Report

Current Biology

Why extinction estimates from extant phylogenies are so often zero

Highlights

- Estimates of extinction rates from phylogenies are suspiciously often zero
- We present a simple explanation for this apparent conundrum
- Our explanation is based on the discovery of congruent diversification scenarios
- We confirm the plausibility of our explanation using simulations and empirical trees

Authors

Stilianos Louca, Matthew W. Pennell

Correspondence

louca.research@gmail.com (S.L.), pennell@zoology.ubc.ca (M.W.P.)

In brief

Extinction rates estimated from phylogenies of extant species are often zero, in contradiction to the fossil record that reveals that extinctions are ubiquitous. Louca and Pennell provide a simple explanation for this conundrum, based on their recent discovery of "congruent" (i.e., statistically indistinguishable) diversification scenarios.

Louca & Pennell, 2021, Current Biology *31*, 1–6
July 26, 2021
https://doi.org/10.1016/j.cub.2021.04.066

CellPress

Please cite this article in press as: Louca and Pennell, Why extinction estimates from extant phylogenies are so often zero, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.04.066

Current Biology

CellPress

Report

Why extinction estimates from extant phylogenies are so often zero

Stilianos Louca[1,2,5,6,*] and Matthew W. Pennell[3,4,5,*]
[1]Department of Biology, University of Oregon, 1210 University of Oregon, Eugene, OR 97403, USA
[2]Institute of Ecology and Evolution, University of Oregon, 5289 University of Oregon, Eugene, OR 97403, USA
[3]Biodiversity Research Centre, University of British Columbia, 2212 Main Mall, Vancouver, BC V6T1Z4, Canada
[4]Department of Zoology, University of British Columbia, 6270 University Boulevard, Vancouver, BC V6T1Z4, Canada
[5]These authors contributed equally
[6]Lead contact
*Correspondence: louca.research@gmail.com (S.L.), pennell@zoology.ubc.ca (M.W.P.)
https://doi.org/10.1016/j.cub.2021.04.066

SUMMARY

Time-calibrated phylogenies of extant species ("extant timetrees") are widely used to estimate historical speciation and extinction rates by fitting stochastic birth-death models.[1] These approaches have long been controversial, as many phylogenetic studies report zero extinction in many taxa, contradicting the high extinction rates seen in the fossil record and the fact that the majority of species ever to have existed are now extinct.[2–9] To date, the causes of this discrepancy remain unresolved. Here, we provide a novel and simple explanation for these "zero-inflated" extinction estimates, based on the recent discovery that there exist many alternative "congruent" diversification scenarios that cannot be distinguished based solely on extant timetrees.[10] Due to such congruencies, estimation methods tend to converge to some scenario congruent to (i.e., statistically indistinguishable from) the true diversification scenario, but not necessarily to the true diversification scenario itself. This congruent scenario may exhibit negative extinction rates, a biologically meaningless but mathematically feasible situation, in which case estimators will tend to stick to the boundary of zero extinction. Based on this explanation, we make multiple testable predictions, which we confirm using analyses of simulated trees and 121 empirical trees. In contrast to other proposed mechanisms for erroneous extinction rate estimates,[5,11–14] our proposed mechanism specifically explains the zero inflation of previous extinction rate estimates in the absence of detectable model violations, even for large trees. Not only do our results likely resolve a long-standing mystery in phylogenetics, they demonstrate that model congruencies can have severe consequences in practice.

RESULTS AND DISCUSSION

Model congruency predicts zero-inflated extinction rate estimates

We consider the general birth-death model with time-dependent speciation rate λ, time-dependent extinction rate μ, and random extant lineage sampling with sampling fraction ρ.[10,12] By "diversification scenario," we mean a specific choice of λ and μ over time and a specific ρ. For any given diversification scenario, there exist many alternative ("congruent") scenarios that generate extant timetrees with the same probability distribution as the given scenario.[10] Congruent scenarios exhibit identical likelihoods and cannot be distinguished using extant timetrees alone, no matter how large and complete. The set of all congruent scenarios, called a "congruence class," contains a myriad of similarly plausible and yet markedly different scenarios. Crucially, when fitting birth-death models to an extant timetree, for example, via maximum-likelihood, estimators will generally converge toward the congruence class of the true diversification scenario, but not necessarily to the true diversification scenario itself. Hence, extinction rate estimates obtained entirely from extant timetrees, no matter how large, will often be very wrong.[10,15] The question is: why are these (wrong) extinction rate estimates often exactly zero, instead of simply "random" positive numbers? In other words, why is the distribution of extinction rate estimates in the literature zero inflated? We propose that the answer to this mystery stems from the fact that the likelihood function of a birth-death model is mathematically well defined for extinction rates that are negative at some or all time points.[16] If one were to consider models with partly or fully negative extinction rates, then the congruence class of the true historical diversification scenario, denoted $\mathbb{H}$, would include many scenarios with partly or fully negative μ and yet positive λ (see Supplement S1.4 in Louca and Pennell[10] for constructing such scenarios). A scenario with $\mu < 0$ is biologically meaningless, but if one were to permit such scenarios, it could often be the case that such a scenario is "closest" to $\mathbb{H}$ (i.e., has the highest likelihood) among the set of considered scenarios, even compared to those scenarios with positive μ. This is not paradoxical once one recognizes that what one is really estimating is the congruence class of the true diversification history and not the true diversification history itself and that any scenario

Please cite this article in press as: Louca and Pennell, Why extinction estimates from extant phylogenies are so often zero, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.04.066

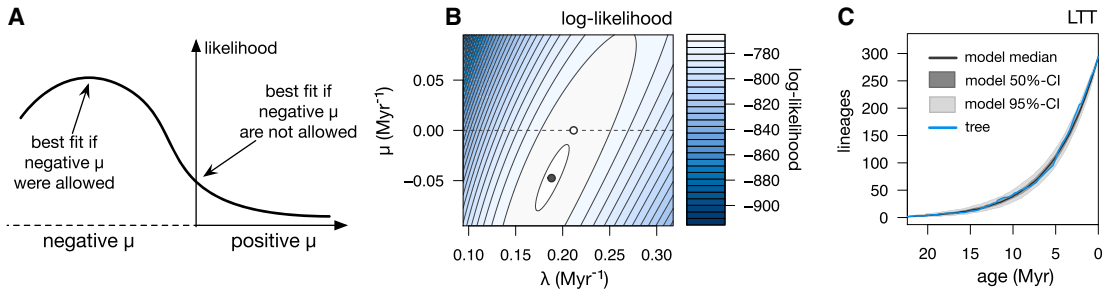

Figure 1. Conceptual illustration

(A) For any extant timetree, the likelihood function will generally be maximized in a parameter region close to the congruence class of the true diversification history (including scenarios with partly or fully negative μ), but not necessarily close to the true diversification history itself. This maximum may even be located in a region where $\mu < 0$. Constraining μ to be positive may merely result in a "compromised" fit, where $\mu = 0$. In the above illustration, we focus on μ as a single fitted parameter for simplicity; however, in practice, the fitted μ and λ may have complex functional forms depending on multiple parameters.

(B) Contour plot of the log-likelihood of an empirical timetree of 293 Trochilidae species,[17] under constant-rate birth-death models with various λ (horizontal axis) and μ (vertical axis) and fixed known sampling proportion. When extinction rates are allowed to be negative, the maximum-likelihood scenario exhibits a negative extinction rate (black dot). Constraining μ to non-negative values will yield a maximum-likelihood fit that is zero (white dot). The dashed line at the $\mu = 0$ boundary is shown for reference.

(C) Lineages-through-time (LTT) curve of the Trochilidae tree (blue curve), compared to LTTs generated by the fitted constant-rates model (black curve shows median; shades show 50% and 95% confidence intervals). Note the good agreement between the model's and the tree's LTTs. The fitted model could not be rejected based on the Sackin[18] and Colless[19] statistics, the distribution of node ages, and the distribution of branch lengths.

For an analogous Bayesian inference, see Figure S1.

with negative μ is congruent to a myriad of scenarios with positive μ. However, imposing the biologically motivated constraint that $\mu \geq 0$, as is typically done in phylogenetic software, places a boundary in parameter space that likelihood optimizers will tend to run up against, thus yielding estimates for μ that are partly or entirely zero (illustration in Figure 1A). Similarly, in a Bayesian context, the posterior distribution may be concentrated in regions where μ is partly or entirely negative, and hence, samplers restricted to non-negative μ will tend toward the zero boundary. Note that whether a fitted model exhibits negative extinction rates, if allowed, depends on the set of considered scenarios (e.g., the specific functional forms fitted). We also mention that the situation differs for the speciation rate (see supporting analysis in Method details).

