Species Delimitation Using Bayes Factors: Simulations and Application to the Sceloporus scalaris Species Group (Squamata: Phrynosomatidae)

JARED A. GRUMMER1,2,*, ROBERT W. BRYSON JR.2, and TOD W. REEDER1

1Department of Biology, San Diego State University, San Diego, CA 92182-4614, USA and 2Department of Biology and Burke Museum of Natural History and Culture, University of Washington, Box 351800, Seattle, WA 98195-1800, USA

*Correspondence to be sent to: Department of Biology, University of Washington, Box 351800, Seattle, WA 98195-1800, USA; E-mail: grummer@uw.edu

Received 25 March 2013; reviews returned 7 June 2013; accepted 11 November 2013

Abstract.—Current molecular methods of species delimitation are limited by the types of species delimitation models and scenarios that can be tested. Bayes factors allow for more flexibility in testing non-nested species delimitation models and hypotheses of individual assignment to alternative lineages. Here, we examined the efficacy of Bayes factors in delimiting species through simulations and empirical data from the Sceloporus scalaris species group. Marginal-likelihood scores of competing species delimitation models, from which Bayes factor values were compared, were estimated with four different methods: harmonic mean estimation (HME), smoothed harmonic mean estimation (SHME), path-sampling/thermodynamic integration (PS), and stepping-stone (SS) analysis. We also performed model selection using a posterior simulation-based analog of the Akaike information criterion through Markov chain Monte Carlo analysis (AICM). Bayes factor species delimitation results from the empirical data were then compared with results from the reversible-jump MCMC (rjMCMC) coalescent-based species delimitation method Bayesian Phylogenetics and Phylogeography (BP&P). Simulation results show that HME and SHME perform poorly compared with PS and SS marginal-likelihood estimators when identifying the true species delimitation model. Furthermore, Bayes factor delimitation (BFD) of species showed improved performance when species limits are tested by reassigning individuals between species, as opposed to either lumping or splitting lineages.

In the empirical data, BFD through PS and SS analyses, as well as the rjMCMC method, each provide support for the recognition of three previously undescribed lineages. In both simulated and empirical data sets, harmonic and smoothed harmonic mean marginal-likelihood estimators provided much higher marginal-likelihood estimates than PS and SS estimators. The AICM displayed poor repeatability in both simulated and empirical data sets, and produced inconsistent model rankings across replicate runs with the empirical data. Our results suggest that species delimitation through the use of Bayes factors with marginal-likelihood estimates via PS or SS analyses provide a useful and complementary alternative to existing species delimitation methods.

Statistical species delimitation using multi-locus DNA sequence data and coalescent-based techniques is a burgeoning field (reviewed by Fujita et al. 2012). Current methods for statistically delimiting species utilize the Bayesian reversible-jump Markov chain Monte Carlo approach (rMCMC; Yang and Rannala 2010), probabilistic modeling or inference of the species history and model selection through likelihood ratio tests (LRTs) or AIC scores (Knowles and Carstens 2007; Ence and Carstens 2011), the minimizing deep coalescences (MDC) measure (Madison 1997), approximate Bayesian computation (Camargo et al. 2012a, 2012b), the general mixed Yule-coalescent model (Pons et al. 2006; Esselstyn et al. 2012; Reid and Carstens 2012), or heuristic approaches (O’Meara 2010). Although AIC and LRT model selection criteria have been implemented in species delimitation, using Bayes factors as a species delimitation model selection tool has not been thoroughly explored.

Bayes factors can be used as a model selection tool when comparing models or phylogenetic hypotheses. The Bayes factor in favor of Hypothesis 1 (H1) over Hypothesis 2 (H2), given the data (D), can be shown as

\[
B_{12} = \frac{\text{pr}(D | H_1)}{\text{pr}(D | H_2)}
\]

where \(\text{pr}(D | H_2)\) represents the marginal likelihood of Hypothesis \(x\) (Kass and Raftery 1995). The Bayes factor can be calculated between two models simply by multiplying twice the difference in the log of the likelihoods when there are no free parameters. However, when many parameters need to be estimated (as in the cases presented in this study), the Bayes factor still resembles the form of a likelihood ratio, but the parameter space must be integrated over with respect to the prior distribution of the parameters given the hypothesis to obtain the probability density of \(\text{pr}(D | H_2)\).

Many authors have advocated the use of Bayes factors as a model selection tool in a variety of contexts (Suchard et al. 2002; Sullivan and Joyce 2005; Ward 2008; Fan et al. 2011; Li and Drummond 2012). The Bayes factor as a model selection tool in phylogenetics is appealing because phylogenetic uncertainty is incorporated when comparing different models, the researcher can make inferences based on the entire set of candidate models (model averaging), and the models to be tested need not be nested (Huelsenbeck and Innemann 2002; Posada and Buckley 2004). Furthermore, other model selection criteria offer different complex models, using the estimated marginal-likelihood values from a Bayesian analysis automatically penalizes the complexity of the model (Spiegelhalter et al. 2002). In the case of
species delimitation, models contain varying numbers of lineages where each individual is assigned to a lineage before analysis. Maximizing the likelihood given the species delimitation model identifies the number of lineages and assignment of individuals to lineages that best explain the data.

To appropriately use the Bayes factor as a model selection tool, the marginal likelihood of competing models must be accurately estimated (Fan et al. 2011; Xie et al. 2011; Li and Drummond 2012). Recent work has shown that harmonic mean estimation (HME; Newton and Raftery 1994) often overestimates the true marginal likelihood (Xie et al. 2011). A more accurate approach to likelihood estimation, smoothed harmonic mean estimation (sHME), was also discussed by Newton and Raftery (1994). This method uses an importance sampling estimator of the marginal likelihood where the importance sampling function uses both prior and posterior samples. An alternative method to rank models was described by Raftery et al. (2007) as the AICM (AIC–Monte Carlo), and has been shown to outperform the HME method in model selection using Bayes factors (Baele et al. 2012). Path-sampling (PS), another marginal-likelihood estimator, explores an almost continuous progression of distributions along a path from the posterior to the prior when calculating the marginal likelihood (Lartillot and Philippe 2006). A similar method, stepping-stone sampling (SS; Xie et al. 2011), bridges the posterior and prior distributions when estimating the marginal likelihood through importance sampling. Baele et al. (2012) concluded that PS and SS methods significantly outperform the HME method when comparing demographic growth and molecular clock models, whereas the performance of the Akaike information criterion through Markov chain Monte Carlo analysis (AICM) method lies between PS/SS and HME methods.

