An evolving view of character macroevolution

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

[1]*Department of Biology and Burke Museum, University of Washington, Seattle, WA 98195*
[2]*Department of Plant and Microbial Biology and Integrative Biology, University of California, Berkeley, CA 94720*
[3]*Department of Biology and Ecology Center, Utah State University, Logan, UT 84322*
[*]*Authors contributed equally*
[†]*tribblec@uw.edu*

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 phylogenetic hypothesis and uses that phylogeny to "correct" for the non-independence of the character data that they wish to analyze. Here we argue that parallel developments in philosophy, data availability, computational capacity, and model development have led away from this classic view of PCMs and towards 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 conceptual trajectory, 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 multiple current topics in character evolution, 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 sequences, 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.

keywords: 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 statistical 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 residuals of a continuous trait evolving on phylogeny are independent under a Brownian model of evolution. This approach of using residuals, 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 (often considered a necessary nuisance) and then subsequently one can perform a comparative analysis "correcting" for the phylogeny in order to illuminate the focal biological feature of interest. There were other practical reasons for the initial prevalence of this step-wise approach. First, comparative biologists were often uninterested in phylogeny itself (the molecular phylogenetic revolution had only just begun, phylogenies were uncommon, and phylogenetics, as a field, was not as strongly integrated with other areas of evolutionary biology as it is today; Losos, 2011). Perhaps more critically, phylogenetic models were in their infancy: we had very limited abilities to 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.

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). 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; Luckow and Bruneau, 1997). 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 a real concern: if one's dataset 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 towards their hypothesis of choice, resulting in inaccurate conclusions (Frumhoff and Reeve, 1994; de Queiroz, 1996).

This two-step paradigm has been long dominant and is still common—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 pedagogical effectiveness, 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 has reduced our dependence on this classic two-step approach. First, greater conceptual clarity, and a move to embrace phylogenetics as an inferential science rather than a necessarily hypothetico-deductive one, mitigated the circularity concerns (see Box 1; Luckow and Bruneau, 1997; Ronquist, 2004). For example, it is more natural, within an inferential framework, to wish 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"; de Queiroz, 1996) with respect to the focal hypothesis. This shift away from the two-step approach was accelerated by increased data availability in the form of molecular datasets—practical concerns about the influence of a few characters are reduced when one's dataset totals thousands of characters. Complementing both the conceptual and data-availability trends were 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. Researchers now had the tools at their disposal to perform joint inference in a reasonable way and had fewer practical or philosophical reasons to prefer the two-step approach. This intersection of increased data availability, modeling options, and computation 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 led naturally to more integrative approaches: a Bayesian framework provides a natural means of incorporating multiple types of data (see Box 2).

Box 1: The shadow of Popper

The paradigm 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. Much of the early period of systematics was marked by the pervasive influence of 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 falsified 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 (*e.g.*, Popper, 1959). While Popper's 1959 *The Logic of Scientific Discovery* also touches on probability theory, the systematics community largely ignored this nuance and responded to Popper's focus on deductive reasoning with mental contortions to show that particular methodologies (*i.e.*, parsimony- or likelihood-based approaches) better fit Popper's logical formalism and thus were more scientific (*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—philosophical details aside—in promoting the general view that hypothesis-testing approaches, such as underlie two-step PCM approaches to understanding character evolution, are superior to purely inferential alternatives. After all, inference (induction) was not logically capable of *demonstrating* (proving) anything (see Rieppel, 2003) and science was based on deduction and falsifiability. The resulting hypothetico-deductive focus led to many of the earlier works on comparative methods 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 towards the focal hypothesis. If any bias, given the outcome of the analysis, runs counter to the focal hypothesis, then you're all good (the analysis is "reliable"); the question of accuracy was secondary (de Queiroz, 1996).

Once clearly the 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)—or conversely, that it is equally applicable to both parsimony- and likelihood-based approaches, and thus useless for choosing between them (de Queiroz, 2014)—and to the rise of unabashedly inferential Bayesian approaches (Ronquist, 2004). While 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 more likely to also recognize that they are *estimates*. This reconceptualization naturally leads us towards a joint-inference framework—if we want our estimates to be as accurate as possible, we presumably want to include all available data that might inform that estimate. This shift, in turn, 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, many of the things we might be interested in learning about character evolution do not naturally fit within a hypothetico-deductive framework. Is the number of times such and such a trait evolved a hypothesis? Is the degree to which such and such a trait is correlated with a particular other trait a hypothesis? These are probabilistic questions—not logical-deductive ones—that are accessible through the tools of inference rather than falsification.

Box 2: The light of Bayes

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

To illustrate this point, imagine we have two datasets, X_1 and X_2. Perhaps these datasets correspond to different DNA sequence alignments, or different morphological characters, or some mix of genetic and morphological data. Suppose we are interested in how these traits evolve together, or how they share information about some underlying variable of interest (for example, 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 datasets.

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:

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

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

However, we may also analyze the two datasets one at a time, treating the posterior from the first dataset as the prior for the second dataset:

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

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

This two-step procedure is 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 sequential analysis is identical to the posterior of the joint analysis (Eq. 1). This equivalency of the step-wise and simultaneous inference approaches, however, only holds when the second step of our inference updates our belief about θ (the posterior of the first step becomes the prior in the second, where it is updated by the data in X_2).

This logic 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 the other parameters of the model. If we instead fix the estimates from the first step (as in conventional PCMs), we obtain biased estimates of the posterior. Rephrasing, and pulling up from the formalism of Bayes theorem, if we believe that different datasets 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: joint inference is required.

We argue that this gradual practical and philosophical transition over the past 40 years, from two-step analysis, where the PCM is conditioned on a pre-existing phylogeny or set of phylogenies, towards a tendency to perform joint inference, has brought about a subtle paradigm shift in phylogenetics: phylogenies are not static structures upon which we can map our characters of interest, or which we should use to "correct" for non-independence. Instead, all our data evolved together and we should not *a priori* partition those data into a set to use for the phylogenetic inference and another set for investigating our comparative questions. This shift is perhaps most usefully illustrated by alternative conceptualizations of "phylogenetic signal." 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, under a Brownian motion (BM) model, and further that if this phylogenetic signal (or "phylogenetic effect") is low, the phylogeny is not useful in understanding the evolution of the trait (Pagel, 1999a; Blomberg et al., 2003; Losos, 2011) and the inclusion of phylogenetic information may even be inappropriate (Losos, 2011). Under the alternative "joint inference" paradigm, however, the lack of phylogenetic signal is simply an indication that our data of interest are poorly explained by Brownian motion—it's not that the phylogeny is irrelevant (do we think that our data are not the result of evolution, or that that evolution did not occur within the lineages of the phylogeny?) but rather that we are not modeling the evolution of our characters adequately. Interesting, a very similar critique—that Brownian motion is a poor model of evolutionary change—was made in the original work (Felsenstein, 1985) and formed the basis of arguments against these sorts of comparative analyses in general (Carpenter, 1992). Now, however, we are not limited to BM models, and have a wealth of approaches available to simultaneously infer multiple inter-related elements of the evolutionary process, including the phylogeny itself and patterns of character evolution.

In this review we provide a brief timeline (Fig. 1) of significant developments in the study of character evolution, a summary of the fundamental "building blocks" of character evolution models, followed by a set of focal discussions of major current topics in this area, with an emphasize on the opportunities provided by the paradigm shift from two-step procedures to joint inference.

