

Integrating Fossil Preservation Biases in the Selection of Calibrations for Molecular Divergence Time Estimation

ALEX DORNBURG^{1,*}, JEREMY M. BEAULIEU¹, JEFFREY C. OLIVER¹, AND THOMAS J. NEAR^{1,2}

¹Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520, USA; and

²Peabody Museum of Natural History, Yale University, New Haven, CT 06520, USA;

*Correspondence to be sent to: Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520, USA. Email: alex.dornburg@yale.edu.

Received 30 July 2009; reviews returned 26 October 2009; accepted 22 October 2010

Associate Editor: Todd Oakley

Abstract.—The selection of fossil data to use as calibration age priors in molecular divergence time estimates inherently links neontological methods with paleontological theory. However, few neontological studies have taken into account the possibility of a taphonomic bias in the fossil record when developing approaches to fossil calibration selection. The Sppil–Rongis effect may bias the first appearance of a lineage toward the recent causing most objective calibration selection approaches to erroneously exclude appropriate calibrations or to incorporate multiple calibrations that are too young to accurately represent the divergence times of target lineages. Using turtles as a case study, we develop a Bayesian extension to the fossil selection approach developed by Marshall (2008). A simple method for bracketing absolute divergence times on molecular phylogenies using multiple fossil calibrations points. *Am. Nat.* 171:726–742) that takes into account this taphonomic bias. Our method has the advantage of identifying calibrations that may bias age estimates to be too recent while incorporating uncertainty in phylogenetic parameter estimates such as tree topology and branch lengths. Additionally, this method is easily adapted to assess the consistency of potential calibrations to any one calibration in the candidate pool. [Bayesian; calibration; Chelonia; molecular clock; prior; relaxed clock; Sppil–Rongis effect; Taphonomy; turtle.]

A challenge inherent to the estimation of molecular divergence times is the incorporation of external age information to calibrate rates of molecular evolution. These calibrations are most often determined from investigations of the fossil record (e.g., Benton and Donoghue 2007), suggesting that proper protocols for dating molecular phylogenies require an integration of neontological and paleontological data (see Brochu et al. 2004). Most of the early molecular divergence time studies treated fossil calibrations as fixed point estimates of the minimal age of a lineage (e.g. Baldwin and Sanderson 1998; Wikstrom et al. 2001; Near, Bolnick, et al. 2005; Near, Meylan, et al. 2005), and it is common for molecular calibration studies to assume that a fossil calibration can be placed on a phylogeny with complete certainty (see Rutschmann et al. 2007; Lee et al. 2009). Both of these methodological points present potential pitfalls, as substantial phylogenetic and temporal uncertainty often surround any given fossil calibration due to a combination of factors including stratigraphic error (Marshall 1990; Springer 1995; Lee 1999; Rieppel 1999; van Tuinen and Hadly 2004), uncertainty in the phylogenetic relationship of a fossil taxon relative to extant lineages (Doyle and Donoghue 1993; Conroy and van Tuinen 2003; Rutschmann et al. 2007; Lee et al. 2009), and underestimates of the true age of a lineage due to the incompleteness of the fossil record (e.g., Raup 1972; Strauss and Sadler 1989; Marshall 1990).

Because molecular age estimates using relaxed-clock methods are dependent on the quality of the external age information, erroneous age information from the fossil record can potentially impact estimates of divergence times. Not surprisingly, the influence of

calibration priors on molecular divergence time methods is an area of increasing focus, with particular interest in the development of methods that attempt to identify fossil calibrations that are potentially inconsistent when multiple calibrations are available. For example, the cross-validation method introduced by Near, Meylan, et al. (2005), extended by Noonan and Chippindale (2006) and Burbrink and Lawson (2007) attempts to objectively identify the most consistent fossils with which to calibrate a molecular phylogeny. In this approach, each fossil is used as a single calibration, and the sum of the squared differences between the estimated molecular age and the known fossil age (FA) is then calculated for every other potential fossil. The fossils that minimize the variance between the molecular age and the FA are identified as the “best” fossils for calibrating a tree.

A key assumption to these methods is that the sampling of fossils in the candidate pool of calibrations is not temporally biased. This is a problematic assumption, given that the taphonomic bias in the fossil record (Sppil–Rongis effect) increases the probability of fossil preservation toward the present, with large gaps often artificially truncating the distribution of lineages at deeper time scales (Marshall 1990; Holland and Patzkowsky 2002; Lu et al. 2006). Based on this preservation bias, it is possible that the majority of samples in the candidate pool of calibrations may reflect fossils that are far younger than the true origin of a lineage. Thus, potential fossil “outliers” so often identified by these cross-validation approaches may actually be the most informative calibrations to use in subsequent analyses (Marshall 2008; Benton et al. 2009).

Marshall (2008) recently proposed a novel way to assess potential calibrations by taking an inverse approach to the problem of cross-validation. Rather than identifying fossils that are “too old,” this method uses the relative branch lengths of a given uncalibrated and ultrametric tree for calculating the empirical scaling factor (s_i), a metric for finding the fossil leading to the oldest root age of the tree. The empirical scaling factor is calculated as the age of the oldest known fossil of a lineage divided by its node-to-tip distance. The s_i values of every lineage that contain a potential fossil calibration are simultaneously compared, and the lineage with the highest s_i is identified as the best fossil with which to calibrate and bracket credibility intervals for the entire tree.

The method proposed by Marshall (2008) is straightforward, useful, and rooted in paleontological theory. However, it assumes *a priori* that the underlying ultrametric tree and associated branch lengths are known with certainty. Alternate and equally probable topologies may lead to conflicting best fossils to calibrate the phylogeny, a result that would call into question the very notion that one fossil calibration is preferred over many. Variance in branch length estimates could also affect the calculation of the scaling factor. Moreover, the scaling factor assumes that fossils can only provide a minimum estimate of the age of a lineage (e.g., Marshall 1990; Martin 1993; Springer 1995; Muller and Reisz 2005; Benton and Donoghue 2007; Donoghue and Benton 2007); however, it is unclear how molecular age estimates are influenced by the incorporation of probabilistic models that accommodate temporal uncertainty around potential fossil calibrations.

