

Estimating the rate of quantitative trait evolution in the presence of gene tree discordance by calculating likelihoods across trees

Yu K. Mo^{1,*} and Matthew W. Hahn^{1,2}

¹Department of Computer Science, Indiana University, Bloomington, IN 47408, United States

²Department of Biology, Indiana University, Bloomington, IN 47405, United States

*Corresponding author. Department of Computer Science, Indiana University, Bloomington, IN, 47408, United States. E-mail: moyu@iu.edu

ABSTRACT

Quantitative traits provide insights into how phenotypes evolve across species. However, standard comparative methods often assume a single species tree and overlook the discordant gene tree histories that may underlie complex traits. Here, we develop a model and software (Spinney) that explicitly incorporate gene tree heterogeneity into rate estimation. Spinney finds the optimal rate of evolution and ancestral states by jointly maximizing the likelihoods across a set of gene trees. Using simulated data, we compare rate estimates from Spinney with those using the species tree alone, test Spinney's ability to distinguish true rate variation from spurious signals caused by gene tree discordance, and evaluate ancestral state reconstruction. Spinney consistently produced more accurate rate estimates and reduced incorrect inferences of rate variation. This method therefore provides a flexible framework to integrate gene tree heterogeneity into comparative methods and to produce reliable inferences of quantitative trait evolution, regardless of the source of discordance.

Keywords: pruning algorithm; phylogenetic comparative methods; rate variation

INTRODUCTION

The rate at which quantitative traits evolve among species can tell us much about the evolutionary forces affecting phenotypes. Previous studies have used quantitative traits to investigate the strength and frequency of adaptive evolution (e.g. Hansen 1997, Khabbazian *et al.* 2016), study the effect of changes in selective regimes on trait evolution (e.g. Rafferty and Nability 2017, Bertram *et al.* 2023), and to understand the traits underlying species diversification (e.g. Uyeda *et al.* 2011, Helmstetter *et al.* 2023). Phylogenetic comparative methods (PCMs) are widely applied to make such inferences on trees (e.g. Felsenstein 1985, Pagel 1999, Blomberg *et al.* 2003, O'Meara *et al.* 2006, Adams 2014).

Most PCMs make the fundamental assumption that the species tree represents the evolutionary history of traits. While there may be issues if the species tree is inferred in error (Martins 1996), a larger problem can occur when a trait's actual history does not match the species history. Gene tree discordance describes topological differences both among individual gene trees and with the species tree. Discordance is ubiquitous among eukaryotes, largely due to incomplete lineage sorting (ILS) and introgression

(Degnan and Rosenberg 2009). Because quantitative traits are determined by genes that may be discordant, such traits do not have to have the same history as the species tree, leading to inaccurate inferences. For example, evolutionary rates are consistently overestimated in the presence of even low levels of discordance due to ILS (Mendes *et al.* 2018). Similarly, Adams *et al.* (2025) demonstrated that phylogenetic regression models inaccurately estimate trait–environment relationships when there is gene tree discordance due to ILS. Discordance due to introgression can also lead to inaccurate inferences when one assumes that traits follow a bifurcating species tree (e.g. Bastide *et al.* 2018, Hibbins and Hahn 2021).

Multiple studies have attempted to model traits that may have discordant histories. Bastide *et al.* (2018) allowed quantitative traits to evolve along both the species tree and an introgression history. While this does not model gene tree discordance directly, it does allow traits to have multiple histories. More recently, Hibbins *et al.* (2023) developed two methods that allowed for discordance due to ILS or introgression, showing that both led to more accurate inferences. The first method incorporated gene tree

discordance into the construction of the covariance matrix, C , used by many PCMs. The second method used the pruning algorithm (Felsenstein 1973) to calculate the weighted likelihood of quantitative trait evolution over the tree topologies possible in a species tree. Although these latter results were limited to three species, the pruning algorithm offers many advantages because of its ability to specify diverse transition matrices, to efficiently calculate likelihoods, and to reconstruct ancestral states (Freckleton 2012).

Here, we introduce a new method for inferring evolutionary rates and ancestral states of quantitative traits. This new method—implemented in the program Spinney—is built on the pruning algorithm framework of Hibbins *et al.* (2023), but accommodates any number of species. In addition, Spinney introduces new methods both for accurately estimating ancestral states and for accommodating multiple evolutionary rate regimes across a species tree in the presence of discordance. We demonstrate the utility and accuracy of Spinney using simulated data, finding the areas of tree space where it is most useful.

NEW APPROACHES

Rate estimates inferred from a set of gene trees

Our method models a continuous trait evolving via Brownian motion. The transition matrix of this model is determined by the probability density function of Brownian motion, where x_i and x_t represent the initial and final trait values over time interval t , and σ^2 is the evolutionary rate parameter:

$$p(x_i, x_t, t) = \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-\frac{(x_i - x_t)^2}{2\sigma^2 t}} \quad (1)$$

To infer the rate parameter, σ^2 , while accounting for discordance, our method uses a group of gene trees (a ‘spinney’) with branch lengths in coalescent units, along with weights for each gene tree (w_i ; Fig. 1). If each of n trees is unique, the weight of each is simply $1/n$. The likelihood of each gene tree given data at the tips is computed individually using Felsenstein’s pruning algorithm (the same data are used for each tree). Our approach discretizes trait values into a probability vector whose length

is the same across nodes within a gene tree and across all gene trees. For tip values, the probability of the observed state is 1 if the value coincides with a grid point; otherwise, its mass is split by linear interpolation between the two nearest grid points (cf. Boucher and Démary 2016). The trait vectors assigned to the leaves serve as the initial states for likelihood calculations for all gene trees, and entries are updated with the pruning algorithm accordingly. Spinney calculates the likelihood for each tree in parallel; this reduces runtime, especially when n is large.

The total likelihood is obtained by summing the weighted likelihoods of individual gene trees, which is then used for numerical optimization to infer σ^2 and ancestral states:

$$L_{\text{total}} = \sum_{i=1}^N w_i \log L_i \quad (2)$$

As in Hibbins *et al.* (2023), our model can take as input ultrametric gene trees either estimated from data or sampled from the species tree under the multispecies coalescent (MSC) model (or any other arbitrary model). We do not assume that we have sampled gene trees at the exact loci underlying our quantitative trait, only that our sample is representative of such trees. Spinney is implemented in Python using several standard libraries. NumPy (Harris *et al.* 2020) provides basic numerical calculations and DendroPy (Moreno *et al.* 2024) supports tree processing. Rate optimization is carried out with the Nelder–Mead algorithm from SciPy (Virtanen *et al.* 2020), and parallelization is implemented with multiprocessing. The source code is freely available on GitHub (<https://github.com/speechless/spinney>).

