

Molecular clock dating using complex mixture models: applied to ancient symbionts

Article

Published Version

Creative Commons: Attribution-Noncommercial 4.0

Open Access

Wang, S. ORCID: <https://orcid.org/0000-0002-7220-7305> and Meade, A. ORCID: <https://orcid.org/0000-0001-7095-7711> (2026) Molecular clock dating using complex mixture models: applied to ancient symbionts. *Molecular Biology and Evolution*, 43 (3). msag039. ISSN 1537-1719 doi: 10.1093/molbev/msag039 Available at <https://centaur.reading.ac.uk/129527/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1093/molbev/msag039>

Publisher: Oxford University Press

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).


www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Molecular clock dating using complex mixture models: applied to ancient symbionts

Sishuo Wang *¹, Andrew Meade²

¹Department of Microbiology, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR, China

²School of Biological Sciences, University of Reading, Reading, UK

*Corresponding author: E-mail: sishuowang@hotmail.ca; sishuowang@cuhk.edu.hk

Associate editor: Nicolas Rodrigue

Abstract

Molecular clocks are a fundamental technique in evolutionary biology for establishing the timing and tempo of organismal divergence. However, currently available molecular clock methods, which often rely on simple homogeneous substitution models, can produce inaccurate time estimates, particularly for deep-time or rapidly evolving lineages where substitution heterogeneity and saturation are common. Hereby, we introduce *phyloHessian* (<https://github.com/evolbeginner/phyloHessianWrapper>), a Julia-based software to enable the use of complex mixture substitution models in molecular dating. *phyloHessian* computes the phylogenetic Hessian matrix and integrates it into PAML-MCMCtree's approximate likelihood framework to conduct dating analyses. Simulations mimicking phylogenies at different timescales demonstrate that complex mixture substitution models significantly enhance the accuracy of divergence time and substitution rate estimates in deep-time phylogenies. This pattern remains consistent across a wide range of uncertainties associated with molecular clock analysis. Additionally, mixture models display greater robustness to model and calibration specifications compared to their homogeneous counterparts. Empirical analysis of ancient symbiont lineages *Microsporidia* and *Rickettsiales* with different substitution models shows that mixture models, compared to homogeneous models, yield accelerated substitution rates and in some cases significantly different divergence times, leading to a revised understanding of their host association origins. Our findings underscore the importance of incorporating complex mixture substitution models for constructing reliable evolutionary timelines and elucidating the evolutionary history of deep-time or fast-evolving lineages.

Keywords molecular clock, PAML, MCMCtree, mixture model, substitution model, deep-time evolution, approximate likelihood, symbiosis zero, zero

Introduction

Historically, our understanding of evolutionary timescales relied on limited fossil records, until a significant change occurred in the 1960s when molecular clock theory was proposed (Zuckerandl and Pauling 1965), making it possible to infer evolutionary time and rate using genetic data. While the original concept of the molecular clock assumed that the rate of gene sequence changes is constant among lineages, further evidence shows that even the same gene can exhibit variations in rate among different species, leading to the development of relaxed molecular clock models which allow for lineage-specific evolutionary rate and a series of relevant models (Thorne et al. 1998; Drummond et al. 2006; Yang and Rannala 2006). Over the years, the molecular clock theory has evolved, continuously increasing in complexity and accuracy, providing a powerful tool for studying biological evolution. Particularly, with the development of Bayesian phylogenetic methods and the rapid growth of

genetic sequence data, Bayesian models based on Markov chain Monte Carlo (MCMC) methods have become the most widely used molecular clock dating method, which allows for a comprehensive examination of various uncertainties in analysis, including factors such as calibrations and across-lineage rate differences (Ho and Duchêne 2014; dos Reis et al. 2016).

While molecular clock approaches offer a powerful framework for inferring evolutionary timescales, the reliability of such inferences heavily depends on accurately modeling the evolution of the molecular data, such as protein sequences, alongside the incorporation of calibration information (Guindon 2020). Protein evolution is typically modeled as site-homogeneous process using empirically derived substitution matrices of amino acid exchangeabilities (e.g. LG or WAG) and fixed equilibrium frequencies, with variation in evolutionary rates modeled with a gamma distribution (Yang 1993). However, while computationally efficient, such models oversimplify real-world biological processes by failing to capture the heterogeneity in amino acid

Received: July 18, 2025. **Revised:** February 1, 2026. **Accepted:** February 4, 2026

© The Author(s) 2026. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

preferences driven by diverse structural and functional constraints across sites. To address these issues, site-heterogeneous substitution models have been proposed since (Lartillot and Philippe 2004; Pagel and Meade 2004). These models account for among-site substitution heterogeneity by incorporating pre-defined substitution matrices for different secondary structures and solvent accessibility (Le et al. 2008; Le and Gascuel 2010), across-site rate variation (Le et al. 2012), or mixtures of amino acid frequency profiles (Quang et al. 2008; Wang et al. 2008). Of particular interest is the last category, which captures different amino acid preferences across sites, as represented by the CAT model (Lartillot and Philippe 2004) and its simplified counterparts for maximum likelihood phylogenetics called Cxx (where xx indicates 10 to 60 profile categories) (Quang et al. 2008). By better approximating the true biological complexity of amino acid substitution, profile mixture models mitigate artifacts like long branch attraction, thereby enhancing the ability to distinguish phylogenetic signal from homoplasy and displaying higher sensitivity to substitution saturation (Wang et al. 2018; Wang et al. 2019; Schrepf et al. 2020; Baños et al. 2024b). These features make profile mixture models especially well-suited for studying microbial evolution, where deep divergences and rapidly evolving sites are prevalent. Indeed, empirical analyses of deep-time phylogenomic data show that profile mixture models often exhibit superior model fit compared to homogeneous alternatives (Williams et al. 2020; Bujaki and Rodrigue 2022).

A particular outcome of using site-homogeneous models like LG+G is that branch lengths are usually underestimated in case of extensive sequence divergence (Venditti et al. 2008; Moody et al. 2022). This happens because proteins may not use all 20 amino acids at each position and homogenous models incorrectly interpret this restricted use of amino acids as evidence of less evolutionary change than what truly occurred over long time (Lartillot et al. 2007; Wang et al. 2008). To see this, consider a site restricted to Asp and Glu. A mixture model would identify a preference for negatively charged amino acids and infer unobserved substitutions (e.g. Asp → Glu → Asp) even if two sequences show the same amino acid, leading to longer branch lengths. Obviously, this pattern is more often observed with an increasing evolutionary scale. Studies have shown that the deepest branches of the tree of life, spanning approximately 4.0 billion years (Gyr), can display branch lengths up to twice as long under profile mixture models compared to the LG+G model (Moody et al. 2022; Wang and Luo 2025). Given that in molecular clock analysis, branch length, the expected substitutions per site, is the product of time and (average) substitution rate of that branch (dos Reis et al. 2016; Guindon 2020), an important question emerges: Do profile mixture models, which often excel in deep-time phylogenetic inference, lead to more accurate divergence time and substitution rate estimates?

To our knowledge, none of the available molecular clock programs have directly implemented complex substitution models including those profile mixture models, with PhyloBayes' CAT model being the exception—yet its high computational demands may limit its use in large-scale molecular clock analyses. We recently developed *bs_inBV* (Wang and Luo 2025), enabling the use of diverse substitution models in the popular MCMCtree molecular dating software by applying a bootstrap approximation of the Hessian matrix, which captures likelihood surface curvature and is required by MCMCtree's fast approximate likelihood dating

procedure (Thorne et al. 1998; dos Reis and Yang 2011). However, the bootstrapping process is computationally intensive, and its approximation sometimes inadequately represents the true Hessian. Additionally, and perhaps more importantly, the relative performance of different substitution models in molecular clock dating remains unexamined through systematic simulation studies, leaving a critical gap in our understanding of how choice in substitution model affects divergence time estimation, particularly in deep-time or fast-evolving phylogenies.

Hereby, we introduce *phyloHessian*, a Julia-based software that uses numerical methods to calculate accurate Hessian matrices for diverse complex substitution models and integrates them directly into MCMCtree's approximate likelihood framework for rapid molecular clock dating. We perform extensive analyses on simulation data to comprehensively compare the performance of different substitution models in divergence time and substitution rate estimation under a wide range of conditions that mimic phylogenies from relatively recent to very ancient divergences. We further analyze empirical data to assess how different substitution models affect our understanding of ancient symbiont lineages and the origins of their host associations.

Results

phyloHessian: enabling molecular clock dating in MCMCtree with complex substitution models

To address the challenges in molecular clock analysis under various substitution models, we developed a Julia-based software *phyloHessian*. *phyloHessian* first calls IQ-Tree (Wong et al. 2025) or PhyML (Guindon et al. 2010), two widely used phylogenetic reconstruction software that implement a range of mixture models, to calculate the maximum likelihood estimates (MLE) of branch length $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_n)^T$ under user-specified substitution model. Let $\theta = (\theta_1, \dots, \theta_n)$ be the branch lengths and $\Delta\theta = \theta - \hat{\theta}$. Next, it applies finite difference numerical methods to calculate the gradient $g = \left(\frac{\partial \ell(\hat{\theta})}{\partial \theta_1}, \dots, \frac{\partial \ell(\hat{\theta})}{\partial \theta_n} \right)^T$ and Hessian matrix $H(\hat{\theta}) = \left(\frac{\partial^2 \ell(\hat{\theta})}{\partial \theta_i \partial \theta_j} \right)$, i.e. the first- and second-order derivatives of the log-likelihood of the phylogenetic data evaluated at the MLEs. Last, the resulting Hessian matrix is integrated into MCMCtree to be used with its approximate likelihood method (Thorne et al. 1998; dos Reis and Yang 2011) for divergence time and rate estimation. MCMCtree's approximate likelihood method approximates the log-likelihood surface for branch lengths using a Taylor expansion to the second order:

$$l(\theta) \approx l(\hat{\theta}) + g^T \Delta\theta + \frac{1}{2} \Delta\theta^T H(\hat{\theta}) \Delta\theta.$$

To demonstrate the reliability of *phyloHessian*, we compared its results with those obtained from well-established phylogenetics software packages. First, we found that *phyloHessian* and IQ-Tree produced identical log-likelihoods under all tested complex substitution models available in IQ-Tree (Data S1). Further, we validated *phyloHessian*'s Hessian calculation by confirming

that the Hessian matrix obtained by phyloHessian and CODEML under site-homogeneous models (such as LG+G) was nearly identical (Data S2).

Profile mixture substitution models enhance time and rate estimation in deep-time and fast-evolving phylogenies: evidence from simulations

To evaluate the impact of mixture substitution models (Materials and Methods: Mixture substitution model) on divergence time and substitution rate estimates and to compare their performance to profile homogeneous models like LG+G, we conducted extensive simulations. To simulate a range of evolutionary scenarios spanning relatively recent to very deep-time divergences, we followed our prior study (Wang and Luo 2025) to generate 30 replicates of 20-tip birth-death trees with protein sequences that were 300 amino acids (aa) in length simulated under the substitution model LG+C60+G4{1.0} for both autocorrelated-rates (AR) and independent-rates (IR) molecular clock rate model. For simplicity, we use “+G” to represent “+G4{1.0}” hereafter [IQ-Tree’s default; four-category discrete Gamma distribution for the relative rate, shape parameter $\alpha=1.0$, i.e. Gamma(1,1)]; see also Materials and Methods: Simulation procedure. We mainly focused on five calibration strategies. These five calibration strategies consisted of a single calibration at the root with both upper and lower bounds (*root_only*), fully calibrated root plus one fully-bounded internal node (*single_interval*), fully calibrated root plus one minimum-bounded internal node (*single_min*), fully calibrated root plus two fully-bounded internal nodes (*two_interval*), and fully calibrated root plus two minimum-bounded internal nodes (*two_mins*) (Fig. 1c).

