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] *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 phylogenetic tools to infer phylogenetic trees and estimate 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, it is currently not possible to integrate these tools seamlessly, as MCMCTree requires pre-calculation of the gradients and Hessian for fast approximate calculation of the likelihood, which is unavailable in IQ-TREE. Furthermore, MCMCTree only implements a small subset of substitution models; and complex models such as the 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 gradients 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 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 time scale, 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; Drummond and Rambaut 2007; Rannala and Yang 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, full MCMC sampling is computationally intensive for large phylogenies. Therefore, approximation methods have been developed. Thorne et al. (1998) proposed to first infer a tree topology using a tree reconstruction method and then sample other parameters on this fixed tree using approximate likelihoods based on a normal approximation of the likelihood function. 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 large phylogenies with hundreds of species such as plants, mammals, insects, and birds (Morris et al. 2018; Álvarez-Carretero et al. 2022; Romiguier et al. 2022; Stiller et al. 2024). However, the major limitation of MCMCtree is that it lacks advanced models of sequence evolution, such as

mixture models frequently used to infer deep divergences (Lartillot and Philippe 2004; Quang et al. 2008).

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). In particular, IQ2MC improves the efficiency of gradients and Hessian calculation (required in the likelihood approximation) compared to the workflow using PAML. An added benefit of IQ2MC is that we can now use all 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 (Le et al. 2008; Quang et al. 2008; Ren et al. 2024). Such models are important for more accurate dating analysis, as model violation might bias time tree inference (Barba-Montoya et al. 2020; Tao et al. 2020), as we will also show in this paper. 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 as follows (Fig. 1). Given an input multiple sequence alignment (MSA), IQ2MC first finds a 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). Users then need to annotate the maximum likelihood tree with fossil/tip date calibrations (which can be done with commonly

used tree editing tools such as FigTree, (Rambaut 2009) or iTOL, (Letunic and Bork 2021)). 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 contains the fossil calibrations), and the gradient and Hessian (stored in the in.BV file) to obtain the MCMC samples of divergence times and evolutionary rates. Currently, step 2 is done using the BaseML and CodeML programs of the PAML package, and thus only those substitution models available in PAML can be used (but see for 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 within a few command lines.

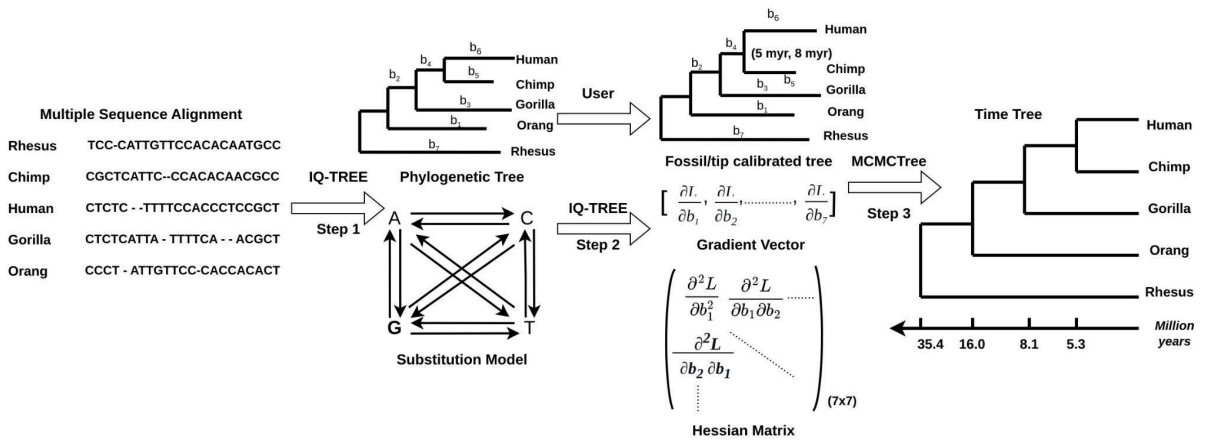

Figure 1. General workflow of IQ2MC for inferring time trees with IQ-TREE 3 and MCMCTree. Step 1 infers the best-fit substitution model and a maximum likelihood phylogenetic tree from a given

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 with single substitution models

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 a 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:

