Evolutionary pathways to lower biomass allocation to the seed coat in crops: insights from allometric scaling

Rubén Milla, Adrianus J Westgeest, Jorge Maestre-Villanueva, Sergio Núñez-Castillo, Alicia Gómez-Fernández, François Vasseur, Cyrille Violle, Jana Balarynová, Petr Smykal

HAL Id: hal-04766013
https://hal.inrae.fr/hal-04766013v1

Submitted on 4 Nov 2024

Evolutionary pathways to lower biomass allocation to the seed coat in crops: insights from allometric scaling

Rubén Milla[1] (iD), Adrianus J. Westgeest[2] (iD), Jorge Maestre-Villanueva[1], Sergio Núñez-Castillo[1], Alicia Gómez-Fernández[1], François Vasseur[3], Cyrille Violle[3], Jana Balarynová[4] and Petr Smykal[4] (iD)

[1]ECOEVO Group, Departamento de Biología, Geología, Física y Química Inorgánica, Instituto de Investigación en Cambio Global (IICG-URJC), Universidad Rey Juan Carlos, Tulipán s/n, Móstoles, 28933, Spain; [2]Département Biologie et Ecologie, Institut Agro, Montpellier, 34060, France; [3]CEFE, Univ Montpellier, CNRS, EPHE-PSL University, IRD, Montpellier, 34090, France; [4]Department of Botany, Faculty of Science, Palacky University, Olomouc, CZ-783 71, Czech Republic

Summary

- Crops generally have seeds larger than their wild progenitors´ and with reduced dormancy. In wild plants, seed mass and allocation to the seed coat (a proxy for physical dormancy) scale allometrically so that larger seeds tend to allocate less to the coats. Larger seeds and light-weight coats might thus have evolved as correlated traits in crops.
- We tested whether 34 crops and 22 of their wild progenitors fit the allometry described in the literature, which would indicate co-selection of both traits during crop evolution. Deviations from the allometry would suggest that other evolutionary processes contribute to explain the emergence of larger, lightweight-coated seeds in crops.
- Crops fitted the scaling slope but deviated from its intercept in a consistent way: Seed coats of crops were lighter than expected by their seed size. The wild progenitors of crops displayed the same trend, indicating that deviations cannot be solely attributed to artificial selection during or after domestication.
- The evolution of seeds with small coats in crops likely resulted from a combination of various pressures, including the selection of wild progenitors with coats smaller than other wild plants, further decreases during early evolution under cultivation, and indirect selection due to the seed coat-seed size allometry.

Introduction

Most agricultural crops have larger seeds with thinner seed coats than their wild progenitors (Harlan *et al.*, 1973; Larson *et al.*, 2014). Larger seeds increase yields, have better nutritional quality, and are better adapted to seed burial during sowing (Sadras, 2007; Purugganan & Fuller, 2009; Garibaldi *et al.*, 2021). Thin coats reduce physical dormancy, allow predictable germination and establishment of crop stands, and easiness of cooking, which are advantages for the farmer and consumer (Smýkal *et al.*, 2014; Fuller & Allaby, 2018). The timing and rates of emergence, and selection pressures on seed size and coats in crops have been discussed profusely (Kluyver *et al.*, 2013, 2017; Fuller *et al.*, 2023). However, both traits are rarely investigated together in the domestication literature, that is increased size and loss of physical dormancy are generally assumed to evolve independently in crop seeds. But seed and coat sizes evolve co-ordinately in wild species (Venable & Brown, 1988; Casas *et al.*, 2017; Chen *et al.*, 2020). This suggests that larger and non-dormant seeds might have been co-selected in crops, without need to invoke different selection pressures for each trait separately.

Physical dormancy in seed plants depends partly on seed size. Seed size and dormancy can be regarded as different adaptations to similar selection pressures, such as seasonality, environmental heterogeneity, and crowding (Venable & Brown, 1988; Casas *et al.*, 2017). Indeed, theoretical models predict covariation of both traits (Venable & Brown, 1988). This concurs with empirical and comparative data. For example, evolutionary divergences in dormancy tend to associate with divergences in seed size (Rees, 1997). Small seeds typically persist for longer in the soil, which is facilitated by physical dormancy (Eriksson & Eriksson, 1997; Bekker *et al.*, 1998), and invest more in physical defence (Fricke & Wright, 2016). In legumes, physical dormancy is less prevalent among larger seeded species (Casas *et al.*, 2017). At the developmental level, the rate of growth of integument cells – the maternal tissue of the flower that will mature into the seed testa – influences seed size (García *et al.*, 2005). Interestingly, larger seeds tend to have less biomass allocation to their coats than expected given their size – that is coat and kernel weights scale allometrically with a slope < 1 (Wu *et al.*, 2019; Chen *et al.*, 2020). This suggests that pleiotropy or genetic linkage may pose constraints to the free variation of kernel and coat sizes (He *et al.*, 2023).

The allometric scaling between coat and kernel sizes might be relevant to the evolution of nondormancy in crops. Given the allometry, it is plausible that selection for large seeds triggered the evolution of reduced investment in the seed coat and thus decreased physical dormancy as a correlated trait. In the archaeological series that report data for both traits, they tend to evolve in parallel. As seeds get larger, seed coats (an archaeological proxy for physical dormancy) get thinner (Smith, 1989; Bruno & Whitehead, 2003; Murphy & Fuller, 2017). This is circumstantial evidence that both traits might have coevolved during domestication. Improved germinability of crops might thus be a correlate of selection for large seeds, rather than a trait directly targeted by ancient farmers. However, it is not known whether crops fit to the allometric evolution of kernel vs coat sizes reported for wild plants. If they fit, a parsimonious explanation for the evolution of reduced dormancy is its correlated evolution with seed size. If crops deviate from the allometry, then it is more plausible that nondormancy and larger seeds might have evolved partially independently in crops, and that different pressures might have selected for each trait separately or at different times.

Here, we explored whether a large set of 34 herbaceous crops fit the kernel-coat biomass allometric scaling observed in wild plants (as described for 940 wild species in Chen *et al.*, 2020). To further investigate how and when lightweight coats emerged in crops, we studied whether the wild progenitors of crops fit the allometry and how their seeds compare with those of early landraces and with recent cultivars. If crop wild progenitors had coats lighter than expected by the allometry, it would indicate that the choice of certain wild species for domestication, instead of others, facilitated the evolution of lightweight seed coats in crops. Comparing wild progenitors with early landraces, and landraces with recent cultivars, will inform on whether early domestication and recent breeding, respectively, had a role in the development of lightweight seed coats in crops. Our questions are as follows: (1) Do the seeds of crops and their wild progenitors fit the coat-kernel biomass allometry observed across wild plants? and (2) How strongly early selection of wild progenitors, further domestication, and recent breeding affected biomass allocation between seed coats and kernels in current crops?

Materials and Methods

Collection of data

To investigate the evolution of seed kernel vs seed coat biomass distribution in crops, in comparison with crop wild progenitors and other wild plants, we put together a database with 1954 individual records of kernel and seed coat data. This database included 1079 observations on crops and their wild progenitors (for 34 species of domesticated crops and 22 species of crop wild progenitors) and 875 observations on other species of wild plants.

To build our database, we used a published compilation, searched for additional data in the primary literature, and measured seeds in the laboratory to generate new unpublished data. Specifically, for wild species we used the Chen *et al.* (2020)

database, which holds 1120 records of coat and kernel mass of 940 wild species. Records in Chen *et al.* (2020) were collapsed per species. The vast majority of these records corresponded unambiguously to wild species. However, those records in Chen *et al.* (2020), which botanical name might indicate an unclear domestication status (i.e. botanical names that matched cultivar names or the names of recognized wild progenitors of crops as listed in Milla (2020a)) or overlapped with our records for crops and their wild progenitors, were removed from that source. The Chen *et al.'s* (2020) database thus filtered included data of 875 wild species.

