Scalable total-evidence inference from molecular and continuous characters in a Bayesian framework

Rong Zhang[1,2], Alexei J. Drummond[1,2,3], and Fábio K. Mendes[1,3*]

[1] *Centre for Computational Evolution, The University of Auckland, 1010, New Zealand*
[2] *School of Computer Science, The University of Auckland, Auckland, 1010, New Zealand*
[3] *School of Biological Sciences, The University of Auckland, Auckland, 1010, New Zealand*
**Correspondence to be sent to: School of Biological Sciences, The University of Auckland, 3A Symonds St., Auckland, New Zealand;*
E-mail: f.mendes@auckland.ac.nz

Abstract

Time-scaled phylogenetic trees are both an ultimate goal of evolutionary biology and a necessary ingredient in comparative studies. While accumulating genomic data has moved the field closer to a full description of the tree of life, the relative timing of certain evolutionary events remains challenging even when this data is abundant, and absolute timing is impossible without external information such as fossil ages and morphology. The field of phylogenetics lacks efficient tools integrating probabilistic models for these kinds of data into unified frameworks for estimating phylogenies. Here, we implement, benchmark and validate popular phylogenetic models for the study of paleontological and neontological continuous trait data, incorporating these models into the BEAST2 platform. Our methods scale well with number of taxa and of characters. We tip-date and estimate the topology of a phylogeny of Carnivora, comparing results from different configurations of integrative models capable of leveraging ages, as well as molecular and continuous morphological data from living and extinct species. Our results illustrate and advance the paradigm of Bayesian, probabilistic total evidence, in which explanatory models are fully defined, and inferential uncertainty in all their dimensions is accounted for.

[Continuous trait, Brownian motion, total evidence, Carnivora]

The advent of molecular sequencing has unquestionably revolutionized comparative biology, giving phylogeneticists unprecedented power to recover species relationships and date important evolutionary events (e.g., Jarvis et al., 2014; Zhang et al., 2014; Suh et al., 2015; Pease et al., 2016; Kawahara et al., 2019; Vanderpool et al., 2020), describe drivers of diversification (Condamine et al., 2013; Morlon, 2014; Sánchez-Reyes et al., 2017; Condamine et al., 2019), and their relationship with ecologically relevant traits (Goldberg and Igić, 2012; Burin et al., 2016; de Alencar et al., 2017). The accumulation of genomic data further allowed the identification of problems or gaps in molecular evolution models (or their usage; e.g., Sullivan and Swofford 1997; Kolaczkowski and Thornton 2004; Mendes and Hahn 2018), which led to improvements in their realism (Yang, 2006; Rannala and Yang, 2003; Degnan and Salter, 2007), as well as the development of a plethora of computational tools for empiricists wishing to use such models (e.g., Lartillot and Philippe, 2004; Stamatakis, 2014; Nguyen et al., 2015; Chifman and Kubatko, 2015; Höhna et al., 2016; Zhang et al., 2018; Suchard et al., 2018; Bouckaert et al., 2019).

Despite all progress, abundant genomic sequences and more complex substitution models have not been a panacea for phylogenetic studies, in which species trees measured in absolute time are either the ultimate goal (Philippe et al., 2011) or a critical ingredient for downstream analyses (Felsenstein, 1985; Uyeda et al., 2018). First, while molecular data informs us on the relative timing of evolutionary events, inferring mutation rates remains challenging (Kong et al., 2012; Besenbacher et al., 2015; Wang et al., 2020), as does reconciling estimates obtained at different evolutionary timescales (Ho et al., 2005; Penny, 2005; Ho et al., 2007). Second, dating the tree of life in absolute time is complicated by the absence of a universal strict molecular clock (Zuckerkandl and Pauling, 1965; Ayala, 1997; Lanfear et al., 2010). Molecular rates have been shown to vary among loci and species (Li, 1997; Larracuente et al., 2008; Bromham, 2009), and to correlate with phenotypic and natural history traits (Martin and Palumbi, 1993; Smith and Donoghue, 2008), the environment (Bleiweiss, 1998; Wright et al., 2006; Gillman et al., 2009), and

even the process of speciation (Webster et al., 2003; Witt and Brumfield, 2003; Venditti and Pagel, 2010). Finally, although non-contemporaneous DNA can help circumvent the aforementioned issues and improve the estimation of substitution rates and divergence times (Rieux and Balloux, 2016), extracting DNA from well-preserved ancient remains has so far been limited to evolutionary young material. This process is also non-trivial and prone to contamination, usually yielding fragmentary data (Cooper and Poinar, 2000; Hagelberg et al., 2015). These complications often lead to phylogenetic trees being reported in lengths of expected substitutions per site – in these "substitution trees", time and evolutionary rates are conflated.

As a reaction to these findings, the past few decades saw improved descriptions of the substitution process from more realistic clock models (Thorne and Kishino, 2005; Ho and Duchêne, 2014), as well as the development of methods for calibrating substitution trees into time-scaled trees. "Node dating" (as dubbed by Ronquist et al. 2012), for example, refers to a collection of techniques whereby a specialist determines an age (range) for a node using fossil occurrence or biogeographical data (Ho and Phillips, 2009). Node dating is complicated by the difficulty in estimating the age of fossils, choosing which fossils to use (Parham et al., 2012) – in many cases information is lost because younger fossils of a group are excluded in favor of the oldest one – and what nodes to assign them to, and choosing probability distributions for their age ranges, a crucial ad hoc step that can introduce bias and circularity to an analysis (Warnock et al., 2011; Field et al., 2020). These issues are further compounded by the analysis sensitivity to node-time priors (Welch et al., 2005), unclear implicit prior probabilities on node times (Heled and Drummond, 2012), and overly simplistic molecular clock models (Berv and Field, 2018).

As an alternative to node dating, the "tip-dating" approach consists of making direct use of heterochronous data – sample ages and character data – in order to calibrate and place taxa in the phylogeny. Tip dating was first employed for divergence time estimation at shorter time scales, in the context of viral phylodynamics (Rambaut, 2000;

Drummond et al., 2002), where sample times are usually known with good precision and molecular data can be abundant. When used at macroevolutionary time scales, tip dating has also been dubbed "total-evidence" dating (Ronquist et al., 2012), likely as a reference to the original total evidence paradigm proposed by Kluge (1989). As in molecular tip dating, total-evidence dating (Pyron, 2011; Ronquist et al., 2012) allows the data – fossil age estimates and morphological characters – to directly inform fossil affinities and calibrate phylogenies, precluding the somewhat arbitrary specialist input that characterizes node dating. For the purposes of the present study, we use the term "total evidence" to mean "probabilistic" total evidence, the analysis of combined data using integrative probabilistic models, as opposed to methods rooted in parsimony or other heuristics (e.g., Giribet et al., 2001; Nylander et al., 2004; Grant et al., 2006; Manos et al., 2007; Arango and Wheeler, 2007).

The success of total-evidence tip dating depends on the quality and size of morphological data sets (number of characters and phylogenetic coverage), and on how well evolutionary models capture the real processes generating the data, i.e., how good the model fit is. Because obtaining molecular data from extinct species is usually hard, one should strive to obtain as many morphological characters as possible from both extinct and extant species, across and along the phylogeny. Crucially, these species will "link" the phylogenetic signal coming from morphological data together with that coming from molecular sequences, allowing a single phylogeny to be informed by both. Furthermore, evolutionary models should meet a delicate balance between realism, utility, and practicality. By being very realistic, models run the risk of being overly complex, hindering the researcher's ability to draw general, useful conclusions. Very complex models also tend to be computationally onerous and technically hard to implement.

Continuous-time Markov models are routinely used in phylogenetics for the study of both discrete and continuous characters. In the case of discrete traits, the 'Mk' and 'Mkv' models (Lewis, 2001) have received the most attention (e.g., Danforth et al., 2006;

Bracken-Grissom et al., 2014) and criticism (e.g., O'Reilly et al., 2016, 2018; Goloboff et al., 2019). Key issues with these models – or with how they are implemented and normally used – include their assumption that discrete characters evolve in uncorrelated fashion and at the same rate, and the fact that autapomorphic characters are usually not represented in character matrices. Solutions for these problems exist, but can be computationally expensive. Continuous characters, on the other hand, are scored at a resolution that usually makes them variable within and across species. Popular continuous character phylogenetic models are based on Brownian motion (BM; Felsenstein, 1973; Hansen and Martins, 1996; but see Blomberg et al., 2020) and can incorporate correlated evolution among traits, which are assumed to evolve as a random walk whose diffusion rate is the evolutionary rate. Using continuous characters in total-evidence tip dating thus not only has the potential to improve phylogenetic inference by enhancing morphological data sets (Parins-Fukuchi, 2018b; Álvarez Carretero et al., 2019; c.f. Varón-González et al., 2020 for some criticism), but also provides natural workarounds for the issues observed under discrete-character models.

While many computational methods exist for the study of morphological character evolution (e.g., Revell, 2012; Pennell et al., 2014; Clavel et al., 2015; Caetano and Harmon, 2017; Mitov et al., 2020), tools capable of jointly modeling molecular and morphological characters are still lacking, particularly those that simultaneously account for uncertainty in species tree topology and branch lengths. With few exceptions, comparative analysis of morphological characters requires a species tree point estimate (e.g., Adams et al., 2009; Lister, 2013; Gibson and Fuentes-G., 2015; or more rarely, a posterior distribution, e.g., Silvestro et al., 2018; Fuentes-G. et al., 2020) to be available and assumed as the truth. Such species trees will have almost invariably been estimated in previous studies using different data sets, often molecular ones. In such cases, the morphological data is then analyzed on a phylogenetic "Procrustean bed", a species tree that might not represent the morphological evolutionary history (Hahn and Nakhleh, 2016). One way forward should be

easily visible in the joint evolutionary modeling of all available data, whereby different sources of data inform on each other's model parameters and on the phylogeny itself.

Reasons why tools do not provide for this joint evolutionary modeling approach include: (i) the technical difficulty of implementing multiple models efficiently under the same statistical framework, (ii) prohibitively slow run times due to model scalability issues, and (iii) lack of available data sets compiling appropriate data from multiple sources. Over time we expect this last reason to become less of a hindrance and more of a motivation for method development in this area. Recent work suggests, however, there is an immediate demand for methods capable of integrating multidimensional data (e.g., Silvestro et al., 2018; Cascini et al., 2019; Koch and Thompson, 2020), as well as work in progress to meet those demands (e.g., Álvarez Carretero et al., 2019; May and Moore, 2020; Gaboriau et al., 2020; Ogilvie et al., 2021).

Here, we implement and validate efficient, general and scalable methods for phylogenetic inference from continuous characters in a hierarchical Bayesian total-evidence framework as part of the BEAST2 platform (Bouckaert et al., 2019). We also implement a birth-death model that conditions on serially sampled occurrence times (such as fossil ages; Stadler and Yang, 2013) to be used as a species tree sampling distribution in our hierarchical model. By leveraging molecular and morphological data from living and extinct Carnivora species, we then illustrate the use of different integrative model configurations in the estimation of the Carnivora species tree topology and branch lengths, comparing different estimates among themselves and with previously published results.