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

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 $H(D, \widehat{B})$. This 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 partition models (-p option in IQ-TREE) 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. Low r_i means that the partition i is slowly evolving, whereas high r_i means that the partition i is fast evolving. For p partitions, $D_1, D_2, \ldots, D_p$, we compute 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-unliked partition models

Edge-unlinked partition models (-Q option in IQ-TREE) 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 $D_1, D_2, \ldots, D_p$, 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:

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

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 (Le et al. 2008, 2012; Quang et al. 2008; 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 includes 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:

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

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:

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

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

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

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

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

We use Eq. (7) to obtain the diagonal elements of the Hessian matrix only, as the off-diagonal elements estimated this way are not stable. For the off-diagonal elements, we used the outer-product-of-scores (OPS) estimator as it is more stable (Porter 2002; Seo et al. 2004). However, the OPS estimator assumes all parameter estimates are within parameter space, and thus the OPS method does not produce the correct diagonal entries for branch lengths with MLEs equal to zero. Thus, our approach here, combining Eq. (7) with the OPS estimator, provides the best estimates (see dos Reis and Yang, 2011, for a discussion of this issue).

Simulated data

We conducted the following simulations to test IQ2MC. We reused six DNA alignments from Tao et al. (2020), simulated using a rooted 16-taxon phylogeny (Supplementary Fig. S1) with 5,000 sites using a strict clock under the GTR+G4 model with $\alpha = 0.25$ (Tavaré 1986; Yang 1994). The six 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, henceforth denoted as 1x, 2x, 4x, 6x, 8x, and 10x, respectively.

We also used the 16-taxon tree to simulate 42 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 phylogeny. 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 analysed under three fossil calibration settings: young, middle, and root. In the young calibration setting, the age of the youngest internal node (t_{15}) (Supplementary Fig. S1) 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). We note MCMCTree conditions the birth-death process on the age of the phylogeny's root to generate the prior on node ages without fossil calibrations. Thus, a calibration on the root is always required by MCMC

Tree. For the six 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 36 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.

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. 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 S1-S4 show the detailed prior distributions.

Table 1: Datasets used in the empirical analyses.

Dataset	Datatype	Taxa	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. For single models, IQ2MC applies 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) is used to determine the best-fit models. 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 IQ2MC analyses, we used the autocorrelated rates model, the same priors on the mean rate, rate diffusion, and calibration densities as the original publications. MCMCTree v4.10.7 (Rannala and Yang 2007) is 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 S10). That is, we ensured the effective sample size (ESS) was at least 200 for all sampled parameters after removing 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 optimisation routines than those implemented in PAML. This is particularly striking for the

optimisation of large amino acid phylogenies, which can take several days with PAML but only a few hours with IQ-TREE. 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 (Le et al. 2008, 2012; Quang et al. 2008; 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 is highly accurate under different simulation conditions

For DNA alignment simulations, branch lengths, gradients, and Hessians estimated 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 optimisation algorithms used in the two programs (Supplementary Fig. S2). Both programs also produced virtually identical 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. S3) and LG+G4 (Supplementary Fig. S8) models, and the results are essentially the same.

Wrong substitution models substantially bias divergence times under simulations

Simulation analyses on both DNA and amino acid data show that model violations substantially affect divergence time estimates, particularly at higher evolutionary rates. Figure 2 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.

