bioRxiv preprint doi: https://doi.org/10.1101/053116; this version posted August 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Classification: Biological Sciences, Evolution

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

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] 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

Keywords:

Seed size, diversification, angiosperms, speciation

bioRxiv preprint doi: https://doi.org/10.1101/053116; this version posted August 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

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 explain why some lineages diversify more than others because it integrates across many key life history traits that influence speciation and extinction, such as dispersal, survival, environmental tolerance and reproductive success. However, the extent and direction of the association between seed mass and diversification has not been assessed across the angiosperm phylogeny. Here, we show for the first time that absolute seed size and the rate of change in seed size are both associated with variation in diversification rates. Based on an unequalled phylogenetic tree that included 4105 angiosperm genera, 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, supporting emerging evidence that rapid morphological change is associated with species divergence. Our study now reveals that variation in morphological traits, as well as the rate at which traits evolve, can contribute to explaining the extremely uneven distribution of diversity across the ToL.

Significance statement

Why species richness varies greatly among different types of organisms is a long-standing interest of biologists. In flowering plants, seed size may play a role in generating different levels of species richness because it integrates across many life history traits that influence speciation and extinction, including environmental

bioRxiv preprint doi: https://doi.org/10.1101/053116; this version posted August 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

tolerance, dispersal ability and survival. Here, we show how seed size and the rate at which it evolves both correlate with rates of species diversification across flowering plants. Faster diversification rates were associated with smaller-seeded species and more rapid changes in seed size, adding to growing evidence linking morphological and species divergence. These results show how phenotypic traits and their rates of evolution can help us understand the mechanisms that generate biodiversity.

Text

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, however, extremely unevenly distributed 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).

Seed mass integrates across many individual life history traits in ways that can influence angiosperm diversification (4). Along with adult plant size, seed mass affects survival, reproductive lifespan and dispersal (5). These traits 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

bioRxiv preprint doi: https://doi.org/10.1101/053116; this version posted August 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

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 traits encapsulated by seed mass are expected to promote speciation or extinction, while others may simultaneously counteract such effects.

The rate of change in key life history traits, such as seed size, can be as important in driving macroevolutionary dynamics as the absolute values of the traits themselves (8). This is because phenotypic divergence may cause reproductive isolation that results in speciation (9). Nevertheless, few empirical studies have detected a correlation between rates of phenotypic evolution and lineage diversification (8, 10). 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" (11)). This may enable greater access to new ecological niches or quicker establishment of reproductive isolation, thereby increasing the rate of speciation (λ) (12). 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 (13), thereby decreasing extinction rates (μ). The net outcome of these processes on diversification ($r = \lambda - \mu$) will ultimately depend upon which of these rates responds more strongly to phenotypic change.

bioRxiv preprint doi: https://doi.org/10.1101/053116; this version posted August 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

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 (14) with an unparalleled dataset of seed mass measurements (15) to obtain mean seed mass values for 24% of all described angiosperm genera (n = 4105). We estimated rates of speciation and extinction using Bayesian Analysis of Macroevolutionary Mixtures (BAMM) (16), which models rate heterogeneity through time and lineages, and accounts for incomplete taxon sampling. Similarly, we analysed the rates of change in seed size across the phylogeny. We then estimated the correlations of the macroevolutionary dynamics with absolute seed size and the rate of seed size evolution.

Results and discussion

As expected, given the high degree of taxonomic imbalance observed in the angiosperm phylogeny, we found strong support for multiple shifts in the rates of diversification. The median number of rate shifts for the speciation/extinction analysis was 123 (95% confidence interval (CI): 108-143) and 187 (95% CI: 148-226) for the seed size evolution analysis. Both posterior distributions for the number of rate shifts were clearly identifiable from their respective priors and were not overly sensitive to the prior settings (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) were significantly correlated with absolute mean genus seed size and rates of seed size evolution by comparing the empirical correlations to a null distribution of correlations generated using STructured Rate Permutations on Phylogenies (STRAPP), which is

bioRxiv preprint doi: https://doi.org/10.1101/053116; this version posted August 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

robust to phylogenetic pseudoreplication (see Methods for details) (17). We focused our analysis on genera to maximise the overlap between the phylogenetic and the seed mass datasets and to make the Bayesian estimation of macroevolutionary dynamics computationally feasible.

