Mol Phylogenet Evol

Comparison of different strategies for using fossil calibrations to generate the time prior in Bayesian molecular clock dating

Jose Barba-Montoya[1], Mario dos Reis[1, 2] and Ziheng Yang[1]

1. *Department of Genetics, Evolution and Environment, University College London, Gower Street, London, WC1E 6BT, UK.*

2. *School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London, E1 4NS, UK.*

Correspondence: *z.yang@ucl.ac.uk*

Abstract

Fossil calibrations are the utmost source of information for resolving the distances between molecular sequences into estimates of absolute times and absolute rates in molecular clock dating analysis. The quality of calibrations is thus expected to have a major impact on divergence time estimates even if a huge amount of molecular data is available. In Bayesian molecular clock dating, fossil calibration information is incorporated in the analysis through the prior on divergence times (the time prior). Here, we evaluate three strategies for converting fossil calibrations (in the form of minimum- and maximum-age bounds) into the prior on times, which differ according to whether they borrow information from the maximum age of ancestral nodes and minimum age of descendent nodes to form constraints for any given node on the phylogeny. We study a simple example that is analytically tractable, and analyse two real datasets (one of 10 primate species and another of 48 seed plant species) using three Bayesian dating programs: MCMCTree, MrBayes and BEAST2. We examine how different calibration strategies, the birth-death process, and automatic truncation (to enforce the constraint that ancestral nodes are older than descendent nodes) interact to determine the time prior. In general, truncation has a great impact on calibrations so that the effective priors on the calibration node ages after the truncation can be very different from the user-specified calibration densities. The different strategies for generating the effective prior also had considerable impact, leading to very different marginal effective priors. Arbitrary parameters used to implement minimum-bound calibrations were found to have a strong impact upon the prior and posterior of the divergence times. Our results highlight the importance of inspecting the joint time prior used by the dating program before any Bayesian dating analysis.

Keywords: Bayesian inference, molecular clock dating, divergence times, fossil calibration, time prior.

1. Introduction

Bayesian inference has become the methodology of choice for molecular clock dating of species divergences because it provides a natural framework for incorporating different sources of information (e.g., from fossils and molecules) (dos Reis, et al. 2016). In a Bayesian dating analysis, one would ideally summarize the relevant prior evidence about species divergence times (say, from the fossil record, geological events, etc.) in a multidimensional joint prior of ages for all nodes on the phylogeny (called the time prior). However, specifying high-dimensional priors with complex correlation structures is a notoriously difficult task, and furthermore, our knowledge of the fossil evidence and of how it informs the species divergence times is very imprecise. The current practice is for the paleontologist to specify minimum- and maximum-age constraints on certain nodes on the tree based on the fossil evidence (Thorne, et al. 1998; Kishino, et al. 2001; Benton, et al. 2009; Ho and Phillips 2009). Such *user-specified* fossil calibrations are then used by the Bayesian dating program to construct the time prior, with the distribution of the ages of non-calibration nodes supplanted by a branching-process model (e.g., a birth-death process) (Yang and Rannala 2006). The user-specified calibration densities are assigned to single nodes on the tree and often do not satisfy the requirement that any ancestral node should be older than its descendants, and thus the dating software must 'truncate' the calibration densities to satisfy this constraint. We refer to the resulting prior of node ages used by the dating software as the *effective prior*, and this may be very different from the original user-specified calibration densities (Inoue, et al. 2010; Warnock, et al. 2015). Furthermore, Bayesian dating programs such as MultiDivTime (Thorne, et al. 1998), MCMCTree (Yang 2007), BEAST2 (Bouckaert, et al. 2014) and MrBayes (Ronquist, et al. 2012b) use different procedures to combine calibration densities with the birth-death process model to generate the time prior, so that different programs may produce very different time priors from the same user-specified fossil calibrations (Inoue, et al. 2010).

Thus, users of dating software are encouraged to run the Markov Chain Monte Carlo (MCMC) algorithm without molecular data to generate the time prior used by the program and to inspect it to ensure that it is a reasonable representation of the fossil evidence. A cross-validation method for assessing the quality of calibrations, based on the consistency between fossils and between fossils and molecules, has also been proposed (Near, et al. 2005). This was noted to sometimes lead to the selection of calibrations of poor reliability (Marshall 2008; Benton, et al. 2009; Warnock, et al. 2015). The problem appears to be partly due to the fact that fossil-calibration constraints provided by the paleontologist are "over-interpreted" by the Bayesian dating program. For example, when fossil evidence suggests that the age of a clade is between 50Ma and 100Ma, the dating software may incorporate that information by assigning a uniform distribution, $t \sim U(50, 100)$, implying, for example, $P\{50 < t < 60\} = P\{90 < t < 100\}$. Such probabilistic statements about the true age may not be intended by the paleontologist. However minimum and maximum bounds alone, in the form of 50

$< t <$ 100, are insufficient to permit a Bayesian dating analysis: a full statistical distribution for the true age has to be specified.

The way that the fossil-based bounds on node ages are converted into statistical distributions in a dating analysis may thus have an important impact on the posterior time estimates. Consider the unbalanced 5-species phylogeny of Figure 1. Suppose that fossil evidence suggests that the age of node 4 should be at least 10 Myrs, while the age of the root is at most 100 Ma, with $t_4 > 10$ and $t_1 < 100$ (Figure 1). Three simple strategies appear possible to construct the calibration densities. In strategy 1 (st1), we apply a minimum-bound calibration on t_4, by using a decay function from 10 Ma to ∞ (such as the offset-exponential), while the age of the root may be assigned a uniform distribution $t_1 \sim U(0, 100)$. Ages of the non-calibration nodes (t_2 and t_3) have densities specified by the birth-death process. In strategy 2 (st2), we propagate the minimum and maximum bounds to all calibration nodes: the root acquires the minimum bound from node 4, while node 4 inherits the maximum age of the root, so that both nodes have joint bounds: $t_4 \sim U(10, 100)$, and $t_1 \sim U(10, 100)$. In strategy 3 (st3), we propagate the minimum and maximum bounds to all nodes on the phylogeny, so that $t_i \sim U(10, 100)$ for $i = 1, 2, 3$ and 4. In all three strategies, the dating program will automatically apply a truncation so that $t_4 < t_3 < t_2 < t_1$. Different programs use different procedures to perform the truncation and to combine the calibration densities with the branching process model (Inoue, et al. 2010). As a result the three strategies should lead to different time priors, and the different programs will also differ even for the same strategy. For simple cases, it is possible to calculate analytically the resulting marginal priors for the node ages after truncation. However, for large phylogenies with dozens of fossil calibrations, analytical calculation is impossible, and the user needs to estimate the prior by running the Bayesian MCMC program without sequence data.

Here we study how the different calibration strategies affect the time prior and the posterior time estimates. We examine two approaches used by Bayesian dating programs to combine calibration densities with the branching process to form a prior density for all node ages (the time prior): the *conditional construction* used by MCMCTree (Yang and Rannala 2006) and the *multiplicative construction* used by BEAST (Bouckaert, et al. 2014) and MrBayes (Ronquist, et al. 2012b) (see Heled and Drummond 2015). We study a simple example that is analytically tractable, and then analyze two real datasets: one of 10 primate species, and another of 48 seed plant species. We show that the different calibration strategies as well as truncation have significant impacts on the time prior and the resulting posterior time estimates. We discuss the implications of our results and give recommendations for the construction of reasonable time priors.

2. Material and Methods

2.1. Fossil calibrations and the time prior

We consider three types of constraints on a node age based on the fossil evidence: minimum-age bound, maximum-age bound, and joint (maximum- and minimum-age) bounds (Figure 2). These are implemented in different Bayesian dating programs using different approaches.

Minimum-age calibrations (Figure 2a). In MCMCTree, a minimum bound is represented using a truncated Cauchy distribution, denoted L(t_L, p, c, p_L) (Inoue, et al. 2010). Here t_L is the minimum age bound, p determines how far the mode of the distribution is from the minimum, c determines how sharply the distribution decays to zero, and p_L is the left tail probability (i.e. the probability that the minimum bound is violated). Smaller values of p and c give a more concentrated calibration density, with a higher probability that the true age is close to the minimum age. For example, $p = 0.1$ means the mode of the distribution is at $(1 + p)t_L = 1.1t_L$. Here we used $p = 0.1$, $c = 0.1$, and $p_L = 0.01$.

In MrBayes and BEAST2, minimum bounds are represented using an offset-exponential distribution (Heled and Drummond 2012; Ronquist, et al. 2012b; Bouckaert, et al. 2014). If y has an exponential distribution with rate parameter θ or mean $1/\theta$, then $t = y + t_L$ has an offset-exponential distribution with parameters θ and t_L, with mean $\theta^{-1} + t_L$. A large θ means that the true age is likely to be close to t_L. In this study, we used $\theta = 10/t_L$, so that the mean of the distribution is 1.1 t_L.

Maximum-age calibrations (Figure 2b). Maximum bounds are represented by a uniform distribution U ~ (0, t_U), where t_U is the maximum age. Bounds are hard (with zero probability for any ages outside the interval) in BEAST2 and MrBayes, and soft in MCMCTree, with p_U to be the error probability that the bound is violated.

Joint (minimum- and maximum-age) calibrations (Figure 2c). Joint bounds are represented by a uniform distribution U(t_L, t_U) in all three programs. Again, bounds are hard in BEAST2 and MrBayes, and soft in MCMCTree, which assigns p_L and p_U as the error probabilities for violations of the bounds (Yang and Rannala 2006). We use $p_L = 0.01$ and $p_U = 0.05$.

2.2 Calibration strategies to generate the time prior

The calibration strategies are different ways of generating the effective prior given the fossil bounds on the calibration nodes on the phylogeny. We consider three strategies.

Calibration strategy st1: Minimum and maximum constraints were applied to calibration nodes as given, without propagating onto other nodes.

Calibration strategy st2: Minimum and maximum constraints are propagated onto all calibration nodes, so that every calibration node has joint minimum and maximum bounds, represented by a uniform distribution. In other words, if a calibration node lacks a minimum bound, the minimum bound of its oldest descendent node is used, and if a calibration node lacks a maximum bound, the maximum bound of its youngest ancestor is used.

Calibration strategy st3. This is like st2 but minimum and maximum bounds are propagated onto all interior nodes on the phylogeny, so that every node has a pair of joint bounds. Note that in st2, every calibration node has a pair of bounds while in st3, every interior node has a pair of bounds.