In this study, we focus our analyses on turtle molecular divergence time estimates using the calibrations presented in Near, Meylan, et al. (2005) and Marshall (2008) to provide an extension to the methods developed by Marshall (2008). This method identifies a consistent set of fossil-based calibration age priors using a Bayesian method that simultaneously accounts for the phylogenetic uncertainty in both tree topology and branch length estimates while taking into account the potential for taphonomic bias in the fossil record. By calculating distributions of scaling factors over the credible interval of branch length and topological estimates, this method can be extended to assess the consistency of all potential calibrations to any calibration in a given pool of potential calibrations.

MATERIALS AND METHODS

Molecular and Paleontological Data Sets

We used the same empirical molecular and paleontological data sets as Near, Meylan, et al. (2005) and Marshall (2008). The molecular data were composed of DNA sequences from three genes, including a single mitochondrial gene (*cytb*) and two nuclear genes that included a protein coding gene (*RAG1*) and an intron

(*RM35*). The DNA sequences were sampled from 23 turtle species that span all major extant lineages of turtles (Shaffer et al. 1997; Fujita et al. 2004; Krenz et al. 2005). We aligned the DNA sequences of each gene separately, using MUSCLE v3.7 (Edgar 2004). All genes were concatenated into a final matrix using Phyutility (Smith and Dunn 2008). We assessed the best-fit model of sequence evolution of each gene and potential data partition by comparing the Akaike weights (Akaike 1973) calculated using MrModeltest v2.2 (Nylander 2004) in conjunction with PAUP* ver.4.0b10 (Swofford 2003).

Extending the Empirical Scaling Factor

The methods presented by Marshall (2008) provide an approach to quantify the potential age estimates from a set of proposed fossil calibrations, allowing for the removal of candidate calibrations that may bias estimates to be too young. As a consequence of the large amount of uncertainty present in phylogenetic parameter estimates, the reliance on a fixed topology and a single set of ultrametric branch lengths are potential drawbacks to this method. Although Marshall (2008) states that a limitation to his method is the reliability of the ultrametric tree, we suggest the implications of relying on a single ultrametric tree have not been fully explored. It is necessary to determine if uncertainty in phylogenetic tree topology and branch lengths could influence the calculation of Marshall's (2008) empirical scaling factor, as this could have significant consequences on the calculation of any metric where the denominator is a node-to-tip distance measure.

To investigate the effect of phylogenetic uncertainty on the empirical scaling factor, we employed recent advances in Bayesian-based divergence time estimation methods to infer the posterior distribution of ultrametric trees and relative branch lengths for our data set (Drummond et al. 2006). The marginal posterior distribution of the ultrametric trees was inferred in BEAST 1.4.8 under a model of uncorrelated rates drawn from a lognormal distribution, assigning a Yule (pure-birth) prior to rates of cladogenesis (Drummond and Rambaut 2007). The turtle nucleotide data set was partitioned by gene and codon, with a separate model for the *RM35* intron, as this partitioning scheme received the highest support when comparing Bayes factor values, with differences greater than 10 being considered as strong support (Kass and Raftery 1995; Brandley et al. 2005; Brown and Lemmon 2007). We ran two independent Markov chain Monte Carlo (MCMC) runs of 20 million generations, sampling every 1000 generations. Convergence and adequate sampling of the likelihood surface (effective sample size >200) of the chains was assessed using Tracer v1.4 and the compare command in AWTY (Nylander et al. 2008). Of the 5000, 20,000 MCMC tree samples were discarded as burn-in for each run, ensuring that the posterior distribution of topologies and branch lengths came from the target distribution. Post-burn-in samples from the marginal posterior

distribution were combined using LogCombiner v1.4.8 (Drummond and Rambaut 2007).

The empirical scaling factor (Equation 1, Marshall 2008) for each fossil is given by,

$$s_i = \frac{FA_i}{NTL_i}, \quad (1)$$

where FA of lineage i is divided by its node-to-tip length (NTL) on an uncalibrated ultrametric tree. We divided the branch lengths of each tree by its total height, as Marshall's (2008) calculation of the empirical scaling factor requires ultrametric trees with a total height of 1. For each lineage, the FA was taken from the minimum ages presented in Marshall (2008). Although there is disagreement surrounding the minimum ages of these fossils (Parham and Irmis 2008), using these previously published dates and phylogenetic placements allows for a direct comparison with Marshall (2008). We calculated s_i across the marginal posterior distribution of ultrametric trees to produce an interval of scaling factors for each fossil using code written for R. If phylogenetic uncertainty had little effect on the calculation of s_i , we would expect the width of the interval to be very close to a single point estimate. Conversely, we expect the interval of s_i to widen as the impact of phylogenetic uncertainty increases. We incorporate the above extension to Marshall's (2008) calculation of the empirical scaling factor as an objective function to assess potential fossil calibrations, allowing for the removal of fossils that are biased to be too young. With this approach, the overlap of all s_i intervals was simultaneously compared with the 95% interval of the calibrations containing the highest mean s_i value to identify a consistent calibration set using code written in R (R Development Core Team 2008).

We used Marshall's (2008, Equation 11) FAc with a confidence level, C , of 0.95 to objectively set the calibration age priors where the hypothesized "true" age for each calibration prior was encompassed by the 95% prior probability distribution of the fossil calibration. We employed the same models and Bayesian dating methods outlined above in a series of analyses that sequentially added fossil calibrations based on the mean age of their s_i interval from oldest to youngest until the entire consistent set of fossil calibration priors were included in the divergence time estimate. Although the order of calibration addition will certainly influence the resulting posterior age estimates, we chose this approach to investigate the effect the number of calibrations had on the variance in age estimates. For each calibration, we applied a lognormal distributed prior age probability to the minimum age of the fossil (Table 1). The shape of the lognormal prior was set to ensure that the 95% tail of the distribution encompassed the maximum age estimated using Marshall's (2008, Equation 11) FAc and a confidence level, C , of 0.95. As the performance of this method to calibration is unknown, we repeated all analyses by modifying the 95% tail of the distribution to extend to one quarter, half, double, or quadruple the temporal coverage of the FAc value for each calibration, though we focus our discussion on the FAc values. We also repeated these analyses using exponential prior age distributions to assess the influence the assumed parametric prior distribution has on age estimates for uncalibrated nodes. For all analyses, we compared the posterior estimates of divergence times inferred under these treatments of the prior age calibrations with the expected estimates of the prior in the absence of sequence data. These empty alignments provide a means to assess the influence of sequence data on divergence time estimates (Drummond et al. 2006).

