Novel Integrative Modeling of Molecules and Morphology Pinpoints Caninae Evolution across Timescales

Huw A. Ogilvie[1,*,†], Fábio K. Mendes[2,3,4,†], Timothy G. Vaughan[5,6], Nicholas J. Matzke[2,4], Tanja Stadler[5,6], David Welch[2,3], and Alexei J. Drummond[2,3,4]

[1]Department of Computer Science, Rice University, Houston TX, 77005, USA

[2]Centre for Computational Evolution, The University of Auckland, Auckland, 1010, New Zealand

[3]School of Computer Science, The University of Auckland, Auckland, 1010, New Zealand

[4]School of Biological Sciences, The University of Auckland, Auckland, 1010, New Zealand

[5]Department of Biosystems Science and Engineering, ETH Zürich, Basel, 4058, Switzerland

[6]SIB Swiss Institute of Bioinformatics, Lausanne, 1015, Switzerland

[*]Correspondence to be sent to: Huw A. Ogilvie, Department of Computer Science – MS-132, Rice University, P.O. Box 1892, Houston, TX 77251-1892, USA; Email: huw.a.ogilvie@rice.edu

[†]These authors contributed equally.

Abstract

Evolutionary models account for either population- or species-level processes, but usually not both. We introduce a new model, the FBD-MSC, which makes it possible for the first time to integrate both the genealogical and fossilization phenomena, by means of the multispecies coalescent (MSC) and the fossilized birth-death (FBD) processes. Using our new model, we reconstruct the phylogeny representing all extant and many fossil Caninae, recovering both the relative and absolute time of speciation events. We quantify known inaccuracy issues with divergence time estimates using the popular strategy of concatenating molecular alignments, and show the FBD-MSC solves them. Our new integrative method and empirical results advance the paradigm and practice of probabilistic total evidence analyses in evolutionary biology.

Introduction

Thanks to increasing availability of molecular and paleontological data, obtaining a high-resolution picture of the tree of life has never been within closer reach. Treating this richness of data in a statistically explicit manner depends on the development of realistic and tractable evolutionary models. The desired result are phylogenies that not only recount how species are related, but that are scaled to absolute time so that species trees can be reconciled with geological and fossil records.

One method for scaling trees into absolute time is to assume a molecular clock [1]. But because a universal clock does not exist, extrapolating clock rates measured in one group of organisms to another can lead to unrealistic evolutionary time estimates [2, 3]. Another method called "node dating" [4], whereby prior distributions are proposed for the ages of internal nodes, is ad hoc and can mislead [5, 6].

The fossilized birth-death (FBD) model solves both complications by introducing probabilistic "tip dating" to paleontology and systematics [4, 6–8]. Using the FBD model, it is possible to combine morphological characters and fossil ages with molecular data in a statistically robust framework, closing the gap between "rocks and clocks" [9].

Studies employing the FBD model have nonetheless invariably assumed morphological and molecular characters evolve along the same phylogeny. This assumption is the core of "concatenation", a practice that simply "pastes" characters together into larger data matrices and that has been shown to produce biased tree estimates in a maximum-likelihood context [10, 11]. In a Bayesian context, concatenation is associated with the overestimation of tip branch lengths by as much as 350%, as well as overly narrow credible intervals, which more often than not exclude true parameter values and tree topologies [12–14].

The multispecies coalescent (MSC) more accurately models the evolution of multiple unlinked loci. Under this model, incomplete lineage sorting (ILS) can occur, where gene lineages fail to coalesce in their immediately common ancestral populations. In such event, depending on how lineages then sort, gene tree

topologies might differ from the species tree topology [15, 16]. The MSC is demonstrably more accurate than concatenation when estimating topologies and relative branch lengths in simulated uncalibrated scenarios [13, 14], but has not yet been put to test with empirical data sets comprised of both multiple unlinked loci as well as morphological data.

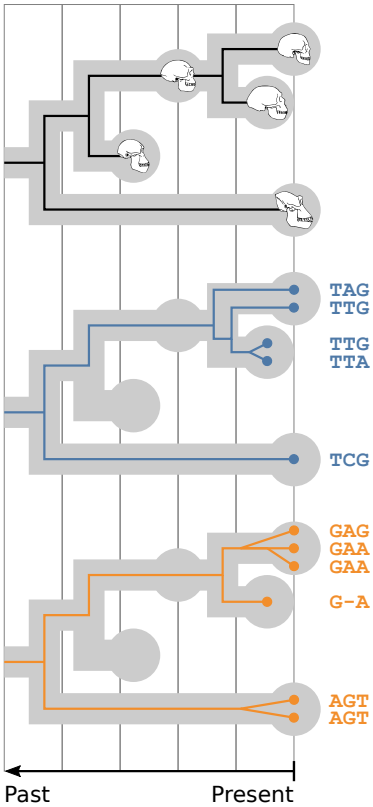

Figure 1: A species tree with a single sampled ancestor—a direct ancestor of other species in the tree—and its relationship to morphological data (top) and multilocus sequence alignments (middle and bottom) in a unified model. The fossilized birth-death (FBD) process is used to model fossilization, speciation and extinction processes, while the multispecies coalescent (MSC) is used to model gene tree evolution within the species tree. GTR family substitution models may be used to model sequence and trait evolution along gene and species trees respectively.

We propose integrative models for species- and population-level evolution, as well as for speciation, extinction and fossilization processes, in order to leverage data of different kinds – molecules, morphology, and the fossil record – in a single probabilistic "total-evidence" [4] analysis. Our model circumvents the known issues caused by concatenation, while explicitly distinguishing the evolutionary processes behind species branching patterns and fossilization, and those behind genealogical branching patterns (Fig. 1). We call our new combined model FBD-MSC, implement it in StarBEAST2 [14] for the BEAST 2 platform [17], demonstrate its correctness, and then compile an exemplar data set of the Caninae (a major canid subfamily) with which we show our model in use.

Materials and Methods

Integrative Model Probability

The integrative model combining the MSC, the FBD process, and morphological evolution can be expressed by combining the component likelihoods. The likelihood of a gene tree is the phylogenetic likelihood [18] $\Pr(D_i|G_i)$ where D_i is the multiple sequence alignment (MSA) for the i-th gene tree G_i. The MSC probability for that gene tree is $P(G_i|S)$ where S is the species tree. The likelihood contribution to the species tree of a morphological character is the phylogenetic likelihood $\Pr(C_j|S)$ where C_j is the vector of states for the j-th character. The prior probability of the species tree under the FBD process is $P(S|\theta)$, where θ is a vector of FBD parameters. By combining the likelihoods for the integrative model we get the probability of the species tree given the molecular, morphological and fossil data:

$$P(S, G, \theta|D, C) = \frac{1}{Z} \prod_i (\Pr(D_i|G_i) \cdot P(G_i|S)) \cdot \prod_j \Pr(C_j|S) \cdot P(S|\theta), \tag{1}$$

where $Z = \Pr(D, C)$ is the marginal likelihood, an unknown normalizing constant that does not need to be computed when using Markov chain Monte Carlo (MCMC) to sample from the posterior distribution.

