

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

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

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

17 The advent of molecular sequencing has unquestionably revolutionized comparative
18 biology, giving phylogeneticists unprecedented power to recover species relationships and
19 date important evolutionary events (e.g., Jarvis et al., 2014; Zhang et al., 2014; Suh et al.,
20 2015; Pease et al., 2016; Kawahara et al., 2019; Vanderpool et al., 2020), describe drivers of
21 diversification (Condamine et al., 2013; Morlon, 2014; Sánchez-Reyes et al., 2017;
22 Condamine et al., 2019), and their relationship with ecologically relevant traits (Goldberg
23 and Igić, 2012; Burin et al., 2016; de Alencar et al., 2017). The accumulation of genomic
24 data further allowed the identification of problems or gaps in molecular evolution models
25 (or their usage; e.g., Sullivan and Swofford 1997; Kolaczkowski and Thornton 2004;
26 Mendes and Hahn 2018), which led to improvements in their realism (Yang, 2006; Rannala
27 and Yang, 2003; Degnan and Salter, 2007), as well as the development of a plethora of
28 computational tools for empiricists wishing to use such models (e.g., Lartillot and
29 Philippe, 2004; Stamatakis, 2014; Nguyen et al., 2015; Chifman and Kubatko, 2015; Höhna
30 et al., 2016; Zhang et al., 2018; Suchard et al., 2018; Bouckaert et al., 2019).

31 Despite all progress, abundant genomic sequences and more complex substitution
32 models have not been a panacea for phylogenetic studies, in which species trees measured
33 in absolute time are either the ultimate goal (Philippe et al., 2011) or a critical ingredient
34 for downstream analyses (Felsenstein, 1985; Uyeda et al., 2018). First, while molecular
35 data informs us on the relative timing of evolutionary events, inferring mutation rates
36 remains challenging (Kong et al., 2012; Besenbacher et al., 2015; Wang et al., 2020), as
37 does reconciling estimates obtained at different evolutionary timescales (Ho et al., 2005;
38 Penny, 2005; Ho et al., 2007). Second, dating the tree of life in absolute time is
39 complicated by the absence of a universal strict molecular clock (Zuckerandl and Pauling,
40 1965; Ayala, 1997; Lanfear et al., 2010). Molecular rates have been shown to vary among
41 loci and species (Li, 1997; Larracuenta et al., 2008; Bromham, 2009), and to correlate with
42 phenotypic and natural history traits (Martin and Palumbi, 1993; Smith and Donoghue,
43 2008), the environment (Bleiweiss, 1998; Wright et al., 2006; Gillman et al., 2009), and

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

53 As a reaction to these findings, the past few decades saw improved descriptions of
54 the substitution process from more realistic clock models (Thorne and Kishino, 2005; Ho
55 and Duchêne, 2014), as well as the development of methods for calibrating substitution
56 trees into time-scaled trees. “Node dating” (as dubbed by Ronquist et al. 2012), for
57 example, refers to a collection of techniques whereby a specialist determines an age (range)
58 for a node using fossil occurrence or biogeographical data (Ho and Phillips, 2009). Node
59 dating is complicated by the difficulty in estimating the age of fossils, choosing which
60 fossils to use (Parham et al., 2012) – in many cases information is lost because younger
61 fossils of a group are excluded in favor of the oldest one – and what nodes to assign them
62 to, and choosing probability distributions for their age ranges, a crucial ad hoc step that
63 can introduce bias and circularity to an analysis (Warnock et al., 2011; Field et al., 2020).
64 These issues are further compounded by the analysis sensitivity to node-time priors (Welch
65 et al., 2005), unclear implicit prior probabilities on node times (Heled and Drummond,
66 2012), and overly simplistic molecular clock models (Berv and Field, 2018).

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

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

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

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

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

113 While many computational methods exist for the study of morphological character
114 evolution (e.g., Revell, 2012; Pennell et al., 2014; Clavel et al., 2015; Caetano and Harmon,
115 2017; Mitov et al., 2020), tools capable of jointly modeling molecular and morphological
116 characters are still lacking, particularly those that simultaneously account for uncertainty
117 in species tree topology and branch lengths. With few exceptions, comparative analysis of
118 morphological characters requires a species tree point estimate (e.g., Adams et al., 2009;
119 Lister, 2013; Gibson and Fuentes-G., 2015; or more rarely, a posterior distribution, e.g.,
120 Silvestro et al., 2018; Fuentes-G. et al., 2020) to be available and assumed as the truth.
121 Such species trees will have almost invariably been estimated in previous studies using
122 different data sets, often molecular ones. In such cases, the morphological data is then
123 analyzed on a phylogenetic “Procrustean bed”, a species tree that might not represent the
124 morphological evolutionary history (Hahn and Nakhleh, 2016). One way forward should be

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

127 Reasons why tools do not provide for this joint evolutionary modeling approach
128 include: (i) the technical difficulty of implementing multiple models efficiently under the
129 same statistical framework, (ii) prohibitively slow run times due to model scalability issues,
130 and (iii) lack of available data sets compiling appropriate data from multiple sources. Over
131 time we expect this last reason to become less of a hindrance and more of a motivation for
132 method development in this area. Recent work suggests, however, there is an immediate
133 demand for methods capable of integrating multidimensional data (e.g., Silvestro et al.,
134 2018; Cascini et al., 2019; Koch and Thompson, 2020), as well as work in progress to meet
135 those demands (e.g., Álvarez Carretero et al., 2019; May and Moore, 2020; Gaboriau et al.,
136 2020; Ogilvie et al., 2021).

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

146 MATERIALS AND METHODS

147 *Integrative model*

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

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

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

154 For the purposes of the present study, the morphological likelihood corresponds to the
 155 probability of observing \mathbf{M} under a phylogenetic BM model (Felsenstein, 1973), and the
 156 molecular likelihood to the probability of a observing \mathbf{S} under a molecular substitution
 157 model (Felsenstein, 1981). Finally, the tree prior $f(\Phi | \theta)$ gives the probability of a specific
 158 topology and node times in phylogenetic tree Φ , with $f(\theta)$ corresponding to the prior
 159 distribution on all remaining parameters $\theta = \{\mathbf{r}, c_m, \mathbf{y}_0, \lambda, \mu, \psi, \rho, t_{\text{mrca}}, \mathbf{c}_s, \boldsymbol{\kappa}, \boldsymbol{\pi}, \boldsymbol{\zeta}\}$ (see
 160 text below, Supplementary Table S11, and Fig. 1 for definitions; in Fig. 1,
 161 $\theta_\Phi = \{\lambda, \mu, \psi, \rho, t_{\text{mrca}}\}$ are the tree prior parameters, and $\theta_s = \{\boldsymbol{\kappa}, \boldsymbol{\zeta}\}$, as $\boldsymbol{\pi}$ is set to
 162 empirical values).

163 The posterior distribution $f(\Phi, \mathbf{b}_m, \mathbf{b}_s, \theta | \mathbf{M}, \mathbf{S})$ under our integrative model is
 164 approximated by Markov Chain Monte Carlo (MCMC) sampling in BEAST2 (Bouckaert
 165 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):

$$f(\mathbf{M}|\Phi, \mathbf{b}_m, \mathbf{r}, c_m, \mathbf{y}_0) = \frac{1}{(2\pi)^{nk/2} |\mathbf{V}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{M} - \mathbf{y}_0)^T \mathbf{V}^{-1}(\mathbf{M} - \mathbf{y}_0)\right), \quad (1.2)$$

where \mathbf{M} is an $n \times k$ matrix of observed continuous characters (n species, k traits), \mathbf{b}_m are the relative branch evolutionary rates, \mathbf{r} are the relative character-specific evolutionary rates, c_m is the global evolutionary rate, and \mathbf{y}_0 corresponds to the trait values from all species at the root of Φ . (Note that we unpack \mathbf{r} , c_m , and \mathbf{y}_0 from $\boldsymbol{\theta}$ in equation 1.1.)

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

$$t_{uw} = \sum_{z \in \text{Path}(u,w)} c_m \mathbf{b}_m^z \ell(z), \quad (1.3)$$

where u and w denote any two species in Φ , $\text{Path}(u, w)$ 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 \mathbf{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:

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

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

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

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

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

191 *Scalability with number of traits and species*

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

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

207 algorithm has been described and generalized in many subsequent studies (e.g., Felsenstein
208 1973; Freckleton 2012; Caetano and Harmon 2017; Álvarez Carretero et al. 2019; Mitov
209 et al. 2020), we point the interested reader to those references and to our supplementary
210 material for more detailed descriptions of the algorithm and a worked example.