Incorporating rate variation using branch segments

The rate of evolution of quantitative traits can vary across a phylogeny, with the σ^2 parameter being higher or lower in different clades. One important use of PCMs is in identifying this rate variation (O’Meara *et al.* 2006). However, because more discordance can result in higher rate estimates, some parts of the species tree may give false signals of rate acceleration simply because of different levels of gene tree heterogeneity (Hibbins *et al.* 2023). Here, our aim is to accommodate different levels of discordance when

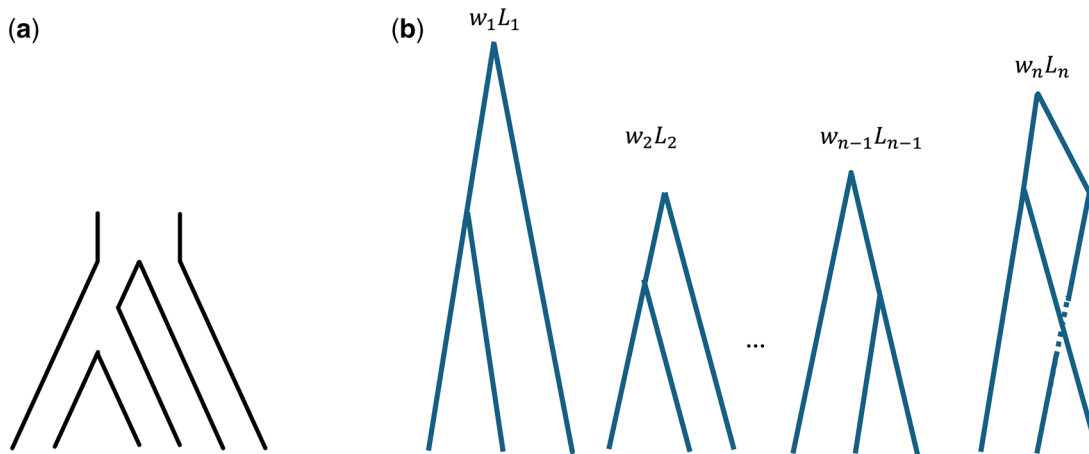


Figure 1. Spinney uses a set of trees for inference. Given (a) a species tree, (b) a set of gene trees is sampled. For the i -th gene tree, the weight is denoted as w_i and the likelihood is L_i . Inferences are carried out based on the summation of weighted likelihood across the gene trees.

testing for the presence of multiple rate parameters across a species tree.

When calculating shifts in rates directly on a species tree, rates are naturally set to change at speciation nodes. However, when modelling traits on gene trees, internal nodes of such trees do not correspond to speciation events. Furthermore, these internal nodes do not have to be the same age across gene trees. To deal with this heterogeneity, we introduce the idea of unary ‘time-slicing’ nodes that correspond to the age of speciation events (Fig. 2a). In this approach, a fully bifurcating species tree is required as a reference tree for assigning time-slicing nodes. Evolutionary rates (i.e. σ^2 parameters) are assigned to branches of the species tree; each gene tree contains time-slicing nodes at times corresponding to speciation events, provided that the lineage at that time has the same set of descendants as in the species tree (cf. Ogilvie *et al.* 2017; note that this does not imply that gene trees must be concordant). These nodes segment each gene tree, ensuring that all segments corresponding to the same branch in the species tree share the same rate (Fig. 2b, c). In this way, rates occurring on different branches of the species tree can correctly be applied to the appropriate branches of each gene tree.

Ancestral state reconstruction

Reconstructing ancestral states traces how traits change through evolutionary history, revealing when shifts occurred and whether similar traits evolved independently in different lineages. However, a challenge again arises because gene trees may differ in both topology and internal-node ages from species trees.

To resolve this problem, we again use the same time-slicing nodes described above. These nodes provide a unified timeframe, allowing discordant gene trees to be combined and ancestral states to be mapped back onto the species tree. Ancestral states are typically estimated at speciation nodes of a phylogeny (e.g. Schluter *et al.* 1997), even though gene trees do not coalesce until after speciation (looking backward in time; Gillespie and Langley 1979). Time-slicing nodes allow us to obtain maximum-likelihood estimates of ancestral states at speciation times averaged across

individual gene trees. For a given speciation event, we identify taxa associated with the event, identify corresponding branches from relevant gene trees, and add time-slicing nodes to those branches. On the species tree, the ancestral state combines time-slicing nodes across relevant lineages from all gene trees, weighting the trait vectors recorded for each gene tree according to tree weights.

METHODS

Simulating traits and gene trees

We conducted a series of simulations to evaluate the performance of our method. We simulated 1000 gene trees with the program *ms* (Hudson 2002) under each species tree described below. We also constructed a covariance matrix, C' , from the simulated gene trees using the R package *seastaR*, in order to incorporate discordant histories (Hibbins *et al.* 2023). For each condition, we then simulated 100 independent traits under a multivariate normal distribution. The trait values across species were drawn as $\mathbf{X} \sim N(\mathbf{0}, \Sigma \odot C')$, where $\mathbf{0}$ is a zero vector for the ancestral values, C' is the covariance matrix derived from simulated gene trees, and Σ is a scaling matrix of the same dimension as C' . Σ controls the rate structure of trait evolution. When $\Sigma = \sigma^2(\mathbf{1}\mathbf{1}^T)$, it corresponds to a scenario with a homogeneous rate σ^2 . More generally, it can take different values across entries, allowing simulation with heterogeneous rates across branches of the species tree.

Gene trees used for likelihood calculations were also sampled from *ms* based on a given species tree; this is equivalent to assuming the MSC model with no introgression. To balance computational efficiency with inference accuracy, we limited the number of sampled gene trees to a sufficiently large yet manageable number (Supporting Information Fig. S1). We used 20 gene trees as the default for inference, where each tree has the same weight, unless otherwise specified. We use σ_y^2 and σ_{st}^2 to denote estimated rate parameters inferred by Spinney and the species tree, respectively.

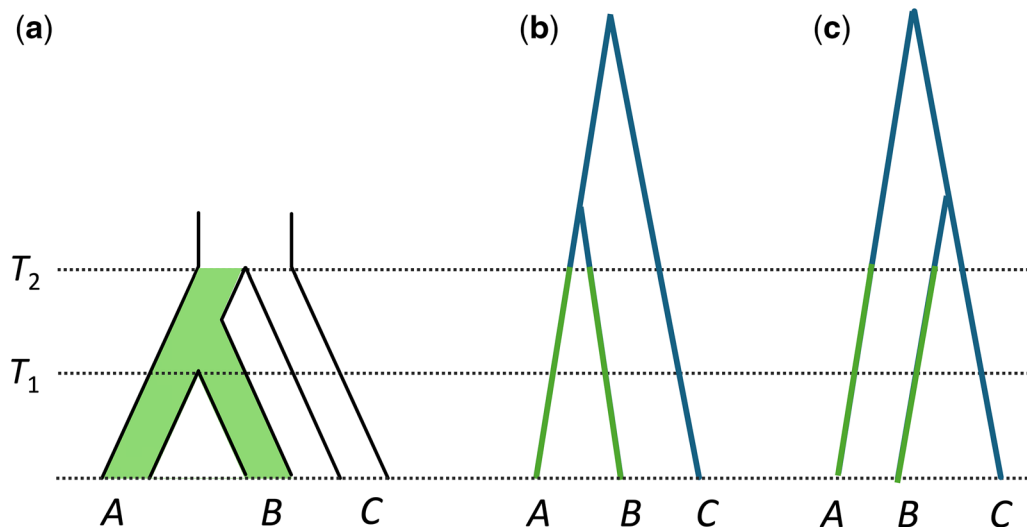


Figure 2. Spinney allows multiple rates within a tree. (a) Given a species tree with speciation time T_1 and T_2 , evolutionary rates are assigned to branches of this tree. Here we imagine that there are two rates in the species tree (green and white branches). Gene tree segments that exist within the same species tree branch share the same evolutionary rate, whether a gene tree is (b) concordant or (c) discordant.