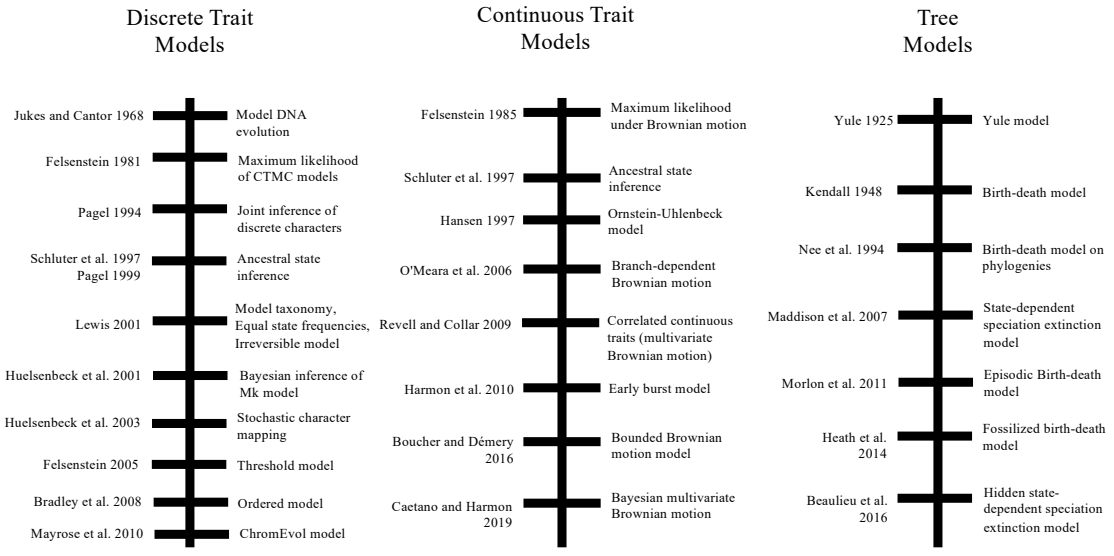

Figure 1: Timeline of select major events in the development of phylogenetic comparative methods, sorted by discrete trait models, continuous trait models, and tree models. We elaborate on some of these methods in the main text.

Fundamental models of trait evolution

Phylogenetic comparative models—whether applied sequentially or jointly—are complex and diverse. In this section we first describe at a high level how stochastic 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.

Phylogenetic comparative methods 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 phylogenetic 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). While 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 some root-state distribution. Next, we simulate the evolution of the trait according to the stochastic model with parameter values θ down 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.

See Figure 2 for an example of this procedure for a discrete-character model with two states (red and blue). In this case, the observed data are that taxon a is blue, while 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), while 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, 1969; Kaplan and Langley, 1979) and were first applied to infer phylogenies by Felsenstein (1973). These models are continuous-time Markov

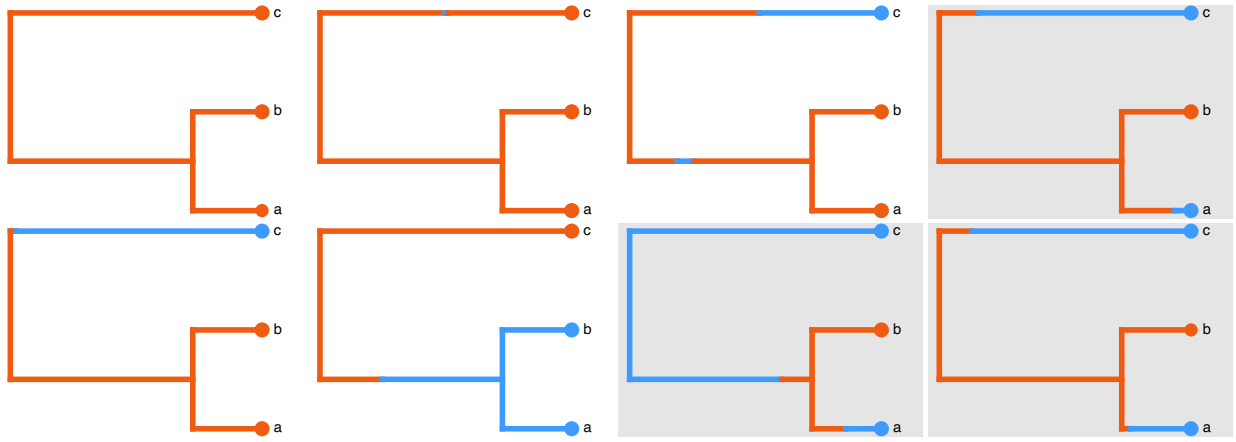

Figure 2: Simulated evolution of a two-state character (the states are red and blue) on a three-tip phylogeny.

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 a *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 as follows:

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

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 (as in the example above), 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), while the all-rates-different (ARD) model (Paradis et al., 2004) allows each transition rate to vary freely ($q_{ij} \neq q_{ji}$). 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 are set to zero, to reflect certain types of changes that may be biologically impossible. In some cases, this can result in "irreversible" models, where once a particular transition has occurred, the character can no longer reverse 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, while all other transition rates are zero (Fig. 3C,D). 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 "gain" rate), and downward transitions occur at another rate (the "loss" rate), though this is by no means required. An example rate matrix for an ordered trait with four 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. 3E,F; 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 $1B$, and create an instantaneous rate matrix reflecting dependent evolution:

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

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

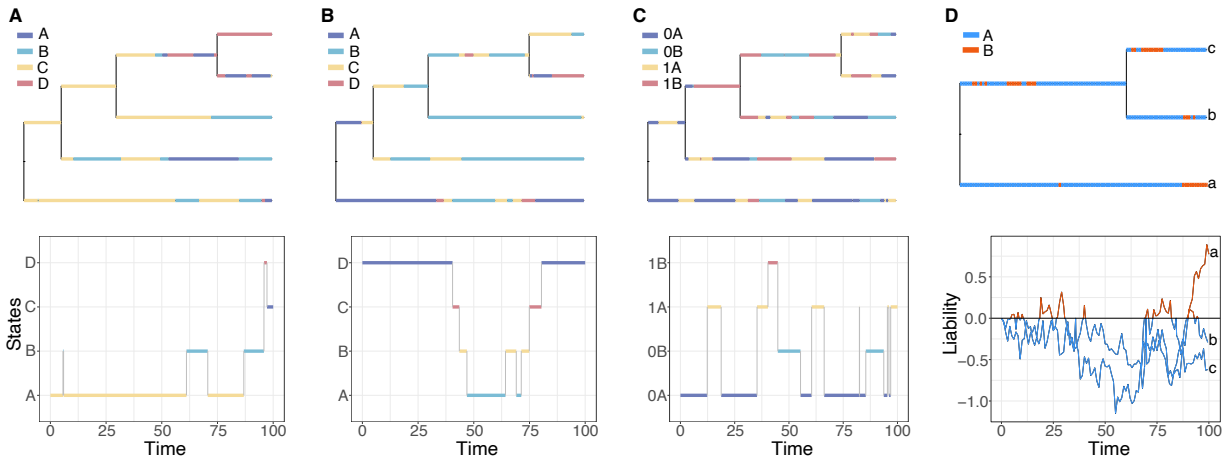

Figure 3: Simulated discrete trait evolution under various models: (A) symmetrical evolution, (B), ordered evolution where only transitions between adjacent states are possible (*e.g.*, $A \leftrightarrow B \leftrightarrow C \leftrightarrow D$), (C) dependent evolution of two binary traits, and (D) the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

transitions from 0 to 1 occur faster when the second character is in state A rather than B. 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, such "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, an alternative to CTMC discrete trait models) 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. 3G,H). 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 develop. 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 or some other continuous-trait model (Fig. 3G,H).

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. Since 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 (Felsenstein, 2005, see 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 need to be inferred (Revell, 2014; Cybis et al., 2015).

Continuous trait models

Most models of continuous trait evolution are based on Brownian motion (a "random walk"), introduced to systematics by Felsenstein (1985). The Brownian motion model is characterized by a single parameter, σ^2 (sometimes called a diffusion parameter; Fig. 4A,B). Under Brownian motion, at any given point in time an increase in the trait value is as likely as a decrease—the trait value wanders (hence "random 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 Brownian motion is the Ornstein–Uhlenbeck (OU) model, which describes the evolution of a trait towards some optimum value θ (Hansen 1997; Butler and King 2004; Fig. 4C,D). Like Brownian motion, 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 towards θ, denoted α. Unlike BM, where variance grows without bound as $t \to \infty$, under the OU model the variance will grow until it reaches an asymptote defined by the stationary distribution $\frac{\sigma^2}{2\alpha}$. Larger values of α result in less variance as the strength of the pull is stronger.

