Systematic Biology (2021), **0**, 0, *pp.* 1–**??**
doi:10.1093/sysbio/FBD-MSC-text

Novel Integrative Modeling of Molecules and Morphology across Evolutionary Timescales

Huw A. Ogilvie[1,*,†], Fábio K. Mendes[2,4,**,†], 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*
***Correspondence to be sent to: Fábio K. Mendes, School of Biological Sciences, The University of Auckland, Private Bag 92019, Auckland Mail Centre, Auckland 1142, New Zealand; Email: f.mendes@auckland.ac.nz*

[†] *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 this 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 that 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.

Key words: Caninae; phylogenetics; fossilized birth-death; species trees; multispecies

coalescent; molecular clock.

Creating a high-resolution picture of the tree of life is an increasingly achievable goal given the ever greater availability of molecular and paleontological data. Realistic and tractable evolutionary models are required to treat this rich data in a statistically sound manner. The end result should be phylogenies that not only explain how species are related, but are also scaled to absolute time, which allows species trees to be reconciled with geological and fossil records.

One method for scaling trees into absolute time is to assume a molecular clock (Zuckerkandl and Pauling, 1965) ticking at a known rate (or rates) per unit time. This strategy is problematic because a universal clock does not exist, and extrapolating clock rates measured in one group of organisms to another can lead to unrealistic evolutionary time estimates (Bromham, 2011; Besenbacher et al., 2019).

Alternatively, the "node dating" method (Ronquist et al., 2012; Ho and Phillips, 2009) proposes prior distributions for divergence times based on fossil ages and morphology. Yet this method faces many problems. Node dating only uses the oldest available fossils, ignoring younger fossils. Fossil affinities and associated node age priors are ultimately specified using expert knowledge (Parham et al., 2012) which, due to its ad hoc nature, can introduce explicit bias and circularity to divergence time estimates (Warnock et al., 2011; Field et al., 2020). Finally, the interaction between priors on "dated" nodes and the overall tree prior used in a hierarchical models creates complex and unintended prior probabilities on node ages throughout the tree (Heled and Drummond, 2012).

The fossilized birth-death (FBD) model introduced probabilistic "tip dating" to paleontology and systematics (Ronquist et al., 2012; Gavryushkina et al., 2014; Zhang et al., 2016; Gavryushkina et al., 2017). This model not only directly solves the shortcomings of node dating, but in providing more accurate model-based uncertainty on divergence times, it also allows relaxed clock models to be less distorted by inadequacies in

the tree prior and calibration scheme. Unless fossil ages are data, relaxed clock models by themselves do not "close the gap between rocks and clocks" (Ronquist et al., 2016; Berv and Field, 2018), and only tell us about relative differences in accumulated evolutionary change, where time and rates are conflated. By using the FBD model, one can combine morphological characters and fossil ages with molecular data in a statistically robust framework, and disentangle absolute time from evolutionary rates.

Studies employing the FBD model have invariably assumed that morphological and molecular characters evolve along the same phylogeny. This assumption is the core of "concatenation" (initially referred to as "total evidence" data combination; Kluge, 1989; Huelsenbeck et al., 1996), a protocol that attempts to harness as much information from as many different data sources as possible. The hope of concatenation is that agreeing signals speak louder than the sampling noise, and that conflicting signals can compete in the resolution of the phylogenetic estimate. The crucial feature of concatenation, as opposed to integrative probabilistic models (discussed below), is that all characters are simply appended together into a single large data matrix.

Since genomes have become central data sources for studying the evolution of living species, concatenation is now often taken to mean "pasting" all sequenced nucleotides together into a single multiple sequence alignment (MSA). This is the meaning we employ here. Within the domain of molecular phylogenetics, concatenation has been shown to produce biased tree estimates in a maximum-likelihood context (Mendes and Hahn, 2016, 2018). In a Bayesian context, concatenation has been associated with the overestimation of tip branch lengths by as much as 350%, as well as inaccurately narrow credible intervals, which often exclude true parameter values and tree topologies (Suzuki et al., 2002; Ogilvie et al., 2016, 2017). By contrast, the multispecies coalescent (MSC) accurately models the evolution of multiple unlinked loci. Concatenation is still used due to the perceived higher computational cost of MSC, which we will show does not exist (relative to Bayesian concatenation) when inferring species trees using tip dating on a real data set.

Under the MSC, incomplete lineage sorting (ILS) can occur, where gene lineages fail to coalesce in their immediately common ancestral populations. In such events, depending on how lineages then sort, gene tree topologies might differ from the species tree topology (Maddison, 1997; Degnan and Rosenberg, 2009). The MSC is demonstrably more accurate than concatenation when estimating topologies and relative branch lengths in simulated uncalibrated scenarios (Ogilvie et al., 2016, 2017), but has not yet been put to test with empirical data sets comprised of both multiple unlinked loci as well as morphological data.

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" (Ronquist et al., 2012) 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 (Ogilvie et al., 2017) for the BEAST 2 platform (Bouckaert et al., 2019), 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 probability mass and density functions (pmf and pdf) characterizing all the component sampling distributions. The probability of the i-th gene tree given the multiple sequence alignment (MSA) of the i-th gene is sometimes referred to as the 'phylogenetic likelihood' (Felsenstein, 1981), and is characterized by pmf $\Pr(\mathbf{D}_i|G_i)$. Under the MSC, the probability of that gene tree given species tree S and

population sizes $\boldsymbol{N_e}$ is $f(G_i|S, \boldsymbol{N_e})$. Note that $S = \{\phi, \boldsymbol{t}^n, \mathbf{t}^{\mathrm{s}}\}$, where $\mathbf{t}^{\mathrm{s}}$ is observed data and correspond to the sample ages (fossil ages and living taxa). Both ϕ and $\boldsymbol{t}^n$ are parameters (random variables) and denote the species tree topology and internal node times, respectively.

The likelihood contribution to the species tree of a morphological character is captured by the phylogenetic likelihood $\Pr(\mathbf{C}_j|S)$ where $\mathbf{C}_j$ is the vector of states for the j-th character. The prior probability of the species tree under the FBD process is $f(S|\boldsymbol{\theta}^{\mathrm{FBD}})$, where $\boldsymbol{\theta}^{\mathrm{FBD}}$ is a vector of FBD parameters. Finally, $f(\boldsymbol{\theta})$ describes the joint distribution over all parameters $\boldsymbol{\theta} = \{\boldsymbol{\theta}^{\mathrm{FBD}}, \boldsymbol{\theta}^r, \boldsymbol{N_e}\}$ (where $\boldsymbol{\theta}^r$ denotes all remaining parameters not explicitly mentioned above). By combining the probability mass and density functions of all sampling distributions comprising the integrative model, we get the probability density of the species tree given the molecular, morphological and fossil age data:

$$f(S, G, \boldsymbol{\theta}|\mathbf{D}, \mathbf{C}, \mathbf{t}^{\mathrm{s}}) = \frac{1}{Z} \prod_i \left(\Pr(\mathbf{D}_i|G_i) \cdot f(G_i|S, \boldsymbol{N_e})\right) \cdot \prod_j \Pr(\mathbf{C}_j|S) \cdot f(S|\boldsymbol{\theta}^{\mathrm{FBD}}) \cdot f(\boldsymbol{\theta}), \tag{0.1}$$

where $Z = \Pr(\mathbf{D}, \mathbf{C})$ 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.

