Khakurel, B., Nikolakis, Z. L., Crother, B. I., & Wright, A. M. (2023). Species Delimitation of Eastern Pinesnake complex (*Pituophis melanoleucus*). *Bulletin of the Society of Systematic Biologists*. https://doi.org/10.18061/bssb.v2i1.9423

Investigations

Species Delimitation of Eastern Pinesnake Complex (*Pituophis melanoleucus*)

Basanta Khakurel[1], Zachary L. Nikolakis[1,2], Brian I. Crother[1], April M. Wright[1]

[1] Department of Biological Sciences, Southeastern Louisiana University, [2] Department of Biology, The University of Texas at Arlington

Keywords: eastern Pinesnake, species delimitation, bpp, Bayesian, sub-species, UCEs

https://doi.org/10.18061/bssb.v2i1.9423

Bulletin of the Society of Systematic Biologists

Abstract

The eastern Pinesnake (*Pituophis melanoleucus*) is found throughout eastern United States. Taxonomy in this group has been controversial with several conflicting species designations. Three subspecies of the eastern Pinesnake have prevailed in the literature to their geographic locations and scale coloration: the northern Pinesnake (*P. m. melanoleucus*), the Florida Pinesnake (*P. m. mugitus*), and the Black Pinesnake (*P. m. lodingi*). Within the region, there are several major barriers to dispersal, particularly major river drainage systems and human modification of the longleaf pine habitat. Consistently, a lack of phylogenetic resolution has plagued these taxa in prior studies. The goal of this study was to examine the taxonomic validity of the eastern Pinesnake complex using single nucleotide polymorphisms (SNPs) isolated from ultra-conserved elements (UCEs) in phylogenetic and population genetic approaches. Molecular species delimitation approaches indicated that the population of eastern Pinesnake exhibits population structure across its range that may rise to the level of being new species.

1 Introduction

The southeastern United States is an area with rich biodiversity. Nearly half of the country's reptiles and amphibians are found in this region (Graham et al., 2010). The drainage of major rivers that flow south to the Gulf of Mexico (e.g., Mississippi, Apalachicola, Suwanee) have created differentiating barriers for the biota in the east and the west (Soltis et al., 2006). These complex barriers have provided opportunities for diversification of various flora and fauna in the region. About 20% of the total population of herpetofauna in the region is considered endemic (Graham et al., 2010; Tuberville et al., 2005).

Longleaf pine (*Pinus palustris*) forest, in particular, provides a critical terrestrial habitat for a number of endemic species (Guyer & Bailey, 1993). One of such species is the eastern Pinesnake, *Pituophis melanoleucus*. It has been hypothesized to have as many as three subspecies (Bonett et al., 2017; Stull, 1940). Longleaf pine barrens are generally disturbed with fire, which makes the habitat suitable for Pinesnake population due to rodent movements following disturbances (Burger & Zappalorti, 1988; Zappalorti & Burger, 1985). With increased human activity, pine barrens are constantly under pressure from development and have become increasingly disturbed over the 20*th* century. Development has lead to habitat fragmentation and isolated patches of pine barrens, thus creating habitat islands for eastern Pinesnake across its range (Baxley et al., 2011; Baxley & Qualls, 2009).

The Family Colubridae is the largest family of snakes found in every continent except Antarctica (Cogger et al., 1998). Among the many unresolved phylogenetic relationships in various genus, *Pituophis* is one of them that has many subspecific designations. *Pituophis melanoleucus* is one of the species with three different subspecies. *Pituophis melanoleucus* occurs across a large range of southeastern United States where many geological barriers that may inhibit gene flow exist (Burbrink et al., 2000; Burbrink & Guiher, 2014; McKelvy & Burbrink, 2017; Myers et al., 2020). Examples of barriers are the Apalachicola and Mississippi river drainage which are believed to have created population differences among many groups of organisms (Pyron & Burbrink, 2009; Soltis et al., 2006). Some prior studies have also supported population structure differences across these barriers in other taxa. For example, some populations of tiger salamanders (Church et al., 2003), rat snakes (Burbrink et al., 2000), musk turtles, and snapping turtles (Thomas et al., 2014) exhibit different population structure in the eastern and western side of the Apalachicola barrier, while some populations of catfish show no genetic differences across the region (Avise et al., 1987).

The eastern Pinesnake has a wide range of habitats across the eastern United States and is thought to contain several distinct populations with high degrees of gene flow

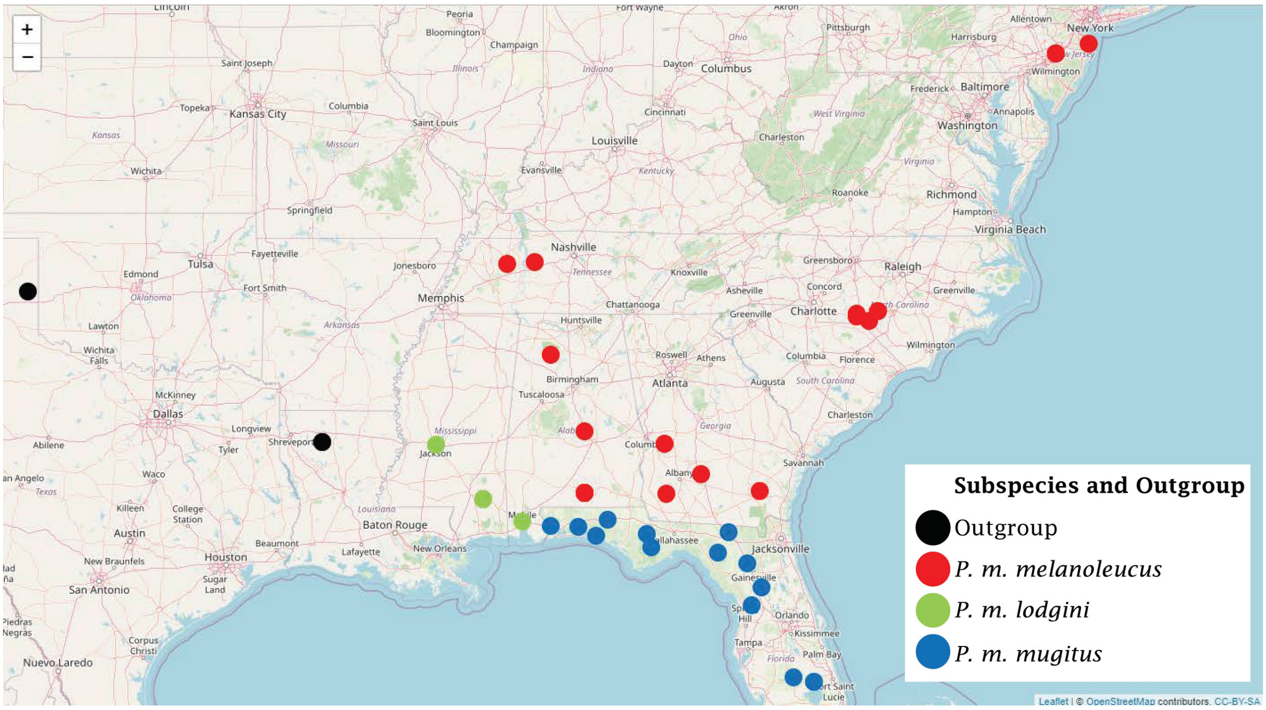


Figure 1. Sampling sites across southeastern United States. The colors indicate the respective subspecies designations commonly used in the Pinesnake literature.

