The LINNEAN SOCIETY of London ZOOLOGICAL Journal of the Linnean Society

Zoological Journal of the Linnean Society, 2014. With 5 figures

Considering gene flow when using coalescent methods to delimit lineages of North American pitvipers of the genus *Agkistrodon*

FRANK T. BURBRINK[1,2]* and TIMOTHY J. GUIHER[3]

[1]*Biology Department 6S-143, 2800 Victory Blvd., College of Staten Island/CUNY, Staten Island, NY 10314, USA*
[2]*Biology Doctoral Program, City University of New York, Graduate Center 365 Fifth Avenue, New York, NY 10016-4309, USA*
[3]*Herpetology Department American Museum of Natural History, New York, NY 10024, USA*

Received 21 May 2014; revised 2 September 2014; accepted for publication 5 September 2014

Examining species diversity and mechanisms of speciation using coalescent models provides a framework for how regional diversity is accrued, even in well-studied areas such as the Nearctic. It is likely, that gene flow among closely-related species with adjacent distributions may be common. However, the absence of gene flow is a primary assumption of many phylogeographical methods that produce species trees and delimit species using Bayesian or likelihood functions in a coalescent framework. In the present study, we examine delimitation when gene flow between species is present using empirical datasets from two species of North American pitvipers of the genus *Agkistrodon*. We also use niche modelling to determine whether these young lineages occur in distinct environmental niches. To manage the problem of gene flow between species, we first identify admixed individuals, demonstrate that gene flow has occurred, and then identify the impact of alternative population assignments of admixed individuals on delimitation posterior probabilities. In addition, we examine the influence of mitochondrial genes relative to other loci combined in coalescent analyses that delimit species. Here, we find that the copperheads (*Agkistrodon contortrix*) and the cottonmouths (*Agkistrodon piscivorus*) are each composed of two distinct species, with each occupying different niches. Importantly, we find that species can be delimited when the amount of gene flow between lineages is low, although the methods are acutely sensitive to population assignment of individuals.

ADDITIONAL KEYWORDS: BPP – migration – phylogeography – species delimitation.

INTRODUCTION

Not only is the known species diversity likely a small fraction of actual total diversity on the earth, but also the processes that generate this diversity remain largely unknown (Dirzo & Raven, 2003; Mora *et al*., 2011). Even in well-explored industrial regions of the world, such as the Nearctic, cryptic species of vertebrates are still being described or properly recognized (Myers *et al*., 2013; Ruane *et al*., 2014). In the eastern Nearctic, for example, numerous biogeographical barriers to dispersal, such as the Appalachian Mountains, Mississippi River, and the historical isolation of the Florida peninsula, have been identified as areas that generate species via allopatry (Burbrink, Lawson & Slowinski, 2000; Soltis *et al*., 2006; Burbrink *et al*., 2008; Pyron & Burbrink, 2009), although continuing gene flow across the barriers remains a possibility (Nosil, 2008). For the most part, properly assessing diversity, particularly with respect to cryptic vertebrates, can be performed using coalescent-based phylogeographical methods, which provide techniques to help delimit and assess processes of species diversification given multilocus data (Burbrink *et al*., 2011; Fujita *et al*., 2012; Ruane *et al*., 2014).

*Corresponding author. E-mail: frank.burbrink@csi.cunby.edu

Properly identifying species boundaries has never been straightforward and is perhaps one of the most controversial enterprises in systematics (De Queiroz, 2007). However, sampling markers throughout the genome and using models that account for different coalescent histories and discordance among loci provides more objective measures for assessing species relationships and delimiting taxa (Carstens & Knowles, 2007; Liu *et al*., 2008, 2009; Brito & Edwards, 2010; Heled & Drummond, 2010; McCormack, Zellmer & Knowles, 2010; Fujita *et al*., 2012; Leaché *et al*., 2014). Unfortunately, these methods usually do not account for alternative sources of gene tree discordance for species delimitation (e.g. gene flow, gene duplication) and thus are not readily considered (but see Yu *et al*., 2011). In addition, most methods that delimit species assume that assignment of individuals to taxa is known; usually derived from phylogeographical studies using single loci, such as mitochondrial DNA (mtDNA) (Birmingham & Moritz, 1998; Zink & Barrowclough, 2008), morphology or biogeography. Therefore, species assessment using multilocus data in a coalescent framework becomes verification of the previous assignment (Harrington & Near, 2012; Ruane *et al*., 2014), although Bayes factors can be used to delimit species and assess alternate assignment of individuals within species (Chen *et al*., 2014; Grummer, Bryson & Reeder, 2014). Still, this may problematic when there is no guide for some or all of the taxonomic groups being studied, or the guide is incorrect (O'Meara, 2010; Leaché & Rannala, 2011).

One assumption shared by all of the currently proposed delimitation methods is that gene flow has ceased upon speciation. Recent studies have suggested that both sympatric and parapatric species may continue to experience limited gene flow after speciation (Coyne & Orr, 2004; Hey, 2006; Nosil, 2008). Alternatively, allopatric populations may exchange genes after the two species come into secondary contact (Pedall *et al*., 2011). For cryptic species, it may be difficult to identify whether there has been gene flow between lineages because this first requires identifying the lineages. Therefore, this creates a protocol problem where species delimitation requires verification that no gene flow is occurring, yet assessment of gene flow requires that separate species or lineages be first delimited. This then raises the question: how do phylogeographical studies incorporate current methods to delimit species while accounting for multiple sources of gene discordance, such as deep coalescence and gene flow?

By examining the phylogeographical history of two viperids in the genus *Agkistrodon* endemic to the eastern Nearctic, we address these issues regarding coalescent species delimitation. The North American copperhead, *Agkistrodon contortrix*, and the cottonmouth, *Agkistrodon piscivorus*, are two closely-related species with largely overlapping ranges in the eastern and central USA (Fig. 1). These viperids are ecologically and medically important because they represent some of the most common venomous snakes in the eastern USA and, particularly the copperhead, account for the majority of human envenomations in the USA (Lavonas *et al*., 2004). Previously, Guiher & Burbrink (2008) and Douglas *et al*. (2009) examined the phylogeographical history of both species using mtDNA and identified two distinct lineages within *A. piscivorus* and three distinct lineages within *A. contortrix* that, in part, fall along previously defined subspecies boundaries (Gloyd & Conant, 1990). No obvious geographical barriers were present between any lineages, despite the presence of the Appalachian Mountains and the Mississippi River within the range of both species. However, we note that lineages of *A. piscivorus* may have been generated during the isolation of Florida into smaller islands from the continent during the Pliocene or Pleistocene and *A. contotrix* as a result of the transitions from eastern forests to the central grasslands (Bailey, 1995; Guiher & Burbrink, 2008); therefore, it is expected that testing niche differences among lineages may provide some evidence about habitat specialization. Without a clear physical boundary between lineages, it is possible that gene flow may be ongoing, obscuring our ability to delimit species. Recent diversification (approximately 0.96–2.50 Mya) among these lineages makes it more likely that there will be discordance between gene trees as a result of less time for genes to sort (Degnan & Rosenberg, 2009). Therefore, 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.

To determine whether these species can be delimited, we provide a framework to infer population structure, identify individuals with mixed ancestry, and, finally, delimit species using the Bayesian Phylogenetics and Phylogeography method (BPP, version 2.0; Yang & Rannala, 2010). Although no migration is a main assumption of this method and it is known gene flow can alter timing of divergence and population sizes (Leaché *et al*., 2014), it has been demonstrated by simulation that limited gene flow does not violate the performance of the software (Zhang *et al*., 2011). Our procedure here for delimiting species is a modified version of the three-step method proposed by Leaché & Fujita (2010) in which we infer population structure, assign individuals to populations, identify putative hybrids, estimate migration, and infer phylogeographical relationships among populations. Next, Bayesian species delimitation in BPP is carried out to test the effect of including genetically admixed

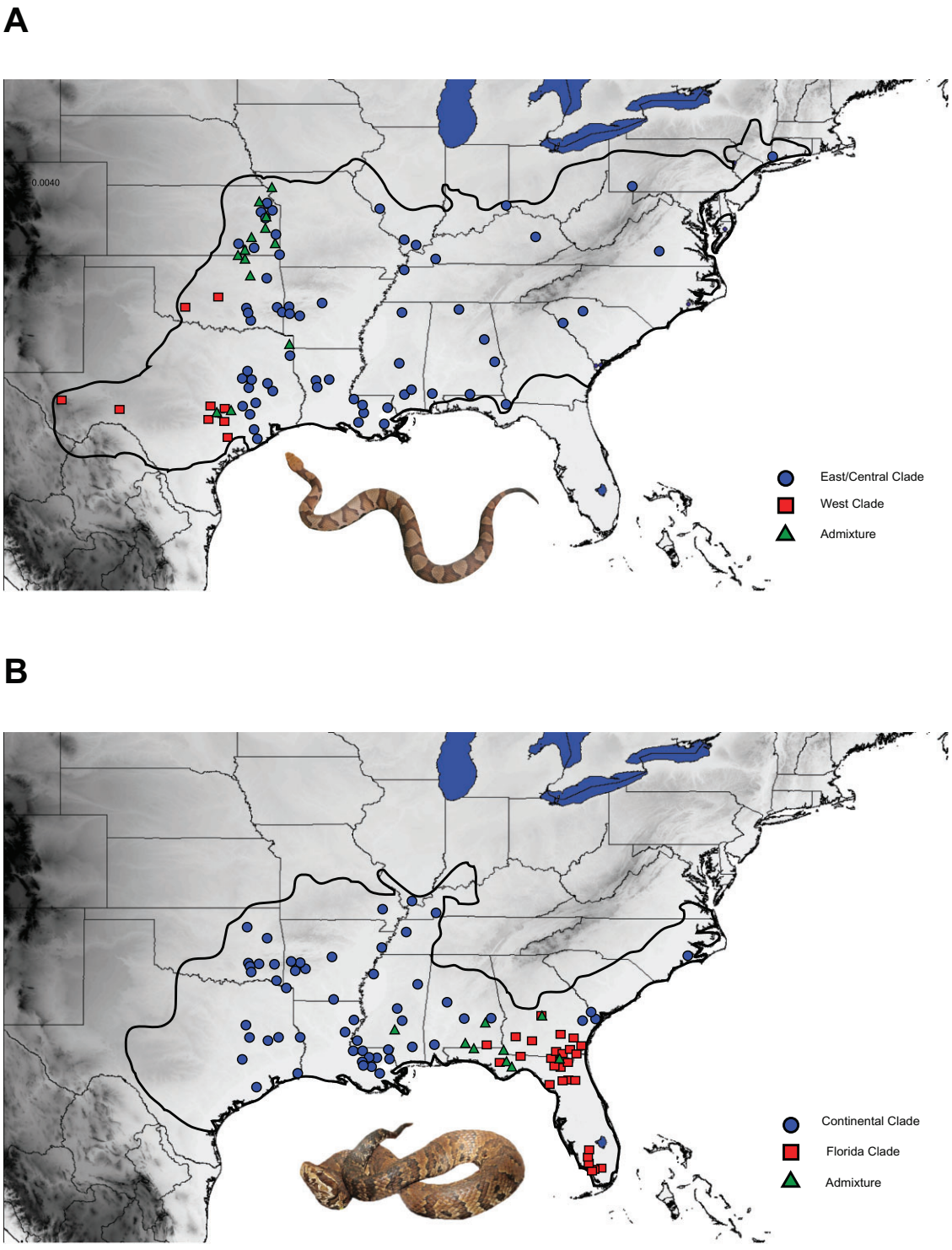

Figure 1. Map of the USA showing the sampling localities and distribution of inferred lineages of *Agkistrodon contortrix* (A) and *Agkistrodon piscivorus* (B). The distribution of both species is indicated by a bold outline. The key to the right of each map indicates the population membership of individuals and also identifies admixture in individuals.

