lane of the Illumina HiSeq PE125.

Bioinformatics

Raw sequence reads were trimmed of adapter contamination using ILLUMIPROCESSOR v.2.0 (Faircloth 2013), a wrapper around the TRIMMOMATIC v.0.39 package (Del Fabbro *et al.* 2013), keeping sequences after trimming with a minimum length of 40 bp and a phred score of 33. These data were then *de novo* assembled using VELVET v.1.2.10 (Zerbino and Birney 2008) to construct contigs from the sequence reads using a k-mer length of 75. These contigs were then mapped to the SqCL probe set

using 'phyluce_assembly_match_contigs_to_probes' and fasta files were extracted from matched contigs using 'phyluce_assembly_get_fastas_from_match_counts' of the phyluce pipeline (Faircloth 2015).

We phased these data by using sequences from each individual sample within these newly created fasta files as its own 'reference' alignment and followed the seqcap_pop pipeline (Harvey *et al.* 2016). Briefly, following this pipeline, we used GATK v.3 (McKenna *et al.* 2010) to mark duplicates, call, realign, and annotate/mask indels, call and annotate SNPs, and then conduct read backed phasing. Finally, we created a sample specific fasta file of all phased loci for every sample using 'add_phased_snps_to_seqs_filter.py' available via the seqcap_pop pipeline (Harvey *et al.* 2016). All samples were combined into locus-specific fasta files using a Perl script developed in Myers *et al.* (2019a). Locus-specific fasta files were aligned using MUSCLE (Edgar 2004) with default parameters implemented in the R package *ape* (Paradis *et al.* 2004), after removing individuals from alignments that fell below the 25% quantile of the distribution of sequence lengths for each locus. All fasta files were conservatively trimmed so that there were no missing sites at the ends of alignments (i.e. trimming by the shortest sequence in the alignment). To avoid including potentially paralogous loci in the dataset, we used the boxplot() function in R to identify loci that were outliers in nucleotide diversity and removed these alignments. This dataset was further filtered to remove alignments with >30% missing individuals per locus. From the remaining fasta files we constructed a vcf file and filtered this for minor allele frequency >10% and retaining only 1 SNP/locus for population assignment analyses.

We also mined off-target sequence capture reads to assemble the mtDNA locus *Cytb* for samples that had not been sequenced in previous studies (e.g. Burbrink *et al.* 2008). We mapped the Illumina sequence capture reads to the *C. constrictor* GenBank accession MF402807 using MITObim v.1.9.1 (Hahn *et al.* 2013). Assembled sequences were then trimmed to the length of *Cytb* in GENEIOUS v.10.2.6 (https://www.geneious.com), combined with *Cytb* data available from GenBank, and aligned using MUSCLE (Edgar 2004) with default parameters.

Population structure

We estimated population structure across the distribution of *C. constrictor* using both sparse Non-Negative Matrix Factorization (sNMF) implemented in the *LEA* R package (Frichot *et al.* 2014, Frichot and François 2015) and by generating a principal component analysis (PCA) plot to visualize population clusters. Prior to estimating population structure, we used VCFTOOLS v.0.1.17 (Danecek *et al.* 2011) to filter the vcf file so that only one random SNP per locus was retained. The method sNMF uses sparse non-negative matrix factorization and computes least-squares estimates of ancestry coefficients (Frichot *et al.* 2014). For this analysis the number of putative ancestral populations tested ranged from $K = 1–6$, with 100 replicate runs for each value of K, and a total of 200 iterations. The cross-validation method using entropy criterion was used to determine the appropriate number of K clusters in the data, with masked genotypes set at 0.05. Values of K with the lowest cross-entropy scores were taken to be the best number of K clusters. Additionally, a PCA was conducted using *adegenet* in R (Jombart 2008) and visualized in *ggplot2* (Wickham *et al.* 2016). Lastly, we used SplitsTree v.4.19.2 (Huson and Bryant 2006) to estimate a network of the concatenated genomic data using the uncorrected *p*-distances.

Species tree, concatenated gene tree, and mtDNA gene tree

We used SNAPP v.1.5 (Bryant *et al.* 2012) implemented in BEAST v.2.6 (Bouckaert *et al.* 2014) to estimate the historical relationships among populations identified by our population clustering analyses. We sampled three individuals from each distinct population inferred from our sNMF analyses, with the exception of the population from south Texas for which only two samples were available. The individuals included in this analysis were chosen to minimize missing data. The same filtered vcf file used for sNMF was converted to nexus format using VCF2PHYLIP v.2.0 (Ortiz 2019) and the R package *phrynomics* (https://github.com/bbanbury/phrynomics). The mutation rates u and v were both set to 1.0, the lambda prior was set to a gamma distribution with alpha = 2 and beta = 200, and we set the 'snapprior' to alpha = 11.75, beta = 109.7, and lambda = 0.01. SNAPP was run for 1 000 000 iterations, sampling the Markov chain Monte Carlo (MCMC) every 1000 steps and the first 10% of samples were discarded as burn-in. We examined convergence using TRACER 1.6 (Rambaut *et al.* 2014) and generated a maximum clade credibility tree from the post-burn-in samples.

The phased loci were concatenated, and we used IQ-Tree v.2.1.3 (Minh *et al.* 2020) to estimate a gene tree based on this alignment with 1000 ultrafast bootstraps to assess support (Hoang *et al.* 2018). Initial IQ-Tree analyses suggested that a partitioned model has a worse fit to these concatenated data than a single evolutionary model. We, therefore, used ModelFinder implemented in IQ-Tree (Kalyaanamoorthy *et al.* 2017) to select a single model of nucleotide evolution for the concatenated matrix. We also used IQ-Tree to estimate a mtDNA gene tree using the *Cytb* data obtained from off-target sequence capture and previously published data on GenBank, support values were estimated with 1000 ultrafast bootstraps.

Lastly, we used BEAST v.2.6 to estimate gene divergence times using (i) the *Cytb* sequence data and (ii) the concatenated genomic data, following the calibrations in Burbrink *et al.* (2008). For the mtDNA analysis we included *Cytb* sampled from GenBank from the following additional taxa used as outgroups for calibrations: *Opheodrys aestivus*, *O. vernalis*, *Sonora semiannulata*, *Tantilla nigriceps*, *Mastigodryas melanolomus*, *Drymarchon couperi*, *Salvadora hexalepis*, *Masticophis mentovarius*, *Masticophis taeniatus*, *Masticophis lateralis*, and *Masticophis bilineatus*. Estimating divergence times for all ingroup specimens along with divergent outgroups can result in incorrectly estimated model parameters due to high variance in such a heterogeneous dataset (Guiher and Burbrink 2008). To avoid this issue, we used only one representative sequence from each of the identified phylogeographic lineages. We used jModelTest v.2.1.10 (Darriba *et al.* 2012) to select the best-fit model of nucleotide evolution based on Akaike information criterion (AIC) scores. We implemented the GTR+I+G model, a log-normal relaxed clock, and a birth–death tree model prior. Two calibrations were used, both of which are derived from the fossil record. First, both *Coluber* and *Masticophis* are known from the Late Miocene and, therefore, a mean date of 11 Mya with a log-normal standard

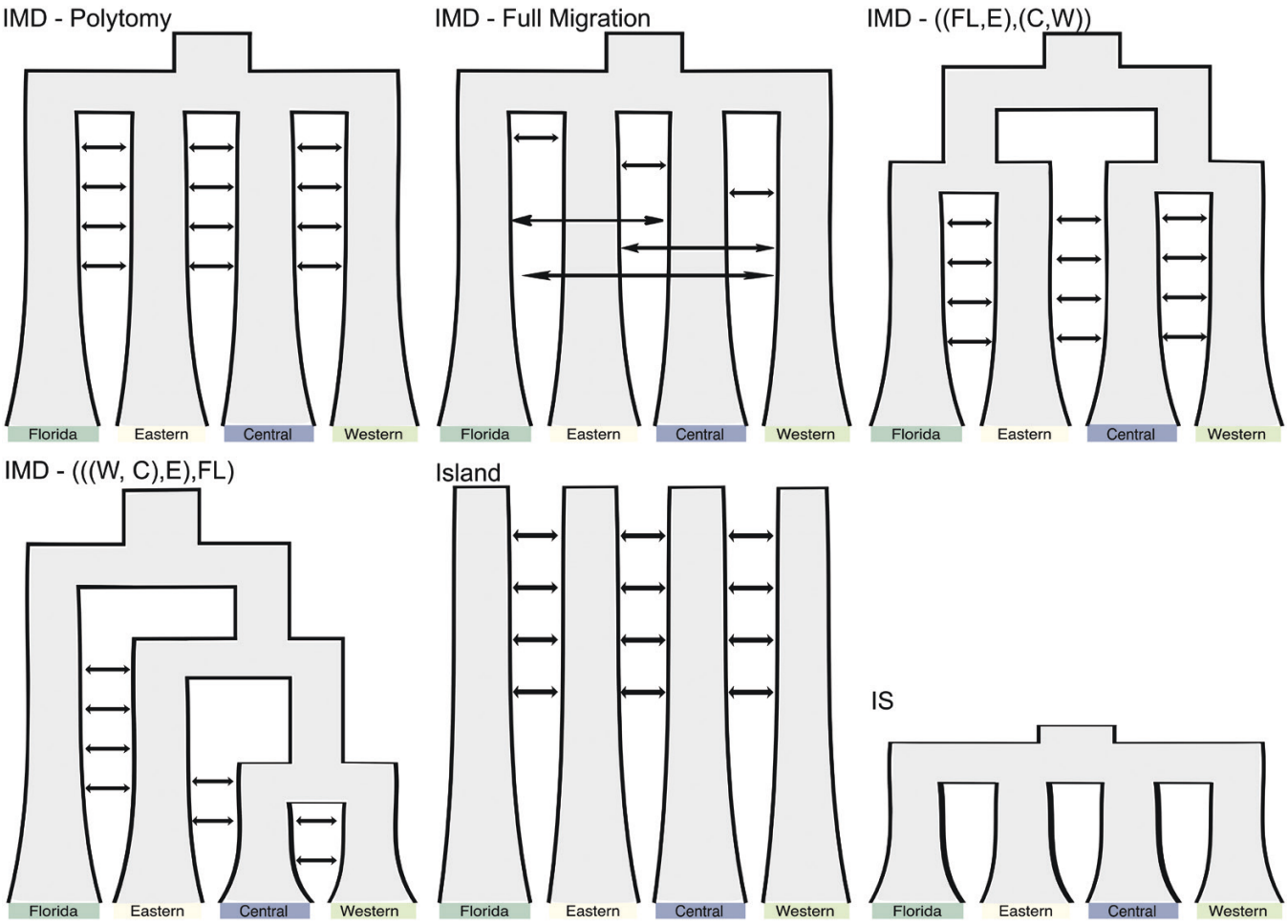

Figure 2. The six demographic models tested in PipeMaster using supervised machine learning. The four isolation-with-migration (IMD) models differ in topology or whether migration was restricted to geographically adjacent lineages only. IS represents a recent isolation without migration model.

deviation of 0.1 was used as a prior on the most recent common ancestor (MRCA) of these two genera (Holman 2000); this resulted in a 95% sampling interval of 9.0–13.3 Mya. The root of the New World Colubrinae radiation was constrained at a mean of 19 Mya with a log-normal standard deviation of 0.2 (95% sampling interval: 12.6–27.6 Mya). This date corresponds to fossils assigned to the genera *Paracoluber* and *Salvadora*, and are the earliest known colubrine snakes associated with this radiation found in North America (Holman 2000). The MCMC chain for the mtDNA analysis was run for 50 000 000 iterations, sampling every 5000 iterations. To date divergence times with the concatenated genomic data we also used BEAST v.2.6 with a dataset consisting of one individual per phylogeographic lineage and *Masticophis fuliginosus* as an outgroup. Due to a lack of additional outgroup sequences, we were only able to use the calibration between *Coluber* and *Masticophis* that was used above for the mtDNA divergence dating. For this analysis we also implemented the GTR+I+G model, a log-normal relaxed clock, and a birth–death tree model prior. The MCMC chain was run for 75 000 000 iterations sampling every 7500 iterations.