(Nikolakis et al., 2021). This species complex is currently classified with three geographic sub-specific taxa: the northern Pinesnake (*P. m. melanoleucus*), the Florida Pinesnake (*P. m. mugitus*), and the Black Pinesnake (*P. m. lodingi*) (Bonett et al., 2017). These snakes range widely in color from uniformly black to having red/bronze patches (Guyer et al., 2019). The Black Pinesnake has brown or black dorsal coloration, the Florida Pinesnake generally has gray anterior color and rusty brown on the posterior, and the Northern Pinesnake is typically yellow in color with dark blotches in the entire body (Guyer et al., 2019). Besides the difference in coloration, molecular studies have shown that, in this complex, there are some populations that are more closely related to geographically closer populations of different subspecies rather than to other snakes of the same assigned subspecies (Rodríguez-Robles & De Jesús-Escobar, 2000).

Given the uncertainty in phylogenetic relationships and conflicting taxonomic hypotheses, the Eastern Pinesnake is a system ripe for study (Nikolakis et al., 2021; Rodríguez-Robles & De Jesús-Escobar, 2000). Morphological characters, particularly synapomorphies, have typically been considered an important component of determining valid species (Assis & Rieppel, 2011; Mayr, 1981). However, Pinesnake subspecies are not observed to have synapomorphies, and those that are suggested to be synapomorphies can be labile within a species. Scale coloration is one example. Individuals of different subspecies appear to show signs of introgression in the wild in terms of coloration patterns (Dye, 2006; Messenger, 2015). In this case, morphology is not conducive to consistent and reproducible taxonomy.

The populations of Pinesnake are geographically separated due to various barriers. Some of them include natural geographic barriers along its range (rivers and drainages), and some are the results of human activities (development and logging). The geographic barrier systems along the range of Pinesnake include the Alabama river system, the Apalachicola river, and the Suwanee river. In a study from (Reichling, 1995), the Tennessee river valley populations of *P. m. melanoleucus*, to the north of most other Pinesnake populations, could be seen as a geographically isolated from other *P. m. melanoleucus* operational taxonomic units (OTUs). The geographic barriers and previous works have allowed us to test several population grouping hypotheses to study the different populations and subspecies of the eastern Pinesnake. Only one study on population genomics of the *Pituophis melanoleucus* complex has been published to this date (Nikolakis et al., 2021), which examined the role of several riverine systems in Pinesnake diversification. The goal of our study was to resolve the phylogeny of Pinesnake and examine the taxonomic validity of the three subspecies of Pinesnake using species delimitation approaches. In this study, we make use of a dataset of Single Nucleotide Polymorphisms (SNPs) obtained from UCE data, which provided us genome wide structural variation. Ultraconserved elements (UCE) are the highly conserved regions within the genome that are shared among evolutionary distinct taxa (Bejerano et al., 2004) and are flanked by more variable regions.

In absence of traditional markers of species distinction according to the morphological or biological species concepts, we can make use of molecular data to identify cryptic species. Molecular phylogenetic data have a long history of application to species delimitation problems (Donoghue, 1985). In the earliest forms, this took the form of phylogenetic species concept (De Queiroz, 2007), which posited that species were independent lineages on a phylogenetic tree. For our data, we used the multi-species coalescent (MSC) as implemented in *bpp* (Yang & Rannala, 2014), which uses both molecular phylogenetics and population genetics to counter problems such as long-branch attraction and the inherent subjectivity of interpreting the phylogeny (Yang, 2015; Yang & Rannala, 2010). Unlike the traditional phylogenetic methods, which assume that the same tree underlies all gene loci, MSC accounts for coalescent processes in ancient and modern species and the resultant species-gene tree conflicts by allowing for multiple gene trees to underlie the data (Yang, 2015), making this approach very appropriate for our genome-wide data. *bpp* is a computationally-efficient software, which made possible complex hypothesis testing in our data.

Using the SNPs, we were able to perform Bayesian phylogenetic inference, which allowed us to infer the phylogenetic relationships among the population of each putative subspecies sampled across the eastern United States. This tree showed higher support and better resolution than previously-published phylogenies (Nikolakis et al., 2021). We also performed several population genetic analyses to examine the genetic structure and the amount of diversity among the recovered lineages. We tested a set of hypotheses based on the traditional taxonomy in the group (Bonett et al., 2017), suggested geographic substructure in the group (Nikolakis et al., 2021), and the groupings suggested by the phylogenetic tree itself. Using *bpp*, we supported a phylogenetically-informed set of putative species in the Pinesnake clade, firmly rejecting the prior dominant taxonomy.

2 Methods

2.1 Sample collection, DNA extraction, and Bioinformatics

Tissue samples were collected from forty-two specimens of *Pituophis melanoleucus* from their geographical distribution (Figure 1) by Nikolakis, Orton, and Crother (2021). The samples provide coverage of most of the Pinesnake's range, including all the recognized subspecies. Five additional samples were collected from specimens of *P. ruthveni*, *P. catenifer*, and *Pantherophis obsoletus* as outgroup taxa. Genomic DNA was isolated and quantified using Qiagen's DNeasy Kit and Qubit 2.0 using standard protocol as stated by the manufacturer. DNA samples were sent to University of Georgia's Department of Genetics for library preparation and sequencing of UCEs. The samples were de-multiplexed, filtered, and processed by removing adaptor sequences and ambiguous bases using the program *Illumiprocessor*, which is incorporated in the software *Phyluce* v.1.5 (Faircloth, 2015), according to the protocol outlined in (Nikolakis et al., 2021). The UCE loci were used to obtain SNPs and were used in further analyses. We used the best practices workflow and suggested parameters from GATK (McKenna et al., 2010) to call and hard-filter variants while also further excluding sites with low-quality scores and read depth (i.e., $GQ < 30$ and $DP < 5$). We then sampled only the first SNP from each UCE loci to avoid potential impacts of genetic linkage.

2.2 Phylogenetic Analyses

Using the concatenated set of SNPs obtained from the UCEs, we conducted phylogenetic analyses using a nucleotide substitution model from *RevBayes* software v.1.1.1 (Höhna et al., 2016). We used the general time reversible (GTR) model of sequence evolution, which allows six exchangeability rates between nucleotide states. The exchangeabilities are drawn from a Dirichlet distribution with an uninformative prior that allows the data to determine their value. We also used Gamma–distributed among site rate variation to allow different sites to evolve at different rates (Yang, 1994). The MCMC was run to replicate 150,000 generations, and the resulting log files were viewed in *Tracer* v.1.7.1 (Rambaut et al., 2018) to check for convergence. The output files were summarized into maximum clade credibility trees (Helfrich et al., 2018) using *RevBayes*. We inferred trees using individual UCE loci which did not provide enough resolution and consisted of a lot of polytomous clades.

2.3 Population structure

In order to generate additional population substructure hypotheses to test, we conducted exploratory Discriminant Analysis of Principal Components (DAPC) analyses using the R package adegenet (Jombart, 2008). Under this methodology, the researcher assesses the BIC scores to obtain the optimal number of populations. We tested the populations suggested by traditional taxonomy as well as by the phylogeny. We also increased k to see which models produce the best BIC score. In our case, the BIC was minimized by a four-population model, corresponding to geographic barriers previously suggested to be important to Pinesnake dispersal (Nikolakis et al., 2021). Additionally, we used the compoplot module in adegenet to determine the probability of assigning each individual to their population as assigned by DAPC. The obtained population groupings were used to further test different delimitation hypotheses as described in **Species delimitation with *bpp***.

2.4 Species delimitation with *bpp*

To validate the population groupings from DAPC and the phylogenetic tree using species delimitation, we used the software *bpp*. All the *bpp* analyses used a guide tree for the clusters, and we used an inverse gamma prior on tau (3, 0.002) and theta (3, 0.004) as those corresponds to broad and uninformative priors (Flouri et al., 2020). The parameter theta is a measure of heterozygosity in the species, and the parameter tau is the age of the root in the species tree; both the priors depend on the species in the dataset used for the analyses (Flouri et al., 2020). For our analyses, we discarded the first 2,000 samples as burn-in and ran the analyses for 20,000 samples, sampling every 2 generations.