As regards crops, we gathered data for a total of 34 crop species, combining literature data (*c.* 8% of the raw data) and laboratory measurements (*c.* 92% of the raw data) as follows. Data for domesticated accessions of bread wheat (*Triticum aestivum*), black seed (*Nigella sativa*), pigeon pea (*Cajanus cajan*), cultivated species of the genus *Vigna* other than *V. unguiculata* (i.e. *V. angularis, V. radiata,* and *V. subterranea*), lima bean (*Phaseolus lunatus*), narrowleaf lupin (*Lupinus angustifolius*), white lupin (*Lupinus albus*), rapeseed (*Brassica napus* and *B. rapa*), vetch (*Vicia* sp.), and winged bean (*Psophocarpus tetragonolobus*) were retrieved from the primary literature (sources available at doi: 10.5281/zenodo.10973414). Fifteen other crops were investigated more in depth, so we measured in the laboratory seeds of at least two contrasting accessions of their wild progenitors, two accessions of landraces, and two accessions of modern varieties. Landraces are traditional domesticated genotypes that have not undergone modern scientific breeding, while modern varieties are the result of recent breeding programmes, often involving scientific methods and technologies. These crops were sunflower (*Helianthus annuus*), soybean (*Glycine max*), sorghum (*Sorghum bicolor*), peas (*Pisum sativum*), oats (*Avena sativa*), maize (*Zea mays*), linseed (*Linum usitatissimum*), lentils (*Lens culinaris*), fava bean (Vicia faba), cowpea (*V. unguiculata*), common bean (*Phaseolus vulgaris*), chickpea (Cicer arietinum), black medick (Medicago lupulina), barley (*Hordeum vulgare*), and amaranth (*Amaranthus cruentus*). For an additional set of six crops, we were only able to obtain seeds of wild progenitor and landrace accessions – sesame (*Sesamum indicum*), rye (*Secale cereale*), grass pea (*Lathyrus sativus*), and borage (*Borago officinalis*) – or of wild progenitor and modern cultivar accessions – peanuts (*Arachis hypogaea*), and narrowleaf lupin (*Lupinus angustifolius*). The botanical identity of the most likely wild progenitor of each crop was taken from the Crop Origins database (Milla, 2020a,b). The specific number of accessions per crop, accession identifiers, and botanical identities of the accessions are available at doi: 10.5281/zenodo.10973414.

For each of the accessions measured, we dissected 5 to 10 seeds. Each seed was separated into kernel and coat, and then, both fractions were oven-dried and weighed separately to the nearest 0.001 mg. of dry mass. To achieve comparability among sources and species, we adopted a comprehensive interpretation of the seed coat as the rigid, outer layer safeguarding the reserves and embryo within, regardless of the specific anatomical tissues contributing to the coat, following Moles (2005) and Chen *et al.* (2020). This implies that not only the testa *sensu stricto* –

that is the seed tissues deriving from the maternal integuments – are included in the coat, but also any other tissues intimately adhered to the dispersed seed – for example pericarp and adhering paleas and lemmas in cereal caryopses. Laboratory measurements were conducted between November 2022 and June 2023.

Phylogeny of the study species

To carry out the comparative analyses described in the Data analyses of the Material and Methods section, we developed a phylogenetic tree for the 1942 species in the database. Botanical names in the database were harmonized following the World Flora Online using the functions WFO.match and WFO.one of the WORLDFLORA R package (Kindt, 2020). To develop the phylogeny, we employed the 'GBOTB.extended.WP.tre' (https://github.com/megatrees/plant_20221117/tree/main) megatree as a baseline, as described in Jin & Qian (2022). This megatree is a time-calibrated molecular phylogeny encompassing 74 531 plant species. It is an enhanced version of the previously established GBOTB.extended phylogeny (Smith & Brown, 2018). We pruned 'GBOTB.extended.WP.tre' to include only the taxa from our database using the phylo.maker function from the V.PHYLOMAKER R package (Jin & Qian, 2019), with the specification scenario = 3. Scenario 3 places the tip for a new genus between the family root node and basal node. This phylogeny did not include the wild progenitors of crops. Since many wild progenitors share binomial with their corresponding cultivated derivatives, we took the wild progenitors into the phylogeny a posteriori as follows. First, we retrieved the antiquity of cultivation for each crop, as the best possible surrogate of divergence time between the wild progenitor and the cultivated varieties, from the Crop Origins database (Milla, 2020a,b). Then, we inserted the wild progenitors as sister edges of their crops, at a height so that the node separating the wild progenitor from its crop is the antiquity of cultivation of the crop. Insertions were done using the function bind.tip from the PHYTOOLS package in R (Revell, 2012). This tree had 931 tips and 811 internal nodes, with 56 unresolved nodes, thus achieving a high level of phylogenetic resolution, at 93.1%. Since the analyses below are robust to soft polytomies (Grafen & Hamilton, 1989), the phylogeny was used as such. The phylogeny was pruned to the taxa involved in each of the analyses described below using the function drop.tip from the APE package in R (Paradis et al., 2004).

Data analyses

Allometric analyses Seed mass, kernel mass, and coat mass data were collapsed to species arithmetic means for allometric analyses. Crop species and their direct wild progenitors were considered as different species, even if they shared binomials. Allometric analyses of interspecific data can be carried out using a diversity of line-fitting approaches (Shingleton, 2019). In this project, we opted to use phylogenetically explicit phylogenetic generalized least squares (PGLS) models, following advice from the recent literature on allometry and the comparative method (Hansen & Bartoszek, 2012; Kilmer & Rodríguez, 2017; Shingleton, 2019). Our PGLS models had seed coat mass as the

response variable, kernel mass as the fixed effects predictor, and the phylogenetic tree of the species involved as the covariance matrix of the generalized squares regression, weighed for the phylogenetic signal of the residuals of the regression (Symonds & Blomberg, 2014). We run three different PGLSs, one for the set of wild species, another for the set of crop species, and another for the set of crop wild progenitors. To test whether the PGLS slopes were isometric (slope = 1), we calculated a test statistic for the PGLS slope against 1, and computed the probability of this statistic to follow a t-distribution under the null hypothesis that the slope is 1. Similarly, to test whether the crops and wild progenitors follow the same scaling slope than wild species, we proceeded the same way, but using the slope of the wild species as reference value for the test. PGLSs were implemented with the pgls function of the CAPER R package (Orme, 2018). Deviations from the allometry were computed as the raw residuals in the y-axis of each species from the baseline fit line (i.e. from the fit line of the 875 wild species in the compilation). We chose this metric of deviation, instead of an orthogonal deviation from the fit line (Chen et al., 2020), because the questions at hand were on the modifications of crop seed coats, not on changes on all components of seed size taken together. The former are best measured by a deviation in the y-axis, while the later would be better measured using an orthogonal deviation.

To test whether crops and their wild progenitors deviate from the allometry more than other wild plants, we ran an additional PGLS analysis with the residuals of each species as the response variable and their domestication status (i.e. crops vs wild progenitors vs other wild species) as a fixed effects factor. Differences among the three categories of species were tested using the phylANOVA function of the PHYTOOLS R package (Revell, 2012).

