Systematic Biology

Branch Lengths, Support, and Congruence: Testing the Phylogenomic Approach with 20 Nuclear Loci in Snakes

John J. Wiens [a]; Caitlin A. Kuczynski [a]; Sarah A. Smith [a]; Daniel G. Mulcahy [b]; Jack W. Sites Jr. [b]; Ted M. Townsend [c]; Tod W. Reeder [c]

[a] Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY, USA
[b] Department of Integrative Biology, Brigham Young University, Provo, UT, USA
[c] Department of Biology, San Diego State University, San Diego, CA, USA

PLEASE SCROLL DOWN FOR ARTICLE

Branch Lengths, Support, and Congruence: Testing the Phylogenomic Approach with 20 Nuclear Loci in Snakes

JOHN J. WIENS,[1] CAITLIN A. KUCZYNSKI,[1] SARAH A. SMITH,[1] DANIEL G. MULCAHY,[2] JACK W. SITES JR.,[2]
TED M. TOWNSEND,[3] AND TOD W. REEDER[3]

[1]*Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY 11794-5245, USA; E-mail: wiensj@life.bio.sunysb.edu (J.J.W.)*
[2]*Department of Integrative Biology, Brigham Young University, Provo, UT 84602, USA*
[3]*Department of Biology, San Diego State University, San Diego, CA 92182-4614, USA*

Abstract.—Many authors have claimed that short branches in the Tree of Life will be very difficult to resolve with strong support, even with the large multilocus data sets now made possible by genomic resources. Short branches may be especially problematic because the underlying gene trees are expected to have discordant phylogenetic histories when the time between branching events is very short. Although there are many examples of short branches that are difficult to resolve, surprisingly, no empirical studies have systematically examined the relationships between branch lengths, branch support, and congruence among genes. Here, we examine these fundamental relationships quantitatively using a data set of 20 nuclear loci for 50 species of snakes (representing most traditionally recognized families). A combined maximum likelihood analysis of the 20 loci gives strong support for 69% of the nodes, but many remain weakly supported, with bootstrap values for 20% ranging from 21% to 66%. For the combined-data tree, we find significant correlations between the length of a branch, levels of bootstrap support, and the proportion of genes that are congruent with that branch in the separate analyses of each gene. We also find that strongly supported conflicts between gene trees over the resolution of individual branches are common (roughly 35% of clades), especially for shorter branches. Overall, our results support the hypothesis that short branches may be very difficult to confidently resolve, even with large, multilocus data sets. Nevertheless, our study provides strong support for many clades, including several that were controversial or poorly resolved in previous studies of snake phylogeny. [Branch length; congruence; gene tree; phylogeny: phylogenomics; snakes.]

In many ways, the genomics revolution seems poised to transform and rapidly accelerate attempts to reconstruct the Tree of Life. Information from the sequencing of whole genomes now makes it possible to develop markers for vast numbers of unlinked nuclear loci. As a result, enormous data sets that combine sequence data from dozens of nuclear loci are now being assembled to address many perplexing phylogenetic questions (e.g., Rokas et al., 2003, 2005; Takezaki et al., 2004; Phillippe et al., 2005; Hallstrom et al., 2007).

But how well will this multilocus combined approach actually work? Recently, some authors have suggested that this approach may be suprisingly unsuccessful for many phylogenetic problems (e.g., Rokas and Carroll, 2006). They have argued that rapid splitting will lead to short branches and conflicting relationships among genes, with no particular tree being expected most frequently among the gene trees from the independent loci (e.g., Slowinski 2001; Poe and Chubb, 2004; Rokas and Carroll, 2006). Under these conditions, adding more loci may not help, and it has been suggested that the level of support in a combined analysis might even be misleading (e.g., Rokas and Carroll, 2006; Degnan and Rosenberg, 2006; Kubatko and Degnan, 2007).

These arguments seem very plausible, and some are supported by theoretical analyses (e.g., Takahata, 1989; Degnan and Rosenberg, 2006; Kubatko and Degnan, 2007). But in fact, most of the empirical literature relating branch lengths to support and congruence is based largely on speculation and anecdotal observations. Most empirical studies have focused solely on problematic branches, without explicit comparison to branches where these problems were not occurring (e.g., Poe and Chubb, 2004; Rokas and Carroll, 2006). Thus, many fundamental questions remain unanswered.

How are branch lengths related to clade support (e.g., bootstrap values) and incongruence among genes? Do shorter branches actually have weaker support and more conflicts among the underlying gene trees? Are the conflicts among genes strongly supported or weakly supported? If shorter branches do tend to be more weakly supported, is this most likely caused by a limited number of informative characters, conflicts among genes (i.e., discordance between gene and species histories), long-branch attraction, or a combination of these factors?

In this paper, we address these questions empirically in snakes, using a data set of 20 nuclear protein-coding loci for 50 ingroup species, including nearly all of the currently recognized families. Most of these data are new to our study (for 19 of 20 genes). Several recent studies have addressed various aspects of snake phylogeny using mitochondrial data or combined nuclear and mitochondrial data (e.g., Slowinski and Lawson, 2002; Wilcox et al., 2002; Lawson et al., 2004, 2005; Noonan and Chippindale, 2006; Vidal et al., 2007a, 2007b). These studies found many relationships at odds with previous taxonomy and many that were at odds with each other (e.g., placement of Aniliidae in Wilcox et al. [2002] versus Vidal et al. [2007b]), suggesting the need for additional study. However, the primary focus of our study is on using nuclear data from snakes to address the relationships between branch lengths, support, and incongruence, rather than producing a major revision of snake phylogeny and taxonomy (such an analysis should include more taxa, as well as mitochondrial and morphological data). In addition, our study also speaks to the critical question of whether the multi-locus, combined-data, "phylogenomic" approach seems likely to be successful for resolving higher-level phylogeny.

MATERIALS AND METHODS

Collection and Analysis of DNA Sequence Data

Taxon sampling was designed to resolve the higher-level relationships of snakes. Thus, for this study, we included one or more representatives of nearly all traditionally recognized families and subfamilies (e.g., Pough et al., 2004). However, we lacked useable samples from the poorly known families Anomochilidae (which is most likely nested within Uropeltidae; Gower et al., 2005) and Xenophiidae (which is most likely the sister taxon of Bolyeriidae; Lawson et al., 2004). We also included seven outgroup species. Recent molecular analyses of higher-level squamate phylogeny suggest that the closest relatives of snakes are the iguanians and anguimorphans (e.g., Townsend et al., 2004; Vidal and Hedges, 2005). We selected two representative species from of each of these clades, with each species representing a major branch within those clades. We also included three species representing more distant outgroups (Gekkonidae: *Gekko gecko*; Teiidae: *Aspidoscelis tigris*; Xantusiidae: *Xantusia vigilis*). Traditional morphological analyses place snakes with or within anguimorphans (e.g., Estes et al., 1988; Lee, 1998). Thus, our choice of outgroups should also be appropriate if the traditional picture of squamate relationships is correct.

Genes sequenced and their basic properties are listed in Table 1. Primers are listed in Appendix 1 (available online at http://www.systematicbiology.org). We developed primers for many relatively novel loci based on comparisons of the nuclear genomes of *Fugu, Gallus*, and *Homo*, using methods described elsewhere (Townsend et al., 2008). In short, gene regions were carefully selected

that were single copy (at least in the *Gallus* and *Homo* genomes), that were contained entirely within a single exon, that could be sequenced (both strands) with a single pair of sequencing reactions (∼500 to 1000 base pairs [bp]), and that were evolving at an appropriate rate (i.e., variable among snake families). Standard methods of DNA extraction, amplification, and sequencing were used. Sequence data were generated in the labs of Reeder, Sites, and Wiens. All sequences for a given gene were generated in the same lab, and generally the same individual specimen was used for a given species in all three labs. Given that all data were from exons of nuclear protein-coding genes, there were relatively few indels, and alignment was done by eye after translating the nucleotide sequences to amino acid sequences (using MacClade, version 4.0; Maddison and Maddison, 2000).