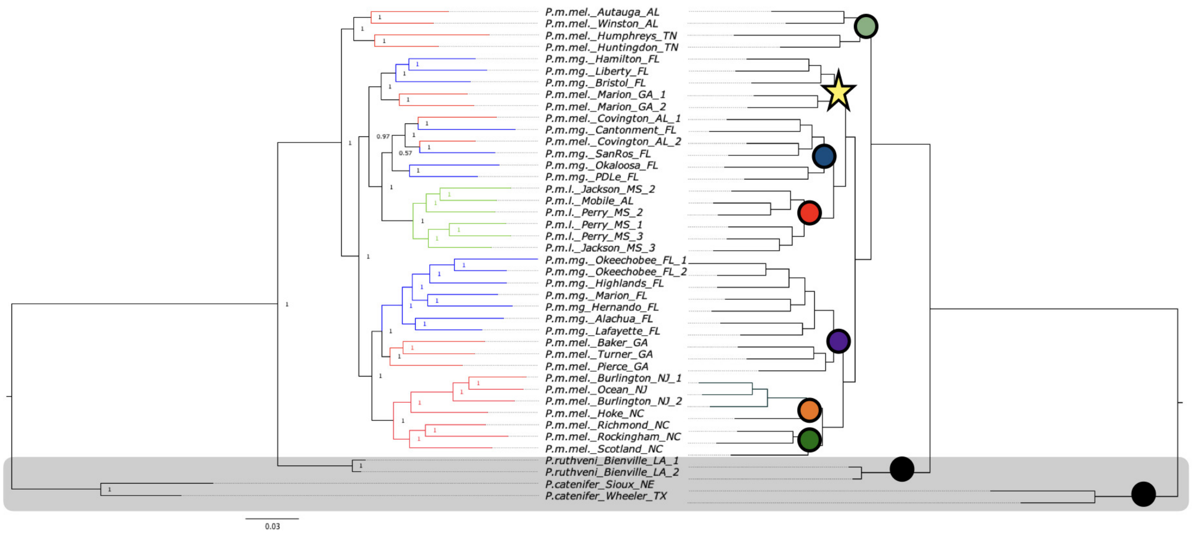
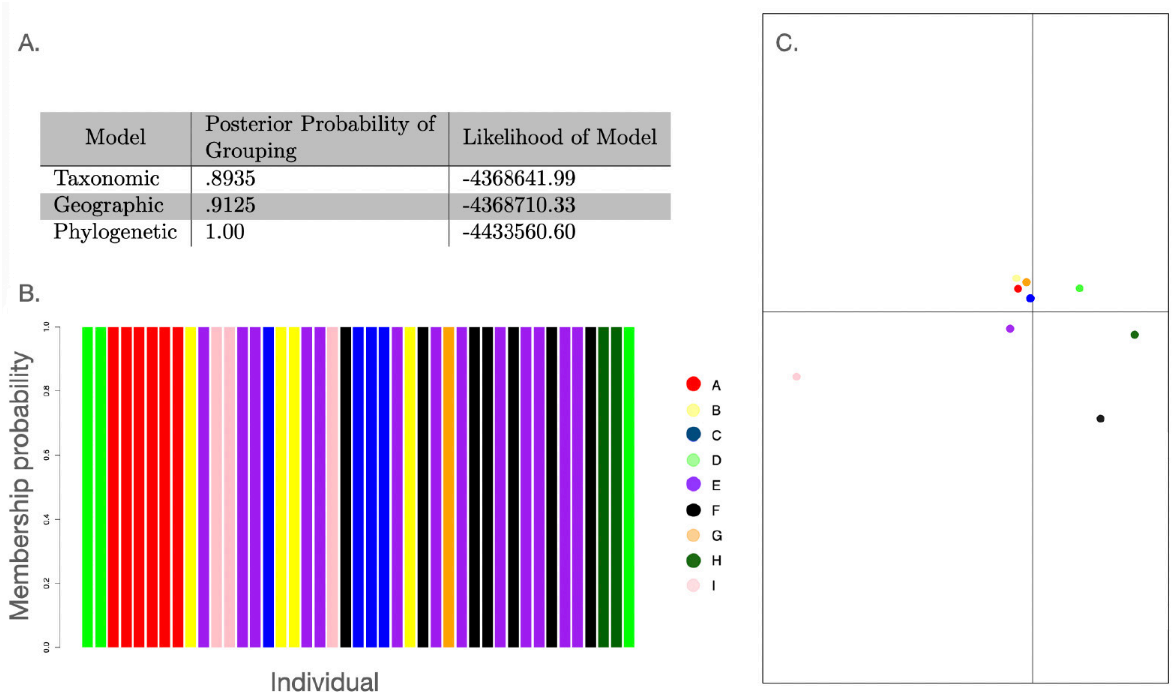


Figure 2. Phylogeny generated using GTR+I+Γ model in *RevBayes*.

At left, The colors of the tips represent current subspecific designations: Black represents the outgroup, red corresponds to *Pituophis melanoleucus melanoleucus* (Northern Pinesnake), blue corresponds to *Pituophis melanoleucus mugitus* (Florida Pinesnake), and green corresponds to *Pituophis melanoleucus lodingi* (Black Pinesnake). The values on the nodes represent the posterior probabilities of the corresponding clade. At right, the colors correspond to designations from BPP and the DAPC analysis. Nodes represented by circles are nodes where DAPC and BPP agree that the clade is distinct. Nodes represented by stars are clades where BPP considers the clade a separate species, but DAPC does not.



Model	Posterior Probability of Grouping	Likelihood of Model
Taxonomic	.8935	-4368641.99
Geographic	.9125	-4368710.33
Phylogenetic	1.00	-4433560.60

Figure 3. Species delimitation results for *Pituophis melanoleucus*.

Panel A shows the support from bpp for different population models with their posterior probabilities and model likelihoods. Panel B shows the population assignments for each sample in the dataset under the phylogenetic model, and Panel C shows a population DAPC of the samples under the phylogenetic model. The legend is common between Panels B and C. These same results for the geographic model can be viewed in Figure S2.

Due to the large size of the data, we tried several different schemas to process the data. First, we performed it using the concatenated SNP dataset. Because *bpp* does not include a variable sites correction (Lewis, 2001), this will tend to overestimate the branch lengths. Therefore, we also did several analyses involving whole UCE loci. We performed the analysis on several subsets of the complete 4,660 UCE dataset with outgroup sequences for *Pituophis catenifer* and *Pituophis ruthveni*. *P. catenifer* and *P. ruthveni* are sister taxa to each other, and, as a whole, they constitute the sister clade to the *Pituophis* complex. We also performed two downsampling experiments (Smith & Carstens, 2019) for our species delimitation analyses. First, we did a random sampling of 100 loci into ten different replicates. Second, we randomly separated the 4,660 loci into four replicates containing 1,165 loci each. *bpp* uses the multi-species coalescent method to estimate the divergence times and population sizes for both extant and ancestral sequences. This information is used to derive the probability distributions indicating whether the lineages can be differentiated from each other. We used the defined algorithm 'A10,' which uses reversible jump MCMC, to test various species delimitation models. We used the clusters obtained from DAPC analyses to group the populations of the subspecies.