The OU model is only one of many derivatives of the BM model; there many others including bounded Brownian motion (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 parameters across the tree, allowing these parameters to vary in a branch-dependent manner (O'Meara et al. 2006; Uyeda and Harmon 2014; Khabbazian et al. 2016; Fig. 4E,F).

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 Brownian motion 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 case the traits are said to be "evolutionary correlated."

Birth-death tree models

The birth-death process models the dichotomous branching process that produces phylogenetic trees, but birth-death models and their derivatives also have 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 "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., 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

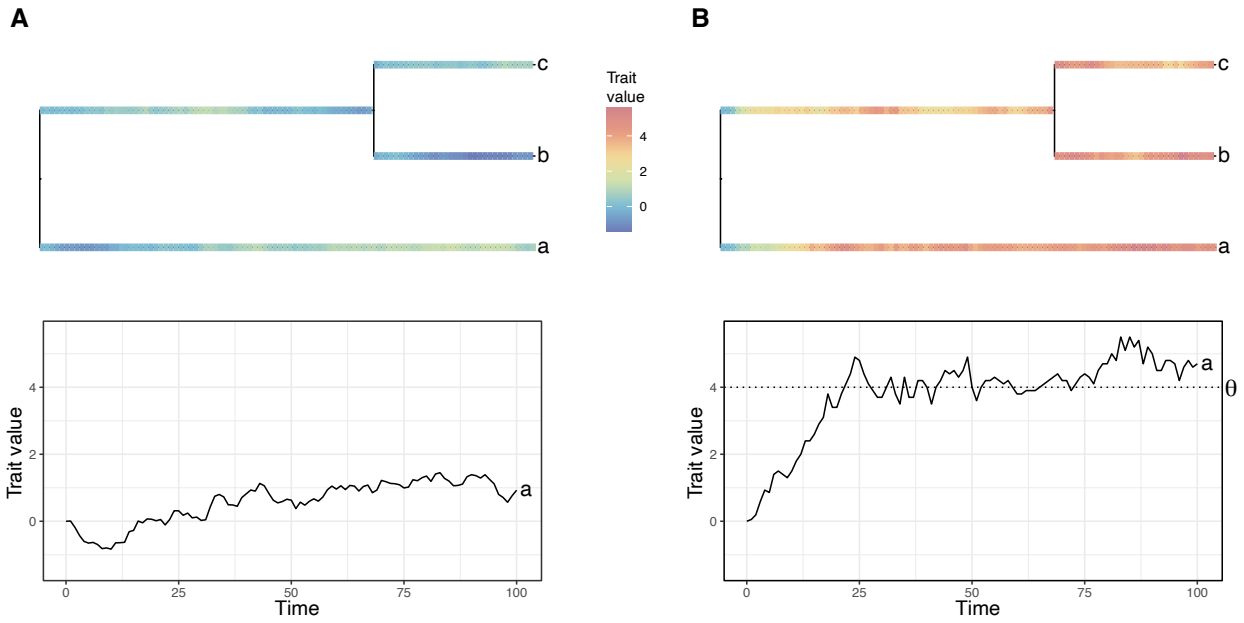

Figure 4: Continuous trait simulated under (A) Brownian motion and (B) OU. For both (A) and (B), the top row shows stochastic maps of the ancestral values of the traits along the phylogeny, while the bottom row shows a sample history—the trait value as a function of time—for the branch leading to species *a*. θ indicates the optimum value for the OU model in (B).

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 datasets that include extinct samples (see the Joint inference in divergence-time section, below). 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 detail below.

Major areas in the study of character evolution

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 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: while Maddison's (2006) insight is most often applied to the inference of trait-associated diversification rates (*i.e.*, state-dependent speciation and extinction rates), it equally applies to inferences of the character evolution parameters themselves (*i.e.*, the transition rates): if a trait affects diversifica-

tion 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 "state-dependent speciation and extinction" (SSE) framework. For example, while the original model, `BiSSE`, was limited to a single binary character (Maddison et al., 2007), that model has been expanded upon 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), since those models need to "keep track" of 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 (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 Popper Box)—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 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. Since 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 "regular" 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 Accounting for evolutionary process variation).

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 models 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 "Accounting for evolution process variation" below).

One important limitation of most SSE approaches from the perspective of joint inference is that while 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 diversification 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 estimates of divergence times, which in turn may compromise our ability to detect state-dependent diversification rates (Fig. 5).

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 fundamental models 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 Ornstein-Uhlenbeck model to include multiple optima (θ): the multi-optima 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 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-optima OU model of body-size evolution (a continuous character). In a Bayesian interpretation of this approach, assigning optima to branches essentially imposes an infinitely strong prior on the history of the discrete state while inferring the continuous trait history (see Box 2).

Shortly after Butler and King (2004), O'Meara et al. (2006) proposed another two-step approach: a multi-rate Brownian motion 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 (σ^2), and "attraction" parameters (α). Col-

