The shape and volume of air, kernels and cracks, in a nutshell

Erik J. Amézquita[1,2,†], Michelle Y. Quigley[3], Patrick Brown[6], Elizabeth Munch[4,5], Daniel H. Chitwood[3,4,†]

[1]Division of Plant Sciences & Technology, [2]Department of Mathematics, University of Missouri, Columbia, MO 65211, USA

[3]Department of Horticulture, [4]Department of Computational Mathematics, Science & Engineering, [5]Department of Mathematics, Michigan State University, East Lansing, MI 48824, USA

[6]Department of Plant Sciences, University of California, Davis, CA 95695, USA

Short title: Measuring the shape of walnuts

Keywords: walnuts; shell crack; mathematical biology; data science; shape

[†]To whom correspondence should be addressed:

Dr. Erik Amézquita
240a Bond Life Sciences Center, Columbia, MO 65211
(573) 884-9601
eah4d@missouri.edu

Dr. Daniel H. Chitwood
1066 Bogue St, East Lansing, MI 48824
(517) 353-0462
chitwoo9@msu.edu

Abstract

Walnuts are the second most produced and consumed tree nut, with over 2.6 million metric tons produced in the 2022-23 harvest cycle alone. The United States is the second largest producer, accounting for 25% of the total global supply. Nonetheless, producers face an ever-growing demand in a more uncertain climate landscape, which requires effective and efficient walnut selection and breeding of new cultivars with increased kernel content and easy-to-open shells. Past and current efforts select for these traits using hand-held calipers and eye-based evaluations. Yet there is plenty of morphology that meets the eye but goes unmeasured, such as the volume of inner air or the convexity of the kernel. Here, we study the shape of walnut fruits based on X-ray CT (Computed Tomography) 3D reconstructions. We compute 49 different morphological phenotypes for 1264 individuals comprising 149 accessions. These phenotypes are complemented by traits of breeding interest such as ease of kernel removal and kernel weight. Through allometric relationships —relative growth of one tissue to another—, we identify possible biophysical constraints at play during development. We explore multiple correlations between all morphological and commercial traits, and identify which morphological traits can explain the most variability of commercial traits. We show that using only volume and thickness-based traits, especially inner air content, we can successfully encode several of the commercial traits.

Core Ideas

- X-ray Computed Tomography (CT) imaging is used to compute a broad array of morphological phenotypes in walnuts.
- These morphological traits suggest biophysical constraints at play during walnut development.
- Relative inner air, shell, and packing tissue volumes are significantly correlated to the rest of shape phenotypes.
- These volumes produce the best prediction models for traits of commercial interest such as shell strength.
- Inexpensive phenotyping platforms that focus solely on volume measurement would enable better walnut breeding.

1 Introduction

There is more than meets the eye in the shape of walnuts (*Juglans regia*). Civilizations originary from modern day Iran have used and traded walnut fruit and tree products since the 7000 BC (Vahdati, 2014). From there and then, walnuts traveled far and wide as they were actively traded through the Silk Road, conquering the Eurasian continent (Pollegioni et al., 2014). The trade of walnuts remains an important part of the global economy. In 2021, the California and the US produced more than 725,000 tons of walnuts valued in more than $1.0B, following a historically increasing trend of both bearing acreage and bearing trees per acre (NASS, 2022). World demand for walnut keeps increasing and it is estimated that the world will consume a record 2.5M tons of walnuts for 2023, and the US is forecasted to satisfy 25% of the global demand (FAS, 2022). The trade is not limited to the food industry, as there is also growing research on additional uses for walnut shell material. This material can be key for more durable batteries (Wahid et al., 2017), lower-cost concrete (Hilal et al., 2020), and stronger epoxy composites (Lala et al., 2018), to name a few examples. As climate change alters weather patterns, and the demand for walnut and its byproducts increases, we must breed walnuts with more suitable traits such as high kernel-to-total weight ratio, adequate shell strength, and easiness to extract both kernel halves intact. Quantitative analyses and comprehensive pheontyping can accelerate current breeding programs by quickly identifying varieties and individuals with desirable characteristics (Fiorani and Schurr, 2013; Rahaman et al., 2015). The rapid selection of potentially desirable progenitors for breeding programs is especially crucial for walnuts, as seedlings are hard to propagate. Even for the fastest-growing accessions, it takes at least 2 years for trees to bear fruit for the first time, and at least 5 more to yield fruit at a commercial scale (Lopez, 2004; Popa et al., 2023).

Most of the current walnut phenotyping follows the measuring guidelines set by the International Plant Genetic Resources Institute (IPGRI, 1994). The morphological phenotyping of the fruit is mainly done using calipers to measure length, width, and height, combined with visual assessments to describe more complicated traits such as texture and sphericity. These simple measurements have proved to be insightful to evaluate and identify promising genotypes. Moderate correlations have been reported between these traditional morphological traits of the walnut tree and fruit with commercial and horticultural traits of interest such as pollen release strategy, yield, shell thickness, kernel weight, and pathogen resistance (Akca and Şen, 1995; Kelc et al., 2007; Khadivi-Khub et al., 2015; Rezaei et al., 2018; Shah et al., 2021; Solar et al., 2003).

However, this caliper- and eye-based approach is time consuming, prone to human error and subjectivity, and fails to capture richer shape nuance observed in the shells and kernels. As next-generation sequencing (NGS) technology advances, we observe an explosion in genomics data collection that must be matched by equally powerful and encompassing phenomics (Araus and Cairns, 2014; Bucksch et al., 2017). We have to look deeper than just nut lengths and widths. To that end, X-ray computed tomography (CT) scanning has proved to be a powerful tool to accurately capture intricate, internal features of a vast array of plant data in a nondestrutive manner. High-resolution, X-ray CT 3D reconstructions have been successfully used to capture and quantify the complex branching architecture of inflorescence in grapevines (Li et al., 2019) and sorghum panicles (Li et al., 2020), identify key morphological traits in barley seeds to distinguish their accession of origin (Amézquita et al., 2021), and determine nuances in soil porosity for diverse wheat root-soil interactions (Zhou et al., 2020)

To the best of our knowledge, Bernard et al. (2020) is the first study that exploits X-ray CT imaging to automatically, accurately, and systematically quantify multiple walnut shape phenotypes. Their results showcase the morphological variability found across the germplasm diversity panel maintained by INRIA, France. There, Bernard et al. measure the absolute and relative volumes of the whole nut and its shell, kernel, and internal air. In particular, they observe that larger fruits are correlated with rougher shell shape and smaller kernel filling ratio, which allows them to select for better genotypes. X-ray CT imaging has also been recently used to document morphological changes of walnut flower bud development (Gao, 2022), estimate kernel weight in an nondestructive way (Gao et al., 2022), and to explore the puzzling diversity and structure of the cell tesselations that conform the hard shell tissue for multiple nuts (Huss et al., 2020).

Here, we study the shape of walnut fruits based on the X-ray CT 3D reconstruction of 1264 individuals comprising 149 accessions maintained by the Walnut Improvement Program at the University of California Davis. We exploit the nondestructiveness of X-rays to isolate individual walnuts and segment out shell, kernel, and packing tissues, as well as the air contained inside every walnut. We first compute 49 different shape- and size-related traits for each walnut such as nut and kernel dimensions, surface areas, volumes, filling ratios, sphericity and convexity indices. We included the computation of the 14 traits used by Bernard et al. (2020). All the image processing tasks were done with an in-house, python-based, open-source script. This

morphological information was combined with values of breeding interest that were collected separately throughout different years. These traits include ease of removing both kernel halves intact, shell strength, and kernel-to-nut weight ratio. Second, we look for allometric relationships of interest across the whole population —the growth rate of a tissue relative to another— which reveal possible biophysical constraints at play during walnut development. These allometric relationships pose theoretical limits on minimum and maximum possible walnut sizes. Third, we examine Spearman correlation coefficients between all the computed morphological phenotypes. We find that the relative content of inner air, an often overlooked trait, is significantly correlated to the rest of tissue sizes and shapes. This suggests that the inner air content plays an important role during tissue development. The air volume is also significantly correlated with several traits of breeding interest. Fourth, we compute several stepwise linear regression models to determine which shape and size phenotypes contribute the most to relative kernel weight, ease of kernel removal, and shell strength. Our results suggest that these traits of breeding interest can be best predicted with only a handful of independent phenotypes, including inner air content. Finally, we examined the distribution of these phenotypes within each accession and through the general population. A principal component analysis using only the 6 most relevant morphological phenotypes reveals that accessions follow a gradient according to their shell strength scores and kernel weight ratio. This suggests that key traits can be successfully predicted if the volume of different tissues is known, which opens the possibility of engineering affordable phenotyping plantforms that only focus on volumetric analyses rather than expensive CT setups. This morphological modeling will allow us to set a new exciting path to explore further the phenotype-genotype relationship in walnuts.

2 Materials and methods

2.1 Plant material and scanning

All plant materials represent walnut breeding lines, germplasm, and cultivars maintained by the Walnut Improvement Program at the University of California, Davis. A total of 150 walnuts accessions were harvested into mesh bags at hull split, oven-dried overnight at 95°F, and then air-dried for several weeks before moving into cold storage at 35°F. 5 to 16 individuals were selected for each accession, for a total of 1301 individual walnuts to be scanned at Michigan State University (Table S1). The walnuts were scanned in 171 batches. The scans were produced

using the the North Star X3000 system and the included efX-DR software. The X-ray source was set at 75 kV and 100 µA, with 720 projections per scan, at 3 frames per second and with 3 frames averaged per projection. The data was obtained in continuous mode. The 3D X-ray CT reconstruction was computed with the efX-CT software, obtaining voxel-based images with voxel size of 75.9 µm.

All the individual walnuts were manually separated with ImageJ (Figure 1A). Densities were rescaled so that all scans share similar air, kernel, and shell density values. Once densities were comparable across samples, the external air and other debris was removed through thresholding and mathematical morphology operations (Figure 1B). Rough estimates for the location of shell, air, kernel, and packing tissues were obtained based on density, connectedness, and object thickness information. These estimates were used then to fully segment the tissues using a watershed segmentation algorithm (Falcao et al., 2004) (Figure 1C-H). We took particular care of tissue labeled as shell, where we distinguished voxels close to the walnut surface, to voxels protruding into the internal cavity (Figure 1D). Some of the scanned walnuts contained incomplete or no kernel at all. These were discarded from further morphological analysis, leaving us with a total of 1264 individual walnuts representing 149 accessions (Table S1). All the image processing above was done automatically with in-house, scipy-based, python scripts.

To make some measurements comparable, all the walnuts were centered on their centers of mass and rotated such that the lateral plane goes through the walnut seal, and the shell tip is the rightmost point of the longitudinal plane (Figure 1I.) The same center and rotation was immediately applied to the kernel. By taking a series of 2D transverse slices across the proximal-distal axis, we approximated the main cavity surrounded by the kernel, located a the proximal side of the walnut, between the two main hemispheres of the kernel (Figure 1J).

2.2 Walnut morphological trait measurements

For each individual we computed the same 14 morphological traits as in Bernard et al. (2020): nut length, height, width, total surface area, total volume, rugosity, sphericity, shape VA3D, equancy, shell volume, shell thickness, kernel volume, kernel volume filling ratio, and the inner air volume. We computed an additional collection of 35 morphological traits for a total of 49 measurements per sample. This includes the volume of packing tissue, as well as the percentage of air, kernel, shell, and packing tissue volume with respect to the total nut volume —the sum of

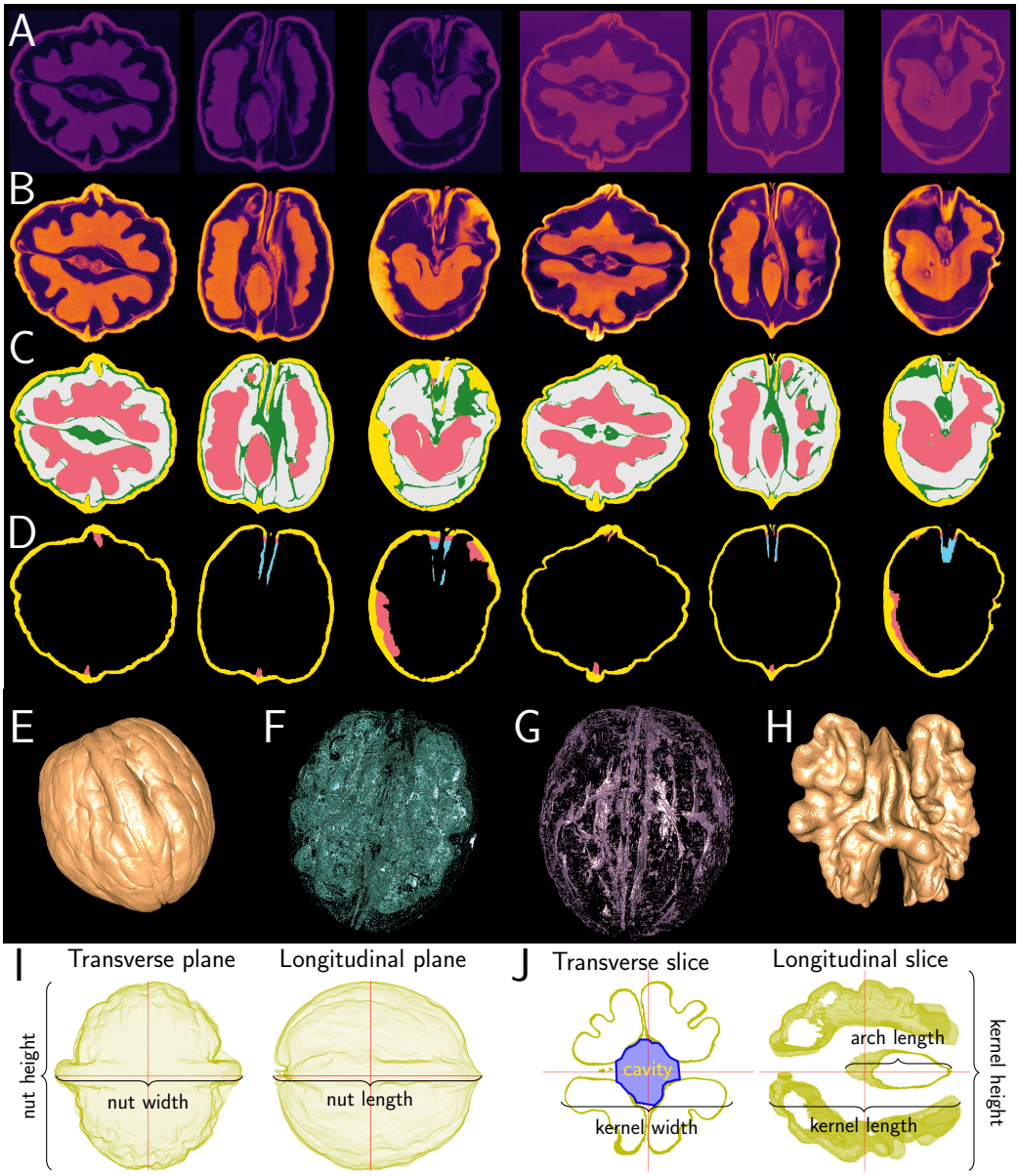


FIGURE 1 Walnut scanning, image processing, and phenotyping. (A) Raw scans of individual walnuts seen from different planes. (B) Densities were standardized across all samples and the external air removed. (C) Shell, air, kernel, and packing tissue were automatically labeled with a combination of basic image morphology operations and watershed segmentation. (D) The tissue labeled as shell was further broken down into external and protruding tissue. (E) 3D renders of shell, (F) air, (G) packing tissue, and (H) kernel. (I) All the walnuts were centered on their center of mass and aligned. (J) The same centering and alignment was applied to the kernels. All the figures above are for illustration purposes only and are not scaled.