Preliminary analyses of aligned sequences were conducted using parsimony (using PAUP* version 4.0b10; Swofford, 2002) to detect possible contamination and other laboratory errors. When a given individual had an identical or nearly identical sequence to another individual, the gene was resequenced for that taxon. However, we did not exclude sequences merely because they conflicted with previous taxonomy or with the phylogeny inferred from other genes. Thus, our data should provide a relatively unbiased estimate of incongruence among genes.

New data were generated from 19 loci, most of which have not been used in previous studies of snake phylogeny. Many previous analyses of snake phylogeny have used the nuclear protein-coding CMOS gene (e.g., Slowinski and Lawson, 2002; Lawson et al., 2005; Vidal et al., 2007a, 2007b). Although we did not generate novel sequences from this gene for our study, we included this gene in our analyses using data from GenBank. Previous authors sequenced CMOS for most of the same genera that we sampled, but in several cases they utilized a different species. We treated these congeneric species as equivalent in our combined analyses and in our analyses of congruence; the phylogenetic scale of our study should be too large to be greatly affected by the non-monophyly of genera. We also supplemented our RAG1 data set with one sequence from GenBank (*Acrochordus*, for which we were unable to obtain sequences for this gene). GenBank numbers are given in Appendix 2 (available online at http://www.systematicbiology.org). Other nuclear genes have been sequenced for some snake species (e.g., Vidal et al., 2007a, 2007b), but these genes were sampled too sparsely to include in our analyses of support and congruence (although they might have been useful if our primary goal was merely phylogeny reconstruction).

The major results of this study are based on separate and combined analyses of each gene using maximum likelihood. In theory, we could have conducted these analyses using parsimony and Bayesian methods instead of or in addition to likelihood. However, estimation of branch lengths from parsimony is problematic (Felsenstein, 2004). Nevertheless, we did conduct limited analyses using parsimony and found that results were

TABLE 1. Basic properties of the 20 nuclear protein-coding genes used in phylogenetic analyses of snakes. Note that the best-fitting model is GTR + I + Γ for all loci but GPR37, for which the best model is SYM + I + Γ. For all loci but NGFB, a separate partition for each codon position is supported by comparison of Bayes factors. For NGFB, two partitions are supported (one for the first and second codon positions, another for the third position).

Gene	Length (bp)	Variable characters	Parsimony-informative characters	Taxa
AHR	457	325	222	50
BDNF	676	221	146	55
BMP2	639	312	235	55
CMOS	573	336	228	52
DNAH3	665	320	252	51
ECEL1	582	315	217	46
FSHR	753	331	244	53
FSTL5	583	251	161	49
GPR37	509	212	168	46
MKL1	948	570	408	48
NGFB	588	377	250	56
NT3	516	338	257	49
PNN	1011	557	342	53
PTGER	471	192	140	42
PTPN12	670	472	362	43
RAG1	1000	473	337	45
SLC30A1	555	280	208	55
TRAF6	633	327	224	52
ZEB2	882	326	190	51
ZFP36L1	611	241	166	52

similar to those from likelihood in terms of bootstrap support for individual clades in the combined analysis (see below). Furthermore, we expect results to be generally similar between likelihood and Bayesian analyses (although Bayesian analyses of the combined data were prohibitively slow), given that both methods are likelihood based (Felsenstein, 2004). However, we did use Bayesian analysis to help select partitions within genes (see below), given that current software for Bayesian analysis provided the best options available for testing both complex models and partitioning strategies.

For each gene, we first determined the best fitting model by comparing likelihoods with the Akaike information criterion (AIC) in MrModelTest (Nylander, 2004). For almost all genes, the GTR+I+Γ model was supported as best (general time reversible with parameters for invariable sites [I] and among-site rate variation at variable sites [Γ]). The single exception was GPR37, for which the model selected (SYM+I+Γ) was almost identical to GTR+I+Γ (except that equal base frequencies are assumed).

We tested whether partitions within genes significantly improved the fit of models to the data using comparisons of Bayes factors (Nylander et al., 2004; Brandley et al., 2005). We analyzed each gene separately using MrBayes (version 3.1.2; Huelsenbeck and Ronquist, 2001) both with and without partitions for separate codon positions, using the GTR+I+Γ model for each gene and each partition. We compared the harmonic mean of the log likelihoods for each partitioning strategy and considered a Bayes factor ≥10 to significantly support partitioning (Nylander et al., 2004; Brown and Lemmon, 2007). We also compared results with a single partition for the first and second positions relative to the third (i.e., synonymous versus nonsynonymous substitutions). These analyses supported the use of three partitions for all genes except NGFB, for which a two-partition model received equivalent support and was used instead. Given that different partitions were significantly supported within genes, we assumed different partitions between genes as well.

Maximum likelihood analyses were implemented in RAxML-VI-HPC (Randomized Accelerated Maximum Likelihood), version 2.2.3 (Stamatakis, 2006). Although many programs can implement phylogenetic analysis using maximum likelihood, RAxML can do so very quickly and also allows one to incorporate data partitions (e.g., for different genes and codon partitions), in contrast to current versions of many other programs (e.g., GARLI, PAUP*). One disadvantage of RAxML is that the current version does not allow the invariant sites parameter "I" (but note that heterogeneity of rates among sites is accounted for by the parameter Γ). The optimal likelihood tree for a given data set was estimated by conducting 20 replicate searches on each data set, using the GTRGAMMA option (i.e., the regular GTR+Γ model). We used the default option of 25 gamma rate categories. We also conducted a nonparametric bootstrap analysis (Felsenstein, 1985) on each data set using 200 pseudoreplicates, using the GTRCAT option (i.e., an approximation that can be used as a reasonable replacement for the GTR+Γ model to increase speed; Stamatakis, 2006). Again, we used the default option of 25 rate categories.

We also conducted a limited set of parsimony analyses to show that our general results were not limited to likelihood. We performed a combined analysis of the 20 loci (using PAUP* version 4.0b10; Swofford, 2002) and compared the levels of bootstrap support to those from the likelihood analyses. We used 200 bootstrap pseudoreplicates, with each pseudoreplicate using a heuristic search with 10 random-addition-sequence replicates and tree bisection-reconnection branch swapping.

We were not able to obtain sequence data for every species for every gene. Many species proved difficult to amplify and/or sequence, even after designing relatively taxon-specific primers. However, in the combined analysis, we included all taxa and genes regardless of the level of completeness. The 20 data sets ranged in their extent of taxonomic completeness from 74% to 98% of the 57 species (mean = 88%; Table 1). Some species had more missing data than others, but even the most incomplete species had data for at least six genes. This level of incompleteness is not unusual for multilocus nuclear analyses of higher-level phylogeny (e.g., Philippe et al., 2004; Rokas et al., 2005) and previous simulation and empirical studies suggest that missing data need not preclude taxa from being accurately placed in a phylogenetic analysis (e.g., Wiens, 2003; Driskell et al., 2004; Phillippe et al., 2004; Wiens et al., 2005). Missing data might also affect the estimated branch lengths, but these potential effects should be ameliorated somewhat by our focus on the lengths of internal branches (i.e., clades of two or more species, such that at least some of the relevant species are likely to be sampled for each gene). Furthermore, we found that the branch lengths in the combined analysis are closely related to the averaged branch lengths of the individual genes (see below), and so presumably the combined-data branch lengths are not strongly misled by artifacts of missing data.