individuals on the probability of support for species delimitation. This approach has the advantage of allowing us to first identify admixture between populations and assign clade membership to individuals and, second, to determine whether populations represent distinct species using an unbiased model based approach. Lastly, we use ecological niche modelling to determine whether lineages occupy distinctly different niches.

Because mtDNA has been the main driver of phylogeographical studies (Hurst & Jiggins, 2005) and thus plays an important role in delimiting cryptic taxa, we also examine how mtDNA genes may influence support for the probability of species delimitation when combined with multiple unlinked loci. Compared to nuclear genes, the effective population (N_e) size of mtDNA is much lower and the mutation rate can be several orders of magnitude higher (Brown, George & Wilson, 1979). This implies that a great number of nuclear genes may be required to obtain the same number of polymorphic sites present in many mitochondrial data sets. Therefore, it is plausible that the phylogenetic signal may be overwhelmed by mitochondrial data when loci are concatenated (Naylor & Brown, 1998). The ability of a gene to sort may be of greater importance than the number of informative sites in a coalescent analysis, provided that an actual gene tree can be inferred. In this case, where N_e is small, the propensity for sorting within shorter time frames is another reason for mtDNA to be the first marker of choice for most studies (Zink & Barrowclough, 2008; Edwards & Bensch, 2009). However, the influence of mitochondrial genes relative to other loci combined in coalescent analyses that delimit species is generally unknown. Therefore, we explore the implications of using datasets that include and exclude mtDNA and ultimately address proper methods for coalescent species delimitation within two species of pitvipers.

MATERIAL AND METHODS

Sampling and sequencing

We obtained 114 individuals of *A. contortrix* and 131 individuals of *A. piscivorus* broadly sampled 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/μL. Five nuclear genes were sequenced for each species specifically for this project and used in conjunction with the mitochondrial dataset (cytochrome *b*; Cyt *b*) reported in Guiher & Burbrink (2008). In addition, Cyt *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 & Diaz, 2010) and two previously identified single copy (scnDNA) nuclear loci (NT3 and SPTBN1). The sequences of *A. piscivorus* also comprised five nuclear genes but used different markers. A single anonymous locus was used (Anon 11; Gibbs & Diaz, 2010) and four scDNA loci (NT3: Noonan & Chippindale, 2006; SPTBN1: Mathee *et al*., 2001; Vimentin Intron 5: Zehner & Paterson, 1983; AHR: Townsend *et al*., 2008). The primers used for amplification were: 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′. In addition, for a broader species tree analyses to determine relationships within *Agkistrodon*, we also sequenced the scnDNA locus Skeletal muscle sodium channel Intron 5 ($NA_v1.4$; Geffeney *et al*. 2005) and included these with the previously mentioned loci using the primers for amplification and sequencing: NAV5F 5′-GGG CAA CGT CTC TGC TCT AC-3′, NAV6R 5′-CGA AGT TCC CCA TGA ACA GT-3′. Amplification of each gene was carried out using GoTaq Green Master Mix (Promega Corp). A 90-s extension time was used for each gene with the annealing temperatures: Anon A (51 °C), Anon 11 (44 °C), Anon 51 (43 °C), NT3 (51 °C), SPTBN1 (51 °C), Vimentin Intron 5 (47 °C), and AHR (45 °C). 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: 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, version 3.6 (Edgar, 2004) in GENEIOUS, version 5.1 (Biomatters Ltd). Heterozygous genotypes were resolved using PHASE, version 2.1.1 (Stephens & Donnelly, 2003) with default setttings and these phased sequences were used for all downstream analyses.

Population structure and phylogenetic inference

Species delimitation methods implemented in BPP, version 2.0 (Yang & Rannala, 2010) 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, we followed the approach outlined by Leaché & 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. To estimate the number of populations and assign individuals to these populations, we used STRUCTURE, version 2.3.3 (Pritchard, Stephens & Donnelly, 2000; Falush, Stephens & Pritchard, 2003, 2007; Hubisz *et al*., 2009) on the phased alleles for each species. We acknowledge the potential shortcomings using STRUCTURE to generate species groups for testing in BPP (Olave, Solà & Knowles, 2013), although we used other estimates such as previous analyses of phylogeographical structure with respect to previously named subspecies (Guiher & Burbrink, 2008) to compare our results from STRUCTURE. To establish the number of populations within each species, five independent runs were performed varying the number of K populations from 2 to 6 for each run. Posterior Bayes Factors (PBF) was used to identify the optimal value for K. Two subsequent runs were performed for 1×10^6 generations with a burn-in period of 500 000 generations. STRUCTURE was also used to identify individuals with mixed ancestry and to determine the proportion of alleles inherited from each population for admixed individuals. We 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 because STRUCTURE includes an 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, we constructed species trees for each species using *BEAST, version 1.7.0 (Heled & Drummond, 2010). In addition, *Crotalus horridus* (two individuals) was included as an outgroup, as well as both Cantil species (*Agkistrodon bilineatus* and *Agkistrodon taylori*), to illustrate relationships within the genus *Agkistrodon*. Total loci used consisted of Cyt *b* and eight nuclear loci mentioned above (Anon A, Anon 11, Anon 51, NT3, SPTBN1, Vimentin Intron 5, AHR, and NAV5).

The most appropriate model of evolution for each gene in the *BEAST runs was determined using Akaike information criterion (AIC) and Bayesian information criteria (BIC) in JMODELTEST (Posada, 2008). Two individuals from each population were used in place of the entire dataset to avoid excessive runtimes. The analysis was run for 200 million generations, sampling every 20 000 generations, and discarding the initial 20% as burn-in. Convergence was determined when estimated sample sizes (ESS) were > 200 for each parameter using TRACER, version 1.5.0 (Rambaut & Drummond, 2007). Bayesian posterior probabilities (*Pp*) greater than 95% were considered strong support for a clade (Felsenstein, 2004). The resulting guide tree and population memberships were subsequently used to parameterize analyses of species delimitation in BPP.

Migration

To estimate migration rate between adjacent lineages, we used IMA (Hey & Nielsen, 2007). It is necessary to provide estimates of mutation rates (substitutions/locus year^{-1}) to scale population size (N_e), time (t), and migration rates in number of years. Priors on mutation rates were based on estimates from the species tree. A generation time of 3.0 years was used for both species (Gloyd & Conant, 1990). We applied the Hasegawa–Kishino–Yano (HKY) model (Hasegawa *et al*., 1985) to each gene and an inheritance scalar of 0.25 and 1.0 was applied to mtDNA and the five nuclear loci, respectively. Independent runs were performed with different seeds in the Markov chain Monte Carlo (MCMC) to estimate the unscaled priors for Θ of each population q_1, q_2, m_1, m_2, and τ, and to determine appropriate values for the geometric heating of 30 chains. We ran the final dataset four times, visiting 8 million trees with the priors: *A. contortrix*: $q_1 = 20$, $q_2 = 20$, $q_a = 25$, $m_1 = 10$, $m_2 = 20$, and $\tau_1 = 10$; *A. piscivorus*: $q_1 = 5$, $q_2 = 5$, $q_a = 40$, $m_1 = 20$, $m_2 = 20$, and $\tau_1 = 10$. We used log-likelihood ratio tests (2LLR) to compare nested

models in the '–L' mode of IMA. This analysis used the chi-squared (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 were evaluated where migration was effectively = 0. We also used AIC to discriminate between multiple models not rejected by 2LLR.

Species delimitation

We used BPP, version 2.0 (Yang & Rannala, 2010) to estimate the probability of species delimitation and explored the effect of including individuals with mixed ancestry by comparing the speciation probabilities between runs comprising an increasing number of admixed individuals. Here, species delimited in accordance with the general lineage species concept (De Queiroz, 2007). It is possible to demonstrate the behaviour of BPP using simulations that include hybridization, and we note that this has been carried out previously (Zhang *et al*., 2011), although the present study demonstrates the ability to delimit with gene flow using empirical data. Each run consisted of 40 samples with admixed individuals representing 0–90% (increasing by 10% from 0%) of each dataset. 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 that represented the lesser proportion of alleles. To ensure that the reversible jump (rj)MCMC was mixing properly, we took two precautions. First, preliminary runs were performed implementing algorithm 0 with values for the fine-tuning parameter $\varepsilon = 2$, 5, 10, 15, and 20 to ensure proper mixing of the rjMCMC between models. The results were consistent among all runs and sufficient mixing of the rjMCMC was achieved with $\varepsilon = 2$. Therefore, algorithm 0 with $\varepsilon = 2$ was 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 (q) and standard deviation (σ) for each parameter were calculated using values for all six loci. We then determined values for α using the equation $\alpha = (m\ s^{-1})^2$ and for β using the equation $\beta = m\ s^{-2}$. The prior distributions used for each species were: *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, we determined whether speciation probabilities were overly influenced by the signal from individual genes, particularly mtDNA. Including all of the individuals in the analyses proved impossible because of a limit of 200 unique alleles in the current version of BPP. Therefore, we 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, first with all six loci (mtDNA and nDNA) and second with only the five nuclear loci. The same conditions and priors specified above were used for all runs.