TABLE 1. Fossil taxa and ages of fossils used as calibrations, empirical scaling factor distributions, and fossil age priors

Node for Calibration	Minimum age (Ma)	Fossil taxa	FA _{0.95} (Ma)	Lognormal (mean, SD)	exponential (mean)	Mean s_i	Marginal posterior distribution of s_i
5	110	<i>Santanachelys gaffneyi</i>	—	—	—	622.9	317.4, 1254.0
9	50	"Ocadia" crassa	61.1	1.58, 0.5	3.69	485.6	252.3, 913.1
10	50	<i>Baltemys</i>	61.1	1.58, 0.5	3.69	417.3	210.8, 772.4
8	52	<i>Hadrianus majusculus</i>	63.5	1.62, 0.5	3.84	319.5	195.1, 518.3
13	90	Lindholmemydidae	109.9	2.17, 0.5	6.64	312.0	205.8, 462.7
12	34	<i>Chrysemys antiqua</i>	41.5	1.20, 0.5	2.51	263.2	131.9, 499.9
6	100	<i>Aspideretes maortuensis</i>	122.1	2.28, 0.5	7.38	256.2	136.6, 452.3
7	65	<i>Hoplochelys</i>	79.4	1.84, 0.5	4.8	252.3	162.0, 405.3
4	110	<i>Cearachelys placidoi</i>	134.3	2.37, 0.5	8.1	239.4	110.0, 393.3
11	71	<i>Yaminuechelus gasparinii</i>	86.7	1.93, 0.5	5.24	176.2	102.8, 283.7
3	110	<i>Araripemys barretoii</i>	134.3	2.37, 0.5	8.12	158.2	110.0, 260.5
17	5	<i>Trachemys inflata</i>	—	—	—	116.2	47.7, 243.7
14	18	<i>Pelusios rusingae</i>	—	—	—	111.9	53.4, 221.1
16	11.6	<i>Chelus</i>	—	—	—	59.9	28.9, 116.4
15	15	<i>Chelodina</i> and <i>Elsaya</i>	—	—	—	48.4	25.4, 88.0

Notes: Candidate pool of minimum prior age calibrations and associated fossils used in this study. Node numbers are identical to those used in Near, Meylan, et al. (2005) and Marshall (2008). Prior age calibrations in bold represent fossil selected by the marginal posterior distribution of empirical scaling factors (s_i) used for further inference in this study. The FA_{0.95} represents the soft upper bound of the calibration prior calculated using Marshall's (2008) FAc. All fossil priors were used as minimum age constraints with the prior age bound following the parametric distributions below. SD = standard deviation.

RESULTS

Selecting Fossils Based on Empirical Scaling Factor Intervals

The intervals of s_i overlapped substantially for 10 potential clade age calibrations and prevented the selection of one best calibration fossil using this criterion (Table 1). Although the calibration fossil for Node 5 had the largest scaling factor, the fossil taxon †*Santanachelys gaffneyi* was identified as an outlier due to incorrect phylogenetic placement in a previous studies (Joyce 2007). This fossil was also identified as an outlier by the methods of Near, Meylan, et al. (2005); however, Marshall (2008) identified this fossil as a potential best calibration.

Although identifying calibrations that will bias estimated ages to be too old by this method is not trivial without prior knowledge, scaling factor distributions can easily be compared between other potential calibrations for consistency if there is reason to doubt the placement of any fossil in the tree (Fig. 1). The fossil that calibrates Node 9, which constrains the crown age of geoemydids, was identified as having the second highest scaling factor (Table 1), estimating the mean age of the most recent common ancestor (MRCA) of all turtles to be 485.6 Ma. This fossil was identified as an outlier by Near, Meylan, et al. (2005), and also not chosen as the best calibration by Marshall (2008). The 95% scaling factor interval on this fossil calibration, however, indicated the MRCA of all turtles to be as recent as 252.3 Ma or as old as 913.1 Ma (Table 1). The large width of the intervals of s_i provides evidence of how phylogenetic uncertainty can have a profound influence on the calculation of the s_i as both variance in tree topology and branch lengths were present in the posterior distribution of ultrametric trees (data not shown). This suggests that using point estimates of parameters such as branch lengths and phylogenetic tree topology may be inappropriate for the calculation of empirical scaling

factors and the selection of optimal fossil calibrations. The broad temporal range encompassed by calibration 9 overlapped with the intervals of s_i for 10 of the 15 potential calibrations. We used this as the consistent set of fossil based calibration priors in subsequent analyses (Table 1).

Testudine Divergence Time Estimates

The posterior distribution trees inferred from our analyses of the combined *cytb*, *R35*, and *RAG1* data set demonstrate considerable topological uncertainty in the deeper level relationships among extant turtle lineages. Although the consensus topology (Fig. 2) resulting from our BEAST analyses suggests cryptodires are paraphyletic with respect to pleurodires, the sister taxon relationship of cryptodires, *sans Carettochelys insculpta* and the softshell turtles, and the pleurodires did not have significant Bayesian posterior support (posterior probability < 0.6). This result is not unexpected as previous analyses assumed monophyly of cryptodires and pleurodires and reflected this relationship in rooting the phylogeny (e.g., Near, Meylan, et al. 2005). Subsequent analyses of turtle molecular phylogenies that were rooted using nonturtle tetrapod species as outgroups have either resulted in paraphyly of the cryptodires, low or conflicting support for cryptodire monophyly (Barley et al. 2010), or incorporated topological phylogenetic constraints to enforce cryptodire monophyly (e.g., Krenz et al. 2005; Huggall et al. 2007; Alfaro et al. 2009).

