Phylogenomic comparative methods: accurate evolutionary inferences in the presence of gene tree discordance

Mark S. Hibbins[1,2], Lara C. Breithaupt[2,3], and Matthew W. Hahn[2, 4]

[1]Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON, CA M5S 3B2

[2]Department of Biology and [4]Department of Computer Science, Indiana University, Bloomington, IN 47405

[3]Department of Computer Science, Duke University, Durham, NC 27710

Abstract

Phylogenetic comparative methods have long been a mainstay of evolutionary biology, allowing for the study of trait evolution across species while accounting for their common ancestry. These analyses typically assume a single, bifurcating phylogenetic tree describing the shared history among species. However, modern phylogenomic analyses have shown that genomes are often composed of mosaic histories that can disagree both with the species tree and with each other—so-called discordant gene trees. These gene trees describe shared histories that are not captured by the species tree, and therefore that are unaccounted for in classic comparative approaches. The application of standard comparative methods to species histories containing discordance leads to incorrect inferences about the timing, direction, and rate of evolution. Here, we develop two approaches for incorporating gene tree histories into comparative methods: one that constructs an updated phylogenetic variance-covariance matrix from gene trees, and another that applies Felsenstein's pruning algorithm over a set of gene trees to calculate trait histories and likelihoods. Using simulation, we demonstrate that our new approaches generate much more accurate estimates of tree-wide rates of trait evolution than standard methods. We apply our methods to two clades of the wild tomato genus *Solanum* with varying rates of discordance, demonstrating the contribution of gene tree discordance to variation in a set of floral traits. Our new approaches have the potential to be applied to a broad range of classic inference problems in phylogenetics, including ancestral state reconstruction and the inference of lineage-specific rate shifts.

Significance statement

Phylogenetic comparative methods allow for the study of trait evolution between species by accounting for their shared evolutionary history. These methods usually assume that species relationships can be described by a single tree. However, different parts of the genome can have their own independent evolutionary histories that can disagree with each other. If these disagreeing histories contribute to trait evolution over time, standard comparative methods can be misled. In this work, we developed two new approaches to phylogenetic comparative methods that account for this variation in histories across the genome. We used these methods to estimate more accurate rates of floral trait evolution in wild tomatoes. Our work opens new approaches for the study of trait evolution among species.

Introduction

A major goal of evolutionary biology is to understand how and why traits vary among species. One of the major sources of this variation is common ancestry. If left unaccounted for, this shared history can lead to pseudoreplication and spurious trait correlations (Felsenstein 1985). Phylogenetic comparative methods have been developed to account for shared history, enabling more accurate inferences about the tempo and mode of trait evolution (Harvey and Pagel 1991). With the statistical toolkit offered by phylogenetic comparative methods, researchers can ask questions about the rate at which traits evolve, whether these rates have changed over time or in different lineages, what traits may have looked like in ancestral or extinct lineages, and whether trait shifts are correlated with historical or environmental factors (Martins and Hansen 1996; Garamszegi 2014; Adams and Collyer 2018; Revell and Harmon 2022).

In classic comparative methods, common ancestry among species is accounted for by using a single species phylogeny. However, genome-scale analyses of phylogenetic history have revealed that individual loci can have their own independent histories (Pollard et al. 2006; White et al. 2009; Fontaine et al. 2015; Pease et al. 2016; Novikova et al. 2016; Copetti et al. 2017; Wu et al. 2018; Edelman et al. 2019; Vanderpool et al. 2020). The result is gene tree discordance—the disagreement of trees at individual loci both with each other and with the species phylogeny. Gene tree discordance has important implications for phylogenetic comparative methods because discordant gene trees contain branches that are not present in the species phylogeny. Evolution along such discordant branches can result in trait similarity among species with no shared history in the species tree (Figure 1). Such patterns of trait variation can mislead standard phylogenetic comparative methods, particularly by resulting in overestimates of the number of trait transitions or the rate of trait evolution (Hahn and Nakhleh 2016; Mendes and Hahn 2016; Mendes et al. 2018; Hibbins et al. 2020; Wang et al. 2021). This effect has been termed "hemiplasy," as single transitions on discordant gene trees can falsely resemble homoplasy when analyzed on the species tree (Avise and Robinson 2008).

Discordance is a concern for evolutionary inference because it has biological causes that cannot be overcome by addressing technical errors or by increasing species sampling (Degnan and Rosenberg 2009). Two primary causes of discordance, incomplete lineage sorting (ILS) and introgression, have different effects on gene tree frequencies and branch lengths and are therefore expected to bias comparative methods in different ways. ILS, a stochastic process that depends on species tree internal branch lengths and population sizes, generates symmetry in the frequencies of possible discordant gene trees (Hudson 1983; Pamilo and Nei 1988). Therefore, higher amounts of ILS lead to broad increases in the occurrence of hemiplasy across multiple possible incongruent trait patterns (Guerrero and Hahn 2018; Mendes et al. 2018). Introgression is a process of historical hybridization and back-crossing that, while widespread in modern phylogenomic datasets (Mallet et al. 2016; Taylor and Larson 2019; Dagilis et al. 2022), is often more limited to specific pairs of taxa. In particular, post-speciation introgression between non-sister lineages leads to an excess of gene trees grouping those lineages as sister (Reich et al. 2009; Green et al. 2010; Durand et al. 2011; Patterson et al. 2012). This pattern should result in an excess of trait-sharing for the species involved in introgression compared to the species not exchanging genes (Hibbins et al. 2020; Hibbins and Hahn 2021; Wang et al. 2021).

While some progress has been made in accounting for discordance in the evolution of discrete traits, especially in nucleotide models (De Maio et al. 2013; De Maio et al. 2015; Schrempf et al. 2016; Ogilvie et al. 2017; Schrempf et al. 2019), many classic phylogenetic comparative methods remain unable to account for gene tree discordance when analyzing quantitative traits. The approaches required to improve these methods will depend on the question being asked. Some tasks, such as maximum-likelihood estimation of the rate of trait evolution under Brownian motion (σ^2) (e.g. Garland and Ives 2000; O'Meara et al. 2006) or phylogenetic regression (Grafen 1989), depend on the specification of a matrix that describes the trait variances and covariances expected from the species phylogeny (often denoted $\boldsymbol{C}$). Other comparative approaches, such as ancestral state reconstruction (Pagel 1999) and inference of lineage-specific rate shifts (Alfaro et al. 2009), can require more sophisticated approaches that calculate state probabilities on different parts of a phylogeny; one such approach is to use Felsenstein's pruning algorithm applied to a species tree with specified branch lengths (Felsenstein 1973). Mendes et al. (2018) showed that failing to account for discordance can bias estimates of σ^2 upwards and can lead to falsely inflated numbers of trait-mean transitions. In general, the development of a comparative framework incorporating gene tree discordance would lead to more accurate evolutionary inferences in a wide variety of systems with ILS and/or introgression, across a wide variety of approaches for making inferences about quantitative traits.

Here, we demonstrate the utility of an updated *phylogenomic* comparative framework, using two distinct approaches to incorporate the summed history of concordant and discordant gene trees into evolutionary inference. In the first approach, we show how to construct an updated phylogenetic variance/covariance matrix (which we denote $\boldsymbol{C}^*$) to include the covariances introduced by discordant gene trees. We provide a new R package, *seastaR*, that can construct this updated matrix for any number of species, either by summing the internal branches of an input set of gene trees or by calculating expected gene tree internal branches from an input species tree using the multispecies coalescent model. We show how estimates of the evolutionary rate are made more accurate by using $\boldsymbol{C}^*$, and suggest how this updated matrix can be passed to other available software packages to make multiple evolutionary inferences more robust to discordance. In the second approach, we develop a method for applying the pruning algorithm over a set of gene trees to return the likelihood of an observed trait across species. Using a pilot implementation of this approach for a rooted three-species tree, we show how it can be used to accurately estimate the rate of quantitative trait evolution. Although currently limited to a smaller number of species, this latter approach has the potential to perform more complicated comparative inferences in the presence of discordance. Finally, we apply our approaches to empirical morphological data from wild tomatoes (Haak et al. 2014), finding a greater discrepancy between species tree and gene tree rate estimates in a clade with a higher rate of gene tree discordance. Overall, our new approaches pave the way towards more accurate evolutionary inferences in the presence of gene tree discordance.

Methods

Building a phylogenetic variance/covariance matrix from data with discordance

As previously discussed, one of the most common ways that phylogeny is incorporated into comparative analyses is by constructing a phylogenetic variance/covariance matrix, $\boldsymbol{C}$. This

square matrix has rows and columns corresponding to the number of taxa in the phylogeny, with the diagonal elements containing the expected trait variances for each species and the off-diagonal elements containing the expected trait covariances between each species pair. Considering three species with the relationship ((A,B),C) (Figure 2A), the standard covariance matrix has the following form:

$$\boldsymbol{C} = \begin{bmatrix} Var(A) & Cov(AB) & 0 \\ Cov(BA) & Var(B) & 0 \\ 0 & 0 & Var(C) \end{bmatrix} \quad \textit{Eq 1}$$

Trait covariances arise from shared internal branches in the phylogeny. As only species A and B share an internal branch in the species tree, the other two species pairs have no expected covariance.

In contrast, if we consider the gene trees that are generated by the species tree in Figure 2A, the two discordant gene trees contain internal branches shared by pairs B-C and A-C. Discordance due to ILS generates all three possible topologies for this species tree, so all off-diagonal entries in the covariance matrix should have non-zero values (Mendes et al. 2018). We are interested in estimating this updated covariance matrix, which we denote ***C*:***

$$\boldsymbol{C}^* = \begin{bmatrix} Var(A) & Cov(AB) & Cov(AC) \\ Cov(BA) & Var(B) & Cov(BC) \\ Cov(CA) & Cov(CB) & Var(C) \end{bmatrix} \quad \textit{Eq 2}$$

To construct ***C****, we provide the R package *seastaR*. *seastaR* uses two approaches for estimating ***C****, both following the same principle: each gene tree topology contributes an internal branch which, after being weighted by that tree's expected frequency, fills an off-diagonal entry in the covariance matrix (Figure 2A). Each gene tree also contributes its total height, weighted by frequency, to the expected trait variances for each species. Both approaches assume that each individual gene tree contributes equally to trait variation among species (i.e. the effect size of mutations that affect trait variation does not differ on average among loci). We also assume that loci contributing to trait variation follow the same distribution of tree topologies as the genome at large, so the specified loci do not have to be explicitly related to the trait in question.