Materials and Methods

Integrative model

The integrative model for Bayesian total-evidence inference (Fig. 1) can be expressed as the product of the probability density and mass functions of its several component sampling distributions. Given continuous morphology and molecular data

matrices $\mathbf{M}$ and $\mathbf{S}$, respectively, the posterior distribution of the time-scaled phylogenetic tree (Φ), morphological and molecular relative branch rates ($\boldsymbol{b}_m$ and $\boldsymbol{b}_s$) and all remaining parameters ($\boldsymbol{\theta}$) is given by:

$$\begin{aligned} f(\Phi, \boldsymbol{b}_m, \boldsymbol{b}_s, \boldsymbol{\theta}|\mathbf{M}, \mathbf{S}) \propto & f(\mathbf{M}|\Phi, \boldsymbol{b}_m, \boldsymbol{\theta}) && \text{(morphological likelihood)} \\ & f(\mathbf{S}|\Phi, \boldsymbol{b}_s, \boldsymbol{\theta}) && \text{(molecular likelihood)} \\ & f(\boldsymbol{b}_s|\boldsymbol{\theta}) f(\boldsymbol{b}_m|\boldsymbol{\theta}) && \text{(molecular and morphological clock models)} \\ & f(\Phi|\boldsymbol{\theta}) && \text{(prior on phylogenetic tree topology and node times)} \\ & f(\boldsymbol{\theta}) && \text{(prior on the remaining parameters)} \end{aligned} \tag{1.1}$$

For the purposes of the present study, the morphological likelihood corresponds to the probability of observing $\mathbf{M}$ under a phylogenetic BM model (Felsenstein, 1973), and the molecular likelihood to the probability of a observing $\mathbf{S}$ under a molecular substitution model (Felsenstein, 1981). Finally, the tree prior $f(\Phi|\boldsymbol{\theta})$ gives the probability of a specific topology and node times in phylogenetic tree Φ, with $f(\boldsymbol{\theta})$ corresponding to the prior distribution on all remaining parameters $\boldsymbol{\theta} = \{\boldsymbol{r}, c_m, \boldsymbol{y_0}, \lambda, \mu, \psi, \rho, t_{\text{mrca}}, \boldsymbol{c}_s, \boldsymbol{\kappa}, \boldsymbol{\pi}, \boldsymbol{\zeta}\}$ (see text below, Supplementary Table S11, and Fig. 1 for definitions; in Fig. 1, $\boldsymbol{\theta}_\Phi = \{\lambda, \mu, \psi, \rho, t_{\text{mrca}}\}$ are the tree prior parameters, and $\boldsymbol{\theta}_s = \{\boldsymbol{\kappa}, \boldsymbol{\zeta}\}$, as $\boldsymbol{\pi}$ is set to empirical values).

The posterior distribution $f(\Phi, \boldsymbol{b}_m, \boldsymbol{b}_s, \boldsymbol{\theta}|\mathbf{M}, \mathbf{S})$ under our integrative model is approximated by Markov Chain Monte Carlo (MCMC) sampling in BEAST2 (Bouckaert et al., 2019).

Continuous character model

Brownian motion is a convenient model for the comparative study of continuous characters because its density function is the same as that of the well known multivariate normal distribution (Felsenstein, 1973):

$$f(\mathbf{M}|\Phi, \boldsymbol{b}_m, \boldsymbol{r}, c_m, \boldsymbol{y_0}) = \frac{1}{(2\pi)^{nk/2}|\boldsymbol{V}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{M}-\boldsymbol{y_0})^{\mathrm{T}}\boldsymbol{V}^{-1}(\mathbf{M}-\boldsymbol{y_0})\right), \tag{1.2}$$

where $\mathbf{M}$ is an $n \times k$ matrix of observed continuous characters (n species, k traits), $\boldsymbol{b}_m$ are the relative branch evolutionary rates, $\boldsymbol{r}$ are the relative character-specific evolutionary rates, c_m is the global evolutionary rate, and $\boldsymbol{y_0}$ corresponds to the trait values from all species at the root of Φ. (Note that we unpack $\boldsymbol{r}$, c_m, and $\boldsymbol{y_0}$ from $\boldsymbol{\theta}$ in equation 1.1.)

The phylogenetic variance-covariance matrix $\boldsymbol{V}$ in equation 1.2 corresponds to the Kronecker product between matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{T}$, i.e., $\boldsymbol{V} = \boldsymbol{\Sigma} \otimes \boldsymbol{T}$. The phylogenetic component of the BM model, $\boldsymbol{T} = (t_{uw})$, is a symmetric $n \times n$ matrix deterministically obtained from phylogenetic tree Φ, relative branch rates $\boldsymbol{b}_m$ and global evolutionary rate c_m :

$$t_{uw} = \sum\nolimits_{z \in \mathrm{Path}(u,w)} c_m \boldsymbol{b}_m^z \ell(z), \tag{1.3}$$

where u and w denote any two species in Φ, $\mathrm{Path}(u, w)$ returns the set of branches on the phylogenetic path shared by u and w (from the root to the most recent common ancestor of u and w), and $\boldsymbol{b}_m^z$ and $\ell(z)$ are the relative rate and length of branch z, respectively.

Matrix $\boldsymbol{\Sigma}$ is the $k \times k$ Hadamard product between $\boldsymbol{\eta}$ and $\boldsymbol{\rho}$, i.e., $\boldsymbol{\Sigma} = \boldsymbol{\eta} \circ \boldsymbol{\rho}$. The symmetric character correlation matrix, $\boldsymbol{\rho}$, is defined as:

$$\boldsymbol{\rho} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1k} \\ \rho_{21} & 1 & \cdots & \rho_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k1} & \rho_{k2} & \cdots & 1 \end{bmatrix}, \tag{1.4}$$

with off-diagonal elements representing the correlation between a pair of different characters. Finally, matrix $\boldsymbol{\eta} = (\eta_{ij})$ is given by:

$$\eta_{ij} = \begin{cases} r_i, & \text{if } i = j \\ \sqrt{r_i r_j}, & \text{otherwise,} \end{cases} \tag{1.5}$$

with $i, j \in \{1, 2, ..., k\}$. Note that if the same relative evolutionary rate is assumed for all characters (i.e., $\boldsymbol{r} = \mathbf{1}$; Fig. 1), then $\boldsymbol{\Sigma} = \boldsymbol{\rho}$.

Considering the the Kronecker product between $k \times k$ matrix $\boldsymbol{\Sigma}$ and $n \times n$ matrix $\boldsymbol{T}$, the phylogenetic variance-covariance matrix $\boldsymbol{V}$ becomes an $nk \times nk$ matrix, which consists of n^2 $k \times k$ matrices, where $\boldsymbol{v}_{uw} = t_{uw}\boldsymbol{\Sigma}$.

Scalability with number of traits and species

It is clear from equation 1.2 and the definition of $\boldsymbol{V}$ that computing the probability density of $\mathbf{M}$ under the BM model for large n and k will be computationally demanding. Not only is a Kronecker product ($\boldsymbol{V} = \boldsymbol{\Sigma} \otimes \boldsymbol{T}$) required, causing the evaluation of 1.2 to slow down proportionally to k^2 and n^2, but also $\boldsymbol{V}$ must be inverted, an operation the lower bound of which is $(nk)^2$ (Freckleton, 2012).

Fortunately, in the same work proposing BM as an evolutionary model for continuous characters, Felsenstein (1973) also introduced the pruning algorithm as the basis for addressing both problems mentioned above. In a nutshell, the original pruning algorithm amounts to computing three quantities, for each of the $(2n - 1)$ nodes in the tree: the variance, the variance-weighted expectation, and the probability density of a multivariate normal distribution given the first two quantities. This algorithm precludes the computation of $\boldsymbol{\Sigma} \otimes \boldsymbol{T}$ and the inversion of $\boldsymbol{V}$, although it is still necessary to invert and calculate the determinant of $\boldsymbol{\rho}$ (this operation can nevertheless be avoided by transforming the data, $\mathbf{M}$; see Eq. 6 in Álvarez Carretero et al., 2019 and the supplementary material for more detail). For the sake of brevity, and because this

algorithm has been described and generalized in many subsequent studies (e.g., Felsenstein 1973; Freckleton 2012; Caetano and Harmon 2017; Álvarez Carretero et al. 2019; Mitov et al. 2020), we point the interested reader to those references and to our supplementary material for more detailed descriptions of the algorithm and a worked example.

More recently, Mitov et al. (2020) proposed a very general pruning-based solution for calculating 1.2, as well as the probability density function of more general models, such as BM with early bursts (Harmon et al., 2008) and accelerating or decelerating rates (Blomberg et al., 2003), BM with trends (Hansen and Martins, 1996) and the Ornstein-Uhlenbeck process (Hansen, 1997; Butler and King, 2004). Unlike the pruning algorithm by Felsenstein (1973), the algorithm in Mitov et al. (2020) does not compute the maximum-likelihood estimate of trait values at internal nodes of the tree, but instead calculates a series of intermediate values (which gives this algorithm its flexibility; see Eq. 2 in Mitov et al. 2020). These intermediate values are then combined at the root node in the calculation of an integral, which then gives the final probability density (Eq. 6 in Mitov et al. 2020). Readers can find a detailed description of this algorithm in Mitov et al. (2020), with it being put to use in Mitov and Stadler (2019). We provide a worked example in the supplementary material.

The second obstacle to carrying out inference for multiple characters under phylogenetic BM is posed by the curse of dimensionality. As k increases, the number of character correlation parameters (the off-diagonal elements of $\boldsymbol{\rho}$) we must estimate increases quadratically; for Bayesian inference, this means long MCMC chains must be employed in order to achieve convergence. Furthermore, unless the number of taxa n also increases in a similar fashion, there will be more parameters to estimate than data points. If there are more parameters to estimate than data points (i.e., $n < k$), then non-identifiability will ensue. This is a problem for which no easy solution exists if one is indeed interested in learning about the evolution of morphological trait correlations (Goswami et al., 2014; Caetano and Harmon, 2017).

Alternatively, although the non-independence among characters should always be accounted for in some manner, it might be of secondary interest relative to the estimation of a time-scaled phylogenetic tree. Different approaches have been explored or suggested for such cases (e.g., Adams, 2014; Goolsby, 2016; Adams and Collyer, 2018), but one simple solution is to employ the unbiased estimator of $\boldsymbol{\rho}$, $\widehat{\boldsymbol{\rho}}$, obtained from multiple character observations (e.g., multiple individuals) within a species in the phylogeny. (Note that here this estimator is unbiased with respect to the population of a single species, and by using $\widehat{\boldsymbol{\rho}}$ we are assuming trait correlations are the same across species and over time.) Unfortunately, when one chooses to do the latter, a third non-obvious issue arises: as $n << k$, the determinant of $\widehat{\boldsymbol{\rho}}$ will approach zero and $\widehat{\boldsymbol{\rho}}$ will be singular and non-invertible. In such cases, it is impossible to evaluate equation 1.2.

One strategy recently employed in a Bayesian context for divergence time estimation (Álvarez Carretero et al., 2019) involves using the linear shrinkage estimate of $\widehat{\boldsymbol{\rho}}$, given by:

$$\boldsymbol{\rho}^* = \delta\mathbf{I} + (1-\delta)\widehat{\boldsymbol{\rho}}, \tag{1.6}$$

which consists of the average between the $k \times k$ identity matrix and $\widehat{\boldsymbol{\rho}}$, weighted by the shrinkage parameter δ. This parameter can be optimized as described in Schäfer and Strimmer (2005).

Tree models for total-evidence dating

Modeling the evolution of molecular and morphological characters is crucial for statistically sound taxon placement across and along a phylogeny. An integrative model for total-evidence inference is nonetheless incomplete without accounting for the phylogenetic process itself: the birth and death of lineages. Total-evidence dating, in particular, further requires addressing the fossilization process underlying heterochronous data sets. The fossilized birth-death process (FBD; Stadler, 2010; Heath et al., 2014; Gavryushkina et al., 2014, 2017) is one tree model that has enjoyed success in the context of total-evidence dating, due to its capacity to account for fossilization simultaneously with speciation and

extinction. Its statistical cousin, the birth-death-sequential-sampling model (BDSS; Stadler, 2010; Stadler and Yang, 2013), is yet another option, differing from the FBD in that it conditions on fossil sampling times rather than using the sample times as data. We compare results obtained from using both models under different configurations (Table 1) in analyses of Carnivora data (see below).

Software

Our integrative model for total-evidence phylogenetic inference (Eq. 1.1) is implemented in the BEAST2 platform (Bouckaert et al., 2019). The molecular components of our integrative model, parametric distributions, MCMC machinery, and the FBD model have already been part of BEAST2 since its first release, or incorporated since then, prior to the present study.

