**TITLE:**

A critique of Thompson and Ramirez-Barahona (2023) or: how I learned to stop worrying and love the fossil record

**AUTHORS:**

Eric R. Hagen1

**AFFILIATIONS:**

1Department of Ecology & Evolutionary Biology, University of Toronto, Toronto, ON M5S 3B2

**Abstract**

Earlier this year, a study published in *Biology Letters* by Thompson and Ramirez-Barahona argued that, according to analyses of diversification on two massive molecular phylogenies comprising thousands of flowering plant species, angiosperms do not show evidence of having suffered decreased diversification during the Cretaceous-Paleogene mass extinction. We show here that their study is methodologically and philosophically flawed. We demonstrate that the methods used in their study possess limitations, identified by the authors of those methods, that are highly likely to fail to detect a true mass extinction event using data similar to those analyzed by Thompson and Ramirez-Barahona. Additionally, we use their study as a springboard to examine the relationship between phylogenetic and fossil evidence in diversification studies.

**Background**

Since the introduction of molecular phylogenies and the comparative methods that employ them, practitioners have defied expectations of what was thought possible. For example, when Felsenstein (1985) famously introduced phylogenetic independent contrasts, reviewers were skeptical that there would ever be sufficient robust phylogenies for the method to prove useful (Huey et al. 2019). Another example is the development of methods to estimate extinction rates from phylogenies that are entirely composed of extant taxa; however, this issue is much stickier. Nee et al. (1994) demonstrated that extinction rates can be estimated, and mass extinction events can be inferred, from lineage-through-time (LTT) plots that only contain information about lineage “births.” Estimation of extinction rates from phylogenies of extant taxa proved controversial: Rabosky (2010) went so far as to argue that extinction rates should not even be estimated from molecular phylogenies, while others, such as Beaulieu and O’Meara (2015), have argued that reliable extinction estimates can be gleaned from phylogenies that are sufficiently large using sophisticated methods that incorporate things like rate heterogeneity. Most recently, Louca and Pennell (2020) demonstrated that diversification rates cannot be reliably inferred from LTTs, inspiring widespread doubt in the degree to which inferences from phylogenies are stretched.

However, optimism remains – perhaps too much optimism. Recently, Thompson and Ramírez-Barahona (2023) attempted to detect a signature of the Cretaceous-Paleogene mass extinction on two very large phylogenies of flowering plants. They failed to do so, arguing that this demonstrated the resilience of angiosperms, likely due to their rich adaptations. This finding was discussed in several popular science articles, including in the New York Times. However, their paper contains two issues. The first is that this finding disagrees with several robust analyses of the angiosperm fossil record (Wilf and Johnson 2004), the fern fossil record (i.e., “fern spikes”, Vajda et al. 2001), and the fossil record of phytophagous insects (Labandeira et al. 2002), which display evidence of plants having undergone widespread extinction after the Cretaceous-Paleogene boundary. While disagreeing with previous evidence is not necessarily an indication of an incorrect result, especially since the fossil record is imperfect, the fact that Thompson and Ramirez-Barahona have offered an answer to a paleobotanical problem using extant-only data can reasonably raise suspicion.

This reply investigates the aspects underlying the second issue with their study: overextending the capabilities of analyses of molecular phylogenies. Despite a veritable pile of papers sounding the alarm about limitations of phylogenetic comparative methods (both warranted and unwarranted), they continue to be applied to problems for which they can offer but limited information and ultimately produce suspect results. Here, we argue that the methods employed by Thompson and Ramirez-Barahona (2023) cannot confidently support the conclusion that flowering plants were not subjected to the ravages of the end-Cretaceous extinction. Specifically, we discuss the lack of phylogenetic information deep in the tree as well as pitfalls of the TESS (Höhna et al. 2016) and CoMET (May et al. 2016) models, and we conclude with a brief discussion of the relative merits of phylogenetic comparative methods (PCMs) and fossils for answering questions about extinction events in the distant geologic past.

