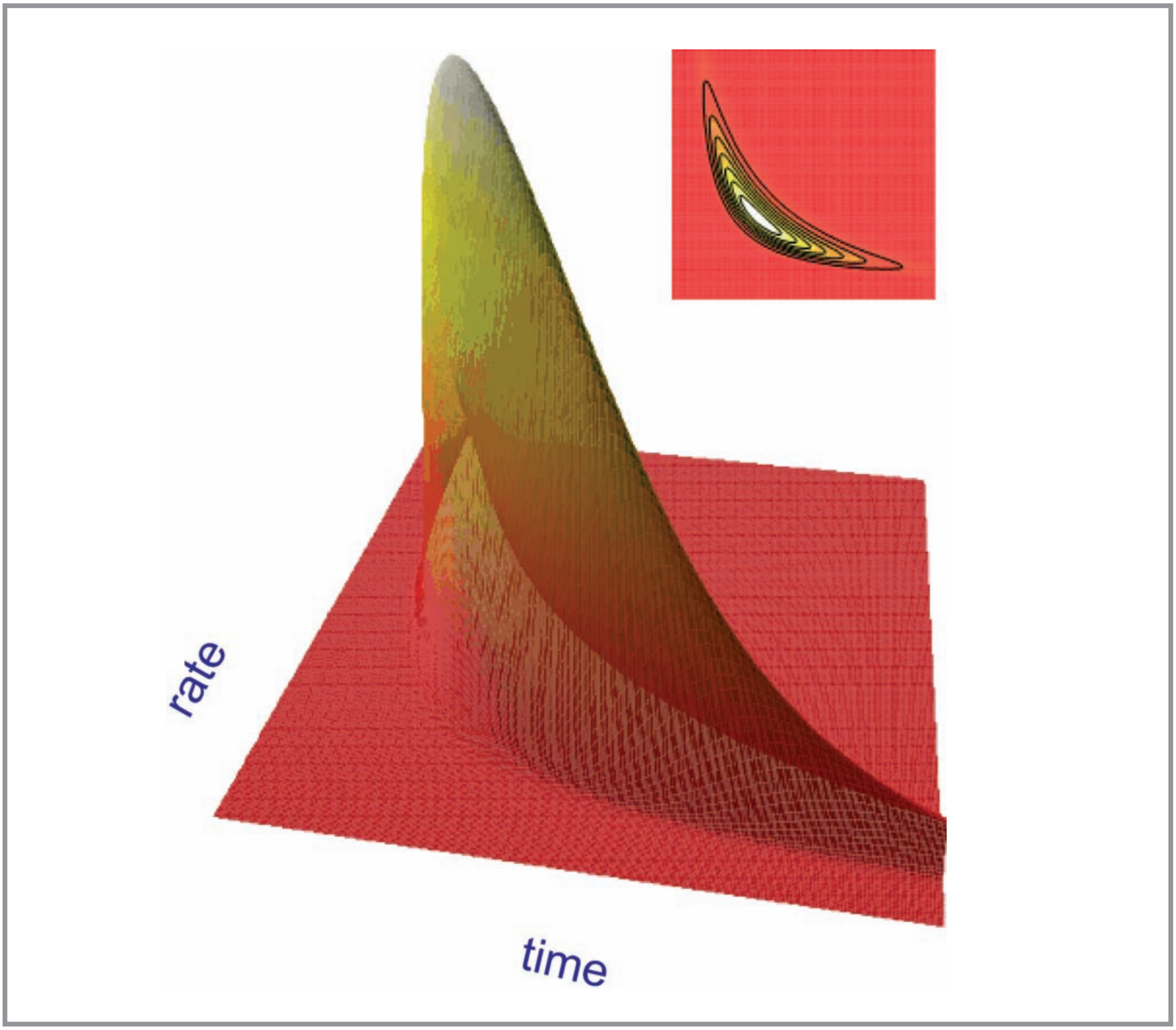
JSE

# Journal of Systematics and Evolution

Volume 51 Number 1 January 2013



Sponsored by Institute of Botany, Chinese Academy of Sciences and Botanical Society of China

Journal of Systematics and Evolution 51 (1): 30–43 (2013) doi: 10.1111/j.1759-6831.2012.00236.x

Research Article

# The unbearable uncertainty of Bayesian divergence time estimation

Mario DOS REIS Ziheng YANG*

(*Department of Genetics, Evolution and Environment*, *University College London*, Darwin Building, Gower Street, London WC1E 6BT, UK)

**Abstract** Divergence time estimation using molecular sequence data relying on uncertain fossil calibrations is an unconventional statistical estimation problem. As the sequence data provide information about the distances only, estimation of absolute times and rates has to rely on information in the prior, so that the model is only semi-identifiable. In this paper, we use a combination of mathematical analysis, computer simulation, and real data analysis to examine the uncertainty in posterior time estimates when the amount of sequence data increases. The analysis extends the infinite-sites theory of Yang and Rannala, which predicts the posterior distribution of divergence times and rate when the amount of data approaches infinity. We found that the posterior credibility interval in general decreases and reaches a non-zero limit when the data size increases. However, for the node with the most precise fossil calibration (as measured by the interval width divided by the mid value), sequence data do not really make the time estimate any more precise. We propose a finite-sites theory which predicts that the square of the posterior interval width approaches its infinite-data limit at the rate $1/n$, where $n$ is the sequence length. We suggest a procedure to partition the uncertainty of posterior time estimates into that due to uncertainties in fossil calibrations and that due to sampling errors in the sequence data. We evaluate the impact of conflicting fossil calibrations on posterior time estimation and point out that narrow credibility intervals or overly precise time estimates can be produced by conflicting or erroneous fossil calibrations.

**Key words** finite-sites theory, fossil calibration, infinite-sites plot, molecular clock.

There has been great interest in estimating the times of species divergences using molecular data since the proposal of the molecular clock hypothesis 40 years ago (Zuckerkandl & Pauling, 1965). The field has moved a long way since then, and several sophisticated computer programs that implement Bayesian estimation of divergence times have been developed, such as MULTIDIVTIME (Thorne et al., 1998; Kishino et al., 2001), BEAST (Drummond & Rambaut, 2007), MrBAYES (Ronquist et al., 2012b), MCMCTREE (Yang, 2007), and PHYLOBAYES (Lartillot et al., 2009), among others. Nevertheless, divergence time estimation using molecular data is a complicated problem. Molecular data provide information only about the distances among species on a phylogeny, but not about the geological ages of clades nor the molecular evolutionary rate. The molecular rate $r$ and the divergence time $t$ always appear as a product (the distance $d = rt$) in the likelihood function (the probability of the sequence data). In other words, $r$ and $t$ are confounded and cannot be estimated separately from sequence data alone. If information on the times of divergence of one or more pairs of species is available (say from the fossil record or some geological event), such information can be used to construct a prior on the divergence times. The Bayesian method can then be used to estimate the molecular rate as well as the divergence times in the phylogeny from a sequence alignment (Thorne et al., 1998).

Overcoming the identifiability problem of rates and times is challenging even in the Bayesian context. Yang & Rannala (2006) and Rannala & Yang (2007) have shown that as the amount of molecular sequence data approaches infinity, the joint posterior distribution of times does not converge to a point mass on the true times as occurs in a conventional estimation problem, but to a one-dimensional distribution. In other words, the root age $t_1$ has a posterior distribution, while the posterior distribution of any other time $t_i$ in the phylogeny is simply a linear transform of the distribution of $t_1$. As a consequence, a plot of posterior credibility interval (CI) widths of times vs. posterior mean times approaches a straight line when the amount of sequence data approaches infinity (Yang & Rannala, 2006; Rannala & Yang, 2007). This plot is known as the infinite-sites plot. An example is shown in Fig. 1, from

Received: 16 July 2012 Accepted: 4 November 2012

* Author for correspondence. E-mail: z.yang@ucl.ac.uk. Tel.: 44-20-76794379. Fax: 44-20-76797096.

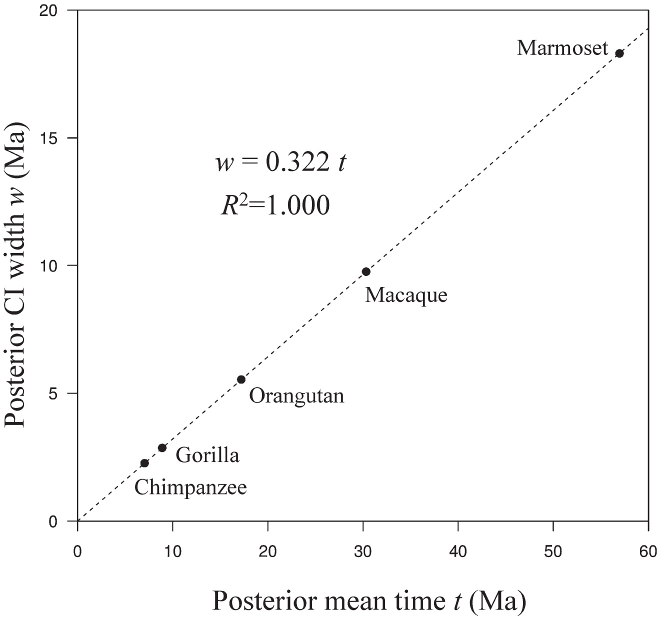


**Fig. 1.** Infinite-sites plot for the divergence times of human versus five other primate species (see phylogeny in Fig. 7 later). Divergence times, in millions of years ago (Ma), were estimated on an alignment of 8 708 584 sites using the Bayesian method with the program MCMCTREE. The Bayesian analysis produces a posterior mean and the 95% posterior credibility interval for each of the 5 interior nodes on the tree. The CI width (the difference of the 2.5% and 97.5% limits) is plotted against the posterior mean in the scatter plot. A line with intercept 0 is fitted to the scatter plot. Despite the very long alignment, the uncertainty of time estimates (as measure by the CI width *w*) does not go to zero, but converges to a limiting value determined by the uncertainty in the fossil calibrations. In this example, for every one million years of divergence, 0.322 million years are added to the CI width of time estimates. The data set is analyzed later in this paper, where the details of the analysis are given (see, e.g., legend to Fig. 7).

an analysis of a large primate dataset. The near perfect fit of the straight line implies that nearly all the uncertainty in the posterior time estimate is due to uncertainties in the fossil calibrations, and adding more sequence data is unlikely to produce more precise estimates. The uncertainty of fossil information therefore imposes a theoretical limit on the precision that can be achieved in divergence time estimation when the amount of sequence data increases.

