Available online at www.sciencedirect.com
ScienceDirect
Molecular Phylogenetics and Evolution 47 (2008) 700–716
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Phylogenetic relationships and heterogeneous evolutionary processes among phrynosomatine sand lizards (Squamata, Iguanidae) revisited

James A. Schulte II *, Kevin de Queiroz

Division of Amphibians and Reptiles, Smithsonian Institution, P.O. Box 37012, MRC 162, Washington, DC 20013-7012, USA

Received 8 August 2007; revised 3 December 2007; accepted 11 January 2008
Available online 24 January 2008

Abstract

Phylogenetic analyses of DNA sequences were conducted to evaluate four alternative hypotheses of phrynosomatine sand lizard relationships. Sequences comprising 2871 aligned base pair positions representing the regions spanning *ND1–COI* and *cyt b-tRNA*^{Thr} of the mitochondrial genome from all recognized sand lizard species were analyzed using unpartitioned parsimony and likelihood methods, likelihood methods with assumed partitions, Bayesian methods with assumed partitions, and Bayesian mixture models. The topology (*Uma*, (*Callisaurus*, (*Cophosaurus*, *Holbrookia*))) and thus monophyly of the "earless" taxa, *Cophosaurus* and *Holbrookia*, is supported by all analyses. Previously proposed topologies in which *Uma* and *Callisaurus* are sister taxa and those in which *Holbrookia* is the sister group to all other sand lizard taxa are rejected using both parsimony and likelihood-based significance tests with the combined, unpartitioned data set. Bayesian hypothesis tests also reject those topologies using six assumed partitioning strategies, and the two partitioning strategies presumably associated with the most powerful tests also reject a third previously proposed topology, in which *Callisaurus* and *Cophosaurus* are sister taxa. For both maximum likelihood and Bayesian methods with assumed partitions, those partitions defined by codon position and tRNA stem and nonstems explained the data better than other strategies examined. Bayes factor estimates comparing results of assumed partitions versus mixture models suggest that mixture models perform better than assumed partitions when the latter were not based on functional characteristics of the data, such as codon position and tRNA stem and nonstems. However, assumed partitions performed better than mixture models when functional differences were incorporated. We reiterate the importance of accounting for heterogeneous evolutionary processes in the analysis of complex data sets and emphasize the importance of implementing mixed model likelihood methods.

Keywords: Phrynosomatinae; *Uma*; *Callisaurus*; *Cophosaurus*; *Holbrookia*; Mixed models; Mixture models; Process heterogeneity; Phylogeny; Mitochondrial DNA

1. Introduction

The phrynosomatine sand lizards are a well-studied clade of squamate reptiles for which several alternative phylogenetic hypotheses have been proposed. The group is composed of 12 currently recognized species that form four mutually exclusive clades traditionally ranked as genera: *Uma* (6 species), *Callisaurus* (1 species), *Cophosaurus* (1 species), and *Holbrookia* (4 species) (Crother, 2000, 2003; Wilgenbusch and de Queiroz, 2000). Although these four clades are well-supported and largely uncontested, the relationships among them have been the subject of considerable debate (reviewed by Wilgenbusch and de Queiroz, 2000).

In an earlier study of sand lizard phylogenetic relationships, Wilgenbusch and de Queiroz (2000) evaluated four previously proposed phylogenetic hypotheses (Fig. 1) using fragments of two mitochondrial genes, *cytochrome b* (~650 bp) and *12S rRNA* (~350 bp). Although these data consistently favored one of the four topologies (Topology I in Fig. 1), support values for the two critical nodes distin-

* Corresponding author. Present address: Department of Biology, 177 Clarkson Science Center, MRC 5805, Clarkson University, Potsdam, NY 13699-5805, USA. Fax: +1 315 268 7118.
E-mail address: jschulte@clarkson.edu (J.A. Schulte II).

1055-7903/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2008.01.010

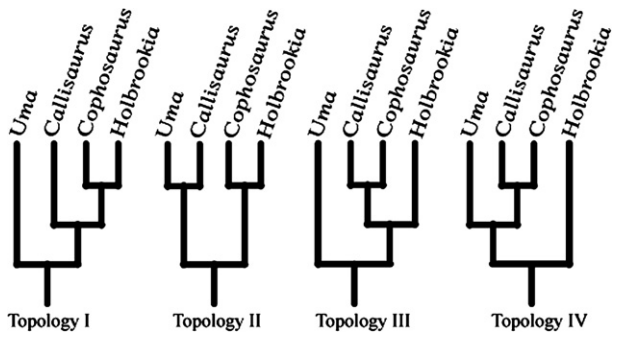

Fig. 1. Alternative hypotheses of higher-level relationships among the phrynosomatine sand lizards (after Wilgenbusch and de Queiroz, 2000). Topology I (Savage, 1958; Cox and Tanner, 1977; Etheridge and de Queiroz, 1988; de Queiroz, 1989, 1992; Reeder and Wiens, 1996; ChangChien, 1996); topology II (Mittleman, 1942; Smith, 1946); topology III (Norris, 1958; Earle, 1961, 1962); topology IV (Axtell, 1958; Clarke, 1965; Adest, 1978).

guishing the favored topology from alternative hypotheses were unimpressive, and the three alternative topologies could not be rejected under the conventional $P \leqslant 0.05$ criterion of significance. Assuming patterns of nucleotide sequence variation similar to those observed in the cytochrome *b* (cyt *b*) gene, Wilgenbusch and de Queiroz (2000) estimated that from 1037 to 2586 base pairs of sequence data would be necessary to reject the various alternative topologies. Therefore, we undertook a project to sequence two fragments of mitochondrial DNA encoding the nearly complete cyt *b* gene (~1130 bp; Kumazawa et al., 1998) and a genomic segment extending from the gene encoding subunit one of nicotinamide adenine dinucleotide dehydrogenase (ND1) through the gene encoding subunit 1 of cytochrome *c* oxidase (COI) (~1700 bp; Macey et al., 1997) in an attempt to distinguish between the alternative hypotheses with a higher degree of confidence. In the process, we also investigated the effects of different methods for accommodating evolutionary heterogeneity among subsets of the data.

Several recent studies (Brandley et al., 2005; Castoe and Parkinson, 2006; Nylander et al., 2004; Strugnell et al., 2005) have shown that partitioning data and using partition specific parameter estimates are associated with higher probabilities of the data, and this result has been shown to hold for previously collected nucleotide sequence data on sand lizards (Wilgenbusch and de Queiroz, 2000). Alternatively, heterogeneity can be accommodated without partitioning the data by applying several models, each weighted by its probability, to every site (Pagel and Meade, 2004; Pagel and Meade, 2005). Here, we use the term *mixed model* (=*assumed partitions* = *a priori partitions*) for analyses that assume predefined data partitions (models) prior to conducting the analysis, with a different model (including parameter estimates) applied to each partition. In contrast, *mixture models* accommodate heterogeneity across sites by applying several models to each site, where the likelihood of a given site is the sum of the likelihoods under the different models, each weighted by its probability as estimated from the data (Pagel and Meade, 2004). In this paper, we analyzed our data using both different general approaches (mixed models versus mixture models) and different strategies within the general approaches (different assumed partitions; different numbers of *Q* matrices), which permitted us to compare the ability of these different approaches and strategies to accommodate heterogeneity in our data.

2. Materials and methods

2.1. Taxon and character sampling

Frozen liver or muscle tissue of 26 phrynosomatine sand lizards, representing all 12 currently recognized species, was used in DNA extractions (Appendix A). These same samples were used by Wilgenbusch and de Queiroz (2000) and cover a wide range of variation within and among sand lizard species with regard to geographic distribution, morphology, and taxonomy. We also used the same outgroup specimens sequenced by Wilgenbusch and de Queiroz (2000): *Phrynosoma platyrhinos*, *P. hernandesi*, *Sceloporus jarrovii, Urosaurus ornatus,* and *Uta stansburiana.* Wilgenbusch and de Queiroz (2000) reported sequencing CAS 174377 representing the taxon *Holbrookia maculata thermophila*; however, our sequences for the sample from CAS 174377 revealed that it is *H. m. flavilenta*. The correct voucher specimen for the data reported for *H. m. thermophila* appears to be CAS 174400, which also was sequenced by Wilgenbusch and de Queiroz (2000) and has an identical cyt *b* sequence to that reported by those authors and was therefore used as the representative of *H. m. thermophila* in the present study. In addition, the museum number for the specimen of *H. m. flavilenta* sequenced by Wilgenbusch and de Queiroz (2000) was published as MVZ 21814; the correct number is MVZ 214814. See Appendix A for museum numbers, localities of voucher specimens from which DNA was extracted, and GenBank accession numbers for DNA sequences.

Genomic DNA was extracted using Qiagen QIAamp tissue kits. Amplification of genomic DNA was conducted in a DNA Engine® (PTC-200™) Peltier Thermal Cycler (MJ Research) using denaturation at 94 °C for 35 s, annealing at 45–55 °C for 35 s, and extension at 72 °C for 150 s for 30–35 cycles. Negative controls were run on all amplifications to check for contamination. Amplified products were visualized on 1.5% SeaKem® agarose gels. Reamplified double-stranded products were purified with AMPure® magnetic beads (Agencourt). Cycle-sequencing reactions were run using the ABI Prism Big Dye Terminator DNA Sequencing Kit (Perkin-Elmer) with denaturation at 96 °C for 10 s, annealing at 50 °C for 10 s, and extension at 60 °C for 4 min for 35–40 cycles. Sequencing reactions were run on an ABI 3100 Genetic Analyzer (Applied Biosystems).

We obtained nucleotide sequences from four protein-coding genes (*ND1, ND2, COI, cyt b*), nine transfer RNA genes ($tRNA^{Ile}$, $tRNA^{Gln}$, $tRNA^{Met}$, $tRNA^{Trp}$, $tRNA^{Ala}$,

tRNA^{Asn}, *tRNA*^{Cys}, *tRNA*^{Tyr}, *tRNA*^{Thr}), and the origin of light-strand replication (O_L) in the mitochondrial DNA. PCR and sequencing primers used in this study are listed in Table 1. DNA sequences were aligned manually, using tRNA secondary structure models and information on codon position. Positions encoding part of *ND1*, complete *ND2*, part of *COI*, and part of *cyt b* were translated to amino acids using MacClade 4.08 (Maddison and Maddison, 2003) for confirmation of alignment. Alignment of protein-coding regions was straightforward, as these regions contained no length variation. Noncoding intergenic regions contained length variation in several instances (see Section 3.1) and gaps were inserted to ensure alignment of flanking coding regions. Alignment of sequences encoding tRNAs was based on secondary structure models (Kumazawa and Nishida, 1993; Macey and Verma, 1997). Secondary structures of tRNAs were inferred from primary structures of the corresponding tRNA genes using these models. Alignment of tRNA stem regions was straightforward; however, several tRNA loop regions were more difficult to align due to significant length variation. Gaps were inserted in these regions to maintain alignment of flanking stem regions by minimizing implied substitutions and indel events, with indels weighted equally to transitions and transversions weighted approximately 4 times indels. A region was considered ambiguously aligned if more than one alternative placement of a gap had the same cost, with equal costs for gap insertions and extensions, while still maintaining alignment of conserved regions of adjacent tRNA stems or coding regions. Gaps were treated as missing data, and regions considered ambiguously aligned were excluded from phylogenetic analyses (see Section 3.1).