Under this model, the MSAs inform the species tree through the gene trees, whereas the morphological characters inform the species tree directly. Canid species trees previously inferred from morphological data alone using an FBD model agreed with previous age estimates and expert opinion [19], suggesting that morphological evolution directly along the species tree is an appropriate model. Ultimately both the MSAs and morphological characters inform the FBD parameters through the species tree (e.g., Supplementary Figs. S4, S5).

Under our integrative model, the likelihood of gene trees and of the discrete morphological tree are computed with classicly used models, namely, the generalized time reversible (GTR) substitution model [20], and the Mk model [21], respectively. The MSC probability density is calculated based on species tree branch lengths, and on a function returning the effective population size for each branch. While in our analyses this function always returned a constant size within a branch, linearly changing population sizes are also supported by StarBEAST2 (and other functions like exponential or stepwise are also possible in principle). Lastly, our FBD prior assumes that the rates of speciation, extinction, and sampling of fossils is constant throughout the species tree.

Compiling molecular data

To maximize the information available for phylogenetic reconstruction, we combined DNA sequences from five previous studies. Four of the studies contained segments of coding and/or non-coding DNA [22–25]. All

sequences from the above studies were retrieved from GenBank (Supplementary Table S5). We excluded all sequences from loci other than nuclear autosomes. We also used only one segment for a given gene where multiple segments were available, avoiding segments from a study for which fewer taxa were available [25].

The fourth study included intron-less taste 2 receptor (*Tas2r* coding sequences [26]. After investigating this data, we identified and removed five sequences with likely erroneous labels, and three sequences that were probably either paralogs, degraded, or contained excessive errors. We also identified two pairs of sequences where the labels had likely been swapped, which we corrected (Supplementary Table S5).

That investigation was partly based on a gene tree (Supplementary Fig. S9), inferred from the unaligned *Tas2r* sequences using `PASTA` [27]. Based on that gene tree we excluded the *Tas2R43* and *Tas2R44* genes, as four of the *Tas2R44* sequences appeared to actually be *Tas2R43* sequences. All *Tas2r* sequences were retrieved from GenBank, other than *Lycaon pictus* sequences, which were extracted from the supplementary material of that study (Supplementary Table S5).

For each locus, we aligned the corresponding sequences using `PRANK` [28]. The resulting MSA was trimmed using the "gappyout" method of `trimAl` [29].

The different data sets were somewhat heterogenous. For one study multiple representative specimens were sometimes available for one species [24], but not for other data sets. For another, multiple haplotypes were sometimes available for one specimen [22], but other data sets apparently used ambiguity codes to represent heterozygosity. To make the data sources more uniform, we chose one sequence per locus at random, and randomly resolved all ambiguity codes. At this stage we also excluded outgroup and domestic dog sequences.

Our final data set included 938 sequences from 58 loci and 31 extant Caninae species. This means the amount of missing data, in terms of the number of sequences for a given taxon and locus that were not available, was 48% (Supplementary Fig. S10).

Compiling morphological and fossil data

Our morphological data set is derived from an existing character matrix [30]. Some of these characters were newly scored by the authors of that study but many had been published previously [31–43]. This matrix included soft tissue, pigmental, ecological, developmental, behavioural, cytogenic and metabolic characters not available for fossil taxa. Since our interest in this data set is to use it for total-evidence inference, we retained 230 characters from the original matrix (indices 12 through 241 in the original study) corresponding to skull, dentition, body proportions and postcranial skeleton characters. These were generally available for both fossil and extant taxa.

Character states were available for all 31 extant taxa in our molecular data set, in addition to five extant or recently extinct species absent from our molecular data set (*Canis rufus*, *Dusicyon australis*, *Vulpes*

bengalensis, *V. pallida* and *V. velox*) and 42 fossil taxa, for a total of 78 species. A total of 11,357 states were known, corresponding to a missing data rate of 37% in terms of the proportion of states which are unknown across all characters and species.

Stratigraphic ranges were extracted from the same publication as the character matrix [30], and the midpoint of each range used as the tip date of the corresponding taxa, rounded to the nearest 0.05 million years. If the range of a species reached the present, time zero was used as the tip date.

Caninae analyses

For both FBD-MSC and FBD-concatenation, we ran 4 independent MCMC chains for 2^{30} (roughly one billion) states each, sampling every 2^{18} states for a total of 4,096 samples per chain. For each method, the chains were concatenated after discarding the first 128 samples (roughly 3%) from each chain as burnin. These chains were then thinned to one in every eight samples, for a total of 1,984 samples per method.

FBD-MSC and FBD-concatenation analyses were also identical in terms of the model, priors and operators used, except for FBD-concatenation the MSAs were assumed to evolve directly along the species tree. Hence StarBEAST 2 operators were disabled for FBD-concatenation since gene trees are no longer part of the model. A full description of the model and prior choice used for the FBD-MSC analysis of Caninae is given in Supplementary Methods.

Summary statistics were calculated for each estimated distribution of trees using `DendroPy`. These included the maximum clade credibility (MCC) tree, branch lengths, internal node ages and branch support. For the purpose of calculating support values and ages, a node is defined as the root of a subtree containing all of, and only, a given set of extant taxa. A branch is defined as the direct connection between parent and child nodes as defined above. Lineages-through-time (LTT) curves for FBD analyses were calculated using a custom script. Summary statistics and LTT plots were visualized using `ggplot2` [44] and `ggtree` [45].

Results

Correctness and well-calibrated validation of the full model

We began testing the correctness of our implementation including the MCMC operators by repeating a validation procedure carried out in a previous study using the FBD model [7]. As in that study, we demonstrated the inference of correct tree topology probabilities by comparing our method to analytically derived probabilities (Table 1). We extended this validation by demonstrating the correct inference of internal node ages when compared to automatic integration using quadrature (Supplementary Fig. S8). Together these results show the correctness of our MCMC operators and tree models.

Table 1: Number of replicate MCMC chains in which a particular topology was found within the 95% interval.

Configuration	T1	T2	T3	T4	T5	T6	T7	T8
SA	95	95	97	91	95	95	95	98
UpDown	94	97	95	97	96	95	94	93
MSC	92	96	95	97	93	95	95	93
Coordinated	96	95	93	95	90	95	97	97
NodeReheight2	95	98	94	96	97	96	95	95
Full	91	92	94	96	97	98	99	96

Full details of configurations are provided in Supplementary Methods.