Figures 1B and 1C illustrate our reasoning using an empirical timetree of 293 hummingbird species.[17] If we only consider scenarios where λ and μ are constant through time, the likelihood has its global maximum at a scenario with a negative μ (Figure 1B). A similar likelihood surface apparently peaking at negative μ was previously reported for *Drosophila* in 1994 by Nee et al.[1] (Figure 5 therein), but this crucial clue has apparently been ignored. As expected, fitting a constant-rates birth-death model to the hummingbird tree via maximum-likelihood yields a negative μ when allowed and a zero μ when constrained to non-negative values (Figure 1B). Importantly, in this case, the fitted model with zero μ explains the data well (Figure 1C) and is not rejected based on any of the model adequacy tests that we considered (using parametric bootstrapping, considering the Sackin[18] and Colless[19] statistics, the distribution of edge lengths based on a Kolmogorov-Smirnov test, and the distribution of node ages based on a Kolmogorov-Smirnov test; $p > 0.05$ in all cases). Hence, the fact that μ is estimated to be zero is unlikely due to an inadequacy of the model to explain the data at hand; rather, the maximum-likelihood scenario with negative μ is probably congruent to another biologically plausible diversification scenario close to the true (but unknown) hummingbird diversification history. Similarly, fitting a constant-rates birth-death model via Bayesian Markov chain Monte Carlo (MCMC) yields a posterior distribution concentrated at negative μ (if allowed) or at $\mu \approx 0$ when constrained to non-negative values (Figure S1).

Based on the above reasoning, we make the following testable predictions for extinction rate estimates obtained via maximum-likelihood estimation. (We do not consider Bayesian estimation here, although similar arguments would apply.) First, erroneously obtaining zero extinction rate estimates should be common even without detectable model violations, i.e., even when the fitted models explain the data well. Second, in almost all cases where μ is erroneously estimated to be zero at one or more time points, one should obtain negative extinction rate estimates if these were allowed. In particular, when allowing negative μ, the distribution of estimated μ should no longer be zero inflated. Third, estimating a negative μ (if allowed) should increase the chances of obtaining a zero extinction rate estimate when constrained to non-negative values, due to the optimization routine getting "trapped" on the $\mu = 0$ boundary. Note that this is not a strict requirement; in some cases, the estimated μ might be negative if allowed and yet strictly positive if constrained, due to the possibility of multiple local maxima in the likelihood and the fact that the optimal value for one parameter generally depends on the values of other parameters. Fourth, in cases where the estimated μ is positive even if allowed to be negative, this estimate should typically be similar to the estimate obtained when constraining μ to be non-negative. Fifth, fixing λ to its true value during fitting, which "collapses" the congruence class to a single scenario, should yield much more accurate estimates of μ and should eliminate their zero inflation. Sixth, even for fitted models with negative μ, these models should be close to the true diversification history's congruence class provided sufficient data for fitting and a sufficiently flexible model.

Please cite this article in press as: Louca and Pennell, Why extinction estimates from extant phylogenies are so often zero, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.04.066

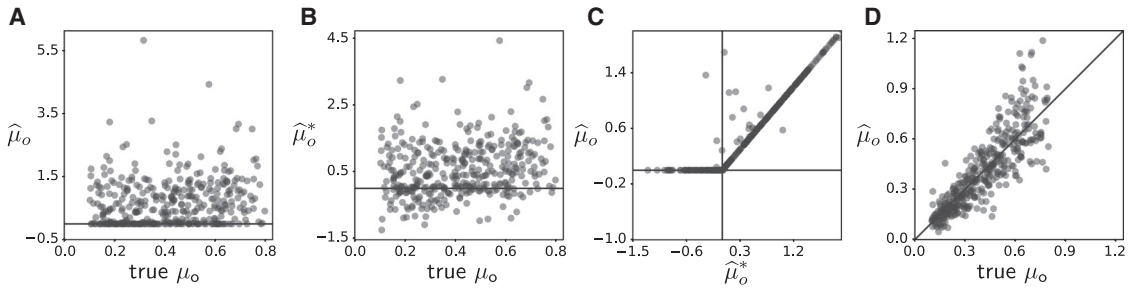

Figure 2. Testing predictions using simulations

(A) Present-day extinction rates fitted to 380 simulated extant timetrees, while requiring the fitted extinction rate to be non-negative (one point per tree). Multiple commonly used models were fitted and the best fitted model was selected based on AIC. Vertical axis: fitted present-day extinction rate ($\widehat{\mu}_o$) is shown. Horizontal axis: true present-day extinction rates (μ_o) are shown.

(B) Present-day extinction rates fitted to the same trees and with the same best models as in (A) but allowing for negative extinction rates.

(C) Present-day extinction rates fitted while allowing negative extinction rates ($\widehat{\mu}_o^*$, horizontal axis) compared to the case where negative rates are not permitted ($\widehat{\mu}_o$), for the same trees as in (A).

(D) Present-day extinction rates $\widehat{\mu}_o$ (vertical axis) fitted while fixing the speciation rate λ to its true profile, compared to the true present-day extinction rate (horizontal axis); a nearly identical result was obtained for $\widehat{\mu}_o^*$ instead of $\widehat{\mu}_o$. The diagonal in (D) is shown for reference. All rates are expressed in Myr^{-1}. For histograms of the rate estimates, see Figures S2C and S2D. Only simulations where the fitted models adequately explained the tree are shown.

For additional details, see Figure S2.

Analysis of simulated data

To test our predictions, we simulated 500 extant timetrees using diversification scenarios with known time-dependent (but lineage-independent) λ and μ and then fitted models while either constraining μ to be non-negative or allowing negative μ (λ was always constrained to be non-negative, and ρ was fixed to its true value). To ensure that μ is not simply misestimated due to stochasticity stemming from small datasets, we simulated large trees with 1,000 tips. We focused on the extinction rate at present day (henceforth μ_o), although all conclusions would apply similarly to other times. The time profiles of λ and μ used in each simulation were defined according to simple stochastic processes, reflecting realistically complex diversification dynamics seen in the fossil record.[20] Importantly, at all times, μ was constrained to be above 10% of the present-day speciation rate and thus was clearly non-zero. We then used maximum likelihood to fit various common functional forms for λ and μ (combinations of exponential, linear, and constant functions, henceforth "ELC" models). Depending on whether we constrain μ to be non-negative or allow μ to be negative, we call a fitted model "constrained" or "unconstrained," respectively. For every tree, we fit all ELC models and selected the best model using AIC, as is common practice.[12] Note that real diversification dynamics will probably never be exactly captured by any human-made mathematical model; instead, the typical objective is to find an approximate model that adequately describes the data. Our analysis resembles this realistic situation, in that the fitted ELC models do not exactly match the generative processes and thus, in some situations, may not be adequate for explaining the generated trees. To illustrate our argument that bad extinction rate estimates remain probable even with good model fits, we only considered simulations where the best fitted model adequately described the data (based on a Kolmogorov-Smirnov test for the distribution of node ages, 380 out of 500 simulations were kept). For every considered simulation, we examined whether the estimated present-day extinction rate in the best fitted constrained model (denoted $\widehat{\mu}_o$) was zero. To each tree, we also fit an unconstrained version of the best model—using the same functional forms for λ and μ but allowing μ to be negative, thus obtaining a potentially new estimate for the present-day extinction rate (denoted $\widehat{\mu}_o^*$).

For all considered simulated trees, the best fitted constrained and fitted unconstrained models closely matched the deterministic lineages-through-time curve (dLTT) of the true diversification scenario as well as the lineages-through-time curve (LTT) of the simulated tree ($R^2 \geq 0.99$; Figure S2J, examples in Figures S2A and S2B). Despite the good fits and large input trees, present-day extinction rate estimates ($\widehat{\mu}_o$) poorly matched the true extinction rates (Figure 2A), consistent with our expectations.[10] For 93 out of 380 considered trees (24.5%), the best constrained model yielded a $\widehat{\mu}_o = 0$ (Figures 2A and S2C). A zero $\widehat{\mu}_o$ is in conflict with the fact that the true extinction rate was at all times clearly non-zero ($\mu \geq 0.1 \cdot \lambda_o$). This fraction of zero extinction estimates is also much larger than what would be expected if estimates were continuously distributed; in other words, the distribution of $\widehat{\mu}_o$ is zero inflated, consistent with our predictions and observations in the literature. Moreover, we did not observe any strong or significant positive correlation between the goodness of fit (in terms of the dLTT's R^2 or in terms of the p value of the Kolmogorov-Smirnov test for node ages) on the one hand and the relative estimation error ($(\widehat{\mu}_o - \mu_o)/\mu_o$) on the other hand ($p > 0.05$ and Spearman's $p < 0.1$; Figures S2I–S2L), suggesting that zero-inflated extinction estimates are unlikely the result of poor model fits. In all cases where $\widehat{\mu}_o$ was zero, the unconstrained variant yielded a negative $\widehat{\mu}_o^*$, consistent with our predictions (Figure 2C). For nearly all trees with negative $\widehat{\mu}_o^*$, we obtained a zero $\widehat{\mu}_o$, consistent with our predictions (Figure 2C). In contrast to $\widehat{\mu}_o$, the distribution of $\widehat{\mu}_o^*$ was not zero inflated, consistent with our predictions (Figure S2D). For trees where the fitted $\widehat{\mu}_o^*$ was positive, the fitted $\widehat{\mu}_o$ was almost always identical to $\widehat{\mu}_o^*$ (up to numerical accuracy), again confirming our predictions (Figure 2C).