In this study, we conduct analyses on simulated data sets to assess the performance of these marginal-likelihood estimators in a multi-species coalescent-based framework. We test the performance of our Bayes factor delimitation (BFD) approach on three general species delimitation scenarios/models: lumping lineages, splitting lineages, and individual reassignment. Alongside marginal-likelihood estimations, we performed species delimitation model selection using the AICM. Subsequently, we used BFD to delimit species of phrynosomatid lizards in the Sceloporus scalaris species group. We then compared these results to a popular method of species delimitation, Bayesian Phylogenetics and Phylogeography (BP&P; Yang and Rannala 2010). Although direct comparison of these two methods is difficult because they cannot evaluate the same types of models (e.g., non-nested), BFD and Bayesian species delimitation provide complementary means to species delimitation. BP&P provides species delimitation probabilities, whereas BFD provides a ranking of the species delimitation models that are tested.

Sceloporus scalaris Species Group

Nine extant taxa are currently recognized in the S. scalaris species group (hereafter referred to as the “scalaris group”). Although some scalaris group taxa occur sympatrically, most members of the group are allopatric and inhabit disjunct mountain ranges throughout México (Fig. 1). Previous higher level phylogenetic studies of Sceloporus and Phrynosomatidae have included only one to a few members of the scalaris group (Reeder and Wiens 1996; Wiens and Reeder 1997; Flores-Villalva et al. 2000; Leaché 2010; Wiens et al. 2010), whereas three molecular systematic studies conducted in the 1990s focused solely on multiple taxa within the scalaris group. Results from these studies using allozymes (Mink and Sites 1996) and mitochondrial DNA (mtDNA; Benabib et al. 1997; Ienabib et al. 1997; Creer et al. 1997) reveal an early split in the scalaris group leading to major “northern” and “southern” clades. In these studies, the northern clade contains S. chaneyi, S. scalaris, and S. goldmani, and the southern clade contains S. slevini and S. bicanthalis. The most recent molecular phylogenetic study of the scalaris group (Bryson et al. 2011) also recovered northern and southern clades. However, the more extensive geographic sampling in this mtDNA-based study resulted in greater phylogenetic resolution and different phylogenetic placements of various taxa as compared with the previous mtDNA and allozyme studies on the group.

The number of species recognized in the scalaris group is far from certain. Subsequent to the early molecular systematic studies on the scalaris group, several taxonomic changes have occurred. One new taxon has been described (S. scalaris bouncerum: Smith et al. 1997) and two subspecies were elevated to species (S. samonotomani: Smith et al. 1996; Watkins-Colwell et al. 1998; S. slevini: Smith et al. 1996; Watkins-Colwell et al. 2003). Furthermore, phylogeographic analysis has revealed several “deeply” divergent mtDNA clades within the scalaris group, some of which appear to represent cryptic and undescribed lineages (Bryson et al. 2011). Here, we apply Bayes factor species delimitation to a multi-locus data set to determine the number of evolutionary independent lineages in the scalaris group.

MATERIALS AND METHODS

BFD of Species

Coupled with Bayesian species tree inference, we implement a Bayesian hypothesis-testing approach similar to that advocated by Knowles and Carstens (2007) and Carstens and Dewey (2010) to statistically test alternate hypotheses of species limits within the scalaris group (i.e., species validation). Bayes factors proffer themselves as versatile and robust in species delimitation because of many of the aforementioned advantages over other Bayesian model selection tools. For instance, individuals can be reassigned and tested as belonging to different lineages. Furthermore, species delimitation
models can be ranked against one another, thus enabling the researcher to make quantitative decisions about species limits.

Our BFD method for testing species limits can be broken down into four steps: (i) assignment of individuals to lineages/populations (species discovery and model generation); (ii) creating alternative groupings by lumping lineages, splitting lineages, or reassigning individuals to alternate lineages in competing models; (iii) generating a marginal-likelihood estimate and species tree for each model; and (iv) calculating Bayes factors to identify the species delimitation model that best explains the data (species validation). After generating species delimitation models (see below), the Bayes factors (2\ln BF) for competing models were evaluated following the recommendations of Kass and Raftery (1995), thus providing assessments of the strength of support for a particular species delimitation hypothesis given multi-locus nDNA data. A 2\ln BF = 0–2 means “not worth more than a bare mention”, 2\ln BF = 2–6 means “positive” support, 2\ln BF = 6–10 provides “strong” support, and 2\ln BF >10 means “decisive” support in distinguishing between competing species delimitation hypotheses. We recognize a lineage as distinct when the Bayes factor support value (2\ln BF) for the leading phylogenetic hypothesis is >10.

We explored four methods of marginal-likelihood estimation in *BEAST to be used in our BFD analysis: harmonic mean estimation (HME; Newton and Raftery 1994), smoothed harmonic mean estimation (sHME), path-sampling/thermodynamic integration (PS; Lartillot and Philippe 2006), and stepping-stone analysis (SS; Xie et al. 2011). Additionally, we used a posterior simulation-based analog of the (AICM; Raftery et al. 2007) to rank species delimitation models.

Accuracy of BFD of Species under the Multi-species Coalescent

The code to implement these marginal-likelihood estimators was originally developed for the program BEAST (Drummond and Rambaut 2007). We therefore performed simulations to assess the accuracy of the marginal-likelihood estimators within a multi-species coalescent framework. The species tree and DNA data sets used in our simulations were designed to emulate the empirical data for the *scalaris group (see Results section). The simulated species tree had 10 ingroup species, each with 4 individuals/sequences, and 1 outgroup species with a single individual. θ and τ values (ancestral population size and divergence times, respectively) used in the simulated species tree were estimated from the empirical data by running MCcoal in the program BP&P (Rannala and Yang 2003; Yang and Rannala 2010). We simulated 600 gene trees within the fixed species tree in the program MCcoal (Rannala and Yang 2003; Yang and Rannala 2010) that were then broken up into 100 sets of 6 loci. A total of 750 bp of

FIGURE 1. Phylogenetic relationships inferred for the *S. scalaris* species group based on maximum-likelihood analysis of the concatenated mtDNA and previously published morphological phylogeny (Wiens and Reeder 1997). The range map indicates sampling localities (Supplementary Table S1) and geographic ranges. The branch leading to *S. goldmani* in the morphological tree is represented in dark gray to reflect that the species is likely extinct and is not sampled for DNA. Black dots on nodes represent bootstrap proportions ≥85, whereas gray dots represent bootstrap proportions between 70 and 85.
Each *BEAST analysis was run for 10^8 generations, under each scenario in *BEAST for a total of 400 analyses. Portions of this study.

Individuals to species; this is the same for the empirical files (one for each scenario) to assign false alignments in the BEAST input files must be generated, each with a different "mapping" file (= "species traits" file in BEAUTi). It should go without stating that this increases analysis time. We then generated four species delimitation scenarios to investigate in *BEAST (Heled and Drummond 2010). In the "True" scenario, individuals (alleles) were assigned to species as the data were simulated (Fig. 2). In the "Lump" scenario, two ingroup species were collapsed into a single species, decreasing the ingroup number of species from 10 to 9. The "Split" scenario had two ingroup species each split into two species, raising the total number of ingroup species from 10 to 12. In the final scenario, "Reassign", a total of four individuals were reassigning individuals, two .xml files through BEAUti and running a separate analysis. For instance, when reassigning individuals, two .xml files were each with a different "mapping" file (= "species traits" file in BEAUTi). It should go without stating that this increases analysis time. In our simulations, we used four different "mapping" files (one for each scenario) to assign individuals to species; this is the same for the empirical portion of this study. Each of the 100 simulated data sets was then analysed under each scenario in *BEAST for a total of 400 analyses. Each *BEAST analysis was run for 10^8 generations, sampling trees every 20,000 steps, and discarding the first 2 x 10^7 trees (20%) as burn-in. HME, sHME, PS, and SS marginal-likelihood estimators were then performed on each analysis, in addition to AICM model selection.