2.2. Phylogenetic analyses

Phylogenetic trees were estimated using both parsimony and probabilistic methods. Parsimony analyses were conducted using PAUP* 4.0b10 (Swofford, 2002) using equal weighting of characters and equal costs for state transformations. Optimal trees were estimated using heuristic searches with 1000 replicates of random stepwise addition and tree bisection and reconnection (TBR) branch swapping. Bootstrap resampling (Felsenstein, 1985a) was applied to assess support for individual nodes using 1000 bootstrap replicates and full heuristic searches with 100 replicates of random stepwise addition and TBR branch swapping.

Maximum-likelihood (ML) analyses also were performed using PAUP*. Searches were conducted using a successive approximations approach (Swofford et al., 1996; Sullivan et al., 2005) for (1) the unpartitioned, combined data set, (2) the fragment encoding the genes spanning ND1–COI, (3) and the fragment spanning cyt *b* to

Table 1
Primers used to amplify and sequence the mtDNA *ND1–COI* and *cyt b-tRNA*^{Thr} regions in this study

Primer name	Human position	Gene	Sequence (5′–3′)	Reference
ND1b[a,s]	L4160	ND1	CGATTCCGATATGACCARCT	Kumazawa and Nishida (1993)
ND1f.3[a,s]	L4178a	ND1	CAACTAATACACCTACTATGAAA	Macey et al. (1997)
ND1f.7[a,s]	L3914	ND1	GCCCCATTTGACCTCACAGAAGG	Macey et al. (1998)
ILEf.1[s]	L4221	tRNA^{Ile}	AAGGATTACTTTGATAGAGT	Macey et al. (1997)
ILEf.6[s]	L4221	tRNA^{Ile}	AAGGGNTACTTTGATAGAGT	Schulte et al. (2003)
METf.6[a,s]	L4437	tRNA^{Met}	AAGCTTTCGGGCCCATACC	Macey et al. (1997)
METr.5[a,s]	H4419	tRNA^{Met}	GGTATGAGCCCGATAGCTT	Macey et al. (1997)
ND2f.4[a,s]	L4645	ND2	ACAGAAGCCGCAACAAAATA	Macey et al. (1997)
ND2r.6[a]	H4980	ND2	ATTTTTCGTAGTTGGGTTTGRTT	Macey et al. (1997)
ND2f.14[a,s]	L4882	ND2	TGACAAAAACTAGCCCC	Macey et al. (1999)
ND2f.54[s]	L5263	ND2	TAACYGGATTTTTACCAAAATGAC	This study
ND2f.57[s]	L4833	ND2	CACAYTTTTGACTNCCAGAAGT	Torres-Carvajal et al. (2006)
ND2r.61[a,s]	H4568	ND2	ATGGCTAGTGTGTTTATTTCTA	This study
TRPf.11[s]	L5550	tRNA^{Trp}	AACCAAGAACCTTCAAAGT	This study
TRPf.12[a,s]	L5549	tRNA^{Trp}	AACCAAGRGCCTTCAAAG	Schulte et al. (2003)
ALAf.4[s]	L5638b	tRNA^{Ala}	CTGAATGCAACTCAGACACTTT	Macey et al. (1997)
ASNr.2[a,s]	H5692	tRNA^{Asn}	TTGGGTGTTTAGCTGTTAA	Macey et al. (1997)
CO1r.1[a,s]	H5934a	COI	AGRGTGCCAATGTCTTTGTGRTT	Macey et al. (1997)
CO1r.8[a,s]	H6159	COI	GCTATGTCTGGGGCTCCAATTAT	Weisrock et al. (2001)
CYT*b*f.1[a,s]	L14842	Cyt *b*	CCATCCAACATCTCAGCATGATGAAA	Kocher et al. (1989)
CYT*b*f.2[a,s]	L15172	Cyt *b*	TGAGGACAAATATCCTTCTGAGG	Fu (1999)
CYT*b*r.3[a,s]	H15916	Cyt *b*	GTCTTCAGTTTTTGGTTTACAAGAC	Kocher et al. (1989)
CYT*b*r.5[a,s]	H15402	Cyt *b*	CTTTGTATGARAAGTATGGGTGRAATGG	This study
CYT*b*f.6[s]	L15429	Cyt *b*	CCATTYCACCCATACTTYTCATACAAAG	This study
CYT*b*r.7[s]	H15216	Cyt *b*	ATTCATTCTACTAGGGTTGTTCC	This study
CYT*b*f.8[s]	L15691	Cyt *b*	ACATCAAAACAACGAAGCAC	This study

Their positions in the human mitochondrial genome (Anderson et al., 1981), target genes, sequences, and references are given. H and L designate primers whose extension produces the heavy and light strands, respectively.

[a] Primers used in amplification.

[s] Primers used in sequencing.

$tRNA^{Thr}$. Both hierarchical likelihood ratio tests and the Akaike Information Criterion as implemented in Modeltest v3.7 (Posada and Crandall, 1998) were used to find the best fitting model of sequence evolution for a tree estimated using neighbor joining (NJ). Parameter values estimated on the NJ tree were fixed in initial heuristic searches under the following conditions: (1) Starting trees were obtained via NJ; (2) TBR branch-swapping; (3) reconnection limit set to eight. Optimal tree(s) obtained from this search were used to estimate new parameter values under an identical model. These new parameter values were then fixed in a second search with the same conditions as the initial run. This process was repeated until the same tree and parameter values were obtained in two successive searches. Bootstrap resampling was applied using ML with 100 pseudoreplicates and heuristic searches as above except that successive approximations were not conducted for each pseudoreplicate. Parameter estimates for bootstrap analyses were fixed at the parameter values obtained from the final iteration of the corresponding successive approximations analysis. In our evaluation of branch support strength, we considered a bootstrap value of $\geqslant 95\%$ as strongly supported (Felsenstein and Kishino, 1993), <95 to $\geqslant 70\%$ as moderately supported, and $<70\%$ as weakly supported.

The impact of using partition-specific (="mixed") evolutionary models on phylogeny estimation has been investigated recently using both likelihood (Wilgenbusch and de Queiroz, 2000) and Bayesian methods (Brandley et al., 2005; Castoe et al., 2004, 2005, 2006; Nylander et al., 2004). We used both approaches under the partitioning strategies (assumed or prior data partitions) described in Table 2. For ML analyses, the program mixed model maximum likelihood analysis (mmml), written by James C. Wilgenbusch, uses PAUP* to find an initial tree using neighbor joining and to perform a heuristic likelihood search saving N (in this case 100) best trees for all data (unpartitioned) and for each partition (Table 2). Trees from all data sets (each partition and the unpartitioned data set) were combined into a single treefile removing duplicate topologies in the process. Likelihood scores for these trees were calculated separately for each partition and then summed to yield a likelihood score for each tree under a given partitioning strategy. For each tree, the partitioned likelihood score is the sum of the likelihood scores for each component partition of a particular partitioning strategy on that tree. The tree with the highest partitioned likelihood score is the best (ML) tree for each partitioning strategy. The current implementation of mmml does not conduct bootstrap analyses, so uncertainty of (support for) the various clades implied by this "point" estimate of the phylogeny could not be evaluated. In addition, the most complicated evolutionary model currently implemented in this program is the general time reversible model with rate variation among sites estimated using a gamma distribution ($GTR + \Gamma$).

Bayesian analyses were performed in MrBayes 3.1–3.1.2 (Ronquist and Huelsenbeck, 2003) using a general time-reversible model of sequence evolution with rate variation among sites estimated using a gamma distribution and an estimated proportion of invariant sites ($GTR + I + \Gamma$; Tavaré, 1986) for the assumed (prior) partitioning strategies outlined in Table 2. This model was chosen for comparison to the results of the ML analyses, in which $GTR + I + \Gamma$ was identified as the most appropriate model for several partitions (see Section 3.1—Phylogeny estimation). Bayesian analyses under the various assumed partitions (Table 2) also were performed using $GTR + \Gamma$ (i.e., without I) to compare likelihood scores under this model to those resulting from both ML analyses under the same assumed partitions performed using mmml and mixture model Bayesian analyses performed using BayesPhylogenies (see below).

For each assumed partition analysis with MrBayes a run of 4×10^6 generations and four Markov chains with default heating values was used. Parameter values for the model were estimated from the data and most were initiated with default uniform priors except that branch lengths were unconstrained (no molecular clock) with default exponential priors. Trees and parameter values were sampled every 1000 generations resulting in 4000 saved trees per analysis, of which 2000 were discarded as "burn-in". Stationarity was assessed by plotting the $-\ln L$ per generation in the program Tracer 1.3 (Rambaut and Drummond, 2005) and using the average standard deviation of split frequencies available in MrBayes. If the two simultaneous runs are converging on the same tree, the average standard deviation of split frequencies are expected to approach zero. After confirming that the analysis appeared to reach stationarity, the resultant 4000 trees (2000 from each simultaneous run) were used to calculate Bayesian credibility values (BC) for each branch in a 50% majority-rule consen-

Table 2
Partitioning strategies (assumed partitions) used in this study

Partition strategy	Abbreviation	Number of characters in partition(s)
All data combined	P1	2763
ND1–COI coding region; cyt *b*-tRNAThr coding region	P2	1697; 1066
All protein-coding genes; tRNAs	P3	2192; 571
Codon positions of all protein coding genes; tRNAs	P4	731; 730; 731; 571
ND1; ND2; COI; cyt *b*; tRNAs	P5	84; 1033; 30; 1045; 571
Codon positions of all protein coding genes; tRNA stem regions; tRNA nonstem regions	P6	731; 730; 731; 378; 193

sus tree. Clades with BC $\geqslant$ 95% were considered strongly supported with the caveat that BC may overestimate support for reasons discussed by Alfaro et al. (2003), Douady et al. (2003) and Lewis et al. (2005).

In addition, we analyzed our complete dataset under the phylogenetic mixture model approach using BayesPhylogenies (Pagel and Meade, 2004). By default, BayesPhylogenies assigns uniform priors on a 0–100 interval for the gamma and substitution rate parameters, an exponential prior for branch lengths with a mean of 10, state frequencies drawn from the Dirichlet distribution, and all trees are considered equally likely a priori. These default settings were used in all mixture model analyses. We performed a total of 12 analyses with nQ and $nQ + \Gamma$ mixture models, where n varied between 1 and 6 independent rate matrices (Qs) and Γ represents among-site rate variation under a gamma distribution with four rate categories. A parameter for the proportion of invariant sites is not implemented in BayesPhylogenies. Each MCMC run consisted of 4×10^6 generations and four Markov chains. We allowed the Markov-chains to reach stationarity before sampling trees at intervals of 1000 generations. Stationarity was assessed by plotting the $-\ln L$ per generation in Microsoft Excel v. X and checking for no average improvement in the likelihood scores. We discarded the first 2000 trees (2×10^6 generations) as "burn-in;" thus, estimates of trees and parameter values are based on 2000 trees sampled from each run.

Conventional likelihood ratio tests used to distinguish maximum likelihood estimates for different partitioning strategies are not applicable in a Bayesian framework because they assume that parameter estimates are at their maximum likelihood values, and MCMC methods sample the posterior density of parameters rather than finding maximum likelihood point estimates (Pagel and Meade, 2005). In a Bayesian context, an alternative approach to model selection uses Bayes factors (Jeffreys, 1935; Kass and Raftery, 1995; Nylander et al., 2004). The Bayes factor (BF) penalizes more complex hypotheses by multiplying the probability of the data given the model parameters of the hypothesis by the set of prior probability terms for each parameter (numbers $\leqslant 1$), which reduces the marginal likelihood. If we consider two alternative partitioning strategies, a and b, as alternative hypotheses, then the Bayes factor (B_{ab}) is the ratio of the marginal likelihood (i.e., probability of the data given all the model parameters for a particular hypothesis scaled by that hypothesis' prior probability and integrated over all parameter values) of hypothesis a to that of hypothesis b and represents a summary of the evidence provided by the data in favor of hypothesis a relative to hypothesis b.