These findings show that the zero inflation of fitted extinction rate stems from "cutting off" fits that would otherwise be

Please cite this article in press as: Louca and Pennell, Why extinction estimates from extant phylogenies are so often zero, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.04.066

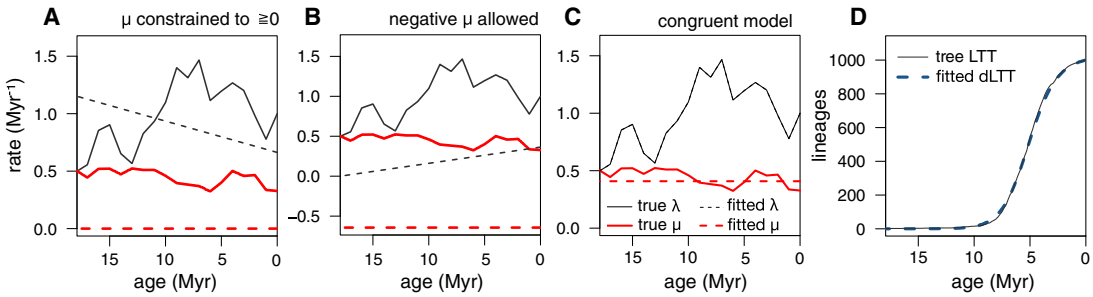

Figure 3. Fitted models are approximately congruent to the true diversification scenarios

Example ELC model fit (and selected according to AIC) to a simulated timetree.

(A) True λ and μ (continuous curves) compared to fitted λ and μ (dashed curves), when the fitted μ is constrained to be non-negative.

(B) True λ and μ (continuous curves) for the same tree as in (A), compared to λ and μ (dashed curves) fitted while allowing a negative μ. The negative fitted $\widehat{\mu}_o^*$ suggests that, in (A), the zero $\widehat{\mu}_o$ was simply "cut off" at that boundary.

(C) Diversification scenario congruent to the model fitted in (B) but exhibiting the correct extinction rate; observe that this scenario is close to the true scenario, hence the congruence class of the model fitted in (B) is close to the congruence class of the true scenario.

(D) LTT curve of the simulated tree (continuous curves) compared to the deterministic LTT of the fitted constrained model in (A). The good match between the tree's LTT and the fitted model's deterministic LTT shows that the model is close to the congruence class of the true scenario.

For an overview of fitting results, see Figure S3.

negative (example in Figure 3). Note that the occurrence of likelihood peaks in biologically implausible parameter regions (e.g., Figure 1B), despite the large tree sizes, makes sense in light of model congruencies: when fitting birth-death models, one is at most estimating the congruence class of the true diversification history, rather than the true diversification history itself, and this congruence class can include scenarios with negative extinction rates. Thus, when specific parameterized functional forms for λ and μ are fitted (as is done here), one will obtain a parameter combination whose corresponding λ and μ curves resemble some member of the true diversification scenario's congruence class, even if the true diversification scenario differed strongly. Consistent with this interpretation, for all considered simulations, the fitted model closely matched the true scenario's dLTT, which fully determines the congruence class,[10] even in cases where $\widehat{\mu}_o$ was zero (Figures S2A and S2B). To further test this interpretation, for each best fitted unconstrained model with negative $\widehat{\mu}_o^*$, we determined the unique congruent diversification scenario with the correct μ. As expected, this congruent scenario typically exhibited a λ close to the truth, confirming our interpretation that the fitted models, even those with negative $\widehat{\mu}_o^*$, are close to the true scenario's congruence class (example in Figure 3C). Similarly, when we re-fitted ELC models to the trees while fixing λ to its true profile, the extinction rate estimates were much more accurate and no longer zero inflated (in fact, none were zero; Figure 2D), consistent with our predictions. Note that the above results cannot be attributed to model inadequacy, i.e., to the possibility that ELC models are not flexible enough to adequately approximate the true profiles of λ and μ over time (explanation in Method details).

Analysis of empirical timetrees

To test the relevance of our arguments to real datasets, we examined 121 previously published timetrees of eukaryotic taxa. To facilitate comparison with our simulation results, we fit the same ELC models and chose the best model based on AIC. The sampling fraction ρ was determined separately based on the literature and fixed during model fitting (Data S1). In each case, we either constrained μ to be non-negative or allowed μ to be negative and compared the obtained present-day extinction rate estimates with or without the constraint ($\widehat{\mu}_o$ and $\widehat{\mu}_o^*$, respectively). To avoid biases due to bad model fits, we omitted 12 trees for which the best fitted model was rejected based on the distribution of node ages and using a Kolmogorov-Smirnov test ($p < 0.05$). For all 109 considered trees, the fitted models explained the data well, with each tree's LTT falling entirely or almost entirely within the 95% confidence interval of LTTs generated by the best fitted constrained model (examples in Figures S3H–S3Q) and with the model's dLTT capturing nearly all of the variance of the tree's LTT ($R^2 > 0.99$; Figure S3G). For 39 out of 109 trees, the fitted model exhibited a present-day extinction rate of exactly zero ($\widehat{\mu}_o = 0$); this high fraction of zero extinction rate estimates is consistent with the zero-inflated estimates in the literature and our simulations. In all cases where $\widehat{\mu}_o = 0$, the fitted $\widehat{\mu}_o^*$ was negative, whereas a positive $\widehat{\mu}_o$ nearly always coincided with a positive and identical $\widehat{\mu}_o^*$ (Figure S3A), consistent with our expectations. We note that model inadequacies, such as non-modeled variation in diversification rates *between* lineages,[11,14] do not provide a plausible alternative explanation for the high frequency of zero (or negative, if allowed) extinction rate estimates in these empirical trees (explanation in Method details).

Conclusions

We showed that birth-death model congruencies can explain the zero-inflated extinction rate estimates from extant timetrees observed in the literature. It is possible that other mechanisms, such as model inadequacy or misspecification[5,11–14] and errors or unresolved polytomies in the input trees,[21] also cause erroneous extinction rate estimates. However, these mechanisms do not explain the zero inflation of extinction rate estimates, and our analysis of empirical trees suggests that model inadequacy was not an important cause of zero-inflated extinction estimates in this dataset (see discussion in Method details). Regardless of whether our proposed mechanism is the main cause of zero-inflated extinction estimates, it is clear that very

erroneous zero-inflated extinction estimates are probable even without detectable model inadequacies and even for large trees. We thus concur with Marshall[9] that most estimates of zero extinction from extant timetrees are almost certainly wrong. More generally, our findings demonstrate that model congruencies have likely been seriously confounding macroevolutionary studies for decades. Whether these issues can ever be resolved remains to be seen; it is clear that timetrees of extant species alone are generally insufficient for estimating μ or even testing simple hypotheses about μ, such as whether μ was non-zero, without additional well-justified constraints. Ultimately, this limitation might be resolved with other sources of information, such as population-genetic data, fossils, or other biological tracers, and models integrating such information. We caution, however, that such models may also exhibit congruencies, for example, if the fossilization rate is an unknown time-dependent parameter also reconstructed from the data.[22] We see asymptotically estimatable parameters, such as the pulled diversification rate,[10,23] as a promising tool for macroevolutionary inferences.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - Simulation analyses
  - Analysis of empirical trees
  - Analysis of empirical tree distributions
  - Model adequacy tests
  - Bayesian inference using the hummingbirds tree
  - Estimation of extinction rates versus present-day speciation rates
  - Checking if simulation results are due to model inadequacy
  - Checking if empirical tree results result from model inadequacies
  - Comparison to other proposed mechanisms for erroneous extinction rate estimates

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2021.04.066.

ACKNOWLEDGMENTS

We thank L. Harmon, J. Rolland, B. Neto-Bradley, K. Kaur, and L.F. Henao Diaz for comments on the manuscript. M.W.P. was supported by a NSERC Discovery Grant. S.L. was supported by a startup grant by the University of Oregon.