Analyses of the effects of evolution under domestication To investigate the effects of ancient domestication and recent breeding on crop seeds, we used the subset of crops for which we had data of at least two accessions of both wild progenitors and landraces (to investigate domestication effects) or of landraces and modern cultivars (to investigate the effects of recent breeding). To compare allocation to seed coats among the different domestication statuses, the seed-to-coat mass ratio (dry mass of coat/dry mass of whole seed; seed coat ratio (SCR), hereon) was used, as it is the commonest metric of relative allocation of seed biomass to the coat (Daws et al., 2006). To assess the magnitude and direction of the effects of domestication and recent breeding on SCR for each crop species separately, we used the Hedges' G metric for effect sizes (Hedges et al., 1999). For the effects of domestication, Hedges' G was calculated as the difference in means between landraces and wild progenitors of each crop divided by the pooled SD of the two groups and multiplied by a correction factor for sample size (Hedges et al., 1999). The effects of recent breeding on SCR were computed in the same way but taking the difference between modern cultivars and landraces. Hegde's G effect sizes and their 95% confidence intervals (CI) were calculated using the cohen.d function of the EFFSIZE R package (Torchiano, 2020).

To investigate the overall effects of domestication and recent breeding on SCR across crop species we performed a linear
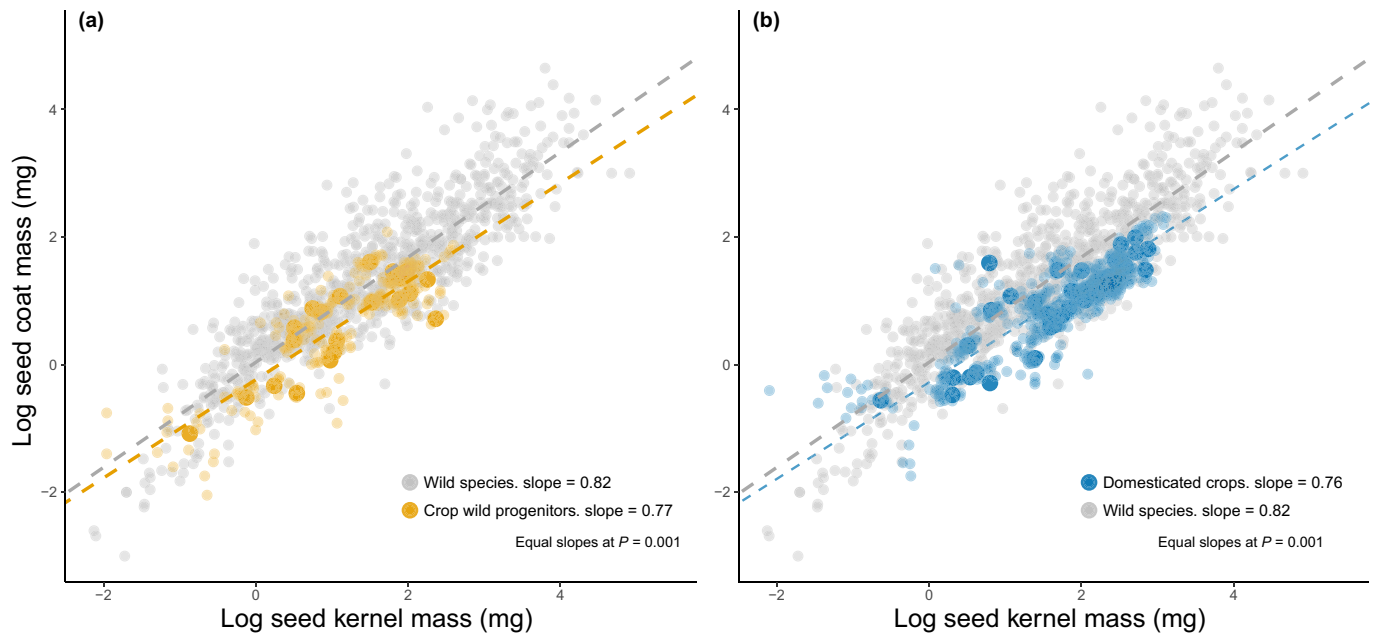
Fig. 1 Kernel to coat mass scaling in wild plants, and in crops and their wild progenitors. Scaling of kernel (embryo and reserve tissues) and seed coat (outer layer safeguarding the reserves and embryo) mass in wild plants (grey, $n = 875$), wild progenitors of crops (yellow, $n = 22$) (a), and crops (blue, $n = 34$) (b). The three fit lines are the result of fitting phylogenetic generalized least squares (PGLS) models to each of the three types of plants separately. The three scaling slopes were allometric (differed from slope = 1 at $P = 0.001$) but did not differ significantly among each other (equal slopes at $P = 0.001$). Different records of a given species were collapsed to species arithmetic means for calculating the allometries. Small semi-transparent yellow and blue points in the background are individual measurements of crop and wild progenitor seeds, while large points are their species arithmetic means.

mixed-effect model (LMMs), using the *lme* function in the NLME R package (Pinheiro *et al.*, 2020). The model included SCR as a response variable, and domestication status (wild progenitor vs landrace vs modern cultivar) as fixed effects predictor. Accession identity, nested within crop species, was included as a random effects factor over the intercept. The residuals of the model were checked for normality, and for homoscedasticity across domestication statuses. Statistical significance of the fixed factor was estimated by using the *anova.lme* function with marginal sums of squares in the NLME R package (Pinheiro *et al.*, 2020). The amount of variance explained by the models was evaluated using the *r.squaredGLMM* function from the MUMIN R package (Barton, 2015). Multiple comparisons tests among domestication statuses were applied with false discovery rate control and using the *glht* function in the MULTCOMP R package (Hothorn, 2016).

Visualization of data and results was developed using GGPLOT2, GGBREAK, and PATCHWORK R packages (Wickham, 2011; Pedersen, 2019; Xu *et al.*, 2021). We used CHATGPT v.3.5 (OpenAI, 2023) to streamline and refine drafts of R code.

Results

Variation in seed size, and kernel and coat sizes among the study species

The species included in our compilation encompassed a wide range of variation in their seed sizes, and in the allocation of biomass to seed kernels and coats. Seed sizes (mg) spanned seven

orders of magnitude across the 931 species included in our analyses, from 0.01 mg of the smallest-seeded species, *Clidemia octona* (Melastomataceae), to the 82 000 mg of the largest-seeded *Idiospermum australiense* (Calycanthaceae). This excludes the tiniest seeds on earth, which can weigh down to *c.* 0.001 mg, but otherwise includes seeds of most sizes, from very small to the largest ones (Moles, 2005). Crops and their wild progenitors also showed a wide variety of seed sizes, spanning *c.* 25% of the variation found among the other wild species of the compilation, despite being much fewer species – 56 species of crops plus wild progenitors vs 875 species of other wild species (Fig. 1). For most of the crops, wild progenitors had smaller seeds than their domesticated counterparts (Supporting Information Fig. S1). The proportional allocation of biomass to coats was also highly variable among the 931 species, ranging from *Eremophila gilesii* (Scrophulariaceae), which allocated 98% of the seed biomass to the coat, to *Couma catingae* (Apocynaceae), which allocated only 0.01% to the coat. The sizes of whole seeds, kernels, and coats of crops and their wild progenitors were within the boundaries of variation found across other wild plants and did not expand further from the phenotypic space set by wild species (Fig. 1).

