Paleobiology

www.cambridge.org/pab

An implementation of the Brownian motion model for Bayesian phylogenetic inference using continuous traits with missing states

Ziye Wang[1] and Chi Zhang[2]

[1]School of Life Sciences, Peking University, Beijing 100871, China
[2]Institute of Vertebrate Paleontology and Paleoanthropology Chinese Academy of Sciences, Beijing 100044, China

Methodological Advances

Cite this article: Wang, Z., and C. Zhang (2026). An implementation of the Brownian motion model for Bayesian phylogenetic inference using continuous traits with missing states. *Paleobiology* **52**, 163–171. https://doi.org/10.1017/pab.2025.10088

Received: 01 August 2025
Revised: 19 November 2025
Accepted: 19 November 2025

Handling Editor: Melanie Hopkins

Corresponding author:
Chi Zhang;
Email: zhangchi@ivpp.ac.cn

Abstract

Continuous morphological traits play a crucial role in phylogenetic inference, yet they are often discretized due to limited software support and challenges of handling missing data efficiently. We present a new implementation of the Brownian motion model for continuous trait evolution in the Bayesian phylogenetics software MrBayes. Our approach efficiently accommodates any proportion of missing data and supports evolutionary rate variation across characters and data partitions. It is compatible with both non-clock and relaxed clock models. We validate the implementation through simulations and apply it to empirical datasets of pterosaurs and ancient humans, demonstrating that continuous traits can improve phylogenetic resolution. This development expands the methodological tool kit for morphological and total-evidence phylogenetics and is applicable across diverse taxonomic groups.

Non-technical Summary

In paleobiology, continuous morphological traits—like limb lengths or skull dimensions—hold rich evolutionary information. However, these traits are often converted into discrete categories due to software limitations and difficulties handling missing data. This study presents a new implementation of the Brownian motion model in the widely used software MrBayes that allows direct analysis of continuous traits, even with substantial missing data. Simulations and analyses of pterosaur and ancient human datasets show that continuous traits, when used directly, can improve phylogenetic resolution compared with discretized versions. This work expands the analytical tool kit available for morphological and total-evidence phylogenetics in paleobiological research.

Introduction

Morphological traits are the primary or sole source of information for inferring the phylogenies of fossil taxa (Wiens 2004; Lee and Palci 2015). For decades, maximum parsimony has been the dominant method in the field (Farris 1970; Fitch 1971). In recent years, Bayesian inference, which better integrates various sources of information and accounts for uncertainties, has been emerging due to its desirable statistical properties (Ronquist and Huelsenbeck 2003). It not only infers tree topologies but also branch lengths in the unit of evolutionary distance. When combined with fossil ages, total-evidence or tip-dating approaches infer timetrees in geological units, evolutionary rates, and diversification patterns coherently (Pyron 2011; Ronquist et al. 2012a; Zhang et al. 2016; Donoghue and Yang 2016; Gavryushkina and Zhang 2020).

Discrete characters are the main type of data for applying these methods. Even though a substantial number of continuous traits are available, they have typically been discretized in morphological phylogenetics due to lack of feasible computational tools. In the parsimony school, Goloboff et al. (2006) added support for continuous traits by treating them as additive characters (Farris 1970) in TNT (Goloboff and Catalano 2016). Statistical models for continuous traits have also been developed for decades. Trait evolution is typically modeled using a Brownian motion (BM) process (Felsenstein 1973, 1985; Freckleton 2012) or an Ornstein–Uhlenbeck (OU) process (Felsenstein 1988; Hansen 1997; Butler and King 2004). However, practical implementations for phylogenetic inference are sparse.

In the general form of the BM model, the variance–covariance matrix, **R,** has $p(p-1)/2$ parameters describing the correlation among p continuous traits (Freckleton 2012), making the computation either too expensive or unidentifiable, even when p is moderate. This involves more parameters when considering diffusion rate variation along the tree. Pre-estimating **R** seems a practical solution (Álvarez-Carretero et al. 2019). When there are no missing or inapplicable states in the data, the implementation in MCMCTree focuses on estimating divergence times (Álvarez-Carretero et al. 2019), while that in BEAST 2 (Zhang et al. 2023) can co-estimate tree

© The Author(s), 2025. Published by Cambridge University Press on behalf of Paleontological Society. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (http://creativecommons.org/licenses/by-nc-nd/4.0), which permits non-commercial re-use, distribution, and reproduction in any medium, provided that no alterations are made and the original article is properly cited. The written permission of Cambridge University Press or the rights holder(s) must be obtained prior to any commercial use and/or adaptation of the article.

PALEOBIOLOGY A PUBLICATION OF THE Paleontological SOCIETY

CAMBRIDGE UNIVERSITY PRESS

topology and the times. In phylogenetic comparative methods, several R packages have been developed to study the traits of interest, assuming the phylogenies are known a priori (Pennell et al. 2014; Revell 2024).

However, empirical datasets typically contain many missing states. Deleting missing cells would result in a great loss of information. The ability to robustly handle missing data is not merely a "nice-to-have" feature, but an absolute "must-have" prerequisite. Apparently, handling missing data imposes great challenges. Even if **R** is known, state-of-the-art methods require either data augmentation or analytical integration (Mitov et al. 2020; Hassler et al. 2022), thus hindering their use in inferring phylogenies with moderate to large datasets with a high proportion of missing states. When the traits are assumed to be independent, RevBayes (Höhna et al. 2016) can handle continuous traits with missing data (Wynd et al. 2025), but the implementation details have not been documented. According to our experience, complex models in RevBayes (especially those not covered in tutorials) can be difficult to set up and are not always reliable.

In this study, we move one step forward and implement the BM model for continuous traits in the widely used software MrBayes (Ronquist et al. 2012b) for Bayesian phylogenetic inference. For simplicity, we do not currently consider correlation among traits. This allows us to incorporate any proportion of missing data and enables efficient likelihood calculation. The implementation can account for evolutionary rate variation across characters, among data partitions, and along the tree branches.

We first describe the implementation of the BM model. Then we perform computer simulations to briefly verify its correctness. Finally, we demonstrate its utility using two empirical datasets, pterosaurs and ancient humans, both of which have combined continuous and discrete characters.

