Molecular Phylogenetics and Evolution 162 (2021) 107194

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

How mitonuclear discordance and geographic variation have confounded species boundaries in a widely studied snake

Thomas L. Marshall *, E. Anne Chambers, Mikhail V. Matz, David M. Hillis

Department of Integrative Biology and Biodiversity Center, The University of Texas at Austin, Austin, TX 78712, USA

ARTICLE INFO

Keywords:
Species delimitation
Population genomics
Mitonuclear discordance
Contact zones
Gene flow
Pantherophis

ABSTRACT

As DNA sequencing technologies and methods for delimiting species with genomic data become more accessible and numerous, researchers have more tools than ever to investigate questions in systematics and phylogeography. However, easy access to sophisticated computational tools is not without its drawbacks. Choosing the right approach for one's question can be challenging when presented with multitudinous options, some of which fail to distinguish between species and intraspecific population structure. Here, we employ a methodology that emphasizes intensive geographic sampling, particularly at contact zones between populations, with a focus on differentiating intraspecific genetic clusters from species in the *Pantherophis guttatus* complex, a group of North American ratsnakes. Using a mitochondrial marker as well as ddRADseq data, we find evidence of mitonuclear discordance which has contributed to historical confusion about the relationships within this group. Additionally, we identify geographically and genetically structured populations within the species *Pantherophis emoryi* that are congruent with previously described morphological variation. Importantly, we find that these structured populations within *P. emoryi* are highly admixed throughout the range of the species and show no evidence of any reproductive isolation. Our data support a revision of the taxonomy of this group, and we recognize two species within the complex and three subspecies within *P. emoryi*. This study illustrates the importance of thorough sampling of contact zones and consideration of gene flow when delimiting species in widespread complexes containing parapatric lineages.

1. Introduction

Species delimitation has benefitted both from technological advances in DNA sequencing as well as innovations in analytical methods in the last several years. Researchers have moved beyond making inferences based on one or a few gene trees to using sophisticated approaches that model gene tree coalescence and divergence with varying degrees of gene flow (Chifman and Kubatko, 2014; Degnan and Rosenberg, 2009; Edwards, 2009; Excoffier et al., 2013; Jackson et al., 2017b; Knowles and Carstens, 2007). As tools for these approaches have become more numerous, the importance of choosing the right method for the task in question has become increasingly important (Chan et al., 2017). Multispecies coalescent approaches that model discordance among gene trees due to incomplete lineage sorting (ILS) have become especially popular (Fujita et al., 2012; Jones et al., 2015; Leaché et al., 2014; Yang and Rannala, 2010). However, these approaches are based on models with implicit and explicit assumptions, some of which are rarely tested by users. One assumption of many multispecies coalescent methods is that gene tree discordance is strictly the result of ILS rather than other causes, such as introgressive hybridization (Chambers and Hillis, 2020; Jackson et al., 2017a). Perhaps most importantly, these methods may lack the ability to differentiate within-species metapopulations from true species, especially in groups with geographically structured populations (Sukumaran and Knowles, 2017). In delimiting species, an important step involves assessment of whether identified populations are on independent evolutionary trajectories (de Queiroz, 2007, 1998; Wiley, 1978). Because genetic divergence can occur among populations within species without leading to speciation (Nosil et al., 2009), it is important to use a methodology that is able to differentiate divergent populations within a species from fully isolated lineages.

Taxa that are geographically widespread and have low vagility—or are isolated due to dispersal barriers (i.e., genetically differentiated through isolation by distance [IBD])—can be heavily structured and pose particular problems for methods based on the multispecies coalescent model (Barley et al., 2018; Hedin, 2015; Mason et al., 2020). Here, we argue that prior to model-based species delimitation, one must

* Corresponding author.
E-mail address: thomm80@utexas.edu (T.L. Marshall).

https://doi.org/10.1016/j.ympev.2021.107194
Received 25 September 2020; Received in revised form 12 March 2021; Accepted 28 April 2021
Available online 1 May 2021

first determine whether the taxonomic system is structured at the population level and the extent to which geographic and genetic distances are correlated (Bradburd et al., 2018; Frantz et al., 2009; Meirmans, 2012). By using population genetics tools and thorough population-level sampling—particularly across potential contact zones—we can determine the extent to which space influences population-level genetic structure and assess whether populations are reproductively isolated or are part of a reproductive continuum (Chambers and Hillis, 2020; Mason et al., 2020).

Although modern approaches to species delimitation incorporate gene tree discordance into their frameworks, the currently accepted taxonomy in many groups is based on earlier studies that relied on more limited data. Mitochondrial markers were particularly popular in studies of phylogeography and species delimitation for many years (Burbrink et al., 2000; Hillis and Wilcox, 2005; Kozak et al., 2005; Lemmon et al., 2007). However, even prior to the wide availability of high-throughput sequence data, the observation of mitonuclear discordance was becoming more frequent (Leaché and McGuire, 2006; Papakostas et al., 2016; Toews and Brelsford, 2012; Wiens et al., 2010). Mitochondrial introgression across species boundaries has often been proposed to explain these patterns of discordance between nuclear and mitochondrial genomes (Bryson et al., 2014; Good et al., 2015; Linnen and Farrell, 2007; McGuire et al., 2007; Ruane et al., 2014; Sullivan et al., 2004), and some cases may be attributable to environmentally-driven selection (Morales et al., 2015; Pavlova et al., 2013; Ribeiro et al., 2011). In either scenario, mitochondrial-based inferences about the evolutionary relationships among species and populations in these systems may be misleading. Because this mitochondrial-based legacy is still present in many taxonomic cases, it is important to revisit groups that are currently described based primarily on mitochondrial divergence.

1.1. The Pantherophis guttatus species complex

Here, we examine the *Pantherophis guttatus* complex, a group of North American ratsnakes distributed from the mid-Atlantic coast through the southeastern and south-central United States, parts of the southwestern U.S., and southward to Veracruz, Mexico (Powell et al., 2016). The latest taxonomic revision (in a relatively unstable taxonomic history) was made by Burbrink (2002), who discovered three distinct mitochondrial haplotype groups and thus concluded that this complex consists of three species (Burbrink 2002). These species include the nominate, brightly colored Cornsnake (*P. guttatus*) found in the U.S. east of the Mississippi River; Slowinski's Cornsnake (*P. slowinskii*), described as inhabiting central Louisiana west into the pine forests of east Texas; and the Great Plains Ratsnake (*P. emoryi*), the most drably colored species found through the plains of the central United States and south through much of Texas into Mexico.

Most recently, Myers et al. (2020) used genome-wide data to investigate historical drivers of diversification among these snakes, but questions remain about the relationships within the *P. guttatus* complex. Although Burbrink (2002) and Myers et al. (2020) portray the three species as geographically isolated, a review of specimens in museum collections reveals connectivity between the distributions of *P. slowinskii* and *P. emoryi*. For example, in east central Texas, the putative gap between these two species, there is no break in the distribution of these snakes (Dixon, 2013), and a lack of distinguishing morphological characters makes it difficult to assign individuals in this region to one species or another without sequencing mitochondrial genes—the only reported diagnostic characters. This contact zone is particularly intriguing because, despite morphological similarity and geographic connectivity of these snakes, the mitochondrial haplotypes of *P. slowinskii* and *P. emoryi* are quite divergent. In fact, the mitochondrial tree portrays a sister relationship between *P. slowinskii* and *P. guttatus* that is exclusive of *P. emoryi* (Burbrink, 2002). In contrast, the results of Myers et al. (2020), who used genome-wide nuclear data, support a sister relationship between *P. slowinskii* and *P. emoryi*, indicating the presence of mitonuclear discordance in this group. Because previous molecular studies did not include individuals from contact zones between these taxa, more representative sampling across the range of both species is needed to better determine both the evolutionary relationships and the presence or extent of reproductive isolation between *P. slowinskii* and *P. emoryi*.

Here, we investigate these questions, with a focus on the western contact zone in the *P. guttatus* complex (i.e., between *P. slowinskii* and *P. emoryi*). We use a dataset with thorough sampling of this contact zone and sequence both a portion of the mitochondrial genome as well as genome-wide nuclear loci to investigate the mitochondrial and nuclear genomic structure across the range of these snakes. We consider admixture, IBD, and barriers to gene flow to assess support for the existence of genetic lineages within this group and from this determine whether these putative species fit the criterion of independently evolving metapopulation lineages (de Queiroz 1998, 2007; Frost and Hillis 1990; Wiley 1978).

2. Materials and methods

2.1. Samples

We obtained muscle tissue, liver tissue, and/or shed skins from snakes in the *P. guttatus* complex, as well as two snakes from the *Pantherophis obsoletus* complex. Tissues were obtained from museum and private collections, as well as field collected individuals. All new specimens for this study were collected following the University of Texas at Austin IACUC protocol AUP-2018–00151. DNA was extracted from all tissues using Qiagen DNeasy Blood & Tissue Kits (Qiagen, Valencia, CA, USA) and visualized on a 0.8% agarose gel. A total of 144 samples was submitted for sequencing. Field-collected individuals were preserved with formalin and deposited at the University of Texas at Austin's Biodiversity Center Collections.

2.2. Genomic sequencing and bioinformatics pipeline

We used double-digest restriction site-associated DNA (ddRAD) sequencing (Peterson et al., 2012) to target genome-wide nuclear loci. We first tested five different restriction enzyme pairs with five samples (including outgroup taxa) and submitted digests to the UT Austin Genomic Sequencing and Analysis Facility (GSAF) to quantify results on an Agilent BioAnalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). We chose the enzyme pair *Eco*RI and *Msp*I for DNA digestion because we estimated that these enzymes would target 1–2% of the genome and because they had been successfully used in another group of colubrine snakes (Cox et al., 2018). We then quantified all DNA samples using either a dsDNA high sensitivity assay on a Qubit 3.0 Fluorometer (Life Technologies, Grand Island, NY, USA) or a NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA). Samples ($n = 144$) were standardized to a concentration of 10 ng/μl and submitted to the GSAF for library preparation and sequencing. Size-selected fragments ranging from 225 to 275 bp (average fragment length of 250 bp) were sequenced on an Illumina NextSeq 500 sequencer with 150 bp paired-end reads; to obtain a 6.5X depth of coverage (with an estimated genome size of 1.79 GB), we requested 700,000 reads/sample.

Sequenced reads were demultiplexed and quality filtered using iPyrad v0.9.18 (Steps 1–2; Eaton and Overcast 2019). We used default quality filters with the exceptions of using a stricter filter for adapters (*filter_adapters* = 2) and trimming the edges of all reads (*trim_reads* = 10, 140, 10, 140). We subsequently used two different analysis pipelines to generate: 1) a dataset of genotype likelihoods, which has the advantage of incorporating site uncertainty for relatively low to medium coverage data; and 2) a dataset of called genotypes, which is required for phylogenomic analyses. To obtain genotype likelihood data, we first concatenated an unannotated draft genome of *P. guttatus* (Ullate-Agote et al.,

2014) into 23 pseudo-chromosomes (custom scripts can be found at https://github.com/z0on/2bRAD_denovo). We then mapped all quality-filtered reads to this reference genome using bowtie2 (Langmead and Salzberg, 2012) and converted resulting SAM files into BAM files using SAMtools (Li et al., 2009). We then used the Analysis of Next Generation Sequencing Data software (ANGSD v.0.929; Korneliussen et al. 2014) to filter sites and generate genotype probability data in a variety of output formats. For the called genotype dataset, we used iPyrad (Steps 3–7) to generate a *de novo* assembly using reads with a minimum depth of coverage of 6X (*mindepth_statistical* = 6). The datasets were assembled using the Stampede 2 and Lonestar 5 high performance computing systems at the University of Texas at Austin's Texas Advanced Computing Center (TACC).