When conducting inference under this model, MSAs are assumed to evolve along gene trees, which then inform the species tree via the MSC, whereas the morphological characters are assumed to have evolved along the species tree itself, and thus inform it directly. 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 (the latter being the species tree, S) are computed with a model in the generalized time reversible (GTR) family (Tavaré, 1986) and the Mk model (Lewis, 2001),

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, the FBD prior assumes that the rates of speciation, extinction, and sampling of fossils are constant throughout the species tree.

Well-calibrated validation of model and operators An integrative (hierarchical) Bayesian model like the one we introduce here consists of a collection of probability mass and density functions characterizing all likelihoods and priors. Although some components of this collection may have been individually validated, the collection itself needs to be validated as a whole. (Where new MCMC operators are introduced, they also need to be validated.) This type of validation can be seen as both an instantiation and probabilistic analog of what software engineers refer to as 'integration testing', whereby software modules are combined and tested together. This type of testing is a mandatory stage in the software development life cycle because it is often hard to predict how different modules will interact.

In the particular case of our FBD-MSC model, this is illustrated by the addition of fossils (via the FBD process) complicating the relationship between the species and gene trees under the MSC – when compared with a species tree model such as the simple birth-death process without fossils. The two key assumptions of a birth-death process that are relaxed under the FBD-MSC are (1) that the species tree must be ultrametric, i.e., all of the n species are sampled at a single point in time, and (2) that the number of nodes is fixed at $(2n-1)$ regardless of how the topology of the tree changes. The relaxation of both assumptions in the FBD-MSC required changes to the implementation of the MSC and related operators in StarBEAST2 (see the Supplementary Material for an example in Algorithm S3).

More specifically, relaxing the second assumption requires fundamental changes to the inferential algorithm. This is because previous implementations of MSC used Metropolis-Hastings MCMC, which does not allow for changing the number of dimensions in the model. But converting a fossil from terminal node to sampled ancestor will decrease (or in the other direction increase) the number of nodes and hence dimensions. In such cases, not only does an additional node age have to be estimated, but so do the parameters of the population size function of the corresponding branch. One possible strategy to sample the additional node age – the one we chose – is to use reversible-jump operators, such as those previously developed for BEAST 2 (Gavryushkina et al., 2017). To sample the additional population size parameters, we implemented a composite model space approach (Carlin and Chib, 1995; Drummond and Suchard, 2010; Wu and Drummond, 2011) whereby population size parameters for the maximum number of species branches are being sampled, but only those corresponding to branches in the current topology are contributing to the likelihood.

Because our full model and related MCMC machinery are new in the ways described above, we verify correctness through a well-calibrated validation study. Here, many independent data sets are simulated under the full model (i.e., all pmf and pdfs), and correctness is deduced from appropriate posterior coverage upon MCMC chain convergence.

Empirical analysis: the canid subfamily Caninae

The diverse family of dogs (Canidae) has a rich fossil record that has made this clade a model for ecological, evolutionary and methodological studies (Zrzavý et al., 2018). Canidae is comprised of three subfamilies – Borophaginae, Hesperocyoninae and Caninae – represented by carnivorans of jackal, fox and wolf semblance (Wang and Tedford, 2008). Borophaginae (∼66–69 species; Wang et al., 1999; Wang and Tedford, 2008) and Hesperocyoninae (∼26–29 species; Wang, 1994; Wang and Tedford, 2008) consist of only

extinct species, with Caninae accounting for the remaining fossils (out of ∼140–178; Slater, 2015; Wang and Tedford, 2008) in addition to 36 living species (Nowak, 1991).

Previous phylogenetic accounts of canids using morphology alone under the FBD model have shown that this type of data can produce sensible age estimates (Matzke and Wright, 2016), but contrasting topologies, particularly in terms of root placement, when compared to molecular trees (in the case of Caninae; Zrzavý and Řičánková, 2004; Zrzavý et al., 2018). Here, we further examine the phylogeny of Caninae by carrying out an integrative statistical analysis where molecular and morphological data jointly inform 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 (Bardeleben et al., 2005a,b; Lindblad-Toh et al., 2005; Koepfli et al., 2015). All sequences from the above studies were retrieved from GenBank (Supplementary Table S9). 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 (Koepfli et al., 2015).

The fourth study included the coding sequences of multiple intron-less taste 2 receptor (*Tas2r*) genes (Shang et al., 2017). After investigating these 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 S9). That investigation was partly based on a gene tree (Supplementary Fig. S12), inferred from the unaligned *Tas2r* sequences using `PASTA` (Mirarab et al., 2014) and available in Supplementary Material. 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 S9).

The different data sets were somewhat heterogenous. For one study multiple representative specimens were sometimes available for one species (Lindblad-Toh et al., 2005), but not for other data sets. For another, multiple haplotypes were sometimes available for one specimen (Bardeleben et al., 2005a), 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.

For each locus, we aligned the corresponding sequences using `PRANK` (Löytynoja and Goldman, 2005). The resulting MSA was trimmed using the "gappyout" method of `trimAl` (Capella-Gutiérrez et al., 2009). Our final data set included 938 sequences from 58 loci and 31 extant Caninae species (out of the 36 known living species; Nowak, 1991). 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. S13). A numerical summary of our molecular data set can be found in Supplementary Table S7.

Compiling morphological and fossil data Our morphological data set is derived from an existing character matrix (Zrzavý et al., 2018). Some of these characters were newly scored by the authors of that study but many had been published previously (Berta, 1988; Christiansen and Adolfssen, 2005; Feldhamer et al., 2007; Finarelli and Flynn, 2009; Friscia et al., 2007; Prevosti, 2010; Slater, 2015; Tedford et al., 1995, 2009; Van Valkenburgh and Koepfli, 1993; Wang, 1993, 1994; Wang et al., 1999). 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 (out of ∼48–80 fossil species; Slater, 2015; Wang and Tedford, 2008), 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. We recorded all character states in a BEAST 2-compatible NEXUS file. Scripts and instructions to create this file are in the Supplementary Material.