The rooted tree topology was fixed in all analyses. This is a requirement for MCMCTree and we did the same for BEAST2 and MrBayes to avoid the confounding effects of alternative phylogenies. A constraint on the root is required in MCMCTree (Yang and Rannala, 2006) and MrBayes (Ronquist, et al. 2012b). BEAST2 does not require a constraint on the root, one or more calibrations on internal nodes may be sufficient (Heled and Drummond 2012, 2015).

The Bayesian analysis requires a prior on the ages of all nodes on the tree. The birth-death branching process is used to provide the prior distribution for the non-calibration nodes, which is combined with the effective prior for the calibration nodes after the truncation, to generate the time prior. Two procedures have been used to achieve this in the current dating programs.

In MCMCTree, the so-called *conditional construction* is used (Yang and Rannala 2006). Let t_C be the ages of the calibration nodes, and $t_{\bar{C}}$ be the ages of the non-calibration nodes. In the example of Figure 1, $t_C = \{t_1, t_4\}$ while $t_{\bar{C}} = \{t_2, t_3\}$. The conditional construction gives the density of all node ages as

$$f(t_C, t_{\bar{C}}) = f(t_C) \cdot f_{\mathrm{BDS}}(t_{\bar{C}} \mid t_C), \quad (1)$$

where $f(t_C)$ is the effective prior on the ages of the calibration nodes, given by the user-specified calibration densities after truncation, while $f_{\mathrm{BDS}}(t_{\bar{C}} \mid t_C)$ is the conditional density of the non-calibration nodes given the calibration node ages, specified by the birth-death-sampling (BDS) process (Yang and Rannala 1997).

Both BEAST2 and MrBayes use the so-called *multiplicative construction*, in which the birth-death process density for all node ages is multiplied with the densities for the calibration nodes to generate the time prior (Heled and Drummond, 2012; Heled and Drummond, 2015).

$$f(t_C, t_{\bar{C}}) \propto f(t_C) \cdot f_{\mathrm{BDS}}(t_C, t_{\bar{C}}) = f(t_C) \cdot \left[f_{\mathrm{BDS}}(t_C) \cdot f_{\mathrm{BDS}}(t_{\bar{C}} \mid t_C) \right], \quad (2)$$

Here t_C is the density of node ages for the calibration nodes based on the user-specified calibration densities (with suitable truncation so that ancestors are older than the descendents), and $f_{\mathrm{BDS}}(t_C)$ is the marginal density of the node ages for the calibration nodes as specified by the birth-death-sampling process, while $f_{\mathrm{BDS}}(t_{\bar{C}} \mid t_C)$ is the conditional density of the ages of the non-calibration nodes given the ages of the calibration nodes as specified by the birth-death-sampling process. As the density of t_C occurs twice in eq. 2, this density is mathematically incorrect and "does not follow the rules of probability calculus" (Heled and Drummond 2012). Here we treat both constructions as heuristic methods for converting user-specified constraints into the time prior.

2.3. Analysis of a simple example with five species

We use a simple and analytically tractable case of five species (Figure 1) to explore the different approaches to constructing the time prior (the prior for all node ages). Nodes 1 and 4 are calibration nodes, with the fossil constraints $t_1 < 100$ Myrs and $t_4 > 10$, while t_2 and t_3 are non-calibration nodes, for which the densities are provided by a branching process such as the birth-death-sampling process. As the birth-death process has no beginning and no ending, it is necessary to condition the process either on the time of origin, or the age of the root, or on the number of sampled extant species (N) (Yang and Rannala 1997). Here we condition on both the number of sampled extant species and the age of the root, as in Yang and Rannala (1997). Let λ be the per-lineage birth (speciation) rate, μ the per-lineage death (extinction) rate, and ρ the sampling fraction. We fix the parameters in the model at $\lambda = \mu = 1$ and $\rho = 0$, so that the ages of the nonroot nodes are order statistics from a uniform kernel (Yang and Rannala 1997). In other words, given the root age t_1, node ages t_2, t_3 and t_4 can be generated by sampling three independent random variables from U(0, t_1) and then ordering them. The joint distribution is

$$f_{\mathrm{BDS}}(t_4, t_3, t_2 \,|\, t_1) = \frac{3!}{t_1^3}, \quad 0 < t_4 < t_3 < t_2 < t_1. \tag{3}$$

This is equivalent to the Dirichlet time prior used by Thorne et al. (1998).

Calibration strategy 1 (st1). We consider the *conditional* construction used by MCMCTree first (Yang and Rannala 2006). The calibration density for t_1 (the root age) is the uniform distribution

$$f_C(t_1) = 1/t_U, \quad 0 < t_1 < t_U, \tag{4}$$

with $t_U = 100$, while that for t_4 is the offset-exponential

$$f_C(t_4) = \theta \mathrm{e}^{-\theta(t_4 - t_L)}, \quad t_L < t_4 < \infty, \tag{5}$$

where $t_L = 10$ and we choose $\theta = 1/t_L$ so that the mean is $2t_L = 20$ Ma.

Multiplying those user-specified calibration densities and removing the unfeasible region (where $t_4 > t_1$) by truncation leads to the effective prior used by the dating program

$$f_C(t_1, t_4) = \frac{1}{k_1} \theta \mathrm{e}^{-\theta(t_4 - t_L)} \cdot \frac{1}{t_U}, \quad t_L < t_4 < t_1 < t_U, \tag{6}$$

where $k_1 = \int_L^{t_U} \int_{t_L}^{t_1} \theta \mathrm{e}^{-\theta(t_4 - t_L)} \cdot \frac{1}{t_U} \mathrm{d}t_4 \mathrm{d}t_1 = 0.80001$ is a normalizing constant, to ensure that $f_C(t_1, t_4)$ integrates to 1.

Under the birth-death-sampling process model, with $\lambda = \mu = 1$ and $\rho = 0$, the joint density for t_2 and t_3, conditioned on the calibration node ages (t_1 and t_4), is given by the fact that t_2 and t_3 are order statistics from U(t_4, t_1), with density

$$f_{\mathrm{BDS}}(t_2, t_3 \,|\, t_1, t_4) = 2/(t_1 - t_4)^2, \quad t_4 < t_3 < t_2 < t_1. \tag{7}$$

The effective time prior or the joint density for all node ages is thus

$$f(t_1,t_2,t_3,t_4)=f_C(t_1,t_4)f_{\mathrm{BDS}}(t_2,t_3|t_1,t_4)=\tfrac{1}{k_1}\theta\,\mathrm{e}^{-\theta(t_4-t_L)}\cdot\tfrac{1}{t_U}\times\tfrac{2}{(t_1-t_4)^2},\ t_L<t_4<t_3<t_2<t_1<t_U, \tag{8}$$

where k_1 is the normalizing constant defined below equation (6).

The marginal prior densities of the calibration node ages (t_1 and t_4) can be obtained by integration.

$$f(t_4)=\int_{t_4}^{t_U} f_C(t_1,t_4)\,\mathrm{d}t_1=\tfrac{1}{k_1 t_U}\theta\,\mathrm{e}^{-\theta(t_4-t_L)}\cdot(t_U-t_4),\ t_L<t_4<t_U, \tag{9}$$

$$f(t_1)=\int_{t_L}^{t_1} f_C(t_1,t_4)\,\mathrm{d}t_4=\tfrac{1}{k_1 t_U}\left[1-\mathrm{e}^{-\theta(t_1-t_L)}\right],\ t_L<t_1<t_U, \tag{10}$$

Note that eq. 9 can also be derived by integrating out t_1, t_2, t_3 from $f(t_1, t_2, t_3, t_4)$, and eq. 10 can be derived by integrating out t_2, t_3, t_4 from $f(t_1, t_2, t_3, t_4)$. Figure 3a (st1) shows the user-specified calibration densities and the effective (marginal) priors after the truncation.

In the *multiplicative* construction used by BEAST and MrBayes, the densities for the calibration nodes of eqs. 4 and 5 are multiplied with the joint density of the ages of all non-root nodes from the birth-death-sampling process (eq. 3) to give

$$\begin{aligned} f(t_1,t_2,t_3,t_4) &\neq \tfrac{1}{k_2}\times f_C(t_1)\cdot f_C(t_4)\cdot f_{\mathrm{BDS}}(t_2,t_3,t_4\,|\,t_1) \\ &= \tfrac{1}{k_2}\times\tfrac{1}{t_U}\cdot\theta\,\mathrm{e}^{-\theta(t_4-t_L)}\cdot\tfrac{3!}{t_1^3}, \qquad t_L<t_4<t_3<t_2<t_1, \end{aligned} \tag{11}$$

where $k_2=\int_{t_L}^{t_U}\int_{t_L}^{t_1}\int_{t_L}^{t_2}\int_{t_L}^{t_3}\tfrac{1}{t_U}\cdot\theta\,\mathrm{e}^{-\theta(t_4-t_L)}\cdot\tfrac{3!}{t_1^3}\,\mathrm{d}t_4\mathrm{d}t_3\mathrm{d}t_2\mathrm{d}t_1=0.0174371$ is a normalizing constant. Note that eq. 11 does not make mathematical sense as two different densities occur for t_4, one in $f_C(t_4)$ and the other in $f_{\mathrm{BDS}}(t_2, t_3, t_4\,|\,t_1)$. The marginal (effective) priors for the calibration node ages (t_1 and t_4) can be obtained by integration

$$\begin{aligned} f(t_1)&=\int_{t_L}^{t_1}\int_{t_L}^{t_2}\int_{t_L}^{t_3} f(t_1,t_2,t_3,t_4)\,\mathrm{d}t_4\mathrm{d}t_3\mathrm{d}t_2=\tfrac{3!}{k_2 t_U t_1^3}\left[\tfrac{(t_1-t_L)^2}{2}+\tfrac{t_L-t_1}{\theta}+\tfrac{1-e^{-\theta(t_1-t_L)}}{\theta^2}\right], \\ f(t_4)&=\int_{t_4}^{t_U}\int_{t_3}^{t_U}\int_{t_2}^{t_U} f(t_1,t_2,t_3,t_4)\,\mathrm{d}t_1\mathrm{d}t_2\mathrm{d}t_3=\tfrac{3\theta e^{-\theta(t_4-t_L)}}{k_2 t_U}\left[\log\tfrac{t_U}{t_4}-1.5+\tfrac{2t_4}{t_U}-\tfrac{t_4^2}{2t_U{}^2}\right], \end{aligned} \tag{12}$$

with $t_L < t_1 < t_U$ and $t_L < t_4 < t_U$. Figure 3b (st1) shows the user-specified calibration densities and the effective (marginal) priors after the truncation.