Demographic model selection and parameter estimation

We performed demographic model selection using supervised machine learning and simulations implemented in the *PipeMaster* R package (Gehara *et al.* 2017). These analyses were conducted for the Florida, eastern, western, and central lineages only; the South Texas lineage was excluded due to low sample size (see Results). We tested six demographic models that differed in the topology, presence or absence of gene flow, and divergence time (Fig. 2). Because of topological uncertainty, five of these models represent different versions of ancient isolation with high migration but are modelled with different topologies. The main contrast among the six models is between a model with recent isolation without migration (Is) and the five ancient divergence models with high migration (five competing IM models). All models included a population size change; because of negative Tajima's D across all lineages a constant size model did not fit the data (results not shown). We simulated 10 000 datasets of 52 summary statistics using *PipeMaster*. These summary statistics were estimated for each population and overall, and include average and variance of nucleotide diversity, the number of segregating sites, Watterson's theta, pairwise F_{ST} between populations, and overall F_{ST}.

Contemporary and ancestral effective population size parameters, divergence times, and timing of population size change parameters were sampled from uniform priors and migration parameters from normal priors truncated at zero (see Supporting Information, Table S2 for prior values). To accommodate

variation in mutation rate of all analysed loci (UCEs, AHE, and traditional Sanger loci), we sampled rates from a normal distribution truncated at 1×10^{-11}, with an average rate of 1×10^{-9} substitutions per site per generation with a standard deviation of 1×10^{-9} (for 5% and 95% quantiles of 2×10^{-11} and 2×10^{-10}, respectively). Before running the analysis, we explored model-fit by visualizing the variance space of the simulations against the observed data using principal components analysis (Supporting Information, Fig. S1). We used the *keras* R package (Arnold 2017) to build a neural network classification model with four hidden layers with 32 nodes each and *relu* activation function, followed by an output layer with *softmax* activation (Goodfellow *et al.* 2016, Gulli and Pal 2017, Agarap 2018). To train the neural network we used an *adam* optimizer with *sparse categorical cross-entropy* loss function and *accuracy* as a metric. We used 80% of the simulations to train the neural network for 1000 epochs, with the remaining 20% for testing. After training and testing we used the neural network to classify our empirical summary statistics and calculate the *softmax* probability for each model (note that the *softmax* probability is a measure of how confident the trained NN is in its resulting classification of the input data, e.g. a *softmax* of 0.95 translates to the NN being 95% confident that the input data can be classified to the correct model).

Because we found strong support for the recent divergence model without migration (Is; see Results), which contradicts age estimates from previous studies using mtDNA data, we also performed a model classification using available mtDNA data only. We tested the same models but using only a subset of summary statistics, because it is only possible to use point estimates for a single locus. We used identical neural network architecture, training, and testing as described above.

For the genomic data, we also estimated demographic parameters on three models that fit our data using a neural network regression. We did not estimate parameters for the mtDNA data due to the uncertainty of single-locus estimates. These models were isolation-with-migration (IMD), recent isolation without migration (IS), and an island model. For each parameter of the model, we used the same neural network architecture as above but with a single node in the output layer with a linear activation function. We used 88 000 simulations to train the neural network for 1000 epochs using the *root mean squared propagation* optimizer and *mean squared error* loss function, with 12 000 simulations for testing. We used the correlation coefficient between estimated and true values as a measure of error. To incorporate error in the parameter estimates we repeated the regression 100 times, randomly varying the training data in each regression. We then calculated the density of the 100 regressions, plotting it against the prior for comparison.

Ecological niche models and divergence

We used ENMs to estimate the potential geographic distributions of our focal taxa under current climatic conditions using the maximum entropy approach implemented in MaxEnt v.3.4.1 (Phillips *et al.* 2006). For locality data we downloaded all georeferenced *C. constrictor* occurrence records from VertNet. These specimen-based localities were assigned to lineage by creating a polygon around the samples used in our population clustering analyses and assigning all localities within a polygon to the corresponding lineage. This was conducted in R using the libraries *spatialEco* (Evans and Ram 2015) and *rgeos* (Bivand *et al.* 2014). Because of the large number of *C. constrictor* specimens held in natural history museums (>14 000 records on VertNet), and because the collection of snakes is probably biased by sampling specimens on roads and near urban centres, we removed duplicate localities and spatially thinned occurrences using *spThin* (Aiello-Lammens *et al.* 2015) in R, resulting in a dataset where nearest neighbours were no less than 25 km apart.

We used the contemporary bioclimatic variables from the CHELSA v.1.2 (Karger *et al.* 2017) dataset, which represent means, extremes, and seasonality of temperature and precipitation sampled at 2.5 arcmin resolution. We removed all but one of the variables that had a Pearson's correlation ≥0.7 using *ENMTools* in R (Warren *et al.* 2017). This resulted in eight variables: annual mean temperature, mean diurnal temperature range, isothermality, temperature seasonality, mean temperature of the wettest quarter, annual precipitation, precipitation of the driest month, and precipitation seasonality. To investigate whether the lineages identified above have non-equivalent environmental niches, we used the enmtools.ecospat.id function in *ENMTools*. This test uses kernel density smoothing to estimate a taxon's niche space and corrects for environment availability when assessing overlap between lineages (Broennimann *et al.* 2012). Using this test, a *D* statistic of 1 indicates complete niche overlap between lineages and a statistic equal to zero suggests no overlap. We ran this test for all geographically adjacent lineage pairs.

To identify optimal model parameters for MaxEnt 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 eight regularization multipliers ranging from 0.5 to 4.0, sampled at an interval of 0.5, using the R package *ENMeval* (Muscarella *et al.* 2014). This was done separately for each lineage and the best-fit parameters were chosen using AIC values. ENMs were constructed with MaxEnt using *Biomod2* (Thuiller *et al.* 2013) in R with 25 evaluation runs, each replicated for 5000 iterations, reserving 25% of sample localities as a training dataset for model evaluation. We created response curves and jack-knifed our data to measure variable importance for each ENM. To identify potential regions of refugia during the Pleistocene, we hindcasted these ENM projections on to the same eight non-correlated variables from the CHELSA Last Glacial Maximum dataset modelled from 21 kya (Karger *et al.* 2021).

Morphological divergence

To investigate if morphological divergence was consistent with population genetic structure, we collected data from eight morphometric measurements to the nearest 0.1 mm (snout-to-vent length, tail length, head width, head length, eye diameter, rostral length, prefrontal length, and internasal length) for a total of 244 specimens from across the species distribution (172 males, 72 females; Fig. 1; Supporting Information, Table S3). The linear measurements that we used to quantify morphological variation are commonly used in snake systematics (Thorpe 1975). Sex was determined by presence or absence of hemipenes and individuals were assumed to have reached sexual maturity with SVL >39 cm in males and >57 cm SVL in females (Brown and Parker 1984).

We used the R package *GroupStruct* (Chan and Grismer 2021) to correct for ontogenetic variation in body size and to visualize population clustering with PCA. To account for allometry, this method uses the allometric equation developed in Thorpe (1975). In *GroupStruct* we used the 'population1' flag in which the assumption is that there is geographic variation in the morphological data and, therefore, a separate allometric slope is calculated for each population. We assigned specimens to population genetic groups based on sampling locality. To test for sexual dimorphism, we used the Student *t*-test on males and females with the adjusted morphometric data. Several characters were found to be significantly different between sexes; therefore, males and females were analysed separately. PCA was used to visualize and assess the degree of morphological differentiation among the phylogeographic lineages. We tested the significance of the PCA using the R package *PCAtest* (Camargo 2022) with 1000 bootstrap replications and 1000 random permutations. Additionally, using *vegan* in R (Oksanen *et al.* 2015), we used a non-parametric permutation multivariate analysis of variance (PERMANOVA) to determine if the centroid location of each lineage in principal component space was statistically different (Skalski *et al.* 2018). This analysis is based on the distances between data points in a dissimilarity matrix and uses a post hoc test to calculate differences between combinations of population pairs to generate a *P*-value. Additionally, to quantify divergence in morphological characters between populations, we fit an ANOVA for each trait with measurements grouped by lineage and used a Tukey's honest significant differences test to calculated adjusted *P*-values for comparisons among group means. Lastly, we used discriminant function analysis with a leave-one-out cross-validation to determine if these phylogeographic lineages can be distinguished using morphological data alone, this analysis was conducted in the R package *MASS* (Venables and Ripley 2002).

RESULTS

Sequence alignments

The final assembled dataset consisted of 3417 loci (3348 UCE loci, 61 AHE loci, and eight of the 'legacy Sanger' loci). The mean proportion of loci recovered within each sample was 94% (range: 65–97%). The mean length of these assembled loci after trimming was 230.2 base pairs (range: 2–1731) and the mean number of segregating sites was 3.9 (range: 0–30). All raw sequence data have been accessioned on the NCBI Sequence Read Archive (BioProject ID PRJNA1082780; Supporting Information, Table S1). The mtDNA sequence assembled from these raw reads ranged in length from 124 to 7083 bp in length and were trimmed to the 1117 bp length of *Cytb*.

Population structure, phylogenetic relationships, and gene divergence times

Results from sNMF analyses indicate that $K = 5$ is the best fit (Fig. 3). These five populations correspond to a population in peninsular Florida, an eastern population in the temperate deciduous forests, a central population within the North American grasslands, a western population isolated across the Rocky Mountains, and a population from south Texas. Hereafter these populations will be referred to as the Florida, eastern, central, western, and south Texas populations, respectively. The PCA plot was visualized with the $K = 5$ population structure inferred from sNMF (Fig. 3). This plot shows clear separation of all five lineages, with only some overlap in PC space between the central and western populations. The network inferred from SplitsTree largely identifies these clusters, albeit with some degree of nestedness (Fig. 3D). For example, the western population is nested within the central group, while the eastern population is divided into two clusters.

