Syst. Biol. 61(2):272–288, 2012

DOI:10.1093/sysbio/syr105
Advance Access publication on November 10, 2011

Accuracy and Precision of Species Trees: Effects of Locus, Individual, and Base Pair Sampling on Inference of Species Trees in Lizards of the *Liolaemus darwinii* Group (Squamata, Liolaemidae)

ARLEY CAMARGO[1,*], LUCIANO J. AVILA[2], MARIANA MORANDO[2], AND JACK W. SITES JR.[1]

[1]*Department of Biology and Bean Life Science Museum, Brigham Young University, Provo, UT 84602, USA; and* [2]*CONICET-CENPAT, Boulevard Almirante Brown 2825, U9120ACF, Puerto Madryn, Chubut, Argentina;*
[*]*Correspondence to be sent to: CONICET-CENPAT, Boulevard Almirante Brown 2825, U9120ACF, Puerto Madryn, Chubut, Argentina; E-mail: arley.camargo@gmail.com.*

Received 21 June 2011; reviews returned 24 August 2011; accepted 1 November 2011
Associate Editor: Elizabeth Jockusch

Abstract.—Molecular phylogenetics has entered a new era in which species trees are estimated from a collection of gene trees using methods that accommodate their heterogeneity and discordance with the species tree. Empirical evaluation of species trees is necessary to assess the performance (i.e., accuracy and precision) of these methods with real data, which consists of gene genealogies likely shaped by different historical and demographic processes. We analyzed 20 loci for 16 species of the South American lizards of the *Liolaemus darwinii* species group and reconstructed a species tree with *BEAST, then compared the performance of this method under different sampling strategies of loci, individuals, and sequence lengths. We found an increase in the accuracy and precision of species trees with the number of loci, but for any number of loci, accuracy substantially decreased only when using only one individual per species or 25% of the full sequence length (~147 bp). In addition, locus "informativeness" was an important factor in the accuracy/precision of species trees when using a few loci, but it became increasingly irrelevant with additional loci. Our empirical results combined with the previous simulation studies suggest that there is an optimal range of sampling effort of loci, individuals, and sequence lengths for a given speciation history and information content of the data. Future studies should be directed toward further assessment of other factors that can impact performance of species trees, including gene flow, locus "informativeness," tree shape, missing data, and errors in species delimitation. [Accuracy; coalescent; Liolaemus; phylogeny; precision; sampling; South America; species trees.]

Molecular phylogenetics has entered a new era in which species trees are estimated from a collection of gene trees by accommodating their heterogeneity and discordance with the species tree (Edwards 2009; Knowles and Kubatko 2010). Over 2 decades ago, it was realized that gene trees could be highly heterogeneous and be discordant with the species tree due to a variety of processes including estimation error, incomplete lineage sorting, horizontal gene transfer, and gene duplication/loss (Pamilo and Nei 1988; Avise 1989; Maddison 1997). Until recently, standard approaches have assumed that all gene trees matched the underlying species tree and relied on sequence concatenation, which was shown to be more accurate than consensus methods (Gadagkar et al. 2005). However, simulation studies have found that concatenation is inconsistent in an "anomaly zone" (Kubatko and Degnan 2007) in which the most frequent gene trees do not match the species tree (anomalous gene trees [AGTs]; Degnan and Rosenberg 2006). As an alternative to concatenation, a gene-tree parsimony approach based on reconciliation of the gene trees with the species tree was proposed over a decade ago (Page 1998; Slowinski and Page 1999), but concatenation remained the preferred choice in practice. Subsequently, Maddison (1997) introduced the idea of a summary-statistic approach based on minimizing deep coalescences (DCs) across multiple gene trees, and more recently, a variety of other approaches have been proposed that use summary statistics (species tree using average ranks of coalescence times [STAR] and species tree using average coalescence times [STEAC], Liu, Yu, Kubatko, et al. 2009; Global LAteSt Split tree, Mossel and Roch 2010), consensus/super tree methods (Degnan et al. 2009), Bayesian concordance factors (Ané et al. 2007), and approximate Bayesian computation approaches (Fan and Kubatko 2011).

A new generation of methods has explicitly incorporated gene-tree heterogeneity due to incomplete lineage sorting into species-tree estimation, based on the multispecies coalescent model (Rannala and Yang 2003; Degnan and Rosenberg 2009). These novel model-based frameworks have led to the development of maximum likelihood (ML) and Bayesian species-tree inference approaches (Liu, Yu, Kubatko, et al. 2009). The ML approach is implemented in the program STEM (Kubatko et al. 2009), which combines user-provided constant population size, estimated gene trees, and relative rates among loci to obtain the ML species tree with branch lengths that accommodate rate variation and differences in ploidy level across loci (Kubatko et al. 2009). A Bayesian approach has been implemented in 2 programs: BEST (Liu et al. 2008) and BEAST (*BEAST; Drummond and Rambaut 2007; Heled and Drummond 2010). BEST applies a hierarchical design to estimate the joint posterior distribution of species trees and gene trees, conditional on the observed sequence data with the restriction that species divergence times cannot predate the coalescence times of alleles. In addition, BEST estimates gene trees without assuming a molecular clock and then ultra-metricizes branch lengths (Castillo-Ramirez et al. 2010). On the other hand, *BEAST relaxes the molecular clock for estimating gene trees and accommodates changing population sizes across the species tree (Heled and Drummond 2010). These methods

further assume that loci are unlinked (free recombination between loci), with no intra-locus recombination, and that gene-tree heterogeneity is due only to incomplete lineage sorting. Newer approaches that incorporate hybridization to the coalescent-based species trees are in active phase of development and the first empirical results are promising (Kubatko 2009; Kubatko and Meng 2010; Chung and Ane 2011; Yu et al. 2011), whereas a recently proposed summary-statistic method appear to be robust to limited gene flow or horizontal gene transfer (STAR and STEAC; Liu, Yu, Kubatko, et al. 2009).

The performance (i.e., accuracy and precision) of multilocus species-tree methods is beginning to be investigated with simulations to assess the impact of sampling strategies (McCormack et al. 2009; Castillo-Ramirez et al. 2010; Heled and Drummond 2010; Leache and Rannala 2011) and to disentangle the relative influence of coalescent versus mutational variance (Huang et al. 2010). The performance of some of these new methods (STEM and BEST) has also been evaluated in the context of species delimitation (Carstens and Dewey 2010). There are at least 3 dimensions in the size of data sets that can be subsampled to evaluate their impact on performance: number of loci, number of individuals, and sequence length (Brito and Edwards 2009). Another dimension of sampling is the variation in locus "informativeness" or phylogenetic signal (Knowles 2009), which, although a dominant factor in empirical studies, has rarely been explored in simulation studies, which typically assume that all loci have the same properties (e.g., same substitution model, same sequence length, etc.).

Herein, we used multilocus sequence data for the *Liolaemus darwinii* group of South American lizards (Squamata, Liolaemidae) to reconstruct a species tree for the group. The genus *Liolaemus* is one of the most ecologically diverse and species-rich genera of lizards on Earth, with 225 recognized species (Lobo et al. 2010), and this is likely a large underestimate (Morando et al. 2003). The genus is distributed over a wide geographic area spanning a large range of latitudinal ($14^\circ \pm 30'$–$52^\circ \pm 30'$S), altitudinal (0–4500 m), and climatic regimes, from the extremely arid Atacama Desert to temperate *Nothofagus* rainforest (Lobo et al. 2010). Species of the *L. darwinii* group inhabit the arid lands of the Monte Desert region of central and northwestern Argentina (Fig. 1), and several morphological characters support its monophyly. The group currently contains 18 recognized species, many of which have been described in the last 2 decades after the most recent taxonomic review (Etheridge 1993). A recent combined molecular/morphological study recovered the first phylogenetic hypothesis for the group as a strongly supported clade nested within the more inclusive *L. boulengeri* clade (Abdala 2007). In this study, we sampled multiple loci from most described species in the *L. darwinii* group to reassess relationships within this clade and to present a working hypothesis that will serve as a framework for ongoing phylogeographic, species delimitation, and speciation studies of these lizards.

The conceptual focus of this study was to empirically evaluate the performance of a multilocus Bayesian species-tree method under varying sampling designs. Here we use the term "performance" when referring to the accuracy and precision of species trees estimated with variable sampling designs and effort. Accuracy and precision estimates in a phylogenetic context (see definitions in Hillis and Bull 1993) can be used for assessing how well a species-tree method recovers the correct tree under varying conditions and sampling effort (Liu, Yu, Kubatko, et al. 2009). Accuracy can be quantified as the distance between the best estimate of the species tree based on all available data and the estimate obtained with a subsample of the data. In addition, Bayesian methods naturally allow for precision estimates via the variation in tree distances among the set of trees in the posterior distribution. The subsampling approach allowed us to evaluate optimal sampling strategies that approximated the best-supported species tree with fewer data (fewer loci, individuals per species, and base pairs).

METHODS

Taxon Sampling

We analyzed 20 loci from 47 individuals representing 16 of the 18 species in the *L. darwinii* group; we could not obtain samples of *L. montanezi* and *L. cinereus* (Appendix 1). For all species except those in the *L. darwinii* complex (*L. darwinii*, *L. grosseorum*, *L. laurenti*, and *L. olongasta*), which were sampled more widely, we collected at or near type localities to include only described lineages. Taxonomic knowledge of the *L. darwinii* group is still incomplete, and there are several candidate species awaiting further study. Based on the cyt *b* topology (see below), 3 individuals with divergent cyt *b* haplotypes were subsampled for each species for all nuclear loci (Appendix 1). Gene flow likely occurs between some species in the *L. darwinii* complex (Morando et al. 2004), which might affect species-tree methods that typically do not accommodate for this source of gene-tree discordance (Leache 2009). Therefore, we excluded individuals from phylogeographic borders or contact zones with other species in the group to minimize the potential impact of intermixed/migrant individuals.

