

Phylogenies of extant species are consistent with an infinite array of diversification histories

Stilianos Louca^{1,2,*} & Matthew W. Pennell^{3,4}

¹*Department of Biology, University of Oregon, USA*

²*Institute of Ecology and Evolution, University of Oregon, USA*

³*Biodiversity Research Centre, University of British Columbia, Vancouver, Canada*

⁴*Department of Zoology, University of British Columbia, Vancouver, Canada*

*Corresponding author. Contact details at www.loucalab.com

Abstract

Time-calibrated molecular phylogenies of extant species ("extant timetrees") are widely used for estimating the dynamics of speciation and extinction rates and reconstructing macroevolutionary events such as mass extinctions. However, there has been considerable debate surrounding the reliability of these inferences in the absence of fossil data, and to date this critical question remains unresolved. Here we mathematically clarify the precise information that can be extracted from extant timetrees under the generalized birth-death model, which underlies the majority of existing estimation methods. We prove that for a given extant timetree and a candidate diversification scenario, there exists an infinite number of alternative diversification scenarios that are equally likely to have generated a given tree. These "congruent" scenarios cannot possibly be distinguished using extant timetrees alone, even in the presence of infinite data. Importantly, congruent diversification scenarios can exhibit markedly different diversification dynamics, suggesting that many previous studies may have over-interpreted phylogenetic evidence. We show that sets of congruent models can be uniquely described using modified identifiable variables, corresponding to an effective speciation and diversification rate in special idealized scenarios that cannot be distinguished from their congruent alternatives. We advocate a shift from thinking about specific diversification scenarios to considering classes of multiple congruent scenarios, and show how such an approach can provide a deeper and more robust view of the processes that shape diversity through time.

Keywords: speciation; extinction; macroevolution; phylogenetic trees; pulled speciation rate

19 Introduction

20 A central challenge in evolutionary biology is to explain why some taxonomic groups and some time periods
21 have so many species while others have so few; ultimately this means estimating and explaining variation in
22 rates of speciation and extinction (1). Estimating these rates is crucial to investigating fundamental questions
23 such as the role of biotic and abiotic processes in shaping patterns of species richness (2), how Earth's
24 biota recover after mass extinction events (3, 4) and whether there are general dynamics that govern how
25 biodiversity accumulates (5). Measuring such rates has taken on a new urgency as we try to understand
26 how anthropogenically induced extinctions compare to "background" rates (6, 7). In the medical domain,
27 "speciation" and extinction rates are key parameters that provide insights into the historical dynamics and
28 future trajectories of viral epidemics (8, 9).

29 Unfortunately, the vast majority of lineages that have ever lived have not left any trace in the fossil
30 record, and hence we are forced to attempt to reconstruct their diversification dynamics from incomplete
31 data. Indeed, for many groups the fossil record is so incomplete that the primary source of information on
32 past diversification dynamics comes from time-calibrated molecular phylogenies of extant lineages ("extant
33 timetrees"). There is now an abundance of increasingly sophisticated methods for extracting this information,
34 with most state-of-the-art methods fitting variants of a birth-death process (10) to extant timetrees (11–18).
35 These methods have, collectively, been used in thousands of studies and have substantially contributed to
36 our understanding of the drivers of diversity through time. Despite their popularity, there has been a long-
37 lingering doubt about many of these inferences. For one, simulation studies have repeatedly shown that some
38 variables, especially extinction rates, are generally difficult to estimate (19–24). But an even more funda-
39 mental issue, which has particularly drawn the attention of paleobiologists, is that there may not be sufficient
40 information in a molecular phylogeny to fully reconstruct historical changes in diversification rates. For ex-
41 ample, when speciation and extinction rates vary through time — and there is abundant evidence from the
42 fossil record that they do (1) — mass extinction events can erode much of the signal of preceding diversi-
43 fication dynamics (25–27), and may themselves even be confused with stagnating speciation rates (28). To
44 date these critical identifiability issues remain poorly understood, and no general theory exists for describ-
45 ing which diversification scenarios can be distinguished from each other and precisely what information on
46 diversification rates is in principle extractable from extant phylogenies.

47 Here, we present a solution to this problem: We develop a mathematical framework for assessing
48 the identifiability of the general stochastic birth-death process, where speciation ("birth") rates (λ) and ex-
49 tinction ("death") rates (μ) can vary over time, and which underlies the majority of existing methods for
50 reconstructing diversification dynamics from phylogenies (16). By considering the full space of possible
51 diversification scenarios (i.e., with arbitrary λ and μ), rather than special cases (as has been done so far),
52 we reveal a fundamental and surprisingly general property of the birth-death process that has far reaching
53 implications for diversification analyses. Specifically, we show that for any given birth-death model there
54 exists an infinite number of alternative birth-death models that can explain any given extant timetree equally
55 well as the candidate model. This ambiguity persists for arbitrarily large trees and cannot be resolved even
56 with an infinite amount of data using any statistical method. Crucially, these alternative models may appear
57 to be similarly plausible and yet exhibit markedly different features, such as different trends through time
58 in both λ and μ . Using simulated and real timetrees as examples, we demonstrate how failing to recognize
59 this immense ambiguity may seriously mislead our inferences about past diversification dynamics, shedding
60 doubt on conclusions from countless previous studies. We further show that these sets of "congruent" models
61 can be uniquely identified based on suitably defined composite variables: the "pulled speciation rate", cor-
62 responding to the effective λ in the hypothetical absence of extinction and under complete species sampling,
63 or equivalently, the "pulled diversification rate", corresponding to the effective net diversification rate in the

64 hypothetical case where λ is time-independent. Based on either one of these variables, it becomes possible
65 to determine whether different diversification scenarios are at all distinguishable, to explore the full range of
66 plausible scenarios that are consistent with the data, and to make inferences about diversification dynamics
67 without knowing λ and μ themselves.

68 Computing the likelihood of diversification models from lineages-through-time curves

69 One of the most important features of extant timetrees is the lineages-through-time curve (LTT), which counts
70 the number of lineages at each time in the past that are represented by at least one extant descending species
71 in the tree. The LTT provides a simple visual overview of a tree's branching density over time and importantly,
72 contains all the information encoded in the tree regarding speciation and extinction rates (29) (see
73 also Supplement S.1.2). This is because the likelihood of a extant timetree under a given birth-death model
74 depends solely on the tree's LTT, but not on any other properties of the tree that do not affect the LTT.

75 Here we show that an elegant analogous relationship also exists between the likelihood of a tree and
76 the LTT that would be predicted by a given birth-death model: Any given speciation and extinction rates over
77 time, λ and μ , and the probability that an extant species will be included in the tree ρ ("sampling fraction"),
78 can be used to define a "deterministic" diversification process, where the number of lineages through time
79 no longer varies stochastically but according to a set of differential equations (Supplement S.1). The LTT
80 predicted by these differential equations ("deterministic LTT", or dLTT) is a mere theoretical property of the
81 model that resembles the LTT of a tree only if the tree is sufficiently large for stochastic effects to become
82 negligible, and assuming the model is an adequate description of the process that generated the tree. It can be
83 shown, however, that the likelihood of a tree under a given birth-death model can be written purely in terms
84 of the tree's LTT and the model's dLTT (Supplement S.1.2). This means that any two models with the same
85 dLTT (conditioned on the number of extant species sampled, M_o) yield identical likelihoods for the tree. In
86 the following, we shall therefore call any two models "congruent" if they have the same dLTT for any given
87 M_o . Note that two models are either congruent or non-congruent regardless of the particular tree considered,
88 meaning that model congruency is a property of models and not the data (Supplement S.1). Furthermore, the
89 probability distribution of tree sizes generated by a birth-death model, when conditioned on the age of the
90 stem (or crown), is identical among congruent models (Supplement S.1.7). Hence, congruent models have
91 equal probabilities of generating any given timetree and LTT, analogous to how congruent geometric objects
92 exhibit similar geometric properties (discussion in Supplement S.1.8). In the absence of further information
93 or constraints, congruent models cannot possibly be distinguished solely based on extant timetrees, neither
94 through the likelihood nor any other test statistic (such as the γ statistic (30)). Note that whether or not
95 a given phylogenetic data set is sufficient to statistically distinguish between non-congruent models is an
96 entirely different matter.

97 Congruent model sets are infinitely large and infinite-dimensional

98 The above considerations lead to an important question: For any birth-death model (i.e., with given λ and μ
99 as functions of time, and a given sampling fraction ρ), how many alternative congruent models are there and
100 how could one possibly construct them? To answer this question, we first present an alternative method for
101 recognizing congruent models. Given a number of sampled species M_o , a model's dLTT is fully determined
102 by its relative slope, $\lambda_p = -M^{-1}dM/d\tau$ (where M is the dLTT and τ is time before present or "age"). It
103 can be shown that λ_p is related to the model's speciation rate as $\lambda_p = P\lambda$, where $P(\tau)$ is the probability
104 that any lineage that existed at age τ survives to the present and is included in the timetree (Supplement
105 S.1.1). In the absence of extinction ($\mu = 0$) and under complete species sampling ($\rho = 1$), λ_p is identical

106 to λ , however in the presence of extinction λ_p is pulled downwards relative to λ at older ages, while under
 107 incomplete sampling λ_p is pulled downwards relative to λ near the present. We thus henceforth refer to λ_p as
 108 the “pulled speciation rate” of a model. Since a model’s dLTT is fully determined by λ_p and, reciprocally, λ_p
 109 is fully determined by the dLTT, it becomes evident that two models are congruent if and only if they have the
 110 same pulled speciation rate. The latter can also be used to calculate a variable called “pulled diversification
 111 rate” (31), defined as:

$$r_p = \lambda - \mu + \frac{1}{\lambda} \frac{d\lambda}{d\tau}. \quad (1)$$

112 The r_p is equal to the net diversification rate ($r = \lambda - \mu$) whenever λ is constant in time ($d\lambda/d\tau = 0$), but
 113 differs from r when λ varies with time. As shown in Supplement S.1.1, λ_p and r_p are linked through the
 114 following differential equation:

$$\frac{d\lambda_p}{d\tau} = \lambda_p \cdot (r_p - \lambda_p), \quad (2)$$

115 with initial condition $\lambda_p(0) = \rho\lambda_o$ (where $\lambda_o = \lambda(0)$ is the present-day speciation rate). Equation (2)
 116 reveals that r_p is completely determined by λ_p (one can just solve for r_p) and, reciprocally, λ_p is completely
 117 determined by r_p and the product $\rho\lambda_o$ (one can just solve the differential equation for λ_p ; see solution in
 118 Supplement S.1.6). We thus conclude that two birth-death models are congruent if, and only if, they have
 119 the same r_p and the same product $\rho\lambda_o$.

120 We are now ready to assess the breadth of congruent model sets. Consider a birth-death model with
 121 speciation rate $\lambda > 0$, extinction rate $\mu \geq 0$ and sampling fraction $\rho \in (0, 1]$. If we denote $\eta_o = \rho\lambda_o$, then
 122 for any alternative chosen extinction rate function $\mu^* \geq 0$, and any alternative assumed sampling fraction
 123 $\rho^* \in (0, 1]$, there exists a speciation rate function $\lambda^* > 0$ such that the alternative model $(\lambda^*, \mu^*, \rho^*)$ is
 124 congruent to the original model. In other words, regardless of the chosen μ^* and ρ^* , we can find a hypothetical
 125 λ^* that satisfies:

$$\lambda^* - \mu^* + \frac{1}{\lambda^*} \frac{d\lambda^*}{d\tau} = r_p, \quad \rho^* \lambda^*(0) = \eta_o. \quad (3)$$

126 Indeed, to construct such a λ^* one merely needs to solve the following differential equation:

$$\frac{d\lambda^*}{d\tau} = \lambda^* \cdot (r_p - \lambda^* + \mu^*), \quad (4)$$

127 with initial condition $\lambda^*(0) = \eta_o/\rho^*$ (solution given in Supplement S.1.4). The above observation implies
 128 that, starting from virtually any birth-death model, we can generate an infinite number of alternative congruent
 129 models simply by modifying the extinction rate μ and/or the assumed sampling fraction ρ . Alternatively,
 130 congruent models can be generated by assuming various ratios of extinction over speciation rates, $\varepsilon = \mu/\lambda$
 131 (formula in Supplement S.1.5). This set of congruent models — henceforth “congruence class” — is thus
 132 infinitely large. The congruence class can have an arbitrary number of dimensions (depending on restrictions
 133 imposed a priori on λ^* and μ^*), since μ^* could depend on an arbitrarily high number of free parameters.

134 For illustration, consider the simulations in Figure 1, showing four markedly distinct and yet congruent
 135 birth-death models. The first scenario exhibits a constant λ and a temporary spike in μ (mass extinction
 136 event), the second scenario instead exhibits a constant μ and a temporary drop in λ (temporary stasis of
 137 speciations) around the same time, the third scenario exhibits a mass extinction event at a completely different
 138 time and a fluctuating λ , while the fourth scenario exhibits an exponentially decaying μ and a fluctuating λ .
 139 These congruent scenarios were obtained simply by assuming alternative extinction rates, and a myriad of
 140 other congruent scenarios also exist. Similar situations can be readily found in the literature. Figure 2, for
 141 example, shows a birth-death model with exponentially varying speciation and extinction rates, $\lambda = \lambda_o e^{\alpha\tau}$
 142 and $\mu = \mu_o e^{\beta\tau}$ (where λ_o , μ_o , α and β are fitted parameters), as commonly considered in other studies
 143 (14, 32), fitted to a massive timetree of 79,874 extant seed plant species via maximum likelihood. Simply

144 by modifying the coefficient β and choosing λ according to Eq. (4), one can obtain an infinite number of
145 congruent and similarly complex scenarios, with even opposite trends over time (Fig. 2B).

146 Such ambiguities were described previously in special cases. For example, Kubo and Iwasa (33)
147 recognized that a variable λ and constant μ can be exchanged for a constant λ and a variable μ to produce
148 the same dLTT; similarly Stadler (28) and Crisp *et al.* (34) observed that simulations of mass extinctions
149 produced similar LTTs as simulations of temporarily stagnating diversification processes. Other previous
150 work on constant-rate birth-death models revealed that alternative combinations of time-independent λ , μ
151 and ρ can yield the same likelihood for a tree (13, 35–37). By generalizing these analyses to the time-variable
152 case, we have revealed that in fact vast expanses of model space are practically indistinguishable.