Single-rate tests

For the single-rate tests, we started with a four-taxon tree (Fig. 3a), keeping the total tree height fixed at 10 coalescent units while adjusting internal branch lengths to obtain discordance levels of 1.6%, 25%, and 50%. To test how Spinney generalizes to larger trees, we added cases with 5, 10, 20 and 50 taxa. The species trees for these larger trees were simulated using the R package *ape* (Paradis and Schliep 2019), with birth and death rates set to 0.1. The discordance levels among these trees are 53%, 66%, 68%, and 98%, respectively. The evolutionary rate was kept constant ($\sigma^2 = 1$) for all single-rate tests.

Two-rate tests

We used several scenarios to evaluate methods for detecting multiple rate parameters. We started with a species tree with clades A and B (Fig. 3b; Supporting Information Fig. S2a). We ensured the internal branches extending to clades A and B were sufficiently long enough to prevent discordance between clades. Clade A has a shorter internal branch in order to introduce higher discordance, whereas clade B has a longer internal branch and shows little discordance. Consequently, almost all discordance occurs in clade A, leading to 50% discordance. Single-rate simulations were run with $\sigma^2 = 1$, and additional scenarios were tested by modifying σ^2 to 0.1 and 10 with the same species tree. Two-rate simulations were conducted on a different tree (Fig. S2b) by having σ_A^2 (the evolutionary rate in clade A) take on the values 0.01, 0.1, 1, 10, and 100, while keeping the rest of the tree at $\sigma^2 = 1$.

To estimate the false positive rate, we estimated models with both one value and two values of σ^2 , using both Spinney and calculations directly on the species tree (all data were produced using simulations with only one rate parameter in this setting). Models with two rates were fit so that clade A was assigned a distinct rate parameter from the rest of the tree (this is the clade that contains most of the discordance). We used a likelihood ratio test (LRT) to assess the false positive rate of both approaches: any result that favours two rate parameters is taken to be a false positive. Given a trait and a set of gene trees, we computed the likelihoods under both one-rate (L_1) and two-rate (L_2) models. The test statistic was computed as:

$$\lambda = -2[\log L_1 - \log L_2] \quad (3)$$

followed by a χ^2 test with one degree of freedom and a significance threshold of 5%.

Evaluating ancestral states

We examined how ancestral state reconstruction is affected by using either the species tree alone or Spinney. To evaluate this, we used the same four-taxon simulations as in the single-rate tests. Rates were inferred using a one-rate model under both the species tree and Spinney, and ancestral states were taken as the estimates associated with the maximum likelihood estimate of σ^2 . Our primary focus was on the speciation event that separates species A and B from species C (at time T_2 ; Fig. 3a).

To visualize ancestral state changes over time, we divided the tree into 100 equally wide bins of time. At each time slice (i.e. bin boundary), we recorded the estimated ancestral state. The ancestral state at the time slice immediately following a speciation event was used as an approximation of the state at that speciation time (because speciation events may not correspond exactly to the boundary of two bins). The ancestral state is represented internally as a vector of the same length as the discretized trait vector; each entry in this vector indicates the probability of observing a specific trait value. Additionally, we computed the 95% confidence interval (CI) for the ancestral state. To determine the CI, we calculated the cumulative distribution function (CDF) from the ancestral state vector. The lower bound for trait x_i was identified as the point where $\text{CDF}(x_i) = 0.025$ and the upper bound for trait x_i was where $\text{CDF}(x_i) = 0.975$. The width of the confidence interval was then calculated as $x_j - x_i$.

RESULTS

Spinney mitigates rate overestimation

For the single-rate case, we compared rate estimates obtained using Spinney with those inferred directly from the species tree. All calculations on the species tree use the standard pruning algorithm to optimize σ^2 on a tree with correctly specified species tree

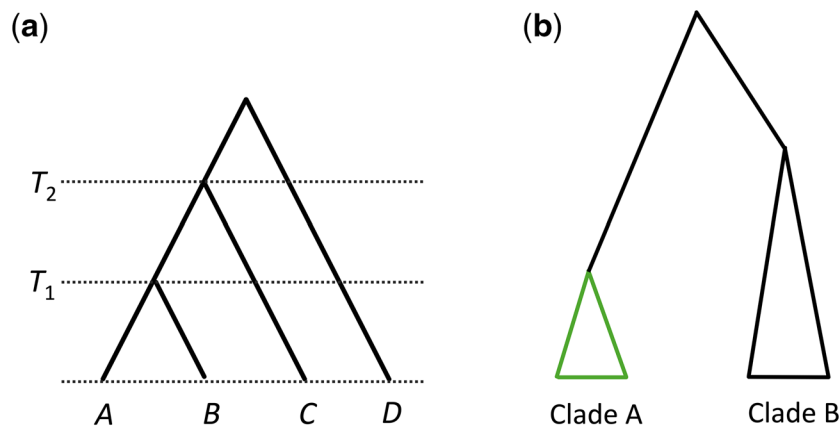


Figure 3. Species trees used for experiments. (a) Four-taxon species tree. Traits are simulated with a single, uniform rate. (b) Two-clade species tree. Traits are simulated either with a single, uniform rate (to test the false positive rate) or with two rates (to test the true positive rate). In all simulations and inferences with two rates, the branches labelled in green have a different rate from the branches labelled in black. Within the species tree, clade A and clade B have different levels of discordance in which clade A contributes all discordance and clade B has no discordance. Detailed trees with branch lengths are shown in Supporting Information Figure S2.

branch lengths. Our primary goal was to assess whether Spinney could accurately estimate the rate of evolution while accounting for gene tree heterogeneity (here caused by ILS), as opposed to using the species tree, which does not incorporate discordance.

For four-taxon trees, inference using the species tree resulted in increasingly severe rate overestimation as discordance increased (Fig. 4a). Across 100 replicates, the mean σ_{st}^2 values were 1.44, 3.66, and 12.47 for discordance levels of 1.6%, 25%, and 50%, respectively (the true value of σ^2 is always 1). Notably, even a small level of discordance led to overestimation. In contrast, Spinney produced more accurate rate estimates. The mean σ_{sy}^2 values were 1.03, 1.60, and 3.30 for discordance levels of 1.6%, 25%, and 50%, respectively. While higher levels of discordance still led to overestimation of rates, this bias was mitigated by using Spinney.

Additionally, the standard deviation of rate estimates from Spinney was smaller compared to those from the species tree (Supporting Information Table S1).

In the above results, Spinney was run by calculating the partial likelihoods across a set of 20 gene trees. We also ran Spinney with larger sets of sampled gene trees (100 or 1000) and with only four gene trees: those expected under the MSC model for each discordance level. In our setting, only the branch subtending species *A* and *B* has discordance, so there are only four gene trees, with weights determined by the MSC model. These expected sets are equivalent to a sufficiently large set of sampled gene trees, which we can see leads to slightly higher accuracy when there is a lot of discordance (Supporting Information Fig. S1). Rate estimates from expected gene trees closely matched those from sampled

