New Phytologist

Research

Global biogeography of seed dormancy is determined by seasonality and seed size: a case study in the legumes

Rafael Rubio de Casas[1,2,3], Charles G. Willis[4], William D. Pearse[5,6,7,8], Carol C. Baskin[9,10], Jerry M. Baskin[9] and Jeannine Cavender-Bares[7]

[1]Department of Ecology, Facultad de Ciencias, Universidad de Granada, Avda. de la Fuentenueva s/n, Granada 18071, Spain; [2]Estación Experimental de Zonas Áridas, EEZA-CSIC, Carretera de Sacramento s/n, La Cañada de San Urbano 04120, Spain; [3]UMR 5175 CEFE – Centre d'Ecologie Fonctionnelle et Evolutive (CNRS), 1919 Route de Mende, Montpellier Cedex 05 F-34293, France; [4]Harvard University Herbaria, 22 Divinity Ave., Cambridge, MA 02138, USA; [5]Department of Biology, McGill University, Stewart Biology Building, 1205 Ave. Docteur Penfield, Montreal, QC Canada, H3A 1B1; [6]Département des Sciences Biologiques, Université du Québec à Montréal, Succursale Centre-ville, Montréal, QC Canada, H3C 3P8; [7]Department of Ecology, Evolution & Behavior, University of Minnesota, 1479 Gortner Ave, Saint Paul, MN 55108, USA; [8]Department of Biology & Ecology Center, Utah State University, Logan, UT 84322, USA; [9]Department of Biology, University of Kentucky, Lexington, KY 40506, USA; [10]Department of Plant and Soil Sciences, University of Kentucky, Lexington, KY 40546, USA

Authors for correspondence:
Rafael Rubio de Casas
Tel: +34 958 249 861
Email: rubiodecasas@ugr.es

Jeannine Cavender-Bares
Tel: +1 612 624 6337
Email: cavender@umn.edu

Received: *18 November 2016*
Accepted: *19 January 2017*

New Phytologist (2017) **214:** 1527–1536
doi: 10.1111/nph.14498

Key words: discrete traits, mixed models, phylogenetic comparative methods, physical dormancy, seed bank, seed germination, seed predation.

Summary

- Seed dormancy is expected to provide ecological advantages by adjusting germination to the favorable growth period. However, many species produce nondormant seeds, particularly in wet tropical forests, a biogeographic pattern that is not well accounted for in current models.
- We hypothesized that the global distribution of dormant seeds derives from their adaptive value in predictably fluctuating (i.e. seasonal) environments. However, the advantage conferred by dormancy might ultimately depend on other seed attributes, particularly size. This general model was tested within a phylogenetically informed framework using a data set comprising > 216 000 world-wide observations of Fabaceae, spanning three orders of magnitude in seed size and including both dormant and nondormant seeds.
- Our results confirmed our hypothesis: nondormant seeds can only evolve in climates with long growing seasons and/or in lineages that produce larger seeds. Conversely, dormancy should be evolutionarily stable in temperate lineages with small seeds. When the favorable season is fleeting, seed dormancy is the only adaptive strategy.
- Based on these results, we predict that, within a given lineage, taxa producing larger, nondormant seeds will necessarily predominate in aseasonal environments, while plants bearing small, dormant seeds will be dominant under short growing seasons.

Introduction

Seeds represent a critical stage in the life history of plants. The timing of germination can determine the environment experienced throughout the life of the plant and thus is crucial for survival and fitness (Donohue *et al.*, 2010). Germination is controlled by factors intrinsic and extrinsic to the seed, and it can only occur under a specific combination of environmental conditions. Seeds of many plants do not germinate even if conditions are favorable: they are dormant (Finch-Savage & Leubner-Metzger, 2006; Baskin & Baskin, 2014). However, other seeds lack dormancy, particularly in tropical rainforests (Moles *et al.*, 2007; Baskin & Baskin, 2014). It has been hypothesized that in constant environments dormancy may actually be maladaptive by precluding early germination and thus limiting the time for growth and development (Ellner, 1987). However, to our knowledge the generality of this hypothesis has never been tested at a global scale.

Seed size is another attribute that can influence plant recruitment. Compared with seedlings from small seeds, large seeds produce better provisioned and larger seedlings that can perform better under unfavorable conditions (Kitajima & Fenner, 2000). Large seeds may not need to undergo dormancy to avoid unfavorable conditions, but rather they may benefit from early germination after maturity, thereby maximizing their growth before the onset of unfavorable conditions. Early germination of large seedlings might enable the development of deeper root systems and defenses against aboveground herbivory and climatic stress, before the onset of the unfavorable season (Brunner *et al.*, 2015). For this reason, it has been hypothesized that dormancy and seed size (i.e. better seed provisioning) may serve as alternative adaptive strategies (Venable & Brown, 1988; Rees, 1996). However, seed size is constrained by factors such as the maximum adult plant size (Venable & Rees, 2009) and larger seeds also require greater resource allocation, which is associated with a lower number of seeds produced, and may lead to higher risk of predation (Gómez & Husband, 2004).

Thus, seed dormancy and size are expected to respond to selective forces in addition to climate. Seed predation, in particular, is expected to be an important selective factor (Janzen, 1971; Brown & Venable, 1991). Because seeds can only remain ungerminated in the soil if they are not eaten (i.e. the predation risk increases with time), Dalling *et al.* (2011) hypothesized that dormancy should depend on the likelihood of seed predation and that nondormancy might evolve as part of a predator avoidance strategy. All else being equal, the risk of predation is thought to be proportional to seed size because larger seeds are easier to detect and are more nutritious than smaller ones (Janzen, 1971; Gómez & Husband, 2004; Long *et al.*, 2015). When the risk of predation of larger seeds is lower than that of smaller seeds, it is usually because size is positively correlated with defense mechanisms (Schutte *et al.*, 2014). Assuming that the probability of survival in the soil is inversely proportional to seed size, Volis & Bohrer (2013) used mathematical modeling to explore the association between seed size and dormancy in annual plants. They predicted that under constantly unfavorable environments (in their case, deserts) large, nondormant seeds would be selected for, whereas in temporally unpredictable environments (e.g. semiarid ecosystems with high inter-annual variation in precipitation conditions) dormant seeds would be selected for, and their size would be dependent upon the likelihood of predation of large seeds. In other words, the adaptive value of dormancy might depend simultaneously on the variability of the environment and on the probability of survival of large seeds in the seedbank.

However, Volis & Bohrer's model, as well as most models of the adaptive value of seed dormancy and size, has focused on their roles as bet-hedging mechanisms for annual plants to cope with habitat unpredictability. Seed dormancy and size clearly provide bet-hedging by spreading the risk of survival to reproduction, which is adaptive under temporal stochastic environments, such as deserts or ruderal habitats (Pake & Venable, 1996; Childs *et al.*, 2010; Metz *et al.*, 2010; Rubio de Casas *et al.*, 2015). Conversely, the linkage of seed dormancy and size to reliable fluctuations in environmental conditions (i.e. seasonality) has been largely unexplored, particularly in the case of perennials.