We were able to link major differences in diversity across angiosperm clades with a single phenotypic trait that should influence speciation and extinction. Specifically, increased speciation was associated with smaller seed sizes (Spearman's ρ = -0.22, p-value = 0.016; Fig. 2a). Increased extinction rates were similarly associated with smaller seeds (ρ = -0.20, p-value = 0.045), but given the relatively weak effect, the net outcome of λ-μ was that diversification rates increased with decreasing seed size (ρ =-0.19, p-value = 0.049). We also identified a stronger positive association between the rate of seed size evolution and both speciation and extinction (ρ = 0.44, p-value < 0.0001 and ρ = 0.35, p-value = 0.003, respectively; Fig. 2b). Again, as the effect of speciation was greater than that of extinction, rates of diversification and phenotypic change were positively correlated (ρ = 0.42, p-value < 0.0001; Fig. 2b). Generally, the observed correlations arose from many phenotypically fast-evolving clades distributed across the phylogeny (Fig. S2).

Our results were unaffected by intra-generic variation in seed mass for two reasons. First, there was no systematic bias in intra-generic variation across the phylogenetic tree. We detected no correlation between the mean and the coefficient of variation (CV) for seed mass of each genus (Fig. S3, PGLS: $F_{1,131}$ = 0.67, p-value = 0.416). Second, we could disregard intra-generic variation in rates of seed mass evolution as influencing our results as we found that genera with larger variation in seed size did

bioRxiv preprint doi: https://doi.org/10.1101/053116; this version posted August 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

not have different macroevolutionary dynamics (Fig. S4, Spearman's ρ_λ = -0.05, p-value = 0.114; ρ_μ = -0.03, p-value = 0.406; ρ_r = -0.067, p-value = 0.058).

Our study supports the idea that variation in seed mass and its rate of evolution can help explain disparity in diversification across the angiosperm phylogeny by playing a central role in plant life history (Fig. S5). Dispersal ability and environmental tolerance are two such aspects of life history that covary with seed mass 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 (18), which can promote speciation by creating isolated populations (19), especially in fragmented landscapes. 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 (20) 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 (21). 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

bioRxiv preprint doi: https://doi.org/10.1101/053116; this version posted August 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

a mere proxy for one of these other traits. For example, genome size positively correlates with seed mass (22), and faster rates of genome size evolution have been linked to increased speciation in angiosperms (23). Shorter, smaller-seeded plants also tend to have faster life cycles, which may accelerate mutation rates (24) and promote diversification (25). 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 (Supplementary Information, Fig. S6). But unlike other traits (23), both absolute seed size as well as its rate of change were correlated with speciation and extinction. Thus, although other traits surely influence diversification (26), we argue that our results generally reflect the role of seed size as a trait that integrates across multiple aspects of life history traits in ways that can predictably influence plant macroevolutionary dynamics (Fig. S5).

Our finding that high rates of phenotypic change correlate with diversification (Fig. 2b) has recently been observed in other taxonomic groups (8, 10), but never across the whole of the angiosperm ToL. Accelerated morphological evolution may allow radiating lineages to occupy more complex adaptive landscapes (27). 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 (28), 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 (8).

bioRxiv preprint doi: https://doi.org/10.1101/053116; this version posted August 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

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 influence other aspects of the evolution and adaptation of flowering plants (e.g. (14)). Clade-specific exceptions arising from local interactions with non-focal traits (29) 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 clearly show that seed size and its rate of evolution correlate with speciation and extinction across the flowering plants, which may help to explain why some clades are much more species-rich than others.

