Title
Evaluating Fossil Calibrations for Dating Phylogenies in Light of Rates of Molecular Evolution: A Comparison of Three Approaches

Journal
Systematic Biology, 61(1)

ISSN
1063-5157

Authors
Lukoschek, Vimoksalehi
Keogh, J Scott
Avise, John C

Publication Date
2012

DOI
10.1093/sysbio/syr075

Syst. Biol. 61(1):22–43, 2012

DOI:10.1093/sysbio/syr075
Advance Access publication on August 11, 2011

Evaluating Fossil Calibrations for Dating Phylogenies in Light of Rates of Molecular Evolution: A Comparison of Three Approaches

VIMOKSALEHI LUKOSCHEK[1,2,*], J. SCOTT KEOGH[3], AND JOHN C. AVISE[1]

[1]*Department of Ecology and Evolutionary Biology, University of California at Irvine, Irvine, CA 92697, USA;* [2]*Australian Research Council Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, QLD 4811, Australia; and* [3]*Division of Evolution, Ecology and Genetics, Research School of Biology, The Australian National University, Canberra, ACT 0200, Australia;*
[*]*Correspondence to be sent to: ARC Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, QLD 4811, Australia; E-mail: vimoksalehi.lukoschek@jcu.edu.au.*

Received 17 November 2010; revises returned 21 January 2011; accepted 4 March 2011
Associate Editor: Frank E. Anderson

Abstract.—Evolutionary and biogeographic studies increasingly rely on calibrated molecular clocks to date key events. Although there has been significant recent progress in development of the techniques used for molecular dating, many issues remain. In particular, controversies abound over the appropriate use and placement of fossils for calibrating molecular clocks. Several methods have been proposed for evaluating candidate fossils; however, few studies have compared the results obtained by different approaches. Moreover, no previous study has incorporated the effects of nucleotide saturation from different data types in the evaluation of candidate fossils. In order to address these issues, we compared three approaches for evaluating fossil calibrations: the single-fossil cross-validation method of Near, Meylan, and Shaffer (2005. Assessing concordance of fossil calibration points in molecular clock studies: an example using turtles. Am. Nat. 165:137–146), the empirical fossil coverage method of Marshall (2008. A simple method for bracketing absolute divergence times on molecular phylogenies using multiple fossil calibration points. Am. Nat. 171:726–742), and the Bayesian multicalibration method of Sanders and Lee (2007. Evaluating molecular clock calibrations using Bayesian analyses with soft and hard bounds. Biol. Lett. 3:275–279) and explicitly incorporate the effects of data type (nuclear vs. mitochondrial DNA) for identifying the most reliable or congruent fossil calibrations. We used advanced (Caenophidian) snakes as a case study; however, our results are applicable to any taxonomic group with multiple candidate fossils, provided appropriate taxon sampling and sufficient molecular sequence data are available. We found that data type strongly influenced which fossil calibrations were identified as outliers, regardless of which method was used. Despite the use of complex partitioned models of sequence evolution and multiple calibrations throughout the tree, saturation severely compressed basal branch lengths obtained from mitochondrial DNA compared with nuclear DNA. The effects of mitochondrial saturation were not ameliorated by analyzing a combined nuclear and mitochondrial data set. Although removing the third codon positions from the mitochondrial coding regions did not ameliorate saturation effects in the single-fossil cross-validations, it did in the Bayesian multicalibration analyses. Saturation significantly influenced the fossils that were selected as most reliable for all three methods evaluated. Our findings highlight the need to critically evaluate the fossils selected by data with different rates of nucleotide substitution and how data with different evolutionary rates affect the results of each method for evaluating fossils. Our empirical evaluation demonstrates that the advantages of using multiple independent fossil calibrations significantly outweigh any disadvantages. [Bayesian dating; Caenophidia; cross-validation; fossil calibrations; Hydrophiinae; molecular clock; nucleotide saturation; snakes.]

Ideally, molecular clock calibrations are obtained from accurately dated fossils that can be assigned to nodes with high phylogenetic precision (Graur and Martin 2004), but reality is generally far from this ideal because of a number of important problems. The incomplete and imperfect nature of the fossil record means that fossils necessarily only provide evidence for the minimum age of a clade. Many clades will be considerably older than the oldest known fossil; thus, nodes may be constrained to erroneously young ages (Benton and Ayala 2003; Donoghue and Benton 2007; Marshall 2008). Incorrect fossil dates also arise from experimental errors in radiometric dating of fossil-bearing rocks or incorrectly assigning fossils to a specific stratum. In addition, misinterpreted character state changes can result in the taxonomic misidentification of fossils or their incorrect placement on the phylogeny (Lee 1999). Ideally, a fossil would date the divergence of two descendant lineages from a common ancestor. In reality, however, fossils rarely represent specific nodes but rather points along a branch (Lee 1999; Conroy and van Tuinen 2003). Thus, although a fossil may appear to be ancestral to a clade, it is impossible to determine how much earlier the fossil existed than the clade's common ancestor. Fossils also may be incorrectly assigned to the crown rather than the stem of a clade (Doyle and Donoghue 1993; Magallon and Sanderson 2001). The most useful fossils are, therefore, geologically well dated, preserved with sufficient morphological characters to be accurately placed on a phylogenetic tree and temporally close to an extant node rather than buried within a stem lineage (van Tuinen and Dyke 2004). However, the fossil records of many, if not most, taxonomic groups fall far short of these criteria. As such, several methods have been developed for evaluating candidate fossil calibrations in order to: determine their internal consistency and identify outliers (Near et al. 2005), identify lineages with the best fossil coverage and identify outliers (Marshall 2008), and evaluate alternative placements of fossils (Rutschmann et al. 2007; Sanders and Lee 2007).

However, fossil calibrations are not the only difficulty in molecular dating. Other factors also contribute to inaccurately calibrated molecular clocks including: incorrectly specified models of evolution (Brandley et al. 2011), inappropriate modelling of rate heterogeneity among lineages (Sanderson 1997; Drummond et al.

2006), and unbalanced taxon sampling potentially resulting in node-density artifacts (Hugall and Lee 2007). In addition, choice of genetic data or gene region can strongly affect estimated divergences (Benton and Ayala 2003). For example, in rapidly evolving genes, such as mitochondrial DNA, saturation has been shown to have the effect of compressing basal branches and artificially pushing shallow nodes toward basal nodes, resulting in overestimated divergence dates (Hugall and Lee 2004; Townsend et al. 2004; Hugall et al. 2007; Phillips 2009). However, the nature of the bias is complicated. For example, underestimating the true rate of hidden substitution results in tree compression: However, if the rate of hidden substitutions were to be overestimated, the reverse would be true. These effects are further complicated by the calibration placement. For example, if only deep splits are calibrated, then recent nodes will be biased to be younger under tree extension and older under tree compression. Slowly evolving genes, as are typical for nuclear DNA, are less prone to such saturation effects; however, nuclear DNA data are not completely immune to these issues; problems of saturation also can emerge for slowly evolving nuclear loci if deeper divergences are being investigated. More importantly, although the effects of saturation have been documented for estimating divergence times (Brandley et al. 2011; Hugall and Lee 2004; Townsend et al. 2004; Hugall et al. 2007; Phillips 2009), the effects of saturation on different approaches for evaluating candidate fossil calibrations have yet to be explored.

Caenophidia ("advanced snakes" comprising acrochordids, elapids, viperids, and colubrids) is a group with a controversial fossil record. Indeed, recent papers using calibrated molecular clocks to date divergences among advanced snake clades highlight the extent of controversy about the placements of certain fossils (Wuster et al. 2007, 2008; Sanders and Lee 2008; Sanders et al. 2008; Kelly et al. 2009). In part, this controversy exists because of the relatively poor nature of the snake fossil record. Well preserved and relatively complete caenophidian fossils date back no further than the Miocene (Rage 1984) and often belong to extant genera (Rage 1988; Szyndlar and Rage 1990, 1999), thus are of little value as calibration points for most studies. Earlier caenophidian fossils mostly comprise isolated vertebrae, the taxonomic affinities of which have been strongly debated (McDowell 1987; Rage 1987). Perhaps, the most controversial calibrations concern the origin of caenophidian snakes themselves, which has been assigned dates of 38 (34–48) myr (Sanders and Lee 2008; Kelly et al. 2009), 57 (47–140) myr (Wuster et al. 2008), and >65 myr (Noonan and Chippindale 2006a, 2006b), based on different interpretations of the fossil record (Table 1). As such, very different dates have been used to calibrate the caenophidian molecular clock (Nagy et al. 2003; Guicking et al. 2006; Burbrink and Lawson 2007; Wuster et al. 2007, 2008; Alfaro et al. 2008; Sanders and Lee 2008; Kelly et al. 2009).

In this paper, we use advanced snakes as a test case to compare three previously published methods for evaluating fossil calibrations: the single-fossil cross-validation method of Near et al. (2005), the empirical fossil coverage method of Marshall (2008), and the Bayesian multicalibration method of Sanders and Lee (2007) and explicitly evaluate the effects of nucleotide saturation on the results of each method. Briefly, the single-fossil cross-validation approach (Near et al. 2005) evaluates candidate fossils, including the alternative ages or placements of fossils at some calibrated nodes, with the aim of identifying a number of plausible reliable calibration sets. The approach of Marshall (2008) aims to identify candidate calibrations with the best fossil coverage and then tests whether these fossils are potential outliers. Finally, the Bayesian multicalibration approach evaluates one or more alternative calibrations in a set by comparing the Bayesian prior and posterior probabilities (PPs) at fossil-calibrated nodes (Sanders and Lee 2007). We explicitly evaluate the effects of using sequence data with different rates of molecular evolution on the best fossils identified by each method using the same mitochondrial and nuclear sequence data set (each with identical taxon sampling) for each method. In addition, we evaluate whether saturation effects can be ameliorated by 1) removing the third codon position of the mitochondrial coding regions and 2) analyzing a combined nuclear and mitochondrial data set. Our study focused on testing alternative placements or ages of controversial fossil calibrations (as is typical for groups with poor fossil records); however, our approach is relevant for any situation where numerous candidate fossil calibrations exist.

Materials and Methods

Fossil Calibrations, Taxon Sampling, Molecular Data, Convergence Diagnostics, and Saturation Plots

Colubroid classification is in flux (Vidal et al. 2007). We use the traditional colubroid classification as comprising viperids, elapids, and colubrids, including colubrid subfamilies recently elevated to higher taxonomic ranks (McDowell 1987; Rage 1987; Lawson et al. 2005). Forty-eight taxa (40 caenophidian and 8 henophidian taxa) were chosen based on the availability of nuclear and mitochondrial sequences (Appendix Table A1) and to appropriately span the various fossil calibrations tested. We specifically selected fossil calibrations that often have been used to date recent caenophidian divergences (Nagy et al. 2003; Guicking et al. 2006; Burbrink and Lawson 2007; Wuster et al. 2007, 2008; Alfaro et al. 2008; Sanders and Lee 2008; Kelly et al. 2009), and for which we could construct nuclear and mitochondrial data sets with appropriate taxon sampling. Details of the fossil calibrations evaluated are given in Table 1. We constructed nuclear and mitochondrial data sets, each with identical taxon sampling, using >100 novel sequences generated for this study and published sequences obtained from GenBank (Appendix Table A1). The mitochondrial data comprised 16S rRNA (454 bp), NADH dehydrogenase subunit 4 (ND4) (672 bp), and cytochrome *b* (1095 bp), and the nuclear data comprised

TABLE 1. Details of fossils tested using three approaches for evaluating candidate fossil calibrations

Fossil calibrations	Node	Calibration priors			Reference
		Mean (95% HPD)	Ln mean (SD)	Zero offset	
Scolecophidians versus alethinophidians	Root	97 (92–100)	2.00 (0.85)	90	The divergence between the Scolecophidia and the Alethinophidia was calibrated based on the earliest alethinophidian fossils: two *Coniphis* vertebrate from Utah from the upper Albanian/lowermost Cenomanian (97–102 Ma) (Gardner and Cifelli 1999) and six *Coniphis* trunk vertebrate from the Cenomanian (94–100 Ma) in Sudan (Rage and Werner 1999). Gardner and Cifelli (1999, p. 95) note that the approximately contemporaneous occurrence of *Coniophis* fossils in geographically distant Sudan and Utah suggests that the Alethenophidia–Scoleophidia split occurred prior to the Cenomanian (99 Ma). This calibration also was used by Kelly et al. (2009) and Sanders and Lee (2008).
Henophidians versus caenophidians	1	68 (65–85)	1.00 (1.20)	65	The divergence between the Henophidia (boids) and Caenophidia (advanced snakes) has been dated using the fossils assigned to the Booidae. Noonan and Chippindale (2006a) dated the Henophidia–Caenophidia split at >75 Ma based on the earliest probable boid fossils from the latest Cretaceous (65–85 Ma) from South America. However, the taxonomic affinities of these older vertebrae were not easy to assign (Albino 2000; Rage 2001). The first vertebrae that are undoubtedly booids occur in the mid-Palaeocene (58.5–56.5 Ma). These vertebrae are assigned to the extant genus *Corallus* (Boinae) and occur contemporaneously with fossil vertebrae from several other boine taxa (Rage 2001) indicating that the Boinae were a separate phylogenetic entity by the mid-Palaeocene, and that extant boine lineages originated early in the Tertiary or late Cretaceous (Rage et al. 2001, p. 146). Based on these fossils, we constrained the Henophidia–Caenophidia split as occurring 68 (65–85) Ma.
Acrochordids versus colubroids	2-*Set A*	38 (34–48)	1.40 (0.75)	34	The MRCA of the Caenophidia (Acrochordidae versus Colubroidea) has been ascribed a range of dates based on different interpretations of the taxonomic affinities of certain fossils. These fossils include six vertebrae from the Cenomanian (93–96 Ma) in Sudan that were assigned to the Colubroidea (Rage and Werner 1999); the oldest *Nigerophis* (Nigeropheidae) vertebra found in Paleocene marine deposits in Nigeria (56–65 Ma) (Rage 1984, 1987); and the oldest undisputed colubroid fossil from the late middle Eocene (37–39 Ma) (Head et al. 2005). We tested the effects of constraining this node with the three different divergence dates previously used based on these fossils: 38 (34–48) myr (Sanders and Lee 2008; Kelly et al. 2009); 57 (47–140) myr (Wuster et al. 2007); and 65 (63–80) myr (Noonan and Chippindale 2006a, 2006b) used >65 myr).
Acrochordids versus colubroids	2-*Set B*	57 (47–140)	2.50 (1.25)	45	
Acrochordids versus colubroids	2-*Set C*	65 (63–80)	1.10 (1.10)	62	

