Syst. Biol. 60(4):393–409, 2011

DOI:10.1093/sysbio/syr011
Advance Access publication on March 9, 2011

Inferring Species-Level Phylogenies and Taxonomic Distinctiveness Using Multilocus Data in *Sistrurus* Rattlesnakes

LAURA S. KUBATKO[1,2,*], H. LISLE GIBBS[2,3], AND ERIK W. BLOOMQUIST[3]

[1]*Department of Statistics, The Ohio State University, Columbus, OH 43210, USA;*
[2]*Departent of Evolution, Ecology and Organismal Biology, The Ohio State University, Columbus, OH 43210, USA; and*
[3]*Mathematical Biosciences Institute, The Ohio State University, Columbus, OH 43210, USA;*
[*]*Correspondence to be sent to: Department of Statistics, The Ohio State University, 404 Cockins Hall, 1958 Neil Avenue, Columbus, OH 43210-1247, USA; E-mail: kubatko.2@osu.edu.*

Received 8 March 2010; reviews returned 5 June 2010; accepted 20 January 2011
Associate Editor: Elizabeth Jockusch

Abstract.—Phylogenetic relationships and taxonomic distinctiveness of closely related species and subspecies are most accurately inferred from data derived from multiple independent loci. Here, we apply several approaches for understanding species-level relationships using data from 18 nuclear DNA loci and 1 mitochondrial DNA locus within currently described species and subspecies of *Sistrurus* rattlesnakes. Collectively, these methods provide evidence that a currently described species, the massasauga rattlesnake (*Sistrurus catenatus*), consists of two well-supported clades, one composed of the two western subspecies (*S. c. tergeminus* and *S. c. edwardsii*) and the other the eastern subspecies (*S. c. catenatus*). Within pigmy rattlesnakes (*S. miliarius*), however, there is not strong support across methods for any particular grouping at the subspecific level. Monophyly based tests for taxonomic distinctiveness show evidence for distinctiveness of all subspecies but this support is strongest by far for the *S. c. catenatus* clade. Because support for the distinctiveness of *S. c. catenatus* is both strong and consistent across methods, and due to its morphological distinctiveness and allopatric distribution, we suggest that this subspecies be elevated to full species status, which has significant conservation implications. Finally, most divergence time estimates based upon a fossil-calibrated species tree are >50% younger than those from a concatenated gene tree analysis and suggest that an active period of speciation within *Sistrurus* occurred within the late Pliocene/Pleistocene eras. [Gene and species trees; genealogical species concept; monophyly-based tests for species delimitation; multilocus phylogenetics; *Sistrurus* rattlesnakes; species tree–based divergence times.]

Delimiting the boundaries and phylogenetic relationships of recently evolved species is critical to understanding the pattern and timing of lineage formation in adaptive radiations (Schluter 2000). Until recently, most analyses in vertebrates have focused on using single-gene trees derived from one or more mitochondrial DNA (mtDNA) genes to identify species using tree-based species criteria (e.g., from reptiles: Parkinson et al. 2000; Serb et al. 2001; Burbrink 2002; Wiens and Penkrot 2002). However, there is increasing recognition that methods designed specifically to handle multiple nuclear loci can provide significant advantages over single or concatenated gene analyses by 1) directly estimating a species tree rather than relying on inferences of species relationships from gene trees (Liu and Pearl 2007; 2) allowing the use of new methods for delimiting species based on the probabilistic assessment of lineage relationships among putative species across multiple gene trees (Knowles and Carstens 2007; Rosenberg 2007; Carstens and Dewey 2010; and 3) enabling more accurate estimates of key parameters (e.g., species divergence times) that are important in understanding the processes of speciation in a particular group (Jennings and Edwards 2005). However, the methods by which information from multiple independent genes is combined and then used to generate such information are only beginning to be widely utilized and several current methods suffer from statistical and methodological drawbacks (as noted by Carstens and Knowles 2007; Liu and Pearl 2007; Edwards 2009).

Species-tree methods seek to handle the potential mismatch between individual gene trees and the actual species phylogeny (Knowles and Kubatko 2010). Although a number of biological processes can lead to incongruence between gene trees and species trees (see Maddison 1997; Carstens and Knowles 2007), the process of incomplete lineage sorting is a potential source of discord in all data sets. Incomplete lineage sorting occurs when recently diverged lineages retain ancestral polymorphism because they have not had sufficient time to achieve reciprocal monophyly (Hudson 1992; Hudson and Coyne 2002; Rosenberg 2003). Compared with mtDNA, this phenomenon is especially problematic for anonymous nuclear DNA loci (Brumfield et al. 2003) and introns (Friesen et al. 1999) because the effective population size of these loci is, on average, four times greater than that of a typical mtDNA locus. As a result, lineage sorting takes much longer to occur (Hudson and Turelli 2003; Rosenberg 2003).

To handle this issue, a number of approaches have been proposed to infer species trees from multiple genes (see reviews in Carstens and Knowles 2007; Liu and Pearl 2007; Liu et al. 2009a). Among these, approaches that directly incorporate the process of lineage sorting into the phylogeny estimation procedure (Carstens and Knowles 2007; Edwards et al. 2007; Liu and Pearl 2007; Kubatko et al. 2009; Liu et al. 2009b; Heled and Drummond 2010) are most promising because they explicitly model the discord between gene trees and species trees that results from the incomplete

lineage sorting process. Here, we use two of these approaches, species tree estimation using maximum likelihood (STEM; Kubatko et al. 2009) and Bayesian Evolutionary Analysis Sampling Trees (*BEAST; Heled and Drummond 2010), to examine species relationships from multilocus data for a recent radiation of rattlesnakes. The two methods differ in the type of input data used to infer the species tree. STEM uses the gene trees estimated from individual genes to compute the maximum likelihood (ML) estimate of the species tree under the coalescent model; *BEAST uses Markov chain Monte Carlo (MCMC) to estimate the posterior distribution of the species tree given the multilocus alignment. *BEAST also assumes the randomness of the effective population sizes and places a hierarchical prior on them, whereas STEM uses a user-supplied value for the effective population sizes.

Here, we explore the use of these methods to analyze species relationships for a relatively large (19 loci) multilocus data set with respect to our ability to recover species relationships and to examine how these patterns compare with those generated from the widely used concatenation method (Kubatko and Degnan 2007). In addition, we test for the taxonomic distinctiveness of species and subspecies (defined here as genealogical exclusivity; Baum and Shaw 1995) using two approaches: 1) A novel statistical test for taxonomic distinctiveness (Rosenberg 2007) based on comparison of observed levels of monophyly across multiple individual gene trees to levels expected under the null hypothesis of a single taxonomic entity and 2) The genealogical sorting index (*gsi*; Cummings et al. 2008) that generates a quantitative measure of the degree to which the ancestry of the various species and subspecies is exclusive. These analyses address an important emerging issue in the field, namely how can phylogenetic information be used to delimit species despite widespread incomplete lineage sorting (Knowles and Carstens 2007; Carstens and Dewey 2010; O'Meara 2010; Yang and Rannala 2010)?

Phylogenetics of Recently Evolved Snake Species and Subspecies

Molecular phylogenetic analyses of closely related species and subspecies of snakes have largely been based upon sequence data from one or more mitochondrial genes (Burbrink et al. 2000; Pook et al. 2000; Burbrink 2002; Douglas et al. 2002, 2006; Bryson et al. 2007; Castoe et al. 2007; Guiher and Burbrink 2008). This approach has yielded new insights into the phylogenetic distinctiveness (or lack thereof) of morphologically defined subspecies (Burbrink et al. 2000; Burbrink 2002; Douglas et al. 2002) and the phylogeography and historical demography of individual species (Douglas et al. 2006; Castoe et al. 2007). On the basis of significant phylogenetic information in mtDNA, new species designations have been proposed based on a genealogical or phylogenetic species concept (Burbrink et al. 2000; Burbrink 2002). However, there is an increasing general recognition that inferences from gene trees based solely on mtDNA need to be corroborated with phylogenetic or population genetic information from nuclear loci (Hudson and Coyne 2002; Schelly et al. 2005; Gibbs et al. 2006; Gonçalvesa et al. 2007; Linnen and Farrell 2007). In the past generating information for sequence-based nuclear loci was difficult because of a lack of available loci for snakes. The development of new methods for rapidly isolating anonymous loci from nonmodel species (Brumfield et al. 2003) and the availability of primers to amplify conserved introns in snakes (Creer et al. 2005) make this much less of a problem (see Gibbs and Diaz 2010). As a result, information from multiple nuclear DNA loci has been increasingly incorporated into studies of higher-level snake systematic relationships (Vidal and Hedges 2002; Wiens et al. 2008). However, multilocus analyses of snakes at species and subspecies levels are rare (but see Douglas et al. 2007) despite the fact that such work would provide much-needed corroboration of the phylogenetic and demographic patterns inferred from mtDNA alone as well as yield additional information on the speciation history of the group of snakes under study. Recent applications of species tree–based analyses to lizards (Leaché 2009, 2010) illustrate the promise of this approach for squamate reptiles.

Here, we use data from 18 recently developed nuclear DNA loci (Gibbs and Diaz 2010) and a mitochondrial gene fragment to analyze phylogenetic relationships among two species comprised of a total of six subspecies of *Sistrurus* rattlesnakes. Rattlesnakes are New World pitvipers (Viperidae; Crotalinae) with species falling into one of two genera that were diagnosed on the basis of distinctive head scale morphology (Gloyd 1940; Klauber 1972): *Crotalus* that contains 20 or more named species (see reviews in Gloyd 1940; Klauber 1972; Murphy et al. 2002) and *Sistrurus* that consists of just three named species: *catenatus*, *miliarius*, and *ravus* (Gloyd 1940). However, recent phylogenetic analyses of rattlesnakes as a whole using mtDNA have suggested that *ravus* is in fact a species of *Crotalus* while confirming the distinctiveness of *S. catenatus* and *S. miliarius* (Murphy et al. 2002; Parkinson et al. 2002). Here, we focus on relationships among subspecies of *catenatus* and *miliarius* alone.

