Syst. Biol. 58(6):547–559, 2009
© The Author(s) 2009. Published by Oxford University Press, on behalf of the Society of Systematic Biologists. All rights reserved.
For Permissions, please email: journals.permissions@oxfordjournals.org
DOI:10.1093/sysbio/syp057
Advance Access publication on September 21, 2009

Species Tree Discordance Traces to Phylogeographic Clade Boundaries in North American Fence Lizards (*Sceloporus*)

ADAM D. LEACHÉ[1,2,*]

[1]*Museum of Vertebrate Zoology, 3101 Valley Life Sciences Building, University of California, Berkeley, CA 94720, USA; and*
[2]*Present address: Genome Center and Section of Evolution and Ecology, One Shields Avenue, University of California, Davis, CA 95616, USA;*
**Correspondence to be sent to: Genome Center and Section of Evolution and Ecology, One Shields Avenue, University of California, Davis, CA 95616, USA E-mail: aleache@ucdavis.edu.*

Abstract.—I investigated the impacts of phylogeographic sampling decisions on species tree estimation in the *Sceloporus undulatus* species group, a recent radiation of small, insectivorous lizards connected by parapatric and peripatric distribution across North America, using a variety of species tree inference methods (Bayesian estimation of species trees, Bayesian untangling of concordance knots, and minimize deep coalescences). Phylogenetic analyses of 16 specimens representing 4 putative species within *S. "undulatus"* using complete (8 loci, >5.5 kb) and incomplete (29 loci, >23.6 kb) nuclear data sets result in species trees that share features with the mitochondrial DNA (mtDNA) genealogy at the phylogeographic level but provide new insights into the evolutionary history of the species group. The concatenated nuclear data and mtDNA data both recover 4 major clades connecting populations across North America; however, instances of discordance are localized at the contact zones between adjacent phylogeographic groups. A random sub-sampling experiment designed to vary the phylogeographic samples included across hundreds of replicate species tree inferences suggests that inaccurate species assignments can result in inferred phylogenetic relationships that are dependent upon which particular populations are used as exemplars to represent species and can lead to increased estimates of effective population size (Θ). For the phylogeographic data presented here, reassigning specimens with introgressed mtDNA genomes to their prospective species, or excluding them from the analysis altogether, produces species tree topologies that are distinctly different from analyses that utilize mtDNA-based species assignments. Evolutionary biologists working at the interface of phylogeography and phylogenetics are likely to encounter multiple processes influencing gene trees congruence, which increases the relevance of estimating species trees with multilocus nuclear data and models that accommodate deep coalescence. [Bayesian analysis; deep coalescence; gene flow; gene trees; introgression; lineage sorting; species delimitation; taxon sampling.]

The growing ease of acquiring genomic scale data sets to study nonmodel organisms enables systematic biologists to assemble the tree of life with increasing detail. However, it is widely recognized that the stochastic process of lineage sorting can cause discordance among gene trees inferred from independent loci, which may provide inaccurate estimations of the species tree in separate and combined phylogenetic analyses (Pamilo and Nei 1988; Maddison 1997; Edwards et al. 2007; Rannala and Yang 2008). New multilocus phylogenetic methods are emerging to reconstruct species trees from independent loci (reviewed by Edwards 2009), and some approaches are integrating population genetics and phylogenetic methods to explicitly accommodate the process of lineage sorting (Liu and Pearl 2007; Liu et al. 2008). The fact that the majority of independent gene trees may be incongruent with the true underlying species tree under some circumstances (Degnan and Rosenberg 2006; Kubatko and Degnan 2007) is a clear warning against relying on a single locus for evolutionary analysis and species delimitation and for a simple concatenation or majority-rule approach to multilocus analysis.

Population dynamics are responsible for many gene tree incongruence problems, and studies centered at the interface of phylogeography and phylogenetics are faced with the difficult task of estimating species histories in situations where incomplete lineage sorting and gene flow among recently diverged species may be common (Belfiore et al. 2008; Brumfield et al. 2008; Carling and Brumfield 2008; Eckert and Carstens 2008). These processes are difficult to differentiate using topological information alone because they result in similar genealogical patterns (Funk and Omland 2003). Spatial patterns of gene tree incongruence can aid in the differentiation of these processes, and the localization of discordance near phylogeographic boundaries may be a signature of current or historical gene flow (Leaché and McGuire 2006; McGuire et al. 2007).

Phylogeographic studies often utilize increased geographic sampling to refine clade boundaries and identify the location of contact zones (e.g., Morando et al. 2003); however, it is unclear how including specimens from the vicinity of species boundaries will impact species tree inference because gene flow is not accommodated by current methods (Eckert and Carstens 2008; Liu et al. 2008). Increasing the number of individuals sampled within species improves species tree accuracy (Maddison and Knowles 2006), but this improvement could come at a cost for phylogeographic studies if individuals from admixed populations are introduced into the analysis. Sampling species from distant allopatric populations, with the assumption that any instances of gene tree incongruence will be the result of deep coalescence and not gene flow, may help mitigate potentially confounding affects of gene flow on species tree inference but would place a major limitation on the scope of phylogeographic studies.

The *Sceloporus undulatus* species group is a radiation of 9 phrynosomatid lizard species connected by parapatric or peripatric distributions across the United States and north central Mexico (Leaché and Reeder 2002).

Until recently, 4 of these species (*Sceloporus consobrinus*, *Sceloporus cowlesi*, *Sceloporus tristichus*, and *S. undulatus*) were treated as a single polytypic species, *S. "undulatus"* (Leaché and Reeder 2002). The current species-level phylogeny, taxonomy, and phylogeographic assessment of *S. "undulatus"* is based on a mitochondrial DNA (mtDNA) genealogy (Fig. 1a), although there are reasons to suspect that this genealogy is not reflective of the true species phylogeny. The probability of mismatch between a gene tree and species tree increases when ancestral population sizes are large and the times between population-splitting events are short (Pamilo and Nei 1988), a combination of demographic characteristics that are likely to apply to species in the *S. undulatus* group, some of which have broad geographic distributions or are connected to other species in the mtDNA genealogy by short internodes (Fig. 1). In addition, mtDNA introgression occurring at a hybrid zone in Arizona between *S. cowlesi* and *S. tristichus* (2 nonsister species in the mtDNA genealogy) introduces gene flow as a potential source of conflict between the mtDNA gene tree and species tree (Leaché and Cole 2007). A primary focus of this study is to determine how the inclusion of specimens from this contact zone, or other putative species boundaries, impacts species tree inference.

Methods

Taxon Sampling

A total of 21 specimens representing 9 species of *Sceloporus* were included in the study (Table 1). The only member of the *S. undulatus* group that was not available for this study was *Sceloporus exsul*, a species with a limited distributional range in central Mexico (Dixon et al. 1972). A total of 16 specimens of *S. "undulatus"* were included to represent each of the 4 mtDNA-based species, *S. consobrinus, S. cowlesi, S. tristichus*, and *S. undulatus* (4 specimens each; Table 1). Phylogeographic samples within *S. "undulatus"* were selected to maximize the geographic coverage within mtDNA clades. In addition, some specimens were sampled from the vicinity of mtDNA clade boundaries, and this included a specimen of *S. cowlesi* from the *S. cowlesi* + *S. tristichus* hybrid zone in Arizona (Fig. 1c).

Multilocus Nuclear Data

I collected sequence data from 29 nuclear loci (Table 2). Four of these loci (*BDNF*, *PNN*, *R*35, and *RAG-1*) represent nuclear exons, whereas the remaining 25 are anonymous nuclear loci isolated from a genomic library constructed from 2 individuals of *S. "undulatus"* from New Mexico (Rosenblum et al. 2007; Supplementary Table S1, available from http://www.sysbio.oxfordjournals.org/). Anonymous nuclear loci have the potential to contain a substantial number of single nucleotide polymorphisms, making them useful targets for reconstructing phylogeny among closely related and recently diverged species (Brumfield et al. 2003). All loci were sequenced in forward and reverse directions using an ABI 3730 capillary sequencer. Contiguous DNA sequences were aligned and edited using Sequencher v4.8, and multiple sequence alignments were generated using Muscle v3.6 (Edgar 2004). Open reading frames for protein coding exons were identified using Mesquite v2.5 (Maddison W.P. and Maddison D.R. 2008). All sequences are deposited in GenBank (Accession Nos. GQ494358-494867). Sequence alignments and trees are deposited in TreeBase (Study Accession No.: S2457).