Analyses of Support and Congruence

We used our combined and separate analyses of these genes to explore the possible correlations between branch length, clade support, and congruence among genes. We hypothesized that the levels of bootstrap support and congruence among genes for each branch would be correlated with the length of that branch in the combined analysis, given that short branches may have too little time for synapomorphies to accumulate and/or for gene histories to coalesce.

Do short branches in the combined analysis reflect short branches in the underlying species phylogeny? There must be a true underlying species history and a true set of time intervals between nodes, and we assume that these underlying lengths will generally be reflected in the branch lengths of both the individual gene trees (at least when averaged across genes) and in the combined analysis of all genes, and that these lengths should

therefore be strongly correlated. This latter assumption was addressed by testing for a correlation between the length of each branch in the combined analysis versus the average of the lengths of the corresponding branch in the analyses of the separate genes.

Two measures of average branch length were used. For the first, we averaged the branch lengths only for those genes that recovered the clade in question (i.e., genes that did not support that clade were not included). For the second, genes that did not support the clade were included and each given a length of 0 when calculating averages. For both analyses, genes that did not include the relevant taxa for that branch were excluded. Both analyses showed a strong correlation between branch lengths in the combined and separate analyses (see Results), and further analyses were therefore conducted using branch lengths from the combined analysis. These and subsequent analyses of correlation were based on non-parametric Spearman rank correlation, conducted in Statview. All results reported as significant remain significant after Bonferroni correction for multiple tests.

We assessed the correlation between branch length and bootstrap support in the combined analysis. Based on the combined analysis, we performed Spearman rank correlation of the estimated length of each branch against the bootstrap value for that branch. The bootstrap consensus tree differed from the optimal tree for three weakly supported clades, but these clades were still represented among the bootstrapped trees, and these bootstrap proportions were used.

The length of a branch isolated from the rest of the tree may not be the most relevant parameter for predicting support and congruence. It may be easier to reconstruct a clade when the immediate descendant branches are close to the length of the branch in question and more difficult when the descendant branches are much longer (i.e., the combination of a short internal branch and long terminal branches has the potential to create long-branch attraction; Felsenstein, 1978; Huelsenbeck, 1995). Therefore, we also used as an index for a given branch the average of the lengths of its two immediate descendant branches, divided by the length of the branch in question. This index was based on branch lengths in the combined-data tree.

Similarly, some authors have argued that short branches that are relatively old will be the most difficult to reconstruct (e.g., Rokas et al., 2005). Given this idea, we developed another index, using the relative depth of the branch within the tree divided by the length of the branch. The relative depth of each branch was calculated as the average of (a) the shortest path from that branch to the present (the sum of the lengths of the left-most series of branches of a right-ladderized tree; note that by "short" here we mean only the number of branches) and (b) the longest path (the sum of the lengths of the right-most series of branches). This index will have relatively high values for short, old branches and smaller values for younger and/or longer branch lengths. Again, this index was based on branch lengths in the combined analysis.

We then assessed the correlations between branch lengths and congruence between gene trees, using various analyses and indices described below. In general, assessing whether a given gene supports a given clade is straightforward if taxon sampling is identical between genes. However, in our study, all genes were missing data for one or more species. Given this limitation, we counted a gene as supporting a given clade that was present in the combined analysis as long as (a) the gene tree supported the monophyly of the clade for all the species of that clade that were included, and (b) the basal species of the clade in the combined analysis was present in the separate gene tree. For example, say that clade (A (B + C)) was present in the combined analysis. If a given gene included taxa A and C but lacked B, and the clade A + C was supported, then we considered this gene to support the clade. If taxon A was lacking data for that gene, we would consider this to be ambiguous, rather than supporting monophyly of the clade. In cases where the basal species of a clade was missing, but the clade was otherwise supported, we counted the next clade up the tree as present. Thus, if a clade (A (B + C)) was present in the combined data, but a gene tree contained only the taxa B + C, we counted this gene as supporting the B + C clade but not the clade (A (B + C)).

We were also interested in whether clades in the combined analysis were strongly supported (or rejected) in the analyses of the separate genes, as potential evidence of concordant or discordant gene histories. Weakly supported congruence or incongruence could be explained by stochastic sampling of characters (i.e., random error) rather than concordant or discordant gene histories, but this possibility seems unlikely if a clade is strongly supported in a given gene tree.

We first evaluated whether a clade in the combined analysis was strongly supported by a given gene based on likelihood bootstrap values. Given that bootstrap values generally appear to be biased but conservative, we arbitrarily considered values ≥70% to indicate strong support (Felsenstein, 2004). Note that we used a more lenient bootstrap value to detect strongly supported congruence and incongruence between genes than we used to determine whether a clade is strongly supported in the overall combined analysis (≥95%). However, we also addressed the effects of using bootstrap values ≥95% as our criterion for strong support for congruence and incongruence between genes (e.g., Taylor and Piel, 2004). To simplify the scoring of results from the separate genes, we assessed strongly supported congruence and incongruence based only on the bootstrap majority-rule consensus trees for the 20 individual gene trees.

We considered a gene to reject a given clade in the combined analysis if alternate relationships were suggested, and we considered there to be strong support for the conflicting clade when the alternate clade had a bootstrap value ≥70%. This assessment was simple when an alternate relationship involved a minor rearrangement of taxa relative to the combined-data tree but was more complicated when the relevant taxa were placed more distantly in the alternate tree. In these cases, we assessed

the strength for the alternate relationship based on the level of support for the deepest clade that was relevant to the alternate placement. For example, given that the

porting a clade minus the number strongly contradicting that clade (no equation given). We also tested for a correlation between branch length and a novel index of strongly supported incongruence among genes:

$$\frac{(\text{No. strongly for} + \text{no. strongly against}) - |(\text{no. strongly for} - \text{no. strongly against})|}{\text{No. genes relevant to that clade}}$$

combined data supported the relationships (A (B (C + D))), and a given gene supported the relationships (D (B (A + C))), we considered that gene's support for the clade A + B + C rather than for A + C. In general, when the differences between the gene tree and the combined-data tree involved the more distant placement of a taxon in the gene tree relative to the combined-data tree, all of the clades involved tended to be weakly supported (making the choice of a particular clade-support value less critical).

We acknowledge that there are many other ways of assessing conflict between trees. However, it is important to remember that we are interested in conflicts involving particular clades, not entire trees or data sets. Furthermore, other methods for assessing conflict might be problematic here because of the differences in taxon sampling between genes and also the overall large number of clades (48) and genes (20) considered.

After tallying the number of genes strongly supporting or rejecting a clade, we assessed the correlation between branch lengths and congruence among genes. Given that not all genes had taxon sampling that made them relevant to a particular clade (e.g., support for a *Cylindrophis* + *Uropeltis* clade in genes lacking data for *Uropeltis*), our indices of congruence were generally standardized by the number of relevant genes.