Methods

We first assume the phylogeny is rooted and binary, with topology τ and branch lengths $\mathbf{t} = \{t_i\}$ ($i = 1, \ldots, 2s - 2$). The morphological traits are assumed to have evolved independently along the phylogeny following a BM process with diffusion rates $\mathbf{r} = \{\sigma_i^2\}$ ($i = 1, \ldots, 2s - 2$). We later show that the root position does not affect the likelihood.

The Likelihood

Let **M** be a matrix of p continuous traits for s species. The general formula of the likelihood function, $f(\mathbf{M} \mid \tau, \mathbf{t}, \mathbf{r})$, involves multivariate normal distributions (Felsenstein 1973; Freckleton 2012). To avoid extensive computation, Felsenstein (1973) introduced a pruning algorithm that computes independent contrasts while traversing the tree in post-order. He further eliminated the calculation of the root branch, resulting in the restricted maximum likelihood (REML; Felsenstein 1981). This is used in our implementation. A working example is shown in Figure 1.

Pulley Principle

Although the pruning algorithm was demonstrated on a rooted tree, Felsenstein (1981) pointed out that the position of the root does not affect the likelihood. In fact, the root ("pulley") can be "rolled onto any branch" of the tree. Thus, we can only infer an unrooted tree without further assumptions (such as a clock model or an out-group).

Missing States

We take advantage of the assumption of trait independence, and for each trait with missing states, we simply prune the branches leading to those states (illustrated in Fig. 1). The operation is trait-wise and can be performed during the application of the pruning algorithm (see the pseudocode in the Appendix). It results in a simple and efficient calculation that can handle any proportion of missing data. According to our knowledge, RevBayes handles continuous traits with missing data in a similar way (S. Höhna personal communication 2021).

Specifically, we use m_k to denote the trait state at node k, and b_k ($\sigma_k^2 t_k$) for its branch length. If k is a tip, let

$$v_k = \begin{cases} 0 & \text{if } m_k \text{ is NA,} \\ b_k & \text{otherwise.} \end{cases} \quad (1)$$

If k is an internal node with two descendant nodes i and j, let

$$v_k = \begin{cases} 0 & \text{if both } m_i \text{ and } m_j \text{ are NA,} \\ b_k + v_i & \text{else if } m_j \text{ is NA,} \\ b_k + v_j & \text{else if } m_i \text{ is NA,} \\ b_k + \dfrac{v_i v_j}{v_i + v_j} & \text{otherwise;} \end{cases} \quad (2)$$

$$m_k = \begin{cases} \text{NA} & \text{if both } m_i \text{ and } m_j \text{ are NA,} \\ m_i & \text{else if } m_j \text{ is NA,} \\ m_j & \text{else if } m_i \text{ is NA,} \\ \dfrac{m_i v_j + m_j v_i}{v_i + v_j} & \text{otherwise.} \end{cases} \quad (3)$$

The log-likelihood is a sum over all internal nodes whose child states are both available.

$$lnL = \sum \left[-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\left(v_i + v_j\right) - \frac{\left(m_i - m_j\right)^2}{2\left(v_i + v_j\right)} \right]. \quad (4)$$

Practices

We highlight a few practices that are important in empirical inferences. They are supported in MrBayes.

First, times and diffusion rates are indistinguishable in non-clock analyses. The tree is unrooted, and the branch lengths are measured by "distance" (i.e., $b_i = t_i \sigma_i^2$; $i = 1, \ldots, 2s - 3$). The times and rates become identifiable in dating analyses with additional time information (either directly from the fossils [tip dating] or from node calibrations [node dating], or both) and clock model assumptions. The inferred timetree is rooted, and the clock rates are the diffusion rates ($\mathbf{r} = \{\sigma_i^2\}$; $i = 1, \ldots, 2s - 2$). Diffusion rate variation along branches is supported by relaxed clock models (e.g., Thorne and Kishino 2002; Drummond et al. 2006; Lepage et al. 2007).

Evolutionary rate variation among data partitions typically should be considered. This feature is supported by a uniform Dirichlet distribution for the relative partition rates (Zhang et al. 2016). Diffusion rate variation among characters is also supported by averaging a discretized gamma (Yang 1994) or lognormal distribution (Wagner 2012; Harrison and Larsson 2015). However, it is more practical to normalize or standardize the continuous traits, so considering character-rate variation would be unnecessary. Our implementation supports both normalized and standardized data.

Normalization has been considered a critical step in the parsimony implementation (Goloboff et al. 2006; Mongiardino Koch et al. 2015). It rescales all the trait values to be between 0 and 1, that

https://doi.org/10.1017/pab.2025.10088 Published online by Cambridge University Press