Calibration strategy 2 (st2). The minimum and maximum bounds are propagated onto all calibration nodes so that the calibration densities are

$$\begin{aligned} f_C(t_1)&=1/(t_U-t_L), \quad t_L<t_1<t_U, \\ f_C(t_4)&=1/(t_U-t_L), \quad t_L<t_4<t_U. \end{aligned} \tag{13}$$

We first consider the *conditional* construction. After truncation, the effective joint prior for t_1 and t_4 becomes, in contrast to eq. 6,

$$f_C(t_1, t_4)=2/(t_U-t_L)^2,\ \ t_L<t_4<t_1<t_U. \tag{14}$$

This is multiplied with the birth-death-sampling process density for the non-calibration nodes of eq. 7 to give the time prior as

$$f(t_1,t_2,t_3,t_4) = f_C(t_1,t_4)\cdot f_{\mathrm{BDS}}(t_2,t_3 \mid t_1,t_4) = \tfrac{2}{(t_U-t_L)^2}\times\tfrac{2}{(t_1-t_4)^2},\ t_L<t_4<t_3<t_2<t_1<t_U. \quad (15)$$

The marginal densities for the calibration node ages are

$$\begin{aligned} f(t_1) &= \int_{t_L}^{t_1} f_C(t_1,t_4)\,\mathrm{d}t_4 = \tfrac{2}{(t_U-t_L)^2}\times(t_1-t_L),\ \ t_L<t_1<t_U, \\ f(t_4) &= \int_{t_4}^{t_U} f_C(t_1,t_4)\,\mathrm{d}t_1 = \tfrac{2}{(t_U-t_L)^2}\times(t_U-t_4),\ \ t_L<t_4<t_U. \end{aligned} \quad (16)$$

Figure 3a (st2) shows the densities.

With the *multiplicative* construction, the time prior is given by multiplying the calibration densities (eq. 13) with the birth-death-sampling density for the noncalibration nodes (eq. 3) and then applying truncation

$$f(t_1,t_2,t_3,t_4) = \tfrac{1}{k_3}\times\tfrac{1}{(t_U-t_L)^2}\cdot\tfrac{3!}{t_1^3},\quad t_L<t_4<t_3<t_2<t_1<t_U, \quad (17)$$

where $k_3 = \int_{t_L}^{t_U}\int_{t_L}^{t_1}\int_{t_L}^{t_2}\int_{t_L}^{t_3}\tfrac{1}{(t_U-t_L)^2}\cdot\tfrac{3!}{t_1^3}\,\mathrm{d}t_4\mathrm{d}t_3\mathrm{d}t_2\mathrm{d}t_1 = 0.00530524$ is a normalizing constant, calculated numerically. The marginal (effective) priors for the calibration nodes (t_1 and t_4) are then

$$\begin{aligned} f(t_1) &= \int_{t_L}^{t_1}\int_{t_L}^{t_2}\int_{t_L}^{t_3} f(t_1,t_2,t_3,t_4)\,\mathrm{d}t_4\mathrm{d}t_3\mathrm{d}t_2 = \tfrac{3}{k_3(t_U-t_L)^2}\left[\log\tfrac{t_U}{t_4} - 1.5 + \tfrac{2t_4}{t_U} - \tfrac{t_4^2}{2t_U^2}\right], \\ f(t_4) &= \int_{t_4}^{t_U}\int_{t_3}^{t_U}\int_{t_2}^{t_U} f(t_1,t_2,t_3,t_4)\,\mathrm{d}t_1\mathrm{d}t_2\mathrm{d}t_3 = \tfrac{(t_1-t_L)^3}{k_3 t_1^3 (t_U-t_L)^2}, \end{aligned} \quad (18)$$

with $t_L < t_1 < t_U$ and $t_L < t_4 < t_U$. Figure 3b (st2) shows the user-specified calibration densities and the effective (marginal) priors after the truncation.

Calibration strategy 3 (st3). The minimum and maximum bounds are propagated onto all nodes on the phylogeny, so that every node has joint bounds: $f_C(t_i) = 1/(t_U - t_L)$, $t_L < t_i < t_U$, for $i = 1, 2, 3, 4$. With the *conditional* construction, the birth-death-sampling model plays no role in the construction of the time prior since all nodes have calibration information. After truncation, the effective time prior is

$$f(t_1,t_2,t_3,t_4) = \tfrac{4!}{(t_U-t_L)^4},\quad t_L<t_4<t_3<t_2<t_1<t_U. \quad (19)$$

Since t_4 is the smallest of four independent and identically distributed (i.i.d.) random variables and t_1 is the largest, their marginal densities are given by the distribution of order statistics

$$\begin{aligned} f(t_1) &= 4\cdot\left(\tfrac{t_1-t_L}{t_U-t_L}\right)^3\cdot\tfrac{1}{t_U-t_L},\ \ t_L<t_1<t_U, \\ f(t_4) &= 4\cdot\left(\tfrac{t_U-t_4}{t_U-t_L}\right)^3\cdot\tfrac{1}{t_U-t_L},\ \ t_L<t_4<t_U. \end{aligned} \quad (20)$$

Figure 3a (st3) shows the densities. Truncation now has a strong effect.

With the *multiplicative* construction, two options seem possible. The first is to ignore the birth-death process density since all the node ages have calibration with this strategy. This is then equivalent to the conditional construction of MCMCTree. The second is to multiply the calibration densities (eq. 19) with the birth-death-sampling density of eq. (3), followed by a truncation to give

$$f(t_1, t_2, t_3, t_4) = \frac{1}{k_4} \frac{4!}{(t_U - t_L)^4} \times \frac{3!}{t_1^3}, \quad t_L < t_4 < t_3 < t_2 < t_1 < t_U, \tag{21}$$

where the normalizing constant $k_4 = \int_{t_L}^{t_U} \int_{t_L}^{t_1} \int_{t_L}^{t_2} \int_{t_L}^{t_3} \frac{4!}{(t_U - t_L)^4} \times \frac{3!}{t_1^3} \mathrm{d}t_4 \mathrm{d}t_3 \mathrm{d}t_2 \mathrm{d}t_1 = 0.000015719$. The marginal priors for t_1 and t_4 are then

$$\begin{aligned} f(t_1) &= \int_{t_L}^{t_1} \int_{t_L}^{t_2} \int_{t_L}^{t_3} f(t_1, t_2, t_3, t_4) \mathrm{d}t_4 \mathrm{d}t_3 \mathrm{d}t_2 = \frac{4! \times 3}{k_4 (t_U - t_L)^4} \left[\log \frac{t_U}{t_4} - 1.5 + \frac{2t_4}{t_U} - \frac{2t_4^2}{2t_U^2} \right], \\ f(t_4) &= \int_{t_4}^{t_U} \int_{t_3}^{t_U} \int_{t_2}^{t_U} f(t_1, t_2, t_3, t_4) \mathrm{d}t_1 \mathrm{d}t_2 \mathrm{d}t_3 = \frac{4! \times (t_1 - t_L)^3}{k_4 t_1^3 (t_U - t_L)^4}, \end{aligned} \tag{22}$$

with $t_L < t_1 < t_U$ and $t_L < t_4 < t_U$. Figure 3b (st3) shows the user-specified calibration densities and the effective (marginal) priors after the truncation.

The calibration densities and the effective time priors generated by the conditional and the multiplicative constructions using the three calibration strategies are plotted in Figure 3. From Figure 3a it is apparent that with the conditional construction strategy st1 generates marginal priors that are closest to the original calibration densities. This is because the youngest node is calibrated with an offset-exponential distribution with a relatively short tail, and so truncation between the two calibration densities is minimal. In Strategy st2 the youngest node inherits the maximum age constraint from the root. This strategy avoids the choice of arbitrary parameters in the Cauchy or shifted-exponential calibrations. In this case truncation is more severe, and the marginal prior densities differ substantially from the calibration densities. In strategy st3, the inclusion of two additional calibration densities for t_3 and t_2 increases the truncation effect, and the result is that the marginal priors on t_4 and t_1 are pushed apart. The multiplicative construction is shown in Figure 3b. Strategy st1 generates marginal priors that are closest to the original calibration densities, while truncation has a major impact in strategies st2 and st3, so that the marginal prior densities differ substantially from the calibration. St2 and st3 generate nearly identical prior densities. Overall Figure 3 shows that the conditional and the multiplicative constructions, as well as the different calibration strategies, generate very different effective time priors.

2.4. Analysis of the primate dataset

We used eight mitochondrial coding genes (Cyt B, CO1, CO2, CO3, ND2, ND3, ND4 and ND4L) and the mitochondrial 12S and 16S ribosomal RNA (rRNA) genes from nine primate species and an outgroup (*Tupaia belangeri*) (Figure 4a) (GenBank accession numbers in Table S1). We partitioned the data into three partitions: (1) 1st and 2nd codon positions; (2) 3rd codon positions and (3) rRNA

genes. The final alignment had 9,361 base pairs. The data were analyzed using the three dating programs (MCMCTree, BEAST2, and MrBayes), under the independent-rates model to construct the prior of the rates. The time unit is set at 100 Myrs. The sequence likelihood was calculated under the HKY+Γ_5 substitution model (Hasegawa, et al. 1985; Yang 1994), with separate substitution-rate parameters assigned and estimated for each partition.

There are nine fossil calibrations (Table 1) (dos Reis, et al. 2012), five of which are joint minimum and maximum bounds, while the other four are minimum bounds only. We implemented calibration strategies st1 and st2 in the programs MCMCTree, BEAST2, and MrBayes. As all nine interior nodes have calibration information, st3 is equivalent to st2. Bounds are soft in MCMCTree, and hard in BEAST2 and MrBayes. Minimum bounds are implemented using the truncated Cauchy distribution in MCMCTree and the offset-exponential distribution in BEAST2 and MrBayes.

In MCMCTree, the parameters of the birth-death-sampling process are fixed at $\lambda = \mu = 1$, and $\rho = 0$. These specify a uniform kernel. The independent-rates model (IR) assumes that the rates for branches are independent variables from the lognormal distribution, specified by the mean of the rate (η) and the variance of the log rate σ^2 (which determines the extent of rate variation across branches) (Rannala and Yang 2007). The mean rate is assigned a gamma prior $\eta \sim G(2, 2)$ with mean 2/2 = 1.0 substitutions per site per time unit (100MY) or 10^{-8} substitutions per site per year, and the rate drift parameter is assigned another gamma prior, $\sigma^2 \sim G(1,10)$, with mean 0.1.