Both species contain three subspecies formally described by Gloyd (1935, 1940, 1955) on the basis of morphological variation in scale characters, body size and coloration, and geographic distribution. Within *S. catenatus*, the eastern subspecies (*S. c. catenatus*) is distinguished from the other two subspecies by its dark ventral coloration, lower number of ventral scales and lower number of dorsal blotches (Gloyd 1940) and has a range in eastern North America from Missouri to central New York (Fig. 1). The other two subspecies (*tergeminus* and *edwardsii*) have ranges in the central and southwestern regions of the United States, respectively (Fig. 1), with *tergeminus* being larger, darker in color, and having

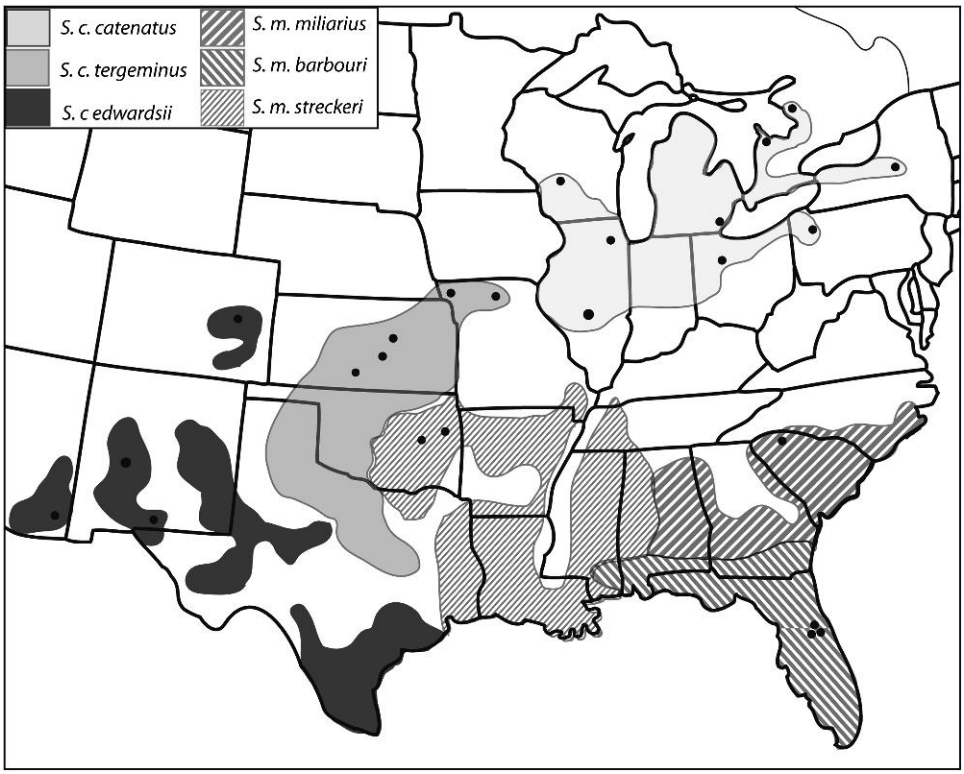

FIGURE 1. Map showing the geographic distributions of each subspecies of *Sistrurus* and the locations of samples used in our analyses. Subspecies abbreviations: Scc (*S. c. catenatus*); Sct (*S. c. tergeminus*); Sce (*S. c. edwardsii*); Smm (*S. m. miliarius*); Smb (*S. m. barbouri*); and Sms (*S. m. streckeri*). Black dots show approximate locations of sampled snakes. Map is modified from Mackessy (2005).

higher numbers of ventral scales and dorsal blotches than *edwardsii* (Gloyd 1955).

The other member of the genus, *S. miliarius*, contains three subspecies (*miliarius, barbouri*, and *streckeri*) found in the south-central and southeastern regions of the United States (Fig. 1). Similar to *catenatus*, these three subspecies are diagnosed on the basis of individuals that are morphologically distinct in scale, coloration, and size and have allopatric distributions (Gloyd 1935, 1940). The *barbouri* subspecies, which is found mainly in the southeastern United States (Fig. 1) is darker, has a more heavily spotted ventral surface, and has 5–10 more ventral scales than *streckeri* or *miliarius*. These subspecies, which are found in the south central United States and in North and South Carolina, northern Georgia, and Alabama, respectively (Fig. 1), differ in scale row counts and the size of ventral blotches (Gloyd 1935).

In addition to an understanding of the evolutionary history of this group of snakes, a phylogenetic analysis is useful because this group is being developed as a model system for understanding venom evolution in viperid snakes (Sanz et al. 2006; Gibbs and Rossiter 2008; Gibbs et al. 2009). Assessing the relationships between named taxa using molecular data would provide a phylogenetic framework for evaluating the evolutionary mechanisms responsible for observed differences in venom proteins and genes between taxa. Equally important there is significant interest in the phylogenetic distinctiveness of one subspecies (*S. c. catenatus*) which is of conservation concern across its range (Szymanski 1998).

MATERIALS AND METHODS

Samples

We analyzed samples from blood, muscle tissue, or shed skins from 24 individual *Sistrurus* and one sample from each of the two outgroups (*Agkistrodon contortrix* and *A. piscivorus*) (Fig. 1 and Table 1). Although our sampling was most complete for the *S. c. catenatus* subspecies ($n = 9$ individuals), we sampled at least two individuals within most subspecies (except for *S. m. miliarius* [$n = 1$]) and more than this in *S. c. tergeminus* ($n = 5$), *S. c. edwardsii* ($n = 4$), and *S. m. barbouri* ($n = 3$). We did not include any samples from putative hybrids between *S. c. catenatus* and *S. c. tergeminus* from populations in central Missouri (Evans and Gloyd 1948). Although the geographic scope of our sampling was limited for some subspecies, possibly leading to undetected phylogeographic structure, we expect our sampling to provide a useful initial evaluation of the phylogenetic relationships with a fairly large number of loci. Our sampling strategy is a reasonable compromise in that we included a fairly large number of loci

TABLE 1. Subspecies designations, locations (country, state/province, and county) and identification numbers of samples used in this study

Designation	Locality	Sample number	Tree ID number[a]
Sistrurus catenatus catenatus	USA: New York, Onondaga Co.	Sca 44	Scc-NY
S. c. catenatus	USA: Pennsylvania, Butler Co.	Sca 39	Scc-PA
S. c. catenatus	USA: Michigan, Oakland Co.	Sca 163	Scc-MI
S. c. catenatus	Canada: Ontario, North Bruce Peninsula District	Sca 348	Scc-ON1
S. c. catenatus	Canada: Ontario, Parry Sound District	Sca 583	Scc-ON2
S. c. catenatus	USA: Ohio, Champaign Co.	Sca 88	Scc-OH
S. c. catenatus	USA: Illinois, Cook Co.	Sca 156	Scc-IL1
S. c. catenatus	USA: Illinois, Clinton Co.	Sca 806	Scc-IL2
S. c. catenatus	USA: Wisconsin, Juneau Co.	Sca 151	Scc-WI
S. c. tergeminus	USA: Missouri, Holt Co.	Scter 49	Sct-MO1
S. c. tergeminus	USA: Missouri, Holt Co.	Scter 83	Sct-MO2
S. c. tergeminus	USA: Kansas, Barber Co.	Scter 02	Sct-KS1
S. c. tergeminus	USA: Kansas, Russell Co.	Scter 16	Sct-KS2
S. c. tergeminus	USA: Kansas., Barton Co.	Scter 115	Sct-KS3
S. c. edwardsii	USA: Colorado, Lincoln Co.	Sced 150	Sce-CO
S. c. edwardsii	USA: Arizona, Cochise Co.	Sced 32	Sce-AZ
S. c. edwardsii	USA: New Mexico, Otero Co.	Sced 127	Sce-NM1
S. c. edwardsii	USA: New Mexico, Valencia Co.	Sced 27	Sce-NM2
S. miliarius miliarius	USA: North Carolina, Pamlico Co.	Smm 10	Smm-NC
S. m. barbouri	USA: Florida, Volusia Co.	Smb 02	Smb-FL1
S. m. barbouri	USA: Florida, Volusia Co.	Smb 100	Smb-FL2
S. m. barbouri	USA: Florida, Volusia Co.	Smb 104	Smb-FL3
S. m. streckeri	USA: Oklahoma, Le Flore Co.	Sms 01	Sms-OK1
S. m. streckeri	USA: Oklahoma, Unknown	Sms 02	Sms-OK2
Agkistrodon contortrix	USA: Ohio, Lawrence Co.	Agc 01	Agc
A. piscivorus	USA: Florida, Volusia Co.	Agp 01	Agp

Note: [a]Tree ID number refers to taxon names in figures showing phylogenetic trees, with 1 and 2 added to the end to indicate the two (phased) sequences within each individual.

while still sampling multiple individuals within most subspecies.

Our choice of the two *Agkistrodon* species as outgroups is based on the repeated finding that this genus is sister to both genera of rattlesnakes (e.g., Parkinson et al. 2002). Preliminary analyses using a number of *Crotalus* species showed that they were not sufficiently distinct to allow the consistent polarization of variable characters within *Sistrurus,* which led us to consider *Agkistrodon*. Preliminary comparisons with these two outgroups indicated that some of the species-tree estimation methods performed better when the more distantly related outgroup was used.

Genetic Analyses

We generated sequence data from each sample for the 18 nuclear DNA loci described by Gibbs and Diaz (2010) and for a single-gene fragment consisting of partial sequence from two mitochondrial genes (*ATP* 6 and 8; Table 2). As described in Gibbs and Diaz (2010), 10 of these loci (locus *A*, 1, 4, 11, 25, 31, 41, 51, 61, and 63) were isolated from a *S. c. catenatus* genomic library, whereas the remaining loci (*TBP*, *CBA*, *OD*, *ETS*, *EF*, *GAPD*, *LAM*, and *FGB*) were generated from introns amplified using conserved primers. For the nDNA loci we used the primers and conditions described by Gibbs and Diaz (2010). For the mtDNA ATP gene fragment, we used the primers and conditions described by Douglas et al. (2002). Briefly, DNA was extracted from each sample using standard phenol–chloroform method, quantified and run on a 1% agarose gel to check for DNA quality. DNA from each sample was then amplified using primers for each locus in combination with a high fidelity polymerase (Platinum Taq—Invitrogen) to minimize polymerase errors and the products sequenced on an ABI 3100 DNA Analyzer. Sequences were then assembled in Sequencher and aligned by eye in Bioedit 7.0.5 (http://www.mbio.ncsu.edu/BioEdit/bioedit.html).

In a small number of cases, base calls at individual sites were ambiguous with respect to one of two possible bases. For further analysis, the chromatograms were imported into CodonCode Aligner (version 1.6.2) and bases were called using the PHRED subroutine. Once quality scores were obtained by base calling, the sequences were assembled into contigs, and the analysis to find heterozygous point mutations was run with all contigs selected. The low-sensitivity option was used when finding mutations to reduce false positives. At sites where heterozygosity was observed, all individuals were visually inspected for confirmation.

When two or more polymorphic sites were present in a sequence, we determined the gametic phase of alleles using the program PHASE 2.1 (Stephens et al. 2001) as implemented in DnaSP ver. 4.9 (Rozas et al. 2003). PHASE uses a Bayesian approach to infer haplotypes from diploid genotypic data accounting for both recombination and linkage disequilibrium. Where necessary, we conducted analyses on three sets of sequences at each locus: *S. c. catenatus*, *S. c. tergeminus*, and *S. c. edwardsii* combined; *S. m. miliarius*, *S. m. barbouri*, and *S. m. streckeri* combined; and the two *Agkistrodon* species

TABLE 2. Aligned length (in base pairs), numbers of phylogenetically informative sites (PI), estimated models of sequence evolution using DT-ModSel (Minin et al. 2003), and average sequence divergences between the ingroup (*Sistrurus*) and outgroup (*Agkistrodon*) for the 18 nuclear DNA loci and 1 mtDNA gene fragment used in this study (sequences have been deposited in GenBank under accession numbers: FJ659860–FJ660411)

Locus[a]	Aligned length	PI sites[b] (ingroup only)	Substitution model	Average divergence
A	296	31	K80 + *I*	0.02527
1	220	12	K80	0.01297
4	267	5	K80 +	0.00726
11	420	14	K80 + *I*	0.01199
25	262	14	TVMef + *I*	0.06160
31	256	8	F81	0.01071
41	274	7	HKY	0.00625
51	260	10	K80	0.02073
61	194	3	HKY	0.00819
63	471	8	HKY + *I*	0.01019
TBP	796	26	HKY + *I*	0.01444
CBA	525	9	HKY + *I*	0.08465
OD	522	16	K81uf + *I*	0.01337
ETS	849	20	HKY + *I*	0.01197
EF	447	5	HKY	0.15541
GAPD	260	6	HKY + *I*	0.01181
LAM[c]	469	20	K80 + *I*	0.16081
FGB	798	15	HKY	0.08364
ATP	665	155	HKY + *I*	0.12264

Notes: [a]Loci *A* to 63 represent anonymous cloned loci, *TBP* to *FGB* are intron-based loci, and *ATP* is an mtDNA gene fragment (see Gibbs and Diaz 2010 for more details).
[b]Excludes gap-based characters.
[c]The *LAM* sequence is a smaller piece of a 684 bp sequence in which a recombination breakpoint occurs at site 214 (see Results). Here, we analyze the larger of the two fragments defined by this breakpoint.

combined. Each data set was analyzed using 500 steps for burn-in, 1 thinning interval and 1000 main iterations and each analysis was repeated at least twice. We inferred the gametic phase of alleles for polymorphic sites with probabilities of $\geqslant$0.7. Based on the results of Harrigan et al. (2008), we assume that sites with phase probabilities this large or greater are accurately inferred when compared with empirical results based on cloning. All polymorphic sites with a probability of $<$0.7 were coded in both alleles with the appropriate IUPAC ambiguity code. For each locus, we calculated the number of parsimony informative sites and the average sequence divergence of each of the ingroup taxa to the two outgroup sequences. DT-ModSel (Minin et al. 2003) was used to select the best-fit evolutionary model for each gene. These results are summarized in Table 2.