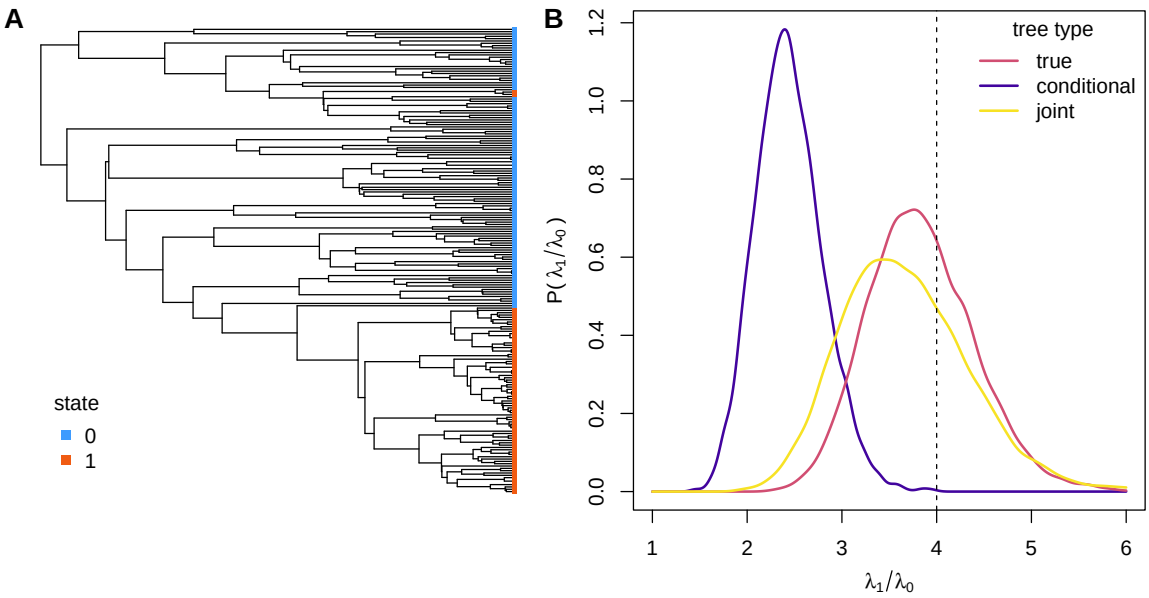

Figure 5: Comparing two-step and joint analysis for state-dependent diversification rates. We simulated a phylogeny with 200 tips under a binary-state-dependent speciation-and-extinction (BiSSE) model, assuming that there were two rates of speciation, λ_0 and λ_1, corresponding to two discrete states, 0 and 1, and that both extinction rates were zero (panel A). We assumed that the speciation rate in state 1 was four times higher than in state 0. We then estimated parameters of the BiSSE model under three scenarios: 1) the "true" scenario, where the true (simulated) tree was used as input for the BiSSE analysis; 2) the "conditional" (two-step) scenario, where we estimated the tree from sequence data and then used the estimated tree as input for the BiSSE analysis, and; 3) the "joint" scenario, where we estimated the parameters of the BiSSE model and the tree simultaneously. For scenarios 2 and 3, we simulated 5000 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 constant-rate Yule tree prior (for the conditional scenario) and a two-rate BiSSE tree prior (for the joint scenario), in both cases assuming the true substitution and relaxed-clock models. Finally, we fit the BiSSE model to the true tree and the MCC tree estimated under the constant-rate Yule model. B) We then compared posterior estimates of the ratio of speciation rates, λ_1/λ_0, estimated for the true scenario (red), the conditional scenario (purple), and the joint scenario (yellow). The true ratio $\lambda_1/\lambda_0 = 4$ (dashed line) is well estimated under the true scenario (red), but the conditional scenario substantially underestimates the ratio, excluding the true value from the 95% HPD (purple). The joint scenario (yellow) results in estimates much more similar to the true scenario, but with more uncertainty, presumably because this scenario incorporates uncertainty in divergence times.

lar 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 (see Box 2 for more on joint inference). Despite this critique, these approaches remained dominant for almost 15 years, until new joint inferences approaches to co-infer the evolutionary history of continuous and discrete characters were developed (May and Moore, 2020; Boyko et al., 2023).

Meanwhile, Felsenstein developed the threshold model (discussed above in Fundamental models of trait evolution: discrete traits) 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 (the "liability") is unobserved, both evolving via Brownian motion and linked via the covariance of the liabilities. Subsequent developments based on the threshold model framework include generalizing to multiple characters and multi-state characters (Cybis et al., 2015) and to enabling dimensionality reduction via a phylogenetic factor analysis to make analysis of large multivariate datasets 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 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-optima 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 et al. (2023) developed an OU-based approach that does jointly infer of the evolution of a discrete character and the evolution of a continuous character under a multi-optima 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 et al. (2023) built on earlier OU-methods that assigned the optima to pre-established dis-

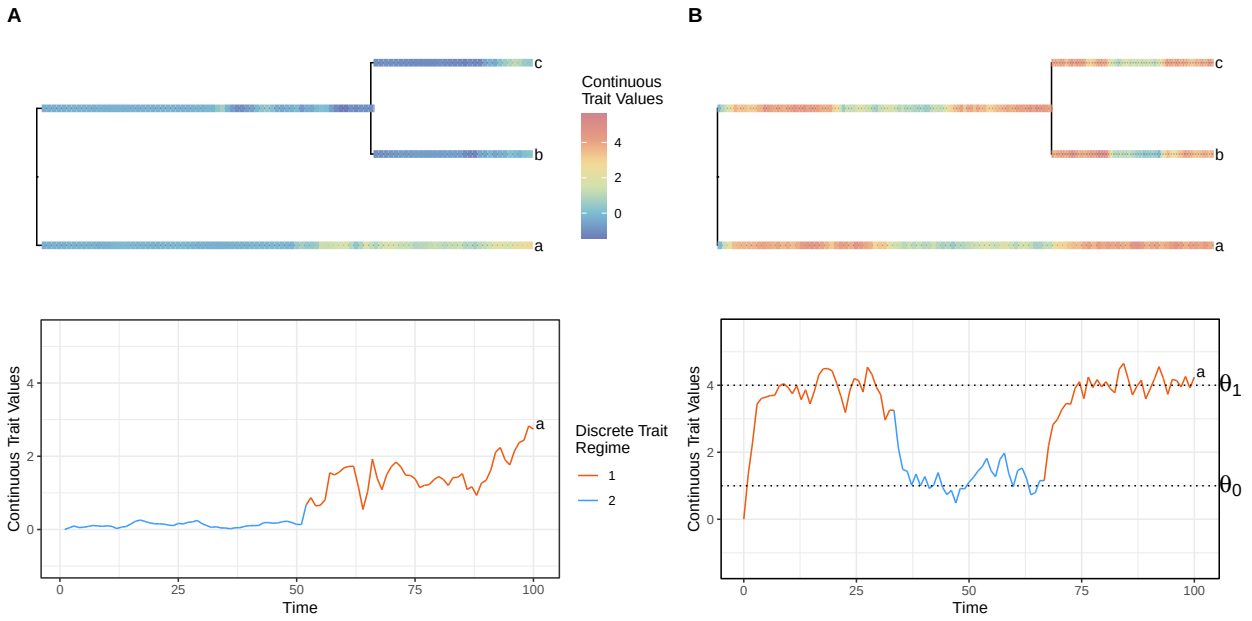

Figure 6: Simulation under models that jointly infer discrete and continuous trait evolution where the rates of the continuous trait are dependent on a discrete trait with states 1 and 2. (A) State-dependent BM; (B) State-dependent OU. Top row: stochastic character map with the continuous trait value plotted along the branches. Bottom row: sample path of the branch leading to species *a*. State-dependent regimes are colored. In (A), the discrete-character regimes differ in the values of the diffusion parameter; for Regime 1 $\sigma_0^2 = 0.001$; for Regime 2 $\sigma_1^2 = 0.1$. In (B), the discrete-character regimes differ in the values of the optimum; for Regime 1 $\theta_0 = 4$; for Regime 2 $\theta_1 = 1$.

crete trait histories on the phylogeny, requiring the evolution of continuous optima to perfectly match the history of the discrete trait (*e.g.*, Beaulieu et al., 2012). Instead, Boyko 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 model discrete and continuous trait evolution, May and Moore (2020) developed a method to test for the effect of a discrete character on the rate of evolution of a continuous character. For example, does the feeding apparatus of reef 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 Brownian motion process while accounting for process variation in the continuous trait(s) that is not linked to the discrete trait of interest. As with the OU-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 et al. (2023) and May and Moore (2020) fully 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 hypothesis involves correlated trait evolution or lineage diversification. For example, most researchers do not use SSE models to accommodate 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 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 hypothesis. 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. While 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 $a \leftrightarrow b$ transitions (q_h). Simultaneous transitions between $0 \leftrightarrow 1$ and $a \leftrightarrow b$ are set to 0.

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

While 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), accommodating process variation while jointly modeling correlated evolution between discrete and continuous variables (Boyko 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).

Most hidden-state model implementations, such as the one described in the above Q-matrix, use a single hidden binary trait. However, the decision to model a single hidden binary variable 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 evolutionary processes of trait evolution can be most appropriately modeled by a single binary 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. However, the classic hidden-states 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).

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) 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 vari-

ation 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, 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 datasets are increasingly available to inform estimation under complex models, and we anticipate and encourage advances in modeling process variation given these datasets.

Joint inference in divergence-time estimation

The classic 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, "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 dated phylogeny from a combined dataset 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 use continuous morphological traits as well (*e.g.*, Álvarez-Carretero et al., 2019; González-Ramírez et al., 2025).

TED approaches are generally discussed specifically in the 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 a fossilized-birth-death 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 (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 the pro-

cess 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 "early diverging" extant lineages *Danaea* and *Christensenia* 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 how the traits themselves have evolved (see also Lee et al., 2009). 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) and variation in rates of fossilization (*e.g.*, from the marattialeans again, where fossilization rates are inferred to be dramatically higher in the Carboniferous, coinciding with the period when "coal balls", the primary source of marattialean fossils, were formed; May and Rothfels, 2023).

Indeed, TED analyses represent the current pinnacle 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 phylogenetic comparative methods 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), while most PCM analyses fail to coestimate 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 datasets 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 a fossilized-birth-death 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 tree 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 "reconstruct" the evolution of the specific characters of choice. This two-step approach runs afoul of the issues described earlier (see Box 2): 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. While 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, we have three main options to infer ancestral states. The first approach is referred to as marginal estimation (Yang et al., 1995; Koshi and Goldstein, 1996). For each character (each site in the case of molecular data) and each internal node we can compute the probability of each possible state given the tree, model, and data. The state with the highest probability is typically considered the best point estimate of the ancestral state. Alternatively, we can summarize the probabilities of all states visually, *e.g.*, using pie charts with wedges proportional to the probability of each state. This approach is named "marginal" estimation because the state estimate at a given node averages over the states at all other nodes.