Niche modelling

The potential distributions for the East, Central, and West 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 MAXENT, version 3.3.3e (Phillips, Anderson & Schapire, 2006; Phillips & 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 environmental niche models (ENMs). Models were trained using georeferenced localities for all individuals employed in the molecular analysis by lineage. Several separate analyses were performed for each lineage to explore the effect of two specific modelling decisions outlined previously (Elith *et al*., 2011): background selection and sampling bias in the training data. Background selection has been shown to impact the distribution predicted by MAXENT (VanDerWal *et al*., 2009). The modelling decision of what landscape to draw background points is often a factor of the question being addressed (e.g., the environmental factors that determine the distribution of a species or how a species utilizes partitioned microclimates), which includes the full range of environmental conditions available to a species (with this possibly being limited by barriers to dispersal) and sampling effort (Elith *et al*., 2011). We explored two candidate backgrounds: (1) 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 (2) the eastern USA, which includes only the environmental conditions inhabited by species of *Agkistrodon*

and may better identify environmental conditions unique to each lineage. We also examined the effect of including genetic hybrids 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. Because neither species included in the present 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, we attempted to correct for sampling bias by downloading every colubroid snake record from the HerpNet database (http://www.herpnet.org/) that included locality data with latitude and longitude within the USA. Environmental data for all nineteen BIOCLIM variables were extracted for all 17 555 samples in DIVA-GIS, version 7.3.0. (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, ultimately having a potentially negative impact on model performance. In addition, we constructed ENMs without correcting for bias by sampling background from the original 19 BIOCLIM layers over both the entire USA 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, we performed 10 replicate runs using a different random seed and cross-validation, 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. We 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). We compared AUC values between the four analyses for each lineage, which included different modelling 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 with models that excluded hybrids. The 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 cross-validated replicates were used.

We also used the niche identity test and background tests in ENMTOOLS, version 1.3 (Warren, Glor & Turelli, 2008, 2010) to determine whether the lineages differed significantly with respect to niche as defined by the nineteen bioclim variables. The niche identity test determines whether two lineages inhabit identical niches, whereas the background test examined whether niches are similar. Niche similarity between all adjacent lineages in each species was assessed using two metrics: Schoener's *D* (Schoener, 1968) and the *I* of Warren *et al*. (2008). The niche identity test determines whether two lineages inhabit identical niches by comparing observed values of *D* and *I* with a distribution of randomized pseudoreplicates generated by randomly assigning samples to either lineage. A one-tailed *t*-test was used to determine whether 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* with 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.

RESULTS

SEQUENCING

A total of 3269 bp obtained from six loci were sequenced for *A. contortrix* with the lengths (Table 1): Anon 11 (340 bp), Anon 51 (322 bp), Anon A (249 bp), Cyt *b* (1107 bp), NT3 (480 bp), and Sptbn1 (771 bp). There were no gaps in the alignments of any of the six loci. The highest number of variable sites was contained by the single mtDNA gene Cyt *b* (71), the five nuclear loci combined contained 38 variable sites ranging from 3–13 per locus. A total of 3666 bp was obtained from the six loci sequenced for *A. piscivorus* with the lengths: Anon (236 bp), Cyt *b* (1107 bp), Nt3 (455 bp), Sptbn1 (938 bp), Vimentin Intron 5 (539 bp), and AHR (391 bp). The Skeletal muscle sodium channel Intron 5 included in Species Tree Inference was 457 bp long. There were no gaps identified in any of the six loci, with the highest number of variable sites contained by the mtDNA gene Cyt *b* (76 bp); the five nuclear loci contained 24 variable sites combined ranging from 1–11 per locus. All sequences produced for the present study are available from the Dryad Digital Repository: http://doi.org/10.5061/dryad.v2n38.

Table 1. Summary statistics for the six loci sequenced for *Agkistrodon contortrix* and *Agkistrodon piscivorus*

Species	Locus	*N*	Number of haplotypes	Nucleotide diversity (π)	Mean number of pairwise differences (k)
Agkistrodon 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
Agkistrodon 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

Population structure and phylogenetic inference

Two populations were identified by STRUCTURE for *A. contortrix* and two for *A. piscivorus*, supported by PBFs and a plateau of marginal likelihoods at $K = 2$. The populations identified by STRUCTURE using all six loci correspond to the geographically distinct mtDNA lineages reported in Guiher & Burbrink (2008), with the exception of the East and Central lineages of *A. contortrix* being combined into a single population (Fig. 1). There were 17 individuals identified with mixed ancestry between the two populations of *A. contortrix*, representing 15% of the dataset, and eight 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 the geographical interface between sister lineages (Fig. 1).

Because analyses of population structure identified only two populations for both species, inferring phylogenetic relationships between populations was unnecessary. The guide tree for both instances would be a single bifurcating tree representing a single speciation event and a subtree resulting from collapsing this node. However, our previous study on *A. contortrix* using mtDNA identified three lineages (Guiher & Burbrink, 2008) and therefore it was necessary to infer phylogenetic relationships to examine the influence of mtDNA on species delimitation. In this case, a guide tree could have been constructed based on the phylogeographical relationships suggested by mtDNA, although we chose to corroborate those relationships with a species tree using multiple nuclear genes in combination with mtDNA. 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, we determined that the most appropriate substitution models were GTR+Γ, HKY+Γ, HKY+Γ, GTR+Γ, HKY, HKY+Γ, HKY+Γ+I, GTR, and HKY+Γ for Cyt *b*, AHR, Anon 11, Anon 51, Anon A, NAV5, NT3, Sptbn1, and Vimentin Intron 5, respectively. Discarding the first 10% (20 million generations) as burn-in resulted in ESS values above 200 for all parameters. The resulting species tree recovered the same relationships between the three lineages previously suggested by mtDNA (Guiher & Burbrink, 2008) with strong support for an East lineage, a Central lineage, and a West lineage (Fig. 2). Therefore, the guide tree ((East, Central), West)) was used for *A. contortrix* (Fig. 2) and a single bifurcating tree for *A. piscivorus* when examining differences in speciation probabilities.

Migration

We combined four independent runs for each analysis resulting in a total of 80 000 samples producing ESS values for all parameters > 200 following a burn-in of 100 000 generations. Likelihood ratio tests failed to reject three models in favour of the full model between the two populations of *A. contortrix*. 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 (see Supporting information, Appendix S1). Divergence between the two lineages was consistent with estimates from species trees, maximum likelihood (ML) estimate of 419 kya (95% highest posterior density = 65 kya to 1.605 Mya), potentially indicating that Pleistocene glacial cycles may have played a role in speciation. Equal migration in both directions was inferred with a ML estimate of 6.6×10^{-6} migrants per generation per gene copy (95% confidence interval = 1.66×10^{-6} to 1.706×10^{-5}). Effective population size

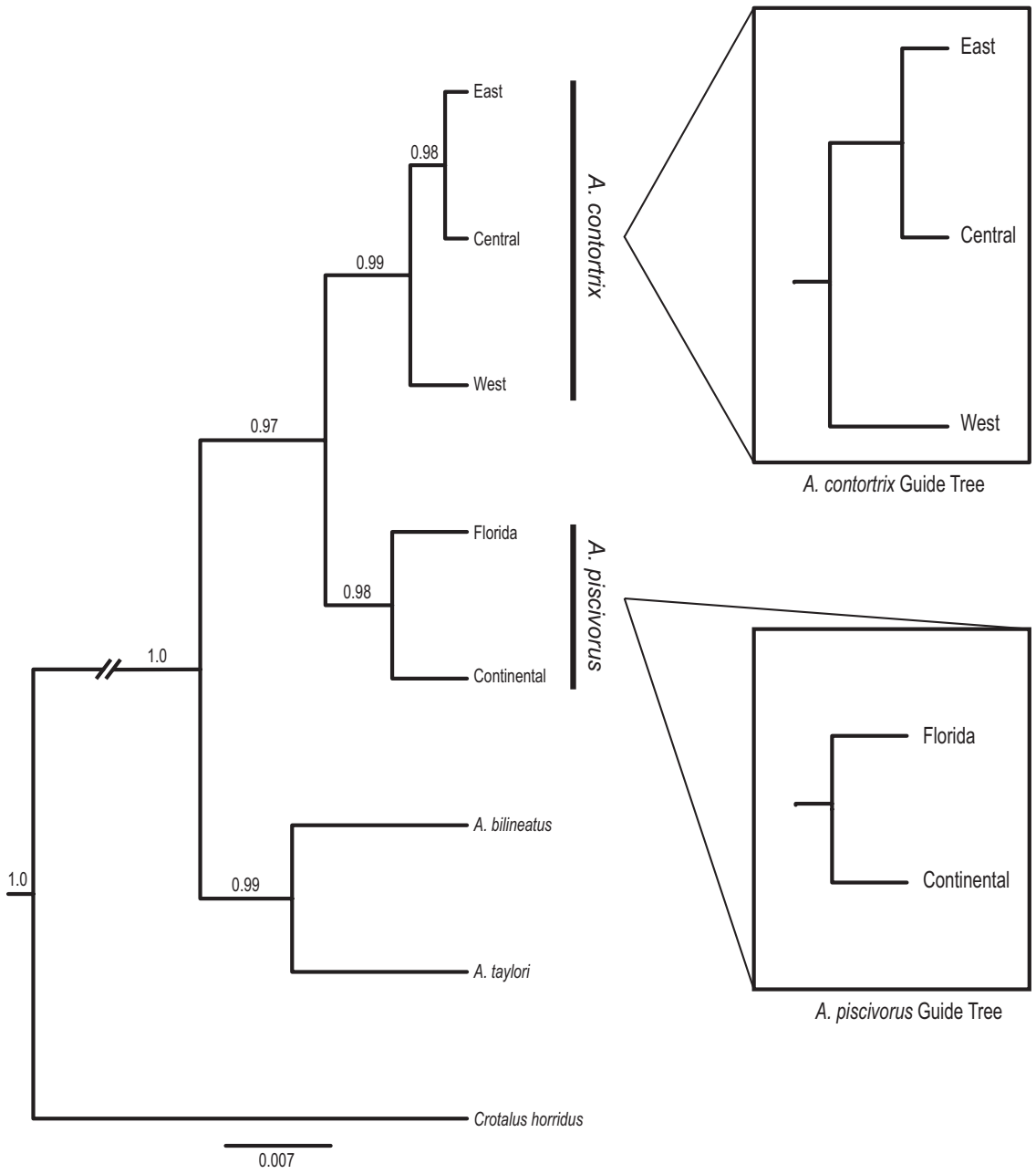

Figure 2. Species tree showing the phylogenetic relationships within the genus *Agkistrodon* implemented in *BEAST (Heled & Drummond, 2010). Posterior probability support for nodes is indicated above branches. The guide trees used for subsequent species delimitation in BPP (Yang & Rannala, 2010) are shown in the insets.

