Syst. Biol. 59(1):74–89, 2010

DOI:10.1093/sysbio/syp078
Advance Access publication on November 25, 2009

The Impact of the Representation of Fossil Calibrations on Bayesian Estimation of Species Divergence Times

JUN INOUE[1], PHILIP C. J. DONOGHUE[2], AND ZIHENG YANG[1,*]

[1]*Department of Biology, Galton Laboratory, University College London, Darwin Building, Gower Street, London WC1E 6BT, UK; and*
[2]*Department of Earth Sciences, University of Bristol, Bristol BS8 1RJ, UK;*
[*]*Correspondence to be sent to: Department of Biology, Galton Laboratory, University College London, Darwin Building, Gower Street, London WC1E 6BT, UK; E-mail: z.yang@ucl.ac.uk.*

Received 14 January 2009; reviews returned 30 March 2009; accepted 9 October 2009
Associate Editor: Susanne S. Renner

Abstract.—Bayesian inference provides a powerful framework for integrating different sources of information (in particular, molecules and fossils) to derive estimates of species divergence times. Indeed, it is currently the only framework that can adequately account for uncertainties in fossil calibrations. We use 2 Bayesian Markov chain Monte Carlo programs, MULTIDIVTIME and MCMCTREE, to analyze 3 empirical data sets to estimate divergence times in amphibians, actinopterygians, and felids. We evaluate the impact of various factors, including the priors on rates and times, fossil calibrations, substitution model, the violation of the molecular clock and the rate-drift model, and the exact and approximate likelihood calculation. Assuming the molecular clock caused seriously biased time estimates when the clock is violated, but 2 different rate-drift models produced similar estimates. The prior on times, which incorporates fossil-calibration information, had the greatest impact on posterior time estimation. In particular, the strategies used by the 2 programs to incorporate minimum- and maximum-age bounds led to very different time priors and were responsible for large differences in posterior time estimates in a previous study. The results highlight the critical importance of fossil calibrations to molecular dating and the need for probabilistic modeling of fossil depositions, preservations, and sampling to provide statistical summaries of information in the fossil record concerning species divergence times. [Bayesian method; divergence time; fossil calibration; molecular clock.]

The molecular clock (rate constancy among lineages) (Zuckerkandl and Pauling 1965) provides a powerful way for dating species divergences. Under the clock assumption, the expected distance between sequences grows linearly with the time of divergence between the species. If the ages of one or more nodes in a phylogenetic tree can be fixed based on the fossil or geological data, molecular branch lengths can be converted into absolute geological times for all the remaining nodes in the tree. Recent work has focused on relaxing the assumption of a molecular clock through rate smoothing (Sanderson 1997, 2002; Yang 2004; Aris-Brosou 2007), likelihood local-clock models (Rambaut and Bromham 1998; Yoder and Yang 2000), or explicit modeling of the rate-drift process (Thorne et al. 1998; Drummond et al. 2006; Rannala and Yang 2007). Most of the recent methods can analyze heterogeneous data from multiple gene loci and accommodate multiple fossil calibrations.

Most early molecular dating studies assumed that fossil calibrations provide known ages of nodes with certainty (Graur and Martin 2004). Although it is now well recognized that the fossil calibrations always underestimate divergence dates, the uncertainties concerning the degree to which fossil calibrations approximate divergence dates are neither easy to codify nor to accommodate. For example, considerable difficulty exists in the use of the likelihood method to account for uncertainties in fossil calibrations (Yang 2006, section 7.3.5). Sanderson (1997) suggested the use of constrained optimization as a way of incorporating fossil calibrations in the form of minimum- and maximum-age bounds. However, this strategy of accounting for uncertainties in fossil calibrations does not appear to be valid as it creates a nonidentifiability problem if none of the node ages is known with certainty (Yang 2006, p. 235–236). Also the use of the nonparametric bootstrap method to assess errors in the time estimates in the penalized likelihood method fails to accommodate uncertainties in fossil calibrations even though it accounts for sampling errors due to finite sequence data (Thorne and Kishino 2005).

The Bayesian method provides a natural framework for integrating different sources of information, including information about divergence times based on the fossil record. Thorne et al. (1998) developed the first Bayesian method for molecular dating that incorporates uncertainties in fossil calibrations by the use of minimum and maximum bounds. Soft bounds or flexible statistical distributions were introduced by Yang and Rannala (2006) and Drummond et al. (2006). In the Bayesian framework, priors are assigned on rates and times and are combined with the likelihood on the sequence data to generate posterior distributions, with the computation achieved through Markov chain Monte Carlo (MCMC) (Thorne et al. 1998). Fossil calibrations are incorporated in the algorithm through the time prior. This methodology has been implemented in MULTIDIVTIME of Thorne et al. (1998) and Kishino et al. (2001), MCMCTREE of Yang and Rannala (2006) and Rannala and Yang (2007), and BEAST of Drummond et al. (2006). Those Bayesian algorithms involve many components, and it is not obvious which of them have the greatest impact on posterior time estimates. In an analysis of a data set of modern cats, Rannala and Yang (2007) used MCMCTREE to obtain time estimates that were about 1.43 times as old as estimates obtained from an earlier

MULTIDIVTIME analysis of the same data. Our efforts to understand such differences led us to the realization that the different strategies for specifying fossil calibrations used by the 2 programs can lead to very different priors on times and thus very different posterior time estimates, even though apparently the same fossil calibrations are used.

In this paper, we first provide a summary of differences between MCMCTREE and MULTIDIVTIME. We then describe a theoretical analysis of a simple case of 4 species to understand the specification of the time prior in the 2 programs under different schemes of fossil calibrations. This analysis motivated us to introduce a modification to the MCMCTREE specification of minimum bounds. We describe 3 analyses of previously published data sets concerning the divergences of *Felidae* (cats), *Amphibia* (frogs and their relatives), and *Actinopterygians* (fishes) to evaluate the impact of various factors on time estimation.

Note that molecular systematists and paleontologists use the terms "lower" and "upper" bounds in exactly opposite ways. To avoid confusion, we use "minimum" and "maximum" bounds in this paper. We also follow the geological convention of using "Ma" to refer to dates in millions of years before present and "Myr" to refer to spans of time in millions of years without reference to a datum.

Materials and Methods

Differences Between MCMCTREE and MULTIDIVTIME

For the sake of later discussion, we provide here a brief summary of the differences between the 2 programs.

(a) Prior on times and fossil calibrations. In MULTIDIVTIME, fossil calibration of a node takes the form of a minimum bound, a maximum bound, or joint minimum and maximum bounds. It is specified by identifying the node number on the rooted tree for the ingroup species. In MCMCTREE, fossil calibration on a node can take the form of a minimum bound, a maximum bound, joint minimum and maximum bounds or a gamma distribution (fig. 2 in Yang and Rannala 2006). It is specified as a label on the node in the rooted tree. Bounds are hard in MULTIDIVTIME and soft in MCMCTREE even though MCMCTREE can emulate hard bounds. The 2 programs differ in their implementation of minimum and maximum bounds and in the specification of the prior on the ages of the noncalibration nodes. Those differences will be analyzed in the next section.

Here, we note the differences in the use of constraints on the age of the root based on fossils, or otherwise; these differences may have a considerable impact on posterior time estimates. MULTIDIVTIME requires a gamma prior on the root age, specified using the mean and standard deviation (rttime, rttimesd), as well as a maximum-age bound (big time). Both the gamma prior and the big-time constraint are applied whether or not there are fossil calibrations on the root.

MCMCTREE requires the root age to be constrained loosely from above (using the control variable RootAge) if no fossil calibration exists on the root or if the fossil is a minimum bound. This loose constraint can be a maximum bound or joint minimum and maximum bounds, and is not used if a fossil calibration exists on the root in the form of a maximum bound, both minimum and maximum bounds, or a gamma distribution.

(b) MULTIDIVTIME requires outgroup species to root the ingroup tree, that is, to break the branch around the root in the unrooted ingroup tree into 2 segments. This is achieved by the ESTBRANCHES program in the package, which calculates the maximum likelihood estimates (MLEs) of branch lengths and their variance–covariance matrix (see Yang 2006, fig. 7.10a). The MCMC analysis uses the rooted ingroup tree only, and the likelihood is calculated using a normal approximation to the MLEs of the branch lengths. In contrast, MCMCTREE does not use outgroups and works with sequence data from the ingroup species only. The likelihood is calculated either exactly or approximately for nucleotide sequences, whereas the normal approximation is always applied for amino acid or codon sequences. In the approximation, the lengths of the branches around the root are resolved through the assumed rate-drift model (see Yang 2006, fig. 7.10b).

(c) Both programs use a gamma prior for the rate at the root (rtrate, rtratesd in MULTIDIVTIME). As the prior means of rates are the same for all nodes, the mean root rate is also the overall rate for the whole tree. Both programs also specify a gamma prior for the rate-drift parameter (ν or brownmean and brownsd in MULTIDIVTIME or σ^2 in MCMCTREE).

Gamma distributions are specified using the mean (m) and standard deviation (s) in MULTIDIVTIME and using the shape (α) and the scale (β) parameters in MCMCTREE. They are related as $m = \alpha/\beta$ and $s = \sqrt{\alpha}/\beta$ or $\alpha = (m/s)^2$ and $\beta = m/s^2$.

Implementation of Fossil Calibrations in MCMCTREE and MULTIDIVTIME

Here, we discuss the specification of the prior on times, especially the implementation of fossil-based minimum and maximum bounds. Given the root age, MULTIDIVTIME specifies the distribution of other node ages by assigning a Dirichlet density on the proportions of the time segments on the path from the root to the tip (Kishino et al. 2001). Minimum and maximum bounds are then applied by truncating the joint density of times, that is, by removing times that violate those bounds. Truncation is achieved by proposing only feasible node ages in the MCMC algorithm.

In MCMCTREE, the joint distribution of the ages for the calibration nodes is generated by multiplying independent densities (see fig. 2 in Yang and Rannala 2006), followed by a truncation, which excludes node ages that violate the intrinsic constraint that any ancestral node should be older than any descendent node. Truncation

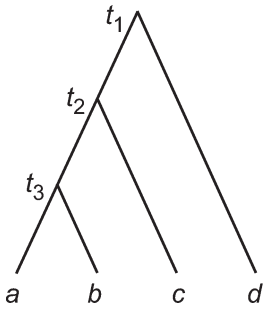

FIGURE 1. A rooted tree of 4 species with node ages t_1, t_2, and t_3.

is again achieved by proposing only feasible node ages in the MCMC. The distribution of the ages of the non-calibration nodes given the ages of the calibration nodes is specified by the birth–death process with species sampling (equation 3 in Yang and Rannala 2006).