153 Model congruency compromises existing reconstruction methods

154 Since the likelihood of an extant timetree can be expressed purely in terms of r_p and the product $\rho\lambda_o$ (Sup-
155 plement S.1.6), or alternatively purely in terms of λ_p (Supplement S.1.3), extant timetrees only provide
156 information about the congruence class of a generating process and not the actual speciation and extinction
157 rates. This identifiability issue can be interpreted as follows: Since all information available on past diversi-
158 fication dynamics (representable by birth-death models, to be precise) is encoded in a single curve, namely
159 the LTT, one should not expect to be able to “extract” from it two independent curves (λ and μ) without
160 additional information, as this would essentially double the amount of information at hand. In order to esti-
161 mate λ and μ , previous phylogenetic studies have been imposing strong and largely arbitrary constraints. For
162 example, many studies assume that λ or μ are constant or vary exponentially through time (38). However
163 these constraints are rarely justified biologically — there is little reason to expect that λ or μ are constant
164 or exponentially varying through time (1). Crucially, considering alternative functional forms may result
165 in drastically different inferences with alternative trends in λ and μ over time, even if the functional forms
166 used are in principle flexible enough to capture the true historical λ and μ (examples in Supplement S.5).
167 We thus conclude that inferences about λ and μ may not even hold qualitatively, if imposed constraints and
168 assumptions are not well-justified. Previous studies have not recognized the breadth of this issue because
169 they typically only consider a limited set of candidate models at a time, both when analyzing real datasets
170 as well as when assessing parameter identifiability via simulations; as a result, previous studies have been
171 (un)lucky enough to not compare two models in the same congruence class (see Supplements S.2 and S.3
172 for reasoning).

173 It is important to realize that congruent scenarios can have markedly different macroevolutionary im-
174 plications. For example, Steeman *et al.* (39) reconstructed past speciation rates of Cetaceans (whales, dol-
175 phins, and porpoises) based on an extant timetree and using maximum-likelihood (assuming $\mu = 0$). Stee-
176 man *et al.* (39) found a temporary increase of λ during the late Miocene-early Pliocene (Fig. 3), suggesting
177 a potential link between Cetacean radiations and concurrent paleoceanographic changes. However, alterna-
178 tively to assuming $\mu = 0$, one could assume that μ was close to λ , consistent with common observations
179 from the fossil record (1). For example, by setting $\mu = 0.9 \cdot \lambda$ one obtains a congruent scenario in which λ
180 no longer peaks during the late Miocene-early Pliocene but instead exhibits a gradual slowdown throughout
181 most of Cetacean evolution (Fig. 3B). Both scenarios are similarly complex and both could have generated
182 the timetree at equal probabilities. Even the common methodological decision to estimating net diversifi-
183 cation ($r = \lambda - \mu$) rather than λ and μ separately (20, 40), is no longer meaningful in light of our results;
184 the shape of r is not conserved across a congruence class. Likewise, the models in a congruence class will
185 not share “average” (however defined) rates; hence absolute rate estimates (11), which have been used to
186 estimate broad macroevolutionary patterns (5) and background rates of extinction (7), are also unlikely to
187 be accurately reconstructed. These issues are likely also present in more complex models with additional

188 free parameters, for example where some clades exhibit distinct diversification regimes (38, 41). In such
189 situations, the tree can always be decomposed into a set of sub-trees with distinct LTT curves, each of which
190 is subject to the same identifiability issues as described here. Our findings thus shed serious doubts over a lot
191 of previous work on past diversification dynamics, including some of the conclusions from our own recent
192 work (5, 31).

193 **Ways forward**

194 Our findings for birth-death models of diversification are closely analogous to classic results from coalescent
195 theory in population genetics (42, 43), where an infinite number of models can give rise to the same drift pro-
196 cess as the idealized Wright-Fisher model. This realization was profoundly important for the field; it focused
197 researchers' attention on the dynamics of the effective population size N_e , a composite but identifiable pa-
198 rameter that is the same for all models with the same Wright-Fisher drift process, rather than on the actual (but
199 non-identifiable) historical demography, sex ratios etc. of the population. Consequently, the field has adapted
200 and developed a plethora of tools to infer changes in N_e through time and across the genome. Here we have
201 discovered an analogous generality that could be similarly transformative to the field of macroevolution. We
202 have mathematically confirmed previous suspicions, voiced notably by paleobiologists, that some historical
203 diversification scenarios may not be distinguishable using reconstructed phylogenies alone (1, 26, 29, 33),
204 and have shown that such congruent scenarios can be defined in terms of the identifiable λ_p (or, equivalently,
205 in terms of the identifiable r_p and $\rho\lambda_o$), suggesting a whole new way to think about diversification analyses.
206 In fact, each congruence class contains exactly one model with $\mu = 0$ and $\rho = 1$, which is also the only model
207 where $\lambda = \lambda_p$; hence the pulled speciation rate can be interpreted as the effective speciation rate generating
208 the congruence class's dLTT in the absence of extinctions and under complete species sampling. Similarly,
209 each congruence class contains an infinite number of models with time-independent λ , and for these models
210 $r_p = r$; hence the pulled diversification rate can be interpreted as the effective net diversification rate if λ
211 was time-independent.

212 Rather than attempting to estimate λ and μ , one can instead estimate λ_p , r_p and $\rho\lambda_o$ (and λ_o if ρ is
213 known) either using likelihood methods (Supplement S.4) or based on the slope and curvature of a tree's LTT
214 (31). Our previous work has shown that r_p itself can yield valuable insight into diversification dynamics and
215 can be useful for testing alternative hypotheses (31). Using simulations, we found that sudden rate transitions,
216 for example due to mass extinction events, usually lead to detectable fluctuations in r_p ; therefore, a relatively
217 constant r_p over time would be indicative of constant — or only slowly changing — speciation and extinction
218 rates. One can also obtain other useful composite variables from λ_p , r_p and $\rho\lambda_o$. For example, in cases where
219 ρ is known one can obtain the “pulled extinction rate”, defined as $\mu_p := \lambda_o - r_p$ (31). Note that $\mu_p(\tau)$ is
220 equal to the extinction rate $\mu(\tau)$ if λ has been constant from τ to the present, but differs from μ in most other
221 cases. The present-day μ_p is related to the present-day μ as follows:

$$\mu_p(0) = \mu(0) - \frac{1}{\lambda_o} \left. \frac{d\lambda}{d\tau} \right|_{\tau=0}. \quad (5)$$

222 Hence if the present-day speciation rate changes only slowly, the present-day μ_p will resemble the present-
223 day μ . Further, since $\mu(0)$ is non-negative, we can obtain the following one-sided bound for the rate at which
224 λ changes at present:

$$\frac{1}{\lambda_o} \left. \frac{d\lambda}{d\tau} \right|_{\tau=0} \geq -\mu_p(0). \quad (6)$$

225 Of course if a macroevolutionary question is only concerned with recent speciation events (2) then one can
226 test hypotheses using λ_o , which can be readily identified if ρ is known. Our intuition on how λ_p , r_p and μ_p

227 behave is based on a limited set of simulations (31), and more work is needed to clarify how to best make
228 use of the information encoded in these variables and what the limits of these inferences are. Critically,
229 we currently lack a sampling theory for r_p , that is, we can estimate r_p accurately given sufficient data (31)
230 (examples in Supplement S.4), but we do not have a closed form solution for the standard error of the estimate
231 when there is stochastic sampling error.

232 An alternative approach to only considering composite parameters is to constrain model space in a
233 biologically justifiable way. Through decades of work in both paleobiology and speciation biology, we have
234 a rich understanding of how the process of speciation works in a variety of taxa, how long it takes to complete,
235 and how variable we expect rates to be through time and space (44, 45). It may be possible to leverage this
236 information to prefer some models over others (26); Bayesian approaches (e.g., model comparison with
237 Bayes Factors; 46) seem promising in this regard. Additionally, trait data combined with phylogenetic data,
238 modeled using trait-dependent diversification models with time-variable rates (e.g., time-dependent Binary
239 State Speciation and Extinction models; 47), could perhaps also resolve the identifiability problem we have
240 uncovered. On the one hand, the additional data (tip character states) could help reduce ambiguities; on the
241 other hand the number of degrees of freedom is also larger, since each character state can exhibit distinct λ
242 and μ over time. This is an important avenue to explore — ultimately we need to find additional data that
243 could be used to resolve the identification issues discussed here, since for the majority of taxa an informative
244 fossil record is lacking.

245 Conclusions

246 We have shown that for virtually any candidate birth-death process, suspected of having generated some extant
247 timetree, there exists an infinite number of alternative and markedly different birth-death processes that
248 could have generated the timetree with the same likelihood. Without further information or prior constraints
249 on plausible diversification scenarios, extant timetrees alone cannot be used to reliably infer speciation rates
250 (except at present-day), extinction rates or net diversification rates, raising serious doubts over a multitude of
251 previous estimates of past diversification dynamics. Our work could thus explain why frequently diversification
252 dynamics observed in the fossil record are in great disagreement with phylogenetics-based inferences
253 (1, 14, 16, 26, 39).

254 On a more positive note, we resolved a long-standing debate and precisely clarified what information
255 can indeed be extracted from extant timetrees alone — namely λ_p , r_p , the product $\rho\lambda_o$ (and λ_o if ρ is known),
256 and any other variables that can be expressed in terms of λ_p , r_p and $\rho\lambda_o$. These identifiable variables not
257 only tell us when two models are in principle distinguishable but they can themselves yield valuable insight
258 into past diversification dynamics. We see these as analogous to the concept of effective population size
259 in population genetics — we cannot uniquely determine the exact sequence of events that led to current
260 diversity, but by blending out some of the historical details we could potentially gain powerful and robust
261 insights into general macroevolutionary phenomena.

262 **Code availability**

263 Computational methods used for this article, including functions for simulating birth-death models, for con-
264 structing models within a given congruence class, for calculating the likelihood of a congruence class, and
265 for directly fitting congruence classes (either in terms of λ_p or in terms of r_p and $\rho\lambda_o$) to extant timetrees
266 (Supplement S.4), are implemented in the R package `castor` (48).

267 **Acknowledgments**

268 S.L. was supported by a startup grant by the University of Oregon, USA. M.W.P. was supported by an NSERC
269 Discovery Grant. We thank L. Harmon, S. Otto, A. MacPherson, D. Schluter, T.J. Davies, L.F. Henao Diaz,
270 K. Kaur, J. Uyeda, D. Caetano, J. Rolland, L. Parfrey, and A. Mooers for insightful comments on this work.

271 **Author contributions**

272 Both authors conceived the project and contributed to the writing of the manuscript. S.L. performed the
273 mathematical calculations and computational analyses.

274 **Competing financial interests**

275 The authors declare that they have no competing interests.

276 **Materials & Correspondence**

277 Correspondence and requests for materials should be addressed to S.L.

278 **References**

- 279 [1] Marshall, C. R. Five palaeobiological laws needed to understand the evolution of the living biota.
280 *Nature Ecology & Evolution* **1**, 165 (2017).
- 281 [2] Schluter, D. & Pennell, M. W. Speciation gradients and the distribution of biodiversity. *Nature* **546**, 48
282 (2017).
- 283 [3] Jablonski, D. Mass extinctions and macroevolution. *Paleobiology* **31**, 192–210 (2005).
- 284 [4] Meredith, R. W. *et al.* Impacts of the Cretaceous terrestrial revolution and KPg extinction on mammal
285 diversification. *Science* **334**, 521–524 (2011).
- 286 [5] Henao Diaz, L. F., Harmon, L. J., Sugawara, M. T. C., Miller, E. T. & Pennell, M. W. Macroevolutionary
287 diversification rates show time dependency. *Proceedings of the National Academy of Sciences* **116**, 7403
288 (2019).
- 289 [6] Pimm, S. L. *et al.* The biodiversity of species and their rates of extinction, distribution, and protection.
290 *Science* **344**, 1246752 (2014).

- 291 [7] De Vos, J. M., Joppa, L. N., Gittleman, J. L., Stephens, P. R. & Pimm, S. L. Estimating the normal
292 background rate of species extinction. *Conservation Biology* **29**, 452–462 (2015).
- 293 [8] Stadler, T., Kühnert, D., Bonhoeffer, S. & Drummond, A. J. Birth–death skyline plot reveals temporal
294 changes of epidemic spread in hiv and hepatitis c virus (hcv). *Proceedings of the National Academy of
295 Sciences* **110**, 228–233 (2013).
- 296 [9] Arora, N. *et al.* Origin of modern syphilis and emergence of a pandemic treponema pallidum cluster.
297 *Nature microbiology* **2**, 16245 (2017).
- 298 [10] Kendall, D. G. *et al.* On the generalized “birth-and-death” process. *The annals of mathematical statistics*
299 **19**, 1–15 (1948).
- 300 [11] Magallon, S. & Sanderson, M. J. Absolute diversification rates in angiosperm clades. *Evolution* **55**,
301 1762–1780 (2001).
- 302 [12] Rabosky, D. L. Likelihood methods for detecting temporal shifts in diversification rates. *Evolution* **60**,
303 1152–1164 (2006).
- 304 [13] Morlon, H., Potts, M. D. & Plotkin, J. B. Inferring the dynamics of diversification: A coalescent
305 approach. *PLOS Biology* **8**, e1000493 (2010).
- 306 [14] Morlon, H., Parsons, T. L. & Plotkin, J. B. Reconciling molecular phylogenies with the fossil record.
307 *Proceedings of the National Academy of Sciences* **108**, 16327–16332 (2011).
- 308 [15] Stadler, T. Recovering speciation and extinction dynamics based on phylogenies. *Journal of evolutionary
309 biology* **26**, 1203–1219 (2013).
- 310 [16] Morlon, H. Phylogenetic approaches for studying diversification. *Ecology Letters* **17**, 508–525 (2014).
- 311 [17] Rabosky, D. L. *et al.* Bamm tools: an r package for the analysis of evolutionary dynamics on phyloge-
312 netic trees. *Methods in Ecology and Evolution* **5**, 701–707 (2014).
- 313 [18] Maliet, O., Hartig, F. & Morlon, H. A model with many small shifts for estimating species-specific
314 diversification rates. *Nature Ecology & Evolution* (2019).
- 315 [19] FitzJohn, R. G., Maddison, W. P. & Otto, S. P. Estimating trait-dependent speciation and extinction
316 rates from incompletely resolved phylogenies. *Systematic Biology* **58**, 595–611 (2009).
- 317 [20] Rabosky, D. L. Extinction rates should not be estimated from molecular phylogenies. *Evolution* **64**,
318 1816–1824 (2010).
- 319 [21] Davis, M. P., Midford, P. E. & Maddison, W. Exploring power and parameter estimation of the bisse
320 method for analyzing species diversification. *BMC Evolutionary Biology* **13**, 38 (2013).
- 321 [22] Beaulieu, J. M. & O’Meara, B. C. Extinction can be estimated from moderately sized molecular phy-
322 logenies. *Evolution* **69**, 1036–1043 (2015).
- 323 [23] Rabosky, D. L. Challenges in the estimation of extinction from molecular phylogenies: a response to
324 beaulieu and o’meara. *Evolution* **70**, 218–228 (2016).
- 325 [24] Moore, B. R., Höhna, S., May, M. R., Rannala, B. & Huelsenbeck, J. P. Critically evaluating the theory
326 and performance of bayesian analysis of macroevolutionary mixtures. *Proceedings of the National
327 Academy of Sciences* **113**, 9569–9574 (2016).

