SPECIES DELIMITATION APPROACHES IN EASTERN PINE SNAKES (*PITUOPHIS MELANOLEUCUS)*

A Thesis

presented to

the Honors Program of

Southeastern Louisiana University

In Partial Fulfillment

of the Requirements for the

Honors Diploma

By

Basanta Khakurel

Dr. April Wright, Thesis Supervisor

Southeastern Louisiana University

Hammond, Louisiana

April 2021

The undersigned, appointed by the Director of the Honors Program,

have examined the thesis

SPECIES DELIMITATION APPROACHES IN EASTERN PINE SNAKES (*PITUOPHIS MELANOLEUCUS)*

presented by

Basanta Khakurel

a candidate for the Honors Diploma

and hereby certify that in their opinion it is worthy of

acceptance.

_____

Thesis Supervisor

Dr. April Wright

_____

Second Reader

Dr. Brian Crother

Name: Basanta Khakurel

Degree: B.S., Southeastern Louisiana University, 2021, Biology

Major: Biological Sciences

Title of Study: SPECIES DELIMITATION APPROACHES IN EASTERN PINE SNAKES (PITUOPHIS MELANOLEUCUS)

The Eastern Pine Snake (*Pituophis melanoleucus*) is found throughout eastern North America with several conflicting subspecific designations. There are three different subspecific taxonomic classifications 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 no resolved 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. Our study is helpful in determining the validity of UCEs in phylogenetic research of the recently evolved species like the eastern pine snake. It also contributes to the knowledge of phylogenetic patterns in the southeastern United States, removing ambiguity in the relationships and lack of geographical structuring in the traditional molecular data.

ACKNOWLEDGEMENTS

First, I would like to thank Dr. April Wright for guiding me in this thesis and taking me as an Undergraduate Research Assistant in the Paleantology Lab. Besides the lessons for the research, I go to learn a lot about thinking scientifically and work hard in every field I step in. I would like to thank the Honors department and Dr. Claire Procopio for always supporting and believing in me and helping me throughout my undergraduate career at Southeastern. I would also like to thank Dr. Brian Crother who provided guidance and suggestions for this thesis which proved valuable in making this thesis a successful one.

I would like to thank Mr. Zachary Lamar Nikolakis whose master's thesis this study is based on. His hard work in collecting and analyzing the data has been very helpful in this study.

Lastly, I would like to express my gratitude to my lab members and all my friends I have made during my time here at Southeastern. This journey has shaped me from a confused international student into the person that I am right now. I am indebted to them as they have made the biggest impact in my life and taught me valuable lessons.

TABLE OF CONTENTS

SECTION 1

INTRODUCTION

Deciphering population genetic structure and using this information to understand population history is a core challenge in biology. There are a number of methods to infer evolutionary relationships using morphological data (Patterson 1982) but due to development of molecular techniques, it has been a more common choice to use molecular data to study the evolution of different organisms. There are a number of molecular markers for such analyses and various different molecular markers which evolve at different rates would lead to different conclusions in phylogenetic analyses. Population structure is defined by the organization of genetic variation and is driven by the combined effects of evolutionary processes that include recombination, mutation, genetic drift, demographic history, and natural selection (Andam et al.). Population structure can be shaped by different environmental differences that regulate gene flow. A population is a distinct group of conspecific individuals without any gene flow whereas a metapopulation is different groups within which gene flow occurs (Woodruff 2001). The concept of species and lineage divergence can also be understood in a different manner in different areas of biology (deQuiroz 2007). deQuiroz (2007) defines species as separately evolving metapopulation lineages and highlights its usefulness in inferring the boundaries and the number of species also known as Species Delimitation.