Although Yang & Rannala (2006) and Rannala & Yang (2007) derived the asymptotic distribution of times for infinite sequence data, the behavior of the posterior distribution of times in large but finite datasets has not been explored. It is unclear whether the uncertainty in posterior time estimates always decreases when the amount of sequence data increases. In any real data analysis, we will also be interested in knowing whether the uncertainty in the posterior time estimates is largely due to uncertainties in fossil calibrations or to finite amount of sequence data. Another interesting question is the asymptotic behavior of the posterior of times when the molecular data and the prior fossil information are in conflict. In a conventional Bayesian analysis, conflicting priors are eventually overruled by the data, and the posterior converges to the true parameter values when the data size increases. In divergence time estimation, the identifiability problem precludes any guarantee that the posterior of times will converge to the true values. In fact, the posterior of times may converge to incorrect values with arbitrarily small uncertainty, especially if some fossil calibrations are wrong (Yang & Rannala, 2006) or if there are conflicts among the calibrations.

In this paper we study the behavior of the posterior distribution of times as the amount of data increases and when the molecular clock holds. The fossil information may either be adequate or conflicting with the molecular data. We develop a finite-sites theory of divergence time estimation, which should apply to all current Bayesian methods, despite their idiosyncratic ways of dealing with fossil calibrations. For example, while we consider the uniform and gamma calibration densities, the theory applies to other densities such as exponential, log-normal, etc. We analyze a few example phylogenies containing from two to nine species, using a combination of mathematical analysis (for the simple cases) and computer simulation. We use the data of six primate genomes to demonstrate the same patterns in real data sets.

## 1 The finite-sites theory of uncertainty in divergence time estimation

Under certain regularity conditions, the variance $V(\hat{\theta})$ of a maximum likelihood estimator $\hat{\theta}$ is asymptotically proportional to $1/n$ with $n$ to be the sample size. As $n \to \infty$, $V(\hat{\theta}) \to 0$ and the estimator converges to the true parameter value. Because for large samples the likelihood function dominates the posterior, the posterior variance is also asymptotically proportional to $1/n$ and the posterior estimate also converges to the true value. In contrast, in divergence time estimation the posterior variance of the divergence time does not converge to zero.

In general, calculating the posterior variance of divergence times seems intractable. Therefore we examine a generic, simpler case of parameter estimation when the parameters are confounded. Although this example is not directly related to divergence time estimation, it sheds light on the asymptotic variance of confounded parameter estimates in the Bayesian setting. We wish to estimate parameters $\mu_1$ and $\mu_2$ when the data and likelihood depend on $\mu = \mu_1 + \mu_2$ only (Yang & Rannala, 2006). The data $\mathbf{y} = (y_i)$ are an independent and identically distributed sample from the normal distribution $N(\mu, 1)$. We assign priors $\mu_1 \sim N(-1, v_1)$ and $\mu_2 \sim N(1, v_2)$. Yang & Rannala (2006) gave the posterior variance of $\mu_1$ as

$$V(\mu_1|\mathbf{y}) = \frac{v_1(1+nv_2)}{1+nv_1+nv_2} = V_\infty(\mu_1|\mathbf{y}) + \frac{v_1^2}{(v_1+v_2)+n(v_1+v_2)^2}, \qquad (1)$$

where

$$V_\infty(\mu_1|\mathbf{y}) = \lim_{n\to\infty} V(\mu_1|\mathbf{y}) = \frac{v_1 v_2}{v_1+v_2}, \qquad (2)$$

is the variance when $n \to \infty$. For large $n$, Equation (1) can be approximated by

$$V(\mu_1|\mathbf{y}) \approx V_\infty(\mu_1|\mathbf{y}) + \frac{1}{n} \times \left(\frac{v_1}{v_1+v_2}\right)^2. \qquad (3)$$

Thus the posterior variance in the finite data approaches its limit at the rate $1/n$:

$$V(\mu_1|\mathbf{y}) - V_\infty(\mu_1|\mathbf{y}) \propto \frac{1}{n}. \qquad (4)$$

The ratio

$$u_{\mu_1} = 1 - \frac{V_\infty(\mu_1|\mathbf{y})}{V(\mu_1|\mathbf{y})} = \frac{1}{\left(1+\frac{v_2}{v_1}\right)(1+nv_2)} \qquad (5)$$

is then the fraction of uncertainty (variance) in the posterior estimate of $\mu_1$ that is due to the finite data and that can be reduced by increasing the amount of data.

In Fig. 2 we plot $u_{\mu_1}$ as a function of $n$ for four sets of ($v_1$, $v_2$): (1, 1), (10, 10), (1, 10), and (10, 1). When the prior on $\mu_1$ is informative ($v_1 = 1$, $v_2 = 10$), the posterior of $\mu_1$ is similar to the prior even in large datasets. Similarly, if $v_2$ has an informative prior and $v_1$ has a diffuse prior, then the posterior distribution will be dominated by the prior on $v_2$ instead.

Translating the results of this simple example to divergence time estimation leads to the following predictions. In large datasets, we expect the square of the CI width $w^2$ to be proportional to the variance. Thus we expect $w^2 - w_\infty^2$ (where $w_\infty$ is the width for infinite data) to approach zero at the rate $1/n$, with $n$ to be the data size or sequence length. Furthermore, $u_F = w_\infty^2/w^2$ will be the fraction of uncertainty in posterior time estimate that is due to uncertainties in the fossil calibration (or in the prior for times and rate) while $u_S = 1 - w_\infty^2/w^2$ will be the fraction due to finite amounts of sequence data. For ages of nodes with a very informative calibration, we expect $u_S$ to be small even in small sequence datasets. For nodes with a diffuse calibration or no calibration, $u_S$ should be large initially but goes to zero quickly when the amount of sequence data increases.

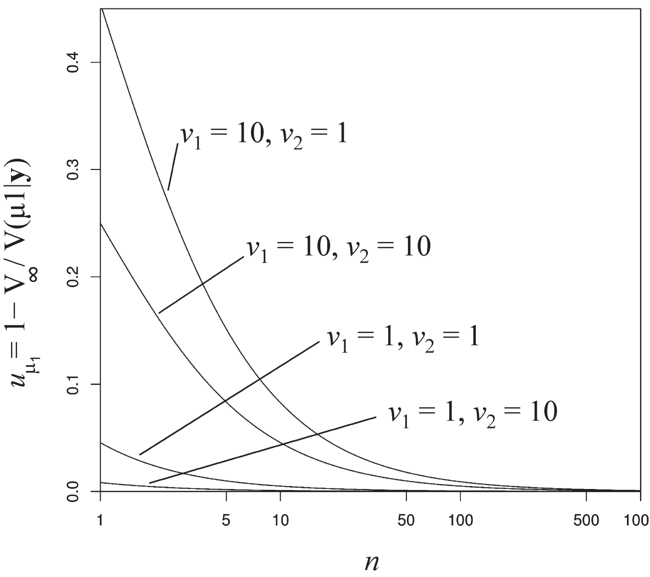


**Fig. 2.** The fraction of uncertainty in the posterior of $\mu_1$, $f(\mu_1|\mathbf{y})$, attributed to limited data when the likelihood depends on $\mu = \mu_1 + \mu_2$.

Below we analyze several simple cases as well as a real dataset to confirm these predictions.

## 2 Uncertainty in posterior time estimates with finite sequence data

### 2.1 The case of two species

Consider an alignment of two nucleotide sequences with $n$ sites and $x$ differences. Suppose the true divergence time is $t = 1$ and the true rate is $r = 0.5$. If one time unit is 100 million years (My), this means a divergence time of 100 My and an evolutionary rate of $10^{-8}$ substitutions per site per year. The true distance, or expected number of changes per site between the two sequences is $d = r \times 2t = 1$. We use the Jukes & Cantor (1969) model to calculate the likelihood of observing the sequence alignment given $r$ and $t$

$$L(r,t) = p^x(1-p)^{n-x} = \left(\frac{3}{4} - \frac{3}{4}\mathrm{e}^{-8rt/3}\right)^x \left(\frac{1}{4} + \frac{3}{4}\mathrm{e}^{-8rt/3}\right)^{n-x}, \qquad (6)$$

or

$$L(d) = \left(\frac{3}{4} - \frac{3}{4}\mathrm{e}^{-4d/3}\right)^{x}\left(\frac{1}{4} + \frac{3}{4}\mathrm{e}^{-4d/3}\right)^{n-x}, \qquad (7)$$

where $p = \frac{3}{4} - \frac{3}{4}\exp(-8rt/3)$ is the expected proportion of differences in the alignment. The maximum likelihood estimate of the distance $d$ is

$$\hat{d} = -\frac{3}{4} \times \log\left(1 - \frac{4}{3}\hat{p}\right), \qquad (8)$$

where $\hat{p} = x/n$. However, neither $r$ nor $t$ has a unique maximum likelihood estimate because the likelihood surface of Equation (6) is maximized by all points along the line $r = \hat{d}/2t$ (Fig. 3: B, B′).

