Systematic Biology Advance Access published August 13, 2011

For Consideration in Systematic Biology

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]ARC Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, QLD 4811, Australia.

[3]*Division of Evolution, Ecology and Genetics, Research School of Biology, The Australian National University, Canberra, ACT 0200, Australia*

Corresponding Author:

Vimoksalehi Lukoschek

ARC Centre of Excellence for Coral Reef Studies,

James Cook University,

Townsville, QLD, 4811, Australia

Phone: +61-7-47816294

Email: vimoksalehi.lukoschek@jcu.edu.au

Running head: Evaluating fossil calibrations

Abstract

Evolutionary and biogeographic studies increasingly rely on calibrated molecular clocks to date key events. While 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 et al. (2005); the empirical fossil coverage method of Marshall (2008); and the Bayesian multi-calibration method of Sanders and Lee (2007), 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 analysing a combined nuclear and mitochondrial dataset. While 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 multi-calibration 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.

Keywords: Bayesian dating, fossil calibrations, cross-validation, nucleotide saturation, molecular clock,

Caenophidia, Hydrophiinae, 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, while 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 (Rutschman 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; Rambaut and Bromham 1998; Drummond et al. 2006); and unbalanced taxon sampling potentially resulting in node

density artefacts (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 towards 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, while the effects of saturation have been documented for estimating divergence times (Hugall and Lee 2004; Townsend et al. 2004; Hugall et al. 2007; Phillips 2009; Brandley et al. 2011), 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; Sanders and Lee 2008; Sanders et al. 2008; Wuster 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,b), 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; Alfaro et al. 2008; Sanders and Lee 2008; Wuster et al. 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 multi-calibration 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 multi-calibration approach evaluates one or more alternative calibrations in a set by comparing the Bayesian prior and posterior probabilities 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 dataset (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) analysing a combined nuclear and mitochondrial dataset. 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 were 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 eight henophidian taxa) were chosen based on the availability of nuclear and mitochondrial sequences

(Appendix 1) 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 datasets with appropriate taxon sampling. Details of the fossil calibrations evaluated are given in Table 1. We constructed nuclear and mitochondrial datasets, each with identical taxon sampling, using >100 novel sequences generated for this study and published sequences obtained from GenBank (Appendix 1). The mitochondrial data comprised 16S rRNA (454 bp), ND4 (672 bp), and cytochrome *b* (1095 bp) and the nuclear data comprised 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 1). For some taxa mitochondrial fragments and/or nuclear genes were concatenated from two individuals or two congeneric species to minimise the amount of missing sequence data, in which case the highest common taxon name was assigned (Appendix 1). 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 non-functional 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 datasets. 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 datasets 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 over-parameterisation. 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, page 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 dataset, but the results were virtually identical to those obtained for the mtDNA data alone, so we do not present the results of the combined dataset.

Bayesian relaxed molecular clocks, which assume rates of molecular evolution are uncorrelated but log-normally 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 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 ten calibrated nodes (Fig. 1 – Tree Root and nodes 1 to 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 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 and fossil age at each node was calculated as $D_i = (\mathrm{MA}_i - \mathrm{FA}_i)$, 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 molecular and fossil ages 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 fossil age for each candidate fossil calibrated node (x) was used as a single calibration prior in the BEAST analysis and $\overline{D}_x$ and its SE were calculated from the remaining nine candidate fossil-dated nodes.

SS values were then calculated as the sum of the squared differences between the molecular (MA) and fossil (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% CI = 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; 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 most recent common ancestor (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 datasets 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 datasets 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 datasets 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 1) for two outgroup taxa (the lizard genera *Varanus* and *Calotes*) to the datasets. 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 semi-parametric penalised likelihood (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; 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

$ESF_i = \frac{FA_i}{NTL_i}$, 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_i values 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, pg 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 for details).

Bayesian Analyses to Evaluating Multi-Calibration 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 multi-calibration framework. This method compares the prior and posterior distributions of the 95% 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 multi-calibration 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 datasets, 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 datasets. 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, while 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 datasets 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; Lukoschek and Keogh 2006; Wuster et al. 2007; Sanders and Lee 2008; Sanders et al. 2008; Kelly et al. 2009; Pyron et al. 2010). Data matrices and relevant trees have been submitted to TreeBASE (#11272). Eight of the ten candidate calibration nodes had extremely high support with ≥99% posterior probabilities (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 dataset (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, while 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 fossil ages 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* 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 sets *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, 9) that returned much older nuclear DNA values for fossil ages at other candidate calibration nodes either produced younger or only slightly older estimates of fossil ages 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 molecular ages for fossil dates at other nodes was more extreme for the mitochondrial than nuclear data, and this was true for both mitochondrial datasets (Figs. 3b & c). By contrast, node 1 produced older ages at other nodes for both mtDNA datasets, 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 (Figs. 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* (Figs. 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 (Figs. 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 (Figs. 5b & c). Interestingly, node 1 was the most incongruent fossil for the younger *set A* calibrations for the entire mtDNA dataset 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 dataset 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).

Fossil Coverage and Fossil Outliers

The four ultrametric trees obtained from the nDNA dataset 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, 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) 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 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 artefact 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, while 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 Multi-Calibration 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). As such, we conducted the multi-calibration analyses using the six fossil calibrated nodes selected by the nuclear data.

Multi-calibration analyses using nuclear DNA revealed similarities and differences between the estimated mean ages and 95% Highest Posterior Densities (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), while 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, while 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

464calibration priors and, with one exception, also for the corresponding nDNA age estimates. By contrast,
465mitochondrial node-1 age estimates for all calibration sets were similar to the calibration prior and to nuclear
466DNA age estimates, and this was true for both mitochondrial datasets (Fig. 6). Nonetheless, the tendency
467for mtDNA to return older age estimates at shallow nodes and the tree root was much pronounced when the
468third codon positions were excluded, with mtDNA3rdExcl age estimates for nodes 6 and 8 age intermediate
469to the nDNA and mtDNA age estimates, and tending to converge on mean nDNA age estimates for node 3
470and the tree root (Fig. 6). While mitochondrial age estimates for node 2 from the entire dataset showed the
471same tendency as nuclear ages to converge on the unconstrained *set D* age estimates (irrespective of the
472calibration prior used), this was not the case for mtDNA with third codon positions excluded (Fig. 6).
473Indeed, with the exception of *set C*, the node-2 mtDNA3rdExcl age estimates tended to converge on the
474calibration prior resulting in age estimates that were younger than the corresponding nDNA estimates, and
475this was also true for the node-3 *set A* age estimate (Fig. 6).
476
477*Discussion*
478Increasing awareness of the importance of identifying reliable fossils to calibrate molecular clocks has
479resulted in the development of several methods for evaluating and employing fossil calibrations (reviewed
480by Ho and Phillips 2009). Each approach has advantages and limitations, as we demonstrate by comparing
481three different approaches with particular emphasis on the impact of nucleotide saturation on the fossils
482selected.
483
484The cross-validation method (Near et al. 2005) discards calibrations until an internally consistent set is
485obtained, and in the process, may discard calibrations with the best temporal coverage because they are
486inconsistent with the remaining calibrations . Nonetheless, the method has been used in several recent
487studies (Near and Sanderson 2004; Noonan and Chippendale 2006b; Rutschmann et al. 2007; Alfaro et al.
4882008). By contrast, the use of empirical scaling factors aims to identify one fossil with the best empirical
489coverage (Marshall 2008); however, accurate results are highly dependant on meeting the assumptions of the
490method (see below). Unlike the cross-validation approach, empirical scaling factors (*ESFs*) have only been
491used in one previous study (Davis et al. 2009). This study obtained an ultrametric tree in r8s using penalised

likelihood with log penalty function (following the advice of Marshall 2008), based on empirical evidence that penalised likelihood (PL) using the log penalty function produces the most reliable ultrametric trees (Smith et al. 2006). However, Davies 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 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 fossil ages 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 multi-calibration 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 modelling of rate variation among lineages. The limitations of using just one calibration in BEAST analyses for modelling rate variation are highlighted in the chronogram from the mitochondrial dataset 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, while 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 dataset (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 dataset 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) while 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 two-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 multi-calibration 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 multi-calibration 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 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 empirical scaling factors. 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 (Davies et al. 2009; Marshall 2008). We also conducted many of the Bayesian single and multi-calibration analyses using a combined mitochondrial and nuclear dataset (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 empirical scaling factors, 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 multi-calibration 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 utilise 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 datasets (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.

Acknowledgements

This study was supported by funds from the University of California, Irvine. 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. JSK thanks the Australian Research Council for ongoing support. VL thanks the ARC Centre of Excellence for Coral Reef Studies at JCU for support while finalizing the manuscript.