The SNAPP-based species tree recovered a pectinate phylogeny with divergences as follows: south Texas, peninsular Florida, eastern, and a sister-relationship between the western and central populations (Fig. 4). All relationships within this tree were strongly supported with posterior probabilities ≥0.99. All ESS values from this analysis were >200 as assessed in TRACER v.1.7.1 (Rambaut *et al.* 2018). The concatenated phylogenomic tree from IQ-Tree, while finding each of these populations as a cluster, differs in their relationships. Here the monophyletic western population is nested within the central lineage, the eastern and Florida populations are sister-taxa, with the south Texas population being the earliest diverging group; again, these relationships are well supported (bootstrap ≥ 90). Within the concatenated tree there are two samples that have 'ladder-like' relationships between the eastern and Florida lineages, and both samples have high admixture proportions from sNMF (LSU-H2967 and FTB2688; both assigned to the eastern population). The mtDNA gene tree recovered the same phylogenetic relationships as the coalescent-based species tree, with moderate to high support (Fig. 4). One sample (LSU H-2967) clustered with the eastern lineage samples using genomic data but mtDNA originated from the eastern lineage.

Divergence times were estimated based on both a single locus mtDNA dataset and the concatenated genomic data (Fig. 4). The mtDNA analysis suggests that *Coluber* and *Masticophis* diverged 10.8 Mya (95% HPD: 8.9–12.7 Mya). In this gene tree, the earliest divergence within *Coluber* occurred 8.2 Mya (6.3–10.2 Mya), corresponding to the south Texas lineage divergence, demonstrating that these lineages began diversifying during the Mid to Late Miocene. The concatenated genomic analysis also suggests that *Coluber* began diversifying in the Mid to Late Miocene with an estimated crown age of 10.7 Mya (8.7–12.8 Mya; Fig. 4). The youngest divergence times between these phylogeographic lineages occurred 6.2 Mya (3.3–8.9 Mya) between the western and central populations.

Demographic model selection and parameter estimation

The trained neural network classified the six models with relatively low accuracy (0.82) due to confusion among the ancient divergence models. This unidentifiability among the ancient divergence models is more evident in the classification using the mtDNA alone where the accuracy was 0.44. When comparing one of the five ancient divergence models with the recent divergence model (Is), the accuracy was 1.0 for the genomic data and 0.88 for the mtDNA data.

The genomic data strongly supports the model with recent divergence without migration (Is), with a probability of ~1.0. Parameter estimates from this model suggest that lineage divergence of all populations occurred ~33 kya (25–43 kya 95% confidence interval) with no gene flow. These estimates also

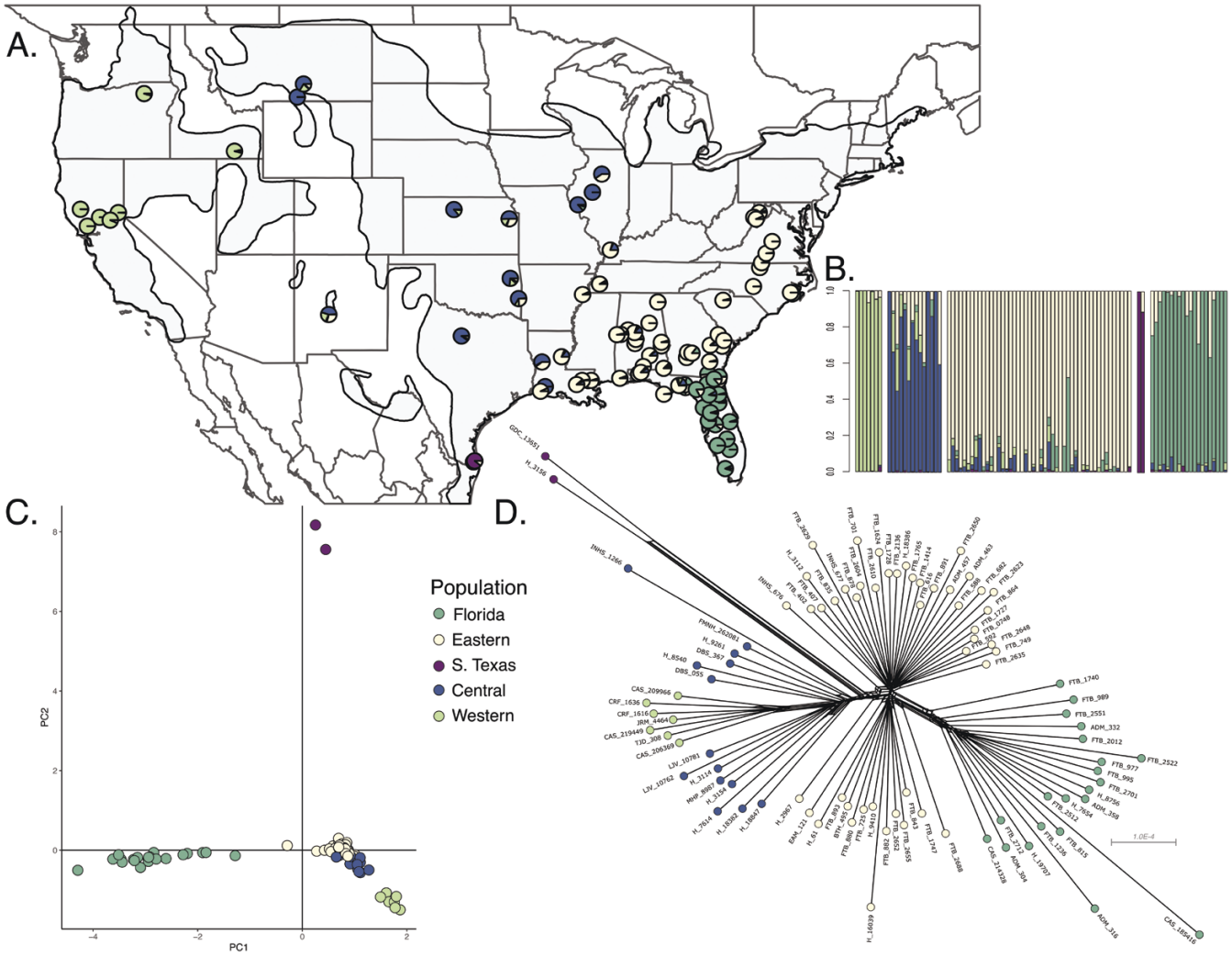

Figure 3. A, population structure inferred from the genomic sequence capture data using sNMF. Different colours within the pie charts correspond to admixture coefficients between each of the five lineages; Florida, eastern, central, south Texas, and western. Light grey highlighting represents the geographic distribution of *Coluber constrictor*. B, results of admixture coefficients from sNMF represented as a bar plot where each bar is a sampled individual. C, PCA of the genomic data demonstrating the differentiation of these phylogeographic lineages. D, SplitsTree network showing population structure and nestedness of these populations.

suggest small effective population sizes for the western and central populations and larger population sizes for the eastern and Florida populations, with a much stronger population size increase in the latter two. The timing of population size change also shows a more recent demographic change for the western population. These parameter estimates are available in the Supporting Information (Table S4; Fig. S2). By contrast, the mtDNA data support one of the ancient IMD models, with the model where the western population is sister to the central population and the eastern population is sister to the Florida population, having the highest probability ($P = 0.43$).

Ecological niche models and divergence

After thinning locality records, we retained 28 specimen localities for the south Texas population, 116 for the Florida population, 138 for the western, 431 for the central, and 463 for the eastern (localities and associated natural history specimen numbers are accessioned on Dryad). All comparisons of niche equivalency were significant but suggest varying levels of divergence in ecological niche between lineages. The greatest differences were between the Florida and the eastern lineages (D statistic = 0.02, P-value = 0.01) and between the central and south Texas lineage (D statistic = 0.03, P-value = 0.01). This test suggests that both the eastern and central lineages (D statistic = 0.14, P-value = 0.01), and the central and western lineages (D statistic = 0.27, P-value = 0.01), occupy more similar environments.

Each lineage had a unique, best-fit combination of feature class and regularization multipliers (Supporting Information, Table S5) that were used to construct ENMs in MaxEnt. All ENMs performed well with high area under the receiver operating characteristic curve values (AUC = 0.89 for south Texas; 0.99 for Florida population; 0.96 for eastern; 0.93 for central; 0.96 for western). All ENMs predict the geographic distribution of the focal lineage with little overlap in predicted geographic distribution between geographically adjacent populations (Fig. 5).

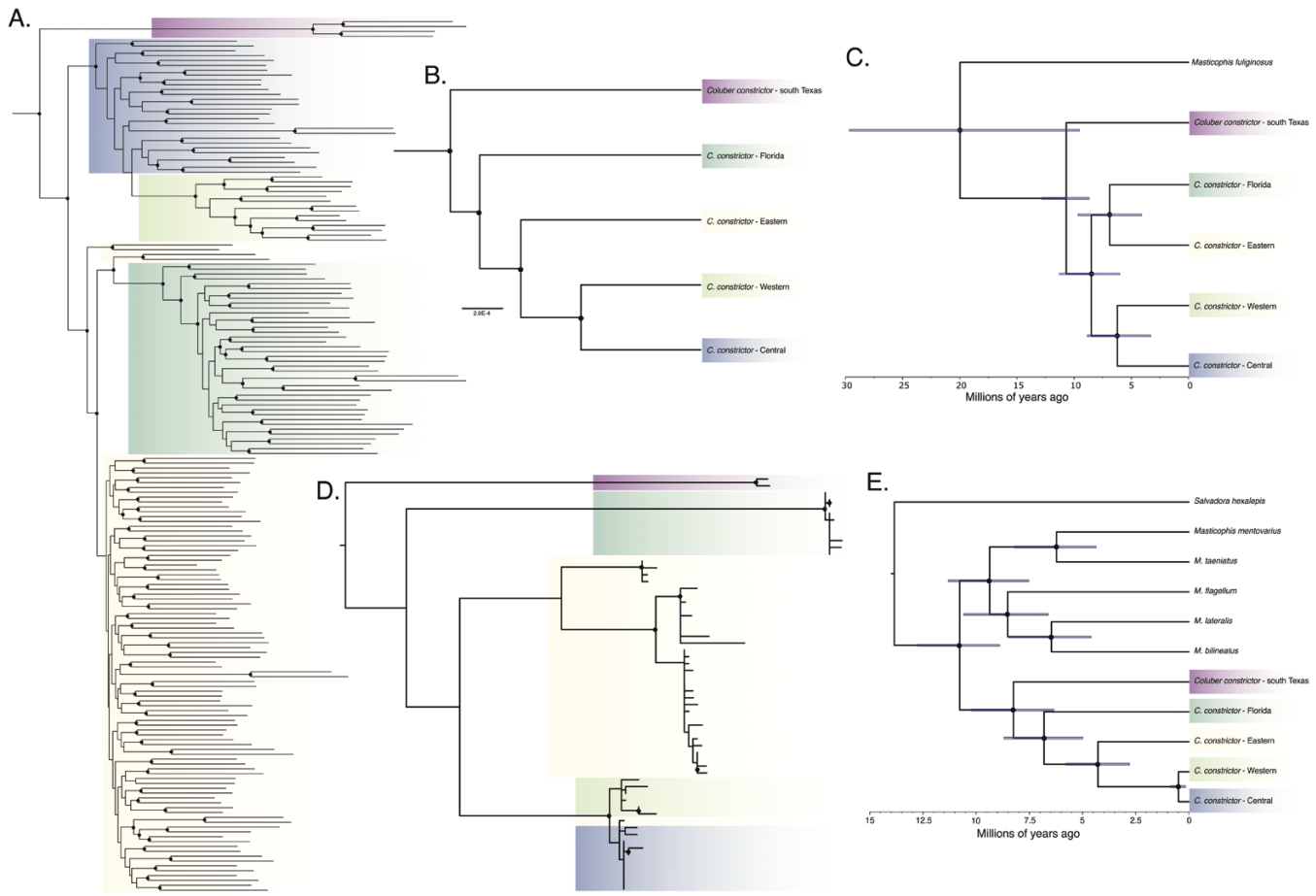

Figure 4. Gene trees and species tree based on sequence capture and by-catch mtDNA data. A, maximum likelihood concatenated tree inferred from phased sequence capture data. B, species' tree based on SNPs from assembled genomic data inferred using SNAPP. C, divergence dated tree from BEAST based on concatenated genomic sequence data. D, maximum likelihood mtDNA phylogeny for nearly all samples for which sequence capture data was generated. E, dated mtDNA gene tree inferred in BEAST including representative species from the sister-genus *Masticophis*. In (A) and (D) black circles at nodes represent bootstrap support values ≥95 and in (C), (D), and (E) represent posterior probabilities ≥0.9.