Sequence Data

Genomic DNA was extracted with the DNA easy kit (Qiagen). We used the Green Go Taq PCR kit (Promega) for all polymerase chain reactions (PCRs) in PTC-200 DNA Engine (MJ Research) or GeneAmp PCR 9700 thermal cyclers (Applied Biosystems, Inc.). Sequencing reactions used the Big-Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Inc.) in a GeneAmp PCR 9700 thermal cycler (Applied Biosystems, Inc.). Sequencing products were cleaned with Sephadex G-50 Fine (GE Healthcare Bio-Sciencies AB) and sequenced

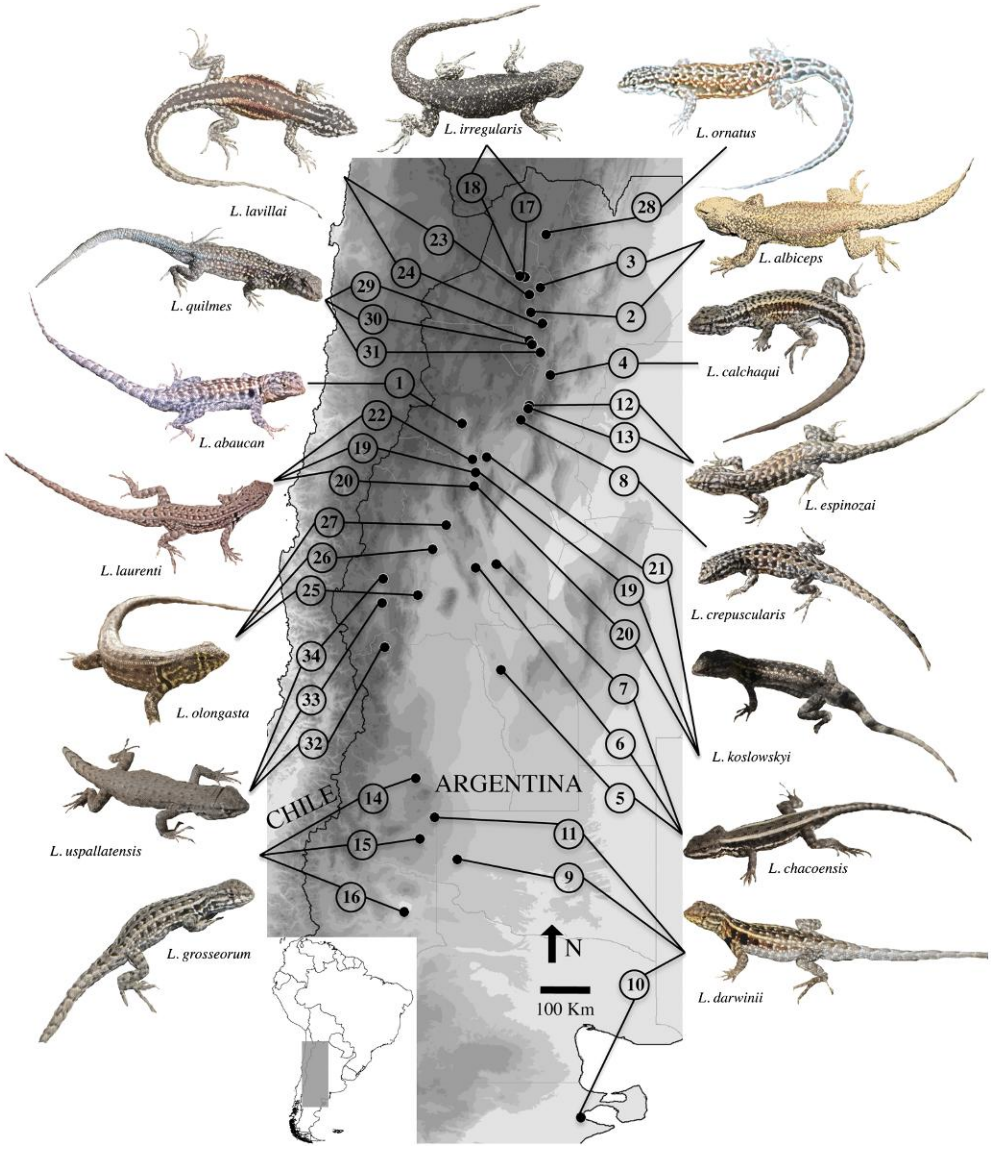

FIGURE 1. Map of Argentina showing sampled localities for species of the *Liolaemus darwinii* group. Numbers in circles refer to the localities listed in Appendix 1, and darker shading indicates higher elevations. The inset shows the map of South America with the study area in gray. Pictures are not scaled. Photos by L.J.A. (*L. darwinii*, *L. laurenti*, *L. calchaqui*, *L. quilmes*, *L. olongasta*, *L. chacoensis*, *L. ornatus*, *L. albiceps*, and *L. abaucan*) and C. Pérez (*L. lavillai*, *L. uspallatensis*, *L. grosseorum*, *L. irregularis*, *L. crepuscularis*, *L. koslowskyi*, and *L. espinozai*).

in an automated sequencer ABI 3730xl DNA Analyzer (Applied Biosystems, Inc.). The cyt *b* mitochondrial DNA (mtDNA) gene was sequenced for a large number of individuals of the *L. darwinii* group (~900 sequences) following methods in Morando et al. (2004). Anonymous nuclear loci (ANL) developed from an individual of *L. darwinii* (LJAMM 7097) were screened for all species included in this study based on the protocols of Noonan and Yoder (2009). From 30 ANL tested in PCR reactions across the sampled individuals, 20 produced positive PCR products for most samples, and the 12 most informative (B9G, A8F, A4B, B3F, A1D, A6D, A9C, B5B, A12D, B8H, B1D, and A9E) were selected for subsequent analyses. One highly variable protein-coding gene (PRLR, Townsend et al. 2008), 5 additional protein-coding genes (CMOS, Saint et al. 1998; ACM4, Gamble et al. 2008; MXRA5, EXPH5, and KIF24, Portik et al. 2011), and 1 intron (BA3, Waltari and Edwards 2002) were also included; this provided a total of 20 loci for analysis (Table 1). PCRs for all nuclear loci except PRLR were performed with the temperature profile of Noonan and Yoder (2009); for PRLR, we used the touchdown cycling protocol for nuclear genes described in Reyes-Velasco and Mulcahy (2010). Sequences were aligned in ClustalX 2.0 (Larkin et al. 2007).

Individuals heterozygous for indels were analyzed with CodonCode Aligner (CodonCode Corp.) to resolve position and length of indels. Ambiguity codes were used to represent polymorphisms in heterozygous individuals in which gap polymorphisms were coded

TABLE 1. List of molecular markers ranked by percentage of variable sites

Locus	Kind of marker	Substitution model	Length (bp)	Variation (%)	DC	GC	DC/GC	Primers	Reference
CYTB	mtDNA	TPM2uf+G	710	35.5	10	48	0.21	Glu, CB3, F1, C2	Morando et al. (2004)
B9G	ANL	K80+G	513	13.7	62	39	1.59	F 5′-AGAGAGGGGAAGGGGTTTG-3′ R 5′-TCCCTTGATATTCACAGACTTAACA-3′	This study
A8F	ANL	HKY	645	12.3	56	39	1.44	F 5′-CTTAACATTTTCAGAACAAGTCTGTC-3′ R 5′-CCCTCCTCATTTACTTTACATGC-3′	This study
A4B	ANL	HKY+G	517	12.0	31	45	0.69	F 5′-ACTCCCGTGGATCTCTGTTG-3′ R 5′-GGGAAAAGGTGGGTGCTTAG-3′	This study
B3F	ANL	K80+I	505	11.9	23	46	0.50	F 5′-CTCACTGCCACAGCAAGAAA-3′ R 5′-TGAGAAAAACTGCAGGTAGCA-3′	This study
A1D	ANL	HKY	783	10.6	51	42	1.21	F 5′-CAATTCTGCAAATCCACCCTA-3′ R 5′-TTGTCAGAAGAATGCTGCAAAT-3′	This study
A6D	ANL	HKY+I	476	9.9	27	47	0.58	F 5′-TGATAGTTACTGCAGGGTCCA-3′ R 5′-GGCTTATTGTTGAGCGGTCT-3′	This study
A9C	ANL	K80+G	745	9.8	100	38	2.63	F 5′-GTGCCCAGTTCTGGTCTTGT-3′ R 5′-AGCCTGAGCCAAACATAGGA-3′	This study
B5B	ANL	HKY	623	9.8	13	44	0.30	F 5′-GCAGAGCCAAAGCCATGT-3′ R 5′-GGTAGCTTGGTGGTAGGTCA-3′	This study
A12D	ANL	HKY	632	9.0	61	37	1.65	F 5′-GCTTCTGGGAAGGCTATGAA-3′ R 5′-TCCAAAATGGCTACACTGAGG-3′	This study
B8H	ANL	HKY	625	8.8	35	38	0.92	F 5′-TTTCTTTTAAGCAGCCAGACA-3′ R 5′-GGTCCCATACGGCTGTGG-3′	This study
EXPH5	Coding	HKY+I	790	7.6	73	43	1.70	F1, R1	Portik et al. (2011)
PRLR	Coding	HKY+I	552	7.6	18	46	0.40	F1, R3	Townsend et al. (2008)
B1D	ANL	HKY	291	7.2	66	44	1.50	F 5′-GATATCGAGGGATTTCAGTTTCC-3′ R 5′-CCAGTGTTTATGAGCAACTGAGTA-3′	This study
KIF24	Coding	HKY+I	551	7.1	15	48	0.31	F1, R1	Portik et al. (2011)
BA3	Intron	K80	347	6.9	65	1	1.59	F, R	Waltari and Edwards (2002)
MXRA5	Intron	HKY	867	5.4	27	42	0.64	F, R	Portik et al. (2011)
ACM4	Coding	TPM1uf+I	453	4.9	22	42	0.52	F, R	Gamble et al. (2008)
A9E	ANL	HKY+I	615	4.2	73	38	1.92	F 5′-TGAACATGCCAGACAGAAACA-3′ R 5′-TCCCGTAGTCCACAAACTGG-3′	This study
CMOS	Coding	K80+I	523	4.2	13	48	0.27	G73, G78	Saint et al. (1998)

as "N." Each locus was analyzed with RDP3 beta35 (Martin et al. 2005) to test for recombination signal, and alignments were also examined to check for fixed heterozygote positions that might suggest the occurrence of multiple-copy loci (Thomson et al. 2010). Best-fit substitution models were obtained in jModeltest 0.1.1 (Posada 2008) with a Bayesian information criterion for model choice (Table 1). We calculated the correlation between locus variation (proportion of variable sites) and the proportion of informative sites for each locus and also calculated the correlation between locus variation and a support index for the corresponding gene tree coestimated with the species tree in *BEAST (see below). This overall support index represents the proportion of nodes in the species tree with posterior probability (PP) $\geqslant 0.95$.

Species Tree

Each locus was included as a separate data partition in an estimate of the species tree using a Bayesian method in *BEAST. We chose this method because it has been shown to outperform other methods under relaxed molecular clock assumptions (Heled and Drummond 2010). In addition, this approach provides an easy and appropriate way to obtain variance estimates from the posterior distribution of trees. First, we ran analyses with all the available data including 20 loci and a maximum of 3 individuals per species per gene. In addition to the species tree, we used the gene trees estimated independently for each locus in *BEAST to evaluate their relative discordance with the species tree using the number of DCs in Mesquite 2.74 (Maddison and Maddison 2010). We calculated the correlation between locus variation and a standardized measure of gene-tree discordance consisting of the ratio between DC and the number of gene copies (GCs).

We used a separate relaxed molecular clock model for each gene with estimation of relative clock rates. We used random starting gene trees under the coalescent model, a Yule process and gamma-distributed population sizes for the species-tree prior, and a continuous population size model with a constant root. Analyses were run for 100 million generations and samples taken every 4000 generations with default prior distributions and operator settings. All analyses were run on the Marylou 5 linux cluster in the Fulton Super computing Lab at Birgham Young University (BYU) https://marylou.byu.edu. Burn-in plots were inspected in Tracer (Rambaut and Drummond 2007) to determine the number of generations necessary to reach stationarity and to reach effective sample sizes (ESSs) higher than 100. Samples taken after this burn-in phase were used for obtaining point estimates and credible intervals of species trees.