Stratigraphic ranges were extracted from the same publication as the character matrix (Zrzavý et al., 2018), and the midpoint of each range used as the tip date of the corresponding taxon, 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. While tip dates could also have been estimated, doing so would increase the number of parameters of our model, and potentially require longer MCMC chains; more importantly, without extensive expert curation it is not immediately clear which prior distributions to use for each of the different fossils in our data set – a problem similar to the characteristic node dating problem.

Caninae analyses We analyzed Caninae data under two models that treat molecular, morphological and temporal information as data; these are the aforementioned FBD-MSC and what we call the FBD-concatenation model. Both were identical in terms of most of the density functions characterizing their sampling distributions, the only difference being that for FBD-concatenation MSAs were assumed to evolve directly along the species tree (i.e., $\prod_i \left(\Pr(\mathbf{D}_i|G_i) \cdot f(G_i|S, \boldsymbol{N_e})\right)$ is replaced with $\Pr(\mathbf{D}|S)$ in Eq. 0.1, where $\mathbf{D}$ represents all MSAs after concatenation). Hence StarBEAST 2 operators were disabled for FBD-concatenation since gene trees are not part of that model; all other operators were shared. A full description of the model and prior choice used for the FBD-MSC analysis of Caninae is given in Supplementary Methods.

Four independent MCMC chains were run for each method, with 4,096 evenly spaced samples being collected. Individual chain lengths were 2^{30} (roughly one billion) states for FBD-MSC and 2^{29} (roughly half a billion) for FBD-concatenation. For each method, the chains were joined after discarding the first 64 samples (roughly 1.6%) from each chain as burn-in. Burn-in was determined by manual inspection of MCMC traces in Tracer 1.7 (Rambaut et al., 2018). These chains were then thinned to one in every eight samples, for a total of 2,016 samples per method. Scripts, BEAST 2 XML files and combined output files for each analysis are available in the Supplementary Material.

Summary statistics were calculated for each estimated distribution of trees using `DendroPy` (Sukumaran and Holder, 2010). These included the maximum clade credibility (MCC) tree, branch lengths, internal node ages and node 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. Lineages-through-time (LTT) curves for FBD analyses were calculated using a custom script. Summary statistics and LTT plots were visualized using `ggplot2` (Wickham, 2016) and `ggtree` (Yu et al., 2017). All scripts for calculating and visualizing summary statistics are available in the Supplementary Material.

Results

Correctness of the full model

We tested the correctness of our model implementation including the MCMC operators by repeating a validation procedure carried out in a previous study using the FBD model (Gavryushkina et al., 2014). 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. S11). Together these results indicate our MCMC operators have been correctly implemented and interact with the FBD model as expected.

To further confirm the correctness of our implementation, we carried out an extensive well-calibrated validation study (e.g., Zhang et al., 2016; Gaboriau et al., 2020; Zhang et al., 2021). 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. S9).

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, the 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 model (as well as MCMC proposal mechanisms and Hastings ratios) has been correctly implemented. 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 ages and methodology for the well-calibrated validation study are given in Supplementary Methods.

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. S14, S15). In order to compare branch length estimates from both models, we parsed their posterior distributions in two steps: (i) we pruned taxa providing only morphological data because their phylogenetic affinity is difficult to estimate, and (ii) we established a branch frequency threshold (see below) that branches had to meet in order to be compared. Both steps minimize topological differences in an attempt to make length comparisons fair; while arbitrary, these steps yielded posteriors whose trees had a reasonable number of branches inducing largely agreeing clades above and below.

A branch was compared only if both its parent and child nodes (the child clade is necessarily a subset of the parent clade) were present in at least 0.5% of the trees in both FBD-concatenation and FBD-MSC posteriors. Every terminal branch meeting this frequency threshold was estimated 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 terminal branch lengths were inferred to be more than twice as long, e.g., those of *Urocyon* species.

Several branches did not meet the frequency threshold, such as the branch 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. S14), *C. alpinus* is more closely related to extant *Canis* than it is to *L. pictus* in the full FBD-concatenation MCC tree (Supplementary Fig. S15). Under both models, the grouping of *Canis*, *Cuon* and *Lycaon* was well supported. The grouping of *Speothos* and *Chrysocyon* by FBD-concatenation was not well supported under the FBD-MSC, whose MCC tree shows *Speothos* as sister to the other extant South American canids (Supplementary Fig. S14, S15).

Mean effective population sizes (Table 2) were estimated well within the empirical range for canine species, which have highly variable census population sizes (Supplementary Table S8). We note that our model integrates out branch-specific N_e

values (Heled and Drummond, 2010), and that what we estimated is the mean of an inverse gamma prior distribution on $\boldsymbol{N_e}$ (see Supplementary Material for more details).

Concatenation estimates significantly higher molecular evolution rates

The mean posterior estimate of the overall molecular clock rate for (nuclear) protein-coding genes was 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^{-3} for the RAG1 gene in mammals (Hugall et al., 2007). Unlike the molecular clock rates, however, the posterior distributions of the morphological clock rates largely overlapped between FBD-concatenation and FBD-MSC (Table 2, Fig. 4b).

Estimates of 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 and 74% using FBD-MSC. This means the rate of extinction is at least 72% or 74% the rate of speciation for this subfamily.

Caninae divergence time estimates are skewed under concatenation, but not under 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 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). (When trees are measured in substitutions per site, however, deep nodes reappear as being older when estimated under FBD-concatenation, as a result of overestimated molecular clock rates; Supplementary Fig. S10.) Plotting estimated node log-ages under FBD-MSC as a function of those from FBD-concatenation (Fig. 6) revealed that younger nodes were consistently estimated as older by concatenation, so much so that speciation events within the past 500 thousand years were not inferred with this approach.

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 (Stadler, 2008). 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.

Computational performance

We report performance summaries for all Caninae analyses in Table 3; these were all run on an 4GHz Intel i7-8086K CPU. The lowest effective sample sizes were observed for the coalescent probability density and the phylogenetic likelihood of the *T2R42* locus under the FBD-MSC and FBD-concatenation, respectively. From our observations, tip-dating methods will require substantially longer chain lengths than node-dating methods, but using the MSC to model gene tree evolution does not incur a substantial computational performance penalty. 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 increases the time to 200 ESS roughly 7-fold (Ogilvie et al., 2016). 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 also carried out critical validation of the FBD-MSC model and related operators, in a thorough validation study. While in an ideal scenario there should be no surprises when previously validated models are combined, there are often fundamental conceptual consequences for combining sampling distributions that only become apparent once we embark on building the composite model.

