Seed size and its rate of evolution correlate with species diversification across angiosperms

Short title: Seed size and angiosperm diversification

J. Igea[1,4], E. F. Miller[1,2], A. S. T. Papadopulos[3], A. J. Tanentzap[1]

Author affiliations:

[1] Department of Plant Sciences, University of Cambridge, Downing St, Cambridge, CB2 3EA, UK

[2] Department of Zoology, University of Cambridge, Downing St, Cambridge, CB2 3EJ, UK

[3] Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, TW9 3AB, UK.

[4]*Corresponding author:*

Dr. Javier Igea

Department of Plant Sciences, University of Cambridge, Downing St, Cambridge,
CB2 3EA, UK

ji247@cam.ac.uk

Abstract

Species diversity varies greatly across the different taxonomic groups that comprise the Tree of Life (ToL). This imbalance is particularly conspicuous within angiosperms, but is largely unexplained. Seed mass is one trait that may help clarify why some lineages diversify more than others because it confers adaptation to different environments, which can subsequently influence speciation and extinction. The rate at which seed mass changes across the phylogeny may also be linked to diversification by increasing reproductive isolation and allowing access to novel ecological niches. However, the magnitude and direction of the association between seed mass and diversification has not been assessed across the angiosperm phylogeny. Here, we show that absolute seed size and the rate of change in seed size are both associated with variation in diversification rates. Based on the largest available angiosperm phylogenetic tree, we found that smaller-seeded plants had higher rates of diversification, possibly due to improved colonisation potential. The rate of phenotypic change in seed size was also strongly positively correlated with speciation rates, providing rare, large-scale evidence that rapid morphological change is associated with species divergence. Our study now reveals that variation in morphological traits and, importantly, the rate at which they evolve can contribute to explaining the extremely uneven distribution of diversity across the ToL.

Introduction

Angiosperms are one of the most species-rich clades on Earth and have dominated terrestrial plant communities since the Late Cretaceous [1]. The astounding diversity of flowering plants is distributed extremely unevenly across the ToL. Each of the five most species-rich angiosperm families contains >10,000 species while more than 200 families contain <100 species each [2]. An enduring pursuit in evolutionary biology is to explain this uneven distribution of biodiversity, not only in angiosperms, but also across the whole ToL [3].

Biological traits offer one way to explain disparity in species diversification if they confer adaptation to different environments. Seed mass is one such trait that is particularly important for angiosperms because it integrates across many characteristics of an individual's life history strategy [4]. Along with adult plant size, seed mass affects survival, reproductive lifespan and dispersal [5]. These life history characteristics contribute to fitness and adaptation, which are the ultimate determinants of whether lineages diversify or go extinct [6]. In support of this idea, seed mass has been shown to correlate negatively with diversification in the Polygoneaceae [7], but this has not been investigated across large taxonomic scales. As seed mass varies over ten orders of magnitude in angiosperms, from the minute 1 µg seeds of some orchids to the >18 kg seeds of the sea coconut (*Lodoicea maldivica*), this huge variation may coincide with variation in species diversity. Generalising the direction and magnitude of a link between seed mass and diversification across taxonomic scales has, however, proved difficult. Some life history characteristics encapsulated by seed mass are expected to promote speciation or extinction, while others may simultaneously counteract such effects [8].

The rate of change in key biological traits, such as seed size, can be as important in driving macroevolutionary dynamics as the absolute values of the traits themselves

[9]. This is because phenotypic divergence may cause reproductive isolation that results in speciation [10]. Nevertheless, few empirical studies have detected a correlation between rates of phenotypic evolution and lineage diversification [9,11,12]. A correlation between the two may be expected where a trait can change more rapidly in some species than others in response to selective pressures (i.e. high “evolvability” [13]). This rapid change may enable greater access to new ecological niches or quicker establishment of reproductive isolation, thereby increasing the rate of speciation (λ) [14]. In the case of seed mass, the ability to switch rapidly from small seeds with high dispersal ability to larger seeds with lower dispersal ability might promote cycles of rapid colonisation and isolation or permit adaptation to new dispersal vectors in novel environments. Rapid evolution of new phenotypes may also allow individuals to escape harsh environmental conditions and competitive interactions [15], thereby decreasing extinction rates (μ). The net outcome of these processes on net diversification ($r = \lambda - \mu$) will ultimately depend upon which of these rates responds more strongly to phenotypic change.

Here, we show that both seed mass and its phenotypic rate of evolution correlate with speciation and extinction across the angiosperm ToL. Our approach combined the most comprehensive phylogenetic timetree available [16] with an unparalleled dataset of seed mass measurements from over 30,000 angiosperm species [17]. We estimated rates of speciation, extinction and seed size evolution across the phylogeny using Bayesian, method-of-moments, and maximum likelihood analyses – each with different assumptions regarding rate variation through time. We then tested whether there was any links between rates of diversification and both seed size and its rate of evolution. Additionally, we examined whether this link was consistent across different methodologies, timescales and taxonomic categories.

Results

Our results point to a strong association between angiosperm diversification and rates of seed size evolution irrespective of analytical method or taxonomic scale, with weaker evidence for a link between macroevolutionary dynamics and absolute seed size. In the first instance, we calculated rates of speciation (λ), extinction (μ) and seed size evolution using Bayesian Analysis of Macroevolutionary Mixtures (BAMM) [18]. BAMM models rate heterogeneity through time and lineages, and accounts for incomplete taxon sampling. We used a genus-level phylogenetic tree for this particular analysis to maximise overlap between the phylogenetic and the seed mass datasets and make computation feasible. The resulting tree contained 4105 genera. As expected, given the high degree of taxonomic imbalance observed in the angiosperm phylogeny, we found strong support for more than 100 shifts in the rates of diversification (Fig. S1). There was also marked heterogeneity in the rates of seed size evolution (Fig. 1), which varied over two orders of magnitude (Fig. S2).