Finally, we assessed evidence for recombination at each locus using the on-line version of the program *G*enetic *A*lgorithm for *R*ecombination *D*etection (GARD; Kosakovsky Pond et al. 2006; see www.datamonkey.org/GARD). This phylogenetically based method uses a genetic algorithm to search multiple sequence alignments for putative recombination breakpoints and then assesses statistical support for their location using an Akaike information criterion (AIC). We conducted this analysis for all *Sistrurus* sequences for the 18 nuclear DNA loci listed in Table 2.

Phylogenetic Analysis

Single-gene phylogenies were estimated in a Bayesian framework using MrBayes (Ronquist and Huelsenbeck 2003). For each gene, 50 million generations with separate HKY + *I* models specified for each gene were used. This model was selected based on the DT-ModSel results for the individual loci, which indicated that this model was generally appropriate (Table 2). The first 10 million iterations were discarded as burn-in, and every 10,000th iteration was sampled from the remaining 40 million, so that a total of 8000 trees were used to estimate the posterior distribution (pooled over 2 independent runs). Convergence for each gene was assessed using the average standard deviation of split frequencies and potential scale reduction factor (PSRF) values. No indication of lack of convergence was found for any of the individual genes (all average standard deviations of split frequencies were $<$0.007, and all PSRFs were very near 1.0).

We used the programs STEM version 1.1 (Kubatko et al. 2009) and *BEAST to obtain species-level phylogenetic estimates. In both analyses, operational taxonomic units used were the subspecies within each of two species identified by Gloyd (1935, 1940, 1955) as well as the two *Agkistrodon* species for the outgroup. Input gene trees for the STEM analysis were taken as the maximum a posteriori (MAP) tree from the single-gene Bayesian analyses in MrBayes described above. ML branch lengths on the MAP trees were obtained using PAUP* with the assumption of a molecular clock. Each gene was additionally given a separate rate multiplier using the following procedure (as originally suggested by Yang 2002). First, the average pairwise distance to the outgroup was determined for each gene. These average pairwise distances were then each divided by their overall mean, and this rate was assigned to each gene.

In addition, the rate for the mtDNA gene was divided by 2 to reflect its haploid status. Finally, the parameter θ was set to 0.0015, which is a reasonable value for these taxa based on empirical estimates of θ using intron data from *S. c. catenatus* analyzed using LAMARC version 2.1 (Kuhner 2006). Varying θ between 0.001 and 0.01 did not change the species-tree estimate.

For the *BEAST analysis, the molecular sequence data were used to obtain a Bayesian estimate of the species tree. Matching the analysis done on the concatenated data set using BEAST (see below), an unlinked HKY + *I* substitution model was assumed for all 19 genes with all related parameters (proportion of invariant sites, transition/transversion ratio) being estimated. This model permits gene-specific mutation rates, which allows for the inclusion of the mtDNA locus; throughout, the mutation rate of *TBP* was set equal to 1.0 so that branch length estimates are scaled in terms of expected substitutions per site of the *TBP* gene. The *TBP* gene was chosen for its relatively high-resolution gene tree estimate in the single-gene analysis. We assumed a hierarchical prior for the effective population sizes (the default) as described by Heled and Drummond (2010). We used *BEAST to generate posterior samples of the 19 individual gene trees and the overall species tree through a MCMC procedure. To assess convergence of the algorithm, we repeated each *BEAST analysis on 4 separate chains for 350 million iterations. In each run, the first 100 million trees were discarded as burn-in, and every 10,000th tree was kept afterward. Each run took ∼10 days on a Unix cluster or desktop computer using a single processor per run. Trace plots and histograms were visually inspected to assess convergence of the MCMC; *BEAST appeared to sample all parameters well except for the effective population sizes (see Discussion).

In terms of approach, *BEAST generates posterior samples from a similar model to that implemented in the widely used species-tree estimation program BEST (Liu and Pearl 2007), whereas the technique used to make inference is different. *BEAST takes a single-stage approach to inference and attempts to sample the 19 gene trees and the species tree simultaneously. In contrast, BEST employs a two-stage algorithm. First, BEST finds the marginal posterior estimates of the 19 gene trees and then uses an importance sampling correction to transform these marginal estimates into joint posterior estimates. Both techniques have the same analytical goal of estimating a species tree and associated parameters; nevertheless, in practice, the current, single-processor, implementation of *BEAST is substantially more computationally efficient than the BEST program.

We also estimated the phylogeny using a concatenated alignment. For estimation in a Bayesian framework, BEAST (Drummond and Rambaut 2007) was run for 100 million generations with separate HKY + models specified for each gene. In each run, the first 10 million trees were discarded as burn-in, and every 10,000th tree was sampled thereafter for a total of 36,000 trees pooled over 4 independent runs. Convergence was assessed by comparing the estimated posterior distributions across the 4 runs through visual inspection of trace plots and summary statistics; poor convergence did not appear to be an issue.

An analysis under the parsimony criterion was also performed for the concatenated data set to gain information concerning the stability of the estimates. Because many of the loci have relatively low rates of evolution, we did not expect substantial disagreement between the parsimony and Bayesian analyses. PAUP* (Swofford 2003) was used to carry out a bootstrap analysis with 100 replicates. For each replicate, heuristic searches were performed using 20 random addition sequence replicates with TBR branch swapping.

Monophyly-Based Tests of Taxonomic Distinctiveness

We also directly assessed the phylogenetic distinctiveness of the subspecies used in the species-tree analysis following an approach developed by Rosenberg (2007). Rosenberg (2007) noted that an observation of monophyly alone is not necessarily evidence for taxonomic distinctiveness because the probability of monophyly for a given collection of lineages may not be low to begin with. To claim taxonomic distinctiveness, he argues that an observation of monophyly should be coupled with a computation of the probability of observing monophyly for the particular taxon sample in the study by chance alone (i.e., when in fact the two groups under consideration are not taxonomically distinct). He then provides methodology to evaluate the probability of taxonomic distinctiveness given an observation of monophyly in the data. As he points out, this should be distinguished from methods (e.g., Huelsenbeck et al. 1996) that are designed to evaluate the strength of the evidence for monophyly in a data set. Rosenberg (2007) provides two versions of the calculations, one that uses monophyly of one group within a larger group and one that uses reciprocal monophyly of the two groups.

In a multilocus setting, these calculations can be used to formally test a hypothesis of phylogenetic distinctiveness (Rosenberg 2007) given observed monophyly across loci. First, we specify the null hypothesis that the lineages are drawn from a single taxonomic group. Under this null model and assuming independence across loci, the probability of observing monophyly (or reciprocal monophyly) in k or more loci in a sample of N loci can be calculated from the Binomial distribution with the probability of "success" given by the equations in Rosenberg (2007) for the probability of monophyly or reciprocal monophyly. When this probability is low, we have evidence against the null hypothesis of a single taxonomic group in favor of taxonomic distinctiveness.

To implement these tests for our data, we considered taxonomic distinctiveness of the subspecies within both the *S. catenatus* clade and the *S. miliarius* clade. Within *S. catenatus*, we considered two separate hypotheses. The first is that *S. c. catenatus* is distinct from a clade containing both *S. c. edwardsii* and *S. c. tergeminus*. The

second is that *S. c. edwardsii* is distinct from *S. c. tergeminus* in the clade that contains them both. Within the *S. miliarius* clade, we examined evidence for distinctiveness of *S. m. streckeri* from a clade containing *S. m. miliarius* and *S. m. barbouri*. Finally, within the clade containing *S. m. miliarius* and *S. m. barbouri*, we examined evidence for distinctiveness of *S. m. barbouri* only as there is only a single *S. m. milarius* sample hence not enough power to conduct the test for this subspecies. All tests were carried out by examining the individually estimated gene trees to determine which either supported or contradicted the particular monophyly relationships of interest. The consensus trees from the single-gene Bayesian analysis described above were used as these gene tree estimates (see Fig. 2). We used consensus trees rather than MAP trees in this case so that only relatively strongly supported observations of monophyly would be used in carrying out these tests. *P* values for the tests were computed in two ways, using only trees that either supported or contradicted the monophyly relationships under consideration, and using all trees, with those that were unresolved for the clades of interest counted as trees that contradicted these relationships.

Note that the species criterion that is invoked here for identifying taxonomically distinct entities is the genealogical species concept (Baum and Shaw 1995) that delimits species as groups of individuals exhibiting concordant patterns of monophyly across unlinked genes. Such approaches often specify a priori that a given percentage of loci must exhibit monophyly for a particular taxonomic group in order for that group to be considered a species (e.g., Hudson and Coyne 2002). The percentage selected is arbitrary, though often at least a majority of loci are required (Hudson and Coyne 2002). The method proposed here takes a different approach in that it assesses the chance of observing the extent of monophyly that occurs in the sample under the null hypothesis of a single taxonomic group. When this chance is low, the sample provides evidence in favor of distinctiveness of the group under consideration.

Genealogical Sorting Index

The *gsi* and *egsi* (Cummings et al. 2008) were used to quantify the degree of exclusive ancestry for all subspecies identified here using the functions implemented in the Genealogical Sorting package in R (available at http://www.genealogicalsorting.org/resources/). The *gsi* and *egsi* provide a measure, for individual genes and for multilocus data, respectively, of the relative degree of exclusive ancestry on a scale from 0 to 1, where 1 indicates complete monophyly. The gene trees used as input for computing the *gsi* and *egsi* were the MAP trees estimated in the single-gene Bayesian analyses described above. We chose to use the MAP trees (rather than the consensus trees used in the monophyly tests above) because, as pointed out by Cummings et al. (2008), the use of consensus trees might diminish the magnitude and significance of the *gsi* measures in comparison with the ML or MAP tree. However, we felt that consensus trees were appropriate in the monophyly tests, as we wanted to use only monophyly relationships that are at least moderately supported by the data in that test. For the *gsi* and *egsi*, a *P* value to evaluate the null hypothesis that the degree of relative exclusive ancestry would be observed by chance alone (i.e., when in fact the two groups under consideration are not taxonomically distinct) was estimated using 1 million permutations.

