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*
[*]Correspondence should be addressed to both authors: www.loucalab.com and pennell@zoology.ubc.ca

Abstract

Time-calibrated molecular phylogenies of extant species ("extant timetrees") are widely used for estimating the dynamics of diversification rates (1–6) and testing for associations between these rates and environmental factors (5, 7) or species traits (8). However, there has been considerable debate surrounding the reliability of these inferences in the absence of fossil data (9–13), 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 and yet plausible 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 composite variables, which contain all available information about past dynamics of diversification (14); this suggests an alternative paradigm for learning about the past from extant timetrees.

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

Introduction

A central challenge in evolutionary biology is to explain why some taxonomic groups and some time periods have so many species while others have so few; ultimately this means estimating and explaining variation in rates of speciation and extinction (13). Estimating these rates is crucial to investigating fundamental questions such as the role of biotic and abiotic processes in shaping patterns of species richness (7), how Earth's biota recover after mass extinction events (15, 16) and whether there are general dynamics that govern how biodiversity accumulates (17). Measuring such rates has taken on a new urgency as we try to understand how anthropogenically induced extinctions compare to "background" rates (18, 19). In the medical domain, "speciation" and extinction rates are key parameters that provide insights into the historical dynamics and future trajectories of viral epidemics (20, 21).

Unfortunately, the vast majority of lineages that have ever lived have not left any trace in the fossil record, and hence we are forced to attempt to reconstruct their diversification dynamics from incomplete data. Indeed, for many groups the fossil record is so incomplete that the primary source of information on past diversification dynamics comes from time-calibrated molecular phylogenies of extant lineages ("extant timetrees"). There is now an abundance of increasingly sophisticated methods for extracting this information, with most state-of-the-art methods fitting variants of a birth-death process (22) to extant timetrees (1–6, 23). These methods have, collectively, been used in thousands of studies and have substantially contributed to our understanding of the drivers of diversity through time. Despite their popularity, there has been a long-lingering doubt about many of these inferences. For one, simulation studies have repeatedly shown that some variables, especially extinction rates, are generally difficult to estimate (11, 24–28). But an even more fundamental issue, which has particularly drawn the attention of paleobiologists, is that there may not be sufficient information in a molecular phylogeny to fully reconstruct historical changes in diversification rates. For example, when speciation and extinction rates vary through time — and there is abundant evidence from the fossil record that they do (13) — mass extinction events can erode much of the signal of preceding diversification dynamics (9, 10, 12), and may themselves even be confused with stagnating speciation rates (29). To date these critical identifiability issues remain poorly understood, and no general theory exists for describing which diversification scenarios can be distinguished from each other and precisely what information on diversification rates is in principle extractable from extant phylogenies.

Here, we present a solution to this problem: We develop a mathematical framework for assessing the identifiability of the general stochastic birth-death process with homogeneous rates, where speciation ("birth") rates (λ) and extinction ("death") rates (μ) can vary over time, and which underlies the majority of existing methods for reconstructing diversification dynamics from phylogenies (5). By considering the full space of possible diversification scenarios (i.e., with arbitrary λ and μ), rather than special cases (as has been done so far), we reveal a fundamental and surprisingly general property of the birth-death process that has far reaching implications for diversification analyses. Specifically, we show that for any given birth-death model there exists an infinite number of alternative birth-death models that can explain any given extant timetree equally well as the candidate model. This ambiguity persists for arbitrarily large trees and cannot be resolved even with an infinite amount of data using any statistical method. Crucially, these alternative models may appear to be similarly plausible and yet exhibit markedly different features, such as different trends through time in both λ and μ. Using simulated and real timetrees as examples, we demonstrate how failing to recognize this immense ambiguity may seriously mislead our inferences about past diversification dynamics, shedding doubt on conclusions from countless previous studies. We further show that these sets of "congruent" models can be uniquely identified based on suitably defined composite variables: the "pulled speciation rate", corresponding to the effective λ in the hypothetical absence of extinction and under complete species sampling, or equivalently, the "pulled diversification rate", corresponding to the effective net diversification

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

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

One of the most important features of extant timetrees is the lineages-through-time curve (LTT), which counts the number of lineages at each time in the past that are represented by at least one extant descending species in the tree. The LTT provides a simple visual overview of a tree's branching density over time and importantly, contains all the information encoded in the tree regarding speciation and extinction rates (30) (see also Supplement S.1.2). This is because the likelihood of a extant timetree under a given birth-death model with homogeneous rates depends solely on the tree's LTT, but not on any other properties of the tree that do not affect the LTT.

Here we show that an elegant analogous relationship also exists between the likelihood of a tree and the LTT that would be predicted by a given birth-death model. Any given speciation and extinction rates over time, λ and μ, and the probability that an extant species will be included in the tree ρ (present-day "sampling fraction"), can be used to define a "deterministic" diversification process, where the number of lineages through time no longer varies stochastically but according to a set of differential equations (3, 31–33; also see Supplement S.1). The LTT predicted by these differential equations ("deterministic LTT", or dLTT) is a mere theoretical property of the model that resembles the LTT of a tree only if the tree is sufficiently large for stochastic effects to become negligible, and assuming the model is an adequate description of the process that generated the tree. It can be shown, however, that the likelihood of a tree under a given birth-death model can be written purely in terms of the tree's LTT and the model's dLTT (Supplement S.1.2). This means that any two models with the same dLTT (conditioned on the number of extant species sampled, M_o) yield identical likelihoods for the tree. In the following, we shall therefore call any two models "congruent" if they have the same dLTT for any given M_o. Note that two models are either congruent or non-congruent regardless of the particular tree considered, meaning that model congruency is a property of models and not the data (Supplement S.1). Furthermore, the probability distribution of tree sizes generated by a birth-death model, when conditioned on the age of the stem (or crown), is identical among congruent models (Supplement S.1.7). Hence, congruent models have equal probabilities of generating any given timetree and LTT, analogous to how congruent geometric objects exhibit similar geometric properties (discussion in Supplement S.1.8). In the absence of further information or constraints, congruent models cannot possibly be distinguished solely based on extant timetrees, neither through the likelihood nor any other test statistic (such as the γ statistic (34)). Note that whether or not a given phylogenetic data set is sufficient to statistically distinguish between non-congruent models is an entirely different matter.

Congruent model sets are infinitely large and infinite-dimensional

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

S.1.1). In the absence of extinction ($\mu = 0$) and under complete species sampling ($\rho = 1$), λ_{p} is identical to λ, however in the presence of extinction λ_{p} is pulled downwards relative to λ at older ages, while under incomplete sampling λ_{p} is pulled downwards relative to λ near the present. We thus henceforth refer to λ_{p} as the "pulled speciation rate" of a model. Since a model's dLTT is fully determined by λ_{p} and, reciprocally, λ_{p} is fully determined by the dLTT, it becomes evident that two models are congruent if and only if the have the same pulled speciation rate. The latter can also be used to calculate a variable called "pulled diversification rate" (14), defined as:

$$r_{\mathrm{p}} = \lambda - \mu + \frac{1}{\lambda}\frac{d\lambda}{d\tau}. \tag{1}$$

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 differs from r when λ varies with time. As shown in Supplement S.1.1, λ_{p} and r_{p} are linked through the following differential equation:

$$\frac{d\lambda_{\mathrm{p}}}{d\tau} = \lambda_{\mathrm{p}} \cdot (r_{\mathrm{p}} - \lambda_{\mathrm{p}}), \tag{2}$$

with initial condition $\lambda_{\mathrm{p}}(0) = \rho\lambda_o$ (where $\lambda_o = \lambda(0)$ is the present-day speciation rate). Equation (2) reveals that r_{p} is completely determined by λ_{p} (one can just solve for r_{p}). Reciprocally, λ_{p} is completely determined by r_{p} and some initial condition (i.e., λ_{p} specified at some fixed time), since one can just solve the differential equation for λ_{p} (see solution in Supplement S.1.6). We thus conclude that two birth-death models are congruent if, and only if, they have the same r_{p} and the same λ_{p} at some time point (for example the same product $\rho\lambda_o$).

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

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

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

$$\frac{d\lambda^*}{d\tau} = \lambda^* \cdot (r_{\mathrm{p}} - \lambda^* + \mu^*), \tag{4}$$

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

For illustration, consider the simulations in Figure 1, showing four markedly distinct and yet congruent birth-death models. The first scenario exhibits a constant λ and a temporary spike in μ (mass extinction event), the second scenario instead exhibits a constant μ and a temporary drop in λ (temporary stasis of speciations) around the same time, the third scenario exhibits a mass extinction event at a completely different time and a fluctuating λ, while the fourth scenario exhibits an exponentially decaying μ and a fluctuating λ. These congruent scenarios were obtained simply by assuming alternative extinction rates, and a myriad of other congruent scenarios also exist. Analogous situations can be readily found in the literature. Figures 2A–C, for example, show a birth-death model with exponentially varying speciation and extinction rates,

$\lambda = \lambda_o e^{\alpha\tau}$ and $\mu = \mu_o e^{\beta\tau}$ (where λ_o, μ_o, α and β are fitted parameters), as commonly considered in other studies (3, 35), fitted to a massive timetree of 79,874 extant seed plant species via maximum likelihood. Simply by modifying the coefficient β and choosing λ according to Eq. (4), one can obtain an infinite number of congruent and similarly complex scenarios, with even opposite trends over time (Fig. 2B). Similarly, Figs. 2D–F show origination and extinction rates of marine animal genera estimated from fossil data (36), compared to two congruent scenarios, one where the linear trend of μ has been reversed (Fig. 2B) and one where μ was set to zero (Fig. 2C).

Such ambiguities were described previously in special cases. For example, Kubo and Iwasa (31) recognized that a variable λ and constant μ can be exchanged for a constant λ and a variable μ to produce the same dLTT; similarly Stadler (29) and Crisp *et al.* (37) observed that simulations of mass extinctions produced similar LTTs as simulations of temporarily stagnating diversification processes. Other previous work on constant-rate birth-death models revealed that alternative combinations of time-independent λ, μ and ρ can yield the same likelihood for a tree (38–41). By generalizing these analyses to the time-variable case, we have revealed that in fact vast expanses of model space are practically indistinguishable.

Model congruency compromises existing reconstruction methods

Since the likelihood of an extant timetree can be expressed purely in terms of r_{p} and the product $\rho\lambda_o$ (Supplement S.1.6), or alternatively purely in terms of λ_{p} (Supplement S.1.3), extant timetrees only provide information about the congruence class of a generating process and not the actual speciation and extinction rates. This identifiability issue can be interpreted as follows: Since all information available on past diversification dynamics (representable by birth-death models, to be precise) is encoded in a single curve, namely the LTT, one should not expect to be able to "extract" from it two independent curves (λ and μ) without additional information, as this would essentially double the amount of information at hand.

In order to estimate λ and μ, previous phylogenetic studies have been imposing strong and largely arbitrary constraints. For example, many studies assume that λ or μ vary exponentially through time (42). However this specific functional form is rarely justified biologically, and alternative functional forms of comparable simplicity and shape (e.g., the logistic function, or the Gauss error function) can be envisioned. Normally one expects that, regardless of which of these functional forms is considered, with sufficient data fitting any of these forms will lead to qualitatively similar trends and shapes. This expectation simply does not hold in our case, because the best-fitting representative within any given model set will generally only be the one closest to the congruence class of the true process, rather than closest to the true process itself (Fig. 3). Consequently, fitting alternative functional forms (i.e., alternative model sets) can result in drastically different inferences with alternative trends in λ and μ over time, even if the each functional form used is in principle adequate for approximating the true historical λ and μ (examples in Supplement S.6 and Supplemental Fig. S5). This conclusion applies to virtually any model set, including birth-death shift models where λ and μ change at discrete time points (43).