We analyze the simple case of 4 species (Fig. 1), where the rooted tree has 3 node ages t_1, t_2, and t_3. A gamma prior on the root age is always assumed: $t_1 \sim G(\alpha, \beta)$. We use $\alpha = \beta = 5$, with the mean equal to 1, representing 100 Myr if 1 time unit is 100 Myr. A few simple cases, which may place fossil bounds on t_2 and t_3, are analyzed both theoretically and by running the 2 programs without sequence data. We set the birth rate, death rate, and sampling fraction to $\lambda = \mu = 1$ and $\rho = 0$, so that the kernel is a uniform distribution (see equation 7.28 in Yang and Rannala 2006) to mimic the Dirichlet prior used in MULTIDIVTIME. We study the joint time prior $f(t_1, t_2, t_3)$ and the marginal prior density $f(t_1)$. The difference between hard and soft bounds is not the concern here, so we use hard bounds in both programs. In MCMCTREE, this is achieved by using a small tail probability 10^{-300} instead of 0.025 in equations 16 and 17 and $\theta = 10^{100}$ in equation 15 of Yang and Rannala (2006). Numerical results are shown in Table 1 and Figure 2.

Case 0: No Fossil Calibration. The root age t_1 has the gamma density

$$g(t_1\,;\alpha,\beta) = \frac{1}{\Gamma(\alpha)}\beta^{\alpha} t_1^{\alpha-1}\,\mathrm{e}^{-\beta t_1}, \qquad (1)$$

and there are no other constraints on the tree.

To derive the joint time prior $f(t_1,t_2,t_3)$ used in MULTIDIVTIME, let the proportions of the 3 time segments from the root to tip species *a* (Fig. 1) be y_1, y_2, y_3, with $y_1+y_2+y_3=1$. The proportions y_2 and y_3 have a Dirichlet distribution, with density

$$f(y_2, y_3) = 2, \quad \text{for } y_2, y_3 > 0,\ \ y_2 + y_3 < 1. \qquad (2)$$

A transform of variables from t_1, y_2, y_3 to t_1, t_2, t_3 produces

$$f_0(t_1, t_2, t_3) = g(t_1) \times 2/t_1^2, \quad 0 < t_3 < t_2 < t_1. \qquad (3)$$

It is easy to see that $E(t_2) = E(t_1) \times 2/3$ and $E(t_3) = E(t_1)/3$. Also the marginal density of t_1 is $\int_0^{t_1}\int_0^{t_2} f_0(t_1, t_2, t_3)\,\mathrm{d}t_3\,\mathrm{d}t_2 = g(t_1)$.

TABLE 1. Prior means and 95% CIs for node ages in the tree of Figure 1 implemented in the 2 programs

Calibration	t_1	t_2	t_3
Case 0, none			
MULTIDIVTIME	1.000 (0.325, 2.055)	0.667	0.333
MCMCTREE v4.1	1.000 (0.325, 2.047)	0.667	0.333
Theory (Equation 1)	*1.000 (0.325, 2.048)*		
Case 1, $t_2 > 0.5$			
MULTIDIVTIME	1.185 (0.607, 2.187)	0.894	0.449
Theory (equation 8)	*1.186 (0.608, 2.185)*		
MCMCTREE			
v4.1	1.374 (0.673, 2.485)	0.936	0.468
v4.1 theory (Equation 11)	*1.374 (0.673, 2.483)*		
v4.2 $L(t_L, 0.1, 2)$	1.309 (0.660, 2.339)	0.856	0.427
v4.2 $L(t_L, 0.1, 1)$	**1.256 (0.645, 2.269)**	**0.787**	**0.394**
v4.2 $L(t_L, 0.1, 0.5)$	1.202 (0.626, 2.195)	0.711	0.356
v4.2 $L(t_L, 0.1, 0.2)$	1.147 (0.600, 2.137)	0.630	0.315
Case 2, $t_2 > 0.5$, $t_3 > 0.3$			
MULTIDIVTIME	1.240 (0.628, 2.262)	0.954	0.627
Theory (Equation 14)	*1.240 (0.628, 2.260)*		
MCMCTREE			
v4.1	1.548 (0.752, 2.745)	1.153	0.726
v4.1 theory (Equation 17)	*1.549 (0.751, 2.745)*		
v4.2 $L(t_L, 0.1, 2)$	1.382 (0.702, 2.423)	0.954	0.573
v4.2 $L(t_L, 0.1, 1)$	**1.290 (0.667, 2.306)**	**0.836**	**0.496**
v4.2 $L(t_L, 0.1, 0.5)$	1.215 (0.630, 2.201)	0.731	0.432
v4.2 $L(t_L, 0.1, 0.2)$	1.150 (0.601, 2.139)	0.633	0.378
Case 3, $0.3 < t_2 < 0.5$, $0.1 < t_3 < 0.2$			
MULTIDIVTIME	0.758 (0.372, 1.546)	0.393	0.150
Theory (Equation 20)	*0.757 (0.371, 1.547)*		
MCMCTREE	1.040 (0.445, 2.068)	0.398	0.150
Theory (Equation 23)	*1.040 (0.446, 2.066)*		
Case 4, $t_2 < 0.5$			
MULTIDIVTIME	0.711 (0.257, 1.519)	0.313	0.156
MCMCTREE	1.017 (0.374, 2.056)	0.246	0.123
Case 5, $t_2 < 0.5$, $t_3 > 0.3$			
MULTIDIVTIME	0.782 (0.412, 1.564)	0.429	0.364
MCMCTREE v4.1	1.048 (0.471, 2.072)	0.427	0.356
MCMCTREE v4.2 $L(t_L, 0.1, 1)$	1.050 (0.475, 2.069)	0.431	0.364
Case 6, $t_2 < 0.5$, $0.3 < t_3 < 0.5$			
MULTIDIVTIME	0.782 (0.412, 1.564)	0.429	0.364
MCMCTREE	1.050 (0.475, 2.068)	0.432	0.366
Case 7, $0.3 < t_2 < 0.5$			
MULTIDIVTIME	0.761 (0.378, 1.547)	0.402	0.201
MCMCTREE	1.040 (0.446, 2.066)	0.398	0.199

Notes: The root age is assigned a gamma prior $t_1 \sim G(5, 5)$, with mean 1, whereas fossil constraints exist on t_2 and/or t_3 in Cases 1–7. Hard bounds are implemented in both programs. Cases 0–3 are studied theoretically, with results shown in italics. The calibration $L(t_L, 0.1, 1)$ is used to analyze the 3 real data sets in this paper and are shown here in bold.

In MCMCTREE, the density $f(t_2, t_3|t_1)$ is specified using the birth–death process with species sampling. With $\lambda = \mu = 1$ and $\rho = 0$, the kernel density is $h(t) = 1/t_1, 0 < t < t_1$, so that

$$f(t_2, t_3|t_1) = 2/t_1^2, \quad 0 < t_3 < t_2 < t_1. \qquad (4)$$

Thus,

$$f(t_1, t_2, t_3) = g(t_1) \times 2/t_1^2, \quad 0 < t_3 < t_2 < t_1. \qquad (5)$$

This is the same as Equation 3. The 2 programs implement the same prior in this case. The gamma density $G(5, 5)$ is shown as curve a in Figure 2a,b, whereas the mean ages of t_2 and t_3 are shown in Table 1, Case 0.

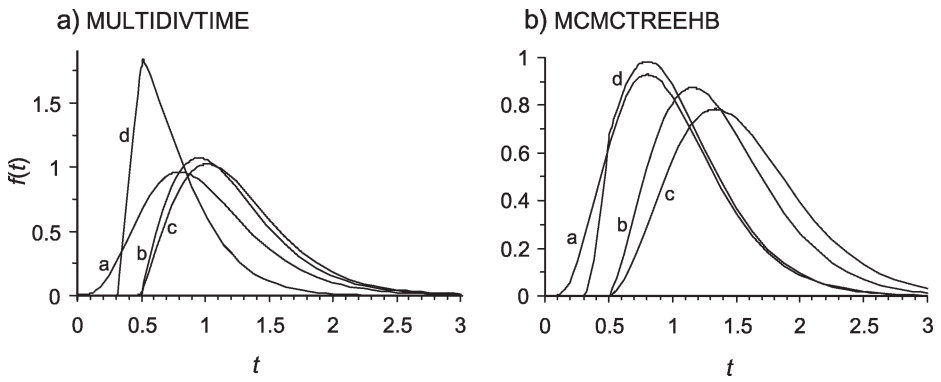

FIGURE 2. The marginal prior density of root age t_1 (Fig. 1) generated by a) MULTIDIVTIME and b) MCMCTREE 4.1. Curve a in both graphs shows the gamma prior $t_1 \sim G(5, 5)$. Curve b is for the minimum bound of Case 1 in Table 1: $t_2 > 0.5$. Curve c is for 2 minimum bounds of Case 2: $t_2 > 0.5, t_3 > 0.3$, whereas curve d is for 2 joint bounds of Case 3: $0.3 < t_2 < 0.5$ and $0.1 < t_3 < 0.2$. The means and 95% intervals for t_1 and the means for t_2 and t_3 are listed in Table 1 for Cases 0–3, corresponding to curves a–d.

Case 1: Minimum Bound $t_2 > t_{2L}$.

In MULTIDIVTIME, the fossil bound is implemented by truncating the density of Case 0 (Equation 3) so that $t_2 > t_{2L}$. The resultant joint time prior is thus

$$f_{\mathrm{I}}(t_1, t_2, t_3) = \frac{1}{J_{\mathrm{I}}} g(t_1) \times 2/t_1^2, \quad 0 < t_3 < t_2 < t_1, \;\; t_2 > t_{2L}, \tag{6}$$

where the normalizing constant is

$$\begin{aligned} J_{\mathrm{I}} &= \int_{t_{2L}}^{\infty} \int_{t_{2L}}^{t_1} \int_{0}^{t_2} f_0(t_1, t_2, t_3) \mathrm{d}t_3 \, \mathrm{d}t_2 \, \mathrm{d}t_1 \\ &= \int_{t_{2L}}^{\infty} g(t_1) \times (1 - t_{2L}^2/t_1^2) \mathrm{d}t_1. \end{aligned} \tag{7}$$

The marginal density of t_1 is

$$\begin{aligned} f_{\mathrm{I}}(t_1) &= \int_{t_{2L}}^{t_1} \int_{0}^{t_2} f_{\mathrm{I}}(t_1, t_2, t_3) \mathrm{d}t_3 \, \mathrm{d}t_2 \\ &= \frac{1}{J_{\mathrm{I}}} g(t_1) \times (1 - t_{2L}^2/t_1^2), \quad t_1 > t_{2L}. \end{aligned} \tag{8}$$

The mean and the 95% credibility interval (CI) for t_1 can be calculated numerically from this density.