2.3. Mitochondrial analyses

To obtain mitochondrial sequence data, we amplified part of the cytochrome *b* (*cytb*) gene from 136 snakes in the *P. guttatus* complex and the two outgroup samples using forward and reverse primers described by Burbrink (2002). Thermocycling conditions for the polymerase chain reaction (PCR) included an initial 3 min of denaturation at 95 °C, followed by 40 cycles consisting of 30 s of denaturation at 95 °C, 30 s of annealing at 47 °C, and 1 min of extension at 72 °C, followed by a final 10 min of extension at 72 °C. Amplicons were visualized on a 0.8% agarose gel and submitted to the DNA Sequencing Facility (DSF) at the University of Texas at Austin for PCR cleanup and Sanger sequencing. Mitochondrial sequence data were trimmed and edited manually and aligned using MUSCLE (Edgar, 2004) in Geneious Prime 2019.0.4 (https://www.geneious.com). To confirm that sequences were edited properly and represented true mitochondrial sequences, we translated them into amino acid sequences in Geneious (this alignment is included in Supplementary Materials). The final alignment consisted of 138 individuals (including two outgroups) and 769 base pairs. We generated a maximum likelihood phylogeny with this alignment using RAxML v.8.2.12 (Stamatakis, 2014) with the gamma-distributed general time reversible model of evolution (*-m* GTRGAMMA). We partitioned the data by codon position and used the automatic bootstrapping criterion *(-N* autoMRE) to select the appropriate number of bootstrap replicates for the dataset, which in this case was 500.

2.4. Phylogenomics

Phylogenetic inferences were conducted using the concatenated dataset that was fully assembled using iPyrad (135 individuals [including two outgroups]; variant and invariant sites). As with our ANGSD-generated dataset, we filtered the data to exclude loci with missing data in over 25% of individuals (iPyrad parameter *min_samples_locus* = 101). We performed a maximum likelihood analysis using RAxML-HPC v.8 (Stamatakis, 2014) on XSEDE on the Cipres Science Gateway (Miller et al., 2011). We used the GTR + G model (*-m* GTRGAMMA) and the automatic bootstrapping criterion (*-N* autoMRE) to select the number of bootstrap replicates (660). Bootstrap values were mapped onto the best tree and visualized with the packages *ape* (Paradis and Schliep, 2018), *cowplot* (Wilke 2017), *ggplot2* (Wickham 2016), and *ggtree* (Yu et al., 2017).

2.5. Population structure

For all population structure analyses, we used the genotype likelihood data produced by ANGSD and filtered the data to include only sites with missing data in no more than 25% of individuals. The program NGSadmix (Skotte et al., 2013) uses a maximum likelihood approach to assign ancestry proportions from each of an *a priori* number of populations to all individuals. Population assignments are calculated based on allele frequencies, and, like other clustering programs, NGSadmix aims to minimize intrapopulation deviations from Hardy-Weinberg equilibrium. We ran NGSadmix on our full nuclear genomic dataset with values of K ranging from 2 to 5. Because genetic clustering analyses can be biased by uneven sampling across populations (Meirmans, 2019), we conducted an additional analysis with even and random sampling of individuals belonging to four populations (36 individuals; see Supplementary Materials for details) and values of K ranging from 2 to 4. We plotted ancestry proportions for each individual using *Pophelper* v2.3.0 (https://github.com/royfrancis/pophelper/blob/master/vignettes/index.Rmd), *ggplot2*, *cowplot*, and *scatterpie* (https://github.com/GuangchuangYu/scatterpie) packages in R (R Core Team 2019).

To assess whether clusters identified by NGSadmix represented discrete populations or were in fact artifacts of IBD, we used conStruct (Bradburd et al., 2018), a clustering method that incorporates a spatial model to account for continuous as well as discrete genetic variation, implemented in R. As input for the analysis, we used a Phylip file produced by iPyrad with one randomly selected SNP per locus (the .usnps file). We removed the two outgroups and then filtered invariant sites using a Python script (https://github.com/btmartin721/raxml_ascbias). We then converted the output file to a STRUCTURE format (.str) using a Perl script (https://github.com/tkchafin/scripts). Because conStruct is sensitive to missing data, we removed two individuals with the highest proportion of missing data in the dataset (referring to the stats.txt file in iPyrad). The resulting input file contained 131 individuals and 3,564 unlinked SNPs. For the spatial model, we input a matrix of pairwise great-circle distances calculated from each sample's GPS coordinates using the *rdist.earth* function in the R package *fields* (Nychka et al., 2015). We performed five independent conStruct runs (values of K ranging from 1 to 5) with each run consisting of 2 chains of 12,000 iterations and an *adapt delta* value set to 0.9. For the run with $K = 5$, we added 3,000 additional iterations per chain to achieve stationarity. To determine the best value of K and test whether spatial or non-spatial models better fit the data, we ran a cross validation analysis with 8 repetitions per K each with 5,000 iterations and a training partition consisting of a random subsampling of 80% of the total loci (*train.prop* = *0.8*). In order to successfully run this analysis, four additional individuals with excess missing data were removed, bringing the total number of individuals in the analysis to 127.

To visualize the genetic dissimilarity among our samples, we performed a Principal Coordinates Analysis (PCoA), or Classical Multidimensional Scaling (MDS). This method uses a pairwise distance matrix among samples as the basis for eigenvector decomposition. In our case, we used a genetic distance (identity by state) matrix generated with ANGSD using the full nuclear genomic dataset. We plotted all individuals along the two eigenvectors (axes) that accounted for the most variation in the data (i.e., had the highest eigenvalues). Eigen decomposition and plotting were performed with the *vegan* package (Oksanen et al., 2019) and visualized using the *ggplot2* and *cowplot* packages in R. We generated two PCoA plots, one in which individuals were color-coded by mitochondrial clade, and another in which individuals were color-coded by population assignment from NGSadmix at $K = 4$ (for admixed individuals, we chose the population of greatest percent ancestry).

To assess relative genetic differentiation between populations identified in the previous analyses, we calculated pairwise F_{ST} values at $K = 4$ using site frequency spectra generated with ANGSD. Individuals were assigned to populations based on ancestry proportions from NGSadmix. For this analysis, we excluded one individual with less than a 0.50 assignment to any one population (TM 164). To filter sites for site frequency spectra, we excluded sites at which greater than 50% of genotypes were considered heterozygotes (likely paralogs) and sites with reads in fewer than 80% of samples. We estimated site frequency likelihoods for each population and generated folded site frequency spectra for all population pairs. Global F_{ST} values were then calculated from each folded site frequency spectrum using ANGSD.

2.6. Species tree estimation

To estimate evolutionary relationships among taxa using a coalescent approach, we generated species trees using SNAPP (Bryant et al., 2012) within BEAST v2.4.7 (Bouckaert et al., 2014). This analysis samples from a posterior distribution of species trees and parameters, which allows for consideration of multiple plausible trees. Because this analysis is computationally intensive and requires unlinked SNPs, we used a subsampled dataset containing five individuals from each of four populations (based on the NGSadmix analysis at $K = 4$) and one randomly sampled SNP per locus (the .usnps output file from iPyrad). Individuals for this dataset were selected based on locality to capture as much variation as possible (see Supplementary Information for details). We conducted the SNAPP analysis with mutation rates μ and ν each set to 1. For the speciation rate prior, we specified a gamma distribution with an alpha shape parameter of 2 and beta scale parameter of 200. The alpha and beta snappriors were set to 1 and 250, respectively. We ran two MCMC chains for 14,000,000 generations each, sampling every 1,000 states, using Beast2 on XSEDE on the Cipres Science Gateway. We combined log files and assessed stationarity with Tracer v1.7.1 (Rambaut et al., 2018). Trees from both chains were combined using LogCombiner and visualized with DensiTree v2.4.7 (Bouckaert and Heled, 2014), using both consensus and all trees. We generated a maximum clade credibility tree with TreeAnnotator (Drummond and Rambaut, 2007) which was then visualized using FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

2.7. Estimating gene flow

Analyses of population structure identify clusters that can either represent populations with distinct evolutionary and demographic histories or simply variation that is a byproduct of geographic distance, known as isolation by distance (IBD; Frantz et al. 2009; Meirmans 2012). To disentangle these two possible scenarios in our dataset, we used the Estimated Effective Migration Surfaces (EEMS) software to visualize geographic areas in which effective migration is lower than predicted by only isolation by distance (Petkova et al., 2016). EEMS partitions the sampling area into a user-defined number of demes and uses Bayesian inference and a stepping-stone model to estimate migration rates between demes based on pairwise genetic distances among samples. Low rates of migration are inferred when genetic distances increase at a relatively high rate across the landscape, accounting for geographic distance between samples. To conduct this analysis, we used the identity by state (IBS) matrix produced by ANGSD and specified coordinates for the boundaries of the sampling area using a custom R script (see Supplementary Materials). We ran two analyses varying deme numbers (200 and 400) to ensure consistency of results. For both runs, we ran one MCMC chain consisting of 10,000,000 iterations following 2,500,000 steps of burn-in, with a thinning interval of 50,000. We performed an additional analysis with a narrowed geographic area encompassing the contact zone between *P. emoryi* and *P. slowinskii* (using an IBS matrix generated with 92 individuals). Parameters for this analysis included 300 demes and one MCMC chain consisting of 15,000,000 iterations following 3,750,000 steps of burn-in, with a thinning interval of 30,000. The results from all EEMS analyses were visualized using the R package *rEEMSplots* (https://github.com/dipetkov/eems/tree/master/plotting/rEEMSplots).

2.8. Population demographics and species delimitation testing

To estimate demographic models for our data, and to test species delimitation hypotheses, we used the software PHRAPL (Jackson et al., 2017a, Jackson et al., 2017b). PHRAPL builds demographic models using parameters specifying migration rates and direction, population size, number of coalescent events, and population growth. Once demographic models have been built, PHRAPL then simulates gene trees based on these models and uses an approximated likelihood framework to compare simulated trees to the empirical gene trees, providing AIC scores based on the number of matches between simulated and empirical trees in order to establish the model of best fit. In this way, PHRAPL is able to incorporate gene flow into models of lineage coalescence.

PHRAPL requires two types of input: an assignment of samples to populations, and rooted gene trees. We used the $K = 4$ results from our NGSadmix analysis (n = 133) for our population assignment. Gene trees for each locus were built using RAxML-ng v.0.8.0 (Stamatakis, 2014; Kozlov et al., 2019) with the GTR + G model on TACC; trees in which there were fewer than five individuals for a given population were removed, and the PHRAPL analysis was performed on 6,430 loci (out of 7,021 total loci).