**Lack of Information**

The supposed ability of Thompson and Ramirez-Barahona to reconstruct the severity of the Cretaceous-Paleogene extinction event on flowering plants from extant species alone relies on the fact that signatures of diversification rate heterogeneity are often retained by LTTs. Nee et al. (1994) note in their seminal paper on the subject that, in all their examples, they assumed 100% sampling, which the phylogenies used by Thompson and Ramirez-Barahona do not have. Another issue not discussed by Nee et al. (1994) is the fact that the evolutionary patterns gleaned from analyses of molecular phylogenies as opposed to those of the fossil record can often differ widely (Marshall 2017). This should not be particularly surprising given the fact that a very small proportion of organisms throughout geologic time are successfully buried and fossilized, preserved to the present, and ultimately excavated by humans (Foote and Miller 2007), as well as the fact that many evolutionary avenues, even ones that produce large clades over time, may ultimately leave no survivors to the present that can be included in molecular phylogenies.

But let’s assume, for the sake of argument, that the tree is roughly accurate in its representation of evolutionary patterns taken by the clade(s) represented, and that it represents all or nearly all the extant species recognized by taxonomists. If this is the case for the phylogenies used by Thompson and Ramirez-Barahona, then they should be able to accurately infer signatures of past mass extinctions at any time along the LTT, right? Not necessarily. As one travels from the tips toward the root of a phylogeny, whether one is performing an ancestral state reconstruction or a diversification study or any other type of phylogenetic comparative analysis, the amount of information available to an investigator deteriorates. O’Meara and Beaulieu (2021) demonstrate this on a simulated tree with one million tips, arguing that analyses and inferences of evolutionary events deep on phylogenetic trees rely on the few lineages that survived to the present and were thus able to be sequenced and included in a molecular phylogeny. This may be a completely unrepresentative example with little information about the evolutionary events that shaped their clade and more inclusive clades in which they are nested.

In the case of Thompson and Ramirez-Barahona (2023), in one of the phylogenies they analyzed, produced by Smith and Brown (2018), only about 0.7% of all speciation events occur at or before 65 mya. Examined in another way, there are only 548 lineages present at the 65 mya time slice, compared to about 77,000 tips in the Smith & Brown tree. While Thompson and Ramirez-Barahona (2023) do include in their discussion an admission of the potential lack of identifiability in any analysis of diversification deep in time, they were not dissuaded from performing (or publishing) their analysis. These facts about the imbalance of information in phylogenies have spurred other scientists to develop and increasingly employ tip rate analyses (e.g., Title and Rabosky 2019; Vasconcelos et al. 2022) that take advantage of the glut of information available near the present, but which unfortunately are unable, at least at this point in time, to provide information about past mass extinction events in the distant past, and thus do not form a viable alternative for analyses like that of Thompson and Ramirez-Barahona (2023).

**Issues with TESS and CoMET**

Thompson and Ramirez-Barahona (2023) fit diversification models to the LTTs of two flowering plant phylogenies using TESS (Höhna et al. 2016) and estimate diversification dynamics on each using the CoMET model (May et al. 2016), which is implemented in the TESS R package. The decision to use these methods to answer their question of whether angiosperms were affected by the KPg mass extinction is a curious one. While the CoMET model is interesting in that it uses compound Poisson process models to distinguish between several different types of diversification rate shifts across phylogenies, the authors clearly state in their discussion that CoMET is not an ideal model for the study conducted by Thompson and Ramirez-Barahona. Specifically, May et al. (2016) claim that, in their simulation study, CoMET almost never showed strong support for the true mass extinction model when the mass extinction was located in the more ancient (i.e., rootward) half of the tree: in these scenarios, the mass extinction model was favored only 8.4% of the time with diversification rate shifts, and only 6% of the time without such shifts. In other words, to quote the authors themselves, “a mass-extinction event that occurs too close to the tips of a study tree will be difficult to detect because too little time will have elapsed for the tree to recover from the event” (May et al. 2016). Since the KPg mass extinction occurs close to the halfway point in time along the Smith and Brown phylogeny, one could reasonably expect Thompson and Ramirez-Barahona (2023) to mention this issue among the other caveats and potential pitfalls they discuss. Alas, they do not. Additionally, May et al. (2016) point out that CoMET inferences of mass extinction event times often do not coincide with the times inferred from fossil record evidence of mass extinctions. This is also unmentioned in the paper.

