

# Factors that shape seed mass evolution

Angela T. Moles<sup>\*†‡</sup>, David D. Ackerly<sup>§</sup>, Campbell O. Webb<sup>¶</sup>, John C. Tweddle<sup>||\*\*</sup>, John B. Dickie<sup>\*\*</sup>, Andy J. Pitman<sup>††</sup>, and Mark Westoby<sup>†</sup>

<sup>\*</sup>National Center for Ecological Analysis and Synthesis, 735 State Street, Santa Barbara, CA 93101-5304; Departments of <sup>†</sup>Biological Sciences and <sup>††</sup>Physical Geography, Macquarie University, NSW 2109, Australia; <sup>§</sup>Department of Biological Sciences, Stanford University, Stanford, CA 94305-5020; <sup>¶</sup>Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520-8106; <sup>||</sup>The Natural History Museum, Cromwell Road, London SW7 5BD, United Kingdom; and <sup>\*\*</sup>Royal Botanic Gardens, Kew, Wakehurst Place, Ardingly, West Sussex RH17 6TN, United Kingdom

Edited by Christopher B. Field, Carnegie Institution of Washington, Stanford, CA, and approved June 8, 2005 (received for review February 22, 2005)

**We used correlated divergence analysis to determine which factors have been most closely associated with changes in seed mass during seed plant evolution. We found that divergences in seed mass have been more consistently associated with divergences in growth form than with divergences in any other variable. This finding is consistent with the strong relationship between seed mass and growth form across present-day species and with the available data from the paleobotanical literature. Divergences in seed mass have also been associated with divergences in latitude, net primary productivity, temperature, precipitation, and leaf area index. However, these environmental variables had much less explanatory power than did plant traits such as seed dispersal syndrome and plant growth form.**

correlated divergence | plant traits | growth form | phylogeny | seed dispersal

As seed plants diversified, they colonized a wide range of habitats and developed a range of growth forms and seed dispersal strategies (1–4). These changes were accompanied by major changes in seed size, a trait that is central to many aspects of plant ecology (5). The fossil record shows a particularly rapid period of change in seed size from 85 million years ago (Ma) until shortly after the Cretaceous–Tertiary boundary (65 Ma). During this period, angiosperms radiated out of the tropics, and they shifted from being predominantly small-seeded to having a much wider range of seed size strategies, with a larger mean size (2, 6). Present-day species have seed masses spanning more than 11 orders of magnitude, from the dust-like seeds of orchids up to the 20-kg seeds of the double coconut (5, 7).

There are two main schools of thought on the factors most likely to have driven these changes in seed size. Tiffney (2, 8) suggested that the radiation of mammals increased the availability of dispersal agents for large seeds and, therefore, allowed plants to radiate into a wider range of seed masses than had been possible previously. This hypothesis is consistent with correlations between large-seededness and animal dispersal in present-day species (9). Eriksson *et al.* (6) argue that it is more likely that a closure of canopies resulting from changes in climate around the Cretaceous–Tertiary favored species with larger seeds. The strong relationship between light environment and seed mass across present-day species (10–12) and the superior survival of large-seeded species when grown under deep shade provide some support for this hypothesis. Eriksson *et al.* (6) also note that changes in growth form around this time might have contributed to the changes in seed mass, because larger seeds are generally associated with larger plants (13).

Here we map a large seed mass database, together with information on other plant traits and environmental data on to the seed plant phylogeny. We then used correlated divergence analysis (14) to determine which factors have been most closely associated with evolutionary divergences of seed mass. We believe this is the beginning of a new and exciting fusion of evolution with ecology, from which we can develop unified accounts of both the present-day spread and the historical radiation of ecological strategies on a world-wide scale.

