SYSTEMATICS AND HISTORICAL BIOGEOGRAPHY OF *AGKISTRODON CONTORTRIX* AND *AGKISTRODON PISCIVORUS*

By

Timothy James Guiher

A dissertation submitted to the Graduate Faculty in Biology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2011

This manuscript has been read and accepted for the
Graduate Faculty in Biology in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy.

Frank T. Burbrink

6/30/2011
Date

Chair of Examining Committee

Laurel A. Eckhardt

Date

Executive Officer

Stéphane Boissinot

Michael J. Hickerson

Christopher J. Raxworthy

Ana A. Carnaval

Supervision Comnmittee

The City University of New York

Abstract

SYSTEMATICS AND HISTORICAL BIOGEOGRAPHY OF *Agkistrodon contortrix* AND *Agkistrodon piscivorus*

by

Timothy James Guiher

Adviser: Frank T. Burbrink, PhD.

Many studies have revealed that lineages currently inhabiting formerly glaciated areas were pushed into southern glacial refugia and have expanded into their modern range since the last glacial maximum. There have been few studies that compare the effects of glacial cycles on lineage diversification, historical demography and migration rates in closely related species with overlapping ranges. In this study I compare phylogeographic structure, historical demography, approximate lineage age, potential distributions, and migration rates in two closely related and broadly co-occurring venomous snakes in eastern North America, the cottonmouth (*Agkistrodon piscivorus*) and copperhead (*A. contortrix*) using multilocus coalescent approaches. It has recently been discovered that gene flow between closely related species with adjacent distributions may be common (Nosil 2008). However, the absence of gene flow is a primary assumption of many phylogeographic methods including species tree inference and Bayesian species delimitation. I provide a framework for examining species delimitation when gene flow between species is present and provide a taxonomic revision of *A. contortrix* and *A. piscivorus*. In addition, I explore whether hybrids

between adjacent species inhabit unique environmental conditions not suitable to one or both species. Finally, I reveal that species diversification was likely a direct result of Pleistocene glacial cycles and that species with the closest proximity to formerly glaciated areas experienced population expansion following the retreat of the Laurentide Ice Sheet. A combination of population expansion out of refugia and niche expansion has resulted in hybridization between adjacent species where species distributions come into contact. It is not clear whether gene flow has persisted during speciation and subsequent interglacial periods or if it has only recently occurred following the last glacial maximum.

ACKNOWLEDGMENTS

I would like to thank the following persons and institutions for providing tissues and invaluable assistance during the preparation of this dissertation: D.B. Shepard & L.J. Vitt (Sam Noble Oklahoma Museum of Natural History), T.W. Taggart, C.J. Schmidt, & J.T. Collins (Sternberg Museum, Fort Hays State University), R. Lawson (California Academy of Sciences), J.A. McGuire & C. Spencer (Museum of Vertebrate Zoology), J. Boundy (Louisiana State University Museum of Zoology), D. R. Frost, C. J. Raxworthy, & D. Kizirian (American Museum of Natural History), T.J. Hibbitts (Texas A&M University), L.J. Rissler (University of Alabama), G. Watkins-Colwell (Yale University), B. Espinoza (California State University, Northridge), H. Lillywhite, R. McCleary, K.L. Krysko (University of Florida), D.C. Cannatella, T.J. LaDuc, & D. Hall (University of Texas, Austin), B.I. Crother (Southeastern Louisiana University), B. Moon (University of Lousiana, Lafayette), C. Phillips (Illinois Natural History Survey), Stan Trauth (Arkansas State University Museum of Zoology), B. Green, R. Combs (Missouri State University), P. Chippendale, C. Rolke (University of Texas at Arlington), C. Guyer, S. Graham (Auburn University), N. Russel (South Florida Collection Management Center Everglades National Park), D. Stevenson (Project Orianne Ltd.) K. Swaim (Swaim Biological Inc.), Kristen Wiley, Jim Harrison (Kentucky Reptile Zoo), F.M. Fontanella, N. Howe, Matt Harris, C.D. Camp, B. Hudson, A. Schwanabeck, K. K.J. Irwin, L.K. Irwin, C.G. Jimenez, S.R.

McNearney, S. Ruane, D. Siegel, and a number of anonymous reviewers. Funding for this research was provided in part by the Professional Staff Congress-CUNY, the American Museum of Natural History (Theodore Roosevelt Memorial Fund), and the College of Staten Island. I would also like to extend special thanks to my family and friends, my wife, Jennifer Costello for support over the last few years, Alex Pyron for providing invaluable advice, my advisor, Frank T. Burbrink, with whom the first chapter of this dissertation was co-authored in print, and my committee: Stéphane Boissinot, Michael J. Hickerson, Christopher J. Raxworthy, and Ana C. Carnaval.

TABLE OF CONTENTS

List of Tables

List of Figures

Introduction

Advances in Phylogenetics and Implications for Inferring the Evolutionary History of Two Species of Venomous Snakes

Combining Phylogenetics and Populations Genetics and the State of Systematics

The field of systematics has undergone considerable change over the last decade. The amount of genomic data available for both model and non-model organisms has steadily increased through the development of new technologies, including second and third generation sequencing, which have substantially reduced the cost, effort and time required to obtain large amounts of sequence data. In parallel, the increase in large multi-loci datasets has seen the appropriate development of phylogenetic theory and computational methodologies that use these data to more explicitly understand species diversity and reveal the processes that have led to this diversity (Edwards 2009). Molecular Systematic studies have traditionally focused on population genetics, phylogeography, or phylogenetics. Many researchers identify themselves with one of these distinct fields and focus on the types of questions and methodologies associated with each discipline. However, inferring the evolutionary history of a group may also require an understanding of how speciation occurs over short time frames and how interactions between species and populations over longer time frames impact our ability to reconstruct phylogenetic histories by influencing population parameters such as population size (N_e), time of divergence, and gene flow (Hey 2006; Edwards 2009). Traditionally, phylogenetics focuses on understanding deeper taxonomic relationships, while population genetics examines populations still in genetic contact (Avis et al., 1987). However, the processes influencing species histories at a node are the same for both types of studies. Recognizing the interplay between these fields requires investigators to combine data and methods from each discipline to

explicitly reconstruct phylogeographic histories and better understand diversity. This creates a somewhat circular dependence of methods from each discipline on parameter estimates from the other. Specifically, gene flow between populations can directly influence estimates of N_e for different genes differentially, adversely affecting inference of coalescence times and species trees (Hey 2006). In addition, species tree estimates assume no gene flow between species and require species to be delimited *a priori*, creating a dependence of species tree analyses on methods of population genetics (Edwards 2009, O'Meara 2010). Conversely, population genetic studies often require estimates of divergence times or mutation rates which can only be provided by phylogenetic tree inference (Hey and Nielsen 2007).

Over the last six years (the time frame of this study) researchers have progressed from inferring phylogeographic history almost exclusively on single mtDNA loci toward studies that incorporate multiple sources of evidence. Phylogeographic explorations now routinely incorporate multiple single copy nuclear (scDNA) markers in addition to mtDNA. In three issues of the journal Molecular Phylogenetics and Evolution, a common repository for phylogeographic and phylogenetic studies, 78% of articles published include multiple independent loci, in contrast to just 38% in the same number of issues in 2005. This transition has been in response in part to both the availability of scDNA markers for a greater amount of taxonomic groups as well as the growing distinction between gene trees and species trees. Moreover, examining the effects of gene tree discordance and developing methods that can utilize multiple unlinked loci to resolve the evolutionary history of species, in contrast to genes, has resulted in an increasing

body of literature on the subject (Edwards 2009; Liu et al. 2009; Leache and Rannala 2011). It is now widely accepted that individual gene trees do not necessarily represent species history and can be misled by several processes, such as incomplete lineage sorting, gene flow, and recombination. Methods have been developed that attempt to resolve gene discordance and infer species trees, examine changes in historical population size, estimate migration rates, and delimit species. The majority of methods have focused on addressing the influence of incomplete lineage sorting on gene discordance. Estimating gene flow has been a central goal for population genetic studies and models have been developed that incorporate multiple loci and combine different types of markers, i.e. microsatellites and STR's, that make it possible to estimate gene flow (i.e. Migrate, Structure and IMa). Tree based methods incorporating gene flow by inferring phylogentic networks have been developed and recently Yu et al. (2011) proposed a model that accommodates both gene flow and ILS. While this is encouraging, further work is required to improve current methods for inferring species relationships and examining evolutionary processes accommodating multiple sources of gene tree discordance.

Future developments are required that can simultaneously assess species delimitation (O'Meara 2010; Leache and Rannala 2011) and infer species trees as well as account for multiple sources of discordance. A major obstacle in the way of achieving this goal is the computational burden that results from the increased number of parameters that such a method requires. The current approach to circumvent this problem is to assume that the assignment of individuals to species is known and that

there has been a single source of discordance, most often deep coalescence (Liu et al. 2009; Heled and Drummond 2010). The recent rapid increase in our ability to produce large multi-loci data sets as a result of second and third generation sequencing technologies has exasperated the need for alternative methods. Studies are now emerging that contain hundreds or even thousands of loci (Morin et al. 2010; Burleigh et al. 2011). In fact, it is foreseeable that studies incorporating complete genomes for a large number of taxa will be possible in the near future, increasing the computational burden of our current methods. This is not to say that we cannot utilize current methods to gain insight into the phylogeographic history of species. What it does mean is that care must be taken to ensure that assumptions of the methods used are met and how violating assumptions can influence the outcome. I attempt to illustrate over the next four chapters how this can be achieved by incorporating methods from population genetics, phylogeography, phylogenetics, and ecological niche modeling to improve our ability to infer evolutionary histories by examining the phylogeographic history of two species of venomous snakes in North America, the North American copperhead (*Agkistrodon contortrix*) and the cottonmouth (*A. piscivorus*).

Taxonomic History of Agkistrodon

Snakes within the genus *Agkistrodon* have received considerable attention from several fields of biology including taxonomy, behavior, ecology, physiology, and recently molecular biology. A query of *Agkistrodon* in google scholar provides a rough estimate of 8,700 papers published on the genus. Behavioral biologists have used both *A.*

contortrix (the North American copperhead) and *A. piscivorus* (the cottonmouth) as model organisms to study the defensive behavior of venomous snakes (Gibbons 2002; Glaudas et al. 2006). Ecologists have focused largely on radio tracking populations of copperheads and cottonmouths to describe home range size and microhabitat use (Roth 2005; Glaudas et al. 2006). Physiologists have contributed to our understanding of reproductive cycles and metabolic rates of snakes by studying both *A. contortrix* and *A. piscivorus* (Zaidan III 2003; Graham et al. 2008). Recently, medical researchers have focused on the use of venom proteins isolated from species of *Agkistrodon* to treat several forms of cancer (Trikha et al. 1994; Zhou et al. 2000; Swenson et al. 2005). Some key factors contribute to the broad interest in these two species: (i) they are common throughout most of their range making large sample sizes easily obtainable, (ii) are responsible for ~37% of the non-lethal human envenomations in the US (iii) are widely distributed, and (iv) cross several barriers to gene flow previously identified in a wide range of taxonomic groups.

Describing the evolutionary history of the species in the genus *Agkistrodon* was the focus of two of most distinguished herpetologists of the twentieth century, Howard K. Gloyd and Roger Conant. Nearly sixty years of research was compiled by the two authors into their monograph; Snakes of the *Agkistrodon* Complex (Gloyd and Conant 1990). Gloyd and Conant included seven Old World species in the genus *Agkistrodon* (*sensu lato*). This taxonomic scheme was challenged by Knight et al. (1992) based on restriction length polymorphisms and a partial 16S rDNA dataset, and later confirmed by Parkinson et al. (1997) using a more complete 16S rDNA dataset. Therefore, the genus

Agkistrodon is restricted to the New Word species *A. contortrix*, *A. piscivorus*, and *A. bilineatus*, while *Gloydius* is now used for the Old World species (Hoge and Romano-Hoge 1981). Subsequently, the Central American species, *A. bilineatus* was split into two species, *A. bilineatus* and *A. taylori* based on mtDNA evidence (Parkinson et al. 2000). Since this study is focused on the copperhead and the cottonmouth I will only discuss the results of Gloyd and Conant (1990) that pertain to *A. contortrix and A. piscivorus*.

The famous American naturalist, Mark Catesby first mentioned the cottonmouth in his account of the taxon in 1743, describing it as *Vipera aquatic* (Catesby et al. 1743). The proper specific names were first used by Lacépède (1789), referring to *Crotalus piscivorus* and then by Linneaus (1766), who described the copperhead as *Boa contortrix*. The generic name, *Agkistrodon* (meaning fishhook tooth), was proposed by Palsiot de Beavois (1799). There are currently five subspecies of *A. contortrix* and three subspecies of *A. piscivorus* (Fig. 1) recognized by the SSAR reptile taxon list (Crother et al. 2011) based on the morphological dataset compiled by Gloyd and Conant; *A. c. contortrix* (Linneaus 1766), *A. c. mokasen* (Palisot de Beauvois 1799), *A. c. phaeogaster* (Gloyd 1969), *A. c. laticinctus* (Gloyd and Conant 1934), *A. c. pictigaster* (Gloyd and Conant 1943), *A. p. piscivorus* (Lacepède 1789), *A. p. conanti* (Gloyd 1969), and *A. p. leucostoma* (Troost 1836). The characters used in their diagnosis were limited to scutellation and color pattern (see Tables 1-11 in Gloyd and Conant 1990). These characters, such as coloration and shape of crossbands, are highly subjective, and with respect to scutellation counts, are variable with overlapping values among taxa. The current diagnosis does not result in a robust taxonomy supported by diagnostic

characters sufficient to reliably assign individuals at the sub-specific level. For example, one of the key characters used for identifying individuals of *A. p. leucostoma* is that specimens tend to darken at an earlier age. This character becomes useless in the absence of a large sampling of all three subspecies. To compound the shortcomings of this dataset, Gloyd and Conant (1990) included a disclaimer in their monograph regarding the compilation of their dataset, "Gloyd followed the typological concept of assigning individual specimens to one race or another, whereas snakes that were morphologically intermediate between adjacent subspecies he labeled as intergrades." The result of this is that their summary of scuttelation and statistics used in diagnosing the subspecies only included individuals that could be assigned to one of the subspecies. The resulting range maps include large areas designated as intergrade zones, which limits the ability of researchers to use geographic origin of specimens as a tool for assigning individuals to the proper sub-specific taxon throughout most of their range in the absence of any useful diagnostic characters (see Figs. 6 and 10 in Gloyd and Conant 1990). The difficulty in diagnosing and properly understanding diversity within these species highlights the need for a molecular approach to examining the phylogeographic histories of *A. contortrix* and *A. piscivorus.*

Modern phylogenetic methods

Modern statistical phylogenetic methods combined with the increasing ease with which molecular datasets can be compiled has revolutionized the field of taxonomy and broadened the scope of questions that can be addressed by evolutionary biologists

(Castoe et al. 2007; Burbrink et al. 2008). Additionally, the concept of recognizing subspecies has been challenged and is becoming less common (Frost and Hillis 1990; Burbrink et al. 2000). This shift in taxonomy away from sub-specific designation is based on the argument that independently evolving lineages supported by either molecular or morphological evidence warrant species recognition. Criterion for sub-specific recognition have never been clearly defined, therefore subspecies represent an arbitrary designation. The current taxonomic proposals for both *A. contortrix* and *A. piscivorus* are hampered by (i) a lack of sufficient diagnostic characters, (ii) the practice of excluding individuals not supported by morphological characters from analysis of those same characters by Gloyd and Conant, and (iii) adherence to an antiquated taxonomic scheme that includes the recognition of subspecies. In the following chapters I present data based on multiple independent loci to address the phylogeographic and biogeographic histories of both species, while providing a strong case for correcting taxonomy based on intersection of genetic, ecological and morphological data.

Specifically, I first present a mtDNA phylogeographic analyses that identifies distinct lineages not congruent with current sub-specific assignments, with exception to *A. p. conanti*. In addition, divergence times and historical demographics are inferred. The results presented were used to guide additional sampling efforts. More importantly, the results from this chapter are contrasted in the chapters that follow that rely on using multiple unlinked loci to explore the pitfalls of interpreting phylogeographic history using single or linked loci. This may seem trivial at this point but as mentioned earlier almost 22% of studies in the most recent issues of MPE were based on mtDNA

data alone. This is followed by a multi-locus analysis that includes increased sampling that attempts to outline protocols for species delimitation in the presence of gene tree discordance when the assumption of “no gene flow” is violated. I infer species trees using two current methods in order to estimate key parameters, such as mutation rates and divergence times, and evaluate the effect parameter estimates have on subsequent analyses, including estimates of migration rates and historical demographics. Ultimately, I propose taxonomic revisions of both species by drawing evidence from phylogeographic analyses, ecological niche models and morphological diversity.

Chapter 1: Inferences from mtDNA

Demographic and Phylogeographic Histories of Two Venomous North American Snakes of the Genus *Agkistrodon*

Introduction

A common goal among researchers examining the phylogeographic and historical demographics of organisms inhabiting temperate regions of the world is to understand population responses to Pliocene and/or Pleistocene glacial cycles. Many of these studies reveal that lineages currently inhabiting previously glaciated areas were pushed into southern glacial refugia (Crespi et al. 2003; Starkey et al. 2003) and have expanded into their modern range after the last glacial maximum (Hewitt 1996; Hewitt et al. 2001; Starkey et al. 2003; Runck and Cook 2005). For many taxa, this ebb and flow has left characteristic genetic signatures that reveal population bottlenecks in the contraction phase and population growth in the expansion phase (Runck and Cook 2005; Soltis et al. 2006). A direct benefit of the coalescent theory revolution in population genetics (Tajima 1989; Kingman 2000) has been the creation of statistical tests designed to examine changes in effective population sizes (N_e) given haplotype or genealogical information. It is then commonly assumed that demographic changes through time in an area of interest result from changes in habitat suitability due to the impact of glacial cycles on regional climate in temperate regions (Waltari and Cook 2005; Douglas et al. 2006; Castoe et al. 2007; Waltari et al. 2007).

In eastern North America, movement of the Laurentide Ice Sheet has been identified as one of the primary factors responsible for shifting the ranges and impacting the population sizes of organisms inhabiting these areas. The magnitude of this event cannot be understated; the ice sheet occupied roughly 13 million square kilometers,

was greater than 2,000 meters thick in many areas, and extended as far south as New York City in the East and central Illinois in the West (Denton and Hughes 1981; Dyke and Prest 1987; Pielou 1991). Since the beginning of the Pleistocene, there were four major glacial and interglacial periods. The last one, the Wisconsinan, began 70,000 years ago (ya) and reached the last maximum 20,000 ya before receding, resulting in current Holocene climatic conditions (Killey 1998). This final glacial episode is often cited as a major contributor to the fluctuating demographic patterns observed in many species with distributions in eastern North America (Loveless and Hamrick 1988; Tegelström et al. 1990; Runck and Cook 2005). It should be noted that not all studies have revealed the same population demographic patterns of decrease in effective population size at glacial maxima and increase during interglacials (Rowe et al. 2004; Burbrink et al. 2008). Many factors may be responsible for variation in these demographic patterns including habitat requirements, location of species in refugia and proximity to the glacial front, population variation in genetic markers, overall ancient effective population size, sampling bias, and responses (or lack thereof) to changing habitats (Ramos-Onsis and Rozas 2002; Rowe et al. 2004). Here, I offer a direct comparison of demographic responses of populations in both formerly glaciated areas as well as unglaciated areas, while considering phylogeographic relatedness, similarity in life history traits, age of lineages, and timing of population size responses to environmental changes.

I provide a comparative analysis of population demographic responses to glacial cycles by conducting phylogeographic analysis of two closely related and broadly co-occurring venomous snakes in North America. The cottonmouth (*Agkistrodon piscivorus*)

and copperhead (*A. contortrix*) are two of the most commonly encountered venomous snakes in the eastern United States. Together, these snakes represent ~30% of the non-lethal human envenomations in this region and are well known to most people living in this region of the United States. In the southern areas of their distributions, they occupy the same range and are nearly syntopic (Fig. 2; Gloyd and Conant 1990). However, major differences exist in their preferred habitats in these areas: *A. piscivorus* tends to be restricted to lowland shaded swamps, whereas *A. contortrix* is much more of a habitat generalist, living in swamps, deciduous hardwoods, pine forests, scrub desert, and high elevation mountain regions (Gloyd and Conant 1990; Conant and Collins 1991). The ranges of the two species overlap in the South, areas considered glacial refugia for many taxa (Waltari et al. 2007). However, the range of *A. contortrix* extends further north occasionally into portions of formerly glaciated regions. The modern range of *A. piscivorus* is generally confined to southern parts of the US and is absent from areas of higher elevation such as the Appalachian Mountains.

I examined the phylogeographic and demographic histories of *A. contortrix* and *A. piscivorus* to test specific hypotheses regarding changes in effective population size (N_e) due to glacial cycles for these species. Considering just their modern distributions, it would be expected that lineages of *A. contortrix*, which currently inhabit formerly glaciated regions, would yield genetic evidence of a population bottleneck during glacial maxima and expansion or population recovery during interglacials (Fig. 2a). In contrast, it is expected that lineages of *A. piscivorus* currently confined to unglaciated areas would show only constant population sizes through time (Fig. 2b). In order to test these

hypotheses I: i) compare the phylogeographic structure of these closely related venomous snakes in areas of overlap and throughout their range to determine if they share similar genetic structure due to the same geological or habitat features, ii) infer divergence dates within the genus *Agkistrodon* as well as for lineages identified within *A. contortrix* and *A. piscivorus* to assess temporal changes in N_e with respect to glacial/environmental change and iii) estimate the demographic history of all lineages in order to identify patterns associated with the proximity of their current distribution relative to formerly glaciated areas.

In this chapter I focus solely on inferences from mtDNA for several reasons. First, phylogeography relied heavily on mtDNA at the time that this study first started. Therefore, obtaining and analyzing mtDNA sequence data was a primary focus of this study from its inception. While recent studies have highlighted reasons why single gene histories may not represent the species history, the phylogeographic and biogeographic histories of a great number of taxa have been inferred over the last decade using mtDNA only. In part, I am illustrating how methodological advances may impact inferences of phylogeographic history. This chapter provides the first part necessary for this comparison by providing interpretations of the phylogeographic structure and biogeogrphic history of both species based on mtDNA.

Methods and Materials

Sequence Acquisition

I obtained 154 tissue samples of *A. contortrix* and 82 tissues of *A. piscivorus* taken throughout their known ranges for my phylogeographic analyses (Fig. 2). I also sequenced two other species of *Agkistrodon* along with sequences obtained form genbank*, (A. bilineatus and A. taylori).* Accession numbers for the two cantil sequences are: AY223614 and AY223613. The standard method of proteinase K digestion in lysis buffer followed by several rounds of phenol/$CHCl_3$ extraction (Sambrock and Russell 2001) was used to obtain total genomic DNA from samples of shed skin, liver or muscle tissue or whole blood. Extracted DNA was precipitated with three volumes of absolute isopropanol and the pelleted DNA was then washed with two changes of 70% EtOH before redissolving in distilled water (Sambrock and Russell 2001). Template material for the polymerase chain reaction (PCR) consisted of samples with DNA/RNA ratios of 1.5-2.1 and DNA concentrations from 10-200ng/ul. Using the PCR with a negative control (water), the complete gene sequence for the mitochondrial gene Cytochrome b (Cytb) was amplified using the specifications included with the AccuTaq Jumpstart Kit (USB Corp.) for a 25ul reaction. The following thermal cycling conditions successfully amplified this gene: 95º C for 2 min followed by 35 cycles of 94º C (60 s), 47º C (30 s) and 68º C (60 s). The PCR products were cleaned using 1ul of ExoSap-it (USB Corp.) per 10 ul of PCR product. The sequencing reaction consisted of 2 ul of DTCS (Beckman-Coulter), 1 ul of 2 uM primer and 3 ul of DNA template. Sequences were purified following the protocol from the Agencourt CleanSeq Dye-terminator removal kit (Beckman Coulter, USA) and analyzed on a Beckman-Coulter CEQ 8000 sequencer. Primers for PCR and cycle sequencing reactions were as follows: Cytb amplification

L14910 (de Queiroz et al. 2002) and H16064 (Burbrink et al. 2000), for sequencing I used L14910 (de Queiroz et al. 2002) and an internal primer designed specifically for this project; Agk SEQ (CGCAATCAATGACCCAACACT). Subsequently, the nucleotide sequences were aligned and examined by eye using the program Sequencher 4.1.2 (Genecodes 2000), and an open reading frame for this gene was determined. There were no gaps in this gene for any of the taxa sampled here.

Phylogeographic Estimation

Phylogeographic relationships of *Agkistrodon contortrix* and *A. piscivorus* were estimated separately, along with appropriate out-groups using Maximum Likelihood (ML) and Bayesian inference (BI). A single sample of *A. piscivorus* was included as an out-group for *A. contortrix* while *A. taylori* was used as an out-group for *A. piscivorus*. The appropriate models for the ML analyses were selected using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) in the program Modeltest version 3.06 (Posada and Crandall 1998; Posada and Buckley 2004) with each starting tree obtained using the neighbor–joining algorithm in PAUP* v4.10b (Swofford 2003). The ML trees and associated support were obtained from 1,000 nonparametric bootstrap pseudoreplicates (Felsenstein 1985) under the preferred AIC and BIC models using the program RAxML-VI-HPC MPI v. 2.2.3 (Stamatakis et al. 2005). Nonparametric bootstrap values above 75% were considered good support for a bipartition (Hillis and Bull 1993).

To infer trees and to assess tree support using models incorporating evolutionary information specific to codon position, I performed partitioned model analyses using BI with MrBayes version 3.1 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003). Prior to tree inference, two substitution models were evaluated for each dataset. The first model accounts for differences in evolutionary rates of each of the three codon positions of the Cytb gene using the GTR + Γ + I model with estimated nucleotide frequencies for each codon position. For this codon–position–specific model, abbreviated 3(GTR + Γ + I), the underlying phylogeny was linked across partitions, but all other model parameters were unlinked among partitions. The second method simply applied the GTR + Γ + I model across all positions with no partitioning among codon positions. For each model, two independent searches were executed to insure convergence of all parameters by comparing the variance across chains within a search to the chain variance among searches using Gelman and Rubin's "r" statistic (Gelman et al. 1995). Searches were considered burned-in when the values for "r" have reached ~1.00. All searches consisted of three "heated" and one "cold" Markov chain estimated for 20 million generations, with every 1000th sample being retained. Uninformative priors were applied to all parameters based on the program's default settings. Parameter stationarity was assumed to have occurred when –lnL values for chains converged on similar values in all replicates for each model yielding a split standard deviation less than 0.005. Trees prior to stationarity were discarded. Using Posterior Bayes factors (PBF) to determine the appropriate model for the sequence data requires that the harmonic mean for the model likelihood $f(X|M_i)$ be estimated from

the values in the stationarity phase of a Markov chain Monte Carlo run following the methodologies of Newton and Raftery (1994). To compare models, PBF followed the form $2\log_e B_{10}$, where B_{10} is the ratio of model likelihoods. A Posterior Bayes factor value greater than 10 was considered strong evidence favoring the more parameter rich model (Kass and Raftery 1995). These BI trees were compared with those obtained using ML and the most credible inferences of relationship were confined to nodes where the posterior probability (Pp) was greater than 95% and the nonparametric bootstraps were greater than 75% (Hillis and Bull 1993).

Divergence Dating

Using the Program BEAST v1.4.6 (Drummond and Rambaut 2007) I estimated the date of origin of all species of *Agkistrodon* and each lineage using a 'relaxed phylogenetics' method (Drummond et al. 2006). This method has the benefit of inferring divergence dates using a Bayesian relaxed molecular clock while incorporating tree uncertainty in the MCMC process. An uncorrelated lognormal tree prior and node constraints obtained from the fossil record with a lognormal prior distribution and mean equal to the age of the fossil were used to estimate divergence dates throughout the tree (Table 1). In order to place calibration references on the tree several additional outgroups were used: *Pareas macularius*, *Bitis nasicornis*, *Atheris nitschei*, *Daboia russelli, Ovophis okinavensis, Crotalus horridus* and *C. viridus*. These samples were taken from GenBank (http://www.ncbi.nlm.nih.gov/) under the respective accession numbers: AF471082, AY188009, AF471070, AF471076, AB175670, AF337057 and AF471066. In

addition, a single specimen of *Sistrurus miliarius* was sequenced for the Cytb gene. When using all of the in-group specimens and very divergent out-groups the model parameters estimated are not appropriate in all parts of the tree due to the variance present in such a heterogeneous dataset. Additionally, when estimating divergence times between species, it is necessary to assume either a coalescent tree prior (i.e. constant population size, exponential growth, logistitic growth, or expansion growth) or a Yule tree prior (a feature in BEAST v1.4.6), neither of which are appropriate. To avoid these imbalances, divergence estimates were inferred for a reduced dataset, which included two individuals from each phylogeographic lineage of *A. contortrix* and *A. piscivorus* along with the out-groups listed above. The analysis was run for 50 million generations, with samples retained every 1000th generation. I applied the GTR+Γ+I model partitioned by codon position for this analysis and a Yule tree prior. The initial 10% was discarded as burn-in.

Four fossil calibrations for the tree were obtained for various groups of viperids: 1) the oldest fossil representing the family Viperidae, found in central Europe, is from the Early Miocene (Late Arikareean; Rage 1982), 2) the first appearance of the subfamily Crotalinae was from the Early Miocene (Hemingfordian; Holman 2000), 3) the oldest fossil from the genus *Sistrurus* dates to the late Miocene (Clarendonian; Parmley and Holman 2007) and 4) the genus *Agkistrodon* first appears in the Late Miocene (Late Hemphillian; Holman 2000). These four fossils listed above were placed at the stem of the origin of the Viperidae, Crotalinae and *Agkistrodon*. The following medians and standard deviations for these constraints were used: Viperidae = 23.8 mya, SD = 0.15,

95% Prior Credible Interval [PCI] = 18.36-30.08; Crotalinae = 16.4 mya, SD = 0.1, PCI = 13.91-19.33; *Sistrurus* = 9.0 mya, SD = 0.3, PCI = 5.50-14.74; *Agkistrodon* = 6.5 mya, SD = 0.2, PCI = 4.678-9.032 (Table 1). The following priors were used for demographic estimates: Eastern clade = 0.105 mya, PCI = 0.020-0.544; Central clade = 0.204 mya, PCI = 0.045-0.896; Western clade = 0.208 mya, PCI = 0.040-1.078; Continental clade = 0.203 mya, PCI = 0.089-0.462; Florida Clade = 0.343 mya, PCI = 0.084-1.388.

Historical Demography

I examined past population dynamics of each phylogeographic lineage of *A. contortrix* and *A. piscivorus* using Bayesian skyline plots (BSP; Drummond et al. 2005). This technique permits the estimation of population size through time and does not require a pre-specified demographic model (e.g, constant size, exponential growth, logistic growth, or expansive growth) prior to the analysis. So that I could estimate relative changes in N_e for each lineage against a time scale in years rather than expected substitutions/site, I estimated the age of each lineage to use as a prior before examining BSPs (see above).

The appropriate phylogenetic model for each clade, as chosen by AIC and BIC in the program Modeltest version 3.06 (Posada and Crandall 1998; Posada and Buckley 2004), along with an uncorrelated lognormal relaxed clock were used to estimate Bayesian skyline plots in BEAST v1.4.6 (Drummond et al. 2003; Drummond et al. 2005) for each lineage. The root age prior for the most recent common ancestor (MRCA) of each lineage was obtained from the divergence dating analysis (see above). The

standard deviation of the root age prior was also chosen to approximate the distribution around the MRCA of each lineage from the divergence dating analysis. I used 15 grouped coalescent intervals (*m*) for all BSP analyses. All priors for the phylogenetic model and the population sizes were uniformly distributed. These analyses estimated genealogies and model parameters every 1000th iteration for $20x10^6$ generations with the initial 10% of samples discarded as burn-in.

To provide other estimates of population size changes, I examined mismatch distributions in the program Arlequin version 3.0 (Schneider et al. 2000). A unimodal mismatch distribution indicates a recent expansion, a multimodal (including bimodal) mismatch distribution indicates diminishing population sizes or structured size, and a ragged distribution suggests that the lineage was widespread (Excoffier et al. 1992; Rogers and Harpending 1992; Rogers et al. 1996; Excoffier and Schneider 1999). The multimodal distribution may indicate that the population is influenced by migration, is subdivided, and/or has undergone historical contraction (Marjoram and Donnelly 1994; Bertorelle and Slatkin 1995; Ray et al. 2003). Statistical significance of these distributions was assessed using sum of squares distances and Harpending's raggedness index (r_g) (Harpending 1994). I also examined demographic patterns in each lineage using Fu's *F* (Fu 1997) in the program DnaSP 4.10.8 (Rozas et al. 2003).

Results

Phylogeographic Structure and Divergence Dates

The Cytb gene for these species was aligned by eye, as there were no gaps and

all 1107bp were in reading frame for every sample. The most appropriate ML substitution model given these data for both species using AIC and BIC in Modeltest version 3.06 (Posada and Crandall 1998) was GTR+ Γ +I. The parameter values for phylogeographic estimates of *A. contortrix* were: Γ = 1.1186 and I = 0.8846, rAC = 1.0000, rAG = 67.5790, rAT = 0.0000, rCG = 0.0000, rCT = 34.3186, rGT = 1.0000. The following paramater values are for phylogeographic estimates of *A. piscivorus*: Γ = 0.8672, I = 0.7574, rAC = 0.0867, rAG = 2.4921, rAT = 0.0000, rCG = 0.0000, rCT = 2.4921, rGT = 1.0000. Additionally, posterior Bayes factors chose the most parameter rich model for BI estimatation for both species (*A. contortrix* PBF = 512.86, partitioned –Ln = -2482.88, unpartitioned –Ln = -2739.31, *A. piscivorus* PBF = 495.84, partitioned –Ln = -2257.17, unpartitioned –Ln = -2505.09), where the GTR+Γ+I model was applied separately to all three codon positions of Cytb. Bayesian and ML trees were congruent and high bootstrap (BS) and posterior probability (Pp) support was returned for all major nodes and phylogeographic lineages of both species (Figs. 3 and 4).

The placement of *A. bilineatus* and *A. taylori* as a monophyletic sister group to *A. piscivorus* received only moderate support in the BI (74% Pp). This relationship is in agreement with previous studies by Parkinson et al. (2000), Parkinson et al. (2002) and Castoe and Parkinson (2006). Each phylogeographic lineage within *A. contortrix* and *A. piscivorus* was well supported as indicated by Pp and BS. The node subtending the Eastern and Central clades received only 64% Pp, indicating that I cannot be certain which lineage was the first to split within *A. contortrix*.

Three geographically distinct lineages were inferred for *Agkistrodon contortrix*

found in the following areas based on current sampling:

1) Eastern: The approximate range of this lineage ranges from Connecticut to Florida along the Atlantic coast, west to include the Mobile Basin in the South and finally extending north into southeastern Illinois (Fig. 2a).

2) Central: This lineage extends from eastern Texas north through eastern Oklahoma into eastern Kansas and continues to just east of the Mississippi River north of the Mississippi embayment and south to western Alabama including the Mobile Basin (Fig. 2a).

3) Western: The distribution of this lineage includes Western and central Texas, central Oklahoma and eastern Kansas where it overlaps with the Central lineage (Fig. 2a).

Divergence dating with error estimation indicates that the origin of *A. contortrix* occurred in the Late Miocene (~6.6 mya; Fig. 5; Table 2) and lineage diversification began in the Early Pleistocene. The first major division was estimated to have occurred between the Western and Central/Eastern lineages in the Early Pleistocene (~1.4 mya). Subsequently, the division between the Eastern and Central lineages occurred in the Early Pleistocene (~1 mya). The phylogenetic uncertainty in the proper placement of the Western lineage of *A. contortrix*, is reflected in the error around the date estimates of

these lineages (Table 2).

In contrast, only two major lineages were discovered for *A. piscivorus.* These lineages are not concordant with geographic areas of genetic discordance found within *A. contortrix*:

1) Florida: The approximate range of this lineage extends from southern Florida to southern Georgia, and west to eastern Alabama (Fig. 2b).

2) Continental: This lineage occupies the remainder of the distribution of *A. piscivorus* from southeastern Virginia to central Georgia, east of the Appalachian Mountains, north to southern Illinois and central Oklahoma, south into central Texas in the west (Fig. 2b)

The origin of *A. piscivorus* was estimated to have taken place in the Late Miocene/Early Pliocene (~5.3 mya; Fig. 5; Table 2) and the separation between the Florida and Continental lineage occurred during the Late Pliocene (~2.5 mya). The initial separation between these two lineages possibly took place during the inundation of the Florida Peninsula resulting in the formation of small islands during an interglacial. This would effectively separate Florida from the continental United States (Webb 1990; Clark et al. 1999).

The MRCA of all extant haplotypes for each lineage of *A. contortrix*, and separately, each lineage of *A. piscivorus* originated between 0.093-0.306 mya (Table 3).

The median date (and SD accounting for the error in date estimation) for the MRCA of all haplotypes for each clade was used as the root time when examining BSP demographics for the respective lineages (see below).