References

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

Alfaro, M.E., D.R. Karns, H.H. Voris, C.D. Brock, B.L. Stuart. 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., F.J. Ayala. 2003. Dating the tree of life. Science 300:1698-1700.

Brandley, M. C., Y. Wang, X. Guo, A. N. Montes de Oca, M. Feria-Ortiz, T. Hikida, H. Ota. 2001. 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., R. Lawson. 2007. How and when did Old World ratsnakes disperse into the New World? Mol. Phylogenet. Evol. 43:173-189.

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

Davis, R. B., S. L. Baldauf, P. J. Mayhew. 2009. Eusociality and the success of the termites: insights from a

supertree of dictyopteran families. J. Evol. Biol. 22:1750-1761.

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

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

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

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

Drummond, A.J., S.Y.W. Ho, N. Rawlence, A. Rambaut. 2007. A Rough Guide to BEAST 1.4. 1-42.

Forstner, M.R.J., S.K. Davis, E. Arevalo. 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., R. L. Cifelli. 1999. A primitive snake from the Cretaceous of Utah. Special Papers Palaeontol. 60:87-100.

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

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

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

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

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

Ho, S.Y.W., M.J. Phillips. 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. Vertebr. Paleontol. 3:223-225.

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

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

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

Kelly, C.M.R., N.P. Barker, M.H. Villet, D.G. Broadley. 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., R. Shine, S. Donnellan, 1998. Phylogenetic relationships of terrestrial Australo-Papuan elapid snakes (Subfamily Hydrophiinae) based on cytochrome *b* and 16S r RNA sequences. Mol. Phylogenet. Evol. 10: 67-81.

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

Labarga, A., F. Valentin, M. Andersson, R. Lopez. 2007. Web Services at the European Bioinformatics Institute. Nucleic Acids Res. Web Services Issue 1-6.

Lawson, R., J.B. Slowinski, B.I. Crother, F.R. Burbrink. 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., J.S. Keogh. 2006. Molecular phylogeny of sea snakes reveals a rapidly diverged adaptive radiation. Biol. J. Linn. Soc. 89:523-539.

Lukoschek, V., M. Waycott, H. Marsh. 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., M.J. Sanderson. 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 calbration points. Am. Nat. 171:726-742.

McDowell, J.R. 1987. Systematics. In: Snakes - ecology and evolutionary biology (eds. R.A. Seigel, J.T. Collins, S.S. Novak), pp. 3-50. New York: Macmillan.

McGuire, J. A., C. C. Witt, D. L. Altshuler, J. V. Remsen 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., U. Joger, M. Wink, F. Glaw, M. Vences. 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., P.A. Meylan, H.B. Shaffer. 2005. Assessing concordance of fossil calibration points in molecular clock studies: an example using turtles. Am. Nat. 165:137-146.

Near, T.J., M.J. Sanderson. 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., P.T. Chippindale. 2006a. Dispersal and vicariance: the complex evolutionary history of boid snakes. Mol. Phylogenet. Evol. 40:347-358.

Noonan, B.P., P.T. Chippindale. 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: Molecular Systematics (eds. D.M. Hillis, C. Moritz, M. B.K.), pp. 205-247. Sunderland, MA: Sinauer Associates, Inc.

Parmley, D., J.A. Holman. 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., K.A. Crandall. 1998. MODELTEST: testing the model of DNA substitutions. Bioinformatics 14:817-818.

Pulquerio, M.J.F., R.A. Nichols. 2006. Dates from the molecular clock: how wrong can we be? Trends

Ecol. Evol. 22:180-184.

Pyron R.A., F.T. Burbrink, G.R. Colli, A.N. Montes de Oca, L.J. Vitt, C.A. Kuczynski, J.J. Wiens. (2010) The phylogeny of advanced snakes (Colubroidea), with discovery of a new subfamily and comparison of support methods for likelihood trees. Mol. Phylogenet. Evol. doi:10.1016/j.ympev.2010.11.006

Rage, J.C. (1984) Serpentes. In: Encyclopedia of Paleoherpetology, pp. 1-79. Stuttgart: Gustav Fischer Verlag.

Rage, J.C. (1987) Fossil history. In: Snakes - ecology and evolutionary biology (eds. R.A. Seigel, J.T. Collins, S.S. Novak), pp. 51-76. New York: Macmillan.

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. II Boidae. Palaeovert. 30:111-150.

Rage, J.C., E. Buffetaut, H. Buffetaut-Tong, Y. Chaimanee, S. Ducrocq, J.J. Jaeger, V. Suteethorn. 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., S.S.Gupta, G.V.R. Prasad. 2001. Amphibians and squamates from the Neogene Siwalik beds of Jammu and Kashmir, India. Palaeontologische Zeitschrift 75:197-208.

Rage, J.C., C. Werner. 1999. Mid-Cretaceous (Cenomanian) snakes from Wadi Abu Hashim, Sudan: The earliest snake assemblage. Palaeontol. Afr. 35:85-110.

Rambaut, A., L. Bromham. 1998. Estimating divergence times from molecular sequences. Mol. Biol. Evol. 15:442-448.

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

Rutschmann, F., T. Eriksson, K.A. Salim, E. Conti. 2007. Assessing calibration uncertainty in molecular dating: the assignment of fossils to alternative calibration points. Syst. Biol. 56:591-608.

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

Sanders, K., M.S.Y. Lee. 2007. Evaluating molecular clock calibrations using Bayesian analyses with soft

and hard bounds. Biol. Lett. 3:275-279.

Sanders, K.L., M.Y.L. Lee. 2008. Molecular evidence for a rapid late-Miocene radiation of Australasian venomous snakes (Elapidae: Colubroidea). Mol. Phylogenet. Evol. 46:1180-1188.

Sanders, K.L., M.Y.L. Lee, R. Leijs, R. Foster, J.S. Keogh. 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., M.S.Y. Lee, M. Archer. 2003. Mid-Tertiary elapid snakes (Squamata, Colubroidea) from Riversleigh, northern Australia: early steps in a continent-wide adaptive radiation. Geobios 36:573-601.

Slowinski, J.B., J.S. Keogh. 2000. Phylogenetic relationships of elapid snakes based on cytochrome *b* mtDNA sequences. Mol. Phylogenet. Evol. 15:157-164.

Slowinski, J.B., A. Knight, A.P. Rooney. 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., D. Pisani, J. A. Mackenzie-Dodds, B. Stockley, B. L. Webster, D. T. J. Littlewood. 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). 4.0 ed. Sunderland, MA: Sinauer Associates, Inc.

Szyndlar, Z., J.C. Rage. 1990. West Palearctic cobras of the genus *Naja* (Serpentes: Elapidae): interrelationships among extinct and extant species. Amphib-Reptilia 11:385-400.

Szyndlar, Z., J.C. Rage. 1999. Oldest fossil vipers (Serpentes: Viperidae) from the Old World. Kaupia 8:9-

20.

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

van Tuinen, M., G.J. Dyke. 2004. Calibration of galliform molecular clocks using multiple fossils and genetic partitions. Mol. Phylogenet. Evol. 30:74-86.

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

Wuster, W., S. Crookes, I. Ineich, Y. Mane, C.E. Pook, J.-F. Trape, D.G. Broadley. 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., L. Peppin, C.E. Pook, D.E. Walker. 2008. A nesting of vipers: phylogeny and historical biogeography of the Viperidae (Squamata: Serpentes). Mol. Phylogenet. Evol. 49:445-459.

Yan, J., H. Li, K. Zhou. 2008. Evolution of the mitochondrial genome in snakes: gene rearrangements and phylogenetic relationships. BMC Genet. 9:569.

Yang, Z., B. Rannala. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Mol. Biol. Evol. 23:212-226.

Figure Legends

Figure 1: Bayesian chronograms from BEAST analyses with tree roots constrained to 97 (92-120) Myr indicating the position of ten candidate fossil calibrated nodes (root and nodes 1-9) evaluated in this study. Solid black dots indicate nodes with ≥98% posterior probabilities. *Fig. 1a* Chronogram from nuclear DNA. *Fig. 1b* Chronogram from entire mitochondrial dataset; and *Fig. 1c* Chronogram from mitochondrial data with third codon positions removed.