Species as defined by deQuiroz (2007) is an ancestral descendant sequence of population evolving separately from others and with its own evolutionary role and tendencies. Different lineages separate from the ancestral lineage to adapt to the evolutionary processes that vary across space. There could be various evolutionary processes operating in different geographical locations. There might be species existing in different environmental circumstances with the

same or different morphology but could be genetically the same. Also, species that are morphologically similar but genetically different that are known as cryptic species also exist (Koshunova et. al. 2019). Species that occupy vast areas of land are more likely blocked with genetic barriers preventing gene flow thus leading to population genetic restructuring (Soltis et. al. 2016). Species Delimitation approaches would be helpful in determining the genetic similarity/dissimilarity among a population. Species could exist as segments of metapopulation lineages regardless of our ability to empirically determine them (Camargo and Sites 2013). There are various tree-based and non-tree-based species delimitation methods that help determine the species boundaries (Camargo and Sites 2013). BPP is one of the softwares that generates the Bayesian posterior distribution of species delimitation models using multispecies coalescent method (Yang and Rannala 2010). It helps in determining the environmental variation amongst different population patches and establish a genetic relation between them if any exist.

The North American continent provides habitat for a diverse group of organisms and due to its vast area, there are many geographical barriers that limit or enhance gene flow among different taxa (Burbrink et al. 2000; Burbrink and Guiher 2015; McKelvy and Burbrink 2017; Myers et al. 2020; Soltis et al. 2006). One of the examples of the barrier is the Apalachicola and Mississippi river drainages which is believed to have created population diversity among scaled reptiles (Pyron and Burbrink 2009). Although this is what is believed to be the case, there might be different organisms exhibiting different population structures across this range. The Eastern Pinesnake has a wide range of habitat across the north eastern United States and contains several distinct populations (Nikolakis 2018). The snakes are currently classified with three geographic sub-specific taxa the Northern Pinesnake (*P. m. melanoleucus*), (2) the Florida Pinesnake (*P. m. mugitus*), and (3) the Black Pinesnake (*P. m. lodingi*) (Crother 2012). These snake range widely

in color from uniformly black to having red/bronze patches (Guyer et al. 2019). This difference in morphology could be due to the different environments they are residing in. This is an ideal group of organisms to perform population structuring as the relationship between different population groups has been ambiguous.

We studied this organism to obtain the evolutionary relation within the group, to determine if UCEs are best suited for this type of analyses and to determine any diversity within the group if any exists. We used *BPP* for delimiting species and *RevBayes* to determine if there is any existing variation within the group. This study will help to provide a better understanding of phylogenetic patterns of snakes in the eastern United States and also provide an insight on the utility of UCEs to study phylogenetic relations for recently evolved organisms.