Both BEAST2 and MrBayes assign hyperpriors to implement the birth-death-sampling model: the net diversification rate $\lambda - \mu \sim U(0, 1)$ and the relative extinction rate $\mu/\lambda \sim U(0, 1)$ (Stadler 2010; Hohna, et al. 2011). In MrBayes the sampling probability (ρ) is fixed at 0.02.

In BEAST2 we specified a *Relaxed Clock Log Normal* (ucld) model, which assumes that the substitution rates for branches are independent variables from a lognormal distribution (Drummond, et al. 2006). The lognormal distribution is parametrized using the mean and the standard deviation. The mean (*ucldMean.c*) was assigned a gamma hyperprior G(2, 0.5) with mean 1.0, and the standard deviation (*ucldStdev.c*) was assigned a gamma hyperprior G(2, 0.05) with mean 0.1.

In MrBayes we used the *Independent Gamma Rate* (IGR) model in where the rates for branches are independent variables from a gamma distribution (Lepage, et al. 2007). The gamma model is parametrized using two parameters: the mean and variance. The mean is assigned a lognormal hyperprior LN(–0.125, 0.5), with the mean $\exp\{-0.125 + 0.5^2/2\} = 1.0$. The variance (*Igrvarpr*) is assigned an exponential prior with mean 0.1.

The MCMC sampling settings were determined through pilot runs and differed among the programs. We ran each program at least twice, and checked for convergence by comparing the posterior mean estimates between runs and by plotting the time series traces of the samples. We then merged the samples from the runs before summarizing the posterior. For MCMCTree, two runs were performed, each consisting of 2×10^6 iterations after a burn-in of 4×10^4 iterations and sampling every

200, resulting in a total of 2×10^4 samples from the two runs. For MrBayes, two runs were performed, each consisting of 2×10^6 iterations, sampling every 100, with the burn-in set to 25% of samples, resulting in a total of 3×10^4 samples from the two runs. For BEAST2 we performed three runs, each consisting of 10^7 iterations, sampling every 1000. The burn-in was set to 30% of samples, resulting in a total of 21,000 samples from all three runs.

2.5. Analysis of the seed plant dataset

We used five plastid genes (*atpB, matK, NdhF, rbcL*, and *rps4*) and two nuclear RNA genes (*18s* and *26s*) for 48 seed plant species (GenBank accession numbers in Table S2) from Barba-Montoya et al. (Submitted). The tree topology of Figure 4b is fixed. The sequence alignment had 13,211 base pairs, with 26% missing data. We treated the data as three partitions: (1) 1st and 2nd codon positions for plastid genes; (2) 3rd positions for plastid genes and (3) nuclear RNA genes. The data were analyzed using the three programs (MCMCTree, BEAST2, and MrBayes), with similar settings as in the analysis of the primate dataset, but some modifications were necessary to accommodate the differences in the time scale and in the rate. The sequence likelihood was calculated under the HKY+Γ_5 substitution model (Hasegawa, et al. 1985; Yang 1994), with separate substitution-rate parameters assigned and estimated for each partition. In MCMCTree the approximate likelihood method (Thorne, et al. 1998; dos Reis and Yang 2011) is used to calculate the sequence likelihood, using the maximum likelihood estimates of branch lengths and the Hessian matrix. In BEAST2 and MrBayes the sequence likelihood was calculated exactly.

There are 15 fossil calibrations on the tree (Figure 4b) (Barba-Montoya, et al. Submitted) Among them seven are joint minimum and maximum bounds and eight are minimum bounds (Table 2). The time unit is set to 100 Myrs. The calibration information is implemented in the three programs using the three strategies as described earlier.

In MCMCTree, the parameters of the birth-death-sampling process are fixed at $\lambda = \mu = 1$, and $\rho = 0$. We used the independent-rates (IR) model, with the overall rate assigned a gamma hyperprior G(2, 30) with mean 2/30 = 0.067 substitutions per site per 100MY, and with the variance of the log-rate assigned a gamma hyperprior $\sigma^2 \sim$ G(2, 20) with mean 0.1. Two runs were performed, each consisting of 10^6 iterations after a burn-in of 40,000 iterations and sampling every 200. The combined sample of 10,000 samples was used to summarize.

In the BEAST2 and MrBayes analyses, hyperpriors are assigned to parameters in the birth-death-sampling model: $\lambda - \mu \sim$ U(0, 1) and $\mu/\lambda \sim$ U(0, 1) (Stadler, 2010; Höhna et al., 2012). In MrBayes, the sampling probability (ρ) is fixed at 0.0002.

In BEAST2 we specified the ucld model. The mean of the lognormal (*ucldMean.c*) was assigned a gamma hyperprior G(2, 0.0335) with mean 0.067, and the standard deviation of the lognormal (*ucldStdev.c*) was assigned a gamma hyperprior G(2, 0.05) with mean 0.1. Three runs were

performed, each consisting of 10^7 iterations, sampling every 1000. The burn-in was set to 30% of samples, resulting in a total of 21,000 samples from the posterior from the three runs.

In MrBayes we used the *Independent Gamma Rate* (IGR) model. The mean of the gamma was assigned a lognormal hyperprior LN(−2.79, 0.5^2), with the mean $\exp\{-2.79 + 0.5^2/2\} = 0.07$, and the variance of the gamma is assigned an exponential hyperprior with mean 0.1. Four runs were performed, each consisting of 1.5×10^6 iterations, sampling every 100. The burn-in was set to 33.3% of samples, resulting in a total of 4×10^4 samples from all four runs.

3. Results

3.1. Analysis of a simple example with five species

Figure 5 shows the results from analysing this example using the three different dating programs. In MCMCTree (Figure 5a) the calibration density used for t_4 in strategy st1, is the Cauchy distribution (shifted-exponential) with parameters t_L =10, p = 0.2, c = 0.5 and p_L = 0.0001. We fix the parameters in the birth-death-sampling model at $\lambda = \mu = 1$ and $\rho = 0$ in all strategies. The prior densities generated by the three calibration strategies using MCMCTree (Figure 5a, st1, st2, st3) are almost identical to those from the conditional construction (Figure 3a, st1, st2, st3).

To examine the implementation in MrBayes and BEAST (Figures 5b, c, and d) we fix the parameters in the birth-death-sampling model at $\lambda = \mu = 1$ and $\rho = 0$. To avoid numerical problems, we used λ= 1.001, μ = 0.999 and ρ = 0.0001. In MrBayes the net diversification rate λ - μ is fixed at 0.002, the relative extinction rate μ /λ is fixed at 0.998 and the sampling probability (ρ) is fixed at 0.0001. In BEAST1 and BEAST2 we use for the net diversification rate λ - μ a uniform distribution U(0.00199,0.00201) and for the relative extinction rate μ /λ U(0.99799,0.99801). In BEAST1 we use U(0.000099,0.000101) for the sampling probability (ρ). None of these programs generated identical results to the multiplicative construction. The prior densities generated by MrBayes and BEAST1 were similar but not identical. Precise reasons for the discrepancies between the analytical example, BEAST1 and MrBayes are unknown. One possible reason is that BEAST1 and MrBayes may not be conditioning the birth-death-sampling age density on both root (t_1) and the number of sampled species (N). Here we emphasize the large differences in the prior generated by the conditional and multiplicative constructions and the priors from the three calibration strategies.

3.2. Analysis of the primate dataset

The calibration densities and the effective time priors generated by the three programs using calibration strategies st1 and st2 are plotted in Figures 6 and 7. The posterior distributions of divergence times are shown in Figures 7 and 8.

First, we note that with both st1 and st2, the user-specified calibration densities are very different from the marginal densities for the node ages in the effective time prior. This difference is mainly

386 caused by the truncation to enforce the constraint that ancestors are older than descendants. In
387 particular, the root age assigned a pair of bounds represented by the uniform distribution, and in the
388 time prior, the density is pushed towards the maximum. Node 18 is a descendent to many other
389 interior nodes but is ancestral to none, so that its density is pushed towards the minimum. The patterns
390 for other nodes are more complex. Second, strategy st2, which uses uniform bounds for all interior
391 nodes, show much greater truncation effect so that the user-specified calibration densities and the
392 marginal prior densities are even more different than under strategy st1. Third, the differences in the
393 prior of node ages are transferred to the differences in the posterior. For example, the prior favoured
394 much older age for the root under st2 than under st1 for all three programs (Figure 7a, b, c, node 11),
395 and this pattern persisted in the posterior.
396 Lastly, the three dating programs produced similar priors and posteriors (Figures 7 and 8),
397 although MCMCTree produced slightly older time estimates and wider intervals, especially for old
398 nodes such as the root.
399

400 ### *3.4. Analysis of seed plant dataset*

401 The calibration densities and the effective time priors generated by the three programs using the
402 three calibration strategies are plotted in Figures 9 and 10. The posterior distributions of node ages are
403 shown in Figures 10 and 11. We see similar patterns to those in the analysis of the primate dataset.
404 First, there are large differences between calibration densities specified by the user on one hand and
405 the (marginal) effective prior densities used by the dating software on the other. The difference is
406 particularly pronounced for nodes with wide joint bounds as the effective prior used by the dating
407 software is much narrower. Furthermore, truncation pushes the ages of old nodes such as the root
408 towards the user-specified maximum bound, or even outside the maximum bound in the case of
409 MCMCTREE, which allows bound violation due to its use of soft bounds (e.g., Figure 10a, b, and c,
410 node 49). At the same time, truncation has the effect of pushing the ages of younger nodes towards
411 the minimum bound in the prior (e.g., Figure 10a, b, and c, nodes 86, 88, and 89).
412 Second, as in the case of the primate dataset, the posterior of the node ages is sensitive to the
413 prior, and differences in the time prior are directly transferred to differences in the posterior. For
414 example, nodes 77 and 78 are older under st2 than under st1 and even older under st3, and exactly the
415 same trend persists in the posterior (Figure 10a, b, c). This pattern holds for all three dating programs.
416 Third, strategies st2 and st3 showed greater truncation effects so that the user-specified calibration
417 densities and the marginal prior densities are even more different than under st1. The large differences
418 in the priors of the three strategies persisted in the posterior. The time estimates tended to be older
419 under st2 than under st1, while st3 produced the oldest time estimates (Figures 10 and 11). For
420 example, the posterior mean estimated using st1 suggests that the eudicots (node 57) originated
421 around 155 Ma, but using st3 the posterior mean was around 195 Ma, with a difference of 40 Myrs.