AUTHOR CONTRIBUTIONS

Both authors contributed equally to this work.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 31, 2020
Revised: March 11, 2021
Accepted: April 26, 2021
Published: May 20, 2021

REFERENCES

1. Nee, S., Holmes, E.C., May, R.M., and Harvey, P.H. (1994). Extinction rates can be estimated from molecular phylogenies. Philos. Trans. R. Soc. Lond. B Biol. Sci. *344*, 77–82.
2. Losos, J.B., and Schluter, D. (2000). Analysis of an evolutionary species-area relationship. Nature *408*, 847–850.
3. Turgeon, J., Stoks, R., Thum, R.A., Brown, J.M., and McPeek, M.A. (2005). Simultaneous quaternary radiations of three damselfly clades across the Holarctic. Am. Nat. *165*, E78–E107.
4. Nee, S. (2006). Birth-death models in macroevolution. Annu. Rev. Ecol. Evol. Syst. *37*, 1–17.
5. Purvis, A. (2008). Phylogenetic approaches to the study of extinction. Annu. Rev. Ecol. Evol. Syst. *39*, 301–319.
6. Steeman, M.E., Hebsgaard, M.B., Fordyce, R.E., Ho, S.Y., Rabosky, D.L., Nielsen, R., Rahbek, C., Glenner, H., Sørensen, M.V., and Willerslev, E. (2009). Radiation of extant cetaceans driven by restructuring of the oceans. Syst. Biol. *58*, 573–585.
7. Quental, T.B., and Marshall, C.R. (2010). Diversity dynamics: molecular phylogenies need the fossil record. Trends Ecol. Evol. *25*, 434–441.
8. Herrera, J.P. (2017). Primate diversification inferred from phylogenies and fossils. Evolution *71*, 2845–2857.
9. Marshall, C.R. (2017). Five palaeobiological laws needed to understand the evolution of the living biota. Nat. Ecol. Evol. *1*, 165.
10. Louca, S., and Pennell, M.W. (2020). Extant timetrees are consistent with a myriad of diversification histories. Nature *580*, 502–505.
11. Rabosky, D.L. (2010). Extinction rates should not be estimated from molecular phylogenies. Evolution *64*, 1816–1824.
12. Morlon, H., Parsons, T.L., and Plotkin, J.B. (2011). Reconciling molecular phylogenies with the fossil record. Proc. Natl. Acad. Sci. USA *108*, 16327–16332.
13. Beaulieu, J.M., and O'Meara, B.C. (2015). Extinction can be estimated from moderately sized molecular phylogenies. Evolution *69*, 1036–1043.
14. Rabosky, D.L. (2016). Challenges in the estimation of extinction from molecular phylogenies: a response to Beaulieu and O'Meara. Evolution *70*, 218–228.
15. Pagel, M. (2020). Evolutionary trees can't reveal speciation and extinction rates. Nature *580*, 461–462.
16. Stadler, T., and Steel, M. (2019). Swapping birth and death: symmetries and transformations in phylodynamic models. Syst. Biol. *68*, 852–858.
17. McGuire, J.A., Witt, C.C., Remsen, J.V., Jr., Corl, A., Rabosky, D.L., Altshuler, D.L., and Dudley, R. (2014). Molecular phylogenetics and the diversification of hummingbirds. Curr. Biol. *24*, 910–916.
18. Sackin, M.J. (1972). 'Good' and 'bad' phenograms. Syst. Biol. *21*, 225–226.
19. Shao, K.-T. (1990). Tree balance. Syst. Biol. *39*, 266–276.
20. Alroy, J. (2008). Colloquium paper: dynamics of origination and extinction in the marine fossil record. Proc. Natl. Acad. Sci. USA *105* (*Suppl 1*), 11536–11542.
21. Kuhn, T.S., Mooers, A.Ø., and Thomas, G.H. (2011). A simple polytomy resolver for dated phylogenies. Methods Ecol. Evol. *2*, 427–436.
22. Louca, S., McLaughlin, A., MacPherson, A., Joy, J.B., and Pennell, M.W. (2021). Fundamental identifiability limits in molecular epidemiology. bioRxiv. https://doi.org/10.1101/2021.01.18.427170.

23. Louca, S., Shih, P.M., Pennell, M.W., Fischer, W.W., Parfrey, L.W., and Doebeli, M. (2018). Bacterial diversification through geological time. Nat. Ecol. Evol. *2*, 1458–1467.
24. Louca, S., and Doebeli, M. (2018). Efficient comparative phylogenetics on large trees. Bioinformatics *34*, 1053–1055.
25. Uhlenbeck, G.E., and Ornstein, L.S. (1930). On the theory of the Brownian motion. Phys. Rev. *36*, 823–841.
26. Louca, S. (2020). Simulating trees with millions of species. Bioinformatics *36*, 2907–2908.
27. Rabosky, D.L., and Lovette, I.J. (2008). Explosive evolutionary radiations: decreasing speciation or increasing extinction through time? Evolution *62*, 1866–1875.
28. Henao Diaz, L.F., Harmon, L.J., Sugawara, M.T.C., Miller, E.T., and Pennell, M.W. (2019). Macroevolutionary diversification rates show time dependency. Proc. Natl. Acad. Sci. USA *116*, 7403–7408.
29. Upham, N.S., Esselstyn, J.A., and Jetz, W. (2019). Inferring the mammal tree: species-level sets of phylogenies for questions in ecology, evolution, and conservation. PLoS Biol. *17*, e3000494.
30. Jetz, W., Thomas, G.H., Joy, J.B., Redding, D.W., Hartmann, K., and Mooers, A.O. (2014). Global distribution and conservation of evolutionary distinctness in birds. Curr. Biol. *24*, 919–930.
31. Akaike, H. (1981). Likelihood of a model and information criteria. J. Econom. *16*, 3–14.
32. Tonini, J.F.R., Beard, K.H., Ferreira, R.B., Jetz, W., and Pyron, R.A. (2016). Fully-sampled phylogenies of squamates reveal evolutionary patterns in threat status. Biol. Conserv. *204*, 23–31.
33. Jetz, W., Thomas, G.H., Joy, J.B., Hartmann, K., and Mooers, A.O. (2012). The global diversity of birds in space and time. Nature *491*, 444–448.
34. Jetz, W., and Pyron, R.A. (2018). The interplay of past diversification and evolutionary isolation with present imperilment across the amphibian tree of life. Nat. Ecol. Evol. *2*, 850–858.
35. Brown, J.M., and Thomson, R.C. (2018). Evaluating model performance in evolutionary biology. Annu. Rev. Ecol. Evol. Syst. *49*, 95–114.
36. Schwery, O., and O'Meara, B.C. (2021). The shape of trees - limits of current diversification models. bioRxiv. https://doi.org/10.1101/2021.01.26.428344.
37. FitzJohn, R.G. (2012). Diversitree: comparative phylogenetic analyses of diversification in R. Methods Ecol. Evol. *3*, 1084–1092.
38. Nee, S., May, R.M., and Harvey, P.H. (1994). The reconstructed evolutionary process. Philos. Trans. R. Soc. Lond. B Biol. Sci. *344*, 305–311.
39. Phillimore, A.B., and Price, T.D. (2008). Density-dependent cladogenesis in birds. PLoS Biol. *6*, e71.
40. Pannetier, T., Martinez, C., Bunnefeld, L., and Etienne, R.S. (2021). Branching patterns in phylogenies cannot distinguish diversity-dependent diversification from time-dependent diversification. Evolution *75*, 25–38.
41. Etienne, R.S., and Rosindell, J. (2012). Prolonging the past counteracts the pull of the present: protracted speciation can explain observed slowdowns in diversification. Syst. Biol. *61*, 204–213.
42. Moen, D., and Morlon, H. (2014). Why does diversification slow down? Trends Ecol. Evol. *29*, 190–197.
43. Foote, M. (2003). Origination and extinction through the Phanerozoic: a new approach. J. Geol. *111*, 125–148.

Please cite this article in press as: Louca and Pennell, Why extinction estimates from extant phylogenies are so often zero, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.04.066

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
castor 1.6.7	Louca and Doebeli[24]	https://cran.r-project.org/web/packages/castor/index.html
R	R project	https://www.r-project.org
Other		
Time-calibrated empirical phylogenies	primary literature	Data S1

RESOURCE AVAILABILITY

Lead contact

Reasonable requests for further information and resources should be directed to and will be fulfilled by the lead contact, Stilianos Louca (louca.research@gmail.com).

Materials availability

This study did not generate new samples or unique reagents.

Data and code availability

R code for performing our analyses is available at: http://www.loucalab.com/archive/ZeroExtinction. All required packages are freely available at the Comprehensive R Archive Network (https://cran.r-project.org). All empirical phylogenies were obtained from the literature; tree sources and literature references are given in Data S1.