Gene Trees and Concatenated Data Phylogeny

The 29 nuclear loci were concatenated to conduct partitioned maximum likelihood (ML) analysis using RAxML-VI-HPC v7.0.4 (Stamatakis 2006) and partitioned Bayesian analysis using MrBayes v3.1.2 (Ronquist and Huelsenbeck 2003). The Akaike information criterion (AIC) in MrModeltest v2.2 (Nylander 2004) was used to determine the best-fit nucleotide substitution model for each locus, resulting in 29 data partitions. Nonparametric bootstrapping (100 replicates) was used to generate support values for ML analyses of the concatenated data and for each nuclear locus. All ML analyses utilized the general time-reversible (GTR) + I + Γ model. Partitioned Bayesian analyses used the substitution models selected using the AIC in MrModeltest and were run for 50 million generations (sampling every 25 000 steps). Convergence was assessed using cumulative posterior probability plots constructed using the program Are We There Yet? (Nylander et al. 2008). The concatenation approach is analogous to the total evidence philosophy in systematic biology advocated by Kluge (1989) and may offer the advantage of revealing a predominant (or underlying) species phylogeny despite the presence of data conflict (e.g., Rokas et al. 2003).

Multilocus Network Approach

Phylogenetic trees are not always the most appropriate tools for visualizing intraspecific relationships because gene evolution does not necessarily follow a strictly bifurcating model (Posada and Crandall 2001). Network representations of data circumvent this problem and can help identify hybrid individuals or sequences that have undergone recombination (Bryant and Moulton 2004; Joly and Bruneau 2006). This characteristic of genetic networks enables them to provide portrayals of species relationships that are distinct from standard phylogenetic trees.

In order to produce a multilocus genetic network linking the 21 specimens of *Sceloporus* included in this study, I first calculated the genetic distance among alleles at each nuclear locus. Only loci with complete data for all specimens were included in the analysis. Heterozygous insertions and deletions were resolved using CodonCode Aligner v2.0.4 (CodonCode Corp., Dedham, MA) Resolving the phase of heterozygous genotypes was accomplished using PHASE v2.1.1 (Stephens and Donnelly 2003). I tested for intragenic

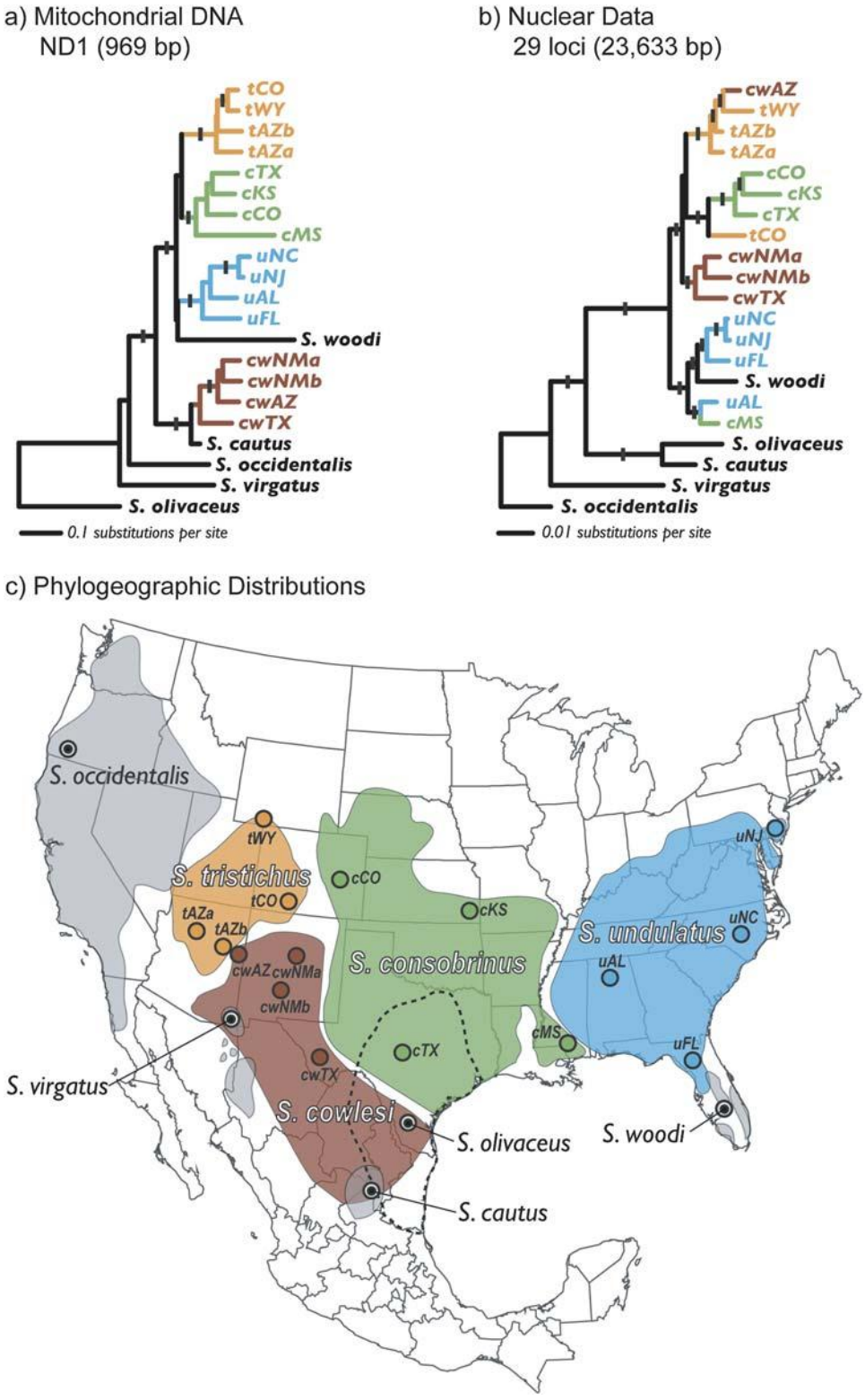


FIGURE 1. Phylogenetic relationships within the *Sceloporus undulatus* species group inferred from a) mtDNA and b) concatenated nuclear data (29 loci). Both trees are based on partitioned Bayesian analyses. For the mtDNA genealogy (a), nodes supported by posterior probability values $\geq$0.95 are indicated with black bars. The partitioned Bayesian analysis of the concatenated nuclear data (b) supports all nodes with posterior probability values $\geq$0.99, and the 16 nodes with ML bootstrap values $\geq$70 are indicated with black bars. c) The 4 major phylogeographic groups within *S. undulatus* are color-coded according to the mtDNA genealogy (Leaché and Reeder 2002; Leaché and Cole 2007).