We stress that common model selection methods based on parsimony or "Occam's razor", such as the Akaike Information Criterion (AIC; 44) and the Bayesian Information Criterion (BIC; 45) that penalize excessive parameters, generally cannot resolve this issue for multiple reasons. First, these information criteria were designed to minimize the error of future model-based predictions by avoiding overfitting to finite data, and not to identify the actual process that generated the data — these are very different scientific goals with well-known trade-offs. There is no reason to believe that the simplest scenario in a congruence class will be the one closest to the truth. Indeed, even if the true model is included in a congruence class, it will almost always be the case that there are both simpler and more complex scenarios within the same congruence class

(e.g., Figs. 2D–F) and, crucially, all of these alternative models remain equally likely even with infinitely large datasets. Second, if one were to apply AIC or BIC, it is unclear how to quantify the complexity of a diversification scenario in comparison with alternative scenarios, which may be described using different functional forms. It is tempting to think that one could simply count the number of parameters. However, any given curve can be written using various alternative functional forms parameterized in distinct ways (recall that ultimately we wish to approximately estimate the curves λ and μ, not the parameters of some functional form); the number of parameters is a property of parameterized sets of curves, not of a single curve. Even if that were not the case, the number of parameters conventionally associated with a given functional form need not necessarily reflect our intuition about complexity: Is a linear extinction rate ($\mu = \alpha + \beta \cdot \tau$, two parameters) more or less complex than an exponentially decaying rate ($\mu = \alpha e^{\beta\tau}$) or an oscillating rate of the form $\mu = \alpha \sin^2(\beta\tau)$? In addition, different members of a congruence class may be described with different functional forms involving the same number of parameters. For example, the diversification scenario with linear extinction rate ($\mu = \alpha + \beta \cdot \tau$) and constant speciation rate ($\lambda = \gamma$) (3 parameters, assuming complete species sampling) is congruent to an alternative and markedly different scenario with zero extinction rate ($\mu^* = 0$) and λ^* defined as the solution to the differential equation $d\lambda^*/d\tau = \lambda^* \cdot (\gamma - \alpha - \beta\tau - \lambda)$ with initial condition $\lambda^*(0) = \gamma$ (again 3 parameters); there is no reason to prefer one congruent scenario over the other based on the number of parameters or biological realism. Third, even when fitting models of the same functional form, as explained above the maximum-likelihood model will a priori tend to be the one closest to the congruence class of the true process, rather than the true process itself, and neither AIC nor BIC would resolve this (since all other allowed models would have the same number of parameters but lower likelihood). Supplemental Fig. S6 shows examples where maximum-likelihood fitted models, chosen among a wide range of model complexities based on AIC, grossly fail to estimate the true rates even when fitting to a massive tree with 1,000,000 tips, despite the fact that the models could in principle have accurately captured the true rates.

Previous studies have not recognized the breadth of this issue because they typically only consider a limited set of candidate models at a time, both when analyzing real datasets as well as when assessing parameter identifiability via simulations; as a result, previous studies have been (un)lucky enough to not compare two models in the same congruence class (see Supplements S.2 and S.3 for reasoning). For example, if a tree was generated by an exponentially decaying λ and μ (e.g., via simulations), then fitting an exponential functional form will of course yield accurate estimates of the exponents; however if the generating process was only approximately exponential and better described by another gradually decaying function, then fitting an exponential curve could even lead to opposite trends (examples in Fig. 2 and Supplement S.6).

It is important to realize that congruent scenarios can have markedly different macroevolutionary implications. For example, Steeman *et al.* (46) reconstructed past speciation rates of Cetaceans (whales, dolphins, and porpoises) based on an extant timetree and using maximum-likelihood (assuming $\mu = 0$). Steeman *et al.* (46) found a temporary increase of λ during the late Miocene-early Pliocene (Fig. 4), suggesting a potential link between Cetacean radiations and concurrent paleoceanographic changes. However, alternatively to assuming $\mu = 0$, one could assume that μ was close to λ, consistent with common observations from the fossil record (13). For example, by setting $\mu = 0.9 \cdot \lambda$ one obtains a congruent scenario in which λ no longer peaks during the late Miocene-early Pliocene but instead exhibits a gradual slowdown throughout most of Cetacean evolution (Fig. 4B). Both scenarios are similarly complex and both could have generated the timetree at equal probabilities. Even the common methodological decision to estimating net diversification ($r = \lambda - \mu$) rather than λ and μ separately (8, 11), is no longer meaningful in light of our results; the shape of r is not conserved across a congruence class. Likewise, the models in a congruence class will not share "average" (however defined) rates; hence absolute rate estimates (1), which have been used to estimate broad macroevolutionary patterns (17) and background rates of extinction (19), are also unlikely to be accurately reconstructed. These issues are likely also present in more complex models with additional free parameters,

for example where some clades exhibit distinct diversification regimes (42, 47). In such situations, the tree can always be decomposed into a set of sub-trees with distinct LTT curves, each of which is subject to the same identifiability issues as described here. Our findings thus shed doubts over a lot of previous work on diversification dynamics, including some of the conclusions from our own work (14, 17).

Ways forward

Our findings for birth-death models of diversification are closely analogous to classic results from coalescent theory in population genetics (48, 49), where an infinite number of models can give rise to the same drift process as the idealized Wright-Fisher model. This realization was profoundly important for the field; it focused researchers' attention on the dynamics of the effective population size N_e, a composite but identifiable parameter that is the same for all models with the same Wright-Fisher drift process, rather than on the actual (but non-identifiable) historical demography, sex ratios etc. of the population. Consequently, the field has adapted and developed a plethora of tools to infer changes in N_e through time and across the genome. Here we have found an analogous generality, and have confirmed previous suspicions that many historical diversification scenarios may not be distinguishable using extant phylogenies alone (10, 13, 30, 31). We have shown that such congruent scenarios can be defined in terms of the λ_p, or equivalently, in terms of the r_p and $\rho\lambda_o$, all of which are identifiable provided sufficient data (indeed, for sufficiently large trees these variables can be directly calculated from the slope and curvature of the LTT; 14). Each congruence class contains exactly one model with $\mu = 0$ and $\rho = 1$, which is also the only model where $\lambda = \lambda_\mathrm{p}$; hence the pulled speciation rate can be interpreted as the effective speciation rate generating the congruence class's dLTT in the absence of extinctions and under complete species sampling. Similarly, each congruence class contains an infinite number of models with time-independent λ, and for these models $r_\mathrm{p} = r$; hence the pulled diversification rate can be interpreted as the effective net diversification rate if λ was time-independent. It is in this way that λ_p and r_p, being identifiable and "effective" rates in idealized scenarios, are analogous to N_e in population genetics.

Of course fossil data could in principle help resolve the issues highlighted here, for example via fossilized-birth-death models (50, 51) or birth-death-chronospecies models (52), however for a large number of taxa (e.g., all prokaryotes and many soft-bodied eukaryotes) fossil data are virtually non-existent. 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 known) either using likelihood methods (Supplement S.5) or based on the slope and curvature of a tree's LTT (14). Our previous work(14) has shown that r_p itself can yield valuable insight into diversification dynamics and can be useful for testing alternative hypotheses (also see Supplement S.4). Using simulations, for example, we found that sudden rate transitions, for example due to mass extinction events, usually lead to detectable fluctuations in r_p; therefore, a relatively constant r_p over time would be indicative of constant — or only slowly changing — speciation and extinction rates (14). One can also obtain other useful composite variables from λ_p, r_p and $\rho\lambda_o$. For example, in cases where ρ is known one can obtain the "pulled extinction rate", defined as $\mu_\mathrm{p} := \lambda_o - r_\mathrm{p}$ (14). Note that $\mu_\mathrm{p}(\tau)$ is equal to the extinction rate $\mu(\tau)$ if λ has been constant from τ to the present, but differs from μ in most other cases. The present-day μ_p is related to the present-day μ as follows:

$$\mu_\mathrm{p}(0) = \mu(0) - \frac{1}{\lambda_o}\frac{d\lambda}{d\tau}\bigg|_{\tau=0}. \tag{5}$$

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

$$\frac{1}{\lambda_o}\frac{d\lambda}{d\tau}\bigg|_{\tau=0} \geq -\mu_\mathrm{p}(0). \tag{6}$$

If a macroevolutionary question is only concerned with recent speciation events (7) then one can test hypotheses using λ_o, which can be readily identified if ρ is known.

Conclusions

We have shown that for virtually any candidate birth-death process, suspected of having generated some extant timetree, there exists an infinite number of alternative and markedly different birth-death processes that could have generated the timetree with the same likelihood. Without further information or prior constraints on plausible diversification scenarios, extant timetrees alone cannot be used to reliably infer speciation rates (except at present-day), extinction rates or net diversification rates, raising serious doubts over a multitude of previous estimates of past diversification dynamics. Our work could thus explain why frequently diversification dynamics observed in the fossil record are in great disagreement with phylogenetics-based inferences (3, 5, 10, 13, 46).

On a more positive note, we resolved a long-standing debate and precisely clarified what information 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), and any other variables that can be expressed in terms of λ_{p}, r_{p} and $\rho\lambda_o$. These identifiable variables not only tell us when two models are in principle distinguishable but they can themselves yield valuable insight into past diversification dynamics. We see these as analogous to the concept of effective population size in population genetics — we cannot uniquely determine the exact sequence of events that led to current diversity, but by blending out some of the historical details we could potentially gain powerful and robust insights into general macroevolutionary phenomena.

Code availability

Computational methods used for this article, including functions for simulating birth-death models, for constructing models within a given congruence class, for calculating the likelihood of a congruence class, and for directly fitting congruence classes (either in terms of λ_{p} or in terms of r_{p} and $\rho\lambda_o$) to extant timetrees (Supplement S.5), are implemented in the R package `castor` (53).

Acknowledgments

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

Author contributions

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

Competing financial interests

The authors declare that they have no competing interests.

Materials & Correspondence

Correspondence and requests for materials should be addressed to both authors.

References

[1] Magallon, S. & Sanderson, M. J. Absolute diversification rates in angiosperm clades. *Evolution* **55**, 1762–1780 (2001).

[2] Rabosky, D. L. Likelihood methods for detecting temporal shifts in diversification rates. *Evolution* **60**, 1152–1164 (2006).

[3] Morlon, H., Parsons, T. L. & Plotkin, J. B. Reconciling molecular phylogenies with the fossil record. *Proceedings of the National Academy of Sciences* **108**, 16327–16332 (2011).

[4] Stadler, T. Recovering speciation and extinction dynamics based on phylogenies. *Journal of evolutionary biology* **26**, 1203–1219 (2013).

[5] Morlon, H. Phylogenetic approaches for studying diversification. *Ecology Letters* **17**, 508–525 (2014).

[6] Rabosky, D. L. *et al.* Bamm tools: an r package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods in Ecology and Evolution* **5**, 701–707 (2014).

[7] Schluter, D. & Pennell, M. W. Speciation gradients and the distribution of biodiversity. *Nature* **546**, 48 (2017).

[8] Maddison, W. P., Midford, P. E., Otto, S. P. & Oakley, T. Estimating a binary character's effect on speciation and extinction. *Systematic Biology* **56**, 701–710 (2007).

[9] Quental, T. B. & Marshall, C. R. Extinction during evolutionary radiations: reconciling the fossil record with molecular phylogenies. *Evolution* **63**, 3158–3167 (2009).

[10] Quental, T. B. & Marshall, C. R. Diversity dynamics: molecular phylogenies need the fossil record. *Trends in Ecology & Evolution* **25**, 434–441 (2010).

[11] Rabosky, D. L. Extinction rates should not be estimated from molecular phylogenies. *Evolution* **64**, 1816–1824 (2010).

[12] Liow, L. H., Quental, T. B. & Marshall, C. R. When can decreasing diversification rates be detected with molecular phylogenies and the fossil record? *Systematic Biology* **59**, 646–659 (2010).

[13] Marshall, C. R. Five palaeobiological laws needed to understand the evolution of the living biota. *Nature Ecology & Evolution* **1**, 165 (2017).

[14] Louca, S. *et al.* Bacterial diversification through geological time. *Nature Ecology & Evolution* **2**, 1458–1467 (2018).

[15] Jablonski, D. Mass extinctions and macroevolution. *Paleobiology* **31**, 192–210 (2005).

[16] Meredith, R. W. *et al.* Impacts of the Cretaceous terrestrial revolution and KPg extinction on mammal diversification. *Science* **334**, 521–524 (2011).

[17] Henao Diaz, L. F., Harmon, L. J., Sugawara, M. T. C., Miller, E. T. & Pennell, M. W. Macroevolutionary diversification rates show time dependency. *Proceedings of the National Academy of Sciences* **116**, 7403 (2019).

[18] Pimm, S. L. *et al.* The biodiversity of species and their rates of extinction, distribution, and protection. *Science* **344**, 1246752 (2014).

[19] De Vos, J. M., Joppa, L. N., Gittleman, J. L., Stephens, P. R. & Pimm, S. L. Estimating the normal background rate of species extinction. *Conservation Biology* **29**, 452–462 (2015).

[20] Stadler, T., Kühnert, D., Bonhoeffer, S. & Drummond, A. J. Birth–death skyline plot reveals temporal changes of epidemic spread in hiv and hepatitis c virus (hcv). *Proceedings of the National Academy of Sciences* **110**, 228–233 (2013).

[21] Arora, N. *et al.* Origin of modern syphilis and emergence of a pandemic treponema pallidum cluster. *Nature microbiology* **2**, 16245 (2017).

[22] Kendall, D. G. *et al.* On the generalized "birth-and-death" process. *The annals of mathematical statistics* **19**, 1–15 (1948).