We then estimated whether shifts in macroevolutionary dynamics (λ, μ and r) estimated with BAMM were significantly correlated with absolute mean genus seed size and rates of seed size evolution by comparing the empirical correlations to a null distribution generated using STructured Rate Permutations on Phylogenies (STRAPP), which is robust to phylogenetic pseudoreplication (see Methods for details) [19]. We were able to link major differences in diversity across angiosperm clades with both the rate of phenotypic evolution and the absolute value of trait itself. Specifically, increased speciation was associated with a faster rate of seed size evolution (Spearman's $\rho = 0.44$, p-value < 0.0001; Fig. 2a). Increased extinction rates were similarly associated with higher evolvability ($\rho = 0.35$, p-value = 0.003), but given the weaker effect, the net outcome of λ-μ was that diversification rates were positively correlated with phenotypic change ($\rho = 0.42$, p-value < 0.0001; Fig. 2b). We also identified an association between seed size and both speciation and extinction ($\rho = -0.22$, p-value = 0.016 and $\rho = -0.20$, p-value = 0.045, respectively; Fig. 2b). Again, as the effect on speciation was greater than that on extinction, diversification rates increased with decreasing seed size ($\rho = -0.19$, p-value = 0.049; Fig. 2b). Generally, the observed correlations arose from many phenotypically fast-evolving clades distributed across the phylogeny (Fig. S2) and were robust to prior choice in the BAMM analyses (Fig. S1).

We also obtained estimates of net diversification rates with other methods to account for intra-generic variation in seed size and test the robustness of our findings. The first method used a moment estimator that assumed diversification rates were constant over time within genera that had a known crown age and number of species [20]. Monophyletic genera were then selected from a phylogenetic tree containing 8448 species present in our initial seed mass dataset so that they each contained ≥4 species with seed size data and more than 30% of their described species, thereby ensuring that the minimum probability of recovering the correct crown age node of any genus was 70% (see Methods). A rate of seed size evolution was then estimated for the 70 resulting genera by fitting a Brownian Motion (BM) model of trait evolution. Mirroring the BAMM analysis, diversification rates were strongly correlated with the rate of seed size evolution (PGLS slope = 0.157, p-value = 0.004; df = 69), and weakly negatively correlated with mean genus seed size (PGLS slope = -0.064, p-value = 0.039; df = 69). Our results were similar when we did not limit the analysis based on available seed size data (PGLS slope = -0.103, p-value = 0.004, df = 127; no analysis was performed for trait evolution with these criteria because of considerable missing data).

The second method that we used to account for intra-generic variation involved estimating macroevolutionary dynamics and phenotypic evolution across sets of clades in our species-level tree that were defined at different timescales. A series of ten, 2 million year-wide time slices from the present up to 20 million years ago was used to identify the most inclusive monophyletic clades of ≥4 species. The minimum intra-clade sampling fraction was set at 30% as above. For each resulting clade in each time interval, diversification rates and seed size evolution were estimated as in our previous non-BAMM analyses. We also fitted a series of time-dependent diversification models to each clade with RPANDA, which uses a maximum likelihood approach to estimate speciation and extinction and allows for incomplete taxon sampling [21]. By measuring the correlation of macroevolutionary rates and morphological rates across sets of similarly aged clades we effectively eliminated any spurious correlations caused by both rates being positively correlated with time. Again, we found a positive correlation between the rate of seed size evolution and speciation rates that was consistent across time slices (Fig. 3a, Fig. S3a). As expected given the weak association between seed size and speciation found in our other analyses, a significant negative correlation was only recovered in some of the analysed time slices (Fig 3b, Fig. S3b).

Discussion

Our study supports the ideas that variation in seed mass and, particularly, its rate of evolution can help explain disparity in diversification across the angiosperm phylogeny by playing a central role in plant life history. As we show with our clade-based analysis, our results are repeatable across many methodologies, varying timescales and do not ascribe to a particular taxonomic category (Figs 2, 3). We also showed that the effect of intra-generic variation in seed mass in the BAMM analysis was negligible (S1 Appendix).

The robust association of high rates of phenotypic change with lineage diversification has recently been observed in other taxonomic groups [9,12], but never across the whole of the angiosperm ToL as we find here. Accelerated morphological evolution may allow radiating lineages to occupy more complex adaptive landscapes [22]. Similarly, species with greater rate of change in their seed mass (i.e., higher evolvability) could shift between adaptive peaks or develop reproductive barriers more rapidly. Alternatively, the theory of punctuated equilibria [23], whereby morphological changes can arise from the speciation process, might also explain the connection of phenotypic evolution with species divergence. However, current methods do not allow us to distinguish whether speciation is responding to morphological change or vice versa when reconstructing 250 million years of evolutionary history [9].