Here, we implement the general pruning-based method of Mitov et al. (2020) for computing the likelihood of phylogenetic Brownian models in BEAST2's `contraband` package (`https://github.com/fkmendes/contraband`). This method can be readily integrated with a variety of BEAST2 clock and epoch models, which allow for among-branch and among-epoch variation of model parameters, such as $\mathbf{\Sigma}$. Under our implementation, the among-trait covariance can either be sampled (as in Caetano and Harmon 2017) or its unbiased estimator can be used (as in Álvarez Carretero et al. 2019). For cases where the number of traits is near to or larger than the number of species, computing the linear shrinkage estimate of the trait variance-covariance matrix is also available as an option. Details on method benchmarking and validation can be found in the supplementary material.

Finally, we implement and validate the BDSS model for its utility both as an alternative to the FBD tree prior in total-evidence dating, and as a necessary component for validating our morphological model implementation against previous work (Álvarez Carretero et al., 2019). Both BDSS and FBD tree models work alongside our

implementation of Mitov et al. (2020)'s method. The BDSS model can be found in the `bdtree` repository (`https://github.com/fkmendes/bdtree`).

Case study: Carnivora phylogeny

We illustrate our total-evidence approach by carrying out several analyses of a published Carnivora data set (Álvarez Carretero et al., 2019) under different integrative model configurations.

Molecular and morphological data The data set in Álvarez Carretero et al. (2019) is comprised of (i) a concatenated alignment of 12 mitochondrial genes from 10 extant species, (ii) 29 three-dimensional cranium landmarks (considered as 87 continuous characters) from the same 10 extant species and an additional nine fossil specimens, and (iii) the same 29 landmarks scored from 21 *Vulpes vulpes* individuals (for estimating $\widehat{\boldsymbol{\rho}}$, see above).

We follow the same protocol in Álvarez Carretero et al. (2019) to prepare the morphological data (cranium landmarks) for analysis, and start by "aligning" the landmarks. In addition to size and shape, raw landmarks carry nuisance information about position and orientation, which preclude their statistical analysis (Mitteroecker et al., 2013). Distilling shape and size from raw landmarks can be done with Procrustes superimposition, commonly used in biological shape analysis to "align" (superimpose) landmarks (Mitteroecker et al., 2013). Procrustes superimposition consists of rotating, translating and scaling landmark configurations relative to their centroid (i.e., their average position) and its size, so as to minimize the Procrustes distance – a measure of how different in shape two landmark configurations are (Gower, 1975; Rohlf and Slice, 1990). The Procrustes distance is given by the summed squared distance over landmarks and their sample average position; if zero, then two landmarks have the same shape.

We further prepare the morphological data so as to address intraspecific character

variation. Comparative methods still traditionally employ a single measurement expected to represent each species in a phylogeny, such as averages from a group of individuals of a species. Individuals within a species, however, will invariably exhibit different phenotypes for a myriad of reasons, such as genetic variability (Lynch and Walsh, 1998), direct effects of environmental factors that differ among populations, variation related to age and sex, seasonal fluctuations (Ives et al., 2007), to name a few. In addition, each data point can be further biased by measurement error due to nonrandom sampling of individuals and instrumental error (Garamszegi and Møller, 2010; Hansen and Bartoszek, 2012). Failing to address intraspecific phenotypic variance can mislead comparative analyses in multiple ways (Kostikova et al., 2016). For example, different modes of evolution can be inferred (e.g., rapid vs. gradual body size changes in vertebrates; Landis and Schraiber, 2017), and both evolutionary rates (Clavel and Morlon, 2017) and divergence times (Álvarez Carretero et al., 2019) can be overestimated.

One way to account for phenotypic variation among conspecifics is to increment trait variances and among-trait covariances by some constant $\mathbf{v}_{\text{err}}$ (Ives et al., 2007). This amounts to using phylogenetic variance-covariance matrix $\boldsymbol{V'} = (\mathbf{v'_{uw}})$, which is updated by $\mathbf{v}_{\text{err}}$ from $\boldsymbol{V} = (\mathbf{v_{uw}})$ (in equation 1.2). We have:

$$\mathbf{v}'_{uw} = \begin{cases} \mathbf{v}_{uw} + \mathbf{v}_{\text{err}}, & \text{if } u = w \\ \mathbf{v}_{uw}, & \text{otherwise,} \end{cases} \tag{1.7}$$

where $\mathbf{v}_{\text{err}}$ is a $k \times k$ matrix given by the Hadamard product between the trait correlation matrix $\boldsymbol{\rho}$ and $\boldsymbol{\epsilon}$, i.e., $\mathbf{v}_{\text{err}} = \boldsymbol{\epsilon} \circ \boldsymbol{\rho}$. Note that in practice $\boldsymbol{\rho}^*$ (equation 1.6) can be used instead of $\boldsymbol{\rho}$ for the aforementioned reasons. The $k \times k$ matrix $\boldsymbol{\epsilon} = (\epsilon_{ij})$ is given by:

$$\epsilon_{ij} = \begin{cases} {\sigma_i}^2, & \text{if } i = j \\ \sqrt{{\sigma_i}^2 {\sigma_j}^2}, & \text{otherwise,} \end{cases} \tag{1.8}$$

where ${\sigma_i}^2$ represents the intraspecific variance of character i.

If the unbiased estimator of ${\sigma_i}^2$, $\widehat{\sigma_i^2}$, is not available or cannot be computed (in the absence of measurements from multiple individuals from a species), $\mathbf{v}_{\text{err}}$ can be inferred at

the cost of longer MCMC chains. As in Álvarez Carretero et al. (2019), we nonetheless use $\widehat{\sigma_i^2}$ calculated for all 87 landmarks from 21 individuals of *V. vulpes*. We also normalize each i-th observed landmarks in $\mathbf{M}$ by their corresponding $\widehat{\sigma_i}$, i.e., $\mathbf{M}^{(s)} = \mathbf{M} \times \text{diag}\{1/\sqrt{\widehat{\boldsymbol{\sigma}}}\}$, where $\widehat{\boldsymbol{\sigma}}$ holds the unbiased estimators of intraspecific standard deviations (of all k landmarks). Using $\mathbf{M}^{(s)}$ is convenient because this normalization leads to unit (co-)variances, i.e., $\boldsymbol{\epsilon}$ becomes a matrix of ones, and then $\mathbf{v}_{\text{err}} = \boldsymbol{\rho}^*$.

Integrative model configurations The general structure of our integrative model can be represented by a probabilistic graphical model (Fig. 1). With the exception of species tree priors in some of our analyses (see below), we matched the model in Álvarez Carretero et al. (2019), which includes molecular and morphological clock models, and all hyperprior distributions. We used the same partitioning scheme for the molecular data – two partitions comprised by 7,331 sites from first and second codon positions, and 3,660 sites from third codon positions, respectively – as well as the same substitution model (HKY+Γ; Hasegawa et al., 1985; Yang, 1994). Equilibrium nucleotide frequencies were set to their empirical values.

Continuous morphological evolution was modelled with a phylogenetic BM model, with all 87 superimposed, standardized characters (see above) sharing the same relative evolutionary rate ($\boldsymbol{r} = \mathbf{1}$; Fig. 1). We estimated the landmark root values ($\boldsymbol{y_0}$) using maximum-likelihood, obtaining $\widehat{\boldsymbol{y_0}}$ as a byproduct of pruning as done in Álvarez Carretero et al., 2019. We did so first because this approach allows for direct comparison with results from that study, and second because it is not immediately obvious how to choose a prior for $\boldsymbol{y_0}$. Using $\widehat{\boldsymbol{y_0}}$ is analogous to employing empirical nucleotide frequencies as the equilibrium distribution under molecular substitution models. We note that both assuming the same evolutionary rates across characters and calculating $\widehat{\boldsymbol{y_0}}$ are not requirements of our implementation; $\boldsymbol{r} = \mathbf{1}$ can be relaxed, and $\boldsymbol{y_0}$ can be sampled during MCMC. Finally, uncorrelated log-normal relaxed clock models were used for both molecular and

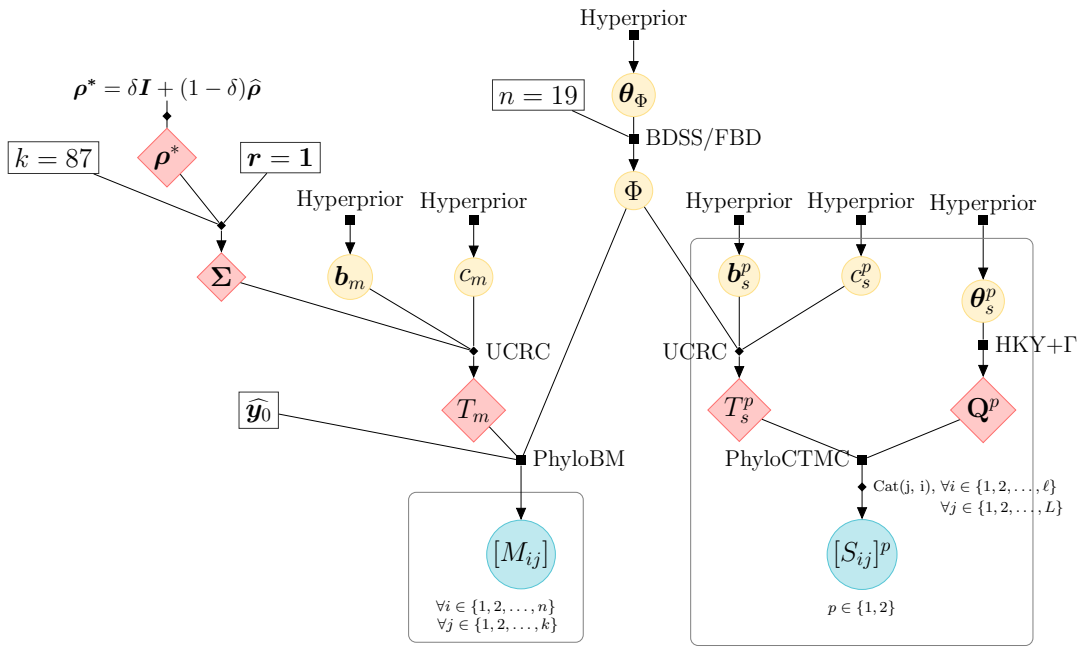

Figure 1. Probabilistic graphical model used in the analyses of the Carnivora data set, composed of 12 concatenated mitochondrial genes ($L = 12$) split into two partitions, and 87 continuous characters ($k = 87$) from 19 Carnivora species ($n = 19$). Filled squares denote sampling distributions, diamonds denote deterministic functions (black) and their outputs (red), and circles denote random variables (yellow) or observed data (blue). Bold symbols represent vectors or matrices, otherwise they are scalars. All symbols are defined in the main text. 'PhyloBM' and 'PhyloCTMC' stand for phylogenetic Brownian motion and phylogenetic continuous-time Markov chain models, and 'UCRC' for uncorrelated relaxed clock models. The sampling distribution for the species tree (Φ) was either the BDSS or FBD model (see Methods section and Table 1). For the sake of clarity, $\boldsymbol{\theta} = \{\boldsymbol{\theta}_\Phi, \boldsymbol{\theta}_s^p\}$ and 'Hyperprior' encompass all parameters and priors not explicitly shown in the graphical model.

morphological evolutionary rates.

In order to explore the influence of species tree priors in total-evidence inference, we carried out three classes of analyses (Table 1). The first analysis constrains the topology of the Carnivora phylogeny to that in Álvarez Carretero et al. (2019) (Fig. 2a). We will henceforth refer to this tree as the "reference tree". Under this first setup, species tree prior parameters were fixed, and only divergence times were estimated. In the second class of analyses, we estimated species tree prior parameters, and both the topology and divergence times of the Carnivora phylogeny, under the BDSS (analysis 3) and FBD (analysis 5) models. The third and final class of analyses (analyses 4 and 6) employed the same models as the second, but we constrained the species tree topology to include the Feliformia and Caniformia clades, both present in the reference tree.