However, computing the marginal likelihoods of the hypotheses (alternative partitioning strategies) is very difficult in practice. We estimated the marginal likelihood of a hypothesis by computing the harmonic mean of the ln-likelihood ($\ln L$) values sampled from the posterior distributions of the two corresponding converged chains being compared (Kass and Raftery, 1995; Raftery, 1996). Although the harmonic mean estimator can be unstable, it is said to be sufficiently accurate for interpretation on a natural logarithmic scale. Harmonic mean calculations were made using the "sump" command in MrBayes of all saved trees after "burn-in" for both MrBayes and BayesPhylogenies runs. Following Pagel and Meade (2005), Bayes factor estimates were calculated by subtracting the harmonic mean $\ln L$ of hypothesis b from the harmonic mean $\ln L$ of hypothesis a, multiplying this value by -2, then subtracting the difference in number of free parameters between models multiplied by ln(0.01), reflecting the uniform 1–100 prior distribution (see Table 6 for an example). BF values from 2 to 5 were considered positive evidence, >5 were strong evidence, and values >10 as very strong evidence favoring one hypothesis over the other (Kass and Raftery, 1995; Pagel and Meade, 2004; Raftery, 1996).

2.3. Hypothesis testing

Significance tests were used to examine support for the optimal tree(s) relative to alternative hypotheses. Phylogenetic hypotheses were tested by comparing the optimal (parsimony or likelihood) topology in the absence of a constraint with the optimal tree under a topological constraint. To find the optimal tree(s) compatible with a particular phylogenetic hypothesis, topologies containing the minimal number of groups necessary to describe the hypothesis of interest (Appendix B) were constructed using MacClade 4.08 (Maddison and Maddison, 2003). These partially resolved topologies were loaded as constraints in PAUP*, which was then used to analyze the sequence data under the parsimony criterion using heuristic searches with the same settings used to estimate the most parsimonious trees in the absence of topological constraints. Phylogenetic hypotheses were tested similarly under likelihood using constraint trees and a successive approximations approach using the same search strategies as described for searches in the absence of topological constraints.

Wilcoxon signed-ranks (WSR) tests (Felsenstein, 1985b; Templeton, 1983) were used with the parsimony trees. They were conducted as two-tailed tests using PAUP*, which incorporates a correction for tied ranks. Goldman et al. (2000) criticized the application of the WSR test as applied in many studies because this test assumes that phylogenetic hypotheses are selected independently of the data analysis (a priori). However, as the optimal trees used in all of our tests corresponded with one of the four alternative hypotheses identified in previous studies (Fig. 1), they may be considered to have been chosen a priori. In any case, Shimodaira–Hasegawa (SH) (Shimodaira and Hasegawa, 1999) and Approximately Unbiased (AU) tests (Shimodaira, 2002) also were performed to test the optimal ML tree in the absence of topological constraints (which corresponded to one of the a priori hypotheses) against the optimal ML trees under each of the alternative hypotheses (see

Appendix B for constraint trees). SH tests were conducted in PAUP* using bootstrap resampling of 10,000 replicate estimated ln-likelihoods of sites (RELL) and AU tests were conducted in the program CONSEL (Shimodaira and Hasegawa, 2001) with 10 sets of 10,000 bootstrap replicates. Note that the constraint topology consistent with Topology I was identical to the optimal ML tree in the absence of constraints from the analyses of the combined data, the *ND1–COI* partition, and the *cyt b-tRNA^{Thr}* partitions. Therefore, SH and AU tests included only the ML topology and the best trees compatible with Topologies II–IV.

Alternative hypotheses also were tested in a Bayesian framework (Huelsenbeck et al., 2002) using the post-burn-in samples of 4000 trees from the analyses with assumed partitions. Topologies corresponding to the alternative hypotheses (Appendix B) were loaded as constraints in PAUP*, and post-burn-in trees for each of the six partitioning strategies (Table 2) were imported. Each of the four alternative topological constraints was then used to filter the trees, saving only those trees that were compatible with a particular constraint. The Bayesian credibility value for the associated hypothesis was calculated by dividing the number of saved trees by the total number in the post-burn-in sample (4000). If the fraction of filtered trees was less than 5% of the total, then the constrained topology in question was outside of the 95% credible interval (i.e., was rejected).

3. Results

3.1. Sequence variation

Prior to alignment, the fragment spanning *ND1–COI* was between 1732 and 1750 base pairs in length and the *cyt b* to *tRNA^{Thr}* fragment was between 1077 and 1093 base pairs across all samples. For the taxa sampled in this study, alignment of sequences for protein-coding genes was unambiguous. Among tRNA genes, several loop regions could not be aligned unambiguously as was the case for some noncoding regions between genes. Regions that could not be aligned unambiguously were parts or all of the dihydrouridine (D) loops for *tRNA^{Ile}* (positions 104–109), *tRNA^{Met}* (positions 250–252), *tRNA^{Trp}* (positions 1356–1361), *tRNA^{Cys}* (positions 1644–1647), and *tRNA^{Thr}* (positions 2852–2863). Nucleotide positions in the TΨC (T) loops for the genes encoding tRNA^{Trp} (positions 1393–1399) and tRNA^{Cys} (positions 1606–1612) were aligned ambiguously as were noncoding sequences between the genes encoding ND1 and tRNA^{Ile} (positions 85–88), tRNA^{Gln} and tRNA^{Met} (positions 231–235), ND2 and tRNA^{Trp} (positions 1337–1339), tRNA^{Trp} and tRNA^{Ala} (positions 1413–1417), tRNA^{Ala} and tRNA^{Asn} (positions 1487–1490), tRNA^{Cys} and tRNA^{Tyr} (positions 1661–1664), tRNA ^{Tyr} and COI (positions 1737–1745), and cyt *b* and tRNA^{Thr} (positions 2821–2838). These ambiguously aligned regions were removed prior to conducting analyses. All taxa used for phylogenetic analysis had a recognizable origin for light-strand replication (O_L) between the *tRNA^{Asn}* and *tRNA^{Cys}* genes according to the criteria outlined in Macey et al. (1997). However, the O_L loop could not be aligned unambiguously; therefore, this region (positions 1574–1584) was excluded. Excluded regions comprise less than 3.8% (108 of 2871) of the aligned sequence positions. Aligned DNA sequences are available in TreeBASE (Study accession number = S1936; Matrix accession number = M3568).

Aligned DNA sequences for all regions comprised 2871 nucleotide positions. Of 2763 inferred unambiguous sites in 31 aligned sequences, 1200 (1003 ingroup only) are variable, 987 (858 ingroup only) are informative under the parsimony criterion, and there are 1075 distinct data (site) patterns under the GTR + Γ + I model in the combined data set. The partition spanning *ND1–COI* has 735 (599 ingroup only) variable characters of which 587 (506 ingroup only) are parsimony informative and 671 distinct data patterns. The region spanning *cyt b* to *tRNA^{Thr}* has 465 (404 ingroup only) variable characters of which 400 (352 ingroup only) are parsimony informative and 452 data patterns. For all protein-coding genes (*ND1*, *ND2*, *COI*, *cyt b*), codon positions vary in their contribution of phylogenetic information with 272 variable first positions (224 parsimony informative; 261 data patterns), 109 variable second positions (73 parsimony informative; 101 data patterns), and 650 variable third positions (571 parsimony informative; 622 data patterns). Among tRNA genes stem regions contribute 119 variable positions (81 parsimony informative; 116 data patterns) and non-stem regions have 50 variable positions (38 parsimony informative; 55 data patterns).

3.2. Phylogeny estimation

3.2.1. Single model and set of parameter values (Unpartitioned Parsimony and ML)

Parsimony analysis of the equally weighted data set including all mtDNA regions identified three most-parsimonious (MP) trees (length = 4200 steps). Two of these trees conform to topology I of Fig. 1, and differ from one another only in the position of the clade containing *Callisaurus* spp. and *C. d. bogerti*, while the third is compatible with topology III. Analysis of 1697 base pairs spanning the region from *ND1–COI* yielded four MP trees (length = 2324 steps), which are all compatible with topology I (Fig. 1). The trees differ in the positions of *Holbrookia propinqua* and the clade containing *Callisaurus spp.* and *C. d. bogerti*. Two MP trees were obtained from analysis of 1066 nucleotide positions spanning genes encoding cyt *b* and tRNA^{Thr} (length = 1866 steps). These trees also conform to topology I (Fig. 1) and differ from each other in the position of the clade containing *Callisaurus spp.* and *C. d. bogerti.* High bootstrap support for most branches was found for trees resulting from each of these partitions and the combined data set (Table 3 and Fig. 2). The com-

Table 3
Bootstrap proportions and Bayesian credibility values for nodes labeled in Fig. 2 based on different analytical methods and character partitions or partitioning strategies

Branch number	Parsimony			Likelihood			Bayesian		
	Combined	ND1–COI	Cytb-tRNA^{Thr}	Combined	ND1–COI	Cytb-tRNA^{Thr}	Assumed partitions (P1–P6)	Mixture models	
								1–6Q	1–6Q + Γ
1	77	98	—	—	95	—	73–94	<50–94	<50–92
2	100	100	54	100	100	69	100	100	100
3	100	100	76	100	100	100	100	100	100
4	98	89	75	100	100	78	100	100	100
5	100	99	73	99	100	96	100	100	100
6	100	100	99	100	100	100	100	100	100
7	100	100	100	100	100	100	100	100	100
8	100	100	100	100	100	100	100	100	100
9	100	100	100	100	100	100	100	100	100
10	70	60	75	72	57	88	98–100	86–100	85–100
11	99	99	60	99	96	72	100	100	100
12	100	100	100	100	100	100	100	100	100
13	100	100	99	100	100	96	100	100	100
14	54	—	—	62	64	—	79–88	82–100	72–90
15	100	100	100	100	100	99	100	100	100
16	100	100	100	100	100	100	100	100	100
17	100	100	100	100	97	94	100	100	100
18	56	56	—	70	—	58	85–97	63–98	69–91
19	100	100	100	100	100	97	100	100	100
20	90	82	66	99	88	77	100	100	100
21	100	100	100	100	100	100	100	100	100
22	100	100	99	100	100	97	100	100	100
23	96	91	84	100	100	92	100	100	100
24	92	53	93	92	—	95	100	100	99–100
25	100	100	100	100	100	99	100	100	100
26	89	83	66	99	98	57	100	100	100
27	95	92	75	100	89	98	100	100	100
28	100	99	98	100	98	96	100	100	100

P1–P6 refer to the partitioning strategies listed in Table 2. Empty cells ("—") refer to bootstrap values less than 50% and thus not present in the bootstrap majority rule consensus tree.

bined data set had the highest average bootstrap values while the cyt *b*-tRNA^{Thr} region had the lowest with the ND1–COI fragment intermediate (Table 3; Fig. 2).