- 328 [25] Quental, T. B. & Marshall, C. R. Extinction during evolutionary radiations: reconciling the fossil record
329 with molecular phylogenies. *Evolution* **63**, 3158–3167 (2009).
- 330 [26] Quental, T. B. & Marshall, C. R. Diversity dynamics: molecular phylogenies need the fossil record.
331 *Trends in Ecology & Evolution* **25**, 434–441 (2010).
- 332 [27] Liow, L. H., Quental, T. B. & Marshall, C. R. When can decreasing diversification rates be detected
333 with molecular phylogenies and the fossil record? *Systematic Biology* **59**, 646–659 (2010).
- 334 [28] Stadler, T. Simulating trees with a fixed number of extant species. *Systematic Biology* **60**, 676–684
335 (2011).
- 336 [29] Nee, S., May, R. M. & Harvey, P. H. The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **344**, 305–311 (1994).
- 338 [30] Pybus, O. G. & Harvey, P. H. Testing macro-evolutionary models using incomplete molecular phylogenies.
339 *Proceedings of the Royal Society of London. Series B: Biological Sciences* **267**, 2267–2272
340 (2000).
- 341 [31] Louca, S. *et al.* Bacterial diversification through geological time. *Nature Ecology & Evolution* **2**,
342 1458–1467 (2018).
- 343 [32] Hallinan, N. The generalized time variable reconstructed birth-death process. *Journal of theoretical
344 biology* **300**, 265–276 (2012).
- 345 [33] Kubo, T. & Iwasa, Y. Inferring the rates of branching and extinction from molecular phylogenies.
346 *Evolution* **49**, 694–704 (1995).
- 347 [34] Crisp, M. D. & Cook, L. G. Explosive radiation or cryptic mass extinction? interpreting signatures in
348 molecular phylogenies. *Evolution* **63**, 2257–2265 (2009).
- 349 [35] Stadler, T. On incomplete sampling under birth–death models and connections to the sampling-based
350 coalescent. *Journal of Theoretical Biology* **261**, 58–66 (2009).
- 351 [36] Stadler, T. How can we improve accuracy of macroevolutionary rate estimates? *Systematic Biology* **62**,
352 321–329 (2013).
- 353 [37] Stadler, T. & Steel, M. Swapping birth and death: Symmetries and transformations in phylodynamic
354 models. *bioRxiv* 494583 (2019).
- 355 [38] Rabosky, D. L. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLOS ONE* **9**, 1–15 (2014).
- 357 [39] Steeman, M. E. *et al.* Radiation of extant cetaceans driven by restructuring of the oceans. *Systematic
358 Biology* **58**, 573–585 (2009).
- 359 [40] Maddison, W. P., Midford, P. E., Otto, S. P. & Oakley, T. Estimating a binary character's effect on
360 speciation and extinction. *Systematic Biology* **56**, 701–710 (2007).
- 361 [41] Alfaro, M. E. *et al.* Nine exceptional radiations plus high turnover explain species diversity in jawed
362 vertebrates. *Proceedings of the National Academy of Sciences* **106**, 13410–13414 (2009).
- 363 [42] Griffiths, R. C. & Tavaré, S. Sampling theory for neutral alleles in a varying environment. *Philosophical
364 Transactions of the Royal Society of London. Series B: Biological Sciences* **344**, 403–410 (1994).

- 365 [43] Donnelly, P. & Tavaré, S. Coalescents and genealogical structure under neutrality. *Annual Review of Genetics* **29**, 401–421 (1995).
- 366
- 367 [44] Sepkoski, J. J. Rates of speciation in the fossil record. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **353**, 315–326 (1998).
- 368
- 369 [45] Alroy, J. Dynamics of origination and extinction in the marine fossil record. *Proceedings of the National Academy of Sciences* **105**, 11536–11542 (2008).
- 370
- 371 [46] Kass, R. E. & Raftery, A. E. Bayes factors. *Journal of the American Statistical Association* **90**, 773–795 (1995).
- 372
- 373 [47] Cantalapiedra, J. L. *et al.* Dietary innovations spurred the diversification of ruminants during the Caenozoic. *Proceedings of the Royal Society B: Biological Sciences* **281**, 20132746 (2014).
- 374
- 375 [48] Louca, S. & Doebeli, M. Efficient comparative phylogenetics on large trees. *Bioinformatics* **34**, 1053–1055 (2017).
- 376
- 377 [49] Smith, S. A. & Brown, J. W. Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany* **105**, 302–314 (2018).
- 378

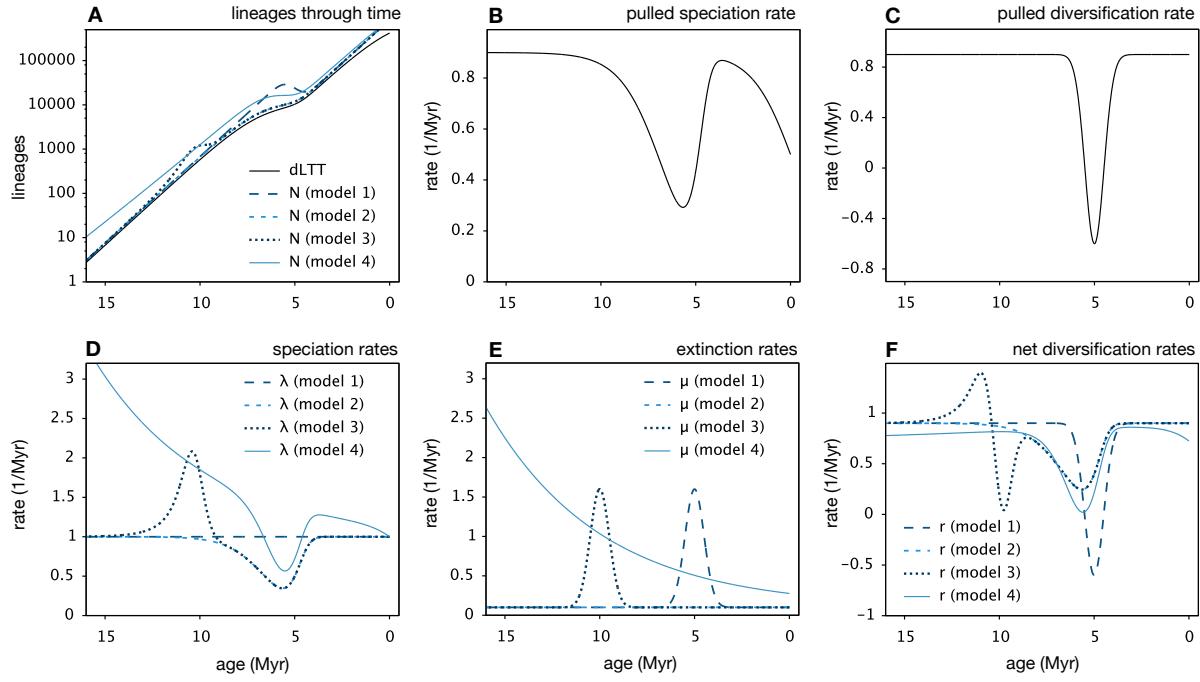


Figure 1: Illustration of congruent birth-death processes. Example of four hypothetical, congruent yet markedly different birth-death models. The first model exhibits a constant speciation rate and a sudden mass extinction event about 5 Myr before present; the second model exhibits a constant extinction rate and a temporary stagnation of the speciation rate about 5–6 Myr before present; the third model exhibits a mass extinction event about 10 Myr before present and a variable speciation rate; the fourth model exhibits an exponentially decreasing extinction rate and a variable speciation rate. In all models the sampling fraction is $\rho = 0.5$. All models exhibit the same deterministic LTT (dLTT), the same pulled speciation rate (λ_p) and the same pulled diversification rate (r_p), and would yield the same likelihood for any given extant timetree. (A) dLTT and deterministic total diversities (N) predicted by the models, plotted over age (time before present). (B) Pulled extinction rate λ_p of the models. (C) Pulled diversification rate r_p of the models. (D) Speciation rates (λ) of the models. (E) Extinction rates (μ) of the models. (F) Net diversification rates ($r = \lambda - \mu$) of the models. For additional examples see Supplemental Fig. S1.

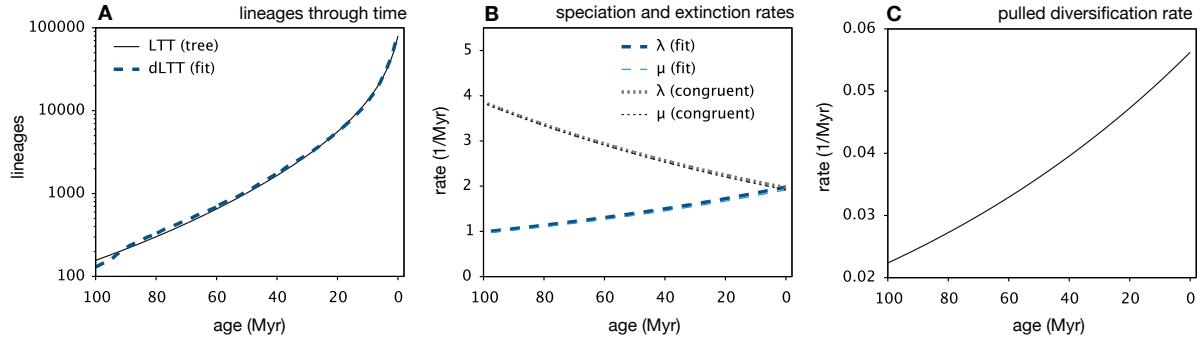


Figure 2: Maximum-likelihood-fitted model versus congruent model. Birth-death model with exponentially varying λ and μ , fitted to a reconstruct timetree of 79,874 seed plant species (49) over the past 100 Myr, compared to a congruent model obtained by simply modifying the exponential coefficient of μ . (A) LTT of the tree, compared to the dLTT predicted by the two models. (B) Speciation rates (λ) and extinction rates (μ) of the two models. (C) Pulled diversification rate of the two models.

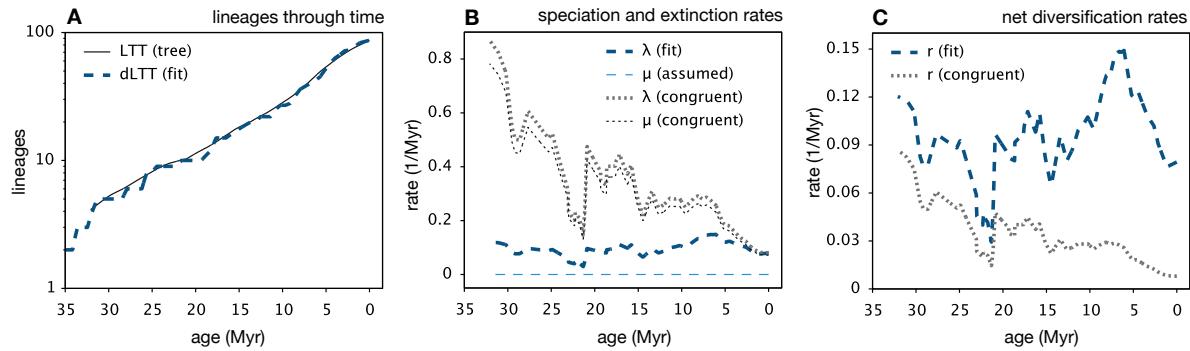


Figure 3: Model congruency invalidates support for macroevolutionary conclusions. Time-dependent birth-death model fitted to a nearly-complete Cetacean timetree by Steeman *et al.* (39) under the assumption of zero extinction rates ($\mu = 0$), compared to a congruent model where the extinction rate is close to the speciation rate ($\mu = 0.9\lambda$). (A) LTT of the tree, compared to the dLTT predicted by the two models. (B) Speciation rates (λ) and extinction rates (μ) of the two models. (C) Net diversification rates ($r = \lambda - \mu$) of the two models. The original fitted rates were used by Steeman *et al.* (39) to link Cetacean diversification dynamics to past paleoceanographic changes.

Phylogenies of extant species are consistent with an infinite array of diversification histories

- Supplemental Information -

Stilianos Louca^{1,2} & Matthew W. Pennell^{3,4}

¹*Department of Biology, University of Oregon, USA*

²*Institute of Ecology and Evolution, University of Oregon, USA*

³*Biodiversity Research Centre, University of British Columbia, Vancouver, Canada*

⁴*Department of Zoology, University of British Columbia, Vancouver, Canada*

S.1 Mathematical derivations

2 In the following, we provide mathematical derivations for various claims made in the main article. Some
3 parts can be found in previous literature (Kendall, 1948, Nee *et al.*, 1994, Morlon *et al.*, 2011, Louca *et al.*,
4 2018), but are included here for completeness.