Bayesian inference with Markov Chain Monte Carlo

Bayesian inference was carried out with the BEAST2 platform, which uses MCMC to approximate the posterior distribution over all model parameters. For each of the six analyses detailed above, we ran two independent 50 million-state MCMC chains, sampling every 5,000 states, and discarded the first 10% of samples as burn-in. All chains converged (i.e., achieved effective sample sizes $\geqslant 200$), and we then combined each pair of chains from each model configuration before further analysis.

Model comparison using nested sampling We compared the fit of the original model in Álvarez Carretero et al. (2019) (analysis 1; Table 1) with that of an almost identical model (analysis 2), the only difference being that the species tree topology was an estimated parameter in the latter. Model comparison was conducted by computing posterior odds between the two models (which we refer to below as $\mathcal{M}_0$ and $\mathcal{M}_1$):

$$\underbrace{\frac{f(\mathcal{M}_0|D)}{f(\mathcal{M}_1|D)}}_{\text{Posterior odds}} = \underbrace{\frac{f(D|\mathcal{M}_0)}{f(D|\mathcal{M}_1)}}_{\text{Bayes factor}} \times \underbrace{\frac{f(\mathcal{M}_0)}{f(\mathcal{M}_1)}}_{\text{Prior odds}}. \tag{1.9}$$

By assuming both models have the same prior probability (i.e., prior odds $= 1$), the posterior odds reduces to the ratio of the marginal likelihoods $\mathcal{Z}_0 = f(D|\mathcal{M}_0)$ and $\mathcal{Z}_1 = f(D|\mathcal{M}_1)$. The marginal likelihood of a model is the probability of observing the data under that model, and so it measures a model's goodness-of-fit. The ratio of two model's marginal likelihoods, known as the Bayes factor (BF), thus allows one to quantify which of the two has the best fit. A log-BF > 0, for example, indicates $\mathcal{M}_0$ fits the data better; when > 2, the log-BF suggests $\mathcal{M}_0$ is decisively supported over $\mathcal{M}_1$; Kass and Raftery, 1995.

Many Bayesian methods exist for calculating marginal likelihoods, such as the harmonic mean approach (Newton and Raftery, 1994), thermodynamic integration (Lartillot and Philippe, 2006), steppingstone sampling (Xie et al., 2011), generalized steppingstone sampling (Fan et al., 2011), and nested sampling (NS; Russel et al., 2019).

We chose NS because it is readily available in BEAST2, the platform in which we implemented our integrative model, but also because this technique overcomes many shortcomings of other methods. Namely, NS is robust to phylogenetic tree spaces with tree "islands", copes better with convex likelihood functions, requires simpler tuning, dispenses with the burn-in stage, and can have the uncertainty around its estimate calculated in a single run (Russel et al., 2019).

In order to calculate posterior odds (Equation 1.9), we carried out two NS analyses (one for each model) with $K = 5,000$ and $N = 55$. Sub-chain lengths of 7,500 produced statistically similar $\mathcal{Z}$ estimates as compared to 5,000-long sub-chains (within twice the sum of their standard deviations), so we deemed a length of 5,000 sufficient for covering the bulk of the marginal likelihood. Given that the (NS) standard deviation of $\mathcal{Z}$ estimates is inversely proportional to the square root of N (Skilling et al., 2006), we adjusted N such that the standard deviation was < 2 (see Supplementary Table 10).

Table 1. Six different integrative model configurations used to estimate the Carnivora phylogeny. Φ indicates both topology and divergence times were sampled, $\mathcal{T}$ indicates just divergence times were sampled. "NA" denotes not applicable.

Analysis (i)	Tree prior	Sampled parameters	Clade constraints
1	BDSS	$\mathcal{T}$	NA
2	BDSS	Φ	NA
3	BDSS	λ, μ, ψ, ρ, Φ	NA
4	BDSS	λ, μ, ψ, ρ, Φ	*Feliformia*, *Caniformia*
5	FBD	λ, μ, ψ, ρ, Φ	NA
6	FBD	λ, μ, ψ, ρ, Φ	*Feliformia*, *Caniformia*

Further comparison of species tree distributions and cranium landmarks using multidimensional scaling In addition to estimating the Carnivora phylogeny using several different models (Table 1), we employed multidimensional scaling (MDS) in order to compare the resulting trees obtained under each model configuration, as well as to further explore the landmark data quantitatively.

For readability purposes, we carried out MDS on a 2-D spatial map while choosing the metric-scaling transformation function that minimized the stress statistic. We used the

`mds()` subroutine from the `smacof` R package (Mair et al., 2019). When exploring phylogenetic space, we compared 100 uniformly sampled trees from each of the six model MCMC chains, and our chosen proximities were (i) Robinson-Foulds distances (Robinson and Foulds, 1981), and (ii) branch scores (Kuhner and Felsenstein, 1994) (see supplementary material for more details). Morphological MDS was conducted on the Euclidean distances among the 19 Carnivora species across all superimposed landmarks.

Results

Carnivora phylogeny

Posterior estimates of Carnivora divergence times matched previous results (Álvarez Carretero et al., 2019) when using the same integrative model configuration (analysis 1; Fig. 2a and Supplementary Fig. 15). Treating the species tree topology as a random variable (analysis 2; Fig. 2b) considerably improved model fit, however, as indicated by a log-BF of 39.07. (A log-BF > 3 is conventionally interpreted as the best model fitting the data substantially better; Kass and Raftery, 1995). In what follows we focus on the summary maximum-clade-credibility (MCC) tree from analysis 2 (Fig. 2b; results from the remaining analyses can be found in the supplementary material).

In terms of fossil placement, there were four notable differences (red asterisks, Fig. 2b) between the reference tree topology and the MCC tree summarized from our posterior samples under the BDSS tree prior, when estimating both topology and divergence times. First, *Smilodon fatalis*, one of the extinct species commonly referred as "saber-toothed cat", was inferred to be more closely related to modern dogs and other canines than to cat-like carnivoran species in Feliformia, the suborder *S. fatalis* is canonically assumed a member of. Similarly, *Hyaenictitherium wongii*, a middle-sized hyaenid from the Late Miocene (Werdelin and Solounias, 1991) was also estimated to be more closely related to canines than to other feliforms. Third, while still placed within caniforms, *Enhydrocyon*

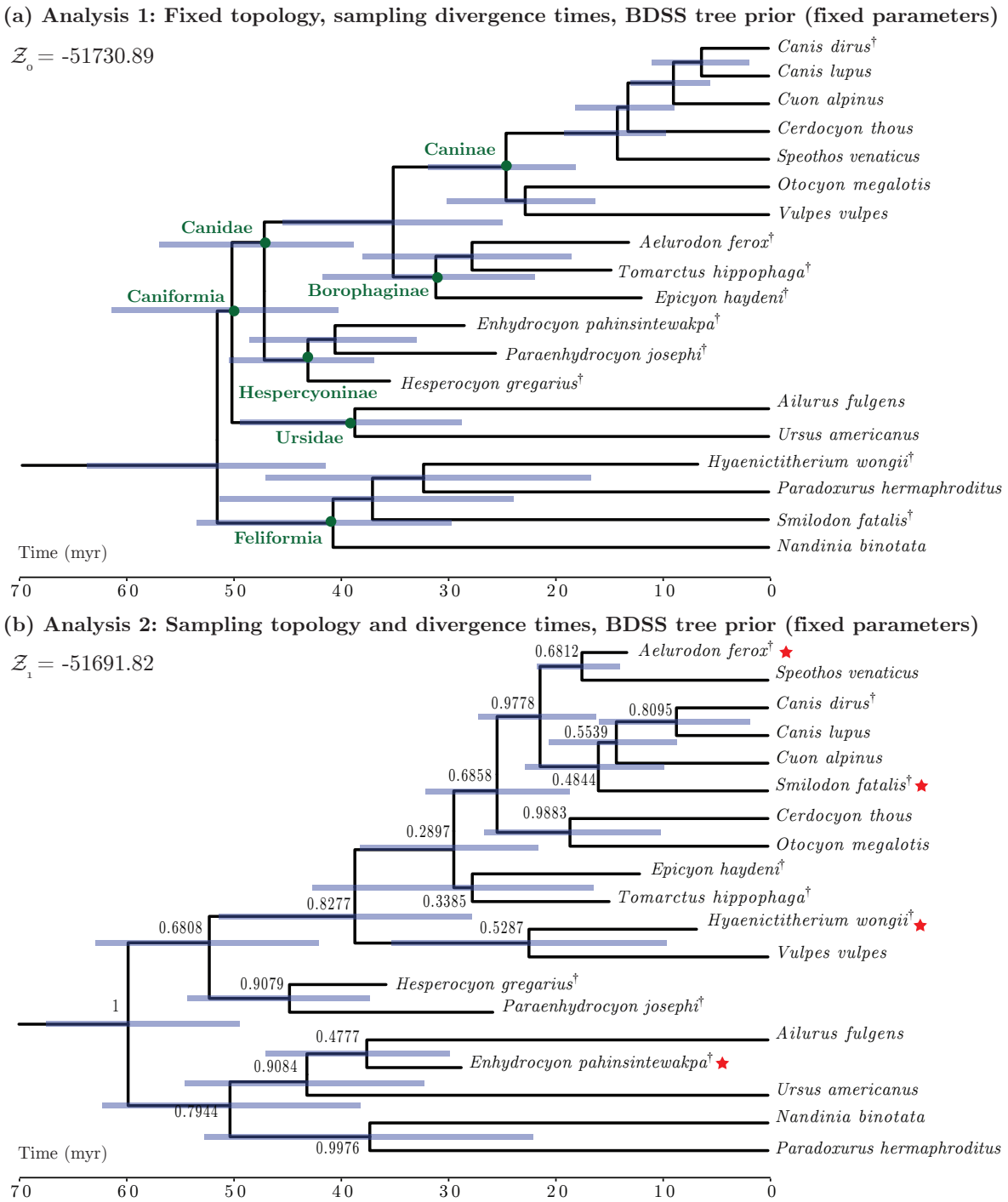

Figure 2. Carnivora maximum-clade-credibility summary trees. The horizontal bars at internal nodes show the 95% credible intervals for their times. Numbers by internal nodes indicate each clade's posterior probability. Carnivora suborders, families and subfamilies are labelled in green next to corresponding internal nodes. (a) Tree estimated from analysis 1 (Table 1; fixed topology from Álvarez Carretero et al., 2019 and other references therein). (b) Tree estimated from analysis 2. Red stars indicate large differences in placement of key taxa.

pahinsintewakpa was inferred to be phylogenetically closer to modern raccoons and bears rather than with the other extinct members of Hesperocyoninae canids. Finally, we did not recover the other extinct canid subfamily, Borophaginae; while two of its members (*Tomarctus hippophaga* and *Epicyon haydeni*) grouped together, *Aelurodon ferox* was inferred as sister to modern bush dogs (*Speothos venaticus*).

The fossil affinities mentioned above meant that none of the more inclusive clades of Carnivora in the reference tree were recovered with high posterior clade probabilities (e.g., Caniformia and Feliformia suborders, Canidae and Ursidae families, Caninae and Borophaginae subfamilies). Another visible difference included the placement of the two ursids and *E. pahinsintewakpa* on the opposite side of the root (relative to the reference tree), closer to feliforms *Nandinia binotata* and *Paradoxurus hermaphroditus* than to other canids, at moderate posterior probability (PP=0.79). This difference was observed in both molecules-only and morphology-only trees (Supplementary Fig. 19).

Topological similarities with the reference tree were nonetheless observed. The clade containing all canids (in addition to *S. fatalis* and *H. wongii*), had relatively high support (posterior probability, PP, of 0.8277). The clades corresponding to Ursidae and Hespercyoninae, the former being expanded by *E. pahinsintewakpa* and the latter missing this taxon, also showed high clade support (PP=0.9 for both). These two feliforms, *N. binotata* and *P. hermaphroditus*, grouped together with very high posterior probability (PP=0.99).