Methods

Seed mass and phylogenetic dataset

Seed mass data for 31,932 species were obtained from the Royal Botanic Gardens Kew Seed Information Database (15). Species names were standardised with The Plant List (TPL) nomenclature (2) and cleaned using the *Taxonstand* R package (30). Further processing at the genus-level was carried out with the *taxonlookup* R package (31), which is a complete genus-family-order mapping for vascular plants that draws from TPL, the Angiosperm Phylogeny website (32) and a higher-level manually curated taxonomic lookup (14). Seed mass mean values for each genus were calculated for a total of 4763 genera.

We used the most comprehensive phylogenetic tree for land plants (14, 33) that

bioRxiv preprint doi: https://doi.org/10.1101/053116; this version posted August 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

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 (34). 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* (31) 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 phylogenetic tree included representatives from 303 plant families (Fig. S7).

Diversification and phenotypic evolution analysis

Speciation, extinction and net diversification rates and rates of seed size evolution were estimated separately on the phylogeny using BAMM version 2.5.0 (16). 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* (35), 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 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

bioRxiv preprint doi: https://doi.org/10.1101/053116; this version posted August 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

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).

Correlation of diversification and trait evolution

We used STRAPP to test for multiple associations between 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 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).

Code availability

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

bioRxiv preprint doi: https://doi.org/10.1101/053116; this version posted August 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

References

1. Crane PR, Friis EM, Pedersen KR (1995) The origin and early diversification of angiosperms. *Nature* 374(6517):27–33.
2. The Plant List. Version 1.1 (2013) Available at: http://www.theplantlist.org/ [Accessed June 20, 2003].
3. Davies T, Barraclough T (2007) 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*, pp 149–163.
4. Moles AT, et al. (2005) A brief history of seed size. *Science (80-)* 307(January):576–580.
5. Moles A, Leishman M (2008) The seedling as part of a plant's life history strategy. *Seedling Ecology and Evolution*, eds Leck MA, Parker VT, Simpson RL (Cambridge University Press, Cambridge), pp 215–235.
6. Salguero-Gómez R, et al. (2016) Fast-slow continuum and reproductive strategies structure plant life-history variation worldwide. *Proc Natl Acad Sci U S A* 113(1):230–5.
7. Kostikova A, Salamin N, Pearman PB (2014) The role of climatic tolerances and seed traits in reduced extinction rates of temperate polygonaceae. *Evolution (N Y)* 68(7):1856–1870.
8. Rabosky DL, et al. (2013) Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nat Commun* 4:1958.
9. Coyne JA, Orr HA (2004) *Speciation* (Sinauer Associates).
10. Price SL, Etienne RS, Powell S (2016) Tightly congruent bursts of lineage and phenotypic diversification identified in a continental ant radiation. *Evolution*. doi:10.1111/evo.12894.
11. Pigliucci M (2008) Is evolvability evolvable? *Nat Rev Genet* 9(1):75–82.
12. Ricklefs RE, Renner SS (1994) Species richness within families of flowering plants. *Evolution (N Y)* 48(5):1619–1636.
13. Rees M, Westoby M (1997) Game-theoretical evolution of seed mass in multi-species ecological models. *Oikos* 78(1):116–126.
14. Zanne AE, et al. (2014) Three keys to the radiation of angiosperms into freezing environments. *Nature* 506(7486):89–92.
15. Kew Seed Information Database (2016) *R Bot Gard Kew*. Available at: http://data.kew.org/sid/ [Accessed March 1, 2016].
16. Rabosky DL (2014) Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS One* 9(2):e89543.
17. Rabosky DL, Huang H (2016) A robust semi-parametric test for detecting trait-dependent diversification. *Syst Biol* 65(2):181–93.
18. Coomes DA, Grubb PJ (2003) Colonization, tolerance, competition and seed-size variation within functional groups. *Trends Ecol Evol* 18(6):283–291.
19. Kisel Y, Barraclough TG (2010) Speciation has a spatial scale that depends on levels of gene flow. *Am Nat* 175(3):316–34.
20. Givnish TJ (2010) Ecology of plant speciation. *Taxon* 59(5):1326–1366.
21. Leishman MR, Wright IJ, Moles AT, Westoby M (2000) The evolutionary ecology of seed size. *The Ecology of Regeneration in Plant Communities*, pp 31–58.
22. Beaulieu JM, et al. (2007) Correlated evolution of genome size and seed mass.

bioRxiv preprint doi: https://doi.org/10.1101/053116; this version posted August 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

New Phytol 173(2):422–37.