Both hierarchical likelihood ratio tests (LRT) and the Akaike information criterion (AIC) selected the GTR + I + Γ model as the most appropriate model for the combined, unpartitioned data. Using these data and that model in a successive approximations ML analysis, a single topology was found ($-\ln L = 21799.01$) conforming to Topology I (Fig. 3). Analysis of the ND1–COI data set with Modeltest selected the TrN + I + Γ model using LRTs and the GTR + I + Γ model using the AIC. In this case, the GTR + I + Γ model was never compared to the TrN + I + Γ model in the path taken by the program for the LRTs. These models differ by 16.77 likelihood units and are significantly different ($P < 0.01$) when a LRT is applied; thus, we used the GTR + I + Γ model for subsequent ML analyses of the ND1–COI data set. Application of the successive approximation ML analysis yielded a single ML tree ($-\ln L = 12685.57$) that conformed to Topology I (Fig. 1). The model selected for the mtDNA fragment spanning *cyt b-tRNA^{Thr}* was GTR + I + Γ using LRTs but HKY + I + Γ using the AIC. These models were not compared by Modeltest using the LRT and differ by only 0.75 likelihood units, which is not significant according to the LRT ($P > 0.05$), so we used the HKY + I + Γ model. The successive approximations ML analysis of this fragment found one tree ($-\ln L = 9043.26$) that conformed to Topology I (Fig. 1). Trees from the combined data and the *ND1–COI* region were almost identical, the only exception being the position of *Uma inornata,* while the topology found using the *cyt b* fragment differed from the other two in the position of the clade containing *C. d. bogerti* and *C. d. spp.* and the relationships of the outgroup species to each other. As with MP analyses, most branches of the ML trees were well supported (Table 3), though about 15.5% of the branches were poorly to moderately supported, and ca. 13% were poorly supported.

3.2.2. Mixed model ML (assumed partitions)

Each of the six partitioning strategies (Table 2) was analyzed using the GTR + Γ model as implemented in mmml. Current implementation of the program allows specification of only one model, though estimated parameter values can differ between partitions. Although the GTR model is not necessarily optimal for all of the various partitions, it was selected by both LRTs and the AIC for both the unpartitioned data set and the ND1–COI partition, and

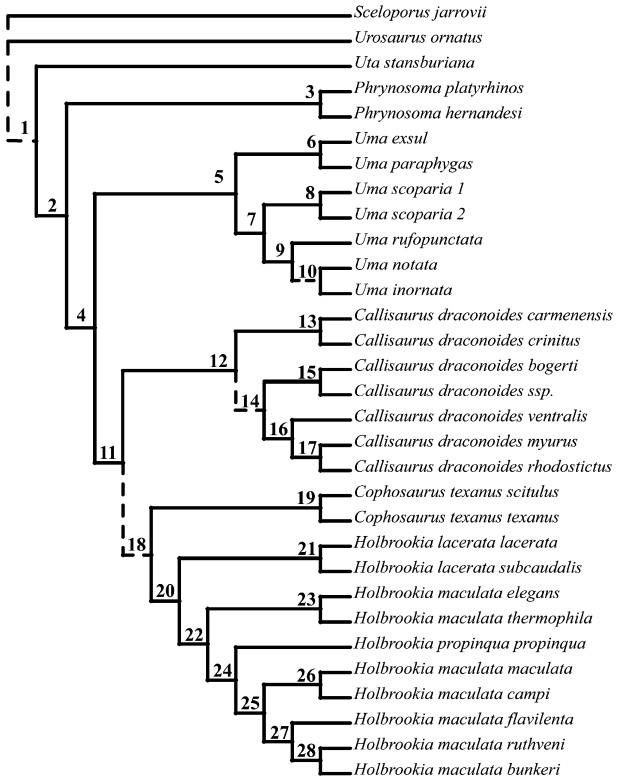

Fig. 2. Phylogenetic relationships of phrynosomatine sand lizards based on parsimony, likelihood, and Bayesian analysis of combined mtDNA sequences. The numbers above each node correspond to the clade numbers in Table 3, which gives bootstrap proportions (MP and ML) and Bayesian credibility (BC) values for the various nodes. The topology shown corresponds to 1 of 3 MP trees for the combined data and 1 of 4 MP trees for the *ND1–COI* fragment, ML trees for the unpartitioned, combined data and the *ND1–COI* fragment, and the Bayesian consensus trees for assumed Bayesian partition strategies P1–P6, and mixture models $2Q$–$5Q$, $1Q+\Gamma$, and $3Q$–$6Q+\Gamma$ (note that the analyses favoring this particular topology are a subset of those favoring Topology I). Dashed branches indicate that alternative topological arrangements were found in the optimal tree resulting from one or more analyses or in one or more of the three most-parsimonious trees.

Fig. 3. ML phylogram of sand lizards relationships based on analysis of unpartitioned data containing 2763 included characters spanning the regions from *ND1* to *COI* and *cyt b* to $tRNA^{Thr}$ under the GTR + I + Γ model. Branches are proportional to their lengths. The ML score of the single topology obtained using the GTR + I + Γ model is $-\ln L = 21799.01$.

by LRTs for the cyt *b* partition. Moreover, using a single model facilitates the comparison of alternative partitioning strategies using LRTs (Table 4). The optimal trees for all partition strategies (Table 2) conformed to Topology I (Fig. 1), with the strategy allowing the three codon positions of all protein coding genes, tRNA stem regions, and tRNA nonstem regions each to have a unique set of parameter values (P6) having the highest likelihood ($-\ln L = 20545.99$; Table 4). According to LRTs (Table 4), P6 was significantly better than all nested partitioning strategies (P1, P3, and P4). For all other comparisons involving nested strategies, strategies with greater numbers of partitions were significantly better than those with fewer partitions (P2 versus P1, P3 versus P1, P4 versus P1 and P3, P5 versus P1 and P3). Finally, for comparisons involving similar numbers of partitions, partitioning strategies based on structural and/or functional properties of the molecules (codon position and RNA secondary structural features) explained the data better than those based on contiguous sequences and/or genes (e.g., P3 versus P2, P6 versus P5) though these differences were not tested for significance because the partitions are not nested. Indeed, P4, which has a separate partition for each codon position, explained the data better than P5, which has a separate partition for each protein-coding gene, despite having fewer partitions (4 versus 5).

3.2.3. Mixed model bayesian (assumed partitions)

The consensus topologies of the 4000 trees in the posterior distributions were identical for all partitioning strategies using both GTR + I + Γ and GTR + Γ and congruent with Topology I, as well as being identical to one of the MP trees and the ML tree for the combined (unpartitioned) data (Fig. 2). The only major effect of the different partitions was that BC values for node 18 (*Cophosaurus* plus *Holbrookia*) increased from 85% to 90% using P1–P3 and 91% with P5 to 96% and 97% using strategies P4 and P6, respectively (Table 3). Most nodal BC values were high (BC > 95%) under all partitioning strategies, and only two nodes (1 and 14) had BC values below 95% under most partitioning strategies.

Table 4
Comparison of alternative partitioning strategies (assumed partitions) based on mixed model maximum likelihood analyses using the ML tree for each partition, all of which correspond to Topology I

Partition(s)	Number of partitions	ML score ($-\ln L$)	$-\ln L$ differences; P values (partition compared)
All data combined (P1)	1	21886.84[a]	
ND1–COI coding region; cyt *b*-tRNA^{Thr} coding region (P2)	2	21821.79	65.05; $P < 0.001$ (P1)
All protein-coding genes; tRNAs (P3)	2	21699.59	187.25; $P < 0.001$ (P1)
Codon positions of all protein coding genes; tRNAs (P4)	4	20610.58	1276.26; $P < 0.001$ (P1)
			1089.01; $P < 0.001$ (P3)
ND1; ND2; COI; cyt *b*; tRNAs (P5)	5	21550.62	336.22; $P < 0.001$ (P1)
			148.97; $P < 0.001$ (P3)
Codon positions of all protein coding genes; tRNA stems; tRNA nonstems (P6)	5	20545.99	1340.85; $P < 0.001$ (P1)
			1153.60; $P < 0.001$ (P3)
			64.59; $P < 0.001$ (P4)

P values for comparisons of nested partitioning strategies are based on likelihood ratio tests.
[a] Current models implemented in the "mmml" program do not include a parameter for the proportion of invariant sites (I), hence the ML score for all data combined (P1) in this analysis is not as good as that obtained for the same data set in analyses that included this parameter (e.g., Table 7).

The harmonic mean $-\ln L$ (Table 5) and approximate Bayes Factors (Table 6) were used to evaluate the ability of the six assumed (prior) partitioning strategies to explain the data set. Among all strategies P6, which separately analyzed each codon position, tRNA stems, and tRNA nonstems, explained the data better than the next best partition (P4) by almost 23 $-\ln L$ units and an approximate BF of 3.85. Partition identity was found to have the greatest effect on improving model performance. Partitions that separated each protein-coding codon position (P4 and P6) were favored over partitions that were defined by fragment or gene boundaries (P2 and P5) or gene type (protein-coding genes and tRNAs—P3) as indicated by both $-\ln L$ (Table 5) and BF estimates (Table 6). Even though strategy P4, which has a partition for each codon position, has fewer partitions (4 versus 5) than P5, which does not, the harmonic mean $-\ln L$ score of P4 is almost 940 units better than that of P5, and the BF estimate of 1918.07 overwhelmingly favors P4. Partitioning tRNAs by stem and non-stem regions (P6 versus P4) also improved the model, though the improvement was far less in terms of both the harmonic mean $-\ln L$ score (ca. 23 $-\ln L$ units) and the BF (3.85).

3.2.4. Mixture model bayesian

The consensus topology of the 2000 trees in the posterior distribution from all mixture model analyses (1–6Q and 1–6$Q + \Gamma$) was congruent with Topology I (Fig. 1). All models yielded identical topologies for sand lizard relationships with the only difference between topologies being the relative position of the outgroup taxa *Sceloporus, Uta,* and *Urosaurus*. BC values for the relationship of *Cophosaurus* and *Holbrookia* (node 18) were above 95% only for the 1Q analysis (98%) and below this cutoff for all other mixture models. The sister-taxon relationship between *Uma notata* and *U. inornata* (node 10) was below 95% using a single rate matrix with or without gamma but the addition of more than one rate matrix yielded BC values above 95% for all other models. Interestingly, node 14 had a BC value above 95% using models 1–4Q but below 95% when models 5–6Q and 1–6$Q + \Gamma$ were used.

Table 5
Arithmetic mean ($\overline{X}_a$) and harmonic mean ($\overline{X}_h$) $-\ln L$s for sets of trees obtained in Bayesian analyses based on assumed (prior) partitions (implemented in MrBayes) and mixture models (implemented in BayesPhylogenies)

Assumed partitions				Mixture models							
Partitioning strategy	$\overline{X}_a - \ln L$	$\overline{X}_h - \ln L$	Free parameters	Model	$\overline{X}_a - \ln L$	$\overline{X}_h - \ln L$	Free parameters	Model	$\overline{X}_a - \ln L$	$\overline{X}_h - \ln L$	Free parameters
P1	21911.82	21940.15	9	1Q	24141.80	24155.32	8	1$Q + \Gamma$	21924.03	21939.13	9
P2	21870.94	21904.18	18	2Q	21943.82	21970.29	15	2$Q + \Gamma$	21747.91	21761.55	16
P3	21752.79	21791.47	18	3Q	21755.12	21770.30	22	3$Q + \Gamma$	21656.37	21674.23	23
P4	20766.30	20803.95	36	4Q	21636.08	21651.64	29	4$Q + \Gamma$	21633.54	21651.82	30
P5	21701.16	21742.26	45	5Q	21594.21	21621.55	36	5$Q + \Gamma$	21633.52	21652.69	37
P6	20736.52	20781.30	45	6Q	21575.72	21600.21	43	6$Q + \Gamma$	21634.24	21651.57	44

Assumed partitions results are from analyses using the GTR + Γ model (not incorporating an estimate of invariant sites) to allow comparison with mixture model analyses results (see text for explanation). Free parameters for assumed partitions are calculated as follows: 5 for rate matrix (G to T rate set to 1), 3 for base frequencies, and 1 for gamma shape parameter; 9 parameters are added for each additional partition as these values are unlinked as implemented in MrBayes. Free parameters for the 1$Q + \Gamma$ mixture model are the same as for P1 under GTR + Γ; however, for each rate matrix above 1Q 7 parameters are added, 6 parameters for the rate matrix (G to T substitution rates for additional matrices are free to vary) and 1 parameter for the weight value of that matrix.