ENMs projected on to the Mid-Pleistocene suggests that all five lineages would have had reduced geographic distributions into refugia that were allopatric from one another (Fig. 5).

Morphological divergence

Sexual dimorphism was found in tail length, head width and length, eye diameter, and rostral length (P-values < 0.05). Principal component analyses are significant (P-value = 0 in both cases); however, there is no clear pattern of separation between lineages (Fig. 6). PC1 accounts for 39.1% of variation and PC2 for 19.7% variation in males, with the highest loadings contributed to head width, length, and eye diameter in PC1 and SVL in PC2. Within females, PC1 accounts for 41.0% of the total variation and PC2 for 16.7%, with the highest loadings contributed to head width, head length, rostral length, prefrontal width, and internasal width on PC1. Using PERMANOVA tests, males of the eastern and Florida lineages are significantly different from the other three lineages but not distinct from one another in morphology; similarly, the central, western, and south Texas lineages are indistinguishable (Table 1). PERMANOVA tests demonstrate that with morphological data from females, the eastern lineage is distinct from the western and south Texas lineages, and the Florida lineage is distinct from the south Texas clade, whereas all other comparisons were not significant (Table 1). Discriminant function analysis with cross-validation demonstrates that these lineages can be differentiated from one another with an accuracy of 56.2% in males and 47.5% in females. In all attempts to classify specimens, there are misclassifications between all lineages using both morphological datasets.

DISCUSSION

Previous studies have suggested morphological variation and extensive population genetic structure across the distribution of the North American racers (e.g. Ortenburger 1928, Anderson 1996, Burbrink *et al.* 2008). Here we corroborate much of this lineage divergence using a combination of genomic sequence capture data, mtDNA, morphology, and ecological niche models. We demonstrate that this structure corresponds to several well-known biogeographic barriers across North America. The genome-scale data and mtDNA show similar geographic lineages and largely agree in the phylogenetic relationships among these lineages. However, demographic models based on the genomic data strongly disagree with the best-fit model

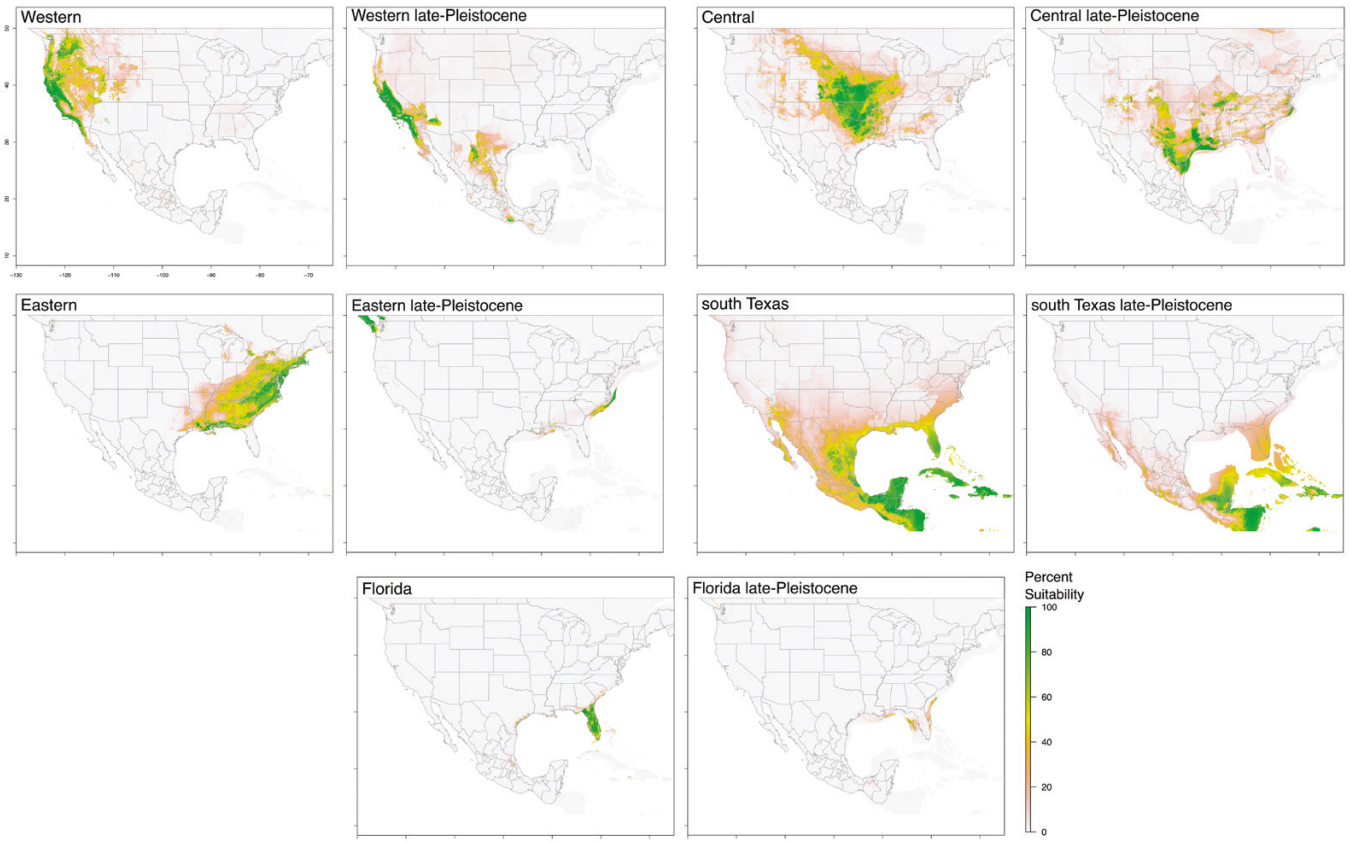

Figure 5. Ecological niche models constructed for each of the five phylogeographic lineages based on current climate and paleo-climate models from ~21 kya. All panels are labelled with their corresponding lineage and if they correspond to late Pleistocene projections. Regions shaded in green are those that were inferred to have higher environmental suitability for each lineage based on locality data used in the ENMs.

to the mtDNA in divergence times and topology. Furthermore, comparing divergence times estimated from mtDNA and concatenated genomic data to the divergence times estimated from demographic models suggests that these discrepancies are the result of differences between gene- vs. coalescent-divergence time estimates. Additionally, we find that several of these lineages are distinguishable morphologically, largely across the Mississippi River, and several adjacent population pairs have diverged in environmental niche. Lastly, hindcast ENMs suggest that all lineages were distributed allopatrically in the Mid-Pleistocene.

Demographic model selection

Model selection approaches were ambiguous to the evolutionary and demographic history of the North American racers. First, the best-fit model, that of recent divergence and no gene flow, could not determine the phylogenetic relationships among these lineages and instead suggested that a polytomy best represented the divergence history. Second, the timing of divergence and amount of gene flow is unclear; the model of recent divergence (~33 kya) obtained the highest support ($P = 1.0$), contradicting results based on mtDNA and concatenation, where divergence was estimated to be several orders of magnitude older. This conundrum is unsatisfying given that the goal of phylogeography is to infer historical relationships and estimate population genetic parameters (Hickerson *et al.* 2010), where the combination of model-based approaches with high throughput sequencing technologies have promised to do this with high accuracy and precision (McCormack *et al.* 2012).

Much of this uncertainty could stem from several historical processes. For example, the rapid diversification of these lineages may have resulted in difficulty discerning relationships and, therefore, a soft polytomy may be the best representation of historical relationships of these lineages. However, this may be unlikely given the inferred relationships based on multispecies coalescent, concatenated, and mtDNA tree-based analyses (Fig. 3). It is well known that rapid and successive lineage divergence can result in high gene tree heterogeneity and this has been shown to be widespread in many taxa (Linkem *et al.* 2016, Pardo-De la Hoz *et al.* 2023). The topological differences between our inferred species' tree and the concatenated tree suggest that there is extensive heterogeneity among gene trees within *C. constrictor*. These differences between gene-tree and coalescent-based analyses are also observed in comparing estimates of divergence times. For example, the timing of diversification largely agreed between the single locus mtDNA analysis and the concatenated genomic analysis. Both of these gene-based approaches suggest that *Coluber* began diversifying during the Late Miocene, which contrasts the coalescent-based estimate suggesting the Late Pleistocene. Whether summary statistic-based methods, like the neural network approach that we implemented here, are influenced by extreme heterogeneity among gene trees is unknown. Future simulation studies are needed to examine this issue (but see: Fan and Kubatko 2011, Alanzi and Degnan 2017).