From a macroevolutionary perspective, seed dormancy and size tend to be largely conserved within major plant lineages (Moles *et al.*, 2005; Willis *et al.*, 2014). Additionally, there are broad biogeographic trends in the global latitudinal distribution of the two traits across latitudes. Specifically, seed dormancy tends to decrease and seed size to increase towards the equator (Moles & Westoby, 2003; Morin & Chuine, 2006; Baskin & Baskin, 2014). Other plant traits have been shown to follow a similar latitudinal trend as a result of seasonality. For example, leaf life span increases towards the equator and decreases towards the poles, as a consequence of changes in the length of the unfavorable season for plant growth (Reich *et al.*, 1997). The dependence of leaf life span on the duration of the favorable season was modeled by Kikuzawa (1996). This approach provides a practical conceptual framework for considering how other traits, such as seed dormancy and size, respond to latitudinal variation in climate seasonality.

Dormancy can be hypothesized to be particularly adaptive in seasonal environments by preventing germination during short and unpredictable favorable periods outside of the growing season, such as a few warm days in winter or a short rainy spell in the dry season (Donohue *et al.*, 2010). In the case of seed size, Moles & Westoby (2003) hypothesized that the length of the growing period might affect the maximum attainable seed mass for a given lineage in a given environment because longer growing seasons provide more time for carbon accumulation. Selection will thus be more likely to favor large seeds in tropical environments, where the high-energy period extends for all or most of the year. Longer growing seasons might simultaneously increase the benefits associated with early germination and enable the development of larger seeds. Although bet-hedging is undoubtedly important for the evolution of dormancy, here we posit that it is not only the predictability of the environment (or lack thereof) but also the length of the favorable season and the risk of predation that determine the adaptive value of seed dormancy and size.

Our working hypothesis (formalized in Supporting Information Methods S1) is that dormancy should be increasingly selected for under more seasonal environments to ensure that germination and early development occur during the window of favorable conditions, but that its adaptive value is contingent on seed size. In any given environment, larger seeds are predicted to be less dependent on dormancy because they produce more vigorous and robust seedlings than smaller seeds (Fig. S1).

To validate these predictions, we generated a large phylogenetically controlled data set of seed dormancy and size in the legume family (Fabaceae). With these data, we tested the association of seed dormancy and size with seasonality at a global scale, and investigated how seed size and seasonality have affected evolutionary transitions between dormancy and nondormancy. Lastly, based on our results, we put forward several general testable hypotheses about the evolutionary ecology of seed size and dormancy at a global scale.

Materials and Methods

Data

We used the legume family (Fabaceae) as a case study. The legumes are a useful system in which to investigate large-scale plant diversity (Yahara *et al.*, 2013) and, more particularly, to test our hypotheses given their global distribution and variation in seed size and dormancy state (Lewis *et al.*, 2005; Baskin & Baskin, 2014). The Fabaceae include three subfamilies (Faboideae, Caesalpinioideae and Mimosoideae) that diversified relatively recently (Lavin *et al.*, 2005) and encompass 730 extant genera and *c.* 19 400 species, *c.* 7% of all angiosperms (Stevens, 2001; Judd *et al.*, 2016). The seed coat of many legume seeds includes a hard, water-impermeable palisade layer of cells that provides mechanical protection for the embryo (Baskin & Baskin, 2014). Additionally, legume seeds contain multiple alkaloids that protect them against predators (Wink & Mohamed, 2003). These defense traits are generally shared among close relatives. Consequently, the differential survival of seeds of related taxa in the soil is probably attributable to differences in nondefensive traits such as size. Legumes are also predominantly perennial (trees and shrubs) which also minimizes the differences potentially caused by alternative life histories.

The source of our dormancy data was Baskin & Baskin (2014). This data set was compiled over the last several decades from several thousand published papers that represent a global distribution of species. The initial data set included information for 1529 Fabaceae species representing *c*. 50% of the recognized genera within the family. Of these species, 1406 are unambiguously classified as perennial and 111 are annuals. The results of the models fitted to the data did not seem to be qualitatively influenced by the inclusion of annual species in the data set, probably because there were so few annuals. Keeping both annuals and perennials should make our analyses more general, and we present only the results obtained with the full data set.

Based on the classification system of Baskin & Baskin (2004), there are three kinds (classes) of dormancy in the Fabaceae: physiological dormancy (PD) – that is, dormancy is broken by specific physiological responses to environmental cues, such as warm and/or cold temperatures, or by dry after-ripening; physical dormancy (PY) – where seeds are water-impermeable as a result of a water-impermeable palisade layer(s) of cells, and dormancy is broken in response to environmental cues (e.g. fluctuating temperatures) that cause a ‘water gap’ to open; and physiophysical dormancy (PY + PD) – dormancy is broken by a combination of physiological and physical environmental cues (Nikolaeva, 1967; Baskin & Baskin, 2004). Of these three dormancy classes, PY is predominant in Fabaceae, being present in > 75% of all species included in this study; 79% of those with Global Biodiversity Information Facility (GBIF) distribution data. In addition, we also recognized species as nondormant. Nondormant seeds lack primary dormancy at seed maturity and germinate under the widest range of conditions possible immediately after dispersal without any dormancy-breaking treatments (Baskin & Baskin, 2004, 2014). For this study, we treated all three kinds of dormancy as a single character state and compared ‘dormant’ (D) and ‘nondormant’ (ND) species. Our final data set contained 1291 dormant and 238 nondormant species.

Geo-referenced data were obtained from the GBIF database (v.2.0.3; January 2011). We compiled data only for species that were present in the dormancy data set and that had at least 20 accessions in the GBIF repository. This resulted in geo-spatial data for 532 species (485 D and 47 ND, representing 170 genera). Of these species, 475 were perennial and 57 were annual. Raw data were scrubbed for obvious mistakes (e.g. accessions recorded as occurring in the sea or for which the sign of the coordinates was inverted) both with automated scripts and by hand. The final, corrected data set included 216 600 records with a median of 415 records per species.

Seed mass, used as a proxy of total seed size, was obtained for all the species with dormancy data, principally from the Kew Gardens Seed Information Database (Royal Botanic Gardens Kew, 2012) and from the primary literature as needed. All seed size data analyses were natural log-transformed.