TABLE 1. Specimens of *Sceloporus* included in the study

Species	Locality	Specimen ID	Nuclear loci	Voucher[a]
Sceloporus "undulatus"				
Sceloporus consobrinus	Colorado, Lincoln Co.	cCO	27	ADL 192
S. consobrinus	Kansas, Cherokee Co.	cKS	25	KU 289053
S. consobrinus	Mississippi, Stone Co.	cMS	24	LSU 55894
S. consobrinus	Texas, Kimble Co.	cTX	26	LVT 00365
Sceloporus cowlesi	Arizona, Apache Co.	cwAZ	27	AMNH R154059
S. cowlesi	New Mexico, Bernallio Co.	cwNMa	27	SDSU 4252
S. cowlesi	New Mexico, Otero Co.	cwNMb	21	SDSU 4218
S. cowlesi	Texas, Brewster Co.	cwTX	27	TWR 947
Sceloporus tristichus	Arizona, Coconino Co.	tAZa	26	LVT 2287
S. tristichus	Arizona, Navajo Co.	tAZb	25	AMNH R154447
S. tristichus	Colorado, Costilla Co.	tCO	28	ADL 271
S. tristichus	Wyoming, Sweetwater Co.	tWY	27	AMNH R154490
Sceloporus undulatus	Alabama, Madison Co.	uAL	28	ADL 303
S. undulatus	Florida, Hamilton Co.	uFL	28	MVZ 150110
S. undulatus	North Carolina, Bladen Co.	uNC	29	MVZ 150089
S. undulatus	New Jersey, Ocean Co.	uNJ	28	SDSU 4181
Sceloporus cautus	Mexico, Nuevo Leon	*S. cautus*	24	MZFC 7414
Sceloporus occidentalis	Oregon, Jackson Co.	*S. occidentalis*	29	SDSU 3956
Sceloporus olivaceus	Texas; La Salle Co.	*S. olivaceus*	9[b]	LSUMZ 48750
Sceloporus virgatus	Arizona, Cochise Co.	*S. virgatus*	21	LSUMZ 48759
Sceloporus woodi	Florida, Highlands Co.	*S. woodi*	28	MVZ 150112

Notes: Species assignments within *S. undulatus* are based on the mtDNA genealogy. The specimen ID corresponds to labels used for terminals in the phylogenetic analyses. The number of nuclear loci sequenced for each sample is indicated.
[a]Only 9 nuclear loci were targeted for *S. olivaceus*.
[b]Personal collector numbers are as follows: ADL, Adam D. Leaché; TWR, Tod W. Reeder.

recombination using the difference of sums of squares test in TOPALi v2.5 (McGuire and Wright 2000). Genetic distance matrices among alleles at each locus were calculated using uncorrected *p* distances and the HKY85 (Hasegawa et al. 1985) model using PAUP v4.0b10 (Swofford 2001). The genetic distance matrices for separate loci were combined into a single distance matrix of specimens using the program POFAD v1.03

TABLE 2. Nuclear loci sequenced for the *Sceloporus undulatus* species group

Locus	Sample size	Characters	Variable sites	Parsimony informative characters	Nucleotide substitution model
BDNF	21	670	7	3	HKY
PNN	21	934	24	7	GTR + Γ
R35	21	658	26	14	GTR + I
RAG-1	21	1055	41	17	HKY + Γ
Sun-006	21	595	45	20	HKY + Γ
Sun-008	21	677	47	12	HKY + Γ
Sun-032	21	449	30	14	HKY
Sun-037	21	478	30	14	HKY + Γ
Sun-020	20	695	22	7	HKY + I
Sun-021	20	974	49	19	HKY
Sun-023	20	929	75	28	GTR + I + Γ
Sun-024	20	1086	14	3	K80
Sun-026	20	1523	53	16	HKY + Γ
Sun-027	20	550	26	19	HKY + I
Sun-033	20	1103	61	24	HKY + I
Sun-035	20	560	34	10	K80
Sun-012	19	468	20	4	HKY
Sun-007	18	549	17	5	GTR
Sun-029	18	680	39	13	HKY
Sun-038	18	1137	52	18	GTR
Sun-036	17	1361	69	16	HKY + Γ
Sun-014	17	1283	65	35	HKY + Γ
Sun-016	17	993	45	21	GTR + I
Sun-009	16	683	38	12	HKY + I
Sun-030	16	707	24	5	HKY
Sun-003	13	254	7	1	K80
Sun-017	13	950	61	26	HKY
Sun-028	12	835	51	27	GTR + I
Sun-019	12	797	46	27	HKY + I
Concatenated nuclear data	21	23 633	1118	437	GTR + I + Γ
mtDNA (ND1)	21	969	332	226	GTR + I + Γ

Notes: Sample size indicates the number of specimens sequenced for a particular locus (maximum = 21). Nucleotide substitution models were selected using the AIC criterion in MrModeltest v2.2 (Nylander 2004).

(Joly and Bruneau 2006). A genetic network among specimens was constructed using the NeighborNet algorithm (Bryant and Moulton 2004) in SplitsTree v4.6 (Huson and Bryant 2006).

Bayesian Estimation of Species Trees

I used the hierarchical Bayesian model implemented in Bayesian estimation of species trees (BEST) v2.2 (Liu et al. 2008) to estimate species trees for the *S. undulatus* group that take into account incomplete lineage sorting and accommodate multiple alleles within species. This Bayesian species tree inference approach estimates the joint posterior distribution of gene trees from unlinked loci, which assumes that loci are correlated by their shared species history (Liu and Pearl 2007). Compared with the concatenation approach, the joint model can provide more robust estimates of species trees using fewer loci (Edwards et al. 2007). This method requires that species assignments are established a priori, and species assignments based on the mtDNA genealogy (Fig. 1a and Table 1) were used as a starting point for species tree analysis.

To avoid complications that could arise from having missing taxa in the component gene trees, only the 8 nuclear loci with complete data for all 21 specimens were used in the BEST analyses. Two Markov chain Monte Carlo (MCMC) runs were initiated at different starting seeds and allowed to proceed for 200 million generations (sampling every 100 000 steps). All analyses were run using genotype data. Convergence was assessed using burn-in plots of likelihood values and parameter estimates. The gene mutation prior was set at (0.5, 1.5) for all analyses. The prior distribution for the effective population size parameter theta (Θ) is modeled using an inverse gamma distribution, which is a 2-parameter probability distribution with a mean $= \beta/(\alpha - 1)$ (when $\alpha > 1$). Small values of Θ should reduce the influence of this prior on the estimated species trees (Liu et al. 2008), whereas high values of Θ should accommodate large effective population sizes and more incomplete lineage sorting. I conducted BEST analyses over a broad range of Θ priors, with mean values ranging from $\Theta = 0.00015$, $\Theta = 0.0015$, $\Theta = 0.015$, and $\Theta = 0.15$ (set using $\beta = 0.0003$, $\beta = 0.003$, $\beta = 0.03$, and $\beta = 0.3$, respectively, while holding α constant at $\alpha = 3$). The harmonic means of analyses using different theta priors were compared using Bayes factors calculated in TRACER v1.4 (Rambaut and Drummond 2007). Posterior probability values for species relationships were obtained by summarizing the posterior distribution of species trees (post burn-in) with a 50% majority-rule consensus tree.

Random Phylogeographic Sampling

To investigate the impacts of phylogeographic sampling on species tree inference, I conducted replicate BEST analyses using data sets that 1) varied the number of specimens sampled to represent species and 2) randomized the selection of specimens with respect to geography. These analyses focused on random sampling within the 4 mtDNA-based species within *S. "undulatus"* (i.e., *S. consobrinus*, *S. cowlesi*, *S. tristichus*, and *S. undulatus*; Table 1), each of which is represented by 4 specimens in the complete data matrix.

I generated 300 replicate data matrices in which either 1, 2, or 3 specimens (100 replicates each) were selected at random to represent each of the 4 species within *S. "undulatus"*. This subsampling approach is analogous to a jackknife test, with deletion factors of 75%, 50%, and 25% applied to each of 4 species represented by 4 specimens. All replicate data matrices were analyzed using BEST v2.2 with similar priors as the full data matrix (inverse gamma prior $= [3, 0.03]$), and analyses were run for 100 million generations (sampling every 100 000 steps). Convergence was assessed using burn-in plots of likelihood values and parameter estimates. A 50% majority-rule consensus tree was used to summarize the 100 species trees resulting from each sampling scheme (i.e., sampling either 1, 2, or 3 specimens per species). The bipartition frequencies equal the number of species trees retaining a particular clade across the 100 replicate analyses. High values indicate that a clade is not sensitive to the particular specimens sampled to represent a species, whereas low values suggest that a clade is sensitive to specimen selection. The posterior distributions of gene trees from these 300 replicate BEST analyses were used in subsequent species tree analyses using other species tree inference methods (see below).

