**ABSTRACT**

The Eastern Pine Snake (*Pituophis melanoleucus*) is found throughout eastern North America. Taxonomy in this group has been controversial with several conflicting subspecific designations. Three subspecies of the Eastern Pine Snakes have been proposed according to their geographical locations: the northern Pine snake (*P. m. melanoleucus*), the Florida Pine snake (*P. m. mugitus*), and the Black Pine snake (*P. m. lodingi*). There are consistently unresolved relationships among these subspecific taxa in previous studies. We analyzed ultra-conserved elements (UCEs) to perform species tree estimation and species delimitation approaches implementing Bayesian inference methods. Species delimitation indicated that the plurality of datasets supported an ingroup of one species rather than three different subspecies. These results confirm prior findings of little divergence between the three putative subspecies and suggesting one single species.

**INTRODUCTION**

Molecular phylogenetic data have a long history of application to species delimitation problems (Donoghue 1985). In the earliest forms, this took the form of the phylogenetic species concept (de Queiroz 2007), which posited that species were independent lineages on a phylogenetic tree. More recently, models that provide a more in-depth look at gene flow among taxa have been used for species delimitation from molecular data (Yang and Rannala 2010). In many ways, this is a return to the biological species concept, placing emphasis on genetic introgression and lineage separation. The multispecies coalescent (MSC) method uses both molecular phylogenetics and population genetics to counter problems, such as long-branch attraction and the inherent subjectivity of interpreting the phylogeny, posed by delimiting species on the basis of phylogenetic trees (Yang and Rannala 2010, Yang 2015). 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). Different evolutionary processes operate at different geographical locations which would lead to population genetic reconstructing over time (Soltis et al. 2006).

The southeastern United States is an area with rich biodiversity consisting of almost half of the country’s reptiles and amphibians (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 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, in particular, provides critical habitat for a number of endemic species (Guyer and Bailey 1993). One of such species is the eastern pine snakes, *Pituophis melanoleucus.* It has been hypothesized to have as many as three subspecies (Stull 1940; Crother 2012). In this manuscript, we use molecular species delimitation methods to examine the taxonomy in this group.

Morphological characters, and particularly synapomorphies, have typically been considered an important component of determining valid species (Assis and Rieppel 2011). However, pine snake subspecies are not observed to have synapomorphies and those that are potential synapomorphies are fairly labile, such as scale coloration. Individuals of different subspecies who appear to show signs of introgression have been observed in the wild (Scott 2008). In the 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. There are multiple types of molecular data that can be brought to bear this question. Each type has different evolutionary properties that lead to the marker capturing different types of variation. Some of such molecular data are DNA barcode (Herbert et. a. 2003), UCEs (Bejerano et al. 2004; Winker et. al. 2018), ddRADseq (Peterson et al. 2012, Reitzel et. al. 2013), Sequence capture (Anderman et. al. 2020). In this study, we make use of a UCE dataset collected for phylogenetics in the pine snakes group (Nikolakis et al. 2021).

*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 and Guiher 2015; McKelvy and Burbrink 2017; Myers et al. 2020). Examples of barriers are the Apalachicola and Mississippi river drainages which are believed to have created population differentiation among many groups of organisms (Pyron and Burbrink 2009; Soltis et. al. 2006). Prior studies have also supported population structure differences across these barriers. For example, some populations of tiger salamanders, rat snakes, musk turtles and snapping turtles exhibit different population structure in the eastern and western side of the Apalachicola barrier (Church et. al. 2003; Burbrink et. al. 2000; Burbrink 2002; Thomas et. al. 2014), but some populations of catfish show no genetic differences across the region (Avise et. al. 1987). The eastern pine snakes have a wide range of habitats across the eastern United States and are thought to contain several distinct populations with high degrees of gene flow (Nikolakis et al. 2021). This species complex is currently classified with three geographic sub-specific taxa, the Northern Pine snake (*P. m. melanoleucus*), (2) the Florida Pine snake (*P. m. mugitus*), and (3) the Black Pine snake (*P. m. lodingi*) (Crother 2012). These snakes range widely in color from uniformly black to having red/bronze patches (Guyer et al. 2019). The Black Pine snake has brown to black dorsal color, the Florida Pine snake has gray anterior color and rusty brown in the posterior with some brownish spots in the dorsal side, and the Northern Pine snake are creamy, yellowish ground in color with dark blotches in the entire body (Guyer et. al. 2019). This species complex is an ideal candidate for testing whether or not the current subspecific taxonomy represents evolutionary history through the use of species delimitation methods given the uncertainty of relationships from previous studies (Nikolakis et al. 2021, Rodrı́guez-Robles and De Jesus-Escobar 2000).