Under the focal simulation scheme (see Materials and Methods: Simulation procedure), the profile mixture model LG+C60+G performed comparably to the traditional LG+G model for divergence time estimation when simulated root ages were set to either 1.0 or 2.0 billion years ago (Ga). However, for root ages ≥ 3.0 Ga, LG+C60+G obtained significantly more accurate time estimates than LG+G, with the performance gap becoming increasingly pronounced at 4.0 Ga (Fig. 1a). For example, under the *root_only* calibration strategy, LG+C60+G demonstrated superior time estimation accuracy (mean relative difference \pm SD: 0.157 ± 0.061 for AR clock model, 0.139 ± 0.046 for IR clock model) compared to LG+G (AR: 0.193 ± 0.062 ; IR: 0.197 ± 0.061). Mean relative difference is defined as $\text{mean}\left(\frac{|b_i^m - b_i^{\text{true}}|}{\max(b_i^m, b_i^{\text{true}})}\right)$, where b_i^m and b_i^{true} denote the inferred posterior mean age and the true value for node i , respectively. The largest discrepancies in divergence time estimation occurred under the calibration scheme *two_min*, where LG+C60+G showed lower error than LG+G (mean relative difference: AR 0.160 vs. 0.232; IR 0.147 vs. 0.259; Fig. 1a). In contrast to divergence time estimates, LG+C60+G provided much more accurate branch-specific substitution rate estimates than LG+G across all tested root ages (Fig. 1b). The error reduction obtained by LG+C60+G compared with LG+G, calculated as $\frac{\text{reldiff}_{\text{LG+G}} - \text{reldiff}_{\text{LG+C60+G}}}{\text{reldiff}_{\text{LG+G}}}$ (i.e. the error reduction was calculated using the node-averaged mean relative differences between

models), averaged across all calibration strategies, ranged from 26% to 53% under AR and 23% to 55% under IR for root ages of 1.0 to 4.0 Ga (Table 1). The pattern held true with alternative metrics (mean branch score distance [BSD] or scatter-plot comparisons) for comparison with the simulation values (Figures S1 to S3).

Next, we explored the patterns for more complex mixture models. We combined structurally constrained substitution models EX2 (exposed/buried sites) and EX3 (highly exposed/intermediate/buried sites) (Le et al. 2008) with the predefined equilibrium frequency C20 to account for site-specific variation in amino acid profiles by 20 categories of equilibrium frequencies. This resulted in 40 and 60 different rate matrices for EX2 and EX3 respectively. When simulation was done under profile-exchangeability mixture models EX2+C20+G and EX3+C20+G, the above patterns became more pronounced (Figure S4). In contrast to profile mixture models, when alignments were simulated under exchangeability mixture models (EX2+G, EX3+G, and LG4M), LG+G and the true simulation model performed similarly in estimating divergence times (Figure S5).

Large sequence divergence may arise from not only long divergence time, but also rapid evolution. To test whether different time estimates are affected by varying evolutionary rates (Figure S6), we simulated alignments on phylogenies with a fixed root age of 1.0 Ga but different mean absolute substitution rates along the phylogeny (0.25 to 1.0 substitutions/site/Gyr). This represents evolutionary rates from relatively low, as in most microbes, to very fast, as observed in some pathogens. We would anticipate comparable sequence divergence whether we fixed the root age at 1.0 Ga and increased the average substitution rate (from 0.25 to 1.0 substitutions/site/Gyr; Figure S6), or fixed the substitution rate at 0.25 substitutions/site/Gyr and increased the root age to 4.0 Ga (Fig. 1). Averaged across all calibration strategies, LG+C60+G displayed error reductions in substitution rate estimate of 65% (AR) and 61% (IR) at a mean substitution rate of 1.0 substitutions/site/Gyr.

Performance of profile mixture models in molecular clock dating under varied simulation conditions

Next, we evaluated the influence of multiple sources of uncertainties in molecular clock dating on the performance of profile mixture and traditional homogenous models when sequence alignments are simulated under profile mixture models. These involve sequence length, Gamma-distributed across-site rate, birth-death process, variance in lineage-specific rates, and number of sequences (see Note S1 for details).

We first investigated how sequence length (ranging from 100 to 2,000 aa) affected the accuracy of the estimates using different substitution models (Figure S7). Although the increase in divergence time estimate accuracy achieved by using LG+C60+G instead of LG+G was minimal for 100-aa alignments, the improvement became increasingly apparent as the alignment length extended to 2,000 aa. Across all analyzed calibration strategies, the LG+C60+G model consistently improved the accuracy of time estimation for the AR clock models compared to LG+G. For a simulated root age of 4.0 Ga, this improvement was

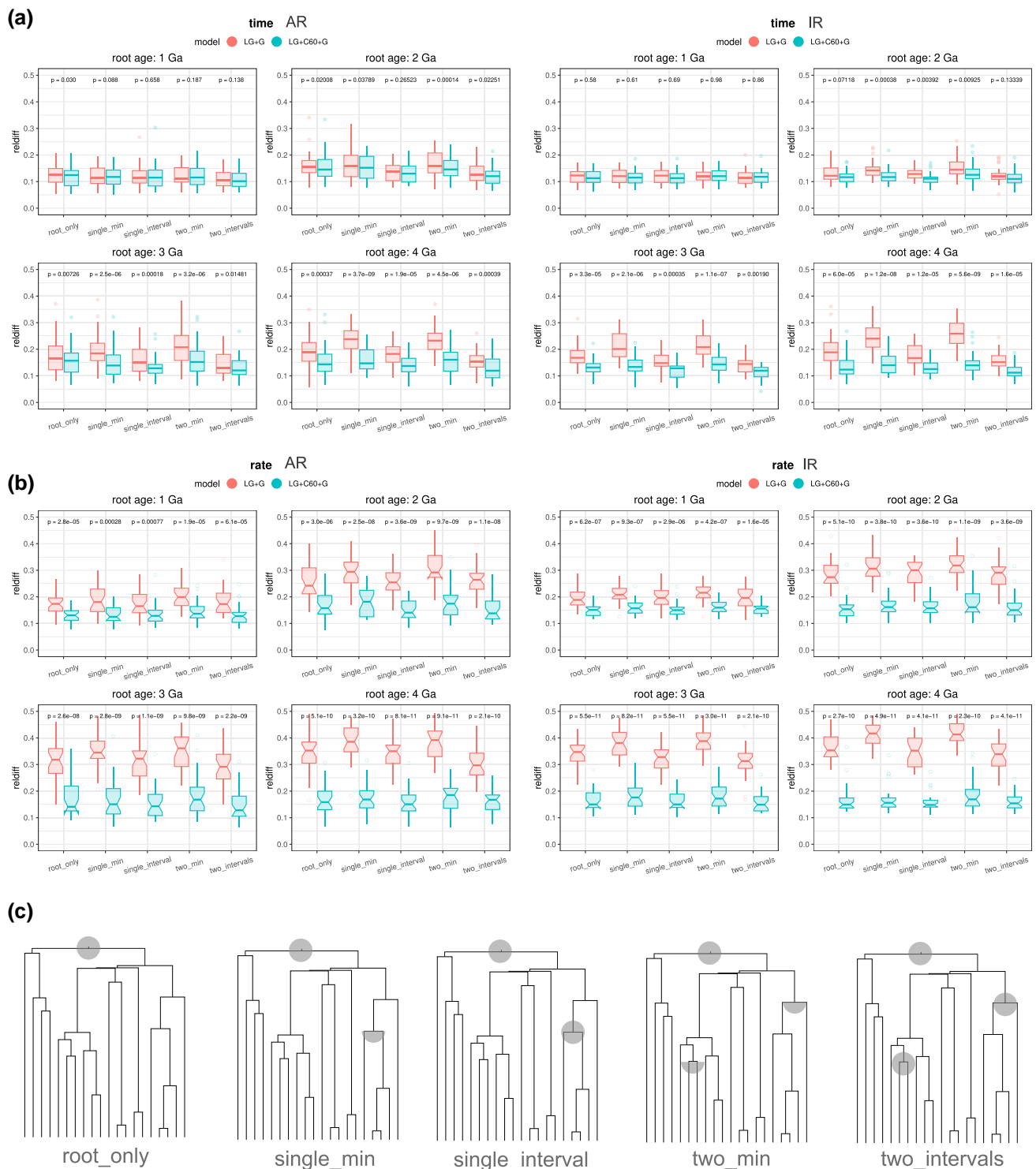


Figure 1 Comparison of the accuracy of divergence time and branch-specific substitution rate estimation under substitution models LG + G and LG + C60 + G by simulation. a and b) Each plot contains 30 dots representing the mean relative difference (compared to the true values used in simulation) based on MCMCtree molecular clock dating with the corresponding calibration strategy on 30 simulated datasets. For each simulation, 30 timetrees each containing 20 tips are simulated using a birth-death process with a mean substitution rate 0.25 substitutions/site/Gyr across four different root ages ranging from 1.0 to 4.0 Ga. Sequence alignments are generated with LG + C60 + G4[1.0] (+G4[1.0] indicates $\alpha = 1$ in a four-category discrete Gamma-distributed across-site relative rates) under AR (left-hand side) and IR (right-hand side) clock models, respectively. *P*-values are obtained from a paired Wilcoxon signed-rank test. The box plots (median, interquartile range, 1.5×IQR whiskers) illustrate the relative differences in time a) and rate b) estimates compared to true values. c) A schematic figure showing the five calibration schemes used for molecular clock analyses on simulated data. *root_only*: a single bounded calibration (both upper and lower bounds); *single_min*: bounded calibration at the root and one minimum-bounded internal node; *single_interval*: bounded calibration at both the root and an internal node; *two_min*: bounded calibration at the root and two minimum-bounded internal nodes; *two_intervals*: bounded calibration at both the root and two internal nodes.

Table 1 Reduction of the relative differences in substitution rate estimates, calculated as $\frac{\text{reldiff}_{\text{LG} + \text{G}} - \text{reldiff}_{\text{LG} + \text{C60} + \text{G}}}{\text{reldiff}_{\text{LG} + \text{G}}}$, under LG + G versus LG + C60 + G under different root ages and calibration strategies.

Calibration strategy	AR (root age)				IR (root age)			
	1.0 Ga	2.0 Ga	3.0 Ga	4.0 Ga	1.0 Ga	2.0 Ga	3.0 Ga	4.0 Ga
<i>root_only</i>	26%	35%	47%	51%	23%	44%	52%	55%
<i>single_min</i>	27%	38%	53%	56%	24%	46%	52%	59%
<i>single_interval</i>	23%	41%	51%	55%	24%	44%	49%	55%
<i>two_min</i>	28%	42%	50%	55%	24%	45%	52%	57%
<i>two_intervals</i>	28%	42%	50%	48%	21%	42%	50%	51%
Average	26%	40%	50%	53%	23%	44%	51%	55%

substantial, resulting in an error reduction (see above) of 30%, 28%, 25%, and 12% for alignment lengths of 2,000, 1,000, 500, and 100 aa, respectively; the benefit was still significant but less pronounced for a younger root age of 3.0 Ga, with corresponding improvements of 22%, 18%, 15%, and 7%. Both LG + C60 + G and LG + G models achieved higher estimation accuracy with longer sequences (i.e. lower values in y axis from 100 to 2,000 aa), indicating that longer sequences are more informative for more accurate molecular clock analyses, consistent with the notion that mixture models perform much better on longer alignments for phylogenetic reconstruction (Baños et al. 2024b). Regarding the substitution rate, mixture models displayed more accurate estimates across timescales for different sequence lengths (Figure S8).

We further adjusted the gamma correction parameter (α) to modulate the variance of the across-site relative rates. Specifically, we set α to 0.5 and 1.5, which scaled the across-site variance to 2-fold ($0.5/0.5^2$) (scheme *Gamma_0.5* in Fig. 2a) and 2/3-fold ($1.5/1.5^2$) (scheme *Gamma_1.5* in Figure S9a), respectively, compared to our focal analysis ($\alpha=1$). These adjustments did not alter the patterns observed in the original analysis. Similarly, when exploring alternative birth-death process parameters, a flat birth-death process (simulation scheme: *flat_time_prior*) produced results comparable to those from the focal analysis (Fig. 2b). Under complete taxon sampling, the difference in time estimates between the LG + G and LG + C60 + G models became more obvious (scheme *complete_taxon_sampling* in Figure S9b). Moreover, increasing the standard deviation of the among-branch sequence absolute substitution rate to twice and three times its default value reduced the differences in divergence time estimation between using LG + G and LG + C60 + G, despite the unchanged overall pattern (schemes *SD_lograte_0.38*, *SD_lograte_0.55* in Figures S2c and S9c). In addition, although all simulations above were performed on 20-tip trees for simplicity, the analysis was extended to 40-, 60-, 80-, and 100-tip trees, with similar patterns found (Figure S10).

Profile mixture models display higher robustness to model misspecification

Further, we applied simulations to test the robustness of profile mixture models to misspecifications in calibrations or substitution models. To explore whether misspecified calibrations would affect estimation under different substitution models when the sequences were simulated under LG + C60 + G, we calibrated the nodes by adjusting their ages. These adjustments consisted

of fixed percentage increases (shifting ages toward the past) or decreases (shifting ages toward the present) (Fig. 3). In general, LG + C60 + G achieved higher accuracy when calibrations were older than the true divergence times (Fig. 3b). For a root age of 4.0 Ga and across all tested calibration strategies, on average, the relative difference in time estimation compared to true values was 0.256 (LG + G) versus 0.195 (LG + C60 + G) for the AR model and 0.278 (LG + G) versus 0.185 (LG + C60 + G) for the IR model. For root age of 3.0 Ga, these values changed to 0.252 (LG) versus 0.202 (LG + C60 + G) for AR and 0.255 (LG) versus 0.192 (LG + C60 + G) for IR (Fig. 3b). Subtle differences between using LG + G and LG + C60 + G were observed for root age ≤ 2.0 Ga. However, when calibrations were biased toward younger ages, the LG + G and LG + C60 + G models generally yielded comparable accuracy for divergence time and substitution rate estimates, irrespective of the root age, with significant differences arising occasionally.