The second approach is referred to as joint estimation. Here we compute the probabilities of states across multiple nodes simultaneously; *e.g.*, what is the probability that ancestral node A is in state 0 *and* ancestral node B is in state 1? In principle, we can evaluate the probability of all possible combinations of ancestral states across nodes, but often researchers will restrict their consideration to the joint probabilities of states at a subset of nodes of particular biological interest. The joint ancestral state point estimate is then the set of states at these nodes that has the highest probability. Note that the joint estimate of the ancestral state for a particular node may differ from the marginal estimate, using the same data—the highest probability configuration of ancestral states when considering all nodes simultaneously may be different from the highest probability state for a node considered in isolation (Yang, 2006; Revell, 2025).

Third, we can take a stochastic mapping approach (see above; Huelsenbeck et al., 2003). Here we simulate many character histories under a chosen model (Q matrix) and track how often each node is in each state. A distinct advantage of this approach is that we obtain character histories along the branches rather than just at the nodes (Fig. 2). As the number of stochastic character maps approaches ∞, the frequencies of states at a particular node will converge to the marginal estimate, while the frequencies of combinations of states at sets of nodes will converge to the joint estimate; in other words, both types of ancestral-state estimates can be equivalent to estimates derived from stochastic character mapping (Revell, 2025).

All three approaches typically condition on a fixed Q matrix, inferred from the trait data. Importantly, this feature means that we are conditioning on both a pre-estimated phylogeny and Q matrix. A partial solution to this problem is fully hierarchical Bayesian approaches, which jointly sample parameters in the Q matrix and ancestral states (Huelsenbeck and Bollback, 2001); nonetheless, these approaches are still two-step, as charter evolution is studied on a known tree. One common approach to the mitigate the risk associated with conditioning on a single tree is to take into account tree uncertainty by fitting the character-evolution model on a distribution of trees (a reflection our tree uncertainty). In a Bayesian framework it it common to use a posterior distribution of trees as a prior distribution of trees for our model of character evolution. Here

the probability of a particular state at a given node is conditional on the frequency of the node in the posterior distribution of trees (Pagel et al., 2004). Put another way, the posterior probability of a node existing represents an upper limit on the probability of a state at the node; if a node is only present in 50% of the trees the probability of a state at that node can maximally be 50%. While this approach allows us to take into account uncertainty in our tree estimate it is not a fully joint inference, as inference of the tree(s) and patterns of trait evolution are separated. As a result, the posterior estimates of the character model are influenced by the posterior estimate of the tree model but not vice-versa. To jointly sample the character history and the tree we need to co-estimate the tree and character histories together (see Box 2; Fig. 7).

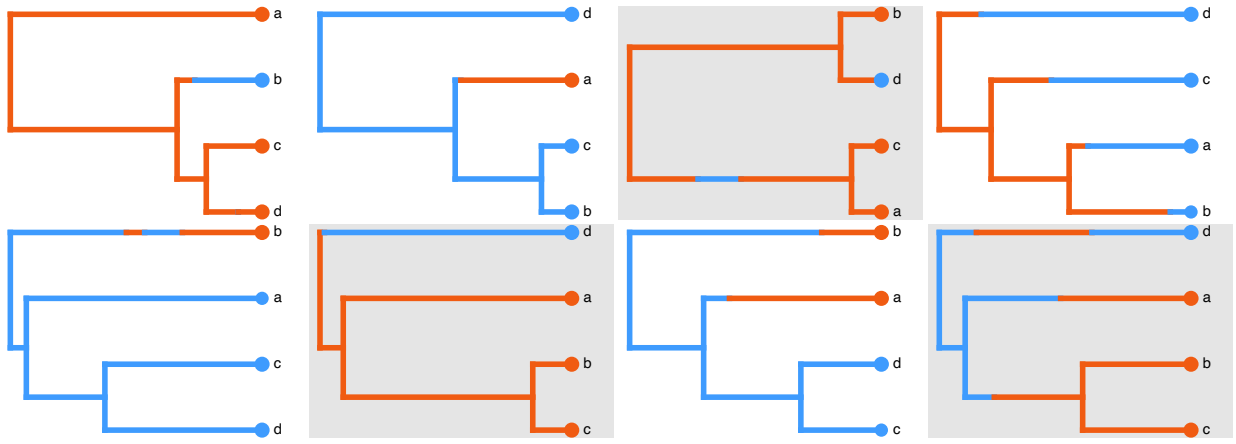

Figure 7: Joint simulation of a two-state character (the states are red and blue) and a four-tip phylogeny. The gray panels are simulations where the character histories match the observed tip states (tips a–b are orange; tip d is blue). Note that the clade (b,c) is present in 2/3 of the simulated histories that match the data (*i.e.*, its posterior probability is estimated to be 2/3).

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. 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 "data sources" 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 (towards 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 may be more appropriate to penalize them once, at their apomorphic origin (Löytynoja and Goldman, 2005). Finally, progressive alignment does not mechanistically

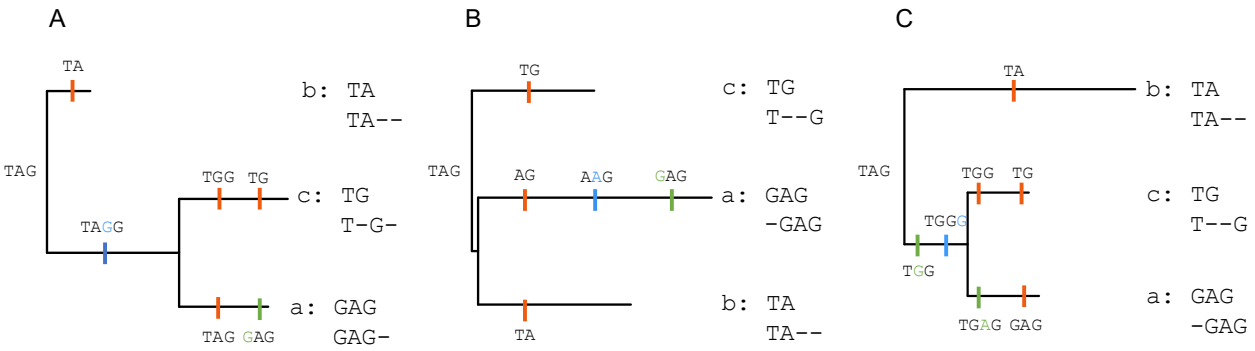

Figure 8: Sequences simulated with a joint model of insertion-deletion and nucleotide substitution. A–C) Simulations of representative nucleotides sequences under an model of indel and nucleotide substitution, conditioned on the root sequence (`TAG`) and observed sequences on different phylogenies. Tip sequences (top) are shown unaligned and (bottom) aligned given the history of sequence evolution. Tree topologies indicate events mapped along branches with tick marks: Blue ticks corresponds 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 coestimates the alignment with the tree.

model the processes that generated the data—it is not possible to derive inferences about, *e.g.*, 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. 8 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 (see ancestral state estimation section, above).