METHOD DETAILS

Simulation analyses

All simulations and maximum-likelihood fitting of birth-death models were performed using the R package castor v1.6.7.[24] All simulated trees had 1000 tips. All times are measured in Myr, and rates are measured in Myr^{-1}. To simulate a variety of diversification scenarios we generated random profiles for λ and μ according to independent Ornstein-Uhlenbeck (OU) stochastic processes.[25] This approach was chosen in order to cover a wide range of diversification scenarios with realistic temporal complexity. For λ, the present-day value λ_o (serving as initial condition for the OU process) was set to $1Myr^{-1}$, the stationary expectation was set to λ_o, the stationary standard deviation was set to $0.5\cdot\lambda_o$ and the relaxation rate was set to $0.1Myr^{-1}$. For μ, the present-day value μ_o was chosen randomly between $0.1\cdot\lambda_o$ and $0.8\cdot\lambda_o$, the stationary expectation was set to μ_o, the stationary standard deviation was set to $0.5\cdot\mu_o$ and the relaxation rate was set to $0.1Myr^{-1}$. Both λ and μ were specified on a discrete time grid spanning 100 Myr and having a time-step of 1 Myr, according to the exact distribution of OU paths, with the exception that values below $0.1Myr^{-1}$ were replaced with $0.1Myr^{-1}$ to ensure that speciation and extinction rates were always above a detectable threshold. The sampling fraction ρ was chosen randomly and uniformly on a logarithmic scale from 0.01 to 1. For every given profile for λ and μ and chosen ρ, we simulated an extant timetree using the castor function generate_tree_hbd_reverse.[26]

Models fitted to the tree were based on functional forms commonly encountered in the literature:[27] λ was either i) assumed to be constant over time, ii) assumed to vary exponentially ($\lambda(t) = \alpha e^{\beta t}$), iii) assumed to vary exponentially plus a constant ($\lambda = \alpha e^{\beta t} + \gamma$), or iv) assumed to vary linearly ($\lambda(t) = \lambda_o + \alpha t$). Similar profiles were considered for μ, thus yielding 4×4 alternative combinations (henceforth "ELC" models, for exponential/linear/constant). We fitted ELC models to the tree using the castor function fit_hbd_model_parametric with appropriately defined functional forms for each ELC model and with options "condition='auto', relative_dt=1e-3, max_start_attempts=100, fit_control=list(eval.max=10000, iter.max=1000, rel.tol=1e-12, step.min=0.00001)." Fitting was agnostic of the simulation parameters, i.e., as if no further information was available apart from the tree itself and the sampling fraction ρ (which was fixed to its true value). As a first guess for the parameters (option "param_guess") we set the exponent (in the case of exponentially varying rates) to zero, and the present-day rates to values obtained by first fitting a constant-rates BD model with known ρ. The bounds for the fitted parameters were chosen such that μ was necessarily non-negative. After fitting all ELC models to a tree, we chose the ELC model that had the smallest AIC, following common practice. We then re-fitted that model with parameter bounds adjusted to allow for negative μ, and compared the estimated present-day μ_o obtained with and without constraining μ to non-negative values ($\widehat{\mu}_o$ and $\widehat{\mu}_o^*$, respectively). To also examine how the extinction rate estimates change if λ was known (Figure 2D), we re-fitted ELC models (and selected the best model via AIC) while setting λ to its true profile over time. To construct

diversification scenarios congruent to the fitted models, but with the correct μ (e.g., Figure 3C), we used the castor function simulate_deterministic_hbd.

To investigate how errors in empirical trees might affect our conclusions, we repeated the above simulation analysis using erroneous trees. Specifically, for each simulated tree we randomly shifted the ages of internal nodes (adding a random number chosen uniformly within the interval $[-\tau/5, \tau/5]$, where τ is the true node age) and subsequently merged a random subset of internal nodes (20% of nodes) with their children nodes into multifurcations. These modifications were done using the castor functions shift_clade_times (with options "shift_descendants = FALSE, negative_edge_lengths = 'move_parent'") and merge_nodes_to_multifurcations (with options "merge_with_parents=FALSE, keep_ancestral_ages=FALSE"). Out of 500 simulated trees, 397 trees were adequately described by the best fitted constrained ELC model (based on a Kolmogorov-Smirnov test for the distribution of node ages, $P > 0.05$). Out of those 397 trees, 123 trees (31.0%) yielded a zero extinction rate estimate ($\widehat{\mu}_o = 0$) when constrained. All trees for which $\widehat{\mu}_o = 0$ yielded a negative extinction rate estimate when allowed ($\widehat{\mu}_o^* < 0$). We also repeated this analysis using trees solely modified by merging nodes into multifurcations, i.e., without previously altering the node ages. Out of 500 simulated trees, 381 trees were adequately described by the best fitted constrained ELC model (based on the distribution of node ages, $P > 0.05$). Out of those 381 trees, 119 trees (31.0%) yielded a zero extinction rate estimate ($\widehat{\mu}_o = 0$) when constrained. These results closely resemble our findings with accurate (non-erroneous) trees, described in the main article.

Analysis of empirical trees

Extant timetrees of 121 eukaryotic taxa were obtained from the literature; many of these timetrees had been previously collected from the literature by Henao Diaz et al.,[28] and some were obtained by extracting sub-trees corresponding to recognized taxa (families or orders) from a large mammal tree[29] or avian tree.[30] Taxon-specific sub-trees were only extracted if the tips associated with the taxon indeed formed a monophyletic clade in the original tree, based on the NCBI taxonomy (status December 3, 2020). The size of the trees ranged from 13 to 11,638 species, and their root age ranged from 6.6 to 431 Myr. The sampling fraction ρ of each tree was calculated based on published estimates of the taxon's total number of extant species; tree sources and literature references are given in Data S1. To each tree, we fit all ELC models while fixing ρ and chose the best supported constrained ELC model according to the AIC.[31]

Analysis of empirical tree distributions

Empirical trees are typically only estimates of the true phylogenetic relationships between species, for example if they have been drawn from a posterior distribution. To examine how sensitive the probability of a zero extinction rate estimate is to variations in tree estimates, we analyzed 100 trees from each of four published posterior timetree distributions: The mammal tree by Upham et al.[29] (4098 tips), the squamate tree by Tonini et al.[32] (5415 tips), the bird tree by Jetz et al.[33] (6670 tips) and the amphibian tree by Jetz et al.[34] (4061 tips). For all mammal trees, the best fitted constrained ELC model yielded a strictly positive $\widehat{\mu}_o^*$ and $\widehat{\mu}_o$. For 34 out of 100 squamate trees, the best fitted constrained ELC model yielded $\widehat{\mu}_o = 0$; in all of those cases, the corresponding $\widehat{\mu}_o^*$ was negative, consistent with our expectations. For 82 out of 100 bird trees, the best fitted constrained ELC model yielded $\widehat{\mu}_o = 0$; in all of those cases, the corresponding $\widehat{\mu}_o^*$ was negative, consistent with our expectations. For 26 out of 100 amphibian trees, the best fitted constrained ELC model yielded $\widehat{\mu}_o = 0$; in all but one of those cases, the corresponding $\widehat{\mu}_o^*$ was negative, largely consistent with our expectations. These examples also demonstrate that the precise location of the likelihood maximum (negative versus positive μ) is generally also affected by the quality of the tree.

Model adequacy tests

Statistical tests for model adequacy were performed using parametric bootstrapping,[35,36] as follows. For a given tree and a given fitted model to be evaluated, we simulated 1000 random trees from the model, fixing the number of tips and the root age to match those of the original tree. Simulations were performed using the castor function generate_tree_hbd_reverse.[26] For every simulated tree, we calculated the Sackin statistic[18] (denoted σ), and then calculated the mean Sackin statistic across all simulated trees (denoted $\overline{\sigma}$). The statistical significance of the tree's Sackin statistic (denoted σ_o) was the fraction of simulated trees for which $|\sigma - \overline{\sigma}|$ was greater than $|\sigma_o - \overline{\sigma}|$. A similar approach was taken for the Colless[19] statistic. To compare the distribution of node ages (times before present) of the fitted model to the tree, we used a Kolmogorov-Smirnov test. Specifically, for every simulated tree we calculated the empirical cumulative distribution function (CDF) of the node ages (denoted F), evaluated at the original tree's node ages via linear interpolation, and then calculated the average of those CDFs, thus obtaining an estimate for the CDF of node ages expected under the model (denoted $\overline{F}$). The Kolmogorov-Smirnov (KS) distance between a tree's CDF F and $\overline{F}$, denoted $D(F,\overline{F})$, is the maximum distance between F and $\overline{F}$ at any age. The statistical significance of the original tree's KS distance $D(F_o,\overline{F})$ was calculated as the fraction of simulated trees for which $D(F,\overline{F})$ was larger than $D(F_o,\overline{F})$. A similar approach was followed for comparing the model's and tree's distribution of edge lengths. The full set of statistical tests is implemented in the castor function model_adequacy_hbd.[24] In all tests, the statistical significance threshold was set to 5%.