## Methods

**Data.** Seed mass data were compiled for 12,987 seed plant species (318 gymnosperms and 12,669 angiosperms) (5). Approximately half of the seed mass data are from the Royal Botanic Gardens Kew's Seed Information Database ([www.rbgekew.org.uk/data/sid/](http://www.rbgekew.org.uk/data/sid/)). The remainder were compiled by A.T.M. In both cases, data were collected opportunistically from the published literature and through personal communications (see acknowledgments for details of contributors). Seed masses reported in the literature were assumed to be dry masses unless otherwise stated. Fresh masses were converted to approximate dry masses by using the following formula: dry mass =  $(0.92 \times \log_{10} \text{fresh mass})^{0.94}$ . This relationship had an  $R^2$  of 0.97 across the 418 species for which we had both fresh and dry weights. It was not possible to distinguish reliably between seeds and diaspores, because of the great variety in what different authors call a "seed." Therefore, no attempt was made to convert "diaspore" masses into "seed" masses. However, we recorded seed mass rather than diaspore mass wherever there was a choice. Obsolete genus names were replaced with valid synonyms from the Vascular Plant Families and Genera database ([www.rbgekew.org.uk/data/vascplnt.html](http://www.rbgekew.org.uk/data/vascplnt.html)). Species names were then checked against the International Plant Names Index ([www.ipni.org/index.html](http://www.ipni.org/index.html)). Subspecies and varieties were not recognized in analyses.

Growth form and dispersal syndrome data were collected opportunistically from the published literature, from the Kew Gardens Seed Information Database, and from the U.S. Department of Agriculture PLANTS Database (<http://plants.usda.gov/>).

Latitude data represent the locations at which species were sampled rather than the midpoints of species' ranges. These data were taken from site descriptions in source papers where possible. Where necessary, latitudes from nearby locations were used in place of exact readings for field sites. These latitude data were entered into BIOME4 (a coupled biogeography and biogeochemistry model; ref. 15) to obtain estimates of net primary productivity (NPP), leaf area index (LAI), mean annual temperature, and mean annual precipitation. The temperature and precipitation data in this model are taken from ground-based measurements. BIOME4 calculates LAI and NPP across the range of plant functional types present in each half-degree grid square, by using climate and soil information linked to an ecophysiological based photosynthesis and stomatal behavior model (15). Species with records from multiple sites were assigned the geometric mean value of climate variables from the sites at which they were present.

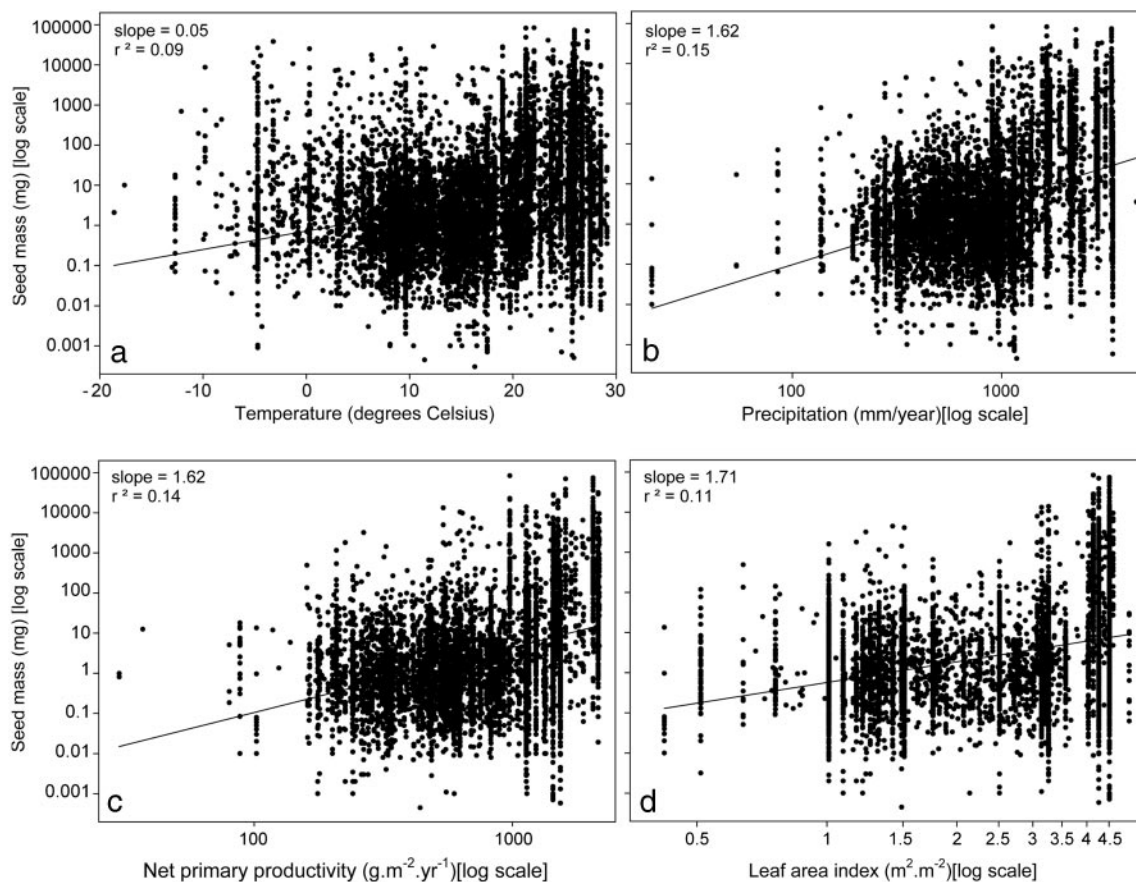
**Statistics.** Species were arrayed on a phylogenetic tree by using PHYLOMATIC (16). This program takes a list of taxa and matches