Seed kernel to coat allometry

Log-transformed seed kernel and coat dry mass scaled with a slope of 0.82 for the wild species of our database (crop wild progenitors excluded). This was statistically significantly different from 0 and significantly lower than 1 (Fig. 1). This indicates an
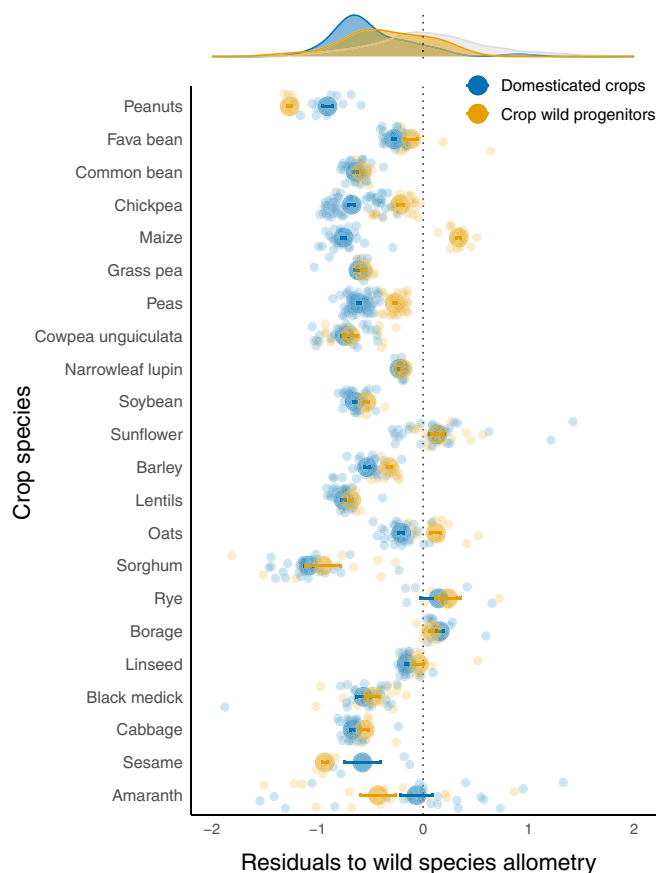
Fig. 2 Deviations from the allometry in crops and in their wild progenitors. Crop-wise deviations (i.e. raw residuals) from the phylogenetic generalized least squares (PGLS) allometry of wild species in Fig. 1, shown separately for wild progenitors and for domesticated accessions of each crop. Small points are raw data, large points are arithmetic means, and horizontal bars are ±1 SEM (SE of the mean). Crops are ordered in the y-axis by increasing seed mass. In the upper margin of the plot, density plots of all deviations are shown, with all domesticated crops together (blue), all crop wild progenitors (yellow), and all other wild species (light grey). According to the results of phylogenetic ANOVA analyses, deviations of the crops were significantly larger (P < 0.05) than those of wild species, whereas there was a statistical trend (P < 0.1) that deviations of wild progenitors tended to be larger than those of other wild species.

allometric scaling of seed kernels and coats, with large seeds investing proportionally less biomass in their coats. Crops and their wild progenitors also showed allometric scaling, with a scaling exponent not significantly different from that of the other wild species (Fig. 1). Deviations – *that is* residuals in the y-axis – from the 0.82 slope were ample among all species in the dataset (Figs 1, 2). Deviations of the crops were significantly (P < 0.05) larger than those of wild species, whereas there was a statistical trend (P < 0.1) for the deviations of wild progenitors to be larger from other wild species (Fig. 2). Most species of crops and their wild progenitors showed deviations consistently skewed towards negative scores (Figs 1, 2). This means that crops´ and wild progenitors´ seeds tended to have a lower allocation to the seed coat than expected for their seed size. Negative deviations were generally of higher magnitude for domesticated accessions of each crop

than for their wild progenitors (Fig. 2). Crops such as peanuts, sorghum, black medick, or lentils had seed coats extremely light for their seed size, whereas the seed coats of other crops (e.g. amaranth, linseed, or oats) had coats only slightly lighter than expected for their seed size (Fig. 2). A few domesticated crops (sunflower, rye, and borage) had seed coats slightly larger than expected by the allometry (Fig. 2).

Effects of evolution under cultivation on biomass distribution in crop seeds

Seed coat ratios (dry mass in coats divided by total seed dry mass) were generally smaller in domesticated accessions of the crops than in their wild progenitors (Figs 3, 4). Hedges' G for the effects of domestication ranged from 1.12 in sesame to −15.58 in maize (Fig. 4). The vast majority of domestication effects on SCR were negative and large. Fourteen, out of 17, crops had their coats reduced during domestication by a Hedges´ G larger than (0.8), which is a threshold for large effect sizes (Cohen, 2013). Within those 14 crops, a few had their coats drastically reduced during domestication, maize being the most extreme example of this (Fig. 4), whereas for most of the other crops, the magnitude of SCR reductions with domestication was large but not extreme (Fig. 4). Reductions in SCR took place in the early stages of crop evolution under cultivation, since wild progenitors had consistently higher SCR than early landraces (but see sesame, cowpea, and borage, which are exceptions to this rule; Fig. 4). By contrast, recent plant breeding had small and inconsistent effects on SCRs (Fig. 4). For a few crops (barley, black medick, and amaranth) SCR increased during recent breeding, and for a few others (peas, soybean, and chickpea) coats became even smaller at this stage. However, for a majority of crops, recent breeding did not trigger major changes in proportional allocation to the seed coats (Fig. 4).

Discussion

We showed that crop seeds generally conform to the allometric scaling of seed kernel and coat mass and lie within the phenotypic variation in these traits found across wild plants globally. However, within that spectrum of variation, most crops have coats lighter than expected by their seed size. This rejects the hypothesis that reduced physical dormancy in crops evolved solely as a genetic correlate of selection for larger seeds. Interestingly, the wild progenitors of crops, like their domesticated relatives, also showed lighter-than-expected coats. Together, these results suggest that at least three different evolutionary processes have contributed to the lightweight seed coats typical of herbaceous crops (Fig. 5): indirect allometric effects of selection for larger seeds (see subsection Crops in the seed kernel-coat allometry below), selection of wild progenitors with seed coats lighter-than-expected given their seed size (see subsection Crop wild progenitors had smaller coats than expected by their seed size below), and evolution of low allocation to the seed coat during the early stages of domestication (see subsection Biomass allocation to coats across the different stages of crop evolution below).

Fig. 3 Effects of evolution under cultivation for the allocation of mass to the seed coat. (a) Effects of domestication on seed coat ratios (SCR) (dry mass in coats divided by total seed dry mass, SCR). Domestication effects are taken as the comparison between crop wild progenitors and landraces. (b) Effects of recent crop breeding on SCR. Effects of recent breeding are taken as the comparison between landraces and modern cultivars of a crop. ns, nonsignificant; ***, $P < 0.001$; are statistical significances of the overall differences among the two plant groups, taken from false discovery rate-corrected Tukey tests after mixed-effects linear models. Points are arithmetic means and error bars are SEs of the mean.
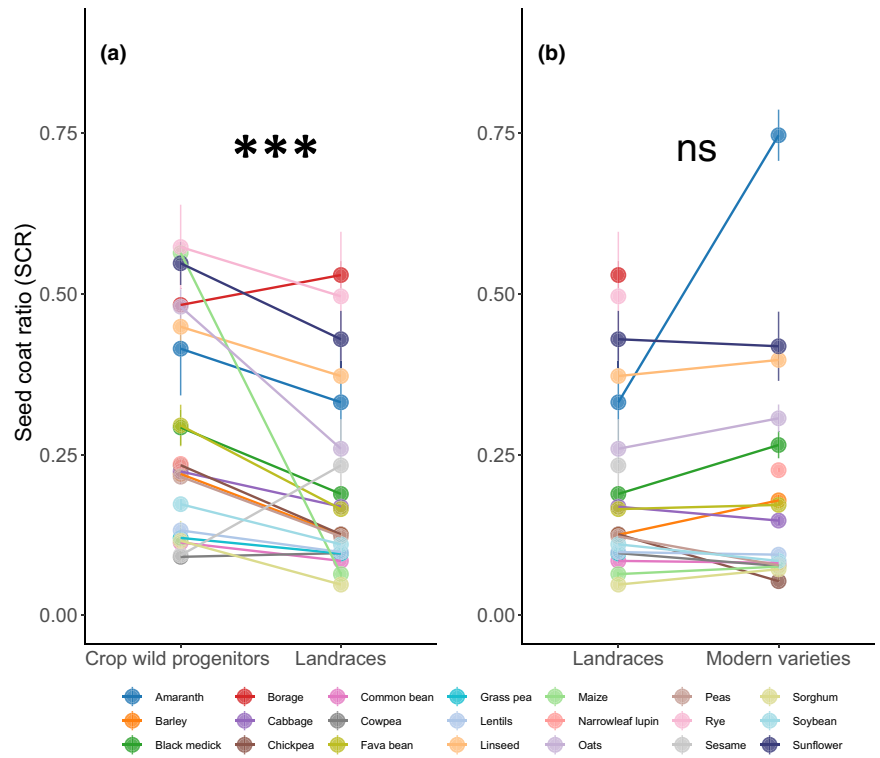