Historical Demography

Median estimates from the Bayesian skyline plots revealed population expansion following the last glacial maximum (~20,000 ya) in all three lineages of *A. contortrix* and one lineage of *A. piscivorus* (the Continental lineage). It should be noted that error around these estimates of N_e through time must be considered and this reduces certainty in demographic trajectories for all lineages excluding the Continental lineage of *A. piscivorus* and the Central lineage of *A. contortrix* (Fig. 6). Contradictory to my initial prediction of constant population size for all lineages of *A. piscivorus*, the Continental clade showed the most extreme population growth of all lineages. The Florida lineage of *A. piscivorus* revealed very weak population growth after approximately 125,000 years before present. Within *A. contortrix*, the Eastern lineage had the strongest signal indicating population expansion after the last glacial maximum. The Central lineage also revealed population expansion while the Western lineage, similar to the Florida lineage, showed weak population growth after 125,000 years before present. The date for the MRCA of the haplotypes for each lineage occurred more recently in lineages that show expansion (i.e., the Eastern, Central and Continental lineage) than in those with little or no growth (i.e., the Florida and Western lineages) (Table 3).

Observed pairwise differences between haplotypes (mismatch distributions) were unimodal, indicating population expansion, for all of the lineages of both species with the exception of the Florida lineage of *A. piscivorus,* which was multimodal. The sudden expansion model was rejected for only the Florida lineage based on the raggedness statistic (HRI = 0.206, $p = 0.03$) and marginally supported for the Western lineage (Table 3). In agreement, Fu's *F* statistic was positive for the Florida lineage suggesting constant or contracting population size (Table 3). The remaining lineages all have negative Fu's *F* values offering further support that these lineages have undergone recent expansion, (Table 3) however, Fu's *F* values were significant only for the Continental and Central lineages. It should be noted that the number of segregating sites has a substantial effect on the power of Fu's *F*, which has been shown to have more power than tests based on the mismatch distribution such as r_g (Ramos-Onsis and Rozas 2002). The number of segregating sites (S) ranged between 6 and 15 for the five lineages tested, which is below the point at which these tests lose power ($S<30$ for $n>50$). Low number of segregating sites also results in similar mismatch distributions for both constant populations and expanding populations (Ramos-Onsis and Rozas 2002). For this reason, coalescent methods such as BSP, which take genealogy into account, may provide a better estimate of demographic history given these data.

Discussion

Phylogeography and Divergence

The major phylogeographic divisions within these closely related large-bodied viperids are not concordant; despite the fact the ranges of the species overlap throughout much of their southern distributions and are in contact with the same barriers. The copperheads (*A. contortrix*) diverged from the ancestor of the remaining three species of *Agkistrodon* during the Late Miocene with the cottonmouths (*A. piscivorus*) originating shortly after this time in the Late Miocene/Early Pliocene (Fig. 5; Table 2). I inferred three phylogeographic lineages of *A. contortrix* and two lineages of *A. piscivorus* (Figs. 3 and 4). Both of the genetic breaks between the three lineages of *A. contortrix* occur within the distribution of the single Continental lineage of *A. piscivorus* (Fig. 2). The other lineage of *A. piscivorus* is found in peninsular Florida, where *A. contortrix* does not occur. The division between the two lineages of *A. piscivorus* occurred approximately one million years before the diversification of the three lineages of *A. contortrix* took place (Fig. 5; Table 2). This indicates that *A. contortrix* has undergone greater diversification in less time than *A. piscivorus*. Although these are two closely related species occupying similar distributions, they occur in slightly different microhabitats, which may offer a possible explanation for their unique responses to changing climates during the Pliocene and Pleistocene (see below).

The divergence dating analysis indicates that the copperheads evolved during the Late Miocene and the principal lineage formation between the Western and Central/Eastern lineages occurred in the late Pliocene/early Pleistocene (Table 2). An ecological transition zone between eastern temperate forests and the Great Plains marks the division between the Western and Central lineages. Elevation increases from

less than 100 meters to over 2000 meters (Bailey 1995). This is combined with a decrease in precipitation from an average of 1,000mm to an average of 255mm (Bailey 1995). In addition, habitat changes in this area from mixed forest and cypress swamp to prairie grassland (Bailey 1995). This prairie steppe habitat, which the Western lineage currently occupies, first appeared in the Early Pliocene and remained stable until the Pleistocene (Emslie 2007). The extent of this habitat subsequently contracted and expanded during Pleistocene glacial events. I place the origin of the Western lineage at approximately 1.4 mya (Fig. 5; Table 2), which roughly corresponds with the beginning of Pleistocene glacial cycles. It is possible that this initial split may have been a result of populations being confined to distinct southern refugia during glacial maxima, isolating western populations from those east of the Great Plains. This ecological transition zone has been identified as a genetic break in passerine birds (Emslie 2007) and cornsnakes (*Pantherophis guttatus* complex, in part; Burbrink 2002). We cannot be certain that the Pliocene distribution of *A. contortrix* was the same as it is today. Other barriers farther east (e.g., the Mississippi River; Burbrink et al. 2000) may have been the primary cause for this initial diversification event with subsequent shifting of geographic lineages into their current distribution.

The Eastern lineage diverged from the Central lineage of *A. contortrix* in the Pleistocene, ~1 mya (Fig. 5; Table 2). This division appears to have occurred at the Mobile Basin in the South. The Mississippi River does not appear to separate these two lineages in the north like other squamates (lizards and snakes) in the Eastern US (Burbrink et al. 2000; Burbrink 2002; Leache and Reeder 2002; Soltis et al. 2006; Howes

et al. 2009). It has been suggested (Burbrink et al. 2008) that the Mississippi River may not be a sufficient barrier to gene flow above the Mississippi Embayment, making colonization across the river after glacial retreat possible. Finally, the current interface between the ranges of the Eastern and Central lineages may be a result of secondary contact following expansion out of stable areas unaffected by glacial advances, thus obscuring the location of the barrier that caused the formation of these lineages.

The cottonmouths (*A. piscivorus*) originated in the Late Miocene/Early Pliocene and the initial division between the Florida and the Continental lineages took place in the mid-Pliocene (~2.5 mya; Fig. 5; Table 2). Rising sea levels separated Peninsular Florida into isolated islands from the Late Miocene to early Pleistocene (Webb 1990; Clark et al. 1999). Inundation of seawater could effectively restrict gene flow between Florida and mainland populations, even in semi-aquatic freshwater species. Several studies have identified genetic breaks between Florida and mainland populations (Avise et al. 1984; Ellsworth et al. 1994; Walker and Avise 1998; Burbrink et al. 2008; Fontanella et al. 2008). Given the placement of the genetic break between the two cottonmouth lineages, it seems likely that gene flow was severed due to isolation of the Florida populations during rising sea levels in the Pliocene.

Results from mtDNA analyses of historical demography and phylogeography suggest that the two species have had unique responses to the same geological or climatic events. This may be explained by unique characteristics in the life history traits of each species. The single Continental clade of *A. piscivorus* overlaps with all three lineages of *A. contortrix* (Fig. 2). Apparently, the Continental lineage of *A. piscivorus* has

not responded to the same barriers to gene flow that promoted lineage diversification in *A. contortrix,* even though this lineage is older than all three of the lineages of *A. contortrix* (Fig. 5; Table 2). A possible reason for this disparity in lineage formation may have to do with the different preferred habitats of these species. The cottonmouths are semi-aquatic, rarely being found more than 1.6 km from the nearest body of water (Gloyd and Conant 1990). Even in the prairie steppe of the western portion of their range they are found in aquatic habitats (Gloyd and Conant 1990). In contrast, the copperheads are most often associated with terrestrial habitats. Changes in habitat, such as vegetational differences, elevation, and aridification, may have resulted in the formation of distinct lineages of *A. contortrix* due to adaptations to distinct ecological factors. Alternatively, geological barriers (e.g. river systems) may have severed gene flow among populations in this primarily terrestrial species. *Agkistrodon piscivorus,* however, tends to live only in aquatic/riparian habitats throughout its continental range, regardless of changing terrestrial habitats. The possibility remains that gene flow has been maintained throughout much of the range of *A. piscivorus* as a function of being restricted to aquatic habitats. Therefore, preference for aquatic habitats may have actually prevented lineage formation throughout the range of this species.

Population Demographics

Estimates of demographic histories did not agree with my initial prediction of expanding populations in *A. contortrix* and stable populations in *A. piscivorus,* based on the current absence of cottonmouths near formerly glaciated areas and the occurrence

of *A. contortrix* close to or in previously glaciated areas. All of the demographic methods suggested population expansion following the last glacial maximum in all lineages of both species with the exception of the Florida lineage of *A. piscivorus* and the Western lineage of *A. contortrix*. Mismatch distributions rejected population expansion only in the Florida and Western lineages. Similarly, negative values, of Fu's *F* suggests expansion in these same lineages,

The use of BSP's permitted me to examine the magnitude of demographic expansion or contraction while estimating the time for this change. Population expansion in the Eastern and Central lineages of *A. contortrix* and the Continental lineage of *A. piscivorus* appears to have co-occured with the retreat of Laurentide Ice Sheet at the end of the Wisconsinan Glacial period (Fig. 6; Killey 1998). Population expansion out of southern refugia following the retreat of the Laurentide Ice Sheet has been recorded in other taxa, such as the ringneck snake, *Diadophis punctatus* (Fontanella et al. 2008), the red-backed vole, *Clethrionomys gapperi* (Runck and Cook 2005) and the seal salamander, *Desmognathus monticola* (Crespi et al. 2003). The magnitude of N_e change varied widely among clades, with the Continental lineage of *A. piscivorus* showing the greatest population growth over the last 20,000 years (Fig. 6). This correlates with the magnitude of increased available habitat following glacial retreat for this lineage. In contrast, the other lineage of *A. piscivorus* has remained stable over the last one million years, despite changes in sea level affecting the size of the Florida Peninsula. Among the lineages of *A. contortrix*, the magnitude of expansion was greatest in the Eastern lineage and smallest in the Western lineage (Fig. 6). The

Eastern lineage has the greatest proportion of its current distribution in or near formerly glaciated areas. This may have allowed the copperhead to maintain populations closer to the glacial front during glacial maxima allowing them to sustain higher population sizes due to more available habitat. No portion of the current range of the Western lineage was glaciated, which also reveals only weak expansion over the last 125,000 years (Fig. 6).

Conclusions

The two closely related species of *Agkistrodon*, the copperheads and the cottonmouths, occupy similar ranges that cross multiple barriers, which have been identified as genetic breaks in a variety of taxa in the Eastern US (Soltis et al. 2006). My data suggest that unique habitat preferences may have shaped both the phylogeographic and demographic histories of each species. This may have resulted in increased lineage diversification within *A. contortrix* in a shorter time period than in *A. piscivorus*. I identified three phylogeographic lineages of *A. contortrix*, and two in *A. piscivorus.* In addition, no common geographic pattern of lineage diversification could be identified for these species. All three genetic breaks identified within *A. contortrix* exist within the range of a single lineage of *A. piscivorus*. The timing of diversification events was also discordant between the two species. Lineage diversification occurred in the Pliocene for *A. piscivorus* and in the Pleistocene for *A. contortrix*. Demographic histories of both species seem to have been shaped by Pleistocene glacial cycles. Population expansion was seen in all lineages, except for the Florida lineage of *A.*

piscivorus, and possibly the Western lineage of *A. contortrix* following the last glacial maximum. The magnitude of expansion among copperhead populations seems to be associated with proximity to formerly glaciated areas. Population expansion in the Continental lineage of *A. piscivorus* was most likely in response to an increase in favorable habitat following the retreat of the Laurentide Ice Sheet. In the following chapters I discuss the implications of using multi-locus coalescent methods and how they relate to the interpretation of the biogeographic histories discussed in this chapter.

Chapter II: Species Delimitation

Species delimitation in the presence of gene flow

Introduction

Properly enumerating species diversity impacts our ability to understand processes associated with speciation, conservation, diversification dynamics, and comparative biology. Central to achieving this goal is the proper identification of species. However, properly identifying species boundaries has never been straight forward and is likely one of the most controversial endeavors in systematics (de Queiroz 2007). Disagreements over how species are defined as well as what types of characters should be analyzed have both contributed to the difficulty in revising species level taxonomies (Eernisse and Kluge 1993). Molecular data has made it easier to examine the phylogeographic history of populations as well as identify cryptic species (Bickford et al. 2007). In turn, recent analytical advances have improved our ability to examine the complex processes involved with speciation. Perhaps most notable are methods that use multi-locus data to account for discordance between gene trees that may result from deep coalescence and incomplete lineage sorting (Carstens and Knowles 2007; Liu and Pearl 2007; Heled and Drummond 2008). However, these methods have also introduced a few obstacles to phylogeography and ultimately species discovery in certain cases. For instance, alternative sources of gene tree discordance (e.g. gene flow) are not readily considered (but see Yu et al. 2011). In addition, most methods assume that species assignment is known. This is not the case in studies that wish to incorporate a taxonomic evaluation at the species level.

Most phylogeographic studies have relied on mtDNA (Birmingham and Moritz 1998; Zink and Barrowclough 2008). However, an increasing availability of genomic data

in non-model organisms has made it possible to use multiple unlinked loci for both phylogenetic reconstruction and phylogeography. Paralleling the increase in data, more complex models have been developed that assess species relationships allowing for different coalescent histories in each unlinked loci (Brumfield et al. 2008; Leache and Rannala 2011). One commonality of these methods is that they require all individuals to be assigned to a species *a priori*. This is problematic when the phylogeographic structure of some or all of the taxonomic groups being studied is not known (O'Meara 2010; Leache and Rannala 2011). Individual species may consist of multiple independently evolving lineages. Therefore, methods of species delimitation are often required prior to assessing phylogenetic relationships, ideally with methods that can use multiple loci without concatenation.

Assessing the phylogeographic structure of a species typically occurs in several stages. First, individuals are sampled throughout the range of the species. Second, informative loci are identified and sequenced for all of the samples. Finally, genetic structuring is assessed by constructing phylogenetic trees. With single gene data sets, like those employing only mtDNA, phylogeographic lineages are typically identified by the presence of geographically distinct, reciprocally monophyletic groups in a single gene tree. This does not require *a priori* hypotheses regarding phylogeographic structure and therefore all gene trees are considered. Assessing individual gene trees for multiple loci increases the computational requirement considerably, particularly when simultaneously inferring a species tree from the posterior distribution of gene trees by incorporating the multispecies coalescent, as is the case with two of the more widely

used species tree methods; BEST (Liu and Pearl 2007) and *BEAST (Heled and Drummond 2010). Both methods have reduced the computational burden by requiring that all samples be assigned to define species *a priori*, reducing the number of trees to be considered. Unfortunately, for many taxa, delimitation prior to species tree inference is not straightforward when a robust species level taxonomy is not available

Recently, several species delimitation methods have been proposed that allow individuals to be assigned to species groupings (Carstens and Dewey 2010; O'Meara 2010; Yang and Rannala 2010). One assumption that all of the currently proposed species delimitation methods have in common is that there has been no gene flow since speciation. Recent studies have suggested that gene flow between closely related species may be common as a result of two processes (Nosil 2008). Sympatric or parapatric populations may continue to experience limited gene flow following speciation (Niemiller et al. 2008). Alternatively, allopatric populations may exchange genes after the two species come into secondary contact (Pedall et al. 2011). It is difficult to identify whether there has been gene flow between lineages since this requires identifying the lineages first. This creates a circular problem; species delimitation requires verification that no gene flow is occurring, which in turn requires that separate species or lineages be delimited. This brings up the question, how do phylogeographic studies incorporate current methods to assess species delimitation while accounting for multiple sources of gene discordance, such as deep coalescence and gene flow? I attempt to answer this question by determining how violating the assumption of no gene flow disrupts coalescent species delimitation, here using the

Bayesian Phylogenetics and Phylogeography method (BPP v.2.0; Yang and Rannala 2010). This method of Bayesian species delimitation estimates the probability of different species delimitations while accounting for uncertainty in the coalescent for each gene tree (Yang and Rannala 2010). I propose a modified version of the three step method proposed by Leaché and Fujita (2010) in which inference of population structure and assignment of individuals to populations is expanded to identify putative hybrids. This is followed by inference of phylogenetic relationships among populations. Finally, Bayesian species delimitation in BPP is carried out testing the effect of including putative hybrids as well as population assignment of putative hybrids on probability support for species delimitation.

In comparison with results presented in Chapter 1, I also examine how mitochondrial genes may influence support for the probability of species delimitation when combined with multiple unlinked loci. Mitochondrial genes have been the marker of choice in phylogeographic studies over the last ten years for several reasons, including readily available primers, fast mutation rates and primarily maternal inheritance which reduces N_e size. Compared to nuclear genes, the mutation rate of mitochondrial genes can be several orders of magnitude higher (Brown et al. 1979). This implies that a great number of nuclear genes are required to obtain the same number of polymorphic sites present in many mitochondrial data sets. Generally in phylogenetics more sites yield the potential for resolving trees if segregating sites support similar relationships. It seems reasonable that the phylogenetic signal may be overwhelmed by mitochondrial genes when data is concatenated (Naylor and Brown 1998). However, the

influence of mitochondrial genes relative to other loci combined in coalescent analyses that delimit species is not known. I explore the implications of using datasets that include and exclude mtDNA on species delimitation

Using two multi-loci datasets from two North American species of venomous snakes in the genus *Agkistrodon*, I address several questions related to species delimitation. The North American copperhead, *A. contortrix*, and the cottonmouth, *A. piscivorus*, are two closely related species with largely overlapping ranges in the Eastern and Central United States (Fig. 2). Previously, I examined the phylogeographic history of both species using mtDNA (Cytb; Chapter1; Guiher and Burbrink 2008). I identified two distinct lineages within *A. piscivorus* and three distinct lineages within *A. contortrix*. No obvious vicariant barriers were present between any lineages, despite the presence of the Appalachian Mountains and the Mississippi River within the range of both species. This suggests that gene flow may be ongoing between adjacent lineages, obscuring our ability to delimit species. Recent diversification among these lineages makes it more likely that there will be discordance between gene trees due to less time for genes to sort (Degnan and Rosenberg 2009). According to mtDNA analysis lineage diversification occurred in the Late Pliocene for *A. piscivorus* (2.5 mya) and began in the Early Pleistocene for *A. contortrix* (1.5 mya). The likelihood of gene flow and recent divergence make these two datasets ideal for addressing two primary concerns using Bayesian Species Delimitation with direct implications for species identification, species delimitation with gene flow and the influence of mtDNA. I outline methods for i) identifying putative hybrid samples and explore the performance of Bayesian Species

Delimitation while violating the assumption of no gene flow and ii) assessing the degree to which this method is influenced by mitochondrial genes by estimating species delimitation both with and without mtDNA included.

Methods

Sampling and sequencing

I obtained 114 individuals of *A. contortrix* and 131 individuals of *A. piscivorus* broadly sampled from throughout their respective ranges. Whole genomic DNA was extracted from ethanol preserved liver, heart, muscle or scale clippings using the DNeasy Kit (Qiagen Inc.). Template material for the polymerase chain reaction (PCR) consisted of samples with DNA/RNA ratios of 1.5–2.1 and DNA concentrations from 10–200 ng/ll. Five nuclear genes were sequenced for each species specifically for this project and used in conjunction with the mitochondrial dataset (Cytb) presented in Guiher and Burbrink (2008). In addition, Cytochrome b was sequenced for any individuals not represented in the previous dataset.

The sequences of *A. contortrix* consisted of three anonymous loci (Anon A, Anon 11, Anon 51) identified for a closely related Crotaline genus, *Sistrurus* (Gibbs and Diaz 2010) and two previously identified single copy (scnDNA) nuclear loci (NT3 and SPTBN1). The sequences of *A. piscivorus* were also comprised of five nuclear genes but used a different choice of markers. A single anonymous locus was used (Anon 11, Gibbs and Diaz 2010) and 4 scDNA loci (NT3, Noonan and Chippindale 2006); (SPTBN1, Mathee et al. 2001); (Vimentin Intron 5, Zehner and Paterson 1983); and (AHR, Wiens 2008). The

following primers were used for amplification: Anon A: AnonAF 5’-AGA ATT GAG CTC CCG TCC TTT-3’, AnonAR 5’-GGG AGC AAT GCC TAG ACC AAG-3’; Anon 11: Anon11F 5’-TCC TTA CTG AGT GAG CAC C-3’, Anon11R 5'-GCA AAG TCA ATG GAG AAA G-3'; Anon 51: Anon51F 5’-ACT TGC CTT CAG AAA TCA TG-3’, Anon51R 5'-ATC AAA GGT TTA AAG AA-3'; ; AHRR 5'-GYR AAC ATS CCA TTR ACT TGC AT -3'; NT3: NT3R 5'-GCG TTT CAT AAA AAT ATT GTT TGA CCG G-3'; NT3F 5-ATA TTT CTG GCT TTT CTC TGT GGC-3'; SPTBN1: SPTBN1SeqF 5’-ATA CAG GCT GAG CGA GTG AGA-3’, SPTBN1SeqR 5’-AGC TGA CAT AGC TCT TGG TAA CA-3’; Vimentin Intron 5: VimExon5F 5’-AAC AAT GAT GCC CTG CGC CA-3’, VimExon6R 5’-CAA TAT CAA GAG CCA TCT TTA CAT T-3’AHR: AHRF 5'-GTC CAC CTG CTT CAA ATA A -3'. Amplification of each gene was carried out using GoTaq Green Master Mix (Promega Corp). A 90s extension time was used for each gene along with the following annealing temperatures; Anon A (51), Anon 11 (44), Anon 51 (43), NT3 (51), SPTBN1 (51), Vimentin Intron 5 (47), and AHR (45). Successful PCR products were purified using 1 µl of ExoSAPIT (USB Corp.) per 10 µl of PCR product. The sequencing reaction for all genes consisted of 2 µl of DTCS (Beckman-Coulter), 1 µl of 2 µm primer and 2 µl of DNA template and 4 µl of water. Sequences were purified using ethanol precipitation and analyzed on a Beckman-Coulter CEQ8000 sequencer. Sequencing primers were as follows: Anon A: AnonAR 5’-GGG AGC AAT GCC TAG ACC AAG-3’; Anon 11: Anon11F 5’-TCC TTA CTG AGT GAG CAC C-3’; Anon 51: Anon51F 5’-ACT TGC CTT CAG AAA TCA TG-3’; NT3: NT3R 5’-GCG TTT CAT AAA AAT ATT GTT TGA CCG G-3’; SPTBN1: SPTBN1SeqF 5’-ATA CAG GCT GAG CGA GTG AGA-3’, SPTBN1SeqR 5’-AGC TGA CAT AGC TCT TGG TAA CA-3’; Vimentin Intron 5: VimExon5F 5’-AAC AAT GAT GCC CTG CGC CA -3’,

VimExon6R 5’-CAA TAT CAA GAG CCA TCT TTA CAT T-3’; AHR; WL325AHR_F5 5’-GTC CAC CTG CTT CAA ATA A-3’. Sequences were edited and aligned using MUSCLE (v. 3.6; Edgar 2004) in the program Geneious v5.1 (Biomatters Ltd.). Heterozygous genotypes were resolved using PHASE v2.1.1 (Stephens and Donnelly 2003) and used for all downstream analyses.

Population structure and Phylogenetic Inference

The species delimitation method implemented in BPP requires the user to specify a guide tree, which represents the most resolved tree and defines a subset of trees by collapsing nodes within the guide tree. To reduce the computational burden species probabilities are assessed over only the subset defined by the guide tree. To construct the guide tree I followed the approach outlined by Leaché and Fujita (2010) in which the number of populations and assignment of individuals to populations is first assessed by inferring population structure followed by inferring phylogenetic relationships among populations. In order to infer the number of populations and assign individuals to these populations I used the program STRUCTURE v2.3.3 (Pritchard et al. 2000; Falush et al. 2003, 2007; Hubisz et al. 2009) with the phased alleles for each species. To establish the number of populations within each species, 5 independent runs were performed varying the number of *K* populations for each run. Posterior Bayes Factors (PBF) was used to identify the optimal value for *K*. Two subsequent runs were performed for 1 million generations with a burn-in period of 500 thousand generations. In addition to determining the number of populations and assigning individuals to

populations, STRUCTURE has the ability to identify individuals with mixed ancestry and reports the proportion of alleles inherited from each population for admixed individuals. I conservatively identified individuals for which less than 85% of their alleles could be assigned to a single population as “hybrids” for downstream analyses. Determining a threshold for defining “hybrids” was somewhat arbitrary since STRUCTURE includes error in the estimate of allele contribution, therefore no individuals are assigned 100% to one population. Preliminary analyses determined that a higher threshold negatively impacted speciation probabilities when single individuals were assigned to the alternate population.

To examine the phylogenetic relationships of the populations inferred by STRUCTURE I constructed species trees for each species individually using the program *BEAST v. 1.6.0 (Heled and Drummond 2010). The most appropriate model of evolution for each gene was determined using AIC and BIC criterion in Jmodeltest (Posada 2008). Phased nuclear alleles for five individuals from each population were used in place of the entire dataset to avoid excessive runtimes. I combined results from two identical analyses run for 50 million generations, sampling every 1000 generations and discarding the initial 20% as burn-in. Convergence was assessed by examining likelihood plots through time in TRACER v. 1.5.0 (Rambaut and Drummond 2007). Bayesian posterior probabilities (Pp) greater than 95% were considered strong support for a clade (Felsenstein 2004). I also produced a genetic network using SplitsTree 4.1.0 (Huson and Bryant 2006) with the NeighborNet algorithm (Bryant and Moulton 2004). The resulting guide tree and population memberships were subsequently used to parameterize

analyses of species delimitation in BPP. I also produced a genetic network using SplitsTree 4.1.0 (Huson and Bryant 2006) with the NeighborNet algorithm (Bryant and Moulton 2004).

Species Delimitation

I use BPP to estimate probability of species delimitation and explored the effect of including individuals with mixed ancestry by comparing the speciation probabilities between runs comprised of an increasing number of admixed individuals. Each run consisted of 40 individuals with admixed individuals representing 0, 10, 20, 30, 40, 50, 60, 70, 80, and 90% of each dataset. In order to account for the uncertainty involved with assigning individuals of mixed ancestry to a population duplicate runs were performed for each dataset, first assigning admixed individuals to the population from which they inherited the greater proportion of alleles and second to the population which represented the lesser proportion of alleles. In order to ensure that the rjMCMC was mixing properly I took two precautions. First, preliminary runs were performed implementing both algorithms with values for the parameter ε =2, 5, 10, 15, and 20. Alogorithm 0 with ε=2 was determined to be the optimal parameterization and used for all subsequent runs. Additionally, all runs were performed with both opposing starting trees, where the starting tree could be either fully collapsed (0) or fully resolved (1), to ensure that the same solution was achieved from the opposite starting point. Each analysis was run for 100,000 generations discarding the first 20,000 as burn-in. The ancestral population size (Θ) and root age (τ_o) both require that a prior gamma Γ (α, β)

distribution be specified. Appropriate priors were determined by performing initial runs with a fixed species tree and diffuse priors. Subsequently, the mean ($\overline{X}$) and standard deviation (σ) for each parameter was calculated using values for all six loci. I then determined values for α using the equation $\alpha=(m/s)^2$ and for β using the equation $\beta=m/s^2$. The following prior distributions were used for each species: *A. contortrix*: Θ (4.5, 1000), τ_o (1, 250); *A. piscivorus*: Θ (2, 100), τ_o (1.6, 360). All fine tuning parameters were optimized to ensure that swapping rates were between 0.30 and 0.70.

In addition, I determined whether speciation probabilities are overly influenced by the signal from individual genes (i.e., mtDNA). Including all of the individuals in the analyses proved impossible due to a limit of 200 unique alleles in the current version of BPP. Therefore, I randomized each dataset to produce 10 datasets for each species and summed the speciation probabilities over all 10 runs. Individuals with admixture were removed and each dataset consisted of 48 (16 per population) and 40 (20 per population) individuals for *A. contortrix* and *A. piscivorus* respectively, ensuring that all individuals were included. Duplicate runs were performed both with all six loci and with the mtDNA loci (Cytb) removed. The same conditions and priors specified above were used for all runs.

Results

Sequencing

A total of 3,269 bp obtained from six loci were sequenced for *A. contortrix* with the following lengths; Anon 11 (340 bp), Anon 51 (322 bp), Anon A (249 bp), Cytb (1107

bp), NT3 (480 bp), and Sptbn1 (771 bp). There were no gaps in any of the six loci. The highest number of variable sites were contained by the single mtDNA gene Cytb (71), the five nuclear loci combined contained 38 variable sites ranging from 3 -13 per locus. A total of 3,666 bp was obtained from the six loci sequenced for *A. piscivorus* with the following lengths; Anon (236 bp), Cytb (1107 bp), Nt3 (455 bp), Sptbn1 (938 bp), Vimentin Intron 5 (539 bp) and AHR (391 bp). There were no gaps identified in any of the six loci with the highest number of variable sites contained by the mtDNA gene Cytb (76 bp), the five nuclear loci contained 24 variable sites combined ranging from 1-11 per locus (Table 4).

Population Structure and Phylogenetic Inference

Two populations were identified by STRUCTURE for *A. contortrix* and for *A. piscivorus* supported by PBF and a plateau of marginal likelihood at *K*=2. The populations identified by STRUCTURE using all six loci correspond to the mtDNA lineages reported in Guiher and Burbrink (2008; chapter 1) with the exception of the Eastern and Central lineages of *A. contortrix* being combined into a single population. There were 17 individuals identified with mixed ancestry between the two populations of *A. contortrix* representing 15% of the dataset and 8 individuals between the two populations of *A. piscivorus* representing 6% of the dataset. Admixed individuals for both species were isolated to the hypothesized contact zone, defined as being the interface between sister lineages (Fig. 7).

Inferring phylogenetic relationships between populations proved to be unnecessary since analyses of population structure identified only two populations for both species. Therefore, the guide tree for both instances was a single bifurcating tree representing a single speciation event and a subtree resulting from collapsing this node. However, phylogenetic inference was necessary to examine the influence of mtDNA on species delimitation in *A. contortrix* since a previous mtDNA study identified three lineages (Chapter 1; Guiher and Burbrink 2008). In this case a guide tree could have been constructed based on the phylogeographic relationships suggested by mtDNA but I chose to corroborate those relationships using multiple nuclear genes in combination with mtDNA to construct a species tree. This was critical to determine how the species delimitation method would resolve the incongruence between the mtDNA and scnDNA loci suggested by the results from STRUCTURE. Using BIC and AIC in Jmodeltest I determined that the most appropriate substitution models were GTR+ Γ +I partitioned by three codon positions, JC+I, HKY, K80+I, K80+G, and GTR+ Γ+I for Cytb, Anon 11, Anon 51, Anon A, NT3, and Sptbn1 respectively. Discarding the first 20% (10 million generations) as burnin resulted in ESS values above 200 for all parameters. The resulting species tree recovered three lineages previously suggested by mtDNA (Chapter 1; Guiher and Burbrink 2008) with strong support, an Eastern lineage, a Central lineage and a Western lineage (Fig. 7). Therefore, the guide tree ((East, Central), West)) was used for *A. contortrix* when examining differences in speciation probabilities between incorporating all six genes versus analyses that excluded mtDNA. Networks of each loci failed to recover distinct clades for any of the loci used for both species (fig 8 and 9).

This may indicate that nuclear loci have not sorted, supporting the need for coalescent methods capable of accounting for ILS.

Species Delimitation with Gene Flow

Including individuals with mixed ancestry affected Bayesian species delimitation methods in three ways and results were largely consistent across both species. First, speciation probabilities decreased as the proportion of admixed individuals in the dataset increased (Table 5). Additionally, speciation probabilities were impacted by which population admixed individuals were assigned. A significant reduction in the probability of delimiting species was observed when admixed individuals were included in the population that contributed the minority of alleles (Table 5). Finally, as the proportion of admixture approached 20% ensuring proper mixing of the rjMCMC proved impossible. This was evident in that runs beginning with different starting trees failed to reach the same conclusion. That is, if a fully resolved (1) starting tree was used then the two species were recovered with a high speciation probability. Conversely, when the starting tree was collapsed (0) there was no support for the existence of two species.

Influence of mtDNA on Bayesian species delimitation

The influence of mtDNA on speciation probabilities differed between the two datasets. Two populations were inferred with 100% support for *A. piscivorus* in all analyses regardless of whether mtDNA was removed. In contrast, assessing species delimitation within *A. contortrix* was significantly influenced by mtDNA. A West lineage

was recovered with 100% support in all analyses. Analyses of the full 6 gene dataset produced moderate support for the existence of a Central and East lineage with a mean speciation probability of 0.948225 (range of 0.81437 and 1.0). A noticeable decrease in support for this speciation event was observed when Cytb was removed from the dataset with a mean speciation probability of 0.458467 (range of 0.11513 and 0.73659).

Discussion

Identifying cryptic species requires identifying genetic patterns of diversification that may not be evident from more traditional characters, such as morphology, behavior or color pattern. Specifically, this requires methods that can delimit species and infer phylogenetic relationships while accounting for multiple sources of gene discordance. Coalescent methods have been developed to delimit species and infer phylogeographic relationships between species (Yang and Rannala 2010). However, these methods are relatively new and it is unclear how they will perform in the face of sources of gene discordance other than incomplete lineage sorting, specifically hybridization. The results presented here suggest that species can be successfully delimited when moderate amounts of gene flow are present but that speciation probabilities are negatively affected by increasing amounts of gene flow.

Investigating population structure prior to species delimitation not only made it possible to assign individuals to respective populations but had the added benefit of identifying putative hybrids. Bayesian species delimitation implemented in BPP assumes that gene flow is absent between species (Yang and Rannala 2010). Therefore, *a priori*

tests of gene flow are a crucial first step. Exploring the effects of violating this assumption reveals that species delimitation in BPP is possible with moderate amounts of gene flow but that several factors impact the outcome on detecting species probabilities. Speciation probabilities decrease when an increasing amount of hybrids are included (Table 5; Fig. 10), although these general results vary given specific model parameterizations. Importantly, results were fairly consistent between the two datasets analyzed here. First, hybrids by definition inherit alleles from both populations and can therefore be assigned to either population. The choice of which population hybrids are assigned to impacted the rate at which the posterior probability of speciation decreased with respect to increased gene flow (Table 5; Fig. 10). For instance, when hybrids were assigned to the population contributing the majority of alleles, speciation probabilities greater than 95% were observed with as much as 30% of the data consisting of putative hybrids for *A. contortrix* and 20% of *A. piscivorus*. In contrast, speciation probabilities decreased more rapidly and with fewer hybrids present when hybrids were assigned to the population which contributed the minority of alleles, 10% and 20% for *A. piscivorus* and *A. contortrix* respectively (Table 5; Fig. 10). Second, the ability to assess mixing was largely dependent on the starting tree used. In both cases only two starting trees were possible, a fully resolved tree (1) or a fully collapsed tree (0). Posterior probabilities of species delimitation were 1.0 for all analyses using the starting tree (1) regardless of the percentage of hybrids included or which population hybrids were assigned to (Table 5; Fig. 10). In contrast, the trend of decreasing speciation probabilities with an increasing percentage of hybrids discussed above was observed when the starting tree was fully

collapsed (Table 5; Fig. 10). This discrepancy is somewhat surprising and requires exploration with a greater number of possible starting trees than examination of two species can provide. However, to avoid inferring potentially wrong or inflated speciation probabilities I suggest performing replicate runs using all possible starting trees to ensure consistency.

It is clear that single gene trees may be poor representations of species histories for many reasons including incomplete lineage sorting, gene flow and recombination (Degnan and Rosenberg 2009; Edwards 2009). However, the field of phylogeography has been dominated by mtDNA studies for the last decade. Determining just how misleading this reliance on a single marker has been for phylogeographic studies requires a much more comprehensive examination than what can be provided here. However, insights can be gained by exploring this question with two empirical datasets presented in this study. Species delimitation within *A. piscivorus* did not appear to be dependent on the presence of mtDNA, recovering 100% support for two species in all 20 replicates (Table 5; Fig. 10). In contrast, BPP suggests that either two or three species of *A. contortrix* depending on whether the mtDNA gene Cytb was included. Both the full dataset and the nuclear only analyses recovered 100% posterior probability support for two species, congruent with the populations recovered by STRUCTURE (Table 5). In addition, both the Central and Eastern lineages were moderately supported with a mean posterior probability of 0.948225 summed over 10 runs using all 6 genes (Table 5). When Cytb was excluded, the speciation probability for this node was reduced (0.458467; Table 5). This suggests that at least in this case, lineage diversity may be overestimated by

mtDNA. Alternatively, mtDNA may represent the phylogeographic history of *A. contortrix* and there has not been ample time for the slower nuclear genes to accumulate substitutions. It is important to note that divergence time estimates from mtDNA (Chapter 1; Guiher and Burbrink 2008) suggest that the two lineages of *A. piscivorus* diverged approximately 2.5mya compared to just 1.38 mya for the East and Central lineages of *A. contortrix*. Certainly, there is a higher likelihood for gene discordance in younger lineages. It is also possible that the discordance observed here is a result of relatively few nuclear genes included. Additional nuclear genes may either corroborate mtDNA or alternatively reduce the influence of the mtDNA signal in subsequent coalescent species delimitation analyses.

A conservative approach to interpreting species delimitation in this case where the single mtDNA marker disagrees with multiple nuclear loci is to only recognize species supported by multiple markers. Ideally, ecological, behavioral, physiological or morphological evidence can be included to delimit species. However, differences among lineages regarding these types of data may not be possible for cryptic species or in species that have diverged rather recently.