There are various tree-based and non-tree-based species delimitation methods that can be used to determine the species boundaries (Camargo and Sites 2013). *bpp* is a software that generates the Bayesian posterior distribution of species delimitation models using the multispecies coalescent framework (Yang and Rannala 2010). It uses reversible-jump MCMC to move between models of the number of populations present in the sample, while calculating the posterior probabilities associated with the model of population differentiation. *bpp* allows us to calculate the model likelihoods for different numbers of species. We used *bpp* to test how many species are present in our ingroup.

We used UCEs to obtain the evolutionary relationships within *Pituophis melanoleucus*, to determine if UCEs are best suited for this type of analyses, and to determine if there is support for multiple subspecies in *Pituophis melanoleucus*. We used *bpp* (Flouri et. al. 2018) for delimiting species under the multispecies coalescent. We also used *RevBayes* (Hӧhna et. al. 2016) to determine if there is any existing phylogenetic structure within the group, as the phylogenetic species concept defines species as a distinct group of conspecific individuals (Woodruff 2001). This study builds on the recent work of Nikolakis et al. (2021) and will help to provide a better understanding of diversity patterns of pine snakes in the eastern United States and may provide additional insight on the utility of UCEs to study phylogenetic relationships in recently diverged clades. In this manuscript, we use both the phylogenetic tree and the MSC to delimit species in the genus *Pituophis*.

**MATERIALS AND METHODS**

**Sample collection, DNA extraction and Bioinformatics**

Tissue Samples were collected from forty-three specimens of *Pituophis melanoleucus* from their geographical distribution (Fig. 1, Table 1). The samples were de-multiplexed, filtered, and processed by removingadapter sequences and ambiguous bases using the program *Illumiprocessor* which is incorporatedin the software *Phyluce* v.1.5 (Faircloth 2015) according to protocol outlines in Nikolakis et. al. (2021).

**Phylogenetic Analyses**

We conducted phylogenetic analyses using a nucleotide substitution model from *RevBayes* software, v.1.1.1 (Hӧhna et. al. 2016). We used the general time reversible (GTR) model (Tavaré 1986) of sequence evolution, which allows six exchangeability rates between nucleotide states. The exchangeabilities are drawn from a Dirichlet distribution with an uninformative prior and allows the data to determine the value of the exchange abilities. 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 50,000 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 then summarized into maximum clade credibility trees (Helfrich et. al. 2018) and majority-rule consensus trees using *RevBayes*.

To delimit species, we used Bayesian Phylogenetics and Evolution, *bpp* (Yang & Rannala 2010), a genealogical method that estimates the time of origin, time of diversification, and the effective population multiplied by the mutation rate, for each species. We used the model A11 which estimates the species delimitation and the species tree (Fluori et. al. 2020). In this analysis the species delimitation model and the species phylogeny both change in the MCMC. The results showing posterior probability distributions indicate whether the lineages can be differentiated from each other. In the analysis, the subspecies of pine snakes were labelled according to their geographical distribution (FE – Far East; ME – Mid East; TN – Tennessee Valley and OG – Outgroup). The burnin was specified at 8000 and each dataset ran for 100000 generations. The output file contained posterior probabilities of the best fit models and the arrangement of species labelled per their geographic location. The line containing all the probabilities for the best fit was extracted from the output files using a UNIX script and a histogram was created to visualize the number of species of pine snakes.

To create a concatenated tree for comparing consensus trees across UCEs, we then used the summarized trees from *RevBayes* (Höhna et. al. 2016) and built a consensus network in *R* (R core team) using the packages *ape* (Paradis & Schliep 2019), *phangorn* (Schliep 2011) and *phytools* (Revell 2012). The code and data for all the analyses are stored in GitHub (<https://github.com/basanta33/Pituophis_>).

**RESULTS**

**Sequence Data**

The majority of thevariation recovered was observed in the extreme regions of each UCE and there was littlecorrelation between the variable sites and locus lengths (see Nikolakis et. al. 2021 for further details).