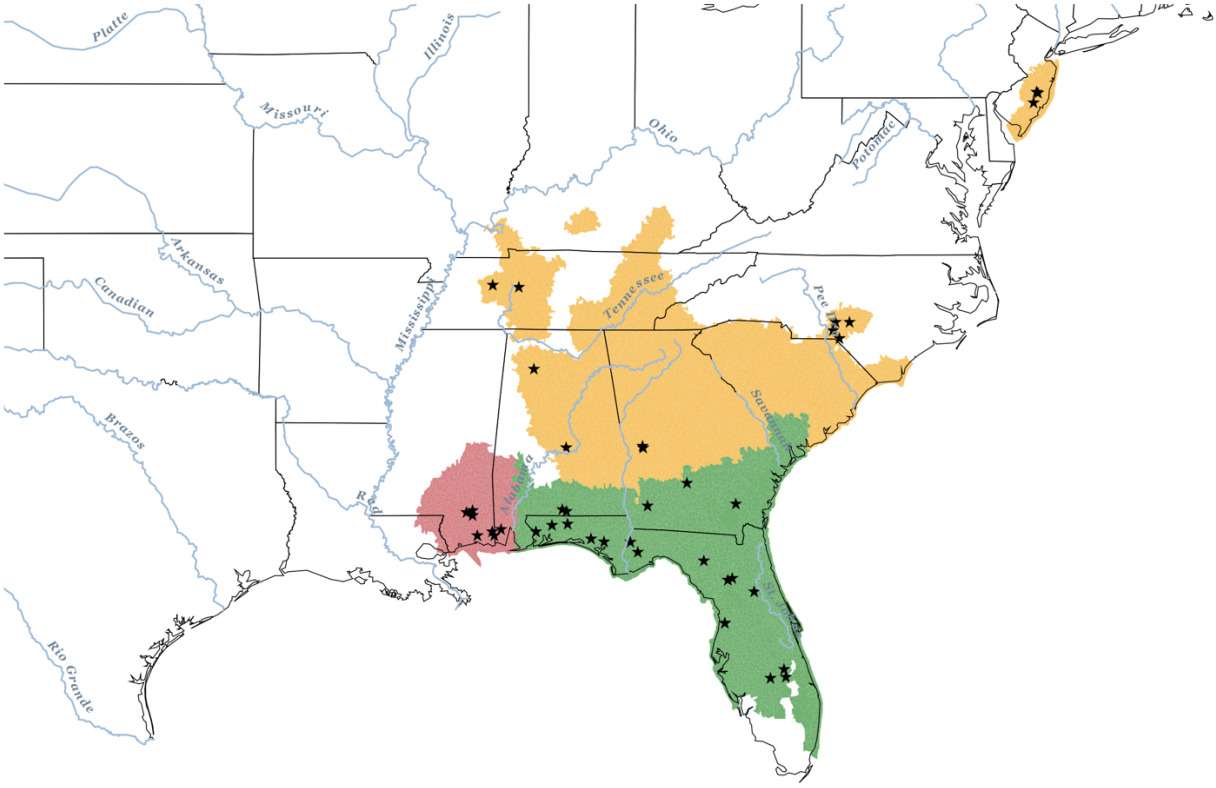

Figure 1: Geographical distribution and sampling locations of *Pituophis melanoleucus*. (Nikolakis 2018).

SECTION 2

MATERIALS AND METHODS

Sample collection and DNA extraction

Tissue Samples were collected from forty-three specimens of *Pituophis melanoleucus* from their geographical distribution (Figure 1). For the outgroup taxa, five samples each were collected from specimens of *Pituophis ruthveni, Pituophis catenifer,* and *Pantherophis obsoletus.* Samples were taken from the liver or muscle tissues and from the shed skins. Some samples were also obtained from ventral scale clips which are used for marking snakes. The collected samples were preserved in 95-100% ethanol until DNA extraction. Genomic DNA was extracted and isolated using Qiagen's DNeasy kit using the manufacturers protocols and quantified using Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA). The DNA samples were diluted to concentrations ranging from 2 to 100 ng/μL and the samples were sent to the University of Georgia's Department of Genetics for library preparation, sequence capture, and sequencing of Ultra conserved Elements (UCEs).

Bioinformatics

The samples were barcoded using Illumina TruSeq adapters with unique 8 base pair sequence tags for each individual. The UCEs were targeted by using a tetrapod 5060-locus probe set (from ultraconserved.org). The samples were de-multiplexed, filtered, and processed by removing adapter sequences and ambiguous bases using the program Illumiprocessor which is incorporated in the software Phyluce v.1.5 (Faircloth 2015). Reads were assembled anew using standard settings in Velvet v.1.1 (Zerbino & Birney 2008) and the assembled contigs were matched against the 5k UCE tetrapod probe kit to identify and extract UCEs using Phyluce. MAFFT

v.7.397 (Katoh & Standley 2013) was used to generate sequence alignments for each individual locus. To examine the effects of missing data on resulting topologies, concatenated data alignment matrices of 50% and 75% were created. The data matrices represent the number of alignments that include every individual (e.g., 75% data matrix would indicate that at least 75% of the alignments contain all individuals). The summary statistics of alignment matrices, total base pair reads, and UCE counts were generated using Python scripts from Phyluce. Three samples that had more than 70% missing data in variable regions in UCE alignments were excluded.

Phylogenetic Analyses

Phylogenetic analyses were conducted using the nucleotide substitution model from RevBayes software, v.1.1.1 (Höhna et. al. 2016). The general time reversible (GTR) model (Tavaré 1996) was used which allows six exchangeability rates to differ among the transition and transversion substitutions. The rates are drawn from a Dirichlet distribution with the prior (1,1,1,1,1,1) which allows for a relatively flat distribution along the substitution rates. We used this priori because we did not expect any substitutions to be more common than others. 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). The output files were then summarized into maximum clade credibility trees (Helfrich et. al. 2018) and 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. The results showing posterior probability distributions indicate whether the lineages can be differentiated

from each other. The species 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 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* (Schliep & Paradis 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_*).