Table 6
Approximate BF estimates used to compare alternative partitioning strategies and mixture models

	Partitioning strategies											
	P1	P2	P3	P4	P5	P6	1*Q*	2*Q*	3*Q*	4*Q*	5*Q*	6*Q*
P1	—	30.49	255.91	2148.06	229.99	2151.91	−4425.73	−87.91	279.83	484.92	512.86	523.30
P2		—	225.42	2117.57	199.50	2121.42	−4456.23	−118.40	249.34	454.42	482.37	492.81
P3			—	1892.15	−25.92	1896.00	−4681.65	−343.82	23.92	229.00	256.95	267.39
P4				—	−1918.07	3.85	−6573.80	−2235.97	−1868.23	−1663.14	−1635.20	−1624.76
P5					—	1921.92	−4655.73	-317.90	49.84	254.92	282.87	293.31
P6						—	−6577.65	−2239.82	−1872.08	−1667.00	−1639.05	−1628.61
1*Q*							—	4337.82	4705.57	4910.65	4938.60	4949.04
2*Q*								—	367.74	572.83	600.77	611.22
3*Q*									—	205.08	233.03	243.47
4*Q*										—	27.94	38.39
5*Q*											—	10.44
6*Q*												—
1*Q* + Γ												
2*Q* + Γ												
3*Q* + Γ												
4*Q* + Γ												
5*Q* + Γ												
6*Q* + Γ												

	1*Q* + Γ	2*Q* + Γ	3*Q* + Γ	4*Q* + Γ	5*Q* + Γ	6*Q* + Γ
P1	2.04	324.96	467.37	479.95	445.98	415.98
P2	−28.45	294.47	436.87	449.46	415.48	385.49
P3	−253.87	69.05	211.45	224.04	190.06	160.07
P4	−2146.02	−1823.10	−1680.69	−1668.11	−1702.09	−1732.08
P5	−227.95	94.97	237.37	249.96	215.98	185.99
P6	−2149.87	−1826.95	−1684.55	−1671.96	−1705.94	−1735.93
1*Q*	4427.77	4750.70	4893.10	4905.69	4871.71	4841.71
2*Q*	89.95	412.87	555.28	567.86	533.89	503.89
3*Q*	−277.79	45.13	187.53	200.12	166.14	136.15
4*Q*	−482.88	−159.95	−17.55	−4.97	−38.94	−68.94
5*Q*	−510.82	−187.90	−45.49	−32.91	−66.89	−96.88
6*Q*	−521.26	−198.34	−55.94	−43.35	−77.33	−107.33
1*Q* + Γ	—	322.92	465.33	477.91	443.94	413.94
2*Q* + Γ		—	142.40	154.99	121.01	91.02
3*Q* + Γ			—	12.58	−21.39	−51.39
4*Q* + Γ				—	−33.98	−63.97
5*Q* + Γ					—	−30.00
6*Q* + Γ						—

Bayes factors were calculated using the following formula: $BF = -2\ (\ln L_a - \ln L_b) - (P_a - P_b)$, where $\ln L_x$ and P_x are the harmonic mean of the ln Likelihoods (from Table 5) and the number of free parameters of model (partition strategy) *x*. For example, comparing models P1 (row) and P2 (column) yields: $-2^{*}(-21940.15 - (-21904.18)) - ((9 - 18)^{*}\ln(0.01)) = 30.49$. BF values from 2 to 5 were considered positive evidence, >5 were strong evidence, and values >10 as very strong evidence favoring one model over the other (Kass and Raftery, 1995; Pagel and Meade, 2004; Raftery, 1996). Positive BF values favor the model in the column heading and negative BF values favor the model in the row heading.

The harmonic mean $-\ln L$ score of the model with one rate matrix and gamma rate-heterogeneity ($1Q + \Gamma$) is about 2216 units better than the score of the model with one rate matrix alone ($1Q$) (Table 5), indicating that there is a large component of rate-heterogeneity in the data. When a second Q matrix without Γ is used, the harmonic mean $-\ln L$ score is only about 30 units below that of the $1Q + \Gamma$ model suggesting that an additional rate matrix (pattern heterogeneity) is an alternative way of accounting for rate heterogeneity (Pagel and Meade, 2004). However, the additional rate matrix comes at the cost of seven additional parameters (6 substitution rate categories plus 1 weight term) versus a single parameter, Γ, and consequently, the simpler model is strongly favored by the approximate BF (Table 6). As the number of rate matrices increases (from 1 to 6), both with and without gamma, the rate of score improvement reaches a plateau (Table 5). For the $5Q$ and $6Q$ mixture models, the scores are higher without Γ than with it (Table 5), which may result from stochastic variation or insufficient mixing of chains due to relatively high numbers of parameters (A. Meade, pers. comm.). These factors may also explain why $1Q + \Gamma$ and the identical P1 (under GTR + Γ) differ slightly with regard to both mean $-\ln L$ scores (Table 5) and approximate Bayes Factor (Table 6).

Determining how many rate matrices to estimate is commonly done using a combination of Bayes factors (Raftery, 1996) and examining the variability in estimated parameters (Pagel and Meade, 2004; Pagel and Meade, 2005). Based on BFs (Table 6) it appears that the model with

six rate matrices and no gamma is justified as the best solution to these data. However, the model with five rate matrices may be more appropriate because, as Pagel and Meade (2005) point out, the Bayes factor test assumes that all sites are independent whereas the true number of independent sites in our data set is likely to be fewer than the 2763 aligned sites, given phenomena such as the pairing of bases in the stem regions of the tRNAs. Non-independence has the effect of inflating the difference in likelihood between models (Pagel and Meade, 2005).

We also can use the matrix weights (provided by BayesPhylogenies) as a guide for evaluating how many rate matrices to estimate (Pagel and Meade, 2004; Pagel and Meade, 2005). If additional rate matrices have much lower weights and high variability (measured by standard deviation) than other matrices in the analysis then it is likely that the additional rate matrices are being poorly estimated. For the 5Q model, rate matrix weights were 0.05,0.15,0.15, 0.56, and 0.09 with all standard deviations below 0.025 suggesting an adequate estimation of these parameters by the data. Other models, 4Q, 6Q, and 4–6$Q + \Gamma$, had either lower rate matrix weights for some matrices or higher variability in those matrix weight estimates; therefore we propose that the 5Q model may be the most appropriate for this data set based on the two criteria outlined above. In any case, none of these mixture models (4–6Q, 4–6$Q + \Gamma$) had $-\ln L$ scores as good as the best scores for assumed partitions (strategies P4 and P6) under the GTR + Γ model (Table 5). Although some of this advantage may result from larger numbers of free parameters in P4 and P6 relative to mixture models with 4–6Q with and without gamma (Table 5), the former models are strongly favored even when penalized for additional parameters in the comparisons involving estimated Bayes factors (Table 6).

3.3. Tests of alternative hypotheses

Optimal topologies (Topology I, Fig. 1) for the combined, unpartitioned data set, the *ND1–COI* fragment, and the *cyt b-tRNA^{Thr}* fragment were tested against each of the three alternative hypotheses (Topologies II–IV, Fig. 1) for each of the same partitions under both parsimony (WSR test) and likelihood (SH and AU tests) (Table 7). The combined data set as well as the *ND1–COI* fragment found Topology I to explain the data significantly better than Topology II with the three tests; however, the *cyt b-tRNA^{Thr}* fragment failed to reject Topology II using all tests (Table 7). All three data sets failed to reject Topology III in favor of Topology I using WSR, SH, and AU tests. Using the combined data set, all three tests rejected Topology IV, and whereas the *ND1–COI* fragment could not reject Topology IV using any of the tests, the AU test (but not the others) rejected this topology using the *cyt b-tRNA^{Thr}* fragment.

The results of Bayesian hypothesis tests of the four alternative hypotheses of sand lizard relationships (Fig. 1) for the six different partitioning strategies based on assumed partitions (Table 2) are presented in Table 8. Topology I was compatible with a substantial majority ($\geqslant$87%) of the trees in the post-burn-in sample for all six strategies, although greater than 95% of the trees were compatible with this constraint only for P4 and P6 (Table 8). Using a 95% credible interval cutoff, Topologies II and IV were

Table 7
Results of Wilcoxon signed-ranks (WSR), Shimodaira–Hasegawa (SH), and approximately-unbiased (AU) tests comparing alternative phylogenetic hypotheses (topologies)

Tree	WSR tests			SH tests			AU tests
	Length	N	P	$-\ln L$	δ $-\ln L$	P	P
Topology I							
Combined	4200	(Best)	N/A	21799.01	(Best)	N/A	N/A
ND1–COI	2324	(Best)	N/A	12685.57	(Best)	N/A	N/A
Cyt *b*-tRNA^{Thr}	1866	(Best)	N/A	9043.26	(Best)	N/A	N/A
Topology II							
Combined	4221	101	0.037*	21822.80	23.79	0.014*	0.002*
ND1–COI	2341	55	0.022*	12704.15	18.58	0.033*	0.011*
Cyt *b*-tRNA^{Thr}	1874	22	0.088	9048.80	5.53	0.188	0.090
Topology III							
Combined	4200	(Best)	N/A	21801.16	2.15	0.631	0.323
ND1–COI	2325	47	0.884	12686.21	0.64	0.736	0.525
Cyt *b*-tRNA^{Thr}	1867	17	0.808	9045.21	1.95	0.518	0.233
Topology IV							
Combined	4214	36	0.019*	21822.39	23.38	0.024*	0.012*
ND1–COI	2335	69	0.185	12700.71	15.14	0.091	0.071
Cyt *b*-tRNA^{Thr}	1874	30	0.144	9051.85	8.59	0.079	0.042*

For each hypothesis and data set, the score of the best tree conforming to the specified hypothesis, the number of characters that discriminate between two trees, and the P-value are given. The four competing hypotheses are illustrated in Fig. 1. Significant values ($P \leqslant 0.05$), indicating rejection of the hypothesis, are indicated with an asterisk.

Table 8
The number and frequency of trees in post burn-in samples from mixed model Bayesian analyses that are compatible with each of the four alternative hypotheses of sand lizard relationships (Fig. 1) under six assumed partitioning strategies (Table 2)

	Topology 1		Topology 2		Topology 3		Topology 4	
	# Compatible trees	Freq	# Compatible trees	Freq	# Compatible trees	Freq	# Compatible trees	Post *P*
P1	3601	=0.900	0	<0.001*	376	=0.094	0	<0.001*
P2	3637	=0.909	0	<0.001*	329	=0.082	0	<0.001*
P3	3511	=0.878	0	<0.001*	453	=0.113	0	<0.001*
P4	3900	=0.975	0	<0.001*	91	=0.023*	0	<0.001*
P5	3544	=0.886	0	<0.001*	429	=0.107	0	<0.001*
P6	3874	=0.969	0	<0.001*	120	=0.030*	0	<0.001*

Frequencies are based on samples of 4000 post-burn-in trees; those below a rejection cutoff of 0.05 are indicated with an asterisk.

rejected using all six assumed partitioning strategies. Topology III could not be rejected using P1, P2, P3, and P5 but was rejected by strategies P4, and P6, the partitions that explain the data best (Table 6) and thus presumably represent the most powerful tests.