In order to evaluate the stability of our recovered tree with the specified conditions, we also estimated the species trees with phased data and different models and priors in *BEAST. We resolved gametic phase of heterozygous sites with the program Phase 2.1.1 (Stephens et al. 2001) when the allelic states had a certainty threshold of 90%, but we maintained ambiguity codes for resolutions below this threshold. Using the unphased data, we first changed the mean birth rate prior of the Yule model, and separately, we used the birth–death model instead of the Yule model. Second, we used a different prior mean population size under the default inverse distribution, and we also implemented the inverse gamma distribution, which is a commonly used prior distribution in other approaches (e.g., BEST). Third, we estimated a *BEAST tree with a longer run using 200 million generations and a sampling frequency of 8000. Fourth, we also estimated species trees after excluding the most variable locus (cyt *b*) and the second most variable locus (B9G) to evaluate their relative impact on the recovered tree in comparison with the analysis based on all 20 loci. In addition, we estimated a species tree using the cyt *b* locus only in order to incorporate a calibration for the mean substitution rate (0.65% substitutions/site/myr; Morando et al. 2004) and to estimate the age of the ancestral split of the *L. darwinii* group following McCormack et al. (2009). Finally, we also analyzed the complete data set with 2 runs in BEST of 4 chains each consisting of 100 million generations sampled every 4000 generations. Input xml files for *BEAST runs were deposited at Dryad (DOI:10.5061/dryad.8m8c0).

Subsampling of Data

We prepared 2 new reduced data sets for each locus by randomly removing one and then 2 individuals per species. To subsample sequence length for each locus, we prepared new data sets for each locus by removing 25%, 50%, and 75% of the sites from all sequences. Based on these new data sets with fewer individuals and base pairs, we randomly subsampled loci to run analyses with 4, 8, 12, and 16 loci (Fig. 2). To minimize the influence of locus-specific effects, subsampling was done in a nested fashion in which 12 loci were sampled from the pool of 16 loci, 8 loci from the pool of 12 loci, and 4 loci from the pool of 8 loci. Five replicate sampling trials were done to evaluate the effect of different locus combinations within each of these 4 subsets, but for analyses with 3 individuals/species, we performed 10 replicates. In beauti (Drummond and Rambaut 2007), we made all possible combinations of number of loci (4, 8, 12, 16, and 20) with number of individuals (1, 2, and 3) and number of loci with proportion of sites (25%, 50% 75%, and 100%).

In addition, because the locus "informativeness" may also have an impact on species-tree performance, we analyzed 3 groups of loci based on their variability (Table 1): (i) most variable (MV) loci, (ii) most conserved (MC) loci, and (iii) a mix of variable and conserved loci (VC). For example, we selected the 2 MV and 2 MC loci for 4 loci, the 4 MV and 4 MC for 8 loci, and so on. Moreover, we also ranked loci based on gene-tree discordance

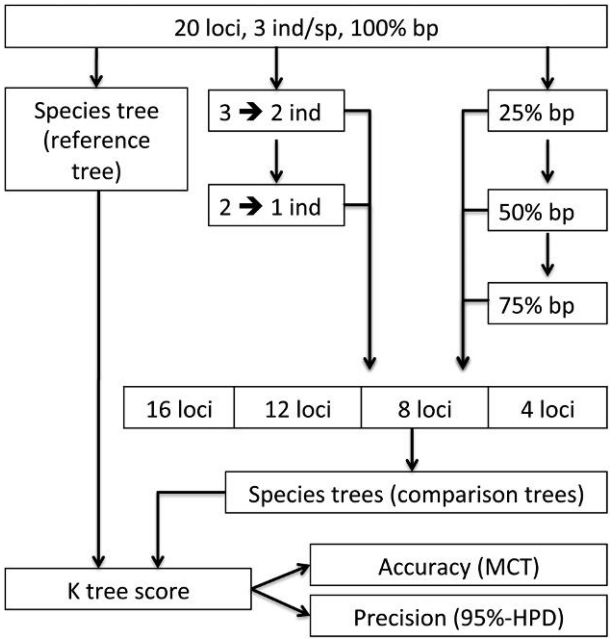

FIGURE 2. Diagram showing the strategy used in this study to sample loci, individuals, and base pairs from 20 loci. MCT, maximum credibility tree; HPD, highest posterior density.

(DC/GC) with the best species tree (Table 1) in 3 classes: (i) most discordant (MD) loci, (ii) least discordant (LD) loci, and (iii) a mix of most and least discordant loci (ML). In order to evaluate the impact of locus "informativeness" in a more general scenario, we simulated 100 gene trees from one symmetric species tree of 8 species using the ms program (Hudson 2002) and assuming a constant population size, 3 alleles sampled per species, and 3 divergence times: 0.25, 0.5, and $1N_e$. We ranked the simulated gene trees based on their discordance with the known species tree using the DCs statistic in Mesquite 2.74 (Maddison and Maddison 2010) and formed 3 sets of gene trees: (i) the 20 LD, (ii) the 20 MD, and (iii) the mix of 10 MD and 10 LD. Next, we simulated sequence data (500 bp) for these 3 sets with SeqGen 1.3.2 (Rambaut and Grassly 1997) using the HKY substitution model and θ per site $= 0.01$, estimated species trees for each set separately in *BEAST, and calculated *K* scores (see below) between the true and estimated species trees. Input nexus file for Mesquite including simulated gene trees and the specified species tree is deposited at Dryad (DOI:10.5061/dryad.8m8c0).

Performance

Liu, Yu, Kubatko, et al. (2009) suggested that the sampling impacts on species-tree estimation could be assessed with a measure of the variance in the tree estimate. More informative data sets are expected to produce more precise estimates coupled with a lower variance. For example, a metric such as the branch length distance (BLD) (Kuhner and Felsenstein 1994), which takes into account topology and branch lengths, could be used to measure distance between trees in the posterior distribution. A modified version of this metric, the minimum BLD or *K* tree score, measures differences in tree topology and *relative* branch lengths, and consequently, the absolute differences in tree depth are scaled to be the same (Soria-Carrasco et al. 2007). This tree distance is not symmetric and therefore is appropriate when one single reference tree (the target "true" tree in our case) is compared with estimates of the reference tree when evaluating the performance of phylogenetic methods (Soria-Carrasco et al. 2007). Because simulations have shown that *BEAST is more accurate when more loci and more base pairs are analyzed (see Figs. 3 and 6 in Heled and Drummond 2010), we assume that the species tree estimated with all data should also represent our best estimate. In this context, we considered "accuracy" not as an estimate of the true species tree but as a measure of approximation toward our best tree based on all the data. In order to calculate accuracy, we summarized the posterior species-tree distributions to obtain the maximum clade credibility (MCC) tree using Tree Annotator (Drummond and Rambaut 2007). The MCC tree estimated from the full data set was used as a reference tree, whereas the best trees obtained with different combinations of loci, sites, and individuals were used as comparison trees to calculate *K* scores with K tree dist v.1 (http://molevol.cmima.csic.es/castresana/Ktreedist.html; Soria-Carrasco et al. 2007). Lower *K* scores were considered as more accurate estimates. To estimate the precision of the species tree, we subsampled 100 trees from the posterior distribution using Log Combiner (Drummond and Rambaut 2007) and used them as comparison trees for calculating *K* scores against the MCC reference tree from the full data set. Based on these 100 scores, we calculated their variance as an estimator of precision since a lower variance in *K* scores represents a more precise estimate of the species tree. The computation of *K* scores using trees from the posterior distribution added the branch support into the tree comparison, in addition to the topological and branch length differences, since the weight of specific branches in the overall *K* value (precision) will be proportional to their frequency in the posterior distribution. In addition, we calculated the correlation between the overall support index and the precision of species trees estimated with varying number of loci and individuals.

RESULTS

Sequence Data

All sequences are available in GenBank (accession numbers JN682514–JN683347). Sequence length varied between 291 bp (B1D) and 867 bp (MXRA5). Percentage of variable sites across the 20 loci ranged between 3% (A9E) and 35% (cyt *b*) (Table 1). Single–base pair indels were rare in ANL sequences, there were a few multiple–base pair indels in some ANL, and alignments were unambiguous except for a 16-bp segment in the A9C locus that was excluded from the analyses

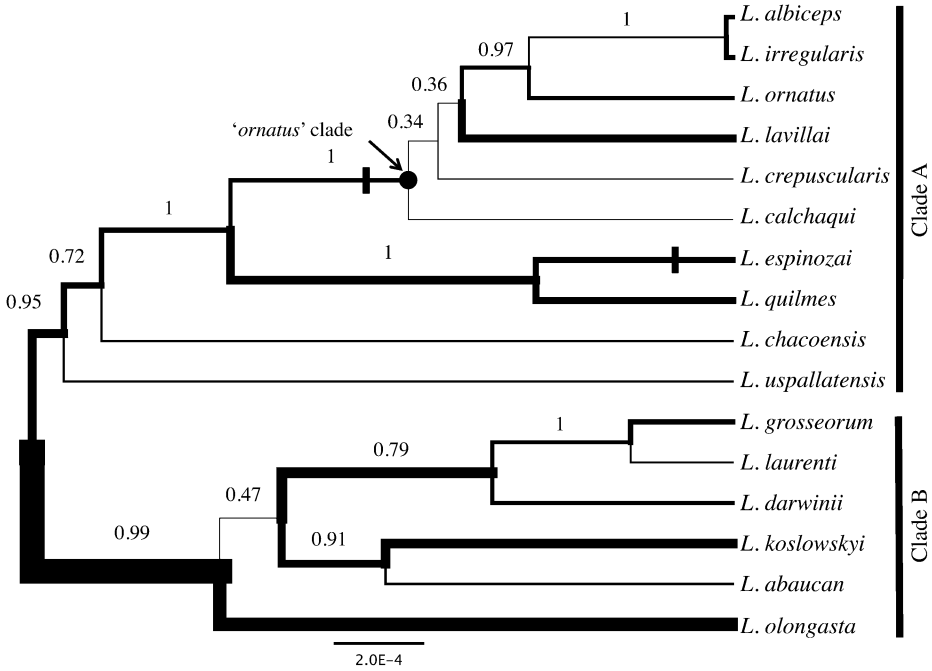

FIGURE 3. Species tree of the *Liolaemus darwinii* group based on 20 loci. Numbers on branches represent PPs. Branch width is proportional to mean relative population size. Transitions from oviparity to viviparity are marked with a transverse bar on branches.

(online Appendix 2, available from http://www.sysbio.oxfordjournals.org/, Dryad DOI:10.5061/dryad.60g211t1). No signal of recombination was detected in any of the data sets. Species represented by only 1 individual occurred in only ~6% of all cases (21 of 320 species/gene combinations [=20 loci × 16 species]), whereas species represented by 2–3 individuals were common across all loci (94%). Only 7% of sequences sampled from within species were identical before phasing the data. Mean percentage of heterozygotes across nuclear loci was 61%, and mean percent site differences between alleles of heterozygotes ranged between 0.32% (A1D) and 1.04% (A9C). The proportion of variable sites was significantly correlated with their proportion of informative sites ($R^2 = 0.75$, $F_{1,17} = 51.6$, $P < 0.01$) and with clade support in their corresponding gene trees ($R^2 = 0.50$, $F_{1,18} = 17.9$, $P < 0.01$) (Figs. 1 and 2 in online Appendix 3, Dryad DOI:10.5061/dryad.60g211t1). Sequence alignments of phased and unphased data as well as output files from Phase are deposited at Dryad (DOI:10.5061/dryad.8m8c0).