The above analyses focused on alignments simulated under complex substitution models, particularly LG + C60 + G. To determine how these models perform in molecular clock analysis when sequences evolve under simpler substitution models, we performed additional simulations using these models and the previously described settings (varying the simulated root ages from 1.0 to 4.0 Ga (Fig. 3c)). These analyses indicate that the profile mixture model LG + C60 + G displayed robust time and rate estimation, regardless of the true underlying substitution process. The robustness of LG + C60 + G is evident in the following comparisons. When sequences were simulated under the basic LG + G model, LG + C60 + G provided estimates of divergence times and absolute substitution rates at least as accurate as those obtained using LG + G. Further, similar results were observed when simulations were conducted with other models with factors not explicitly modeled by LG + C60 + G, such as invariant sites (+I) and free rate across-site heterogeneity (+R) (Note S1). These findings suggest that LG + C60 + G's robustness was maintained even when the true evolutionary process involves additional complexities in substitution models.

Comparison between profile mixture models with posterior mean site frequency and bootstrapped Hessian approximations in molecular clock dating

The computational intensity of the LG + C60 + G model often necessitates the posterior mean site frequency (PMSF)

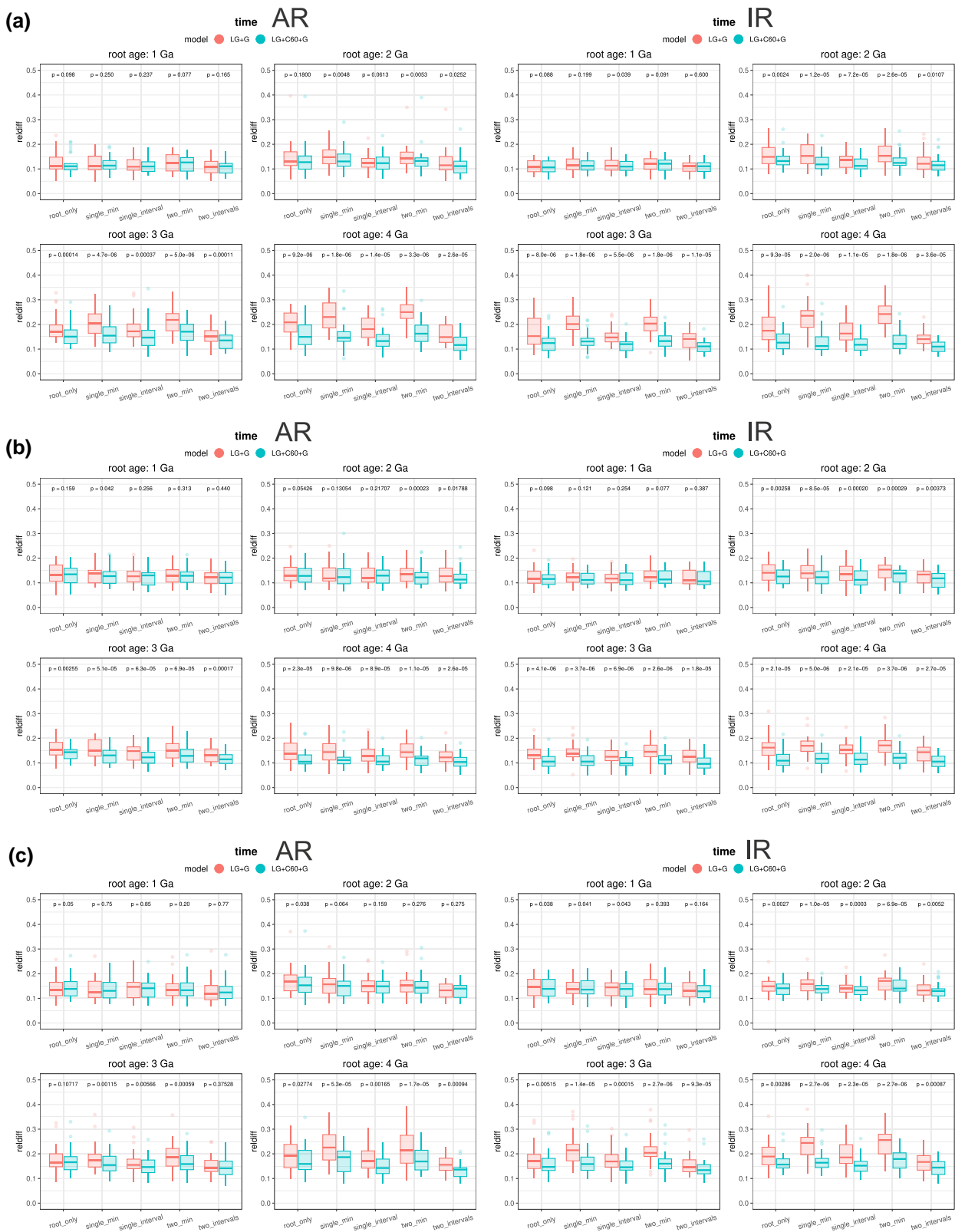


Figure 2 Investigating the robustness of divergence time estimation to alternative settings in simulation. a) Sequence simulated under substitution model LG + C60 + G4{0.5} (*Gamma*_{0.5}; the parameter alpha in the discrete Gamma distribution equal to 0.5), thus more rate heterogeneity across sites than the focal analysis (LG + C60 + G4{1.0}). b) Timetree simulated under a flat time prior (*flat_time_prior*). c) Timetree simulated under a molecular clock with larger across-branch rate variance (*var_lograte*_{0.38}). The plots on the left and right correspond to analyses using the AR and IR clock models, respectively. More detailed information about the above alternative simulation schemes can be found in [Note S1](#). Calibration strategies and other settings are the same as those used in [Fig. 1](#).

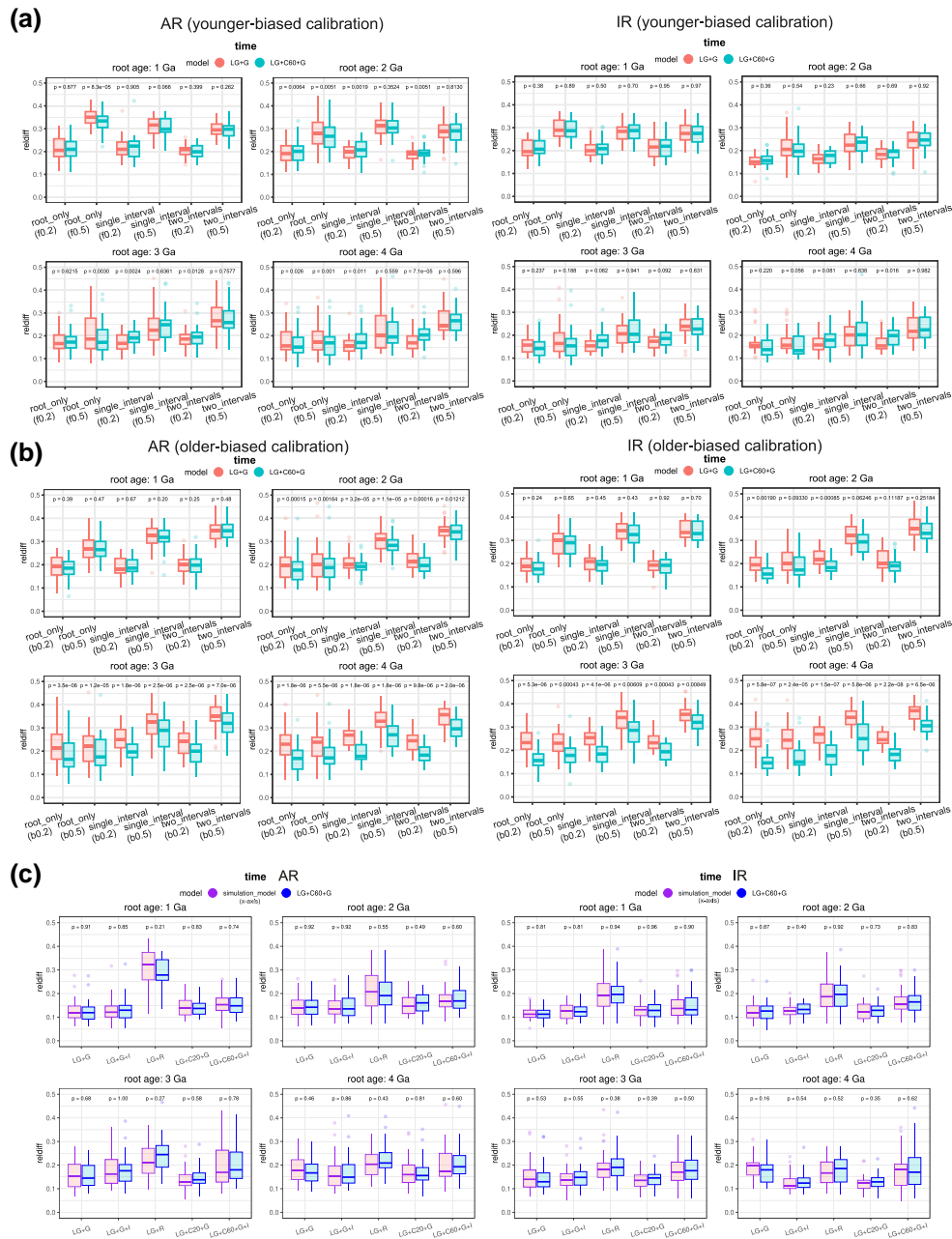


Figure 3 Evaluation of the impacts of calibration and model misspecification on molecular clock dating. Accuracy of divergence time estimation (relative difference) is shown as the y-axis. a and b) The lower and upper time bounds of all calibrations are shifted forward a) and backward b) in time by 20% (f0.2, b0.2) or 50% (f0.5, b0.5) of their true values, respectively. Three calibration schemes are considered: (i) a root-only calibration (*root_only*), (ii) the root plus one fully calibrated internal node (*single_interval*), and (iii) the root plus two fully calibrated internal nodes (*two_intervals*), each with time ranges of the true value $\pm 20\%$ or $\pm 50\%$. c) Assessment of the accuracy of divergence time estimation under alternative substitution models used in simulation (LG + C60 + G used in most other analyses): LG + G, LG + G + I, LG + R, LG + C20 + G, and LG + C60 + G + I. Boxplots compare the accuracy obtained by the true model (purple) and by LG + C60 + G (blue). Each plot contains 30 dots representing the mean relative difference based on MCMCtree molecular clock dating with the corresponding calibration strategy on 30 simulated datasets. For each simulation, 30 timetrees each containing 20 tips are simulated using a birth-death process with a mean substitution rate 0.25 substitutions/site/Gyr across four different root ages ranging from 1.0 to 4.0 Ga. The plots on the left and right correspond to analyses using the AR and IR clock models, respectively. Calibration strategies are the same as those used in Fig. 1.

approximation to accelerate phylogenetic reconstruction, especially for large datasets (Wang et al. 2018). This occurs because PMSF requires the computationally intensive pruning algorithm to run only once per site pattern using an approximated equilibrium frequency for each site, versus c times for a mixture model

with c classes. A distinct approximation, previously employed in molecular clock analyses under complex substitution models, uses a bootstrap-based method to approximate the Hessian matrix and is implemented in the software bs_inBV (Wang and Luo 2025). To assess the impact of these two approximations on

