

1 Putting the F in FBD analyses: tree constraints or morphological 2 data ?

3 Joëlle Barido-Sottani^{1,*}, Alexander Pohle², Kenneth De Baets³, Duncan Murdoch⁴, and
4 Rachel C. M. Warnock⁵

5 ¹*Institut de Biologie de l'ENS (IBENS), École normale supérieure, CNRS, INSERM,
6 Université PSL, 75005 Paris, France*

7 ²*Paläontologisches Institut und Museum, Universität Zürich, 8006 Zürich, Switzerland*

8 ³*Institute of Evolutionary Biology, Faculty of Biology, Biological and Chemical Research
9 Centre, University of Warsaw, 02-089 Warsaw, Poland*

10 ⁴*Oxford University Museum of Natural History, Oxford, United Kingdom*

11 ⁵*GeoZentrum Nordbayern, Department of Geography and Geosciences, Friedrich-Alexander
12 Universität Erlangen-Nürnberg, Erlangen, Germany*

13 ^{*}*Correspondence to be addressed to: joelle.barido-sottani@m4x.org*

14 1 Abstract

15 The fossilized birth-death (FBD) process provides an ideal model for inferring phylogenies from
16 both extant and fossil taxa. Using this approach, fossils (with or without character data) are
17 directly considered as part of the tree. This leads to a statistically coherent prior on divergence
18 times, where the variance associated with node ages reflects uncertainty in the placement of fossil
19 taxa in the phylogeny. Since fossils are typically not associated with molecular sequences, additional
20 information is required to place fossils in the tree. Previously, this information has been provided
21 in two different forms: using topological constraints, where the user specifies monophyletic clades
22 based on established taxonomy, or so-called total-evidence analyses, which use a morphological data
23 matrix with data for both fossil and extant specimens in addition to the molecular alignment. In
24 this work, we use simulations to evaluate these different approaches to handling fossil placement in

25 FBD analyses, both in ideal conditions and in datasets including uncertainty or even errors. We also
26 explore how rate variation in fossil recovery or diversification rates impacts these approaches. We
27 find that the extant topology is well recovered under all methods of fossil placement. Divergence
28 times are similarly well recovered across all methods, with the exception of constraints which
29 contain errors. These results are consistent with expectations: in FBD inferences, divergence times
30 are mostly informed by fossil ages, so variations in the position of fossils strongly impact these
31 estimates. On the other hand, the placement of extant taxa in the phylogeny is driven primarily
32 by the molecular alignment. We see similar patterns in datasets which include rate variation,
33 however one notable difference is that relative errors in extant divergence times increase when
34 more variation is included in the dataset, for all approaches using topological constraints, and
35 particularly for constraints with errors. Finally, we show that trees recovered under the FBD
36 model are more accurate than those estimated using non-FBD (i.e., non-time calibrated) inference.
37 This result holds even with the use of erroneous fossil constraints and model misspecification under
38 the FBD. Overall, our results underscore the importance of core taxonomic research, including
39 morphological data collection and species descriptions, irrespective of the approach to handling
40 phylogenetic uncertainty using the FBD process.

41 2 Introduction

42 Time-calibrated trees provide a crucial basis for hypothesis-testing in the life and earth sciences.
43 Phylogenetic dating combines molecular and fossil evidence, allowing us to reconstruct a timeline of
44 events that are otherwise not directly observable. Within a Bayesian framework, temporal evidence
45 is incorporated via the tree prior or tree model. The fossilised birth-death (FBD) process explic-
46 itly combines the lineage diversification and fossil recovery processes, providing an ideal model for
47 inferring phylogenies from both extant species and fossil specimens (Stadler, 2010). Using this
48 approach, fossils (with or without character data) are directly considered as part of the tree (Heath
49 et al., 2014; Gavryushkina et al., 2014, 2017; Zhang et al., 2016). This leads to a statistically
50 coherent prior on divergence times, where the variance associated with node ages reflects the in-
51 completeness of the fossil record, as well as uncertainty associated with the placement of fossil taxa
52 in the phylogeny. Bayesian inference using the FBD process as a tree prior also allows for reliable
53 estimation of the diversification and sampling parameters. This model has been successfully applied
54 to datasets of living and fossil taxa (Schuster et al., 2018; Šmíd and Tolley, 2019; Thomas et al.,

55 2020; Pohle et al., 2022).

56 The initial implementation of the FBD model assumed constant diversification (birth and death)
57 and fossil recovery rates through the entire phylogeny (Heath et al., 2014). However, a wide range of
58 factors, from biological and geological processes to collection practices, contribute to the probability
59 a given organism will be sampled in the fossil record (Kidwell and Holland, 2002; Smith and
60 McGowan, 2011; Benson et al., 2021; Raja et al., 2022). As a consequence, fossil recovery potential
61 varies substantially across time, space and taxa. Other variables, such as environmental conditions
62 and phenotypic traits, contribute to variation in diversification (speciation and extinction) rates.
63 Extensions of the FBD process have integrated these variations into the model (Gavryushkina et al.,
64 2014; Zhang et al., 2016), however, many empirical analyses are still done under the constant-
65 rate assumption, for several reasons such as computational cost, lack of precise knowledge of rate
66 changes, and ease of setup and interpretation.

67 One challenge in integrating fossil specimens in FBD analyses is that unlike extant species, fossils are
68 only exceptionally associated with molecular sequences. As a result, additional information needs
69 to be added to the inference to allow fossils to be placed in the tree topology. In previous research,
70 this information has been provided in two different forms, which can be used separately or in
71 combination. The first is topological constraints, which are constraints added by the user specifying
72 that certain subsets of tips, extant or extinct, need to be monophyletic clades in the inferred
73 phylogeny. These constraints are generally based on the taxonomy, with the constrained clades
74 corresponding to genera and/or higher classifications. Topological constraints use information
75 which is usually already available, and do not add computational complexity or cost to the inference.
76 However, they do not easily accommodate uncertainty in the taxonomy, which is present even for
77 well-known crown groups (Marx et al., 2016). Another approach is so-called total-evidence analyses
78 (Ronquist et al., 2012), which uses a morphological data matrix, with data for both fossil and
79 extant specimens, in addition to the molecular alignment. The morphological matrix is added to
80 the inference along with a morphological substitution model and a morphological clock model, and
81 contributes to the phylogenetic likelihood. Although total-evidence approaches better account for
82 the underlying empirical data, they are more costly both in the time and effort required to assemble
83 the matrix and in the added computational cost of the inference. In addition, the accuracy and
84 precision of the inference has been found to be strongly dependent on the size of the morphological
85 matrix (Barido-Sottani et al., 2020). Although these two approaches can in theory be combined

86 in the same analysis, in practice they are often viewed as separate alternatives (for instance, Šmíd
87 and Tolley (2019) uses constraints, while Thomas et al. (2020) uses a total-evidence approach).