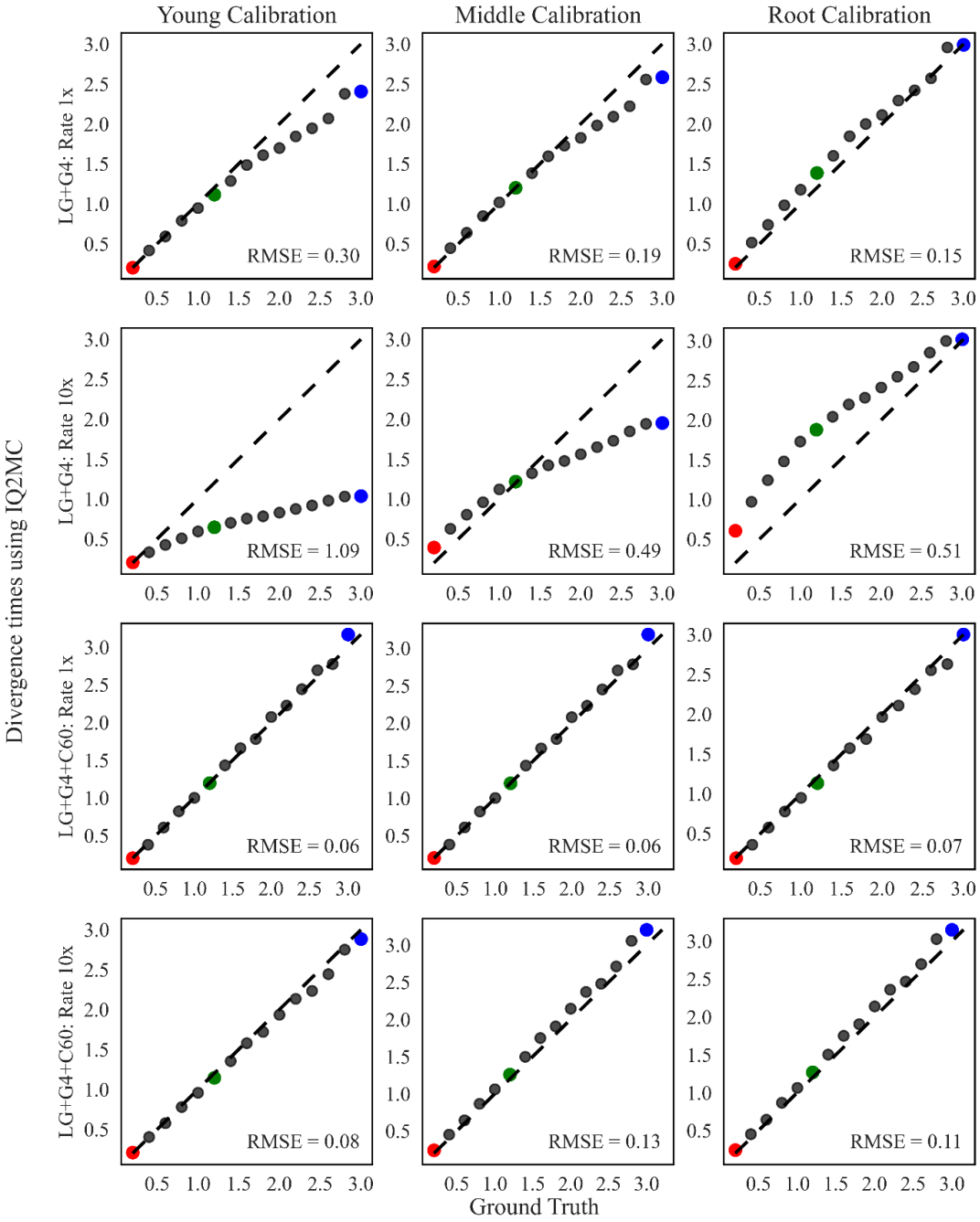

Figure 2: Comparison of estimated divergence times under LG+G4 (misspecified) and LG+G4+C60 (true) model to ground truth times under the IQ2MC workflow. Three different calibration settings are used with varying rates (1x and 10x). The columns represent the young, middle, and root calibration

settings. Young calibration setting (young node (red) t_{15} ~ 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 rates 1x and 10x rates for the LG+G4 model and third and fourth rows represent 1x and 10x rates for the LG+G4+C60 model. RMSE: root mean square error.

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. 2, 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. 2, 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, both times estimated under true and misspecified models are more similar 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. 2, 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 Figures S4 and S5). 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 Figures S9-S14).

Impact of partition and mixture models for time estimates in empirical data

DNA mixture models for divergence time estimation: Placental Mammals and Plant Data Analysis

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 250 million years of tree depth. This is a subset of data under the Afrotheria clade analyzed by (Álvarez-Carretero et al. 2022) using the sequential Bayesian sub-tree dating approach. We did not remove the taxa representing the outgroup, as in the sequential Bayesian sub-tree dating approach, the out-group is commonly used until the merging step of sub-trees to the backbone trees and is useful for transferring phylogenetic and calibration information from the backbone tree to the sub-tree (Álvarez-Carretero et al. 2022; Stiller et al. 2024).