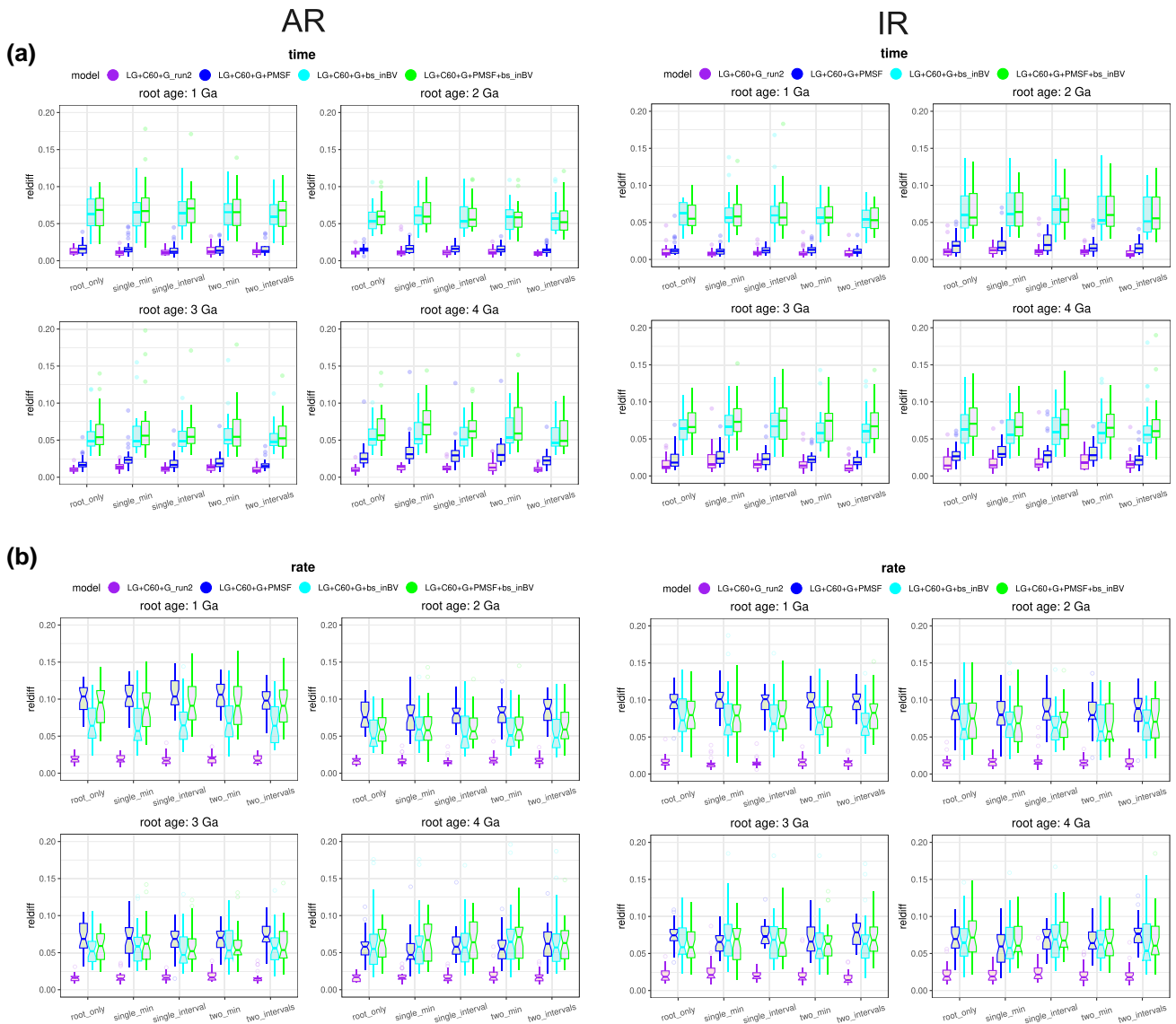


Figure 4. Evaluation of the approximation of PMSF and bootstrapped Hessian on molecular clock analysis by simulation. Phylogenies are simulated under the same strategy as the focal dataset. The sequence alignment is simulated under LG + C60 + G{1.0}. The bootstrapped Hessian matrix is calculated using the software *bs_inBV*. The y axis indicates the relative accuracy of the posterior mean of the divergence times a) or substitution rates b) compared to that obtained by the first MCMC run under LG + C60 + G (note that this is different from other figures where the baseline in comparison is the true simulated times and rates). Each plot contains 30 dots representing the mean relative difference based on MCMCtree molecular clock dating with the corresponding calibration strategy on 30 simulated datasets. For each simulation, 30 timetrees each containing 20 tips are simulated using a birth-death process with a mean substitution rate 0.25 substitutions/site/Gyr across four different root ages ranging from 1.0 to 4.0 Ga. LG + C60 + G_run2: the second independent MCMC run of MCMCtree under LG + C60 + G. LG + C60 + G + PMSF: PMSF approximation with LG + C60 + G. LG + C60 + G + *bs_inBV*: bootstrapped Hessian under LG + C60 + G. LG + C60 + G + PMSF + *bs_inBV*: bootstrapped Hessian under LG + C60 + G + PMSF. Calibration strategies are the same as those used in Fig. 1.

molecular clock dating under profile mixture models, we compared both divergence time and substitution rate estimates obtained under these approximations by evaluating their relative differences from the true simulation parameters. We found that both approximations provided accurate estimates of divergence times, no matter which root age was used in simulation (Fig. 4a). In terms of the time estimates, across all simulated root ages, compared to LG + C60 + G estimated rates, the average differences for AR were $3.6 \pm 1.0\%$ (PMSF approximation; LG + C60 + G + PMSF), $7.1 \pm 0.6\%$ (bootstrapped Hessian; LG + C60 + G +

bs_inBV), and $7.3 \pm 0.8\%$ (combined use; LG + C60 + G + PMSF + *bs_inBV*); for IR, the differences were $1.8 \pm 0.4\%$, $5.9 \pm 0.4\%$, and $6.5 \pm 0.5\%$, respectively. The deviations in time estimates from these approximations, particularly PMSF, were close enough to the stochastic variation observed across independent MCMC runs of the original profile mixture model (LG + C60 + G_run2 in Fig. 4; comparisons in Fig. 4 are made relative to the initial MCMC run under LG + C60 + G). The estimates of evolutionary rates exhibited slightly greater divergence from the original model with either approximation (Fig. 4b). The mean relative

differences under both AR and IR models were around 5% to 10% for either approximation, or their combination, compared to the true parameters used in simulation.

We further compared the log-likelihood calculated by phyloHessian using Hessian computed by the finite-difference method (see [Materials and Methods](#)) and approximated via bootstrap with `bs_inBV`. Roughly speaking, both methods approximated the true log-likelihood of the data reasonably well. phyloHessian achieved higher Pearson correlation coefficients with the exact likelihood than the bootstrap-based method (P -value = 0.00026, paired Wilcoxon test; [Figure S11](#)), on the data tested, while running ~ 10 times faster (assuming 1,000 bootstraps for `bs_inBV`). More details are given in [Discussion](#).

Divergence times and substitution rates under mixture models for the ancient symbionts microsporidia and Rickettsiales: an empirical study

To evaluate mixture models on empirical data, we selected two ancient symbiont lineages with distinct evolutionary significance and accelerated evolutionary rates: Microsporidia, highly reduced, fast-evolving obligate intracellular parasites of animals related to fungi ([Capella-Gutiérrez et al. 2012](#)), and Rickettsiales, an alphaproteobacterial lineage that shares a close evolutionary relationship with the ancestors of mitochondria ([Martijn et al. 2018](#); [Wang and Luo 2021](#)). Their rapid evolution combined with ancient divergence makes them ideal test cases for our analysis. Note that our primary goal was to assess how complex substitution models influence the molecular clock analysis itself, leading us to test different settings rather than selecting a single optimal configuration (e.g. partitions, clock model).

For both datasets, we partitioned the alignments into three subsets (see [Materials and Methods](#)) according to a Gaussian mixture model-based clustering according to the logarithm of the estimated substitution rates of each gene and selected the best-fitting substitution models for each partition under both site-homogeneous and mixture substitution models respectively. In all cases, profile mixture models were associated with the lowest AIC values, with LG+Cxx+F+I+R4 (where $xx = 40, 50, 60$) being the most frequently selected ([Data S3 to S4](#)). Microsporidia do not have fossil records. Thus, we applied different calibrations to only the root node based on previous estimate of the occurrence time of its LCA (last common ancestor) ([Note S2](#)). We found that the time estimates for Microsporidia lineages were largely consistent between the best-fitting profile mixture and homogenous models ([Fig. 5](#)). The LCA of Microsporidia was estimated to occur at approximately 0.41 Ga (95% HPD: 0.74 to 0.45 Ga) and 0.63 Ga (95% HPD: 0.73 to 0.52 Ga) under the AR and IR clock models, respectively ([Fig. 5a](#)). In contrast to the similar divergence time estimates between homogeneous and mixture models, the absolute substitution rate was at least 1/3 higher under the mixture model ([Fig. 5b](#)). All microsporidian lineages experienced accelerated sequence changes compared to the outgroup lineages. While the highest substitution rate was predicted for the branch leading to the LCA of Microsporidia

under the IR clock model, the AR clock model instead predicted the highest rate on the branch leading to *Enterocytozoon*, a major microsporidian pathogen in human and animals. The above pattern was consistent with alternative calibrations and different numbers of partitions ([Figure S12](#)).

Rickettsiales also lack direct fossil evidence; however, due to their close phylogenetic relationship with mitochondria, we employed the mitochondrial endosymbiosis dating strategy as used in our prior study ([Wang and Luo 2021](#)). This approach allows estimating Rickettsiales' evolutionary timeline by incorporating eukaryotic fossils, based on the evolutionary scenario that mitochondria evolved from a lineage closely related to Alphaproteobacteria (and thus Rickettsiales). Particularly, under the AR clock model, the time estimates of Rickettsiales lineages under the mixture substitution model differed by an average of >20% compared to those from the homogeneous alternatives ([Fig. 6a](#)). In general, the ages of most lineages were estimated to be older under the profile mixture model ([Fig. 6a](#)). For example, under the AR clock model, the LCA of Rickettsiales was estimated to occur at 2.67 Ga (95% HPD: 2.97 to 2.35 Ga) by the mixture model, whereas the homogeneous substitution model estimated it at 2.26 Ga (95% HPD: 2.57 to 1.94 Ga). Similarly, eukaryotes were estimated to originate at 2.29 Ga (95% HPD: 2.57 to 1.98 Ga) under the mixture model, compared to 1.74 Ga (95% HPD: 1.98 to 1.50 Ga) under the homogeneous model ([Fig. 6a](#)). The above pattern holds under alternative calibration settings ([Figure S13](#)). With a single partition ([Figure S13a and b](#)), the difference in time estimates between different substitution models was less obvious. In all calibration schemes, the substitution rates under the mixture models were roughly 1.5 to 2 times as fast as those estimated under homogeneous models ([Fig. 6b](#)).

A recent study ([Baños et al. 2024b](#)) suggests that the empirical frequency estimated from the alignment (+F) might mislead phylogenetic reconstruction for profile mixture models. As an additional analysis, we re-performed molecular clock dating with the best-fitting mixture model without the empirical frequency option. Similar patterns were found ([Figures S14 and S15](#)).

Discussion

Approximate likelihood dating with complex models

The present study introduces phyloHessian, enabling Hessian calculation under diverse mixture and non-mixture amino acid substitution models for molecular clock dating using MCMCtree's approximate likelihood method ([Thorne et al. 1998](#); [dos Reis and Yang 2011](#)). A key question is whether MCMCtree's approximate likelihood method performs equally well as exact likelihood under mixture substitution models in terms of divergence time estimation. While previous research suggests that this holds true for homogeneous models ([dos Reis and Yang 2011](#)), a full comparison is impossible due to the lack of software directly implementing complex substitution models. Therefore, we compared the phylogenetic likelihoods of bootstrap trees calculated by the approximate and exact