The origin of monocots (node 78) was dated to ~136 Ma under st1 in BEAST2 and MrBayes and 150 Ma in MCMCTree, but using st3 the posterior mean for this node was around 190 Ma, with again a difference of ~40 Myrs. These differences in the posterior reflect the differences in the time prior generated under the three strategies (Figures 10 and 11).

Differences in posterior time estimates exist among the three dating programs, reflecting their different procedures to construct the time prior using the same fossil-calibration information (Figures 9 and 10). BEAST2 produced slightly younger estimates of root age (node 49) and MCMCTree produced narrower intervals than BEAST2 and MrBayes. The differences among the dating programs in both the prior and the posterior are the smallest for calibration strategy st3. This is because with st3 all nodes on the phylogeny were calibrated, so that the birth-death-sampling process plays no or little role in specifying the time prior.

4. Discussion

In a conventional Bayesian analysis, the posterior distribution of the parameters converge to a point mass (the true value of the parameter) and the prior becomes less and less important when the amount of data approaches infinity. Bayesian molecular clock dating is an unconventional estimation problem in the sense that such convergence to truth does not occur (Yang and Rannala 2006). If the amount of molecular data increases and the fossil calibration information is fixed, the posterior will not converge to a point or to the true node ages, and furthermore the prior will continue to exert a large impact on the posterior. Even if we use whole genomes in the dating analysis so that sequence distances and branch lengths are estimated with virtually no random sampling errors, fossil calibrations and the time prior constructed using the fossil calibrations will remain important to the posterior time estimates. The fundamental difficulty faced by the dating analysis is the confounding effect of time and rate in sequence comparisons: molecular data provide information about the genetic distances, and only fossil calibrations (or dated geological events) can resolve the distances into absolute times and absolute rates. The asymptotic dynamics of the dating problem has been characterized in the infinite-sites theory (Yang and Rannala 2006; Rannala and Yang 2007; dos Reis and Yang 2013; Zhu, et al. 2015).

Our analyses highlight the fact that the different dating programs such as MrBayes, BEAST, and MCMCTree use different and somewhat arbitrary procedures to construct the prior on divergence times and the resulting time priors may be very different among the programs even if exactly the same fossil calibration information is specified. We suggest that the user should be aware of such differences and always inspect the time prior by running the program without using the sequence data. The differences in the time prior may and may not have a large impact on the posterior time estimates, depending on the number, nature and locations of the fossil calibrations on the phylogeny, the amount of sequence data, and the seriousness of the violation of the clock, among other things. Similarly it is not possible to make a general recommendation as to which procedure is more appropriate for all

datasets (perhaps beyond the fact that the 'multiplicative construction' is a mathematical mistake and should be avoided). A procedure that produces time priors that better match the original calibration densities should make it easier for the user to summarize the fossil evidence, but we note that such a requirement may not be achievable because truncation can have a very large effect so that the effective priors are very different from the calibration densities whatever procedure is followed to convert the calibration densities to the effective time prior. In the future, we see probabilistic modeling and statistical analysis of fossil data (including both fossil presence/absence data and morphological measurements) as an important approach to summarizing the fossil evidence to generate distributions of divergence times for use as molecular clock calibrations (Tavaré, et al. 2002; Wilkinson, et al. 2011; Ronquist, et al. 2012a; Bracken-Grissom, et al. 2014; Heath, et al. 2014). For the present, we suggest that the palaeontologist should take a proactive role in constructing calibration densities, by making subjective judgments regarding the quality of the fossil and its placement on the phylogeny. We also encourage the use of the error probabilities in soft-bound calibrations as an approach to represent the uncertainties in the soft maximum bounds. It should be stressed that decisions will be made arbitrarily by the computer program if not subjectively by the palaeontologist. Given that in many cases the resulting time prior can be quite counterintuitively different to the calibration densities, we cannot emphasize enough how important it is for the user to explicitly calculate the time prior by running the MCMC analysis without data.

In this paper, we have focused on divergence time estimation when fossil calibration information is available on certain nodes on the tree, a procedure called *node calibration*. Recently *tip-calibration* methods have been developed, which analyze fossil data jointly with molecular data, in the so-called fossilized birth-death process model (Heath, et al. 2014; Zhang, et al. 2016). Morphological characters for both extant and extinct (fossil) species can be incorporated into a joint analysis with the molecular data for extant species (Ronquist, et al. 2012a; O'Reilly, et al. 2015).The dates for the fossil species provide the calibration information that resolves the morphological distances into absolute times and rates, which are propagated to the other nodes on the phylogeny represented by the molecular data. While the approach shows great promise, it has its own set of challenges (dos Reis et al., 2016; Ronquist et al., 2016). First, morphological characters, driven by natural selection and adaptation to environment and occasionally undergoing convergent evolution, rarely evolving in a clock-like fashion (Kimura 1983). Second, morphological characters may be strongly correlated. Thus current models (Lewis 2011), which ignore the correlation, are overstating the information content in the data. Third, without constraints on the interior nodes, the Bayesian dating analysis tends to be very sensitive to the birth-death-sampling process used to specify the time prior. Changing the parameters in the branching process may change the shape of the tree (reflected in the relative of internal versus external branch lengths), leading to drastically different posterior time estimates (Drummond and Stadler 2016; Ronquist, et al. 2016; Zhang, et al. 2016).We believe that both node calibrations and tip calibrations will have a major role to play in the foreseeable future (O'Reilly, et al. 2015).

Author contributions

M.d.R, and Z.Y. conceived the project and designed the analysis. J.B.-M. prepared the data sets and carried out the real data analysis. M.d.R carried out the theoretical 5-species analysis. All authors contributed to the interpretation of results and worked on the manuscript.

Acknowledgments

This research was funded by Biotechnology and Biosciences Research Council (UK) grant (BB/N000609/1) and Natural Environment Research Council grant (NE/N002067/1). J.B.-M. was supported by a CONACyT-Mexico and UCL scholarship. We thank Sebastian Höhna, Tanja Stadler and Chi Zhang for their help with implementations in MrBayes and BEAST.

Appendix A. Supplementary material

The molecular sequence alignments and five phylogenetic trees in Newick format with fossil calibrations have been deposited in Figshare: https://figshare.com/s/2d1ac059646932e74525 Supplementary data associated with this article can be found, in the online version, at….

References

Barba-Montoya, J., dos Reis, M., Schneider, H., Donoghue, P. C. J. and Yang, Z. Submitted. Constraining uncertainty in the timescale of angiosperm evolution and the veracity of a cretaceous terrestrial revolution. New Phytol.

Benton, M. J., Donoghue, P. C. J. and Asher, R. J. 2009. Calibrating and constraining molecular clocks. In: Hedges, S. B., Kumar, S., editors. The timetree of life. Oxford, UK: Oxford University Press. p. 35-86.

Bouckaert, R., J., Heled, D., Kuhnert, T., Vaughan, C. H., Wu, D. X., Suchard, M. A., Rambaut, A. and J., D. A. 2014. Beast 2: A software platform for bayesian evolutionary analysis. PLoS Comput Biol doi: 10.1371/journal.pcbi.1003537.g001

10.1371/journal.pcbi.1003537.g002

Bracken-Grissom, H. D., Ahyong, S. T., Wilkinson, R. D., Feldmann, R. M., Schweitzer, C. E., Breinholt, J. W., Bendall, M., Palero, F., Chan, T.-Y., Felder, D. L., Robles, R., Chu, K.-H., Tsang, L.-M., Kim, J. D., Martin, W. and A., C. K. 2014. The emergence of lobsters: Phylogenetic relationships, morphological evolution and divergence time comparisons of an ancient group (decapoda: Achelata, astacidea, glypheidea, polychelida). Syst Biol 63: 457-479.

dos Reis, M., Donoghue, P. C. and Yang, Z. 2016. Bayesian molecular clock dating of species divergences in the genomics era. Nat Rev Genet 17: 71-80. doi: 10.1038/nrg.2015.8

dos Reis, M., Inoue, J., Hasegawa, M., Asher, R. J., Donoghue, P. C. and Yang, Z. 2012. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. Proc Biol Sci 279: 3491-3500. doi: 10.1098/rspb.2012.0683

dos Reis, M. and Yang, Z. 2011. Approximate likelihood calculation on a phylogeny for bayesian estimation of divergence times. Mol Biol Evol 28: 2161-2172. doi: 10.1093/molbev/msr045

dos Reis, M. and Yang, Z. 2013. The unbearable uncertainty of bayesian divergence time estimation. J Syst Evol 51: 30-43. doi: 10.1111/j.1759-6831.2012.00236.x

Drummond, A. J., Ho, S. Y. W., Phillips, M. J. and Rambaut, A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biology 4: e88. doi: 10.1371/journal.pbio.0040088.g001