SECTION 3

RESULTS

Sequence Data

The average reads for each individual were 1,886,033 with a contig range of 7,136 to 19,784. A range of 3,383 to 4,156 UCEs were recovered per individual with an average length of 596 bp. The number of variable sites between individual UCEs ranged from 0-74 with an average of 7 and 18% of the total UCEs containing no variable sites. The majority of the variation recovered was observed in the extreme regions of each UCE and there was little correlation between the variable sites and locus lengths.

Species Delimitation with BPP

A study was carried to examine whether the current taxonomic classification of Pine Snakes offers a just explanation in the classification of the subspecies of the snakes. The model A11 estimated the joint species delimitation and species tree estimation. The model accommodated for the gene tree uncertainty and variable population sizes to explore different species delimitation models and different species phylogenies. We processed the output from all 4661 datafiles 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).

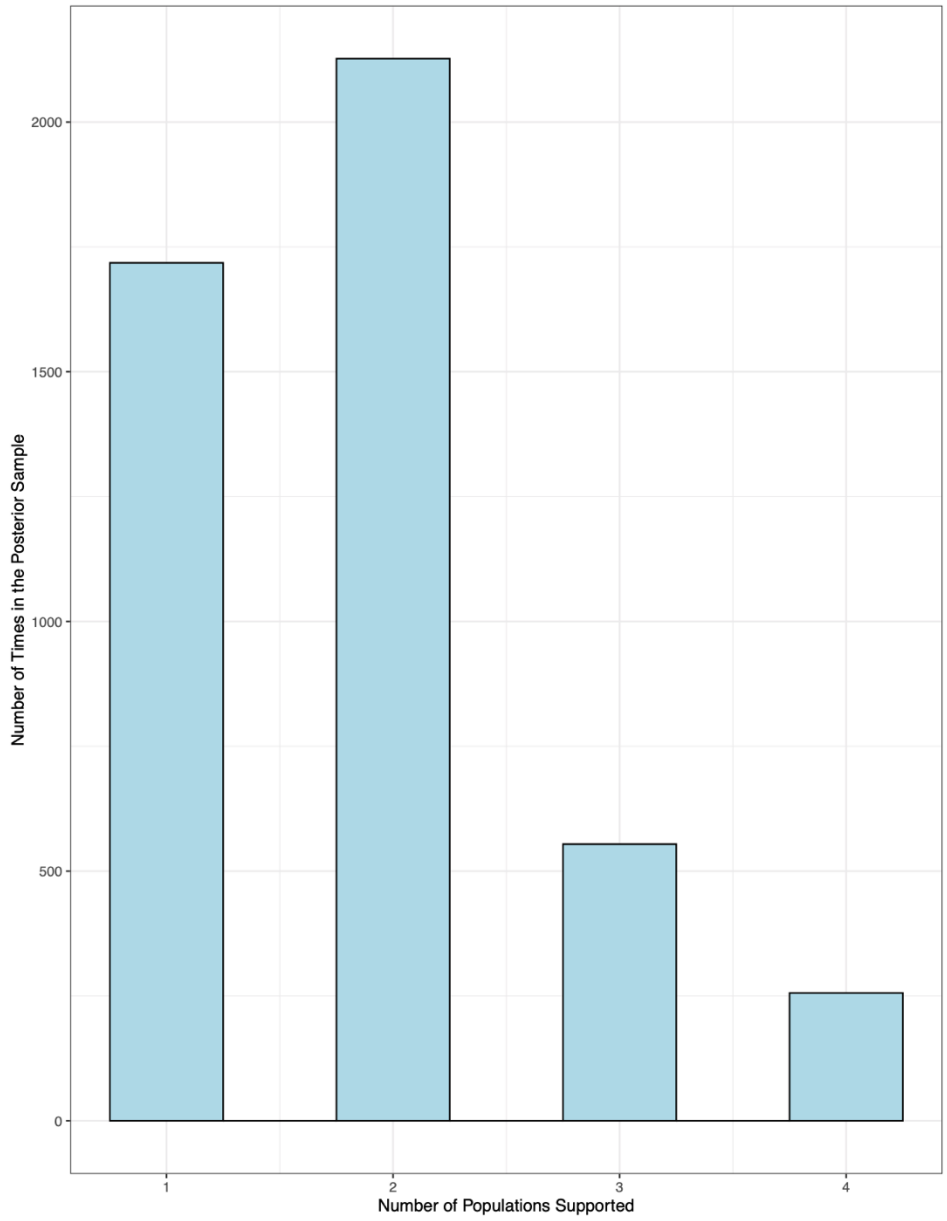

Figure 2: A histogram showing the possible number of species as indicated by the probabilities of the best fit model.