88 In this work, we compare and evaluate different approaches to add fossil placement information in
89 FBD analyses, both in ideal conditions and in datasets including uncertainty or even errors. We
90 also use datasets containing variation in fossil recovery rates or speciation rates to explore whether
91 ignoring rate variations impacts fossil placement approaches.

92 **3 Methods**

93 **3.1 Simulations**

94 **3.1.1 Trees and fossils**

95 Our goal was to assess the performance of the FBD process, examining the impact of phylogenetic
96 uncertainty and model misspecification. The parameters of the simulation were constrained to
97 reflect values obtained for marine invertebrates.

98 Trees and fossils were simulated using the R packages TreeSim and FossilSim. The simulations
99 were conditioned on the number of extant tips ($n = 25$), with constant speciation rate ($\lambda = 0.11$),
100 extinction rate ($\mu = 0.1$) and fossil recovery rate ($\psi = 0.03$). We assumed complete sampling at
101 the present (i.e. the probability of extant species sampling $\rho = 1$). Parameters were selected such
102 that the expected origin time was 250 Myr and the expected number of fossils was 100. Simulated
103 data sets were filtered to select for trees with an origin time between 225 and 275 Myr and with
104 between 80 and 120 fossils, and simulations which did not fit these two criteria were discarded.

105 We also simulated two sets of trees with (a) variable fossil recovery rates ($\psi_1 = 0.02$ and $\psi_2 = 0.04$),
106 and (b) variable speciation rates ($\lambda_1 = 0.12$ and $\lambda_2 = 0.06$) and variable fossil recovery rates
107 ($\psi_1 = 0.02$ and $\psi_2 = 0.04$). The speciation rate in set (a) and the extinction rate in sets (a) and (b)
108 were fixed to the same values as in the constant birth-death simulation. In both variable sets, the
109 variation in rates was linked to a trait that was simulated along each tree under a Brownian motion
110 process. Traits were simulated with an initial value of 2.25 (for set (a)) or 1.5 (for set (b)) and
111 variance = 0.01. Trait values were then assigned to two discrete types: values < 0 were assigned to
112 type 1 (with rates ψ_1 and λ_1) and values > 0 were assigned to type 2 (with rates ψ_2 and λ_2). The

113 final trees were kept if each trait value was assigned to at least 10% of the tips. The parameters
114 of the BM process were calibrated so the final sampled trees contained an average of 8 to 11 trait
115 changes.

116 We generated 50 replicate phylogenies for each of the three simulation conditions.

117 **3.1.2 Characters**

118 Molecular sequence alignments of 1000 sites were simulated under the HKY + Γ model with five
119 discrete gamma rate categories ($\alpha = 0.25$). Branch rates were simulated under a lognormal un-
120 correlated clock model. For each tree replicate the average substitution rate was sampled from
121 a gamma distribution with an expected value = 1, and shape and scale parameters = 2 and 0.5,
122 respectively. The log of this rate was then used to define the mean of a lognormal distribution with
123 variance = 0.01 from which branch specific rates were independently drawn.

124 Morphological data matrices of 50 or 300 characters were simulated under a binary state Mk model
125 with a strict clock and a rate = 0.1.

126 **3.2 Inference**

127 We analysed each replicate using Bayesian phylogenetic inference in the BEAST2 framework
128 (Bouckaert et al., 2019), under the constant rate FBD process implemented in the package Sampled
129 Ancestors (SA) (Gavryushkina et al., 2014).

130 We examined the impact of five different ways of incorporating the phylogenetic uncertainty asso-
131 ciated with fossils:

- 132 1. Fossil samples without character data were assigned to the nearest extant ancestral node
133 using clade constraints (designated as “correct constraints”)(Fig. 1C).
- 134 2. As in (1) but with 2% of fossils assigned to the wrong node, selected at random (designated
135 as “constraints with errors”)(Fig. 1D).
- 136 3. Fossil samples without character data were assigned to the node above the nearest ancestral
137 node (designated as “imprecise constraints”)(Fig. 1E).

138 4. As in (1) but 5 nodes picked at random in the tree were collapsed, meaning all constraints
139 below that node were removed (designated as “collapsed constraints”)(Fig. 1F).

140 5. Character data (= 50 characters) was included for both fossil and extant samples, with no
141 additional constraints (designated as “total-evidence with n=50”)(Fig. 1G).

142 6. Character data (= 300 characters) was included for both fossil and extant samples, with no
143 additional constraints (designated as “total-evidence with n=300”)(Fig. 1G).

144 Configuration (1) corresponds to a situation with perfect information on fossil taxonomy (Fig. 1C),
145 while configuration (2) includes a low amount of misplaced fossils (Fig. 1D). Configurations (3)
146 and (4) include uncertainty in fossil placement in two different ways: in configuration (3) all clades
147 have some amount of uncertainty in fossil assignments (Fig. 1E), in configuration (4) a few clades
148 have no information on fossil placement within the clade, while the other clades are known perfectly
149 (Fig. 1F). Configurations (5) and (6) are total-evidence analyses, with respectively a low or high
150 amount of morphological characters for each fossil (Fig. 1G).

151 The FBD process was parameterized using the ‘canonical’ (speciation, extinction, sampling) pa-
152 rameterization, with an exponential prior (mean = 1.0) on the birth, death, sampling migration
153 parameters. A uniform prior $U(0, 1000)$ was used for the origin time of the process. We deliberately
154 chose uninformative priors for these parameters in order to minimize the influence of the priors on
155 the results. A lognormal prior for HKY parameter κ (mean = 1 and sd = 1.25), with a uniform
156 prior on the state frequencies. An exponential prior for alpha for gamma distributed rates (mean
157 = 1). The Mk model was applied to the binary character data. An exponential prior (mean = 1)
158 was used for the mean of the uncorrelated relaxed clock model and a gamma prior ($\alpha = 0.5396$ and
159 $\beta = 0.3819$) on the standard deviation.