Using the set of consistent fossil calibrations identified by the scaling factor distributions, we estimated the mean crown age of turtles at approximately 249 Ma, with the 95% highest posterior density (HPD) interval spanning 182 Ma through 324 Ma (Node 1; Fig. 2). The mean crown age of pleurodires was approximately 147

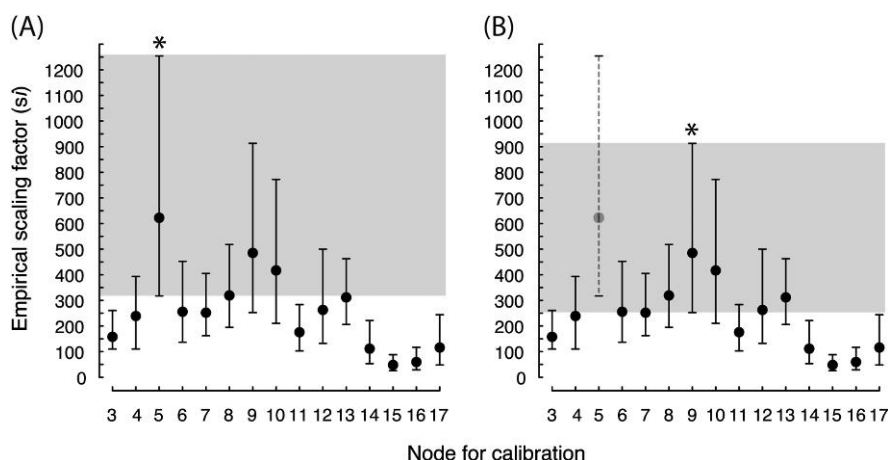


FIGURE 1. Selection of a consistent set of fossils using posterior distributions of empirical scaling factors. The mean (circle) and 95% highest posterior density interval (lines) of the empirical scaling factor (s_i) is shown for each proposed calibration. The fossil calibration with the highest scaling factor is identified with an asterisk. Node numbers are identical to those in Figure 2. Gray shading depicts the overlap of scaling factors with the posterior distribution of the calibration with the highest mean value of s_i . Panel A are the results when the candidate calibration at Node 5 is included. Panel B are the results when the calibration at Node 5 is excluded from the pool of candidate calibrations (see text).

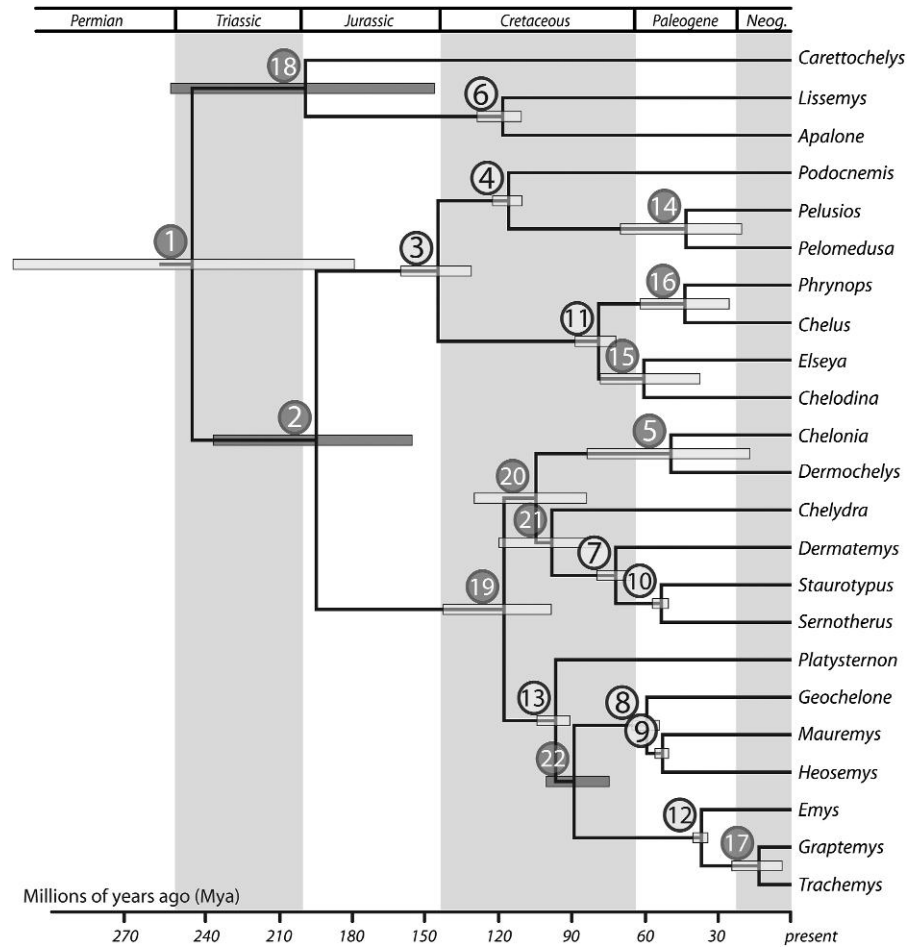


FIGURE 2. Divergence time estimates of turtle lineages. Light 95% HPD interval bars around nodes indicate posterior probabilities greater than 0.95, dark bars indicate posterior probabilities less than 0.95. Light colored circles identify nodes calibrated by fossils identified by the distribution of empirical scaling factors (Table 1). Dark circles identify nodes with no fossil calibration. Node numbers are the same as used in Near, Meylan, et al. (2005) and Marshall (2008).

Ma, with the 95% HPD interval spanning 133 and 162 Ma (Node 3; Fig. 2). Within the pleurodires, the pelomedusids (Node 14; Fig. 2) were estimated to have a mean crown age of approximately 44 Ma, with the 95% HPD interval placing the timing of their divergence sometime between 20 and 71 Ma. The age of the MRCA of Emydidae was estimated at approximately 37 Ma, with the 95% HPD interval spanning 34–41 Ma (Node 12; Fig. 2). The sea turtles were monophyletic, and the estimated divergence between the two sea turtle lineages, Cheloniidae and Dermochelyidae, was between 17 and 85 Ma (Node 5; Fig. 2). The estimated crown age of the Geoemydidae was approximately 51–57 Ma (Node 9; Fig. 2), with this clade sharing an MRCA with Testudinidae approximately 55–66 Ma (Node 8; Fig. 2). The 95% HPD intervals and corresponding mean age estimates are presented in Table 2 for all nodes in the turtle phylogeny.

Our sequential fossil additions also suggest that incorporating just a few calibration age priors can dramatically influence the precision of divergence time estimates globally. In the case of the 95% HPD interval

for the age of the MRCA of all turtles, the addition of the first three calibrations led to a reduction of consistent age estimates of almost 250 myr regardless of how the distribution of prior ages was modeled. The asymptotic decline of the 95% HPD intervals (Fig. 3, see also Appendix 1) illustrates a pattern of diminishing returns, where increasing the number of calibration age priors past a certain threshold has a decreasing influence on the increased precision of the posterior age estimates.