The posterior probabilities of delimitation were obtained based on a guide tree which accounts for the phylogenetic uncertainty. Our results would indicate that there is just a single species of *P. melanoleucus* which might have different morphological characteristics which led to the current classification.

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* package enabled us to evaluate the conflicting phylogenetic signals from the collected datasets. The tree (Figure 3) indicates that there is no variation among the topologies estimated from different pine snakes over the eastern side of the United States. A presence of variation would have the nodes of the trees connected to form a web-like structure.

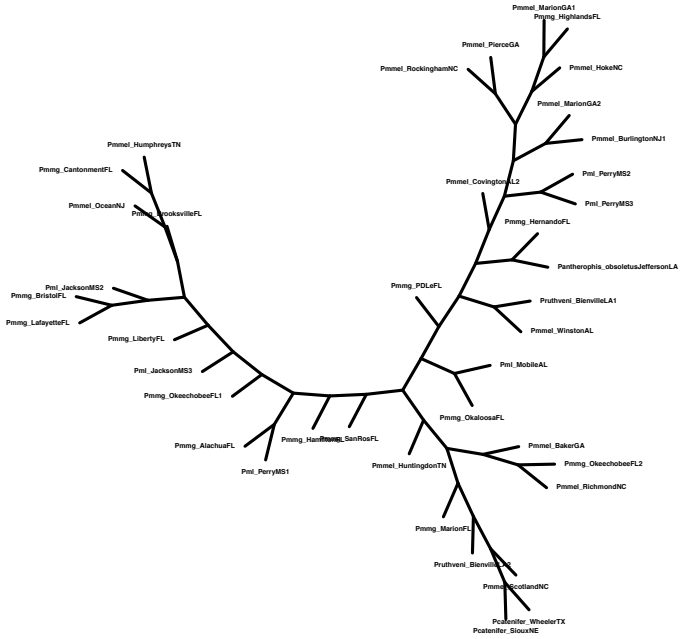

Figure 3: The consensus network of phylogenetic trees indicating no variation amongst the different subspecific classification of *P. melanoleucus*

SECTION 4

DISCUSSION

Confirmation 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 2) indicates that there is little to no variation among the subspecies of *P. melanoleucus*. This result is in agreement with Nikolakis's study (2018) which uses sequence-capture based approach to explore the classification of *P. melanoleucus.* The Eastern Pine Snakes diverged about 6 to 3 million years ago according to the study in 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 there is almost no variation amongst the subspecies of *P. melanoleucus* (Figure). The consensus network compiled using all the 4600 data files show that there is no connection between multiple tips and each node thus indicating no variation among the subspecies of pine snakes in the eastern region.

Gene flow across a geographic barrier

The three different geographical lineages far-eastern, mid-eastern and Tennessee clades (Figure) would be separated by the Apalachicola/Chattahoochee River Basin acting as a geographical barrier (Nikolakis 2018). Previous phylogeographic studies across that region (Burbrink et al., 2008, 2000; Weinell and Austin, 2017) 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. 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 phylogeographic analyses of the snakes could be due to 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 2018).

Use of UCE for recently evolved 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 reconstruction 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. This is due to recent evolution of the pine snakes from other *Pituophis* complexes. 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). This 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 been recently diverged from its common ancestor (Winker et. al. 2018).