160 Analyses were run for a minimum of 200,000,000 generations, sampling every 10,000 generations
161 and discarding 10% as burnin. We assessed convergence by calculating the ESS values for all model
162 parameters – if any ESS values were < 200, analyses were longer (this was only necessary for 19
163 replicates). Eight analyses that failed to converge after 2,000,000,000 generations were excluded.
164 We note that all of these are analyses in which fossils have no character data and are assigned to
165 the next nearest ancestor (Fig. 1F). The maximum number excluded for a given simulation and
166 inference scenario was 3.

167 For each replicate, we calculated the relative error of divergence time estimates (as the absolute

168 difference between the median estimate and the true value, divided by the true value), averaged
169 over all extant nodes, and the 95% HPD coverage, i.e. the proportion of extant nodes for which the
170 true age was contained within the 95% Highest Posterior Density interval. We also measured the
171 normalised Robinson-Foulds distance between the estimated phylogeny and the truth, averaged
172 over all posterior samples, for both the full tree including fossils and the reconstructed extant
173 phylogeny.

174 We also performed a separate unconstrained (i.e., non-FBD) tree inference using RevBayes. Since
175 these analyses contain no age information the trees produced are non-time calibrated and have
176 branch lengths in substitutions per site rather than in units of time. We used a uniform tree prior
177 on the topology, with an exponential prior ($\lambda = 10$, mean = 0.1) on the branch lengths. We
178 used the same settings as above for substitution models. We ran 4 independent chains for 200,000
179 generations, sampling every 200 generations, discarding 10% as burnin and combining the output.
180 We assessed convergence by calculating the ESS values for all model parameters as above and
181 examined the variance across chains. All runs converged. We measured the normalised Robinson-
182 Foulds distance between the estimated phylogeny and the truth, averaged over all posterior samples,
183 and compared those to the results obtained using the FBD inference.

184 4 Results

185 Figure 2 shows the results on the datasets simulated under constant rates, variable fossilization
186 rates, and variable birth and fossilization rates.

187 Under a constant-rate process, we can see that the extant topology is well recovered under all
188 methods of fossil placement. Divergence times are similarly well recovered across all methods, with
189 the marked exception of constraints which contain errors. Indeed, the relative error is much higher
190 and the coverage much lower in the presence of errors. Overall, these results are quite consistent
191 with expectations: in FBD inferences, divergence times are mostly informed by fossil ages, so
192 variations in the position of fossils will strongly impact these estimates. On the other hand, the
193 placement of extant taxa in the phylogeny is driven primarily by the molecular alignment, and thus
194 is mostly independent from the handling of fossil specimens. As expected, fossil placement methods
195 also impact the accuracy of the estimate of the full phylogeny. The best performance is obtained
196 with correct constraints, followed by imprecise constraints and total-evidence with a high number

197 of morphological characters, whereas the worst estimates are obtained using total-evidence with a
198 low number of morphological characters.

199 We see very similar patterns in datasets which include rate variation. One notable difference
200 is that relative errors in extant divergence times are higher the more variation is included in the
201 dataset for all approaches using topological constraints, and particularly for constraints with errors.
202 Interestingly, total-evidence approaches show the same levels of relative error in extant divergence
203 times regardless of rate variation in the simulation models. One possible explanation is that total-
204 evidence approaches include more uncertainty than topological constraints and so are able to better
205 compensate for model mismatches between the simulation and the inference. The effect of simulated
206 rate variation is almost null on the other metrics, i.e. the coverage of extant divergence times and
207 the RF distances for extant and full phylogenies.

208 Figure. 3 shows the results obtained using the FBD versus non-FBD (i.e., without fossil age in-
209 formation) inference. The comparison shows that trees recovered under the FBD model are more
210 accurate, in terms of Robinson-Foulds distance – this applies to both the extant topology estimated
211 using the molecular data only or the total-evidence matrix and the full topology. This result holds
212 even with the use of erroneous fossil constraints and model misspecification under the FBD. Inter-
213 estingly, the unconstrained topology estimates are more accurate for trees simulated using constant
214 diversification and sampling rates than trees simulated with rate variation. This might reflect a
215 better correspondence between the uniform tree model and a constant birth-death sampling pro-
216 cess. The inclusion of fossil taxa increases the accuracy of the extant tree. More morphological
217 character data also leads to improved accuracy of both the extant and full trees.