Gene Trees and Species Tree

Burn-in plots suggested that stationarity was reached after about 5000 samples (20%); these were discarded and posterior distributions of gene and species trees were estimated with the remaining 20,000 samples. In spite of intraspecific variation, gene trees show little paraphyly within species except for some interdigitation of samples between *L. irregularis* and *L. albiceps*, *L. espinozai* and *L. quilmes*, and *L. laurenti* and *L. grosseorum* (see online Appendix 4, Dryad DOI:10.5061/dryad.60g211t1). In addition, lineages representing species pairs in the best-supported species tree were divergent in all cases except for the (*L. irregularis* + *L. albiceps*) clade, suggesting good support for species-level differentiation. The best-supported species tree derived from complete data sets for 20 loci recovers 8 of 15 nodes with PP $\geqslant$0.95 (Fig. 3), and its topology is discordant with all 20 gene trees (see online Appendix 4). The species tree has 2 well-supported sister and most inclusive clades: Clade A, which groups *L. albiceps*, *L. irregularis*, *L. ornatus*, *L. lavillai*, *L. crepuscularis*, *L. calchaqui*, *L. espinozai*, *L. quilmes*, *L. chacoensis*, and *L. uspallatensis*, and Clade B, including *L. grosseorum*, *L. laurenti*, *L. darwinii*, *L. koslowskyi*, *L. abaucan*, and *L. olongasta* (Fig. 3). The A9C locus was the MC with the species-tree topology, whereas cyt *b* was the LC, based on the ratio between DCs and number of alleles (Table 1, online Appendix 4). The proportion of variable sites was not correlated with the discordance of gene trees ($R^2 = 0.02$, $F_{1,17} = 0.30$, $P = 0.59$), but the precision of species trees and the clade support index were significantly correlated ($R^2 = 0.84$, $F_{1,13} = 66.52$, $P < 0.01$; Figs. 3 and 4 in online Appendix 3).

The species tree estimated with all phased data is similar and does not have any highly supported conflict with our reference tree based on unphased data (Fig. 3) except for the placement of *L. uspallatensis*, which was grouped within Clade B with PP = 0.945 (see Fig. 1 in online Appendix 5, Dryad DOI:10.5061/dryad.60g211t1). The topologies of the alternative species trees estimated

with (i) different prior distributions and models, (ii) a longer run, and (iii) the program BEST are similar and do not have any strongly supported clades that are incongruent with the reference species tree (see Figs. 2–7 in online Appendix 5). The species trees estimated with 19 loci (excluding cyt *b* or B9G) do not have any highly supported conflict (i.e., incongruent clades with PP $\geqslant$ 0.95) with the reference tree estimated with all 20 loci (compare Fig. 3 with Figs. 8 and 9 in online Appendix 5). The ancestral split of the *L. darwinii* group was dated to 14.4 Ma based on the species tree estimated with the calibrated cyt *b* locus, and the mean population size was 476,000, which equals the effective population size (N_e) times the generation length (see BEAST manual). Assuming an average generation length of 2 years for these lizards, this divergence time corresponds to 7.2 million generations and $N_e = 238{,}000$, and therefore, the scaled divergence time is $\sim 30N_e$. The full data set and associated species trees are deposited at TreeBASE (http://purl.org/phylo/treebase/phylows/study/TB2:S11925).

Sampling of Loci and Individuals

Accuracy increased when sampling more than 4 loci but remained constant when analyzing 8 or more loci, for all numbers of individuals sampled per species, and accuracy was consistently higher when sampling 2–3 individuals (Fig. 4a). Precision also shows a steep increase between 4 and 12 loci but did not increase beyond 12 loci, and again sampling of 2–3 individuals per species is always better than sampling only 1 individual (Fig. 4b). The magnitude of the differences in *K* score between the best-supported tree and trees estimated with fewer loci involved differences in topology: trees based on 16 and 12 loci had different relationships within Clade B, and the 8-locus and 4-locus trees grouped *L. uspallatensis* and *L. chacoensis* within Clade A (Figs. 1–4 in online Appendix 6, Dryad DOI:10.5061/dryad.60g211t1). The highest *K* score was obtained for the comparison between the reference tree and the species tree estimated with 4 loci and 1 individual/species (0.00353), which corresponds to a Robinson–Foulds distance of 16. In all cases, branch support, and consequently precision of species trees, declined with fewer loci. Increases in accuracy and precision with number of loci were also seen when the species tree estimated with 16 loci was used as the reference tree (Fig. 5 in online Appendix 6). When we subsampled loci consisting of phased data, we also find a sustained increase in accuracy and precision with number of loci with a gradual approximation to the reference tree based on 20 loci (Fig. 6 in online Appendix 6). In addition, the standard errors in accuracy were similar or lower when using 10 instead of 5 replicates in analyses with variable number of loci (Fig. 7 in online Appendix 6). Mean ESS of posterior, prior, and likelihood parameters for analyses with 2 and 3 individuals/species (1200, 1220, and 488, respectively) were higher than ESS for analyses with 1 individual/species (354, 344, and 443).

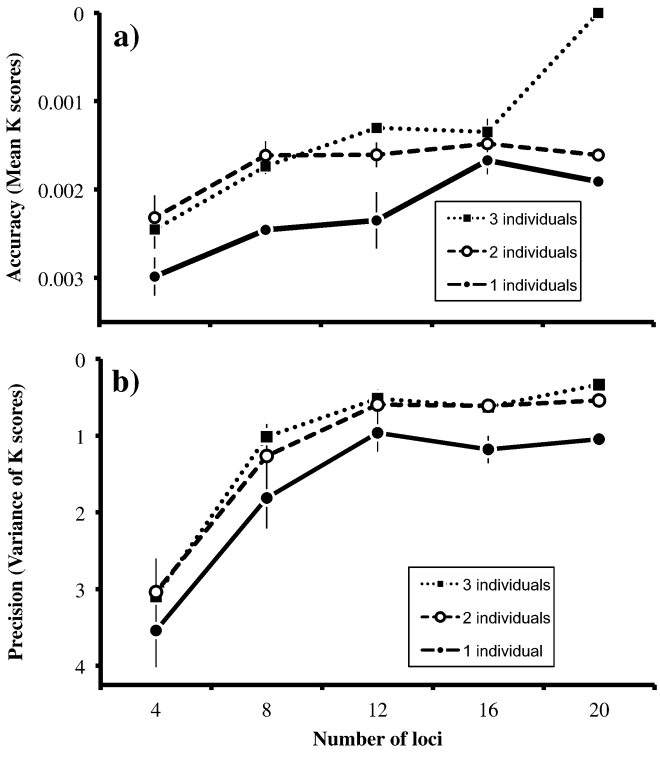

FIGURE 4. Accuracy (a) and precision (b) of species trees estimated from different numbers of loci (4, 8, 12, 16, and 20) and different numbers of individuals (1, 2, and 3).

Sampling of Base Pairs

The accuracy of species trees estimated from data sets with variable number of sites show a similar pattern of improvement with increases in the number of loci and reached a plateau between 12 and 16 loci (Fig. 5a). Accuracy was almost indistinguishable for data sets with full sequence length, 75% (~440 bp/locus), and 50% (~295 bp/locus) of the sites, but analyses run with only 25% of the sites (~147 bp/locus) had lower accuracy for any given number of loci. The highest *K* score was obtained for the comparison between the reference tree and the species tree estimated with 4 loci and 25% of the sites (0.00439), which corresponds to a Robinson–Foulds distance of 26. Precision showed a similar pattern with a large increase between 4 and 12 loci and an apparent plateau after 16 loci. For 8 or more loci, sequences with only 25% of the sites led to lower precision (Fig. 5b). When we compared the accuracy of 7 sampling strategies of loci and sequence length that represented the same total sampling effort of base pairs, we found that in 6 of them, the strategy that sampled more of the shorter loci produced more accurate species trees than analyses with fewer but longer loci (Table 2).

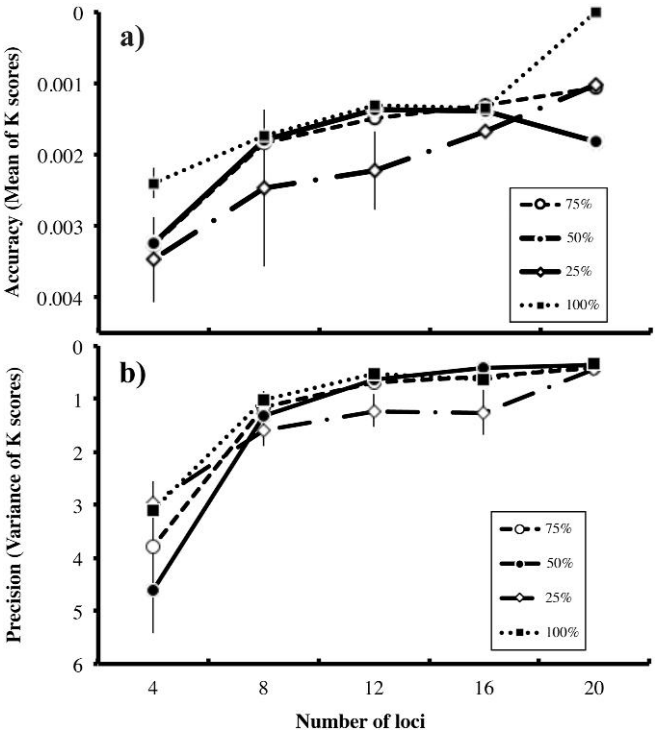

FIGURE 5. Accuracy (a) and precision (b) of species trees estimated from different numbers of loci (4, 8, 12, 16, and 20) and different numbers of base pairs (100%, 75%, 50%, and 25% of full sequences).