Continued

TABLE 1. (*Continued*)

Fossil calibrations	Node	Calibration priors			Reference
		Mean (95% HPD)	Ln mean (SD)	Zero offset	
Viperids versus colubrids + elapids	3-*Set A*	34 (31–43)	1.40 (0.70)	30	The MRCA of the Colubroidea (viperids versus colubrids and elapids) has been dated based on the oldest colubrid fossils from the Late Eocene (34–37 Ma) in Thailand (Rage et al. 1992); however, the oldest putative colubroid fossils from the Cenomanian (93–96 Ma) (Rage and Werner 1999) have also been used to constrain the upper bound of this clade. Head et al.(2005, p. 249) and Parmley and Holman (2003, p. 6) argue that taxonomically and geographically divergent colubrid fossils found the late Eocene in Krabi Basin (Rage et al. 1992), Pondaung (Head et al. 2005), and North America (Parmley and Holman 2003) indicate that colubroids had started diverging pre-Late Eocene, possibly even in the early Paleogene (43–60 Ma). We tested two divergence dates previously used: 34 (31–43) myr (Sanders and Lee 2008; Kelly et al. 2009; Wiens et al. 2006); and 47 (40–95) myr (Wuster et al. 2008). We also tested a constraint of 40 (37–60) myr based on the geographically and taxonomically divergent colubrid fossils from the Late Eocene.
Viperids versus colubrids + elapids	3-*Set B*	47 (40–95)	2.00 (1.20)	40	
Viperids versus colubrids + elapids	3-*Set C*	40 (37–60)	1.10 (1.25)	37	
Natricines versus colubrines (Stem)	4	36 (35–45)	0.50 (1.10)	35	Fossils assigned to *Coluber cadurci* and *Natrix mlynarskii*, extinct species that belong the extant subfamilies Colubrinae and Natricinae respectively, have been described from the early Oligocene (30–34 Ma) in Europe (Rage 1988). A third colubrid, *Texasophis galbreathi*, has been described from the early Orellan to Whitneyan ages of the Oligocene (30–31 Ma) in North American (Holman 1984, p. 225). Based on these fossils, the crown natricine–colubrine divergence has been constrained at 35–45 myr (Guicking et al. 2006; Alfaro et al. 2008). However, fossils with colubrine and natricine morphology appear almost immediately after the first appearance of intedeterminate colubrids, suggesting that these primitive fossils may be more appropriate for dating the stem natricine–colubrine clade (in our case, the divergence between the xenodontines, natricines, colubrines). We tested the effect of constraining the stem (node 4) and crown (node 5) colubrine–natricine clades at 37 (35–45) myr.
Natricines versus colubrines (Crown)	5				

Continued

TABLE 1. (*Continued*)

Fossil calibrations	Node	Calibration priors			Reference
		Mean (95% HPD)	Ln mean (SD)	Zero offset	
Elapines versus hydrophines Crown hydrophiines	6 7	23 (21–30)	1.00 (0.80)	20	A fossil vertebra from the late Oligocene/early Miocene (20–23 Ma) has been assigned to *Laticauda* and, based on its similarity to *L. colubrina* but differences from *L. laticaudata* and other elapids, Scanlon et al. (2003, p. 579) suggested that this fossil is nested within (not basal to) the genus *Laticauda*. Based on this taxonomic assignment, Wuster et al. (2007) used a minimum age of 24 myr to calibrate the divergence between Laticauda and all other hydrophiines (crown hydrophiines). However, the taxonomic affinity and/or stratigraphic age of this fossil have recently been questioned (Sanders and Lee 2008, p. 1186). This vertebra is one of the oldest elapid fossil known and might, therefore, be basal to (rather than nested within) the extant elapid group. Apart from this fossil, the earliest appearances of modern elapids in the first fossil record are proteroglyphous fangs from Germany dated at 20–23 Ma (Kuch et al. 2006). We tested the effects of constraining the crown hydrophiines and crown elapids with dates of 23 (21–30) myr.
African versus Asian *Naja* (Stem) African versus Asian *Naja* (Crown)	8 9	19 (17–30)	1.00 (1.00)	16	Fossils of 3 extinct European *Naja* species with apomorphies that distinguish Asian and African *Naja* occur 16 Ma (Szyndlar and Rage 1990). These fossils have been used to date the divergence between the crown African and Asian *Naja* (Wuster et al. 2007, 2008; Kelly et al. 2009). However, these extinct fossil species display primitive conditions that are very rare among living cobras (Szyndlar and Rage 1990) suggesting that they should be used to calibrate the stem rather than the crown *Naja* clade. We assigned the divergence between *Naja* and the closely related *Bungarus* as the stem clade and explored the effects of constraining the crown and stem *Naja* with dates of 19 (17–30) myr.

Notes: Constraints are given as absolute values (millions of years before present) and the corresponding lognormal mean, standard deviation, and zero offset of the calibration prior used in BEAST analyses. Phylogenetic placement of nodes is shown on Figure 1.

the oocyte maturation factor gene (c-*mos*—864 bp) and the recombination activating gene 1 (RAG-1—2400 bp). Novel cytochrome *b*, 16S rRNA, and ND4 fragments were amplified and sequenced using the primers published in Lukoschek and Keogh (2006), Palumbi (1996), and Forstner et al. (1995), respectively, and the protocols of Lukoschek and Keogh (2006) and Lukoschek et al. (2007). Amplifications of RAG-1 and c-*mos* used the primers and protocols of Groth and Barrowclough (1999) and Saint et al. (1998). Newly generated sequences were submitted to GenBank (Appendix Table A1). For some taxa, mitochondrial fragments and/or nuclear genes were concatenated from two individuals or two congeneric species to minimize the amount of missing sequence data, in which case, the highest common taxon name was assigned (Appendix Table A1). Sequences were edited in SeqMan (Lasergene v.6; DNASTAR Inc.), aligned with Clustal W2 (default parameters) (Labarga et al. 2007), and visually refined. Following alignment, coding region sequences were translated into amino acid sequences in MacClade v.4.06 (Sinauer Inc.) using the vertebrate mitochondrial and nuclear genetic codes as appropriate. No premature stop codons were observed, so we are confident that the mitochondrial sequences obtained were mitochondrial in origin, and that the nuclear genes were not nonfunctional nuclear copies (pseudogenes). Saturation plots comparing uncorrected "p" genetic distances with General Time Reversible plus invariant plus gamma (GTRig) distances were constructed for the nuclear and mitochondrial data sets. In order to evaluate saturation in each of the mitochondrial codon positions, we also constructed saturation plots for the first, second, and third codon positions of the ND4 and cytochrome *b* genes.

The best-fit models of molecular evolution for the nuclear and mitochondrial data sets were selected based on Akaike Information Criteria (AIC) implemented in ModelTest 3.06 (Posada and Crandall 1998) using model scores (−lnL) obtained from PAUP* (Swofford 2000). We evaluated alternative partitioning strategies using a modified version of the Akaike information criterion for small sample sizes (AIC_c) and Bayesian information criterion (BIC) (McGuire et al. 2007). AIC_c and BIC values incorporate a penalty for increasing the number of parameters in the model, thus potentially avoiding problems with model overparameterization. Three partitioning strategies were evaluated for the mitochondrial (mtCode, mtRNA; mtCode1+2, mtCode3, mtRNA; mtCode1, mtCode2, mtCode3, mtRNA) and nuclear data (nDNA; nDNA1+2, nDNA3; nDNA1, nDNA2, nDNA3). Bayesian analyses (four incrementally heated chains run for 2,000,000 generations sampled every 100th generation with all substitution parameters and rates allowed to vary across partitions) were conducted in MrBayes (Ronquist and Huelsenbeck 2003) and used to evaluate combinations of character partition and evolutionary model. AIC_c and BIC values were calculated using the equations of (McGuire et al. 2007, p. 841). AIC_c and BIC criteria selected the same optimal partitions as follows: mitochondrial—mtCode1-GTRig, mtCode2-GTRig, mtCode3-GTRig, mtRNA-GTRig; mtDNA excluding third codon positions (mtDNA3rdExcl)—mtCode1-GTRig, mtCode2-GTRig, mtRNA-GTRig; and nuclear—nDNA1-GTRig, nDNA2-GTRig, nDNA3-GTRig with model parameters allowed to vary independently across partitions. However, MrBayes returned unrealistic estimates of alpha for the nDNA1 gamma distribution of rate heterogeneity (66.74 ± 4006.05), so we used the next best nDNA model (nDNA1+2-GTRig, nDNA3-GTRig) and the best mtDNA model for all Bayesian analyses (BEAST and MrBayes). We also conducted extensive preliminary analyses of all three methods using a combined nDNA + mtDNA data set, but the results were virtually identical to those obtained for the mtDNA data alone, so we do not present the results of the combined data set.

Bayesian relaxed molecular clocks, which assume rates of molecular evolution are uncorrelated but lognormally distributed among lineages (Drummond et al. 2006), as implemented in BEAST v1.4.8 (Drummond and Rambaut 2007) were used for all dating analyses. Yule and birth–death models performed similarly in all preliminary analyses, so the birth–death model (Gernhard 2008) with a uniform prior was used to model cladogenesis for all final analyses. We summarized the outputs of all MrBayes and BEAST Markov chain Monte Carlo (MCMC) analyses using TRACER (version 1.4) in order to obtain parameter estimates, as well as evaluate effective sample sizes (ESSs) and convergence. ESS values greater than 100 are generally regarded as being sufficient to obtain a reliable posterior distribution (Drummond et al. 2007), and we adjusted the numbers of MCMC runs to ensure that ESSs were greater than 100 for all relevant parameters in each set of analyses conducted (numbers of MCMC runs for different analyses are specified in relevant sections). ESS values typically were much larger than 100 for most parameters in each analysis. Graphical exploration of trace files for tree likelihoods and other tree-specific parameters using TRACER (version 1.4) indicated that convergence had been reached in all cases.

Single-Fossil Cross-Validations

The agreement or consistency between single-fossil calibration dates and other available fossil calibrations for 10 calibrated nodes (Fig. 1—Tree Root and nodes 1–9) was evaluated using a modified version of the single-fossil cross-validations developed by Near et al. (2005). There were two main differences in our approach. First, rather than using fixed points for each calibration, we used lognormal distributions that placed a hard minimum bound and soft maximum bound on each calibration (Table 1), thereby allowing for uncertainty in the fossil dates (Yang and Rannala 2006; Ho and Phillips 2009). For each single-fossil calibration (i), we calculated the metrics $\overline{D}_x$, SS_x, and s (Near et al. 2005) for the other nine fossil-calibrated nodes on the tree using age estimates obtained from BEAST. We conducted the

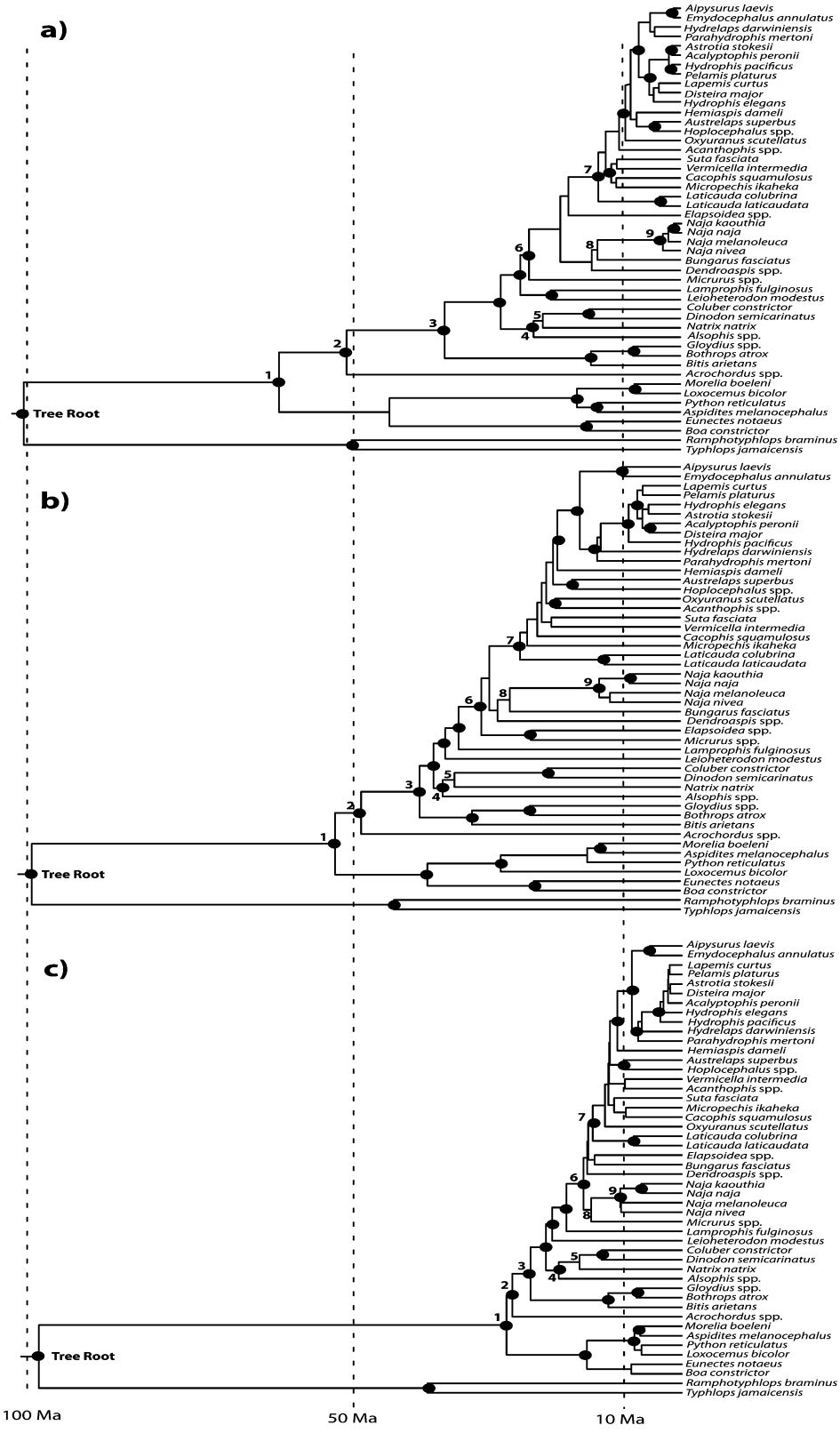

FIGURE 1. Bayesian chronograms from BEAST analyses with tree roots constrained to 97 (92–120) myr indicating the position of 10 candidate fossil-calibrated nodes (root and nodes 1–9) evaluated in this study. Solid black dots indicate nodes with $\geqslant$98% PPs. a) Chronogram from nuclear DNA. b) Chronogram from entire mitochondrial data set and c) Chronogram from mitochondrial data with third codon positions removed.