This extreme discordance seen between the mtDNA and sequence capture data in reconstructing the demographic history of this taxon could also be the result of sex-biased dispersal

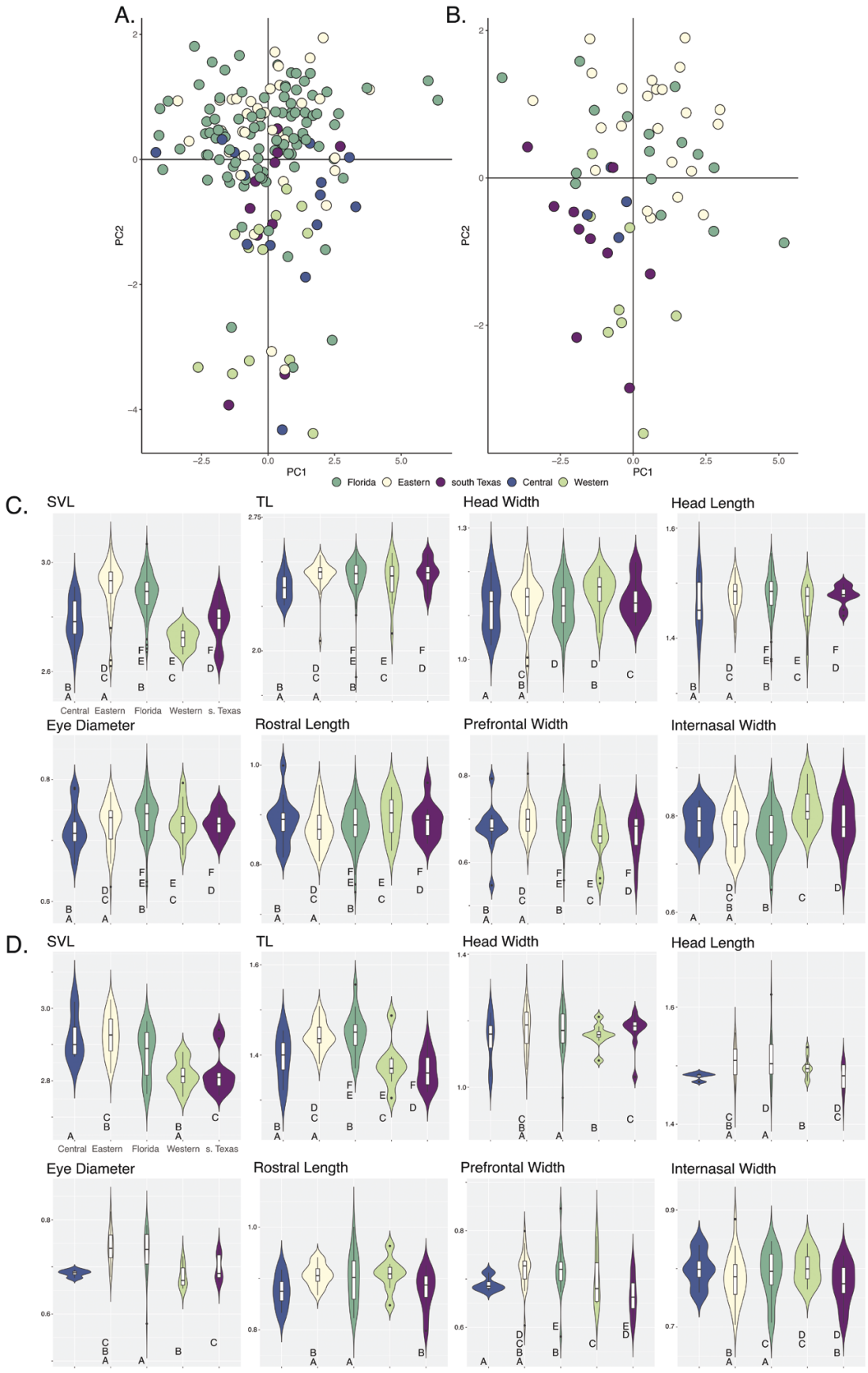

Figure 6. Morphological differentiation between phylogeographic lineages in *Coluber constrictor*. A, B, PCA of morphological variables; A, males; B, females. C, D, are boxplots of each linear measurement showing the differentiation of each of these traits between the five lineages: C, represents males; D, females. In panels (C) and (D), pairwise comparisons significant at an adjusted $P < .05$ with a Tukey's honest significant difference test are indicated by letters, where shared letters indicate that pair-wise comparisons are significantly different.

Table 1. Morphological differences from PERMANOVA.

	F-statistic	R^2	*P*-value
Male			
East—Florida	0.62	0.009	NS
East—Central	4.12	0.14	0.01
East—Western	6.93	0.22	0.01
East—S. Texas	5.37	0.20	0.01
Florida—Central	3.19	0.06	0.05
Florida—Western	5.56	0.1	0.01
Florida—S. Texas	4.16	0.08	0.05
Central—Western	0.67	0.06	NS
Central—S. Texas	1.42	0.15	NS
Western—S. Texas	0.78	0.08	NS
Female			
East—Florida	1.04	0.06	NS
East—Central	1.31	0.10	NS
East—Western	4.56	0.28	0.01
East—s. Texas	5.17	0.27	0.01
Florida—Central	0.62	0.06	NS
Florida—Western	1.88	0.16	NS
Florida—S. Texas	3.20	0.21	0.05
Central—Western	1.10	0.16	NS
Central—S. Texas	1.65	0.22	NS
Western—S. Texas	0.14	0.15	NS

(Toews and Brelsford 2012). In this case it is possible that females are highly philopatric, with high dispersal in males. This could result in a situation where ancient mtDNA lineages are maintained with much shallower divergence throughout the nuclear genome. Mark–recapture (Glaudas and Rodriguez-Robles 2011), as well as population genetic and phylogeographic analyses (Dubey *et al.* 2008, Pernetta *et al.* 2011), have suggested that male-biased dispersal is common within snakes. While this could explain the observed pattern here, it is expected that we would find a signal of gene flow in the nuclear genome, which is not the case in our demographic modelling. Additionally, when estimating divergence times based on the concatenated genomic data, we do not find the expected shallow divergence times.

It is also possible that the history of this group is too complex for the most commonly used isolation-migration models, where repeated bouts of isolation and contact are the cause of model selection uncertainty. Complex demographic histories of rapid divergence coupled with gene flow during glacial and interglacial periods is common in many temperate zone taxa (DeRaad *et al.* 2022, Harrington and Burbrink 2023). Similar histories of divergence have been suggested to be widespread and have been called the 'mixing-isolation-mixing' model of divergence (He *et al.* 2019). If this is the true underlying history of the group, then the expectation might be that model selection approaches would default to a history of ancient divergence with high gene-flow. However, previous studies have demonstrated that discriminating between isolation-only and secondary-contact models is difficult. For example, data simulated under a secondary-contact model could only be confidently assigned back to this true model when the period of isolation was sufficiently long. When shorter periods of isolation were simulated, ABC approaches were inconclusive (Roux *et al.* 2016).

How then can we choose among competing models of historical demography? Our demographic analyses suggest that the phylogeographic relationships are best represented as a polytomy with very recent divergence and no migration. However, this disagrees with the best-fit model based on mtDNA, estimated gene-divergence times, and with previous phylogeographic estimates (Burbrink *et al.* 2008). Here, we have a strong signal of divergence in the mtDNA phylogeographic analyses that matches both the population structure and SNAPP-based species' tree topology reconstructed from the genomic data, all dating the origin of diversification to the Miocene. We note, however, that both the mtDNA and concatenated genomic divergence times are gene-divergence, not lineage-divergence times (Edwards and Beerli 2000) and we do not take them to be reflective of species divergence times. A caveat of overly relying on gene-tree based (e.g. mtDNA alone) inferences is that increased levels of genetic diversity may be explained by the persistence of large N_e at mutation-drift equilibrium (Charlesworth 2009, Morgan-Richards *et al.* 2017). A pattern of high genetic polymorphism within sampling sites has been demonstrated to be the result of elevated levels of N_e maintained through time (Morgan-Richards *et al.* 2017). However, this seems less likely here given that the deep mtDNA lineages of racers are geographically structured and match population genetic structure observed with the genomic data.

In cases where demographic model selection and divergence-time estimates between datasets strongly disagree, there are two additional methods that may prove useful in future analyses. First, whole genome resequencing data could provide more accurate and precise model selection and demographic parameter estimates over commonly used sequence capture loci like UCEs. It is now possible to generate low-coverage, whole-genome resequencing data (~2× coverage) for similar costs compared to reduced representation sequencing methods, and by spreading sequencing effort across more samples, albeit at a lower per base pair coverage, while increasing the accuracy of many population genetic inferences (Lou *et al.* 2021, Reid and Pinsky 2022). Second, ancient DNA sampled from Pleistocene subfossils will improve our understanding of the demographic histories and evolutionary relationships of many taxa (Orlando *et al.* 2009, van der Valk *et al.* 2021). These approaches have also recently been used to better understand both taxonomy and biogeographic history in squamate reptiles and, in some cases, DNA has been obtained with minimally destructive sampling methods (Torres-Roig *et al.* 2021, Scarsbrook *et al.* 2022). Throughout North America, Pleistocene snake subfossils are remarkably common (Brattstrom 1967, Holman 2000) and future efforts to extract and sequence DNA from these fossils may further improve our understanding of species' diversification and demographic histories.

Coluber constrictor phylogeography

The observed population genetic structure within *Coluber constrictor* corroborate previously identified phylogeographic breaks within this taxon to known barriers across North America, and, in some cases, subspecific taxonomy. This structure that we find is associated with the Florida peninsula, across the transition

between the forested region of eastern North America and the central grasslands, across the Rocky Mountains, and a south Texas group that may extend as far south as southern Mexico and Guatemala.

Divergence between the Florida peninsula and continental North America, as well as across the transition between the forested region of eastern North America and the central grasslands, is commonly found in many taxa (Soltis *et al.* 2006, Burbrink *et al.* 2022). Although both regions have been shown to be important for promoting population genetic divergence across entire communities of organisms, the timing and causes of divergence are not well known. For example, across the Florida-continental discontinuity several non-mutually exclusive hypotheses have been proposed, including Pliocene sea-level rises resulting in Florida being mostly submerged, consisting of only a series of islands in the centre of the peninsula (Webb 1990, Clark *et al.* 1999), Pleistocene refugia (Waltari *et al.* 2007), which is also supported in our hindcast ENMs, or, as we have demonstrated here, the potential for divergent ecological selection on environmental niche. Similarly, population genetic structure across the middle Nearctic has contributed to both the ecotonal transition from deciduous, broadleaf forests into grasslands, as well as vicariance across the Mississippi River (Burbrink *et al.* 2000, Leaché and Reeder 2002, Pyron and Burbrink 2010, Satler and Carstens 2017, Myers *et al.* 2020). Future comparative studies should focus on the interaction of vicariance and adaptation in shaping shared patterns of biodiversity at these two regions.

Two lineages identified here also correspond to previously named taxonomic units, *C. c. mormon* and *C. c. oaxaca*. Both of these taxa were originally described as species (Baird and Girard 1852, Jan 1863) and there has been disagreement in the literature regarding the specific status of these taxa (Fitch *et al.* 1981, Greene 1984). The geographic distribution of the North American racers is allopatric across the Rocky Mountains and, therefore, it is not surprising to find population genetic structure across this region (but see: Corn and Bury 1986). The two samples from southern Texas that form a distinct population fall within the geographic range of *C. c. oaxaca*; however, the full distribution of this subspecies is discontinuous from southern Texas along the east coast of Mexico to Guatemala and Belize. Whether the entirety of this distribution belongs to this lineage is unknown. Furthermore, with our sampling of only two individuals, we were unable to include this lineage in our demographic analyses, so whether there is gene flow between this lineage and adjacent populations is also unknown.

Model selection with genomic data supported population size expansion in all lineages. These demographic size changes probably reflect population expansion associated with the end of the last glacial maxima in the Late Pleistocene, a common finding in phylogeographic studies of taxa in eastern North America (Hewitt 2000), and is in agreement with previous single-locus phylogeographic analyses of this taxon (Burbrink *et al.* 2008). The inferred population expansions also agree with the ENMs projected to the LGM where all lineages have large reductions in suitable habitat.