To reduce computational time, PHRAPL subsamples individuals per population and loci per population. We ran PHRAPL by taking two samples per population (*popAssignments* call within the *GridSearch* function), set seven initial values for divergence time (*CollapseStarts*) and six for migration rate (*MigrationStarts*), 10,000 trees, and 5 subsample replicates per gene. Allowing for symmetrical migration, fixed population sizes and growth, considering only fully resolved trees, a single migration rate, and migration only between geographically adjacent populations (i.e., between *emoryi* "north" and *emoryi* "south", *emoryi* "north" and *slowinskii*, *emoryi* "south" and *slowinskii*, and *slowinskii* and *guttatus*), there were 2,944 possible demographic models. We tested 579 of these models, exploring all possible topologies. All PHRAPL analyses were run on TACC (see Supplementary Materials for all PHRAPL scripts, input data, and output files).

3. Results

3.1. Sampling and bioinformatic analysis

Our final mitochondrial dataset consisted of 136 individuals in the *P. guttatus* complex and two outgroups (Table S1). Nine of these individuals were collected within the range of *P. guttatus*, and the rest of our samples (127 individuals) were from within the ranges of *P. slowinskii* and *P. emoryi*.

From our ddRAD sequencing, we obtained 149,200,844 raw reads for 142 samples, some of which were removed during processing (see Supplementary Materials for details). After processing with the iPyrad pipeline, we had 1,323,981 total sites for 135 samples (including two outgroups), with an average depth of coverage of 26.8. From this dataset, we recovered 72,770 SNPs and 7,021 loci, with 13% missing data (proportion of missing cells). After mapping reads to the *P. guttatus* genome and processing with the ANGSD pipeline, we recovered 65,182 SNPs for 133 samples (no outgroups included).

3.2. Analysis of mitochondrial data

The maximum likelihood analysis of the *cytb* data revealed four distinct mitochondrial clades (Fig. 1b), indicating greater mitochondrial structure than has been previously described in the group (Burbrink 2002). The *P. guttatus* and *P. slowinskii* samples formed a monophyletic clade exclusive of *P. emoryi*, thus exhibiting relationships consistent with Burbrink (2002). Our increased sampling of this group west of the Mississippi River, however, revealed new insight into the distribution of the *P. slowinskii* group. Previously described as inhabiting Louisiana west of the Mississippi River and the woodlands of east Texas, our analysis indicated that the mitochondrial haplotype group associated with *P. slowinskii* extends farther north into Arkansas, eastern Oklahoma, and eastern Kansas. Additionally, we found two distinct mitochondrial clades within *P. emoryi*, with one group distributed across south Texas and another inhabiting the southern Rocky Mountains and central plains of the United States. We refer to these two mitochondrial clades going forward as the *P. emoryi* "north" and *P. emoryi* "south" groups (Fig. 1). Although contact zones among these mitochondrial

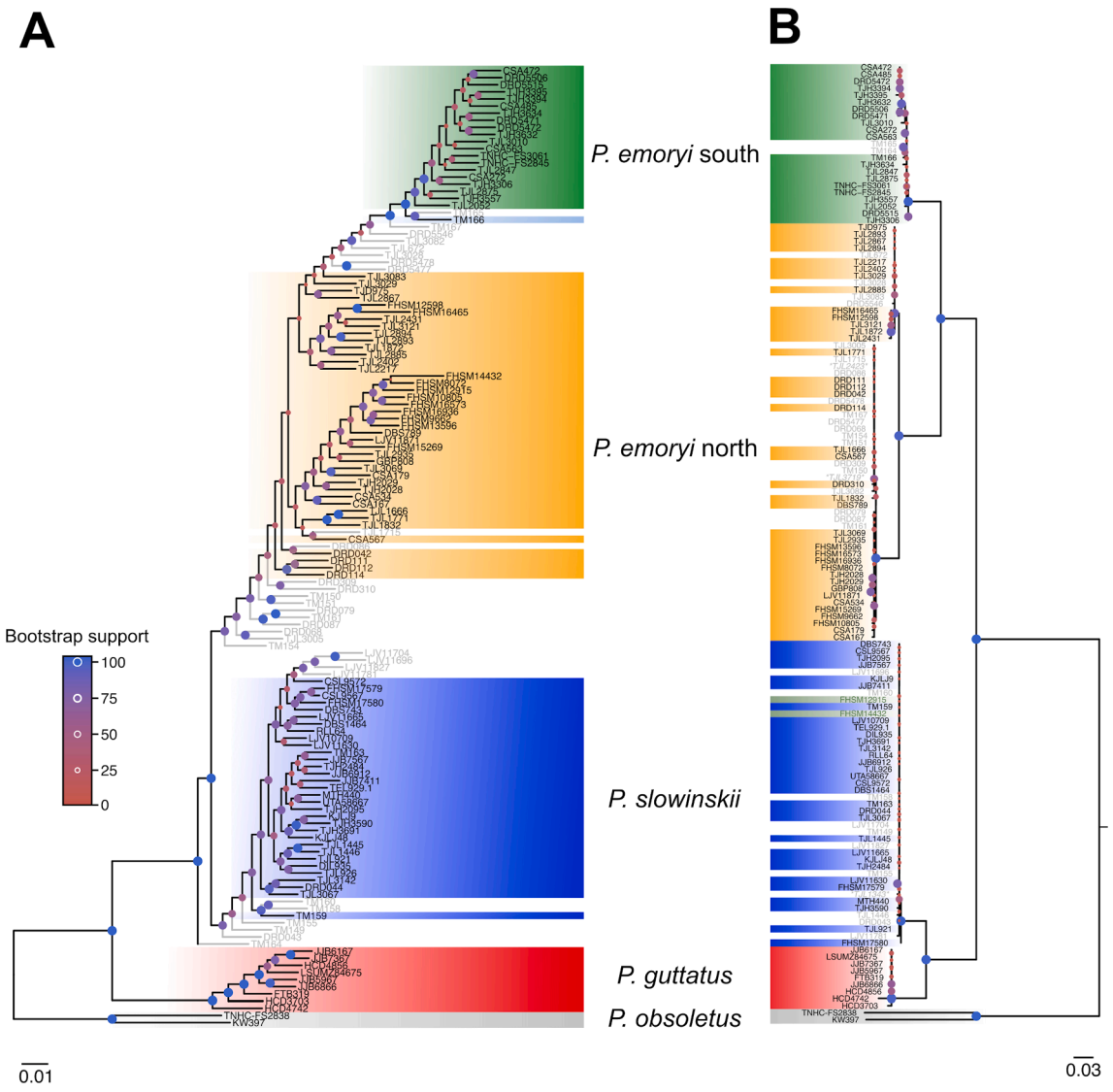

Fig. 1. Maximum likelihood trees reconstructed in RAxML (Stamatakis 2014) using a) 7,021 loci (including variant and invariant sites) from ddRAD data in 135 individuals; and b) 769 bp of the mitochondrial *cytb* gene and 138 individuals. Bootstrap support values are indicated by colored circles at nodes. Samples are colored based on membership to NGSadmix clusters at $K = 4$, with grey samples those that had less than 75% membership to any given cluster. Italicized samples on the *cytb* tree are those which were not included in the ddRAD analysis.

clades exist (e.g., east central Texas), the geographic distributions of these groups appear to be well-defined with little overlap.

3.3. Phylogenomics

The ML analysis of the concatenated ddRAD loci showed two maximally supported monophyletic clades (Fig. 1a), with one clade consisting of individuals of the *P. guttatus* group, and another consisting of all individuals west of the Mississippi River. Within this western clade, a secondary and reasonably well-supported (over 70% bootstrap support for both clades) split was found between many individuals of the *P. slowinskii* and *P. emoryi* groups. However, two eastern Kansas individuals in the *P. slowinskii* mitochondrial group clustered with *P. emoryi*, and individuals sampled near contact zones of these two groups were positioned at the roots of these clades, indicating possible conflicting signal among loci due to introgression (McDade, 1990). Within the *P. emoryi* clade, further division associated with mitochondrial structure was evident, with one particularly well supported (89%) clade consisting of individuals of the *P. emoryi* "south" group. No monophyly was observed for individuals belonging to the *P. emoryi* "north" mitochondrial clade; these individuals were paraphyletic with respect to the *P. emoryi* "south" mitochondrial clade.

3.4. Population structure

Because population clustering programs can yield inaccurate results with uneven sampling (Meirmans 2019), we performed NGSadmix analyses on two different sets of our data to examine how robust our results were to sample input alterations. For our full dataset (133 individuals), we conducted four NGSadmix runs with K ranging from 2 to 5. At values of K below 4, *P. guttatus* individuals clustered with *P. slowinskii* individuals (Supplementary Fig. S1a). At $K = 5$, individuals from the *P. emoryi* "north" group were split into two populations, but most individuals were heavily admixed (Supplementary Fig. S1a). At $K = 4$, clusters were consistent with 1) all individuals from the *P. guttatus* group; 2) most individuals identified as *P. slowinskii* (excluding two eastern Kansas individuals that clustered with *P. emoryi*, congruent with the maximum likelihood phylogeny), with some shared ancestry with *P. guttatus* to the southeast and considerable shared ancestry with *P. emoryi* to the west; 3) most individuals from the *P. emoryi* "north"

group, with some shared ancestry with *P. slowinskii* to the east and *P. emoryi* "south"; and 4) individuals from the *P. emoryi* "south" group, with limited shared ancestry with both *P. slowinskii* and the other *P. emoryi* population at regions of contact (Fig. 2a & S2a).

Because the results from our analysis with the full dataset for values of *K* less than 4 conflicted with prior knowledge of these snakes (e.g., the split between *P. emoryi* populations occurred before *P. guttatus* split from *P. slowinskii*), we suspected that these results might be biased due to uneven sampling. The sampling efforts for this study were focused on collecting snakes west of the Mississippi River, and thus *P. guttatus* was underrepresented compared to other taxa in the complex. We therefore performed an additional analysis with nine randomly subsampled individuals from each of four populations (determined from the results of the analysis with the complete dataset at $K = 4$). In this analysis, the population split at $K = 2$ occurred at the Mississippi River, with all *P. guttatus* individuals assigned to one cluster while all individuals west of the Mississippi River assigned to the other (Supp Fig. S1b). The next splits at $K = 3$ and $K = 4$ revealed considerable admixture between snakes belonging to the *P. slowinskii* and *P. emoryi* groups.

The cross-validation analysis in conStruct assigned the highest likelihood scores to spatial models with four and five layers (Fig. 2b). Because the likelihood curve levels at $K = 4$, we selected a spatial model with $K = 4$ as the ideal fit for the dataset. Although spatial distance does appear to explain some of the genetic variation among individuals, indicating the presence of IBD in this complex, these results nevertheless support the existence of four discrete populations. Results from both chains of the conStruct run using a spatial model with four layers were nearly identical. Although they were similar to the NGSadmix analysis in their geographic placement of clusters, they differ significantly in the degree of admixture seen in the *P. emoryi* "north" and *P. slowinskii* clusters. All individuals belonging to these clusters show significant admixture, with *P. emoryi* "north" individuals sharing a high proportion of ancestry with *P. emoryi* "south", and *P. slowinskii* individuals sharing ancestry with *P. emoryi* "north" (Fig. 2a).