MCMCTREE represents the minimum bound $t_2 > t_{2L}$ by the improper density $f(t_2) = 1, t_2 > t_{2L}$ (see equation 15 in Yang and Rannala 2006). This is multiplied with the gamma density $g(t_1)$, followed by the truncation $t_1 > t_2$. The resulting density $f(t_1,t_2)$ is multiplied by $f_{\mathrm{BD}}(t_3|t_1, t_2)$, the conditional density of t_3 given t_1 and t_2, specified by the birth–death process with species sampling (see equation 12 in Yang and Rannala 2006), to give rise to the joint time prior as

$$f_{\mathrm{I}}(t_1, t_2, t_3) = \frac{1}{Z_{\mathrm{I}}} g(t_1), \quad 0 < t_3 < t_2 < t_1, \;\; t_2 > t_{2L}, \tag{9}$$

where the normalizing constant is

$$Z_{\mathrm{I}} = \int_{t_{2L}}^{\infty} \int_{t_{2L}}^{t_1} g(t_1) \mathrm{d}t_2 \, \mathrm{d}t_1 = \int_{t_{2L}}^{\infty} g(t_1) \times (t_1 - t_{2L}) \mathrm{d}t_1. \tag{10}$$

The marginal density of t_1 is

$$f_{\mathrm{I}}(t_1) = \frac{1}{Z_{\mathrm{I}}} g(t_1) \times (t_1 - t_{2L}), \quad t_1 > t_{2L}. \tag{11}$$

Figure 2a,b, curve b, shows the marginal densities for t_1 generated by the 2 programs (Equations 8 and 11) for the minimum bound $t_2 > 0.5$. The prior means for t_1, t_2, and t_3 are shown in Table 1, Case 1. The truncation pushes the node ages upward, so that the mean of t_1 becomes 1.186 in MULTIDIVTIME and 1.374 in MCMCTREE. The effect on the MCMCTREE implementation is quite dramatic. The numerical results obtained from running the 2 programs agree well with the theoretical calculations.

Case 2. Two Minimum Bounds $t_2 > t_{2L}$ and $t_3 > t_{3L}$.

In MULTIDIVTIME, the constraints are again implemented by truncating the density of Case 0 (Equation 3). The joint time prior is thus

$$\begin{aligned} f_{\mathrm{II}}(t_1, t_2, t_3) = \frac{1}{J_{\mathrm{II}}} g(t_1) \times 2/t_1^2, \quad & 0 < t_3 < t_2 < t_1, \\ & t_2 > t_{2L}, \;\; t_3 > t_{3L}, \end{aligned} \tag{12}$$

where

$$\begin{aligned} J_{\mathrm{II}} &= \int_{t_{2L}}^{\infty} \int_{t_{2L}}^{t_1} \int_{t_{3L}}^{t_2} f_0(t_1, t_2, t_3) \mathrm{d}t_3 \, \mathrm{d}t_2 \, \mathrm{d}t_1 \\ &= \int_{t_{2L}}^{\infty} g(t_1) \times (t_1 + t_{2L} - 2t_{3L})(t_1 - t_{2L})/t_1^2 \, \mathrm{d}t_1, \end{aligned} \tag{13}$$

and the marginal density of t_1 is

$$\begin{aligned} f_{\mathrm{II}}(t_1) = \int_{t_{2L}}^{t_1} \int_{t_{3L}}^{t_2} f_{\mathrm{II}}(t_1, t_2, t_3) \mathrm{d}t_3 \, \mathrm{d}t_2 = \frac{1}{J_{\mathrm{II}}} g(t_1) \\ \times (t_1 + t_{2L} - 2t_{3L})(t_1 - t_{2L})/t_1^2, \quad t_1 > t_{2L}. \end{aligned} \tag{14}$$

In MCMCTREE, the gamma density $g(t_1)$ and the 2 improper flat densities for t_2 and t_3 are multiplied, followed by the truncation $t_3 < t_2 < t_1$. The resulting joint

time prior is

$$f_{\mathrm{II}}(t_1,t_2,t_3)=\frac{1}{Z_{\mathrm{II}}}g(t_1),\quad 0<t_3<t_2<t_1,\ \ t_2>t_{2L},\ \ t_3>t_{3L}, \tag{15}$$

where

$$\begin{aligned}Z_{\mathrm{II}}&=\int_{t_{2L}}^{\infty}\int_{t_{2L}}^{t_1}\int_{t_{3L}}^{t_2}g(t_1)\mathrm{d}t_3\,\mathrm{d}t_2\,\mathrm{d}t_1\\&=\int_{t_{2L}}^{\infty}g(t_1)\times\frac{1}{2}(t_1+t_{2L}-2t_{3L})(t_1-t_{2L})\mathrm{d}t_1.\end{aligned} \tag{16}$$

The marginal density of t_1 is

$$f_{\mathrm{II}}(t_1)=\frac{1}{Z_{\mathrm{II}}}\int_{t_{2L}}^{t_1}\int_{t_{2L}}^{t_2}g(t_1)\mathrm{d}t_3\,\mathrm{d}t_2=\frac{1}{Z_{\mathrm{II}}}g(t_1)\times\frac{1}{2}(t_1+t_{2L}-2t_{3L})(t_1-t_{2L}),\quad t_1<t_{2L}. \tag{17}$$

Numerical results are generated for the minimum bounds $t_2>0.5$ and $t_3>0.3$. Figure 2a,b, curve c, shows the marginal densities for t_1 generated by the 2 programs (Equations 14 and 17). The prior means for t_1, t_2, and t_3 are shown in Table 1, Case 2. The prior mean of t_1 becomes 1.240 in MULTIDIVTIME and 1.549 in MCMCTREE. It may appear surprising that given the constraint $t_2>0.5$, applying the additional constraint $t_3>0.3$ further increases the prior mean of t_1.

Case 3: Two Joint Bounds $t_{2L}<t_2<t_{2U}$ and $t_{3L}<t_3<t_{3U}$, with $t_{2L}>t_{3U}$.

MULTIDIVTIME implements the bounds by truncating the joint density of Case 0 (Equation 3). The resulting joint time prior is

$$f_{\mathrm{III}}(t_1,t_2,t_3)=\frac{1}{J_{\mathrm{III}}}g(t_1)\times 2/t_1^2,\quad 0<t_3<t_2<t_1,\ \ t_{2L}<t_2<t_{2U},\ \ t_{3L}<t_3<t_{3U}, \tag{18}$$

where

$$\begin{aligned}J_{\mathrm{III}}&=\int_{t_{2L}}^{\infty}\int_{t_{2L}}^{\min(t_1,t_{2U})}\int_{t_{3L}}^{t_{3U}}f_0(t_1,t_2,t_3)\mathrm{d}t_3\,\mathrm{d}t_2\,\mathrm{d}t_1\\&=\int_{t_{2L}}^{t_{2U}}g(t_1)/t_1^2\times(t_{3U}-t_{3L})(t_1-t_{2L})\mathrm{d}t_1\\&\quad+\int_{t_{2U}}^{\infty}g(t_1)/t_1^2\times(t_{3U}-t_{3L})(t_{2U}-t_{2L})\mathrm{d}t_1.\end{aligned} \tag{19}$$

The marginal density of t_1 is

$$f_{\mathrm{III}}(t_1)=\frac{1}{J_{\mathrm{III}}}g(t_1)/t_1^2\times(t_{3U}-t_{3L})(\min(t_1,t_{2U})-t_{2L})\quad t_1>t_{2L}. \tag{20}$$

In MCMCTREE, the 2 fossil calibrations are represented by uniform distributions $t_2\sim U(t_{2L},\,t_{2U})$ and $t_3\sim U(t_{3L},\,t_{3U})$. These densities are multiplied with the gamma density for t_1, followed by the truncation $t_3<t_2<t_1$ to generate the joint time prior

$$f_{\mathrm{III}}(t_1,t_2,t_3)=\frac{1}{Z_{\mathrm{III}}}g(t_1),\quad 0<t_3<t_2<t_1,\ \ t_{2L}<t_2<t_{2U},\ \ t_{3L}<t_3<t_{3U}, \tag{21}$$

where

$$\begin{aligned}Z_{\mathrm{III}}&=\int_{t_{2L}}^{\infty}\int_{t_{2L}}^{\min(t_1,t_{2U})}\int_{t_{3L}}^{t_{3U}}g(t_1)\mathrm{d}t_3\,\mathrm{d}t_2\,\mathrm{d}t_1\\&=\int_{t_{2L}}^{t_{2U}}g(t_1)\times(t_{3U}-t_{3L})(t_1-t_{2L})\mathrm{d}t_1\\&\quad+\int_{t_{2U}}^{\infty}g(t_1)\times(t_{3U}-t_{3L})(t_{2U}-t_{2L})\mathrm{d}t_1.\end{aligned} \tag{22}$$

The marginal density of t_1 is

$$f_{\mathrm{III}}(t_1)=\frac{1}{Z_{\mathrm{III}}}g(t_1)\times(t_{3U}-t_{3L})(\min(t_1,t_{2U})-t_{2L}),\quad t_1<t_{2L}. \tag{23}$$

Numerical results are generated for the bounds $0.3<t_2<0.5$ and $0.1<t_3<0.2$. Figure 2a,b, curve d, shows the marginal densities for t_1 generated by the 2 programs (Equations 20 and 23). The prior means for t_1, t_2, and t_3 are shown in Table 1, Case 3. In this case, the prior mean of t_1 is 1.040 for MCMCTREE, almost identical to the mean from the gamma $g(t_1)$, but is 0.757 for MULTIDIVTIME, much smaller than 1.

A few additional cases (Cases 4–7 in Table 1) are examined by running the programs, without the theoretical analysis. Application of maximum bounds (e.g., $t_2<0.5$ in Case 4) leads to reduction of the prior mean for t_1 (e.g., from 1 to 0.711) in MULTIDIVTIME, whereas it has little impact in MCMCTREE. Note also that in Case 6, the bound $t_3<0.5$ should be uninformative given that $t_2<0.5$ (as we know $t_3<t_2$). However, its use affects the prior in MCMCTREE, although not in MULTIDIVTIME (cf. Cases 5 and 6 in Table 1).

The differences discussed here do not indicate a superiority of one program over the other. In theory, each program allows the user to change the fossil specifications as well as the root age constraint until the joint time prior is a good summary of our knowledge of the divergence times among those species based on relevant fossil data. Nevertheless, specifying multidimensional priors is notoriously difficult, and prior solicitation may be made easier if the prior densities specified by the user (before the truncation) are close to those actually implemented (after the truncation). It is important to note that these two can be very different. For example, the prior on t_1 actually implemented in Case 1 is not the gamma in either program (see Equations 8 and 11). In this sense, the effects of minimum bounds in MCMCTREE and of maximum and joint bounds in MULTIDIVTIME on the

prior of root age may be disconcerting. To address this issue, we have introduced a modification to the implementation of the minimum bound in MCMCTREE.

Modification to the MCMCTREE Implementation of Minimum Bound

A number of authors have discussed statistical distributions of lineage divergence times that may serve as a suitable summary of fossil evidence (Hedges and Kumar 2004; Barnett et al. 2005; Yang and Rannala 2006; Ho 2007). A well-recognized feature of the fossil data is that they should provide hard or nearly hard minimum bounds but soft maximum bounds (Benton and Donoghue 2007). The most probable time of divergence should be older than the fossil minimum because the acquisition of fossilizable apomorphies reflecting divergence of descendent lineages will significantly postdate the actual divergence (Smith and Peterson 2002; Steiper and Young 2008). Thus, the probability density should increase with increasing age from the minimum, peak at the most probable time, and decay at a rate commensurate with the strength of evidence supporting the peak value. The most likely time could be informed by phylogenetic bracketing (Reisz and Muller 2004; Marshall 2008) or knowledge of the gaps in the rock record, which underpin the gaps in the fossil record.