DISCUSSION

Our study provides a novel approach to calibration selection that integrates both phylogenetic uncertainty and the taphonomic bias of the fossil record. We demonstrate that incorporating phylogenetic uncertainty into Marshall's (2008) method of fossil calibration selection leads to ambiguity in the effort to choose a single best fossil calibration. Rather than selecting a single best fossil, our extension of this method selects a consistent set

TABLE 2. Divergence time estimates of turtles

Node	Calibrated?	Lognormal priors		Exponential priors	
		Mean (Myr)	95% HPD	Mean (Myr)	95% HPD
1	No	249.8	182.2–324.5	233.9	169.2–314.4
2	No	198.0	158.1–240.9	191.2	155.6–231.9
3	Yes	147.3	133.5–162.8	143.8	129.0–160.4
4	Yes	117.9	112.3–124.6	113.5	110.0–120.6
5	No	50.2	17.3–85.2	50.2	20.5–83.9
6	Yes	120.3	112.6–131.0	106.0	100.0–118.4
7	Yes	73.2	67.1–81.1	71.7	65.0–82.4
8	Yes	60.3	55.0–66.4	59.0	52.7–66.1
9	Yes	53.7	51.2–56.9	51.7	50.0–54.9
10	Yes	54.2	51.3–57.9	51.9	50.0–55.6
11	Yes	80.4	73.2–90.2	78.5	71.0–91.6
12	Yes	37.5	34.8–41.0	36.2	34.0–40.4
13	Yes	98.3	92.3–106.1	94.5	90.0–103.1
14	No	44.0	20.7–71.3	43.2	21.4–67.5
15	No	61.5	38.1–79.7	60.8	36.8–80.0
16	No	44.4	25.9–63.0	43.0	25.4–61.6
17	No	13.5	3.85–24.9	13.1	4.55–23.8
18	No	202.6	148.8–258.7	180.3	125.0–237.9
19	No	119.8	100.1–145.2	116.4	97.1–142.7
20	No	106.5	85.4–132.2	103.7	81.4–129.3
21	No	99.9	79.0–121.9	97.6	76.3–120.8
22	No	90.6	76.0–102.3	87.6	74.0–99.9

Notes: Node ages estimated using all 10 fossil calibrations with overlapping scaling factor distributions and the parametric distributions on calibration age priors from Table 1. Node numbers correspond with the nodes in Figure 1 and the numbering from Near, Meylan, et al. (2005). Nodes marked “yes” represent the presence of fossil calibrations. Mean ages and 95% HPD values were estimated using both lognormal and exponential calibration prior densities. All ages are in units of millions of years (Myr).

of calibrations that take the nonrandom process of fossilization into account. Our approach has the advantage of not being limited to identifying calibration sets that are too old as consistent calibration sets can be identified by comparing the overlap of scaling factors between any reference fossil with the entire candidate pool of calibrations.

Selecting Fossils Based on Empirical Scaling Factor Intervals

Even if the true phylogeny of a clade were known, calibrating a molecular phylogeny would not be trivial as a major challenge shared by both studies of molecular divergence times and paleontology is inferring the origination and duration of a lineage. The fossil record is far from complete (see Kalmar and Currie 2010) and the Spill–Rongis effect posits the origination of a lineage in the fossil record to occur far later than the true origin of the lineage (e.g., Marshall 1990; Sepkoski 1998; Kirchner and Weil 2000). This bias favoring preservation toward the recent is complicated by a growing body of literature that suggests preservation potential to be highly heterogeneous over time (Alroy et al. 2001; Holland and Patzkowsky 2002; Peters and Foote 2002; Lu et al. 2006). Moreover, this nonuniform preservation process often produces a false signal of clustered lineage originations (Holland 1995; Kidwell and Holland 2002; Peters 2006) that could greatly mislead divergence time studies. For example, recent simulations by Lu et al. (2006) suggested that some species may have gaps in the fossil record extending more than half the true duration of a species, a result supported empirically by Foote (2003). Because the use of fossil calibrations

in molecular divergence time analyses requires age constraints that reflect the true origin of a given lineage, calibrating molecular phylogenies is not as simple as finding “good” and “bad” calibrations, as the natural processes of diversification and fossilization cause potential fossil calibrations for node ages to lay on a spectrum of accuracy (Sanders and Lee 2007).

Marshall’s (2008) study represents an important step toward integrating neontological and paleontological approaches to critically assess the fossil information used to calibrate molecular phylogenies. By taking potential preservation biases into account, Marshall (2008) suggests that lineages that appear too old may not always be outliers. However, this method relies on a single best fossil, while we demonstrate that even for a data set with a well-studied fossil record, using neontological data to find a best fossil can be ambiguous. This is a consequence of the s_i calculation being highly sensitive to variance in phylogenetic parameter estimates present in the ultrametric tree. Our extension of Marshall’s (2008) method incorporates variance in phylogenetic parameter estimates to make it a very useful and computationally efficient method of assessing the accuracy of specific fossil inferred ages contained in a candidate pool of calibrations relative to the age of the oldest fossil. Rather than repeating multiple analyses with individual fossils (e.g., Near, Meylan, et al. 2005), our approach provides a means of assessing the distribution of empirical scaling factors based on the posterior distribution of phylogenetic tree topologies and branch length estimates. A caveat to this approach is that it still requires an assessment of the quality of the fossils selected as tree topological support and relative molecular branch length estimates are affected by