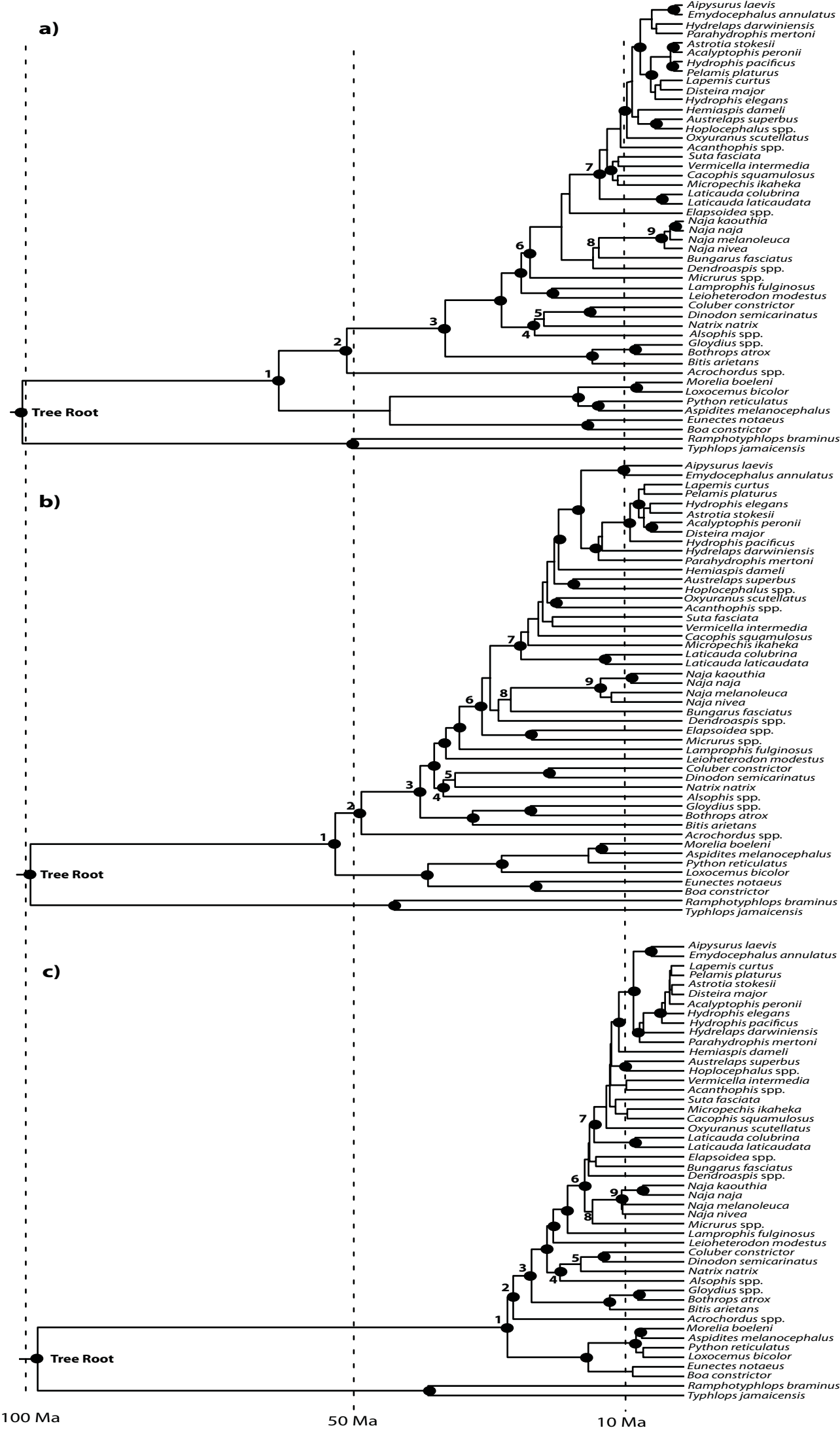
Figure 2: Saturation plots of genetic distances corrected for multiple substitutions versus uncorrected 'p' distances. Corrected genetic distances were calculated using the estimated best-fit models of sequence evolution obtained from AIC criterion in ModelTest. *Fig. 2a*: Saturation plots of the entire mitochondrial DNA dataset (black circles) versus nuclear DNA (gray diamonds). Note the different axis scales for the nuclear and mitochondrial datasets. 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.

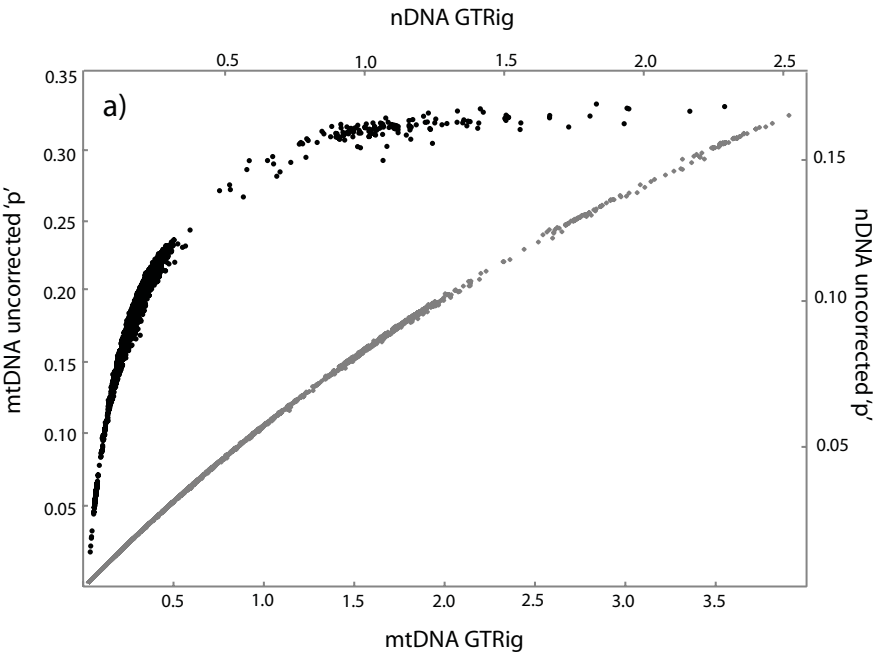
Figure 3: Histogram of the mean differences $\overline{D}_x$ and standard errors (SE) between fossil and estimated molecular ages (Myr) for each of three sets of ten single-fossil calibrated nodes from *a)* nuclear DNA; *b)* mitochondrial DNA; and *c)* mitochondrial DNA with third codon position removed. Fossil ages for eight of the ten candidate nodes were identical for each set, differing only for nodes 2 and 3 (see Fig. 1). Fossil ages used as constraints are given in Table 1. For a single node (*x*) the fossil age at node *x* was used as a single calibration prior. Molecular age estimates were obtained for the nine other candidate nodes for which fossil ages were available.

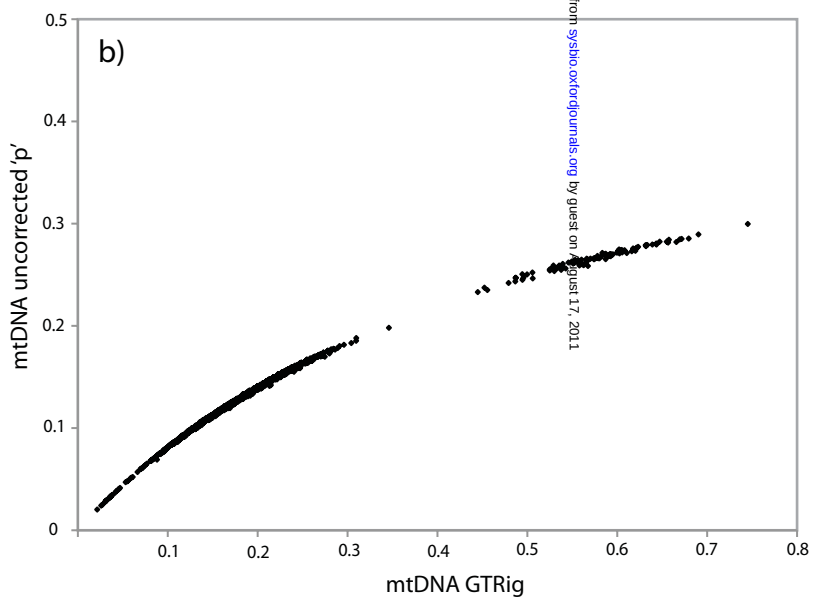
Figure 4: *SS* values for each candidate fossil calibration node when used as the single calibration prior in each of the three calibration sets for *a)* nuclear DNA; *b)* mitochondrial DNA; and *c)* mitochondrial DNA with third codon position removed.

