

# **Is photosynthetic capacity associated with diversification of C<sub>3</sub> plants?**

Christopher D. Muir<sup>1,\*</sup>

<sup>1</sup> Department of Botany, University of Wisconsin, Madison, WI, USA 53706

\* Corresponding author: Christopher D. Muir, [cdmuir@wisc.edu](mailto:cdmuir@wisc.edu)

Author contributions: CDM conceived of the ideas and wrote the manuscript.

Data availability statement: Data are available from previously published sources. The code to reproduce analyses is available on GitHub (<https://github.com/cdmuir/photosynthesis-diversification>) and archived on Zenodo (<https://doi.org/10.5281/zenodo.17835088>).

Word count: 63 in Abstract; 1,038 in Main text

Running head: Photosynthetic capacity and diversification

Keywords: ancestral state reconstruction, diversification, macroevolution, photosynthesis, phylogenetic comparative methods

## Abstract

Schweiger & Schweiger (2024) recently proposed that evolution of greater photosynthetic capacity in  $C_3$  plant species may be linked to increased species diversification. This conclusion is premature because their methods cannot reliably estimate diversification rate or ancestral trait values. Reanalyzing the data set using model-based phylogenetic comparative methods reveals no evidence that photosynthetic capacity is associated with diversification rates among  $C_3$  plant genera.

## Main text

Phylogenetic comparative methods (PCMs) are widely used to infer historical patterns of species diversification and phenotypic evolution based on the relationships among extant species and their trait values (Pennell & Harmon 2013). Recently Schweiger & Schweiger (2024) purport to show that greater photosynthetic capacity is linked to greater diversification among  $C_3$  vascular land plants over the past 80 million years. Specifically, they argue that diversification rates increased along with photosynthetic capacity ( $V_{\text{cmax}}$  and  $J_{\text{max}}$ ) as atmospheric  $\text{CO}_2$  decreased throughout the Miocene. This conclusion is premature because the PCMs Schweiger & Schweiger (2024) use to estimate diversification rates and ancestral states do not account for and are highly sensitive to incomplete taxon sampling.

### *Diversification analysis*

Diversification rates are typically inferred using the distribution of branch lengths in a molecular phylogeny. The reconstructed evolutionary process, even with complete sampling of extant taxa and perfect information, cannot infer the presence of lineages that left no descendants because of extinction (Nee *et al.* 1994b). Incomplete taxon sampling acts like a recent extinction event because nodes leading to unsampled taxa cannot be inferred. Valid inference of diversification rates must therefore condition on both extinction and incomplete sampling. Schweiger & Schweiger (2024) analyze a molecular phylogeny of vascular land plants containing 232 species, 0.065% of the 359,208 currently accepted species as of December 2024 (Kindt 2020). They estimate diversification rate as  $\log(n)/t$ , where  $n$  is the number of tips and  $t$  is the stem group age (Magallon & Sanderson 2001). Since apparent  $n$  is a small fraction of its true value, this statistic cannot estimate diversification rate. The Kendall-Moran estimate of speciation rate used by Schweiger & Schweiger (2024) works for completely sampled phylogenies with no extinction (Baldwin & Sanderson 1998), an untenable assumption in this case. Schweiger & Schweiger (2024) compound these errors by pruning tips and branches younger than each time interval for which they estimate diversification rates. Thus, at each interval they artificially remove nodes in the phylogeny, further biasing downward estimates of diversification.

Random incomplete sampling biases downward estimates of diversification rates in the recent past (Nee *et al.* 1994a). The bias caused by random incomplete sampling is opposite the pattern Schweiger & Schweiger (2024) observe of more rapid diversification toward the present. Apparent increases in diversification in younger clades is a general phenomenon not confined to  $C_3$  plants (Harmon *et al.*

2021) that may be a result of bias toward sampling species-rich clades (Louca *et al.* 2022). Reliable inference must address both incomplete and nonrandom sampling.