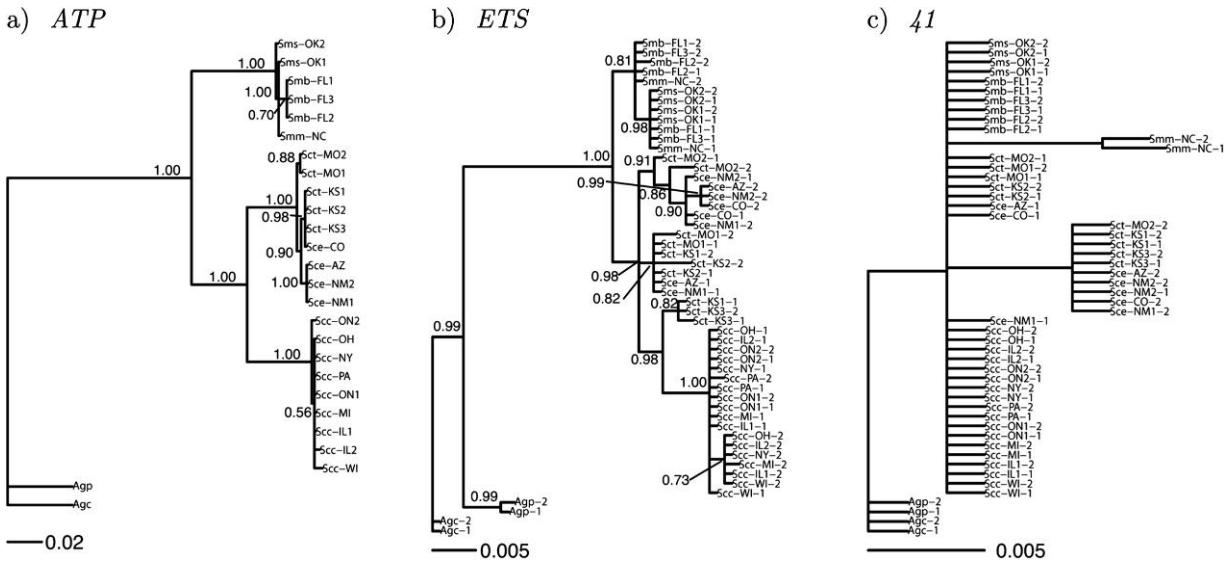

FIGURE 2. Examples of individual gene trees inferred using Bayesian analysis that showed a) significant—*ATP*; b) moderate—*ETS*; or c) limited—*Clone 41* phylogenetic signal. Branch lengths are in units of number of substitutions per site. See the Supplemental Information for single-gene phylogenetic estimates of the additional loci in this study.

Estimation of Divergence Times From Gene and Species Trees

We were interested in comparing divergence time estimates for nodes in the *Sistrurus* phylogeny based on the concatenated sequence data as compared with species-level divergences based on the species-tree topology generated using *BEAST. Theoretically, the species-tree estimates should be more recent because gene divergences must predate species divergences in the absence of gene flow following speciation (Edwards and Beerli 2000) but there are few empirical estimates of the degree to which species and gene tree-based estimates differ (Jennings and Edwards 2005). A widely used approach to generating divergence time estimates using concatenated data is to apply the program BEAST (Drummond and Rambaut 2007) in combination with one or more fossil calibration dates and estimate divergence times based on sequence divergence (e.g., Belfiore et al. 2008). This approach can also be used to generate comparable species divergence times in *BEAST by calibrating a particular speciation time (on the species tree) to a fossil date.

To generate divergence times using both programs, we used a single fossil calibration date of 9 Ma based on the most recently described *Sistrurus* fossil (Parmley and Holman 2007). For both BEAST and *BEAST, we set a lognormal prior for the minimum age of *Sistrurus* clade with a minimum age of 9 Ma, a mean of 10 Ma, and a standard deviation of 1.73 Ma. We ran both fossil calibrated analyses using the same number of iterations, burn-in samples and subsampling frequency as the two analyses done without a fossil calibration. We also replicated these runs 4 times and visually compared convergence through trace plots and histograms. Convergence did not appear to be an issue for either analysis, except for estimates of the ancestral population sizes.

RESULTS

Analysis of Recombination

We found little evidence of recombination in our nuclear DNA loci, likely because the loci were relatively short in length. Based on the GARD results, 17 of 18 loci showed no statistically significant support for internal recombination breakpoints in the aligned sequences. Only the *LAM* locus showed significant support ($\Delta AIC - c = 120.1$ between no breakpoint and single breakpoint models) for a breakpoint at position 214 within the 684 bp sequence. With the goal of analyzing a single nonrecombining block of sequence at this locus, we only analyzed sequence from positions 215 to 684 at this locus.

Phylogenetic Analysis

Based on our multilocus data set, one of our primary goals was to estimate a species phylogeny for these recently diverged taxa. Examination of the phylogenetic estimates for individual genes obtained through a Bayesian analysis reveals a lack of resolution in and substantial incongruence between the single-gene phylogenies (see Fig. 2 for examples of individual loci that vary in phylogenetic informativeness; all 19 individual gene tree estimates are provided in the Supplementary material [available from http://www.sysbio.oxfordjournals.org/], and the complete data matrix and gene trees are available in TreeBASE [http://purl.org/phylo/treebase/phylows/study/TB2:S11174]). Thus, we expect inference of a species phylogeny to be challenging, as is the case with most recently diverged groups. Figure 3 shows the MAP tree from the *BEAST analysis, which we use as the estimate of the species tree. The species-tree estimate obtained using STEM is similar (Fig. 4a), except that the relationships within the *S. miliarius* clade are altered, so that *S. m. barbouri* and *S. m. streckeri* form a clade.

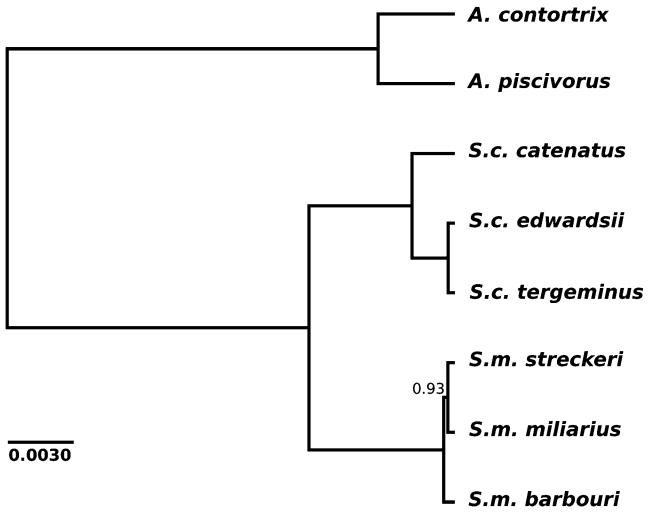

FIGURE 3. Most probable species tree from *BEAST. Branch lengths are scaled according to expected number of substitutions at the *TBP* locus. The 0.93 represents the posterior probability of the *S. m. miliarius* and *S. m. streckeri* clade. All other clades have posterior probabilities in excess of 0.99.

We also performed analyses of the concatenated data sets in both a Bayesian framework (using BEAST) and under the parsimony criterion. These two analyses provide similar inferences at the deeper nodes but differ in the groupings within the *S. miliarius* clade as well as in the placement of some of the tip taxa (Fig. 5;

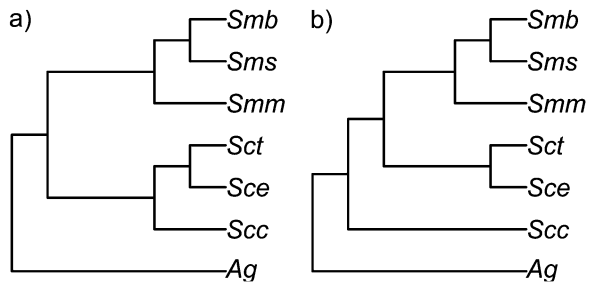

FIGURE 4. The species-tree estimates obtained by a) STEM and b) PhyloNet.

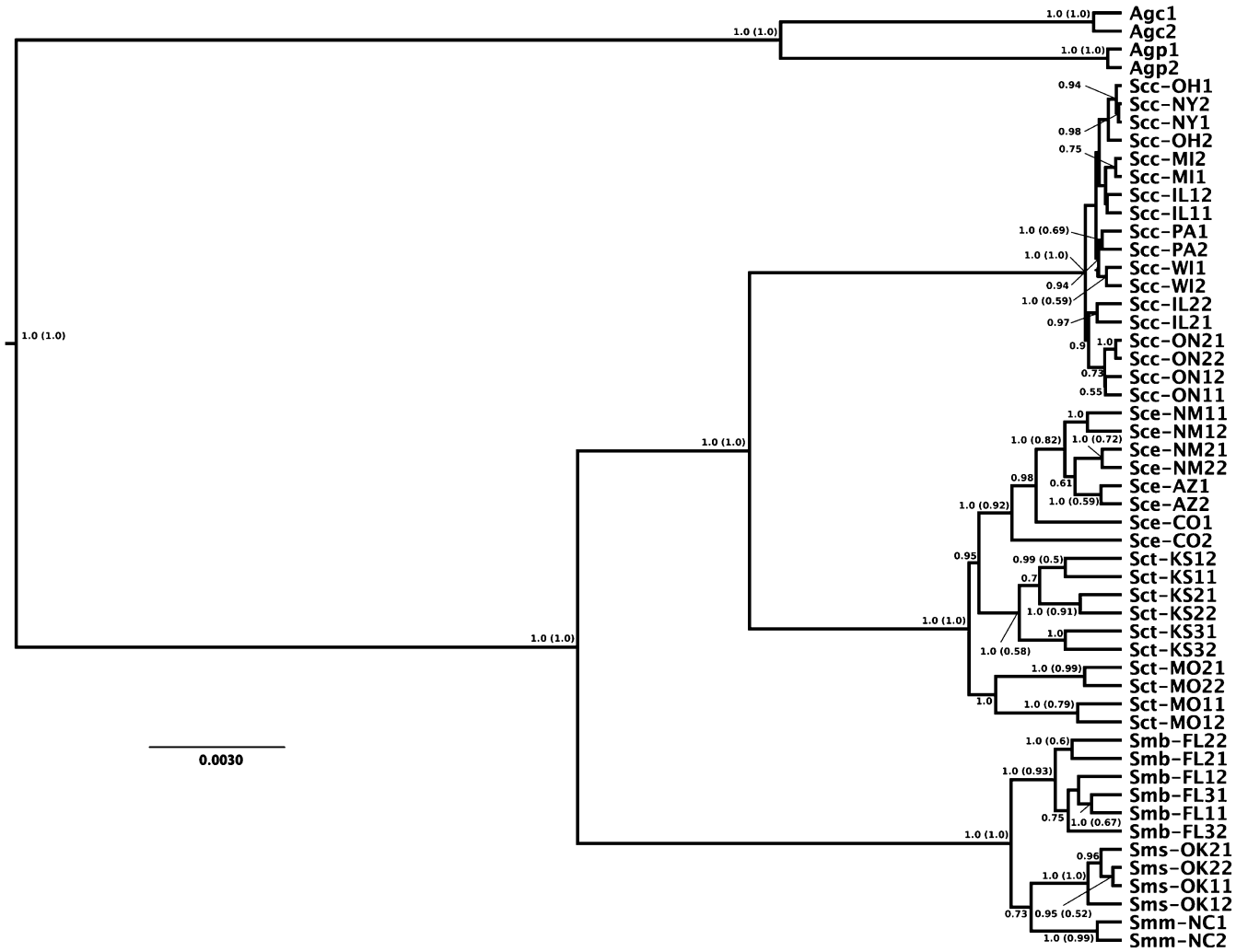

FIGURE 5. Maximum clade credibility tree from concatenated analysis in BEAST. This tree represents the posterior sample with the maximum sum of clade posterior probabilities at the internal nodes. This particular tree has branch lengths equal to expected substitutions per site of the *TBP* locus. Posterior probabilities of each clade above 0.5 are shown; BPs above 0.5 are shown in parentheses next to the posterior probabilities. Any missing probability indicates a value below 0.5.