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: Ma, million years ago; NPP, net primary productivity; LAI, leaf area index.

<sup>†</sup>To whom correspondence should be addressed. E-mail: [amoles@bio.mq.edu.au](mailto:amoles@bio.mq.edu.au).

© 2005 by The National Academy of Sciences of the USA



**Fig. 1.** Cross-species relationships between seed mass and mean annual temperature (a), mean annual precipitation (b), NPP (c), and LAI (d). All regressions were significant ( $P < 0.001$ ).

them by genus or family name to a megatree constructed from published phylogenies (PHYLOMATIC tree version: R20040402). If any genus is missing from the megatree, the program returns a polytomy of genera within that family. Because of the extensive nature of our dataset, and the sparsity of genus-level resolution in PHYLOMATIC's megatree, most genera were arrayed as polytomies within families. We added some within-family resolution for particularly large families (5), but most polytomies remained unresolved. Species were always arranged as polytomies within genera.

Age estimates for major nodes in the tree were taken from Wikström *et al.* (17). These authors applied a nonparametric rate-smoothing algorithm (which allows for different clades evolving at different rates) to DNA sequence data. These estimates were calibrated against the fossil record at a single point. Although this technique is not perfect, it does afford the best estimates of divergence time available at present. The tree used by Wikström *et al.* (17) coincided with our tree at 150 of 2,229 nodes. Ages were estimated for other nodes in our tree by using the BLADJ algorithm in PHYLOCOM, a program developed by C.O.W., D.D.A., and S. Kembel (program and documentation available upon request). This program distributes undated nodes evenly between nodes of known ages. Our age estimates should be treated as very rough approximations only.

Analyses were performed by using the "Analysis of Traits (aot)" module of PHYLOCOM (written by D.D.A.). PHYLOCOM-aot calculates internal node averages for continuous traits by using the method described by Felsenstein (14). Briefly, the value of a trait at node  $X_3$  which is ancestral to nodes  $X_1$  and  $X_2$ , is calculated as follows:  $X_3 = (X_1/b'1 + X_2/b'2)/(1/b'1 +$

$1/b'2)$ , where  $b'1$  and  $b'2$  are transformed branch lengths between  $X_3$  and  $X_1$  and  $X_2$ , respectively (see ref. 5 for details).

PHYLOCOM-aot calculates contrast sizes in different ways, according to the nature of the trait being analyzed and the structure of the phylogenetic tree.