211 More recently, Mitov et al. (2020) proposed a very general pruning-based solution
212 for calculating 1.2, as well as the probability density function of more general models, such
213 as BM with early bursts (Harmon et al., 2008) and accelerating or decelerating rates
214 (Blomberg et al., 2003), BM with trends (Hansen and Martins, 1996) and the
215 Ornstein-Uhlenbeck process (Hansen, 1997; Butler and King, 2004). Unlike the pruning
216 algorithm by Felsenstein (1973), the algorithm in Mitov et al. (2020) does not compute the
217 maximum-likelihood estimate of trait values at internal nodes of the tree, but instead
218 calculates a series of intermediate values (which gives this algorithm its flexibility; see Eq.
219 2 in Mitov et al. 2020). These intermediate values are then combined at the root node in
220 the calculation of an integral, which then gives the final probability density (Eq. 6 in Mitov
221 et al. 2020). Readers can find a detailed description of this algorithm in Mitov et al.
222 (2020), with it being put to use in Mitov and Stadler (2019). We provide a worked example
223 in the supplementary material.

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

234 Alternatively, although the non-independence among characters should always be
235 accounted for in some manner, it might be of secondary interest relative to the estimation
236 of a time-scaled phylogenetic tree. Different approaches have been explored or suggested
237 for such cases (e.g., Adams, 2014; Goolsby, 2016; Adams and Collyer, 2018), but one
238 simple solution is to employ the unbiased estimator of ρ , $\hat{\rho}$, obtained from multiple
239 character observations (e.g., multiple individuals) within a species in the phylogeny. (Note
240 that here this estimator is unbiased with respect to the population of a single species, and
241 by using $\hat{\rho}$ we are assuming trait correlations are the same across species and over time.)
242 Unfortunately, when one chooses to do the latter, a third non-obvious issue arises: as
243 $n \ll k$, the determinant of $\hat{\rho}$ will approach zero and $\hat{\rho}$ will be singular and non-invertible.
244 In such cases, it is impossible to evaluate equation 1.2.

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

$$\rho^* = \delta \mathbf{I} + (1 - \delta) \hat{\rho}, \quad (1.6)$$

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

250 *Tree models for total-evidence dating*

251 Modeling the evolution of molecular and morphological characters is crucial for
252 statistically sound taxon placement across and along a phylogeny. An integrative model for
253 total-evidence inference is nonetheless incomplete without accounting for the phylogenetic
254 process itself: the birth and death of lineages. Total-evidence dating, in particular, further
255 requires addressing the fossilization process underlying heterochronous data sets. The
256 fossilized birth-death process (FBD; Stadler, 2010; Heath et al., 2014; Gavryushkina et al.,
257 2014, 2017) is one tree model that has enjoyed success in the context of total-evidence
258 dating, due to its capacity to account for fossilization simultaneously with speciation and

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

264 *Software*

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

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

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

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

287 *Case study: Carnivora phylogeny*

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

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

297 We follow the same protocol in Álvarez Carretero et al. (2019) to prepare the
298 morphological data (cranium landmarks) for analysis, and start by “aligning” the
299 landmarks. In addition to size and shape, raw landmarks carry nuisance information about
300 position and orientation, which preclude their statistical analysis (Mitteroecker et al.,
301 2013). Distilling shape and size from raw landmarks can be done with Procrustes
302 superimposition, commonly used in biological shape analysis to “align” (superimpose)
303 landmarks (Mitteroecker et al., 2013). Procrustes superimposition consists of rotating,
304 translating and scaling landmark configurations relative to their centroid (i.e., their
305 average position) and its size, so as to minimize the Procrustes distance – a measure of
306 how different in shape two landmark configurations are (Gower, 1975; Rohlf and Slice,
307 1990). The Procrustes distance is given by the summed squared distance over landmarks
308 and their sample average position; if zero, then two landmarks have the same shape.

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

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

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

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

327 where \mathbf{v}_{err} is a $k \times k$ matrix given by the Hadamard product between the trait correlation
328 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
329 instead of $\boldsymbol{\rho}$ for the aforementioned reasons. The $k \times k$ matrix $\boldsymbol{\epsilon} = (\epsilon_{ij})$ is given by:

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

330 where σ_i^2 represents the intraspecific variance of character i .

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

333 the cost of longer MCMC chains. As in Álvarez Carretero et al. (2019), we nonetheless use
334 $\widehat{\sigma}_i^2$ calculated for all 87 landmarks from 21 individuals of *V. vulpes*. We also normalize each
335 i -th observed landmarks in \mathbf{M} by their corresponding $\widehat{\sigma}_i$, i.e., $\mathbf{M}^{(s)} = \mathbf{M} \times \text{diag}\{1/\sqrt{\widehat{\sigma}}\}$,
336 where $\widehat{\sigma}$ holds the unbiased estimators of intraspecific standard deviations (of all k
337 landmarks). Using $\mathbf{M}^{(s)}$ is convenient because this normalization leads to unit
338 (co-)variances, i.e., $\boldsymbol{\epsilon}$ becomes a matrix of ones, and then $\mathbf{v}_{\text{err}} = \boldsymbol{\rho}^*$.

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

348 Continuous morphological evolution was modelled with a phylogenetic BM model,
349 with all 87 superimposed, standardized characters (see above) sharing the same relative
350 evolutionary rate ($\mathbf{r} = \mathbf{1}$; Fig. 1). We estimated the landmark root values (\mathbf{y}_0) using
351 maximum-likelihood, obtaining $\widehat{\mathbf{y}}_0$ as a byproduct of pruning as done in Álvarez Carretero
352 et al., 2019. We did so first because this approach allows for direct comparison with results
353 from that study, and second because it is not immediately obvious how to choose a prior
354 for \mathbf{y}_0 . Using $\widehat{\mathbf{y}}_0$ is analogous to employing empirical nucleotide frequencies as the
355 equilibrium distribution under molecular substitution models. We note that both assuming
356 the same evolutionary rates across characters and calculating $\widehat{\mathbf{y}}_0$ are not requirements of
357 our implementation; $\mathbf{r} = \mathbf{1}$ can be relaxed, and \mathbf{y}_0 can be sampled during MCMC. Finally,
358 uncorrelated log-normal relaxed clock models were used for both molecular and

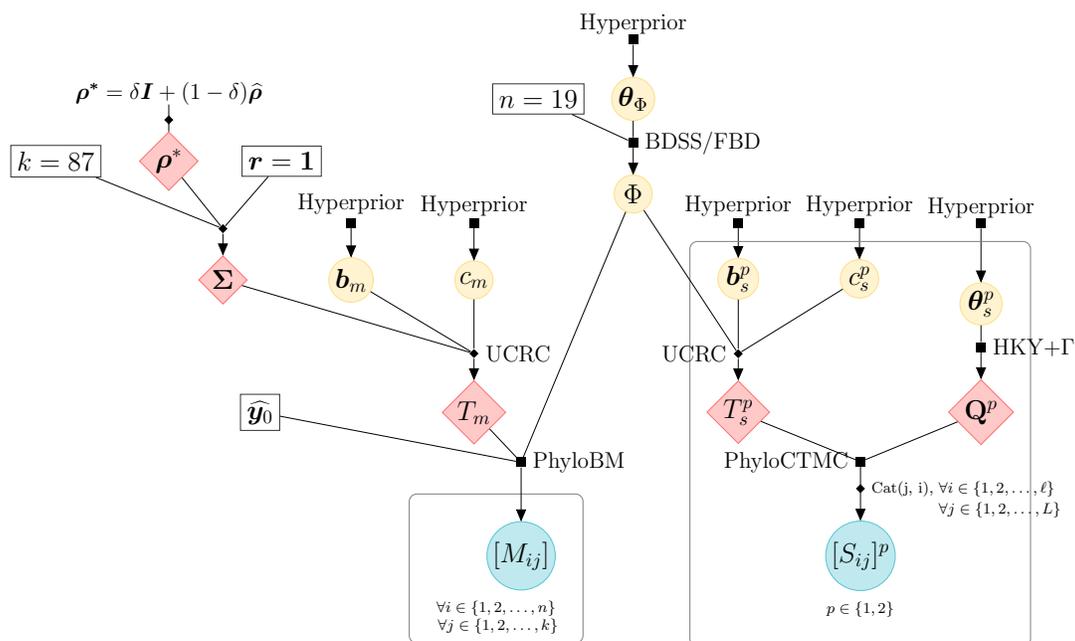


Figure 1. Probabilistic graphical model used in the analyses of the Carnivora data set, composed of 12 concatenated mitochondrial genes ($L = 12$) split into two partitions, and 87 continuous characters ($k = 87$) from 19 Carnivora species ($n = 19$). Filled squares denote sampling distributions, diamonds 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, $\theta = \{\theta_\Phi, \theta_s^p\}$ and ‘Hyperprior’ encompass all parameters and priors not explicitly shown in the graphical model.

359 morphological evolutionary rates.

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

Bayesian inference with Markov Chain Monte Carlo

370

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

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

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

382 By assuming both models have the same prior probability (i.e., prior odds = 1), the
383 posterior odds reduces to the ratio of the marginal likelihoods $\mathcal{Z}_0 = f(D|\mathcal{M}_0)$ and
384 $\mathcal{Z}_1 = f(D|\mathcal{M}_1)$. The marginal likelihood of a model is the probability of observing the data
385 under that model, and so it measures a model's goodness-of-fit. The ratio of two model's
386 marginal likelihoods, known as the Bayes factor (BF), thus allows one to quantify which of
387 the two has the best fit. A log-BF > 0 , for example, indicates \mathcal{M}_0 fits the data better; when
388 > 2 , the log-BF suggests \mathcal{M}_0 is decisively supported over \mathcal{M}_1 ; Kass and Raftery, 1995.