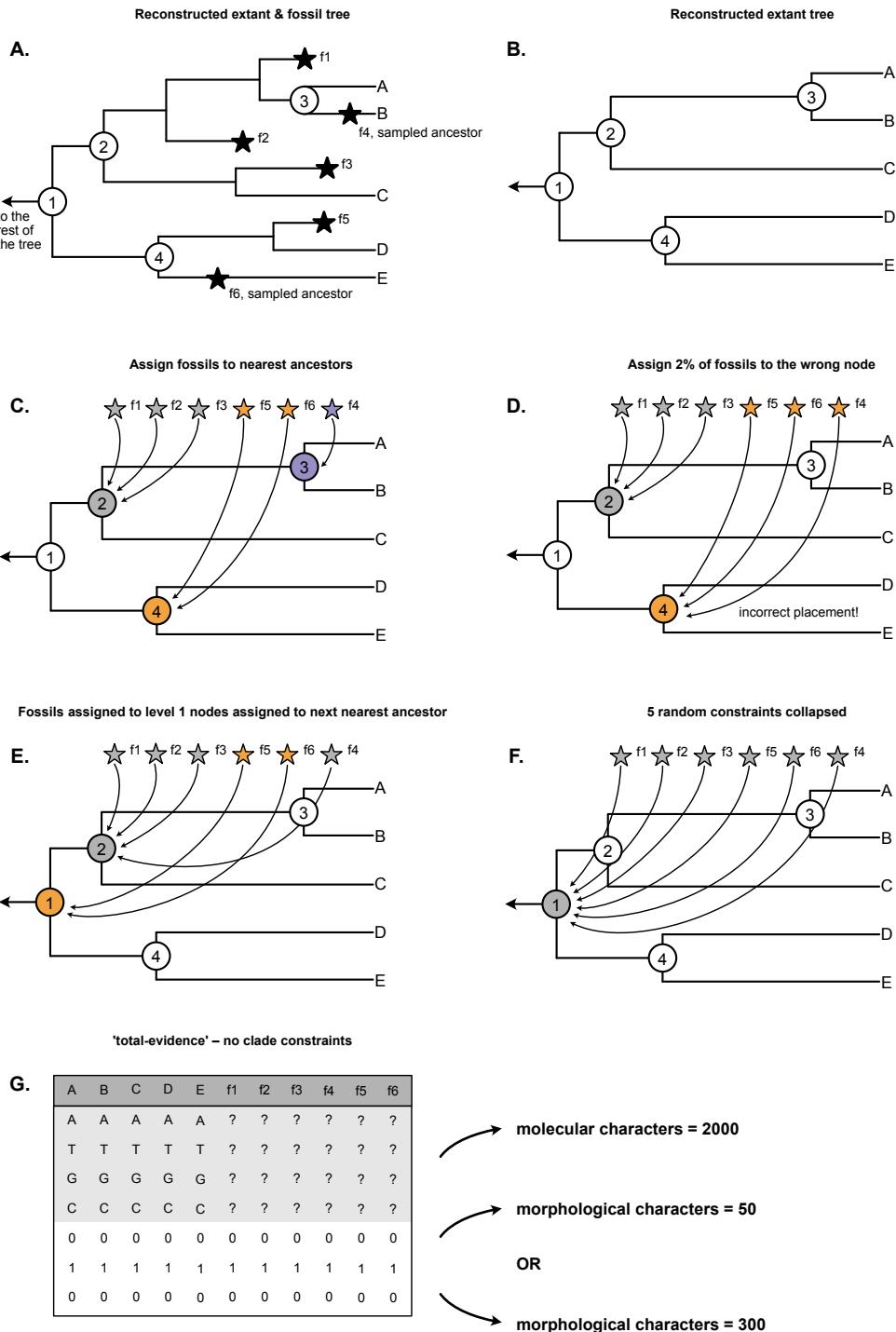


Figure 1: Schematic representation of the different analyses performed in this study. First, we simulate a full tree with fossil samples (A), from which the true extant tree can be obtained (B). We then set up clade constraints according to different rules: in (C), all clade constraints are correct and complete. In (D), some fossils are assigned to the wrong clade. In (E), all clades have a low level of taxonomic uncertainty. In (F), some subclades are fully unknown while the rest of the tree is fully known. Finally, we also perform two total-evidence analyses (G).

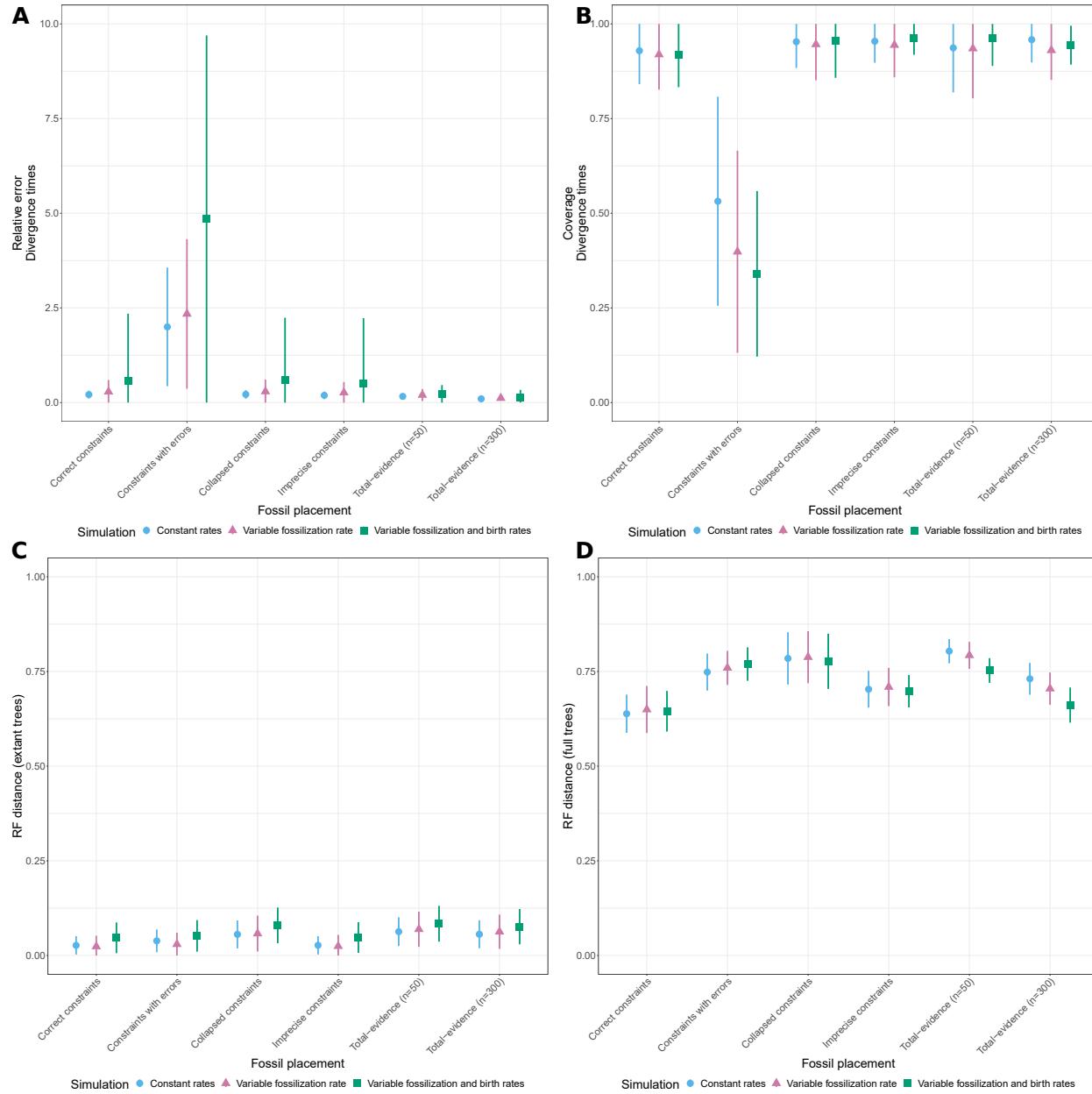


Figure 2: Results of FBD analyses for simulations under a constant birth-death process (red), variation in fossilization rates (green) or variation in birth and fossilization rates (blue). Absolute relative error (A) and 95% HPD coverage (B) of divergence time estimates, averaged over all nodes in the extant phylogeny. Average Robinson-Foulds distance between posterior samples and the true tree, for the extant tree only (C) or the full tree including fossil samples (D). All measures show the average and standard deviation across all replicates for different fossil placement approaches.