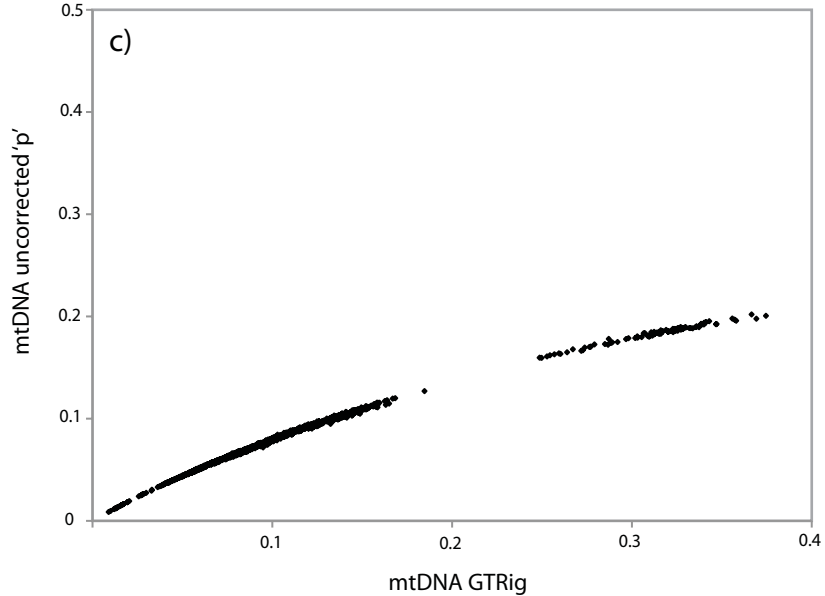
Figure 5: Effect of sequentially removing candidate fossil calibrated nodes on *s*, the average squared deviation of D_i values for the remaining fossil calibrations in each set. *5a:* Nuclear DNA *s* values for three calibration sets. Fossils were removed based on highest to lowest *SS* values calculated from all ten 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. *5b:* MtDNA *s* values for three calibration sets when fossils were removed based on highest to lowest *SS* values calculated for all ten fossil calibrated nodes.

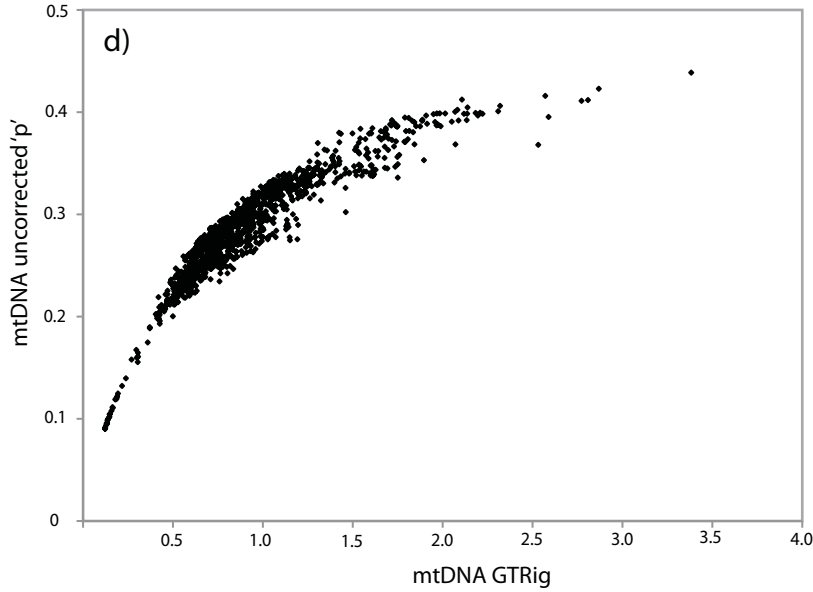
Figure 6: Bayesian multi-fossil calibration analyses showing fossil calibration priors and posterior distributions of molecular age estimates (mean and 95% HPD intervals) at six fossil calibrated nodes using four calibration sets (*A, B, C, D*). Each calibration set comprised four calibration priors 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. Molecular age 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 molecular age 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.