With the aim of better reflecting the nature of fossil minima in establishing probability distributions of divergence times, we used a truncated Cauchy distribution to represent the minimum bound, in place of the improper distribution of equation 15 in Yang and Rannala (2006). The Cauchy distribution with location parameter t_0 and scale parameter s has the density

$$f(t; t_0, s) = \frac{1}{\pi s\left[1 + \left(\frac{t-t_0}{s}\right)^2\right]}, \quad -\infty < t < \infty, \qquad (24)$$

and the cumulative distribution function

$$F(t; t_0, s) = \frac{1}{\pi}\tan^{-1}\left(\frac{t-t_0}{s}\right) + \frac{1}{2}, \quad -\infty < t < \infty. \qquad (25)$$

This is Student's t distribution with df = 1. The density is symmetrical around t_0, the mode and median, and is very heavy tailed. We expect the true node age to be close to the minimum bound t_L, but the distance may depend on the quality of the fossil data. Thus, we place the mode at $t_0 = t_L(1 + p)$, where the offset proportion p should be small (say 0.1) if the fossil estimate is a good approximation to the true divergence time, whereas p should be large (say 0.5) if the fossil estimate is a poor estimate of the true age. The scale parameter is specified as $s = ct_L$, where a smaller c means that the density drops off more rapidly away from the mode. We then truncate the distribution so that $t > t_L$. To make the bound soft, we assign a small probability left of t_L, represented by a rapid power decay, as in Yang and Rannala (2006). The minimum bound $t > t_L$ is thus represented by the following density, specified by parameters p and c:

$$f(t; t_L, p, c) = \begin{cases} 0.975 \times \dfrac{1}{A\pi c t_L\left[1+\left(\frac{t-t_L(1+p)}{ct_L}\right)^2\right]}, & \text{if } t > t_L, \\ 0.025 \times \dfrac{\theta}{t_L}\left(\dfrac{t}{t_L}\right)^{\theta-1}, & \text{if } 0 < t \leqslant t_L, \end{cases} \qquad (26)$$

where the normalizing constant due to the truncation of the Cauchy distribution is $A = 1 - F(t_L, t_0, s) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{p}{c}\right)$, and where $\theta = \frac{0.975}{0.025} \times \frac{1}{\pi A c[1+(p/c)^2]}$ is chosen to make the density continuous at t_L. The distribution has mode at $t_L(1+p)$, and the 2.5% and 97.5% limits at t_L and $t_L\left[1 + p + c\tan\left(\pi\left(\frac{1}{2} - \frac{0.025A}{0.975}\right)\right)\right]$. Both the mean and the variance are infinite.

The new minimum bound is implemented in the MCMCTREE program in PAML version 4.2, using the format $L(t_L, p, c)$. In this paper, we will refer to the old and new implementations as MCMCTREE 4.1 and MCMCTREE 4.2. Figure 3a,b shows the calibration $t > t_L = 1$ for different values of parameters p (0.1 or 0.5) and c (0.2, 0.5, 1, or 2). The 2.5% limit is at 1, whereas the 97.5% limits for those values of c are 5.8, 12.8, 24.4, and 47.8, respectively, when $p = 0.1$, and are 6.2, 13.2, 24.8, and 48.2, when $p = 0.5$. If 1 time unit is 100 Myr, then $L(1, 0.1, 1)$ implements the constraint

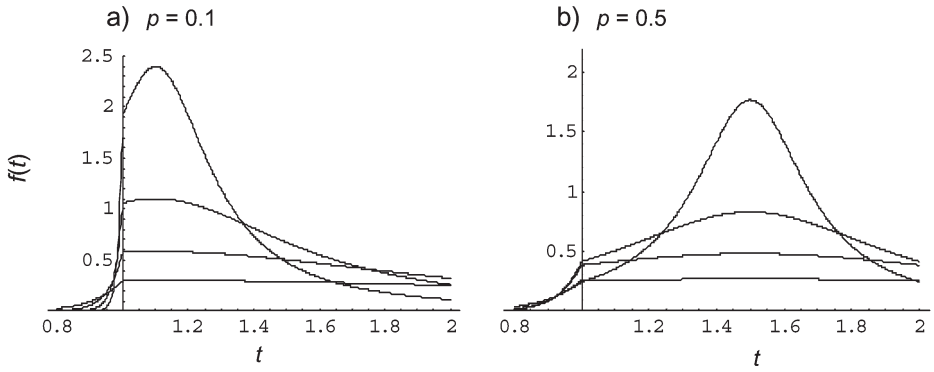

FIGURE 3. The minimum-bound density $L(t_L, p, c)$ implemented in MCMCTREE 4.2. For each value of p (0.1 or 0.5), the 4 curves from top to bottom correspond to $c = 0.2$, 0.5, 1, and 2. The node is at 1.1 for $p = 0.1$ and at 1.5 for $p = 0.5$.

$t > 100$ Ma by the 95% interval (100 Ma, 2440 Ma). Thus, although the distribution bounds the node age from above, the bound is very weak. Although both large p and large c push up the node ages, c has a far greater impact. When $c = \infty$, the density should converge to that implemented in version 4.1. We used $p = 0.1$ and $c = 0.2$, 0.5, 1, and 2 to implement the minimum bounds for the 4-species example, with the results shown in Table 1.

We do not apply the truncated Cauchy density (Equation 26) for calibration at the root. If a fossil minimum bound is used for the root, we insist that a maximum bound is specified as well (using the control variable RootAge in the control file) in which case the joint bound (fig. 2c in Yang and Rannala 2006) is used.

We also tested a few distributions to represent the maximum bound $t < t_U$, alternative to the soft uniform distribution over $(0, t_U)$ implemented in MCMCTREE 4.1 (Yang and Rannala 2006; Fig. 2c), but we decided not to change this distribution. Although we expect that the true age is more likely to be intermediate of 0 and t_U (rather than equally likely to be anywhere in the interval), it is unclear what intermediate ages are mostly likely to be true, making it difficult to construct a reasonable prior. Furthermore, maximum bounds are uncommon (because they are difficult to justify on the basis of fossil data), and minimal bounds often exist on daughter nodes on the tree, so that the uniform distribution $U(0, t_U)$ is automatically truncated from below. This is particularly the case for the loose maximum bound applied to the root.

For fixed values of p and c, the prior density (Equation 26) implies larger absolute errors for older calibrations. We do not imply, however, that the quality of fossils deteriorates proportionally with divergence time. Indeed, the degree to which fossil estimates approximate divergence times may not differ materially for events that differ in age by an order of magnitude. We stress, therefore, that in real data analysis, different values of p and c should be used for each minimum-bound calibration based on a careful assessment of the fossil and geological data on which the bound is based. However, we do not make such an attempt in this study.

Sequence Data sets

Three previously published data sets were analyzed to estimate divergence times, representing different time scales and different scenarios of fossil calibrations. GenBank accession numbers for the sequences can be found in the original publications.

The first data set consists of the nuclear RAG1 genes from 52 amphibian species, originally analyzed by San Mauro et al. (2005). The alignment included 1368 sites. The rooted ingroup tree is shown in Figures 4 and 5, from San Mauro et al. Two maximum and 8 minimum bounds were used as calibrations (Table 2). In the MULTIDIVTIME analysis, a coelacanth, *Latimeria chalumnae*, was used as the outgroup.

The second data set consists of the mitochondrial genomes from 28 actinopterygian bony fishes and 2 coelacanths, analyzed previously by Inoue et al. (2005). The rooted tree is shown in Figure 6. As in Inoue et al., 2 maximum and 12 minimum bounds were used (Table 2). The alignment has 10,327 sites, including the first and second codon positions of 12 protein-coding genes encoded by the same strand of the genome, 22 transfer RNA genes, and 2 ribosomal RNA genes. These were analyzed as 4 partitions in both programs, with independent rates and substitution parameters estimated for each partition. For the MULTIDIVTIME analysis, the catshark, *Scyliorhinus canicula*, was used as the outgroup.

The third data set consists of nuclear genes from 38 species of modern cats (family Felidae), analyzed by Johnson et al. (2006) and Rannala and Yang (2007). For the MULTIDIVTIME analysis, the banded linsang, *Prionodon linsang*, was used as the outgroup. The rooted ingroup species tree is shown in Figure 7, extracted from the phylogeny of Johnson et al. We used 1 maximum and 11 minimum bounds as fossil calibrations according to Johnson et al. (Table 2). However, 2 minimum calibrations, 5 Ma for the *Leopardus geoffroyi—Felis catus* split and 1 Ma for the *Puma concolor—F. catus* split (nodes 4 and 6 in Johnson et al. 2006) are redundant and not used. The alignment included 19,984 sites, from 30 nuclear genes (19 autosomal, 5 X-linked, and 6 Y-linked genes), all of which were analyzed as one partition, as in previous studies.

MCMCTREE 4.1 and 4.2 and MULTIDIVTIME were used in the analysis, with the settings chosen to be as similar as possible. In both programs, the likelihood is calculated using a normal approximation of the MLEs of branch lengths, obtained using the BASEML program under the F84 + Γ_5 substitution model (Yang 1994) (see Yang 2006, p. 246–247, for details). The gamma priors for the overall rate and for the rate-drift parameter (ν or σ^2) are summarized in Table 3. The prior mean for the overall rate (rtrate) is set to a rough estimate obtained by fitting a molecular clock to the sequence data, using point calibrations. A gamma prior is also specified on the root age for MULTIDIVTIME (rttime), whereas it is specified for MCMCTREE for the amphibian data set only, which does not have fossil calibrations on the root. Other prior parameters in MULTIDIVTIME were fixed as follows: rttmsd = rttm/2, rtratesd = rtrate, and brownmean = brownsd = 1/rttm, and the same were applied in MCMCTREE.

In MCMCTREE, the auto-correlated rates model (clock = 3) was used to specify the prior of rates, as in MULTIDIVTIME. The parameters of the birth–death process with species sampling were fixed at $\lambda = \mu = 1$ and $\rho = 0$, so that the prior is similar to that used in MULTIDIVTIME. A loose maximum bound for the root age of 1000 Ma is used for MULTIDIVTIME.

The number of iterations, the burn-in, and the sampling frequency were determined in pilot runs of the programs. Every analysis was conducted at least twice to ensure consistency between runs.