S.1.1 General considerations

6 We begin with listing some basic mathematical properties of deterministic birth-death models that will be of
7 use at various later stages. Our starting point is some time-dependent speciation rate λ , some time-dependent
8 extinction rate μ and some sampling fraction ρ (fraction of extant species included in the tree). Let τ denote
9 time before present (“age”). The deterministic total diversity, i.e. the number of species predicted at any point
10 in time according to the deterministic model, and conditional upon M_o extant species having been sampled
11 at present-day, is obtained by solving the following differential equation backward in time:

$$\frac{dN}{d\tau} = N \cdot (\mu - \lambda), \quad (1)$$

12 with initial condition $N(0) = M_o/\rho$, i.e.:

$$N(\tau) := \frac{M_o}{\rho} \exp \left[\int_0^\tau du [\mu(u) - \lambda(u)] \right]. \quad (2)$$

13 The deterministic LTT (dLTT), i.e. the number of lineages represented in the final extant timetree at any time
14 point according to the deterministic model, is given by:

$$M(\tau) = N(\tau) \cdot (1 - E(\tau)), \quad (3)$$

15 where $E(\tau)$ is the probability that a lineage extant at age τ will be missing from the timetree (either due to
16 extinction or not having been sampled). As explained by Morlon *et al.* (2011), the extinction probability E

17 satisfies the differential equation:

$$\frac{dE}{d\tau} = \mu - E \cdot (\lambda + \mu) + E^2 \lambda, \quad E(0) = 1 - \rho. \quad (4)$$

18 Taking the derivative of both sides in Eq. (3), and then using Eq. (4) to replace $dE/d\tau$ as well as Eq. (1) to
19 replace $dN/d\tau$ quickly leads to the differential equation:

$$\frac{dM}{d\tau} = M \lambda \cdot (E - 1), \quad (5)$$

20 with initial condition $M(0) = M_o$. The solution to this differential equation is:

$$M(\tau) = M_o \cdot \exp \left[\int_0^\tau du \lambda(u) \cdot [E(u) - 1] \right]. \quad (6)$$

21 Observe that E is a property purely of the model, and does not depend on the particular tree considered;
22 together with Eq. (6), this shows that any two models either have equal dLTMs for any given tree or they have
23 non-equal dLTMs for any given tree. Hence, model congruency is a property of two models, regardless of
24 tree.

25 Defining the relative slope of the dLTT:

$$\lambda_p := -\frac{1}{M} \frac{dM}{d\tau} \quad (7)$$

26 allows us to write Eq. (5) as follows:

$$\lambda_p = \lambda \cdot (1 - E). \quad (8)$$

27 As becomes clear in Eq. (8), in the absence of extinction and if $\rho = 1$, the relative slope λ_p becomes equal
28 to the speciation rate λ ; in the presence of extinction λ_p is artificially pulled upwards near the present, an
29 effect known as the “pull of the present”. Reciprocally, under incomplete sampling λ is artificially pulled
30 downwards near the present. We shall therefore henceforth call λ_p the “pulled speciation rate”.

31 Taking the derivative on both sides of Eq. (8) and using Eq. (4) to replace $dE/d\tau$ leads to:

$$\frac{d\lambda_p}{d\tau} = \lambda_p \cdot \left[\frac{1}{\lambda} \frac{d\lambda}{d\tau} - \mu + \lambda E \right] = \lambda_p \cdot \left[\frac{1}{\lambda} \frac{d\lambda}{d\tau} + \lambda - \mu - \lambda \cdot (1 - E) \right] = \lambda_p \cdot (r_p - \lambda_p), \quad (9)$$

32 where we defined the “pulled diversification rate”:

$$r_p := \lambda - \mu + \frac{1}{\lambda} \frac{d\lambda}{d\tau}. \quad (10)$$

33 Rearranging terms in Eq. (9) yields:

$$r_p = \lambda_p + \frac{1}{\lambda_p} \frac{d\lambda_p}{d\tau}, \quad (11)$$

34 which shows that r_p can be directly calculated from the dLTT.

35 S.1.2 The likelihood in terms of the LTT and dLTT

36 In the following we show how the likelihood of an extant timetree under a birth-death model can be expressed
 37 purely in terms of the tree's LTT and the model's dLTT. We begin with the case where the stem age is known
 38 and the likelihood is conditioned on the survival of the stem lineage; the alternative case where only the
 39 crown age is known is very similar and will be discussed at the end.

40 Our starting point is the likelihood formula described by [Morlon et al. \(2011\)](#):

$$L = \frac{\rho^{n+1} \Psi(\tau_1, \tau_o)}{1 - E(\tau_o)} \prod_{i=1}^n \lambda(\tau_i) \Psi(s_{i,1}, \tau_i) \Psi(s_{i,2}, \tau_i), \quad (12)$$

41 where n is the number of branching points (internal nodes), τ_o is the age of the stem, $\tau_1 > \tau_2 > \dots > \tau_n$ are
 42 the ages (time before present) of the branching points, $s_{i,1}, s_{i,2}$ are the ages at which the daughter lineages
 43 originating at age τ_i themselves branch (or end at a tip), ρ is the tree's sampling fraction (fraction of present-
 44 day extant species included in the tree), $E(\tau)$ is the probability that a single lineage that existed at age τ
 45 would survive to the present and be represented in the tree ([Morlon et al., 2011](#), Eq. 2 therein), Ψ is defined
 46 as:

$$\Psi(s, \tau) := e^{R(\tau) - R(s)} \left[\frac{1 + \rho \int_0^s du \lambda(u) e^{R(u)}}{1 + \rho \int_0^\tau du \lambda(u) e^{R(u)}} \right]^2, \quad (13)$$

47 and $R(\tau)$ is defined as:

$$R(\tau) := \int_0^\tau du [\lambda(u) - \mu(u)]. \quad (14)$$

48 It is straightforward to confirm that Ψ satisfies the property $\Psi(s, \tau) = \Psi(0, \tau)/\Psi(0, s)$; using this property
 49 in Eq. (12) leads to:

$$L = \frac{\rho^{n+1}}{1 - E(\tau_o)} \cdot \frac{\Psi(0, \tau_o)}{\Psi(0, \tau_1)} \prod_{i=1}^n \frac{\lambda(\tau_i) \Psi(0, \tau_i)^2}{\Psi(0, s_{i,1}) \Psi(0, s_{i,2})}. \quad (15)$$

50 Since each internal node except for the root is the child of another internal node, the numerator and denominator
 51 in Eq. (15) partly cancel out, eventually leading to:

$$L = \frac{\rho^{n+1} \Psi(0, \tau_o)}{1 - E(\tau_o)} \prod_{i=1}^n \lambda(\tau_i) \Psi(0, \tau_i). \quad (16)$$

52 Since the set of branching times τ_i is completely determined by the LTT (branching events correspond to
 53 jumps in the LTT), we conclude that the likelihood of a tree is entirely determined by its LTT.

54 Further, from Eq. (11) we know that the model's dLTT satisfies:

$$\lambda - \mu + \frac{d \ln \lambda}{d \tau} = \frac{d \ln \lambda_p}{d \tau} - \frac{d \ln M}{d \tau}. \quad (17)$$

55 Integrating both sides of Eq. (17) yields:

$$R(\tau) + \ln \frac{\lambda(\tau)}{\lambda_o} = \int_0^\tau du \left[\lambda - \mu + \frac{d \ln \lambda}{du} \right] = \int_0^\tau du \left[\frac{d \ln \lambda_p}{du} - \frac{d \ln M}{du} \right] = \ln \frac{\lambda_p(\tau)}{\lambda_p(0)} - \ln \frac{M(\tau)}{M_o}, \quad (18)$$

56 where M_o is the number of extant species included in the timetree. Hence:

$$e^{R(\tau)} \frac{\lambda(\tau)}{\lambda_o} = \frac{\lambda_p(\tau) M_o}{\lambda_p(0) M(\tau)}. \quad (19)$$

57 Using Eq. (19) in Eq. (13) yields:

$$\Psi(0, \tau) = \frac{\lambda_o}{\lambda(\tau)} \cdot \frac{\lambda_p(\tau) M_o}{\lambda_p(0) M(\tau)} \cdot \left[1 + \frac{\rho \lambda_o}{\lambda_p(0)} M_o \int_0^\tau du \frac{\lambda_p(u)}{M(u)} \right]^{-2} \quad (20)$$

58 Recall that $\rho \lambda_o = \lambda_p(0)$ according to standard theory (Sanmartín and Meseguer, 2016, Louca *et al.*, 2018),
59 so that Eq. (20) can be written as:

$$\Psi(0, \tau) = \frac{1}{\rho \lambda(\tau)} \cdot \frac{\lambda_p(\tau) M_o}{M(\tau)} \cdot \left[1 + M_o \int_0^\tau du \frac{\lambda_p(u)}{M(u)} \right]^{-2}. \quad (21)$$

60 Note that:

$$\frac{\lambda_p}{M} = \frac{d}{d\tau} \frac{1}{M}. \quad (22)$$

61 Hence, Eq. (21) can be further simplified to:

$$\begin{aligned} \Psi(0, \tau) &= \frac{1}{\rho \lambda(\tau)} \cdot \frac{\lambda_p(\tau) M_o}{M(\tau)} \cdot \left[1 + M_o \int_0^\tau du \frac{d}{du} \left(\frac{1}{M} \right) \right]^{-2} \\ &= \frac{1}{\rho \lambda(\tau)} \cdot \frac{\lambda_p(\tau) M_o}{M(\tau)} \cdot \left[1 + M_o \left(\frac{1}{M(\tau)} - \frac{1}{M_o} \right) \right]^{-2} \\ &= \frac{\lambda_p(\tau) M(\tau)}{\rho \lambda(\tau) M_o}. \end{aligned} \quad (23)$$

62 Inserting Eq. (23) into the likelihood formula (16) yields:

$$L = \frac{1}{[1 - E(\tau_o)] \lambda(\tau_o)} \cdot \frac{\lambda_p(\tau_o) M(\tau_o)}{M_o^{n+1}} \prod_{i=1}^n \lambda_p(\tau_i) M(\tau_i). \quad (24)$$

63 Recall that $(1 - E) \lambda = \lambda_p$ according to Eq. (8), which when inserted into (24) yields:

$$L = \frac{M(\tau_o)}{M_o^{n+1}} \prod_{i=1}^n \lambda_p(\tau_i) M(\tau_i). \quad (25)$$

64 Since $\lambda_p M = -dM/d\tau$, Eq. (25) becomes:

$$L = \frac{M(\tau_o)}{M_o^{n+1}} \prod_{i=1}^n \left[-\frac{dM}{d\tau} \Big|_{\tau_i} \right]. \quad (26)$$

65 A corollary of Eq. (26) is that for any given extant timetree, any two models with the same dLTT will also
66 yield the same likelihood.

67 Note that the likelihood in Eq. (12) or equivalently Eq. (26) is conditioned upon the survival of the stem
68 lineage, assuming that the stem age is known. If the stem age is unknown the likelihood should be conditioned

69 upon the splitting at the root and the survival of the root's two daughter-lineages, as follows:

$$L_r = \frac{\rho^{n+1}}{\lambda(\tau_1) \cdot [1 - E(\tau_1)]^2} \prod_{i=1}^n \lambda(\tau_i) \Psi(s_{i,1}, \tau_i) \Psi(s_{i,2}, \tau_i). \quad (27)$$

70 Note that Eq. (27) can be obtained from (12) by setting the stem age equal to the crown age ($\tau_o = \tau_1$)
 71 and adjusting the conditioning. Following a similar procedure as above, it is easy to show that L_r can be
 72 expressed in the following alternative forms:

$$L_r = \frac{\rho^{n+1} \Psi(0, \tau_1)}{\lambda(\tau_1) \cdot [1 - E(\tau_1)]^2} \prod_{i=1}^n \lambda(\tau_i) \Psi(0, \tau_i), \quad (28)$$

73 and

$$L_r = \frac{M^2(\tau_1)}{M_o^{n+1}} \prod_{i=2}^n \left[-\frac{dM}{d\tau} \Big|_{\tau_i} \right]. \quad (29)$$

74 \square

75 S.1.3 The likelihood in terms of λ_p

76 In the following we show how the likelihood of an extant timetree under a birth-death model can be expressed
 77 purely in terms of the tree's LTT and the model's pulled speciation rate λ_p .

78 We begin with the case where the stem age is known and the likelihood is conditioned on the survival of the
 79 stem lineage. Our starting point is the likelihood formula in Eq. (26):

$$L = \frac{M(\tau_o)}{M_o^{n+1}} \prod_{i=1}^n \left[-\frac{dM}{d\tau} \Big|_{\tau_i} \right], \quad (30)$$

80 where M is the dLTT and $M_o := M(0)$. From Eq. (7) it is easy to obtain the following relationship between
 81 M and λ_p :

$$M(\tau) = M_o e^{-\Lambda_p(\tau)}, \quad (31)$$

82 where we defined:

$$\Lambda_p(\tau) := \int_0^\tau ds \lambda_p(s). \quad (32)$$

83 Inserting Eq. (31) into Eq. (30) yields:

$$L = e^{-\Lambda_p(\tau_o)} \prod_{i=1}^n \underbrace{\frac{-1}{M(\tau_i)} \frac{dM}{d\tau} \Big|_{\tau_i}}_{\lambda_p(\tau_i)} \cdot e^{-\Lambda_p(\tau_i)}, \quad (33)$$

84 and hence:

$$L = e^{-\Lambda_p(\tau_o)} \prod_{i=1}^n \lambda_p(\tau_i) \cdot e^{-\Lambda_p(\tau_i)}. \quad (34)$$

85 If only the crown age is known and the likelihood is conditioned on the splitting at the root and the survival
 86 of the root's two daughter-lineages (likelihood formula in Eq. (29)), we instead obtain the expression:

$$L_r = \frac{e^{-\Lambda_p(\tau_1)}}{\lambda_p(\tau_1)} \prod_{i=1}^n \lambda_p(\tau_i) \cdot e^{-\Lambda_p(\tau_i)}. \quad (35)$$

87 **S.1.4 Calculating λ from r_p and μ**

88 In the following we provide the general solution to the differential equation (4) in the main article:

$$\frac{d\lambda}{d\tau} = \lambda \cdot (r_p + \mu^* - \lambda), \quad (36)$$

89 with initial condition:

$$\lambda(0) = \eta_o/\rho > 0. \quad (37)$$

90 We assume that r_p and μ^* are sufficiently “well-behaved”, specifically that they are integrable over any finite
 91 interval. Observe that Eq. (36) is an example of a Bernoulli-type differential equation, as it can be written
 92 in the standard form:

$$\frac{d\lambda}{d\tau} = P(\tau)\lambda(\tau) + Q(\tau)\lambda^\alpha(\tau), \quad (38)$$

93 where $\alpha = 2$, $P = r_p + \mu^*$ and $Q = -1$. Using the standard technique for solving Bernoulli differential
 94 equations (i.e., substituting $u = \lambda^{1-\alpha}$ to obtain a linear differential equation for u), it is straightforward to
 95 obtain the solution:

$$\lambda(\tau) = \frac{\eta_o e^{\Lambda(\tau)}}{\rho + \eta_o \int_0^\tau ds e^{\Lambda(s)}}, \quad (39)$$

96 where we defined:

$$\Lambda(\tau) := \int_0^\tau ds [r_p(s) + \mu^*(s)]. \quad (40)$$

97 Note that the solution in Eq. (39) is strictly positive and continuous, and hence λ is indeed a valid speciation
 98 rate.