Since we do not know the truth, 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 gradients and the Hessian under both substitution models and estimated divergence times under four different calibration settings (supplementary Tables S1 and S6). The Bayesian Information Criterion (BIC) score for the GTR+G4 model is 1,672,596.65, and the BIC score for the GTR mixture model is 1,661,753.49, which indicates a much better fit of the mixture model to the data. The linear regression slopes ranged from 1.05 to 1.13 with low dispersion ($R^2 = 0.99$) (Fig. 3a-3d). The highest linear regression slope (1.13) was obtained when the middle calibration setting was utilized. Under all analyses, the root age was underestimated by 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. The divergence time estimates for the outgroup showed substantial discrepancies when young, middle, and root calibration settings were used, specifically the GTR+G4 model underestimating divergence times (Fig. 3a–3c). However, these discrepancies disappeared when all calibrations were applied (Fig. 3d). This suggests that DNA mixture models can have a substantial impact on divergence time estimation, particularly in sequential Bayesian subtree dating approaches. Therefore, their effects should be thoroughly investigated for the robustness and accuracy of time tree inference.

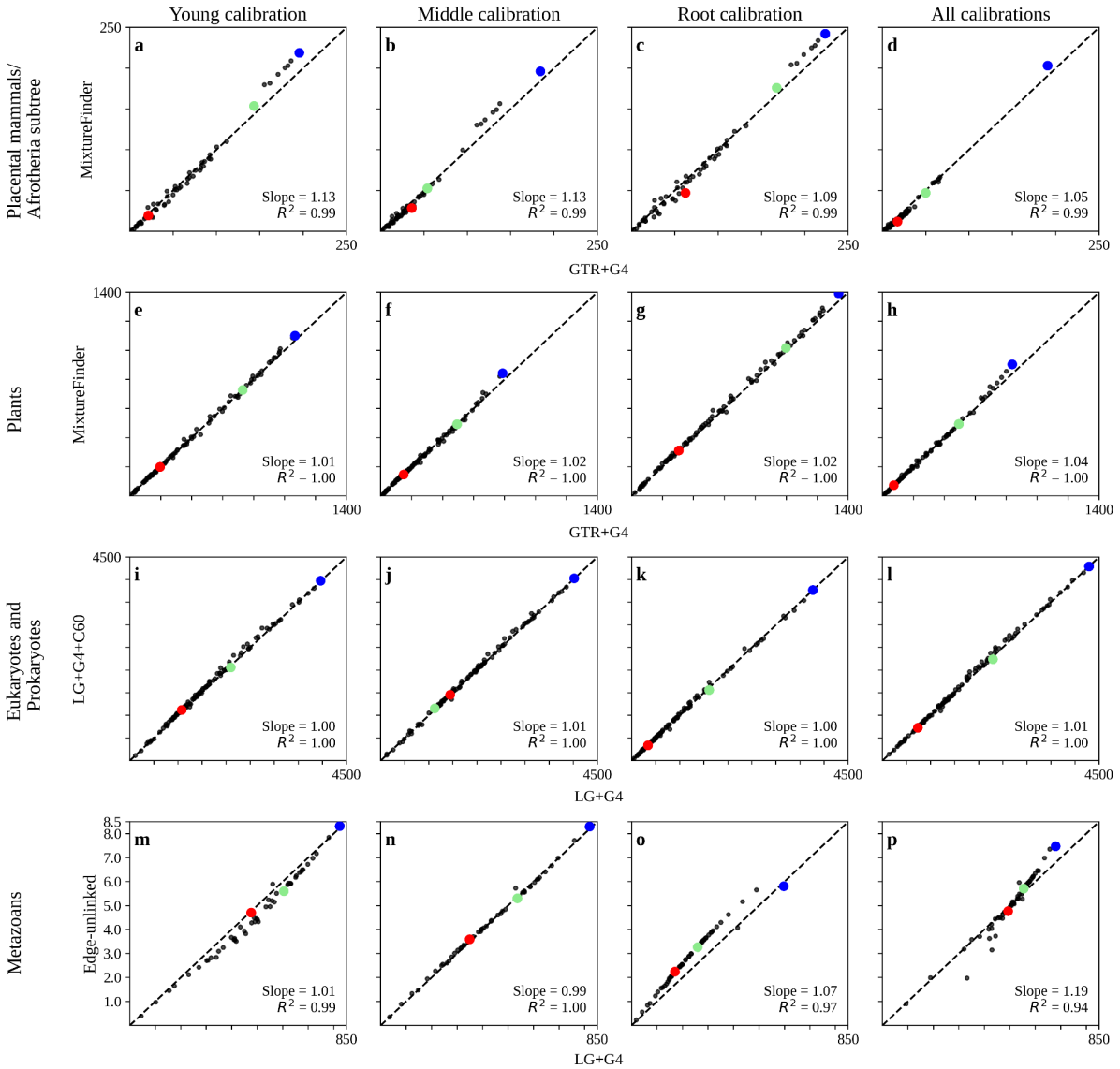

Figure 3: Comparison of estimated divergence times under complex and simple substitution models for empirical datasets. (a-d) divergence times comparison for the placental mammal/afrotheria subtree dataset under GTR+G4 and MxitureFinder with concatenated alignments. (e-h) divergence times comparison 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 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 Table S1-S4).

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 compared the times obtained with the GTR+G4 model (on a concatenated MSA) 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. S15) when using partition models instead of concatenated MSA.

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 MSA, 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 MSA as the simple model and a 4-component DNA mixture model estimated by the MixtureFinder as the complex model

(Supplementary Table S7) for the time estimates comparison. The BIC score for the GTR+G4 model is 217,754.25, and the BIC score for the DNA mixture model is 211,087.20, demonstrating the better fit of the DNA mixture model for 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 ($R^2 = 1$, Fig. 3e-3h).

For all the calibration settings, the GTR+G4 model underestimated the root age compared to the DNA mixture model, but fell within 95% CIs of the root age estimated by the DNA mixture model. 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 suggests that the effect of the substitution model on time estimates is data and context-dependent (Tao et al. 2020), demonstrating the importance of thoroughly evaluating model choice for accurate divergence time estimation.

Protein Mixture Models for Eukaryotes and Prokaryotes Data

