Of traits and trees: probabilistic distances under continuous trait models for dissecting the interplay among phylogeny, model, and data

Richard H. Adams[1,*], Heath Blackmon[2], and Michael DeGiorgio[1,*]

[1]Department of Computer and Electrical Engineering and Computer Science, Florida Atlantic University, Boca Raton, FL 33431, USA

[2]Department of Biology, Texas A&M University, College Station, TX 77843, USA

[*]**Correspondence**: adamsr@fau.edu (R.H.A.), mdegiorg@fau.edu (M.D.)

Running Head: Probabilistic phylogenetic distances under continuous trait models

Abstract

Stochastic models of character trait evolution have become a cornerstone of evolutionary biology in an array of contexts. While probabilistic models have been used extensively for statistical inference, they have largely been ignored for the purpose of measuring distances between phylogeny-aware models. Recent contributions to the problem of phylogenetic distance computation have highlighted the importance of explicitly considering evolutionary model parameters and their impacts on molecular sequence data when quantifying dissimilarity between trees. By comparing two phylogenies in terms of their induced probability distributions that are functions of many model parameters, these distances can be more informative than traditional approaches that rely strictly on differences in topology or branch lengths alone. Currently, however, these approaches are designed for comparing models of nucleotide substitution and gene tree distributions, and thus, are unable to address other classes of traits and associated models that may be of interest to evolutionary biologists. Here we expand the principles of probabilistic phylogenetic distances to compute tree distances under models of continuous trait evolution along a phylogeny. By explicitly considering both the degree of relatedness among species and the evolutionary processes that collectively give rise to character traits, these distances provide a foundation for comparing models and their predictions, and for quantifying the impacts of assuming one phylogenetic background over another while studying the evolution of a particular trait. We demonstrate the properties of these approaches using theory, simulations, and several empirical datasets that highlight potential uses of probabilistic distances in many scenarios. We also introduce an open-source R package named PRDATR for easy application by the scientific community for computing phylogenetic distances under models of character trait evolution.

INTRODUCTION

Probabilistic models of character trait evolution have become invaluable tools across many fields of evolutionary biology. Indeed, stochastic evolutionary models are the heart of comparative methods (e.g., Felsenstein 1985), and an incredibly diverse body of literature now exists that includes numerous applications of such models for phylogenetic reconstruction (e.g, Liò and Goldman 1998), ancestral state reconstruction (e.g., Schluter et al. 1997), and evolutionary rate estimation (e.g., Martins 1994), as well as for studies of coevolution (e.g., Ronquist 1997), adaptation (e.g., Revell et al. 2010), lineage diversification (e.g., O'Meara and Beaulieu 2016), and correlated trait evolution (e.g., Bawa et al. 2018). At a fundamental level, these models are designed to parameterize the probability distributions of character traits conditioned upon a particular phylogenetic tree and set of evolutionary parameters, which themselves are designed to capture pertinent processes that influence traits over time. Thus, probabilistic models of trait evolution provide a vehicle for interpreting biodiversity in light of both the processes and the phylogenetic history of organisms that collectively shape biological variation observed in nature.

Given such widespread adoption of probabilistic models for studying evolution, it is somewhat surprising that these same models have been relatively ignored for the purpose of measuring distances between trees conditioning on such phylogeny-aware models. Tree comparisons are a routine yet essential part of phylogenetic analysis that can be useful for elucidating methodological shortcomings and statistical biases in tree reconstruction methods (e.g., Reddy et al. 2017), as well as for more general study of macroevolutionary (e.g., Watanabe and Slice 2014) and microevolutionary processes (e.g., Yahara et al. 2014). There is now a wealth of frameworks for computing tree distances, including the Robinson-Foulds metric (Robinson and Foulds 1979), the Billera-Holmes-Vogtmann (BHV) or geodesic metric (Billera et al. 2001), and the path-length-difference metric (Penny et al. 1993), among others (Estabrook et al. 1985; Lin et al. 2012; Kuhner and Yamato 2015; Colijn and Plazzotta 2018).

Though widely employed throughout the literature, these more traditional approaches are primarily concerned with measuring differences in the branching structure (i.e., topology) and/or branch lengths of trees, and they do not explicitly consider any particular evolutionary process that may act on genotypic or phenotypic variation.

From a modeling perspective, however, phylogenies are more than just a topology and set of branch lengths: they define the degree of covariation in character traits expected among lineages, and thus, provide a fundamental framework for studying trait evolutionary processes, which has broad relevance for many fields and applications (e.g., O'Meara 2012; Nunn 2013; Pennell and Harmon 2013). Coupled with a model of evolution, trees can therefore be identified as points on a space of distributions over characters traits, which have been referred to as "phylogenetic oranges" (Kim 2020; Moulton and Steel 2004). While likelihood-based model selection is often conducted before or alongside parameter estimation for both continuous (e.g., Eastman et al. 2011; Uyeda and Harmon, 2014) and discrete traits (e.g., Huelsenbeck et al. 2004; Drummond and Suchard 2010), post-inference comparison of fitted trees is typically conducted without reference to the models themselves using the Robinson-Foulds or BHV metric, for example. Importantly, most of these classical measures of phylogenetic distance ignore this information, such that new approaches that more effectively incorporate aspects of the evolutionary process alongside knowledge of organismal relationships hold promise for conducting comparisons of trees at finer resolutions.

Recently, two probabilistic frameworks have been proposed for computing phylogenetic distances—one for comparing trees in terms of their underlying probability distributions over nucleotide site patterns for genetic sequence data (Garba et al. 2018), and another for quantifying distances between gene tree distributions under the multispecies coalescent model (Adams and Castoe 2019b). From this model-based perspective, the distance between two trees is measured as the distance between their induced probability distributions, which are functions of all relevant parameters specified

in the evolutionary models (e.g., substitution rates, base equilibrium frequencies, and random mating) in addition to properties of the gene tree or species tree (i.e., topology and branch lengths). For example, the distances of Garba et al. (2018) assume generalized time-reversible (GTR; Tavaré 1986) substitution models to compare two trees in terms of their underlying probability distributions, and thus, these distances can detect underlying differences in substitution parameters, topology, and/or branch lengths that influence nucleotide site pattern probabilities. Similarly, the species tree distances proposed by Adams and Castoe (2019b) employ coalescent theory to measure the distance between two multispecies coalescent models (i.e., two species trees) in terms of their gene tree probability distributions, which are influenced by demographic parameters such as effective population sizes, divergence times, and species topologies. A key advantage of using a probabilistic approach to tree distance is that it provides a natural means for assessing model identifiability, which is required for inference to be possible. Two models that induce identical probability distributions will yield a corresponding distance of zero, such that even an infinite amount of data will be unable to distinguish between the two (Zhu and Degnan 2017). Recently, probabilistic distances have also been leveraged to define new spaces to model phylogenies (Garba et al. 2020). Collectively, these new approaches represent a targeted effort to more effectively leverage a longstanding model-based perspective that has been used extensively for decades to both study evolutionary process and *estimate* trees, but not necessarily to *compare* them with one another in terms of their probability distributions over traits.

A critical limitation of these newly-proposed, probabilistic-based distance approaches is that, in their current form, they are not readily applicable to the many other types of traits and models that may be important in evolutionary studies. In particular, these probabilistic-based distances of phylogenetic trees do not currently consider continuous traits and associated models. However, these previous approaches do suggest a promising opportunity for comparing phylogeny-aware models of continuous trait evolution in a similar manner. In this study, we expand the framework of probabilistic

phylogenetic distances to incorporate these traits and associated models in an effort to provide more meaningful measures of tree distance. These distances seek to compare two phylogeny-aware evolutionary models in terms of their underlying probability distributions over continuous character traits, rather than on their topology and/or branch lengths alone. We demonstrate these probabilistic phylogenetic distances using theory, simulations, and empirical analyses, which collectively highlight the value of this approach for investigating continuous trait models and their predictions under an array of conditions. We examine the application of these measures across a range of diverse phylogenetic frameworks with different tree topologies and sizes (i.e., numbers of taxa), which provide insight into both their theoretical properties and empirical applications.

METHODS

Background: Probabilistic Phylogenetic Distances and Continuous Trait Models of Evolution

Though originally described in the context of nucleotide sequence data and models, it is relatively straightforward to extend the framework of Garba et al. (2018) for the purpose of measuring distances between other types of traits and models. We first note that the distance equations of Garba et al. (2018) can be used with only minimal modification for the purpose of comparing discrete trait models because they are based on standard four-state models of nucleotide substitution. These principles can therefore be adopted for computing distances under similar k-state ($k \in \{2,3,...\}$) discrete trait models by incorporating an appropriate k-state Markov model, such as the Cavender-Farris-Neyman (Neyman 1971; Farris 1973; Cavender 1978) for binary traits, the Mk model (Pagel 1994; Lewis 2001) for $k = 2$ or more models, or even 20-state models of amino acid substitution (e.g., Dayhoff et al. 1978). However, these distances are based on assumptions underlying discrete trait evolution, and therefore cannot be applied for continuous traits in their current form.

We therefore primarily focus on deriving and applying probabilistic phylogenetic distances under models of continuous trait evolution that are not currently considered by these approaches. A multitude of different models have been proposed for studying the evolution of continuous traits along a phylogeny, with many of them developed as extensions to the familiar Brownian motion (BM) model (Cavalli-Sforza and Edwards 1967; Felsenstein 1973), which includes an evolutionary rate parameter σ^2 measuring the rate of character trait change through time and the mean character trait value θ that typically represents the ancestral state of the root node (in Felsenstein 1973, the likelihood is computed using contrasts, such that the root state is not used). BM describes the process of continuous trait change occurring along branches of a phylogeny, with differences in trait values being drawn from a normal distribution with a mean equal to the ancestral state and variance proportional to σ^2 and time.

To demonstrate the properties of probabilistic distances under continuous trait models, we primarily focus on six models that are commonly used in evolutionary studies—though we note that many related models and combinations or variations of models can likely be used to compute distances in a similar manner. The six focal models considered in this study include the standard constant-rate BM model, the stationary-peak, or single optimum, Ornstein-Uhlenbeck model (OU; Lande 1976; Hansen 1997), and the early-burst model (EB; Harmon et al. 2010; Blomberg et al. 2003), as well as Pagel's lambda (L), delta (D), and kappa (K) models (Pagel 1999a,b). These models differ in their numbers and types of parameters, which are designed to capture the effects of particular evolutionary processes (Table 1). For example, the EB model incorporates a parameter a that determines whether the evolutionary rate $\sigma^2 = \sigma_0^2 e^{at}$ increases ($a > 0$) or decreases ($a < 0$) exponentially through time t from an initial value σ_0^2, whereas the OU model includes a parameter α that is proportional to the strength of attraction toward an optimum trait value θ. The OU model can also be described in terms of the "phylogenetic half-life" $t_{1/2} = \ln(2/\alpha)$, which measures the mean length of time required for the trait value to move halfway toward the optimum (Hansen 1997). Therefore, the trait value moves

toward its optimum faster for larger values of α. It is also worth noting that, unlike the BM model, the OU model has a stationary mean. For the L model, the $\lambda \in [0,1]$ parameter transforms the tree to become more star-like when λ is close to zero (i.e., species are statistically independent) and branch lengths are unaltered when $\lambda = 1$. The K model represents a punctuated model of trait change that raises all branch lengths in the phylogeny to the power of $\kappa \geq 0$. Evolution is directly proportional to branch lengths when $\kappa = 1$, while evolution is independent of branch lengths when $\kappa = 0$ (i.e., branches have the same length of 1.0). When $\kappa > 1$, traits evolve proportionally faster on longer branches compared to shorter branches, and conversely, evolution occurs proportionally slower for longer branches when $\kappa \in (0,1)$. The parameter $\delta \geq 0$ of the D model is designed to capture rate variation through time by raising all node heights in a tree to the power of δ. When $\delta > 1$, evolution has been fast in the recent past, and conversely, recent evolution has slowed down when $\delta \in (0,1)$. Finally, all branch lengths collapse to zero when $\delta = 0$.

Deriving Probabilistic Phylogenetic Distances Under Macroevolutionary Models of Continuous Trait Evolution

Parameters in these models (i.e., BM, OU, EB, L, K, and D) directly influence the probability distribution of traits, and therefore, we wish to incorporate this information when quantifying distances between models. For example, consider a phylogenetic model of continuous character evolution $\varphi^{\mathrm{BM}} = \{T, \upsilon, \sigma^2, \vec{\mu}\}$, which implements a BM model according to an n species tree with topology T and set of branch lengths υ, an evolutionary rate parameter σ^2, and a n-length vector containing the mean trait value for each tip $\vec{\mu} = (\theta, \theta, ..., \theta)$ (i.e., the expected trait value is the same as the ancestral state for each tip). A continuous trait χ that evolves according to this model will yield a vector $\vec{\chi}$ of length n containing the trait values observed for each of the n species, and we state that the distribution of $\vec{\chi}$

follows model φ^{BM} (i.e., $\vec{\chi} \sim \varphi^{\text{BM}}$), such that the probability density function of $\vec{\chi}$ given φ^{BM} is P $(\vec{\chi} \mid \varphi^{\text{BM}})$.

Due to the hierarchically structured nature of phylogenetic trees, $\vec{\chi}$ is distributed as multivariate normal (MVN) specified according to the parameters of φ^{BM} (e.g., Rohlf 2001; Revell and Harmon 2008). The topology T and branch lengths υ define the variance-covariance matrix, which is scaled by the rate parameter σ^2, and the ancestral state θ provides the expected trait value for each tip. We can therefore denote the probability distribution of $\vec{\chi}$ as $\vec{\chi} \sim N(\vec{\mu}, \Sigma)$, where $\vec{\mu}$ is an n-dimensional vector containing the mean trait value (i.e., ancestral state) at each tip, and Σ is the $n \times n$-dimensional phylogenetic variance-covariance matrix, which is a function of T, υ, and σ^2, and defines the covariance of trait values within and between species. Thus, it is straightforward to translate the probability distribution of $\vec{\chi}$ under any of the six focal models examined in this study (φ^{BM}, φ^{OU}, φ^{EB}, φ^{L}, φ^{K}, and φ^{D}) by rescaling Σ according to the model parameters. For example, we could derive the probability distribution of $\vec{\chi}$ under a single optimum OU-based phylogenetic model $\varphi^{\text{OU}} = \{T, \upsilon, \sigma^2, \theta, \alpha\}$ by transforming the variance-covariance matrix Σ according to both σ^2 and α.