Sequence data were then evolved along each gene tree in the program Seq-Gen (Rambaut and Grassly 1997) under the HKY model of sequence evolution with no site rate heterogeneity, thus generating 100 data sets each composed of 6 loci and 750 bp of sequence data per locus.

We then generated four species delimitation scenarios to investigate in *BEAST (Heled and Drummond 2010). In the "True" scenario, individuals (alleles) were assigned to species as the data were simulated (Fig. 2). In the "Lump" scenario, two ingroup species were collapsed into a single species, decreasing the ingroup number of species from 10 to 9. The "Split" scenario had two ingroup species each split into two species, raising the total number of ingroup species from 10 to 12. In the final scenario, "Reassign", a total of four individuals were reassigning individuals, two .xml files through BEAUti and running a separate analysis. For instance, when reassigning individuals, two .xml files were each with a different "mapping" file (= "species traits" file in BEAUTi). It should go without stating that this increases analysis time. In our simulations, we used four different "mapping" files (one for each scenario) to assign individuals to species; this is the same for the empirical portion of this study. Each of the 100 simulated data sets was then analysed under each scenario in *BEAST for a total of 400 analyses. Each *BEAST analysis was run for 10^8 generations, sampling trees every 20,000 steps, and discarding the first 2 x 10^7 trees (20%) as burn-in. HME, sHME, PS, and SS marginal-likelihood estimators were then performed on each analysis, in addition to AICM model selection.

Sequence data were then evolved along each gene tree in the program Seq-Gen (Rambaut and Grassly 1997) under the HKY model of sequence evolution with no site rate heterogeneity, thus generating 100 data sets each composed of 6 loci and 750 bp of sequence data per locus. We then generated four species delimitation scenarios to investigate in *BEAST (Heled and Drummond 2010). In the "True" scenario, individuals (alleles) were assigned to species as the data were simulated (Fig. 2). In the "Lump" scenario, two ingroup species were collapsed into a single species, decreasing the ingroup number of species from 10 to 9. The "Split" scenario had two ingroup species each split into two species, raising the total number of ingroup species from 10 to 12. In the final scenario, "Reassign", a total of four individuals were reassigning individuals, two .xml files through BEAUti and running a separate analysis. For instance, when reassigning individuals, two .xml files were each with a different "mapping" file (= "species traits" file in BEAUTi). It should go without stating that this increases analysis time. In our simulations, we used four different "mapping" files (one for each scenario) to assign individuals to species; this is the same for the empirical portion of this study. Each of the 100 simulated data sets was then analysed under each scenario in *BEAST for a total of 400 analyses. Each *BEAST analysis was run for 10^8 generations, sampling trees every 20,000 steps, and discarding the first 2 x 10^7 trees (20%) as burn-in. HME, sHME, PS, and SS marginal-likelihood estimators were then performed on each analysis, in addition to AICM model selection.

sequence data were then evolved along each gene tree in the program Seq-Gen (Rambaut and Grassly 1997) under the HKY model of sequence evolution with no site rate heterogeneity, thus generating 100 data sets each composed of 6 loci and 750 bp of sequence data per locus. We then generated four species delimitation scenarios to investigate in *BEAST (Heled and Drummond 2010). In the "True" scenario, individuals (alleles) were assigned to species as the data were simulated (Fig. 2). In the "Lump" scenario, two ingroup species were collapsed into a single species, decreasing the ingroup number of species from 10 to 9. The "Split" scenario had two ingroup species each split into two species, raising the total number of ingroup species from 10 to 12. In the final scenario, "Reassign", a total of four individuals were reassigning individuals, two .xml files through BEAUti and running a separate analysis. For instance, when reassigning individuals, two .xml files were each with a different "mapping" file (= "species traits" file in BEAUTi). It should go without stating that this increases analysis time. In our simulations, we used four different "mapping" files (one for each scenario) to assign individuals to species; this is the same for the empirical portion of this study. Each of the 100 simulated data sets was then analysed under each scenario in *BEAST for a total of 400 analyses. Each *BEAST analysis was run for 10^8 generations, sampling trees every 20,000 steps, and discarding the first 2 x 10^7 trees (20%) as burn-in. HME, sHME, PS, and SS marginal-likelihood estimators were then performed on each analysis, in addition to AICM model selection.

Molecular Data Collection

Genomic DNA was extracted from tissues using the NucleoSpin Tissue kit (Macherey-Nagel, Bethlehem, PA) and PCR was used to amplify six nuclear gene regions (five protein-coding and one intron) and two mtDNA gene regions (Supplementary Table S2). Standard PCR protocols were used to amplify mtDNA, whereas a "touch-down" protocol was used to amplify the nDNA regions (94°C for 1:00, [0:30 at 94°C, 0:30 at 61°C, 1:30 at 68°C] × 5 cycles, [0:30 at 94°C, 0:30 at 57°C, 1:30 at 68°C] × 5 cycles, [0:30 at 94°C, 0:30 at 50°C, 1:30 at 68°C] for 25 cycles). PCR templates were purified and subjected to automated DNA sequencing with an ABI 3730 capillary sequencer. Sequences were edited using Sequencher v4.7 (Gene Codes Corp., Ann Arbor, MI). Heterozygous individuals for a given nuclear gene region were identified as having two alleles of the same length (i.e., alleles containing nucleotide substitutions) or having alleles differing in length because of insertions/deletions (indels; such alleles could also have differences because of nucleotide substitutions). For heterozygous individuals with alleles of the same length, gene sequences were phased using PHASE (Stephens et al. 2001) where heterozygous positions were considered phased if they were inferred with >90% probability; if heterozygous positions were resolved with <90% probability, the appropriate IUPAC ambiguity code was used in the phylogenetic analyses. Length heterozygous genotypes were resolved using the program CodonCode Aligner v3.7 (CodonCode Corp., Dedham, MA).

