IQ2MC: A New Framework to Infer Phylogenetic Time Trees Using IQ-TREE 3 and MCMCtree with Mixture Models

Piyumal Demotte[1], Muthukumaran Panchaksaram[2], Hashara Kumarasinghe[1], Nhan Ly-Trong[1], Mario dos Reis[2,$], Bui Quang Minh[1,$]

[1] *School of Computing, College of Systems and Society, Australian National University, Canberra, ACT 2600, Australia*

[2] *Center for Evolutionary and Functional Genomics, School of Biological and Behavioural Sciences, Queen Mary University of London, London E1 4NS, United Kingdom*

[$] Corresponding authors:

Bui Quang Minh, m.bui@anu.edu.au

Mario dos Reis, m.dosreisbarros@qmul.ac.uk

ABSTRACT

IQ-TREE and MCMCtree are two widely used tools for inferring phylogenetic trees and estimating divergence times, respectively. As MCMCtree performs fast approximate Markov Chain Monte Carlo sampling to obtain the times along a fixed tree topology, it would be natural to use IQ-TREE to obtain the tree. However, integrating these tools seamlessly was not previously possible, as MCMCtree requires pre-calculation of the gradient and Hessian for fast approximate calculation of the likelihood, which were unavailable in IQ-TREE. Furthermore, MCMCtree only implements a small subset of substitution models, and complex models such as mixture models are not available. This is an important limitation because complex substitution models are required for reliable estimates of divergence times in deep phylogenies. Here, we introduce a new pipeline, IQ2MC, which facilitates the integration of IQ-TREE 3 and MCMCtree, substantially speeds up the pre-calculation steps, and allows the use of a wide range of IQ-TREE's complex models in divergence time inference. IQ2MC provides an updated version IQ-TREE 3.0.1 to calculate the gradient and Hessian at the maximum likelihood branch lengths, and then MCMCtree is used for MCMC sampling of divergence times. IQ2MC also provides several advanced partition models and mixture models not available in the current MCMCtree workflow. We finally show the applications of IQ2MC on simulated and four empirical datasets of placental mammals, plants, eukaryotes/prokaryotes, and metazoan. A tutorial on how to use IQ2MC is available at https://iqtree.github.io/doc/Dating.

Keywords: Bayesian phylogenetic dating, Markov Chain Monte Carlo, Maximum likelihood, Mixture models

Estimation of species divergence times is an integral part of understanding the evolutionary process among species (Marin et al. 2017; Betts et al. 2018; Álvarez-Carretero et al. 2022; Stiller et al. 2024). By placing speciation events in a geological timescale, we can obtain much richer information about patterns of evolution and diversification through time that cannot be obtained by using uncalibrated molecular trees. The current state-of-the-art is Bayesian molecular dating approaches (Thorne et al. 1998; Yang and Rannala 2006; Ronquist et al. 2012; Drummond and Bouckaert 2015), which allow incorporation of arbitrary fossil calibration densities to construct the prior on node ages, while integrating over sophisticated models of rate variation among branches of the phylogeny, such as the independent log-normal and geometric Brownian rate models (Thorne et al. 1998; Rannala and Yang 2007; Drummond and Rambaut 2007). Typically, Markov Chain Monte Carlo (MCMC) is used to approximate the posterior distributions of tree topologies, times, rates of evolution, and other model parameters.

Despite its flexibility, MCMC sampling is computationally expensive, particularly for phylogenies with hundreds or more species or when using molecular alignments that are very long. Therefore, approximation methods have been developed (Thorne et al. 1998; Guindon 2010; dos Reis and Yang 2011). Thorne et al. (1998) proposed to first infer the tree topology and then sample the divergence times on this fixed tree by using a fast approximation of the likelihood function based on the normal distribution. dos Reis and Yang (2011) further refined this approach by using the full Taylor approximation of the likelihood and by implementing transformations of the branch lengths to improve the accuracy of the approximation. The approximate method is currently implemented in the MCMCtree software as part of the PAML package (Yang 2007), and has been widely used to date phylogenomic datasets with hundreds of species such as plants, mammals, insects, and birds (Morris et al. 2018; Romiguier et al. 2022; Álvarez-Carretero et al. 2022; Stiller et al. 2024).

A typical MCMC analysis of a phylogenomic alignment with millions of sites using approximate likelihood requires weeks of computation time (e.g., Alvarez-Carretero et al. 2022), while using exact likelihood would require years (Battistuzzi et al. 2011; Alvarez-Carretero et al. 2022). However, an important limitation of the approximate method implemented in MCMCtree is that complex substitution models, such as mixture models (Lartillot and Philippe 2004; Quang et al. 2008). However, these models have been shown to be critical in the inference of deep divergences. For example, Wang and Luo (2025) showed that simple amino-acid substitution models recovered branch lengths that were shorter than those estimated under mixture models, thus impacting divergence time estimates when reconstructing the bacterial tree of life. (See also Wang and Meade 2026; Barba-Montoya et al. 2020; Tao et al. 2020). Therefore, there is a clear need for making complex models available for divergence time inference. Incidentally, because complex mixture models are much more computationally demanding than simpler models, they would benefit the most by being implemented under fast approximations, and thus making them readily available for phylogenomic analyses.

Here we introduce a new streamlined workflow, IQ2MC, to efficiently infer time trees by combining IQ-TREE 3 (Wong et al. 2025) and MCMCtree (dos Reis and Yang 2011). IQ2MC improves the efficiency of gradient and Hessian calculation (required in the likelihood approximation) compared to the workflow using PAML and, more importantly, IQ2MC allows the use of models of sequence evolution available in IQ-TREE 3 that are not supported in PAML (Yang 2007), including partition (Chernomor et al. 2016) and mixture models (Quang et al. 2008; Le et al. 2008; Ren et al. 2024). We discuss the IQ2MC workflow for time tree inference, present the technical details of the implementation, especially for partition and mixture models, and validate IQ2MC using simulated and four empirical

datasets. Our new IQ2MC workflow demonstrates the application and impact of advanced substitution models in IQ-TREE 3 for time tree inference using MCMCtree.

MATERIALS AND METHODS

IQ2MC workflow to infer time trees

We introduce IQ2MC, a streamlined workflow for efficient inference of time trees by combining IQ-TREE 3 (Wong et al. 2025) and MCMCtree. The workflow involves three steps: (1) phylogeny estimation; (2) estimation of gradient and Hessian; and (3) estimation of divergence times (Fig. 1). In the first step, the user provides a multiple sequence alignment (MSA) and IQ2MC first finds the best-fit model (Kalyaanamoorthy et al. 2017) and reconstructs a maximum likelihood (ML) tree with branch lengths in number of substitutions per site (Wong et al. 2025) (Alternatively, instead of constructing an ML tree, users can also provide a tree topology, and IQ2MC will estimate the branch lengths). In the second step, IQ2MC computes the gradient vector and Hessian matrix of the log-likelihood function at the maximum likelihood estimates (MLEs) of the branch lengths on the fixed tree topology (this is done using the newly implemented option --dating mcmctree in IQ-TREE v3.0.1). Step 2 will automatically generate a control input file for MCMCtree, which users can optionally modify to set appropriate parameter values such as those required for MCMC convergence (Yang and Rodríguez 2013), the birth-death-process prior (Rannala and Yang 1996), or the evolutionary rate model (Rannala and Yang 2007; dos Reis et al. 2014). In the third step, users can then run MCMCtree using the control file, tree file (which the user edits to add the appropriate fossil calibrations), and the gradient and Hessian (stored in the in.BV/.Hessian file) to obtain the MCMC samples of divergence times and evolutionary rates. Previously, step 2 was done using the BaseML and CodeML programs of the PAML package, and thus only those substitution models available in PAML could be used (but see Wang and Luo

(2025) for an approximate alternative). IQ2MC is a convenient and much faster drop-in replacement for PAML and allows users to do everything with a few command lines.

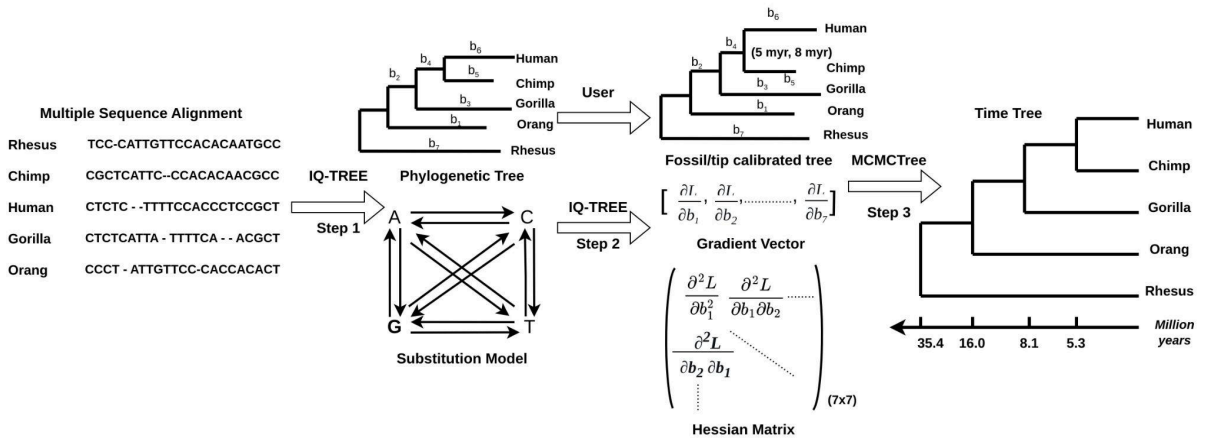

Figure 1. General workflow of IQ2MC for inferring time trees with IQ-TREE 3 and MCMCtree. Step 1 infers the best-fit substitution model and a maximum likelihood phylogenetic tree from a given multiple sequence alignment using IQ-TREE 3. Step 2 computes the gradient vector and the Hessian matrix of the log-likelihood function at the maximum likelihood parameter estimates using IQ-TREE 3. Step 3 constructs a time tree using MCMCtree with the user-provided time calibrations of internal nodes/tips.

Taylor expansion of the Log-likelihood

Calculation of Felsenstein's phylogenetic likelihood over site-patterns in an MSA is computationally expensive, and the computational cost grows proportionally with the number of taxa and site-patterns in the MSA, becoming prohibitive for phylogenomic alignments with millions of sites. Therefore, MCMCtree approximates the log-likelihood during MCMC sampling by its second-order Taylor expansion evaluated at the MLEs of the phylogeny's branch lengths (dos Reis and Yang, 2011). Computation of the approximate likelihood only depends on the number of taxa in the phylogeny and not on the length of the MSA, and it is therefore much faster than exact likelihood in timetree inference (Battistuzzi et al., 2011).

The approximation works on a fixed tree topology and by fixing the substitution model parameters to their maximum-likelihood estimates.

Let T be an unrooted tree topology with n taxa, and let $B = (b_1, b_2, \ldots, b_{2n-3})$ be the vector of branch lengths of T in units of substitutions per site. Let $\widehat{B}$ be the vector of maximum likelihood estimates (MLEs) of B, obtained under a substitution rate matrix Q, which applies to all sites in the MSA. The matrix Q is of size 4×4 for DNA and 20×20 for protein alignments. We can apply Taylor expansion (Thorne et al. 1998; dos Reis and Yang 2011) to approximate the log-likelihood of B given the alignment D:

$$\log L(D;B) \approx \log L(D;\widehat{B}) + g(D,\widehat{B})^T(B-\widehat{B}) + 1/2\,(B-\widehat{B})^T H(D,\widehat{B})(B-\widehat{B}). \quad (1)$$

Where $g(D,\widehat{B})$ and $H(D,\widehat{B})$ are the gradient vector and Hessian matrix of the log-likelihood function evaluated at the branch length MLEs.