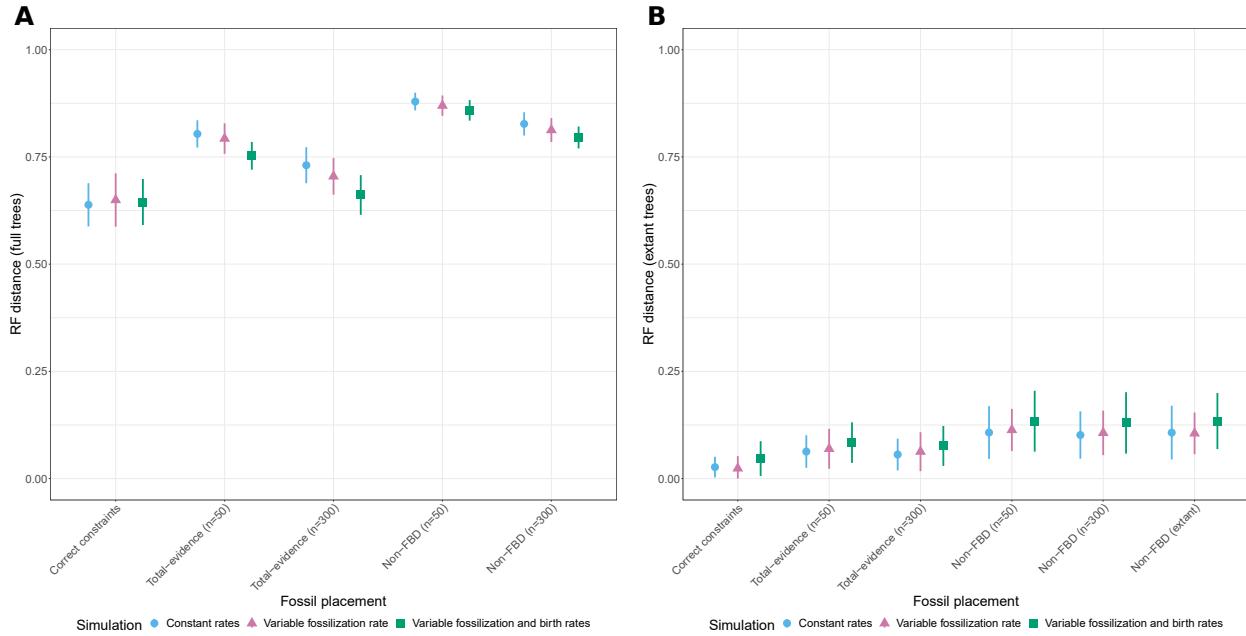


Figure 3: Comparison with non-FBD analysis (i.e., with no fossil age information) for simulations under a constant birth-death process (red), variation in fossilization rates (green) or variation in birth and fossilization rates (blue). Average Robinson-Foulds distance between posterior samples and the true tree, for the extant tree only (C) or the full tree including fossil samples (D). All measures show the average and standard deviation across all replicates for different fossil placement approaches.

218 5 Discussion

219 The fossilised birth-death process offers flexible opportunities to reconstruct dated phylogenies in-
220 corporating fossils under a mechanistic framework. Here, we explored different options for handling
221 the phylogenetic or taxonomic uncertainty associated with fossil samples using simulations.

222 We found that alternative approaches to including fossils had relatively little impact on the accuracy
223 of the extant topology (based on RF distances, Fig. 2C-D). As noted above, the signal for the extant
224 topology largely comes from the molecular sequence alignment, which explains why the extant
225 phylogeny was reasonable even when a small portion of fossils were placed erroneously in the tree.
226 Using fixed constraints to assign fossils to nodes produced more accurate trees, compared to trees
227 recovered using morphological characters to place the fossils. However, this represents an idealised
228 scenario, as reliable and precise taxonomic information used to inform constraints will often not be
229 available (see below for further discussion on data quality and availability). The results obtained
230 using non-time calibrated tree inference show that trees recovered with the inclusion of fossils with
231 character data are more accurate than trees based on the extant taxa only (Fig. 3A-B). We also
232 confirm that results are even better when the fossil age information is taken into account using
233 the FBD process. Previous simulations studies have also shown that the inclusion of fossils and
234 age information can improve the accuracy of the topology among extant taxa (Mongiardino Koch
235 et al., 2021).

236 Different approaches to handling taxonomic uncertainty had a more notable impact on divergence
237 times. We found that the best possible scenario, in terms of recovering accurate node ages (relative
238 error and coverage), was to assign fossils to the correct nearest node in the extant tree (Fig. 2A-B).
239 This is unsurprising, as this is equivalent to fixing large portions of the topology using correct
240 monophyletic clade constraints. But we further show that the use of less precise clade constraints
241 (i.e., assigning fossils to larger or more inclusive clades, Fig. 1E) also recovers accurate estimates,
242 similar to those obtained inferring the position of fossils based on morphological character data.
243 Similarly, Heath et al. (2014) found using simulations that less precise constraints increased the
244 variance but did not reduce accuracy of posterior divergence estimates (see also O'Reilly and
245 Donoghue (2020)). However, our results show that when only a small proportion (2%) of fossils
246 are assigned to incorrect nodes, overall accuracy decreases. Assigning larger proportions of fossils
247 to incorrect nodes will reduce accuracy further. Preliminary runs using 10% of fossils assigned

248 to incorrect nodes lead to coverage < 0.1 (results not shown). Together, our results suggest that
249 uncertainty is a much less critical issue for inference than outright errors. Thus, we recommend that
250 when in doubt, it is better to err on the side of caution and to either use larger and more inclusive
251 clade constraints, since this does not compromise the overall accuracy of results and only leads to a
252 small overall decrease in precision, or to infer the position of fossils using morphological data when
253 available. The latter is preferential since the posterior will best reflect uncertainty associated with
254 the placement of fossil taxa. In addition, if clade constraints are overly conservative (i.e., based on
255 very large clades or higher level taxonomic divisions) there will be too much uncertainty and the
256 analyses might fail to converge. ‘Imprecise constraints’ was the only simulation scenario in which
257 we had to exclude replicates due to lack of convergence.