Two extant canines for which both morphological and molecular data were available presented aberrant species relationships relative to the reference tree: the bush dog (*S. venaticus*) and the red fox (*V. vulpes*). The bush dog grouped with *A. ferox*, a fossil member of Borophaginae, at moderate posterior probability (PP=0.68). Both these species were inferred to be more closely related to *Canis* species and *Cuon alpinus* than to *Cerdocyon thous*. The phylogenetic affinity of the bush dog seems to be supported by both molecular and cranium landmark data (Supplementary Fig. 19). The placement of the red

fox was quite uncertain, with this species forming a clade with one of the four outlier fossils mentioned above, *H. wongii* (PP=0.5287). Curiously, this grouping was not observed in either the molecule-only or the morphology-only trees (Supplementary Fig. 19).

Phylogenetic constraints and rogue fossils One way to incorporate prior knowledge about species relationships in phylogenetic inference is to constrain the monophyly of specific groups. Holding the entire topology constant as done by Álvarez Carretero et al. (2019), for example, takes this strategy to an extreme. The rationale behind monophyletic constraints is that sometimes experts agree the veracity of that clade is beyond doubt. In practice, using monophyletic priors might make sense when a researcher does not have easy access to (or cannot use) the data upon which the confident monophyly belief is predicated. Because three of the four "rogue" fossils (marked with a red star; Fig. 2b) were placed on the wrong side of the root, in the wrong suborder, we reasoned constraining the monophyly of Caniformia and Feliformia could help us further scrutinize the behavior of our model.

The main non-trivial topological difference with respect to the unconstrained analysis was the placement of the red fox, *V. vulpes*, as an outgroup of the remaining canines (Supplementary Figs. 17a and 17c). Apart from the still intrusive *A. ferox*, this analysis recovered Caninae with considerable posterior probability (PP=0.85). Moreover, if one were to ignore *E. pahinsintewakpa*, the posterior probability of Ursidae and Hespercyoninae increased from 0.9 (in both cases) to 0.97 and 0.99, respectively.

We also repeated the unconstrained analysis with the model used in analysis 2 while removing either (i) three rogue fossils, *S. fatalis*, *H. wongii*, and *E. pahinsintewakpa* (Supplementary Fig. 18a), or (ii) all four rogue fossils (adding *A. ferox* to the three aforementioned specimens; Supplementary Fig. 18b). Our hope was to determine whether data from these "rogue" fossils were driving the topological differences between the reference and estimated topologies (Fig. 2).

If we consider the reference tree to be the desired goal, removing both three or all of the four rogue fossils improved the placement of canine species relative to the analysis

including all fossils (Fig. 2b). *S. venaticus* was inferred to be more closely related to other canine species, and *V. vulpes* was inferred to be a canine instead of being placed outside of both Caninae and Borophaginae (Supplementary Fig. 18b).

On the other hand, removing rogue fossils split the *Canis* genus and resulted in the dire wolf, *Canis dirus*, being grouped with either taxa from Ursidae (Supplementary Fig. 18a) or from Feliformia (Supplementary Fig. 18b). In both cases, the placement of *C. dirus* attained high posterior probability: PP=0.9515 with *A. fulgens* when removing three rogue fossils, and PP=0.9867 with the remaining feliforms when dropping all rogue fossils. Removal of problematic fossils also affected members of subfamily Borophaginae: *Tomarctus hippophaga* was placed with high posterior probability among members of the Caninae subfamily, and *Epicyon haydeni* grouped with species outside of Canidae at moderate to high posterior probability (Supplementary Fig. 18a).

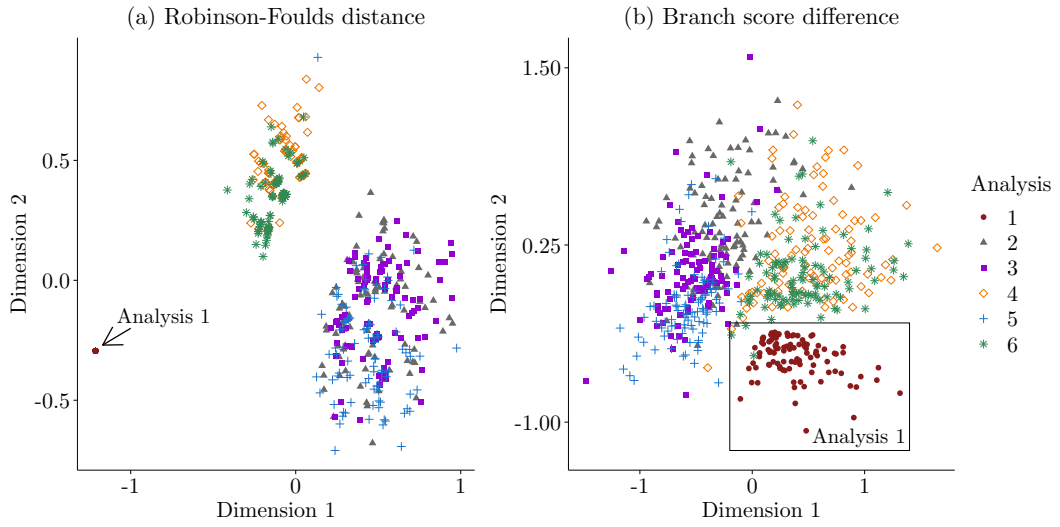

Figure 3. Comparison of Carnivora species tree posterior distributions using multidimensional scaling (MDS) of tree distance measures: (a) Robinson-Foulds distance, and (b) Branch score distance. Each point consists of one of 100 equally spaced (in the MCMC chain) posterior tree samples after discarding the burn-in. Note that Robinson-Foulds distances do not take branch lengths into account and that we fix the species tree topology in analysis 1, so this analysis is represented by a single point in (a).

Further comparison of species tree distributions and cranium landmarks

Species tree posterior distributions were not markedly shifted in topological space (disregarding branch lengths) by choice of tree prior nor by the fixing of their hyperparameters, regardless of the monophyly constraints of the Caniformia and Feliformia suborders (Fig. 3a). Such constraints (analyses 4 and 6; Table 1) did, unsurprisingly, yield a separate topological cluster of tree distributions for both tree distance measures (Fig. 3), as did constraining the entire topology (analysis 1). Unconstrained tree posterior distributions (analyses 2, 3 and 5) were not distinguishable as MDS of Robinson-Foulds distances suggests (Fig. 3a), and still largely overlapped when branch lengths were accounted for (Fig. 3b).

Results from MDS of Euclidean distances between species landmarks (Fig. 4) largely agreed with our reconstruction of the species tree topology (Fig. 2b). For example, cranium landmarks from feliforms *S. fatalis* and *H. wongii* proved to be very different from each other and from the other two feliforms who grouped together (Fig. 2b); *S. fatalis*, in particular, is an outlier relative to all other specimens. Cranium landmarks of *E. pahinsintewakpa* (Hespercyoninae) were more similar to those of ursids *Ailurus fulgens* and *Ursus americanus*, and of feliforms *N. binotata* and *P. hermaphroditus*, than to other caniforms'. MDS also placed *A. ferox* closer to canines than to other members of Borophaginae. All these observations are in line with the inferred MCC topology (Fig. 2b).

Discussion

While the principle behind the "total evidence" approach – simultaneously leveraging multiple sources of data in phylogenetic reconstruction – is over 30 years old (Kluge, 1989), it was only given a statistically principled treatment in the last decade (Pyron, 2011; Ronquist et al., 2012). Even the more recent total-evidence example studies, however, still limit themselves to discrete or discretized morphological traits (Lee and Palci,

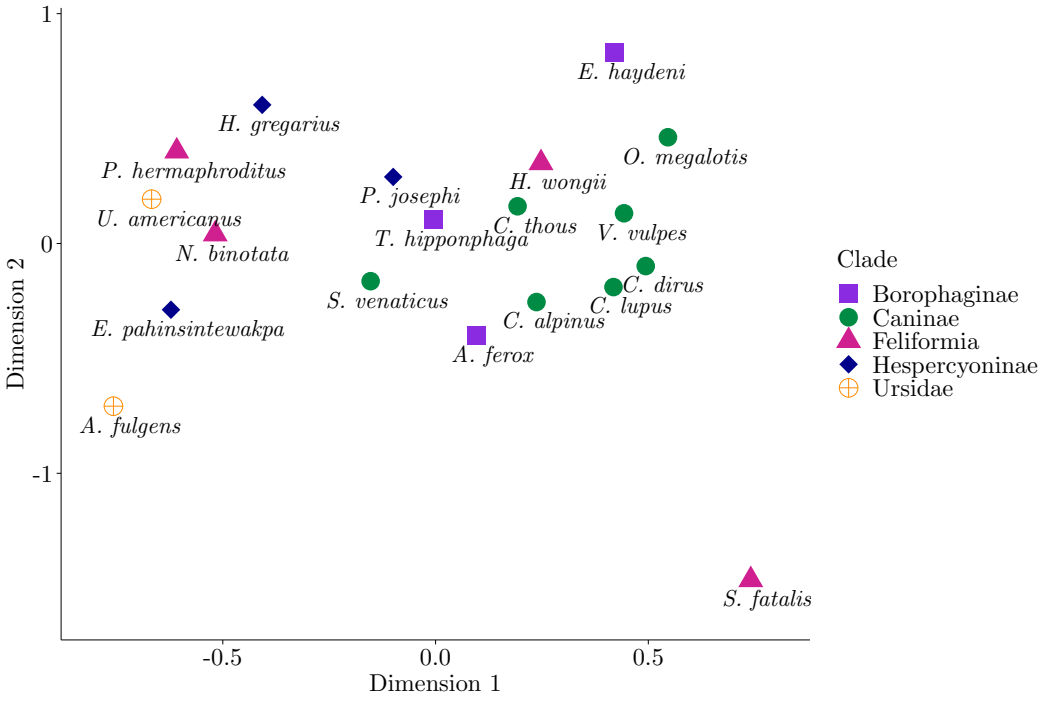

Figure 4. Multidimensional scaling (MDS) of Euclidean distances between Carnivora species cranium landmarks.

2015), and do not always model trait evolution statistically (e.g., O'Leary, 1999; Farias et al., 2000; Seiffert, 2007; Schuh et al., 2009; Arrigo et al., 2013; Polotow et al., 2015).

This is likely not being driven, first, by a perceived superiority of discrete or discretized morphological traits over their continuous counterparts, in terms of their usefulness to phylogenetic inference. Conventionally used discrete trait models are known to face challenges such as accounting for among-trait correlation, and making assumptions about the stationary distribution over character states (Parins-Fukuchi, 2018a; Álvarez Carretero et al., 2019). Discrete trait data sets also suffer from subjectivity in the inclusion and scoring of characters, and from loss of information caused by the discretization of continuous traits (Goloboff et al., 2006).

Second, the common use of maximum-parsimony in total-evidence inference is also not due to a consensus on the superiority of this criterion; rather, tools for the joint modeling of continuous morphological and molecular evolution are still lacking (but see

Álvarez Carretero et al., 2019; May and Moore, 2020; Gaboriau et al., 2020). It was not until recently that careful simulation studies investigated the use of phylogenetic BM (Parins-Fukuchi, 2018b; Varón-González et al., 2020) and implemented statistical tools with the purpose of placing fossils and inferring phylogenies (Parins-Fukuchi, 2018a; Álvarez Carretero et al., 2019; May and Moore, 2020).

We introduce a new total-evidence method that allows comparative biologists to add quantitative traits to molecular sequences in the joint inference of phylogenetic relationships and divergence times. Our BEAST2 implementation was extensively benchmarked and validated; it is correct, fast, scales linearly with the number of species, and supports both the inference of among-trait correlations with MCMC, as well as the use of the linear shrinkage method. Because it follows a general mathematical framework (Mitov and Stadler, 2019), our method can be readily extended to include models such as BM with trends (Hansen and Martins, 1996), Ornstein-Uhlenbeck (OU; Hansen and Martins, 1996; O'Meara et al., 2006), early burst (EB; Harmon et al., 2010), accelerated or decelerated-rate models (ACDC; Blomberg et al., 2003), to name a few. The BEAST2 platform also provides a series of clock and other evolutionary models that can be used in conjunction with (or replace components of) the integrative model we use here. These extensions will be the subject of future studies.