The Principal Coordinates Analysis (PCoA) based on the SNP distance matrix showed the greatest differentiation between *P. guttatus* and all other groups (Fig. S2b). Among individuals west of the Mississippi River, three genetic clusters were evident and corresponded to the clusters identified by NGSadmix and conStruct. Individuals with high proportions of shared ancestry according to NGSadmix grouped closely together in this analysis, while those with less than a 75% assignment to any one cluster were generally intermediate in position between these groupings.

Pairwise F_{ST} values were calculated for populations as defined by NGSadmix at $K = 4$ (Table 1). The highest values were consistently between *P. guttatus* and all other populations, providing further support that the most significant genetic break in this complex occurs at the Mississippi River. Pairwise comparisons between *P. emoryi* "north" and both *P. slowinskii* and *P. emoryi* "south" were comparable.

3.5. Species tree estimation

The consensus species trees as well as the maximum clade credibility tree produced by the SNAPP analysis were well-supported (posterior probabilities for all clades were equal to 1) and closely resembled the ML tree in that *P. slowinskii*, *P. emoryi* "north", and *P. emoryi* "south" formed a clade exclusive of *P. guttatus*. However, these trees differed from the ML topology in that the closest relationship was between *P. slowinskii* and *P. emoryi* "north", with *P. emoryi* "south" falling outside of this clade (Supplementary Fig. S3).

Table 1
Weighted F_{ST} between populations defined by NGSadmix at $K = 4$, calculated from folded site frequency spectra using ANGSD.

From	To	Weighted F_{ST}
P. guttatus	*P. slowinskii*	0.457705
P. guttatus	*P. emoryi* "north"	0.440084
P. guttatus	*P. emoryi* "south"	0.508385
P. slowinskii	*P. emoryi* "north"	0.177738
P. slowinskii	*P. emoryi* "south"	0.339417
P. emoryi "north"	*P. emoryi* "south"	0.201851

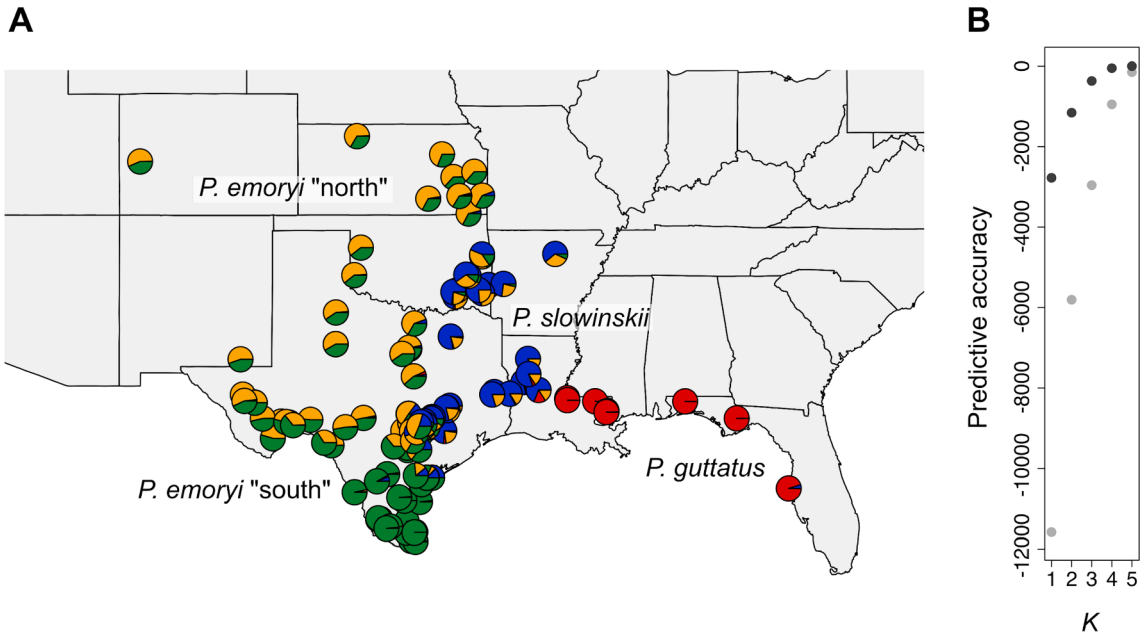

Fig. 2. (a) Population structure of 131 individuals (3,564 SNPs) inferred by conStruct using a spatial model with four layers. Colors correspond to unique ancestry from *P. guttatus* (red), *P. slowinskii* (blue), *P. emoryi* "north" (yellow), and *P. emoryi* "south" (green). (b) Results of the cross-validation analysis in conStruct showing model fit of $K = 1$ to 5 using a spatial (blue) and non-spatial (green) model. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.6. Estimating barriers to gene flow

We performed three different EEMS analyses, including two with different deme numbers (200 and 400) and one with a subsampled dataset focusing on the contact zone between *P. emoryi* and *P. slowinskii*. Plots were consistent across all three analyses, suggesting robust results, with relatively low effective migration across boundaries of the major genetic clusters identified by other population structure analyses (Fig. 3 & Supplementary Fig. S4). Particularly low migration rates were inferred across the Mississippi River, between the range of the *P. emoryi* and *P. slowinskii* clusters, and in south central Texas, separating the *P. emoryi* "north" and *P. emoryi* "south" groups. Effective migration rates within the boundaries of population clusters were generally high, but slightly reduced migration was inferred within the *P. emoryi* "north" group, separating western and central Texas individuals from individuals from the northern Great Plains. The analysis with 400 demes also identified some relatively weak barriers to gene flow within the range of *P. slowinskii* (Fig. S4a).

EEMS also estimates effective diversity within demes by comparing observed to expected genetic similarity between individuals from the same deme. Diversity plots consistently showed the highest within-deme diversity in this complex to be in central Texas (Fig. S4). Areas of low diversity included the range of the *P. guttatus* samples, much of the range of *P. slowinskii*, and areas of higher latitude.

3.7. Estimating population demographics and species delimitation

Using PHRAPL, we ran 579 demographic models against all our loci trees and ranked support for resulting models based on AIC values. There was very little consistency in the PHRAPL results, meaning that many of the ten best-supported demographic models differed widely in their placement and timing of coalescent events. Additionally, the four best-supported models were biologically unfeasible; these models had topologies in which the most recent coalescent event was between one of the *emoryi* populations ("north" or "south") and *guttatus*. The only consistency in our PHRAPL results was that all ten best-supported models all included migration as a parameter, indicating that an isolation-migration model, rather than isolation-only, best fits our data (Table S2). Migration rates were high, with the top five ranked models having minimum migration rates of 4.54, indicating that substantial gene flow likely exists among all lineages.

4. Discussion

Wide-ranging species complexes with varying degrees of regional differentiation are common throughout the tree of life (Bowen et al., 2007; Fontaine et al., 2015; Hillis et al., 1983; Owusu et al., 2015). Those with limited vagility are particularly conducive to the formation of highly structured populations. This strong genetic structuring over geographic space provides an interesting problem for species delimitation because interpopulation structure can often be conflated with interspecific divergence (Sukumaran and Knowles 2017). Genetic divergence in wide-ranging taxa can arise for different reasons, including geographic isolation (either through vicariance or IBD), adaptation to different environments, or a combination of these factors (Coyne and Orr, 2004). In some cases, these drivers of divergence may lead to reproductive isolation, and subsequently speciation, among lineages (Coyne and Orr, 2004). On the other hand, genetic structure at some loci can develop in the presence of continued gene exchange throughout most of the genome (Irwin, 2002, 2012). One solution to this problem is intensive sampling at contact zones of putative species, and the inclusion of genome-wide loci in a population genetics framework to identify whether gene flow is limited, or ongoing, between supposed species. Here, by sampling thoroughly in regions of contact within the *P. guttatus* complex, we assessed whether identified populations could be best described as independently evolving lineages (i.e., species) or intraspecific genetic clusters.

4.1. Mitonuclear discordance

Genome-wide SNPs from extensive sampling across the range of the *P. guttatus* complex confirmed that the major phylogeographic break in this group occurs at the Mississippi River, between *P. guttatus* and all other individuals (i.e., those currently described as either *P. slowinskii* or *P. emoryi*). Although the conclusion of the Mississippi River as a strong phylogeographic barrier supports previous work (Al-Rabab'ah and Williams, 2002; Burbrink et al., 2020; Hoffman and Blouin, 2004; Myers et al., 2020; Near et al., 2001; Soltis et al., 2006), this finding is surprising within this complex because it conflicts with the relationship portrayed by the mitochondrial tree, first illustrated by Burbrink (2002) and supported here, in which *P. slowinskii* clusters with *P. guttatus*, rather than *P. emoryi*. Because the current taxonomy of this complex is based on this discordant mitochondrial gene tree, a renewed look at species boundaries in this group is warranted.

Mitochondrial introgression has been documented in many organisms (Bryson et al., 2014; Good et al., 2015; Linnen and Farrell, 2007; McGuire et al., 2007; Ruane et al., 2014; Sullivan et al., 2004), and the geographic pattern of discordance in the *P. guttatus* complex is more consistent with a pattern of introgression from *P. guttatus* into *P. slowinskii* rather than incomplete lineage sorting (Toews and

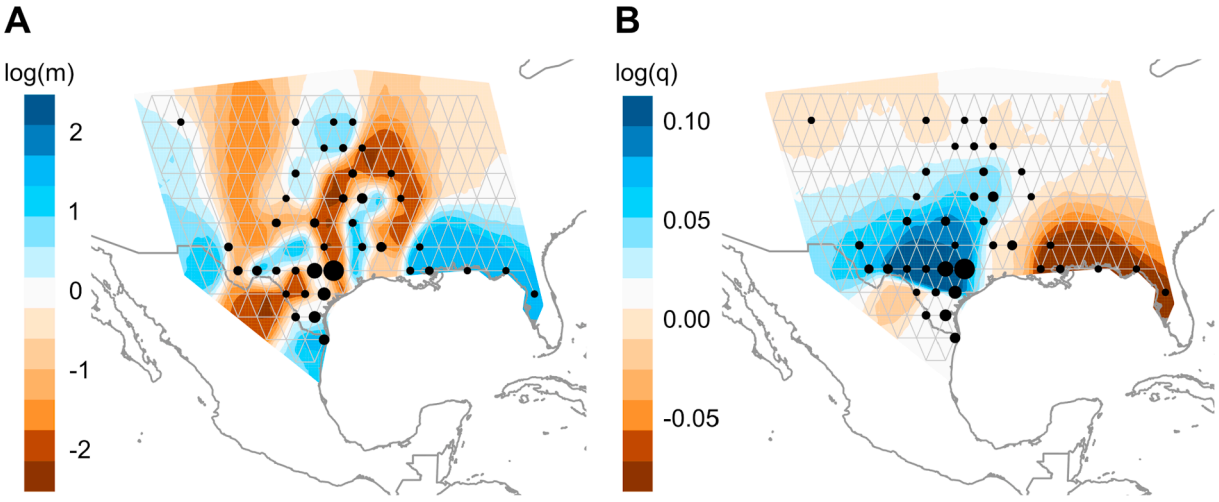

Fig. 3. EEMS (Petkova et al. 2016) analysis for 200 demes using the full nuclear dataset. a) Migration rate surface. b) Diversity rate surface.