[23] Maliet, O., Hartig, F. & Morlon, H. A model with many small shifts for estimating species-specific diversification rates. *Nature Ecology & Evolution* (2019).

[24] FitzJohn, R. G., Maddison, W. P. & Otto, S. P. Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Systematic Biology* **58**, 595–611 (2009).

[25] Davis, M. P., Midford, P. E. & Maddison, W. Exploring power and parameter estimation of the bisse method for analyzing species diversification. *BMC Evolutionary Biology* **13**, 38 (2013).

[26] Beaulieu, J. M. & O'Meara, B. C. Extinction can be estimated from moderately sized molecular phylogenies. *Evolution* **69**, 1036–1043 (2015).

[27] Rabosky, D. L. Challenges in the estimation of extinction from molecular phylogenies: a response to beaulieu and o'meara. *Evolution* **70**, 218–228 (2016).

[28] Moore, B. R., Höhna, S., May, M. R., Rannala, B. & Huelsenbeck, J. P. Critically evaluating the theory and performance of bayesian analysis of macroevolutionary mixtures. *Proceedings of the National Academy of Sciences* **113**, 9569–9574 (2016).

[29] Stadler, T. Simulating trees with a fixed number of extant species. *Systematic Biology* **60**, 676–684 (2011).

[30] 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).

[31] Kubo, T. & Iwasa, Y. Inferring the rates of branching and extinction from molecular phylogenies. *Evolution* **49**, 694–704 (1995).

[32] Harvey, P. H., May, R. M. & Nee, S. Phylogenies without fossils. *Evolution* **48**, 523–529 (1994).

[33] Billaud, O., Moen, D., Parsons, T. & Morlon, H. Estimating diversity through time using molecular phylogenies: Old and species-poor frog families are the remnants of a diverse past. *Systematic Biology* (2019).

[34] Pybus, O. G. & Harvey, P. H. Testing macro–evolutionary models using incomplete molecular phylogenies. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **267**, 2267–2272 (2000).

[35] Hallinan, N. The generalized time variable reconstructed birth-death process. *Journal of theoretical biology* **300**, 265–276 (2012).

[36] Alroy, J. Dynamics of origination and extinction in the marine fossil record. *Proceedings of the National Academy of Sciences* **105**, 11536–11542 (2008).

[37] Crisp, M. D. & Cook, L. G. Explosive radiation or cryptic mass extinction? interpreting signatures in molecular phylogenies. *Evolution* **63**, 2257–2265 (2009).

[38] Stadler, T. On incomplete sampling under birth–death models and connections to the sampling-based coalescent. *Journal of Theoretical Biology* **261**, 58–66 (2009).

[39] Morlon, H., Potts, M. D. & Plotkin, J. B. Inferring the dynamics of diversification: A coalescent approach. *PLOS Biology* **8**, e1000493 (2010).

[40] Stadler, T. How can we improve accuracy of macroevolutionary rate estimates? *Systematic Biology* **62**, 321–329 (2013).

[41] Stadler, T. & Steel, M. Swapping birth and death: Symmetries and transformations in phylodynamic models. *Systematic Biology* **68**, 852–858 (2019).

[42] Rabosky, D. L. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLOS ONE* **9**, 1–15 (2014).

[43] Stadler, T. Mammalian phylogeny reveals recent diversification rate shifts. *Proceedings of the National Academy of Sciences* **108**, 6187–6192 (2011).

[44] Akaike, H. Likelihood of a model and information criteria. *Journal of Econometrics* **16**, 3–14 (1981).

[45] Schwarz, G. Estimating the dimension of a model. *Annals of Statistics* **6**, 461–464 (1978).

[46] Steeman, M. E. *et al.* Radiation of extant cetaceans driven by restructuring of the oceans. *Systematic Biology* **58**, 573–585 (2009).

[47] Alfaro, M. E. *et al.* Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proceedings of the National Academy of Sciences* **106**, 13410–13414 (2009).

[48] Griffiths, R. C. & Tavare, S. Sampling theory for neutral alleles in a varying environment. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **344**, 403–410 (1994).

[49] Donnelly, P. & Tavare, S. Coalescents and genealogical structure under neutrality. *Annual Review of Genetics* **29**, 401–421 (1995).

[50] Heath, T. A., Huelsenbeck, J. P. & Stadler, T. The fossilized birth-death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences* **111**, E2957–E2966 (2014).

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

[52] Silvestro, D., Warnock, R. C. M., Gavryushkina, A. & Stadler, T. Closing the gap between palaeontological and neontological speciation and extinction rate estimates. *Nature Communications* **9**, 5237 (2018).

[53] Louca, S. & Doebeli, M. Efficient comparative phylogenetics on large trees. *Bioinformatics* **34**, 1053–1055 (2017).

[54] Smith, S. A. & Brown, J. W. Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany* **105**, 302–314 (2018).

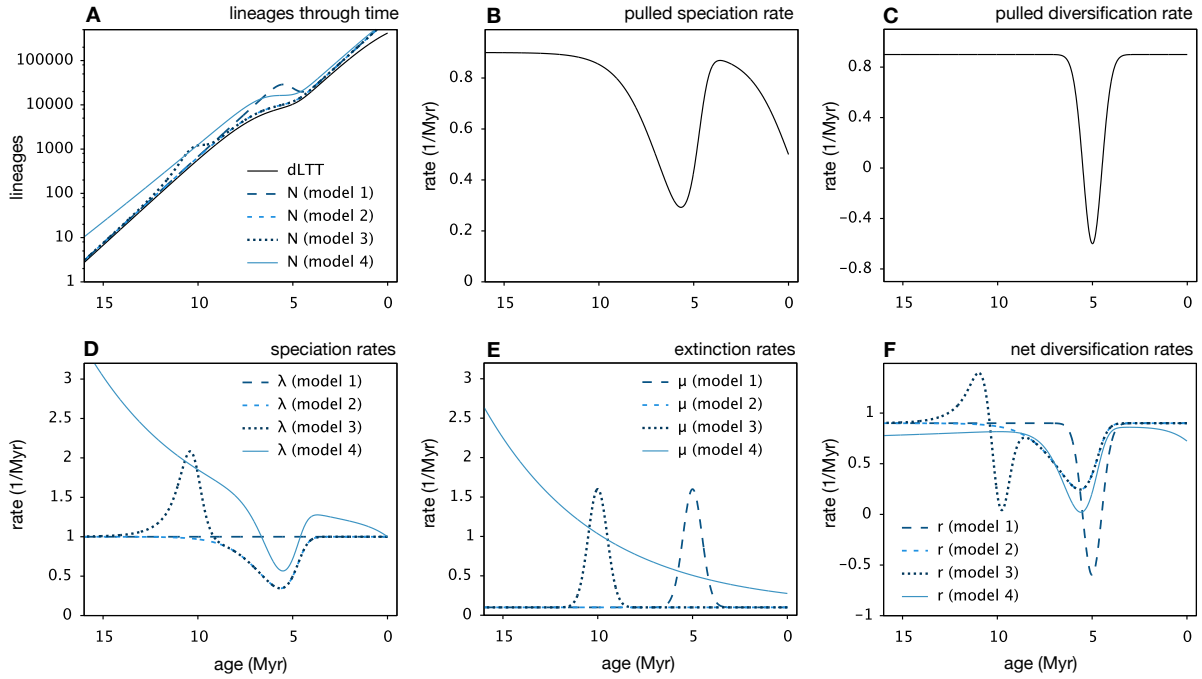

Figure 1: **Illustration of congruent birth-death processes (simulations).** Example of four hypothetical, congruent yet markedly different birth-death models. The first model exhibits a constant speciation rate and a sudden mass extinction event about 5 Myr before present; the second model exhibits 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.

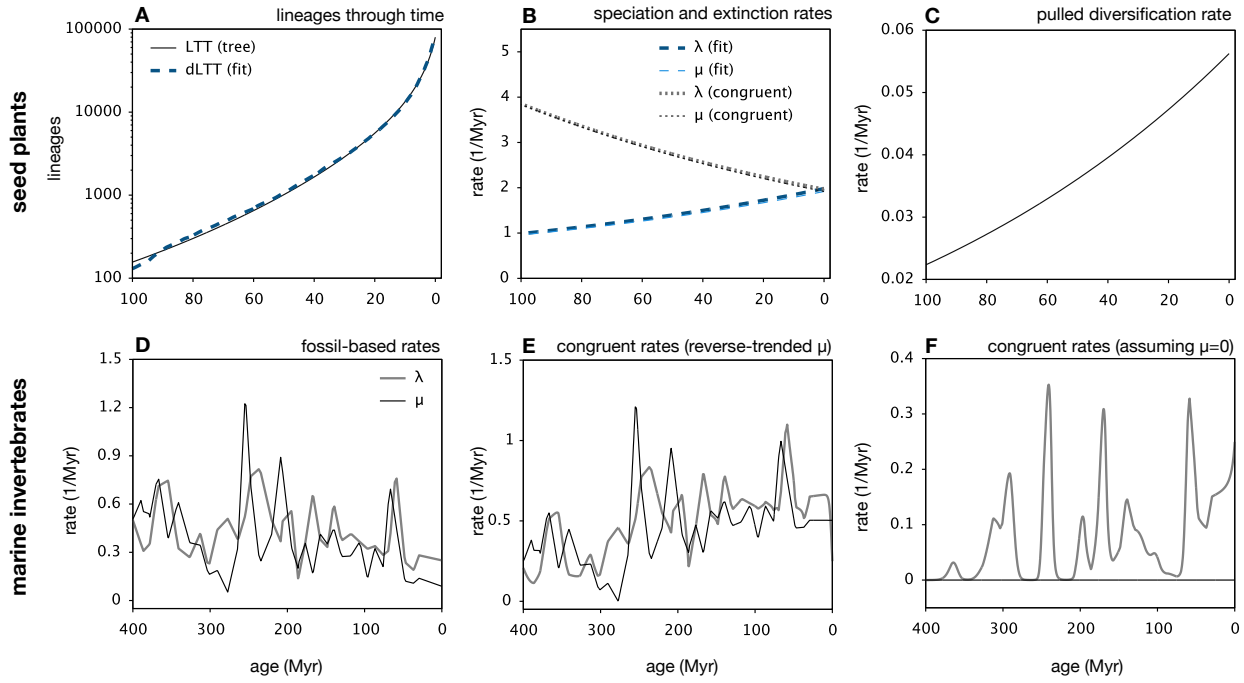

Figure 2: **Illustration of congruent birth-death processes (real data).** Top row: Birth-death model with exponentially varying λ and μ, fitted to a reconstruct timetree of 79,874 seed plant species (54) over the past 100 Myr, compared to a congruent model obtained by simply modifying the exponential 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. Bottom row: (D) Origination and extinction rates of marine invertebrate genera, estimated from fossil data by Alroy (36). (E) Congruent scenario to D, after reversing the linear trend of μ. (F) Congruent scenario to D, assuming zero extinction rate.

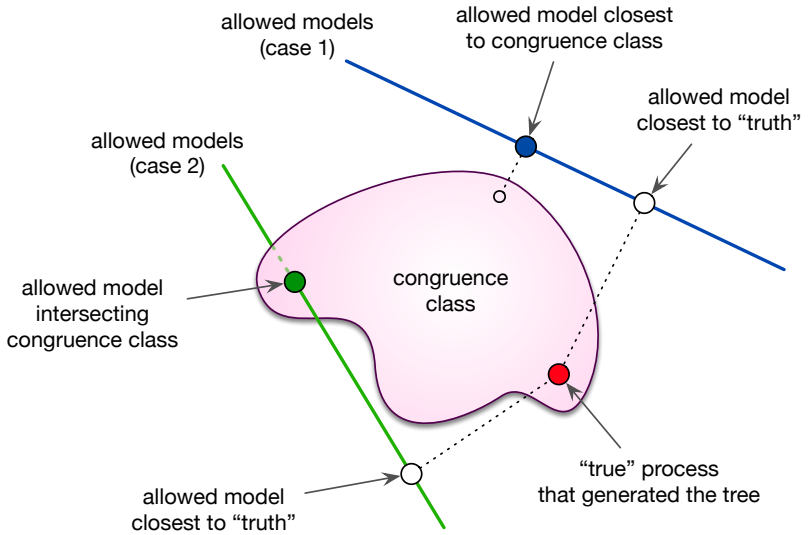

Figure 3: **Conceptual implications for reconstructing diversification history.** Conceptual illustration of the limited identifiability of a diversification process, assumed to be adequately described by some unknown birth-death model (red circle, henceforth "true process"). The congruence class of the true process is shown as a sub-space comprising a continuum of alternative models (pink area). In practice, maximum-likelihood model selection is performed among a parameterized low-dimensional set of allowed models, the precise nature of which can vary from case to case, for example depending on assumed functional forms for λ and μ or the number of allowed rate shifts (43). The two continuous lines shown here represent two alternative cases of allowed model sets (e.g., considered in two alternative studies), from within each the model closest to the truth is (ideally) sought. In each case, however, likelihood-based model selection will converge towards the allowed model closest to the congruence class (blue and green filled circles), which in general is not the allowed model actually closest to the true process (white circles). This identifiability issue persists even for infinitely large datasets.