Conclusion

In this review, we argue that the systematics community has gradually moved away from studying character evolution via a classical two-step "phylogenetic comparative methods" approach where the focal character(s) of interest are analyzed on a fixed pre-estimated phylogeny, and towards 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. We further argue that joint inference is generally more conceptually sound and accurate than the canonical two-step approach, and that it opens up entirely new types of analyses that are not possible under the two-step PCM paradigm. Related to the movement towards joint inference is a more general movement towards inference (estimation) and away from more "Popperian" and hypothesis-testing related approaches, which we believe 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, 2020; 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, 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 an practical perspective, there are relatively few "pre-packaged" joint models available 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 datasets typically applied to tree inference, the exclusion of a small set of focal characters from that inference will mostly 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 promoted 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 so-called "universal probabilistic programming language", or PPL; Senderov et al., 2024) promise to increase the range of joint phylogenetic models even further. As access to these methods expands, joint inference will enable testing increasingly complex hypotheses of 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 for 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, the SSB Legacy Committee for inviting us to contribute this article, and XXX reviewers for their helpful comments. This material is based on work supported by the NSF Postdoctoral Research Fellowships in Biology Program under Grant No. 2209159 to JM-G.

References

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

Arcila, D., Pyron, R. A., Tyler, J. C., Ortí, G., and Betancur-R, R. (2015). An evaluation of fossil tip-dating versus node-age calibrations in tetraodontiform fishes (Teleostei: Percomorphaceae). *Molecular Phylogenetics and Evolution*, 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., and Suchard, M. A. (2021). Markov-modulated continuous-time Markov chains to identify site-and branch-specific evolutionary variation in BEAST. *Systematic Biology*, 70(1):181–189.

Baum, D. A. and Smith, S. D. (2013). Tree thinking. *An Introduction to Phylogenetic Biology. Roberts and Company Publishers*.

Beaulieu, J. M., Jhwueng, D.-C., Boettiger, C., and O'Meara, B. C. (2012). Modeling stabilizing selection: Expanding the Ornstein–Uhlenbeck model of adaptive evolution. *Evolution*, 66(8):2369–2383.

Beaulieu, J. M., O'Meara, B. C., and Donoghue, M. J. (2013). Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. *Systematic Biology*, 62(5):725–737.

Beaulieu, J. M. and O'Meara, B. C. (2016). Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. *Systematic Biology*, 65(4):583–601.

Blomberg, S. P., Garland Jr, T., and Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution*, 57(4):717–745.

Boucher, F. C. and Démery, V. (2016). Inferring bounded evolution in phenotypic characters from phylogenetic comparative data. *Systematic Biology*, 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., et al. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, 15(4):e1006650.

Boyko, J. D., O'Meara, B. C., and Beaulieu, J. M. (2023). A novel method for jointly modeling the evolution of discrete and continuous traits. *Evolution*, 77(3):836–851.

Brandley, M. C., Huelsenbeck, J. P., and 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. and McLennan, D. A. (1991). *Phylogeny, Ecology and Behaviour: A Research Program in Comparative Biology*. University of Chicago Press, Chicago.

Brown, J. M. and Lemmon, A. R. (2007). The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. *Systematic Biology*, 56(4):643–655.

Butler, M. A. and King, A. A. (2004). Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *The American Naturalist*, 164(6):683–695.

Caetano, D. S. and Harmon, L. J. (2019). Estimating correlated rates of trait evolution with uncertainty. *Systematic Biology*, 68(3):412–429.

Caetano, D. S., O'Meara, B. C., and Beaulieu, J. M. (2018). Hidden state models improve state-dependent diversification approaches, including biogeographical models. *Evolution*, 72(11):2308–2324.

Carpenter, J. M. (1992). Comparing methods. *Cladistics*, 8(2):191–195.

Chatzou, M., Magis, C., Chang, J.-M., Kemena, C., Bussotti, G., Erb, I., and Notredame, C. (2016). Multiple sequence alignment modeling: methods and applications. *Briefings in Bioinformatics*, 17(6):1009–1023.

Cobbett, A., Wilkinson, M., and Wills, M. A. (2007). Fossils impact as hard as living taxa in parsimony analyses of morphology. *Systematic Biology*, 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., and Near, T. J. (2009). Piscivory limits diversification of feeding morphology in centrarchid fishes. *Evolution*, 63(6):1557–1573.

Cornwallis, C. K. and Griffin, A. S. (2024). A guided tour of phylogenetic comparative methods for studying trait evolution. *Annual Review of Ecology, Evolution, and Systematics*, 55.

Cybis, G. B., Sinsheimer, J. S., Bedford, T., Mather, A. E., Lemey, P., and Suchard, M. A. (2015). Assessing phenotypic correlation through the multivariate phylogenetic latent liability model. *The Annals of Applied Statistics*, 9(2):969–991.

de Queiroz, A., Donoghue, M. J., and Kim, J. (1995). Separate versus combined analysis of phylogenetic evidence. *Annual Review of Ecology and Systematics*, pages 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. *The American Naturalist*, 148(4):700–708.

de Queiroz, K. (2014). Popperian corroboration and phylogenetics. *Systematic Biology*, 63(6):1018–1022.

de Queiroz, K. and Poe, S. (2001). Philosophy and phylogenetic inference: a comparison of likelihood and parsimony methods in the context of Karl Popper's writings on corroboration. *Systematic Biology*, 50(3):305–321.

Dos Reis, M. and Yang, Z. (2013). The unbearable uncertainty of Bayesian divergence time estimation. *Journal of Systematics and Evolution*, 51(1):30–43.

Faith, D. P. and Trueman, J. W. (2001). Towards an inclusive philosophy for phylogenetic inference. *Systematic Biology*, 50(3):331–350.

Farris, J. S., Kluge, A. G., and Carpenter, J. M. (2001). Popper and likelihood versus "Popper*". *Systematic Biology*, 50(3):438–444.

Felsenstein, J. (1973). Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. *Systematic Biology*, 22(3):240–249.

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

Felsenstein, J. (2003). *Inferring phylogenies*. Sinauer Associates.

Felsenstein, J. (2005). Using the quantitative genetic threshold model for inferences between and within species. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1459):1427–1434.

Fitch, W. M. and 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. *Biochemical Genetics*, 4(5):579–593.

FitzJohn, R. G. (2010). Quantitative traits and diversification. *Systematic Biology*, 59(6):619–633.

FitzJohn, R. G. (2012). Diversitree: comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution*, 3(6):1084–1092.

Freyman, W. A. and Höhna, S. (2018). Cladogenetic and anagenetic models of chromosome number evolution: a Bayesian model averaging approach. *Systematic Biology*, 67(2):195–215.

Freyman, W. A., Johnson, M. G., and Rothfels, C. J. (2023). homologizer: Phylogenetic phasing of gene copies into polyploid subgenomes. *Methods in Ecology and Evolution*, 14(5):1230–1244.

Frumhoff, P. C. and Reeve, H. K. (1994). Using phylogenies to test hypotheses of adaptation: a critique of some current proposals. *Evolution*, pages 172–180.

Gaya, E., Redelings, B. D., Navarro-Rosinés, P., Llimona, X., De Cáceres, M., and 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. and Foo, J. (2020). Memory in trait macroevolution. *The American Naturalist*, 195(2):300–314.

Goldberg, E. E. and Igić, B. (2008). On phylogenetic tests of irreversible evolution. *Evolution*, 62(11):2727–2741.

Goldberg, E. E. and Igić, B. (2012). Tempo and mode in plant breeding system evolution. *Evolution*, 66(12):3701–3709.

Goldberg, E. E., Lancaster, L. T., and Ree, R. H. (2011). Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. *Systematic Biology*, 60(4):451–465.

Goldman, N., Anderson, J. P., and Rodrigo, A. G. (2000). Likelihood-based tests of topologies in phylogenetics. *Systematic Biology*, 49(4):652–670.

González-Ramírez, I. S., Deanna, R., and Smith, S. D. (2025). Late Cretaceous origins for major nightshade lineages from total evidence timetree analysis. *bioRxiv*, pages 2025–07.

Hand, S. J., Maugoust, J., Beck, R. M., and Orliac, M. J. (2023). A 50-million-year-old, three-dimensionally preserved bat skull supports an early origin for modern echolocation. *Current Biology*, 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*. 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., et al. (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., and Stadler, T. (2014). The fossilized birth–death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences*, 111(29):E2957–E2966.