Seed mass itself will also covary with dispersal ability and environmental tolerance in ways that can change speciation and extinction. For example, we found that smaller-seeded genera had faster speciation rates. This may be because smaller-seeded genera generally disperse over larger distances [24], which can promote speciation by creating isolated populations [25]. However, the relationship between dispersal and speciation is highly context dependent. The permeability of a landscape to dispersal determines the values of dispersal that may promote species divergence [26]. The weak correlation that we observed between seed size and diversification might therefore reflect contradicting patterns of different strength operating at different clades in the angiosperm phylogeny. Dispersal syndromes may also modify the effect of seed size on speciation. For instance, species with larger seeds are generally associated with biotic dispersal that distributes seeds over greater distances than wind or gravity dispersal [5]. However, broad-scale predictions for the effects of dispersal syndromes on diversification may be inaccurate, since the former depend on landscape connectivity [8] and can sometimes be inconsistent, e.g. a wind-dispersed seed might be transported by an animal. We also found that smaller-seeded genera had higher rates of extinction, possibly due to smaller nutritional reserves that constrain establishment, environmental tolerance and access to limiting resources [27]. Detailed contextual data will be necessary to expand upon the mechanisms underlying our findings in specific regions and clades.

Although seed mass is associated with other traits that can affect diversification, there is little evidence that these better explain our observed correlations or that seed size is a mere proxy for one of these other traits. For example, genome size positively correlates with seed mass [28], and faster rates of genome size evolution have been linked to increased speciation in angiosperms [29]. Shorter, smaller-seeded plants also tend to have faster life cycles, which may accelerate mutation rates [30] and promote diversification [31]. By comparing the effects of genome size and life cycle across a subset of >900 genera in our dataset, we found that only the distinction between strictly annual versus perennial genera influenced macroevolutionary dynamics in a similar way as seed size (S1 Appendix, Fig. S4). But unlike other traits [29], both absolute seed size as well as its rate of change were correlated with speciation and extinction. Thus, although other traits surely influence diversification [32], we argue that our results generally reflect the role of seed size as a trait that integrates across multiple aspects of life history characteristics in ways that can predictably influence plant macroevolutionary dynamics (Fig. S5).

The approach applied here can help to unravel the processes responsible for generating large-scale asymmetries in biodiversity. It also offers the potential to test how widely-varying traits and their rate of morphological evolution influence other aspects of the evolution and adaptation of flowering plants (e.g. [16]). Clade-specific exceptions arising from local interactions with non-focal traits [33] and specific spatio-temporal contexts will undoubtedly interact with broad-scale macroevolutionary patterns and may modulate the effects of seed mass on diversification. Regardless, our results show that seed size and its rate of evolution correlate with speciation and extinction across the flowering plants. This finding may help to explain why some clades are much more species-rich than others and points to a key role of rapid morphological evolution in generating greater levels of diversity.

Materials and methods

Seed mass and phylogenetic dataset

Seed mass data for 31,932 species were obtained from the Royal Botanic Gardens Kew Seed Information Database [17]. Species names were standardised with The Plant List (TPL) nomenclature [2] and cleaned using the *Taxonstand* R package [34].

Further processing at the genus-level was performed with the *taxonlookup* R package [35], which is a complete genus-family-order mapping for vascular plants that draws from TPL, the Angiosperm Phylogeny website [36] and a higher-level manually-curated taxonomic lookup [16]. Mean seed mass values were subsequently calculated for a total of 4763 genera.

We used the most comprehensive phylogenetic tree for land plants [16,37] that comprises 31,389 species. Taxonomic information for our phylogenetic tree was run through *Taxonstand* and *taxonlookup* as described above to make it as comparable as possible to the seed mass dataset. Monophyly of the genera in the tree was assessed using the *Monophy* package [38]. The initial estimate was 16%, but we removed taxa classified as outliers by *Monophy* (i.e., taxa that lay outside of an established “core clade” for each genus) and this resulted in only 8% of the genera not being monophyletic. The *Phyndr* [35] package was then used to generate a genus-level tree with as much overlap as possible between the phylogeny and the trait data. The final genus-level phylogenetic tree included representatives from 303 plant families (Fig. S6). We also obtained a species-level tree with 8448 representatives from the initial phylogenetic tree that were also present in the Kew Seed Information Database.

Diversification and phenotypic evolution analyses

First, speciation, extinction and net diversification rates and rates of seed size evolution were estimated separately on the genus-level phylogeny using BAMM version 2.5.0 [18]. BAMM models shifts in macroevolutionary regimes across a phylogenetic tree using reversible-jump Markov chain Monte Carlo (rjMCMC) sampling. Initial prior settings were calculated with the *setBAMMpriors* function in

BAMMtools [39], and the *expectedNumberOfShifts* parameter was set at 25 for both the speciation/extinction and trait evolution analyses, respectively. We incorporated non-random incomplete sampling information following BAMM protocols by calculating the proportion of genera sampled inside each family and estimated the backbone sampling as the overall proportion of sampled genera. *Taxonlookup* was used as a reference for these calculations. All BAMM analyses were run for 50 million generations. We verified convergence by plotting chain traces and ensuring that the effective sample sizes of all relevant parameters exceeded 200. The first 10 million generations were discarded as burn-in. The *plotPrior* function in BAMMtools was used to check the difference between model priors and posteriors. We also analysed BAMM prior sensitivity by re-running both the diversification and the trait evolution analyses with different settings for the *expectedNumberOfShifts* parameter of either 10, 25, 50 or 100 (Fig. S1).

