UC Santa Barbara

UC Santa Barbara Previously Published Works

Title

An Evolving View of Character Macroevolution

Permalink

https://escholarship.org/uc/item/3z88j7jd

Journal

SYSTEMATIC BIOLOGY

ISSN

1063-5157

Authors

Tribble, Carrie M

Martinez-Gomez, Jesus

Rothfels, Carl J

et al.

Publication Date

2026

DOI

10.1093/sysbio/syag034

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial License, available at https://creativecommons.org/licenses/by-nc/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Systematic Biology
75th Anniversary

Syst. Biol. 00(0):1–25, 2026
© The Author(s) 2026. Published by Oxford University Press on behalf of the Society of Systematic Biologists. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site-for further information please contact journals.permissions@oup.com
https://doi.org/10.1093/sysbio/syag034
Advance Access publication April 10, 2026

An Evolving View of Character Macroevolution

Carrie M. Tribble [1,2,*,†], Jesús Martínez-Gómez [3,4,†], Carl J. Rothfels [5], and Michael R. May [5]

[1]*Department of Biology, University of Washington, Seattle, WA 98195, USA*
[2]*Burke Museum of Natural History and Culture, University of Washington, Seattle, WA 98195, USA.*
[3]*Department of Plant and Microbial Biology and Integrative Biology, University of California, Berkeley, CA 94720, USA*
[4]*Department of Ecology, Evolution and Marine Biology, University of California, Santa Barbara, CA 93106, USA*
[5]*Department of Biology and Ecology Center, Utah State University, Logan, UT 84322, USA*
[*]*Correspondence to be sent to: Department of Biology, University of Washington, Box 351800 Seattle, WA 98195, USA. E-mail tribblec@uw.edu*
[†]*Authors contributed equally.*

Received 1 September 2025; reviews returned 9 March 2026; accepted 16 March 2026
Associate Editor: Stacey Smith

Abstract.—Phenotypes serve as the interface between organisms and their environments and are thus pivotal for comprehensive biological understanding. However, comparative analyses of species' phenotypes must account for the non-independence of characters imposed by the branching pattern of macroevolution. Methods to account for this phylogenetic non-independence have historically been conceived of as their own subfield: phylogenetic comparative methods (PCMs). In this conceptualization, a researcher takes a pre-existing phylogeny and uses it to correct for the non-independence of the character data that they wish to analyze. Here, we argue that parallel developments in scientific philosophy, data availability, computational capacity, and model development have enabled a new paradigm of character evolution, where patterns of character evolution are seen as intrinsically related to the diversification process, and thus should be inferred jointly with the tree rather than reconstructed on an existing phylogeny in a two-step process. In the context of this paradigm shift, we review historical milestones in studies of character macroevolution and discuss major recent conceptual and methodological advances, with an emphasis on the opportunities and insights provided by the joint-inference perspective. We include primers on current topics in character evolution where joint inference is particularly effective, including: (1) state-dependent speciation and extinction models and the importance of cladogenetic change; (2) jointly modeling discrete and continuous characters; (3) accounting for hidden process variation in character evolution; (4) joint inference in divergence-time estimation; (5) joint inference of phylogeny and ancestral states; and (6) joint inference of alignment and phylogeny. The article concludes with a reflection on the future trajectory of these methods, emphasizing the interconnectedness of character evolution with broader processes in biology. [Alignment inference; ancestral-state estimation; Bayesian inference; Brownian motion; hidden-state models; indel models; joint inference; phylogenetic comparative methods; SSE models; stochastic character maps; structured Markov models; threshold model; total-evidence dating.]

The Evolution of the Study of Character Macroevolution

The birth of modern approaches to studying character evolution across species is often credited to Felsenstein (1985). Prior to this work, such comparative studies were typically ahistoric—they did not account for phylogenetic structure (reviewed in Huey et al. 2019). Felsenstein (1985) clearly articulated the problem of not accounting for phylogenetic relationships: trait data from species are not statistically independent because species share greater or lesser degrees of evolutionary history, as determined by the underlying phylogeny. Felsenstein (1985) also provided a solution to this problem by showing that the contrasts (differences in trait values between sister nodes) of a continuous trait evolving on a phylogeny are independent under a Brownian model of evolution. This approach of using contrasts, now commonly referred to as phylogenetic independent contrasts (PIC), became the basis of many empirical studies and is still used today. In order to apply PIC, one needs a pre-existing phylogeny with branch lengths, necessitating a two-step inference: First one infers a phylogeny (using a variety of methods) and then subsequently one can perform a comparative analysis correcting for the phylogeny in order to illuminate the focal biological feature of interest. We note that the use of the term "correcting for" in this context implies that the phylogeny is a source of error rather than a representation of the generative evolutionary process (Westoby and Leishman 1995; Desdevises et al. 2003; Beauchamp and Fernández-Juricic 2004; see Rohle 2006). We disagree with this implication, but use "correcting for" throughout this manuscript where appropriate to represent this viewpoint. There were other practical reasons for the initial prevalence of this two-step approach. Perhaps most critically, phylogenetics was in its infancy: we had very limited abilities to generate or simultaneously analyze heterogeneous data sets (basic partitioned analyses, where some parameters are independently estimated for different data subsets, had not been developed; see, e.g., de Queiroz et al. 1995), and few tenable alternatives to the two-step approach existed.

Downloaded from https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syag034/8651289 by guest on 21 May 2026

Further supporting these practical motivations for a two-step procedure were broader philosophical and conceptual concerns. Specifically, it was considered a cardinal rule that one should "never use the characters that are part of the evolutionary hypothesis under investigation to build your phylogenetic tree" (Brooks and McLennan 1991). In this view, the two-step procedure is not just a practical necessity, it is the optimal approach—to do otherwise and include the data relating to the focal comparative hypothesis in the tree inference itself was considered to either be circular (Coddington 1988) or to violate the requirement of character independence necessary for proper hypothesis testing in the first place (see Box 1; Armbruster 1992; Frumhoff and Reeve 1994). As argued by Luckow and Bruneau (1997), some of this concern about circularity and non-independence was due to conflating two potential problems: (1) real statistical concerns regarding character independence during inference (an assumption of most models) and (2) philosophical concerns regarding independence between the tree and the focal hypothesis (which is not possible, hence correcting for the phylogeny in the first place). However, even given that distinction, both the circularity and independence-related arguments were made in the context of a period of history when phylogenetic data were scarce and issues of bias introduced by potential non-independence among characters were of greater concern: if one's data set comprises only a few dozen hard-earned characters, correlation (non-independence) among a handful of those characters can introduce significant bias. Phrased differently, even if a researcher believed that the inclusion of the focal data might be expected to produce a more accurate phylogeny in general, they might still be concerned that correlated patterns of stochastic variation in those (non-independent) characters might bias the phylogenetic inference toward their hypothesis of choice (Frumhoff and Reeve 1994; de Queiroz 1996). However, we note that these philosophical stances were not necessarily shared by those studying molecular evolution (e.g., Yang 1998), as they commonly used the same sequence data to infer a phylogeny as to make inferences about the sequence evolution (e.g., rates of evolution, tests of selection, ancestral-sequence inference) when species trees were not readily available.

This two-step paradigm is still dominant—approaches based on this paradigm are what we now call phylogenetic comparative methods (PCMs)—with major texts (and research communities) focused on either phylogeny inference (e.g., Felsenstein 2003) or the subsequent application of those phylogenies to studies of trait evolution (e.g., Harmon 2018; Revell and Harmon 2022). In part, this continuing dichotomy is likely due to—or at least reinforced by—the ease of teaching tree estimation and the inference-using-trees separately, and to the practical difficulties in developing and using joint models, but it also reflects the historical idea that the inference of phylogeny and phylogenetic comparative methods are separate enterprises (e.g., Kapli et al. 2020). However, conceptual developments and progress in the areas of data availability, model complexity, and computational power have reduced our dependence on this classic two-step approach and enabled the application of joint inference to many phylogenetic comparative questions.

First, greater conceptual clarity, and a move to embrace phylogenetics as an inferential science rather than a necessarily hypothetico-deductive one, mitigates the circularity concerns (see Box 1; Luckow and Bruneau 1997; Ronquist 2004). For example, it is more natural, within an inferential framework, to include all data that might inform that inference in order to derive as accurate an inference as possible, whereas a hypothetico-deductive framework lends to greater concern about conservativeness ("reliability" *sensu* de Queiroz 1996) with respect to the focal hypothesis. Second, from a practical perspective, the ability to apply joint inference has been facilitated by increased data availability in the form of molecular data sets—practical concerns about the influence of a few characters are reduced when one's data set totals thousands of characters. Complementing both the conceptual and data-availability trends are increases in model availability, particularly in the form of likelihood-based approaches (e.g., Schluter et al. 1997; Pagel 1999b; Lewis 2001; Butler and King 2004) and data partitioning (e.g., Nylander et al. 2004; Brown and Lemmon 2007), and rapid increases in raw computational power. In many contexts, researchers now have the tools at their disposal to perform joint inference in a reasonable way and fewer practical or philosophical reasons to prefer the two-step approach. This intersection of increased data availability, modeling options, and computational power also facilitated the development of Bayesian inference methods (Rannala and Yang 1996; Yang and Rannala 1997; Huelsenbeck and Bollback 2001; Huelsenbeck et al. 2002), which lead naturally to more integrative approaches: a Bayesian framework provides a clear means of incorporating multiple types of data (see the *Joint Inference and the Bayesian Paradigm* section below).

Box 1. The influence of hypothetico-deductive reasoning in comparative phylogenetics

The shift from understanding character evolution via two-step PCMs to deriving understanding from joint inference connects to a related shift in our conception of the nature of phylogenetic inquiry in general. The early period of systematics was influenced by the philosopher of science, Karl Popper (reviewed in Helfenbein and DeSalle 2005). Phrased simplistically, Popper held that science differs from metaphysics in that science operates by deductive reasoning: it produces predictions that can be fal-

Downloaded from https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syag034/8651289 by guest on 21 May 2026

sified by the formal rules of logic. Specifically, the *modus tollens* (confirming the consequent) deductive structure states that if it is true that when A holds, B also holds (i.e., "if A, then B"), and we observe something other than B (i.e., "not B"), then we can conclude that A is not true (i.e., "therefore, not A"), and it is only through this falsification that science moves forward (Popper 1959). Although Popper's *The Logic of Scientific Discovery* (1959) also touches on probability theory, much of the systematics community responded to Popper's emphasis on deductive reasoning with extensive debates as to whether particular methodologies (i.e., parsimony) better fit Popper's logical formalism and thus were more scientific than the alternative (i.e., likelihood) (e.g., Siddall and Kluge 1997; de Queiroz and Poe 2001; Faith and Trueman 2001; Farris et al. 2001; see discussion in Rieppel 2003; Helfenbein and DeSalle 2005).

Popper's ideas were also influential in promoting the view that formal hypothesis-testing approaches (i.e., rejecting a null hypothesis at a particular *p*-value) are superior to purely inferential (estimation-based) alternatives. The resulting hypothetico-deductive focus led to many comparative studies being explicitly framed in the context of falsifiable hypotheses. For example, in discussing whether it is appropriate to include the comparative data related to one's phenomenon of interest in the underlying phylogenetic inference, de Queiroz (1996) focused specifically on the question of whether those data could bias the conclusions toward the focal hypothesis. If any bias ran counter to the focal hypothesis then the analysis was considered reliable; the question of accuracy was secondary (de Queiroz 1996).

Once a dominant paradigm in systematics (e.g., see Olmstead 2001), the influence of Popperianism has declined, due in part to arguments that it is inapplicable to historically contingent fields like phylogenetics (Sober 1991; Rieppel 2003; Vogt 2008), and to the rise of unabashedly inferential Bayesian approaches (Ronquist 2004). Although we still have an awkward tendency to refer to phylogenies as *hypotheses* (suggesting, perhaps, an unspecified null that we were able to reject?), we are also more likely to recognize that they are *estimates*.

We argue, given the availability of joint inference approaches, for a subtle paradigm shift in studies of character evolution: phylogenies are not static structures upon which to map our characters of interest. Instead, all our data evolved together and we should not distinguish between data we use for phylogenetic inference and data we use for phylogenetic comparative questions. Indeed, in the study of molecular evolution, practitioners will often estimate the phylogeny of their sequences and the parameters of the model of sequence evolution simultaneously (Yang 1998). From a modeling perspective, there is nothing conceptually different between molecular sequences and traits, which suggests that joint inference of phylogeny and trait models is conceptually coherent.

Neither are phylogenies a "nuisance" source of non-independence, and phylogenetic signal is not something we should seek to "correct" via a two-step approach. The classic view of phylogenetic signal in PCMs is that the distribution of trait values may be more or less determined by the structure of the (pre-existing) phylogeny, and further that if this phylogenetic signal (or "phylogenetic effect"; Losos 2011) is low, the phylogeny is not needed for tests of correlated traits (Pagel 1999a; Blomberg et al. 2003; Losos 2011) and the inclusion of phylogenetic information may even be inappropriate (Losos 2011). We suggest that the lack of phylogenetic signal is simply an indication that our data are poorly explained by the assumed model, that the data are evolving too quickly for the phylogeny to matter very much, or that the phylogeny is a bad fit for the data. Joint inference may help disentangle these various scenarios by simultaneously quantifying model fit, parameter uncertainty, and phylogenetic informativeness.

Related to this argument for joint inference is our broader argument in favor of greater emphasis on inference (estimation) in general, versus focusing on hypothetico-deductive approaches. This shift mirrors recent trends in the statistical community toward an emphasis on effect sizes (parameter values) rather than relying on *p*-values (Wasserstein and Lazar 2016). An emphasis on estimation also relieves us of some practical and conceptual difficulties associated with the hypothesis-testing framework. For one thing, hypothesis testing requires that we be ignorant of our data before performing the test (e.g., see Goldman et al. 2000), a condition that is rarely met in phylogenetics. Perhaps more fundamentally, much of what we might be interested in learning about character evolution does not naturally fit within a formal hypothetico-deductive framework. The number of times a trait evolved, or the degree to which a trait is correlated with a particular other trait, can be framed as formal hypotheses, but only awkwardly; these are fundamentally probabilistic questions—not logical-deductive ones—that are accessible through the tools of estimation rather than falsification. Note that an argument in favor of embracing estimation is *not* an argument against making careful observations of the natural world, using those observations to formulate questions (hypotheses in the informal sense), and then designing a study (either estimation or hypothesis-testing based) to investigate those questions. Indeed, we feel that that approach to understanding character evolution is particularly compelling.

In this review we advocate for the conceptual and practical advantages of the joint-inference paradigm,

Downloaded from https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syag034/8651289 by guest on 21 May 2026

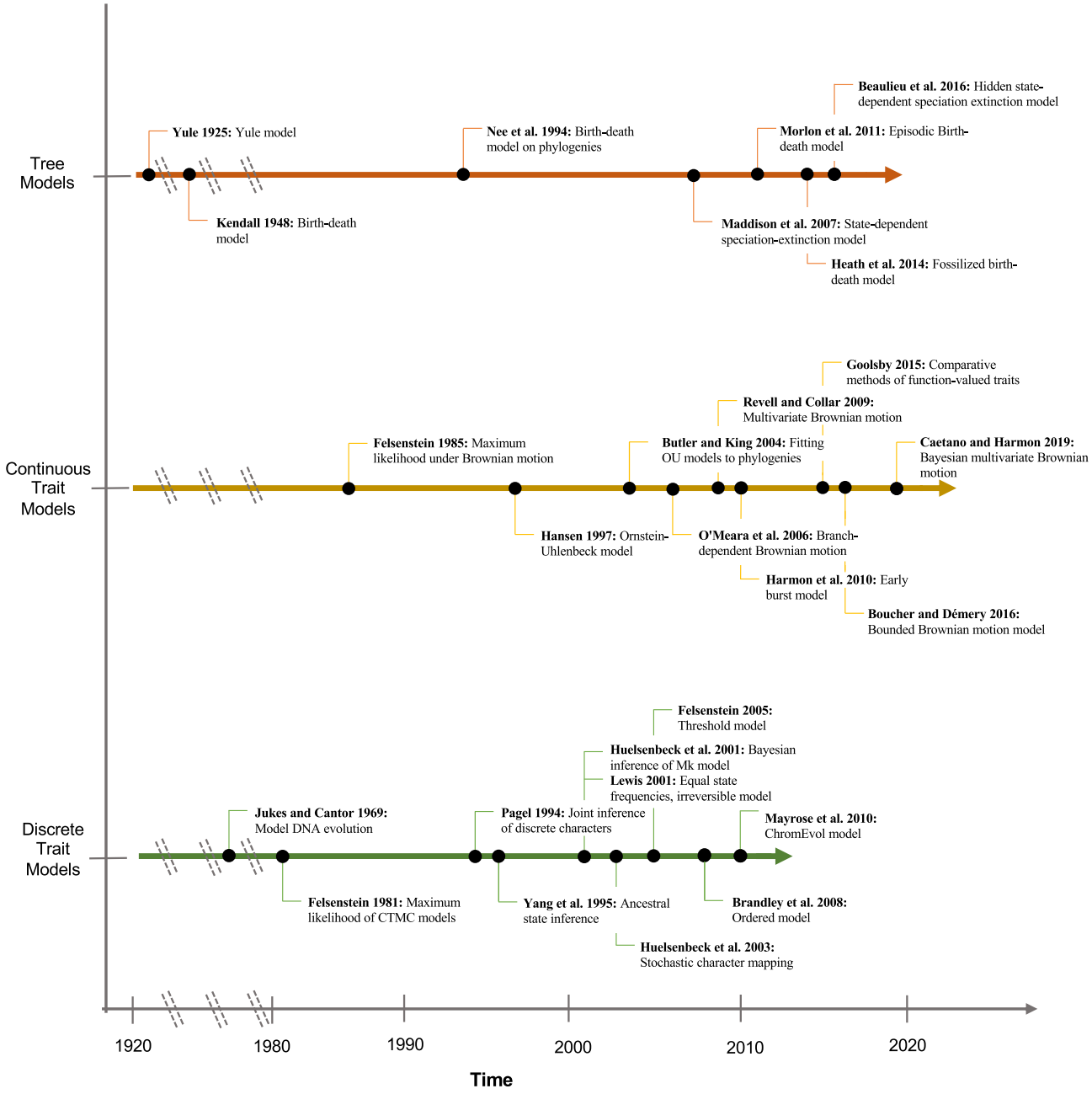

FIGURE 1. Timeline of select major events in the development of phylogenetic comparative methods, sorted by discrete trait models, continuous trait models, and tree models. We elaborate on some of these methods in the main text.

while recognizing that, in many cases, results from joint inference and a two-step approach will be effectively identical (e.g., inclusion of a comparative data set is unlikely to influence estimates of topology derived from genomic data). We first provide a brief timeline (Fig. 1) of significant developments in the study of character evolution and a summary of the fundamental building blocks of character-evolution models. We then highlight a set of major current topics where the application of joint inference is particularly effective and conclude with a discussion of potential future directions for the study of character evolution.