### *Ancestral trait analysis*

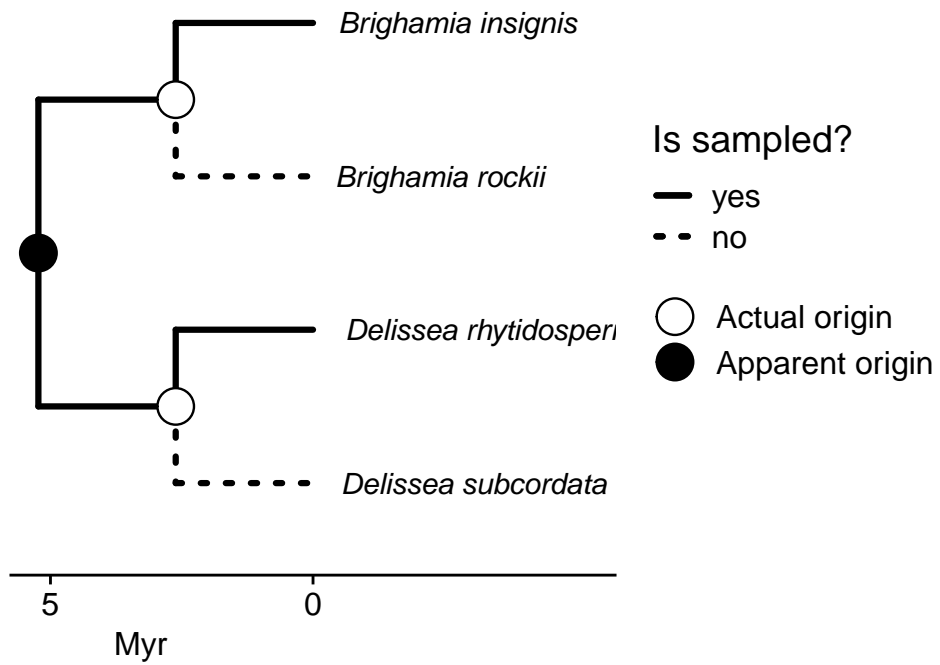
Unobserved ancestral states must be reconstructed using a probabilistic model of trait evolution, but inferred states are often impossible to independently verify (but see Polly 2001). There is a high likelihood that estimates from misspecified models are misleading (Revell 2025). Schweiger & Schweiger (2024) do not estimate ancestral states based on a trait model, but instead calculate it as the “median trait values from all species that originated in the respective time period,” where time period refers to a 5-million-year interval. There are two fundamental errors with this approach. First, because the phylogeny contains a small fraction of all species, the origin time of a species (when it split from its sister species) will be biased upward by an unpredictable amount (Figure 1a). The nodes used by Schweiger & Schweiger (2024) are on average 315% older than and modestly correlated with ( $\log\text{-log } r^2 = 0.43$ ) the nodes separating nearest species in the unpruned Zanne *et al.* (2014) tree, which itself contains 31,389 species, representing only 8.7% of described vascular land plant species.

Second, the method for estimating average trait values in past intervals implies strong and unrealistic assumptions about trait evolution, is inappropriate for a sparsely sampled phylogeny, and is highly sensitive to small differences in taxon sampling. Even if species’ origination dates were inferred correctly, estimating ancestral states by assuming they are equal to the state of the extant descendant implies all trait evolution occurs during speciation events (cladogenic) and not along lineages (anagenic). Assuming cladogenic evolution is especially problematic in this case because most nodes in the phylogeny are not sampled, thereby missing most cladogenic events when trait change would occur by their implied assumptions. A median of 7 nodes per species were pruned from the Zanne *et al.* (2014) phylogeny in their analysis, but this is a severe undercount of missing nodes because of both incomplete sampling in the unpruned tree and extinction. Finally, estimates of past trait values using this method are highly sensitive to small changes in taxon sampling. A minor change in sampling can result in quite different estimates even with the same underlying evolutionary scenario (see Figure 2 for a worked example).

### *No evidence that photosynthetic capacity is linked to diversification in seed plants*

Several PCMs have been developed in the past decade to estimate whether diversification rates vary among lineages and through time in association with traits (Martínez-Gómez *et al.* 2024; Morlon *et al.* 2024). These methods account for extinction and incomplete sampling and therefore can, in principle, test whether diversification is associated with traits like  $V_{\text{cmax}}$  or  $J_{\text{max}}$ . When extinction and incomplete sampling are accounted for using BAMM version 2.5.0 (Rabosky 2014) on a nearly-complete phylogeny of seed plant genera (Dimitrov *et al.* 2023), there is no evidence that speciation or net diversification rates are associated with  $V_{\text{cmax}}$  or  $J_{\text{max}}$  (Figure 3) based on permutation tests developed by Rabosky & Huang (2016) (see Supporting Information). This method answers a slightly different question than that posed by Schweiger & Schweiger (2024) because it tests for an association between tip rates and traits, not ancestral rates and traits. However, the tip rates of many genera encompass much of

A)



B)