Environmental conditions

We used the coordinates of GBIF accessions to estimate the geographic distribution and climatic conditions of each species. We quantified seasonality using three alternative metrics. First, we computed a unique measure of seasonality based on a principal components analysis (PCA) of a subset of bioclimatic variables from the WorldClim Global Climate Database (www.worldclim.org). These variables were used because they capture different aspects of seasonal variation and included isothermality (WorldClim variable BIO3), temperature seasonality (BIO4), minimum temperature of the coldest month (BIO6), annual temperature range (BIO7), precipitation of the driest month (BIO14), precipitation seasonality (BIO15) and precipitation of the driest quarter (BIO17). Data for each variable were extracted based on species occurrence records and combined with a PCA, using the ‘PRINCIPAL’ function with ‘VARIMAX’ rotation in R v.2.15 (R Development Core Team, 2009). To control for artifacts associated with PC calculation, we also reduced environmental variation to two axes by means of a nonmetric multidimensional scaling (NMDS) of the median bioclimatic values for each species using the function ‘metaMDS’ of the VEGAN R package with Euclidean distances, 100 independent starts and a maximum number of iterations of 500. As additional seasonality metrics, we used two temperature-specific indices: growing-degree days (GDD) and frost day frequency (FDF). GDD was obtained from the Atlas of the Biosphere (http://nelson.wisc.edu/sage/; based on data from CRU TS 2.1; see below), which bases its estimates on the formula of New *et al.* (1999) with a 5°C base temperature. FDF was obtained from the Climatic Research Unit (CRU) of the University of East Anglia Global Climate Database (CGIAR-CSI CRU TS 2.1; Mitchell & Jones, 2005) and represents the average number of days in a year that experience frost.

Phylogenetic construction

All models fitted to the data were corrected for the influence of co-ancestry in trait distribution by incorporating phylogenetic hypotheses of relatedness based on 650 bootstrap composite phylogenies. These phylogenies were constructed from an original set of 650 bootstrap phylogenies published by the Legume Phylogeny Working Group (LPWG) (The Legume Phylogeny Working Group, 2013). The LPWG phylogenies were based on a data set of three genes (*matK*, *rbcL* and ITS) for 1276 taxa and estimated using maximum likelihood (RAxML v.7.27; Stamatakis *et al.*, 2008). For these 650 bootstrap trees, we estimated divergence times using penalized maximum likelihood in the program TreePL (Smith & O’Meara, 2012), based on five fossil constraints taken from Lavin *et al.* (2005) and one previously estimated divergence date for the Fabaceae crown group (Bell *et al.*, 2010). To adapt these phylogenetic hypotheses to our data, we grafted the species present in our dormancy data set onto the revised backbone phylogenies. Species were grafted on at the generic level, with the crown age of the genus set as the node age of the most recent common ancestor. Taxa with dormancy data that were not present in the LPWG phylogeny were excluded.

We created two sets of composite phylogenies to address genera with multiple species. For the first set, genera with multiple species were left as polytomies. For the second set, these genera were randomly resolved as a bifurcating clade using the function 'RCOAL' in the R package APE v.3.2 (Paradis *et al.*, 2004).

Association of seed dormancy and size with seasonality

To examine whether the distribution of seed dormancy is significantly associated with differences in seed size and the duration of the favorable season, we regressed seed dormancy against the environmental and life history data. As species' shared evolutionary history can bias estimates of trait correlations, all models were fitted controlling for phylogenetic nonindependence among species. Full models were of the form:

$$\text{Dormancy} \sim \text{seasonality} \times \text{seed size} \mid \text{phylogeny}$$

where dormancy was considered to be a logistic response variable (0: nondormant; 1: dormant), seasonality was estimated with either the axes of the PCA of bioclimatic variables, GDD or FFD, and seed size was the log of seed size of each species. Then, to estimate the relative effect of seasonality and seed size on dormancy, the full models were compared with univariate models incorporating only seasonality or seed size as the predictor of dormancy. A significant effect of either or both of the predictors can be taken as an indication of the adaptive value of dormancy; that is, dormancy has evolved in response to seed size and/or seasonality. We first fitted the data using a phylogenetic logistic model as implemented in the PHYLOLM R package (Tung Ho & Ané, 2014) using median size and seasonality estimates for each species. These logistic models were run over the 650 bootstrap replicates. We tried to approximate intraspecific variation in ecological conditions by fitting models incorporating variation in climatic variables at the population scale. For this analysis, we used MCMCGLMM (Hadfield, 2010), a Bayesian approach that accounts for nonindependence among species by including the phylogenetic relationships as a random variable. In all analyses, we used weakly informative Gelman priors (Gelman *et al.*, 2008; Hadfield, 2010) and determined the number of iterations, thinning and the burn-in period using diagnostics in the CODA package (Plummer *et al.*, 2006). Final analyses were run over the consensus tree of the 650 replicates using four chains of 2×10^6 generations with a burn-in of 2×10^4 and a thinning of 25. Parameter estimates were computed by combining the output of the four chains. In MCMCGLMM analyses, all 216 600 observations were used but phylogenetic autocorrelation was corrected for using a single phylogeny, whereas in PHYLOLM analyses we used the 650 trees but based the PCA only on the median values per species of the bioclimatic variables ($n = 522$). The fit of the models to the data was estimated based on Akaike information criterion (AIC) (PHYLOLM) or deviance information criterion (DIC) (MCMCGLMM) values. In the latter case, we also evaluated the explanatory power of our models by computing the conditional R^2 (i.e. the proportion of variance explained by both the fixed and random factors – the phylogeny) using the approach described in Nakagawa & Schielzeth (2013) with code available at S. Nakagawa's website (http://www.i-deel.org/) modified for MCMCGLMM output.

Evolutionary shifts between dormancy and nondormancy

Comparing the dormancy of extant species across environments while controlling for phylogeny is distinct from comparing the past environmental conditions under which dormancy or nondormancy evolved. If lack of dormancy is adaptive under specific combinations of environmental conditions and seed sizes, then it should have evolved not only in present-day habitats but also in the past, probably repeatedly, whenever the appropriate conditions were met (i.e. low seasonality and large seeds).

To analyze the evolution of shifts in dormancy in the Fabaceae, we developed a new technique that can identify associations between evolutionary shifts in dormancy states and the other variables across the phylogeny. In this analysis, we focused on the past environmental conditions associated with transitions between dormancy and nondormancy. First, we reconstructed the evolution of a reduced subset of dormancy categories (D and ND) using the stochastic character mapping approach (SIMMAP; Huelsenbeck *et al.*, 2003; Bollback, 2006). We ran 1000 iterations of the MAKE.SIMMAP function in the R package PHYTOOLS (Revell, 2012). Second, we reconstructed seed size and environmental conditions throughout the phylogeny under Brownian motion (BM) using maximum likelihood (1000 iterations of the ANC.ML function in PHYTOOLS). Additionally, to control for the potential bias caused by the evolutionary model used to reconstruct traits, we repeated the analyses using Ohrstein–Uhlenbeck models considering that evolution might deviate from pure BM to attain a 'phenotypic optimum' (OU; Butler & King, 2004; Beaulieu *et al.*, 2012). Finally, we calculated the number of times a shift between dormancy states was reconstructed across each branch of the phylogeny. Initially, the ancestral 'root' state was not coerced and we let SIMMAP determine the most likely state. Then, to control for the potential influence of the root state on shifts across the tree, the analyses were repeated forcing the root state to be either 'dormant' or 'nondormant'. We emphasize that these are analyses of *shifts* between dormancy states and not regression-like models of seed dormancy as a function of seed size and climate accounting for phylogenetic nonindependence, as in the PHYLOLM and MCMCGLMM models described earlier. All analyses were run across all 650 bootstrap trees.