We first assessed the simple proportion of genes supporting a clade:

$$\frac{\text{No. genes supporting clade}}{\text{No. genes relevant to that clade}}$$

We tested for a correlation between this index and branch lengths in the combined analysis using Spearman's rank correlation. We also tested for a correlation between branch length and the proportion of genes strongly supporting a clade, for which we used:

$$\frac{\text{No. genes with strong support for clade}}{\text{No. genes relevant to that clade}}$$

Furthermore, we assessed the correlation between branch length and the proportion of genes that showed strong support for a relationship that contradicted the branch in question:

$$\frac{\text{No. genes strongly contradicting clade}}{\text{No. genes relevant to that clade}}$$

We also assessed the correlation between branch length and the absolute number of genes strongly sup-

This index will take its highest values (close to or equal to 1) when the number of genes that strongly favor a clade is equal to the number strongly rejecting it and when these two classes of genes make up nearly all of the relevant genes. Conversely, it will take low values (close or equal to 0) when the relevant clades are weakly supported in the analyses of the separate genes or when all or most of the genes strongly support the branch found in the combined analysis. An important difference between this index and the previous index is that the simple difference between the number of genes supporting and rejecting a given clade does not necessarily reflect the overall number of genes in strong conflict.

It may seem strange that a clade could appear in the combined analysis that had a larger number of genes strongly rejecting it than supporting it. However, it is important to remember that for any three taxa, there are three possible rooted trees. Thus, even though 50% of the genes may support a clade and 50% reject it, if the genes rejecting the clade do not agree, then there may still be twice as many genes supporting the clade than supporting any particular alternate topology (i.e., imagine that 50% of the genes support tree 1, 25% support tree 2, and 25% support tree 3). Similarly, even though more genes may reject than support a given clade, that clade could still be supported by more genes than any particular alternate topology. Although it would have been preferable to assess the relative frequency of support for different alternate resolutions of each clade, this was complicated in many cases by large differences between topologies and the correspondingly large number of alternate relationships.

We then tested for a correlation between bootstrap support (in the combined likelihood analysis) and the number of genes strongly supporting a clade minus the number strongly rejecting it. We also tested a correlation between bootstrap support and our index of conflict described above. The former analysis was then repeated after using a more conservative criterion for assessing strong support and rejection of clades from individual genes (values ≥95% rather than 70%).

Taxonomy

Higher-level snake taxonomy is somewhat unstable at present. Therefore, it is important to clearly define our usage of higher-taxon names. Overall, we follow the taxonomy of Vidal et al. (2007b) for noncolubroid snakes, and for colubroid snakes we follow Lawson et al. (2005), who included the generic content of each of their higher taxa, unlike Vidal et al. (2007a). However, based on the tree of Vidal et al. (2007a) and our own results,

we differ from Lawson et al. (2005) in recognizing Xenodermatidae as distinct from Colubridae. We also prefer to recognize Boodontidae and Atractaspididae as separate families rather than parts of Elapidae, to preserve the traditional meaning of the medically important Elapidae.

RESULTS

The combined data set of 20 genes consists of 13,322 characters, of which 6783 are variable and 4766 are parsimony informative. Combined likelihood analysis yields a phylogeny in which 33 of 48 internal nodes are very strongly supported by bootstrap values ($\geq$95%, and 31 are 100%), whereas 15 nodes are more weakly supported (Fig. 1), with bootstrap values for 10 of these clades ranging from 21% to 66%. There is a strong correlation between likelihood and parsimony bootstrap values for each branch ($r = 0.929$; $P < 0.0001$), and all branch lengths and bootstrap values mentioned hereafter are for likelihood. Basic data for each branch, including lengths, bootstrap values, supporting genes, and related indices are given in Appendix 3 (available online at http://www.systematicbiology.org).

Branch lengths in the combined analysis are correlated with the average lengths of comparable branches in analyses of the separate genes, regardless of whether genes that did not support the clade were excluded ($r = 0.870$; $P < 0.0001$) or included ($r = 0.950$; $P < 0.0001$). There is a significant correlation between branch length and bootstrap support in the combined analysis ($r = 0.810$; $P < 0.0001$); all weakly supported clades are associated with relatively short branch lengths (length = 0.010 estimated substitutions/site or less), whereas strongly supported clades are both short and long (Fig. 2a). There is a strong correlation between bootstrap support and the lengths of branches adjusted for the lengths of the immediate descendant branches ($r = 0.721$; $P < 0.0001$), and between bootstrap support and the average depth of the branch divided by the length of the branch ($r = -0.517$; $P = 0.0004$), but these correlations are no stronger than those considering the length of the branch alone. There is no correlation between bootstrap support and the average depth of branches ($r = 0.026$; $P = 0.8581$). Thus, deep nodes do not appear to be especially hard to reconstruct with strong support in snakes.

Almost all branches have at least one gene that strongly ($\geq$70%) supports or rejects the clade (47 of 48). There are 27 branches for which all genes strongly support the branch, and none strongly reject it. There are strongly supported conflicts between genes associated with 17 of the 48 branches. However, this does not include three branches on which there are one or more genes that strongly reject a clade but none that strongly support it. There are nine branches on which there are more genes strongly contradicting the branch than supporting it.

There is a significant correlation between branch length and the proportion of genes supporting the branch ($r = 0.849$; $P < 0.0001$); longer branches tend to be supported by more genes (Fig. 2b). There is also a strong correlation between the length of a branch and the proportion of genes that strongly support that branch ($r = 0.868$; $P < 0.0001$; Fig. 2c). There is a weak negative correlation between the length of a branch and the proportion of genes that strongly reject it ($r = -0.263$; $P = 0.0718$). There is extensive strongly supported conflict between genes with regard to some of the short branches, but many short branches lack such conflicts, and such conflicts are only entirely absent on the longest branches. We suspect that the weak relationship between branch length and strong incongruence is caused by very short branches having too little time to accumulate the mutations needed for strong support of the relevant clades in most of the separate gene trees. Thus, incongruent gene histories may be present but difficult to detect for the shortest branches but are absent at the longest branches; these two factors might together explain the hump-shaped pattern in the relationship between branch length and incongruent gene trees (Fig. 2c).

There is a strong correlation between branch length and the number of genes strongly supporting a clade minus the number strongly rejecting it ($r = 0.819$; $P < 0.0001$; Fig. 2d). On short branches, the number of genes strongly rejecting the branch may be equal to or exceed the number strongly supporting it. There is no significant correlation between our index of strongly supported conflict and branch lengths ($r = -0.106$; $P = 0.4695$).

Similarly, there is a strong correlation between the number of genes strongly supporting minus rejecting the clade and bootstrap support ($r = 0.796$; $P < 0.0001$) but not between our index of conflict and support ($r = 0.137$; $P = 0.3479$). Most branches in which the number of genes strongly rejecting a clade exceeds the number supporting it are weakly supported in the combined analysis (67%). Furthermore, all branches are strongly supported in which the number of genes strongly supporting a clade exceeds the number rejecting it by a value of three or more. Nevertheless, there are some branches that are strongly supported in the combined analysis despite having equal numbers of genes that strongly support and reject them and some that remain weakly supported despite having more genes that strongly support them than reject them (6%, or 3/48).