In practice, we would like to be able to compare two evolutionary models, such as $\varphi_1^{\text{BM}} = \{T_1, \upsilon_1, \sigma_1^2, \vec{\mu}_1\}$ and $\varphi_2^{\text{OU}} = \{T_2, \upsilon_2, \sigma_2^2, \theta_2, \alpha_2\}$, or alternatively, a BM-based model φ_1^{BM} and an EB-based model φ_2^{EB}, or another pair of models. Thus, we are interested in measuring distances between two models φ_1 and φ_2 in terms of their probability distributions over $\vec{\chi}$, rather than between only their topologies (i.e., T_1 versus T_2) and/or branch lengths (i.e., υ_1 versus υ_2). The probabilistic distance between two phylogenetic models of continuous trait evolution can be denoted as:

$$d(\varphi_1, \varphi_2) = d(N(\vec{\mu}_1, \Sigma_1), N(\vec{\mu}_2, \Sigma_2)) \tag{1}$$

where $\vec{\mu}_i$ are the mean trait vectors and Σ_i are the transformed variance-covariance matrices that have been rescaled according to the evolutionary parameters of the two models φ_i, $i \in \{1,2\}$. For example, the Σ_i for a given model φ_i can be obtained using the *vcv* function provided in the R package GEIGER

(Pennell et al. 2014) or the *PCMVar* function from the R package PCMBASE (Mitov et al. 2019). A conceptual framework for computing these model distances is provided in Figure 1. In this study, we propose three probability distances that are based on the same distances that have previously been used for phylogenetic model distances (i.e., Garba et al. 2018): the Hellinger distance (d_H), the Kullback–Leibler divergence (d_{KL}), and the Jensen-Shannon distance (d_{JS}^2):

$$d_H(\varphi_1, \varphi_2) = d_H(N(\vec{\mu}_1, \Sigma_1), N(\vec{\mu}_2, \Sigma_2))$$

$$= 1 - 2^{n/2} \frac{|\Sigma_1|^{1/4}|\Sigma_2|^{1/4}}{|\Sigma_1 + \Sigma_2|^{1/2}} \exp\left\{-\frac{1}{4}(\vec{\mu}_1 - \vec{\mu}_2)^T(\Sigma_1 + \Sigma_2)^{-1}(\vec{\mu}_1 - \vec{\mu}_2)\right\} \quad (2) \text{ (Pardo 2005)}$$

$$d_{KL}(\varphi_1, \varphi_2) = d_{KL}(N(\vec{\mu}_1, \Sigma_1), N(\vec{\mu}_2, \Sigma_2))$$

$$= \frac{1}{2}\left(\text{tr}(\Sigma_2^{-1}\Sigma_1) + (\vec{\mu}_2 - \vec{\mu}_1)^T\Sigma_2^{-1}(\vec{\mu}_2 - \vec{\mu}_1) - n + \ln\left(\frac{|\Sigma_2|}{|\Sigma_1|}\right)\right) \quad (3) \text{ (Duchi 2007)}$$

$$d_{JS}(\varphi_1, \varphi_2) = d_{JS}(N(\vec{\mu}_1, \Sigma_1), N(\vec{\mu}_2, \Sigma_2))$$

$$= \frac{1}{2}d_{KL}\left(\varphi_1, \frac{\varphi_1 + \varphi_2}{2}\right) + \frac{1}{2}d_{KL}\left(\varphi_2, \frac{\varphi_1 + \varphi_2}{2}\right) \quad (4) \text{ (Lin 1991)}$$

where |A| and tr(A) are respectively the determinant and trace of matrix A, $\vec{x}^T$ is the transpose of vector $\vec{x}$, n is the number of tips on the tree, and $\frac{\varphi_1 + \varphi_2}{2}$ denotes a mixture of the two models φ_1 and φ_2. The Jensen-Shannon divergence is a metric with an upper bound of $\sqrt{\log(2)}$ (Garba et al. 2018), and we also note that no closed form solution exists for the Jensen-Shannon divergence between two MVN distributions because the mixture of two Gaussian distributions with distinct components is not Gaussian itself (i.e., the distributions will not be of the same Gaussian family; Nielsen 2019), but the Jensen-Shannon distance may be approximated using simulations (e.g., Abou-Moustafa and Ferrie 2012; Garba et al. 2018). The Kullback–Leibler divergence is notable for its role in model selection, as it forms the theoretical basis for the Akaike information criterion (AIC), which is designed to approximate the Kullback–Leibler distance between the true generating model and a fitted model (Akaike 1973). In our example demonstrations, we primarily focus on computing the Hellinger distance

(Eq. 2), which is a bounded metric with a maximum value of one (two models are completely divergent) and a minimum of zero (two models induce identical probability distributions and are mathematically indistinguishable). The Hellinger distance is a metric that satisfies the triangle inequality and has symmetry, with a distance of zero between two models indicating identical distributions. In contrast, the Kullback–Leibler divergence is not a metric because it is asymmetric (Johnson and Sinanović 2001).

Models of multiple trait coevolution can also be incorporated into these distances by including an evolutionary rate matrix that specifies the rate for each trait and the covariance between each pair of traits. For example, bivariate BM models can be implemented by including the evolutionary rate matrix $R = \begin{pmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_2^2 \end{pmatrix}$, where ρ represents the evolutionary covariance between the two traits and σ_i^2, $i \in \{1,2\}$, specifies a rate for each trait, and by setting $\vec{\mu}$ as a $2n$-dimensional vector containing the expected tip value (i.e., ancestral states) for both traits at each tip in the tree. With Σ representing an unscaled phylogenetic variance-covariance matrix (i.e., branch lengths are not already scaled by R), we can use the Kronecker product $V = R \otimes \Sigma$ of the evolutionary rate matrix R and Σ to compute d_H, d_{KL} and d_{JS}^2 with the following equations:

$$d_H(\varphi_1,\varphi_2) = d_H(N(\vec{\mu}_1, V_1 = R_1 \otimes \Sigma_1), N(\vec{\mu}_2, V_2 = R_2 \otimes \Sigma_2))$$

$$= 1 - 2^{n/2}\frac{|V_1|^{1/4}|V_2|^{1/4}}{|V_1 + V_2|^{1/2}}\exp\left\{-\frac{1}{4}(\vec{\mu}_1 - \vec{\mu}_2)^T(V_1 + V_2)^{-1}(\vec{\mu}_1 - \vec{\mu}_2)\right\} \tag{5}$$

$$d_{KL}(\varphi_1,\varphi_2) = d_{KL}(N(\vec{\mu}_1, V_1 = R_1 \otimes \Sigma_1), N(\vec{\mu}_2, V_2 = R_2 \otimes \Sigma_2))$$

$$= \frac{1}{2}\left(tr(V_2^{-1}V_1) + (\vec{\mu}_2 - \vec{\mu}_1)^TV_2^{-1}(\vec{\mu}_2 - \vec{\mu}_1) - 2n + \ln\left(\frac{|V_2|}{|V_1|}\right)\right) \tag{6}$$

$$d_{JS}(\varphi_1,\varphi_2) = d_{JS}(N(\vec{\mu}_1, V_1 = R_1 \otimes \Sigma_1), N(\vec{\mu}_2, V_2 = R_2 \otimes \Sigma_2))$$

$$= \frac{1}{2}d_{KL}\left(\varphi_1,\frac{\varphi_1 + \varphi_2}{2}\right) + \frac{1}{2}d_{KL}\left(\varphi_2,\frac{\varphi_1 + \varphi_2}{2}\right) \tag{7}$$

All features of both the tree and evolutionary model are likely to influence trait distributions, and therefore probabilistic distances. That is, perturbations to any model components are likely to be captured by probabilistic distances, including differences in evolutionary parameters (i.e., Table 1) and ancestral states, as well as the tree topology and branch lengths, which collectively determine the phylogenetic covariance structure. The effect of differences in mean trait values can be illustrated in an example of two univariate normal models $\varphi_1 : N(\mu = 0, \sigma^2 = 1)$ and $\varphi_2 : N(\mu = 2, \sigma^2 = 1)$, which can also be viewed as BM processes acting on a single species that diverged from an ancestor one unit length of time in the past (i.e., branch length $v = 1$) with mean trait values of either $\mu = 0$ or $\mu = 2$, respectively (Fig. S1a). Using Equation (2), the Hellinger distance between these two distributions is 0.39. If we increase the expectation to $\mu = 5$, then we obtain a third model $\varphi_3 : N(\mu = 5, \sigma^2 = 1)$ that is a Hellinger distance of 0.95 from model φ_1.

The particular timing and structure of phylogenetic relationships also influences trait distributions and therefore probabilistic distances by determining elements of the covariance matrices Σ, which can be demonstrated in another example with two multivariate normal models $\varphi_1 : MVN$ $(\vec{\mu} = 0, \Sigma_1 = \text{vcv}["(A:1,(B:2/3,(C:1/3,D:1/3):1/3):1/3)"])$ and $\varphi_2 : MVN$ $(\vec{\mu} = 0, \Sigma_2 = \text{vcv}["((A:1/2,B:1/2):1/2,(C:1/2,D:1/2):1/2)"])$. Here 0 is a vector of length four containing only zeros (i.e., mean trait values assumed to be zero, and under BM indicates ancestral state of zero), and $\text{vcv}[\cdot]$ indicates a function used to extract the phylogenetic covariance matrix according to the two different trees specified in quoted newick format. In this case, the covariance matrix Σ_1 reflects a "balanced" tree shape for model φ_1, while an "unbalanced" tree is used for φ_2. Given clear differences in both tree shape and branch lengths, we expect different trait distributions, and these differences are reflected by measuring a Hellinger distance of 0.65 between these two models (Fig. S1b).

Probabilistic distances are also influenced by tree size (i.e., number of taxa), which can be illustrated by computing distances between small versus large star phylogenies (Fig. S1c). In this case, the Hellinger distance between a pair of three-tip star phylogenies is smaller ($d_H = 0.08$; left example in Fig. S1c) than when computed between two larger four-tip star phylogenies ($d_H = 0.11$; right example in Fig. S1c), which each include an additional element of the variance-covariance matrix that must be considered. By adding this fourth species we have expanded the dimension of the probability distribution, requiring that the density be more spread out than in the setting with fewer species.

Computing Probabilistic Distances Under Evolutionary Models for Both Bifurcating Trees and Phylogenetic Networks

We demonstrated the properties of these distances under a range of evolutionary scenarios that modulate the magnitude of important model parameters defined by each of the six models (BM, OU, EB, L, K, and D; Table 1). We first used an example phylogeny and a randomly generated set of branch lengths (v are sampled according to an exponential distribution with a rate of one) for eight taxa (Fig. 2a) that was first employed by Felsenstein (1985) to illustrate the variance-covariance structure of phylogenetic trees and the comparative method (i.e., Fig. 8 in Felsenstein 1985). In this demonstration, we sought to leverage probabilistic distances to capture and quantify differences in evolutionary model parameters that influence the probability distribution over $\vec{\chi}$, and thus the tree topologies and branch lengths were identical for both φ_1 and φ_2 (i.e., $T_1 = T_2$ and $v_1 = v_2$) in our bifurcating tree examples depicted in Figure 2. For legibility, we drop the T and v notation from φ except where noted.

For each phylogenetic model comparison, we computed $d_H(\varphi_1, \varphi_2)$ between pairs of models for which φ_1 represented a BM-based model ($\varphi_1 = \varphi_1^{BM}$), and the second model $\varphi_2 \in \{\varphi_2^{BM}, \varphi_2^{OU}, \varphi_2^{EB}, \varphi_2^L, \varphi_2^K, \varphi_2^D\}$ was varied according to one of the six models. To explore the properties of the distances under different baseline evolutionary rate scenarios, we varied the rate parameter between either $\sigma_1^2 = 1$

or $\sigma_1^2 = 2$ for φ_1, while fixing $\sigma_2^2 = 1$ for the second model φ_2, and we set the ancestral root state as $\theta_1 = \theta_2 = 0$ in these example demonstrations. For each variant of φ_2, we selected a single parameter to be scaled by a factor $\gamma \in [0,5]$ (scaled parameters shown in Table 1).

Phylogenetic networks pose a number of unique challenges to evolutionary inference, including identifiability issues for certain types of models (Zhu and Degnan 2017). For phenotypic data, many network-based models treat the trait value of a reticulation point as a weighted mean of its respective parental lineages (Bastide et al. 2018). To understand the dynamics of evolutionary model distances under more complex tree structures, we also computed probabilistic distances for a pair of hybridization networks (Fig. 3a) and additionally, between a network and a strictly bifurcating tree (Fig. 3c). The network topology was simulated using the *SimulateNetwork* function provided in the R package BMHYD (Jhwueng and O'Meara 2015) using a birth-death model (Nee et al. 1994) and a single unidirectional pulse migration of proportion $m = 1/2$, with birth rate of 1 and death rate of $1/2$; this network was also pruned to generate a second bifurcating topology by removing the migration edge (Fig. 3c). For the first model φ_1, we applied a standard BM model with $\sigma_1^2 = 1$, and for the second model φ_2, we then scaled either the rate $\sigma_2^2 = \gamma\sigma_1^2$ or the migration proportion $m_2 = \gamma m$ by a factor of $\gamma \in [0,2]$, which allowed us to investigate the impacts of scaling these two parameters when comparing model distances.