4. Discussion

4.1. Sand lizard relationships

Previous studies have identified four competing hypotheses of higher-level relationships among phrynosomatine sand lizards (Fig. 1). Although the results of most recent studies have favored one of these alternative hypotheses (Topology I; ChangChien, 1996; de Queiroz, 1989, 1992; Reeder and Wiens, 1996; Wilgenbusch and de Queiroz, 2000; but see Leaché and McGuire, 2006), previously collected data have not been sufficient to reject the alternative hypotheses according to conventional statistical criteria.

The results of the present study provide substantially stronger support for Topology I over alternative phylogenetic hypotheses. Two out of three optimal trees resulting from parsimony analysis of the combined data set conformed to this topology, as did all optimal trees resulting from parsimony analysis of the *ND1–COI* fragment, parsimony analysis of the *cyt b-tRNA*^{Thr} fragment, successive approximations likelihood analysis under a single set of parameter values for the combined data set, the *ND1–COI* fragment, and the *cyt b-tRNA*^{Thr} fragment, mixed model maximum likelihood analysis with partition-specific sets of parameter values under six different partitioning strategies (P1–P6) applied to the combined data set, Bayesian analysis with partition-specific sets of parameter values under six different partitioning strategies (P1–P6) applied to the combined data set, and Bayesian mixture models with one to six rate matrices both with and without a parameter for rate heterogeneity applied to the combined data set. In addition, explicit tests of alternative phylogenetic hypotheses based on parsimony and likelihood (Table 7) as well as Bayesian (Table 8) approaches indicate that the combined data set significantly favors Topology I over both Topology II and Topology IV (failure to reject these topologies using the separate gene fragments under parsimony and likelihood presumably reflects decreased test power associated with smaller sample sizes when the fragments are analyzed separately).

Of the three alternatives to Topology I, only Topology III is not consistently rejected by our data. One of the three optimal trees resulting from the parsimony analysis of the combined data conformed to this topology, which was ranked a close second in the likelihood analysis of the combined data set and in both parsimony and likelihood analyses of the separate *ND1–COI* and the *cyt b-tRNA*^{Thr} fragments (Table 7). Moreover, Topology III was not even close to being rejected by the various explicit parsimony- and likelihood-based tests of alternative phylogenetic hypotheses for the separate gene fragments and the combined (but unpartitioned) data (P values ranging from 0.518 to 1.00). Topology III and Topology I are similar in that both have *Uma* as the sister group of the remaining sand lizards, and they differ only in placing *Cophosaurus* as the sister group of *Callisaurus* (Topology III) versus *Holbrookia* (Topology I). Consistent with the weak support for Topology I over Topology III, the node uniting *Cophosaurus* and *Holbrookia* (node 18) is relatively weakly supported (Table 3). On the other hand, Topology I is favored over Topology III not only by all likelihood and Bayesian analyses of our combined mtDNA data set, but also by previous analyses of allozymes, morphology, immunology, and mtDNA (ChangChien, 1996; de Queiroz, 1989, 1992; Reeder and Wiens, 1996; Wilgenbusch and de Queiroz, 2000). Moreover, under our Bayesian hypothesis tests (Table 8), which employed partition specific parameter estimates, Topology III is much closer to being rejected even in the worst case ($P = 0.113$ for strategy P3) than in the case of the parsimony and likelihood tests based on combined but unpartitioned data (Table 7), and it is rejected by the partitioning strategies that explain the data best (P4 and P6, Tables 5 and 6) and thus are expected to be associated with the most powerful tests. Thus, in contrast to the situation before the present study, Topology I can now be considered highly corroborated (strongly supported) relative to the three alternative phylogenetic hypotheses (Fig. 1).

4.2. Phylogenetic analysis under heterogeneous processes

Accounting for heterogeneity in nucleotide sequence evolution is an active area of investigation in phylogenetics

(Brandley et al., 2005; Castoe et al., 2004; Huelsenbeck and Nielsen, 1999; Nylander et al., 2004; Pagel and Meade, 2004; Wilgenbusch and de Queiroz, 2000). Studies utilizing sequences from different parts of the genome have to account for variation in evolutionary processes and the resulting patterns that exist both among and within those sequences. For example, protein-coding genes are expected to exhibit different patterns of sequence evolution than genes coding for ribosomal or transfer RNAs; similarly, different codon positions within protein coding genes and positions that code for stem versus non-stem regions in RNAs are likewise expected to exhibit different patterns. Application of a single set of model parameter estimates to heterogeneous data partitions will often result in a poor fit to the data. A better strategy is to assess the ability of different models and/or parameter estimates to explain the data (Swofford et al., 1996).

Two primary approaches have been developed for assessing and incorporating heterogeneity into phylogenetic analyses. The first is to partition data into biologically meaningful classes based on prior knowledge of the characters, and then to assume a different evolutionary model for each class. In the case of nucleotide sequences, partitioning is commonly based on transcriptional or translational products, such as protein-coding genes, tRNA genes, or non-transcribed DNA. This is by far the most common strategy and is easily implemented in some existing computer programs, such as MrBayes (Ronquist and Huelsenbeck, 2003). A second approach is to invoke an increasing number of evolutionary models without a priori partitioning of the data, allowing the data to estimate the appropriate number of models while attempting to avoid overparameterization (=mixture-model approach; Pagel and Meade, 2004). An advantage of the latter method is that variation within a partition that may have been missed by the former strategy can be taken into account (Pagel and Meade, 2004). It is also advantageous when biologically meaningful partitions are not evident.

Consistent with previous studies (Brandley et al., 2005; Nylander et al., 2004), partition-specific parameter estimates generally improved the fit of our data by allowing different components of the data set to have distinct values. For example, substitution rates among 1st, 2nd, and 3rd codon positions for protein coding genes each had different parameter values that exhibited limited overlap in their 95% credible intervals. This result was found using both mixed model ML and Bayesian analyses with assumed partitions (results not shown). Brandley et al. (2005) noted that data partitioning greatly reduces systematic error and improves mean $-\ln L$s and posterior probabilities. Our results support this conclusion with regard to tree scores but not necessarily with regard to clade posterior probabilities. Tree scores for ML mixed model and Bayesian analyses using assumed partitions generally improved with increasing numbers of partitions (except P4 was better than P5; Table 4), as they did for mixture models with increasing numbers of Q matrices up to $6Q$ without Γ and $4Q$ with Γ (Table 5). For Bayesian credibility values, however, alternative partitioning strategies showed no obvious trend toward improvement. Partitioning strategies P4 and P6, among the most complex partitions assumed, showed an increase in BC values for node 18 on our phylogenetic tree, but for node 1 there was a decrease in BC values and node 14 showed no improvement. In mixture model analyses, BC values for nodes 1, 14, and 18 generally decreased with increasing numbers of rate matrices, whereas node 10 showed higher BC values with more rate matrices.

Most studies utilizing prior (assumed) partitions in phylogenetic analyses of nucleotide sequence data have implemented four basic strategies. One is to partition by contiguous DNA fragments, such as the region spanning *ND1–COI* versus that spanning *cyt b* through $tRNA^{Thr}$ (as in our P2). Another strategy is to partition by classes of genes based on their transcriptional and/or translational products, such as tRNA versus rRNA versus protein-coding genes (P3, P4 in part, P5 in part). Yet another strategy is to partition by individual genes, such as *ND2* versus *COI* versus $tRNA^{Thr}$ (P5 in part). The fourth strategy is to partition based on structurally and/or functionally defined classes of sites, such as codon positions and RNA stems versus nonstems, regardless of the gene boundaries (P4 in part, P6). These four common alternatives were examined in the present study. In addition, other more complicated partitioning strategies may be applied, including classes of nucleotides that code for functional properties of the encoded protein, such as hydrophobicity, charge, or chemical properties of R-groups (Naylor et al., 1995; Savolainen et al., 2002).

Using both likelihood and Bayesian methods (Tables 4–6), all of the assumed partitions implemented here (P2–P6) are better than no partitioning at all (P1). Partitioning by classes of genes (P3, P4, P5 in part), by individual genes (P5 in part), and by classes of sites (P4 in part, P6), is each better than partitioning by contiguous DNA fragments (P2). Partitioning by individual genes (P5 in part) or by classes of sites within genes (P4 in part, P6) is better than partitioning by classes of genes (P3), though Bayes factors (Table 6) favored partitioning by classes of genes (P3) over partitioning by individual genes (P5 in part). Finally, partitioning by classes of sites within genes (P4 in part, P6) is better than partitioning by individual genes (P5). These results are congruent with those of several previous studies (Brandley et al., 2005; Castoe et al., 2005; Castoe and Parkinson, 2006; Nylander et al., 2004; Strugnell et al., 2005).

Nylander et al. (2004) suggested that although accounting for data heterogeneity across partitions greatly improved the fit of the data, allowing heterogeneity within partitions might yield even greater improvements. Further partitioning might lead to improvements in $-\ln L$ scores, but at some point the advantages of further partitioning (improvement in $-\ln L$) are outweighed by the disadvantages (inaccurate parameter estimation). All data sets examined to date have found that partitioning by codon

position and/or tRNA secondary structure (stem versus nonstem) improved mean $-\ln L$ scores; however, Strugnell et al. (2005) found that in some cases partitioning by codon for all genes together produced a better fit to the data using AIC than partitioning by both codon and gene. These results indicate that the addition of substitution parameters with increasing partitions in some cases is not justified. As the number of parameters increases, the amount of sequence data to be evaluated per parameter decreases, affecting the ability of probabilistic methods to estimate model parameter values, as well as inferences derived from them (e.g., topologies, branch lengths, nodal support), accurately.

An alternative approach to prior (assumed) partitioning is to estimate heterogeneity using mixture models (Pagel and Meade, 2004). Phylogenetic mixture models allow different models to be applied to the same site with varying probabilities, which are estimated from the data themselves. For our data (Tables 5, 6), mixture models both with and without rate heterogeneity (Γ) performed better than unpartitioned data (P1) and assumed partitions based on contiguous gene fragments (P2) and classes of genes (P3). On the other hand, assumed partitions based on codon position and tRNA stems and non-stems (P4 in part, P6) were between 800 and 900 $-\ln L$ units better than the preferred mixture model (5Q) or the best mixture model score (6Q). This improvement comes only at the cost of a single free parameter difference between P6 and 6Q, and thus P6 (and P4) was very strongly favored over 6Q (with or without Γ) according to Bayes factors (Table 6).

Development and implementation of mixed model maximum likelihood, perhaps using genetic algorithms (Zwickl, 2006) or reversible jump (RJ) Markov chain Monte Carlo methods (Huelsenbeck et al., 2004; Pagel and Meade, 2006) that can simultaneously estimate substitution rate parameters and a phylogenetic hypothesis, are the next stage in the evolution of these analytical methods. We anticipate that these developments will serve two very important functions for the phylogenetic community. First, they will simplify the steps researchers must take to perform the multitude of analyses currently available. Second, these methods will greatly facilitate the ability to account for heterogeneous evolutionary processes in complex phylogenetic data sets.