To further confirm the correctness of our implementation, we carried out an extensive well-calibrated validation study [e.g., 8, 46]. This tests the implementation and illustrates statistical power and parameter identifiability. The simulations used for this study were conditioned on 40 extant taxa for each species tree, but covered a broad range of tree sizes when including fossil taxa (Supplementary Fig. S6).

As can be seen in Fig. 2, all parameters had appropriate coverage, i.e., approximately 95% of 95%-HPD intervals contained the true simulated value. Furthermore, a high correlation between posterior estimates and true values was obtained for all parameters, with the exception of birth rate for which only a moderate correlation was observed. Apart from birth rates, FBD-MSC had the power to estimate the other parameters in Fig. 2 with useful precision. In the parameter space covered by our simulations, larger trees are likely necessary if one hopes to accurately estimate birth rates.

Overall, results from our analytical validation and from Fig. 2 allow us to conclude that our models (and MCMC proposal mechanisms and Hastings ratios) are correct. Our conclusions are further strengthened by the fact that all the well-calibrated simulation code is independent of BEAST 2 code used in inference, and in large part written in different programming languages (Supplementary Table S4). Full details of the MCMC operators, methodology for the comparison of topology probabilities and internal node heights and methodology for the well-calibrated validation study are given in Supplementary Methods.

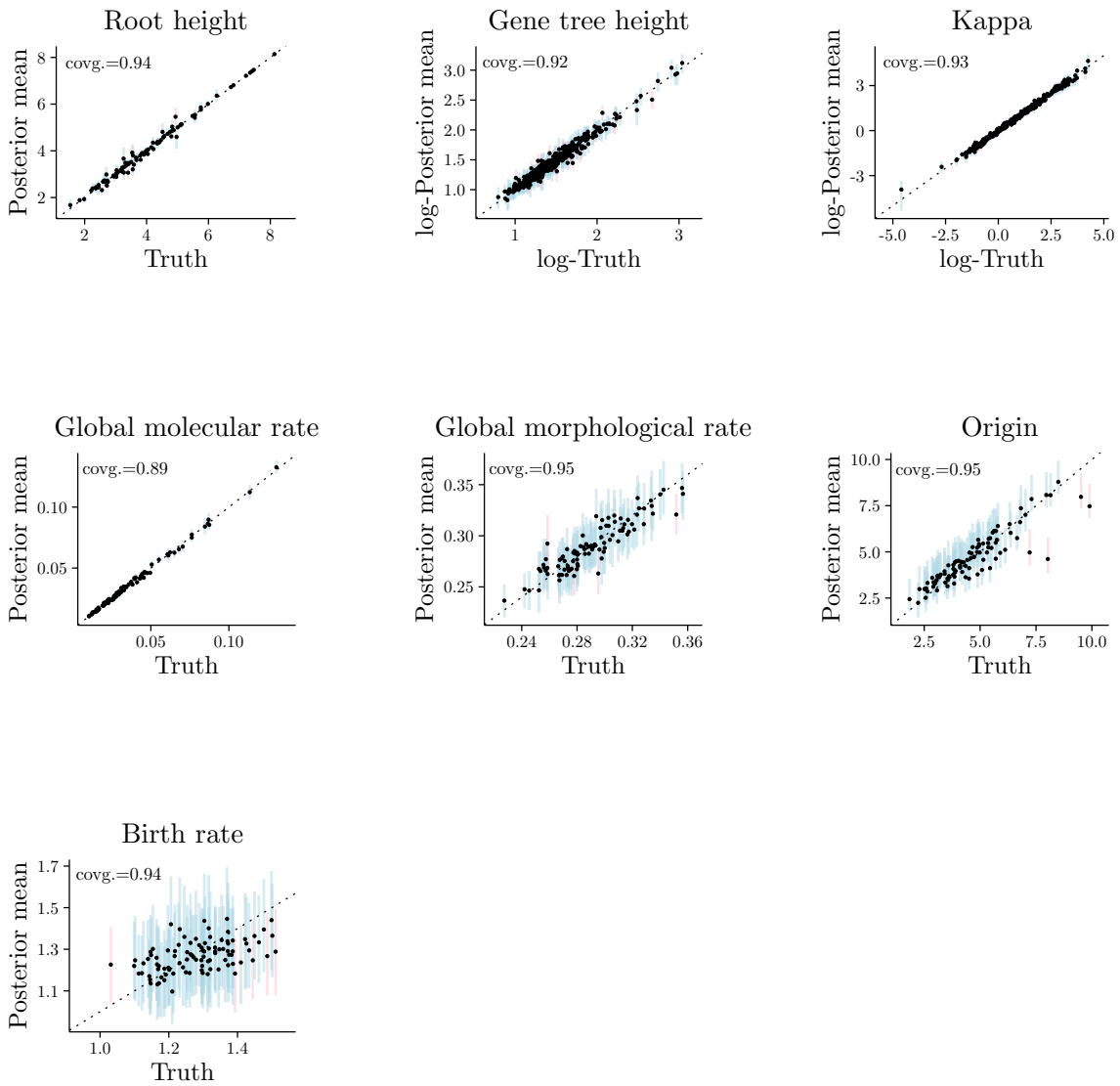

Figure 2: Parameter posterior means against their true simulated values, for 97 simulations (3 were excluded due to convergence issues). Blue lines correspond to the 95%-HPD intervals for a parameter, one line per simulation, when the true value was contained within the interval. When the true value was outside the interval, red lines were used instead. Panels for parameters "Kappa" and "Gene tree height" contain 388 data points (4 loci times 97 simulations). The coverage of each parameter is shown at the top-left corner of the corresponding panel.

Inferring time-trees of Caninae species

We summarized the posterior distributions of Caninae species trees inferred under FBD-concatenation and FBD-MSC models as maximum clade credibility (MCC) trees (Supplementary Figs. S11, S12). Since taxon placement is less certain when only morphological data is available, we pruned taxa with only morphological data before comparing differences in branch lengths between the two distributions. To minimize the effect of topological differences when quantifying differences in length, we define a branch as an edge that directly connects two nodes, each defining a clade (the child clade is necessarily a subset of the parent clade).

Every tip branch present above a frequency threshold of 0.5% in both posterior distributions was inferred to be longer using FBD-concatenation compared with FBD-MSC, and internal branches present above the threshold in both distributions were generally inferred to be shorter (Fig. 3). In some cases the tip branch lengths were inferred to be more than twice as long, e.g., those of *Urocyon* species.