Technically, one needs to compute the log-likelihood, first derivative, and second derivative of the log-likelihood for each site pattern in the MSA, which are then combined into $\log L(D;\widehat{B})$, $g(D,\widehat{B})$ and $H(D,\widehat{B})$. The Taylor approximation works well when B is close enough to $\widehat{B}$ during the MCMC sampling. If not, one needs to apply transformations of the branch lengths (dos Reis and Yang 2011). As this equation only applies to homogeneous substitution models over sites, we now extend it to partition and mixture models.

Edge-linked partitions model with proportional branch lengths

For phylogenomic data with multiple genes or loci, it is common to use partition models (Lanfear et al. 2014; Kainer and Lanfear 2015; Chernomor et al. 2016), where each partition represents a subset of sites and different partitions may have different substitution models. The likelihood for each partition needs to be calculated separately for dating under MCMCtree.

Edge-linked (or branch-linked) partition models (-p option in IQ-TREE 3) estimate a single set of branch lengths B shared across all partitions. However, each partition i is assigned a partition-specific rate r_i that rescales the shared branch lengths. The mean of the r_i is fixed so that it is equal to one. Thus, a low r_i means the partition is slowly evolving, whereas a high r_i means the partition is fast evolving. For each of the p partitions, D_1, D_2, ..., D_p, we compute the required gradient vector $g(D_i, r_i\widehat{B})$ and Hessian matrix $H(D_i, r_i\widehat{B})$. Since all partitions share the same set of branch lengths, we can obtain a single gradient vector and Hessian matrix by scaling each partition-specific gradient and Hessian by r_i and r_i^2 respectively, and summing them: $g(D, \widehat{B}) = \sum_i r_i\, g(D_i, r_i\widehat{B})$ and $H(D, \widehat{B}) = \sum_i r_i^2\, H(D_i, r_i\widehat{B})$. We then apply equation (1) to approximate the log-likelihood under the edge-linked partition model.

Edge-unlinked partition models

Edge-unlinked (or branch-unliked) partition models (-Q option in IQ-TREE 3) allow a different set of branch lengths for each partition: B_i for partition D_i. Consequently, we estimate a separate gradient vector $g(D_i, \widehat{B_i})$ and a Hessian matrix $H(D_i, \widehat{B_i})$ for each partition i. For p partitions D_1, D_2, ..., D_p, we approximate the per-partition log-likelihood $\log L(D_i, B_i)$ for each partition i (Eq. 1), and sum them up to obtain the total approximate log-likelihood:

$$\log L(D; B) \approx \sum_{i=1}^{p} \{\log L(D_i; \widehat{B_i}) + g(D_i, \widehat{B_i})^T (B_i - \widehat{B_i}) + 1/2\, (B_i - \widehat{B_i})^T H(D_i, \widehat{B_i})\, (B_i - \widehat{B_i})\}. \quad (2)$$

Note the in.BV/.Hessian file generated with edge-unlinked models will contain p blocks, corresponding to the required p gradient vectors and p Hessian matrices for each partition.

Mixture models

Single and partition models assume a single rate matrix Q for all sites in a partition or an MSA. However, this is known to be biologically unrealistic and can cause biased phylogenetic estimates (Lartillot et al. 2007), and thus a number of mixture models have been introduced (Quang et al. 2008; Le et al. 2008, 2012; Wang et al. 2018; Ren et al. 2024). A mixture model typically consists of several substitution matrices $\{Q_1, Q_2, \ldots, Q_k\}$, the so-called classes. Each site in the MSA has a certain probability (or weight) of belonging to a class. Because we do not know which class a site belongs to, site likelihoods are marginalized by their weighted sum over the substitution classes in the mixture. A benefit of mixture models compared with partition models is that users do not need to manually partition the sites into subsets.

Mixture models in phylogenetics include models of rate variation among sites (Yang 1994, 1995; Soubrier et al. 2012), variation in amino acid stationary frequencies among sites (Quang et al. 2008; Wang et al. 2018), and variation of amino acid or DNA substitution rate matrices among sites (Le et al. 2008, 2012; Ren et al. 2024). In this section, we mainly focus on the computation of g and H for the latter ones, but similar calculations also apply to other types of mixture models.

Let $\{Q_1, Q_2, \ldots, Q_k\}$ be the k rate matrices in the mixture model with corresponding weights $\{w_1, w_2, \ldots, w_k\}$, where $\sum_{j=1}^{k} w_j = 1$. The log-likelihood under the mixture model is the sum over the log-likelihoods for sites and substitution rate matrices:

$$\log L(D;B) = \sum_s \log \{ \sum_{j=1}^{k} w_j L(D_s; B, Q_j) \} \quad (3)$$

, where D_s is the s-th site in the alignment D.

Yang (2000) gives the derivatives of the log-likelihood function under a single substitution matrix. Here we provide the derivation of the gradient and the Hessian under a mixture of rate matrices. Let $f_s = L(D_s, B) = \sum_{j=1}^{k} w_j L(D_s; B, Q_j)$ be the site-likelihood under a mixture model. The first and second-order partial derivatives of f_s with respect to the branch lengths B are

$$\frac{\partial f_s}{\partial B} = \sum_{j=1}^{k} w_j \frac{\partial L(D_s;B,Q_j)}{\partial B}, \quad (4)$$

and

$$\frac{\partial^2 f_s}{\partial B \partial B^T} = \sum_{j=1}^{k} w_j \frac{\partial^2 L(D_s;B,Q_j)}{\partial B \partial B^T}. \quad (5)$$

From Eq. (3), $\log L(D;B) = \sum_s \log \{f_s\}$. Therefore, the gradient and Hessian of the total log-likelihood are

$$g(D,B) = \sum_s \frac{\frac{\partial f_s}{\partial B}}{f_s}, \quad (6)$$

and

$$H(D,B) = \sum_s \frac{f_s(\frac{\partial^2 f_s}{\partial B \partial B^T}) - (\frac{\partial f_s}{\partial B})^2}{(f_s)^2}. \quad (7)$$

Implementation in IQ-TREE 3

As part of the maximum likelihood branch length optimization, IQ-TREE 3 already computes the gradient and the diagonal of the Hessian matrix using Eqs. (6) and (7),

respectively. IQ-TREE 3 now additionally calculates the off-diagonal elements of the Hessian matrix by using the outer-product-of-scores (OPS) estimator (Porter 2002; Seo et al. 2004). The OPS estimator is necessary because estimates of the off-diagonal elements by the difference method (Eq. 7) are not stable (Porter 2002; Seo et al. 2004). Here, we first construct a matrix, G, whose columns correspond to the gradient vectors of each site in the alignment calculated at maximum-likelihood estimates of branch lengths using Eq (6). Thus, if the unrooted phylogeny has n species and the alignment has s sites, then matrix G has 2n – 3 rows and *s* columns. The OPS estimator of the Hessian is:

$$H \approx - GG^T. \quad (8)$$

However, we note the OPS estimator does not produce the correct diagonal elements of H for branch lengths with MLEs equal to zero (dos Reis and Yang, 2011), thus we use Eq. (7) to calculate the diagonal elements of the H using the difference method. This approach of using Eq. (7) for the diagonal and (8) for the off-diagonal of the Hessian matrix provides the best estimate of the complete Hessian matrix (dos Reis and Yang, 2011). Given the alignment of n species, s sites and a mixture of k Q matrices with having p states, the time complexity of estimating the gradient and Hessian is $O(nskp^2 + n^2 s)$.

For partition models, we implemented Eqs. (1) and (2) for edge-linked and edge-unlinked models. For m partitions on a fixed topology of n species, where partition j has n_j species, s_j sites, and a single p_j state rate matrix, edge-linked partitions models have a time complexity of $O(\sum_j n_j s_j p_j^2 + n^2 \sum_j s_j)$ while edge-unlinked partition models have a time complexity of $O(\sum_j n_j s_j p_j^2 + \sum_j n_j^2 s_j)$ for gradient and Hessian computation.

Simulated data

We conducted the following simulations to test IQ2MC. We used four tree topologies: two unbalanced trees and two birth-death trees, each with 16 and 64 taxa (Supplementary Figs. S1- S4). Birth-death trees were simulated using the TreeSim package (Stadler 2011) with speciation and extinction rates of 2 and a sampling fraction of 0.1. For both the unbalanced and birth-death trees, the root age was fixed at 3 time units. These topologies were used to simulate 24 DNA alignments with 5,000 sites using AliSim-HPC (Ly-Trong et al. 2023) under a strict clock and the GTR+G4 model with $\alpha = 0.25$ (Tavaré 1986; Yang 1994). The 24 alignments were generated under six different sequence divergences: 0.1, 0.2, 0.4, 0.6, 0.8, and 1.0 nucleotide substitutions per site per unit time for each tree topology, henceforth denoted as 1x, 2x, 4x, 6x, 8x, and 10x, respectively.

We also used the four tree topologies to simulate 168 amino acid alignments under a combination of the 6 divergence levels above and 7 substitution models: LG+G4 (Le and Gascuel 2008) and six profile mixture models; LG+G4+C10 to LG+G4+C60 (Quang et al. 2008) using AliSim-HPC (Ly-Trong et al. 2023). All alignments have 5,000 sites and the same gamma shape ($\alpha = 0.25$). We then validated IQ2MC by running IQ2MC and PAML on all DNA and amino acid alignments using GTR+G4 and LG+G4 models, respectively, to verify that both tools produce the same results, including estimated branch lengths, gradient vector, and Hessian matrix.

Next, we assessed the impact of model violation and varying the placement of informative calibrations on the nodes of the phylogenies. We hypothesize that if a phylogeny only has an informative calibration on a young node, then large discrepancies will be observed in the estimated age of the root when using simple vs. complex substitution models, because simpler models tend to underestimate longer branches more than shorter ones. Each alignment was analyzed under three fossil calibration settings, depending on where in the tree

the most informative calibration is placed: young calibration (placed on a young node), middle calibration (placed in the middle node), and root calibration (placed on the root node). Note that MCMCtree always requires a calibration on the root (to condition the birth-death process on the root age, see Yang and Rannala, 2006), therefore, for the young and middle calibration settings, the root node has a diffuse calibration. For example, in the 16-taxon unbalanced tree (Supplementary Fig. S1), under the young calibration setting, the age of the youngest internal node (t_{15}) is calibrated with a narrow uniform distribution B(0.18, 0.22), and the root node is (t_1) calibrated with a gamma density G(2, 0.66), which is a diffuse calibration with a large variance. Thus, in this case, the youngest node has the most informative calibration. For the middle calibration, the middle node (t_{10}) is calibrated with a narrow uniform distribution B(1.08, 1.32), and the root node (t_1) is calibrated with the same diffuse gamma distribution as above, G(2, 0.66). Thus, in this case, the middle node has the most informative calibration. Finally, for the root calibration setting, the root (t_1) is calibrated with a narrow uniform distribution, B(2.7, 3.3). These calibrations were applied to mimic different calibration scenarios seen in real data analysis, in which the most informative node calibration may be present in different parts of the tree. For all simulations under each tree topology, we used $t \pm t/10$ to define the lower and upper bounds of the uniform calibration, where t is the true age of the calibrated node (Supplementary Tables S1- S4). For the birth-death trees, multiple calibrations were applied to randomly selected internal nodes under the young and middle calibration schemes, to emulate common scenarios in real data analysis in which many calibrations are present throughout the phylogeny (Supplementary Figs. S2 and S4).

For the 24 DNA alignments, we then ran IQ2MC under the GTR+G4 model (which was used to simulate the alignments) and compared the results with those under the simpler and incorrect JC69 model (Jukes and Cantor 1969). For the 168 amino acid alignments

simulated under profile mixture models, we compare the results obtained using IQ2MC with the correct profile mixture model vs using the simpler LG+G4 model. All simulations were performed with five replicate runs to account for uncertainty in the estimates.