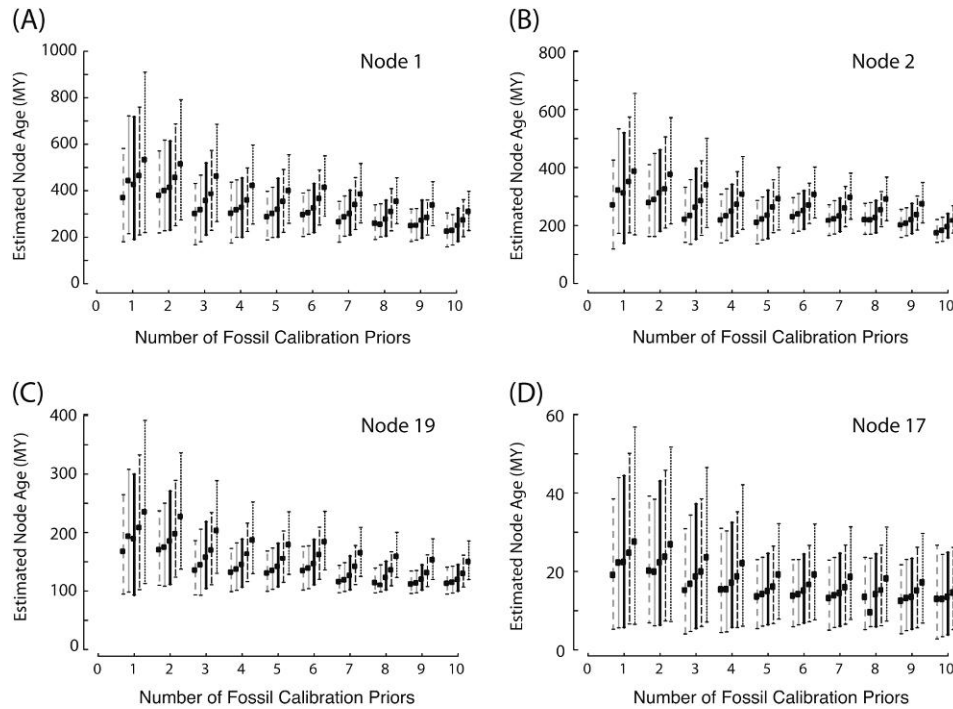


FIGURE 3. The effect of the width of the credible 95% prior age interval on molecular divergence time estimates as plotted by the molecular age estimate versus the number of lognormal distributed calibration age priors used in the analysis. Vertical lines represent the credibility interval of molecular age estimates. Filled points represent the mean of the marginal posterior distribution. Solid black lines represent the posterior distribution of credible age estimates using Marshall's (2008) FA₉₅. Other lines represent treatments of the clade age prior: dotted grey lines when the FA₉₅ was divided by 4, solid grey lines the FA₉₅ was divided by 2, thin solid black lines the FA₉₅ was multiplied by 2, and for dotted black lines the FA₉₅ was multiplied by 4. Node numbers are as given in Figure 1. (A) Node 1, MRCA of all extant turtles. (B) Node 2. (C) Node 19. (D) Node 17.

clade-specific molecular rate heterogeneity and substitution model selection (e.g., Posada and Crandall 2001; Abdo et al. 2005; Alfaro and Huelsenbeck 2006; Dornburg et al. 2008; Venditti et al. 2008; Brown et al. 2009; Marshall 2010), which could bias the results of this method.

Reanalysis of Testudine Divergence Times

Although we incorporate the same molecular data as Marshall (2008) and Near, Meylan, et al. (2005), our posterior age estimates provide evidence for an alternate timing of testudine diversification, highlighting the importance of calibration age priors in studies of molecular divergence times. Although Marshall (2008) suggested between one and four cladogenic events to have occurred among sampled taxa during the Cenozoic, we instead find evidence for as many as nine. This is similar to the results of Near, Meylan, et al. (2005), though our divergence time estimates often differ appreciably from Near, Meylan, et al. (2005). For example, our inferences of testudine divergence times imply 19 of the 22 nodes to be older than originally estimated by Near, Meylan, et al. (2005), with 11 estimates excluded from our posterior distribution of clade ages (Table 2). These differences in age estimates are most likely the consequence of four factors: 1) our use of an

alternate fossil set; 2) our use of soft upper bounds on the distribution of prior ages in contrast to fixed age calibrations; 3) allowing for phylogenetic uncertainty in the underlying tree topology and branch lengths; and 4) our incorporation of a relaxed clock model that does not assume rate autocorrelation a priori.

We find many of the divergence dates inferred by Marshall (2008) to be far older than those we infer. For example, Marshall dates the split of *Chelodina* and *Elseya* (Node 14; Fig. 1) to have occurred during the early Cretaceous (~130 Ma) or the Jurassic–Cretaceous boundary 145 Ma. This is in sharp contrast to our inference of a divergence between the early Miocene (~20 Ma) and the Late Cretaceous (~71 Ma). A variety of factors underlie these conflicting age estimates. For example, although we use the same approach to model the distribution of prior ages, our departure from Marshall's (2008) analysis may be partially caused by his use of a single calibration. Our results indicate that the temporal uncertainty associated with the use of a single fossil to calibrate the molecular rate of evolution for a data set with a demonstrated departure from the molecular clock (Near, Meylan, et al. 2005) may cause the credible distribution of age estimates to grow unrealistically large. When coupled with uncertainty in the phylogenetic tree topology and branch lengths estimates, our use of a single calibration allowed the distribution of

turtle divergence times to extend beyond the Cambrian (>700 Ma). Although these age estimates become more reasonable as additional calibrations are added, the conflicting age estimates inferred between our study and Marshall (2008) and Near, Meylan, et al. (2005) highlight that the timing of testudine divergences are far from resolved.

The relative paucity of the fossil record, combined with the potential for the Spill–Rongis effect to bias the true origination of a lineage toward the present (e.g., Marshall 1990; Sepkoski 1998; Kirchner and Weil 2000) presents challenges in assigning a credible prior age distribution to a fossil calibration. Marshall's (2008) method for absolute age bracketing has an advantage over other methods such as the bounding of a crown lineage by the first appearance of a stem taxon (Behrensmeyer et al. 2000; Drummond et al. 2006) or modeling the expected age of a lineage based on a birth–death process (see Tavaré et al. 2002) because it can be modified to accommodate a nonrandom process of fossilization. Recent advances in the available methods for molecular divergence time estimation provide an unprecedented opportunity to incorporate many aspects of the information present in the fossil record to develop informed calibration age priors that will ultimately lead to more reliable molecular age estimates. Given the tight association of the distribution of prior ages with the resulting estimates of divergence times (e.g., Yang and Rannala 2006; Rannala and Yang 2007; Inoue et al. 2010), this will clearly become an area of increased study that will undoubtedly require a collaborative effort by both neontologists and paleontologists as evolutionary biology moves toward the goal of a time-calibrated Tree of Life.

SUPPLEMENTARY MATERIAL

Supplementary material can be found at <http://www.sysbio.oxfordjournals.org/>.