cross-validations using both the mean and median age estimates in order to evaluate whether the posterior age distributions (rather than point age estimates) influenced which fossil calibrations were identified as incongruent. The difference between the molecular (MA) and fossil age (FA) at each node was calculated as $D_i = (MA_i - FA_i)$, where FA_i is the fossil age and MA_i is the mean or median molecular age estimate for node i using the candidate fossil calibration at node x. The average difference $\overline{D}_x$ between the MA and FA across the nine other fossil-calibrated nodes for the fossil calibration at node x was then calculated as

$$\overline{D}_x = \frac{\sum_{i \neq x} D_i}{n-1}.$$

The FA for each candidate fossil-calibrated node (x) was used as a single calibration prior in the BEAST analysis, and its standard errors were calculated from the remaining nine candidate fossil-dated nodes. *SS* values were then calculated as the sum of the squared differences between the MA and FA age estimates at all other fossil-dated nodes using the formula

$$SS_x = \sum_{i \neq x} D_i^2.$$

Finally, the average squared deviations, s, were calculated using the formula $s = \frac{\sum_{x=1}^{n} \sum_{i \neq x} D_i^2}{n(n-1)}$, where n is equal to the total number of observations of D_i (i.e., the number of fossil calibrations remaining). For more details about the single-fossil cross-validation analyses, see Near et al. (2005).

The second difference in our approach was that, rather than using the cross-validations to exclude specific fossils, we used them in a more exploratory fashion to evaluate the alternative placements of three fossils as calibrations for their respective stem (nodes 4, 6, and 8) and crown (nodes 5, 7, and 9) clades (Table 1). We also evaluated three different pairs (referred to as calibration sets) of fossil dates for two nodes, the most recent common ancestor (MRCA) of Caenophidia (Fig. 1—node 2), and the MRCA of Colubroidea (Fig. 1—node3), based on their previous use in other studies (Table 1). Each alternative set of fossil dates for nodes 2 and 3 (Table 1: *Sets A, B, C*) was evaluated by conducting a separate iteration of the cross-validation exercise (i.e., three separate iterations). In each case, the calibration set and the corresponding molecular dates from the single-fossil dating analyses were used to calculate $\overline{D}_x$, SS_x, and s. The molecular and fossil dates for the other eight single–fossil-calibrated nodes were the same for the three calibration sets.

Preliminary analyses revealed that the shallower calibrations (Fig. 1, nodes 4–9) artificially inflated age estimates at deeper nodes to unrealistically high values. In order to stabilize estimated ages at deeper nodes, we constrained the root using a normal prior (mean = 110 MA, 95% confidence interval = 85–135 MA) spanning a wide range of plausible dates for this node (Table 1) in all single-fossil calibration analyses. BEAST runs for single-fossil cross-validations were conducted as follows: nDNA—4,000,000 generations sampled every 100 generations, mtDNA—5,000,000 generations sampled every 100 generations, and mtDNA3rdExcl—10,000,000 generations sampled every 100 generations.

Evaluating Fossil Coverage and Identifying Outliers

The approach of Marshall (2008) involves generating an ultrametric tree that is uncalibrated with respect to the fossil record and then mapping all candidate fossil calibrations onto the tree to determine which of the calibrated lineages has the best temporal fossil coverage. Specifically, the method aims to identify the lineage, for which the oldest fossil (for that lineage) sits proportionally closest to the node of its MRCA (true time of origin) and therefore has the best temporal coverage. Marshall (2008) emphasizes two assumptions of the method: 1) the proportional branch lengths of the ultrametric tree are accurate and 2) fossilization is random: However, the method also assumes that fossils are accurately dated and assigned correctly to their respective lineages (see below for further discussion).

The first and arguably most important step in the approach of Marshall (2008) is to generate a reliable ultrametric phylogeny that is uncalibrated with respect to the fossil record using an appropriate relaxed clock algorithm. Given that obtaining accurate proportional branch lengths of the ultrametric tree is critical to the success of this method, we generated a number of ultrametric trees using different approaches and compared the results. Specifically, we generated ultrametric trees for the mtDNA and nDNA data sets in BEAST by constraining the tree root with a fixed value (arbitrarily set to 100). However, MCMC runs of 20,000,000 generations were needed to obtain ESSs >100 for the calibrated nodes using nDNA and convergence could not be achieved for mtDNA. As such, we followed the approach of Marshall (2008) and obtained ultrametric trees using r8s (Sanderson 2003). r8s requires user-specified input trees, so we used MrBayes (MCMC chains of 2,000,000 generations sampling every 100 generations and all default settings) to obtain optimal Bayesian phylogenies for the nDNA and mtDNA data sets using the same partitioning strategies and models of evolution used for the BEAST analyses. As there is evidence that branch lengths are more accurately estimated by maximum-likelihood (ML) than Bayesian criteria (Schwartz and Mueller 2010), we also generated ML trees for the nDNA, mtDNA, and mtDNA3rdExcl data sets in PAUP (Swofford 2000) under optimal models of sequence evolution obtained from AIC in ModelTest (Posada and Crandall 1998). We generated rooted input trees (required by r8s) by adding sequences obtained from GenBank (Appendix Table A1) for two outgroup taxa (the lizard genera *Varanus* and *Calotes*)

to the data sets. The lizard taxa were pruned from the optimal ML and Bayesian trees and the resulting rooted trees used to obtain ultrametric trees in r8s, again fixing the root age to an arbitrary value of 100. We used semiparametric penalized likelihood (PL) (Sanderson 2002) and optimal smoothing parameters identified from the cross-validation procedure in r8s as follows: MrBayes tree—smoothing parameter of 3200 with log penalty function and ML tree—smoothing parameter of 3200 with additive penalty function. Given that Smith et al. (2006) demonstrated that the log penalty function better estimated branch lengths than the additive penalty function for calibrated ultrametric trees, we also generated an ML ultrametric tree using the log penalty function and optimal smoothing parameter of 320 (note, however, that the sum of squares obtained from the cross-validations for the log penalty function were much higher than the additive penalty function, suggesting that the additive penalty was more appropriate).

We used the resultant ultrametric trees to calculate the empirical scaling factor (ESF) for each candidate fossil calibration (including the three alternative fossil dates for nodes 2 and 3 and the alternative placements of three fossils, Table 1) using the equation $\mathrm{ESF}_i = \frac{\mathrm{FA}_i}{\mathrm{NTL}}$, where FA_i is the age of the oldest fossil of the lineage and NTL_i is the relative node to tip length of the branch of that lineage on the ultrametric phylogeny (Marshall 2008). The fossil with the largest ESF_i is regarded as having the best temporal coverage; however, fossils that have been incorrectly assigned and/or incorrectly dated may also have the highest ESF values and these outliers need to be identified. We tested for possible fossil outliers by comparing the distribution of ESF_i values to a uniform distribution using the Kolmorgorov–Smirnov test, on the assumption that ESF_ivalues for fossil outliers lie outside a uniform distribution (Marshall 2008). One limitation of this approach is that it is most effective if there is just one outlier (Marshall 2008, p. 732). We were testing the alternative stem and crown placements of three fossils. As such, the ESF_i values for the crown placements (that inevitably will be larger than the ESF_i values for their stem placements) might potentially cluster together, thereby making it impossible to identify them as outliers. In order to address this issue, we modified the approach of Marshall (2008) to test the alternative placements of these fossils (see Results section for details).

Bayesian Analyses to Evaluating Multicalibration Sets

We used the method of Sanders and Lee (2007) to evaluate three alternative dates for two nodes with controversial fossil calibrations in a Bayesian multicalibration framework. This method compares the prior and posterior distributions of the 95% highest posterior densities (HPD) intervals for each candidate calibration, particularly focusing on potentially controversial calibrations of interest. In our case, the single-fossil cross-validations identified plausible congruent calibration sets comprising six fossil-calibrated nodes that included nodes 2 and 3 but could not distinguish between the different possible ages assigned to these two nodes (Table 1—*Sets A, B*, and *C*). In addition, the ESF_i values for the same six fossil-calibrated nodes indicated that none were outliers. However, ESF_i values cannot be used to evaluate alternative dates for the same node because the oldest date will inevitably have the highest empirical coverage, even if that date is not correct. Moreover, ESF_i values from different ultrametric trees identified different fossils as having the highest empirical coverage (see below for details). We evaluated the alternative ages for nodes 2 and 3 using three sets of BEAST multicalibration analyses that incorporated the four congruent calibrations and the *Set A, B,* and *C* node 2 and 3 calibration ages in turn. For each analysis, we compared the prior and posterior distributions of all six fossil-calibrated nodes, with the expectation that the node 2 and 3 calibration set most consistent with the other four fossil-dated nodes would return posterior distributions for all six calibrated nodes that were similar to their prior constraints (Sanders and Lee 2007). We also conducted a fourth set of analyses using the four congruent fossils with no constraints on nodes 2 and 3 (*Set D*) and compared the unconstrained and constrained node 2 and 3 age estimates. These four sets of BEAST analyses were conducted for nDNA, mtDNA, and mtDNA3rdExcl data sets, using the same lognormal priors, relaxed molecular clocks, and partitioned evolutionary models as the single-fossil dating analyses. MCMC runs comprised 4,000,000 generations for the nuclear data and 10,000,000 generations for both mitochondrial data sets. In each case, MCMC runs were sampled every 100 generations.

Given that certain combinations of priors can interact to generate unexpected effective joint priors, we also performed an analysis for each calibration set without data (empty alignments) to ensure that the effective priors were similar to the original priors. We assessed how informative the data were by comparing the effective priors with posteriors obtained using data (Drummond et al. 2006). These analyses indicated that the effective priors were similar to the original priors and the posteriors obtained from the data departed from the priors (indicating informative data).

RESULTS

The final nDNA alignment had 3264 characters of which 870 were variable and 421 were parsimony informative, whereas the mtDNA alignment had 2221 characters of which 1368 were variable and 1193 were parsimony informative, and the mtDNA3rdExcl had 1632 characters of which 884 were variable and 578 were parsimony informative. All tree topologies from PAUP* ML analyses and Bayesian MCMC searches (MrBayes and BEAST) of the nuclear and mitochondrial data sets converged on a topology (Fig. 1) highly congruent with published molecular phylogenies for the the elapid taxa (Slowinski et al. 1997; Keogh 1998; Keogh et al. 1998;

Slowinski and Keogh 2000; Lukoschek and Keogh 2006; Wuster et al. 2007; Sanders and Lee 2008; Sanders et al. 2008; Kelly et al. 2009; Pyron et al. 2011). Data matrices and relevant trees have been submitted to TreeBASE (#11272). Eight of the 10 candidate calibration nodes had extremely high support with ≥99% PPs for all analyses conducted (Fig. 1). The two nodes with poor support were node 5 (typically with ∼80% PPs for mtDNA and <50% PPs for nDNA) and node 8 (typically with ∼55% PPs for mtDNA and <50% PPs for mtDNA). Other nodes with PPs >98% are also shown on the trees (Fig. 1). Saturation plots revealed an abundance of hidden substitutions in all three codon positions of the mitochondrial data set (Fig. 2a–d) but particularly in the third codon position (Fig. 2d).