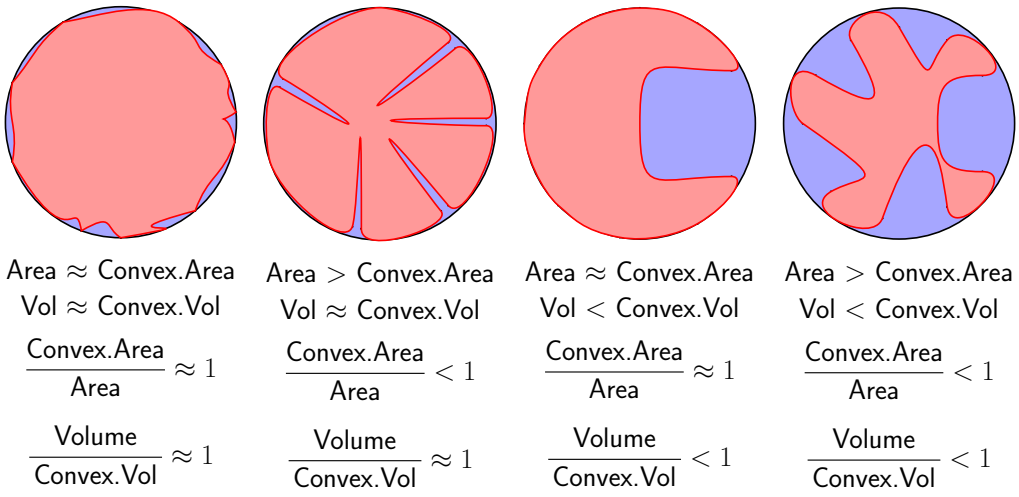


FIGURE 2 Comparing the surface area and volume of an object to the ones from its convex hull can reveal different kinds of lobeyness. For illustrative purposes, a 2D example is presented below, with a red object and its blue convex hull. As this is a 2D example instead of a 3D one, area refers to perimeter, while volume refers to area instead. Both convex area and volume ratios are bounded between 0 and 1, where 1 indicates perfectly convex shape. A low convex area ratio with a higher convex volume ratio suggests deep, narrow troughs across the object's surface. The opposite case suggests wider, square-like cavities.

air, kernel, shell, and packing tissue volumes. We also considered the airless volume ratio —where now the total nut volume is just the sum of kernel, shell, and packing tissue volumes. Sphericity was calculated with the Krumbein and Sneed indices for every walnut —indices between 0 and 1, where 1 indicates a perfect sphere (Blott and Pye, 2008). The surface area and volume of the nut's convex hull was compared to the actual nut surface area and volume respectively as a proxy for different kinds of lobeyness —the ratios are between 0 and 1, where 1 indicates a perfectly convex shape (Figure 2). These computations were repeated for the kernel to assess its sphericity and lobeyness as well. For the main cavity at the proximal side of the kernel, we measured its depth, surface area, and volume. We computed the percentage of cavity area and volume comprised by the total kernel surface area and kernel's convex hull volume respectively as a proxy for relative size of the cavity. We also measured the length of the arch-like structure of the kernel that bridges the two main hemispheres. As indicated before, for the shell and tissue with shell-like density, we computed both the volume and the percentage of it protruding into the main walnut cavity. Finally, we computed the average density value for shell, kernel, and packing tissues. Since the X-ray CT scans only measure relative density, we considered the three possible ratios of these averages as a proxy for absolute density. (Table 1).

TABLE 1 Morphological traits measured for each walnut based on its X-ray CT scan. A dash denotes not applicable. A dagger (†) denotes a trait that has not been measured before to the best of our knowledge.

Trait	Formula	Description	Unit
Whole walnut			
Length	L_w	Distance from base to tip	mm
Width	W_w	Longest distance across the seal	mm
Height	H_w	Longest distance perpendicular to the seal	mm
Surface	A_w	Surface area of the actual nut	mm^2
Convex surface†	A_{cw}	Surface area of the convex hull of the nut	mm^2
Volume	V_w	Total volume of the actual nut, including air	mm^3
Volume w/o air†	$\hat{V}_w$	Volume of the nut if we ignore the air contained	mm^3
Convex volume†	V_{cw}	Volume of the convex hull of the nut	mm^3
VA3D	$A_w^3/(36\pi V_w^2)$	Shape factor	-
Feret Ratio	L_w/H_w	Inverse index of roundness	-
Krumbein†	$\sqrt[3]{W_w H_w/L_w^2}$	Index of roundness	-
Sneed†	$\sqrt[3]{H_w^2/W_w L_w}$	Index of roundness	-
Sphericity	$\sqrt[3]{36\pi V_w^2}/A_w$	Wadell's index of roundness	-
Rugosity	$1/\text{sphericity}$	Index of surface roughness	-
Convex area ratio†	A_{cw}/A_w	Index of nonconvexity	-
Convex volume ratio†	V_w/V_{cw}	Index of nonconvexity	-
Shell			
Total Volume	V_s	Total volume of the shell	mm^3
Protruding Volume†	V_p	Volume of shell-like bits protruding into the walnut	mm^3
Thickness	T_s	Average thickness of the external section of the shell	mm
Volume ratio	V_s/V_w	Percentage of shell with respect to the whole walnut	%
Volume ratio w/o air†	$V_s/\hat{V}_w$	Percentage of shell with respect to the airless walnut	%
Protruding ratio†	V_p/V_s	Percentage of shell-like tissue protruding into the walnut	%
Kernel			
Length	L_k	Longest distance perpendicular to the transverse plane	mm
Width	W_k	Longest distance perpendicular to the longitudinal plane	mm
Height	H_k	Longest distance perpendicular to the lateral plane	mm
Surface†	A_k	Surface area of the actual kernel	mm^2
Convex surface†	A_{ck}	Surface area of the convex hull of the kernel	mm^2
Volume	V_k	Total volume of the actual kernel	mm^3
Convex volume†	V_{ck}	Volume of the convex hull of the kernel	mm^3
Arc length†	a_k	Length of the tissue connecting both hemispheres	mm
Cavity length†	c_k	Length of the main cavity at the proximal end	mm
Cavity surface†	s_k	Surface area of the cavity at proximal end	mm^2
Cavity volume†	v_k	Volume of the main cavity at the proximal end	mm^3
Volume ratio	V_k/V_w	Percentage of kernel with respect to the whole walnut	%
Volume ratio w/o air†	$V_k/\hat{V}_w$	Percentage of kernel with respect to the airless walnut	%
Feret ratio†	L_k/H_k	Inverse index of roundness	-
Sphericity†	$\sqrt[3]{36\pi V_k^2}/A_k$	Wadell's index of roundness	-
Convex area ratio†	A_{ck}/A_k	Index of nonconvexity	-
Convex volume ratio†	V_k/V_{ck}	Index of nonconvexity	-
Cavity area ratio†	s_k/A_k	Percentage of area represented by the main cavity	%
Cavity volume ratio†	v_k/V_{ck}	Percentage of volume represented by the main cavity	%
Density vs shell	ρ_k/ρ_s	Average density relative to shell density	-
Packing tissue			
Volume†	V_t	Total volume of the packing tissue	mm^3
Volume ratio†	V_t/V_w	Percentage of packing tissue with respect to the whole walnut	%
Volume ratio w/o air†	$V_t/\hat{V}_w$	Percentage of tissue with respect to the airless walnut	%
Density vs shell†	ρ_t/ρ_s	Average density relative to shell density	-
Density vs kernel†	ρ_t/ρ_k	Average density relative to kernel density	-
Air			
Volume	V_a	Total volume of the air contained by the nut	mm^3
Volume ratio†	V_a/V_w	Percentage of air with respect to the whole walnut	%

TABLE 2 Traits of breeding interest. Measured and evaluated per IPGRI (1994) guidelines. Dash indicates not applicable.

Trait	Range	Description	Unit
Blank ratio	0–100	Percentage of nuts with blank kernel	%
Ease of Removal	4–8	How easy is to remove both kernel halves intact	-
Kernel Fill	3–8	If kernel fills poorly or well the walnut's interior	-
Kernel Plumpness	2–8	If kernel is thin or plump	-
Kernel Weight Ratio	0–100	Kernel-to-nut weight ratio	%
Packing Tissue Thickness	4–8	If the tissue is sparse or very thick	-
Shell Color	4–8	How light or dark is the color	-
Shell Integrity	4–8	How complete is the shell	-
Shell Seal	4–8	If the seal is weak/open or very strong	-
Shell Strength	4–8	If shell is papery or strong	-
Shell Texture	3-8	How smooth or rough is the shell	-
Shrivel—major	0–100	Percentage of kernels exhibiting >50% shrivel	%
Shrivel—minor	0–100	Percentage of kernels exhibiting <50% shrivel	%
Shrivel—tip	0–100	Percentage of kernels exhibiting tip shrivel	%

2.3 Measurement of additional traits of breeding interest

Separately from the walnuts scanned at Michigan State University, accessions were evaluated for 14 traits of breeding interest (Table 2). Ten walnuts per accession were collected across multiple years and locations managed by the Walnut Improvement Program at the University of California, Davis. These walnuts were later cracked open from each sample using a hammer at Davis, California. The assessment and scoring was done following the IPGRI (1994) guidelines. All the trait scores were averaged across all the samples of the same accession to obtain scores representative of said accession.

2.4 Allometric relationships

We measured the variability of each morphological phenotype by computing their quartile coefficient of variation (QCD) across the 1264 scans. We preferred the QCD as it only depends on the 25th and 75th quartiles, making it robust against outliers compared to the coefficient of variation (CV) (Bonett, 2006). We studied allometric relationships between all size-specific morphological traits —the relative growth of one feature with respect to another one. Since different plant tissues grow relative to each other following a power law rather than a simple linear relationship (Niklas, 2004; West et al., 1999), we plotted size-related data in log-log plots and calculated the best fit line (Figure 3). The slope of this line indicates whether a tissue grows at a faster or slower pace than another, while the intercept allows to compute theoretical limits on tissue size.

2.5 Accession summaries and correlations

We averaged each morphological phenotype for walnuts within the same accession so we could compare them to the accession-based traits of breeding interest. Since size-related phenotypes follow nonlinear relationships, we favored the computation of Spearman rather Pearson correlation coefficients between all the different morphological and breeding phenotypes (Figures 4, S1, S2). After applying a Bonferroni correction, all p-values smaller than 0.01 were deemed highly significant.

2.6 Prediction of traits of breeding interest

A stepwise linear regression was performed to determine the most relevant morphological phenotypes to predict ease of removal. We only considered phenotypes significantly correlated to ease of removal as potential explanatory variables. Data was standardized to zero mean and unit variance. A train-test split was followed to avoid overfitting results. That is, an ordinary least squares (OLS) linear model was fitted using 70% of the data while predictions of ease of removal were made for the remaining 30%. The R^2 coefficient of determination was computed between the predicted and actual ease of removal values as a measure of prediction accuracy. The 70-30 training-test split was randomized and repeated 100 times and the average R^2 coefficient was taken as the overall accuracy score. To determine the contribution of individual phenotypes to the prediction of ease of removal, we computed numerous linear models where we varied the number and the specific phenotypes used as explanatory variables. First we computed models using a single explanatory phenotype and recorded the model with the highest accuracy. Second, we considered models using all the possible pairs of phenotypes as explanatory variables and recorded the pair that provided the model with the highest accuracy. We subsequently repeated this exploration of models using all possible combinations of three to six phenotypes as explanatory variables (Figure 5A). The above procedure was repeated to identify the morphological features that are the most relevant to predict shell strength and kernel weight ratio (Figure 5B-C). For all the models described above, we noticed that prediction results improved considerably if the Earliest Himalayan accession (UCACCSD 85-023-2) was discarded (Figure S3). All the R^2 scores reported are thus considering only 148 of the 149 accessions scanned.

Finally, we performed a principal component analysis (PCA) of the most predictive morphological phenotypes. We then plotted the two principal components and observed how

these also reflect the traits of breeding interest. The PCA was repeated using all 49 morphological phenotypes as well.

3 Results

3.1 Distribution and variability of phenotypes

Values of nut length, width, height, sphericity, convexity, and density were overall stable with low QCD values (0.05 or less). This suggests that walnuts by and large have similar overall shell shape, rugosity, and shell lobeyness. A high nut convex volume ratio (0.95±0.01) combined with a low convex surface area ratio (0.63±0.01) indicates that most of the shell' surfaces are covered by numerous, narrow, deep grooves. The relative average densities are also stable (0.03), where the reported average density across the whole kernel and packing tissue is 86% and 61% of the shell density respectively. This is contrasted by higher variability of tissue volume (0.15), which agrees with reports of highly variable walnut weight values (Cosmulescu and Stefanescu, 2018; Rashnodi et al., 2019). At the same time, walnuts reveal a especially large variability of shell tissue. In particular, the amount of shell tissue that protrudes into the walnut ranges from 12 to 1077 mm^3. This corresponds to a QCD of almost 0.4, where the 75th percentile (237 mm^3) is more than double of the 25th one (106 mm^3). There is similar variability when it comes to all the measurements related to the main proximal-side cavity (Table 3).

3.2 Allometry and possible biophysical constraints

We observe that most of the size-specific traits follow power laws with respect to the total nut volume V_w, as our allometric log-log plots exhibit large R^2 coefficients of determination, most of them above 0.5 (Figure 3). The size-related measurement that exhibits the most superlinear growth rate is the total air volume contained inside the nut V_a. Our data suggests that these two volumes follow the power law $V_a \approx \exp(-3.17)V_w^{1.22}$. That is, as the nut volume increases, biophysical constraints require an air volume increase by a larger factor. However, the air volume must always be lower than the total nut volume. Evaluating the extreme case of a hypothetical walnut consisting entirely of air, we find that $V_w = \exp(-3.17)V_w^{1.22}$ when $V_w \approx 2.3 \times 10^6$ mm^3. This is the same volume of a 16cm diameter sphere. The volume of the convex hull of the nut V_{cw} also follows a superlinear growth rate with respect to V_w, with a power law $V_{cw} = \exp(-0.16)V_w^{1.02}$. Do notice that if $V_w < 2063\text{mm}^3$, then $V_{cw} < V_w$, which is impossible.

TABLE 3 Morphological trait values. Standard deviation, 25th quartile, 75th quartile, quartile coefficient of dispersion, and coefficient of variance are indicated by SD, $\mathbf{Q_{25}}$, $\mathbf{Q_{75}}$, QCD, and CV respectively. Traits sorted by QCD. A dash denotes not applicable.