Joint Inference and the Bayesian Paradigm

The explicitly inferential nature of Bayesian inference lends itself to joint estimation (compared with hypothetico-deductive approaches; see Box 1). However, Bayes' theorem also provides a formalization demonstrating the importance of joint inference: if two

Downloaded from https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syaf034/8651289 by guest on 21 May 2026

data sets share aspects of the same generating process (e.g., they share a phylogenetic history), then accurate inference *depends on* the two data sets jointly informing model parameters.

To illustrate this point, imagine we have two data sets, X_1 and X_2. Perhaps these data sets correspond to different DNA sequence alignments, or different morphological characters, or some mix of genetic and morphological data. Suppose we are interested in how they share information about some underlying variable of interest (e.g., a phylogenetic tree). To that end, we specify a model that describes how both X_1 and X_2 evolve together (i.e., a joint model) with parameters θ. The parameters θ represent one or more unknown variables that we hope to estimate (e.g., phylogeny or rate of morphological change), while the model describes how specific values of those parameters affect the probabilities of different possible observed data sets.

The joint probability of X_1 and X_2 under this model—given specific values of the parameters θ—is $P(X_1, X_2 \mid \theta)$, conventionally called the *likelihood function*. In a Bayesian context, the *joint posterior distribution* of θ is

$$\overbrace{P(\theta \mid X_1, X_2)}^{\text{posterior}} = \frac{\overbrace{P(X_1, X_2 \mid \theta)}^{\text{likelihood}}\overbrace{P(\theta)}^{\text{prior}}}{\underbrace{P(X_1, X_2)}_{\text{marginal likelihood}}} \quad (1)$$

where $P(\theta)$ is our belief about the specific values of θ before observing the two data sets, and $P(\theta \mid X_1, X_2)$ is our posterior belief about the specific values of θ after observing the two data sets.

An alternative, and equivalent, approach to joint inference is *Bayesian sequential inference*. Here, we also analyze the two data sets, one at a time, treating the posterior from the first data set as the prior for the second data set:

$$P(\theta \mid X_1) = \frac{P(X_1 \mid \theta)P(\theta)}{P(X_1)} \quad \text{(prior updated by first dataset)}$$

$$P(\theta \mid X_1, X_2) = \frac{P(X_2 \mid X_1, \theta)P(\theta \mid X_1)}{P(X_2 \mid X_1)} \quad \text{(prior updated by second dataset)}$$

This sequential procedure is mathematically equivalent to doing a joint analysis because

$$\begin{aligned} P(\theta \mid X_1, X_2) &= \frac{P(X_2 \mid X_1, \theta)P(\theta \mid X_1)}{P(X_2 \mid X_1)} && \text{(from above)} \\ &= \frac{P(X_2 \mid X_1, \theta)}{P(X_2 \mid X_1)} \times \frac{P(X_1 \mid \theta)P(\theta)}{P(X_1)} && \text{(using posterior from step 1 as prior)} \\ &= \frac{P(X_1, X_2 \mid \theta)P(\theta)}{P(X_1, X_2)} && \text{(combining terms)} \end{aligned}$$

where the last line follows because $P(A, B) = P(A \mid B)P(B)$. Notice that the final posterior distribution of the Bayesian sequential analysis is identical to the posterior of the joint analysis (Equation 1). The equivalency of the Bayesian sequential and joint-inference approaches, however, only holds when the second step of our inference updates our belief about θ (the full posterior of the first step becomes the prior in the second, where it is updated by the data in X_2).

The logic of Bayesian sequential analysis also implies that if we study the evolution of a trait on a phylogeny, that trait should in turn affect our estimate of the tree itself, as well as the other parameters of the model. If we instead fix the estimates from the first step (as in conventional PCMs), we may obtain biased estimates of the posterior. Rephrasing, and pulling up from the formalism of Bayes theorem, if we believe that different data sets evolved following some sort of shared process (such as evolving on the same phylogeny), then accurate inferences of that process require that one data subset be allowed to update our inferences from the other data subsets and vice versa: in principle, joint inference (or its mathematical equivalent, Bayesian sequential inference) is required.

Bayesian sequential analysis provides a powerful alternative to simultaneous joint inference by allowing researchers to perform portions of their inference at different times or across different computers, effectively breaking one large problem into easier-to-handle pieces. For example, comparative biologists have long recognized the value of incorporating phylogenetic uncertainty, which has motivated procedures that average parameter estimates uniformly over a sample of trees (e.g., Pryer et al. 2004; Organ et al. 2011)—a procedure that is at least spiritually Bayesian—or sampling over a distribution of trees using Markov chain Monte Carlo (MCMC) so that parameter estimates are averaged in proportion to the updated posterior probability of each tree (i.e., a true Bayesian sequential approach; Pagel and Lutzoni 2002; Pagel and Meade 2006). However, there is more to the joint (/Bayesian sequential) paradigm than accommodating phylogenetic uncertainty. Joint phylogenetic models unlock a broad range of interesting evolutionary questions, for example: How do changes in one trait affect the process of evolution for another trait? How do traits shape the processes of diversification? And how are extant and extinct lineages related? We will explore these interesting questions below, following a description of the fundamental components of most models used to study character evolution.

Fundamental Models of Trait Evolution

Phylogenetic comparative models—whether applied on a fixed topology or in a joint inference—are complex and diverse. In this section we first describe at a high level how phylogenetic comparative models of evolution work, by way of simulation. We then review a selection of major models and their extensions (also see O'Meara 2012; Baum and Smith 2013; Harmon 2018; Cornwallis and Griffin 2024); these models are the main building blocks for macroevolutionary study of character evolution.

Downloaded from https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syag034/8651289 by guest on 21 May 2026

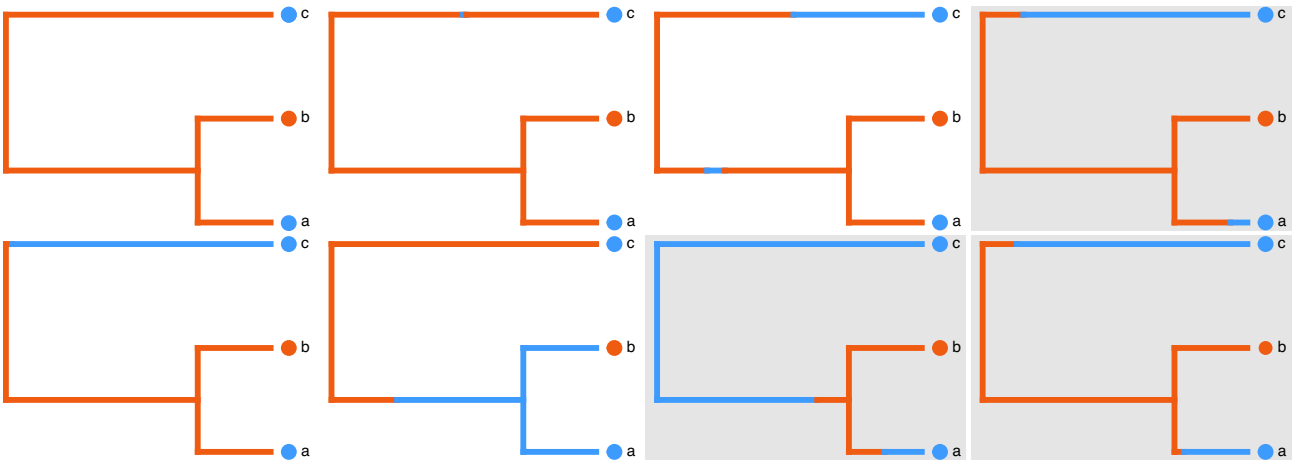

FIGURE 2. Simulated evolution of a two-state character (the states are red and blue) on a three-tip phylogeny. Each panel represents a single simulation. The gray panels are simulations where the character histories match the observed tip states (tips a and c are blue; tip b is red). Circles indicate observed data.

PCMs are typically based on stochastic (random) models of evolution. Generically, a stochastic model describes how a state variable at time t, denoted $X(t)$, changes over time; the process of change is governed by a set of parameter values denoted θ. In a traditional PCM study, we often assume that the phylogeny is known, and denote the state variable for branch i at time t as $X_i(t)$. The state variable may be a continuous trait value, such as body size, or a discrete trait value, such as fruit color. However, the state variable may also be more complicated; for example, $X_i(t)$ may be a set of continuous and/or discrete traits that are co-evolving.

The main utility of a phylogenetic stochastic model is that it allows us to compute the probability of observing a particular set of character states at the tips of the tree (the observed character data) given a set of parameter values θ—this probability of the data given the model is the likelihood of the model. Thus, the stochastic model allows us to estimate the parameter values θ in a model-based inference framework (e.g., maximum likelihood or Bayesian inference). Although the details of computing these probabilities are different for different models, in general we can understand how they work using a simulation analogy.

The simulation procedure works as follows: First, we choose an initial state at the root of the tree according to a root-state distribution; in theory we could draw from any distribution but often we start our simulation with biologically informed values from a known distribution (e.g., a known ancestral state or the stationary frequency of the model). Next, we simulate the evolution of the trait according to the stochastic model with parameter values θ along the branches of the tree until we reach the present. Finally, we collect the states at the tips of the tree to serve as the outcome. If we repeat this procedure many times, the fraction of outcomes that match the observed data is the probability of the data given the model, otherwise known as the *likelihood* (of the model). The *ancestral-state estimate* at a particular node reflects the fraction of times that node is in each state among outcomes that match the tip data (see also the section, below, on joint inference of phylogeny and ancestral character states). Likewise, the histories of change along branches for outcomes that match the observed data are called *stochastic character maps*, which can be used to estimate the number of changes across the tree, or the amount of time each branch spends in each state, etc. Although this simulation analogy is useful for conceptualizing likelihood, in practice, we typically calculate these probabilities analytically using mathematical functions rather than relying on computationally intensive simulations.

See Figure 2 for an example of this simulation procedure for a discrete-state model with two states (red and blue). In this case, the observed data are that taxon a is blue, whereas taxa b and c are red. The fraction of outcomes matching the observed data (indicated in the gray boxes) is the probability of the data (in this case, 3/10) given the model and a specific set of parameter values. These simulations are also useful for estimating ancestral states and stochastic maps. The common ancestor of a and b is inferred to be red with probability 1 (3 of 3 outcomes), whereas the root of the tree is inferred to be red with probability 2/3. The probability that blue evolved twice (based on the stochastic maps) is 2/3 (Fig. 2). This same basic logic also applies to continuous trait models.

Discrete Trait Models

Models of evolution of discrete characters (characters with discrete character states) have their origin in models of nucleotide substitution (Jukes and

Downloaded from https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syag034/8651289 by guest on 21 May 2026

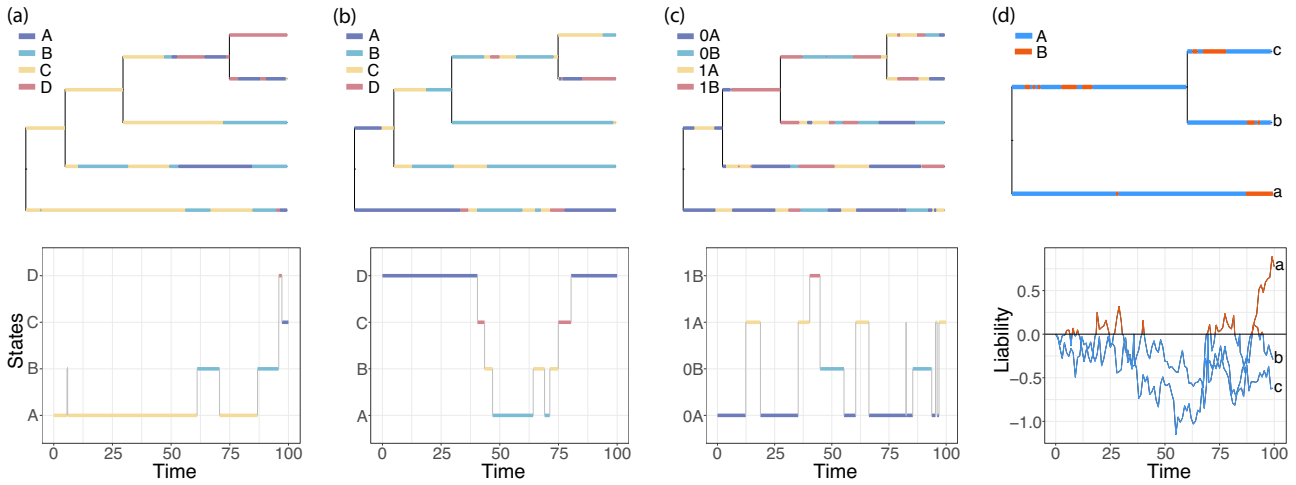

FIGURE 3. Simulated discrete trait evolution under various models: (a) symmetrical evolution, (b) ordered evolution where only transitions between adjacent states are possible (e.g., $A \leftrightarrow B \leftrightarrow C \leftrightarrow D$), (c) dependent evolution of two binary traits, and (d) the threshold model. For all models, the top row shows a stochastic map of the discrete trait history and for panels a–c the bottom row shows a single sample history—trait value plotted as a function of time. The bottom of panel d shows the sample history for all three branches of the tree.

Cantor 1969; Neyman 1971; Farris 1973; Cavender 1978; Kaplan and Langley 1979) and were first applied to infer phylogenies by Felsenstein (1973, 1981). These models are continuous-time Markov chains (CTMCs)—models where the probability of changing to a new state depends solely on the current state (and on the parameter values of the model) rather than being influenced by past states—that can be generally referred to as Mk models, where k signifies the number of states (Lewis 2001). These models can be represented by an *instantaneous rate matrix*, Q, that describes the instantaneous rate of change between states. The rows of a Q matrix represent the starting states and the columns represent the ending states, and elements of the matrix represent the instantaneous rate of change between those particular states. For a simple, symmetric two-state model (with states A and B), the Q matrix is

$$Q = \begin{bmatrix} - & q \\ q & - \end{bmatrix},$$

where the diagonal elements (indicated by $-$) are such that the rows of the matrix sum to zero. Therefore, in the Q matrix above, the transition rate from $A \rightarrow B$ and the transition rate from $B \rightarrow A$ both occur at the instantaneous rate q.

Common extensions of the core Mk model allow transition rates to vary. When transition rates are not equal, the rate of change between state i and state j is typically denoted q_{ij}. A symmetrical model (often referred to as SYM; Zharkikh 1994) fixes the transition rates to and from a given pair of states to be equal (i.e., $q_{ij} = q_{ji}$; see Fig. 3a,b for a four-state example), whereas the all-rates-different (ARD) model (Paradis et al. 2004) allows each transition rate to vary freely. Applying these models to a trait with three states (A, B, and C) results in these instantaneous rate matrices:

$$Q_{\text{SYM}} = \begin{bmatrix} - & q_{AB} & q_{AC} \\ q_{AB} & - & q_{BC} \\ q_{AC} & q_{BC} & - \end{bmatrix}$$

$$Q_{\text{ARD}} = \begin{bmatrix} - & q_{AB} & q_{AC} \\ q_{BA} & - & q_{BC} \\ q_{CA} & q_{CB} & - \end{bmatrix}$$

It is also possible to specify models where some transition rates are set to zero, to reflect certain types of changes that may be biologically impossible. In some cases, this approach can result in irreversible models, where once a particular transition has occurred, the character can no longer revert to the original state (see Goldberg and Igić 2008 and references therein), for example,

$$Q_{\text{IRR}} = \begin{bmatrix} - & q_{AB} & q_{AC} \\ q_{BA} & - & q_{BC} \\ 0 & 0 & - \end{bmatrix},$$

where it is impossible to evolve out of state C once it is reached.