of the extant species is predicted to be much smaller than the ancestral population (see Supporting information, Appendix S2). Ten models were not rejected by LLR investigating migration between the two lineages of *A. piscivorus*; AIC could not easily discriminate between five models given low ΔAIC values (see Supporting information, Appendix S1). 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 the Continental lineage is equal to the ancestral population size, and the population size of the Florida lineage equals the ancestral population size. The fifth model suggests that distinctly different population sizes were attributed to each lineage and migration occurred in one direction from the Florida lineage into the Continental lineage (see Supporting information,

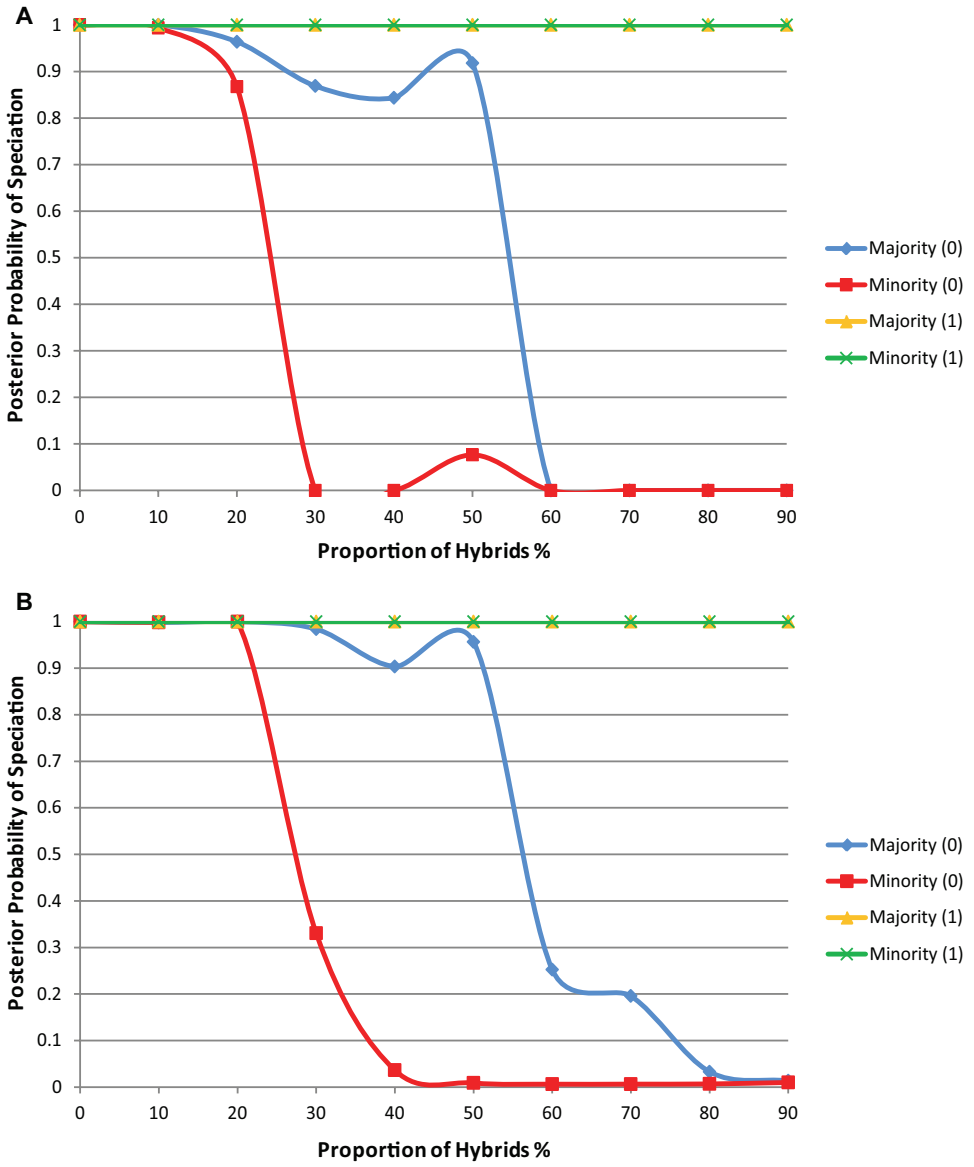

Figure 3. Posterior probability of speciation inferred by BPP, version 2.0 (Yang & Rannala, 2010) with an increasing proportion of hybrids included in the dataset for *Agkistrodon contortrix* (A) and *Agkistrodon piscivorus* (B). Results are shown for both species with hybrids assigned to the population that contributed the majority of alleles with a fully resolved starting tree (1) and a fully collapsed starting tree (0), as well as hybrids assigned to the population that contributed the minority of alleles with both starting trees.

Appendix S1). A migration rate of 8.6 6×10^{-6} gene copies per generation (95% confidence interval = 2.26×10^{-6} to 2.586×10^{-5}) was inferred; however, it was not possible to discriminate between equal migration in both directions and only migration from the Florida lineage into the Continental lineage. The Ancestral population size was estimated to be much larger than the population size of either extant species (see Supporting information, Appendix S2).

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 (Fig. 3; see also Supporting information, Appendix S3). Additionally, speciation probabilities were altered dependent upon which population the admixed

individuals were assigned. A noticeable reduction in the probability of delimiting species was observed when admixed individuals were included in the population that contributed the minority of alleles (see Supporting information, Appendix S3). Finally, as the proportion of admixture approached 20%, we found that ensuring proper mixing of the rjMCMC proved impossible. This was evident where runs beginning with different starting trees failed to reach the same conclusion with respect to the number of species supported. Therefore, 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), then results indicated 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. By contrast, assessing species delimitation within *A. contortrix* was significantly influenced by mtDNA. A western lineage was recovered with 100% support in all analyses with or without mtDNA. Analyses of the full six-gene dataset produced moderate support for the existence of a Central and East lineage with a mean speciation probability of 0.95 (range 0.81–1.0). A noticeable decrease in support for this speciation event was observed when Cyt *b* was removed from the dataset with a mean speciation probability of 0.46 (range 0.11–0.74).

Niche modelling

Using MAXENT, ecological niche models for both lineages produced test AUC values > 0.7 for all replicates (see Supporting information, Appendix S4). Higher AUCs 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 (see Supporting information, Appendix S4). This was likely a result of bias in colubroid sampling, which is oriented towards the western USA, far outside of the range of the Florida lineage (Fig. 4F). In addition, colubroid sampling within Florida was biased toward southern Florida (Fig. 4F). This was opposite to the sampling bias observed in the Florida lineage (Fig. 4C, D). Therefore, the method applied here was unlikely to account for the bias present in the sampling of this lineage. Sampling background points from the entire USA resulted in higher AUCs for every lineage (see Supporting information, Appendix S4). This was in contrast to restricting background points to the eastern USA, 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, we applied the minimum training presence (MTP) threshold to convert distributions to binary predictions of suitability because it had the added benefit of being biologically interpretable (Pearson *et al*., 2007). Binomial omission tests were not significant for several thresholds for ENMs 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 and maximum test sensitivity plus specificity, 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 was not unexpected given the low sample size ($N = 9$) available for this lineage, resulting in the use of only linear features. Including hybrids for the West lineage improved model performance. However, binomial omission tests were not significant for several threshold rules and exhibited a greater amount of variance in fractional area predicted compared to ENMs for the other lineages. Significant binary thresholds ($P > 0.05$) tended to either over fit or provide little discriminatory power for this lineage. 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 affected by the inclusion of hybrid samples resulting in a predicted hybrid zone extending across the mid-Atlantic coastal plains in southern North Carolina to the southern coastal plains in south-eastern Louisiana (Fig. 4C). Potential distributions for the Continental lineage of *A. piscivorus* and the combined East/Central lineage of *A. contortrix* were not affected by including hybrid samples. In both cases, potential hybrid zones were predicted as well, regardless of whether hybrids were used to train the model (Fig. 4). Comparisons could not be made for the East and Central lineages because genetic data did not identify putative hybrids. The variables with the greatest contribution to each model were dependent on parameterization of the model (Table 1).

Observed values of Schoener's *D* and Warren *et al*.'s *I* values were significantly lower than expected from a random distribution (see Supporting information, Appendix S5). 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