All previous descriptions of the MSC (e.g., Pamilo and Nei, 1988; Rannala and Yang, 2003; Heled and Drummond, 2010), for example, have assumed the species tree is ultrametric. By combining the MSC with the FBD, a number of assumptions that were valid for ultrametric species trees no longer held. The most glaring idiosyncrasy of the FBD-MSC is that the number of branches, and therefore the number of population sizes, becomes a random variable of the composite model. As outlined in the above sections, there are multiple technical solutions to this problem. Further extensions of our integrative model could introduce an oriented species tree formalism (Stadler et al., 2018), which could allow the population size to remain the same through successive fossils from the same morphological species.

These considerations highlight the point that the construction of novel joint models often requires, and leads to, new thinking. Their implementation needs more than just good book-keeping. As the size and complexity of phylogenetic models continues to grow,

so too should we expect to discover emergent properties. We believe that careful construction and validation of joint models is a fundamental contributor to scientific novelty in molecular evolution and phylogenetics.

Caninae taxonomy

We tested our new method – and compared it to the popular alternative of concatenation – with a data set of molecular and morphological characters of canine fossil and living taxa. We still do not fully understand how concatenation (of both molecular and morphological characters) can bias species tree inference by not capturing the possible topological independence between characters. Unlike the MSC, concatenation "channels" the information from all characters into supporting a single tree topology, which is then taken as a proxy for the species tree. Under the MSC, conversely, sets of sites are allowed to evolve along their own gene trees, whose topology might on average be less resolved than the single tree proxy estimated through concatenation. These more uncertain gene trees must then inform the species tree through the MSC density function. Morphological characters can thus perhaps be seen as having relatively greater influence in FBD-MSC inference when compared to FBD-concatenation.

Indeed, support for the topology we obtained with the FBD-MSC is echoed by morphological phylogenetic studies of Caninae (Tedford et al., 2009; Prevosti, 2010), probably as a result of their specialized dentitions. A previous study of Canidae (the family to which Caninae belongs) that combined morphological characters and mitochondrial sequence alignments found that support for (*L. pictus*, *C. alpinus*), for example, came only from the morphological data, and proposed that the responsible characters are likely convergent due to the hypercarnivory of these two species (Zrzavý and Řičánková, 2004). It could be due to the lower relative signal of morphological characters (with respect to nucleotide sites) under FBD-concatenation that the aforementioned clade is not recovered. Similarly, the inference of *Speothos* as sister to the other extant South American canids

under the FBD-MSC model might reflect the same phenomenon. These results suggest 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 estimated markedly higher global (mean) molecular evolutionary rates than the FBD-MSC. 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 to, but usually greater than, the latter. As a result, estimated tree lengths (and internal node ages) 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" (Mendes et al., 2018). 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 (Mendes and Hahn, 2016).

Concatenation also estimated larger terminal branch lengths (i.e., older divergence times of contemporary species). Again, we believe this is due to internal nodes under this procedure representing coalescent times rather than speciation times. Under the MSC, coalescent times are always deeper than corresponding speciation times. Furthermore, terminal branch lengths can be inferred to be larger because of hemiplasy when concatenation is carried out in the presence of ILS (Mendes and Hahn, 2016). Irrespective of the cause, these results suggest that using the FBD-MSC model can significantly improve branch length estimation from real data.

Our findings agree with a previous empirical study that also demonstrated concatenation can lead to longer terminal branch lengths relative to the MSC model (Bragg et al., 2018). That study was nonetheless limited to molecular data from extant species and was therefore missing critical dating information that only serially timed data

(like fossils) can provide, and that can only be incorporated through an integrative model like ours. We note that while it is possible that FBD-concatenation is correctly estimating the empirical molecular rate (the truth is unknown), with FBD-MSC underestimating it, we find this unlikely. Previous theoretical and simulation work reveal biases in line with what we observed here, and that these biases are corrected (or should be corrected) by the MSC as a result of modeling population-level processes (Mendes and Hahn, 2016; Ogilvie et al., 2017).

Finally, another novel finding included the fact that estimates of younger node divergence times could be substantially biased upwards by concatenation, but those of deeper nodes were less affected. The causes for this result are likely manifold. Mendes and Hahn (2016) 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 ages 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 (Heled et al., 2013) or between lineages (Wen and Nakhleh, 2018; Zhang et al., 2018). 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 (Sukumaran and Knowles, 2017), or incorporating budding speciation (Stadler et al., 2018). 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).

Supplementary Material

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.xwdbrv1d1.

References

Bardeleben, C., R. L. Moore, and R. K. Wayne. 2005a. 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.

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

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

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

Besenbacher, S., C. Hvilsom, T. Marques–Bonet, T. Mailund, and M. H. Schierup. 2019. Direct estimation of mutations in great apes reconciles phylogenetic dating. Nature Ecology and Evolution 3:286–292.

Bouckaert, R., T. G. Vaughan, J. Barido-Sottani, S. Duchêne, M. Fourment, A. Gavryushkina, J. Heled, G. Jones, D. Kühnert, N. De Maio, M. Matschiner, F. K. Mendes, N. F. Müller, H. A. Ogilvie, L. du Plessis, A. Popinga, A. Rambaut, D. Rasmussen, I. Siveroni, M. A. Suchard, C.-H. Wu, D. Xie, C. Zhang, T. Stadler, and A. J. Drummond. 2019. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. PLOS Computational Biology 15:1–28.

Bragg, J. G., S. Potter, A. C. Afonso Silva, C. J. Hoskin, B. Y. H. Bai, and C. Moritz. 2018. Phylogenomics of a rapid radiation: the Australian rainbow skinks. BMC Evolutionary Biology 18:15.

Bromham, L. 2011. 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.

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

Carlin, B. P. and S. Chib. 1995. Bayesian model choice via Markov chain Monte Carlo methods. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 57:473–484.

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

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

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

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

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. Journal of Molecular Evolution 17:368–376.

Field, D. J., J. S. Berv, A. Y. Hsiang, R. Lanfear, M. J. Landis, and A. Dornburg. 2020. Timing the extant avian radiation: the rise of modern birds, and the importance of

555 modeling molecular rate variation. Pages 159–181 in Pennaraptoran theropod dinosaurs:
556 past progress and new frontiers (M. Pittman and X. Xu, eds.) vol. 440. Bulletin of the
557 American Museum of Natural History, New York, NY.

558 Finarelli, J. A. and J. J. Flynn. 2009. Brain-size evolution and sociality in Carnivora.
559 Proceedings of the National Academy of Sciences 106:9345–9349.

560 Friscia, A. R., B. Van Valkenburgh, and A. R. Biknevicius. 2007. An ecomorphological
561 analysis of extant small carnivorans. Journal of Zoology 272:82–100.