Single-Fossil Cross-Validations

In all cases, the results of single-fossil cross-validations using mean and median age estimates from BEAST were highly consistent, so we present only the results from the mean age estimates. Nuclear DNA cross-validations produced similar results for each calibration set, with $\overline{D}_x$ values indicating that four fossils consistently produced older molecular divergence estimates for other candidate fossil-calibrated nodes, whereas the other six fossils produced younger divergence estimates; however, the relative magnitude of these tendencies differed between calibration sets (Fig. 3a). Specifically, the youngest fossil dates for nodes 2 and 3 (*set A*) resulted in larger molecular overestimates and smaller underestimates of fossil dates than *sets B* and *C*, which returned similar mean differences ($\overline{D}_x$) between the fossil and molecular dates (Fig. 3a). *SS* values ranked the four node calibrations that consistently produced older molecular divergence estimates for other FAs as the most incongruent fossils (Fig. 4a). *Set A* calibrations produced consistently larger *SS* values for all fossil calibrated nodes than *sets B* and C (Fig. 4a), reflecting the larger differences ($\overline{D}_x$) between the molecular and fossil dates using the younger *set A* calibrations (Fig. 3a). By contrast, *SS* values for *sets B*

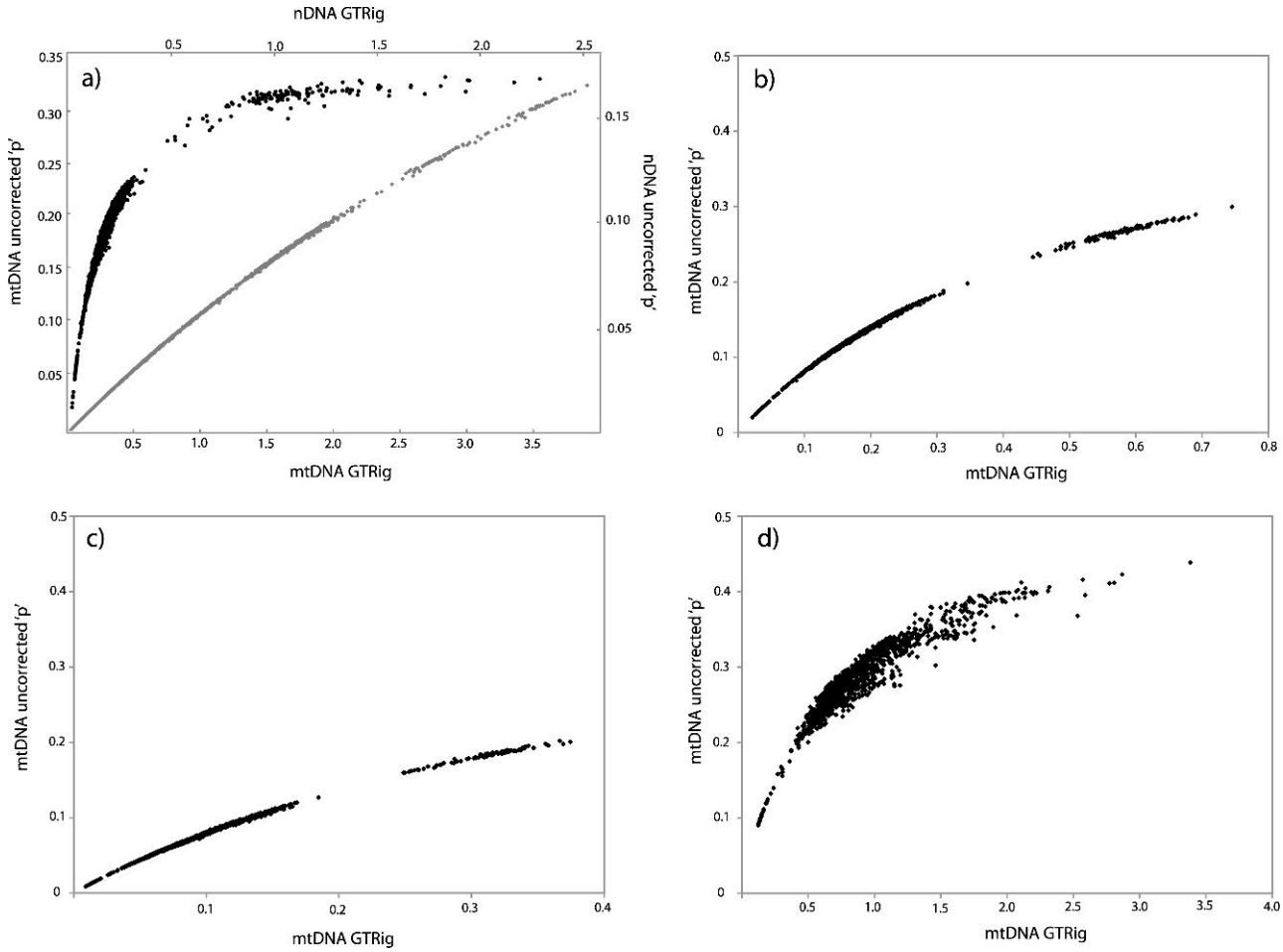

FIGURE 2. Saturation plots of genetic distances corrected for multiple substitutions versus uncorrected "p" distances. Corrected genetic distances were calculated using the estimated best-fit models of sequence evolution obtained from AIC criterion in ModelTest. a) Saturation plots of the entire mitochondrial DNA data set (black circles) versus nuclear DNA (gray diamonds). Note the different axis scales for the nuclear and mitochondrial data sets. Saturation plots are also shown for b) mtDNA first codon position, c) mtDNA second codon position, and d) mtDNA third codon position for the combined ND4 and cytochrome *b* genes. Note the different *x*-axis scales for b, c, and d.

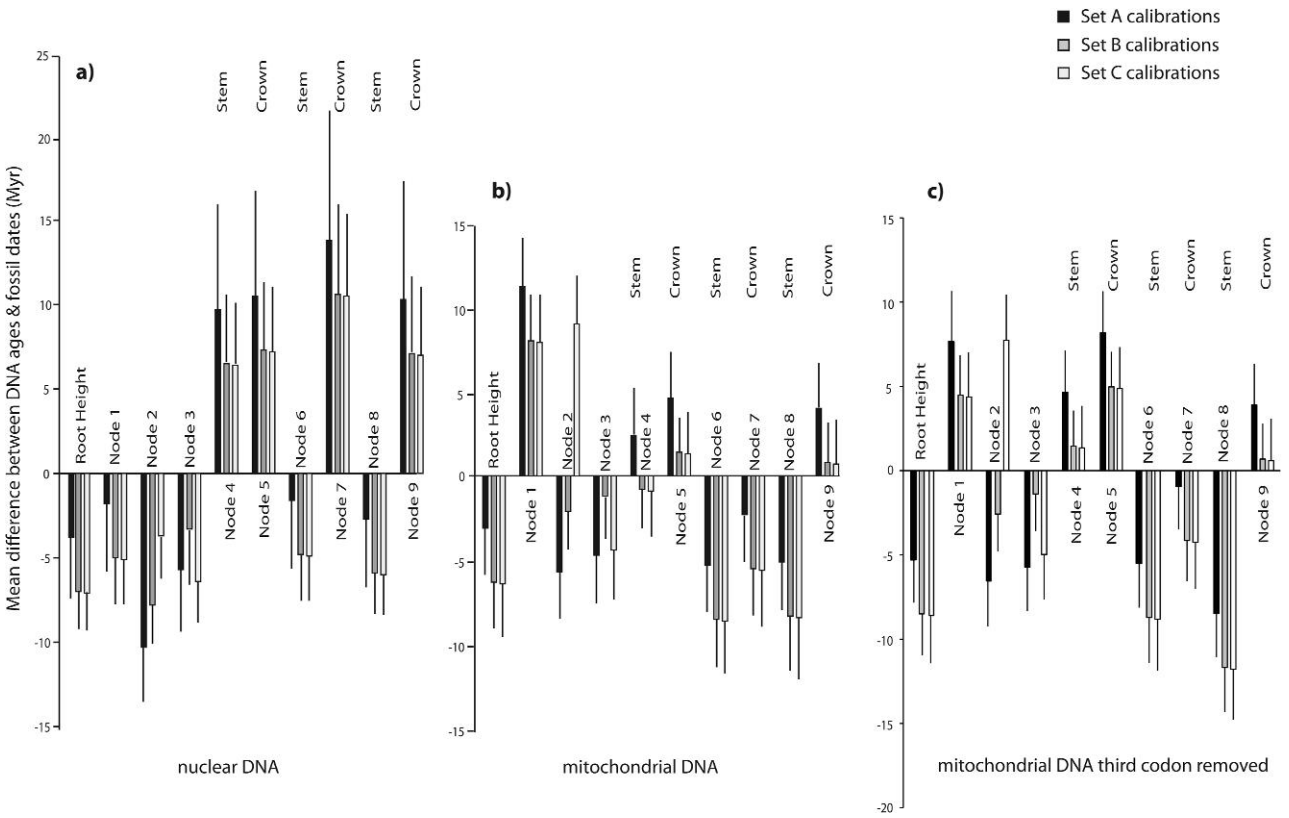

FIGURE 3. Histogram of the mean differences and standard errors between fossil and estimated MAs for each of three sets of 10 single–fossil-calibrated nodes from a) nuclear DNA; b) mitochondrial DNA; and c) mitochondrial DNA with third codon position removed. FAs for 8 of the 10 candidate nodes were identical for each set, differing only for nodes 2 and 3 (see Fig. 1). FAs used as constraints are given in Table 1. For a single node (x), the FA at node x was used as a single calibration prior. MA estimates were obtained for the nine other candidate nodes, for which FAs were available.

and *C* were very similar (Fig. 4a). Sequential removal of fossil calibrations from most to least divergent, as ranked by *SS* values (Fig. 4a), resulted in steep incremental declines in *s* values for the subsequent removal of nodes 7, 9, 5, and 4 for all calibration sets (Fig. 5a). At this point, *s* values for se*ts B* and *C* were small and subsequent removal of fossils did not markedly decrease *s* values (Fig. 5a). Starting *s* values for *set A* were much larger than for *sets B* and *C* and did not drop to low values until the fifth fossil calibration (node 2) was removed and then remained low (Fig. 5a).

Mitochondrial DNA produced a markedly different pattern of mean differences ($\overline{D}_x$) between the molecular and fossil dates than nuclear DNA (Fig. 3). Most notably, the four fossil calibrations (nodes 4, 5, 7, and 9) that returned much older nuclear DNA values for FAs at other candidate calibration nodes either produced younger or only slightly older estimates of FAs for mtDNA (Fig. 3b) and this remained the case even when the third codon positions were removed (Fig. 3c). In addition, the tendency for nodes 6 and 8 to produce younger MAs for fossil dates at other nodes was more extreme for the mitochondrial than nuclear data, and this was true for both mitochondrial data sets (Fig. 3b,c). By contrast, node 1 produced older ages at other nodes for both mtDNA data sets, whereas this node produced younger dates for nuclear DNA. Given these differences, it is not surprising that mitochondrial *SS* values ranked fossils differently than nuclear *SS* values (Fig. 4b,c). In addition, $\overline{D}_x$ values for the younger *set A* calibrations (at nodes 2 and 3) did not follow the same pattern as for *sets B* and *C* (Fig. 3b,c) and the mitochondrial rank order of candidate calibrations was different for *set A* calibrations than for *sets B* and *C*, which were similar (Fig. 4b,c). *Sets B* and*C* had highest *SS* values at nodes 6 and 8; however, removing these nodes only slightly decreased *s* values, which did not decline sharply until subsequent removals of the third and fourth ranked fossils and then remained low (Fig. 5b,c). Interestingly, node 1 was the most incongruent fossil for the younger *set A* calibrations for the entire mtDNA data set and *s* values dropped sharply when it was removed. Subsequent removal of the three next most incongruent fossils did not produce further decreases in *s*, but *s* decreased with the removal of the fifth and subsequent fossils (Fig. 5b). By contrast, node 8 was the most incongruent fossil for all three calibration sets for the mtDNA data set with third codon position excluded, and *s* values did not drop sharply until the first two most incongruent nodes were excluded in each case (Fig. 5c).

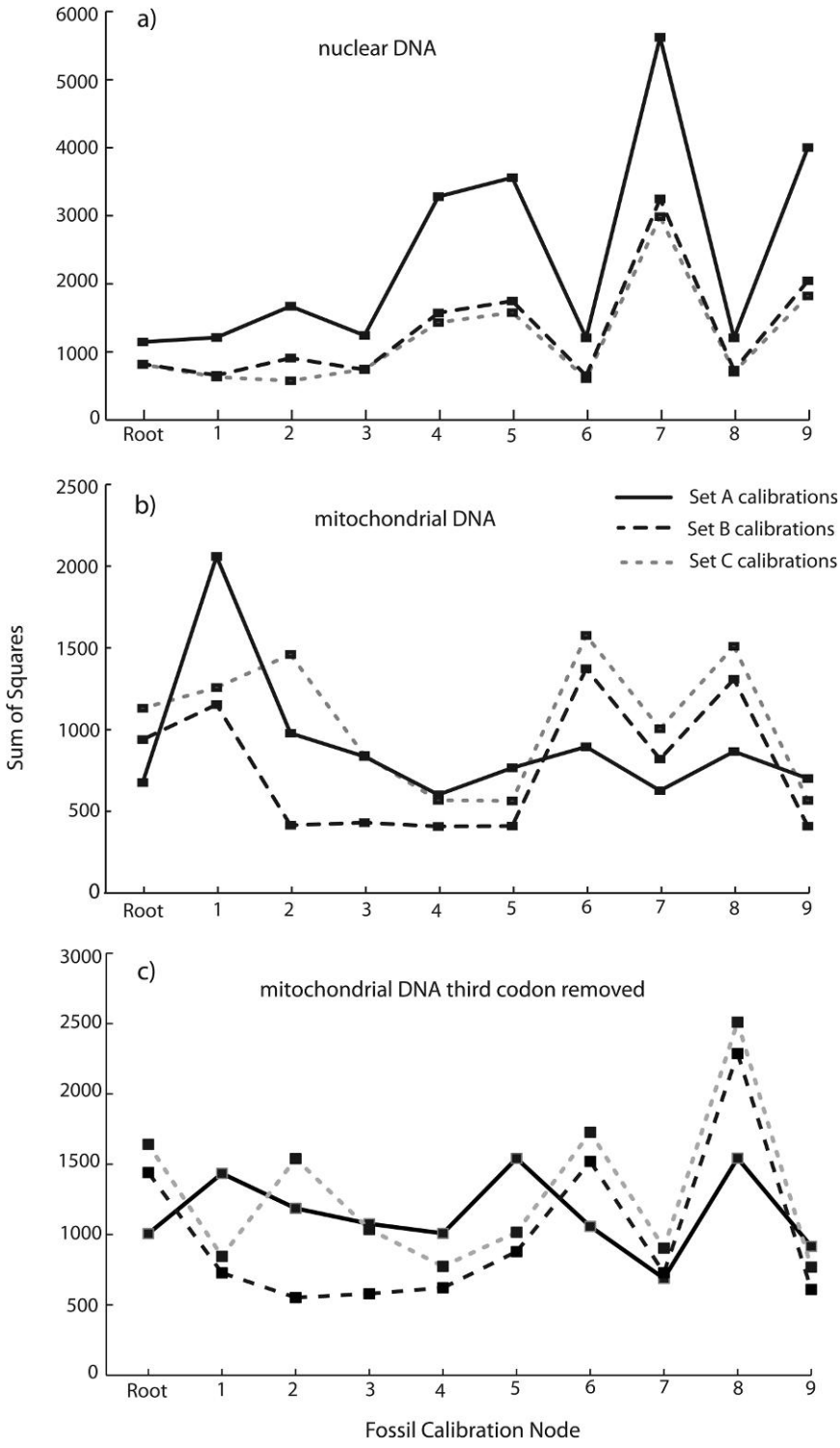

FIGURE 4. *SS* values for each candidate fossil calibration node when used as the single calibration prior in each of the three calibration sets for a) nuclear DNA; b) mitochondrial DNA; and c) mitochondrial DNA with third codon position removed.