Ordered Markov models (e.g., Brandley et al. 2008) allow transitions between adjacent states, whereas all other transition rates are zero. These types of models are often applied to traits that have a natural ordering, for example the loss of teeth (Paluh et al. 2021) or the evolution of the number of vertebrae (Spear et al. 2023). Typically, upward transitions occur at one rate (the rate of gain), and downward transitions occur at another rate (the rate of loss), though this is by no means required. An example rate matrix for an ordered trait with four

Downloaded from https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syag034/8651289 by guest on 21 May 2026

states (1, 2, 3, and 4) would be

$$Q_{\text{ORD}} = \begin{bmatrix} - & \lambda & 0 & 0 \\ \mu & - & \lambda & 0 \\ 0 & \mu & - & \lambda \\ 0 & 0 & \mu & - \end{bmatrix},$$

where λ and μ are the gain and loss rates, respectively.

A bespoke example of an ordered Markov model is the ChromEvol model of chromosome evolution (Mayrose et al. 2010). Chromosome evolution is governed by a variety of processes including chromosome gain, loss, and polyploidy. For simplicity, we show an example Q matrix that describes the evolution of a group with four different chromosome numbers (1, 2, 3, and 4):

$$Q_{\text{Chromo}} = \begin{bmatrix} - & \lambda + \rho & 0 & 0 \\ \mu & - & \lambda & \rho \\ 0 & \mu & - & \lambda \\ 0 & 0 & \mu & - \end{bmatrix}$$

Here, λ and μ represent chromosome gain and loss, and ρ represents polyploidization events. This general ChromEvol model then provides the foundation for additional elaborations tailored to specific questions (e.g., Zenil-Ferguson et al. 2017; Freyman and Höhna 2018; Mayrose and Lysak 2021; Tribble et al. 2025).

The Mk model framework can also be used to study the joint evolution of two discrete characters. This approach is commonly used to test if traits are evolutionarily correlated (Fig. 3c; Pagel 1994). For example, we can have traits each with two states: Trait One with states 0 and 1, and Trait Two with states A and B. One way to jointly model the evolution of these traits is to combine the two binary characters into a single character with four states ($0A$, $1A$, $0B$, and $1B$):

$$Q_{\text{Dep}} = \begin{bmatrix} - & q_{0A,1A} & q_{0A,0B} & 0 \\ q_{1A,0A} & - & 0 & q_{1A,1B} \\ q_{0B,0A} & 0 & - & q_{0B,1B} \\ 0 & q_{1B,1A} & q_{1B,0B} & - \end{bmatrix}$$

Note that many elements of this rate matrix are 0 because, by assumption, only a single trait can change at a time. Dependent evolution is inferred by comparing transition rates for one character as a function of the state of the other character: if $q_{0A,1A}$ is inferred to be higher than $q_{0B,1B}$, then transitions from 0 to 1 occur faster when the second character is in state A rather than B. Although this model is commonly used, it was shown to fail under cases of phylogenetic pseudoreplication—where a trait is inferred to only evolve once (Maddison and FitzJohn 2015). Under these scenarios we should be unable to test for correlated evolution as there should be no statistical power in a single event, yet this method—and other methods—will find evidence for correlated evolution (Uyeda et al. 2018; Gardner and Organ 2021).

The flexibility of the Q matrix allows users to specify their own model of trait evolution informed by prior knowledge of their group or to formulate and test specific hypotheses of evolutionary rates. For example, structured Markov models, implemented in software such as `PARAMO` and `ontophylo` (Tarasov et al. 2019; Porto et al. 2024), allow researchers to include developmental (ontogenetic) constraints in their character evolution models (e.g., Tarasov 2019; Tribble et al. 2023) or to properly account for the difference between character states that are missing because they are unknown rather than being inapplicable to the specific taxon in question (Tarasov 2023; see Maddison 1993).

Threshold Model

The threshold model (Wright 1934) represents discrete trait evolution via an unobserved continuous character—when the trait value of the continuous character crosses a particular threshold, the observed discrete character changes state (Fig. 3d). The threshold model was inspired by models in quantitative genetics, where additive variation of an (unobserved) polygenic trait can contribute to discrete changes in phenotype. Wright focused on the development of guinea pig toes, which is determined by an underlying physiological character that varies approximately continuously. If that continuous character—referred to as the liability—crosses some threshold, the guinea pig will develop an additional toe; below that threshold, the normal number of toes develops. Felsenstein (2005) adapted this model to macroevolution, permitting the modeling of discrete trait evolution (such as number of toes) across a phylogeny by assuming that the liability, typically unobserved, evolves via Brownian motion (BM, see below) with rate σ^2 or some other continuous-trait model (Fig. 3d).

Generally, we are interested in inferring the parameters of the liability (e.g., BM parameters) as well as the placement of the threshold, with some simplification. Because the liability at the tips is not observed (unlike a typical continuous trait model; Fig. 4), we lack any information of the liability scale (e.g., does the liability range from 0−1 or 5−100). As such, we can fix the value of σ^2 (see the *Continuous Trait Models* section, Felsenstein 2005, and Revell 2014 for further discussion). Furthermore in a simple two-state case, where there is a single threshold, we can arbitrarily set the threshold value, typically to 0 (Felsenstein 2005; Revell 2014; Cybis et al. 2015). In cases with $n > 2$ states, the liability value of $n - 1$ thresholds needs to be inferred (Revell 2014; Cybis et al. 2015).

Continuous Trait Models

Most models of continuous trait evolution are based on BM (a so-called random walk), introduced to systematics by Felsenstein (1985). The BM model is characterized by a single parameter, σ^2 (sometimes called a diffusion parameter; Fig. 4a). Under BM, at any given point in time an increase in the trait value is as likely as a decrease—the trait value wanders (hence random

Downloaded from https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syag034/8651289 by guest on 21 May 2026

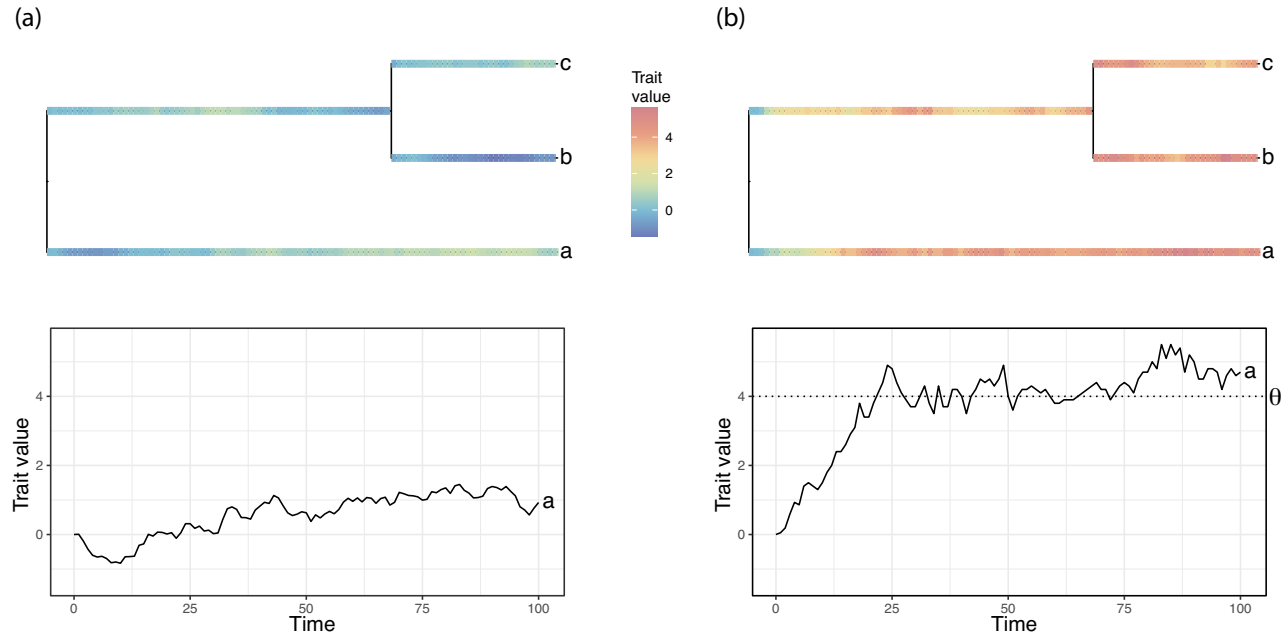

FIGURE 4. Continuous trait simulated under (a) Brownian motion and (b) an OU process. For both panels a and b, the top row shows stochastic maps of the ancestral values of the traits along the phylogeny, whereas the bottom row shows a sample history—the trait value as a function of time—for the branch leading to species *a*. θ indicates the optimum value for the OU model in panel b.

walk). After a set amount of time, t, the new values will be normally distributed with a mean equal to the starting value X_0 and the variance determined by $t\sigma^2$.

A common variation on BM is the Ornstein–Uhlenbeck (OU) model, which describes the evolution of a trait toward some optimum value θ (Hansen 1997; Butler and King 2004; Fig. 4b). Like BM, the OU model assumes normally distributed changes along branches of a phylogeny described by σ^2, but it also includes two additional parameters, the aforementioned optimum value, θ, and the magnitude of the pull toward θ, denoted α. Unlike BM, where variance grows without bound as $t \to \infty$, under the OU model the variance will converge to $\frac{\sigma^2}{2\alpha}$. Larger values of α result in less variance as the strength of the pull is stronger.

The OU model is only one of many elaborations of the BM model; there are many others including bounded BM (Boucher and Démery 2016), those that model early bursts (Blomberg et al. 2003; Harmon et al. 2010), or those that model jumps in trait space (a Lévy process, Landis et al. 2013). These models are often motivated by attempting to capture a unique evolutionary process. An additional elaboration is to allow for heterogeneity in the parameter values across the tree, allowing them to vary in a branch-dependent manner (O'Meara et al. 2006; Thomas et al. 2006; Uyeda and Harmon 2014; Khabbazian et al. 2016).

This class of models can also have multivariate extensions that aim to jointly estimate the evolution of two or more continuous traits. For example, the multivariate BM model (Revell et al. 2008; Caetano and Harmon 2019) is commonly used. Similar to a univariate BM model, under a multivariate BM model each trait has its own diffusion parameter (e.g., σ_1^2 and σ_2^2), but each pair of traits will also have an evolutionary covariance parameter, σ_{12}, that indicates the degree and nature of correlated evolution between traits. If $\sigma_{12} = 0$, the traits evolve independently, equivalent to individually modeling each trait in a univariate framework. For $\sigma_{12} > 0$, if X_1 increases X_2 will tend to increase as well, and for $\sigma_{12} < 0$, if X_1 increases X_2 will decrease. In these cases the traits are said to be evolutionary correlated.

Finally, recent developments have provided reasonable alternatives to BM-based models for continuous traits for the first time. For example, Blomberg et al. (2020) introduce non-Gaussian stochastic diffusion models, which may better capture underlying evolutionary processes than the Gaussian models discussed above and provide clearer biological interpretations of parameter estimates.

Birth-Death Tree Models

The birth-death process models the dichotomous branching process that produces phylogenetic trees, but birth-death models and their derivatives have also become prominent in the study of trait evolution. Birth-death processes are continuous-time Markov chains where, at any point in time, each lineage has a certain rate, λ, of producing (birthing) a new lineage (a speciation event that results in a node in the phylogeny) and a certain rate, μ, of going extinct (death; Nee et al.

Downloaded from https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syag034/8651289 by guest on 21 May 2026

TABLE 1. Examples of joint methods

Model and example question	Implementation(s)	Empirical examples
Threshold model: Are continuous and discrete traits correlated?	`R::threshBayes` (Revell 2012), `Threshml` (Felsenstein 2012), `BEAST` (Cybis et al. 2015), `R::mcmcglmm` (Hadfield 2015)	Ogburn and Edwards (2015), Iles et al. (2017), Esquerré et al. (2019), and Suissa and Friedman (2022)
hOUwie: Are continuous and discrete traits correlated?	`R::OUwie` (Boyko, O'Meara, et al. 2023)	Boyko, Hagen, et al. (2023) and Sperotto et al. (2025)
MuSSCRat model: Do rates of continuous-trait evolution depend on a discrete trait?	`RevBayes` (May and Moore 2020)	Burress et al. (2020), Wolff et al. (2021), Corn et al. (2022), Garland et al. (2025), Liu et al. (2025). RevBayes tutorial.
SSE models: How do character states affect diversification rates? What is the history of a character that might affect diversification rates?	`Mesquite` (Maddison et al. 2007), `R::diversitree` (FitzJohn 2012), `R::castor` (Louca and Pennell 2020b), `RevBayes` (Freyman and Höhna 2018), `TensorPhylo` (May and Meyer 2025)	Goldberg et al. (2010), Mayrose et al. (2011), Beaulieu and Donoghue (2013). R::diversitree tutorial; RevBayes tutorial.
HiSSE models: How do I infer trait-related diversification dynamics while accommodating process variation?	`R::HiSSE` (Beaulieu and O'Meara 2016), `RevBayes` (Höhna et al. 2016; Tribble et al. 2025)	Ng and Smith (2018), Xue et al. (2020), Meseguer et al. (2022), Tribble et al. (2025). R::HiSSE tutorial; RevBayes tutorial.
TED models: How are living and extinct lineages related and what is the timescale of evolution in my group?	`MrBayes` (Ronquist et al. 2012), `BEAST` (Bouckaert et al. 2019; Baele et al. 2025), `RevBayes` (Höhna et al. 2016)	Lee (2016), Gavryushkina et al. (2017), May et al. (2021), Tejada et al. (2024). BEAST 2 tutorial; RevBayes tutorial.
Ancestral-state reconstruction: What is the evolutionary history of a trait on an unknown phylogeny?	`BayesTraits` (Pagel et al. 2004), `BEAST` (Bouckaert et al. 2019; Baele et al. 2025), `RevBayes` (Höhna et al. 2016; Freyman and Höhna 2019)	Pagel and Lutzoni (2002), Lemey et al. (2009), Braga et al. (2020), Nyathi et al. (2025). BayesTraits tutorial; BEAST 2 tutorial; RevBayes tutorial.
Indel models: Can I estimate my phylogeny without conditioning on a single estimate of the alignment? What is the ancestral sequence of my gene of interest?	`BAli-Phy` (Suchard and Redelings 2006; Redelings 2021), `BayesCAT` (Shim and Larget 2018), `BEAST` (Lunter et al. 2005)	Gaya et al. (2011), Westesson et al. (2012), Redelings (2014), Borstein et al. (2024), Sharma (2025). BAli-Phy tutorial.

1994). These models produce probability distributions of trees. There are many variations of birth-death tree models, for example time-dependent models that allow speciation and/or extinction rates to vary over time (e.g., Morlon et al. 2011), or fossilized birth-death (FBD) models, which model fossil recovery alongside speciation and extinction (Heath et al. 2014), and are thus useful for inferring time-calibrated phylogenies from data sets that include extinct samples (see the *Joint Inference in Divergence-Time Estimation* section, below). Some birth-death models are non-identifiable, as multiple sets of parameter values produce the same likelihood (Kubo and Iwasa 1995; Louca and Pennell 2020a). We direct interested readers to Title et al. (2025) for a more in-depth explanation of these methods and their limitations. One of the most prominent extensions of the birth-death process links speciation and extinction rates to character states (state-dependent speciation and extinction [SSE] models; Maddison et al. 2007). We review these methods and other current areas in the study of character evolution in more detail below.

Major Areas in the Application of Joint Inference to Studies of Character Evolution

In this section we highlight areas of phylogenetic comparative study where the application of joint-inference methods is particularly effective or where the joint-inference paradigm allows for comparative investigations that would not otherwise be possible (we additionally provide examples of joint-inference approaches in Table 1). This section is not intended to be a comprehensive review of all available comparative approaches. Rather, we provide a series of case studies to demonstrate the value, feasibility, and challenges of joint inference. We present the topics in no particular order. We begin with two topics: SSE models and the joint modeling of discrete and continuous traits. These sections focus on capturing the complexity of trait evolution, typically on a fixed tree. In our third deep dive, we introduce evolutionary process variation, examining how heterogeneity across the tree influences our inference. The final three topics focus on active fields that exemplify joint inference: jointly estimating phylogeny alongside divergence times, inference of ancestral states, and alignment-phylogeny co-estimation.

State-Dependent Speciation and Extinction Models and the Importance of Cladogenetic Change

A good example of the importance of joint inference comes from Maddison's (2006) demonstration that the distribution of character states at the tips of a phylogeny is the result of the combined effects of differential rates of character change (e.g., maybe state A transitions to state B more frequently than the reverse) and of differential rates of speciation and/or extinction associated with those character states (e.g., maybe lineages in state B speciate more quickly than do those in state A). It is only by jointly inferring the parameters of the two generating

Downloaded from https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syag034/8651289 by guest on 21 May 2026

processes (i.e., the transition rates among states and the diversification rates associated with those states) that either can be inferred accurately (Maddison et al. 2007). This later point is worth emphasizing: although Maddison's (2006) insight is most often applied to the inference of trait-associated diversification rates (i.e., SSE rates), it equally applies to inferences of the character evolution parameters themselves (i.e., the transition rates): if a trait affects diversification rates, accurate reconstruction of the patterns of trait evolution requires that character-change parameters be co-estimated with the diversification rates. This insight profoundly altered our understanding of how to approach questions of character evolution and diversification-rate inference and prompted the development of a broad suite of methods elaborating on this framework (state-dependent speciation and extinction: SSE). For example, although the original model, `BiSSE`, was limited to a single binary character (Maddison et al. 2007), that model has been expanded to allow for multi-state characters (`MuSSE`; FitzJohn 2012), continuous characters (`QuaSSE`; FitzJohn 2010), and biogeographic characters (`GeoSSE`; Goldberg et al. 2011).