We evaluated the effects of particular fixed trees shape while increasing the number of sampled species n when computing probabilistic distances. For each number of sampled species $n \in$ {4, 8, 16, 32, 64, 128, 512, 1024, 2048}, we simulated three different trees using a fixed shape ("balanced", "left unbalanced", or "star"; example topologies shown in Fig. 4) and we computed pairwise Hellinger distances (d_H) between a BM model φ_1^{BM} and either the OU model with $\alpha \in$ {0.01, 0.02,..., 0.10, 0.20, ..., 1, 2,...,10}, the EB model with $a \in$ {0.01, 0.02,..., 0.10, 0.20, ..., 1, 2,...,10}, or the D model with $\delta \in$

{0.01, 0.02,..., 0.10, 0.20, ..., 1, 2,..., 10}; we applied a rate $\sigma^2 = 1.0$ for all four models (BM, OU, EB, and D). We repeated this analysis twice using two different models of branch lengths on trees where lineage splits are evenly distributed from the time of sampling to the root: (1) all internal branches and the shortest external branches are of equal length and scaled to give a total tree height of 1.0, and (2) all internal branches and the shorted external branches are of equal length and scaled to give a total tree length of 3.0. Additionally, we evaluated the effects of increasing the number of sampled species n when computing probabilistic distances under three different models (BM, OU, and EB) by simulating phylogenetic trees using a pure-birth model (Yule 1925) for $n \in$ {4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048} taxa. For each value of n, we simulated 100 random Yule trees with a birth rate of 10 using the *pbtree* function from the R package PHYTOOLS (Revell 2012), and we computed all $\binom{100}{2} = 4950$ pairwise distances between these sets of 100 trees using either a BM model with varying $\sigma^2 \in \{1,10,100\}$, an OU model with varying $\alpha \in \{1,10,100\}$, or an EB model with varying $a \in \{1,10,100\}$ to all simulated trees. We repeated this analysis for the same model parameters (i.e., 1, 10, and 100 for the σ^2, α, and a parameters) and tree sizes (i.e., $n \in \{4,8,...,1024\}$ tips) for trees simulated under the Aldous' Branching model, which is defined by a symmetric split distribution: $q(n, i) = \frac{n}{2h(n-1)} \times \frac{1}{i(n-1)}$, where $h(n)$ is the nth harmonic number (Aldous 1996). Trees simulated under this model were generated using the *rtreeshape* function provided in the R package APTREESHAPE (Bortolussi et al. 2006).

Investigating the Interplay Between Probabilistic Distances and Likelihood Ratio Test Significance
We conducted an array of simulations to investigate the interplay between probabilistic model distances and the significance of the likelihood ratio test used for model selection. First, we simulated a random $n = 100$ tip phylogeny based on a pure-birth Yule process (Yule 1925) with a birth rate of 10 (Fig. S6a). Next, we simulated datasets under an OU model with $\sigma^2 = 1$ and $\alpha \in [0,1]$. For each value of α

in this range, we simulated 100 replicate datasets with this model $\varphi^{OU}(\sigma^2 = 1, \alpha)$, and for each replicate, we fit two alternative models to the simulated data: (i) a BM model and (ii) an OU model. Model parameters were estimated using maximum likelihood with the *fitContinuous* function provided in GEIGER (Harmon et al. 2007). We used the results of these two fitted models to compute a likelihood ratio test with significance assessed assuming a chi-squared distribution with one degree of freedom. Because all observations were simulated under an OU model, fitting a BM model in this case represents a scenario of model-misspecification, and we used these simulations to characterize the relationship between the significance of the likelihood ratio test between two models and the probabilistic distance between them. We repeated this same analysis for a larger tree with $n = 1000$ tips (Fig. S7a).

We also expanded these simulations to investigate the impacts of tree shape, tree size, and evolutionary parameters on the significance of the likelihood ratio test alongside probabilistic distances. We simulated datasets under an OU model using three different tree sizes ($n = 128$, 512, and 1024 tips) and three different tree shapes ("balanced", "left unbalanced", and randomly-generated Yule tree with birth rate $\lambda = 1$; example tree shapes shown at the top of Fig. 5) using branch lengths scaled to give a total tree height of 1.0. For the "balanced" and "left unbalanced" shapes, lineage splits are evenly distributed from the time of sampling to the root of the tree, and all internal branches or shortest external branches are of equal length. For each tree size and shape, we simulated character trait datasets using an OU model that varied in the parameter $\alpha \in [0,1]$, and computed both the likelihood ratio test and Hellinger distance between an OU and BM model that have each been fit to the simulated dataset using the *fitContinuous* function of GEIGER.

Leveraging Probabilistic Distances to Compute Distances Between Fitted Evolutionary Models

For our first empirical demonstration, we applied probabilistic distances to compare the fit of each of the six evolutionary models investigated in this study (BM, OU, EB, L, K, and D) for a dataset consisting of mean genome size estimates (mean c-value) for $n = 465$ amphibian species obtained from previous studies (Pyron 2014; Liedtke et al. 2018). Amphibian genomes are known to exhibit extreme variation in size among all vertebrates, and a recent study highlighted the extraordinary complex evolutionary dynamics of this trait within this clade (i.e., Liedtke et al. 2018). Though these six models are likely highly oversimplified for this particular trait and clade, we can nonetheless use this system as an example for computing probabilistic distances to dissect relative evolutionary model fit and its impacts on continuous trait distributions (i.e., as measured via model distances; Fig. 1). We employed a phylogenetic tree estimate alongside genome size data obtained from Pyron (2014) and Liedtke et al. (2018), respectively, and fit each of the six models to the dataset. We then computed probabilistic distances (d_H here) between each pair of fitted models to demonstrate the utility of d_H for comparing fitted models to one another.

Computing Pairwise Distributions of Probabilistic Distances to Investigate Phylogenetic Uncertainty

We demonstrated the application of our probabilistic distance approach for probing the impacts of phylogenetic uncertainty when studying the evolution of the total genomic transposable element (TE) content for a dataset of $n = 48$ bird genomes (Jarvis et al. 2014). In the context of this study, we refer to "phylogenetic uncertainty" as a lack of knowledge pertaining to which particular phylogenetic background is more or less appropriate for modeling the evolution of a given trait (i.e., Maddison 1997), rather than uncertainty in the resolution of species relationships for any one tree (i.e., Huelsenbeck et al. 2000). Because the phylogeny defines the degree of relatedness among species and in turn, covariation in their biological characteristics (i.e., Figs. 1 and S1), understanding the specific

phylogenetic framework that a trait evolved under may be essential for accurate inferences—yet, there are many reasons why one tree estimate differs from another (Degnan and Rosenberg 2009; Liu et al. 2015). Thus, it can be difficult to decide which tree is "best" when confronted with a set of plausible phylogenetic hypotheses, particularly when studying complex traits. Importantly, choosing any one tree, such as an overall species tree estimate, may not be the best choice (i.e., Hahn and Nakhleh 2016). Indeed, Hahn and Nakhleh (2016) noted that every accompanying analysis of avian trait evolution based on this dataset of 48 bird genomes assumed a single species tree, despite notoriously high levels of gene tree conflict within this dataset (e.g., none of the individual gene trees matched the species topology).

We chose a particularly complex trait (percent genomic TE content) to demonstrate properties of probabilistic distances in this context. In this case, it is not immediately clear how the evolution genomic TE content should be most appropriately modeled (i.e., species trees versus gene trees), and we seek to apply probabilistic distances to investigate such uncertainty. To account for uncertainty in phylogenetic framework, we computed pairwise model distances under BM and OU models for two independent sets of trees that have been estimated for these 48 bird genomes: (i) 6144 individual gene trees (2136 exons, 329 introns, and 3679 UCEs) and (ii) 31 species trees (Reddy et al. 2017). We downloaded these sets of trees (Mirarab et al. 2014), we fit both a BM or an OU model independently to each tree (i.e., each gene and species trees), and trait value (i.e., point estimate of percent genomic TE content for each species) using the function *fitContinuous* function in GEIGER. We computed pairwise distributions among all fitted model distances for both tree sets, and conducted MDS using *cmdscale* to visualize these pairwise distances projected into multidimensional space and among trees and between BM and OU models.

Applying Probabilistic Distances to Multi-trait Models of Continuous Trait Evolution

We demonstrated the application of probabilistic distances in two different scenarios of multiple trait evolution. In the first scenario, we simulated a 25-tip phylogeny with a birth rate of one using the function *pbtree* provided in PHYTOOLS (tree shown in Fig. 8a); this tree was used to simulate bivariate datasets with differing degrees of covariation between the two traits using a BM model with an evolutionary rate matrix $R = \begin{pmatrix} \sigma_1^2 = 1 & \rho \\ \rho & \sigma_2^2 = 1 \end{pmatrix}$, where ρ represents the evolutionary covariance between the two traits and $\sigma^2 = 1$ for both traits. For each simulated dataset, we fit two BM models: $\varphi_1^{\text{BM}}\begin{pmatrix} \hat{\sigma}_1^2 & \hat{\rho} \\ \hat{\rho} & \hat{\sigma}_2^2 \end{pmatrix}$ and $\varphi_2^{\text{BM}}\begin{pmatrix} \hat{\sigma}_1^2 & \rho = 0 \\ \rho = 0 & \hat{\sigma}_2^2 \end{pmatrix}$, where $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ indicates that the rates for both traits are estimated from the data, whereas $\hat{\rho}$ denotes that the covariance between the traits is also estimated for the first model. That is, φ_1^{BM} represents an unconstrained model for which both the covariance ρ and the rates σ_1^2 and σ_2^2 are estimated, whereas φ_2^{BM} represents a constrained model that assumes $\rho = 0$ (i.e., φ_2^{BM} assumes traits evolve independently). We used the function *mvBM* provided in the R package mvMORPH (Clavel et al. 2015) to fit both models and compute the likelihood ratio, and we measured the Hellinger distance (d_{H}) between the two models across a non-negative range of covariance $\rho \in [0,1]$.

In the second scenario, we applied probabilistic distances to characterize model divergence when studying traits that have experienced singular evolutionary events, which have recently been shown to mislead phylogenetic comparative methods when neglected (Uyeda et al. 2018). Indeed, Uyeda et al. (2018) used such a scenario to demonstrate how instantaneous character trait shifts can yield misleading evidence of apparent correlations between two traits that, in fact, evolved independently of one another. Following the approach of Uyeda et al. (2018), we recreated a version of "Felsenstein's worse-case scenario" by simulating bivariate datasets of independent traits (i.e., $\rho = 0$ for all simulations) along a 40-tip phylogeny (i.e., Fig. 5 in Felsenstein 1985, Fig. 2a in Uyeda et al.

2018, and Fig. 8c in this study). On one of the two stem branches subtending the root, we simulate an instantaneous shift in the character trait value that is drawn from a MVN distribution with zero covariance and equal variances that are a scalar value $\sigma^2 = 1$. That is, a large variance of this MVN distribution used to randomly draw a shift value is likely to produce larger magnitude shifts or jumps in trait value, and for smaller variances is likely to yield smaller magnitude shifts or jumps in trait value. For each simulated dataset, we fit two BM models: $\varphi_1^{\text{BM}}\begin{pmatrix} \hat{\sigma}_1{}^2 & \hat{\rho} \\ \hat{\rho} & \hat{\sigma}_2{}^2 \end{pmatrix}$ and $\varphi_2^{\text{BM}}\begin{pmatrix} \hat{\sigma}_1{}^2 & \rho = 0 \\ \rho = 0 & \hat{\sigma}_2{}^2 \end{pmatrix}$, and we computed both the likelihood ratio test as well as Hellinger distance (d_{H}) between them.

To further investigate distances between models that differ in their underlying assumptions of complex traits coevolution, we also applied probabilistic distances to an alignment of 29 morphometric characters taken from a dataset of 87 cranium landmarks (each trait spans three landmarks) measured for 10 extant and nine extinct carnivoran mammals from a recent study (Álvarez-Carretero et al. 2019). We obtained a dated phylogeny hypothesis from Álvarez-Carretero et al. (2019) referred to as "morph+molec B" that was inferred using both a set of morphological and molecular characters (see Fig. 8 in Álvarez-Carretero et al. 2019 for details). Each of the 29 morphometric characters consists of a triplet set of landmark measurements, and thus, we fit two different BM models to each of the 29 triplet traits: a constrained model (i.e., landmarks assumed to evolve idependently) and an unconstrained model for which between-landmark covariances are estimated for each of the three landmarks associated with a given trait. We computed pairwise Kullback–Leibler divergences (d_{KL}) between all 29 fitted models for each set (i.e., constrained vs. unconstrained model sets) independently. Next, we conducted MDS of each set of 29 traits independently to project and visualize these pairwise model distances to the first two principle coordinates of variation.

Investigating Identifiability of Mixed Gaussian Models

Recently, a number of models have been proposed that relax the assumption of model homogeneity across a phylogeny (Mitov et al. 2019, 2020), and studies have highlighted intrinsic difficulties of inferring multiple rate optima of the OU model (Ho and Ané 2014). We applied the Hellinger distance to investigate the identifiability for three scenarios of mixed OU models on an example phylogeny of flowering plants (Davis et al. 2007). For the first scenario, we computed distances using two examples of a single shift model discussed in Ho and Ané (2014) and depicted in the left and right trees shown in Figure 9a that varied in the parameters of the shift. For the left tree model, we applied a background OU model with parameters $\sigma^2 = 1$ and $\alpha = 1$ using an ancestral state θ and selection optimum θ_{opt} of one (i.e., $\theta = \theta_{\text{opt}} = 1$), and we then shifted the optimum $\theta_{\text{shift}} = 2$ at the branch marked by an asterisk (Fig. 9a, left tree). Alternatively, we varied the ancestral state $\theta \in [-2, 3]$ and set the shift optimum $\theta_{\text{shift}} = \theta$, while varying the background optimum $\theta_{\text{opt}} \in [-2, 3]$, to compute distances for the right tree in Figure 9a to compute the Hellinger distance across a range of values for these parameters, which have been shown to be unidentifiable (Ho and Ané 2014).

For the second scenario, we computed distances between mixed models with two shifts in α and depicted in the left and right trees shown in Figure 9c that both applied a background OU model with parameters $\sigma^2 = 1$ and $\alpha = 1$, while varying the location of one of the two OU shifts (location of asterisks in left vs. right trees of Fig. 9c). For the left tree model, we used $\sigma^2 = 1$ and $\alpha = 2$ for the first shift of the OU and $\sigma^2 = 1$ and $\alpha = 3$ for the second shift. We computed distances between this shift model and a separate shift model with a different location for the second shift (gray asterisk shown in the right tree of Fig. 9c) for which we varied the $\sigma^2 \in [0,5]$ and $\alpha \in [0,5]$ parameters. Similarly, for a third scenario, we computed probabilistic distances between two different shift models depicted in the left and right trees of Figure 9e, with the same values for all shift models described in the second scenario, but different locations in the trees (Fig. 9c vs. Fig. 9e).