Fig. 4 Effect sizes of domestication and recent breeding on seed coat mass ratios (SCR). Effect sizes of domestication (left) and recent breeding (right) on SCR for each crop. Points are estimates of Hedges´ G (metric of effect size) and their 95% confidence intervals (CIs). Overlap of the 95% CI with zero was taken as evidence of negligible effects and nonoverlap as a significant effect. Note that x-axis breaks were introduced between −5 and −10 to facilitate visualization of CIs overlap with zero. Effect sizes were missing for crops for which we did not obtain seeds of any of the two domestication stages involved in each of the two effects (e.g. peanuts, for which we have data of its wild progenitor and of modern cultivars of the crop, but not of landraces).
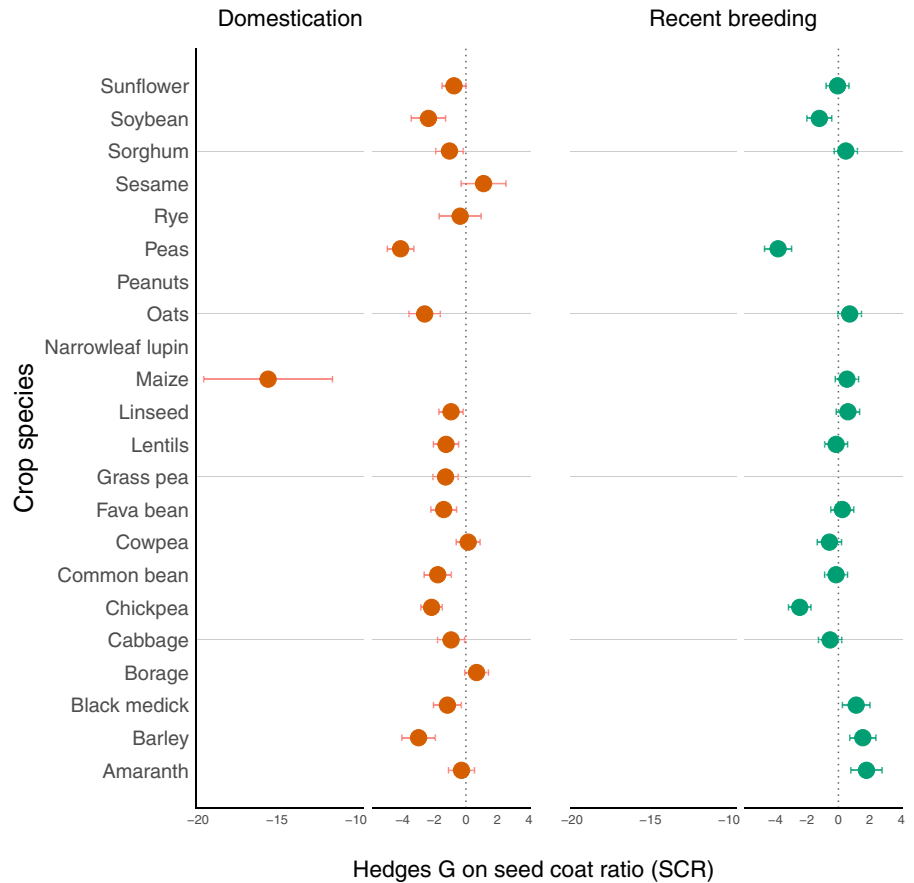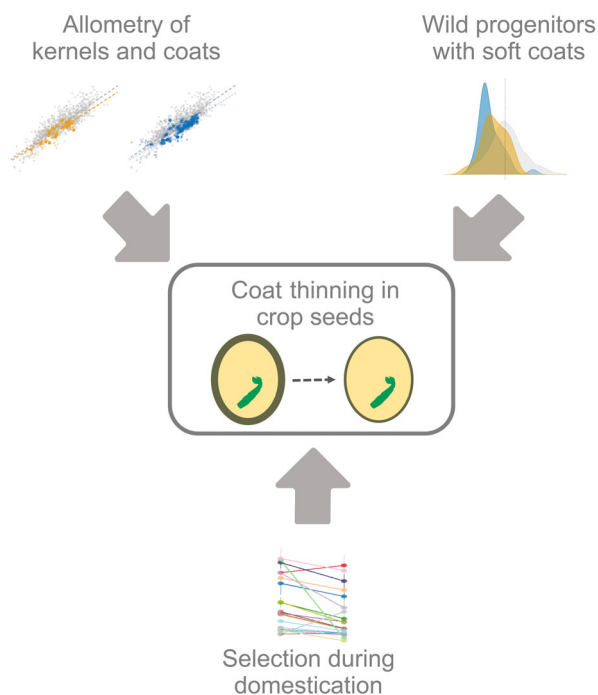
Fig. 5 Conceptual diagram of the mechanisms influencing the evolution of seed coats in crops. In this paper, we report evidence that at least three different evolutionary processes have influenced the emergence of lightweight seed coats in herbaceous crops. Allometric effects of selection for larger seeds (see results in Fig. 1), selection of wild progenitors with seed coats lighter than expected given their seed size (see results in Fig. 2), and evolution of low allocation to the seed coat during the early stages of domestication (see results in Figs 3, 4). [Correction added on 26 May 2024, after first online publication: a labelling error in Fig. 5 has been corrected.]

Crops in the seed kernel-coat allometry

Our PGLS slope for wild species, the SMA slopes reported in Wu *et al.* (2019), and the PGLS slopes reported in Chen *et al.* (2020) are all allometric, even if ours (0.82) deviated from isometry more strongly than Wu's (0.88–0.93) and Chen's (0.89–0.91). Also, crops conformed to the allometric slopes (Fig. 1). Several explanations might account for a negative allometry between kernel and coat sizes. Small seeds may need more investment in defence, since they can be eaten by a wider range of seed predators, and tend to persist for longer in the seed bank (Eriksson & Eriksson, 1997; Bekker *et al.*, 1998; Wu *et al.*, 2019). Also, small seeds tend to be packaged in infructescence with many seeds, and these can be targeted by large herbivores as part of their foraging strategies (i.e. the so-called fruit as foliage foraging strategy, Janzen, 1984). This may select for strong coats in small seeds to endure the transit through animals' guts (Spengler III & Mueller, 2019). In large seeds, more reserve tissue is needed to attain the high seedling vigour typical of large-seeded species (Venable & Brown, 1988; Wu *et al.*, 2019). High investment in reserve tissue seems also adaptive for the seeds of crops, both as an ecological adaptation to the agricultural environment – high seedling vigour and adaptation to burial – and as a consequence of human selection for high yields

(Kluyver *et al.*, 2013; Preece *et al.*, 2021). Importantly, none of the crops stepped out of the global variation observed for the relationship of kernel to seed coat masses in wild plants (Fig. 1). Thus, evolution under cultivation did not move crops to phenotypic regions beyond the boundaries of variation in the wild for these traits.