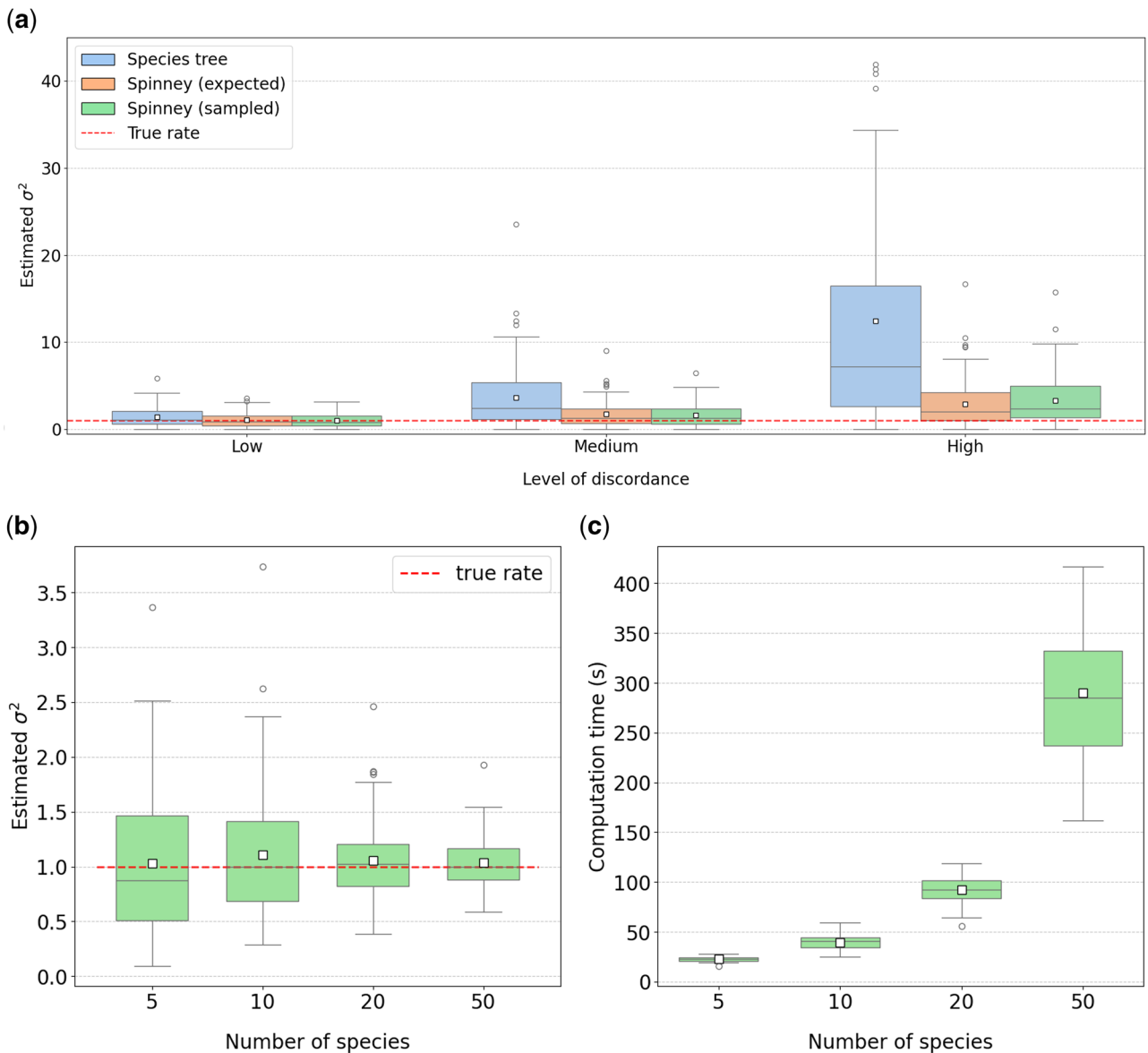


Figure 4. Rate estimates for the single-rate case. Rates were inferred using the species tree and Spinney. (a) Four-taxon settings under discordance levels of 1.6% (Low), 25% (Medium), and 50% (High). Spinney used both expected gene trees derived from the MSC (orange boxes) and sampled gene trees (green boxes). (b) Rates inferred by Spinney for species trees with 5, 10, 20, and 50 taxa, with 53%, 66%, 68% and 98% discordance, respectively. Spinney only used sampled gene trees in these calculations. (c) Computation time for the analyses in (b). Runtime increases with the number of species.

gene trees, consistently correcting overestimation. These also showed lower standard deviations and fewer outliers, as with sampled gene trees (Fig. 4a).

Generalization to larger species trees (5, 10, 20, and 50 taxa) revealed the same pattern: Spinney corrected overestimation relative to species-tree rates (Fig. 4b). For example, for the 10-taxon case, the mean rate estimates by the species tree and Spinney are 2.00 and 1.11 respectively. Although discordance levels frequently exceeded 50% in these larger trees, the degree of overestimation was less severe than in the four-taxon scenario. Unlike in the four-taxon trees, it is generally infeasible to enumerate all possible topologies under the MSC model for larger trees, as the number of topologies increases double-factorially with the number of taxa (Felsenstein 2004). Spinney must therefore rely on a finite sample of gene trees; however, even with moderate sample sizes (e.g. 20 gene trees), it produced accurate and stable estimates.

We recorded the computer wall-clock time for large trees. As expected, runtimes increased with the number of species. For 5, 10, and 20 taxa, each replicate finished in 28, 98, and 119 s respectively (Fig. 4c). For 50 taxa, wall-clock time ranged from 162 to 417 s. All optimizations ran on HPE Cray EX (SLES 15) CPU nodes (AMD EPYC 7742). Spinney used one node per condition, with 10 threads and 2 GB of memory.

Spinney reduces false inferences of rate variation

We are particularly interested in whether discordance could lead to incorrectly favouring a model with two rate parameters when there is more discordance in one part of a tree. Specifically, we predict that calculations using the species tree would favour the two-rate model because high discordance in (for instance) clade A will falsely lead to a distinct σ^2 compared to the rest of the tree (Fig. 3b). We also predict that Spinney will correct this overestimation by integrating discordant histories, thereby reducing the tendency to favour the two-rate model.

As expected, inference by the species tree overestimated rates. Starting with the one-rate model applied to the tree in Figure 3b, the mean σ_{st}^2 was 8.56, whereas the mean σ_{sy}^2 was 1.93. When estimating a two-rate model (only one rate was used to simulate these data), both methods inferred a higher rate for clade A and recovered an accurate rate for clade B. However, the species tree overestimated σ_A^2 by 18 \times , while Spinney was only 5 \times higher (Fig. 5; Supporting Information Table S3).

To test whether these overestimates would falsely lead to inferences of rate variation, we recorded the number of cases where the two-rate model was incorrectly favoured over the one-rate model via an LRT. We then compared these results across different values of σ^2 with 100 replicates per condition (Fig. 6). For $\sigma^2 = 0.1$, inferences using the species tree incorrectly favoured the two-rate model in 31 out of 100 replicates (i.e. a false positive rate of 31%), while Spinney did so in only 14 cases. The pattern was more pronounced at $\sigma^2 = 1$, where the species tree inference resulted in 48 incorrect selections (false positive rate (FPR) = 48%), compared to 15 for Spinney (FPR = 15%). At $\sigma^2 = 10$, the species tree inference incorrectly favoured the two-rate model in 33 cases, whereas Spinney did so in just 12 cases. These results suggest that Spinney effectively reduces the false positive rate by largely accounting for discordant histories.

The results above show that Spinney reduces the false positive rate, but it may do so simply by reducing the overall power of the LRT. To examine the power of both methods when there truly are two rates, we simulated a case in which clade A has a higher rate (i.e. σ_A^2 is higher) and also has 25% discordance. Both Spinney and the species tree have similar true positive rates (TPRs) for the two-rate model (Supporting Information Table S4). When clade A differed by 100 \times from clade B, TPRs were more than 90% for both methods. When clade A differed by 10 \times from clade B, TPRs dropped to 40%. We observed similar patterns when σ_A^2 was less than σ_B^2 (Supporting Information Table S4).