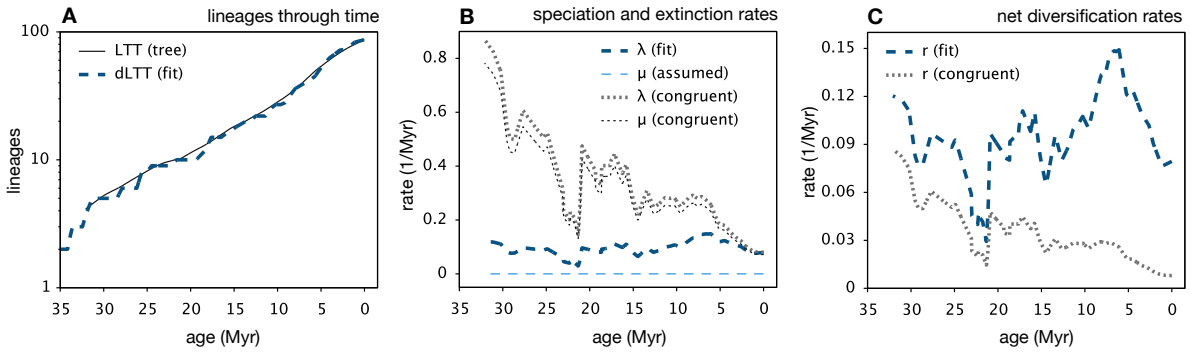

Figure 4: **Previous studies have likely over-interpreted phylogenetic data.** Time-dependent birth-death model fitted to a nearly-complete Cetacean timetree by Steeman *et al.* (46) under the assumption of zero extinction rates ($\mu = 0$), compared to a congruent model where the extinction rate is 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.* (46) 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*

Contents

S.1 Mathematical derivations

In the following, we provide mathematical derivations for various claims made in the main article. Some parts can be found in previous literature (1, 2, 3, 4, 5, 6), but are included here for completeness.

S.1.1 General considerations

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

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

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

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

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

$$M(\tau) = N(\tau) \cdot (1 - E(\tau)), \tag{3}$$

where $E(\tau)$ is the probability that a lineage extant at age τ will be missing from the timetree (either due to extinction or not having been sampled). As explained by Morlon *et al.* (5), the extinction probability E satisfies the differential equation:

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

We mention that the solution to Eq. (4) is provided by Morlon *et al.* (5, Eq. 2). 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 replace $dN/d\tau$ quickly leads to the differential equation:

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

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

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

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

Defining the relative slope of the dLTT:

$$\lambda_{\mathrm{p}} := -\frac{1}{M}\frac{dM}{d\tau} \tag{7}$$

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

$$\lambda_{\mathrm{p}} = \lambda \cdot (1 - E). \tag{8}$$

We note that $P(\tau) := 1 - E(\tau)$ is the probability that a lineage extant at age τ is represented in the extant timetree. P can thus be interpreted as a generalization of the present-day sampling fraction ρ to previous times. In fact, trimming a timetree at some age $\tau_o > 0$ (i.e., omitting anything younger than τ_o) would yield a new (shorter) timetree, whose tips are a random subset of the lineages that existed at age τ_1, each included at probability $P(\tau_o)$.

As becomes clear in Eq. (8), in the absence of extinction and if $\rho = 1$, the relative slope λ_{p} becomes equal to the speciation rate λ; in the presence of extinction λ_{p} is artificially pulled downwards relative to λ towards the past. Reciprocally, under incomplete sampling λ is artificially pulled downwards near the present. We shall therefore henceforth call λ_{p} the "pulled speciation rate".

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

$$\frac{d\lambda_{\mathrm{p}}}{d\tau} = \lambda_{\mathrm{p}} \cdot \left[\frac{1}{\lambda}\frac{d\lambda}{d\tau} - \mu + \lambda E\right] = \lambda_{\mathrm{p}} \cdot \left[\frac{1}{\lambda}\frac{d\lambda}{d\tau} + \lambda - \mu - \lambda \cdot (1 - E)\right] = \lambda_{\mathrm{p}} \cdot (r_{\mathrm{p}} - \lambda_{\mathrm{p}}), \tag{9}$$

where we defined the "pulled diversification rate":

$$r_{\mathrm{p}} := \lambda - \mu + \frac{1}{\lambda}\frac{d\lambda}{d\tau}. \tag{10}$$

Rearranging terms in Eq. (9) yields:

$$r_{\mathrm{p}} = \lambda_{\mathrm{p}} + \frac{1}{\lambda_{\mathrm{p}}}\frac{d\lambda_{\mathrm{p}}}{d\tau}, \tag{11}$$

which shows that r_{p} can be directly calculated from the dLTT.

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

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

Our starting point is the likelihood formula described by Morlon *et al.* (5):

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

where n is the number of branching points (internal nodes), τ_o is the age of the stem, $\tau_1 > \tau_2 > .. > \tau_n$ are the ages (time before present) of the branching points, $s_{i,1}$, $s_{i,2}$ are the ages at which the daughter lineages originating at age τ_i themselves branch (or end at a tip), ρ is the tree's sampling fraction (fraction of present-

day extant species included in the tree), $E(\tau)$ is the probability that a single lineage that existed at age τ would survive to the present and be represented in the tree (5, Eq. 2 therein), Ψ is defined as:

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

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

$$R(\tau) := \int_0^\tau du\ \left[\lambda(u)-\mu(u)\right]. \tag{14}$$

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

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

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

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

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

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

$$\lambda - \mu + \frac{d\ln\lambda}{d\tau} = \frac{d\ln\lambda_{\mathrm{p}}}{d\tau} - \frac{d\ln M}{d\tau}. \tag{17}$$

Integrating both sides of Eq. (17) yields:

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

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

$$e^{R(\tau)}\frac{\lambda(\tau)}{\lambda_o} = \frac{\lambda_{\mathrm{p}}(\tau)M_o}{\lambda_{\mathrm{p}}(0)M(\tau)}. \tag{19}$$

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

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

Recall that $\rho\lambda_o = \lambda_{\mathrm{p}}(0)$ according to Eq. (8), so that Eq. (20) can be written as:

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

Note that:

$$\frac{\lambda_{\mathrm{p}}}{M} = \frac{d}{d\tau}\frac{1}{M}. \tag{22}$$

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

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

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

$$L = \frac{1}{\left[1 - E(\tau_o)\right]\lambda(\tau_o)} \cdot \frac{\lambda_{\mathrm{p}}(\tau_o)M(\tau_o)}{M_o^{n+1}} \prod_{i=1}^{n} \lambda_{\mathrm{p}}(\tau_i)M(\tau_i). \tag{24}$$

Recall that $(1 - E)\,\lambda = \lambda_{\mathrm{p}}$ according to Eq. (8), which when inserted into (24) yields:

$$L = \frac{M(\tau_o)}{M_o^{n+1}} \prod_{i=1}^{n} \lambda_{\mathrm{p}}(\tau_i)M(\tau_i). \tag{25}$$

Since $\lambda_{\mathrm{p}}M = -dM/d\tau$, Eq. (25) becomes:

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

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

Note that the likelihood in Eq. (12) or equivalently Eq. (26) is conditioned upon the survival of the stem lineage, assuming that the stem age is known. If the stem age is unknown the likelihood should be conditioned upon the splitting at the root and the survival of the root's two daughter-lineages, as follows:

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

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

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

and

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

□

S.1.3 The likelihood in terms of λ_{p}

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

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

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

where M is the dLTT and $M_o := M(0)$. From Eq. (7) it is easy to obtain the following relationship between M and λ_{p}:

$$M(\tau) = M_o e^{-\Lambda_{\mathrm{p}}(\tau)}, \tag{31}$$

where we defined:

$$\Lambda_{\mathrm{p}}(\tau) := \int_0^{\tau} ds\ \lambda_{\mathrm{p}}(s). \tag{32}$$

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

$$L = e^{-\Lambda_{\mathrm{p}}(\tau_o)} \prod_{i=1}^{n} \underbrace{\frac{-1}{M(\tau_i)} \frac{dM}{d\tau}\bigg|_{\tau_i}}_{\lambda_{\mathrm{p}}(\tau_i)} \cdot e^{-\Lambda_{\mathrm{p}}(\tau_i)}, \tag{33}$$

and hence:

$$\boxed{L = e^{-\Lambda_{\mathrm{p}}(\tau_o)} \prod_{i=1}^{n} \lambda_{\mathrm{p}}(\tau_i) \cdot e^{-\Lambda_{\mathrm{p}}(\tau_i)}.} \tag{34}$$

If 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 (likelihood formula in Eq. (29)), we instead obtain the expression:

$$\boxed{L_r = \frac{e^{-\Lambda_{\mathrm{p}}(\tau_1)}}{\lambda_{\mathrm{p}}(\tau_1)} \prod_{i=1}^{n} \lambda_{\mathrm{p}}(\tau_i) \cdot e^{-\Lambda_{\mathrm{p}}(\tau_i)}.} \tag{35}$$

S.1.4 Calculating λ from r_{p} and μ

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

$$\frac{d\lambda}{d\tau} = \lambda \cdot (r_{\mathrm{p}} + \mu^* - \lambda), \tag{36}$$

with initial condition:

$$\lambda(0) = \eta_o/\rho > 0. \tag{37}$$

We assume that $r_{\rm p}$ and μ^* are sufficiently "well-behaved", specifically that they are integrable over any finite interval. Observe that Eq. (36) is an example of a Bernoulli-type differential equation, as it can be written in the standard form:

$$\frac{d\lambda}{d\tau} = p(\tau)\lambda(\tau) + q(\tau)\lambda^{\alpha}(\tau), \tag{38}$$

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

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

where we defined:

$$\Lambda(\tau) := \int_0^{\tau} ds\ \left[r_{\rm p}(s) + \mu^*(s)\right]. \tag{40}$$

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

For future reference, we mention that the above solution can be easily generalized to the case where the "initial condition" for λ is given at some arbitrary age τ_1, rather than at present-day. Specifically, the solution to the differential equation:

$$\frac{d\lambda}{d\tau} = \lambda \cdot \left(r_{\rm p} + \mu^* - \lambda\right), \tag{41}$$

with condition:

$$\lambda(\tau_1) = \lambda_1, \tag{42}$$

is given by:

$$\lambda(\tau) = \frac{\lambda_1 e^{\Lambda(\tau)}}{e^{\Lambda(\tau_1)} + \lambda_1 \int_0^{\tau} ds\, e^{\Lambda(s)} - \lambda_1 \int_0^{\tau_1} ds\, e^{\Lambda(s)}}. \tag{43}$$

Special cases:

- In the special case where $r_{\rm p}$ and μ^* are time-independent and $r_{\rm p} + \mu^* \neq 0$, the solution in Eq. (39) takes the form:

$$\lambda(\tau) = \frac{P}{(P\rho/\eta_o - 1)e^{-P\tau} + 1}, \tag{44}$$

  where $P = r_{\rm p} + \mu^*$.

- If and only if $\mu^*(\tau) = \eta_o/\rho - r_{\rm p}(\tau)$, the solution in Eq. (39) is time-independent:

$$\lambda(\tau) = \frac{\eta_o}{\rho}. \tag{45}$$

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

S.1.5 Calculating λ from r_{p} and ε

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

$$\frac{d\lambda}{d\tau} = \lambda \cdot [r_{\mathrm{p}} + (\varepsilon - 1)\lambda]\,. \tag{46}$$

We assume that r_{p} and ε are sufficiently "well-behaved", specifically that they are integrable over any finite interval. Observe that Eq. (46) is an example of a Bernoulli-type differential equation, as it can be written in the standard form:

$$\frac{d\lambda}{d\tau} = p(\tau)\lambda(\tau) + q(\tau)\lambda^{\alpha}(\tau), \tag{47}$$

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

$$\boxed{\lambda(\tau) = \frac{\lambda_o e^{R_{\mathrm{p}}(\tau)}}{1 + (1-\varepsilon)\cdot\lambda_o \int_0^\tau ds\; e^{R_{\mathrm{p}}(s)}},} \tag{48}$$

where we defined:

$$R_{\mathrm{p}}(\tau) := \int_0^\tau ds\; r_{\mathrm{p}}(s). \tag{49}$$

In the special case where r_{p} is time-independent and non-zero, the solution in Eq. (48) simplifies to:

$$\lambda(\tau) = \frac{\lambda_o e^{r_{\mathrm{p}}\tau}}{1 + (1-\varepsilon)\cdot\dfrac{\lambda_o}{r_{\mathrm{p}}}\left(e^{r_{\mathrm{p}}s} - 1\right)}. \tag{50}$$

S.1.6 The likelihood in terms of the r_{p}

In the following we show how the likelihood of a tree under a birth-death model can be expressed solely in terms of the model's pulled diversification rate r_{p} and the product $\rho\lambda_o$. We first consider the case where the stem age is known and the likelihood is conditioned on the survival of the stem lineage (5); the alternative case where only the crown age is known and the likelihood is conditioned upon the survival of the root's two daughter lineages (Eq. 28) can be treated similarly and is briefly mentioned at the end.