23. Puttick MN, Clark J, Donoghue PCJ (2015) Size is not everything: rates of genome size evolution, not C-value, correlate with speciation in angiosperms. *Proc Biol Sci* 282(1820):20152289–.
24. Lanfear R, Ho SYW, Love D, Bromham L (2010) Mutation rate is linked to diversification in birds. *Proc Natl Acad Sci U S A* 107(47):20423–8.
25. Lanfear R, et al. (2013) Taller plants have lower rates of molecular evolution. *Nat Commun* 4:1879.
26. Beaulieu JM, O'Meara BC (2016) Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. *Syst Biol*:syw022–.
27. Lovette IJ, Bermingham E, Ricklefs RE (2002) Clade-specific morphological diversification and adaptive radiation in Hawaiian songbirds. *Proc Biol Sci* 269(1486):37–42.
28. Eldredge N, Gould SJ (1972) Punctuated equilibria: an alternative to phyletic gradualism. *Models in Paleobiology*, ed Schopf T (Freeman Cooper, San Francisco), pp 82–115.
29. Donoghue MJ, Sanderson MJ (2015) Confluence, synnovation, and depauperons in plant diversification. *New Phytol* 207(2):260–274.
30. Cayuela L, Granzow-de la Cerda Í, Albuquerque FS, Golicher DJ (2012) taxonstand: An r package for species names standardisation in vegetation databases. *Methods Ecol Evol* 3(6):1078–1083.
31. Pennell MW, FitzJohn RG, Cornwell WK (2015) A simple approach for maximizing the overlap of phylogenetic and comparative data. *Methods Ecol Evol*:n/a–n/a.
32. Stevens P. (2012) Angiosperm Phylogeny website. *Angiosperm Phylogeny Website Version 12*. Available at: http://www.mobot.org/MOBOT/research/APweb/.
33. Qian H, Jin Y (2015) An updated megaphylogeny of plants, a tool for generating plant phylogenies and an analysis of phylogenetic community structure. *J Plant Ecol*:rtv047.
34. Schwery O, O'Meara BC (2015) MonoPhy: A simple R package to find and visualize monophyly issues. doi:10.7287/peerj.preprints.1600v1.
35. Rabosky DL, et al. (2014) BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods Ecol Evol* 5(7):701–707.
36. Bennett M, Leitch IJ (2012) Plant DNA C-values database (release 6.0, Dec 2012). Available at: http://www.kew.org/cvalues/.

Acknowledgments

We thank D. A. Coomes, A. J. Helmstetter, T. Jucker and W. G. Lee for useful comments that helped improve the manuscript. J.I. and A.J.T. thank the Gatsby Charitable Foundation, Wellcome Trust and Newton Trust for funding. E.F.M was

bioRxiv preprint doi: https://doi.org/10.1101/053116; this version posted August 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

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.

Figure Legends

Figure 1. Seed mass and macroevolutionary rates 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. Correlation of seed mass and seed mass rate with macroevolutionary dynamics. STRAPP correlations between speciation (λ), extinction (μ), and net diversification (r) and each of a) seed mass and b) rate of seed mass change. Coloured lines are Spearman correlations between λ (shown in red), μ (blue), r (purple) and each trait 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).

bioRxiv preprint doi: https://doi.org/10.1101/053116; this version posted August 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Supporting Information

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. 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. S4. 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.

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.

bioRxiv preprint doi: https://doi.org/10.1101/053116; this version posted August 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Fig. S6. 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. S7. Phylogenetic tree of 303 angiosperm families with representatives in this study. The red bars indicate the levels of sampling for each family.