562 Gaboriau, T., F. K. Mendes, S. Joly, D. Silvestro, and N. Salamin. 2020. A multi-platform
563 package for the analysis of intra- and interspecific trait evolution. Methods in Ecology
564 and Evolution 11:1439–1447.

565 Gavryushkina, A., T. A. Heath, D. T. Ksepka, T. Stadler, D. Welch, and A. J. Drummond.
566 2017. Bayesian total-evidence dating reveals the recent crown radiation of penguins.
567 Systematic Biology 66:57–73.

568 Gavryushkina, A., D. Welch, T. Stadler, and A. J. Drummond. 2014. Bayesian inference of
569 sampled ancestor trees for epidemiology and fossil calibration. PLOS Computational
570 Biology 10:e1003919.

571 Heled, J., D. Bryant, and A. J. Drummond. 2013. Simulating gene trees under the
572 multispecies coalescent and time-dependent migration. BMC Evolutionary Biology
573 13:44.

574 Heled, J. and A. J. Drummond. 2010. Bayesian inference of species trees from multilocus
575 data. Molecular Biology and Evolution 27:570–580.

576 Heled, J. and A. J. Drummond. 2012. Calibrated tree priors for relaxed phylogenetics and
577 divergence time estimation. Systematic Biology 61:138–149.

578 Ho, S. Y. W. and M. J. Phillips. 2009. Accounting for calibration uncertainty in

phylogenetic estimation of evolutionary divergence times. Systematic Biology 58:367–380.

Huelsenbeck, J. P., J. J. Bull, and C. W. Cunningham. 1996. Combining data in phylogenetic analysis. Trends in Ecology and Evolution 11:152–157.

Hugall, A. F., R. Foster, M. S. Y. Lee, and M. Hedin. 2007. Calibration choice, rate smoothing, and the pattern of tetrapod diversification according to the long nuclear gene RAG-1. Systematic Biology 56:543–563.

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

Koepfli, K.-P., J. Pollinger, R. Godinho, J. Robinson, A. Lea, S. Hendricks, R. M. Schweizer, O. Thalmann, P. Silva, Z. Fan, et al. 2015. Genome-wide evidence reveals that African and Eurasian golden jackals are distinct species. Current Biology 25:2158–2165.

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

Lindblad-Toh, K., C. M. Wade, T. S. Mikkelsen, E. K. Karlsson, D. B. Jaffe, M. Kamal, M. Clamp, J. L. Chang, E. J. Kulbokas, M. C. Zody, et al. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature 438:803–819.

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

Maddison, W. P. 1997. Gene trees in species trees. Systematic Biology 46:523–536.

Matzke, N. J. and A. Wright. 2016. Inferring node dates from tip dates in fossil canidae: the importance of tree priors. Biology Letters 12:20160328.

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

Mendes, F. K. and M. W. Hahn. 2016. Gene tree discordance causes apparent substitution rate variation. Systematic Biology 65:711–721.

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

Mirarab, S., N. Nguyen, and T. Warnow. 2014. PASTA: ultra-large multiple sequence alignment. Pages 177–191 in Research in Computational Molecular Biology (R. Sharan, ed.) Springer International Publishing, Cham.

Nowak, R. M. 1991. Walker's mammals of the world vol. 2. Johns Hopkins University Press, Baltimore, Maryland.

Ogilvie, H. A., R. R. Bouckaert, and A. J. Drummond. 2017. StarBEAST2 brings faster species tree inference and accurate estimates of substitution rates. Molecular Biology and Evolution 34:2101–2114.

Ogilvie, H. A., J. Heled, D. Xie, and A. J. Drummond. 2016. Computational performance and statistical accuracy of *BEAST and comparisons with other methods. Systematic Biology 65:381–396.

Pamilo, P. and M. Nei. 1988. Relationships between gene trees and species trees. Molecular Biology and Evolution 5:568–583.

Parham, J. F., P. C. J. Donoghue, C. J. Bell, T. D. Calway, J. J. Head, P. A. Holroyd, J. G. Inoue, R. B. Irmis, W. G. Joyce, D. T. Ksepka, J. S. L. Patané, N. D. Smith, J. E. Tarver, M. V. Tuinen, Z. Yang, K. D. Angielczyk, J. M. Greenwood, C. A. Hipsley, L. Jacobs, P. J. Makovicky, J. Müller, K. T. Smith, J. M. Theodor, R. C. M. Warnock, and M. J. Benton. 2012. Best practices for justifying fossil calibrations. Systematic Biology 61:346–359.

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

Rambaut, A., A. J. Drummond, D. Xie, G. Baele, and M. A. Suchard. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Systematic Biology 67:901–904.

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

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

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

Shang, S., X. Wu, J. Chen, H. Zhang, H. Zhong, Q. Wei, J. Yan, H. Li, G. Liu, W. Sha, and H. Zhang. 2017. The repertoire of bitter taste receptor genes in canids. Amino Acids 49:1159–1167.

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

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

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

Sukumaran, J. and M. T. Holder. 2010. DendroPy: a Python library for phylogenetic computing. Bioinformatics 26:1569–1571.

Sukumaran, J. and L. L. Knowles. 2017. Multispecies coalescent delimits structure, not species. Proceedings of the National Academy of Sciences 114:1607–1612.

Suzuki, Y., G. V. Glazko, and M. Nei. 2002. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. Proceedings of the National Academy of Sciences 99:16138–16143.

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

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

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

Van Valkenburgh, B. and K. P. Koepfli. 1993. Cranial and dental adaptation to predation in canids. Symposia of the Zoological Society of London 65:15–37.

Wang, X. 1993. Transformation from plantigrady to digitigrady: functional morphology of locomotion in *Hesperocyon* (Canidae: Carnivora). American Museum novitates 3069:1–23.

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

Wang, X. and R. H. Tedford. 2008. Dogs: their fossil relatives and evolutionary history. Columbia University Press, New York, NY.

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

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

Wen, D. and L. Nakhleh. 2018. Coestimating reticulate phylogenies and gene trees from multilocus sequence data. Systematic Biology 67:439–457.

Wickham, H. 2016. ggplot2: Elegant Graphics for Data Analysis. 2nd ed. Springer-Verlag, New York.

Wu, C.-H. and A. J. Drummond. 2011. Joint inference of microsatellite mutation models, population history and genealogies using transdimensional Markov chain Monte Carlo. Genetics 188:151–164.

Yu, G., D. K. Smith, H. Zhu, Y. Guan, and T. T. Lam. 2017. 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.

Zhang, C., H. A. Ogilvie, A. J. Drummond, and T. Stadler. 2018. Bayesian inference of species networks from multilocus sequence data. Molecular Biology and Evolution 35:504–517.