In *bpp*, we tested three main hypotheses. The first was a **taxonomic hypothesis**, in which all individuals were binned into populations based on their expected subspecies identity. The second was a **geographic hypothesis** suggested by the DAPC results and prior work (Nikolakis et al., 2021). This hypothesis had four groupings: Far-East (FE), Mid-East (ME), Mid-Atlantic (MA), and outgroup populations as suggested by the DAPC analysis. The group FE consisted of samples from northern Alabama, Tennessee, and Florida; ME consisted of Alabama and Mississippi samples (separated from the FE group by the Alabama riverine system); and MA consisted of North Carolina and New Jersey samples. Finally, we also tested a **phylogenetic hypothesis**, placing major clades into their own categories.

3 Results

3.1 Phylogenetic inference

The full dataset, including the outgroups, had 39,077 SNPs obtained from the UCEs. A phylogenetic tree rooted with the outgroup is depicted in Figure 2. In the resulting maximum clade credibility tree, posterior probabilities for almost all clades were ≥ 0.97, a marked improvement over previously-published trees. The clade with the lowest posterior probability (PP = .57) has vexed other prior authors (Nikolakis et al., 2021). Our result shows a dramatic increase in support for this clade as well as for the separation between the North Carolina and New Jersey taxa from other members of the *P. melanoleucus melanoleucus*. We also supported the Tennessee *P. melanoleucus melanoleucus* samples as grouping with some Alabama *P. melanoleucus melanoleucus* samples, a novel phylogenetic result. While the OTUs that have been referred to as *Pituophis melanoleucus lodingi* are monophyletic, the other two proposed subspecies are not.

3.2 Population Structure and Species Delimitation

3.2.1 Taxonomic and Geographic Hypotheses

We used population genetic metrics to assess genetic variability among the individuals and clusters of the *P. melanoleucus* complex. The first method we employed was a DAPC, which recovered three ingroup clusters. These clusters are distinct and non-overlapping (Figure 2). However, the clusters notably do not align to the subspecies proposed in the literature. These results, instead, suggest that spatial structure is a better descriptor of the species complex, with the Alabama riverine system forming a significant barrier for these snakes. The compoplot indicates that individuals in these clusters typically have a 100% posterior probability of being assigned to their DAPC cluster. Using the results from the DAPC clusters, we separated the populations accordingly to perform species delimitation under the three-species DAPC model.

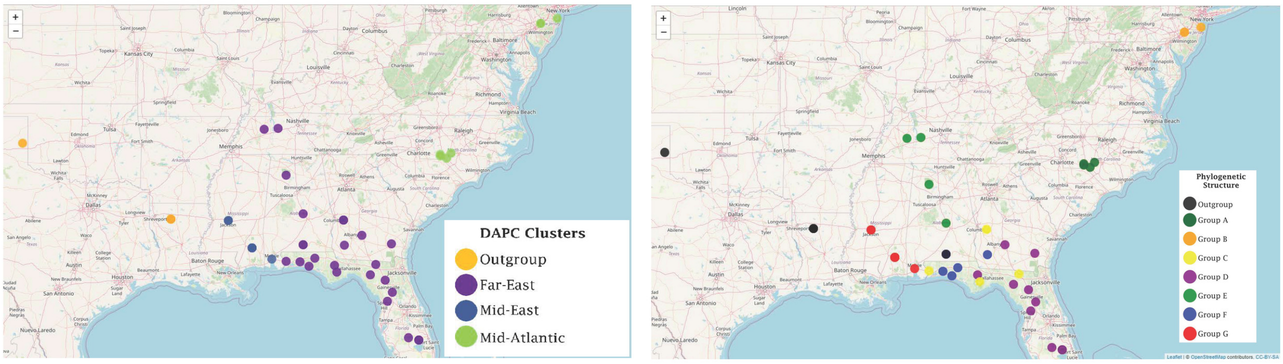
bpp was able to detect some structure in the clusters that we provided from our DAPC analyses. For the 100 UCE runs, there were mixed results among the 10 replicates. Amongst the 10 randomly sampled dataset, only two of them indicated that the three clusters can be considered separate species. Within the eight remaining replicates that did not differentiate all three clusters as different species, two of them grouped FE and ME clusters together; the remaining six of the replicates grouped FE with MA, indicating that ME has already diverged from the complex. All of the replicates have posterior probability ranging from 0.5 to 0.8.

In the four replicates with 1,165 UCEs, two replicates show that FE and MA are sister to each other and that ME is sister to the (FE, MA) group with high posterior probability (< 0.92). One replicate showed that FE is sister to ME and that MA is sister to (FE, ME) with a posterior probability of 0.566. The remaining folder grouped FE and MA together as sister to the ME cluster with posterior probability 0.626.

In the analysis with all 4,660 UCEs, the results indicate the presence of three different ingroup species of Pinesnake, with an Eastern clade (the FE and MA) that is sister to the ME clade. The posterior probability of the presence of three different species (FE, ME, MA) is 0.913. These results can be seen in Figure S2. Because the replicates with all the UCEs included were computationally tractable to run, we opted to use all the data in all further analyses.

3.2.2 Phylogenetic Hypothesis

Upon relaxing the possible number of populations (k) in our DAPC analyses, we observed potential substructures in the overall population of the Pinesnake as suggested by the phylogeny. The groupings suggested by the DAPC analysis match mostly with the phylogeny except for the clade containing Florida panhandle samples and southern Georgia samples as indicated in Figure 3. We grouped each clade into a possible species and performed a *bpp* analyses using the phylogeny as the guide tree. The results from *bpp* agrees with the phylogeny, suggesting that there is a significant structuring in the Pinesnake population. The poste-



A

B

Figure 4. A. The samples map colored according to the geographic hypothesis suggested by the DAPC results. B. The samples map colored per the clusters that can be inferred from the phylogeny. (Figures made using R package *leaflet*)

rior probability of 8 possible species as indicated by our *bpp* analysis is 1.0. This result is shown in Figure 3A. It should be noted that the geographic and phylogentic hypotheses are largely compatible. The phylogenetic hypothesis merely indicates that there are clades within each geographic unit that may be their own species.

4 Discussion

Our analyses suggested that the Pinesnake species complex exhibits levels of population structure across its range. The population structuring could be due to the presence of natural barriers (the Alabama Riverine system) to gene flow and/or due to anthropogenic activities that have isolated the populations. The subspecific designations for the Pinesnake complex have historically been assigned based on the minor differences in body coloration and scale number (Barbour, 1921; Blanchard, 1920). But various studies have concluded that the geographically closer individuals of supposedly different subspecies are more closely related to other individuals of same subspecies (Rodríguez-Robles & De Jesús-Escobar, 2000). Our analyses indicate that there is little validity to the historical subspecific designations used in the literature. We suggest that natural barriers in the Southeast are likely to be responsible for genetic differentiation in the group.

4.1 Phylogeny, DAPC, and bpp

We performed a number of tests of different population models using DAPC, the phylogeny, and *bpp*. The phylogenetic tree strongly contradicted the traditional taxonomy (Bonett et al., 2017) in the group. Many individuals of different subspecies that were geographically closer to each other formed a monophyletic group rather than individuals of same subspecies being grouped together. On our phylogeny, the current taxonomy of Pinesnake is muddled in the tree with different subspecies being sister to other subspecies rather than the ones of the same subspecies. Only *Pituophis melanoleucus lodingi* (Black Pinesnake) is monophyletic on the tree. In addition to that, there is a group of *P. m. melanoleucus* (Northern Pinesnake) in the North Carolina and New Jersey region that appears monophyletic in the tree. All the other *P. m. melanoleucus* and *P. m. mugitus* (Florida Pinesnake) have been placed in various places all across the phylogeny.