Helfenbein, K. G. and DeSalle, R. (2005). Falsifications and corroborations: Karl Popper's influence on systematics. *Molecular Phylogenetics and Evolution*, 35(1):271–280.

Höhna, S., Landis, M. J., Heath, T. A., Boussau, B., Lartillot, N., Moore, B. R., Huelsenbeck, J. P., and Ronquist, F. (2016). RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic Biology*, 65(4):726–736.

Holmes, I. and Bruno, W. J. (2001). Evolutionary HMMs: a Bayesian approach to multiple alignment. *Bioinformatics*, 17(9):803–820.

Huelsenbeck, J. P. and Bollback, J. P. (2001). Empirical and hierarchical Bayesian estimation of ancestral states. *Systematic Biology*, 50(3):351–366.

Huelsenbeck, J. P., Larget, B., Miller, R. E., and Ronquist, F. (2002). Potential applications and pitfalls of Bayesian inference of phylogeny. *Systematic Biology*, 51(5):673–688.

Huelsenbeck, J. P., Nielsen, R., and Bollback, J. P. (2003). Stochastic mapping of morphological characters. *Systematic Biology*, 52(2):131–158.

Huey, R. B., Garland Jr, T., and Turelli, M. (2019). Revisiting a key innovation in evolutionary biology: Felsenstein's "phylogenies and the comparative method". *The American Naturalist*, 193(6):755–772.

Jukes, T. (1969). Evolution of protein molecules. *Mammalian Protein Metabolism*, 3.

Kaplan, N. and Langley, C. H. (1979). A new estimate of sequence divergence of mitochondrial DNA using restriction endonuclease mappings. *Journal of Molecular Evolution*, 13:295–304.

Kapli, P., Yang, Z., and Telford, M. J. (2020). Phylogenetic tree building in the genomic age. *Nature Reviews Genetics*, 21(7):428–444.

Katoh, K., Asimenos, G., and Toh, H. (2009). Multiple alignment of DNA sequences with MAFFT. *Bioinformatics for DNA Sequence Analysis*, pages 39–64.

Kemena, C. and 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., and Ané, C. (2016). Fast and accurate detection of evolutionary shifts in Ornstein–Uhlenbeck models. *Methods in Ecology and Evolution*, 7(7):811–824.

Koshi, J. M. and Goldstein, R. A. (1996). Probabilistic reconstruction of ancestral protein sequences. *Journal of Molecular Evolution*, 42:313–320.

Landis, M. J., Schraiber, J. G., and Liang, M. (2013). Phylogenetic analysis using Lévy processes: finding jumps in the evolution of continuous traits. *Systematic Biology*, 62(2):193–204.

Lee, M., Oliver, P., and Hutchinson, M. (2009). Phylogenetic uncertainty and molecular clock calibrations: a case study of legless lizards (Pygopodidae, Gekkota). *Molecular Phylogenetics and Evolution*, 50(3):661–666.

Lewis, P. O. (2001). A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology*, 50(6):913–925.

Losos, J. B. (2011). Seeing the forest for the trees: The limitations of phylogenies in comparative biology: (American Society of Naturalists address). *The American Naturalist*, 177(6):709–727.

Louca, S. and Pennell, M. W. (2020). Extant timetrees are consistent with a myriad of diversification histories. *Nature*, 580(7804):502–505.

Löytynoja, A. and Goldman, N. (2005). An algorithm for progressive multiple alignment of sequences with insertions. *Proceedings of the National Academy of Sciences*, 102(30):10557–10562.

Luckow, M. and Bruneau, A. (1997). Circularity and independence in phylogenetic tests of ecological hypotheses. *Cladistics*, 13(1-2):145–151.

Maddison, W. P. (1993). Missing data versus missing characters in phylogenetic analysis. *Systematic Biology*, 42(4):576–581.

Maddison, W. P. (2006). Confounding asymmetries in evolutionary diversification and character change. *Evolution*, 60(8):1743–1746.

Maddison, W. P. and FitzJohn, R. G. (2015). The unsolved challenge to phylogenetic correlation tests for categorical characters. *Systematic Biology*, 64(1):127–136.

Maddison, W. P., Midford, P. E., and Otto, S. P. (2007). Estimating a binary character's effect on speciation and extinction. *Systematic Biology*, 56(5):701–710.

Magnuson-Ford, K. and Otto, S. P. (2012). Linking the investigations of character evolution and species diversification. *The American Naturalist*, 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., and Irimia, M. (2024). Evolution of tissue-specific expression of ancestral genes across vertebrates and insects. *Nature Ecology & Evolution*, 8(6):1140–1153.

Martins, E. P. and Hansen, T. F. (1997). Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *The American Naturalist*, 149(4):646–667.

May, M. R., Contreras, D. L., Sundue, M. A., Nagalingum, N. S., Looy, C. V., and Rothfels, C. J. (2021). Inferring the total-evidence timescale of marattialean fern evolution in the face of model sensitivity. *Systematic Biology*, 70(6):1232–1255.

May, M. R. and 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. *Systematic Biology*, 69(3):530–544.

May, M. R. and Rothfels, C. J. (2023). Diversification models conflate likelihood and prior, and cannot be compared using conventional model-comparison tools. *Systematic Biology*, 72(3):713–722.

Mayrose, I., Barker, M. S., and Otto, S. P. (2010). Probabilistic models of chromosome number evolution and the inference of polyploidy. *Systematic Biology*, 59(2):132–144.

Mayrose, I. and Lysak, M. A. (2021). The evolution of chromosome numbers: mechanistic models and experimental approaches. *Genome Biology and Evolution*, 13(2):evaa220.

Mayrose, I., Zhan, S. H., Rothfels, C. J., Magnuson-Ford, K., Barker, M. S., Rieseberg, L. H., and Otto, S. P. (2011). Recently formed polyploid plants diversify at lower rates. *Science*, 333(6047):1257–1257.

Mika, K., Whittington, C. M., McAllan, B. M., and Lynch, V. J. (2022). Gene expression phylogenies and ancestral transcriptome reconstruction resolves major transitions in the origins of pregnancy. *eLife*, 11:e74297.

Miklós, I., Lunter, G. A., and Holmes, I. (2004). A "long indel" model for evolutionary sequence alignment. *Molecular Biology and Evolution*, 21(3):529–540.

Mongiardino Koch, N., Garwood, R. J., and Parry, L. A. (2021). Fossils improve phylogenetic analyses of morphological characters. *Proceedings of the Royal Society B*, 288(1950):20210044.

Mongiardino Koch, N. and Parry, L. A. (2020). Death is on our side: paleontological data drastically modify phylogenetic hypotheses. *Systematic Biology*, 69(6):1052–1067.

Morlon, H., Parsons, T. L., and Plotkin, J. B. (2011). Reconciling molecular phylogenies with the fossil record. *Proceedings of the National Academy of Sciences*, 108(39):16327–16332.

Nee, S., May, R. M., and Harvey, P. H. (1994). The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 344(1309):305–311.

Nylander, J. A., Ronquist, F., Huelsenbeck, J. P., and Nieves-Aldrey, J. (2004). Bayesian phylogenetic analysis of combined data. *Systematic Biology*, 53(1):47–67.

Olmstead, R. (2001). Phylogenetic inference and the writings of karl popper. *Systematic Biology*, 50(3):304.

O'Meara, B. C. (2012). Evolutionary inferences from phylogenies: a review of methods. *Annual Review of Ecology, Evolution, and Systematics*, 43(1):267–285.

O'Meara, B. C., Ané, C., Sanderson, M. J., and Wainwright, P. C. (2006). Testing for different rates of continuous trait evolution using likelihood. *Evolution*, 60(5):922–933.

Pagel, M. (1994). Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 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. *Systematic Biology*, 48(3):612–622.

Pagel, M., Meade, A., and Barker, D. (2004). Bayesian estimation of ancestral character states on phylogenies. *Systematic Biology*, 53(5):673–684.