Alongside our mixed OU scenarios, we also investigated model divergences between two different applications of mixed BM models (left trees with asterisks shown in Figs. S9a and c) and the K model (right trees without asterisks in Figs. S9a and c), which varied both the parameter values (σ_1^2 for the BM marked by gray asterisk, and the κ parameter of the K model) and the tree topology for five taxa (Fig. S9a vs. Fig. S9c). For the first mixed BM model (Fig. S9a), we applied a background BM with $\sigma^2 = 1$, and used $\sigma_1^2 \in [0,5]$ (gray asterisk), $\sigma_2^2 = 3$, and $\sigma_3^2 = 2$, for the three shifts marked in Figure S9a, respectively. For the second mixed BM model, we also applied a background BM with $\sigma^2 = 1$ while varying the rate parameter $\sigma_1^2 \in [0,5]$ for the BM shift marked by a gray asterisk in Figure S9c. In both cases, we varied the $\kappa \in [0,5]$ parameter of the K model. We repeated this analysis for larger trees with $n = 1027$ tips (Fig. S10) that were generated by appending a "balanced" subtree with 1024 tips (Fig. S10a) or an "unbalanced" subtree with 1024 tips (Fig. S10c) each with equal branch lengths of one (i.e., different tree heights and a length of one for each branch of the subtree).

In addition to scenarios of mixed models, we applied the Hellinger distance (d_H) to investigate identifiability (or lack thereof) between the OU model and the generalized EB model for which the rate parameter is positive (sometimes referred to as the Acceleration-Deceleration model; Blomberg et al. 2003). We downloaded a phylogeny of $n = 100$ Anolis lizards (Mahler et al. 2010) that was also analyzed in a recent study that highlighted unidentifiability of the OU and EB models (Uyeda et al. 2015), which can occur when the initial rate parameter of the EB model is parameterized as $\sigma_0^2 = e^{-2\alpha}$, where α corresponds to the OU model parameter α (Uyeda et al. 2015). We computed the Hellinger distance across a range of values for a using an EB model with $\sigma_0^2 = e^{-2\alpha} = e^{-2(0.25)} = 0.607$ and $a \in [0,2]$, and an OU model with parameters $\sigma^2 = 1$ and $\alpha = 0.25$. Using these results, we also computed the absolute difference in Hellinger distances measured across the range of a to characterize the rate of change in model divergence.

RESULTS

Probabilistic Distances Between Capture Impacts of Model Parameter Scaling

Computing probabilistic distances across models highlights the impacts of scaling particular model parameters designed to capture different evolutionary processes acting on trait distributions for both bifurcating trees (Fig. 2) and phylogenetic networks (Fig. 3). By scaling these model parameters by a factor $\gamma \in [0,5]$, probabilistic distances provide a clear means for assessing model identifiability (or lack thereof) and for dissecting the relative impacts of different parameters on phenotypic trait distributions. In all cases, we find substantial differences in slopes of the curves generated by computing Hellinger distances between a standard BM model ($\varphi_1 = \varphi_1^{BM}$) with $\sigma_1^2 = 1$ (Fig. 2b) or when $\sigma_1^2 = 2$ (Fig. 2c) against the six alternative models with scaled parameters $\varphi_2 \in \{\varphi_2^{BM}, \varphi_2^{OU}, \varphi_2^{EB}, \varphi_2^{L}, \varphi_2^{K}, \varphi_2^{D}\}$. Computing a maximum approximate derivative (maximum rate of change in these slopes) highlights particular model parameters with especially strong influence on model distances, and therefore the underlying continuous trait probability distributions when scaled by a factor γ (i.e., comparing the maximum approximate derivatives shown in Figs. 2b and c). We also find that as we scale each parameter by larger γ values (as $\gamma \rightarrow 5$ in Figs. 2b and c), several distances exhibit increasing trends and/or asymptotic behaviors toward the maximum d_H of 1, suggesting that the underlying probability distributions of continuous traits become increasingly divergent when parameters are scaled in such a fashion.

In our bifurcating tree scenarios (Fig. 2), probabilistic distances measured between φ_1 and each of the respective scaled BM $\varphi_2^{BM}(\sigma^2 = 1\gamma)$, kappa $\varphi_2^{K}(\kappa = 1\gamma)$, delta $\varphi_2^{D}(\delta = 1\gamma)$, and lambda φ_2^{L} ($\lambda = 1\gamma$) models approach zero when their respective parameters are essentially unscaled (i.e., these distance curves approach zero when $\gamma = 1$ in Fig. 2a), indicating that these models become unidentifiable under these conditions because they induce identical distributions over $\vec{\chi}$. In other words, the K, D, and L models collapse to a simple BM model as γ approaches 1 for each model. Conversely,

scaling the a and α parameters for the EB and OU models, respectively, did not result in model unidentifiability when computed against a standard BM model φ_1^{BM} when $\sigma^2 = 1$ (Fig. 2b) or when $\sigma^2 = 2$ (Fig. 2c). Indeed, only the scaled BM and delta models approach unidentifiability when γ approaches 2 for scenarios with φ_1^{BM} and an evolutionary rate of two (i.e., $\sigma^2 = 2$), while the other four models (OU, EB, K, and L) remain identifiable for all values of γ (Fig. 2c).

Similarly, computing both the Hellinger (d_{H}) distance and the Kullback–Leibler (d_{KL}) divergence between BM models evolving according to a pair of hybridization networks (Fig. 3a) or a hybridization network and a strictly bifurcating tree (Fig. 3c) highlighted the utility of such approaches for probing the effects of scaling evolutionary rate or the migration proportion m on trait distributions underlying model distances and complex phylogenetic structures (Figs. 3b and d). We observed fundamentally different curves for d_{H} and d_{KL} across a range of scaling for either m or σ^2, as well as apparent shifts in model identifiability under these conditions. For example, when $\gamma = 0$ yielding $m = 0.5\gamma = 0$, both the d_{H} and d_{KL} distance curves approach zero when comparing the network and tree shown in Figure 3c, such that the hybridization network essentially collapses to the tree as the migration proportion goes to zero. In many cases, scaling the evolutionary rate parameter σ^2 has a particularly strong impact on model distances under both networks and trees.

Tree shape, number of taxa, branch length model, and the particular parameter values had substantial influence over probabilistic distances (Figs. 4 and S1). In particular, the specific model used to generate branch lengths had a strong and synergistic role in shaping probabilistic distances between different models and tip numbers (Fig. 4 vs. Fig. S1). When using branch lengths are scaled to a tree height of one, we found that increasing the number of sampled species n on a fixed topology shape yields trends of increasingly larger pairwise distances between a BM model and either the OU (Figs. 4a-c), EB (Figs. 4g-i), or D (Figs. 4m-o) models. However, we observed an opposite trend towards smaller Hellinger distances for the largest trees when branches are equal in length and scaled such that

the total tree length is three for the OU and EB model comparisons (Figs. S1a-f). We also found that distances computed under the D model were less sensitive (compared to the OU and EB comparisons) to the branch length model (Fig. 4g-i vs. Fig. S1g-i). One reason that may explain these differences for the D model is that the tree becomes more star-like as the δ parameter increases, which in turn, increases the total length of the tree as well as the total amount of evolution occurring along branches. For a given number of taxa n and height h, the total tree length approaches nh as the δ parameter increases, thereby resulting in greater evolution along the tree with a larger number of taxa. Specifically, with large δ, for our balanced tree scenario (Fig. S1g with original fixed length of three) the total tree length increase logarithmically as a function of the number of tips, and for our left unbalanced scenario (Fig. S1h with original fixed length of three) the total tree length approaches six with increasing tip number. On our examples simulated under the Yule (Fig. S3) or Aldous' branching (Fig. S4) models, we observed similar trends toward increasingly larger pairwise distances under the BM (Figs. S3a and S4a), OU (Figs. S3b and S4b), and EB (Figs. S3c and S4c) models for larger trees, respectively (see *Methods* for details). However, the influence of increasing the number of sampled species n depended on the relative scaling of each parameter (dark to light distributions indicated in Figs. S3 and S4). In all examples, modulating specific evolutionary parameters tended to have a strong and synergistic influence with tree shape and taxa number on probabilistic distances (Figs. 4 and S2-S4). We also found that probabilistic distances measured between even the largest trees in our simulations (2048 species) required less than a minute of computation using a single thread of a 2.6 GHz Intel Core i7 CPU (Fig. S5).

Probabilistic Distances and Likelihood Ratio Test Significance

The fit of nested evolutionary models (i.e., BM and OU) to a dataset of continuous characters is often compared using the likelihood ratio tests (i.e., O'Meara et al. 2006). In this context, our simulation

analyses identified a positive relationship between increasing model distances and significance (i.e., smaller *p*-values) for the likelihood ratio test between fitted OU and BM models, which varied according to the size and shape of the tree, as well as model parameters (Figs. 5 and S6-S7; see *Methods* for details). That is, the ability to identify statistically significant differences between the fitted models increased as the models became more divergent, because the data were generated under increasingly larger α values for the OU model. Conversely, smaller α models tended to decrease the ability to detect statistically significant differences between the models because the parameter estimates of the BM and OU models were more similar, and thus, they induce similar underlying distributions over character traits $\vec{\chi}$ that were quantified via the Hellinger distance (Figs. 5, S6, and S7).

Comparing Fitted Models with Probabilistic Distances

Size is a fundamental characteristic of genomes, and yet, accurately modeling the evolution of genome size can be challenging (e.g., Liedtke et al. 2018). Comparing the fit of the six evolutionary models to the amphibian genome size dataset consisting of 465 species provided a detailed depiction of the relative distance between each estimated model (Fig. 6; see *Methods* for details). For example, the distance between the fitted BM and OU models (5.7×10^{-6}) was substantially smaller than any other pairwise model distance (see the BM-OU edge in Fig. 6). This small distance between the BM and OU models suggests that the estimated α parameter for the OU model does not strongly influence the underlying distribution of trait values, and thus the estimate α is close to zero, indicating a lack of stabilizing selection for genome size. Conversely, the largest pairwise distance of any two models was observed between the EB and K models for this dataset (i.e., the large EB-K edge in Fig. 6).

Investigating the Impacts of Phylogenetic Uncertainty with Probabilistic Distances

Rigorous modeling of trait evolution often requires accounting for uncertainty in the phylogenetic relationships assumed to underlie a given trait (Huelsenbeck et al. 2000). Our applications of probabilistic distances to the 6144 gene trees (2136 exons, 329 introns, and 3679 UCEs) and the 31 species trees estimated for a 48 bird genome dataset highlights the utility of this approach for probing the impact of such uncertainty on evolutionary model comparisons (Fig. 7; see *Methods* for details). For example, our analyses of the avian dataset revealed substantial differences among distances computed using the individual gene trees or the reconstructed species-level trees (i.e., Figs. 7a vs. b). Similarly, we found major differences in the MDS projections for both the gene and species tree analyses depending on whether a simple BM or OU model was used. In particular, we see substantial divergence in pairwise model distances between the 6144 gene trees that were based on the BM or OU models, respectively (i.e., Figs. 7a vs. c).

Applying Probabilistic Distances to Models of Multi-trait Evolution

Determining the degree of covariation (or lack thereof) between two traits is a major goal of phylogenetic comparative methods (Uyeda et al. 2018), and probabilistic distances can be useful in this context for quantifying model divergence and its relation to model testing (Fig. 8). We found that increasing the degree of evolutionary covariation between two traits yields corresponding increases in probabilistic distances between models that differ on whether traits are or are not assumed to be independent (Fig. 8b). In particular, our application of the Hellinger distance (d_H) highlights the divergence between such models when analyzing independent traits that have evolved under scenarios of instantaneous character trait change (i.e., Figs. 8c and d). Though the two traits were simulated independently from one another, we found that as the magnitude of the instantaneous shift increases, so does the Hellinger distance, as well as statistical support (i.e., lower *p*-values) for an incorrect model

of trait coevolution (Fig. 8d). Applying MDS of pairwise probabilistic distances computed between multi-trait models fit to morphometric landmark characters yielded differences in model-space depending on whether landmarks of each trait are assumed to evolve independently or in a correlated fashion (Fig. S8). Most multi-trait models are more tightly clustered together (i.e., pairwise model distances tend to be smaller) when fitting models that assume independence in comparison to models that estimate between-landmark evolutionary covariance (Fig. S8a vs. S8b), and we also see that the particular traits driving most of the variance in the coordinates differ between these analyses (i.e., the placement of models fitted to trait 27 vs. trait 25 in Fig. S8a vs. S8b driving variation in the first coordinate).

Investigating Probabilistic Distances Between Complex Gaussian Models

Our applications of the Hellinger distance to multiple OU models with shifted parameters underscored the utility of this approach for investigating the degree of identifiability between complex models (Fig. 9). For example, distances measured between two models depicted in Figure 9a that varied in the locations and parameter values of OU shifts highlighted regions of parameter space for which the models were mathematically unidentifiable, as well as nearby regions with particularly small distances that may be unidentifiable in practice. We also find that overall levels of identifiability were lower between the two shift models shown in Figure 9c (left vs. right tree), with the Hellinger distance ranging from zero to a maximum of 0.92 for these parameter values (Fig. 9d), when contrasted to the third example scenario depicted in Figure 9f. Indeed, when comparing the two scenarios (i.e., middle vs. bottom rows of Fig. 9), we observed a more diffuse range of Hellinger distances (i.e., larger band of gray shading in Fig. 9d vs. Fig. 9f) in the first scenario, indicating that these models may be more difficult to distinguish from one another as measured by the lower Hellinger distances, at least for many of the parameter values explored here.

Measuring model divergences between our two applications of mixed BM models and the K model highlights regions of parameter space for which these two models become unidentifiable (Figs. S9 and S10). The identifiability (or lack thereof) of mixed BM and K models depends on the particular shape of the tree (Fig. S9a vs. Fig. S9c and Fig. S10a vs. Fig. S10c), with a larger region of unidentifiability (or nearly so) depicted in our second example (Fig. 9b vs. Fig. 9d and Fig. S10b vs. Fig. S10d). Additionally, our applications of probabilistic distances to the generalized EB model highlighted regions of parameter space that may lead to identifiability issues with an OU model (Fig. S11a), as well as the rate of change in model distances as a function of the parameter a (Fig. S11b). We find that the rate of model convergence increases as the parameter a approaches 0.5, while eventually trending toward an asymptote of zero as a increases and the models diverge from one another (right-hand side of the curves shown in Fig. S11b).