Table 1: Taxon and sample locality for each individual Pinesnake

Taxon	Sampling Region
Pantherophis obsoletus	Jefferson LA
Pituophis catenifer	Sioux NE
Pituophis catenifer	Wheeler TX
Pituophis ruthveni	Bienville LA_1
Pituophis ruthveni	Jackson MS_1
Pituophis m. lodingi	Bienville LA_2
Pituophis m. lodingi	Jackson MS_2
Pituophis m. lodingi	Jackson MS_3
Pituophis m. lodingi	Mobile AL
Pituophis m. lodingi	Perry MS_1
Pituophis m. lodingi	Perry MS_2
Pituophis m. lodingi	Perry MS_3
Pituophis m. melanoleucus	Autauga AL
Pituophis m. melanoleucus	Burlington NJ_1
Pituophis m. melanoleucus	Burlington NJ_2
Pituophis m. melanoleucus	Humphreys TN
Pituophis m. melanoleucus	Huntingdon TN
Pituophis m. melanoleucus	Marion GA_1
Pituophis m. melanoleucus	Marion GA_2
Pituophis m. melanoleucus	Ocean NJ
Pituophis m. melanoleucus	Richmond NC
Pituophis m. melanoleucus	Rockingham NC
Pituophis m. melanoleucus	Scotland NC
Pituophis m. melanoleucus	Winston AL
Pituophis m. melanoleucus	Hoke NC
Pituophis m. mugitus	Baker GA
Pituophis m. mugitus	Covington AL_1
Pituophis m. mugitus	Covington AL_2
Pituophis m. mugitus	Pierce GA
Pituophis m. mugitus	Turner GA
Pituophis m. mugitus	Alachua FL
Pituophis m. mugitus	Bristol FL
Pituophis m. mugitus	Brooksville FL
Pituophis m. mugitus	Cantonment FL
Pituophis m. mugitus	Hamilton FL
Pituophis m. mugitus	Hernando FL
Pituophis m. mugitus	Highlands FL
Pituophis m. mugitus	Lafayette FL
Pituophis m. mugitus	Liberty FL
Pituophis m. mugitus	Marion FL
Pituophis m. mugitus	Okaloosa FL
Pituophis m. mugitus	Okeechobee FL_1
Pituophis m. mugitus	Okeechobee FL_2
Pituophis m. mugitus	PonceDe Leon FL
Pituophis m. mugitus	Santa Rosa FL

REFERENCES

Andam, C. P., et al. "Population Structure of Pathogenic Bacteria." *Genetics and Evolution of Infectious Diseases*, Elsevier, 2017, pp. 51–70. doi:10.1016/B978-0-12-799942-5.00003-2.

Camargo, Arley, and Jack Sites. "Species Delimitation: A Decade After the Renaissance." *The Species Problem - Ongoing Issues*, edited by Igor Pavlinov, InTech, 2013. doi:10.5772/52664.

Crother, Brian I. *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, Kevin. "Species Concepts and Species Delimitation." *Systematic Biology*, vol. 56, no. 6, Dec. 2007, pp. 879–86. doi:10.1080/10635150701701083.

Devitt, Thomas J., et al. "Species Delimitation in Endangered Groundwater Salamanders: Implications for Aquifer Management and Biodiversity Conservation." *Proceedings of the National Academy of Sciences*, vol. 116, no. 7, Feb. 2019, pp. 2624–33. doi:10.1073/pnas.1815014116.

Felsenstein, Joseph. "Evolutionary Trees from DNA Sequences: A Maximum Likelihood Approach." *Journal of Molecular Evolution*, vol. 17, no. 6, Nov. 1981, pp. 368–76. doi:10.1007/BF01734359.

Flouri, Tomáš, Xiyun Jiao, et al. "A Bayesian Implementation of the Multispecies Coalescent Model with Introgression for Phylogenomic Analysis." *Molecular Biology and Evolution*,

edited by Michael Rosenberg, vol. 37, no. 4, Apr. 2020, pp. 1211–23. doi:10.1093/molbev/msz296.

Flouri, Tomáš, Bruce Rannala, et al. *A Tutorial on the Use of BPP for Species Tree Estimation and Species Delimitation*. p. 17.

Flouri, Tomáš, Xiyun Jiao, et al. "Species Tree Inference with BPP Using Genomic Sequences and the Multispecies Coalescent." *Molecular Biology and Evolution*, edited by Anne D Yoder, vol. 35, no. 10, Oct. 2018, pp. 2585–93. doi:10.1093/molbev/msy147.