Bayesian inference using the hummingbirds tree

To illustrate the consequences for Bayesian inference when the likelihood function has a maximum at negative μ, we performed Bayesian MCMC of a constant-rates birth-death model using the empirical hummingbirds tree,[17] discussed in the main article, as input. The MCMC was ran using the function mcmc in the R package diversitree,[37] with options "nsteps=10000, w=.1," and using the castor function loglikelihood_hbd as a custom likelihood function. A flat prior was used for both λ and μ.

Please cite this article in press as: Louca and Pennell, Why extinction estimates from extant phylogenies are so often zero, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.04.066

Estimation of extinction rates versus present-day speciation rates

In this work we have demonstrated using simulated and empirical trees that extinction rates cannot be reliably estimated entirely from extant species timetrees alone, no matter how large and accurate these trees are. In fact the errors can be so large that — if allowed — extinction rates are often estimated to be negative with strong statistical support, a biologically impossible situation. It should be noted, however, that the situation is not as severe for speciation rates, for two reasons. First, in contrast to μ_o, the product of the present-day speciation rate and sampling fraction, $\lambda_o\rho$, can asymptotically (i.e., provided a sufficiently large tree) be accurately estimated. A mathematical proof for the general time-dependent birth-death model was provided by Louca et al.[10] An intuitive way to explain this is that near the present extinction has not yet had time to erase much information from the tree, and hence the per-lineage branching density resembles $\lambda_o\rho$. If ρ is somehow independently known, then we can in fact estimate λ_o if given a sufficiently large tree. Second, the congruence class of any given biologically realistic birth-death process, i.e., with sufficiently smooth and non-negative rates λ, μ, does not contain any members with negative speciation rate (unless we also allow for negative ρ). Indeed, as seen in Equation (39) in Supplement S1.4 of Louca et al.,[10] the speciation rate of any congruent model is always non-negative even if μ was negative. Hence, when fitting birth-death models to extant timetrees, it will be rare to encounter a likelihood function that is maximized at a negative λ, and in fact the probability of this occurring will tend to zero as the size of the tree tends to infinity.

To illustrate this difference in our ability to estimate λ_o (when ρ is known) versus μ_o, we plotted the relative estimation error of λ_o and μ_o, obtained from our unconstrained as well as constrained ELC models, in Figures S2E–S2H. As can be seen, λ_o is generally estimated much more accurately than μ_o. In fact, the mean modulus of the relative estimation error is about one order of magnitude greater for μ_o compared to λ_o.

Checking if simulation results are due to model inadequacy

The simulation results presented in the main text cannot be attributed to model-inadequacy, i.e., to the possibility that ELC models are not flexible enough to adequately approximate the true profiles of λ and μ over time used in the simulations, for three reasons. First, there is no reason to expect that this type of model inadequacy should lead to zero-inflated extinction rates, instead of simply a mix of over- and under-estimation errors. Second, and more importantly, in all considered simulations the best fitted constrained model matched the data well, based on a Kolmogorov-Smirnov test on the distribution of node ages and based on the fact that each tree's LTT (which contains all information in the context of birth-death models[10,38]) was completely or nearly completely contained within the 95% confidence interval of LTTs generated by the best fitted model. Similarly, the dLTTs of the best fitted models closely resembled the dLTTs of the true scenarios used in the simulations ($R^2 > 0.99$ in all cases, examples in Figures S2A and S2B). Since the dLTT of a birth-death model fully determines its congruence class[10] and the probability distribution of generated trees (when μ is non-negative), this means that the best fitted constrained and unconstrained models indeed converged toward the true congruence class and that the best fitted constrained models would generate trees similar to those of the true diversification scenario; this pattern is fully consistent with our theory and rules out the possibility of serious model inadequacies. Third, as mentioned earlier, there was no significant positive correlation between the considered goodness of fit measures and the relative estimation error ($(\widehat{\mu}_o - \mu_o)/\mu_o$, Figures S2I and S2J), contrary to what would be expected if model inadequacy caused a deflation of extinction rate estimates. Similarly, we did not observe a negative correlation between the considered goodness of fit measures and the modulus of the relative estimation error ($|\widehat{\mu}_o - \mu_o|/\mu_o$, Figures S2K and S2L), contrary to what would be expected if model inadequacy was the cause of erroneous (in any direction) extinction rate estimates.

Checking if empirical tree results result from model inadequacies

Erroneous extinction estimates have previously been attributed to non-modeled variation in diversification rates *between* different lineages[11,14] existing at any given time point. While previous research has not identified any mechanism by which such model inadequacy should generally lead to zero-inflated rate estimates, we nevertheless consider this possibility in our analysis of empirical trees. Specifically, for every tree we examined whether the best fitted model adequately explained not only variation through time (captured by the LTT) but also variation across lineages using parametric bootstrapping and four test statistics: the Sackin[18] and Colless[19] statistics, the distribution of edge lengths using a Kolmogorov-Smirnov test, and the distribution of node ages using a Kolmogorov-Smirnov test. Perhaps not surprisingly, a large fraction of models were rejected based on at least one of these tests, with only 40 out of 109 trees passing all tests — that is to say, many of the empirical datasets were not fully consistent with the fitted homogeneous birth-death models. The majority of rejected models were rejected on the basis of the Sackin and/or Colless test statistics, which are specifically designed to detect tree imbalances, for example caused by rate heterogeneities across lineages, although it is possible that erroneous trees and unresolved polytomies also contributed to bad model fits.[21] For 16 of the 40 remaining trees, the best fitted models exhibited a zero extinction rate, which corresponds to a similar fraction as that observed in the larger tree set. Moreover, we did not observe any significant positive correlation between the statistical significance (P value) of any of the deployed adequacy tests and $\widehat{\mu}_o^*$, which one would expect if model inadequacies were the main cause of deflated extinction rate estimates (Spearman's rank correlations were either negative, or non-significant, see Figures S3C–S3F). Hence, processes leading to deviations from the assumptions of birth-death models do not seem to provide a plausible alternative explanation for the high frequency of zero (or negative, if allowed) extinction rate estimates in these empirical trees.

Please cite this article in press as: Louca and Pennell, Why extinction estimates from extant phylogenies are so often zero, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.04.066

Comparison to other proposed mechanisms for erroneous extinction rate estimates

Various explanations have been proposed for why extinction rates may be misestimated from phylogenies, all of which essentially invoke some type of model inadequacy.[5,11–14] However, the proposed mechanisms do not explain the observation that extinction rate estimates based on extant phylogenies are suspiciously often zero. Indeed, Rabosky[11] and Beaulieu et al.[13] showed that heterogeneity in rates across lineages (a violation of the assumption of homogeneity in birth-death models) often leads to an inflation — rather than deflation — of extinction rate estimates when these are obtained via birth-death model fitting (Figure 3 in Rabosky[11]). Purvis[5] suggested that biased species sampling or incompletely resolved phylogenies could lead to low extinction rate estimates. However, the mechanisms proposed by Purvis[5] are expected to alter extinction estimates by some factor, rather than completely erase any evidence for extinction. Morlon et al.[12] proposed that failing to adequately account for variation in rates over time can lead to erroneous extinction rate estimates. A slow-down of speciation rates over time is expected, for example, under adaptive radiation with diversity-dependent cladogenesis.[39,40] An apparent slow-down of λ near the present is also possible under protracted speciation, which decreases the probability of recent speciation events being detected and included in the phylogeny.[41,42] A true or apparent slow-down in λ would counteract the "pull of the present" expected in the presence of extinction,[38] and if not accounted for, could lead to an underestimation of present-day μ. However, it can be shown using simulations as well as empirical datasets (see our main text) that even when fitting birth-death models accounting for temporal variation, and even if trees are simulated under a birth-death model with homogeneous rates across lineages and no sampling bias, extinction rate estimates are zero-inflated. A zero-inflation of extinction rate estimates even in the absence of any model violations can even be seen in the simulation results by Rabosky[11] and Beaulieu et al.,[13] although these studies did not provide any explanation for that phenomenon. A zero-inflation of extinction rate estimates even without any detectable model violations cannot be explained by the above mechanisms. It is also notable that birth-death models are routinely fit to paleontological data without considering across-lineage heterogeneity and assuming a fairly simple model of time-variability e.g., piecewise constant across different time-bins;[43] and yet typically estimate extinction rates to be similarly high as speciation rates.[9] Also note that our observation that the proportion of zero extinction estimates was similar for those empirical trees that passed all four of our model adequacy tests versus those that didn't, suggesting that model inadequacy was not an important additional cause of zero-inflated extinction estimates in our dataset.