The first approach for estimating ***C****, *trees_to_vcv*, constructs this matrix from a list of provided gene trees (with branch lengths) and their observed frequencies. The method works by obtaining all the internal branch lengths present in each gene tree, as well as the height of each gene tree, and averaging them to get ***C****. A major advantage of this approach is that it can easily account for both ILS and introgression as sources of gene tree discordance, as the effects of both are captured in the distribution of observed gene tree topologies and branch lengths. On the other hand, individual gene trees may be inferred with error, making their branch lengths and frequencies less reliable. If accurately estimated gene trees are unavailable, our second approach, *get_full_matrix*, constructs ***C**** solely from an input species tree in coalescent units. This method breaks the input phylogeny down into each possible triplet, and for each triplet uses expectations from the multispecies coalescent model to calculate the expected internal branches and frequencies for each possible gene tree (see Mendes et al. 2018). For an exemplar five-taxon tree specified in

coalescent units, we compared the standard $\boldsymbol{C}$ matrix to a $\boldsymbol{C}^*$ matrix computed using *get_full_matrix* (Figure 2B). The test tree has three internal branches, each of length of 0.1 coalescent units. Given these branch lengths, we expect 60% of trees to be discordant for each of these three branches, meaning that only 40% of gene trees will have (for instance) the clade containing species 5 and 4 sister to the clade containing species 3 and 2. As expected, $\boldsymbol{C}^*$ contains covariance entries for species pairs that do not share an internal branch in the species tree, but that share internal branches in at least one discordant gene tree (Figure 2B). In addition, the sister lineages in the species tree have smaller covariances in $\boldsymbol{C}^*$ than in $\boldsymbol{C}$, because they do not share an internal branch in many discordant trees.

Our package, *seastaR*, contains several other utilities, including a parser for an input set of estimated gene trees, a simulator that can simulate trait evolution using $\boldsymbol{C}^*$, and a function to obtain the maximum-likelihood estimate of σ^2 using $\boldsymbol{C}^*$ (see Results). Also note that, although not currently implemented, *seastaR* could be extended to construct $\boldsymbol{C}^*$ from an input species network specified in coalescent units, using expectations from the multispecies network coalescent model (Hibbins and Hahn 2021).

Calculating trait likelihoods over a set of gene trees using Felsenstein's pruning algorithm

Updating the phylogenetic variance/covariance matrix provides a straightforward solution to accounting for gene tree discordance that works for several important inference tasks in comparative methods. However, many questions require more sophisticated models that do not have straightforward solutions making use of this matrix. For these questions, the field would benefit from a general approach to calculating likelihoods given a set of gene trees and a model of trait evolution. Our solution makes use of Felsenstein's pruning algorithm (Felsenstein 1973), a dynamic programming algorithm that calculates probabilities for a set of character states across all nodes in a phylogeny. A tree-wide likelihood can be calculated from the probabilities at the root, which can be used in conjunction with numerical optimization methods to estimate model parameters.

We developed an approach to apply the pruning algorithm to a specified set of gene trees, rather than to a single tree. This approach is implemented in C++ and draws heavily on the infrastructure of *CAFE* (Hahn et al. 2005, Mendes et al. 2020), a program that uses the pruning algorithm to calculate likelihoods for a birth-death model of gene family evolution. We make several modifications based on the methods presented in Boucher and Démery (2016) and implemented in *CAGEE* (https://github.com/hahnlab/CAGEE) that allow *CAFE*'s implementation of the pruning algorithm to be applied to continuous traits rather than integer counts of gene families (see also Bertram et al. 2022). First, the pruning algorithm requires a vector of possible discrete character states over which probabilities can be calculated. To obtain this vector from an observed continuous trait, we take the range $(-2(max(|\boldsymbol{X}|), 2(max(|\boldsymbol{X}|))$ where $\boldsymbol{X}$ is the vector of observed characters for each species. The vector of character states is then filled with 100 equidistant steps from the lower bound to the higher bound. Second, we need to assign probabilities to all the character states at the tips of the phylogeny, so that the pruning algorithm has a place to start. This is straightforward for integer count data, as the observed value can simply be assigned a probability of 1 at the tip. However, for continuous traits it will often be the case that none of the values in our discretized trait vector exactly match the observed values at the tips. Therefore, we

implement an approach that distributes the probability at the tip over the two states in the discretized vector closest to each of the observed values, proportional to how distant they are from the observed value (Equation 18 in the appendix of Boucher and Démery 2016). Third, to calculate the transition probability between each pair of discretized trait values over a branch in the phylogeny, we use the Brownian motion model. The probability density for Brownian motion is:

$$p(x, x_0, t) = \frac{1}{\sqrt{2\pi t}\sigma} e^{-\frac{(x-x_0)^2}{2\sigma^2 t}}$$

Eq 3

where x_0 is the initial trait value, x is the trait value after time t, and σ^2 is the evolutionary rate per unit time. With these methods in place, we can apply the standard pruning algorithm to an individual tree with observed character states and a specified σ^2 value.

To estimate a single likelihood over a set of gene trees, we initially apply the standard pruning algorithm to each gene tree individually. Like the covariance matrix approach, we assume that individual loci contribute equally to trait variation among species, and that trait loci follow the same distribution of tree topologies as the genome at large. These gene trees with branch lengths are given to the method directly, and must be ultrametric. Any set of trees can be specified, but the manner in which they are specified will depend on the size of the species tree (i.e. number of tips). For a large species tree, individual gene trees can be inferred or predicted by theory, similarly to the two approaches used by *seastaR*. Because it may not be possible to sample every possible topology, we recommend sampling a reasonable number of individual gene trees (see Discussion). For a small species tree, the most efficient approach will be to specify one tree for each possible topology, along with its frequency. Again, the branch lengths and frequencies of each tree topology can be averaged from a set of inferred trees or predicted from theory. The total likelihood is then calculated as:

$$L = \sum_{\tau} f(\tau_i) \left(-log \left(max(\boldsymbol{p}_{\boldsymbol{\tau}_i}) \right) \right)$$

Eq 4

where τ is the set of gene trees, $f(\tau_i)$ is the frequency of gene tree *i,* and $\boldsymbol{p}_{\boldsymbol{\tau}_i}$ is the vector of character state probabilities at the root for gene tree *i*. In words: we obtain a partial negative log-likelihood for each individual gene tree, these partial likelihoods are then weighted by each gene tree's observed frequency, and finally the weighted partial likelihoods are summed together to produce the total likelihood (Figure 3).

A major advantage of the pruning algorithm method is that maximum-likelihood inference can be used to estimate parameters for a wide variety of models. In addition, like the *trees_to_vcv* method of *seastaR*, this approach can easily handle introgression events if the signals of introgression are contained in the specified gene trees, or if expected gene trees under the multispecies network coalescent could be specified by the user. Currently, our implementation uses the Nelder-Mead algorithm (Nelder and Mead 1965) to find the optimal Brownian motion evolutionary rate parameter, σ^2. In the future, we would like the software to also perform more

sophisticated inferences, such as ancestral state reconstruction or lineage-specific rate shifts. Our approach could also be extended to any evolutionary model, not just Brownian motion.

Simulating complex traits with discordance

To demonstrate the utility of our new phylogenomic comparative approaches, we used simulations to evaluate their performance on a simple inference task: estimating the evolutionary rate parameter, σ^2. We simulated traits from a phylogenetic history with increasing rates of gene tree discordance by making random draws from a multivariate normal distribution (where $\boldsymbol{C^*}$ specifies the covariance structure). This simulation approach assumes an infinitesimal contribution to the trait from all genomic loci, an approximation that holds reasonably well for many complex quantitative traits. For each simulated dataset, we applied both standard inference of σ^2 using the species tree and our updated inferences that account for gene tree discordance. See the Supplementary Methods for the exact conditions and parameters used in our simulations.

We simulated our traits under the model parameterization of Mendes et al. (2018). Levels of discordance in this model are altered by changing the effective population size, *N,* allowing us to increase the level of discordance by increasing *N*. The equations for trait variances and covariances are also scaled by *N*, such that branch lengths in units of absolute time are divided by 2*N*. The evolutionary rate is a compound parameter, $2N\mu\sigma_M^2$, where μ is the mutation rate and σ_M^2 is the variance in mutational effect sizes. A consequence of this formulation of the evolutionary rate is that the true rate used to simulate the data increases as we increase the rate of discordance by increasing *N*. This model is akin to the one shown in Figure 1B, in which mutations occur on gene tree branches with normally distributed effect sizes. Given enough mutations and enough time, these cumulative effects resemble Brownian motion of trait means along each lineage (Figure 1C; Mendes et al. 2018).

Data availability

Source code and analysis scripts related to *seastaR* and the covariance matrix method can be found in https://github.com/larabreithaupt/seastaR. Code and scripts related to the pruning algorithm method can be found in https://github.com/mhibbins/genetreepruningalg.

Results

Phylogenomic comparative approaches yield more accurate evolutionary rate estimates in the presence of discordance

We applied both of our new phylogenomic comparative approaches to data simulated with discordance in order to evaluate their accuracy in estimating the evolutionary rate parameter, σ^2. For the approach that uses the updated the variance/covariance matrix, $\boldsymbol{C^*}$, we use the maximum-likelihood estimator of σ^2:

$$\widehat{\sigma}^2 = \frac{[\boldsymbol{X} - E(\boldsymbol{X})]^{\mathrm{T}} \boldsymbol{C}^{-1} [\boldsymbol{X} - E(\boldsymbol{X})]}{n} \qquad \textit{Eq 5}$$

(O'Meara et al. 2006), where $\boldsymbol{X}$ is the vector of observed trait values at the tips, $\boldsymbol{C}$ is the phylogenetic variance-covariance matrix, and n is the number of tips. $E(\boldsymbol{X})$ is the vector containing the expected trait value at the root, calculated as follows:

$$E(\boldsymbol{X}) = \mathbf{1}[(\mathbf{1}^{\mathrm{T}}\boldsymbol{C}^{-1}\mathbf{1})^{-1}(\mathbf{1}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{X})] \qquad \textit{Eq 6}$$

Where $\mathbf{1}$ is a column vector of ones of size n x 1. To account for gene tree discordance with this estimator, we simply use $\boldsymbol{C}^*$ in place of $\boldsymbol{C}$ in equations 5 and 6. We have implemented this method in *seastaR* to allow users to estimate σ^2. For this approach, we simulated 1000 trait datasets for each condition of increasing gene tree discordance, estimating σ^2 using both $\boldsymbol{C}$ and $\boldsymbol{C}^*$ for each dataset.