Mass of kernels and coats scaled positively, which concurs with developmental patterns of coordinated growth of seed constituents (García *et al.*, 2005; Sundaresan, 2005). Even if this positive scaling was a constraint to the free variation of coat and kernel mass (PGLS $R^2 = 0.64$), there was a wide deviation from the allometry among individual species. In wild species, deviations were highly varied, either positive, negative, and of different magnitudes. This indicates that diverse mechanisms and selection forces, other than the allometry, influence biomass distribution in the seeds. For example, similarly large-seeded species may bear either soft- or hard-coated seeds, depending on their ecological strategies (Baskin & Baskin, 1998). In the linages with soft seeds, the enlargement of reserve tissues evolves faster than that of coats, which generates consistently negative deviations from the allometry (Chen *et al.*, 2020). Fire frequency and intensity select for coats thicker than expected for a given seed size, while thin-coated seeds might be favoured in dimorphic seeds to bet hedge the survival to secondary dispersal by seed predators (Moreira & Pausas, 2012; Paulsen *et al.*, 2013). At the developmental level, the relative dominance of the parental and maternal genomes during embryogenesis generates variation in seed size and differential growth of the tissues in the seed (Sundaresan, 2005). These and other mechanisms, together with the allometry, generate variation in coat and kernel mass distributions among wild species. But deviations of the allometry in most crops were consistently large and negative, in contrast with the combination of positive and negative deviations found for wild species. This means that crop seeds had similarly smaller-than-expected coats, at all ranges of seed size. It is worth mentioning that crops and wild plants were a widely different subsample of the seed plants. For instance, wild plants in Chen *et al.* (2020) included woody plants and plants with recalcitrant seeds, which were absent in our dataset on crops. This heterogeneity among the two groups of plants might partly explain biases in the magnitude and direction of deviations for wilds and crops, which merits further research. In any case, provided systematic large and negative deviations from the allometry in crops, their reduced coats cannot be explained solely as an allometric correlate of selection for larger seeds.

It is important to note that a reduction in biomass allocation to the coat (the metric used in this manuscript) does not necessarily imply a decrease in coat thickness (a geometrical trait). In fact, modelling seed geometry as spheres of diverse sizes gives that an allometric scaling of kernel and coat sizes (taken as volumes) is compatible with thinning, thickening, or constancy in the width of the seed coats (Supporting Information Table S1). Assuming that geometric volumes can be transferred to dry mass, this means that seed coat thinning in crops cannot be attributed solely to a decrease in biomass allocation to the coat. However, this argumentation must be taken with care. Taking volume as mass assumes that tissue-specific density stays constant, which may not hold, because tissue proportions and palisade cell sizes change between coats

of different thicknesses (Hradilová *et al.*, 2017; Zablatzká *et al.*, 2021). Investigating changes in seed coat tissue density alongside mass and thickness can shed light on this. Also, even if coat thickness were partly independent from biomass allocation to the coat, the proportionally larger kernel of large seeds has a greater potential for imbibition, swelling and potentially breaking the coat, which might also facilitate the release of physical dormancy.

Crop wild progenitors had smaller coats than expected by their seed size

A majority of the crop wild progenitors had coats lighter than expected, provided their seed sizes. This effect was larger for peanuts, sesame, lentils, or sorghum, but was otherwise generalized, except in a few crops – maize, sunflower, oats, and rye. Previous work on the evolution of crop seed coats has focused on the changes occurring during or after domestication, comparing wild progenitors and cultivated varieties, or examining archaeological series across the Holocene (Smith, 1989; Janská *et al.*, 2019). We speculate that, even if agricultural practices had started with dormant genotypes, the wild progenitors of crops had lower physical dormancy than other wild plants, which provided a baseline for the evolution of the lightweight seed coats of crops. Similar preadaptations have been advocated earlier to explain the acquisitive physiological profile of crops (Gómez-Fernández & Milla, 2024). Previous work on pulses has also suggested that the existence of populations of pulse wild progenitors with dimorphic individuals, harbouring both thick- and thin-coated genotypes, was a prerequisite for their adoption as crops (Ladizinsky, 1987; Abbo *et al.*, 2011). Seed polymorphism is widespread in plants (Silvertown, 1984), and its relevance to domestication may extend to pseudocereals originating in the Americas (Mueller, 2017; Belcher *et al.*, 2023). However, this may not generalize to other crops, which coat thinning series in the archaeological record tend to proceed slowly and might not predate domestication (e.g. *Macrotyloma uniflorum*, Murphy & Fuller, 2017). Overall, the role of wild progenitors' coats on the evolution of reduced dormancy in crops still needs more research. Further, a few of the wild progenitors had seed coats heavier than expected by their seed size. The most remarkable example of this is maize. The wild teosintes originating maize had a strong seed encapsulation, which extreme reduction happened entirely during domestication (Stitzer & Ross-Ibarra, 2018). But for most crops, reduced physical dormancy might have been facilitated by the choice of wild species for cultivation that had coats lighter than other wild species.

Biomass allocation to coats across the different stages of crop evolution

For most crops, we found that primitive landraces allocate less biomass to their seed coats than their wild progenitors. By contrast, landraces differed little, and idiosyncratically – that is different crop species behaving differently – from their modern cultivar counterparts. This indicates that the stages of initial domestication and early geographic spread of domesticates were key to the evolution

of crop seeds. This is in wide accordance with the archaeological record of many crops (Smith, 1989; Murphy & Fuller, 2017; Fuller *et al.*, 2023). In fact, coat thinning of archaeological samples is one of the signatures of domestication in several crops (e.g. in American chenopods; Smith, 1989), together with other traits like seed size and morphology (Bruno *et al.*, 2018). Our results are also in line with previous studies comparing contemporary accessions of wild progenitors with cultivars (Maass *et al.*, 2007; Sedláková *et al.*, 2021; Balarynová *et al.*, 2022). Our contribution extends this pattern to biomass partitioning and to types of crops poorly covered in the seed coat vs domestication literature like linseed, cabbage, medick, and others.

Direct human selection seems an unlikely mechanism for seed coat thinning in early domestication. Some archaeological records show that coats thinned gradually across millennia (Murphy & Fuller, 2017), which may be incompatible with farmers' timescale for selection. Indirect mechanisms, like selection for yield, or relaxation of natural selection (e.g. no need for resistance to endozoochory, Spengler III & Mueller, 2019), might be involved. In any case, more research should be carried to unveil the selection forces leading to low SCRs and thin coats during the early stages of domestication. As for modern breeding, this recent stage of crop evolution impacted SCRs in diverse ways, probably reflecting that modern selection on coats did not pursue advances in the physiological performance of crops. Instead, modern breeding has targeted seed coats for traits as diverse as seed oil content (Zhang *et al.*, 2018), polyphenols levels (Zhong *et al.*, 2019; Balarynová *et al.*, 2022), seed coat pigmentation (Smýkal *et al.*, 2018), seed longevity (Tiwari & Hariprasad, 1997), postharvest endurance (Elsadr *et al.*, 2011), and other industrial applications (Moïse *et al.*, 2005). Such diversification of breeding aims probably had diverse but small effects on the allocation of mass between kernels and coats in recent times, which is in line with our results.

Conclusions

Our analyses suggest that a collection of mechanisms is needed to explain the seed coats of crops (Fig. 5). Crops fit the allometric scaling slopes of kernel and coat biomass. Thus, the allometry partially contributes to explain their lightweight coats, since crop seeds are larger than those of their wild progenitors. Still, additional mechanisms are needed because crops had consistently smaller seed coats than expected by the allometry. We show evidence that at least two other mechanisms are at play. First, crop wild progenitors tend to be lightweight-coated wild species, and thus, their selection as agricultural species provided a head start for the evolution of soft-coated seeds. Second, during early domestication there was a further reduction of biomass allocation to the coat, which is in line with previous archaeological evidence. This work showcases that crop traits are the result of multiple evolutionary processes, and thus provides new insight on the emergence of a key domestication trait. Furthermore, we showed that an allometric approach is useful to address challenges in crop biology, and suggest that applications like breeding on allometric slopes, intercepts, and deviations from the scaling are future avenues for targeting valuable trait combinations in crops.