Brelsford, 2012). The study of mitonuclear coevolution and its relevance in shaping species boundaries is a growing area of research (Burton et al., 2013; Morales et al., 2018; Sloan et al., 2017), and an interesting question in this system is whether certain nuclear genes associated with mitochondrial function may have co-introgressed with the mitochondrial genome. Further study could investigate whether this introgression might be adaptive.

4.2. Species delimitation

There are many ways to approach species delimitation. Although multispecies coalescent-based approaches have gained popularity in recent years (Fujita et al., 2012; Jones et al., 2015; Leaché et al., 2014; Yang and Rannala, 2010), they are also known to over-split taxa in certain conditions (Chambers and Hillis, 2020; Mason et al., 2020; Sukumaran and Knowles, 2017). We are following the lead of recent authors in choosing a relatively conservative approach (Chan et al., 2017; Devitt et al., 2019; Georges et al., 2018; Natusch et al., 2020) to identify lineages based on independent evolutionary trajectories. Although some of these authors incorporated coalescent-based approaches as part of their species delimitation process, they ultimately recognized fewer taxa than these methods suggested based on the results of other analyses. Here, we chose not to use a multispecies coalescent-based method and instead incorporated the results of several population structure and phylogenomic analyses to clarify taxonomic relationships in the *P. guttatus* complex.

Whereas the major split in this group occurs at the Mississippi River, our results also indicate the presence of substantial population structure within the western group. Among the populations west of the Mississippi River, we identified three distinct mitochondrial clades, and nuclear genomic structure largely corresponded with the mitochondrial structure, with the caveat that some individuals exhibited mitonuclear discordance and most individuals had some degree of admixed ancestry (Fig. 2). Nevertheless, the results of the PCoA and particularly the EEMS analysis strongly supported the presence of three structured populations west of the Mississippi River, consisting largely of 1) individuals in the *P. slowinskii* group, 2) individuals in the *P. emoryi* "north" group, and 3) individuals in the *P. emoryi* "south" group.

At first glance, results of the NGSadmix and, particularly, the conStruct analyses, with geographically widespread admixture in areas where populations converge (and beyond), and the EEMS analysis, with strong barriers to gene flow congruent with population boundaries, appear contradictory. However, it is important to note that EEMS estimates effective, not actual, migration, and assumes equilibrium across time (Petkova et al., 2016). One explanation for the apparent discrepancy between these analyses is that the genetic distances between individuals in each population that lead to inferences of low effective migration could be the result of historical isolation among populations. In such a scenario, admixture proportions may reflect more recent gene flow after populations have come into contact. An alternative explanation is that the three recognized populations represent lineages that have diverged in the presence of gene flow perhaps due to environmental adaptations. Low effective migration may also be inferred when population densities vary across geographic space (Petkova et al. 2016). Areas of low population density can create the appearance of migration barriers because individuals on either side of these areas will be more genetically distant than expected under isolation by distance. We also note that heavily sampled contact zones tended to show weaker barriers to gene flow compared to less densely sampled areas, which could be an artifact of sampling bias. Our PHRAPL results, which were conducted to estimate demographic history given our gene trees, were largely internally inconsistent across models, and so we cannot confidently draw conclusions from these analyses. It is possible that this inconsistency may have resulted from each locus having low phylogenetic signal, thus resulting in PHRAPL being unable to determine the model of best fit given low-signal trees. In any case, we do not see sufficient evidence to consider the lineages west of the Mississippi River to be on independent evolutionary trajectories. Even *P. slowinskii* fails to maintain any degree of reproductive isolation when in contact with *P. emoryi* populations. This can be seen in the broad area of admixture between *P. slowinskii* and *P. emoryi* populations that extends from the Texas coast northward through east central Texas, Oklahoma, and into eastern Kansas, and northern Arkansas (Fig. 4). We found no individuals of *P. slowinski* that were not extensively admixed with genes from the other populations of *P. emoryi*. There is extensive intergradation among *P. emoryi* "north," *P. emoryi* "south," and *P. slowinskii*. The *P. emoryi* "north" metapopulation appears roughly equidistant to both *P. slowinskii* and *P. emoryi* "south" in the PCoA plot, an observation that is supported by similar F_{ST} values between each population pair. Although *P. emoryi* "north" and *P. emoryi* "south" form a monophyletic clade in the ML tree, the species trees produced by SNAPP show a closer relationship between *P. emoryi* "north" and *P. slowinskii*. These results make it difficult to consider *P. emoryi* and *P. slowinskii* as separate species without considering a further split within *P. emoryi*, which does not appear warranted at the species level given the widespread intergradation among the various geographically defined groups. We also have no evidence in our dataset of individuals of pure ancestry from any of the three populations existing in sympatry. This indicates that these regions of contact should not be considered hybrid zones between species, in which pure parental forms mate and produce F1 hybrids, which in turn produce F2 hybrids as well as backcrosses to the parental populations. Rather, these contact zones are more indicative of broad bands of intergradation, in which ancestry proportions gradually shift from one population to another.

4.3. Consistency with morphology

Our recognition of three genetic clusters west of the Mississippi River, including the discovery of a divergent population in south Texas, is strikingly congruent with morphological variation observed in this complex across Texas (Vaughan et al., 1996). This variation has previously led to the description of three subspecies in the state based on dorsal blotches, ventral and caudal scales, ventral pigmentation, and subcaudal stripes. Although these snakes were assigned to *Elaphe guttata* at the time of the study and these authors characterized the east Texas subspecies as *E. g. guttata*, the geographic distribution of the three subspecies is in strong agreement with the results of our molecular analyses. Furthermore, this morphological analysis also recognized a region of introgression between the north and central Texas subspecies, *E. g. emoryi*, and the south Texas subspecies, *E. g. meahllmorum*, along the southern edge of the Edwards Plateau and in western Texas; this is also congruent with results from our molecular analyses.

4.4. Taxonomic revision

The concept of subspecies has had a contentious history. Criticisms regarding the use of subspecies (named with trinomials) to designate sections of continuous clines into more-or-less arbitrary taxa (e.g., Frost and Hillis, 1990; Frost et al., 1992; Wilson and Brown, 1953) led to the decline in the use and recognition of subspecies over the past several decades. Some authors found little genetic support for subspecies that had been described based on morphology (Ball and Avise, 1992; Burbrink et al., 2000; Zink, 1996), although other authors continued to argue for the utility of recognizing intraspecific geographic variation (e.g., Mayr, 1982; Phillimore and Owens, 2006). There has been a recent resurgence of papers that have argued for the utility of subspecies as a means of recognizing geographically circumscribed lineages that may have been temporarily isolated in the past, but which have since merged over broad zones of intergradation that show no evidence of reproductive isolation (e.g., Georges et al., 2018; Hillis, 2020; Hillis and Wüster, 2021; Natusch et al., 2020). De Queiroz (2020) argued that such lineages within species might also be considered species (which would result in species within species), but noted that some taxonomic means of

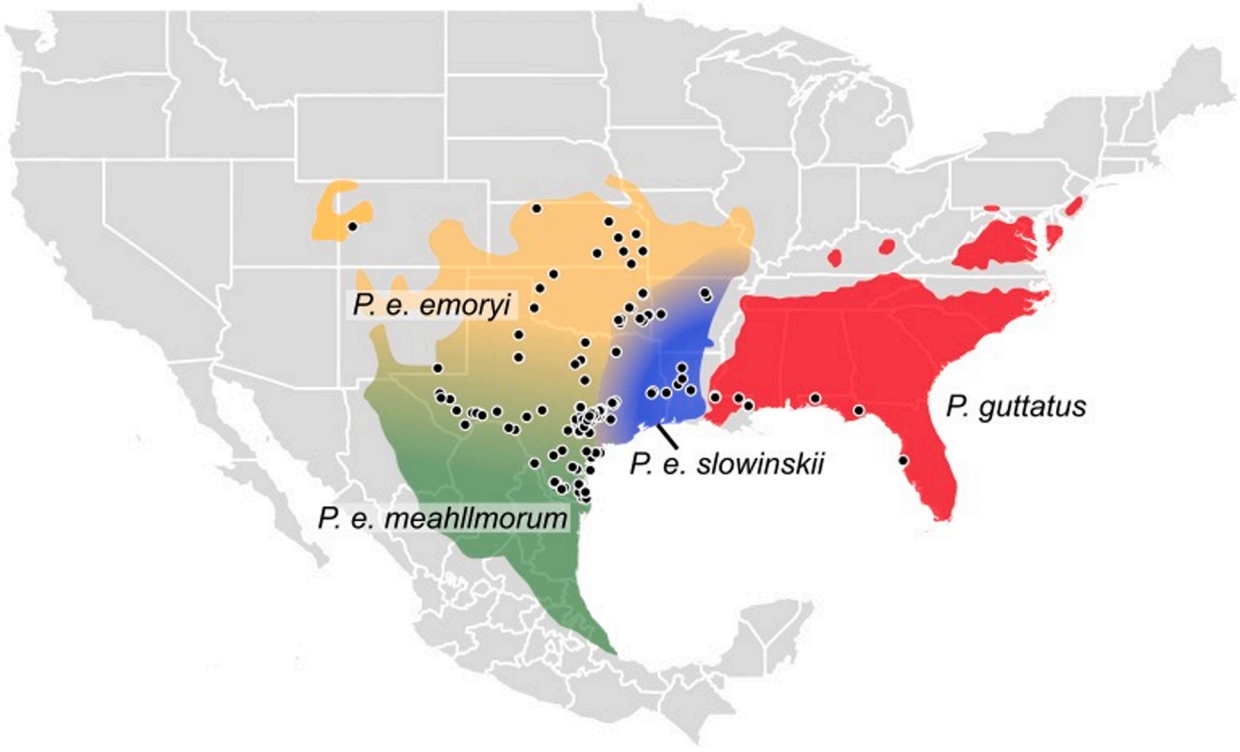

Fig. 4. Suggested range maps and taxonomic revisions based on our analysis for all purported species within the *P. guttatus* complex; gradients between colors represent potential areas of intergradation. Our analysis shows little evidence for *P. slowinskii* being retained as a separate species from *P. emoryi;* thus, we suggest that *P. emoryi* consists of three geographically and genetically structured subspecies: *P. emoryi emoryi, P. e. slowinskii,* and *P. e. meahllmorum.* This revision is also supported by morphological data.

distinguishing these incompletely isolated lineages (such as trinomials) is necessary. Hillis (2021) agreed with de Queiroz (2020) on the utility of designating incompletely isolated lineages within species using trinomials, but argued that the term "species" (and binomial names) should continue to be reserved for the largest metapopulation-level lineages that are evolving independently.