Trait	Units	Mean + SD	Range	$\mathbf{Q_{25}}$	$\mathbf{Q_{75}}$	CV	QCD
Nut Ratio Convex Volume	%	0.96 ± 0.01	0.86 — 0.98	0.95	0.97	0.02	0.01
Nut Ratio Convex Area	%	0.63 ± 0.01	0.49 — 0.66	0.63	0.64	0.02	0.01
Shell Rugosity	-	1.66± 0.05	1.56 — 2.30	1.62	1.69	0.03	0.02
Nut Sphericity	-	0.60± 0.02	0.44 — 0.64	0.59	0.62	0.03	0.02
Density Kernel vs Shell	%	0.86 ± 0.04	0.79 — 1.09	0.84	0.88	0.04	0.02
Sneed Index	-	0.92± 0.03	0.83 — 1.00	0.90	0.94	0.03	0.02
Density Packing vs Kernel	%	0.71 ± 0.03	0.61 — 0.85	0.68	0.73	0.05	0.03
Density Packing vs Shell	%	0.61 ± 0.04	0.54 — 0.77	0.58	0.63	0.06	0.04
Krumbein Index	-	0.90± 0.05	0.75 — 1.00	0.87	0.93	0.05	0.04
Kernel Sphericity	-	0.25± 0.02	0.20 — 0.36	0.24	0.26	0.06	0.04
Kernel Ratio Convex Area	%	0.38 ± 0.02	0.30 — 0.53	0.36	0.40	0.06	0.04
Nut Feret Ratio	-	1.21± 0.10	1.00 — 1.56	1.14	1.27	0.08	0.05
Kernel Ratio Convex Volume	%	0.56 ± 0.05	0.38 — 0.69	0.54	0.60	0.09	0.05
Nut Width	mm	32.0 ±2.75	23.1 — 41.0	30.2	33.7	0.09	0.06
Kernel Feret Ratio	-	1.25± 0.11	1.01 — 1.70	1.18	1.31	0.08	0.06
Nut Height	mm	33.4 ±2.96	25.6 — 44.1	31.4	35.2	0.09	0.06
Nut VA3D	-	4.59± 0.50	3.78 — 12.1	4.27	4.83	0.11	0.06
Kernel Height	mm	28.3 ±2.75	20.8 — 40.1	26.5	29.9	0.10	0.06
Kernel Length	mm	30.4 ±2.96	18.2 — 40.4	28.4	32.4	0.10	0.07
Kernel Width	mm	24.8 ±2.64	15.7 — 34.0	23.0	26.3	0.11	0.07
Nut Length	mm	38.5 ±4.07	26.3 — 53.2	35.7	41.2	0.11	0.07
Packing Ratio	%	0.13 ± 0.02	0.07 — 0.23	0.11	0.14	0.15	0.08
Kernel Ratio	%	0.34 ± 0.05	0.18 — 0.48	0.31	0.37	0.15	0.09
Kernel Arc Height	mm	11.7 ±2.26	0.10 — 19.6	10.7	12.8	0.19	0.09
Kernel Cavity Height	mm	15.3 ±2.83	4.92 — 36.8	13.7	16.8	0.18	0.10
Kernel Ratio Cavity Area	%	0.03 ± 0.01	0.01 — 0.07	0.03	0.03	0.20	0.11
Nut Convex Surface Area	mm^2	3674 ± 609	2057 — 6067	3257	4044	0.17	0.11
Kernel Convex Area	mm^2	2572± 427	1277 — 3934	2281	2847	0.17	0.11
Nut Surface Area	mm^2	5798± 1010	3316 — 9978	5116	6400	0.17	0.11
Air Ratio	%	0.37± 0.07	0.14 — 0.61	0.32	0.41	0.18	0.12
Shell Thickness	mm	0.87 ± 0.16	0.50 — 1.61	0.76	0.97	0.19	0.12
Kernel Surface Area	mm^2	6773 ± 1276	2827 — 10991	5905	7600	0.19	0.13
Shell Ratio	%	0.16± 0.03	0.09 — 0.33	0.14	0.18	0.20	0.13
Nut Volume	mm^3	19560± 4758	7905 — 41132	16300	22235	0.24	0.15
Kernel Ratio Cavity Volume	%	0.02 ± 0.01	0.00 — 0.06	0.02	0.03	0.25	0.15
Kernel Volume	mm^3	6565±1589	2087 — 12232	5542	7650	0.24	0.16
Nut Convex Volume	mm^3	20511± 5102	8470 — 43346	16932	23441	0.25	0.16
Packing Volume	mm^3	2475± 676	954 — 6321	2032	2814	0.27	0.16
Kernel Convex Volume	mm^3	11680± 2900	4007 — 21761	9710	13501	0.25	0.16
Kernel Cavity Area	mm^2	201± 53.9	39.9 — 476	165	232	0.27	0.17
Shell Volume	mm^3	3188± 886	1139 — 7446	2568	3724	0.28	0.18
Air Volume	mm^3	7332± 2628	2126 — 21812	5520	8681	0.36	0.22
Kernel Cavity Volume	mm^3	288± 114	12.1 — 836	207	347	0.40	0.25
Protruding Shell Ratio	%	0.06 ± 0.03	0.01 — 0.19	0.04	0.07	0.46	0.30
Protruding Shell Volume	mm^3	187 ± 118	12.1 — 1077	106	237	0.63	0.38

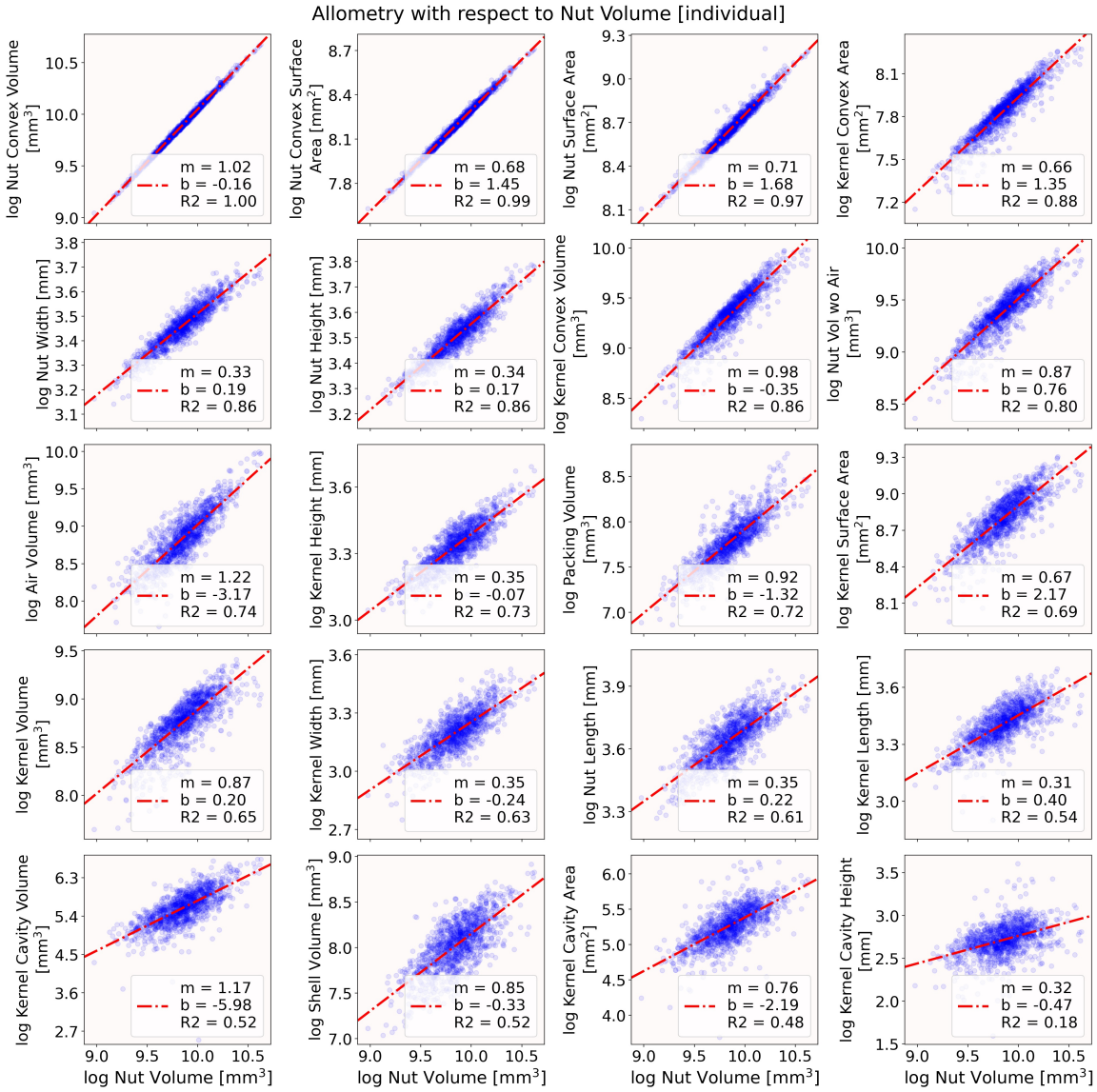


FIGURE 3 Various allometry plots between different logarithmic values of tissue volumes, areas, and lengths compared to the total walnut volume. An ordinary least squares linear model was computed for each case. The slope, intercept, and coefficient of determination for each linear model is indicated by m, b, and R^2 respectively.

In other words, the allometric power law only holds for walnuts comparable to a 1.6cm diameter sphere. This agrees with the fact that the walnut seed is very lobed during early developmental stages (Sartorius and Stösser, 1997). Of important note is the fact that kernel volume grows at a slightly sublinear rate with respect to total nut volume, with a power law $V_k = \exp(0.19)V_w^{0.87}$. For instance, if the total nut volume is duplicated, then the kernel volume will only increase by a factor of $2^{0.87} \approx 1.8$, which already indicates that larger walnuts tend to have smaller kernel percentages, while they also tend to contain higher air percentages.

3.3 Correlation between several morphological phenotypes

The last observation is also supported by a significant negative Spearman correlation between the kernel and air volume ratios (-0.59), and nut volume (-0.27). (Figures 4). If we ignore the air when taking into account volume ratios, we also observe very negative correlations with shell volume, shell volume ratios, and shell thickness (between -0.46 and -0.88). For shell thickness, we observe unsurprisingly high correlations with shell volume percentage (0.93), shell total volume (0.79), and other shell-related measurements. There are negative correlations with kernel density relative to shell density (-0.83) and with air percentage (-0.61) (Figure 4). Whenever air was discarded from the volume ratio considerations, we observed that the ratios of shell, kernel, and packing tissue were significantly and negatively correlated between each other. This suggests that increasing one ratio diminishes the other two, and that these three ratios can be manipulated independently. On the other hand, whenever air was included into the volume ratio considerations, we observed no significant correlations among percentages of shell, kernel and packing tissues. However, these three were all negatively correlated to the air ratio. This suggests that the air ratio determines the volume ratio for the rest of tissues. Another interesting observation is that the volume ratio of shell and packing tissue when the air is excluded is negatively (-0.30) and positively (0.27) correlated respectively to the volume ratio of air. This might indicate that increasing the content of air also increases the content of packing tissue at the expense of shell (Figure 4). There is also significant correlation between convexity and sphericity indices (Figure S1). Finally, we note that air ratio is positively correlated (0.54) with kernel convex area ratio while negatively correlated (-0.81) with kernel convex volume ratio, which suggests that walnuts with large air content tend to have large and wide troughs (Figure S2A). However, there was no strong correlation between air ratio and the main cavity at the proximal side of the kernel (Figure S2A).

3.4 Correlation between morphology and traits of breeding interest

Ease of kernel removal is negatively correlated to air (-0.40) and kernel (-0.33) volume ratios, and to kernel weight (-0.41) and density (-0.33) ratios. Ease of removal is positively correlated to shell (0.43) and packing tissue (0.33) volume ratios, and to shell thickness (0.40). The shell strength unsurprisingly is positively correlated with shell volume ratio (0.73) and shell thickness (0.71). It is also negatively correlated to the volume (-0.56), weight (-0.74), and density (-0.58) ratios of the kernel, and to the air volume ratio (-0.60). Finally, we observe that weight ratio of the kernel is negatively correlated with shell volume ratio (-0.82) and thickness (-0.81). We highlight that there is no correlation (0.07) between kernel weight and kernel volume ratios whenever the latter takes air volume into consideration. The correlation improves significantly (0.78) when the volume ratio does not include air, but this is not a one-to-one relationship between kernel volume and weight. This could be explained by observing also a significant correlation (0.65) between relative kernel weight and density, suggesting that as kernel increases in size, so does in density (Figure 4). In other words, nuts with a higher relative content of air tend to have weaker shells, more relative kernel weight, and the kernel is easier to extract.

3.5 Modeling and explaining certain traits of breeding interest

Whenever we limited an OLS linear model to use a single explanatory variable to predict ease of kernel removal scores, the best possible prediction on average was achieved by using air volume ratio. The model in this case reported an average R^2 coefficient between predicted scores versus true scores was 0.26 when predicting the same data used to train the model, and 0.18 when using test data instead. We observe that through the 100 repetitions, the R^2 coefficient is highly variable, which indicates that the model is quite sensitive to the initial train-test split. This large variability remains present regardless of the number of explanatory variables used by the model, which reflects that the ease of removal of scores are not normally distributed for the accessions at hand. If the OLS model is allowed to use two explanatory variables, the average R^2 coefficient improves to 0.40 and 0.29 when predicting the train and test data respectively when using shell and packing tissue volume ratios. A slight improvement is observed when allowing the model to use three explanatory variables, with mean R^2 scores of 0.42 and 0.30 for predicting training and testing data respectively. The best trio consists again of shell and packing tissue volume ratio, and relative kernel density. Finally, we observe that adding more explanatory variables to

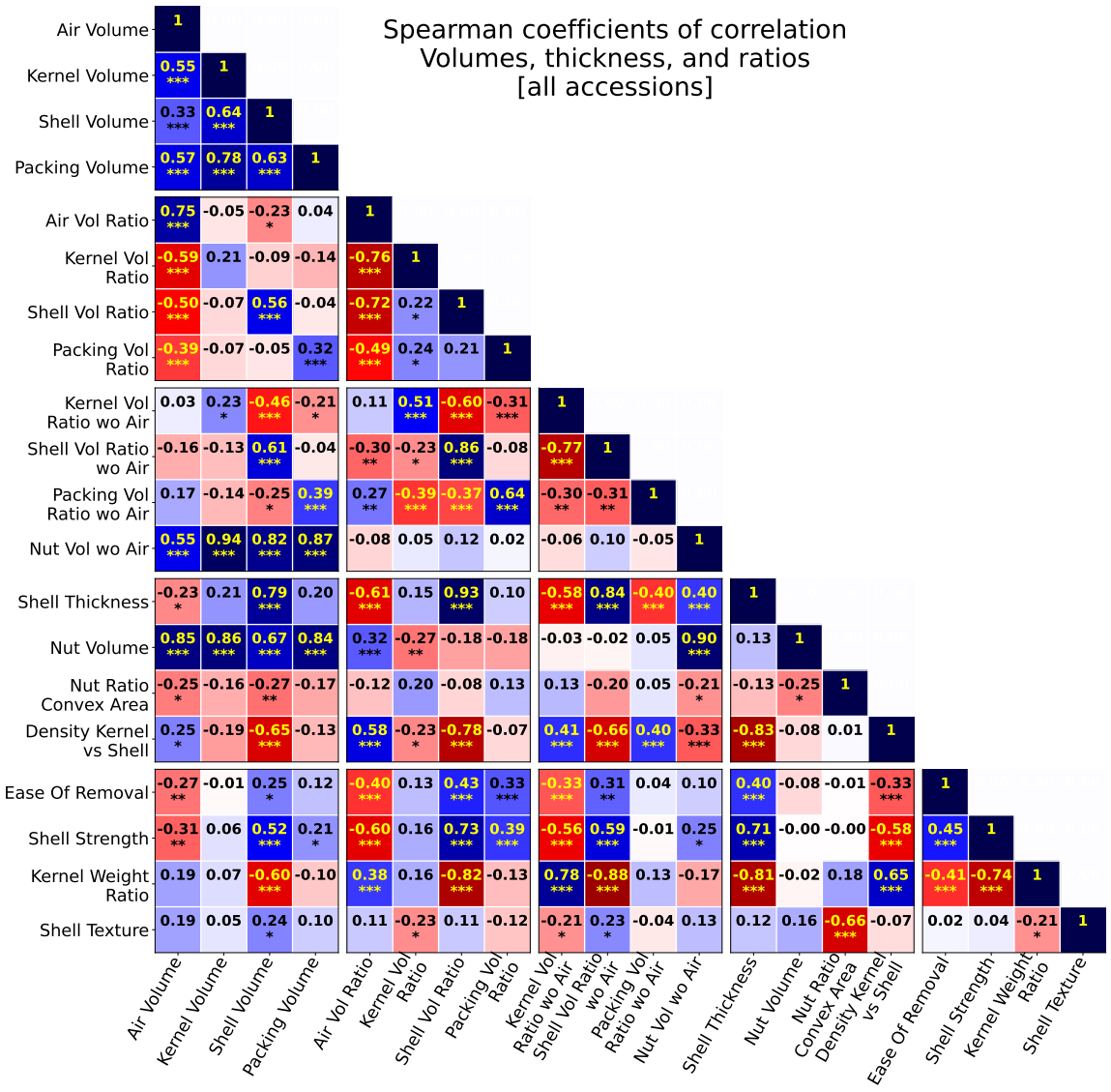


FIGURE 4 Spearman correlation coefficients between shape, size, and traits of breeding interest. The correlation was computed using phenotype values that were averaged across all individuals of the same accession. Three stars (***) denote an associated p-value of 10^{-4} or smaller. Two stars denote a p-value between 10^{-4} and 10^{-3}. One star denotes it between 10^{-3} and 10^{-2}.

the OLS does not improve the results. Moreover, the variability of R^2 coefficients increases for the test data, which indicates that the model tends to overfit as more variables are taken into account (Figure 5A-B).