The SSE framework is additionally powerful for the study of character evolution because it inherently requires the incorporation (within the likelihood calculation) of events that happen on branches that ultimately go extinct and thus are not observed (Maddison et al. 2007). This feature permits the development of models where character-state transitions occur cladogenetically (i.e., in association with a speciation event) rather than just anagenetically (i.e., along the branches of the tree), because those models need to monitor all cladogenesis events, whether or not both daughter lineages leave extant descendants (Goldberg and Igić 2012). The earliest such cladogenetic model (`BiSSE-ness`; Mayrose et al. 2011; Magnuson-Ford and Otto 2012) was a direct conceptual extension of `BiSSE` and was followed by approaches that utilized the SSE framework for the cladogenetic functionality specifically (e.g., `ChromoSSE`, which incorporates cladogenetic changes in chromosome number but not differential trait-associated speciation and extinction rates; Freyman and Höhna 2018) and that included both cladogenetic state change and differential diversification rates (e.g., to infer the evolution of lineages' geographic range, where, e.g., the colonization of new areas might be expected to be linked to speciation; Goldberg et al. 2011).

As somewhat of an aside—but relevant, given our focus on inference versus hypothesis-testing (see Box 1)—one of the main challenges to SSE models came from Rabosky and Goldberg (2015), who claimed, using an hypothesis-testing framework, that SSE models had high rates of type I error (i.e., that such models concluded that focal character states were associated with different diversification rates when in fact no such association existed). To reach this conclusion, Rabosky and Goldberg (2015) simulated phylogenies with heterogeneous rates of diversification (but those rates were unrelated to the focal trait) and then used likelihood-ratio tests to compare a model of rates being associated with the focal trait against a null model of no rate variation. But this—the absence of rate variation—is the wrong null model; the true generating model *does* have rate variation, it is just not related to the focal trait. Because the model being rejected is not the true model, this result does not demonstrate type I error (Beaulieu and O'Meara 2016): it is the wrong hypothesis test. Hidden-state SSE models (`HiSSE`; Beaulieu and O'Meara 2016), which co-estimate the rates associated with both the focal trait and a set of unobserved character states (and all the respective transition rates among character states, both hidden and observed), attempt to resolve this issue—it is not that the co-estimation is flawed but rather that standard, non-hidden SSE models are not co-inferring enough of the relevant processes. These methods—and modeling hidden process variation in general—are discussed in more depth below (see the *Accounting for Evolutionary Process Variation* section).

Adding to their power to investigate previously inaccessible problems through co-estimation, different SSE components—including hidden states and cladogenetic state-change—can be combined together. For example, Zenil-Ferguson et al. (2019) performed a series of combined multi-state and hidden-state SSE analyses to investigate the interaction of polyploidy and the evolution of self-compatibility (the ability of a plant to fertilize itself), and their consequences, both individually and in concert, on diversification rates in the tomato family, Solanaceae. Similarly, Tribble et al. (2025) implemented a ChromoHiSSE model to explore the role of chromosomal changes as a speciation mechanism in the species-rich plant genus *Carex*. This model allows for different rates of cladogenetic chromosome number change and cladogenesis without chromosome number change, where those rates depend on a hidden state (the different hidden states have different rate parameters). By jointly modeling hidden states and cladogenetic change, ChromoHiSSE can detect not only how chromosome number change affects speciation rates, but also how the *relationship* between chromosome number change and speciation varies across the phylogeny (e.g., process variation, see the *Accounting for Evolutionary Process Variation* section below).

One important limitation of most SSE approaches from the perspective of joint inference is that although patterns of character change and rates of speciation and extinction are co-estimated, the tree itself is fixed. In this respect the application of SSE models is similar to the two-step classical PCM approaches. This limitation is important because speciation and extinction rates are inextricably related to the shape and branch lengths of the tree itself; ideally, we would want to co-estimate the tree along with the diversification parameters and rates of character-state change. Failing to accommodate diversification-rate variation potentially misleads the

Downloaded from https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syag034/8651289 by guest on 21 May 2026

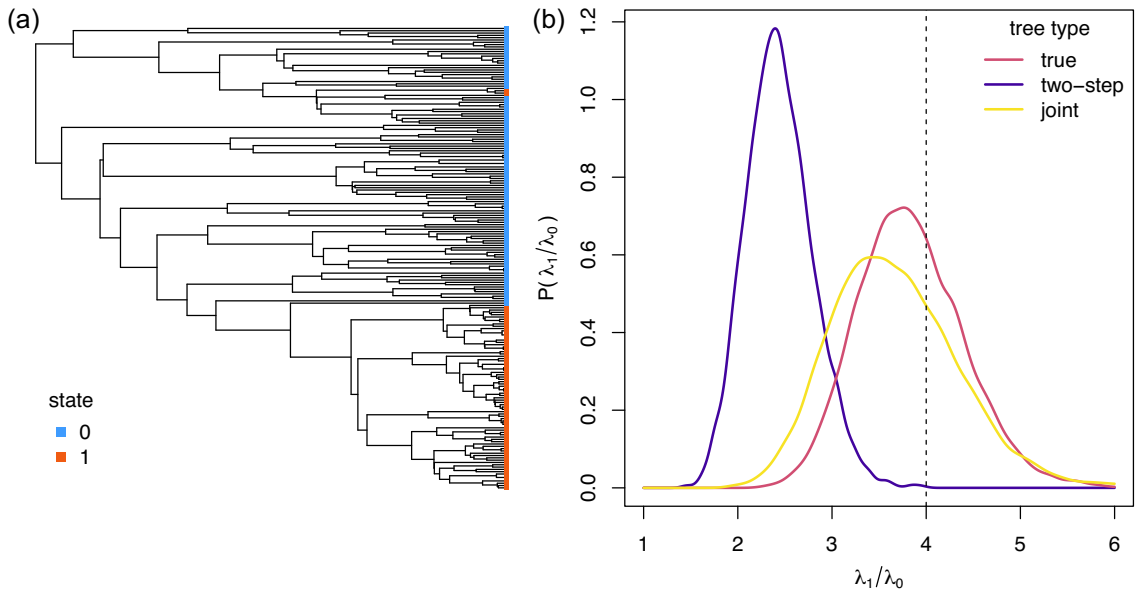

FIGURE 5. Comparing two-step and joint analysis for state-dependent diversification rates. (a) A phylogeny simulated under the binary-state pure-birth (BSPB) model with 200 tips. Tips are colored by the simulated state. (b) Posterior estimates of the speciation-rate ratio λ_1/λ_0 using the true tree (red), the conditional (two-step) tree (i.e., where the tree is first estimated under a binary-state-independent tree prior; purple), and the joint analysis (i.e., where the tree and the parameters of the BSPB model are jointly estimated; yellow). The true ratio, $\lambda_1/\lambda_0 = 4$ (dashed line), is well estimated under the true scenario, but the conditional scenario substantially underestimates the ratio, excluding the true value from the 95% HPD. The joint scenario results in estimates much more similar to the true scenario, but with more uncertainty, presumably because this scenario incorporates uncertainty in divergence times.

initial estimates of divergence times (SSE models require a chronogram), which in turn may compromise our ability to subsequently detect state-dependent diversification rates. To demonstrate this risk, we simulated a phylogeny under a simple binary-state-dependent pure-birth (BSPB) model with two speciation rates, λ_0 and λ_1, corresponding to two discrete states (i.e., a BiSSE model with $\mu_i = 0$ for all states), where the speciation rate in state 1 was four times higher than in state 0. We then estimated parameters of the BSPB model under three scenarios: (1) the "true" scenario, where the true (simulated) tree was used as input for a BSPB analysis; (2) the "two-step" scenario, where we estimated the tree from sequence data and then used the estimated tree as input for a BSPB analysis; and (3) the "joint" scenario, where we estimated the parameters of the BSPB model and the tree simultaneously from sequence data. For scenarios 2 and 3, we simulated 1000 sites on the true tree under a Jukes–Cantor substitution model and an uncorrelated lognormal relaxed clock. We then estimated Bayesian posterior distributions of divergence times (assuming the true tree topology was known) under a binary-state-independent Yule tree prior (equivalent to a BSPB model with $\lambda_0 = \lambda_1$) for the two-step scenario and a BSPB tree prior for the joint scenario, in both cases assuming the true substitution and relaxed-clock models. Finally, we fit the BSPB model to the true tree and to the Maximum Clade Credibility (MCC) tree estimated under the constant-rate Yule model (as the second part of the two-step scenario; in contrast, the BSPB model was fit jointly with the tree inference in the joint scenario). The results demonstrate that the true ratio of speciation rates is well estimated using the true tree and in the joint inference, but is significantly biased in the two-step scenario (Fig. 5). This result also provides an example of the strong sensitivity of divergence-time estimates to details of the clock and tree models (Dos Reis and Yang 2013; Condamine et al. 2015; May et al. 2021; see Supplemental Materials S1 for more details).

Jointly Modeling Discrete and Continuous Traits

Historically, the models used for estimating processes of evolution of continuous versus discrete characters were different, thus advances in modeling these character types have largely proceeded independently (see the *Fundamental Models of Trait Evolution* section; Fig. 1). The first methods for correlating continuous and discrete traits were modifications of generalized linear models (GLMs), which treat the phylogeny as the covariance matrix for residual errors (i.e., departures from expected trait values) in a linear modeling framework (Martins and Hansen 1997; Symonds and Blomberg 2014). Recently, however, mechanistic models have been developed that link the evolution of continuous and discrete traits.

In most of these approaches, a joint model links the underlying process of trait evolution among characters. Butler and King (2004) extended the OU model to in-

Downloaded from https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syaf034/8651289 by guest on 21 May 2026

clude multiple optima, θ: the multi-optimum OU model. These multiple optima correspond to different lineages on the phylogeny and thus to different optimal continuous trait values across different subclades of the phylogeny in question. Furthermore, Butler and King (2004) assigned the phylogenetic location of these different optima based on the values of a discrete trait. This approach allowed them to test if a continuous trait has different optimal values depending on the value of a discrete state. Specifically, they assigned the terminal branches of the anole phylogeny to the discrete states of the habitat type of the species at the tip. Each habitat type was then assigned a different optimum (θ parameter) to be inferred under a multi-optimum OU model of body-size evolution (a continuous character).

Shortly after Butler and King (2004), O'Meara et al. (2006) proposed another two-step approach: a multi-rate BM model that could be used to model shifts in *rates* of continuous trait evolution (the diffusion parameter σ^2), which laid the groundwork for linking such shifts to the evolutionary history of a discrete trait. Almost simultaneously, Thomas et al. (2006) published a similar set of statistical innovations.

Subsequent developments based on these works have expanded the toolbox of two-step, OU-derived methods for modeling correlated evolution between discrete and continuous traits. Beaulieu et al. (2012) expanded on Butler and King (2004) by increasing the parameters that vary across regimes to include optima, diffusion parameters, and the so-called attraction parameters,α. Collar et al. (2009), among others, took up where O'Meara et al. (2006) left off and used stochastic mapping to infer the full evolutionary history of the discrete trait across all internal and terminal branches of a phylogeny. Revell (2013) provides an important critique of these methods: by relying on a two-step approach, the continuous trait inference depends on a pre-inferred discrete trait history, thus failing to correctly model the codependency between discrete and continuous traits. Despite this critique, these approaches remained dominant for almost 15 years, until new joint inference approaches to co-infer the evolutionary history of continuous and discrete characters were developed (Fig. 6; May and Moore 2020; Boyko, O'Meara, et al. 2023).

Meanwhile, Felsenstein developed the threshold model (Fig. 3d; discussed above in *Fundamental Models of Trait Evolution: Discrete Traits Models*) independent of the OU approaches, extending it to jointly model two traits: a continuous trait that is observed and a discrete trait where the underlying continuous trait (referred to as the liability) is unobserved, both evolving via BM and linked via the covariance of the liabilities. Subsequent developments based on the threshold model framework include generalizing to multiple characters and multistate characters (Cybis et al. 2015) and enabling dimensionality reduction via a phylogenetic factor analysis to make analysis of large multivariate data sets tractable (Tolkoff et al. 2018).

The threshold model approach differs from other methods for jointly modeling discrete and continuous traits because it estimates the correlation between the traits themselves (or, at least, the trait liabilities) rather than inferring a link between the underlying trait-evolution processes (i.e., their rates, or optima). This feature is important because if we assume the processes themselves are linked (e.g. Boyko, O'Meara, et al. 2023), then we are also assuming bidirectional information flow between the traits: information about the evolution of the discrete state will and should change our inference for the evolution of the continuous trait, and vice versa. The threshold model, on the other hand, does not assume the evolutionary processes generating the discrete and continuous characters are linked, but rather proposes a way to estimate their covariances. The inferred evolutionary history of the traits would be the same if inferred together or separately. Furthermore, the threshold model has memory; the discrete character will change more frequently when the liability is close to the threshold. This quality of the model is likely desirable in some empirical cases but not in others (see Goldberg and Foo 2020).

Most recent developments in the estimation of correlated evolution between discrete and continuous traits have built off of the OU framework from Butler and King (2004) and O'Meara et al. (2006), among others. Tribble et al. (2023) developed a method to test for correlated evolution between continuous and discrete traits, where the discrete trait states are linked to the inferred optima of a continuous trait while accommodating process variation: additional variation in the evolutionary processes that is not due to the traits directly included in the model (described in detail in the following section). Under the multi-optimum OU approaches, discrete traits are linked to inferred optima of continuous traits, but those models assume a perfect link between the evolution of discrete states and optima—no other cause for variation in trait evolution beyond the modeled traits is included. In Tribble et al. (2023), the authors fit a multi-optimum OU model to the continuous traits, then correlate the per-branch optima with per-branch discrete states using stochastic mapping. This approach notably does *not* jointly estimate the continuous and discrete states, but instead infers their evolutionary histories independently, subsequently testing for a statistically significant overlap in those evolutionary histories.

In parallel, Boyko, O'Meara, et al. (2023) developed a state-dependent OU-based approach that does joint inference of the evolution of a discrete character and of a continuous character under a multi-optimum OU model with process variation. Under this approach, the evolutionary processes of both characters are linked; the discrete trait states may correspond to the inferred continuous state optima. Boyko, O'Meara, et al. (2023) built on earlier OU methods that assigned the optima to pre-established discrete trait histories on the phylogeny, requiring the evolution of continuous op-

Downloaded from https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syaq034/8651289 by guest on 21 May 2026

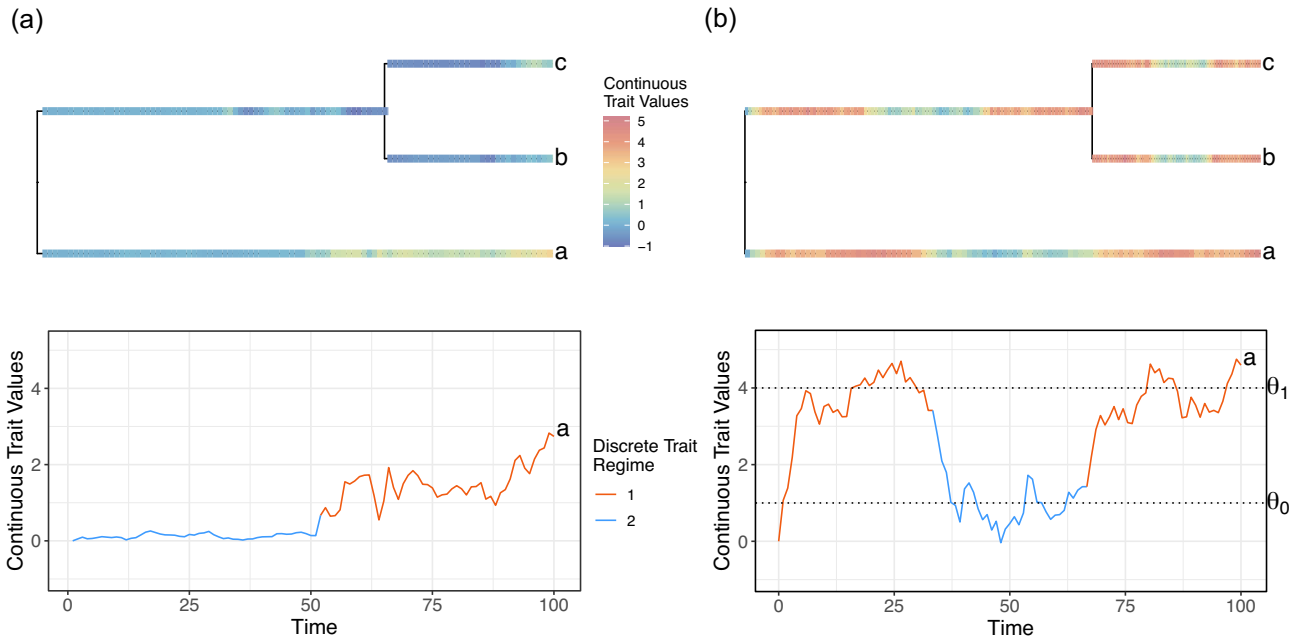

FIGURE 6. Simulation under models that jointly infer discrete and continuous trait evolution where the rates of the continuous trait are dependent on a discrete trait with states 1 and 2. (a) State-dependent BM and (b) state-dependent OU. Top row: stochastic character map with the continuous trait value plotted along the branches. Bottom row: sample path of the branch leading to species *a*. State-dependent regimes are colored. In panel a, the discrete-character regimes differ in the values of the diffusion parameter; for Regime 1 $\sigma^2_0 = 0.001$; for Regime 2 $\sigma^2_1 = 0.1$. In panel b, the discrete-character regimes differ in the values of the optimum; for Regime 1 $\theta_0 = 1$; for Regime 2 $\theta_{1=4}$.

tima to perfectly match the history of the discrete trait (e.g., Beaulieu et al. 2012). Instead, Boyko, O'Meara, et al. (2023) jointly infer the evolutionary histories of the discrete and continuous traits, while also modeling hidden process variation in their shared underlying regime.