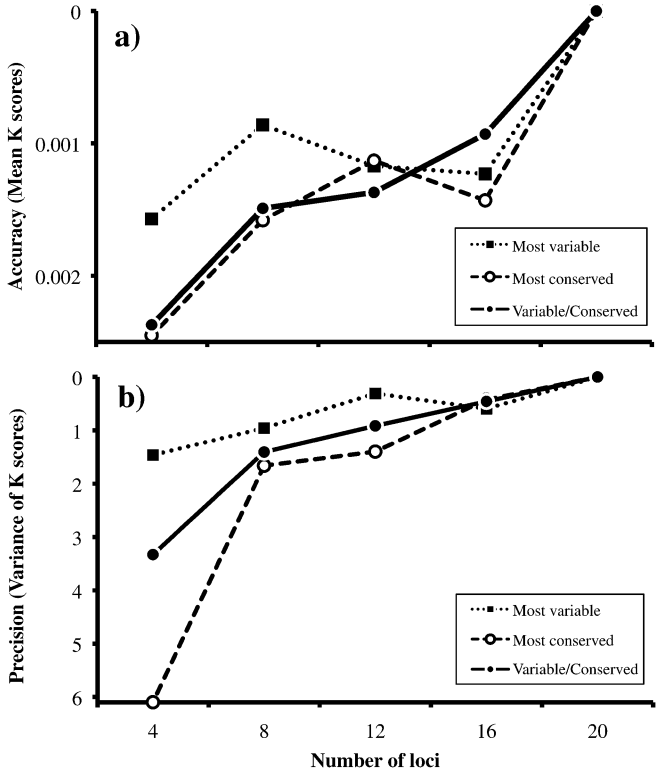

FIGURE 6. Accuracy (a) and precision (b) of species trees estimated from different numbers of loci (4, 8, 12, 16, and 20) and different locus combinations based on their proportion of variable sites (see Table 1).

Locus "Informativeness"

Sets of loci with different sequence variation generated species trees with roughly the same accuracy when 12 and 16 loci were included in the analyses, but when only 4 and 8 loci were analyzed, performance was better with the MV loci (Fig. 6a). An increase of precision with number of loci was clear for all 3 sampling designs of loci, and consistently, the set of MV loci gave more precise estimates, the set of MC loci gave the least precise estimates, and the mixed set of loci produced an expected intermediate pattern (Fig. 6b). When the discordance of gene trees was used as an indicator of locus "informativeness," the LD loci had the highest accuracy for all numbers of loci followed by the MD loci, and the ML loci had the lowest accuracy (Fig. 7a). However, precision was similar among these groups of loci with different levels of discordance with the species tree (Fig. 7b).

The simulation results that compared accuracy of species trees using 3 sets of loci show the same pattern to that observed in the analysis of our empirical data. The mix of heterogeneous gene trees (LD and MD) resulted in the most inaccurate species trees, whereas the LD gene tree set produced the most accurate species tree and the MD gene tree set was intermediate. The distribution of gene trees in "tree space" shows that the LD group occupies a central portion, the MD is restricted to the periphery, and the mix is evenly distributed across this multidimensional space (online Appendix 7, Dryad DOI:10.5061/dryad.60g211t1).

TABLE 2. Accuracy of alternative sampling strategies with more loci or more base pairs (bp) but with the same total sampling effort (total number of bp)

Sampling strategy		Sampling effort
More loci (*K* score)	More bp (*K* score)	# loci × # bp
75%, 16 loci (0.001315)	**100%, 12 loci (0.001302)**	7200
50%, 16 loci (0.001378)	100%, 8 loci (0.001738)	4800
50%, 12 loci (0.001362)	75%, 8 loci (0.001833)	3600
50%, 8 loci (0.001802)	100%, 4 loci (0.002400)	2400
25%, 16 loci (0.001675)	100%, 4 loci (0.002400)	2400
25%, 12 loci (0.002230)	75%, 4 loci (0.003268)	1800
25%, 8 loci (0.002467)	50%, 4 loci (0.003246)	1200

Notes: Sequence length ranged between ~600 bp (100%) and ~150 bp (25%). The strategy with higher accuracy (lower *K* score) is bolded. These comparisons are also shown in Figure 5a.

DISCUSSION

In empirical phylogenetics, researchers often have finite data at hand for estimating species trees because time and/or technical constraints make it difficult to obtain sequence data from a large number of loci. Limited budgets also usually force researchers to sample either more loci or more individuals for a given sequencing

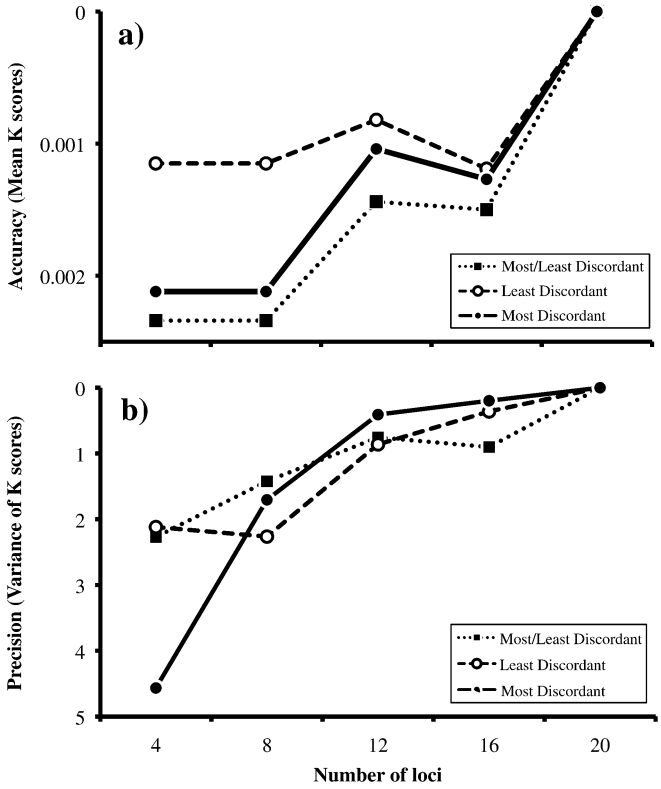

FIGURE 7. Accuracy (a) and precision (b) of species trees estimated from different numbers of loci (4, 8, 12, 16, and 20) and different locus combinations based on their discordance with the species tree (see Table 1).

cost (Felsenstein 2006). Moreover, many potentially useful markers with variability sufficient for resolution of recent species radiations (e.g., ANL) might be poorly informative if very short sequences free of recombination are used in analyses. In this study, we assembled 20 loci for 16 species, which represents a large data set compared with sample sizes often used in empirical studies of species trees (Belfiore et al. 2008; Linnen and Farrell 2008; Liu et al. 2008; Espregueira-Themudo et al. 2009; Leache 2009; Kubatko and Meng 2010; Kubatko et al. 2011). These data coverage allowed us to evaluate the performance of one species-tree method with decreasing number of loci, individuals, and sequence lengths. We measured performance in terms of both accuracy and precision because higher precision is correlated with higher overall support of the species tree, which is of interest for evaluating confidence of inferred trees in empirical phylogenetics. We found an increase in accuracy and precision with the number of loci and, for any number of loci, accuracy declined when using only one individual per species or 25% of the full sequence length (~147 bp/locus). Locus "informativeness" was an important factor when using a limited number of loci, but it was increasingly irrelevant with more loci and accuracy decreased with higher gene-tree heterogeneity.

Performance

Rather than compare performance of different species-tree methods, here we focused on one method that relaxes many assumptions of other methods (Heled and Drummond 2010). Our choice for a fully probabilistic method was based on findings that these methods usually outperform summary-statistic methods (e.g., Liu, Yu, Pearl, and Edwards 2009). Further, *BEAST has been shown to outperform BEST (Heled and Drummond 2010), which in turn is more accurate than STEM and concatenation over a range of divergence times, population sizes, and species-tree topologies (Leache and Rannala 2011). However, summary-statistic methods are more efficient when dealing with genomic data sets that are beyond computational capabilities for heavily parameterized coalescent methods (Liu, Yu, Pearl, and Edwards 2009; Knowles and Kubatko 2010). One advantage of *BEAST over other methods with practical implications is the possibility of estimating the root position without using outgroups, which sometimes are difficult to sample with sequence-based markers due to the annealing specificity of PCR primers (as in this study).

Evaluation of species-tree methods requires simulation studies, but empirical studies are also necessary to assess performance with real data, which include gene genealogies shaped by different but unknown historical and demographic processes. On the other hand, simulation studies use a common, simple evolutionary model for all loci, and sometimes data set sizes are unrealistically large at least for Sanger sequencing techniques (i.e., 100 loci in Leache and Rannala (2011); 6400 bp/locus in Heled and Drummond (2010)). In addition, more empirical studies based on realistic data and divergence scenarios are needed to apply recommendations about sampling strategies derived from theoretical and simulation studies (Castillo-Ramirez et al. 2010; Knowles 2010). Another important contribution of this study was our measurement of performance based on topology and also branch lengths using the *K* scores since most studies to date have quantified accuracy in terms of topology only (Liu and Edwards 2009; McCormack et al. 2009; Leache and Rannala 2011).

Our evaluation of species-tree inference using a coalescent-based method is valid as long as gene-tree heterogeneity was a result of incomplete lineage sorting only. When including multiple individuals for reconstruction of species trees, sampling from phylogeographic lineage boundaries might obfuscate phylogenetic signal due to gene flow between species (Leache 2009). Our geographic sampling strategy probably minimized the impact of gene flow as a source of discordance because we avoided regions of known or suspected contact zones. However, the distributions of several of the species in the *L. darwinii* group are not fully known, and hybridization/introgression of mtDNA is known in some closely related species within the *L. darwinii* complex (Morando et al. 2004). If this has occurred in our samples, it was likely restricted to closely related or

sister species, which should not lead to major phylogenetic estimation error (Brumfield et al. 2008; Eckert and Carstens 2008).

Number of Loci and Individuals

Performance patterns found with subsampling of individuals and loci in our study mimic simulation results obtained for deep divergence histories. In our analysis, larger gains in performance were obtained with an increase in the number of loci (up to 12 loci when convergence was reached in topology). Besides providing more accurate estimates of species trees (by the criterion applied in this study), a higher number of loci produces more robust estimates of population sizes (Felsenstein 2006), which is a critical parameter for accommodating gene-tree discordance with the species tree in the multispecies coalescent model (Castillo-Ramirez et al. 2010). In contrast, there was an increase in accuracy when more than 1 individual per species was sampled but no further gain between 2 and 3 individuals. In one simulation study, a larger gain in accuracy was obtained with more loci instead of more individuals for deep divergences and a given species sampling effort, but the biggest gains in accuracy for shallow histories were obtained when more individuals were sampled (McCormack et al. 2009). Gains in performance with more sampled loci, but not with more sampled individuals, suggest that our species tree represents a relatively deep diversification history within the *L. darwinii* group. Indeed, the total tree depth of ~14.4 Ma for the ancestral split within the *L. darwinii* group corresponds to a deep tree of $\sim 30N_e$, a value times larger than the maximum tree depth simulated in the previous studies (see McCormack et al. 2009). Multiple individuals are preferred over multiple loci in recent radiations because more GCs are likely to cross the species boundary and coalesce in ancestral populations (Heled and Drummond 2010; Knowles 2010). Consequently, more individuals should be sampled when gene lineages within species have not yet sorted to monophyly, but when species are recovered as well-supported clades in older divergences (as in our gene trees, online Appendix 4), only additional loci contain phylogenetic signal about species relationships (Knowles 2010). However, it should be noted that recognition of monophyletic gene lineages within species may be contingent upon sampling density, and it is possible that with additional sampling, we could have found species that were not recovered as monophyletic in individual gene trees within our study group, especially in species that have diverged recently (e.g., the pair *L. albiceps*–*L. irregularis*).