A similar analysis was performed on OLS linear models to predict shell strength, with shell volume ratio being the most predictive single trait. This produced an R^2 coefficient of 0.53 and 0.49 if predicting with training and test data respectively. The variability of R^2 coefficients is much lower than in the prior modeling of ease of kernel removal. The model can improve its R^2 up to 0.66 and 0.62 respectively if it uses two explanatory variables instead, namely shell thickness and packing tissue volume ratio. The model shows no improvements, and actually performance worsens, when more explanatory variables are added to it, which indicates overfitting (Figure 5C-D). To predict kernel weight ratio, using simply the shell volume airless ratio, the model reports an average R^2 score of 0.75 and 0.73 with training and testing data respectively. The variability of R^2 is even lower than in the previous two models. This performances improves up to 0.84 and 0.82 if the model uses two explanatory variables instead. The best pair consists of shell thickness and kernel airless volume ratio. No improvements are reported whenever the model employs more explanatory variables (Figure 5E-F).

Finally, we perfomed a PCA using only the 6 most predictive morphological phenotypes discussed above: shell thickness, shell and packing tissue volume ratio, and shell and kernel airless volume ratio. The first two PCs explain more than 85% of the variance. We observe that accessions follow a gradient based on their scores of traits of breeding interest. If all the 49 morphological phenotypes are considered instead, the first two PCs now only explain 50% of the total variance. As with the OLS models, this suggests that more phenotypes tend to muddle the overall picture. Similarly, the gradient of breeding values is less clear. We highlight that the Earliest Himalayan accession (85-023-2) has a very small kernel which is notoriously hard to extract. Nonetheless, it is morphologically average, even when limited to the most predictive phenotypes, as it lies close to the center of the PCA (Figure 6). This averageness is also observed when examining the z-scores of all morphological phenotypes for every accession (Figure S5).

4 Discussion

There is plenty of observed phenotypic diversity within a fixed walnut population, but surprisingly these variations are usually not sufficient to distinguish geographically distant populations (Roor

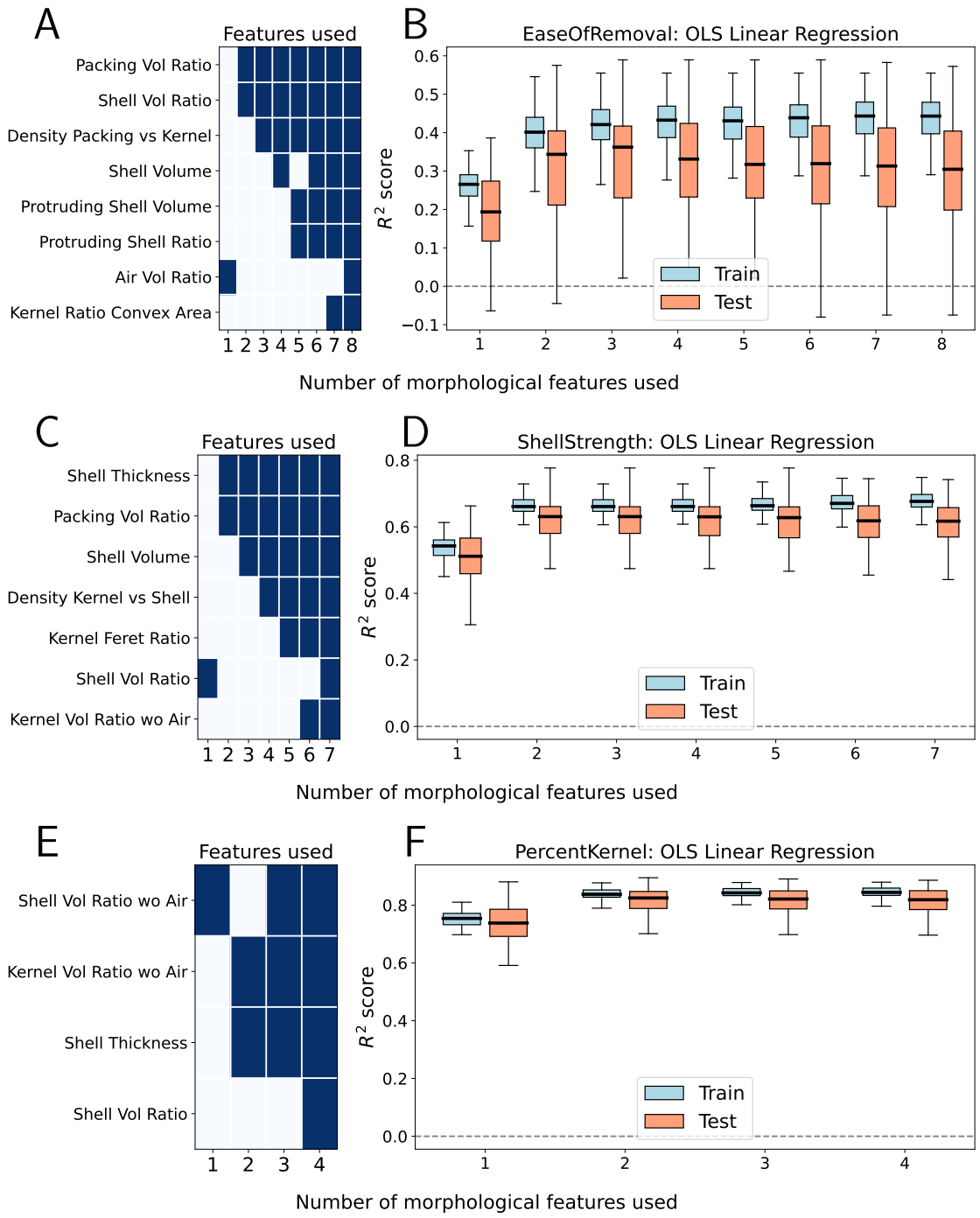


FIGURE 5 Morphological traits that can predict the best certain traits of breeding interest using an OLS linear regression. (A) If the model is limited to a single explanatory variable, then air volume ratio produces the best possible prediction of ease of kernel removal. (B) The R^2 coefficient between predicted versus the true ease of removal value is 0.2. This can improve up to 0.32 if the OLS model uses two explanatory variables instead, these two being shell and packing tissue volume ratio. A moderate improvement is made by considering three explanatory variable, adding relative kernel density. No improvement is made by adding more explanatory variables. (C) - (F) Analogous analyses were made with models predicting shell strength, relative kernel weight.

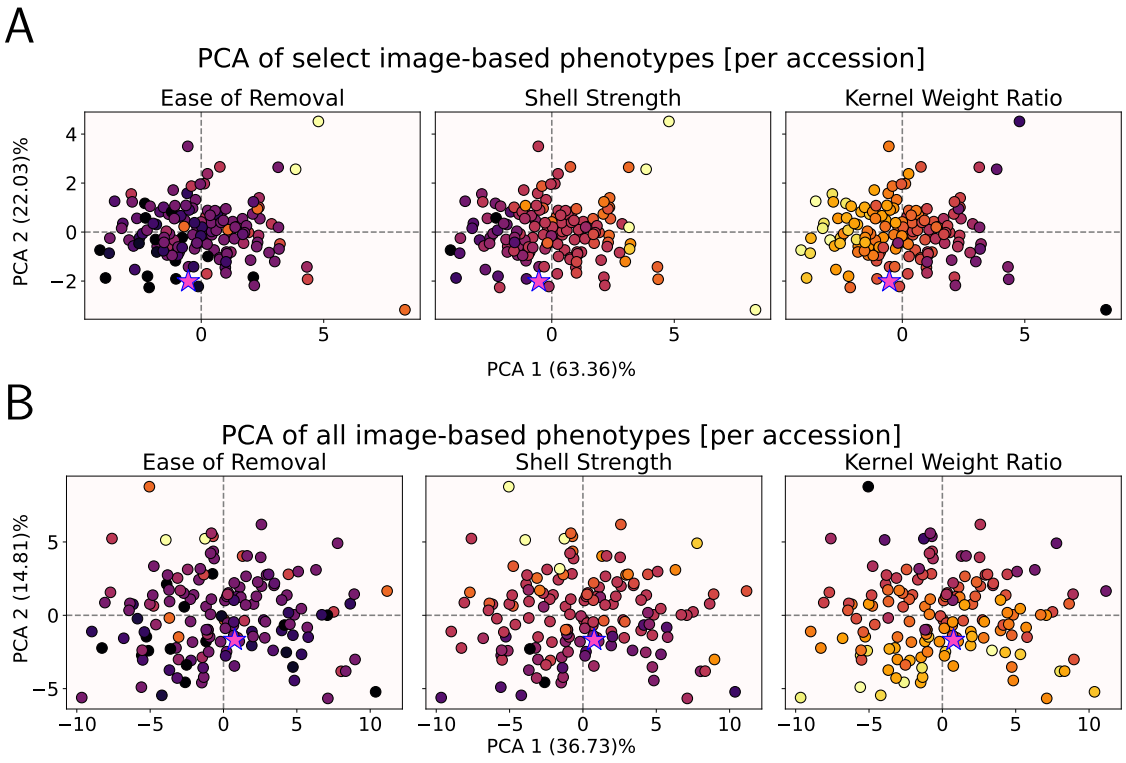


FIGURE 6 Principal Component Analysis using solely morphological phenotypes. The scatter plot is the same for each row, however the color varies depending on the trait of interest indicated on top. Darker colors represent lower values. The Earliest accession (85-023-2) is represented by a magenta star instead. (A) The PCA uses only the 6 most predictive morphological phenotypes as suggested by Figure 5. (B) The PCA uses all 49 morphological phenotypes from Table 3.

et al., 2017). This difficulty to comprehensively measure walnut morphology is more pressing when trying to understand the fine-grained details that determine important traits of commercial concern (Bernard et al., 2021). Three such traits are the ratio of kernel-to-total walnut weight, the kernel ease of removal —how easy is to remove its main two halves intact—, and shell strength. Multiple-pronged strategies have been proposed and developed to unravel the underlying mechanisms that regulate shell thickness, shape, and strength. Through genome-wide association studies (GWAS), transcription factors and pathways that affect the shell and seal formation have been identified (Sideli et al., 2020; Wang et al., 2022). Walnut shell physics have been explored with numerical simulations, where walnuts are modeled as thin spheres and biophysical mechanical properties are tested under unidirectional loads based on finite-element analyses (Bao et al., 2022; Koyuncu et al., 2004). Novel work has focused on the shape of the polylobate sclerid individual cells that tesselate and conform the walnut shell while forming intricate puzzles that confer remarkable toughness and strength (Antreich et al., 2019; Zhang et al., 2014). These tough puzzles in turn are determined by biochemical processes during walnut shell development, as individual cells go from a soft to hard state (Antreich et al., 2021).

X-ray CT scans allows us to accurately extract more nuanced shape and size features from our sampled walnuts, providing new avenues to explore subtle morphological changes and its implications. For example, with more size-related features, we can compute better allometric relationships that point to biophysical constraints in walnut growth development. For example, the growth of empty space within a walnut outpaces the overall nut growth rate, while the kernel growth rate falters (Figure 3). This indicates that larger walnuts tend to contain a higher proportion of air and a lower proportion of kernel, which agrees with past observations of inner air increase during walnut development Xiao et al. (2020). At the same time, we observed that for small nuts, their total volume was almost identical to the volume of their convex hulls. This implies that smooth, groove-free nuts must be small. Moreover, this allometric relationship only holds for nut that are larger than a certain size (16mm in diameter), which indicates that the growth dynamics of the nut undergo a regime change as the nut develops (Pinney and Polito, 1983; Zhao et al., 2016). All of the allometric observations and correlations above suggest that walnut and kernel sizes and smoothness are not just dependent on genes and environment, but there are also unavoidable biophysical constraints at play that should be explored further and considered by breeding programs (Niklas and Hammond, 2019).

Our extended list of measured phenotypes also offers new insight to qualitatively assessed traits. The biophysical constraints around air content appear to also play a role for several traits of interest. Walnuts with relatively low air content tend to be smaller in volume, have thicker and stronger shells, more packing tissue, the kernels are denser and harder to extract, with a lower weight ratio. The effects of air content can also affect correlation coefficients for other tissues. For example, if we consider air as part of the total walnut volume, then the relative volume of packing tissue is positively correlated to shell strength and ease of kernel removal. This agrees with reports that walnuts with thick packing tissue tend to be difficult to open (Kouhi et al., 2020; Mirmahdi and Khadivi, 2021; Sarikhani Khorami et al., 2014). These correlations vanish if we disregard air as part of the total walnut volume. However, new correlations emerge, as the airless ratio of packing tissue volume is now negatively correlated to shell thickness, volume, and density. Unlike packing tissue, the shell volume appears to be more independent from inner air content. Regardless if we consider air as part of the total walnut volume or not, we observe that walnuts with large shell volume ratio tend to have thicker, stronger, and denser shells, a low ratio of kernel volume and weight, and make kernel removal more difficult. This agrees with reports that thin shells allow easier kernel extraction (Amiri et al., 2010; Arzani et al., 2008; Fallah et al., 2022). Moreover, inner air may play a key role in walnut development. Examining the relative air volume compared to airless relative content of shell and packing tissue, the correlations suggest that as a walnut grows, packing tissue develops at the expense of shell tissue. This could be explained by the fact that the shell undergoes a careful biochemical balance between insulation and permeability of air and water as the walnut develops (Antreich et al., 2022) (Figure 4).

This nuanced relationship between air, shell, packing tissue during shell development is further highlighted when we consider their individual influence over ease of kernel removal. The single trait that can best explain and predict ease of removal is relative air volume. However, when the model is allowed to use any two morphological traits, the best results are produced when considering relative shell and packing tissue volumes, as opposed to relative air volume and something else. From a statistical point of view, the pair of shell-packing tissue volume ratios makes sense: both traits are completely uncorrelated and both are strongly correlated to air relative volume (Figure 4). This suggests that the influence of inner air into the ease of removal score can be understood as a combination of the relative shell and packing tissue volume influence (Figure 5A). We can make a similar argument for shell thickness and relative packing tissue volume encoding the same information as relative shell volume when it comes

to understand each of these traits influence over shell strength (Figure 5C). We can also argue that the relationship between the kernel weight ratio and the shell airless volume ratio is already encoded by the kernel airless volume ratio and shell thickness. The inner air content, and the air ratio in a walnut are traits that often overlooked, yet they seem to play crucial roles in walnut development and appear to be intrinsically linked to traits of high commercial interest. Moreover, we can successfully encode commercially relevant scores using only volume- and thickness-based morphological traits (Figure 6). This opens a potential avenue for new phenotyping platforms focusing on just volumetric measurements.

Nonetheless, the exact link between broad morphological phenotypes and commercially relevant traits remains elusive. We highlight the Earliest accession (85-023-2) is morphologically average when compared to the general population as its corresponding z-scores are small for most of phenotypes (Figure S5D). Yet, its shell is notoriously strong and it is extremely difficult to extract its kernel intact. This accession is originary from the Himalaya region, which is one of the hotspots for walnut diversity (Aradhya et al., 2017) and the only wild accession that formed part of our study. This in turn poses exciting questions on morphological changes during walnut domestication.

Walnuts offer a especially unique opportunity to analyze domestication in perennial crops. Despite their long history with humans, current research suggests that walnut domestication happened less than 100 years ago (Mapelli et al., 2018). Even today, due to economical and horticultural reasons, walnut is mainly propagated via seeds instead of grafting throughout most of Southwest Asia (Rezaee et al., 2008; Thapa et al., 2021). This makes walnut an exciting organism to study the immediate effects of domestication and breeding in real time across multiple populations. We can especially focus on inner air, kernel, shell, and packing tissue volumes, as these can successfully encode other commercial traits. While X-ray CT scanning offers a powerful and precise way to quantify these volumes, there are inexpensive alternatives if we just want to measure volume. Kernel, shell, and packing tissue volumes can be measured via toluene (C_7H_8) displacement (Aydin, 2003; Gharibzahedi et al., 2012). Inner air volume can be estimated following the Archimedean' principle and a vacuum flask partially filled with a surfactant solution (Marquard, 1989; Raskin, 1983). This is an exciting opportunity to design affordable phenotyping platforms. A careful, nuanced study of walnut morphology combined with extended low-cost phenotyping might provide us key insights into other domestication-induced

phenotypical and genotypical changes, and accelerate the selection of progenitors in breeding programs.