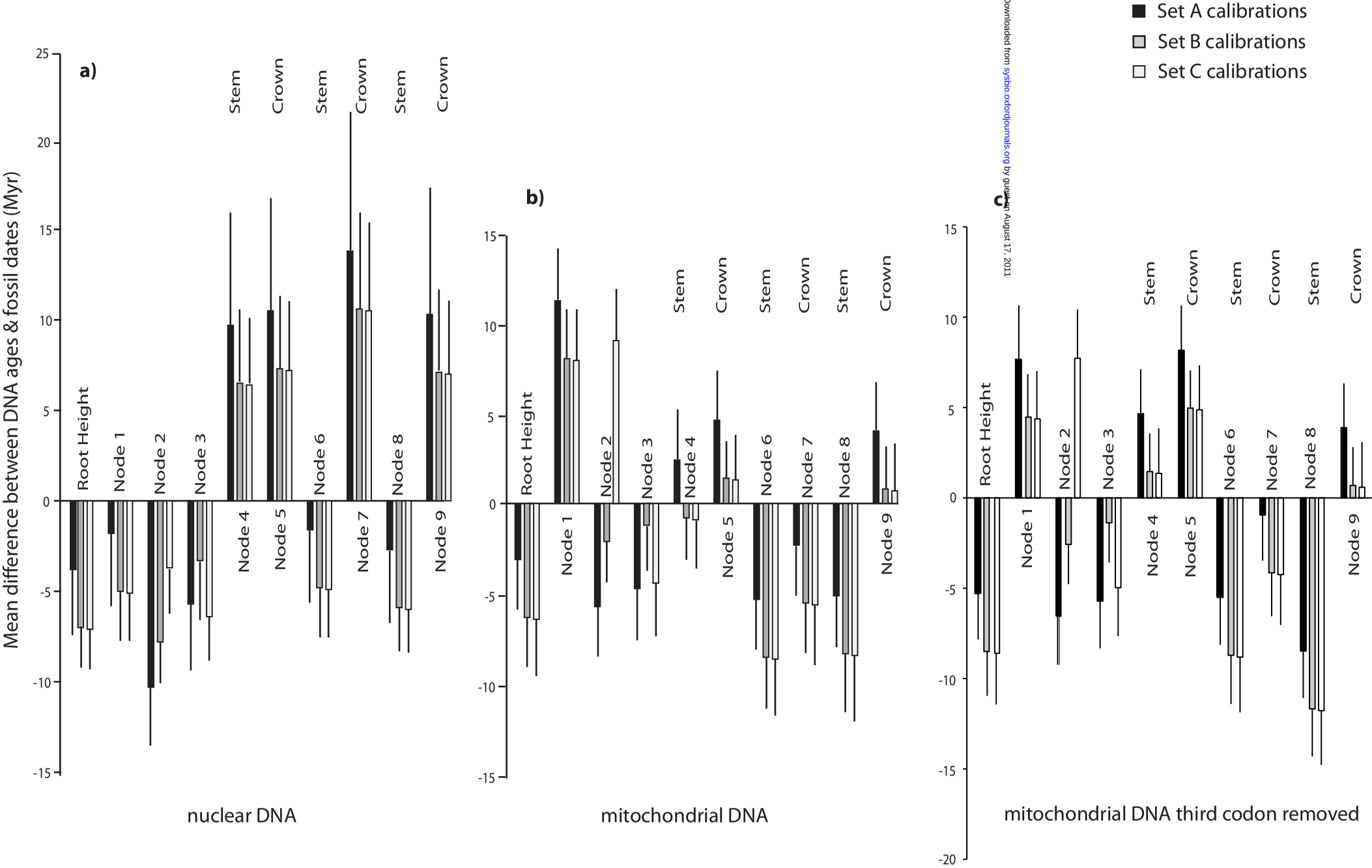

a)
Aipysurus laevis
Emydocephalus annulatus
Hydrelaps darwiniensis
Parahydrophis mertoni
Astrotia stokesii
Acalyptophis peronii
Hydrophis pacificus
Pelamis platurus
Lapemis curtus
Disteira major
Hydrophis elegans
Hemiaspis dameli
Austrelaps superbus
Hoplocephalus spp.
Oxyuranus scutellatus
Acanthophis spp.
Suta fasciata
Vermicella intermedia
Cacophis squamulosus
Micropechis ikaheka
Laticauda colubrina
Laticauda laticaudata
Elapsoidea spp.
Naja kaouthia
Naja naja
Naja melanoleuca
Naja nivea
Bungarus fasciatus
Dendroaspis spp.
Micrurus spp.
Lamprophis fulginosus
Leioheterodon modestus
Coluber constrictor
Dinodon semicarinatus
Natrix natrix
Alsophis spp.
Gloydius spp.
Bothrops atrox
Bitis arietans
Acrochordus spp.
Morelia boeleni
Loxocemus bicolor
Python reticulatus
Aspidites melanocephalus
Eunectes notaeus
Boa constrictor
Ramphotyphlops braminus
Typhlops jamaicensis
Tree Root
1
2
3
4
5
6
7
8
9
b)
Aipysurus laevis
Emydocephalus annulatus
Lapemis curtus
Pelamis platurus
Hydrophis elegans
Astrotia stokesii
Acalyptophis peronii
Disteira major
Hydrophis pacificus
Hydrelaps darwiniensis
Parahydrophis mertoni
Hemiaspis dameli
Austrelaps superbus
Hoplocephalus spp.
Oxyuranus scutellatus
Acanthophis spp.
Suta fasciata
Vermicella intermedia
Cacophis squamulosus
Micropechis ikaheka
Laticauda colubrina
Laticauda laticaudata
Naja kaouthia
Naja naja
Naja melanoleuca
Naja nivea
Bungarus fasciatus
Dendroaspis spp.
Elapsoidea spp.
Micrurus spp.
Lamprophis fulginosus
Leioheterodon modestus
Coluber constrictor
Dinodon semicarinatus
Natrix natrix
Alsophis spp.
Gloydius spp.
Bothrops atrox
Bitis arietans
Acrochordus spp.
Morelia boeleni
Aspidites melanocephalus
Python reticulatus
Loxocemus bicolor
Eunectes notaeus
Boa constrictor
Ramphotyphlops braminus
Typhlops jamaicensis
Tree Root
1
2
3
4
5
6
7
8
9
c)
Aipysurus laevis
Emydocephalus annulatus
Lapemis curtus
Pelamis platurus
Astrotia stokesii
Disteira major
Acalyptophis peronii
Hydrophis elegans
Hydrophis pacificus
Hydrelaps darwiniensis
Parahydrophis mertoni
Hemiaspis dameli
Austrelaps superbus
Hoplocephalus spp.
Vermicella intermedia
Acanthophis spp.
Suta fasciata
Micropechis ikaheka
Cacophis squamulosus
Oxyuranus scutellatus
Laticauda colubrina
Laticauda laticaudata
Elapsoidea spp.
Bungarus fasciatus
Dendroaspis spp.
Naja kaouthia
Naja naja
Naja melanoleuca
Naja nivea
Micrurus spp.
Lamprophis fulginosus
Leioheterodon modestus
Coluber constrictor
Dinodon semicarinatus
Natrix natrix
Alsophis spp.
Gloydius spp.
Bothrops atrox
Bitis arietans
Acrochordus spp.
Morelia boeleni
Aspidites melanocephalus
Python reticulatus
Loxocemus bicolor
Eunectes notaeus
Boa constrictor
Ramphotyphlops braminus
Typhlops jamaicensis
Tree Root
1
2
3
4
5
6
7
8
9
100 Ma
50 Ma
10 Ma