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

$$R_{\mathrm{p}}(\tau) := \int_0^\tau du\; r_{\mathrm{p}}(u). \tag{51}$$

Then from the definition of r_{p} (Eq. 1 in the main article) we have:

$$R_{\mathrm{p}}(\tau) = \int_0^\tau du\ [\lambda(u) - \mu(u)] + \int_0^\tau du\ \frac{d\ln\lambda}{du} = R(\tau) + \ln\frac{\lambda(\tau)}{\lambda_o}. \tag{52}$$

Exponentiating (52) and rearranging yields:

$$e^{R(\tau)} = e^{R_{\mathrm{p}}(\tau)}\frac{\lambda_o}{\lambda(\tau)}. \tag{53}$$

Inserting Eq. (53) into the definition of Ψ in Eq. (13) yields:

$$\Psi(0,\tau) = e^{R_{\mathrm{p}}(\tau)}\frac{\lambda_o}{\lambda(\tau)}\left[1 + \rho\lambda_o\int_0^\tau du\ e^{R_{\mathrm{p}}(u)}\right]^{-2}. \tag{54}$$

Inserting Eq. (54) into the likelihood formula (16) yields:

$$L = \frac{(\rho\lambda_o)^{n+1}e^{R_{\mathrm{p}}(\tau_o)}}{[1-E(\tau_o)]\,\lambda(\tau_o)}\left[1 + \rho\lambda_o\int_0^{\tau_o} du\ e^{R_{\mathrm{p}}(u)}\right]^{-2}\prod_{i=1}^{n} e^{R_{\mathrm{p}}(\tau)}\left[1 + \rho\lambda_o\int_0^{\tau} du\ e^{R_{\mathrm{p}}(u)}\right]^{-2}. \tag{55}$$

Recall that $(1-E)\lambda = \lambda_{\mathrm{p}}$ according to Eq. (8), which when inserted into Eq. (55) yields:

$$L = \frac{(\rho\lambda_o)^{n+1}e^{R_{\mathrm{p}}(\tau_o)}}{\lambda_{\mathrm{p}}(\tau_o)}\left[1 + \rho\lambda_o\int_0^{\tau_o} du\ e^{R_{\mathrm{p}}(u)}\right]^{-2}\prod_{i=1}^{n} e^{R_{\mathrm{p}}(\tau)}\left[1 + \rho\lambda_o\int_0^{\tau} du\ e^{R_{\mathrm{p}}(u)}\right]^{-2}. \tag{56}$$

From Eqs. (8) and (11) we know that λ_{p} satisfies the initial value problem (Bernoulli differential equation):

$$\frac{d\lambda_{\mathrm{p}}}{d\tau} = \lambda_{\mathrm{p}}\cdot(r_{\mathrm{p}} - \lambda_{\mathrm{p}}), \quad \lambda_{\mathrm{p}}(0) = \rho\lambda_o. \tag{57}$$

It is straightforward to verify that the solution to Eq. (57) is given by:

$$\lambda_{\mathrm{p}}(\tau) = \frac{\rho\lambda_o e^{R_{\mathrm{p}}(\tau)}}{1 + \rho\lambda_o\int_0^\tau e^{R_{\mathrm{p}}(u)}du}. \tag{58}$$

Inserting the solution (58) into Eq. (56) yields the following expression for the likelihood:

$$\boxed{L = \left[1 + \rho\lambda_o\int_0^{\tau_o} du\ e^{R_{\mathrm{p}}(u)}\right]^{-1}(\rho\lambda_o)^n\prod_{i=1}^{n} e^{R_{\mathrm{p}}(\tau)}\left[1 + \rho\lambda_o\int_0^{\tau} du\ e^{R_{\mathrm{p}}(u)}\right]^{-2}.} \tag{59}$$

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:

$$\boxed{L_r = e^{-R_{\mathrm{p}}(\tau_o)}(\rho\lambda_o)^{n-1}\prod_{i=1}^{n} e^{R_{\mathrm{p}}(\tau)}\left[1 + \rho\lambda_o\int_0^{\tau} du\ e^{R_{\mathrm{p}}(u)}\right]^{-2}.} \tag{60}$$

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

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

Consider a birth-death process with parameters (λ, μ, ρ), starting from a single lineage at some time before present τ_o and ultimately resulting in a timetree at age 0, comprising only extant species that are included at some probability ρ. The probability that the timetree will comprise n tips can be expressed using formulas first derived by Kendall *et al.* (7):

$$\begin{aligned} P(n) &= (1 - E(\tau_o)) \cdot (1 - H) \cdot H^{n-1}, \quad n \geq 1 \\ P(0) &= E(\tau_o), \end{aligned} \tag{61}$$

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

$$H := \frac{\rho \int_0^{\tau_o} ds\, e^{R(s)} \lambda(s)}{1 + \rho \int_0^{\tau_o} ds\, e^{R(s)} \lambda(s)}, \tag{62}$$

and R was previously defined in Eq. (14). Note that the formula in Eq. (61) can be readily obtained using equations 8, 10b and 11 in (7), after setting the time variable therein equal to τ_o (i.e. $t = \tau_o$), switching from time to age ($\tau = \tau_o - t$), and adding the term $-\delta(\tau) \ln \rho$ to the extinction rate (where δ is the Dirac distribution, peaking at age 0) to account for incomplete species sampling. As shown previously in Eq. (53), we have

$$e^{R(\tau)} = e^{R_{\mathrm{p}}(\tau)} \frac{\lambda_o}{\lambda(\tau)}, \tag{63}$$

where R_{p} is defined as:

$$R_{\mathrm{p}}(\tau) := \int_0^{\tau} du\, r_{\mathrm{p}}(u), \tag{64}$$

and r_{p} is the pulled diversification rate. Inserting Eq. (63) into Eq. (62) allows us to write H as follows:

$$H = \frac{\rho\lambda_o \int_0^{\tau_o} ds\, e^{R_{\mathrm{p}}(s)}}{1 + \rho\lambda_o \int_0^{\tau_o} ds\, e^{R_{\mathrm{p}}(s)}}. \tag{65}$$

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 congruence class.

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

$$\boxed{P_{\mathrm{stem}}(n) = (1 - H) \cdot H^{n-1}.} \tag{66}$$

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

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 two daughter lineages, denoted $P_{\text{root}}(n)$, can be derived in a similar way, as follows. The probability that the two daughter lineages survive, conditional upon the split at age τ_o, is given by the product:

$$P(\text{daughter lineages survive} \mid \text{split at } \tau_o) = (1 - E(\tau_o))^2. \tag{67}$$

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

$$\begin{aligned}
&P(\text{daughter lineages survive and tree has size } n \mid \text{split at } \tau_o) \\
&= \sum_{k=1}^{n-1} P(k)P(n-k) \\
&= (1 - E(\tau_o))^2 \sum_{k=1}^{n-1} (1-H) \cdot H^{k-1} \cdot (1-H) \cdot H^{n-k-1} \\
&= (1 - E(\tau_o))^2 (1-H)^2 \sum_{k=1}^{n-1} H^{n-2} \\
&= (n-1) \cdot (1 - E(\tau_o))^2 (1-H)^2 H^{n-2}.
\end{aligned} \tag{68}$$

Dividing Eq. (68) by Eq. (67) yields the desired probability:

$$\boxed{P_{\text{root}}(n) = (n-1) \cdot (1-H)^2 H^{n-2}.} \tag{69}$$

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

□

S.1.8 The geometric nature of congruence classes

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

In geometry, two objects are called congruent if they exhibit similar geometric properties, such as identical angles between corresponding lines and identical distances between corresponding points. More precisely, two geometric objects (sets of points in Euclidean space $\mathbb{R}^n$) are called congruent if one set can be transformed into the other set by means of an isometry, i.e. a mapping that preserves distances between pairs of points (via translations, rotations, and/or reflections). Object congruency is a type of equivalence relation, and hence the set of models congruent to some focal object is an equivalence class. The set of all isometries is itself a group (known as "Euclidean group") that acts on the set of geometric objects, and congruence classes of objects correspond to "orbits" under the action of isometries (8).

By analogy, two birth-death models are called "congruent" if they exhibit similar statistical properties in terms of their generated extant timetrees and LTTs (see main text and Supplement S.1). In fact, congruence classes can be interpreted as the orbits of a group of mappings acting on model space that preserve dLTTs (just as isometries preserve distances in Euclidean space). For technical reasons, we shall henceforth only consider the space of birth-death models (denoted $\mathcal{B}$) with strictly positive λ, μ and ρ and continuously differentiable λ 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 real-valued strictly positive functions defined on the interval $[0, \tau_o]$. For any $S_o \in (0, \infty)$ and any $f \in$

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

$$\frac{dS[S_o,f]}{d\tau}=S[S_o,f](\tau)\cdot\left[f(\tau)-S[S_o,f](\tau)\right],\quad S[S_o,f](0)=S_o. \tag{70}$$

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

$$S[S_o,f](\tau)=\frac{S_o e^{F(\tau)}}{1+S_o\int_0^\tau ds\; e^{F(s)}}, \tag{71}$$

where we denoted:

$$F(\tau):=\int_0^\tau ds\; f(s). \tag{72}$$

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

$$g_{\alpha,\beta}(\lambda,\mu,\rho):=\left(S\left[\lambda/\alpha,\lambda-\mu+\frac{1}{\lambda}\frac{d\lambda}{d\tau}+\beta\mu\right],\beta\mu,\alpha\rho\right). \tag{73}$$

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

$$\lambda^*:=S\left[\lambda/\alpha,\lambda-\mu+\frac{1}{\lambda}\frac{d\lambda}{d\tau}+\beta\mu\right] \tag{74}$$

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

$$G:=\left\{g_{\alpha,\beta}:\alpha\in(0,\infty),\beta\in\mathcal{C}^1_+[0,\tau_o]\right\}, \tag{75}$$

constitutes a group with group operation:

$$g_{\alpha,\beta}\circ g_{\tilde\alpha,\tilde\beta}:=g_{\alpha\tilde\alpha,\beta\tilde\beta} \tag{76}$$

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

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

$$G(\lambda,\mu,\rho):=\{g(\lambda,\mu,\rho):g\in G\} \tag{77}$$

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

S.2 Why previous studies failed to detect model congruencies

In practice, reconstructions of λ and μ over time are typically performed by selecting among a limited set of allowed models, i.e., considering specific functional forms described by a finite number of parameters (9, 10, 11, 12, 13, 5). In these situations it is generally unlikely that the allowed model set intersects a given congruence class more than once (see Supplement S.3 for mathematical justification). For example, when considering only constant-rate birth-death models and assuming that ρ is fixed (as is usually the case; 14), each congruence class reduces to a single combination of λ and μ. Likelihood functions defined over a limited allowed model set thus generally don't exhibit ridges associated with congruence classes, and may even exhibit a unique global maximum in the space of considered parameters, leaving the impression that λ and μ have been estimated close to their true values. Our findings suggest that this impression is almost certainly false. Instead, obtained estimates for λ and μ are almost always going to be a random outcome that depends on the particular choice of allowed models, such as the functional forms considered for λ and μ, and will be as close as possible to the congruence class of the truth rather than close to the truth itself. Unless one has reasons to prefer specific functional forms for λ and μ (e.g., based on a mechanistic macroevolutionary model; 15), fitted λ and μ are unlikely to resemble the true rates even if in principle the functional forms considered are flexible enough to resemble the true λ and μ (see Supplement S.6 for examples using simulations and real data).

By analogy, studies that test whether diversification dynamics are influenced by some environmental or geological variable X (e.g., temperature), either by testing for correlations between X and the estimated λ or μ (16, 17) or by fitting models in which λ or μ are explicit functions of X (18, 19, 20), will generally lead to unreliable conclusions. Indeed, specifying λ or μ as functions of X (e.g., assuming $\mu = \alpha X + \beta$ and fitting the coefficients α and β) is essentially equivalent to choosing particular functional forms for λ or μ. Incidentally, evaluations of estimation methods based on simulations of the same limited model set as considered by the very estimation method evaluated (9, 11, 12, 13, 5), for example simulating trees with linearly varying λ and μ and then fitting models with linearly varying λ and μ to check if linear coefficients are accurately estimated, can generate the false impression that λ and μ can in general be reliably identified. Indeed, any given simulated model is typically going to be the sole representative of its congruence class that's also in the method's allowed model set.