Tests of the performance of CoMET have previously been performed (see Culshaw et al. 2019, which found high rates of Type II error when mass extinction survival rates were not very low and when trees were not quite large). Therefore, we performed a limited number of tests involving the Smith & Brown tree and simulated trees of similar size.To test the degree to which TESS and CoMET can accurately identify the generating models of simulated trees, we repeated the analysis by Thompson and Ramirez-Barahona (though we cannot be certain we replicated their methods exactly, as the Dryad repository for their study is not [yet?] available to the public, and I have not emailed them to ask any questions) of the Smith and Brown phylogeny (2018). We assumed a sampling fraction of about 22.5%, calculated by dividing the number of tips in the Smith and Brown tree (about 77,000) by the estimate of about 350,000 total angiosperm species published by Paton et al. (2008). After extracting branching times from the tree, we created three simulated trees in TESS to which we could fit our LTT using the following branching-process models: constant birth-death, episodic birth-death, and a birth-death model including a mass extinction. The marginal likelihoods of each model were estimated using stepping-stone sampling in TESS, and the relative support of the models were compared using Bayes factors (see Jeffreys 1961). Although the mass extinction model was decisively favored over episodic birth-death, this analysis yielded extremely strong support for the constant birth-death model over both the episodic and mass extinction models according to Bayes factors exceeding 1.9 million, astronomically higher than the “decisive” threshold of 100 advocated by Jeffreys (1961).

To test the degree to which TESS and CoMET can accurately identify mass extinction events in an analysis like that of Thompson and Ramirez-Barahona (2023), we used TreeSim (Stadler 2010) to simulate phylogenies of similar size and age to the Smith and Brown phylogeny (each possessing a speciation rate of 0.2 and extinction rate of 0.1 prior to the mass extinction event, and lambda of 0.3 and mu of 0.2 post-extinction event). Once again assuming a sampling fraction of 0.22 as in the Smith and Brown tree, we estimated diversification rates using the tess.analysis function (see Fig. 1) to test whether a mass extinction would be detected by the CoMET model. We opted for the function to run 10,000 iterations and to estimate hyperpriors empirically rather than setting them manually.

Our analyses were successfully able to detect mass extinction on a phylogeny possessing 70,000 extant tips (originally with 73,062 additional extinct taxa) and a mass extinction at 66 mya through which only 10% of lineages survived. After increasing the extinction rate from 0.2 to 0.25, TESSS also succeeded on a phylogeny of 70,000 extant tips (originally with 71,807 additional extinct taxa) with a root age of 143.2 mya, a mass extinction at 66 mya which 20% of lineages survived (the estimate calculated by Johnson 1992), and an extinction rate raised from 0.2 to 0.25. However, as the percentage of lineages that survive the extinction event increases, the signal disappears. No mass extinction was detected on a phylogeny of 142,833 extant/extinct tips with a root age of 149.68 mya in which 30% of species survive the KPg, nor on a 141,651-tip tree with a root age of 152.76 mya in which 40% of species survive the KPg (see Fig. 1).

A group of graphs showing different types of exclamation

Description automatically generated with medium confidence

*Figure 1.* CoMET fails to detect a mass extinction at 66 mya on a simulated phylogeny of similar size and age to the Smith and Brown tree used in the analysis by Thompson and Ramirez-Barahona (2023). In this simulation, 40% of lineages survive the mass extinction; CoMET successfully detected mass extinction events on similar simulated trees with lower survival rates.