For the approach using the pruning algorithm, we implemented the Nelder-Mead optimization algorithm. Given a set of input gene trees and tree frequencies, our optimization approach proposes a new value of σ^2 in each iteration, returning a single likelihood value over the set of gene trees each time; the optimal value of σ^2 is the one that maximizes this total likelihood. Owing to longer computation times, we simulated 100 trait datasets for each set of parameters with this method, using either a single tree specified (the species tree) or multiple trees specified (the gene trees).

As expected, we found that increasing the level of discordance results in an increasingly upward bias in estimates of the evolutionary rate from the species tree (Figure 4, green lines). As there are no internal branches in the species tree that can explain the increased trait covariances between non-sister taxa, such methods must propose a higher evolutionary rate to explain the data. In contrast, we found that both the covariance matrix approach (i.e. $\boldsymbol{C}^*$; Figure 4A) and pruning algorithm approach (Figure 4B) yielded more accurate evolutionary rate estimates, ones that closely tracked the true population-scaled evolutionary rate as the level of discordance increased. Both phylogenomic comparative approaches can model the increased covariances generated by the increasing frequencies of discordant gene trees. As can be observed, both approaches tend to slightly underestimate the true evolutionary rate, but they are much closer to the true value than standard species tree estimates, especially at higher rates of discordance.

Phylogenomic comparative approaches are robust to the effects of gene tree estimation error

In empirical datasets, it is reasonable to expect gene trees to be estimated with some degree of error, especially in the limits of short sequence length (such as ultra-conserved elements), long periods of evolutionary divergence, or high rates of sequencing error. In general, these sources of technical error should not be biased towards specific lineages, so their effect should be to cause general overestimation of gene tree discordance. This may in turn result in lower evolutionary rate estimates when using our approaches, as they might "overcorrect" the problem. More generally, we were concerned that increasing the rate of discordance might always leads to a lower evolutionary rate estimate, regardless of the true history that generated the data. Such behavior would present a potential problem for the application of our approaches to empirical datasets.

To address these concerns, we simulated traits from gene trees under a single, fixed rate of gene tree discordance (of approximately 15%) using the methods described in the previous section.

We then applied our approaches to estimating σ^2 to this dataset, varying the specified rate of gene tree discordance from 0 (in which case we used the standard species tree inference) to approximately 60%. In contrast to our initial concerns, we found that in both the covariance matrix (Figure 5A) and pruning algorithm (Figure 5B) approaches: 1) the effect of mis-specifying the rate of gene tree discordance is relatively small compared to the effect of using the species tree in place of gene trees; 2) increasing the specified rate of gene tree discordance leads to a small increase, rather than decrease, in the estimated evolutionary rate, but still closely tracked the true value. This latter effect may occur because increasing the specified rate of gene tree discordance requires branch lengths to be scaled down in accordance with *N*, resulting in less proposed time over which evolutionary changes can occur. Overall, these results suggest that gene tree estimation error should not be a major concern for our approaches, as long as the correct set of tree topologies is specified.

Rate estimates for floral traits in the wild tomato clade Solanum *are consistent with evolution on discordant gene trees*

Our simulations show that when traits evolve on discordant gene trees, standard species tree approaches tend to greatly overestimate the true value, while our gene tree approaches slightly underestimate the true value but are much more accurate. The degree of discrepancy between species tree and gene tree approaches grows larger as the rate of discordance increases. To further test these expectations and to highlight the application of our methods to empirical data, we estimated the evolutionary rates of several floral traits (anther length, corolla diameter, and stigma length) measured in wild tomatoes (*Solanum*) (Haak et al. 2014). We obtained the time-scaled phylogeny of this clade from Pease et al. (2016) and converted from time in years to coalescent units assuming $N = 100{,}000$ and one generation every two years (Hamlin et al. 2020). We then pruned the phylogenetic tree into high and low ILS triplets, each consisting of three taxa. The high ILS group consisted of the following accessions (IDs from the Tomato Genetics Resource Center): *S. galapagense* LA0436*, S. cheesmaniae* LA3124*,* and *S. pimpinellifolium* LA1269; the low ILS group consisted of *S. pennellii* LA3778, *S. pennellii* LA0716, and *S. pimpinellifolium* LA1589. Based on the internal branch lengths in coalescent units, the high ILS and low ILS triplets had expected rates of discordance of approximately 47% and 0.9%, respectively. These rates correspond to the rates of discordance seen in empirically estimated gene trees in Pease et al. (2016). Based on our simulation results, if discordant gene trees contribute to tomato floral trait variation, we should see a greater discrepancy between species tree methods and our gene tree methods in the high ILS triplet.

In both of our approaches, we used the multispecies coalescent model to calculate the expected gene tree frequencies and branch lengths in each triplet. For the covariance matrix method, we used these expectations to construct the covariance matrix $\boldsymbol{C^*}$ for each triplet using the *get_full_matrix*() method. For the pruning algorithm method, we specified a representative gene tree of each of the possible topologies with expected branch lengths, and weighted each tree by its expected frequency. For both methods, we used the standard approach of specifying a single species tree and our gene tree approaches to estimate the evolutionary rate using the mean trait values within each accession if multiple individuals were measured.

In line with our expectations, rate estimates obtained from standard species tree approaches are much higher than those obtained from both of our gene tree methods in the high ILS triplet, for all three traits (Figure 6). The discrepancy is much smaller in the low ILS triplets (Figure 6). The bias due to discordance was very large for the estimates obtained from the covariance matrix method in *seastaR* (Figure 6A), where the species tree rate estimates were several orders of magnitude higher than the gene tree estimates in the high ILS triplet. This is consistent with our simulation finding that the estimated evolutionary rate is more biased under discordance when using covariance methods (Figure 4A). Even when accounting for gene tree discordance, the rate estimates obtained from the covariance matrix method were substantially higher than those obtained from the pruning algorithm method (compare Figure 6A and 6B). This discrepancy can be explained by flat/undefined likelihood surfaces for the proposed values of σ^2 (Supplementary Figure 1): the pruning algorithm method, which employs a likelihood search, reaches a likelihood plateau and does not propose further improvements, whereas the covariance matrix method uses the analytical likelihood estimator to obtain the maximum value, regardless of the shape of the likelihood surface. This problem may arise when a small number of taxa are studied, as less information is available to discern the rate of trait evolution. To help users evaluate this problem in their datasets, we have implemented a function for calculating trait likelihoods over a range of proposed σ^2 values in *seastaR*. In this case, we believe the pruning algorithm estimates represent more biologically realistic rates of evolution.

Discussion

There has been much phylogenetic research focused on the accurate estimation of species trees in the face of gene tree discordance (e.g. Degnan and Rosenberg 2009; Bryant et al. 2012; Chifman and Kubatko 2014; Mirarab et al. 2014; Mendes and Hahn 2018; Zhang et al. 2018). Despite this focus on inferring trees in the face of discordance, standard phylogenetic comparative methods still rely on a single "resolved" tree to describe the shared history of species. Recent work has made it clear that, if only a single tree is used, gene tree discordance can shape trait variation and mislead comparative methods (e.g. Mendes et al. 2018; Hibbins et al. 2020). However, few solutions have been proposed to solve these problems, especially for quantitative traits evolving on clades containing discordance. Here, we have developed two approaches, which we refer to as *phylogenomic* comparative methods, that can incorporate gene tree discordance into comparative inference. One approach uses a more complete phylogenetic variance-covariance matrix that includes the covariance present in discordant gene trees. We have developed an R package, *seastaR*, for building this matrix using the frequencies and branch lengths of relevant gene trees. The second approach applies the pruning algorithm over a set of gene trees—concordant and discordant—to estimate likelihoods. Using simulation, we demonstrate that these methods generate more accurate evolutionary rate estimates for traits evolving in the presence of discordance, and are generally robust to the effects of gene tree estimation error. Finally, we demonstrate that empirical floral traits in the wild tomato clade *Solanum* are consistent with evolution on discordant gene trees, with the clade with a higher rate of gene tree discordance exhibiting a greater discrepancy in rate estimates between traditional approaches and our new methods.

Many phylogenetic comparative methods take the variance-covariance matrix, $\boldsymbol{C}$, as input (e.g. Pagel 1999; Housworth et al. 2004; O'Meara et al. 2006; Revell and Harmon 2008). Because of the wide use of $\boldsymbol{C}$, we anticipate that a more complete variance-covariance matrix, $\boldsymbol{C^*}$, will be easy to incorporate into many comparative analyses. The *seastaR* package provides an easy way for users to generate $\boldsymbol{C^*}$, either from a set of specified gene trees or from a specified species tree (assuming a multispecies coalescent process). Here, we have demonstrated how $\boldsymbol{C^*}$ can be used to obtain a maximum-likelihood estimate of the rate of quantitative trait evolution under Brownian motion, a method that is also implemented in *seastaR*. One obvious extension of the use of $\boldsymbol{C^*}$ is in phylogenetic generalized linear mixed models (PGLMMs), where the covariance matrix is often specified directly in packages such as *MCMCglmm* (Hadfield 2010). However, many popular packages for implementing comparative methods—such as *phytools* (Revell 2012), *ape* (Paradis and Schliep 2019), and *Geiger* (Pennell et al. 2014) —do not take a matrix directly, instead turning an input species tree into a matrix. Furthermore, they require a strictly bifurcating tree as input to construct a *phylo* class object. Integrating the ability to accept $\boldsymbol{C^*}$ (or equivalent sets of gene trees) into these methods would enable a much larger array of inference tasks to take discordance into account.