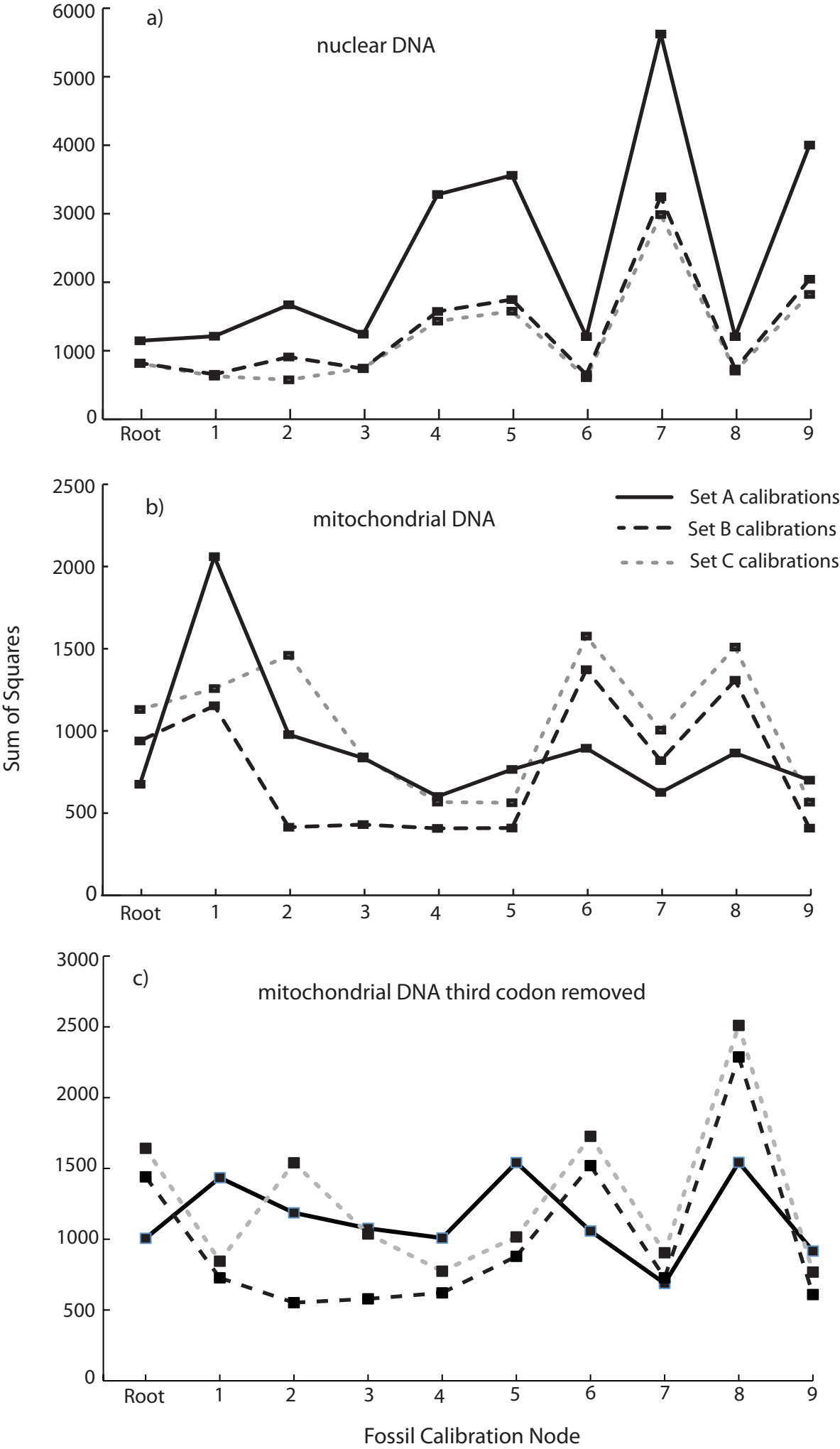


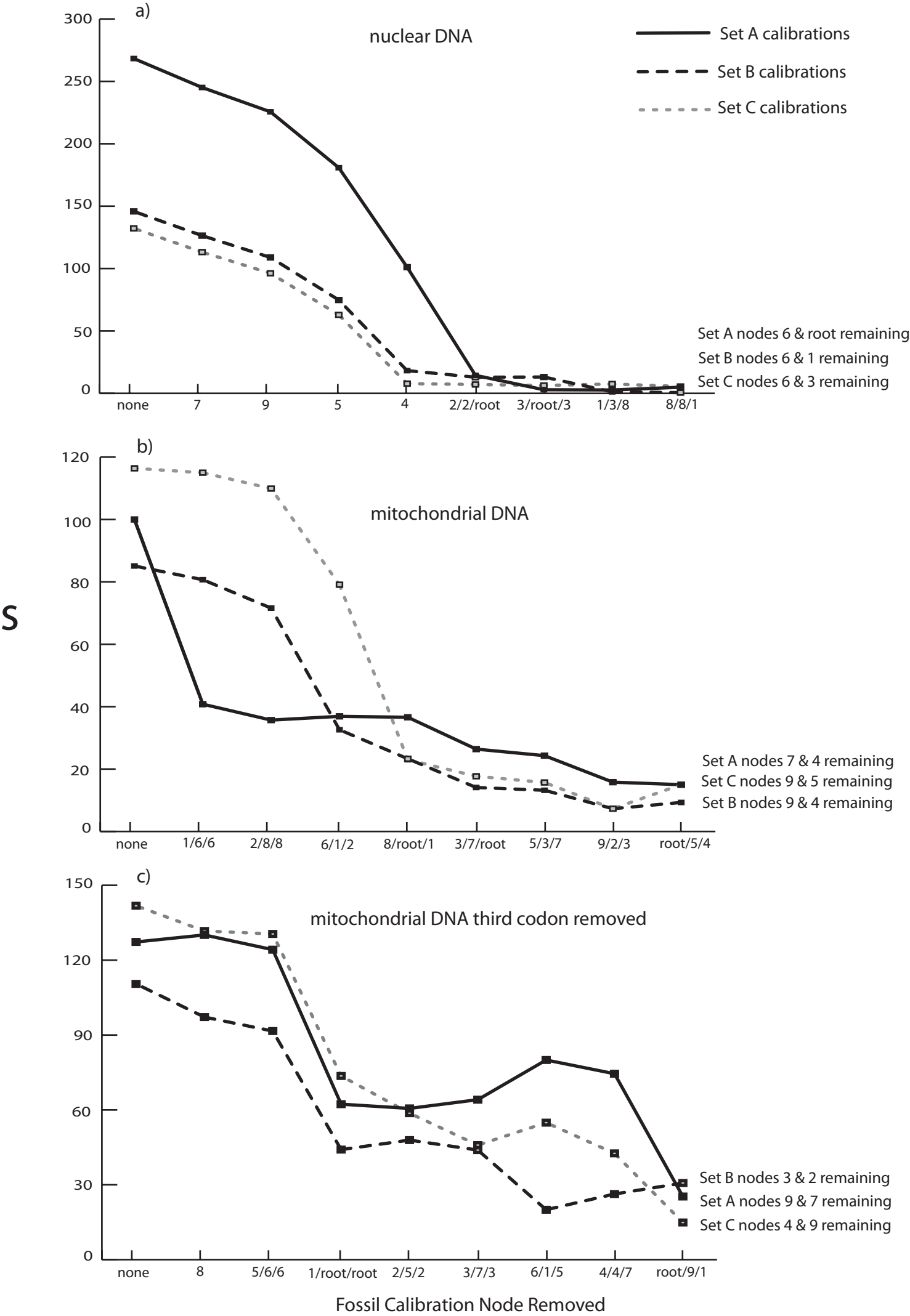









a)
nuclear DNA
b)
mitochondrial DNA
c)
mitochondrial DNA third codon removed
Set A calibrations
Set B calibrations
Set C calibrations
Sum of Squares
Fossil Calibration Node
Root
1
2
3
4
5
6
7
8
9
0
500
1000
1500
2000
2500
3000
4000
5000
6000


a)
nuclear DNA
Set A calibrations
Set B calibrations
Set C calibrations
300
250
200
150
100
50
0
none
7
9
5
4
2/2/root
3/root/3
1/3/8
8/8/1
Set A nodes 6 & root remaining
Set B nodes 6 & 1 remaining
Set C nodes 6 & 3 remaining
b)
mitochondrial DNA
120
100
80
60
40
20
0
none
1/6/6
2/8/8
6/1/2
8/root/1
3/7/root
5/3/7
9/2/3
root/5/4
Set A nodes 7 & 4 remaining
Set C nodes 9 & 5 remaining
Set B nodes 9 & 4 remaining
c)
mitochondrial DNA third codon removed
150
120
90
60
30
0
none
8
5/6/6
1/root/root
2/5/2
3/7/3
6/1/5
4/4/7
root/9/1
Set B nodes 3 & 2 remaining
Set A nodes 9 & 7 remaining
Set C nodes 4 & 9 remaining
Fossil Calibration Node Removed

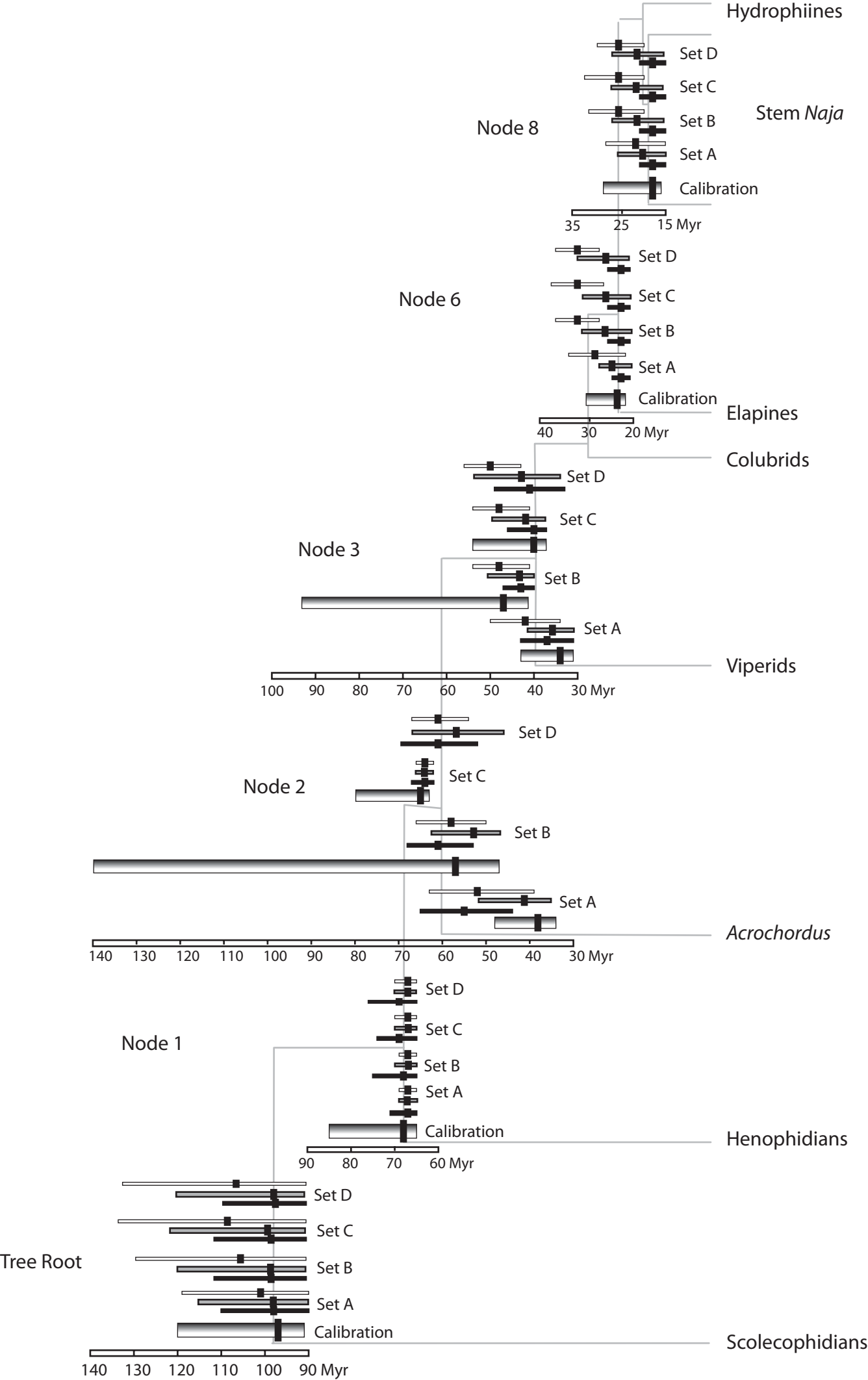

Hydrophiines
Set D
Set C
Set B
Set A
Calibration
Node 8
Stem *Naja*
35
25
15 Myr
Set D
Set C
Set B
Set A
Calibration
Node 6
Elapines
40
30
20 Myr
Colubrids
Set D
Set C
Node 3
Set B
Set A
Viperids
100
90
80
70
60
50
40
30 Myr
Set D
Set C
Node 2
Set B
Set A
Acrochordus
140
130
120
110
100
90
80
70
60
50
40
30 Myr
Set D
Set C
Node 1
Set B
Set A
Calibration
Henophidians
90
80
70
60 Myr
Set D
Set C
Tree Root
Set B
Set A
Calibration
Scolecophidians
140
130
120
110
100
90 Myr

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

Fossil Calibrations	Node	Calibration Priors			Reference
		Mean (95% HPD)	Ln Mean (stdev)	Zero offset	
Scolecophidians vs. 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 Mya) (Gardner and Cifelli, 1999) and six *Coniphis* trunk vertebrate from the Cenomanian (94-100 Mya) in Sudan (Rage and Werner, 1999). Gardner and Cifelli (1999, pg. 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 Mya). This calibration also was used by (Kelly et al., 2009; Sanders and Lee, 2008).
Henophidians vs. 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 Mya based on the earliest probable boid fossils from the latest Cretaceous (65-85 Mya) 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 Mya). 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, pg. 146). Based on these fossils we constrained the Henophidia-Caenophidia split as occurring 68 (65-85) Mya.
Acrochordids vs. colubroids	**2 - *Set A***	38 (34-48)	1.40 (0.75)	34	The MRCA of the Caenophidia (Acrochordidae vs. 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 Mya) 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 Mya) (Rage, 1984, 1987); and the oldest undisputed colubroid fossil from the late-middle Eocene (37-39 Mya) (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) My (Kelly et al., 2009; Sanders and Lee, 2008); 57 (47-140) My (Wuster et al., 2008); and 65 (63-80) My (Noonan and Chippindale, 2006a, b) used > 65 My).
Acrochordids vs. colubroids	**2 - *Set B***	57 (47-140)	2.50 (1.25)	45	
Acrochordids vs. colubroids	**2 - *Set C***	65 (63-80)	1.10 (1.10)	62	
Viperids vs. colubrids + elapids	**3 - *Set A***	34 (31-43)	1.40 (0.70)	30	The MRCA of the Colubroidea (viperids vs. colubrids and elapids) has been dated based on the oldest colubrid fossils from the Late Eocene (34-37 Mya) in Thailand (Rage et al., 1992), however, the oldest putative colubroid fossils from the Cenomanian (93-96 Mya) (Rage and Werner, 1999) have also been used to constrain the upper bound of this clade. Head et al., (2005) (pg.249) and Parmley and Holman, (2003) (pg. 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
Viperids vs. colubrids + elapids	**3 - *Set B***	47 (40-95)	2.00 (1.20)	40	