- (i) At dichotomous nodes, the contrast size for a continuous trait is the difference in trait values, divided by the square root of the total branch length between the two descendent nodes (18). The direction of subtraction must be maintained for all traits at each node to allow for negative correlations in trait evolution.
- (ii) For binary traits at dichotomous nodes, the contrast size is simply one. Contrasts on discrete traits can be calculated only at a limited set of nodes, where contrasting states of the binary trait occur on at least two of the descendent nodes and where the paths connecting taxa that form a contrast do not cross.
- (iii) For polytomies, the program uses a method introduced by Pagel (19), in which daughter nodes are ranked by the trait value of the independent variable. If the independent variable is continuous, the daughter nodes are then split into equal-sized high and low groups. Taxa are ranked by the trait value of the independent variable. If there are an odd number of daughter nodes, the median value is assigned to the lower group if its value is lower than the mean across all daughter nodes, or to the upper group if its value is higher than the mean. If the independent variable is binary, then the daughter nodes are split into two groups corresponding to the two states. For both dependent and independent

variables, a harmonic mean branch length and a mean trait value is calculated for each group, and the contrast size is calculated as for a dichotomous node (*i*).

Seed mass, latitude, NPP, LAI, precipitation, and temperature were analyzed as continuous variables. Seed mass, NPP, LAI, and precipitation were log-transformed before analysis. Growth form was analyzed as an ordered multistate categorical variable (0, herb; 1, shrub; 2, tree). PHYLOCOM-aot presently treats ordered multistate categorical variables in the same manner as continuous variables. Dispersal syndrome was coded as binary (0, abiotic dispersal; 1, biotic dispersal).

Correlated divergence analyses are performed by using linear regression of the contrast in trait 1 against the contrast in trait 2. These regressions are forced through the origin, because a given contrast could have been calculated by subtracting the trait value for species A from the trait value of species B, or vice versa. (Here, seed mass contrasts were arbitrarily made such that the contrast value was positive.) That is, each data point effectively exists half in one sector of the graph and half in the opposite sector. Forcing the regression through the origin accounts for this property of the points (20).

Correlated divergence analyses are based on differences in traits between descendant lineages. They do not attempt to infer the ancestral values of climate or trait variables from the past. Thus, contrast analyses are not able to distinguish between no change in a trait and the case where both descendant lineages have undergone change in the same direction (for example, if both lineages were to experience a halving of seed mass in response to climate cooling). Empirical analysis has shown that correlated divergence analysis can accurately estimate the correlations between evolutionary changes even when estimates of ancestral traits are not accurate (21). In using correlated divergence analysis on the climatic conditions under which species occur, we assume that there is a heritable basis to species' environmental tolerances.

We performed cross-species analyses using SPSS version 13.0. In these analyses, species are treated as replicates.