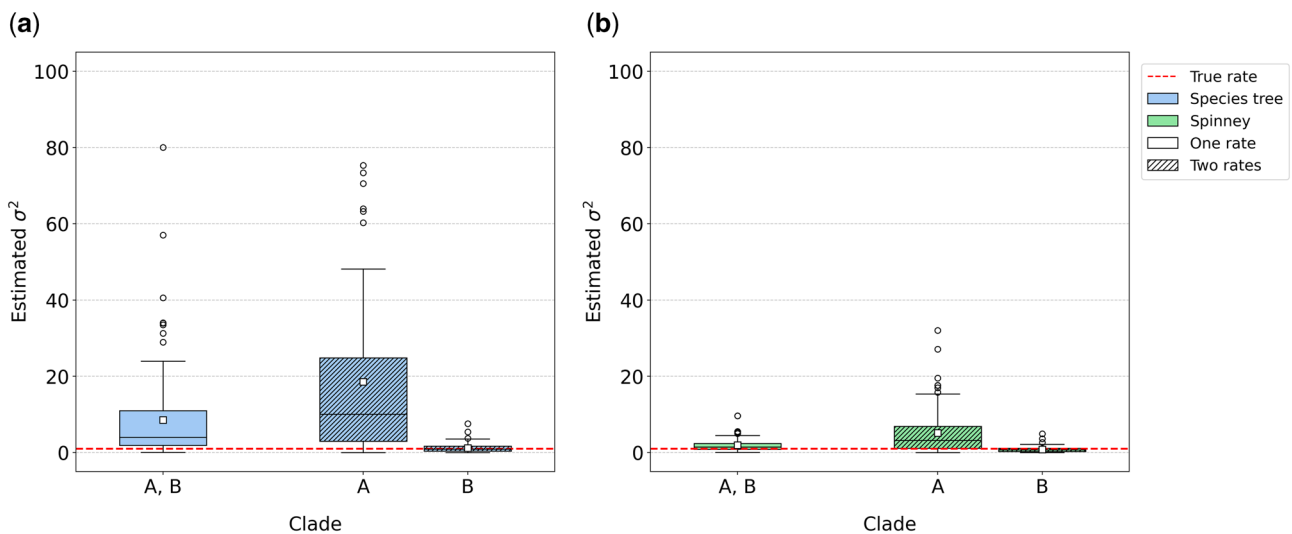


Figure 5. Rate estimates for the six-taxon case with only one rate simulated. Rates are inferred with (a) the species tree or (b) Spinney, with a one-rate model (plain boxes) or two-rate model (shaded boxes). Summary statistics are provided in Supporting Information Table S3.

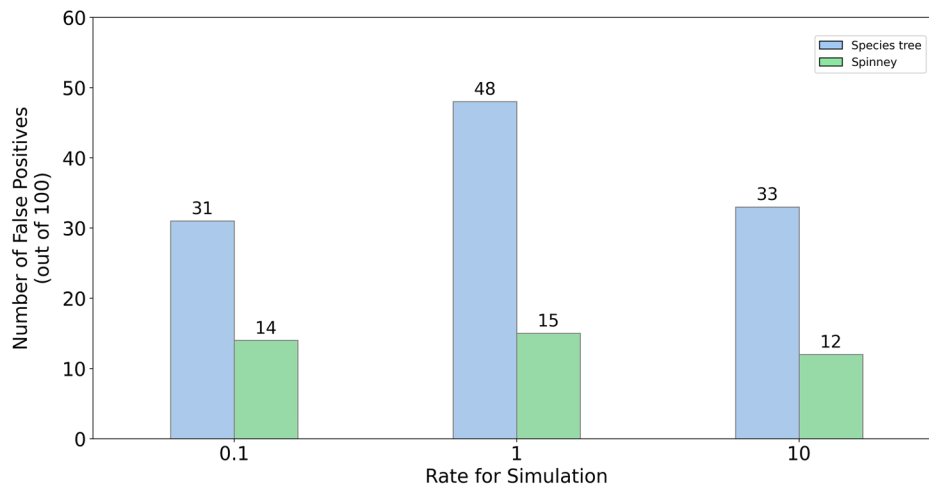


Figure 6. Number of cases in which the two-rate model was incorrectly selected by the LRT when one rate was simulated. The single simulated rate varies along the x -axis, and inference was performed for each condition using both the species tree and Spinney. Spinney reduces the FPR in all conditions, but does not eliminate it entirely.

Ancestral state reconstruction

We evaluated ancestral states across time, focusing on speciation event T_2 , which splits taxon C from taxa A and B (Fig. 3). Although there is only a single species tree topology, ILS generates four possible genealogical histories at T_2 (corresponding to the expected topologies in Fig. 4a).

Although the 95% CIs of ancestral states at time T_2 increased for both methods as ILS increased, inferences using Spinney produced relatively narrower values (Fig. 7a). For low discordance, the mean CI widths were 4.11 with a negligible difference from the species tree. For medium discordance, the means were 6.84 and 6.38 for the species tree and Spinney, respectively. For high discordance, the means were 12.82 and 7.40 for the species tree and Spinney, respectively ($P = 8.9 \times 10^{-18}$; Wilcoxon test). The main cause of the larger CIs using the species tree appears to be the higher (incorrect) value of σ^2 used by this method—higher rates mean more spread in the possible ancestral states.

We also compared species-tree and Spinney reconstructions of ancestral states across the whole species history, using a trait simulated under medium discordance (Fig. 7b, c). For $0 \leq t < T_1$, both methods behave similarly: the bright regions represent observed trait values that gradually spread under Brownian motion. At T_1 , the species tree assumes that species A and B merge instantly, producing an ancestral state centred between their trait values. In contrast, Spinney allows lineages to coalesce at different times across gene trees, so the spread of ancestral states persists longer into the past. These results indicate that Spinney captures varying coalescent histories across gene trees and thus reflects variation in genealogical histories across loci in a more realistic way. For $T_1 < t < T_2$, the species tree traces only a single, concentrated trajectory, while Spinney aggregates multiple coalescent histories and produces a broader, more uniform distribution of states. At T_2 , the species tree again merges species C instantaneously, generating a sharp transition with a wider distribution. Beyond T_2 , the species tree continues its diffusion process, whereas Spinney maintains a flatter, less concentrated distribution. In general, Spinney accounts for gene tree heterogeneity more correctly and conveys this in inferred ancestral states.

DISCUSSION

An increasingly common finding in phylogenomics is that a large fraction of gene trees do not match the species tree (e.g. Pollard *et al.* 2006, White *et al.* 2009, Jarvis *et al.* 2014, Copetti *et al.* 2017, Rouard *et al.* 2018, Hime *et al.* 2021, Morales-Briones *et al.* 2021). Despite the fact that discordant trees are located across the genome and can underlie important traits (e.g. Fontaine *et al.* 2015, Lamichhaney *et al.* 2016, Li *et al.* 2017, Pease *et al.* 2016, Han *et al.* 2017, Palesch *et al.* 2018, Wu *et al.* 2018, Hibbins *et al.* 2020, Urban *et al.* 2021, Feng *et al.* 2022), PCMs have largely relied on a single, ‘resolved’ species tree to carry out analyses (Hahn and Nakhleh 2016). The use of only a single tree topology when many different topologies exist can cause problems for comparative analyses, including in cases involving continuous traits (Mendes *et al.* 2018, Bastide *et al.* 2018, Hibbins and Hahn 2021, Hibbins *et al.* 2023, Schraiber *et al.* 2024, Adams *et al.* 2025).