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-value < 0.05) was detected.

Fig. S9. 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).

bioRxiv preprint doi: https://doi.org/10.1101/053116; this version posted August 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Supplementary Information

Seed mass is central to a network of inter-correlated life history traits that can impact diversification. Two such traits are genome size or plant C-value (measured as picograms of DNA per haploid nucleus) and life cycle. We compared the correlation between macroevolutionary parameters (λ, μ, r) and each of seed mass, C-value and life cycle across a dataset of 938 plant genera (106 strictly annuals and 832 strictly perennials). Genome content data and life cycle data were downloaded from the Plant DNA C-values database (36). Surprisingly, mean seed mass did not differ between strictly annual and perennial plants when accounting for phylogenetic relationships (Fig. S9, PGLS: $F_{1,937} < 0.01$, p-value = 0.972).

STRAPP requires large phylogenetic trees (n > 1000 species) to detect correlations (17). In this reduced dataset, we ran STRAPP correlations for each focal trait with the diversification parameters calculated from our 4105 genera BAMM analysis. We then calculated the absolute differences in the observed and the null correlations between the macroevolutionary parameters and seed mass, C-value and "annuality" (a binary variable specifying whether the genus was annual or not). These differences were similar and positive for seed mass and life cycle, while smaller for C-value (Fig. S6). However, with this reduced dataset, only life-cycle showed significant correlations with macroevolutionary dynamics (Table S1). These correlations were weaker than those observed with seed mass in the main text (Fig. 2), and considerably weaker than correlations with the rate of seed size evolution.

Table S1. STRAPP correlations for 938 genera of angiosperms with seed mass,

bioRxiv preprint doi: https://doi.org/10.1101/053116; this version posted August 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

genome size (i.e., C-value) and life cycle data. Macroevolutionary parameters were obtained from an analysis with 4105 genera (see Main text).

parameter	Seed mass *rho*	Seed mass p-value	Life Cycle *rho*	Life Cycle p-value	C-value *rho*	C-value p-value
λ	-0.166	0.070	0.171	0.017	0.042	0.828
μ	-0.162	0.099	0.147	0.055	0.109	0.540
r	-0.141	0.142	0.168	0.012	-0.090	0.885

bioRxiv preprint doi: https://doi.org/10.1101/053116; this version posted August 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