Simulation with site-specific amino-acid profiles

To further assess the accuracy of amino acid profile mixture models (Quang et al. 2008) for time-tree inference relative to site-homogeneous models under realistic simulation conditions, we simulated alignments in which each site had its own amino acid profile. These simulations are important because amino acid mixture models have been shown to produce different branch lengths for ancient phylogenies and thus impact divergence time estimates (Wang and Luo 2025), which in turn has been shown to be critical for reconstructing ancient evolutionary history, such as the origin of bacteria (Davín et al. 2025). For these simulations, we reused the four tree topologies and fossil calibration schemes: young, middle, and root, used in the previous simulations.

We extracted 500, 1,000, 2,000, and 5,000 independent empirical amino acid profiles, estimated using the posterior mean site frequency (PMSF) method (Wang et al. 2018), from Giacomelli et al. (2025). Using these profiles, we simulated alignments of 500, 1,000, 2,000, and 5,000 sites with AliSim-HPC (Ly-Trong et al. 2023) under each tree topology, the LG+G4+PMSF model, and two levels of sequence divergence: 0.1 and 1.0. We refer to these as the 1x and 10x rate settings, respectively. In total, 32 amino acid alignments were simulated, and divergence times were then estimated under the LG+G4 and LG+G4+C60 models. These analyses allowed us to evaluate whether profile mixture models improve divergence-time estimation relative to the site-homogeneous LG+G4 model under challenging conditions intended to resemble real phylogenomic datasets.

Evaluating the quality of the Taylor approximation to the log-likelihood under mixture models

Under standard substitution models, the second-order Taylor expansion approximates the log-likelihood curve well (dos Reis and Yang 2011). Here we assess the quality of the approximation for the more complex mixture models. We simulated a two-taxon alignment of 500 sites under the LG+G4+C60 model with a gamma shape ($\alpha = 0.5$), then expanding it to obtain two other alignments of 5,000 and 50,000 sites. Simulations were performed using three branch lengths (pairwise distances): 0, 0.05, and 1.0. The aim was to investigate the accuracy of the approximate log-likelihood under a profile mixture model, as well as the effect of the branch-length transformations proposed by dos Reis and Yang (2011). For each simulated alignment, we computed the approximate likelihood curve under no transform (NT), using the arcsine-based transform (ARCSIN) and the square-root transform (SQRT). We compared the approximations to the exact likelihood calculated under the profile mixture model using IQ-TREE 3.

We tested the approximation with larger alignments of more than two taxa. We used a 16-taxon unbalanced tree and a 16-taxon birth–death tree, with uniform fossil calibrations (Supplementary Tables S1 and S3). For each tree, we simulated eight alignments ranging in length from 500 to 1,000,000 sites under the LG+G4+C60 model with a gamma shape parameter of 0.5. For each MCMC run, we randomly obtained 100 subsamples under the independent log-normal rate model, taken from the stationary phase of the MCMC chain. These samples were used to calculate the approximate log-likelihood under the ARCSIN-based transform and the exact log-likelihood under the LG+G4+C60 model in IQ-TREE 3. We measure the relative mean error between the approximate (l^{approx}) and the exact log-likelihoods (l^{exact}), i.e., $|l^{approx} - l^{exact}|/|l^{exact}|$ across m=100 samples. We also

compute the Spearman rank correlations of the log-likelihoods between the exact and the approximation strategy. In all simulation analyses, convergence was assessed by ensuring that effective sample sizes (ESS) exceeded 200 for all sampled parameters after discarding the first 10% of samples as burn-in, and by confirming that two independent MCMC runs produced concordant posterior histograms for the sampled parameters.

Empirical data

To demonstrate the impact of using the new partition and mixture models available in IQ2MC, we used IQ2MC to reanalyze four empirical datasets (Table 1) originally analyzed with the existing PAML workflow. We selected these four empirical datasets as they span a broad range of clades and evolutionary timescales across the tree of life, including plants (~798 Ma), animals (~ 600- 800 Ma), mammals (~250 Ma; Afrotheria subtree), and a deep phylogeny of bacteria and early eukaryotes (~4,500 Ma). This diversity allows us to evaluate the performance of IQ2MC under varying levels of sequence divergence, rate heterogeneity, and phylogenetic depth. In addition, we included two nucleotide (plants and mammals) and two amino acid (metazoan and bacteria and early eukaryote) datasets to demonstrate the applicability of the method across different data types and substitution models. For all analyses, we used the published tree topologies and four calibration settings: young, middle, root, and all calibrations as used in the original papers. For each setting, we always set a calibration on the root node of the tree as required by MCMCtree, whereas in the young and middle calibrations, a calibration is additionally added to either one node closest to the tips or one node in the middle of the tree, respectively. Supplementary Tables S5-S8 show the detailed calibrations used.

Table 1: Datasets used in the empirical analyses.

Dataset	Datatype	Taxa	Alignment partitions	Sites	Fossil calibrations	Reference
Placental mammals/ Afrotheria subtree	DNA	60	5	262,910	10	(Álvarez-Carretero et al. 2022)
Plants	DNA	103	1	2,217	37	(Morris et al. 2018)
Eukaryotes and Prokaryotes	Amino Acid	102	29	9,874	11	(Betts et al. 2018)
Metazoans	Amino Acid	54	10	38,577	34	(dos Reis et al. 2015)

For three datasets (placental mammals/afrotheria subtree, metazoans, eukaryotes and prokaryotes), we ran IQ2MC under combinations of two substitution models and two partition models (giving four model combinations): single and mixture models, edge-linked partition models, and edge-unlinked partition models. Whereas for the plant dataset, we used single and mixture models (Supplementary Tables S9- S13). For single models, we used the GTR+G4 and LG+G4 for DNA and amino acid alignments, respectively. For edge-linked and edge-unlinked partition models, ModelFinder (Chernomor et al. 2016; Kalyaanamoorthy et al. 2017) was used to determine the best-fit models for each partition. That is, different partitions may use different substitution models (say HKY+G4 in one partition and GTR+G4 in another). For mixture models, we used MixtureFinder (Ren et al. 2024) for DNA alignments and LG+G4+C60 for amino acid alignments.

For comparison, we also ran PAML under single and edge-unlinked partition models because it does not support edge-linked partition and mixture models. The PAML pipeline

always produced the same results as the IQ2MC pipeline when the models matched, and thus, we do not report the PAML results. For all the above IQ2MC analyses, we used the autocorrelated rates model, the same priors on the mean rate, rate diffusion, and calibration densities as the original publications.

We further assessed the accuracy of the log-likelihood approximation under profile mixture models using the metazoans, and eukaryotes and prokaryotes empirical datasets. For each dataset, we extracted 200 posterior samples from the stationary phase of the MCMCtree analyses under the LG+G4+C60 model, using both the autocorrelated-rates clock and the independent-rates clock. The analyses used the same fossil calibration settings as in the original dating analyses. For each sampled time tree, we then computed the exact log-likelihood under the LG+G4+C60 model using IQ-TREE 3 and compared it with the corresponding approximate log-likelihood.

For all the above analyses, MCMCtree v4.10.7 (Rannala and Yang 2007) was used with the default ARCSINE transform on the branch lengths (dos Reis and Yang 2011). For each combination of datasets and models, we carried out two independent MCMC runs, where the number of MCMC iterations was manually determined to ensure convergence (supplementary Table S14). That is, we ensured the effective sample size (ESS) was at least 200 for all sampled parameters after removing at least 10% of the samples as burn-in, and that the two MCMC independent runs resulted in congruent histograms for sampled parameters. For each comparison of divergence times estimated using partition/mixture models versus the single models, we calculated the slope and R^2 from a linear regression.

RESULTS

IQ2MC allows parallel gradient and Hessian calculation for advanced models of sequence evolution such as mixture models

The IQ2MC workflow (Fig. 1) implements multi-threading calculations of the gradient and Hessian matrix for all models of sequence evolution in IQ-TREE 3, including all DNA, protein, codon, binary, and morphological models, which results in much faster optimization routines than those implemented in PAML. This is particularly striking for the optimization of large amino acid phylogenies, which can take several days with PAML but only a few hours with IQ-TREE (Supplementary Table S9). Moreover, IQ2MC supports edge-linked (EL) and edge-unlinked (EUL) partition models (Chernomor et al. 2016), whereas BaseML/CodeML is limited to edge-unlinked models. More importantly, IQ2MC supports mixture models (Quang et al. 2008; Le et al. 2008, 2012; Wang et al. 2018; Ren et al. 2024), which are especially useful for deep divergence dating but not available in BaseML/CodeML. Compared with the current PAML workflow (Yang 2007; Álvarez-Carretero et al. 2022), this integration significantly broadens the scope and applicability of advanced models in phylogenetic dating with MCMCtree.

IQ2MC accurately computes the gradient and Hessian under different simulation conditions

For DNA alignment simulations, estimated branch lengths, gradients, and Hessians using the GTR+G4 model in the IQ2MC workflow were virtually identical to those from PAML, except for tiny numerical differences due to the different optimization algorithms used in the two programs and due to the use of random starting branch lengths (Supplementary Fig. S5). Both programs also produced virtually identical parameter estimates for the simulated amino acid alignments under the LG+G4 model (Supplementary Fig. S6). For profile mixture models (LG+G4+C10 to LG+G4+C60), branch lengths were estimated only with IQ2MC, as PAML lacks support for these models. These estimates also show strong agreement with the true values used in the simulation (Supplementary Fig. S7).

Furthermore, we compared the divergence times estimated between the two workflows for GTR+G4 (Supplementary Fig. S8) and LG+G4 (Supplementary Fig. S9) models, and the results are essentially the same.

Likelihood approximation works well under mixture models

dos Reis and Yang (2011) analyzed the accuracy of the approximation for simple substitution models under transforms for the case of a branch length equal to zero, a branch length within parameter space, and a long branch that generates a saturated tail in the likelihood. Here we repeat those analyses but using the LG+G4+C60 mixture model. Our results show the approximation, particularly under the ARCSINE-based transform, is excellent for the long alignments. We note the Taylor expansion is an asymptotic approximation, and the approximation improves as the sample size (in our case the number of sites in the alignment) increases (Thorne et al. 1998; Seo et al. 2004). For all three branch lengths (short, Figs. 2A-A'', median, Figs. 2B-B'', and long, Figs. 2C-C''), the approximation improves with sample size, although the rate of asymptotic improvement is slower for the longer branches, as expected. The approximation is also best near the likelihood peak, as expected. We note that when carrying out MCMC sampling, proposals on the tails of the likelihood tend to be rejected, as the proposed branch lengths are far from the likelihood peak and are thus unlikely. This means errors in the tails tend to be less important (see discussion in dos Reis and Yang, 2011). These results demonstrate the Taylor approximation is suitable for MCMC inference under the LG+G4+C60 mixture model under phylogenomic alignments. We note that for short alignments (i.e., on the order of 500- 1,000 sites), the approximation error may be noticeable and may affect the posterior sampling. For those cases, exact sampling may not be too slow, and we recommend users use one of the exact mixture model

implementations available in, for example, PhyloBayes for divergence time estimation (Lartillot and Philippe 2004).