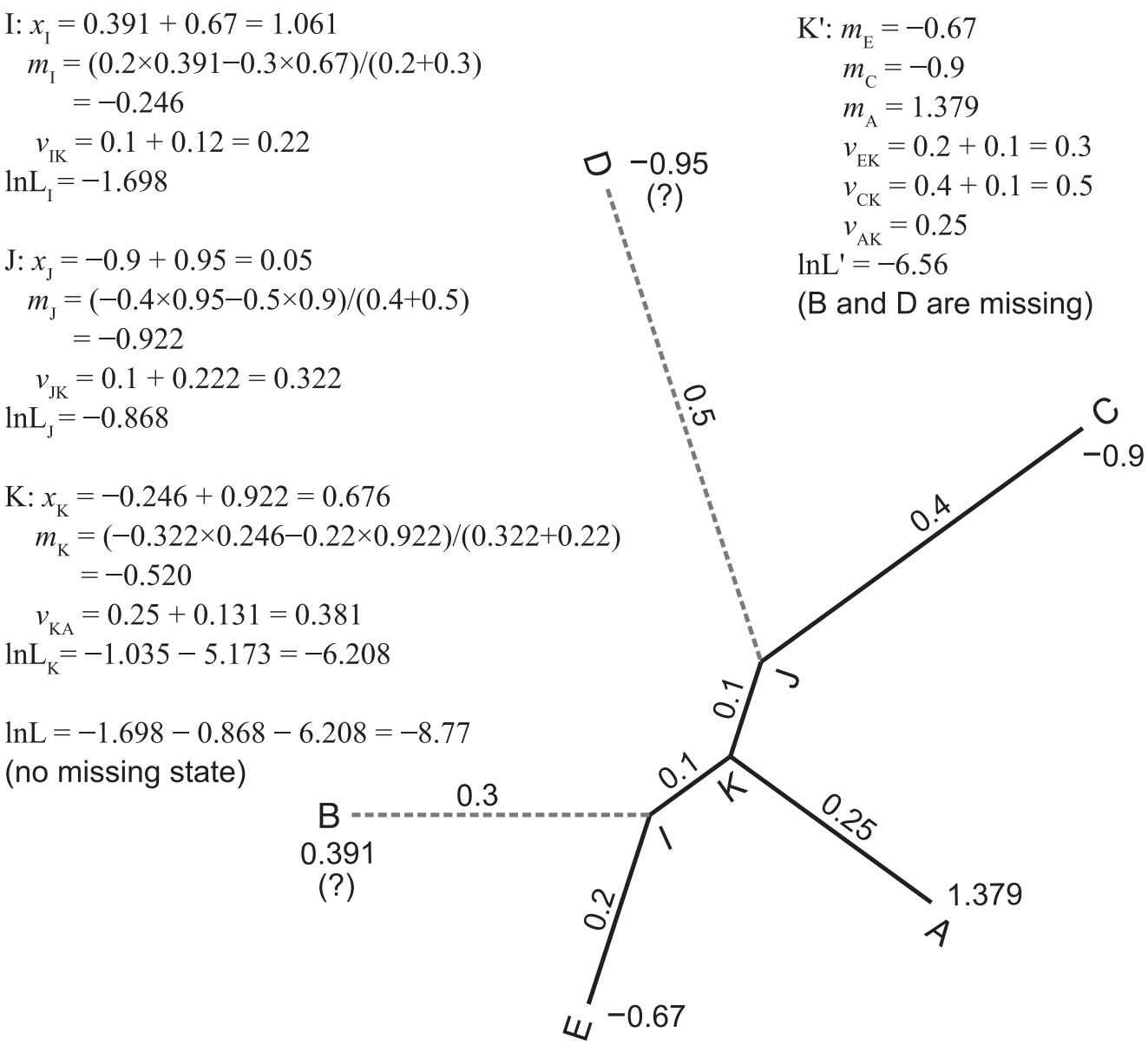


Figure 1. Pruning algorithm without and with missing states. One continuous trait with state values at the tips of a five-taxa tree is used for illustration. We assume the root is at the branch leading to taxon A. The algorithm traverses the tree (internal nodes) in post-order, that is, I, J, K, and the root. For internal node k with two descendant nodes i and j, we calculate the contrast ($x_k = m_i - m_j$), ancestral state (m_k), and transformed branch length (v_k). The contrasts follow independent normal distributions. According to the pulley principle, the root position does not affect the likelihood and can be placed anywhere in the tree. When the states at taxa B and D are missing, we simply prune the branches leading to them, resulting in a star tree with three tips, A, C, and E. Refer to the pseudocode in the Appendix for a more rigorous calculation.

is, $m_k' = (m_k - m_{\min}) / (m_{\max} - m_{\min} + \varepsilon)$, where $m_{\max}$ and $m_{\min}$ are the maximum and minimum observed values of the trait, and ε is a very small number (e.g., 0.0001). Relative to this, discretizing a trait into n states is simply taking the integer part of nm_k' (Thiele 1993). Alternatively, standardization is a common technique in statistical analyses that brings different traits to the same scale. It subtracts m_k by the sample mean and then divides by the standard deviation. For a trait with skewed values, it appears reasonable to log-transform the values before normalizing or standardizing them, to approximately conform them to normality and potentially improve model fit.

Simulation

We performed computer simulations to verify our implementation. We first generated variable timetrees from a birth–death process using TreeSim (Stadler 2011) in R (R Core Team 2021) with a birth rate of 5.0 and a death rate of 4.0, conditioned on a root age of 1.0. The ages are on a relative scale and can be arbitrarily rescaled depending on the chosen time unit. We kept 100 trees with no more than 50 tips to make sure the data size is manageable.

For each tree, we simulated the evolution of continuous traits under a BM process with a diffusion rate of 1.0. Each trait is an independent draw from the same BM process. The starting state at the root was set to 0.0. We recorded a moderate number of 200 continuous traits in the data matrix. In addition to the complete data, we also generated datasets with 50% missing states in the extinct taxa and 10% missing in the extant taxa, mimicking observations in empirical datasets. Specifically, we replaced each state by a question mark with the corresponding probability.

https://doi.org/10.1017/pab.2025.10088 Published online by Cambridge University Press

For comparisons, we also generated 200 variable binary discrete characters under the Mkv model (Lewis 2001). The substitution rate was set to 1.0 (strict clock; Zuckerkandl and Pauling 1965). And to test the functionality of supporting mixed data types, we generated data matrices each containing 100 continuous and 100 discrete characters.

We performed both non-clock and tip-dating analyses for each dataset using MrBayes (Ronquist et al. 2012b). In the non-clock analyses, we used uniform prior for the tree topologies and gamma-Dirichlet prior for the branch lengths (Zhang et al. 2012). The inferred trees are unrooted. In the tip-dating analyses, we used the constant-rate fossilized birth–death prior (Stadler 2010; Heath et al. 2014) for the timetrees and strict clock model for the evolutionary rates. The tip ages were assigned their true values assuming they are perfectly known. The inferred trees are rooted due to the clock assumption. For each inference, two independent chains were executed for enough generations to ensure convergence and sufficient effective sample sizes (ESS > 100). The posterior tree samples were summarized as a 50% majority-rule consensus tree.

We employed the Quartet metric (Estabrook et al. 1985) to compare the inferred tree topology with the true one that generated the data. It has several advantages over the traditional Robinson-Foulds metric (Robinson and Foulds 1981). Additionally, we calculated the Strict Joint Assertion (which is the number of quartets that are resolved identically in both trees over that resolved either identically or differently in both trees; Estabrook et al. 1985) as a measure of accuracy, and the percentage of resolved internal branches (the number of internal branches in the estimated consensus tree over that in the true tree) as precision.