536 Drummond, A. J. and Stadler, T. 2016. Bayesian phylogenetic estimation of fossil ages. Philos Trans
537 R Soc Lond B Biol Sci 371. doi: 10.1098/rstb.2015.0129
538 Hasegawa, M., Kishino, H. and Yano, T. 1985. Dating the human-ape splitting by a molecular clock
539 of mitochondrial DNA J Mol Evol 22: 160-174.
540 Heath, T. A., Huelsenbeck, J. P. and Stadler, T. 2014. The fossilized birth-death process for coherent
541 calibration of divergence-time estimates. Proc Natl Acad Sci U S A 111: E2957-2966. doi:
542 10.1073/pnas.1319091111
543 Heled, J. and Drummond, A. J. 2012. Calibrated tree priors for relaxed phylogenetics and divergence
544 time estimation. Syst Biol 61: 138-149. doi: 10.1093/sysbio/syr087
545 Heled, J. and Drummond, A. J. 2015. Calibrated birth-death phylogenetic time-tree priors for
546 bayesian inference. Syst Biol 64: 369-383. doi: 10.1093/sysbio/syu089
547 Ho, S. Y. and Phillips, M. J. 2009. Accounting for calibration uncertainty in phylogenetic estimation
548 of evolutionary divergence times. Syst Biol 58: 367-380. doi: 10.1093/sysbio/syp035
549 Hohna, S., Stadler, T., Ronquist, F. and Britton, T. 2011. Inferring speciation and extinction rates
550 under different sampling schemes. Mol Biol Evol 28: 2577-2589. doi:
551 10.1093/molbev/msr095
552 Inoue, J., Donoghue, P. C. and Yang, Z. 2010. The impact of the representation of fossil calibrations
553 on bayesian estimation of species divergence times. Syst Biol 59: 74-89. doi:
554 10.1093/sysbio/syp078
555 Kimura, M. 1983. The neutral theory of molecular evolution. In. Molecular evolutionary rates
556 contrasted with phenotypic evolutionary rates: Cambride University Press. p. 55-97.
557 Kishino, H., Thorne, J. L. and Bruno, W. J. 2001. Performance of a divergence time estimation
558 method under a probabilistic model of rate evolution. Mol Biol Evol 18: 352–361.
559 Lepage, T., Bryant, D., Philippe, H. and Lartillot, N. 2007. A general comparison of relaxed
560 molecular clock models. Mol Biol Evol 24: 2669-2680. doi: 10.1093/molbev/msm193
561 Lewis, P. O. 2011. Alikelihood approach to estimating phylogeny from discrete morphological
562 character data. Syst Biol 50: 913-925.
563 Marshall, C. R. 2008. A simple method for bracketing absolute divergence times on molecular
564 phylogenies using multiple fossil calibration points. Am Nat 171: 726-742. doi:
565 10.1086/587523
566 Near, T. J., Bolnick, D. I. and Wainwright, P. 2005. Fossil calibrations and molecular divergence
567 time estimates in centrachid fishes (telostei: Centrarchidae). Evolution 59: 1768-1782. doi:
568 10.1554/05-030.1.s1
569 O'Reilly, J. E., dos Reis, M. and Donoghue, P. C. 2015. Dating tips for divergence-time estimation.
570 Trends Genet 31: 637-650. doi: 10.1016/j.tig.2015.08.001
571 Rannala, B. and Yang, Z. 2007. Inferring speciation times under an episodic molecular clock. Syst
572 Biol 56: 453-466. doi: 10.1080/10635150701420643
573 Ronquist, F., Klopfstein, S., Vilhelmsen, L., Schulmeister, S., Murray, D. L. and Rasnitsyn, A. P.
574 2012a. A total-evidence approach to dating with fossils, applied to the early radiation of the
575 hymenoptera. Syst Biol 61: 973-999. doi: 10.1093/sysbio/sys058
576 Ronquist, F., Lartillot, N. and Phillips, M. J. 2016. Closing the gap between rocks and clocks using
577 total-evidence dating. Philos Trans R Soc Lond B Biol Sci 371. doi: 10.1098/rstb.2015.0136
578 Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Hohna, S., Larget, B., Liu,
579 L., Suchard, M. A. and Huelsenbeck, J. P. 2012b. Mrbayes 3.2: Efficient bayesian
580 phylogenetic inference and model choice across a large model space. Syst Biol 61: 539-542.
581 doi: 10.1093/sysbio/sys029
582 Stadler, T. 2010. Sampling-through-time in birth-death trees. J Theor Biol 267: 396-404. doi:
583 10.1016/j.jtbi.2010.09.010
584 Tavaré, S., R., M. C., Will, O., Soligos, C. and D, M. R. 2002. Using the fossil record to estimate the
585 age of the last common ancestor of extant primates. Nature 416: 726-729.
586 Thorne, J. L., Kishino, H. and Painter, I. S. 1998. Estimating the rate of evolution of the rate of
587 molecular evolution. Mol Biol Evol 15: 1647-1657.
588 Warnock, R. C., Parham, J. F., Joyce, W. G., Lyson, T. R. and Donoghue, P. C. 2015. Calibration
589 uncertainty in molecular dating analyses: There is no substitute for the prior evaluation of
590 time priors. *Proc Biol Sci* 282: 20141013. doi: 10.1098/rspb.2014.1013

591 Wilkinson, R. D., Steiper, M. E., Soligo, C., Martin, R. D., Yang, Z. and Tavare, S. 2011. Dating
592 primate divergences through an integrated analysis of palaeontological and molecular data.
593 Syst Biol 60: 16-31. doi: 10.1093/sysbio/syq054
594 Yang, Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable
595 rates over sites: Approximate methods. J Mol Evol 39: 306-314.
596 Yang, Z. 2007. Paml 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol 24: 1586-
597 1591. doi: 10.1093/molbev/msm088
598 Yang, Z. and Rannala, B. 1997. Bayesian phylogenetic inference using DNA sequences: A markov
599 chain monte carlo method. Mol Biol Evol 14: 717-724.
600 Yang, Z. and Rannala, B. 2006. Bayesian estimation of species divergence times under a molecular
601 clock using multiple fossil calibrations with soft bounds. Mol Biol Evol 23: 212-226. doi:
602 10.1093/molbev/msj024
603 Zhang, C., Stadler, T., Klopfstein, S., Heath, T. A. and Ronquist, F. 2016. Total-evidence dating
604 under the fossilized birth-death process. Syst Biol 65.
605 Zhu, T., dos Reis, M. and Yang, Z. 2015. Characterization of the uncertainty of divergence time
606 estimation under relaxed molecular clock models using multiple loci. Syst Biol 64: 267-280.
607
608

Figure Legends

Figure 1. A five-species phylogeny used in the analytical example of fossil calibration strategies.

Figure 2. Probability densities for describing uncertainties in fossil calibrations: (a) soft minimum bound represented by a shifted-exponential distribution specified as $t_L = 20$, $p = 0.1$, $c = 0.1$, $p_L = 0.01$; (b) soft maximum bound specified as "$t_U = 80$, $p_R = 0.05$"; and (c) soft lower and upper bound, specified as "$t_L = 20$, $t_U = 80$, $p_L = 0.01$, $p_U = 0.05$". Black solid lines represent calibration densities. Red dashed lines represent a) minimum age (t_L,), b) maximum age (t_U) and c) both (t_L, t_U).

Figure 3. User-specified calibrations and effective priors for node ages t_1 and t_4 under three calibration strategies (st1, st2, st3) in a simple example of five species (Figure 1), generated using the (a) conditional and (b) the multiplicative construction. Dashed lines represent the user-specified calibration densities, while dotted lines represent the effective prior densities.

Figure 4. Phylogenies for a) 10 primate species, and b) 48 seed plant species. Calibration nodes are indicated by black solid circles.

Figure 5. User-specified calibrations and effective priors for node ages t_1 and t_4 under three calibration strategies (st1, st2, st3) in a simple example of five species (Figure. 1), generated using a) MCMCTree; b) MrBayes; c) BEAST1 and d) BEAST2. Dashed lines represent the user-specified calibration densities, while dotted lines represent the effective prior densities.

Figure 6. Means and 95% CIs in the time prior (the prior for node ages) on the primate phylogeny (Figure 5a) generated using calibration strategies st1 and st2 and three dating programs: MCMCTree, BEAST2 and MrBayes.

Figure 7. User-specified calibration densities (dashed lines), effective time priors (dotted lines), and the posterior (solid lines) for the primate dataset, under calibration strategies st1 (red) and st2 (black), implemented in MCMCTree, BEAST2 and MrBayes.

Figure 8. Timetrees showing posterior divergence time estimates for the primates. The branches are drawn to reflect the posterior means of node ages and the bars represent 95% HPD intervals. The dataset was analysed using MCMCTree, MrBayes amd BEAST2 under the independent-rates model, using calibration strategies st1 and st2.

Figure 9. Means and 95% CIs in the time prior for node ages on the seed plant phylogeny (Figure 5b) generated using three calibration strategies (st1-3) and three dating programs: MCMCTree, BEAST2 and MrBayes. Calibration nodes are highlighted in red.

Figure 10. User-specified calibration densities (dashed lines), effective time priors (dotted lines), and the posterior (solid lines) for the seed plant dataset, under calibration strategies st1 (red), st2 (black), and st3 (blue), implemented in MCMCTree, BEAST2 and MrBayes. Only the 15 calibration nodes are used in the plots.

Figure 11. Timetrees showing posterior divergence time estimates for major seed plant groups. The branches are drawn to reflect the posterior means of node ages and the bars represent 95% HPD intervals. The dataset was analysed using MCMCTree, MrBayes amd BEAST2 under the independent-rates model, using three calibration strategies: st1, st2, and st3.

Tables

Table 1. Primate fossil calibrations used in this study

Node	Clade	Minimum (Ma)	Maximum (Ma)
11	Scandentia-Primates	61.5 (†*Carpolestidae*)	130 (absence of placentals)
12	Primates (*Otolemur*-Human)	55.6 (†*Altiatlasius*)	—
13	Haplorhini (*Tarsius*-Human)	45 (†*Tarsius*)	—
14	Anthropoidea (*Callithrix*-Human)	33.7 (†*Catopithecus*)	—
15	Catarrhini (*Macaca*-Human)	23.5 (†*Proconsul*)	34 (absence of hominoids)
16	Hominidae (*Pongo*-Human)	11.2 (†*Sivapithecus*)	33.7 (absence of pongines)
17	Ponginae (*Gorilla-Pan*/Human)	7.25 (†*Chorapithecus*)	—
18	Homininae (*Pan*-Human)	5.7 (†*Orrorin*)	10 (absence of hominines)
19	Lorisoidea (*Otolemur- Microcebeus*)	33.7(†*Karanisia*)	55.6 (absence of strepsirrhines)

Note: All calibrations are derived from dos Reis et al. (2012). Fossil taxa are indicated by a dagger (†) before their names.