FUNDING

This work was supported by the National Science Foundation [DEB-0716155, ANT-0839007].

ACKNOWLEDGMENTS

Matthew C. Brandley, Chad D. Brock, Stephen Smith, Todd Oakley, and Jack Sullivan provided valuable comments on earlier versions of this manuscript. Michael J. Donoghue, Charles R. Marshall, Peter A. Meylan, and H. Bradley Shaffer provided important help and discussion regarding our work on turtle divergence times.

REFERENCES

- Abdo Z., Minin V.N., Joyce P., Sullivan J. 2005. Accounting for uncertainty in the tree topology has little effect on the decision-theoretic approach to model selection in phylogeny estimation. *Mol. Biol. Evol.* 22:691–703.
- Akaike H. 1973. Information theory as an extension of the maximum likelihood principle. In: Petrov B.N., Csaki F. editors. Second Annual Symposium on information theory. Budapest (Hungary): Akademi Kiado. p. 267–281.
- Alfaro M.E., Huelsenbeck J.P. 2006. Comparative performance of Bayesian and AIC-based measures of phylogenetic model uncertainty. *Syst. Biol.* 55:89–96.
- Alfaro M.E., Santini F., Brock C., Alamillo H., Dornburg A., Rabosky D.L., Carnevale G., Harmon L.J. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proc. Nat. Acad. Sci.* 106:13410–13414.
- Alroy J., Marshall C.R., Bambach R.K., Beszko K., Foote M., Fürsich F.T., Hansen T.A., Holland S.M., Ivany L.C., Jablonski D., Jacobs D.K., Jones D.C., Kosnik M.A., Lidgard S., Low S., Miller A.I., Novack-Gottshall P.M., Olszewski T.D., Patzkowsky M.E., Raup D.M., Roy K., Sepkoski J.J. Jr., Sommer M.G., Wagner P.J., Webber A. 2001. Effects of sampling standardization on estimates of Phanerozoic marine diversification. *Proc. Nat. Acad. Sci. U.S.A.* 98:6261–6266.
- Baldwin B.G., Sanderson M.J. 1998. Age and rate of diversification of the Hawaiian silversword alliance (Compositae). *Proc. Natl. Acad. Sci. U.S.A.* 95:9402–9406.
- Barley A., Spinks P.Q., Thomson R.C., Shaffer H.B. 2010. Fourteen nuclear genes provide phylogenetic resolution for difficult nodes in the turtle tree of life. *Mol. Phylog. Evol.* 55:1189–1194.
- Behrensmeyer A.K., Kidwell S.M., Gastaldo R.A. 2000. Taphonomy and paleobiology. *Paleobiology* 26(suppl.): 103–147.
- Benton M.J., Donoghue P.C.J. 2007. Paleontological evidence to date the tree of life. *Mol. Biol. Evol.* 24:26–53.
- Benton M.J., Donoghue P.C.J., Asher R.J. 2009. Calibrating and constraining molecular clocks. In: S. B. Hedges, and S. Kumar, editors. *The timetree of life*. Oxford: Oxford University Press. p. 35–86.
- Brandley M., Schmitz A., Reeder T.W. 2005. Partitioned bayesian analyses, partition choice, and phylogenetic relationships of scincid lizards. *Syst. Biol.* 54:373–390.
- Brochu C.A., Sumrall C.D., Theodor J.M. 2004. When clocks (and communities) collide: estimating divergence time from molecules and the fossil record. *J. Paleont.* 78:1–6.
- Brown J.M., Lemmon A.R. 2007. The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. *Syst. Biol.* 56:643–655.
- Brown J.M., Hedtke S.H., Lemmon A.R., Lemmon E.M. 2009. When trees grow too long: investigating the causes of highly inaccurate Bayesian branch-length estimates. *Syst. Biol.* 59:145–161.
- Burbrink F.T., Lawson R. 2007. How and when did old world rat-snakes disperse into the New World? *Mol. Phylogenet. Evol.* 43:173–189.
- Conroy C.J., van Tuinen M. 2003. Extracting time from phylogenies: positive interplay between fossil and genetic data. *J. Mammol.* 84:444–455.
- Donoghue P.C.J., Benton M.J. 2007. Rocks and clocks: calibrating the Tree of Life using fossils and molecules. *Trends Ecol. Evol.* 22:424–431.
- Dornburg A., Santini F., Alfaro M.E. 2008. The influence of model averaging on clade posteriors: an example using the triggerfishes (family Balistidae). *Syst. Biol.* 57:905–919.
- Doyle J.A., Donoghue M.J. 1993. Phylogenies and Angiosperm diversification. *Paleobiology*. 19:141–167.
- Drummond A.J., Ho S.Y.W., Phillips M.J., Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* 4:e88.
- Drummond A.J., Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214.
- Edgar R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.
- Foote M. 2003. Origination and extinction through the Phanerozoic: a new approach. *J. Geol.* 111:125–148.
- Fujita M.K., Engstrom T.N., Starkey D.E., Shaffer H.B. 2004. Turtle phylogeny: insights from a novel nuclear intron. *Mol. Phylogenet. Evol.* 31:1031–1040.
- Holland S.M. 1995. The stratigraphic distribution of fossils. *Paleobiology*. 21:92–109.
- Holland S.M., Patzkowsky S.E. 2002. Stratigraphic variation in the timing of first and last occurrences. *Palaios*. 17:134–146.