Minimizing Deep Coalescences across Multiple Loci

I used Maddison's deep coalescence (MDC) measure (Maddison 1997) to reconstruct the species phylogeny that minimized the number of deep coalescences across the 8 nuclear loci with complete sampling. The deep coalescence measure is a count of the number of extra gene lineages (per branch) that result from fitting a gene tree into a species tree, assuming that discordance is the result of incomplete lineage sorting (Maddison 1997). These extra gene lineages are summed across the tree to quantify the discordance between a particular gene tree and species tree, and the method can be extended to search for the species tree that minimizes the number of deep coalescences across multiple loci.

The MDC measure was calculated using Mesquite v2.5 (Maddison W.P. and Maddison D.R. 2008). Instead of representing each independent locus by a single genealogy, gene tree uncertainty was accommodated into the species tree inference procedure by calculating the MDC measure using random resampling from the posterior probability distribution of gene trees obtained from the BEST analysis. This resampling procedure was repeated 500 times using the Mesquite software module AUGIST (Oliver 2008). When fitting gene trees into a species tree, gene trees were considered unrooted, and a heuristic search utilizing subtree pruning and regrafting was used to search for the species tree topology that minimized the number of deep coalescences across loci. All equally parsimonious species trees were retained for

each replicate. The inferred species trees were summarized using a 50% majority-rule consensus tree, and the bipartition frequencies for nodes were used as measures of species tree uncertainty.

The sensitivity of the MDC approach to phylogeographic sampling was evaluated by analyzing the 300 replicate data set results generated from the BEST analyses. Although accounting for gene tree uncertainty by resampling from the posterior distribution of gene trees is desirable (Oliver 2008), this approach was not feasible given the large number of replicate data sets. Therefore, the MDC analyses of the 300 replicate data sets utilized the consensus trees summarizing the posterior probability distributions of gene trees generated from the BEST analyses. The MDC multiple loci species tree was calculated for each replicate data set, and the 100 species trees resulting from each sampling scheme (i.e., sampling either 1, 2, or 3 specimens per species) were summarized using a 50% majority-rule consensus tree.

Bayesian Untangling of Concordance Knots

The 2-stage Bayesian model implemented in Bayesian untangling of concordance knots (BUCKy) v1.2b (Ané et al. 2007) was used to reconstruct the primary concordance tree (Baum 2007) for the *S. undulatus* group. This method estimates gene tree concordance by constructing the posterior distribution of gene-to-tree maps and calculating the proportion of loci that support a particular clade. The first step of the analysis involves estimating posterior distributions of genealogies for each locus, and the second step constructs the posterior distributions of gene-to-tree maps (Ané et al. 2007). Gene trees are not assumed to share a common history, and a Dirichlet process prior controlled by a single parameter (α) is incorporated into the method to model the degree of gene tree clustering (Ané et al. 2007). At the 2 extremes, setting $\alpha = 0$ forces all loci to share a single underlying species tree, whereas $\alpha = \infty$ is analogous to complete independence for each locus (Ané et al. 2007). Revised posterior distributions for each gene tree are obtained after considering the posterior distribution of gene-to-tree maps in conjunction with the Dirichlet process prior (Ané et al. 2007).

Primary concordance trees were estimated using BUCKy for each of the 100 replicate data sets constructed by randomly sampling 1 specimen per each species within *S. consobrinus*, *S. cowlesi*, *S. tristichus*, and *S. undulatus*. The replicate analyses that sampled 2 or 3 specimens per species were not analyzed with BUCKy because the nonexclusivity of species hindered the construction of species-level consensus trees. Each BUCKy analysis utilized MCMC sampling with 1 000 000 generations, 4 concurrent chains, and a 10% burn-in factor. Analyses were conducted using 2 different Dirichlet process priors, $\alpha = 0.1$ and $\alpha = 1.0$. These values were selected using a web-based utility (http://bigfork.botany.wisc.edu/concordance/) that graphs the prior distribution of the number of distinct trees based on the number of species ($n = 9$) and loci ($n = 8$) used in the analysis. These alpha values result in fundamentally different prior distributions; $\alpha = 0.1$ places high prior density for 1 shared tree, whereas $\alpha = 1.0$ accommodates a greater number of distinct trees (3 or 4) underlying the data. The primary concordance tree was recorded for each replicate analysis, and these 100 trees were summarized using a 50% majority-rule consensus tree.

RESULTS

Multilocus Nuclear Data

A total of 29 nuclear loci were sequenced for 21 specimens in the *S. undulatus* group, 8 of which amplified and sequenced for all 21 specimens (Table 2). The 29 nuclear loci range in size from 254 to 1523 bp (mean = 972 bp), contain 7–75 variable sites (mean = 38.5), and include 1–35 parsimony informative characters (mean = 15; Table 2). ML analyses of these nuclear loci result in a set of genealogies that support conflicting phylogenetic relationships among phylogeographic samples of *S. "undulatus"* (Supplementary Figure S1, available from http://www.sysbio.oxfordjournals.org/). Although no single gene tree is fully resolved, most of the loci with complete data provide strong support for placing *Sceloporus cautus*, *Sceloporus occidentalis*, *Sceloporus olivaceus*, and *Sceloporus virgatus* outside of a clade containing the remaining members of the *undulatus* group. Similar to the mtDNA genealogy, most of the nuclear gene trees place *Sceloporus woodi* with eastern populations of *S. "undulatus"*.

Concatenated Data and Multilocus Network

Concatenation of the 29 nuclear loci results in a data set containing 23 633 bp of aligned nucleotide positions, 1118 variable sites, and 437 parsimony informative sites (Table 2). The partitioned ML phylogenetic analysis of the concatenated data supports 16 of 18 nodes with bootstrap values $\geq$70% (Fig. 1b). The partitioned Bayesian phylogenetic analysis results in the same topology as the ML analysis and supports all 18 nodes with posterior probability values $\geq$0.99 (Fig. 1b). The concatenated nuclear data phylogeny (Fig. 1b) is distinctive from the mtDNA genealogy (Fig. 1a) in several respects: 1) *S. cautus* and *S. olivaceus* form a clade that is sister to all *S. "undulatus"* samples, 2) *S. woodi* is placed within *S. undulatus* as opposed to being sister to this species, and 3) *S. cowlesi* is sister to *S. tristichus* and *S. consobrinus*.

The multilocus network supports 4 phylogeographic groups within *S. "undulatus"* (Fig. 2). Four instances of discordance are supported with respect to the mtDNA genealogy, and each point of disagreement traces to an mtDNA clade boundary (Fig. 2): 1) the specimen of *S. cowlesi* from eastern AZ (cwAZ) is placed within *S. tristichus*, 2) the *S. tristichus* specimen from southern Colorado (tCO) is more closely related to *S. consobrinus*, 3) the specimen of *S. consobrinus* from Mississippi (cMS) is more closely related to populations of *S. undulatus* in