Table 2. Seed plant fossil calibrations used in this study

Node	Clade	Minimum divergence time (Ma)	Maximum divergence time (Ma)
49	Spermatophytes (*Ginkgo-Quercus*)	308.14 (†*Cordaites iowensis*)	365.63 (base of Vco zone which contains the first seeds)
50	Angiosperms (*Amborella-Quercus*)	125.9 (tricolpate pollen)	247.3 (sediments below the oldest occurrence of angiosperm like pollen which are devoid of such pollen)
57	Eudicots without Ceratophyllm (*Nandina-Quercus)*	119.6 (†*Hyrcantha decussata*)	—
65*	No name *(Arabidopsis-Quercus)*	82.8 (†*Paleoclusia chevaliery* and †*Dressiantha bicarpellata*)	127.2 (oldest potential age of tricolpate pollen)
70	Vitales (*Vitis-Leea*)	65.6 (†*Indovitis chitaleyae)*	—
76	Cornales (*Petalonix-Cornus*)	85.8 (†*Tylerianthus crossmanensis*)	—
77	Proteales (*Nelumbo-Platanus*)	107.59 (†*Sapindopsis variabilis*, †*Aquia brookensis* and †*Palatonocarpus brookensis*)	—
78	Monocots (*Acorus-Musa*)	112.6 (†*Liliacidites*)	—
84	Chloroanthaeles (*Cloranthus-Hedyosmum*)	92.8 (†*Pennipolis*)	—
86	No name (*Trimenia-Illicium*)	107.59 (†*Anacostia virginiensis*)	—
88	Cabombaceae (*Cabomba-Numphaea*)	111 (†*Pluricarpellatia peltata*)	—
89	Acrogymnospermae (*Ginkgo-Pinus*)	308.14 (†*Cordaties iowensis*)	365.7 (base of Vco zone which contains the first seeds)
90	Conifers (*Pinus-Metasequoia*)	147 (†*Rissikia media*)	312.38 (sediments bearing †*Cordaites iowensis*)
92	Gnetales (*Gnetum-Welwitschia*)	119.6 (†*Eoantha zherkihinii*)	312.38 (sediments bearing †*Cordaites iowensis*)
94	No name (*Ginkgo-Cycas*)	264.7 (†*Crossozamia*)	365.63 (base of Vco zone which contains the first seeds)

Note: Calibrations are derived from Barba-Montoya et al. (submitted) and (*) from Clarke et al. (2011). Fossil taxa are indicated by a dagger (†) before their names.

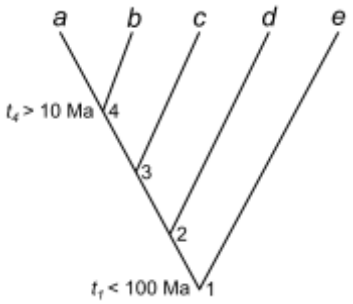
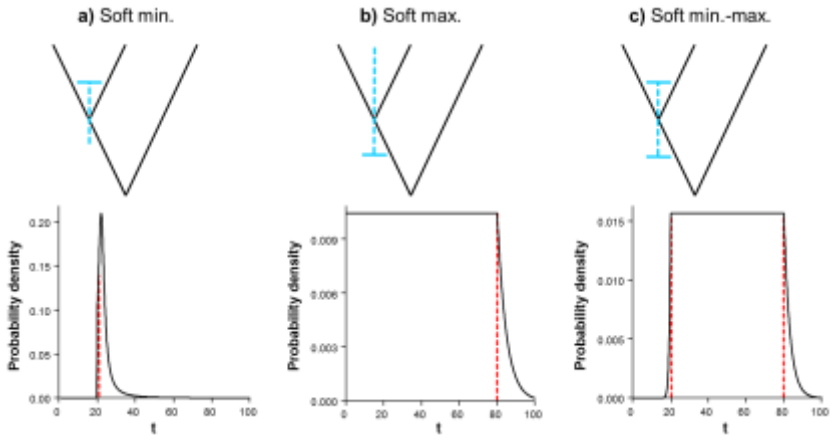


Figure 1

665



666

667

668 Figure 2

669

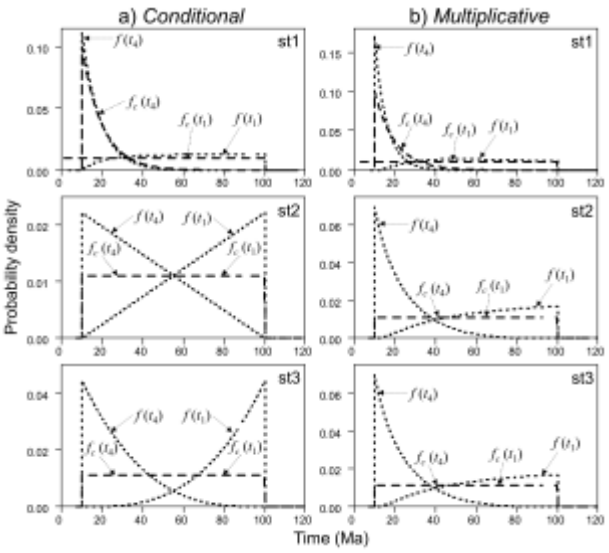
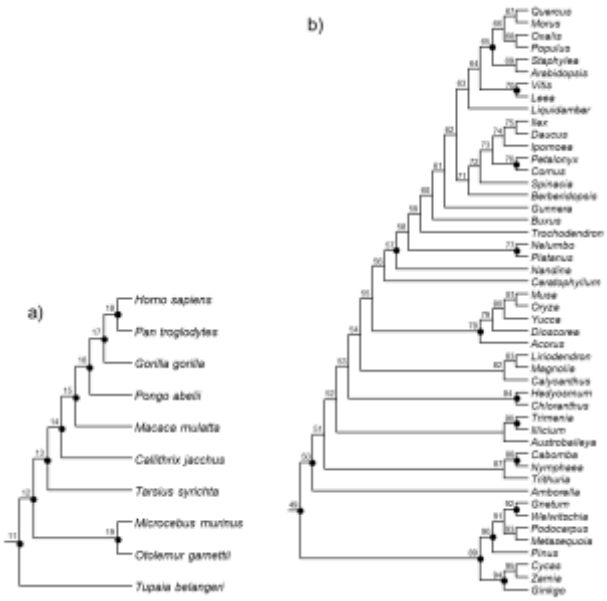

a) Conditional
b) Multiplicative
st1
st2
st3
Probability density
Time (Ma)