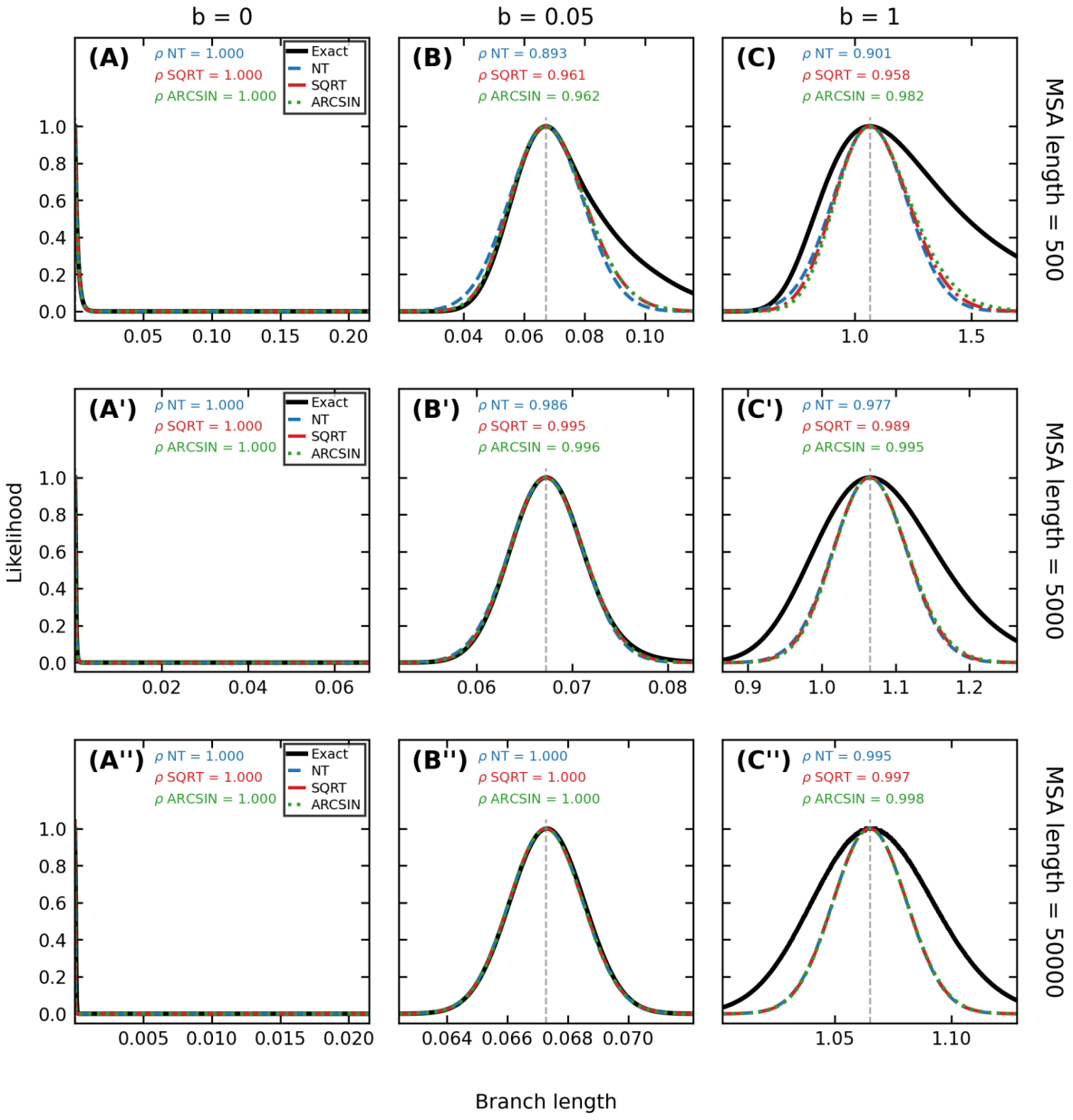

Figure 2. Accuracy of likelihood approximation under the LG+G4+C60 profile mixture model for two-taxon simulations. Exact and approximate likelihood curves were compared for two-taxon amino acid alignments simulated under three branch-length settings, b = 0, b = 0.05, and b = 1, with alignment lengths of 500, 5,000, and 50,000 sites. Approximate likelihoods were computed without transformation (NT), with the square-root transformation (SQRT), and with the arcsine transformation

(ARCSIN). The vertical dashed line indicates the maximum-likelihood estimate of the branch length. ρ represents Spearman rank correlation.

We next evaluated the accuracy of the approximation under the LG+G4+C60 profile mixture model for large trees (Fig. 3). For both the birth–death and unbalanced tree simulations, the relative mean absolute error (RMAE) decreased consistently as the alignment length increased from 500 to 1 million amino acid sites, as expected. The largest approximation errors were observed for the shortest alignments, whereas substantially lower errors were obtained for longer alignments. At 500,000 to 1,000,000 sites, both tree shapes showed similar and low RMAE values, indicating that the approximation is, for practical purposes, essentially identical to the exact values.

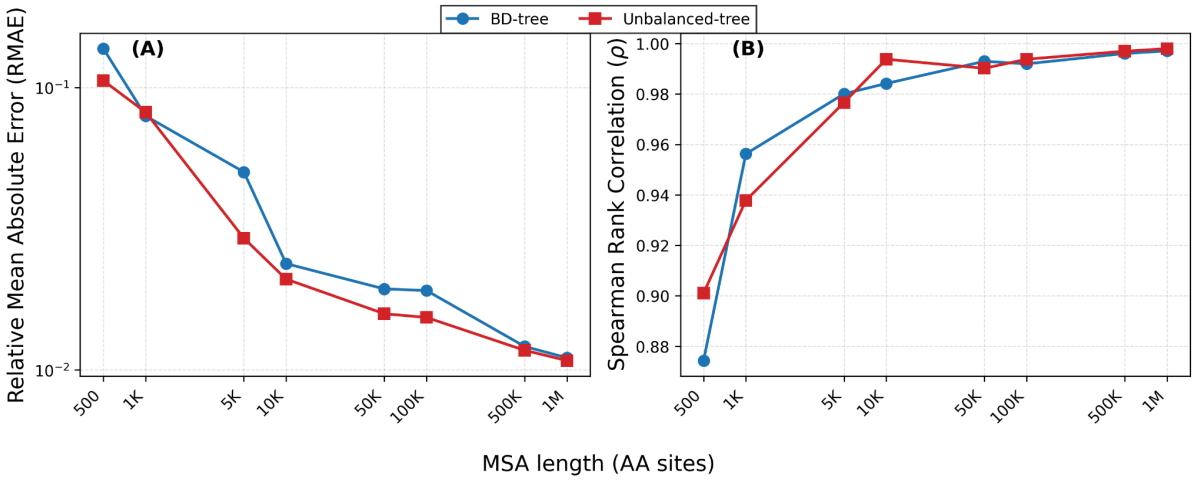

Figure 3. Effect of alignment length on log-likelihood approximation error under the LG+G4+C60 model for phylogenies of many species. (A). Relative mean absolute error (RMAE) (B). Spearman rank correlation between the approximate and exact log-likelihoods is shown as a function of alignment length for the birth–death (BD) (Blue curves) and unbalanced tree (red curves) simulations. Alignments were simulated under the LG+G4+C60 model across lengths ranging from 500 to 1,000,000 amino acid sites. For each alignment length and tree type, approximate log-likelihoods were evaluated for posterior samples and compared with the exact log-likelihoods computed using IQ-TREE 3.

This pattern was also evident from the direct comparison between approximate and exact log-likelihoods. For the 500-site alignments (Figs. 4A, B), the scatter plots showed greater dispersion around the diagonal for both the birth–death tree and the unbalanced tree, indicating reduced agreement between the approximate and exact likelihoods when sequence information was limited. However, for the 1-million-site alignments (Fig. 4C, D), the points closely followed the diagonal in both tree shapes, with very high correlation and R^2 values. This indicates near one-to-one agreement between the approximate likelihoods and the exact likelihoods computed using IQ-TREE 3. These results recapitulate the findings of dos Reis and Yang (2011) for simpler models under relaxed clocks.

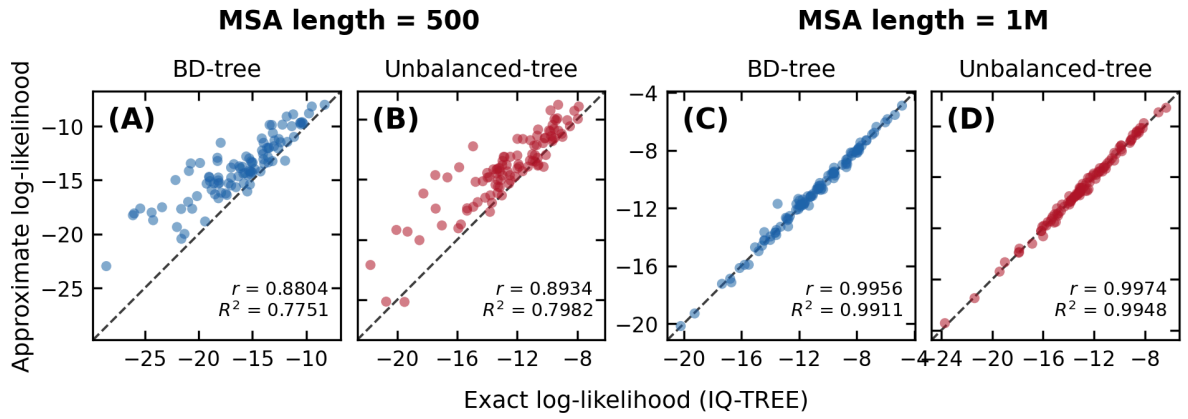

Figure 4. Comparison of approximate versus exact log-likelihoods under the LG+G4+C60 model for MCMC samples obtained from the stationary phase of MCMC chains. Panels A and B show results for 500-site alignments simulated on the birth–death (BD) tree and unbalanced tree, respectively. Panels C and D show the results for 1-million-site alignments simulated on the BD tree and unbalanced tree, respectively. The dashed diagonal line represents perfect agreement between the approximate and exact log-likelihoods.

Likelihood approximation for empirical datasets under different clock models

We further evaluated the accuracy of the approximation in empirical amino acid datasets (metazoan and, eukaryotes and prokaryotes data) analyzed under the LG+G4+C60 profile mixture model. Across both empirical datasets, the autocorrelated-rates clock and the independent-rates clock models, the approximate and exact log-likelihoods showed strong positive agreement with high correlation (Figs. 5A-D). The agreement was particularly clear for the metazoan dataset, whereas the eukaryote dataset showed slightly greater spread, suggesting that approximation accuracy can vary among empirical datasets.

These results recapitulate the trends observed by dos Reis and Yang (2011) in mammal and plant datasets using simpler nucleotide models (Fig. 5 in their paper). Overall, these results indicate that the approximate likelihood provides a reliable representation of the exact LG+G4+C60 likelihood in the posterior region sampled by MCMCtree. This consistency is expected to improve with the availability of larger alignments, as shown in simulations. Also, the consistency of this pattern indicates the approximation is robust to the rate model used, supporting the use of IQ2MC for efficient Bayesian time-tree inference under computationally demanding profile mixture models in empirical phylogenomic analyses. However, we do stress that users should attempt to construct reasonably long alignments to guarantee the approximation is asymptotically appropriate.