S.3 Typical model sets do not exhibit congruence ridges

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

$$\lambda = \lambda(\tau, \mathbf{q}), \quad \mu = \mu(\tau, \mathbf{q}). \tag{78}$$

For simplicity, assume that the sampling fraction ρ is given (identifiability issues associated with uncertainties in ρ are already well known; 21, 22, 23, 24).

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

$$r_{\mathrm{p}}(\tau, \mathbf{q}) = \lambda(\tau, \mathbf{q}) - \mu(\tau, \mathbf{q}) + \frac{1}{\lambda(\tau, \mathbf{q})} \frac{\partial \lambda(\tau, \mathbf{q})}{\partial \tau}, \tag{79}$$

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

$$\frac{\partial \lambda(\tau, \mathbf{h})}{\partial \tau} = \lambda(\tau, \mathbf{h}) \cdot \left[r_{\mathrm{p}}(\tau, \mathbf{q}) - \lambda(\tau, \mathbf{h}) + \mu(\tau, \mathbf{h}) \right], \quad \lambda(0, \mathbf{h}) = \lambda(0, \mathbf{q}). \tag{80}$$

Unless the functional forms of λ and μ have been specifically designed for this purpose, it is generally unlikely that Eq. (80) will be satisfied for some $\mathbf{h} \neq \mathbf{q}$.

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

$$r_{\mathrm{p}}(\tau, \mathbf{h}(s)) = r_{\mathrm{p}}(\tau, \mathbf{q}), \tag{81}$$

and such that:

$$\lambda(0, \mathbf{h}(s)) = \lambda(0, \mathbf{q}), \tag{82}$$

for all $s \in [-\varepsilon, \varepsilon]$ and all $\tau \geq 0$. Taking the derivative of Eq. (81) with respect to s at 0 yields:

$$\sum_{i=1}^{n} \left. \frac{\partial r_{\mathrm{p}}}{\partial q_i} \right|_{(\tau, \mathbf{q})} \cdot \left. \frac{dh_i}{ds} \right|_{s=0} = 0. \tag{83}$$

Denote $\mathbf{H} := \left. \frac{d\mathbf{h}}{ds} \right|_{s=0}$ and $\mathbf{R}(\tau) := \left. \frac{\partial r_{\mathrm{p}}}{\partial \mathbf{q}} \right|_{\tau, \mathbf{q}}$. Then the condition in Eq. (83) can be written in vector notation:

$$\mathbf{R}(\tau)^{\mathrm{T}} \cdot \mathbf{H} = 0. \tag{84}$$

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 (84) means that the ridge must move perpendicular to the direction of steepest descent of r_{p}. Observe that condition (84) must be satisfied for all ages $\tau \geq 0$. Hence, for any arbitrary choice $\tau_1, \tau_2, .., \tau_m \geq 0$, we obtain the following m linear equations that must be satisfied by $\mathbf{H}$:

$$\begin{aligned} &\mathbf{R}(\tau_1)^{\mathrm{T}} \cdot \mathbf{H} = 0. \\ &\vdots \\ &\mathbf{R}(\tau_m)^{\mathrm{T}} \cdot \mathbf{H} = 0. \end{aligned} \tag{85}$$

Unless the functional forms of λ and μ are specifically designed for this purpose, the system in Eq. (85) will almost certainly be over-determined if m is chosen sufficiently high ($m \gg n$). Hence, in practice, for a chosen set of allowed models and a given point $\mathbf{q}$ in parameter space, a congruence ridge will almost never exist at that point.

□

S.4 Interpreting the PDR

Given that the PDR is a composite quantity that depends on both λ and μ (Eq. 1), properly interpreting the estimated PDR in terms of actual speciation/extinction rates remains the responsibility of the investigator. Previous work has shown that the PDR can indeed yield valuable insight into diversification dynamics and can be useful for testing alternative hypotheses (6). For example, sudden rate transitions (e.g., due to mass extinction events) almost always lead to fluctuations in the PDR; thus, a relatively constant PDR over time would be indicative of constant — or only slowly changing — speciation and extinction rates.

The PDR can be used to obtain other useful variables. For example, it is straightforward to confirm that the PDR and the total diversity N satisfy the following relationship:

$$\frac{N(\tau)}{N(0)} \cdot \frac{\lambda_o}{\lambda(\tau)} = \exp\left[-\int_0^\tau du\, r_{\mathrm{p}}(u)\right]. \tag{86}$$

Observe that the left hand side of this equation, henceforth called deterministic "pulled normalized diversity" (dPND), corresponds to the ratio of deterministic total diversity at some age τ over the assumed present-day total diversity $N(0)$, modulated by the factor $\lambda_o/\lambda(\tau)$. Like the PDR, the dPND is the same for all models in a congruence class, and can be readily estimated from extant timetrees (Fig. S4C). As becomes apparent from Eq. (86), while the dPND can yield information on variations of past diversity, the amount of information depends on how well λ can be constrained a priori.

Another useful derived variable is the "pulled extinction rate", or PER (6), defined as:

$$\mu_{\mathrm{p}} := \lambda_o - r_{\mathrm{p}}. \tag{87}$$

The PER is equal to the extinction rate μ if λ is time-independent, but differs from μ in most other cases. Note that calculating the PER requires knowing the present-day speciation rate λ_o, which can be estimated from the timetree if the sampling fraction ρ is known (simply divide the estimated $\rho\lambda_o$ by ρ). The present-day PER is related to the present-day extinction rate as follows:

$$\mu_{\mathrm{p}}(0) = \mu(0) - \frac{1}{\lambda_o}\frac{d\lambda}{d\tau}\bigg|_{\tau=0}. \tag{88}$$

Observe that if the present-day speciation rate changes only slowly, the present-day PER will resemble the present-day extinction rate. Further, since $\mu(0)$ is non-negative, we can obtain the following lower bound for the exponential rate at which λ changes:

$$\frac{1}{\lambda_o}\frac{d\lambda}{d\tau}\bigg|_{\tau=0} \geq -\mu_{\mathrm{p}}(0). \tag{89}$$

In particular, if the estimated $\mu_{\mathrm{p}}(0)$ is negative, this is evidence that λ is currently decreasing over time.

S.5 Fitting congruence classes instead of models

The discussion in the main article revealed that speciation and extinction rates constitute partly interchangeable (and thus partly redundant) parameters that cannot be completely resolved from extant timetrees alone, no matter how large the dataset. Extant timetrees do, however, contain the proper information to estimate 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

congruence class from which a tree was likely generated. In fact, for sufficiently large trees r_{p} and η_o can be directly calculated from the slope and curvature of the tree's LTT (6). Since each congruence 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 classes; on this space the likelihood function no longer exhibits the highly problematic ridges seen in the original model space. We thus suggest describing birth-death models in terms of r_{p} and η_o, rather than λ and μ, when fitting models to timetrees. Since the likelihood function can be expressed directly in terms of r_{p} and η_o (Supplement S.1.6), such a parameterization is suitable for maximum-likelihood or Bayesian estimation methods. Reciprocally, since every given r_{p} and η_o correspond to a unique and non-empty congruence class (as shown in the main article), any r_{p} and η_o estimated from an extant timetree will represent at least one biologically meaningful scenario. It is thus possible to directly fit congruence classes, rather than individual models, via maximum-likelihood. A similar reasoning can also be applied to the pulled speciation rate λ_{p}, which provides an alternative representation of congruence classes.

To demonstrate this approach, we created software for fitting r_{p} and η_o to extant timetrees via maximum likelihood. The code is integrated into the R package `castor` (25) as function `fit_hbd_pdr_on_grid`. The function accepts as input an extant timetree, and an arbitrary number of discrete ages at which to estimate r_{p}, assuming r_{p} varies linearly or polynomially between those ages. The function also accepts optional lower and upper bounds for the fitted r_{p} and/or η_o. The code then maximizes the likelihood of the tree, given by Eq. (56) in the Supplement, by iteratively refining the r_{p} values on the age grid and/or η_o. Optionally, one can limit the evaluation of the likelihood function to a smaller "truncated" age interval than covered by the tree, i.e. some age interval $[0, \tau^*]$, where τ^* may be smaller than the root age. This may be useful for avoid estimation errors towards older ages due to a small number of lineages in the tree. The likelihood formula for the "truncated" case can be easily obtained by assuming that the tree is split into multiple sub-trees, each originating at the truncation age, and considering each sub-tree an independent realization of the same birth-death process and subject to the same sampling fraction ρ. To avoid non-global local optima, the fitting can be repeated multiple times, each time starting at random start values for the fitted parameters, and the best fit among all repeats is kept. We also developed similar computer code for fitting the pulled speciation rate λ_{p} to extant timetrees, implemented as function `fit_hbd_psr_on_grid` in the R package `castor`.

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

S.6 Fitting birth-death models to trees yields unreliable results

To illustrate the identifiability issues discussed in the main article, we simulated and analyzed two massive extant timetrees ($\sim$114,000 and $\sim$785,000 tips) via a birth-death process, subject to a mass extinction event (both trees) and a rapid radiation event (second tree). Instead of fitting models of the exact same functional form as used in the simulations, we fitted generic piecewise-linear curves for λ and μ that could in principle take various alternative shapes (including approximately the shapes used for the simulations), and visually compared the estimated profiles to their true profiles (Supplemental Figs. S5A–F). Specifically, we fitted λ and μ at multiple discrete time points, treating the rates at each time point as free parameters, while assuming a known ρ. Despite the enormous tree sizes, and despite the fact that the fitted models reproduced the trees' LTTs and the true r_{p} extremely well (Supplemental Figs. S5A,C,D,F), the estimated λ and μ were far from their true values and even exhibited spurious trends (Supplemental Figs. S5BE). This is consistent with our expectation that the particular combination of fitted λ and μ is essentially a random pick from the periphery

of the true process's congruence class. In contrast, when we fixed μ to its true profile, λ was accurately estimated (Supplemental Fig. S2), consistent with the expectation that any given μ and ρ fully determine the corresponding λ in the congruence class.

We also examined a large extant timetree of 79,874 seed plant species (Supplemental Fig. S3) published by Smith *et al.* (26), and estimated λ and μ over the last 100 million years using two alternative approaches (methods details in Supplement S.7). In the first approach, we fitted generic piecewise-linear curves for λ and μ, similarly to the previous example. In the second approach, we fitted parameterized time curves for λ and μ that included an exponential as well as a polynomial term (5). Even though both approaches yielded similar estimates for r_{p}, and both accurately reproduced the tree's LTT, they yielded markedly different λ and μ (Supplemental Figs. S5D–F). This observation is consistent with our argument that, depending on the precise set of models considered, the estimated λ and μ will generally be a random pick from the underlying (true or close-to-true) congruence class.

To illustrate our point that common model selection approaches such as minimizing the Akaike Information Criterion (AIC) (27) cannot resolve the identifiability issues discussed, we also fitted a series of models of variable complexity to a massive timetree of 1,000,000 tips. The tree was simulated based on origination and extinction rates of marine invertebrate genera, previously estimated from marine invertebrate fossil data (28) (Fig. 2D in the main article). We fitted two types of models: piecewise constant models and piecewise linear models. In piecewise constant models (sometimes also referred to as "birth-death-shift" models; 13) the rates λ and μ have constant values within discrete time intervals, with every time interval exhibiting distinct λ and μ. In piecewise linear models λ and μ vary linearly between discrete time points. For both model types we considered various temporal grid sizes, ranging from 5 up to 15 grid points, thus including sufficient model complexity for approximating the true rates. In all cases the time grid points where located at equidistant intervals between the present and the tree's root. For each model type (piecewise constant or piecewise linear) and for each grid size we estimated the free parameters (either the rates within each interval, or the rates at each grid point, respectively) via maximum likelihood using the function `fit_hbd_model_on_grid` in the R package `castor`. Only the most recent 100 million years were considered for fitting, in order to focus estimations on times with greater lineage density in the simulated tree (towards the root estimated rates will be inaccurate regardless of the arguments presented in this paper). Fitting was repeated 20 times with random start parameters to avoid local non-global optima. Among each model type, we then kept the maximum likelihood model with smallest AIC value, shown in Supplemental Fig. S6. As expected, estimated rates were highly inaccurate and missed important features, despite the fact that we were using a massive tree of 1,000,000 tips, and the fact that the tree's LTT was almost perfectly matched by the models' dLTTs.