**Species Delimitation with BPP**

We carried out a study to examine whether there is any support for the recognition of subspecies within *Pituophis melanoleucus*. The model A11 accommodated for the gene tree uncertainty and variable population sizes over time to explore different species delimitation models and different species phylogenies. We processed the output from the datafiles with variable sites to obtain the posterior probabilities of different species groupings of pine snakes. When we combined the probabilities obtained from all the datasets to a single file and made a histogram, the plurality of the datasets indicated that there is an ingroup of one species of *Pituophis melanoleucus* and the outgroup of *P. ruthveni*, *P. catenifer*, and *Pantherophis obsoletus* (Figure 2). The posterior probabilities of delimitation were obtained based on a guide tree which accounts for the phylogenetic uncertainty.

**Consensus Network**

The output trees obtained from *RevBayes* were summarized to obtain maximum credibility trees and consensus trees. The consensus trees obtained were composed into a network of phylogenetic trees which reflected little to no variation amongst the subspecies of *P. melanoleucus*. The tree obtained from the *phangorn* (Schliep 2011) package enabled us to evaluate the conflicting phylogenetic signals from the collected datasets. The tree (Figure 4) indicates that there is little variation among the topologies estimated from different pine snakes over the eastern side of the United States. If different loci had supported different phylogenetic trees, we would have expected the nodes of the trees to connect to the tips in a web-like structure.

**DISCUSSION**

**Corroboration of a single species of Eastern Pine Snakes**

Our results show that *Pituophis melanoleucus* is not composed of various distinct geographic lineages within the eastern United States. The Bayesian consensus tree (Figure 4) indicates that there is little to no variation among the subspecies of *P. melanoleucus*. Figure 3 also indicates that there is little to no phylogenetic structure among the samples collected from different parts of the eastern United States. This result agrees with Nikolakis’s study (2021) which used sequence-capture based approach to understand the factors that correlate with genetic diversity and distance in the lineages of *P. melanoleucus*. The eastern pine snake diverged within the *Pituophis* about 6 to 3 million years ago (Pyron and Burbrink 2009). The complex of *P. melanoleucus* appears to be of a single species with very little genetic differentiation, as indicated by the posterior probabilities of the best fit model for the species tree generated by *bpp*. In addition to the results from *bpp*, the majority rule consensus tree obtained from *RevBayes* also indicated that we get the same trees across the sites, a near-polytomy with little geographic structure (Figure 3). The consensus network compiled using all the 4600 data files show that there is no connection between multiple tips and each node thus indicating little topology variation among the subspecies of pine snakes in the eastern region (Figure 4).

**Gene flow across a geographic barrier**

The three different geographical lineages: far-eastern, mid-eastern and Tennessee clades are separated by the Apalachicola/Chattahoochee River Basin acting as a geographical barrier (Nikolakis 2021). Previous squamate phylogeographic studies across that region (Burbrink et al., 2000, 2008; Soltis et. al. 2006; Weinell and Austin, 2017; Myers et. al. 2020) indicated that there is significant genetic variation among the clades existing across the barrier. Despite the geographical barrier, molecular species delimitation indicates that the three lineages of pine snakes are not distinct from each other. *bpp* should detect if there is gene flow across the barrier. In our analysis, bpp indicated that there was only one population 4650 times and two populations (one being the outgroup) 4654 times during the two runs of the datasets (Fig. 2). The geographic barrier did not seem to have much effect on the isolation of the population of pine snakes across the mid-eastern and far-eastern sides. Due to the indication of little genetic differentiation, it can be concluded that gene flow has been maintained in the population of eastern pine snakes across that region. This disunity in the previous phylogeographic analyses (Burbrink et al. 2008, 2000; Weinell and Austin 2017) of the snakes and our study could be because of the distribution of the species of the pine snakes across the eastern United States. Previous movement studies indicate that *P. melanoleucus* is a very mobile species with their home ranges spanning from approximately 35 hectares to over 105 hectares (Nikolakis et. al. 2021) which could contribute to the gene flow among the different populations.

**Use of UCE for recently diverged lineages**

Ultra-conserved Elements (UCEs) are highly conserved regions within the genome that are shared among evolutionarily distant taxa (Bejerano et al. 2004). The use of UCEs has been increasing in phylogeny inference across many vertebrate taxa (Gustafson et. al. 2019). Although UCE is an important molecular marker, for this complex the use of molecular markers other than UCE would be a better option. Some UCEs in this study did not even distinguish between the outgroup and the ingroup taxa, so we had very little information in those sites. Of the sites with enough information to discriminate in- from the outgroup, 83% supported one in-group population (Figure 3). This is because of hypothesized recent evolution of the pine snakes from other *Pituophis* (Pyron and Burbrink 2009). The oldest fossils of *P. melanoleucus* have been found in Florida dating 0.8 to 2.5 mya and more northern fossils from Pennsylvania have been dated from 0.1 mya (Holman 2000). These data could indicate that the lineage diverged during the late Pleistocene when there were environmental fluctuations leading to periods of isolation and connection. These periods led to the maintenance of gene flow through the contacts of the different populations. As UCEs are conserved sequences, they evolve very slowly, thus decreasing the power to detect variation among the organisms that have recently diverged from its common ancestor (Winker et. al. 2018). Thus, further study with larger sample sizes and broader genomic coverage will be needed to determine the species limits between these taxa (Sukumaran and Knowles 2017).