Second, estimates of net diversification rates were obtained for a curated set of angiosperm genera using the species-level phylogeny. Only monophyletic genera with four or more species in the tree were considered for further analysis. A minimum threshold value for genus specific sampling fraction was set at 0.3. This ensured that the minimum probability of recovering the true root node of any genera was 0.7 ("crown sampling probability", following [40,41]). Crown ages for the selected genera were then calculated and used to estimate net diversification rates using the method-of-moments estimator [20]. The total number of species in each genus was obtained from *taxonlookup*. Following standard practice, we assumed three values of relative extinction fraction, $\varepsilon = 0$, 0.5 and 0.9. Different values did not affect our conclusions; therefore we present the results of the intermediate extinction fraction (ε

= 0.5). The rate of seed size evolution was estimated with by fitting a Brownian Motion (BM) model of evolution to the seed size data within each genus using *fitContinuous* from the package *geiger* [42].

Finally, we obtained clade-based measures of diversification and seed size evolution across the species-level tree. Clade-specific sampling fractions were calculated by weighting the genus-specific sampling fractions (i.e., the number of congeneric species in the 8448 species tree divided by the total number of described species for that genus) by the number of species from each genus present in the clade. These weighted sampling fractions were used to estimate the total clade species richness. A minimum threshold of 0.3 was used for the clade-specific sampling fraction, again ensuring that the crown sampling probability for each clade was at least 0.7 (the actual median clade crown sampling probabilities for the time slices ranged between 0.84 and 0.95). We estimated net diversification rates with the method-of-moments estimator as described above. We also used RPANDA to fit a series diversification models that estimated time-dependent rates to each clade. We fitted six different models of diversification: (i) pure birth model with constant λ (speciation rate); (ii) pure birth model with exponential λ; (iii) birth-death model with constant λ and μ (extinction); (iv) birth-death model with constant λ and exponential μ; (v) birth-death model with exponential λ and constant μ; and (vi) birth-death model with exponential λ and exponential μ. We then used AIC-based model selection to select the best fitting model and obtain the corresponding macroevolutionary parameters. Additionally, to account for possible biases when analysing clades with many non-congeneric species, we confirmed the results of our clade-based analysis but only considered clades that only contained congeneric species (Fig. S7).

Correlation of diversification and trait evolution

All rate variables were log-transformed for the correlation analyses. We used STRAPP to test for multiple associations between BAMM-estimated macroevolutionary dynamics and each of seed mass (using genus mean values as character states) and seed mass rate of evolution (using seed mass evolutionary rates at the genus-level tips of the phylogeny as character states). STRAPP compares the correlation between a focal trait and a macroevolutionary parameter (λ, μ or r) to a null distribution of correlations. The null correlations are generated by permuting the evolutionary rates in the tips of the phylogenetic tree while maintaining the location of rate shift events in the phylogeny. In each case, we calculated the absolute difference between the observed correlation of the macroevolutionary rate and the trait state and the null correlation obtained by the structured permutations across 5000 samples from the BAMM posterior. The reported p-value was the proportion of replicates where the null correlation coefficient was greater than the observed correlation. We found a low type I error associated with our STRAPP correlation analysis (p-value = 0.11, Fig. S8).

For both the curated genera dataset and the clade based analyses, we estimated the relationship between speciation rates with seed size rate of evolution and mean genus seed size with phylogenetic generalised least square (PGLS) as implemented in the R package *caper* [43] (Fig. S9, Fig. S10). Similar results as presented in the main text were obtained when analysing net diversification instead of speciation rates (Fig. S11). Finally, we similarly analysed the correlation between speciation and each of seed size and its rate of change, this time selecting only clades consisting of congeneric species. Again, this analysis resulted in a similar pattern as the one presented in the main text (Fig. S7).

Code availability

Scripts used to carry out the analysis described in the paper and generate the figures will be deposited in Github upon acceptance.

Acknowledgments

We thank D. A. Coomes, A. J. Helmstetter, T. Jucker and W. G. Lee for useful comments that helped improve the manuscript; and V. Soria-Carrasco for help with analyses. J.I. and A.J.T. thank the Gatsby Charitable Foundation, Wellcome Trust and Newton Trust for funding. E.F.M was funded by the BBSRC DTP at the University of Cambridge.

Author contributions

J.I and A.S.T.P conceived the study. J.I. and E.F.M. performed the analysis. J.I. and A.J.T interpreted the analysis and wrote the manuscript. All authors edited the manuscript.

References

1. Crane PR, Friis EM, Pedersen KR. The origin and early diversification of angiosperms. Nature. 1995;374: 27–33. doi:10.1038/374027a0
2. The Plant List. Version 1.1 [Internet]. 2013 [cited 20 Jun 2003]. Available: http://www.theplantlist.org/
3. Davies T, Barraclough T. The diversification of flowering plants through time and space: Key innovations, climate and chance. Reconstructing the Tree of Life: Taxonomy and Systematics of Species Rich Taxa. 2007. pp. 149–163. doi:DOI: 10.1201/9781420009538.ch10