If prior information about $r$ and $t$ is available, we may obtain estimates for these parameters using the Bayesian method. Consider a gamma prior on the rate $r \sim \mathrm{G}(2,4)$ with mean 0.5, and a gamma prior on the time $t \sim \mathrm{G}(2,2)$ with mean 1. Note that the gamma density with parameters $\alpha$ and $\beta$ is

$$g(t|\alpha,\beta) = \frac{\beta^{\alpha}\mathrm{e}^{-\beta t}t^{\alpha-1}}{\Gamma(\alpha)},$$

with mean $\alpha/\beta$, variance $\alpha/\beta^2$ and mode $(\alpha-1)/\beta$ (if $\alpha > 1$). The joint prior of rate and time is thus

$$f_{RT}(r,t) = f_R(r)f_T(t) = 16r\mathrm{e}^{-4r} \times 4t\mathrm{e}^{-2t}. \qquad (9)$$

The joint prior is rather diffuse, with a mode at $t = 0.5$ and $r = 0.25$ (Fig. 3: A, A′).

The joint posterior distribution of $r$ and $t$ is

$$f_{RT}(r,t|x) = \frac{1}{C}f_{RT}(r,t)L(r,t), \qquad (10)$$

where $C$ is a normalizing constant

$$C = \int_0^\infty \int_0^\infty f_{RT}(r,t)L(r,t)\mathrm{d}t\mathrm{d}r. \qquad (11)$$

We use an adaptive numerical algorithm (cubature package in R) to calculate $C$.

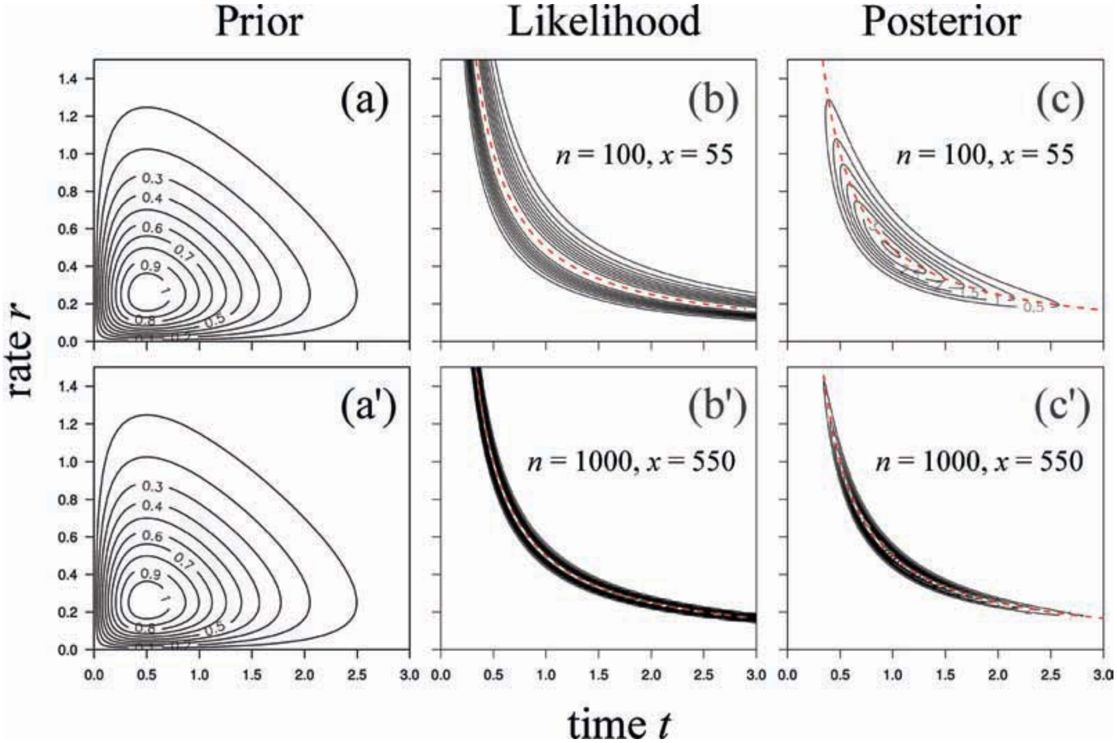


**Fig. 3.** The prior, likelihood and posterior distribution of rate $r$ and time $t$ for two datasets of a pairwise sequence alignment. In (A) and (A′) the joint prior of Equation (9) is shown. In (B) and (B′) the likelihood of Equation (6) is shown for $x = 55$, $n = 100$ and for $x = 550$, $n = 1000$. In (C) and (C′) the joint posterior distribution of Equation (10) is shown. The true values are $r = 0.5$, $t = 1$, $d = 1$ and the expected proportion of differences in the alignment is $p = \frac{3}{4} - \frac{3}{4}\exp(-4/3) = 0.5523$. In both (B) and (B′) $\hat{d} = 0.9913$. The likelihood surface resembles an L-shaped ridge, with the maximum along the $r = \hat{d}/2t$ line (dashed line).

The joint posterior distribution is shown in Fig. 3C and C′ for two example data sets, with $n = 100$, $x = 55$, or $n = 1000$, $x = 550$.

The marginal posterior distributions of $r$ and $t$ are

$$f(r|x) = \frac{1}{C}\int_0^\infty f_{RT}(r,t)L(r,t)\mathrm{d}t, \qquad (12)$$

$$f(t|x) = \frac{1}{C}\int_0^\infty f_{RT}(r,t)L(r,t)\mathrm{d}r. \qquad (13)$$

The posterior for infinite data can be obtained following Yang & Rannala (2006), as the prior density $f_{RT}(r,t)$ conditioned on $d = \hat{d}$. Using the variable transform $r = d/2t$, and noting that the Jacobian of the transform is $|\partial(r,t)/\partial(d,t)| = 1/(2t)$, we have

$$f_\infty(t|d = \hat{d}\,) = \frac{f_{RT}\left(\frac{\hat{d}}{2t}, t\right) \times \frac{1}{2t}}{\int_0^\infty f_{RT}\left(\frac{\hat{d}}{2t}, t\right) \times \frac{1}{2t}\,\mathrm{d}t}. \qquad (14)$$

Similarly, the posterior of $r$ is

$$f_\infty(r|d = \hat{d}) = \frac{f_{RT}\left(r, \frac{\hat{d}}{2r}\right) \times \frac{1}{2r}}{\int_0^\infty f_{RT}\left(r, \frac{\hat{d}}{2r}\right) \times \frac{1}{2r}\,\mathrm{d}r}. \qquad (15)$$

Fig. 4 shows the marginal prior and marginal posterior densities of $t$ and $r$ for example data and for infinite data. It can be seen that the posterior variances of $t$ and $r$ are reduced with the increase of sequence data, approaching a non-zero limit. The posterior for $n = 1000$ and $n = \infty$ are indistinguishable, and even that for $n = 100$ is very close. An important implication of Equation (14) is that if the prior on the rate is diffuse (so that $f_R(r) \approx a$, constant around $\hat{d}/2t$), the posterior variance of $t$ is essentially the same as the prior variance, and no amount of molecular data will affect the inference of $t$. Note that the marginals (Fig. 4) appear to approach their limits faster than the joint distribution of $t$ and $r$ (Fig. 3). Also note that there is a single posterior distribution for infinite data, so that if $t$ is given, then $r$ is known with absolute precision.

We now calculate the expected credibility interval (CI) width $w$, for the posterior time $t$, averaging over possible data sets. This is tedious to do analytically so for each sequence length $n = 10, 10^2, 10^3, 10^4$, and $10^5$, we use the program EVOLVER to simulate 1000 pairwise sequence alignments. We then use the program MCMCTREE to calculate the posterior of $t$, $r$ and the CI-width $w$ for each simulated alignment. EVOLVER and MCMCTREE are part of the PAML software package (Yang, 2007). Table 1 shows that as the number of sites approaches infinity, the CI width decreases until reaching a limiting value. At $d = r \times 2t = 1$, the sequence data are very informative so that with only 100 or 1000 sites in the alignment, the CI width is close to the infinite-data limit.

## 2.2 The case of three species

To consider $s > 2$ species, evolving under the strict clock, note that for infinite data the marginal

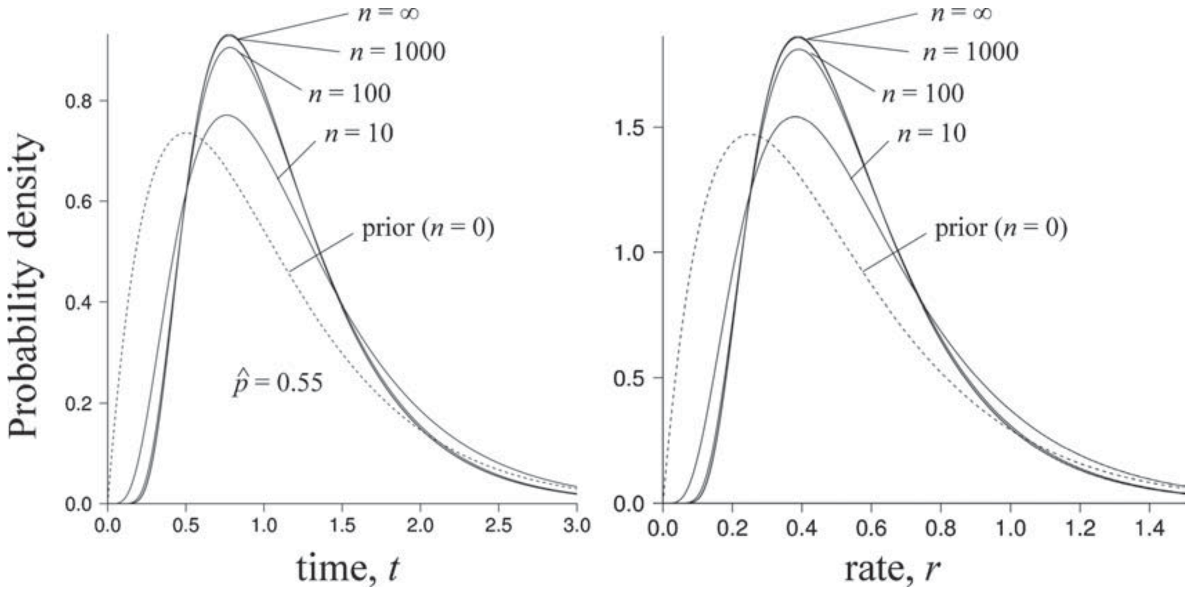


**Fig. 4.** The marginal prior (dashed line) and posterior (solid lines) of $t$ (left) and $r$ (right). The prior densities are $f_T(t) = g(t|\alpha = 2, \beta = 2)$ and $f_R(r) = g(r|\alpha = 2, \beta = 4)$. The posterior densities for finite data are calculated using Equations (12) and (13), and for infinite data using Equations (14) and (15). In all cases $\hat{p} = 0.55$ and $\hat{d} = 0.9913$. The posterior densities for $n = 1000$ and $n = \infty$ are almost identical in both cases.