Acknowledgments

We thank the Smithsonian Institution (postdoctoral fellowship to J.A. Schulte), the Laboratories of Analytical Biology, and the National Science Foundation (postdoctoral fellowship to J.A. Schulte II) for financial support; L. Weigt and J. Hunt for laboratory support; O. Torres-Carvajal, A. Meade, and M. Pagel for analytical advice; J. Wilgenbusch for information concerning tissue samples, and providing invaluable support concerning phylogenetic analyses, particularly those involving mmml. Thanks also are owed to the many people who assisted with obtaining specimens, whose names are listed in the Acknowledgments of Wilgenbusch and de Queiroz (2000).

Appendix A

Collection numbers and localities for voucher specimens from which DNA was obtained and GenBank accession numbers are presented. Abbreviations for collections are as follows: California Academy of Sciences (CAS), San Francisco, California; Louisiana State University Museum of Zoology (LSUMZ), Baton Rouge, Louisiana; Museum of Vertebrate Zoology (MVZ), University of California, Berkeley, California; Royal Ontario Museum (ROM), Toronto, Canada; National Museum of Natural History (USNM), Smithsonian Institution, Washington, District of Columbia; field specimen numbers of Kevin de Queiroz (KdQ); field specimen numbers of Richard R. Montanucci (RRM); field specimen numbers of Christopher A. Phillips (CAP). Subspecies taxonomy for *Holbrookia maculata* follows Axtell (1958). **Outgroups:** *Phrynosoma hernandesi,* USA; Colorado; Weld Co.; vic. Stoneham (RRM 2470, **EU543778, EU543747**); *Phrynosoma platyrhinos,* USA; California; no further data (KdQ 057, **EU543777, EU543746**); *Sceloporus jarrovii,* USA; Arizona; Cochise Co.; Huachuca Mts., vic. Carr Canyon (KdQ 036, **EU543774, EU543743**); *Urosaurus ornatus,* USA; Arizona; Yavapai Co.; 3 mi N Hillside (MVZ 214658, **EU543776, EU543745**); *Uta stansburiana,* USA; Arizona; Pima Co.; Santa Catalina Mts,, mouth of Pima Canyon (CAS 178153, **EU543775, EU543744**); **Sand Lizards:** *Callisaurus draconoides bogerti,* México; Sinaloa; 14 km S Dimas (ROM 14970, **EU543788, EU543757**); *Callisaurus draconoides carmenensis,* México; Baja California; San Ignacio (ROM 14182, **EU543786, EU543755**); *Callisaurus draconoides crinitis,* México; Baja California; Dunes N of Guerrero Negro (MVZ 214832, **EU543787, EU543756**); *Callisaurus draconoides myurus,* USA; Nevada; Mineral Co.; W shore of Walker Lake (MVZ 214751, **EU543791, EU543760**); *Callisaurus draconoides rhodostictus,* USA; California; Inyo Co.; Argus Mts., Darwin Wash (MVZ 214739, **EU543792, EU543761**); *Callisaurus draconoides ssp.,* México; Sonora; 4.0 mi. NW of Bahía Kino (ROM 15034, **EU543789, EU543758**); *Callisaurus draconoides ventralis,* USA; New Mexico; Hidalgo Co.; 12.4 mi N and 0.9 mi W of jct. US Hwy 80 and Portal Road (NM 533) (LSUMZ 48811, **EU543790, EU543759**); *Cophosaurus texanus scitulus,* USA; Arizona; Pima Co.; Santa Catalina Mts., mouth of Pima Canyon (MVZ 214711, **EU543793, EU543762**); *Cophosaurus texanus texanus,* USA; Texas; Blanco Co.; Flat Creek, 3.8 mi N of U. S. Rte. 290 (at Henly) on Farm Road 3232, then 0.5 mi E on dirt road (USNM 315499, **EU543794, EU543763**); *Holbrookia elegans elegans,* México; Sinaloa; Mazatlán (ROM 14964, **EU543797, EU543766**); *Holbrookia elegans thermophila,* USA; Arizona; Pima Co.; 0.6 mi E of Hwy 83 on Old Sonoita Hwy (CAS 174400, **EU543798, EU543767**);*Holbrookia lacerata lacerata,* USA; Texas; Concho Co.; 5.8 (by air) miles

E of Eden on dirt road paralleling U. S. Rte. 87 (USNM 315500, **EU543795, EU543764**); *Holbrookia lacerata subcaudalis,* USA; Texas; McMullen Co.; Charles Caron Ranch, 9.5 mi E of Texas Rte. 16 on Farm Road 1962 (USNM 315503, **EU543796, EU543765**); *Holbrookia propinqua propinqua,* USA; Texas; Nueces Co.; dunes just N of E end of Access Rd. 3, just S of Mustang Island State Park (MVZ 214863, **EU543799, EU543768**); *Holbrookia maculata bunkeri,* USA; New Mexico; Luna Co.; 2.0 mi W of Cambray on NM Hwy 549 (USNM 337740, **EU543803, EU543772**); *Holbrookia maculata campi,* USA; New Mexico; McKinley Co.; Zuni Indian Reservation; jct. of Sand Spring Canyon Creek and Zuni River (MVZ 214801, **EU543801, EU543770**); *Holbrookia maculata flavilenta,* USA; Arizona; Cochise Co.; 0.1 mi S Hwy I-10 at junction 666N (MVZ 214814, **EU543804, EU543773**); *Holbrookia maculata maculata,* USA; New Mexico; Chaves Co.; Dune area 0.9 mi S Hwy 380 and 37.0 mi E jct. Hwy 285 at Roswell (USNM 214806, **EU543800, EU543769**); *Holbrookia maculata ruthveni,* USA; New Mexico; Otero Co.; White Sands National Monument, 5.0 mi W of U.S. Hwy 70/82 on The Dunes Drive (USNM 337752, **EU543802, EU543771**); *Uma exsul,* México; Coahuila, Bilbao Dunes (ROM 15315, **EU543779, EU543748**);*Uma inornata,* USA; California; Riverside Co.; Windy Point (CAP 1752, **EU543785, EU543754**); *Uma notata,* USA; California; Imperial Co.; E edge Algodones Dunes, 0.9 mi W jct. of Hwy I-10 and S-34 (MVZ 214799, **EU543784, EU543753**);*Uma rufopunctata,* México; Sonora; 17 km NE Puerto Peñasco (ROM 13919, **EU543783, EU543752**); *Uma paraphygas,* México; Durango; Bolsón de Mapimí, vic. Laboratorio del Desierto (ROM 15089, **EU543780, EU543749**); *Uma scoparia* (1), USA; California; San Bernardino Co.; Pisgah Crater (ROM 14637, **EU543781, EU543750**); *Uma scoparia* (2), USA; California; San Bernardino Co.; S side Kelso Dunes (MVZ 214784, **EU543782, EU543751**).

Appendix B

Trees used as constraints for tests of alternative hypotheses of sand lizard relationships (note that these topologies do not constrain the traditional sand lizard genera to be monophyletic, which makes them more difficult to reject than topologies that constrain monophyly of the genera):

Topology 1: (*Sceloporus jarrovii, Uta stansburiana, Urosaurus ornatus, Phrynosoma platyrhinos, Phrynosoma hernandesi, (Uma exsul, Uma paraphygas, Uma scoparia 1, Uma scoparia 2, Uma rufopunctata, Uma notata, Uma inornata, (Callisaurus draconoides carmenensis, Callisaurus draconoides crinitus, Callisaurus draconoides bogerti, Callisaurus draconoides ssp., Callisaurus draconoides ventralis, Callisaurus draconoides myurus, Callisaurus draconoides rhodostictus, (Cophosaurus texanus scitulus, Cophosaurus texanus texanus, Holbrookia lacerata lacerata, Holbrookia lacerata subcaudalis, Holbrookia elegans elegans, Holbrookia elegans thermophila, Holbrookia propinqua propinqua, Holbrookia maculata maculata, Holbrookia maculata campi, Holbrookia maculata ruthveni, Holbrookia maculata bunkeri, Holbrookia maculata flavilenta*))))).

Topology 2: (*Sceloporus jarrovii, Uta stansburiana, Urosaurus ornatus, Phrynosoma platyrhinos, Phrynosoma hernandesi, ((Uma exsul, Uma paraphygas, Uma scoparia 1, Uma scoparia 2, Uma rufopunctata, Uma notata, Uma inornata, Callisaurus draconoides carmenensis, Callisaurus draconoides crinitus, Callisaurus draconoides bogerti, Callisaurus draconoides ssp, Callisaurus draconoides ventralis, Callisaurus draconoides myurus, Callisaurus draconoides rhodostictus), (Cophosaurus texanus scitulus, Cophosaurus texanus texanus, Holbrookia lacerata lacerata, Holbrookia lacerata subcaudalis, Holbrookia elegans elegans, Holbrookia elegans thermophila, Holbrookia propinqua propinqua, Holbrookia maculata maculata, Holbrookia maculata campi, Holbrookia maculata ruthveni, Holbrookia maculata bunkeri, Holbrookia maculata flavilenta*))).

Topology 3: (*Sceloporus jarrovii, Uta stansburiana, Urosaurus ornatus, Phrynosoma platyrhinos, Phrynosoma hernandesi, (Uma exsul, Uma paraphygas, Uma scoparia 1, Uma scoparia 2, Uma rufopunctata, Uma notata, Uma inornata, ((Callisaurus draconoides carmenensis, Callisaurus draconoides crinitus, Callisaurus draconoides bogerti, Callisaurus draconoides ssp, Callisaurus draconoides ventralis, Callisaurus draconoides myurus, Callisaurus draconoides rhodostictus, Cophosaurus texanus scitulus, Cophosaurus texanus texanus), Holbrookia lacerata lacerata, Holbrookia lacerata subcaudalis, Holbrookia elegans elegans, Holbrookia elegans thermophila, Holbrookia propinqua propinqua, Holbrookia maculata maculata, Holbrookia maculata campi, Holbrookia maculata ruthveni, Holbrookia maculata bunkeri, Holbrookia maculata flavilenta*))).

Topology 4: (*Sceloporus jarrovii, Uta stansburiana, Urosaurus ornatus, Phrynosoma platyrhinos, Phrynosoma hernandesi, ((Uma exsul, Uma paraphygas, Uma scoparia 1, Uma scoparia 2, Uma rufopunctata, Uma notata, Uma inornata, (Callisaurus draconoides carmenensis, Callisaurus draconoides crinitus, Callisaurus draconoides bogerti, Callisaurus draconoides ssp, Callisaurus draconoides ventralis, Callisaurus draconoides myurus, Callisaurus draconoides rhodostictus, Cophosaurus texanus scitulus, Cophosaurus texanus texanus)), Holbrookia lacerata lacerata, Holbrookia lacerata subcaudalis, Holbrookia elegans elegans, Holbrookia elegans thermophila, Holbrookia propinqua propinqua, Holbrookia maculata maculata, Holbrookia maculata campi, Holbrookia maculata ruthveni, Holbrookia maculata bunkeri, Holbrookia maculata flavilenta*)).

References

Adest, G.A., 1978. The relationships of the sand lizards *Uma*, *Callisaurus*, and *Holbrookia* (Sauria: Iguanidae): an electrophoretic study. Unpublished Ph.D. dissertation, University of California, Los Angeles.

Alfaro, M.E., Zoller, S., Lutzoni, F., 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. Mol. Biol. Evol. 20, 255–266.

Anderson, S., Bankier, A.T., Barrell, B.G., De Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R., Young, I.G., 1981. Sequence and organization of the human mitochondrial genome. Nature 290, 457–465.