After the failure of the traditional taxonomic hypothesis, we used DAPC to formulate alternative hypotheses. The DAPC that minimized the BIC score ($k = 4$) grouped populations such that the major barrier in the range of the eastern Pinesnake could be the Apalachicola-Chattahoochee drainage system in the southern part and anthropogenic activities in the northern (Figure S1). The DAPC analysis with $k = 4$ resulted in three ingroup clusters, with a group in the Atlantic side comprising New Jersey and North Carolina populations; another group of populations from Alabama and Mississippi; and the third group containing populations from Alabama, Tennessee, Florida, and Georgia. The phylogenetic tree also agrees with the clusters from DAPC, rendering population groups of Mid-Atlantic (NJ and NC) and Mid-East (MS and some AL) monophyletic. On the other hand, the Far-East group is rendered non-monophyletic in the tree. This suggests that the smaller groups (Mid-East and Mid-Atlantic) are diverging from a larger (Far-East) group. According to the results from DAPC and the phylogeny, the population structure maps with the Apalachicola-Chattahoochee drainage systems, which splits the monophyletic group of *P. m. lodingi* (Mid-East) from the Far-East group, and the small population of *P. m. melanoleucus* (Mid-Atlantic), which is being isolated due to anthropogenic activities such as logging and urbanization. This hypothesis has been suggested by other authors, as well (Burbrink et al., 2000; Pyron & Burbrink, 2009; Weinell & Austin, 2017). A phylogenetic tree colored according to these groupings can be seen on Figure S1}.

While this hypothesis is feasible, the phylogenetic analysis using *RevBayes* demonstrates highly-supported monophyletic groups within each geographic region (Figure 2). Thus, we decided to relax the number of DAPC clusters to see if there might be more substructure in each group. Upon the relaxation of the number of possible DAPC clusters ($k = 9$), each major group is separated by DAPC, except for one, which is indicated by a star in Fig 3. This clade contains the samples from southern Georgia and the Florida panhandle. This clade has puzzled other studies (Nikolakis et al., 2021) as well. This conflict amongst the Florida and Georgia samples could possibly indicate that there is a significant level of hybridization amongst the individuals of different subspecies in that region. Nearly all of our other clades are supported with a posterior probability of 1.

When examining for possible sub-structuring after evaluating the geographic hypothesis, we relaxed the number of possible DAPC clusters. This further differentiated the clades we had obtained from previous DAPC analyses ($k = 4$). The Mid-East group consisting of Mississippi and some of the Alabama sample remains the same as a cluster. This cluster has historically been referred to as *Pituophis melanoleucus lodingi*, the only subspecies under the taxonomic hypothesis to be monophyletic in our study. The Far-East group (containing samples from Tennessee, Alabama, Florida, and Georgia) is separated into four separate clusters. Group G (Figure 4) is west of the Chattahoochee-Apalachicola and the Alabama Riverine systems. Group E contains samples from Alabama and Tennessee. This group has previously been suggested by Reichling (1995) but is contradicted (with poor support) by other recent taxonomic work (Nikolakis et al., 2021). Florida panhandle and Georgia samples are obtained as a separate cluster (Group F, within and just west of the Chattahoochee-Apalachicola Riverine system), and some Florida panhandle samples are grouped together with Alabama samples (Group C). The remaining clusters contained samples from Florida peninsula (Group D), east of the Chattahoochee-Apalachicola Riverine system. All the samples had a membership probability of 1 as shown in the compoplot (Figure 3B). Using the clusters obtained from the DAPC analyses ($k = 9$), the *bpp* analysis indicated a posterior probability of 1.0 for eight species (both outgroup clusters grouped into one cluster). This suggests that there is further population structure in the Pinesnakes, related to the river drainages in the Southeast and habitat fragmentation to the north.

We find significant structure related to the loss of longleaf habitat. What was called Mid-Atlantic group under the geographic hypothesis is separated into the New Jersey and North Carolina clades under the phylogenetic hypothesis. This separation makes a good deal of sense due to habitat fragmentation and loss of longleaf habitat, especially in Virginia, Maryland, and Delaware (Noss, 1989). Lack of prescribed burns have been impacting the health of remaining pine forests which has in turn been affecting the populations of *P. melanoleucus*. (Malik et al., 2014; Noss, 1989). Our results suggest that habitat fragmentation and loss of pine forests has become a major driving force in population structure in the Northern reaches of the *P. melanoleucus* complex.

Our phylogenetic analyses refute the traditional taxonomic hypotheses, suggesting more gene flow among the members of different subspecies that are geographically closer. Previous squamate phylogeographic studies across these regions (Burbrink et al., 2000; Myers et al., 2020; Soltis et al., 2006) indicated that there is significant genetic variation among clades that existed across a variety of barriers in the region. From our work, we can infer that there is population structure in the Pinesnake, some of which may rise to the level of naming a new species. For example, the groups Mid-East and Mid-Atlantic containing populations of Black and Northern Pinesnake are supported as a distinct group in every analysis that we ran, suggesting that this group may be a separate species. However, final delimitation of these species requires further knowledge about the morphology, ecology, and movement of the organisms (Sukumaran & Knowles, 2017), we cannot be certain that the population structure we observe in the Pinesnake range are distinct species. In particular, more sampling in the Alabama and Chattahoochee Riverine systems would be an illuminating look at some of the more muddled clades.

4.2 UCE in Species Delimitation

The utility of UCEs has been well-demonstrated in deeper timescales (Faircloth, 2015; Gilbert et al., 2015; McCormack et al., 2013) but has been less explored in reptiles for recent evolutionary events (Harvey et al., 2016; Mason et al., 2018; Winker et al., 2018). As UCEs are markers that are conserved across a large range of taxa, they are highly conserved regions within the genome that are shared among evolutionarily distant taxa (Bejerano et al., 2004). Having a large set of UCE loci led us to first attempt using SNPs for some analyses which proved valuable but for species delimitation using *bpp*, due to the lack of invariant sites, the branch lengths were inflated in our trees. As suggested by the scientific literature (Bryant & Hahn, 2020; Roch & Steel, 2015), concatenation approaches do not produce a reliable result. This is something we experienced with our study; concatenating all UCEs and even using concatenated variable sites only data (SNPs) produced results that are inconclusive, whereas treating each UCE locus as separate showed some structure in the populations of Pinesnake across their range (Figure S2A).

bpp can accommodate many loci without concatenation, making it an ideal tool for exploring this problem. Therefore, we decided to try down-samples of different numbers of loci to test how it would influence the results we had. In addition to concatenating the dataset, we treated each UCE as a separate locus. We found that, as more data are added, more population structure is uncovered using *bpp* (Figure S2A). Most of the results in the species delimitation analyses recognize each cluster (Far-East, Mid-East, and Mid-Atlantic) as separate evolutionary entities. In the analyses with ten replicates of 100 loci each, there was indication of Far-East and Mid-Atlantic as a single cluster and Mid-East being a separate population group. This separation reflects the barrier caused by Apalachicola-Chattahoochee drainage system as a major separation in the range of *Pituophis*. In the analysis with four replicates of 1,165 loci each, there was an indication of both Mid-East and Mid-Atlantic being a separate evolution entity in different analyses. In one of the analyses, Mid-Atlantic is grouped together with Far-East, with Mid-East as the sister to the group of MA and FE. All the results from these analyses indicate that both the Apalachicola-Chattahoochee drainage system and urbanization act as a major barrier in the range of Pinesnake. One thing to keep in mind is the type of data we used for these studies. The UCE data are fairly conserved across a long range of taxa which could lead to different results.

As depicted in Figure S2A, using fewer loci caused more uncertainty in our results. In a recently diverged group of Pinesnake, addition of more data is crucial to have enough variability among sites to support divergences on the tree. The posterior probabilities of proposed clusters were also higher when we added more loci to our analyses. Because of these results, for testing our phylogenetic hypothesis, we used the full, non-concatenated set of loci, and we suggest other authors do the same.