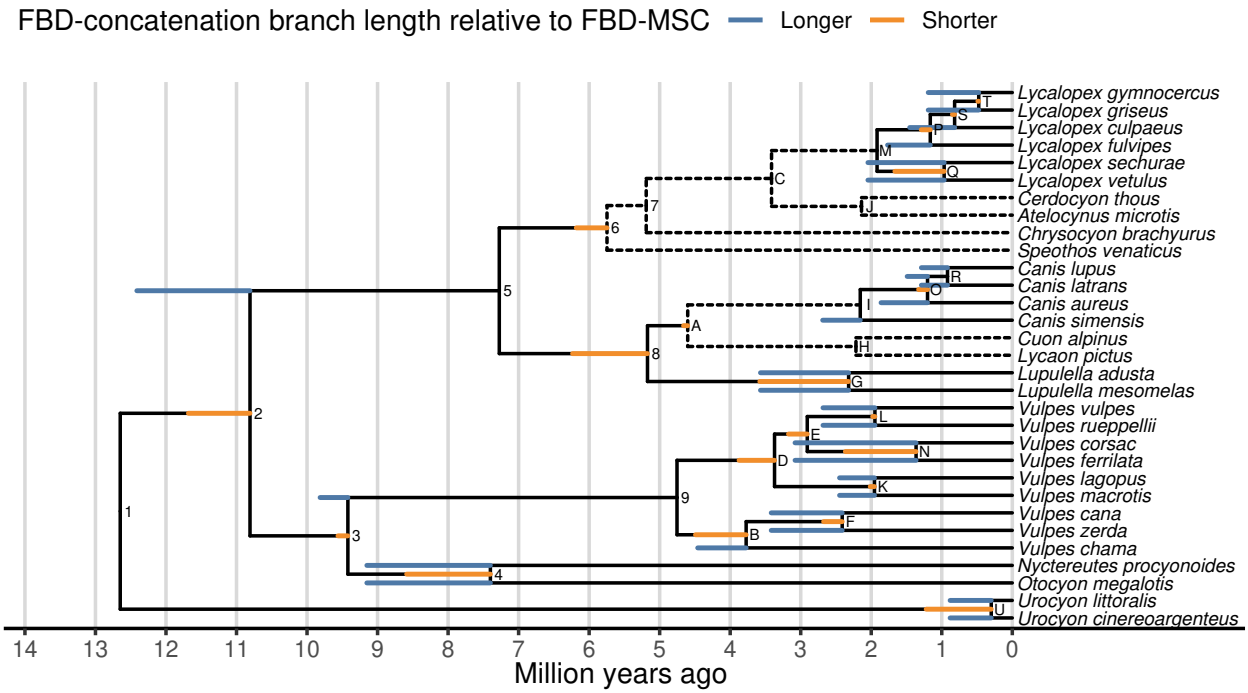

Figure 3: Branch length changes resulting from concatenation. The tree shown is the maximum clade credibility (MCC) tree with mean internal node heights from the FBD-MSC posterior distribution. When the length estimated by FBD-concatenation was longer than for FBD-MSC, the additional length is shown as an extension in blue. When the length was shorter, the reduction is shown as a truncation in orange. The difference in branch lengths is the mean among FBD-concatenation samples including that branch, less the FBD-MSC mean. Dashed lines represent branches with less than 0.5% support using FBD-concatenation.

Several branches (strictly defined as above) from the pruned FBD-MSC MCC topology were not supported by FBD-concatenation, including the branches connecting *Cuon alpinus*, *Lycaon pictus* and extant *Canis* up to their common ancestor. While *C. alpinus* and *L. pictus* belong to a clade sister to all extant and some

stem *Canis* species in the full FBD-MSC MCC tree (Supplementary Fig. S11), *C. alpinus* is more closely related to extant *Canis* than it is to *L. pictus* in the full FBD-concatenation MCC tree (Supplementary Fig. S12). Both methods support the grouping of *Canis*, *Cuon* and *Lycaon*. The grouping of *Speothos* and *Chrysocyon* by FBD-concatenation conflicts with the FBD-MSC topology which makes *Speothos* sister to the other extant South American canids (Supplementary Fig. S11, S12).

Concatenation leads to significant overestimation of molecular evolution rates

Table 2: Macroevolutionary parameter estimates.

Parameter	FBD-concatenation	FBD-MSC
Molecular clock rate ($\times 10^{-3}$)*	0.91 (0.73–1.08)	0.62 (0.52–0.73)
Morphological clock rate ($\times 10^{-2}$)*	3.10 (2.54–3.72)	3.29 (2.79–3.85)
Mean population size**	NA	1.51 (1.04–1.98)
Diversification rate ($\lambda - \mu$)*	0.09 (0.02–0.18)	0.10 (0.01–0.18)
Turnover ($\mu \div \lambda$)	0.85 (0.72–0.98)	0.86 (0.74–0.99)
Sampling proportion ($\psi \div (\psi + \mu)$)	0.16 (0.08–0.26)	0.15 (0.07–0.25)

All values are posterior mean estimates followed in brackets by 95% highest posterior densities.
* Per million years.
** The mean of the inverse gamma distribution fit to per-branch $N_e g$ values, which are effective population sizes N_e scaled by generation time in millions of years g.

The overall molecular clock rate for the (nuclear) protein-coding genes in our data set was inferred to be 6.2×10^{-4} substitutions per site per million years using FBD-MSC, but a broader peak centered at the higher rate of 9.1×10^{-4} was observed using FBD-concatenation (Table 2, Fig. 4a). Our FBD-concatenation estimated average clock rate was consistent with previous rate estimates obtained without accounting for ILS, e.g., 10×10^{-4} for the RAG1 gene in mammals [47].

Unlike the molecular clock rates, however, the posterior distributions of the morphological clock rates are similar between FBD-concatenation and FBD-MSC (Table 2, Fig. 4b). Estimates of other macroevolutionary parameters were also similar for FBD-MSC or FBD-concatenation (Table 2). We find the rate of extinction within Caninae to be high, with a lower bound on turnover of 72% using FBD-concatenation or 74% using FBD-MSC. This means the rate of extinction is at least 72% or 74% the rate of speciation for this subfamily.