Combining a multilocus approach to identify putative hybrids with Bayesian species delimitation identifies two species of both *A. contortrix* and *A. piscivorus*. Admixture was identified along the contact zone between adjacent species in both species complexes. Discriminating between recent gene flow after species come into secondary contact and ongoing gene flow during speciation may require many thousands of loci (Nosil 2008), therefore it is not possible to determine if gene flow is

recent or has been ongoing since divergence in either species complex. I have shown that Bayesian species delimitation is robust to gene flow when as much as 20% of the dataset is represented by admixed individuals. However, increasing the proportion of admixed individuals beyond this level results in a steep decline in the probability of supporting a speciation event (Fig. 10). Furthermore, I demonstrate that phylogeographic structure unique to a single fast evolving locus (such as mtDNA) can yield strong support for divergence. This may lead to discordance in the number of species supported by different loci, particularly in recently diverged lineages. Therefore, it may be necessary to examine gene discordance to determine how well recent speciation events are supported by the data. Unexpectedly, the starting tree used proved to have a significant impact on model performance. Beginning with a fully resolved starting tree made it impossible to achieve proper mixing resulting in a failure to explore alternative tree space. Therefore, it is crucial to perform replicate runs with fully collapsed or partially resolved starting trees to avoid inflated speciation probabilities.

Bayesian species delimitation has provided evidence that both *A. contortrix* and *A. piscivorus* are both comprised of multiple species. In the next chapter I present species descriptions and examine ecological and morphological data to determine if additional diagnosable characters can be provided. Moreover, I explore whether addition of ecological and morphological data can provide support for the recognition of either two or three species of *A. contortrix*.

Chapter III: Taxonomy

A taxonomic revision of *A. contortrix* and *A. piscivorus* based on ecology, morphology and genetic data.

Introduction

The importance of discovering new species and accurately tallying the number of taxa in a region is particularly critical during the earth's sixth mass extinction. The current rate of extinction is estimated to be 1,000 to 10,000 times higher than background rates under species turn over (Chivian and Bernstein 2008; Barnosky et al. 2011). If the current trend continues as much as half of the worlds' species are predicted to go extinct within the next 100 years (Thomas et al. 2004). Conservation efforts depend on the ability of taxonomists to discover and describe species before they go extinct. This includes examining the phylogeographic history of currently known taxa to assess the probability that cryptic species (i.e., those not recognized by traditional morphological characters) are present. Moreover, underestimating the total number of species in an area has negative implications on our ability to properly use taxa in comparative biology in order to study speciation processes, community assemblage rules, and changes in ecology, genetics, and morphology over time (Barraclough and Nee 2001).

Important to all methods that rely on correct taxonomic identification is the use of a proper species concept. Over the last 20 years, the general species concept is likely favored among evolutionary biologists that integrate disparate sources of information because it outlines explicit criterion including but not limited to ecology, physiology, morphology and phylogenetic history that can be used to support species recognition (de Queiroz 2007). It has been proposed that species may acquire attributes that permit the use of specific criterion along a gradient of diversification following speciation

resulting in a “gray zone” (de Queiroz 2007). On one end of this “gray zone” populations have not diversified along any of the species criterion and there is uniform agreement that a single species is represented. On the other end, populations have acquired divergent characteristics along all species criterion traditionally used for species delimitation and supporting speciation is nontrivial. However, this ideal scenario is rarely achieved, particularly with cryptic species or when divergence is recent. This has resulted in a debate over which and how many criterion are sufficient for species delimitation (de Queiroz 2007; Bauer et al. 2011; Fujita and Leaché 2011). Fujita and Leaché (2011) have argued that the method of Bayesian species delimitation implemented in Chapter 2 provides an objective method of species delimitation. Therefore, based on the evidence presented in Chapter 2, I recognize 2 species of *A. piscivorus* and either 2 or 3 species of *A. contortrix*.

Genetic analyses presented in Chapters 1 and 2 suggest that 2 or 3 unrecognized species of *Agkistrodon* occur in the US. The phylogeographic analysis of mtDNA presented in chapter 1 and Bayesian species delimitation with 6 unlinked loci presented in chapter 2 suggest 2 species of *A. piscivorus*; a Florida species with a distribution including Florida and parts of the Gulf coastal plain in southern Georgia, Alabama and eastern Mississippi, and a Northern species which inhabits the Atlantic coastal plain from southeastern Virginia to Florida, North to southern Illinois east of the Appalachians west into central Oklahoma and Texas (Figs. 2b and 7b). Results presented in the previous two chapters suggest 3 species of *A. contortrix*; an Eastern species ranging from Connecticut south to Florida and west to the eastern portion of the Mobile Basin, a

Central species extending approximately from the east portion of the Mobile basin west to eastern Texas, Oklahoma and Kansas, and a Western species inhabiting the more xeric habitat of southwestern and central Texas, central Oklahoma and eastern Kansas. However, excluding the mtDNA locus (Cytb) from Bayesian species delimitation supports only two species of *A. contortrix*, collapsing the East and Central lineages into a single species. I attempt to resolve the uncertainty in the number of species of *A. contortrix* and corroborate results for *A. piscivorus* by examining two additional species criteria, niche similarity (Ecological species concept; de Queiroz 2007) and morphological diagnosability (Evolutionary species concept; de Queiroz 2007) for the five lineages of *A. contortrix* and *A. piscivorus*. Specifically, I investigate whether lineages have evolved distinct ecological requirements by presenting a comprehensive exploration of ecological niche modeling (ENM). I compare niche similarity between adjacent lineages using three metrics, the niche identity test and niche background test of Warren et al. (2008) to determine if adjacent lineages occupy identical or similar niches, and comparison of geographic overlap of predicted distributions to explore whether ENM's predict geographically distinct distributions. I also provide statistical analysis of morphological measurements to determine if the proposed species can be differentiated morphologically and attempt to identify characters useful for species diagnosis. Divergent ecological requirements or morphologies will lend additional justification for the recognition of separate East and Central species of *A. contortrix*, while the absence of such results supports a single species. I propose taxonomic revisions for both species based on analysis of the three criteria examined.

Methods

Niche modeling

The potential distributions for the East, Central and Western lineages of *A. contortrix* and the Florida and Continental lineages of *A. piscivorus* were estimated by generating ecological niche models using a maximum entropy method in the program Maxent 3.3.3e (Phillips et al. 2006; Phillips and Dudik 2008). The 19 BIOCLIM variables describing temperature and precipitation from the WorldClim data set (Hijmans et al. 2005) at 30-s spatial resolution were used to construct ENM's. Models were trained using georeferenced localities for all individuals used in the molecular analysis by lineage. Several separate analyses were performed for each lineage to explore the effect of two specific modeling decisions outlined in Elith et al. (2011), background selection and sampling bias in the training data. Background selection has been shown to impact the predicted distribution predicted by Maxent (VanDerWal et al. 2009). The modeling decision of what landscape to draw background points from is often a factor of the question being addressed (e.g. what are the environmental factors that determine the distribution of a species or how does a species utilize partitioned microclimates?), including the full range of environmental conditions available to a species (which may be limited by barriers to dispersal), and sampling effort (Elith et al. 2011). I explore two candidate backgrounds; (i) the entire USA which represents all environmental conditions available and seeks to addresses what variables are critical in defining the distribution of each lineage, and (ii) the eastern US which includes only the environmental conditions

inhabited by species of *Agkistrodon* and may better identify environmental conditions unique to each lineage. I also examined the effect of including hybrids identified in chapter 2 on the prediction of potential distributions for each lineage. A combination of hybrid assignment, bias correction and background selection resulted in a total of eight models per lineage. Since neither species included in this study is rare, sampling bias should be similar to collecting biases encountered in other snakes. The method of collection is consistent for most all snake species, by encounter on roads or by chance encounter in suitable habitat. Therefore, collection effort for target species may be adequately represented by collection localities for other species of snakes collected in the USA. To test this, I attempted to correct for sampling bias by downloading every colubroid snake record from the HerpNet database that included locality data with Latitude and Longitude within the United States. Environmental data for all nineteen BIOCLIM variables was extracted for all 17,555 samples in DIVA-GIS v7.3.0. (Hijmans et al. 2001). Models were constructed by sampling 10,000 random background points from these points. Ideally, background points could have been restricted to crotaline collection localities, however, the limited number of crotaline localities available would significantly limit the number of background points having a potentially negative impact on model performance. In addition, I constructed ENM's without correcting for bias by sampling background from the original 19 BIOCLIM layers over both the entire US and the reduced distribution outlined above. Two sets of analyses were run under the above parameterizations, one including hybrids in the training samples for each lineage and one excluding hybrids to examine whether hybrids were inhabiting environmental

conditions not suitable to one or both lineages. All analyses used auto features along with the default regularization multiplier (1.0). The number of iterations was increased to 5000 to allow the algorithm to run to the default convergence threshold (10^{-5}). Initially, I performed 10 replicate runs using a different random seed and crossvalidate which randomly divided the samples into replicate folds. This provided an estimate of the sensitivity of the predicted distribution to the samples used to train and test the model. I evaluated model performance using two criteria, the threshold-independent receiver operating characteristic curve (AUC) and the threshold-dependent binomial omission tests. Sufficient discrimination between 'presence' and 'absence' is indicated by AUC values greater than 0.7 (Swets 1988). I compared AUC values between the four analyses for each lineage which included different modeling decisions (i.e., combinations of sampling bias and background choice). Model fit was determined only between models that included identical candidate samples; models that included hybrids were not compared to models that excluded hybrids. Resulting distributions were projected in DIVA-GIS using the appropriate threshold for each model to the produce binary predictions of suitability. Mean values of thresholds over the 10 crossvalidated replicates were used.

I also used the niche identity test and background tests in ENMTools v1.3 (Warren et al. 2008, 2010) to determine if the lineages differed significantly with respect to niche as defined by the nineteen bioclim variables. The niche identity test determines if two lineages inhabit identical niches while the background test examined if niches were similar. Niche similarity between all adjacent lineages in each species was assessed

using two metrics, Schoener's *D* (Schoener 1968) and "Warren et al.'s" *I* (Warren et al. 2008). The niche identity test determines if two lineages inhabit identical niches by comparing observed values of *D* and *I* to a distribution of randomized pseudoreplicates generated by randomly assigning samples to either lineage. A one-tailed t-test was used to determine if observed values differed significantly from randomized distributions of *D* and *I*. The background test examined whether lineages occupy similar niches by comparing observed values of *D* and *I* to a randomized sample of background points drawn from the other lineage being compared equal to the number of points used to train that lineage. Statistical significance was determined using a two-tailed t-test.

Morphology

A total of 20 mensural characters (Appendix 1) were measured for 100 individuals. I examined ~20 individuals for each mtDNA lineage with the exception of the Central (21 individuals) and West (19 individuals) lineages. Samples for both species were obtained throughout their entire respective ranges including representatives of all eight currently recognized subspecies (Fig. 11). Bias in size of mensural characters can occur between juveniles and adults (Thorpe et al. 1983) confounding predictions of morphological dissimilarity among lineages. Gloyd and Conant (1990) considered individuals below 300mm juveniles. Therefore, I excluded any individuals with a snout-vent (SVL) less than 350mm. There was no steep slope observed when examining the HL/SVL ratio for individuals with a SVL greater than 350mm indicating that this was sufficient (Fig. 12). All characters were transformed logarithmically to ensure a linear

relationship between all characters (Hillis 1978; Thorpe et al. 1983; Sokal and Rohlf 1995).

All statistical analysis of morphological data was carried out in the program STATISTICA v10. I used principal components analysis (PCA) with varimax rotation and correlation matrix to determine if lineages could be separated morphologically without *a priori* assigning clade membership. Several tests were performed to determine if differences could be detected between *a priori* defined lineages in the morphological characters examined. First, I used a MANOVA test of all characters to identify differences in gross morphology. Since morphological characters can be correlated with size (Burbrink 2001), I also conducted a MANCOVA test identifying SVL as a covariate. This corrected for the effect of size on the remaining characters and allowed me to determine how effective the data is at discriminating between groups in the absence of size differences. Tukey's HSD post hoc test was performed to identify if means calculated in both the MANOVA and MANCOVA differed significantly among lineages. Finally, I performed canonical discriminate function analysis (DFA) to classify lineages and identify which characters best describe differences between lineages.

Results

Niche Modeling

Using Maxent, ecological niche models for both lineages produced test AUC values >0.7 for all replicates (Table 6). Higher AUC's indicate that correcting for sampling bias by using a large dataset of colubroid collection localities to match bias in the

background sampling improved model performance in every case except for the Florida lineage (Table 6). This is likely due to that the bias in colubroid sampling is towards the western US, far outside of the range of the Florida lineage (Fig. 13f). In addition, colubroid sampling within Florida is biased toward southern Florida (Fig. 13f). This is opposite of the sampling bias observed in the Florida lineage (Figs. 13c and 13d). Therefore the method applied here is unlikely to account for the bias present in the sampling of this lineage. Sampling background points from the entire US resulted in higher AUC's for every lineage (Table 6). This is in contrast to restricting background points to the eastern US, the approximate distribution of both species. Binomial omission tests were significant for all models for all threshholds ($P < 0.001$) with the exception of models for the West lineage of *A. contortrix*. For all other lineages I applied the minimum training presence (MTP) threshold to convert distributions to binary predictions of suitability since it had the added benefit of being biologically interpretable (Pearson et al. 2007). Binomial omission tests were not significant for several thresholds for ENM's of the West lineage and exhibited a much greater amount of variance in fractional area predicted compared to models for the other lineages. Only two threshold rules, equal test sensitivity and specificity (ESS) and maximum test sensitivity plus specificity (MSS), provided significant results for the West lineage when hybrid samples were excluded. Logistic thresholds and fractional predicted area were identical for both rules and were substituted for MTP in this model. Poor performance for this model is not surprising given the low sample size (9) available for this lineage resulting in the use of only linear features. Including hybrids for the West lineage

resulted in 25 individuals and improved model performance. However, binomial omission tests were not significant for several threshold rules and exhibited a much greater amount of variance in fractional area predicted compared to ENM's for the other lineages. Significant binary thresholds (P>0.05) tended to either over fit or provide little discriminatory power for this lineage and did not seem to provide good representation of the underlying model. The equate entropy of thresholded and original distributions rule provided the best balance between overfit predictions and distributions with little predictive power. The predicted distribution of the Florida lineage was also impacted by the inclusion of hybrid samples resulting in a potential hybrid zone extending across the mid-Atlantic coastal plains in southern North Carolina to the southern coastal plains in south-eastern Louisiana (Fig. 13c). Potential distributions for the Continental lineage of *A. piscivorus* and the combined East/Central lineage of *A. contortrix* were not impacted by including hybrid samples. In both cases potential hybrid zones were predicted well regardless of whether hybrids were used to train the model (Fig. 13). Comparisons could not be made for the East and Central lineages since genetic data did not identify putative hybrids. The variables with the greatest contribution to each model were dependent on parameterization of the model (Table 6).

Observed values of Schoener's *D* and Warren et al.'s *I* values were significantly lower than expected from a random distribution (Table 7). This indicates that the null hypothesis of identical niches can be rejected for all pairwise lineage comparisons considered. However, observed values were not significantly different from the

background of the adjacent lineage in all comparisons of the background test (Table 8). This indicates that sister lineages inhabit similar but not identical niches. As a second metric of niche similarity, I calculated predicted ranges and their area of overlap by multiplying the number of pixels representing suitable habitat by 0.86 (30 s of arc equals 0.93 km, thus a 30-s pixel equals 0.86 km^2). This differed from the metrics provided by ENMtools in two ways. First, estimates of range overlap compares distributions in geographic space while *D* and *I* reflect differences in niche space. Second, calculating the amount of geographic overlap is dependent on binomial thresholds to convert distributions to binary predictions of suitability. While, ENMtools does provide an option to apply binomial thresholds the authors caution against it. Predicted distributions of all adjacent lineages overlapped substantially with the exception of the Continental lineage when hybrids were not used to train the model (Fig. 13, Table 9). As described above, hybrids had virtually no effect on the predicted potential distribution of the Continental lineage. Therefore, differences calculated for the amount of overlap between the two lineages of *A. piscivorus* was a result of variation in the Florida models not the ENM's for the Continental lineage. Estimates of overlap between the West lineage and combined East/Central lineage were similarly influenced by variation in ENM's for the West lineage (Fig. 13a and 13b, Table 9). The Central and East lineages overlapped considerably with one another geographically (Fig. 13e, Table 9).

Morphology

There was no separation between lineages within a species as demonstrated by PCA, however there was clear separation between *A. contortrix* and *A. piscivorus* (Kruskal Wallace; P < 0.00001). The first and second PCA axis accounted for 85.22% and 3.68% of the variation respectively. All variables loaded heavily on PC1 indicative of the contribution of most character to the variation along the first axis (Table 10). Loadings on PC2 were considerably lower with the highest being eye diameter left (15) and eye diameter right (14; Table 10). A MANOVA using all 20 characters suggests that there are significant morphological differences among the five lineages (Wilk's $\lambda = 0.007$, $F_{4,94} = 9.29$, $P < 0.001$). However, the Tukey post hoc test reveals that only the West lineage can be distinguished from the East lineage of *A. contrortix* (P=.0485). Similarly, a MANCOVA specifying SVL as a covariate identified significant differences in morphology among all lineages (Wilk's $\lambda = 0.008$, $F_{4,93} = 9.4$, $P < 0.001$). Removing the effect of size on the remaining variables increases the number of lineages that can be differentiated morphologically. A Tukey post hoc test significantly separated *A. contortrix* from *A. piscivorus* and the East lineage from the Central and West lineages. Classification matrix of DFA shows that the West, Continental, and Florida lineage can be correctly classified greater than 90% of the time. Classification of the Central and East lineage was less successful (76% and 85% respectively). Two variables, SVL (42.1) and width across parietals (12.8) had the highest loadings using DFA (Table 11).

Discussion

Recent diversification and cryptic speciation provide unique challenges to species delimitation (Rissler and Apodaca 2007). Relatively young lineages may not have

had sufficient time to acquire unique morphological or ecological properties making corroboration from multiple lines of evidence difficult (de Queiroz 2007). This is confounded in cryptic species which retain morphologies similar to the ancestral condition as a result of niche conservatism (Bickford et al, 2007). Examination of genealogical (exclusive coalescence of alleles to a single species), ecological (unique niche), and evolutionary (diagnosability) species criterion for *A. contortrix* and *A. piscivorus* reveals that both species complexes are comprised of two cryptic species or that there has not been sufficient time since divergence for species to develop unique morphologies or entirely different ecological requirements.

Ecological niche models and statistical analysis of morphological characters provide additional support for taxonomic revision of *A. piscivorus* and *A. contortrix*. Two phylogeographic lineages of *A. piscivorus* identified by genetic coalescent analyses have non-identical but similar ecological niches, supported by both *D* and *I*. Estimates of geographic overlap in the potential distributions of the two lineages suggest that the two lineages inhabit geographically distinct niches. Estimates of overlap are influenced by whether hybrid samples are included in the sampling of the Florida lineage, which is estimated to overlap considerably with the Continental lineage (Table 9). However, it is likely that this value is inflated by the small and restricted distribution of the Florida lineage. In fact, the inferred area of overlap is restricted to the contact zone between the two lineages and includes all of the putative hybrid samples. This suggests that the Florida lineage inhabits a unique environment in Peninsular Florida. This is not surprising given that a number of studies have identified genetic breaks between Florida and

mainland populations (Burbrink et al. 2008; Fontanella et al. 2008; Ellsworth et al. 1994; Avise et al. 1984; Walker and Avis 1998). Consistency between ENM's for the Continental lineage combined with the dependency of estimates of potential overlap on the Florida lineage may suggest that hybridization is a result of migration from the Florida lineage north. However, coalescent based estimates of migration are required to corroborate such conjecture (see chapter 4).

The two lineages of *A. piscivorus* can be distinguished morphologically; however results were not consistent across analyses. Low sample sizes may hamper discrimination between lineages and suggest caution when interpreting the morphological analyses presented. Both the Florida and Continental lineage could be correctly classified 90% and 95% of the time by DFA. Significant morphological differences were not detected by ANOVA or MANOVA, although both lineages were different from the three lineages of *A. contortrix* according to MANOVA.

Either two or three lineages of *A. contortrix* were detected by scnDNA and mtDNA respectively. Morphological analysis and ENM's support only two lineageas of *A. contortrix*, corroborating results from Bayesian species delimitation for scnDNA presented in Chapter 2. This suggests that mtDNA reflects population structure that is not representative of current phylogeographic structure, possibly a result of populations segregated to unique glacial refugia with subsequent mixing during interglacials (Hewitt 2000). Alternatively, the East and Central lineage may represent lineages that are in the beginning of the "gray zone" proposed by de Queiroz (2007) and have not had sufficient time to acquire differences among multiple species criteria or for slower evolving

nuclear genes to sort. The West lineage of *A. contortrix* was correctly classified ~95% of the time morphologically using DFA. However, results from the ANOVA and MANOVA reveal that the West lineage can be separated morphologically from the east lineage but not the Central lineage suggesting that morphological differences in *A. contortrix* may exist along a cline from east to west. ANOVA suggests that morphological differences between the West and East lineages exist but that neither can be differentiated from the Central lineage. Contradictory results from MANOVA indicate that the East and Central lineages can be differentiated morphologically but that the West lineage does not differ morphologically from any of the lineages including the two lineages of *A. piscivorus*. These two contrasting results suggest that significant differences in the morphology of the West lineages are related to size. All three lineages were found to inhabit similar but not identical ecological niches according to both Schoener's *D* and Warren et al.'s *I*. However considerable geographic overlap between the potential distributions of the East lineage and Central lineage was estimated; representing 53% and 69% of their distributions respectively (Fig. 13e). This is quite different from the situation discussed for the Florida lineage in that the amount of overlap is significant for both lineages and is not restricted to a contact zone. Both *D* and *I* reveal that the West lineage and the combined East/Central lineage have similar but not identical ecological requirements. The potential distributions of both lineages overlap along a north south contact zone that includes all of the putative hybrids identified in Chapter 2.

Based on the phylogenetic, ecological, and morphological evidence presented I propose that two species of *A. piscivorus* and two species of *A. contortrix* should be

recognized. The two lineages of *A. piscivorus* were found to have distinct morphologies by DFA. However, morphological analysis was unable to provide individual characters that could be used for diagnosis. Similarly, both lineages were found to inhabit distinct ecological niches which overlap at a hybrid zone. Analysis of morphological data and ENM's failed to provide compelling support for recognizing the East and Central lineages of *A. contortrix* as distinct species. As discussed above, the East and Central lineages were both distinct from the West lineage morphologically but not from one another. The niche identity and background tests suggest that the East and Central lineages inhabit similar but non-identical niches, however this was not corroborated examining geographic overlap which suggests that the potential distributions of the two lineages are not geographically separate. Therefore, I suggest that *A. contortrix* be separated into two species, one comprised of the East and Central lineages and the other represented by the West lineage. Detailed history of synonymy for both *A. controtrix* and *A. piscivorus* are provided in (Gloyd and Conant 1990).

***Agkistrodon contortrix* (Linnaeus 1766)**

Eastern Copperhead

Holotype: Unknown.

Type Locality: "Carolina"(Linnaeus 1766), restricted to Charleston, SC by Schmidt (1953).

Etymology: Specific epithet refers to female contortionist, possible reference to dorsal pattern.

Synonymy: This species comprises the previously recognized subspecies *A. c. contortrix* (Linnaeus 1766)*, A. c. mokasen* (Palisot de Beauvois 1799)*,* and *A. c. phaeogaster* (in part; Gloyd 1969).

Diagnosis: The Eastern Copperhead (*A. contortrix*) is a medium sized pit viper with an average adult size 61-90cm and a maximum size of 132.1cm (Gloyd and Conant 1990; Conant and Collins 1991). There is a single anal plate, keeled dorsal scales and typically 23 midbody scale rows (range: 21-25; Gloyd and Conant 1990). Subcaudals range from 38-52 in males and 37-48 in females, while ventral scales number 139-157 with no variation between sexes (Gloyd and Conant 1990). A combination of geography and color pattern distinguishes the Eastern Copperhead from related species. There are 10-21 dark crossbands on a brown, tan or gray background. Cross bands are narrow at midbody and widen toward the sides, described as hourglass or dumbbell shaped, and may be broken at middorsum (Gloyd and Conant 1990). The head is coppery brown or reddish brown with a pale cheek stripe (Gloyd and Conant 1990). The approximate range of this species extends from Connecticut to Florida along the Atlantic coast, west to include eastern Texas north through eastern Oklahoma into eastern Kansas (Fig. 14a). Hybridization with the Broad-Banded Copperhead apparently occurs along an ecological transition zone from south-eastern Texas north to eastern Kansas defined by increasing elevation, decreasing precipitation and transition from mixed forest and cypress swamp to prairie grassland (Bailey 1995). Diagnosis may be difficult for some individuals in this area (Fig. 14a).

***Agkistrodon laticinctus* (Gloyd and Conant 1934)**

Broad-Banded Copperhead

Holotype: UMMZ75599, collected by William A, Bevan and R. F. Harvey.

Type Locality: Twenty six miles northwest of San Antonio, TX

Etymology: Specific epithet refers to color pattern, derived from Latin *latus* and *cinctus*, translated as “side” or “broad” and “banded” respectively.

Synonymy: This species comprises the previously recognized subspecies *A. c. laticinctus* (Gloyd and Conant 1934)*, A. c. pictigaster* (Gloyd and Conant 1943)*,* and *A. c. phaeogaster* (in part; Gloyd 1969).

Diagnosis: The Broad-Banded Copperhead (*A. laticinctus*) is a medium sized pit viper with an average adult size 56-76cm and a maximum size of 95cm (Gloyd and Conant 1990; Conant and Collins 1991). There is a single anal plate, keeled dorsal scales and typically 23 midbody scale rows (range: 21-25; Gloyd and Conant 1990). Subcaudals range from 38-62 in males and 39-57 in females, while ventral scales number 138-155 with no variation between sexes (Gloyd and Conant 1990). A combination of geography and color pattern distinguishes the Broad-Banded Copperhead from related species. There are 10-18 brown to red crossbands on a light to medium to light brown background. Cross bands do not narrow at midbody (Gloyd and Conant 1990). Head coloration typically matches that of the crossbands with a lighter cheek stripe (Gloyd and Conant 1990). The approximate distribution of the *A. laticinctus* includes western and central Texas, central Oklahoma and eastern Kansas (Fig. 14a). Hybridization with the Eastern Copperhead occurs along an ecological transition zone from south eastern

Texas north to eastern Kansas defined by increasing elevation, decreasing precipitation and transition from mixed forest and cypress swamp to prairie grassland. Diagnosis may be difficult for some individuals in this area (Fig. 14a). The Broad-Banded copperhead can be distinguished from the Eastern copperhead by shape of the crossbands which do not narrow at the center.

Agkistrodon piscivorus (Lacepède 1789)

Northern Cottonmouth

Holotype: Unknown.

Type Locality: "Carolina" (Lacepède 1789), restricted to Charleston, SC by Schmidt (1953).

Etymology: Specific epithet refers to dietary habit, derived from Latin *pisces* and *vorare*, translated as 'fish' and 'to devour' respectively

Synonymy: This species comprises the previously recognized subspecies *A. p. piscivorus* (Lacepède 1789) and *A. p. leucostoma* (Troost 1836).

Diagnosis: The Northern Cottonmouth (*A. piscivorus*) is a medium to large bodied semi-aquatic pit viper with an average adult size 76-114cm and a maximum size of 188cm (Gloyd and Conant 1990; Conant and Collins 1991). They possess a single anal plate, keeled dorsal scales and typically 25 midbody scale rows (range: 23-27; Gloyd and Conant 1990). Subcaudals range from 38-53 in males and 36-50 in females, while ventral scales number 128-142 with no variation between sexes (Gloyd and Conant 1990). A combination of geography and color pattern distinguishes the Northern Cottonmouth

from the Florida cottonmouth. There are 10-17 dark crossbands on an olive, brown or black background. Cross bands are often indistinguishable in adults. The head is typically black or brown lacking vertical rostral stripes; a dark cheek stripe is present in juveniles but subdued or indistinguishable in adults (Gloyd and Conant 1990; Conant and Collins 1991). The Northern Cottonmouth occupies lower elevations throughout the south-eastern US from south-eastern Virginia to central Georgia, east of the Appalachian Mountains, north to southern Illinois and eastern Kansas, south into central Texas in the west (Fig. 14b). Hybridization with the Florida cottonmouth does seem to occur in the mid-Atlantic coastal plains in southern North Carolina to the southern coastal plains in south-eastern Louisiana and diagnosis may be difficult for some individuals in this area (Fig. 14b).

Agkistrodon conanti **(Gloyd 1969)**

Florida Cottonmouth

Holotype: USNM165962, collected by R. P. Elliot, J. Wariner, and P. Pinnel.

Type Locality: Seven miles southeast Gainseville, FL

Etymology: Specific epithet is a patronym honoring Roger Conant, a prominent American herpetologist of the 20th century.

Synonymy: This species comprises the previously recognized subspecies *A. p. conanti*.

Diagnosis: The Florida Cottonmouth (*A. conanti*) is a medium to large bodied semi-aquatic pit viper with an average adult size 76-122cm and a maximum size of 189.2cm (Gloyd and Conant 1990; Conant and Collins 1991). There is a single anal plate, keeled

dorsal scales and typically 25 midbody scale rows (range: 23-27; Gloyd and Conant 1990). Subcaudals range from 45-54 in males and 41-49 in females, while ventral scales number 135-145 in males and 132-144 in females (Gloyd and Conant 1990). A combination of geography and color pattern distinguishes the Florida Cottonmouth from related species. There are 11-16 dark crossbands on an olive, brown or black background which may become subdued in adults (Gloyd and Conant 1990). The head is typically brown with vertical stripes along the snout on the rostrals, prenasals, and first supralabials (Gloyd and Conant 1990). Dark stripes appear on the lower jaw extending from the mental to the first four or five infralabials (Gloyd and Conant 1990). A dark cheek stripe is present is bordered above and below by pale stripes (Gloyd and Conant 1990; Conant and Collins 1991). The distribution of the Florida Cottonmouth extends from southern Florida to approximately Savannah Georgia, and west to south-eastern Alabama (Fig. 14b). Hybridization with the Northern Cottonmouth occurs in the mid-Atlantic coastal plains in southern North Carolina to the southern coastal plains in south-eastern Louisiana and diagnosis to species may be difficult for some individuals in this area (Fig. 14b). The Florida Cottonmouth can be distinguished from the Northern Cottonmouth by the presence of rostral stripes and a prominent dark cheek stripe that persists into adulthood.

Chapter IV: Biogeography

Can gene tree/species tree discordance influence inference of biogeographic history of species?

Introduction

Understanding the processes that have influenced diversification and distributional patterns of extant species requires detailed examinations of phylogeographic and biogeographic histories (Pagel 1999). This includes investigating speciation processes, historical patterns of population size changes, and species interactions such as migration and hybridization. Robust inferences of the species tree may be required to provide credible estimates of species relationships, time of divergence, migration, and population size changes through time (Edwards 2009; Heled and Drummond 2008, 2010; Hey 2006). Often gene trees (or concatenated gene trees) are used to represent the underlying species tree assuming that all genes share the same topology. However, recent work has demonstrated that incomplete lineage sorting, horizontal gene transfer, recombination and gene flow can result in topological discord between genes (Degnan et al. 2009; Edwards 2009). In addition, it has been shown that branch length heterogeneity can result in discordance in divergence time estimates from independent loci (Thorne and Kishino 2002), leading to poor estimates of species divergence times. More importantly gene tree and concatenated methods reflect the approximate time that alleles coalesce, opposed to the coalescence of species. This means that gene trees overestimate species divergence since gene divergences predates the actual speciation event (Edwards and Beerli 2000; Carstens and Knowles 2007), although, this is not expected to be as important for older divergences, where the ratio of gene/species divergence times converge to 1.0. One recently proposed method addresses this problem inferring species divergence times by

incorporating fossil calibrations using multi-locus, species-tree coalescent methods (*BEAST; Heled and Drummond 2010). While this is a step forward for estimates of species relationships and time of divergence, additional methods are still required to understand demographic history, and migration dynamics in a multi-locus coalescent framework.

Phylogenetic methods that require multiple species to assess species relationships and divergence times do not provide estimates of population size changes and migration rates. Instead, additional analyses are required to construct a complete picture of the speciation process and population dynamics after divergence (Hey 2006; Hey and Nielsen 2007). Multi-locus coalescent methods have been developed for investigating historical changes in N_e (EBSP; Heled and Drummond 2008) and migration (IMa; Hey and Nielsen 2007). Both methods require estimates for mutation rates for at least one locus to provide estimates in demographic quantities (i.e. N_e in number of individuals, time in years). The most common source of mutation rates are those estimated by phylogenetic inference *a priori* or derived from the literature. However, estimates may depend on the phylogenetic method used. Species tree methods based on the multi-gene coalescent optimize the species tree and gene tree jointly (Liu and Pearl 2007; Heled and Drummond 2010) providing the potential for not only topological discordance compared to gene trees but also for discordance in parameter estimates such as mutation rates for individual genes. This creates a mechanism by which secondary analyses, such as estimates of migration and historical changes in N_e can be affected by gene tree/ species tree discordance. This is a topic that has received less

attention than comparisons of topology and branch lengths. Burbrink and Pyron (2011) recently compared estimates of diversification rates using the tree based statistic γ (Pybus and Harvey 2000) from gene trees and specie trees. Using both simulated and empirical data they demonstrated that gene trees estimated that diversification occurred earlier when estimates of θ were large but tended to be accurate for most realistic estimates of N_e. It is not surprising that tree based statistics can be impacted by discordance in phylogenetic tree reconstruction. However, it is unclear how sensitive analyses that require parameter estimates provided by phylogenetic tree inference will be to discordance between gene trees and species trees.

I attempt to address this question by exploring two analyses that rely on estimate of mutation rates. First, I explore the effect of using mutation rates estimated by gene trees versus species trees as priors to investigate changes in historical population size using a multi-locus coalescent method, extended Bayesian skyline plots (EBSP; Heled and Drummond 2008). The historical demographic model implemented in EBSP provides estimates for the number of population size changes as well as the direction (expansion or contraction) and approximate timing of shifts in N_e. Researchers commonly attempt to provide possible explanations for shifts in population size by correlating the inferred timing with geographic events, such as Pleistocene glacial cycles (Shapiro et al. 2004; Barnes et al. 2007; Beck et al. 2008). Therefore, it is important to know whether population size has remained constant or experienced expansion/contraction. Additionally, providing credible estimates of when shifts in population sizes occurred is also important because it will have a direct impact on the

interpretation of how species histories are shaped by historical events. Similarly, mutation rate estimates have the potential to influence estimates of migration rates using the isolation with migration model implemented in IMa. This method scales parameter estimates by the geometric mean of mutation rates among all loci. Variation between loci is accommodated using mutation rate scalars which are defined by the ratio between mutation rates. Therefore, parameter estimates are directly influenced by mutation rate estimates leading to potentially different conclusions if estimates are sufficiently different.

I examine historical changes in population size and migration rates for four species of *Agkistrodon* defined in Chapter 3; *A. contortrix*, *A. laticinctus*, *A. piscivorus* and *A. conanti*. Specifically, I address whether the hypotheses that the species in closest proximity to glaciated areas (e.g. *A. contortrix* and *A. piscivorus*) experienced population expansion following the last glacial maximum, while *A. laticinctus* and *A. conanti* retained constant population sizes throughout the Wisconsinan glacial cycle . I examine whether including admixed individuals alter inferences of historical demography through time. It is not obvious whether including hybrids is appropriate. The method of EBSP (Heled and Drummond 2008) assumes that populations are panmictic, suggesting that hybrids should be excluded. Therefore, hybridization results in some individuals that carry alleles of both species, violating the assumption of panmixia. It is unclear whether this will inform estimates of historical demographics or obscure signals of expansion or contraction. Finally, I explore whether migration between *A. piscivorus* and *A. conanti* and between *A. contortrix* and *A. laticinctus* occurs at equal rates in both

directions. As discussed in chapter 3, the potential distributions of *A. piscivorus* and *A. laticinctus* were dependent on whether hybrids were included while ENM's for *A. contortrix* and *A. piscivorus* were not sensitive to hybrid sampling. This suggests that hybridization may be a result of expanding realized niches of *A. conanti* and *A. laticinctus* into environments inhabited by adjacent species. This would be supported by asymmetrical migration rates out of *A. conanti* and *A. laticinctus* into *A. piscivorus* and *A. contortrix* respectively. I test this hypothesis by estimating migration rates for each species using the multi-locus coalescent approach of IMa (Hey and Nielsen 2007) to determine if migration occurs in a single direction out of *A. laticinctus* into *A. contortrix* and out of *A. conanti* into *A. piscivorus*.

I also examine the time of divergence using multi-locus coalescent methods. Estimates of time of divergence presented in Chapter 1 suggest that the origin of *Agkistrodon* is in the Late Miocene with subsequent diversification in the Late Pliocene and Early Pleistocene (Table 2; Fig. 5). This would suggest that Pleistocene glacial cycles were not a contributing factor to diversification within *Agkistrodon*. It has been suggested that Pleistocene glacial cycles played a major role driving diversification of many extant species (Johnson and Cicero 2004) However, some argue that processes during the Pliocene were just as important (Zink and Klicka 2006, Lovette 2005). As already mentioned, gene divergences necessarily pre-date species divergences (Edwards and Beerli 2000; Carstens and Knowles 2007). Therefore, estimates of Pliocene or Early Pleistocene gene divergence may actually reflect species divergence after the onset of Pleistocene climate change. I test this hypothesis using coalescent methods to estimate

time of species divergence and compare results to estimates of gene divergence from mtDNA and concatenated data.