In another approach to jointly modeling discrete and continuous trait evolution, May and Moore (2020) developed a state-dependent BM method to test for the effect of a discrete character on the rate of evolution of a continuous character (Fig. 6a). For example, does the feeding apparatus of reef-dwelling fish evolve faster than that of non-reef-dwelling fish? The method, `MuSSCRat`, jointly models a discrete character and one or more continuous characters evolving under a state-dependent multivariate BM process while accounting for process variation in the continuous trait(s) that is not linked to the discrete trait of interest. As with the BM-based approaches (e.g., O'Meara et al. 2006; Thomas et al. 2006), and unlike the approaches based on the threshold model, `MuSSCRat` models the relationship between the evolutionary processes (the rates) of the traits rather than the relationship between the trait values themselves. Both Boyko, O'Meara, et al. (2023) and May and Moore (2020) jointly model the evolution of a continuous and a discrete trait via an inferred co-evolutionary process, but differ in what that process is and thus the biological questions they are able to address. Future developments in the joint evolution of discrete and continuous traits may expand the suite of underlying evolutionary mechanisms linking trait evolution, or expand the number of traits that are included in such models.

Accounting for Evolutionary Process Variation

The goal of many character evolution models is to link trait evolution with a factor possibly contributing to variation in its evolutionary process, so-called process variation. For example, the evolution of a continuous trait such as body size may depend on the evolution of a discrete trait like life history. However, the evolution of body size likely depends on many additional factors not included in the model, which may themselves also be evolving. For example, geography could influence overall body size, environment might mediate life history, and genetic or functional linkage with other traits could influence body size, life history, or the relationship between them.

Despite this complexity, approaches to studying character evolution are often limited to modeling the influence of one or a few factors on a given trait. Additionally, methods for accounting for the influence of other factors are rarely used unless the focal question involves correlated trait evolution or lineage diversification. For example, most researchers do not use SSE models to accom-

Downloaded from https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syag034/8651289 by guest on 21 May 2026

modate the relationship between lineage diversification and trait evolution unless they are explicitly interested in how the trait influences diversification. However, as Maddison (2006) pointed out, if a trait influences lineage diversification dynamics, ancestral-state estimates will be incorrect without jointly estimating lineage diversification.

When discussing process variation, it is helpful to distinguish between focal and nuisance dependencies. A focal dependence is of direct interest to the study in question; for example, the relationship between dysploidy and speciation is one of the focal dependencies in the study of how karyotype changes generate macroevolutionary outcomes (e.g., Tribble et al. 2025). Alternatively, a nuisance dependency is a factor that influences the variable of interest but is itself not part of the primary scientific question. For example, the phasing of loci into distinct polyploid subgenomes may be a nuisance dependency when modeling the phylogenetic relationships of a clade (Freyman et al. 2023); even if a researcher is not explicitly interested in the subgenome identities of each sequence used in phylogenetic inference, that phasing will nonetheless inform the inferred topology and thus must be included in the model. This phenomenon is well established in the statistical framework of classic linear models, where nuisance dependencies may be included as indirect effects and focal dependencies are included as direct effects. Failure to include indirect effects in a linear model can lead to mistaken inference of the relationship(s) between the independent and dependent variables. Similarly, avoiding nuisance dependencies in phylogenetic analyses may lead to mistaken outcomes (Beaulieu et al. 2012).

A major challenge to appropriately model these nuisance dependencies is that they are often unknown. To extend our linear modeling analogy, unknown sources of variation are captured in an error term, but there is no error term in most stochastic phylogenetic models. Although jointly modeling multiple traits, or traits and lineage diversification, can account for the interdependency of those processes, it cannot account for the possible influence of other unknown factors.

A popular solution to the problem of unknown nuisance dependencies is to include so-called hidden states (Beaulieu et al. 2013; Beaulieu and O'Meara 2016). Hidden states are akin to error terms for phylogenetic models, metaphorical sponges that absorb additional process variation not captured by the explicitly modeled nuisance and focal dependencies. Hidden state models accommodate hidden process variation by expanding the state space of a standard Mk model. They include an additional trait(s) with its own states in addition to the focal trait. The evolutionary process of the focal trait depends in some way on the state of the hidden trait—perhaps the transition rates are different for different hidden states. Take the example of a binary trait with states $0, 1$ with one transition rate between states, q:

$$Q = \begin{array}{c} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{bmatrix} - & q \\ q & - \end{bmatrix} \end{array}$$

In a hidden-state variant of this model, we include an additional hidden binary trait h with states A, B. The inclusion of the hidden state doubles the state space to $0A, 1A, 0B, 1B$, where A, B represent the influence of the hidden trait on 0,1. The resulting model includes a rate for $0 \leftrightarrow 1$ transitions when the hidden state is A (denoted q_a), for when the hidden state is B (q_b), and for transitions $A \leftrightarrow B$ (q_h). Simultaneous transitions between $0 \leftrightarrow 1$ and $A \leftrightarrow B$ are set to 0.

$$Q_h = \begin{array}{c} \\ 0A \\ 1A \\ 0B \\ 1B \end{array} \begin{array}{c} \begin{array}{cccc} 0A & 1A & 0B & 1B \end{array} \\ \begin{bmatrix} - & q_a & q_h & 0 \\ q_a & - & 0 & q_h \\ q_h & 0 & - & q_b \\ 0 & q_h & q_b & - \end{bmatrix} \end{array}$$

Although introduced into the character-evolution modeling world by Beaulieu et al. (2013), hidden states were first developed for implementation in phylogenetics for molecular evolution (e.g., the covarion model; Fitch and Markowitz 1970; Tuffley and Steel 1998). Since these initial developments, variants of hidden-state models have been proposed as solutions to various problems in phylogenetic modeling, including spurious associations in SSE models (Beaulieu and O'Meara 2016; Caetano et al. 2018) and correlated discrete-character models (Boyko and Beaulieu 2023), accommodating process variation while jointly modeling correlated evolution between discrete and continuous variables (Boyko, O'Meara, et al. 2023), detecting multiple modes of state-dependent chromosome evolution (Tribble et al. 2025), and extending the covarion model to accommodate lineage-specific variation via Markov-modulated CTMCs (Baele et al. 2021). Similarly, the term "tip fog" has been introduced to describe the contribution of process variation and error (or uncertainty in tip estimates; Beaulieu and O'Meara 2025); it refers to the "variance that is not modeled by the evolutionary process applied to the rest of the tree." Tip fog is thus somewhat analogous to our definition of process variation, and we elaborate on tip-fog approaches below.

Many hidden-state model implementations, such as the one described in the above Q matrix, use a single hidden discrete trait. However, the decision to model a single hidden discrete trait is somewhat arbitrary, or based on computational limitations, rather than coming from our interpretation of biological reality. Do we think that all the variation in processes of trait evolution can be most appropriately modeled by a single discrete variable? One alternative approach is to vary the number of hidden traits and/or hidden states in the model, as more traits/states could accommodate more complex patterns of process variation (Beaulieu and O'Meara 2025).

Downloaded from https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syag034/8651289 by guest on 21 May 2026

However, the classic hidden-state model is only appropriate for discrete traits, and alternative solutions are necessary for continuous traits. For example, `MuSSCRat` (discussed above) jointly models the evolution of a discrete trait and a continuous trait, while also modeling hidden process variation (specifically, continuous rate heterogeneity) via a relaxed clock model (May and Moore 2020). Alternatively, Beaulieu and O'Meara (2025) suggest incorporating a parameter ζ (zeta) into the variance-covariance matrix of a BM model to accommodate additional variance from tip fog (a distinct approach from the discrete character solution described above; Beaulieu et al. 2012).

Uyeda et al. (2018) pointed out that appropriately modeling process variation is critical for all comparative methods and is a key part of the solution to Maddison and FitzJohn's "unsolved challenge": the impact of unreplicated events (evolutionary changes that occur singularly on the tree) on accurate inference (Maddison and FitzJohn 2015). Uyeda et al. (2018) also argue that researchers must account for unreplicated events while assessing relationships between focal traits and processes or risk being misled by mistaken associations between variables. While Uyeda et al. (2018) describe this as part of "phylogenetic natural history" (where do major shifts in tempo and mode occur in the phylogeny?), the idea is largely congruent with our description of evolutionary process variation. One important lesson from Uyeda et al. (2018) is that some forms of process variation are more easily addressed than others. In particular, unreplicated events are challenging to account for using standard hidden-state approaches.

More creative approaches to modeling hidden process variation are needed, especially methods that focus on biological realism, to capture the underlying complexity of character macroevolution. For example, Uyeda et al. (2018) suggest Bayesian mixture modeling as an alternative method for modeling process variation. Computational limits—which have previously restricted the state space of models or the number of parameters that can reasonably be inferred—are continuously relaxed as new hardware and better algorithms are developed. Large phylogenetic and phenotypic data sets are increasingly available to inform estimation under complex models, and we anticipate and encourage advances in modeling process variation given these data sets. In parallel, we recognize that large phenotypic data sets are hard won, often requiring a wealth of natural history information to compile. Computational improvements will not replace careful morphological and anatomical examination of organisms as an integral part of understanding process variation and other complex evolutionary processes.

Joint Inference in Divergence-Time Estimation

The canonical approach to inferring a macroevolutionary timescale—a phylogeny with branch lengths in units of time—is node-dating. In this approach, a researcher, based on the age of a fossil and on their implicit understanding of the processes of character evolution and of the probability of fossil recovery, attaches a probability distribution (a calibration density) to the age of particular nodes in a tree. In conjunction with a tree model and a clock model, this age information can be propagated throughout the phylogeny, thus time-calibrating it. This approach is very similar to a two-step PCM in that the focal data (here, the age data of the fossil) are superimposed on a more-or-less pre-existing phylogeny: the fossil ages correct the branch lengths of the phylogeny rather than being incorporated with the character data to jointly infer the phylogeny and the timescale of evolution.

In the last decade, so-called Bayesian total-evidence dating (TED, also known as tip dating or combined-evidence dating; Ronquist et al. 2012) has emerged as an alternative to the node-dating approach. In this framework, researchers estimate a time-calibrated phylogeny from a combined data set of extant and extinct species, with the extinct species treated as tips and their phylogenetic position and branch lengths estimated from morphological character data. In turn, the fossil ages provide the temporal data necessary to disentangle rate and time. Early applications of TED used discrete morphological data (e.g., Ronquist et al. 2012; Arcila et al. 2015), but more recent studies have begun to also use continuous morphological traits (e.g., Álvarez-Carretero et al. 2019; González-Ramírez et al. 2026).

TED approaches are generally discussed specifically in the context of divergence-time estimation, where they have several key advantages over node-dating. First, TED frees us from relying on *a priori* assignment of fossils to nodes in the extant subtree—the position of the fossils (including the full uncertainty in that position) is inferred from the data instead of being fixed in advance by the researcher. Second, and related, TED approaches allow us to infer the processes of morphological evolution and, when incorporating an FBD tree model (Heath et al. 2014), the process of fossilization, removing the need to specify a calibration density in advance of the analysis. Third, TED analyses can incorporate a richer array of fossil data than can node-dating approaches, which are limited to including only the oldest fossil that is attributed to any particular node (Sauquet et al. 2012). However, these advantages come at a cost, as the TED model is complex, requiring the user to specify models of molecular evolution, morphological evolution, and tree evolution, and achieving MCMC convergence is often difficult (Warnock and Wright 2020; May et al. 2021).

Less commonly discussed is the fact that the joint-inference nature of TED additionally provides for a series of inferences (beyond the divergence times themselves) that would not otherwise be possible, including those directly pertaining to character macroevolution. For example, by jointly inferring the placement of the fossils alongside the timescale of evolution and

Downloaded from https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syag034/8651289 by guest on 21 May 2026

the process of morphological change, there is an opportunity to infer that particular character states are plesiomorphic instead of synapomorphic. A good illustration of this outcome comes from a study of marattialean ferns (May et al. 2021) where a series of fossils long treated as members of the extant genus *Marattia* were included in a TED analysis. However, instead of being resolved in crown Marattiales with *Marattia*, the TED joint inference strongly preferred a placement of these fossils deep along the Marattiales stem branch. In conjunction with the different topological inference, the TED analysis resulted in different conclusions of character evolution: *Marattia* and several other extant genera exhibit a conserved suite of ancestral character states and the extant lineages *Danaea* and *Christensenia*—the more phylogenetically isolated (*Danaea*) and depauperate (*Christensenia*) marattialean genera—are characterized primarily by apomorphies, rather than the other way around. To rephrase this conclusion slightly, a fixed tree may bias the estimates of morphological model parameters if the phylogenetic positions of fossils are incorrect; jointly estimating the tree not only allows more accurate placement of fossils, but may also provide more reliable and realistic estimates of topology and of how the traits themselves have evolved (see also Lee et al. 2009; González-Ramírez et al. 2026). Moreover, the novel macroevolutionary inferences enabled by the TED joint inference extend beyond phylogenetic relationships and patterns of character evolution. For example, TED analyses can estimate patterns of species richness over deep geological time (the marattialean ferns are estimated to have had a peak species richness of ≈2500 species, at the end of the Carboniferous—this diversity declined precipitously with the onset of global drying in the Permian to the present extant diversity of ≈100 species; May et al. 2021).

Indeed, TED analyses exemplify the current power of joint phylogenetic and macroevolutionary inference: by jointly estimating the phylogeny and the model of trait evolution, these analyses naturally accommodate phylogenetic uncertainty about fossil placement, topology, branch lengths, and all other relevant model parameters. Despite the apparent advantages of TED for joint evolutionary inference, the communication between the subfields of divergence-time estimation and PCMs remains limited. TED analyses still tend to rely on models that could be viewed as overly simplistic or uninteresting by the PCM community (e.g., simple Mk1 models of morphological change; Lewis 2001), whereas most PCM analyses fail to co-estimate the phylogeny jointly with the focal patterns of character evolution. To some degree, this lack of communication may be due to the framing of TED as being specifically a method for estimating divergence times incorporating fossils, whereas most phylogenetic comparative studies involve only extant species. However, there is nothing inherent to the total-evidence framework that requires the use of fossils: practitioners may still jointly estimate the phylogeny and models of morphological evolution for data sets of extant taxa. Additionally, the value of fossils in studies of trait evolution has long been recognized (e.g., Cobbett et al. 2007; Mongiardino Koch and Parry 2020; Mongiardino Koch et al. 2021; Hand et al. 2023), and TED (with an FBD tree model) offers a coherent way to include fossils when they might be helpful. As more flexible computational frameworks for joint phylogenetic analysis continue to emerge and mature (e.g., `RevBayes`, `BEAST`; Höhna et al. 2016; Bouckaert et al. 2019), we foresee great opportunity for the merging of complex models of morphological evolution and phylogenetic inference in one common framework, including using complex state-dependent diversification models as tree priors.

Joint Inference of Phylogeny and Ancestral Character States

Inference of ancestral states, whether they be of morphological characters or nucleotides/amino acids (ancestral-sequence inference), is still often undertaken in a two-step PCM framework, where first a phylogeny is inferred and subsequently that phylogeny is used to estimate the patterns of evolution of the specific characters of choice. This two-step approach runs afoul of the issues described earlier (see the *Joint Inference and the Bayesian Paradigm* section): given that this two-step approach assumes that the focal characters evolved on the phylogeny, information in those characters should be allowed to influence the topology through joint inference. However, ancestral-state estimation additionally presents some specific challenges. Although early PCM researchers treated ancestral states as parameters, which could thus be subject to hypothesis-testing via likelihood-ratio tests (e.g., Pagel 1999b), this treatment is incorrect (e.g., see Yang et al. 1995). Ancestral states are not parameters—settings of the evolutionary model that produced our data—rather they are data themselves (random variables produced by the model): if we were to travel back in time we could even determine what those states were! Unfortunately, we cannot know those states and must treat them as unobserved (just as we might treat character data as missing for an extant sample that we were not able to score).

Given that a likelihood-ratio-based approach is inappropriate, our best option is to use the rules of probability to estimate ancestral states, recognizing them as random variables drawn from a conditional distribution (conditional on the tree, observed data, and model parameters; Yang et al. 1995; Koshi and Goldstein 1996). In this framing, we can derive the distributions of states across nodes ("joint" ancestral states), as well as for individual nodes ("marginal" ancestral states) exactly. Abstractly, the distribution of tip and node states can be represented as

$$P(X_{\text{tips}}, X_{\text{nodes}} \mid T, \theta),$$

Downloaded from https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syag034/8651289 by guest on 21 May 2026

where T is the tree, θ are the parameters of the assumed evolutionary model, and X_{nodes} are the states at some pre-defined set of one or more internal nodes. The rules of conditional probability can then be used to write down the distribution of node states given the tip states

$$P(X_{\text{nodes}} \mid X_{\text{tips}}, T, \theta) = \frac{P(X_{\text{tips}}, X_{\text{nodes}} \mid T, \theta)}{P(X_{\text{tips}} \mid T, \theta)}.$$

Note that $P(X_{\text{tips}} \mid T, \theta)$ is simply the probability of the observed data under the model (i.e., the likelihood). This equation describes the probability that one or more internal nodes have the set of states X_{nodes}. For a single node, this equation describes the *marginal* ancestral-state distribution, whereas for multiple nodes it describes the *joint* ancestral-state distribution. We use the term marginal in this case because we are marginalizing over states at all other nodes, whereas we use joint because we are looking at the distribution of more than one node. (Note that this use of "joint" differs from the way we have used this term elsewhere, as we are referring to the probability of states at two or more nodes under one model, rather than jointly modeling multiple traits and/or the tree.) Although fast algorithms exist for computing the joint probability (e.g., Pupko et al. 2000), they are generally more computationally expensive and less intuitive than the marginal approach (described below); it is harder to visualize the joint estimate and its associated uncertainty on a tree.