99 **Special cases:**

100 • In the special case where r_p and μ^* are time-independent and $r_p + \mu^* \neq 0$, the solution in Eq. (39)
 101 takes the form:

$$\lambda(\tau) = \frac{P}{(P\rho/\eta_o - 1)e^{-P\tau} + 1}, \quad (41)$$

102 where $P = r_p + \mu^*$.

- 103 • If and only if $\mu^*(\tau) = \eta_o/\rho - r_p(\tau)$, the solution in Eq. (39) is time-independent:

$$\lambda(\tau) = \frac{\eta_o}{\rho}. \quad (42)$$

104 Hence, for a fixed ρ , a congruence class can include at most one model with constant speciation rate;
 105 it includes exactly one model with constant speciation rate if and only if $\eta_o/\rho \geq \max_\tau r_p(\tau)$.

106 **S.1.5 Calculating λ from r_p and ε**

107 In the following we show how the speciation rate λ can be calculated from the pulled diversification rate r_p ,
 108 the present-day speciation rate λ_o and the ratio of extinction over speciation rate, $\varepsilon := \mu/\lambda$. Specifically, we
 109 provide the general solution to the following differential equation:

$$\frac{d\lambda}{d\tau} = \lambda \cdot [r_p + (\varepsilon - 1)\lambda]. \quad (43)$$

110 We assume that r_p and ε are sufficiently “well-behaved”, specifically that they are integrable over any finite
 111 interval. Observe that Eq. (43) is an example of a Bernoulli-type differential equation, as it can be written
 112 in the standard form:

$$\frac{d\lambda}{d\tau} = P(\tau)\lambda(\tau) + Q(\tau)\lambda^\alpha(\tau), \quad (44)$$

113 where $\alpha = 2$, $P = r_p$ and $Q = \varepsilon - 1$. Using the standard technique for solving Bernoulli differential
 114 equations (i.e., substituting $u = \lambda^{1-\alpha}$ to obtain a linear differential equation for u), it is straightforward to
 115 obtain the solution:

$$\boxed{\lambda(\tau) = \frac{\lambda_o e^{R_p(\tau)}}{1 + (1 - \varepsilon) \cdot \lambda_o \int_0^\tau ds e^{R_p(s)}}}, \quad (45)$$

116 where we defined:

$$R_p(\tau) := \int_0^\tau ds r_p(s). \quad (46)$$

117 In the special case where r_p is time-independent and non-zero, the solution in Eq. (45) simplifies to:

$$\lambda(\tau) = \frac{\lambda_o e^{r_p \tau}}{1 + (1 - \varepsilon) \cdot \frac{\lambda_o}{r_p} (e^{r_p s} - 1)}. \quad (47)$$

118 **S.1.6 The likelihood in terms of the r_p**

119 In the following we show how the likelihood of a tree under a birth-death model can be expressed solely in
 120 terms of the model’s pulled diversification rate r_p and the product $\rho\lambda_o$. We first consider the case where the
 121 stem age is known and the likelihood is conditioned on the survival of the stem lineage (Morlon *et al.*, 2011);
 122 the alternative case where only the crown age is known and the likelihood is conditioned upon the survival
 123 of the root’s two daughter lineages (Eq. 28) can be treated similarly and is briefly mentioned at the end.

¹²⁴ Our starting point is the likelihood formula in Eq. (16), Supplement S.1.2. Define:

$$R_p(\tau) := \int_0^\tau du r_p(u). \quad (48)$$

¹²⁵ Then from the definition of r_p (Eq. 1 in the main article) we have:

$$R_p(\tau) = \int_0^\tau du [\lambda(u) - \mu(u)] + \int_0^\tau du \frac{d \ln \lambda}{du} = R(\tau) + \ln \frac{\lambda(\tau)}{\lambda_o}. \quad (49)$$

¹²⁶ Exponentiating (49) and rearranging yields:

$$e^{R(\tau)} = e^{R_p(\tau)} \frac{\lambda_o}{\lambda(\tau)}. \quad (50)$$

¹²⁷ Inserting Eq. (50) into the definition of Ψ in Eq. (13) yields:

$$\Psi(0, \tau) = e^{R_p(\tau)} \frac{\lambda_o}{\lambda(\tau)} \left[1 + \rho \lambda_o \int_0^\tau du e^{R_p(u)} \right]^{-2}. \quad (51)$$

¹²⁸ Inserting Eq. (51) into the likelihood formula (16) yields:

$$L = \frac{(\rho \lambda_o)^{n+1} e^{R_p(\tau_o)}}{[1 - E(\tau_o)] \lambda(\tau_o)} \left[1 + \rho \lambda_o \int_0^{\tau_o} du e^{R_p(u)} \right]^{-2} \prod_{i=1}^n e^{R_p(\tau)} \left[1 + \rho \lambda_o \int_0^\tau du e^{R_p(u)} \right]^{-2}. \quad (52)$$

¹²⁹ Recall that $(1 - E)\lambda = \lambda_p$ according to Eq. (8), which when inserted into Eq. (52) yields:

$$L = \frac{(\rho \lambda_o)^{n+1} e^{R_p(\tau_o)}}{\lambda_p(\tau_o)} \left[1 + \rho \lambda_o \int_0^{\tau_o} du e^{R_p(u)} \right]^{-2} \prod_{i=1}^n e^{R_p(\tau)} \left[1 + \rho \lambda_o \int_0^\tau du e^{R_p(u)} \right]^{-2}. \quad (53)$$

¹³⁰ From Eq. (??) we know that λ_p satisfies the initial value problem (Bernoulli differential equation):

$$\frac{d\lambda_p}{d\tau} = \lambda_p \cdot (r_p - \lambda_p), \quad \lambda_p(0) = \rho \lambda_o, \quad (54)$$

¹³¹ It is straightforward to verify that the solution to Eq. (54) is given by:

$$\lambda_p(\tau) = \frac{\rho \lambda_o e^{R_p(\tau)}}{1 + \rho \lambda_o \int_0^\tau e^{R_p(u)} du}. \quad (55)$$

¹³² Inserting the solution (55) into Eq. (53) yields the following expression for the likelihood:

$$L = \left[1 + \rho \lambda_o \int_0^{\tau_o} du e^{R_p(u)} \right]^{-1} (\rho \lambda_o)^n \prod_{i=1}^n e^{R_p(\tau)} \left[1 + \rho \lambda_o \int_0^\tau du e^{R_p(u)} \right]^{-2}. \quad (56)$$

¹³³ In the alternative case where only the crown age is known, and the likelihood is conditioned on the splitting at the root and the survival of the root's two daughter lineages, we obtain the following expression for the likelihood:

$$L_r = e^{-R_p(\tau_o)} (\rho \lambda_o)^{n-1} \prod_{i=1}^n e^{R_p(\tau)} \left[1 + \rho \lambda_o \int_0^\tau du e^{R_p(u)} \right]^{-2}. \quad (57)$$

136 S.1.7 Congruent models have the same probability distribution of generated tree sizes

137 In the following, we show that the distribution of extant timetree sizes generated by a birth-death model,
 138 either conditional upon the age and survival of the stem, or conditional upon the age of the root and the
 139 survival of its two daughter lineages, is the same for all models in a congruence class.

140 Consider a birth-death process with parameters (λ, μ, ρ) , starting from a single lineage at some time before
 141 present τ_o and ultimately resulting in a timetree at age 0, comprising only extant species that are included at
 142 some probability ρ . The probability that the timetree will comprise n tips can be expressed using formulas
 143 first derived by [Kendall et al. \(1948\)](#):

$$\begin{aligned} P(n) &= (1 - E(\tau_o)) \cdot (1 - H) \cdot H^{n-1}, \quad n \geq 1 \\ P(0) &= E(\tau_o), \end{aligned} \tag{58}$$

144 where $E(\tau_o)$ is the probability that a lineage existing at age τ_o will be missing from the timetree (as defined
 145 previously), H is defined as:

$$H := \frac{\rho \int_0^{\tau_o} ds e^{R(s)} \lambda(s)}{1 + \rho \int_0^{\tau_o} ds e^{R(s)} \lambda(s)}, \tag{59}$$

146 and R was previously defined in Eq. (14). Note that the formula in Eq. (58) can be readily obtained using
 147 equations 8, 10b and 11 in [\(Kendall et al., 1948\)](#), after setting the time variable therein equal to τ_o (i.e.
 148 $t = \tau_o$), switching from time to age ($\tau = \tau_o - t$), and adding the term $-\delta(\tau) \ln \rho$ to the extinction rate
 149 (where δ is the Dirac distribution, peaking at age 0) to account for incomplete species sampling. As shown
 150 previously in Eq. (50), we have

$$e^{R(\tau)} = e^{R_p(\tau)} \frac{\lambda_o}{\lambda(\tau)}, \tag{60}$$

151 where R_p is defined as:

$$R_p(\tau) := \int_0^\tau du r_p(u), \tag{61}$$

152 and r_p is the pulled diversification rate. Inserting Eq. (60) into Eq. (59) allows us to write H as follows:

$$H = \frac{\rho \lambda_o \int_0^{\tau_o} ds e^{R_p(s)}}{1 + \rho \lambda_o \int_0^{\tau_o} ds e^{R_p(s)}}. \tag{62}$$

153 Since $\rho \lambda_o$, r_p and R_p are the same for all models in a congruence class, H is also constant across the
 154 congruence class.

155 The probability of obtaining a tree of size $n \geq 1$ conditional upon the age of the stem lineage (τ_o) and its
 156 survival to the present, denoted $P_{\text{stem}}(n)$, is given by the ratio $P(n)/(1 - E(\tau_o))$, i.e.:

$$P_{\text{stem}}(n) = (1 - H) \cdot H^{n-1}. \tag{63}$$

157 Since H is constant across a congruence class, the same also holds for $P_{\text{stem}}(n)$ for any n . The probability

158 of obtaining a tree of size $n \geq 1$ conditional upon the splitting of the root at age τ_o and the survival of its
 159 two daughter lineages, denoted $P_{\text{root}}(n)$, can be derived in a similar way, as follows. The probability that
 160 the two daughter lineages survive, conditional upon the split at age τ_o , is given by the product:

$$P(\text{daughter lineages survive} \mid \text{split at } \tau_o) = (1 - E(\tau_o))^2. \quad (64)$$

161 The probability that the two daughter lineages survive and the timetree has size $n \geq 1$, conditional upon the
 162 split at age τ_o , is given by the following sum of probabilities:

$$\begin{aligned} & P(\text{daughter lineages survive and tree has size } n \mid \text{split at } \tau_o) \\ &= \sum_{k=1}^{n-1} P(k)P(n-k) \\ &= (1 - E(\tau_o))^2 \sum_{k=1}^{n-1} (1 - H) \cdot H^{k-1} \cdot (1 - H) \cdot H^{n-k-1} \\ &= (1 - E(\tau_o))^2 (1 - H)^2 \sum_{k=1}^{n-1} H^{n-2} \\ &= (n - 1) \cdot (1 - E(\tau_o))^2 (1 - H)^2 H^{n-2}. \end{aligned} \quad (65)$$

163 Dividing Eq. (65) by Eq. (64) yields the desired probability:

$$P_{\text{root}}(n) = (n - 1) \cdot (1 - H)^2 H^{n-2}. \quad (66)$$

164 Since H is constant across the congruence class, the same also holds for $P_{\text{root}}(n)$.

165 \square

166 S.1.8 The geometric nature of congruence classes

167 In the following, we provide a geometric interpretation of model congruence classes, by pointing out an
 168 analogy to the concept of object congruency in geometry. A basic background in abstract algebra is assumed.

169 In geometry, two objects are called congruent if they exhibit similar geometric properties, such as identical
 170 angles between corresponding lines and identical distances between corresponding points. More precisely,
 171 two geometric objects (sets of points in Euclidean space \mathbb{R}^n) are called congruent if one set can be trans-
 172 formed into the other set by means of an isometry, i.e. a mapping that preserves distances between pairs of
 173 points (via translations, rotations, and/or reflections). Object congruency is a type of equivalence relation,
 174 and hence the set of models congruent to some focal object is an equivalence class. The set of all isometries
 175 is itself a group (known as “Euclidean group”) that acts on the set of geometric objects, and congruence
 176 classes of objects correspond to “orbits” under the action of isometries (Climenhaga and Katok, 2017).