Ecological divergence despite little morphological differentiation

An important axis of differentiation between *C. constrictor* lineages is divergence in environmental niche. Both the south Texas and Florida lineage are significantly different in their respective ecological niche when compared to adjacent, parapatric lineages. However, none of the three lineages distributed across temperate North America have a signature of divergence in the environmental space in which they occur. Both the south Texas and Florida lineages occur in subtropical climates; therefore, adaptations to these environments may be an important barrier in keeping these lineages distinct. Divergence across environmental gradients has been shown to be important in promoting and maintaining lineage divergence within many regions (Myers *et al.* 2019b, Freedman *et al.* 2023). This process might be common, particularly with divergence between lineages in the Florida peninsula and continental North America. This region has long been shown to be an important contact zone between divergent lineages (Remington 1968, Soltis *et al.* 2006), but the cause of lineage divergence in many taxa is not known. It has been hypothesized that this could be the result of sea-level fluctuations, which historically have resulted in present-day Florida being a series of islands, or, alternatively, it could be divergence due to environmental differences.

Similarly to the population genetic structure within this taxon, there is some signal of morphological divergence. The measurements that we collected to examine morphology primarily focused on body size and linear measurements that describe head shape. Both the eastern and Florida lineages were found to be statistically differentiated from populations primarily west of the Mississippi River with morphometric data collected from males. There is some evidence that a similar pattern is found in females. However, a number of these lineages had small sample sizes that may have precluded detecting this differentiation. These differences were detectable using several linear measurements and, therefore, more fine-scale analyses designed to quantify shape differences may further differentiate these phylogeographic lineages (e.g. geometric morphometrics; Rohlf and Marcus 1993). Such approaches have been successfully applied to differentiate recently diverged squamate species (Ruane 2015, Pavón-Vázquez *et al.* 2022) and, more recently, μCT scans have further increased precision in identifying divergent lineages (Melville *et al.* 2019).

Competition between closely related snake species may be an important factor driving differences in body size and diet between populations of *C. constrictor*. Body size of North American racers has been hypothesized to decrease when they co-occur with the closely related species, the eastern coachwhip (*Masticophis flagellum*; Steen *et al.* 2013). Additionally, where *C. constrictor* lineages and *Masticophis* species co-occur, they partition resources by preying on different sizes and species of prey (Brown and Parker 1982, Halstead *et al.* 2008). Such ecological pressures are probably important for driving morphological differences, particularly in head shape and size in snakes (Vincent *et al.* 2004, Fabre *et al.* 2016). Character displacement via interspecific competition may be an important factor in driving morphological evolution in this group. Furthermore, there are additional morphological characters, such as differences in hemipenis morphology, that were not examined here and that may be important in maintaining population divergence, which is discussed below (Auffenberg 1955, Anderson 1996).

Our results corroborate several previous studies based on combinations of morphology, limited genetic data, and ecology that have identified all these lineages as distinct, although there are discrepancies between our results and some of these earlier

studies. To give a brief overview: (i) the distinction of the Florida lineage has been previously supported with fixed allelic differences from analysis of allozymes (Anderson 1996) and hemipenial basal spine morphology that differs from other populations throughout the eastern Nearctic (note that variation in the hemipenis may not be fixed between the eastern and Florida lineages according to: Dunn and Wood 1939, Auffenberg 1955, Anderson 1996); (ii) lineage structure that is loosely associated with the Mississippi River (here the distinction between the eastern and central lineages) is supported by allozyme analyses (Anderson 1996) and also corroborated by body coloration (Wright and Wright 1957); (iii) populations from south Texas through Central America are genetically distinct and possess unique hemipenis morphology (corresponding to *C. c. oaxaca*; Ortenburger 1928, Anderson 1996); and (iv) body size differences, scale counts, and ecology between the western and central lineage (the western lineage corresponds to *C. c. mormon*; Fitch *et al.* 1981; but see: Greene 1984). Future directions to better understand evolutionary history of this taxon should increase genetic sampling in several regions to better characterize the extent of hybrid zones to understand how population structure is being maintained and to better describe the geographic distributions of each lineage (e.g. in the upper Midwest of the US). Second, future work should focus on collecting additional morphological data with larger sample sizes to quantify variations that may distinguish these groups from one another.

CONCLUSION

We have demonstrated that the widely distributed North American racers show clear phylogeographic structure at several well-documented biogeographic barriers. These include divergence between the Florida peninsula and continental North America, at habitat transitions between eastern broadleaf forests to grasslands, and across the Rocky Mountains. ENMs suggest that there is little environmental differentiation among continentally distributed lineages, yet strong differentiation between these lineages and populations distributed in subtropical climates. Furthermore, there is evidence to support morphological divergence between lineages separated roughly by the Mississippi River. While diversification within this group has resulted in genetically, morphologically, and ecologically distinct lineages, the timing of this radiation differs dramatically between gene-based estimates of 8.2–10.7 Mya and coalescent-based estimates of ~33 kya.

SUPPLEMENTARY DATA

Supplementary data are available at *Zoological Journal of the Linnean Society* online.

ACKNOWLEDGEMENTS

The authors acknowledge the following institutions and individuals for providing access to tissues samples: American Museum of Natural History (M. Arnold), California Academy of Sciences (J. Vindum), Florida Museum of Natural History, University of Florida (K. Krysko, D. Blackburn), Louisiana State University Museum of Natural Sciences (J. Boundy, D. Dittman, R. Brumfeld, F. Sheldon), the Illinois Natural History Survey (C. Phillips), Sternberg Collection (T. Taggart), Texas Natural History Collections, Genetic Diversity Collection (T. LaDuc), Museum of Vertebrate Zoology (C. Spencer, J. McGuire), B. Hamilton, C. Feldman, D. Shepard, J. Mendelson, and L. Vitt, as well as the numerous people who helped us collect samples in the field. We thank Juan Esteban Martínez for translating the abstract to Spanish. Travis Taggart provided access to literature. We also thank K. de Queiroz, E. Langan, S. Gotte, and A. Wynn for access to specimens at the National Museum of Natural History, J. Sheridan and S. Kennedy-Gold for measurements of the type specimen of *C. constrictor helvigularis*. EAM was funded by the Gerstner Scholar and Theodore Roosevelt (AMNH) fellowships.

FUNDING

Funding was provided by NSF-DEB 1831241 and NSF-DEB 2323125.

CONFLICT OF INTEREST

None declared.

DATA AVAILABILITY

All demultiplexed Illumina reads are available via the NCBI SRA (BioProject ID:PRJNA1082780). All genomic and mtDNA data are available in fasta, phylip, and vcf format on Dryad (doi: 10.5061/dryad.rxwdbrvhh). Additionally, all locality data from the ENMs have been accessioned on Dryad as csv files.

REFERENCES

Agarap AF. Deep learning using rectified linear units (relu). arXiv preprint, arXiv:1803.08375, 2018.

Aiello-Lammens ME, Boria RA, Radosavljevic A *et al.* spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography* 2015;**38**:541–5. https://doi.org/10.1111/ecog.01132

Alanzi ARA, Degnan JH. Inferring rooted species trees from unrooted gene trees using approximate Bayesian computation. *Molecular Phylogenetics and Evolution* 2017;**116**:13–24. https://doi.org/10.1016/j.ympev.2017.07.017

Anderson RE. *Biochemical and Hemipenial Systematics of Coluber constrictor (Reptilia: Serpentes)*. Unpublished Master's Thesis, Southeastern Louisiana University, 1996.

Arnold TB. kerasR: R interface to the keras deep learning library. *Journal of Open Source Software* 2017;**2**:296.

Auffenberg W. A reconsideration of the racer, *Coluber constrictor*, in Eastern United States. *Tulane Studies in Zoology* 1955;**2**:89–155.

Baird SF, Girard C. Characteristics of some new reptiles in the Museum of the Smithsonian Institution. *Proceedings of the Academy of Natural Sciences of Philadelphia* 1852;**6**:68–70.

Beaumont MA. Approximate Bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics* 2010;**41**:379–406. https://doi.org/10.1146/annurev-ecolsys-102209-144621

Bivand R, Keitt T, Rowlingson B. *rgdal: Bindings for the Geospatial Data Abstraction Library. R package v.0.8-16*, 2014. Available online: https://cran.r-project.org/web/packages/rgdal/index.html.

Bouckaert R, Heled J, Kühnert D *et al.* BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 2014;**10**:e1003537. https://doi.org/10.1371/journal.pcbi.1003537

Brattstrom BH. A succession of pliocene and pleistocene snake faunas from the high plains of the United States. *Copeia* 1967;**1967**:188–202. https://doi.org/10.2307/1442194

Broennimann O, Fitzpatrick MC, Pearman PB *et al*. Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecology and Biogeography* 2012;**21**:481–97. https://doi.org/10.1111/j.1466-8238.2011.00698.x

Brown WS, Parker WS. Niche dimensions and resource partitioning in a great basin desert snake community. In: S Norman (ed.), *US Department of the Interior*. Washington, D.C.: Herpetological Communities, 1982. 59–81.

Brown WS, Parker WS. Growth, reproduction and demography of the racer, *Coluber constrictor mormon*, in northern Utah. *Vertebrate Ecology and Systematics: A Tribute to Henry S. Fitch*, 1984, 13–40.

Bryant D, Bouckaert R, Felsenstein J *et al*. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution* 2012;**29**:1917–1932.

Burbrink FT, Lawson R, Slowinski JB. Mitochondrial DNA phylogeography of the polytypic North American rat snake (*Elaphe obsoleta*): a critique of the subspecies concept. *Evolution; International Journal of Organic Evolution* 2000;**54**:2107–18. https://doi.org/10.1554/0014-3820(2000)054[2107:MDPOTP]2.0.CO;2

Burbrink FT, Fontanella F, Pyron RA *et al*. Phylogeography across a continent: the evolutionary and demographic history of the North American racer (Serpentes: Colubridae: *Coluber constrictor*). *Molecular Phylogenetics and Evolution* 2008;**47**:274–88.

Burbrink FT, Bernstein JM, Kuhn A *et al*. Ecological divergence and the history of gene flow in the Nearctic milksnakes (*Lampropeltis triangulum* complex). *Systematic Biology* 2022;**71**:839–58. https://doi.org/10.1093/sysbio/syab093

Camargo A. PCAtest: testing the statistical significance of principal component analysis in R. *PeerJ* 2022;**10**:e12967. https://doi.org/10.7717/peerj.12967

Carstens B, Lemmon AR, Lemmon EM. The promises and pitfalls of next-generation sequencing data in phylogeography. *Systematic Biology* 2012;**61**:713–5. https://doi.org/10.1093/sysbio/sys050

Chan KO, Grismer LL. A standardized and statistically defensible framework for quantitative morphological analyses in taxonomic studies. *Zootaxa* 2021;**5023**:293–300. https://doi.org/10.11646/zootaxa.5023.2.9

Charlesworth B. Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics* 2009;**10**:195–205. https://doi.org/10.1038/nrg2526

Clark AM, Bowen BW, Branch LC. Effects of natural habitat fragmentation on an endemic scrub lizard (*Sceloporus woodi*): an historical perspective based on a mitochondrial DNA gene genealogy. *Molecular Ecology* 1999;**8**:1093–104. https://doi.org/10.1046/j.1365-294x.1999.00653.x

Corn PS, Bury RB. Morphological variation and zoogeography of racers (*Coluber constrictor*) in the Central Rocky Mountains. *Herpetologica* 1986;**42**:258–64.