Results

Global distribution of seed size, dormancy and seasonality

The distribution of the legume seed dormancy and size data at a global scale is shown in Fig. 1. Dormancy is more prevalent in the temperate zones. Seed size also exhibited a latitudinal distribution; species with the largest seeds were concentrated in the lower latitudes. As expected, environmental variables used to approximate seasonality also followed a latitudinal pattern. GDD had a clear maximum in the tropical region, between 20°N and 25°S, which is also largely frost-free (Fig. 1). The results of the PCA of bioclimatic variables are shown in Table 1. According to

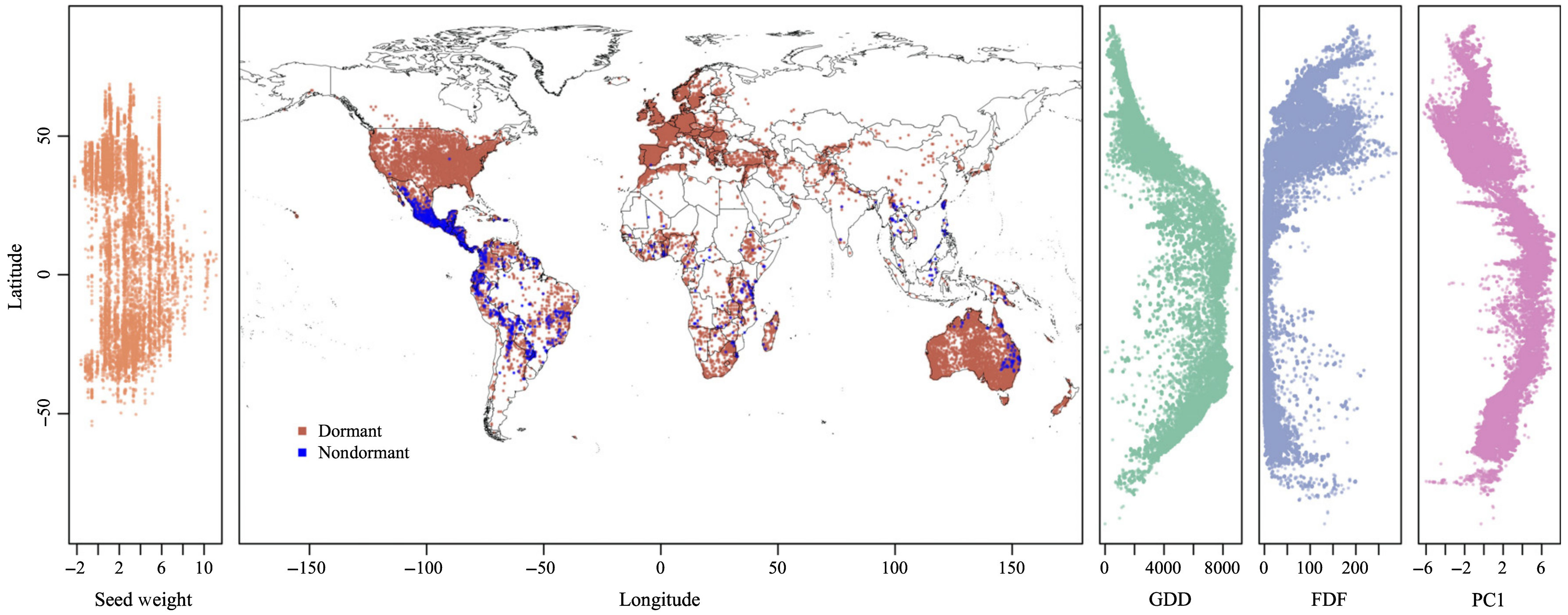

Fig. 1 Global distribution of dormancy, seed size and seasonality among the legumes (Fabaceae). The following traits are graphed along a latitudinal gradient using all the available individual observations for each species. Left to right: seed size ($\log_{10}(g)$); dormancy; growing-degree days (GDD); frost day frequency (FDF); and the first principal component (PC1) of multiple temperature and precipitation indicators of seasonality acquired from the Bioclim data set.

Table 1 Results of the principal components analysis (PCA) of bioclimatic variables

	Full data set		Median values	
Variable	PC1 49.0%	PC2 37.2%	PC1 55.8%	PC2 36.4%
BIO3	0.470	–	0.475	–
BIO4	−0.449	−0.317	−0.498	–
BIO6	0.493	0.130	0.485	–
BIO7	−0.259	−0.463	−0.483	–
BIO14	−0.241	0.528	0.125	−0.592
BIO15	0.400	−0.310	0.148	0.535
BIO17	−0.221	0.534	0.146	−0.586

Values indicate the loadings on the first two axes of each of the variables included in the analyses. These were the only axes found to have eigenvalues > 1. The percentage of variance absorbed by each axis is indicated below its name. BIO3, isothermality; BIO4, temperature seasonality; BIO6, minimum temperature of the coldest month; BIO7, annual temperature range; BIO14, precipitation of the driest month; BIO15, precipitation seasonality; BIO17, precipitation of the driest quarter.

these analyses, the first two axes had eigenvalues > 1 and explained > 86% of the total variance (49% PC1 and 33% PC2; Table 1; Fig. 1). Therefore, we decided to use PC1 as a multidimensional proxy of seasonality. This variable also had a clear latitudinal pattern, with maximum values around the equator and minimum values at high latitudes. Nevertheless, PC1 displayed more variation than the other seasonality metrics, with higher values in equivalent latitudes for locations that have less extreme seasons (e.g. PC1 values were higher for Europe than for North America). PC1 was probably more nuanced than GDD or FDF because it incorporated variation in both temperature and precipitation. Variables that contributed the most to this axis were minimum temperature of the coldest month (BIO6), isothermality (BIO3), temperature seasonality (BIO4) and precipitation seasonality (BIO15; Table 1). The PCA of median species values resulted in two axes with eigenvalues > 1 which explained > 91% of the variance (Table 1). The first axis explained 55% of the variance and was determined primarily by the temperature variables, while the second axis (36%) was determined exclusively by the precipitation variables (Table 1). The results of the other technique used to reduce the dimensionality of our data (NMDS) yielded very similar results to those of PCA. The first score also exhibited a clear latitudinal pattern closely related to that of PC1 ($R^2 > 97\%$; Fig. S2).