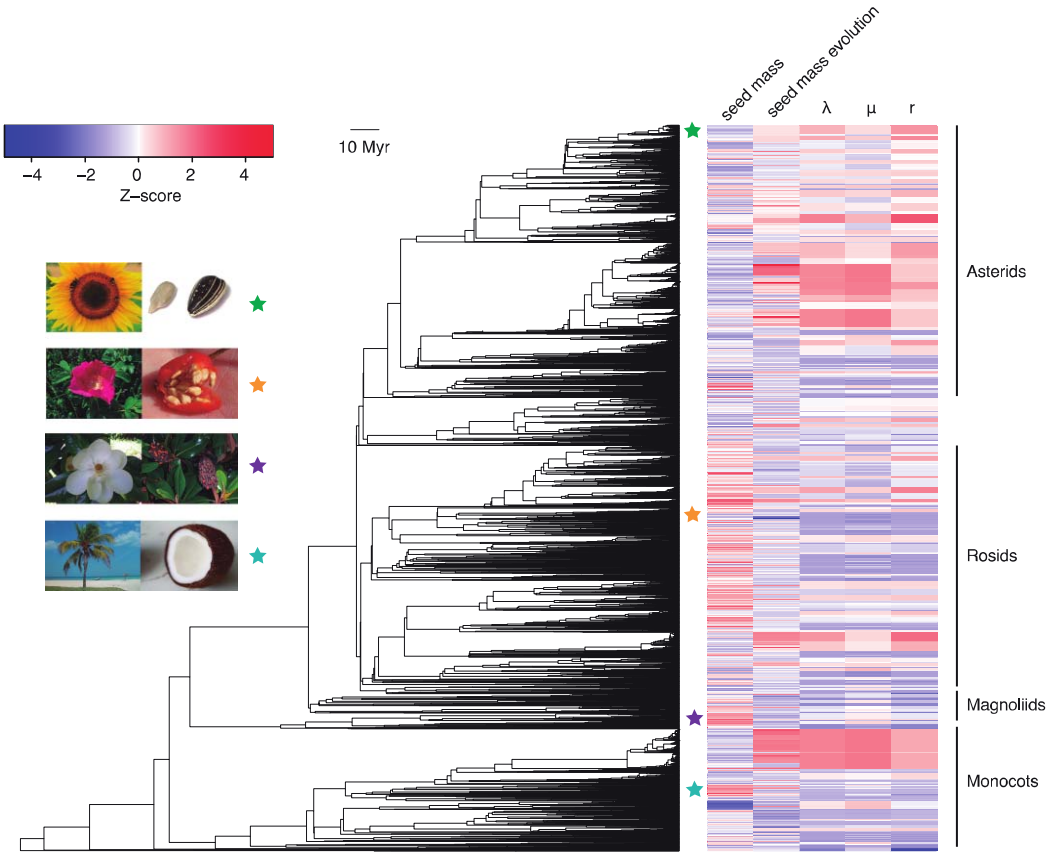

Figure 1

bioRxiv preprint doi: https://doi.org/10.1101/053116; this version posted August 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

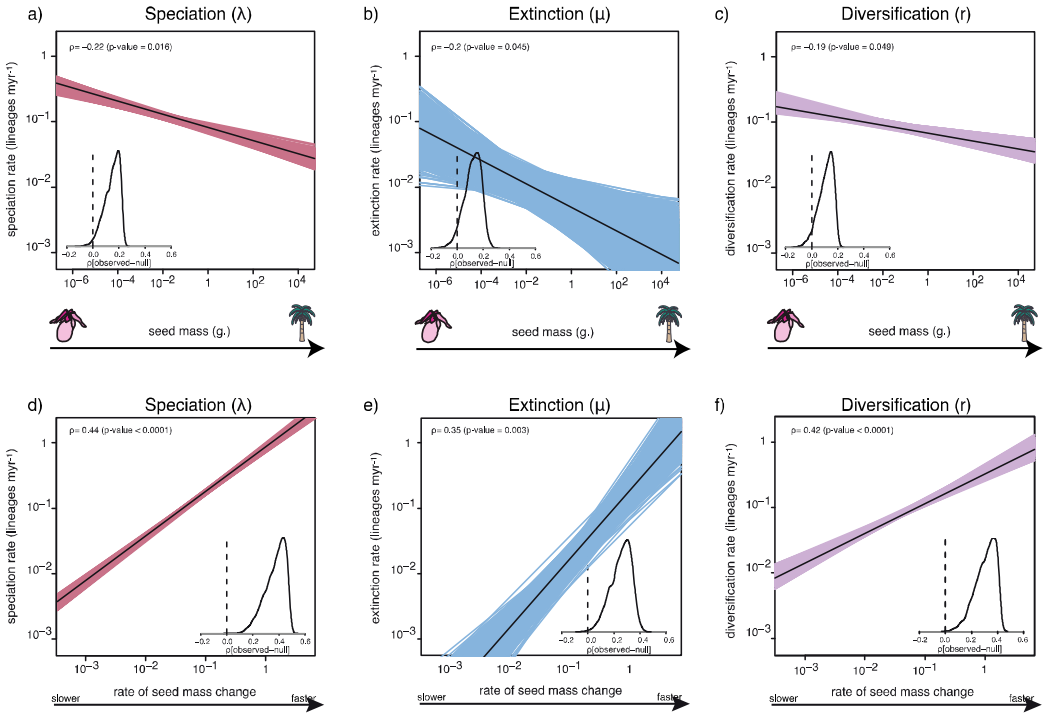

Figure 2