Zhang, C., T. Stadler, S. Klopfstein, T. A. Heath, and F. Ronquist. 2016. Total-evidence dating under the fossilized birth-death process. Systematic Biology 65:228–249.

Zhang, R., A. J. Drummond, and F. K. Mendes. 2021. Scalable total-evidence inference from molecular and continuous characters in a Bayesian framework. bioRxiv.

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

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

Zuckerkandl, E. and L. Pauling. 1965. Molecules as documents of evolutionary history. Journal of Theoretical Biology 8:357–366.

Fig. 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.

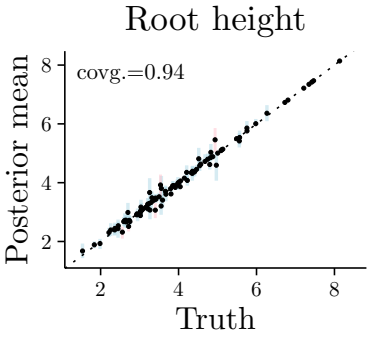
Fig. 2. Parameter posterior means against their true simulated values, for 97 simulations (3 were excluded due to convergence issues). Blue lines correspond to the 95%-HPD intervals for a parameter, one line per simulation, when the true value was contained within the interval. When the true value was 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.

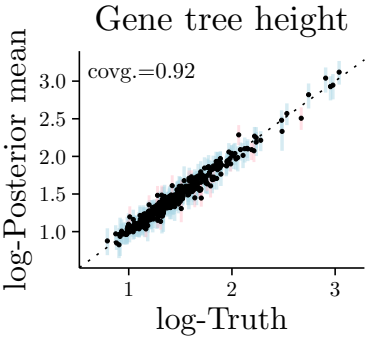
Fig. 3. Branch length changes resulting from concatenation. The tree shown is the maximum clade credibility (MCC) tree with mean internal node ages from the FBD-MSC posterior distribution. When the length estimated by FBD-concatenation was longer than for FBD-MSC, the additional length is shown as an 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 not meeting the 0.5% frequency threshold described in the main text.

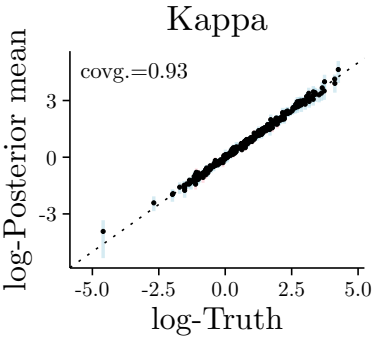
Fig. 4. Posterior distribution of clock rates. Posterior probabilities of molecular clock rates (a) and morphological clock rates (b) were calculated using bin widths of 2×10^{-5} and 5×10^{-4} respectively.

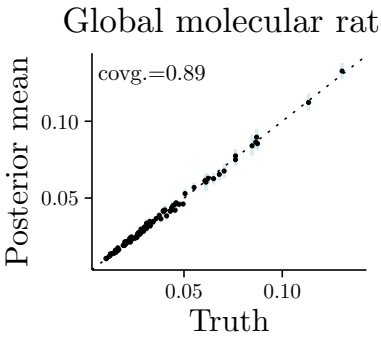
Fig. 5. Tempo of Caninae evolution. Crown ages estimated by fossilized birth-death with multispecies coalescent (FBD-MSC) and with concatenation (FBD-concatenation) models (a), compared with lineages-through-time (LTT) curves including extinct lineages (b). Posterior mean internal node ages (solid circles) 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.

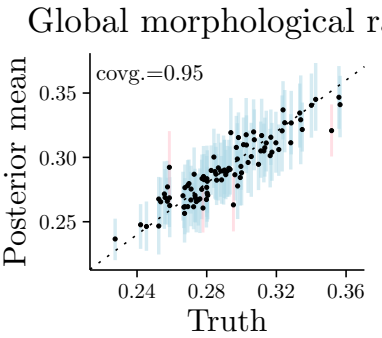
Fig. 6. Correlation between log-node ages from the posterior distributions of species trees pruned of morphology-only taxa. Internal nodes from the pruned FBD-MSC MCC tree are drawn as ellipses with their labels from Fig. 3. Ellipses are centered on the mean estimate of the log-node age for both methods.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