UCEs are shared across a large range of evolutionarily distant taxa, and this sharing of markers in evolutionarily distant taxa might have influenced our study. As shown on Figure S2A, use of all the data is especially crucial for this particular question. This is because of hypothesized recent evolution of eastern Pinesnake from other *Pituophis* (Pyron & Burbrink, 2009). The oldest fossils of *P. melanoleucus* have been found in Florida dating 0.8 to 2.5 million years (myr), and more northern fossils from Pennsylvania have been dated from 0.1 myr (Holman, 2000). These dates indicate that the lineage diverged during the late Pleistocene when there were environmental fluctuations leading to periods of isolation and connection. This could also be an important factor that maintained gene flow through the contacts of different populations. Another possible factor that contributed in the gene flow of the Pinesnake are the large home ranges of these snakes (Kapfer et al., 2010). Our study indicates that the eastern Pinesnake has shown some population structuring due to the geographic barriers existing across its range.

While UCEs were the tool available to us for this study, a sequencing technology that captures more variability might be helpful for discriminating between the phylogenetic and geographic hypotheses for a full taxonomic revision.

5 Conclusions

Using a dataset of UCEs, we found evidence that there is geographic and phylogenetic substructure in populations of *Pituophis melanoleucus* across its range. Using the phylogenetic tree, we reject the traditional three-subspecies taxonomy in *P. melanoleucus*. We find that the groups typically referred to as the Florida Pinesnake (*P. m. mugitus*), the Black Pinesnake (*P. m. lodingi*), and the Northern Pinesnake (*P. m. melanoleucus*) do not form monophyletic clades. We also find some evidence that riverine systems contribute more to population structure in the range of *Pituophis melanoleucus* than previously appreciated (Nikolakis et al., 2021). In particular, the Alabama and the Apalachicola-Chattahoochee Riverine systems seem to be barriers within the group. To the northern end of the Pinesnake range, loss of the longleaf pine habitat may also be a major natural barrier. Using the multi-species coalescent and the phylogenetic species concept, we substantiate that there could be as many as eight species of Pinesnake in what has been called *P. melanoleucus*. We suggest further study of the snake's patterns of introgression using more labile markers than UCEs.

……

Funding

This study was funded from National Science Foundation (grant number DEB2045842) to AMW and BK. AMW and BK were covered on an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant number P2O GM103424-21. AMW was additionally supported on NSF DBI 2113425. BK was supported by DiGS Fellowship from the College of Science and Technology, Southeastern Louisiana University.

Acknowledgements

We thank Bruce Rannala and Frank T. Burbrink for their feedback in the early stages of this project, which helped improve the overall structure of the study.

Submitted: February 27, 2023 EST, Accepted: September 11, 2023 EST

References

Assis, L. C. S., & Rieppel, O. (2011). Are monophyly and synapomorphy the same or different? Revisiting the role of morphology in phylogenetics. *Cladistics*, *27*(1), 94–102. https://doi.org/10.1111/j.1096-0031.2010.00317.x

Avise, J. C., Reeb, C. A., & Saunders, N. C. (1987). Geographic population structure and species differences in mitochondrial DNA of mouthbrooding marine catfishes (ariidae) and demersal spawning toadfishes (batrachoididae). *Evolution*, *41*(5), 991. https://doi.org/10.2307/2409187

Barbour, T. (1921). The florida pine snake. *Proc. New England Zool. Club*, *7*, 117–118.

Baxley, D. L., Lipps, G. J., Jr., & Qualls, C. P. (2011). Multiscale Habitat Selection by Black Pine Snakes (*Pituophis melanoleucus lodingi*) in Southern Mississippi. *Herpetologica*, *67*(2), 154–166. https://doi.org/10.1655/herpetologica-d-10-00029.1

Baxley, D. L., & Qualls, C. P. (2009). Black pine snake (*Pituophis melanoleucus lodingi*): Spatial ecology and associations between habitat use and prey dynamics. *Journal of Herpetology*, *43*(2), 284–293. https://doi.org/10.1670/08-074r1.1

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W. J., Mattick, J. S., & Haussler, D. (2004). Ultraconserved elements in the human genome. *Science*, *304*(5675), 1321–1325. https://doi.org/10.1126/science.1098119

Blanchard, F. N. (1920). A black *Pituophis*. *Copeia*, *81*, 30. https://doi.org/10.2307/1435931

Bonett, R. M., Boundy, J., Burbrink, F. T., Crother, B. I., Queiroz, K. de, Frost, D. R., Highton, R., Iverson, J. B., Jockusch, E. L., Kraus, F., & others. (2017). *Scientific and standard english names of amphibians and reptiles of north america north of mexico, with comments regarding confidence in our understanding*. Society for the Study of Amphibians and Reptiles Lawrence, Kansas.

Bryant, D., & Hahn, M. W. (2020). The Concatenation Question. In C. Scornavacca, F. Delsuc, & N. Galtier (Eds.), *Phylogenetics in the Genomic Era* (p. 3.4:1-3.4:23). No commercial publisher | Authors open access book.

Burbrink, F. T., & Guiher, T. J. (2014). Considering gene flow when using coalescent methods to delimit lineages of North American pitvipers of the genus *Agkistrodon*. *Zoological Journal of the Linnean Society*, *173*(2), 505–526. https://doi.org/10.1111/zoj.12211

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/j.0014-3820.2000.tb01253.x

Burger, J., & Zappalorti, R. T. (1988). Habitat use in free-ranging pine snakes, *Pituophis melanoleucus*, in new jersey pine barrens. *Herpetologica*, *44*(1), 48–55. https://doi.org/10.1007/s12039-015-0801-6

Church, S. A., Kraus, J. M., Mitchell, J. C., Church, D. R., & Taylor, D. R. (2003). Evidence for multiple pleistocene refugia in the postglacial expansion of the eastern tiger salamander, *Ambystoma tigrinum tigrinum*. *Evolution*, *57*(2), 372–383. https://doi.org/10.1554/0014-3820(2003)057[0372:EFMPRI]2.0.CO;2

Cogger, H. G., Zweifel, R. G., Kirshner, D., & others. (1998). *Encyclopedia of reptiles & amphibians*. Weldon Owen.

De Queiroz, K. (2007). Species concepts and species delimitation. *Systematic Biology*, *56*(6), 879–886. https://doi.org/10.1080/10635150701701083

Donoghue, M. J. (1985). A critique of the biological species concept and recommendations for a phylogenetic alternative. *The Bryologist*, *88*(3), 172. https://doi.org/10.2307/3243026

Dye, D. D. (2006). Black pine snake. In *Florida Backyard Snakes*. https://floridabackyardsnakes.com/black-pine-snake-2.html

Faircloth, B. C. (2015). PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics*, *32*(5), 786–788. https://doi.org/10.1093/bioinformatics/btv646

Flouri, T., Rannala, B., & Yang, Z. (2020). A Tutorial on the Use of BPP for Species Tree Estimation and Species Delimitation. In C. Scornavacca, F. Delsuc, & N. Galtier (Eds.), *Phylogenetics in the Genomic Era*. No commercial publisher | Authors open access book.

Gilbert, P. S., Chang, J., Pan, C., Sobel, E. M., Sinsheimer, J. S., Faircloth, B. C., & Alfaro, M. E. (2015). Genome-wide ultraconserved elements exhibit higher phylogenetic informativeness than traditional gene markers in percomorph fishes. *Molecular Phylogenetics and Evolution*, *92*, 140–146. https://doi.org/10.1016/j.ympev.2015.05.027

Graham, S. P., Steen, D. A., Nelson, K. T., Durso, A. M., & Maerz, J. C. (2010). An overlooked hotspot? Rapid biodiversity assessment reveals a region of exceptional herpetofaunal richness in the southeastern united states. *Southeastern Naturalist*, *9*(1), 19–34. https://doi.org/10.1656/058.009.0102

Guyer, C., & Bailey, M. A. (1993). Amphibians and reptiles of longleaf pine communities. *Proceedings of the Tall Timbers Fire Ecology Conference*, *18*, 139–158.