**How I Learned to Stop Worrying and Love the Fossil Record**

We have demonstrated that Thompson and Ramirez-Barahona (2023) erred in their methodological choices and interpretations. Reading the peer review history files reveals that the anonymous reviewers who handled the manuscript for *Biology Letters* did not make substantive critiques of the manuscript’s methods as described here. Instead, the reviewers mainly took issue with the authors’ seemingly unwarranted distrust of the flowering plant fossil record. While we are not aware of the exact specialties of the two reviewers, it seems likely that both were paleontologists unfamiliar with PCMs. This is an issue that has been discussed from comparative biology (Griffin and Yapuncich 2015) to pure mathematics (Bordg 2021): as disciplines become more specialized and adopt more complex tools, journals ought to exercise greater caution in assigning reviewers. We believe that Thompson and Ramirez-Barahona (2023) would not have been published in its current form had at least one reviewer been a specialist on PCMs.

While our issues with the paper are methodological ones, we also think it’s likely that this paper would not have appeared if the authors were not so skeptical of the fossil record. Thompson and Ramirez-Barahona mention numerous aspects of fossil evidence indicating a mass extinction for plants at the KPg boundary, and they argue for the integration of fossil and phylogenetic evidence in addressing their research question. However, they conclude their paper by arguing that the contradiction they observed between their results and fossil evidence can be partially explained by the fossil record recording species-level extinction while the phylogenies they analyzed record the survival of higher taxa. We disagree with this interpretation. First, this argument contradicts their introductory remarks that studies which assign plant fossils to species-level taxa, an extremely difficult task, find no variation in extinction rates across the KPg. Second, we believe that the proper conclusion would have been deferring to the fossil evidence and pleading for further phylogenetic research (to say nothing of their study’s methodological problems). This is largely practiced in the literature surrounding estimates of the origin of angiosperms, where molecular clock analyses based on extant taxa frequently estimate much earlier origination times than fossil evidence (see Coiro et al. 2019). On this front, the trustworthiness of the fossil record has been aggressively defended (see Herendeen et al. 2017), whereas molecular clock research is, by the admission of its architects, still evolving (e.g., Smith et al. 2010; Beaulieu et al. 2015; Brown and Smith 2018).

Philosophically, it may seem suspect to favor the fossil record over phylogenetic evidence: fossil data *can* be biased, and inferences about the past from phylogenies of extant taxa *can* be accurate. But we believe that studies like that of Thompson and Ramirez-Barahona fall into a conceptual pitfall described by O’Meara and Beaulieu (2021), who advise biologists that “it is important to recognize that *our methods are better suited for using the past to learn about the present survivors, not using the present survivors to learn about the past*” (p. 21; emphasis in the original). Instead of attempting to “corroborate” fossil evidence with phylogenetic analyses, it may be time for comparative biologists to stop worrying and love the fossil record.