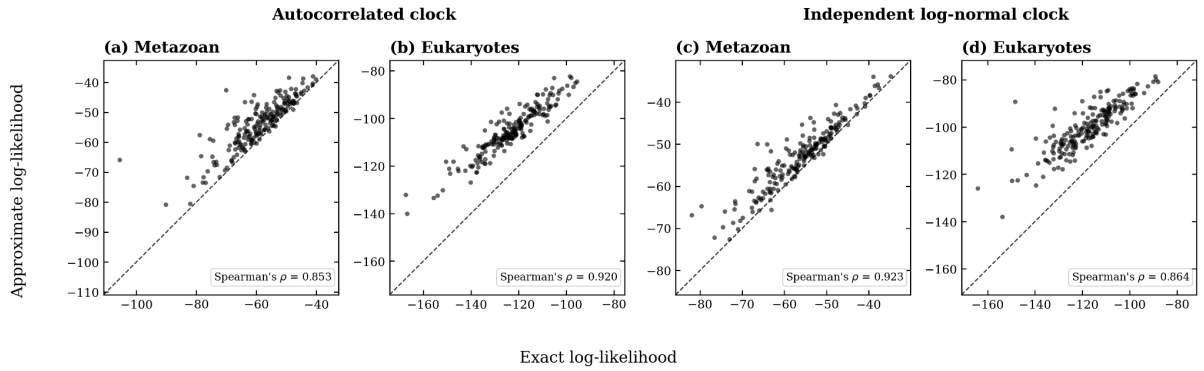

Figure 5. Comparison of approximate and exact LG+G4+C60 log-likelihoods in empirical datasets under autocorrelated and independent log-normal rate models. Panels A and B show the comparison

between approximate and exact log-likelihoods under the autocorrelated-rates for the metazoan and eukaryote–prokaryote datasets, respectively. Panels C and D show the corresponding comparisons under the independent log-normal rates for the metazoan and eukaryote–prokaryote datasets, respectively. The dashed diagonal line is the x=y line.

Inference under wrong substitution models substantially biases divergence time estimates under simulations

Our simulation analyses on both DNA and amino acid data show that model violations substantially affect divergence time estimates, particularly at higher evolutionary rates. Figure 6 compares divergence times from LG+G4 and LG+G4+C60 models to ground truth across two divergence levels (1x and 10x), with time estimates from each model on the y-axis and simulation ground truths on the x-axis.

Under the young calibration setting, the misspecified LG+G4 model leads to a substantial underestimation of divergence times compared to the true LG+G4+C60 model, with discrepancies increasing at higher evolutionary rates (Fig. 6, left column). In the middle calibration setting, placing the calibration on an older node improves agreement between LG+G4 estimates and ground truth at lower rates (Fig. 6, middle column). However, discrepancies remain at higher divergence levels (10x) when using LG+G4 compared to the LG+G4+C60 profile mixture model.

In the root calibration, times estimated under the true and misspecified models are close to the ground truth at lower rates. In other words, a precise calibration on the root pins the inferred timespan of the phylogeny, making the analysis robust to substitution model misspecification. Despite this improvement, estimation bias remains and is stronger under faster evolutionary rates, with the LG+G4 model substantially overestimating divergence times for the middle nodes at higher divergence levels. (see Fig. 6, right column).

Similar patterns were observed in other comparisons. For DNA alignments, time estimates obtained using the misspecified JC69 model tended to be similarly biased, with substantial underestimation of the root age for the young calibration setting, whereas estimates under the true GTR+G4 model were centered on the true values (Supplementary Figs. S10- S13). For amino acid alignments, the time estimates follow the same trend, with the simpler LG+G4 model providing larger estimation errors compared to the LG+G4+C10 to LG+G4+C60 models (Supplementary Figs. S14- S21).

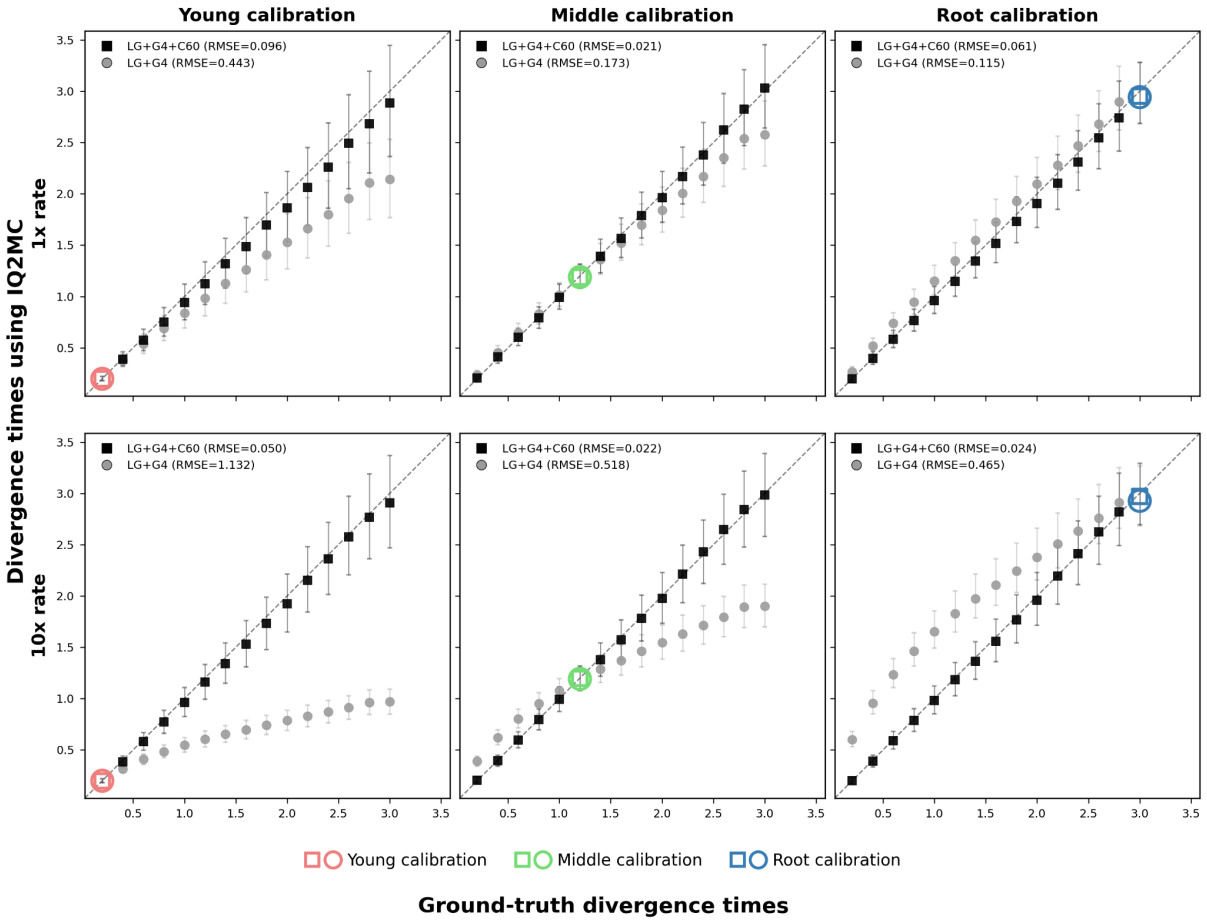

Figure 6: Comparison of estimated divergence times under LG+G4 (misspecified) and LG+G4+C60 (true) models to ground truth times under the IQ2MC workflow. Three different calibration settings are used with varying rates (1x and 10x). The columns represent the young, middle, and root calibration settings. Young calibration setting (young node (red) $t_{15} \sim B(0.18,022)$, root node (blue) t_1

~ G(2,0.66)), middle calibration setting (middle node (green) t_{10} ~ B(1.08,1.32), root node (blue) t_1 ~ G(2,0.66)), and root calibration setting (root node (blue) t_1 ~ B(2.7, 3.3)) while the first and second rows represent the time estimates under the 1x and 10x rates simulations respectively. RMSE: root mean square error.

Mixture models improve accuracy under complex simulation conditions

We assessed whether profile mixture models improve divergence-time estimation under more realistic simulation conditions, where the amino acid alignments are simulated using published site-specific profiles obtained under the posterior mean site frequency (PMSF) method (Giacomelli et al., 2025results). Figure 7 shows RMSE summaries of divergence times estimated using the LG+G4 and LG+G4+C60 models for the simulation under the 64-taxon birth-death tree. Here, the LG+G4+C60 always produced lower RMSE than LG+G4 across all sequence lengths, calibration schemes, and divergence-rate settings. The improvement was especially clear under the 10× rate setting, where the site-homogeneous LG+G4 model showed substantially higher error, while the profile mixture model maintained lower and more stable RMSE values. Under the 1× rate setting, the difference between models was smaller, but LG+G4+C60 still produced lower RMSE, indicating that modeling site-specific amino acid preferences improved divergence-time estimation even when sequence divergence was modest. Similar trends were observed in the remaining PMSF simulation settings (Supplementary Figs. S22- S24).

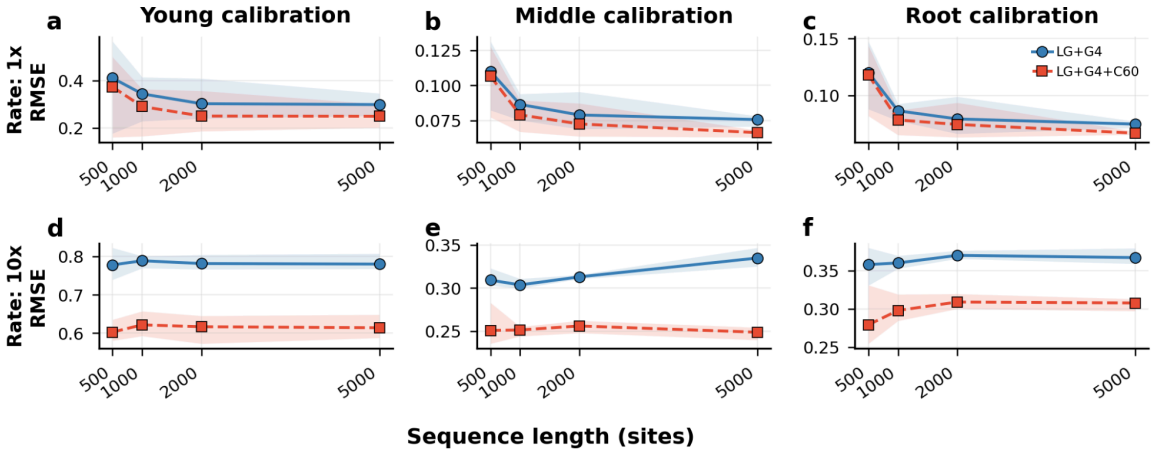

Figure 7: Comparison of root mean square error (RMSE) between ground-truth and estimated divergence times under the LG+G4 and LG+G4+C60 substitution models for realistic amino acid simulations using the PMSF profile approach and the 64-taxon birth-death tree. The x-axis represents the sequence length of the simulated multiple sequence alignments, and the y-axis represents the RMSE of the estimated divergence times. The first and second rows correspond to simulations with the 1× and 10× rates, respectively, while the first, second, and third columns correspond to the young, middle, and root calibration settings, respectively. Blue circles and red squares show the mean RMSE across five replicate analyses under the LG+G4 and LG+G4+C60 models, respectively. The shaded bands show the minimum-to-maximum RMSE range across the five replicate analyses for each sequence length, substitution model, and calibration setting.

Impact of partition and mixture models on time estimates in empirical datasets

DNA mixture models for divergence time estimation in Placental Mammals and Plants

To analyze the effects of complex DNA models on divergence time estimations, we reanalyzed two datasets of placental mammals and plants. The placental mammal dataset consists of 60 taxa with about 250 million years of tree depth. This is a subset of the

Afrotheria clade analyzed by (Álvarez-Carretero et al. 2022). Since we do not know the true divergence times, we compared the divergence times inferred between the simpler model (GTR+G4 on a concatenated alignment) and the complex model (GTR mixture model estimated with MixtureFinder on a concatenated alignment). We computed the gradient and the Hessian under both substitution models and estimated divergence times under four different calibration settings (supplementary Tables S5 and S10). Under the Bayesian Information Criterion (BIC) score, the GTR mixture model indicated a much better fit to the data over the GTR+G4 model. The linear regression slopes for comparison of mean divergence times ranged from 1.05 to 1.13 with low dispersion ($R^2 = 0.99$) (Figs. 8a-8d). The highest linear regression slope (1.13) was obtained when the middle calibration setting was utilized. Under all analyses, the root age was younger under the GTR+G4 model compared with the GTR mixture model but fell within the 95% credible interval (CIs) of root age estimated by the mixture model (Supplementary Fig. S25). The divergence time estimates for the outgroup showed substantial discrepancies when young, middle, and root calibration settings were used, specifically the GTR+G4 model producing consistently younger divergence times (Figs. 8a–8c). However, these discrepancies disappeared when all calibrations were applied (Fig. 8d). This suggests that DNA mixture models can have a substantial impact on divergence time estimation, particularly when few calibrations are present in the phylogeny. We recommend users study the impact of substitution models on the robustness of time tree inference in their own datasets.