Multiple sequence alignments were performed using Opal (Wheeler and Kickeciglu 2007) as implemented in Mesquite (Maddison and Maddison 2010). Alignments were performed with the following default gap costs: Open = 260, Terminal Open = 100, Extension = 69,
Terminal Extension = 66. A single ambiguously aligned region of 11 bp (as determined by changing alignment parameters that resulted in different alignments for this region) was excluded from the intron region NOS1. Indels were treated as missing data in all phylogenetic analyses. Testing for intragenic recombination was accomplished using the difference in sum-of-squares test (McGuire and Wright 2000) as implemented in TOPALi (Milne et al. 2009) with a 10 bp increment and window size of 100 for 500 parametric bootstraps. To test for selection in the protein-coding nuclear genes, we used a codon-based Z-test of positive selection averaged across all sequence pairs. We computed the probability of rejecting the null hypothesis of strict-neutrality ($dN = dS$) in favor of positive selection using MEGA5 (Tamura et al. 2011). Analyses were conducted using the Nei–Gojobori method (Nei and Gojobori 1986) and the variance of the difference was computed using 1000 bootstrap replicates.

**Generation of Species Delimitation Models**

To evaluate or test alternative species delimitation models and explore a priori assignment of individuals to lineages (species discovery), we generated competing species delimitation models based on three tree-based sources of phylogenetic information, and a network approach, in conjunction with current taxonomy: (i) concatenated nDNA tree, (ii) multi-locus nDNA network, (iii) mtDNA gene tree, and (iv) phylogenetic relationships based on morphological characters (Figs. 1 and 3; Supplementary Table S3; Wiens and Reeder 1997). Based on the results of these analyses (see Results section), we generated eight species delimitation models to test (Fig. 3).

To infer our concatenated nDNA phylogeny, we selected one allele at each locus for heterozygous individuals (Carstens and Dewey 2010; McCormack et al. 2011). Exons were partitioned by codon position, whereas non-coding regions were each given their own partition. Each gene region was analysed under maximum likelihood using RAxML v7.2.7 (Stamatakis 2006) with the GTR + Γ nucleotide substitution model for each partition for 500 non-parametric bootstrap iterations. We also analysed the concatenated nDNA data set in MrBayes v3.1.2 for a run length of $5 \times 10^7$ generations, where each of the 16 partitions (nDNA only) was given its own nucleotide substitution model determined by jModelTest (Posada 2008). We constructed the multi-locus nDNA network in POFA (Joly and Bruneau 2006), where we used PAUP* (Swofford 2002) to create an uncorrected $P$-distance matrix for each locus. POFA subsequently converted these distance matrices of alleles into a single
were run for 108 generations with the first 20 million same (informative) priors across all analyses. Analyses value of 0.015, shape parameter of 3 and scale of 0.3. model, while the species population mean hyperprior and constant root was used for the population size used for the species tree prior, and the piecewise linear were estimated relative to this gene. A Yule process was fixed for the gene BACH1 and rates for the other loci were estimated by *BEAST. Protein-coding genes were partitioned by codon position, and the intron was given a single partition. All analyses in *BEAST were performed including limited gene flow (Zhang et al. 2011; Camargo 2007), so the simpler HKY model of DNA substitution (evaluated in Tracer v1.5; Rambaut and Drummond 2010). Analyses with the nucleotide tree inference program *BEAST v1.7.4 (Heled and Rambaut 2010) where we used the median node heights to determine if the analyses had converged onto similar tree posterior distributions from replicate analyses were performed on each of the four replicate runs. We combined the log files from separate runs with the program LogCombiner v1.7.4 (Rambaut and Drummond 2007) and performed marginal-likelihood calculations on the combined log files as well. It is not possible to compare the results of our BFD method to a commonly used method of species delimitation, we analysed each species delimitation model in the program BP&P (Yang and Rannala 2010). This method utilizes a rjMCMC algorithm to assign probabilities to speciation events while accounting for incomplete lineage sorting and uncertainty in gene tree reconstruction, and assumes no admixture following speciation. Results from simulation studies show that this method produces robust results under a variety of demographic scenarios, including limited gene flow (Zhang et al. 2011; Camargo et al. 2012b). We tested eight species delimitation models with this method, where four of the models examined subsets of three models that served as guide trees (Fig. 3). In all BP&P analyses, Pp ≥ 0.95 were interpreted as strong support in favor of a speciation event (Leaché and Fujita 2010). Running the complete data set resulted in poor mixing of the rjMCMC chain. We therefore broke up the larger guide trees into subtrees to test them separately, and then combined information from the separate analyses into a single summary figure (Supplementary Fig. S1). We believe this should not affect our results because the base of each subclade was supported with a Pp of 1.0 in BP&P, and support values necessarily only increase toward the base of the guide tree in BP&P.

**Species Tree and Marginal-Likelihood Estimation**

We estimated species trees for each species delimitation model using the coalescent-based species tree inference program *BEAST v1.7.4 (Heled and Drummond 2010). Analyses with the nucleotide substitution models determined by jModelTest resulted in over-parameterization and subsequent low effective sample size (ESS) values for the posterior, prior, likelihood, and partitions with little information (evaluated in Tracer v1.5; Rambaut and Drummond 2007), so the simpler HKY model of DNA substitution was used (with invariant sites or Γ as determined by jModelTest) for all partitions, and base frequencies were estimated by *BEAST. Protein-coding genes were partitioned by codon position, and the intron was given a single partition. All analyses in *BEAST were performed under an uncorrelated lognormal relaxed molecular clock for each locus where the mean clock rate of 1.0 was fixed for the gene BACH1 and rates for the other loci were estimated relative to this gene. A Yule process was used for the species tree prior, and the piecewise linear and constant root was used for the population size model, while the species population mean hyperprior was given an inverse gamma distribution with an initial value of 0.015, shape parameter of 3 and scale of 0.3. All other parameters not mentioned were given the same (informative) priors across all analyses. Analyses were run for 108 generations with the first 20 million generations (20%) discarded as burn-in, saving every 20 000th tree.

Four replicates using different random starting trees were performed on each species delimitation model, and the post burn-in portion of the posterior distribution from each replicate run was combined using LogCombiner (Drummond and Rambaut 2007). In addition to assessing the ESS values in Tracer, the species tree posterior distributions from replicate analyses were also examined in AWTY (Wilgenbusch et al. 2004) to determine if the analyses had converged onto similar posterior distributions. Species trees were produced from the combined files composed of the four replicate runs using TreeAnnotator (Drummond and Rambaut 2007) where we used the median node heights to construct the maximum clade credibility tree with a minimum clade credibility value of 0.5. Following *BEAST analyses, AICM, HME, and sHME estimations were performed on the collected samples following a 20% burn-in period. PS and SS analyses were each run for a chain length of 108 generations for 100 path steps (totaling 109 generations). The four marginal-likelihood estimators and AICM analysis were performed on each of the four replicate runs. We combined the log files from separate runs with the program LogCombiner v1.7.4 (Rambaut and Drummond 2007) and performed marginal-likelihood calculations on the combined log files as well. It is not possible to determine if the independent rjMCMC chains were converging onto similar posterior distributions, and
nodal probabilities of the four independent runs were averaged for the reported guide trees.