Methods

Sequence Acquisition

To provide estimates of divergence time and mutation rates among the four species of *Agkistrodon* a total of nine single copy nuclear loci (scnDNA) and one mitochondrial locus (mtDNA) were sequenced for 26 Crotaline taxa and 1 Viperinae outgroup (*Bitis nasicornis*) including two individuals of each species of *Agkistrodon*. Whole genomic DNA was extracted from ethanol preserved liver, heart, muscle or scale clippings using the DNeasy Kit (Qiagen Inc.). Template material for the polymerase chain reaction (PCR) consisted of samples with DNA/RNA ratios of 1.5–2.1 and DNA concentrations from 10–200 ng/ll. Nuclear loci consisted of four anonymous loci (Anon A, Anon 11, Anon 51, Anon 63) identified for a closely related Crotaline genus, *Sistrurus* (Gibbs and Diaz 2010) and five previously identified scnDNA loci: Aryl Hydrocarbon Recepter (AHR, Wiens 2008); Neurotophin (NT3, Noonan and Chippindale 2006); Skeletal muscle sodium channel Intron 5 ($NA_v1.4$, Geffeney et al. 2005); Spectrin beta non-erythrocytic 1 Intron (SPTBN1, Mathee et al. 2001); Vimentin Intron 5 (Zehner and Paterson 1983). In addition, a single mitochondrial marker (Cytb) was sequenced. The following primers were used for amplification: Anon A: AnonAF 5’-AGA ATT GAG CTC CCG TCC TTT-3’, AnonAR 5’-GGG AGC AAT GCC TAG ACC AAG-3’; Anon 11: Anon11F 5’-TCC TTA CTG AGT GAG CAC C-3’, Anon11R 5'-GCA AAG TCA ATG GAG AAA G-3'; Anon 51:

Anon51F 5’-ACT TGC CTT CAG AAA TCA TG-3’, Anon51R 5'-ATC AAA GGT TTA AAG AA-3'; Anon 63: Anon63F 5’-ATT AGC CCA GAA CTG TGC TTA-3’, Anon63R 5’- AAA GAT TCT GGG AAG CCA AA-3’; AHR: AHRF 5'-GTC CAC CTG CTT CAA ATA A -3'; AHRR 5'-GYR AAC ATS CCA TTR ACT TGC AT -3'; $NA_v1.4$: NAV5F 5’- GGG CAA CGT CTC TGC TCT AC -3’, NAV6R 5’- CGA AGT TCC CCA TGA ACA GT -3’; NT3: NT3R 5'-GCG TTT CAT AAA AAT ATT GTT TGA CCG G-3'; NT3F 5-ATA TTT CTG GCT TTT CTC TGT GGC-3'; SPTBN1: SPTBN1SeqF 5’-ATA CAG GCT GAG CGA GTG AGA-3’, SPTBN1SeqR 5’-AGC TGA CAT AGC TCT TGG TAA CA-3’; Vimentin Intron 5: VimExon5F 5’-AAC AAT GAT GCC CTG CGC CA-3’, VimExon6R 5’-CAA TAT CAA GAG CCA TCT TTA CAT T-3’. Amplification of each gene was carried out using GoTaq Green Master Mix (Promega Corp). A 90s extension time was used for each gene along with the following annealing temperatures; Anon A (51), Anon 11 (44), Anon 51 (43), Anon 63(47), AHR (45), $NA_v1.4$(49), NT3 (51), SPTBN1 (51), and Vimentin Intron 5 (47), and. Successful PCR products were purified using 1 µl of ExoSAPIT (USB Corp.) per 10 µl of PCR product. The sequencing reaction for all genes consisted of 2 µl of DTCS (Beckman-Coulter), 1 µl of 2 µm primer and 2 µl of DNA template and 4 µl of water. Sequences were purified using ethanol precipitation and analyzed on a Beckman-Coulter CEQ8000 sequencer. Sequencing primers were as follows: Anon A: AnonAR 5’-GGG AGC AAT GCC TAG ACC AAG-3’; Anon 11: Anon11F 5’-TCC TTA CTG AGT GAG CAC C-3’; Anon 51: Anon51F 5’-ACT TGC CTT CAG AAA TCA TG-3’; Anon 63: Anon63F 5’-ATT AGC CCA GAA CTG TGC TTA-3’, Anon63R 5’- AAA GAT TCT GGG AAG CCA AA-3’; $NA_v1.4$: NAV5F 5’-GGG CAA CGT CTC TGC TCT AC -3’, NAV6R 5’- CGA AGT TCC CCA TGA ACA GT -3; NT3: NT3R 5’-GCG TTT CAT AAA AAT ATT GTT TGA CCG G-3’; SPTBN1: SPTBN1SeqF 5’-ATA

CAG GCT GAG CGA GTG AGA-3', SPTBN1SeqR 5'-AGC TGA CAT AGC TCT TGG TAA CA-3'; Vimentin Intron 5: VimExon5F 5'-AAC AAT GAT GCC CTG CGC CA -3', VimExon6R 5'-CAA TAT CAA GAG CCA TCT TTA CAT T-3'; AHR; WL325AHR_F5 5'-GTC CAC CTG CTT CAA ATA A-3'. Sequences were edited and aligned using MUSCLE (v. 3.6; Edgar 2004) in the program Genious v5.1 (Biomatters Ltd.).

Divergence Dating and mutation rate estimates

I estimate time of divergence for the origin and subsequent diversification within the genus *Agkistrodon*. Results presented in chapter 1 suggest a Miocene origin with diversification occurring in the Pleistocene prior to glaciations. I address whether these results are robust by incorporating a multi-locus approach to estimate time of species divergence. To examine discordance in divergence dates derived from different methods of tree inference I inferred individual gene trees for each locus, a gene tree using a concatenated dataset and species trees using a coalescent based method implemented in *BEAST (Heled and Drummond 2010). All tree inference was performed in the program BEAST v1.6.1 (Drummond and Rambaut 2007). This made it possible to incorporated tree uncertainty in the MCMC process, however only the coalescent based species tree method accounts for divergent coalescent histories of loci resulting from incomplete lineage sorting (ILS). The appropriate substitution model for each gene was determined using AIC and BIC in the program Jmodeltest (Posada 2008, Guindon and Gascuel 2003). Subsequently I examined three partitioning schemes for each gene; unpartitioned, partitioned by codon position, and combined first and second codon

position partitioned from the third. Posterior Bayes Factors (PBF) was used to estimate the optimal partitioning scheme for each gene independently. I examined a strict clock, relaxed uncorrelated lognormal clock and a random local clock model for each gene separately; using PBF's to determine the optimal clock model for each locus. Individual gene trees and the concatenated dataset included all individuals sequenced for the 10 loci. However, to provide estimates of population size (N_e), a necessary parameter for inferring gene coalescence, Bayesian species tree inference requires two individuals to be included for each species. Therefore, Bayesian species trees were inferred using a reduced dataset including only species which were represented by two individuals. The same node constraints used in chapter 1 (Table 1) were applied here, however the oldest Viperidae fossil was excluded due to limited sampling of extant viperids in this dataset and the absence of the Old World Viperidae outgroup. All fossil constraints used a lognormal prior distribution with the mean equal to the age of the fossil and SD equal to the time span of the North American Land Mammal Age encompassing the fossil. A Yule tree prior was used in all analyses with a piecewise linear with constant root population size prior for Bayesian species tree analysis. I determined burnin by examining traces by eye along with ESS values in Tracer v.1.4 (Rambaut and Drummond 2007) after $500x10^6$ generations.

Migration Rates

Migration rates between species for both *A. contortrix* and *A. piscivorus* were estimated in the program IMa (Hey and Nielsen 2007). Estimates of migration rates are

explored to determine if hybridization between adjacent species identified in Chapter 2 are a result of migration of one species into the other or if migration occurs in both directions. Analyses in IMa took advantage of the datasets presented in chapter 2, consisting of six genes (1 mtDNA and 5 phased scnDNA loci) for each species. It is necessary to provide estimates of mutation rates (substitutions/locus/year) to scale population size (N_e), time (t) and migration rates in number of years. Two identical analyses were performed differing only in the priors placed on mutation rates. The first analysis used priors on mutation rates for each gene estimated by individual gene trees (Table 12) while priors on mutation rates for the second analysis were based on estimates from the species tree (Table 12). A generation time of 3.0 years was used for both species (Gloyd and Conant 1990). I applied the HKY model (Hasegawa et al. 1985) to each gene and an inheritance scalar of 0.25 and 1.0 were applied to mtDNA and the five nuclear loci, respecively. Independent runs were performed with different seeds in the MCMC to estimate the unscaled priors for Θ of each population (q_1 q_2), m_1 m_2 and τ and determine appropriate values for the geometric heating of 30 chains. I ran the final dataset four times visiting 8 million trees with the following priors: *A. contortrix*: q_1=20, q_2 =20, q_a=25, m_1=10, m_2 =20, and τ_1=10; *A. piscivorus*: q_1=5, q_2 =5, q_a=40, m_1=20, m_2 =20, and τ_1=10. I used log-likelihood ratio tests (2LLR) to compare nested models in the "-L" mode of IMa. This analysis used the chi square (or mixed) distribution from the 2LLR to examine differences in 16 models against a full model where all five parameters differ: N_{e1}, N_{e2}, N_{ea}, m_1, and m_2. Additionally, m_1 and m_2 are evaluated where migration

is effectively = 0. I also used Akaike Information Criterion (AIC) to discriminate between multiple models not rejected by 2LLR.

Historical Demography

I examine historical demographics for each species using the method of extended Bayesian skyline plots (EBSP; Heled and Drummond 2008) in Beast 1.6.1 (Drummond and Rambaut 2007) to explore whether Pleistocene glacial cycles influenced population sizes of the four species of *Agkistrodon* distributed in the US. Changes in N_e over time for each species are estimated by inferring the coalescent history of each gene simultaneously. The same datasets presented in Chapter 2, consisting of six genes (1 mtDNA and 5 phased scnDNA loci) for each species were used. Four separate analyses were performed for each species to examine the affect of mutation rate priors and gene flow on estimates of historical demography. The first analysis incorporated mutation rates estimated by individual gene trees (Table 12) excluding individuals identified as putative hybrids in Chapter 2. Mutation rates for the second analysis were provided by Bayesian species tree inference. Both analyses were repeated including putative hybrids. To achieve reliable estimate samples sizes (ESS) greater than 200 replicate runs were combined for each analysis resulting in $1.9x10^8$ generations. Burnin was determined for each replicate by examining traces by eye along with ESS values in Tracer v.1.4 (Rambaut and Drummond 2007). I also determined the probable number of population size shifts in Tracer by examining the frequency

distribution of these changes under the parameter demographic.populationSizeChanges.

Results

Divergence Dating and mutation rate estimates

Sequences were obtained for 40 Viperid taxa and aligned by eye resulting in a total of 5237 bp. Indels (ranging between 0 and 12 per gene) were not used to inform the phylogenetic analyses but instead coded as missing data. Both the complete dataset and the reduced dataset used for species tree inference were 67 % complete. The concatenated dataset was missing 132 sequences out of a potential 400 over the 10 loci with each locus ranging from 47.5% to 95% complete (Table 13). The reduced dataset was missing 75 sequences out of a potential 230 with each loci being between 43.5% and 100% complete (Table 13). The mtDNA marker (Cytb) had the greatest number of variable sites with the nuclear markers showing a significant decrease in the number of variable sites (770 vs. >100; Table 13). Using AIC in jModeltest and PBF's, I determined the following substitution models to be most appropriate for each gene: Cytb: GTR+Γ partitioned among the three codons; Anon A: unpartitioned; Anon 11: unpartitioned; Anon 51: fist and second codon positions combined; Anon 63: unpartitioned; AHR: partitioned among the three codons; $NA_v1.4$: unpartitioned; NT3: partitioned among the three codons; SPTBN1: unpartitioned; Vimentin Intron 5: unpartitioned. Using PBF's I determined that a strict clock was the most appropriate clock model for all genes with the exception of Cytb and $NA_v1.4$ (relaxed lognormal clock) and SPTBN1 (random local

clock). Discarding the initial 20% as burnin resulted in ESS values above 200 for all parameters with the exception of prior and posterior in both the concatenated analysis and Bayesian species tree inference. This is potentially problematic, however results are consistent across repeated analysis indicating that results may be robust to low ESS values.

Relationships within *Agkistrodon* inferred by concatenated analysis and Bayesian species tree inference were congruent with phylogeographic structure inferred by mtDNA (Guiher and Burbrink 2008). Bayesian species tree inference provided high support (96% Pp) for the placement of *A. bilineatus* and *A. taylori* as a monophyletic sister group to the two species in the *A. piscivorus* complex. Both mtDNA analysis and concatenated analysis failed to provide adequate support for this node (Guiher and Burbrink 2008). There was discordance between the two methods on the placement of the Central and South American Crotalines. The concatenated tree placed the Central South American genera sister to the rattlesnake genera with *Agkistrodon* on a separate branch sister to this clade (Fig. 15a). Species trees inferred the Central and South American genera (*Bothreichis lateralis*, *B. schlegelii*, *Bothrops leucurus*, *Lachesis muta*, and *L. stenophrys)* to be sister to *Agkistrodon*, although neither method provided high support for this node (Fig. 15).

Estimates of divergence dates for species of *Agkistrodon* were consistently lower for both concatenated analysis and Bayesian species tree inference compared to divergence estimates provided by Cytb (Tables 2, 13). To explore how the discordance could be explained by increased taxon sampling within Crotalinae, I re-estimated

divergence dates inferred by Cytb alone including all available taxa. This resulted in dates much more similar to those inferred with the concatenated dataset (Table 14). However, dates inferred by species trees remain lower; in some instances mtDNA dates were three times higher than those inferred by species trees (Table 14). However, dates inferred by species trees should reflect speciation events as opposed to gene divergence dates (as estimated by gene tree and concatenation methods). In the absence of gene flow, divergence among genes always occurs prior to speciation (Edwards and Beerli 2000; Carstens and Knowles 2007; McCormack et al. 2010).

The origin of *Agkistrodon* is estimated to have occurred in the Late Miocene, although dates from species trees place this closer to the boundary between the Pliocene and Miocene, compared to the Mid Miocene estimate in Chapter 1 (Table 14; Fig 15; Guiher and Burbrink 2008). All three methods infer a Pliocene origin for the *A. contortrix* complex (Table 14; Fig. 15), lower than the Miocene origin estimated previously (Guiher and Burbrink 2008). Similarly, both concatenation and the revised Cytb analyses estimate a Miocene origin for the *A. piscivorus* complex, while species tree estimates suggest a Pleistocene origin. All three methods suggest that species diversification within the *A. piscivorus* and *A. contortrix* complexes occurred in the Pleistocene, although species tree estimates are considerably younger than concatenated methods (Guiher and Burbrink 2008).

Three of the gene trees (Anon 11, SPTBN1, and Vimentin Intron 5) failed to resolve monophyletic copperhead and cottonmouth species complexes, making mutation rate estimates for those nodes potentially ambiguous. However, either a strict

or local clock was used in all three cases resulting in a single mutation rate inferred over the tree. Mutation rates inferred by gene trees and species trees were remarkably similar in most cases for both species (Table 12). The greatest discordance was in estimates for Cytb and then SPTBN1 in both species (Table 12).

Estimates of Migration

I combined four independent runs for each analysis resulting in a total of 80,000 samples producing ESS values for all paramaters >200 following a burnin of 100,000 generations. Results were largely consistent between the two analyses estimating demographic parameters for the two species of copperheads. Likelihood ratio tests failed to reject five models in favor of the full model when using mutation rates estimated by the gene trees (Table 15a). Similarly, using mutation rates estimated by the species tree LLR tests failed to reject three models (Table 15b). The model where migration is equal in both directions and extant population sizes are equal but both are different from the ancestral population size received the lowest AIC for both analyses (Table 15a and 15b). Parameter estimates were similar between the two analyses. Estimates of divergence dates between the two species were consistent with estimates from species trees suggesting Pleistocene glacial cycles may have played a role in speciation. Equal migration in both directions was inferred with an ML estimate 0.0000065 migrants per generation per gene copy (95% CI = 0.0000016-0.0000170). Effective population size of the extant species is predicted to be much smaller than the ancestral population (Table 16). Results were less consistent between the two analyses

comparing demographic parameters between two species of cottonmouths. Nine models were not rejected by the LLR tests when mutation rates were estimated by the gene trees (Table 15c). However, models could be discriminated by AIC, which favored a model with equal migration rates in both directions and equal extant population sizes both smaller than the ancestral population size (Table 15c). Similarly, ten models were not rejected by LLR when mutation rates were estimated by the species tree; AIC could not easily discriminate between five models given low ΔAIC values (Table 15d). Four of these models have equal migration in both directions and include several combinations of population size parameters where all population sizes are equal, all population sizes are unique, the population size of *A. piscivorus* is equal to the ancestral population size, and the population size of *A. conanti* equals the ancestral population size. The fifth model suggests that distinctly different sizes are attributed to each species and migration occurred in one direction from *A. conanti* into *A. piscivorus* (Table 15d). The gene tree mutation rate analysis estimates equal migration rates in both directions at a rate of 0.0000092 gene copies per generation (95% CI = 0.0000021-0.0000282). Analysis using the species tree mutation rate estimates a migration rate of 0.0000086 gene copies per generation (95% CI = 0.0000022-0.0000258), however it was not possible to discriminate between equal migration in both directions and only migration from *A. conanti* into *A. piscivorus*. The Ancestral population size is estimated to be much larger than the population size of either extant species (Table 16).

Historical Demographics

I determined the appropriate model of evolution for each species and each gene using BIC in jModeltest, approximating the closest model available in BEAST v1.6.1 (Drummond and Rambaut 2007). The following models were used for each species: *A. contortrix*: Cytb=GTR+Γ+I, Anon 11=HKY, Anon 51=HKY, Anon A=HKY+I, NT3=HKY+ Γ, and SPTBN1=GTR+ Γ; *A. laticinctus*: Cytb=HKY+Γ, Anon 11=HKY+ Γ, Anon 51=HKY, Anon A=HKY, NT3=HKY, and SPTBN1=HKY; *A. piscivorus*: Cytb=GTR+Γ+I, Anon 11=HKY+ Γ+I, AHR=HKY, NT3=HKY, and SPTBN1=HKY+I, and Vimentin Intron 5=HKY+I; PRLR=HKY+ Γ; *A. conanti*: Cytb=HKY+Γ, Anon 11=HKY+ Γ, AHR=HKY, NT3=HKY, and SPTBN1=HKY+ Γ, and Vimentin Intron 5=HKY. Population changes through time were estimated with EBSP and combined runs resulted in between 200 million and 1.9 billion generations for each lineage. Obtaining ESS >200 for all paramaters was not possible for analyses of *A. contortrix* and *A. piscivorus*. Examination of the four analyses for each species reveals that using mutation rates from gene trees consistently resulted in older dates (increase of 14% -312%) for the coalescence of alleles within the lineage. Including hybrid individuals resulted in an even more pronounced increase (2.5-37.5 times older) in time of population coalescence. A similar trend was also found for estimates of the timing of population size changes resulting in estimates of the timing of expansion that were 1.2-2.5 times older when hybrid samples were included. Estimates of population size through time suggest that both *A. contortrix* and *A. piscivorus* experienced expansion coinciding with the retreat of the Laurentide ice sheet (Figs. 16 and 17) in all four analyses for each species. These results are supported by estimates of the parameter demographic.populationSizeChanges indicating two population size changes for all

analyses of *A. contortrix* and one population size change for all analyses of *A. piscivorus*. Examination of the EBSP's does not reveal a signal of expansion for *A. laticinctus* or *A. conanti* (Figs. 16 and 18). A mean of zero population size changes was estimated for *A. conanti* for all analyses except where the species tree mutation rate was used and hybrids were included (mean=1). Similarly, a population size change equal to zero was inferred for *A. laticinctus* when hybrids were excluded; one and two population size changes were inferred when hybrids were included with a gene tree mutation rate and the species mutation rate respectively.

Discussion

I have demonstrated that discordance between gene trees and species trees can alter the interpretation of species history by influencing analyses dependent on phylogenetic tree inference for priors on parameter estimates. Estimates of population sizes, timing of changes in population size and migration rates are susceptible to variation in mutation rates.

Previous work has demonstrated that divergence time estimates are sensitive to taxon sampling (Linder et al. 2005) resulting in a bias toward younger ages when taxa are under sampled. A comparison of divergence estimates inferred by mtDNA with increased sampling to estimates provided in Chapter 1 support the conclusion that divergence estimates are sensitive to sampling (Table 2 and 13). However, increasing sampling within Crotalinae resulted in the opposite trend of increased node ages with decreased sampling. This discordance may not be significant since the placement of the

additional taxa was not generally well supported (Fig. 15). Underscoring this, the placement of the South American and Central American species was not supported by the single mtDNA gene tree, concatenated gene tree or the species tree. Moreover, all three analyses suggested different relationships with respect to the CA and SA taxa (Fig. 15). Resolving relationships among New World Crotalines has proved problematic for previous studies as well (Castoe and Parkinson 2006; Wuster et al. 2008) making it difficult to be certain whether the rattlesnakes (*Sistrurus* and *Crotalus*) or Central and South American taxa represent the sister lineage to *Agkistrodon*. This uncertainty may influence dating divergences within *Agkistrodon*, however since relationships among *Agkistrodon* species are consistent across analyses the effect should be most evident at the origin of the genus.

Divergence estimates were consistently younger in species trees compared to concatenated trees (Table 14; Fig. 15). This is not surprising since gene divergence necessarily predates species divergence (Edwards and Beerli 2000). In addition, divergence dates for most nodes were younger in species trees and concatenated analysis than dates suggested by divergence dating with a single mtDNA gene (Tables 2 and 13; Fig 15; Chapter 1; Guiher and Burbrink 2008). The origin of *Agkistrodon* likely occurred earlier than previously suggested (Chapter 1; Table 2; Table 14) but considering the uncertainty around the closest relative of *Agkistrodon* this is speculative at this time. However, it is evident that the timing of diversification within *Agkistrodon* was overestimated by mtDNA. Species trees suggest a Pliocene origin for all three species complexes, as compared to Miocene estimates using a single gene tree. In

addition, species diversification within *A. contortrix* and *A. piscivorus* most likely occurred during the Pleistocene. This suggests that Pleistocene glacial cycles may have played a role in speciation, restricting populations into distinct refugia, a result not supported by Pliocene diversification suggested in Chapter 1. This conclusion is supported for the two copperhead species by estimates for the splitting time parameter *t* provided by IMa (Table 16). Similarly, species tree estimates of divergence time suggest that Pleistocene glaciations may have played a major role in the speciation of *A. piscivorus* and *A. conanti* Table 14; Fig 15), contrasting results presented in chapter 1. This may have been a result of rising sea levels restricting Peninsular Florida to a series of islands isolating populations of *A. piscivorus* in Florida from those north of Florida (Webb 1990; Clark et al. 1999). Estimates of the parameter *t* in IMa do not support this conclusion (Table 16) however, estimates of splitting time from IMa include a wide confidence interval spanning 115kya to ~8.2mya.

Estimates of population size, migration rates and splitting time for species of copperheads appear robust to variation in mutation rates. Likelihood ratio test and AIC support a model with equal population sizes (N_e =83,369, 95% CI=35,503-144,743). This is perhaps surprising given the much larger range of *A. contortrix* than *A. laticinctus*. Migration is estimated to occur equally in both directions. Therefore, it does not appear that hybridization is solely a result of *A. laticinctus* expanding into novel environments as ENM's presented in Chapter 3 might suggest. A possible scenario that would account for this pattern is the combination of expansion of the realized niche of *A. laticinctus* and population growth of *A. contortrix* following the last glacial maximum. Expansion

out of potential refugia likely resulted in secondary contact since hybrid samples appear to be limited to an ecological transition zone along the contact zone between both species.

It was not possible to discriminate between several nested models considered by IMa for the two species of cottonmouths. The best model identified by LLR and AIC was different between analyses using gene tree mutation rates or species tree mutation rates. Furthermore, a greater number of models were considered plausible when species tree mutation rates were used. Therefore, I cannot be certain whether population sizes are different for both extant species or if either is different from the ancestral population (N_{eAp}, Ne_{Ac}, and N_{eA}). Regardless, estimates of population sizes for *A. piscivorus* and *A. conanti* suffered from the same shortcomings discussed above. The majority of models supported by both analyses indicate that migration occurs equally in both directions. However, analyses with species mutation rates could not reject a single model which inferred unidirectional migration from *A. conanti* into *A. piscivorus*. Examination of ecological niche models in Chapter 3 supported this hypothesis by demonstrating that a putative hybrid zone could only be predicted by including hybrid individuals in the modeling of *A. conanti* and not *A. piscivorus*. However, results are inconclusive regarding the direction migration between the two species.

Extended Bayesian skyline plots suggest that both *A. contortrix* and *A. piscivorus* experienced population expansion following the last glacial maximum (LGM). This is in agreement with conclusions drawn from mtDNA in Chapter 1. However, the addition of multiple scDNA loci revealed that *A. contortrix* experienced two shifts in population size

in response to Pleistocene glaciations, population contraction after approximately 250Kya and recent population expansion. Analyses using mutation rates estimates by gene trees resulted in older mean estimates for the timing of the expansion. Species tree rates inferred possible Holocene expansion after 10 kya in *A. piscivorus* and 25kya in *A. contortrix* compared to 25Kya and 30Kya inferred by gene tree rates for each species respectively. The difference was not sufficient to influence our conclusions regarding species response to Pleistocene glacial cycles for these two species. However, differences of this magnitude could potentially impact conclusions over older time scales.

Using multi-locus coalescent methods to examine timing of divergence, historical changes in population sizes and migration rates between adjacent species revealed that Pleistocene glacial cycles played an important role in shaping the phylogeographic and biogeographic histories of *Agkistrodon* species within the US. Estimates of divergence times from species trees suggest that the genus *Agkistrodon* first diverged during the Miocene with subsequent diversification of the three species complexes occurring during the Pliocene and early Pleistocene. This is considerably younger than the Miocene origin of the three species complexes suggested by mtDNA (chapter 1; Guiher and Burbrink 2008). In addition, divergence estimates from species trees suggest that diversification within the *A. contortrix* and *A. piscivorus* complexes occurred more recently during the Pleistocene than previously suggested, likely as a result of glaciations of North America during the Pleistocene. This was most likely caused by populations being restricted to distinct refugia during periods of glacial maxima. In addition, rising

sea levels would have restricted populations of cottonmouths in Florida to a series of islands limiting gene flow between *A. conanti* and *A. piscivorus*. Following the retreat of the Laurentide ice sheet *A. contortrix* and *A. piscivorus* experienced population expansion out of refugia, resulting in gene flow at the contact zone. The hypothesis that gene flow between *A. contortrix* and *A. laticinctus* is a result of *A. laticinctus* expanding into novel environments inhabited by *A. contortrix* was not supported by estimates of migration rates, indicating that migration occurs at equal rates in both directions. However, I was unable to reject the hypothesis that gene flow between the two species of cottonmouths is a result of *A. conanti* expanding into environmental conditions suitable for *A. piscivorus* since LLR and AIC failed to discriminate between a single model which included unidirectional migration out of *A. conanti* into *A. piscivorus* from two models which included equal migration in both directions.

I believe that this study has demonstrated that discordance between gene trees and species trees can shape our understanding of species histories beyond phylogenetic relationships and time of divergence. Variation in parameter estimates should be taken into consideration as a source of error when conducting analyses of historical demographics (Figs. 16-19) or migration (Table 15) that are dependent on phylogenetic inference to define priors on parameters.

Tables

Tables

Table 1. Fossil Calibrations used for divergence dating of mtDNA gene tree

Fossil Calibration	Node	Date	95% PCI
C1	Viperidae	23.8my	(18.36-30.08my)
C2	Crotalinae	16.4my	(13.91-19.33my)
C3	*Sistrurus*/*Crotalus*	9.0my	(5.50-14.74my)
C4	Agkistrodon MRCA	6.5my	(4.68-9.03my)

Fossil Calibrations used for divergence dating of the mtDNA gene tree in the Program Beast v1.4.6 (Drummond and Rambout 2003). Median dates are provided along with 95% Posterior Confidence Interval (PCI) in parentheses.

Table 2. Estimates of divergence dates for key nodes of mtDNA gene tree

Node	Taxon/Lineage	Divergence Time	95% HPD
1	*Crotalus*/*Agkistrodon*	15.56mya	(13.04-18.17mya)
2	*A. contortrix*	6.60mya	(4.84-8.53mya)
3	Western clade	1.38mya	(0.70-2.19mya)
4	Eastern clade/Cenral Clade	0.96mya	(0.46-1.56mya)
5	*A. piscivorus*	5.30mya	(3.64-7.14mya)

Mean divergence time inferred for each lineage and outgroups in the program Beast v1.4.6 (Drummond and Rambaut 2003; Drummond et al. 2005). The 95% higher posterior density around each estimate is shown in parentheses.

Table 3. Demographic population parameters estimated for mtDNA

Clade	n	Fu's F	Modality	SSD	HRI	Range within Clade Pairwise Distance (pi)	Median of TMRCA
Eastern	49	-2.674	Unimodal	0.002	0.019	0.00000-0.00562	0.029mya (0.004-0.089mya)
Central	58	-5.181*	Unimodal	0.049	0.152	0.00000-0.00672	0.095mya (0.021-0.229mya)
Western	30	-2.147	Unimodal	0.072	0.194	0.00000-0.00961	0.128mya (0.035-0.297mya)
Continental	77	-5.354*	Unimodal	0.074	0.159	0.00000-0.00737	0.078mya (0.016-0.207mya)
Florida	24	1.342	Multimodal	0.087	0.206*	0.00000-0.00723	0.176mya (0.041-0.436mya)

P-values <0.005 for Fu's F statistic and Harpending's Raggedness Index (HRI) are indicated by "*". Estimates for time since the most recent common ancestor (TMRCA) (95% higher posterior density shown in parentheses) were calculated using Bayesian Skyline Plots in the program BEAST v1.4.6 (Drummond and Rambaut 2006; Drummond et al. 2005).

Table 4. Sequence summary statistics for the six loci sequenced for A. contortrix and A. piscivorus

Species	Locus	n	No. Hap.	Nucleotide Diversity (π)	Mean number of pairwise differences (k)
A. contortrix	Cyt b	93	39	0.01313±0.00109	14.530±4.969
	Anon A	74	7	0.01297±0.00138	2.166±1.460
	Anon 11	70	5	0.00349±0.00075	0.767±0.320
	Anon 51	81	5	0.00345±0.00025	0.972±0.441
	NT3	77	5	0.00493±0.00280	1.932±1.223
	SPTBN1	56	5	0.00387±0.00087	2.175±1.486
A. piscivorus	Cyt b	129	23	0.01824±0.00175	17.513±5.963
	Anon 11	120	8	0.01104±0.00061	2.572±0.866
	AHR	115	2	0.00131±0.00009	0.394±0.133
	NT3	116	2	0.00194±0.00011	0.420±0.141
	SPTBN1	102	12	0.00266±0.00032	1.894±0.640
	VimIntron5	117	5	0.00034±0.00012	0.097±0.033

Table 5. Posterior probability of speciation inferred by BPP v.2.0 (Yang & Rannala 2010) when the number of individuals with admixture comprised between 10-90% of the dataset.

	A. piscivorus				A. contortrix			
Percent Hybrids	Majority Start Tree 0	Majority Start Tree 1	Minority Start Tree 0	Minority Start Tree 1	Majority Start Tree 0	Majority Start Tree 1	Minority Start Tree 0	Minority Start Tree 1
10	1.00	1.00	0.99	1.00	0.99	1.00	0.99	1.00
20	0.96	1.00	0.86	1.00	1.00	1.00	1.00	1.00
30	0.86	1.00	0.00	1.00	0.98	1.00	0.33	1.00
40	0.84	1.00	0.00	1.00	0.90	1.00	0.03	1.00
50	0.91	1.00	0.07	1.00	0.95	1.00	0.01	1.00
60	0.00	1.00	0.00	1.00	0.25	1.00	0.00	1.00
70	0.00	1.00	0.00	1.00	0.19	1.00	0.00	1.00
80	0.00	1.00	0.00	1.00	0.03	1.00	0.00	1.00
90	0.00	1.00	0.00	1.00	0.13	1.00	0.00	1.00

Speciation probabilities, reflecting the posterior probability of a speciation event between two lineages, are shown for four runs with different model parameters of both *A. piscivorus* and *A. contortrix*: (i) hybrids assigned to the lineage which contributed the majority of alleles with a fully collapsed starting tree (0), (ii) hybrids assigned to the lineage which contributed the majority of alleles with a fully resolved starting tree (1), (iii) hybrids assigned to the lineage which contributed the minority of alleles with a fully collapsed starting tree (0), (iv) hybrids assigned to the lineage which contributed the minority of alleles with a fully resolved starting tree (1).

Table 6. Threshold independent model performance calculated by AUC for ENM's of the five lineages of A. contortrix and A. piscivorus

Lineage	Background/Bias correction	AUC	Variable contributions for the best model in order of contribution
East	USA/Colubroid	0.977*	Bio 14 (67.7%), Bio 17 (17.6%), Bio 2 (3.9%), Bio 9 (2.4%), Bio 3 (2.4%),
	East/ Colubroid	0.894	Bio 15 (2.2%), Bio 6 (1.9%), Bio 1 (0.6%), Bio 11 (0.5%), Bio 8 (0.4%),
	USA/None	0.920	Bio 13 (0.1%), Bio 7 (0.1%), Bio 18 (0.1%), Bio 5 (0.1%), Bio 19 (0.0%),
	East/None	0.784	Bio 16 (0.0%), Bio 10 (0.0%), Bio 4 (0.0%), Bio 12 (0.0%)
Central	USA/ Colubroid	0.985*	Bio 14 (32.2%) Bio 9 (20.3%) Bio 18 (13.5%) Bio 13 (9.3%),
	East/ Colubroid	0.948	Bio 10 (5.6%), Bio 1 (4.8%), Bio 12 (4.3%), Bio 15 (2.6%), Bio 17 (2.1%),
	USA/None	0.973	Bio 4 (1.6%), Bio 5 (1.4%), Bio 2 (0.7%), Bio 3 (0.4%), Bio 19 (0.3%),
	East/None	0.939	Bio 6 (0.3%), Bio 8 (0.3%), Bio 16 (0.0%), Bio 7 (0.0%), Bio 11 (0.0%)
East/Central	USA/ Colubroid	0.975*	Bio 14 (41.1%) Bio 9 (11.5%) Bio 18 (9.3%) Bio 12 (7.1%), Bio 1 (6.9%),
	East/ Colubroid	0.872	Bio 17 (5.5%), Bio 13 (5.0%), Bio 2 (3.9%), Bio 4 (2.8%), Bio 10 (1.7%),
	USA/ None	0.840	Bio 15 (1.4%), Bio 3 (0.9%), Bio 11 (0.8%), Bio 7 (0.6%), Bio 5 (0.5%),
	East/ None	0.947	Bio 8 (0.4%), Bio 19 (0.3%), Bio 6 (0.3%), Bio 16 (0.1%)
East/Central	USA/ Colubroid	0.978*	Bio 12 (19.4%) Bio 18 (12.6%) Bio 3 (10.8%) Bio 13 (9.4%),
With hybrids	East/ Colubroid	0.887	Bio 10 (8.5%), Bio 9 (7.9%), Bio 15 (6.7%), Bio 14 (6.6%), Bio 5 (4.7%),
	USA/ None	0.946	Bio 2 (4.5%), Bio 16 (3.0%), Bio 17 (1.6%), Bio 7 (1.3%), Bio 8 (1.0%),
	East/ None	0.854	Bio 1 (0.7%), Bio 11 (0.6%), Bio 19 (0.2%), Bio 4 (0.2%), Bio 6 (0.1%)

Table 6. continued

Lineage	Background/Bias correction	AUC	Variable contributions for the best model in order of contribution
West	USA/ Colubroid	0.949*	Bio 8 (44.0%) Bio 9 (21.4%) Bio 1 (17.3%) Bio 18 (5.4%), Bio 10 (5.1%),
	East/ Colubroid	0.922	Bio 15 (3.1%), Bio 17 (2.3%), Bio 13 (1.1%), Bio 5 (0.2%), Bio 14 (0.0%),
	USA/ None	0.902	Bio 12 (0.0%), Bio 11 (0.0%), Bio 6 (0.0%), Bio 4 (0.0%), Bio 3 (0.0%),
	East/None	0.791	Bio 2 (0.0%), Bio 19 (0.0%), Bio 16 (0.0%), Bio 7 (0.0%)
West	USA/ Colubroid	0.962*	Bio 18 (48.5%) Bio 15 (12.3%) Bio 9 (11.6%) Bio 14 (11.4%),
With Hybrids	East/ Colubroid	0.878	Bio 10 (5.8%), Bio 13 (2.4%), Bio 2 (1.6%), Bio 8 (1.6%), Bio 12 (1.5%),
	USA/ None	0.956	Bio 1 (1.4%), Bio 19 (0.9%), Bio 5 (0.4%), Bio 16 (0.4%), Bio 6 (0.2%),
	East/ None	0.900	Bio 3 (0.1%), Bio 11 (0.0%), Bio 17 (0.0%), Bio 7 (0.0%), Bio 4 (0.0%)
Continental	USA/ Colubroid	0.992*	Bio 14 (41.3%) Bio 17 (17.1%) Bio 11 (9.6%) Bio 9 (7.4%),
	East/ Colubroid	0.965	Bio 2 (4.0%), Bio 15 (3.9%), Bio 7 (3.6%), Bio 18 (2.8%), Bio 13 (2.4%),
	USA/ None	0.965	Bio 8 (1.3%), Bio 5 (1.2%), Bio 1 (1.2%), Bio 6 (1.2%), Bio 10 (1.1%),
	East/ None	0.919	Bio 16 (0.7%), Bio 3 (0.6%), Bio 4 (0.5%), Bio 12 (0.1%), Bio 19 (0.0%)
Continental	USA/ Colubroid	0.989*	Bio 14 (40.7%) Bio 17 (19.8%) Bio 11 (11.3%) Bio 2 (5.5%),
With Hybrids	East/ Colubroid	0.961	Bio 9 (5.3%), Bio 15 (4.3%), Bio 7 (3.4%), Bio 13 (2.8%), Bio 6 (1.3%),
	USA/ None	0.965	Bio 18 (1.0%), Bio 10 (1.1%), Bio 8 (1.0%), Bio 1 (0.9%), Bio 3 (0.5%),
	East/ None	0.928	Bio 5 (0.4%), Bio 4 (0.3%), Bio 16 (0.1%), Bio 12 (0.1%), Bio 19 (0.0%)
Florida	USA/ Colubroid	0.971	Bio 18 (48.1%) Bio 11 (12.3%) Bio 14 (12.3%) Bio 3 (11.6%),
	East/ Colubroid	0.939	Bio 1 (4.3%), Bio 15 (4.0%), Bio 4 (3.1%), Bio 10 (1.2%), Bio 9 (1.0%),
	USA/ None	0.992*	Bio 17 (0.6%), Bio 6 (0.3%), Bio 19 (0.2%), Bio 8 (0.2%), Bio 16 (0.2%),
	East/ None	0.973	Bio 12 (0.2%), Bio 7 (0.1%), Bio 5 (0.1%), Bio 2 (0.0%), Bio 13 (0.0%)

Table 6. continued

Lineage	Background/Bias correction	AUC	Variable contributions for the best model in order of contribution
Florida	USA/ Colubroid	0.990	Bio 18 (45.7%) Bio 3 (17.6%) Bio 14 (15.3%) Bio 11 (14.1%),
With Hybrids	East/ Colubroid	0.962	Bio 4 (3.2%), Bio 15 (1.1%), Bio 1 (0.5%), Bio 10 (0.5%), Bio 9 (0.4%),
	USA/ None	0.992*	Bio 8 (0.4%), Bio 19 (0.3%), Bio 6 (0.2%), Bio 17 (0.1%), Bio 5 (0.1%),
	East/ None	0.983	Bio 13 (0.1%), Bio 16 (0.1%), Bio 12 (0.1%), Bio 7 (0.1%), Bio 2 (0.0%)

Model performance was assessed for four backgrounds sampling schemes: (i) sampled randomly from the entire USA, (ii) sampled randomly from an East background restricted to the approximate distribution of Agkistrodonin the USA, (iii) sampled throughout the USA with bias correction using colubroid sampling, and (iv) restricted East sampling with bias correction using colubroid sampling in the eastern USA. Models with the highest AUC's between the four models are signified by "*" for each lineage both with and without hybrids and were used to calculate niche overlap in Table 6. Model contribution of the 19 bioclim variables are shown in parentheses for the best model chosen for each sampling scheme.