We also estimated the divergence times with EUL and EL partition models, where the best-fitted substitution model for each partition was obtained with ModelFinder. We then compared the times obtained with the GTR+G4 model (on a concatenated alignment) with EUL and EL partition models. The analyses reveal the significant impact of the substitution

model and the partition scheme on divergence time estimation (Supplementary Fig. S26) when using partition models instead of a concatenated alignment.

We further investigated the effects of DNA mixture models on divergence time estimation using a plant dataset. The dataset originally consisted of 103 taxa, 856,439 sites, and 798 million years of tree depth. Following Tao et al. (2020), we used the 2,217-site subsampled alignment, shown to yield comparable time estimates to the full alignment (Morris et al. 2018) and also suitable for mixture model estimation under MixtureFinder (Ren et al. 2024). We selected the GTR+G4 model on a concatenated alignment as the simple model and a 4-component DNA mixture model estimated by MixtureFinder as the complex model (Supplementary Table S11) for the time estimates comparison. The DNA mixture model produced a lower BIC score than the GTR+G4 model, demonstrating the better fit of the DNA mixture model to the data. However, the time estimates obtained by both simpler and complex substitution models were very similar, as observed by Tao et al. (2020) previously for the GTR+G4 and JC69 model comparison. The maximum linear regression slope was 1.04 with low dispersion ($R^2 = 1$, Figs. 8e-8h).

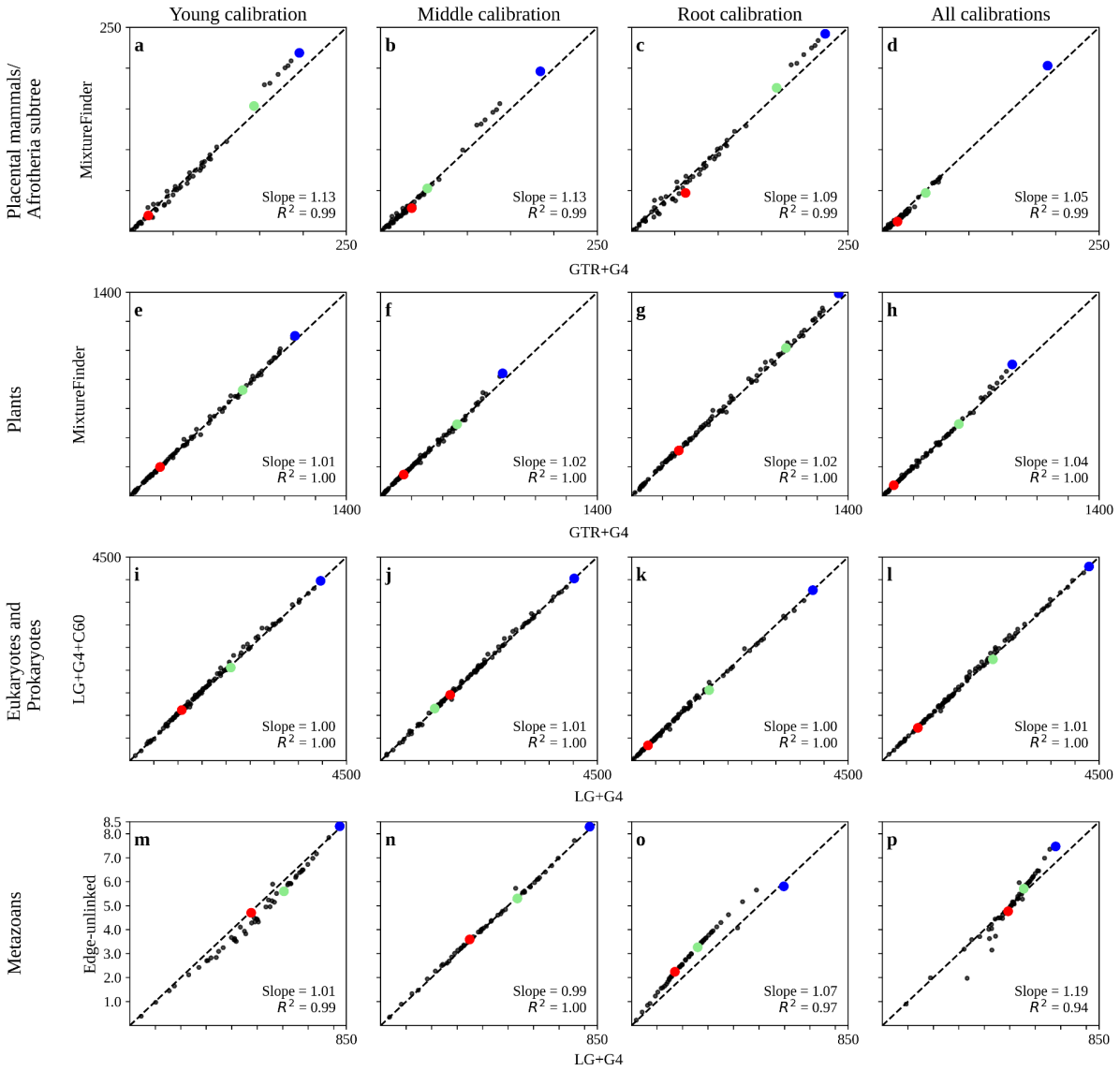

Figure 8: Comparison of estimated divergence times under complex and simple substitution models for empirical datasets. (a-d) Divergence time comparison for the placental mammal/afrotheria subtree dataset under GTR+G4 and MixtureFinder with concatenated alignments. (e-h) divergence times comparison for the plant dataset under GTR+G4 and MixtureFinder with concatenated alignments. (i-l) Divergence times comparison for the eukaryotes and prokaryotes dataset under LG+G4 and LG+G4+C60 models with concatenated alignments. (m-p) Divergence times comparison for the metazoan dataset under LG+G4 with concatenated alignment and ModelFinder with edge-unlinked partition model. (a, e, i, m) Only young and root calibrations are used. (b, f, j, n) Only the middle and root calibrations are used. (c, g, k, o) Only the root calibration is used. (d, h, l, p) All calibrations in

the source publication are used. Each axis represents divergence times in million years. In each plot, the red dot represents the young calibration, the green dot represents the middle calibration, and the blue dot represents the root calibration (Supplementary Tables S5-S8).

For all the calibration settings, the GTR+G4 model provided younger estimates of the root age compared to the DNA mixture model, but fell within 95% CIs of the root age estimated by the DNA mixture model (Supplementary Fig. S25). Therefore, the impact on the DNA mixture model for the time tree is negligible for the plant data compared to the placental mammal data. This contrasts with the placental mammal and other deep phylogenomic datasets, suggesting that the effect of substitution-model complexity on divergence-time estimates depends on the dataset, node depth of the phylogeny, and the calibrations used. The similar rate estimates obtained under GTR+G4 and the DNA mixture model (Supplementary Figs. S27 and S28) under both independent log-normal and autocorrelated clock models resulted in correspondingly similar node ages, whereas mixture and partition models can have much larger effects when they substantially alter branch-length or rate estimates, as also observed for Rickettsiales in recent molecular-dating analyses (Wang and Meade 2026).

Protein Mixture Models for divergence time estimation in Eukaryotes and Prokaryotes

We next reanalyzed the dataset of Eukaryotes and Prokaryotes (Betts et al. 2018; Tao et al. 2020) with 102 species and 4.5 billion years of tree depth to demonstrate the usage of complex amino acid substitution models available in IQ2MC for phylogenetic dating. We compare the divergence time estimated using the LG+G4 model and the LG+G4+C60 profile mixture model on a concatenated alignment. We calculated the gradient and Hessian matrix for each substitution model (Supplementary Table S12) and estimated divergence times under four different calibration settings (Supplementary Table S7). The BIC score for the

LG+G4+C60 model was lower than for the LG+G4 model, suggesting a better fit of the LG+G4+C60 profile mixture model to the data.

Divergence time estimates obtained using the LG+G4 model were highly consistent with those from the profile mixture model (LG+G4+C60) across all four calibration settings, yielding a maximum slope of 1.01 and an R^2 of 1 (Figs. 8i-8l). Under the young, middle, and all-calibration settings, the LG+G4 model exhibited a slight tendency to produce younger divergence times compared to the profile mixture model, though the estimates remained within 95% CIs of the profile mixture model (Supplementary Fig. S25). These findings suggest that the profile mixture model may exert a minor influence on divergence time estimation for deep phylogenies, which requires further investigation.

We further analyzed the divergence times estimated for Eukaryotes and Prokaryotes data with EUL and EL partition models. ModelFinder was used to estimate the best-fit substitution model for each partition (Supplementary Table S12). The source publication of the dataset contained 29 partitions, and the time estimates in the original research showed a high influence on partition schemes (Betts et al. 2018; Tao et al. 2020). Moreover, we observe the same influence on time estimates when the divergence times are estimated under the EUL and EL partition models with ModelFinder (Supplementary Figs. S29 and S30), where EUL and EL models produced older time estimates for part of the nodes while producing younger estimates for the rest compared to the LG+G4 model on a concatenated alignment.

Partition models with ModelFinder for Metazoan Data

We further investigated the effects of partition models for divergence time estimation using the Metazoan data (dos Reis et al. 2015), which includes 54 species and 757 million years of tree depth. Here we compare the divergence time estimated using the LG+G4 model and the best-fitted substitution models estimated with ModelFinder for the EUL partition

model. We calculated the gradients and Hessian matrix for the LG+G4 model with a concatenated alignment. For the EUL model with ModelFinder, gradient vectors and Hessian matrices were estimated for each partition separately (Supplementary Table S13). The EUL model had a lower BIC score compared to the LG+G4 model, indicating that the EUL model provides a better fit for Metazoan data. We then used MCMCtree for divergence time estimation with four calibration settings (Supplementary Table S8).

The divergence times estimated with the young, root, and all calibration settings produced highly variable time estimates when the EUL partition model and rather simple LG+G4 model were compared (Figs. 8m, 8o, and 8p). Especially, under the young calibration setting, the LG+G4 model produced older time estimates compared to the EUL partition model (Fig. 8m) with a slope of 1.01 and R^2 of 0.99. Furthermore, under the root-only calibration setting (Fig. 8o), the LG+G4 model produced younger divergence time estimates (slope of 1.07 and R^2 of 0.97, Fig. 8o). However, both models produced similar root age estimates across the two calibration settings. The highest discrepancies in time estimates between the two models occurred under the all-calibration setting (slope of 1.19 and R^2 0.94, Fig. 8p). The LG+G4 model produced younger time estimates for the root age and the divergence times for some nodes while producing older estimates for the rest compared to the EUL model.