Axtell, R.W., 1958. A monographic revision of the iguanid genus *Holbrookia*. Ph.D. dissertation, University of Texas, Austin.

Brandley, M.C., Schmitz, A., Reeder, T.W., 2005. Partitioned Bayesian analyses, partition choice, and the phylogenetic relationships of scincid lizards. Syst. Biol. 54, 373–390.

Castoe, T.A., Doan, T.M., Parkinson, C.L., 2004. Data partitions and complex models in Bayesian analysis: the phylogeny of gymnophthalmid lizards. Syst. Biol. 53, 448–469.

Castoe, T.A., Sasa, M.M., Parkinson, C.L., 2005. Modeling nucleotide evolution at the mesoscale: the phylogeny of the Neotropical pitvipers of the *Porthidium* group (Viperidae: Crotalinae). Mol. Phylogenet. Evol. 37, 881–898.

Castoe, T.A., Parkinson, C.L., 2006. Bayesian mixed models and the phylogeny of pitvipers (Viperidae: Serpentes). Mol. Phylogenet. Evol. 39, 91–110.

ChangChien, L.-L., 1996. A phylogenetic study of sceloporine lizards and their relationships with other iguanid lizards based on DNA/DNA hybridization. Ph.D. dissertation, University of Wisconsin, Madison.

Clarke, R.F., 1965. An ethological study of the iguanid lizard genera *Callisaurus*, *Cophosaurus*, and *Holbrookia*. Emporia St. Res. Stud. 13, 1–66.

Cox, D.C., Tanner, W.W., 1977. Osteology and myology of the head and neck regions of *Callisaurus*, *Cophosaurus*, *Holbrookia*, and *Uma* (Reptilia: Iguanidae). Great Basin Nat. 37, 35–56.

Crother, B.I., 2000. Scientific and standard English names of amphibians and reptiles of North America north of Mexico, with comments regarding confidence in our understanding. Herpetolog. Circ. 29, Society for the Study of Amphibians and Reptiles.

Crother, B.I., 2003. Scientific and standard English names of amphibians and reptiles of North America north of Mexico: update. Herpetol. Rev. 34, 196–203.

de Queiroz, K., 1989. Morphological and biochemical evolution in the sand lizards. Ph.D. dissertation, University of California, Berkeley.

de Queiroz, K., 1992. Phylogenetic relationship and rates of allozyme evolution among the lineages of sceloporine sand lizards. Biol. J. Linn. Soc. 45, 333–362.

Douady, C.J., Delsuc, F., Boucher, Y., Doolittle, W.F., Douzery, E.J.P., 2003. Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. Mol. Biol. Evol. 20, 248–254.

Earle, A.M., 1961. The middle ear of *Holbrookia* and *Callisaurus*. Copeia 1961, 405–410.

Earle, A.M., 1962. The middle ear of the genus *Uma* compared to those of the other sand lizards. Copeia 1962, 185–188.

Etheridge, R., de Queiroz, K., 1988. A phylogeny of Iguanidae. In: Estes, R., Pregill, G. (Eds.), Phylogenetic Relationships of the Lizard Families. Stanford Univ. Press, Stanford, California, pp. 283–367.

Felsenstein, J., 1985a. Confidence limits on phylogenies an approach using the bootstrap. Evolution 39, 783–791.

Felsenstein, J., 1985b. Confidence limits on phylogenies with a molecular clock. Syst. Zool. 34, 152–161.

Felsenstein, J., Kishino, H., 1993. Is there something wrong with the bootstrap on phylogenies? A reply to Hillis and Bull. Syst. Biol. 42, 193–200.

Fu, J., 1999. Phylogeny of Lacertid Lizards (Squamata: Lacertidae) and the Evolution of Unisexuality, Ph.D. dissertation. University of Toronto, Toronto.

Goldman, N., Anderson, J.P., Rodrigo, A.G., 2000. Likelihood-based tests of topologies in phylogenetics. Syst. Biol. 49, 652–670.

Huelsenbeck, J.P., Nielsen, R., 1999. The effect of non-independent substitution on phylogenetic accuracy. Syst. Biol. 48, 317–328.

Huelsenbeck, J.P., Larget, B., Miller, R.E., Ronquist, F., 2002. Potential applications and pitfalls of Bayesian inference of phylogeny. Syst. Biol. 51, 673–688.

Huelsenbeck, J.P., Larget, B., Alfaro, M.E., 2004. Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. Mol. Biol. Evol. 21, 1123–1133.

Jeffreys, H., 1935. Some tests of significance, treated by the theory of probability. Proc. Camb. Phil. Soc. 31, 203–222.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795.

Kocher, T.D., Thomas, W.K., Meyer, A., Edwards, S.V., Paabo, S., Villablanca, F.X., Wilson, A.C., 1989. Dynamics of mitochondrial DNA evolution in animals: Amplification and sequencing with conserved primers. Proc. Natl. Acad. Sci. USA 86, 6196–6200.

Kumazawa, Y., Nishida, M., 1993. Sequence evolution of mitochondrial tRNA genes and deep-branch animal phylogenetics. J. Mol. Evol. 37, 380–398.

Kumazawa, Y., Ota, H., Nishida, M., Ozawa, T., 1998. The complete nucleotide sequence of a snake (*Dinodon semicarinatus*) mitochondrial genome with two identical control regions. Genetics 150, 313–329.

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

Lewis, P.O., Holder, M.T., Holsinger, K.E., 2005. Polytomies and Bayesian phylogenetic inference. Syst. Biol. 54, 241–253.

Macey, J.R., Verma, A., 1997. Homology in phylogenetic analysis: alignment of transfer RNA genes and the phylogenetic position of snakes. Mol. Phylogenet. Evol. 7, 272–279.

Macey, J.R., Larson, A., Ananjeva, N.B., Fang, Z., Papenfuss, T.J., 1997. Two novel gene orders and the role of light-strand replication in rearrangement of the vertebrate mitochondrial genome. Mol. Biol. Evol. 14, 91–104.

Macey, J.R., Schulte II, J.A., Larson, A., Fang, Z., Wang, Y., Tuniyev, B.S., Papenfuss, T.J., 1998. Phylogenetic relationships of toads in the *Bufo bufo* species group from the eastern escarpment of the Tibetan Plateau: a case of vicariance and dispersal. Mol. Phylogenet. Evol. 9, 80–87.

Macey, J.R., Schulte II, J.A., Larson, A., Tuniyev, B.S., Orlov, N., Papenfuss, T.J., 1999. Molecular phylogenetics, tRNA evolution, and historical biogeography in anguid lizards and related taxonomic families. Mol. Phylogenet. Evol. 12, 250–272.

Maddison, D.R., Maddison, W.P., 2003. MacClade: Analysis of Phylogeny and Character Evolution. Version 4.08. Sinauer Associates, Sunderland, MA.

Mittleman, M.B., 1942. A summary of the iguanid genus *Urosaurus*. Bull. Mus. Comp. Zool. Harvard Univ. 91, 106–176.

Naylor, G.J.P., Collins, T.M., Wesley, M.B., 1995. Hydrophobicity and phylogeny. Nature 373, 565–566.

Norris, K.S., 1958. The evolution and systematics of the iguanid genus *Uma* and its relation to the evolution of other North American desert reptiles. Bull. Am. Mus. Nat. Hist. 114, 253–326.

Nylander, J.A.A., Ronquist, F., Huelsenbeck, J.P., Nieves-Aldrey, J.L., 2004. Bayesian phylogenetic analysis of combined data. Syst. Biol. 53, 47–67.

Pagel, M., Meade, A., 2004. A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. Syst. Biol. 53, 571–581.

Pagel, M., Meade, A., 2005. Mixture models in phylogenetic inference. In: Gascuel, O. (Ed.), Mathematics of Evolution and Phylogeny. Clarendon Press, Oxford, pp. 121–142.

Pagel, M., Meade, A., 2006. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. Am. Nat. 167, 808–825.

Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Raftery, A.E., 1996. Hypothesis testing and model selection. In: Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (Eds.), Markov Chain Monte Carlo in Practice. Chapman & Hall, London, pp. 163–187.

Rambaut, A., Drummond, A.J., 2005. Tracer. Version 1.3, Available from: http://tree.bio.ed.ac.uk/software/tracer/.

Reeder, T.W., Wiens, J.J., 1996. Evolution of the lizard family Phrynosomatidae as inferred from diverse types of data. Herpetol. Monogr. 10, 43–84.

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

Savolainen, V., Chase, M.W., Salamin, N., Soltis, D.E., Soltis, P.S., López, A.J., Fédrigo, O., Naylor, G.J.P., 2002. Phylogeny reconstruction and functional constraints in organellar genomes: plastid *atpB* and *rbcL* sequences versus animal mitochondrion. Syst. Biol. 51, 638–647.

Savage, J.M., 1958. The iguanid lizard genera *Urosaurus* and *Uta*, with remarks on related groups. Zoologica 43, 41–54.

Schulte II, J.A., Valladares, J.P., Larson, A., 2003. Phylogenetic relationships within Iguanidae inferred using molecular and morphological data and a phylogenetic taxonomy of iguanian lizards. Herpetologica 59, 399–419.

Shimodaira, H., 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51, 492–508.

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114–1116.

Shimodaira, H., Hasegawa, M., 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17, 1246–1247.

Smith, H.M., 1946. Handbook of lizards. Comstock, Ithaca, NY.

Strugnell, J., Norman, M., Jackson, J., Drummond, A.J., Cooper, A., 2005. Molecular phylogeny of coleoid cephalopods (Mollusca: Cephalopoda) using a multigene approach; the effect of data partitioning on resolving phylogenies in a Bayesian framework. Mol. Phylogenet. Evol. 37, 426–441.

Sullivan, J., Abdo, Z., Joyce, P., Swofford, D.L., 2005. Evaluating the performance of a successive-approximations approach to parameter optimization in maximum-likelihood phylogeny estimation. Mol. Biol. Evol. 22, 1386–1392.

Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony* (and other methods). Version 4.0. Sinauer Associates, Sunderland, MA.

Swofford, D.L., Olsen, G.J., Waddell, P.J., Hillis, D.M., 1996. Phylogenetic inference. In: Hillis, D.M., Moritz, C., Mable, B.K. (Eds.), Molecular Systematics, 2nd ed. Sinauer Associates, Sunderland, MA, pp. 407–543.

Tavaré, S., 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. Lec. Math. Life Sci. 17, 57–86.

Templeton, A.R., 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. Evolution 37, 221–244.

Torres-Carvajal, O., Schulte II, J.A., Cadle, J.E., 2006. Phylogenetic relationships of South American lizards of the genus *Stenocercus* (Squamata: Iguania): a new approach using a general mixture model for gene sequence data. Mol. Phylogenet. Evol. 39, 171–185.

Weisrock, D.W., Macey, J.R., Ugurtas, I.H., Larson, A., Papenfuss, T.J., 2001. Molecular phylogenetics and historical biogeography among salamandrids of the "true" salamander clade: rapid branching of numerous highly divergent lineages in *Mertensiella luschani* associated with the rise of Anatolia. Mol. Phylogenet. Evol. 18, 434–448.

Wilgenbusch, J., de Queiroz, K., 2000. Phylogenetic relationships among the phrynosomatid sand lizards inferred from mitochondrial DNA sequences generated by heterogeneous evolutionary processes. Syst. Biol. 49, 592–612.

Zwickl, D.J., 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation, The University of Texas at Austin.