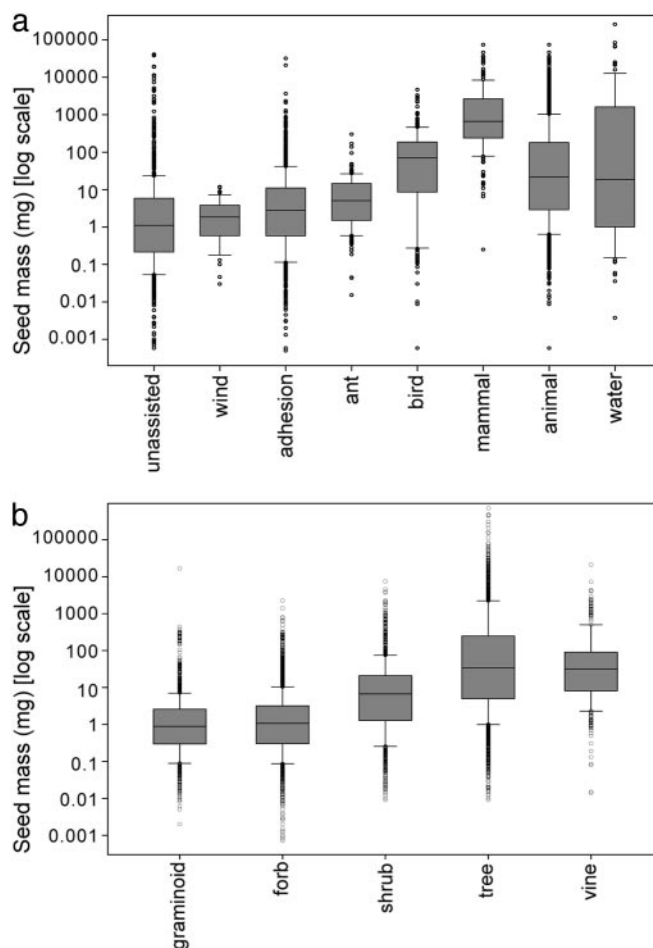
## Results

We begin by considering correlations across present-day species as background for the divergence analyses that follow. Seed mass was positively correlated with NPP, LAI, precipitation, and temperature ( $P < 0.001$ ; Fig. 1). The correlation with precipitation was strongest ( $R^2 = 0.15$ ), followed by NPP ( $R^2 = 0.14$ ), LAI ( $R^2 = 0.11$ ), and then temperature ( $R^2 = 0.09$ ). The slopes of these relationships were moderately steep: there was a 41-fold increase in seed mass with every 10-fold increase in precipitation (a 5,500-fold shift in mean seed mass across the range of precipitation represented in our dataset), an 11-fold increase in seed mass with every  $10^\circ$  increase in temperature, and 41- and 51-fold increases in seed mass with 10-fold increases in NPP and LAI, respectively.

As in previous studies (22, 23), we found that species with unassisted dispersal or wind dispersal had smaller seeds than species dispersed by animals or water (Fig. 2*a*). We also confirmed that herbs and grasses generally make smaller seeds than shrubs, which generally make smaller seeds than trees or vines (Fig. 2*b*).

Correlations among LAI, NPP, plant height, latitude, temperature, precipitation, growth form, and seed dispersal syndrome are summarized in Fig. 4, which is published as supporting information on the PNAS web site.

Next, we performed correlated divergence analysis to determine how tightly divergences in seed mass have been associated with divergences in other plant traits and climate variables. Divergences in seed mass were positively correlated with divergences in temperature, NPP, LAI, precipitation, and growth