Seed dormancy dependence on seasonality and seed size

The simple mathematical model presented in Methods S1 predicted a fitness increase associated with dormancy as the length of the favorable season decreased. The length of the favorable season at which dormant and nondormant strategies were equivalent (i.e. the adaptive equilibrium) depended on seed size. In general, larger seeds were less dependent on dormancy (Methods S1; Table S1; Fig. S1). Our empirical results supported these findings. The global biogeographic distribution of seed traits and the associated environmental variables were strongly correlated (Fig. 1; Table 2). Although this pattern could be observed across the phylogeny (Fig. S3), this association appeared not to be

Table 2 Results from the maximum likelihood analysis of seed dormancy controlling for phylogeny in the legumes

	Intercept			Log_{10} (seed mass)			Seasonality			Interaction				
	Est.	SE	*P*	Est.	SE	*P*	Est.	SE	*P*	Est.	SE	*P*	α	AIC
FDF														
Mean	2.045	0.977	0.036	−0.403	0.278	0.147	0.822	0.320	0.010	−0.323	0.123	0.009	0.017	209.531
SD	0.007	0.002	0.001	0.002	0.001	0.001	0.004	0.001	0.000	0.002	0.000	< 0.0001	0.000	0.372
GDD														
Mean	12.748	42.744	0.003	−43.453	16.805	0.010	−0.001	0.001	0.013	0.001	0.000	0.026	0.013	215.725
SD	0.000	0.003	0.000	0.000	0.001	< 0.0001	0.000	0.000	0.000	0.000	0.000	< 0.0001	0.000	0.372
PC1														
Mean	6.526	1.713	0.000	−2.136	0.584	< 0.0001	−1.685	0.494	0.001	0.611	0.187	0.001	0.018	199.873
SD	0.302	0.243	0.000	0.163	0.061	< 0.0001	0.150	0.046	0.000	0.065	0.016	< 0.0001	0.006	0.372
Seed mass alone														
Mean	3.992	0.686	0.000	−0.980	0.212	< 0.0001	–	–	–	–	–	–	0.033	215.724
SD	0.011	0.009	0.000	0.003	0.001	< 0.0001	–	–	–	–	–	–	0.001	0.381

Four model types are presented in the table. The first four model the likelihood of seed dormancy across Fabaceae species as a function of the interaction between $\log_{10}$(seed mass) and four metrics of seasonal variation in environmental conditions (frost day frequency (FDF), growing degree days (GDD), and principal component 1 (PC1) and nonmetric multidimensional scaling 1 (NMDS1) of WorldClim environmental data. The fifth models seed dormancy as a function of seed mass alone. Each model was fitted across 650 bootstrap phylogenies to control for phylogenetic uncertainty; we therefore present, for each model type, both the mean estimate (first row) and SD (second row) for each explanatory variable's estimate (Est.), SE, and *P*-value (*P*) across the bootstraps. Finally, we also present the estimate of α, which measures the degree of phylogenetic correlation in the data (Tung Ho & Ané, 2014), and the fit of the models to the data estimated by the Akaike information criterion corrected for finite sample sizes (AIC). The results provide strong evidence for an interactive effect of seed mass and environment: in all cases the interactions between environmental conditions and seed size were significant. In keeping with our Bayesian modeling results (see Table 3), however, these results do suggest that seed mass alone also has strong explanatory power. The models containing seed mass alone have AIC values comparable to, but still greater than, those of the other models (indicating comparable model support), and their AIC values are are lower than those of models containing environmental data alone (see Supporting Information Table S2).

affected by phylogenetic nonindependence. The results obtained with maximum likelihood (ML) using a single (median) value of the seasonality estimates per species showed that seed dormancy is determined by the interaction between seed size and seasonality. These variables were highly significant in the full models (i.e. models incorporating both terms as predictors). Moreover, including seasonality and seed size improved the fit of the models to the data, as indicated by the lower AIC values (Table 2). The results provide strong evidence for an interactive effect of seed size and environment: in all cases the interactions between environmental conditions and seed size were significant. However, these results do suggest that seed size alone also has strong explanatory power. The models containing seed size alone have AIC values close to those of the other models (indicating comparable model support; Table S2). The Bayesian mixed models incorporating intraspecific variation in environmental conditions fitted with MCMCglmm were also congruent with our forecasts and broadly agreed with our ML results (Table 3). All models exhibited high explanatory power (conditional $R^2 > 0.8$), although in this case one of the seasonality estimates was nonsignificant when accounting for seed size (FDF; Table 3). Moreover, the full models including FDF and PC1 as well as seed size had a worse fit to the data than the univariate model including only seed size as indicated by the DIC. The other seasonality estimate, GDD, seemed to have more explanatory power and was significant both alone and in interaction with seed size. Nevertheless, the fit to the data of the model including both GDD and seed size was only marginally better than that of the seed size univariate model (Table 3). Taken together, these results appear to indicate that, although both environmental conditions and seed size are needed for a proper explanation of the distribution of seed dormancy, seed size is a better predictor of dormancy than seasonality.

Transitions between dormancy and nondormancy in the evolution of the Fabaceae

Inference of evolutionary transitions towards nondormancy supported the hypothesis that both the environment and seed size influence the evolution of nondormancy. Reconstructed estimates of seed size and GDD were consistent with a high likelihood of evolution towards nondormancy only in larger seeds exposed to longer favorable seasons. Moreover, no transition towards nondormancy was reconstructed with a seed size $< 1 \log_{(g)}$ or with < 4000 GDD (Fig. 2). This result did not seem to be affected by the evolutionary model considered or by uncertainty in the root state. All analyses gave the same results, regardless of whether trait evolution was modeled following a BM or OU process and independently of the root state (Fig. S4). This reinforces the results of our regression models, and shows that the evolution of seed dormancy is contingent on the characteristics of the environment but also on seed size.

Discussion

Our results support the working hypothesis that seed dormancy is adaptive in seasonal environments. However, whether dormancy or nondormancy is favored by selection depends ultimately on seed size. Under certain conditions (i.e. when survival in the soil is not a function of seed size), dormancy and size might constitute alternative strategies to adapt to seasonality. The phenological tuning provided by dormancy might be compensated for by the postgermination vigor of seedlings coming from larger seeds. However, our results also indicate an environmental limit to this trade-off: where the growing season is particularly short, dormancy is always favored regardless of seed size.