posterior probabilities from BEAST are given above the nodes, with bootstrap proportions [BPs] from the parsimony analysis in parentheses). The difference in branch lengths largely results from the concatenation assumption and the information present in the *ATP* locus. The concatenated analyses show mixed support for the patterns observed in the species-tree analyses. Within *S. catenatus*, both the parsimony and Bayesian analyses continue to show strong support (PP 1.0 and BP $\geqslant$ 0.92) for separate *S. c. catenatus* and *S. c. edwardsii* clades, but *S. c. tergeminus* sequences are paraphyletic with sequences from the Kansas (KS) samples clustering with *S c. edwardsii* sequences, whereas sequences from Missouri (MO) form a separate clade. Within *S. miliarius*, the subspecific relationships are poorly resolved. The Bayesian analysis of the concatenated data using BEAST suggests that *S. m. streckeri* and *S. m. miliarius* form a clade, whereas the parsimony analysis of the concatenated data suggests that *S. m. miliarius* and *S. m. barbouri* are sister groups. The grouping from the Bayesian concatenated analysis agrees with that found by *BEAST. Finally, STEM provides strongest support for the third of these groupings, that is, *S. m. barbouri* and *S. m. streckeri* as sister groups; however, the alternate resolutions within this clade found by the other methods also have high likelihood under this model in STEM.

The tree generated from the concatenated data set also allows us to look for evidence of phylogeographic structure within the best-sampled subspecies (*S. c. catenatus*). Inspection of the BEAST tree shows no evidence for strong phylogeographic structure in this subspecies: Although some clades have high support, they consist of clusters of individuals from noncontiguous locations (e.g., clades consisting of NY and OH samples and PA and WI samples, respectively). We tentatively conclude that *S. c. catenatus* is not further subdivided into geographically distinct lineages. However, this conclusion

needs to be confirmed with more extensive samples than analyzed here.

Monophyly Tests of Taxonomic Distinctiveness

Results of the tests for taxonomic distinctiveness are shown in Table 3. There is strong evidence for the distinctiveness of *S. c. catenatus* from the *S. c. edwardsii* to *S. c. tergeminus clade*, regardless of which method (monophyly vs. reciprocal monophyly) and which collection of gene trees (reduced vs. full) are used for the computations. In general, using the full set of gene trees with unresolved genes counted as trees that contradict monophyly relationships will be a conservative method because such trees actually do not provide evidence either way. Even under this stringent test, the probability of observing the number of trees that maintain particular levels of monophyly would be very low under the null model of a single taxonomic entity (<0.0001 in all cases).

Within the clade containing *S. c. edwardsii* and *S. c. tergeminus*, there is limited support for distinctiveness of the two groups, although *S. c. edwardsii* is monophyletic for the *TBP* gene. In many of the gene trees, the subspecies are nearly monophyletic with the exception of only one or two of the lineages, so perhaps these subspecies are in the early stages of becoming distinct lineages. Within *S. miliarius*, there is some support for the distinctiveness of *S. m. streckeri* and the clade containing *S. m. barbouri* and *S. m. miliarius*, with the *P* values in all comparisons (monophyly and reciprocal monophyly as well as full vs. reduced data set) below 0.01. When only monophyly is used as the criterion, the *P* values are much smaller for both the full and reduced data sets. In addition, there is moderate support for distinctiveness of *S. m. barbouri* from *S. m. miliarius* in the clade that contains them both.

Genealogical Sorting

The *gsi* and *egsi* values indicate a high degree of exclusive ancestry within species and subspecies overall (Table 4). When examining all 19 loci combined, the *egsi* values range from a little over 0.4 (for *S. c. tergeminus* and *S. c. edwardsii*) to about 0.74 (for *S. c. catenatus*). Although some of these values are not necessarily very close to 1, which means that the groups do not display a high degree of monophyly, all the *P* values indicate that the degree of exclusivity observed would be very unusual if in fact the groups were not distinct. These results are congruent with the observation made earlier for the individual genes: Although complete monophyly was somewhat rare, most genes display clear indications of nonrandom clustering at the subspecies level.

We also compared *gsi* values across genes. Some genes (e.g., the mitochondrial gene, *ATP*, as well as *CBA*, *OD*, and *TBP*) show very strong indications of shared ancestry across all subspecies. In other cases, the support is mixed, with a strong signal in some groups but not others. Across all genes, the *S. c. catenatus* subspecies has

TABLE 3. Tests of taxonomic distinctiveness based on patterns of monophyly in individual gene trees following Rosenberg (2007)

Distinctiveness of	Method	*a*	*b*	Trees supporting monophyly	Trees contradicting monophyly	P_A or P_{AB}	*P* value reduced	*P* value full
Scc vs. (Sce-Sct)	Monophyly	18	18	a, b, d, g, h, n, p	c, e, j, k, o	2.32×10^{-11}	2.86×10^{-72}	1.82×10^{-70}
(Sce-Sct) vs. Scc	Monophyly	18	18	a, d, p, s	c, e, g, j, k, n, o	2.32×10^{-11}	9.56×10^{-41}	1.12×10^{-39}
Scc vs. (Sce-Sct)	Reciprocal monophyly	18	18	a, d, p	c, e, g, j, k, n, o	6.30×10^{-12}	3.00×10^{-32}	2.42×10^{-31}
Sce vs. Sct	Monophyly	10	8	g	a, b, c, d, e, f, j, m, n, p, r	7.48×10^{-6}	8.22×10^{-5}	0.00014
Sct vs. Sce	Monophyly	8	10		a, b, c, d, e, f, g, j, m, n, p, r	1.14×10^{-5}	1.0	1.0
Sce vs. Sct	Reciprocal monophyly	10	8		a, b, c, d, e, f, g, j, n, m, p, r	2.69×10^{-6}	1.0	1.0
Sms vs. (Smm-Smb)	Monophyly	4	8	c, d, f, g, k	j, l, n	0.00242	4.66×10^{-12}	9.46×10^{-10}
(Smm-Smb) vs. Sms	Monophyly	8	4	c, p	e, g, j, l, n	0.00067	9.50×10^{-6}	7.70×10^{-5}
Sms vs. (Smm-Smb)	Reciprocal monophyly	4	8	c	e, g, j, l, n	0.00037	0.00220	0.00696
Smb vs. Smm	Monophyly	6	2	a, j	c, g, e, i, l, n	0.01361	0.00491	0.02715

Notes: The column labeled "Distinctiveness of" gives the taxa under consideration and the column labeled "Method" gives the method on which the calculations were based. When "Monophyly" was used to carry out the test, calculations were performed to examine monophyly of the first group listed in the first column within the larger clade defined by all groups listed in the first column. When "Reciprocal monophyly" was used for the calculations, then the order in which groups are listed does not affect the test. The columns labeled "*a*" and "*b*" give the number of lineages used to carry out the test, using the notation in Rosenberg (2007). The columns labeled "Trees supporting monophyly" and "Trees contradicting monophyly" specify the gene trees (see the Supplementary material) that support and contradict, respectively, the particular type of monophyly examined. The column labeled "P_A or P_{AB}" gives the probability of the particular monophyly relationships using the equations in Rosenberg (2007) (see Materials and Methods section). The last two columns give *P* values for specific tests. The column labeled "*P* value reduced" gives the *P* value using only gene trees which either support or contradict the monophyly relationships under consideration (trees that are unresolved with respect to these relationships are excluded). The column labeled "*P* value full" includes all gene trees, with those unresolved for the relationships of interest counted as trees that contradict these relationships.

TABLE 4. The *gsi* is given for each subspecies separately for each of the 19 loci and for the 19 loci combined (*P* values based on 1 million permutations are given in parentheses below each value)

Locus	*Sistrurus catenatus catenatus*	*S. c. edwardsii*	*S. c. tergeminus*	*S. miliarius miliarius*	*S. m. barbouri*	*S. m. sterckeri*
A	0.9167	0.5170	0.6264	1.0000	0.5072	0.3929
	(<0.000001)	(<0.000001)	(<0.000001)	(0.0099)	(0.0001)	(0.0034)
1	0.4808	0.2917	0.2825	0.3200	0.4457	0.3929
	(<0.00001)	(0.0045)	(0.0048)	(0.0581)	(0.0003)	(0.0030)
4	0.4444	0.2273	0.1272	0.2350	0.5072	0.3923
	(<0.00001)	(0.0300)	(0.4255)	(0.0982)	(<0.00001)	(0.0031)
11	0.7750	0.3182	0.3609	0.3200	0.6832	1.0000
	(<0.000001)	(0.0029)	(0.0006)	(0.0570)	(<0.0001)	(<0.0001)
25	0.7750	0.5785	0.2825	0.2350	1.0000	0.4688
	(<0.000001)	(<0.000001)	(0.0023)	(0.1104)	(<0.000001)	(0.0015)
31	0.7143	0.3818	0.3061	0.4900	0.5842	1.0000
	(<0.000001)	(0.0004)	(0.0024)	(0.0286)	(<0.0001)	(<0.00001)
41	0.5200	0.4650	0.2609	1.0000	0.2174	0.7344
	(<0.00001)	(<0.00001)	(0.0083)	(0.0060)	(0.0319)	(0.0001)
51	1.0000	0.5170	0.5142	0.3200	0.5842	1.0000
	(<0.000001)	(<0.00001)	(<0.00001)	(0.0627)	(<0.00001)	(<0.00001)
61	0.2286	0.2097	0.1905	1.0000	0.3177	0.2917
	(0.0372)	(0.0620)	(0.1116)	(0.1115)	(0.0041)	(0.0120)
63	0.8421	0.3182	0.5143	1.0000	0.5072	0.2917
	(<0.000001)	(0.0020)	(<0.00001)	(0.0106)	(0.0001)	(0.0140)
TBP	1.0000	1.0000	0.8786	1.0000	0.5842	1.0000
	(<0.000001)	(<0.000001)	(<0.000001)	(0.0114)	(0.00001)	(0.00001)
CBA	0.9166	0.6523	0.4688	0.4900	0.3953	0.3359
	(<0.000001)	(<0.000001)	(<0.0001)	(0.0288)	(0.0007)	(0.0071)
OD	1.0000	0.5170	0.5663	0.4900	1.0000	0.7344
	(<0.000001)	(<0.00001)	(<0.00001)	(0.0311)	(<0.000001)	(0.0001)
ETS	1.0000	0.4650	0.6264	0.1075	0.5842	0.5750
	(0.0002)	(<0.000001)	(<0.00001)	(<0.000001)	(0.3313)	(<0.00001)
EF	0.2286	0.1655	0.1626	0.1075	0.6832	0.7343
	(0.0120)	(0.0675)	(0.0747)	(0.2978)	(<0.00001)	(0.3374)
GAPD	0.4444	0.1655	0.1905	0.0438	0.0825	0.4688
	(<0.00001)	(0.1356)	(0.0715)	(0.7864)	(0.7088)	(0.0013)
LAM	0.9167	0.5170	0.2609	1.0000	0.3177	0.7343
	(<0.000001)	(<0.00001)	(0.0189)	(0.0113)	(0.0043)	(0.0001)
FGB	0.7750	0.3181	0.4688	1.0000	0.4457	1.0000
	(<0.000001)	(0.0030)	(<0.00001)	(0.0099)	(0.0003)	(<0.00001)
ATP	1.0000	0.5785	0.6964	0.2350	1.0000	0.7344
	(<0.000001)	(<0.0001)	(<0.000001)	(0.0903)	(<0.000001)	(0.0001)
All combined	0.7357	0.4318	0.4097	0.5470	0.5498	0.6464
	(<0.000001)	(<0.000001)	(<0.000001)	(0.0015)	(<0.000001)	(<0.0001)

the strongest and most consistent support, with all 19 of the *P* values less than the typically used 0.05 cutoff (17 of the 19 *P* values are <0.002).