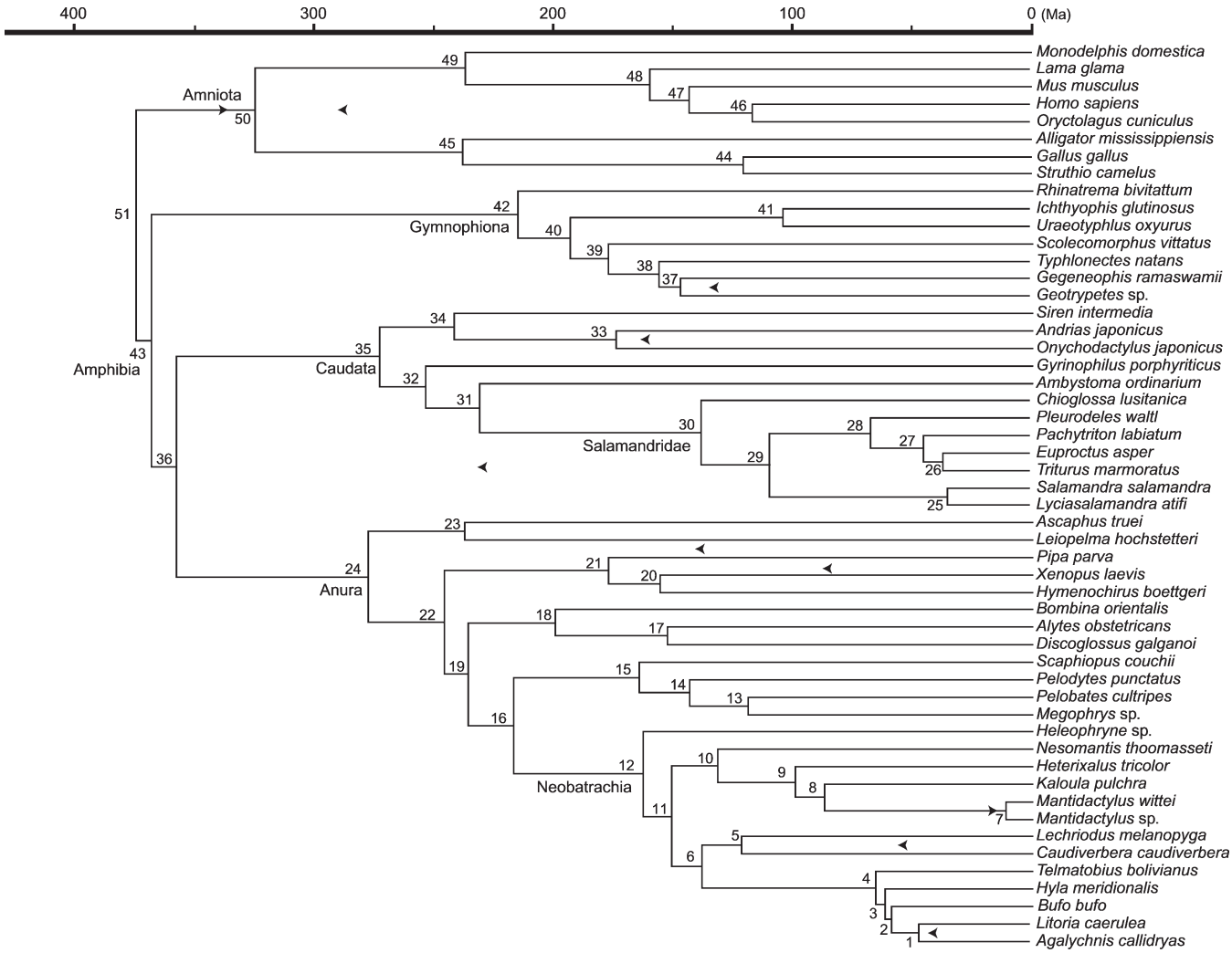

FIGURE 4. The rooted tree for the amphibian data set, showing fossil calibrations. The branches are drawn to show the estimates of divergence times in San Mauro et al. (2005).

RESULTS

The Implementation of Minimum-Bound Calibration

We evaluated the impact of parameters p and c in the minimum-bound distribution (Equation 26) on the prior and posterior of divergence times. We analyzed the 3 data sets with 2 different values for p (0.1 and 0.5) and 4 different values for c (0.2, 0.5, 1, and 2). The results, shown in supplementary Figures S1–S3 (available from http://sysbio.oxfordjournals.org/), match closely the patterns for the 4-species case of Table 1. For example, increasing either p or c made node ages older, with c having a larger effect. Importantly, the effect was present in the posterior as well as in the prior. The only exception is that in the amphibian data set, the posterior did not seem to be sensitive to the values of p and c. The reason for this exception is not known.

Based on those tests and on the results for the 4-species example, we used $p = 0.1$ and $c = 1$ as default values for the analysis of the 3 data sets in this paper. At those values, MCMCTREE 4.2 typically produce an older age for the root but younger ages for the young nodes compared with MULTIDIVTIME.

The Amphibian Data set

The priors and posteriors of divergence times obtained from the analysis using MULTIDIVTIME as well as versions 4.1 and 4.2 of MCMCTREE are shown in the chronograms of Figure 5 in which the branch lengths represent the prior/posterior means of the node ages. The 95% CIs were shown for the root node only. Several nodes are connected by lines across the analyses for easy comparison.

The 2 versions of MCMCTREE produced more diffuse priors and wider prior intervals than MULTIDIVTIME. The prior means of node ages from MCMCTREE 4.1 were much older than those from MULTIDIVTIME, whereas those from MCMCTREE 4.2 were more similar. Compared with MULTIDIVTIME, MCMCTREE 4.2 produced older age estimates for 3 basal nodes (Nodes 36, 43, and 51 [the root]) but younger age estimates for

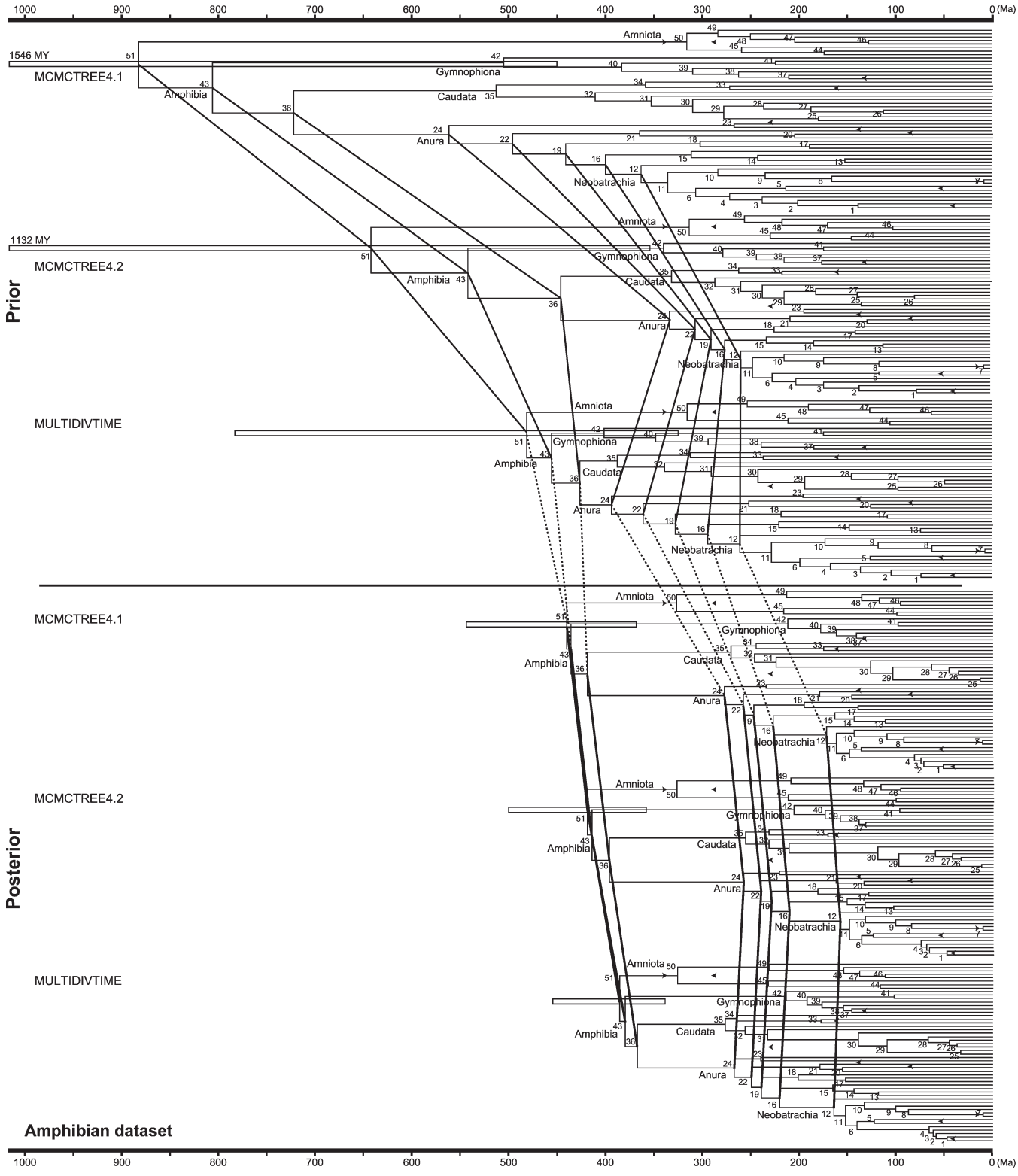

FIGURE 5. The amphibian tree showing the prior or posterior means of node ages estimated using MCMCTREE and MULTIDIVTIME.

other nodes (such as 4 key nodes within the Anura). The pattern is very similar to that seen in the analysis of the 4-species case (Table 1), where the root age t_1 was older in MCMCTREE 4.2 than in MULTIDIVTIME, whereas ages t_2 and t_3 were younger. However, those differences did not persist in the posterior.

The posterior time estimates obtained from MCMCTREE 4.1 were slightly older than those from

TABLE 2. Fossil constraints used in the 3 data sets (Myr)

Node		Bounds
Amphibian data set (San Mauro et al. 2005)		
50	*Oryctolagus–Struthio*	<338
		>288
37	*Gegeneophis–Geotrypetes*	>130
36	*Lyciasalamandra–Agalychnis*	>230
33	*Andrias–Onychodactylus*	>161
24	*Leiopelma–Agalychnis*	>140
21	*Pipa–Hymenochirus*	>86
7	*Mantidactylus wittei–Mantidactylus* sp.	<15
5	*Lechriodus–Caudiverbera*	>53
1	*Litoria–Agalychnis*	>42
Actinopterygian data set (Inoue et al. 2005)		
27	*Latimeria–Pagrus* (root)	<528
		>411
25	*Polypterus–Pagrus*	<450
21	*Atractosteus–Amia*	>155
20	*Scaphirhynchus–Atractosteus*	>141
19	*Polyodon–Scaphirhynchus*	>89
15	*Hiodon–Pantodon*	>155
14	*Osteoglossum–Pantodon*	>112
13	*Anguilla–Pagrus*	>97
12	*Notacanthus–Anguilla*	>90
10	*Conger–Anguilla*	>50
7	*Engraulis–Sardinops*	>57
6	*Cyprinus–Crossostoma*	>50
1	*Polymixia–Pagrus*	>90
Felid data set (Johnson et al. 2006)		
37	*Panthera–Felis catus* (root)	<16
36	*Neofelis nebulosa–Panthera*	>3.8
34	*Panthera uncia–P. tigris*	>1
27	*Caracal serval–C. caracal*	>3.8
24	*Leopardus pardalis–L. geoffroyi*	<5
22	*Leopardus jacobita–L. geoffroyi*	>1
18	*Lynx pardinus–Felis catus*	>5.3
17	*Lynx rufus–L. pardinus*	>2.5
13	*Acinonyx jubatus–Puma Concolor*	>3.8
12	*Puma yagouaroundi–P. Concolor*	>1.8
11	*Prionailurus bengalensis–Felis catus*	>4.2
10	*Otocolobus manul–Prionailurus bengalensis*	>1
1	*Felis silvestris–F. catus*	>1

Note: The node numbers refer to those in the trees of Figures 4 (amphibian), 6 (actinopterygian), and 7 (felid).