Table 7. Niche identity statistics for pairwise comparisons of adjacent clades.

Lineage	Hybrids	I ($\overline{X}$,σ, P)	D ($\overline{X}$,σ, P)
East vs. Central	NA	0.576 (0.935, 0.026, <0.0000001)	0.283 (0.731, 0.055, <0.0000001)
East/Central vs. West	YES	0.631 (0.912, 0.033, <0.0000001)	0.348 (0.684, 0.046, <0.0000001)
East/Central vs. West	NO	0.551 (0.871, 0.036, <0.0000001)	0.291 (0.729, 0.053, <0.0000001)
Continental vs. Florida	YES	0.514 (0.959, 0.017, <0.001)	0.251 (0.787, 0.043, <0.0000001)
Continental vs. Florida	NO	0.400 (0.952, 0.017, <0.001)	0.151 (0.766, 0.042, <0.001)

Observed values of D and I from the niche equivalency test of Warren et al. (2008) implemented in ENMTools with the mean and standard deviation of the null distribution shown in parentheses.

Table 8. Statistics from Niche background test for pairwise comparisons of adjacent clades

Lineage	Hybrids	I ($\overline{X}$,σ, P)	D ($\overline{X}$,σ, P)
East vs. Central	NA	0.576 (0.639, 0.023, =0.289)	0.283 (0.330, 0.018, =0.297)
Central vs. East	NA	0.576 (0.511, 0.039, =0.352)	0.283 (0.236, 0.028, =0.342)
East/Central vs. West	YES	0.631 (0.457, 0.079, =0.410)	0.348 (0.217, 0.059, =0.386)
West vs. East/Central	YES	0.631 (0.456, 0.030, =0.411)	0.348 (0.175, 0.021, =0.410)
East/Central vs. West	NO	0.551 (0.454, 0.075, =0.368)	0.291 (0.215, 0.057, =0.357)
West vs. East/Central	NO	0.551 (0.514, 0.015, =0.319)	0.291 (0.242, 0.011, =0.324)
Continental vs. Florida	YES	0.514 (0.482, 0.048, =0.335)	0.251 (0.243, 0.029, =0.323)
Florida vs. Continental	YES	0.514 (0.449, 0.023, =0.331)	0.251 (0.202, 0.017, =0.318)
Continental vs. Florida	NO	0.400 (0.415, 0.051, =0.312)	0.151 (0.190, 0.027, =0.301)
Florida vs Continental	NO	0.400 (0.413, 0.022, =0.313)	0.151 (0.169, 0.014, =0.310)

Observed values of D and I from the niche background test of Warren et al. (2008) implemented in ENMTools with the mean and standard deviation of the null distribution shown in parentheses.

Table 9. Area of predicted habitat suitability for the inferred lineages of *A. contortrix* and *A. piscivorus* with and without hybrids included in training data. Niche overlap calculations for pairwise comparisons of adjacent clades

Lineage	Hybrids	Area
Eastern	NA	2,174,030 km^2
Central	NA	1,667,736 km^2
East/Central	YES	2,913,464 km^2
East/Central	NO	2,481,939 km^2
Western	YES	1,306,498 km^2
Western	NO	603,212 km^2
Continental	YES	1,649,205 km^2
Continental	NO	1,582,128 km^2
Florida	YES	507,596 km^2
Florida	NO	272,776 km^2
East vs. Central	NA	1,155,547 km^2
East/Central vs. West	YES	766,848 km^2
East/Central vs. West	NO	131,061 km^2
Continental vs. Florida	YES	329,096 km^2
Continental vs. Florida	NO	126,941 km^2

Area calculated as the total number of 30 arc second pixels predicted as suitable using the MTP threshold, multiplied by 0.86 (30 s of arc = 0.86 km2). Overlap between potential distributions for pairwise comparisons of adjacent clades calculated as the total area of the zone of predicted overlap in environmental suitability from the ENMs.

Table 10. Principle component loadings for all variables on the first three axes.

Character	Axis one	Axis two	Axis three
SVL	-0.852	0.420	0.127
TL	-0.887	0.097	0.043
CN	-0.945	-0.059	0.102
CM	-0.953	0.015	0.201
CV	-0.898	0.251	0.202
HLL	-0.983	-0.043	-0.071
HLR	-0.972	-0.020	-0.094
HW	-0.926	-0.105	0.073
HH	-0.943	-0.193	0.071
BW	-0.944	-0.022	0.183
BH	-0.945	-0.005	0.192
EYE R	-0.914	-0.281	-0.090
EYE L	-0.895	-0.212	-0.145
NER	-0.950	0.027	-0.037
NEL	-0.964	0.030	0.014
NSR	-0.815	0.349	-0.038
NSL	-0.828	0.308	-0.218
WBN	-0.954	-0.021	-0.083
WBE	-0.930	-0.211	-0.056
PW	-0.935	-0.202	-0.102

Table 11. Canonical discriminant functions based on all variables. The proportion of the sum of the eigen values are shown in parentheses.

Character	Axis one (23.9%)	Axis two (1.1%)	Axis three (0.6%)
SVL	-4.844	1.014	2.391
TL	-0.204	0.547	0.191
CN	-0.069	-0.478	-0.545
CM	0.491	0.689	2.287
CV	0.054	-1.147	-1.648
HLR	0.127	0.170	0.002
HLL	2.822	-0.786	-1.333
HW	-0.068	0.043	-0.165
HH	0.714	0.295	-0.156
BW	0.601	0.418	-1.734
BH	-0.520	0.891	-1.010
EYE R	0.455	1.057	0.452
EYE L	0.154	-0.593	-0.051
NER	-0.088	-0.244	0.102
NEL	-0.814	-0.959	-0.698
NSR	-0.147	0.256	-0.254
NSL	0.362	-0.712	0.484
WBN	-0.118	0.495	-0.007
WBE	-0.030	0.688	0.136
PW	1.491	-1.255	1.598

Table 12. Mutation rates estimated by gene trees and species trees

	A. piscivorus Complex		A. contortrix Complex	
Locus	Gene Tree	Species Tree	Gene Tree	Species Tree
Anon A	___	___	0.0058 (0.0031-0.0088)	0.0059 (0.0032-0.0013)
Anon 11	0.0026 (0.0013-0.0040)	0.0025 (0.0013-0.0037)	0.0026 (0.0013-0.0040)	0.0025 (0.0013-0.0037)
Anon 51	___	___	0.0037 (0.0021-0.0054)	0.0036 (0.0022-0.0054)
AHR	0.0009 (0.0005-0.0013)	0.0012 (0.0007-0.0017)	___	___
Cytb	0.0176 (0.0060-0.0305)	0.0393 (0.0064-0.0855)	0.0146 (0.0057-0.0234)	0.0444 (0.0076-0.1053)
NT3	0.0022 (0.0013-0.0031)	0.0018 (0.0011-0.0026)	0.0022 (0.0013-0.0031)	0.0018 (0.0011-0.0026)
SPTBN1	0.0007 (0.0003-0.0010)	0.0013 (0.0009-0.0018)	0.0006 (0.0001-0.0010)	0.0013 (0.0009-0.0018)
VIM INTRON5	0.0017 (0.0011-0.0024)	0.0014 (0.0009-0.0019)	___	___

Mutation rate estimates, expressed in mutations/site/my, provided by Bayesian Inference in the program BEAST v1.6.1 (Drummond and Rambaut 2006; Drummond et al. 2005) with the *BEAST algorithm implemented to infer species trees (Heled and Drummond 2010). The 95% HPD are shown in parentheses.

Table 13. Sequence summaries for 10 loci used in divergence dating

Locus	N Concatenated Tree	N Species Tree	Sites (bp)	Variable Sites (bp)	Gaps	Pariwise identity %
Anon A	22	10	194	53	4	92.2
Anon 11	27	15	306	48	3	96.5
Anon 51	23	13	322	72	3	92.2
Anon 63	21	13	438	47	3	98.0
AHR	34	21	391	53	0	98.2
Cyt b	38	23	1107	770	0	81.5
$NA_v1.4$	22	16	470	52	16	97.1
NT3	27	16	434	64	1	97.2
SPTBN1	19	12	938	86	10	97.4
VimIntron5	26	16	566	94	17	95.2

The number of taxa sequenced for each locus for a total of 40 taxa for the concatenated analysis and 23 taxa for the species tree. Values for the size of each locus, the number of variable sites, the number of gaps, and the pairwise sequence identity observed between 40 Viperid taxa.

Table 14. Divergence dates for key nodes estimated by the mtDNA gene tree, concatenated dataset and species tree inference.

Node	Taxon	Cytb	Concatenated	Species Tree
1	Agkistrodon	8.89mya (6.33-10.80mya)	9.77mya (7.87-11.73mya)	6.55mya (5.03-8.34mya)
2	Crotalus/Agkistrodon	11.54mya (7.52-16.51mya)	9.77mya (7.87-11.73mya)	5.94mya (4.53-7.47mya)
3	A. contortrix complex	5.28mya (3.44-7.27mya)	3.75mya (2.68-4.79mya)	2.23mya (1.35-3.16mya)
4	A. laticinctus/A. contortrix	0.99mya (0.50-1.59mya)	1.25mya (0.77-1.83mya)	0.61mya (0.18-1.09mya)
5	A. piscivorus complex	3.51mya (2.20-4.98mya)	2.82mya (1.97-3.73mya)	1.54mya (0.87-2.32mya)
6	A. piscivorus/A. conanti	1.66mya (0.79-2.68mya)	1.32mya (0.73-1.89mya)	0.54mya (0.14-1.04mya)

Mean divergence time inferred for each lineage and outgroups in the program Beast v1.6.1 (Drummond and Rambaut 2003; Drummond et al. 2005). The 95% higher posterior density around each estimate is shown in parentheses. Divergence estimates for the MRCA for Crotalus and Agkistrodonis shown for comparison to divergence estimates in Table 2 and reflects differences in the placement of Agkistrodonwithin Crotalinae by the three analyses.

Table 15a. Tests of nested models for *A. contortrix* and *A. laticinctus* with priors on mutation rates estimated by gene trees.

Model (Θ)	log(p'($\widehat{\Theta}$/*X*))	-2$\widehat{\Lambda}$	*P*	df	AIC	ΔAIC
$\Theta_1\ \Theta_2\ \Theta_A\ m_1\ m_2$	-3.153	-	-	1	16.307	2.317
$\Theta_1\ \Theta_2\ \Theta_A\ m_1{=}m_2$	-3.853	1.399	0.2367*	1	15.706	1.716
$\Theta_1\ \Theta_2\ \Theta_A\ m_1\ m_2{=}0$	-11.188	16.070	0.0001	1†	30.377	16.387
$\Theta_1\ \Theta_2\ \Theta_A\ m_1{=}0\ m_2$	-24.961	43.615	<0.0000	1†	57.922	43.932
$\Theta_1\ \Theta_2\ \Theta_A\ m_1{=}0\ m_2{=}0$	-100.048	193.789	<0.0000	2†	206.096	192.106
$\Theta_1{=}\Theta_2\ \Theta_A\ m_1\ m_2$	-3.797	1.288	0.2564*	1	15.595	1.605
$\Theta_1{=}\Theta_2{=}\Theta_A\ m_1\ m_2$	-6.999	7.691	0.0214	2	19.998	6.008
$\Theta_1{=}\Theta_2\ \Theta_A\ m_1{=}m_2$	-3.995	1.683	0.4311*	2	13.990	-
$\Theta_1{=}\Theta_2\ \Theta_A\ m_1{=}0\ m_2{=}0$	-100.049	193.789	<0.0000	3†	204.097	190.106
$\Theta_1{=}\Theta_2{=}\Theta_A\ m_1{=}m_2$	-7.418	8.529	0.0363	3	18.836	4.846
$\Theta_1{=}\Theta_2{=}\Theta_A\ m_1{=}0\ m_2{=}0$	-126.720	247.123	<0.0000	4†	255.439	241.449
$\Theta_2\ \Theta_1{=}\Theta_A\ m_1\ m_2$	-5.597	4.887	0.0271	1	19.195	5.204
$\Theta_2\ \Theta_1{=}\Theta_A\ m_1{=}m_2$	-5.880	5.453	0.0654*	2	17.760	3.770
$\Theta_2\ \Theta_1{=}\Theta_A\ m_1{=}0\ m_2{=}0$	-125.561	244.814	<0.0000	3†	255.121	241.131
$\Theta_1\ \Theta_2{=}\Theta_A\ m_1\ m_2$	-5.610	4.914	0.0266	1	19.221	5.231
$\Theta_1\ \Theta_2{=}\Theta_A\ m_1{=}m_2$	-6.009	5.711	0.0575*	2	18.018	4.028
$\Theta_1\ \Theta_2{=}\Theta_A\ m1{=}0\ m_2{=}0$	-112.692	219.076	<0.0000	3†	229.383	215.393

*The model was not rejected by the 2LLR test in favor of the full model $\Theta_1\ \Theta_2\ \Theta_A\ m_1\ m_2$ with p<0.05

†The resulting chi square distribution is a mixture.

Table 15b. Tests of nested models for *A. contortrix* and *A. laticinctus* with priors on mutation rates estimated by the species tree.

Model (Θ)	log(p'($\widehat{\Theta}$/*X*))	-2$\widehat{\Lambda}$	*P*	df	AIC	ΔAIC
$\Theta_1\ \Theta_2\ \Theta_A\ m_1\ m_2$	-2.929	-	-	1	15.859	1.641
$\Theta_1\ \Theta_2\ \Theta_A\ m_1{=}m_2$	-3.799	1.738	0.1874*	1	15.598	1.380
$\Theta_1\ \Theta_2\ \Theta_A\ m_1\ m_2{=}0$	-12.359	18.858	<0.0000	1†	32.718	18.499
$\Theta_1\ \Theta_2\ \Theta_A\ m_1{=}0\ m_2$	-18.858	31.857	<0.0000	1†	45.716	31.498
$\Theta_1\ \Theta_2\ \Theta_A\ m_1{=}0\ m_2{=}0$	-100.893	195.926	<0.0000	2†	207.786	193.568
$\Theta_1{=}\Theta_2\ \Theta_A\ m_1\ m_2$	-3.709	1.560	0.2117*	1	15.419	1.201
$\Theta_1{=}\Theta_2{=}\Theta_A\ m_1\ m_2$	-7.394	8.929	0.0115	2	20.788	6.570
$\Theta_1{=}\Theta_2\ \Theta_A\ m_1{=}m_2$	-4.109	2.358	0.3076*	2	14.218	-
$\Theta_1{=}\Theta_2\ \Theta_A\ m_1{=}0\ m_2{=}0$	-101.212	196.563	<0.0000	3†	206.423	192.205
$\Theta_1{=}\Theta_2{=}\Theta_A\ m_1{=}m_2$	-7.498	9.137	0.0275	3	18.996	4.778
$\Theta_1{=}\Theta_2{=}\Theta_A\ m_1{=}0\ m_2{=}0$	-127.883	249.906	<0.0000	4†	257.765	243.547
$\Theta_2\ \Theta_1{=}\Theta_A\ m_1\ m_2$	-5.748	5.638	0.0176	1	19.497	5.279
$\Theta_2\ \Theta_1{=}\Theta_A\ m_1{=}m_2$	-6.288	6.716	0.0348	2	18.576	4.358
$\Theta_2\ \Theta_1{=}\Theta_A\ m_1{=}0\ m_2{=}0$	-126.724	247.588	<0.0000	3†	257.447	243.229
$\Theta_1\ \Theta_2{=}\Theta_A\ m_1\ m_2$	-6.625	7.391	0.0066	1	21.251	7.033
$\Theta_1\ \Theta_2{=}\Theta_A\ m_1{=}m_2$	-6.900	7.942	0.0189	2	19.801	5.583
$\Theta_1\ \Theta_2{=}\Theta_A\ m1{=}0\ m_2{=}0$	-113.855	221.850	<0.0000	3†	231.709	217.491

*The model was not rejected by the 2LLR test in favor of the full model $\Theta_1\ \Theta_2\ \Theta_A\ m_1\ m_2$ with p<0.05

†The resulting chi square distribution is a mixture.

Table 15c. Tests of nested models for *A. piscivorus* and *A. conanti* with priors on mutation rates estimated by gene trees.

Model (Θ)	log(p'($\widehat{\Theta}$/*X*))	-2$\widehat{\Lambda}$	*P*	df	AIC	ΔAIC
$\Theta_1\ \Theta_2\ \Theta_A\ m_1\ m_2$	-2.284	-	-	1	14.569	4.792
$\Theta_1\ \Theta_2\ \Theta_A\ m_1{=}m_2$	-2.427	0.285	0.5934*	1	12.855	3.078
$\Theta_1\ \Theta_2\ \Theta_A\ m_1\ m_2{=}0$	-18.023	31.477	<0.0000	1†	44.047	34.270
$\Theta_1\ \Theta_2\ \Theta_A\ m_1{=}0\ m_2$	-17.277	29.985	<0.0000	1†	42.554	32.777
$\Theta_1\ \Theta_2\ \Theta_A\ m_1{=}0\ m_2{=}0$	-108.492	212.414	<0.0000	2†	222.984	213.207
$\Theta_1{=}\Theta_2\ \Theta_A\ m_1\ m_2$	-1.657	-1.254	1.0000*	1	11.315	1.538
$\Theta_1{=}\Theta_2{=}\Theta_A\ m_1\ m_2$	-3.747	2.925	0.2317*	2	13.494	3.717
$\Theta_1{=}\Theta_2\ \Theta_A\ m_1{=}m_2$	-1.888	-0.792	1.0000*	2	9.777	-
$\Theta_1{=}\Theta_2\ \Theta_A\ m_1{=}0\ m_2{=}0$	-150.208	295.845	<0.0000	3†	304.415	294.638
$\Theta_1{=}\Theta_2{=}\Theta_A\ m_1{=}m_2$	-4.699	4.828	0.1848*	3	13.398	3.621
$\Theta_1{=}\Theta_2{=}\Theta_A\ m_1{=}0\ m_2{=}0$	-175.865	347.159	<0.0000	4†	353.729	343.952
$\Theta_2\ \Theta_1{=}\Theta_A\ m_1\ m_2$	-3.247	1.924	0.1654*	1	14.494	4.717
$\Theta_2\ \Theta_1{=}\Theta_A\ m_1{=}m_2$	-3.449	2.329	0.3121*	2	12.899	3.122
$\Theta_2\ \Theta_1{=}\Theta_A\ m_1{=}0\ m_2{=}0$	-131.785	258.999	<0.0000	3†	267.569	257.792
$\Theta_1\ \Theta_2{=}\Theta_A\ m_1\ m_2$	-2.947	1.325	0.2497*	1	13.894	4.117
$\Theta_1\ \Theta_2{=}\Theta_A\ m_1{=}m_2$	-3.284	2.000	0.3679*	2	12.569	2.792
$\Theta_1\ \Theta_2{=}\Theta_A\ m1{=}0\ m_2{=}0$	-145.507	286.443	<0.0000	3†	295.013	285.236

*The model was not rejected by the 2LLR test in favor of the full model $\Theta_1\ \Theta_2\ \Theta_A\ m_1\ m_2$ with p<0.05

†The resulting chi square distribution is a mixture.

Table 15d. Tests of nested models for *A. piscivorus* and *A. conanti* with priors on mutation rates estimated by the species tree.

Model (Θ)	log(p'($\widehat{\Theta}$/*X*))	-2$\widehat{\Lambda}$	*P*	df	AIC	ΔAIC
$\Theta_1\ \Theta_2\ \Theta_A\ m_1\ m_2$	-2.356	-	-	1	14.712	1.861
$\Theta_1\ \Theta_2\ \Theta_A\ m_1{=}m_2$	-2.496	0.280	0.5967*	1	12.993	0.142
$\Theta_1\ \Theta_2\ \Theta_A\ m_1\ m_2{=}0$	-2.667	0.621	0.8614*	1†	13.334	0.483
$\Theta_1\ \Theta_2\ \Theta_A\ m_1{=}0\ m_2$	-18.036	31.360	<0.0000	1†	44.072	31.221
$\Theta_1\ \Theta_2\ \Theta_A\ m_1{=}0\ m_2{=}0$	-112.622	220.531	<0.0000	2†	231.243	218.392
$\Theta_1{=}\Theta_2\ \Theta_A\ m_1\ m_2$	-3.765	2.819	0.0932*	1	15.531	2.680
$\Theta_1{=}\Theta_2{=}\Theta_A\ m_1\ m_2$	-4.299	3.886	0.1433*	2	14.598	1.747
$\Theta_1{=}\Theta_2\ \Theta_A\ m_1{=}m_2$	-3.972	3.231	0.1988*	2	13.944	1.093
$\Theta_1{=}\Theta_2\ \Theta_A\ m_1{=}0\ m_2{=}0$	-115.544	226.375	<0.0000	3†	235.087	222.236
$\Theta_1{=}\Theta_2{=}\Theta_A\ m_1{=}m_2$	-4.555	4.399	0.2215*	3	13.111	0.260
$\Theta_1{=}\Theta_2{=}\Theta_A\ m_1{=}0\ m_2{=}0$	-166.336	327.959	<0.0000	4†	334.672	321.821
$\Theta_2\ \Theta_1{=}\Theta_A\ m_1\ m_2$	-3.418	2.124	0.1450*	1	14.836	1.985
$\Theta_2\ \Theta_1{=}\Theta_A\ m_1{=}m_2$	-3.425	2.138	0.3434*	2	12.851	-
$\Theta_2\ \Theta_1{=}\Theta_A\ m_1{=}0\ m_2{=}0$	-128.833	252.953	<0.0000	3†	261.665	248.814
$\Theta_1\ \Theta_2{=}\Theta_A\ m_1\ m_2$	-3.458	2.204	0.1377*	1	14.916	2.065
$\Theta_1\ \Theta_2{=}\Theta_A\ m_1{=}m_2$	-3.625	2.537	0.2813*	2	13.520	0.399
$\Theta_1\ \Theta_2{=}\Theta_A\ m1{=}0\ m_2{=}0$	-152.223	299.734	<0.0000	3†	308.446	295.595

*The model was not rejected by the 2LLR test in favor of the full model $\Theta_1\ \Theta_2\ \Theta_A\ m_1\ m_2$ with p<0.05

†The resulting chi square distribution is a mixture.

Table 16. Population parameter estimates in IMa with mutation rate priors provided by gene trees and species tree for the A. contortrix complex and A. piscivorus complex.

Model	N_1	N_2	N_A	m_1	m_2	t
A. contortrix X A. laticinctus						
Gene Tree Mutation Rates						
MLE	83,369	57,676	509,023	0.0000065	0.0000164	345,099
Lower 95% HPD	35,503	19,663	204,852	0.0000016	0.0000040	55,000
Upper 95% HPD	144,743	115,794	1,111,521	0.0000170	0.0000414	1,095,000
Species Tree Mutation Rates						
MLE	83,361	58,760	541,271	0.0000066	0.0000160	419,568
Lower 95% HPD	37,141	24,032	320,287	0.0000016	0.0000041	64,999
Upper 95% HPD	146,382	115,764	1,242,609	0.0000170	0.0000398	1,604,999
A. piscivorus X A. conanti						
Gene Tree Mutation Rates						
MLE	31,438	16,860	1,773,268	0.0000092	0.0000142	1,419,769
Lower 95% HPD	8,308	4,315	124,960	0.0000021	0.0000021	104,999
Upper 95% HPD	65,471	38,267	5,043,614	0.0000282	0.0000452	7,674,999
Species Tree Mutation Rates						
MLE	33,584	17,348	1,826,407	0.0000086	0.0000133	1,699,145
Lower 95% HPD	9,219	4,438	129,702	0.0000022	0.0000021	115,000
Upper 95% HP	68,348	38,764	4,854,951	0.0000258	0.0000424	8,225,000

The units on population size estimates for extant species (N_1 and N_2) and the ancestral population (N_A) are individuals. m_1 and m_2 are the migration rate per generation per gene into species 1 and species 2 respectively. Time since splitting (t) is in number of years.

Figures

A. Northern Copperhead (*A. c. mokasen*)

B. Southern Copperhead (A. c. contortrix)

C. Osage Copperhead (*A. c. phaeogaster*)

D. Broad-banded Copperhead (*A. c. laticinctus*)

E. Trans-Pecos Copperhead (*A. c. pictigaster*)

F. Eastern Cottonmouth (*A. p. piscivorus*)

G. Florida Cottonmouth (*A. p. conanti*)

H. Western Cottonmouth (*A. p.leucostoma*)

Fig 1

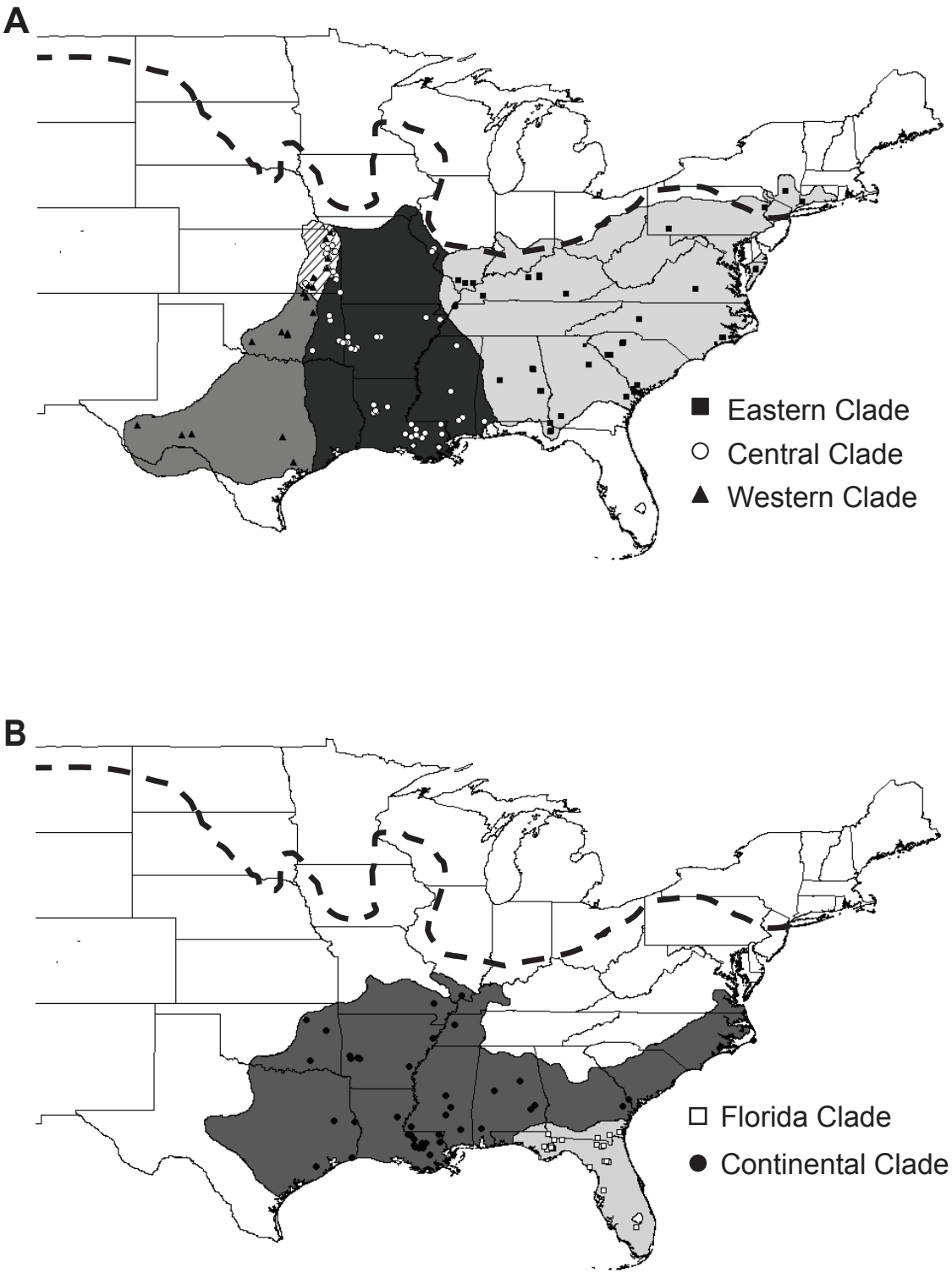

A
Eastern Clade
Central Clade
Western Clade
B
Florida Clade
Continental Clade

Fig 2

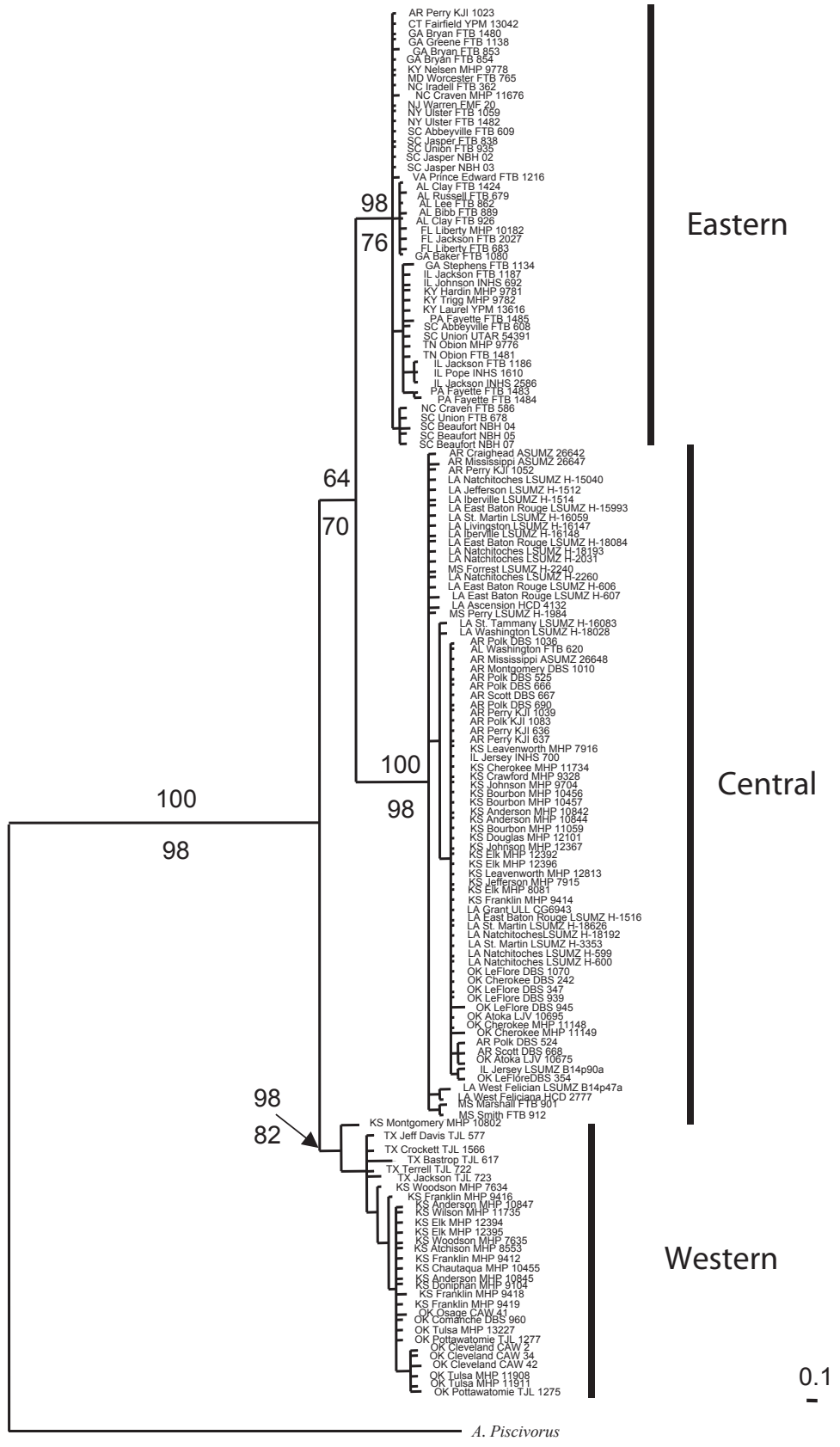

AR Perry KJI 1023
CT Fairfield YPM 13042
GA Bryan FTB 1480
GA Greene FTB 1138
GA Bryan FTB 853
GA Bryan FTB 854
KY Nelsen MHP 9778
MD Worcester FTB 765
NC Iradell FTB 362
NC Craven MHP 11676
NJ Warren EMF 20
NY Ulster FTB 1059
NY Ulster FTB 1482
SC Abbeyville FTB 609
SC Jasper FTB 838
SC Union FTB 935
SC Jasper NBH 02
SC Jasper NBH 03
VA Prince Edward FTB 1216
AL Clay FTB 1424
AL Russell FTB 679
AL Lee FTB 862
AL Bibb FTB 889
AL Clay FTB 926
FL Liberty MHP 10182
FL Jackson FTB 2027
FL Liberty FTB 683
GA Baker FTB 1080
GA Stephens FTB 1134
IL Jackson FTB 1187
IL Johnson INHS 692
KY Hardin MHP 9781
KY Trigg MHP 9782
KY Laurel YPM 13616
PA Fayette FTB 1485
SC Abbeyville FTB 608
SC Union UTAR 54391
TN Obion MHP 9776
TN Obion FTB 1481
IL Jackson FTB 1186
IL Pope INHS 1610
IL Jackson INHS 2586
PA Fayette FTB 1483
PA Fayette FTB 1484
NC Craven FTB 586
SC Union FTB 678
SC Beaufort NBH 04
SC Beaufort NBH 05
SC Beaufort NBH 07
AR Craighead ASUMZ 26642
AR Mississippi ASUMZ 26647
AR Perry KJI 1052
LA Natchitoches LSUMZ H-15040
LA Jefferson LSUMZ H-1512
LA Iberville LSUMZ H-1514
LA East Baton Rouge LSUMZ H-15993
LA St. Martin LSUMZ H-16059
LA Livingston LSUMZ H-16147
LA Iberville LSUMZ H-16148
LA East Baton Rouge LSUMZ H-18084
LA Natchitoches LSUMZ H-18193
LA Natchitoches LSUMZ H-2031
MS Forrest LSUMZ H-2240
LA Natchitoches LSUMZ H-2260
LA East Baton Rouge LSUMZ H-606
LA East Baton Rouge LSUMZ H-607
LA Ascension HCD 4132
MS Perry LSUMZ H-1984
LA St. Tammany LSUMZ H-16083
LA Washington LSUMZ H-18028
AR Polk DBS 1036
AL Washington FTB 620
AR Mississippi ASUMZ 26648
AR Montgomery DBS 1010
AR Polk DBS 525
AR Polk DBS 666
AR Scott DBS 667
AR Polk DBS 690
AR Perry KJI 1039
AR Polk KJI 1083
AR Perry KJI 636
AR Perry KJI 637
KS Leavenworth MHP 7916
IL Jersey INHS 700
KS Cherokee MHP 11734
KS Crawford MHP 9328
KS Johnson MHP 9704
KS Bourbon MHP 10456
KS Bourbon MHP 10457
KS Anderson MHP 10842
KS Anderson MHP 10844
KS Bourbon MHP 11059
KS Douglas MHP 12101
KS Johnson MHP 12367
KS Elk MHP 12392
KS Elk MHP 12396
KS Leavenworth MHP 12813
KS Jefferson MHP 7915
KS Elk MHP 8081
KS Franklin MHP 9414
LA Grant ULL CG6943
LA East Baton Rouge LSUMZ H-1516
LA St. Martin LSUMZ H-18626
LA NatchitochesLSUMZ H-18192
LA St. Martin LSUMZ H-3353
LA Natchitoches LSUMZ H-599
LA Natchitoches LSUMZ H-600
OK LeFlore DBS 1070
OK Cherokee DBS 242
OK LeFlore DBS 347
OK LeFlore DBS 939
OK LeFlore DBS 945
OK Atoka LJV 10695
OK Cherokee MHP 11148
OK Cherokee MHP 11149
AR Polk DBS 524
AR Scott DBS 668
OK Atoka LJV 10675
IL Jersey LSUMZ B14p90a
OK LeFloreDBS 354
LA West Feliciana LSUMZ B14p47a
LA West Feliciana HCD 2777
MS Marshall FTB 901
MS Smith FTB 912
KS Montgomery MHP 10802
TX Jeff Davis TJL 577
TX Crockett TJL 1566
TX Bastrop TJL 617
TX Terrell TJL 722
TX Jackson TJL 723
KS Woodson MHP 7634
KS Franklin MHP 9416
KS Anderson MHP 10847
KS Wilson MHP 11735
KS Elk MHP 12394
KS Elk MHP 12395
KS Woodson MHP 7635
KS Atchison MHP 8553
KS Franklin MHP 9412
KS Chautaqua MHP 10455
KS Anderson MHP 10845
KS Doniphan MHP 9704
KS Franklin MHP 9418
KS Franklin MHP 9419
OK Osage CAW 41
OK Comanche DBS 960
OK Tulsa MHP 13227
OK Pottawatomie TJL 1277
OK Cleveland CAW 2
OK Cleveland CAW 34
OK Cleveland CAW 42
OK Tulsa MHP 11908
OK Tulsa MHP 11911
OK Pottawatomie TJL 1275
A. Piscivorus
98
76
64
70
100
98
100
98
98
82
Eastern
Central
Western
0.1