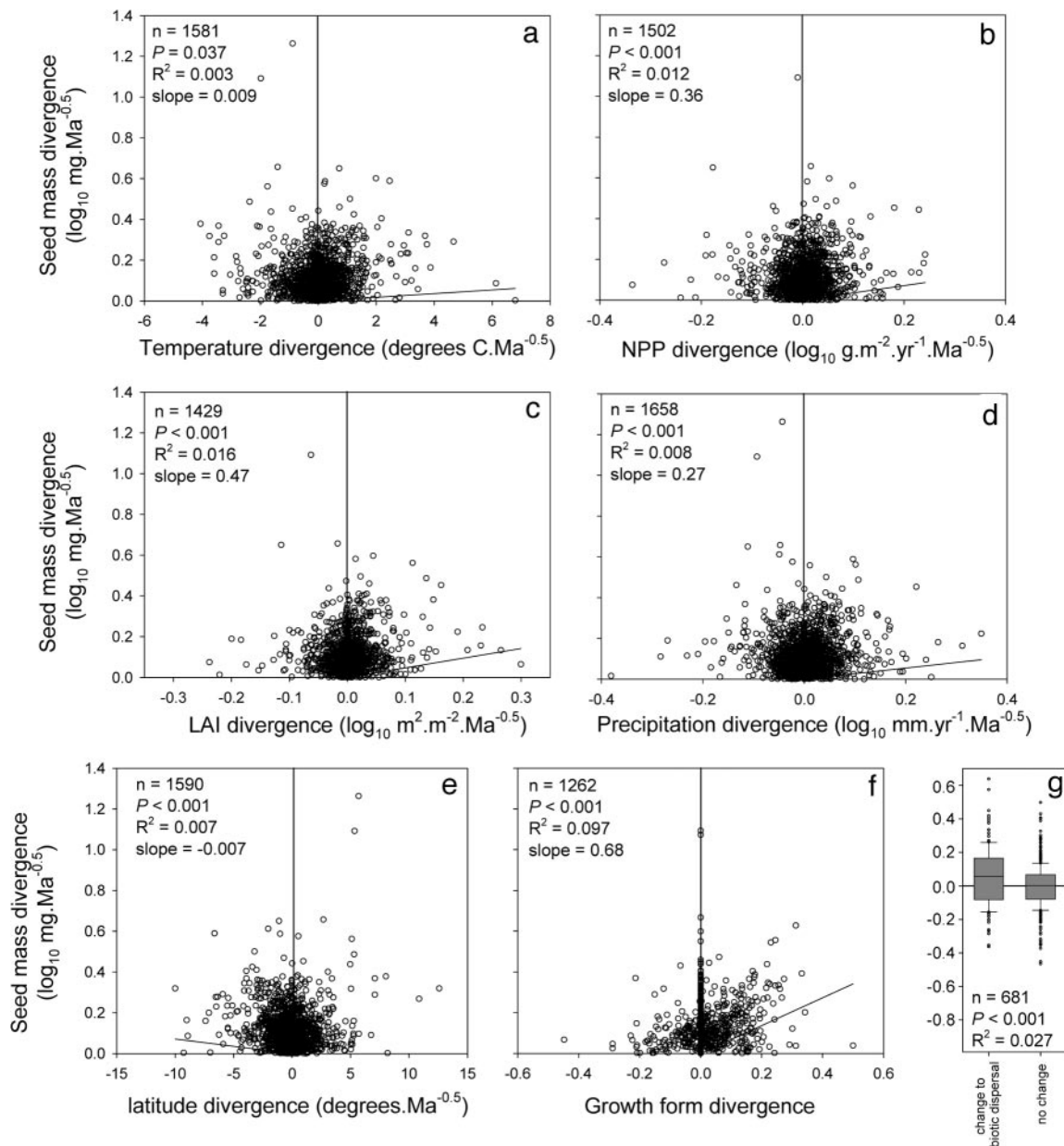


**Fig. 2.** Cross-species relationships between seed mass and dispersal syndrome (*a*) ( $n = 4,159$  species) and growth form (*b*) ( $n = 7,211$  species). We have presented dispersal syndrome data in eight categories here to illustrate the relationship between the mass of animal-dispersed seeds and the types of animals that disperse them. The "animal" category in this figure was used for species whose dispersal agent was not identified beyond "animal" in the source paper. We had to pool these data to two categories; biotic and abiotic dispersal, for phylogenetic analyses (because the computer software cannot cope with unordered multistate categorical variables yet).

form, and they were negatively correlated with divergences in latitude (Fig. 3). Although all of these relationships were statistically significant ( $P < 0.001$ ), none of the climate variables explained more than 2% of the variation in seed mass contrasts. By comparison, contrast in dispersal syndrome accounted for 3% of the variation, and contrast in growth form explained 10% of the variation in seed mass contrasts. Thus, divergences in seed mass have been much more strongly associated with divergences in other plant traits than with divergences in the physical environment in which species occur.

The fact that the cross-species relationships have more predictive power than the phylogenetic relationships indicates that seed mass and associated traits exhibited a high degree of phylogenetic signal, such that a relatively small number of divergences deep in the phylogeny have a strong influence on the cross-species relationship. These influential divergences are discussed in ref. 5.

Finally, we constructed a general linear model, with seed mass contrast as the dependent variable, and contrasts in latitude, temperature, precipitation, LAI, NPP, dispersal syndrome, and growth form as predictors (using only the 567 divergences for



**Fig. 3.** We used correlated divergence analysis to determine whether divergences in seed mass have been associated with divergences in environmental conditions or other plant traits. Relationships between divergence in  $\log_{10}$  seed mass and divergence in mean annual temperature (a),  $\log_{10}$  NPP (b),  $\log_{10}$  LAI (c),  $\log_{10}$  mean annual precipitation (d), latitude (e), and growth form (f). The box plot (g) shows the magnitude of the contrast in  $\log_{10}$  seed mass for divergences in which there was a shift from abiotic to biotic dispersal. Regression statistics are the results of linear regression through the origin, except those in g, which are from logistic regression, with a Cox and Snell  $R^2$  (SPSS version 11.01). Sample sizes are the number of divergences.

which all of the above data were available). The full model (including all predictor variables and all interaction terms) had an  $R^2$  value of 0.53 ( $P < 0.001$ ). Thus, about half of the variation in divergence in seed mass is explained by divergences in our seven predictor variables. The remaining variation is likely to be attributable to microenvironmental factors that are not captured by site information and to interactions between species.