**Table 1** Posterior means, 95% credibility intervals and interval widths of divergence time between two species

| $n$ | $t = 1$ | 95% CI | $w$ |
|---|---|---|---|
| 0 | 1.000 | (0.121, 2.784) | 2.663 |
| 10 | 1.186 | (0.324, 2.836) | 2.514 |
| $10^2$ | 1.140 | (0.393, 2.571) | 2.179 |
| $10^3$ | 1.119 | (0.397, 2.516) | 2.121 |
| $10^4$ | 1.118 | (0.397, 2.513) | 2.117 |
| $10^5$ | 1.119 | (0.398, 2.514) | 2.117 |
| $\infty$ | 1.119 | (0.398, 2.516) | 2.118 |

Note: The results are averages over 1000 replicate datasets.

posterior distribution of the root age $t_1$ in a phylogeny is given by Yang & Rannala (2006)

$$f(t_1|\mathbf{d} = \hat{\mathbf{d}}) \propto f_R\left(\frac{\hat{d}_1}{t_1}\right) \times f_{\mathbf{T}}\left(t_1, \frac{\hat{d}_2}{\hat{d}_1}t_1, \ldots, \frac{\hat{d}_{s-1}}{\hat{d}_1}t_1\right) \times \left(\frac{t_1}{\hat{d}_1}\right)^{s-2} \times \frac{1}{t_1}, \qquad (16)$$

where $f_R(r)$ is the prior for rate $r$, $f_{\mathbf{T}}(t_1, \ldots, t_{s-1})$ is the joint prior of node ages $(t_1, \ldots, t_{s-1})$, and $\hat{\mathbf{d}} = (\hat{d}_i)$ are the maximum likelihood estimates of the distances from interior nodes of the phylogeny to the present time (note that $d_i$'s here differ from the two-sequence case above by a factor of 2).

We now consider the three-species phylogeny of Fig. 5A, with root age $t_1 = 1$, the age of the internal node $t_2 = 0.5$, and the rate $r = 1$. We use an exponential prior with mean 1 for the rate, $f_R(r) = \exp(-r)$, and a gamma prior on the root age $t_1 \sim \mathrm{G}(100, 100)$. This is nearly the same as a normal distribution with mean 1 and standard deviation 0.1, and represents a fairly informative calibration with the 95% prior interval to be (0.814, 1.205). The prior on $t_2$ is conditioned on $t_1$, and is uniformly distributed between 0 and $t_1$. The joint time prior is

$$f_{\mathbf{T}}(t_1, t_2) = g(t_1|100, 100) \times \frac{1}{t_1}. \qquad (17)$$

From Equation (16), the posterior of the root age for infinite data, given the distances $\hat{\mathbf{d}} = (\hat{d}_1, \hat{d}_2)$ is

$$f(t_1|\hat{\mathbf{d}}) \propto f_R\left(\frac{\hat{d}_1}{t_1}\right) f_{\mathbf{T}}\left(t_1, \frac{\hat{d}_2}{\hat{d}_1}t_1\right) \times \frac{1}{\hat{d}_1.} \qquad (18)$$

For finite data we use computer simulation. The program EVOLVER is used to generate sequence alignments of three species using the phylogeny of Fig. 5A under the Jukes–Cantor model. One thousand alignments were simulated for each of $n = 10^2$, $10^3$, $10^4$, $10^5$, and $10^6$ sites. The program MCMCTREE was used to estimate the posterior distribution of the rate and the times.

Table 2 shows the posterior CI widths for $t_1$ and $t_2$ for the different alignment lengths, averaged over the 1000 replicates. Because the prior on the rate is diffuse, the prior CI width for $t_1$ is essentially the same as the posterior CI width, so that the molecular data do not really reduce the uncertainty in the posterior estimate of $t_1$. Molecular data are informative only about the relative ages of the nodes (i.e., about $t_2/t_1$) and as the amount of molecular data increases, the posterior CI width of $t_2$ is progressively reduced until it is exactly half that of $t_1$ for infinite data. We note that the width $w_2$ for $t_2$ decreases very rapidly, and $10^4$ sites are nearly as informative as infinite data. Fig. 6 shows the plot of $w_2^2 - w_{2,\infty}^2$ versus $1/n$. The trend is expectedly close to a straight line.

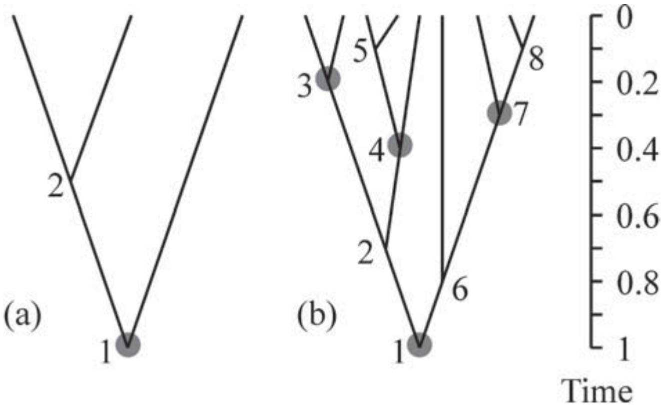


**Fig. 5.** The two tree topologies used in the simulations. Grey circles denote nodes with calibrations. (A) Tree of three species. (B) Tree of nine species. In (A), the calibration for the root is $t_1 \sim \mathrm{G}(100, 100)$. In (B) the calibrations are $t_1 \sim B(0.5, 1.5)$, $t_3 \sim B(0.1, 0.3)$, $t_4 \sim B(0.3, 0.5)$ and $t_7 \sim B(0.2, 0.4)$. Here $t \sim B(a, b)$ means uniform over $a < t < b$, although soft-bounds are used; that is, there is a 2.5% probability mass for $t < a$, and 2.5% for $t > b$.

## 2.3 The case of nine species

We now consider the nine-species phylogeny of Fig. 5B. The rate is $r = 1$, with prior $f_R(r) = \exp(-r)$. The age of the root is $t_1 = 1$ and the ages of the other nodes are $t_2 = 0.7$, $t_3 = 0.2$, $t_4 = 0.4$, $t_5 = 0.1$, $t_6 = 0.8$, $t_7 = 0.3$, and $t_8 = 0.1$. Four nodes have fossil calibrations: $t_1 \sim B(0.5, 1.5)$, $t_3 \sim B(0.1, 0.3)$, $t_4 \sim B(0.3, 0.5)$, and $t_7 \sim B(0.2, 0.4)$. We use a soft uniform distribution, $B(a,b)$, with 2.5% probability mass added at each tail (for details see Yang & Rannala, 2006).

We use MCMCTREE to obtain the posterior distribution of the rate and times for simulated data sets. We simulated 1000 alignments for each of $n = 10^2$, $10^3$, $10^4$, $10^5$, and $10^6$ sites. Table 3 shows the average

**Table 2** Posterior means, 95% credibility intervals and interval widths of divergence times $t_1$ and $t_2$ for the three-species tree of Fig. 5A

| $n$ | $t_1 = 1$ | 95% CI | $w_1$ | $t_2 = 0.50$ | 95% CI | $w_2$ |
|---|---|---|---|---|---|---|
| 0 | 1.000 | (0.814, 1.205) | 0.392 | 0.500 | (0.025, 1.026) | 1.001 |
| $10^2$ | 1.003 | (0.817, 1.207) | 0.390 | 0.510 | (0.211, 0.859) | 0.648 |
| $10^3$ | 1.001 | (0.815, 1.205) | 0.390 | 0.496 | (0.363, 0.646) | 0.283 |
| $10^4$ | 1.000 | (0.815, 1.204) | 0.390 | 0.501 | (0.404, 0.608) | 0.204 |
| $10^5$ | 1.000 | (0.815, 1.204) | 0.390 | 0.500 | (0.407, 0.603) | 0.196 |
| $10^6$ | 1.000 | (0.815, 1.204) | 0.390 | 0.500 | (0.407, 0.602) | 0.195 |
| $\infty$ | 1.000 | (0.815, 1.204) | 0.390 | 0.500 | (0.407, 0.602) | 0.195 |

Note: There is a fossil calibration for the age of the root $t_1$.

posterior CI width $w_i$ for each of the eight node ages. Note that for node 4 with the most precise calibration, the interval width stayed largely constant (minor differences may be due to random sampling errors). For all other nodes the posterior CI widths of the node ages are progressively reduced until reaching their limits at $n = \infty$. In the limit the posterior CI widths are proportional to the times.