The pruning algorithm is widely used in likelihood-based inference of phylogenetic trees (Felsenstein 1981) and for some applications in quantitative trait evolution (e.g. Hahn et al. 2005; FitzJohn 2012; Freckleton 2012; Ho and Ané 2014; Uyeda and Harmon 2014; Hiscott et al. 2016; Mitov et al. 2020). Our method using the pruning algorithm across a set of gene trees makes many of the same assumptions as previous implementations, but models each trait as the combined result of a large number of loci; these loci were represented in our calculations by a smaller number of exemplar gene tree topologies, each with the mean set of branch lengths for a given topology. Although it is not as straightforward to incorporate our method into other approaches as with $\boldsymbol{C^*}$, because the pruning algorithm is a general method for calculating likelihoods, it has enormous potential to be applied to a wide variety of inference problems. As trees get larger, the computational cost of the matrix operations in equations 5 and 6 grows exponentially with the number of taxa. In contrast, the number of calculations in the pruning algorithm only grows linearly, and therefore trees with thousands of tips can be analyzed (Mitov and Stadler 2019). Furthermore, even though several methods for dealing with sparse matrices make it possible to analyze larger numbers of taxa (e.g. Hadfield and Nakagawa 2010), $\boldsymbol{C^*}$ has more covariance entries and is therefore less sparse than $\boldsymbol{C}$; this again limits matrix-based approaches in phylogenomic comparative methods.

Both of our approaches can be extended in multiple ways. While we have only considered Brownian motion models here, there are multiple other trait models that could be used. The Ornstein-Uhlenbeck process is a popular model for trait evolution, with estimators available using both matrix (Hansen 1997; Butler and King 2004; Beaulieu et al. 2012; Rohlfs et al. 2014) and pruning algorithm approaches (FitzJohn 2012; Ho and Ané 2014; Uyeda and Harmon 2014; Mitov et al. 2020). Additional models for continuous traits include "early burst" (Harmon et al. 2010) and Lévy ("jump") processes (Landis and Schraiber 2017). All of these models should be able to be accommodated by phylogenomic comparative methods. In addition, although we have described the covariances in our models with a particular set of gene trees in mind, both methods can be used with any weighted mixture of trees. This means that users do not have to assume a particular model of species tree evolution (e.g. the multispecies coalescent model) and can even

ignore ILS altogether in favor of phylogenetic network models (e.g. Bastide et al. 2018). This should also allow our approaches to accommodate unequal contributions to trait variation across individual loci, for example, if some loci are known to be functionally related to the trait of interest and therefore expected to have mutations of larger effect on average.

There are also multiple caveats that come with our proposed approaches, and some important technical limitations to consider. First, errors in gene tree or species tree specification might bias inferences. This is especially true if gene trees are being used as inputs, as we require both accurate and ultrametric trees. Our methods assume that loci controlling trait variation and the genome at large follow the same distribution of trees; however, if trait loci experience stronger than average selection, these loci could have proportionally fewer discordant gene trees than the genomic background (cite something here?). Gene trees may also be mis-specified due to technical errors in their estimation. We found that error in gene tree frequencies and branch lengths is relatively inconsequential for our approaches, under the conditions considered here. Specifying a set of incorrect gene tree topologies may have more of an effect, but since ILS is expected to generate all possible topologies with respect to a single branch, we do not expect this to be a significant issue. Obtaining ultrametric gene trees remains challenging due to variation in rates of evolution among loci and small amounts of data per locus. Even when species trees are used to generate gene tree frequencies (i.e. *get_full_matrix*), many coalescent-based methods for inferring species trees do not estimate tip branch lengths (e.g. Liu et al. 2010; Mirarab et al. 2014), further limiting accurate inferences (but see Bastide et al. 2018; Hibbins et al. 2020). If there is uncertainty in the species tree topology or branch lengths, a straightforward solution would be to embed the approaches used here within a Bayesian framework (e.g. Huelsenbeck et al. 2000; Pagel and Meade 2006). It is important to note, however, that gene tree discordance is not equivalent to species tree uncertainty: averaging over each gene tree topology on its own in a Bayesian framework would simply mean averaging over many incorrect trees. Instead, a proper Bayesian approach to accommodating discordance would have to sum over a new set of gene trees (or covariances) for each species tree topology proposed, as was done here with a single topology.

A second caveat is that large numbers of taxa make it harder to accurately estimate both the matrix used in *seastaR* and the gene trees used within the pruning algorithm. If gene trees are predicted from theory, *seastaR* calculates $\boldsymbol{C}^*$ from the species tree by breaking the tree into triplets. While this will return approximately correct covariances for all pairs of species, it necessarily ignores any covariance structures that might only be possible in trees with four or more taxa. The problem for the pruning algorithm approach could be even worse, as separate gene tree topologies must be specified: specifying representative gene trees for all possible topologies becomes prohibitive with more taxa because the number of gene trees grows super-exponentially. Even if gene trees are estimated from the data, with only a few dozen taxa there are more possible gene tree topologies than independent loci in a genome. Two solutions suggest ways around these issues. First, the problems can be somewhat ameliorated by recognizing that it is not the number of taxa that is the issue, but instead the number of lineages within "knots" (cf. Ané et al. 2007) on the larger phylogeny that are prone to gene tree discordance. For instance, even in a tree with 100 species, if only three are undergoing ILS, then only three topologies must be considered. Judicious choices as to the number of different topologies that must be considered in any particular analysis could save a lot of computational effort. Second, as mentioned in the Methods, one approach that can be applied to the pruning algorithm method is to sample a limited number of individual gene

trees, either directly from the inferred trees or from the multispecies coalescent model applied to the species tree. Even if we have to sample 100 trees, the likelihood calculations on each are relatively fast and can be parallelized. Such a sampling scheme will also naturally recapitulate the degree of discordance associated with every branch in the species tree.

Throughout our analyses, we found that rate estimates using a single species tree differed from those accounting for gene trees, even when the level of discordance was very low or zero (Figure 4A). This result occurs because the two modes of inference are fundamentally different: even with no discordance, "gene tree" analyses are based on gene tree branch lengths, not species tree branch lengths. Gene tree branch lengths are always longer than species tree branch lengths because each pair of lineages is expected to coalesce $2N$ generations before their time of speciation (Gillespie and Langley 1979; Edwards and Beerli 2000). These longer gene tree branch lengths result in higher trait variances in the traits, such that a higher evolutionary rate must be proposed to explain the same data when using the species tree for analysis. This distinction highlights an additional challenge for a potential application of our pruning algorithm approach – ancestral state reconstruction. Because the internal nodes of gene trees—including concordant gene trees—do not exist at the same moment of time as the internal nodes of species trees, reconstructing ancestral states at the time of speciation requires knowledge of the contribution of each gene tree branch to trait evolution at that particular time point. This could be accomplished by the insertion of single-descendant nodes on gene tree lineages that are concurrent with ancestral nodes on the species tree. Inferring lineage-specific rate shifts will likewise require that each gene tree branch, or segment of a gene tree branch, be assigned to specific species tree lineages (cf. Ogilvie et al. 2017). In general, these considerations highlight the fact that using gene trees in place of a species tree is a fundamentally different mode of inference, and that standard comparative methods using the species tree may yield incorrect inferences even if there is no discordance.

We consistently found that evolutionary rate estimates for tomato floral traits were much greater using species tree approaches than our gene tree approaches (Figure 6). For the high ILS triplet, there was even more bias than in the low ILS knot. These results are consistent with a contribution of gene trees, rather than a single species phylogeny, to variation in these traits. Our analysis of gene tree error suggests that this result is not simply an artifact of increasing the specified rate of gene tree discordance, but is the result of biological variation in the floral traits. Furthermore, our findings have implications for the study of evolutionary rate variation among clades. For example, imagine that researchers wished to investigate whether the evolutionary rate of corolla diameter differed between our high ILS and low ILS triplets. Applying standard species tree methods, they would find that the corolla diameter of the high ILS species evolves at a much faster rate than in their low ILS counterparts. However, from our results in Figure 6, after correcting for the contribution of discordant gene trees, this difference disappears and the trait appears to evolve at approximately the same rate in both clades. This result highlights how variation in the rate of gene tree discordance among clades is a confounding factor when studying the evolution of lineage-specific rate shifts.

An increasingly common finding in phylogenomics is that of rapid and/or highly parallel trait evolution associated with rapid species radiations (Schluter et al. 1997; Boughman et al. 2005; Sun et al. 2012; Parins-Fukuchi et al. 2021; Urban et al. 2022). The application of classic comparative approaches in these systems has suggested that many radiations violate constant-rate

Brownian motion models, with more complex models being proposed instead (Simpson 1944; Blomberg et al. 2003; Freckleton and Harvey 2006). However, adaptive radiations often have very little time between speciation events, resulting in high rates of gene tree discordance and therefore high potential for hemiplasy (Pease et al. 2016). Here we have found that apparently higher rates of trait evolution in rapid radiations may be perfectly consistent with a standard Brownian motion model with a constant evolutionary rate. In this circumstance, higher apparent rates of evolution are simply the result of a stronger contribution of discordant gene trees to covariance among species. Our proposed phylogenomic comparative methods help to address these issues, providing more accurate evolutionary inferences in systems with high rates of discordance.

Acknowledgements

We would like to thank Ben Fulton and Jason Bertram for offering helpful programming and statistical advice in implementing the pruning algorithm approach, as well as David Haak for sharing floral trait data. Two reviewers provided helpful comments on the manuscript. This work was supported by the EEB Postdoctoral Fellowship awarded to Mark Hibbins by the Department of Ecology and Evolutionary Biology at the University of Toronto, as well as National Science Foundation grant DEB-1936187 awarded to Matthew Hahn.