DISCUSSION

Statistical models of evolutionary change convey information encoded within their parameters to define probability distributions over character traits given a phylogeny. We have found that computing distances under such models can be useful for elucidating the impacts of excluding, including, and/or modulating various parameters (i.e., Figs. 2-4) for both strictly bifurcating trees (i.e., Fig. 2), as well as more complex phylogenetic structures and models, such as hybridization networks (Fig. 3), scenarios of trait coevolution (Figs. 8 and S8), and mixed Gaussian models that allow shifts in the models and parameters along different lineages (Figs. 9, S9, and S10). Throughout our demonstrations, we explored diverse phylogenetic backgrounds across a range of tree sizes and shapes to gain insight into the properties of these measures in both data-limited (e.g., small trees) and data-rich (e.g., large trees) scenarios that are both of relevance to empirical studies. Collectively, we also find that probabilistic

distances measures provide a toolset for investigating model divergences in a way that is complementary with both likelihood-based model selection criterion (e.g., Figs. 5, S6, and S7) and mathematical proofs of model identifiability (or lack thereof; e.g., Fig. S11).

Phylogenetic trees and networks represent powerful and complementary frameworks for modeling the evolution of molecular sequence data (Blair and Ané 2019), and our results suggest the same when considering continuous traits. Probabilistic distances offer a means for quantifying the impacts of assuming one phylogenetic background or another on inferences of character evolution (i.e., Figs. 3 and 7), and for investigating regions of parameter space that yield identical trait distributions for models (e.g., Figs. 2, 3, 9, and S9-S11). Fundamentally, these distances can be interpreted as capturing underlying differences in character trait distributions between two models. For example, scaling the overall BM rate parameter σ^2 resulted in particularly strong impacts on model distances in many cases, and therefore, these distances were able to quantify impacts of this parameter on predicted trait distributions under models with faster or slower rates (i.e., Figs. 2-4). Scaling the migration proportion m of a hybridization network model permitted investigation of the effects of including or excluding a migration edge in the tree for continuous trait distributions (Fig. 3). Likewise, increasing the number of sampled taxa n in the model yields increasing trends in pairwise model distances, but these trends are modulated by the particular parameters included in the models, as well as the particular shape of the tree (i.e., Figs. 4 and S2-S4). Projections of multi-model space based on these distances depend on whether models incorporate trait coevolution, or alternatively, models that assume independence have been fit to morphometric landmark data (Fig. S8a vs. Fig. S8b). Taken together, these results underscore the complexities of comparing phylogeny-aware models, such that scaling and including or excluding parameters can have complex implications for character trait evolution and tree distances.

A growing concern in comparative trait studies is the identifiability (or lack thereof) of many commonly-used models of evolutionary inference (e.g., Ho and Ané 2014; Zhu and Degnan 2017), with recent evidence implicating the non-identifiability of entire classes of popular birth-death models that are often used for studying trait evolution (Louca and Pennell 2020). Relevant to these concerns is the distinction between mathematical and practical identifiability: two models are mathematically indistinguishable when they induce identical probability distributions, whereas two models that are mathematically distinguishable can still be "practically indistinguishable" when model inference is unreliable for finite datasets of reasonable size (Zhu and Degnan 2017). Thus, the recent emergence of probabilistic distances that can be used to directly measure model identifiability, such as those proposed in this study and others (i.e., Garba et al. 2018; Adams and Castoe 2019b), provide a timely and relevant framework for assessing these important properties of phylogenetic models. As two models converge toward the same probability distribution and therefore mathematical unidentifiability, the probabilistic distance between them will correspondingly decrease toward zero. Thus, our applications have shown that probabilistic distances can shed light on particular regions of parameter space that yield indistinguishable models, including more complex scenarios and mixed models that incorporate shifts in parameters along particular branches or clades of a tree (i.e., Figs. 9, S9, and S10). Finally, we have found that probabilistic distances provide a clear connection with likelihood-based tests of model fit (i.e., Figs. 5, S6, and S7). Significance of statistical tests of evolutionary model fit are likely a complex function of many parameters, such as the number of tips, the divergence times, the tree shape, number of traits, and the particular model analyzed (Ho and Ané 2014), and these features may be reflected in our analysis (e.g., Fig. S6 vs. Fig. S7).

Studying the evolution of complex traits often requires the use of parameter-rich models that more adequately accommodate these complexities. In our avian empirical example, we deliberately chose such a trait (percent genomic TE content) to demonstrate the utility of pairwise probabilistic

distances when considering uncertainty in the phylogenetic background assumed for a given trait. As in this case, it is not always straightforward to decide how the phylogenetic framework for a particular trait should best be modeled (i.e., a single gene tree, multiple gene trees, or a single species tree). For example, if a particular trait is known to be encoded by a single locus, then it may be preferable to assume the specific gene tree of that locus, rather than an overall species tree. Conversely, some traits may be better modeled as a function of the overall species tree, or perhaps multiple gene trees for polygenic traits. In practice, however, it is seldom known which or even how many gene trees underlie a particular trait, and recent studies have demonstrated that focusing only on one particular tree (usually the species tree) can be problematic when modeling trait evolution (i.e., Hahn and Nakhleh 2016). Traits that evolved along discordant gene trees may yield patterns of "hemiplasy" when forced to a species tree, leading to incorrect inferences (i.e., Avise and Robinson 2008; Mendes and Hahn 2016; Guerrero and Hahn 2018). In our case, modeling total genomic TE content is particularly challenging: should we only use gene trees located in or nearby TE-rich regions in the genome? Should we use multiple gene trees? The overall species tree? For example, many studies of TE evolution typically assume a single species-level phylogeny (e.g., Malmstrøm et al. 2018), and most analyses of avian trait evolution based on these 48 genomes have assumed only a single tree, despite widespread phylogenetic discordance within this dataset (i.e., Zhang et al. 2014; Hahn and Nakhleh 2016). Importantly, the distances applied here can be used to investigate the effects of such uncertainty on character trait distributions and models (i.e., Fig. 7).

There are a number of limitations to our applications of probabilistic distances explored in this study. We primarily focused on the Hellinger distance, because closed-form solutions exist for this distance that is also a true metric bounded by zero (identical models) and one (maximum divergence between models), which proved useful throughout our demonstrations. We also applied the Kullback–Leibler divergence in two scenarios (Figs. 3 and S8), which is also of interest given its connections

with model selection, as several information-theoretic approaches, such as AIC, are designed to approximate this divergence between a fitted model and a true underlying probability generating model (Akaike 1973). It is likely that other distances may prove useful in this context, such as the Fréchet (Dowson and Landau 1982) and Bhattacharyya (Bhattacharyya 1943) distances. We also examined only a handful of different models, and there is now a wealth of models that may prove useful for future distance computations (e.g., Slater 2013; Puttick 2018; Guerrero and Hahn 2018; Mendes et al. 2018). From this perspective, the distances discussed in this study provide a flexible foundation for incorporating new models and novel phylogenetic distances. As a practical consideration, we also note that the computational complexity of computing probabilistic distances may be higher than more traditional approaches that only consider topologies and/or branch lengths alone (Fig. S5). Collectively, we have shown a number of insightful uses of this approach for contrasting phylogeny-aware models with one another and believe that our distance framework advances the toolkit for future studies seeking to understand the evolutionary history of organisms and their traits.

SOFTWARE AVAILABILITY

The R package PRDATR (**PR**obabilistic **D**istances under models of **A**daptive **T**rait evolution in **R)** was written in R v3.6.1 and is available on Github at: github.com/radamsRHA/PRDATR. PRDATR includes a number of functions for computing probabilistic model distances for bifurcating trees, hybridization networks, and an array of continuous trait models, and it also provides scripts for replicating experiments demonstrated in this study.

FUNDING AND ACKNOWLEDGEMENTS

REFERENCES

Abou-Moustafa K.T., Ferrie F.P. 2012. A note on metric properties for some divergence measures: The Gaussian case. J. Mach. Learn. Res. 15:1–15.

Adams R.H., Castoe T.A. 2019a. Statistical binning leads to profound model violation due to gene tree error incurred by trying to avoid gene tree error. Mol. Phylogenet. Evol. 134:164–171.

Adams R.H., Castoe T.A. 2019b. Probabilistic Species Tree Distances: Implementing the Multispecies Coalescent to Compare Species Trees Within the Same Model-Based Framework Used to Estimate Them. Syst. Biol. 61:194–207.

Akaike H. 1973. Information theory as an extension of the the maximum likelihood principle. 2nd Int. Symp. Inf. Theory. 267–281.

Aldous, D., 1996. Probability distributions on cladograms. In Random discrete structures. Springer, New York, NY. 1–18.

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

Bawa K.S., Ingty T., Revell L.J., Shivaprakash K.N. 2018. Correlated evolution of flower size and seed number in flowering plants (monocotyledons). Ann. Bot. 123:181–190.

Bastide, P., Solís-Lemus, C., Kriebel, R., William Sparks, K. Ané, C., 2018. Phylogenetic comparative methods on phylogenetic networks with reticulations. Syst. Biol. 67: 800–820.

Bhattacharyya A. 1943. On A Measure of Divergence Between Two Statistical Populations Defined

by their Probability Distributions. Bull. Calcutta Methematical Soc. 35:99–109.

Billera L.J., Holmes S.P., Vogtmann K. 2001. Geometry of the space of phylogenetic trees. Adv. Appl. Math. 27:733–767.

Blair C., Ané C. 2019. Phylogenetic Trees and Networks Can Serve as Powerful and Complementary Approaches for Analysis of Genomic Data. Syst. Biol. 69:593–601.

Blomberg S.P., Garland T., Ives A.R. 2003. Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. Evolution. 57:717–745.

Bortolussi, N., Durand, E., Blum, M., François, O., 2006. apTreeshape: statistical analysis of phylogenetic tree shape. Bioinformatics. 22:363–364.

Butler M.A., King A.A. 2004. Phylogenetic comparative analysis: A modeling approach for adaptive evolution. Am. Nat. 164:683–695.

Cavalli-Sforza L.L., Edwards A.W. 1967. Phylogenetic analysis. Models and estimation procedures. Am. J. Hum. Genet. 21:550–570.

Cavender J.A. 1978. Taxonomy with confidence. Math. Biosci. 40:271–280.

Chira A.M., Thomas G.H. 2016. The impact of rate heterogeneity on inference of phylogenetic models of trait evolution. J. Evol. Biol. 29:2502–2518.

Clavel, J., Escarguel, G. Merceron, G. 2015. mvMORPH: an R package for fitting multivariate evolutionary models to morphometric data. Methods Ecol. Evol. 6:1311–1319.

Colijn C., Plazzotta G. 2018. A Metric on Phylogenetic Tree Shapes. Syst. Biol. 67:113–126.

Davis, C.C., Latvis, M., Nickrent, D.L., Wurdack, K.J., Baum, D.A., 2007. Floral gigantism in Rafflesiaceae. Science. 315:1812.

Dayhoff M., Schwartz R., Orcutt B. 1978. "A model of evolutionary change in proteins". Atlas of protein sequence and structure. National Biomedical Research Foundation Silver Spring MD. 5:345–352.

Degnan J.H., Rosenberg N.A. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 6:332–340.

Dowson D.C., Landau B. V. 1982. The Fréchet distance between multivariate normal distributions. J. Multivar. Anal. 12:450–455.

Drummond, A.J., Suchard, M.A., 2010. Bayesian random local clocks, or one rate to rule them all. BMC Biol., 8:1–12.

Duchi J. 2007. Derivations for linear algebra and optimization. Berkeley, California, 3:2325–5870.

Eastman, J.M., Alfaro, M.E., Joyce, P., Hipp, A.L., L.J., 2011. A novel comparative method for identifying shifts in the rate of character evolution on trees. Evolution. 65:3578–3589.

Edwards S. V., Xi Z., Janke A., Faircloth B.C., McCormack J.E., Glenn T.C., Zhong B., Wu S., Lemmon E.M., Lemmon A.R., Leaché A.D., Liu L., Davis C.C. 2016. Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. Mol. Phylogenet. Evol. 94:447–462.

Estabrook G.F., Mc Morris F.R., Meacham C.A. 1985. Comparison of undirected phylogenetic trees based on subtrees of four evolutionary units. Syst. Zool. 34:193–200.

Farris J.S. 1973. A probability model for inferring evolutionary trees. Syst. Zool. 22:250–256.

Felsenstein J. 1973. Maximum likelihood estimation of evolutionary trees from continuous characters. Am. J. Hum. Genet. 25:471.

Felsenstein J. 1985. Phylogenies and the comparative method. Am. Nat. 125:1–15.

Garba M.K., Nye T.M.W., Boys R.J. 2018. Probabilistic Distances between Trees. Syst. Biol. 67:320–327.

Garba, M.K., Nye, T.M., Lueg, J., Huckemann, S.F. 2020. Information geometry for phylogenetic trees. arXiv:2003, https://arxiv.org/abs/2003.130044.

Guerrero R.F., Hahn M.W. 2018. Quantifying the risk of hemiplasy in phylogenetic inference. Proc.

Natl. Acad. Sci. 115:12787–12792.

Hahn M.W., Nakhleh L. 2016. Irrational exuberance for resolved species trees. Evolution. 70:7–17.

Hansen T.F. 1997. Stabilizing Selection and the Comparative Analysis of Adaptation. Evolution. 51:1341–1351.

Harmon L.J., Losos J.B., Jonathan Davies T., Gillespie R.G., Gittleman J.L., Bryan Jennings W., Kozak K.H., McPeek M.A., Moreno-Roark F., Near T.J., Purvis A., Ricklefs R.E., Schluter D., Schulte J.A., Seehausen O., Sidlauskas B.L., Torres-Carvajal O., Weir J.T., Mooers A.T. 2010. Early bursts of body size and shape evolution are rare in comparative data. Evolution. 64:2385–2396.

Harmon L.J., Weir J.T., Brock C.D., Glor R.E., Challenger W. 2007. GEIGER: investigating evolutionary radiations. Bioinformatics. 24:129–131.

Ho, L.S.T., Ané, C. 2014. Intrinsic inference difficulties for trait evolution with Ornstein-Uhlenbeck models. Methods Ecol. Evol. 5:1133–1146.

Hua X., Lanfear R. 2018. The influence of non-random species sampling on macroevolutionary and macroecological inference from phylogenies. Methods Ecol. Evol. 9:1353–1362.

Huelsenbeck, J.P., Larget, B., Alfaro, M.E. 2004. Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. Mol. Biol. Evol. 21:1123–1133.

Huelsenbeck J.P., Rannala B., Masly J.P. 2000. Accommodating phylogenetic uncertainty in evolutionary studies. Science. 288:2349–2350.

Ives A.R., Midford P.E., Garland T. 2007. Within-species variation and measurement error in phylogenetic comparative methods. Syst. Biol. 56:252–270.