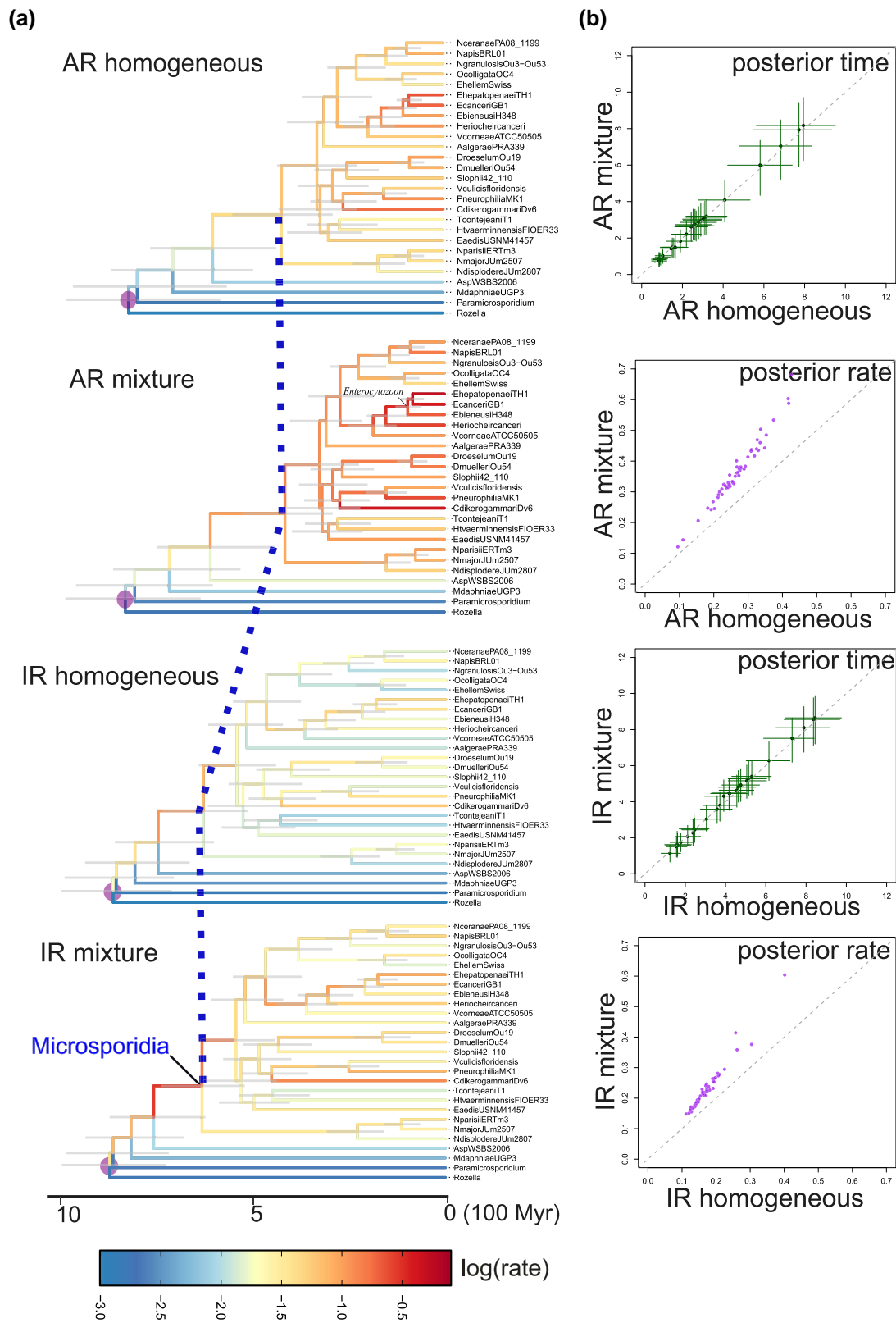


Figure 5 Evolutionary timeline and substitution rates estimated by MCMCtree under different substitution models for Microsporidia. a) A uniform calibration prior between 0.5 and 0.9 Ga was applied to the root indicated by a circle in purple (see Note S2), with soft bounds extending 2.5% beyond each time bound. Branch colors indicate the logarithm of the posterior mean substitution rate (estimated by MCMCtree), ranging from blue (low) to red (high). b) Posterior divergence times and branch-specific substitution rates estimated by the mixture substitution model (y axis) versus those estimated by the homogeneous substitution model (x axis) under the AR or IR clock models. The bars indicate the 95% HPD.

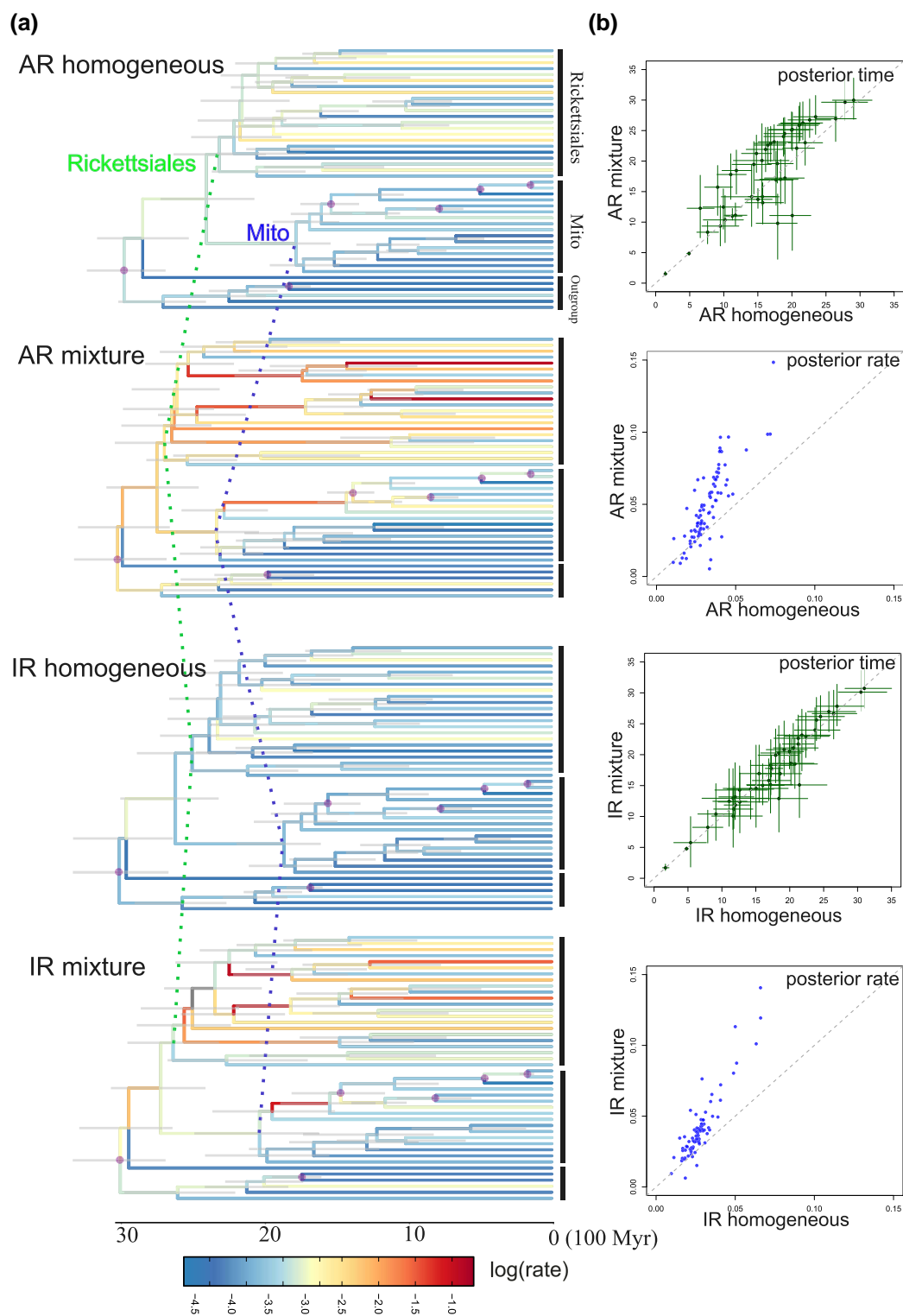


Figure 6 Evolutionary timeline and substitution rates estimated by MCMCtree under different substitution models for Rickettsiales. This is based on the mitochondrial endosymbiosis dating strategy (Wang and Luo 2021), using eukaryotic fossils to calibrate Rickettsiales evolution, based on the evolutionary scenario that mitochondria originated from a lineage closely related to α -Proteobacteria. a) The estimated evolutionary timeline of Rickettsiales under different substitution (the best-fitting homogeneous or mixture) and clock (IR or AR) models. All calibrated nodes are denoted by a circle. In addition to the calibration at the root, all four calibrations are placed within eukaryotes in the clade of mitochondria (see Note S2). Branch colors indicate the logarithm of the posterior mean substitution rate estimated by MCMCtree, ranging from blue (low) to red (high). b) Posterior divergence times and branch-specific substitution rates estimated by the mixture substitution model (y axis) versus those estimated by the homogeneous substitution model (x axis) under the AR or IR clock models. The bars indicate the 95% HPD.

likelihood methods under LG + C60 + G using phyloHessian. The approximate likelihood values obtained from phyloHessian (using finite differences) generally showed good agreement with those derived from both the exact method and bs_inBV, despite minor discrepancies (Figure S11). Further, phyloHessian sometimes outperformed bs_inBV, with bs_inBV showing markedly lower log-likelihood in occasional cases (Figure S11). Discrepancies in time and rate estimates between the Hessian-based method (phyloHessian) and the bootstrap approximation (bs_inBV) stem from three factors (Fig. 4): the inherent limitation of bs_inBV due to the finite number of bootstrap replicates; the statistical assumption in bs_inBV that the gradient is zero when approximating the Hessian via the covariance matrix, an assumption that is invalid when the MLEs of branch lengths are near zero (dos Reis and Yang 2011); and the difference in parameter handling, as MCMCtree fixes substitution model parameters (such as the parameter α for the gamma distribution for relative rates) at their MLEs while bs_inBV re-estimates all parameters in each bootstrap, resulting in phyloHessian using only the branch length Hessian versus bs_inBV approximating the full parameter Hessian. A more thorough examination in time estimates, instead of the phylogenetic likelihood, is necessary to fully address the question in future.

It was when we were finishing writing this manuscript that we became aware, through personal communication, of another computational software, IQ2MC (Demotte et al. 2025), implemented in the newest version of IQ-Tree (v3.0.1). IQ2MC leverages the highly optimized computation of IQ-Tree's phylogenetic library enabling remarkably efficient Hessian matrix estimation under not only amino acid, but nucleic acid and partitioned models, unlike phyloHessian which is restricted to amino acid models. Due to these computational advantages, IQ2MC was markedly faster than CODEML and phyloHessian (Data S5). However, our software, phyloHessian, offers more flexible settings in molecular dating by alternative Hessian calculation algorithms and interactions with other software like PhyML and bs_inBV. Its code for phylogenetic likelihood calculation under complex substitution models would be a valuable resource for phylogenetic software development within the Julia ecosystem. The application of IQ2MC suggests that mixture substitution models help prevent biased date estimates, particularly if sequences were highly diverged (Demotte et al. 2025). This is well consistent with the findings in the present study. Collectively, our studies underscore the importance of selecting appropriate substitution models in (deep-time) molecular dating analyses while providing researchers with different options for practical applications.

Other methods, such as McmcDate (Mahendrarajah et al. 2023), also accelerate molecular dating analyses by approximating the phylogenetic likelihood. It is in concept similar to bs_inBV (Wang and Luo 2025), but operates in a Bayesian framework. A key advantage of bs_inBV and McmcDate over phyloHessian and IQ2MC is their ability to accommodate substitution models not yet available in specific phylogenetics software, such as FunDi (Gaston et al. 2011) or CAT-PMSF (Szánthó et al. 2023). However, these approximations via bootstrap or MCMC sampling may sacrifice computational speed and accuracy to some extent (Fig. 4), as discussed above.

Deep-time molecular clock dating: the influence of complex mixture models

Our extensive simulations demonstrate that the advantage of using the complex mixture substitution model for divergence time estimation becomes increasingly significant as evolutionary timescales deepen (Figs. 1 to 4). Particularly, when the root age was earlier than 2.0 Ga, the difference became obvious. This highlights the importance of substitution models that better fit the data for dating deep-time phylogenies. On the other side, the substitution rate was almost always estimated higher and more accurate under mixture models than under homogeneous models. The underlying logic is straightforward. Mixture models typically estimate longer branch lengths than homogeneous models (Venditti et al. 2008; Moody et al. 2022; Wang and Luo 2025). Because branch length equals substitution rate multiplied by time (dos Reis et al. 2016; Guindon 2020), the underestimated branch lengths under simpler substitution models must be attributable to changes in either the rate, the time, or both. Given that molecular clock analyses typically constrain the time estimates through fossil-based calibrations, the increased branch lengths under mixture models primarily result in higher rate estimates, with a smaller impact on the time estimates themselves.

The case in empirical data analysis is more complex. Unlike simulations with known parameters, empirical data could arise from processes so complex that even the best-fitting models are inadequate. This notion is supported by the frequent selection of the most highly parameterized models as the best fit, implying that the complexity of the real process the sequence is generated far exceeds the capabilities of our current models. Consequently, they could make little difference in the time estimates compared to homogeneous models, because both might be an oversimplification of the true model. In such cases, it would be worth applying models like UDM (Schrempf et al. 2020) with more profile mixtures than C60, or the GTRpmix exchangeability matrix and its derived models (Baños et al. 2024a). While some of these models have been integrated into phyloHessian, a comprehensive evaluation of their applications in molecular dating analysis on empirical data remains a task for future research.