Current Biology, Volume *31*

Supplemental Information

Why extinction estimates from extant phylogenies are so often zero

Stilianos Louca and Matthew W. Pennell

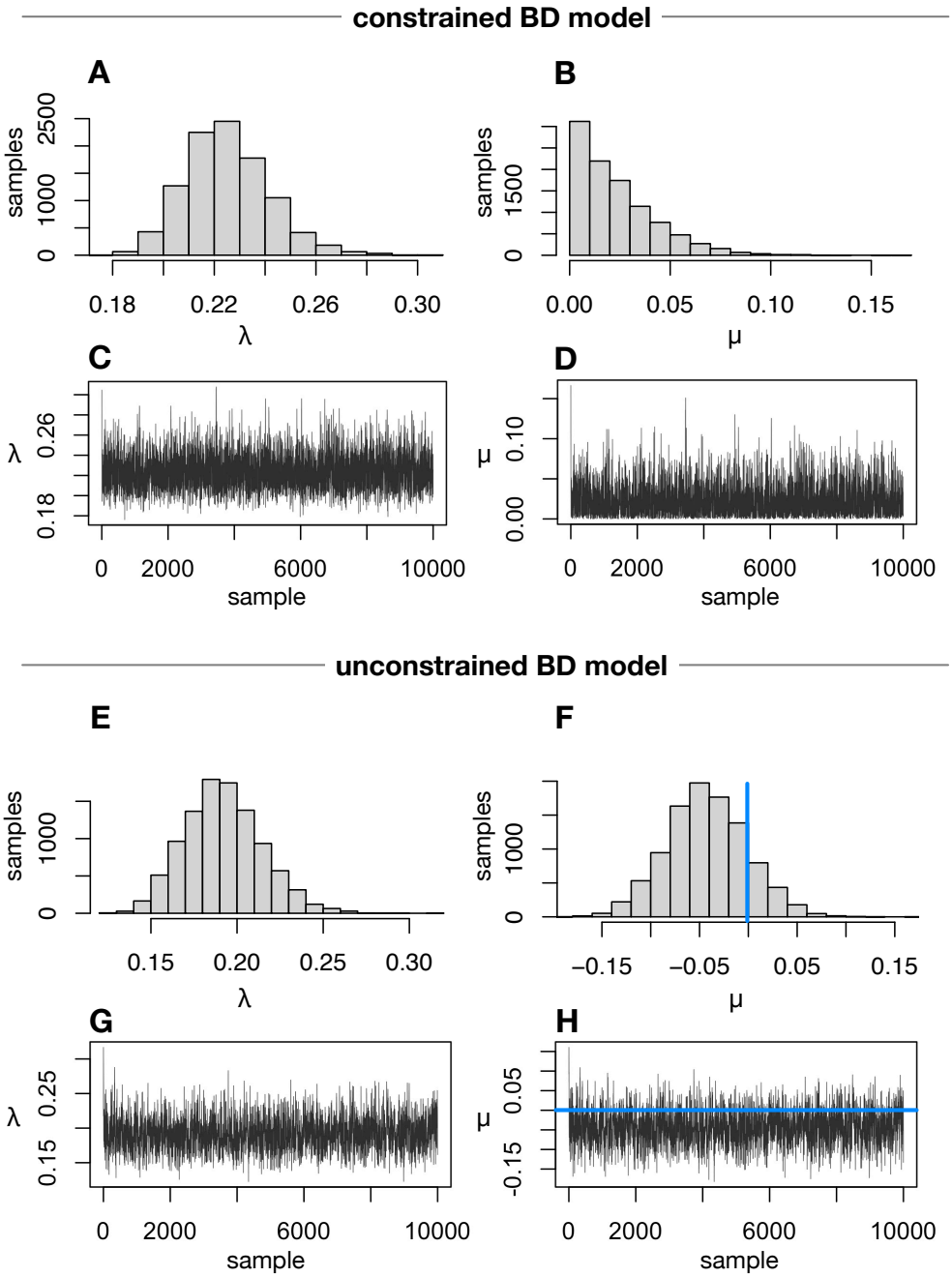

Figure S1: **Fitting constant-rate BD models to the Hummingbird timetree via Bayesian MCMC (related to Figure 1).** (A,B) Posterior distributions of λ (A) and μ (B) sampled through Bayesian MCMC using a constant-rates birth-death model, constrained to non-negative μ, for an empirical timetree of 293 Trochilidae (hummingbird) species (see Methods for details). (C,D) Trace plots of λ and μ. Note the strong skew of μ towards zero. (E–H) Posterior distributions and trace plots of λ and μ sampled through Bayesian MCMC using a constant-rates birth-death model, allowing negative μ, for the same timetree as in A–D. The blue lines in F and H indicate the value zero. Note the concentration of the posterior μ to mostly negative values. A flat prior was used for both λ and μ.