When strong congruence and incongruence are assessed using a more conservative criterion (bootstrap values $\geq$95%), the correlation between the length of a branch and the proportion of genes strongly supporting that branch remains significant (r = 0.819; $P < 0.0001$). However, there are no strongly supported conflicts between genes using this criterion; all the genes either strongly support the clade, weakly support the clade, or offer only weak support for alternate relationships. At the same time, nearly half of the branches (21 of 48) lack any genes that strongly support them using this criterion. Of these 21 branches that are not strongly supported by any genes, 11 had genes that strongly supported conflicting relationships based on the more lenient bootstrap criterion ($\geq$70%).
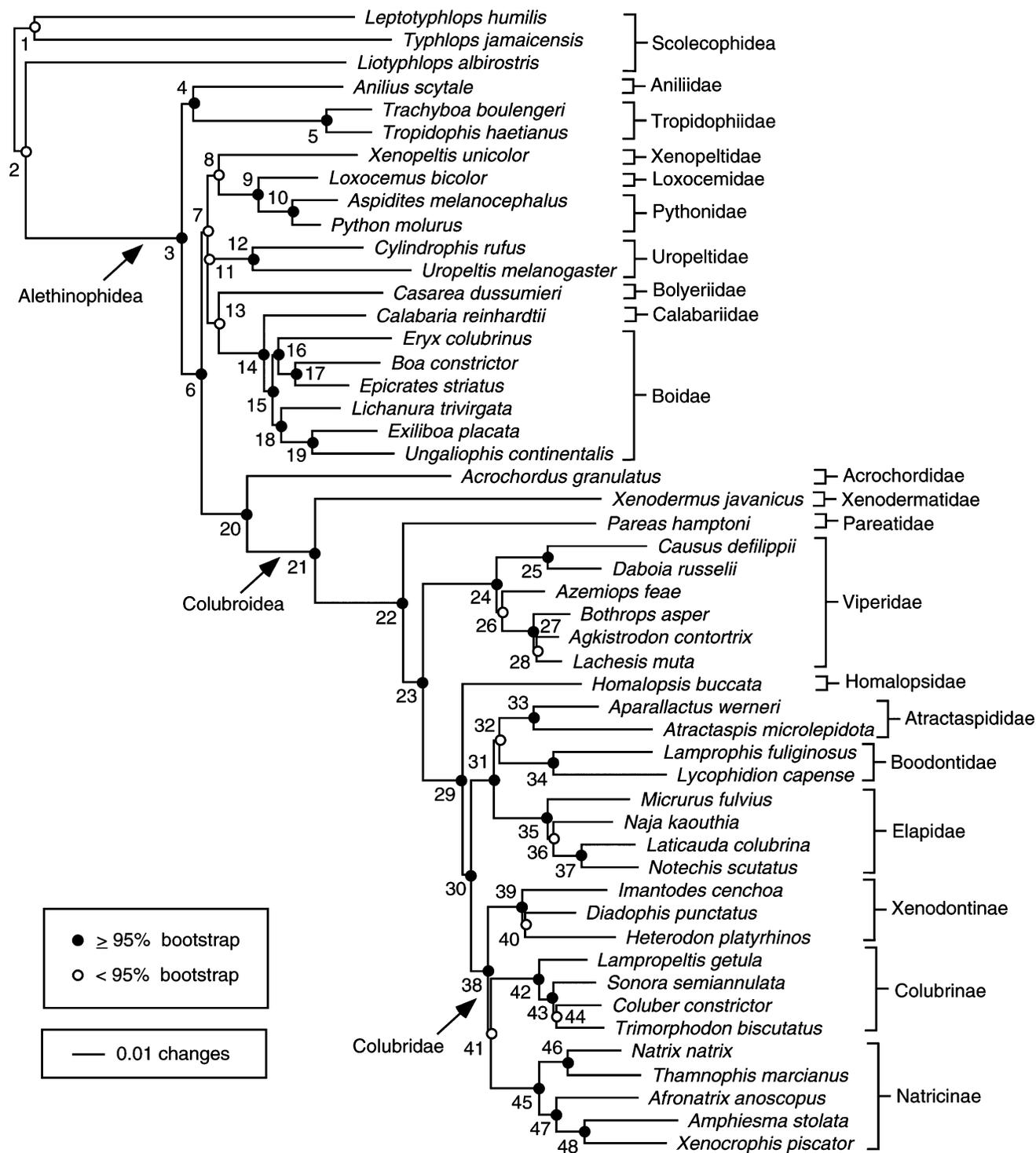
FIGURE 1.   Higher-level snake phylogeny based on combined maximum likelihood analysis of 20 nuclear loci (likelihood = −119,868.9391). Branch lengths and bootstrap values (indicated with open and filled circles) are estimated from the combined data. The parsimony topology, support values, and branch lengths are very similar. A summary of branch lengths, bootstrap values, and congruence among genes for each of the numbered clades is provided in Appendix 3 (http://www.systematicbiology.org). The seven nonsnake outgroup taxa are not shown. The higher taxon that each species belongs to is indicated at right; see Materials and Methods for a justification of the taxonomy used.
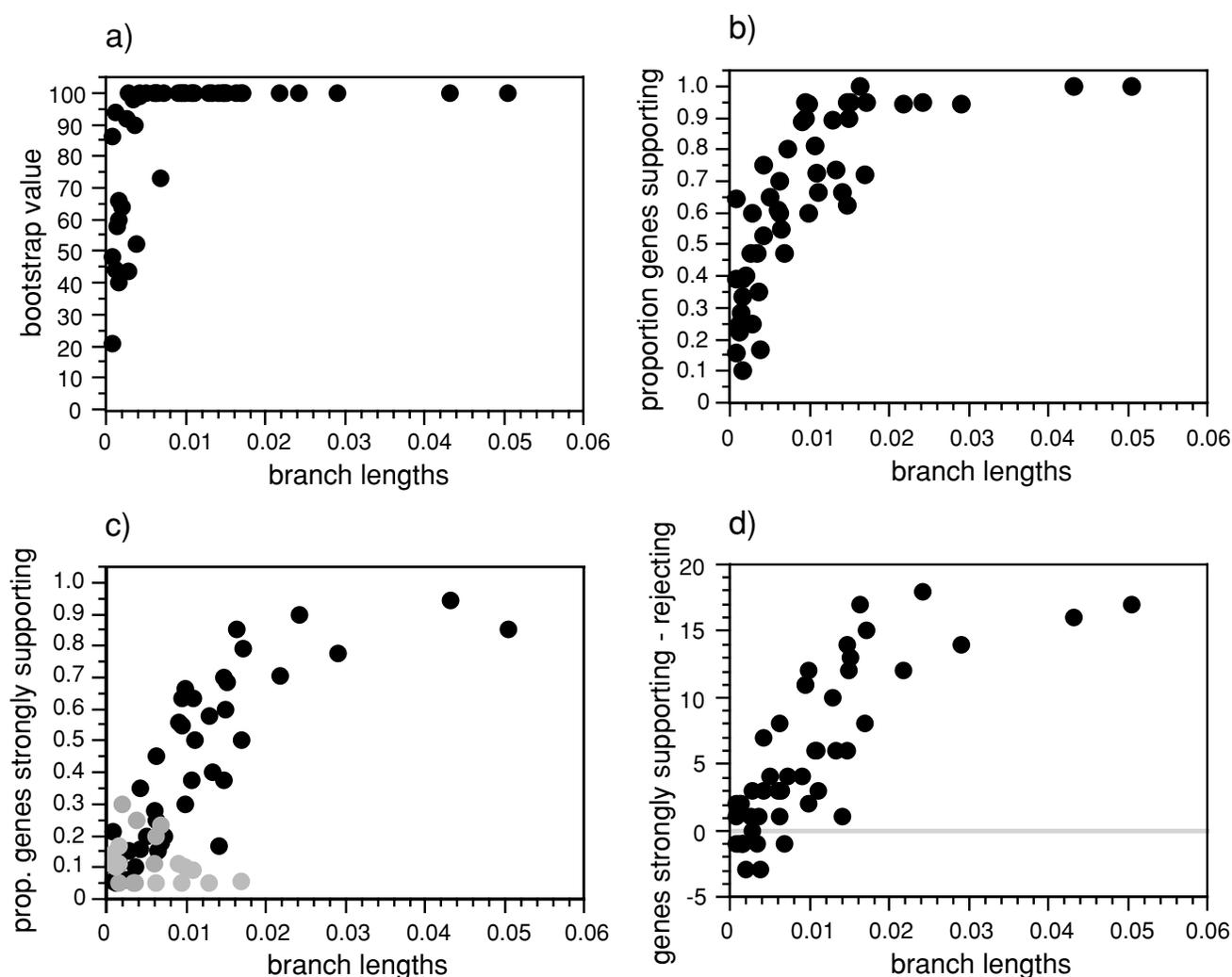
FIGURE 2.   Relationships between the estimated lengths of individual branches in the combined maximum likelihood analysis and (a) bootstrap support in the combined analysis; (b) proportion of genes supporting the clade in the separate analyses of the 20 genes; (c) proportion of genes strongly supporting the clade (≥70% bootstrap support) shown with black dots, with proportion strongly supporting an alternate relationship shown with gray dots; and (d) number of genes strongly supporting the clade—number strongly supporting an alternate clade.