4. Moles AT, Ackerly DD, Webb CO, Tweddle JC, Dickie JB, Westoby M. A brief history of seed size. Science (80-). 2005;307: 576–580. doi:10.1126/science.1104863
5. Moles A, Leishman M. The seedling as part of a plant's life history strategy. In: Leck MA, Parker VT, Simpson RL, editors. Seedling ecology and evolution. Cambridge: Cambridge University Press; 2008. pp. 215–235.
6. Salguero-Gómez R, Jones OR, Jongejans E, Blomberg SP, Hodgson DJ, Mbeau-Ache C, et al. Fast-slow continuum and reproductive strategies structure plant life-history variation worldwide. Proc Natl Acad Sci U S A. 2016;113: 230–5. doi:10.1073/pnas.1506215112
7. Kostikova A, Salamin N, Pearman PB. The role of climatic tolerances and seed traits in reduced extinction rates of temperate polygonaceae. Evolution (N Y). 2014;68: 1856–1870. doi:10.1111/evo.12400
8. Givnish TJ. Ecology of plant speciation. Taxon. 2010;59: 1326–1366.
9. Rabosky DL, Santini F, Eastman J, Smith SA, Sidlauskas B, Chang J, et al. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. Nat Commun. Nature Publishing Group; 2013;4: 1958. doi:10.1038/ncomms2958
10. Coyne JA, Orr HA. Speciation. Sinauer Associates; 2004.
11. Adams DC, Berns CM, Kozak KH, Wiens JJ. Are rates of species diversification correlated with rates of morphological evolution? Proc R Soc B Biol Sci. 2009;276: 2729–38. doi:10.1098/rspb.2009.0543
12. Price SL, Etienne RS, Powell S. Tightly congruent bursts of lineage and phenotypic diversification identified in a continental ant radiation. Evolution. 2016; doi:10.1111/evo.12894
13. Pigliucci M. Is evolvability evolvable? Nat Rev Genet. Nature Publishing Group; 2008;9: 75–82. doi:10.1038/nrg2278
14. Ricklefs RE, Renner SS. Species richness within families of flowering plants. Evolution (N Y). [Society for the Study of Evolution, Wiley]; 1994;48: 1619–1636. doi:10.2307/2410252
15. Rees M, Westoby M. Game-theoretical evolution of seed mass in multi-species ecological models. Oikos. 1997;78: 116–126. doi:10.2307/3545807
16. Zanne AE, Tank DC, Cornwell WK, Eastman JM, Smith SA, FitzJohn RG, et al. Three keys to the radiation of angiosperms into freezing environments. Nature. 2014;506: 89–92. doi:10.1038/nature12872
17. Kew Seed Information Database. In: Royal Botanic Gardens Kew [Internet]. 2016 [cited 1 Mar 2016]. Available: http://data.kew.org/sid/
18. Rabosky DL. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. PLoS One. Public Library of Science; 2014;9: e89543. doi:10.1371/journal.pone.0089543
19. Rabosky DL, Huang H. A robust semi-parametric test for detecting trait-dependent diversification. Syst Biol. Oxford University Press; 2016;65: 181–93. doi:10.1093/sysbio/syv066
20. Magallón S, Sanderson MJ. Absolute diversification rates in angiosperm clades. Evolution (N Y). 2001;55: 1762–1780. doi:10.1111/j.0014-3820.2001.tb00826.x
21. Morlon H, Lewitus E, Condamine FL, Manceau M, Clavel J, Drury J. RPANDA: An R package for macroevolutionary analyses on phylogenetic trees. Fitzjohn R, editor. Methods Ecol Evol. 2016;7: 589–597. doi:10.1111/2041-210X.12526

22. Lovette IJ, Bermingham E, Ricklefs RE. Clade-specific morphological diversification and adaptive radiation in Hawaiian songbirds. Proc Biol Sci. 2002;269: 37–42. doi:10.1098/rspb.2001.1789
23. Eldredge N, Gould SJ. Punctuated equilibria: an alternative to phyletic gradualism. In: Schopf T, editor. Models in Paleobiology. San Francisco: Freeman Cooper; 1972. pp. 82–115.
24. Coomes DA, Grubb PJ. Colonization, tolerance, competition and seed-size variation within functional groups. Trends Ecol Evol. 2003;18: 283–291. doi:10.1016/S0169-5347(03)00072-7
25. Kisel Y, Barraclough TG. Speciation has a spatial scale that depends on levels of gene flow. Am Nat. 2010;175: 316–34. doi:10.1086/650369
26. Claramunt S, Derryberry EP, Remsen J V, Brumfield RT. High dispersal ability inhibits speciation in a continental radiation of passerine birds. Proc Biol Sci. 2012;279: 1567–74. doi:10.1098/rspb.2011.1922
27. Leishman MR, Wright IJ, Moles AT, Westoby M. The evolutionary ecology of seed size. The Ecology of Regeneration in Plant Communities. 2000. pp. 31–58.
28. Beaulieu JM, Moles AT, Leitch IJ, Bennett MD, Dickie JB, Knight CA. Correlated evolution of genome size and seed mass. New Phytol. 2007;173: 422–37. doi:10.1111/j.1469-8137.2006.01919.x
29. Puttick MN, Clark J, Donoghue PCJ. Size is not everything: rates of genome size evolution, not C-value, correlate with speciation in angiosperms. Proc Biol Sci. 2015;282: 20152289-. doi:10.1098/rspb.2015.2289
30. Lanfear R, Ho SYW, Love D, Bromham L. Mutation rate is linked to diversification in birds. Proc Natl Acad Sci U S A. 2010;107: 20423–8. doi:10.1073/pnas.1007888107
31. Lanfear R, Ho SYW, Jonathan Davies T, Moles AT, Aarssen L, Swenson NG, et al. Taller plants have lower rates of molecular evolution. Nat Commun. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2013;4: 1879. doi:10.1038/ncomms2836
32. Beaulieu JM, O'Meara BC. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. Syst Biol. 2016; syw022-. doi:10.1093/sysbio/syw022
33. Donoghue MJ, Sanderson MJ. Confluence, synnovation, and depauperons in plant diversification. New Phytol. 2015;207: 260–274. doi:10.1111/nph.13367
34. Cayuela L, Granzow-de la Cerda Í, Albuquerque FS, Golicher DJ. taxonstand: An r package for species names standardisation in vegetation databases. Methods Ecol Evol. 2012;3: 1078–1083. doi:10.1111/j.2041-210X.2012.00232.x
35. Pennell MW, FitzJohn RG, Cornwell WK. A simple approach for maximizing the overlap of phylogenetic and comparative data. Methods Ecol Evol. 2015; n/a-n/a. doi:10.1111/2041-210X.12517
36. Stevens P. Angiosperm Phylogeny website. In: Angiosperm Phylogeny Website. Version 12 [Internet]. 2012. Available: http://www.mobot.org/MOBOT/research/APweb/
37. Qian H, Jin Y. An updated megaphylogeny of plants, a tool for generating plant phylogenies and an analysis of phylogenetic community structure. J Plant Ecol. 2015; rtv047. doi:10.1093/jpe/rtv047
38. Schwery O, O'Meara BC. MonoPhy: A simple R package to find and visualize monophyly issues. PeerJ Inc.; 2015; doi:10.7287/peerj.preprints.1600v1