To investigate the branch length estimates, we fixed the tree topology to the one used to simulate the data in another round of inference. The branch lengths in the tip-dating analyses are directly comparable; while in the non-clock analyses, the true tree was unrooted for comparison (by joining the two branches connecting the root). We calculated the root-mean-square error to assess the accuracy with which the branch lengths were estimated, and the mean widths of 95% credibility intervals as precision.

Simulation scripts and MrBayes commands can be found in Supplementary Material. Because RevBayes also supports phylogenetic inference using continuous traits under the BM model, as a brief verification, we also included a test comparing the non-clock inference results between MrBayes and RevBayes using a continuous trait matrix of 10 taxa generated in the simulation.

Empirical Data

We applied the new implementation to datasets of pterosaurs (Andres 2021) and ancient humans (Ni et al. 2021); each contains both continuous and discrete characters.

Pterosaurs

The pterosaur dataset has 52 continuous traits and 223 discrete ones for 177 species (Andres 2021). However, we excluded 65 of them that have more than 70% missing states in both data types, as well as one out-group species with a highly uncertain age, resulting in 111 species (three of them are treated as an out-group). The original traits are unscaled measurements, thus we normalized them to be within the range of 0 and 1. For comparison, we also discretized the continuous traits into six ordered states, which was the usual treatment before the availability of the BM model.

The continuous and discrete characters are naturally divided as two partitions. Rate variation between the partitions was accounted for using a uniform Dirichlet distribution. In the tip-dating analyses, fossil ages are gathered from the Paleobiology Database. As MrBayes does not fully support analyzing purely fossil taxa, each fossil age was reduced by 66, that is, treating the two youngest species as extant. The original data include 12 species that extend to 66 Ma, and considering potentially unsampled species, we set the sampling proportion of "extant" species to 0.1. The prior for the root age was assigned an offset-exponential distribution with mean 186 (the boundary between Permian and Triassic, minus 66) and offset 176 (close to the age of the oldest out-group fossil, *Euparkeria capensis*). We tested both the independent-lognormal (Drummond et al. 2006) and autocorrelated-lognormal (Thorne and Kishino 2002) relaxed clock models. In each case, we either linked or unlinked the clock rate variation for the two partitions.

We executed four independent runs and four Metropolis-coupled chains per run for 40 million generations in the non-clock analysis, or for 80 million generations in the tip-dating analyses, and sampled every 1000 generations. The beginning 25% samples were discarded as burn-in, and the remaining samples from the four runs were combined after confirming consistency. We checked the parameter traces, effective sample sizes, average standard deviation of split frequencies and treespaces using the R package RWTY (Warren et al. 2017).

Ancient Humans

The human dataset has 400 continuous traits and 234 discrete ones for 55 operational taxonomic units (Ni et al. 2021). The original study discretized the continuous traits into six ordered states. Here, we compare those results with analyses using continuous traits modeled under the BM process. The continuous traits have already been normalized.

We focused on the tip-dating analyses and used the white noise relaxed clock model (Lepage et al. 2007) to match the original study. We either linked or unlinked the clock rate variation for the two partitions. The root age prior was assigned Uniform(2.8, 3.6) (time is measured by million years). *Homo habilis* was used as the out-group, and we did not enforce extra topological constraints to see if the BM model could improve resolution of the resulting tree. We executed four independent runs and four Metropolis-coupled chains per run for 50 million generations and sampled every 500 generations.

Data, commands, and results (consensus trees in Newick format) can be found in the Supplementary Material.

Results

Simulation

The tree topology is estimated more reliably using continuous traits than using discrete characters, as measured by the Quartet distances (Fig. 2). This results from both increased accuracy and precision (Supplementary Fig. S1). The branch length estimation also has higher accuracy and precision (less bias and narrower credibility interval; Fig. 2). Similarly, using mixed data is more beneficial than using discrete characters alone (with an equal number of characters). Having missing data (50% in fossil taxa) reduces the accuracy and precision, resulting in similar performance to that using discrete data without missing states (Fig. 2).

https://doi.org/10.1017/pab.2025.10088 Published online by Cambridge University Press