Figures

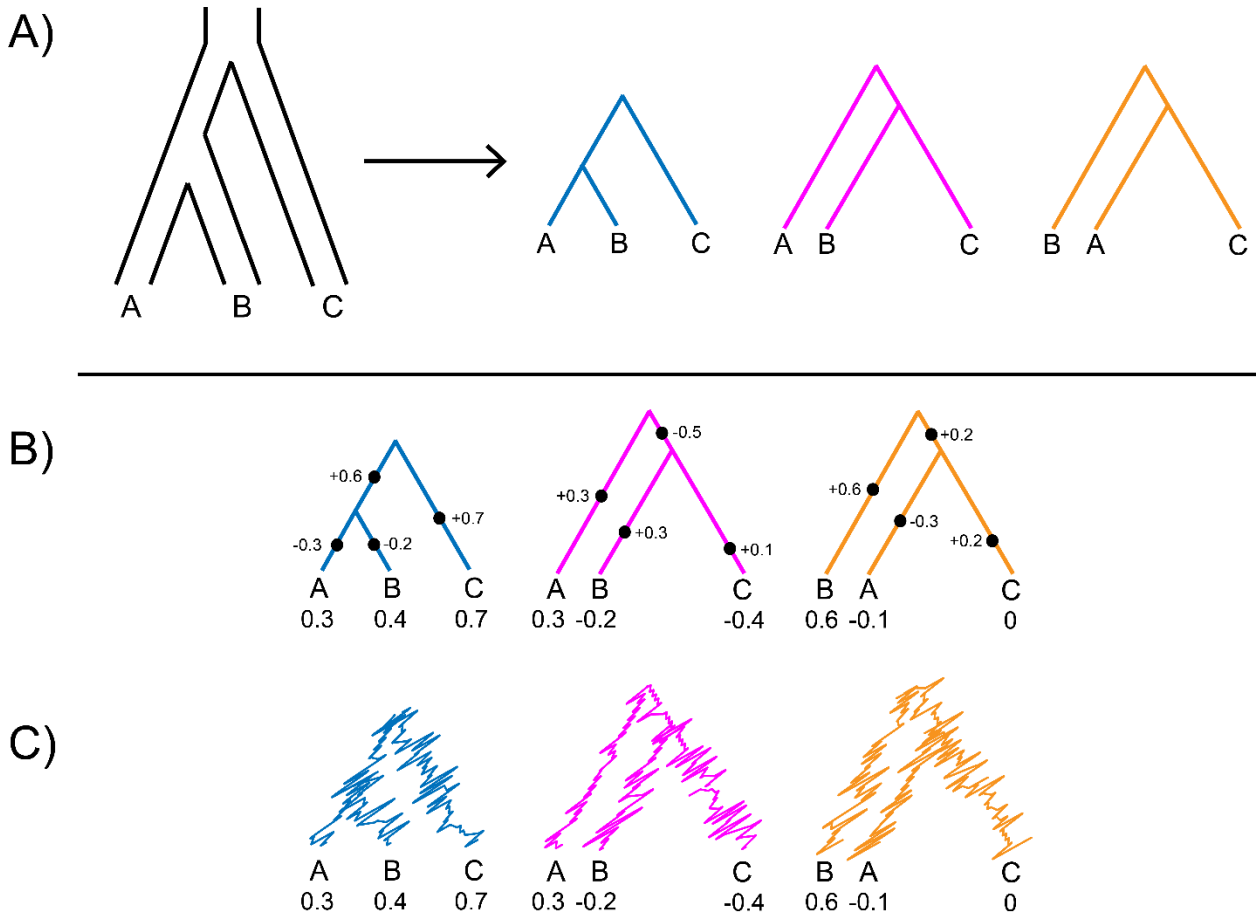

Figure 1: Conceptualizing quantitative trait evolution with discordant gene trees. A) Given a species tree (far left), we model gene trees as arising under the multispecies coalescent process. One topology is concordant with the species tree (blue), while the other two possible topologies are discordant (red and yellow). Under ILS these two discordant trees have the same topology and frequency. B) Over the course of evolution, mutations occur at loci that affect quantitative traits, each of which has a topology drawn from the multispecies coalescent process. Mutations on the internal branches of discordant gene trees can introduce shared trait history that is not captured by the species tree. Here we summarize mutations occurring at different loci on a single tree if the loci had the same topology, with each mutation at each locus contributing positively or negatively to the trait value in each species. In the example here, species pairs B-C and A-C might covary in quantitative trait values due to mutations on shared branches in gene trees, despite sharing no common ancestor in the species tree. For example, a mutation on the internal branch of the magenta tree causes species B and C to have more similar trait values. C) Given a large number of mutations and loci, trait evolution over time can be modelled by Brownian motion on each gene tree topology. This stochastic process models the trait value as a random walk over time, with species trait values calculated as the weighted average of the values on each gene tree (Mendes et al. 2018; Hibbins and Hahn 2021).

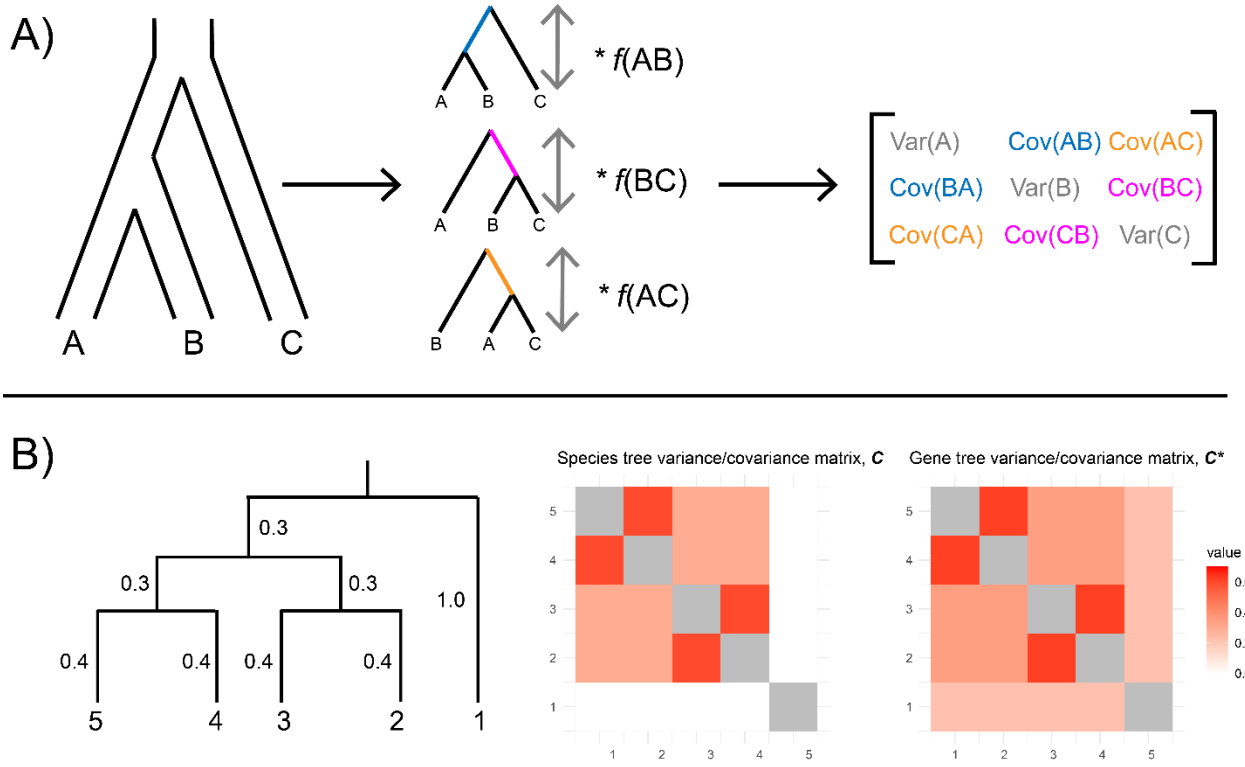

Figure 2: Inferring the gene tree variance/covariance matrix, ***C****. A) Gene trees are generated from a species tree under the multispecies coalescent process (note that introgression can readily be incorporated, but is not shown here for clarity). Each gene tree contributes its internal branch length (for covariance terms) and its total height (for variance terms) to ***C****. The contribution of each tree to ***C**** is weighted by its expected or observed frequency, depending on the approach taken. Frequencies are denoted as *f*(XY), where X and Y are the taxa sister in the gene tree of interest. B) A comparison of ***C*** and ***C**** for a five-taxon species tree with branch lengths labelled in coalescent units (not precisely to scale). Each internal branch has a length of 0.3, corresponding to a level of discordance of approximately 50%. This level of discordance means that each clade descended from these internal branches (5/4/3/2, 5/4, and 3/2) will be present in ~50% of gene trees. The standard phylogenetic covariance matrix, ***C***, contains no covariance between species 1 and the other taxa, because they do not share an internal branch in the species tree. In contrast, species 1 covaries with all other species in the tree using ***C****, because multiple discordant gene trees have species 2-5 sharing an internal branch with species 1.

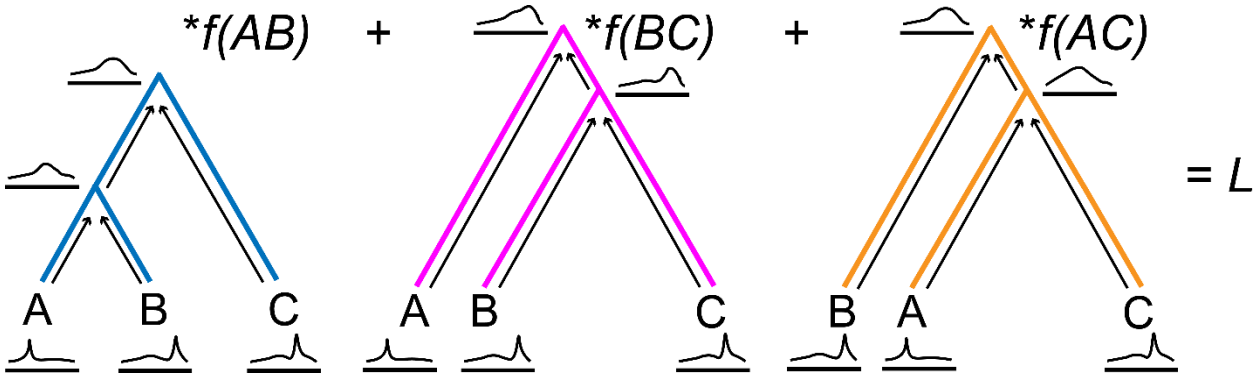

Figure 3: Applying the pruning algorithm to sets of gene trees. In our proposed approach, the pruning algorithm (shown as upward arrows) is applied to each individual gene tree to obtain character state probabilities at each node (curved lines) up to the root for a quantitative trait. These root probabilities are then used to obtain a partial likelihood from each gene tree, which are then summed together weighted by the gene tree frequency to obtain the final likelihood.