While the “mean relative difference” metric used to compare between models takes the absolute value, it is possible to tell a presumable bias in the time estimates from Figures S2 and S3: The posterior ages estimated by the homogeneous model seemed to overestimate the divergence times in scenarios mimicking deep-dating evolution in simulation. As such, it would be very important to have appropriate maximum time bounds on internal nodes (Wang and Luo 2021; Moody et al. 2022). If indeed in current molecular clock models there is a systematic bias in overestimating the divergence times, it might also explain the observation that mixture models displayed greater improvements in time estimation for alignment simulated under mixture models in the absence of maximum time bounds (*single_min*, *two_min*) compared to the case where internal nodes are fully calibrated (*single_interval*, *two_intervals*; Fig. 1a). In contrast to simulations, the Rickettsiales data reveals a contrasting pattern where mixture models led to older age estimates (Fig. 6). We suspect that one important factor lies in the calibration settings. In simulations, we have complete control over the calibration settings.

Real-world data introduce complexities absent in simulation because the available calibrations are not uniform in their quality. While some might accurately reflect the true age of the node, others could be overly young, overly old, or simply uninformative. This complex interplay between substitution models and the quality and number of time calibrations, along with other variables such as the clock model and sequence length, likely makes the impact of mixture substitution models on posterior age estimates case dependent.

In summary, while the present study indicates significant impacts of complex profile mixture substitution models on divergence time inferences for deep-time or fast-evolving lineages, the importance of these models on time estimates cannot be overstated; their effect likely requires consideration on a dataset-specific basis. Therefore, we recommend rigorous molecular clock analyses employing different substitution models to assess their impact on divergence time and substitution rate estimates in future research. Note also that while our simulations used an LG+C60+G model, the LG exchangeabilities themselves were not estimated in the presence of profile mixtures (Le and Gascuel 2008). Ideally, for practical applications, it would be more appropriate to use models that directly estimate exchangeabilities under a corresponding profile mixture model for such cases (Baños et al. 2024a).

Implications for the evolution of ancient symbionts

For empirical datasets of Microsporidia and Rickettsiales, profile mixture models, which display much higher model fit to data than their homogeneous counterparts, produced significantly higher substitution rates than estimated by homogeneous models (Figs. 5b and 6b). The difference suggests that bacterial pathogens/symbionts could experience relatively high evolutionary rates up to 1.0 or 2.0 Gyr. This hints that factors like elevated mutation rates, selection, or persistently reduced effective population sizes, which are often responsible for rapid sequence divergence, can have more profound and long-lasting effects on the evolutionary trajectory of symbiotic microorganisms than previously thought. Studying the substitution rates will also improve our understanding of microbial lifestyle evolution.

One of the merits of the mitochondria symbiosis-based dating method for Rickettsiales evolution is that it co-estimates the divergence times for both Rickettsiales and their eukaryotic hosts with a unified dataset (Wang and Luo 2021). Notably, under profile mixture models, the origin time of Rickettsiales was estimated at 2.5 to 2.0 Ga, roughly 0.5 Gyr earlier than that of the last eukaryotic common ancestor (LECA) using homogeneous models estimated in both the current (Fig. 6) and previous studies (Parfrey et al. 2011; Betts et al. 2018; Wang and Luo 2021; Moody et al. 2024). Hence, if the revised time estimate is accurate, it implies that the Rickettsiales' LCA existed before LECA emerged. This interpretation is consistent with recent comparative genomics findings suggesting that early Rickettsiales initially adopted a free-living or non-obligate symbiotic lifestyle (Castelli et al. 2024), with subsequent evolutionary events leading to independent

colonizations of eukaryotic hosts—first protists, then animals (Castelli et al. 2016; Wang and Luo 2021). Alternatively, it is possible that early Rickettsiales were associated with unknown host organisms belonging to eukaryotic stem groups (i.e. extinct or unknown lineages that diverged before LECA but after the split with the closest living relatives of eukaryotes). This hypothesis points to previously unrecognized interactions between stem-group eukaryotes and prokaryotes and hints at complex prokaryotic engagement as a foundational trait of eukaryotic ancestors, offering insights into the origin of eukaryotes.

Materials and methods

Mixture substitution model

In a mixture model, each category consists of a rate matrix M , an equilibrium frequency vector F , and a relative substitution rate R . Additionally, the computation of the likelihood necessitates the inclusion of branch lengths θ and the tree topology τ . Let the full parameter vector be denoted by $\Theta = (R, F, V, \theta, \tau)$. Denote by \mathcal{R} , \mathcal{F} , and \mathcal{V} the number of categories of exchangeability matrices, equilibrium frequencies, and relative substitution rates, respectively. Denote by S the number of alignment sites. Hence, the full likelihood of the phylogeny is given as

$$P(D|\Theta) = \prod_{s=1}^S \sum_{r=1}^{\mathcal{R}} \sum_{f=1}^{\mathcal{F}} \sum_{v=1}^{\mathcal{V}} P(D_s|R_r, F_f, V_v, \theta, \tau) P(R_r, F_f, V_v), \quad (1)$$

where $P(R_r, F_f, V_v)$ indicate the mixture weights (Guindon 2019). Model constraints are enforced by restricting the set of allowed combinations (r, f, v) , with the mixture weights normalized over this restricted set. Under independence, this gives $P(R_r, F_f, V_v) = P(R_r)P(F_f)P(V_v)$, whereas for linked models such as LG4M or LG4X, combinations with $f \neq v$ are excluded.

Numerical calculation of the Hessian matrix

Two primary numerical methods are commonly employed for computing the Hessian matrix. Let $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_n)$ indicate the vector of the MLEs of branch lengths and $\ell(\hat{\theta})$ be the log-likelihood of the phylogeny. The outer product of scores (OPS) method (Seo et al. 2004; dos Reis and Yang 2011) estimates the Hessian matrix indirectly using the first derivatives (scores) of the log-likelihood to sum the contributions of the log-likelihood at each site:

$$H(\hat{\theta}) \approx - \sum_{s=1}^S \left(\frac{\partial}{\partial \theta} \ell_s(\hat{\theta}) \frac{\partial}{\partial \theta^T} \ell_s(\hat{\theta}) \right). \quad (2)$$

Each component of the score vector $\frac{\partial}{\partial \theta} \ell_s(\hat{\theta})$ is numerically calculated by the central finite difference method as

$$\frac{\partial \ell_s}{\partial \theta_i} \approx \frac{\ell_s(\hat{\theta} + h e_i) - \ell_s(\hat{\theta} - h e_i)}{2h} + \mathcal{O}(h^2),$$

where e_i is the i th standard basis vector and h is a small step size, typically the cube or square root of machine precision.

The second way is by a direct second-order derivative method. It calculates $H(\hat{\theta})$ via finite differences of the log-likelihood function by

$$H(\hat{\theta}) = \sum_{s=1}^S \frac{\partial^2}{\partial \theta \partial \theta^T} \ell_s(\hat{\theta}), \quad (3)$$

where diagonal elements and off-diagonal elements are respectively calculated by the central finite difference method as

$$H_{ii}(\hat{\theta}) \approx \frac{\ell(\hat{\theta} + he_i) - 2\ell(\hat{\theta}) + \ell(\hat{\theta} - he_i)}{h^2},$$

$$H_{ij}(\hat{\theta}) \approx$$

$$\frac{\ell(\hat{\theta} + he_i + he_j) - \ell(\hat{\theta} + he_i - he_j) - \ell(\hat{\theta} - he_i + he_j) + \ell(\hat{\theta} - he_i - he_j)}{4h^2}.$$

Under regularity conditions, $H(\hat{\theta})/S$ converges to the expected Fisher information for a single site $I(\theta) = -E\left[\frac{\partial^2}{\partial \theta \partial \theta^T} \ell_s(\theta)\right]$ regardless of whether it is calculated using Eq. (2) or (3) (Pawitan 2001):

$$\frac{1}{S} \sum_{s=1}^S \frac{\partial^2}{\partial \theta \partial \theta^T} \ell_s(\hat{\theta}) \simeq -\frac{1}{S} \sum_{s=1}^S \left(\frac{\partial}{\partial \theta} \ell_s(\hat{\theta}) \frac{\partial}{\partial \theta^T} \ell_s(\hat{\theta}) \right). \quad (4)$$

In both methods described as above, the error term is of order $\mathcal{O}(h^2)$. To further improve the numerical accuracy of the second-order derivative method, particularly when dealing with very short branches, we derive the central finite difference calculation of the Hessian at $\mathcal{O}(h^4)$. For the OPS method, this is simply

$$\frac{\partial \ell}{\partial \theta_i} \approx \frac{-\ell(\hat{\theta} + 2he_i) + 8\ell(\hat{\theta} + he_i) - 8\ell(\hat{\theta} - he_i) + \ell(\hat{\theta} - 2he_i)}{12h} + \mathcal{O}(h^4) \quad (5)$$

followed by summing the contributions of the log-likelihood at each site according to Eq. (2). Regarding the direct second-order derivative method, the result is more complex and is detailed in Note S3.3.

In phyloHessian, the default algorithm is the OPS method, which is employed by CODEML in PAML v4.3+. Alternatively, users can specify the second-order derivative method, as used in PAML 4.1 and earlier versions. Users can also specify the order of the error term, allowing balancing computational accuracy with speed based on their needs. If not otherwise indicated, the computation of all Hessian matrices in the present study was performed using the OPS algorithm.

Simulation procedure

We evaluated the performance of different substitution models using simulated data that mimics deep-time evolution (Figure S16). In our focal simulation scheme, for each substitution model used in sequence evolution simulation (details provided below), we generated 30 timetrees each with 20 tips, using the R package TreeSim (Stadler 2011). Following our previous work (Wang and Luo 2025), timetrees were created under a birth-death model with birth and death rates of 0.4 and 0.2 lineages per 0.1 Gyr as estimated in a prior study of prokaryotic lineage diversification (Scholl and Wiens 2016) and with a taxon sampling proportion of 10% (see Note S1 for alternative settings). Simulated trees were generated with true root ages of 1.0, 2.0, 3.0, and 4.0 Ga to represent a range of evolutionary timescales approaching the age of Earth (~4.5 Ga).

For each timetree, branch-specific substitution rates were sampled under both the independent rates (IR) and autocorrelated rates (AR) model. For the IR mode, following our previous study (Wang and Luo 2021), the rate follows lognormal distribution with parameters $\mu_{\text{IR}} = -3.7$ and $\sigma_{\text{IR}} = 0.2$, such that the mean of rate is $e^{\mu_{\text{IR}} + \frac{\sigma_{\text{IR}}^2}{2}} \approx 0.025$ substitutions/site/0.1 Gyr, with a standard deviation of $\sqrt{(e^{\sigma_{\text{IR}}^2} - 1)e^{2(\mu_{\text{IR}} + \frac{\sigma_{\text{IR}}^2}{2})}} \approx 0.005$ substitutions/site/0.1 Gyr. This roughly corresponds to empirical observations that tip-to-root distances in the bacterial tree of life, constructed with universally conserved orthologs, are roughly 1.0 amino acid substitutions/site (Moody et al. 2022; Wang and Luo 2025), assuming an origin of early life at around 4.0 Ga (Battistuzzi and Hedges 2009; Marin et al. 2016; Betts et al. 2018; Moody et al. 2024). The methodology used to determine the simulation parameters for the AR model is described in detail in the next section.

Alignments of 300 amino acids (aa) were simulated using AliSim (Ly-Trong et al. 2022) according to the timetree generated by simulation, under various substitution models. In the focal analysis, we employed the LG + C60 + G4{1.0} model. This indicates LG substitution model plus C60 profile mixture models (60 empirically derived amino acid site frequency profiles) (Quang et al. 2008) and across-site rate variation under a discrete gamma distribution with four relative rate categories parameterized by $\alpha = 1.0$, hence Gamma(1,1), meaning that the mean and variance of the across-site relative rates are $\frac{1}{\alpha} = 1$ and $\frac{1}{\alpha^2} = 1$, respectively (Yang 1994). Further, we varied several simulation parameters, including sequence length, number of taxa, calibration strategies, birth-death process parameters, across-site rate heterogeneity, molecular clock settings, and model misspecifications. These are detailed in Note S1.

Determining the parameters for simulation under the AR clock rate model

Denote the absolute rate of sequence evolution (measured in substitutions per site per time unit) under IR and AR respectively by the vectors $X_{\text{IR}} = (X_{1,\text{IR}}, \dots, X_{n,\text{IR}})$ and $X_{\text{AR}} = (X_{1,\text{AR}}, \dots, X_{n,\text{AR}})$ where n represents the number of branches. Their log-transformed values are given by $Y_{\text{IR}} = (Y_{1,\text{IR}}, \dots, Y_{n,\text{IR}})$ and $Y_{\text{AR}} = (Y_{1,\text{AR}}, \dots, Y_{n,\text{AR}})$. To make simulations under IR and AR models more comparable, we require that the expectations of the sample mean and sample variance of the log-transformed rate under the two models to be equal, respectively, given the same phylogeny. Mathematically, this requires that

$$E(\bar{Y}_{\text{AR}}) = E(\bar{Y}_{\text{IR}}), \quad E(s_{\text{AR}}^2) = E(s_{\text{IR}}^2) \quad (6)$$

where the sample variances of the log-transformed rates are denoted by s_{AR}^2 and s_{IR}^2 .