DISCUSSION

Branch Lengths, Support, and Congruence

Many authors have suggested that short branches will be difficult for phylogenetic methods to reconstruct, in part because incongruent gene histories are expected to be common on very short branches (e.g., Slowinski, 2002; Poe and Chubb, 2004; Rokas and Carroll, 2006). However, to our knowledge, no previous studies have empirically examined the relationships between branch lengths, support, and congruence within a given clade. Our results from snakes show that longer branches tend to have higher bootstrap support and greater congruence among gene trees, whereas shorter branches have lower support and greater incongruence among genes. Many previous studies on this topic have focused on the idea of "hard polytomies," branches that are so short that they are essentially unresolvable polytomies (e.g., Slowinski,

2002; Poe and Chubb, 2004; Rokas and Carroll, 2006). Here we show that there is a continuum between hard polytomies and easily resolved clades that is generally related to branch lengths.

Short branches present at least three potential sources of error. First, very short branches may simply have too little time to accumulate substitutions. Thus, they may be difficult to resolve because there is too little informative variation. A second, related problem is that short branches may exacerbate the problem of long-branch attraction (Felsenstein 1978, 2004; Huelsenbeck 1995). Very long branches tend to accumulate parallel changes and if the intervening branches are very short, there will be too few actual synapomorphies to prevent the long-branches from being placed together (even if these long branches are not actually closely related). Third, short branches may be especially prone to the problem of discordance between the gene and species trees. Although

this discordance may occur through at least three processes (paralogy, introgression, incomplete lineage sorting; Maddison, 1997), discordance associated with short branches may be most likely to be caused by incomplete lineage sorting (ILS). When speciation splits one species into two, the gene histories of the two new species are thought to go through a gradual process of drift leading from polyphyly to paraphyly to reciprocal monophyly of the lineages (e.g., Neigel and Avise, 1986). However, if splits occur very rapidly, anomalous gene histories that are inconsistent with the species phylogeny (e.g., gene histories reflecting the time before species monophyly) may be retained over long evolutionary time scales (e.g., Poe and Chubb, 2004; Edwards et al., 2005).

Our results suggest that the first and third sources of error may apply to our data for snakes. On the shortest branches, relatively few genes yield strongly supported clades, either supporting or rejecting the clade, suggesting that there is too little time for these genes to accumulate mutations that would strongly support their gene histories. Long-branch attraction may be problematic for some clades within snakes (e.g., placement of *Liotyphlops* with Alethinophidia, rendering Scolecophidia paraphyletic; Fig. 1). However, we found that bootstrap support in the combined analysis was related to the length of the individual branches and not the depth of branches within the tree (but note that branch depth might be far more important in groups that are much older than snakes [e.g., Rokas et al., 2005] or for genes that are more fast-evolving [e.g., mitochondrial genes]). In contrast, we found evidence suggesting considerable incongruence among gene trees for many of the short branches. In some cases, there were many more genes that strongly rejected a clade than supported it. Furthermore, we found that low bootstrap support was significantly associated with this incongruence (i.e., number of genes strongly supporting a clade minus those rejecting it).

If the primary cause of weak support for the clades in this study is related to incongruence among gene histories, then this may be a very difficult problem to resolve. The problem of weak support alone can potentially be solved simply by adding more genes (especially genes that are evolving more quickly). Long-branch attraction can potentially be avoided by adding species that can subdivide and shorten the long branches (e.g., Poe, 2003). However, a solution to the problem of incongruent gene histories on short branches is not immediately obvious. If branches are very short, then the different gene histories may occur at nearly equal frequencies, and the gene history that is sampled most frequently may simply reflect stochastic sampling (e.g., Rokas and Carroll, 2006). In some cases, a misleading gene tree may even be significantly more common (Degnan and Rosenberg, 2006). Adding ingroup taxa should generally reduce branch lengths on average, and so there is no reason to expect this to help (e.g., Edwards et al., 2005). In fact, in our analyses of higher-level snake phylogeny, we found that support values for two clades improved considerably after removing the bolyeriid *Casarea dussumieri* (clade 7 goes from 40% to 80%, clade 11 goes from 21% to 85%,

and weakly supported clade 13 [44%] is eliminated). This species was sequenced for 15 of the 20 genes but is nevertheless ambiguously placed on our tree. Sampling additional individuals within a species seems unlikely to be helpful at this deep phylogenetic scale; presumably, the discordant gene histories are no longer maintained as polymorphisms within a species. The most promising solution may be to utilize phylogenetic methods that can incorporate differences in gene histories to help resolve species phylogeny (e.g., Edwards et al., 2007), although such methods are not yet widely available (i.e., functioning versions of programs to implement such methods were not available to us at the time this paper was submitted).

Limitations and Caveats

Our study has several limitations that should be pointed out, although we believe that these problems do not necessarily invalidate our conclusions. First, because our study is based on empirical data, we do not know the true species phylogeny. Although we consider clades with strong bootstrap values that are supported by many genes to be well resolved, it is possible that these clades are nevertheless incorrect, either because of long-branch attraction or because the gene trees have converged on an incorrect answer (e.g., Degnan and Rosenberg, 2006; Kubatko and Degnan, 2007). However, the correlations we found between branch length, support, and incongruence make these possibilities seem unlikely (e.g., in our study, gene trees are highly concordant on longer branches, and concordantly misleading gene trees are not expected on long branches). In theory, we could have performed simulations to address these questions instead, but our overall goal was to test these relationships empirically.

Second, we acknowledge that even our estimates of gene histories are subject to some error. Although we considered clades with relatively strong support ($\geq$70%) to indicate concordant or discordant gene trees, this need not always be true. In fact, we found no cases of strongly supported incongruence between genes where the conflicting clades each had bootstrap support $\geq$95%. One optimistic interpretation of this discrepancy is that there are no real conflicts between gene trees using this more stringent criterion (e.g., Taylor and Piel, 2004). Unfortunately, we think it is more likely that for short branches there is too little time for the gene trees to become strongly supported, regardless of whether they are concordant or discordant with the species tree (as indicated by the fact that 21 of 48 of branches lack any genes that strongly support them with bootstrap values $\geq$95%). We acknowledge that these issues of weak support and gene tree congruence might also be resolved by using longer fragments and faster evolving genes (i.e., both yielding more informative characters per gene), such that all genes had stronger support for their underlying trees. It should also be noted that there could be other causes of strongly supported conflict between topologies apart from ILS, including paralogy, introgression, or even lab error. Furthermore,

the level of congruence and support among genes is not the only factor impacting support in the combined analysis. For example, many genes that weakly support or reject a clade might, when taken together, have a strong influence on the resolution and level of support for a clade in the combined analysis (e.g., Gatesy and Baker, 2005).