## 2.4 Analysis of genomic data from six primate species

We estimate the divergence times on a phylogeny of six primate species (Fig. 7). We work with two sequence alignments for the six species: (1) the 1st and 2nd codon positions and (2) the 3rd codon positions from 14 631 orthologous protein-coding genes. After removing ambiguous sites the alignments are 8 708 584 and 4 354 292 sites long, respectively. This is a subset of a larger 36 species alignment analyzed by dos Reis et al. (2012). The fossil calibrations (Fig. 7) are also from dos Reis et al. (2012). We estimate the five divergence times assuming the molecular clock. The clock can be easily rejected using a likelihood ratio test because of the large amount of data (results not shown). However the deviations from the clock are small (Fig. 8), and are not expected to have a great impact on time estimation. The divergence times were estimated with MCMCTREE, using the HKY + $\Gamma_5$ substitution model (Hasegawa et al., 1985; Yang, 1994). Table 4 shows the priors used for the rate $r$, the transition-transversion rate ratio $\kappa$, and the shape parameter $\alpha$ for the gamma model of among-site rate variation. Table 5 shows the posterior time estimates for the two alignments. The alignments are so long that they can be effectively treated as infinite data, and the posterior CI widths $w$, are nearly perfectly proportional to the posterior times $t$ (Fig. 1).

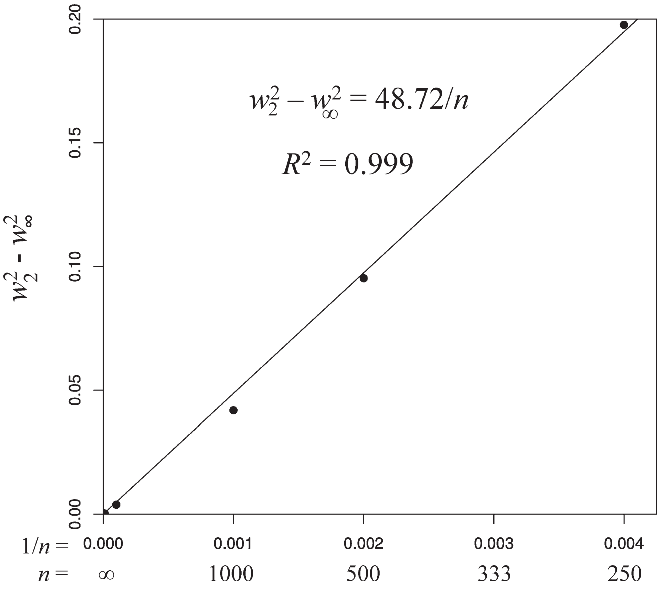


**Fig. 6.** Plot of $w^2 - w_\infty^2$ versus $1/n$ for the internal node in the three species phylogeny.

We are interested in the behavior of the posterior CI width, $w$, as the sequence length increases. We generated alignments of lengths $n = 10^2, 5 \times 10^2, 10^3, 5 \times 10^3, 10^4, 10^5$, and $10^6$ sites by sampling sites randomly without replacement from each of the two large alignments. The number of replicates is 500 for lengths $n = 10^2$–$10^5$, 250 for $10^5$ and 100 for $10^6$. We then estimated the divergence times for each replicate sample with MCMCTREE, using the HKY + $\Gamma_5$ model and the priors of Table 4. The results are summarized in Fig. 9. The CI widths for all nodes are reduced with the increase of sequence length. However for nodes 8 and 11 (Catarrhini and Hominini) the changes in $w_8$ and $w_{11}$ are very small, apparently because these two nodes have the most precise calibrations. The calibration uncertainty, which may be measured by the (interval width)/(mid value) is 0.91, 0.37, 1.00, and 0.56 for nodes 7, 8, 9, and 11. The most precise calibrations on $t_8$ and $t_{11}$ largely determine the limit of precision achievable with infinite data. As the sequence length is increased, the slope of the infinite-sites plot decreases.

In Fig. 10 we plot $w^2 - w_\infty^2$ against $1/n$ for nodes 7–11 for the two alignments. Here we treat the interval width $w$ for the full data (Table 5) as $w_\infty$ as the original datasets are large. All the points fall on a straight line as

**Table 3** Average 95% interval widths of divergence times among nine species in the tree of Fig. 5B

| $n$ | $w_1$ ($t_1 = 1$) | $w_2$ ($t_2 = 0.7$) | $w_3$ ($t_3 = 0.2$) | $w_4$ ($t_4 = 0.4$) | $w_5$ ($t_5 = 0.1$) | $w_6$ ($t_6 = 0.8$) | $w_7$ ($t_7 = 0.3$) | $w_8$ ($t_8 = 0.1$) |
|---|---|---|---|---|---|---|---|---|
| 0 | 0.997 | 0.949 | 0.200 | 0.200 | 0.421 | 0.994 | 0.200 | 0.352 |
| $10^2$ | 0.724 | 0.594 | 0.154 | 0.191 | 0.114 | 0.683 | 0.180 | 0.113 |
| $10^3$ | 0.565 | 0.396 | 0.118 | 0.192 | 0.062 | 0.486 | 0.163 | 0.063 |
| $10^4$ | 0.503 | 0.352 | 0.102 | 0.195 | 0.051 | 0.409 | 0.152 | 0.052 |
| $10^5$ | 0.490 | 0.343 | 0.098 | 0.195 | 0.049 | 0.393 | 0.147 | 0.049 |
| $10^6$ | 0.487 | 0.341 | 0.098 | 0.195 | 0.049 | 0.390 | 0.146 | 0.049 |
| $\infty$ | 0.487 | 0.341 | 0.097 | 0.195 | 0.049 | 0.390 | 0.146 | 0.049 |

Note: Nodes 1, 3, 4 and 7 have fossil calibrations.

expected, except for nodes 8 and 11. For those two nodes $w^2 - w_\infty^2$ is close to zero, so that the estimates are prone to sampling errors, and there appears to have more scatter. We can calculate the uncertainty in posterior time estimates due to sequence data, $u_S$, for the root age for different sample lengths. For example, for the 1st and 2nd codon position data, we have $w_\infty^2 = 335.15$ (Fig. 1). For samples of length $n = 10^2$, the average $w^2 = 3076.21$, and the fraction of uncertainty due to finite sequence length is then $u_S = 1 - w_\infty^2/w^2 = 1 - 335.15/3076.21 = 89.1\%$. In contrast, the proportion is much smaller in larger datasets. For example, $u_S = 7.3\%$ when sequence length $n = 10^5$, and $u_S$ is effectively zero for $n = 10^6$. The sequence dataset is so large that essentially all the uncertainty in the posterior time estimates is due to uncertainties in the fossil calibrations.

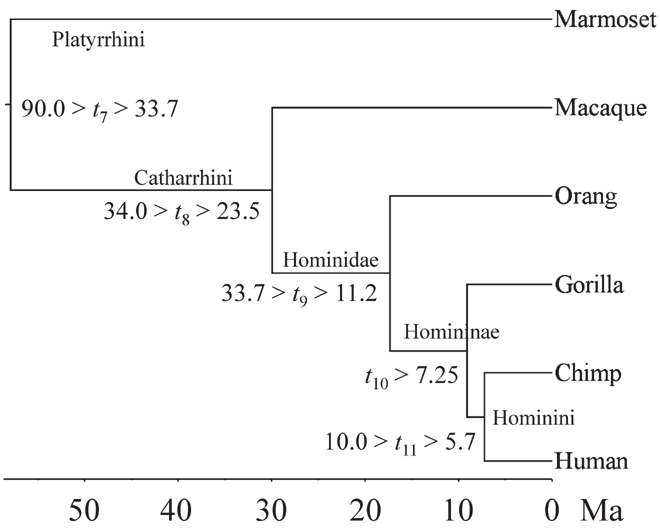


**Fig. 7.** The tree of six primate species showing fossil calibrations, and divergence times estimated with MCMCTREE, strict clock, HKY + $\Gamma_5$ using the alignment of 1st and 2nd codon positions. The alignment is over 8.7 million sites long, with all ambiguous sites removed. Species are: human (*Homo sapiens*), chimpanzee (*Pan troglodytes*), gorilla (*Gorilla gorilla*), orangutan (*Pongo abelii*), macaque (*Macaca mulatta*), and marmoset (*Callithrix jacchus*). Minimum bounds are 1% soft and maximum bounds are 5% soft. For node 10 (*Homininae*), a truncated Cauchy distribution is used as the calibration density with $p = 0.1$ and $c = 1$ (for details see Inoue et al., 2010).

To do similar calculations in other datasets, note that the $w$s are generated by a Bayesian analysis of the real data, and the $w_\infty$s can be calculated using the estimated distances $\hat{d}_i$ using Equation (16) (see Yang & Rannala, 2006). This equation may be hard to calculate analytically for most datasets, and the program MCMCTREE can be used to obtain $w_\infty$ using MCMC sampling (see the program's manual for details).

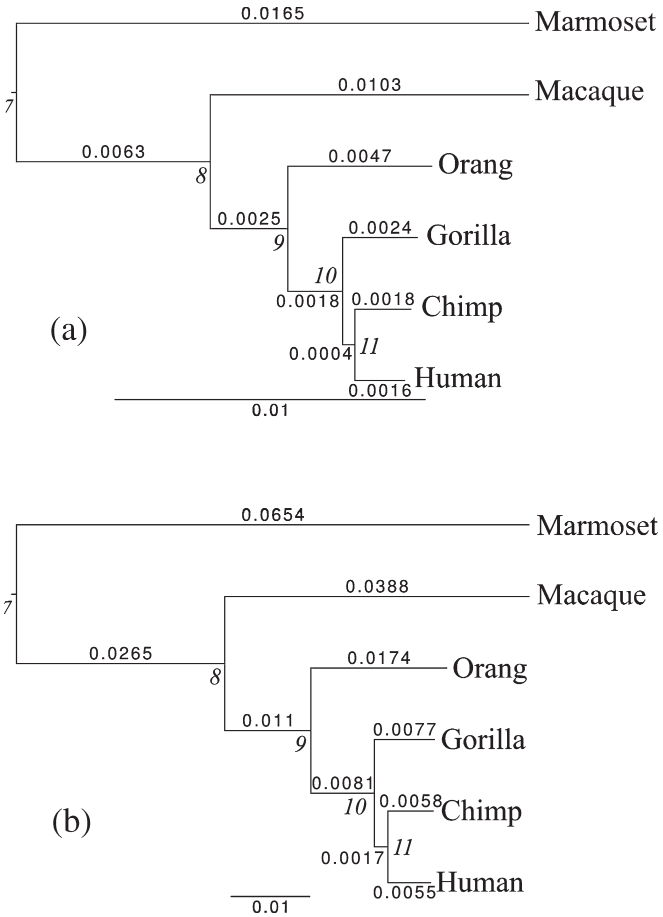


**Fig. 8.** The maximum likelihood estimates of branch lengths for the six-primate tree estimated under the HKY + $\Gamma_5$ model without assuming the clock, for (A) the alignment of 1st and 2nd codon positions, and (B) the alignment of 3rd codon positions. Branch lengths (in substitutions per site) are estimated using unrooted trees but the trees are here midpoint-rooted for clarity.