For the IR model, it is assumed that the rate for each branch, denoted $X_{\text{IR}} = (X_{1,\text{IR}}, \dots, X_{n,\text{IR}})$, is independently and identically (*i.i.d.*) distributed according to a lognormal distribution with parameters μ_{IR} and σ_{IR}^2 , such that the rate for each branch can be expressed as $X_{i,\text{IR}} \sim \text{Lognormal}(\mu_{\text{IR}}, \sigma_{\text{IR}}^2)$. In other words, we have

$$E(\bar{Y}_{\text{IR}}) = \mu_{\text{IR}}, \quad E(s_{\text{IR}}^2) = \sigma_{\text{IR}}^2. \quad (7)$$

The autocorrelated rates (AR) model in molecular clock dating is a special case of the geometric Brownian motion (GBM) process, where lineage-specific substitution rates evolve according to

GBM with the drift parameter equal to zero (Rannala and Yang 2007). It is thus parameterized by two parameters, μ_{AR} , the log-transformed rate at the root, and σ_{AR}^2 , which determines the volatility of the rate changes over time. Accordingly, it can be shown that the log-transformed rate under AR model after time t follows $\mathcal{N}(\mu_{AR} - \frac{1}{2}t\sigma_{AR}^2, t\sigma_{AR}^2)$ (Rannala and Yang 2007; Panchaksaram et al. 2025). Let $C = (C_{ij})_{n \times n}$ denote the branch-wise covariance matrix, where n is the number of branches in a rooted tree (note that this is different from what is so-called phylogenetic covariance matrix which is related to only the tips [Figure S17]). Further, the log-transformed rates Y_{AR} follow the following multivariate distribution:

$$Y_{AR} \sim \mathcal{N}\left(\left(\mu_{AR} - \frac{1}{2}\sigma_{AR}^2 c_{11}, \dots, \mu_{AR} - \frac{1}{2}\sigma_{AR}^2 c_{nn}\right), \sigma_{AR}^2 C\right).$$

Thus, it is easy to see that

$$E(Y_{i,AR}) = \mu_{AR} - \frac{1}{2}\sigma_{AR}^2 c_{ii}. \tag{8}$$

The expectation of the sample variance s_{AR}^2 of log-transformed rate Y_{AR} across all n branches is calculated as

$$E(s_{AR}^2) = \frac{1}{n-1} E\left\{ \sum_i \left(Y_{i,AR} - \frac{1}{n} \sum_j Y_{j,AR} \right)^2 \right\}. \tag{9}$$

$$\begin{aligned} & \sum_{i=1}^n E(Y_{i,AR}^2) - \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n E(Y_{i,AR} Y_{j,AR}) \\ &= \sum_i \left[c_{ii}\sigma_{AR}^2 + \left(\mu_{AR} - \frac{1}{2}c_{ii}\sigma_{AR}^2 \right)^2 \right] - \frac{1}{n} \sum_{i,j} \left[c_{ij}\sigma_{AR}^2 + \left(\mu_{AR} - \frac{1}{2}c_{ii}\sigma_{AR}^2 \right) \left(\mu_{AR} - \frac{1}{2}c_{jj}\sigma_{AR}^2 \right) \right] \\ &= \sigma_{AR}^2 \underbrace{\left[\sum_i c_{ii} - \frac{1}{n} \sum_{i,j} c_{ij} \right]}_{\text{Term 1}} \\ &+ \underbrace{\left[\sum_i \left(\mu_{AR} - \frac{1}{2}c_{ii}\sigma_{AR}^2 \right)^2 - \frac{1}{n} \sum_{i,j} \left(\mu_{AR} - \frac{1}{2}c_{ii}\sigma_{AR}^2 \right) \left(\mu_{AR} - \frac{1}{2}c_{jj}\sigma_{AR}^2 \right) \right]}_{\text{Term 2}}. \end{aligned}$$

Substituting $\sum_i c_{ii} = \text{tr}(C)$ and $\sum_{i,j} c_{ij} = \frac{1}{n} \mathbf{1}' C \mathbf{1}$ into Term 1, we can rewrite it as

$$\sigma_{AR}^2 \left[\sum_i c_{ii} - \frac{1}{n} \sum_{i,j} c_{ij} \right] = \sigma_{AR}^2 \left[\text{tr}(C) - \frac{1}{n} \mathbf{1}' C \mathbf{1} \right].$$

For Term 2, noting that terms involving μ_{AR} cancel out, we have

$$\begin{aligned} & \sum_i \left(\mu_{AR} - \frac{1}{2}c_{ii}\sigma_{AR}^2 \right)^2 - \frac{1}{n} \sum_{i,j} \left(\mu_{AR} - \frac{1}{2}c_{ii}\sigma_{AR}^2 \right) \left(\mu_{AR} - \frac{1}{2}c_{jj}\sigma_{AR}^2 \right) \\ &= \left(n\mu_{AR}^2 - \mu_{AR}\sigma_{AR}^2 \text{tr}(C) + \frac{1}{4} \sum_i c_{ii}^2 \sigma_{AR}^4 \right) \\ &+ \left(-n\mu_{AR}^2 + \mu_{AR}\sigma_{AR}^2 \text{tr}(C) - \frac{1}{4n} \sum_{i,j} c_{ii} c_{jj} \sigma_{AR}^4 \right) \\ &= \frac{\sigma_{AR}^4}{4} \left[\sum_i c_{ii}^2 - \frac{1}{n} (\text{tr}(C))^2 \right]. \end{aligned}$$

Expanding the right-hand side, we obtain

$$\begin{aligned} & E\left[\sum_{i=1}^n \left(Y_{i,AR} - \frac{1}{n} \sum_{j=1}^n Y_{j,AR} \right)^2 \right] \\ &= \sum_{i=1}^n E(Y_{i,AR}^2) \\ &- \frac{1}{n} \sum_{i=1}^n \left(\sum_{j=1}^n 2E(Y_{i,AR} Y_{j,AR}) - \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^n E(Y_{j,AR} Y_{k,AR}) \right) \\ &= \sum_{i=1}^n E(Y_{i,AR}^2) - \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n E(Y_{i,AR} Y_{j,AR}). \end{aligned} \tag{10}$$

Further, note the following relationship:

$$E(Y_{i,AR}^2) = \text{Var}(Y_{i,AR}) + [E(Y_{i,AR})]^2 = c_{ii}\sigma_{AR}^2 + \left(\mu_{AR} - \frac{1}{2}\sigma_{AR}^2 c_{ii} \right)^2,$$

$$\begin{aligned} E(Y_{i,AR} Y_{j,AR}) &= \text{Cov}(Y_{i,AR}, Y_{j,AR}) + E(Y_{i,AR})E(Y_{j,AR}) \\ &= \sigma_{AR}^2 c_{ij} + \left(\mu_{AR} - \frac{1}{2}\sigma_{AR}^2 c_{ii} \right) \left(\mu_{AR} - \frac{1}{2}\sigma_{AR}^2 c_{jj} \right). \end{aligned}$$

Introducing the above into Eq. (10), we have

Hence,

$$E(s_{AR}^2) = \frac{\sigma_{AR}^2}{n-1} \left[\left(\text{tr}(C) - \frac{1}{n} \mathbf{1}' C \mathbf{1} \right) + \frac{\sigma_{AR}^2}{4} \left(\sum_i c_{ii}^2 - \frac{1}{n} (\text{tr}(C))^2 \right) \right].$$

Because $E(s_{IR}^2) = \sigma_{IR}^2$, equating $E(s_{IR}^2)$ with $E(s_{AR}^2)$ and rearranging terms gives the quadratic equation for σ_{AR}^2

$$\frac{1}{4} \left[\sum_i c_{ii}^2 - \frac{1}{n} (\text{tr}(C))^2 \right] \sigma_{AR}^4 + \left[\text{tr}(C) - \frac{1}{n} \mathbf{1}' C \mathbf{1} \right] \sigma_{AR}^2 - (n-1)\sigma_{IR}^2 = 0. \tag{11}$$

Discarding the negative root, along with solving Eq. (8), we ultimately get

$$\hat{\sigma}_{AR}^2 = \frac{2(n-1)\sigma_{IR}^2}{\sqrt{\sigma^2 + 4b(n-1)\sigma_{IR}^2} + a}, \quad \hat{\mu}_{AR} = \mu_{IR} + \frac{\hat{\sigma}_{AR}^2 \text{tr}(C)}{2n} \tag{12}$$

where $a = \text{tr}(C) - \frac{1}{n} \mathbf{1}'C\mathbf{1}$, $b = \frac{1}{4} [\sum_i c_{ii}^2 - \frac{1}{n} (\text{tr}(C))^2]$, and $\hat{\sigma}_{\text{AR}}^2$ and $\hat{\mu}_{\text{AR}}$ are the values of the parameters for the AR model that satisfy Eq. (6) given the parameters for the IR model (σ_{IR}^2 and μ_{IR}).

Calibrations used in molecular dating analysis on simulated datasets

Molecular dating analysis was conducted using MCMCtree implemented in PAML v4.10.7 (Yang 2007) (see also Note S3) on phylogenies simulated with procedures described above under varying calibration strategies: a root-only calibration, or root calibration plus one internal calibration, chosen as the node at the median age of all nodes in the true timetree, or two internal calibrations which targeted the 1/3 and 2/3 age quantiles. Time priors for calibrated nodes were set to be uniform within the interval $[\text{true_age} - (\text{true_age}/5), \text{true_age} + (\text{true_age}/5)]$, with soft bounds (MCMCtree's default: 2.5% probability of age outside bounds). Other information about MCMCtree analysis can be found in Note S3. The five calibration schemes mainly used in the present study are *root_only*: a single bounded calibration (both upper and lower bounds); *single_min*: bounded calibration at the root and one minimum-bounded internal node; *single_interval*: bounded calibration at both the root and an internal node; *two_min*: bounded calibration at the root and two minimum-bounded internal nodes; and *two_intervals*: bounded calibration at both the root and two internal nodes.

Comparing the performance of different substitution models on simulated datasets

We evaluated the performance of various substitution models mainly using the relative difference (reldiff) that compared the posterior mean divergence times and substitution rates estimated by MCMCtree (using each model) to the true values used in simulations. Relative difference: calculated as $\text{mean}\left(\frac{|b_i^m - b_i^{\text{true}}|}{\max(b_i^m, b_i^{\text{true}})}\right)$, where b_i^m and b_i^{true} denote the posterior mean divergence times or branch-specific substitution rates inferred by MCMCtree, and those used in simulation (true values), respectively.

Empirical data analysis

Genomes of Microsporidia and Rickettsiales were downloaded from MicrosporidiaDB (Aurrecochea et al. 2011) and GenBank (last accessed July 2024), respectively. Sequences of five additional Proteobacteria as the outgroup were retrieved from our prior study (Wang and Luo 2025). TreeCluster v1.0.4 (Balaban et al. 2019) was used to select representative taxa with a cutoff of phylogenetic depth (which takes phylogenetic distance and topology into account to cluster taxa into subclades) of 0.5 and 2.5 respectively for Microsporidia (yielding 27 members) and Rickettsiales (22 members). Genes used to reconstruct phylogeny and molecular dating for Microsporidia were based on 27 orthologs identified by OrthoFinder v3.0.1 (Emms and Kelly 2019) present in all members allowing ≤ 2 copies per genome, while for Rickettsiales we

adopted the mitochondrial endosymbiosis-based strategy (Wang and Luo 2021) by using 24 mitochondria-encoded genes conserved across Alphaproteobacteria (Wang and Wu 2015) (see Data availability). For both datasets, we partitioned the alignments into one, two, or three subsets by fitting the logarithm of the average substitution rates estimated by CODEML into a Gaussian mixture model. We selected the best-fitting substitution models for each partition under both site-homogeneous and site-heterogeneous models according to AIC (Note S3). Calibration information along with alternatives is available in Data S6.

Acknowledgments

We particularly thank Minh Bui and Mario dos Reis for comments on early versions of the manuscript and for sharing with us their preprint (Demotte et al. 2025). We also thank Xiyun Jiao, Jianhao Lv, Gergely Szöllösi, Ziheng Yang, and Tianqi Zhu for the discussion, Charley McCarthey for testing an early version of the software, Hui Li for proofreading, and Margaret Ip and Haiwei Luo for their support.