In order to showcase our method, we carried out Bayesian total-evidence estimation of a phylogeny of Carnivora from both extant and extinct species data. When constraining our integrative model to match that in a previous study (Álvarez Carretero et al., 2019) – where the species tree topology was held constant at a "reference" configuration (Finarelli and Goswami, 2009; Martín-Serra et al., 2014; Álvarez Carretero et al., 2019) – our analysis yielded the same results. But an almost identical model in which the species tree topology was a random variable fit the data significantly better. This result bears out how treating the topology as data can be misguided, even if not necessarily so in Álvarez Carretero et al. (2019)'s (and our analogous) analysis. At best, the habit of assuming a

known topology will always mask the phylogenetic incongruence between data and tree, and can have unwanted consequences to evolutionary inference (Mendes et al., 2016). When all available data can be used in a single analysis and nothing else is known about a phylogeny, one should strive to estimate both divergence times and topology, steering clear from the alluring comfort of the phylogenetic Procrustean bed (Hahn and Nakhleh, 2016).

When estimating the topology of the Carnivora species tree, the molecular phylogenetic signal alone supported most clades with high posterior probability (Supplementary Fig. 19a) as opposed to the much noisier morphological signal (Supplementary Fig. 19b). This is unsurprising because molecular alignments harbor a much larger number of topologically informative characters (Lee and Palci, 2015), and because the molecular tree space was also smaller (we did not have DNA from fossils). Neither data set decidedly supported the reference tree topology, however, with the phylogenetic signal coming from molecules and morphology both agreeing and disagreeing in different parts of the tree. For example, adding morphological data decreased the posterior probability of *Canis* by 0.1 posterior probability units relative to the molecules-only tree, while increasing the support of (*N. binotata*, *P. hermaphroditus*) by 0.0436 units.

Despite the noisy phylogenetic signal carried by the cranium landmarks, adding them to the molecular alignments had a considerable effect on the posterior distribution of species trees. Topologically, the effect was largely driven by four rogue fossils. Among those, *S. fatalis* had a particularly strong phylogenetic affinity to other canine species and *A. ferox* in the morphology-only tree (Supplementary Fig. 19b), which was less pronounced but still present in the full data analysis (Fig. 2b). The aberrant placement of rogue fossils relative to the reference tree was clearly echoed by further exploration of the cranium landmarks using multidimensional scaling (Fig. 4). These results are in agreement with principal component analyses of cranium landmarks carried out by Álvarez Carretero et al. (2019), which also revealed *S. fatalis* to be an outlier specimen.

While adding cranium landmark data to the analysis decreased node support overall, the phylogenetic signal coming from this data partition was not merely diffuse, but strongly supported certain clades. The *Canis* genus comprised by the extant gray wolf (*C. lupus*) and the extinct dire wolf (*C. dirus*), for example, was well supported by the continuous trait data set. The morphological phylogenetic signal also trumped the signal contained in the molecular alignments; *Cerdocyon thous*, the South American crab-eating fox, was confidently inferred as sister to *Otocyon megalotis*, the African bat-eared fox. This result suggests that the number and size of molecular alignments used in this study were not large enough to overwhelm the signal from the morphological data.

At this point, we should clarify that the goal of our Carnivora analyses was not to contribute to the understanding of this group's systematics, but to illustrate the estimation of a species tree topology and divergence times from both molecular and many continuous morphological traits. While *C. dirus* being more closely related to *C. lupus* than to *Cuon alpinus* is plausible (Tedford et al., 2009; Slater, 2015; but see Perri et al., 2021), *S. fatalis* earned the soubriquet "saber-toothed cat" – instead of "saber-toothed canid" – for good reason (Werdelin, 1996; Turner and Antón, 1997; Anton et al., 2004; Werdelin et al., 2010; Flynn et al., 2010; Christiansen, 2013). This latter result motivated us to further explore the empirical consequences of excluding *S. fatalis* and other rogue taxa, as well as employing monophyletic constraint priors. Neither strategy proved a silver bullet in producing a match between our estimated species and the reference tree, but some improvements were observed, especially with monophyletic constraints.

Many modeling approaches not explored in this study remain open, some of which might help remedy the issues we and others have observed. The placement of species with outlier morphology, for example, could be improved by binning continuous traits into different partitions, each with its own relative evolutionary rate(s). This strategy has been long recognized as an improvement to molecular substitution models (Sullivan and Swofford, 1997), and in our and similar cases might prevent the grouping of lineages that

have undergone large amounts of phenotypic change. Local morphological clock models applied to evolutionary rates and adaptive optima regimes (in the case of OU models) could further accommodate ecologically relevant traits evolving under selection (Eastman et al., 2011; Uyeda and Harmon, 2014; Gaboriau et al., 2020). Accounting for rampant gene tree discordance due to incomplete lineage sorting and introgression may also prove necessary if it is shown that population-level processes cannot be merely buffered out through an additional variance term, but by being carefully modeled instead (Mendes et al., 2018). Progress made on this front (Bastide et al., 2018) and on the phylogenetic modeling of intraspecific trait variance (Gaboriau et al., 2020) might hold the key to capturing additional dimensions of phenotypic evolution.

The performance of our method as described here should not be seen as a set standard because it is likely to be highly dependent on the size, phylogenetic scope of, and phylogenetic signal within a data set. While it did not greatly matter to species tree estimation here, the choice and configuration of tree priors and hyperpriors can be a critical component in total-evidence analyses (Ronquist et al., 2016). Future analyses might additionally consider discrete morphological traits, or even attempt to analyze those traits under threshold models (Wright, 1934; Felsenstein, 2005) by adapting the implementation we introduced here. It is still unclear how the signal among different morphological and molecular data partitions should interact in data sets of different size.

We are only beginning to understand the power and utility of leveraging discrete and continuous morphology in addition to molecules within a robust statistical framework. Recent studies suggest this approach holds promise (e.g., Ronquist et al., 2016; Gavryushkina et al., 2017; Ogilvie et al., 2021), and the future looks bright from many angles. Even if many clades are not prone to fossilization, the vast majority of species to ever roam the planet have gone extinct (Lee and Palci, 2015), and obtaining their DNA (but not measuring their morphology) is challenging at best (Cooper and Poinar, 2000; Hagelberg et al., 2015), and impossible in most cases (Austin et al., 1997). Ongoing efforts

to make different kinds of morphological data easily available (O'Leary and Kaufman, 2011; Cunningham et al., 2015) have only started to scratch the surface of the tree of life's canopy. We are confident that methods such as ours will motivate the curation, expansion and publication of rich morphological data sets, and fuel the probabilistic total evidence paradigm.

Supplementary Material

Data available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.YYYY

Acknowledgements

We thank Julien Clavel and Venelin Mitov for useful discussions on the implementation of continuous trait evolutionary models, and Sandra Álvarez-Carretero for assistance with MCMCTree analyses. This work was supported by the Royal Society of New Zealand (16-UOA-277 to FKM and AJD), and by China Scholarship Council (No.201706990021 to RZ).

Competing interests

The authors declare that they have no competing interests.

References

Adams, D. C. 2014. A method for assessing phylogenetic least squares models for shape and other high-dimensional multivariate data. Evolution 68:2675–2688.

Adams, D. C., C. M. Berns, K. H. Kozak, and J. J. Wiens. 2009. Are rates of species diversification correlated with rates of morphological evolution? Proc. Royal Soc. B 276:2729–2738.

Adams, D. C. and M. L. Collyer. 2018. Multivariate phylogenetic comparative methods: evaluations, comparisons, and recommendations. Syst. Biol. 67:14–31.

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

Anton, M., M. J. Salesa, J. Morales, and A. Turner. 2004. First known complete skulls of the scimitar-toothed cat *Machairodus aphanistus* (Felidae, Carnivora) from the Spanish late Miocene site of Batallones-1. J. Vertebr. Paleontol. 24:957–969.

Arango, C. P. and W. C. Wheeler. 2007. Phylogeny of the sea spiders (arthropoda, pycnogonida) based on direct optimization of six loci and morphology. Cladistics 23:255–293.

Arrigo, N., J. Therrien, C. L. Anderson, M. D. Windham, C. H. Haufler, and M. S. Barker. 2013. A total evidence approach to understanding phylogenetic relationships and ecological diversity in *Selaginella subg. Tetragonostachys*. Am. J. Bot. 100:1672–1682.

Austin, J. J., A. J. Ross, A. B. Smith, R. A. Fortey, and R. H. Thomas. 1997. Problems of reproducibility–does geologically ancient DNA survive in amber-preserved insects? Proc. Royal Soc. B. 264:467–474.

Ayala, F. J. 1997. Vagaries of the molecular clock. Proc. Natl. Acad. Sci. U.S.A 94:7776–7783.

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

Berv, J. S. and D. J. Field. 2018. Genomic signature of an avian Lilliput effect across the K-Pg extinction. Syst. Biol. 67:1–13.

Besenbacher, S., S. Liu, J. M. Izarzugaza, J. Grove, K. Belling, J. Bork-Jensen, S. Huang, T. D. Als, S. Li, R. Yadav, et al. 2015. Novel variation and de novo mutation rates in population-wide de novo assembled danish trios. Nat. Commun. 6:1–9.

Bleiweiss, R. 1998. Slow rate of molecular evolution in high-elevation hummingbirds. Proc. Natl. Acad. Sci. U.S.A 95:612–16.

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

Blomberg, S. P., S. I. Rathnayake, and C. M. Moreau. 2020. Beyond Brownian motion and the Ornstein-Uhlenbeck process: stochastic diffusion models for the evolution of quantitative characters. Am. Nat. 195:145–165.

Bouckaert, R., T. G. Vaughan, J. Barido-Sottani, S. Duchêne, M. Fourment, A. Gavryushkina, J. Heled, G. Jones, D. Kühnert, N. de Maio, et al. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 15:e1006650.

Bracken-Grissom, H. D., S. T. Ahyong, R. D. Wilkinson, R. M. Feldmann, C. E. Schweitzer, J. W. Breinholt, M. Bendall, F. Palero, T.-Y. Chan, D. L. Felder, et al. 2014. The emergence of lobsters: phylogenetic relationships, morphological evolution and divergence time comparisons of an ancient group (Decapoda: Achelata, Astacidea, Glypheidea, Polychelida). Syst. Biol. 63:457–479.

Bromham, L. 2009. Why do species vary in their rate of molecular evolution? Biol. Lett. 5:401–404.

Burin, G., W. D. Kissling, P. R. Guimarães, Ç. H. Şekercioğlu, and T. B. Quental. 2016. Omnivory in birds is a macroevolutionary sink. Nat. Commun. 7:1–10.

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

Caetano, D. S. and L. J. Harmon. 2017. ratematrix: an R package for studying evolutionary integration among several traits on phylogenetic trees. Methods. Ecol. Evol. 8:1920–1927.

Cascini, M., K. J. Mitchell, A. Cooper, and M. J. Phillips. 2019. Reconstructing the evolution of giant extinct kangaroos: comparing the utility of dna, morphology, and total evidence. Syst. Biol. 68:520–537.

Chifman, J. and L. Kubatko. 2015. Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites. J. Theor. Biol. 374:35–47.

Christiansen, P. 2013. Phylogeny of the sabertoothed felids (Carnivora: Felidae: Machairodontinae). Cladistics 29:543–559.

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

Clavel, J. and H. Morlon. 2017. Accelerated body size evolution during cold climatic periods in the cenozoic. Proc. Natl. Acad. Sci. U.S.A 114:4183–4188.

Condamine, F. L., J. Rolland, and H. Morlon. 2013. Macroevolutionary perspectives to environmental change. Ecol. Lett. 16:72–85.