Here, we have addressed one known issue when using only a single tree in the presence of discordance—that rates of evolution are overestimated overall—and have explored the related, but less-studied, problem of detecting shifts in rates on different parts of a tree. Our approach builds on the work of Hibbins *et al.* (2023), who applied the pruning algorithm to calculate trait likelihoods across a set of different tree topologies to account for discordance. While their initial work was limited to a simple three-taxon tree, here we have introduced a program, Spinney, to deal with larger species trees. The results presented above have demonstrated that Spinney reduces, but does not completely remove, the problems of discordance in a principled manner; it should therefore be useful to the research community. Nonetheless, we discuss several caveats, similar approaches, and future directions below.

The sample of gene trees

For trees with smaller numbers of tips, it is possible to exactly calculate every single relevant gene tree topology (and their weights) to use as input to Spinney. While this is not possible for trees with larger numbers of tips, it is really the number of phylogenetic ‘knots’ (cf. Ané *et al.* 2007) on the tree that determine how

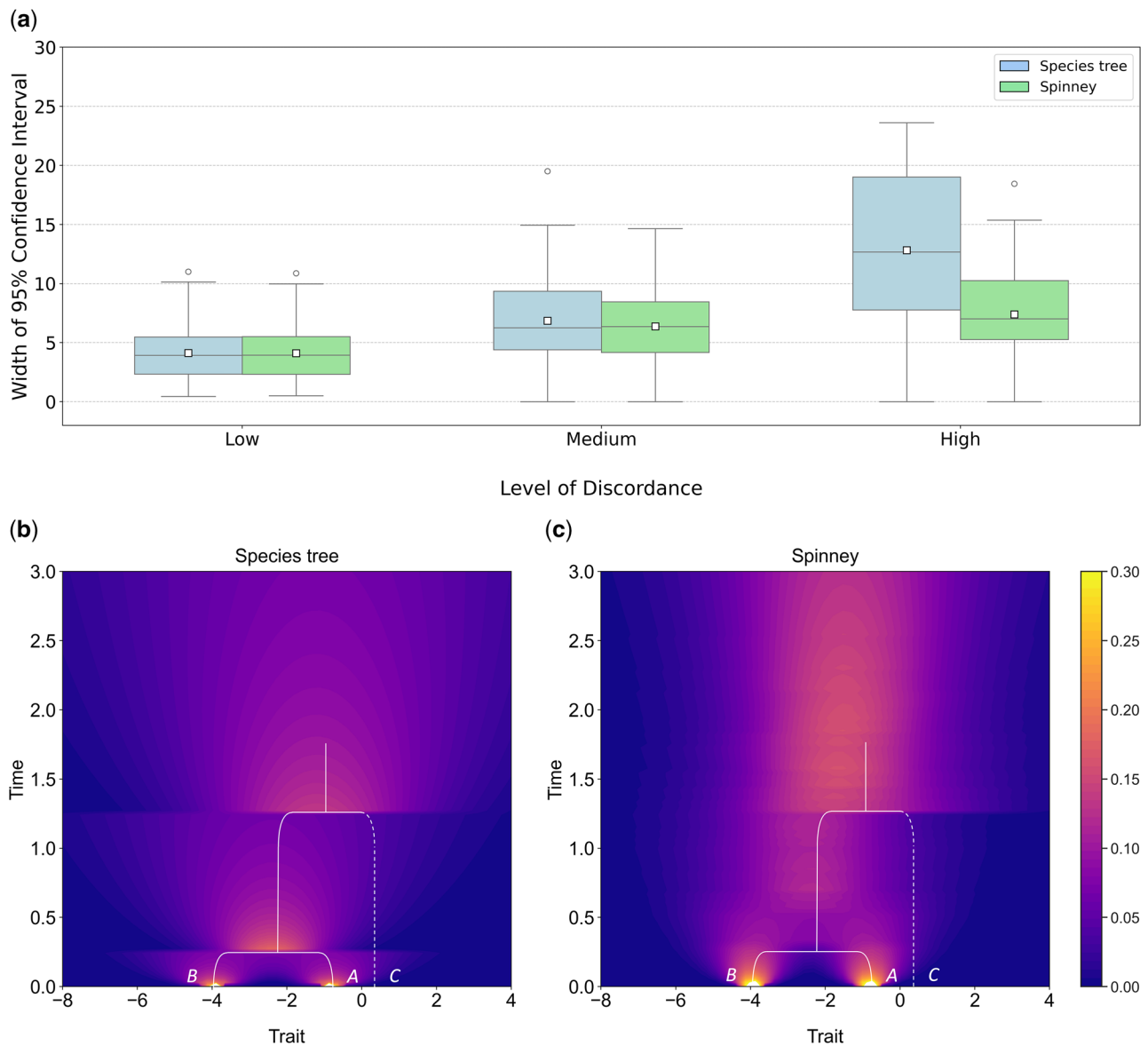


Figure 7. Ancestral states inferred using the species tree and Spinney. (a) The width of the 95% confidence interval at time T_2 . (b)-(c) Ancestral state maps of branches extending from taxa A and B in the species tree. Note that ancestral states for species C are not shown, which is why a dashed line is used (species D is omitted completely). Ancestral states were inferred with (b) the species tree and (c) Spinney. Colours encode the probability of the ancestral trait value in a specific time bin. The species trees are embedded inside the maps for clarity. Speciation times are $T_1 = 0.25$ and $T_2 = 1.25$. Species C contributes to ancestral states only for $t > 1.25$. The trait was simulated with 25% (medium) discordance. Trait values at the tips for taxa A, B, C, and D were -0.82 , -3.95 , 0.11 , and -3.50 , respectively.

many discordant topologies exist. For instance, a species tree with 100 tips might only show three topologies if ILS was restricted to a single internal branch. However, with even slightly more discordance, the number of possible gene trees becomes extremely large, requiring that we use only a sample of them as input.

While using more gene trees will always be more precise, this comes at a computational cost because likelihoods must be calculated on each tree (though this can be done in parallel). In our experiments, we used 20 sampled gene trees, which performed well for the species trees we considered, even though the number of possible gene trees was in the millions for some topologies (Fig. 4b). Although it may seem surprising in these cases that 20 trees are sufficient, two related aspects of discordance work in our

favour. First, many unique tree topologies have very low expected frequencies, and therefore contribute very little to observed trait data. As a result, not using these topologies has little effect on the results. Second, even when there are many phylogenetic knots spread across a species tree, each sampled gene tree provides information about each knot. That is, each gene tree used as input to Spinney can tell us about the dominant concordant and discordant topologies associated with each branch of the species tree, even if each unique tree topology has an individually low probability.

We have suggested two methods that can be used to sample gene trees. The first uses empirical gene trees taken from a dataset. The advantage of such an approach is that one can capture complex causes of discordance in the distribution of trees: for

instance, multiple overlapping introgression events. However, input gene trees must be ultrametric and in coalescent units, which is challenging to infer (we have previously found that gene tree inference error has only minor effects; Hibbins *et al.* 2023). An alternative approach is to infer a species tree whose branch lengths are in coalescent units (e.g. using the approach of Tabatabaee *et al.* 2023), and then to use this tree to simulate a sample of gene trees. While this will necessarily be a simplified species history, all sampled gene trees can be used directly. In both cases, an accurate species tree is still required for accurate inference. Furthermore, in both cases we also assume that our sample of gene trees reflects the set of gene trees at loci that actually contribute to the trait.

Testing for multiple rates

A common use of PCMs is to detect shifts in the rate of evolution along a species tree (O’Meara *et al.* 2006, Eastman *et al.* 2011, Rabosky *et al.* 2014, Revell 2021, Martin *et al.* 2023). As previous work had already shown that gene tree discordance could drive artefactual patterns of increased evolutionary rates, we reasoned that methods testing for multiple rates might also be misled when levels of discordance differ among clades.