RESULTS

BFD of Species: Simulations

Computational time for running the MCMC chain on each replicate in *BEAST ranged from 6–10 h, whereas PS and SS marginal-likelihood estimators took an additional 8–10 h on a desktop iMac for each replicate (8 cores, 2.7 GHz/processor). Our simulations demonstrate that the marginal-likelihood estimators employed in BEAST by Baele et al. (2012) work within a multi-species coalescent framework in *BEAST, particularly the PS and SS estimators. This conclusion is confirmed by the fact that the “True” scenario was favored over the alternative scenario (e.g., “Lump, Split, Reassign”) by strong Bayes factor support (2\(\ln B_f > 10\)) using the more accurate PS and SS estimators. The results using the HME and shME estimators were more equivocal in terms of recovering the True scenario under which the data were simulated (Fig. 4).

HME and shME estimators provided much less discrimination than PS and SS in identifying the true species delimitation scenario (Fig. 4; Table 1). However, the variance in marginal-likelihood estimates was much greater for PS and SS estimators than HME and shME estimators. When considering accuracy by species delimitation scenario, defined as a 2\(\ln B_f \) value > 0, all marginal-likelihood estimators were most accurate under the reassignment scenario, with PS and SS estimators obtaining 100% and 99% accuracy, respectively (Table 1). All four estimators performed moderately well when identifying “Lump” (≥45%) and “Split” (≥57%) scenarios (Fig. 4; Table 1). When considering accuracy at a 2\(\ln B_f \) value > 0, the HME and shME estimators approached or exceeded that of the PS and SS estimators in “Split” and “Reassign” scenarios. However, when considering accuracy at a 2\(\ln B_f \) value > 10, the HME and shME were much less accurate than PS and SS estimators (Table 1).

The AICM model selection tool performed well when investigating the “Reassign” scenario (92% accuracy; Table 1). However, model selection performance was weak (41%) and moderate (68%) under “Lump” and “Split” scenarios, respectively.

scalars Group Gene Sequence Data

DNA sequence data were obtained from six nuclear loci (five protein-coding and one intron) ranging in length from 521 to 1254 bp, yielding a total of 4635 bp with the number of parsimony-informative sites within the *scalaris* group ranging from 51 to 92 per locus (437 in total; Table 2; Dryad doi:10.5061/dryad.c7s77; GenBank accession nos. KF436951-KF437289). The mtDNA gene

| Table 1. Species delimitation results based on the simulated data sets |

<table>
<thead>
<tr>
<th>Method</th>
<th>Species delimitation scenario</th>
<th>True</th>
<th>Lump</th>
<th>Split</th>
<th>Reassign</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>AICM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>21 919.6</td>
<td>N/A</td>
<td>21 971.0</td>
<td>41 411 21 933.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HME</td>
<td>-10 864.6</td>
<td>N/A</td>
<td>-10 864.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>shME</td>
<td>-10 851.1</td>
<td>N/A</td>
<td>-10 851.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PS</td>
<td>-11 142.0</td>
<td>N/A</td>
<td>-11 170.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SS</td>
<td>-11 123.3</td>
<td>N/A</td>
<td>-11 157.8</td>
</tr>
</tbody>
</table>

For each scenario, the marginal-likelihood estimate (or model score for the AICM) is given in the first column. The second column provides the number of times out of 100 that the correct scenario is favored (i.e., BF-value > 0), followed by how many times out of 100 that the correct scenario is favored with a BF-value ≥ 10. Species delimitation scenarios are shown in Figure 2.
TABLE 2. Locus-specific information for the nDNA and mtDNA used in this study

<table>
<thead>
<tr>
<th>Locus</th>
<th>Protein coding</th>
<th>Length (bp)</th>
<th>Variable sites</th>
<th>Percentage of variable positions</th>
<th>Parsimony-informative sites</th>
<th>Alleles</th>
<th>DNA substitution model</th>
</tr>
</thead>
<tbody>
<tr>
<td>BACH1</td>
<td>Yes</td>
<td>1254</td>
<td>100</td>
<td>8.0</td>
<td>92</td>
<td>42</td>
<td>GTR+I</td>
</tr>
<tr>
<td>EXPH5</td>
<td>Yes</td>
<td>977</td>
<td>100</td>
<td>10.2</td>
<td>92</td>
<td>39</td>
<td>HKY+I+Gamma1</td>
</tr>
<tr>
<td>KIF24</td>
<td>Yes</td>
<td>521</td>
<td>75</td>
<td>14.4</td>
<td>70</td>
<td>37</td>
<td>GTR+I</td>
</tr>
<tr>
<td>NKTR</td>
<td>Yes</td>
<td>657</td>
<td>61</td>
<td>9.9</td>
<td>56</td>
<td>37</td>
<td>GTR+I</td>
</tr>
<tr>
<td>NOS1</td>
<td>No</td>
<td>608</td>
<td>78</td>
<td>12.8</td>
<td>76</td>
<td>36</td>
<td>GTR+I+Gamma1</td>
</tr>
<tr>
<td>R35</td>
<td>Yes</td>
<td>658</td>
<td>64</td>
<td>9.7</td>
<td>51</td>
<td>39</td>
<td>GTR+I</td>
</tr>
<tr>
<td>mtDNA</td>
<td>Yes*</td>
<td>1698</td>
<td>564</td>
<td>34.4</td>
<td>435</td>
<td>—</td>
<td>GTR+I+Gamma1</td>
</tr>
<tr>
<td>Total</td>
<td>—</td>
<td>6333</td>
<td>1062</td>
<td>—</td>
<td>892</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Information from outgroup sequences not included.

*For the mtDNA data, a single partition (ND4 tRNAs) was not protein coding and was assigned the HKY+I+Gamma1 DNA substitution model.

regions totaled 1698 bp and yielded 584 parsimony-informative sites (Table 2). No evidence of intragenic recombination was detected for any locus. The P-values from our Z-tests for selection were all non-significant (P > 0.5), indicating that the protein-coding nuclear genes were not under positive selection. The single gene and concatenated analyses provided strong support for the monophyly of the scalaris group, but low support for interspecific relationships among outgroup taxa (Supplementary Figs. S3–S5).

Species Delimitation Model Groupings

The composition of major clades is in disagreement between the trees inferred from the mtDNA, concatenated nDNA, and morphological data sets (Figs. 1 and 5). Specifically, in terms of the molecular data, sample no. 12 (Figs. 1 and 5) has the nDNA of the northern population of S. s. brownorum (the “correct” assignment based on morphology and habitat) and mtDNA of southern S. s. brownorum populations. Similarly, sample no. 23 has the nDNA of S. a. subniger (“correct” assignment based on morphology and habitat) and mtDNA of S. s. scalaris. Given the morphological distinctiveness of these populations (S. s. brownorum North vs. South, and S. a. subniger vs. S. s. scalaris) and the geographic locations of these populations where the respective clades overlap, we infer these disagreements between mtDNA and nDNA to represent cases of localized mitochondrial introgression. Mitochondrial introgression (and gene flow) has previously been posited as the source of conflict between mtDNA and nDNA relationships in Sceloporus lizards (Leaché 2010), and studies of other closely related phrynosomatids have also inferred historic introgression between species (Leaché and McGuire 2006). Given that this conflict is likely due to historic introgression and not incomplete lineage sorting, we did not combine the mtDNA and nDNA data sets for our Bayes factor species delimitation analyses. Our study thus provides a multi-locus nDNA perspective to species delimitation within the scalaris group.