177 By analogy, two birth-death models are called “congruent” if they exhibit similar statistical properties in
 178 terms of their generated extant timetrees and LTTs (see main text and Supplement S.1). In fact, congruence
 179 classes can be interpreted as the orbits of a group of mappings acting on model space that preserve dLTTs (just
 180 as isometries preserve distances in Euclidean space). For technical reasons, we shall henceforth only consider
 181 the space of birth-death models (denoted \mathcal{B}) with strictly positive λ, μ and ρ and continuously differentiable λ
 182 and μ defined over some age interval $[0, \tau_o] \subseteq \mathbb{R}$. Let $\mathcal{C}_+^1[0, \tau_o]$ denote the set of all continuously differentiable
 183 real-valued strictly positive functions defined on the interval $[0, \tau_o]$. For any $S_o \in (0, \infty)$ and any $f \in$

¹⁸⁴ $\mathcal{C}_+^1[0, \tau_o]$, define $S[S_o, f] \in \mathcal{C}_+^1[0, \tau_o]$ as the solution to the following initial value problem:

$$\frac{dS[S_o, f]}{d\tau} = S[S_o, f](\tau) \cdot [f(\tau) - S[S_o, f](\tau)], \quad S[S_o, f](0) = S_o. \quad (67)$$

¹⁸⁵ It is straightforward to verify that the solution to the above problem is given by:

$$S[S_o, f](\tau) = \frac{S_o e^{F(\tau)}}{1 + S_o \int_0^\tau ds e^{F(s)}}, \quad (68)$$

¹⁸⁶ where we denoted:

$$F(\tau) := \int_0^\tau ds f(s). \quad (69)$$

¹⁸⁷ For any arbitrary $\alpha \in (0, \infty)$ and $\beta \in \mathcal{C}_+^1[0, \tau_o]$, let $g_{\alpha, \beta} : \mathcal{B} \rightarrow \mathcal{B}$ be a transformation of birth-death models
¹⁸⁸ defined as follows:

$$g_{\alpha, \beta}(\lambda, \mu, \rho) := \left(S \left[\lambda/\alpha, \lambda - \mu + \frac{1}{\lambda} \frac{d\lambda}{d\tau} + \beta\mu \right], \beta\mu, \alpha\rho \right). \quad (70)$$

¹⁸⁹ Note that $g_{\alpha, \beta}$ is dLTT-preserving, that is, it maps models to models within the same congruence class.
¹⁹⁰ Indeed, the variable

$$\lambda^* := S \left[\lambda/\alpha, \lambda - \mu + \frac{1}{\lambda} \frac{d\lambda}{d\tau} + \beta\mu \right] \quad (71)$$

¹⁹¹ is exactly the speciation rate of a model with extinction rate $\mu^* := \beta\mu \in \mathcal{C}_+^1[0, \tau_o]$ and sampling fraction
¹⁹² $\rho^* := \alpha\rho \in (0, \infty)$, congruent to the original model (λ, μ, ρ) . The set of all such transformations,

$$G := \left\{ g_{\alpha, \beta} : \alpha \in (0, \infty), \beta \in \mathcal{C}_+^1[0, \tau_o] \right\}, \quad (72)$$

¹⁹³ constitutes a group with group operation:

$$g_{\alpha, \beta} \circ g_{\tilde{\alpha}, \tilde{\beta}} := g_{\alpha\tilde{\alpha}, \beta\tilde{\beta}} \quad (73)$$

¹⁹⁴ and identity element $g_{1,1}$. The group G acts on the set of birth-death models, while preserving dLTTs.
¹⁹⁵ Abstractly, each mapping $g \in G$ corresponds to an “isometric” transformation in model space that preserves
¹⁹⁶ the statistics of generated extant timetrees and dLTTs, in analogy to how rotations, translations or reflections
¹⁹⁷ preserve distances in Euclidean space.

¹⁹⁸ Note that not all dLTT-preserving mappings defined on \mathcal{B} are members of G . It turns out, however, that G is
¹⁹⁹ large enough to completely generate congruence classes in \mathcal{B} . In other words, for any model $(\lambda, \mu, \rho) \in \mathcal{B}$,
²⁰⁰ the orbit:

$$G(\lambda, \mu, \rho) := \{g(\lambda, \mu, \rho) : g \in G\} \quad (74)$$

²⁰¹ is exactly the congruence class of the model; indeed, for any congruent model $(\lambda^*, \mu^*, \rho^*) \in \mathcal{B}$ one can find
²⁰² a transformation $g_{\alpha, \beta} \in G$ such that $(\lambda^*, \mu^*, \rho^*) = g_{\alpha, \beta}(\lambda, \mu, \rho)$, by choosing $\alpha := \rho^*/\rho$ and $\beta := \mu^*/\mu$.

203 S.2 Why previous studies failed to detect model congruencies

204 In practice, reconstructions of λ and μ over time are typically performed by selecting among a limited set
205 of allowed models, i.e., considering specific functional forms described by a finite number of parameters
206 ([Rabosky, 2006b,a](#), [Rabosky and Lovette, 2008](#), [Silvestro et al., 2011](#), [Stadler, 2011](#), [Morlon et al., 2011](#)).
207 In these situations it is generally unlikely that the allowed model set intersects a given congruence class
208 more than once (see Supplement S.3 for mathematical justification). For example, when considering only
209 constant-rate birth-death models and assuming that ρ is fixed (as is usually the case; [Morlon, 2014](#)), each
210 congruence class reduces to a single combination of λ and μ . Likelihood functions defined over a limited
211 allowed model set thus generally don't exhibit ridges associated with congruence classes, and may even
212 exhibit a unique global maximum in the space of considered parameters, leaving the impression that λ and μ
213 have been estimated close to their true values. Our findings suggest that this impression is almost certainly
214 false. Instead, obtained estimates for λ and μ are almost always going to be a random outcome that depends
215 on the particular choice of allowed models, such as the functional forms considered for λ and μ , and will
216 be as close as possible to the congruence class of the truth rather than close to the truth itself. Unless one
217 has reasons to prefer specific functional forms for λ and μ (e.g., based on a mechanistic macroevolutionary
218 model; [McPeek, 2008](#)), fitted λ and μ are unlikely to resemble the true rates even if in principle the functional
219 forms considered are flexible enough to resemble the true λ and μ (see Supplement S.5 for examples using
220 simulations and real data).

221 By analogy, studies that test whether diversification dynamics are influenced by some environmental or ge-
222 ological variable X (e.g., temperature), either by testing for correlations between X and the estimated λ or
223 μ ([Mayhew et al., 2008](#), [Esselstyn et al., 2009](#)) or by fitting models in which λ or μ are explicit functions
224 of X ([Egan and Crandall, 2008](#), [Cantalapiedra et al., 2014](#), [Condamine et al., 2013](#)), will generally lead to
225 unreliable conclusions. Indeed, specifying λ or μ as functions of X (e.g., assuming $\mu = \alpha X + \beta$ and fitting
226 the coefficients α and β) is essentially equivalent to choosing particular functional forms for λ or μ . Inciden-
227 tally, evaluations of estimation methods based on simulations of the same limited model set as considered by
228 the very estimation method evaluated ([Rabosky, 2006b](#), [Rabosky and Lovette, 2008](#), [Silvestro et al., 2011](#),
229 [Stadler, 2011](#), [Morlon et al., 2011](#)), for example simulating trees with linearly varying λ and μ and then fitting
230 models with linearly varying λ and μ to check if linear coefficients are accurately estimated, can generate
231 the false impression that λ and μ can in general be reliably identified. Indeed, any given simulated model is
232 typically going to be the sole representative of its congruence class that's also in the method's allowed model
233 set.

234 S.3 Typical model sets do not exhibit congruence ridges

235 In the following we explain why it is unlikely in practice that a limited set of allowed models (e.g., considered
236 for maximum-likelihood estimation) will intersect any given congruence class more than once, and that it
237 is especially unlikely that multiple intersections of a congruence class form a sub-manifold in parameter
238 space (i.e., a “congruence ridge”). Consider a set of allowed models, parameterized through n independent
239 parameters $q_1, \dots, q_n \in \mathbb{R}$, i.e. such that the speciation and extinction rates of a model are given as functions
240 of age (τ) and the chosen parameters ($\mathbf{q} \in \mathbb{R}^n$):

$$\lambda = \lambda(\tau, \mathbf{q}), \quad \mu = \mu(\tau, \mathbf{q}). \quad (75)$$

²⁴¹ For simplicity, assume that the sampling fraction ρ is given (identifiability issues associated with uncertainties
²⁴² in ρ are already well known; [Stadler, 2009](#), [Morlon et al., 2010](#), [Stadler, 2013](#), [Stadler and Steel, 2019](#)).

²⁴³ Now consider some particular choice of parameters, \mathbf{q} , with corresponding PDR:

$$r_p(\tau, \mathbf{q}) = \lambda(\tau, \mathbf{q}) - \mu(\tau, \mathbf{q}) + \frac{1}{\lambda(\tau, \mathbf{q})} \frac{\partial \lambda(\tau, \mathbf{q})}{\partial \tau}, \quad (76)$$

²⁴⁴ and present-day speciation rate $\lambda(0, \mathbf{q})$. For any other choice of parameters $\mathbf{h} \in \mathbb{R}^n$, the corresponding model
²⁴⁵ would be in the same congruence class as the first model if and only if $\lambda(0, \mathbf{h}) = \lambda(0, \mathbf{q})$ and $r_p(\tau, \mathbf{h}) =$
²⁴⁶ $r_p(\tau, \mathbf{q})$ for all ages $\tau \geq 0$, in other words $\lambda(\cdot, \mathbf{h})$ must be a solution to the initial value problem:

$$\frac{\partial \lambda(\tau, \mathbf{h})}{\partial \tau} = \lambda(\tau, \mathbf{h}) \cdot [r_p(\tau, \mathbf{q}) - \lambda(\tau, \mathbf{h}) + \mu(\tau, \mathbf{h})], \quad \lambda(0, \mathbf{h}) = \lambda(0, \mathbf{q}). \quad (77)$$

²⁴⁷ Unless the functional forms of λ and μ have been specifically designed for this purpose, it is generally unlikely
²⁴⁸ that Eq. (77) will be satisfied for some $\mathbf{h} \neq \mathbf{q}$.

²⁴⁹ A stronger argument for the low probability of congruence ridges can be made as follows. Suppose that \mathbf{q}
²⁵⁰ was part of a congruence ridge, i.e. a sub-manifold in parameter space belonging to the same congruence
²⁵¹ class. Then there must exist a curve in parameter space, i.e. a one-parameter function $\mathbf{h} : [-\varepsilon, \varepsilon] \rightarrow \mathbb{R}^n$,
²⁵² passing through \mathbf{q} (e.g., say $\mathbf{h}(0) = \mathbf{q}$), such that:

$$r_p(\tau, \mathbf{h}(s)) = r_p(\tau, \mathbf{q}), \quad (78)$$

²⁵³ and such that:

$$\lambda(0, \mathbf{h}(s)) = \lambda(0, \mathbf{q}), \quad (79)$$

²⁵⁴ for all $s \in [-\varepsilon, \varepsilon]$ and all $\tau \geq 0$. Taking the derivative of Eq. (78) with respect to s at 0 yields:

$$\sum_{i=1}^n \frac{\partial r_p}{\partial q_i} \Big|_{(\tau, \mathbf{q})} \cdot \frac{dh_i}{ds} \Big|_{s=0} = 0. \quad (80)$$

²⁵⁵ Denote $\mathbf{H} := \frac{d\mathbf{h}}{ds}|_{s=0}$ and $\mathbf{R}(\tau) := \frac{\partial r_p}{\partial \mathbf{q}}|_{(\tau, \mathbf{q})}$. Then the condition in Eq. (80) can be written in vector notation:

$$\mathbf{R}(\tau)^T \cdot \mathbf{H} = 0. \quad (81)$$

²⁵⁶ Note that \mathbf{H} can be interpreted as the “velocity vector” along the ridge curve \mathbf{h} at the point \mathbf{q} , and hence condition
²⁵⁷ (81) means that the ridge must move perpendicular to the direction of steepest descent of r_p . Observe
²⁵⁸ that condition (81) must be satisfied for all ages $\tau \geq 0$. Hence, for any arbitrary choice $\tau_1, \tau_2, \dots, \tau_m \geq 0$, we
²⁵⁹ obtain the following m linear equations that must be satisfied by \mathbf{H} :

$$\begin{aligned} \mathbf{R}(\tau_1)^T \cdot \mathbf{H} &= 0. \\ &\vdots \\ \mathbf{R}(\tau_m)^T \cdot \mathbf{H} &= 0. \end{aligned} \quad (82)$$

²⁶⁰ Unless the functional forms of λ and μ are specifically designed for this purpose, the system in Eq. (82)
²⁶¹ will almost certainly be over-determined if m is chosen sufficiently high ($m \gg n$). Hence, in practice, for a
²⁶² chosen set of allowed models and a given point \mathbf{q} in parameter space, a congruence ridge will almost never

263 exist at that point.

264

□

265 S.4 Fitting congruence classes instead of models

266 The discussion in the main article revealed that speciation and extinction rates constitute partly interchangeable
267 (and thus partly redundant) parameters that cannot be completely resolved from extant timetrees alone,
268 no matter how large the dataset. Extant timetrees do, however, contain the proper information to estimate
269 the pulled diversification rate r_p and η_o (recall that $\eta_o = \rho\lambda_o$), and may thus be used to at least identify the
270 congruence class from which a tree was likely generated. In fact, for sufficiently large trees r_p and η_o can be
271 directly calculated from the slope and curvature of the tree's LTT ([Louca et al., 2018](#)). Since each congruence
272 class corresponds to a unique r_p and η_o , the r_p and η_o can be used to parameterize the space of congruence
273 classes; on this space the likelihood function no longer exhibits the highly problematic ridges seen in the
274 original model space. We thus suggest describing birth-death models in terms of r_p and η_o , rather than λ and
275 μ , when fitting models to timetrees. Since the likelihood function can be expressed directly in terms of r_p and
276 η_o (Supplement S.1.6), such a parameterization is suitable for maximum-likelihood or Bayesian estimation
277 methods. Reciprocally, since every given r_p and η_o correspond to a unique and non-empty congruence class
278 (as shown in the main article), any r_p and η_o estimated from an extant timetree will represent at least one
279 biologically meaningful scenario. It is thus possible to directly fit congruence classes, rather than individual
280 models, via maximum-likelihood. A similar reasoning can also be applied to the pulled speciation rate λ_p ,
281 which provides an alternative representation of congruence classes.