**CONCLUSION**

Using a dataset of ultra-conserved elements, we find no evidence for multiple subspecies in *Pituophis melanoleucus.* Using the multispecies coalescent and the phylogenetic species concept, we substantiate *Pituophis melanoleucus* as a single species across its whole range, a hypothesis suggested by previous authors. However, given low informativity of many UCEs, we suggest that other data types would be preferable for a study on this time scale.

**LITERATURE CITED**

Assis, L.C.S., and O. Rieppel. 2011. Are monophyly and synapomorphy the same or different? Revisiting the role of morphology in phylogenetics. Cladistics 27(1):94-102.

Camargo, A., and J. Sites. 2013. Species Delimitation: A Decade After the Renaissance, p. 225- 247. *In:* The Species Problem - Ongoing Issues. Igor Pavlinov (eds.). InTech, 2013.

Avise, J.C., A.C. Reeb, and N.C. Saunders. 1987. Geographic population structure and species differences in mitochondrial DNA of mouthbrooding marine catfishes (Ariidae) and demersal spawning toadfishes (Batrachoididae). Evolution 41:991–1002.

Bejerano G, M. Pheasant, I. Makunin, S. Stephen, W.J. Kent, J.S. Mattick, and D. Haussler. Ultraconserved elements in the human genome. 2004 May. Science 28;304(5675):1321-5.

Burbrink, F.T. 2002. Phylogeographic analysis of the cornsnake (*Elaphe guttata*) complex as inferred from maximum likelihood and Bayesian analyses. Molecular Phylogenetics and Evolution 25:465–476.

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

Church, S.A.,J.M. Kraus, J.C. Mitchell, D.R. Church, and D.R. Taylor. 2003. Evidence for multiple Pleistocene refugia in the postglacial expansion of the eastern tiger salamander, *Ambystoma tigrinum tigrinum*. Evolution 57:372–383.

Crother, B. I. 2012. 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, 2012.

De Queiroz, K. 2007. Species Concepts and Species Delimitation. Systematic Biology 56:879–886.

Donoghue, M.J. 1985. A Critique of the Biological Species Concept and Recommendations for a Phylogenetic Alternative. The Bryologist 88(3):172-181.

Flouri, T., J. Xiyun, B. Rannala, and Z. Yang. 2019. A Bayesian Implementation of the Multispecies Coalescent Model with Introgression for Phylogenomic Analysis. Molecular Biology and Evolution 37:1211–1223.

Flouri, T., B. Rannala, and Z. Yang. 2020. A Tutorial on the Use of BPP for Species Tree Estimation and Species Delimitation p.5.6:1-5.6:16. *In:* Phylogenetics in the Genomic Era. HAL-archives.

Flouri, T., J. Xiyun, B. Rannala, and Z. Yang. 2018. Species Tree Inference with BPP Using Genomic Sequences and the Multispecies Coalescent. Molecular Biology and Evolution 35:2585–2593.

Graham, S.P., D.A. Steen, K.T. Nelson, A.M. Durso, and J.C. Maerz. 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.

Gustafson, G.T., A. Alexander, J.S. Sproul, J.M. Pflug, D.R. Maddison, and A.E.Z. Short. 2019. Ultraconserved Element (UCE) Probe Set Design: Base Genome and Initial Design Parameters Critical for Optimization. Ecology and Evolution 9:6933–6948.

Guyer, C., and M.A. Bailey. 1993. Amphibians and reptiles of longleaf pine communities. p. 139–159. *In:* Proceedings of the Tall Timbers Fire Ecology Conference, No. 18. The Longleaf Pine Ecosystem: Ecology, Restoration and Management. Tall Timbers Research Station, Tallahassee, FL.

Guyer, C., M.A. Alexander, and R.H. Mount. 2019. Lizards and Snakes of Alabama. University of Alabama Press.