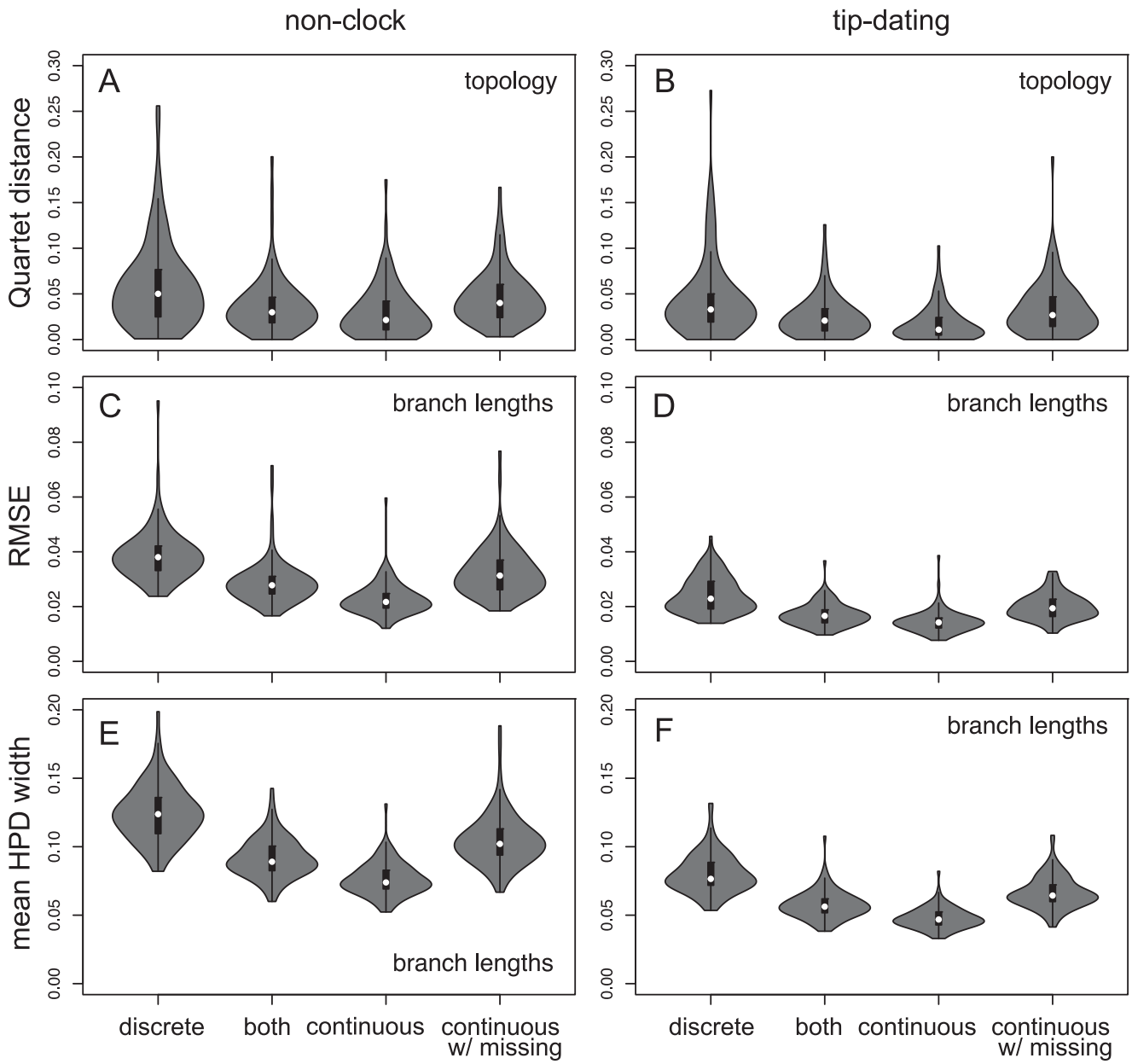


Figure 2. Quartet distance metric comparing the inferred tree topology with the true tree (**A** and **B**), and root-mean-square error (RMSE; **C** and **D**) and mean widths of 95% highest posterior density (HPD) intervals (**E** and **F**) of the estimated branch lengths. Each violin plot contains 100 replicates. The four data types are: (1) 200 variable binary discrete characters; (2) 100 variable binary discrete characters and 100 continuous traits; (3) 200 continuous traits; and (4) 200 continuous traits with 50% missing states in the extinct taxa and 10% missing in the extant taxa.

The branch lengths are more accurately and precisely estimated in the tip-dating analyses; this is due to the use of tip dates, which increases the information in the data, and of simple strict clock assumption, which just introduces one extra parameter. Overall, the simulation results indicate that the relevant functionalities are correctly implemented.

Pterosaurs

The inferred trees have good resolution in all the analyses performed after removing the species with more than 70% missing states (Supplementary Material, tree files). The inferred topologies generally agree with the most parsimonious tree (Fig. 3). The average substitution rate of the discrete data partition is about four times the diffusion rate of the continuous data partition (Supplementary Material, log files).

For the tip-dating analyses, convergence and mixing are good under the independent-lognormal relaxed clock model with either shared or unlinked rates between the two partitions. However, given the limited number of characters, particularly continuous ones, unlinking the clock rates does not appear beneficial for time-tree inference, especially considering it doubles the number of parameters. Convergence and mixing are significantly more difficult under the autocorrelated-lognormal relaxed clock model, especially when the clock rates are unlinked (Supplementary Material, log files). We have to rerun the analyses multiple times to achieve

https://doi.org/10.1017/pab.2025.10088 Published online by Cambridge University Press