Padial and de la Riva (2020) recently argued that putative taxa should be considered species if they "are found to correspond to phylogenetic lineages, while they are rejected as fabricated taxa if they reflect arbitrary partitions of continuous…variation." If we consider the *P. guttatus* complex using these criteria, only the taxa *Pantherophis guttatus* (Cornsnake or Red Ratsnake, largely east of the Mississippi River) and *Pantherophis emoryi* (Emory's Ratsnake, largely west of the Mississippi River) would be recognized as species, as suggested by Dowling (1951). Although we found evidence of small amounts of admixture between these two lineages in Louisiana, we note that this admixture is only evident on one side of the contact zone (i.e., a few *P. emoryi* individuals in Louisiana have some *P. guttatus* ancestry, but we found no *P. guttatus* individuals with evidence of *P. emoryi* ancestry), and admixture between *P. emoryi* and *P. guttatus* is geographically limited in scope, with both parental types occurring around the contact zone. Other than a potential area of contact in southern Louisiana, these two species appear to be geographically isolated, as *P. emoryi* is seemingly absent from the Mississippi Embayment. All individuals we sampled clearly fell into one of two distinct phylogenetic groups, which we recognize as *P. guttatus* and *P. emoryi*.

Within *P. emoryi*, there are three geographically circumscribed subgroups that do exhibit some degree of genetic divergence and phylogenetic structure, but many sampled individuals are intermediate, and there is clearly continuous genetic variation where these groups come into contact (Fig. 2 & Supplementary Figs. S1 & S2a). Many individuals, especially near contact zones, cannot be classified into one of the three subgroups of *P. emoryi*, because of their intermediacy (Fig. S2). Most eastern individuals of *P. emoryi* exhibit the mitochondrial discordance (closer relationship of mtDNA to *P. guttatus* than to other *P. emoryi*) that was the original basis for the recognition of the proposed species *P. slowinskii* by Burbrink (2002). This introgressed mtDNA is found much further west and north in the range of *P. emoryi* than was reported by Burbrink (2002), and increased sampling of *P. emoryi* shows that the populations called *P. slowinskii* by Burbrink (2002) are neither geographically nor genetically isolated from other populations of *P. emoryi.* Based on the criteria presented by Padial and de la Riva (2020), those authors would not recognize these continuously intergrading geographic and genetic groups within *P. emoryi* as formal taxa. In contrast, we see some utility in recognizing these groups as subspecies of *P. emoryi*, for those who wish to designate the geographic differences in morphology and genetics within the species. The recognition and use of subspecies is optional, so authors who do not wish to recognize subspecies are not required to do so. The geographic clusters within *P. emoryi* have already been named and described, and they exhibit some consistent morphological and genetic divergence, except where they come into contact and gradually intergrade. For those who wish to recognize these three genetically structured groups within *P. emoryi* with formal names, we recommend subspecies names as follows (Fig. 4): (1) Great Plains Ratsnake, *Pantherophis emoryi emoryi*; (2) Thorn-scrub Ratsnake, *Pantherophis emoryi meahllmorum* (originally described by Smith et al., 1994, as *Elaphe guttata meahllmorum*); and (3) Slowinski's Ratsnake, *Pantherophis emoryi slowinskii* (originally described by Burbrink, 2002, as *Elaphe slowinskii*). As is typical with subspecies, many individuals near the intergradation zones cannot be classified to an individual subspecies, but instead can be classified as intergrades between two or more of the subspecies.

4.5. Future work

Future work is needed to better understand the drivers of

diversification within *P. emoryi*. Is divergence between subspecies simply the result of historical separation, or, as indicated by Myers et al. (2020), the result of adaptation to particular ecological niches? The sampling conducted here, which has provided a clearer picture of the geographic distribution of *P. e. slowinskii* and includes individuals from contact zones, is well suited to further investigation of this question. Ecological niche modeling with all three subspecies, including assessment of changes or overlap of niches at contact zones, can provide further clarification on the role of ecological divergence within this species. Full genome or exome sequencing along with cline analyses can also be used to identify genes with steep clines at contact zones and may be involved in environmentally driven selection.

Areas within the range of *P. emoryi* that could benefit from additional sampling include Missouri, New Mexico, Colorado, Utah, and northern Mexico. Based on our current results, we predict that Missouri may represent a zone of intergradation between *P. e. slowinskii* and *P. e. emoryi* (Fig. 4). The western part of the range of *P. emoryi* has also been understudied. Although a putatively isolated population in western Colorado and eastern Utah was previously named as a subspecies (Woodbury and Woodbury, 1942), some have argued that there is no morphological basis for this designation (Smith et al., 1994; Thomas, 1974). We did not find evidence of molecular differentiation between an individual from western Colorado and other *P. e. emoryi* individuals, but we acknowledge that our sampling of this area was limited and could have biased our results. There is also little known about the population structure of *P. emoryi* within Mexico, and sampling in this country could determine whether all individuals belong to *P. e. meahllmorum* or whether there is additional undiscovered diversity within this region.

5. Conclusions

This study illustrates the importance of intensive sampling, particularly at contact zones, and the use of population genomic techniques to clarify species boundaries in wide-ranging species complexes. By examining how lineages interact with one another in contact, we can better determine whether these lineages are independently evolving and should be recognized as species. To remain independent, closely related species must exhibit some degree of reproductive isolation where they come into contact, even if they exhibit narrow geographically-limited tension zones, in which hybrids are produced through direct contact between pure parental forms (Barton and Hewitt 1985). When population ancestry proportions instead shift gradually across geographic space along a large band of contact, with no pure parental classes present, this is indicative of intergradation among geographically differentiated populations, or subspecies.

The *P. guttatus* complex represents a strong case of mitonuclear discordance and could lend itself to further study of the evolution of this phenomenon. Particularly intriguing is the possibility of adaptive introgression of mitochondrial and nuclear genes from *P. guttatus* into *P. emoryi*. Whatever the cause of the phenomenon, this study demonstrates how mitochondrial gene trees can be misleading when used to make inferences about species boundaries and evolutionary relationships. Because much of our currently accepted taxonomy is still based on earlier studies that relied on mitochondrial markers to delimit species, and because this phenomenon appears to be more common than previously thought, it is important to revisit these groups with the sequencing and analytical tools that are now available.

Finally, as with many widespread groups, the *P. guttatus* complex contains extensive genetically structured populations within species, and it is worth discussing how we might choose to recognize such diversity. Though the subspecies concept has had a controversial past, it is a useful system for describing within-species diversity. Recognizing such diversity can be important both for conservation purposes (to recognize populations that are adapted to particular regional environments) and evolutionary studies (to gain insight into how populations diverge and potentially speciate over time). We observed a strong consistency between our molecular results and those of previous morphological studies and argue that high throughput sequencing methods may provide us with the tools to delineate evolutionarily meaningful subspecies within species that contain genetically continuous and connected, but geographically structured, populations.

CRediT authorship contribution statement

Thomas L. Marshall: Conceptualization, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Visualization, Funding acquisition. **E. Anne Chambers:** Formal analysis, Writing - original draft, Writing - review & editing, Visualization. **Mikhail V. Matz:** Software, Resources, Writing - review & editing, Supervision. **David M. Hillis:** Writing - original draft, Writing - review & editing, Supervision, Funding acquisition.

Acknowledgements

We thank Travis J. LaDuc (University of Texas at Austin's Biodiversity Collections), Toby J. Hibbitts (Texas A&M University's Biodiversity Research and Teaching Collections), Curtis J. Schmidt (Fort Hays State University's Sternberg Museum of Natural History), Drew R. Davis (University of Texas Rio Grande Valley), Donna L. Dittman (Louisiana State University's Museum of Natural Science), Brandi S. Coyner (University of Oklahoma's Sam Noble Museum of Natural History), Vicky Zhuang (University of Texas at El Paso), Carl J. Franklin (University of Texas at Arlington Herpetological Collections), Thomas Lott, and Kenneth P. Wray for providing tissue samples. We thank Aaron Short, Cat Bachhofer, and Bill Pierce for providing found specimens, and Analisa Shields-Estrada, Shashwat Sirsi, Amanda Driscoe, and Mackenzie Johnson for assistance with field collection. We are grateful to Jinlin Duan for assistance with specimen preparation and DNA extraction, Isaac Overcast for assistance with the use of iPyrad, Gideon Bradburd for help with conStruct, Brian O'Meara for help running the PHRAPL software, and Spencer Fox for help running PHRAPL on TACC. This work was funded by a National Science Foundation Graduate Research Fellowship (DGE-1144466), the University of Texas at Austin's Blair Scholarship Fund in Zoology, the Texas Herpetological Society, the Texas Academy of Science, and the Alfred W. Roark Centennial Professorship in Natural Sciences at the University of Texas at Austin.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2021.107194. Data files and code can be found at https://doi.org/10.5061/dryad.cc2fqz64d.

References

Al-Rabab'ah, M.A., Williams, C.G., 2002. Population dynamics of *Pinus taeda* L. based on nuclear microsatellites. For. Ecol. Manag. 163 (1-3), 263–271. https://doi.org/10.1016/S0378-1127(01)00584-9.

Ball, R.M., Avise, J.C., 1992. Mitochondrial DNA phylogeographic differentiation among avian populations and the evolutionary significance of subspecies. Auk 109, 626–636. https://doi.org/10.1093/auk/109.3.626.

Barley, A.J., Brown, J.M., Thomson, R.C., 2018. Impact of model violations on the inference of species boundaries under the multispecies coalescent. Syst. Biol. 67, 269–284. https://doi.org/10.1093/sysbio/syx073.

Barton, N.H., Hewitt, G.M., 1985. Analysis of hybrid zones. Annu. Rev. Ecol. Syst. 16 (1), 113–148. https://doi.org/10.1146/annurev.es.16.110185.000553.

Bouckaert, R., Heled, J., 2014. DensiTree 2: Seeing Trees Through the Forest. bioRxiv 012401. doi:10.1101/012401.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., Prlic, A., 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Comput. Biol. 10 (4), e1003537. https://doi.org/10.1371/journal.pcbi.1003537.

Bowen, B.W., Bass, A.L., Rocha, L.A., Grant, W.S., Robertson, D.R., 2007. Phylogeography of the trumpetfishes (*Aulostomus*): ring species complex on a global scale. Evolution 55, 1029–1039. https://doi.org/10.1111/j.0014-3820.2001.tb00619.x.

Bradburd, G.S., Coop, G.M., Ralph, P.L., 2018. Inferring continuous and discrete population genetic structure across space. Genetics 210 (1), 33–52. https://doi.org/10.1534/genetics.118.301333.

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A., RoyChoudhury, A., 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol. Biol. Evol. 29, 1917–1932. https://doi.org/10.1093/molbev/mss086.

Bryson, R.W., Smith, B.T., Nieto-Montes de Oca, A., García-Vázquez, U.O., Riddle, B.R., 2014. The role of mitochondrial introgression in illuminating the evolutionary history of Nearctic treefrogs. Zool. J. Linn. Soc. 172 (1), 103–116. https://doi.org/10.1111/zoj.2014.172.issue-110.1111/zoj.12169.

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

Burbrink, F.T., Gehara, M., Myers, E.A., 2020. Resolving spatial complexities of hybridization in the context of the gray zone of speciation in North American ratsnakes (*Pantherophis obsoletus* complex). Biorxiv 1–64.