Jarvis E.D., Mirarab S., Aberer A.J., Li B., Houde P., Li C., Ho S.Y.W., Faircloth B.C., Nabholz B., Howard J.T., Suh A., Weber C.C., Da Fonseca R.R., Li J., Zhang F., Li H., Zhou L., Narula N., Liu L., Ganapathy G., Boussau B., Bayzid M.S., Zavidovych V., Subramanian S., Gabaldón T., Capella-Gutiérrez S., Huerta-Cepas J., Rekepalli B., Munch K., Schierup M., Lindow B., Warren

W.C., Ray D., Green R.E., Bruford M.W., Zhan X., Dixon A., Li S., Li N., Huang Y., Derryberry E.P., Bertelsen M.F., Sheldon F.H., Brumfield R.T., Mello C. V., Lovell P. V., Wirthlin M., Schneider M.P.C., Prosdocimi F., Samaniego J.A., Velazquez A.M.V., Alfaro-Núñez A., Campos P.F., Petersen B., Sicheritz-Ponten T., Pas A., Bailey T., Scofield P., Bunce M., Lambert D.M., Zhou Q., Perelman P., Driskell A.C., Shapiro B., Xiong Z., Zeng Y., Liu S., Li Z., Liu B., Wu K., Xiao J., Yinqi X., Zheng Q., Zhang Y., Yang H., Wang J., Smeds L., Rheindt F.E., Braun M., Fjeldsa J., Orlando L., Barker F.K., Jønsson K.A., Johnson W., Koepfli K.P., O'Brien S., Haussler D., Ryder O.A., Rahbek C., Willerslev E., Graves G.R., Glenn T.C., McCormack J., Burt D., Ellegren H., Alström P., Edwards S. V., Stamatakis A., Mindell D.P., Cracraft J., Braun E.L., Warnow T., Jun W., Gilbert M.T.P., Zhang G. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science. 346:1320–1331.

Jhwueng D.C., O'Meara B. 2015. Trait evolution on phylogenetic networks. BioRxiv. https://doi.org/10.1101/023986.

Johnson D.H., Sinanović S. 2001. Symmetrizing the Kullback-Leibler Distance. IEEE Trans. Inf. Theory. 78:96.

Kim, J. 2000. Slicing hyperdimensional oranges: the geometry of phylogenetic estimation. Mol. Phylogenet. Evol. 17:58–75.

Kuhner M.K., Yamato J. 2015. Practical performance of tree comparison metrics. Syst. Biol. 64:205–214.

Lande R. 1976. Natural Selection and Random Genetic Drift in Phenotypic Evolution. Evolution. 30:314–334.

Landis M.J., Schraiber J.G. 2017. Pulsed evolution shaped modern vertebrate body sizes. Proc. Natl. Acad. Sci. 114:13224–13229.

Lewis P.O. 2001. A likelihood approach to estimating phylogeny from discrete morphological

character data. Syst. Biol. 50:913–925.

Liberles D.A. 2007. Ancestral sequence reconstruction. Oxford University Press on Demand.

Liedtke H.C., Gower D.J., Wilkinson M., Gomez-Mestre I. 2018. Macroevolutionary shift in the size of amphibian genomes and the role of life history and climate. Nat. Ecol. Evol. 2:1792.

Lin J. 1991. Divergence Measures Based on the Shannon Entropy. IEEE Trans. Inf. Theory. 37:145–151.

Lin Y., Rajan V., Moret B.M.E. 2012. A metric for phylogenetic trees based on matching. IEEE/ACM Trans. Comput. Biol. Bioinforma. 9:1014–1022.

Liò P., Goldman N. 1998. Review: Models of molecular evolution and phylogeny. Genome Res. 8:1233–1244.

Liu L., Xi Z., Wu S., Davis C.C., Edwards S. V. 2015. Estimating phylogenetic trees from genome-scale data. Ann. N. Y. Acad. Sci. 1360:36–53.

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

Mahler, D.L., Revell, L.J., Glor, R.E., Losos, J.B., 2010. Ecological opportunity and the rate of morphological evolution in the diversification of Greater Antillean anoles. Evolution. 64:2731–2745.

Malmstrøm M., Britz R., Matschiner M., Tørresen O.K., Hadiaty R.K., Yaakob N., Tan H.H., Jakobsen K.S., Salzburger W., Rüber L. 2018. The most developmentally truncated fishes show extensive Hox gene loss and miniaturized genomes. Genome Biol. Evol. 10:1088–1103.

Martins E.P. 1994. Estimating the rate of phenotypic evolution from comparative data. Am. Nat. 144:193–209.

Mendes F.K., Fuentes-González J.A., Schraiber J.G., Hahn M.W. 2018. A multispecies coalescent model for quantitative traits. Elife. 7:e36482.

Mirarab S., Bayzid M.S., Boussau B., Warnow T. 2014. Statistical binning enables an accurate coalescent-based estimation of the avian tree. Science. 346:1250463.

Mitov, V., Bartoszek, K.,Stadler, T., 2019. Automatic generation of evolutionary hypotheses using mixed Gaussian phylogenetic models. Proc. Natl. Acad. Sci. 116:16921–16926.

Mitov, V., Bartoszek, K., Asimomitis, G., Stadler, T., 2020. Fast likelihood calculation for multivariate Gaussian phylogenetic models with shifts. Theor Popul Biol. 131:66–78.

Moulton, V., Steel, M. 2004. Peeling phylogenetic 'oranges'. Adv. Appl. Math. 33:710–727.

Nee S., May R.M., Harvey P.H. 1994. The reconstructed evolutionary process. Philos. Trans. R. Soc. B Biol. Sci. 344:305–311.

Neyman J. 1971. Molecular studies of evolution: a source of novel statistical problems. Statistical decision theory and related topics. Elsevier. 1–27.

Nielsen, F., 2019. On the Jensen–Shannon Symmetrization of Distances Relying on Abstract Means. Entropy. 21:485.

Nunn C.L. 2013. The Comparative Approach in Evolutionary Anthropology and Biology. University of Chicago Press.

O'Meara B.C. 2012. Evolutionary Inferences from Phylogenies: A Review of Methods. Annu. Rev. Ecol. Evol. Syst. 43:267–285.

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

O'Meara B.C., Beaulieu J.M. 2016. Past, future, and present of state-dependent models of diversification. Am. J. Bot. 103:792–795.

Pagel M. 1994. Detecting correlated evolution on phylogenies: A general method for the comparative analysis of discrete characters. Proc. R. Soc. B Biol. Sci. 255:37–45.

Pagel M. 1999a. Inferring the historical patterns of biological evolution. Nature. 401:877.

Pagel M. 1999b. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. Syst. Biol. 48:612–622.

Pardo L. 2005. Statistical inference based on divergence measures. Chapman and Hall/CRC, Boca Raton, FL.

Pennell M.W., Harmon L.J. 2013. An integrative view of phylogenetic comparative methods: Connections to population genetics, community ecology, and paleobiology. Ann. N. Y. Acad. Sci. 1289:90–105.

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

Penny D., Watson E.E., Steel M.A. 1993. Trees from languages and genes are very similar. Syst. Biol. 42:382–384.

Puttick M.N. 2018. Mixed evidence for early bursts of morphological evolution in extant clades. J. Evol. Biol. 31:502–515.

Pyron R.A. 2014. Biogeographic analysis reveals ancient continental vicariance and recent oceanic dispersal in amphibians. Syst. Biol. 63:779–797.

Reddy S., Kimball R.T., Pandey A., Hosner P.A., Braun M.J., Hackett S.J., Han K.L., Harshman J., Huddleston C.J., Kingston S., Marks B.D., Miglia K.J., Moore W.S., Sheldon F.H., Witt C.C., Yuri T., Braun E.L. 2017. Why do phylogenomic data sets yield conflicting trees? Data type influences the avian tree of life more than taxon sampling. Syst. Biol. 66:857–879.

Revell L.J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3:217–223.

Revell L.J. 2014. Ancestral character estimation under the threshold model from quantitative genetics. Evolution. 68:743–759.

Revell L.J., Harmon L.J. 2008. Testing quantitative genetic hypotheses about the evolutionary rate matrix for continuous characters. Evol. Ecol. Res. 10:311–331.

Revell L.J., Mahler D.L., Sweeney J.R., Sobotka M., Fancher V.E., Losos J.B. 2010. Nonlinear selection and the evolution of variances and covariances for continuous characters in an anole. J. Evol. Biol. 23:407–421.

Robinson D.F., Foulds L.R. 1979. Comparison of weighted labelled trees. In Combinatorial Mathematics VI (Horadam A.F., and Wallis W.D eds). Berlin, Springer. pp119–126.

Rohlf F.J. 2001. Comparative methods for the analysis of continuous variables: Geometric interpretations. Evolution. 55:2143–2160.

Ronquist F. 1997. Phylogenetic approaches in coevolution and biogeography. Zool. Scr. 26:313–322.

Schliep K.P. 2011. phangorn: Phylogenetic analysis in R. Bioinformatics. 27:592–593.

Schluter, D., Price, T., Mooers, A.Ø., Ludwig, D. 1997. Likelihood of ancestor states in adaptive radiation. Evolution. 51:1699–1711.

Slater G.J. 2013. Phylogenetic evidence for a shift in the mode of mammalian body size evolution at the Cretaceous-Palaeogene boundary. Methods Ecol. Evol. 4:734–744.

Tavaré S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. Am. Math. Soc. Lect. Math. Life Sci. 17:57–86.

Uyeda, J.C., Caetano, D.S., Pennell, M.W., 2015. Comparative analysis of principal components can be misleading. Syst. Biol. 64:677–689.

Uyeda, J.C., Harmon, L.J., 2014. A novel Bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. Syst. Biol., 63:902–918.

Uyeda, J.C., Zenil-Ferguson, R. Pennell, M.W., 2018. Rethinking phylogenetic comparative methods. Syst. Biol. 67:1091–1109.

Watanabe A., Slice D.E. 2014. The utility of cranial ontogeny for phylogenetic inference: A case study

in crocodylians using geometric morphometrics. J. Evol. Biol. 27:1078–1092.

Yahara K., Didelot X., Ansari M.A., Sheppard S.K., Falush D. 2014. Efficient inference of recombination hot regions in bacterial genomes. Mol. Biol. Evol. 31:1593–1605.

Yule G. 1925. A mathematical theory of evolution, based on the conclusions of Dr. JC Willis, FRS. Philos. Trans. R. Soc. London Ser. B. 213:21–87.

Zhu, S., Degnan, J.H. 2017. Displayed trees do not determine distinguishability under the network multispecies coalescent. Syst. Biol. 66:283–298.

TABLE LEGENDS

TABLE 1. Summary of the six models used to demonstrate the properties of probabilistic phylogenetic distances under models of continuous trait evolution. Asterisk (*) denotes the particular scaled parameters used in simulation analyses.

FIGURE CAPTIONS

FIGURE 1. Conceptual schematic depicting an example set of distance computations for a simple phylogenetic model with $n = 2$ taxa (top left). Coupled with a particular model (i.e., Brownian motion [BM], Ornstein-Uhlenbeck [OU], or early-burst [EB]), this phylogenetic tree model provides a variance-covariance matrix that is scaled by model parameters. In this example, the first model φ_1 (lower left) represents a standard BM model with $\sigma^2 = 1$, and there are three alternative models possible for φ_2: BM($\sigma^2 = 1$), OU($\sigma^2 = 1, \alpha = 1$), or EB($\sigma^2 = 1, a = 1$). For each model under this phylogenetic scenario, the probability distribution of trait values $\vec{\chi}$ sampled at the tips can be formulated as a bivariate (i.e., $n = 2$) normal distribution, which is depicted by each respective model as a heatmap overlaid by a contour plot, with darker colors representing higher probabilities. Distances are computed by comparing these bivariate normal distributions with one another (arrows from φ_1 to each φ_2 indicate pairs of model distances to be computed).

FIGURE 2. Probabilistic phylogenetic distances under models of discrete trait evolution computed across a range of scaling values. (a) Symmetric topology phylogenetic tree with $n = 8$ taxa that continuous trait evolutionary models are condition on. (b) Hellinger distances (d_H) computed using the tree in panel (a) for Brownian motion (BM), Ornstein-Uhlenbeck (OU), early-burst (EB), Pagel's lambda (L), Pagel's kappa (K), and Pagel's delta (D) continuous trait models, with the first model representing a standard BM model with $\sigma^2 = 1$, and the respective parameters of the second model

scaled by γ. See Table 1 for description of each model and scaled parameters. (c) Hellinger distances computed using the tree in panel (a) for BM, OU, EB, L, K, and D models, with the first model representing a standard BM model with $\sigma^2 = 2$, and the respective parameters of the second model scaled by γ.

FIGURE 3. Hellinger distance (d_H) and Kullback–Leibler divergence (d_{KL}) between a pair of hybridization networks (a) shown in (b), or between a bifurcating tree and a hybridization network (c) shown in (d), which were computed across a range of values for either the evolutionary rate parameter σ^2 (i.e., $d_H(\sigma^2\gamma)$) or the migration proportion m (i.e., $d_H(m\gamma)$) for two BM models.

FIGURE 4. The synergistic influence of tree shape, taxa number, and evolutionary model parameter on probabilistic distances. Results shown for the Hellinger distance (d_H) computed between a Brownian motion (BM) model and either the Ornstein-Uhlenebck (OU; panels a-c), early-burst (EB; panels d-f), or Pagel's delta (D; panels g-i) model for simulations using different numbers of taxa on three different tree shapes: "balanced" (left column), "left unbalanced" (center column), and "star" (right column). Branch lengths are chosen such that the total tree height is scaled to 1.0. For each plot, the particular parameter values are indicated with arrows pointing to the specific lines, such that each line represents a different parameter value on a log-scale from 0.01 to 10.0.

FIGURE 5. Investigating the relationship between model distances and the significance of likelihood ratio tests between fitted Brownian motion (BM) and Ornstein-Uhlenbeck (OU) models (traits simulated under an OU model). Results shown for three different tree shapes: "balanced" (left panels), "left unbalanced" (center), and trees simulated under a Yule model with birth rate $\lambda = 1$ (right) with equal branch lengths that are scaled to give a total tree height of 1.0. *P*-values for a likelihood ratio test

comparing the OU and BM models as a function of their Hellinger distance (d_H) are shown for three different tree sizes: 128 (panels a-c), 512 (panels d-f), and 1024 (panels g-i) tips. The mean (circle) and standard deviations (bars) of the distribution of 10 replicate p-values (subtracted from one). Each simulation replicate was computed by incrementally increasing the α parameter of the OU model from 0 to 5 (from left to right in each panel colored in the blue scale shown), at increments of 0.01.