Gustafson, Grey T., et al. "Ultraconserved Element (UCE) Probe Set Design: Base Genome and Initial Design Parameters Critical for Optimization." *Ecology and Evolution*, vol. 9, no. 12, June 2019, pp. 6933–48. doi:10.1002/ece3.5260.

Holland, B. R. "Using Consensus Networks to Visualize Contradictory Evidence for Species Phylogeny." *Molecular Biology and Evolution*, vol. 21, no. 7, Mar. 2004, pp. 1459–61. doi:10.1093/molbev/msh145.

Höhna, et. al. 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic Biology*, 65:726-736.

Korshunova, Tatiana, et al. "Multilevel Fine-Scale Diversity Challenges the 'Cryptic Species' Concept." *Scientific Reports*, vol. 9, no. 1, Dec. 2019, p. 6732. doi:10.1038/s41598-019-42297-5.

Nikolakis, Zachary Lamar. "Phylogenomics of the Eastern Pinesnake." Southeastern Louisiana University. July 2018.

Pyron, R. Alexander, and Frank T. Burbrink. "Neogene Diversification and Taxonomic Stability in the Snake Tribe Lampropeltini (Serpentes: Colubridae)." *Molecular Phylogenetics and Evolution*, vol. 52, no. 2, Aug. 2009, pp. 524–29. doi:10.1016/j.ympev.2009.02.008.

Rodríguez-Robles, Javier A., and José M. De Jesús-Escobar. "Molecular Systematics of New World Gopher, Bull, and Pinesnakes (Pituophis: Colubridae), a Transcontinental Species Complex." *Molecular Phylogenetics and Evolution*, vol. 14, no. 1, Jan. 2000, pp. 35–50. doi:10.1006/mpev.1999.0698.

Ruane, S., et al. "Coalescent Species Delimitation in Milksnakes (Genus Lampropeltis) and Impacts on Phylogenetic Comparative Analyses." *Systematic Biology*, vol. 63, no. 2, Mar. 2014, pp. 231–50. doi:10.1093/sysbio/syt099.

Schliep, Klaus, et al. "Intertwining Phylogenetic Trees and Networks." *Methods in Ecology and Evolution*, edited by Richard Fitzjohn, vol. 8, no. 10, Oct. 2017, pp. 1212–20. doi:10.1111/2041-210X.12760.

Schliep, Klaus Peter. "Phangorn: Phylogenetic Analysis in R." *Bioinformatics*, vol. 27, no. 4, Feb. 2011, pp. 592–93. doi:10.1093/bioinformatics/btq706.

Solís-Lemus, Claudia, et al. "Bayesian Species Delimitation Combining Multiple Genes and Traits in a Unified Framework: SPECIES DELIMITATION INTEGRATING GENES AND TRAITS." *Evolution*, vol. 69, no. 2, Feb. 2015, pp. 492–507. doi:10.1111/evo.12582.

Sukumaran, Jeet, and L. Lacey Knowles. "Multispecies Coalescent Delimits Structure, Not Species." *Proceedings of the National Academy of Sciences*, vol. 114, no. 7, Feb. 2017, pp. 1607–12. doi:10.1073/pnas.1607921114.

Winker, Kevin, et al. “Ultraconserved Elements (UCEs) Illuminate the Population Genomics of a Recent, High-Latitude Avian Speciation Event.” *PeerJ*, vol. 6, Oct. 2018, p. e5735. doi:10.7717/peerj.5735.

Woodruff, David S. “Populations, Species, and Conservation Genetics.” *Encyclopedia of Biodiversity*, Elsevier, 2001, pp. 811–29, doi:10.1016/B0-12-226865-2/00355-2.

Yang, Z., and B. Rannala. “Unguided Species Delimitation Using DNA Sequence Data from Multiple Loci.” *Molecular Biology and Evolution*, vol. 31, no. 12, Dec. 2014, pp. 3125–35. doi:10.1093/molbev/msu279.

Zhao, Yan, et al. “Species Delimitation for the Molecular Taxonomy and Ecology of the Widely Distributed Microbial Eukaryote Genus *Euplotes* (Alveolata, Ciliophora).” *Proceedings of the Royal Society B: Biological Sciences*, vol. 285, no. 1871, Jan. 2018, p. 20172159. doi:10.1098/rspb.2017.2159.