258 We also examined the impact of model violation on the accuracy of topology and divergence times
259 recovered under the FBD model, by varying sampling rate only or both sampling and diversification
260 rates. Not accounting for rate variation in these parameters has a modest impact on the accuracy
261 of the topology – again reflecting the fact that the signal for topology predominantly comes from
262 the character data (Fig. 2C-D). Model misspecification has a much more discernible impact on the
263 divergence estimates (Fig. 2A-B). This is because the signal for divergence times largely comes
264 from the distribution of fossil sampling times and relies more on the birth-death sampling model
265 being correct. However, in most scenarios coverage for node ages remains high. These general
266 results also match the findings of previous simulation studies that explored model misspecification
267 under the FBD process (Heath et al., 2014). There are challenging identifiability issues associated
268 with birth-death processes (Louca and Pennell, 2020; Louca et al., 2021). In particular, Louca and
269 Pennell (2020) identified ‘congruence classes’, within which infinitely many diversification histories
270 can have the same likelihood. This means that applying oversimplified birth-death models can result
271 in highly misleading results. Our findings show that reliable phylogenies can be obtained under the
272 FBD process, even if the underlying process is more complex than the model used for inference. The
273 most notable exception are the results obtained with erroneous fossil constraints – when a portion
274 of fossils are incorrectly placed in the tree, the relative negative impact of model misspecification
275 is worse. Since we can rarely be certain of the underlying diversification or sampling process, this
276 further underscores the need to use taxonomic constraints with extreme caution. Furthermore, our
277 results demonstrate that total-evidence analyses are more robust against model misspecification,
278 adding to the list of benefits associated with this approach.

279 We note that our simulation design represents an idealised scenario in several ways. For instance,
280 we only simulated binary characters for our morphological alignments, whereas empirical data
281 matrices often include multi-state characters or characters with hierarchical state dependencies. In
282 addition, our matrices had no missing data, which is unrealistic for empirical datasets including
283 fossils. These two factors likely contribute to the good overall performance of the total-evidence
284 analyses in our study.

285 Ordinarily clade constraints and ‘total-evidence’ analyses are treated as two distinct approaches,
286 but they can easily be combined. In fact this might be desirable, since several previous simulation
287 studies have shown that the accuracy of parameters estimated under the FBD process increases
288 with fossil sampling (Heath et al., 2014; Barido-Sottani et al., 2020). In some cases, we might only
289 have abundant morphological character data associated with a subset of fossils. For example, we
290 might have fossils with fewer traits (e.g., fossil cephalopod shells), alongside rarer but exceptionally
291 preserved specimens with more informative traits (e.g., fossil cephalopods associated with jaws and
292 or soft-tissues). Shell characteristics that are diagnostic of higher-level taxonomic groups could be
293 used to define clade constraints for specimens associated with a smaller amount of data. Similarly,
294 in palaeobotany, rarely preserved tissues (e.g., flowers or seeds) can have a small number of traits
295 (i.e., 2-3) that would be considered too few to construct a matrix for phylogenetic analysis, but
296 that are nonetheless considered definitive synapomorphies of certain plant groups. Another example
297 could be ammonoid aptychi or lower jaws, which have been mapped to phylogenies based on a small
298 number of specimens with known jaws, relative to those where shells/moulds are known (Engeser
299 and Keupp, 2002).

300 We might also have different types of evidence associated with the presence of a group. For instance,
301 worm eggs can sometimes be assigned to higher taxonomic ranks based on the presence of specific
302 structures (Hugot et al., 2014), but these traits can have little to do with the adult morphology
303 of exceptionally preserved worms. Other examples could include trace fossils (e.g., where they
304 predate body fossils), molecular fossils (e.g., biomarkers), or exceptional preservation of different
305 ontogenetic stages (e.g., small larvae in early phosphatic or Orsten-type preservation, see Maas et al.
306 (2006)). In this way, we can partition different types of evidence to use in different ways. Parasites
307 might be an especially good example in this context. Many groups of parasites are highly diverse
308 today but most have poor preservation in the past, associated with two distinct fossil datasets –
309 rare exceptionally preserved specimens that can be used to build a morphological matrix versus

310 more common eggs or trace fossils that can only be assigned to higher groupings (De Baets et al.,
311 2021), potentially better suited to the use of clade constraints (Warnock and Engelstädter, 2021).

312 Here, we considered a scenario in which we have an extant clade. For fossil only datasets, typically
313 we use the FBD model with morphological data only and no topological constraints. Previous
314 simulations have shown that trees generated with matrices that are typical of many fossil groups
315 (i.e., 30 characters) will be highly uncertain (Barido-Sottani et al., 2020). Clade constraints could
316 also be useful in this context, provided these can be defined with a high degree of confidence. As an
317 example, a backbone constraint based on the analysis of molecular data could be applied to a tree
318 that consists mostly fossils or is based on morphology, but where the morphology never recovers
319 the established molecular phylogeny. For example, cyclostomes or many mammalian superorders
320 are never recovered as monophyletic on the basis of morphology alone. In other cases, we might
321 want to look at a large clade that contains too many fossil species for all of them to be included
322 in a single analyses, but where previous studies show unequivocal support for certain subclades.
323 In this scenario, clade constraints could help stabilising the topology, without the need to collect
324 additional morphological data while remaining computationally efficient.

325 For some groups (e.g., soft-bodied worms or unicellular organisms), we it might not be possible
326 or practical to collect much more data, due to both incomplete preservation and/or the labour
327 associated with collecting morphological data, or obtain reliable information about broader level
328 taxonomy. In this situation, extended versions of the FBD process can help make the best use of
329 the available data. For instance, the occurrence birth-death process allows fossils associated with
330 morphological character data and fossils associated with age data only to be assigned different
331 sampling rates. Under this model, occurrence data do not need to be constrained to any part of
332 the tree, but can improve overall FBD parameter estimates.