We next reanalyze a 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 MSA. We calculated the gradients and Hessian matrix for each substitution model (Supplementary Table S8) and estimated divergence times under four different calibration settings (Supplementary Table S3). The BIC score for the LG+G4 model is 2,329,433.11, and the BIC score for the LG+G4+C60 model is 2,312,129.03, suggesting a better fit of the 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 (Fig. 3i-3l). Under the young, middle, and all-calibration settings, the LG+G4 model exhibited a slight tendency to underestimate divergence times compared to the profile mixture model, though the estimates remained within 95% CIs of the profile mixture model. 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 S8). 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 Fig. S16 and S17).

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 S9). The LG+G4 model yielded a BIC score of 2,152,878.01 while the EUL model had a lower BIC score of

2,145,256.68, 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 S4).

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 (Fig. 3m, 3o, and 3p). Especially, under the young calibration setting, the LG+G4 model overestimates the time estimates against the EUL partition model (Fig. 3m) with a slope of 1.01 and R^2 of 0.99. Furthermore, under the root-only calibration settings (Fig. 3o), the LG+G4 model underestimates divergence time with a slope of 1.07 and R^2 of 0.97. However, both models produced similar root age estimates across the two calibration settings. The highest discrepancies in time estimates occur under the all-calibration setting (Fig. 3p), with a slope of 1.19 and R^2 0.94. The LG+G4 model underestimates the root age and the divergence times for some nodes while overestimating others compared to the EUL model.

In addition to the comparison of the LG+G4 model to the EUL partition model, we compared the time estimates from the LG+G4 model against the LG+G4+C60 on a concatenated MSA 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 with a slope of 1.03 and R^2 of 0.99, even with all calibrations applied. The time estimates of the LG+G4 model were consistent with the estimates obtained from the EL model. However, LG+G4 underestimated the times of the internal nodes with a slope of 1.05 and R^2 of 1 under the root-only calibration setting (Supplementary Fig. S18).

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 to optimize and parallelize computations across multiple CPU cores. All analyses are 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 speed up under a single-threaded mode, while with 10 threads, a 92.5x speed up 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 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 data (300 taxa × 500k sites)			Simulated amino acid data (300 taxa × 500k sites)		
	Runtime* (hh:mm:ss)	Memory	Speedup	Runtime (hh:mm:ss)	Memory	Speedup
PAML	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 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 MSAs with varying numbers of sites and taxa (Supplementary Fig. S19 and S20). 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 S5). 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 MSA took approximately 60 hours (Supplementary Table S11). 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 pre-calculation step of the gradient and the Hessian 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 on simulated and empirical data analysis, we show that IQ2MC can facilitate more reliable time tree estimates.