MULTIDIVTIME. The MCMCTREE 4.2 estimates were even more similar. The posterior distributions of the times produced by MCMCTREE were more diffuse than those generated by MULTIDIVTIME, as indicated by the wider CI for the root age. Overall, the CIs overlap substantially between the analyses. As the gamma prior on the root age is not based on any fossil data, we removed it in an MCMCTREE 4.2 analysis to assess its impact. This analysis produced very similar time estimates (see supplementary Fig. S4).

Our time estimates using MULTIDIVTIME were slightly older than those obtained by San Mauro et al. (2005) using the same program. For example, Node 43 (Gymnophiona/Anura) was dated to 380 Ma with 95% CI (334, 443) in our analysis, whereas San Mauro et al. obtained 367 Ma (328, 417). The time of divergence of salamanders (Caudata) and frogs (Anura) (Node 36) was 367 Ma (324, 427) in our analysis but 357 Ma (317, 405) by San Mauro et al. Although the CIs overlap between the 2 analyses, the differences are systematic and appear to be due to the use of different priors on the rate-drift parameter (ν). We used 0.24 for both brownmean and brownsd (Table 3), whereas San Mauro et al. used 0.01 for both values. All these clock estimates are in substantial discord with fossil minima because the earliest records of crown-lissamphibians and batrachians are latest Permian-early Triassic (ca. 251 Ma, Marjanovic and Laurin 2007; Ruta and Coates 2007).

The Actinopterygian Data set

The prior and posterior means of divergence times obtained using MULTIDIVTIME, and MCMCTREE 4.1 and 4.2 are shown in the trees of Figure 6. Compared with MULTIDIVTIME, MCMCTREE 4.1 produced much older prior mean ages, whereas MCMCTREE 4.2 produced older root age but younger ages for other nodes, similar to the analysis of the amphibian data set. Similar differences are visible in the posterior, but the posteriors were even more similar among the 3 analyses. The MULTIDIVTIME estimates are almost identical between this study and Inoue et al. (2005).

MCMCTREE 4.2 dated the basal actinopteran (Node 22) to 334 Ma (280, 383) and the basal teleostean (Node 16) to 296 Ma (248, 341). The MULTIDIVTIME estimates were older, at 376 Ma (336, 413) and 333 Ma (295, 371), respectively (see also Inoue et al. 2005). Recently, Hurley et al. (2007) reported a stem-amiid, †*Brachydegma*, from the Early Permian (>276 Ma). The age of †*Brachydegma* aligned closely with the MCMCTREE 4.2 estimate for the Lepisosteidae/*Amia* divergence (Node 21), at 333 Ma (279, 382). Our estimate for the date of divergence of stem-teleosts from Amiidae (and other members of the "Ancient fish clade") suggests a Triassic divergence (ca. 300 Ma) contrasting sharply with the oldest fossil record of stem-teleosts (Late Jurassic, >155 Ma; Arratia and Schultze 1999). The early fossil record of neopterygians and stem-teleosts has not been actively investigated and so this may go some way to explain the disparity in date estimates (Hurley et al. 2007). However, the topology of our tree—specifically with regard to the relationship between teleosts and the nonteleost actinoptergians, is not compatible with the scheme followed by most morphologists (Hurley et al. 2007), and this impacts materially upon the interpretation of the fossil minimum bounds. Furthermore, mitochondrial data sets have been shown to support much older divergence dates for Actinopteri and Teleostei than do nuclear data sets (Hurley et al. 2007).

The Felid Data set

The prior and posterior means obtained in different analyses of the felid data set are shown in the trees of Figure 7. MCMCTREE 4.1 produced much older (and nearly proportionally older) prior node ages than MULTIDIVTIME. MCMCTREE 4.2 is more similar to MULTIDIVTIME, with old nodes to be older and young nodes to be slightly younger in MCMCTREE 4.2 than in MULTIDIVTIME. The patterns are similar to those

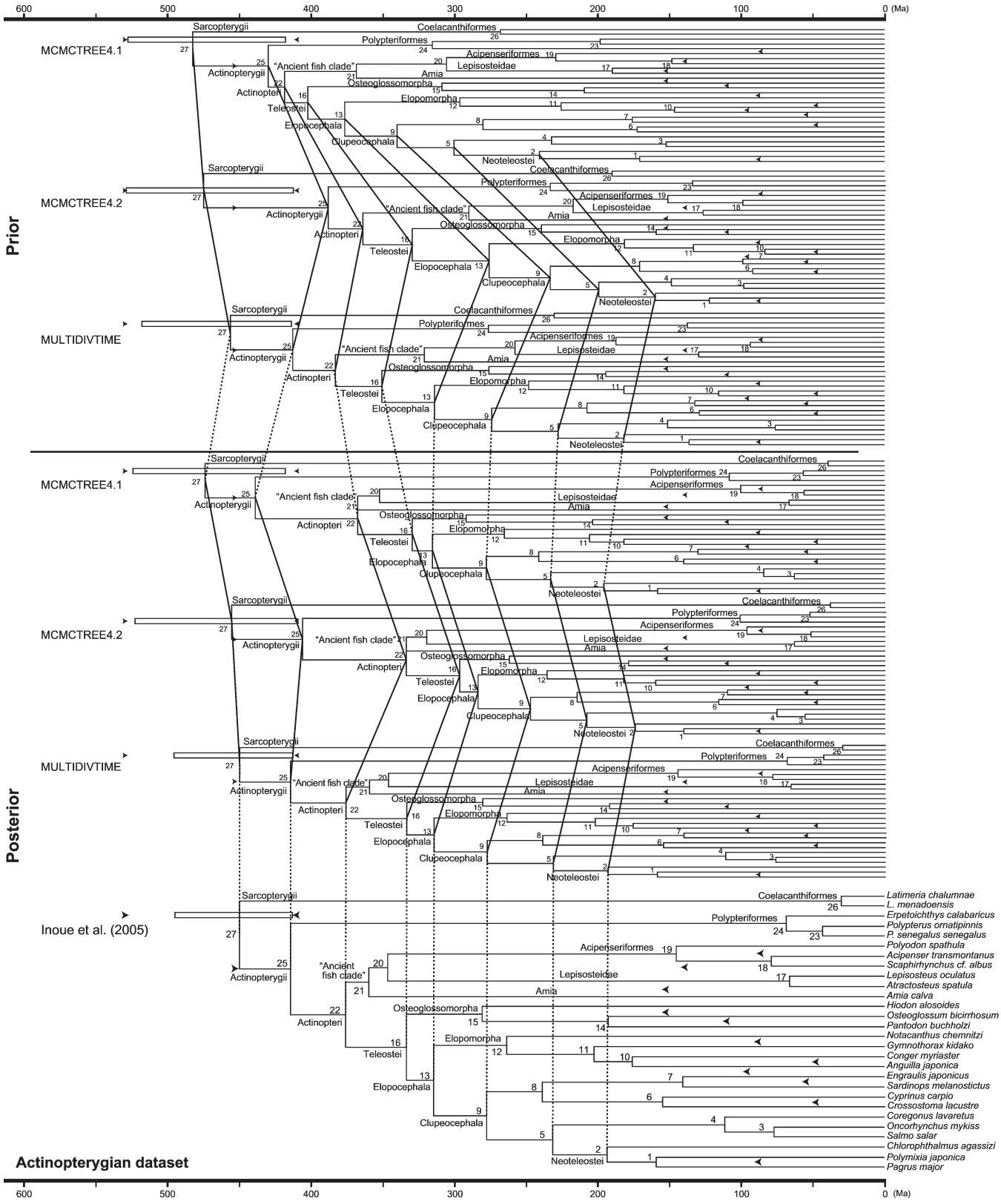

FIGURE 6. The rooted tree for the actinopterygian data set, showing fossil calibrations, and the prior and posterior means of node ages estimated using MCMCTREE and MULTIDIVTIME.

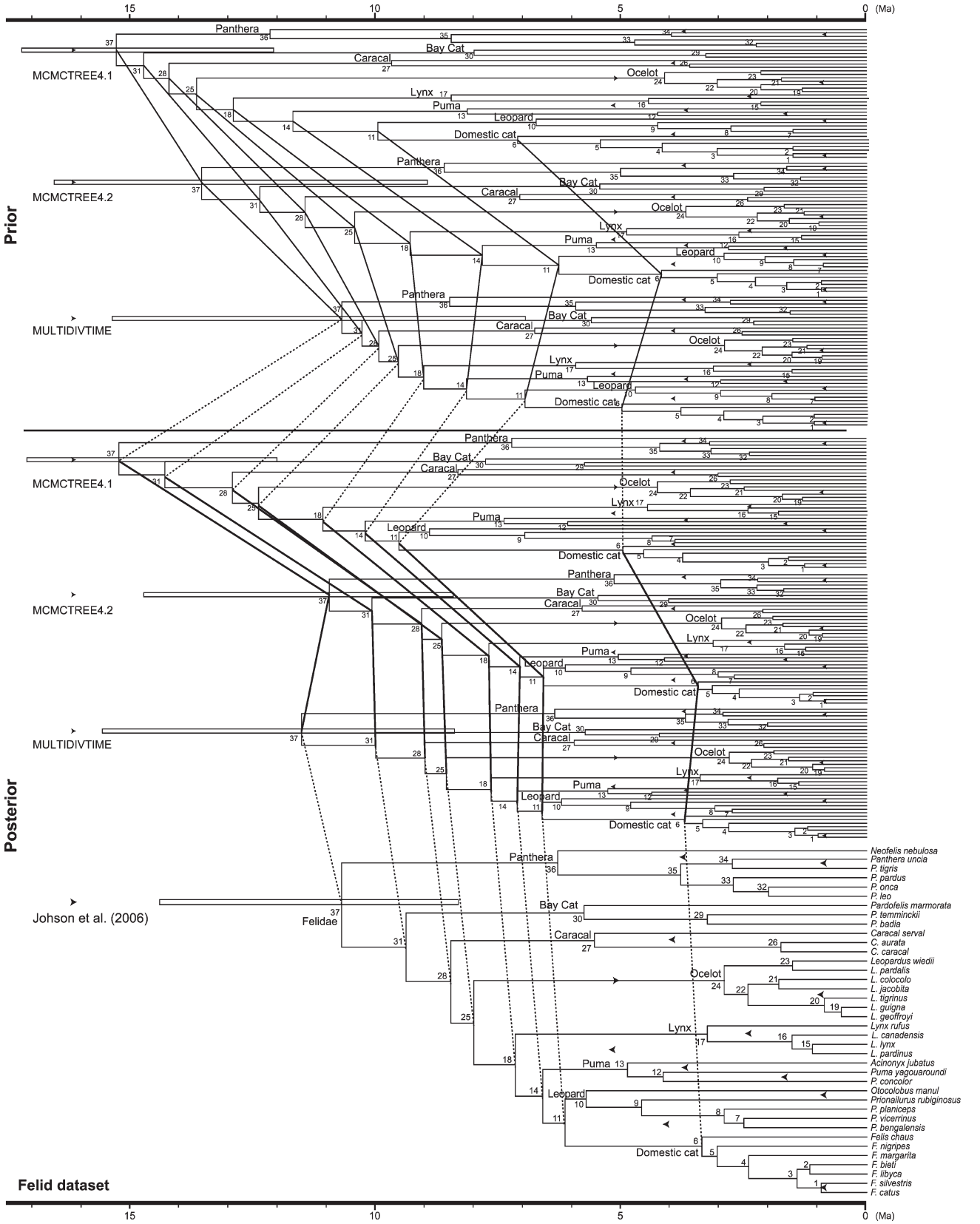

FIGURE 7. The rooted tree for the felid data set, showing fossil calibrations, and the priors and posterior means of node ages estimated using MCMCTREE and MULTIDIVTIME.