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., et al. (2021). Rampant tooth loss across 200 million years of frog evolution. *eLife*, 10:e66926.

Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20(2):289–290.

Popper, K. (1959). *The Logic of Scientific Discovery*. Julius Springer, Hutchinson & Co.

Porto, D. S., Uyeda, J., Mikó, I., and Tarasov, S. (2024). ontophylo: Reconstructing the evolutionary dynamics of phenomes using new ontology-informed phylogenetic methods. *Methods in Ecology and Evolution*, 15(2):290–300. _eprint: https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.14283.

Quental, T. B. and Marshall, C. R. (2010). Diversity dynamics: molecular phylogenies need the fossil record. *Trends in Ecology & Evolution*, 25(8):434–441.

Rabosky, D. L. and Goldberg, E. E. (2015). Model inadequacy and mistaken inferences of trait-dependent speciation. *Systematic Biology*, 64(2):340–355.

Rannala, B. and Yang, Z. (1996). Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *Journal of Molecular Evolution*, 43:304–311.

Redelings, B. D., Holmes, I., Lunter, G., Pupko, T., and Anisimova, M. (2024). Insertions and deletions: Computational methods, evolutionary dynamics, and biological applications. *Molecular Biology and Evolution*, 41(9):msae177.

Redelings, B. D. and Suchard, M. A. (2005). Joint bayesian estimation of alignment and phylogeny. *Systematic Biology*, 54(3):401–418.

Revell, L. J. (2013). A comment on the use of stochastic character maps to estimate evolutionary rate variation in a continuously valued trait. *Systematic Biology*, 62(2):339–345.

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. *Evolutionary Biology*, pages 1–25.

Revell, L. J. and Harmon, L. J. (2022). *Phylogenetic comparative methods in R*. Princeton University Press.

Revell, L. J., Harmon, L. J., and Collar, D. C. (2008). Phylogenetic signal, evolutionary process, and rate. *Systematic Biology*, 57(4):591–601.

Rieppel, O. (2003). Popper and systematics. *Systematic Biology*, 52(2):259–271.

Ronquist, F. (2004). Bayesian inference of character evolution. *Trends in Ecology & Evolution*, 19(9):475–481.

Ronquist, F., Klopfstein, S., Vilhelmsen, L., Schulmeister, S., Murray, D. L., and Rasnitsyn, A. P. (2012). A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. *Systematic Biology*, 61(6):973–999.

Rothfels, C. J. and Schuettpelz, E. (2014). Accelerated rate of molecular evolution for vittarioid ferns is strong and not driven by selection. *Systematic Biology*, 63(1):31–54.

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., et al. (2012). Testing the impact of calibration on molecular divergence times using a fossil-rich group: the case of *Nothofagus* (Fagales). *Systematic Biology*, 61(2):289–313.

Schluter, D., Price, T., Mooers, A. Ø., and 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., and Ronquist, F. (2024). Treeppl: A universal probabilistic programming language for phylogenetics. *bioRxiv*.

Siddall, M. E. and Kluge, A. G. (1997). Probabilism and phylogenetic inference. *Cladistics*, 13(4):313–336.

Sober, E. (1991). *Reconstructing the past: Parsimony, evolution, and inference*. 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., et al. (2023). Evolution of vertebral numbers in primates, with a focus on hominoids and the last common ancestor of hominins and panins. *Journal of Human Evolution*, 179:103359.

Swiston, S. K. and Landis, M. J. (2025). Testing relationships between multiple regional features and biogeographic processes of speciation, extinction, and dispersal. *Systematic Biology*, 74(2):282–300.

Symonds, M. R. E. and 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*, pages 105–130. Springer, Berlin, Heidelberg.

Tarasov, S. (2019). Integration of anatomy ontologies and evo-devo using structured Markov models suggests a new framework for modeling discrete phenotypic traits. *Systematic Biology*, 68(5):698–716.

Tarasov, S. (2023). New phylogenetic Markov models for inapplicable morphological characters. *Systematic Biology*, 72(3):681–693.

Tarasov, S., Mikó, I., Yoder, M. J., and Uyeda, J. C. (2019). PARAMO: A pipeline for reconstructing ancestral anatomies using ontologies and stochastic mapping. *Insect Systematics and Diversity*, 3(6):1.

Thomas, G. H., Freckleton, R. P., and Székely, T. (2006). Comparative analyses of the influence of developmental mode on phenotypic diversification rates in shorebirds. *Proceedings of the Royal Society B: Biological Sciences*, 273(1594):1619–1624.

Thompson, A., May, M. R., Hopkins, B., Riedl, N., Barmina, O., Liebeskind, B. J., Zhao, L., Begun, D., and Kopp, A. (2024). Quantifying transcriptome turnover on phylogenies by modeling gene expression as a binary trait. *bioRxiv*, page 2024.10.03.616564.

Thorne, J. and Kishino, H. (1992). Freeing phylogenies from artifacts of alignment. *Molecular Biology and Evolution*, 9(6):1148–1162.

Thorne, J. L., Kishino, H., and Felsenstein, J. (1991). An evolutionary model for maximum likelihood alignment of DNA sequences. *Journal of Molecular Evolution*, 33:114–124.

Tolkoff, M. R., Alfaro, M. E., Baele, G., Lemey, P., and Suchard, M. A. (2018). Phylogenetic factor analysis. *Systematic Biology*, 67(3):384–399.

Tribble, C. M., Márquez-Corro, J. I., May, M. R., Hipp, A. L., Escudero, M., and Zenil-Ferguson, R. (2025). Macroevolutionary inference of complex modes of chromosomal speciation in a cosmopolitan plant lineage. *New Phytologist*, 245(5):2350–2361.

Tribble, C. M., May, M. R., Jackson-Gain, A., Zenil-Ferguson, R., Specht, C. D., and Rothfels, C. J. (2023). Unearthing modes of climatic adaptation in underground storage organs across Liliales. *Systematic Biology*, 72(1):198–212.

Tuffley, C. and Steel, M. (1998). Modeling the covarion hypothesis of nucleotide substitution. *Mathematical Biosciences*, 147(1):63–91.

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

Uyeda, J. C., Zenil-Ferguson, R., and Pennell, M. W. (2018). Rethinking phylogenetic comparative methods. *Systematic Biology*, 67(6):1091–1109.

Vialle, R. A., Tamuri, A. U., and Goldman, N. (2018). Alignment modulates ancestral sequence reconstruction accuracy. *Molecular Biology and Evolution*, 35(7):1783–1797.

Vogt, L. (2008). The unfalsifiability of cladograms and its consequences. *Cladistics*, 24(1):62–73.

Warnock, R. C. and Wright, A. M. (2020). *Understanding the tripartite approach to Bayesian divergence time estimation*. Cambridge University Press.

Westesson, O., Lunter, G., Paten, B., and Holmes, I. (2012). Accurate reconstruction of insertion-deletion histories by statistical phylogenetics. *PLoS One*, 7(4):e34572.

Wong, K. M., Suchard, M. A., and 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.

Yang, Z. (2006). *Computational Molecular Evolution*. OUP Oxford.

Yang, Z., Kumar, S., and Nei, M. (1995). A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics*, 141(4):1641–1650.

Yang, Z. and Rannala, B. (1997). Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo method. *Molecular Biology and Evolution*, 14(7):717–724.

Zenil-Ferguson, R., Burleigh, J. G., Freyman, W. A., Igić, B., Mayrose, I., and Goldberg, E. E. (2019). Interaction among ploidy, breeding system and lineage diversification. *New Phytologist*, 224(3):1252–1265.

Zenil-Ferguson, R., Ponciano, J. M., and Burleigh, J. G. (2017). Testing the association of phenotypes with polyploidy: an example using herbaceous and woody eudicots. *Evolution*, 71(5):1138–1148.

Zharkikh, A. (1994). Estimation of evolutionary distances between nucleotide sequences. *Journal of Molecular Evolution*, 39(3):315–329.