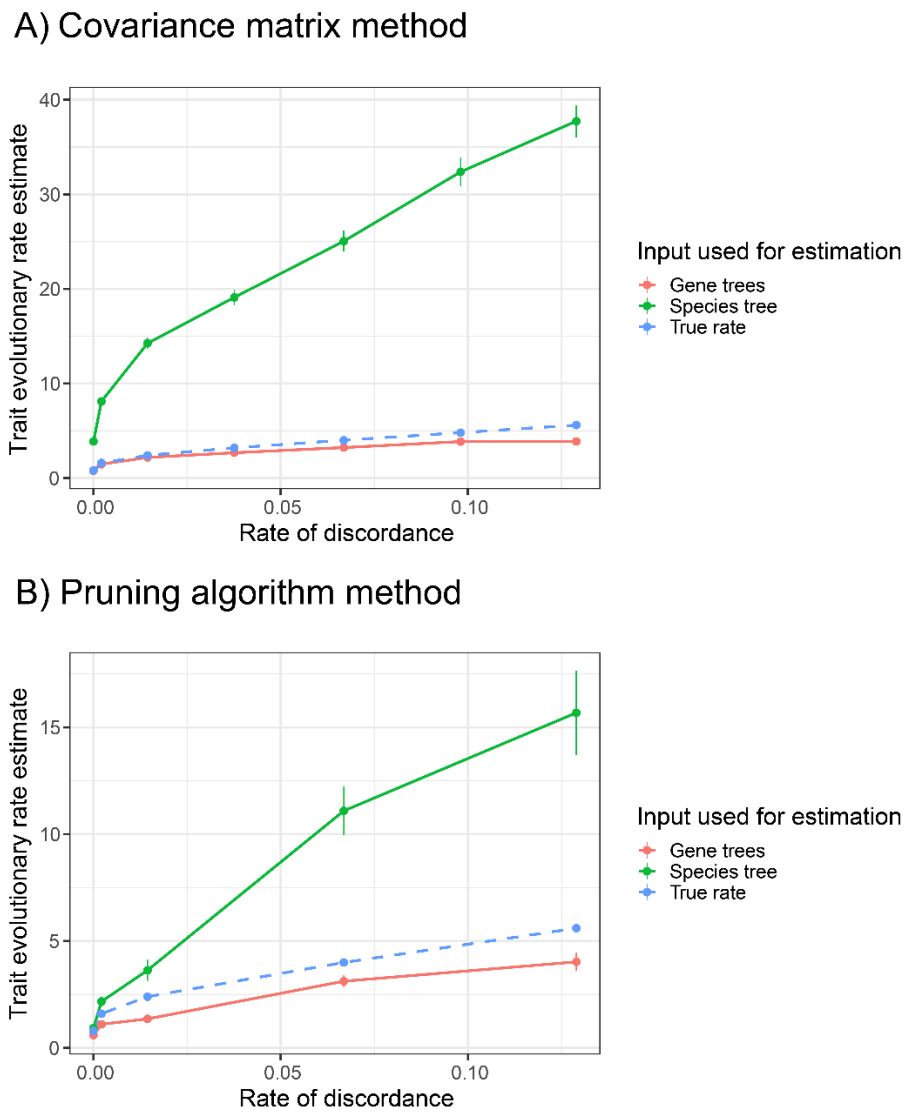

Figure 4: Phylogenomic comparative methods produce more accurate evolutionary rate estimates. A) Rate estimates obtained using a maximum-likelihood estimator applied to the covariance matrix (equation 5). B) Rate estimates obtained using numerical optimization of the likelihood with the pruning algorithm. In both panels, the green line shows inferences from methods using only the species tree, the red line shows the inferences from methods accounting for gene tree discordance, and the blue dashed line shows the true simulated value of the evolutionary rate. The level of gene tree discordance expected from each simulated species tree (see Supplementary Methods) is shown on the x-axis. Note that panels A and B have different y-axis scales.

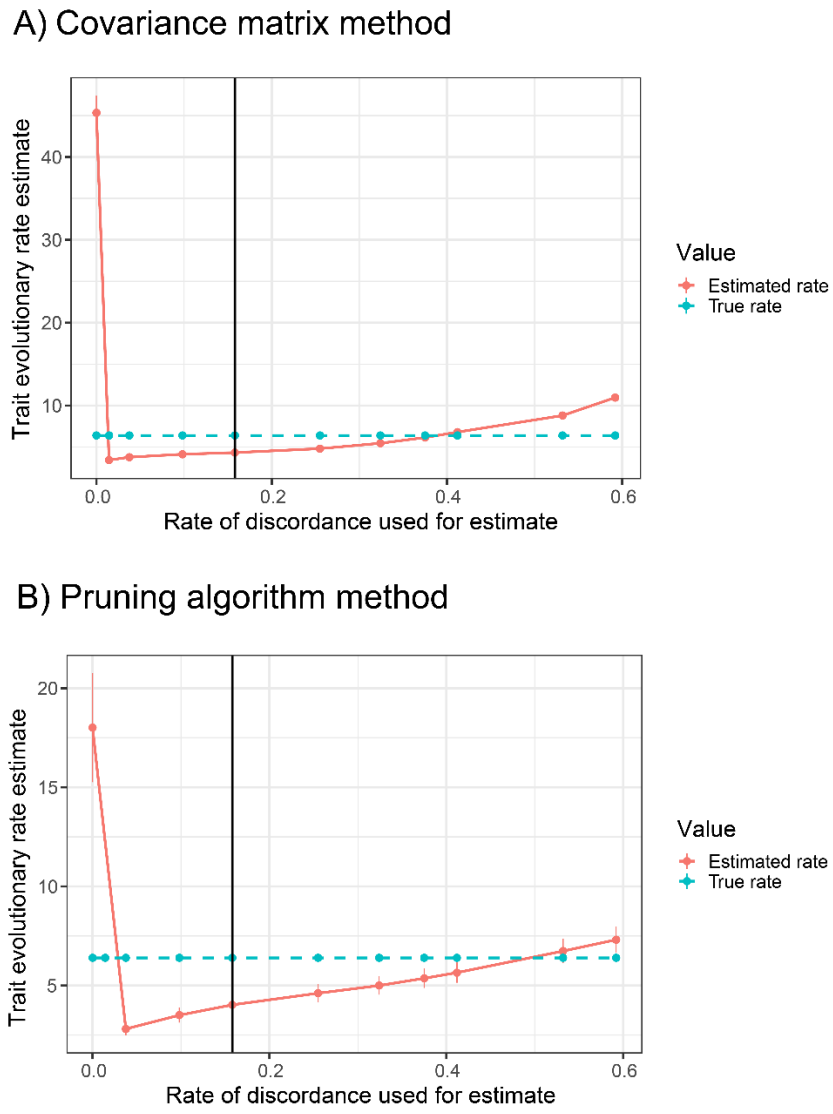

Figure 5: Phylogenomic comparative methods are robust to gene tree estimation error. In both panels, the solid vertical line denotes the true rate of discordance used to simulate the trait data, and the horizontal blue line denotes the true evolutionary rate. The x-axis shows the rate of discordance supplied to each approach when estimating the evolutionary rate from the simulated data. For a rate of discordance equal to 0, we used the standard species tree inference rather than gene tree inference. Specifying too much discordance can also cause overestimation of the evolutionary rate. Note that panels A and B have different y-axis scales.

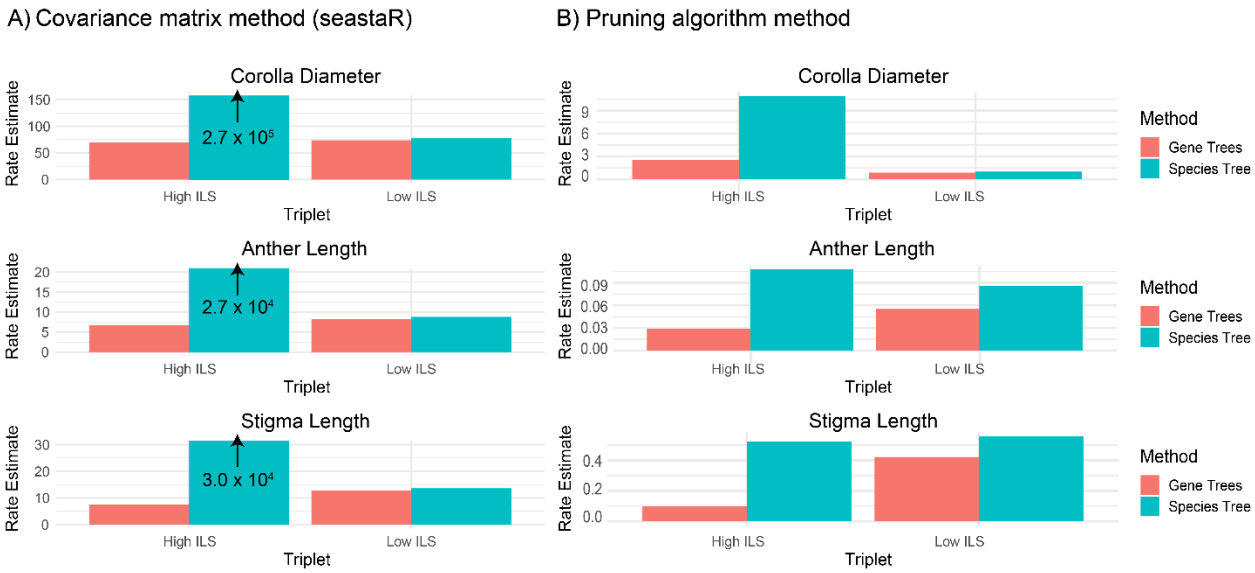

Figure 6: Evolutionary rate estimates for three floral traits in *Solanum* using our newly developed approaches (red bars), in comparison to standard species tree methods (blue bars). A) rate estimates (σ^2) obtained using the analytical maximum-likelihood estimator as implemented in our R package *seastaR*. In the high ILS triplet, the species tree estimates (blue) go far above the scale of the y-axis, so these bars are labelled with an up arrow and the true estimated values for clarity. B) Rate estimates (σ^2) obtained via maximum likelihood optimization using our pruning algorithm implementation for calculating likelihoods on gene trees. Note that the values on the y-axes are not the same in the two panels.