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

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

399 In order to calculate posterior odds (Equation 1.9), we carried out two NS analyses
400 (one for each model) with $K = 5,000$ and $N = 55$. Sub-chain lengths of 7,500 produced
401 statistically similar \mathcal{Z} estimates as compared to 5,000-long sub-chains (within twice the
402 sum of their standard deviations), so we deemed a length of 5,000 sufficient for covering
403 the bulk of the marginal likelihood. Given that the (NS) standard deviation of \mathcal{Z} estimates
404 is inversely proportional to the square root of N (Skilling et al., 2006), we adjusted N such
405 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	$\lambda, \mu, \psi, \rho, \Phi$	NA
4	BDSS	$\lambda, \mu, \psi, \rho, \Phi$	<i>Feliformia, Caniformia</i>
5	FBD	$\lambda, \mu, \psi, \rho, \Phi$	NA
6	FBD	$\lambda, \mu, \psi, \rho, \Phi$	<i>Feliformia, Caniformia</i>

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

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

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

419 RESULTS

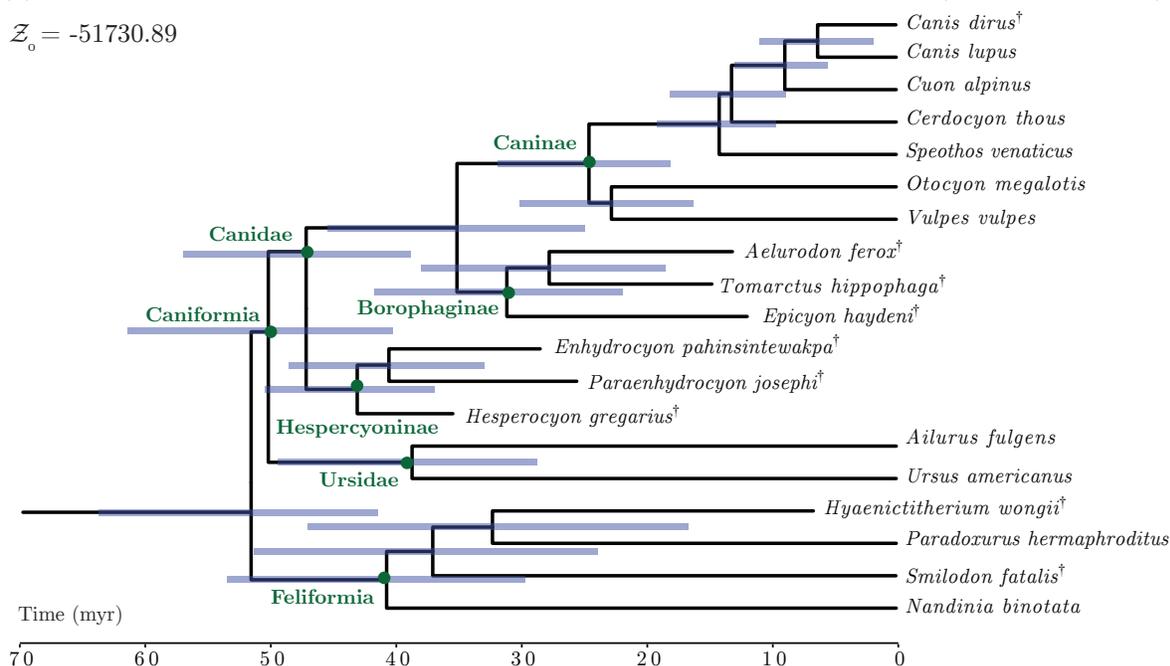
420 *Carnivora phylogeny*

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

429 In terms of fossil placement, there were four notable differences (red asterisks, Fig.
430 2b) between the reference tree topology and the MCC tree summarized from our posterior
431 samples under the BDSS tree prior, when estimating both topology and divergence times.
432 First, *Smilodon fatalis*, one of the extinct species commonly referred as “saber-toothed
433 cat”, was inferred to be more closely related to modern dogs and other canines than to
434 cat-like carnivoran species in Feliformia, the suborder *S. fatalis* is canonically assumed a
435 member of. Similarly, *Hyaenictitherium wongii*, a middle-sized hyaenid from the Late
436 Miocene (Werdelin and Solounias, 1991) was also estimated to be more closely related to
437 canines than to other feliforms. Third, while still placed within caniforms, *Enhydrocyon*

(a) Analysis 1: Fixed topology, sampling divergence times, BDSS tree prior (fixed parameters)

$$\mathcal{Z}_0 = -51730.89$$



(b) Analysis 2: Sampling topology and divergence times, BDSS tree prior (fixed parameters)

$$\mathcal{Z}_1 = -51691.82$$

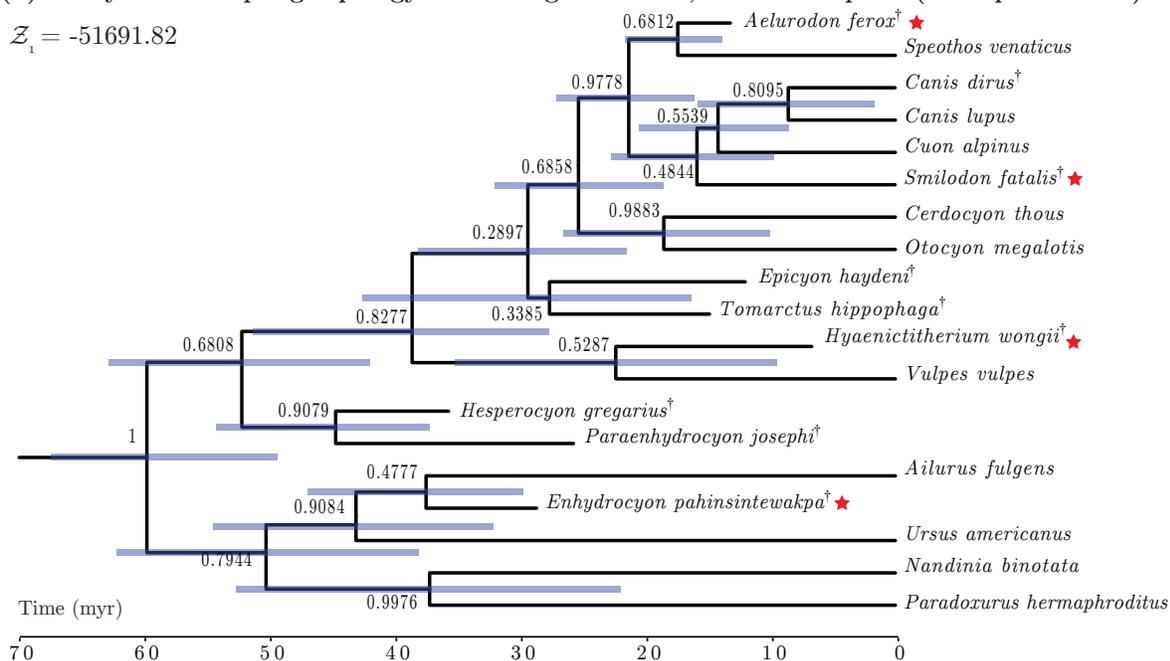


Figure 2. Carnivora maximum-clade-credibility summary trees. The horizontal bars at internal nodes show the 95% credible intervals for their times. Numbers by internal nodes indicate each clade's posterior probability. Carnivora suborders, families and subfamilies are labelled in green next to corresponding 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.

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

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

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

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

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

468 *Phylogenetic constraints and rogue fossils* One way to incorporate prior knowledge
469 about species relationships in phylogenetic inference is to constrain the monophyly of
470 specific groups. Holding the entire topology constant as done by Álvarez Carretero et al.
471 (2019), for example, takes this strategy to an extreme. The rationale behind monophyletic
472 constraints is that sometimes experts agree the veracity of that clade is beyond doubt. In
473 practice, using monophyletic priors might make sense when a researcher does not have easy
474 access to (or cannot use) the data upon which the confident monophyly belief is predicated.
475 Because three of the four “rogue” fossils (marked with a red star; Fig. 2b) were placed on
476 the wrong side of the root, in the wrong suborder, we reasoned constraining the monophyly
477 of Caniformia and Feliformia could help us further scrutinize the behavior of our model.

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

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

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

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

495 On the other hand, removing rogue fossils split the *Canis* genus and resulted in the
496 dire wolf, *Canis dirus*, being grouped with either taxa from Ursidae (Supplementary Fig.
497 18a) or from Feliformia (Supplementary Fig. 18b). In both cases, the placement of *C. dirus*
498 attained high posterior probability: PP=0.9515 with *A. fulgens* when removing three rogue
499 fossils, and PP=0.9867 with the remaining feliforms when dropping all rogue fossils.
500 Removal of problematic fossils also affected members of subfamily Borophaginae:
501 *Tomarctus hippophaga* was placed with high posterior probability among members of the
502 Caninae subfamily, and *Epicyon haydeni* grouped with species outside of Canidae at
503 moderate to high posterior probability (Supplementary Fig. 18a).