Fig 3

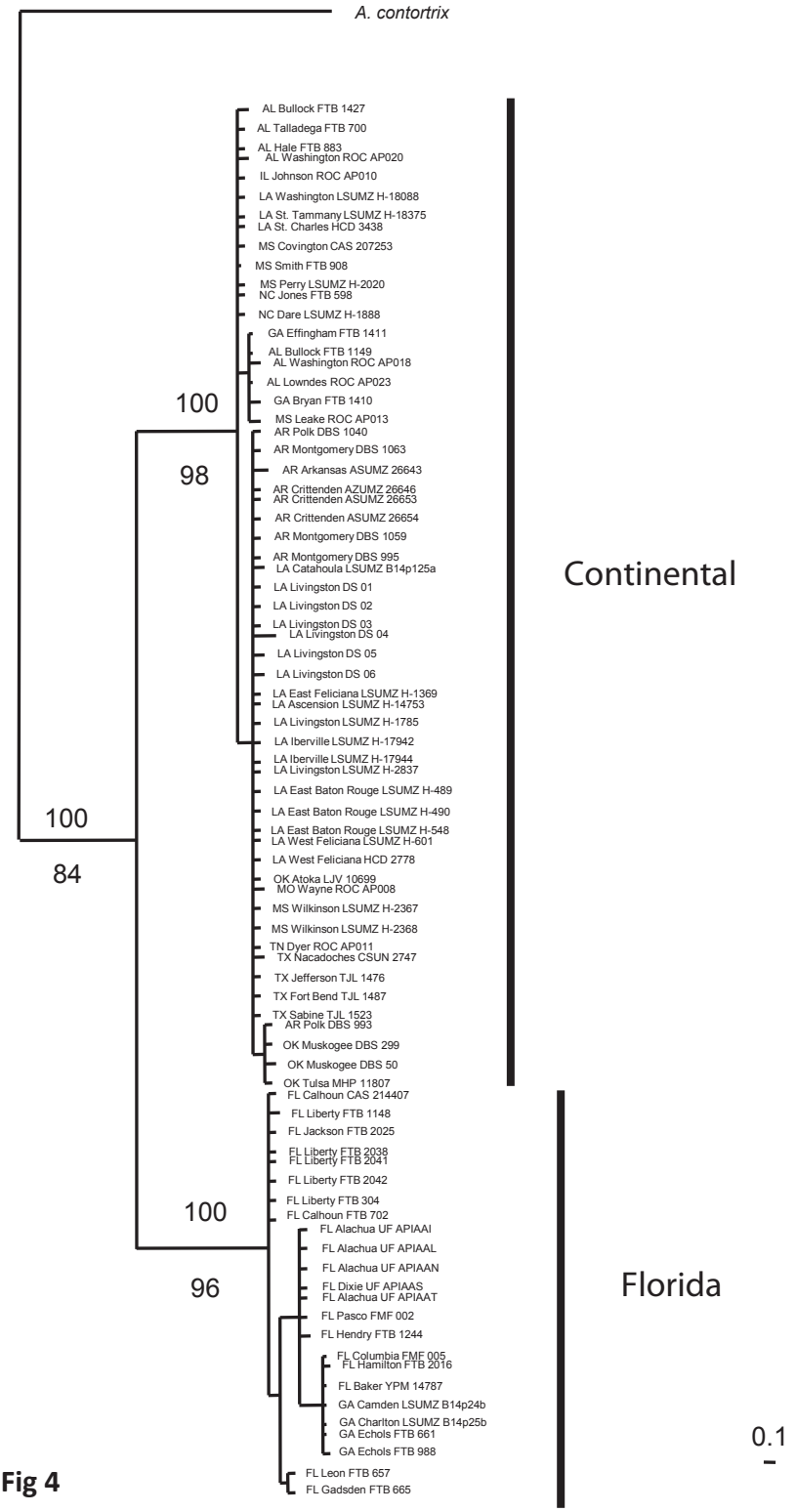

A. contortrix
AL Bullock FTB 1427
AL Talladega FTB 700
AL Hale FTB 883
AL Washington ROC AP020
IL Johnson ROC AP010
LA Washington LSUMZ H-18088
LA St. Tammany LSUMZ H-18375
LA St. Charles HCD 3438
MS Covington CAS 207253
MS Smith FTB 908
MS Perry LSUMZ H-2020
NC Jones FTB 598
NC Dare LSUMZ H-1888
GA Effingham FTB 1411
AL Bullock FTB 1149
AL Washington ROC AP018
AL Lowndes ROC AP023
GA Bryan FTB 1410
MS Leake ROC AP013
AR Polk DBS 1040
AR Montgomery DBS 1063
AR Arkansas ASUMZ 26643
AR Crittenden AZUMZ 26646
AR Crittenden ASUMZ 26653
AR Crittenden ASUMZ 26654
AR Montgomery DBS 1059
AR Montgomery DBS 995
LA Catahoula LSUMZ B14p125a
LA Livingston DS 01
LA Livingston DS 02
LA Livingston DS 03
LA Livingston DS 04
LA Livingston DS 05
LA Livingston DS 06
LA East Feliciana LSUMZ H-1369
LA Ascension LSUMZ H-14753
LA Livingston LSUMZ H-1785
LA Iberville LSUMZ H-17942
LA Iberville LSUMZ H-17944
LA Livingston LSUMZ H-2837
LA East Baton Rouge LSUMZ H-489
LA East Baton Rouge LSUMZ H-490
LA East Baton Rouge LSUMZ H-548
LA West Feliciana LSUMZ H-601
LA West Feliciana HCD 2778
OK Atoka LJV 10699
MO Wayne ROC AP008
MS Wilkinson LSUMZ H-2367
MS Wilkinson LSUMZ H-2368
TN Dyer ROC AP011
TX Nacadoches CSUN 2747
TX Jefferson TJL 1476
TX Fort Bend TJL 1487
TX Sabine TJL 1523
AR Polk DBS 993
OK Muskogee DBS 299
OK Muskogee DBS 50
OK Tulsa MHP 11807
FL Calhoun CAS 214407
FL Liberty FTB 1148
FL Jackson FTB 2025
FL Liberty FTB 2038
FL Liberty FTB 2041
FL Liberty FTB 2042
FL Liberty FTB 304
FL Calhoun FTB 702
FL Alachua UF APIAAI
FL Alachua UF APIAAL
FL Alachua UF APIAAN
FL Dixie UF APIAAS
FL Alachua UF APIAAT
FL Pasco FMF 002
FL Hendry FTB 1244
FL Columbia FMF 005
FL Hamilton FTB 2016
FL Baker YPM 14787
GA Camden LSUMZ B14p24b
GA Charlton LSUMZ B14p25b
GA Echols FTB 661
GA Echols FTB 988
FL Leon FTB 657
FL Gadsden FTB 665
100
98
100
84
100
96
Continental
Florida
0.1

Fig 4

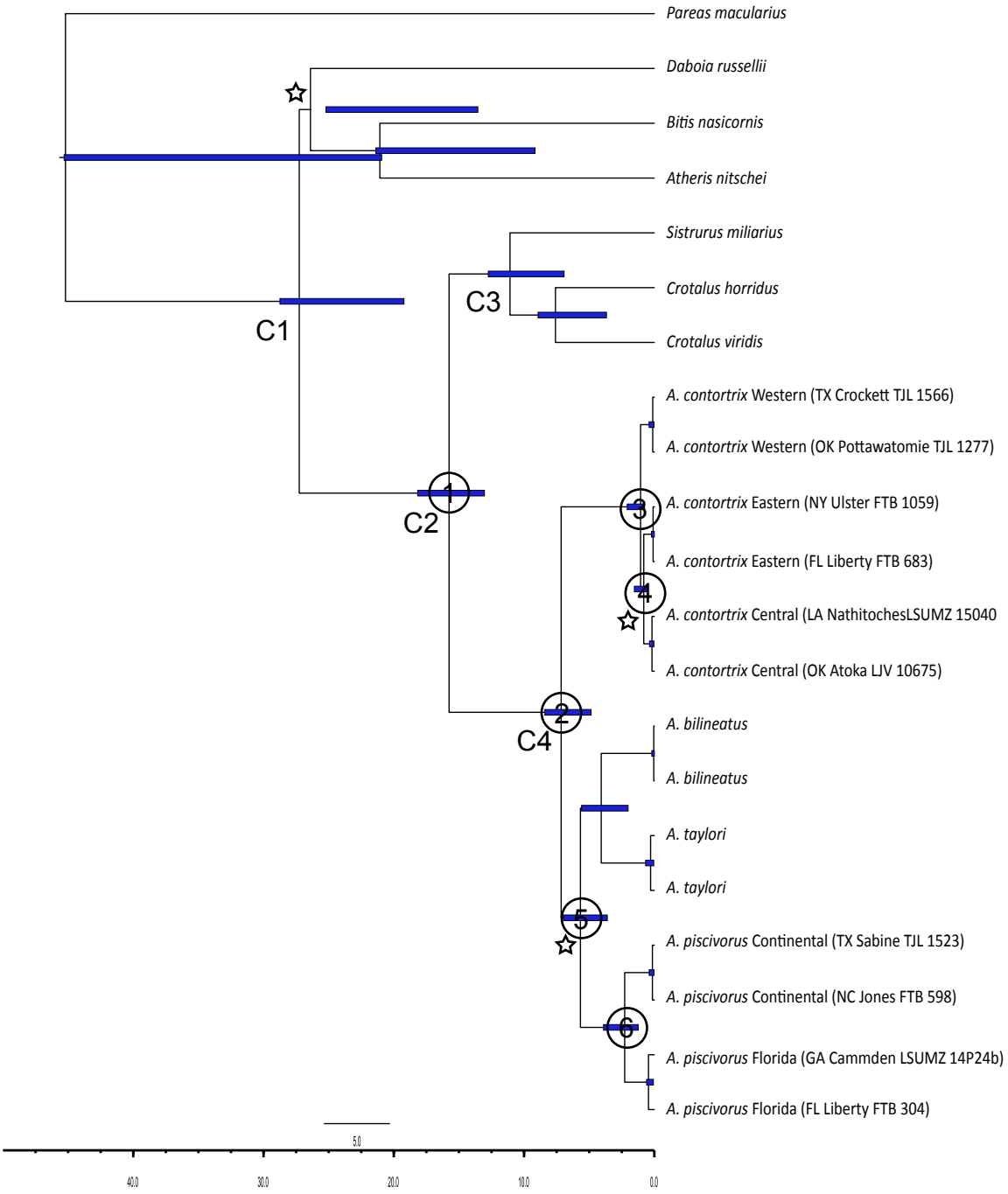

Pareas macularius
Daboia russellii
Bitis nasicornis
Atheris nitschei
Sistrurus miliarius
Crotalus horridus
Crotalus viridis
C1
C3
C2
C4
A. contortrix Western (TX Crockett TJL 1566)
A. contortrix Western (OK Pottawatomie TJL 1277)
A. contortrix Eastern (NY Ulster FTB 1059)
A. contortrix Eastern (FL Liberty FTB 683)
A. contortrix Central (LA NathitochesLSUMZ 15040
A. contortrix Central (OK Atoka LJV 10675)
A. bilineatus
A. bilineatus
A. taylori
A. taylori
A. piscivorus Continental (TX Sabine TJL 1523)
A. piscivorus Continental (NC Jones FTB 598)
A. piscivorus Florida (GA Cammden LSUMZ 14P24b)
A. piscivorus Florida (FL Liberty FTB 304)
1
2
3
4
5
6
5.0
40.0
30.0
20.0
10.0
0.0

Fig 5

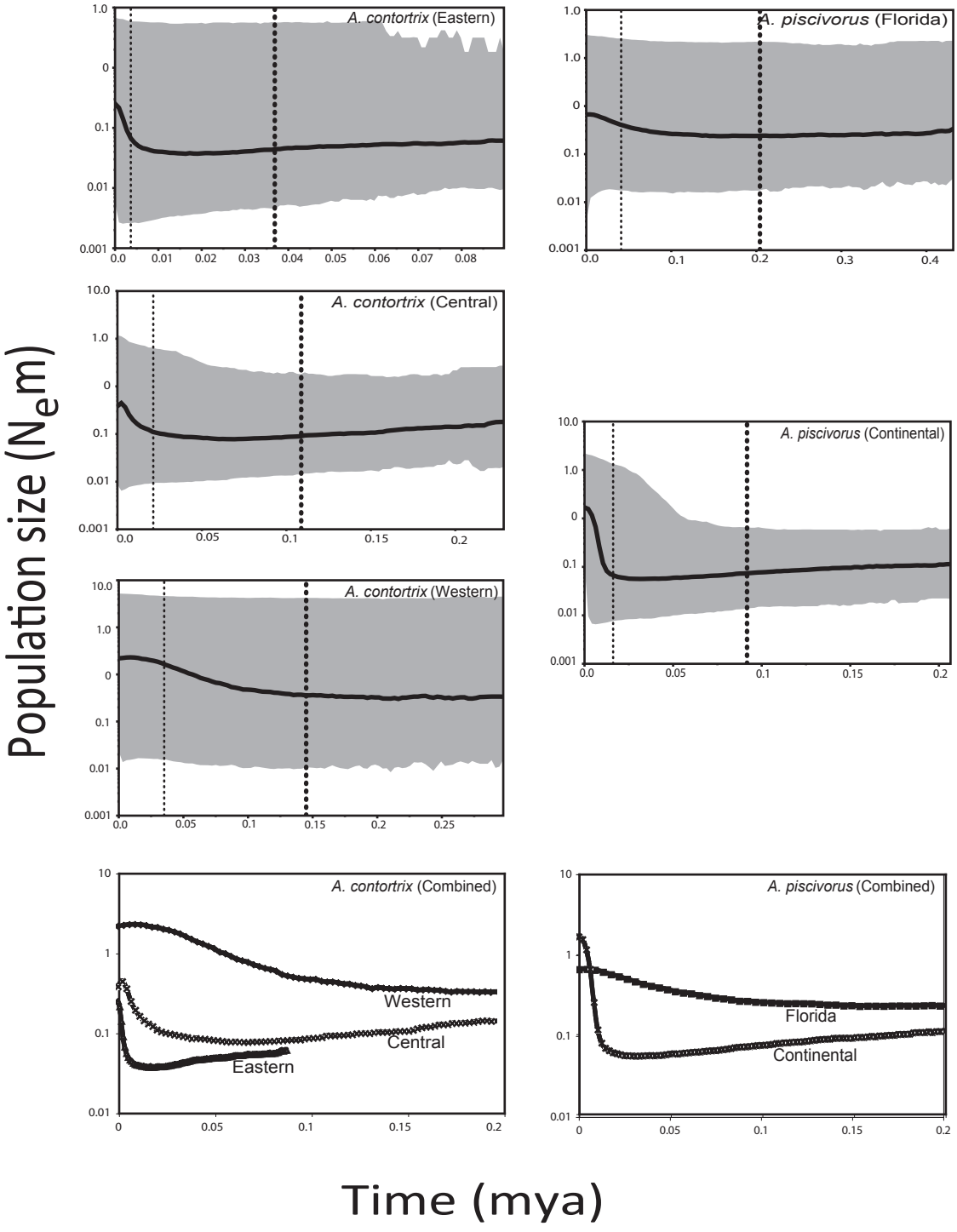

A. contortrix (Eastern)
A. piscivorus (Florida)
A. contortrix (Central)
A. piscivorus (Continental)
A. contortrix (Western)
A. contortrix (Combined)
A. piscivorus (Combined)
Western
Central
Eastern
Florida
Continental
Population size (NeM)
Time (mya)

Fig 6


A
East/Central Clade
West Clade
Hybrids
Continental Clade
East Clade
Central Clade
West Clade
1.0
1.0
0.0040
B
Florida Clade
Continental Clade
Hybrids

Fig 7

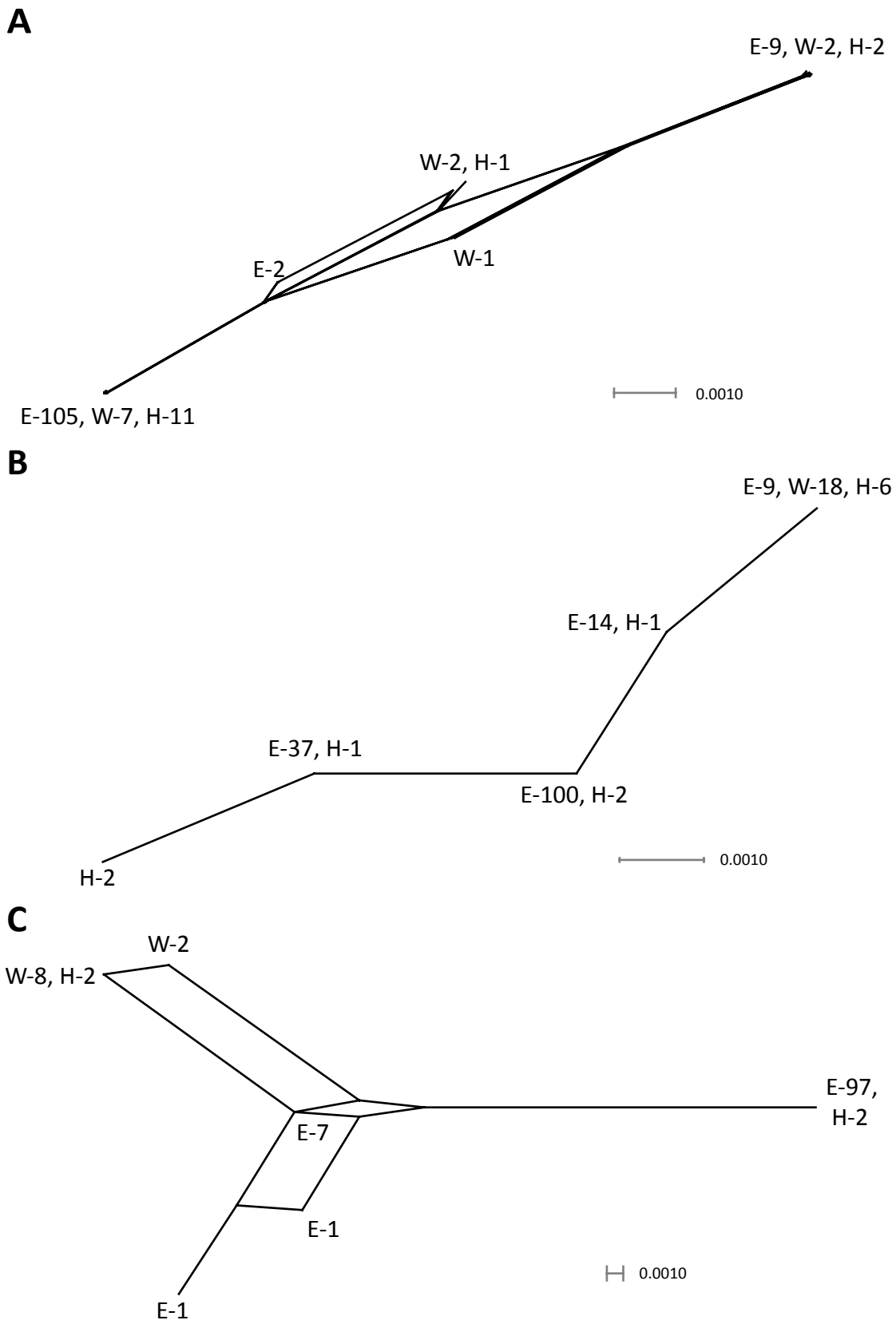
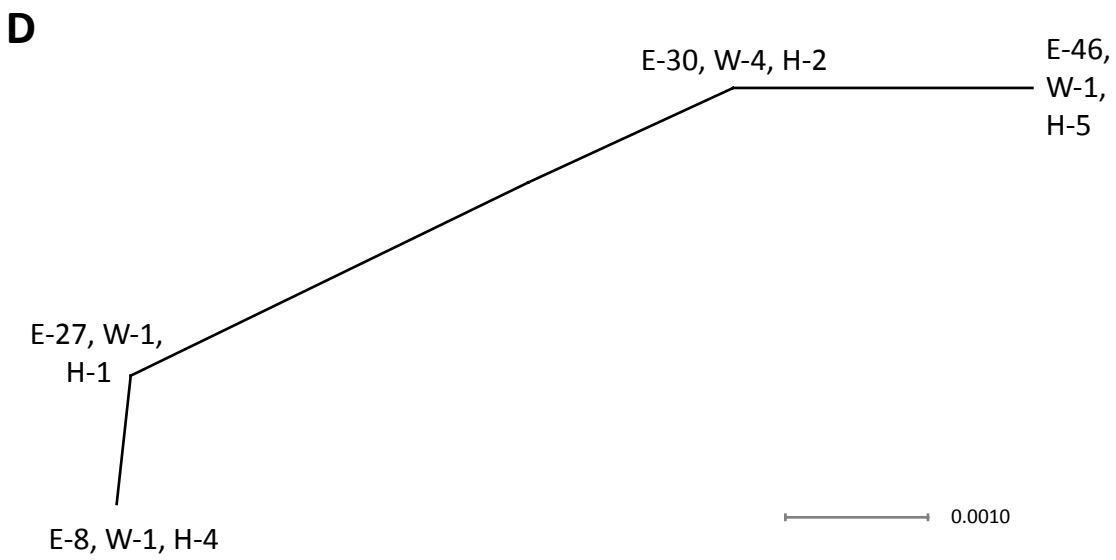

A
E-9, W-2, H-2
W-2, H-1
W-1
E-2
E-105, W-7, H-11
0.0010
B
E-9, W-18, H-6
E-14, H-1
E-37, H-1
E-100, H-2
H-2
0.0010
C
W-2
W-8, H-2
E-97,
H-2
E-7
E-1
E-1
0.0010

Fig 8


D
E-30, W-4, H-2
E-46,
W-1,
H-5
E-27, W-1,
H-1
E-8, W-1, H-4
0.0010

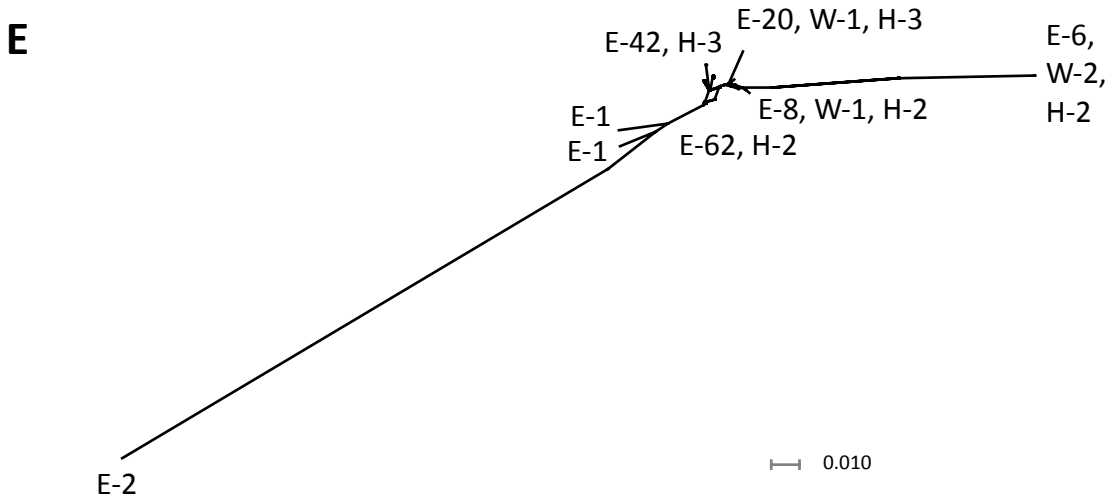

E
E-20, W-1, H-3
E-42, H-3
E-6,
W-2,
H-2
E-8, W-1, H-2
E-1
E-1
E-62, H-2
E-2
0.010

Fig 8 cont.

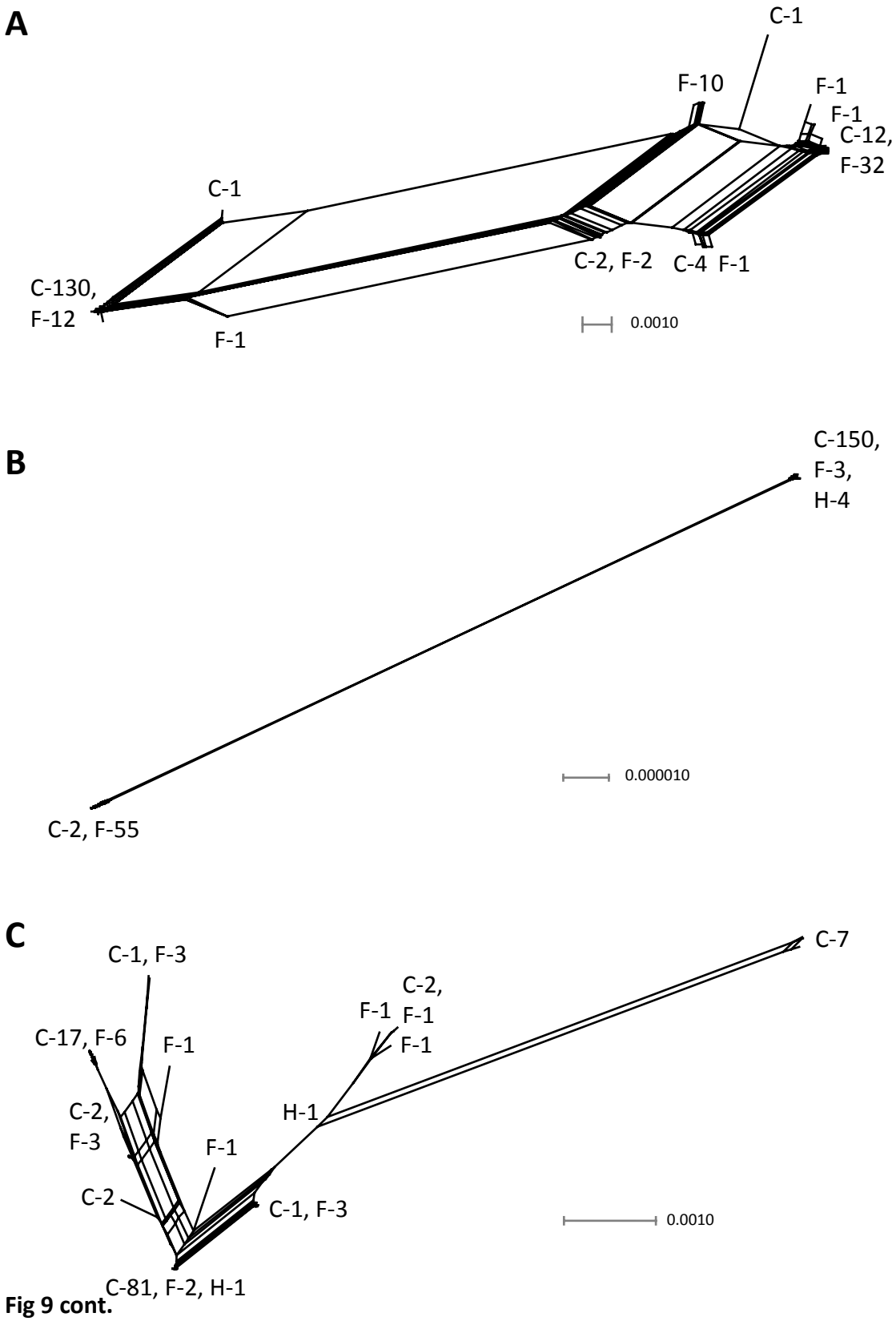
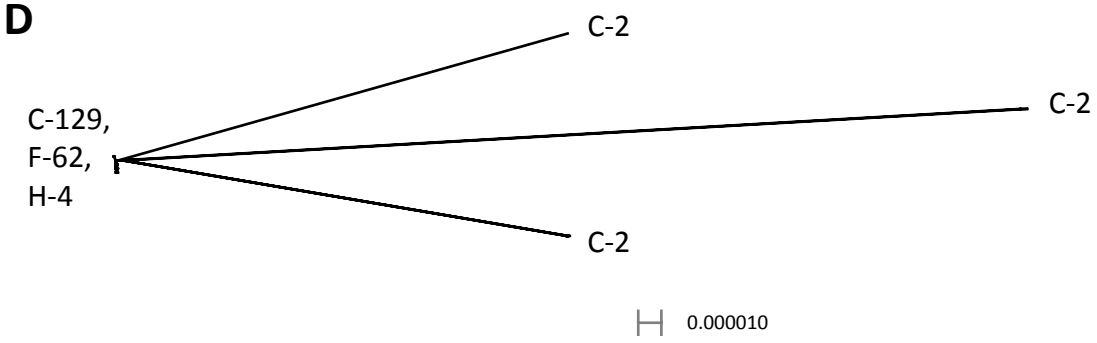

A
C-1
F-10
F-1
F-1
C-12,
F-32
C-1
C-130,
F-12
F-1
C-2, F-2
C-4
F-1
0.0010
B
C-150,
F-3,
H-4
0.000010
C-2, F-55
C
C-1, F-3
C-7
C-2,
F-1
F-1
F-1
C-17, F-6
F-1
C-2,
F-3
H-1
F-1
C-2
C-1, F-3
0.0010
C-81, F-2, H-1

Fig 9 cont.

D



E

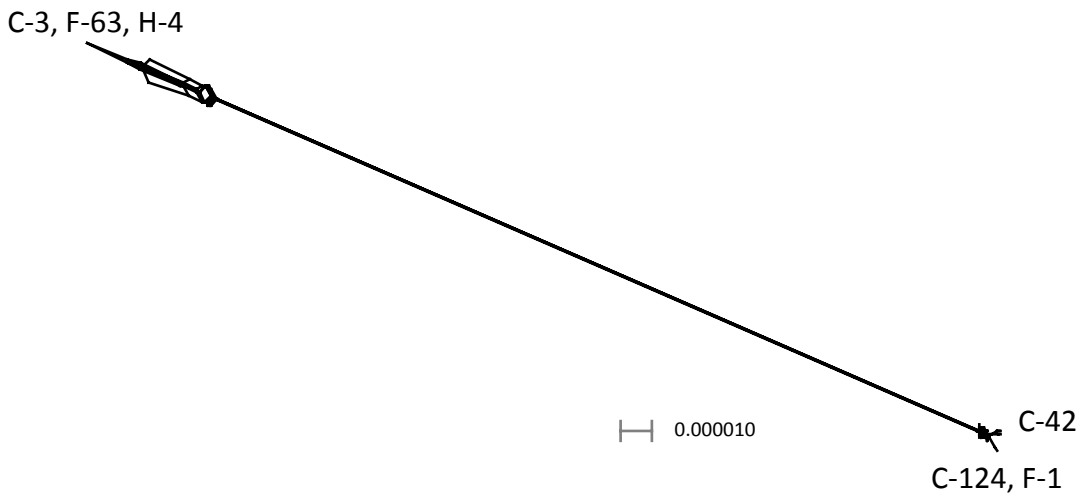


Fig 9 cont.