Skewed Caninae divergence time estimates under concatenation are fixed by the MSC

For clades with crown ages younger than 4 million years ago in the FBD-MSC MCC tree, and with at least 0.5% posterior support using FBD-concatenation, the crown age was estimated to be older using FBD-concatenation, often with little overlap in the posterior distributions. The most extreme example is node

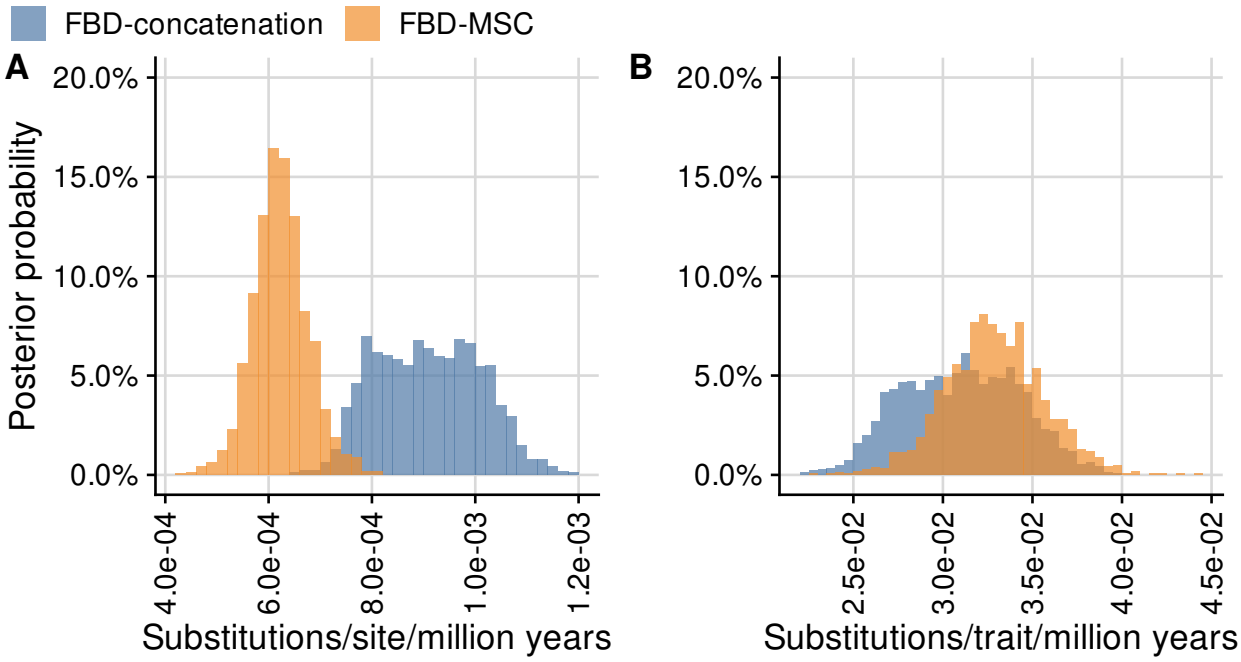

Figure 4: Posterior distribution of clock rates. Posterior probabilities of molecular clock rates (A) and morphological clock rates (B) were calculated using bin widths of 2×10^{-5} and 5×10^{-4} respectively.

N, the common ancestor of *Vulpes corsac* and *V. ferrilata*, inferred to be under 1.5 million years old using FBD-MSC, but around 3 million years old using FBD-concatenation (Fig. 5a).

Crown absolute ages estimated with FBD-concatenation for older clades were more in line with FBD-MSC estimates than those of younger clades; while posterior means could diverge, credible intervals were usually substantially overlapping (Fig. 5a). The correlation between log-ages estimated using the two methods shows younger node heights are consistently overestimated by concatenation, and the exclusion of any speciation events from the concatenation posterior distribution within the past 500 thousand years (Fig. 6). However, as a consequence of the overestimated molecular clock rate, deep nodes are also estimated to be far older using FBD-concatenation compared with FBD-MSC when measured in substitutions per site (Supplementary Fig. S7).

The skewed ages of younger nodes inferred using FBD-concatenation will affect macroevolutionary analyses, including analyses of lineages-through-time (LTT). To demonstrate this, we computed LTT curves based on the species tree posterior distributions inferred using FBD-concatenation and FBD-MSC (Fig. 5b). For both methods the curves are concave upwards, as expected for a birth-death model of evolution with good taxon sampling [48]. However the curves diverge towards the present, so that the burst of speciation leading to the abundance of extant Caninae species is estimated to occur earlier using FBD-concatenation compared with FBD-MSC.