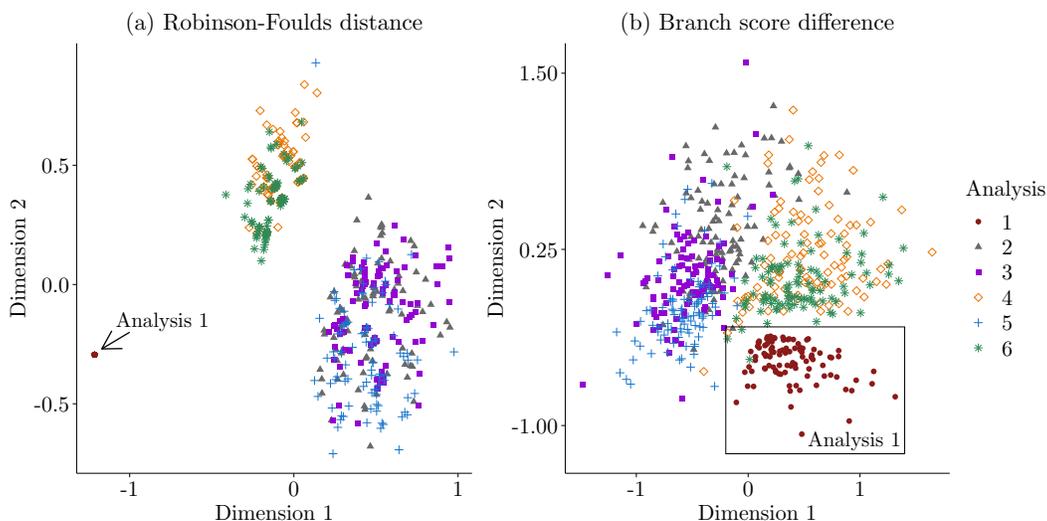


Figure 3. Comparison of Carnivora species tree posterior distributions using multidimensional scaling (MDS) of tree distance measures: (a) Robinson-Foulds distance, and (b) Branch score distance. Each point consists of one of 100 equally spaced (in the MCMC chain) posterior tree samples after discarding 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).

504 *Further comparison of species tree distributions and cranium landmarks*

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

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

523 DISCUSSION

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

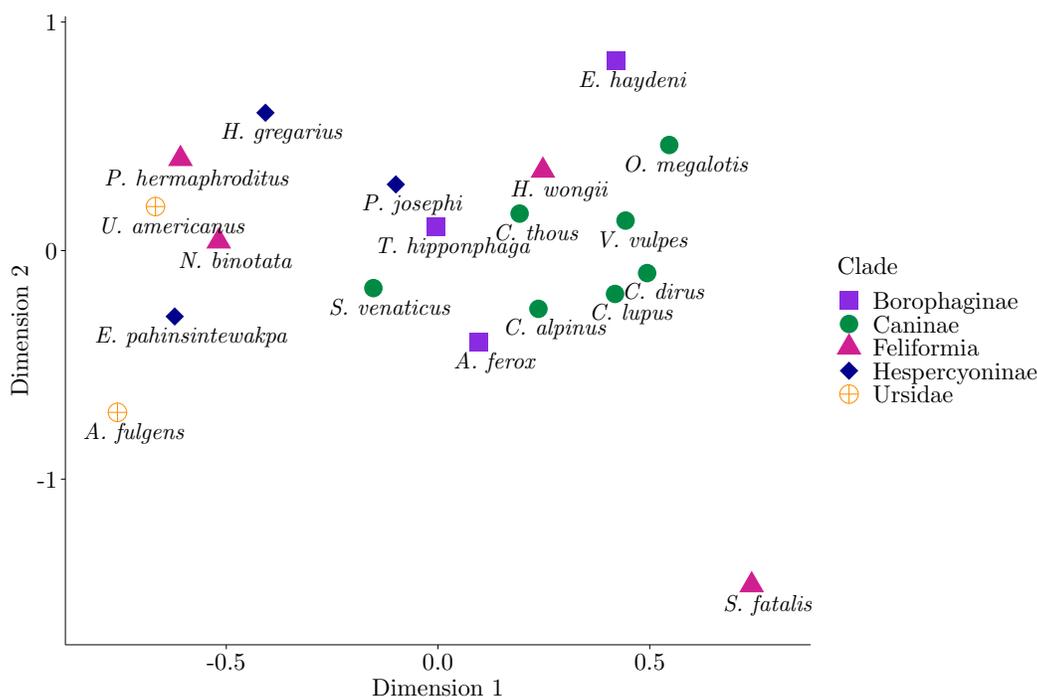


Figure 4. Multidimensional scaling (MDS) of Euclidean distances between Carnivora species cranium landmarks.

529 2015), and do not always model trait evolution statistically (e.g., O’Leary, 1999; Farias
530 et al., 2000; Seiffert, 2007; Schuh et al., 2009; Arrigo et al., 2013; Polotow et al., 2015).

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

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

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

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

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

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

574 When estimating the topology of the Carnivora species tree, the molecular
575 phylogenetic signal alone supported most clades with high posterior probability
576 (Supplementary Fig. 19a) as opposed to the much noisier morphological signal
577 (Supplementary Fig. 19b). This is unsurprising because molecular alignments harbor a
578 much larger number of topologically informative characters (Lee and Palci, 2015), and
579 because the molecular tree space was also smaller (we did not have DNA from fossils).
580 Neither data set decidedly supported the reference tree topology, however, with the
581 phylogenetic signal coming from molecules and morphology both agreeing and disagreeing
582 in different parts of the tree. For example, adding morphological data decreased the
583 posterior probability of *Canis* by 0.1 posterior probability units relative to the
584 molecules-only tree, while increasing the support of (*N. binotata*, *P. hermaphroditus*) by
585 0.0436 units.

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

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

605 At this point, we should clarify that the goal of our Carnivora analyses was not to
606 contribute to the understanding of this group's systematics, but to illustrate the
607 estimation of a species tree topology and divergence times from both molecular and many
608 continuous morphological traits. While *C. dirus* being more closely related to *C. lupus*
609 than to *Cuon alpinus* is plausible (Tedford et al., 2009; Slater, 2015; but see Perri et al.,
610 2021), *S. fatalis* earned the soubriquet “saber-toothed cat” – instead of “saber-toothed
611 canid” – for good reason (Werdelin, 1996; Turner and Antón, 1997; Anton et al., 2004;
612 Werdelin et al., 2010; Flynn et al., 2010; Christiansen, 2013). This latter result motivated
613 us to further explore the empirical consequences of excluding *S. fatalis* and other rogue
614 taxa, as well as employing monophyletic constraint priors. Neither strategy proved a silver
615 bullet in producing a match between our estimated species and the reference tree, but
616 some improvements were observed, especially with monophyletic constraints.

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

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

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

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

650 to make different kinds of morphological data easily available (O’Leary and Kaufman,
651 2011; Cunningham et al., 2015) have only started to scratch the surface of the tree of life’s
652 canopy. We are confident that methods such as ours will motivate the curation, expansion
653 and publication of rich morphological data sets, and fuel the probabilistic total evidence
654 paradigm.

655 SUPPLEMENTARY MATERIAL

656 Data available from the Dryad Digital Repository:

657 <https://doi.org/10.5061/dryad.YYYY>

658 ACKNOWLEDGEMENTS

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

664 COMPETING INTERESTS

665 The authors declare that they have no competing interests.

666

REFERENCES

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

669 Adams, D. C., C. M. Berns, K. H. Kozak, and J. J. Wiens. 2009. Are rates of species
670 diversification correlated with rates of morphological evolution? *Proc. Royal Soc. B*
671 276:2729–2738.

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

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

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

679 Arango, C. P. and W. C. Wheeler. 2007. Phylogeny of the sea spiders (arthropoda,
680 pycnogonida) based on direct optimization of six loci and morphology. *Cladistics*
681 23:255–293.

682 Arrigo, N., J. Therrien, C. L. Anderson, M. D. Windham, C. H. Haufler, and M. S. Barker.
683 2013. A total evidence approach to understanding phylogenetic relationships and
684 ecological diversity in *Selaginella* subg. *Tetragonostachys*. *Am. J. Bot.* 100:1672–1682.