TABLE 3. Gamma priors used in analysis of the 3 data sets (1 time unit = 100 Myr)

Variable	Frog		Fish		Cat	
	MULTIDIVTIME	MCMCTREE	MULTIDIVTIME	MCMCTREE	MULTIDIVTIME	MCMCTREE
Root age (rttime, rttimesd)	4.2, 2.1	$G(4, 0.95)$	4.5, 2.25	None	0.1, 0.05	None
Substitution rate of the root node (rtrate, rtratesd) or rgene_gamma	0.068, 0.068	$G(1, 14.7)$	0.05, 0.05	$G(1, 20)$	0.126, 0.126	$G(1, 7.94)$
Rate-drift parameter ν or σ^2 (brownmean, brownsd) or sigma2_gamma	0.24, 0.24	$G(1, 4.17)$	0.22, 0.22	$G(1, 4.5)$	9.28, 9.28	$G(1, 0.11)$

Notes: MULTIDIVTIME specifies the gamma prior by the mean (m) and standard deviation (s), whereas MCMCTREE uses the shape (α) and scale (β) parameters, that is, $G(\alpha, \beta)$. These are related as follows: $\alpha = (m/s)^2$, $\beta = m/s^2$. Whenever possible, the same gamma priors are used in the 2 programs. MULTIDIVTIME always requires a gamma prior on the root age, whereas MCMCTREE can use it only if no fossil calibration exists on the root. For the fish and felid data sets, a gamma prior is used for the root age in MULTIDIVTIME but not in MCMCTREE as there are fossil calibrations on the root.

noted in the other 2 data sets and in the analysis of the 4-species example.

The huge differences in the prior between MCMCTREE 4.1 and MULTIDIVTIME persisted in the posterior, so that the posterior mean ages estimated by MCMCTREE 4.1 were nearly proportionally older than those from MULTIDIVTIME, as noted in the analysis of Rannala and Yang (2007). Posterior time estimates from MCMCTREE 4.2 were very similar to those from MULTIDIVTIME. The root age is the most different, with posterior means to be 11.0 Ma (8.4, 14.8) and 11.6 Ma (8.5, 15.6) from the 2 programs, although the 95% CIs were wide and overlapped considerably. It is interesting to note that the prior mean of root age from MCMCTREE 4.2 was older but the posterior mean was younger than corresponding means from MULTIDIVTIME.

It may be noted that our approximate likelihood calculation using MCMCTREE 4.1 produced similar time estimates to those obtained using the exact calculation (Rannala and Yang 2007). Those and other results (not shown) suggest that the normal approximation is quite reliable, at least for such large data sets.

The MCMCTREE 4.2 estimates were slightly older than those obtained by Johnson et al. (2006) using MULTIDIVTIME. For example, the divergence of Panthera from the other felids (Node 37) was dated to 11.0 Ma (8.4, 14.8) in our analysis and to 10.8 Ma (8.4, 14.5) by Johnson et al. The separation between leopard and domestic cat (Node 11) was dated to 6.6 Myr (5.1, 8.9) in our analysis and to 6.2 Ma (4.8, 8.6) by Johnson et al. These authors suggested that modern felid arose in Asia with the divergence of the *Panthera* lineage (Eurasia) at 10.8 Ma (8.4, 14.5) (Node 37) and the divergence of the Bay Cat lineage (Eurasia) at 9.4 Ma (7.4, 12.8) (Node 31). Based on the estimated time of the Ocelot lineage at 8.1 Ma (6.3, 11.0) (Node 25), they suggested that the common ancestor to 5 felid lineages: Ocelot (America), *Lynx* (America), *Puma* (America), *Leopard Cat* (Eurasia), and *Domestic Cat* (Eurasia), crossed the Bering land bridge to North America for the first time, at 8.5–8.0 Ma. Our time estimation for the separation for the Ocelot lineage was 8.7 Ma (6.7, 11.6), older than the geology-based time estimate for the Bering land bridge but the 95% CIs overlap.

The felid data set showed larger differences in posterior time estimates among methods than the other 2 data sets. This appears to reflect the imprecise fossil calibrations in the felid data set, as revealed by the infinite-sites plots (see below).

Infinite-Sites Plots

The "infinite-sites" theory developed by Yang and Rannala (2006) predicts that when the amount of sequence data approaches infinity, the posterior means of times and the 95% CIs for different nodes will fall on a straight line. The theory is general and applies to all current methods of Bayesian divergence time estimation, including that of Thorne et al. (1998). Figure 8 shows the infinite-sites plots for the 3 data sets in which the 95% CI widths of node ages are plotted against the posterior means.

The amphibian data set is the smallest among the 3 data sets analyzed. The weak correlation ($R^2 = 0.52$ by MCMCTREE and 0.33 by MULTIDIVTIME) indicates that the sequence data are far from saturation (Figure 8a,a′). The slope of 0.44 (or 0.40 by MULTIDIVTIME) is a measure of fossil precision and means that every 1 Myr of divergence time adds 0.44 Myr of uncertainty in the posterior estimate (or adds 0.40 Myr to the 95% CI interval). The outlier in the plots corresponds to Node 50 in the tree of Figure 4, which has joint fossil bounds, leading to a much narrower posterior CI than for other nodes.

In the larger actinopterygian data set, which consists of almost the whole mitochondrial genome, the correlation ($R^2 = 0.75$ by MCMCTREE or 0.66 by MULTIDIVTIME) is much stronger (Fig. 8b,b′). The smaller slope (0.34 by MCMCTREE or 0.26 by MULTIDIVTIME) means that the fossils are slightly more informative than in the amphibian data set, with every 1 Myr of divergence adding only 0.34 Myr of uncertainty in the posterior CI. The scatter plot indicates that old nodes were estimated more precisely relatively than young nodes because the only 2 maximum bounds are on old nodes (Node 27 [the root] and Node 25) and the only joint bounds are at the root (Fig. 6).

In the felid data set, the linear regression is nearly perfect, with $R^2 = 0.98$ by both programs (Fig. 8c,c′),

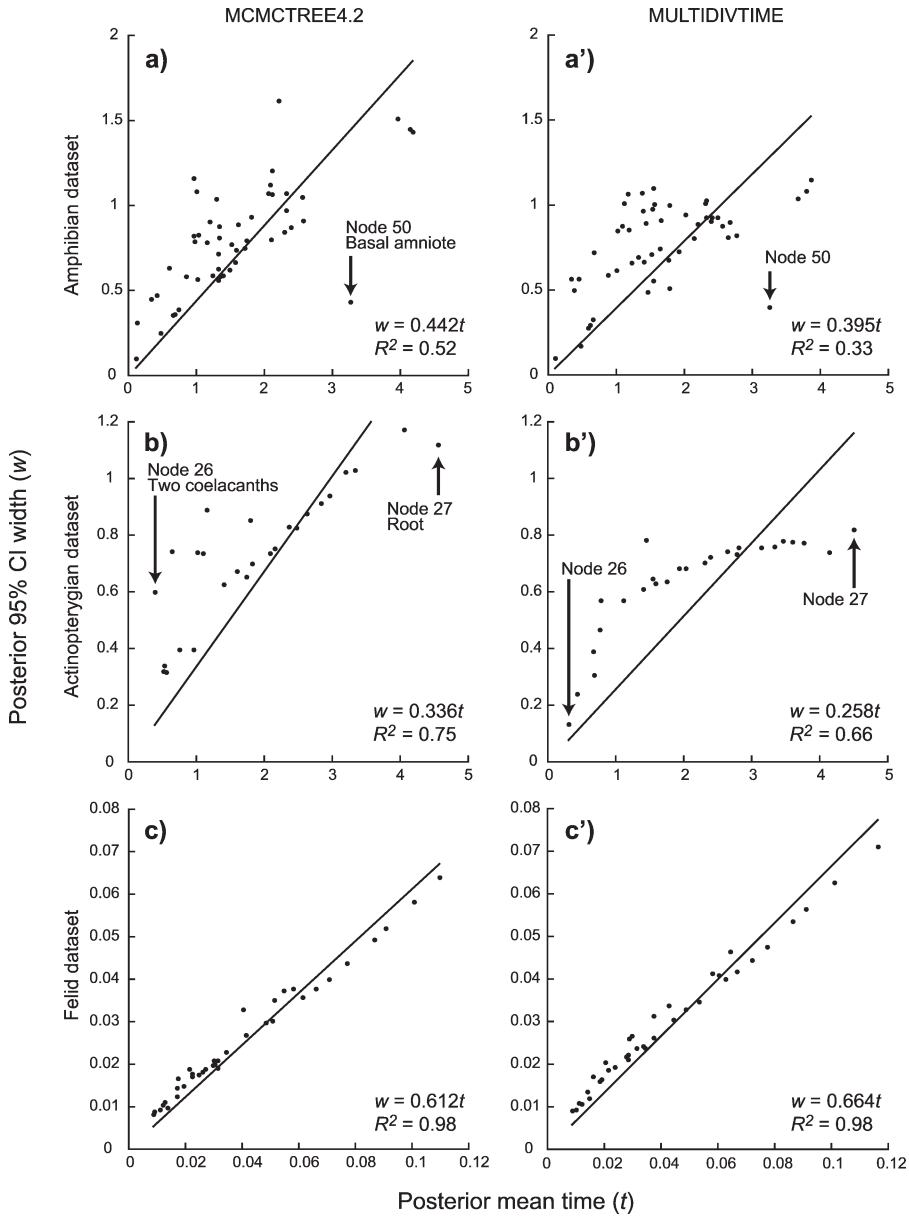

FIGURE 8. The infinite-sites plot for the 3 data sets. Each point corresponds to one internal node on the tree. The x-axis is the posterior mean of the node age, whereas the y-axis is the 95% posterior CI width, that is, the difference of the 97.5% and 2.5% limits. The time unit is 100 Myr. The posterior means of the node ages were used to draw the trees in Figures 5, 6, and 7.

indicating that the amount of sequence data (at ~20 Kb) had nearly reached saturation, and adding more sequence data are unlikely to improve the precision of posterior time estimates. The fossil calibrations are the most imprecise among the 3 data sets, indicated by the large slopes (0.61 by MCMCTREE or 0.66 by MULTIDIVTIME).