Supplementary material

Supplementary material is available at *Molecular Biology and Evolution* online.

Funding

The work is supported by the Natural Science Foundation of China (32400493, 42293294), Hong Kong Research Grants Council (RGC) General Research Fund (GRF) (14112024), and CUHK Direct Grant (4054912).

Conflicts of interest

None declared.

Data availability

phyloHessian is available at <https://github.com/evolbeginner/phyloHessianWrapper>. Other data and codes (Goto et al. 2010) are made available at <https://doi.org/10.6084/m9.figshare.29589881>.

References

- Aurrecochea C et al. AmoebaDB and MicrosporidiaDB: functional genomic resources for amoebozoa and microsporidia species. *Nucleic Acids Res.* 2011;39:D612–D619. <https://doi.org/10.1093/nar/gkq1006>.
- Balaban M, Moshiri N, Mai U, Jia X, Mirarab S. TreeCluster: clustering biological sequences using phylogenetic trees. *PLoS One.* 2019;14:e0221068. <https://doi.org/10.1371/journal.pone.0221068>.
- Baños H et al. GTRpmix: a linked general time-reversible model for profile mixture models. *Mol Biol Evol.* 2024a;41:msae174. <https://doi.org/10.1093/molbev/msae174>.

- Baños H, Susko E, Roger AJ. Is over-parameterization a problem for profile mixture models? *Syst Biol.* 2024b;73:53–75. <https://doi.org/10.1093/sysbio/syad063>.
- Battistuzzi FU, Hedges SB. A major clade of prokaryotes with ancient adaptations to life on land. *Mol Biol Evol.* 2009;26:335–343. <https://doi.org/10.1093/molbev/msn247>.
- Betts HC *et al.* Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. *Nat Ecol Evol.* 2018;2:1556–1562. <https://doi.org/10.1038/s41559-018-0644-x>.
- Bujaki T, Rodrigue N. Bayesian cross-validation comparison of amino acid replacement models: contrasting profile mixtures, pairwise exchangeabilities, and gamma-distributed rates-across-sites. *J Mol Evol.* 2022;90:468–475. <https://doi.org/10.1007/s00239-022-10076-y>.
- Capella-Gutiérrez S, Marcet-Houben M, Gabaldón T. Phylogenomics supports microsporidia as the earliest diverging clade of sequenced fungi. *BMC Biol.* 2012;10:47. <https://doi.org/10.1186/1741-7007-10-47>.
- Castelli M *et al.* Host association and intracellularity evolved multiple times independently in the Rickettsiales. *Nat Commun.* 2024;15:1093. <https://doi.org/10.1038/s41467-024-45351-7>.
- Castelli M, Sasser D, Petroni G. Biodiversity of “non-model” Rickettsiales and their association with aquatic organisms. In: Thomas S, editor. *Rickettsiales: biology, molecular biology, epidemiology, and vaccine development*. Springer International Publishing AG; 2016. p. 59–91.
- Demotte P *et al.* IQ2MC: A New Framework to Infer Phylogenetic Time Trees Using IQ-TREE 3 and MCMCTree with Mixture Models. 2025. Available from: <https://ecoevorxiv.org/repository/view/9125/>
- dos Reis M, Donoghue PCJ, Yang Z. Bayesian molecular clock dating of species divergences in the genomics era. *Nat Rev Genet.* 2016;17:71–80. <https://doi.org/10.1038/nrg.2015.8>.
- dos Reis M, Yang Z. Approximate likelihood calculation on a phylogeny for Bayesian estimation of divergence times. *Mol Biol Evol.* 2011;28:2161–2172. <https://doi.org/10.1093/molbev/msr045>.
- Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* 2006;4:699–710. <https://doi.org/10.1371/journal.pbio.0040088>.
- Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 2019;20:238. <https://doi.org/10.1186/s13059-019-1832-y>.
- Gaston D, Susko E, Roger AJ. A phylogenetic mixture model for the identification of functionally divergent protein residues. *Bioinformatics.* 2011;27:2655–2663. <https://doi.org/10.1093/bioinformatics/btr470>.
- Goto N *et al.* BioRuby: bioinformatics software for the Ruby programming language. *Bioinformatics.* 2010;26:2617–2619. <https://doi.org/10.1093/bioinformatics/btq475>.
- Guindon S. PhyML Manual. 2019. Available from: <https://gensoft.pasteur.fr/docs/phyml/3.3.20190909/phyml-manual.pdf>
- Guindon S *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 2010;59:307–321. <https://doi.org/10.1093/sysbio/syq010>.
- Guindon S. Rates and rocks: strengths and weaknesses of molecular dating methods. *Front Genet.* 2020;11:526. <https://doi.org/10.3389/fgene.2020.00526>.
- Ho SYW, Duchêne S. Molecular-clock methods for estimating evolutionary rates and timescales. *Mol Ecol.* 2014;23:5947–5965. <https://doi.org/10.1111/mec.12953>.
- Lartillot N, Brinkmann H, Philippe H. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol Biol.* 2007;7:S4. <https://doi.org/10.1186/1471-2148-7-S1-S4>.
- Lartillot N, Philippe H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol.* 2004;21:1095–1109. <https://doi.org/10.1093/molbev/msh112>.
- Le SQ, Dang CC, Gascuel O. Modeling protein evolution with several amino acid replacement matrices depending on site rates. *Mol Biol Evol.* 2012;29:2921–2936. <https://doi.org/10.1093/molbev/mss112>.
- Le SQ, Gascuel O. An improved general amino acid replacement matrix. *Mol Biol Evol.* 2008;25:1307–1320. <https://doi.org/10.1093/molbev/msn067>.
- Le SQ, Gascuel O. Accounting for solvent accessibility and secondary structure in protein phylogenetics is clearly beneficial. *Syst Biol.* 2010;59:277–287. <https://doi.org/10.1093/sysbio/syq002>.
- Le SQ, Lartillot N, Gascuel O. Phylogenetic mixture models for proteins. *Philos. Trans. R. Soc. B Biol. Sci.* 2008;363:3965–3976. <https://doi.org/10.1098/rstb.2008.0180>.
- Ly-Trong N, Naser-Khdour S, Lanfear R, Minh BQ, AliSim: a fast and versatile phylogenetic sequence simulator for the genomic era. *Mol Biol Evol.* 2022;39:masac092. <https://doi.org/10.1093/molbev/msac092>.
- Mahendrarajah TA *et al.* ATP synthase evolution on a cross-braced dated tree of life. *Nat Commun.* 2023;14:7456. <https://doi.org/10.1038/s41467-023-42924-w>.
- Marin J, Battistuzzi FU, Brown AC, Hedges SB. The Timetree of prokaryotes: new insights into their evolution and speciation. *Mol Biol Evol.* 2016;34:msw245–msw446. <https://doi.org/10.1093/molbev/msw245>.
- Martijn J, Vosseberg J, Guy L, Offre P, Ettema TJG. Deep mitochondrial origin outside the sampled Alphaproteobacteria. *Nature.* 2018;557:101–105. <https://doi.org/10.1038/s41586-018-0059-5>.
- Moody ERR *et al.* An estimate of the deepest branches of the tree of life from ancient vertically evolving genes. *eLife.* 2022;11:e66695. <https://doi.org/10.7554/eLife.66695>.
- Moody ERR *et al.* The nature of the last universal common ancestor and its impact on the early earth system. *Nat Ecol Evol.* 2024;8:1654–1666. <https://doi.org/10.1038/s41559-024-02461-1>.
- Pagel M, Meade A. A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. *Syst Biol.* 2004;53:571–581. <https://doi.org/10.1080/10635150490468675>.
- Panchaksaram M, Freitas L, Dos Reis M. Bayesian selection of relaxed-clock models: distinguishing between independent and autocorrelated rates. *Syst Biol.* 2025;74:323–334. <https://doi.org/10.1093/sysbio/syae066>.
- Parfrey LW, Lahr DJG, Knoll AH, Katz LA. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proc Natl Acad Sci U S A.* 2011;108:13624–13629. <https://doi.org/10.1073/pnas.1110633108>.
- Pawitan Y. *In all likelihood: statistical modelling and inference using likelihood*. Oxford University Press; 2001.

- Quang LS, Gascuel O, Lartillot N. Empirical profile mixture models for phylogenetic reconstruction. *Bioinformatics*. 2008;24:2317–2323. <https://doi.org/10.1093/bioinformatics/btn445>.
- Rannala B, Yang Z. Inferring speciation times under an episodic molecular clock. *Syst Biol*. 2007;56:453–466. <https://doi.org/10.1080/10635150701420643>.
- Scholl JP, Wiens JJ. Diversification rates and species richness across the Tree of Life. *Proc. R. Soc. B Biol. Sci.* 2016;283:20161335. <https://doi.org/10.1098/rspb.2016.1334>.
- Schrempf D, Lartillot N, Szöllösi G. Scalable empirical mixture models that account for across-site compositional heterogeneity. *Mol Biol Evol*. 2020;37:3616–3631. <https://doi.org/10.1093/molbev/msaa145>.
- Seo TK, Kishino H, Thorne JL. Estimating absolute rates of synonymous and nonsynonymous nucleotide substitution in order to characterize natural selection and date species divergences. *Mol Biol Evol*. 2004;21:1201–1213. <https://doi.org/10.1093/molbev/msh088>.
- Stadler T. Simulating trees with a fixed number of extant species. *Syst Biol*. 2011;60:676–684. <https://doi.org/10.1093/sysbio/syr029>.
- Szánthó LL, Lartillot N, Szöllösi GJ, Schrempf D. Compositionally constrained sites drive long-branch attraction. *Syst Biol*. 2023;72:767–780. <https://doi.org/10.1093/sysbio/syad013>.
- Thorne JL, Kishino H, Painter IS. Estimating the rate of evolution of the rate of molecular evolution. *Mol Biol Evol*. 1998;15:1647–1657. <https://doi.org/10.1093/oxfordjournals.molbev.a025892>.
- Venditti C, Meade A, Pagel M. Phylogenetic mixture models can reduce node-density artifacts. *Syst Biol*. 2008;57:286–293. <https://doi.org/10.1080/10635150802044045>.
- Wang HC, Li K, Susko E, Roger AJ. A class frequency mixture model that adjusts for site-specific amino acid frequencies and improves inference of protein phylogeny. *BMC Evol Biol*. 2008;8:331. <https://doi.org/10.1186/1471-2148-8-331>.
- Wang HC, Minh BQ, Susko E, Roger AJ. Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. *Syst Biol*. 2018;67:216–235. <https://doi.org/10.1093/sysbio/syx068>.
- Wang HC, Susko E, Roger AJ. The relative importance of modeling site pattern heterogeneity versus partition-wise heterotachy in phylogenomic inference. *Syst Biol*. 2019;68:1003–1019. <https://doi.org/10.1093/sysbio/syz021>.
- Wang S, Luo H. Dating Alphaproteobacteria evolution with eukaryotic fossils. *Nat Commun*. 2021;12:3324. <https://doi.org/10.1038/s41467-021-23645-4>.
- Wang S, Luo H. Dating the bacterial tree of life based on ancient symbiosis. *Syst Biol*. 2025;74:639–655. <https://doi.org/10.1093/sysbio/syae071>.
- Wang Z, Wu M. An integrated phylogenomic approach toward pinpointing the origin of mitochondria. *Sci Rep*. 2015;5:7949. <https://doi.org/10.1038/srep07949>.
- Williams TA, Cox CJ, Foster PG, Szöllösi GJ, Embley TM. Phylogenomics provides robust support for a two-domains tree of life. *Nat Ecol Evol*. 2020;4:138–147. <https://doi.org/10.1038/s41559-019-1040-x>.
- Wong TKF et al. IQ-TREE 3: Phylogenomic Inference Software using Complex Evolutionary Models. 2025. Available from: <https://ecoevorxiv.org/repository/view/8916/>
- Yang Z. Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Mol Biol Evol*. 1993;10:1396–1401. <https://doi.org/10.1093/oxfordjournals.molbev.a040082>.
- Yang Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J Mol Evol*. 1994;39:306–314. <https://doi.org/10.1007/BF00160154>.
- Yang Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol Biol Evol*. 2007;24:1586–1591. <https://doi.org/10.1093/molbev/msm088>.
- Yang Z, Rannala B. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol Biol Evol*. 2006;23:212–226. <https://doi.org/10.1093/molbev/msj024>.
- Zuckermandl E, Pauling L. Evolving genes and proteins. In: Bryson V, Vogel H, editors. *Evolutionary divergence and convergence in proteins*. Academic Press; 1965.