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

688 Ayala, F. J. 1997. Vagaries of the molecular clock. *Proc. Natl. Acad. Sci. U.S.A*
689 94:7776–7783.

- 690 Bastide, P., C. Solís-Lemus, R. Kriebel, K. W. Sparks, and C. Ané. 2018. Phylogenetic
691 comparative methods on phylogenetic networks with reticulations. *Syst. Biol.*
692 67:800–820.
- 693 Berv, J. S. and D. J. Field. 2018. Genomic signature of an avian Lilliput effect across the
694 K-Pg extinction. *Syst. Biol.* 67:1–13.
- 695 Besenbacher, S., S. Liu, J. M. Izarzugaza, J. Grove, K. Belling, J. Bork-Jensen, S. Huang,
696 T. D. Als, S. Li, R. Yadav, et al. 2015. Novel variation and de novo mutation rates in
697 population-wide de novo assembled danish trios. *Nat. Commun.* 6:1–9.
- 698 Bleiweiss, R. 1998. Slow rate of molecular evolution in high-elevation hummingbirds. *Proc.*
699 *Natl. Acad. Sci. U.S.A* 95:612–16.
- 700 Blomberg, S. P., T. Garland Jr., and A. R. Ives. 2003. Testing for phylogenetic signal in
701 comparative data: behavioral traits are more labile. *Evolution* 57:717–745.
- 702 Blomberg, S. P., S. I. Rathnayake, and C. M. Moreau. 2020. Beyond Brownian motion and
703 the Ornstein-Uhlenbeck process: stochastic diffusion models for the evolution of
704 quantitative characters. *Am. Nat.* 195:145–165.
- 705 Bouckaert, R., T. G. Vaughan, J. Barido-Sottani, S. Duchêne, M. Fourment,
706 A. Gavryushkina, J. Heled, G. Jones, D. Kühnert, N. de Maio, et al. 2019. BEAST 2.5:
707 An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.*
708 15:e1006650.
- 709 Bracken-Grissom, H. D., S. T. Ahyong, R. D. Wilkinson, R. M. Feldmann, C. E.
710 Schweitzer, J. W. Breinholt, M. Bendall, F. Palero, T.-Y. Chan, D. L. Felder, et al.
711 2014. The emergence of lobsters: phylogenetic relationships, morphological evolution and
712 divergence time comparisons of an ancient group (Decapoda: Achelata, Astacidea,
713 Glypheidea, Polychelida). *Syst. Biol.* 63:457–479.

- 714 Bromham, L. 2009. Why do species vary in their rate of molecular evolution? *Biol. Lett.*
715 5:401–404.
- 716 Burin, G., W. D. Kissling, P. R. Guimarães, Ç. H. Şekercioğlu, and T. B. Quental. 2016.
717 Omnivory in birds is a macroevolutionary sink. *Nat. Commun.* 7:1–10.
- 718 Butler, M. A. and A. A. King. 2004. Phylogenetic comparative analysis: a modeling
719 approach for adaptive evolution. *Am. Nat.* 164:683–695.
- 720 Caetano, D. S. and L. J. Harmon. 2017. ratematrix: an R package for studying
721 evolutionary integration among several traits on phylogenetic trees. *Methods. Ecol. Evol.*
722 8:1920–1927.
- 723 Cascini, M., K. J. Mitchell, A. Cooper, and M. J. Phillips. 2019. Reconstructing the
724 evolution of giant extinct kangaroos: comparing the utility of dna, morphology, and total
725 evidence. *Syst. Biol.* 68:520–537.
- 726 Chifman, J. and L. Kubatko. 2015. Identifiability of the unrooted species tree topology
727 under the coalescent model with time-reversible substitution processes, site-specific rate
728 variation, and invariable sites. *J. Theor. Biol.* 374:35–47.
- 729 Christiansen, P. 2013. Phylogeny of the sabertoothed felids (Carnivora: Felidae:
730 Machairodontinae). *Cladistics* 29:543–559.
- 731 Clavel, J., G. Escarguel, and G. Merceron. 2015. mvMORPH: an R package for fitting
732 multivariate evolutionary models to morphometric data. *Methods Ecol. Evol.*
733 6:1311–1319.
- 734 Clavel, J. and H. Morlon. 2017. Accelerated body size evolution during cold climatic
735 periods in the cenozoic. *Proc. Natl. Acad. Sci. U.S.A* 114:4183–4188.
- 736 Condamine, F. L., J. Rolland, and H. Morlon. 2013. Macroevolutionary perspectives to
737 environmental change. *Ecol. Lett.* 16:72–85.

- 738 Condamine, F. L., J. Rolland, and H. Morlon. 2019. Assessing the causes of diversification
739 slowdowns: temperature-dependent and diversity-dependent models receive equivalent
740 support. *Ecol. Lett.* 22:1900–1912.
- 741 Cooper, A. and H. N. Poinar. 2000. Ancient DNA: do it right or not at all. *Science* 289.
- 742 Cunningham, J. A., I. A. Rahman, S. Lautenschlager, E. J. Rayfield, and P. C. J.
743 Donoghue. 2015. A virtual world of paleontology. *Trends Ecol. Evol.* 29:347–357.
- 744 Danforth, B. N., S. Sipes, J. Fang, and S. G. Brady. 2006. The history of early bee
745 diversification based on five genes plus morphology. *Proc. Natl. Acad. Sci. U.S.A*
746 103:15118–15123.
- 747 de Alencar, L. R. V., M. Martins, G. Burin, and T. B. Quental. 2017. Arboreality
748 constrains morphological evolution but not species diversification in vipers. *Proc. Royal*
749 *Soc. B.* 284:20171775.
- 750 Degnan, J. H. and L. A. Salter. 2007. Gene tree distributions under the coalescent process.
751 *Evolution* 59:24–37.
- 752 Drummond, A. J., O. G. Pybus, A. Rambaut, R. Forsberg, and A. G. Rodrigo. 2002.
753 Measurably evolving populations. *Trends Ecol. Evol.* 18:481–488.
- 754 Eastman, J. M., M. E. Alfaro, P. Joyce, A. L. Hipp, and L. J. Harmon. 2011. A novel
755 comparative method for identifying shifts in the rate of character evolution on trees.
756 *Evolution* 65:3578–3589.
- 757 Fan, Y., R. Wu, M.-H. Chen, L. Kuo, and P. O. Lewis. 2011. Choosing among partition
758 models in bayesian phylogenetics. *Mol. Biol. Evol.* 28:523–532.
- 759 Farias, I. P., G. Ortí, and A. Meyer. 2000. Total evidence: molecules, morphology, and the
760 phylogenetics of cichlid fishes. *J. Exp. Zool.* 288:76–92.
- 761 Felsenstein, J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous
762 characters. *Am. J. Hum. Genet.* 25:471–492.

- 763 Felsenstein, J. 1981. Evolutionary trees from gene frequencies and quantitative characters:
764 finding maximum likelihood estimates. *Evolution* 35:1229–1242.
- 765 Felsenstein, J. 1985. Phylogenies and the comparative method. *Am. Nat.* 125:1–15.
- 766 Felsenstein, J. 2005. Using the quantitative genetic threshold model for inferences between
767 and within species. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 360:1427–1434.
- 768 Field, D. J., J. S. Berv, A. Y. Hsiang, R. Lanfear, M. J. Landis, and A. Dornburg. 2020.
769 Timing the extant avian radiation: the rise of modern birds, and the importance of
770 modeling molecular rate variation. Pages 159–181 *in* Pennaraptoran theropod dinosaurs:
771 past progress and new frontiers (M. Pittman and X. Xu, eds.) vol. 440. Bulletin of the
772 American Museum of Natural History, New York, NY.
- 773 Finarelli, J. A. and A. Goswami. 2009. The evolution of orbit orientation and
774 encephalization in the Carnivora (Mammalia). *J. Anat.* 214:671–678.
- 775 Flynn, J. J., J. A. Finarelli, and M. Spaulding. 2010. Phylogeny of the Carnivora and
776 Carnivoramorpha, and the use of the fossil record to enhance understanding of
777 evolutionary transformations Page 25–63. *Cambridge Studies in Morphology and*
778 *Molecules: New Paradigms in Evolutionary Bio* Cambridge University Press.
- 779 Freckleton, R. P. 2012. Fast likelihood calculations for comparative analyses. *Methods*
780 *Ecol. Evol.* 3:940–947.
- 781 Fuentes-G., J. A., P. D. Polly, and E. P. Martins. 2020. A bayesian extension of
782 phylogenetic generalized least squares: Incorporating uncertainty in the comparative
783 study of trait relationships and evolutionary rates. *Evolution* 74:311–325.
- 784 Gaboriau, T., F. K. Mendes, S. Joly, D. Silvestro, and N. Salamin. 2020. A multi-platform
785 package for the analysis of intra- and interspecific trait evolution. *Methods Ecol. Evol.*
786 Pages 1–9.