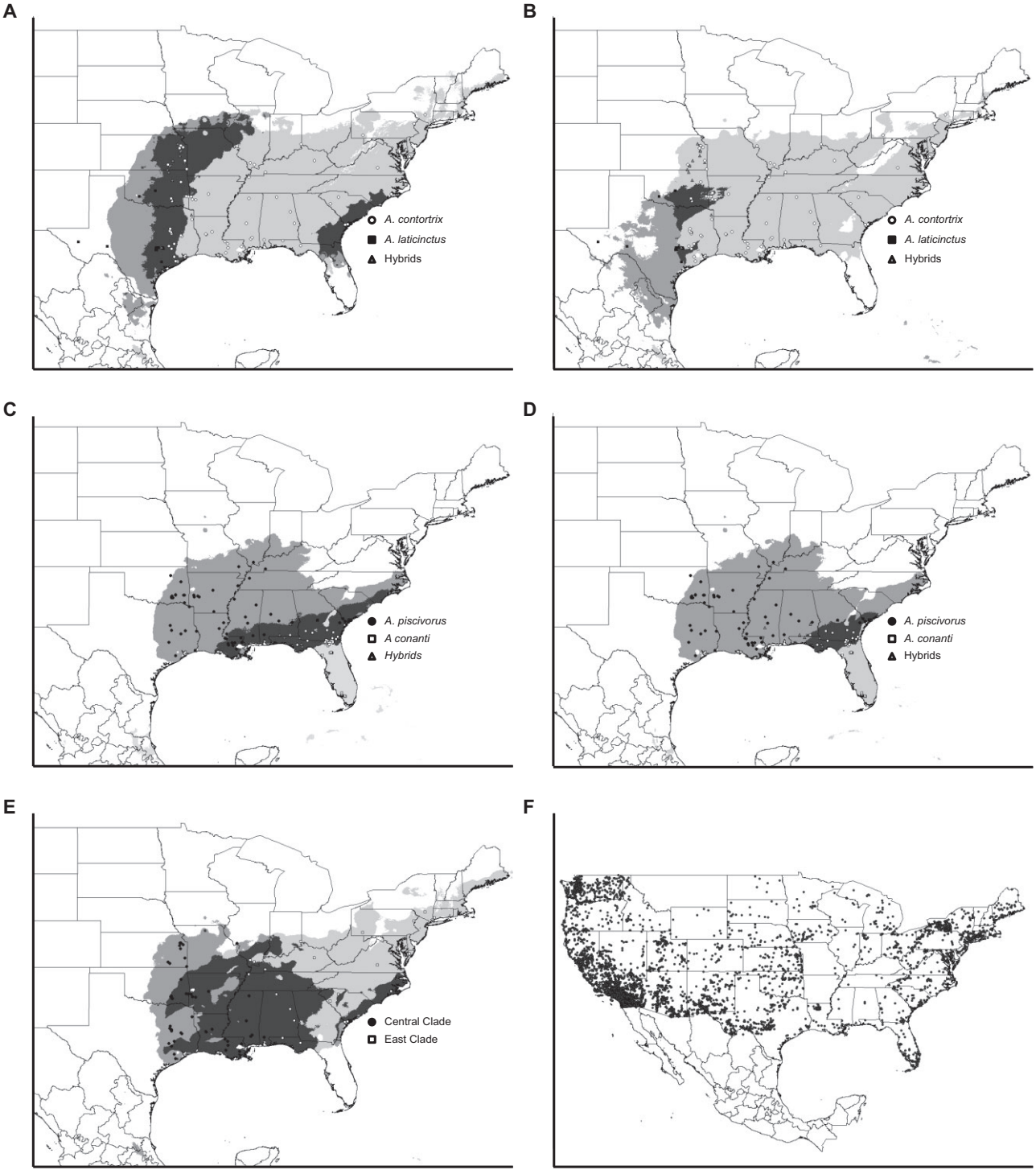

Figure 4. Niche predictions for each species of the *Agkistrodon contortrix* complex with hybrids included in training the model (A), *Agkistrodon contortrix* complex with hybrids excluded (B), *Agkistrodon piscivorus* complex with hybrids included in training the model (C), *Agkistrodon piscivorus* complex with hybrids excluded (D), and East and Central lineages of *Agkistrodon contortrix* inferred by mtDNA (E). The distribution of colubriod samples obtained from the Herpnet database used to correct for sampling bias is shown in (F). Overlap between predicted suitable habitat is indicated by the darkest intermediate shade.

Table 2. Statistics from the Niche background test for pairwise comparisons of adjacent clades within *Agkistrodon*

Lineage	Hybrids	$I(\bar{X}, \sigma, P)$	$D(\bar{X}, \sigma, P)$
East versus Central	NA	0.576 (0.639, 0.023, =0.289)	0.283 (0.330, 0.018, =0.297)
Central versus East	NA	0.576 (0.511, 0.039, =0.352)	0.283 (0.236, 0.028, =0.342)
East/Central versus West	Yes	0.631 (0.457, 0.079, =0.410)	0.348 (0.217, 0.059, =0.386)
West versus East/Central	Yes	0.631 (0.456, 0.030, =0.411)	0.348 (0.175, 0.021, =0.410)
East/Central versus West	No	0.551 (0.454, 0.075, =0.368)	0.291 (0.215, 0.057, =0.357)
West versus East/Central	No	0.551 (0.514, 0.015, =0.319)	0.291 (0.242, 0.011, =0.324)
Continental versus Florida	Yes	0.514 (0.482, 0.048, =0.335)	0.251 (0.243, 0.029, =0.323)
Florida versus Continental	Yes	0.514 (0.449, 0.023, =0.331)	0.251 (0.202, 0.017, =0.318)
Continental versus Florida	No	0.400 (0.415, 0.051, =0.312)	0.151 (0.190, 0.027, =0.301)
Florida versus 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 ± SD of the null distribution shown in parentheses. NA, not applicable.

test (Table 2). This indicates that sister lineages inhabit similar but not identical niches. As a second metric of niche similarity, we 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 compare distributions in geographical space, whereas D and I reflect differences in niche space. Second, calculating the amount of geographical overlap is dependent on binomial thresholds to convert distributions to binary predictions of suitability. Although ENMTOOLS does provide an option to apply binomial thresholds, caution is warranted. 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. 4, Table 3). As described above, hybrids had almost 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* were a result of variation in the Florida models, and not the ENMs for the Continental lineage. Estimates of overlap between the West lineage and combined East/Central lineage were similarly influenced by variation in ENMs for the West lineage (Fig. 4A, B, Table 3). The Central and East lineages overlapped considerably with one another geographically (Fig. 4E, Table 3).

Table 3. Area of predicted habitat suitability for the inferred lineages of *Agkistrodon contortrix* and *Agkistrodon piscivorus* with and without hybrids included in training data

Lineage	Hybrids	Area
East	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
West	Yes	1 306 498 km^2
West	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 versus Central	No	1 155 547 km^2
East/Central versus West	Yes	766 848 km^2
East/Central versus West	No	131 061 km^2
Continental versus Florida	Yes	329 096 km^2
Continental versus Florida	No	126 941 km^2

Niche overlap calculations for pairwise comparisons of adjacent clades.

Area calculated as the total number of 30 arc second pixels predicted as suitable using the minimum training presence threshold, multiplied by 0.86 (30 s of arc = 0.86 km^2). Overlap between potential distributions for pairwise comparisons of adjacent clades was calculated as the total area of the zone of predicted overlap in environmental suitability from the environmental niche models. NA, not applicable.

DISCUSSION

Delimiting species using molecular data can be complicated by the lack of lineage sorting across the genome, gene flow, the amount of admixture in individuals, and the size of hybrid zones. Coalescent methods to delimit species and infer phylogeographical relationships between species when incomplete lineage sorting is expected are becoming more common (Yang & Rannala, 2010). However, these methods are relatively new and it is unclear how they will perform when other sources of gene discordance are present, specifically hybridization. In addition, this creates a methodological problem where species delimitation requires that little or no

gene flow has occurred and yet assessment of gene flow requires that separate species or lineages be first delimited. We examine a framework that addresses this problem by first identifying individuals with mixed ancestry and finally delimiting species using coalescent delimitation methods. The empirical results presented here for the venomous snake genus *Agkistrodon* suggest that species can be successfully delimited when low or 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 made it possible to assign individuals to respective populations and had the added benefit of identifying admixed individuals. Bayesian species delimitation implemented in BPP assumes that gene flow is absent between species (Yang & Rannala, 2010). Therefore, a priori tests of gene flow are a crucial first step. Here, the results suggest that migration between lineages within copperheads and within is generally much lower than a single individual/generation and admixture was not common among lineages. As expected from simulations (Zhang *et al*., 2011), our results indicate that the effects of violating the assumption of gene flow reveal that species delimitation in BPP is possible even with moderate amounts of gene flow, although several factors affect the outcome on detecting species probabilities. Speciation probabilities decrease when an increasing amount of hybrids are included (Fig. 3; see also Supporting information, Appendix S3), although these general results vary given specific model parameterizations, which include population assignment of individuals with admixture and the starting tree used. Importantly, the 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 admixed individuals were assigned also affected the rate at which the posterior probability of speciation decreased with respect to increased gene flow (Fig. 3; see also Supporting information, Appendix S3). Specifically, 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*. By contrast, speciation probabilities decreased more rapidly and with fewer hybrids present when hybrids were assigned to the population that contributed the minority of alleles, 10% and 20% for *A. piscivorus* and *A. contortrix*, respectively (Fig. 3; see also Supporting information, Appendix S3).

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 to which population hybrids were assigned (Fig. 3; see also Supporting information, Appendix S3). By contrast, the trend of decreasing speciation probabilities with an increasing percentage of hybrids was observed when the starting tree was fully collapsed (Fig. 3; see also Supporting information, Appendix S3). This discrepancy is somewhat unexpected and requires exploration with a greater number of possible starting trees with more nodes than used here. However, to avoid inferring potentially inflated speciation probabilities, we 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 (Edwards, 2009; Degnan & Rosenberg, 2009). However, phylogeography has been dominated by mtDNA studies (Zink & Barrowclough, 2008) for the last two decades and understanding just how misleading reliance on this single marker has been for these studies is generally unknown, at least until such studies are revisited using multilocus data. We explored this question with two empirical datasets in the present 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 (Fig. 3; see also Supporting information, Appendix S3). By contrast, BPP suggests either two or three species of *A. contortrix* depending on whether the mtDNA gene Cyt *b* 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 (see Supporting information, Appendix S3). In addition, both the Central and East lineages were moderately supported with a mean posterior probability of 0.95 summed over 10 runs using all six genes (see Supporting information, Appendix S3). When Cyt *b* was excluded, the speciation probability for this node was reduced (0.46; see Supporting information, Appendix S3). This suggests that at least in the copperheads, lineage diversity may be overestimated by mtDNA. Alternatively, mtDNA may represent the correct phylogeographical history of *A. contortrix* but there just has not been sufficient time for the slower nuclear genes to accumulate substitutions or sort (Moore & Aug, 1995; Palumbi, Cipriano & Hare, 2001). It is important to note that divergence time estimates from mtDNA (Guiher & Burbrink, 2008; Douglas *et al*., 2009) suggest that the two lineages of *A. piscivorus* diverged at the end of the Pliocene or early Pleistocene (approximately 2.5–1.9 Mya) 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 (Wakeley, 2008; Degnan & Rosenberg, 2009). It is also possible that the discordance observed here is a result of relatively few nuclear genes used. 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 (Padial *et al*., 2010). However, differences among lineages regarding these types of data may not be possible for cryptic species or in taxa that have diverged rather recently (Fujita *et al*., 2012). Therefore, even with molecular data alone, challenges for delimiting species using coalescent methods with numerous unlinked loci in the face of lineage sorting, hybridization, recombination, and potentially gene duplication and loss still exist. Although no single method has been developed to simultaneously ameliorate all sources of discordance, we have shown that understanding and combining estimates of admixture with current coalescent delimitation methods can reduce the negative outcome of underestimating taxonomic diversity of cryptic species.