Table 3 Results from the logistic regressions controlling for phylogeny considering the variation in seasonality across species ranges

	Posterior mean	95% CI	R^2	95% CI	Effective samples	pMCMC	DIC
PC1 of Bioclim vars.	−0.926	−1.592; −0.419	0.746	0.486; 0.960	1723	< 0.0001	209.29
	−2.073	−3.452; −0.809	–	–	1456.9	< 0.0001	–
FDF	8.4×10^{-4}	3.0×10^{-4}; 1.3×10^{-3}	0.839	0.685; 0.965	2094	< 0.0001	220.58
	8.3×10^{-4}	-1.0×10^{-4}; 1.8×10^{-3}	–	–	2133	0.061	–
GDD	-9×10^{-3}	-0.021; -1.5×10^{-4}	0.832	0.587; 0.993	3638.2	0.035	194.52
	-1.5×10^{-3}	-2.8×10^{-3}; -2.3×10^{-3}	–	–	3748	0.01	–
Log(seed size)	−0.937	−1.554; −0.433	0.802	0.575; 0.968	1174	< 0.0001	191.05
PC1 × seed size	0.293	0.081; 0.526	0.871	0.729; 0.976	2139.5	0.005	191.16
FDF × seed size	-2.47×10^{-5}	-2.4×10^{-4}; 1.9×10^{-4}	0.880	0.734; 0.979	4109	0.83	194.70
GDD × seed size	2.49×10^{-4}	-1.4×10^{-5}; 5.3×10^{-4}	0.876	0.720; 0.981	5135	0.055	186.72

The table displays the coefficients of fixed factors of the Bayesian mixed-models fit with MCMCglmm incorporating intraspecific variation in climatic conditions and controlling for phylogenetic autocorrelation within the Fabaceae by including the consensus phylogeny as a random factor in the model. The variables used are seed size ($\log_{10}$ of seed mass, g), dormancy assignment (dormant/nondormant) based on Baskin & Baskin (2014) and environmental heterogeneity estimates obtained from principal component 1 (PC1) of the bioclimatic data, the average yearly frequency of frost days (FDF) and average yearly number of growth degree days (GDD). See text, Table 1 and Fig. 1 for details. The first row for each variable corresponds to the results of the univariate model including only that variable as an independent predictor, while the second row shows the results for the full model including each variable, seed size and their interaction. Seed size had a highly significant effect in every model. PC1: first axis of the PCA of the Bioclim variables considered to best approximate seasonality (see text for details). R^2, conditional R^2 computed based on the methods described by Nakagawa & Schielzeth (2013); DIC, deviance information criterion, an estimate of the goodness of fit of each model. A difference of more than five units in DIC can be seen as substantial (Spiegelhalter *et al.*, 2002).

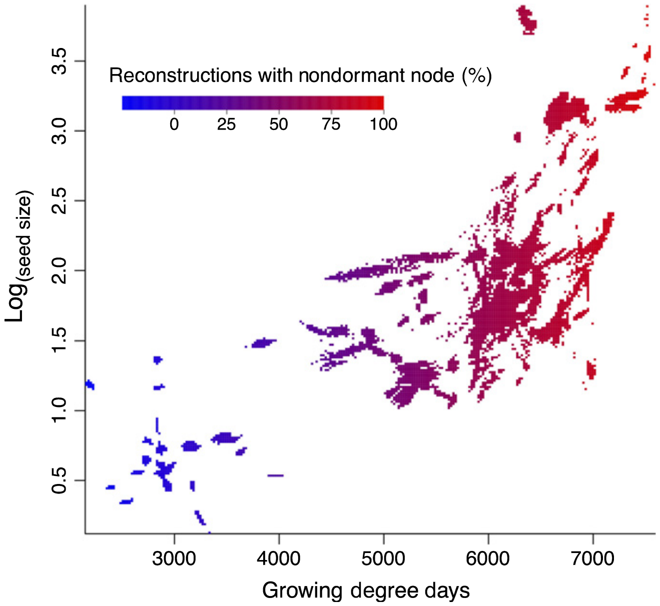

Fig. 2 Evolutionary transitions towards nondormancy across the Fabaceae. The plot is of the average reconstructed probability of transition towards nondormancy, growing degree-days and seed size across all 650 bootstrap phylogenies. The colors of the points represent the probability that a node is nondormant; the more red the point the more likely ($n = 650$ bootstraps × 198 nodes = 128 700 points). In all simulations, no nondormant node was reconstructed with a seed size < 1 $\log_{10}$(g) or with < 4000 growing degree days. No jitter has been added to the plot; the clusters of points represent clades that were aged differently in the bootstrapped phylogenies.

To date, the theoretical framework for the adaptive value of seed dormancy hinges mainly on its role as a bet-hedging strategy for annuals in unpredictable environments, such as deserts (e.g. Volis & Bohrer, 2013). Clearly, change in geometric fitness through bet-hedging might determine the micro- and macroevolution of any trait in general and of seed dormancy in particular (Lee & Doughty, 2003). At this scale of analysis, we cannot definitively tease apart the relative effect of unpredictable versus seasonal variation in environmental conditions in the evolution of seeds. Nevertheless, our results indicate, at least for the Fabaceae, that adaptation to seasonality is also a significant factor in the evolution of dormancy. The association between diapause eggs and predictable environmental fluctuations has been studied in some detail in insects (Cohen, 1970; Bradford & Roff, 1997), but to our knowledge the role of seed dormancy across broad latitudinal gradients in response to seasonal cycles had not been explicitly addressed before.

Analyses of a large data set of legumes ($n = 216\,000$ observations of 532 species) revealed a clear association between seed dormancy, seed size and seasonality. Dormancy was prevalent in temperate, seasonal environments, while nondormant seeds were concentrated in the tropics. Seed size also showed a similar latitudinal gradient, with larger seeded species concentrated in tropical regions. With the help of phylogenetic models, we tested whether this correlation was indicative of an evolutionary process linking environmental conditions and seed size and dormancy at a global scale. The results of these models showed that dormancy is associated with seasonality; the distribution of dormancy states depended on the interaction between seed size and seasonality; and seed size was a better predictor of dormancy than environmental conditions, such that small seeds were more likely to be dormant, although the models that best fitted the data incorporated both terms. In other words, both environmental variation/seasonality and seed size are critical to explaining the occurrence of dormancy and nondormancy and the evolutionary transitions between them.

Our simulations of the evolutionary shifts between dormant and nondormant seeds using the legume data, estimating the probability of a change in dormancy state (i.e. from dormant to nondormant or vice versa), suggest evolutionary hysteresis in these phylogenetic transitions. While the shift from dormancy to nondormancy was labile in lineages with larger seeds from aseasonal climates, nondormancy never evolved in small-seeded lineages, and particularly not in climates with shorter growing seasons. These results imply that nondormancy can only evolve in climates with long growing seasons and/or in taxa that produce larger seeds. Conversely, dormancy should be evolutionarily stable in temperate lineages with small seeds.

As an additional step, we formalized the association among seed dormancy and size and seasonal variation in environmental favorability in the model presented in Methods S1. In general, larger seeds were less dependent on dormancy, in agreement with our empirical results and previous theoretical work on the trade-off between seed dormancy and size (e.g. Venable & Brown, 1988). Based on our model results, however, there is a limit to this trade-off that is governed by the seasonality of the environment. Specifically, the importance of seed size diminished as the favorable season shortened; in environments where the growing season was particularly short, dormancy was always adaptive, the same pattern that was recovered by our simulations of the evolution of shifts in the dormancy state.