Costa GC, Wolfe C, Shepard DB *et al*. Detecting the influence of climatic variables on species distributions: a test using GIS niche-based models along a steep longitudinal environmental gradient. *Journal of Biogeography* 2008;**35**:637–46. https://doi.org/10.1111/j.1365-2699.2007.01809.x

Danecek P, Auton A, Abecasis G *et al*.; 1000 Genomes Project Analysis Group. The variant call format and VCFtools. *Bioinformatics* 2011;**27**:2156–8. https://doi.org/10.1093/bioinformatics/btr330

Darriba D, Taboada GL, Doallo R *et al*. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 2012;**9**:772–772. https://doi.org/10.1038/nmeth.2109

Del Fabbro C, Scalabrin S, Morgante M *et al*. An extensive evaluation of read trimming effects on Illumina NGS data analysis. *PLoS One* 2013;**8**:e85024. https://doi.org/10.1371/journal.pone.0085024

DeRaad DA, McCormack JE, Chen N *et al*. Combining species delimitation, species trees, and tests for gene flow clarifies complex speciation in scrub-jays. *Systematic Biology* 2022;**71**:1453–70. https://doi.org/10.1093/sysbio/syac034

Dubey S, Brown GP, Madsen T *et al*. Male-biased dispersal in a tropical Australian snake (*Stegonotus cucullatus*, Colubridae). *Molecular Ecology* 2008;**17**:3506–14. https://doi.org/10.1111/j.1365-294x.2008.03859.x

Dunn ER, Wood GC. Notes on eastern snakes of the genus *Coluber*. *Notulae Naturae* 1939;**5**:1–4.

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 2004;**32**:1792–7. https://doi.org/10.1093/nar/gkh340

Edwards SV, Beerli P. Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution; International Journal of Organic Evolution* 2000;**54**:1839–54. https://doi.org/10.1111/j.0014-3820.2000.tb01231.x

Ernst CH, Ernst EM. *Snakes of the United States and Canada*. Washington, DC: Smithsonian Books, 2003.

Evans JS, Ram K. *Package 'spatialEco'. R Package*, 2015. https://github.com/jeffreyevans/spatialEco.

Ewing GB, Jensen JD. The consequences of not accounting for background selection in demographic inference. *Molecular Ecology* 2016;**25**:135–41. https://doi.org/10.1111/mec.13390

Fabre AC, Bickford D, Segall M *et al*. The impact of diet, habitat use, and behaviour on head shape evolution in homalopsid snakes. *Biological Journal of the Linnean Society* 2016;**118**:634–47. https://doi.org/10.1111/bij.12753

Faircloth BC. illumiprocessor: a trimmomatic wrapper for parallel adapter and quality trimming, 2013. https://github.com/faircloth-lab/illumiprocessor.

Faircloth BC. PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics* 2015;**32**:786–8. https://doi.org/10.1093/bioinformatics/btv646

Faircloth BC, McCormack JE, Crawford NG *et al*. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology* 2012;**61**:717–26. https://doi.org/10.1093/sysbio/sys004

Fan HH, Kubatko LS. Estimating species trees using approximate Bayesian computation. *Molecular Phylogenetics and Evolution* 2011;**59**:354–63. https://doi.org/10.1016/j.ympev.2011.02.019

Fetter KC, Weakley A. Reduced gene flow from mainland populations of *Liriodendron tulipifera* into the Florida Peninsula promotes diversification. *International Journal of Plant Sciences* 2019;**180**:253–69. https://doi.org/10.1086/702267

Fitch HS, Brown WS, Parker WS. *Coluber mormon*, a species distinct from *C. constrictor*. *Transactions of the Kansas Academy of Science* 1981;**84**:196–203. https://doi.org/10.2307/3628274

Freedman AH, Harrigan RJ, Zhen Y *et al*. Evidence for ecotone speciation across an African rainforest–savanna gradient. *Molecular Ecology* 2023;**32**:2287–300. https://doi.org/10.1111/mec.16867

Frichot E, François O. LEA: an R package for landscape and ecological association studies. *Methods in Ecology and Evolution* 2015;**6**:925–9. https://doi.org/10.1111/2041-210x.12382

Frichot E, Mathieu F, Trouillon T *et al*. Fast and efficient estimation of individual ancestry coefficients. *Genetics* 2014;**196**:973–83. https://doi.org/10.1534/genetics.113.160572

Gehara M, Garda AA, Werneck FP *et al*. Estimating synchronous demographic changes across populations using hABC, and its application for a herpetological community from Northeastern Brazil. *Molecular Ecology* 2017;**26**:4756–4771.

Glaudas X, Rodriguez-Robles JA. Vagabond males and sedentary females: spatial ecology and mating system of the speckled rattlesnake (*Crotalus mitchellii*). *Biological Journal of the Linnean Society* 2011;**103**:681–95.

Goodfellow I, Bengio Y, Courville A. *Deep Learning*. Cambridge, MA: MIT Press, 2016.

Greene HW. Taxonomic status of the western racer, *Coluber constrictor mormon*. *Journal of Herpetology* 1984;**18**:210–1. https://doi.org/10.2307/1563756

Guiher TJ, Burbrink FT. Demographic and phylogeographic histories of two venomous North American snakes of the genus *Agkistrodon*. *Molecular Phylogenetics and Evolution* 2008;**48**:543–53. https://doi.org/10.1016/j.ympev.2008.04.008

Gulli A, Pal S. *Deep Learning with Keras*. Birmingham, UK: Packt Publishing, 2017.

Hahn C, Bachmann L, Chevreux B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Research* 2013;**41**:e129–e129. https://doi.org/10.1093/nar/gkt371

Halstead BJ, Mushinsky HR, McCoy ED. Sympatric *Masticophis flagellum* and *Coluber constrictor* select vertebrate prey at different levels of taxonomy. *Copeia* 2008;**2008**:897–908. https://doi.org/10.1643/ce-07-221

Hamilton CA, Hendrixson BE, Brewer MS *et al.* An evaluation of sampling effects on multiple DNA barcoding methods leads to an integrative approach for delimiting species: a case study of the North American tarantula genus *Aphonopelma* (Araneae, Mygalomorphae, Theraphosidae). *Molecular Phylogenetics and Evolution* 2014;**71**:79–93. https://doi.org/10.1016/j.ympev.2013.11.007

Harrington S, Burbrink FT. Complex cycles of divergence and migration shape lineage structure in the common kingsnake species complex. *Journal of Biogeography* 2023;**50**:341–51. https://doi.org/10.1111/jbi.14536

Harvey MG, Smith BT, Glenn TC *et al.* Sequence capture versus restriction site associated DNA sequencing for shallow systematics. *Systematic Biology* 2016;**65**:910–24. https://doi.org/10.1093/sysbio/syw036

He Z, Li X, Yang M *et al.* Speciation with gene flow via cycles of isolation and migration: insights from multiple mangrove taxa. *National Science Review* 2019;**6**:275–88. https://doi.org/10.1093/nsr/nwy078

Hewitt G. The genetic legacy of the quaternary ice ages. *Nature* 2000;**405**:907–13. https://doi.org/10.1038/35016000

Hickerson MJ. All models are wrong. *Molecular Ecology* 2014;**23**:2887–9. https://doi.org/10.1111/mec.12794

Hickerson MJ, Carstens BC, Cavender-Bares J *et al.* Phylogeography's past, present, and future: 10 years after Avise, 2000. *Molecular Phylogenetics and Evolution* 2010;**54**:291–301. https://doi.org/10.1016/j.ympev.2009.09.016

Hoang DT, Chernomor O, Von Haeseler A *et al.* UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* 2018;**35**:518–22. https://doi.org/10.1093/molbev/msx281

Holman JA. *Fossil Snakes of North America: Origin, Evolution, Distribution, Paleoecology (Life of the Past)*. Bloomington, IN: Indiana University Press, 2000.

Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 2006;**23**:254–67. https://doi.org/10.1093/molbev/msj030

Jan G. *Elenco Sistematico degli Ofidi Descritti e Disegnati per L'iconografia Generale*. Milano: Tip. di A. Lombardi, 1863.

Jaynes KE, Myers EA, Gvoždík V *et al.* Giant tree frog diversification in West and Central Africa: isolation by physical barriers, climate, and reproductive traits. *Molecular Ecology* 2022;**31**:3979–98. https://doi.org/10.1111/mec.16169

Jombart T. Adegenet: An R package for the multivariate analysis of genetic markers. *Bioinformatics* 2008;**24**:1403–5. https://doi.org/10.1093/bioinformatics/btn129

Kalyaanamoorthy S, Minh BQ, Wong TKF *et al.* ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* 2017;**14**:587–9. https://doi.org/10.1038/nmeth.4285

Karger DN, Conrad O, Böhner J *et al.* Climatologies at high resolution for the earth's land surface areas. *Scientific Data* 2017;**4**:1–20.

Karger DN, Nobis MP, Normand S *et al.* CHELSA-TraCE21k v1.0. Downscaled transient temperature and precipitation data since the last glacial maximum. *Climate of the Past Discussions* 2021;**2021**:1–27.

Kim D, Bauer BH, Near TJ. Introgression and species delimitation in the Longear Sunfish *Lepomis megalotis* (Teleostei: Percomorpha: Centrarchidae). *Systematic Biology* 2022;**71**:273–85. https://doi.org/10.1093/sysbio/syab029

Leaché AD, Reeder TW. Molecular systematics of the Eastern fence lizard (*Sceloporus undulatus*): a comparison of parsimony, likelihood, and Bayesian approaches. *Systematic Biology* 2002;**51**:44–68. https://doi.org/10.1080/106351502753475871

Lemmon AR, Emme SA, Lemmon EM. Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology* 2012;**61**:727–44. https://doi.org/10.1093/sysbio/sys049

Linkem CW, Minin VN, Leaché AD. Detecting the anomaly zone in species trees and evidence for a misleading signal in higher-level skink phylogeny (Squamata: Scincidae). *Systematic Biology* 2016;**65**:465–77. https://doi.org/10.1093/sysbio/syw001

Lou RN, Jacobs A, Wilder AP *et al.* A beginner's guide to low-coverage whole genome sequencing for population genomics. *Molecular Ecology* 2021;**30**:5966–93. https://doi.org/10.1111/mec.16077

Martin MD, Olsen MT, Samaniego JA *et al.* The population genomic basis of geographic differentiation in North American common ragweed (*Ambrosia artemisiifolia* L.). *Ecology and Evolution* 2016;**6**:3760–71. https://doi.org/10.1002/ece3.2143

McCormack JE, Maley JM, Hird SM *et al.* Next-generation sequencing reveals phylogeographic structure and a species tree for recent bird divergences. *Molecular Phylogenetics and Evolution* 2012;**62**:397–406. https://doi.org/10.1016/j.ympev.2011.10.012

McKenna A, Hanna M, Banks E *et al.* The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 2010;**20**:1297–303. https://doi.org/10.1101/gr.107524.110

Melville J, Chaplin K, Hipsley CA *et al.* Integrating phylogeography and high-resolution X-ray CT reveals five new cryptic species and multiple hybrid zones among Australian earless dragons. *Royal Society Open Science* 2019;**6**:191166. https://doi.org/10.1098/rsos.191166

Minh BQ, Schmidt HA, Chernomor O *et al.* IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 2020;**37**:1530–4. https://doi.org/10.1093/molbev/msaa015

Momigliano P, Florin AB, Merilä J. Biases in demographic modeling affect our understanding of recent divergence. *Molecular Biology and Evolution* 2021;**38**:2967–85. https://doi.org/10.1093/molbev/msab047

Morales AE, Carstens BC. Evidence that *Myotis lucifugus* 'subspecies' are five nonsister species, despite gene flow. *Systematic Biology* 2018;**67**:756–769.