39. Rabosky DL, Grundler M, Anderson C, Title P, Shi JJ, Brown JW, et al. BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. Kembel S, editor. Methods Ecol Evol. 2014;5: 701–707. doi:10.1111/2041-210X.12199
40. Sanderson MJ. How many taxa must be sampled to identify the root node of a large clade? Syst Biol. Oxford University Press; 1996;45: 168–173. doi:10.2307/2413613
41. Rabosky DL. Challenges in the estimation of extinction from molecular phylogenies: A response to Beaulieu and O'Meara. Evolution. 2016;70: 218–28. doi:10.1111/evo.12820
42. Pennell MW, Eastman JM, Slater GJ, Brown JW, Uyeda JC, Fitzjohn RG, et al. Geiger v2.0: An expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. Bioinformatics. Oxford University Press; 2014;30: 2216–2218. doi:10.1093/bioinformatics/btu181
43. Orme D. The caper package: comparative analysis of phylogenetics and evolution in R. R Packag version. 2013;5.
44. Bennett M, Leitch IJ. Plant DNA C-values database (release 6.0, Dec 2012) [Internet]. 2012. Available: http://www.kew.org/cvalues/

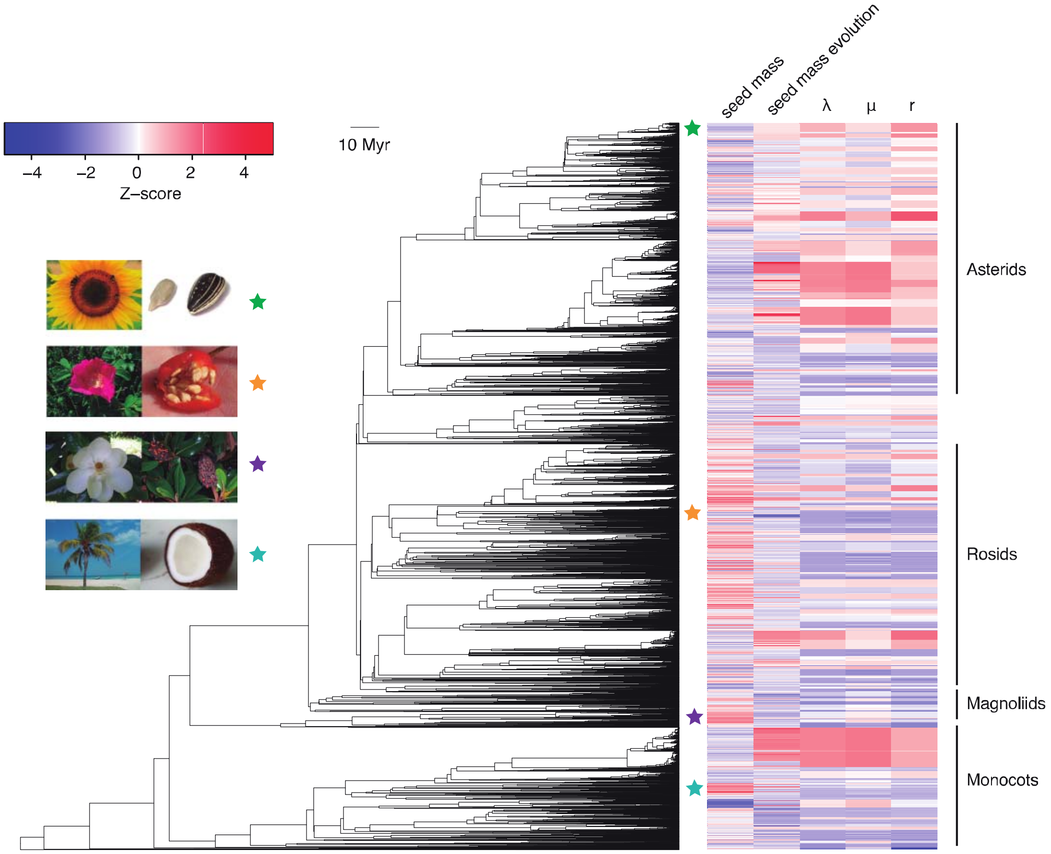

Figure 1. Seed mass and macroevolutionary rates estimated with BAMM across the angiosperm tree of life. Phylogenetic tree of 4105 genera of plants with seed mass, rate of seed mass change, and speciation (λ), extinction (μ) and net diversification (r) rates estimated by BAMM. Seed mass and rate data were standardised to Z-scores so that variation could be directly compared.