Acknowledgements

Daniel Chitwood is supported by the USDA National Institute of Food and Agriculture, and by Michigan State University AgBioResearch. The work of Elizabeth Munch is supported in part by the National Science Foundation through grants CCF-1907591, CCF-2106578, and CCF-2142713.

Author contributions

EA, PB, EM, and DC conceived the experiment. PB selected the accessions to be scanned and collected the plant material, ensuring that walnut accessions were broadly represented. MQ and DC collected the digital data. EA developed the necessary scripts to process the scans and extracted their shape descriptors. EA analyzed the data and wrote the manuscript. All authors contributed, reviewed, and revised the manuscript.

Software and data availability

The processed and cleaned walnut X-ray CT 3D reconstructions can be found in the Dryad repository https://doi.org/10.5061/dryad.ngf1vhj09, along with their separated tissues.

All our code in the form of python jupyter notebooks is available at the https://github.com/amezqui3/walnut_tda repository. This includes the image processing pipeline to clean the raw scans and segment the walnut tissues, the computation of all the evaluated phenotypes.

Conflict of interests

None declared

Supplementary Figure Captions

FIGURE S1 Spearman correlation for all phenotypes. Most of the overall nut size-related traits are only positively correlated with kernel size-related traits as expected. All the shell-specific traits are only positively

correlated between themselves. All the sphericity, aspect ratio, and rugosity are highly correlated only among themselves.

FIGURE S2 Detailed Spearman correlation indices for more pairs of traits. Indices have been multiplied by 100. Indices in yellow are highly significant (p-value $>$ 0.001) with an absolute value larger than 0.35. (A) Correlations between select morphological phenotypes with all the 49 traits. (B) Correlations between all the traits of breeding interest and the 49 morphological phenotypes.

FIGURE S3 Ridge regression using all the 49 morphological phenotypes as descriptive variables to predict (A) Ease of Removal, (B) Shell Strength, (C) Kernel-to-total weight ratio, and (D) Kernel fill score. For all cases, we used an $\alpha = 10$ regularization value for the ridge regression. Individual points represent individual accessions. Red diamonds represent predicted values when considering all 149 accessions, while blue circles are predicted values when excluding the Himalayan Earliest accession (UCACCSD 85-023-2). The R^2 coefficient of determination is indicated at the bottom of each plot. Notice that excluding the Earliest accession improves considerably the R^2 score.

FIGURE S4 The values were centered at 0 and rescaled to unit variance. A normal Gaussian bell is drawn in red for comparison. The Fisher-corrected kurtosis (k) and skewness (s) are indicated for each trait distribution. For reference, a normal distribution has both $k = 0$ and $s = 0$.

FIGURE S5 Variability and uniqueness of morphological traits for every accession. For every measured phenotype, we computed the mean and QCD values for individuals of the same accession. (A) Since most of the morphological phenotypes follow a normal distribution (Figure S4), the z-score of these means was computed. A large z-score would suggest that the accession shows a very different phenotypical value compared to the rest of the population. We were particularly interested to explore the trait distribution for accessions associated with the lowest and highest ease of kernel removal scores. (B) The QCD of the traits for individuals in the same accession quantifies how variable is every shape and size parameter within individuals of the same accession. (C) Mann-Whitney U-tests were performed to determine if certain morphological traits were sufficiently different to tell apart one accession from the rest of the population. (D) The mean absolute z-score provides an overall estimate on how distinctive are the morphological features of an accession. Accessions with the lowest ease of kernel removal scores are depicted in blue, while highest scores are in red and orange. (E) Similar to (C), the average QCD provides an overall estimate on how variable is the shape of individuals within the same accession.

References

Akca Y, Şen SM (1995). The relationship between dichogamy and yield–nut characteristics in *Juglans regia* L. *In Acta Horticulturae*, Volume 442, pp. 215–216. Leuven, Belgium: International Society for Horticultural Science (ISHS). doi:10.17660/ActaHortic.1997.442.31.

Amézquita EJ, Quigley MY, Ophelders T, Landis JB, Koenig D, Munch E, Chitwood DH (2021). Measuring hidden phenotype: quantifying the shape of barley seeds using the Euler characteristic transform. *in silico Plants* **4**(1): diab033. doi:10.1093/insilicoplants/diab033.

Amiri R, Vahdati K, Mohsenipoor S, Mozaffari MR, Leslie C (2010). Correlations between some horticultural traits in walnut. *HortScience* **45**(11): 1690–1694. doi:10.21273/HORTSCI.45.11.1690.

Antreich SJ, Huss JC, Xiao N, Singh A, Gierlinger N (2022). The walnut shell network: 3D visualisation of symplastic and apoplastic transport routes in sclerenchyma tissue. *Planta* **256**(3): 49. doi:10.1007/s00425-022-03960-w.

Antreich SJ, Xiao N, Huss JC, Gierlinger N (2021). A belt for the cell: cellulosic wall thickenings and their role in morphogenesis of the 3D puzzle cells in walnut shells. *Journal of Experimental Botany* **72**(13): 4744–4756. doi:10.1093/jxb/erab197.

Antreich SJ, Xiao N, Huss JC, Horbelt N, Eder M, Weinkamer R, Gierlinger N (2019). The puzzle of the walnut shell: A novel cell type with interlocked packing. *Advanced Science* **6**(16): 1900644. doi:10.1002/advs.201900644.

Aradhya M, Velasco D, Ibrahimov Z, Toktoraliev B, Maghradze D, Musayev M, Bobokashvili Z, Preece JE (2017). Genetic and ecological insights into glacial refugia of walnut (*Juglans regia* L.). *PLOS ONE* **12**(10): 1–27. doi:10.1371/journal.pone.0185974.

Araus JL, Cairns JE (2014). Field high-throughput phenotyping: the new crop breeding frontier. *Trends in Plant Science* **19**(1): 52–61. doi:10.1016/j.tplants.2013.09.008.

Arzani K, Mansouri-Ardakan H, Vezvaei A, Roozban MR (2008). Morphological variation among Persian walnut (*Juglans regia*) genotypes from central Iran. *New Zealand Journal of Crop and Horticultural Science* **36**(3): 159–168. doi:10.1080/01140670809510232.

Aydin C (2003). Physical properties of almond nut and kernel. *Journal of Food Engineering* **60**(3): 315–320. doi:10.1016/S0260-8774(03)00053-0.

Bao X, Chen B, Dai P, Li Y, Mao J (2022). Construction and verification of spherical thin shell model for revealing walnut shell crack initiation and expansion mechanism. *Agriculture* **12**(9). doi:10.3390/agriculture12091446.

Bernard A, Crabier J, Donkpegan ASL, Marrano A, Lheureux F, Dirlewanger E (2021). Genome-Wide Association Study Reveals Candidate Genes Involved in Fruit Trait Variation in Persian Walnut (*Juglans regia* L.). *Frontiers in Plant Science* **11**. doi:10.3389/fpls.2020.607213.

Bernard A, Hamdy S, Le Corre L, Dirlewanger E, Lheureux F (2020). 3D characterization of walnut morphological traits using X-ray computed tomography. *Plant Methods* **16**(1): 115. doi:10.1186/s13007-020-00657-7.

Blott SJ, Pye K (2008). Particle shape: a review and new methods of characterization and classification. *Sedimentology* **55**(1): 31–63. doi:10.1111/j.1365-3091.2007.00892.x.

Bonett DG (2006). Confidence interval for a coefficient of quartile variation. *Computational Statistics & Data Analysis* **50**(11): 2953–2957. doi:10.1016/j.csda.2005.05.007.

Bucksch A, Atta-Boateng A, Azihou AF, Battogtokh D, Baumgartner A, Binder BM, Braybrook SA, Chang C, Coneva V, DeWitt TJ et al. (2017). Morphological plant modeling: Unleashing geometric and topological potential within the plant sciences. *Frontiers in Plant Science* **8**. doi:10.3389/fpls.2017.00900.

Cosmulescu S, Stefanescu D (2018). Morphological variation among Persian walnut (*Juglans regia*) genotypes within the population and depending on climatic year. *Scientia Horticulturae* **242**: 20–24. doi:10.1016/j.scienta.2018.07.018.

Falcao AX, Stolfi J, de Alencar Lotufo R (2004). The image foresting transform: theory, algorithms, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **26**(1): 19–29. doi:10.1109/TPAMI.2004.1261076.

Fallah M, Vahdati K, Hasani D, Rasouli M, Sarikhani S (2022). Breeding of Persian walnut: Aiming to introduce late-leafing and early-harvesting varieties by targeted hybridization. *Scientia Horticulturae* **295**: 110885. doi:10.1016/j.scienta.2022.110885.

FAS (2022). Tree nuts: World markets and trade. Technical report, Foreign Agricultural Service. US Department of Agriculture, Washington, DC. URL: https://apps.fas.usda.gov/psdonline/circulars/TreeNuts.pdf.

Fiorani F, Schurr U (2013). Future scenarios for plant phenotyping. *Annual Review of Plant Biology* **64**(1): 267–291. doi:10.1146/annurev-arplant-050312-120137.

Gao T, Zhang S, Sun H, Ren R (2022). Mass detection of walnut based on X-ray imaging technology. *Journal of Food Process Engineering* **45**(8): e14034. doi:10.1111/jfpe.14034.

Gao Z (2022). Three-dimensional virtual flower bud of walnut tree based on micro-computed tomography. *Agronomy Journal* **114**(4): 1935–1943. doi:10.1002/agj2.20931.

Gharibzahedi SMT, Mousavi SM, Hamedi M, Khodaiyan F (2012). Comparative analysis of new Persian walnut cultivars: nut/kernel geometrical, gravimetrical, frictional and mechanical attributes and kernel chemical composition. *Scientia Horticulturae* **135**: 202–209. doi:10.1016/j.scienta.2011.11.030.

Hilal N, Mohammed Ali TK, Tayeh BA (2020). Properties of environmental concrete that contains crushed walnut shell as partial replacement for aggregates. *Arabian Journal of Geosciences* **13**(16): 812. doi:10.1007/s12517-020-05733-9.

Huss JC, Antreich SJ, Bachmayr J, Xiao N, Eder M, Konnerth J, Gierlinger N (2020). Topological interlocking and geometric stiffening as complementary strategies for strong plant shells. *Advanced Materials* **32**(48): 2004519. doi:doi.org/10.1002/adma.202004519.

IPGRI (1994). *Descriptors for Walnut (*Juglans *Spp.)*. Rome, Italy: International Plant Genetic Resources Institute.

Kelc D, Štampar F, Solar A (2007). Fruiting behaviour of walnut trees influences the relationship between the morphometric traits of the parent wood and nut weight. *The Journal of Horticultural Science and Biotechnology* **82**(3): 439–445. doi:10.1080/14620316.2007.11512256.

Khadivi-Khub A, Ebrahimi A, Sheibani F, Esmaeili A (2015). Phenological and pomological characterization of Persian walnut to select promising trees. *Euphytica* **205**(2): 557–567. doi:10.1007/s10681-015-1429-9.

Kouhi M, Rezaei A, Hassani D, Sarikhani S, Vahdati K (2020). Phenotypic Evaluation and Identification of Superior Persian Walnut (*Juglans regia* L.) Genotypes in Mazandaran Province, Iran. *Journal of Nuts* **11**(4): 315–326. doi:10.22034/jon.2021.1919541.1104.

Koyuncu MA, Ekinci K, Savran E (2004). Cracking characteristics of walnut. *Biosystems Engineering* **87**(3): 305–311. doi:10.1016/j.biosystemseng.2003.11.001.

Lala SD, Deoghare AB, Chatterjee S (2018). Mechanical and morphological characterization of walnut shell reinforced epoxy composite. *IOP Conference Series: Materials Science and Engineering* **377**(1): 012011. doi:10.1088/1757-899X/377/1/012011.

Li M, Klein LL, Duncan KE, Jiang N, Chitwood DH, Londo JP, Miller AJ, Topp CN (2019). Characterizing 3D inflorescence architecture in grapevine using X-ray imaging and advanced morphometrics: implications for understanding cluster density. *Journal of Experimental Botany* **70**(21): 6261–6276. doi:10.1093/jxb/erz394.

Li M, Shao MR, Zeng D, Ju T, Kellogg EA, Topp CN (2020). Comprehensive 3D phenotyping reveals continuous morphological variation across genetically diverse sorghum inflorescences. *New Phytologist* **226**(6): 1873–1885. doi:10.1111/nph.16533.

Lopez JM (2004). Walnut tissue culture: research and field applications. *In* Michler CH, Pijut PM, Van Sambeek JW, Coggeshall MV, Seifert J, Woeste K, Overton R, Ponder FJ (Eds.), *Black walnut in a new century*, 6th Walnut Council Research Symposium, pp. 146–152. St. Paul, MN: USDA, North Central Research Station.

Mapelli S, Pollegioni P, Woeste KE, Chiocchini F, Lungo SD, Olimpieri I, Tortolano V, Clark J, Hemery GE, Malvolti ME (2018). Spatial genetic structure of common walnut (*Juglans regia* L.) in central Asia. *In Acta Horticulturae*, Number 1190, pp. 27–34. International Society for Horticultural Science (ISHS), Leuven, Belgium. doi:10.17660/ActaHortic.2018.1190.5.

Marquard RD (1989). A rapid method for determination of nut volume of pecan. *HortScience* **24**(3): 520. doi:10.21273/HORTSCI.24.3.520.

Mirmahdi NS, Khadivi A (2021). Identification of the promising Persian walnut (*Juglans regia* L.) genotypes among seedling-originated trees. *Food Science & Nutrition* **9**(4): 2217–2226. doi:10.1002/fsn3.2193.

NASS (2022). 2022 California Walnut. Objective Measurement Report. Technical report, National Agricultural Statistics Service. US Department of Agriculture, Sacramento, CA. URL: https://www.nass.usda.gov/Statistics_by_State/California/Publications/Specialty_and_Other_Releases/Walnut/index.php.

606 **Niklas KJ** (2004). Plant allometry: is there a grand unifying theory? *Biological Reviews* **79**(4): 871–889.
607 doi:10.1017/S1464793104006499.

608 **Niklas KJ, Hammond ST** (2019). Biophysical effects on the scaling of plant growth, form, and ecology.
609 *Integrative and Comparative Biology* **59**(5): 1312–1323. doi:10.1093/icb/icz028.

610 **Pinney K, Polito VS** (1983). English walnut fruit growth and development. *Scientia Horticulturae* **21**(1): 19–28.
611 doi:10.1016/0304-4238(83)90182-6.

612 **Pollegioni P, Woeste KE, Chiocchini F, Olimpieri I, Tortolano V, Clark J, Hemery GE, Mapelli S, Malvolti**
613 **ME** (2014). Landscape genetics of Persian walnut (*Juglans regia* L.) across its Asian range. *Tree Genetics &*
614 *Genomes* **10**(4): 1027–1043. doi:10.1007/s11295-014-0740-2.

615 **Popa RG, Bălăcescu A, Popescu LG** (2023). Organic Walnut Cultivation in Intensive and Super-Intensive System—
616 Sustainable Investment. Case Study: Gorj County, Romania. *Sustainability* **15**(2). doi:10.3390/su15021244.

617 **Rahaman MM, Chen D, Gillani Z, Klukas C, Chen M** (2015). Advanced phenotyping and phenotype data analysis
618 for the study of plant growth and development. *Frontiers in Plant Science* **6**. doi:10.3389/fpls.2015.00619.

619 **Rashnodi NZ, Moghadam JE, Fazeli A** (2019). The Screening of Persian Walnut Genotypes Based on the
620 Quantitative and Qualitative Characters and the Investigation of Genetic Diversity among Promising Samples
621 Using ISSR Marker. *Plant Productions* **42**(2): 279–294. doi:10.22055/ppd.2019.24861.1569.

622 **Raskin I** (1983). A method for measuring leaf volume, density, thickness, and internal gas volume.
623 *HortScience* **18**(5): 698–699. doi:10.21273/HORTSCI.18.5.698.