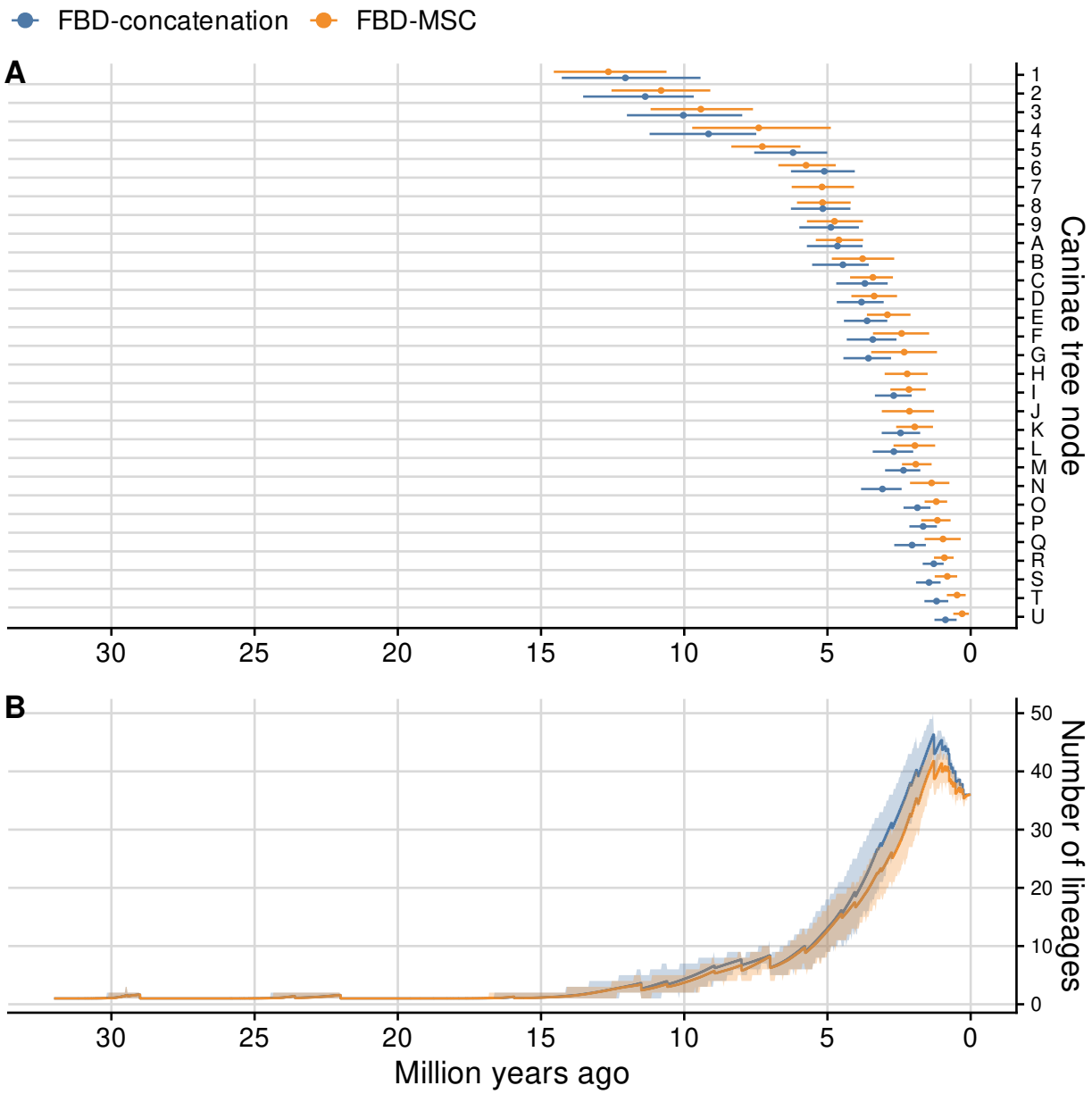

Figure 5: Tempo of caninae evolution. Crown ages estimated by fossilized birth-death with multispecies coalescent (FBD-MSC) and with concatenation (FBD-concatenation) models (A), compared with lineages-through-time (LTT) curves including extinct lineages (B). Posterior mean internal node ages (solid circles) with 95% highest posterior density (HPD) intervals are estimated from samples where that clade is present after pruning all morphology-only taxa. Internal node labels correspond to those in Figure 3. Posterior mean estimates (solid lines) of LTT are calculated for 1,024 evenly spaced time steps spanning 0 to 32Ma. 95% highest posterior density (HPD) intervals calculated for each step are shown as ribbons.

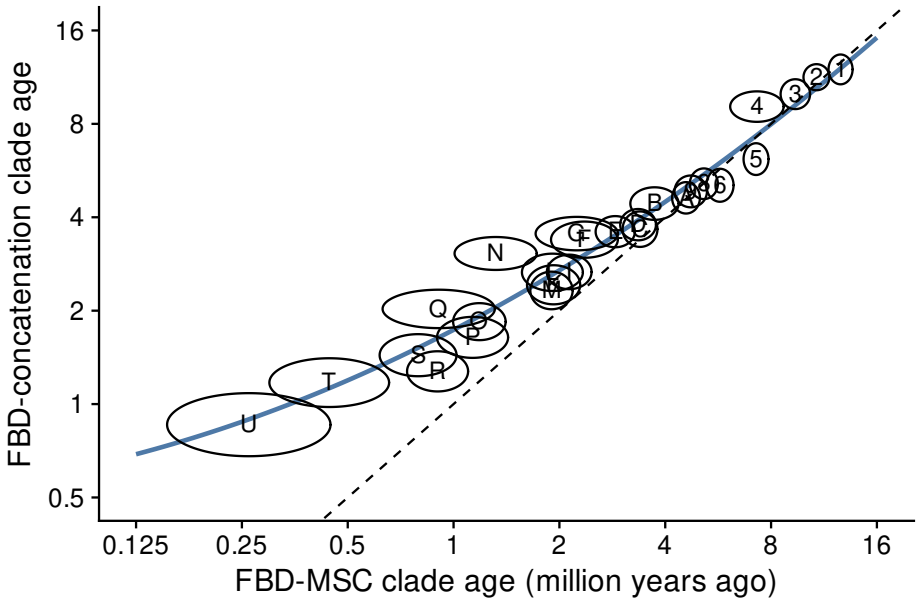

Figure 6: Correlation between log-node heights from the posterior distributions of species trees pruned of morphology-only taxa. Internal nodes from the pruned FBD-MSC MCC tree are drawn as ellipses with their labels from Fig. 3. Ellipses are centered on the mean estimate of the log-node height for both methods. The width and height of each ellipse corresponds to the standard deviation of the log-node heights for FBD-MSC and FBD-concatenation respectively. The dashed black line shows the 1:1 line along which estimates are equal, and the solid blue line is the quadratic line of best fit.

Computational performance

All Caninae analyses were run on an 4GHz Intel i7-8086K CPU. FBD-MSC chains implemented in StarBEAST 2 took a combined 413.8 CPU hours to run, and the smallest effective sample size (ESS) was 335, for the coalescent probability density. FBD-concatenation chains implemented in BEAST 2 took a combined 641.6 CPU hours to run, and the smallest ESS was 555, for the molecular clock rate. The computational performance in ESS per hour was therefore 0.81 for our implementation of FBD-MSC, and 0.87 ESS per hour for FBD-concatenation, a negligible difference. For each method four chains were run in parallel, so the actual time spent waiting on chains to finish was roughly one quarter the combined CPU hours.

A previous study on MSC inference using MCMC found that doubling the number of loci increasing the time to 200 ESS roughly 7-fold [13]. Thus even by running multiple chains in parallel as done in this study, we anticipate our MCMC-based implementation will be applied to data sets of fewer than 1,000 loci.

Discussion

We introduced a new integrative model, the FBD-MSC, for total-evidence analysis of data from extinct and extant species. Because we model the population-level phenomenon of ILS under the FBD-MSC, we make it possible for comparative biologists to carry out statistical inference across evolutionary time scales for the first time. We validated the FBD-MSC model extensively, in what is perhaps the largest, most thorough validation study to date; we expect our effort to satisfy the most demanding users and method developers. We then tested our new method – as well as compared it to the popular alternative of concatenation – with a data set of molecular and morphological characters of canine fossil and living taxa.

Caninae taxonomy

Support for the topology we obtained with the FBD-MSC is echoed by morphological phylogenetic studies of Caninae. [36, 39], probably as a result of their specialized dentitions. A previous study of Caninae that combined morphological characters and mitochondrial sequence alignments found that support for this clade came only from the morphological data, and proposed that the responsible characters are likely convergent due to the hypercarnivory of these two species [49]. Concatenation methods give too much weight to individual sites because (i) neighboring sites are not necessarily independent due to linkage, and (ii) recombination among sites allows the genealogical history of a site to be correlated with, but not identical with the species tree. Due to this overweighting, the morphological evidence for *L. pictus* and *C. alpinus* being sisters relative to extant *Canis* may be overridden by the molecular evidence when FBD-concatenation is used. Similarly, the FBD-MSC topology which makes *Speothos* sister to the other extant South American canids may be driven by similar reasons, suggesting that improved morphological models will be a fruitful avenue of ongoing research.

Caninae evolutionary rates and divergence times

When analyzing our data sets with concatenation, we observed this method overestimated the global (mean) molecular evolutionary rate by a clear margin. We expect this outcome for at least two reasons. First, concatenation treats coalescent times as speciation times, when the former must always be at least equal, but usually greater than the latter. As a result, estimated tree lengths (and internal node heights) in substitutions per site under FBD-concatenation will be larger than under FBD-MSC. The reconciliation of the calendar ages of deeper nodes with the fossil record then manifests as a higher overall molecular rate, a phenomenon that in the context of morphological models was dubbed the "deep coalescence effect" [50]. Second, molecular rates are spuriously inflated as a result of sites within the concatenated alignment having different genealogies because of ILS, which leads to hemiplasy [10].