Divergence Time Estimates

Table 5 shows that the estimate of the date of the divergence between *S. catenatus* and *S. miliarius* is similar for both gene and species tree–based estimates, likely reflecting the fact that the same fossil calibration was used in both analyses. In contrast, all other species tree–based divergence times are >50% less than those inferred from the concatenated gene tree-based estimates. For example, the gene tree-based point estimate of the date for the splitting off of *S. c. catenatus* from the other two subspecies (6.06 Ma) is more than twice as large as the point estimate from the species-tree analysis (2.93 Ma). Overall, these differences imply very different time courses of diversification in these snakes: Gene tree estimates suggest that the currently described subspecies originated over the Pliocene to early Pleistocene, whereas species-tree estimates suggest a much more recent origin to the subspecies starting in the late Pliocene with substantial diversification occurring in the mid to late Pleistocene.

DISCUSSION

Multilocus Phylogenetic Analysis of Sistrurus

Despite substantial variability in the single-gene phylogenetic estimates, the methods we employed all strongly supported the relationships observed at the species level. However, at the subspecies level, different methods provide support for alternative relationships. In all analyses, *S. c. catenatus* and *S. c. edwardsii* form strongly supported monophyletic clades. However, in the trees constructed from the concatenated data, *S. c. tergeminus* is paraphyletic, with two of the samples from northwestern Missouri inferred to be sister to the entire *S. c. edwardsii*–*S. c. tergeminus* clade. Within *S. miliarius*, all three possible topologies for the three subspecies (*miliarius, barbouri*, and *streckeri*) are supported by different analyses: In the parsimony analysis, *S. m. miliarius* and *S. m. barbouri* are sister groups, whereas in the BEAST and *BEAST analyses, *S. m. miliarius* is sister to *S. m. sterckeri*. The species-tree estimate obtained

TABLE 5. Comparison of divergence estimates for specific nodes based on branch lengths calibrated using a fossil date from the concatenated gene tree analysis in BEAST and the species tree analysis in *BEAST, based on data from all 19 genes scaled to expected substitutions per site of the *TBP* gene (see Materials and Methods section)

Dated node	Divergence estimates from concatenated gene tree (Ma)[a]	Divergence estimates from species tree (Ma)[a]	Percent difference[b] (%)
(Scc (Sce,Sct)) vs. (Sms(Smb, Smm))	9.45 (9.14, 10.24)	10.04 (9.25, 12.97)	+6
Scc vs. (Sce, Sct)	6.06 (5.22, 7.02)	2.92 (1.58,4.90)	−52
Sce vs. Sct	2.41 (2.01, 2.88)	0.47 (0.24, 0.86)	−79
Smb vs. (Smb, Sms)	1.98 (1.60, 2.47)	0.77 (0.44,1.31)	−62
Sms vs. Smm	1.60 (1.23, 2.06)	0.49 (0.25, 0.92)	−69

Notes: [a]Dates (50% median value from the posterior distribution with the 2.5% and 97.5% values shown below) are presented as branch lengths translated into millions of years before present (Ma). Values are conditional on monophyly of all six of the clades in the row headings. For the species-tree analysis, the probability of all clades being monophyletic is ∼91%; for the gene tree analysis, this probability is ∼74%.
[b]Percent difference gives the change in speciation times from the concatenation analysis to the species-tree analysis.

using STEM shows the third possible relationship, with *S. m. barbouri* sister to *S. m. streckeri*. Nevertheless, all analyses give an indication of the uncertainty in this placement: In STEM, this can seen in that the likelihood scores for trees with alternative arrangements within this clade are not much lower than that of the ML tree; in the concatenated analysis using parsimony, several of the BP values are somewhat small (<0.90) on nodes in this part of the tree; and in BEAST and *BEAST, the branch length delineating the three subspecies in *S. miliarius* is extremely short, and the posterior probabilities are not highly significant (<95%).

We found estimation of the species tree with *BEAST to be relatively straightforward and that most parameters showed strong evidence of convergence. However, we had issues with a lack of convergence of the estimates of effective population sizes, particularly for tip taxa. However, our estimates of all other parameters were very robust to this lack of convergence in this single class of parameters and trace plots of all other parameters were stable across runs. This lack of convergence implies that we can only generate samples from a posterior conditional distribution given the effective population sizes. Thus, we must assume that our estimated posterior conditional distribution converges to the true joint distribution to a high degree of accuracy. Nevertheless, we feel comfortable making this assumption due to the strong convergence found between runs. We also used BEST (Liu and Pearl 2007) for these data and found similar results—for example, convergence across all parameters except branch lengths with stability in the species-tree estimates despite this. However, BEST required substantially longer run times making thorough experimentation more difficult. It remains unclear why *BEAST and BEST have opposite convergence behavior: Some possibilities include the prior distributions and the differences in posterior estimation, importance sampling versus MCMC. In the future, we hope that our results induce further work on this issue, especially the development of more robust MCMC transition kernels.

We also attempted to use the program BUCKy (Ane et al. 2007) to estimate a primary concordance tree with some success. BUCKy was able to complete an analysis using all 19 genes of 52 tips each in about 36 h (utilizing the previous analysis carried out by MrBayes). However, as described by Baum (2007), care must be taken in using phased data in a Bayesian concordance analysis. To properly deal with the phased nature of the data (specifically, the fact that each individual is represented by two distinct tips in each tree) would require subsampling of individuals repeatedly and thus substantial additional computational effort. Finally, we considered the minimize deep coalescences (MDC) method, and in particular, the implementation in PhyloNet (Than and Nakhleh 2009) with the MAP trees from our single-gene Bayesian analyses as input. The tree found using this criterion indicated paraphyly at the species level (Fig. 4b), which we viewed to be unrealistic in that it was not supported in any of the other analyses. Upon further examination, we found that this result was driven by two genes in the data; when these two genes were removed, the tree found by PhyloNet matched the tree returned by STEM (Fig. 4). We have described our analyses of this data set using both MDC and BUCKy in more detail in Kubatko and Gibbs (2010).

We note that all the analyses included here assume that there is no gene flow or other horizontal transfer between subspecies. The effect of gene flow on the ability to correctly infer the species tree was recently examined for several methods by Eckert and Carstens (2008). They found that the coalescent-based methods they examined (ESP-COAL and MDC) were somewhat robust to gene flow provided that the rate of gene flow was relatively low, whereas concatenation was affected to a larger extent. Due to the heterogeneous geographic distribution and high level of population genetic structure of several of the subspecies we examined (Gibbs et al. 1997; Anderson et al. 2009), we do not expect gene flow, if present at all, to be occurring at very high rates. This is confirmed by results using the program IMa (Hey and Nielsen 2007) to estimate levels of gene flow between subspecies of *S. catenatus* based on the loci described here—all estimates of migration rates were nonzero but low and of the magnitude that Eckert and Carstens (2008) demonstrate will have minimal effects on correctly inferring the species tree.

Finally, as discussed by Leaché (2009), a lack of samples from geographic areas that represent subspecies boundaries may lead to an overestimate of genealogical exclusively (see below) if gene flow leading to shared genotypes is present in these regions. The preliminary analyses of gene flow between our best-sampled taxa

(*S. c. catenatus* and *S. c. tergeminus*), which include samples close to the boundary between these subspecies, suggest that this is not the case but the possible impact of this bias for other subspecies is unknown at present. This possible bias could be especially important for subspecies of *S. miliarius* which are genetically similar yet for which only limited samples were collected.

Observations of Monophyly and Genealogical Sorting

Here, we have invoked a genealogical species concept (Baum and Shaw 1995) by implementing a series of tests for taxonomic distinctiveness and exclusive ancestry based on groups that were observed to be monophyletic in our analyses. These methods provide a statistical formalization of the recently invoked concept of phylogenetic species as lineages exhibiting monophyletic patterns in a majority of sampled loci, which are not contradicted by phylogenetic patterns at other loci (Dettman et al. 2003; Weisrock et al. 2006). Although some proposals for using observed monophyly to delineate genealogical species require monophyly to be well supported in a certain arbitrarily selected proportion of the loci (Hudson and Coyne 2002), we adopt the probabilistic approach proposed by Rosenberg (2007) and Cummings et al. (2008) for assessing the extent of the evidence against a single taxonomic group in the sample of loci. These approaches trade off the need to determine a "cutoff" value for the proportion of loci that must be observed to be monophyletic with the consideration of a *P* value for the null hypothesis of a single taxonomic entity. Because interpretation of *P* values (and more generally, probabilities of events) are familiar, this interpretation is more straightforward. In addition, when sample sizes within the groups of interest are large enough, it allows identification of taxonomic distinctiveness even before a majority of loci show monophyly, a desirable property for recent, rapid radiations (Knowles and Carstens 2007).

We note that although we have chosen to implement the tests suggested by Rosenberg (2007) and Cummings et al. (2008) for this purpose, other possibilities have been proposed. Among the most promising of these is a method by Carstens and Knowles (2007) in which gene tree topology probabilities under the coalescent model are used to compare support for a single taxonomic unit to support for distinct taxa using likelihood ratio tests. Although this test may be appropriate for our data, we have chosen not to implement it for several reasons. First, their test requires specification of the times of the speciation events in the tree so that gene tree probabilities can be computed. Because our times are estimated with fairly large confidence intervals, assigning specific times is difficult in this case. In addition, Rosenberg (2007) shows that his tests will be most powerful whenever the number of lineages sampled within each putative taxonomic unit is relatively large and evenly distributed across groups. This is certainly the case within the *S. catenatus* group, where 9 individuals within *S. c. catenatus* are included, and 4 and 5 individuals are included for *S. c. edwardsii* and *S. c. tergeminus*, respectively. Within *S. miliarius*, our sample sizes are much smaller, which perhaps contributes to the larger *P* values observed in this group. Future examination of relationships at the subspecies level within this group would be enhanced with more extensive sampling.

It is worth commenting on some additional features of the monophyly tests of Rosenberg (2007). In particular, it may seem unintuitive that observing monophyly at only one or a few genes,whereas several genes show nonmonophyletic relationships (e.g., see Sce vs. Sct comparison) can still lead to a relatively small *P* value. This occurs because a large number of lineages within a combined group of two taxa makes it very unlikely to observe monophyly of one of the groups *if there was actually no distinction between the two taxa.* Thus, even a single observation of monophyly can provide substantial evidence that two groups are distinct.