FIGURE 6. Computing probabilistic Hellinger distances (d_H) between the Brownian motion (BM), Ornstein-Uhlenbeck (OU), early-burst (EB), Pagel's lambda (L), Pagel's kappa (K), and Pagel's delta (D) continuous trait models that were fit to the amphibian genome size dataset of $n = 465$ taxa. Graphical network showing the six models (BM, OU, EB, L, K, and D) as nodes connected by edges, with the widths of edges scaled by their respective probabilistic distances (shown beside each edge).

FIGURE 7. Multidimensional scaling (MDS) based on pairwise Hellinger distances (d_H) estimated assuming a Brownian motion (BM) model (a) or an Ornstein-Uhlenbeck (OU) model (c) for a set of 6144 avian gene trees that comprise 2136 exons (dark gray), 329 introns (black), and 3679 UCEs (light gray). Analogous, plots (b) and (d) depict pairwise distances projected using MDS for the 31 avian species trees assuming BM (b) or OU (d) models, respectively.

FIGURE 8. Applying the Hellinger distance (d_H) to multivariate models of continuous trait evolution. Results for simulation analyses using the phylogeny depicted in panel (a) are shown in panel (b), where ρ is the covariance between the two traits. Phylogeny depicting "Felsenstein's worst case" scenario is shown in panel (c), which was used to simulate datasets in which an instantaneous shift occurs on one of the ancestral branches (location of shift depicted as a tick mark on the tree in panel (c)), and results shown in panel (d) with $\log_{10}$ ratio of the shift to Brownian motion (BM) variance (x-axis) and the
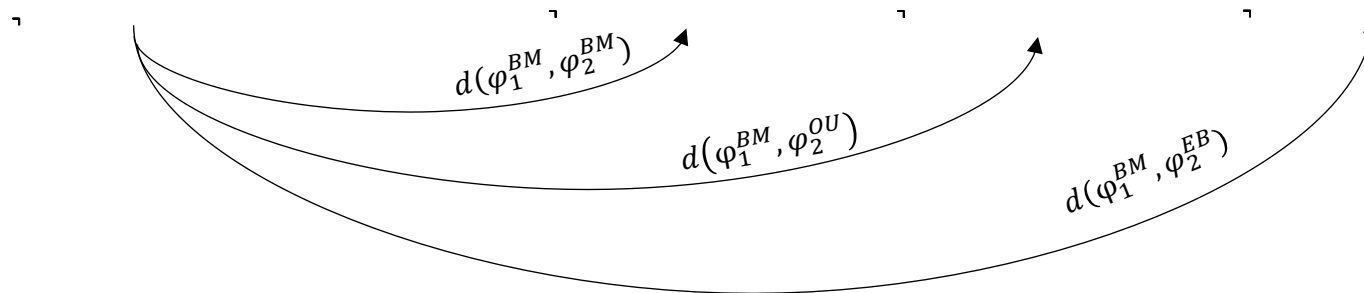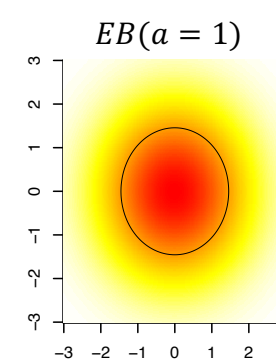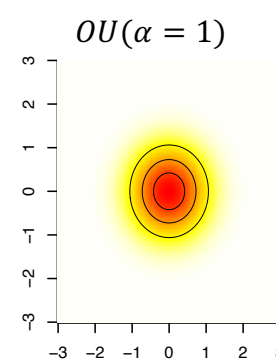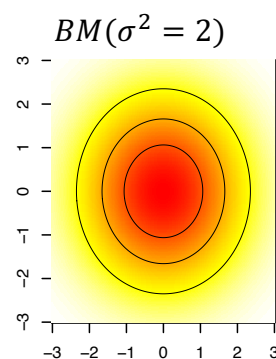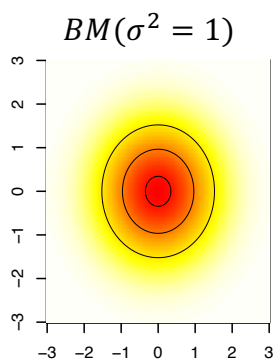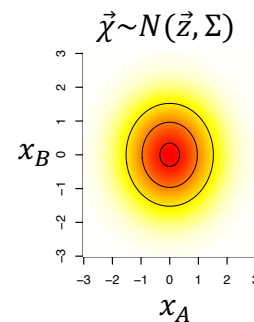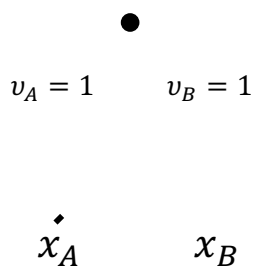
Hellinger distance (*y*-axis) computed between the unconstrained and constrained models that have been fit to the simulated data. Color of points in panels (b) and (c) indicate 1–*p*-value of the likelihood ratio test between an unconstrained model (i.e., ρ is estimated) and a constrained model (i.e., $\rho = 0$ such that traits are assumed to be independent) that have been fit to the data.
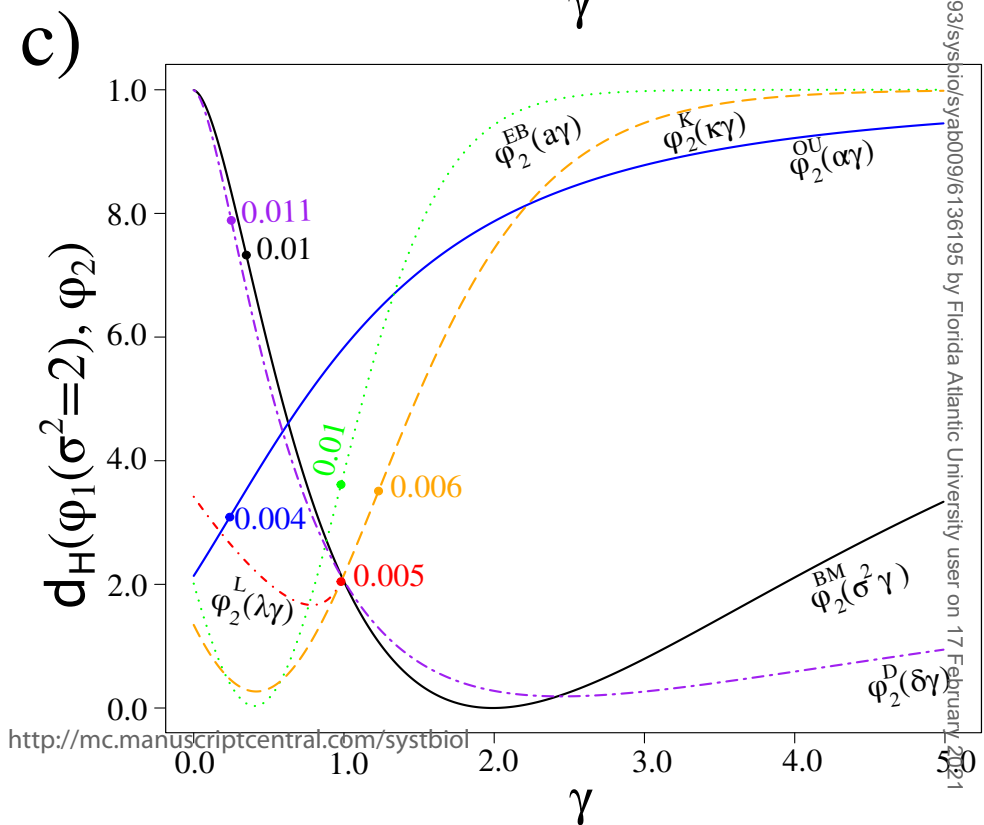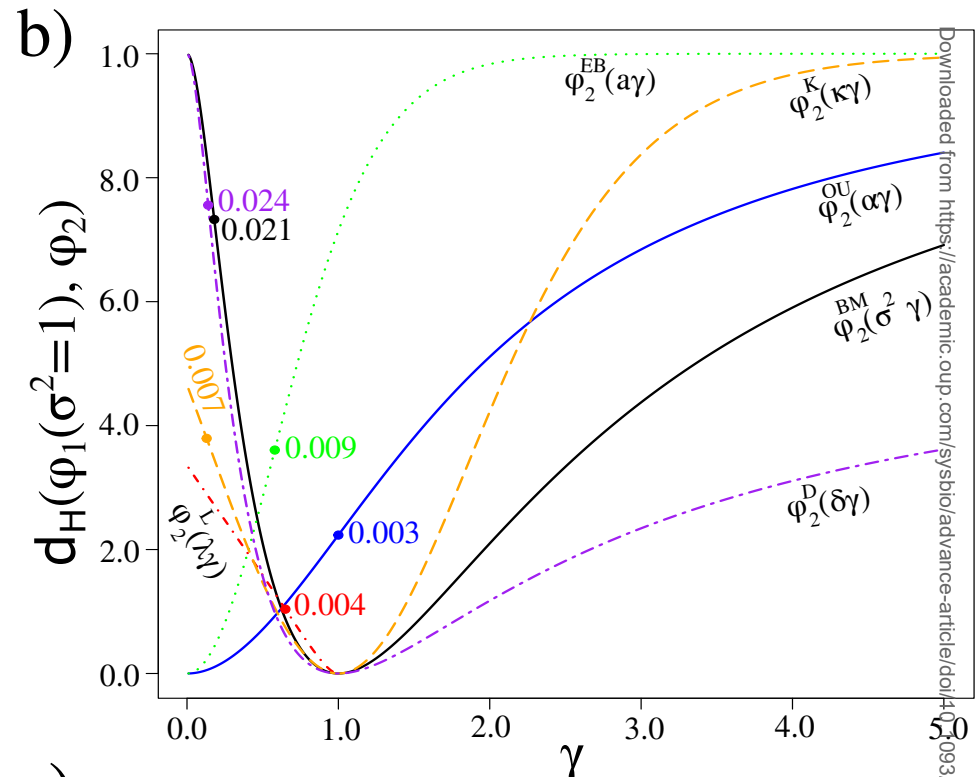
FIGURE 9. Investigating identifiability of mixed Ornstein-Uhlenbeck (OU) models using the Hellinger distance (d_H). Asterisks (*) indicate the location of shift points for OU model parameters in the tree pairs shown in panels (a), (c) and (e). Heatmap shown in panel (b) represents the Hellinger distance computed between the left and right tree models displayed in panel (a) across a range of values for the ancestral state θ and the background optimum θ_{opt} using a shift optimum $\theta_{shift} = \theta$ of the right tree, while using $\theta = 1$, $\theta_{shift} = 2$, and $\theta_{opt} = 1$ for the left tree in panel (a). Panel (d) depicts the distance between the two tree models shown in panel (c) across a range of σ^2 and α parameter values of the OU model marked with a gray asterisk in the left tree of panel (c). Similarly, results for the Hellinger distance between the two tree models displayed in panel (e) are shown in panel (f), with a range of σ^2 and α parameter values for the OU model represented by a gray asterisk in the right tree of panel (e).

Model	Abbreviation	Model parameters	Interpretation
Brownian motion	BM	σ^{2} *	Evolutionary rate
Ornstein-Uhlenbeck	OU	σ^{2},α *	Pull toward optimum
Early burst	EB	σ^{2},a *	Rate acceleration (positive) or deceleration (negative)
Pagel's lambda	L	σ^{2},λ *	Tree is star-like when λ=0
Pagel's kappa	K	σ^{2},κ *	Raise branches to the power κ
Pagel's delta	D	σ^{2},δ *	Raise node depths to the power δ

TABLE 1. Summary of the six models used to demonstrate the properties of probabilistic phylogenetic distances under models of continuous trait evolution. Asterisk (*) denotes the particular scaled parameters used in simulation analyses.

Phylogeny for both φ_1 and φ_2

●

$v_A = 1$ $\quad$ $v_B = 1$

$\dot{x}_A$ $\qquad$ x_B

$\vec{\chi} \sim N(\vec{z}, \Sigma)$

$BM(\sigma^2 = 1)$ $\qquad$ $BM(\sigma^2 = 2)$ $\qquad$ $OU(\alpha = 1)$ $\qquad$ $EB(a = 1)$

$d(\varphi_1^{BM}, \varphi_2^{BM})$

$d(\varphi_1^{BM}, \varphi_2^{OU})$

$d(\varphi_1^{BM}, \varphi_2^{EB})$

a)

$\varphi 1$
$(\sigma^2=1, m=0.5)$

$\varphi 2$
$(\sigma^2=1, m=0.5\gamma)$
or
$(\sigma^2=1\gamma, m=0.5)$

b)

c)