In contrast to joint ancestral states, marginal ancestral-state estimates are relatively easy to compute, visualize, and interpret. There are various algorithms for computing marginal ancestral-state probabilities (which are generally variants of Felsenstein's pruning algorithm), with notable examples that involve only considering data downstream of the node (Paradis and Schliep 2019), rerooting the tree at the node (theoretically correct for time-reversible models; Yang 2014), or directly evaluating the full marginal probability analytically (using so-called sandwich algorithms, e.g., Minin and Suchard 2008). For any given node, the state that has the highest marginal probability is typically considered the best point estimate of the ancestral state at that node (Yang 2014; Revell 2025), but it is also common to represent the whole distribution as a pie chart (with slices proportional to the state probability) on a phylogenetic tree. We note that it is common practice to reconstruct the marginal ancestral state at each internal node, then plot all of those marginal estimates on one tree (either as point estimates or pie charts), an approach we deem "full-tree marginal reconstruction." Although this is generally a very useful way to visualize ancestral states, such plots should be interpreted with caution when it comes to making statements about states at multiple nodes because that joint information has been removed by the marginal reconstruction.

The logic of ancestral-state estimation can be extended to the *entire evolutionary history* of the trait along the tree, which is referred to as stochastic character mapping (Huelsenbeck et al. 2003). Rather than focusing on the distribution of states at internal nodes, X_{nodes}, here we focus on the distribution of the entire history of states (over all nodes and branches), H. The distribution of the history H conditional on the observed data and model is

$$P(H \mid X_{\text{tips}}, T, \theta) = \frac{P(H, X_{\text{tips}} \mid T, \theta)}{P(X_{\text{tips}} \mid T, \theta)}.$$

This distribution is generally analytically intractable. However, many algorithms have been developed to generate samples from this distribution (Nielsen 2002; Huelsenbeck et al. 2003; Hobolth 2008; Minin and Suchard 2008; Irvahn and Minin 2014), including for state-dependent birth-death models (Freyman and Höhna 2019) and continuous-trait models (Martin and Weber 2026). An individual sample of H is referred to as a stochastic character map (Fig. 2). Importantly, the sample of histories H can be used to estimate ancestral states at nodes; these estimates are, in principle, equivalent to the results from standard ancestral-state algorithms (Revell 2025).

So far we have focused on ancestral-state estimation under a fixed model (with, for example, maximum-likelihood parameter estimates) on a pre-estimated tree. In reality, there is often significant uncertainty about both the model parameter values and the tree. A partial solution to this problem is to use Bayesian inference to jointly sample parameter values and ancestral states (Huelsenbeck and Bollback 2001); nonetheless, many Bayesian approaches are still two-step, as character evolution is studied on a known tree. One common approach to mitigate the risk associated with conditioning on a single tree is to take into account tree uncertainty by fitting the character-evolution model by maximum likelihood to each tree in a distribution of trees (a reflection of our uncertainty about the tree), and averaging the result and/or reporting confidence intervals. Alternatively, in a Bayesian framework, one can use a posterior distribution of trees (e.g., previously inferred from molecular data) as a prior distribution of trees for our model of character evolution (a Bayesian sequential approach) or co-estimate the tree and ancestral states in a fully Bayesian joint analysis. Although there are some difficulties in summarizing the states over the posterior distribution of trees (e.g., some nodes do not appear in all trees; Pagel et al. 2004), Bayesian approaches (either sequential or joint) fully accommodate uncertainty in the tree, model parameters, and ancestral states, and properly allow the character data to update our belief about different trees. Such analyses are possible in many standard Bayesian implementations of ancestral-state estimation, for example, `BayesTraits` (Pagel and Meade 2006), `BEAST` (Bouckaert et al. 2019), and `RevBayes` (Höhna et al. 2016).

Downloaded from https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syag034/8651289 by guest on 21 May 2026

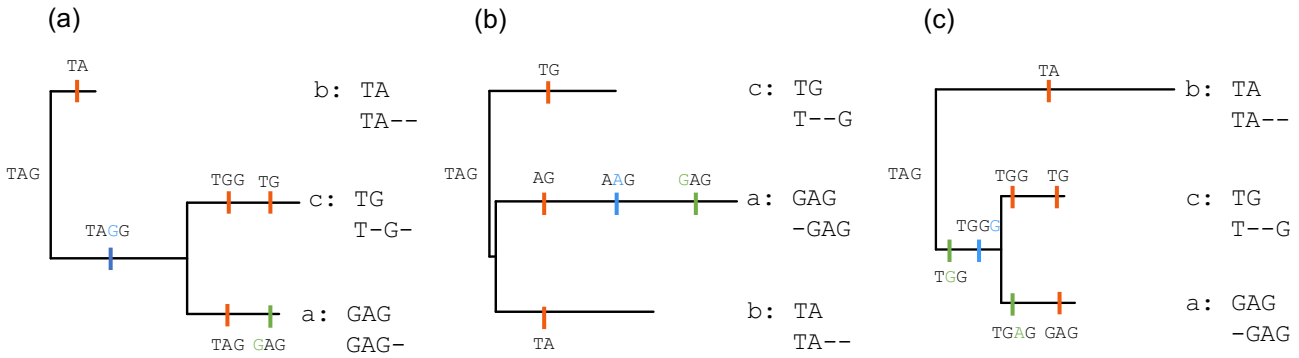

FIGURE 7. Sequences simulated with a joint model of insertion-deletion and nucleotide substitution. (a–c) Three randomly selected simulations of representative nucleotide sequences under an indel and substitution model, conditioned on the root sequence (TAG) and observed tip sequences (TA, TG, GAG, in no particular order) on different phylogenies. Tip sequences are shown unaligned (top) and aligned (bottom) given the history of sequence evolution. Tree topologies indicate events mapped along branches with tick marks: Blue ticks correspond to deletions, orange to insertions, and green to substitutions. Colored alignment positions represent the event with respect to the ancestral sequence. A progressive aligner (e.g., `mafft`) would likely infer a different alignment (a: GAG, b: TA-, and c: TG-) than the indel model because the indel model co-estimates the alignment with the tree.

Joint Inference of Alignment and Phylogeny

One particularly foundational area where joint inference can be applied is in the inference of alignments. The vast majority of phylogenetic studies take a two-step approach by applying a piece of progressive alignment software to their sequences (reviewed in Löytynoja and Goldman 2005; Katoh et al. 2009; Kemena and Notredame 2009; Chatzou et al. 2016) and subsequently treating those alignments as if they were data. But those alignments are not data—they are inferences, *estimates* of homology—and as with all estimates, they have associated uncertainties (Sankoff et al. 1973). In addition to the failure to account for alignment uncertainty, which can compromise downstream phylogenetic inferences (e.g., Vialle et al. 2018), this two-step approach has several major weaknesses. First, the alignment evolved on the phylogeny (substitutions and insertion and deletion events happened along the branches of the phylogeny) and, as we repeatedly emphasize in this review, that calls for joint estimation: the phylogeny contains information about the alignment and vice versa and so the two need to be able to influence each other. Second, and more practically, phylogenetic inference with multiple sequence alignments generated by progressive alignment can be biased (toward the guide tree; Thorne and Kishino 1992). Third, the score-based approach generally underlying alignment algorithms is not informed by evolutionary relationships—indels are repeatably penalized when it is more appropriate to penalize them once, at their apomorphic origin (Löytynoja and Goldman 2005). Finally, progressive alignment does not mechanistically model the processes that generated the data—it is not possible to derive inferences about, for example, insertion rates or other aspects of the process that generated the alignment (Wong et al. 2008).

However, joint estimation of the alignment and phylogeny is possible thanks to the development of models of insertion and deletion (indel models, reviewed in Redelings et al. 2024). These indel models aim to describe the molecular processes that can produce variation in the length of homologous loci; they generally include parameters for insertion rate, deletion rate, length distribution of indels, and total alignment length (see Fig. 7 for how this approach works in a simulation; Thorne et al. 1991; Holmes and Bruno 2001; Miklós et al. 2004; Redelings and Suchard 2005). The joint estimation of alignment and phylogeny allows users to account for alignment uncertainty and thus to avoid conditioning their phylogenetic inferences on a single homology estimate (a single alignment; e.g., Gaya et al. 2011; Westesson et al. 2012), to produce more accurate point estimates of alignment and/or phylogeny (including producing point estimates of the alignment for other downstream applications in situations where two-step tree inference is unreliable, e.g., Rothfels and Schuettpelz 2014), and provides mechanistic inferences relating to the indel process, thus allowing, for example, for the inference of ancestral gaps in DNA sequence data (Sharma 2025; see *Joint Inference of Phylogeny and Ancestral Character States* section, above).

CONCLUSION

In this review, we argue for a paradigm shift away from studying character evolution via a classical two-step PCM approach where the focal character(s) of interest are analyzed on a fixed pre-estimated phylogeny, and toward a joint-inference paradigm, where all data are incorporated to jointly infer details of the generating process; this approach integrates the historically separate subfields of phylogenetic inference and phylogenetic comparative methods. Specifically, we argue that joint inference is generally more conceptually sound and accurate than the canonical two-step approach, and that

Downloaded from https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syag034/8651289 by guest on 21 May 2026

it opens up entirely new types of analyses that are not possible under the two-step PCM paradigm. Related to our arguments for joint inference is our broader argument in favor of an increased focus on inference (estimation) in general, versus more formal hypothesis-testing-related approaches; we believe this shift has conceptual and practical advantages (e.g., by focusing more on parameter estimation than model comparison we can mitigate some of the issues raised by Maddison and FitzJohn 2015).

It is important to acknowledge, however, that this joint-inference paradigm is not the be-all and end-all. There are many situations in which parameters of interest may be non-identifiable (e.g., Dos Reis and Yang 2013; Louca and Pennell 2020a; May et al. 2021) and the structure of our models may preclude particular inferences (e.g., May and Rothfels 2023). Perhaps most relevantly, there are many evolutionary processes that are difficult to model accurately (e.g., ancestral-transcriptome reconstruction and biogeography, among others; Mika et al. 2022; Mantica et al. 2024; Thompson et al. 2024; Swiston and Landis 2025) and questions that are effectively inaccessible given the data at hand (Quental and Marshall 2010; see discussion in Losos 2011). Moreover, from a practical perspective, there are relatively few pre-packaged joint models available (but see available methods listed in Table 1), and developing a bespoke model for a particular application is beyond the reach of most systematic biologists. Fortunately, in many cases we can expect any inaccuracies introduced by adopting a two-step PCM approach to be minimal: given the size of the sequence data sets typically applied to tree inference, the exclusion of a small set of focal characters from that inference will most likely have a negligible impact.

However, we anticipate that the barriers introduced by these practical issues to decline with time, and that the trends that have allowed for joint inference—trends in the conceptualization of phylogenetics and increases in computation power, data availability, and model complexity—will continue. Already, platforms like `RevBayes` (Höhna et al. 2016) and `BEAST` (Bouckaert et al. 2019) provide relatively easy access to a wide variety of flexible joint-inference models, and emerging platforms like `TreePPL` (a universal probabilistic programming language, or PPL; Senderov et al. 2024) promise to increase the available range of joint phylogenetic models even further. As access to these methods expands, joint inference will enable explorations of increasingly complex questions involving the interconnections among diversification dynamics, character evolution, molecular evolution, and more. For example, a joint model of phylogeny, nucleotide sequences, and morphological characters could infer correlated evolution between morphology, genes, and speciation rates. Joint inference positions statistical systematics as a unified framework for integrating diverse organismal data and building a more complete understanding of complex macroevolutionary processes.

Acknowledgments

We thank Rose Novick, Stacey Smith, Brian O'Meara, and the extended Rothfels lab for helpful feedback on earlier drafts of this paper, Boris Igić, Matt Pennell, Dave Swofford, Will Freyman, Stacey Smith, Josef Uyeda, Luke Harmon, Rosana Zenil-Ferguson, and the Berkeley, Beer, and Biosystematics crew for discussions that have informed our understanding of this topic, and the SSB Legacy Committee for inviting us to contribute this article.

Supplementary Material

Supplementary material is available at *Systematic Biology* online.

Funding

This material is based on work supported by the NSF Postdoctoral Research Fellowships in Biology Program [grant #2209159 to J.M.G.] and by NSF CAREER [grant #2144011 to C.J.R.].

Data Availability

The codes for figures in the manuscript are available at https://github.com/Jesusthebotanist/SysBio_EvolvingViews_CharacterEvolution, archived on Zenodo at https://doi.org/10.5281/zenodo.17982676.

References

Álvarez-Carretero S., Goswami A., Yang Z., Dos Reis M. 2019. Bayesian estimation of species divergence times using correlated quantitative characters. Syst. Biol. 68(6):967–986.

Arcila D., Pyron R.A., Tyler J.C., Ortí G., Betancur-R R. 2015. An evaluation of fossil tip-dating versus node-age calibrations in tetraodontiform fishes (Teleostei: Percomorphaceae). Mol. Phylogenet. Evol. 82:131–145.

Armbruster W.S. 1992. Phylogeny and the evolution of plant-animal interactions. BioScience 42(1):12–20.

Baele G., Gill M.S., Bastide P., Lemey P., Suchard M.A. 2021. Markov-modulated continuous-time Markov chains to identify site- and branch-specific evolutionary variation in BEAST. Syst. Biol. 70(1):181–189.

Baele G., Ji X., Hassler G.W., McCrone J.T., Shao Y., Zhang Z., Holbrook A.J., Lemey P., Drummond A.J., Rambaut A., Suchard M.A. 2025. Beast x for bayesian phylogenetic, phylogeographic and phylodynamic inference. Nat. Methods 22(8):1653–1656.

Baum D.A., Smith S.D. 2013. Tree thinking: An Introduction to Phylogenetic Biology. Greenwood Village: Roberts and Company Publishers.

Beauchamp G., Fernández-Juricic E. 2004. Is there a relationship between forebrain size and group size in birds? Evol. Ecol. Res. 6(6):833–842.

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. https://doi.org/10.1111/j.1558-5646.2012.01619.x

Downloaded from https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syag034/8651289 by guest on 21 May 2026