Fossil Coverage and Fossil Outliers

The four ultrametric trees obtained from the nDNA data set differed in their proportional branch lengths, resulting in differing ESF_i values for the candidate fossil calibrations (Table 2). Nonetheless, the four highest ESF_i values (in decreasing order) for the ML and MrBayes ultrametric trees were for nodes 9, 7, 5, and 4 (Table 2), the same nodes identified as least congruent by the cross-validation analyses. These four nodes also had the highest ESF_i values for the BEAST ultrametric tree

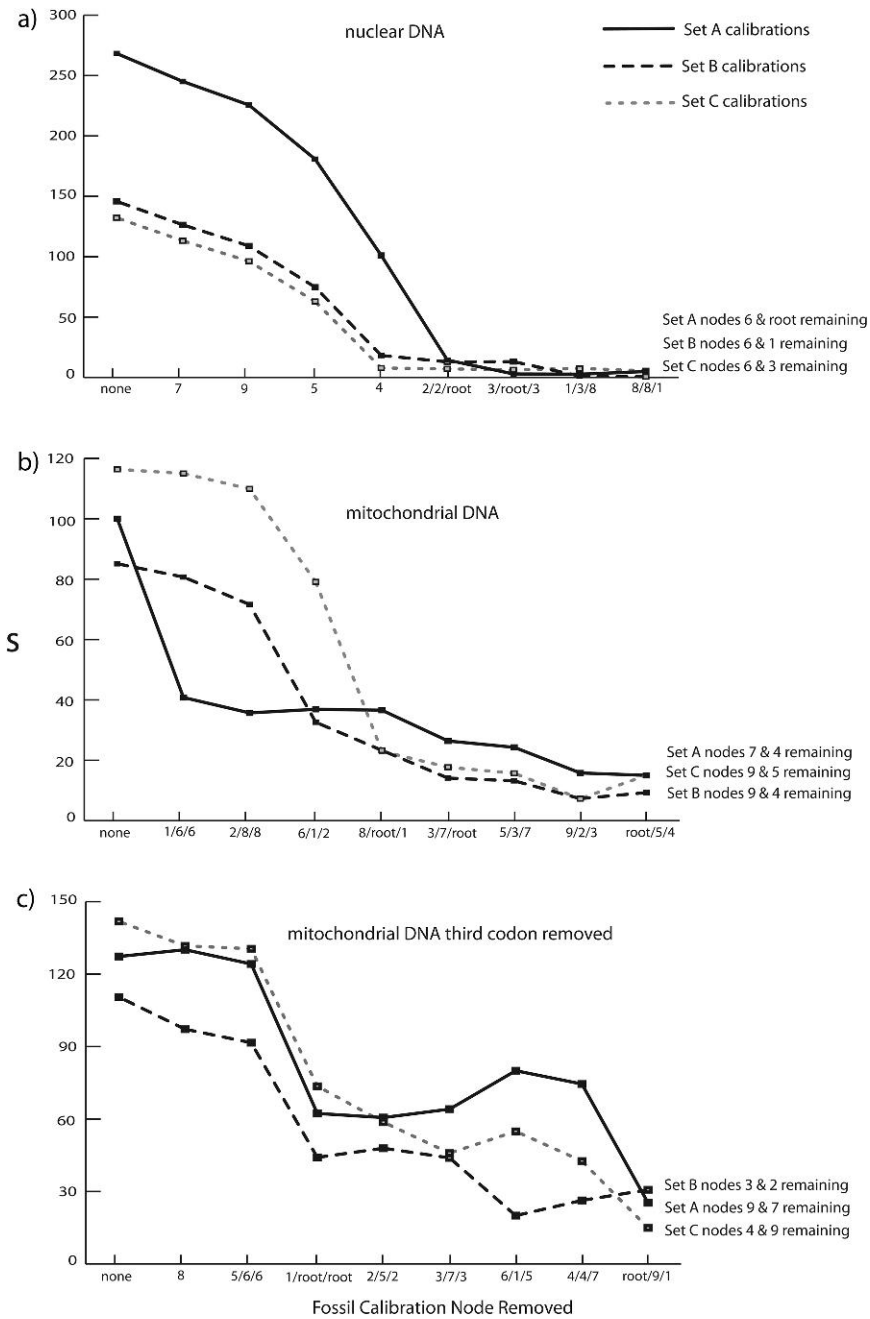

FIGURE 5. Effect of sequentially removing candidate fossil calibrated nodes on *s*, the average squared deviation of D_i values for the remaining fossil calibrations in each set. a) Nuclear DNA *s* values for three calibration sets. Fossils were removed based on highest to lowest *SS* values calculated from all 10 fossil-calibrated nodes. Removal order (shown on the *x*-axis) of the first four most incongruent fossils was identical for each calibration set but then differed between sets. b) mtDNAs values for three calibration sets when fossils were removed based on highest to lowest *SS* values calculated for all 10 fossil-calibrated nodes.

but in different decreasing order (Table 2). Lack of resolution in the ML and Bayesian nDNA trees resulted in nodes 4 and 5 forming a polytomy: As such, it was not possible to evaluate the alternative placements of this fossil calibration (as the ESF_i values for the stem and crown placement were identical). Moreover, issues regarding the taxonomic affinities of these fossils (Table 1; and Supplementary material A available from http://www.sysbio.oxfordjournals.org/) suggest that it is not possible to accurately place them on the phylogeny (despite their use to date caenophidian divergences in previous studies: Guicking et al. 2006; Alfaro et al. 2008). As such, we excluded them from the outlier analysis.

Nodes 7 and 9 were the shallower crown placements of the two candidate fossil calibrations, for which the

TABLE 2. ESF_i for candidate fossil calibrations for nuclear and mitochondrial data sets calculated using proportional branch lengths obtained from uncalibrated ultrametric trees produced using different methods

	Nuclear DNA							Mitochondrial DNA				mtDNA noThird	
	MrBayes - log 3200		ML - add 3200		ML - log 320		BEAST		MrBayes - log 1		ML - log 10		ML - log 10	
Node no	ESF	Node no	ESF	Node no	ESF	Node no	ESF	Node no	ESF	Node no	ESF	Node no	ESF
2-*Set A*	76	2-*Set A*	79	2-*Set A*	86	2-*Set A*	57	Root	97	Root	97	Root	97
Root	97	Root	97	Root	97	3-*Set A*	72	2-*Set A*	148	2-*Set A*	165	2-*Set A*	128
3-*Set A*	113	3-*Set A*	115	2 - *Set B*	130	3-*Set C*	84	3-*Set A*	178	3-*Set A*	193	3-*Set A*	137
2-*Set B*	114	2-*Set B*	118	1	138	1	85	3-*Set C*	209	3-*Set C*	227	8	140
8	122	8	120	3-*Set A*	140	2-*Set B*	85	1	213	8	232	6	148
1	123	1	128	2-*Set C*	148	6	85	8	221	1	241	3-*Set C*	161
2-*Set C*	130	2-*Set C*	135	3-*Set C*	165	8	88	2-*Set B*	223	6	245	4	185
3-*Set C*	132	3-*Set C*	136	8	168	2-*Set C*	97	6	235	2-*Set B*	247	7	187
6	137	6	152	3-*Set B*	193	Root	97	3-*Set B*	246	4	265	3-*Set B*	190
3-*Set B*	**156**	**3-*Set B***	**159**	**6**	**224**	3-*Set B*	99	4	247	3-*Set B*	267	2-*Set B*	192
4	196	4	194	4	250	**7**	**130**	2-*Set C*	254	2-*Set C*	281	1	197
5	196	5	194	5	250	4	134	5	271	5	286	9	203
7	235	7	242	7	361	9	168	7	299	7	324	5	205
9	250	9	370	9	494	5	171	**9**	**373**	**9**	**404**	**2- *Set C***	**218**

Notes: The BEAST nDNA chronogram was obtained by fixing the root to an arbitrary value of 100. The remaining ultrametric trees were produced in r8s using the optimal ML and Bayesian (MrBayes) phylogenies. Uncalibrated ultrametric trees were obtained by fixing the root to an arbitrary value of 100 and using PL with the logarithmic (log) or additive (add) penalty function and the optimal smoothing parameter obtained from cross-validation (shown in the column heading). Nodes and corresponding ESF values highlighted in bold for each ultrametric tree indicate the fossil with the highest empirical coverage after removing fossils identified as outliers (i.e., not conforming to a uniform distribution). See text for more details.

alternative deeper stem placements also were evaluated. Obviously, the candidate fossils cannot correctly be assigned to both the stem and crown nodes so, prior to testing whether the distributions of ESF_i values conformed to uniform distributions, we removed the ESF_i values for the corresponding stem placements of each fossil (nodes 6 and 8). The resulting distributions of ESF_i values for the BEAST and ML ultrametric trees (under both the additive and log penalty functions) were strongly rejected as belonging to uniform distributions (BEAST $P < 0.05$; ML trees $P < 0.005$ in both cases); however, this was not the case for the MrBayes tree ($0.20 < P > 0.10$). These inconsistent results highlight the sensitivity of this approach to differences in proportional branch lengths obtained from ultrametric trees obtained using different methods (see below for further discussion). Given that the weight of evidence suggested that crown placement of the *Naja* fossil was an outlier, we removed the ESF_i values for node 9 and reinserted the ESF_i values for the corresponding stem placement of the fossil (node 8). The resulting distributions of ESF_i values for the MrBayes and ML ultrametric trees also were rejected as belonging to uniform distributions, suggesting that the crown placement of the putative *Laticauda* fossil at node 7 also is an outlier. However, this was not the case for the BEAST ultrametric tree (Table 2). We then removed the ESF_i values for node 7 (from the ML and MrBayes ESF_i distributions) and inserted the ESF_i values for the stem placement of the fossil at node 6. The resulting distributions of ESF_i values were not rejected as belonging to uniform distributions. In terms of the MrBayes tree, the inclusion of ESF_i values for both potential outliers (nodes 7 and 9) may have resulted in the artifact mentioned by Marshall (2008), whereby the larger ESF_i values of outliers group together making it impossible to distinguish the resultant distribution from a uniform distribution (thereby failing to identify node 9 as an outlier). In order to explore this possibility, we removed the ESF_i for node 7 and retained the ESF_i of the corresponding stem placement at node 6. The resulting distribution of ESF_i values did not conform to a uniform distribution, supporting node 9 as an outlier. Overestimation of shorter branches has recently been demonstrated for Bayesian approaches (Schwartz and Mueller 2010), and the smaller difference between ESF_i values for nodes 9 and 7 for the Bayesian than ML trees may reflect overestimation of short branches in the crown *Naja* clade by MrBayes.

The proportional branch lengths and corresponding ordering of ESF_i values for the ultrametric trees obtained from optimal mtDNA ML and MrBayes and the mtDNA3rdExcl ML trees were different from those obtained from nDNA (Table 2). For the both mtDNA trees, the crown nodes 5, 7, and 9 still had the highest ESF_i values, whereas for the mtDNA3rdExcl tree, the node 2 *Set C* had the highest ESF_i value (Table 2). However, the distributions of ESF_i values conformed to uniformity for all three mitochondrial ultrametric trees (ML and MrBayes), and this result was true for distributions

including just one potential crown node outlier (and the corresponding stem placement of the other fossil): Thus, no outliers were identified.

Evaluating Multicalibration Sets Using Bayesian Analyses

There were consistent differences in the plausible sets of congruent fossil calibrations identified from the cross-validations from nuclear and mitochondrial DNA, and the fossil outliers identified from nuclear but not mitochondrial data based on ESF_i values. These differences are almost certainly due to the effects of nucleotide saturation for mtDNA (see Discussion section). As such, we conducted the multicalibration analyses using the six fossil-calibrated nodes selected by the nuclear data.