333 Beyond inferring dated trees, Soul and Friedman (2015) showed that trees constructed using higher
334 level taxonomy and dated using time-scaling methods can be successfully used for phylogenetic
335 comparative analyses. Although we emphasise the need to take a very cautious approach when
336 using clade constraints, the FBD model could also be used in this context, with the added benefit
337 that the output better reflects uncertainty in fossil ages, node ages and phylogenetic uncertainty.

338 Irrespective of the approach to handling taxonomic or phylogenetic uncertainty using the FBD
339 process, core taxonomic research, including morphological data collection and species descriptions,
340 remain essential. The issues we observed with the erroneous placement of fossils mirror those

341 identified previously in node dating studies, where node calibrations based on inaccurate fossil
342 assignments have been shown to result in large errors in divergence time estimation (e.g. Phillips
343 et al., 2009). These issues spurred the development of ‘best practices’ for justifying fossil calibra-
344 tions for node dating, taking into account both phylogenetic and fossil age uncertainty (Parham
345 et al., 2012). Authors are recommended to provide an explicit set of statements justifying the
346 assignment of a fossil to a given node, with reference to up-to-date phylogenetic analyses incorpo-
347 rating the relevant taxa or to a set of unequivocal synapomorphies. This rigorous and transparent
348 approach to defining clade constraints is directly applicable to analyses that employ the FBD pro-
349 cess, along with the criteria used to justify fossil ages. How age uncertainty is handled in analyses
350 using the FBD process also has important implications for the accuracy of both divergence times
351 and topology (Barido-Sottani et al., 2019, 2020). Developments of phylogenetic models used in
352 palaeobiology are no substitute for the expertise contributed by fundamental systematics, taxo-
353 nomic and stratigraphic research. Our results reiterate the need for increased and direct support
354 for taxonomy-based projects (Britz et al., 2020; Engel et al., 2021). Improving approaches to
355 phylogenetic dating requires both advanced methodological and empirical perspectives.

356 6 Acknowledgments

357 JBS was supported by funds from the National Science Foundation (USA), grants DBI-1759909 and
358 DEB-1556615 and from the European Union’s Horizon 2020 Research and Innovation Programme
359 under the Marie Skłodowska-Curie grant agreement No. 101022928. KDB was supported by the
360 I.3.4 Action of the Excellence Initiative – Research University Programme at the University of
361 Warsaw (Project: PARADIVE).

362 References

363 Joëlle Barido-Sottani, Gabriel Aguirre-Fernández, Melanie J Hopkins, Tanja Stadler, and Rachel
364 Warnock. Ignoring stratigraphic age uncertainty leads to erroneous estimates of species diver-
365 gence times under the fossilized birth–death process. *Proceedings of the Royal Society B*, 286
366 (1902):20190685, 2019.

367 Joëlle Barido-Sottani, Nina van Tiel, Melanie J Hopkins, David F Wright, Tanja Stadler, and
368 Rachel Warnock. Ignoring fossil age uncertainty leads to inaccurate topology and divergence
369 time estimates in time calibrated tree inference. *Frontiers in Ecology and Evolution*, 8:183, 2020.

370 Roger BJ Benson, Richard Butler, Roger A Close, Erin Saupe, and Daniel L Rabosky. Biodiversity
371 across space and time in the fossil record. *Current Biology*, 31(19):R1225–R1236, 2021.

372 Remco Bouckaert, Timothy G. Vaughan, Joëlle Barido-Sottani, Sebastián Duchêne, Mathieu Four-
373 ment, Alexandra Gavryushkina, Joseph Heled, Graham Jones, Denise Kühnert, Nicola De Maio,
374 Michael Matschiner, Fábio K. Mendes, Nicola F. Müller, Huw A. Ogilvie, Louis Du Plessis, Alex
375 Popinga, Andrew Rambaut, David Rasmussen, Igor Siveroni, Marc A. Suchard, Chieh Hsi Wu,
376 Dong Xie, Chi Zhang, Tanja Stadler, and Alexei J. Drummond. BEAST 2.5: An advanced soft-
377 ware platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, 15(4):e1006650,
378 2019. ISSN 15537358. doi: 10.1371/journal.pcbi.1006650.

379 Ralf Britz, Anna Hundsdörfer, and Uwe Fritz. Funding, training, permits—the three big challenges
380 of taxonomy. *Megataxa*, 1(1):49–52, 2020.

381 Kenneth De Baets, John Warren Huntley, Adiël A. Klompmaker, James D. Schiffbauer, and A. D.
382 Muscente. The fossil record of parasitism: Its extent and taphonomic constraints. In Kenneth
383 De Baets and John Warren Huntley, editors, *The Evolution and Fossil Record of Parasitism:
384 Coevolution and Paleoparasitological Techniques*, pages 1–50. Springer International Publishing,
385 Cham, 2021. ISBN 978-3-030-52233-9. doi: 10.1007/978-3-030-52233-9_1.

386 Michael S Engel, Luis MP Ceríaco, Gimo M Daniel, Pablo M Dellapé, Ivan Löbl, Milen Marinov,
387 Roberto E Reis, Mark T Young, Alain Dubois, Ishan Agarwal, et al. The taxonomic impediment:
388 a shortage of taxonomists, not the lack of technical approaches. *Zoological Journal of the Linnean
389 Society*, 193(2):381–387, 2021.

390 Theo Engeser and Helmut Keupp. Phylogeny of the aptychi-possessing neoammonoidea (aptychophora nov., cephalopoda). *Lethaia*, 35(1):79–96, 2002.

391

392 Alexandra Gavryushkina, David Welch, Tanja Stadler, and Alexei J Drummond. Bayesian inference
393 of sampled ancestor trees for epidemiology and fossil calibration. *PLoS Computational Biology*,
394 10(12):e1003919, 2014.

395 Alexandra Gavryushkina, Tracy A Heath, Daniel T Ksepka, Tanja Stadler, David Welch, and
396 Alexei J Drummond. Bayesian total-evidence dating reveals the recent crown radiation of penguins.
397 *Systematic biology*, 66(1):57–73, 2017.

398 Tracy A Heath, John P Huelsenbeck, and Tanja Stadler. The fossilized birth–death process for co-
399 herent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences*,
400 111(29):E2957–E2966, 2014.