Importantly, IQ2MC introduces the usage of DNA mixture models and amino acid profile mixture models, which were previously unavailable with MCMCTree. These models, designed to account for site-specific variations 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. 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, as has also been pointed out by Wang and Luo (2025). Finally, the impact of advanced substitution models on divergence times with different clock models should be further investigated, especially in the context of mixture models, which leads to a potential future research direction.

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: `doi.org/10.5281/zenodo.15321227`

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 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 S., Tamuri A.U., Battini M., Nascimento F.F., Carlisle E., Asher R.J., Yang Z., Donoghue P.C.J., Dos Reis M. 2022. A species-level timeline of mammal evolution integrating phylogenomic data. Nature. 602:263–267.

Barba-Montoya J., Tao Q., Kumar S. 2020. Using a GTR+Γ substitution model for dating sequence divergence when stationarity and time-reversibility assumptions are violated. Bioinformatics. 36:i884–i894.

Battistuzzi F.U., Billing-Ross P., Paliwal A., Kumar S. 2011. Fast and slow implementations of relaxed-clock methods show similar patterns of accuracy in estimating divergence times. Mol. Biol. Evol. 28:2439–2442.

Betts H.C., Puttick M.N., Clark J.W., Williams T.A., Donoghue P.C.J., Pisani D. 2018. Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. Nat. Ecol. Evol. 2:1556–1562.

Chernomor O., von Haeseler A., Minh B.Q. 2016. Terrace aware data structure for phylogenomic inference from supermatrices. Syst. Biol. 65:997–1008.

Drummond A.J., Bouckaert R.R. 2015. Bayesian Evolutionary Analysis with BEAST. Cambridge University Press.

Drummond A.J., Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7:214.

Jukes T.H., Cantor C.R. 1969. Evolution of Protein Molecules. Mammalian Protein Metabolism. Elsevier. p. 21–132.

Kainer D., Lanfear R. 2015. The effects of partitioning on phylogenetic inference. Mol. Biol. Evol. 32:1611–1627.

Kalyaanamoorthy S., Minh B., Wong T.K.F., von Haeseler A., Jermiin L. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat. Methods. 14:587–589.

Lanfear R., Calcott B., Kainer D., Mayer C., Stamatakis A. 2014. Selecting optimal partitioning schemes for phylogenomic datasets. BMC Evol. Biol. 14:82.

Lartillot N., Brinkmann H., Philippe H. 2007. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. BMC Evol. Biol. 7 Suppl 1:S4.

Lartillot N., Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol. Biol. Evol. 21:1095–1109.

Le S.Q., Dang C.C., Gascuel O. 2012. Modeling protein evolution with several amino acid replacement matrices depending on site rates. Mol. Biol. Evol. 29:2921–2936.

Le S.Q., Gascuel O. 2008. An improved general amino acid replacement matrix. Mol. Biol. Evol. 25:1307–1320.

Le S.Q., Lartillot N., Gascuel O. 2008. Phylogenetic mixture models for proteins. Philos. Trans. R. Soc. Lond. B Biol. Sci. 363:3965–3976.

Letunic I., Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Research. 49:W293–W296.

Ly-Trong N., Barca G.M.J., Minh B.Q. 2023. AliSim-HPC: parallel sequence simulator for phylogenetics. Bioinformatics. 39:btad540.

Marin J., Battistuzzi F.U., Brown A.C., Hedges S.B. 2017. The timetree of prokaryotes: New insights into their evolution and speciation. Mol. Biol. Evol. 34:437–446.

Morris J.L., Puttick M.N., Clark J.W., Edwards D., Kenrick P., Pressel S., Wellman C.H., Yang Z., Schneider H., Donoghue P.C.J. 2018. The timescale of early land plant evolution. Proc. Natl. Acad. Sci. U. S. A. 115:E2274–E2283.

Porter J. 2002. Effciency of covariance matrix estimators for maximum likelihood estimation. J. Bus. Econ. Stat. 20:431–440.

Quang L.S., Gascuel O., Lartillot N. 2008. Empirical profile mixture models for phylogenetic reconstruction. Bioinformatics. 24:2317–2323.

Rambaut A. 2009. FigTree. Tree figure drawing tool. http://tree. bio. ed. ac. uk/software/figtree/.