While Hibbins *et al.* (2023) had to separately analyse parts of a larger tree to make such comparisons, Spinney makes it possible to directly test for multiple rates on a larger tree, while taking into account variable discordance. As shown in our results, Spinney greatly reduced the false positive rate relative to using the species tree alone, but kept the true positive rate on par with the species tree. Nevertheless, Spinney does not completely reduce false positives to their expected level under the null model. We are not sure why this occurs, nor why Spinney overestimates the rate associated with higher discordance, $\sigma^2_{A'}$ more in this scenario than in the single-rate simulations (see Fig. 5b). It may be that something about the shape of the species tree or the number of tips makes it harder to correct for the effect of discordance.

Similar approaches

There are a few similar approaches to the one used here, for both continuous and discrete traits. For continuous traits, the method of Osmond and Coop (2024) applied to population-level data on ancestral ranges is a clear antecedent. In their original approach, a large number of gene trees from across the genome are used to calculate the dispersal rate and ancestral locations of individuals back in time; this was further extended to the full ancestral recombination graph in Deraje *et al.* (2025). Osmond and Coop’s approach to calculating dispersal rates is very similar to ours, maximizing the likelihood across trees. Although the resulting estimates are slightly biased, this is probably due to the fact that the underlying model of uniform dispersal does not match actual geographical dispersal routes. Interestingly, for the inference of ancestral locations, Osmond and Coop allow each tree to find its own maximum-likelihood value. This highlights a difference in fine-grained population genetic data, where each locus really might have existed in a distinct ancestor living in a distinct geographical location.

A second similar approach, applied to DNA or amino acid sequence data, is the “mixtures across sites and trees” (MAST) model of Wong *et al.* (2024). In this model, site-likelihoods are calculated under a limited set of alternative tree topologies. The

total likelihood of an alignment is maximized over these tree topologies (inferring branch lengths simultaneously) and tree weights. This is very similar to the approach taken by Spinney, with one important distinction: the default setting in MAST also estimates the tree weights from the data. While these weights are not equivalent to gene tree frequencies, they may be highly correlated with these frequencies.

Future directions

The approach we take here calculates likelihoods on trees using the pruning algorithm (Felsenstein 1973). Although sometimes overlooked as an approach to PCMs, it has found wide use in a number of programs (e.g. Hahn *et al.* 2005, FitzJohn 2012, Ho and Ané 2014, Uyeda and Harmon 2014, Hiscott *et al.* 2016, Mitov *et al.* 2020, Bertram *et al.* 2023, Martin *et al.* 2023). The pruning algorithm has many advantages over matrix approaches, including the ability to analyse very large trees (e.g. Freckleton 2012), to easily implement a variety of different (non-Brownian motion) models of trait evolution (e.g. Boucher and Démercy 2016), and to take into account measurement error (e.g. Han *et al.* 2013). This last feature is extremely important, as measurement error can have drastic consequences for comparative analyses (e.g. Cooper *et al.* 2016, Beaulieu and O’Meara 2025). In the future, we can imagine combining many of these features with calculations across sets of gene trees—even gene trees possibly inferred in error—to allow for more accurate evolutionary inferences.

ACKNOWLEDGEMENTS

We thank Mark Hibbins for very helpful advice early in this project.

AUTHOR CONTRIBUTIONS

Yu Mo (Data curation [Equal], Formal analysis [Equal], Software [Lead], Visualization [Lead], Writing—original draft [Equal], Writing—review & editing [Equal]), Matthew W. Hahn (Conceptualization [Equal], Funding acquisition [Lead], Project administration [Lead], Supervision [Lead], Writing—original draft [Equal], Writing—review & editing [Equal]).

SUPPLEMENTARY DATA

Supplementary data are available at *Evolutionary Journal of the Linnean Society* online.

CONFLICT OF INTEREST

None declared.

FUNDING

This research was funded by National Science Foundation grant DBI-2146866.

DATA AVAILABILITY

The data underlying this article are available in GitHub.