624 **Rezaee R, Vahdati K, Grigoorian V, Valizadeh M** (2008). Walnut grafting success and bleeding rate as affected
625 by different grafting methods and seedling vigour. *The Journal of Horticultural Science and Biotechnology* **83**(1):
626 94–99. doi:10.1080/14620316.2008.11512352.

627 **Rezaei Z, Khadivi A, ValizadehKaji B, Abbasifar A** (2018). The selection of superior walnut (*Juglans*
628 *regia* L.) genotypes as revealed by morphological characterization. *Euphytica* **214**(4): 69. doi:10.1007/
629 s10681-018-2153-z.

630 **Roor W, Konrad H, Mamadjanov D, Geburek T** (2017). Population Differentiation in Common Walnut
631 (*Juglans regia* L.) across Major Parts of Its Native Range—Insights from Molecular and Morphometric Data.
632 *Journal of Heredity* **108**(4): 391–404. doi:10.1093/jhered/esw122.

633 **Sarikhani Khorami S, Arzani K, Roozban MR** (2014). Correlations of certain high-heritability horticultural
634 traits in Persian walnut (*Juglans regia* L.). *In Acta Horticulturae*, Number 1050, pp. 61–68. International
635 Society for Horticultural Science (ISHS), Leuven, Belgium. doi:10.17660/ActaHortic.2014.1050.6.

636 **Sartorius R, Stösser R** (1997). On the apomictic seed development in the walnut (*Juglans Regia* L.). *In Acta*
637 *Horticulturae*, Volume 442, pp. 225–229. Leuven, Belgium: International Society for Horticultural Science
638 (ISHS). doi:10.17660/ActaHortic.1997.442.33.

Shah RA, Bakshi P, Sharma N, Jasrotia A, Itoo H, Gupta R, Singh A (2021). Diversity assessment and selection of superior Persian walnut (*Juglans regia* L.) trees of seedling origin from North-Western Himalayan region. *Resources, Environment and Sustainability* **3**: 100015. doi:10.1016/j.resenv.2021.100015.

Sideli GM, Marrano A, Montanari S, Leslie CA, Allen BJ, Cheng H, Brown PJ, Neale DB (2020). Quantitative phenotyping of shell suture strength in walnut (*Juglans regia* L.) enhances precision for detection of QTL and genome-wide association mapping. *PLOS ONE* **15**(4): 1–21. doi:10.1371/journal.pone.0231144.

Solar A, Ivančič A, Štampar F (2003). Morphometric characteristics of fruit bearing shoots in persian walnut (*Juglans regia* l.) — potential selection criteria for breeding. *European Journal of Horticultural Science* **2**.

Thapa R, Thapa P, Ahamad K, Vahdati K (2021). Effect of grafting methods and dates on the graft take rate of Persian walnut in open field condition. *International Journal of Horticultural Science and Technology* **8**(2): 133–147. doi:10.22059/ijhst.2020.311553.401.

Vahdati K (2014). Traditions and folks for walnut growing around the Silk Road. *In Acta Horticulturae*, Volume 1032, pp. 19–24. Leuven, Belgium: International Society for Horticultural Science (ISHS). doi:10.17660/ActaHortic.2014.1032.1.

Wahid M, Gawli Y, Puthusseri D, Kumar A, Shelke MV, Ogale S (2017). Nutty Carbon: Morphology Replicating Hard Carbon from Walnut Shell for Na Ion Battery Anode. *ACS Omega* **2**(7): 3601–3609. doi:10.1021/acsomega.7b00633.

Wang J, Ye H, Zhou H, Chen P, Liu H, Xi R, Wang G, Hou N, Zhao P (2022). Genome-wide association analysis of 101 accessions dissects the genetic basis of shell thickness for genetic improvement in Persian walnut (*Juglans regia* L.). *BMC Plant Biology* **22**(1): 436. doi:10.1186/s12870-022-03824-1.

West GB, Brown JH, Enquist BJ (1999). A general model for the structure and allometry of plant vascular systems. *Nature* **400**(6745): 664–667. doi:10.1038/23251.

Xiao N, Bock P, Antreich SJ, Staedler YM, Schönenberger J, Gierlinger N (2020). From the soft to the hard: Changes in microchemistry during cell wall maturation of walnut shells. *Frontiers in Plant Science* **11**. doi:10.3389/fpls.2020.00466.

Zhang H, Lan H, Tang Y, Li Y (2014). Study on fracture mechanism of walnut shell according to brittle fracture area. *In 2014 Fifth International Conference on Intelligent Systems Design and Engineering Applications*, pp. 954–957. doi:10.1109/ISDEA.2014.211.

Zhao S, Wen J, Wang H, Zhang Z, Li X (2016). Changes in lignin content and activity of related enzymes in the endocarp during the walnut shell development period. *Horticultural Plant Journal* **2**(3): 141–146. doi:10.1016/j.hpj.2016.08.003.

Zhou H, Whalley WR, Hawkesford MJ, Ashton RW, Atkinson B, Atkinson JA, Sturrock CJ, Bennett MJ, Mooney SJ (2020). The interaction between wheat roots and soil pores in structured field soil. *Journal of Experimental Botany* **72**(2): 747–756. doi:10.1093/jxb/eraa475.

673 # Supplementary data

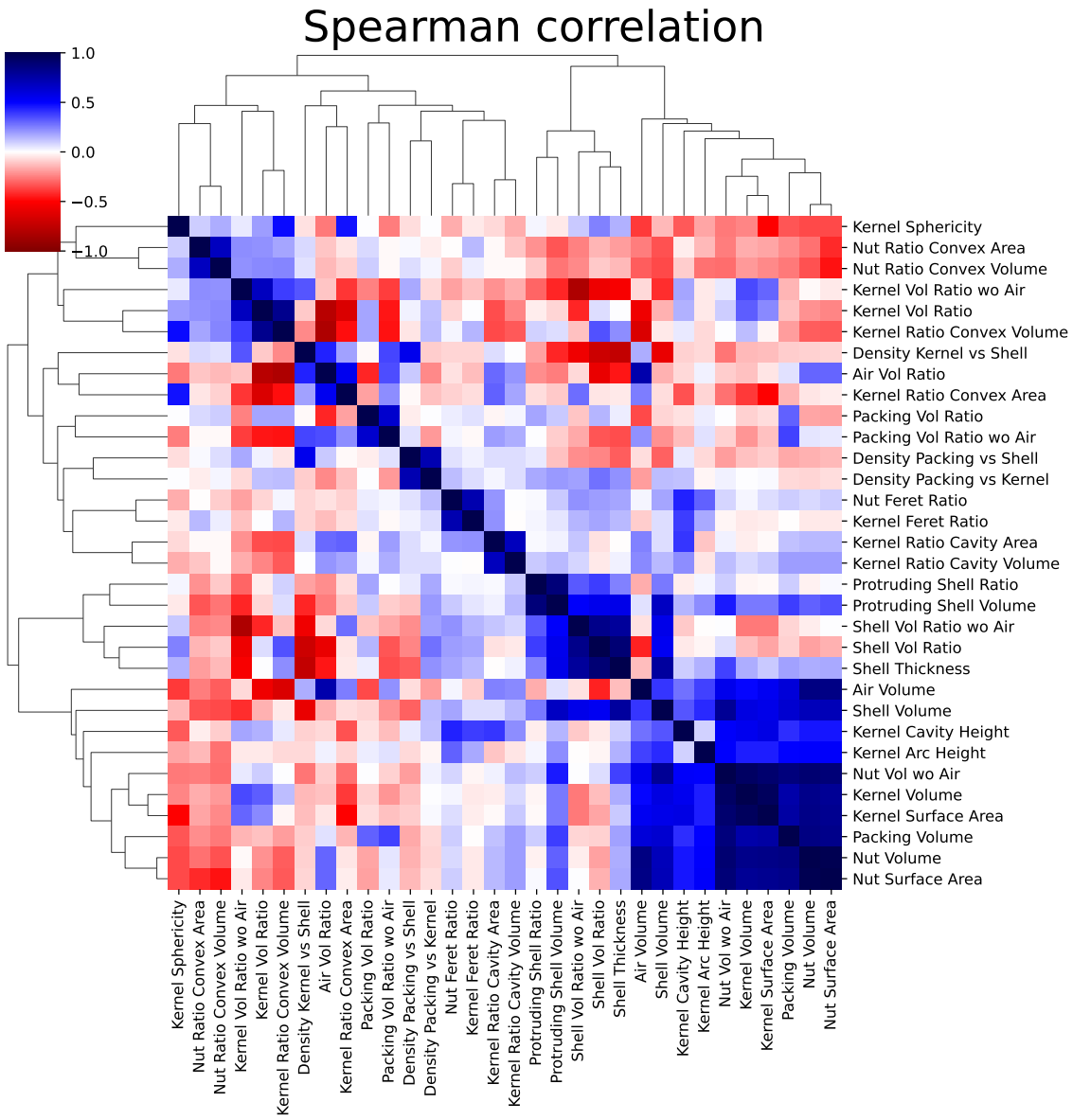


FIGURE S1 Spearman correlation for all phenotypes. Most of the overall nut size-related traits are only positively correlated with kernel size-related traits as expected. All the shell-specific traits are only positively correlated between themselves. All the sphericity, aspect ratio, and rugosity are highly correlated only among themselves.

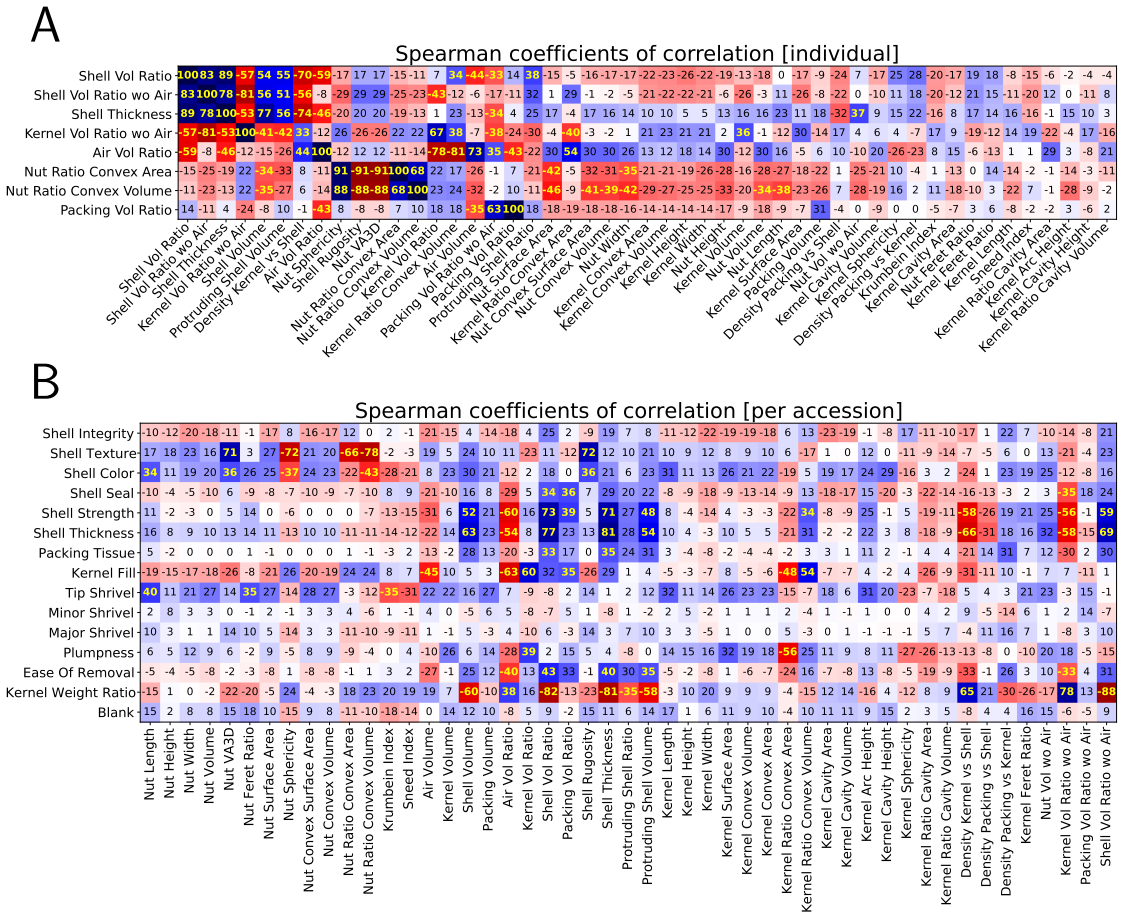


FIGURE S2 Detailed Spearman correlation indices for more pairs of traits. Indices have been multiplied by 100. Indices in yellow are highly significant (p-value > 0.001) with an absolute value larger than 0.35. (A) Correlations between select morphological phenotypes with all the 49 traits. (B) Correlations between all the traits of breeding interest and the 49 morphological phenotypes.

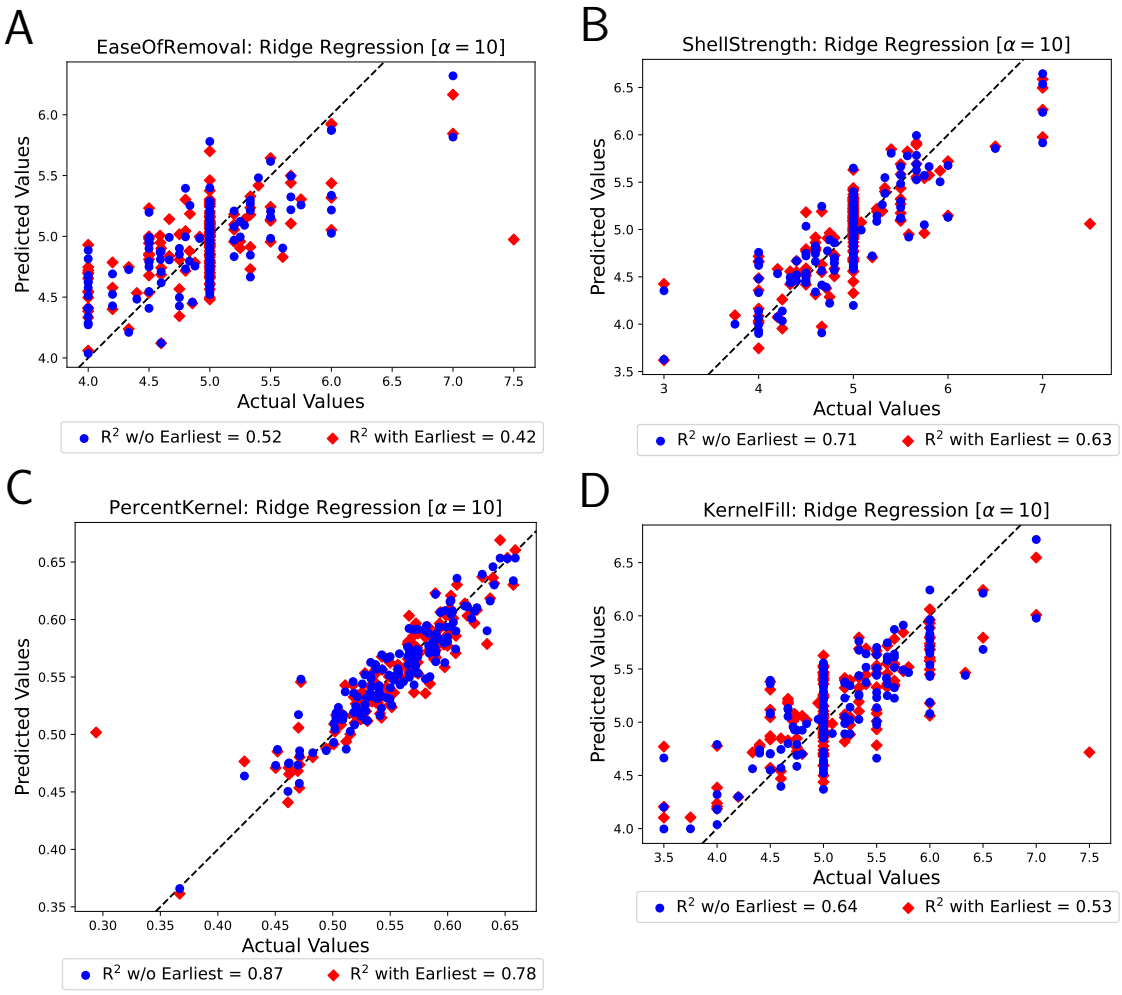


FIGURE S3 Ridge regression using all the 49 morphological phenotypes as descriptive variables to predict (A) Ease of Removal, (B) Shell Strength, (C) Kernel-to-total weight ratio, and (D) Kernel fill score. For all cases, we used an $\alpha = 10$ regularization value for the ridge regression. Individual points represent individual accessions. Red diamonds represent predicted values when considering all 149 accessions, while blue circles are predicted values when excluding the Himalayan Earliest accession (UCACCSD 85-023-2). The R^2 coefficient of determination is indicated at the bottom of each plot. Notice that excluding the Earliest accession improves considerably the R^2 score.