Multicalibration analyses using nuclear DNA revealed similarities and differences between the estimated mean ages and 95% HPD intervals for the six calibrated nodes across calibration *sets A, B, C,* and *D*. The most striking similarities were for the four fossil calibrations common to each calibration set (tree root and nodes 1, 6, and 8), for which the means and minimum 95% HPD intervals were very similar to their respective calibration priors (<5% in all cases), whereas maximum 95% HPD intervals invariably were smaller than the calibrations (Fig. 6). By contrast, age estimates for nodes 2 and 3 differed considerably between calibration sets, in part reflecting the influence of their calibration priors but also reflecting inconsistencies between these priors and the other four fossil calibrations (Fig. 6). Moreover, age estimates for nodes 2 and 3 tended to converge on ages estimated by *set D* (Fig. 6), in which nodes 2 and 3 were not constrained. This tendency was most pronounced for node 2, for which the *set A* age estimate was far more similar to the *set D* estimate than to the *set A* calibration prior. Indeed, the *set A* prior and posterior distributions barely overlapped (Fig. 6). Similarly, the *set B* estimated age for node 2 also was closer to the *set D* estimate than to the *set B* calibration prior, with the *set B* maximum age estimate 70 million years younger than its calibration prior (Fig. 6). *Set C* returned a node 2 age estimate that was similar to both its calibration prior and the *set D* age estimate for this node, although its minimum 95% HPD interval was younger than the hard minimum bound of the prior. The node 3 age nDNA estimates were more similar to their respective calibration priors, but again, posterior distributions diverged from priors towards the unconstrained *set D* age estimate. The *set A* estimated mean age was slightly older than its calibration prior, but posterior and prior distributions were identical, whereas the *set C* age estimate also was identical to the mean and minimum bounds of the calibration prior (Fig. 6). The *set B* estimated mean age and minimum 95% HPD were younger than the calibration prior (Fig. 6).

Mitochondrial age estimates were invariably older for the shallower nodes 3, 6, and 8 than their respective calibration priors and, with one exception, also for the corresponding nDNA age estimates. By contrast, mitochondrial node 1 age estimates for all calibration sets were similar to the calibration prior and to nuclear DNA age estimates, and this was true for both mitochondrial data sets (Fig. 6). Nonetheless, the tendency for mtDNA to return older age estimates at shallow nodes and the tree root was much pronounced when the third codon positions were excluded, with mtDNA3rdExcl age estimates for nodes 6 and 8 age intermediate to the nDNA and mtDNA age estimates and tending to converge on mean nDNA age estimates for node 3 and the tree root (Fig. 6). Although mitochondrial age estimates for node 2 from the entire data set showed the same tendency as nuclear ages to converge on the unconstrained *set D* age estimates (irrespective of the calibration prior used), this was not the case for mtDNA with third codon positions excluded (Fig. 6). Indeed, with the exception of *set C,* the node 2 mtDNA3rdExcl age estimates tended to converge on the calibration prior resulting in age estimates that were younger than the corresponding nDNA estimates, and this was also true for the node 3 *set A* age estimate (Fig. 6).

Discussion

Increasing awareness of the importance of identifying reliable fossils to calibrate molecular clocks has resulted in the development of several methods for evaluating and employing fossil calibrations (reviewed by Ho and Phillips 2009). Each approach has advantages and limitations, as we demonstrate by comparing three different approaches with particular emphasis on the impact of nucleotide saturation on the fossils selected.

The cross-validation method (Near et al. 2005) discards calibrations until an internally consistent set is obtained, and in the process, may discard calibrations with the best temporal coverage because they are inconsistent with the remaining calibrations (see Marshall 2008 for indepth discussion of this issue). Nonetheless, the method has been used in several recent studies (Near and Sanderson 2004; Noonan and Chippindale 2006b; Rutschmann et al. 2007; Alfaro et al. 2008). By contrast, the use of ESF aims to identify one fossil with the best empirical coverage (Marshall 2008); however, accurate results are highly dependant on meeting the assumptions of the method (see below). Unlike the cross-validation approach, ESFs have only been used in one previous study (Davis et al. 2009). This study obtained an ultrametric tree in r8s using PL with log penalty function (following the advice of Marshall 2008), based on empirical evidence that PL using the log penalty function produces the most reliable ultrametric trees (Smith et al. 2006). However, Davis et al. (2009) comment that their resultant dates were much older than expected for several lineages. Our study demonstrated that ultrametric trees generated from ML and Bayesian nDNA phylogenies using the log penalty function were incongruent in terms of the magnitude and order of the ESFs (Table 2) and the fossil outliers identified. By contrast, results from the ML ultrametric tree using the

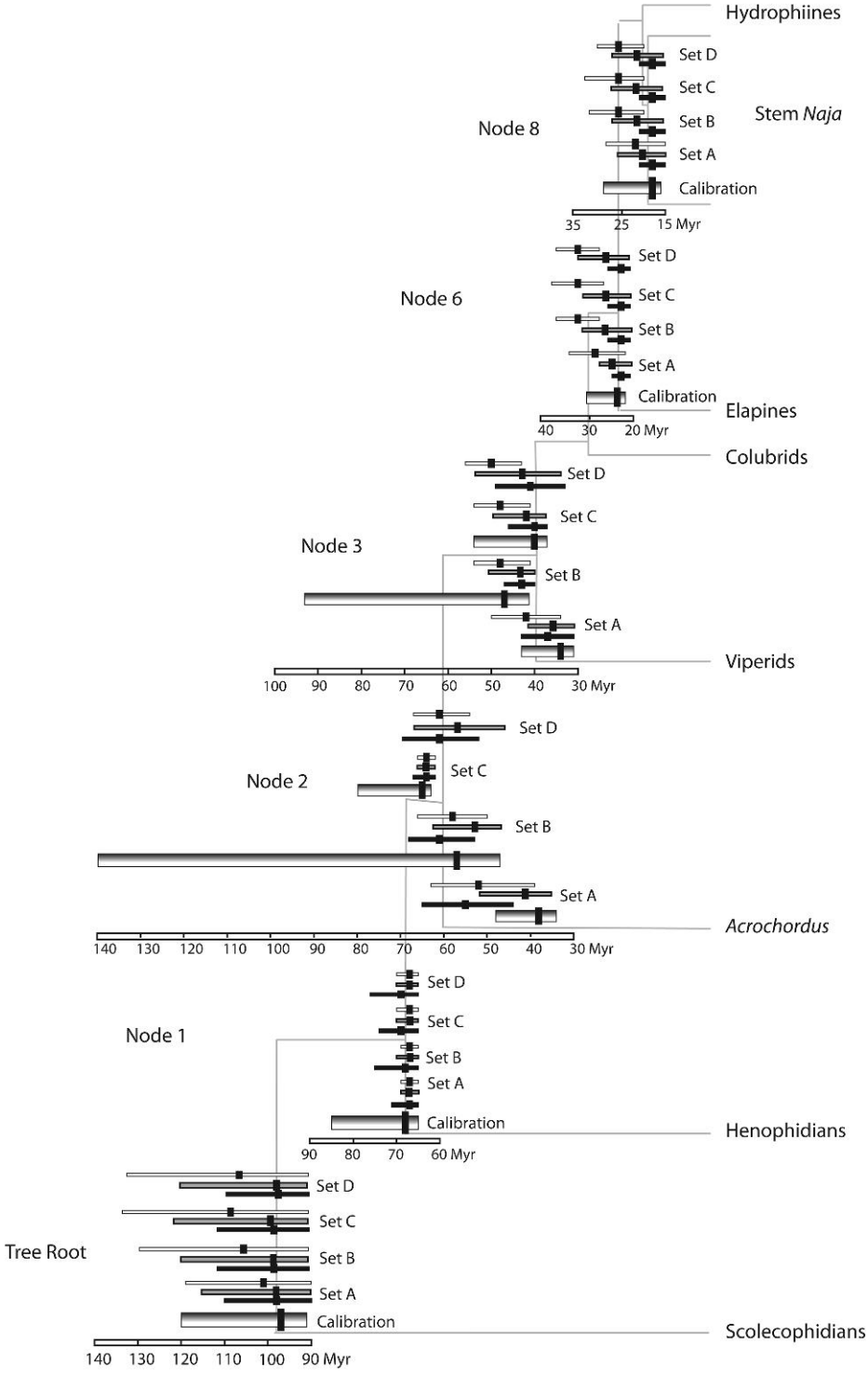

FIGURE 6. Bayesian multifossil calibration analyses showing fossil calibration priors and posterior distributions of MA estimates (mean and 95% HPD intervals) at six fossil-calibrated nodes using four calibration sets (*A*, *B*, *C*, *D*). Each calibration set comprised four calibration priors that were identical among sets (tree root, nodes 1, 6, and 8) and two priors that differed among sets (nodes 2 and 3). Lognormal calibration priors are shown as wider shaded bars with the lognormal mean shown as a black square on the bar. MA estimates for nuclear (black bars), mitochondrial (white bars), and mitochondrial DNA with third codon position removed (gray bars) are shown in pairs for each calibration set at each node. Bars indicate 95% HPDs with estimated mean ages indicated by black squares. At nodes 2 and 3, the respective calibration prior is shown immediately below the corresponding nuclear and mitochondrial age estimates. Calibration priors at the other four nodes are shown below all four sets of MA estimates. Prior and posterior distributions are shown on a diagrammatic chronogram depicting the backbone of the phylogeny; however, this chronogram does not represent the results of any specific analysis.

additive penalty function were more similar to those obtained for the MrBayes tree. At the very least, these results suggest that the findings of Smith et al. (2006) are not universal and various approaches for obtaining uncalibrated ultrametric trees need to be evaluated for reliability and consistency of results.

These conflicting results highlight a major limitation of ESFs, which is the reliance on accurate proportional branch lengths (which we do not know, or the entire dating process would be considerably easier). The final step of Marshall's (2008) approach uses the lineage with the highest coverage to calibrate the tree and estimate divergences. Our nuclear DNA results suggest that the *set B* date for node 3 had the highest coverage (Table 2). However, we were evaluating several controversial FAs for this node (Table 1) and, by default, the highest coverage will be assigned to the oldest fossil so ESFs cannot be used for this task.

The third method we evaluated, which uses a Bayesian framework to evaluate several candidate fossils in a multicalibration framework (Sanders and Lee 2007), is ideally suited for the task. However, one limitation of this method is that at least some of the candidate calibrations are assumed to be reliable, with just one or two calibrations being evaluated. In addition, multiple calibrations can interact with each other to generate different effective priors; However, the extent of this effect can be evaluated explicitly (Drummond et al. 2006), and our analyses of priors with empty alignments indicated that this was not an issue in our study. Nonetheless, one limitation of our study was that the calibrations for nodes 2 and 3 were evaluated in pairs based on their previous use in other studies and, as such, the best combination may not have been included in our analyses. Rutschmann et al. (2007) recently presented an alternative approach for evaluating the internal consistency of fossil calibrations that compared *s* values from all possible combinations of dates and nodes (72 combinations in our case) (Rutschmann et al. 2007). However, this approach will be subject to the same saturation effects demonstrated in our study and, as such, the effects of using rapidly and slowly evolving gene regions or codon positions for evaluating the internal consistency of calibrations will need to be considered.

There is a growing consensus that the advantages of using multiple independent fossil calibrations significantly outweigh any disadvantages (Ho and Phillips 2009). Multiple calibrations can ameliorate the effects of errors in fossil dates and/or the assignment of fossils to certain nodes (Conroy and van Tuinen 2003; van Tuinen and Dyke 2004), provided that errors are not biased in the same direction. Moreover, the use of multiple calibrations allows the explicit modeling of rate variation among lineages. The limitations of using just one calibration in BEAST analyses for modeling rate variation are highlighted in the chronogram from the mitochondrial data set with third codon positions removed: The two basal branches extending from the tree root on the BEAST chronogram were massively stretched and the remaining internal branches overly compressed (Fig. 1c). The addition of multiple calibrations ameliorated this effect (Fig. 6), presumably resulting in more accurately estimated branch lengths (time) throughout the chronogram. Although the mtDNA3rdExcl ultrametric tree generated in r8s did not suffer from similarly stretched basal branches (results not shown), the approach of Marshall (2008) ultimately relies on just one calibration to date the phylogeny, and our analyses demonstrated the highly variable results that could be obtained using different methods to generate the ultrametric tree (Table 2). Moreover, although this approach might be realistic for groups with exceptionally good fossil records (provided that the hurdle of obtaining a reliable ultrametric tree can be overcome), on its own, it is likely to produce highly misleading results in the majority of cases where the fossil record is less than ideal.

Evaluating the Effects of Saturation on Identifying Reliable Calibrations

The differences in the plausible sets of congruent fossil calibrations identified from the cross-validations from nuclear and mitochondrial DNA, as well as fossil outliers identified from nuclear but not mitochondrial data based on ESF_i values, can be entirely accounted for by saturation effects. The saturation plots revealed strong mitochondrial saturation in the data set (Fig. 2), particularly the third codon position (Fig. 2d). The saturation effects on tree topology, and corresponding age estimates of fossil-calibrated nodes, are clearly evident in Figure 1. Compared with the nuclear chronogram (Fig. 1a), the chronogram from the entire mitochondrial data set had compressed internal branches, which essentially reduced the total distance (time) between nodes 1 and 9 on the chronogram (Fig. 1b). This result was also true for the nDNA and mtDNA ultrametric trees generated in r8s (not shown).

In terms of the cross-validations, the three sets of nuclear cross-validations identified the same four shallow fossil-calibrated nodes (4, 5, 7, and 9) as least congruent with the six other candidate calibrations tested. These nodes also had the highest ESF_i values (Table 2), with nodes 7 and 9 being identified as outliers by three of the four nuclear DNA ultrametric trees. By contrast, mitochondrial cross-validations identified nodes 6 and 8 as least congruent for *sets B* and *C* (and also *set A* when the third codon positions were removed). Thus, for two fossils (*Naja* and *Laticauda*), nuclear DNA favored stem placement (nodes 8 and 8), whereas mtDNA favored crown placement (nodes 7 and 9), directly as the result of saturation effects. Specifically, if a crown group is constrained with the same fossil calibration as its respective stem group, the placement of a fossil at the shallower crown node will return older estimates at other nodes than stem placement, irrespective of data type. However, because mitochondrial distances were artificially shortened (due to compression of internal branches resulting from nucleotide saturation),

the tendency for crown placement to produce much older age estimates for other fossil-calibrated nodes, which was so strongly apparent for nuclear DNA, disappeared for mtDNA: Instead, stem placement resulted in younger age estimates at deeper fossil-calibrated nodes. Similarly, the compressed internal branches for mtDNA resulted in smaller differences between the larger ESF_i values; thus, ESF_i distributions did not deviate from uniformity with the result that fossil outliers were not identified. Evaluating these results in terms of the actual fossils (Table 1 and Supplementary material A) further suggests that misleading results were obtained from the mitochondrial data due to the effects of saturation.