Condamine, F. L., J. Rolland, and H. Morlon. 2019. Assessing the causes of diversification slowdowns: temperature-dependent and diversity-dependent models receive equivalent support. Ecol. Lett. 22:1900–1912.

Cooper, A. and H. N. Poinar. 2000. Ancient DNA: do it right or not at all. Science 289.

Cunningham, J. A., I. A. Rahman, S. Lautenschlager, E. J. Rayfield, and P. C. J. Donoghue. 2015. A virtual world of paleontology. Trends Ecol. Evol. 29:347–357.

Danforth, B. N., S. Sipes, J. Fang, and S. G. Brady. 2006. The history of early bee diversification based on five genes plus morphology. Proc. Natl. Acad. Sci. U.S.A 103:15118–15123.

de Alencar, L. R. V., M. Martins, G. Burin, and T. B. Quental. 2017. Arboreality constrains morphological evolution but not species diversification in vipers. Proc. Royal Soc. B. 284:20171775.

Degnan, J. H. and L. A. Salter. 2007. Gene tree distributions under the coalescent process. Evolution 59:24–37.

Drummond, A. J., O. G. Pybus, A. Rambaut, R. Forsberg, and A. G. Rodrigo. 2002. Measurably evolving populations. Trends Ecol. Evol. 18:481–488.

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

Fan, Y., R. Wu, M.-H. Chen, L. Kuo, and P. O. Lewis. 2011. Choosing among partition models in bayesian phylogenetics. Mol. Biol. Evol. 28:523–532.

Farias, I. P., G. Ortí, and A. Meyer. 2000. Total evidence: molecules, morphology, and the phylogenetics of cichlid fishes. J. Exp. Zool. 288:76–92.

Felsenstein, J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. Am. J. Hum. Genet. 25:471–492.

Felsenstein, J. 1981. Evolutionary trees from gene frequencies and quantitative characters: finding maximum likelihood estimates. Evolution 35:1229–1242.

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

Felsenstein, J. 2005. Using the quantitative genetic threshold model for inferences between and within species. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 360:1427–1434.

Field, D. J., J. S. Berv, A. Y. Hsiang, R. Lanfear, M. J. Landis, and A. Dornburg. 2020. Timing the extant avian radiation: the rise of modern birds, and the importance of modeling molecular rate variation. Pages 159–181 *in* Pennaraptoran theropod dinosaurs: past progress and new frontiers (M. Pittman and X. Xu, eds.) vol. 440. Bulletin of the American Museum of Natural History, New York, NY.

Finarelli, J. A. and A. Goswami. 2009. The evolution of orbit orientation and encephalization in the Carnivora (Mammalia). J. Anat. 214:671–678.

Flynn, J. J., J. A. Finarelli, and M. Spaulding. 2010. Phylogeny of the Carnivora and Carnivoramorpha, and the use of the fossil record to enhance understanding of evolutionary transformations Page 25–63. Cambridge Studies in Morphology and Molecules: New Paradigms in Evolutionary Bio Cambridge University Press.

Freckleton, R. P. 2012. Fast likelihood calculations for comparative analyses. Methods Ecol. Evol. 3:940–947.

Fuentes-G., J. A., P. D. Polly, and E. P. Martins. 2020. A bayesian extension of phylogenetic generalized least squares: Incorporating uncertainty in the comparative study of trait relationships and evolutionary rates. Evolution 74:311–325.

Gaboriau, T., F. K. Mendes, S. Joly, D. Silvestro, and N. Salamin. 2020. A multi-platform package for the analysis of intra- and interspecific trait evolution. Methods Ecol. Evol. Pages 1–9.

Garamszegi, L. Z. and A. P. Møller. 2010. Effects of sample size and intraspecific variation in phylogenetic comparative studies: a meta-analytic review. Biol. Rev. 85:797–805.

Gavryushkina, A., T. A. Heath, D. T. Ksepka, T. Stadler, D. Welch, and A. J. Drummond. 2017. Bayesian total-evidence dating reveals the recent crown radiation of penguins. Syst. Biol. 66:57–73.

Gavryushkina, A., D. Welch, T. Stadler, and A. J. Drummond. 2014. Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. PLoS Comput. Biol. 10:e1003919.

Gibson, A. K. and J. A. Fuentes-G. 2015. A phylogenetic test of the Red Queen hypothesis: outcrossing and parasitism in the nematode phylum. Evolution 69:530–540.

Gillman, L. N., D. J. Keeling, H. A. Ross, and S. D. Wright. 2009. Latitude, elevation and the tempo of molecular evolution in mammals. Proc. Royal Soc. B 276:3353–3359.

Giribet, G., G. D. Edgecombe, and W. C. Wheeler. 2001. Arthropod phylogeny based on eight molecular loci and morphology. Nature 413:157–161.

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

Goloboff, P. A., C. I. Mattoni, and A. S. Quinteros. 2006. Continuous characters analyzed as such. Cladistics 22:589–601.

Goloboff, P. A., M. Pittman, D. Pol, and X. Xu. 2019. Morphological data sets fit a common mechanism much more poorly than DNA sequences and call into question the Mkv model. Syst. Biol. 68:494–504.

Goolsby, E. W. 2016. Likelihood-based parameter estimation for high-dimensional phylogenetic comparative models: overcoming the limitations of “distance-based” methods. Syst. Biol. 65:852–870.

Goswami, A., J. B. Smaers, C. Soligo, and P. D. Polly. 2014. The macroevolutionary consequences of phenotypic integration: from development to deep time. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 369:20130254.

Gower, J. C. 1975. Generalized procrustes analysis. Psychometrika 40:33–51.

Grant, T., D. R. Frost, J. P. Caldwell, R. Gagliardo, C. F. Haddad, P. J. Kok, D. B. Means, B. P. Noonan, W. E. Schargel, and W. C. Wheeler. 2006. Phylogenetic systematics of dart-poison frogs and their relatives (Amphibia: Athesphatanura: Dendrobatidae). Bull. Am. Mus. Nat. 2006:1–262.

Hagelberg, E., M. Hofreiter, and C. Keyser. 2015. Ancient DNA: the first three decades. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 370:20130371.

Hahn, M. W. and L. Nakhleh. 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.

Hansen, T. F. and K. Bartoszek. 2012. Interpreting the evolutionary regression: the interplay between observational and biological errors in phylogenetic comparative studies. Syst. Biol. 61:413–425.

Hansen, T. F. and E. P. Martins. 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. Evolution 50:1404–1417.

Harmon, L. J., J. B. Losos, T. Jonathan Davies, R. G. Gillespie, J. L. Gittleman, W. Bryan Jennings, K. H. Kozak, M. A. McPeek, F. Moreno-Roark, T. J. Near, et al. 2010. Early bursts of body size and shape evolution are rare in comparative data. Evolution 64:2385–2396.

Harmon, L. J., J. T. Weir, C. D. Brock, R. E. Glor, and W. Challenger. 2008. GEIGER: investigating evolutionary radiations. Bioinformatics 24:129–131.

Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol Evol. 22:160–174.

Heath, T. A., J. P. Huelsenbeck, and T. Stadler. 2014. The fossilized birth–death process for coherent calibration of divergence-time estimates. Proc. Natl. Acad. Sci. U.S.A 111:E2957–E2966.

Heled, J. and A. J. Drummond. 2012. Calibrated tree priors for relaxed phylogenetics and divergence time estimation. Syst. Biol. 61:138–149.

Ho, S. Y. W. and S. Duchêne. 2014. Molecular-clock methods for estimating evolutionary rates and timescales. Mol. Ecol. 23:5947–5965.

Ho, S. Y. W., S.-O. Kolokotronis, and R. G. Allaby. 2007. Elevated substitution rates estimated from ancient dna sequences. Biol. Lett. 3:702–705.

Ho, S. Y. W. and M. J. Phillips. 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. Syst. Biol. 58:367–380.

Ho, S. Y. W., M. J. Phillips, A. Cooper, and A. J. Drummond. 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. Mol. Biol. Evol. 22:1561–1568.

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

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

Jarvis, E. D., S. Mirarab, A. J. Aberer, B. Li, P. Houde, C. Li, S. Y. Ho, B. C. Faircloth, B. Nabholz, J. T. Howard, et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science 346:1320–1331.

Kass, R. E. and A. E. Raftery. 1995. Bayes factors. J. Am. Stat. 90:773–795.

Kawahara, A. Y., D. Plotkin, M. Espeland, K. Meusemann, E. F. Toussaint, A. Donath, F. Gimnich, P. B. Frandsen, A. Zwick, M. dos Reis, et al. 2019. Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths. Proc. Natl. Acad. Sci. U.S.A 116:22657–22663.

Kluge, A. G. 1989. A concern for evidence and a phylogenetic hypothesis of relationships among Epicrates (Boidae, Serpentes). Syst. Zool. 38:7–25.

Koch, N. M. and J. R. Thompson. 2020. A total-evidence dated phylogeny of Echinoidea combining phylogenomic and paleontological data. Syst. Biol. Syaa069.

Kolaczkowski, B. and J. W. Thornton. 2004. Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. Nature 431:980–984.

Kong, A., M. L. Frigge, G. Masson, S. Besenbacher, P. Sulem, G. Magnusson, S. A. Gudjonsson, A. Sigurdsson, A. Jonasdottir, A. Jonasdottir, et al. 2012. Rate of de novo mutations and the importance of father's age to disease risk. Nature 488:471–475.

Kostikova, A., D. Silvestro, P. B. Pearman, and N. Salamin. 2016. Bridging inter-and intraspecific trait evolution with a hierarchical bayesian approach. Syst. Biol. 65:417–431.

Kuhner, M. K. and J. Felsenstein. 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. Mol. Biol. Evol. 11:459–468.

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

Lanfear, R., J. J. Welch, and L. Bromham. 2010. Watching the clock: studying variation in rates of molecular evolution between species. Trends Ecol. Evol. 25:495–503.

Larracuente, A. M., T. B. Sackton, A. J. Greenberg, A. Wong, N. D. Singh, D. Sturgill, Y. Zhang, B. Oliver, and A. G. Clark. 2008. Evolution of protein-coding genes in *drosophila*. Trends Genet. 24:114–123.

Lartillot, N. and H. Philippe. 2004. Bayesian phylogenetic software based on mixture models. Mol. Biol. Evol. 21:1095–1109.

Lartillot, N. and H. Philippe. 2006. Computing bayes factors using thermodynamic integration. Syst. Biol. 55:195–207.

Lee, M. S. Y. and A. Palci. 2015. Morphological phylogenetics in the genomics age. Curr. Biol. 25:R922–R929.

Lewis, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst. Biol. 50:913–925.

Li, W. 1997. Molecular Evolution. Sinauer Associates Inc, Sunderland, Massachusetts.

Lister, A. M. 2013. The role of behaviour in adaptive morphological evolution of African proboscideans. Nature 500:331–334.

Lynch, M. and B. Walsh. 1998. Genetics and analysis of quantitative traits vol. 1. Sinauer Sunderland, MA.

Mair, P., P. J. Groenen, and J. de Leeuw. 2019. More on multidimensional scaling and unfolding in R: smacof version 2. J. Stat. Softw. .

Manos, P. S., P. S. Soltis, D. E. Soltis, S. R. Manchester, S.-H. Oh, C. D. Bell, D. L. Dilcher, and D. E. Stone. 2007. Phylogeny of extant and fossil juglandaceae inferred from the integration of molecular and morphological data sets. Syst. Biol. 56:412–430.

Martin, A. P. and S. R. Palumbi. 1993. Body size, metabolic rate, generation time, and the molecular clock. Proc. Natl. Acad. Sci. U.S.A 90:4087–4091.

Martín-Serra, A., B. Figueirido, and P. Palmqvist. 2014. A three-dimensional analysis of the morphological evolution and locomotor behaviour of the carnivoran hind limb. BMC Evol. Biol. 14:129.