Third, we have not sampled all genes for all taxa. This makes calculation of some relevant indices complicated (e.g., we consider a gene as supporting a clade, even though not every species in that clade has been included) and might influence the estimation of branch lengths in some cases. However, we think that our major results are not artifacts of incomplete sampling or missing data. In particular, we strongly suspect that weakly supported branches are not necessarily associated with incomplete taxa (see also Wiens et al. 2005). For example, *Uropeltis melanogaster*, one of the species for which we have the fewest genes sampled (6 of 20), is placed next to its sister taxon on the combined-data tree with a bootstrap value of 100%. We also remind readers that the levels of missing data in our study do not appear to be unusual for empirical studies (e.g., Phillippe et al., 2004; Rokas et al., 2005). Finally, branch lengths from the combined data seem to be strongly correlated with those estimated from the individual genes, suggesting that these combined-data branch lengths reflect the underlying branch lengths of the species tree rather than artifacts of missing data.

We acknowledge that although we have examined the correlations between branch lengths, support, and congruence, we have not necessarily established causal relationships among these variables. We assume that the time between branching events will generally determine levels of support and congruence; it seems unlikely that the converse is true. Furthermore, levels of support in the combined analysis cannot influence congruence among separately analyzed genes. However, we have not fully established a causal relationship between levels of incongruence and levels of support.

Implications for Snake Phylogeny

We consider the phylogeny of snakes presented here to be somewhat preliminary. In the future, we will include additional characters (e.g., morphology, mitochondrial DNA sequences) and many additional taxa for which a smaller subset of characters have been sampled. Nevertheless, our phylogeny provides strong support for many interesting and controversial phylogenetic results suggested in previous studies with a smaller sampling of genes and/or taxa (many of which were discordant between previous studies). Many of the relationships differ strongly from hypotheses based on morphology, particularly the widely cited study by Cundall et al. (1993) and more recent studies by Lee and Scanlon (2002) and Lee et al. (2007).

First, our results support the polyphyly of the traditional Tropidophiidae (*Exiliboa, Trachyboa, Tropidophis, Ungaliophis*), with Aniliidae as sister taxon of *Tropidophis + Trachyboa* at the base of Alethinophidea and *Exili-*

boa and *Ungaliophis* placed in a more derived position with boine and erycine boids (e.g., Slowinski and Lawson 2002; Wilcox et al., 2002; Noonan and Chippindale, 2006; Vidal et al., 2007b). Traditional hypotheses based on morphology (e.g., Cundall et al., 1993) supported a monophyletic Tropidophiidae, but there has also been morphological support (Zaher, 1994; Lee and Scanlon, 2002; Lee et al., 2007) for polyphyly of the family and recognition of Tropidophiidae (for *Trachyboa* and *Tropidophis*) and Ungaliophiidae (for *Exiliboa* and *Ungaliophis*). Morphological results also differ in placing tropidophiids and ungaliophiids relatively close to the advanced snakes (e.g., Cundall et al., 1993; Zaher, 1994; Lee and Scanlon, 2002; Lee et al., 2007).

Second, our results support the placement of *Xenopeltis* and *Loxocemus* with Pythonidae (Slowinski and Lawson, 2002; Wilcox et al., 2002; Lee et al., 2007 Vidal et al., 2007b). Analyses based on morphological data have placed pythonids with boids rather than with xenopeltids and loxocemids (e.g., Cundall et al., 1993; Lee and Scanlon 2002; Lee et al., 2007). One molecular analysis (Noonan and Chippindale, 2006) placed *Xenopeltis* with *Cylindrophis* rather than *Loxocemus* and the pythons.

Third, our results support placement of *Cylindrophis* (Cylindrophidae of some authors) and *Uropeltis* (Uropeltidae) as sister taxa (Wilcox et al., 2002; Lawson et al., 2004). Some analyses of morphological data do not place these genera as sister taxa (e.g., Lee et al., 2007).

Fourth, our results support nonmonophyly of the traditionally recognized erycine boids (e.g., *Eryx, Lichanura*) relative to boine boids and the ungaliophine tropidophiids (Noonan and Chippindale, 2006; Vidal et al., 2007b). Similarly, we place *Calabaria* (Calabariidae) as sister taxon of the clade including the former ungaliophiine tropidophiids (*Exiliboa, Ungaliophis*), Boinae, and Erycinae (now Boidae; Vidal et al., 2007b), whereas morphological data typically place *Calabaria* as the sister taxon of the Erycinae (e.g., Lee et al., 2007). These surprising results suggest that there may have been repeated evolution of a similar burrowing ecomorph of alethinophidian snakes in different continental regions (e.g., *Calabaria* in West Africa; *Eryx* in northern and eastern Africa, Europe, and Asia; and *Charina* and *Lichanura* in North America).

Finally, our results do not support the monophyly of Scolecophidia, although this clade has been supported in most previous morphological and molecular studies. Specifically our representative of Anomalepidae does not group with our representatives of Leptotyphlopidae and Typhlopidae. However, this result is poorly supported, and we suspect that analyses including additional taxon sampling for this clade will break up the long branches among the included species and support its monophyly. Most previous studies supporting scolecophidian monophyly have not included an anomalepid, and so are not strictly comparable (e.g., Slowinski and Lawson, 2002; Wilcox et al., 2002), but the analysis of Lee et al. (2007) did include all three families for both nuclear and mitochondrial genes and supported scolecophidian monophyly.

Not all of our results are discordant with previous phylogeny or taxonomy. For example, we find strong support for the placement of Acrochordidae as sister group to Colubroidea as have many previous morphological (e.g., Cundall et al., 1993; Lee and Scanlon, 2002; Lee et al., 2007) and molecular studies (e.g., Slowinski and Lawson, 2002; Wilcox et al., 2002; Lee et al., 2007; Vidal et al., 2007b). Finally, our hypothesis for the basal relationships within Colubroidea are largely concordant with other recent molecular studies (Lawson et al., 2005; Vidal et al., 2007a), including the successively derived positions of *Xenodermus*, *Pareas*, Viperidae, Homolopsidae; a clade consisting of Elapidae, Atractaspididae, and Boodontidae; and a clade including colubrine, natricine, and xenodontine colubrids. Although these results are somewhat discordant with traditional taxonomy, that taxonomy was not based on a rigorous phylogenetic analysis of morphological data. We also found strong support for numerous groups that have been recognized traditionally, including elapids, viperids, colubrines, natricines, and xenodontines.

REFERENCES

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

Brown, J. M., and A. R. Lemmon. 2007. The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. Syst. Biol. 56:643–655.

Cundall, D., V. Wallach, and D. A. Rossman. 1993. The systematic relationships of the snake genus *Anomochilus*. Zool. J. Linn. Soc. 109:275–299.

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

Driskell, A. C., C. Ané, J. G. Burleigh, M. M. McMahon, B. C. O'Meara, and M. J. Sanderson. 2004. Prospects for building the Tree of Life from large sequence databases. Science 306:1172–1174.

Edwards, S. V., W. B. Jennings, and A. M. Shedlock. 2005. Phylogenetics of modern birds in the era of genomics. Proc. R. Soc. Lond. B 272:979–992.