DISCUSSION

Factors Affecting Molecular Dating

We analyzed the 3 data sets to evaluate the impact of several other factors, such as the approximate and exact likelihood calculations, the substitution model, the amount of rate variation among lineages assumed in the prior, and so forth. For each data set, we changed only one aspect of the major analysis. Some of the results of this sensitivity analysis are shown in supplementary Figures S4–S6 for the 3 data sets, respectively. For example, we ran the analysis using the JC69 model (Jukes and Cantor 1969) instead of F84+Γ_5. The posterior estimates were very similar between the 2 models. This appears to be due to the use of multiple fossil calibrations in the analysis. It was previously noted that with the use of one single calibration, nonproportional underestimation of sequence distances by simplistic models can lead to systematically biased time estimates (e.g., Yang 1996).

We then examined the impact of the prior on the rate-drift parameter σ^2 (Rannala and Yang 2007), by multiplying parameter α in the prior $G(\alpha, \beta)$ by 0.1 or 10, so that both the mean and the variance of the prior distribution are reduced or increased by 10. In all 3 data sets, increasing the mean σ^2 in the prior (so that the prior assumes more variable rates among lineages) led to younger node ages. The effect is particularly dramatic for the actinopterygian data set, presumably because the large σ^2 allowed large changes in the rate across branches. Reducing the mean σ^2 by 10-fold had minimal impact in the 3 data sets (see supplementary Figs. S4–S6).

In sum, in our analysis of the 3 data sets, the exact and approximate likelihood calculation, the substitution model, and the 2 models of rate drift (results not shown) had minimal impact on the posterior time estimation. The prior on the rate-drift parameter σ^2 was more important. By far the most important factor in our analysis was the fossil calibrations or the different ways of representing them.

The Importance of Fossil Calibrations

We found the impact of the strategy for incorporating the same minimum and maximum bounds in the 2 programs to be surprising. It is important to note that the uncertainties in the posterior and the impact of the prior will not disappear with the addition of sequence data. The infinite-sites theory (Yang and Rannala 2006; Rannala and Yang 2007) predicts that when the sequence length approaches infinity, the posterior distribution will become 1D, so that the posterior means (and other quantities such as the standard deviations or the 95% limits) of node ages will become proportional across nodes on the tree, but the absolute times have to be resolved by the prior. The fossil record can provide only uncertain calibrations with no node age known with certainty, so that the posterior time estimates will always involve uncertainties. The consistency of Bayesian estimation in the conventional setting does not apply to the dating problem. Because of the importance of the prior, we recommend that all molecular dating analyses should carefully assess and report the priors on times and rates.

The sensitivity of the posterior time estimates to different strategies for representing minimum and maximum fossil bounds underscores the critical importance of fossil calibrations to molecular dating and highlights an urgent need for research into ways of summarizing the fossil data to provide useful calibrations. Currently, most bounds derived from the fossil record are minimum bounds, but these alone are insufficient for effective calibration. Confidence intervals on fossil distribution data have been advocated as a means of determining the true time of origin of a species in question (Marshall 1990, 2008). However, they suffer from the difficulty of distinguishing the earliest members of divergent lineages from members of the ancestral lineage before the split (Donoghue and Benton 2007).

In this paper, we implement a more flexible and potentially more realistic distribution to represent minimum bounds based on a soft-truncated Cauchy distribution with 2 parameters p and c. We used the values $p = 0.1$ and $c = 1$ in the analysis of the 3 data sets. However, we stress that those values are very unlikely to represent all fossil minimum bounds well. Ideally, different values for p and c should be used for different minimum bounds, chosen to reflect the differing confidence in the degree to which paleontological minima reflect lineage divergence times. This will require researchers to revisit the fossil data on which the calibrations are based. For example, variation in the amount of sediments representative of different physical and biotic environments through geological time has been identified as the principal bias in the record of fossil biodiversity (Raup 1972; Smith and McGowan 2007). Terrestrial sediments are essentially unrepresented in the Ordovician System (McGowan and Smith 2008), so it is no surprise that the earliest unequivocal records of terrestrial animals and plants are from the ensuing Silurian System (Labandeira 2005). Consideration of such large biases in the fossil record may provide a means for constructing probabilistic descriptions of clade divergence times. We leave such work to future research. Nevertheless, the method implemented here allows one to use flexible distributions to represent the information in the fossil record.

Our main objectives in this paper have been to assess the impact of the prior and other factors on Bayesian estimation of divergence times. In a recent study, Lepage et al. (2007) used the Bayes factor to compare the fit to the sequence data of different models for generating priors on times and rates. Their analysis did not use any calibrations, so that it is unclear how relevant the results are to practical dating analysis. Dating species divergences without any calibration does not appear to be a very meaningful exercise. We suggest that the sensitivity of time estimation to the prior is a more important question than the fit of the prior to the sequence data. We also suggest that the appropriateness of the time prior and of the fossil calibrations should ideally be assessed by a careful appraisal of previous data, especially the fossil record, rather than by the fit of the prior to the sequence data being analyzed.

SUPPLEMENTARY MATERIAL

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

FUNDING

This study is supported by a grant from the Biotechnological and Biological Sciences Research Council (UK) to Z.Y. and P.D.

ACKNOWLEDGMENTS

We thank Jeff Thorne, Bodil Svennblad, Céline Poux and Michael E. Steiper for many constructive comments.

REFERENCES

Aris-Brosou S. 2007. Dating phylogenies with hybrid local molecular clocks. PLoS One. 2:e879.

Arratia G., Schultze H.P. 1999. Mesozoic fishes from Chile. In: Arratia G., Schultze H.P., editors. Mesozoic fishes 2: systematics and the fossil record. München (Germany): Pfeil. p. 265–334.

Barnett R., Barnes I., Phillips M.J., Martin L.D., Harington C.R., Leonard J.A., Cooper A. 2005. Evolution of the extinct Sabretooths and the American cheetah-like cat. Curr. Biol. 15:R589–R590.

Benton M.J., Donoghue P.C.J. 2007. Paleontological evidence to date the tree of life. Mol. Biol. Evol. 24:26–53.

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

Drummond A.J., Ho S.Y.W., Phillips M.J., Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4:e88.

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

Hedges S.B., Kumar S. 2004. Precision of molecular time estimates. Trends Genet. 20:242–247.

Ho S.Y.W. 2007. Calibrating molecular estimates of substitution rates and divergence times in birds. J. Avian Biol. 38:409–414.

Hurley I.A., Mueller R.L., Dunn K.A., Schmidt E.J., Friedman M., Ho R.K., Prince V.E., Yang Z., Thomas M.G., Coates M.I. 2007. A new time-scale for ray-finned fish evolution. Proc. R. Soc. Lond. B. Biol. Sci. 274:489–498.

Inoue J.G., Miya M., Venkatesh B., Nishida M. 2005. The mitochondrial genome of Indonesian coelacanth Latimeria menadoensis (*Sarcopterygii: Coelacanthiformes*) and divergence time estimation between the two coelacanths. Gene. 349:227–235.

Johnson W.E., Eizirik E., Pecon-Slattery J., Murphy W.J., Antunes A., Teeling E., O'Brien S.J. 2006. The late Miocene radiation of modern Felidae: a genetic assessment. Science. 311:73–77.

Jukes T.H., Cantor C.R. 1969. Evolution of protein molecules. In: Munro H.N., editor. Mammalian protein metabolism. New York: Academic Press. p. 21–123.

Kishino H., Thorne J.L., Bruno W.J. 2001. Performance of a divergence time estimation method under a probabilistic model of rate evolution. Mol. Biol. Evol. 18:352–361.

Labandeira C.C. 2005. Invasion of the continents: cyanobacterial crusts to tree-inhabiting arthropods. Trends Ecol. Evol. 20:253–262.

Lepage T., Bryant D., Philippe H., Lartillot N. 2007. A general comparison of relaxed molecular clock models. Mol. Biol. Evol. 24:2669–2680.

Marjanovic D., Laurin M. 2007. Fossils, molecules, divergence times, and the origin of Lissamphibians. Syst. Biol. 56:369–388.

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 calibration points. Am. Nat. 171:726–742.

McGowan A.J., Smith A.B. 2008. Are global Phanerozoic marine diversity curves truly global? A study of the relationship between regional rock records and global Phanerozoic marine diversity. Paleobiology. 34:80–103.

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

Rannala B., Yang Z. 2007. Inferring speciation times under an episodic molecular clock. Syst. Biol. 56:453–466.

Raup D.M. 1972. Taxonomic diversity during the Phanerozoic. Science. 177:1065–1071.

Reisz R.R., Muller J. 2004. Molecular timescales and the fossil record: a paleontological perspective. Trends Genet. 20:237–241.

Ruta M., Coates M.I. 2007. Dates, nodes and character conflict: addressing the lissamphibian origin problem. J. Syst. Palaeontol. 5: 67–122.

San Mauro D., Vences M., Alcobendas M., Zardoya R., Meyer A. 2005. Initial diversification of living amphibians predated the breakup of *Pangaea*. Am. Nat. 165:590–599.

Sanderson M.J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol. Biol. Evol. 14:1218–1232.

Sanderson M.J. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol. Biol. Evol. 19:101–109.

Smith A.B., McGowan A.J. 2007. The shape of the Phanerozoic marine palaeodiversity curve: how much can be predicted from the sedimentary rock record of Western Europe? Paleontology. 50:765–774.

Smith A.B., Peterson K.J. 2002. Dating the time of origin of major clades: molecular clocks and the fossil record. Ann. Rev. Earth Planet. Sci. 30:65–88.

Steiper M.E., Young N.M. 2008. Timing primate evolution: lessons from the discordance between molecular and paleontological estimates. Evol. Anthropol. 17:179–188.

Thorne J.L., Kishino H. 2005. Estimation of divergence times from molecular sequence data. In: Nielsen R., editor. Statistical methods in molecular evolution. New York: Springer-Verlag. p. 233–256.

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.

Yang Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J. Mol. Evol. 39:306–314.

Yang Z. 1996. Among-site rate variation and its impact on phylogenetic analyses. Trends Ecol. Evol. 11:367–372.

Yang Z. 2004. A heuristic rate smoothing procedure for maximum likelihood estimation of species divergence times. Acta Zool. Sin. 50:645–656.

Yang Z. 2006. Computational molecular evolution. Oxford: Oxford University Press.

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.

Yoder A.D., Yang Z. 2000. Estimation of primate speciation dates using local molecular clocks. Mol. Biol. Evol. 17:1081–1090.

Zuckerkandl E., Pauling L. 1965. Evolutionary divergence and convergence in proteins. In: Bryson V., Vogel H.J., editors. Evolving genes and proteins. New York: Academic Press. p. 97–166.