**Table 4** Priors used in the analysis of the primate data

| | 1st and 2nd | 3rd |
|---|---|---|
| $r$ | G(2, 8000) | G(2, 2000) |
| $\kappa$ | G(2, 0.5) | G(2, 0.3) |
| $\alpha$ | G(2, 20) | G(2, 4) |

## 3 Uncertainty of time estimation in the presence of conflicting fossil information

Our analysis above has assumed that correct fossil calibration information is used, and there is no conflict among fossils and between fossils and molecules. In this section we examine how the posterior CI width is affected by conflicts between fossil calibrations and by the use of one versus two fossil calibrations. Because of the identifiability problem of times and rates, conflicting fossil calibrations are expected to cause a great deal of problem.

We consider infinite sequence data evolving under a strict clock and use the infinite-sites theory (Eq. 16, see also Yang & Rannala, 2006) to calculate the limiting posterior distribution of times $t_1$ (the root age) and $t_2$ on the three-species tree of Fig. 5A. The true rate is $r = 1$, and the true ages of nodes are $t_1 = 2$ and $t_2 = 1$. In all cases the prior of the rate is $f_R(r) = \exp(-r)$ with mean 1. We consider uniform calibrations first with four scenarios of fossil calibrations (Fig. 11: A–D; Table 6): (a) Both nodes have good calibrations, that is, the mean of the calibration densities are the true times. (b) Only the root has a (good) calibration and node 2 has no calibration, so the prior on $t_2$ is uniform and conditioned on $t_1$: $t_2|t_1 \sim \mathrm{U}(0, t_1)$. (c) Node 2 has a good calibration and the root has a bad calibration (i.e., the mean of the calibration density is younger than the true root age). (d) Both nodes have bad calibrations. The mean of the calibration density for $t_2$ is too young, and that for $t_1$ is too old. We then consider gamma calibrations, with four corresponding cases (a′) to (d′) (Fig. 11: A′–D′; Table 6).

We first discuss the limiting posterior distributions for the four cases of uniform calibrations.

In case (a) (Fig. 11: A), the limiting posterior distribution of $t_1$ is

$$f(t_1|d_1 = 2) \propto \mathrm{e}^{-2/t_1} \times \mathrm{U}(t_1|1.8, 2.2) \times \mathrm{U}(t_1/2|0.9, 1.1), \qquad (19)$$

where $\mathrm{U}(t|a, b) = 1/(b - a)$ is the uniform density for $t$. Note that both calibrations on $t_1$ and $t_2$ include the true values right at the mid-points of the bounds, and also the interval width for $t_2$ is exactly half that for $t_1$. The (marginal) posterior distribution of times is very similar to the prior.

In case (b) (Fig. 11: B), the limiting posterior distribution of $t_1$ is

$$f(t_1|d_1 = 2) \propto \mathrm{e}^{-2/t_1} \times \frac{1}{t_1} \times \mathrm{U}(t_1|1.8, 2.2). \qquad (20)$$

Here only $t_1$ has a calibration. The posterior distribution of times is very similar to that in case (a), indicating that a single calibration is as good as two consistent calibrations. However we suggest that this result may be due to our assumption of the clock and of an infinite amount of sequence data. Without the clock assumption or with finite amount of data, it should be better to use multiple calibrations.

In case (c) (Fig. 11: C), the limiting posterior distribution of $t_1$ is

$$f(t_1|d_1 = 2) \propto \mathrm{e}^{-2/t_1} \times \mathrm{U}(t_1|1.6, 2.0) \times \mathrm{U}(t_1/2|0.9, 1.1). \qquad (21)$$

In this case, there is a good calibration on $t_2$, but the calibration on $t_1$ is too young and causes the posterior of both $t_1$ and $t_2$ to be highly concentrated compared with (a).

In case (d) (Fig. 11: D), the limiting posterior distribution of $t_1$ is

$$f(t_1|d_1 = 2) \propto \mathrm{e}^{-2/t_1} \times \mathrm{U}(t_1|1.99, 2.41) \times \mathrm{U}(t_1/2|0.81, 1.01). \qquad (22)$$

**Table 5** Posterior means and 95% CI's for divergence times for the primate dataset

| Node | | 1st and 2nd position | | | 3rd position | | |
|---|---|---|---|---|---|---|---|
| | | $t$ | 95% CI | $w$ | $t$ | 95% CI | $w$ |
| 7 | root | 57.0 | (46.8, 65.2) | 18.3 | 62.5 | (58.6, 66.8) | 8.21 |
| 8 | human/macaque | 30.3 | (24.9, 34.7) | 9.76 | 32.8 | (30.8, 35.1) | 4.30 |
| 9 | human/orang | 17.2 | (14.1, 19.7) | 5.54 | 17.6 | (16.5, 18.8) | 2.29 |
| 10 | human/gorilla | 8.89 | (7.31, 10.2) | 2.86 | 7.99 | (7.50, 8.55) | 1.05 |
| 11 | human/chimp | 7.05 | (5.79, 8.06) | 2.27 | 6.03 | (5.66, 6.45) | 0.79 |
| | rate ($\times 10^{-9}$/site/year) | 0.270 | (0.236, 0.327) | | 0.982 | (0.917, 1.040) | |

Note: Times are in My before present.

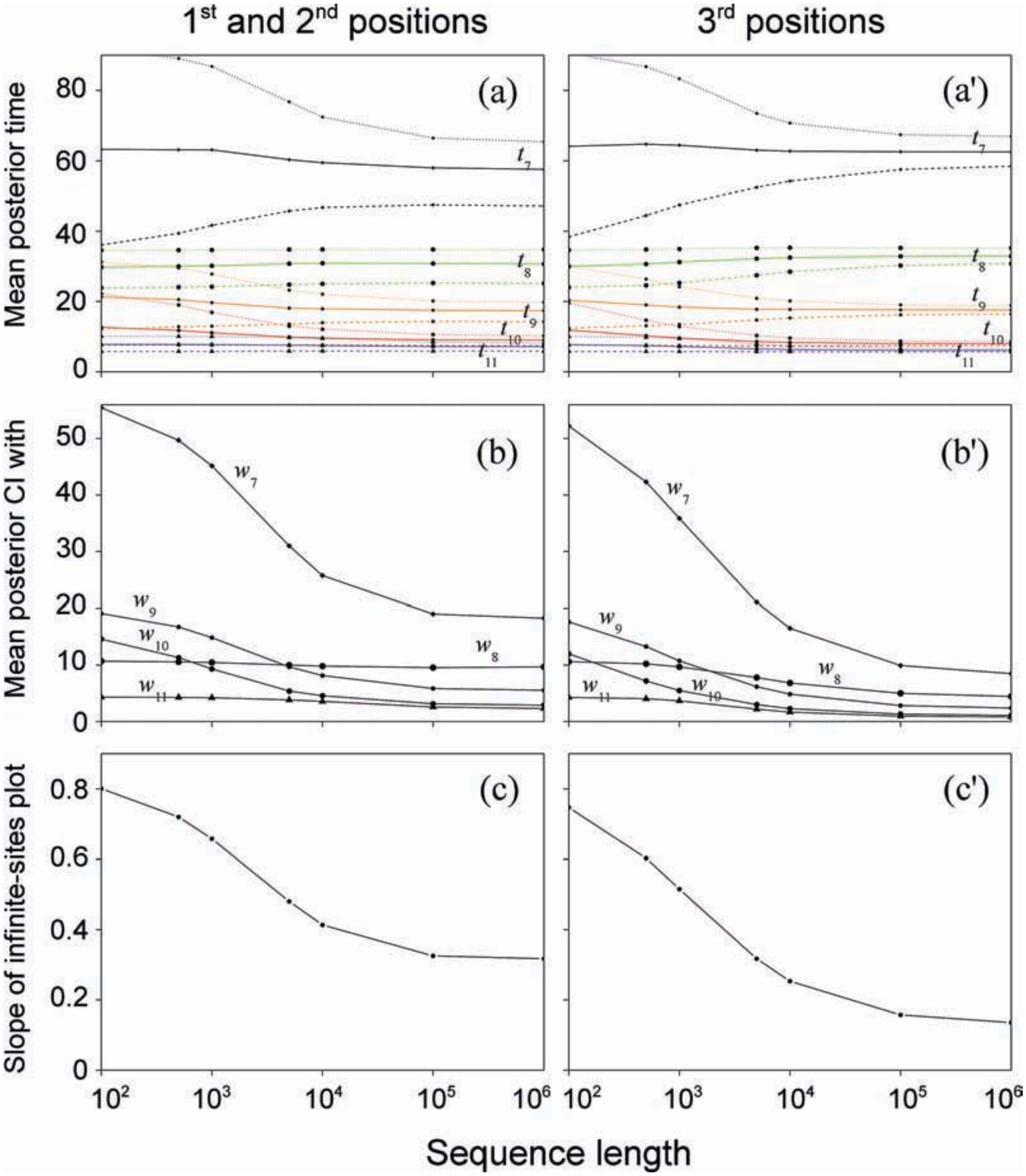


**Fig. 9.** Posterior distribution of divergence times in the primate tree of six species in random samples of sites from the original large alignments. In (A) and (A′) the posterior mean (continuous line) and 95% CI (dashed lines) for each time $t_i$ are plotted as a function of the sequence length. In (B) and (B′) the posterior CI width, $w$, is plotted as a function of the sequence length; and in (C) and (C′) the slope of the infinite-sites plot is plotted against the sequence length.