Figure 3

675



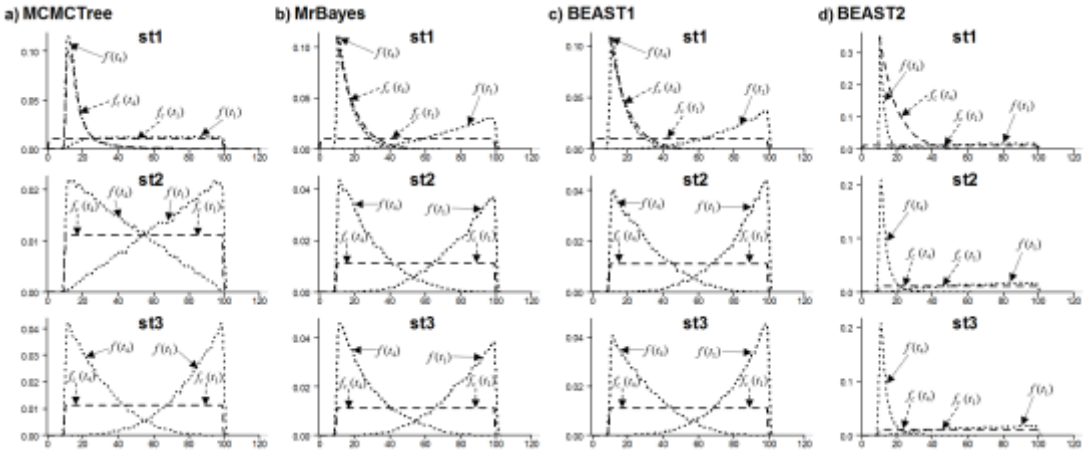
676

677

678 Figure 4

679

680



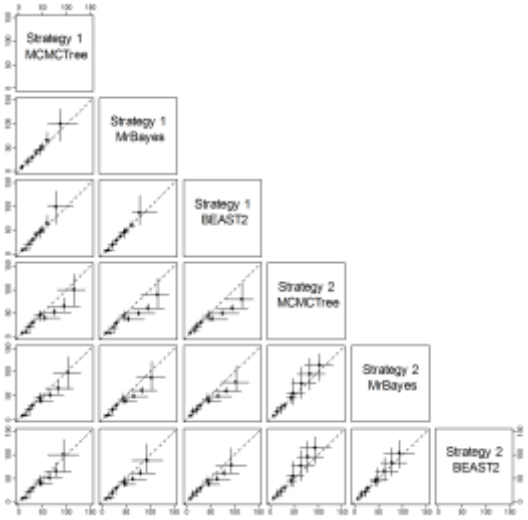
681

682

683 Figure 5

684

685



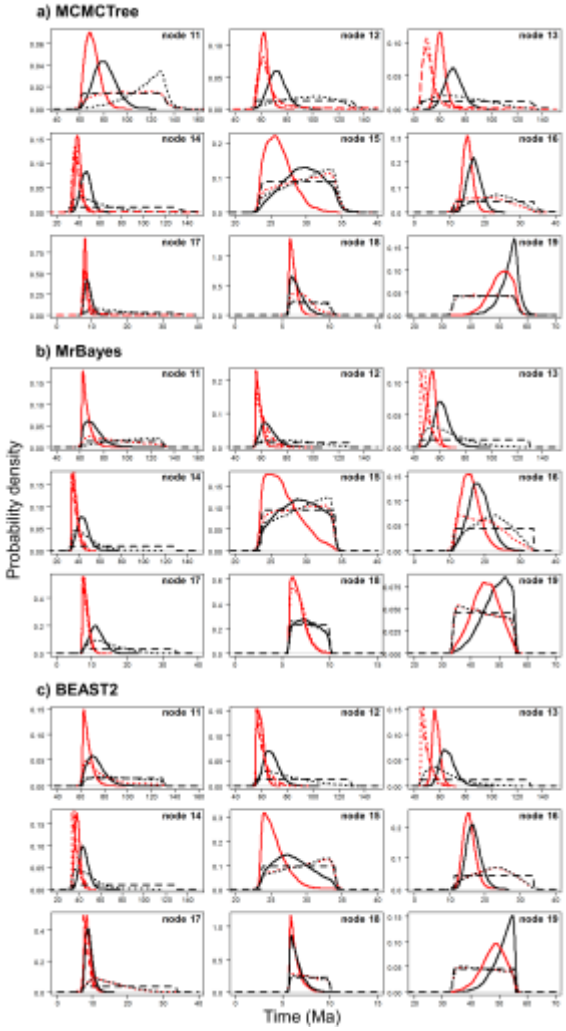
686

687

688 Figure 6

689

690



691

692

693 Figure 7

694

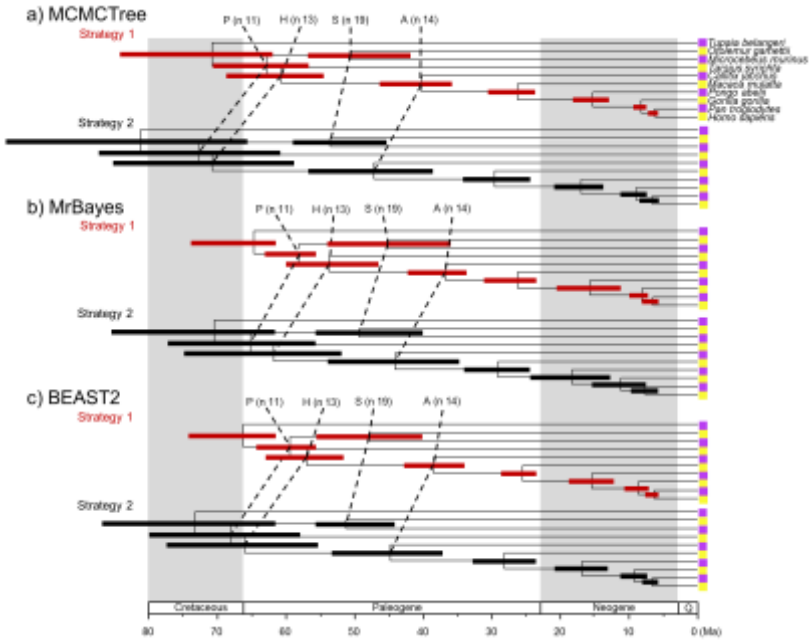
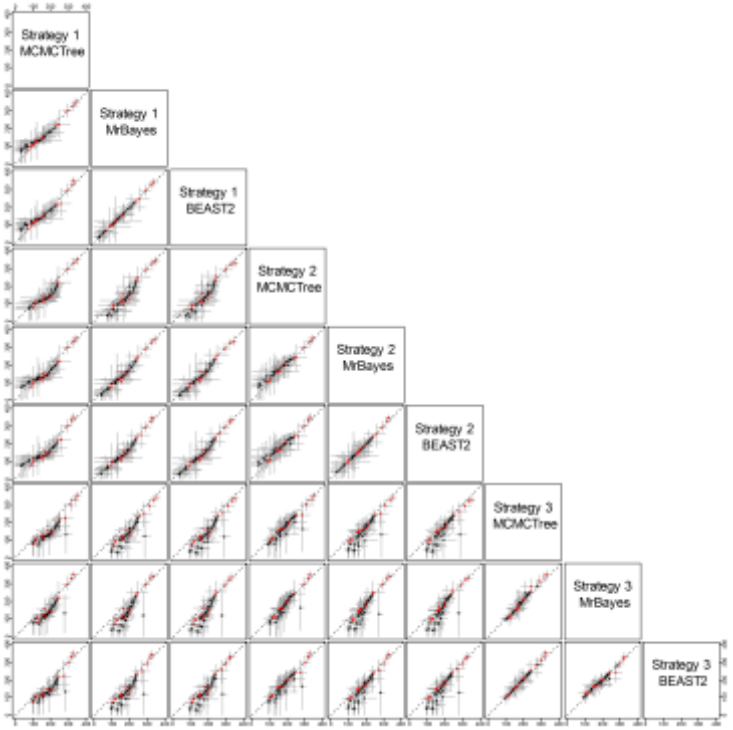

a) MCMCTree
Strategy 1
Strategy 2
P (n 11)
H (n 13)
S (n 18)
A (n 14)
b) MrBayes
Strategy 1
Strategy 2
P (n 11)
H (n 13)
S (n 19)
A (n 14)
c) BEAST2
Strategy 1
Strategy 2
P (n 11)
H (n 13)
S (n 19)
A (n 14)
Cretaceous
Paleogene
Neogene
Q
80
70
60
50
40
30
20
10
0 (Ma)

Figure 8

700



701

702

703 Figure 9

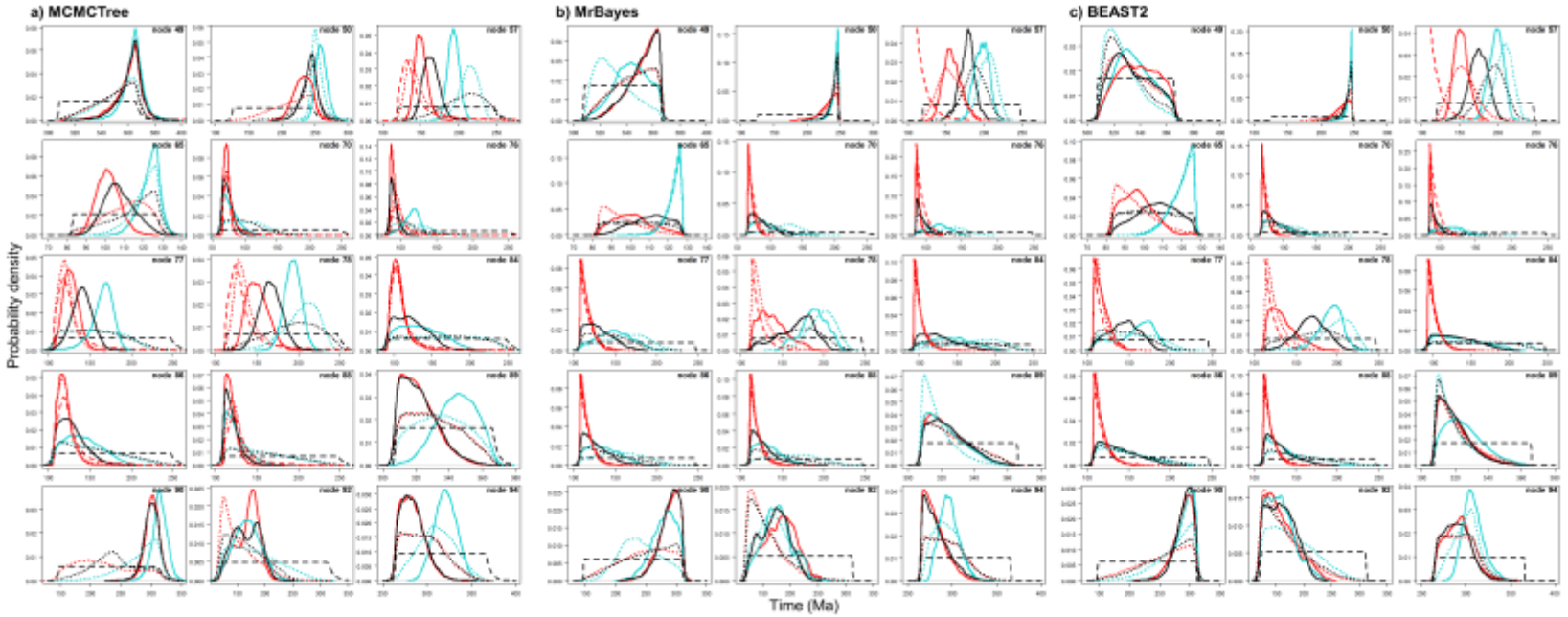


Figure 10

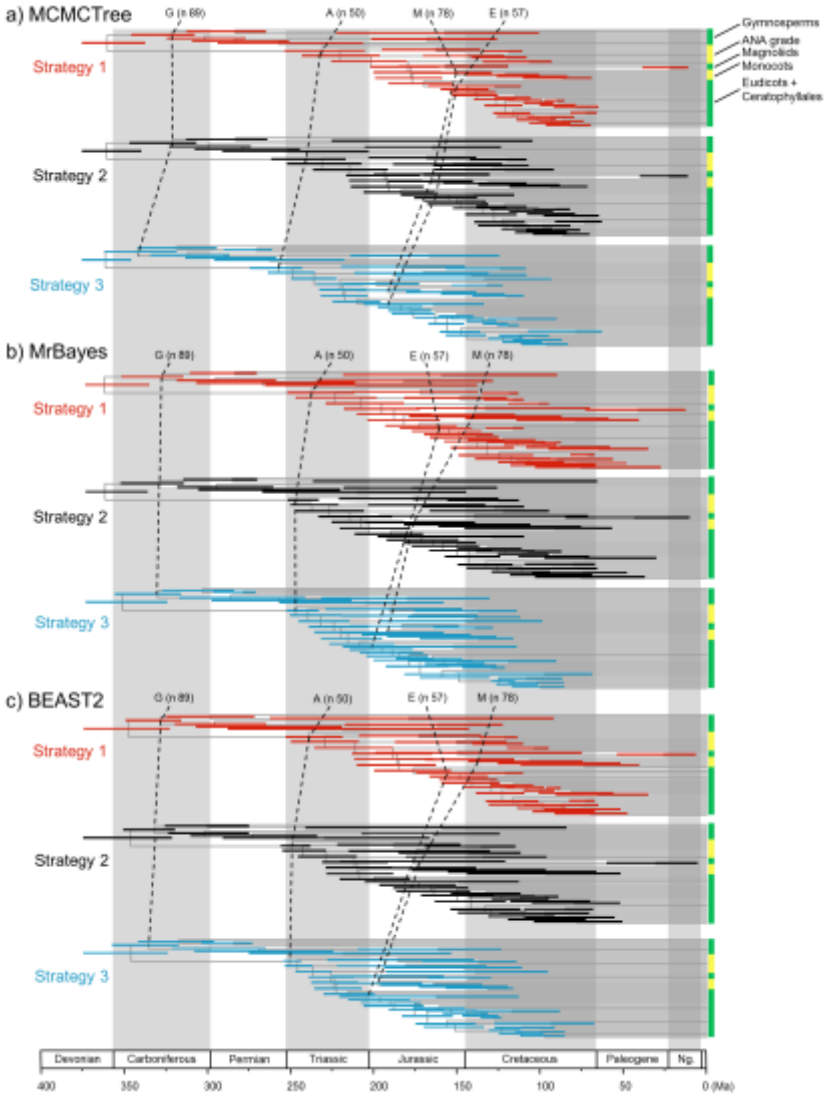

a) MCMCTree
G (n 89)
A (n 50)
M (n 78)
E (n 57)
Gymnosperms
ANA grade
Magnoliids
Monocots
Eudicots +
Ceratophyllales
Strategy 1
Strategy 2
Strategy 3
b) MrBayes
G (n 89)
A (n 50)
E (n 57)
M (n 78)
Strategy 1
Strategy 2
Strategy 3
c) BEAST2
G (n 89)
A (n 50)
E (n 57)
M (n 78)
Strategy 1
Strategy 2
Strategy 3
Devonian
Carboniferous
Permian
Triassic
Jurassic
Cretaceous
Paleogene
Ng
400
350
300
250
200
150
100
50
0 (Ma)

Figure 11