- 787 Garamszegi, L. Z. and A. P. Møller. 2010. Effects of sample size and intraspecific variation
788 in phylogenetic comparative studies: a meta-analytic review. *Biol. Rev.* 85:797–805.
- 789 Gavryushkina, A., T. A. Heath, D. T. Ksepka, T. Stadler, D. Welch, and A. J. Drummond.
790 2017. Bayesian total-evidence dating reveals the recent crown radiation of penguins.
791 *Syst. Biol.* 66:57–73.
- 792 Gavryushkina, A., D. Welch, T. Stadler, and A. J. Drummond. 2014. Bayesian inference of
793 sampled ancestor trees for epidemiology and fossil calibration. *PLoS Comput. Biol.*
794 10:e1003919.
- 795 Gibson, A. K. and J. A. Fuentes-G. 2015. A phylogenetic test of the Red Queen
796 hypothesis: outcrossing and parasitism in the nematode phylum. *Evolution* 69:530–540.
- 797 Gillman, L. N., D. J. Keeling, H. A. Ross, and S. D. Wright. 2009. Latitude, elevation and
798 the tempo of molecular evolution in mammals. *Proc. Royal Soc. B* 276:3353–3359.
- 799 Giribet, G., G. D. Edgecombe, and W. C. Wheeler. 2001. Arthropod phylogeny based on
800 eight molecular loci and morphology. *Nature* 413:157–161.
- 801 Goldberg, E. E. and B. Igić. 2012. Tempo and mode in plant breeding system evolution.
802 *Evolution* 66:3701–3709.
- 803 Goloboff, P. A., C. I. Mattoni, and A. S. Quinteros. 2006. Continuous characters analyzed
804 as such. *Cladistics* 22:589–601.
- 805 Goloboff, P. A., M. Pittman, D. Pol, and X. Xu. 2019. Morphological data sets fit a
806 common mechanism much more poorly than DNA sequences and call into question the
807 Mk_v model. *Syst. Biol.* 68:494–504.
- 808 Goolsby, E. W. 2016. Likelihood-based parameter estimation for high-dimensional
809 phylogenetic comparative models: overcoming the limitations of “distance-based”
810 methods. *Syst. Biol.* 65:852–870.

- 811 Goswami, A., J. B. Smaers, C. Soligo, and P. D. Polly. 2014. The macroevolutionary
812 consequences of phenotypic integration: from development to deep time. *Philos. Trans.*
813 *R. Soc. Lond., B, Biol. Sci.* 369:20130254.
- 814 Gower, J. C. 1975. Generalized procrustes analysis. *Psychometrika* 40:33–51.
- 815 Grant, T., D. R. Frost, J. P. Caldwell, R. Gagliardo, C. F. Haddad, P. J. Kok, D. B.
816 Means, B. P. Noonan, W. E. Schargel, and W. C. Wheeler. 2006. Phylogenetic
817 systematics of dart-poison frogs and their relatives (Amphibia: Athesphatanura:
818 *Dendrobatidae*). *Bull. Am. Mus. Nat.* 2006:1–262.
- 819 Hagelberg, E., M. Hofreiter, and C. Keyser. 2015. Ancient DNA: the first three decades.
820 *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 370:20130371.
- 821 Hahn, M. W. and L. Nakhleh. 2016. Irrational exuberance for resolved species trees.
822 *Evolution* 70:7–17.
- 823 Hansen, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation.
824 *Evolution* 51:1341–1351.
- 825 Hansen, T. F. and K. Bartoszek. 2012. Interpreting the evolutionary regression: the
826 interplay between observational and biological errors in phylogenetic comparative
827 studies. *Syst. Biol.* 61:413–425.
- 828 Hansen, T. F. and E. P. Martins. 1996. Translating between microevolutionary process and
829 macroevolutionary patterns: the correlation structure of interspecific data. *Evolution*
830 50:1404–1417.
- 831 Harmon, L. J., J. B. Losos, T. Jonathan Davies, R. G. Gillespie, J. L. Gittleman,
832 W. Bryan Jennings, K. H. Kozak, M. A. McPeck, F. Moreno-Roark, T. J. Near, et al.
833 2010. Early bursts of body size and shape evolution are rare in comparative data.
834 *Evolution* 64:2385–2396.

- 835 Harmon, L. J., J. T. Weir, C. D. Brock, R. E. Glor, and W. Challenger. 2008. GEIGER:
836 investigating evolutionary radiations. *Bioinformatics* 24:129–131.
- 837 Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating of the human-ape splitting by a
838 molecular clock of mitochondrial DNA. *J. Mol Evol.* 22:160–174.
- 839 Heath, T. A., J. P. Huelsenbeck, and T. Stadler. 2014. The fossilized birth–death process
840 for coherent calibration of divergence-time estimates. *Proc. Natl. Acad. Sci. U.S.A*
841 111:E2957–E2966.
- 842 Heled, J. and A. J. Drummond. 2012. Calibrated tree priors for relaxed phylogenetics and
843 divergence time estimation. *Syst. Biol.* 61:138–149.
- 844 Ho, S. Y. W. and S. Duchêne. 2014. Molecular-clock methods for estimating evolutionary
845 rates and timescales. *Mol. Ecol.* 23:5947–5965.
- 846 Ho, S. Y. W., S.-O. Kolokotronis, and R. G. Allaby. 2007. Elevated substitution rates
847 estimated from ancient dna sequences. *Biol. Lett.* 3:702–705.
- 848 Ho, S. Y. W. and M. J. Phillips. 2009. Accounting for calibration uncertainty in
849 phylogenetic estimation of evolutionary divergence times. *Syst. Biol.* 58:367–380.
- 850 Ho, S. Y. W., M. J. Phillips, A. Cooper, and A. J. Drummond. 2005. Time dependency of
851 molecular rate estimates and systematic overestimation of recent divergence times. *Mol.*
852 *Biol. Evol.* 22:1561–1568.
- 853 Höhna, S., M. J. Landis, T. A. Heath, B. Boussau, N. Lartillot, B. R. Moore, J. P.
854 Huelsenbeck, and F. Ronquist. 2016. RevBayes: Bayesian phylogenetic inference using
855 graphical models and an interactive model-specification language. *Syst. Biol.* 65:726–736.
- 856
- 857 Ives, A. R., P. E. Midford, and T. Garland Jr. 2007. Within-species variation and
858 measurement error in phylogenetic comparative methods. *Syst. Biol.* 56:252–270.

- 859 Jarvis, E. D., S. Mirarab, A. J. Aberer, B. Li, P. Houde, C. Li, S. Y. Ho, B. C. Faircloth,
860 B. Nabholz, J. T. Howard, et al. 2014. Whole-genome analyses resolve early branches in
861 the tree of life of modern birds. *Science* 346:1320–1331.
- 862 Kass, R. E. and A. E. Raftery. 1995. Bayes factors. *J. Am. Stat.* 90:773–795.
- 863 Kawahara, A. Y., D. Plotkin, M. Espeland, K. Meusemann, E. F. Toussaint, A. Donath,
864 F. Gimnich, P. B. Frandsen, A. Zwick, M. dos Reis, et al. 2019. Phylogenomics reveals
865 the evolutionary timing and pattern of butterflies and moths. *Proc. Natl. Acad. Sci.*
866 U.S.A 116:22657–22663.
- 867 Kluge, A. G. 1989. A concern for evidence and a phylogenetic hypothesis of relationships
868 among *Epicrates* (Boidae, Serpentes). *Syst. Zool.* 38:7–25.
- 869 Koch, N. M. and J. R. Thompson. 2020. A total-evidence dated phylogeny of Echinoidea
870 combining phylogenomic and paleontological data. *Syst. Biol.* Syaa069.
- 871 Kolaczkowski, B. and J. W. Thornton. 2004. Performance of maximum parsimony and
872 likelihood phylogenetics when evolution is heterogeneous. *Nature* 431:980–984.
- 873 Kong, A., M. L. Frigge, G. Masson, S. Besenbacher, P. Sulem, G. Magnusson, S. A.
874 Gudjonsson, A. Sigurdsson, A. Jonasdottir, A. Jonasdottir, et al. 2012. Rate of de novo
875 mutations and the importance of father’s age to disease risk. *Nature* 488:471–475.
- 876 Kostikova, A., D. Silvestro, P. B. Pearman, and N. Salamin. 2016. Bridging inter-and
877 intraspecific trait evolution with a hierarchical bayesian approach. *Syst. Biol.*
878 65:417–431.
- 879 Kuhner, M. K. and J. Felsenstein. 1994. A simulation comparison of phylogeny algorithms
880 under equal and unequal evolutionary rates. *Mol. Biol. Evol.* 11:459–468.
- 881 Landis, M. J. and J. G. Schraiber. 2017. Pulsed evolution shaped modern vertebrate body
882 sizes. *Proc. Natl. Acad. Sci. U.S.A* 114:13224–13229.