Concatenation also overestimated tip branch lengths (i.e., older divergence times of contemporary species). Again, we believe this is due to internal nodes under this procedure represent coalescent times rather than speciation times. Under the MSC, coalescent times are always deeper than corresponding speciation times. Furthermore, tip branch lengths can be overestimated because of hemiplasy when concatenation is carried out in the presence of ILS [10]. Irrespective of the cause, these results suggest that using the FBD-MSC model can significantly improve branch estimation from real data.

We note that a previous empirical study demonstrated that concatenation overestimates tip branch lengths, but their analyses were limited to molecular data from extant species [51]. Thus, to our knowledge, this is the first time the effect of concatenation on divergence times has been quantified empirically with multiple extinct and extant species, relative to correctly accounting for coalescent processes.

Another pattern we observed was that estimates of younger divergence times were biased substantially upwards by concatenation, but deeper nodes were less affected. The causes for what we think is a novel result are likely manifold. Mendes & Hahn [10] showed that hemiplasy spuriously increases terminal branch rates while decreasing internal branch rates, possibly to a lesser degree when ILS happens among more species (see Supplementary Fig. 5 from that study). In the absence of calibrations, this effect translates to longer terminal branches and shorter internal branches. Because a larger number of (shortened) internal branches is expected between deeper nodes and the present, the net effect of ILS on deeper node heights should be less pronounced. Future studies may examine the contribution of ILS to the convergence of FBD-concatenation and FBD-MSC estimates at deeper time scales, and its interplay with tip dating and different molecular clock models.

Conclusion

We have demonstrated that failing to model population-level processes when inferring species trees using an FBD model will substantially shift estimates of branch lengths, species divergence times and clock rates, and exclude the possibility of recent speciation. This is the first time that such biases are quantified in a real data set, as well as addressed using superior modeling. Our newly implemented FBD-MSC model accounts for the coalescent process while still being powerful enough to precisely recover rates and times. We also found topological differences between FBD-MSC and FBD-concatenation, but these may be due to traits being treated independently, when they can evolve in concert, e.g., towards hypercarnivory. New models could either rule in or out the *Lycaon+Cuon* grouping by ascribing their similar morphology to homology or convergent evolution. Alternatively, support for this putative clade could be further scrutinized through expanded sampling of fossil taxa, traits or genes.

A number of other avenues for further development of FBD-MSC are open. These include accounting for gene flow after speciation [52] or between lineages [53, 54]. Recent advances in speciation models beyond

cladogenesis can also be applied within an FBD-MSC framework, for example treating species as a kind of trait evolving along a population tree [55], or incorporating budding speciation [56]. The modular architecture of BEAST 2 makes FBD-MSC analyses with the many substitution models in BEAST 2 immediately available, and will make future extensions as in the examples above relatively straightforward to implement. As it is, our method can be used today to infer time-scaled species trees and rates through total-evidence tip dating, without the problems caused by concatenation.

Acknowledgements

We thank Craig Moritz for his advice on preparing the manuscript, and the late Colin Groves for his insight into the Caninae fossil record. Graham Slater provided expertise and insight regarding Caninae taxonomy. Constructive comments from Jeet Sukumaran, David Swofford and Gavin Naylor helped us improve the rigor and clarity of our manuscript. This work was supported by the Royal Society of New Zealand (16-UOA-277 to FKM, AJD, DW, NJM, TGV and TS); the Australian Research Council (FL110100104 to Craig Moritz supported HAO); and the European Research Council (PhyPD:335529 to TS).

References

1. Zuckerkandl, E. & Pauling, L. Molecules as documents of evolutionary history. *Journal of Theoretical Biology* **8,** 357–366 (1965).
2. Bromham, L. The genome as a life-history character: why rate of molecular evolution varies between mammal species. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **366,** 2503–2513 (2011).
3. Besenbacher, S., Hvilsom, C., Marques–Bonet, T., Mailund, T. & Schierup, M. H. Direct estimation of mutations in great apes reconciles phylogenetic dating. *Nature Ecology and Evolution* **3,** 286–292 (2019).
4. Ronquist, F. *et al.* A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. *Systematic Biology* **61,** 973–999 (2012).
5. Parham, J. F. *et al.* Best practices for justifying fossil calibrations. *Systematic Biology* **61** (12 2012).
6. Gavryushkina, A. *et al.* Bayesian Total-Evidence Dating Reveals the Recent Crown Radiation of Penguins. *Systematic Biology* **66,** 57–73 (2017).
7. Gavryushkina, A., Welch, D., Stadler, T. & Drummond, A. J. Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. *PLOS Computational Biology* **10,** e1003919 (2014).

8. Zhang, C., Stadler, T., Klopfstein, S., Heath, T. A. & Ronquist, F. Total-Evidence Dating under the Fossilized Birth-Death Process. *Systematic Biology* **65,** 228–249 (2016).

9. Ronquist, F., Lartillot, N. & Phillips, M. J. Closing the gap between rocks and clocks using total-evidence dating. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **371,** 20150136 (1699 2016).

10. Mendes, F. K. & Hahn, M. W. Gene tree discordance causes apparent substitution rate variation. *Systematic Biology* **65,** 711–721 (2016).

11. Mendes, F. K. & Hahn, M. W. Why concatenation fails near the anomaly zone. *Systematic Biology* **67,** 158–169 (2018).

12. Suzuki, Y., Glazko, G. V. & Nei, M. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. *Proceedings of the National Academy of Sciences* **99,** 16138–16143 (2002).

13. Ogilvie, H. A., Heled, J., Xie, D. & Drummond, A. J. Computational Performance and Statistical Accuracy of *BEAST and Comparisons with Other Methods. *Systematic Biology* **65,** 381–396 (2016).

14. Ogilvie, H. A., Bouckaert, R. R. & Drummond, A. J. StarBEAST2 Brings Faster Species Tree Inference and Accurate Estimates of Substitution Rates. *Molecular Biology and Evolution* **34,** 2101–2114 (2017).

15. Maddison, W. P. Gene Trees in Species Trees. *Systematic Biology* **46,** 523–536 (1997).

16. Degnan, J. H. & Rosenberg, N. A. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution* **24,** 332–340 (2009).

17. Bouckaert, R. *et al.* BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Computational Biology* **15,** 1–28 (2019).

18. Felsenstein, J. Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution* **17,** 368–376 (1981).

19. Matzke, N. J. & Wright, A. Inferring node dates from tip dates in fossil Canidae: the importance of tree priors. *Biology Letters* **12,** 20160328 (2016).