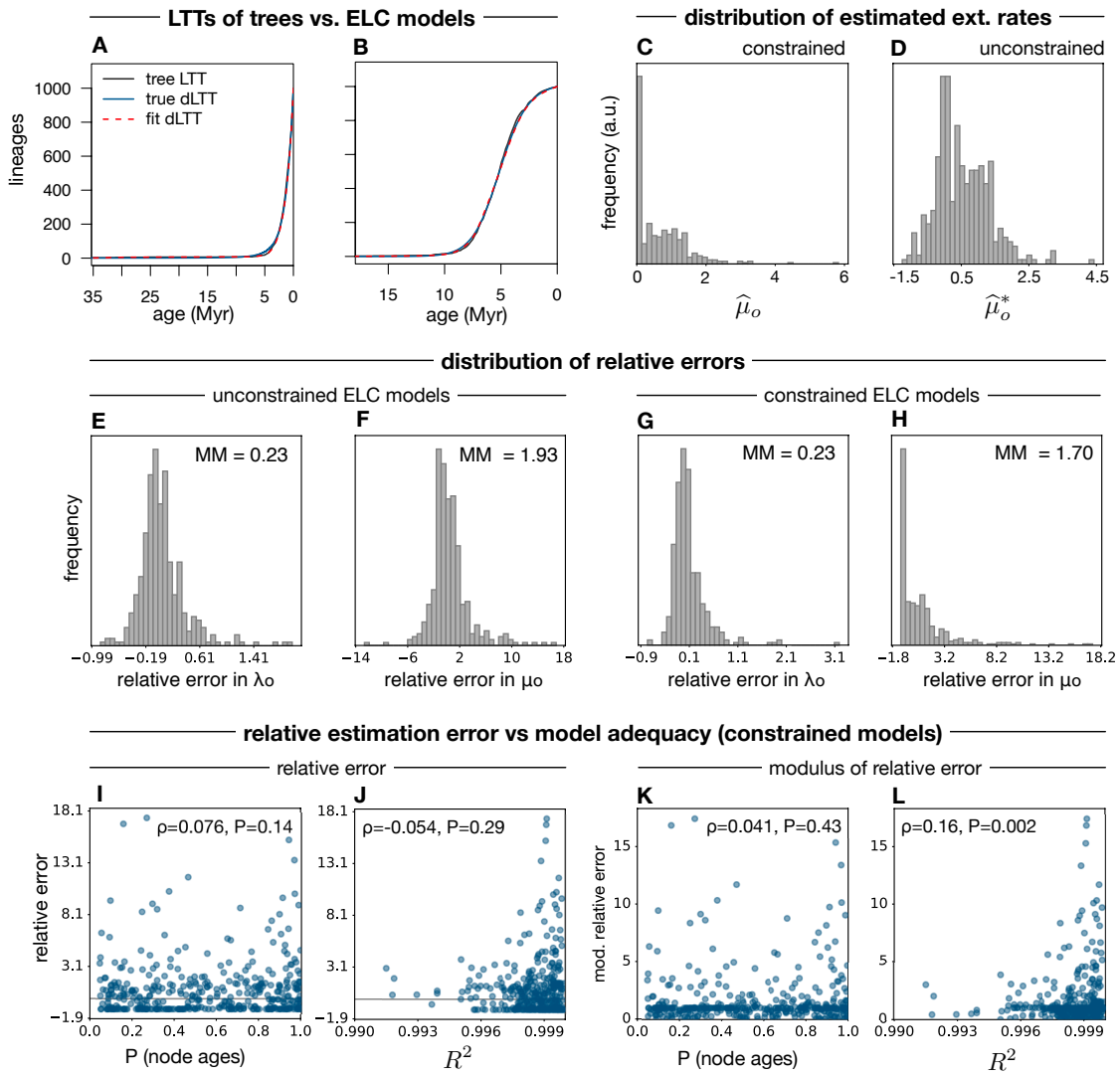

Figure S2: **Details on fitting ELC models to 380 simulated trees (related to Figure 2).** (A,B) Typical lineages-through-time (LTT) curves of timetrees simulated under various hypothetical diversification scenarios (continuous black curves), compared to the deterministic LTTs (dLTTs) of the true generating diversification scenario (blue continuous curves) and the dLTTs of the best fitted constrained ELC models (red dashed curves). In both examples the estimated present-day extinction rate ($\hat{\mu}_o$) was zero. (C) Histogram of present-day extinction rates ($\hat{\mu}_o$) estimated for 380 simulated timetrees (as in Figure 2A) based on the best fitted ELC model, while constraining extinction rates to be non-negative. Observe the clear zero-inflation of present-day extinction rate estimates. (D) Histogram of estimated present-day extinction rates ($\hat{\mu}_o^*$) across the same timetrees, based on the same fitted ELC models as in (C) but allowing for negative extinction rates. (E,F) Distribution of relative errors in (E) the estimated present-day speciation rate ($(\hat{\lambda}_o^* - \lambda_o)/\lambda_o$) and (F) the estimated present-day extinction rate ($(\hat{\mu}_o^* - \mu_o)/\mu_o$), when fitting ELC models to the trees while allowing for negative extinction rates. (G,H) Distribution of relative errors in (G) the estimated present-day speciation rate ($(\hat{\lambda}_o - \lambda_o)/\lambda_o$) and (H) the estimated present-day extinction rate ($(\hat{\mu}_o - \mu_o)/\mu_o$), when fitting ELC models to the trees while constraining μ to be non-negative. The mean modulus (MM) of the estimation errors is indicated in the figures. (I) Statistical significance of the Kolmogorov-Smirnov test for the distribution of node ages (horizontal axis) in the trees under their best fitted constrained ELC models (one point per tree), compared to the relative error of the present-day extinction rate estimated under that model ($(\hat{\mu}_o - \mu_o)/\mu_o$, where μ_o is the present-day extinction rate of the true diversification scenario used in the simulations). The Spearman's rank correlation between the relative error and the Kolmogorov-Smirnov test significance, as well as the two-sided statistical significance of that correlation, are shown inside the figure. (J) Similar to I, but with the horizontal axis showing the fraction of variance (R^2) in the true scenario's dLTT captured by the fitted model's dLTT. (K,L) Similar to (I,J), but sowing the modulus of the relative errors on the vertical axis.

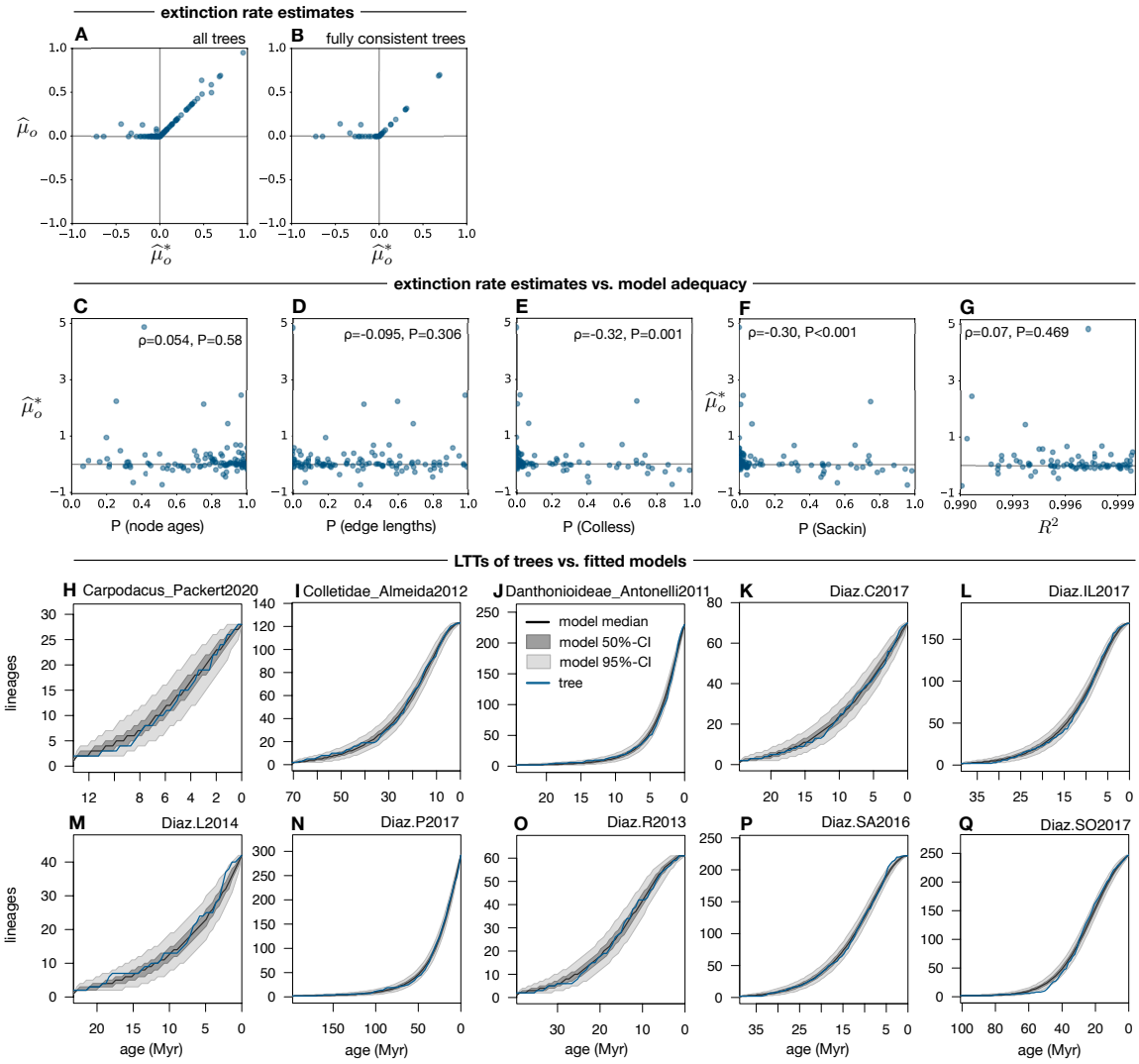

Figure S3: **Details on fitting ELC models to empirical trees (related to Figure 3).** (A,B) Present-day extinction rates estimated from empirical extant timetrees of various taxa (one point per tree), based on the best fitted ELC model. Horizontal axes show the estimated extinction rate while allowing for negative extinction rates ($\hat{\mu}_o^*$), vertical axes show the fitted extinction rate when constrained to non-negative values ($\hat{\mu}_o$). All units are in Myr^{-1}. (A) Showing all 109 trees where the best fitted constrained model adequately explains the distribution of node ages (using a Kolmogorov-Smirnov test). (B) Showing only the 40 trees where the model also adequately explains the distribution of edge lengths (using a Kolmogorov-Smirnov test), the tree's Sacking and Colless imbalance statistics. Only trees with a $\hat{\mu}_o < 1\ \mathrm{Myr}^{-1}$ are shown in order to focus on near-zero and negative rate estimates. (C) Statistical significance of the Kolmogorov-Smirnov test for the node ages (horizontal axis) in empirical trees under their best fitted constrained ELC model (one point per tree), compared to the corresponding $\hat{\mu}_o^*$. The Spearman's rank correlation between $\hat{\mu}_o^*$ and the Kolmogorov-Smirnov test significance, as well as the two-sided statistical significance of that correlation, are shown in the figure. (D) Similar to C, but for the tree's edge lengths. (E) Statistical significance of each tree's Colless statistic (horizontal axis) under the best fitted constrained ELC model, compared to the corresponding $\hat{\mu}_o^*$ (one point per tree). (F) Similar to E, but for the Sackin statistic. (G) Coefficient of determination (R^2, aka. "fraction of explained variance") of the best fitted constrained model's dLTT versus the tree's LTT, compared to the corresponding $\hat{\mu}_o^*$. None of the metrics shown in the horizontal axes in C–G (P-values and R^2) were found to significantly positively correlate with $\hat{\mu}_o^*$, suggesting that model inadequacy unlikely drives the many negative (or zero, if constrained) extinction rate estimates. (H–Q) LTTs for the first 10 out of 39 empirical trees (in alphabetical order) for which the best fitted constrained ELC model adequately described the tree's node distribution (based on a Kolmogorov-Smirnov test) and for which $\hat{\mu}_o = 0$. Blue curves show the LTTs of the trees, black curves show the median LTT generated by the fitted model, and dark and light shades show the 50% and 95% equal-tailed confidence intervals of LTTs generated by the fitted model.