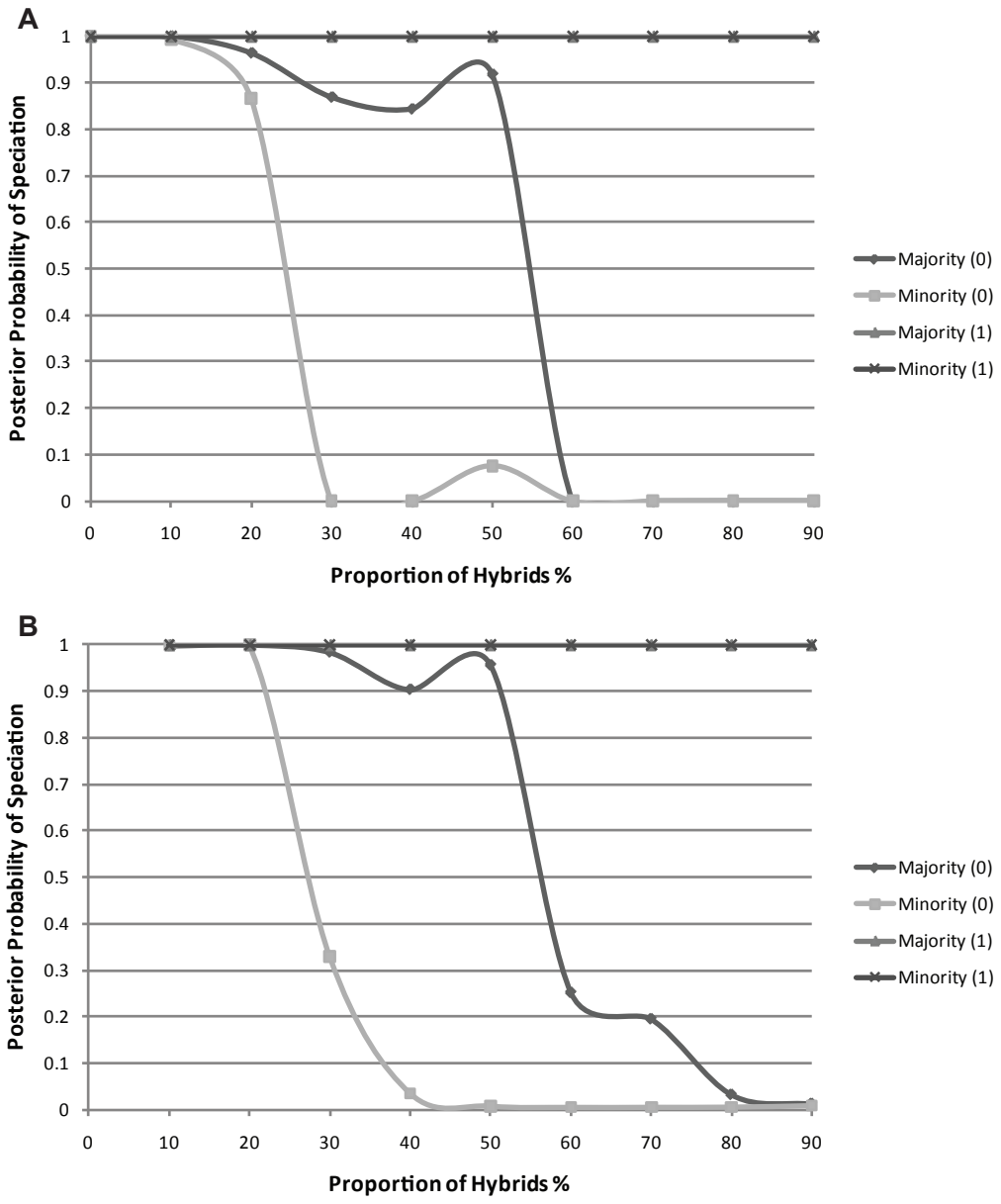
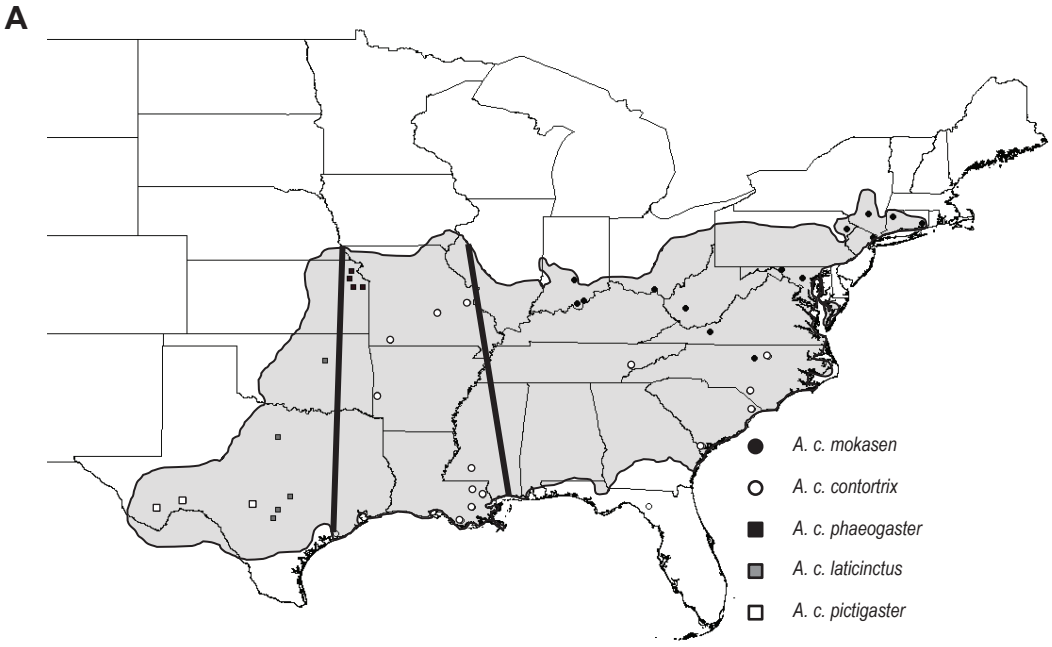

A
Posterior Probability of Speciation
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
10
20
30
40
50
60
70
80
90
Proportion of Hybrids %
Majority (0)
Minority (0)
Majority (1)
Minority (1)
B
Posterior Probability of Speciation
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
10
20
30
40
50
60
70
80
90
Proportion of Hybrids %
Majority (0)
Minority (0)
Majority (1)
Minority (1)

Fig 10


A
A. c. mokasen
A. c. contortrix
A. c. phaeogaster
A. c. laticinctus
A. c. pictigaster

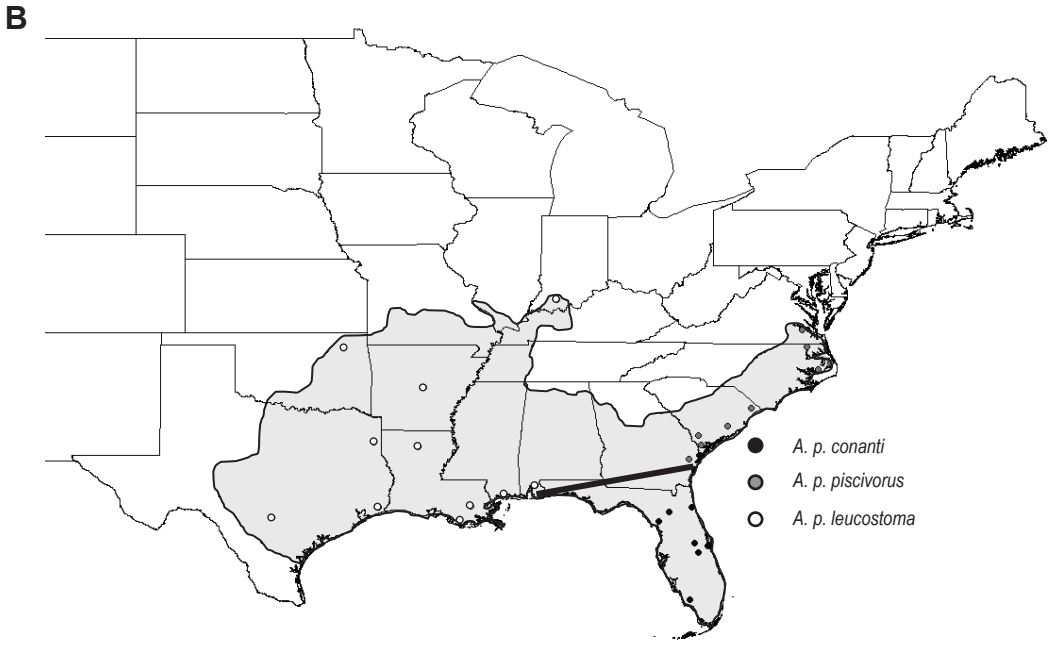

B
A. p. conanti
A. p. piscivorus
A. p. leucostoma

Fig 11

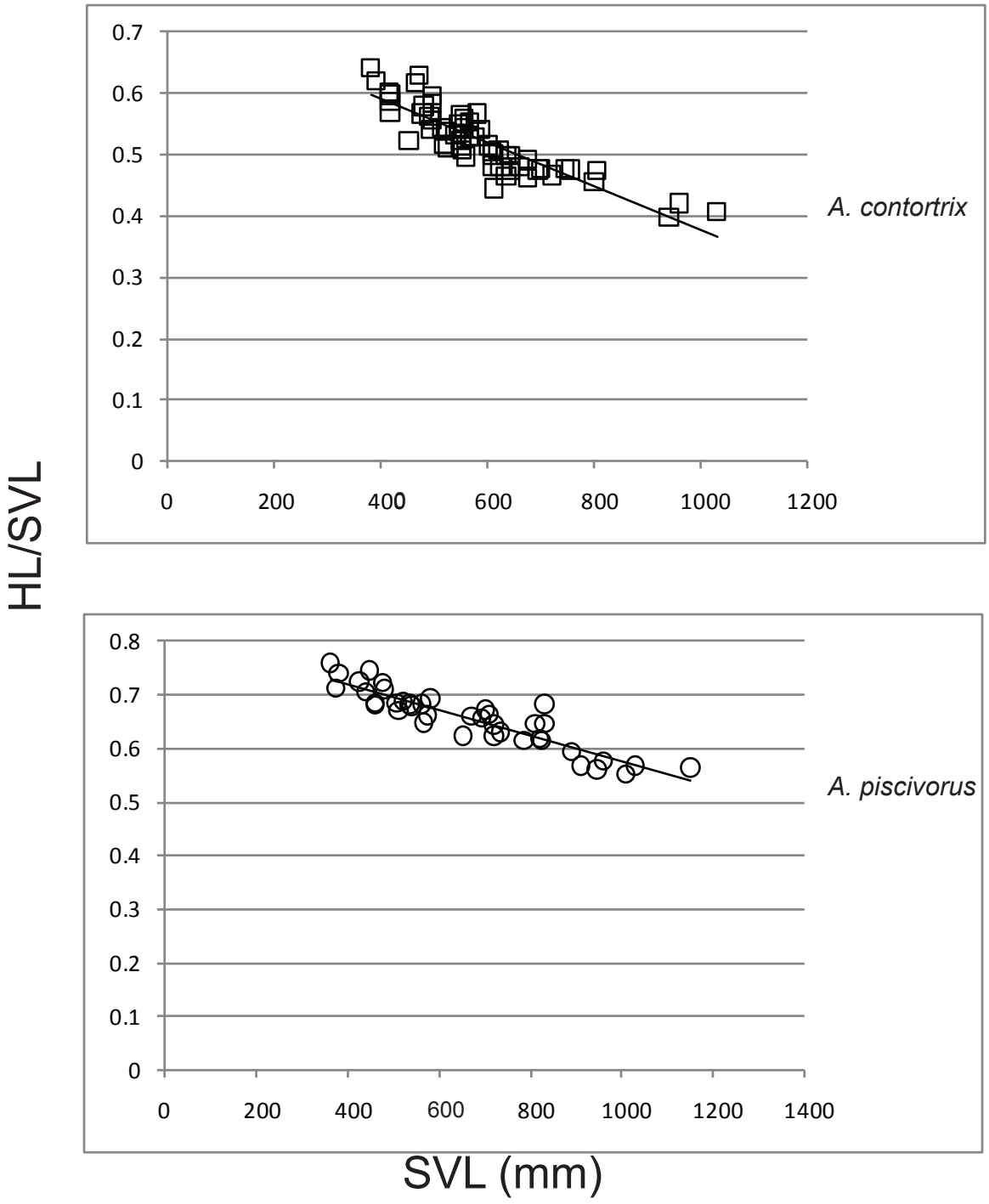

HL/SVL
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
200
400
600
800
1000
1200
A. contortrix
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
200
400
600
800
1000
1200
1400
A. piscivorus
SVL (mm)

Fig 12

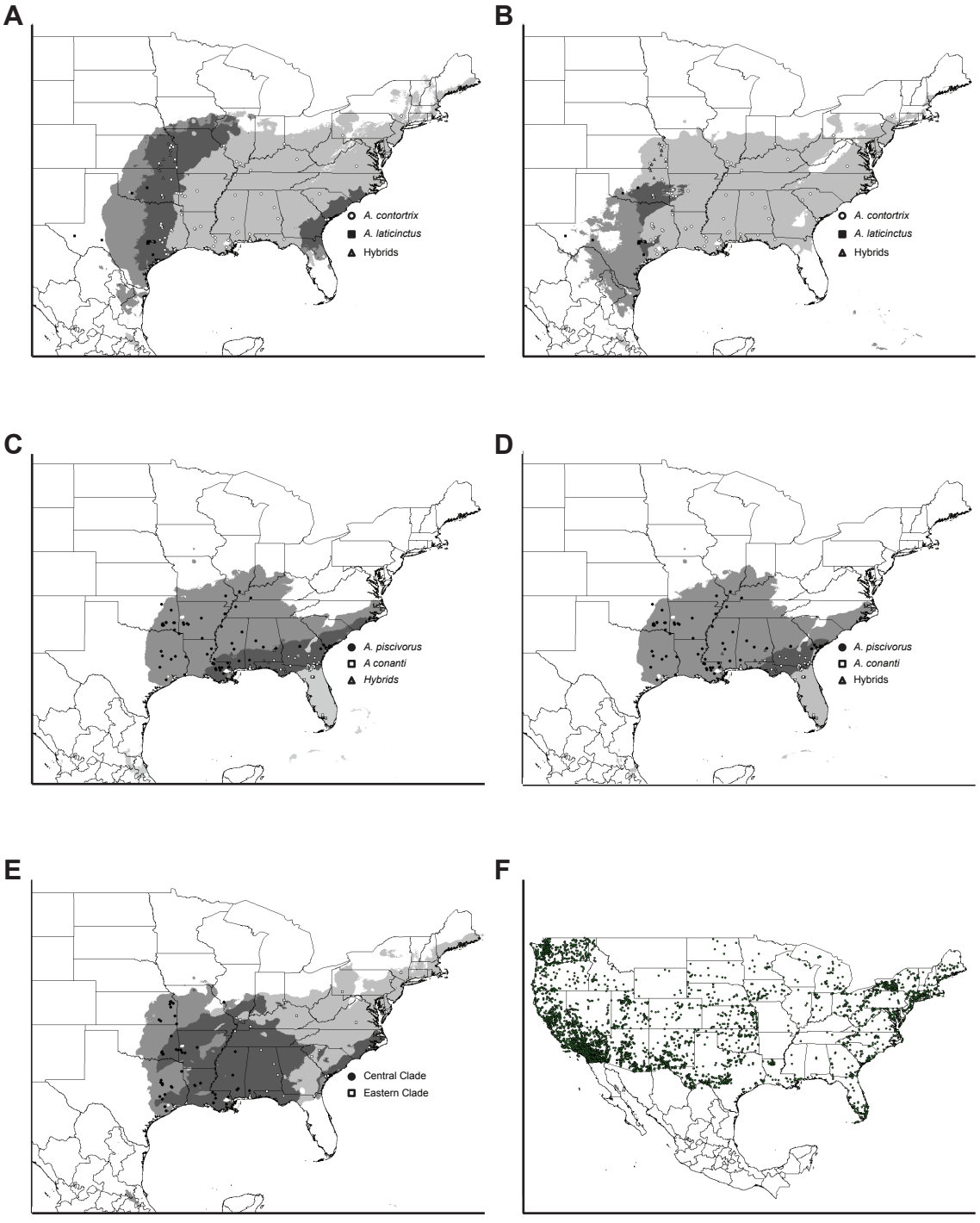

A
A. contortrix
A. laticinctus
Hybrids
B
A. contortrix
A. laticinctus
Hybrids
C
A. piscivorus
A conanti
Hybrids
D
A. piscivorus
A. conanti
Hybrids
E
Central Clade
Eastern Clade
F

Fig 13

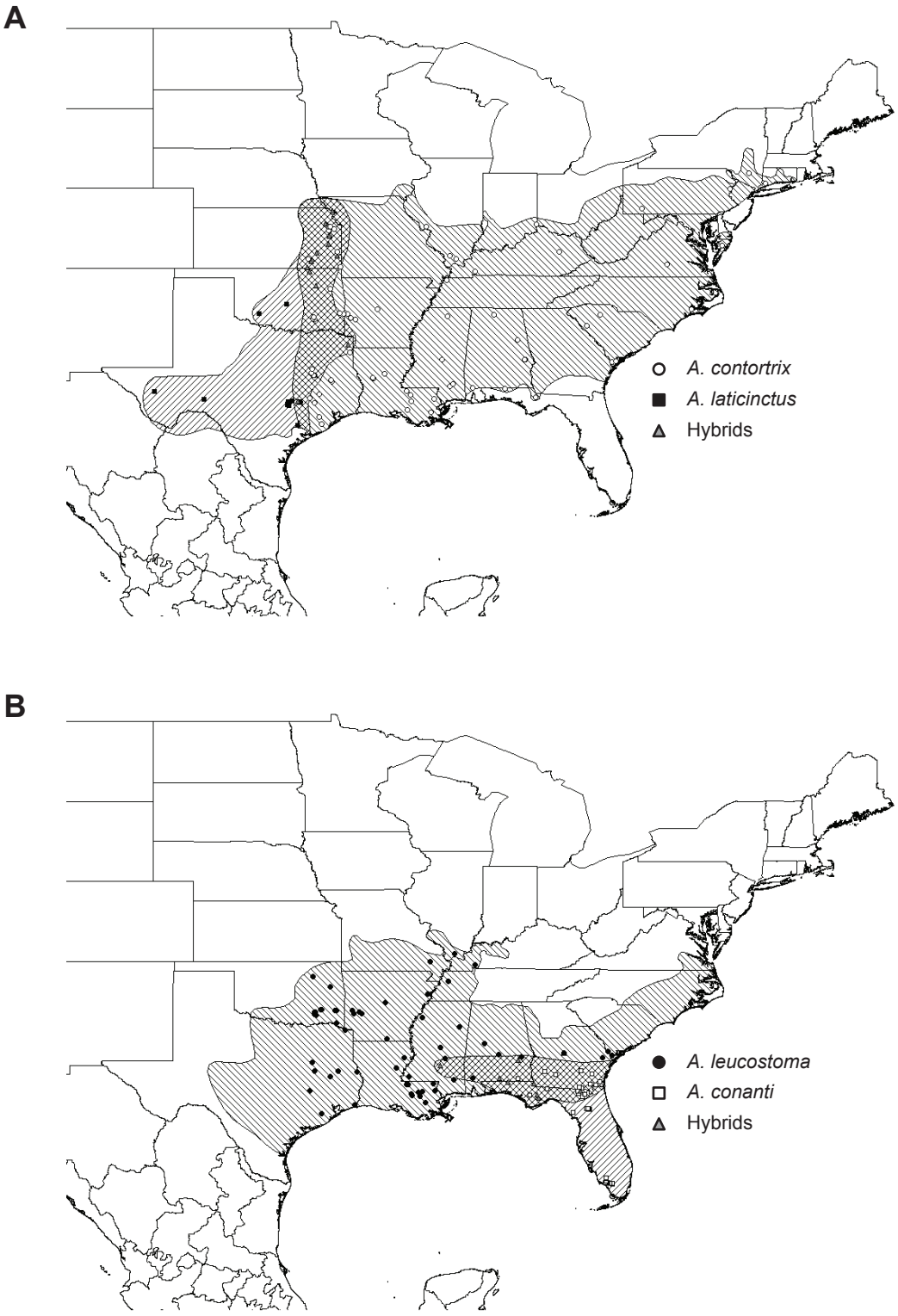

A
A. contortrix
A. laticinctus
Hybrids
B
A. leucostoma
A. conanti
Hybrids

Fig 14

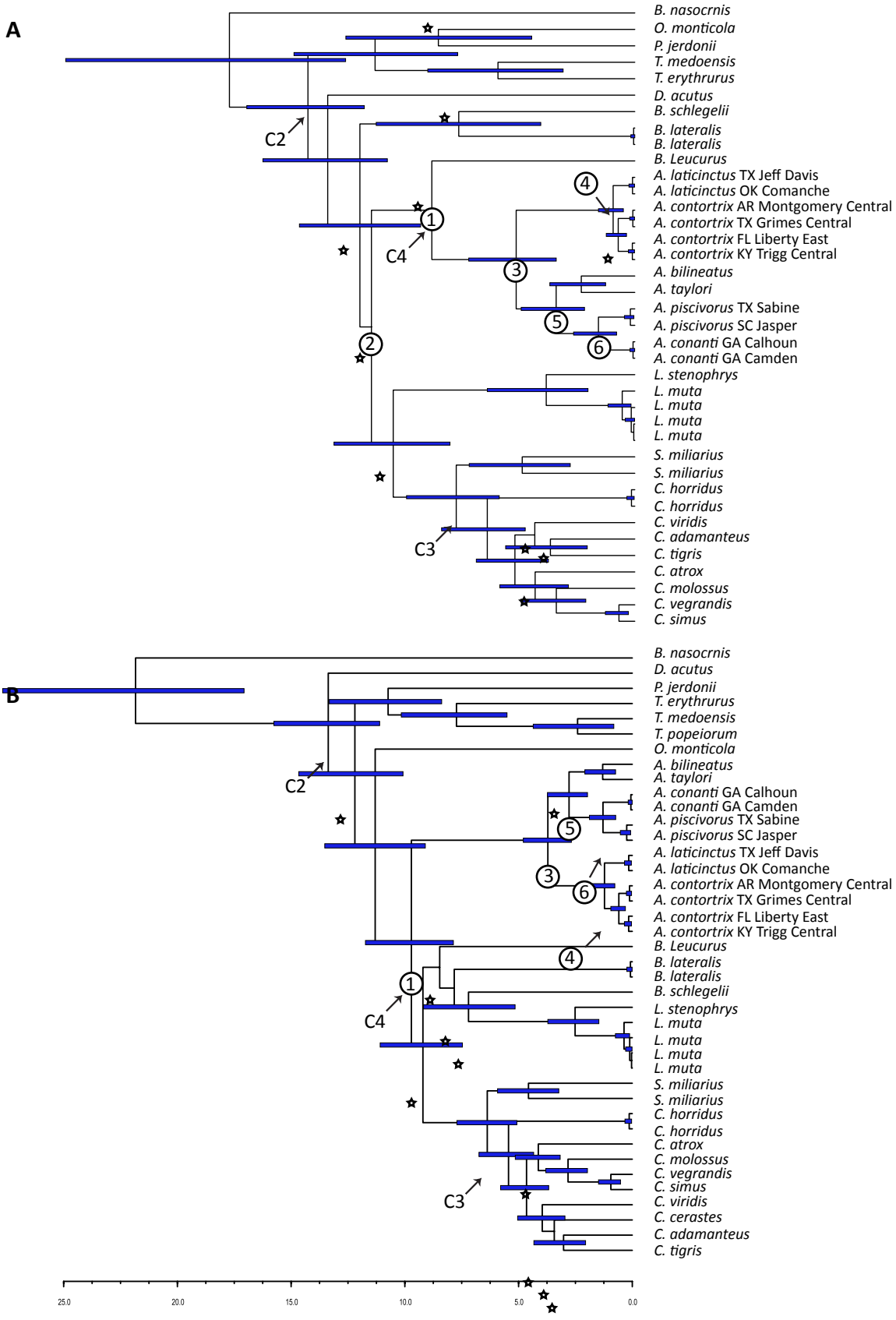

A
B. nasocrnis
O. monticola
P. jerdonii
T. medoensis
T. erythrurus
D. acutus
B. schlegelii
B. lateralis
B. lateralis
B. Leucurus
A. laticinctus TX Jeff Davis
A. laticinctus OK Comanche
A. contortrix AR Montgomery Central
A. contortrix TX Grimes Central
A. contortrix FL Liberty East
A. contortrix KY Trigg Central
A. bilineatus
A. taylori
A. piscivorus TX Sabine
A. piscivorus SC Jasper
A. conanti GA Calhoun
A. conanti GA Camden
L. stenophrys
L. muta
L. muta
L. muta
L. muta
S. miliarius
S. miliarius
C. horridus
C. horridus
C. viridis
C. adamanteus
C. tigris
C. atrox
C. molossus
C. vegrandis
C. simus
C2
C4
C3
1
2
3
4
5
6
B
B. nasocrnis
D. acutus
P. jerdonii
T. erythrurus
T. medoensis
T. popeiorum
O. monticola
A. bilineatus
A. taylori
A. conanti GA Calhoun
A. conanti GA Camden
A. piscivorus TX Sabine
A. piscivorus SC Jasper
A. laticinctus TX Jeff Davis
A. laticinctus OK Comanche
A. contortrix AR Montgomery Central
A. contortrix TX Grimes Central
A. contortrix FL Liberty East
A. contortrix KY Trigg Central
B. Leucurus
B. lateralis
B. lateralis
B. schlegelii
L. stenophrys
L. muta
L. muta
L. muta
L. muta
S. miliarius
S. miliarius
C. horridus
C. horridus
C. atrox
C. molossus
C. vegrandis
C. simus
C. viridis
C. cerastes
C. adamanteus
C. tigris
C2
C4
C3
1
3
4
5
6
25.0
20.0
15.0
10.0
5.0
0.0

Fig 15

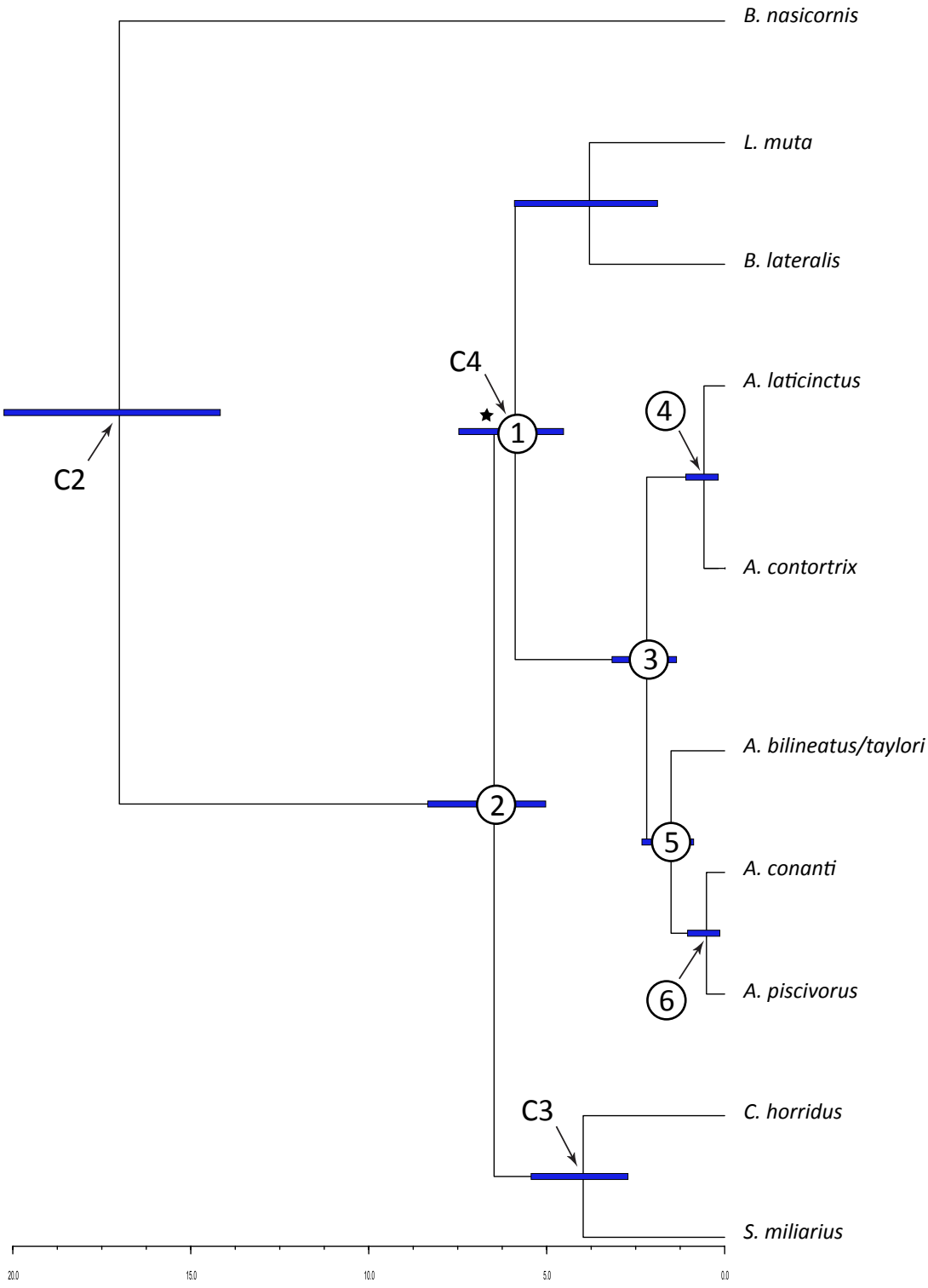

B. nasicornis
L. muta
B. lateralis
C4
A. laticinctus
4
1
C2
A. contortrix
3
A. bilineatus/taylori
2
5
A. conanti
6
A. piscivorus
C3
C. horridus
S. miliarius
20.0
15.0
10.0
5.0
0.0

Fig 15 cont.

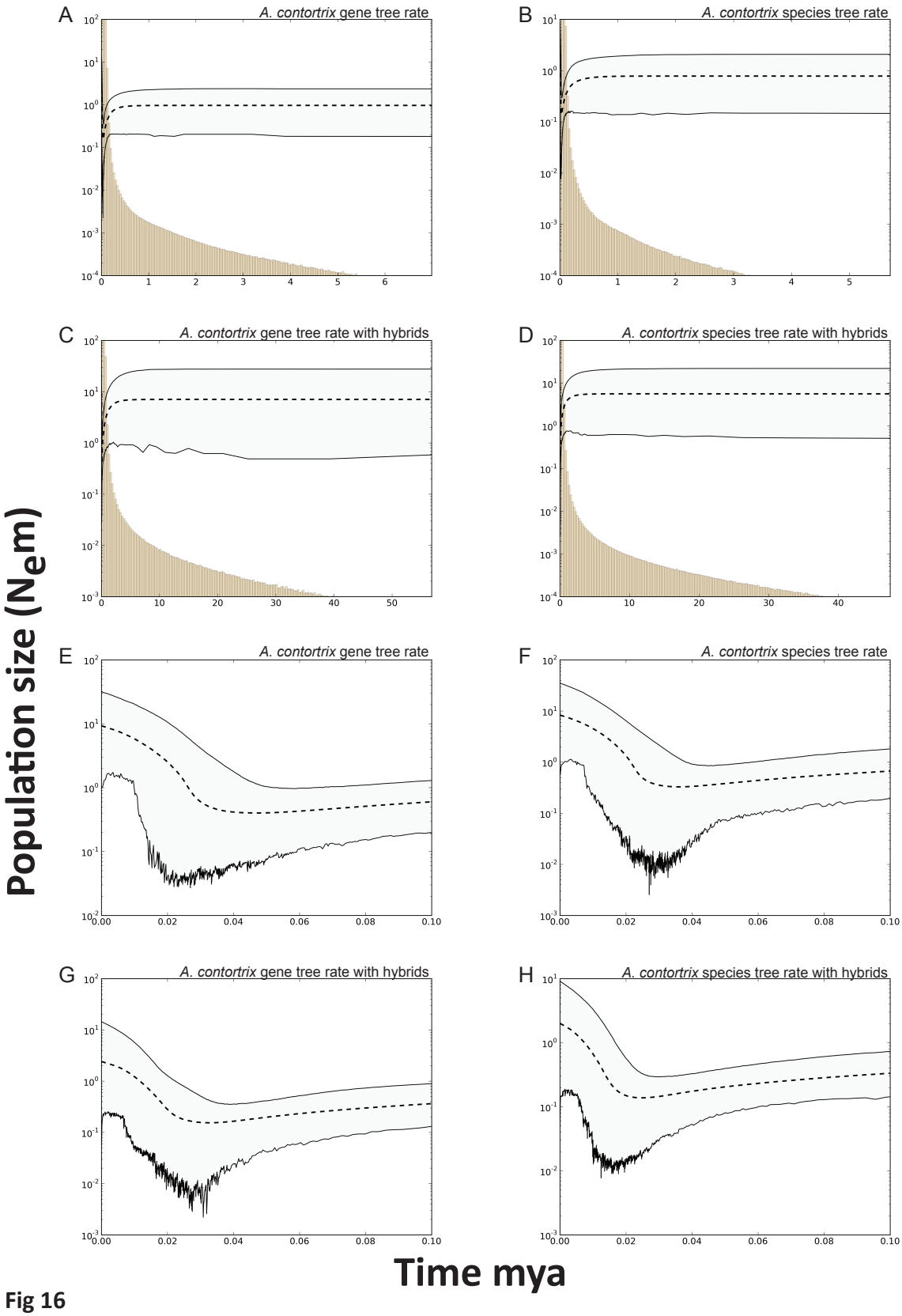

A
A. contortrix gene tree rate
B
A. contortrix species tree rate
C
A. contortrix gene tree rate with hybrids
D
A. contortrix species tree rate with hybrids
E
A. contortrix gene tree rate
F
A. contortrix species tree rate
G
A. contortrix gene tree rate with hybrids
H
A. contortrix species tree rate with hybrids
Population size (NeM)
Time mya

Fig 16

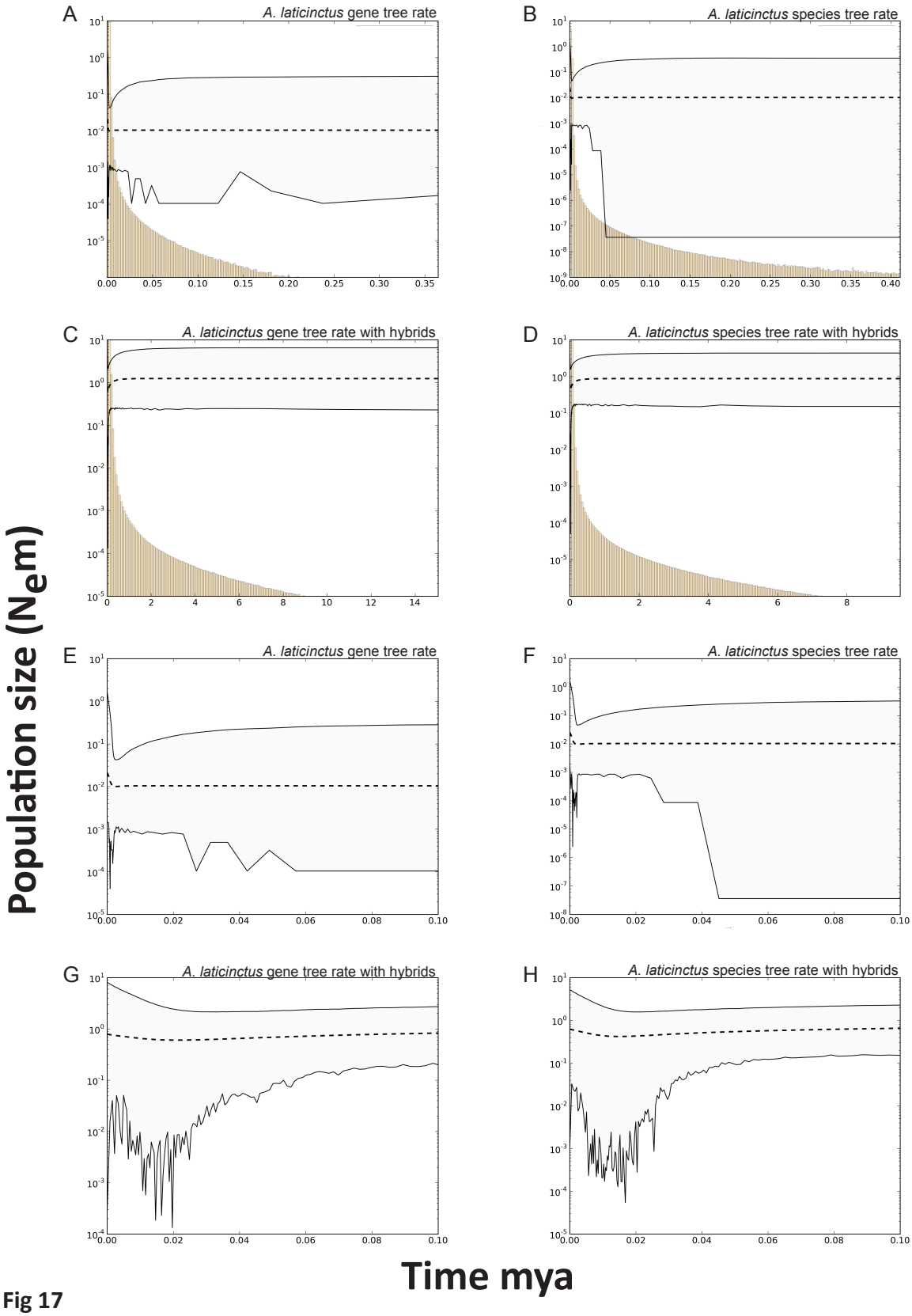

A
A. laticinctus gene tree rate
B
A. laticinctus species tree rate
C
A. laticinctus gene tree rate with hybrids
D
A. laticinctus species tree rate with hybrids
E
A. laticinctus gene tree rate
F
A. laticinctus species tree rate
G
A. laticinctus gene tree rate with hybrids
H
A. laticinctus species tree rate with hybrids
Population size (NeM)
Time mya

Fig 17

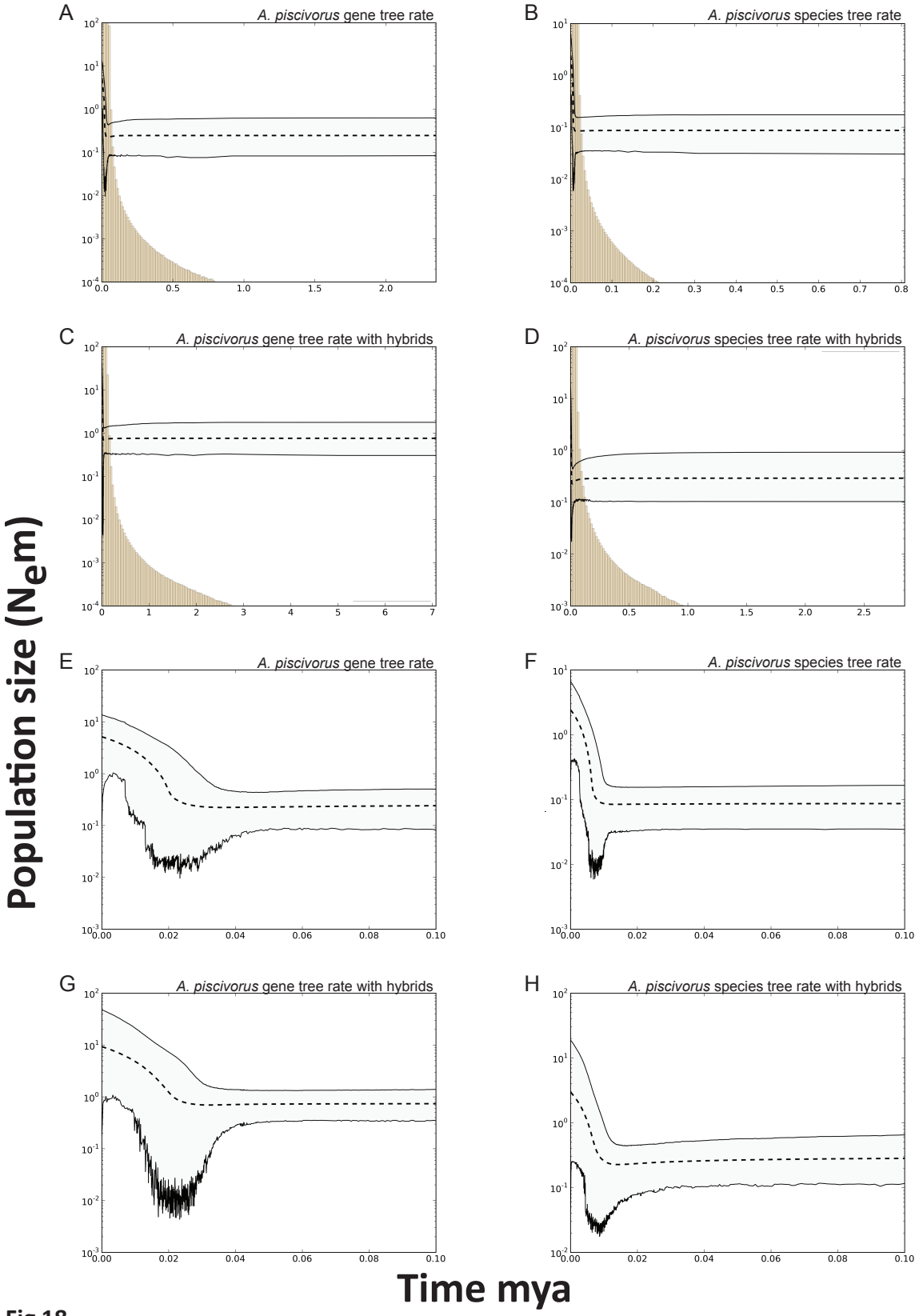

A
A. piscivorus gene tree rate
B
A. piscivorus species tree rate
C
A. piscivorus gene tree rate with hybrids
D
A. piscivorus species tree rate with hybrids
E
A. piscivorus gene tree rate
F
A. piscivorus species tree rate
G
A. piscivorus gene tree rate with hybrids
H
A. piscivorus species tree rate with hybrids
Population size (Nem)
Time mya

Fig 18

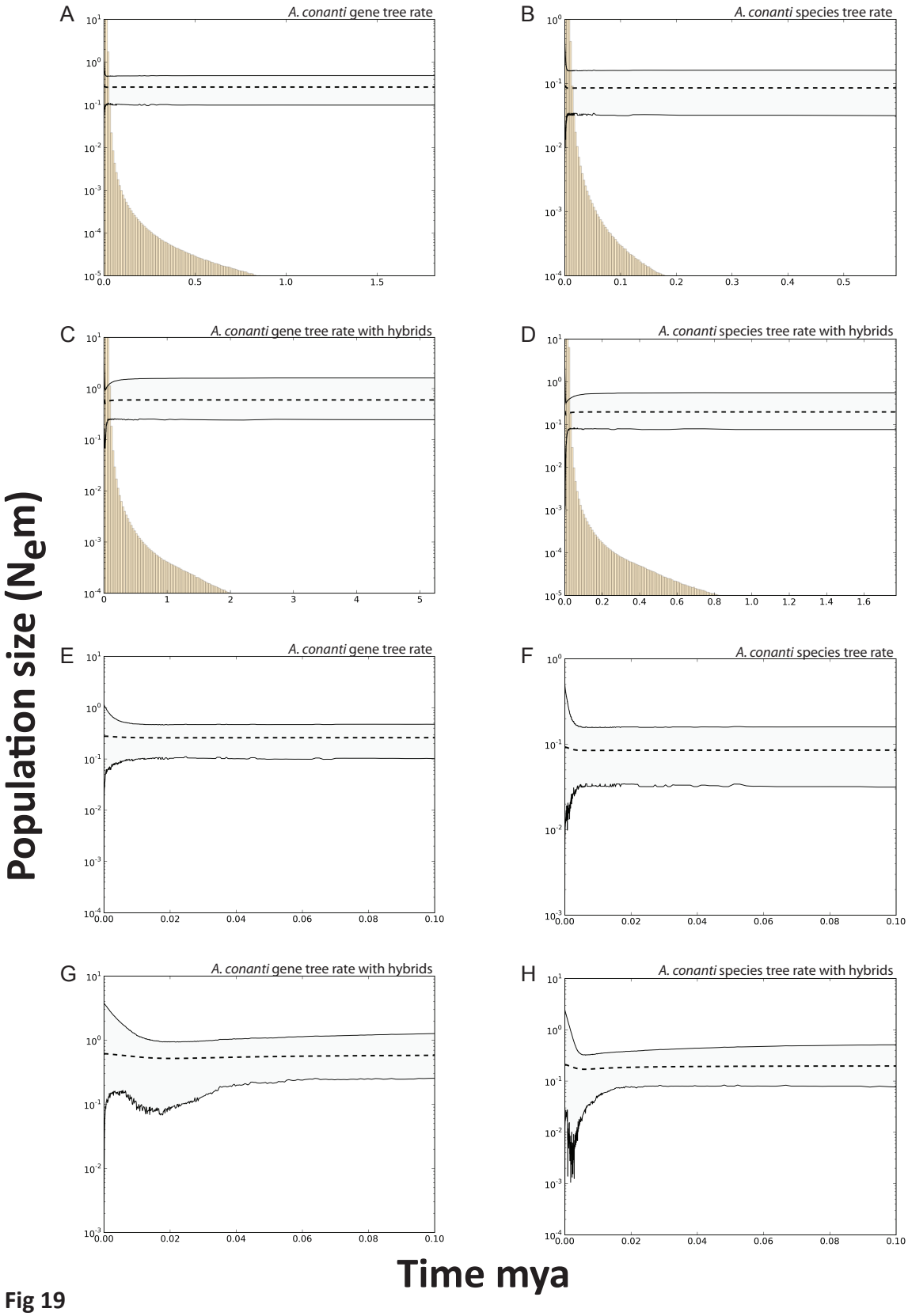

A
A. conanti gene tree rate
B
A. conanti species tree rate
C
A. conanti gene tree rate with hybrids
D
A. conanti species tree rate with hybrids
E
A. conanti gene tree rate
F
A. conanti species tree rate
G
A. conanti gene tree rate with hybrids
H
A. conanti species tree rate with hybrids
Population size (Nem)
Time mya

Fig 19

Appendix 1

List of Morphological Characters used in this study

1. Snout-Vent Length (SVL). Measured from the anterior rostral tip to the posterior margin of the anal plate.

2. Tail Length (TL). Measured from the posterior margin of the anal plate to the tip of the tail spine.

3. Circumference of Neck (CN). Measured directly posterior to the posterior apex of the retroarticular process of the compound bone, at the narrowest part of the neck.