REFERENCES

- Adams DC. Quantifying and comparing phylogenetic evolutionary rates for shape and other high-dimensional phenotypic data. *Systematic Biology* 2014;**63**:166–77.
- Adams R, Lozano JR, Duncan M *et al*. A tale of too many trees: a conundrum for phylogenetic regression. *Molecular Biology and Evolution* 2025;**42**:msaf032.
- Ané C, Larget B, Baum DA *et al*. Bayesian estimation of concordance among gene trees. *Molecular Biology and Evolution* 2007;**24**:412–26.
- Bastide P, Solís-Lemus C, Kriebel R *et al*. Phylogenetic comparative methods on phylogenetic networks with reticulations. *Systematic Biology* 2018;**67**:800–20.
- Beaulieu JM, O’Meara BC. Navigating “tip fog”: embracing uncertainty in tip measurements. *Evolution* 2025;**79**:1131–42.
- Bertram J, Fulton B, Tourigny JP *et al*. CAGEE: computational analysis of gene expression evolution. *Molecular Biology and Evolution* 2023;**40**:msad106.
- Blomberg SP, Garland J, Ives AR. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 2003;**57**:717–45.
- Boucher FC, Démetry V. Inferring bounded evolution in phenotypic characters from phylogenetic comparative data. *Systematic Biology* 2016;**65**:651–61.
- Cooper N, Thomas GH, Venditti C *et al*. A cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies. *Biological Journal of the Linnean Society* 2016;**118**:64–77.
- Copetti D, Búrquez A, Bustamante E *et al*. Extensive gene tree discordance and hemiplasy shaped the genomes of North American columnar cacti. *Proceedings of the National Academy of Sciences* 2017;**114**:12003–8.
- Degnan JH, Rosenberg NA. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution* 2009;**24**:332–40.
- Deraje P, Kitchens J, Coop G *et al*. The promise and challenge of spatial inference with the full ancestral recombination graph under Brownian motion. *Genes|Genomes|Genetics* 2025;**15**:jkaf214.
- Eastman JM, Alfaro ME, Joyce P *et al*. A novel comparative method for identifying shifts in the rate of character evolution on trees. *Evolution* 2011;**65**:3578–89.
- Felsenstein J. Maximum-likelihood estimation of evolutionary trees from continuous characters. *American Journal of Human Genetics* 1973;**25**:471–92.
- Felsenstein J. Phylogenies and the comparative method. *The American Naturalist* 1985;**125**:1–15.
- Felsenstein J. *Inferring Phylogenies*. Sunderland, MA: Sinauer Associates, 2004.
- Feng S, Bai M, Rivas-González I *et al*. Incomplete lineage sorting and phenotypic evolution in marsupials. *Cell* 2022;**185**:1646–60.e18.
- FitzJohn RG. Diversitree: comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution* 2012;**3**:1084–92.
- Fontaine MC, Pease JB, Steele A *et al*. Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science (New York, N.Y.)* 2015;**347**:1258524.
- Freckleton RP. Fast likelihood calculations for comparative analyses. *Methods in Ecology and Evolution* 2012;**3**:940–7.
- Gillespie JH, Langley CH. Are evolutionary rates really variable? *Journal of Molecular Evolution* 1979;**13**:27–34.
- Hahn MW, De Bie T, Stajich JE *et al*. Estimating the tempo and mode of gene family evolution from comparative genomic data. *Genome Research* 2005;**15**:1153–60.
- Hahn MW, Nakhleh L. Irrational exuberance for resolved species trees. *Evolution* 2016;**70**:7–17.
- Han F, Lamichhaney S, Grant BR *et al*. Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin’s finches. *Genome Research* 2017;**27**:1004–15.
- Han MV, Thomas GW, Lugo-Martinez J *et al*. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Molecular Biology and Evolution* 2013;**30**:1987–97.
- Hansen TF. Stabilizing selection and the comparative analysis of adaptation. *Evolution* 1997;**51**:1341–51.
- Harris CR, Millman KJ, Van Der Walt SJ *et al*. Array programming with NumPy. *Nature* 2020;**585**:357–62.
- Helmstetter AJ, Zenil-Ferguson R, Sauquet H *et al*. Trait-dependent diversification in angiosperms: patterns, models and data. *Ecology Letters* 2023;**26**:640–57.
- Hibbins MS, Breithaupt LC, Hahn MW. Phylogenomic comparative methods: accurate evolutionary inferences in the presence of gene tree discordance. *Proceedings of the National Academy of Sciences of the United States of America* 2023;**120**:e2220389120.
- Hibbins MS, Gibson MJ, Hahn MW. Determining the probability of hemiplasy in the presence of incomplete lineage sorting and introgression. *eLife* 2020;**9**:e63753.
- Hibbins MS, Hahn MW. The effects of introgression across thousands of quantitative traits revealed by gene expression in wild tomatoes. *PLoS Genetics* 2021;**17**:e1009892.
- Hime PM, Lemmon AR, Lemmon ECM *et al*. Phylogenomics reveals ancient gene tree discordance in the amphibian tree of life. *Systematic Biology* 2021;**70**:49–66.
- Hiscott G, Fox C, Parry M *et al*. Efficient recycled algorithms for quantitative trait models on phylogenies. *Genome Biology and Evolution* 2016;**8**:1338–50.
- Ho L, Ané C. A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Systematic Biology* 2014;**63**:397–408.
- Hudson RR. Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics* 2002;**18**:337–8.
- Jarvis ED, Mirarab S, Aberer AJ *et al*. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science (New York, N.Y.)* 2014;**346**:1320–31.
- Khabbazian M, Kriebel R, Rohe K *et al*. Fast and accurate detection of evolutionary shifts in Ornstein–Uhlenbeck models. *Methods in Ecology and Evolution* 2016;**7**:811–24.
- Lamichhaney S, Han F, Berglund J *et al*. A beak size locus in Darwin’s finches facilitated character displacement during a drought. *Science (New York, N.Y.)* 2016;**352**:470–4.
- Li YY, Liu Z, Qi FY *et al*. Functional effects of a retained ancestral polymorphism in *Prestin*. *Molecular Biology and Evolution* 2017;**34**:88–92.
- Martin BS, Bradburd GS, Harmon LJ *et al*. Modeling the evolution of rates of continuous trait evolution. *Systematic Biology* 2023;**72**:590–605.
- Martins EP. Conducting phylogenetic comparative studies when the phylogeny is not known. *Evolution* 1996;**50**:12–22.
- Mendes FK, Fuentes-González JA, Schraiber JG *et al*. A multispecies coalescent model for quantitative traits. *eLife* 2018;**7**:e36482.
- Mitov V, Bartoszek K, Asimomitis G *et al*. Fast likelihood calculation for multivariate Gaussian phylogenetic models with shifts. *Theoretical Population Biology* 2020;**131**:66–78.
- Morales-Briones DF, Kadereit G, Tefarikis DT *et al*. Disentangling sources of gene tree discordance in phylogenomic data sets: testing ancient hybridizations in *Amaranthaceae* s.l. *Systematic Biology* 2021;**70**:219–35.
- Moreno MA, Holder MT, Sukumaran J. DendroPy 5: a mature Python library for phylogenetic computing. *Journal of Open Source Software* 2024;**9**:6943.
- O’Meara BC, Ané C, Sanderson MJ *et al*. Testing for different rates of continuous trait evolution using likelihood. *Evolution* 2006;**60**:922–33.
- Ogilvie HA, Bouckaert RR, Drummond AJ. StarBEAST2 brings faster species tree inference and accurate estimates of substitution rates. *Molecular Biology and Evolution* 2017;**34**:2101–14.
- Osmond M, Coop G. Estimating dispersal rates and locating genetic ancestors with genome-wide genealogies. *eLife* 2024;**13**:e72177.
- Pagel M. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Systematic Biology* 1999;**48**:612–22.
- Palesch D, Bosinger SE, Tharp GK *et al*. Sooty mangabey genome sequence provides insight into AIDS resistance in a natural SIV host. *Nature* 2018;**553**:77–81.
- Paradis E, Schliep K. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics (Oxford, England)* 2019;**35**:526–8.

- Pease JB, Haak DC, Hahn MW *et al.* Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLoS Biology* 2016;**14**:e1002379.
- Pollard DA, Iyer VN, Moses AM *et al.* Widespread discordance of gene trees with species tree in *Drosophila*: evidence for incomplete lineage sorting. *PLoS Genetics* 2006;**2**:e173.
- Rabosky DL, Donnellan SC, Grundler M *et al.* Analysis and visualization of complex macroevolutionary dynamics: an example from Australian scincid lizards. *Systematic Biology* 2014;**63**:610–27.
- Rafferty NE, Nabity PD. A global test for phylogenetic signal in shifts in flowering time under climate change. *Journal of Ecology* 2017;**105**:627–33.
- Revell LJ. A variable-rate quantitative trait evolution model using penalized-likelihood. *PeerJ* 2021;**9**:e11997.
- Rouard M, Droc G, Martin G *et al.* Three new genome assemblies support a rapid radiation in *Musa acuminata* (wild banana). *Genome Biology and Evolution* 2018;**10**:3129–40.
- Schluter D, Price T, Mooers AO *et al.* Likelihood of ancestor states in adaptive radiation. *Evolution* 1997;**51**:1699–711.
- Schraiber JG, Edge MD, Pennell M. Unifying approaches from statistical genetics and phylogenetics for mapping phenotypes in structured populations. *PLoS Biology* 2024;**22**:e3002847.
- Tabatabaee Y, Zhang C, Warnow T, Mirarab S. Phylogenomic branch length estimation using quartets. *Bioinformatics* 2023;**39** (Supplement_1): i185–193.
- Urban S, Nater A, Meyer A *et al.* Different sources of allelic variation drove repeated color pattern divergence in cichlid fishes. *Molecular Biology and Evolution* 2021;**38**:465–77.
- Uyeda JC, Hansen TF, Arnold SJ *et al.* The million-year wait for macroevolutionary bursts. *Proceedings of the National Academy of Sciences* 2011;**108**:15908–13.
- Uyeda JC, Harmon LJ. A novel Bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. *Systematic Biology* 2014;**63**:902–18.
- Virtanen P, Gommers R, Oliphant TE; SciPy 1.0 Contributors *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* 2020;**17**:261–72.
- White MA, Ané C, Dewey CN *et al.* Fine-scale phylogenetic discordance across the house mouse genome. *PLoS Genetics* 2009;**5**:e1000729.
- Wong TK, Cherryh C, Rodrigo AG *et al.* MAST: phylogenetic inference with mixtures across sites and trees. *Systematic Biology* 2024;**73**:375–91.
- Wu M, Kostyun JL, Hahn MW *et al.* Dissecting the basis of novel trait evolution in a radiation with widespread phylogenetic discordance. *Molecular Ecology* 2018;**27**:3301–16.