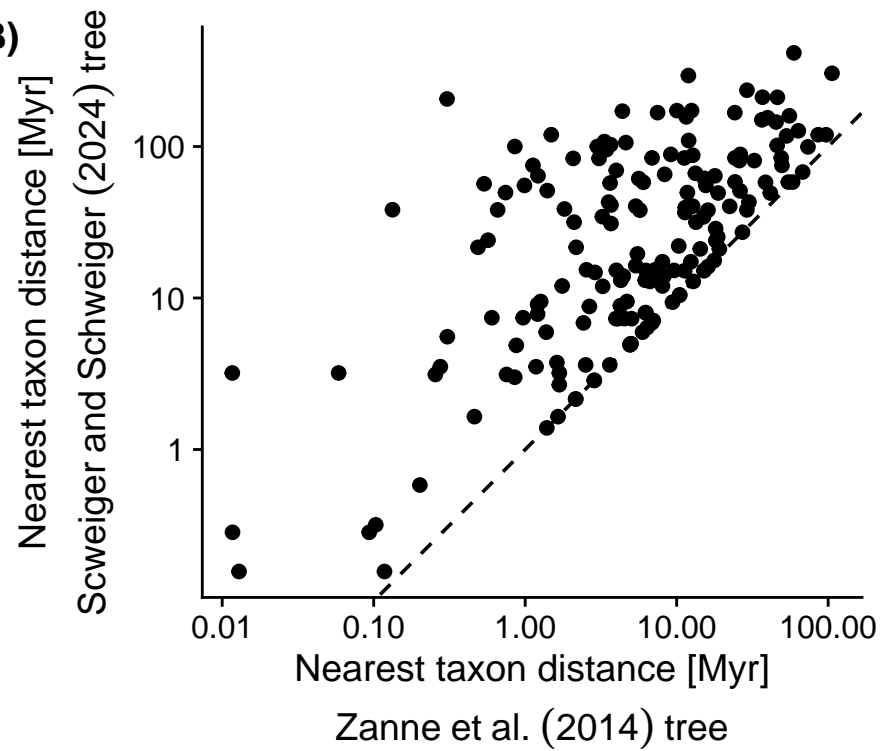


Figure 1: (Caption next page.)

Figure 1: (Previous page). The method Schweiger & Schweiger (2024) use to infer when a species originated is misleading when taxon sampling is incomplete. (A) For example, the Hawaiian lobeliads *Brighamia insignis* A. Gray and *Delissea rhytidosperra* H. Mann in their data set are inferred by this method to have “originated” 5.23 mya, as shown in the phylogeny above. This apparent origin (black point on more ancient node) is an artifact of more closely related congeneric species (*B. rockii* H. St. John and *D. subcordata* Gaudich.) not being sampled (dashed branches leading to unsampled species). The node of the apparent origin represents the most recent common ancestor of these species, but both species actually diverged from more closely related species much more recently (white points). The timing of divergence from putative sister species is approximate and for illustration purposes only (based on Givnish *et al.* 2009). (B) The consequences of this methodological choice are that the divergence times used by Schweiger & Schweiger (2024) (*y*-axis) are biased upward by an average of 315% and modestly correlated ( $r^2 = 0.43$  on log-log scale) with the divergence times estimated from the full phylogeny (*x*-axis) from Zanne *et al.* (2014). Those divergence times are themselves only loosely related to the true divergence time between sister species because of mostly incomplete sampling even in the unpruned Zanne *et al.* (2014) phylogeny.

the Miocene when global CO<sub>2</sub> declined and tip rates suffer from fewer identifiability issues (Louca & Pennell 2020). Alternative methods not considered here are CLaDS2 (Barido-Sottani & Morlon 2023) and QuaSSE (FitzJohn 2010). While these PCMs account for extinction and incomplete sampling, no method performs well when taxon sampling is extremely sparse, samples are biased, or when making inferences about the deep past from extant taxa. Certain questions may be better suited to specific clades with (near-)complete sampling (Donoghue & Edwards 2019), well resolved and fossil-calibrated nodes, and to relatively recent geological epochs.

## Acknowledgments

Andreas Schweiger and two anonymous reviewers provided constructive feedback on an earlier draft of this manuscript. CDM is supported by the National Science Foundation (DEB-2131817). Support for this research was provided by the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison with funding from the Wisconsin Alumni Research Foundation. The author declares no conflicts of interest.