Ecological niche models provide additional support for taxonomic revision of *A. piscivorus* and *A. contortrix*. The two lineages of *A. piscivorus* have non-identical but similar ecological niches, supported by both *D* and *I*. Estimates of geographical overlap in the potential distributions of the two lineages suggest that they 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 (Figs 4, 5; Tables 2, 3). However, it is likely that this value is inflated by the small and restricted distribution of the Florida lineage. Indeed, 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 unexpected given that a number of studies have identified genetic breaks between Florida and mainland populations (Avise & Saunders, 1984; Ellsworth *et al*., 1994; Walker *et al*., 1998; Burbrink *et al*., 2008; Fontanella *et al*., 2008). Consistency between ENMs 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, although both Guiher & Burbrink (2008) and Douglas *et al*. (2009) found population sizes within this lineage to be stable through time. All three initial lineages of *A. contortrix* were found to inhabit similar but not identical ecological niches according to both Schoener's *D* and Warren *et al*.'s *I*. However, considerable geographical overlap between the potential distributions of the East lineage and Central lineage was estimated; representing 53% and 69% of their distributions, respectively (Fig. 4E), supporting our coalescent delimitation results that these may not be distinct species. 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 (Figs 4, 5). By contrast to Guiher & Burbrink (2008) and Douglas *et al*. (2009), ENMs and phylogeographical analyses suggest that the Mississippi River likely had no part in generating these lineages, and rather that habitat transitions from eastern forests to central grasslands probably drove diversification within the copperheads.

Based on the phylogenetic and ecological evidence presented, we propose that two species of cottonmouth and two species of copperhead should be recognized. The two lineages of *A. piscivorus* were found to inhabit distinct ecological niches that overlap along a putative hybrid zone from northern Florida to southern Georgia, Alabama, and Mississippi (Fig. 5). Ecological niche models and coalescent analyses failed to provide compelling support for recognizing the East and Central lineages of *A. contortrix* as distinct species. 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 geographical overlap, which suggests that the potential distributions of the two lineages are not geographically separate. Therefore, we suggest that *A. contortrix* be separated into two species: one comprising the combined East and Central lineages and the other represented by the West lineage, with a potential hybrid zone from Texas to Kansas (Fig. 5).

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 contortionists, possible reference to dorsal pattern.

Synonymy: This species comprises the previously recognized subspecies *Agkistrodon contortrix contortrix*

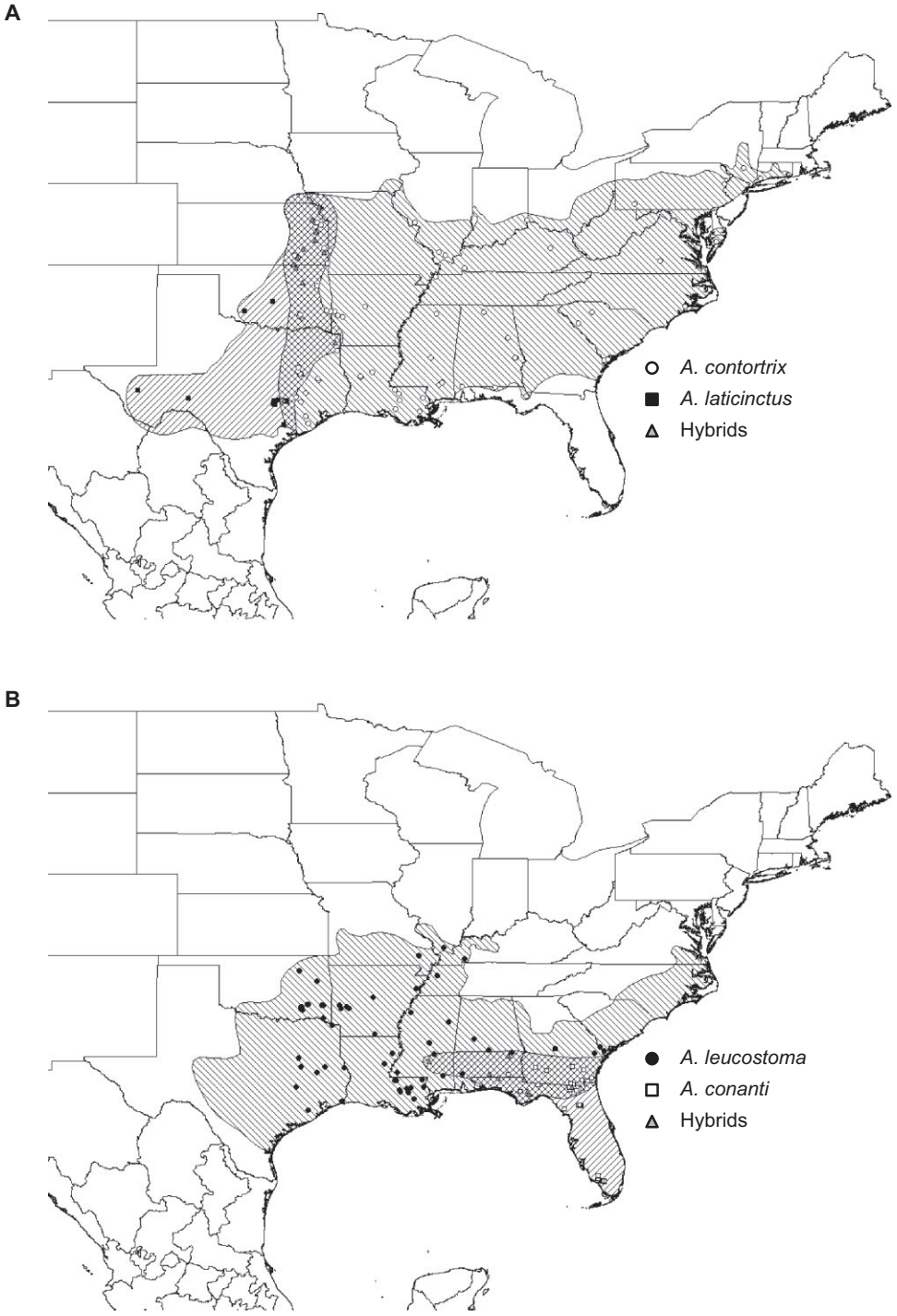

Figure 5. The distribution of species within the *Agkistrodon contortrix* complex (A) and *Agkistrodon piscivorus* complex (B) inferred by Bayesian species delimitation. Putative hybrids zones are identified with cross-hatches. Distributions are adapted from Gloyd & Conant (1990) and the results from ecological niche modelling.

(Linnaeus, 1766), *Agkistrodon contortrix mokasen* (Palisot de Beauvois, 1799), and *Agkistrodon contortrix phaeogaster* (in part; Gloyd, 1969).

Diagnosis: Combining characteristics for the subspecies *A. c. contortrix, A. c. mokason*, and *A. phaegaster*, we provide a diagnosis for *A. contortrix*. We find that the eastern copperhead (*A. contortrix*) is a medium-sized pit viper with an average adult size of 61–90 cm and a maximum size of 132.1 cm (Gloyd & Conant, 1990; Conant & Collins, 1991), with a ratio of tail to total length of 0.11–0.17 in males and 0.11–0.16 in females. There is a single anal plate, keeled dorsal scales and typically 23 midbody scale rows (range 21–25;Gloyd & Conant, 1990). Subcaudals range from 38 to 59 in males and from 37 to 49 in females, whereas ventral scales number 139–157 with no variation between sexes (Gloyd & Conant, 1990). Supralabials and infralabials range from 6 to 10 (mode 8) and from 8 to 12 (mode 10), respectively, and total postoculars + suboculars range from 3 to 6 (mode 4; Gloyd & Conant, 1990). A combination of geography and colour pattern distinguishes the eastern copperhead from related species. There are 10–21 dark cross-bands on a brown, tan or grey background (or very dark or black individuals) in males (10–17 in *A. laticinctus*) and 11–20 in females (10–18 in *A. laticinctus*). Cross-bands are narrow at the midbody and widen toward the sides, described as hourglass or dumbbell shaped, and may be broken at middorsum (Gloyd & Conant, 1990), whereas they generally do not narrow at the midbody in *A. laticinctus*. The head is coppery brown or reddish–brown with a pale cheek stripe (Gloyd & 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, where *A. laticinctus* ranges from eastern Texas north to eastern Kansas (Fig. 5A). Hybridization with the broad-banded copperhead apparently occurs along an ecological transition zone from south-eastern Texas north to eastern Kansas, defined by an 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 without additional molecular or morphological data and analyses (Fig. 5A).

AGKISTRODON LATICINCTUS (GLOYD & 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 colour pattern, derived from Latin *latus* and *cinctus*, translated as 'side' or 'broad' and 'banded', respectively.

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

Diagnosis: Combining characteristics for the subspecies *A. c. laticinctus* and *A. c. pictigaster*, we provide a diagnosis for *A. laticinctus*. The broad-banded copperhead (*A. laticinctus*) is a medium-sized pit viper with an average adult size of 56–76 cm and a maximum size of 95 cm (Gloyd & Conant, 1990; Conant & Collins, 1991), with a ratio of tail to total length of 0.11–0.18 in males and 0.11–0.16 in females. There is a single anal plate, keeled dorsal scales, and typically 23 midbody scale rows (range 21–25; Gloyd & Conant, 1990). Subcaudals range from 38 to 62 in males and from 39 to 57 in females, whereas ventral scales number 138–155 with no variation between sexes (Gloyd & Conant, 1990). Supralabials and infralabials range from 7 to 10 (mode 8) and 8 to 12 (mode 10), respectively, and total postoculars + suboculars range from 2 to 6 (mode 4; Gloyd & Conant, 1990). A combination of geography and colour pattern distinguishes the broad-banded copperhead from related species. There are 10–18 brown to red cross-bands in males (10–21 in *A. contortrix*) and 10–18 in females (11–20 in *A. contortrix*) on a light to medium- to light-brown background. Cross-bands do not narrow at midbody (Gloyd & Conant, 1990), whereas they generally narrow at the midbody in *A. contortrix*. Head coloration typically matches that of the cross-bands with a lighter cheek stripe (Gloyd & Conant, 1990). The approximate distribution of the *A. laticinctus* includes western and central Texas, central Oklahoma, and eastern Kansas, whereas the range of *A. contortrix* extends from Connecticut to Florida along the Atlantic coast, west to include eastern Texas north through eastern Oklahoma into eastern Kansas (Fig. 5A). Hybridization between the broad-banded copperhead and the eastern copperhead occurs along an ecological transition zone from south eastern Texas north to eastern Kansas, defined by an 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 without additional morphological data (Fig. 5A).

AGKISTRODON PISCIVORUS (LACEPÈDE 1789)

Northern Cottonmouth
Holotype: Unknown.

Type locality: 'Carolina' (Lacépède, 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* (Lacépède, 1789) and *A. p. leucostoma* (Troost, 1836).