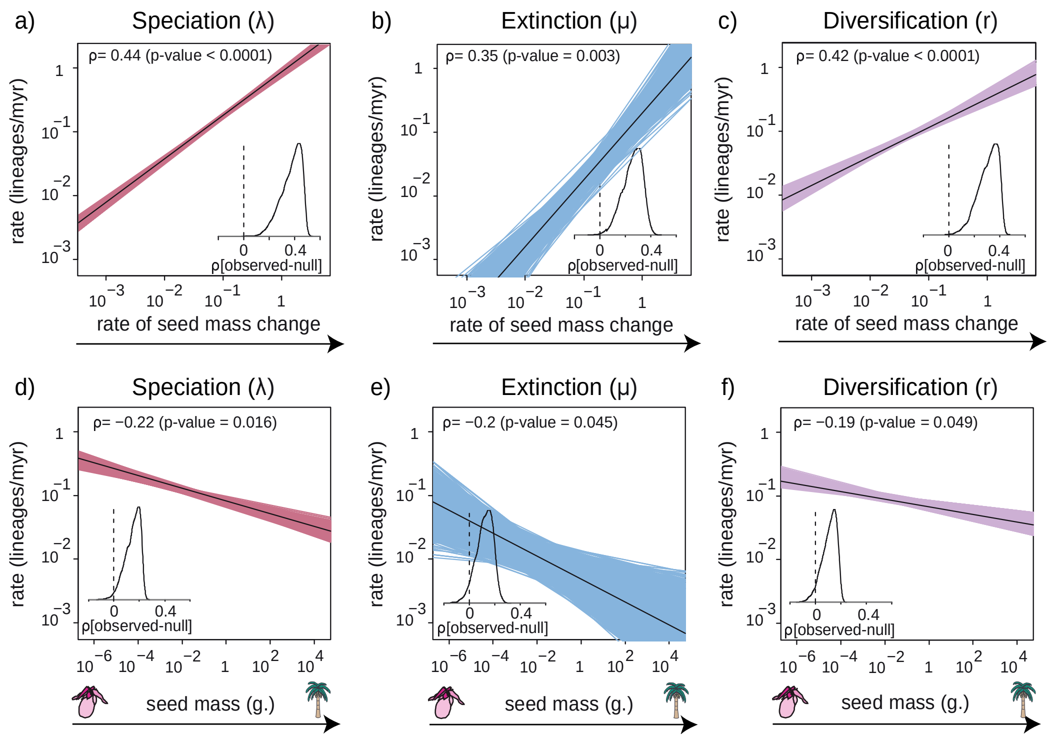

Figure 2. Seed mass and seed mass rate of evolution correlate with macroevolutionary dynamics estimated with BAMM. Spearman correlations were calculated between speciation (λ), extinction (μ), and net diversification (r) and each of a) rate of seed mass change and b) seed mass. Coloured lines are correlations for one sample of the BAMM posterior distribution, bold line is the median. The insets show the density plots of the absolute difference between the observed and null correlation calculated across 5000 structured permutations of the evolutionary rates on the phylogenetic tree (myr = million years).

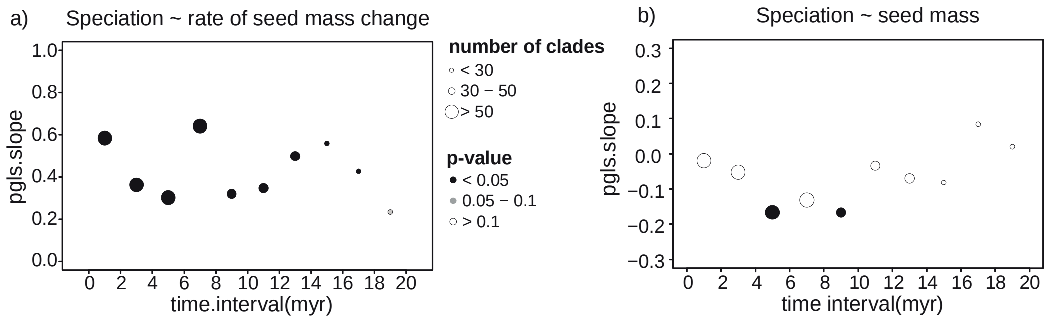

Figure 3. Seed mass and seed mass rate of evolution correlate with speciation in the clade-based analysis. Correlation of (a) rate of seed mass evolution and (b) seed mass with speciation rate (λ) estimated using RPANDA in the clade-based analysis. The strength of correlations is shown as PGLS slopes and were calculated using mean clade-level seed mass across 10 time slices in the species-level phylogenetic tree. The size of the circles represents the number of clades in each time slice while the colour indicates the significance of the slope. A detailed representation of the results in each time slice is given in figures S9 and S10.

Supporting Information captions

Fig. S1. Prior and posterior distribution of the number of rate shifts in BAMM for a) the speciation/extinction and b) phenotypic evolution analyses for *expectedNumberOfShifts* = 10, 25, 50 and 100. The analyses in the main text were carried out with *expectedNumberOfShifts* = 25 for both speciation/extinction and phenotypic evolution analyses.

Fig. S2. Phylogenetic tree of 4105 angiosperm genera with branches scaled by diversification rate as determined by BAMM. The colours of the branches indicate the rate of seed mass evolution, also estimated with BAMM.