In addition to the comparison of the LG+G4 model on concatenated alignment to the EUL partition model, we compared the time estimates from the LG+G4 model on concatenated alignment against the LG+G4+C60 on a concatenated alignment and the EL partition model. The time estimates from the LG+G4 model showed dissimilarities to the estimates of the LG+G4+C60 profile mixture model (slope of 1.03 and R^2 of 0.99, Fig. S31), even when all calibrations were used. The time estimates of the LG+G4 model were consistent with the estimates obtained from the EL model. However, under the root-only

calibration setting, LG+G4 estimated younger divergence times for the internal nodes (slope of 1.05 and R^2 of 1, Supplementary Fig. S31).

Wall-clock time and RAM usage

We evaluated the computational efficiency of IQ-TREE 3 for gradient and Hessian calculations by benchmarking IQ-TREE's runtime and memory usage against PAML (BaseML and CodeML for DNA and amino acid data, respectively). We simulated two large DNA and protein alignments using Alisim-HPC (Ly-Trong et al. 2023) with 300 taxa and 500,000 sites under GTR+G4 and LG+G4 substitution models. We evaluated two threading configurations for IQ-TREE: a single-threaded mode and a 10-threaded mode. For PAML, we used only a single thread, as it does not support multithreading. In the IQ2MC implementation of gradient and Hessian computation, we leveraged vector operations and OpenMP (Open Multi-Processing) within IQ-TREE 3 to optimize and parallelize computations across multiple CPU cores. All analyses were conducted on a server with an AMD EPYC 7551 32-core Processor and 500 GB of RAM. The wall-clock runtimes and memory usage reported in Table 2 represent the averages from three independent runs.

Across all analyses, IQ-TREE significantly outperformed PAML in both single-threaded and multi-threaded modes (Table 2). For DNA data, IQ-TREE was able to achieve a 9.45x speedup under a single-threaded mode, while with 10 threads, a 92.5x speedup was achieved. The same observation was visible for amino acid data, where IQ-TREE was 14.71x times faster under a single-threaded mode than PAML, while a 148.82x speed-up is achieved with 10 threads, making large-scale Hessian computations for phylogenomic data feasible within practical timeframes. The peak memory usage of IQ-TREE remains comparable to PAML in the case of DNA data (18.3 GB vs. 14 GB) and

amino acid data (68 GB vs. 64 GB). However, considering the substantial reductions in runtime, this increase is a reasonable trade-off for large-scale phylogenomic analyses.

Table 2: Runtime comparison between PAML (CodeML and BaseML as existing tools) and IQ-TREE 3 (new implementation) for computing the gradient and Hessian matrix at the maximum likelihood estimates of the branch lengths.

Method	Simulated DNA alignment (300 taxa × 500K sites)			Simulated amino acid alignment (300 taxa × 500K sites)		
	Runtime* (hh:mm:ss)	Memory	Speedup	Runtime (hh:mm:ss)	Memory	Speedup
PAML (1 thread)	13:03:23	14 GB	-	247:32:02	64 GB	-
IQ-TREE (1 thread)	1:22:53	18.3 GB	9.45	16:49:59	68 GB	14.71
IQ-TREE (10 threads)	0:08:28	18.3 GB	92.5	1:39:48	68 GB	148.82

***** Analyses are conducted on an AMD EPYC 7551 32-core server with 500 GB RAM.

We further benchmarked IQ-TREE 3 and PAML for gradient and Hessian estimation using simulated alignments with varying numbers of sites and taxa (Supplementary Figs. S32 and S33). Our results show that, on average, IQ-TREE achieves a 100x speed-up compared to PAML. We also measured the runtime and peak memory consumption for gradient and Hessian calculations across all empirical datasets analyzed (Supplementary Table S9). The runtime ranged from a few seconds to several hours. Notably, when using ModelFinder or MixtureFinder for model selection, the runtime can increase to several hours, depending on the dataset size and number of substitution models evaluated.

For Bayesian dating with the approximate likelihood method, MCMCtree typically requires several hours to multiple days to achieve MCMC convergence, depending on the size of the dataset. For instance, the Eukaryotes and Prokaryotes dataset analysis using a concatenated alignment took approximately 60 hours (Supplementary Table S15). Under the EUL partition model with 29 partitions, MCMC convergence required 183 hours. These results indicate that the primary computational bottleneck now shifts to the MCMCtree algorithm for divergence time estimation. This underscores the need for faster MCMC algorithms to keep pace with growing phylogenomic datasets.

DISCUSSION

We present a novel framework, IQ2MC, for time tree inference that seamlessly integrates IQ-TREE 3 and MCMCtree. IQ2MC enables significant speed-ups (up to 2 orders of magnitude) in the calculation of the gradient and the Hessian of the log-likelihood function evaluated at the MLEs of branch lengths, a key step for fast MCMC sampling of divergence times. This approach takes full advantage of the wide range of advanced substitution models and model selection tools available in IQ-TREE 3, making molecular dating faster, more accurate, and user-friendly. By accommodating more sophisticated models of sequence evolution and by testing them on simulated and empirical data, we show IQ2MC can facilitate more reliable time tree inference.

Importantly, IQ2MC introduces the usage of DNA mixture models and amino acid profile mixture models, which were previously unavailable for dating with MCMCtree. These models, designed to account for site-specific variation in substitution patterns, significantly improve the realism of evolutionary models and help to identify potential biased date estimates that can arise when simpler models are used. Through comprehensive analyses of

simulated and empirical datasets, we reveal the impact of adopting mixture models compared to simpler substitution models, highlighting their significance in time tree estimation.

Our site-heterogeneous simulations show that when calibrations are sparse, using simple models leads to biased time estimates (Fig. 6), whereas those under mixture models are unbiased. Further analysis on real data also revealed discrepancies between the dates estimated under these two models. A recent study on deep divergence of Fungi and Bacteria (Wang and Meade 2026) also corroborated this result. Therefore, when there is scant calibration information (such as in large areas of the early Tree of Life), we recommend users carry out molecular dating using mixture models to increase the robustness of the analysis.

Furthermore, the analysis of four empirical datasets suggests that one should carefully select the substitution models for divergence time estimation. As we show here, when phylogenies are densely calibrated, estimated divergence times between simple and complex substitution models are not markedly different (as noticed by Tao et al. 2020), arguably because the dense calibrations and the relaxed clocks interact to place node ages in their right geological context. On the other hand, when calibrations across the phylogeny are sparse, differences in time estimates between simple and complex substitution models can be very dramatic, as shown here for our simulation analyses and in the Metazoan dataset. Thus, correct model selection appears critical for robust divergence time estimation in deep phylogenies (see also Wang and Luo, 2025).

Wang and Meade (2026) provided important evidence in this context for the ancient symbiont groups Microsporidia and Rickettsiales. They showed that profile mixture models can produce higher evolutionary-rate estimates and, for Rickettsiales, markedly older divergence-time estimates under some combinations of rate models and calibrations. Therefore, similar time estimates across the presented empirical datasets should not be interpreted as evidence that mixture models have little influence on molecular dating more

generally. Their effect is expected to depend on the distribution and informativeness of calibrations, the magnitude of model misspecification, branch lengths, and the flexibility of the rate model.

Finally, the impact of advanced substitution models on divergence times with different rate models should be further investigated, especially in the context of mixture models, which leads to a potential future research direction. Because rate models can lead to substantially different time estimates. Bayesian model selection approaches using approximate likelihood have been recommended to select the best rate model in empirical datasets (Panchaksaram et al. 2025). These rate model selection approaches can now be used with complex substitution models thanks to IQ2MC.

An equally important question is whether the approximation used by MCMCtree remains reliable for mixture models. Our analysis confirmed that it is the case, except when branches are too long and alignments relatively short. We also showed that branch length transformations under mixture models work well and improve the approximation accuracy, including the default ARCSIN transformation in MCMCtree.

We note the approximated method works on a fixed tree topology. In some cases, substantial uncertainties in topological relationships between clades may exist. For those cases, we recommend users carry out several MCMC runs on different topologies and compare the results (e.g., see dos Reis et al. (2015), Alvarez-Carretero et al.(2022)). Users can, for example, combine MCMCtree analyses with phylogenetic bootstrapping methods, such as the standard bootstrap (Felsenstein 1985) and the ultrafast bootstrap (Hoang et al. 2018). Timetrees with different topologies can then be combined according to their posterior probabilities. The exact details on the best approaches to achieve this under the approximate method should be the subject of future studies.

Other Bayesian methods such as BEAST (Bouckaert et al. 2019) and PhyloBayes (Lartillot et al. 2009) have the advantage of being able to sample divergence times and tree topologies simultaneously, thus allowing the user to automatically assess topological uncertainty on the time estimates. However, these methods are currently too expensive to run on the large datasets analyzed here. For example, we ran BEAST 2.7.8 with the BEAGLE library on the Metazoan dataset with the standard LG+G4 model. Using 32 threads, BEAST achieved 1.66 ESS/hour on the tree height posterior, whereas on the same machine, MCMCtree achieved 1745 ESS/hour on the same posterior, using a single thread. That means MCMCtree is 1051x times faster than BEAST on wall time, and 33,639x times faster on CPU time. We note the current development version of MCMCtree is multi-threaded, and thus the speed advantages of the approximation will be much more marked.

Furthermore, BEAST currently does not provide mixture models. For deep time divergence inference, we recommend users use PhyloBayes with the CAT mixture models (Lartillot and Philippe 2004) if they need to assess the effect of tree topology and if the alignments are relatively short, as PhyloBayes calculates the likelihood exactly using data augmentation. Furthermore, this is especially advisable when there is evidence of simple model violations (Szánthó et al. 2023), which may impact tree topologies and in turn bias divergence times. Our suggested approach of using mixture models and potentially combining with the bootstrap procedure above may solve the concern of fixed topology, while still being able to deliver the result within reasonable time.

The combination of increased computational efficiency and support for a large array of substitution models in IQ-TREE 3 for the approximate likelihood technique represents a substantial step forward in large-scale phylogenomic dating. IQ-TREE highly optimizes the necessary precalculation of the derivatives, and the computational bottleneck now shifts to MCMCtree. Therefore, we will investigate efficient MCMC algorithms, including parallel

MCMC techniques such as Metropolis-coupled MCMC, in the future to improve the divergence time estimation.

SUPPLEMENTARY MATERIAL

Supplementary data is available at: `https:doi.org/10.5281/zenodo.22239741`

ACKNOWLEDGMENTS

We thank Thomas Wong and Robert Lanfear for their useful comments.

AVAILABILITY

IQ2MC is available under the IQ-TREE release version **3.0.1**, freely available at https://iqtree.github.io/, and the comprehensive user manual is available at https://iqtree.github.io/doc/Dating. An updated version of PAML that supports IQ2MC workflow is available at: https://github.com/iqtree/paml.

FUNDING

This work was partly supported by the Australian Research Council Training Centre for Accelerated Future Crops Development (IC210100047 to P.D.); Biotechnology and Biological Sciences Research Council (UK) grants (BB/T01282X/1 and BB/Y003624/1 to M.d.R.); Chan-Zuckerberg Initiative grants for essential open-source software for science (EOSS4-0000000312 to B.Q.M.). Computational resources were provided by the Australian Government through the National Computational Infrastructure under the ANU Merit Allocation Scheme and the Center for Integrative Bioinformatics Vienna (CIBIV) cluster by Arndt von Haeseler.

REFERENCES

Álvarez-Carretero, Sandra, Asif U. Tamuri, Matteo Battini, et al. 2022. "A Species-Level Timeline of Mammal Evolution Integrating Phylogenomic Data." *Nature* 602 (7896): 263–267.

Battistuzzi, Fabia U., Paul Billing-Ross, Aditya Paliwal, and Sudhir Kumar. 2011. "Fast and Slow Implementations of Relaxed-Clock Methods Show Similar Patterns of Accuracy in Estimating Divergence Times." *Molecular Biology and Evolution* 28 (9): 2439–2442.