401 Jean-Pierre Hugot, Scott L Gardner, Victor Borba, Priscilla Araujo, Daniela Leles, Átila Augusto
402 Stock Da-Rosa, Juliana Dutra, Luiz Fernando Ferreira, and Adauto Araújo. Discovery of a
403 240 million year old nematode parasite egg in a cynodont coprolite sheds light on the early
404 origin of pinworms in vertebrates. *Parasites & Vectors*, 7(1):486, 2014. ISSN 1756-3305. doi:
405 10.1186/s13071-014-0486-6. URL <https://doi.org/10.1186/s13071-014-0486-6>.

406 Susan Kidwell and Steven Holland. The quality of the fossil record: Implications for evolutionary
407 analyses. *Annual Review of Ecology and Systematics*, 11:561–88, 11 2002. doi: 10.1146/ann-
408 nurev.ecolsys.33.030602.152151.

409 Stilianos Louca and Matthew W Pennell. Extant timetrees are consistent with a myriad of diver-
410 sification histories. *Nature*, 580(7804):502–505, 2020.

411 Stilianos Louca, Angela McLaughlin, Ailene MacPherson, Jeffrey B Joy, and Matthew W Pennell.
412 Fundamental identifiability limits in molecular epidemiology. *Molecular biology and evolution*,
413 38(9):4010–4024, 2021.

414 Andreas Maas, Andreas Braun, Xi-Ping Dong, Philip C.J. Donoghue, Klaus J. Müller, Ewa Olemp-
415 ska, John E. Repetski, David J. Siveter, Martin Stein, and Dieter Waloszek. The ‘orsten’—more
416 than a cambrian konservat-lagerstätte yielding exceptional preservation. *Palaeoworld*, 15(3):
417 266–282, 2006. ISSN 1871-174X. doi: <https://doi.org/10.1016/j.palwor.2006.10.005>. URL

418 <https://www.sciencedirect.com/science/article/pii/S1871174X06000382>. The Fourth
419 International Symposium on the Cambrian System.

420 Felix Marx, Olivier Lambert, and Mark D Uhen. *Cetacean Paleobiology*. Wiley-Blackwell, 05 2016.
421 ISBN 978-1-118-56127-0.

422 Nicolás Mongiardino Koch, Russell J Garwood, and Luke A Parry. Fossils improve phylogenetic
423 analyses of morphological characters. *Proceedings of the Royal Society B*, 288(1950):20210044,
424 2021.

425 Joseph E O'Reilly and Philip CJ Donoghue. The effect of fossil sampling on the estimation of
426 divergence times with the fossilized birth-death process. *Systematic biology*, 69(1):124–138, 2020.

427 James F Parham, Philip CJ Donoghue, Christopher J Bell, Tyler D Calway, Jason J Head, Patri-
428 cia A Holroyd, Jun G Inoue, Randall B Irmis, Walter G Joyce, Daniel T Ksepka, et al. Best
429 practices for justifying fossil calibrations. *Systematic Biology*, 61(2):346–359, 2012.

430 Matthew J Phillips, Thomas H Bennett, and Michael SY Lee. Molecules, morphology, and ecology
431 indicate a recent, amphibious ancestry for echidnas. *Proceedings of the National Academy of
432 Sciences*, 106(40):17089–17094, 2009.

433 Alexander Pohle, Björn Kröger, Rachel Warnock, Andy H King, David H Evans, Martina Aubrech-
434 tová, Marcela Cichowolski, Xiang Fang, and Christian Klug. Early cephalopod evolution clarified
435 through bayesian phylogenetic inference. *BMC biology*, 20(1):1–30, 2022.

436 Nussaibah B Raja, Emma M Dunne, Aviwe Matiwane, Tasnuva Ming Khan, Paulina S Nätscher,
437 Aline M Ghilardi, and Devapriya Chattopadhyay. Colonial history and global economics distort
438 our understanding of deep-time biodiversity. *Nature ecology & evolution*, 6(2):145–154, 2022.

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

442 Astrid Schuster, Sergio Vargas, Ingrid S Knapp, Shirley A Pomponi, Robert J Toonen, Dirk Er-
443 penbeck, and Gert Wörheide. Divergence times in demosponges (Porifera): first insights from
444 new mitogenomes and the inclusion of fossils in a birth-death clock model. *BMC Evolutionary
445 Biology*, 18(1):1–11, 2018.

446 Jiří Šmíd and Krystal A Tolley. Calibrating the tree of vipers under the fossilized birth-death
447 model. *Scientific Reports*, 9(1):1–10, 2019.

448 Andrew B. Smith and Alistair J. McGowan. The ties linking rock and fossil records and
449 why they are important for palaeobiodiversity studies. *Geological Society, London, Spe-*
450 *cial Publications*, 358(1):1–7, 2011. ISSN 0305-8719. doi: 10.1144/SP358.1. URL
451 <https://sp.lyellcollection.org/content/358/1/1>.

452 Laura C Soul and Matt Friedman. Taxonomy and phylogeny can yield comparable results in
453 comparative paleontological analyses. *Systematic Biology*, 64(4):608–620, 2015.

454 Tanja Stadler. Sampling-through-time in birth–death trees. *Journal of Theoretical Biology*, 267(3):
455 396–404, 2010.

456 Daniel B. Thomas, Alan J. D. Tennyson, R. Paul Scofield, Tracy A. Heath,
457 Walker Pett, and Daniel T. Ksepka. Ancient crested penguin constrains tim-
458 ing of recruitment into seabird hotspot. *Proceedings of the Royal Society B: Bi-*
459 *ological Sciences*, 287(1932):20201497, 2020. doi: 10.1098/rspb.2020.1497. URL
460 <https://royalsocietypublishing.org/doi/abs/10.1098/rspb.2020.1497>.

461 Rachel Warnock and Jan Engelstädter. The molecular clock as a tool for understanding host-
462 parasite evolution. In Kenneth De Baets and John Warren Huntley, editors, *The Evolution and*
463 *Fossil Record of Parasitism: Coevolution and Paleoparasitological Techniques*, pages 417–450.
464 Springer International Publishing, Cham, 2021. ISBN 978-3-030-52233-9. doi: 10.1007/978-3-
465 030-52233-9_13.

466 Chi Zhang, Tanja Stadler, Seraina Klopstein, Tracy A Heath, and Fredrik Ronquist. Total-evidence
467 dating under the fossilized birth–death process. *Systematic Biology*, 65(2):228–249, 2016.