282 To demonstrate this approach, we created software for fitting r_p and η_o to extant timetrees via maximum
283 likelihood. The code is integrated into the R package `castor` ([Louca and Doebeli, 2017](#)) as function
284 `fit_hbd_pdr_on_grid`. The function accepts as input an extant timetree, and an arbitrary number of dis-
285 crete ages at which to estimate r_p , assuming r_p varies linearly or polynomially between those ages. The
286 function also accepts optional lower and upper bounds for the fitted r_p and/or η_o . The code then maximizes
287 the likelihood of the tree, given by Eq. (53) in the Supplement, by iteratively refining the r_p values on the age
288 grid and/or η_o . Optionally, one can limit the evaluation of the likelihood function to a smaller “truncated”
289 age interval than covered by the tree, i.e. some age interval $[0, \tau^*]$, where τ^* may be smaller than the root age.
290 This may be useful for avoid estimation errors towards older ages due to a small number of lineages in the
291 tree. The likelihood formula for the “truncated” case can be easily obtained by assuming that the tree is split
292 into multiple sub-trees, each originating at the truncation age, and considering each sub-tree an independent
293 realization of the same birth-death process and subject to the same sampling fraction ρ . To avoid non-global
294 local optima, the fitting can be repeated multiple times, each time starting at random start values for the fitted
295 parameters, and the best fit among all repeats is kept. We also developed similar computer code for fitting
296 the pulled speciation rate λ_p to extant timetrees, implemented as function `fit_hbd_ps_r_on_grid` in the R
297 package `castor`.

298 Supplemental Fig. S4 shows an example where either the r_p or λ_p were accurately fitted to an extant timetree,
299 simulated under a birth-death scenario subject to an early rapid radiation event and followed by a mass
300 extinction event. In this example, we limited fitting to ages where the LTT was over 500 lineages (i.e.,
301 $M(\tau^*) = 500$), and repeated the fitting 100 times to avoid non-global local optima.

302 S.5 Fitting birth-death models to trees yields unreliable results

303 To illustrate the identifiability issues discussed in the main article, we simulated and analyzed two massive
304 extant timetrees ($\sim 114,000$ and $\sim 785,000$ tips) via a birth-death process, subject to a mass extinction event
305 (both trees) and a rapid radiation event (second tree). Instead of fitting models of the same functional form
306 as used in the simulations, we fitted generic piecewise-linear curves for λ and μ that could in principle take
307 various alternative shapes, and visually compared the estimated profiles to their true profiles (Supplemental
308 Figs. S5A–F). Specifically, we fitted λ and μ at multiple discrete time points, treating the rates at each time
309 point as free parameters, while assuming a known ρ . Despite the enormous tree sizes, and despite the fact that
310 the fitted models reproduced the trees' LTTs and the true r_p extremely well (Supplemental Figs. S5A,C,D,F),
311 the estimated λ and μ were far from their true values and even exhibited spurious trends (Supplemental Figs.
312 S5BE). This is consistent with our expectation that the particular combination of fitted λ and μ is essentially
313 a random pick from the periphery of the true process's congruence class. In contrast, when we fixed μ to
314 its true profile, λ was accurately estimated (Supplemental Fig. S2), consistent with the expectation that any
315 given μ and ρ fully determine the corresponding λ in the congruence class.

316 We also examined a large extant timetree of 79,874 seed plant species (Supplemental Fig. S3) published
317 by [Smith and Brown \(2018\)](#), and estimated λ and μ over the last 100 million years using two alternative
318 approaches (methods details in Supplement S.6). In the first approach, we fitted generic piecewise-linear
319 curves for λ and μ , similarly to the previous example. In the second approach, we fitted parameterized time
320 curves for λ and μ that included an exponential as well as a polynomial term ([Morlon et al., 2011](#)). Even
321 though both approaches yielded similar estimates for r_p , and both accurately reproduced the tree's LTT, they
322 yielded markedly different λ and μ (Supplemental Figs. S5D–F). This observation is consistent with our
323 argument that, depending on the precise set of models considered, the estimated λ and μ will generally be a
324 random pick from the underlying (true or close-to-true) congruence class.

325 S.6 Fitting birth-death models to seed plants

326 An extant timetree of 79,874 seed plant species, constructed using GenBank sequence data with a back-
327 bone provided by [Magallón et al. \(2015\)](#), was obtained from the Supplemental Material published by [Smith](#)
328 and [Brown \(2018\)](#), tree "CBMB"). The tree is shown in Supplemental Fig. S3. The sampling fraction
329 was calculated based on the tree size and the number of extant seed plant species estimated at 422,127
330 by [Govaerts \(2001\)](#). As mentioned in Supplement S.5, two approaches were used to fit λ and μ over
331 time. In the first approach, λ and μ were allowed to vary independently at 8 discrete and equidistant time
332 points (assuming piecewise linearity between grid points) and were estimated via maximum-likelihood us-
333 ing the function `fit_hb_model_on_grid` in the R package `castor` ([Louca and Doebeli, 2017](#)) (options
334 "condition='stem'", `relative_dt=0.001`). Fitting was repeated 100 times using random start param-
335 eters to avoid local non-global optima in the likelihood function. The PDR was then estimated from the fitted
336 λ and μ using the formula in Eq. (1) and using central finite differences to calculate derivatives on the time
337 grid. In the second approach, λ and μ were assumed to be of the following general functional forms:

$$\lambda(\tau) = \max \left(0, p_1 \cdot e^{-p_2 \cdot \tau / \tau_r} + p_3 + p_4 \cdot \frac{\tau}{\tau_r} + p_5 \cdot \left(\frac{\tau}{\tau_r} \right)^2 + p_6 \cdot \left(\frac{\tau}{\tau_r} \right)^3 + p_7 \cdot \left(\frac{\tau}{\tau_r} \right)^4 \right) \quad (83)$$

$$\mu(\tau) = \max \left(0, q_1 \cdot e^{-q_2 \cdot \tau / \tau_r} + q_3 + q_4 \cdot \frac{\tau}{\tau_r} + q_5 \cdot \left(\frac{\tau}{\tau_r} \right)^2 + q_6 \cdot \left(\frac{\tau}{\tau_r} \right)^3 + q_7 \cdot \left(\frac{\tau}{\tau_r} \right)^4 \right), \quad (84)$$

339 where τ_r is the age of the root and $p_1, \dots, p_7, q_1, \dots, q_7$ are fitted parameters. Parameters were fitted using the
340 `castor` function `fit_hbd_model_parametric` (options “`condition='stem'`, `relative_dt=0.001`,
341 `param_min=-100`, `param_max=100`”). As in the first approach, fitting was repeated 100 times to avoid
342 local non-global optima. In both approaches, the likelihood only incorporated branching events at ages
343 between 0 and 130 Myr, since the LTT and any parameter estimates become much less reliable at older ages.

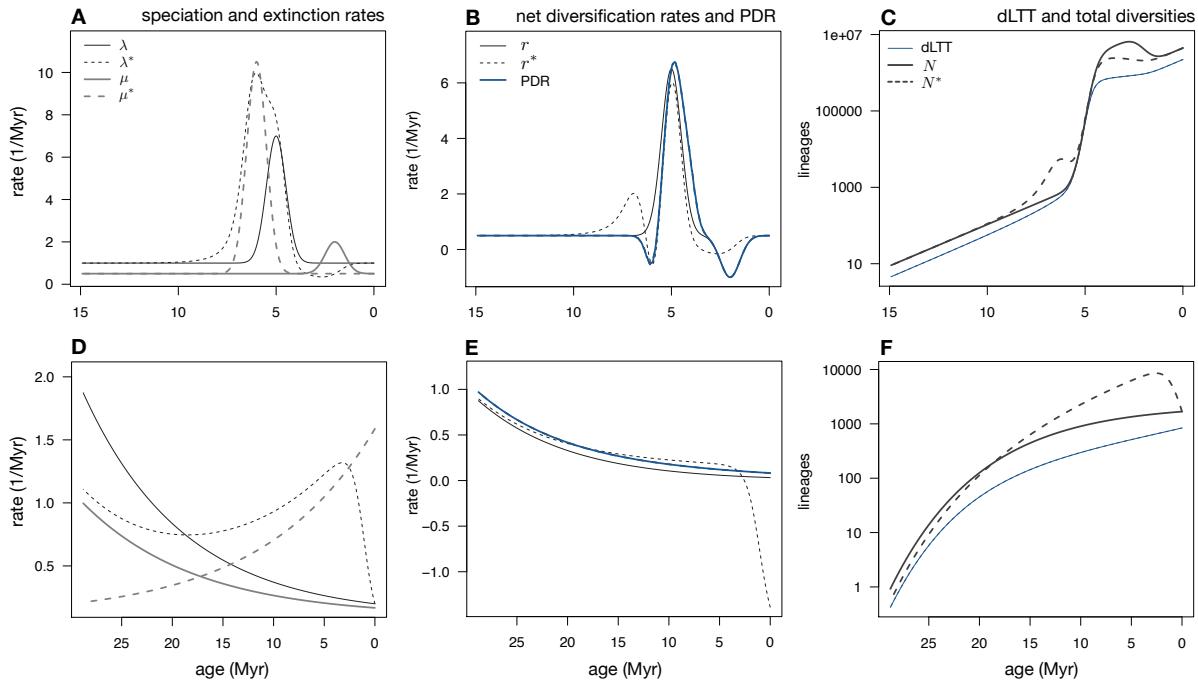


Figure S1: Examples of congruent birth-death processes. (A–C) Example of two congruent yet markedly different birth-death models. Both models exhibit a temporary spike in the extinction rate and a temporary spike in the speciation rate, however the timings of these events differ substantially between the two models. Both models exhibit the same dLTT and the same pulled diversification rate r_p , and would yield identical likelihoods for any given extant timetree. (A) Speciation rates (λ and λ^*) and extinction rates (μ and μ^*) of the two models, plotted over time. Continuous curves correspond to the first model, dashed curves to the second model. (B) Net diversification rates (r and r^*) and pulled diversification rate r_p of the two models. (C) Deterministic LTT (dLTT) and deterministic total diversities (N and N^*) predicted by the two models. (D–F) Another example of two congruent models. In the first model, the speciation and extinction rates both decrease exponentially over time, whereas in the second model the extinction rate increases exponentially over time and the speciation rate exhibits variable directions of change over time. In all models the sampling fraction is $\rho = 0.5$.

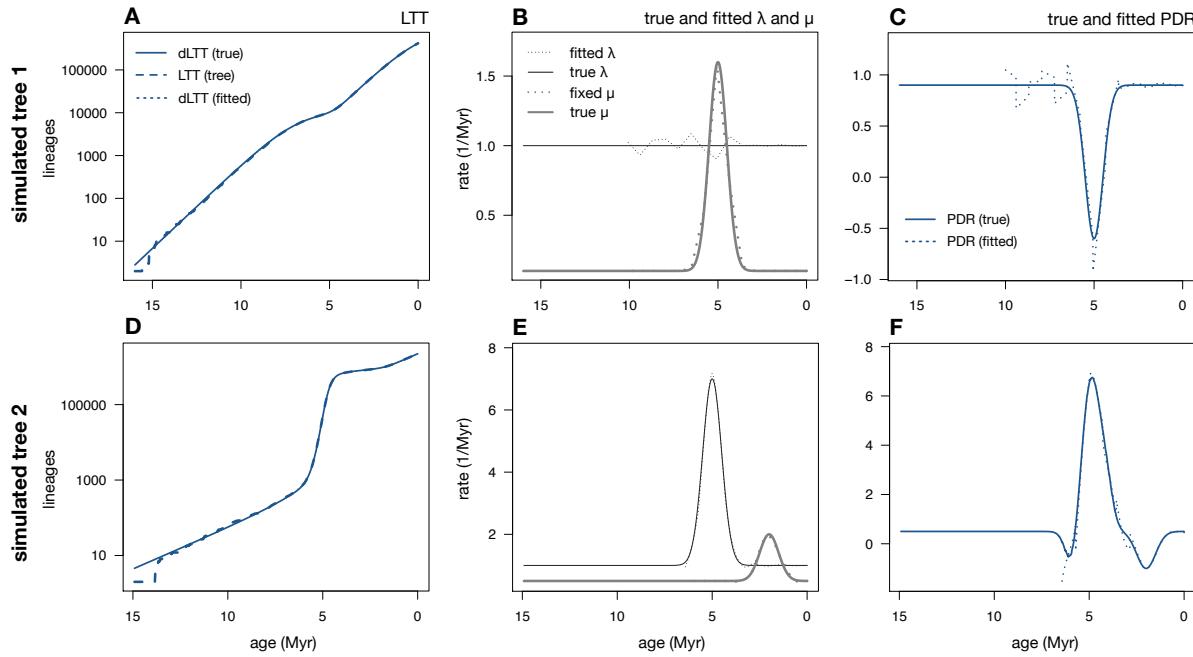


Figure S2: Estimating λ when μ and ρ are fixed. (A–C) Example analysis of a simulated extant timetree ($\sim 114,000$ tips), exhibiting a mass extinction event at ~ 5 Myr before present. A birth-death model was fitted while fixing μ and ρ to their true values; λ was fitted at 15 discrete time points. (A) Lineages through time curve (LTT) of the generated tree (long-dashed curve), deterministic LTT (dLTT) of the true model that generated the tree (continuous curve), and dLTT of a maximum-likelihood fitted model (short-dashed curve). The fitted dLTT is practically identical to the true dLTT and is thus covered by the latter. (B) True speciation and extinction rates (continuous curves), along with the fitted speciation rate and fixed extinction rate (dashed curves). (C) Pulled diversification rate (PDR) of the true model (r_p , continuous curve), compared to the PDR of the fitted model (dashed curve). (D–F) Example analysis of a simulated extant timetree ($\sim 785,000$ tips), exhibiting a rapid radiation event at ~ 5 Myr before present and a mass extinction event at ~ 2 Myr before present. A birth-death model was fitted similarly to the previous example, and D–F are analogous to A–C. In both cases, rate estimation was restricted to ages where the LTT included at least 500 lineages.