## Hypothetical phylogeny of four species

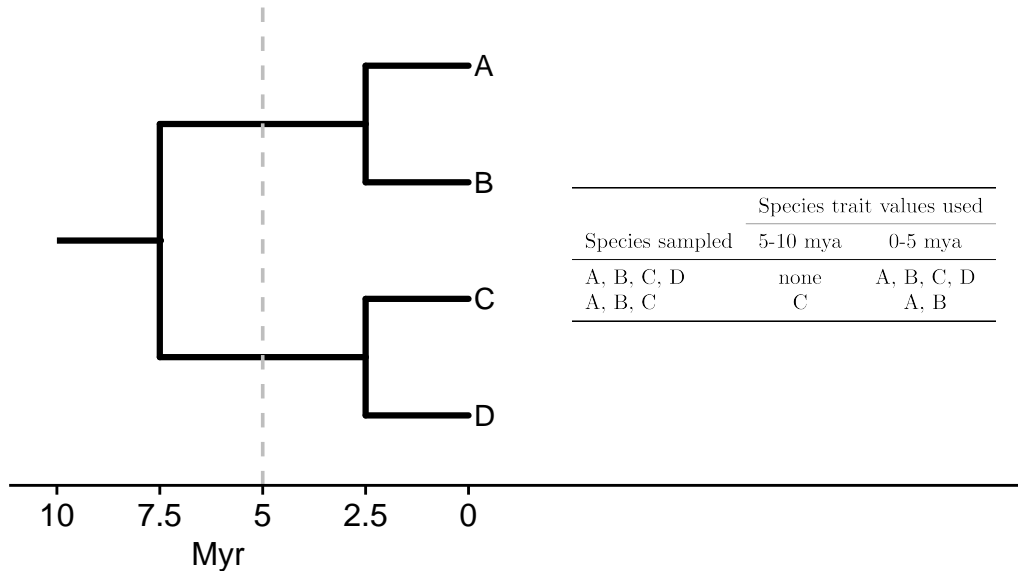


Figure 2: The method Schweiger & Schweiger (2024) use to estimate ancestral trait values is sensitive to small differences in taxon sampling. Assume we know the true phylogeny of four species A, B, C, and D (left). The species A and B are sister species that split 2.5 mya, likewise for C and D. The common ancestor of all four split 7.5 mya. If we sampled all four species, their trait values would contribute to the estimated average trait value in the interval between 0 and 5 mya. This clade would not contribute to the estimated average trait value in the interval between 5 and 10 mya because the species did not originate in that interval. However, it is clear that the ancestors of these species must have existed during that interval. Now consider the effect of incomplete sampling. If species D were not sampled, then the trait values for A and B would contribute to the 0-5 mya interval and species C would contribute to the 5-10 mya interval. The species used to calculate average trait values at different time intervals changes depending on whether all species are sampled or a subset are sampled (right), even though the underlying biological reality is unchanged.