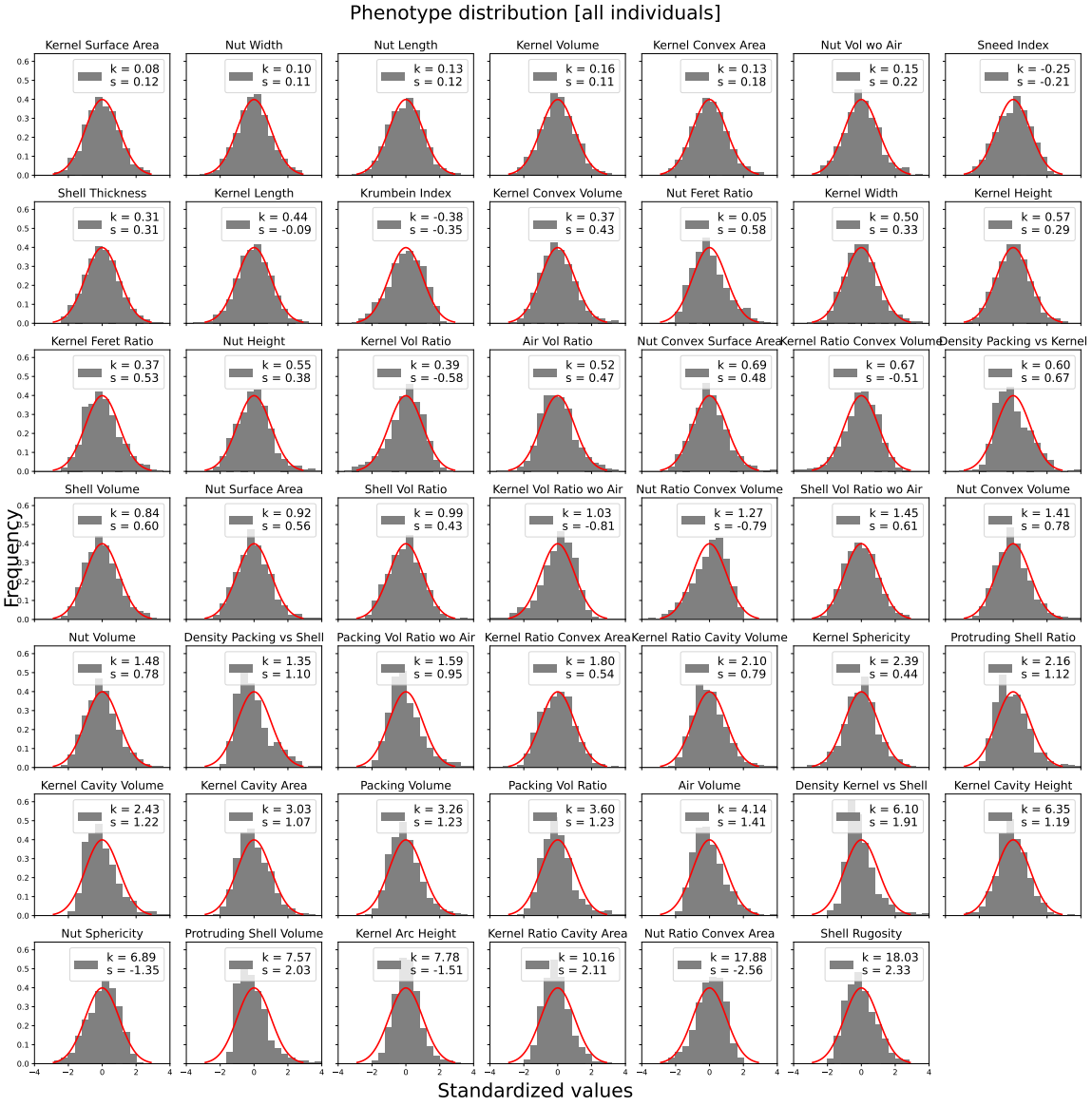


FIGURE S4 Distribution of all the shape related phenotypes described in Table 1. The values were centered at 0 and rescaled to unit variance. A normal Gaussian bell is drawn in red for comparison. The Fisher-corrected kurtosis (k) and skewness (s) are indicated for each trait distribution. For reference, a normal distribution has both $k = 0$ and $s = 0$.

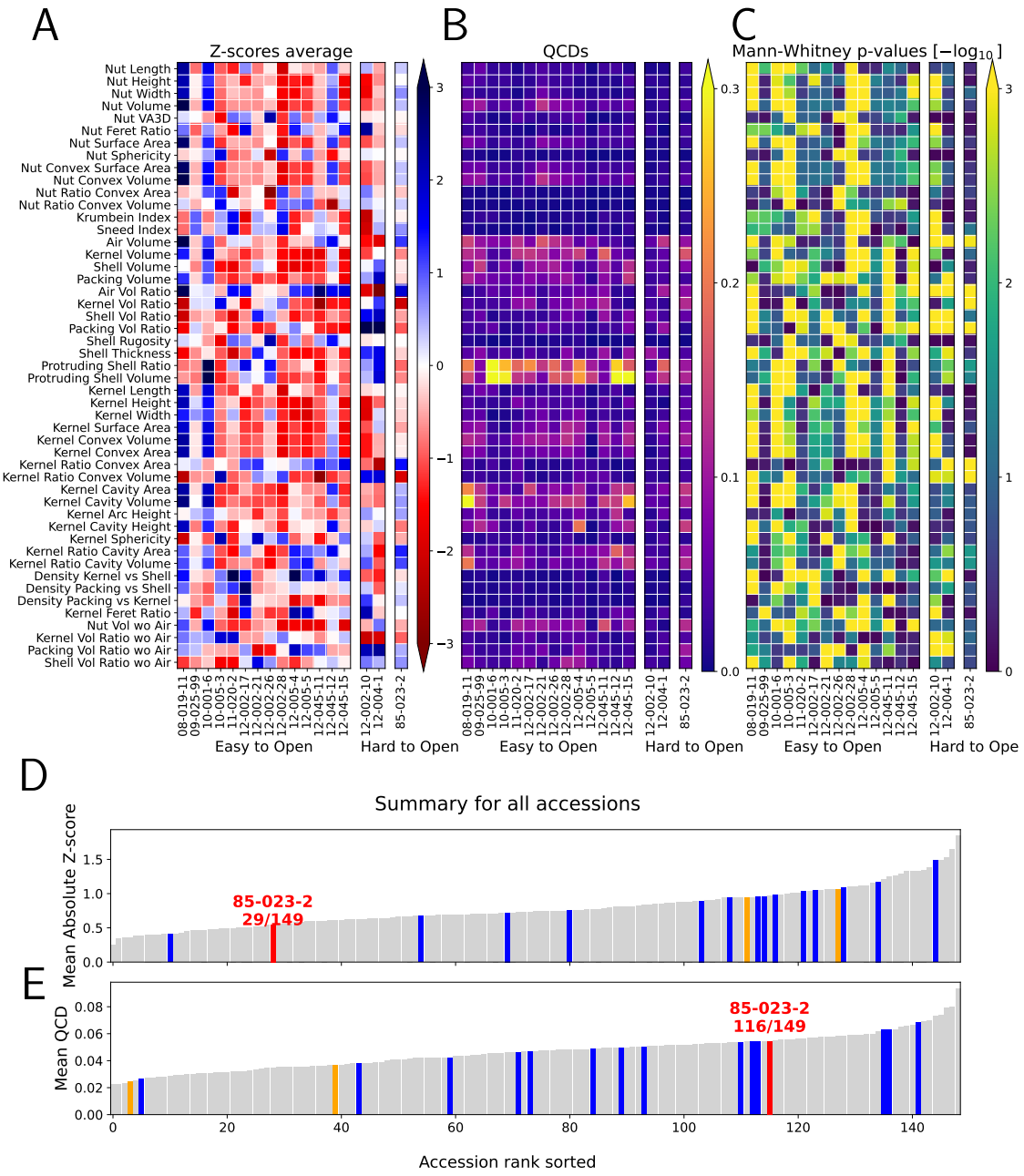
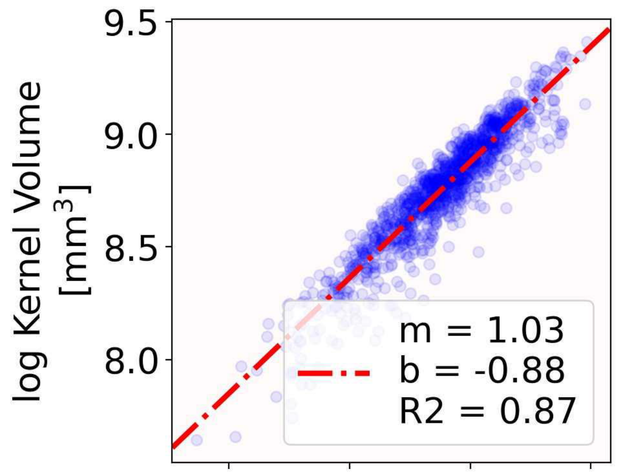
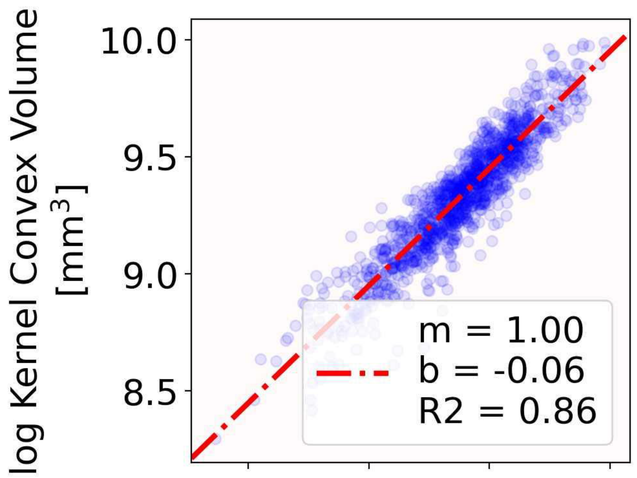
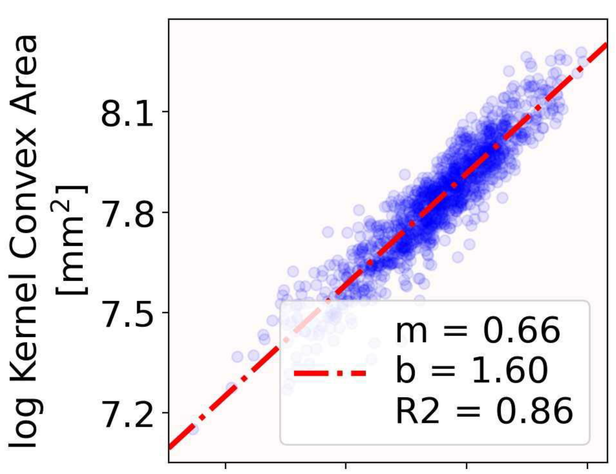


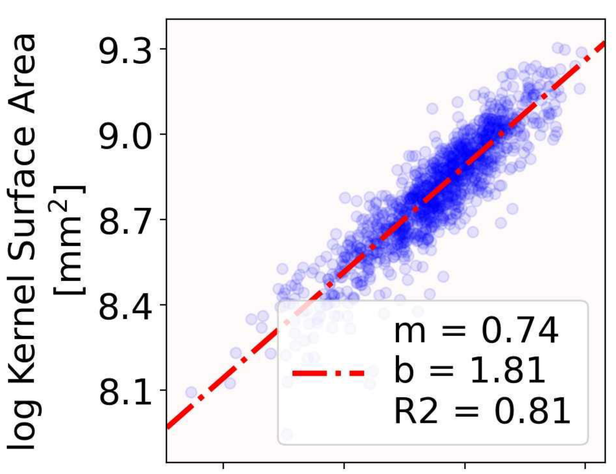
FIGURE S5 Variability and uniqueness of morphological traits for every accession. For every measured phenotype, we computed the mean and QCD values for individuals of the same accession. (A) Since most of the morphological phenotypes follow a normal distribution (Figure S4), the z-score of these means was computed. A large z-score would suggest that the accession shows a very different phenotypical value compared to the rest of the population. We were particularly interested to explore the trait distribution for accessions associated with the lowest and highest ease of kernel removal scores. (B) The QCD of the traits for individuals in the same accession quantifies how variable is every shape and size parameter within individuals of the same accession. (C) Mann-Whitney U-tests were performed to determine if certain morphological traits were sufficiently different to tell apart one accession from the rest of the population. (D) The mean absolute z-score provides an overall estimate on how distinctive are the morphological features of an accession. Accessions with the lowest ease of kernel removal scores are depicted in blue, while highest scores are in red and orange. (E) Similar to (C), the average QCD provides an overall estimate on how variable is the shape of individuals within the same accession.