- Huggall A.F., Foster R., Lee M.S. 2007. Calibration choice, rate smoothing, and the pattern of tetrapod diversification according to the long nuclear gene RAG-1. *Syst. Biol.* 56:543–563.
- Inoue J., Donoghue P.C.J., Yang Z. 2010. The impact of the representation of fossil calibrations on Bayesian estimation of species divergence times. *Syst. Biol.* 59:74–89.
- Joyce W.G. 2007. Phylogenetic relationships of Mesozoic turtles. *Bull. Peabody Mus. Nat. Hist.* 48:3–102.
- Kalmar A., Currie D.J. 2010. The completeness of the fossil record and its impact on patterns of diversification. *Paleobiology.* 36:51–60.
- Kass, R.E., Raftery A.E. 1995. Bayes factors. *J. Am. Stat. Assoc.* 90:773–795.
- Kidwell S.M., Holland S.M. 2002. The quality of the fossil record: implications for evolutionary analysis. *Annu. Rev. Ecol. Syst.* 33:561–588.
- Kirchner J.W., Weil A. 2000. Delayed biological recovery from extinctions throughout the fossil record. *Nature.* 404:177–179.
- Krenz J.B., Naylor G.J.P., Shaffer H.P., Janzen F.J. 2005. Molecular phylogenetics and evolution of turtles. *Mol. Phylog. Evol.* 37:178–191.
- Lee M.S.Y. 1999. Molecular clock calibrations and metazoan divergence dates. *J. Mol. Evol.* 49:385–391.
- Lee M.S.Y., Oliver P.M., Hutchinson M.N. 2009. Phylogenetic uncertainty and molecular clock calibrations: a case study of legless lizards (Pygopodidae, Gekkota). *Mol. Phylogenet. Evol.* 50:661–666.
- Lu P.J., Yogo M., Marshall C.R. 2006. Phanerozoic marine biodiversity dynamics in light of the incompleteness of the fossil record. *Proc. Nat. Acad. Sci. U.S.A.* 103:2736–2739.
- Marshall C.R. 1990. Confidence intervals on stratigraphic ranges. *Paleobiology.* 16:1–10.
- Marshall C.R. 2008. A simple method for bracketing absolute divergence times on molecular phylogenies using multiple fossil calibrations points. *Am. Nat.* 171:726–742.
- Marshall D.C. 2010. Cryptic failure of partitioned Bayesian phylogenetic analyses: lost in the land of long trees. *Syst. Biol.* 59:108–117.
- Martin, R.D. 1993. Primate origins: plugging the gaps. *Nature* 363:223–234.
- Muller J., Reisz R.R. 2005. Four well-constrained calibration points from the vertebrate fossil record for molecular clock estimates. *Bioessays.* 27:1069–75.
- Near T.J., Meylan P.A., Shaffer H.B. 2005. Assessing concordance of fossil calibration points in molecular clock studies: an example using turtles. *Am. Nat.* 165:137–146.
- Near T.J., Bolnick D.I., Wainwright P.C. 2005. Fossil calibrations and molecular divergence time estimates in centrarchid fishes (Teleostei: Centrarchidae). *Evolution.* 59:1768–1782.
- Noonan B.P., Chippindale P.T. 2006. Vicariant origin of Malagasy reptiles supports late cretaceous Antarctic land bridge. *Am. Nat.* 168:730–741.
- Nylander J.A.A. 2004. MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University, Sweden.
- Nylander J.A.A., Wilgenbusch J.C., Warren D.L., Swofford D.L. 2008. AWTY (Are We There Yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. *Bioinformatics.* 24(4):581–583.
- Parham J.F., Irmis R.B. 2008. Caveats on the use of fossil calibrations for molecular dating: a comment on Near et al. *Am. Nat.* 171:132–136.
- Peters S.E. 2006. Genus extinction, origination, and the durations of sedimentary hiatuses. *Paleobiology.* 32:387–407.
- Peters S.E., Foote M. 2002. Determinants of extinction in the fossil record. *Nature.* 416:420–424.
- Posada D., Crandall K.A. 2001. Selecting the best-fit model of nucleotide substitution. *Syst. Biol.* 50:580–601.
- Rannala B., Yang Z. 2007. Bayesian estimation of species divergence times from multiple loci using multiple calibrations. *Syst. Biol.* 56:453–466.
- Raup D. 1972. Taxonomic diversity during the Phanerozoic. *Science.* 177:1065–1071.
- Rieppel O. 1999. Phylogeny and paleobiogeography of Triassic Sauropterygia: problems solved and unresolved. *Paleoecol. Paleoclim. Paleoecol.* 153:1–15.
- Rutschmann F., Eriksson T., Abu Salim K., Conti E. 2007. Assessing calibration uncertainty in molecular dating: the assignment of fossils to alternative calibration points. *Syst. Biol.* 56:591–608.
- Sanders K.L., Lee M.S.Y. 2007. Evaluating molecular clock calibrations using Bayesian analyses with soft and hard bounds. *Biol. Lett.* 3:275–279.
- Sepkoski J.J. Jr. 1998. Rates of speciation in the fossil record. *Phil. Trans. R. Soc. Lond. B.* 353:315–326.
- Shaffer H.B., Meylan P., McKnight Mark L. 1997. Tests of turtle phylogeny: molecular, morphological, and paleontological approaches. *Syst. Biol.* 46:235–268.
- Smith S., Dunn C. 2008. Phyutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics.* 24:715–716.
- Springer M.S. 1995. Molecular clocks and the incompleteness of the fossil record. *J. Mol. Evol.* 41:531–538.
- Sterli J., de la Fuente M.S., Rougier G.W. 2007. Anatomy and relationships of *Palaeochersis talampayensis*, a Late Triassic turtle from Argentina. *Palaeontogr. Abt. A.* 281:1–61.
- Sterli J., Joyce W.G. 2007. The cranial anatomy of the Early Jurassic turtle *Kayentachelys aprix*. *Acta Palaeontol. Pol.* 52:675–694.
- Strauss D., Sadler P.M. 1989. Classical confidence intervals and Bayesian probability estimates for the ends of local taxon ranges. *Math. Geol.* 21:411–427.
- Swofford D.L. 2003. PAUP* 4.0b: phylogenetic analysis using parsimony (*and other methods). version 4.0. Sinauer Associates, Sunderland, MA.
- Tavaré S., Marshall C.R., Will O., Soligo C., Martin R.D. 2002. Using the fossil record to estimate the age of the last common ancestor of extant primates. *Nature.* 416:726–729.
- Thorne J.L., Kishino H., Painter I.S. 1998. Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol. Evol.* 15:1647–1657.
- van Tuinen M., Hadly E. 2004. Calibration and error in placental molecular clocks: a conservative approach using the Cetartiodactyl fossil record. *J. Hered.* 95:200–208.
- Venditti C., Meade A., Pagel M. 2008. Phylogenetic mixture models can reduce node-density artifacts. *Syst. Biol.* 57:286–293.
- Wikstrom N., Savolainen V., Chase W. 2001. Evolution of the angiosperms: calibrating the family tree. *Proc. R. Soc. B.* 268:2211–2220.
- Yang Z., Rannala B. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol. Biol. Evol.* 23:212–226.