References

Adams, D. C., & Collyer, M. L. (2018). Multivariate phylogenetic comparative methods: evaluations, comparisons, and recommendations. *Systematic Biology, 67*(1), 14-31. doi:10.1093/sysbio/syx055

Alfaro, M. E., Santini, F., Brock, C., Alamillo, H., Dornburg, A., Rabosky, D. L., . . . Harmon, L. J. (2009). Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proceedings of the National Academy of Sciences of the United States of America, 106*(32), 13410-13414. doi:10.1073/pnas.0811087106

Ane, C., Larget, B., Baum, D. A., Smith, S. D., & Rokas, A. (2007). Bayesian estimation of concordance among gene trees. *Molecular Biology and Evolution, 24*(2), 412-426. doi:10.1093/molbev/msl170

Avise, J. C., & Robinson, T. J. (2008). Hemiplasy: a new term in the lexicon of phylogenetics. *Systematic Biology, 57*(3), 503-507. doi:10.1080/10635150802164587

Bastide, P., Solis-Lemus, C., Kriebel, R., William Sparks, K., & Ane, C. (2018). Phylogenetic comparative methods on phylogenetic networks with reticulations. *Systematic Biology, 67*(5), 800-820. doi:10.1093/sysbio/syy033

Beaulieu, J. M., Jhwueng, D. C., Boettiger, C., & O'Meara, B. C. (2012). Modeling stabilizing selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution. *Evolution, 66*(8), 2369-2383. doi:10.1111/j.1558-5646.2012.01619.x

Blomberg, S. P., Garland, T., Jr., & Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution, 57*(4), 717-745. doi:10.1111/j.0014-3820.2003.tb00285.x

Boucher, F. C., & Demery, V. (2016). Inferring bounded evolution in phenotypic characters from phylogenetic comparative data. *Systematic Biology, 65*(4), 651-661. doi:10.1093/sysbio/syw015

Boughman, J. W., Rundle, H. D., & Schluter, D. (2005). Parallel evolution of sexual isolation in sticklebacks. *Evolution, 59*(2), 361-373. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/15807421

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., & RoyChoudhury, A. (2012). Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution, 29*(8), 1917-1932. doi:10.1093/molbev/mss086

Butler, M. A., & King, A. A. (2004). Phylogenetic comparative analysis: A modeling approach for adaptive evolution. *American Naturalist, 164*(6), 683-695. doi:10.1086/426002

Chifman, J., & Kubatko, L. (2014). Quartet inference from SNP data under the coalescent model.

Bioinformatics, 30(23), 3317-3324. doi:10.1093/bioinformatics/btu530

Copetti, D., Burquez, A., Bustamante, E., Charboneau, J. L. M., Childs, K. L., Eguiarte, L. E., . . . Sanderson, M. J. (2017). Extensive gene tree discordance and hemiplasy shaped the genomes of North American columnar cacti. *Proceedings of the National Academy of Sciences of the United States of America, 114*(45), 12003-12008. doi:10.1073/pnas.1706367114

Dagilis, A. J., Peede, D., Coughlan, J. M., Jofre, G. I., D'Agostino, E. R. R., Mavengere, H., . . . Matute, D. R. (2022). A need for standardized reporting of introgression: Insights from studies across eukaryotes. *Evolution Letters, 6*(5), 344-357. doi:10.1002/evl3.294

De Maio, N., Schlotterer, C., & Kosiol, C. (2013). Linking great apes genome evolution across time scales using polymorphism-aware phylogenetic models. *Molecular Biology and Evolution, 30*(10), 2249-2262. doi:10.1093/molbev/mst131

De Maio, N., Wu, C. H., O'Reilly, K. M., & Wilson, D. (2015). New routes to phylogeography: a Bayesian structured coalescent approximation. *PLoS Genetics, 11*(8), e1005421. doi:10.1371/journal.pgen.1005421

Degnan, J. H., & Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology and Evolution, 24*(6), 332-340. doi:10.1016/j.tree.2009.01.009

Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution, 28*(8), 2239-2252. doi:10.1093/molbev/msr048

Edelman, N. B., Frandsen, P. B., Miyagi, M., Clavijo, B., Davey, J., Dikow, R. B., . . . Mallet, J. (2019). Genomic architecture and introgression shape a butterfly radiation. *Science, 366*(6465), 594-599. doi:10.1126/science.aaw2090

Edwards, S. V., & Beerli, P. (2000). Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution, 54*(6), 1839-1854. doi:10.1111/j.0014-3820.2000.tb01231.x

Felsenstein, J. (1973). Maximum-likelihood estimation of evolutionary trees from continuous characters. *American Journal of Human Genetics, 25*(5), 471-492. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/4741844

Felsenstein, J. (1981). Evolutionary trees from DNA-sequences - a maximum-likelihood approach. *Journal of Molecular Evolution, 17*(6), 368-376. doi:10.1007/Bf01734359

Felsenstein, J. (1985). Phylogenies and the comparative method. *American Naturalist, 125*(1), 1-15. doi:10.1086/284325

FitzJohn, R. G. (2012). DiversiTree: comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution, 3*(6), 1084-1092. doi:10.1111/j.2041-

859 210X.2012.00234.x

860 Fontaine, M. C., Pease, J. B., Steele, A., Waterhouse, R. M., Neafsey, D. E., Sharakhov, I. V., . . .
861 Besansky, N. J. (2015). Extensive introgression in a malaria vector species complex
862 revealed by phylogenomics. *Science, 347*(6217), 1258524. doi:10.1126/science.1258524

863 Freckleton, R. P. (2012). Fast likelihood calculations for comparative analyses. *Methods in*
864 *Ecology and Evolution, 3*(5), 940-947. doi:10.1111/j.2041-210X.2012.00220.x

865 Freckleton, R. P., & Harvey, P. H. (2006). Detecting non-Brownian trait evolution in adaptive
866 radiations. *PLoS Biology, 4*(11), e373. doi:10.1371/journal.pbio.0040373

867 Garamszegi, L. Z. (2014). *Modern phylogenetic comparative methods and their application in*
868 *evolutionary biology*. Berlin: Springer.

869 Garland, T., & Ives, A. R. (2000). Using the past to predict the present: Confidence intervals for
870 regression equations in phylogenetic comparative methods. *American Naturalist, 155*(3),
871 346-364. doi:10.1086/303327

872 Gillespie, J. H., & Langley, C. H. (1979). Are evolutionary rates really variable? *Journal of*
873 *Molecular Evolution, 13*(1), 27-34. doi:10.1007/BF01732751

874 Grafen, A. (1989). The phylogenetic regression. *Philosophical Transactions of the Royal Society*
875 *of London B: Biological Sciences, 326*(1233), 119-157. doi:10.1098/rstb.1989.0106

876 Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., . . . Paabo, S. (2010).
877 A draft sequence of the Neandertal genome. *Science, 328*(5979), 710-722.
878 doi:10.1126/science.1188021

879 Guerrero, R. F., & Hahn, M. W. (2018). Quantifying the risk of hemiplasy in phylogenetic
880 inference. *Proceedings of the National Academy of Sciences of the United States of*
881 *America, 115*(50), 12787-12792. doi:10.1073/pnas.1811268115

882 Haak, D. C., Ballenger, B. A., & Moyle, L. C. (2014). No evidence for phylogenetic constraint on
883 natural defense evolution among wild tomatoes. *Ecology, 95*(6), 1633-1641.
884 doi:10.1890/13-1145.1

885 Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: the
886 MCMCglmm R package. *Journal of Statistical Software, 33*(2), 1-22.
887 doi:10.18637/jss.v033.i02

888 Hadfield, J. D., & Nakagawa, S. (2010). General quantitative genetic methods for comparative
889 biology: phylogenies, taxonomies and multi-trait models for continuous and categorical
890 characters. *Journal of Evolutionary Biology, 23*(3), 494-508. doi:10.1111/j.1420-
891 9101.2009.01915.x

892 Hahn, M. W., De Bie, T., Stajich, J. E., Nguyen, C., & Cristianini, N. (2005). Estimating the tempo
893 and mode of gene family evolution from comparative genomic data. *Genome Research,*

15(8), 1153-1160. doi:10.1101/gr.3567505

Hahn, M. W., & Nakhleh, L. (2016). Irrational exuberance for resolved species trees. *Evolution, 70*(1), 7-17. doi:10.1111/evo.12832

Hansen, T. F. (1997). Stabilizing selection and the comparative analysis of adaptation. *Evolution, 51*(5), 1341-1351. doi:10.1111/j.1558-5646.1997.tb01457.x

Harmon, L. J., Losos, J. B., Jonathan Davies, T., Gillespie, R. G., Gittleman, J. L., Bryan Jennings, W., . . . Mooers, A. O. (2010). Early bursts of body size and shape evolution are rare in comparative data. *Evolution, 64*(8), 2385-2396. doi:10.1111/j.1558-5646.2010.01025.x

Harvey, P. H., & Pagel, M. (1991). *The comparative method in evolutionary biology*. Oxford: Oxford University Press.