Taxonomic Implications

In our analysis, the tests of taxonomic distinctiveness show the strongest support for the phylogenetic distinctiveness of *S. c. catenatus* from *S. c. edwardsii* and *S. c. tergeminus*. Based on these results, morphological differences described by Gloyd (1940), and the allopatric distributions of each subspecies, we believe that there are multiple lines of evidence that strongly support the elevation of *S. c. catenatus* to full species status. In addition, there are indications of distinctiveness within many of the other groups. When considering the *egsi*, all subspecies show a strong indication of shared exclusive ancestry. Support for taxonomic distinctiveness is less pronounced using the monophyly based tests of Rosenberg (2007), however, which may be an indication that insufficient time has passed to allow sorting of lineages into their respective clades.

Estimates of Divergence Time

Estimates of divergence times based on the concatenated data set are roughly consistent with other gene analyses using fossil dates for molecular clock calibrations. Douglas et al. (2006) estimated the date of the origin of the two *Sistrurus* species as between 9.58 Ma and 10.23 Ma based on a gene tree for North American pitvipers as estimated from a single mtDNA gene (*ATP 6-8*) and multiple fossil calibrations. This is similar to our gene tree-based estimate for the same event of 9.45 Ma. Our estimate of 10.05 Ma from a fossil calibrated *BEAST analysis also agrees with the ages estimated by Douglas et al. (2006). The similarity of our two BEAST-based estimates is not surprising, however, because we are essentially recovering the prior distribution.

In theory, gene tree-based estimates should in general overestimate species or population divergences because most gene tree divergences will predate species and/or population divergences, with the degree to which these

estimates differ depending on the effective population size of ancestral taxa (Edwards and Beerli 2000). Our results provide an empirical example of the significant differences between these time estimates that likely arise because of the relatively large effective ancestral population sizes of these snakes in nature. Other examples of this pattern which draw on multilocus data are few but Jennings and Edwards (2005) found, based on a data set consisting of anonymous nuclear DNA loci, that species divergence time estimates were roughly three times smaller than those in which gene-based divergence estimates were used as proxies for species-level estimates. Our approach also represents a methodological advance by illustrating how a widely used approach for estimating node ages (use of fossil calibration points with BEAST) can now be extended to a species tree framework as represented by the program *BEAST.

Given that species-tree estimates are more appropriate for inferences about the timing of diversification within taxa, this implies that speciation events within North American snakes have occurred more recently and more rapidly than suggested by gene tree estimates that have been widely used. In particular, these analyses have suggested that a substantial amount of differentiation occurred in North American snakes including rattlesnakes during the late Miocene, Pliocene, and early Pleistocene periods (Burbrink et al. 2000; Pook et al. 2000; Douglas et al. 2006, 2007; Bryson et al. 2007; Castoe et al. 2007; Fontanella et al. 2008; Guiher and Burbrink 2008). Our results suggest that recalibration of these dates using species tree–based estimates would focus attention on geological and biological factors during the Pleistocene rather than earlier geological eras as key drivers in the diversification of North American snakes.

Insights into Sistrurus *Evolution*

The geographic locations of some (but not all) splits between genetically distinct forms coincide with phylogenetic breaks observed in other vertebrates, implicating a role for "hard vicariance" as a mechanism leading to differentiation in these snakes. The locations of boundaries between a number of the *Sistrurus* subspecies now shown to be genetically distinct coincide with a number of well-known geographic features that are associated with phylogeographic splits in other taxa. For example, if we assume that "intergrade" samples in Missouri are *S c. tergeminus* individuals (Gibbs et al. 2011), the boundary between *S. c. catenatus* and *S. c. tergeminus* coincides with the Mississippi River, which is associated with splits in other vertebrate species (see Soltis et al. 2006 for review) including some snakes (e.g., Burbrink et al. 2000; Burbrink 2002). Likewise, the boundary between *S. m. streckeri* and *S. m. barbouri* in western Alabama (Campbell and Lamar 2004) falls close to the Tombigbee River which, again, has been associated with phylogeographic structure in a number of species (Soltis et al. 2006) including two species of *Nerodia* water snakes (Lawson 1987). As argued for other vertebrates, these associations imply that vicariant effects due to these rivers acting as large-scale barriers to gene flow combined with divergence in isolated refugia on either sides of these barriers have likely been major modes of speciation in *Sistrurus*.

However, not all genetically distinct subspecies have boundaries that coincide with an obvious biogeographic barrier. The split between *S. c. tergeminus* and *S. c. edwardsii* occurs in North Central Texas (Campbell and Lamar 2004) where there is no obvious large-scale geographical feature that could act as an isolating barrier. Consistent with this lack of such feature, there are a number of reptiles with continuous distributions through this area that show no genetic discontinuities (Burbrink 2002; Leaché and Reeder 2002; Leaché and McGuire 2006; but see Fontanella et al. 2008). However, there is a major change in habitat type used by *Sistrurus* through this area: *S. c. tergeminus* is found in mesic grasslands and marsh habitats, whereas *S. c. edwardsii* occurs in xeric grassland habitat (Holycross and Mackessy 2002; Campbell and Lamar 2004). This shift in habitat is associated with a shift in diet (Holycross and Mackessy 2002) and so at least raises the possibility that some form of ecological speciation associated with this habitat shift may be responsible for differentiation between these subspecies.

Conservation Implications

One of the subspecies studied here (*S. c. catenatus*) is seriously threatened due to significant population declines (Szymanski 1998) and has been named by the US Fish and Wildlife Service as a candidate species for listing under the US Endangered Species Act (ESA) (US Federal Register 1999). Populations of this subspecies have been significantly impacted by habitat destruction and degradation and many are showing substantial declines in numbers. This study provides clear genetic evidence that this subspecies is highly distinct taxonomically, represents a phylogenetically distinct lineage relative to *S. c. tergeminus* and *S. c. edwardsii* and therefore should be elevated to full species status under a genealogical species concept. This reinforces the need to classify *S. c. catenatus* as a "distinct population segment" under the ESA because our results provide support for both criteria used to identify such units (US Federal Register 1996), namely that the evidence for reciprocal monophyly indicates that *S. c. catenatus* is both reproductively isolated and represents a distinct and unique portion of the evolutionary history of this group of snakes.

SUPPLEMENTARY MATERIAL

Supplementary material, including data files and/or online-only appendices, can be found at http://www.sysbio.oxfordjournals.org/.

Funding

This research was supported by funds from the National Science Foundation (DMS-0702277 to L.S.K.) and the Ohio State University and the Ohio Division of Wildlife (H.L.G.). The research of E.W. Bloomquist was partially supported by the National Science Foundation under Agreement No. 0635561.

Acknowledgments

This study would not have been possible without assistance from a number of people with the collection of samples. In particular, we thank J. Chiucchi, D. Wynn, G. Johnson, M. Kowalski, K. Prior, D. Harvey, M. Dreslik, F. Durbian, J. Collins, T. Taggart, S. Mackessy, A. Holycross, Z. Orr, T. Farrell, and D. Bender for either help in the field or for providing samples. We acknowledge the pioneering work on *Sistrurus* phylogenetics by M.E. Douglas and M.R. Douglas that stimulated our study. We also thank J. Diaz for help in the lab, D. Pearl, L. Liu, L. Nakhleh, J. Heled, and C. Ane for helpful discussion regarding aspects of species tree estimation, R. King for comments, and the Ohio Supercomputer Center and the Mathematical Biosciences Institute at Ohio State University for access to computing resources.

References

Anderson C.D., Gibbs H.L., Douglas M.E., Holycross A.T. 2009. Conservation genetics of the Desert massasauga rattlesnake (*Sistrurus catenatus edwardsii*). Copeia. 2009:740–747.

Ane C., Larget B., Baum D.A., Smith S.D., Rokas A. 2007. Bayesian estimation of concordance among gene trees. Mol. Biol. Evol. 24: 412–426.

Baum D.A. 2007. Concordance trees, concordance factors, and the exploration of reticulate genealogy. Taxon. 56:417–426.

Baum D.A., Shaw K.L. 1995. Genealogical perspectives on the species problem. In: Hoch P.C., Stephenson A.G., editors. Experimental and molecular approaches to plant biosystematics. St. Louis (MO): Missouri Botanical Garden. p. 289–303.

Belfiore N.M., Liu L., Moritz C. 2008. Multi-locus phylogenetics of a rapid radiation in the Genus *Thomomys* (Rodentia: Geomyidae). Syst. Biol. 57:294–310.

Brumfield R.T., Beerli P., Nickerson D.A., Edwards S.V. 2003. The utility of single nucleotide polymorphisms in inferences of population history. Trends Ecol. Evol. 18:249–256.

Brumfield R.T., Liu L., Lum D.E., Edwards S.V. 2008. Comparison of species tree methods for reconstructing the phylogeny of Bearded Manakins (Aves: Pipridae, Manacus) from multi-locus sequence data. Syst. Biol. 57:719–731.

Bryson R.W., Pastorini J., Burbrink F.T., Forstner M.R. 2007. A phylogeny of the *Lampropeltis mexicana* complex (Serpentes: Colubridae) based on mitochondrial DNA sequences suggests evidence for species-level polyphyly within *Lampropeltis*. Mol. Phylogenet. Evol. 43:674–684.

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

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

Campbell J.A., Lamar W. 2004. The venomous reptiles of the western hemisphere. Ithaca (NY): Cornell University Press.

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

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

Castoe T.A., Spencer C., Parkinson C.L. 2007. Phylogeographic structure and historical demography of the western diamondback rattlesnake (*Crotalus atrox*): a perspective on North American desert biogeography. Mol. Phylogenet. Evol. 42:193–212.

Creer S., Malhotra A., Thorpe R.S., Pook C.E. 2005. Targeting optimal introns for phylogenetic analyses in non-model taxa: experimental results in Asian pitvipers. Cladistics. 21:390–395.

Cummings M.P., Neel M.C., Shaw K.L. 2008. A genealogical approach to quantifying lineage divergence. Evolution. 62:2411–2422.

Dettman J.R., Jacobson D.J., Taylor J.W. 2003. A multi-locus genealogical approach to phylogenetic species recognition in the model eukaryote *Neurospora*. Evolution. 57:2703–2720.

Douglas M.E., Douglas M.R., Schuett G.W., Porras L.W., Holycross A.T. 2002. Phylogeography of the western rattlesnake (*Crotalus viridis*) complex, with emphasis on the Colorado Plateau. In: Schuett G. W., Hoggren M., Douglas M. E., Greene H. W., editors.) Biology of the vipers. Eagle Mountain (Utah): Eagle Mountain Publishing, LC. p. 11–50.

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

Douglas M.E., Douglas M.R., Schuett G.W., Porras L.W., Thomason B.L. 2007. Genealogical concordance between mitochondrial and nuclear DNAs supports species recognition of the Panamint rattlesnake (*Crotalus mitchellii stephensi*). Copeia. 2007:920–932.

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

Eckert A.J., Carstens B.C. 2008. Does gene flow destroy phylogenetic signal? The performance of three methods for estimating species phylogenies in the presence of gene flow. Mol. Phylogenet. Evol. 49:832–842.

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

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

Edwards S.V., Liu L., Pearl D.K. 2007. High-resolution species trees without concatenation. Proc. Natl. Acad. Sci. U.S.A. 104:5936–5941.

Evans P.D., Gloyd H.K. 1948. The subspecies of the massasauga, *Sistrurus catenatus*, in Missouri. Bull. Chicago Acad. Sci. 8:225–232.