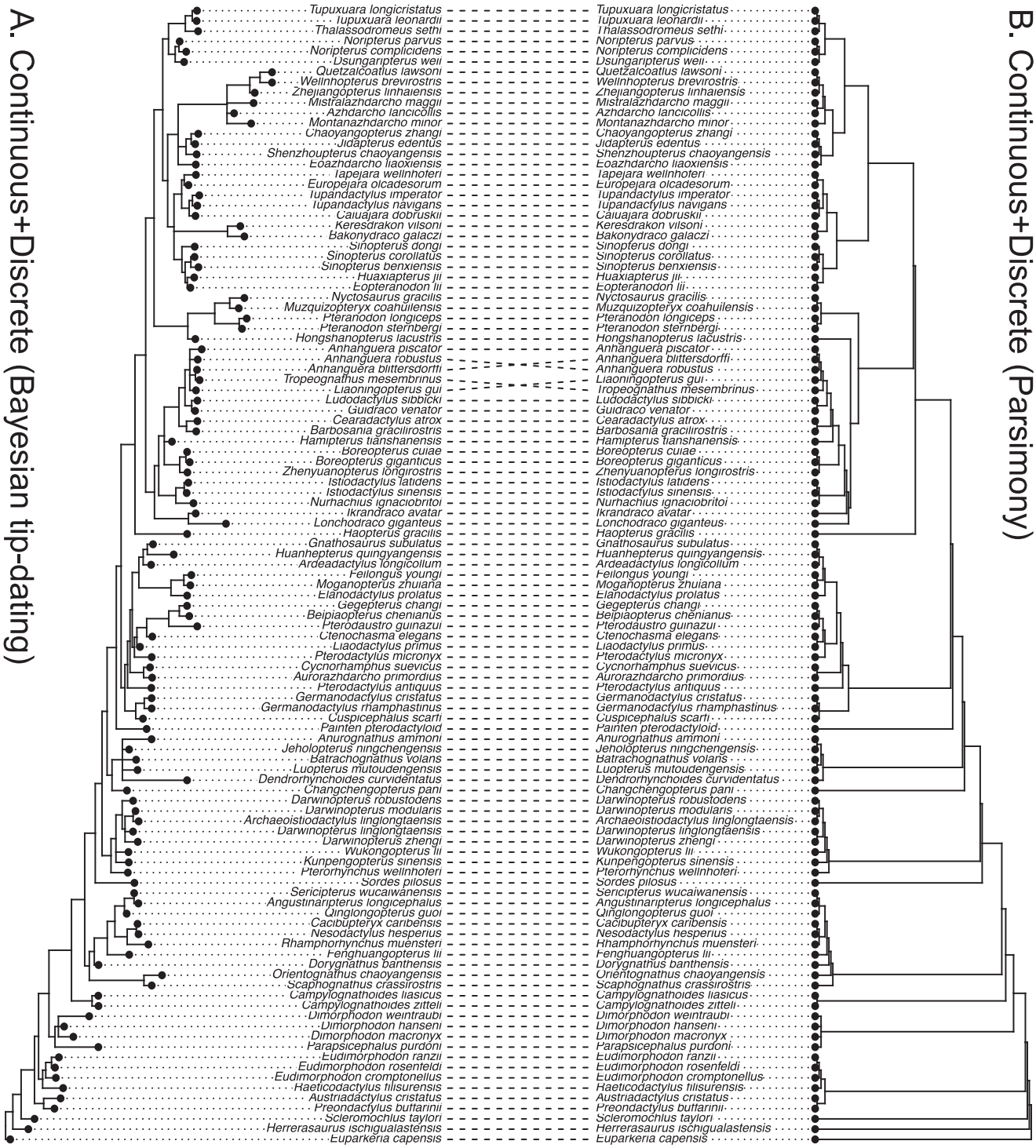


Figure 3. Tanglegram comparing the 50% majority-rule consensus tree from the Bayesian tip-dating analysis (**A**) to the strict consensus of the most parsimonious trees from TNT (**B**) using both continuous (normalized) and discrete morphological characters of pterosaurs. The relaxed clock model is independent-lognormal shared between the two partitions in the tip-dating analysis.

convergence of all the chains. The marginal likelihood under the autocorrelated-lognormal model is much lower than that under the independent-lognormal model (roughly assessed by the harmonic mean estimates; Supplementary Material, log files). Therefore, we prefer the results obtained under the linked independent-lognormal relaxed clock model.

Ancient Humans

The dated phylogeny using both continuous and discrete characters shows improved resolution compared with that obtained when the continuous characters are discretized (Supplementary Material, tree files). Discretization appears to have caused a loss of information. The average substitution rate of the discrete data partition is

https://doi.org/10.1017/pab.2025.10088 Published online by Cambridge University Press

about eight times the diffusion rate of the continuous data partition. Compared with the pterosaur dataset, this dataset includes many more continuous traits; as a result, we can infer different patterns of branch rate variation across partitions more reliably (Supplementary Material).

The original study enforced several topological constraints based on the most parsimonious tree (Ni et al. 2021). Without these constraints, the present results show unresolved relationships among some lineages (Fig. 4). Note that the most parsimonious tree only represents a point estimate without considering uncertainties. The treespace appears multimodal in the human data compared with the treespace in the pterosaur data (Supplementary Figs. S2, S3). Nevertheless, our aim is to just illustrate the usage of the BM model rather than to clarify human evolution, which is not possible based on this single dataset.

Discussion

The ability to leverage continuous traits in phylogenetic inference is an appealing feature. Our study provides an efficient implementation