In this case, the calibration on $t_2$ is too young while that on $t_1$ is too old. The conflicting calibrations lead to extremely narrow posterior CIs.

The corresponding results for the gamma calibrations are shown in Fig. 11: A′–D′.

In case (a′) (Fig. 11: A′), the limiting posterior distribution of $t_1$ is

$$f(t_1|d_1 = 2) \propto \mathrm{e}^{-2/t_1} \times g(t_1|200,\ 100) \times g(t_2|200, 200). \tag{23}$$

In this case, both calibrations (on $t_1$ and $t_2$) are good and consistent. The posterior is more concentrated than the prior, which is different from the uniform case (a).

In case (b′) (Fig. 11: B′), the limiting posterior distribution of $t_1$ is

$$f(t_1|d_1 = 2) \propto \mathrm{e}^{-2/t_1} \times \frac{1}{t_1} \times g(t_1|200, 100). \tag{24}$$

Here there is only one calibration on $t_1$ and for that node the prior and posterior are nearly identical. Compared

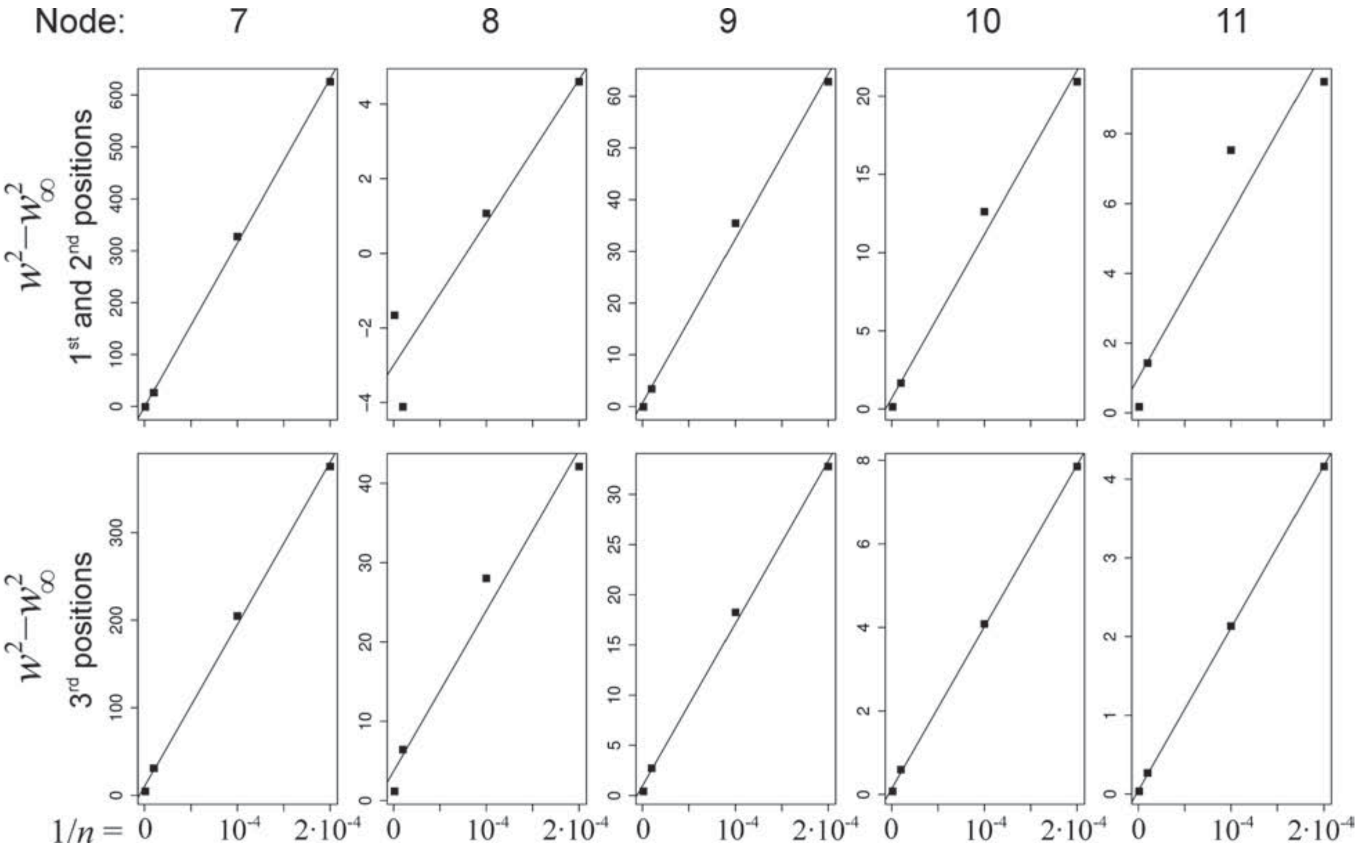


**Fig. 10.** Plot of $w^2 - w_\infty^2$ versus $1/n$ for the primate data set. All nodes show a good linear trend.

with case (a′), the posterior intervals are wider and use of one calibration is not as good as use of two. This pattern is different from that for the uniform calibration cases (a) and (b).

In case (c′) (Fig. 11: C′), the limiting posterior distribution of $t_1$ is

$$f(t_1|d_1 = 2) \propto \mathrm{e}^{-2/t_1} \times g(t_1|180, 100) \\ \times g(t_1/2|200, 200). \tag{25}$$

In this case, there is a good calibration on $t_2$ but the calibration on $t_1$ is too young. The poor calibration causes narrow CIs (compared with a′). However, the situation is not as bad as in the uniform case (c).

In case (d′) (Fig. 11: D′), the limiting posterior distribution of $t_1$ is

$$f(t_1|d_1 = 2) \propto \mathrm{e}^{-2/t_1} \times g(t_1|230, 100) \\ \times g(t_1/2|170, 200). \tag{26}$$

In this case the calibration on $t_1$ is too young, while the calibration on $t_2$ is too old. Interestingly, the posteriors look very reasonable, and centered around the true values. They also appear slightly more precise than in case (a′), where two good calibrations are used. The pattern is very different from the uniform case (d).

## 4 Discussion

Due to the confounding effect of rates and times, the posterior estimates of times will involve uncertainties even if an infinite amount of sequence data is used. In this paper we have developed a procedure for partitioning the uncertainty in posterior time estimates into two components, due to the uncertainty in the fossil calibrations (the prior for times and rates) and due to the finite nature of the sequence data, respectively. We have also suggested a measure $u_S$, which is the fraction of the uncertainty (variance) in the posterior time estimates attributable to limited sequence data. While $u_S$ goes to zero with the increase of the sequence data for all nodes in the tree, different nodes may behave very differently. For nodes with very precise calibrations the prior and posterior intervals are similar and $u_S$ is close to zero even for small amounts of sequence data. For nodes with very diffuse calibrations or with no calibrations, $u_S$ is initially large, but decreases very quickly with the increase of sequence data. The limiting posterior