$\varphi 1$
$(\sigma^2=1, m=0.5)$

$\varphi 2$
$(\sigma^2=1, m=0.5\gamma)$
or
$(\sigma^2=1\gamma, m=0.5)$

d)
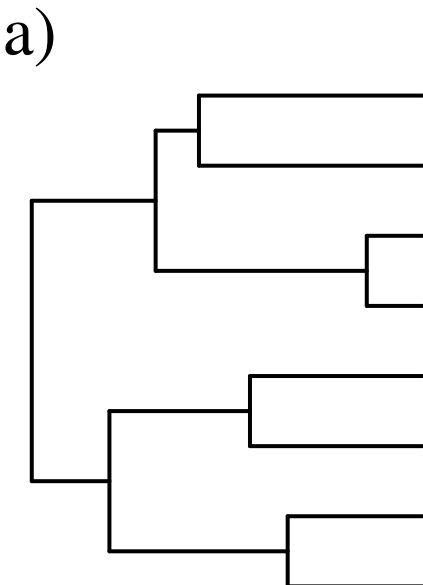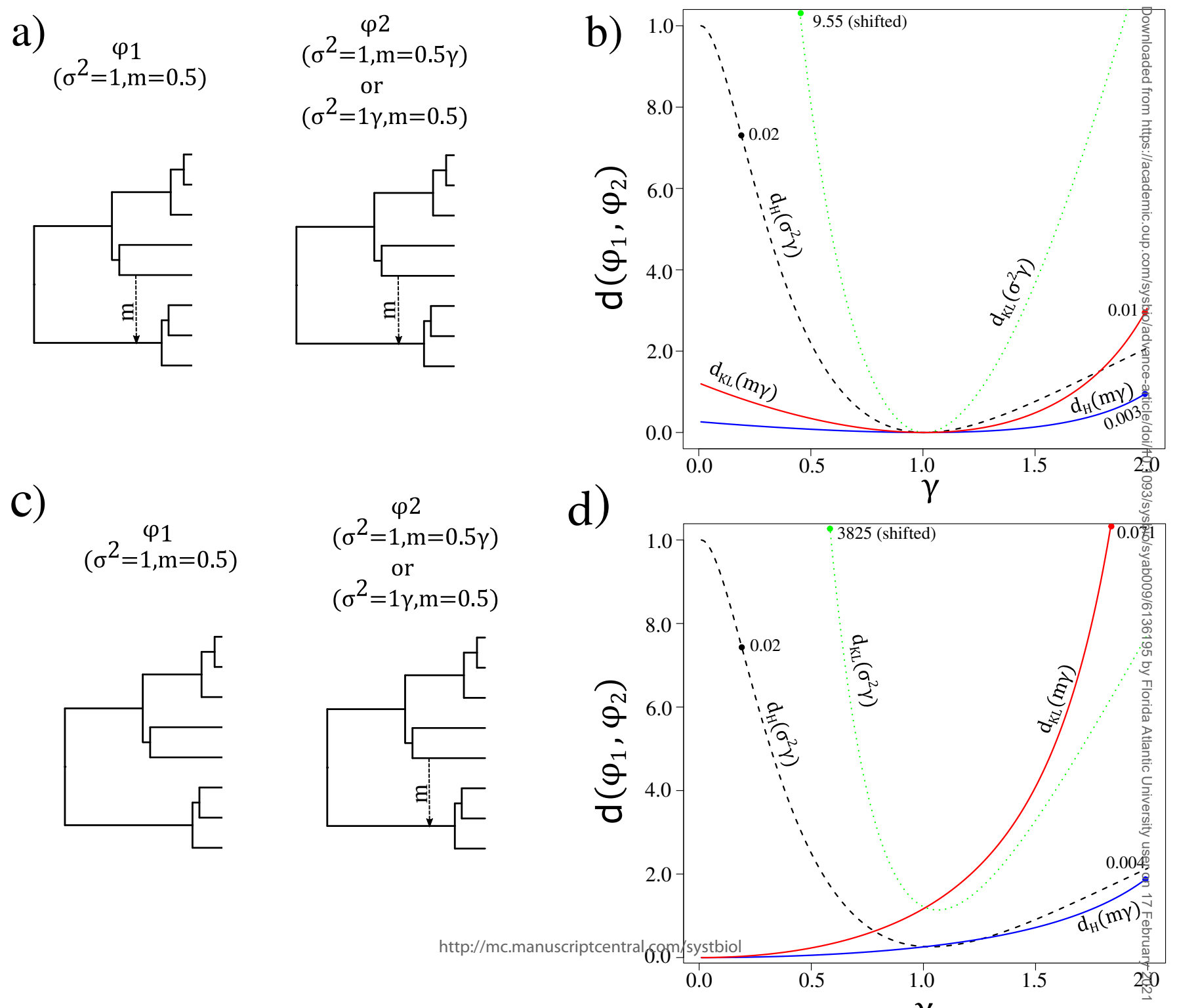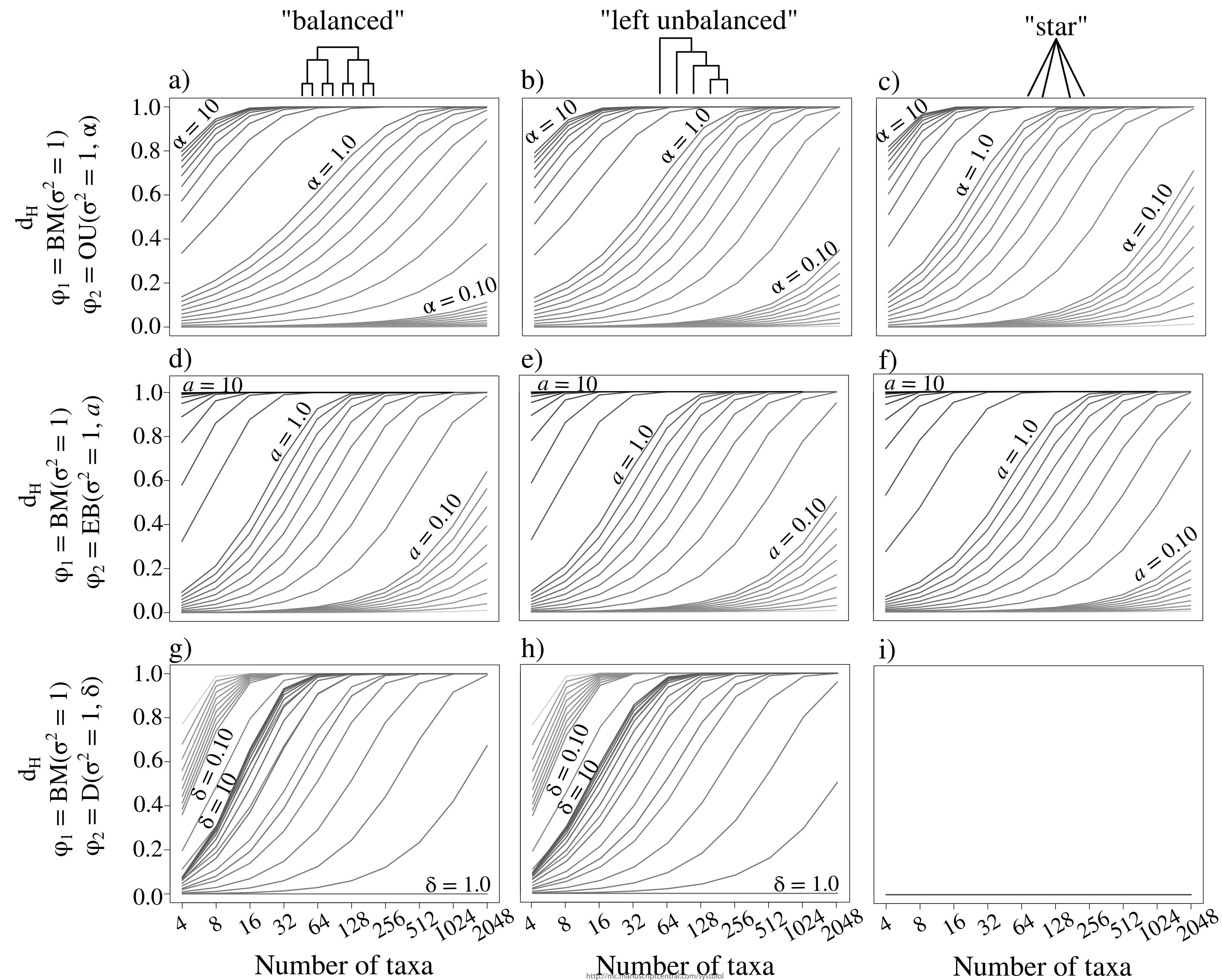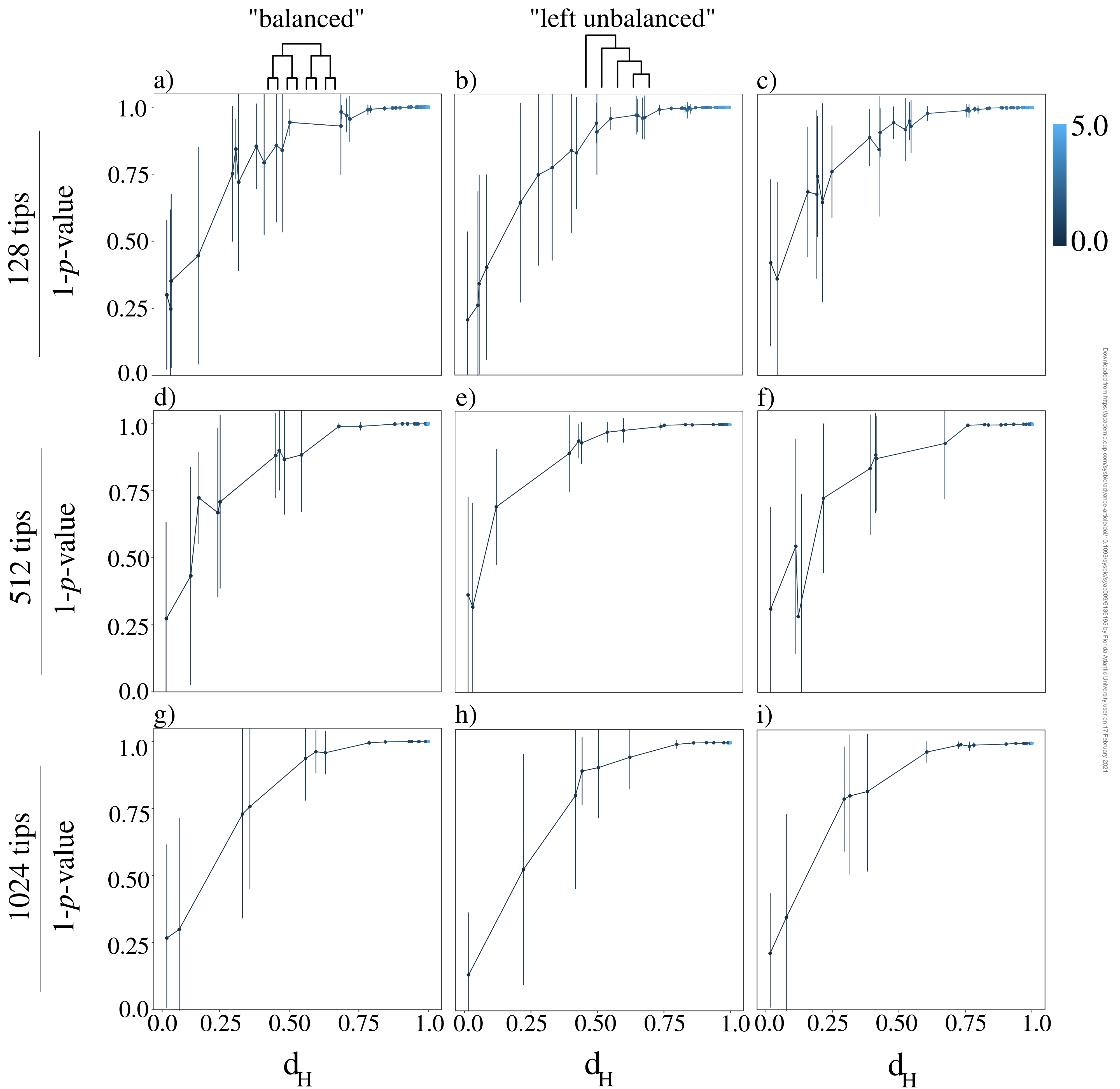
a) BM for 6144 gene trees

b) BM for 31 species trees

- Exon.c12.ExaML.unpartitioned
- Exon.c12.ExaML.partitioned
- Exon.c1.ExaML.unpartitioned
- Intron.MP−EST.binned
- Exon.c2.ExaML.unpartitioned
- Intron.RAxML.Heterogenous.c123−RY.ExaML.unpartitioned
- Exon.AminoAcid.ExaML.partitioned23
- Exon.RAxML.Homogenous.c123
- TENT.MP−EST.binned
- Exon.RAxML.unpartitioned
- Literature.DNAxDNA.SibleyAhlquist
- Exon.RAxML.Homogenous.c12
- Exon.c3.ExaML.unpartitioned
- TEIT.RAxML
- Exon.RAxML.Heterogenous.c123
- Exon.c123.ExaML.partitioned
- Intron.RAxML.unpartitioned
- Literature.Mitochondrial.Pacheco
- TENT.ExaML.25% TENT.c3.ExaML
- TENT.ExaML
- TENT.ExaML.75%
- UCE.RAxML.unpartitioned
- Literature.Nuclear_genes.Hackett
- TENT.MP−EST.unbinned
- Intron.MP−EST.unbinned
- TENT.ExaML.50%
- Intron.RAxML.partitioned

c) OU for 6144 gene trees

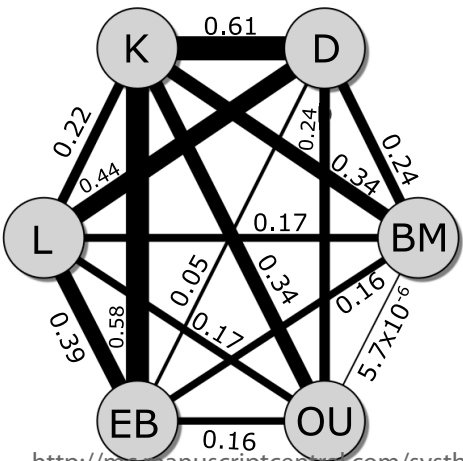d) OU for 31 species trees

- Literature.DNAxDNA.SibleyAhlquist
- Intron.MP−EST.binned
- Intron.RAxML.Heterogenous
- Intron.RAxML.Homogenous
- Exon.c123.ExaML.unpartitioned
- Intron.RAxML.unpartitioned
- TENT.ExaML
- TENT.ExaML.25%
- TENT.ExaML.50%
- TENT.ExaML.75%
- Exon.c123−RY.ExaML.unpartitioned
- TEIT.RAxML
- Exon.c12.ExaML.partitioned
- Exon.c12.ExaML.unpartitioned
- Exon.c2.ExaML.unpartitioned
- UCE.RAxML.unpartitioned
- TENT.MP−EST.binned
- TENT+c3.ExaML
- Exon.RAxML.Homogenous.c12
- Exon.c3.ExaML.unpartitioned
- Literature.Nuclear_genes.Hackett
- Intron.MP−EST.unbinned
- TENT.MP−EST.unbinned
- Intron.RAxML.partitioned
- Exon.c1.ExaML.unpartitioned
- Exon.RAxML.Homogenous.c123
- Exon.AminoAcid.ExaML.partitioned
- Literature.Mitochondrial.Pacheco
- Exon.RAxML.Heterogenous.c123
- Exon.RAxML.Heterogenous.c12
- Exon.c123.ExaML.partitioned

coordinate 1

coordinate 2

a)

$\theta_{shift} = \theta$

b)

d_H

$\theta_{opt} \neq \theta$

c)

d)

d_H

σ^2

α

e)

f)

d_H

σ^2

α