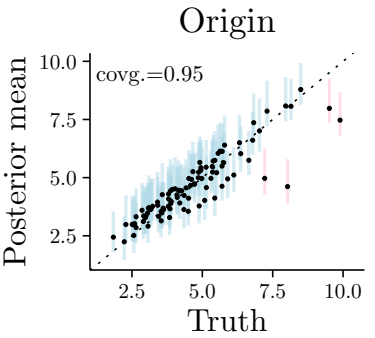


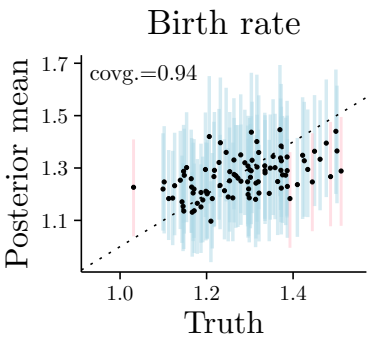


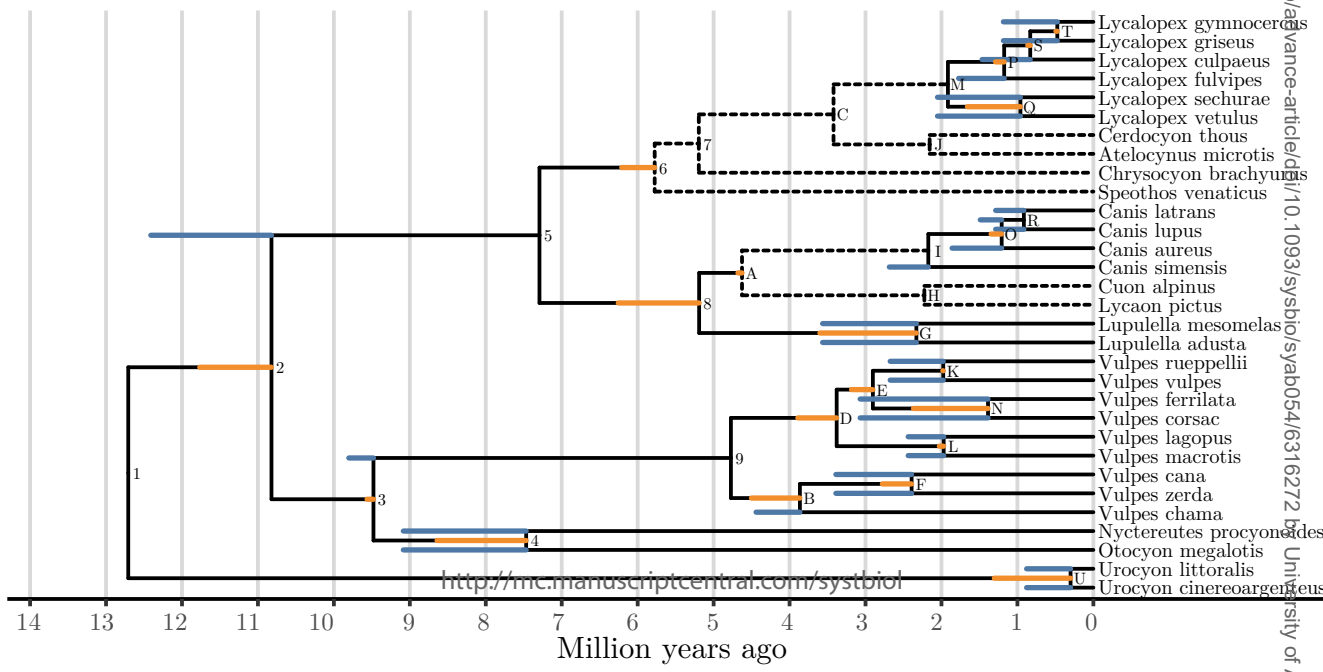


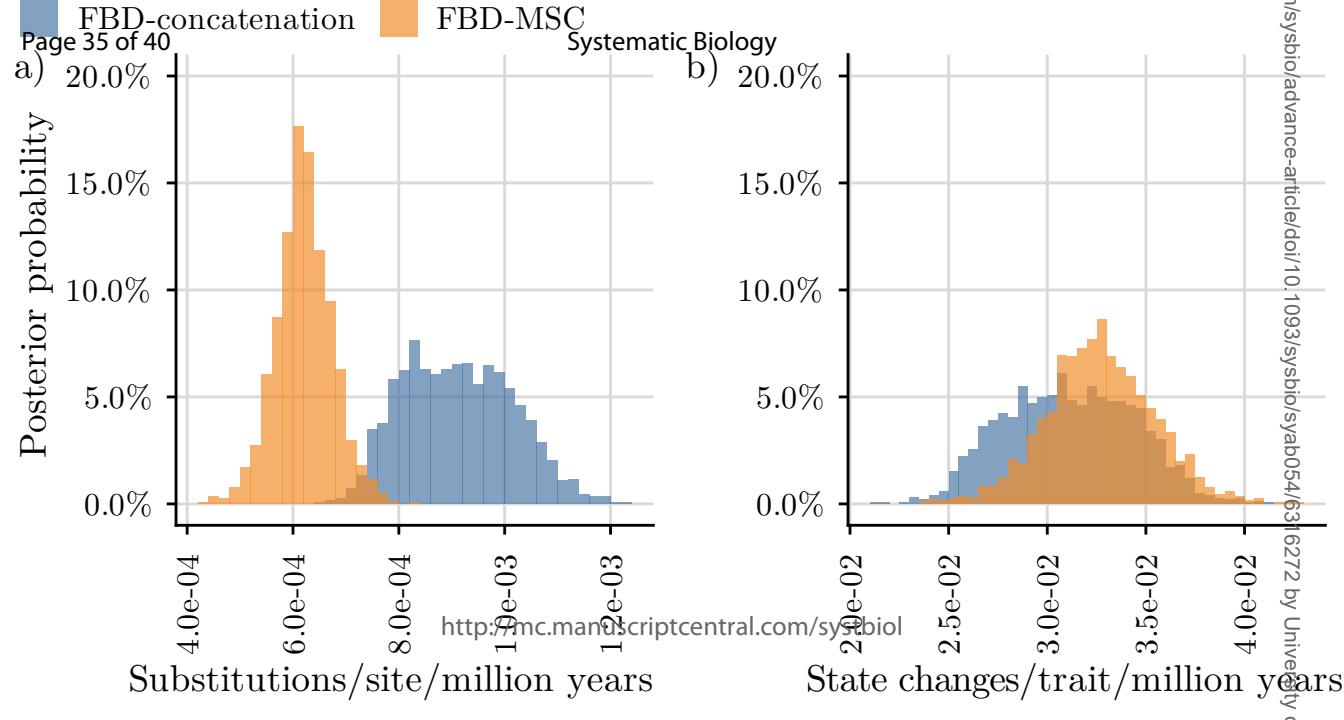


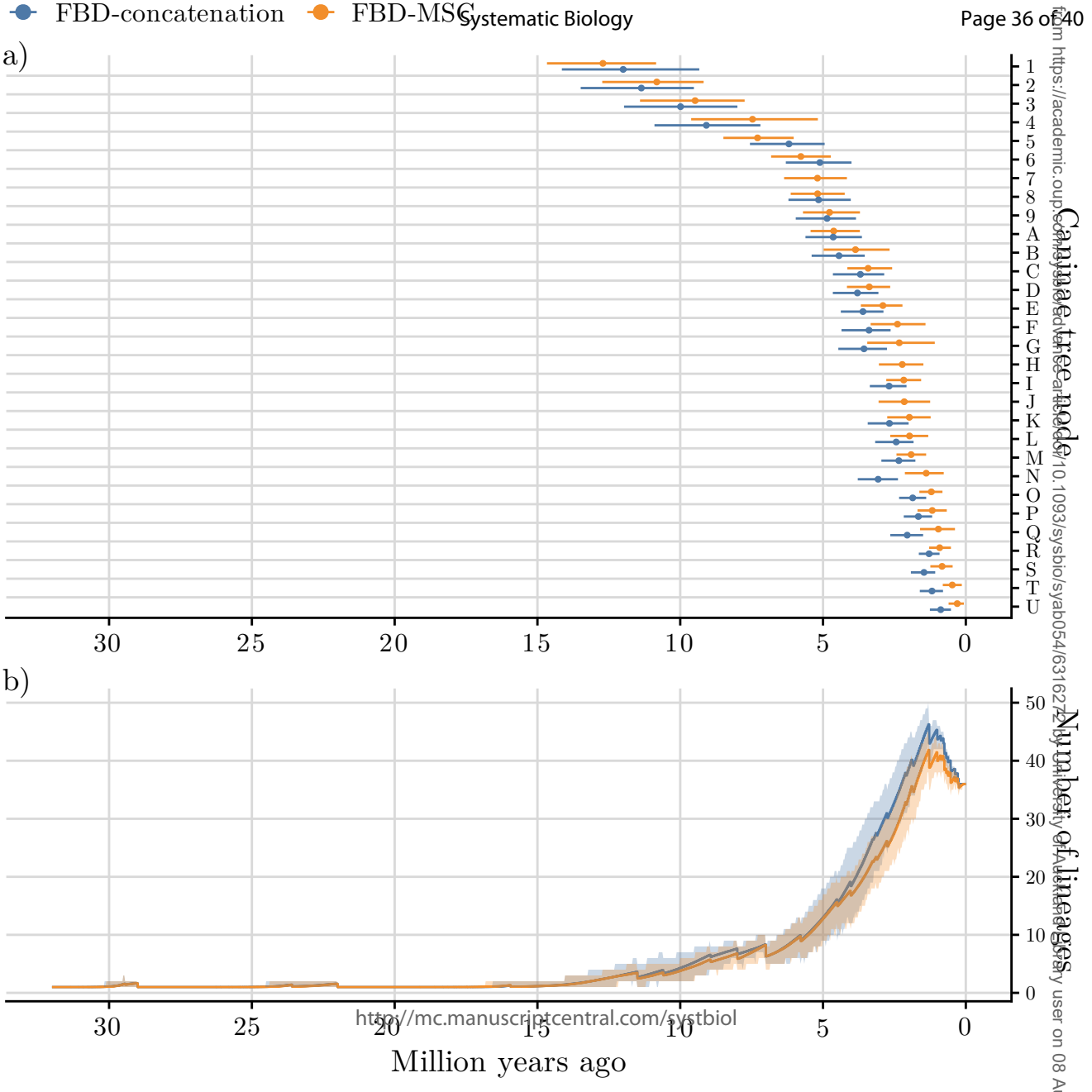












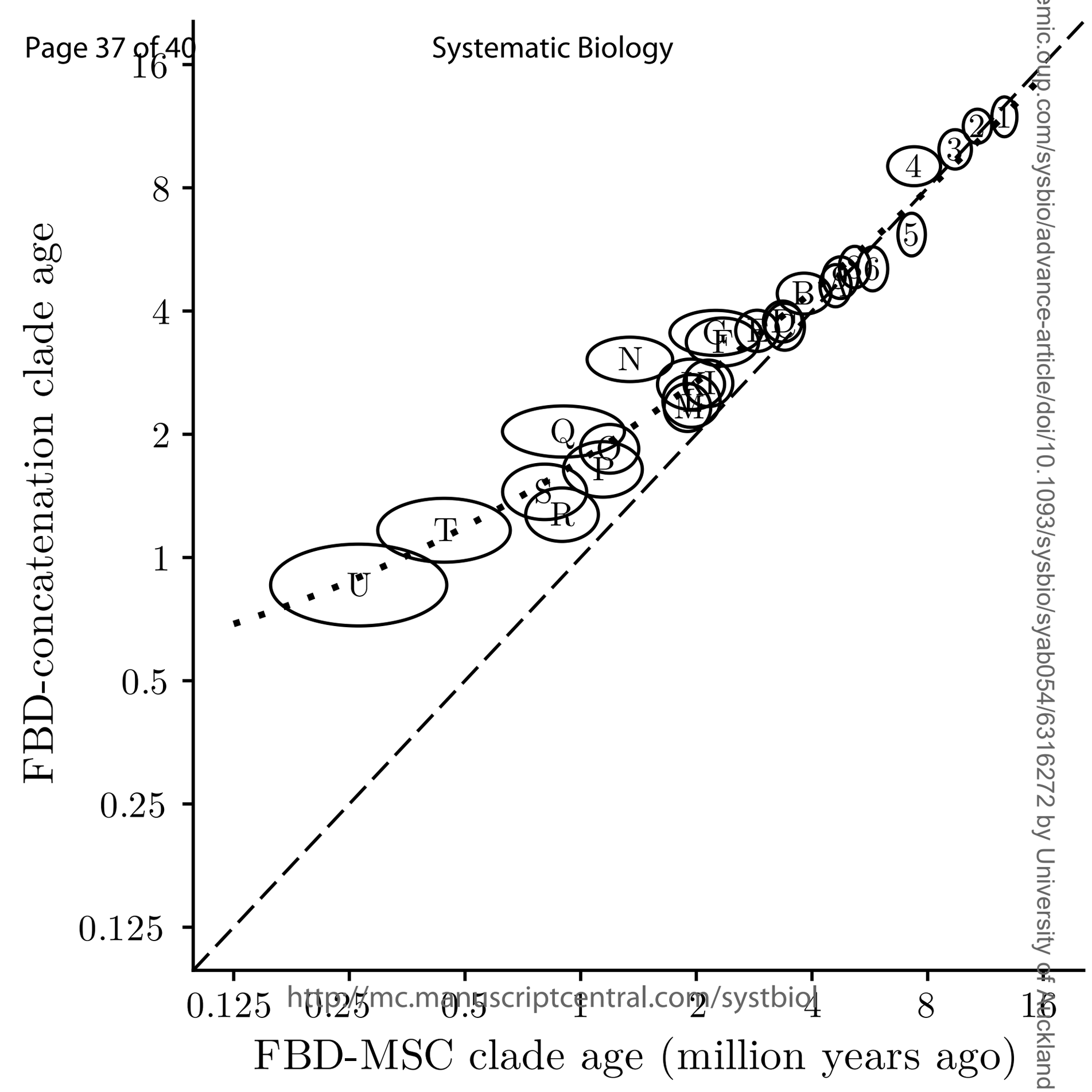


Table 1: Number of replicate MCMC chains in which a particular topology was found within the 95% HPD interval (the expected number under a correct implementation is 95). Full details of operator configurations are provided in the Supplementary Material.

Configuration \ Topology	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

Table 2: Parameter estimates. All values are posterior mean estimates followed in brackets by the bounds of 95% highest-posterior-density intervals. 'Molecular clock rate' is the global (mean) rate that scales the relative locus-specific molecular rates. *expected number of character state changes (i.e., substitutions for molecular data) 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 (see also Supplementary Table S8).

Parameter	FBD-concatenation	FBD-MSC
Molecular clock rate* ($\times 10^{-3}$)	0.91 (0.74-1.11)	0.62 (0.52-0.73)
Morphological clock rate* ($\times 10^{-2}$)	3.11 (2.50-3.68)	3.27 (2.69-3.78)
Mean population size**	NA	1.51 (1.08-1.97)
Diversification rate ($\lambda - \mu$)	0.09 (0.01-0.18)	0.10 (0.01-0.19)
Turnover ($\mu \div \lambda$)	0.85 (0.72-0.98)	0.86 (0.73-0.99)
Sampling proportion ($\psi \div (\psi + \mu)$)	0.16 (0.07-0.25)	0.15 (0.07-0.25)

Table 3: Computational performance of Caninae analyses under the two models.

Model	Combined CPU hours	Min. ESS	ESS per hour
FBD-MSC	399	274	0.69
FBD-concatenation	326	229	0.70