The effects of mitochondrial saturation are also evident in many studies estimating divergence times in snakes. Studies that have relied primarily or entirely on mitochondrial data (Nagy et al. 2003; Guicking et al. 2006; Burbrink and Lawson 2007; Wuster et al. 2007, 2008; Alfaro et al. 2008; Kelly et al. 2009) have recovered 2-fold older age estimates for some advanced snake clades from mitochondrial sequence data (see table 1 in Kelly et al. 2009) than from nuclear sequence data (Sanders and Lee 2008), even when almost exactly the same calibrations were used (Sanders and Lee 2008; Kelly et al. 2009). Jiang et al. (2007) demonstrated accelerated rates of mitochondrial evolution in advanced snakes, suggesting that the extent of nucleotide saturation may be more pronounced than in other taxonomic groups. Nonetheless, the effects of mitochondrial saturation for estimating branch lengths and dating divergences have been well documented for other vertebrate groups such as agamid lizards (Hugall and Lee 2004), squamates (Townsend et al. 2004), tetrapods (Hugall et al. 2007), rodents (Jansa et al. 2006), and across all vertebrates (Phillips 2009). In addition, Brandley et al. (2011) recently demonstrated the importance of data partitioning for obtaining accurate divergence estimates in lizards, particularly drawing attention to the effects of highly saturated mitochondrial third codon positions. We found similarly high levels of third codon mitochondrial saturation (Fig. 2d), yet removing third codon nucleotides did not ameliorate saturation effects for the single-fossil cross-validations (Figs. 2–4). However, removing third codon nucleotides improved the multicalibration BEAST analyses. Posterior distributions for the six fossil-calibrated nodes (with third codon nucleotides excluded) typically converged on age estimates from nuclear DNA (deeper nodes) or were intermediate between the mitochondrial (entire) and nuclear DNA results (shallower nodes). These results suggest that removing highly saturated third codon positions in multicalibration Bayesian analyses might provide a way forward for dealing with mitochondrial saturation, both for evaluating fossil calibrations and for estimating divergences. More importantly, our study demonstrates that saturation strongly influenced different approaches for evaluating candidate calibrations and highlights the need to carefully consider the effects of data type when evaluating fossils.

How Wrong Can We Be?

Pulquerio and Nichols (2006) explored the many factors that can contribute to the highly variable dates obtained using calibrated molecular clocks and posed the question "*how wrong can we be?*" Given that the choice of fossil calibrations is fundamental in obtaining accurate dates, it is vital that the methods used to evaluate candidate fossils are used with a clear understanding of the advantages and disadvantages of each, as well as the effects of data type and other factors on the results. Our study highlighted the fact that nucleotide saturation strongly influences which fossil calibrations are identified as outliers by the cross-validations and ESFs. Previous studies that used cross-validations to evaluate fossil calibrations have tended to use some combination of nuclear and mitochondrial DNA (Near and Sanderson 2004; Near et al. 2005; Noonan and Chippindale 2006b; Alfaro et al. 2008) or nuclear and plastid DNA (Rutschmann et al. 2007), and this was also the case for the fossil coverage approach (Marshall 2008; Davis et al. 2009). We also conducted many of the Bayesian single- and multicalibration analyses using a combined mitochondrial and nuclear data set (with appropriate partitioning), and the results were very similar to those of the mitochondrial data (results not shown), indicating that combining nuclear and mitochondrial data does not inevitably counteract the effects of mitochondrial saturation (but see Brandley et al. 2011). Given that nucleotide saturation typically has the effect of compressing basal branches, it is most likely that older calibrations at shallow nodes will be identified as more congruent with candidate calibrations at deeper nodes by cross-validations using sequence data with high levels of saturation yet not be identified as outliers based on the distribution of ESFs, as was the case in our study. If these calibrations subsequently are used in dating analyses that also rely partially or entirely on saturated DNA, the resultant age estimates will suffer from the compounded effects of two sources of error from nucleotide saturation. Recent studies have demonstrated the potential benefits of appropriate data partitioning (Brandley et al. 2011) and the use of RY coding for mitochondrial data (Phillips 2009) for ameliorating saturation effects on estimating divergence dates. We demonstrate that excluding third codon positions can also ameliorate saturation effects in Bayesian multicalibration analyses, with relevance both for evaluating fossil calibrations and estimating divergences.

To our knowledge, there has been no previous evaluation of the effects of data type (saturation) on approaches for evaluating fossil calibrations (Near and Sanderson 2004; Rutschmann et al. 2007; Sanders and Lee 2007). Given that these approaches are in their infancy, further exploration of the effects of using sequence data with different evolutionary rates for evaluating candidate fossils and their most appropriate placement on a phylogeny is obviously needed. In the meantime, we urge researchers evaluating candidate fossil calibrations to utilize several of the methods

currently available and critically compare the results. Moreover, we think it imperative that researches conduct these analyses using separate nuclear and mitochondrial data sets (rather than combining the data) and use one or more of the various approaches for ameliorating mitochondrial saturation and compare the results, particularly when evaluating fossils that span very different temporal depths on the tree.

Supplementary Material

Supplementary material, including data files and/or online-only appendices, can be found at http://www.sysbio.oxfordjournals.org/.

Funding

This study was supported by funds from the University of California, Irvine.

Acknowledgments

We thank Matt Phillips for insightful comments and for invaluable assistance in generating ultrametric trees in r8s. We thank Andrei Tatarenkov, Jim McGuire, Frank Anderson, and one anonymous reviewer for extensive constructive comments that significantly improved the manuscript. J.S.K. thanks the Australian Research Council for ongoing support. V.L. thanks the ARC Centre of Excellence for Coral Reef Studies at James Cook University for support while finalizing the manuscript.

References

Albino A.M. 2000. New record of snakes from the Cretaceous of Patagonia (Argentina). Geodiversitas. 22:247–253.

Alfaro M.E., Karns D.R., Voris H.H., Brock C.D, Stuart B.L. 2008. Phylogeny, evolutionary history, and biogeography of Oriental-Australian rear-fanged water snakes (Colubroidea: Homalopsidae) inferred from mitochondrial and nuclear DNA sequences. Mol. Phylogenet. Evol. 46:576–593.

Benton M.J., Ayala F.J. 2003. Dating the tree of life. Science. 300:1698–1700.

Brandley M.C., Wang Y., Guo X., Montes de Oca A.N., Feria-Ortiz M., Hikida T., Ota H. 2011. Accommodating heterogeneous rates of evolution in molecular divergence dating methods: an example using intercontinental dispersal of *Plestiodon* (*Eumeces*) lizards. Syst. Biol. 60:3–15.

Burbrink F.R., Lawson R. 2007. How and when did Old World ratsnakes disperse into the New World? Mol. Phylogenet. Evol. 43:173–189.

Conroy C.J., van Tuinen M. 2003. Extracting time from phylogenies: positive interplay between fossil and genetic data. J. Mammal. 84:444–455.

Davis R.B., Baldauf S.L., Mayhew P.J. 2009. Eusociality and the success of the termites: insights from a supertree of dictyopteran families. J. Evol. Biol. 22:1750–1761.

Donoghue M.J., Benton M.J. 2007. Rocks and clocks: calibrating the Tree of Life using fossils and molecules. Trends Ecol. Evol. 22:424–431.

Doyle J.A., Donoghue, M.J. 1993. Phylogenies and angiosperm diversification. Paleobiology. 19:141–167.

Drummond A.J., Ho S.Y.W., Phillips M.J., Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4:699–710.

Drummond A.J., Ho S.Y.W., Rawlence N., Rambaut A. 2007. A rough guide to BEAST 1.4. 1–42.

Drummond A.J., Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7:214.

Forstner M.R.J., Davis S.K., Arevalo E. 1995. Support for the hypothesis of anguimorph ancestry for the suborder serpentes from phylogenetic analysis of mitochondrial DNA sequences. Mol. Phylogenet. Evol. 4:93–102.

Gardner J.D., Cifelli R.L. 1999. A primitive snake from the Cretaceous of Utah. Special Papers Palaeontology. Volume 60. p. 87–100.

Gernhard T. 2008. The conditioned reconstructed response. J. Theor. Biol. 253:769–778.

Graur D., Martin W. 2004. Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. Trends Genet. 20:80–86.

Groth J.G., Barrowclough G.F. 1999. Basal divergences in birds and the phylogenetic utility of the nuclear RAG-1 gene. Mol. Phylogenet. Evol. 12:115–123.

Guicking D., Lawson R., Joger U., Wink M. 2006. Evolution and phylogeny of the genus *Natrix* (Serpentes: Colubridae). Biol. J. Linn. Soc. 87:127–143.

Head J., Holroyd P.A., Hutchison J.H., Ciochon R.L. 2005. First report of snakes (Serpentes) from the late middle Eocene Pondaung formation, Myanmar. J. Vert. Paleontol. 25:246–250.

Ho S.Y.W., Phillips M.J. 2009. Accounting for calibration uncertaintly in phylogenetic estimation of evolutionary divergence times. Syst. Biol. 58:367–380.

Holman J.A. 1984. *Texasophis galbreathi*, new species, the earliest New World colubrid snake. J. Vert. Paleontol. 3:223–225.

Hugall A.F., Foster R., Lee M.S.Y. 2007. Calibration choice, rate smoothing, and the pattern of tetrapod diversification according to the long nuclear gene RAG-1. Syst. Biol. 56:543–563.

Hugall A.F., Lee M.S.Y. 2004. Molecular claims of Gondwanan age for Australian agamid lizards are untenable. Mol. Biol. Evol. 21:2102–2110.

Hugall A.F., Lee M.Y.L. 2007. The likelihood node density effect and consequences for evolutionary studies of molecular rates. Evolution. 61:2293–2307.

Jansa S.A., Barker F.K., Heaney L.R. 2006. The pattern and timing of diversification of Philippine endemic rodents: evidence from mitochondrial and nuclear gene sequences. Syst. Biol. 55:73–88.

Jiang Z.J., Castoe T.A., Austin C.C., Burbrink F.T., Herron M.D., McGuire J.A., Parkinson C.L., Pollock D.D. 2007. Comparative mitochondrial genomics of snakes: extraordinary substitution rate dynamics and functionality of the duplicate control region. BMC Evol. Biol. 7:123

Kelly C.M.R., Barker N.P., Villet M.H., Broadley D.G. 2009. Phylogeny, biogeography and classification of the superfamily Elapoidea: a rapid radiation in the late Eocence. Cladistics. 25:38–63.

Keogh J.S. 1998. Molecular phylogeny of elapid snakes and a consideration of their biogeographic history. Biol. J. Linn. Soc. 63:177–203.

Keogh J.S., Shine R., Donnellan S. 1998. Phylogenetic relationships of terrestrial Australo-Papuan elapid snakes (Subfamily Hydrophiinae) based on cytochrome *b* and 16S rRNA sequences. Mol. Phylogenet. Evol. 10:67–81.

Kuch U., Muller J., Modden C., Mebs D. 2006. Snake fangs from the lower Miocene of Germany: evolutionary stability of perfect weapons. Naturewissenschaften. 93:84–87.

Labarga A., Valentin F., Andersson M., Lopez R. 2007. Web services at the European bioinformatics institute. Nucleic Acids Res. (Web Services Issue):1–6.

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

Lee M.S.Y. 1999. Molecular clock calibrations and metozoan divergence dates. J. Mol. Evol. 49:385–391.

Lukoschek V., Keogh J.S. 2006. Molecular phylogeny of sea snakes reveals a rapidly diverged adaptive radiation. Biol. J. Linn. Soc. 89:523–539.

Lukoschek V., Waycott M., Marsh H. 2007. Phylogeographic structure of the olive sea snake, *Aipysurus laevis* (Hydrophiinae) indicates recent Pleistocene range expansion but low contemporary gene flow. Mol. Ecol. 16:3406–3422.