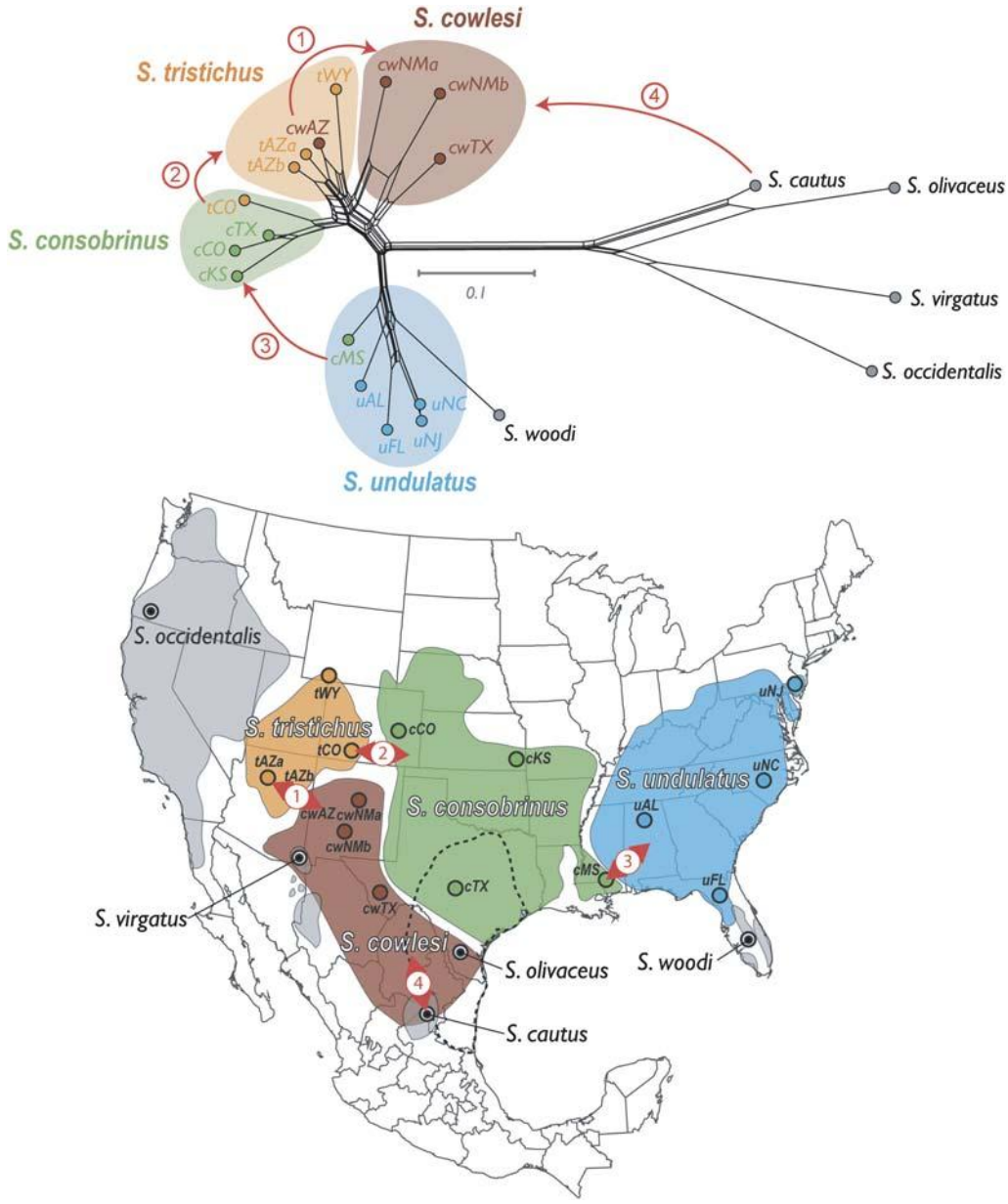


FIGURE 2. Phylogeographic comparison of nuclear and mtDNA clade boundaries. The network is constructed from the complete multilocus nuclear data set (8 loci). Phylogeographic samples are color-coded according to mtDNA clade membership (Fig. 1). Instances of discordance between the nuclear and mtDNA data are identified with red arrows.

the eastern United States, and 4) the nuclear data place *S. cautus* well outside of the genetic network of *S. undulatus*, whereas the mtDNA genealogy places *S. cautus* sister to *S. cowlesi* (Fig. 1a). These 4 instances of discordance are also evident in the concatenated nuclear data phylogeny (Fig. 1b).

Bayesian Estimation of Species Trees

The BEST analyses incorporating the smallest prior value for Θ (mean Θ = 0.00015) failed to converge after 200 million generations (Fig. 3), indicating that this prior value may be unrealistically small for these data. The analyses utilizing a mean Θ = 0.0015 appeared to get trapped on different local optima before eventually converging after 175 million generations (Fig. 3). The remaining analyses (mean Θ = 0.015 or Θ = 0.15) converged before 20 million generations, and the analysis with Θ = 0.015 ($\alpha = 3$, $\beta = 0.03$) had the highest log likelihood (Fig. 3). The species tree resulting from this analysis (mean Θ = 0.015) is fully resolved and similar to the concatenation tree at deeper phylogenetic levels (Fig. 4). The BEST tree provides strong support for a clade containing *S. cautus* and *S. olivaceus* (clade C, posterior probability = 1.0), which is placed sister to a clade containing *S. consobrinus*, *S. cowlesi*, *S. tristichus*, *S. woodi*,

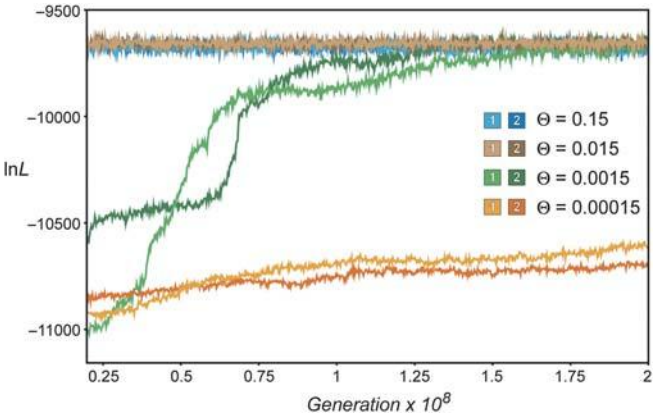


FIGURE 3. Likelihood burn-in plots for replicate BEST analyses using alternative theta priors and different starting seeds. The Bayes factor calculated from the analyses using $\Theta = 0.015$ (harmonic mean = −9714.338) and $\Theta = 0.15$ (harmonic mean = − 9724.891) is 10.553.

and *S. undulatus* (Fig. 4). Strong support is provided for a clade containing *S. cowlesi* and *S. tristichus* (clade G, posterior probability = 0.98; Fig. 4), a result that is at odds with the mtDNA genealogy (Fig. 1a). The sister taxon relationship between *S. woodi* and *S. undulatus* received the lowest support (clade E, posterior probability = 0.82; Fig. 4), but the concatenated data phylogeny provided strong support (ML bootstrap = 100; posterior probability = 1.0; Fig. 1b) for the placement of *S. woodi* within *S. undulatus*.

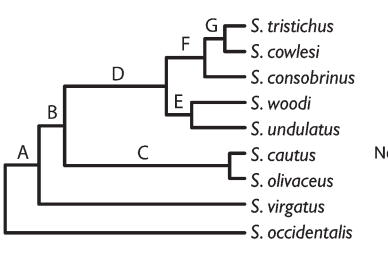
Random Phylogeographic Sampling

Varying the number of specimens sampled to represent species and randomizing the selection of specimens with respect to geography within *S. "undulatus"* have an impact on species tree estimation using BEST, MDC, and BUCKy (Fig. 4). The 50% majority-rule consensus tree topologies obtained after summarizing replicate BEST analyses of randomly sampled data are identical to the species tree estimated with complete data (with the exception of a polytomy between *S. cowlesi, S. consobrinus*, and *S. tristichus* that results when only one specimen is randomly sampled per species); however, there is ambiguity concerning the phylogenetic relationships among the 4 species that were subjected to random sampling (i.e., clades E, F, and G; Fig. 4). These clades are sensitive to phylogeographic sampling and were only supported in <50% to 84% of the replicate analyses (Fig. 4). The support for these clades varies drastically across the individual analyses (Table 3), with posterior probability values ranging from 0.05 to 1.0 for clades F and G and standard deviations exceeding 0.30 (Table 3). Some of these observed differences in the mean posterior probability values for clades are significantly different across various sampling intensities (Table 3). The deepest nodes in the phylogeny are recovered with a frequency of 100% (Fig. 4), indicating that these relationships are not sensitive to phylogeographic sampling artifacts within *S. "undulatus"*. Increasing the number of specimens sampled per species within *S. consobrinus, S. cowlesi*, and *S. tristichus* increased the chance of obtaining the complete data topology for clade G (Fig. 4), but the support for clades E and F do not show a clear relationship between sample size and the mean posterior probability.

Minimizing Deep Coalescences across Multiple Loci

The species tree that minimizes the number of deep coalescences across the 8 nuclear loci is identical to the BEST tree (Fig. 4). All clades received high support (bipartition frequency $\geq$ 98%; Fig. 4), which was estimated by summarizing 500 species tree inferences using random resampling from the posterior distribution of gene trees produced from the BEST analysis.