Acknowledgements

Competing interests

None declared.

Authors contributions

RM conceived the study. RM, PS, JB, JM-V and SN-C contributed to the materials and data. AJW contributed a modelling exercise. RM analysed the data and wrote the first draft of the manuscript. All authors contributed to the subsequent rounds of writing and gave the approval for submission of the final version.

ORCID

Rubén Milla https://orcid.org/0000-0001-8912-4373
Petr Smykal https://orcid.org/0000-0002-6117-8510
Adrianus J. Westgeest https://orcid.org/0000-0002-6128-6909

Data availability

The sources of data for each species in the database are available at doi: 10.5281/zenodo.10973414. Unpublished raw data of this manuscript are available also at doi: 10.5281/zenodo.10973414. Raw data compiled from the literature and from Chen *et al.* (2020) are not provided to avoid duplicate deposit of primary data. The analyses carried out in this manuscript did not require the development of custom code. Functions were run as provided by the R packages mentioned in the Materials and Methods section.

References

Abbo S, Rachamim E, Zehavi Y, Zezak I, Lev-Yadun S, Gopher A. 2011. Experimental growing of wild pea in Israel and its bearing on Near Eastern plant domestication. *Annals of Botany* **107**: 1399–1404.

Balarynová J, Klčová B, Sekaninová J, Kobrlová L, Cechová MZ, Krejčí P, Leonova T, Gorbach D, Ihling C, Smržová L et al. 2022. The loss of polyphenol oxidase function is associated with hilum pigmentation and has been selected during pea domestication. *New Phytologist* **235**: 1807–1821.

Barton K. 2015. Package 'MuMIn'. Model selection and model averaging based on information criteria. [WWW document] URL https://cran.r-project.org/web/packages/MuMIn/ [accessed 31 January 2024].

Baskin CC, Baskin JM. 1998. *Seeds: ecology, biogeography, and, evolution of dormancy and germination.* San Diego, CA, USA: Elsevier.

Bekker RM, Bakker JP, Grandin U, Kalamees R, Milberg P, Poschlod P, Thompson K, Willems JH. 1998. Seed size, shape and vertical distribution in the soil: indicators of seed longevity. *Functional Ecology* **12**: 834–842.

Belcher ME, Williams D, Mueller NG. 2023. Turning over a new leaf: experimental investigations into the role of developmental plasticity in the domestication of goosefoot (*Chenopodium berlandieri*) in Eastern North America. *American Antiquity* **88**: 554–569.

Bruno MC, Pinto M, Rojas W. 2018. Identifying domesticated and wild kañawa (*Chenopodium pallidicaule*) in the archeobotanical record of the Lake Titicaca Basin of the Andes. *Economic Botany* **72**: 137–149.

Bruno MC, Whitehead WT. 2003. Chenopodium cultivation and formative period agriculture at Chiripa, Bolivia. *Latin American Antiquity* **14**: 339–355.

Casas RR d, Willis CG, Pearse WD, Baskin CC, Baskin JM, Cavender-Bares J. 2017. Global biogeography of seed dormancy is determined by seasonality and seed size: a case study in the legumes. *New Phytologist* **214**: 1527–1536.

Chen SC, Wu LM, Wang B, Dickie JB. 2020. Macroevolutionary patterns in seed component mass and different evolutionary trajectories across seed desiccation responses. *New Phytologist* **228**: 770–777.

Cohen J. 2013. *Statistical power analysis for the behavioral sciences.* Cambridge, MA, USA: Academic press.

Daws MI, Garwood NC, Pritchard HW. 2006. Prediction of desiccation sensitivity in seeds of woody species: a probabilistic model based on two seed traits and 104 species. *Annals of Botany* **97**: 667–674.

Elsadr HT, Wright LC, Peter Pauls K, Bett KE. 2011. Characterization of seed coat post harvest darkening in common bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics* **123**: 1467–1472.

Eriksson Å, Eriksson O. 1997. Seedling recruitment in semi-natural pastures: the effects of disturbance, seed size, phenology and seed bank. *Nordic Journal of Botany* **17**: 469–482.

Fricke EC, Wright SJ. 2016. The mechanical defence advantage of small seeds. *Ecology Letters* **19**: 987–991.

Fuller DQ, Allaby R. 2018. Seed dispersal and crop domestication: shattering, germination and seasonality in evolution under cultivation. *Annual Plant Reviews Online* **38**: 238–295.

Fuller DQ, Denham T, Allaby R. 2023. Plant domestication and agricultural ecologies. *Current Biology* **33**: R636–R649.

García D, Gerald JNF, Berger F. 2005. Maternal control of integument cell elongation and zygotic control of endosperm growth are coordinated to determine seed size in Arabidopsis. *Plant Cell* **17**: 52–60.

Garibaldi LA, Aizen MA, Sáez A, Gleiser G, Strelin MM, Harder LD. 2021. The influences of progenitor filtering, domestication selection and the boundaries of nature on the domestication of grain crops. *Functional Ecology* **35**: 1998–2011.

Gómez-Fernández AI, Milla R. 2024. Early human selection of crops' wild progenitors explains the acquisitive physiology of modern cultivars. *Nature Plants* **10**: 25–36.

Grafen A, Hamilton WD. 1989. The phylogenetic regression. *Philosophical Transactions of the Royal Society of London. B: Biological Sciences* **326**: 119–157.

Hansen TF, Bartoszek K. 2012. Interpreting the evolutionary regression: the interplay between observational and biological errors in phylogenetic comparative studies. *Systematic Biology* **61**: 413–425.

Harlan JR, de Wet JM, Price EG. 1973. Comparative evolution of cereals. *Evolution* **27**: 311–325.

He T, Angessa TT, Li C. 2023. Pleiotropy structures plant height and seed weight scaling in barley despite long history of domestication and breeding selection. *Plant Phenomics* **5**: 15.

Hedges LV, Gurevitch J, Curtis PS. 1999. The meta-analysis of response ratios in experimental ecology. *Ecology* **80**: 1150–1156.

Hothorn T. 2016. Package 'MULTCOMP'. Simultaneous inference in general parametric models. [WWW document] URL https://cran.r-project.org/web/packages/multcomp [accessed 31 January 2024].

Hradilová I, Trněný O, Válková M, Cechová M, Janská A, Prokešová L, Aamir K, Krezdorn N, Rotter B, Winter P et al. 2017. A combined comparative transcriptomic, metabolomic, and anatomical analyses of two key domestication traits: pod dehiscence and seed dormancy in Pea (*Pisum* sp.). *Frontiers in Plant Science* **8**: 542.

Janská A, Pecková E, Sczepaniak B, Smýkal P, Soukup A. 2019. The role of the testa during the establishment of physical dormancy in the pea seed. *Annals of Botany* 123: 815–829.

Janzen D. 1984. Dispersal of small seeds by big herbivores: foliage is the fruit. *The American Naturalist* 123: 338–353.

Jin Y, Qian H. 2019. V.PHYLOMAKER: an R package that can generate very large phylogenies for vascular plants. *Ecography* 42: 1353–1359.

Jin Y, Qian H. 2022. V.PHYLOMAKER2: an updated and enlarged R package that can generate very large phylogenies for vascular plants. *Plant Diversity* 44: 335–339.