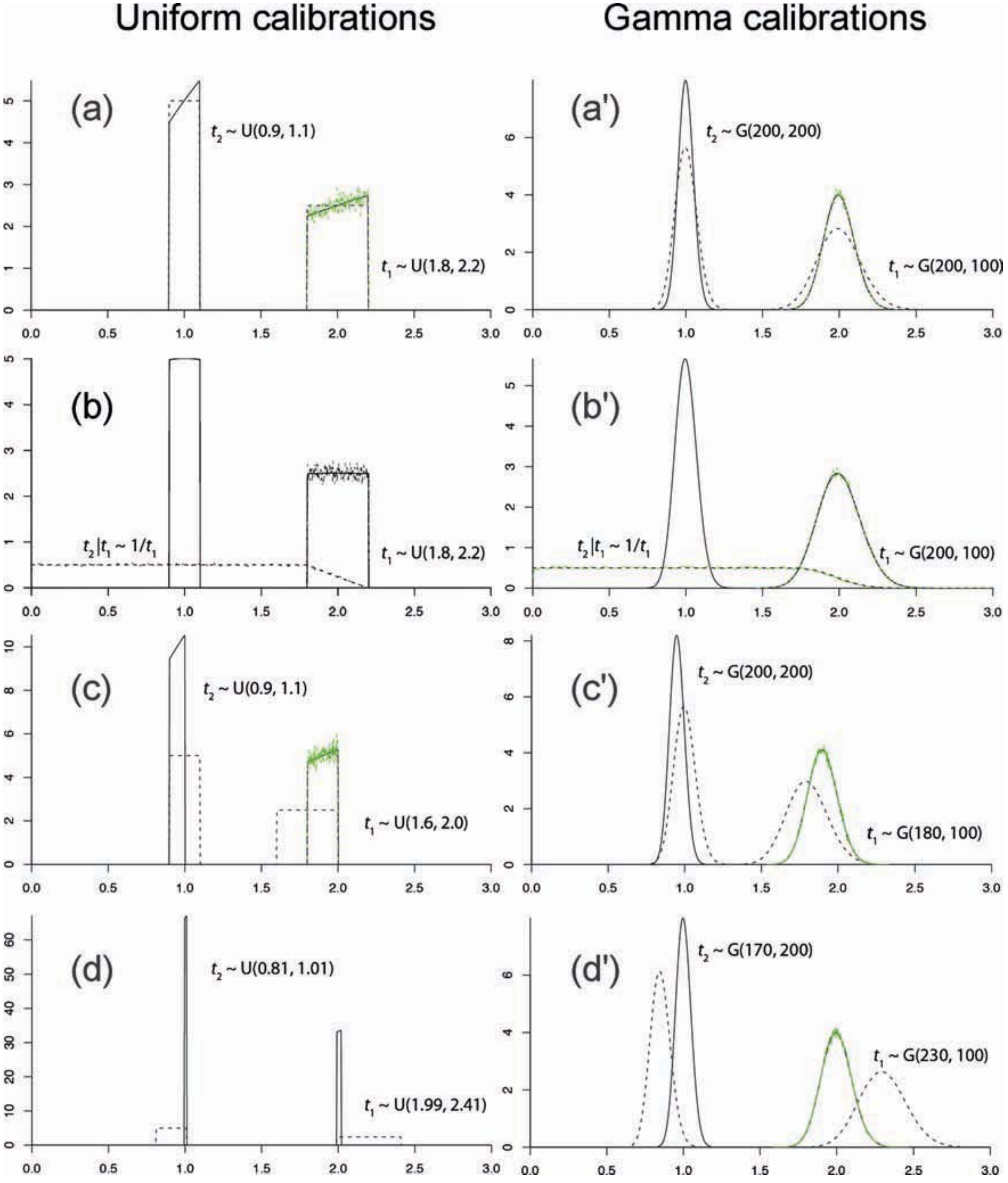


**Fig. 11.** Uncertainty in divergence time estimation for the three-species tree of Fig. 5A in the presence of conflicting fossil calibrations (uniform or gamma calibrations). The dashed line indicates the marginal prior distribution for a given time ($t_1$ or $t_2$), while the solid line indicates the corresponding marginal posterior distribution. The green wiggly line shows the results of calculating some of the priors (or posteriors) numerically by MCMC sampling (MCMCTREE). This was done for some of the distributions to confirm the accuracy of the analytical calculations. The true ages are $t_1 = 2$ and $t_2 = 1$. In (A) and (A′), two good calibrations are used on $t_1$ and $t_2$. In (B) and (B′) a single calibration on $t_1$ is used. In (C) and (C′), there is a bad calibration on $t_1$ and a good calibration on $t_2$. In (D) and (D′) both calibrations are poor: the one on $t_1$ is too old and the one on $t_2$ too young.

distribution of times, which is one-dimensional, is thus dominated by the nodes with precise calibrations, so that their accuracy is critical.

The finite-sites theory we developed in this paper is based on the molecular clock. Under relaxed-clock models (e.g., the auto-correlated and independent-rates models, Thorne et al., 1998; Drummond et al., 2006; Rannala & Yang, 2007; Linder et al., 2011), there is more uncertainty due to rate variation over lineages, which can be reduced by the use of a huge number of

**Table 6** Calibration densities used in the analysis of the three-species phylogeny of Fig. 5A

| Case | Uniform calibrations | Case | Gamma calibrations |
|---|---|---|---|
| (a) | $t_1 \sim \mathrm{U}(1.80, 2.20), t_2 \sim \mathrm{U}(0.9, 1.1)$ | (a′) | $t_1 \sim \mathrm{G}(200, 100), t_2 \sim \mathrm{G}(200, 200)$ |
| (b) | $t_1 \sim \mathrm{U}(1.80, 2.20)$ | (b′) | $t_1 \sim \mathrm{G}(200, 100)$ |
| (c) | $t_1 \sim \mathrm{U}(1.60, 2.00), t_2 \sim \mathrm{U}(0.9, 1.1)$ | (c′) | $t_1 \sim \mathrm{G}(180, 100), t_2 \sim \mathrm{G}(200, 200)$ |
| (d) | $t_1 \sim \mathrm{U}(1.99, 2.41), t_2 \sim \mathrm{U}(0.81, 1.01)$ | (d′) | $t_1 \sim \mathrm{G}(230, 100), t_2 \sim \mathrm{G}(200, 200)$ |

Note: The true node ages are $t_1 = 2$ and $t_2 = 1$.

loci evolving with different rate trajectories. In this regard, the molecular clock case may be considered a best-case scenario. Rannala & Yang (2007) showed that when both the number of sites and the number of loci increase to infinity, the posterior distribution of divergence times approaches a theoretical limit, as in the case of the molecular clock. The dynamics when the number of loci or the number of sites goes to infinity merits further study.

We have also examined whether using two consistent fossil calibrations leads to more precise posterior time estimates, compared with using only one calibration; two fossil calibrations are consistent if the calibration means are close to the true times, and if the calibrations have similar uncertainties (measured by the prior CI width to prior mean ratio). We found that the behavior is different for the uniform and the gamma calibrations. Use of multiple consistent uniform calibrations does not seem to improve the precision compared with use of a single calibration, but use of multiple gamma calibrations does. Conflicting fossil bounds are found to lead to very precise and overconfident posterior time estimates, and the bias is not corrected by the use of a huge amount of sequence data. In contrast, conflicting gamma calibrations lead to more reasonable posterior time estimates. Our results suggest that one should not automatically use uniform bounds as calibrations and it may be beneficial to use nonuniform probability curves such as the gamma, which may capture the information in the fossil record more accurately.

Despite the development of the inifinite-sites theory (Yang & Rannala, 2006; Rannala & Yang, 2007) which gives the limit of precision in Bayesian time estimation, the nature of the estimation problem does not appear to be well appreciated in many empirical dating analyses. For example, Mulcahy et al. (2012) observed that confidence intervals on ages estimated using the program BEAST were not significantly different when sampling 2 versus 25 loci for the reptile dataset they analyzed. The authors considered the result to be disturbing. Nevertheless it is not surprising. Even with infinite amount of sequence data, we will not reach full precision if the fossil calibrations involve uncertainties. In fact, infinite-sites plots in most dating analyses (e.g., Inoue et al., 2010; dos Reis et al., 2012) suggest that in a typical analysis much of the uncertainty is due to uncertainties in fossils, rather than limited amount of sequence data. Deriving reliable and precise calibration densities is thus extremely important to molecular dating analyses, and probabilistic modelling and statistical analysis of the fossil data appears to be the most promising approach (Wilkinson et al., 2011; Ronquist et al., 2012a). Similarly statistical methods that aim to investigate continuous trait evolution or species divergence rates should take into account the considerable uncertainty in divergence time estimates, rather than relying on point estimates.

**Acknowledgements** We thank two anonymous referees for many constructive comments, which have helped us to improve the readability of the paper. This study is supported by a grant from the Biotechnological and Biological Sciences Research Council (BBSRC). Z.Y. is a Royal Society-Wolfson Merit Award holder.

## References

dos Reis M, Inoue J, Hasegawa M, Asher RJ, Donoghue PC, Yang Z. 2012. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. Proceedings of the Royal Society B: Biological Sciences 279: 3491–3500.

Drummond AJ, Ho SY, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biology 4: e88.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7: 214.

Hasegawa M, Kishino H, Yano T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. Journal of Molecular Evolution 22: 160–174.

Inoue J, Donoghue PC, Yang Z. 2010. The impact of the representation of fossil calibrations on Bayesian estimation of species divergence times. Systematic Biology 59: 74–89.

Jukes TH, Cantor CR. 1969. Evolution of protein molecules. In: Munro HN ed. Mammalian protein metabolism. New York: Academic Press. 21–123.

Kishino H, Thorne J, Bruno W. 2001. Performance of a divergence time estimation method under a probabilistic model of rate evolution. Molecular Biology and Evolution 18: 352.

Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics 25: 2286–2288.

Linder M, Britton T, Sennblad B. 2011. Evaluation of Bayesian models of substitution rate evolution—parental guidance versus mutual independence. Systematic Biology 60: 329–342.

Mulcahy DG, Noonan BP, Moss T, Townsend TM, Reeder TW, Sites JW Jr, Wiens JJ. 2012. Estimating divergence dates and evaluating dating methods using phylogenomic and mitochondrial data in squamate reptiles. Molecular Phylogenetics and Evolution 65: 974–991.

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

Ronquist F, Klopfstein S, Vilhelmsen L, Schulmeister S, Murray DL, Rasnitsyn AP. 2012a. A total-evidence approach to dating with fossils, applied to the early radiation of the hymenoptera. Systematic Biology 61: 973–999.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012b. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61: 539–542.

Thorne JL, Kishino H, Painter IS. 1998. Estimating the rate of evolution of the rate of molecular evolution. Molecular Biology and Evolution 15: 1647–1657.

Wilkinson RD, Steiper ME, Soligo C, Martin RD, Yang Z, Tavare S. 2011. Dating primate divergences through an integrated analysis of palaeontological and molecular data. Systematic Biology 60: 16–31.

Yang Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. Journal of Molecular Evolution 39: 306–314.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Molecular Biology Evolution 24: 1586–1591.

Yang Z, Rannala B. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Molecular Biology and Evolution 23: 212–226.

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

# JSE

# Journal of Systematics and Evolution

Volume 51 Number 1 January 2013

Journal of Systematics and Evolution

Volume 51 Number 1 January 2013 Pages 1–116

## REVIEWS

## RESEARCH ARTICLES

**Cover illustration:** The posterior distribution of time *t* and rate *r* for a dataset of two sequences with 100 sites and 55 differences. Gamma priors are used for time and rate, while sequence distance is calculated under the Jukes-Cantor model. With more and more sequence data, the posterior of time and rate will converge to a curve (instead of a point). Because the sequence data provide information about distance only, posterior estimates of time and rate will have uncertainties even with an infinite amount of sequence data. The situation is the same when we date species divergences using uncertain fossil calibrations. See figure 3c in DOS REIS & YANG, pp. 30–43 in this issue.

ISSN 1674-4918
CN 11-5779/Q

WILEY-
BLACKWELL

ISSN 1674-4918
01>
9 771674 491135

Wiley-Blackwell