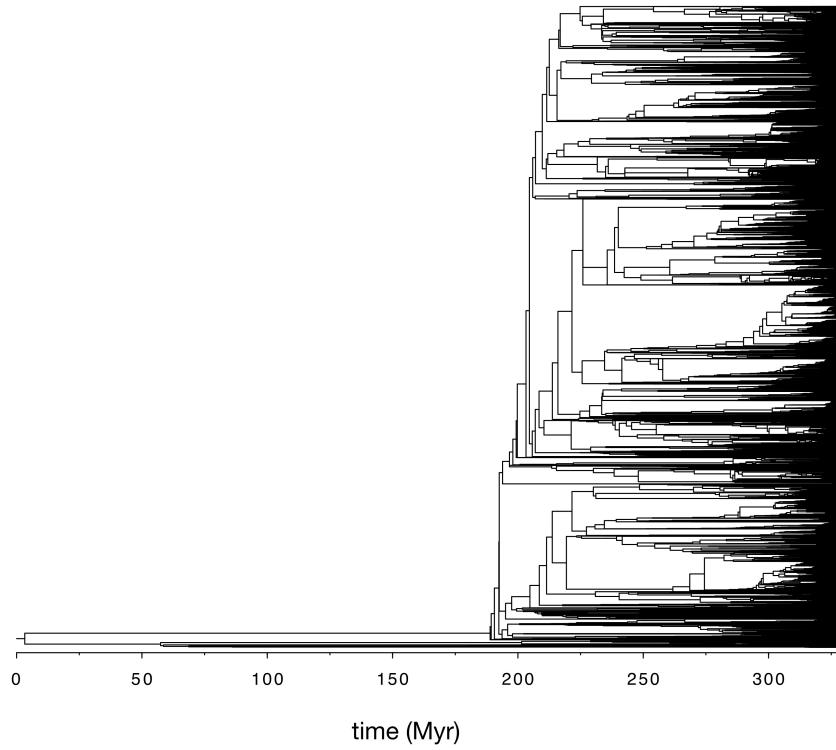


Figure S3: Seed plant tree. Extant timetree of 79,874 seed plant species, discussed in the main article. The tree was constructed and made available by [Smith and Brown \(2018\)](#) (tree “GBMB”).

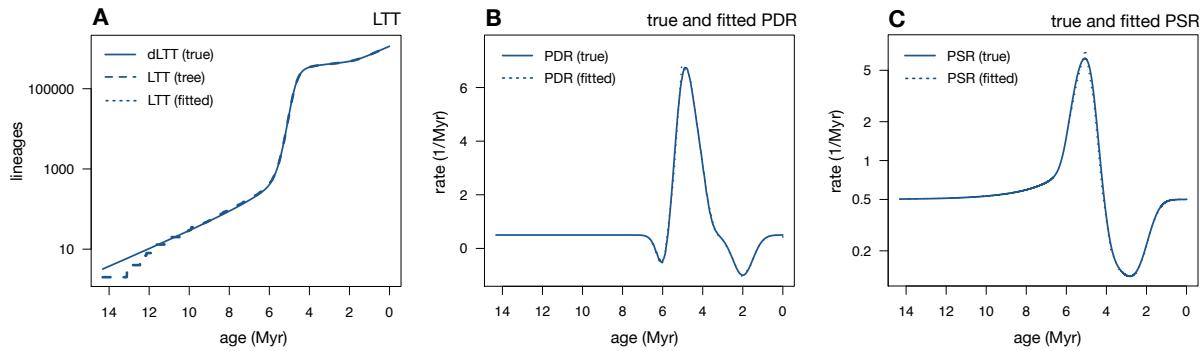


Figure S4: Fitting congruence classes instead of models. Analysis of an extant timetree generated by a birth-death model, exhibiting a temporary rapid radiation event about 5 Myr before present and a mass extinction event about 2 Myr before present. A congruence class was fitted to the timetree either in terms of the pulled diversification rate (PDR, r_p) and the product $\rho\lambda_o$, or in terms of the pulled speciation rate (PSR, λ_p). (A) Lineages through time curve (LT) of the tree (long-dashed curve), together with the deterministic LTT (dLTT) of the true model (continuous curve) and the dLTT of the fitted congruence classes (short-dashed curve); in both cases the fitted dLTT was virtually identical to the true dLTT, and is thus completely covered by the latter. (B) PDR of the true model (continuous curve), compared to the fitted PDR (short-dashed curve). (C) PSR of the true model (continuous curve), compared to the fitted PSR (short-dashed curve). The PDR and PSR were fitted via maximum-likelihood using the R package `castor` ([Louca and Doeblei, 2017](#)), allowing the PDR or PSR to vary freely at 15 discrete equidistant time points (Supplement S.4).

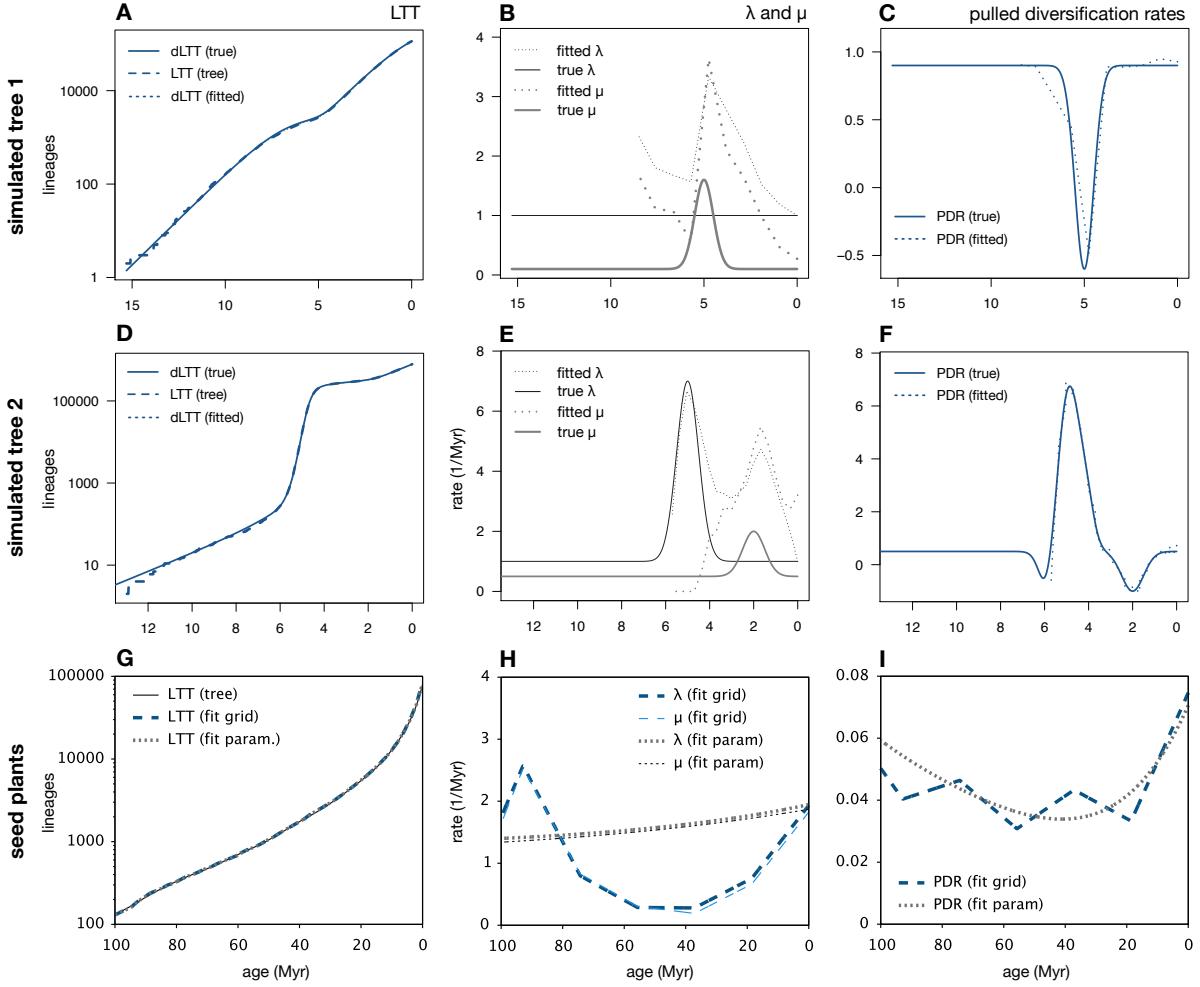


Figure S5: Identifiability issues persist in large trees. (A–C) Diversification analysis of a timetree ($\sim 114,000$ tips) simulated from a birth-death process exhibiting a mass extinction event at ~ 5 Myr before present. (A) Lineages through time curve (LTT) of the generated tree (long-dashed curve), deterministic LTT (dLTT) of the true model that generated the tree (continuous curve), and dLTT of a maximum-likelihood fitted model (short-dashed curve). The fitted dLTT is practically identical to the true dLTT and is thus covered by the latter. (B) True speciation and extinction rates (continuous curves), compared to fitted speciation and extinction rates (dashed curves). (C) Pulled diversification rate (PDR) of the true model (continuous curve), compared to the PDR of the fitted model (dashed curve). (D–F) Diversification analysis of a timetree ($\sim 785,000$ tips) simulated from a birth-death process exhibiting a rapid radiation event at ~ 5 Myr before present and a mass extinction event at ~ 2 Myr before present. Sub-figures D–F are analogous to A–C. See Supplemental Fig. S2 for fitting results when μ is fixed to its true value. (G–I) Diversification analyses of an extant timetree of 79,874 seed plant species (Smith and Brown, 2018), performed either by fitting λ and μ on a grid of discrete time points or by fitting the parameters of generic polynomial/exponential functions for λ and μ . (G) LTT of the tree, dLTT of the grid-fitted model and dLTT of the fitted parametric model. (H) Speciation and extinction rates predicted by the grid-fitted model or the fitted parametric model. (I) PDR predicted by the grid-fitted model and the fitted parametric model.

344 References

- 345 Cantalapiedra, J.L., FitzJohn, R.G., Kuhn, T.S., Fernández, M.H., DeMiguel, D., Azanza, B., *et al.* (2014)
346 Dietary innovations spurred the diversification of ruminants during the Caenozoic. *Proceedings of the*
347 *Royal Society B: Biological Sciences* **281**: 20132746.
- 348 Climenhaga, V. and Katok, A. (2017) From Groups to Geometry and Back. Student Mathematical Library.
349 Providence, Rhode Island, USA: American Mathematical Society.
- 350 Condamine, F.L., Rolland, J., and Morlon, H. (2013) Macroevolutionary perspectives to environmental
351 change. *Ecology Letters* **16**: 72–85.
- 352 Egan, A.N. and Crandall, K.A. (2008) Divergence and diversification in North American Psoraleeae
353 (Fabaceae) due to climate change. *BMC Biology* **6**: 55.
- 354 Esselstyn, J.A., Timm, R.M., and Brown, R.M. (2009) Do geological or climatic processes drive speciation
355 in dynamic archipelagos? the tempo and mode of diversification in southeast asian shrews. *Evolution* **63**:
356 2595–2610.
- 357 Govaerts, R. (2001) How many species of seed plants are there? *Taxon* **50**: 1085–1090.
- 358 Kendall, D.G. (1948) On some modes of population growth leading to RA Fisher's logarithmic series distri-
359 bution. *Biometrika* **35**: 6–15.
- 360 Kendall, D.G. *et al.* (1948) On the generalized “birth-and-death” process. *The annals of mathematical statis-
361 tics* **19**: 1–15.
- 362 Louca, S. and Doebeli, M. (2017) Efficient comparative phylogenetics on large trees. *Bioinformatics* **34**:
363 1053–1055.
- 364 Louca, S., Shih, P.M., Pennell, M.W., Fischer, W.W., Parfrey, L.W., and Doebeli, M. (2018) Bacterial diver-
365 sification through geological time. *Nature Ecology & Evolution* **2**: 1458–1467.
- 366 Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L.L., and Hernández-Hernández, T. (2015) A metacali-
367 brated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist* **207**:
368 437–453.
- 369 Mayhew, P.J., Jenkins, G.B., and Benton, T.G. (2008) A long-term association between global temperature
370 and biodiversity, origination and extinction in the fossil record. *Proceedings of the Royal Society B:
371 Biological Sciences* **275**: 47–53.
- 372 McPeek, M.A. (2008) The ecological dynamics of clade diversification and community assembly. *The Amer-
373 ican Naturalist* **172**: E270–E284.
- 374 Morlon, H. (2014) Phylogenetic approaches for studying diversification. *Ecology Letters* **17**: 508–525.
- 375 Morlon, H., Parsons, T.L., and Plotkin, J.B. (2011) Reconciling molecular phylogenies with the fossil record.
376 *Proceedings of the National Academy of Sciences* **108**: 16327–16332.
- 377 Morlon, H., Potts, M.D., and Plotkin, J.B. (2010) Inferring the dynamics of diversification: A coalescent
378 approach. *PLOS Biology* **8**: e1000493.
- 379 Nee, S., May, R.M., and Harvey, P.H. (1994) The reconstructed evolutionary process. *Philosophical Trans-
380 actions of the Royal Society of London B: Biological Sciences* **344**: 305–311.

- 381 Rabosky, D.L. (2006a) Laser: a maximum likelihood toolkit for detecting temporal shifts in diversification
382 rates from molecular phylogenies. *Evolutionary Bioinformatics* **2**.
- 383 Rabosky, D.L. (2006b) Likelihood methods for detecting temporal shifts in diversification rates. *Evolution*
384 **60**: 1152–1164.
- 385 Rabosky, D.L. and Lovette, I.J. (2008) Explosive evolutionary radiations: decreasing speciation or increasing
386 extinction through time? *Evolution: International Journal of Organic Evolution* **62**: 1866–1875.
- 387 Sanmartín, I. and Meseguer, A.S. (2016) Extinction in phylogenetics and biogeography: From timetrees to
388 patterns of biotic assemblage. *Frontiers in Genetics* **7**: 35.
- 389 Silvestro, D., Schnitzler, J., and Zizka, G. (2011) A bayesian framework to estimate diversification rates and
390 their variation through time and space. *BMC Evolutionary Biology* **11**: 311.
- 391 Smith, S.A. and Brown, J.W. (2018) Constructing a broadly inclusive seed plant phylogeny. *American Journal*
392 *of Botany* **105**: 302–314.
- 393 Stadler, T. (2009) On incomplete sampling under birth–death models and connections to the sampling-based
394 coalescent. *Journal of Theoretical Biology* **261**: 58–66.
- 395 Stadler, T. (2011) Mammalian phylogeny reveals recent diversification rate shifts. *Proceedings of the National*
396 *Academy of Sciences* **108**: 6187–6192.
- 397 Stadler, T. (2013) How can we improve accuracy of macroevolutionary rate estimates? *Systematic Biology*
398 **62**: 321–329.
- 399 Stadler, T. and Steel, M. (2019) Swapping birth and death: Symmetries and transformations in phylodynamic
400 models. *bioRxiv* p. 494583.