Kilmer JT, Rodríguez RL. 2017. Ordinary least squares regression is indicated for studies of allometry. *Journal of Evolutionary Biology* 30: 4–12.

Kindt R. 2020. WORLDFLORA: an R package for exact and fuzzy matching of plant names against the World Flora Online taxonomic backbone data. *Applications in Plant Sciences* 8: e11388.

Kluyver TA, Charles M, Jones G, Rees M, Osborne CP. 2013. Did greater burial depth increase the seed size of domesticated legumes? *Journal of Experimental Botany* 64: 4101–4108.

Kluyver TA, Jones G, Pujol B, Bennett C, Mockford EJ, Charles M, Rees M, Osborne CP. 2017. Unconscious selection drove seed enlargement in vegetable crops. *Evolution Letters* 1: 64–72.

Ladizinsky G. 1987. Pulse domestication before cultivation. *Economic Botany* 41: 60–65.

Larson G, Piperno DR, Allaby RG, Purugganan MD, Andersson L, Arroyo-Kalin M, Barton L, Vigueira CC, Denham T, Dobney K *et al.* 2014. Current perspectives and the future of domestication studies. *Proceedings of the National Academy of Sciences, USA* 111: 6139–6146.

Maass BL, Usongo MF, Maass BL, Usongo MF. 2007. Changes in seed characteristics during the domestication of the lablab bean (*Lablab purpureus* (L.) Sweet: Papilionoideae). *Australian Journal of Agricultural Research* 58: 9–19.

Milla R. 2020a. Crop origins and phylo food. GitHub. [WWW document] URL https://github.com/rubenmilla/Crop_Origins_Phylo [accessed 31 January 2024].

Milla R. 2020b. Crop origins and phylo food: a database and a phylogenetic tree to stimulate comparative analyses on the origins of food crops. *Global Ecology and Biogeography* 29: 606–614.

Moïse JA, Han S, Gudynaitę-Savitch L, Johnson DA, Miki BLA. 2005. Seed coats: structure, development, composition, and biotechnology. *In Vitro Cellular & Developmental Biology. Plant* 41: 620–644.

Moles AT. 2005. A brief history of seed size. *Science* 307: 576–580.

Moreira B, Pausas JG. 2012. Tanned or burned: the role of fire in shaping physical seed dormancy. *PLoS ONE* 7: e51523.

Mueller NG. 2017. Documenting domestication in a lost crop (*Polygonum erectum* L.): evolutionary bet-hedgers under cultivation. *Vegetation History and Archaeobotany* 26: 313–327.

Murphy C, Fuller DQ. 2017. Seed coat thinning during horsegram (*Macrotyloma uniflorum*) domestication documented through synchrotron tomography of archaeological seeds. *Scientific Reports* 7: 5369.

OpenAI. 2023. GPT-3.5 Chat Model.

Orme D. 2018. The CAPER package: comparative analysis of phylogenetics and evolution in R.

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

Paulsen TR, Colville L, Kranner I, Daws MI, Högstedt G, Vandvik V, Thompson K. 2013. Physical dormancy in seeds: a game of hide and seek? *New Phytologist* 198: 496–503.

Pedersen TL. 2019. Package 'PATCHWORK'. [WWW document] URL https://cran.r-project.org/web/packages/patchwork [accessed 31 January 2024].

Pinheiro J, Bates D, DebRoy S, Sarkar D, Core RT. 2020. NLME: linear and nonlinear mixed effects models. [WWW document] URL https://cran.r-project.org/web/packages/nlme [accessed 31 January 2024].

Preece C, Jones G, Rees M, Osborne CP. 2021. Fertile Crescent crop progenitors gained a competitive advantage from large seedlings. *Ecology and Evolution* 11: 3300–3312.

Purugganan MD, Fuller DQ. 2009. The nature of selection during plant domestication. *Nature* 457: 843–848.

Rees M. 1997. Evolutionary ecology of seed dormancy and seed size. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 351: 1299–1308.

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

Sadras VO. 2007. Evolutionary aspects of the trade-off between seed size and number in crops. *Field Crops Research* 100: 125–138.

Sedláková V, Hanáček P, Grulichová M, Zablatzká L, Smýkal P. 2021. Evaluation of seed dormancy, one of the key domestication traits in Chickpea. *Agronomy* 11: 2292.

Shingleton AW. 2019. Which line to follow? The utility of different line-fitting methods to capture the mechanism of morphological scaling. *Integrative and Comparative Biology* 59: 1399–1410.

Silvertown JW. 1984. Phenotypic variety in seed germination behavior: the ontogeny and evolution of somatic polymorphism in seeds. *The American Naturalist* 124: 1–16.

Smith BD. 1989. Origins of agriculture in Eastern North America. *Science* 246: 1566–1571.

Smith SA, Brown JW. 2018. Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany* 105: 302–314.

Smýkal P, Nelson MN, Berger JD, Von Wettberg EJB. 2018. The impact of genetic changes during crop domestication on healthy food development. *Agronomy* 8: 26.

Smýkal P, Vernoud V, Blair MW, Soukup A, Thompson RD. 2014. The role of the testa during development and in establishment of dormancy of the legume seed. *Frontiers in Plant Science* 5: 351.

Spengler R III, Mueller N. 2019. Grazing animals drove domestication of grain crops. *Nature Plants* 5: 656–662.

Stitzer MC, Ross-Ibarra J. 2018. Maize domestication and gene interaction. *New Phytologist* 220: 395–408.

Sundaresan V. 2005. Control of seed size in plants. *Proceedings of the National Academy of Sciences, USA* 102: 17887–17888.

Symonds MRE, Blomberg SP. 2014. A primer on phylogenetic generalised least squares. In: Garamszegi LZ, ed. *Modern phylogenetic comparative methods and their application in evolutionary biology: concepts and practice.* Berlin, Heidelberg, Germany: Springer, 105–130.

Tiwari SP, Hariprasad AS. 1997. Selection criteria for seed longevity in soya bean. *Tropical Agriculture* 47: 70–72.

Torchiano M. 2020. Package 'EFFSIZE'. [WWW document] URL https://cran.r-project.org/web/packages/effsize [accessed 31 January 2024].

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

Wickham H. 2011. GGPLOT2. *WIREs Computational Statistics* 3: 180–185.

Wu LM, Chen SC, Wang B. 2019. An allometry between seed kernel and seed coat shows greater investment in physical defense in small seeds. *American Journal of Botany* 106: 371–376.

Xu S, Chen M, Feng T, Zhan L, Zhou L, Yu G. 2021. Use ggbreak to effectively utilize plotting space to deal with large datasets and outliers. *Frontiers in Genetics* 12: 77486.

Zablatzká L, Balarynová J, Klčová B, Kopecký P, Smýkal P. 2021. Anatomy and histochemistry of seed coat development of wild (*Pisum sativum* subsp. *elatius* (M. Bieb.) Asch. et Graebn. and domesticated pea (*Pisum sativum* subsp. *Sativum* L.)). *International Journal of Molecular Sciences* 22: 4602.

Zhang D, Sun L, Li S, Wang W, Ding Y, Swarm SA, Li L, Wang X, Tang X, Zhang Z *et al.* 2018. Elevation of soybean seed oil content through selection for seed coat shininess. *Nature Plants* 4: 30–35.

Zhong L, Wu G, Fang Z, Wahlqvist ML, Hodgson JM, Clarke MW, Junaldi E, Johnson SK. 2019. Characterization of polyphenols in Australian sweet lupin (*Lupinus angustifolius*) seed coat by HPLC-DAD-ESI-MS/MS. *Food Research International* 116: 1153–1162.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Seed mass of domesticated and wild progenitor accessions of each crop species.

Table S1 Modelling of volumetric size of kernels and coats and their scaling with increasing seed size.