### Discussion

Divergences in growth form explained more than three times as much of the variation in divergences in seed mass as did divergences in dispersal syndrome, LAI, or temperature. Thus, our results do not support Tiffney's (2) hypothesis (changes in dispersal syndrome have been the major driver of seed evolution)

or the hypothesis of Eriksson *et al.* (ref. 6; changes in seed mass were driven by canopy closure). However, our results do support one hypothesis of Eriksson *et al.* (6), that changes in seed mass are driven by changes in growth form.

The idea that changes in seed mass are predominantly driven by changes in growth form is consistent with the fact that plant size is the strongest correlate of seed mass across present-day species (9). It is also consistent with the fact that 9 of the 10 largest divergences in seed mass in the history of plants were associated with a divergence in growth form (5).

Our finding that divergences in seed mass are most closely associated with divergences in growth form is also consistent with the fossil record. Early angiosperm seeds were mostly small (2, 6), and the earliest angiosperms are currently thought to have

been short-lived, small, woody, or sometimes herbaceous species that grew in highly disturbed forest understory environments (3, 24). Median and maximum seed mass increased dramatically from  $\approx 85$  Ma to shortly after the Cretaceous–Tertiary boundary (65 Ma; ref. 6), and it appears that angiosperms radiated into a wider range of growth forms over this same time period. Fossil wood deposits are almost exclusively from gymnosperms before the Turonian (93.5 Ma), but records of dicotyledon wood increase through the later Cretaceous (25). By the late Campanian or early Maastrichtian (around 75 Ma), there were at least some angiosperm-dominated forests in existence, including dicotyledonous trees with trunks up to 1 m in diameter (3). Most Cenozoic floras encompass a wide range of seed mass strategies and a wide range of plant growth forms (2, 3, 6).

The relatively weak association between divergences in seed mass and divergences in climate might seem surprising. However, this finding is consistent with the observation that coexisting species commonly span five or six orders of magnitude of variation in seed mass (26). Our results are remarkably similar to those of Wright *et al.* (27), who showed significant, yet relatively weak effects of climate on a suite of leaf traits across 2,548 species from around the world, with correlations between different leaf traits far outweighing the effects of climate.

So why have plant size and seed mass been correlated through plant evolution? One theory is that these variables are part of a

spectrum of coordinated life history traits, including plant size, plant lifespan, time to first reproduction, seedling survival, and seed mass (as in Charnov's formulation for mammals; ref. 28). These variables are correlated because large plants require longer to reach their adult size (13). To survive this lengthy juvenile period, species with large adult size need to have high rates of seedling survival, which are achieved by producing larger seeds (29). The correlations between these life history variables suggest that plant trait evolution might be usefully understood through models that incorporate species' life history speed, such as the *r-K* spectrum (30) or metabolic rate (31).

Rick Condit, Sandra Díaz, Peter Juniper, Michelle Leishman, Janice Lord, Margaret Mayfield, Barbara Rice, Ken Thompson, Ian Wright, and S. Joseph Wright all gave us access to unpublished data. Thanks to Bruce Tiffney and Erin Leakey for helpful discussion of the paleobotanical literature and to Rick Reeves for help with data compilation. This work was supported by the National Center for Ecological Analysis and Synthesis, Australian Research Council Funding (to M.W.), and a National Science Foundation grant (to C.O.W., M. J. Donoghue, and D.D.A.). The Millennium Seed Bank Project is funded by the U.K. Millennium Commission, The Wellcome Trust, and Orange PLC. The Royal Botanic Gardens, Kew, are partially funded by the U.K. Department for Environment, Food, and Rural Affairs.