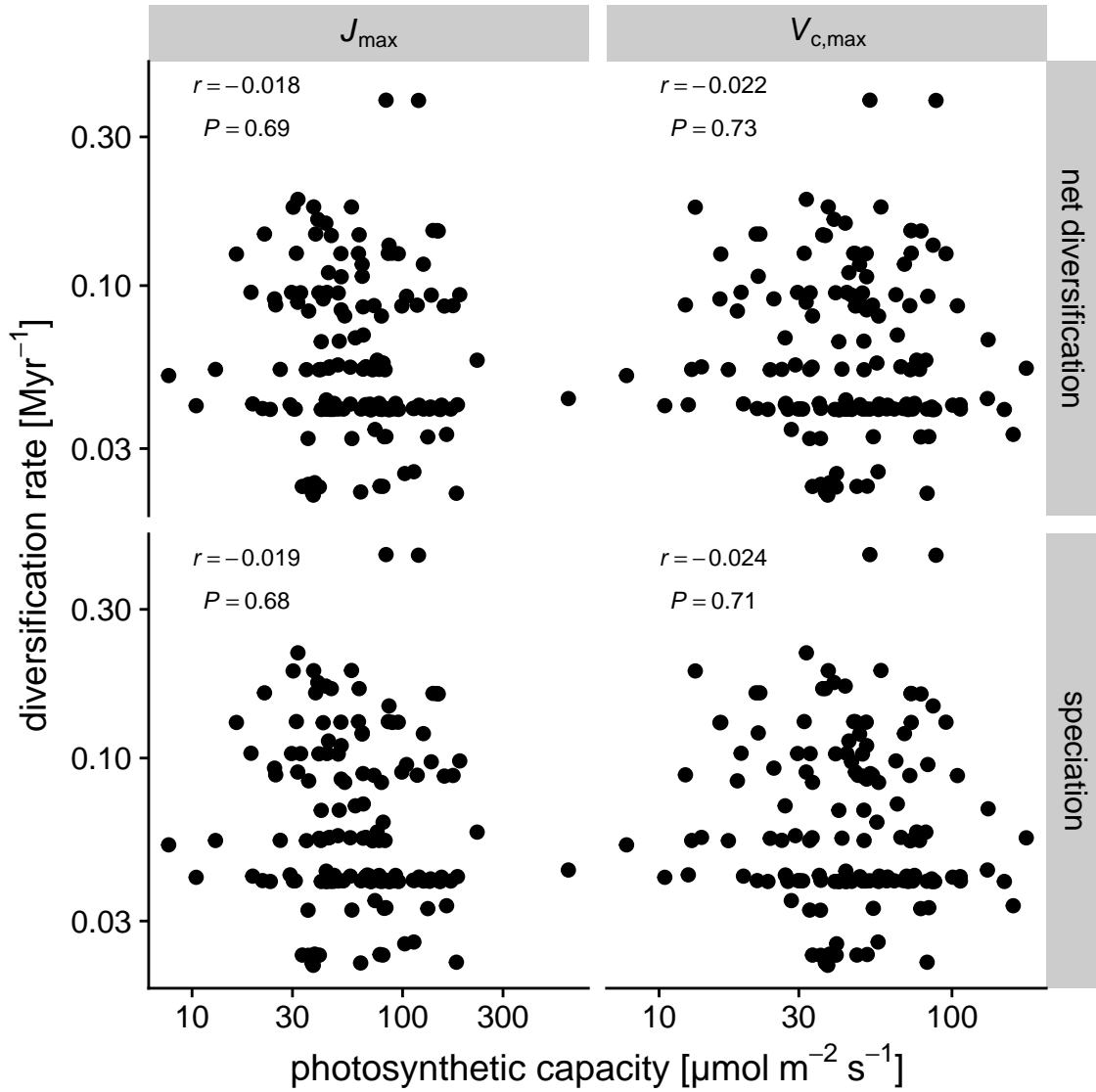


Figure 3: No evidence that photosynthetic capacity explains variation in speciation or net diversification rates among  $C_3$  seed plant genera. Net diversification (top row) and speciation (bottom row) rates estimated using BAMM version 2.5.0 on a nearly-complete phylogeny of seed plant genera (Dimitrov *et al.* 2023) are not significantly associated with average  $J_{\text{max}}$  (left column) or  $V_{\text{c,max}}$  (right column) based on STRAPP permutation tests (Rabosky & Huang 2016). Both axes are log-scaled. The estimated Pearson product moment correlation  $r$  and permutation based  $P$ -values from two-tailed tests are in the upper-left of each facet.