Viperids vs. colubrids + elapids	**3 - *Set C***	40 (37-60)	1.10 (1.25)	37	(43-60 Mya). We tested two divergence dates previously used: 34 (31-43) My (Kelly et al., 2009; Sanders and Lee, 2008; Wiens et al., 2006); and 47 (40-95) My (Wuster et al., 2008). We also tested a constraint of 40 (37-60) My based on the geographically and taxonomically divergent colubrid fossils from the Late
Natricines vs. colubrines (Stem)	**4**	36 (35-45)	0.50 (1.10)	35	Fossils assigned to *Coluber caduricl* and *Natrix mlynarskii*, extinct species that belong the extant sub-families Colubrinae and Natricinae respectively, have been described from the early Oligocene (30-34 Mya) in Europe (Rage, 1988). A third colubrid, *Texasophis galbreathi*, has been described from the early Orellan to Whitneyan ages of the Oligocene (30-31 Mya) in North American (Holman, 1984) (pg. 225). Based on these fossils the crown natricine-colubrine divergence has been constrained at 35-45 My (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) My.
Natricines vs. colubrines (Crown)	**5**				
Elapines vs. hydrophines	**6**	23 (21-30)	1.00 (0.80)	20	A fossil vertebra from the late Oligocene/early Miocene (20-23 Mya) 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) (pg. 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 My 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 & Lee, 2008, pg. 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 Mya (Kuch et al., 2006). We tested the effects of constraining the crown hydrophiines and crown elapids with dates of 23 (21-30) My.
Crown hydrophiines	**7**				
African vs. Asian *Naja* (Stem)	**8**	19 (17-30)	1.00 (1.00)	16	Fossils of three extinct European *Naja* species with apomorphies that distinguish Asian and African *Naja* occur 16 Mya (Szyndlar and Rage, 1990). These fossils have been used to date the divergence between the crown African and Asian *Naja* (Kelly et al., 2009; Wuster et al., 2007, 2008). 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) My.
African vs. Asian *Naja* (Crown)	**9**				

Table 2. Empirical Scaling Factors (ESF_i) for candidate fossil calibrations for nuclear and mitochondrial datasets calculated using proportional branch lengths obtained from uncalibrated ultrametric trees produced using different methods. 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 maximum likelihood (ML) and Bayesian (MrBayes) phylogenies. Uncalibrated ultrametric trees were obtained by fixing the root to an arbitrary value of 100 and using penalised likelihood 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.

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 #	***ESF***	**Node #**	***ESF***	**Node #**	***ESF***	**Node #**	***ESF***	**Node #**	***ESF***	**Node #**	***ESF***	**Node #**	***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**

Lukoschek et al., *Supplementary Material A.*

Evaluating Fossil Calibrations

Cross-validations and empirical coverage

Nuclear DNA identified the crown placements of the *Naja* and *Laticauda* fossils as outliers using both the cross-validation analyses and empirical scaling factors; however, this was not the case for the mitochondrial data. We accounted for these differences in terms of the effects of mitochondrial saturation, however, it is also important to evaluate these results in terms of the fossils themselves.

The *Naja* fossils comprise three extinct species with characters that distinguish Asian and African *Naja* (Szyndlar and Rage, 1990). However, these fossils also have primitive characters very rare among living cobras (Szyndlar and Rage, 1990, pg 398) suggesting that they belong to the stem *Naja*. When used to constrain the divergence between extant African and Asian *Naja* species (crown *Naja* - node 9) (Kelly et al., 2009; Wuster et al., 2007; Wuster et al., 2008), these fossils overestimated dates for other fossil-calibrated nodes (Fig. 2a) and were identified as outliers by empirical scaling factors (Table 2). However, our alternative placement for dating the *Naja-Bungarus* divergence (node 8), based on previous evidence of the close relationships between *Naja* and *Bungarus* (Slowinski and Keogh, 2000; Wuster et al., 2007), consistently produced much younger divergence dates (Fig. 2) and had intermediate empirical coverage (Table 2). *Naja* is paraphyletic with *Boulengerina* and *Paranaja* and its relationships with other cobra genera are poorly resolved (Slowinski and Keogh, 2000; Wuster et al., 2007); thus, its sister group is difficult to identify as is the most appropriate placement of this fossil on the tree. However, the *Naja* fossils include well-preserved skull elements with well-defined morphological characters (Szyndlar and Rage, 1990); thus, the best nodal placement might be identified from cladistic analysis of extinct and extant taxa (Doyle and Donoghue, 1993). The relative completeness of these fossils also means they could be used in a Bayesian approach that incorporates morphological data from fossils and extant species (Lee et al., 2009) to evaluate the

effects of alternative nodal placements on estimated divergence times.

An elapid fossil from the Australian late Oligocene/early Miocene (20-23 MA) also consistently overestimated dates at other fossil calibrated nodes when used to calibrate the crown hydrophiines (Fig. 2a - node 7) and was identified as an outlier by empirical scaling factors (Table 2). This juvenile vertebra was described as being nested within *Laticauda* rather than within or basal to any other elapid clade (Scanlon et al., 2003); however, there have been calls for taxonomic and/or stratigraphic revisions of this fossil (J. Scanlon pers. comm.), partly in response to younger molecular dates obtained for the divergence between *Laticauda* and the remaining hydrophiines (Sanders and Lee, 2008). Our analyses clearly are unable to resolve either the taxonomic affinities or correct stratigraphy of this fossil. Nonetheless, the alternative placement of this fossil to constrain the crown elapids (node 6) tended to produce much younger estimates of other fossil dates (Fig. 2a), though it had amongst the highest empirical scaling factors (Table 2). Proteroglyphous fangs very similar to those of modern elapids first appear in the fossil record in Germany 20-23 MA (Kuch et al., 2006) suggesting that earlier constraints probably might be more realistic for the crown elapids.

The nuclear DNA cross-validations indicated that constraining the stem and crown natricine-colubrine clades (nodes 4 & 5 respectively) at 36 (35-45) My (Alfaro et al., 2008; Guicking et al., 2006) overestimated dates at other fossil calibrated nodes (Fig. 3a). This result is not surprising given that the natricine-colubrine divergence is much shallower than the crown Colubroidea (node 3) in the nuclear gene tree (Figs. 1a), yet their calibrations overlapped to greater or lesser extents (Table 1). Indeed, the colubrine-natricine constraint was almost identical to the *set A* constraint for the MRCA of all advanced snakes (Table 1 - node 2). This discrepancy is not confined to our study: Alfaro et al., (2008) used 35-55 MA to constrain the tree-root (crown colubroids) and 35-45 MA for the deeply nested natricine-colubrine divergence (see their Fig. 2). These overlapping constraints reflect the fact that the oldest known colubrine fossil, a vertebra

assigned to *Nebraskophis* from the Late Eocene in North America (Parmley and Holman, 2003), occurs simultaneously with the oldest undoubted colubrid fossil from the late Eocene in Thailand (34-37 MA) (Rage et al., 1992), while earliest natricine fossil (*Natrix mlynarskii)* appears in Europe soon after in the early Oligocene (32-34 MA), where it co-occurs with colubrine (*Coluber cadurci)* fossils (Rage, 1988). The nodal age for the crown Colubroidea has been inferred from the oldest undoubted colubrid fossil (see below), while the colubrine-natricine divergence has been inferred from the oldest natricine fossil. However molecular phylogenetic appraisals invariably infer nested positions for the colubrines and natricines among the viperids, elapids, atractaspids, and other colubrid subfamilies that comprise the Colubroidea (Kelly et al., 2009; Lawson et al., 2005; Vidal et al., 2007; Yan et al., 2008; Zaher et al., 2009).

How can these discrepancies be resolved? Firstly, Rage (1988, pg. 467) questioned the taxonomic affinities of the natricine fossil in its description. Apart from this fossil, the next appearance of natricine morphology in the fossil record does not occur until the early Miocene (20-23 MA), when several natricine species appear (Rage and Auge, 1993; Ivanov, 2001). The earliest 'natricine' fossil may, therefore, be so deeply buried in the stem lineage as to be irrelevant for dating the divergence between modern colubrines and natricines. Secondly, molecular phylogenies do not resolve the natricines and colubrines as sister taxa. Instead, basal divergences among extant colubrid clades comprise one or more poorly resolved polytomies (Lawson et al., 2005; Vidal et al., 2007; Kelly et al., 2009) and a sister-group relationship among colubrines and natricines is only recovered if other clades in the polytomy are not sampled (as in our study and also in Alfaro et al., 2008). Thus, even if the Oligocene fossil is a natricine, its appropriate placement is deeper in the tree. Finally, the first appearance of an undoubted colubrid probably does not represent the earliest divergences among the Colubroidea. Molecular phylogenetic hypotheses indicate that other clades (e.g. viperids, homalopsids, pareatids) diverged earlier in the colubroid radiation (Lawson et al., 2005; Vidal et al., 2007; Alfaro et al., 2008; Wuster et al., 2008); however, the earliest known fossils from these clades date to the early

Miocene (Rage and Auge, 1993; Szyndlar and Rage, 1999; Ivanov, 2001).