Diagnosis: Combining characteristics for the subspecies *Agkistrodon piscivorus piscivorus* and *Agkistrodon piscivorus leucostoma*, we provide a diagnosis for *A. piscivorus*. The Northern Cottonmouth (*A. piscivorus*) is a medium- to large-bodied semi-aquatic pit viper with an average adult size of 76–114 cm and a maximum size of 188 cm (Gloyd & Conant, 1990; Conant & Collins, 1991), with a ratio of tail to total length of 0.13–0.19 in males and 0.12–0.18 in females. They possess a single anal plate, keeled dorsal scales and typically 25 midbody scale rows (range 23–27; Gloyd & Conant, 1990). Subcaudals range from 38 to 53 in males and from 42 to 53 in females, whereas ventral scales number 128–142 with no variation between sexes (Gloyd & Conant, 1990). Supralabials and infralabials range from 6 to 10 (mode 8) and 8 to 12 (mode 11), respectively and total postoculars + suboculars range from 1 to 4 (mode 3; Gloyd & Conant, 1990). A combination of geography and colour pattern distinguishes the Northern Cottonmouth from the Florida Cottonmouth. There are 10–17 dark cross-bands on an olive, brown or black background, whereas *A. conanti* features 11–16 crossbands. Crossbands in *A. piscivorus* are often indistinguishable from the ground colour in adults, yet may be prominent or subdued in adult *A. conanti*. The head is typically black or brown, lacking vertical rostral stripes; a dark cheek stripe is present in juveniles but may be subdued or indistinguishable from the ground colour in adults yet often present in adult *A. conanti* (Gloyd & Conant, 1990; Conant & Collins, 1991). The Northern Cottonmouth occupies lower elevations throughout the south-eastern USA 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, whereas the Florida cottonmouth (*A. conanti*) ranges from southern Florida to approximately Savannah Georgia, and west to south-eastern Alabama (Fig. 5B). Hybridization with the Florida Cottonmouth occurs 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 without additional morphological and molecular data (Fig. 5B).

AGKISTRODON CONANTI (GLOYD, 1969)

Florida Cottonmouth

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

Type locality: Seven miles south-east 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 *Agkistrodon piscivorus conanti*.

Diagnosis: The Florida Cottonmouth (*A. conanti*) is diagnosed from the subspecies *A. p. conanti*. This species is a medium- to large-bodied semi-aquatic pit viper with an average adult size 76–122 cm and a maximum size of 189.2 cm (Gloyd & Conant, 1990; Conant & Collins, 1991), with a ratio of tail to total length of 0.15–0.19 in males and 0.13–0.18 in females. There is a single anal plate, keeled dorsal scales and typically 25 midbody scale rows (range 23–27; Gloyd & Conant, 1990). Subcaudals range from 45 to 54 in males and from 41 to 49 in females, whereas ventral scales number 135–145 in males and 132–144 in females (Gloyd & Conant, 1990). Supralabials and infralabials range from 6 to 10 (mode 8) and 9 to 12 (mode 10), respectively, and total postoculars + suboculars range from 2 to 4 (mode 3; Gloyd & Conant, 1990). A combination of geography and colour pattern distinguishes the Florida cottonmouth from related species. There are 11–16 dark cross-bands on an olive, brown or black background, which may become subdued in adults, whereas *A. piscivorus* generally has 10–17 cross-bands that often become indistinguishable from the ground colour in adults (Gloyd & Conant, 1990). The head is typically brown with vertical stripes along the snout on the rostrals, prenasals, and first supralabials (Gloyd & Conant, 1990). Dark stripes appear on the lower jaw extending from the mental to the first four or five infralabials (Gloyd & Conant, 1990). A dark cheek stripe is present bordered above and below by pale stripes and often present in adults, although it is often indistinguishable from the ground colour in adult *A. piscivorus* (Gloyd & Conant, 1990; Conant & Collins, 1991). The distribution of the Florida Cottonmouth extends from southern Florida to approximately Savannah, Georgia, and west to south-eastern Alabama (Fig. 5B), whereas the Northern Cottonmouth ranges in the USA from south-eastern Virginia to central Georgia, east of the Appalachian Mountains, north to southern Illinois and eastern Kansas. Hybridization between the Florida and Northern cottonmouth occurs in the mid-Atlantic coastal plains in southern North

Carolina to the southern coastal plains in southeastern Louisiana and diagnosis may be difficult without additional morphological and molecular data for some individuals in this area.

ACKNOWLEDGEMENTS

We would like to thank the following individuals and institutions for providing tissues and invaluable assistance: D. B. Shepard and L. J. Vitt (Sam Noble Oklahoma Museum of Natural History); T. W. Taggart, C. J. Schmidt, and J. T. Collins (Sternberg Museum, Fort Hays State University); R. Lawson (California Academy of Sciences); J. A. McGuire and C. Spencer (Museum of Vertebrate Zoology); J. Boundy (Louisiana State University Museum of Zoology); D. R. Frost, C. J. Raxworthy, and 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, and K. L. Krysko (University of Florida); D. C. Cannatella, T. J. LaDuc, and D. Hall (University of Texas, Austin); B. I. Crother (Southeastern Louisiana University); B. Moon (University of Lousiana, Lafayette); C. Phillips (Illinois Natural History Survey); S. Trauth (Arkansas State University Museum of Zoology); B. Green and R. Combs (Missouri State University); P. Chippendale and C. Rolke (University of Texas at Arlington); C. Guyer and S. Graham (Auburn University); N. Russel (South Florida Collection Management Center Everglades National Park); D. Stevenson (Project Orianne Ltd); K. Swaim (Swaim Biological Inc.); K. Wiley and J. Harrison (Kentucky Reptile Zoo); F. M. Fontanella, N. Howe, M. Harris, C. D. Camp, B. Hudson, A. Schwanabeck, K. K. J. Irwin, L. K. Irwin, C. G. Jimenez, S. R. McNearney, S. Ruane, and 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. This research was supported, in part, under National Science Foundation Grants CNS-0958379 and CNS-0855217, as well as the City University of New York High Performance Computing Center at the College of Staten Island. We would also like to extend special thanks to the following individuals for providing useful criticisms and suggestions during the research and completion of the manuscript: A. Pyron, S. Boissinot, M. Hickerson, C. J. Raxworthy, and A. Carnaval. Funding was provided, in part, by L. Clampitt, S. Harris, and D. Rosenberg.

REFERENCES

Avise JC, Saunders NC. 1984. Hybridization and introgression among species of sunfish (*Lepomis*) – analysis by mitochondrial-DNA and allozyme markers. *Genetics* **108:** 237–255.

Bailey RG. 1995. *Description of the ecoregions of the United States*. Washington, DC: USDA Forest Service.

Birmingham E, Moritz C. 1998. Comparative phylogeography: concepts and applications. *Molecular Ecology* **7:** 367–369.

Brito PH, Edwards SV. 2010. Multilocus phylogeography and phylogenetics using sequence-based markers (vol 135, pg 439, 2009). *Genetica* **138:** 465.

Brown WM, George M, Wilson AC. 1979. Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America* **76:** 1967–1971.

Burbrink FT, Fontanella F, Pyron RA, Guiher TJ, Jimenez C. 2008. Phylogeography across a continent: the evolutionary and demographic history of the North American racer (Serpentes?: Colubridae?: *Coluber constrictor*). *Molecular Phylogenetics and Evolution* **47:** 274–288.

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

Burbrink FT, Yao H, Ingrasci M, Bryson R, Guiher TJ, Ruane S. 2011. Speciation at the Mogollon Rim in the Arizona mountain kingsnake (*Lampropeltis pyromelana*). *Molecular Phylogenetics and Evolution* **60:** 445–454.

Carstens BC, Knowles LL. 2007. Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: an example from Melanoplus grasshoppers. *Systematic Biology* **56:** 400–411.

Chen X, Jiang K, Guo P, Huang S, Rao D, Ding L, Takeuchi H, Che J, Zhang YP, Myers EA, Burbrink FT. 2014. Assessing species boundaries and the phylogenetic position of the rare Szechwan ratsnake, *Euprepiophis perlaceus* (Serpentes: Colubridae), using coalescent-based methods. *Molecular Phylogenetics and Evolution* **70:** 130–136.

Conant R, Collins JT. 1991. *A field guide to reptiles and amphibians?: eastern and central North America*. Boston, MA: Houghton Mifflin.

Coyne JA, Orr HA. 2004. *Speciation*. Sunderland, MA: Sinnauer Associates Inc.

De Queiroz K. 2007. Species concepts and species delimitation. *Systematic Biology* **56:** 879–886.

Degnan JH, Rosenberg NA. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology and Evolution* **24:** 332–340.

Dirzo R, Raven PH. 2003. Global state and biodiversity loss. *Annual Review of Environment and Resources* **28:** 137–167.

Douglas ME, Douglas MR, Schuett G, Porras L. 2009. Global climate change and evolution of the New World pitviper genus Agkistrodon (Viperidae). *Journal of Biogeography* **36:** 1164–1180.

Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *Bmc Bioinformatics* **5:** 1–19.

Edwards S, Bensch S. 2009. Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough 2008. *Molecular Ecology* **18:** 2930–2933.

Edwards SV. 2009. Is a new and general theory of molecular systematics emerging? *Evolution* **63:** 1–19.

Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ. 2011. A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions* **17:** 43–57.

Ellsworth DL, Honeycutt RL, Silvy NJ, Bickham JW, Klimstra WD. 1994. Historical biogeography and contemporary patterns of mitochondrial DNA variation in white-tailed deer from the Southeastern United States. *Evolution* **48:** 122–136.

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

Falush D, Stephens M, Pritchard JK. 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes* **7:** 574–578.

Felsenstein J. 2004. *Inferring Phylogenies*. Sunderland, MA: Sinauer Associates, 664.

Fontanella FM, Feldman CR, Siddall ME, Burbrink FT. 2008. Phylogeography of *Diadophis punctatus*: extensive lineage diversity and repeated patterns of historical demography in a trans-continental snake. *Molecular Phylogenetics and Evolution* **46:** 1049–1070.

Fujita MK, LeachÈ AD, Burbrink FT, McGuire JA, Moritz C. 2012. Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology & Evolution* **27:** 480–488.

Geffeney SL, Fujimoto E, Brodie ED, Ruben PC. 2005. Evolutionary diversification of TTX-resistant sodium channels in a predator-prey interaction. *Nature* **434:** 759–763.