The 12 lineages labeled in Figure 5 represent the most divided set of relationships based on the concatenated nDNA data set for validation via species delimitation (= Model A, Fig. 3). The multi-locus nDNA network inferred with POFAD recovered an identical set of groupings as in the concatenated nDNA tree (Supplementary Fig. S6), and we therefore restrict our discussion to the concatenated nDNA tree. Species delimitation Models B and C represent lumping lineages S. a. aeneus and S. a. subniger West, respectively, with S. a. subniger Central, Southern, and Eastern. The monophyly of the scalaris group is strongly supported, with the root of the scalaris group strongly placed between the
northern and southern clades (Fig. 5). The northern clade consisted of two subclades, one being weakly supported and containing *S. s. scalaris, S. scalaris unicanthalis, and the southern populations of *S. s. brownorum. The other subclade was strongly supported and contained the northern populations of *S. s. brownorum, *S. aenei, *S. chaneyi, and *S. samcolemani. The three taxa belonging to the southern clade include *S. a. aeneus, *S. a. subniger, and *S. bicantalalis; *S. a. aeneus is paraphyletic and includes *S. a. aeneus. Though the specific placement of the single *S. a. aeneus individual among the *S. a. subniger individuals is weakly supported, this lineage was treated as its own lineage in a subset of our analyses due to historical taxonomy. Geographic structuring was recovered within *S. a. subniger where two individuals in the western portion of this taxon’s range were strongly supported as sister to the remaining samples. These clades were subsequently tested as separate lineages (*S. a. subniger West, *S. a. subniger) in the species delimitation analyses. Similarly, though weakly placed among other subspecies of *S. scalaris, the *S. s. scalaris individual from Tapalpa was treated as a separate lineage due to high genetic divergence (e.g., a long branch separates it from other *S. s. scalaris individuals) and distinctiveness in the mtDNA tree (Figs. 1 and 5; Supplementary Fig. S1). Although some lineages were comprised of 1–2 individuals in a subset of the species delimitation models examined, recent work has shown that this low level of sampling can produce moderate levels of species tree accuracy and precision in *BEAST (Camargo et al. 2012a).

A high level of phylogeographic structuring was recovered in the mtDNA tree (Fig. 1; Supplementary Fig. S1). However, many of the basal relationships between geographically structured clades were weakly supported. The clades labeled in Figure 1 represent the most divided set of relationships based on the mtDNA tree (= Model D; Fig. 3). Models E and F contain the lumped lineages of *S. a. aeneus with *S. a. subniger Central, and *S. a. aeneus with *S. a. subniger Central and Southern, respectively. Whereas Model G contains the same species as Model C, some individuals have been reassigned to different species in Model G (see Fig. 3 caption). The tree inferred from the morphological data set of (Wiens and Reeder 1997) is shown in Figure 1, and represents species delimitation Model H (Fig. 3).

**BFD of Species: scalaris Group**

Computational time for running the MCMC chain in *BEAST ranged from 12 to 14 h, whereas PS and SS marginal-likelihood estimators took an additional 9–11 h on a desktop iMac for each replicate (8 cores, 2.7 GHz/processor). AICM results were characterized by low repeatability and high variance between replicates, and the ordering of species delimitation models was not consistent across independent runs (Supplementary Fig. S2). Therefore, the AICM method was not used in the final species delimitation model selection analyses for the *scalaris* group.

Other than Model A, the rankings among species delimitation models by HME/sHME were the same, and PS and SS analyses were in exact agreement about the ordering of the eight species delimitation models (Fig. 6; Table 3). Furthermore, PS and SS estimations provided greater discrimination between competing species delimitation models. Using BFD, Model A, which was based on groupings from the concatenated nuclear phylogeny (and multi-locus nDNA network) and recognizes all *scalaris* group taxa as distinct, including three previously undescribed lineages, received “decisive” support over all other models (2lnBF = 29.33–1414.05 and 30.93–1405.59 in PS and SS analyses, respectively). The consistently least-favored species delimitation models were based on morphological groupings (Model H) and the maximum-likelihood mtDNA tree (Model D; Fig. 6, Table 3).

The HME and sHME consistently produced the highest marginal-likelihood values out of the four marginal-likelihood estimator methods examined, and also provided little discrimination between many of the competing species delimitation models (Fig. 6;
TABLE 3. Marginal-likelihood estimates and Bayes factor testing results (2lnBf) from the combined analyses of four independent runs with sHME, HME, PS, and SS methods

<table>
<thead>
<tr>
<th>Model</th>
<th>sHME</th>
<th>HME</th>
<th>PS</th>
<th>SS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MLE</td>
<td>2lnBf</td>
<td>MLE</td>
<td>2lnBf</td>
</tr>
<tr>
<td>A</td>
<td>-11 485.08</td>
<td>59.37</td>
<td>-11 457.45</td>
<td>3.26</td>
</tr>
<tr>
<td>B</td>
<td>-11 455.40</td>
<td>N/A</td>
<td>-11 455.82</td>
<td>N/A</td>
</tr>
<tr>
<td>C</td>
<td>-11 460.09</td>
<td>11.33</td>
<td>-11 461.19</td>
<td>10.75</td>
</tr>
<tr>
<td>D</td>
<td>-12 128.89</td>
<td>1346.99</td>
<td>-12 128.80</td>
<td>1345.97</td>
</tr>
<tr>
<td>E</td>
<td>-11 462.35</td>
<td>13.91</td>
<td>-11 462.83</td>
<td>14.02</td>
</tr>
<tr>
<td>F</td>
<td>-11 461.78</td>
<td>12.75</td>
<td>-11 462.60</td>
<td>13.56</td>
</tr>
<tr>
<td>G</td>
<td>-11 464.56</td>
<td>18.33</td>
<td>-11 465.04</td>
<td>18.44</td>
</tr>
<tr>
<td>H</td>
<td>-11 459.50</td>
<td>8.21</td>
<td>-11 460.44</td>
<td>9.24</td>
</tr>
</tbody>
</table>

The model receiving the best marginal-likelihood score for each estimation method is indicated by a 2lnBf score = N/A, and its associated marginal likelihood is in bold.

FIGURE 7. Species delimitation results from BP&P. The concatenated a) nDNA, b) mtDNA, and c) morphology guide trees are equivalent to Models A, D, and H in Figure 3, respectively. Letters at nodes represent the species delimitation model resulting from collapsing daughter taxa from that node into a single composite lineage. Model G is not shown because lineage compositions differ from Model D due to individual reassignments. Black dots represent nodes that received a Pp-value ≥ 0.95 across the three prior combinations examined (see Methods section for details). Nodes that did not receive Pp ≥ 0.95 across all prior combinations examined are not labeled. The dashed branches represent where the subclades that served as guide trees were joined together (Supplementary Fig. S1).