S.7 Fitting birth-death models to seed plants

An extant timetree of 79,874 seed plant species, constructed using GenBank sequence data with a backbone provided by Magallón *et al.* (29), was obtained from the Supplemental Material published by Smith *et al.* (26, tree "CBMB"). The tree is shown in Supplemental Fig. S3. The sampling fraction was calculated based on the tree size and the number of extant seed plant species estimated at 422,127 by Govaerts (30). As mentioned in Supplement S.6, two approaches were used to fit λ and μ over time. In the first approach, λ and μ were allowed to vary independently at 8 discrete and equidistant time points (assuming piecewise linearity between grid points) and were estimated via maximum-likelihood using the function `fit_hb_model_on_grid` in the R package `castor` (25) (options "`condition='stem', relative_dt=0.001`"). Fitting was repeated 100 times using random start parameters to avoid local non-global optima in the likelihood function. The PDR was then estimated from the fitted λ and μ using the formula in Eq. (1) and using central finite differences to

calculate derivatives on the time grid. In the second approach, λ and μ were assumed to be of the following general functional forms:

$$\lambda(\tau) = \max\left(0, p_1 \cdot e^{-p_2 \cdot \tau/\tau_r} + p_3 + p_4 \cdot \frac{\tau}{\tau_r} + p_5 \cdot \left(\frac{\tau}{\tau_r}\right)^2 + p_6 \cdot \left(\frac{\tau}{\tau_r}\right)^3 + p_7 \cdot \left(\frac{\tau}{\tau_r}\right)^4\right) \quad (90)$$

$$\mu(\tau) = \max\left(0, q_1 \cdot e^{-q_2 \cdot \tau/\tau_r} + q_3 + q_4 \cdot \frac{\tau}{\tau_r} + q_5 \cdot \left(\frac{\tau}{\tau_r}\right)^2 + q_6 \cdot \left(\frac{\tau}{\tau_r}\right)^3 + q_7 \cdot \left(\frac{\tau}{\tau_r}\right)^4\right), \quad (91)$$

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

S.8 Supplemental figures

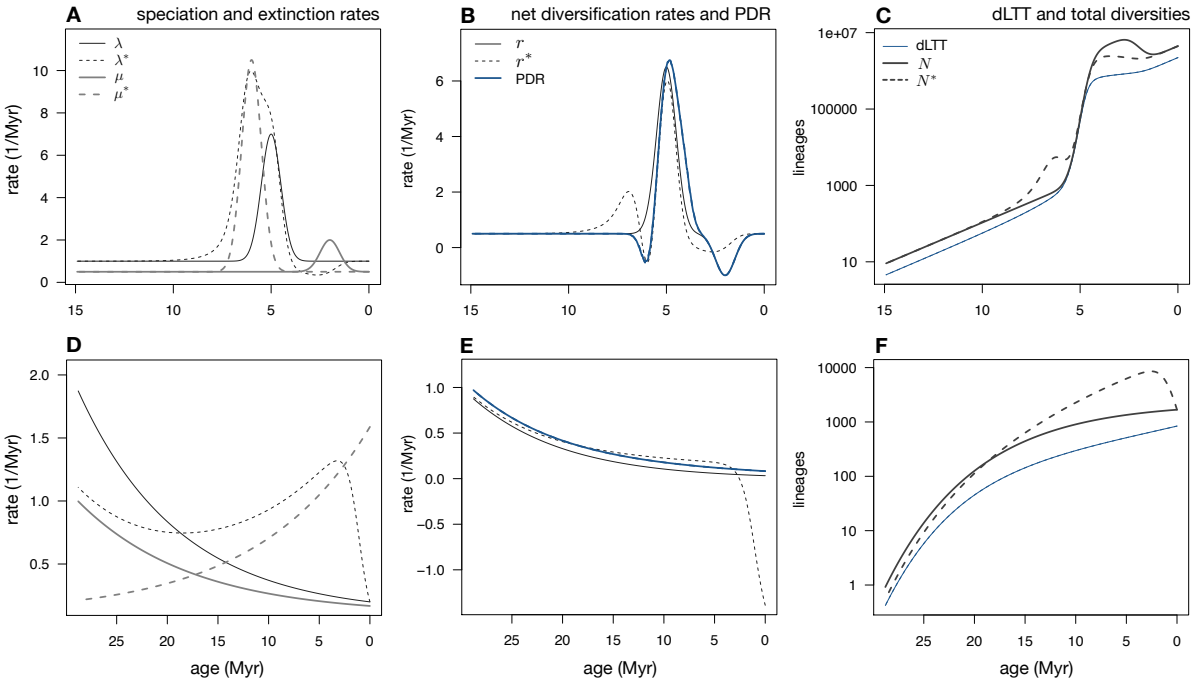

Figure S1: **Examples of congruent birth-death processes.** (A–C) Example of two congruent yet markedly different birth-death models. Both models exhibit a temporary spike in the extinction rate and a temporary spike in the speciation rate, however the timings of these events differ substantially between 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 $\rho = 0.5$.

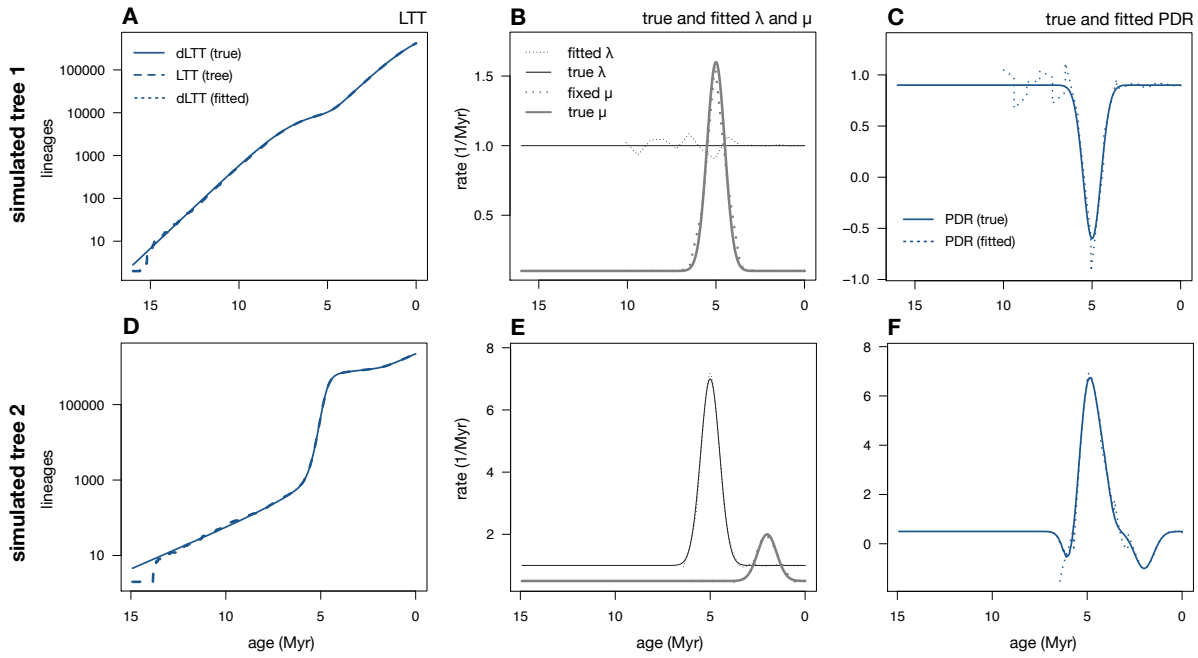

Figure S2: **Estimating λ when μ and ρ are fixed.** (A–C) Example analysis of a simulated extant timetree (∼114,000 tips), exhibiting a mass extinction event at ∼5 Myr before present. A birth-death model was fitted while fixing μ and ρ to their true values; λ was fitted at 15 discrete time points. (A) 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 (∼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.

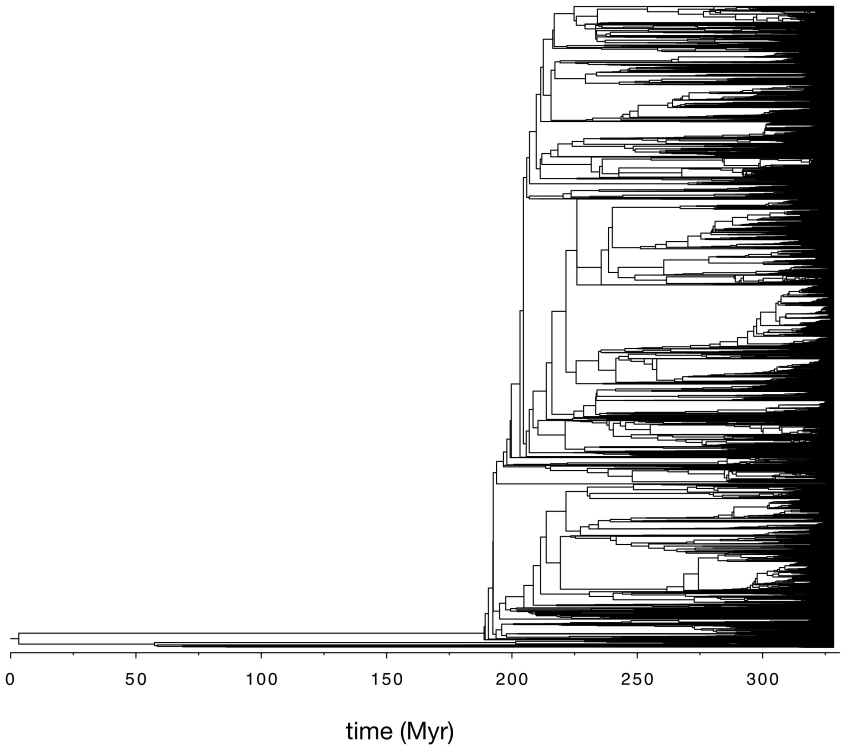

Figure S3: **Seed plant tree.** Extant timetree of 79,874 seed plant species, discussed in the main article. The tree was constructed and made available by Smith *et al.* (26) (tree "GBMB").

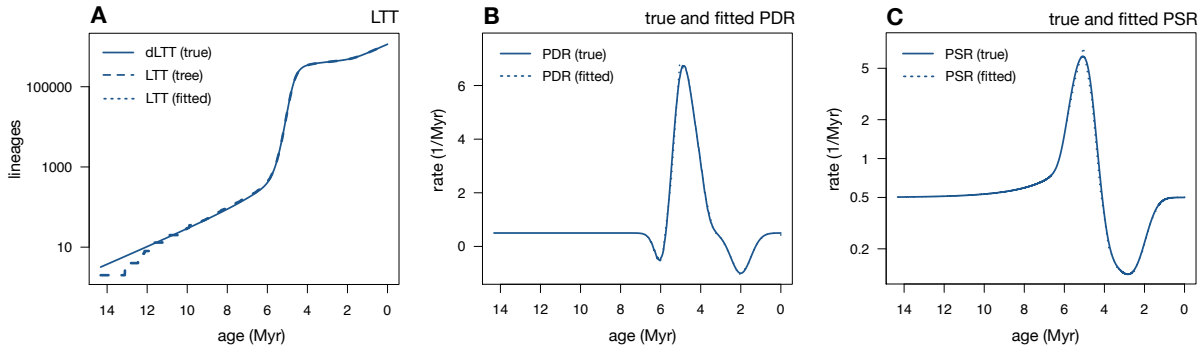

Figure S4: **Fitting congruence classes instead of models.** Analysis of an extant timetree generated by a birth-death model, exhibiting a temporary rapid radiation event about 5 Myr before present and a mass extinction event about 2 Myr before present. A congruence class was fitted to the timetree either 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 (LTT) 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` (25), allowing the PDR or PSR to vary freely at 15 discrete equidistant time points (Supplement S.5).