Betts, Holly C., Mark N. Puttick, James W. Clark, Tom A. Williams, Philip C. J. Donoghue, and Davide Pisani. 2018. "Integrated Genomic and Fossil Evidence Illuminates Life's Early Evolution and Eukaryote Origin." *Nature Ecology & Evolution* 2 (10): 1556–1562.

Bouckaert, Remco, Timothy G. Vaughan, Joëlle Barido-Sottani, et al. 2019. "BEAST 2.5: An Advanced Software Platform for Bayesian Evolutionary Analysis." *PLoS Computational Biology* 15 (4): e1006650.

Chernomor, Olga, Arndt von Haeseler, and Bui Quang Minh. 2016. "Terrace Aware Data Structure for Phylogenomic Inference from Supermatrices." *Systematic Biology* 65 (6): 997–1008.

Davín, Adrián A., Ben J. Woodcroft, Rochelle M. Soo, et al. 2025. "A Geological Timescale for Bacterial Evolution and Oxygen Adaptation." *Science (New York, N.Y.)* 388 (6742): eadp1853.

Drummond, Alexei J., and Remco R. Bouckaert. 2015. *Bayesian Evolutionary Analysis with BEAST*. Cambridge University Press.

Drummond, Alexei J., and Andrew Rambaut. 2007. "BEAST: Bayesian Evolutionary Analysis by Sampling Trees." *BMC Evolutionary Biology* 7 (1): 214.

Felsenstein, Joseph. 1985. "Confidence Limits on Phylogenies: An Approach Using the Bootstrap." *Evolution; International Journal of Organic Evolution* 39 (4): 783–791.

Giacomelli, Mattia, Matteo Vecchi, Roberto Guidetti, et al. 2025. "CAT-Posterior Mean Site Frequencies Improves Phylogenetic Modeling under Maximum Likelihood and Resolves Tardigrada as the Sister of Arthropoda plus Onychophora." *Genome Biology and Evolution* 17 (1): evae273.

Guindon, Stéphane. 2010. "Bayesian Estimation of Divergence Times from Large Sequence Alignments." *Molecular Biology and Evolution* 27 (8): 1768–1781.

Hoang, Diep Thi, Olga Chernomor, Arndt von Haeseler, Bui Quang Minh, and Le Sy Vinh. 2018. "UFBoot2: Improving the Ultrafast Bootstrap Approximation." *Molecular Biology and Evolution* 35 (2): 518–522.

Jukes, Thomas H., and Charles R. Cantor. 1969. "Evolution of Protein Molecules." In *Mammalian Protein Metabolism*. Elsevier.

Kainer, D., and R. Lanfear. 2015. "The Effects of Partitioning on Phylogenetic Inference."

Molecular Biology and Evolution 32 (6): 1611–1627.

Kalyaanamoorthy, S., B. Minh, Thomas K. F. Wong, A. von Haeseler, and L. Jermiin. 2017. "ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates." *Nature Methods* 14 (6): 587–589.

Lanfear, Robert, Brett Calcott, David Kainer, Christoph Mayer, and Alexandros Stamatakis. 2014. "Selecting Optimal Partitioning Schemes for Phylogenomic Datasets." *BMC Evolutionary Biology* 14 (April): 82.

Lartillot, Nicolas, Henner Brinkmann, and Hervé Philippe. 2007. "Suppression of Long-Branch Attraction Artefacts in the Animal Phylogeny Using a Site-Heterogeneous Model." *BMC Evolutionary Biology* 7 Suppl 1 (February): S4.

Lartillot, Nicolas, and Hervé Philippe. 2004. "A Bayesian Mixture Model for across-Site Heterogeneities in the Amino-Acid Replacement Process." *Molecular Biology and Evolution* 21 (6): 1095–1109.

Lartillot, N., T. Lepage, and S. Blanquart. 2009. "PhyloBayes 3: A Bayesian Software Package for Phylogenetic Reconstruction and Molecular Dating." *Bioinformatics* 25 (17): 2286–2288.

Le, Si Quang, Cuong Cao Dang, and Olivier Gascuel. 2012. "Modeling Protein Evolution with Several Amino Acid Replacement Matrices Depending on Site Rates." *Molecular Biology and Evolution* 29 (10): 2921–2936.

Le, Si Quang, and Olivier Gascuel. 2008. "An Improved General Amino Acid Replacement Matrix." *Molecular Biology and Evolution* 25 (7): 1307–1320.

Le, Si Quang, Nicolas Lartillot, and Olivier Gascuel. 2008. "Phylogenetic Mixture Models for Proteins." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 363 (1512): 3965–3976.

Ly-Trong, Nhan, Giuseppe M. J. Barca, and Bui Quang Minh. 2023. "AliSim-HPC: Parallel Sequence Simulator for Phylogenetics." *Bioinformatics* 39 (9): btad540.

Marin, Julie, Fabia U. Battistuzzi, Anais C. Brown, and S. Blair Hedges. 2017. "The Timetree of Prokaryotes: New Insights into Their Evolution and Speciation." *Molecular Biology and Evolution* 34 (2): 437–446.

Morris, Jennifer L., Mark N. Puttick, James W. Clark, et al. 2018. "The Timescale of Early Land Plant Evolution." *Proceedings of the National Academy of Sciences of the United States of America* 115 (10): E2274–E2283.

Panchaksaram, Muthukumaran, Lucas Freitas, and Mario Dos Reis. 2025. "Bayesian Selection of Relaxed-Clock Models: Distinguishing between Independent and Autocorrelated Rates." *Systematic Biology* 74 (2): 323–334.

Quang, Le Si, Olivier Gascuel, and Nicolas Lartillot. 2008. "Empirical Profile Mixture Models for Phylogenetic Reconstruction." *Bioinformatics (Oxford, England)* 24 (20): 2317–2323.

Rannala, B., and Z. Yang. 1996. “Probability Distribution of Molecular Evolutionary Trees: A New Method of Phylogenetic Inference.” *Journal of Molecular Evolution* 43 (3): 304–311.

Rannala, B., and Ziheng Yang. 2007. “Inferring Speciation Times under an Episodic Molecular Clock.” *Systematic Biology* 56 (3): 453–466.

Reis, Mario dos, Yuttapong Thawornwattana, Konstantinos Angelis, Maximilian J. Telford, Philip C. J. Donoghue, and Ziheng Yang. 2015. “Uncertainty in the Timing of Origin of Animals and the Limits of Precision in Molecular Timescales.” *Current Biology: CB* 25 (22): 2939–2950.

Reis, Mario dos, and Ziheng Yang. 2011. “Approximate Likelihood Calculation on a Phylogeny for Bayesian Estimation of Divergence Times.” *Molecular Biology and Evolution* 28 (7): 2161–2172.

Reis, Mario dos, Tianqi Zhu, and Ziheng Yang. 2014. “The Impact of the Rate Prior on Bayesian Estimation of Divergence Times with Multiple Loci.” *Systematic Biology* 63 (March): 555–565.

Ren, Huaiyan, Thomas K. F. Wong, Bui Quang Minh, and Robert Lanfear. 2024. “MixtureFinder: Estimating DNA Mixture Models for Phylogenetic Analyses.” In *bioRxiv*. March 21. https://doi.org/10.1101/2024.03.20.586035.

Romiguier, Jonathan, Marek L. Borowiec, Arthur Weyna, et al. 2022. “Ant Phylogenomics Reveals a Natural Selection Hotspot Preceding the Origin of Complex Eusociality.” *Current Biology: CB* 32 (13): 2942–2947.e4.

Ronquist, Fredrik, Maxim Teslenko, Paul van der Mark, et al. 2012. “MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice across a Large Model Space.” *Systematic Biology* 61 (3): 539–542.

Seo, Tae-Kun, Hirohisa Kishino, and Jeffrey L. Thorne. 2004. “Estimating Absolute Rates of Synonymous and Nonsynonymous Nucleotide Substitution in Order to Characterize Natural Selection and Date Species Divergences.” *Molecular Biology and Evolution* 21 (7): 1201–1213.

Soubrier, Julien, Mike Steel, Michael S. Y. Lee, et al. 2012. “The Influence of Rate Heterogeneity among Sites on the Time Dependence of Molecular Rates.” *Molecular Biology and Evolution* 29 (11): 3345–3358.

Stadler, Tanja. 2011. “Simulating Trees with a Fixed Number of Extant Species.” *Systematic Biology* 60 (5): 676–684.

Stiller, Josefin, Shaohong Feng, Al-Aabid Chowdhury, et al. 2024. “Complexity of Avian Evolution Revealed by Family-Level Genomes.” *Nature* 629 (8013): 851–860.

Szánthó, Lénárd L., Nicolas Lartillot, Gergely J. Szöllősi, and Dominik Schrempf. 2023. “Compositionally Constrained Sites Drive Long-Branch Attraction.” *Systematic Biology* 72 (4): 767–780.

Tao, Qiqing, Jose Barba-Montoya, Louise A. Huuki, Mary Kathleen Durnan, and Sudhir

Kumar. 2020. “Relative Efficiencies of Simple and Complex Substitution Models in Estimating Divergence Times in Phylogenomics.” *Molecular Biology and Evolution* 37 (6): 1819–1831.

Tavaré, S. 1986. “Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences.” *Lecture of Mathematics for Life Science*. https://cir.nii.ac.jp/crid/1370846644340782223.

Thorne, J. L., H. Kishino, and I. S. Painter. 1998. “Estimating the Rate of Evolution of the Rate of Molecular Evolution.” *Molecular Biology and Evolution* 15 (12): 1647–1657.

Wang, Huai-Chun, Bui Quang Minh, Edward Susko, and Andrew J. Roger. 2018. “Modeling Site Heterogeneity with Posterior Mean Site Frequency Profiles Accelerates Accurate Phylogenomic Estimation.” *Systematic Biology* 67 (2): 216–235.

Wang, Sishuo, and Haiwei Luo. 2025. “Dating the Bacterial Tree of Life Based on Ancient Symbiosis.” *Systematic Biology*, ahead of print, January 23. https://doi.org/10.1093/sysbio/syae071.

Wang, Sishuo, and Andrew Meade. 2026. “Molecular Clock Dating Using Complex Mixture Models: Applied to Ancient Symbionts.” *Molecular Biology and Evolution* 43 (3): msag039.

Wong, Thomas K. F., Nhan Ly-Trong, Huaiyan Ren, et al. 2025. *IQ-TREE 3: Phylogenomic Inference Software Using Complex Evolutionary Models*. https://ecoevorxiv.org/repository/view/8916/.

Yang, Z. 1995. “A Space-Time Process Model for the Evolution of DNA Sequences.” *Genetics* 139 (2): 993–1005.

Yang, Z. 2000. “Maximum Likelihood Estimation on Large Phylogenies and Analysis of Adaptive Evolution in Human Influenza Virus A.” *Journal of Molecular Evolution* 51 (5): 423–432.

Yang, Ziheng. 1994. “Maximum Likelihood Phylogenetic Estimation from DNA Sequences with Variable Rates over Sites: Approximate Methods.” *Journal of Molecular Evolution* 39 (3): 306–314.

Yang, Ziheng. 2007. “PAML 4: Phylogenetic Analysis by Maximum Likelihood.” *Molecular Biology and Evolution* 24 (8): 1586–1591.

Yang, Ziheng, and B. Rannala. 2006. “Bayesian Estimation of Species Divergence Times under a Molecular Clock Using Multiple Fossil Calibrations with Soft Bounds.” *Molecular Biology and Evolution* 23 (1): 212–226.

Yang, Ziheng, and Carlos E. Rodríguez. 2013. “Searching for Efficient Markov Chain Monte Carlo Proposal Kernels.” *Proceedings of the National Academy of Sciences of the United States of America* 110 (48): 19307–19312.