Table 3). According to these two estimators, Model B, based on the concatenated nDNA phylogeny (but with S. a. aeneus + S. a. subniger East collapsed into a single lineage), received the best marginal-likelihood score. Model D, based on the maximum-likelihood mtDNA tree, provided the worst explanation of the data (Figs. 3 and 6; Table 3).

Bayesian Species Delimitation

The BP&P results are shown in Figure 7. Although the northern clade in the mtDNA gene tree was fully resolved at the base, basal relationships within the southern clade were weakly supported (Fig. 1; Supplementary Fig. S5), thus making it difficult to generate a fully bifurcating guide tree for BP&P analysis. To deal with this problem, we generated the BP&P guide tree based on the fully resolved ML tree (including weakly supported nodes); when bootstrap values were <70 in this tree, we considered multiple guide trees to reflect the uncertainty of these relationships. This resulted in five subclades that served as guide trees and were tested in separate analyses. Results from these independent analyses were combined into a single tree (Supplementary Fig. S1). In the groupings based on the most divided set of mtDNA relationships (Model D), all lineages except S. a. subniger South, S. a. subniger Central and S. a. aeneis
were recognized as distinct with Pp ≥ 0.95. For the concatenated nDNA guide tree (Model A), all nodes received a Pp of 1.0 for all θ and τc combinations, except for the node shared by S. a. aeneus and S. a. submigere Centra (-0.95). In the guide tree based on the morphological relationships (Model H), all lineages were recognized as distinct with nodal Pp = 1.0.

**DISCUSSION**

Bayes factors are a common model selection tool in phylogenetics (Sullivan and Joyce 2005), and their use has been demonstrated in discriminating between data partitioning models or the conformity of a data set to a molecular clock (e.g., Fan et al. 2011; Baele et al. 2012). In this article, we extend the use of Bayes factors to the problem of species delimitation within the context of Bayesian species tree inference. Using Bayes factors, we can compare the marginal likelihoods of species delimitation models that differ with respect to the number of species and/or the assignment of individuals to species. Marginal-likelihood scores can vary depending on the estimator used to calculate them, so we compared the impacts of four readily available marginal-likelihood estimators and the AIC technique for ranking models.

A number of benefits exist when using the BFD approach to testing models of species limits. First, in terms of species tree inference, topological uncertainty is accounted for at the level of gene trees and the species tree. This is highly beneficial since gene trees are often difficult to estimate with high accuracy, particularly for taxa characterized by rapid radiations and/or recent divergences (both scenarios producing few phylogenetically informative characters), or simply with large sample numbers that are typical in phylogeographic and species delimitation studies (e.g., Belfiore et al. 2008; Harrington and Near 2012). Furthermore, some species delimitation approaches assume that the gene trees are inferred without error (e.g., SpedestEM), which is an assumption that is not commonly met. A related benefit of using BFD is that the user does not have to define a set of phylogenetic relationships between lineages a priori (e.g., the guide tree in BF&P; Yang and Rannala 2010). In many cases, including when the study system is poorly known, the researcher may have little prior knowledge on relationships within the group, or otherwise would not want to fix the topology of the species tree. In this case, assigning individuals to lineages can be done through population assignment methods, or phylogenetic inference as we have done, with no constraint on relationships between lineages. Allowing *BEAST to reconstruct the evolutionary relationships between these lineages then accommodates phylogenetic uncertainty in the species-level phylogeny.

A third advantage of BFD is that it can rank non-nested species delimitation and reassignment models in a way that other currently available species delimitation methods cannot feasibly perform. Specifically, our approach simultaneously tests species limits (i.e., lineage assignments) and reconstructs the species tree for the given model. Whereas SpedestEM (Ence and Carstens 2011) can rank non-nested models, it has the critical assumption that the gene trees have been inferred without error. Similarly, although it is possible to investigate reassignment scenarios in BF&P, the assumption of a fixed guide tree can mislead species delimitation in these cases.

A similar approach to BFD was used when delimiting species of *Myotis* bats (Carstens and Dewey 2010). In this method, two lineages were separated or collapsed for analysis in BEST (Liu 2008; Liu et al. 2008), and Baele factors were computed between alternate species delimitation models to determine the model that best explained the data. Although we use a similar approach of probabilistic modeling, our BFD method obviates convergence issues these authors confronted with BEST (e.g., high standard deviation of split frequencies between independent runs). Furthermore, our method explores the critical issue of using alternative (and more accurate) marginal-likelihood estimators when comparing species delimitation models through Bayes factors.

Although results across independent replicates for any particular species delimitation model were largely congruent with one another in HME, sHME, PS, and SS analyses, the AICM method showed erratic behavior and high variance between runs with the empirical and simulated data alike (Supplementary Fig. S2). Importantly, the ranking of species delimitation models was not the same between the four independent runs with the empirical data. Similarly, AICM results from analysing the simulated data showed high variance and little discrimination between competing species delimitation models (Supplementary Fig. S2). The AICM method multiplies the likelihood value by –2 (Raftery et al. 2007), therefore the variance is automatically expected to be doubled with this method. Baele et al. (2012) also reported high levels of variance between independent AICM runs when analysing simulated and empirical HIV data sets. We do not recommend using the AICM for Bayes factor model selection during species delimitation.