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

Harvey, M. G., Smith, B. T., Glenn, T. C., Faircloth, B. C., & Brumfield, R. T. (2016). Sequence capture versus restriction site associated DNA sequencing for shallow systematics. *Systematic Biology*, *65*(5), 910–924. https://doi.org/10.1093/sysbio/syw036

Helfrich, P., Rieb, E., Abrami, G., Lücking, A., & Mehler, A. (2018). TreeAnnotator: Versatile visual annotation of hierarchical text relations. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

Höhna, S., Landis, M. J., Heath, T. A., Boussau, B., Lartillot, N., Moore, B. R., Huelsenbeck, J. P., & Ronquist, F. (2016). RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic Biology*, *65*(4), 726–736. https://doi.org/10.1093/sysbio/syw021

Holman, J. A. (2000). *Fossil snakes of north america: Origin, evolution, distribution, paleoecology*. Indiana University Press.

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

Kapfer, J. M., Pekar, C. W., Reineke, D. M., Coggins, J. R., & Hay, R. (2010). Modeling the relationship between habitat preferences and home-range size: A case study on a large mobile colubrid snake from north america. *Journal of Zoology*, *282*(1), 13–20. https://doi.org/10.1111/j.1469-7998.2010.00706.x

Lewis, P. O. (2001). A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology*, *50*(6), 913–925. https://doi.org/10.1080/106351501753462876

Malik, M. A. R., Butt, A. N., & Choi, J. N. (2014). Rewards and employee creative performance: Moderating effects of creative self-efficacy, reward importance, and locus of control. *Journal of Organizational Behavior*, *36*(1), 59–74. https://doi.org/10.1002/job.1943

Mason, N. A., Olvera-Vital, A., Lovette, I. J., & Navarro-Sigüenza, A. G. (2018). Hidden endemism, deep polyphyly, and repeated dispersal across the Isthmus of Tehuantepec: Diversification of the White-collared Seedeater complex (Thraupidae: *Sporophila torqueola*). *Ecology and Evolution*, *8*(3), 1867–1881. https://doi.org/10.1002/ece3.3799

Mayr, E. (1981). Biological classification: Toward a synthesis of opposing methodologies. *Science*, *214*(4520), 510–516. https://doi.org/10.1126/science.214.4520.510

McCormack, J. E., Harvey, M. G., Faircloth, B. C., Crawford, N. G., Glenn, T. C., & Brumfield, R. T. (2013). A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. *PloS One*, *8*(1), e54848. https://doi.org/10.1371/journal.pone.0054848

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

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Messenger, K. (2015). *Northern x florida pine intergrade*. https://www.flickr.com/photos/kevinmessenger/21613549684/in/photostream/

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). *Molecular Ecology*, *29*(4), 797–811. https://doi.org/10.1111/mec.15358

Nikolakis, Z. L., Orton, R. W., & Crother, B. I. (2021). Fine-scale population structure within an Eastern Nearctic snake complex (*Pituophis melanoleucus*). *Zoologica Scripta*, *51*(2), 133–146. https://doi.org/10.1111/zsc.12522

Noss, R. F. (1989). Longleaf pine and wiregrass: Keystone components of an endangered ecosystem. *Natural Areas Journal*, *9*(4), 211–213.

Pyron, R. A., & Burbrink, F. T. (2009). Neogene diversification and taxonomic stability in the snake tribe lampropeltini (serpentes: colubridae). *Molecular Phylogenetics and Evolution*, *52*(2), 524–529. https://doi.org/10.1016/j.ympev.2009.02.008

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904. https://doi.org/10.1093/sysbio/syy032

Reichling, S. B. (1995). The taxonomic status of the louisiana pine snake (*Pituophis melanoleucus ruthveni*) and its relevance to the evolutionary species concept. *Journal of Herpetology*, *29*(2), 186. https://doi.org/10.2307/1564556

Roch, S., & Steel, M. (2015). Likelihood-based tree reconstruction on a concatenation of aligned sequence data sets can be statistically inconsistent. *Theoretical Population Biology*, *100*, 56–62. https://doi.org/10.1016/j.tpb.2014.12.005

Rodríguez-Robles, J. A., & De Jesús-Escobar, J. M. (2000). Molecular systematics of new world gopher, bull, and pinesnakes (*Pituophis*: Colubridae), a transcontinental species complex. *Molecular Phylogenetics and Evolution*, *14*(1), 35–50. https://doi.org/10.1006/mpev.1999.0698

Smith, M. L., & Carstens, B. C. (2019). Process-based species delimitation leads to identification of more biologically relevant species*. *Evolution*, *74*(2), 216–229. https://doi.org/10.1111/evo.13878

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

Stull, O. G. (1940). Variations and relationships in the snakes of the genus *Pituophis*. *Bulletin of the United States National Museum*, *175*, 1–225. https://doi.org/10.5479/si.03629236.175.1

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

Thomas, T. M., Granatosky, M. C., Bourque, J. R., Krysko, K. L., Moler, P. E., Gamble, T., Suarez, E., Leone, E., & Roman, J. (2014). Taxonomic assessment of alligator snapping turtles (chelydridae: Macrochelys), with the description of two new species from the southeastern united states. *Zootaxa*, *3786*(2), 141. https://doi.org/10.11646/zootaxa.3786.2.4

Tuberville, T. D., Willson, J. D., Dorcas, M. E., & Gibbons, J. W. (2005). Herpetofaunal species richness of southeastern national parks. *Southeastern Naturalist*, *4*(3), 537–569. https://doi.org/10.1656/1528-7092(2005)004[0537:HSROSN]2.0.CO;2

Weinell, J. L., & Austin, C. C. (2017). Refugia and Speciation in North American Scarlet Snakes (*Cemophora*). *Journal of Herpetology*, *51*(1), 161–171. https://doi.org/10.1670/15-125

Winker, K., Glenn, T. C., & Faircloth, B. C. (2018). Ultraconserved elements (UCEs) illuminate the population genomics of a recent, high-latitude avian speciation event. *PeerJ*, *6*, e5735. https://doi.org/10.7717/peerj.5735

Yang, Z. (1994). Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. *Journal of Molecular Evolution*, *39*(3), 306–314. https://doi.org/10.1007/bf00160154

Yang, Z. (2015). The BPP program for species tree estimation and species delimitation. *Current Zoology*, *61*(5), 854–865. https://doi.org/10.1093/czoolo/61.5.854

Yang, Z., & Rannala, B. (2010). Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences*, *107*(20), 9264–9269. https://doi.org/10.1073/pnas.0913022107

Yang, Z., & Rannala, B. (2014). Unguided species delimitation using DNA sequence data from multiple loci. *Molecular Biology and Evolution*, *31*(12), 3125–3135. https://doi.org/10.1093/molbev/msu279

Zappalorti, R. T., & Burger, J. (1985). On the importance of disturbed sites to habitat selection by pine snakes in the pine barrens of new jersey. *Environmental Conservation*, *12*(4), 358–361. https://doi.org/10.1017/s0376892900034500