Burbrink, F.T., Lawson, R., Slowinski, J.B., 2000. Mitochondrial DNA phylogeography of the polytypic North American rat snake (*Elaphe obsoleta*): A critique of the subspecies concept. Evolution 54 (6), 2107–2118. https://doi.org/10.1111/evo.2000.54.issue-610.1111/j.0014-3820.2000.tb01253.x.

Burton, R.S., Pereira, R.J., Barreto, F.S., 2013. Cytonuclear genomic interactions and hybrid breakdown. Annu. Rev. Ecol. Evol. Syst. 44 (1), 281–302. https://doi.org/10.1146/annurev-ecolsys-110512-135758.

Chambers, E.A., Hillis, D.M., 2020. The multispecies coalescent over-splits species in the case of geographically widespread taxa. Syst. Biol. 69 (1), 184–193. https://doi.org/10.1093/sysbio/syz042.

Chan, K.O., Alexander, A.M., Grismer, L.L., Su, Y.-C., Grismer, J.L., Quah, E.S.H., Brown, R.M., 2017. Species delimitation with gene flow: A methodological comparison and population genomics approach to elucidate cryptic species boundaries in Malaysian Torrent Frogs. Mol. Ecol. 26 (20), 5435–5450. https://doi.org/10.1111/mec.14296.

Chifman, J., Kubatko, L., 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics 30, 3317–3324. https://doi.org/10.1093/bioinformatics/btu530.

Cox, C.L., Davis Rabosky, A.R., Holmes, I.A., Reyes-Velasco, J., Roelke, C.E., Smith, E.N., Flores-Villela, O., McGuire, J.A., Campbell, J.A., 2018. Synopsis and taxonomic revision of three genera in the snake tribe Sonorini. J. Nat. Hist. 52 (13-16), 945–988. https://doi.org/10.1080/00222933.2018.1449912.

Coyne, J.A., Orr, H.A., 2004. Speciation. Sinauer Associates Inc, Sunderland, MA.

de Queiroz, K., 1998. The general lineage concept of species, species crteria, and the process of speciation and terminological recommendations. Endless Forms Species Speciat. 57–75 https://doi.org/10.1080/10635150701701083.

de Queiroz, K., 2007. Species concepts and species delimitation. Syst. Biol. 56, 879–886. https://doi.org/10.1080/10635150701701083.

de Queiroz, K., 2020. An updated concept of subspecies resolves a dispute about the taxonomy of incompletely separated lineages. Herpetol. Rev. 51, 459–461.

Degnan, J.H., Rosenberg, N.A., 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24 (6), 332–340. https://doi.org/10.1016/j.tree.2009.01.009.

Devitt, T.J., Wright, A.M., Cannatella, D.C., Hillis, D.M., 2019. Species delimitation in endangered groundwater salamanders: Implications for aquifer management and biodiversity conservation. Proc. Natl. Acad. Sci. USA 116 (7), 2624–2633. https://doi.org/10.1073/pnas.1815014116.

Dixon, J.R., 2013. Amphibians & reptiles of Texas: with keys, taxonomic synopses, bibliography, and distribution maps. Texas A&M University Press, College Station.

Dowling, H.G., 1951. A taxonomic study of the ratsnakes, genus *Elaphe* Fitzinger. I. The status of the name *Scotophis laetus* Baird and Girard (1853). Copeia 1951 (1), 39. https://doi.org/10.2307/1438049.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7 (1), 214. https://doi.org/10.1186/1471-2148-7-214.

Eaton, D., Overcast, I., 2019. ipyrad: Interactive assembly and analysis of RAD-seq data sets. Available at https://github.com/dereneaton/ipyrad.

Edgar, R.C., 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32 (5), 1792–1797. https://doi.org/10.1093/nar/gkh340.

Edwards, S.V., 2009. Is a new and general theory of molecular systematics emerging? Evolution 63, 1–19. https://doi.org/10.1111/j.1558-5646.2008.00549.x.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V.C., Foll, M., Akey, J.M., 2013. Robust demographic inference from genomic and SNP data. PLoS Genet. 9 (10), e1003905. https://doi.org/10.1371/journal.pgen.1003905.

Fontaine, M.C., Pease, J.B., Steele, A., Waterhouse, R.M., Neafsey, D.E., Sharak, I.V., 2015. Extensive introgression in a malaria vector species complex revealed by phylogenomics. Science 347, 1258524. doi:10.1126/science.1258524.

Frantz, A.C., Cellina, S., Krier, A., Schley, L., Burke, T., 2009. Using spatial Bayesian method to determine the genetic structure of a continuously distributed population: Clusters or isolation by distance? J. Appl. Ecol. 46, 493–505. https://doi.org/10.1111/j.1365-2664.2008.01606.x.

Frost, D.R., Hillis, D.M., 1990. Species in concept and practice: herpetological applications. Herpetologica. 46, 86–104. https://www.jstor.org/stable/3892607.

Frost, D.R., Kluge, A.G., Hillis, D.M., 1992. Species in contemporary herpetology: Comments on phylogenetic inference and taxonomy. Herpetol. Rev. 23, 46–54.

Fujita, M.K., Leaché, A.D., Burbrink, F.T., McGuire, J.A., Moritz, C., 2012. Coalescent-based species delimitation in an integrative taxonomy. Trends Ecol. Evol. 27 (9), 480–488. https://doi.org/10.1016/j.tree.2012.04.012.

Georges, A., Gruber, B., Pauly, G.B., White, D., Adams, M., Young, M.J., Kilian, A., Zhang, X., Shaffer, H.B., Unmack, P.J., 2018. Genomewide SNP markers breathe new life into phylogeography and species delimitation for the problematic short-necked turtles (Chelidae: *Emydura*) of eastern Australia. Mol. Ecol. 27, 5195–5213. https://doi.org/10.1111/mec.14925.

Good, J.M., Vanderpool, D., Keeble, S., Bi, K.e., 2015. Negligible nuclear introgression despite complete mitochondrial capture between two species of chipmunks. Evolution 69 (8), 1961–1972. https://doi.org/10.1111/evo.2015.69.issue-810.1111/evo.12712.

Hedin, M., 2015. High-stakes species delimitation in eyeless cave spiders (*Cicurina*, Dictynidae, Araneae) from central Texas. Mol. Ecol. 24 (2), 346–361. https://doi.org/10.1111/mec.13036.

Hillis, D.M., 2020. The detection and naming of geographic variation within species. Herpetol. Rev. 51, 52–56. https://doi.org/10.1016/S0022-3182(70)80093-0.

Hillis, D.M., 2021. New and not-so-new conceptualizations of species and subspecies: A reply to the "It's species all the way down" view. Herpetol. Rev. 52, 49–50.

Hillis, D.M., Frost, J.S., Wright, D.A., 1983. Phylogeny and biogeography of the *Rana pipiens* complex: a biochemical evaluation. Syst. Biol. 32 (2), 132–143. https://doi.org/10.1093/sysbio/32.2.132.

Hillis, D.M., Wilcox, T.P., 2005. Phylogeny of the New World true frogs (*Rana*). Mol. Phylogenet. Evol. 34 (2), 299–314. https://doi.org/10.1016/j.ympev.2004.10.007.

Hillis, D.M., Wüster, W., 2021. Taxonomy and nomenclature of the *Pantherophis obsoletus* complex. Herpetol. Rev. 52, 51–52.

Hoffman, E.A., Blouin, M.S., 2004. Evolutionary history of the northern leopard frog: reconstruction of phylogeny, phylogeography, and historical changes in population demography from mitochondrial DNA. Evolution 58 (1), 145–159. https://doi.org/10.1111/evo.2004.58.issue-110.1111/j.0014-3820.2004.tb01581.x.

Irwin, D.E., 2002. Phylogeographic breaks without geographic barriers to gene flow. Evolution 56 (12), 2383–2394. https://doi.org/10.1111/evo.2002.56.issue-1210.1111/j.0014-3820.2002.tb00164.x.

Irwin, D.E., 2012. Local adaptation along smooth ecological gradients causes phylogeographic breaks and phenotypic clustering. Am. Nat. 180 (1), 35–49. https://doi.org/10.1086/666002.

Jackson, N.D., Carstens, B.C., Morales, A.E., O'Meara, B.C., 2017a. Species delimitation with gene flow. Syst. Biol. 66, 799–812. https://doi.org/10.1093/sysbio/syw117.

Jackson, N.D., Morales, A.E., Carstens, B.C., O'Meara, B.C., 2017b. PHRAPL: Phylogeographic Inference Using Approximate Likelihoods. Syst. Biol. 66, 1045–1053. https://doi.org/10.1093/sysbio/syx001.

Jones, G., Aydin, Z., Oxelman, B., 2015. DISSECT: An assignment-free Bayesian discovery method for species delimitation under the multispecies coalescent. Bioinformatics 31, 991–998. https://doi.org/10.1093/bioinformatics/btu770.

Knowles, L.L., Carstens, B.C., Weins, J., 2007. Delimiting species without monophyletic gene trees. Syst. Biol. 56 (6), 887–895. https://doi.org/10.1080/10635150701701091.

Korneliussen, T.S., Albrechtsen, A., Nielsen, R., 2014. ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics 15, 1–13. https://doi.org/10.1186/s12859-014-0356-4.

Kozak, K.H., Larson, A., Bonett, R.M., Harmon, L.J., 2005. Phylogenetic analysis of ecomorphological divergence, community structure, and diversification rates in dusky salamanders (Plethodontidae: *Desmognathus*). Evolution 59 (9), 2000–2016. https://doi.org/10.1111/evo.2005.59.issue-910.1111/j.0014-3820.2005.tb01069.x.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9 (4), 357–359. https://doi.org/10.1038/nmeth.1923.

Leaché, A.D., Fujita, M.K., Minin, V.N., Bouckaert, R.R., 2014. Species delimitation using genome-wide SNP Data. Syst. Biol. 63, 534–542. https://doi.org/10.1093/sysbio/syu018.

Leaché, A.D., McGuire, J.A., 2006. Phylogenetic relationships of horned lizards (*Phrynosoma*) based on nuclear and mitochondrial data: Evidence for a misleading mitochondrial gene tree. Mol. Phylogenet. Evol. 39 (3), 628–644. https://doi.org/10.1016/j.ympev.2005.12.016.

Lemmon, E.M., Lemmon, A.R., Collins, J.T., Lee-Yaw, J.A., Cannatella, D.C., 2007. Phylogeny-based delimitation of species boundaries and contact zones in the trilling chorus frogs (*Pseudacris*). Mol. Phylogenet. Evol. 44 (3), 1068–1082. https://doi.org/10.1016/j.ympev.2007.04.010.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25 (16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Linnen, C.R., Farrell, B.D., 2007. Mitonuclear discordance is caused by rampant mitochondrial introgression in *Neodiprion* (Hymenoptera: Diprionidae) sawflies. Evolution 61 (6), 1417–1438. https://doi.org/10.1111/evo.2007.61.issue-610.1111/j.1558-5646.2007.00114.x.

Mason, N.A., Fletcher, N.K., Gill, B.A., Funk, W.C., Zamudio, K.R., 2020. Coalescent-based species delimitation is sensitive to geographic sampling and isolation by distance. Syst. Biodivers. 18 (3), 269–280. https://doi.org/10.1080/14772000.2020.1730475.