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

Friesen V.L., Congdon B.C., Kidd M.G., Birt T.P. 1999. Polymerase chain reaction (PCR) primers for the amplification of five nuclear introns in vertebrates. Mol. Ecol. 8:2141–2152.

Gibbs H.L., Corey S.J., Blouin-Demers G., Prior K.A., Weatherhead P.J. 2006. Hybridization between mtDNA-defined phylogeographic lineages of black ratsnakes (*Pantherophis* sp.). Mol. Ecol. 15:3755–3767.

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

Gibbs H.L., Murphy M., Chiucchi J. 2011. Genetic identity of endangered massasauga rattlesnakes (*Sistrurus sp.*) in Missouri. Conserv. Genet. doi:10.1007/s10592-010-0151-3.

Gibbs H.L., Prior K., Weatherhead P.J., Johnnson G. 1997. Genetic structure of populations of the threatened eastern massasauga rattlesnake, *Sistrurus c. catenatus*: evidence from microsatellite DNA markers. Mol. Ecol. 6:1123–1132.

Gibbs H.L., Rossiter W. 2008. Rapid evolution by positive selection and gene gain and loss: PLA_2 venom genes in closely related *Sistrurus* rattlesnakes with divergent diets. J. Mol. Evol. 66:151–166.

Gibbs H.L., Sanz L., Calvete J.J. 2009. Snake population venomics: proteomics-based analyses of individual variation reveals significant gene regulation effects on venom protein expression in *Sistrurus* rattlesnakes. J. Mol. Evol. 68:113–125.

Gloyd H.K. 1935. The subspecies of *Sistrurus miliarius*. Occasional Papers, Museum of Zoology, University of Michigan 322:1–7.

Gloyd H.K. 1940. The Rattlesnakes, Genera *Sistrurus* and *Crotalus*. Special Publ. Chicago Acad. Sci. 4:104–118.

Gloyd H.K. 1955. A review of the massasaugas, *Sistrurus catenatus*, of the southwestern United States (Serpentes: Crotalidae). Bull. Chicago Acad. Sci. 10:83–98.

Gonçalvesa H., Martínez-Solanod I., Ferranda N., García-Parísc M. 2007. Conflicting phylogenetic signal of nuclear vs mitochondrial DNA markers in midwife toads (Anura, Discoglossidae, Alytes): deep coalescence or ancestral hybridization? Mol. Phylogenet. Evol. 44:494–500.

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

Harrigan R.J., Mazza M.E., Sorenson M.D. 2008. Computation versus cloning: evaluation of two methods for haplotype determination. Mol. Ecol. Resour. 8:1239–1248.

Heled J., Drummond A.J. 2010. Bayesian inference of species trees from multilocus data. Mol. Biol. Evol. 27:570–580.

Hey J., Nielsen R. 2007. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. Proc. Natl. Acad. Sci. U.S.A. 104:2785–2790.

Holycross A.T., Mackessy S.P. 2002. Variation in the diet of *Sistrurus catenatus* (massasauga) with emphasis on *Sistrurus catenatus edwardsii* (desert massasauga). J. Herpetol. 36:454–464.

Hudson R.R. 1992. Gene trees, species trees and the segregation of ancestral alleles. Genetics. 131:509–512.

Hudson R.R., Coyne J.A. 2002. Mathematical consequences of the genealogical species concept. Evolution. 56:1557–1565.

Hudson R.R., Turelli M. 2003. Stochasticity overrules the "three-times"rule: genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. Evolution. 57:182–190.

Huelsenbeck J.P., Hillis D.M., Nielsen R. 1996. A likelihood ratio test of monophyly. Syst. Biol. 45:546–558.

Jennings W.B., Edwards S.V. 2005. Speciational history of Australian grass finches (*Poephila*) inferred from 30 gene trees. Evolution. 59:2033–2047.

Klauber L.M. 1972. Rattlesnakes: their habits, life histories, and influence on mankind. 2nd ed. Berkeley (CA): University of Califoria Press.

Knowles L.L., Carstens B.C. 2007. Delimiting species without monophyletic gene trees. Syst. Biol. 56:887–895.

Knowles L.L., Kubatko L.S. 2010. Estimating species trees: practical and theoretical aspects. Hoboken (NJ): Wiley-Blackwell.

Kosakovsky Pond S.L., Posada D., Gravenor M.B., Woelk C.H., Frost S.D.W. 2006. Automated phylogenetic detection of recombination using a genetic algorithm. Mol. Biol. Evol. 23:1891–1901.

Kubatko L.S., Carstens B.C., Knowles L.L. 2009. STEM: species tree estimation using maximum likelihood for gene trees under coalescence. Bioinformatics. 25:971–973.

Kubatko L., Degnan J. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56(1): 17–24.

Kubatko L.S., Gibbs H.L. 2010. Estimating species relationships and taxon distinctiveness in Sistrurus rattlesnakes using multilocus data. In: Knowles L.L., Kubatko L.S., editors. Estimating species trees: practical and theoretical aspects. Hoboken (NJ): Wiley-Blackwell. p. 193–207.

Kuhner M.K. 2006. LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. Bioinformatics. 22:768–770.

Lawson R. 1987. Molecular studies of thamnophiine snakes. 1. The phylogeny of the genus *Nerodia*. J. Herpetol. 21:140–157.

Leaché A.D. 2009. Species tree discordance traces to phylogeographic clade boundaries in North American fence lizards (*Sceloporus*). Syst. Biol. 58:547–559.

Leaché A.D. 2010. Species trees for spiny lizards (Genus *Sceloporus*): identifying points of concordance and conflict between nuclear and mitochondrial data. Mol. Phylogenet. Evol. 54:162–171.

Leaché A.D., McGuire J. 2006. Phylogenetic relationships of horned lizards (*Phrynosoma*) based on nuclear and mitochondrial data: evidence for a misleading mitochondrial gene tree. Mol. Phylogenet. Evol. 39:628–644.

Leaché A.D., Reeder T.W. 2002. Molecular systematics of the eastern fence lizard (*Sceloporus undulatus*): A comparison of parsimony, likelihood, and Bayesian approaches. Syst. Biol. 51:44–68.

Linnen C.R., Farrell B.D. 2007. Mitonuclear discordance is caused by rampant mitochondrial introgression in *Neodiprion* (Hymenoptera: Diprionidae) sawflies. Evolution. 61:1417–1438.

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

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

Liu L., Yu L., Pearl D.K., Edwards S.V. 2009b. Estimating species phylogenies using coalescence times among sequences. Syst. Biol. 58: 468–477.

Mackessy S.P. 2005. Desert Massasauga Rattlesnake (*Sistrurus catenatus edwardsii*): a technical conservation assessment. USDA Forest Service, Rocky Mountain Region. Available from: http://www.fs.fed.us/r2/projects/scp/assessments/massasauga.pdf.

Maddison W.P. 1997. Gene trees in species trees. Syst. Biol. 46:523–536.

Minin V., Abdo Z., Joyce P., Sullivan J. 2003. Performance-based selection of likelihood models for phylogeny estimation. Syst. Biol. 52:674–683.

Murphy R.W., Fu J., Lathrop A., Feltham J.V., Kovak V. 2002. Phylogeny of the rattlesnakes (*Crotalus* and *Sistrurus*) inferred from sequences of five mitochondrial DNA genes. In: Schuett G.W., Hoggren M., Douglas M.E., Greene H.W., editors. Biology of the vipers. Eagle Mountain (Utah): Eagle Mountain Publishing, LC. p. 69–92.

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

Parkinson C.L., Chippindale P., Campbell J. 2002. Multigene analyses of pitviper phylogeny with comments on their biogeographical history. In: Schuett G.W., Hoggren M., Douglas M.E., Greene H.W., editors. Biology of the vipers. Eagle Mountain (Utah): Eagle Mountain Publishing, LC. p. 93–110.

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

Parmley D., Holman J.A. 2007. Earliest fossil record of a Pigmy rattlesnake (Viperidae: *Sistrurus* Garman). J. Herpet. 41:141–144.

Pook C.E., Wuster W., Thorpe R.S. 2000. Historical biogeography of the western rattlesnake (Serpentes: Viperidae: *Crotalus viridis*) inferred from mitochondrial DNA sequence information. Mol. Phylogenet. Evol. 15:269–282.

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

Rosenberg N.A. 2003. The shapes of neutral gene genealogies in two species: probabilities of monophyly, paraphyly, and polyphyly in a coalescent model. Evolution. 57:1465–1477.

Rosenberg N.A. 2007. Statistical tests for taxonomic distinctiveness from observations of monophyly. Evolution. 61:317–323.

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

Sanz L., Gibbs H.L., Mackessy S.P., Calvete J.J. 2006. Venom proteomes of closely related *Sistrurus* rattlesnakes with divergent diets. J. Proteome Res. 5:2098–2112.

Schelly R., Salzburger W., Koblmuller S., Duftner N., Sturmbauer C. 2005. Phylogenetic relationships of the lamprologine cichlid genus *Lepidiolamprologus* (Teleostei: Perciformes) based on mitochondrial and nuclear sequences, suggesting introgressive hybridization. Mol. Phylogenet. Evol. 38:426–438.

Schluter D. 2000. The ecology of adaptive radiation. Oxford (UK): Oxford University Press.

Serb J.M., Phillips C.A., Iverson J.B. 2001. Molecular phylogeny and biogeography of *Kinosternon avescens* based on complete mitochondrial control region sequences. Mol. Phylogenet. Evol. 18:149–162.

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

Stephens M., Smith N., Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. 68:978–989.

Swofford D.L. 2003. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4. Sunderland (MA): Sinauer Associates.

Szymanski J. 1998. Status assessment for the Eastern Massasauga (*Sistrurus c. catenatus*) 1998. Fort Snelling (MN): U.S. Fish and Wildlife Service, Endangered Species Division .

Than C., Nakhleh L. 2009. Species tree inference by minimizing deep coalescences. PLOS Comput. Biol. 5(9):e1000501.

US Federal Register 1996. Policy regarding the recognition of distinct vertebrate population segments under the Endangered Species Act. 61:4722.

US Federal Register 1999. Endangered and threatened wildlife and plants. 64:57534.

Vidal N., Hedges S.B. 2002. Higher-level relationships of caenophidian snakes inferred from four nuclear and mitochondrial genes. Comptes Rendus Biologies 325:987–995.

Weisrock D.W., Shaffer H.B., Storz B.L., Storz S.R., Voss S.R. 2006. Multiple nuclear gene sequences identify phylogenetic species boundaries in the rapidly radiating clade of Mexican ambystomatid salamanders. Mol. Ecol. 15:2489–2503.

Wiens J.J., Kuczynski C.A., Smith S.A., Mulcahy D.G., Sites Jr. J.W., Townsend T.M., Reeder T.W. 2008. Branch lengths, support, and congruence: testing the phylogenomic approach with 20 nuclear loci in snakes. Syst. Biol. 57:420–431.

Wiens J.J., Penkrot T.A. 2002. Delimiting species using DNA and morphological variation and discordant species limits in spiny lizards (*Sceloporus*). Syst. Biol. 51:69–91.

Yang Z. 2002. Likelihood and Bayes estimation of ancestral population sizes in Hominoids using data from multiple loci. Genetics. 162:1811–1823.

Yang Z., Rannala B. 2010. Bayesian species delimitation using multilocus sequence data. Proc. Natl. Acad. Sci. U.S.A. 107:9264–9269.