1. Axelrod, D. I. (1959) *Science* **130**, 203–207.
2. Tiffney, B. H. (1984) *Ann. Mo. Bot. Gard.* **71**, 551–576.
3. Wing, S. L. & Boucher, L. D. (1998) *Annu. Rev. Earth Planet. Sci.* **26**, 379–421.
4. Crane, P. R. & Lidgard, S. (1989) *Science* **246**, 675–678.
5. Moles, A. T., Ackerly, D. D., Webb, C. O., Tweddle, J. C., Dickie, J. B. & Westoby, M. (2005) *Science* **307**, 576–580.
6. Eriksson, O., Friis, E. M. & Lofgren, P. (2000) *Am. Nat.* **156**, 47–58.
7. Harper, J. L., Lovell, P. H. & Moore, K. G. (1970) *Annu. Rev. Ecol. Syst.* **1**, 327–356.
8. Tiffney, B. H. (2004) *Annu. Rev. Ecol. Syst.* **35**, 1–29.
9. Leishman, M. R., Westoby, M. & Jurado, E. (1995) *J. Ecol.* **83**, 517–529.
10. Foster, S. A. & Janson, C. H. (1985) *Ecology* **66**, 773–780.
11. Salisbury, E. (1974) *Proc. R. Soc. London Ser. B* **186**, 83–88.
12. Mazer, S. J. (1989) *Ecol. Monogr.* **59**, 153–175.
13. Moles, A. T., Falster, D. S., Leishman, M. R. & Westoby, M. (2004) *J. Ecol.* **92**, 384–396.
14. Felsenstein, J. (1985) *Am. Nat.* **125**, 1–15.
15. Kaplan, J. O., Bigelow, N. H., Prentice, I. C., Harrison, S. P., Bartlein, P. J., Christensen, T. R., Cramer, W., Matveyeva, N. V., McGuire, A. D., Murray, D. F., *et al.* (2003) *J. Geophys. Res.* **108**, 1–17.
16. Webb, C. O. & Donoghue, M. J. (2005) *Mol. Ecol. Notes* **5**, 181–183.
17. Wikström, N., Savolainen, V. & Chase, M. W. (2001) *Proc. R. Soc. London Ser. B* **268**, 2211–2220.
18. Oakley, T. H. & Cunningham, C. W. (2000) *Evolution* **54**, 397–405.
19. Levin, D. A. (1974) *Am. Nat.* **108**, 193–206.
20. Garland, T., Harvey, P. H. & Ives, T. R. (1992) *Syst. Biol.* **41**, 18–32.
21. Pagel, M. D. (1992) *J. Theor. Biol.* **156**, 431–442.
22. Grafen, A. (1989) *Philos. Trans. R. Soc. London B* **119**, 119–157.
23. Lord, J., Egan, J., Clifford, T., Jurado, E., Leishman, M., Williams, D. & Westoby, M. (1997) *J. Biogeogr.* **24**, 205–211.
24. Feild, T. S., Arens, N. C. & Dawson, T. E. (2003) *Int. J. Plant Sci.* **164**, S129–S142.
25. Wheeler, E. A. & Baas, P. (1991) *Int. Assoc. Wood Anat. Bull.* **12**, 275–332.
26. Leishman, M. R., Wright, I. J., Moles, A. T. & Westoby, M. (2000) in *Seeds: The Ecology of Regeneration in Plant Communities*, ed. Fenner, M. (CAB International, Wallingford, U.K.), pp. 31–57.
27. Wright, I. J., Reich, P. B., Westoby, M., Ackerly, D. D., Baruch, Z., Bongers, F., Cavender-Bares, J., Chapin, F. S., Cornelissen, J. H. C., Diemer, M., *et al.* (2004) *Nature* **428**, 821–827.
28. Charnov, E. L. (1993) *Life History Invariants: Some Explorations of Symmetry in Evolutionary Ecology* (Oxford Univ. Press, Oxford).
29. Moles, A. T. & Westoby, M. (2004) *J. Ecol.* **92**, 372–383.
30. Pianka, E. R. (1970) *Am. Nat.* **104**, 492–597.
31. Enquist, B. J., West, G. B., Charnov, E. L. & Brown, J. H. (1999) *Nature* **401**, 907–911.