In addition to very low performance, sampling of only one individual/species caused poor convergence and mixing of Markov chain Monte Carlo (MCMC) chains in *BEAST based on ESS values, suggesting low information content in the data. The impact of using single individual/species could be more serious for *BEAST because this method also estimates population sizes of extant lineages (Heled and Drummond 2010), whereas methods such as BEST only estimate population sizes for ancestral lineages where multiple alleles can coexist (Castillo-Ramirez et al. 2010). Even though there is an increase in performance with more loci (especially in deep divergences) and individuals (at shallow divergences) because of reduced coalescence variance, there is a concomitant increase of mutational variance with trade-offs for the relative gains of increased sampling effort (Huang et al. 2010). For example, when sampling additional individuals without increasing the information content of the data, the search through a tree space with more alternative topologies becomes more difficult and leads to more uncertainty in gene and species trees (Huang et al. 2010).

Sequence Length

Our subsampling of base pairs to assess the effect of sequence length shows that sequences ~150 bp long are probably too short for accurate estimates of species trees using our data. Simulations have found larger gains in accuracy with loci of 500 versus 250 bp, but there were no additional improvements with loci longer than 500 bp, implying that this number was the optimal locus length for the simulated conditions (Castillo-Ramirez et al. 2010). These simulations are consistent with the substantial decrease in accuracy (and precision) that we found when using 25% of original sequences (~147 bp/locus) but convergence in accuracy when analyzing 50% of sites or more (>295 bp/locus). In addition, Castillo-Ramirez et al. (2010) suggested that increasing locus length is a better strategy than sequencing additional loci for a given total number of base pairs with BEST, but, in contrast, our results with *BEAST support the strategy of sampling more loci instead of increasing locus length. Even though the number of loci (2–24 loci) and sequence length (250–1000 bp) simulated were similar to the ranges subsampled in our study (4–20 loci and 150–600 bp), differences in optimal sampling effort could be contingent on the species tree and the information content of the data. These comparisons suggest that, for a given speciation history, there may exist a minimum threshold in sequence length below which the mutational variation is too low for robust estimation of gene trees, and this results in inaccurate and poorly supported species trees. The impact of mutational variance could potentially be reduced by methods that incorporate gene-tree uncertainty (i.e., BEST and *BEAST), although these heavily parameterized methods might perform poorly with limited genetic variation (Huang et al. 2010). However, recent simulations have shown that BEST is more accurate than other methods based on point estimates of gene trees when there is low genetic variation (Leache and Rannala 2011).

Locus "Informativeness"

Another sampling dimension relevant in empirical phylogenetics but that has been poorly explored in

simulation studies is the relative information content of loci (Knowles 2009, 2010). In simulations, all loci usually have the same length and a common, sometimes simplified, substitution model is employed that reduces the rate variation across loci. Here, we evaluated performance with loci that differed in variability, length, substitution model, and discordance with the species tree (Table 1). As expected, our results show reduced performance with conserved loci, but this improved with MV loci and was highest when the MV loci were used. These results are intuitive and agree with simulations demonstrating that the low number of informative sites is the most relevant factor decreasing accuracy under some simulation conditions (Castillo-Ramirez et al. 2010), probably as a result of limited phylogenetic signal for estimating well-supported gene trees (McCormack et al. 2009; Huang et al. 2010). Because both higher number and "informativeness" of loci increased performance, inclusion (to increase quantity) versus exclusion (to increase "informativeness") of a locus can be justified depending on which strategy provides a larger gain in performance. However, excluding conserved loci could impact estimates of branch lengths and population sizes due to ascertainment bias (Knowles 2010), which results from discarding loci with low-frequency alleles in the population because of their apparent low variability in the sampled data (Nielsen 2004; Rosenblum and Novembre 2007; Guillot and Foll 2009).

After the number of informative sites, the next most relevant factor impacting the accuracy of species trees in simulations analyzed with BEST was gene-tree heterogeneity (Castillo-Ramirez et al. 2010). In this empirical data set, we grouped loci with varying amounts of discordance with the species tree and found that the most heterogeneous mix of discordant and concordant loci had the lowest accuracy, the LD loci were the most accurate, and the MD loci had intermediate accuracy. In addition, when we simulated gene trees from one fixed species tree and sampled these gene trees in groups of low, high, and mixed discordance, we obtained the same patterns of accuracy as found in our empirical data, suggesting that this is probably a more general phenomenon (online Appendix 7). The lowest accuracy of the mixed group seems counterintuitive, but the more even distribution of this group across tree space, in contrast to the smaller portions occupied by the other two groups, suggests that increasing incongruence among gene trees might lead to higher uncertainty in species-tree estimates (Fig. 4 in online Appendix 7). High levels of gene-tree heterogeneity are common not only in recent species radiations but also when short branches in the species tree generate frequent AGTs, and consequently, more loci are required to estimate species trees accurately (Knowles 2010). Even though our species tree seems to reflect a deep speciation history, short internodes might be responsible for the high degree of heterogeneity observed in our gene trees since none of them matches the species-tree topology (online Appendix 4).

Systematics

Our recovered species tree for the *L. darwinii* group was robust and shows a stable topology in comparison with *BEAST analyses using different priors, models, and MCMC settings and also in comparison with a BEST analysis. In addition, analyses with 19 loci show that the 2 MV loci (cyt *b* and B9G) did not have a major influence on the reference tree estimated with all 20 loci. Previous phylogenetic studies of the *L. darwinii* group recovered clades similar to Clades A and B of our species tree but with some differences in species composition. All previous studies found support for the "*ornatus*" clade (nested within Clade A in Fig. 3) that includes *L. albiceps*, *L. irregularis*, *L. ornatus*, *L. lavillai*, *L. calchaqui*, and *L. crepuscularis*. In addition, these studies also grouped *L. darwinii*, *L. laurenti*, *L. grosseorum*, *L. chacoensis*, and *L. olongasta* into the "*grosseorum*" clade (Abdala 2007). These studies differed in the kind of data used to infer the phylogeny and their species sampling of the *L. darwinii* group: Etheridge (2000) used morphological and behavioral characters of 11 species, Schulte et al. (2000) sequenced 3 mtDNA genes plus several tRNAs of 11 species, Morando (2004) analyzed 3 mtDNA and 2 nuclear genes of 12 species, and Abdala (2007) inferred a morphological + molecular phylogeny for 16 species (the same as those used in this study) (online Appendix 8, Dryad DOI:10.5061/dryad.60g211t1). Our species tree placed the well-supported (*L. quilmes*, *L. espinozai*) clade as sister to the "*ornatus*" clade within Clade A and the well-supported (*L. koslowskyi*, *L. abaucan*) clade within Clade B (Fig. 3), whereas these species were often placed outside of Clades A and B in the previous studies. The relationship of *L. uspallatensis* is ambiguous in our analyses with unphased (to Clade A) or phased data (to Clade B) as well as among previous studies (grouped with Clade B or outside Clades A and B) (online Appendix 8).

The major difference is the placement of *L. chacoensis* within Clade A in our species tree, which was always assigned to Clade B in the previous studies (online Appendix 8). This conflicting topology might be a result of shared ancestral polymorphisms of *L. grosseorum* with Clade B, which would bias the concatenated analyses used in the previous studies that do not account for the process of incomplete lineage sorting. In addition, *L. chacoensis* is morphologically more similar to members of the *L. wiegmannii* group, an outgroup of the *L. darwinii* group (Abdala 2007), which could also have impacted the morphology-based or combined morphological–molecular analyses.

From a biogeographic perspective, our species tree implies a clear association between clades and ecoregions. Clade B occurs mostly in the Monte Desert of south-central and west-central Argentina at lower altitudes (Etheridge 1993). Within Clade A, the *ornatus* clade occupies the Puna and Prepuna of northwestern Argentina at higher altitudes (Abdala 2007). Sister to the *ornatus* clade, the (*L. espinozai*, *L. quilmes*) clade inhabits the Prepun—Monte ecotone and the

northernmost region of the Monte Desert, respectively (see maps in Roig-Junent et al. 2001; Abdala 2005). The species external to these clades occur in the isolated Uspallata–Calingasta valley in west-central Argentina (*L. uspallatensis*) and the Chaco lowlands of central and northern Argentina (*L. chacoensis*). In addition, the topology of our species tree is also consistent with the natural history of the group (Abdala and Diaz Gomez 2006) with viviparity evolving twice, once in *L. espinozai* and once in the *ornatus* clade (assuming no reversals from viviparity to oviparity). Although our goal was not to address species limits in the group, the virtual lack of genetic divergence and slight morphological differentiation between *L. albiceps* and *L. irregularis* (Lobo and Laurent 1995) suggests that these species warrant further sampling and phylogeographic analysis to elucidate their potential conspecificity.

Conclusions

Diversification in the *L. darwinii* group appears to be old but with episodes of rapid speciation that have resulted in short internal branches in the species tree and high gene-tree heterogeneity. The diversification of this clade represents a unique and specific speciation history, but our results and previous simulation studies consistently suggest that there are optimal ranges of sampling effort for estimating species trees that depend on the speciation history, the kinds of data, and the specific inference method used (Castillo-Ramirez et al. 2010; Huang et al. 2010). The choice of an appropriate species-tree method and sampling design will depend on the data available and an unknown speciation history. Although generalizations about optimal sampling designs will be difficult, a potential approach could consist of the subsampling strategy used in this study to assess convergence of species-tree estimates (topology and branch lengths) and branch support (precision). In addition, one recent study has also suggested the use of simulations to explore optimal sampling efforts and designs for a particular empirical study based on the consistency of the estimated species tree (Knowles 2010).

There are several aspects of the speciation history, the data, and the inference methods that should be further investigated including tree shape (symmetric vs. pectinate) and the variance of branch lengths since both could have substantial effects on species-tree accuracy (McCormack et al. 2009), particularly in the AGT parameter space (Degnan and Rosenberg 2006). Second, the impact of gene flow has not been adequately explored, but it probably is very influential on species-tree methods that only model incomplete lineage sorting, especially if gene flow has occurred deep in the species tree and involved lineages that are not closely related (Eckert and Carstens 2008) and/or if gene flow is distributed unevenly across the species tree (Chung and Ane 2011). Finally, the impact of missing data on species-tree estimation has not been assessed with empirical and simulated data, although this factor is of substantial interest because complete data sets containing many loci and individuals are difficult to obtain in practice.

Supplementary Material

Supplementary material, including data files and/or online-only appendices, can be found in the Dryad data repository (DOI:10.5061/dryad.8m8c0 and DOI:10.5061/dryad.60g211t1).

Funding

A.C. acknowledges financial support from the BYU Department of Biology, the Bean Life Science Museum, and Graduate Studies Office through research, mentoring, and travel scholarships. He also thanks the Society of Systematic Biologists and the Society for the Study of Amphibians and Reptiles for research and travel grants. Major funding for this study was provided by the National Science Foundation (NSF) "Partnership for International Research and Education" award (OISE 0530267) for collaborative research on Patagonian biodiversity, granted to the following institutions (listed alphabetically): BYU, Centro Nacional Patagónico, Dalhousie University, Darwinion Botanical Institute, George Washington University, Universidad Nacional de Cordoba, Universidad Austral de Chile, Universidad Nacional del Comahue, and the Universidad de Concepción, and also by the NSF "Assembling the Tree-of-Life—Deep Scaly" project on squamate reptiles (subaward EF 0334966) to J.W.S. Jr., L.J.A. and M.M. acknowledge several grants under programs PIP & IBol (CONICET) and FONCYT PICT (ANPCYT) (Argentina).