Despite the congruence of the theoretical and empirical results in support of our hypotheses, there are limitations in our data and analytical tools that will be important areas for future studies to address. For example, our analyses are limited by incomplete sampling of phylogenetic and environmental diversity as well as by factors affecting seed survival in the soil other than seed size, such as physical or chemical defensive traits. Seed defenses might alter the association between dormancy state and environment, by facilitating post-dispersal survival of larger seeds in the seedbank (Schutte *et al.*, 2014). Although we base our conclusions on the strong association between dormancy and seed size found in a family known for its primary and secondary seed defenses, detailed examination of the role of seed defensive properties would be an important avenue for future research. Also, our data set was particularly enriched in a single class of dormancy (physical) and our models did not fully account for intraspecific variation in this trait, seed size or genetic structure. However, variation within species in seed dormancy and size is predicted to follow a similar geographic trend as that observed across species while encompassing lower variation (Moles & Westoby, 2003; Montesinos-Navarro *et al.*, 2012). These and other shortcomings might limit the applicability of our results in certain cases. However, our conclusions were validated by a three-pronged approach: extensive analyses of empirical data, computational models of trait evolution and a theoretical model. As such, we feel that they provide meaningful insights into the biogeographic and macroevolutionary patterns of dormancy and seed size.

In view of our results, the global distribution of dormant seeds seems to follow reliable patterns that establish the likelihood of seed dormancy in any given evolutionary lineage or ecological community. Considering that dormancy, by definition, only limits germination under favorable conditions, and seeds of any species can be present in their local seedbank if conditions are unfavorable for their germination, we can posit explicit predictions across ecological and phylogenetic scales (all else being equal).

1 Within lineages, taxa occurring in more seasonal environments should have smaller, more dormant seeds.

2 Within communities, smaller seeded species should dominate the soil seedbank. By contrast, larger seeded species ensure the presence of their propagules locally across seasons by an ‘aerial’ bank, that is, by longer reproductive lifespans and/or delayed release from the maternal plant.

To conclude, we propose that a general pattern should be observable at a global scale: aseasonal environments are expected to select for and be dominated by taxa producing larger, nondormant seeds, while smaller, dormant seeds should be adaptive and predominate in seasonal habitats. Future research addressing these hypotheses will clarify the role of seed size and environmental conditions in the evolution of dormancy.

Acknowledgements

This paper is the product of the working group ‘Germination, trait coevolution and niche limits in changing environments’ organized by Kathleen Donohue and R.R.d.C. and sponsored by the National Evolutionary Synthesis Center, National Science Foundation EF-0905606. R.R.d.C. was additionally funded by the European Commission (MC-IIF-2011-300026 ‘TEE-OFF’) and the Talentia program (Junta de Andalucia/EC – FP7 CFUND; grant agreement no. 267226, ‘Bet-hedging, trade-offs and the evolution of seed dispersal and dormancy’). J.C-B. would like to thank the EEB Theory Group at the University of Minnesota, particularly Tanjona Ramiadantsoa, Allison Shaw and Xiaojing Wei, for feedback on the mathematical model.

Author contributions

C.G.W., J.C-B. and R.R.d.C. designed the study. C.C.B., C.G.W., J.M.B. and R.R.d.C. compiled and curated the data. W.D.P. designed the trait evolution model. J.C-B. developed the mathematical model. C.G.W., R.R.d.C. and W.D.P. analyzed the data. R.R.d.C. and J.C-B. wrote the first draft of the manuscript. All authors contributed revisions.

References

Baskin JM, Baskin CC. 2004. A classification system for seed dormancy. *Seed Science Research* 14: 1–16.

Baskin CC, Baskin JM. 2014. *Seeds: ecology, biogeography, and evolution of dormancy and germination.* San Diego, CA, USA: Elsevier/Academic Press.

Beaulieu JM, Jhwueng D-C, Boettiger C, O’Meara BC. 2012. Modeling stabilizing selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution. *Evolution* 66: 2369–2383.

Bell CD, Soltis DE, Soltis PS. 2010. The age and diversification of the angiosperms re-revisited. *American Journal of Botany* 97: 1296–1303.

Bollback JP. 2006. SIMMAP: stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics* 7: 88.

Bradford MJ, Roff DA. 1997. An empirical model of diapause strategies of the cricket *Allonemobius socius*. *Ecology* 78: 442–451.

Brown JS, Venable DL. 1991. Life history evolution of seed-bank annuals in response to seed predation. *Evolutionary Ecology* 5: 12–29.

Brunner I, Herzog C, Dawes MA, Arend M, Sperisen C. 2015. How tree roots respond to drought. *Frontiers in Plant Science* 6: 547.

Butler MA, King AA. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *American Naturalist* 164: 683–695.

Childs DZ, Metcalf CJE, Rees M. 2010. Evolutionary bet-hedging in the real world: empirical evidence and challenges revealed by plants. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 277: 3055–3064.

Cohen D. 1970. A theoretical model for the optimal timing of diapause. *American Naturalist* 104: 389–400.

Dalling JW, Davis AS, Schutte BJ, Arnold AE. 2011. Seed survival in soil: interacting effects of predation, dormancy and the soil microbial community. *Journal of Ecology* 99: 89–95.

Donohue K, Rubio de Casas R, Burghardt L, Kovach K, Willis CG. 2010. Germination, post-germination adaptation, and species ecological ranges. *Annual Review of Ecology, Evolution, and Systematics* 41: 293–319.

Ellner S. 1987. Competition and dormancy: a reanalysis and review. *American Naturalist* 130: 798–803.

Finch-Savage W, Leubner-Metzger G. 2006. Seed dormancy and the control of germination. *New Phytologist* 171: 501–523.

Gelman A, Jakulin A, Pittau MG, Su Y-S. 2008. A weakly informative default prior distribution for logistic and other regression models. *Annals of Applied Statistics* 2: 1360–1383.

Gómez JM, Husband B. 2004. Bigger is not always better: conflicting selective pressures on seed size in *Quercus ilex*. *Evolution* 58: 71–80.

Hadfield JD. 2010. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software* 33: 1–22.

Huelsenbeck JP, Nielsen R, Bollback JP. 2003. Stochastic mapping of morphological characters. *Systematic Biology* 52: 131–158.

Janzen DH. 1971. Seed predation by animals. *Annual Review of Ecology Evolution and Systematics* 2: 465–492.

Judd WS, Campbell CS, Kellogg E, Stevens PF, Donoghue MJ. 2016. *Plant systematics: a phylogenetic approach.* Sunderland, MA, USA: Sinauer Associates.

Kikuzawa K. 1996. Geographical distribution of leaf life span and species diversity of trees simulated by a leaf-longevity model. *Vegetatio* 122: 61–67.

Kitajima K, Fenner M. 2000. Ecology of seedling regeneration. In: Fenner M, ed. *Seeds: ecology of regeneration in plant communities.* Wallingford, UK: CABI Publishing, 331–360.