**Comparison to an Existing Species Delimitation Method**

Although we wanted to evaluate the performance of BF&P with our simulated data sets, we were unable to test one of our three species delimitation scenarios, “Reassign.” BF&P is not able to reassign individuals between lineages during analysis. Furthermore, BFD can be applied to situations where the implementation of BF&P would be difficult, such as in cases where the relationships are not known between populations, and when competing species delimitation models are non-nested (e.g., testing reassignment of individuals). These points further emphasize the differences in
methodology between our novel Bayes factor method and Bayesian species delimitation. Because of the analytical differences between these two methods, we suggest exploring species delimitation through both means. Our simulation studies showed higher accuracy and Bayes factor discrimination among competing models with PS and SS methods, so we restrict our interpretations to the results from these marginal-likelihood estimators. Studies by Fan et al. (2011), Xie et al. (2011), and Baele et al. (2012) also showed the higher accuracy of these marginal-likelihood estimators. Comparison of BFD and BPP & P species delimitation produced contradictory results. Specifically, all lineages in Models D and H (based on the ML mtDNA tree and morphological relationships, respectively) were supported as distinct through analysis in BPP & P. However, these models were strongly selected against in BFD. The models based on mtDNA relationships were particularly difficult to test in BPP & P because of the weak support at the base of the southern clade (Fig. 1, Supplementary Fig. S5). No clear way exists of how to handle polytomies in constructing the BPP & P guide tree (an issue that does not have to be confronted with our Bayes factor approach), but our method of resolution (Supplementary Fig. S1) may give erroneous support to these nodes and augment the discordant results between our BFD method and BPP & P. The strong support provided by BPP & P for the independence of all morphological lineages and nearly all mtDNA lineages could be due to four factors. First, the pattern of evolutionary relationships as inferred from the mtDNA may not reflect the pattern inferred from the nDNA (Bensch et al. 2006; Leaché and McGuire 2006; Leaché 2010). With maternal inheritance and effective population size one-fourth of that of nDNA, processes such as sex-biased dispersal, gene flow, and the stochastic process of lineage sorting may lead to conflicting evolutionary histories between DNA types (Ballard and Whitlock 2004). Second, if the user-specified BPP & P guide tree does not accurately reflect historical relationships and lineages are arbitrarily placed sister to one another, as could be the case with analysing nDNA using a guide tree based on morphological relationships, support for the nodes separating these genetically distant related lineages will increase (i.e., the analyses will tend to not support the collapsing of two genetically divergent lineages into a single species; Leaché and Fujita 2010). Third, the program may simply be incorrectly inferring speciation events given our data. This, however, seems unlikely as false-positive rates were inferred to be low with simulated data sets of a comparable size, i.e., fewer than five individuals sampled per lineage (Zhang et al. 2011). Finally, undetected hybridization would violate the assumptions of our coalescent-based analyses. However, simulation studies have suggested that coalescent-based methods are robust to relatively low levels of gene flow and hybridization (Eckert and Carstens 2008; Zhang et al. 2011). Our current genetic sampling is inadequate to perform any rigorous tests for hybridization or estimate levels of gene flow between scalaris group populations. However, we would expect our coalescent-based “BEAST analyses to collapse sister populations into a single lineage if gene flow among them is relatively high (Leaché et al. 2013).

**Taxonomic Recommendations**

In our study, we statistically tested the status of all extant S. scalaris species group taxa and populations as distinct independent evolutionary lineages (= species). Because of our limited sampling, particularly for lineages currently recognized subspecies, we have identified these three cryptic lineages: S. s. brownorum from the western portion of the Trans-Mexican Volcanic Belt, S. s. scalaris from Tapalpa, Jalisco, and S. s. bicanthalis from the Sierra Laurel in Aguascalientes and adjacent Jalisco. Reassigning individuals from the southern portion of the S. s. brownorum range to the northern S. s. brownorum populations provided decisive Bayes factor support to recognize these two populations as distinct lineages. Formal taxonomic descriptions of these lineages are needed and are forthcoming. Finally, S. bicanthalis, S. chaneyi, S. sanselemani, and S. sicvini are confirmed as distinct species.

**Future Work**

One potential shortcoming of our BFD method is that depending on outgroup species sampling, it cannot be run on a single species as currently implemented. However, if the researcher has sequence data for at least one outgroup species, BFD can be performed. The utility of our method in conjunction with population assignment methods as a means to generate a priori individual assignment to species warranting investigation. This would be particularly valuable with admixed individuals and the exploration of individual reassignment. The idea of individual assignment highlights the issue of the lack of species
FIGURE 8. Multi-locus nDNA-based phylogenetic relationships inferred in *BEAST from the highest ranked and strongest supported species delimitation model (Tables 2 and 4). The outgroup (*S. clarkii*) is included, and Bayesian Pp support values are given at each node. Three new independent evolutionary lineages were supported by Bayes factor species delimitation: *Sceloporus* sp. nov. "Laurel", *Sceloporus* sp. nov. "Tapalpa", and *Sceloporus* sp. nov. "subniger West".

TABLE 4. Previous taxonomic designations for *S. scalaris* group taxa and new taxonomic proposals based on Bayes factor species delimitation results

<table>
<thead>
<tr>
<th>Previous taxon name</th>
<th>New taxonomic designation</th>
<th>Sample nos.</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Sceloporus</em> a. aeneus</td>
<td><em>Sceloporus</em> aeneus</td>
<td>34</td>
</tr>
<tr>
<td><em>Sceloporus</em> a. subniger</td>
<td><em>Sceloporus</em> sp. nov. &quot;subniger West&quot;</td>
<td>24, 25</td>
</tr>
<tr>
<td><em>Sceloporus</em> a. subniger</td>
<td><em>Sceloporus</em> a. subniger</td>
<td>20–23, 26–28</td>
</tr>
<tr>
<td><em>Sceloporus</em> bicanthalis</td>
<td><em>Sceloporus</em> bicanthalis</td>
<td>29–33</td>
</tr>
<tr>
<td><em>Sceloporus</em> chaneyi</td>
<td><em>Sceloporus</em> chaneyi</td>
<td>9, 10</td>
</tr>
<tr>
<td><em>Sceloporus</em> samcolemani</td>
<td><em>Sceloporus</em> samcolemani</td>
<td>6–8</td>
</tr>
<tr>
<td><em>Sceloporus</em> s. brownorum</td>
<td><em>Sceloporus</em> brownorum</td>
<td>11, 12</td>
</tr>
<tr>
<td><em>Sceloporus</em> s. brownorum</td>
<td><em>Sceloporus</em> sp. nov. &quot;Laurel&quot;</td>
<td>13, 14</td>
</tr>
<tr>
<td><em>Sceloporus</em> s. unicanthalis</td>
<td><em>Sceloporus</em> unicanthalis</td>
<td>15</td>
</tr>
<tr>
<td><em>Sceloporus</em> s. scalaris</td>
<td><em>Sceloporus</em> scalaris</td>
<td>17–19</td>
</tr>
<tr>
<td><em>Sceloporus</em> s. scalaris</td>
<td><em>Sceloporus</em> sp. nov. &quot;Tapalpa&quot;</td>
<td>16</td>
</tr>
<tr>
<td><em>Sceloporus</em> slevini</td>
<td><em>Sceloporus</em> slevini</td>
<td>1–5</td>
</tr>
</tbody>
</table>

Sample numbers correspond to sampling localities in Figure 1 and Supplementary Table 1.

Determining the impacts of missing data, and how partitioning the data set into few individuals per lineage might affect species delimitation results, should also be addressed through further simulations. Regarding the *scalaris* group, we would like to collect more molecular data to assess putative past and current levels of gene flow between *S. a. aeneus*/*S. a. subniger* and *S. s. brownorum* North and South. These data would help support or refute our hypothesis of historic introgression between lineages, and potentially lend insight into the use of our method of BFD of species, in spite of gene flow, within a multi-species coalescent framework.

SUPPLEMENTARY MATERIAL

Data files and/or other supplementary information related to this article have been deposited at Dryad under doi:10.5061/dryad.c7s77.

FUNDING

We thank the American Museum of Natural History for the Theodore Roosevelt Memorial Fund awarded to J.A.G. and the National Science Foundation for the ATOL (Assembling the Tree of Life) award to TW.R. (EF 0334967).

ACKNOWLEDGMENTS

For help with marginal-likelihood estimations, we would like to thank G. Baele. For comments and suggestions on earlier versions of the manuscript, we thank the Reeder and Leaché labs, B. Carstens, and one anonymous reviewer.

REFERENCES