Supporting Information

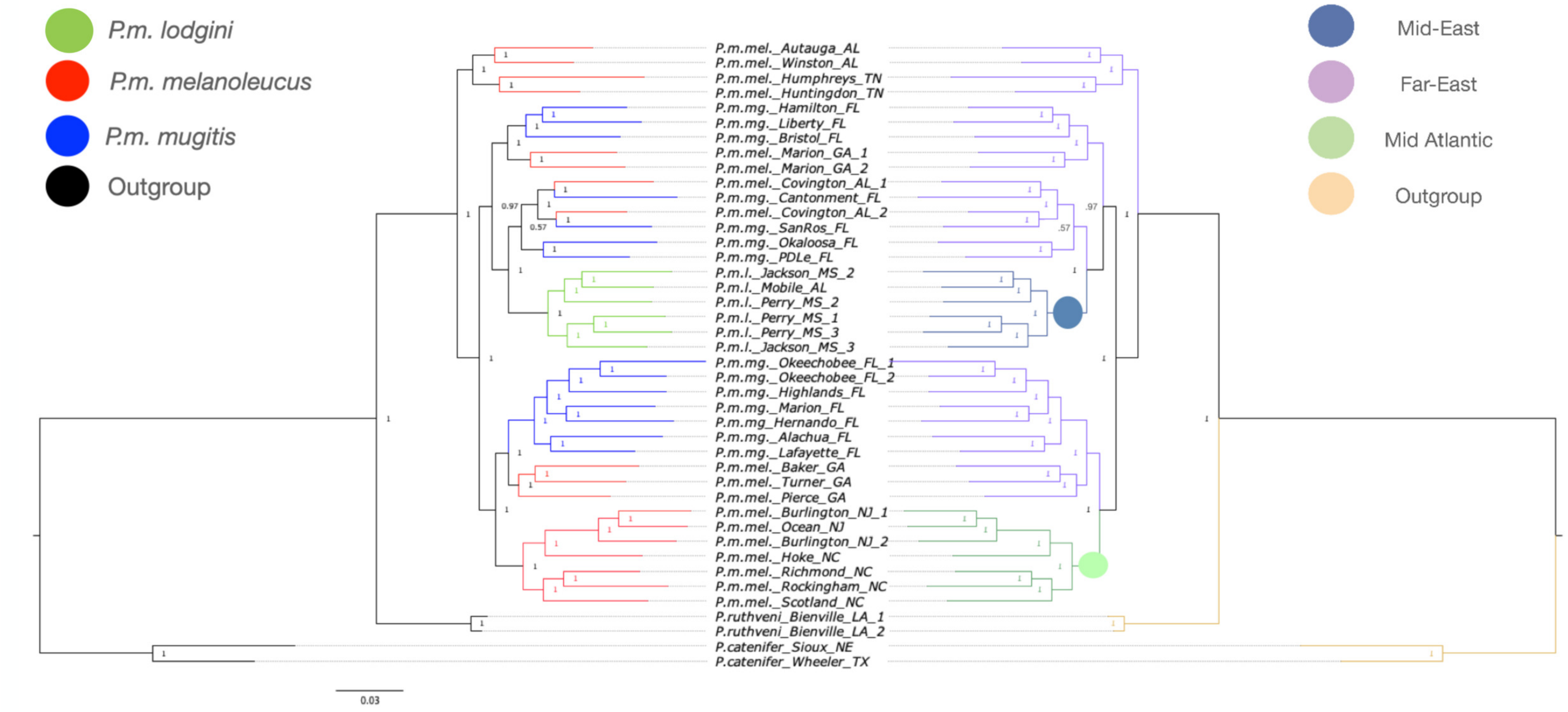
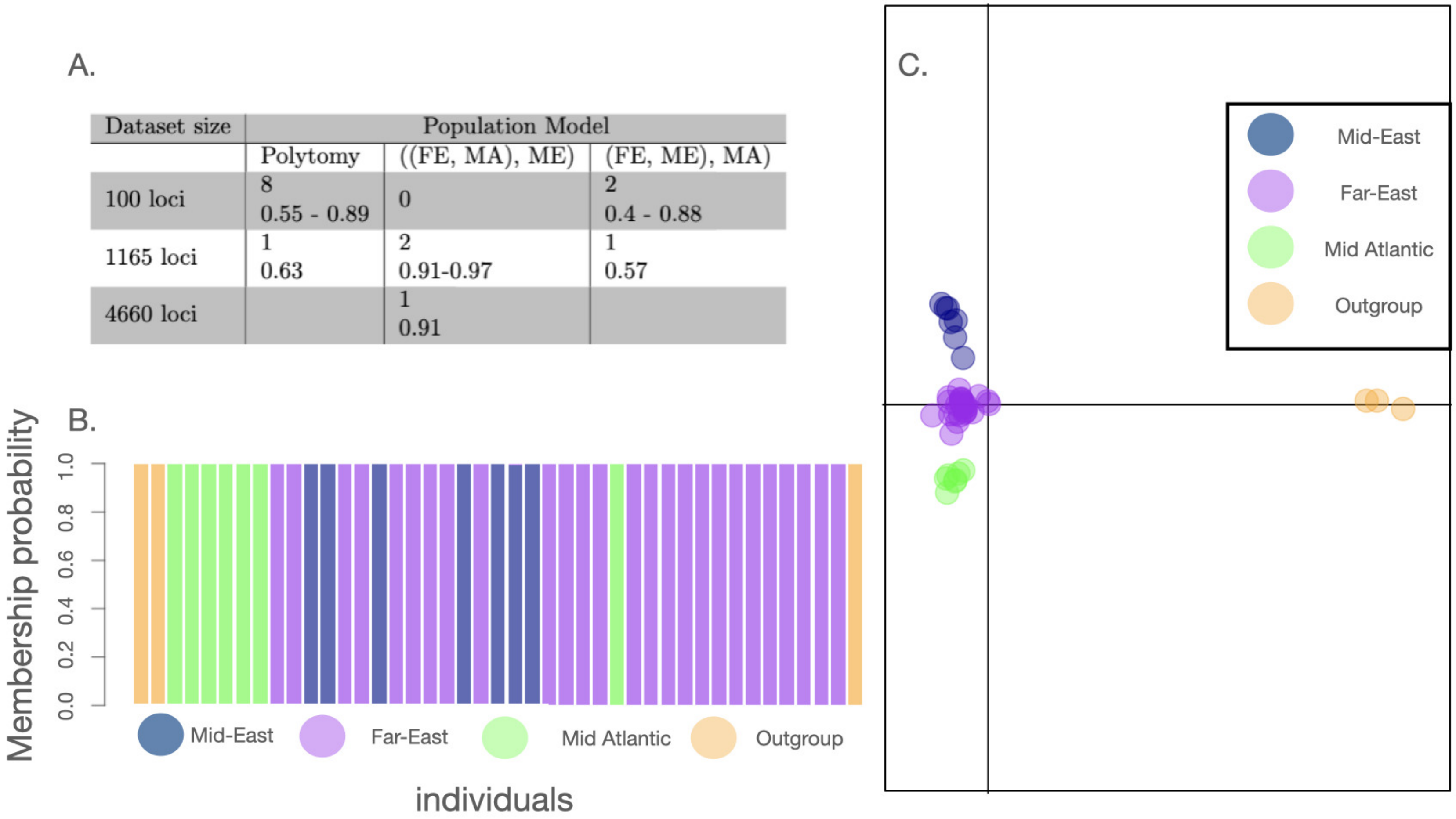


Figure S1. Phylogenetic Trees colored according to the subspecific designations (in the right) and nodes and branches colored according to the DAPC clusters obtained with the minimized BIC scores ($k = 4$).



Dataset size	Population Model		
	Polytomy	((FE, MA), ME)	(FE, ME), MA)
100 loci	8 0.55 - 0.89	0	2 0.4 - 0.88
1165 loci	1 0.63	2 0.91-0.97	1 0.57
4660 loci		1 0.91	

Figure S2. A. The posterior probabilities for different population models obtained from different downsampling schemes. B. Compoplot showing membership probabilities for the samples colored according to the DAPC samples. C. Scatter-plot depicting the clusters obtained from DAPC.