The species trees inferred using the MDC approach are sensitive to phylogeographic sampling at shallow and deep levels in the phylogeny, resulting in low bipartition frequencies for most clades (Fig. 4). Similar to the BEST analyses, the 50% majority-rule consensus tree topologies are identical to the species tree estimated with complete data, with the exception of the collapse of the *S. woodi* + *S. undulatus* clade in one instance (Fig. 4). Increasing the number of specimens sampled within species generally decreases the amount



		BEST				Minimize Deep Coal.					
		Random Sampling				Random Sampling				BUCKy	
		N = 1	*N* = 2	*N* = 3	*N* = 4	*N* = 1	*N* = 2	*N* = 3	*N* = 4	α = 0.1	α = 1
Node	A	100	100	100	1.0	-	-	-	-	-	-
	B	100	100	100	0.89	64	64	82	99	99	99
	C	100	100	100	1.0	94	97	100	100	100	100
	D	100	100	100	1.0	70	82	68	100	100	100
	E	81	73	76	0.82	69	63	<50	98	84	82
	F	72	63	84	0.98	71	85	94	99	80	80
	G	<50	60	76	0.98	57	96	100	100	<50	<50

FIGURE 4. Species tree estimate for the *Sceloporus undulatus* group based on the BEST analysis of the 8 nuclear loci with complete data (mean $\Theta = 0.015$). The bipartition frequencies are based on consensus trees from the 100 replicate analyses using BEST, BUCKy, and MDC, each of which incorporated random sampling from the full data matrix within *Sceloporus cowlesi, Sceloporus consobrinus, Sceloporus tristichus*, and *S. undulatus*. The BEST analysis of the full data matrix ($N = 4$) supports 5 nodes with posterior probability values $\geq$0.98. Nodal support for the MDC analysis of the full data matrix is based on 500 replicate analyses utilizing random resampling from the posterior distribution of gene trees from the BEST analysis.

TABLE 3. Posterior probability values (mean, minimum, maximum, variance, and standard deviation) for relationships within the *Sceloporus undulatus* species group calculated from 100 replicate BEST analyses utilizing random sampling within species (at varying sampling levels)

Clade	Sampling intensity	Mean	Minimum	Maximum	Variance	Standard deviation	$P < 0.05$
Sceloporus woodi + *S. undulatus* (Clade E)	$N = 1$	0.769	0.105	0.999	0.0906	0.3009	
	$N = 2$	0.678	0.100	0.999	0.0722	0.2688	1 versus 2
	$N = 3$	0.699	0.174	0.999	0.0638	0.2527	
Sceloporus tristichus + *Sceloporus cowlesi* + *Sceloporus consobrinus* (Clade F)	$N = 1$	0.777	0.345	0.996	0.0365	0.1911	1 versus 3
	$N = 2$	0.747	0.059	1.000	0.0919	0.3032	2 versus 3
	$N = 3$	0.840	0.120	1.000	0.0525	0.2290	
S. tristichus + *S. cowlesi* (Clade G)	$N = 1$	0.507	0.084	1.000	0.1190	0.3449	1 versus 2
	$N = 2$	0.689	0.055	1.000	0.0975	0.3123	1 versus 3
	$N = 3$	0.754	0.055	1.000	0.1112	0.3335	

Notes: Significant differences ($P < 0.05$) in the mean posterior probability values for different sampling intensities were calculated using Welch's *t*-test (for unequal variance).

of conflict among species trees (as reflected by higher bipartition frequencies), but this pattern is unclear for clade D and the opposite for clade E (Fig. 4).

Bayesian Untangling of Concordance Knots

The primary concordance trees estimated using BUCKy for each of the 100 replicate data sets constructed by randomly sampling 1 specimen per species within *S. consobrinus*, *S. cowlesi*, *S. tristichus*, and *S. undulatus* were similar to those constructed using BEST and MDC, with the only exception being the lack of support for a clade containing *S. cowlesi* and *S. tristichus* (clade G; Fig. 4). Analyses conducted using Dirichlet process priors of $\alpha = 0.1$ and $\alpha = 1.0$ produced similar results (Fig. 4). The low bipartition frequencies supporting the shallowest nodes in the species tree (clades E, F, and G; Fig. 4) indicate that the species topology is sensitive to the particular specimens selected to represent each species. Similar to the BEST analyses, the deeper nodes of the inferred species trees (B, C, and D; Fig. 4) are recovered across the replicate analyses.

DISCUSSION

Phylogeographic Discordance at mtDNA Clade Boundaries

This study provides a new perspective on the phylogenetic relationships of lizards in the *S. undulatus* species group. Whereas previous phylogenetic studies utilized data sets consisting of morphology (Wiens and Reeder 1997), allozymes (Miles et al. 2002), or mtDNA (Wiens and Reeder 1997; Leaché and Reeder 2002), the results presented here are based on phylogenetic analyses of multiple independent nuclear loci (up to 29 loci). The probability of estimating an accurate species tree typically increases with the number of loci, and several of the analyses presented here take an additional step toward improving phylogenetic accuracy by incorporating coalescent-based inference methods (Edwards et al. 2007).

A comparison of the mtDNA genealogy and multilocus nuclear data reveals an intriguing pattern of discordance at the phylogeographic level within *S. "undulatus"*. Although the nuclear and mtDNA data both recover 4 major phylogeographic groups, the composition of these groups differs with respect to the phylogenetic placement of 4 specimens (Fig. 2). Interestingly, each instance of conflict traces to a population sampled from the vicinity of a contact zone between adjacent phylogeographic groups (Fig. 2). This spatial pattern of discordance suggests that mitochondrial introgression operating across phylogeographic boundaries is responsible for at least some of the observed discordance between the mtDNA genealogy and multilocus nuclear data. The phylogenetic placement of 1 specimen of *S. cowlesi* (specimen cwAZ; Fig. 2) within *S. tristichus* (Figs. 1b and 2) by the multilocus nuclear data provides a compelling example of mtDNA introgression. This specimen is from a *S. cowlesi* × *S. tristichus* hybrid zone in Arizona, where mtDNA introgression is ongoing (Leaché and Cole 2007). From the perspective of the multilocus nuclear data, this specimen is more appropriately considered a representative of *S. tristichus* that is carrying a *S. cowlesi* mtDNA genome. The remaining cases of discordance each involve populations from the vicinity of mtDNA clade boundaries (Fig. 2), which suggests that mtDNA introgression may be pervasive across the current species boundaries established for this group of lizards. Although these new contact zones remain to be investigated more thoroughly, these data provide strong evidence that reproductive isolation is not complete across many of the currently established species boundaries within *S. "undulatus"*.

Phylogeographic Sampling and Species Tree Inference

Despite the increased computational challenges presented by analyzing data sets containing large numbers of taxa, numerous studies have demonstrated that dense taxon sampling can improve phylogenetic accuracy (reviewed by Heath et al. 2008). This tenet appears to hold true for species tree inference using the BEST method as well, although an increased number of loci is typically required to achieve an improvement in phylogenetic accuracy as species sampling increases (Liu et al. 2008). For the species tree analyses of the *S. undulatus* species group presented here, the number of species included in the analyses remained constant, although the number of individuals representing some species

TABLE 4. Support for alternative phylogenetic relationships within the *Sceloporus undulatus* species group inferred from mtDNA-introgressed (shown in bold) and nonintrogressed specimens

Hypothesis	Replicates including specimen	Trees supporting hypothesis	Mean posterior probability
Sceloporus cowlesi + Sceloporus tristichus			
cwAZ	**24**	**19 (79%)**	**0.96**
cwNMa	26	16 (62%)	0.67
cwNMb	27	4 (15%)	0.87
cwTX	23	2 (8%)	0.69
S. tristichus + Sceloporus consobrinus			
tCO	**26**	**18 (69%)**	**0.85**
tAZa	27	10 (37%)	0.67
tAZb	20	5 (25%)	0.73
tWY	27	0 (0%)	—
S. consobrinus + S. undulatus			
cMS	**19**	**19 (100%)**	**0.66**
cKS	29	0 (0%)	—
cTX	24	0 (0%)	—
cCO	28	0 (0%)	—

Notes: Values are calculated from the 100 replicate BEST analyses utilizing random sampling of 1 specimen per species within *S. consobrinus*, *S. cowlesi*, *S. tristichus*, and *S. undulatus*.

was varied (Fig. 4). Maddison and Knowles (2006) conducted species tree simulations using the minimize deep coalescence measure and demonstrated that sampling more individuals typically improves species tree accuracy, even under scenarios with considerable incomplete lineage sorting. Using empirical data for pocket gophers in the genus *Thomomys* to estimate species trees using the BEST method, Belfiore et al. (2008) found an increase in phylogenetic resolution when multiple individuals were sampled for each species. The empirical data presented here for *Sceloporus* lizards suggest that species tree inference methods are also sensitive to phylogeographic sampling and that species relationships can change depending upon which particular specimens are used to represent species (Fig. 4). This result does not change the current sampling paradigms in molecular systematics (i.e., more sampling is better), but it does reveal that even a limited number of samples may be sufficient to identify incongruent species boundaries using multilocus phylogeographic analysis.