Morgan-Richards M, Bulgarella M, Sivyer L *et al.* Explaining large mitochondrial sequence differences within a population sample. *Royal Society Open Science* 2017;**4**:170730. https://doi.org/10.1098/rsos.170730

Muscarella R, Galante PJ, Soley-Guardia M *et al.* ENMeval: an R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods in Ecology and Evolution* 2014;**5**:1198–205. https://doi.org/10.1111/2041-210x.12261

Myers EA, Burgoon JL, Ray JM *et al.* Coalescent species tree inference of *Coluber* and *Masticophis*. *Copeia* 2017;**105**:640–8. https://doi.org/10.1643/ch-16-552

Myers EA, Bryson Jr RW, Hansen RW *et al.* Exploring Chihuahuan Desert diversification in the gray-banded kingsnake, *Lampropeltis alterna* (Serpentes: Colubridae). *Molecular Phylogenetics and Evolution* 2019a;**131**:211–8.

Myers EA, Xue AT, Gehara M *et al.* Environmental heterogeneity and not vicariant biogeographic barriers generate community-wide population structure in desert-adapted snakes. *Molecular Ecology* 2019b;**28**:4535–48. https://doi.org/10.1111/mec.15182

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

Nosil P. *Ecoloigcal Speciation*. Oxford: Oxford University Press, 2012.

O'Connell KA, Streicher JW, Smith EN *et al.* Geographical features are the predominant driver of molecular diversification in widely distributed North American whipsnakes. *Molecular Ecology* 2017;**26**:5729–51. https://doi.org/10.1111/mec.14295

Oksanen J, Blanchet FG, Kindt R *et al. vegan: Community Ecology Package. R package v.2.3-1*, 2015, 264. https://github.com/vegandevs/vegan.

Orlando L, Metcalf JL, Alberdi MT *et al.* Revising the recent evolutionary history of equids using ancient DNA. *Proceedings of the National Academy of Sciences of the United States of America* 2009;**106**:21754–9. https://doi.org/10.1073/pnas.0903672106

Ortenburger AI. *The Whip Snakes and Racers: Genera Masticophis and Coluber*. Ann Arbor, MI: Memoirs of the University of Michigan Museums, 1928.

Ortiz EM. *vcf2phylip v.2. 0: Convert a VCF Matrix into Several Matrix Formats for Phylogenetic Analysis*. 2019. https://doi. org/10.5281/zenodo **2540861** (1 January 2022, date last accessed).
Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 2004;**20**:289–90. https://doi.org/10.1093/bioinformatics/btg412
Pardo-De la Hoz CJ, Magain N, Piatkowski B *et al*. Ancient rapid radiation explains most conflicts among gene trees and well-supported phylogenomic trees of Nostocalean Cyanobacteria. *Systematic Biology* 2023;**72**:694–712. https://doi.org/10.1093/sysbio/syad008
Pavón-Vázquez CJ, Esquerré D, Fitch AJ *et al*. Between a rock and a dry place: phylogenomics, biogeography, and systematics of ridge-tailed monitors (Squamata: Varanidae: *Varanus acanthurus* complex). *Molecular Phylogenetics and Evolution* 2022;**173**:107516. https://doi.org/10.1016/j.ympev.2022.107516
Pernetta AP, Allen JA, Beebee TJC *et al*. Fine-scale population genetic structure and sex-biased dispersal in the smooth snake (*Coronella austriaca*) in southern England. *Heredity* 2011;**107**:231–8. https://doi.org/10.1038/hdy.2011.7
Phillips SJ, Anderson RP, Schapire RE. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 2006;**190**:231–59. https://doi.org/10.1016/j.ecolmodel.2005.03.026
Pinho C, Hey J. Divergence with gene flow: models and data. *Annual Review of Ecology, Evolution, and Systematics* 2010;**41**:215–30. https://doi.org/10.1146/annurev-ecolsys-102209-144644
Portik DM, Leaché AD, Rivera D *et al*. Evaluating mechanisms of diversification in a Guineo-Congolian tropical forest frog using demographic model selection. *Molecular Ecology* 2017;**26**:5245–63. https://doi.org/10.1111/mec.14266
Provost KL, Myers EA, Smith BT. Community phylogeographic patterns reveal how a barrier filters and structures taxa in North American warm deserts. *Journal of Biogeography* 2021;**48**:1267–83. https://doi.org/10.1111/jbi.14115
Pyron RA, Burbrink FT. Lineage diversification in a widespread species: roles for niche divergence and conservatism in the common kingsnake, *Lampropeltis getula*. *Molecular Ecology* 2009;**18**:3443–57.
Pyron RA, Burbrink FT. Hard and soft allopatry: physically and ecologically mediated modes of geographic speciation. *Journal of Biogeography* 2010;**37**:2005–15. https://doi.org/10.1111/j.1365-2699.2010.02336.x
Rambaut A, Suchard MA, Xie D *et al*. *Tracer v.1. 6. Computer Program and Documentation Distributed by the Author*. 2014. http://beast. bio. ed. ac. uk/Tracer (27 July 2014, date last accessed).
Rambaut A, Drummond AJ, Xie D *et al*. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology* 2018;**67**:901–4. https://doi.org/10.1093/sysbio/syy032
Reid BN, Pinsky ML. Simulation-based evaluation of methods, data types, and temporal sampling schemes for detecting recent population declines. *Integrative and Comparative Biology* 2022;**62**:1849–63. https://doi.org/10.1093/icb/icac144
Remington CL. Suture zones of hybrid interaction between recently joined biotas. *Evolutionary Biology* 1968;**2**:231–428.
Rissler LJ, Apodaca JJ. Adding more ecology into species delimitation: ecological niche models and phylogeography help define cryptic species in the black salamander (*Aneides flavipunctatus*). *Systematic Biology* 2007;**56**:924–42. https://doi.org/10.1080/10635150701703063
Rohlf FJ, Marcus LF. A revolution morphometrics. *Trends in Ecology & Evolution* 1993;**8**:129–32.
Roux C, Fraïsse C, Romiguier J *et al*. Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS Biology* 2016;**14**:e2000234. https://doi.org/10.1371/journal.pbio.2000234
Ruane S. Using geometric morphometrics for integrative taxonomy: an examination of head shapes of milksnakes (genus *Lampropeltis*). *Zoological Journal of the Linnean Society* 2015;**174**:394–413. https://doi.org/10.1111/zoj.12245
Satler JD, Carstens BC. Do ecological communities disperse across biogeographic barriers as a unit? *Molecular Ecology* 2017;**26**:3533–45. https://doi.org/10.1111/mec.14137
Scarsbrook L, Verry AJF, Walton K *et al*. Ancient mitochondrial genomes recovered from small vertebrate bones through minimally destructive DNA extraction: phylogeography of the New Zealand gecko genus *Hoplodactylus*. *Molecular Ecology* 2022;**32**:2964–2984.
Schrider DR, Kern AD. Supervised machine learning for population genetics: a new paradigm. *Trends in Genetics* 2018;**34**:301–12. https://doi.org/10.1016/j.tig.2017.12.005
Singhal S, Grundler M, Colli G *et al*. Squamate conserved loci (SqCL): a unified set of conserved loci for phylogenomics and population genetics of squamate reptiles. *Molecular Ecology Resources* 2017;**17**:e12–e24.
Skalski JR, Richins SM, Townsend RL. A statistical test and sample size recommendations for comparing community composition following PCA. *PLoS One* 2018;**13**:e0206033. https://doi.org/10.1371/journal.pone.0206033
Soltis DE, Morris AB, McLachlan JS *et al*. Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology* 2006;**15**:4261–93. https://doi.org/10.1111/j.1365-294X.2006.03061.x
Steen DA, McClure CJW, Smith LL *et al*. The effect of coachwhip presence on body size of North American racers suggests competition between these sympatric snakes. *Journal of Zoology* 2013;**289**:86–93.
Thorpe RS. Quantitative handling of characters useful in snake systematics with particular reference to intraspecific variation in the ringed snake *Natrix natrix* (L.). *Biological Journal of the Linnean Society* 1975;**7**:27–43. https://doi.org/10.1111/j.1095-8312.1975.tb00732.x
Thuiller W, Georges D, Engler R. *biomod2: Ensemble Platform for Species Distribution Modeling. R package version*, 2013. https://cran.r-project.org/web/packages/biomod2/index.html.
Toews DPL, Brelsford A. The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology* 2012;**21**:3907–30. https://doi.org/10.1111/j.1365-294X.2012.05664.x
Torres-Roig E, Mitchell KJ, Alcover JA *et al*. Origin, extinction and ancient DNA of a new fossil insular viper: molecular clues of overseas immigration. *Zoological Journal of the Linnean Society* 2021;**192**:144–68. https://doi.org/10.1093/zoolinnean/zlaa094
van der Valk T, Pečnerová P, Díez-del-Molino D *et al*. Million-year-old DNA sheds light on the genomic history of mammoths. *Nature* 2021;**591**:265–9. https://doi.org/10.1038/s41586-021-03224-9
Venables WN, Ripley BD. *Modern Applied Statistics with S*. New York NY: Springer, 2002.
Vincent SE, Herrel A, Irschick DJ. Sexual dimorphism in head shape and diet in the cottonmouth snake (*Agkistrodon piscivorus*). *Journal of Zoology* 2004;**264**:53–9. https://doi.org/10.1017/s0952836904005503
Walker MJ, Stockman AK, Marek PE *et al*. 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* 2009;**9**:1–11.
Waltari E, Hijmans RJ, Peterson AT *et al*. Locating Pleistocene Refugia: comparing phylogeographic and ecological niche model predictions. *PLoS One* 2007;**2**:e563. https://doi.org/10.1371/journal.pone.0000563
Warren D, Dinnage R, Matzke N. *ENMTools: Analysis of Niche Evolution Using Niche and Distribution Models. R package v.0.2*, 2017. https://github.com/danlwarren/ENMTools.
Webb SD. Historical biogeography. *Ecosystems of Florida*. Orlando, FL: University of Central Florida Press, 1990, 70–102.
Wickham H, Chang W, Wickham MH. *Package 'ggplot2'. Create Elegant Data Visualisations Using the Grammar of Graphics, v.2*, 2016. 1–189. https://ggplot2.tidyverse.org/reference/ggplot2-package.html.
Wilson JS, Pitts JP. Identifying Pleistocene refugia in North American cold deserts using phylogeographic analyses and ecological niche modelling. *Diversity and Distributions* 2012;**18**:1139–52. https://doi.org/10.1111/j.1472-4642.2012.00902.x
Wright AH, Wright AA. *Handbook of Snakes of the United States and Canada*. Ithaca, NYY: Comstock Publishers Associates.
Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* 2008;**18**:821–9. https://doi.org/10.1101/gr.074492.107