Höhna, S., M.J. Landis, T.A. Heath, B. Bissau, N. Lartillot, B.R. Moore, J.P. Huelsenbeck, and F. Ronquist. 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. Systematic Biology 65:726-736.

Myers, E.A., A.D. McKelvy, and F.T. Burbrink. 2020. Biogoegraphic barriers, Pleistocene refugia, and climatic gradients in the southeastern Nearctic drive diversification in cornsnakes (*Pantherophis guttatus* complex). Molecular Ecology 29(4):797-811.

Nikolakis, Zachary L., Richard W. Orton, and Brian I. Crother. 2021. Fine Scale Population Structure and Extensive Gene Flow within an Eastern Nearctic Snake Complex (*Pituophis melanoleucus*). BioRxiv 2021.09.28.462151

Paradis E., and K. Schliep. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35(3):526–528.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E.. 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.

Pyron, A.R., and Frank B.T. 2009. Neogene Diversification and Taxonomic Stability in the Snake Tribe Lampropeltini (Serpentes: Colubridae). Molecular Phylogenetics and Evolution 52:524–529.

Rodrı́guez-Robles, J.A., and J.M. De Jesús-Escobar. 2000. Molecular Systematics of New World Gopher, Bull, and Pinesnakes (Pituophis: Colubridae), a Transcontinental Species Complex. Molecular Phylogenetics and Evolution 14:35–50.

Reitzel, A.M., S. Herrera, M.J. Layden, M.Q. Martindale, and T.M. Shank. 2013. Going Where Traditional Markers Have Not Gone before: Utility of and promise for RAD Sequencing in Marine Invertebrate Phylogeography and Population Genomics. Molecular Ecology 22:2953–2970.

Schliep, K.P. 2011. Phangorn: Phylogenetic Analysis in R. Bioinformatics 27:592–593.

Scot, J. (2008). [Black Pine and Florida Pine Snake] [Photograph]. Flickr. https://www.flickr.com/photos/tamers1/3980065515

Soltis, D.E., A.B Morris, J.S. McLachlan, P.S. Maons, and P.S. Soltis. 2006. Comparative phylogeography of unglaciated eastern North America. Molecular Ecology 15: 4261-4293.

Tavaré S. 1986. Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences. Some Mathematical Questions in Biology: DNA Sequence Analysis. 17:57–86.

Thomas, T. M., Granatosky, M. C., Bourque, J. R., Krysko, K. L., Moler, P. E., Gamble, T., Suarez, E., Leone, E., and 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.

Tuberville, T.D., J.D. Willson, M.E. Dorcas, and J.W. Gibbons. 2005. Herpetofaunal Species Richness of Southeastern National Parks. Southeastern Naturalist 4(3):537-569.

Stull, O.G. Variations and Relationships in the Snakes of the Genus *Pituophis*. United States, U.S. Government Printing Office, 1940.

Sukumaran, J., and L.L. Knowles. 2017. Multispecies Coalescent Delimits Structure, Not Species. Proceedings of the National Academy of Sciences 114:1607–1612.

Winker, K., T.C. Glenn, and B.C. Faircloth. 2018. Ultraconserved Elements (UCEs) illuminate the population genomics of a recent, high-latitude avian speciation event. PeerJ 6:e5735.

Woodruff, D.S. 2001. Populations, Species, and Conservation Genetics. Encyclopedia of Biodiversity, Elsevier:811–829.

Yang, Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. J Mol Evol 39:306–314.

Yang, Z., Rannala B. 2010. Bayesian species delimitation using multilocus sequence data. Proc Natl Acad Sci USA 107(20):9264–9269.

Yang, Z. 2015. The BPP program for species tree estimation and species delimitation. Current Zoology 61(5):854-865.

Yang, Z., and B. Rannala. 2014. Unguided Species Delimitation Using DNA Sequence Data from Multiple Loci. Molecular Biology and Evolution 31:3125–3135.

**Figure and Table Captions:**

* Figure 1: Sampling sites across southeastern United States. The different colored regions represent habitat of the different subspecies of pine snakes. Red: Black Pine Snake; Yellow: North Pine Snake; Green: Florida Pine Snake.
* Figure 2: Number of times the posterior probabilities of the number of populations suppported appear in the *bpp* output.
* Figure 3. Consensus Phylogeny. We see no consistent phylogenetic structure to the tree.
* Figure 4: Consensus Network of the pine snakes samples. A lack of web-like structure indicates that there is little to no variation amongst the samples collected.
* Table 1: A table containing all the subspecies sample and their collection locations.