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

Estes, R., K. de Queiroz, and J. Gauthier. 1988. Phylogenetic relationships within Squamata. Pages 119–281 *in* Phylogenetic relationships of the lizard families (R. Estes and G. Pregill, eds.). Stanford University Press, Stanford, California.

Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. 27:401–410.

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

Felsenstein, J. 2004. Inferring phylogenies. Sinauer Associates, Sunderland, Massachusetts.

Gatesy, J., and R. Baker. 2005. Hidden likelihood support in genomic data: Can forty-five wrongs make a right? Syst. Biol. 54:483–492.

Gower, D. J., N. Vidal, J. N. Spinks, and C. J. McCarthy. 2005. The phylogenetic position of Anomochilidae (Reptilia: Serpentes): First evidence from DNA sequences. J. Zool. Syst. Evol. Res. 43:315–320.

Hallstrom, B. M., M. Kullberg, M. A. Nilsson, and A. Janke. 2007. Phylogenomic data analyses provide evidence that Xenarthra and Afrotheria are sister groups. Mol. Biol. Evol. 24:2059–2068.

Huelsenbeck, J. P. 1995. The performance of phylogenetic methods in simulation. Syst. Biol. 44:17–48.

Huelsenbeck, J. P., and F. Ronquist. 2001. MrBayes: Bayesian inference of phylogeny. Bioinformatics 17:754–755.

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

Lawson, R., J. Slowinski, and F. T. Burbrink. 2004. A molecular approach to discerning the phylogenetic placement of the enigmatic snake *Xenophidion schaeferi* among the Alethinophidia. J. Zool. 263:285–294.

Lawson, R., J. B. Slowinski, B. I. Crother, and F. T. Burbrink. 2005. Phylogeny of the Colubroidea (Serpentes): New evidence from mitochondrial and nuclear genes. Mol. Phylogenet. Evol. 37:581–601.

Lee, M. S. Y. 1998. Convergent evolution and character correlation in burrowing reptiles: Towards a resolution of squamate relationships. Biol. J. Linn. Soc. 65:369–453.

Lee, M. S. Y., A. F. Hugall, R. Lawson, and J. D. Scanlon. 2007. Phylogeny of snakes (Serpentes): Combining morphological and molecular data in likelihood, Bayesian, and parsimony analyses. Syst. Biodiv. 4:371–389.

Lee, M. S. Y., and J. D. Scanlon. 2002. Snake phylogeny based on osteology, soft anatomy, and ecology. Biol. Rev. 77:333–401.

Maddison, D. R., and W. P. Maddison. 2000. MacClade 4.0. Sinauer Associates, Sunderland, Massachusetts.

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

Neigel, J. E., and J. C. Avise. 1986. Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation. Pages 515–534 *in* Evolutionary processes and theory (E. Nevo and S. Karlin, eds.). Academic Press, New York.

Noonan, B. P., and P. T. Chippindale. 2006. Dispersal and vicariance: The complex evolutionary history of boid snakes. Mol. Phylogenet. Evol. 40:347–358.

Nylander, J. A. A. 2004. MrModelTest 2.0. Program distributed by the author. Evolutionary Biology Centre, Uppsala University (http://www.ebc.uu.se/systzoo/staff/nylander.html).

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

Philippe, H., N. Lartillot, and H. Brinkmann. 2005. Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa and Protostomia. Mol. Biol. Evol. 22:1246–1253.

Philippe, H., E. A. Snell, E. Bapteste, P. Lopez, P. W. H. Holland, and D. Casane. 2004. Phylogenomics of eukaryotes: Impact of missing data on large alignments. Mol. Biol. Evol. 21:1740–1752.

Poe, S. 2003. Evaluation of the strategy of long-branch subdivision to improve the accuracy of phylogenetic methods. Syst. Biol. 52:423–428.

Poe, S., and A. L. Chubb. 2004. Birds in a bush: Five genes indicate explosive evolution of avian orders. Evolution 58:404–415.

Pough, F. H., R. M. Andrews, J. E. Cadle, M. L. Crump, A. H. Savitzky, and K. D. Wells. 2004. Herpetology, 3rd edition. Pearson-Prentice Hall, Upper Saddle River, New Jersey.

Rokas, A., and S. B. Carroll. 2006. Bushes in the Tree of Life. PLoS Biol. 4:e352.

Rokas, A., D. Krueger, and S. B. Carroll. 2005. Animal evolution and the molecular signature of radiations compressed in time. Science 310:1933–1938.

Rokas, A., B. L. Williams, N. King, and S. B. Carroll. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature 425:798–804.

Slowinski, J. B. 2001. Molecular polytomies. Mol. Phylogenet. Evol. 19:114–120.

Slowinski, J. B., and R. Lawson. 2002. Snake phylogeny: Evidence from nuclear and mitochondrial genes. Mol. Phylogenet. Evol. 24:194–202.

Stamatakis, A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.

Swofford, D. L. 2002. PAUP*: Phylogenetic analysis using parsimony (*and other methods). Version 4.0b10. Sinauer, Sunderland, Massachusetts.

Takahata, N. 1989. Gene genealogy in three related populations: Consistency probability between gene and population trees. Genetics 122:957–966.

Takezaki, N., F. Figueroa, Z. Zaleska-Rutczynska, N. Takahata, and J. Klein. 2004. The phylogenetic relationship of tetrapod, coelacanth, and lungfish revealed by the sequences of 44 nuclear genes. Mol. Biol. Evol. 21:1512–1524.

Taylor, D. J., and W. H. Piel. 2004. An assessment of accuracy, error, and conflict with support values from genome-scale phylogenetic data. Mol. Biol. Evol. 21:1534–1537.

Townsend, T., A. Larson, E. J. Louis, and J. R. Macey. 2004. Molecular phylogenetics of Squamata: The position of snakes, amphisbaenians, and dibamids, and the root of the squamate tree. Syst. Biol. 53:735–757.

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

Vidal, N., A.-S. Delmas, P. David, C. Cruaud, A. Couloux, and S. B. Hedges. 2007a. The phylogeny and classification of caenophidian snakes inferred from seven nuclear protein-coding genes. C. R. Biologies 330:182–187.

Vidal, N., A.-S. Delmas, and S. B. Hedges. 2007b. The higher-level relationships of alethinophidian snakes inferred from seven nuclear and mitochondrial genes. Pages 27–33 in Biology of the boas and pythons (R. W. Henderson and R. Powell, eds.). Eagle Mountain Publishing, Eagle Mountain, Utah.

Vidal, N., and S. B. Hedges. 2005. The phylogeny of squamate reptiles (lizards, snakes, and amphisbaenians) inferred from nine nuclear protein-coding genes. C. R. Biologies 328:1000–1008.

Wiens, J. J. 2003. Missing data, incomplete taxa, and phylogenetic accuracy. Syst. Biol. 52:528–538.

Wiens, J. J., J. W. Fetzner, C. L. Parkinson, and T. W. Reeder. 2005. Hylid frog phylogeny and sampling strategies for speciose clades. Syst. Biol. 54:719–748.

Wilcox, T. P., D. J. Zwickl, T. A. Heath, and D. M. Hillis. 2002. Phylogenetic relationships of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. Mol. Phylogenet. Evol. 25:361–371.

Zaher, H. 1994. Les Tropidopheoidea (Serpentes; Alethinophidea) sont-ils réellemeent monophylétiques? Arguments en faveur de leur polyphylétisme. C. R. Acad. Sci. Paris 317:471–478.

An Okinawan pitviper (*Ovophis okinavensis*) from Amami Island in the Ryuku Archipelago of Japan. Photo by John Wiens.