Mayr, E., 1982. Of what use are subspecies? Auk 99, 593–595. https://doi.org/10.1093/auk/99.3.593a.

McDade, L., 1990. Hybrids and phylogenetic systematics. I. Patterns of character expression in hybrids and their implications for cladistic analysis. Evolution 44 (6), 1685–1700.

McGuire, J.A., Linkem, C.W., Koo, M.S., Hutchison, D.W., Lappin, A.K., Orange, D.I., Lemos-Espinal, J., Riddle, B.R., Jaeger, J.R., 2007. Mitochondrial introgression and incomplete lineage sorting through space and time: Phylogenetics of crotaphytid

lizards. Evolution 61 (12), 2879–2897. https://doi.org/10.1111/evo.2007.61.issue-1210.1111/j.1558-5646.2007.00239.x.
Meirmans, P.G., 2019. Subsampling reveals that unbalanced sampling affects Structure results in a multi-species dataset. Heredity 122 (3), 276–287. https://doi.org/10.1038/s41437-018-0124-8.
Meirmans, P.G., 2012. The trouble with isolation by distance. Mol. Ecol. 21, 2839–2846. https://doi.org/10.1111/j.1365-294X.2012.05578.x.
Miller, M.A., Pfeiffer, W., Schwartz, T., 2011. The CIPRES science gateway: A community resource for phylogenetic analyses. Proc. TeraGrid 2011 Conf. Extrem. Digit. Discov. TG'11 https://doi.org/10.1145/2016741.2016785.
Morales, H.E., Pavlova, A., Amos, N., Major, R., Kilian, A., Greening, C., Sunnucks, P., 2018. Concordant divergence of mitogenomes and a mitonuclear gene cluster in bird lineages inhabiting different climates. Nat. Ecol. Evol. 2 (8), 1258–1267. https://doi.org/10.1038/s41559-018-0606-3.
Morales, H.E., Pavlova, A., Joseph, L., Sunnucks, P., 2015. Positive and purifying selection in mitochondrial genomes of a bird with mitonuclear discordance. Mol. Ecol. 24 (11), 2820–2837. https://doi.org/10.1111/mec.13203.
Myers, E.A., McKelvy, A.D., Burbrink, F.T., 2020. Biogeographic barriers, Pleistocene refugia, and climatic gradients in the southeastern Nearctic drive diversification in cornsnakes (*Pantherophis guttatus* complex). Mol. Ecol. 29 (4), 797–811. https://doi.org/10.1111/mec.v29.410.1111/mec.15358.
Natusch, D.J.D., Esquerré, D., Lyons, J.A., Hamidy, A., Lemmon, A.R., Moriarty Lemmon, E., Riyanto, A., Keogh, J.S., Donnellan, S., 2020. Species delimitation and systematics of the green pythons (*Morelia viridis* complex) of melanesia and Australia. Mol. Phylogenet. Evol. 142, 106640. https://doi.org/10.1016/j.ympev.2019.106640.
Near, T.J., Page, L.M., Mayden, R.L., 2001. Intraspecific phylogeography of *Percina evides* (Percidae: Etheostomatinae): an additional test of the Central Highlands pre-Pleistocene vicariance hypothesis. Mol. Ecol. 10 (9), 2235–2240. https://doi.org/10.1046/j.1365-294X.2001.01362.x.
Nosil, P., Harmon, L.J., Seehausen, O., 2009. Ecological explanations for (incomplete) speciation. Trends Ecol. Evol. 24 (3), 145–156. https://doi.org/10.1016/j.tree.2008.10.011.
Nychka, D., Furrer, R., Paige, J., Sain, S., 2015. fields: Tools for spatial data. R package version 8.4-1, www.image.ucar.edu/fields. doi:10.5065/D6W957CT.
Oksanen, J., F. Guillaume Blanchet, R.K., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., M. Henry H. Stevens, H.W., 2019. vegan: Community ecology package. R Packag. version 2.5-5.
Owusu, S.A., Sullivan, A. R., Weber, J.A., Hipp, A.L., Gailing, O. Taxonomic relationships and gene flow in four North American Quercus species (Quercus section Lobatae). Syst. Bot. 40, 510-521. doi:10.1600/036364415X688754.
Padial, J.M., de la Riva, I., 2020. A paradigm shift in our view of species drives current trends in biological classification. Biological Reviews (Early view. https://doi.org/10.1111/brv.12676).
Papakostas, S., Michaloudi, E., Proios, K., Brehm, M., Verhage, L., Rota, J., Peña, C., Stamou, G., Pritchard, V.L., Fontaneto, D., Declerck, S.A.J., 2016. Integrative taxonomy recognizes evolutionary units despite widespread mitonuclear discordance: Evidence from a rotifer cryptic species complex. Syst. Biol. 65 (3), 508–524. https://doi.org/10.1093/sysbio/syw016.
Paradis, E., Schliep, K., Schwartz, R., 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35 (3), 526–528. https://doi.org/10.1093/bioinformatics/bty633.
Pavlova, A., Amos, J.N., Joseph, L., Loynes, K., Austin, J.J., Keogh, J.S., Stone, G.N., Nicholls, J.A., Sunnucks, P., 2013. Perched at the mito-nuclear crossroads: Divergent mitochondrial lineages correlate with environment in the face of ongoing nuclear gene flow in an Australian bird. Evolution 67 (12), 3412–3428. https://doi.org/10.1111/evo.12107.
Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E., Orlando, L., 2012. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One 7 (5), e37135. https://doi.org/10.1371/journal.pone.0037135.
Petkova, D., Novembre, J., Stephens, M., 2016. Visualizing spatial population structure with estimated effective migration surfaces. Nat. Genet. 48 (1), 94–100. https://doi.org/10.1038/ng.3464.
Phillimore, A.B., Owens, I.P.F., 2006. Are subspecies useful in evolutionary and conservation biology? Proc. R. Soc. B Biol. Sci. 273 (1590), 1049–1053. https://doi.org/10.1098/rspb.2005.3425.
Powell, R., Conant, R., Collins, J.T., 2016. Peterson Field Guide to Reptiles and Amphibians of Eastern and Central North America. Houghton Mifflin Harcourt Publishing Company, New York.
Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., Susko, E., 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67 (5), 901–904. https://doi.org/10.1093/sysbio/syy032.
Ribeiro, Â.M., Lloyd, P., Bowie, R.C.K., 2011. A tight balance between natural selection and gene flow in a Southern African arid-zone endemic bird. Evolution 65, 3499–3514. https://doi.org/10.1111/j.1558-5646.2011.01397.x.
R Core Team, 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.
Ruane, S., Bryson, R.W., Pyron, R.A., Burbrink, F.T., 2014. Coalescent species delimitation in Milksnakes (Genus *Lampropeltis*) and impacts on phylogenetic comparative analyses. Syst. Biol. 63 (2), 231–250. https://doi.org/10.1093/sysbio/syt099.
Skotte, L., Korneliussen, T.S., Albrechtsen, A., 2013. Estimating individual admixture proportions from next generation sequencing data. Genetics 195, 693–702. https://doi.org/10.1534/genetics.113.154138.
Sloan, D.B., Havird, J.C., Sharbrough, J., 2017. The on-again, off-again relationship between mitochondrial genomes and species boundaries. Mol. Ecol. 26 (8), 2212–2236. https://doi.org/10.1111/mec.2017.26.issue-810.1111/mec.13959.
Smith, H.M., Chiszar, D., Staley II, J.R., Tepedelen, K., 1994. Population relationships in the corn snake *Elaphe guttata* (Reptilia: Serpentes). Texas J. Sci. 46, 259–292.
Soltis, D.E., Morris, A.B., McLachlan, J.S., Manos, P.S., Soltis, P.S., 2006. Comparative phylogeography of unglaciated eastern North America. Mol Ecol. 15, 4261–4293. doi: 10.1111/j.1365-294X.2006.03061.x. Stamatakis, A., 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033.
Sukumaran, J., Knowles, L.L., 2017. Multispecies coalescent delimits structure, not species. Proc. Natl. Acad. Sci. 114 (7), 1607–1612. https://doi.org/10.1073/pnas.1607921114.
Sullivan, J.P., Lavoué, S., Arnegard, M.E., Hopkins, C.D., 2004. AFLPs resolve phylogeny and reveal mitochondrial introgression within a species flock of African electric fish (Mormyroidea: Teleostei). Evolution 58 (4), 825–841. https://doi.org/10.1111/evo.2004.58.issue-410.1111/j.0014-3820.2004.tb00415.x.
Thomas, R.A., 1974. Geographic variation in *Elaphe guttata* (Linnaeus). Unpublished M.S. dissertation. Texas A&M University, College Station.
Toews, D.P.L., Brelsford, A., 2012. The biogeography of mitochondrial and nuclear discordance in animals. Mol. Ecol. 21, 3907–3930. https://doi.org/10.1111/j.1365-294X.2012.05664.x.
Ullate-Agote, A., Milinkovitch, M.C., Tzika, A.C., 2014. The genome sequence of the corn snake (*Pantherophis guttatus*), a valuable resource for EvoDevo studies in squamates. Int. J. Dev. Biol. 58 (10-11-12), 881–888. https://doi.org/10.1387/ijdb.150060at.
Vaughan, R.K., Dixon, J.R., Thomas, R.A., 1996. A reevaluation of populations of the corn snake *Elaphe guttata* (Reptilia: Serpentes: Colubridae) in Texas. Texas J. Sci. 48, 175–190.
Wickham, H., 2016. ggplot2: Elegant graphics for data analysis. Springer-Verlag, New York, New York, NY.
Wiens, J.J., Kuczynski, C.A., Stephens, P.R., 2010. Discordant mitochondrial and nuclear gene phylogenies in emydid turtles: Implications for speciation and conservation. Biol. J. Linn. Soc. 99, 445–461. https://doi.org/10.1111/j.1095-8312.2009.01342.x.
Wiley, E.O., 1978. The evolutionary species concept reconsidered. Syst. Zool. 27, 17–26. https://doi.org/10.2307/2412809.
Wilke, C.O., 2017. cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. R package version (9), 2. https://CRAN.R-project.org/package=cowplot.
Wilson, E.O., Brown, W.L., 1953. The subspecies concept and its taxonomic application. Syst. Biol. 2, 97–111. https://doi.org/10.2307/2411818.
Woodbury, A.H., Woodbury, D.M., 1942. Studies of the rat snake, *Elaphe laeta*, with description of a new subspecies. Proc. Biol. Soc. Washingt. 55, 133–142.
Yang, Z., Rannala, B., 2010. Bayesian species delimitation using multilocus sequence data. Proc. Natl. Acad. Sci. U. S. A. 107 (20), 9264–9269. https://doi.org/10.1073/pnas.0913022107.
Yu, G., Smith, D., Zhu, H., Guan, Y., Lam, T.T., 2017. ggtree: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol. Evol. 8, 28–36.
Zink, R.M., 1996. Comparative phylogeography in North American birds. Evolution 50 (1), 308–317. https://doi.org/10.1111/evo.1996.50.issue-110.1111/j.1558-5646.1996.tb04494.x.