Rannala B., Yang Z. 1996. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. J. Mol. Evol. 43:304–311.

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

dos Reis M., Thawornwattana Y., Angelis K., Telford M.J., Donoghue P.C.J., Yang Z. 2015. Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. Curr. Biol. 25:2939–2950.

dos Reis M., Yang Z. 2011. Approximate likelihood calculation on a phylogeny for Bayesian estimation of divergence times. Mol. Biol. Evol. 28:2161–2172.

dos Reis M., Zhu T., Yang Z. 2014. The impact of the rate prior on Bayesian estimation of divergence times with multiple loci. Syst. Biol. 63:555–565.

Ren H., Wong T.K.F., Minh B.Q., Lanfear R. 2024. MixtureFinder: Estimating DNA mixture models for phylogenetic analyses. bioRxiv.:2024.03. 20.586035.

Romiguier J., Borowiec M.L., Weyna A., Helleu Q., Loire E., La Mendola C., Rabeling C., Fisher B.L., Ward P.S., Keller L. 2022. Ant phylogenomics reveals a natural selection hotspot preceding the origin of complex eusociality. Curr. Biol. 32:2942–2947.e4.

Ronquist F., Teslenko M., van der Mark P., Ayres D.L., Darling A., Höhna S., Larget B., Liu L., Suchard M.A., Huelsenbeck J.P. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61:539–542.

Seo T.-K., Kishino H., Thorne J.L. 2004. Estimating absolute rates of synonymous and nonsynonymous nucleotide substitution in order to characterize natural selection and date species divergences. Mol. Biol. Evol. 21:1201–1213.

Soubrier J., Steel M., Lee M.S.Y., Der Sarkissian C., Guindon S., Ho S.Y.W., Cooper A. 2012. The influence of rate heterogeneity among sites on the time dependence of molecular rates. Mol. Biol. Evol. 29:3345–3358.

Stiller J., Feng S., Chowdhury A.-A., Rivas-González I., Duchêne D.A., Fang Q., Deng Y., Kozlov A., Stamatakis A., Claramunt S., Nguyen J.M.T., Ho S.Y.W., Faircloth B.C., Haag J., Houde P., Cracraft J., Balaban M., Mai U., Chen G., Gao R., Zhou C., Xie Y., Huang Z., Cao Z., Yan Z., Ogilvie H.A., Nakhleh L., Lindow B., Morel B., Fjeldså J., Hosner P.A., da Fonseca R.R., Petersen B., Tobias J.A., Székely T., Kennedy J.D., Reeve A.H., Liker A., Stervander M., Antunes A., Tietze D.T., Bertelsen M.F., Lei F., Rahbek C., Graves G.R., Schierup M.H., Warnow T., Braun E.L., Gilbert M.T.P., Jarvis E.D., Mirarab S., Zhang G. 2024. Complexity of avian evolution revealed by family-level genomes. Nature. 629:851–860.

Tao Q., Barba-Montoya J., Huuki L.A., Durnan M.K., Kumar S. 2020. Relative efficiencies of simple and complex substitution models in estimating divergence times in phylogenomics. Mol. Biol. Evol. 37:1819–1831.

Tavaré S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. Lecture of mathematics for life science.

Thorne J.L., Kishino H., Painter I.S. 1998. Estimating the rate of evolution of the rate of molecular evolution. Mol. Biol. Evol. 15:1647–1657.

Wang H.-C., Minh B.Q., Susko E., Roger A.J. 2018. Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. Syst. Biol. 67:216–235.

Wang S., Luo H. 2025. Dating the bacterial tree of life based on ancient symbiosis. Syst. Biol.

Wong T.K.F., Ly-Trong N., Ren H., Baños H., Roger A.J., Susko E., Bielow C., De Maio N., Goldman N., Hahn M.W., Others. 2025. IQ-TREE 3: Phylogenomic Inference Software using Complex Evolutionary Models. .

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

Yang Z. 1995. A space-time process model for the evolution of DNA sequences. Genetics. 139:993–1005.

Yang Z. 2000. Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A. J. Mol. Evol. 51:423–432.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 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. Mol. Biol. Evol. 23:212–226.

Yang Z., Rodríguez C.E. 2013. Searching for efficient Markov chain Monte Carlo proposal kernels. Proc. Natl. Acad. Sci. U. S. A. 110:19307–19312.