Multi-calibration Bayesian evaluation of node 2 and 3 calibration sets

The alternative calibration for nodes 2 and 3 were evaluated by comparing the prior and posterior distributions from multi-calibration Bayesian analyses (Fig. 3). The lognormal priors we used model the probability distribution of the actual emergence dates of a clade using the fossil date to inform the hard minimum bound for the youngest possible age, specifying a mean age somewhat older than the fossil, and have a soft maximum bound that allows for the clade to be considerably older than the fossil record (Ho and Phillips, 2009; Yang and Rannala, 2006). It is important to note that the maximum bound of a lognormal prior accounts for a relatively small proportion of the total prior probability and thus has a much smaller effect on the posterior distribution than the log-normal mean or mode, which accounts for a much larger proportion of the prior probability and thus exerts a far stronger constraint (see Fig. 2f in Ho and Phillips 2009). Moreover, the hard minimum bound gives zero probability to the nodal age actually being younger than the oldest fossil known (Ho and Phillips, 2009). As such, posterior distributions that are younger than their lognormal calibration priors strongly suggest that their respective calibrations were too old.

Comparing the *set B* posterior distributions with their respective priors suggests that the *set B* calibrations are too old (Fig. 3). In particular, the estimated maximum 95% HPDs are tens of millions of years younger than their respective calibration priors, with the mean and minimum bound for the crown colubroids (node 3) also younger than its calibration prior (Fig. 3). By contrast, the *set A* posterior distribution for the crown caenophidians (node 2) was much older than its prior (in fact they barely overlapped) indicating that the *set A* calibration was too young (Fig. 3). The young *set A* constraint of 38 (34-48) My for crown caenophidians (node 2) (Kelly et al., 2009; Sanders and Lee, 2008) was based on the first appearance of undisputed colubroids (37-39 My) in the fossil record (Head et al., 2005). However, these same fossils also were used to constrain the minimum bound of the crown colubroids (*set B* - node 3) at 40 My (Wuster et al.,

2008), with an maximum bound of 95 My based on possible colubroid fossils from the Cenomanian (93-96 MA) (Rage and Werner, 1999). They were then further used to extrapolate even older dates of 57 (47-140) MA for the origins of the Caenophidia (*set B* – node 2) (Wuster et al., 2008). Ignoring the contentious taxonomic affinities of the oldest Cenomanian colubroid fossils (Head et al., 2005; Sanders and Lee, 2008), the first primitive colubroids to appear in the fossil record almost certainly belong to the stem colubroids; thus, are inappropriate for dating the crown colubroid radiation (node 3) (Doyle and Donoghue, 1993; Magallon and Sanderson, 2001).

The *set C* estimated mean and minimum HPD for caenophidian origins (node 2) also were younger than its calibration prior of 65 (63-80) MA (Noonan and Chippindale, 2006a; Noonan and Chippindale, 2006b). This constraint was based on the first *Nigerophis* fossil dated at 56-65 My (Rage, 1984). However, the relevance of this fossil for dating caenophidian origins relies on the Nigeropheidae belonging to the Acrochordoidea (McDowell, 1987; Rage, 1987) but morphological similarities of this fossil to several disparate groups renders its taxonomic affinities uncertain (Rage, 1984, pg. 71). The maximum bound of this calibration ignores an indeterminate colubroid fossil dated at 49-56 My (Rage et al., 2003) that suggests the Caenophidia started diverging in the early to mid-Eocene. (Head et al., 2005) suggested that this ‘colubroid’ fossil might be an acrochordid: nonetheless, even if reassigned the fossil remains relevant for dating caenophidian origins (acrochordoid-colubroid divergence - node 2).

Alfaro, M. E., D. R. Karns, H. H. Voris, C. D. Brock, and B. L. Stuart. 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.

Doyle, J. A., and M. J. Donoghue. 1993. Phylogenies and angiosperm diversification.

Paleobiology 19:141-167.

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

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

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

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

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

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

Lee, M. S. Y., P. G. Oliver, and M. N. Hutchison. 2009. Phylogenetic uncertainty and molecular clock calibrations: a case study of legless lizards (Pygopodidae, Gekkota). Mol Phylogenet Evol 50:661-666.

Magallon, S., and M. J. Sanderson. 2001. Absolute diversification rates in angiosperm clades. Evolution 55:1762-1780.

McDowell, J. R. 1987. Systematics. Pages 3-50 *in* Snakes - ecology and evolutionary biology (R. A. Seigel, J. T. Collins, and S. S. Novak, eds.). Macmillan, New York.

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

Noonan, B. P., and P. T. Chippindale. 2006b. Vicariant origin of Malagasy reptiles supports late Cretaceous Antarctic land bridge. The American Naturalist 168:730-741.

Parmley, D., and J. A. Holman. 2003. *Nebraskophis* HOLMAN from the Late Eocene of Georgia (USA), the oldest known North American colubrid snake. Acta Zool Cracov 46:1-8.

Rage, J. C. 1984. Serpentes. Pages 1-79 *in* Encyclopedia of Paleoherpetology Gustav Fischer Verlag, Stuttgart.

Rage, J. C. 1987. Fossil history. Pages 51-76 *in* Snakes - ecology and evolutionary biology (R. A. Seigel, J. T. Collins, and S. S. Novak, eds.). Macmillan, New York.

Rage, J. C., S. Bajpai, J. G. M. Thewissen, and B. N. Tiwari. 2003. Early Eocene snakes from Kutch, Western India, with a review of the Palaeophiidae. Geodiversitas 25: 695-716.

Rage, J. C., E. Buffetaut, H. Buffetaut-Tong, Y. Chaimanee, S. Ducrocq, J. J. Jaeger, and V. Suteethorn. 1992. A colubrid in the late Eocene of Thailand: The oldest known Colubridae (Reptilia, Serpentes). Comptes Rendus de l'Académie des sciences 314:1085-1089.

Rage, J. C., and C. Werner. 1999. Mid-Cretaceous (Cenomanian) snakes from Wadi Abu Hashim, Sudan: The earliest snake assemblage. Palaeontol Afr 35:85-110.

Sanders, K. L., and M. Y. L. Lee. 2008. Molecular evidence for a rapid late-Miocene radiation of Australasian venomous snakes (Elapidae: Colubroidea). Mol Phylogenet Evol 46:1180-1188.

Scanlon, J. D., M. S. Y. Lee, and M. Archer. 2003. Mid-Tertiary elapid snakes (Squamata, Colubroidea) from Riversleigh, northern Australia: early steps in a continent-wide adaptive radiation. Geobios 36:573-601.

Slowinski, J. B., and J. S. Keogh. 2000. Phylogenetic relationships of elapid snakes based on cytochrome *b* mtDNA sequences. Mol Phylogenet Evol 15:157-164.

Szyndlar, Z., and J. C. Rage. 1990. West Palearctic cobras of the genus *Naja* (Serpentes: Elapidae): interrelationships among extinct and extant species. Amphib-Reptilia 11:385-400.

Vidal, N., A. S. Delmas, P. David, C. Cruaud, A. Couloux, and S. B. Hedges. 2007. The

phylogeny and classification of caenophidian snakes inferred from seven nuclear protein-coding genes. C R Biol 330:182-187.

Wuster, W., S. Crookes, I. Ineich, Y. Mane, C. E. Pook, J.-F. Trape, and D. G. Broadley. 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., L. Peppin, C. E. Pook, and D. E. Walker. 2008. A nesting of vipers: phylogeny and historical biogeography of the Viperidae (Squamata: Serpentes). Mol Phylogenet Evol 49:445-459.

Yan, J., H. Li, and K. Zhou. 2008. Evolution of the mitochondrial genome in snakes: gene rearrangements and phylogenetic relationships. BMC Genet 9:569.

Yang, Z., and B. Rannala. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Mol Biol Evol 23:212-226.

Zaher, H., F. G. Grazziotin, J. E. Cadle, R. W. Murphy, J. C. de Moura-Leite, and S. L. Bonatto. 2009. Molecular phylogeny of advanced snakes (Serpentes, Caenophidai) with an emphasis on South American Xenodontines: a revised classification and descriptions of new taxa. Papeis Avulsos de Zoologia 49:115-153.