20. Tavaré, S. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences* **17,** 57–86 (1986).

21. Lewis, P. O. A Likelihood Approach to Estimating Phylogeny from Discrete Morphological Character Data. *Systematic Biology* **50,** 913–925 (2001).

22. Bardeleben, C., Moore, R. L. & Wayne, R. K. Isolation and Molecular Evolution of the Selenocysteine tRNA (Cf *TRSP*) and RNase P RNA (Cf *RPPH1*) Genes in the Dog Family, Canidae. *Molecular Biology and Evolution* **22,** 347–359 (2005).

23. Bardeleben, C., Moore, R. L. & Wayne, R. K. A molecular phylogeny of the Canidae based on six nuclear loci. *Molecular Phylogenetics and Evolution* **37,** 815–831 (2005).

24. Lindblad-Toh, K. *et al.* Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438,** 803–819 (2005).

25. Koepfli, K.-P. *et al.* Genome-wide evidence reveals that African and Eurasian golden jackals are distinct species. *Current Biology* **25,** 2158–2165 (2015).

26. Shang, S. *et al.* The repertoire of bitter taste receptor genes in canids. *Amino Acids* **49,** 1159–1167 (2017).

27. Mirarab, S., Nguyen, N. & Warnow, T. *PASTA: Ultra-Large Multiple Sequence Alignment* in *Research in Computational Molecular Biology* (ed Sharan, R.) (Springer International Publishing, Cham, 2014), 177–191.

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

29. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25,** 1972–1973 (2009).

30. Zrzavý, J., Duda, P., Robovský, J., Okřinová, I. & Pavelková Řičánková, V. Phylogeny of the Caninae (Carnivora): Combining morphology, behaviour, genes and fossils. *Zoologica Scripta* **47,** 373–389 (2018).

31. Berta, A. Quaternary evolution and biogeography of the large South American Canidae (Mammalia: Carnivora). *University of California Publications in Geological Sciences* **132** (1988).

32. Christiansen, P. & Adolfssen, J. S. Bite forces, canine strength and skull allometry in carnivores (Mammalia, Carnivora). *Journal of Zoology* **266,** 133–151 (2005).

33. Feldhamer, G., Drickamer, L., Vessey, S., Merritt, J. & Krajewski, C. *Mammalogy: Adaptation, Diversity, Ecology* (Johns Hopkins University Press, Baltimore, Maryland, 2007).

34. Finarelli, J. A. & Flynn, J. J. Brain-size evolution and sociality in Carnivora. *Proceedings of the National Academy of Sciences* **106,** 9345–9349 (2009).

35. Friscia, A. R., Van Valkenburgh, B. & Biknevicius, A. R. An ecomorphological analysis of extant small carnivorans. *Journal of Zoology* **272,** 82–100 (2007).

36. Prevosti, F. J. Phylogeny of the large extinct South American Canids (Mammalia, Carnivora, Canidae) using a "total evidence" approach. *Cladistics* **26,** 456–481 (2010).

37. Slater, G. J. Iterative adaptive radiations of fossil canids show no evidence for diversity-dependent trait evolution. *Proceedings of the National Academy of Sciences* **112,** 4897–4902 (2015).

38. Tedford, R. H., Taylor, B. E. & Wang, X. Phylogeny of the Caninae (Carnivora, Canidae): the living taxa. *American Museum Novitates,* 3146 (1995).

39. Tedford, R. H., Wang, X. & Taylor, B. E. Phylogenetic systematics of the North American fossil Caninae (Carnivora: Canidae). *Bulletin of the American Museum of Natural History* **325** (2009).

40. Van Valkenburgh, B. & Koepfli, K. P. Cranial and Dental Adaptation to Predation in Canids. *Symposia of the Zoological Society of London* **65,** 15–37 (1993).

41. Wang, X. Transformation from plantigrady to digitigrady: functional morphology of locomotion in *Hesperocyon* (Canidae: Carnivora). *American Museum Novitates,* 3069 (1993).

42. Wang, X. Phylogenetic systematics of the Hesperocyoninae (Carnivora, Canidae). *Bulletin of the American Museum of Natural History* **221** (1994).

43. Wang, X., Tedford, R. H. & Taylor, B. E. Phylogenetic systematics of the Borophaginae (Carnivora, Canidae). *Bulletin of the American Museum of Natural History* **243** (1999).

44. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* 2nd (Springer-Verlag, New York, 2016).

45. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* **8,** 28–36 (2017).

46. Gaboriau, T., Mendes, F. K., Joly, S., Silvestro, D. & Salamin, N. A multi-platform package for the analysis of intra- and interspecific trait evolution. *Methods Ecol. Evol.* **11,** 1439–1447 (11 2020).

47. Hugall, A. F., Foster, R., Lee, M. S. Y. & Hedin, M. Calibration Choice, Rate Smoothing, and the Pattern of Tetrapod Diversification According to the Long Nuclear Gene RAG-1. *Systematic Biology* **56,** 543–563 (2007).

48. Stadler, T. Lineages-through-time plots of neutral models for speciation. *Mathematical Biosciences* **216,** 163–171 (2008).

49. Zrzavý, J. & Řičánková, V. Phylogeny of Recent Canidae (Mammalia, Carnivora): relative reliability and utility of morphological and molecular datasets. *Zoologica Scripta* **33,** 311–333 (2004).

50. Mendes, F. K., Fuentes-González, J. A., Schraiber, J. & Hahn, M. W. A multispecies coalescent model for quantitative traits. *eLife* **7,** e36482 (2018).

51. Bragg, J. G. *et al.* Phylogenomics of a rapid radiation: the Australian rainbow skinks. *BMC Evolutionary Biology* **18,** 15 (2018).

52. Heled, J., Bryant, D. & Drummond, A. J. Simulating gene trees under the multispecies coalescent and time-dependent migration. *BMC Evolutionary Biology* **13,** 44 (2013).

53. Wen, D. & Nakhleh, L. Coestimating Reticulate Phylogenies and Gene Trees from Multilocus Sequence Data. *Systematic Biology.* Advance article (2017).

54. Zhang, C., Ogilvie, H. A., Drummond, A. J. & Stadler, T. Bayesian Inference of Species Networks from Multilocus Sequence Data. *Molecular Biology and Evolution.* Advance article (2017).

55. Sukumaran, J. & Knowles, L. L. Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences* **114,** 1607–1612 (2017).

56. Stadler, T., Gavryushkina, A., Warnock, R. C., Drummond, A. J. & Heath, T. A. The fossilized birth-death model for the analysis of stratigraphic range data under different speciation modes. *Journal of Theoretical Biology* **447,** 41–55 (2018).