Hibbins, M. S., Gibson, M. J., & Hahn, M. W. (2020). Determining the probability of hemiplasy in the presence of incomplete lineage sorting and introgression. *eLife, 9*. doi:10.7554/eLife.63753

Hibbins, M. S., & Hahn, M. W. (2021). The effects of introgression across thousands of quantitative traits revealed by gene expression in wild tomatoes. *PLoS Genetics, 17*(11), e1009892. doi:10.1371/journal.pgen.1009892

Hiscott, G., Fox, C., Parry, M., & Bryant, D. (2016). Efficient recycled algorithms for quantitative trait models on phylogenies. *Genome Biology and Evolution, 8*(5), 1338-1350. doi:10.1093/gbe/evw064

Ho, L., & Ane, C. (2014). A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Systematic Biology, 63*(3), 397-408. doi:10.1093/sysbio/syu005

Housworth, E. A., Martins, E. P., & Lynch, M. (2004). The phylogenetic mixed model. *American Naturalist, 163*(1), 84-96. doi:10.1086/380570

Hudson, R. R. (1983). Testing the constant-rate neutral allele model with protein sequence data. *Evolution, 37*(1), 203-217. doi:10.1111/j.1558-5646.1983.tb05528.x

Huelsenbeck, J. P., Rannala, B., & Masly, J. P. (2000). Accommodating phylogenetic uncertainty in evolutionary studies. *Science, 288*(5475), 2349-2350. doi:10.1126/science.288.5475.2349

Landis, M. J., & Schraiber, J. G. (2017). Pulsed evolution shaped modern vertebrate body sizes. *Proceedings of the National Academy of Sciences of the United States of America, 114*(50), 13224-13229. doi:10.1073/pnas.1710920114

Liu, L., Yu, L., & Edwards, S. V. (2010). A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evolutionary Biology, 10*, 302. doi:10.1186/1471-2148-10-302

928 Mallet, J., Besansky, N., & Hahn, M. W. (2016). How reticulated are species? *BioEssays, 38*(2),
929 140-149. doi:10.1002/bies.201500149

930 Martins, E. P., & Hansen, T. F. (1996). The statistical analysis of interspecific data: a review and
931 evaluation of phylogenetic comparative methods. In E. P. Martins (Ed.), *Phylogenies and*
932 *the comparative method in animal behavior.* New York: Oxford University Press.

933 Mendes, F. K., Fuentes-Gonzalez, J. A., Schraiber, J. G., & Hahn, M. W. (2018). A multispecies
934 coalescent model for quantitative traits. *eLife, 7*. doi:10.7554/eLife.36482

935 Mendes, F. K., & Hahn, M. W. (2016). Gene tree discordance causes apparent substitution rate
936 variation. *Systematic Biology, 65*(4), 711-721. doi:10.1093/sysbio/syw018

937 Mendes, F. K., Vanderpool, D., Fulton, B., & Hahn, M. W. (2020). CAFE 5 models variation in
938 evolutionary rates among gene families. *Bioinformatics*, *36*(23), 5516-5518.
939 doi:10.1093/bioinformatics/btaa1022

940 Mirarab, S., Reaz, R., Bayzid, M. S., Zimmermann, T., Swenson, M. S., & Warnow, T. (2014).
941 ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics, 30*(17),
942 i541-548. doi:10.1093/bioinformatics/btu462

943 Mitov, V., Bartoszek, K., Asimomitis, G., & Stadler, T. (2020). Fast likelihood calculation for
944 multivariate Gaussian phylogenetic models with shifts. *Theoretical Population Biology,*
945 *131*, 66-78. doi:10.1016/j.tpb.2019.11.005

946 Mitov, V., & Stadler, T. (2019). Parallel likelihood calculation for phylogenetic comparative
947 models: The SPLITT C++ library. *Methods in Ecology and Evolution, 10*(4), 493-506.
948 doi:10.1111/2041-210x.13136

949 Nelder, J. A., & Mead, R. (1965). A simplex-method for function minimization. *Computer*
950 *Journal, 7*(4), 308-313. doi:10.1093/comjnl/7.4.308

951 Novikova, P. Y., Hohmann, N., Nizhynska, V., Tsuchimatsu, T., Ali, J., Muir, G., . . . Nordborg,
952 M. (2016). Sequencing of the genus *Arabidopsis* identifies a complex history of
953 nonbifurcating speciation and abundant trans-specific polymorphism. *Nature Genetics,*
954 *48*(9), 1077-1082. doi:10.1038/ng.3617

955 Ogilvie, H. A., Bouckaert, R. R., & Drummond, A. J. (2017). StarBEAST2 brings faster species
956 tree inference and accurate estimates of substitution rates. *Molecular Biology and*
957 *Evolution, 34*(8), 2101-2114. doi:10.1093/molbev/msx126

958 O'Meara, B. C., Ane, C., Sanderson, M. J., & Wainwright, P. C. (2006). Testing for different rates
959 of continuous trait evolution using likelihood. *Evolution, 60*(5), 922-933. Retrieved from
960 https://www.ncbi.nlm.nih.gov/pubmed/16817533

961 Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature, 401*(6756), 877-
962 884. doi:10.1038/44766

963 Pagel, M., & Meade, A. (2006). Bayesian analysis of correlated evolution of discrete characters
964 by reversible-jump Markov chain Monte Carlo. *American Naturalist, 167*(6), 808-825.
965 doi:10.1086/503444

966 Pamilo, P., & Nei, M. (1988). Relationships between gene trees and species trees. *Molecular*
967 *Biology and Evolution, 5*(5), 568-583. doi:10.1093/oxfordjournals.molbev.a040517

968 Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and
969 evolutionary analyses in R. *Bioinformatics, 35*(3), 526-528.
970 doi:10.1093/bioinformatics/bty633

971 Parins-Fukuchi, C., Stull, G. W., & Smith, S. A. (2021). Phylogenomic conflict coincides with
972 rapid morphological innovation. *Proceedings of the National Academy of Sciences of the*
973 *United States of America, 118*(19). doi:10.1073/pnas.2023058118

974 Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., . . . Reich, D. (2012).
975 Ancient admixture in human history. *Genetics, 192*(3), 1065-1093.
976 doi:10.1534/genetics.112.145037

977 Pease, J. B., Haak, D. C., Hahn, M. W., & Moyle, L. C. (2016). Phylogenomics reveals three
978 sources of adaptive variation during a rapid radiation. *PLoS Biology, 14*(2), e1002379.
979 doi:10.1371/journal.pbio.1002379

980 Pennell, M. W., Eastman, J. M., Slater, G. J., Brown, J. W., Uyeda, J. C., FitzJohn, R. G., . . .
981 Harmon, L. J. (2014). geiger v2.0: an expanded suite of methods for fitting
982 macroevolutionary models to phylogenetic trees. *Bioinformatics, 30*(15), 2216-2218.
983 doi:10.1093/bioinformatics/btu181

984 Pollard, D. A., Iyer, V. N., Moses, A. M., & Eisen, M. B. (2006). Widespread discordance of gene
985 trees with species tree in *Drosophila*: evidence for incomplete lineage sorting. *PLoS*
986 *Genetics, 2*(10), e173. doi:10.1371/journal.pgen.0020173

987 Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian
988 population history. *Nature, 461*(7263), 489-494. doi:10.1038/nature08365

989 Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other
990 things). *Methods in Ecology and Evolution, 3*(2), 217-223. doi:10.1111/j.2041-
991 210X.2011.00169.x

992 Revell, L. J., & Harmon, L. J. (2008). Testing quantitative genetic hypotheses about the
993 evolutionary rate matrix for continuous characters. *Evolutionary Ecology Research, 10*(3),
994 311-331. Retrieved from <Go to ISI>://WOS:000256015000001

995 Revell, L. J., & Harmon, L. J. (2022). *Phylogenetic comparative methods in R*: Princeton
996 University Press.

997 Rohlfs, R. V., Harrigan, P., & Nielsen, R. (2014). Modeling gene expression evolution with an
998 extended Ornstein-Uhlenbeck process accounting for within-species variation. *Molecular*

Biology and Evolution, 31(1), 201-211. doi:10.1093/molbev/mst190

Schluter, D., Price, T., Mooers, A. O., & Ludwig, D. (1997). Likelihood of ancestor states in adaptive radiation. *Evolution, 51*(6), 1699-1711. doi:10.1111/j.1558-5646.1997.tb05095.x

Schrempf, D., Minh, B. Q., De Maio, N., von Haeseler, A., & Kosiol, C. (2016). Reversible polymorphism-aware phylogenetic models and their application to tree inference. *Journal of Theoretical Biology, 407*, 362-370. doi:10.1016/j.jtbi.2016.07.042

Schrempf, D., Minh, B. Q., von Haeseler, A., & Kosiol, C. (2019). Polymorphism-aware species trees with advanced mutation models, bootstrap, and rate heterogeneity. *Molecular Biology and Evolution, 36*(6), 1294-1301. doi:10.1093/molbev/msz043

Simpson, G. G. (1944). *Tempo and mode in evolution*. New York: Columbia University Press.

Sun, Y., Wang, A., Wan, D., Wang, Q., & Liu, J. (2012). Rapid radiation of Rheum (Polygonaceae) and parallel evolution of morphological traits. *Molecular Phylogenetics and Evolution, 63*(1), 150-158. doi:10.1016/j.ympev.2012.01.002

Taylor, S. A., & Larson, E. L. (2019). Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nature Ecology & Evolution, 3*(2), 170-177. doi:10.1038/s41559-018-0777-y

Urban, S., Gerwin, J., Hulsey, C. D., Meyer, A., & Kratochwil, C. F. (2022). The repeated evolution of stripe patterns is correlated with body morphology in the adaptive radiations of East African cichlid fishes. *Ecology and Evolution, 12*(2), e8568. doi:10.1002/ece3.8568

Uyeda, J. C., & Harmon, L. J. (2014). A novel Bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. *Systematic Biology, 63*(6), 902-918. doi:10.1093/sysbio/syu057

Vanderpool, D., Minh, B. Q., Lanfear, R., Hughes, D., Murali, S., Harris, R. A., . . . Hahn, M. W. (2020). Primate phylogenomics uncovers multiple rapid radiations and ancient interspecific introgression. *PLoS Biology, 18*(12), e3000954. doi:10.1371/journal.pbio.3000954

Wang, Y., Cao, Z., Ogilvie, H. A., & Nakhleh, L. (2021). Phylogenomic assessment of the role of hybridization and introgression in trait evolution. *PLoS Genetics, 17*(8), e1009701. doi:10.1371/journal.pgen.1009701

White, M. A., Ane, C., Dewey, C. N., Larget, B. R., & Payseur, B. A. (2009). Fine-scale phylogenetic discordance across the house mouse genome. *PLoS Genetics, 5*(11), e1000729. doi:10.1371/journal.pgen.1000729

Wu, M., Kostyun, J. L., Hahn, M. W., & Moyle, L. C. (2018). Dissecting the basis of novel trait evolution in a radiation with widespread phylogenetic discordance. *Molecular Ecology*, *27*(16), 3301-3316. doi:10.1111/mec.14780

Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics, 19*(Suppl 6), 153. doi:10.1186/s12859-018-2129-y