Gibbs HL, Diaz J. 2010. Identification of single copy nuclear DNA markers for North American pit vipers. *Molecular Ecology Resources* **10:** 177–180.

Gloyd HK. 1969. Two additional subspecies of North American crotalid snakes, genus *Agkistrodon*. *Proceedings of the Biological Society of Washington* **90:** 1002–1015.

Gloyd HK, Conant R. 1934. The broad-banded copperhead: a new subspecies of *Agkistrodon* mokasen. *Occasional papers of the University of Michigan Museum of Zoology* **183:** 1–6.

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

Gloyd HK, Conant R. 1990. *Snakes of the* Agkistrodon *complex: a monographic review*. St Louis, MO: Society for the Study of Amphibians and Reptiles.

Grummer JA, Bryson RW, Reeder TW. 2014. Species delimitation using bayes factors: simulations and application to the *Sceloporus scalaris* species group (Squamata: Phrynosomatidae). *Systematic Biology* **63:** 119–133.

Guiher TJ, Burbrink FT. 2008. Demographic and phylogeographic histories of two venomous North American snakes of the genus *Agkistrodon*. *Molecular Phylogenetics and Evolution* **48:** 543–553.

Harrington RC, Near JT. 2012. Phylogentic and coalescent strategies of species delimitation in snubnose darters (Percidae: *Etheostoma*). *Systematic Biology* **61:** 63–79.

Hasegawa M, Kishino H, Yano TA. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **22:** 160–174.

Heled J, Drummond AJ. 2010. Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution* **27:** 570–580.

Hey J. 2006. Recent advances in assessing gene flow between diverging populations and species. *Current Opinion in Genetics & Development* **16:** 592–596.

Hey J, Nielsen R. 2007. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences of the United States of America* **104:** 2785–2790.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25:** 1965–1978.

Hijmans RJ, Guarino L, Cruz M, Rojas E. 2001. Computer tools for spatial analysis of plant genetic resources data?: 1 . DIVA-GIS. *Plants Genetic Resources Newsletter* **127:** 15–19.

Hubisz MJ, Falush D, Stephens M, Pritchard JK. 2009. Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources* **9:** 1322–1332.

Hurst GD, Jiggins FM. 2005. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings Biological Sciences* **272:** 1525–1534.

Lacépède BGE d l v. 1789. *Histoire naturelle des quadrpedes ovipares et des serpens*. Hotel de Thou: Par.

Lavonas EJ, Gerardo CJ, O'Malley G, Arnold TC, Bush SP, Banner W, Steffens M, Kerns WP. 2004. Initial experience with Crotalidae polyvalent immune Fab (ovine) antivenom in the treatment of copperhead snakebite. *Annals of Emergency Medicine* **43:** 200–206.

Leaché AD, Fujita MK. 2010. Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). *Proceedings of the Royal Society of London Series B, Biological Sciences* **277:** 3071–3077.

Leaché AD, Harris RB, Rannala B, Yang Z. 2014. The influence of gene flow on species tree estimation: a simulation study. *Systematic Biology* **63:** 17–30.

Leaché AD, Rannala B. 2011. The accuracy of species tree estimation under simulation: a comparison of methods. *Systematic Biology* **60:** 126–137.

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

Liu L, Pearl DK, Brumfield RT, Edwards SV. 2008. Estimating species trees using multiple-allele DNA sequence data. *Evolution* **62:** 2080–2091.

Liu L, Yu L, Kubatko L, Pearl DK, Edwards SV. 2009. Coalescent methods for estimating phylogenetic trees. *Molecular Phylogenetics and Evolution* **53:** 320–328.

Mathee CA, Burzlaff JD, Taylor JF, Davis SK. 2001. Mining the mammalian genome for artiodactyl systematics. *Systematic Biology* **50:** 367–390.

McCormack JE, Zellmer AJ, Knowles LL. 2010. Does niche divergence accompany allopatric divergence in aphelocoma jays as predicted under ecological speciation?: insights from tests with niche models. *Evolution* **64:** 1231–1244.

Moore WS, Aug N. 1995. Inferring phylogenies from mtDNA variation – mitochondrial gene trees versus nuclear gene trees. *Evolution* **49:** 718–726.

Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B. 2011. How many species are there on earth and in the ocean? *PLoS Biology* **9:** e1001127.

Myers EA, Rodriguez-Robles JA, Denardo DF, Staub RE, Stropoli A, Ruane S, Burbrink FT. 2013. Multilocus phylogeographic assessment of the California Mountain Kingsnake (*Lampropeltis zonata*) suggests alternative patterns of diversification for the California Floristic Province. *Molecular Ecology* **22:** 5418–5429.

Naylor GJ, Brown WM. 1998. Amphioxus mitochondrial DNA, chordate phylogeny, and the limits of inference based on comparisons of sequences. *Systematic Biology* **47:** 61–76.

Noonan BP, Chippindale PT. 2006. Dispersal and vicariance: the complex evolutionary history of boid snakes. *Molecular Phylogenetics and Evolution* **40:** 347–358.

Nosil P. 2008. Speciation with gene flow could be common. *Molecular Ecology* **17:** 2103–2106.

O'Meara BC. 2010. New heuristic methods for joint species delimitation and species tree inference. *Systematic Biology* **59:** 59–73.

Olave M, Solà E, Knowles LL. 2013. Upstream analyses create problems with DNA-based species delimitation. *Systematic Biology* **63:** 263–271.

Padial JM, Miralles A, De la Riva I, Vences M. 2010. The integrative future of taxonomy. *Frontiers in Zoology* **7:** 16.

Palisot de Beauvois AMFJ. 1799. Memoir on Amphibia. Serpents. *Transactions of the American Philosophican Society* **4:** 362–381.

Palumbi SR, Cipriano F, Hare MP. 2001. Predicting nuclear gene coalescence from mitochondrial data: the three-times rule. *Evolution; International Journal of Organic Evolution* **55:** 859–868.

Pearson RG, Raxworthy CJ, Nakamura M, Peterson AT. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography* **34:** 102–117.

Pedall I, Fritz U, Stuckas A, Valdeon A, Wink M. 2011. Gene flow across secondary contact zones of the Emys orbicularis complex in the Western Mediterranean and evidence for extinction and re-introduction of pond turtles on Corsica and Sardina (Testudines: Emydidae). *Journal of Zoologic Systematics & Evolutionary Research* **49:** 44–57.

Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190:** 231–259.

Phillips SJ, Dudik M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* **31:** 161–175.

Posada D. 2008. jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution* **25:** 1253–1256.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155:** 945–959.

Pyron RA, Burbrink FT. 2009. Lineage diversification in a widespread species: roles for niche divergence and conservatism in the common kingsnake, *Lampropeltis getula*. *Molecular Ecology* **18:** 3443–3457.

Rambaut A, Drummond AJ. 2007. *Tracer v.1.5*. Available at: http://beast.bio.ed.ac.uk/Tracer

Ruane S, Bryson RW, Pyron RA, Burbrink FT. 2014. Coalescent species delimitation in milksnakes (genus *Lampropeltis*) and impacts on phylogenetic comparative analyses. *Systematic Biology* **63:** 231–250.

Schmidt KP. 1953. *A check list of North American amphibians and reptiles*. Chicago, IL: University of Chicago Press.

Schoener TW. 1968. Anolis lizards of bimini – resource partitioning in a complex fauna. *Ecology* **49:** 704–727.

Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS. 2006. Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology* **15:** 4261–4293.

Stephens M, Donnelly P. 2003. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics* **73:** 1162–1169.

Swets JA. 1988. Measuring the accuracy of diagnostic systems. *Science* **240:** 1285–1293.

Townsend TM, Alegre RE, Kelley ST, Wiens JJ, Reeder TW. 2008. Rapid development of multiple nuclear loci for phylogenetic analysis using genomic resources: an example from squamate reptiles. *Molecular Phylogenetics and Evolution* **47:** 129–142.

Troost G. 1836. On a new genus of serpents, and two new species of the genus Heterodon, inhabiting Tennessee. *Annals of Lyceum of Natural History of New York* **3:** 174–190.

VanDerWal J, Shoo LP, Graham C, Williams SE. 2009. Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? *Ecological Modelling* **220:** 589–594.

Wakeley J. 2008. *Coalescent theory: an introduction*. Greenwood Village, CO: Roberts & Company Publishers.

Walker D, Moler PE, Buhlmann KA, Avise JC. 1998. Phylogeographic patterns in *Kinosternon subrubrum* and *K. baurii* based on mitochondrial DNA restriction analyses. *Herpetologica* **54:** 174–184.

Warren DL, Glor RE, Turelli M. 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* **62:** 2868–2883.

Warren DL, Glor RE, Turelli M. 2010. ENMTools: a toolbox for comparative studies of environmental niche models. *Ecography* **33:** 607–611.

Yang Z, Rannala B. 2010. Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences of the United States of America* **107:** 9264–9269.

Yu Y, Than C, Degnan JH, Nakleh L. 2011. Coalescent histories on phylogenetic networks and detection of hybridization despite incomplete lineage sorting. *Systematic Biology* **60:** 138–149.

Zehner ZE, Paterson BM. 1983. Characterization of the chicken vimentin gene: single copy gene producing multiple mRNAs. *Proceedings of the National Academy of Sciences of the United States of America* **80:** 911–915.

Zhang C, Zhang DX, Zhu T, Yang Z. 2011. Evaluation of a bayesian coalescent method of species delimitation. *Systematic Biology* **60:** 747–761.

Zink RM, Barrowclough GF. 2008. Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology* **17:** 2107–2121.

ARCHIVED DATA

DNA sequences and locality data are archived in the Dryad Digital Repository: http://doi.org/10.5061/dryad.v2n38.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Appendix S1. a, tests of nested models in IMa for the East and West lineages of *Agkistrodon contortrix*. b, tests of nested models in IMa for the Continental and Florida lineages of *Agkistrodon piscivorus*.

Appendix S2. Population parameter estimates in IMA for the *Agkistrodon contortrix* complex and *Agkistrodon piscivorus* complex.

Appendix S3. Posterior probability of speciation inferred by BPP, version 2.0 when the number of individuals with admixture comprised between 10% and 90% of the dataset.

Appendix S4. Threshold independent model performance calculated by AUC for ENMs of the five lineages of *Agkistrodon contortrix* and *Agkistrodon piscivorus*.

Appendix S5. Niche identity statistics for pairwise comparisons of adjacent clades of *Agkistrodon contortrix* and *Agkistrodon piscivorus*.