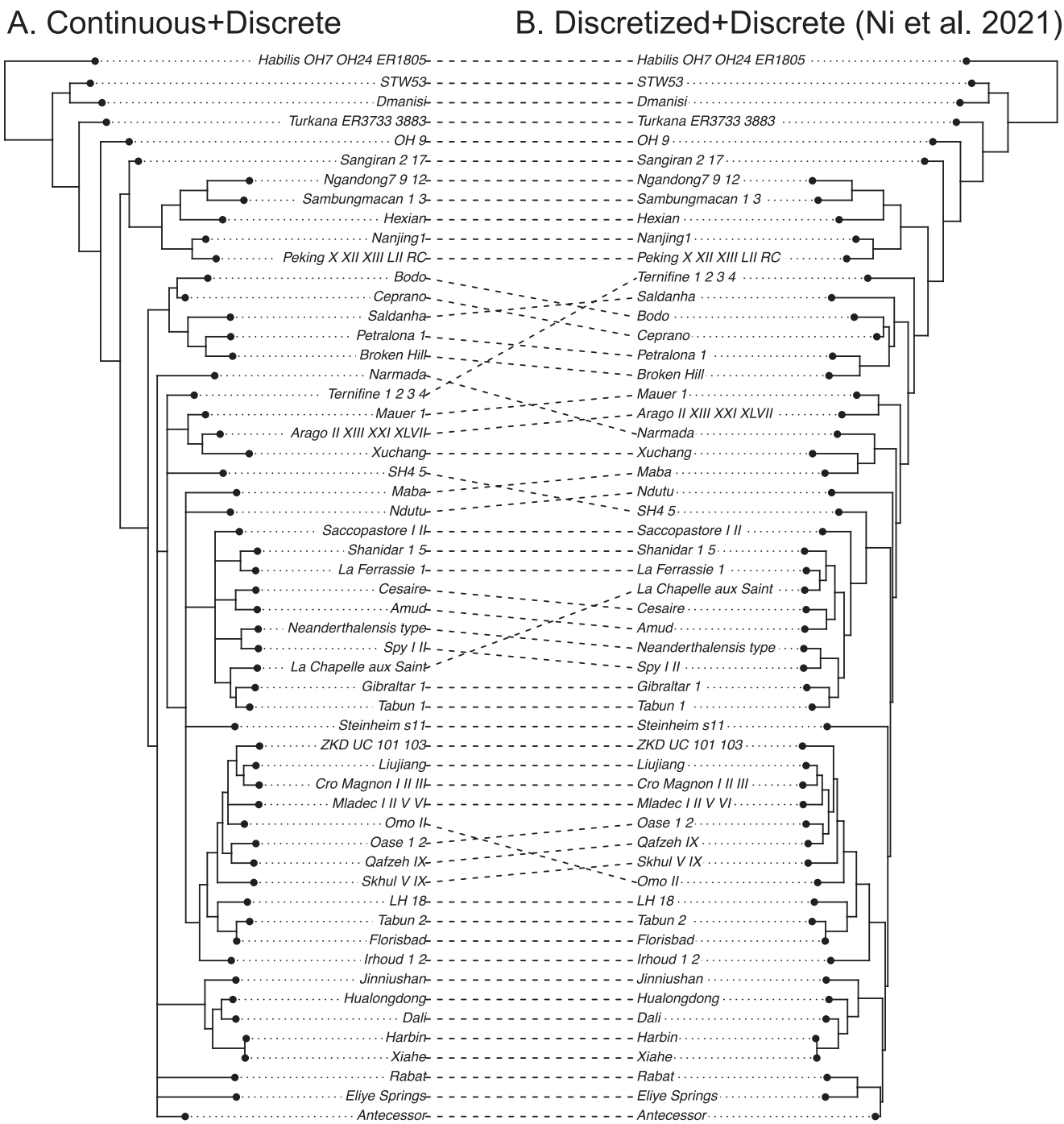


Figure 4. Tanglegram comparing the 50% majority-rule consensus tree using both continuous (normalized) and discrete morphological characters (**A**) with that using discretized continuous and discrete characters (**B**) of ancient humans. Both analyses used the white noise relaxed clock unlinked between the two partitions. The original study (tree on the right) enforced 10 topological constraints referring to the most parsimonious tree from TNT using both continuous and discrete characters (Ni et al. 2021).

https://doi.org/10.1017/pab.2025.10088 Published online by Cambridge University Press

of the BM model for Bayesian phylogenetic inference using continuous traits with missing states, integrated into the widely used software MrBayes. The implementation extends a suite of existing features toward morphological and total-evidence phylogenetics, combining continuous and discrete characters as well as molecular sequences.

Our simulation results indicate that continuous traits contain more phylogenetic information than discrete traits evolved at the same rate. Discrete characters could suffer from multiple hits, especially when the substitution rate is high; by contrast, continuous traits are less saturated. Similar patterns have been shown in a previous simulation study comparing continuous and discrete characters evolved at several different rates (Parins-Fukuchi 2018). Continuous traits should preferably be directly incorporated when available rather than discarded or discretized.

We implemented an approach that efficiently handles missing states by pruning branches. However, we still estimate the same number of branch lengths, as there were no missing data (e.g., $2s - 3$ for unrooted tree). We do not expect a taxon with completely missing states or with states that do not overlap with those of other taxa, so there is information to infer each length. But the missing states in a real dataset are unlikely randomly distributed, thus the reliability of each branch length estimate is likely data dependent.

One limitation of our implementation is that correlation among traits is not considered. Simulation studies have shown that topological inference is relatively robust to violations of the trait-independence assumption (Parins-Fukuchi 2018), but evolutionary rates (or branch lengths in non-clock analyses) are largely overestimated (Álvarez-Carretero et al. 2019). Thus, we advise performing subsequent comparative analyses that properly account for among-trait correlations to study the evolution of particular traits of interest.

Acknowledgments. This research was supported by the National Key Research and Development Program of China (2023YFF0804502 to C.Z.) and the National Natural Science Foundation of China (42172006 to C.Z.).

Author Contribution. C.Z. designed the study and led the analyses. Both authors contributed to the implementation in MrBayes. C.Z. led the interpretation of the results and writing of the manuscript. Both authors revised the manuscript and gave final approval for publication.

Competing Interests. The authors declare no competing interests.

Data Availability Statement. Electronic Supplementary Material is available from the Zenodo Digital Repository (https://doi.org/10.5281/zenodo.17647357). The Brownian model is implemented in a feature branch of MrBayes (https://github.com/zhangchicool/MrBayes) and will be merged to the main branch for the upcoming release (version 3.2.8).

Literature Cited

Álvarez-Carretero, S., A. Goswami, Z. Yang, and M. dos Reis. 2019. Bayesian estimation of species divergence times using correlated quantitative characters. *Systematic Biology* **68**:967–986.

Andres, B. 2021. Phylogenetic systematics of *Quetzalcoatlus* Lawson 1975 (Pterodactyloidea: Azhdarchoidea). *Journal of Vertebrate Paleontology* **41**: 203–217.

Butler, M. A., and A. A. King. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *American Naturalist* **164**:683–695.

Donoghue, P. C. J., and Z. Yang. 2016. The evolution of methods for establishing evolutionary timescales. *Philosophical Transactions of the Royal Society B* **371**:20160020.

Drummond, A., S. Ho, M. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology* **4**:e88.

Estabrook, G. F., F. R. McMorris, and C. A. Meacham. 1985. Comparison of undirected phylogenetic trees based on subtrees of four evolutionary units. *Systematic Zoology* **34**:193–200.

Farris, J. S. 1970. Methods for computing Wagner trees. *Systematic Biology* **19**: 83–92.

Felsenstein, J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. *American Journal of Human Genetics* **25**:471–492.

Felsenstein, J. 1981. Evolutionary trees from gene frequencies and quantitative characters: finding maximum likelihood estimates. *Evolution* **35**:1229–1242.

Felsenstein, J. 1985. Phylogenies and the comparative method. *American Naturalist* **125**:1–15.

Felsenstein, J. 1988. Phylogenies and quantitative characters. *Annual Review of Ecology and Systematics* **19**:445–471.

Fitch, W. M. 1971. Toward defining the course of evolution: minimum change for a specified tree topology. *Systematic Zoology*. **20**:406–416.

Freckleton, R. P. 2012. Fast likelihood calculations for comparative analyses. *Methods in Ecology and Evolution* **3**:940–947.

Gavryushkina, A., and C. Zhang. 2020. Total-evidence dating and the fossilized birth–death model. Pp. 175–193 *in* S. Y. W. Ho, ed. *The molecular evolutionary clock*. Springer, Cham, Switzerland.

Goloboff, P. A., and S. A. Catalano. 2016. TNT version 1.5, including a full implementation of phylogenetic morphometrics. *Cladistics* **32**:221–238.

Goloboff, P. A., C. I. Mattoni, and A. S. Quinteros. 2006. Continuous characters analyzed as such. *Cladistics* **22**:589–601.

Hansen, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation. *Evolution* **51**:1341–1351.

Harrison, L. B., and H. C. E. Larsson. 2015. Among-character rate variation distributions in phylogenetic analysis of discrete morphological characters. *Systematic Biology* **64**:307–324.