4. Circumference at Midbody (CM). Measured at the 55th ventral scale.

5. Circumference at Vent (CV). Measured at the posterior margin of the anal plate.

6. Head Length Left (HLL). Measured from the rostral tip to the posterior apex of the retroarticular process of the compound bone.

7. Head Length Right (HLR). Measured from the rostral tip to the posterior apex of the retroarticular process of the compound bone.

8. Head Width (HW). Measured at the posterior apex of the retroarticular process of the compound bone, at the widest part of the head.

9. Head Height (HH). Measured at the frontal midway between the prefrontals and parietals.

10. Body Width (BW) Measured at the 55th ventral scale.

11. Body Height (BH) Measured at the 55th ventral scale.

12. Right Eye Diameter (EYE R). Measured at the widest horizontal point between the right preocular and right postocular.

13. Left Eye Diameter (EYE L). Measured at the widest horizontal point between the left preocular and left postocular.

14. Nostril to Eye Right (NER). Measured from the anterior margin of the right posterior nasal to the posterior suture of the right preoculars.

15. Nostril to Eye Left (NEL). Measured from the anterior margin of the left posterior nasal to the posterior suture of the left preoculars.

16. Nostril to Snout Right (NSR). Measured from the tip of the rostral to the posterior margin of the right anterior nasal.

17. Nostril to Snout Left (NSL). Measured from the tip of the rostral to the posterior margin of the right anterior nasal.

18. Width Between Nostrils (WBN). Measured from the dorsal suture of the left posterior nasal and anterior nasal to the dorsal suture of the right posterior nasal and anterior nasal.

19. Width Between Eyes (WBE). Measured at the anterior suture of the supraoculars and postoculars.

20. Parietal Width (PW). Measured at the suture between the parietal, postocular and anterior most temporal.

Bibliography

Avise, J. C., J. F. Shapira, and S. W. Daniel. 1984. Mitochondrial DNA differentiation during the speciation process in *Peromyscus*. Mol Biol Evol 1:38-56.

Avis, J. C., J. Arnold, R. M. Ball, E. Birmingham, T. Lamb, J. E. Neigel, C. A. Reeb, N. C. Saunders. 1987. Intraspecific phylogeography: The mitochondrial DNA bridge between populaiton genetics and systematics. Ann. Rev. Ecol. Syst. 18:489-522.

Bailey, R. G. 1995. Description of the ecoregions of the United States. 2n ed. (rev. and expanded 1st ed. 1980). Misc. Publ. No 1391 (rev). U.S. Forest Service, Washington, DC.

Barnes, I., B. Shapiro, A. Lister, T. Kuznetsov, A. Sher, D. Guthrie, and M. G. Thomas. 2007. Genetic structure and extinction of the wooly mammoth, *Mammuthus primigenius*. Curr Biol 17:1072-1075.

Barnosky, A. D., N. Matzke, S. Tomiya, G. O. Wogan, B. Swartz, T. B. Quental, C. Marshall, J. L. McGuire, E. L. Lindsey, K. C. Maguire, B. Mersey, and E. A. Ferrer. 2011. Has the Earth's sixth mass extinction already arrived? Nature 471:51-57.

Barraclough, T. G., and S. Nee. 2001. Phylogenetics and speciation. Trends Ecol Evol 16:391-399.

Bauer, A. M., J. F. Parham, R. M. Brown, B. L. Stuart, L. Grismer, T. J. Papenfuss, W. Bohme, J. M. Savage, S. Carranza, J. L. Grismer, P. Wagner, A. Schmitz, N. B. Ananjeva, and R. F. Inger. 2011. Availability of new Bayesian-delimited gecko names and the importance of character-based species descriptions. Proc Biol Sci 278:490-492; discussion 493-495.

Beck, J. B., H. Schmuths, and B. A. Schaal. 2008. Native range genetic variation in Arabidopsis thaliana is strongly geographically structured and reflects Pleistocene glacial dynamics. Mol Ecol 17:902-915.

Bertorelle, G., and M. Slatkin. 1995. The number of segregating sites in expanding human populations, with implications for estimates of demographics parameters. Mol Biol Evol 12:887-892.

Bickford, D., D. J. Lohman, N. S. Sodhi, P. K. Ng, R. Meier, K. Winker, K. K. Ingram, and I. Das. 2007. Cryptic species as a window on diversity and conservation. Trends Ecol Evol 22:148-155.

Birmingham, E., and C. Moritz. 1998. Comparative phylogeography: concepts and applications. Mol Ecol 7:367-369.

Brown, W. M., M. George, Jr., and A. C. Wilson. 1979. Rapid evolution of animal mitochondrial DNA. Proc Natl Acad Sci U S A 76:1967-1971.

Brumfield, R. T., L. Liu, D. E. Lum, and S. V. Edwards. 2008. Comparison of species tree methods for reconstructing the phylogeny of bearded manakins (Aves: Pipridae, *Manacus*) from Multilocus Sequence Data. Syst Biol 57:719-731.

Burbrink, F. T. 2001. Systematics of the eastern ratsnake complex (*Elaphe obsoleta*). Herpetol Monogr 15:1-53.

Burbrink, F. T. 2002. Phylogeographic analysis of the cornsnake (*Elaphe guttata*) complex as inferred from maximum likelihood and Bayesian analyses. Mol Phylogenet Evol 25:465-476.

Burbrink, F. T., and R. Alexander Pyron. 2011. The impact of gene-tree/species-tree discordance on diversification-rate estimation. Evolution: (In Press)

Burbrink, F. T., F. Fontanella, R. A. Pyron, T. J. Guiher, and C. Jimenez. 2008. Phylogeography across a continent: The evolutionary and demographic history of the North American racer (Serpentes: Colubridae*: Coluber constrictor*). Mol Phylogenet Evol 47:274-288.

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.

Burleigh, J. G., M. S. Bansal, O. Eulenstein, S. Hartmann, A. Wehe, and T. J. Vision. 2011. Genome-Scale phylogenetics: inferring the plant tree of life from 18,896 gene trees. Syst Biol 60:117-125.

Carstens, B. C., and T. A. Dewey. 2010. Species delimitation using a combined coalescent and information-theoretic approach: An example from North American myotis bats. Syst Biol 59:400-414.

Carstens, B. C., and L. L. Knowles. 2007. Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: An example from *Melanoplus* grasshoppers. Syst Biol 56:400-411.

Castoe, T. A., and C. L. Parkinson. 2006. Bayesian mixed models and the phylogeny of pitvipers (Serpentes: Viperidae). Mol Phylogenet Evol 39:91-110.

Castoe, T. A., C. L. Spencer, and C. L. Parkinson. 2007. Phylogeographic structure and historical demography of the western diamondback rattlesnake (*Crotalus atrox*):

A perspective on North American desert biogeography. Mol Phylogenet Evol 42:193-212.

Catesby, M., E. C. Hartopp, J. F. Hunnewell, and Jay I. Kislak Collection (Library of Congress). 1743. The natural history of Carolina, Florida and the Bahama Islands : containing the figures of birds, beasts, fishes, serpents, insects, and plants : particularly the forest-trees, shrubs, and other plants, not hitherto described, or very incorrectly figured by authors : together with their descriptions in English and French, to which, are added observations on the air, soil, and waters, with remarks upon agriculture, grain, pulse, roots, &c. : to the whole, is prefixed a new and correct map of the countries treated of. Printed at the expence of the author, and sold by W. Innys and R. Manby ... by Mr. Hauksbee, at the Royal Society House, and by the author, at Mr. Bacon's in Hoxton, London.

Chivian, E., and A. Bernstein. 2008. Sustaining life: How human health depends on biodiversity. Oxford University, New York.

Clark, A. M., B. W. Bowen, and C. L. Branch. 1999. Effects of habitat fragmentation on an endemic scrub lizard (*Sceloporus woodi*): an historical perspective based on a mitochondrial gene geneology. Mol Ecol 8:1093-1104.

Conant, R., and J. T. Collins. 1991. A field guide to reptiles and amphibians: Eastern and Central North America. Houghton Mifflin, Boston.

Crespi, E. J., L. J. Rissler, and R. A. Browne. 2003. Testing Pleistocene refugia theory: phylogeographical analysis of *Desmognathus wrighti*, a high-elevation salamander in the southern Appalachians. Mol Ecol 12:969-984.

Crother, B. I., J. Boundy, F. T. Burbrink, J. A. Campbell, and R. A. Pyron. 2011. Squamata - Snakes. *IN* B. I. Crother (ed.), Scientific and Standard English Names of Amphibians and Reptiles of North America North of Mexico. . SSAR Herpetological Circular 37:24-45.

de Queiroz, A., R. Lawson, and J. A. Lemos-Espinal. 2002. Phylogenetic relationships of North American garter snakes (*Thamnophis*) based on four mitochondrial genes: How much DNA is enough? Mol Phylogenet Evol 22:315-329.

de Queiroz, K. 2007. Species concepts and species delimitation. Syst Biol 56:879-886.

Degnan, J. H., M. DeGiorgio, D. Bryant, and N. A. Rosenberg. 2009. Properties of consensus methods for inferring species trees from gene trees. Syst Biol 58:35-54.

Degnan, J. H., and N. A. Rosenberg. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol Evol 24:332-340.

Denton, G. H., and T. J. Hughes. 1981. The Last Great Ice Sheets. John Wiley and Sons, New York.

Douglas, M. E., M. R. Douglas, G. W. Schuett, and L. W. Porras. 2006. Evolution of rattlesnakes (Viperidae; *Crotalus*) in the warm deserts of western North America shaped by Neogene vicariance and Quaternary climate change. Mol Ecol 15:3353-3374.

Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. Plos Biol 4:699-710.

Drummond, A. J., O. G. Pybus, A. Rambaut, R. Forsberg, and A. G. Rodrigo. 2003. Measurably evolving populations. Trends Ecol Evol 18:481-488.

Drummond, A. J., and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. Bmc Evol Biol 7:214.

Drummond, A. J., A. Rambaut, B. Shapiro, and O. G. Pybus. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol 22:1185-1192.

Dyke, A. S., and V. K. Prest. 1987. The Late Wisconsinan and Holocene history of the Laurentide Ice Sheet. Geographie physique et Quaternaire 41:237-263.

Edgar, R. C. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. Bmc Bioinformatics 5:1-19.

Edwards, S. V. 2009. Is a new and general theory of molecular systematics emerging? Evolution 63:1-19.

Edwards, S. V., and P. Beerli. 2000. Perspective: Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. Evolution 54:1839-1854.

Eernisse, D. J., and A. G. Kluge. 1993. Taxonomic congruence versus total evidence, and amniote phylogeny inferred from fossils, molecules, and morphology. Mol Biol Evol 10:1170-1195.

Elith, J., S. J. Phillips, T. Hastie, M. Dudik, Y. E. Chee, and C. J. Yates. 2011. A statistical explanation of MaxEnt for ecologists. Divers Distrib 17:43-57.

Ellsworth, D. L., R. L. Honeycutt, and N. J. Silvy. 1994. Historical biogeography and contemporary patterns of mitochondrial DNA variation in white-tailed deer from the southeastern United States. Evolution 48:122-136.

Emslie, S. D. 2007. Fossil Passerines from the early Pliocene of Kansas and the evolution of songbirds in North America. Auk 124.

Excoffier, L., and S. Schneider. 1999. Why hunter-gatherer populations do not showsign of Pleistocene demographics expansions. Proceedings of the National Academy of Sciences 96:10597-10602.

Excoffier, L., P. E. Smouse, and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genet Mol Res 131.

Falush, D., M. Stephens, and J. K. Pritchard. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567-1587.

Falush, D., M. Stephens, and J. K. Pritchard. 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. Mol Ecol 16:895-908.

Felsenstein. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783-791.

Felsenstein, J. 2004. Inferring phylogenies. Sinauer Associates, Sunderland, Mass.

Fontanella, F. M., C. R. Feldman, M. E. Siddall, and F. T. Burbrink. 2008. Phylogeography of *Diadophis punctatus*: Extensive lineage diversity and repeated patterns of historical demography in a trans-continental snake. Mol Phylogenet Evol 46:1049-1070.

Frost, D. R., and D. M. Hillis. 1990. Species in concept and practice - Herpetological Applications. Herpetologica 46:87-104.

Fu, Y. X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147:915-925.

Fujita, M. K., and A. D. Leache. 2011. A coalescent perspective on delimiting and naming species: a reply to Baure et al. Proceedings of the Royal Society B: Biological Sciences 278:493-495.

Geffeney, S. L., E. Fujimoto, E. D. Brodie, 3rd, E. D. Brodie, Jr., and P. C. Ruben. 2005. Evolutionary diversification of TTX-resistant sodium channels in a predator-prey interaction. Nature 434:759-763.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 1995. Bayesian data analysis. Chapman and Hall, London.

Genecodes. 2000. SEQUENCHER 4.5 Genecodes.

Gibbons, J. W. D., M. E. . 2002. Defensive behavior of cottonmouths (*Agkistrodon piscivorus*) towards humans. Copeia 1:195-198.

Gibbs, H. L., and J. Diaz. 2010. Identification of single copy nuclear DNA markers for North American pit vipers. Mol Ecol Resour 10:177-180.

Glaudas, X., C. T. Winne, and L. A. Fedewa. 2006. Ontogeny of anti-predator behavioral habituation in cottonmouths (*Agkistrodon piscivorus*). Ethology 112:608-615.

Gloyd, H. K. 1969. Two additional subspecies of North American crotalid snakes, genus *Agkistrodon*. P Biol Soc Wash 90:1002-1015.

Gloyd, H. K., and R. Conant. 1934. The broad-banded copperhead: a new subspecies of *Agkistrodon mokasen*. Occas. Papers Mus. Zool. University of Michigan 183:1-6.

Gloyd, H. K., and R. Conant. 1943. A synopsis of the American forms of *Agkistrodon* (copperheads and moccasins). Bulletin of the Chicago Academy of Science 7:147-170.

Gloyd, H. K., and R. Conant. 1990. Snakes of the *Agkistrodon* complex. A monographic review. Society for the Study of Amphibians and Reptiles, Oxford, Ohio.

Graham, S. P., R. L. Earley, S. K. Hoss, G. W. Schuett, and M. S. Grober. 2008. The reproductive biology of male cottonmouths (*Agkistrodon piscivorus*): Do plasma steroid hormones predict the mating season? Gen Comp Endocr 159:226-235.

Guiher, T. J., and F. T. Burbrink. 2008. Demographic and phylogeographic histories of two venomous North American snakes of the genus *Agkistrodon*. Mol Phylogenet Evol 48:543-553.

Harpending, H. C. 1994. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. Hum Biol 66:591-600.

Hasegawa, M., H. Kishino, and T. A. Yano. 1985. Dating of the Human Ape Splitting by a Molecular Clock of Mitochondrial-DNA. J Mol Evol 22:160-174.

Heled, J., and A. J. Drummond. 2008. Bayesian inference of population size history from multiple loci. Bmc Evol Biol 8:289.

Heled, J., and A. J. Drummond. 2010. Bayesian Inference of Species Trees from Multilocus Data. Mol Biol Evol 27:570-580.

Hewitt, C. D., A. J. Broccoli, J. F. B. Mitchell, and R. J. Stouffer. 2001. A coupled model study of the last glacial maximum: Was part of the North Atlantic relatively warm? Geophys Res Lett 28:1571-1574.

Hewitt, G. M. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. Biol J Linn Soc 58:247-276.

Hewitt, G. M. 2000. The genetic legacy of the Quaternary ice ages. Nature 405:907-913.

Hey, J. 2006. Recent advances in assessing gene flow between diverging populations and species. Curr Opin Genet Dev 16:592-596.

Hey, J., and R. Nielsen. 2007. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. P Natl Acad Sci USA 104:2785-2790.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. Int J Climatol 25:1965-1978.

Hijmans, R. J., M. Cruz, E. Rojas, and L. Guarino. 2001. DIVA-GIS, version 1.4. A geographic information system for the management and analysis of genetic resources data.

Hillis, D. M., and J. J. Bull. 1993. An empirical-test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst Biol 42:182-192.

Hillis, M. 1978. On ratios-a response to Atchley, Gaskins and Anderson. Syst Zool 27:61-62.

Hoge, A. R., and S. A. R. W. D. L. Romano-Hoge. 1981. Poisonous snakes of the world, Part 1: Check list of the pit vipers Viperoidea, Viperidae, Crotalinae. Mem. Inst. Butantan 42/43:179-283.

Holman, J. A. 2000. Fossil snakes of North America: origin, evolution, distribution, paleoecology. Indiana University Press, Bloomington.

Howes, B. J., J. W. Brown, H. L. Gibbs, T. B. Herman, S. W. Mockford, K. A. Prior, and P. J. Weatherhead. 2009. Directional gene flow patterns in disjunct populations of the black ratsnake (*Pantheropis obsoletus*) and the Blanding's turtle (*Emydoidea blandingii*). Conserv Genet 10:407-417.

Hubisz, M. J., D. Falush, M. Stephens, and J. K. Pritchard. 2009. Inferring weak population structure with the assistance of sample group information. Mol Ecol Resour 9:1322-1332.

Huelsenbeck, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754-755.

Huson, D. H., D. Bryant. 2006. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23:254-267.

Kass, R. E., and A. E. Raftery. 1995. Bayes factors and model uncertainty. Journal of American Statistics Association 90.

Killey, M. M. 1998. Illinois' ice age legacy. . Illinois State Geological Survey GeoScience Education Series 14:66.

Kingman, J. F. 2000. Origins of the coalescent. 1974-1982. Genetics 156:1461-1463.

Knight, A., L. D. Densmore, and E. D. Rael. 1992. Molecular systematics of the *Agkistrodon* complex. . Selva, Tyler, TX.

Lacepède, B. G. E. d. l. v. 1789. Histoire naturelle des quadrupèdes ovipares et des serpens, vol 2, Table Methodique. Hotel de Thou, Par.

Leache, A. D., and M. K. Fujita. 2010. Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). Proc Biol Sci 277:3071-3077.

Leache, A. D., and B. Rannala. 2011. The accuracy of species tree estimation under simulation: A comparison of methods. Syst Biol 60:126-137.

Leache, A. D., and T. W. Reeder. 2002. Molecular systematics of the Eastern Fence Lizard (*Sceloporus undulatus*): A comparison of parsimony, likelihood, and Bayesian approaches. Syst Biol 51:44-68.

Linder, H. P., C. R. Hardy, and F. Rutschmann. 2005. Taxon sampling effects in molecular clock dating: an example from the African Restionaceae. Mol Phylogenet Evol 35:569-582.

Linnaeus, C. 1766. Systema Naturae...Editio duodecima.. . Herpetological section reprinted 1963 by Ohio Herpetol. Soc., now Soc. Study Amphib. Reptiles.

Liu, L., and D. K. Pearl. 2007. Species trees from gene trees: reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. Syst Biol 56:504-514.

Liu, L., L. L. Yu, L. Kubatko, D. K. Pearl, and S. V. Edwards. 2009. Coalescent methods for estimating phylogenetic trees. Mol Phylogenet Evol 53:320-328.

Loveless, M. D., and J. L. Hamrick. 1988. Genetic organization and evolutionary history in two North American species of *Cirsium*. Evolution 42:254-265.

Marjoram, P., and P. Donnelly. 1994. Pairwise comparisons of mitochondrial DNA sequences in subdivided populations and implications for early human evolution. Genetics 136:673-683.

Mathee, C. A., J. D. Burzlaff, J. F. Taylor, and S. K. Davis. 2001. Mining the mammalian genome for artiodactyl systematics. Syst Biol 50:367-390.

McCormack, J. E., J. Heled, K. S. Delaney, A. T. Peterson, and L. L. Knowles. 2010. Calibrating divergence times of species trees versus gene trees:implications for speciation history of *Aphelocoma* jays. Evolution 65:184-202.

Morin, P. A., F. I. Archer, A. D. Foote, J. Vilstrup, E. E. Allen, P. Wade, J. Durban, K. Parsons, R. Pitman, L. Li, P. Bouffard, S. C. A. Nielsen, M. Rasmussen, E. Willerslev, M. T. P. Gilbert, and T. Harkins. 2010. Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. Genome Res 20:908-916.

Naylor, G. J., and W. M. Brown. 1998. Amphioxus mitochondrial DNA, chordate phylogeny, and the limits of inference based on comparisons of sequences. Syst Biol 47:61-76.

Newton, M. A., and A. E. Raftery. 1994. Approximate Bayesian inference with the weighted likelihood bootstrap. Journal of Royal Statistics Society. B 56:3-48.

Niemiller, M. L., P. Nosil, and B. M. Fitzpatrick. 2008. Recent divergence-with-gene-flow in Tennessee cave salamanders (Plethodontidae; *Gyrinophilus*) inferred from gene genealogies (vol 17, pg 2258, 2008). Mol Ecol 19:1513-1514.

Noonan, B. P., and P. T. Chippindale. 2006. Vicariant origin of Malagasy reptiles supports late cretaceous antarctic land bridge. Am Nat 168:730-741.

Nosil, P. 2008. Speciation with gene flow could be common. Mol Ecol 17:2103-2106.

O'Meara, B. C. 2010. New heuristic methods for joint species delimitation and species tree inference. Syst Biol 59:59-73.

Pagel, M. 1999. Inferring the historical patterns of biological evolution. Nature 401:877-884.

Palisot de Beauvois, A. M. F. J. 1799. Memoir on Amphibia. Serpents. Trans Amer. Philos. Soc. 4:362-381.

Parkinson, C. L., P. T. Chippindale, and J. Campbell. 2002. Multigene analyses of pitviper phylogeny with comments on their biogeographical history. *IN* Biology of the Vipers. Eagle Mountain Publishing, LC. Eagle Mountain, Utah.

Parkinson, C. L., S. M. Moody, and J. E. Ahlquist. 1997. Phylogenetic relationships of the '*Agkistrodon* complex' based on mitochondrial DNA sequence data. Oxford University, London.

Parkinson, C. L., K. R. Zamudio, and H. W. Greene. 2000. Phylogeography of the pitviper clade *Agkistrodon*: historical ecology, species status, and conservation of cantils. Mol Ecol 9:411-420.

Parmley, D., and J. A. Holman. 2007. Earliest Fossil Record of a Pigmy Rattlesnake (Viperidae: *Sistrurus Garman*). . J Herpetol 41:141-144.

Pearson, R. G., C. J. Raxworthy, M. Nakamura, and A. T. Peterson. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. J Biogeogr 34:102-117.

Pedall, I., U. Fritz, H. Stuckas, A. Valdeon, and M. Wink. 2011. Gene flow across secondary contact zones of the *Emys orbicularis* complex in the Western Miditerranean and evidence for extinction and re-introduction of pond turtles on Corsica and Sardina (Testudines: Emydidae). J Zool Syst Evol Res 49:44-57.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecol Model 190:231-259.

Phillips, S. J., and M. Dudik. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31:161-175.

Pielou, E. C. 1991. After the Ice Age: The Return of Life to Glaciated North America. University of Chicago Press, Chicago.

Posada, D. 2008. jModelTest: Phylogenetic model averaging. Mol Biol Evol 25:1253-1256.

Posada, D., and T. R. Buckley. 2004. Model selection and model averaging in phylogenetics: advantages of the AIC and Bayesian approaches over likelihood ratio tests. Syst Biol 53.

Posada, D., and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817-818.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945-959.

Pybus, O. G., and P. H. Harvey. 2000. Testing macro-evolutionary models using incomplete molecular phylogenies. P Roy Soc Lond B Bio 267:2267-2272.

Rage, J. C. 1982. L'histoire des serpents. Pour la Science 54:16-27.

Rambaut, A., and A. J. Drummond. 2007. Tracer v.1.5, Available from http://beast.bio.ed.ac.uk/Tracer.

Ramos-Onsis, S. E., and J. Rozas. 2002. Statistical properties of new neutrality tests against population growth. Mol Biol Evol 19:2092-2100.

Ray, N., M. Currat, and L. Excoffier. 2003. Intra-deme molecular diversity in spatially expanding populations. Mol Biol Evol 20:76-86.

Rissler, L. J., and J. J. Apodaca. 2007. Adding more ecology into species delimitation: ecological niche models and phylogeography help define cryptic species in the black salamander (*Aneides flavipunctatus*). Syst Biol 56:924-942.

Rogers, A. R., A. E. Fraley, M. J. Bamshad, W. S. Watkins, and L. B. Jorde. 1996. Mitochondrial mismatch analysis is insensitive to the mutational process. Mol Biol Evol 13:895-902.

Rogers, A. R., and H. Harpending. 1992. Population growth makes waves in the distribution of pairwise genetic differences. Mol Biol Evol 9:552-569.

Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.

Roth, E. D. 2005. Spatial ecology of a cottonmouth (*Agkistrodon piscivorus*) population in East Texas. J Herpetol 39:308-312.

Rowe, K. C., E. J. Heske, P. W. Brown, and K. N. Paige. 2004. Surviving the ice: Northern refugia and postglacial colonization. Proc Natl Acad Sci U S A 101:10355-10359.

Rozas, J., J. C. Sánchez-DelBarrio, X. Messeguer, and R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496-2497..

Runck, A. M., and J. A. Cook. 2005. Postglacial expansion of the southern red-backed vole (*Clethrionomys gapperi*) in North America. Mol Ecol 14:1445-1456.

Sambrock, J., and D. W. Russell. 2001. Molecular Cloning: A Laboratory Manual. Cold Spring Laboratory, Cold Spring Harbor

Schmidt, K. P. 1953. A check list of North American amphibians and reptiles. Sixth edition. American Society of Ichthyology and Herpetology, Chicago, Illinois.

Schneider, S., D. Roessli, and L. Excoffier. 2000. Arlequin: A software for population genetics data analysis. Version 2.000. Genetics and Biometry Lab, Department of Anthroplogy, University of Geneva.

Schoener, T. W. 1968. Anolis lizards of Bimini - resource partitioning in a complex fauna. Ecology 49:704-726.

Shapiro, B., A. J. Drummond, A. Rambaut, M. C. Wilson, P. E. Matheus, A. V. Sher, O. G. Pybus, M. T. P. Gilbert, I. Barnes, J. Binladen, E. Willerslev, A. J. Hansen, G. F. Baryshnikov, J. A. Burns, S. Davydov, J. C. Driver, D. G. Froese, C. R. Harington, G. Keddie, P. Kosintsev, M. L. Kunz, L. D. Martin, R. O. Stephenson, J. Storer, R. Tedford, S. Zimov, and A. Cooper. 2004. Rise and fall of the Beringian steppe bison. Science 306:1561-1565.

Sokal, R. R., and F. J. Rohlf. 1995. Biometry: the Principal and practice of statistics in biological research. W. H. Freeman and Company, New York.

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

Stamatakis, A., T. Ludwig, and H. Meier. 2005. RAxML-II: a program for sequential, parallel and distributed inference of large phylogenetic. Concurr Comp-Pract E 17:1705-1723.

Starkey, D. E., H. B. Shaffer, R. L. Burke, M. R. J. Forstner, J. B. Iverson, F. J. Janzen, A. G. J. Rhodin, and G. R. Ultsch. 2003. Molecular systematics, phylogeography, and the effects of Pleistocene glaciation in the painted turtle (*Chrysemys picta*) complex. Evolution 57:119-128.

Stephens, M., and P. Donnelly. 2003. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet 73:1162-1169.

Swenson, S., F. Costa, W. Ernst, G. Fujii, and F. S. Markland. 2005. Contortrostatin, a snake venom disintegrin with anti-angiogenic and anti-tumor activity. Pathophysiol Haemost Thromb 34:169-176.

Swets, J. A. 1988. Measuring the accuracy of diagnostic systems. Science 240:1285-1293.

Swofford, D. l. 2003. PAUP*bv10: Phylogenetic analysis using parsimony (*and other methods) v. 4. . Sinauer and Associates, Sunderland.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585-595.

Tegelström, H., H. P. Gelter, and M. Jaarola. 1990. Variation in Pied Flycatcher (*Ficedula hypoleuca*) mitochondrial DNA. Auk 107:730-736.

Thomas, C. D., A. Cameron, R. E. Green, M. Bakkenes, L. J. Beaumont, Y. C. Collingham, B. F. Erasmus, M. F. De Siqueira, A. Grainger, L. Hannah, L. Hughes, B. Huntley, A. S. Van Jaarsveld, G. F. Midgley, L. Miles, M. A. Ortega-Huerta, A. T. Peterson, O. L. Phillips, and S. E. Williams. 2004. Extinction risk from climate change. Nature 427:145-148.

Thorne, J. L., and H. Kishino. 2002. Divergence time and evolutionary rate estimation with multilocus data. Syst Biol 51:689-702.

Thorpe, R. S., H. Black, and A. Malhorta. 1983. Morphometric studies in inbred and hybrid house mice (*Mus* sp.): multivariate analysis of size and shape. Journal of Zoology London 199:421-432.

Trikha, M., Y. De Clerck, and F. S. Markland. 1994. Contortrostatin, A Snake Venom Disintegrin, Inhibits beta1 Integrin-mediated Human Metastatic Melanoma Cell Adhesion and Blocks Experimental Metastasis. Cancer Res 54:5.

Troost, G. 1836. On a new genus of serpents, and two new species of the genus *Heterodon*, inhabiting Tennessee. Ann. Lyc. Nat. Hist., New York 3:174-190.

VanDerWal, J., L. P. Shoo, C. H. Graham, and S. E. Williams. 2009. Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? Ecological Modelling 220:589-594.

Walker, D., and J. C. Avise. 1998. Principles of phylogeography as illustrated by freshwater and terrestrial turtles in the southeastern United States. Annu Rev Ecol Syst 29:23-58.

Waltari, E., and J. A. Cook. 2005. Hares on ice: phylogeography and historical demographics of *Lepus arcticus*, *L. othus*, and *L. timidus* (Mammalia: Lagomorpha). . Mol Ecol 14:3005-3016.

Waltari, E., R. J. Hijmans, A. T. Peterson, A. S. Nyari, S. L. Perkins, and R. P. Guralnick. 2007. Locating Pleistocene refugia: Comparing phylogeographic and ecological niche model predictions. Plos One 2(7): e536.

Warren, D. L., R. E. Glor, and M. Turelli. 2008. Environmental niche equivalency versus conservatism: Quantitative approaches to niche evolution. Evolution 62:2868-2883.

Warren, D. L., R. E. Glor, and M. Turelli. 2010. ENMTools: a toolbox for comparative studies of environmental niche models. Ecography 33:607-611.

Webb, S. D. 1990. Historical biogeography. *IN* Ecosystems of Florida. University of Central Florida Press, Orlando.

Wiens, J. J. 2008. Systematics and herpetology in the age of genomics. Bioscience 58:297-307.

Wuster, W., L. Peppin, C. E. Pook, and D. E. Walker. 2008. A nesting of vipers: phylogeny and historical biogeography of the Viperidae (Squamata: Serpentes). Mol Phylogenet Evol 49:445-459.

Yang, Z. H., and B. Rannala. 2010. Bayesian species delimitation using multilocus sequence data. P Natl Acad Sci USA 107:9264-9269.

Yu, Y., C. Than, J. H. Degnan, and L. Nakhleh. 2011. Coalescent histories on phylogenetic networks and detection of hybridization despite incomplete lineage sorting. Syst Biol 60:138-149.

Zaidan III, F. 2003. Variation in cottonmouth (*Agkistrodon piscivorus leucostoma*) resting metabolic rates. Comparative Biochemistry and Physiology 134:511-523.

Zehner, Z. E., and B. M. Paterson. 1983. Characterization of the chicken vimentin gene: single copy gene producing multiple mRNAs. Proceedings of the National Academy of Sciences 80:911-915.

Zhou, Q., C. P. Hu, M. R. Ritter, S. D. Swenson, S. Argounova, A. L. Epstein, and F. S. Markland. 2000. Molecular cloning and functional expression of contortrostatin, a homodimeric disintegrin from southern copperhead snake venom. Arch Biochem Biophys 375:278-288.

Zink, R. M., and G. F. Barrowclough. 2008. Mitochondrial DNA under siege in avian phylogeography. Mol Ecol 17:2107-2121.