Beaulieu J.M., O'Meara B.C., Donoghue M.J. 2013. Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. Syst. Biol. 62(5):725–737.
Beaulieu J.M., Donoghue M.J. 2013. Fruit evolution and diversification in campanulid angiosperms. Evolution 67(11):3132–3144.
Beaulieu J.M., O'Meara B.C. 2016. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. Syst. Biol. 65(4):583–601.
Beaulieu J.M., O'Meara B.C. 2025. Navigating "tip fog": embracing uncertainty in tip measurements. Evolution 79(7):1131–1142.
Blomberg S.P., Garland T., Ives A.R. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution 57(4):717–745.
Blomberg S.P., Rathnayake S.I., Moreau C.M. 2020. Beyond Brownian motion and the Ornstein-Uhlenbeck process: stochastic diffusion models for the evolution of quantitative characters. Am. Nat. 195(2):145–165.
Borstein S.R., Hammer M.P., O'Meara B.C., McGee M.D. 2024. The macroevolutionary dynamics of pharyngognathy in fishes fail to support the key innovation hypothesis. Nat. Commun. 15(1):10325.
Boucher F.C., Démery V. 2016. Inferring bounded evolution in phenotypic characters from phylogenetic comparative data. Syst. Biol. 65(4):651–661.
Bouckaert R., Vaughan T.G., Barido-Sottani J., Duchêne S., Fourment M., Gavryushkina A., Heled J., Jones G., Kühnert D., De Maio N., Matschiner M., Mendes F.K., Müller N.F., Ogilvie H.A., du Plessis L., Popinga A., Rambaut A., Rasmussen D., Siveroni I., Suchard M.A., Wu C.H., Xie D., Zhang C., Stadler T., Drummond A.J. 2019. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 15(4):e1006650.
Boyko J.D., Hagen E.R., Beaulieu J.M., Vasconcelos T. 2023. The evolutionary responses of life-history strategies to climatic variability in flowering plants. New Phytol. 240(4):1587–1600.
Boyko J.D., O'Meara B.C., Beaulieu J.M. 2023. A novel method for jointly modeling the evolution of discrete and continuous traits. Evolution 77(3):836–851. https://doi.org/10.1093/evolut/qpad002
Boyko J.D., Beaulieu J.M. 2023. Reducing the biases in false correlations between discrete characters. Syst. Biol. 72(2):476–488.
Braga M.P., Landis M.J., Nylin S., Janz N., Ronquist F. 2020. Bayesian inference of ancestral host–parasite interactions under a phylogenetic model of host repertoire evolution. Syst. Biol. 69(6):1149–1162.
Brandley M.C., Huelsenbeck J.P., Wiens J.J. 2008. Rates and patterns in the evolution of snake-like body form in squamate reptiles: evidence for repeated re-evolution of lost digits and long-term persistence of intermediate body forms. Evolution 62(8):2042–2064.
Brooks D.R., McLennan D.A. 1991. Phylogeny, ecology and behaviour: A research program in comparative biology. Chicago (IL): University of Chicago Press.
Brown J.M., Lemmon A.R. 2007. The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. Syst. Biol. 56(4):643–655.
Burress E.D., Martinez C.M., Wainwright P.C. 2020. Decoupled jaws promote trophic diversity in cichlid fishes. Evolution 74(5):950–961.
Butler M.A., King A.A. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. Am. Nat. 164(6):683–695.
Caetano D.S., O'Meara B.C., Beaulieu J.M. 2018. Hidden state models improve state-dependent diversification approaches, including biogeographical models. Evolution 72(11):2308–2324.
Caetano D.S., Harmon L.J. 2019. Estimating correlated rates of trait evolution with uncertainty. Syst. Biol. 68(3):412–429.
Cavender J.A. 1978. Taxonomy with confidence. Math. Biosci. 40(3–4):271–280.
Chatzou M., Magis C., Chang J.M., Kemena C., Bussotti G., Erb I., Notredame C. 2016. Multiple sequence alignment modeling: methods and applications. Brief. Bioinform. 17(6):1009–1023.
Cobbett A., Wilkinson M., Wills M.A. 2007. Fossils impact as hard as living taxa in parsimony analyses of morphology. Syst. Biol. 56(5):753–766.
Coddington J.A. 1988. Cladistic tests of adaptational hypotheses. Cladistics 4(1):3–22.
Collar D.C., O'Meara B.C., Wainwright P.C., Near T.J. 2009. Piscivory limits diversification of feeding morphology in centrarchid fishes. Evolution 63(6):1557–1573. https://doi.org/10.1111/j.1558-5646.2009.00626.x
Condamine F.L., Nagalingum N.S., Marshall C.R., Morlon H. 2015. Origin and diversification of living cycads: a cautionary tale on the impact of the branching process prior in Bayesian molecular dating. BMC Evol. Biol. 15(1):65.
Corn K.A., Friedman S.T., Burress E.D., Martinez C.M., Larouche O., Price S.A., Wainwright P.C. 2022. The rise of biting during the Cenozoic fueled reef fish body shape diversification. Proc. Natl. Acad. Sci. USA 119(31):e2119828119.
Cornwallis C.K., Griffin A.S. 2024. A guided tour of phylogenetic comparative methods for studying trait evolution. Annu. Rev. Ecol. Evol. Syst. 55:181–204.
Cybis G.B., Sinsheimer J.S., Bedford T., Mather A.E., Lemey P., Suchard M.A. 2015. Assessing phenotypic correlation through the multivariate phylogenetic latent liability model. Annu. Appl. Stat. 9(2):969–991. https://doi.org/10.1214/15-AOAS821
de Queiroz A., Donoghue M.J., Kim J. 1995. Separate versus combined analysis of phylogenetic evidence. Annu. Rev. Ecol. Syst. 26:657–681.
de Queiroz K. 1996. Including the characters of interest during tree reconstruction and the problems of circularity and bias in studies of character evolution. Am. Nat. 148(4):700–708.
de Queiroz K., Poe S. 2001. Philosophy and phylogenetic inference: a comparison of likelihood and parsimony methods in the context of Karl Popper's writings on corroboration. Syst. Biol. 50(3):305–321.
Desdevises Y., Legendre P., Azouzi L., Morand S. 2003. Quantifying phylogenetically structured environmental variation. Evolution 57(11):2647–2652.
Dos Reis M., Yang Z. 2013. The unbearable uncertainty of Bayesian divergence time estimation. J. Syst. Evol. 51(1):30–43.
Esquerré D., Brennan I.G., Catullo R.A., Torres-Pérez F., Keogh J.S. 2019. How mountains shape biodiversity: The role of the ANdes in biogeography, diversification, and reproductive biology in south America's most species-rich lizard radiation (Squamata: Liolaemidae). Evolution 73(2):214–230.
Faith D.P., Trueman J.W. 2001. Towards an inclusive philosophy for phylogenetic inference. Syst. Biol. 50(3):331–350.
Farris J.S. 1973. A probability model for inferring evolutionary trees. Syst. Biol. 22(3):250–256.
Farris J.S., Kluge A.G., Carpenter J.M. 2001. Popper and likelihood versus "Popper". Syst. Biol. 50(3):438–444.
Felsenstein J. 1973. Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. Syst. Biol. 22(3):240–249.
Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17:368–376.
Felsenstein J. 1985. Phylogenies and the comparative method. Am. Nat. 125(1):1–15.
Felsenstein J. 2004. Inferring phylogenies. Sunderland, MA: Sinauer Associates.
Felsenstein J. 2005. Using the quantitative genetic threshold model for inferences between and within species. Philos. T. R. Soc. B Biol. Sci. 360(1459):1427–1434.
Felsenstein J. 2012. A comparative method for both discrete and continuous characters using the threshold model. Am. Nat. 179(2):145–156. https://doi.org/10.1086/663681
Fitch W.M., Markowitz E. 1970. An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution. Biochem. Genet. 4(5):579–593.
FitzJohn R.G. 2010. Quantitative traits and diversification. Syst. Biol. 59(6):619–633.
FitzJohn R.G. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. Methods Ecol. Evol. 3(6):1084–1092.
Freyman W.A., Johnson M.G., Rothfels C.J. 2023. homologizer: phylogenetic phasing of gene copies into polyploid subgenomes. Methods Ecol. Evol. 14(5):1230–1244.

Downloaded from https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syag034/8651289 by guest on 21 May 2026

Freyman W.A., Höhna S. 2018. Cladogenetic and anagenetic models of chromosome number evolution: a Bayesian model averaging approach. Syst. Biol. 67(2):195–215.
Freyman W.A., Höhna S. 2019. Stochastic character mapping of state-dependent diversification reveals the tempo of evolutionary decline in self-compatible Onagraceae lineages. Syst. Biol. 68(3):505–519.
Frumhoff P.C., Reeve H.K. 1994. Using phylogenies to test hypotheses of adaptation: a critique of some current proposals. Evolution 48(1):172–180.
Gardner J.D., Organ C.L. 2021. Evolutionary sample size and consilience in phylogenetic comparative analysis. Syst. Biol. 70(5):1061–1075.
Garland K.L., Hay E.M., Field D.J., Evans A.R. 2025. Common developmental origins of beak shapes and evolution in theropods. iScience 28(4):112246.
Gavryushkina A., Heath T.A., Ksepka D.T., Stadler T., Welch D., Drummond A.J. 2017. Bayesian total-evidence dating reveals the recent crown radiation of penguins. Syst. Biol. 66(1):57–73.
Gaya E., Redelings B.D., Navarro-Rosinés P., Llimona X., De Cáceres M., Lutzoni F. 2011. Align or not to align? Resolving species complexes within the *Caloplaca saxicola* group as a case study. Mycologia 103(2):361–378.
Goldberg E.E., Kohn J.R., Lande R., Robertson K.A., Smith S.A., Igić B. 2010. Species selection maintains self-incompatibility. Science 330(6003):493–495.
Goldberg E.E., Lancaster L.T., Ree R.H. 2011. Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. Syst. Biol. 60(4):451–465.
Goldberg E.E., Foo J. 2020. Memory in trait macroevolution. Am. Nat. 195(2):300–314.
Goldberg E.E., Igić B. 2008. On phylogenetic tests of irreversible evolution. Evolution 62(11):2727–2741.
Goldberg E.E., Igić B. 2012. Tempo and mode in plant breeding system evolution. Evolution 66(12):3701–3709.
Goldman N., Anderson J.P., Rodrigo A.G. 2000. Likelihood-based tests of topologies in phylogenetics. Syst. Biol. 49(4):652–670.
González-Ramírez I.S., Deanna R., Smith S.D. 2026. Late Cretaceous origins for major nightshade lineages from total evidence timetree analysis. Annals of Botany. https://doi.org/10.1093/aob/mcag011
Goolsby E.W. 2015. Phylogenetic comparative methods for evaluating the evolutionary history of function-valued traits. Syst. Biol. 64(4):568–578.
Hadfield J.D. 2015. Increasing the efficiency of MCMC for hierarchical phylogenetic models of categorical traits using reduced mixed models. Methods Ecol. Evol. 6(6):706–714.
Hand S.J., Maugoust J., Beck R.M., Orliac M.J. 2023. A 50-million-year-old, three-dimensionally preserved bat skull supports an early origin for modern echolocation. Curr. Biol. 33(21):4624–4640.
Hansen T.F. 1997. Stabilizing selection and the comparative analysis of adaptation. Evolution 51(5):1341–1351.
Harmon L.J. 2018. Phylogenetic comparative methods: learning from trees. Minneapolis, MN Open Textbook Library: CreateSpace Independent Publishing Platform.
Harmon L.J., Losos J.B., Jonathan Davies T., Gillespie R.G., Gittleman J.L., Bryan Jennings W., Kozak K.H., McPeek M.A., Moreno-Roark F., Near T.J., Purvis A., Ricklefs R.E., Schluter D, Schulte Ii J.A., Seehausen O., Sidlauskas B.L., Torres-Carvajal O., Weir J.T., Mooers A.Ø. 2010. Early bursts of body size and shape evolution are rare in comparative data. Evolution 64(8):2385–2396.
Heath T.A., Huelsenbeck J.P., Stadler T. 2014. The fossilized birth–death process for coherent calibration of divergence-time estimates. Proc. Natl. Acad. Sci. USA 111(29):E2957–E2966.
Helfenbein K.G., DeSalle R. 2005. Falsifications and corroborations: Karl Popper's influence on systematics. Mol. Phylogenet. Evol. 35(1):271–280.
Hobolth A. 2008. A Markov chain Monte Carlo expectation maximization algorithm for statistical analysis of DNA sequence evolution with neighbor-dependent substitution rates. J. Comput. Graph. Stat. 17(1):138–162.
Höhna S., Landis M.J., Heath T.A., Boussau B., Lartillot N., Moore B.R., Huelsenbeck J.P., Ronquist F. 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. Syst. Biol. 65(4):726–736.
Holmes I., Bruno W.J. 2001. Evolutionary HMMs: a Bayesian approach to multiple alignment. Bioinformatics 17(9):803–820.
Huelsenbeck J.P., Larget B., Miller R.E., Ronquist F. 2002. Potential applications and pitfalls of Bayesian inference of phylogeny. Syst. Biol. 51(5):673–688.
Huelsenbeck J.P., Nielsen R., Bollback J.P. 2003. Stochastic mapping of morphological characters. Syst. Biol. 52(2):131–158.
Huelsenbeck J.P., Bollback J.P. 2001. Empirical and hierarchical Bayesian estimation of ancestral states. Syst. Biol. 50(3):351–366.
Huey R.B., Garland T., Turelli M. 2019. Revisiting a key innovation in evolutionary biology: Felsenstein's "phylogenies and the comparative method". Am. Nat. 193(6):755–772.
Iles W.J., Sass C., Lagomarsino L., Benson-Martin G., Driscoll H., Specht C.D. 2017. The phylogeny of Heliconia (Heliconiaceae) and the evolution of floral presentation. Mol. Phylogenet. Evol. 117:150–167.
Irvahn J., Minin V.N. 2014. Phylogenetic stochastic mapping without matrix exponentiation. J. Comput. Biol. 21(9):676–690.
Jukes T.H., Cantor C.R. 1969. Evolution of protein molecules. Mammalian Protein Metabolism 3(21):132.
Kaplan N., Langley C.H. 1979. A new estimate of sequence divergence of mitochondrial DNA using restriction endonuclease mappings. J. Mol. Evol. 13:295–304.
Kapli P., Yang Z., Telford M.J. 2020. Phylogenetic tree building in the genomic age. Nat. Rev. Genet. 21(7):428–444.
Katoh K., Asimenos G., Toh H. 2009. Multiple alignment of DNA sequences with MAFFT. Bioinformatics for DNA Sequence Analysis 537:39–64.
Kemena C., Notredame C. 2009. Upcoming challenges for multiple sequence alignment methods in the high-throughput era. Bioinformatics 25(19):2455–2465.
Khabbazian M., Kriebel R., Rohe K., Ané C. 2016. Fast and accurate detection of evolutionary shifts in Ornstein–Uhlenbeck models. Methods Ecol. Evol. 7(7):811–824.
Koshi J.M., Goldstein R.A. 1996. Probabilistic reconstruction of ancestral protein sequences. J. Mol. Evol. 42:313–320.
Kubo T., Iwasa Y. 1995. Inferring the rates of branching and extinction from molecular phylogenies. Evolution 49(4):694–704.
Landis M.J., Schraiber J.G., Liang M. 2013. Phylogenetic analysis using Lévy processes: finding jumps in the evolution of continuous traits. Syst. Biol. 62(2):193–204.
Lee M., Oliver P., Hutchinson M. 2009. Phylogenetic uncertainty and molecular clock calibrations: a case study of legless lizards (Pygopodidae, Gekkota). Mol. Phylogenet. Evol. 50(3):661–666.
Lee M.S. 2016. Multiple morphological clocks and total-evidence tip-dating in mammals. Biol. Lett. 12(7):20160033.
Lemey P., Rambaut A., Drummond A.J., Suchard M.A. 2009. Bayesian phylogeography finds its roots. PLoS Comput. Biol. 5(9): e1000520.
Lewis P.O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst. Biol. 50(6): 913–925.
Liu J.W., Bogarín D., Pérez-Escobar O.A., Pupulin F., Karremans A.P., Serracín Z., Xie Y., Restrepo E., Ramírez S.R. 2025. Exceptional diversification of floral form in a specialized orchid pollination system. bioRxiv. https://doi.org/10.1101/2025.07.25.666443
Losos J.B. 2011. Seeing the forest for the trees: the limitations of phylogenies in comparative biology: (American Society of Naturalists address). Am. Nat. 177(6):709–727.
Louca S., Pennell M.W. 2020a. Extant timetrees are consistent with a myriad of diversification histories. Nature 580(7804):502–505.
Louca S., Pennell M.W. 2020b. A general and efficient algorithm for the likelihood of diversification and discrete-trait evolutionary models. Syst. Biol. 69(3):545–556.
Löytynoja A., Goldman N. 2005. An algorithm for progressive multiple alignment of sequences with insertions. Proc. Natl. Acad. Sci. USA 102(30):10557–10562.
Luckow M., Bruneau A. 1997. Circularity and independence in phylogenetic tests of ecological hypotheses. Cladistics 13(1–2):145–151.

Downloaded from https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syag034/8651289 by guest on 21 May 2026