Hassler, G., M. R. Tolkoff, W. L. Allen, L. S. T. Ho, P. Lemey, and M. A. Suchard. 2022. Inferring phenotypic trait evolution on large trees with many incomplete measurements. *Journal of the American Statistical Association* **117**:678–692.

Heath, T. A., J. P. Huelsenbeck, and T. Stadler. 2014. The fossilized birth–death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences USA* **111**:E2957–E2966.

Höhna, S., M. J. Landis, T. A. Heath, B. Boussau, N. Lartillot, B. R. Moore, J. P. Huelsenbeck, and F. Ronquist. 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic Biology* **65**:726–736.

Lee, M. S. Y., and A. Palci. 2015. Morphological phylogenetics in the genomic age. *Current Biology* **25**:R922–R929.

Lepage, T., D. Bryant, H. Philippe, and N. Lartillot. 2007. A general comparison of relaxed molecular clock models. *Molecular Biology and Evolution* **24**: 2669–2680.

Lewis, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* **50**:913–925.

Mitov, V., K. Bartoszek, G. Asimomitis, and T. Stadler. 2020. Fast likelihood calculation for multivariate Gaussian phylogenetic models with shifts. *Theoretical Population Biology* **131**:66–78.

Mongiardino Koch, N., Soto, I. M., and Ramírez, M. J. 2015. First phylogenetic analysis of the family Neriidae (Diptera), with a study on the issue of scaling continuous characters. *Cladistics* **31**:142–165.

Ni, X., Q. Ji, W. Wu, Q. Shao, Y. Ji, C. Zhang, L. Liang, et al. 2021. Massive cranium from Harbin in northeastern China establishes a new Middle Pleistocene human lineage. *The Innovation* **2**:100130.

Parins-Fukuchi, C. 2018. Use of continuous traits can improve morphological phylogenetics. *Systematic Biology* **67**:328–339.

Pennell, M. W., J. M. Eastman, G. J. Slater, J. W. Brown, J. C. Uyeda, R. G. FitzJohn, M. E. Alfaro, and L. J. Harmon. 2014. geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics* **30**:2216–2218.

Pyron, R. A. 2011. Divergence time estimation using fossils as terminal taxa and the origins of Lissamphibia. *Systematic Biology* **60**:466–481.

R Core Team. 2021. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Revell, L. J. 2024. phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and other things). *PeerJ* **12**:e16505.

Robinson, D. F., and L. R. Foulds. 1981. Comparison of phylogenetic trees. *Mathematical Biosciences* **53**:131–147.

Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**:1572–1574.

https://doi.org/10.1017/pab.2025.10088 Published online by Cambridge University Press

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

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

Stadler, T. 2010. Sampling-through-time in birth–death trees. *Journal of Theoretical Biology* **267**:396–404.

Stadler, T. 2011. Simulating trees with a fixed number of extant species. *Systematic Biology* **60**:676–684.

Thiele, K. 1993. The Holy Grail of the perfect character: the cladistic treatment of morphometric data. *Cladistics* **9**:275–304.

Thorne, J. L., and H. Kishino. 2002. Divergence time and evolutionary rate estimation with multilocus data. *Systematic Biology* **51**:689–702.

Wagner, P. J. 2012. Modelling rate distributions using character compatibility: implications for morphological evolution among fossil invertebrates. *Biology Letters* **8**:143–146.

Warren D. L., A. J. Geneva, and R. Lanfear. 2017. RWTY (R We There Yet): an R package for examining convergence of Bayesian phylogenetic analyses. *Molecular Biology and Evolution* **34**:1016–1020.

Wiens, J. J. 2004. The role of morphological data in phylogeny reconstruction. *Systematic Biology* **53**:653–661.

Wynd, B. M., B. Khakurel, C. F. Kammerer, P. J. Wagner, and A. M. Wright. 2025. Incorporating continuous characters in joint estimation of dicynodont phylogeny. *Systematic Biology*, syaf078.

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

Zhang, C., B. Rannala, and Z. Yang. 2012. Robustness of compound Dirichlet priors for Bayesian inference of branch lengths. *Systematic Biology* **61**:779–784.

Zhang, C., T. Stadler, S. Klopfstein, T. A. Heath, and F. Ronquist. 2016. Total-evidence dating under the fossilized birth-death process. *Systematic Biology* **65**:228–249.

Zhang, R., A. J. Drummond, and F. K. Mendes. 2023. Fast Bayesian inference of phylogenies from multiple continuous characters. *Systematic Biology* **73**: 102–124.

Zuckerkandl, E., and L. Pauling. 1965. Evolutionary divergence and convergence in proteins. Pp. 97–166 *in* V. Bryson and H. J. Vogel, eds. *Evolving genes and proteins*. Academic Press, New York.

Appendix

Pruning algorithm

```
1: Initialize log-likelihood: lnL ← 0
2: For each trait c:
3:    For each node k in T (post-order):
4:       If k is a tip node:
5:          If m[c,k] is NA:
6:             v[c,k] ← 0
7:          Else:
8:             v[c,k] ← b[k] // branch length of
               k in T
9:       Else (k is an internal node):
10:         Let i and j be the two child nodes
            of k
11:         If m[c,i] and m[c,j] are both NA:
12:            v[c,k] ← 0
13:            m[c,k] ← NA
14:         Else if m[c,i] is observed and m[c,j]
            is NA:
15:            v[c,k] ← b[k] + v[c,i]
16:            m[c,k] ← m[c,i]
17:         Else if m[c,i] is NA and m[c,j] is
            observed:
18:            v[c,k] ← b[k] + v[c,j]
19:            m[c,k] ← m[c,j]
20:         Else (both m[c,i] and m[c,j] are
            observed):
21:            v[c,k] ← b[k] + v[c,i]*v[c,j] / (v[c,i]
                        + v[c,j])
22:            m[c,k] ← (v[c,j]*m[c,i] + v[c,i]
                        *m[c,j]) / (v[c,i] + v[c,j])
23:            lnL ← lnL -0.5*log(2*pi)
                         -0.5*log(v[c,i] + v[c,j])
                         -0.5*(m[c,i] -m[c,j])
                          ^2 / (v[c,i]+v[c,j])
24: Return lnL
```

https://doi.org/10.1017/pab.2025.10088 Published online by Cambridge University Press