Acknowledgments

We appreciate the valuable comments and suggestions from two anonymous referees, the associate editor E. Jockusch, and the editor-in-chief R. DeBry; all have greatly improved the original manuscript. We also thank A. Leaché, S. V. Edwards, E. Benavides, and F. Werneck for their feedback on earlier drafts. We thank students and postdocs of Sites, Morando, and Avila labs for their assistance and support throughout the completion of this study. A.C. thanks his graduate committee for feedback on his dissertation research/manuscripts. Specimens were collected following local regulations and permits issued by Administración de Parques Nacionales, Direcciones Provinciales de Fauna de Río Negro, Chubut, Neuquen, La Pampa, Mendoza, Catamarca, and Administración de Áreas Protegidas de Neuquen y Mendoza (Argentina).

References

Abdala C.S. 2005. Una nueva especie del género *Liolaemus* perteneciente al complejo *darwinii* (Iguania: Liolaemidae) de la provincia de Catamarca, Argentina. Rev. Esp. Herpetol. 19:5–17.

Abdala C.S. 2007. Phylogeny of the *boulengeri* group (Iguania: Liolaemidae, *Liolaemus*) based on morphological and molecular characters. Zootaxa. 1538:21–33.
Abdala C.S., Díaz Gómez J.M. 2006. A new species of the *Liolaemus darwinii* group (Iguania: Liolaemidae) from Catamarca Province, Argentina. Zootaxa. 1317:21–33.
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.
Avise J.C. 1989. Gene trees and organismal histories: a phylogenetic approach to population biology. Evolution. 43:1192–1208.
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.
Brito P.H., Edwards S.V. 2009. Multilocus phylogeography and phylogenetics using sequence-based markers. Genetica. 135:439–455.
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.
Carstens B.C., Dewey T.A. 2010. Species delimitation using a combined coalescent and information-theoretic approach: an example from North American *Myotis* bats. Syst. Biol. 59:400–414.
Castillo-Ramírez S., Liu L., Pearl D., Edwards S.V. 2010. Bayesian estimation of species trees: a practical guide to optimal sampling and analysis. In: Knowles L.L., Kubatko L.S., editors. Estimating species trees: practical and theoretical aspects. Hoboken (NJ): Wiley-Blackwell. p. 15–33.
Chung Y., Ané C. 2011. Comparing two Bayesian methods for gene tree/species tree reconstruction: simulations with incomplete lineage sorting and horizontal gene transfer. Syst. Biol. 60:261–275.
Degnan J.H., DeGiorgio M., Bryant D., Rosenberg N.A. 2009. Properties of consensus methods for inferring species trees from gene trees. Syst. Biol. 58:35–54.
Degnan J.H., Rosenberg N.A. 2006. Discordance of species trees with their most likely gene trees. PLoS Genet. 2:e68.
Degnan J.H., Rosenberg N.A. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24:332–340.
Drummond A.J., Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7:214.
Eckert A.J., Carstens B.C. 2008. Does gene flow destroy phylogenetic signal? The performance of three methods for estimating species phylogenies in the presence of gene flow. Mol. Phylogenet. Evol. 49:832–842.
Edwards S.V. 2009. Is a new and general theory of molecular systematics emerging? Evolution. 63:1–19.
Espregueira-Themudo G., Wielstra B., Arntzen J.W. 2009. Multiple nuclear and mitochondrial genes resolve the branching order of a rapid radiation of crested newts (*Triturus*, Salamandridae). Mol. Phylogenet. Evol. 52:321–328.
Etheridge R. 1993. Lizards of the *Liolaemus darwinii* complex (Squamata: Iguania: Tropiduridae) in Northern Argentina. Boll. Museo Reg. Sci. Nat. 11:137–199.
Etheridge R. 2000. A review of lizards of the *Liolaemus wiegmannii* group (Squamata, Iguania, Tropiduridae) and a history of morphological change in the sand-dwelling species. Herpetol. Monogr. 14:293–352.
Fan H., Kubatko L.S. 2011. Estimating species trees using approximate Bayesian computation. Mol. Phylogenet. Evol. 59:354–363.
Felsenstein J. 2006. Accuracy of coalescent likelihood estimates: do we need more sites, more sequences, or more loci? Mol. Biol. Evol. 23:691–700.
Gadagkar S.R., Rosenberg M.S., Kumar S. 2005. Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. J. Exp. Zool. B. 304B:64–74.
Gamble T., Bauer A.M., Greenbaum E., Jackman T.R. 2008. Evidence for Gondwanan vicariance in an ancient clade of gecko lizards. J. Biogeogr. 35:88–104.
Guillot G., Foll M. 2009. Correcting for ascertainment bias in the inference of population structure. Bioinformatics. 25:552–554.
Heled J., Drummond A.J. 2010. Bayesian inference of species trees from multilocus data. Mol. Biol. Evol. 27:570–580.
Hillis D.M., Bull J.J. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42:182–192.
Huang H., He Q., Kubatko L.S., Knowles L.L. 2010. Sources of error inherent in species-tree estimation: impact of mutational and coalescent effects on accuracy and implications for choosing among different methods. Syst. Biol. 59:573–583.
Hudson R.R. 2002. Generating samples under a Wright-Fisher neutral model. Bioinformatics. 18:337–338.
Knowles L.L. 2009. Estimating species trees: methods of phylogenetic analysis when there is incongruence across genes. Syst. Biol. 58: 463–467.
Knowles L.L. 2010. Sampling strategies for species tree estimation. In: Knowles L.L., Kubatko L.S., editors. Estimating species trees: practical and theoretical aspects. Hoboken (NJ): Wiley-Blackwell. p. 163–73.
Knowles L.L., Kubatko L.S. 2010. Estimating species trees: an introduction to concepts and models. In: Knowles L.L., Kubatko L.S., editors. Estimating species trees: practical and theoretical aspects. Hoboken (NJ): Wiley-Blackwell. p. 1–14.
Kubatko L.S. 2009. Identifying hybridization events in the presence of coalescence via model selection. Syst. Biol. 58:478–488.
Kubatko L.S., Carstens B.C., Knowles L.L. 2009. STEM: species tree estimation using maximum likelihood for gene trees under coalescence. Bioinformatics. 25:971–973.
Kubatko L.S., Degnan J.H. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56: 17–24.
Kubatko L.S., Gibbs H.L., Bloomquist E.W. 2011. Inferring species-level phylogenies and taxonomic distinctiveness using multilocus data in *Sistrurus* rattlesnakes. Syst. Biol. 60:393–409.
Kubatko L.S., Meng C. 2010. Accommodating hybridization in a multilocus phylogenetic framework. In: Knowles L.L., Kubatko L.S., editors. Estimating species trees: practical and theoretical aspects. Hoboken (NJ): Wiley-Blackwell. p. 99–113.
Kuhner M.K., Felsenstein J. 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. Mol. Biol. Evol. 11:459–468.
Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J., Higgins D.G. 2007. Clustal W and Clustal X version 2.0. Bioinformatics. 23:2947–2948.
Leaché A.D. 2009. Species tree discordance traces to phylogeographic clade boundaries in North American fence lizards (*Sceloporus*). Syst. Biol. 58:547–559.
Leaché A.D., Rannala B. 2011. The accuracy of species tree estimation under simulation: a comparison of methods. Syst. Biol. 60:126–137.
Linnen C.R., Farrell B.D. 2008. Comparison of methods of species-tree inference in the sawfly genus *Neodiprion* (Hymenoptera: Diprionidae). Syst. Biol. 57:876–890.
Liu L. 2008. BEST: Bayesian estimation of species trees under the coalescent model. Bioinformatics. 24:2542–2543.
Liu L., Edwards S.V. 2009. Phylogenetic analysis in the anomaly zone. Syst. Biol. 58:452–460.
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.
Liu L., Yu L., Kubatko L., Pearl D.K., Edwards S.V. 2009. Coalescent methods for estimating phylogenetic trees. Mol. Phylogenet. Evol. 53:320–328.
Liu L., Yu L., Pearl D.K., Edwards S.V. 2009. Estimating species phylogenies using coalescence times among sequences. Syst. Biol. 58:468–477.
Lobo F., Espinoza R.E., Quinteros S. 2010. A critical review and systematic discussion of recent classification proposals for liolaemid lizards. Zootaxa. 2549:1–30.
Lobo F., Laurent R.F. 1995. Un nouveau *Liolaemus* Andin (Tropiduridae). Rev. Fr. Aquariol. 22:107–116.
Maddison W.P. 1997. Gene trees in species trees. Syst. Biol. 46: 523–536.
Maddison W.P., Maddison D.R. 2010. Mesquite: a modular system for evolutionary analysis. Version 2.73. Available from: http://mesquiteproject.org.

Martin D.P., Williamson C., Posada D. 2005. RDP2: recombination detection and analysis from sequence alignments. Bioinformatics. 21:260–262.

McCormack J.E., Huang H., Knowles L.L. 2009. Maximum likelihood estimates of species trees: how accuracy of phylogenetic inference depends upon the divergence history and sampling design. Syst. Biol. 58:501–508.

Morando M. 2004. Sistemática y Filogenia de Grupos de Especies de los Géneros *Phymaturus y Liolaemus* (Squamata: Tropiduridae: Liolaemidae). Tucumán: Universidad Nacional de Tucumán. p. 249.

Morando M., Avila L.J., Baker J., Sites J.W. Jr. 2004. Phylogeny and phylogeography of the *Liolaemus darwinii* complex (Squamata: Liolaemidae): evidence for introgression and incomplete lineage sorting. Evolution. 58:842–861.

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. 52:159–185.

Mossel E., Roch S. 2010. Incomplete lineage sorting: consistent phylogeny estimation from multiple loci. IEEE ACM Trans. Comput. Biol. 7:166–171.

Nielsen R. 2004. Population genetic analysis of ascertained SNP data. Hum. Genom. 1:218–224.

Noonan B., Yoder A.D. 2009. Anonymous nuclear markers for Malagasy plated lizards (*Zonosaurus*). Mol. Ecol. Resour. 9:402–404.

Page R.D.M. 1998. GeneTree: comparing gene and species phylogenies using reconciled trees. Bioinformatics. 14:819–820.

Pamilo P., Nei M. 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5:568–583.

Portik D.M., Wood P.L., Grismer J.L., Stanley E.L., Jackman T.R. 2011. Identification of 104 rapidly-evolving nuclear protein-coding markers for amplification across scaled reptiles using genomic resources. Conserv. Genet. Resour. doi: 10.1007/s12686-011-9460-1.

Posada D. 2008. jModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25:1253–1256.

Rambaut A., Drummond A.J. 2007. Tracer v1.4. Available from: http://beast.bio.ed.ac.uk/Tracer.