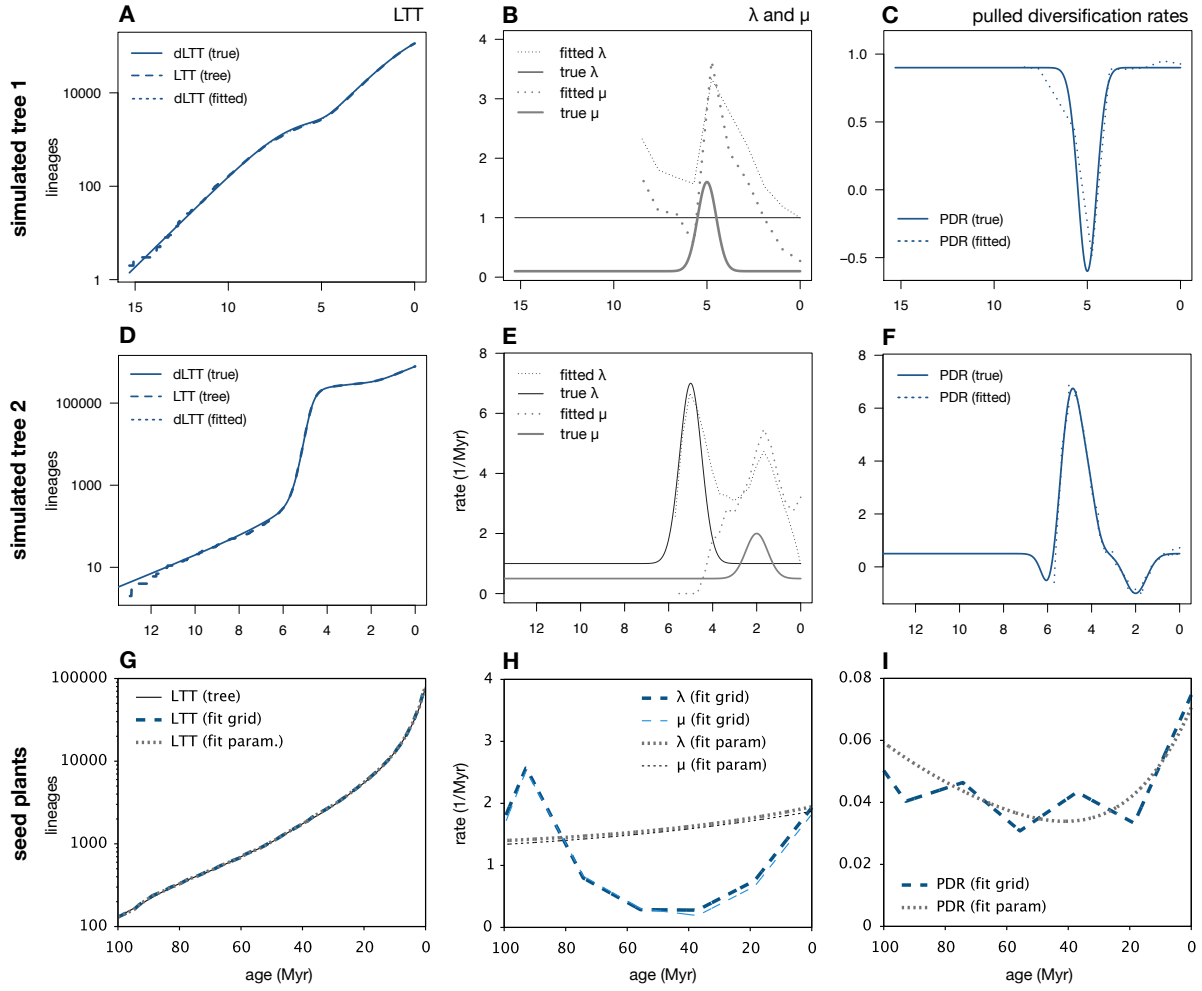

Figure S5: **Identifiability issues persist in large trees.** (A–C) Diversification analysis of a timetree (∼114,000 tips) simulated from a birth-death process exhibiting a mass extinction event at ∼5 Myr before present. (A) Lineages through time curve (LTT) of the generated tree (long-dashed curve), 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). Observe the great disagreement between the fitted and true λ and μ, despite the fact that the allowed model set could in principle approximate the true rates reasonably well. (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 (∼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. Again, observe the great disagreement between the fitted and true λ and μ, despite the fact that the allowed model set could in principle approximate the true rates reasonably well. 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 (26), 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.

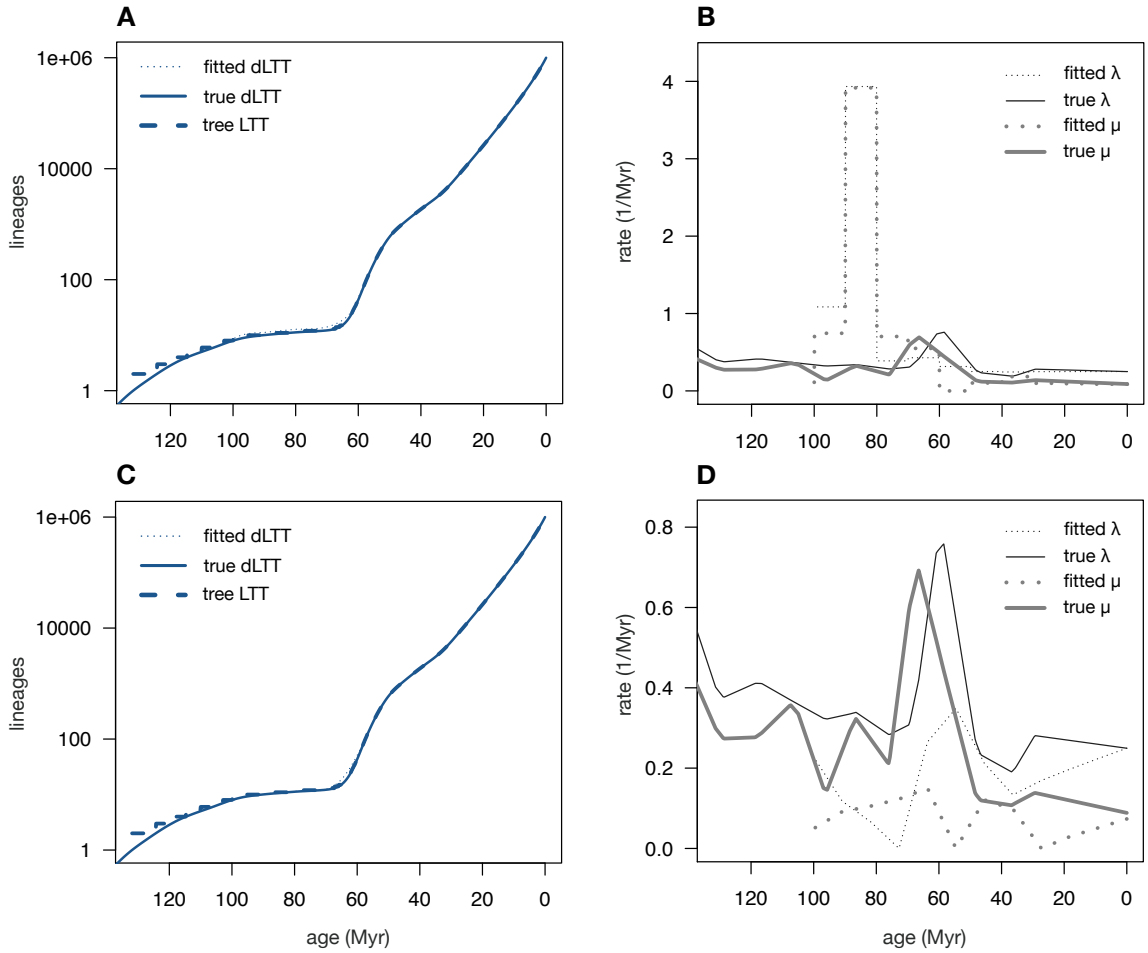

Figure S6: **Identifiability issues cannot be resolved with AIC.** Maximum-likelihood birth-death models fitted to a tree comprising 1,000,000 tips, simulated based on the origination and extinction rates of marine invertebrate genera estimated from fossil data (28). Top row: Maximum-likelihood-fitted piecewise constant model (also known as birth-death-shift model), with grid size ($N = 11$) chosen by minimizing the AIC. Bottom row: Maximum-likelihood-fitted piecewise linear model, with grid size ($N = 12$) chosen by minimizing the AIC. Left column: dLTTs of the fitted models compared to the true dLTT and the tree's LTT. Right column: Fitted speciation and extinction rates, compared to the true rates used to generate the tree. Observe that in both cases the maximum-likelihood models poorly reflect the true rates despite a near-perfect match of the LTT, even when the complexity of the models was optimized based on the AIC. For Methods details see Supplement S.6.

References

[1] D. G. Kendall, On some modes of population growth leading to RA Fisher's logarithmic series distribution. *Biometrika* **35**, 6–15 (1948).

[2] P. H. Harvey, R. M. May, S. Nee, Phylogenies without fossils. *Evolution* **48**, 523–529 (1994).

[3] S. Nee, R. M. May, P. H. Harvey, The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **344**, 305–311 (1994).

[4] T. Kubo, Y. Iwasa, Inferring the rates of branching and extinction from molecular phylogenies. *Evolution* **49**, 694–704 (1995).

[5] H. Morlon, T. L. Parsons, J. B. Plotkin, Reconciling molecular phylogenies with the fossil record. *Proceedings of the National Academy of Sciences* **108**, 16327–16332 (2011).

[6] S. Louca, *et al.*, Bacterial diversification through geological time. *Nature Ecology & Evolution* **2**, 1458–1467 (2018).

[7] D. G. Kendall, *et al.*, On the generalized "birth-and-death" process. *The annals of mathematical statistics* **19**, 1–15 (1948).

[8] V. Climenhaga, A. Katok, *From Groups to Geometry and Back*, Student Mathematical Library (American Mathematical Society, Providence, Rhode Island, USA, 2017).

[9] D. L. Rabosky, Likelihood methods for detecting temporal shifts in diversification rates. *Evolution* **60**, 1152–1164 (2006).

[10] D. L. Rabosky, Laser: a maximum likelihood toolkit for detecting temporal shifts in diversification rates from molecular phylogenies. *Evolutionary Bioinformatics* **2** (2006).

[11] D. L. Rabosky, I. J. Lovette, Explosive evolutionary radiations: decreasing speciation or increasing extinction through time? *Evolution: International Journal of Organic Evolution* **62**, 1866–1875 (2008).

[12] D. Silvestro, J. Schnitzler, G. Zizka, A bayesian framework to estimate diversification rates and their variation through time and space. *BMC Evolutionary Biology* **11**, 311 (2011).

[13] T. Stadler, Mammalian phylogeny reveals recent diversification rate shifts. *Proceedings of the National Academy of Sciences* **108**, 6187–6192 (2011).

[14] H. Morlon, Phylogenetic approaches for studying diversification. *Ecology Letters* **17**, 508–525 (2014).

[15] M. A. McPeek, The ecological dynamics of clade diversification and community assembly. *The American Naturalist* **172**, E270–E284 (2008).

[16] P. J. Mayhew, G. B. Jenkins, T. G. Benton, A long-term association between global temperature and biodiversity, origination and extinction in the fossil record. *Proceedings of the Royal Society B: Biological Sciences* **275**, 47–53 (2008).

[17] J. A. Esselstyn, R. M. Timm, R. M. Brown, Do geological or climatic processes drive speciation in dynamic archipelagos? the tempo and mode of diversification in southeast asian shrews. *Evolution* **63**, 2595–2610 (2009).

[18] A. N. Egan, K. A. Crandall, Divergence and diversification in North American Psoraleeae (Fabaceae) due to climate change. *BMC Biology* **6**, 55 (2008).

[19] J. L. Cantalapiedra, *et al.*, Dietary innovations spurred the diversification of ruminants during the Caenozoic. *Proceedings of the Royal Society B: Biological Sciences* **281**, 20132746 (2014).

[20] F. L. Condamine, J. Rolland, H. Morlon, Macroevolutionary perspectives to environmental change. *Ecology Letters* **16**, 72–85 (2013).

[21] T. Stadler, On incomplete sampling under birth–death models and connections to the sampling-based coalescent. *Journal of Theoretical Biology* **261**, 58–66 (2009).

[22] H. Morlon, M. D. Potts, J. B. Plotkin, Inferring the dynamics of diversification: A coalescent approach. *PLOS Biology* **8**, e1000493 (2010).

[23] T. Stadler, How can we improve accuracy of macroevolutionary rate estimates? *Systematic Biology* **62**, 321–329 (2013).

[24] T. Stadler, M. Steel, Swapping birth and death: Symmetries and transformations in phylodynamic models. *Systematic Biology* **68**, 852–858 (2019).

[25] S. Louca, M. Doebeli, Efficient comparative phylogenetics on large trees. *Bioinformatics* **34**, 1053–1055 (2017).

[26] S. A. Smith, J. W. Brown, Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany* **105**, 302–314 (2018).

[27] H. Akaike, Likelihood of a model and information criteria. *Journal of Econometrics* **16**, 3–14 (1981).

[28] J. Alroy, Dynamics of origination and extinction in the marine fossil record. *Proceedings of the National Academy of Sciences* **105**, 11536–11542 (2008).

[29] S. Magallón, S. Gómez-Acevedo, L. L. Sánchez-Reyes, T. Hernández-Hernández, A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist* **207**, 437–453 (2015).

[30] R. Govaerts, How many species of seed plants are there? *Taxon* **50**, 1085–1090 (2001).