- 883 Lanfear, R., J. J. Welch, and L. Bromham. 2010. Watching the clock: studying variation in
884 rates of molecular evolution between species. *Trends Ecol. Evol.* 25:495–503.
- 885 Larracuenta, A. M., T. B. Sackton, A. J. Greenberg, A. Wong, N. D. Singh, D. Sturgill,
886 Y. Zhang, B. Oliver, and A. G. Clark. 2008. Evolution of protein-coding genes in
887 *drosophila*. *Trends Genet.* 24:114–123.
- 888 Lartillot, N. and H. Philippe. 2004. Bayesian phylogenetic software based on mixture
889 models. *Mol. Biol. Evol.* 21:1095–1109.
- 890 Lartillot, N. and H. Philippe. 2006. Computing bayes factors using thermodynamic
891 integration. *Syst. Biol.* 55:195–207.
- 892 Lee, M. S. Y. and A. Palci. 2015. Morphological phylogenetics in the genomics age. *Curr.*
893 *Biol.* 25:R922–R929.
- 894 Lewis, P. O. 2001. A likelihood approach to estimating phylogeny from discrete
895 morphological character data. *Syst. Biol.* 50:913–925.
- 896 Li, W. 1997. *Molecular Evolution*. Sinauer Associates Inc, Sunderland, Massachusetts.
- 897 Lister, A. M. 2013. The role of behaviour in adaptive morphological evolution of African
898 proboscideans. *Nature* 500:331–334.
- 899 Lynch, M. and B. Walsh. 1998. *Genetics and analysis of quantitative traits vol. 1*. Sinauer
900 Sunderland, MA.
- 901 Mair, P., P. J. Groenen, and J. de Leeuw. 2019. More on multidimensional scaling and
902 unfolding in R: smacof version 2. *J. Stat. Softw.* .
- 903 Manos, P. S., P. S. Soltis, D. E. Soltis, S. R. Manchester, S.-H. Oh, C. D. Bell, D. L.
904 Dilcher, and D. E. Stone. 2007. Phylogeny of extant and fossil juglandaceae inferred
905 from the integration of molecular and morphological data sets. *Syst. Biol.* 56:412–430.

- 906 Martin, A. P. and S. R. Palumbi. 1993. Body size, metabolic rate, generation time, and the
907 molecular clock. *Proc. Natl. Acad. Sci. U.S.A* 90:4087–4091.
- 908 Martín-Serra, A., B. Figueirido, and P. Palmqvist. 2014. A three-dimensional analysis of
909 the morphological evolution and locomotor behaviour of the carnivoran hind limb. *BMC*
910 *Evol. Biol.* 14:129.
- 911 May, M. R. and B. R. Moore. 2020. A Bayesian approach for inferring the impact of a
912 discrete character on rates of continuous-character evolution in the presence of
913 background-rate variation. *Syst. Biol.* 69:530–544.
- 914 Mendes, F. K., J. A. Fuentes-González, J. G. Schraiber, and M. W. Hahn. 2018. A
915 multispecies coalescent model for quantitative traits. *eLife* 7:e36482.
- 916 Mendes, F. K. and M. W. Hahn. 2018. Why concatenation fails near the anomaly zone.
917 *Syst. Biol.* 67:158–169.
- 918 Mendes, F. K., Y. Hahn, and M. W. Hahn. 2016. Gene tree discordance can generate
919 patterns of diminishing convergence over time. *Mol. Biol. Evol.* 33:3299–3307.
- 920 Mitov, V., K. Bartoszek, G. Asimomitis, and T. Stadler. 2020. Fast likelihood calculation
921 for multivariate Gaussian phylogenetic models with shifts. *Theor. Popul. Biol.*
922 131:66–78.
- 923 Mitov, V. and T. Stadler. 2019. Parallel likelihood calculation for phylogenetic
924 comparative models: The SPLIT C++ library. *Methods Ecol. Evol.* 10:493–506.
- 925 Mitteroecker, P., P. Gunz, S. Windhager, and K. Schaefer. 2013. A brief review of shape,
926 form, and allometry in geometric morphometrics, with applications to human facial
927 morphology. *HYSTRIX* 24:59–66.
- 928 Morlon, H. 2014. Phylogenetic approaches for studying diversification. *Ecol. Lett.*
929 17:508–525.

- 930 Newton, M. A. and A. E. Raftery. 1994. Approximate Bayesian inference with the
931 weighted likelihood bootstrap. *J. R. Stat. Soc. Series B Stat. Methodol.* 56:3–26.
- 932 Nguyen, L.-T., H. A. Schmidt, A. Von Haeseler, and B. Q. Minh. 2015. IQ-TREE: a fast
933 and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol.*
934 *Biol. Evol.* 32:268–274.
- 935 Nylander, J. A. A., F. Ronquist, J. P. Huelsenbeck, and J. Nieves-Aldrey. 2004. Bayesian
936 phylogenetic analysis of combined data. *Syst. Biol.* 53:47–67.
- 937 Ogilvie, H. A., F. K. Mendes, T. G. Vaughan, N. J. Matzke, T. Stadler, D. Welch, and
938 A. J. Drummond. 2021. Novel integrative modeling of molecules and morphology
939 pinpoints Caninae evolution across timescales. *bioRxiv* .
- 940 O’Leary, M. A. 1999. Parsimony analysis of total evidence from extinct and extant taxa
941 and the cetacean-artiodactyl question (mammalia, ungulata). *Cladistics* 15:315–330.
- 942 O’Leary, M. A. and S. G. Kaufman. 2011. MorphoBank: phylophenomics in the “cloud”.
943 *Cladistics* 27:1–9.
- 944 O’Meara, B. C., C. Ané, M. J. Sanderson, and P. C. Wainwright. 2006. Testing for
945 different rates of continuous trait evolution using likelihood. *Evolution* 60:922–933.
- 946 O’Reilly, J. E., M. N. Puttick, L. Parry, A. R. Tanner, J. E. Tarver, J. Fleming, D. Pisani,
947 and P. C. J. Donoghue. 2016. Bayesian methods outperform parsimony but at the
948 expense of precision in the estimation of phylogeny from discrete morphological data.
949 *Biol. Lett.* 12:20160081.
- 950 O’Reilly, J. E., M. N. Puttick, D. Pisani, and P. C. J. Donoghue. 2018. Empirical realism
951 of simulated data is more important than the model used to generate it: a reply to
952 Goloboff *et al.* *Paleontology* 61:631–635.
- 953 Parham, J. F., P. C. J. Donoghue, C. J. Bell, T. D. Calway, J. J. Head, P. A. Holroyd,
954 J. G. Inoue, R. B. Irmis, W. G. Joyce, D. T. Ksepka, J. S. L. Patané, N. D. Smith, J. E.

- 955 Tarver, M. V. Tuinen, Z. Yang, K. D. Angielczyk, J. M. Greenwood, C. A. Hipsley,
956 L. Jacobs, P. J. Makovicky, J. Müller, K. T. Smith, J. M. Theodor, R. C. M. Warnock,
957 and M. J. Benton. 2012. Best practices for justifying fossil calibrations. *Syst. Biol.*
958 61:346–359.
- 959 Parins-Fukuchi, C. 2018a. Bayesian placement of fossils on phylogenies using quantitative
960 morphometric data. *Evolution* 72:1801–1814.
- 961 Parins-Fukuchi, C. 2018b. Use of continuous traits can improve morphological
962 phylogenetics. *Syst. Biol.* 67:328–339.
- 963 Pease, J. B., D. C. Haak, M. W. Hahn, and L. C. Moyle. 2016. Phylogenomics reveals
964 three sources of adaptive variation during a rapid radiation. *PLoS Biol.* 14:e1002379.
- 965 Pennell, M. W., J. M. Eastman, G. J. Slater, J. W. Brown, J. C. Uyeda, R. G. FitzJohn,
966 M. E. Alfaro, and L. J. Harmon. 2014. geiger v2.0: an expanded suite of methods for
967 fitting macroevolutionary models to phylogenetic trees. *Bioinformatics* 30:2216–2218.
- 968 Penny, D. 2005. Relativity for molecular clocks. *Nature* 436:183–184.
- 969 Perri, A. R., K. J. Mitchell, A. Mouton, et al. 2021. Dire wolves were the last of an ancient
970 New World canid lineage. *Nature* 591:87–91.
- 971 Philippe, H., H. Brinkmann, D. V. Lavrov, D. T. J. Littlewood, M. Manuel, G. Wörheide,
972 and D. Baurain. 2011. Resolving difficult phylogenetic questions: why more sequences
973 are not enough. *PLoS Biol.* 9:e1000602.
- 974 Polotow, D., A. Carmichael, and C. E. Griswold. 2015. Total evidence analysis of the
975 phylogenetic relationships of Lycosoidea spiders (Araneae, Entelegynae). *Invertebr. Syst.*
976 29:124–163.
- 977 Pyron, R. A. 2011. Divergence time estimation using fossils as terminal taxa and the
978 origins of Lissamphibia. *Syst. Biol.* 60:466–481.