## References

- Baldwin, B.G. & Sanderson, M.J. (1998). [Age and rate of diversification of the Hawaiian silversword alliance \(Compositae\)](#). *Proceedings of the National Academy of Sciences*, 95, 9402–9406.
- Barido-Sottani, J. & Morlon, H. (2023). [The ClaDS rate-heterogeneous birth–death prior for full phylogenetic inference in BEAST2](#). *Systematic Biology*, 72, 1180–1187.
- Dimitrov, D., Xu, X., Su, X., Shrestha, N., Liu, Y., Kennedy, J.D., *et al.* (2023). [Diversification of flowering plants in space and time](#). *Nature Communications*, 14, 7609.
- Donoghue, M.J. & Edwards, E.J. (2019). [Model clades are vital for comparative biology, and ascertainment bias is not a problem in practice: A response to Beaulieu and O’Meara \(2018\)](#). *American Journal of Botany*, 106, 327–330.
- FitzJohn, R.G. (2010). [Quantitative Traits and Diversification](#). *Systematic Biology*, 59, 619–633.
- Givnish, T.J., Millam, K.C., Mast, A.R., Paterson, T.B., Theim, T.J., Hipp, A.L., *et al.* (2009). [Origin, adaptive radiation and diversification of the Hawaiian lobeliads \(Asterales: Campanulaceae\)](#). *Proceedings of the Royal Society B: Biological Sciences*, 276, 407–416.
- Harmon, L.J., Pennell, M.W., Henao-Diaz, L.F., Rolland, J., Siple, B.N. & Uyeda, J.C. (2021). [Causes and Consequences of Apparent Timescaling Across All Estimated Evolutionary Rates](#). *Annual Review of Ecology, Evolution, and Systematics*, 52, 587–609.
- Kindt, R. (2020). [WorldFlora: An R package for exact and fuzzy matching of plant names against the World Flora Online taxonomic backbone data](#). *Applications in Plant Sciences*, 8, e11388.
- Louca, S., Henao-Diaz, L.F. & Pennell, M. (2022). [The scaling of diversification rates with age is likely explained by sampling bias](#). *Evolution*, 76, 1625–1637.
- Louca, S. & Pennell, M.W. (2020). [Extant timetrees are consistent with a myriad of diversification histories](#). *Nature*, 580, 502–505.
- Magallon, S. & Sanderson, M.J. (2001). [Absolute diversification rates in angiosperm clades](#). *Evolution*, 55, 1762–1780.
- Martínez-Gómez, J., Song, M.J., Tribble, C.M., Kopperud, B.T., Freyman, W.A., Höhna, S., *et al.* (2024). [Commonly used Bayesian diversification methods lead to biologically meaningful differences in branch-specific rates on empirical phylogenies](#). *Evolution Letters*, 8, 189–199.
- Morlon, H., Andréoletti, J., Barido-Sottani, J., Lambert, S., Perez-Lamarque, B., Quintero, I., *et al.* (2024). [Phylogenetic Insights into Diversification](#). *Annual Review of Ecology, Evolution, and Systematics*, 55, 1–21.
- Nee, S., Holmes, E.C., May, R.M. & Harvey, P.H. (1994a). [Extinction rates can be estimated from molecular phylogenies](#). *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 344, 77–82.
- Nee, S., May, R.M. & Harvey, P.H. (1994b). [The Reconstructed Evolutionary Process](#). *Philosophical Transactions: Biological Sciences*, 344, 305–311.
- Pennell, M.W. & Harmon, L.J. (2013). [An integrative view of phylogenetic comparative methods: Connections to population genetics, community ecology, and paleobiology](#). *Annals of the New York Academy of Sciences*, 1289, 90–105.
- Polly, P.D. (2001). [Paleontology and the Comparative Method: Ancestral Node Reconstructions versus Observed Node Values](#). *The American Naturalist*, 157, 596–609.



- Rabosky, D.L. (2014). [Automatic Detection of Key Innovations, Rate Shifts, and Diversity-Dependence on Phylogenetic Trees](#). *PLoS ONE*, 9, e89543.
- Rabosky, D.L. & Huang, H. (2016). [A Robust Semi-Parametric Test for Detecting Trait-Dependent Diversification](#). *Systematic Biology*, 65, 181–193.
- Revell, L.J. (2025). [Ancestral State Reconstruction of Phenotypic Characters](#). *Evolutionary Biology*, 52, 1–25.
- Schweiger, A.H. & Schweiger, J.M.-I. (2024). [Significant Links Between Photosynthetic Capacity, Atmospheric CO<sub>2</sub> and the Diversification of C<sub>3</sub> Plants During the Last 80 Million Years](#). *Ecology Letters*, 27, e14523.
- Zanne, A.E., Tank, D.C., Cornwell, W.K., Eastman, J.M., Smith, S.A., FitzJohn, R.G., *et al.* (2014). [Three keys to the radiation of angiosperms into freezing environments](#). *Nature*, 506, 89–92.

## Supporting Information

### **Estimating the relationship between photosynthetic capacity and diversification rates using BAMM**

The speciation and net diversification rates estimated by Dimitrov *et al.* (2023) were unavailable from the authors, but I reanalyzed their molecular phylogeny of seed plant genera using their sampling fractions and following the same procedures in BAMM as described in their paper. For the STRAPP analysis, I estimated the correlation between log-transformed rates and traits using 1000 permutations to generate the null distribution of the Pearson product-moment correlation. The  $V_{c,max}$  and  $J_{max}$  trait data from Schweiger & Schweiger (2024) were averaged at the genus level when more than one species per genus was present. This was necessary because Dimitrov *et al.* (2023) estimated a single diversification rate per genus. Although this method uses macroevolutionary models to account for extinction and incomplete sampling, the results should be interpreted with extreme caution. For example, it does not account for uncertainty in the phylogeny, the species sampling is extremely sparse, variation within genera is ignored, and temporal variation in diversification rates may not be identifiable from extant taxa only (Louca & Pennell 2020).