Rambaut A., Grassly N.C. 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. Comput. Appl. Biosci. 13:235–238.

Rannala B., Yang Z. 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. Genetics. 164:1645–1656.

Reyes-Velasco J., Mulcahy D.G. 2010. Additional taxonomic remarks on the genus *Pseudoleptodeira* (Serpentes: Colubridae) and the phylogenetic placement of "*P. uribei*". Herpetologica. 66:99–110.

Roig-Juñent S., Flores G., Claver S., Debandi G., Marvaldi A. 2001. Monte Desert (Argentina): insect biodiversity and natural areas. J. Arid Environ. 47:77–94.

Rosenblum E.B., Novembre J. 2007. Ascertainment bias in spatially structured populations: a case study in the eastern fence lizard. J. Hered. 98:331–336.

Saint K.M., Austin C.C., Donnellan S.C., Hutchinson M.N. 1998. C-mos, a nuclear marker useful for Squamate phylogenetic analysis. Mol. Phylogenet. Evol. 10:259–263.

Schulte J.A., Macey J.R., Espinoza R.E., Larson A. 2000. Phylogenetic relationships in the iguanid lizard genus *Liolaemus*: multiple origins of viviparous reproduction and evidence for recurring andean vicariance and dispersal. Biol. J. Linn. Soc. 69:75–102.

Slowinski J.B., Page R.D.M. 1999. How should species phylogenies be inferred from sequence data? Syst. Biol. 48:814–825.

Soria-Carrasco V., Talavera G., Igea J., Castresana J. 2007. The K tree score: quantification of differences in the relative branch length and topology of phylogenetic trees. Bioinformatics. 23:2954–2956.

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

Thomson R.C., Wang I.J., Johnson J.R. 2010. Genome-enabled development of DNA markers for ecology, evolution and conservation. Mol. Ecol. 19:2184–2195.

Townsend T.M., Alegre R.E., Kelley S.T., Wiens J.J., Reeder T.W. 2008. Rapid development of multiple nuclear loci for phylogenetic analysis using genomic resources: an example from squamate reptiles. Mol. Phylogenet. Evol. 47:129–142.

Waltari E., Edwards S.V. 2002. Evolutionary dynamics of intron size, genome size, and physiological correlates in archosaurs. Am. Nat. 160:539–552.

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

APPENDIX

TABLE A1. List of specimens sequenced for this study

Specimen	Locality	Coordinates	Province
Liolaemus abaucan			
LJAMM 2359	1	27°26′50″S, 67°40′44″W	Ruta Provincial 36, 16 Km South Palo Blanco, Tinogasta, Catamarca
LJAMM 2362	1	27°26′50″S, 67°40′44″W	Ruta Provincial 36, 16 Km South Palo Blanco, Tinogasta, Catamarca
LJAMM 2371	1	27°26′50″S, 67°40′44″W	Ruta Provincial 36, 16 Km South Palo Blanco, Tinogasta, Catamarca
Liolaemus albiceps			
LJAMM 12040	2	24°59′57″S, 66°09′15″W	Road toward Nevado del Acay, 5 km South Estacion Muñano, from Ruta Nacional 51, Rosario de Lerma, Salta
LJAMM 12046	2	24°59′57″S, 66°09′15″W	Road toward Nevado del Acay, 5 km South Estacion Muñano, from Ruta Nacional 51, Rosario de Lerma, Salta
LJAMM 2646	3	24°27′02″S, 65°57′05″W	Santa Rosa de Tastil, Rosario de Lerma, Salta
Liolaemus calchaqui			
LJAMM 12834	4	26°22′45″S, 65°43′54″W	Ruta Provincial 352, 38.3 km West Hualinchay, Cumbres Calchaquies, Trancas, Tucumán
LJAMM 12837	4	27°22′45″S, 65°43′54″W	Ruta Provincial 352, 38.3 km West Hualinchay, Cumbres Calchaquies, Trancas, Tucumán
LJAMM 12842	4	28°22′45″S, 65°43′54″W	Ruta Provincial 352, 38.3 km West Hualinchay, Cumbres Calchaquies, Trancas, Tucumán
Liolaemus chacoensis			
LJAMM 10649	5	32°53′19″S, 66°49′04″W	1 km NE La Calera, 8.7 km SW Ruta Nacional 147, Sierra del Gigante, Belgrano, San Luis
LJAMM 10854	6	30°38′20″S, 67°22′38″W	Ruta Provincial 511, 7.6 km East San Agustin del Valle Fertil, Valle Fertil, San Juan
LJAMM 5042	7	30°32′57″S, 66°55′40″W	Ruta Provincial 27, 41.6 Km North El Portezuelo; 6.6 Km North San Ramon, Gral. Juan F. Quiroga, La Rioja
Liolaemus crepuscularis			
LJAMM 12635	8	27°21′58″S, 66°22′25″W	Puesto Lopez, Ruta Provincial 47, 2 km South Mina Capillitas, Andalgalá, Catamarca
LJAMM 12642	8	27°21′58″S, 66°22′25″W	Puesto Lopez, Ruta Provincial 47, 2 km South Mina Capillitas, Andalgalá, Catamarca
LJAMM 12644	8	28°21′58″S, 66°22′25″W	Puesto Lopez, Ruta Provincial 47, 2 km South Mina Capillitas, Andalgalá, Catamarca
Liolaemus darwinii			
LJAMM 10582	9	37°04′29″S, 67°47′07″W	Ruta Provincial 16, 23.6 km West junction with Ruta Nacional 151, Puelén, La Pampa
LJAMM 11022	10	42°46′00″S, 65°03′00″W	Puerto Madryn, Biedma, Chubut
LJAMM 5104	11	36°08′19″S, 68°17′23″W	Ruta Provincial 190, 2 Km North Agua Escondida, Malargüe, Mendoza
Liolaemus espinozai			
LJAMM 12666	12	27°03′19″S, 66°11′51″W	Ruta Provincial 47, 21 km South El Arenal, Santa Maria, Catamarca
LJAMM 12668	12	28°03′19″S, 66°11′51″W	Ruta Provincial 47, 21 km South El Arenal, Santa Maria, Catamarca
LJAMM 4338	13	27°07′30″S, 66°13′03″W	Ruta Provincial 47, 20 Km South Punta de Balasto, Campo Arenal, Santa Maria, Catamarca
Liolaemus grosseorum			
LJAMM 4019	14	35°17′09″S, 68°41′52″W	Ruta Provincial 180, 30 Km South El Nihuil, San Rafael, Mendoza
LJAMM 4046	15	36°37′17″S, 68°36′38″W	Ruta Provincial 180, 28.1 Km North south entrance to La Matancilla, Malargüe, Mendoza
LJAMM 7825	16	38°13′49″S, 68°57′36″W	Ruta Provincial 8, 23 km North Añelo, Añelo, Neuquén
Liolaemus irregularis			
LJAMM 12795	17	24°13′13″S, 66°16′59″W	Ruta Nacional 51, 6 km East San Antonio de los Cobres, San Antonio de los Cobres, Salta
LJAMM 12798	17	25°13′13″S, 66°16′59″W	Ruta Nacional 51, 6 km East San Antonio de los Cobres, San Antonio de los Cobres, Salta
LJAMM 2629	18	24°12′02″S, 66°24′04″W	5 Km NW San Antonio de los Cobres, Paraje Pompeya, Los Andes, Salta
Liolaemus koslowskyi			
LJAMM 4159	19	28°32′09″S, 67°22′21″W	Ruta Nacional 40, Km 657. 9 Km East Pituil, Famatina, La Rioja
LJAMM 4206	20	28°50′19″S, 67°24′47″W	Entrance to Antinaco, 3.8 Km East Ruta Nacional 40, Famatina, La Rioja
LJAMM 5011	21	28°11′35″S, 67°08′03″W	10 Km North Cerro Negro, Belén, Catamarca
Liolaemus laurenti			
LJAMM 2334	22	28°14′44″S, 67°27′11″W	Ruta Nacional 40 and La Puerta River, Km 1298, Tinogasta, Catamarca
LJAMM 4160	19	28°32′09″S, 67°22′21″W	Ruta Nacional 40, Km 657. 9 Km East Pituil, Famatina, La Rioja
LJAMM 4210	20	28°50′19″S, 67°24′47″W	Entrance to Antinaco, 3.8 Km East Ruta Nacional 40, Famatina, La Rioja
Liolaemus lavillai			
LJAMM 12735	23	24°36′25″S, 66°11′41″W	Ruta Nacional 40, 13 km North La Toma, 2 km South El Saladillo, La Poma, Salta
LJAMM 12812	24	25°14′14″S, 65°54′02″W	Cumbres del Obispo, Ruta Provincial 33, 46.8 km East Cachi, Parque Nacional Los Cardones, Salta
LJAMM 12815	24	26°14′14″S, 65°54′02″W	Cumbres del Obispo, Ruta Provincial 33, 46.8 km East Cachi, Parque Nacional Los Cardones, Salta
Liolaemus olongasta			
LJAMM 10751	25	31°14′20″S, 68°39′04″W	Ruta Nacional 40, Matagusanos, Ullum, San Juan
LJAMM 10783	26	30°13′43″S, 68°19′35″W	Ruta Provincial 150 and Rio Huaco, Jachal, San Juan
LJAMM 10821	27	29°41′17″S, 68°01′41″W	Ruta Nacional 76, 16.2 km South Pagancillo, 42 km South Villa Union, Felipe Varela, La Rioja

(Continued)

TABLE A1. *Continued*

Specimen	Locality	Coordinates	Province
Liolaemus ornatus			
LJAMM 12019	28	23°16′38″S, 65°49′09″W	Ruta Nacional 40, 58.8 km South junction Ruta nacional 9, between Agua de Castilla and Quebraleña, Cochinoca, Jujuy
LJAMM 12021	28	23°16′38″S, 65°49′09″W	Ruta Nacional 40, 58.8 km South junction Ruta nacional 9, between Agua de Castilla and Quebraleña, Cochinoca, Jujuy
Liolaemus quilmes			
LJAMM 12713	29	25°36′53″S, 66°11′43″W	Ruta Nacional 40, 10 km North Angastaco, between La Arcadia and El Carmen, San Carlos, Salta
LJAMM 4346	30	25°42′19″S, 66°07′42″W	Ruta Nacional 40, 6 Km South Angastaco, San Carlos, Salta
LJAMM 4404	31	25°52′44″S, 65°56′38″W	Ruta Nacional 40, 2.7 Km North San Carlos, San Carlos, Salta
Liolaemus uspallatensis			
LJAMM 12500	32	32°23′26″S, 69°23′20″W	Los Tambillos, Ruta Nacional 149, 24 km North de Uspallata, in front of Estancia Los Tambillos, Las Heras, Mendoza
LJAMM 12506	33	31°25′06″S, 69°26′31″W	Ruta Provincial 406, 4 km North Tamberias, Calingasta, San Juan
LJAMM 12630	34	30°52′43″S, 69°25′12″W	Ruta Provincial 412, junction with Ruta Provincial 425, 26 km North Villa Nueva, Calingasta, San Juan