Fig. S3. Seed mass and its rate of evolution are associated with speciation in the clade-based analyses. (a) PGLS slope of the relationship between speciation rate (λ) from the method-of-moments estimator and and the rate of seed mass evolution across 10 time slices in the species level phylogenetic tree. Circles are scaled to the number of clades in each time slice while colour indicates the significance of the slope. (b) PGLS slope of the relationship between speciation rate and mean clade seed mass. For a detailed representation of the results in each time slice, see figure S12 and S13.

Fig. S4. STRAPP correlations of diversification and phenotypic traits for 938 angiosperm genera. The distribution of the absolute difference in the observed correlation minus the null correlation is plotted for each trait. The coloured dotted lines indicate the mean of that distribution, and the black dotted line indicates 0; a distribution with mean = 0 would show no association between a focal trait and speciation. STRAPP correlation of seed mass (shown in blue), C-value (shown in red) and life cycle (shown in green) with a) speciation rate (λ), b) extinction rate (μ), and c) net diversification rate (r).

Fig. S5. Proposed effects of seed mass and other life history traits on diversification (solid lines). Dashed lines indicate correlations between life history traits. Numbers indicate reference where the link is proposed.

Fig. S6. Phylogenetic tree of 303 angiosperm families with representatives in this study. The red bars indicate the levels of sampling for each family.

Figure S7. Correlation of speciation with seed mass and seed mass rate of evolution in the clade-based analysis only considering congeneric species. (a) PGLS slope of the relationship of speciation rate - estimated with the method-of-moments estimator - with mean clade seed mass across 10 time slices in the species level phylogenetic tree. The size of the circles represents the number of clades in each time slice while the colour indicates the significance of the slope. (b) PGLS slope of the relationship of speciation rate and the rate of seed mass evolution.

Fig. S8. Type I error analysis. We estimated the type I error rate of our analysis by simulating neutral traits on the angiosperm phylogenetic tree. We performed 1000 simulations and then ran 1000 STRAPP tests with each simulated dataset. We estimated the corresponding p-values for the association between traits and diversification and calculated the type I error as the proportion of datasets where a significant association ($p\text{-value} < 0.05$) was detected.

Fig. S9. Correlations between clade rate of seed size evolution and speciation rate (estimated with RPANDA) across time slices from: (a) 0 to 2 million years (myr); (b) 2 to 4 myr; (c) 4 to 6 myr; (d) 6 to 8 myr; (e) 8 to 10 myr; (f) 10 to 12 myr; (g) 12 to 14 myr; (h) 14 to 16 myr; (i) 16 to 18 myr; and (j) 18 to 20 myr.

Fig. S10. Correlations between mean clade seed mass and speciation rate (estimated with RPANDA) across time slices from: (a) 0 to 2 million years (myr); (b) 2 to 4 myr; (c) 4 to 6 myr; (d) 6 to 8 myr; (e) 8 to 10 myr; (f) 10 to 12 myr; (g) 12 to 14 myr; (h) 14 to 16 myr; (i) 16 to 18 myr; and (j) 18 to 20 myr.

Figure S11. Correlation of (a) rate of seed mass evolution and (b) seed mass with net diversification rate (r) estimated using RPANDA in the clade-based analysis. The strength of correlations is shown as PGLS slopes and were calculated using mean clade-level seed mass across 10 time slices in the species-level phylogenetic tree. The size of the circles represents the number of clades in each time slice while the colour indicates the significance of the slope.

Fig. S12. Correlations between clade rate of seed size evolution and speciation rate (estimated with the method-of-moments estimator) across time slices from: (a) 0 to 2 million years (myr); (b) 2 to 4 myr; (c) 4 to 6 myr; (d) 6 to 8 myr; (e) 8 to 10 myr; (f) 10 to 12 myr; (g) 12 to 14 myr; (h) 14 to 16 myr; (i) 16 to 18 myr; and (j) 18 to 20 myr.

Fig. S13. Correlations between mean clade seed mass and speciation rate (estimated with the method-of-moments estimator) across time slices from: (a) 0 to 2 million years (myr); (b) 2 to 4 myr; (c) 4 to 6 myr; (d) 6 to 8 myr; (e) 8 to 10 myr; (f) 10 to 12 myr; (g) 12 to 14 myr; (h) 14 to 16 myr; (i) 16 to 18 myr; and (j) 18 to 20 myr.

Fig. S14. Mean genus seed mass of strict annual (n = 106) and perennial (n = 832) genera. No significant difference between the means of the two groups was found when phylogeny was accounted for (PGLS, $F_{1,937} < 0.01$, p-value = 0.972).

Fig. S15. Mean genus seed mass and coefficient of variation for 2613 angiosperm genera. The two variables were not significantly correlated (PGLS test, $F_{1,131} = 0.67$, p-value = 0.416).

Fig. S16. Correlation of coefficient of variation in seed mass with macroevolutionary dynamics. STRAPP correlations with speciation (λ), extinction (μ) and net diversification (r) and the genus-level coefficient of variation in seed mass for 2613 angiosperm genera (i.e. each with seed mass data for >1 species in the Kew SID). Correlations and insets calculated as in Fig. 2.

Table S1. STRAPP correlations for 938 genera of angiosperms with seed mass, genome size (i.e., C-value) and life cycle data. Macroevolutionary parameters were obtained from an analysis with 4105 genera (see Main text).