May, M. R. and B. R. Moore. 2020. A Bayesian approach for inferring the impact of a discrete character on rates of continuous-character evolution in the presence of background-rate variation. Syst. Biol. 69:530–544.

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

Mendes, F. K. and M. W. Hahn. 2018. Why concatenation fails near the anomaly zone. Syst. Biol. 67:158–169.

Mendes, F. K., Y. Hahn, and M. W. Hahn. 2016. Gene tree discordance can generate patterns of diminishing convergence over time. Mol. Biol. Evol. 33:3299–3307.

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

Mitov, V. and T. Stadler. 2019. Parallel likelihood calculation for phylogenetic comparative models: The SPLITT C++ library. Methods Ecol. Evol. 10:493–506.

Mitteroecker, P., P. Gunz, S. Windhager, and K. Schaefer. 2013. A brief review of shape, form, and allometry in geometric morphometrics, with applications to human facial morphology. HYSTRIX 24:59–66.

Morlon, H. 2014. Phylogenetic approaches for studying diversification. Ecol. Lett. 17:508–525.

Newton, M. A. and A. E. Raftery. 1994. Approximate Bayesian inference with the weighted likelihood bootstrap. J. R. Stat. Soc. Series B Stat. Methodol. 56:3–26.

Nguyen, L.-T., H. A. Schmidt, A. Von Haeseler, and B. Q. Minh. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32:268–274.

Nylander, J. A. A., F. Ronquist, J. P. Huelsenbeck, and J. Nieves-Aldrey. 2004. Bayesian phylogenetic analysis of combined data. Syst. Biol. 53:47–67.

Ogilvie, H. A., F. K. Mendes, T. G. Vaughan, N. J. Matzke, T. Stadler, D. Welch, and A. J. Drummond. 2021. Novel integrative modeling of molecules and morphology pinpoints Caninae evolution across timescales. bioRxiv .

O'Leary, M. A. 1999. Parsimony analysis of total evidence from extinct and extant taxa and the cetacean-artiodactyl question (mammalia, ungulata). Cladistics 15:315–330.

O'Leary, M. A. and S. G. Kaufman. 2011. MorphoBank: phylophenomics in the "cloud". Cladistics 27:1–9.

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

O'Reilly, J. E., M. N. Puttick, L. Parry, A. R. Tanner, J. E. Tarver, J. Fleming, D. Pisani, and P. C. J. Donoghue. 2016. Bayesian methods outperform parsimony but at the expense of precision in the estimation of phylogeny from discrete morphological data. Biol. Lett. 12:20160081.

O'Reilly, J. E., M. N. Puttick, D. Pisani, and P. C. J. Donoghue. 2018. Empirical realism of simulated data is more important than the model used to generate it: a reply to Goloboff *et al.* Paleontology 61:631–635.

Parham, J. F., P. C. J. Donoghue, C. J. Bell, T. D. Calway, J. J. Head, P. A. Holroyd, J. G. Inoue, R. B. Irmis, W. G. Joyce, D. T. Ksepka, J. S. L. Patané, N. D. Smith, J. E.

Tarver, M. V. Tuinen, Z. Yang, K. D. Angielczyk, J. M. Greenwood, C. A. Hipsley, L. Jacobs, P. J. Makovicky, J. Müller, K. T. Smith, J. M. Theodor, R. C. M. Warnock, and M. J. Benton. 2012. Best practices for justifying fossil calibrations. Syst. Biol. 61:346–359.

Parins-Fukuchi, C. 2018a. Bayesian placement of fossils on phylogenies using quantitative morphometric data. Evolution 72:1801–1814.

Parins-Fukuchi, C. 2018b. Use of continuous traits can improve morphological phylogenetics. Syst. Biol. 67:328–339.

Pease, J. B., D. C. Haak, M. W. Hahn, and L. C. Moyle. 2016. Phylogenomics reveals three sources of adaptive variation during a rapid radiation. PLoS Biol. 14:e1002379.

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

Penny, D. 2005. Relativity for molecular clocks. Nature 436:183–184.

Perri, A. R., K. J. Mitchell, A. Mouton, et al. 2021. Dire wolves were the last of an ancient New World canid lineage. Nature 591:87–91.

Philippe, H., H. Brinkmann, D. V. Lavrov, D. T. J. Littlewood, M. Manuel, G. Wörheide, and D. Baurain. 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biol. 9:e1000602.

Polotow, D., A. Carmichael, and C. E. Griswold. 2015. Total evidence analysis of the phylogenetic relationships of Lycosoidea spiders (Araneae, Entelegynae). Invertebr. Syst. 29:124–163.

Pyron, R. A. 2011. Divergence time estimation using fossils as terminal taxa and the origins of Lissamphibia. Syst. Biol. 60:466–481.

Rambaut, A. 2000. Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into maximum likelihood phylogenies. Bioinformatics 16:395–399.

Rannala, B. and Z. Yang. 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. Genetics 164:1645–1656.

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

Rieux, A. and F. Balloux. 2016. Inferences from tip-calibrated phylogenies: a review and a practical guide. Mol. Ecol. 25:1911–1924.

Robinson, D. F. and L. R. Foulds. 1981. Comparison of phylogenetic trees. Math. Biosci. 53:131–147.

Rohlf, F. J. and D. Slice. 1990. Extensions of the Procrustes method for the optimal superimposition of landmarks. Syst. Biol. 39:40–59.

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

Ronquist, F., N. Lartillot, and M. J. Phillips. 2016. Closing the gap between rocks and clocks using total-evidence dating. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 371:20150136.

Russel, P. M., B. J. Brewer, S. Klaere, and R. R. Bouckaert. 2019. Model selection and parameter inference in phylogenetics using nested sampling. Syst. Biol. 68:219–233.

Sánchez-Reyes, L. L., H. Morlon, and S. Magallón. 2017. Uncovering higher-taxon diversification dynamics from clade age and species-richness data. Syst. Biol. 66:367–378.

Schäfer, J. and K. Strimmer. 2005. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. Stat. Appl. Genet. Mol. Biol. 4.

Schuh, R. T., C. Weirauch, and W. C. Wheeler. 2009. Phylogenetic relationships within the Cimicomorpha (Hemiptera: Heteroptera): a total-evidence analysis. Syst. Entomol. 34:15–48.

Seiffert, E. R. 2007. A new estimate of afrotherian phylogeny based on simultaneous analysis of genomic, morphological, and fossil evidence. BMC Evol. Biol. 7:1–13.

Silvestro, D., M. F. Tejedor, O. L. Martha L Serrano-Serrano, V. Rossier, J. Rolland, A. Zizka, S. Höhna, A. Antonelli, and N. Salamin. 2018. Early arrival and climatically-linked geographic expansion of New World monkeys from tiny African ancestors. Syst. Biol. 68:78–92.

Skilling, J. et al. 2006. Nested sampling for general bayesian computation. Bayesian Anal. 1:833–859.

Slater, G. J. 2015. Iterative adaptive radiations of fossil canids show no evidence for diversity-dependent trait evolution. Proc. Natl. Acad. Sci. U.S.A 112:4897–4902.

Smith, S. and M. Donoghue. 2008. Rates of molecular evolution are linked to life history in flowering plants. Science 322:86–89.

Stadler, T. 2010. Sampling-through-time in birth–death trees. J. Theor. Biol. 267:396–404.

Stadler, T. and Z. Yang. 2013. Dating phylogenies with sequentially sampled tips. Syst. Biol. 62:674–688.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.

Suchard, M. A., P. Lemey, G. Baele, D. L. Ayres, A. J. Drummond, and A. Rambaut. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. 4:vey016.

Suh, A., L. Smeds, and H. Ellegren. 2015. The dynamics of incomplete lineage sorting across the ancient adaptive radiation of neoavian birds. PLoS Biol. 13:e1002224.

Sullivan, J. and D. L. Swofford. 1997. Are guinea pigs rodents? The importance of adequate models in molecular phylogenetics. J. Mamm. Evol. 4:77–86.

Tedford, R. H., X. Wang, and B. E. Taylor. 2009. Phylogenetic systematics of the North American fossil caninae (Carnivora: Canidae). Bull. Am. Mus. Nat. 2009:1–218.

Thorne, J. L. and H. Kishino. 2005. Estimation of divergence times from molecular sequence data. Pages 233–256 *in* Statistical methods in molecular evolution (R. Nielsen, ed.). Springer, New York, NY.

Turner, A. and M. Antón. 1997. The big cats and their fossil relatives: an illustrated guide to their evolution and natural history. Columbia University Press, New York.

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

Vanderpool, D., B. Q. Minh, R. Lanfear, D. Hughes, S. Murali, R. A. Harris, M. Raveendran, D. M. Muzny, R. A. Gibbs, K. C. Worley, J. Rogers, and M. W. Hahn. 2020. Primate phylogenomics uncovers multiple rapid radiations and ancient interspecific introgression. PLoS Biol. 18:e3000954.

Varón-González, C., S. Whelan, and C. P. Klingenberg. 2020. Estimating phylogenies from shape and similar multidimensional data: why it is not reliable. Syst. Biol. 69:863–883.

Venditti, C. and M. Pagel. 2010. Speciation as an active force in promoting genetic evolution. Science 25:14–20.

Wang, R. J., G. W. Thomas, M. Raveendran, R. A. Harris, H. Doddapaneni, D. M. Muzny, J. P. Capitanio, P. Radivojac, J. Rogers, and M. W. Hahn. 2020. Paternal age in rhesus macaques is positively associated with germline mutation accumulation but not with measures of offspring sociability. Genome Res. 30:826–834.

Warnock, R. C. M., Z. Yang, and P. C. J. Donoghue. 2011. Exploring uncertainty in the calibration of the molecular clock. Biol. Lett. 8:156–159.

Webster, A., R. Payne, and M. Pagel. 2003. Molecular phylogenies link rates of evolution and speciation. Science 301:478.

Welch, J. J., E. Fontanillas, and L. Bromham. 2005. Molecular dates for the "Cambrian explosion": the influence of prior assumptions. Syst. Biol. 54:672–678.

Werdelin, L. 1996. Carnivoran ecomorphology: a phylogenetic perspective. Pages 582–624 *in* Carnivore Behavior, Ecology and Evolution (J. Gittleman, ed.) vol. 2. Cornell University Press, Ithaca.

Werdelin, L. and N. Solounias. 1991. The Hyaenidae: taxonomy, systematics and evolution. Foss. Strat. 30:1–104.

Werdelin, L., N. Yamaguchi, W. E. Johnson, and S. J. O'Brien. 2010. Phylogeny and evolution of cats (Felidae). Oxford University Press, Oxford.

Witt, C. and R. Brumfield. 2003. Comment on molecular phylogenies link rates of evolution and speciation. Science 303:173b.

Wright, S. 1934. An analysis of variability in the number of digits in an inbred strain of guinea pigs. Genetics 19:506–536.

Wright, S., J. Keeling, and L. Gillman. 2006. The road from Santa Rosalia: a faster tempo of evolution in tropical climates. Proc. Natl. Acad. Sci. U.S.A 103:7718–7722.

Xie, W., P. O. Lewis, Y. Fan, L. Kuo, and M.-H. Chen. 2011. Improving marginal likelihood estimation for bayesian phylogenetic model selection. Syst. Biol. 60:150–160.

Yang, Z. 1994. Maximum likelihood phylogenetic estimation from dna sequences with variable rates over sites: approximate methods. J. Mol. Evol. 39:306–314.

Yang, Z. 2006. Computational molecular evolution. Oxford University Press, Oxford, United Kingdom.

Zhang, C., M. Rabiee, E. Sayyari, and S. Mirarab. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinf. 19:153.

Zhang, G., C. Li, Q. Li, B. Li, D. M. Larkin, C. Lee, J. F. Storz, A. Antunes, M. J. Greenwold, R. W. Meredith, et al. 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. Science 346:1311–1320.

Zuckerkandl, E. and L. Pauling. 1965. Molecules as documents of evolutionary history. J. Theor. Biol. 8:357–366.