- 979 Rambaut, A. 2000. Estimating the rate of molecular evolution: incorporating
980 non-contemporaneous sequences into maximum likelihood phylogenies. *Bioinformatics*
981 16:395–399.
- 982 Rannala, B. and Z. Yang. 2003. Bayes estimation of species divergence times and ancestral
983 population sizes using DNA sequences from multiple loci. *Genetics* 164:1645–1656.
- 984 Revell, L. J. 2012. phytools: An R package for phylogenetic comparative biology (and other
985 things). *Methods Ecol. Evol.* 3:217–223.
- 986 Rieux, A. and F. Balloux. 2016. Inferences from tip-calibrated phylogenies: a review and a
987 practical guide. *Mol. Ecol.* 25:1911–1924.
- 988 Robinson, D. F. and L. R. Foulds. 1981. Comparison of phylogenetic trees. *Math. Biosci.*
989 53:131–147.
- 990 Rohlf, F. J. and D. Slice. 1990. Extensions of the Procrustes method for the optimal
991 superimposition of landmarks. *Syst. Biol.* 39:40–59.
- 992 Ronquist, F., S. Klopfstein, L. Vilhelmsen, S. Schulmeister, D. L. Murray, and A. P.
993 Rasnitsyn. 2012. A total-evidence approach to dating with fossils, applied to the early
994 radiation of the Hymenoptera. *Syst. Biol.* 61:973–999.
- 995 Ronquist, F., N. Lartillot, and M. J. Phillips. 2016. Closing the gap between rocks and
996 clocks using total-evidence dating. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*
997 371:20150136.
- 998 Russel, P. M., B. J. Brewer, S. Klaere, and R. R. Bouckaert. 2019. Model selection and
999 parameter inference in phylogenetics using nested sampling. *Syst. Biol.* 68:219–233.
- 1000 Sánchez-Reyes, L. L., H. Morlon, and S. Magallón. 2017. Uncovering higher-taxon
1001 diversification dynamics from clade age and species-richness data. *Syst. Biol.* 66:367–378.

- 1003 Schäfer, J. and K. Strimmer. 2005. A shrinkage approach to large-scale covariance matrix
1004 estimation and implications for functional genomics. *Stat. Appl. Genet. Mol. Biol.* 4.
- 1005 Schuh, R. T., C. Weirauch, and W. C. Wheeler. 2009. Phylogenetic relationships within
1006 the Cimicomorpha (Hemiptera: Heteroptera): a total-evidence analysis. *Syst. Entomol.*
1007 34:15–48.
- 1008 Seiffert, E. R. 2007. A new estimate of afrotherian phylogeny based on simultaneous
1009 analysis of genomic, morphological, and fossil evidence. *BMC Evol. Biol.* 7:1–13.
- 1010 Silvestro, D., M. F. Tejedor, O. L. Martha L Serrano-Serrano, V. Rossier, J. Rolland,
1011 A. Zizka, S. Höhna, A. Antonelli, and N. Salamin. 2018. Early arrival and
1012 climatically-linked geographic expansion of New World monkeys from tiny African
1013 ancestors. *Syst. Biol.* 68:78–92.
- 1014 Skilling, J. et al. 2006. Nested sampling for general bayesian computation. *Bayesian Anal.*
1015 1:833–859.
- 1016 Slater, G. J. 2015. Iterative adaptive radiations of fossil canids show no evidence for
1017 diversity-dependent trait evolution. *Proc. Natl. Acad. Sci. U.S.A* 112:4897–4902.
- 1018 Smith, S. and M. Donoghue. 2008. Rates of molecular evolution are linked to life history in
1019 flowering plants. *Science* 322:86–89.
- 1020 Stadler, T. 2010. Sampling-through-time in birth–death trees. *J. Theor. Biol.* 267:396–404.
- 1021 Stadler, T. and Z. Yang. 2013. Dating phylogenies with sequentially sampled tips. *Syst.*
1022 *Biol.* 62:674–688.
- 1023 Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis
1024 of large phylogenies. *Bioinformatics* 30:1312–1313.
- 1025 Suchard, M. A., P. Lemey, G. Baele, D. L. Ayres, A. J. Drummond, and A. Rambaut.
1026 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10.
1027 *Virus Evol.* 4:vey016.

- 1028 Suh, A., L. Smeds, and H. Ellegren. 2015. The dynamics of incomplete lineage sorting
1029 across the ancient adaptive radiation of neoavian birds. *PLoS Biol.* 13:e1002224.
- 1030 Sullivan, J. and D. L. Swofford. 1997. Are guinea pigs rodents? The importance of
1031 adequate models in molecular phylogenetics. *J. Mamm. Evol.* 4:77–86.
- 1032 Tedford, R. H., X. Wang, and B. E. Taylor. 2009. Phylogenetic systematics of the North
1033 American fossil caninae (Carnivora: Canidae). *Bull. Am. Mus. Nat.* 2009:1–218.
- 1034 Thorne, J. L. and H. Kishino. 2005. Estimation of divergence times from molecular
1035 sequence data. Pages 233–256 *in* *Statistical methods in molecular evolution* (R. Nielsen,
1036 ed.). Springer, New York, NY.
- 1037 Turner, A. and M. Antón. 1997. *The big cats and their fossil relatives: an illustrated guide*
1038 *to their evolution and natural history*. Columbia University Press, New York.
- 1039 Uyeda, J. C. and L. J. Harmon. 2014. A novel bayesian method for inferring and
1040 interpreting the dynamics of adaptive landscapes from phylogenetic comparative data.
1041 *Syst. Biol.* 63:902–918.
- 1042 Uyeda, J. C., R. Zenil-Ferguson, and M. W. Pennell. 2018. Rethinking phylogenetic
1043 comparative methods. *Syst. Biol.* 67:1091–1109.
- 1044 Vanderpool, D., B. Q. Minh, R. Lanfear, D. Hughes, S. Murali, R. A. Harris,
1045 M. Raveendran, D. M. Muzny, R. A. Gibbs, K. C. Worley, J. Rogers, and M. W. Hahn.
1046 2020. Primate phylogenomics uncovers multiple rapid radiations and ancient
1047 interspecific introgression. *PLoS Biol.* 18:e3000954.
- 1048 Varón-González, C., S. Whelan, and C. P. Klingenberg. 2020. Estimating phylogenies from
1049 shape and similar multidimensional data: why it is not reliable. *Syst. Biol.* 69:863–883.
- 1050 Venditti, C. and M. Pagel. 2010. Speciation as an active force in promoting genetic
1051 evolution. *Science* 25:14–20.

- 1052 Wang, R. J., G. W. Thomas, M. Raveendran, R. A. Harris, H. Doddapaneni, D. M.
1053 Muzny, J. P. Capitanio, P. Radivojac, J. Rogers, and M. W. Hahn. 2020. Paternal age in
1054 rhesus macaques is positively associated with germline mutation accumulation but not
1055 with measures of offspring sociability. *Genome Res.* 30:826–834.
- 1056 Warnock, R. C. M., Z. Yang, and P. C. J. Donoghue. 2011. Exploring uncertainty in the
1057 calibration of the molecular clock. *Biol. Lett.* 8:156–159.
- 1058 Webster, A., R. Payne, and M. Pagel. 2003. Molecular phylogenies link rates of evolution
1059 and speciation. *Science* 301:478.
- 1060 Welch, J. J., E. Fontanillas, and L. Bromham. 2005. Molecular dates for the “Cambrian
1061 explosion”: the influence of prior assumptions. *Syst. Biol.* 54:672–678.
- 1062 Werdelin, L. 1996. Carnivoran ecomorphology: a phylogenetic perspective. Pages 582–624
1063 *in* *Carnivore Behavior, Ecology and Evolution* (J. Gittleman, ed.) vol. 2. Cornell
1064 University Press, Ithaca.
- 1065 Werdelin, L. and N. Solounias. 1991. The Hyaenidae: taxonomy, systematics and evolution.
1066 *Foss. Strat.* 30:1–104.
- 1067 Werdelin, L., N. Yamaguchi, W. E. Johnson, and S. J. O’Brien. 2010. Phylogeny and
1068 evolution of cats (Felidae). Oxford University Press, Oxford.
- 1069 Witt, C. and R. Brumfield. 2003. Comment on molecular phylogenies link rates of
1070 evolution and speciation. *Science* 303:173b.
- 1071 Wright, S. 1934. An analysis of variability in the number of digits in an inbred strain of
1072 guinea pigs. *Genetics* 19:506–536.
- 1073 Wright, S., J. Keeling, and L. Gillman. 2006. The road from Santa Rosalia: a faster tempo
1074 of evolution in tropical climates. *Proc. Natl. Acad. Sci. U.S.A* 103:7718–7722.
- 1075 Xie, W., P. O. Lewis, Y. Fan, L. Kuo, and M.-H. Chen. 2011. Improving marginal
1076 likelihood estimation for bayesian phylogenetic model selection. *Syst. Biol.* 60:150–160.

- 1077 Yang, Z. 1994. Maximum likelihood phylogenetic estimation from dna sequences with
1078 variable rates over sites: approximate methods. *J. Mol. Evol.* 39:306–314.
- 1079 Yang, Z. 2006. *Computational molecular evolution*. Oxford University Press, Oxford,
1080 United Kingdom.
- 1081 Zhang, C., M. Rabiee, E. Sayyari, and S. Mirarab. 2018. ASTRAL-III: polynomial time
1082 species tree reconstruction from partially resolved gene trees. *BMC Bioinf.* 19:153.
- 1083 Zhang, G., C. Li, Q. Li, B. Li, D. M. Larkin, C. Lee, J. F. Storz, A. Antunes, M. J.
1084 Greenwold, R. W. Meredith, et al. 2014. Comparative genomics reveals insights into
1085 avian genome evolution and adaptation. *Science* 346:1311–1320.
- 1086 Zuckerkandl, E. and L. Pauling. 1965. Molecules as documents of evolutionary history. *J.*
1087 *Theor. Biol.* 8:357–366.