An assumption that is shared among 2 of the species tree inference methods used here (BEST and MDC) is that species assignments are known a priori. This assumption is difficult to satisfy in study systems that span the fuzzy boundary between populations and species that is typically encountered in phylogeographic studies of recently diverged lineages (Wake 2006; Bond and Stockman 2008). Lineage discovery is often the goal of phylogeographic analysis, and assigning populations to species prior to the analysis may be difficult or impossible in some cases and could prevent the discovery of new lineages. In this study, the results from previous analyses of mtDNA data (Leaché and Reeder 2002; Leaché and Cole 2007) dictated how many distinct species to recognize and how groups of populations were assigned to those species. A comparison of the mtDNA genealogy to the concatenated data phylogeny (Fig. 1) and multilocus genetic network (Fig. 2) reveals several likely instances of mtDNA introgression, which would result in specimens being assigned to the wrong species and lead to inaccurate species tree inference.

If mtDNA introgression is the source of species tree discordance observed in this study, then species tree analyses including "introgressed specimens" (i.e., those crossing the mtDNA clade boundaries) should support species topologies that are distinct from analyses utilizing nonintrogressed specimens. The 100 replicate BEST analyses that randomly sampled 1 specimen per species within *S. consobrinus*, *S. cowlesi*, *S. tristichus*, and *S. undulatus* provide an opportunity to test this hypothesis directly because each species is represented by just 1 specimen. Indeed, the species tree analyses that included the introgressed specimens sampled from mtDNA clade boundaries tend to support topologies that received either low or no support compared with analyses that sampled nonintrogressed specimens (Table 4). For example, all analyses that sampled specimen cMS (a specimen of *S. consobrinus* from an mtDNA clade boundary in Mississippi) to represent *S. consobrinus* placed this species sister to *S. undulatus*, a relationship that was not supported when any other specimen of *S. consobrinus* was sampled (Table 4). Thus, sampling populations from the vicinity of phylogeographic clade borders can result in distinctly different species tree inferences. More specifically, inaccurate species assignments, in this case resulting from misclassifying species using mtDNA, can impact species tree inference.

How should species tree studies deal with the potential problems caused by inaccurate species assignments? Two alternatives for the phylogeographic data presented here include 1) focusing on sampling species from distant allopatric populations and exclude from the analysis any specimens sampled from species boundaries and 2) reassigning specimens with introgressed mtDNA copies to their prospective species. A reanalysis of the empirical data under these 2 alternatives (Fig. 5) results in species tree topologies similar to each other but distinct from analyses including misidentified

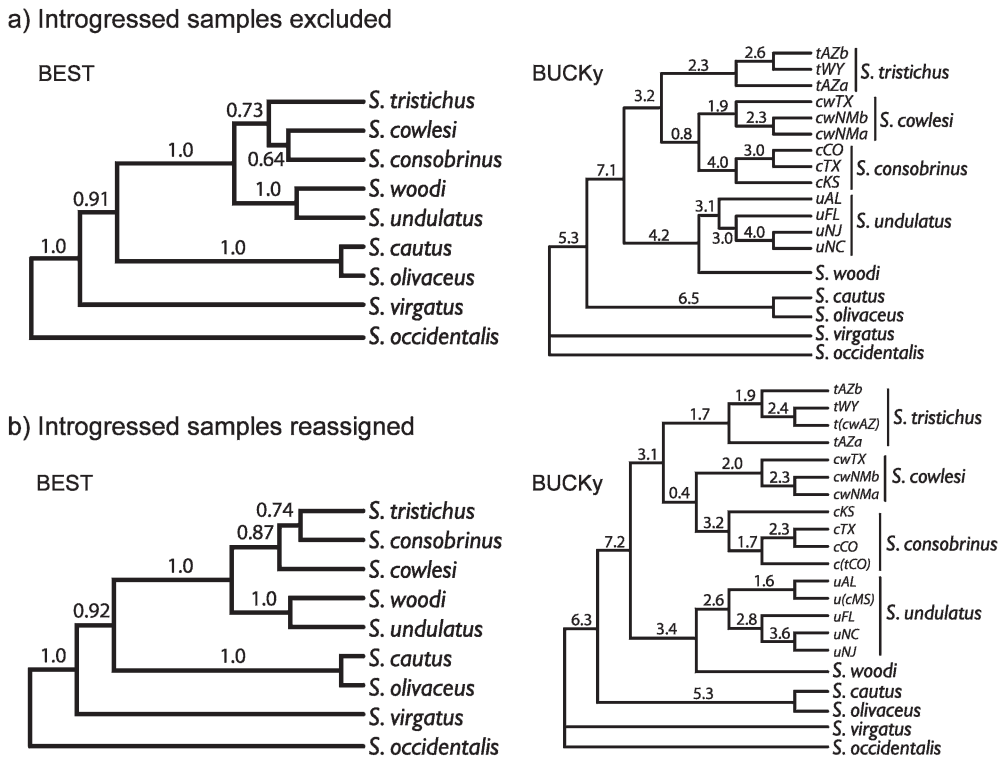


FIGURE 5. Species tree estimates for the *Sceloporus undulatus* species group based on analyses that a) exclud specimens with introgressed copies of mtDNA and b) reassign introgressed specimens to their prospective species. The species trees estimated using the MDC approach (not shown) are identical to the BEST topologies. Primary concordance factors are mapped onto nodes for the BUCKy trees.

specimens (due to mtDNA introgression; Fig. 4). This supports the hypothesis that inaccurate species assignments will mislead species tree inference. The new species tree topologies obtained after excluding or reassigning introgressed specimens no longer support a sister taxon relationship between *S. cowlesi* and *S. tristichus* with strong support (posterior probability = 0.99; Fig. 4) and instead place *S. cowlesi* (or *S. tristichus*) sister to *S. consobrinus* with weak support (posterior probability $<$ 0.75; Fig. 5). In addition, the new species trees (Fig. 5) provide strong support for a sister group relationship between *S. undulatus* and *S. woodi* (posterior probability = 1.0), which received low support in the previous analyses (Fig. 4). Thus, excluding or reassigning specimens both appear to be viable options for dealing with "rogue" specimens when using species tree inference methods that require the a priori designation of species assignments and may produce more accurate posterior probability estimates for inferred species relationships.