**References**

1. Beaulieu, J.M., and B.C. O’Meara. 2023. Fossils do not substantially improve, and may even harm, estimates of diversification rate heterogeneity. *Systematic Biology* 72: 50-61.
2. Beaulieu, J.M., B.C. O’Meara, P. Crane, and M.J. Donoghue. 2015. Heterogeneous rates of molecular evolution and diversification could explain the Triassic age estimate for angiosperms. *Systematic Biology* 64: 869-878.
3. Bordg, A. 2021. A replication crisis in mathematics? *The Mathematical Intelligencer* 43: 48-52.
4. Brown, J.W., and S.A. Smith. 2018. The past sure is tense: on interpreting phylogenetic divergence time estimates. *Systematic Biology* 67: 340-352.
5. Coiro, M., J.A. Doyle, and J. Hilton. 2019. How deep is the conflict between molecular and fossil evidence on the age of angiosperms? *New Phytologist* 223: 83-99.
6. Culshaw, V., T. Stadler, and I. Sanmartín. 2019. Exploring the power of Bayesian birth-death skyline models to detect mass extinction events from phylogenies with only extant taxa. *Evolution* 73: 1133-1150.
7. Felsenstein, J. 1985. Phylogenies and the comparative method. *The American Naturalist* 125: 1-15.
8. FitzJohn, R.G. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution* 3: 1084-1092.
9. Foote, M., and A.I. Miller. 2007. Principles of Paleontology (3rd edition). Freeman: New York, USA.
10. Griffin, R.H., and G.S. Yapuncich. 2015. The independent evolution method is not a viable phylogenetic comparative method. *PLoS ONE* 10: e0144147.
11. Herendeen, P.S., E.M. Friis, K.R. Pedersen, and P.R. Crane. 2017. Palaeobotanical redux: revisiting the age of the angiosperms. *Nature Plants* 3: 1-8.
12. Höhna, S., M.R. May, and B.R. Moore. 2016. TESS: an R package for efficiently simulating phylogenetic trees and performing Bayesian inference of lineage diversification rates. *Bioinformatics* 32: 789-791.
13. Huey, R.B., T. Garland, and M. Turelli. 2019. Revisiting a key innovation in evolutionary biology: Felsenstein’s “Phylogenies and the comparative method”. *The American Naturalist* 193: 755-772.
14. Jeffreys, H. 1961. Theory of Probability (3rd edition). Oxford University Press: Oxford, UK.
15. Labandeira, C.C., K.R. Johnson, and P. Wilf. 2002. Impact of the terminal Cretaceous event on plant-insect associations. *Proceedings of the National Academy of Sciences* 99: 2061-2066.
16. Louca, S., and M.W. Pennell. 2020. Extant timetrees are consistent with a myriad of diversification histories. *Nature* 580: 502-505.
17. Marshall, C.R. 2017. Five palaeobiological laws needed to understand the evolution of the living biota. *Nature Ecology & Evolution* 1: 0165.
18. May, M.R., S. Höhna, and B.R. Moore. 2016. A Bayesian approach for detecting the impact of mass-extinction events on molecular phylogenies when rates of lineage diversification may vary. *Methods in Ecology and Evolution* 7: 947-959.
19. Nee, S., R.M. May, and P.H. Harvey. 1994. The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society B* 344: 305-311.
20. O’Meara, B.C., and J.M. Beaulieu. 2021. Potential survival of some, but not all, diversification methods. *EcoEvoRxiv* doi: 10.32942/osf.io/w5nvd
21. Paton, A.J., N. Brummitt, R. Govaerts, K. Harman, S. Hinchcliffe, B. Allkin, and E.N. Lughadha. 2008. Towards Target 1 of the Global Strategy for Plant Conservation: a working list of all known plant species—progress and prospects. *Taxon* 57: 602-611.
22. Smith, S.A., J.M. Beaulieu, and M.J. Donoghue. 2010. An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. *Proceedings of the National Academy of Sciences* 107: 5897-5902.
23. Smith, S.A., and J.W. Brown. 2018. Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany* 105: 302-314.
24. Stadler, T. 2010. TreeSim in R—simulating trees under the birth-death model. [R package]
25. Thompson, J.B., and S. Ramírez-Barahona. 2023. No phylogenetic evidence for angiosperm mass extinction at the Cretaceous–Palaeogene (K-Pg) boundary. *Biology Letters* 19: 20230314.
26. Title, P.O., and D.L. Rabosky. 2019. Tip rates, phylogenies and diversification: what are we estimating, and how good are the estimates? *Methods in Ecology and Evolution* 10: 821-834.
27. Vajda, V., J.I. Raine, and C.J. Hollis. 2001. Indication of global deforestation at the Cretaceous-Tertiary boundary by New Zealand fern spike. *Science* 294: 1700-1702.
28. Vasconcelos, T., B.C. O’Meara, and J.M. Beaulieu. 2022. A flexible method for estimating tip diversification rates across a range of speciation and extinction scenarios. *Evolution* 76: 1420-1433.
29. Wilf, P., and K.R. Johnson. 2004. Land plant extinction at the end of the Cretaceous: a quantitative analysis of the North Dakota megafloral record. *Paleobiology* 30: 347-368.