Lunter G., Miklós I., Drummond A., Jensen J.L., Hein J. 2005. Bayesian coestimation of phylogeny and sequence alignment. BMC Bioinformatics 6(1):83.
Maddison W.P. 1993. Missing data versus missing characters in phylogenetic analysis. Syst. Biol. 42(4):576–581. https://doi.org/10.2307/2992490
Maddison W.P. 2006. Confounding asymmetries in evolutionary diversification and character change. Evolution 60(8):1743–1746.
Maddison W.P., Midford P.E., Otto S.P. 2007. Estimating a binary character's effect on speciation and extinction. Syst. Biol. 56(5):701–710.
Maddison W.P., FitzJohn R.G. 2015. The unsolved challenge to phylogenetic correlation tests for categorical characters. Syst. Biol. 64(1):127–136.
Magnuson-Ford K., Otto S.P. 2012. Linking the investigations of character evolution and species diversification. Am. Nat. 180(2):225–245.
Mantica F., Iñiguez L.P., Marquez Y., Permanyer J., Torres-Mendez A., Cruz J., Franch-Marro X., Tulenko F., Burguera D., Bertrand S., Doyle T., Nouzova M., Currie P.D., Noriega F.G., Escriva H., Arnone M.I., Albertin C.B., Wotton K.R., Almudi I., Martin D., Irimia M. 2024. Evolution of tissue-specific expression of ancestral genes across vertebrates and insects. Nat. Ecol. Evol. 8(6):1140–1153. https://doi.org/10.1038/s41559-024-02398-5
Martin B., Weber M. 2026. Stochastic character mapping of continuous traits on phylogenies. Systematic Biologyhttps://doi.org/10.1093/sysbio/syag031
Martins E.P., Hansen T.F. 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. Am. Nat. 149(4):646–667.
May M.R., Contreras D.L., Sundue M.A., Nagalingum N.S., Looy C.V., Rothfels C.J. 2021. Inferring the total-evidence timescale of marattialean fern evolution in the face of model sensitivity. Syst. Biol. 70(6):1232–1255.
May M.R., Meyer X. 2025. TensorPhylo RevBayes plugin. Available from: https://bitbucket.org/mrmay/tensorphylo/src/master
May M.R., Moore B.R. 2020. A Bayesian approach for inferring the impact of a discrete character on rates of continuous-character evolution in the presence of background-rate variation. Syst. Biol. 69(3):530–544. https://doi.org/10.1093/sysbio/syz069
May M.R., Rothfels C.J. 2023. Diversification models conflate likelihood and prior, and cannot be compared using conventional model-comparison tools. Syst. Biol. 72(3):713–722.
Mayrose I., Barker M.S., Otto S.P. 2010. Probabilistic models of chromosome number evolution and the inference of polyploidy. Syst. Biol. 59(2):132–144.
Mayrose I., Zhan S.H., Rothfels C.J., Magnuson-Ford K., Barker M.S., Rieseberg L.H., Otto S.P. 2011. Recently formed polyploid plants diversify at lower rates. Science 333(6047):1257–1257.
Mayrose I., Lysak M.A. 2021. The evolution of chromosome numbers: mechanistic models and experimental approaches. Genome Biol. Evol. 13(2):evaa220.
Meseguer A.S., Carrillo R., Graham S.W., Sanmartín I. 2022. Macroevolutionary dynamics in the transition of angiosperms to aquatic environments. New Phytol. 235(1):344–355.
Mika K., Whittington C.M., McAllan B.M., Lynch V.J. 2022. Gene expression phylogenies and ancestral transcriptome reconstruction resolves major transitions in the origins of pregnancy. eLife 11:e74297. https://doi.org/10.7554/eLife.74297
Miklós I., Lunter G.A., Holmes I. 2004. A "long indel" model for evolutionary sequence alignment. Mol. Biol. Evol. 21(3):529–540.
Minin V.N., Suchard M.A. 2008. Fast, accurate and simulation-free stochastic mapping. Philos. T. R. Soc. B: Biol. Sci. 363(1512):3985–3995.
Mongiardino Koch N., Garwood R.J., Parry L.A. 2021. Fossils improve phylogenetic analyses of morphological characters. Proc. Roy. Soc. B 288(1950):20210044.
Mongiardino Koch N., Parry L.A. 2020. Death is on our side: paleontological data drastically modify phylogenetic hypotheses. Syst. Biol. 69(6):1052–1067.
Morlon H., Parsons T.L., Plotkin J.B. 2011. Reconciling molecular phylogenies with the fossil record. Proc. Natl. Acad. Sci. USA 108(39):16327–16332.
Nee S., May R.M., Harvey P.H. 1994. The reconstructed evolutionary process. Philos. T. R. Soc. London Ser. B Biol. Sci. 344(1309):305–311.
Neyman J. 1971. Molecular studies of evolution: a source of novel statistical problems. In: Gupta S.S., Yackel J., editors. Statistical Decision Theory and Related Topics New York: Academic Press. p. 1–27.
Ng J., Smith S.D. 2018. Why are red flowers so rare? Testing the macroevolutionary causes of tippiness. J. Evol. Biol. 31(12):1863–1875.
Nielsen R. 2002. Mapping mutations on phylogenies. Syst. Biol. 51(5):729–739.
Nyathi S., Rezende I.M., Walter K.S., Thongsripong P., Mutuku F., Ndenga B., Mbakaya J.O., Agola G., Vu D.M., Bennett S., Mordecai E.A., Andrews J.R., LaBeaud A.D. 2025. Geographic origin and evolution of dengue virus serotypes 1 and 3 circulating in Africa. Virus Evol. 11(1):veae116.
Nylander J.A., Ronquist F., Huelsenbeck J.P., Nieves-Aldrey J. 2004. Bayesian phylogenetic analysis of combined data. Syst. Biol. 53(1):47–67.
Ogburn R.M., Edwards E.J. 2015. Life history lability underlies rapid climate niche evolution in the angiosperm clade Montiaceae. Mol. Phylogenet. Evol. 92:181–192.
Olmstead R. 2001. Phylogenetic inference and the writings of Karl Popper. Syst. Biol. 50(3):304.
O'Meara B.C. 2012. Evolutionary inferences from phylogenies: a review of methods. Annu. Rev. Ecol. Evol. Syst. 43(1):267–285.
O'Meara B.C., Ané C., Sanderson M.J., Wainwright P.C. 2006. Testing for different rates of continuous trait evolution using likelihood. Evolution 60(5):922–933.
Organ C., Nunn C.L., Machanda Z., Wrangham R.W. 2011. Phylogenetic rate shifts in feeding time during the evolution of *Homo*. Proc. Natl. Acad. Sci. USA 108(35):14555–14559.
Pagel M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proc. Roy. Soc. London Ser. B Biol. Sci. 255(1342):37–45.
Pagel M. 1999a. Inferring the historical patterns of biological evolution. Nature 401(6756):877–884.
Pagel M. 1999b. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. Syst. Biol. 48(3):612–622.
Pagel M., Meade A., Barker D. 2004. Bayesian estimation of ancestral character states on phylogenies. Syst. Biol. 53(5):673–684.
Pagel M., Lutzoni F. 2002. Accounting for phylogenetic uncertainty in comparative studies of evolution and adaptation. Biological evolution and statistical physics. London: Springer. p. 148–161.
Pagel M., Meade A. 2006. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. Am. Nat. 167(6):808–825.
Paluh D.J., Riddell K., Early C.M., Hantak M.M., Jongsma G.F., Keeffe R.M., Magalhães Silva F., Nielsen S.V., Vallejo-Pareja M.C., Stanley E.L., Blackburn D.C. 2021. Rampant tooth loss across 200 million years of frog evolution. eLife 10:e66926.
Paradis E., Claude J., Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20(2):289–290.
Paradis E., Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35:526–528. https://doi.org/10.1093/bioinformatics/bty633
Popper K. 1959. The logic of scientific discovery. London: Julius Springer, Hutchinson & Co.
Porto D.S., Uyeda J., Mikó I., Tarasov S. 2024. ontophylo: reconstructing the evolutionary dynamics of phenomes using new ontology-informed phylogenetic methods. Methods Ecol. Evol. 15(2):290–300. https://doi.org/10.1111/2041-210X.14283
Pryer K.M., Schuettpelz E., Wolf P.G., Schneider H., Smith A.R., Cranfill R. 2004. Phylogeny and evolution of ferns (monilophytes) with a focus on the early leptosporangiate divergences. Am. J. Bot. 91(10):1582–1598.
Pupko T., Pe I., Shamir R., Graur D. 2000. A fast algorithm for joint reconstruction of ancestral amino acid sequences. Mol. Biol. Evol. 17(6):890–896.

Downloaded from https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syag034/8651289 by guest on 21 May 2026

Quental T.B., Marshall C.R. 2010. Diversity dynamics: molecular phylogenies need the fossil record. Trends Ecol. Evol. 25(8): 434–441.
Rabosky D.L., Goldberg E.E. 2015. Model inadequacy and mistaken inferences of trait-dependent speciation. Syst. Biol. 64(2):340–355.
Rannala B., Yang Z. 1996. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. J. Mol. Evol. 43:304–311.
Redelings B.D. 2014. Erasing errors due to alignment ambiguity when estimating positive selection. Mol. Biol. Evol. 31(8):1979–1993.
Redelings B.D. 2021. BAli-Phy version 3: model-based co-estimation of alignment and phylogeny. Bioinformatics 37(18):3032–3034.
Redelings B.D., Holmes I., Lunter G., Pupko T., Anisimova M. 2024. Insertions and deletions: Computational methods, evolutionary dynamics, and biological applications. Mol. Biol. Evol. 41(9): msae177.
Redelings B.D., Suchard M.A. 2005. Joint Bayesian estimation of alignment and phylogeny. Syst. Biol. 54(3):401–418.
Revell L.J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3:217–223.
Revell L.J. 2013. A comment on the use of stochastic character maps to estimate evolutionary rate variation in a continuously valued trait. Syst. Biol. 62(2):339–345. https://doi.org/10.1093/sysbio/sys084
Revell L.J. 2014. Ancestral character estimation under the threshold model from quantitative genetics. Evolution 68(3):743–759.
Revell L.J. 2025. Ancestral state reconstruction of phenotypic characters. Evol. Biol. 52:1–25.
Revell L.J., Harmon L.J., Collar D.C. 2008. Phylogenetic signal, evolutionary process, and rate. Syst. Biol. 57(4):591–601.
Revell L.J., Harmon L.J. 2022. Phylogenetic comparative methods in R. Princeton, NJ; Oxford: Princeton University Press.
Rieppel O. 2003. Popper and systematics. Syst. Biol. 52(2):259–271.
Rohle F.J. 2006. A comment on phylogenetic correction. Evolution 60(7):1509–1515.
Ronquist F. 2004. Bayesian inference of character evolution. Trends Ecol. Evol. 19(9):475–481.
Ronquist F., Klopfstein S., Vilhelmsen L., Schulmeister S., Murray D.L., Rasnitsyn A.P. 2012. A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. Syst. Biol. 61(6):973–999.
Rothfels C.J., Schuettpelz E. 2014. Accelerated rate of molecular evolution for vittarioid ferns is strong and not driven by selection. Syst. Biol. 63(1):31–54.
Sankoff D., Morel C., Cedergren R.J. 1973. Evolution of 5S RNA and the non-randomness of base replacement. Nat. New Biol. 245(147):232–234.
Sauquet H., Ho S.Y., Gandolfo M.A., Jordan G.J., Wilf P., Cantrill D.J., Bayly M.J., Bromham L., Brown G.K., Carpenter R.J., Lee D.M., Murphy D.J., Sniderman J.M.L., Udovicic F. 2012. Testing the impact of calibration on molecular divergence times using a fossil-rich group: the case of *Nothofagus* (Fagales). Syst. Biol. 61(2):289–313.
Schluter D., Price T., Mooers A.Ø., Ludwig D. 1997. Likelihood of ancestor states in adaptive radiation. Evolution 51(6):1699–1711.
Senderov V., Kudlicka J., Lundén D., Palmkvist V., Braga M.P., Granqvist E., Çaylak G., Virgoulay T., Broman D., Ronquist F. 2024. Treeppl: a universal probabilistic programming language for phylogenetics. bioRxiv. https://doi.org/10.1101/2023.10.10.561673
Sharma C.V.R. 2025. Beyond the Matrix: Assessing the Impact of Multiple Sequence Alignment and Codon Sequence Models on the Accuracy of Ancestral Sequence Reconstructions. Ph.D. thesis, Brandeis University.
Shim H., Larget B. 2018. Bayescat: Bayesian co-estimation of alignment and tree. Biometrics 74(1):270–279.
Siddall M.E., Kluge A.G. 1997. Probabilism and phylogenetic inference. Cladistics 13(4):313–336.
Sober E. 1991. Reconstructing the past: parsimony, evolution, and inference. Cambridge, MA; London: MIT Press.
Spear J.K., Grabowski M., Sekhavati Y., Costa C.E., Goldstein D.M., Petrullo L.A., Peterson A.L., Lee A.B., Shattuck M.R., Gómez-Olivencia A., Williams S.A. 2023. Evolution of vertebral numbers in primates, with a focus on hominoids and the last common ancestor of hominins and panins. J. Hum. Evol. 179:103359.
Sperotto P., Vasconcelos T., Kriebel R., Reginato M. 2025. Evolutionary shift in petal function from pollinator attraction to water drainage in small buzz-pollinated flowers. Evolution 79:qpaf183.
Suchard M.A., Redelings B.D. 2006. BAli-Phy: simultaneous Bayesian inference of alignment and phylogeny. Bioinformatics 22(16):2047–2048.
Suissa J.S., Friedman W.E. 2022. Rapid diversification of vascular architecture underlies the carboniferous fern radiation. Proc. Roy. Soc. B 289(1973):20212209.
Swiston S.K., Landis M.J. 2025. Testing relationships between multiple regional features and biogeographic processes of speciation, extinction, and dispersal. Syst. Biol. 74(2):282–300.
Symonds M.R.E., Blomberg S.P. 2014. A primer on phylogenetic generalised least squares. In: Garamszegi L.Z., editor. Modern phylogenetic comparative methods and their application in evolutionary biology: concepts and practice. Berlin, Heidelberg: Springer. p. 105–130.
Tarasov S. 2019. Integration of anatomy ontologies and evo-devo using structured Markov models suggests a new framework for modeling discrete phenotypic traits. Syst. Biol. 68(5):698–716.
Tarasov S. 2023. New phylogenetic Markov models for inapplicable morphological characters. Syst. Biol. 72(3):681–693.
Tarasov S., Mikó I., Yoder M.J., Uyeda J.C. 2019. PARAMO: a pipeline for reconstructing ancestral anatomies using ontologies and stochastic mapping. Insect Syst. Divers. 3(6):1.
Tejada J.V., Antoine P.O., Münch P., Billet G., Hautier L., Delsuc F., Condamine F.L. 2024. Bayesian total-evidence dating revisits sloth phylogeny and biogeography: a cautionary tale on morphological clock analyses. Syst. Biol. 73(1):125–139.
Thomas G.H., Freckleton R.P., Székely T. 2006. Comparative analyses of the influence of developmental mode on phenotypic diversification rates in shorebirds. Proc. Roy. Soc. B Biol. Sci. 273(1594):1619–1624.
Thompson A., May M.R., Hopkins B., Riedl N., Barmina O., Liebeskind B.J., Zhao L., Begun D., Kopp A. 2024. Quantifying transcriptome turnover on phylogenies by modeling gene expression as a binary trait. bioRxiv. https://doi.org/10.1101/2024.10.03.616564
Thorne J., Kishino H. 1992. Freeing phylogenies from artifacts of alignment. Mol. Biol. Evol. 9(6):1148–1162.
Thorne J.L., Kishino H., Felsenstein J. 1991. An evolutionary model for maximum likelihood alignment of DNA sequences. J. Mol. Evol. 33:114–124.
Title P.O., Henao-Díaz L.F., Zenil-Ferguson R., Vasconcelos T. 2025. An evolving view of lineage diversification. Syst. Biol. 75: syaf086.
Tolkoff M.R., Alfaro M.E., Baele G., Lemey P., Suchard M.A. 2018. Phylogenetic factor analysis. Syst. Biol. 67(3):384–399. https://doi.org/10.1093/sysbio/syx066
Tribble C.M., May M.R., Jackson-Gain A., Zenil-Ferguson R., Specht C.D., Rothfels C.J. 2023. Unearthing modes of climatic adaptation in underground storage organs across Liliales. Syst. Biol. 72(1):198–212.
Tribble C.M., Márquez-Corro J.I., May M.R., Hipp A.L., Escudero M., Zenil-Ferguson R. 2025. Macroevolutionary inference of complex modes of chromosomal speciation in a cosmopolitan plant lineage. New Phytol. 245(5):2350–2361.
Tuffley C., Steel M. 1998. Modeling the covarion hypothesis of nucleotide substitution. Math. Biosci. 147(1):63–91.
Uyeda J.C., Zenil-Ferguson R., Pennell M.W. 2018. Rethinking phylogenetic comparative methods. Syst. Biol. 67(6):1091–1109. https://doi.org/10.1093/sysbio/syy031
Uyeda J.C., Harmon L.J. 2014. A novel Bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. Syst. Biol. 63(6): 902–918.
Vialle R.A., Tamuri A.U., Goldman N. 2018. Alignment modulates ancestral sequence reconstruction accuracy. Mol. Biol. Evol. 35(7):1783–1797.

Downloaded from https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syag034/8651289 by guest on 21 May 2026

Vogt L. 2008. The unfalsifiability of cladograms and its consequences. Cladistics 24(1):62–73.
Warnock R.C., Wright A.M. 2020. Understanding the tripartite approach to Bayesian divergence time estimation. Cambridge: Cambridge University Press.
Wasserstein R.L., Lazar N.A. 2016. The ASA statement on *p*-values: context, process, and purpose.The American Statistician 70(2):129–133. https://doi.org/10.1080/00031305.2016.1154108
Westesson O., Lunter G., Paten B., Holmes I. 2012. Accurate reconstruction of insertion-deletion histories by statistical phylogenetics. PLoS One 7(4):e34572.
Westoby M., Leishman M. 1995. Issues of interpretation after relating comparative datasets to phylogeny. J. Ecol. 83(5):892–893.
Wolff J.O., Wierucka K., Uhl G., Herberstein M.E. 2021. Building behavior does not drive rates of phenotypic evolution in spiders. Proc. Natl. Acad. Sci. USA 118(33):e2102693118.
Wong K.M., Suchard M.A., Huelsenbeck J.P. 2008. Alignment uncertainty and genomic analysis. Science 319(5862):473–476.
Wright S. 1934. An analysis of variability in number of digits in an inbred strain of guinea pigs. Genetics 19(6):506–536.
Xue B., Guo X., Landis J., Sun M., Tang C., Soltis P., Soltis D., Saunders R. 2020. Accelerated diversification correlated with functional traits shapes extant diversity of the early divergent angiosperm family Annonaceae. Mol. Phylogenet. Evol. 142:106659.
Yang Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol. Biol. Evol. 15(5):568–573.
Yang Z. 2014. Molecular evolution: a statistical approach. Oxford: Oxford University Press.
Yang Z., Kumar S., Nei M. 1995. A new method of inference of ancestral nucleotide and amino acid sequences. Genetics 141(4): 1641–1650.
Yang Z., Rannala B. 1997. Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo method. Mol. Biol. Evol. 14(7):717–724.
Zenil-Ferguson R., Ponciano J.M., Burleigh J.G. 2017. Testing the association of phenotypes with polyploidy: an example using herbaceous and woody eudicots. Evolution 71(5):1138–1148.
Zenil-Ferguson R., Burleigh J.G., Freyman W.A., Igić B., Mayrose I., Goldberg E.E. 2019. Interaction among ploidy, breeding system and lineage diversification. New Phytol. 224(3):1252–1265.
Zharkikh A. 1994. Estimation of evolutionary distances between nucleotide sequences. J. Mol. Evol. 39(3):315–329.

Downloaded from https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syag034/8651289 by guest on 21 May 2026

Syst. Biol. 00(0):1–25, 2026
© The Author(s) 2026. Published by Oxford University Press on behalf of the Society of Systematic Biologists. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site-for further information please contact journals.permissions@oup.com
https://doi.org/10.1093/sysbio/syag034
Advance Access publication April 10, 2026