Posterior probability estimates of the effective population size parameter Θ obtained from BEST are also impacted by inaccurate species assignments (Table 5). Analyses that utilized species assignments based on mtDNA clade membership produced posterior estimates of Θ that are up to an order of magnitude higher compared with analyses that reassigned introgressed specimens to their prospective species (Table 5). This pattern suggests that posterior estimates of Θ may be

TABLE 5. Inaccurate species assignments result in higher posterior probability estimates of effective population size (Θ) in BEST analyses

Species/clade	mtDNA assignments	Introgressed samples reassigned	ΔΘ
Sceloporus consobrinus	0.0182	0.0089	−0.0093
Sceloporus tristichus	0.0184	0.0096	−0.0088
Sceloporus cowlesi	0.0164	0.0128	−0.0036
Sceloporus undulatus + *Sceloporus woodi*	0.0072	0.0057	−0.0015
S. undulatus	0.0106	0.0119	0.0013
Clade F	0.0116	0.0105	−0.0011
Clade D	0.0046	0.0040	−0.0006
Sceloporus cautus + *Sceloporus olivaceus*	0.0075	0.0071	−0.0004
Clade A	0.0037	0.0037	0.0000
Clade B	0.0061	0.0061	0.0000
S. tristichus + *S. cowlesi*	0.0077	—	—
S. tristichus + *S. consobrinus*	—	0.0097	—

Notes: Posterior estimates of Θ are estimated from analyses using the same prior mean (Θ = 0.015), which is specified using an inverse gamma probability distribution of (α = 3, β = 0.03). Clade names correspond to Figure 4.

useful for identifying species that contain misidentified specimens or those species that may be composed of multiple independent evolutionary lineages.

Species delimitation is often challenging in studies of adaptive radiations or recently diverged lineages, and this places investigations of these and other understudied clades at a disadvantage when using species tree inference methods that rely upon the accurate identification of species prior to the analysis. Systematists have a long history of using mtDNA for reconstructing phylogeny (reviewed by Brito and Edwards 2009), and utilizing species boundaries inferred from an mtDNA genealogy will represent the best starting point for many multilocus phylogenetic and phylogeographic studies. Although relying upon mtDNA for accurate species identification is the likely cause for the species tree discordance revealed in this study, establishing a preliminary set of species boundaries using multilocus network, concatenation, or even population-clustering approaches is also a reasonable starting point for species tree inference.

Supplementary Material

Supplementary material can be found at: http://www. sysbio.oxfordjournals.org/.

Funding

Funding was provided by the National Science Foundation (DEB-0508929 and DBI-0805455) and the UC Berkeley Chang-Lin Tien Graduate Fellowship in Biodiversity.

Acknowledgments

For comments and discussion, I thank T. Belfiore, D. Buckley, F. Burbrink, C. Cole, M. Hickerson, E. Jockusch, L. Liu, the McGuire Lab, J. MacKenzie, C. Moritz, G. Pauly, A. Pyron, B. Rannala, T. Reeder, E. Rosenblum, D. Vieites, R. Gillespie, Y. Wang, D. Wake, Z. Yang, and the Phylogenetics Discussion Group at University of California, Davis. I thank M. Hafner, J. Sullivan, and 2 anonymous reviewers for their insights and suggestions.

References

Ané 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.

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

Bond J.E., Stockman A.K. 2008. An integrative method for delimiting cohesion species: finding the population-species interface in a group of Californian trapdoor spiders with extreme genetic divergence and geographic structuring. Syst. Biol. 57(4):628–646.

Brito P.H., Edwards S.V. 2009. Multilocus phylogeography and phylogenetics using sequence-based markers. Genetica. 135(3):439–455.

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 multilocus sequence data. Syst. Biol. 57:719–731.

Bryant D., Moulton V. 2004. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. Mol. Biol. Evol. 21:255–265.

Carling M.D., Brumfield R.T. 2008. Integrating phylogenetic and population genetic analyses of multiple loci to test species divergence hypotheses in passerina buntings. Genetics. 178:363–377.

Degnan J.H., Rosenberg N.A. 2006. Discordance of species trees with their most likely gene trees. PLoS Genet. 2:762–768.

Dixon J.R., Ketchersid C.A., Lieb C.S. 1972. A new species of *Sceloporus* (*undulatus* group; Sauria, Iguanidae) from Mexico. Proc. Biol. Soc. Wash. 84:307–312.

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.

Edgar R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

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

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

Funk D.J., Omland K.E. 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annu. Rev. Ecol. Syst. 34:397–423.

Hasegawa M., Kishino H., Yano T. 1985. Dating the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22:160–174.

Heath T.A., Hedtke S.M., Hillis D.M. 2008. Taxon sampling and the accuracy of phylogenetic analyses. J. Syst. Evol. 46:239–257.

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

Joly S., Bruneau A. 2006. Incorporating allelic variation for reconstructing the evolutionary history of organisms from multiple genes: an example from *Rosa* in North America. Syst. Biol. 55:623–636.

Kluge A.G. 1989. A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). Syst. Zool. 38:7–25.

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

Leaché A.D., Cole C.J. 2007. Hybridization between multiple fence lizard lineages in an ecotone: locally discordant variation in mitochondrial DNA, chromosomes, and morphology. Mol. Ecol. 16:1035–1054.

Leaché A.D., McGuire J.A. 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.

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., Pearl D.K., Brumfield R.T., Edwards S.V. 2008. Estimating species trees using multiple-allele DNA sequence data. Evolution. 62:2080–2091.

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

Maddison W.P., Knowles L.L. 2006. Inferring phylogeny despite incomplete lineage sorting. Syst. Biol. 55:21–30.

Maddison W.P., Maddison D.R. 2008. Mesquite: a modular system for evolutionary analysis. Version 2.5. Available from: http://mesquiteproject.org.

McGuire J.A., Linkem C.W., Koo M.S., Hutchison D.W., Lappin A.K., Orange D.I., Lemos-Espinal J.A., Riddle B.R., Jaeger J.R. 2007. Mitochondrial introgression and incomplete lineage sorting through

space and time: phylogenetics of crotaphytid lizards. Evolution. 61:2879–2897.
McGuire G., Wright F. 2000. TOPAL 2.0: improved detection of mosaic sequences within multiple alignments. Bioinformatics. 16:130–134.
Miles D.B., Noecker R., Roosenburg W.M., White M.N. 2002. Genetic relationships among populations of *Sceloporus undulatus* fail to support present subspecific designations. Herpetologica. 58: 277–282.
Morando M., Avila L.J., Sites J.W. Jr. 2003. Sampling strategies for delimiting species: genes, individuals, and populations in the *Liolaemus elongatus-kriegi* complex (Squamata: Liolaemidae) in Andean-Patagonian South America. Syst. Biol. 52(2):159–185.
Nylander J.A.A. 2004. MrModeltest v2. Evolutionary Biology Centre, Uppsala University. Available from: http://www.abc.se/∼nylander/.
Nylander J.A.A., Wilgenbusch J.C., Warren D.L., Swofford D.L. 2008. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. Bioinformatics. 24(4):581–583.
Oliver J.C. 2008. AUGIST: inferring species trees while accommodating gene tree uncertainty. Bioinformatics. 24:2932–2933.
Pamilo P., Nei M. 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5:568–583.
Posada D., Crandall K.A. 2001. Intraspecific gene genealogies: trees grafting into networks. Trends Ecol. Evol. 16:37–45.
Rambaut A., Drummond A.J. 2007. Tracer v1.4. Available from: http://beast.bio.ed.ac.uk/Tracer.
Rannala B., Yang Z. 2008. Phylogenetic inference using whole genomes. Annu. Rev. Genomics Hum. Genet. 9:217–231.
Rokas A., Williams B.L., King N., Carroll S.B. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature. 425:798–804.
Ronquist F., Huelsenbeck J.P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19: 1572–1574.
Rosenblum E.B., Belfiore N.M., Moritz C. 2007. Anonymous nuclear markers for the eastern fence lizard, *Sceloporus undulatus*. Mol. Ecol. Notes. 7:113–116.
Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analysis with thousands of taxa and mixed models. Bioinformatics. 22:2688–2690.
Stephens M., Donnelly P. 2003. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. Am. J. Hum. Genet. 73:1162–1169.
Swofford D.L. 2001. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sunderland (MA): Sinauer Associates.
Wake D.B. 2006. Problems with species: patterns and processes of species formation in salamanders. Ann. Mo. Bot. Gard. 93(1): 8–23.
Wiens J.J., Reeder T.W. 1997. Phylogeny of the spiny lizards (*Sceloporus*) based on molecular and morphological evidence. Herpetol. Monogr. 11:1–101.

Received 8 January 2009; reviews returned 13 March 2009; accepted 17 August 2009
Associate Editor: Mark Hafner