Magallon S., Sanderson M.J. 2001. Absolute diversification rates in angiosperm clades. Evolution. 55:1762–1780.
Marshall C.R. 2008. A simple method for bracketing absolute divergence times on molecular phylogenies using multiple fossil calibration points. Am. Nat. 171:726–742.
McDowell J.R. 1987. Systematics. In: Seigel R.A., Collins J.T., Novak S.S., editors. Snakes—ecology and evolutionary biology. New York: Macmillan. p. 3–50.
McGuire J.A., Witt C.C., Altshuler D.L., Remsen J.V. Jr. 2007. Phylogenetic systematics and biogeography of hummingbirds: Bayesian and maximum likelihood analyses of partitioned data and selection of an appropriate partitioning strategy. Syst. Biol. 56:837–856.
Nagy Z.T., Joger U., Wink M., Glaw F., Vences M. 2003. Multiple colonisation of Madagascar and Socotra by colubrid snakes: evidence from nuclear and mitochondrial gene phylogenies. Proc. R. Soc. Lond. B. Biol. Sci. 270:2613–2621.
Near T.J., Meylan P.A., Shaffer H.B. 2005. Assessing concordance of fossil calibration points in molecular clock studies: an example using turtles. Am. Nat. 165:137–146.
Near T.J., Sanderson M.J. 2004. Assessing the quality of molecular divergence time estimates by fossil calibrations and fossil-based model selection. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 359:1477–1483.
Noonan B.P., Chippindale P.T. 2006a. Dispersal and vicariance: the complex evolutionary history of boid snakes. Mol. Phylogenet. Evol. 40:347–358.
Noonan B.P., Chippindale P.T. 2006b. Vicariant origin of Malagasy reptiles supports late Cretaceous Antarctic land bridge. Am. Nat. 168:730–741.
Palumbi S.R. 1996. Nucleic acids II: the polymerase chain reaction. In: Hillis D.M., Moritz C., Mable B.K., editors. Molecular systematics. Sunderland (MA): Sinauer Associates, Inc. p. 205–247.
Parmley D., Holman J.A. 2003. *Nebraskophis* HOLMAN from the Late Eocene of Georgia (USA), the oldest known North American colubrid snake. Acta Zool. Cracov. 46:1–8.
Phillips M.J. 2009. Branch-length estimation bias misleads molecular dating for a vertebrate mitochondrial phylogeny. Gene. 441: 132–140.
Posada D., Crandall K.A. 1998. MODELTEST: testing the model of DNA substitutions. Bioinformatics. 14:817–818.
Pulquerio M.J.F., Nichols R.A. 2006. Dates from the molecular clock: how wrong can we be? Trends Ecol. Evol. 22:180–184.
Pyron R.A., Burbrink F.T., Colli G.R., Montes A.N., de Oca Vitt L.J., Kuczynski C.A., Wiens J.J. 2011. The phylogeny of advanced snakes (Colubroidea), with discovery of a new subfamily and comparison of support methods for likelihood trees. Mol. Phylogenet. Evol. 58:329–342.
Rage J.C. 1984. Serpentes. In: Encyclopedia of paleoherpetology. Stuttgart (Germany): Gustav Fischer Verlag. p. 1–79.
Rage J.C. 1987. Fossil history. In: Seigel R.A., Collins J.T., Novak S.S., editors. Snakes—ecology and evolutionary biology. New York: Macmillan. p. 51–76.
Rage J.C. 1988. The oldest known colubrid snakes. State of the art. Acta Zool. Cracov. 31:457–474.
Rage J.C. 2001. Fossil snakes from the Paleocene of Sao Jose de Itaborai, Brazil. Part II. Boidae. Palaeovertebrata. 30:111–150.
Rage J.C., Buffetaut E., Buffetaut-Tong H., Chaimanee Y., Ducrocq S., Jaeger J.J., Suteethorn V. 1992. A colubrid in the late Eocene of Thailand: the oldest known Colubridae (Reptilia, Serpentes). C. R. Acad. Sci. 314:1085–1089.
Rage J.C., Gupta S.S., Prasad G.V.R. 2001. Amphibians and squamates from the Neogene Siwalik beds of Jammu and Kashmir, India. Palaont. Z. 75:197–208.
Rage J.C., Werner C. 1999. Mid-Cretaceous (Cenomanian) snakes from Wadi Abu Hashim, Sudan: the earliest snake assemblage. Palaeontol. Afr. 35:85–110.
Rambaut A., Bromham L. 1998. Estimating divergence times from molecular sequences. Mol. Biol. Evol. 15:442–448.
Ronquist F., Huelsenbeck J.P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19:1572–1574.
Rutschmann F., Eriksson T., Salim K.A., Conti E. 2007. Assessing calibration uncertainty in molecular dating: the assignment of fossils to alternative calibration points. Syst. Biol. 56:591–608.
Saint K.M., Austin C.C., Donnellan S., Hutchison M.N. 1998. C-mos, a nuclear marker useful for squamate phylogenetic analysis. Mol. Phylogenet. Evol. 10:259–263.
Sanders K., Lee M.S.Y. 2007. Evaluating molecular clock calibrations using Bayesian analyses with soft and hard bounds. Biol. Lett. 3:275–279.
Sanders K.L., Lee M.Y.L. 2008. Molecular evidence for a rapid late-Miocene radiation of Australasian venomous snakes (Elapidae: Colubroidea). Mol. Phylogenet. Evol. 46:1180–1188.
Sanders K.L., Lee M.Y.L., Leijs R., Foster R., Keogh J.S. 2008. Molecular phylogeny and divergence dates for Australasian elapids and sea snakes (Hydrophiinae): evidence from seven genes for rapid evolutionary radiations. J. Evol. Biol. 21:682–695.
Sanderson M.J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol. Biol. Evol. 14:1218–1231.
Sanderson M.J. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol. Biol. Evol. 19:101–109.
Sanderson M.J. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics. 19:301–302.
Scanlon J.D., Lee M.S.Y., Archer M. 2003. Mid-Tertiary elapid snakes (Squamata, Colubroidea) from Riversleigh, northern Australia: early steps in a continent-wide adaptive radiation. Geobios. 36:573–601.
Schwartz R.S., Mueller R.L. 2010. Branch length estimation and divergence dating: estimates of error in Bayesian and maximum likelihood frameworks. BMC Evol. Biol. 10:5
Slowinski J.B., Keogh J.S. 2000. Phylogenetic relationships of elapid snakes based on cytochrome *b* mtDNA sequences. Mol. Phylogenet. Evol. 15:157–164.
Slowinski J.B., Knight A., Rooney A.P. 1997. Inferring species trees from gene trees: a phylogenetic analysis of the Elapidae (Serpentes) based on the amino acid sequences of venom proteins. Mol. Phylogenet. Evol. 8:349–362.
Smith A.B., Pisani D., Mackenzie-Dodds J.A., Stockley B., Webster B.L., Littlewood D.T.J. 2006. Testing the molecular clock: molecular and paleontological estimates of divergence times in the Echinoidea (Echinodermata). Mol. Biol. Evol. 23:1832–1851.
Swofford D. 2000. PAUP*: phylogenetic analysis using parsimony (*and other methods). 4th ed. Sunderland (MA): Sinauer Associates, Inc.
Szyndlar Z., Rage J.C. 1990. West Palearctic cobras of the genus *Naja* (Serpentes: Elapidae): interrelationships among extinct and extant species. Amphibia-Reptilia. 11:385–400.
Szyndlar Z., Rage J.C. 1999. Oldest fossil vipers (Serpentes: Viperidae) from the Old World. Kaupia. 8:9–20.
Townsend T.M., Larson A., Louis E., Macey J.R. 2004. Molecular phylogenetics of squamata: the position of snakes, amphisbaenians, and dibamids, and the root of the squamate tree. Syst. Biol. 53: 735–757.
van Tuinen M., Dyke G.J. 2004. Calibration of galliform molecular clocks using multiple fossils and genetic partitions. Mol. Phylogenet. Evol. 30:74–86.
Vidal N., Delmas A.S., David P., Cruaud C., Couloux A., Hedges S.B. 2007. The phylogeny and classification of caenophidian snakes inferred from seven nuclear protein-coding genes. C. R. Biol. 330: 182–187.
Wiens J.J., Brandley, M.C., Reeder, T.W. 2006. Why does a trait evolve multiple times within a clade? Repeated evolution of snakelike body form in squamate reptiles. Evolution. 60:123–141.
Wuster W., Crookes S., Ineich I., Mane Y., Pook C.E., Trape J.-F., Broadley D.G. 2007. The phylogeny of cobras inferred from mitochondrial DNA sequences: evolution of venom spitting and the phylogeography of the African spitting cobras (Serpentes: Elapidae: *Naja nigricollis* complex). Mol. Phylogenet. Evol. 45:437–453.
Wuster W., Peppin L., Pook C.E., Walker D.E. 2008. A nesting of vipers: phylogeny and historical biogeography of the Viperidae (Squamata: Serpentes). Mol. Phylogenet. Evol. 49:445–459.
Yang Z., Rannala B. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Mol. Biol. Evol. 23:212–226.

APPENDIX A1. Taxa used in three approaches for evaluating candidate fossil calibrations with details of sequences generated or retrieved from GenBank

	Family or subfamily	16S rRNA	ND4	Cyt *b*	c-*mos*	RAG-1
Lizard Taxa						
Varanus spp.	Varanidae	EF193687	DQ525172	NC010974	EU366456	AY662608
Calotes spp.	Iguanidae	NC009683	NC009683	NC009683	AF137525	AY662584
Scoliophidian Taxa						
Ramphotyphlops braminus	Typhlopidae	DQ343649	DQ343649	DQ343649	AF544717	AY444062
Typhlops jamaicensis	Typhlopidae	AM236345	AM236345	AM236345	AF544733	—
Henophidian Taxa						
Boa constrictor	Boidae	AM236348	AM236348	AM236348	AF471115/AF544676	AY988064
Eunectes notaeus	Boidae	AM236347	AM236347	AM236347	AY099964	AY988063
Aspidites melanocephalus	Pythonidae	PT EF545061	—	U69741	DQ465557	DQ465560
Loxocemus bicolor	Pythonidae	AF544828	—	AY099993	AF544730/AY444035	AY444061
Morelia spp.	Pythonidae	EF545049	—	EF545099	AF544723	AY988069
Python spp.	Pythonidae	EF545066	73760200	73760200	AF544675/AY099968	EU624119
Caenophidian Taxa						
Acrochordus spp.	Acrochordidae	AB177879	AB177879	AB177879	EU366454/EU366443	AY487388/AY988070
Elapidae						
Aipysurus laevis	Hydrophiinae (Marine)	DQ233989	EF506658	DQ233913	FJ587169	FJ587087
Emydocephalus annulatus	Hydrophiinae (Marine)	DQ234000	FJ593195	DQ233939	FJ587171	FJ587093
Hydrelaps darwiniensis	Hydrophiinae (Marine)	DQ234046	FJ593199	DQ233948	FJ587175	FJ587097
Parahydrophis mertoni	Hydrophiinae (Marine)	DQ234047	FJ593201	DQ233974	FJ587177	FJ587099
Acalyptophis peronii	Hydrophiinae (Marine)	DQ234005	FJ593202	DQ233924	FJ587178	FJ587100
Astrotia stokesii	Hydrophiinae (Marine)	DQ234008	FJ593205	DQ233927	FJ587181	FJ587103
Disteira major	Hydrophiinae (Marine)	DQ234017	FJ593209	DQ233936	FJ587185	FJ587107
Hydrophis elegans	Hydrophiinae (Marine)	DQ234021	FJ593216	DQ233950	FJ587190	FJ587113
Hydrophis pacificus	Hydrophiinae (Marine)	DQ234035	FJ593224	DQ233963	FJ587196	FJ587119
Lapemis curtis	Hydrophiinae (Marine)	DQ234040	FJ593227	DQ233968	FJ587199	FJ587122
Pelamis platurus	Hydrophiinae (Marine)	DQ234051	FJ593233	DQ233977	FJ587201	FJ587124
Acanthophis spp.	Hydrophiinae (Terrestrial)	EU547161	EU547016	EU547063	EU546926	EU546887
Austrelaps superbus	Hydrophiinae (Terrestrial)	EU547176	EU547030	EU547078	EU546940	EU546901
Cacophis squamulosus	Hydrophiinae (Terrestrial)	EU547150	EU547007	EU547052	EU366451	EU366440
Hemiaspis dameli	Hydrophiinae (Terrestrial)	EU547171	EU547025	EU547073	EU546935	EU546896
Hoplocephalus spp.	Hydrophiinae (Terrestrial)	EU547177	EU547031	EU547079	EU546941	EU546902
Laticauda colubrina	Hydrophiinae (Terrestrial)	EU547138	EU546998	EU547040	EU366446/AF544702/AY05893	EU366433/AY487404
Laticauda laticaudata	Hydrophiinae (Terrestrial)	FJ587203	FJ593191	FJ587153	FJ587158	FJ587082
Micropechis ikaheka	Hydrophiinae (Terrestrial)	EU547140	EU547000	EU547042	EU366449	EU366435

Continued

APPENDIX A1. (*Continued*)

	Family or subfamily	16S rRNA	ND4	Cyt *b*	c-*mos*	RAG-1
Elapidae						
Oxyuranus scutellatus	Hydrophiinae (Terrestrial)	EU547149	EU547003	EU547051	EU546916	EU546877
Suta fasciata	Hydrophiinae (Terrestrial)	EU547162	EU547017	EU547064	EU546927	EU546888
Vermicella intermedia	Hydrophiinae (Terrestrial)	EU547153	EU547010	EU547055	EU546919	EU546880
Bungarus fasciatus	Elapinae	JF357944	JF357926	EU547086/AF217830	JF357954	EU366438/AY487389
Dendroaspis spp.	Elapinae	JF357945	JF357927	JF357936	AF544735/FJ387197	AY487395
Elapsoidea spp.	Elapinae	JF357946	JF357928	JF357937	AF544678/AY058930	AY487373
Micrurus spp.	Elapinae	JF357947	JF357929	JF357938	EF137421/AF544708	AY487411
Naja kaouthia	Elapinae	JF357948	JF357930	JF357939	AY058938	JF412633
Naja naja	Elapinae	DQ343648	DQ343648	DQ343648	EU366445	EU366432
Naja melanoleuca	Elapinae	JF357949	DQ897689	DQ897732	JF357955	JF357956
Naja nivea	Elapinae	EU624272	AY058983	AF217827	AY058939	—
Colubridae						
Alsophis spp.	Xenodontinae	AF158517	U49308	AF471085	AF471126/AF544694	AY487376
Coluber spp.	Colubrinae	L01770/AY188081	AF138746	EU180489	AY486937	—
Dinodon semicarinatus	Colubrinae	AB008539	AB008539	AB008539	AF471163	AY662611
Lamprophis fuliginosus	Boodontinae	JF357950	JF357932	JF357940	AF544686	AY487378
Leioheterodon modestus	Pseudoxyrhophiinae	JF357951	JF357933	JF357941	AY058933	AY487377
Natrix natrix	Natricinae	AF158530	AY487800	AY487756	AF471121/AF544697	—
Viperidae						
Bitis arietans	Viperinae	JF357952	JF357934	JF357942	AF471130	—
Bothrops/Bothriechis	Crotalinae	AY223672	AY223636	AY223598/AF191587	DQ469788/AF544680	DQ469790/AY487374
Gloydius spp.	Crotalinae	JF357953	JF357935	JF357943	AF435019	AY662614