Lavin M, Herendeen PS, Wojciechowski MF. 2005. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. *Systematic Biology* 54: 575–594.

Lee MSY, Doughty P. 2003. The geometric meaning of macroevolution. *Trends in Ecology & Evolution* 18: 263–266.

Lewis G, Schrire B, MacKinder B, Lock M (Eds.). 2005. *Legumes of the World.* Kew, UK: Royal Botanic Gardens, Kew.

Long RL, Gorecki MJ, Renton M, Scott JK, Colville L, Goggin DE, Commander LE, Westcott DA, Cherry H, Finch-Savage WE. 2015. The ecophysiology of seed persistence: a mechanistic view of the journey to germination or demise. *Biological Reviews* 90: 31–59.

Metz J, Liancourt P, Kigel J, Harel D, Sternberg M, Tielbörger K. 2010. Plant survival in relation to seed size along environmental gradients: a long-term study from semi-arid and Mediterranean annual plant communities. *Journal of Ecology* **98**: 697–704.

Mitchell TD, Jones PD. 2005. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *International Journal of Climatology* **25**: 693–712.

Moles AT, Ackerly DD, Tweddle JC, Dickie JB, Smith R, Leishman MR, Mayfield MM, Pitman A, Wood JT, Westoby M. 2007. Global patterns in seed size. *Global Ecology and Biogeography* **16**: 109–116.

Moles AT, Ackerly DD, Webb CO, Tweddle JC, Dickie JB, Pitman AJ, Westoby M. 2005. Factors that shape seed mass evolution. *Proceedings of the National Academy of Sciences, USA* **102**: 10540–10544.

Moles AT, Westoby M. 2003. Latitude, seed predation and seed mass. *Journal of Biogeography* **30**: 105–128.

Montesinos-Navarro A, Picó FX, Tonsor SJ. 2012. Clinal variation in seed traits influencing life cycle timing in *Arabidopsis thaliana*. *Evolution* **66**: 3417–3431.

Morin X, Chuine I. 2006. Niche breadth, competitive strength and range size of tree species: a trade-off based framework to understand species distribution. *Ecology Letters* **9**: 185–195.

Nakagawa S, Schielzeth H. 2013. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* **4**: 133–142.

New M, Hulme M, Jones P. 1999. Representing twentieth-century space–time climate variability. Part I: development of a 1961–90 mean monthly terrestrial climatology. *Journal of Climate* **12**: 829–856.

Nikolaeva MG. 1967. *Physiology of deep dormancy in seeds*. Leningrad, Russia: Izdatel'stvo Nauka.

Pake CE, Venable DL. 1996. Seed banks in desert annuals: implications for persistence and coexistence in variable environments. *Ecology* **77**: 1427–1435.

Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**: 289–290.

Plummer M, Best N, Cowles K, Vines K. 2006. CODA: convergence diagnosis and output analysis for MCMC. *R News* **6**: 7–11.

R Development Core Team. 2009. *R: a language and environment for statistical computing. Version 3.3.2*. Vienna, Austria: R Foundation for Statistical Computing.

Rees M. 1996. Evolutionary ecology of seed dormancy and seed size. *Philosophical Transactions of the Royal Society of London, Series B Biological Sciences* **351**: 1299–1308.

Reich PB, Walters MB, Ellsworth DS. 1997. From tropics to tundra: global convergence in plant functioning. *Proceedings of the National Academy of Sciences, USA* **94**: 13730–13734.

Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* **3**: 217–223.

Royal Botanic Gardens Kew. 2012. *Seed information database (SID) version 7.1*. [WWW document] URL http://data.kew.org/sid/ [accessed April 2012].

Rubio de Casas R, Donohue K, Venable DL, Cheptou P-O. 2015. Gene-flow through space and time: dispersal, dormancy and adaptation to changing environments. *Evolutionary Ecology* **29**: 813–831.

Schutte BJ, Davis AS, Peinado SAJ, Ashigh J. 2014. Seed-coat thickness data clarify seed size–seed-bank persistence trade-offs in *Abutilon theophrasti* (Malvaceae). *Seed Science Research* **24**: 119–131.

Smith SA, O'Meara BC. 2012. treePL: Divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* **28**: 2689–2690.

Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**: 483–689.

Stamatakis A, Hoover P, Rougemont J, Renner S. 2008. A rapid bootstrap algorithm for the RAxML web servers. *Systematic Biology* **57**: 758–771.

Stevens PF. 2001. *Angiosperm Phylogeny Website. Version 12*. July 2012 (and more or less updated since). [WWW document] URL http://www.mobot.org/MOBOT/research/APweb/ [accessed December 2012].

The Legume Phylogeny Working Group. 2013. Legume phylogeny and classification in the 21st century: progress, prospects and lessons for other species-rich clades. *Taxon* **62**: 217–248.

Tung Ho L si, Ané C. 2014. A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Systematic Biology* **63**: 397–408.

Venable DL, Brown JS. 1988. The selective interactions of dispersal, dormancy, and seed size as adaptations for reducing risk in variable environments. *American Naturalist* **131**: 360–384.

Venable DL, Rees M. 2009. The scaling of seed size. *Journal of Ecology* **97**: 27–31.

Volis S, Bohrer G. 2013. Joint evolution of seed traits along an aridity gradient: seed size and dormancy are not two substitutable evolutionary traits in temporally heterogeneous environment. *New Phytologist* **197**: 655–667.

Willis CG, Baskin CC, Baskin JM, Auld JR, Venable DL, Cavender-Bares J, Donohue K, Rubio de Casas R, The NESCent Germination Working Group. 2014. The evolution of seed dormancy: environmental cues, evolutionary hubs, and diversification of the seed plants. *New Phytologist* **203**: 300–309.

Wink M, Mohamed GIA. 2003. Evolution of chemical defense traits in the Leguminosae: mapping of distribution patterns of secondary metabolites on a molecular phylogeny inferred from nucleotide sequences of the *rbcL* gene. *Biochemical Systematics and Ecology* **31**: 897–917.

Yahara T, Javadi F, Onoda Y, de Queiroz LP, Faith DP, Prado DE, Akasaka M, Kadoya T, Ishihama F, Davies S *et al.* 2013. Global legume diversity assessment: concepts, key indicators, and strategies. *Taxon* **62**: 249–266.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information tab for this article:

Fig. S1 Graphical representation of the predictions of the mathematical model.

Fig. S2 Results of the NMDS of bioclimatic seasonality variables.

Fig. S3 Phylogenetic distribution of seed dormancy, seed size and seasonality in the Fabaceae.

Fig. S4 Shifts towards nondormancy in the legumes considering alternative evolutionary hypotheses.

Table S1 Parameter values used in the mathematical model

Table S2 Results of the univariate PHYLOLM models describing the likelihood of seed dormancy as a function of environmental conditions

Methods S1 Simple mathematical model relating dormancy to season length.