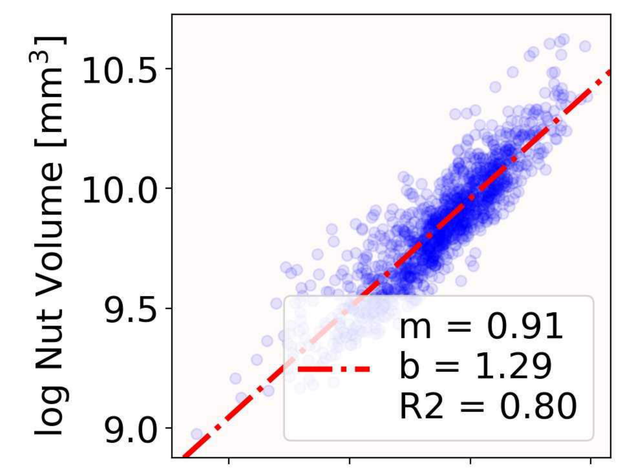
Allometry with respect to Nut Vol wo Air [individual]

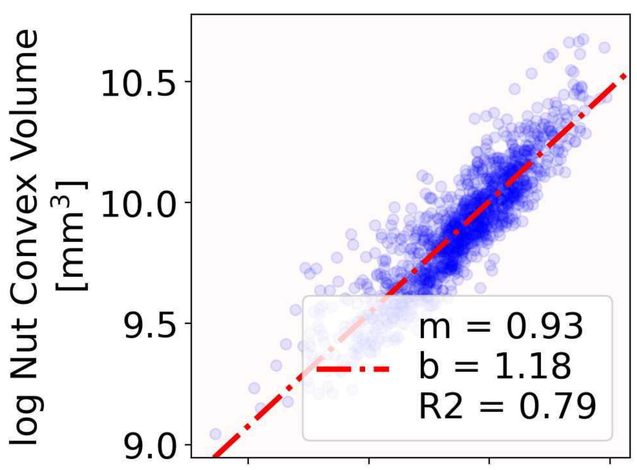


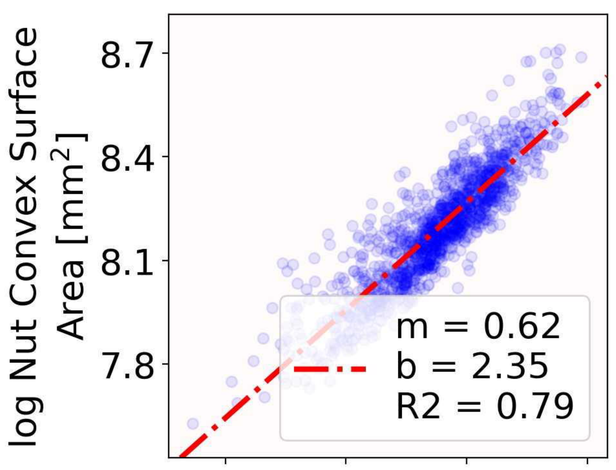


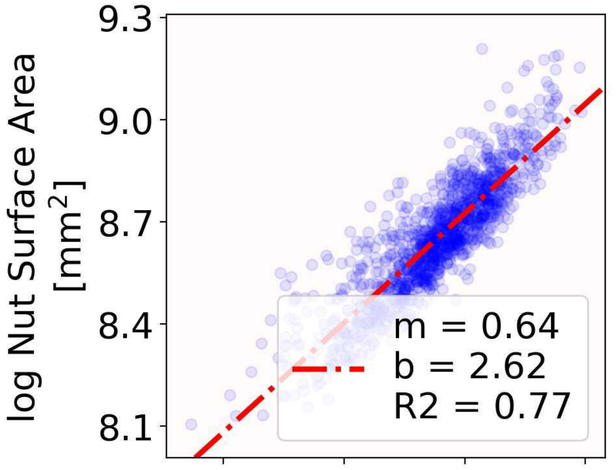


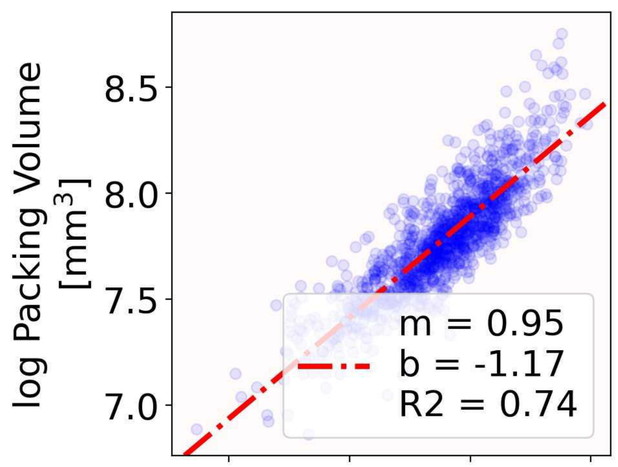


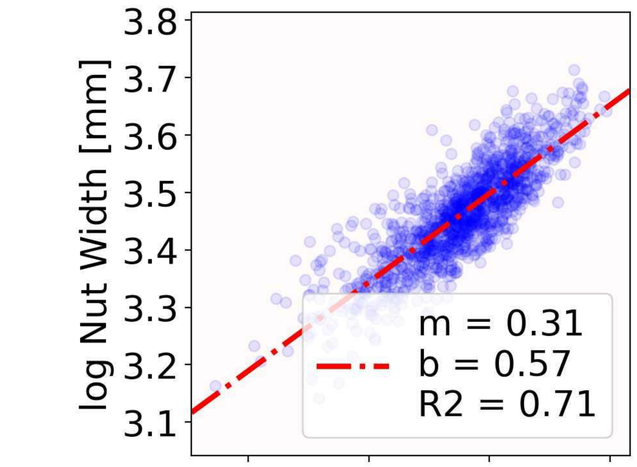


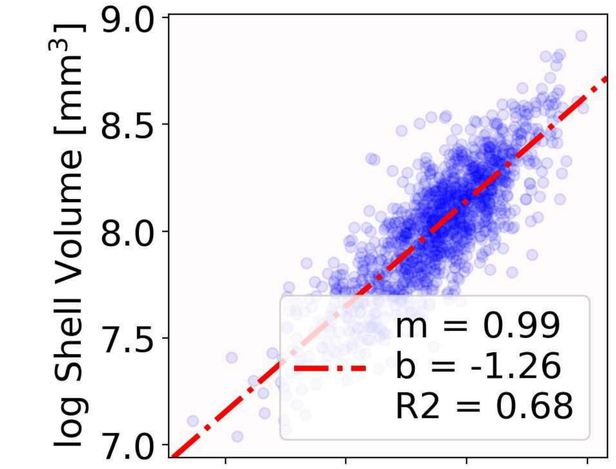


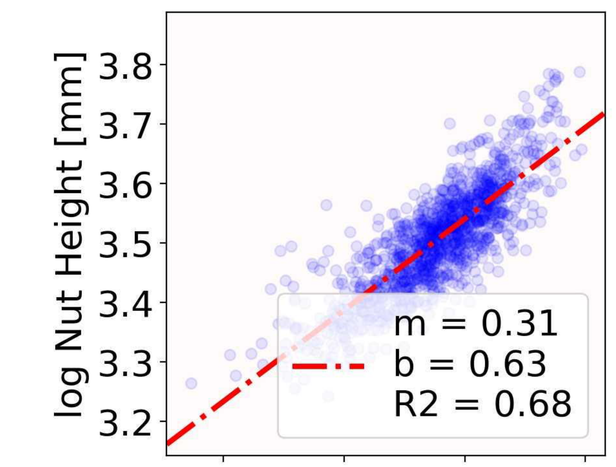


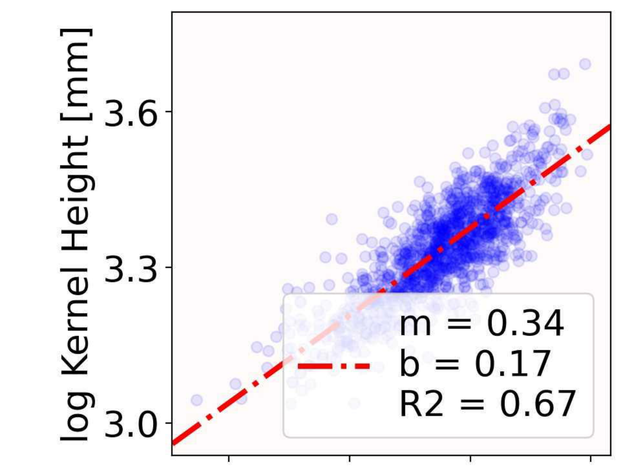


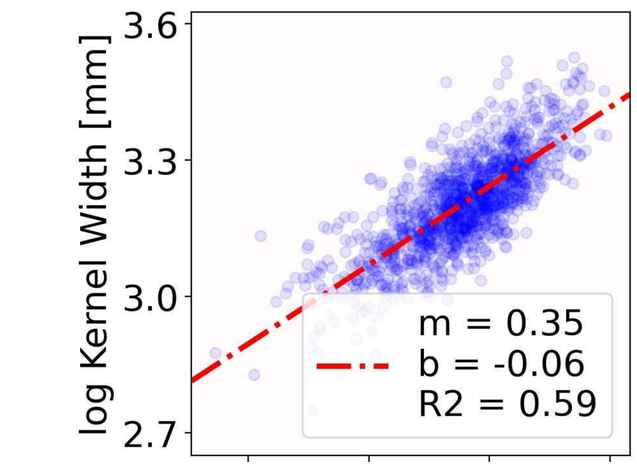


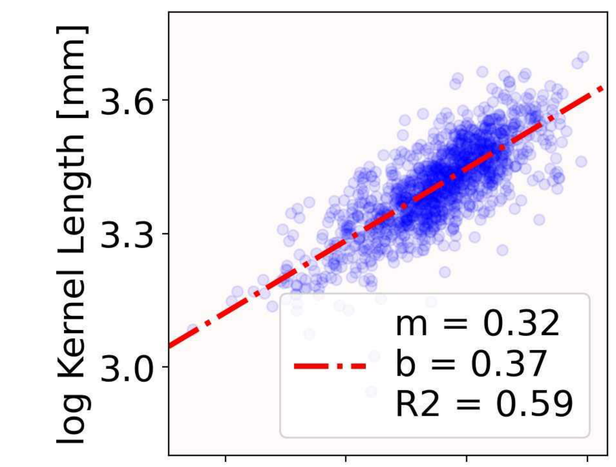


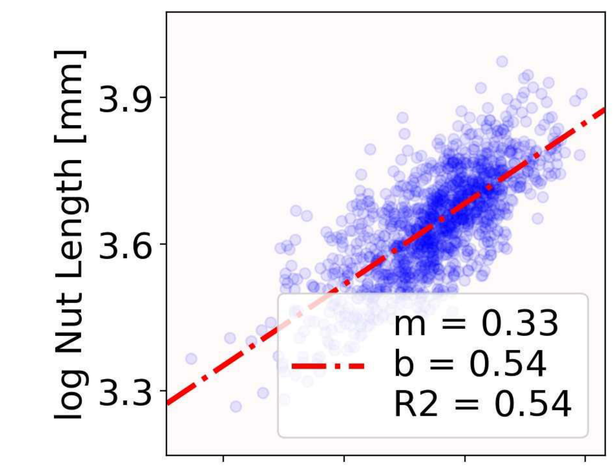


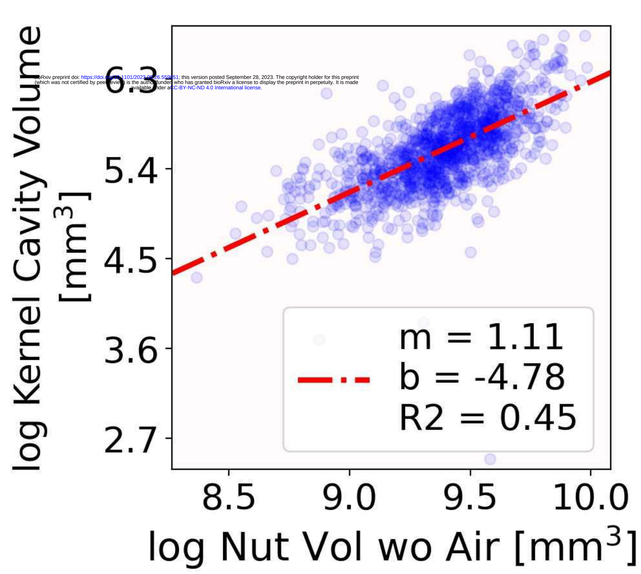


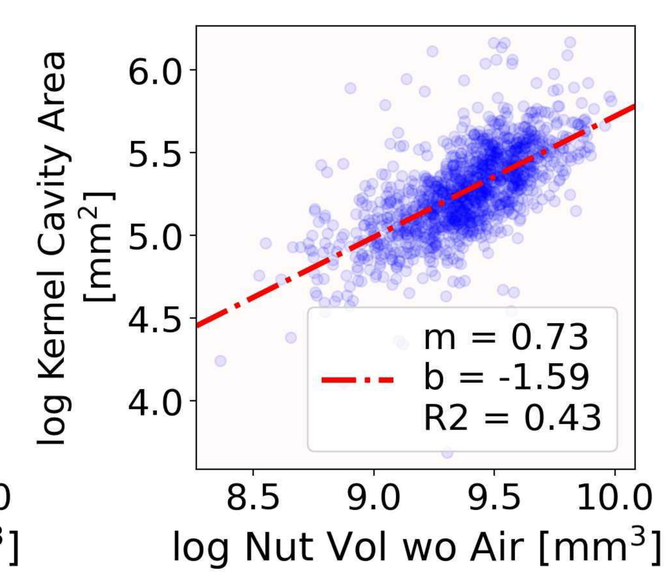


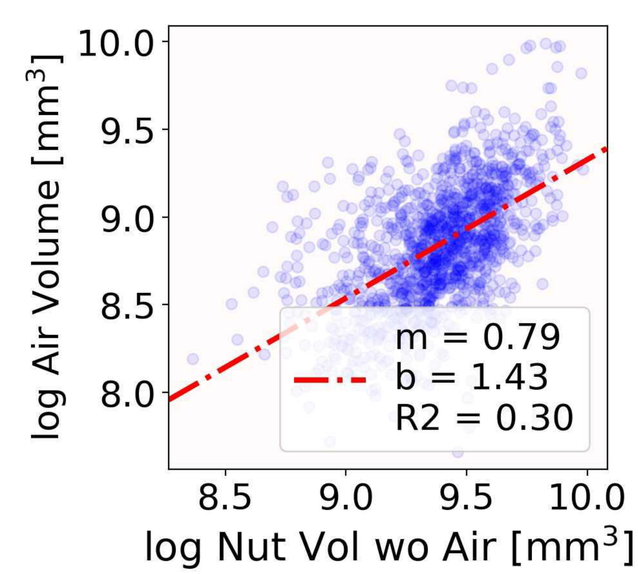


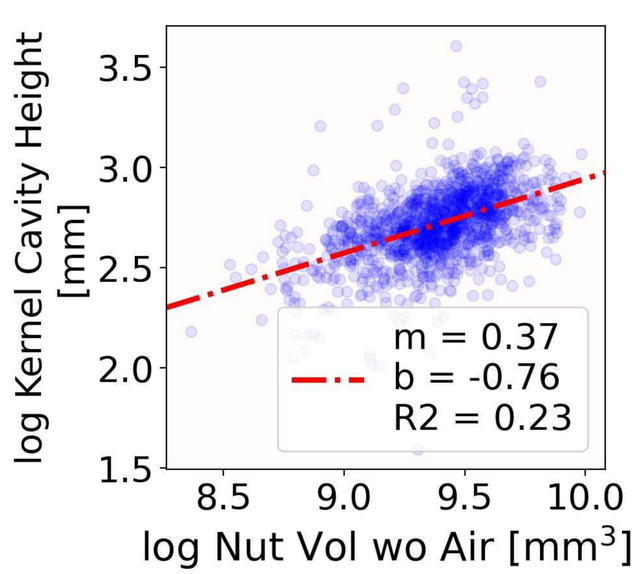












A

B

C

D

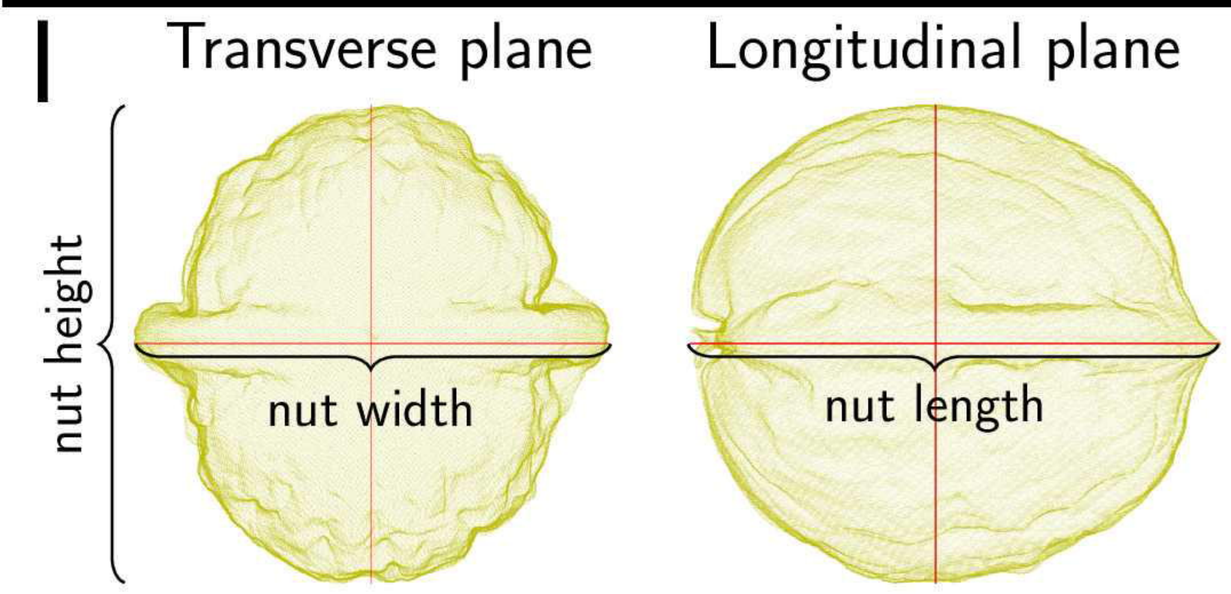
E

F

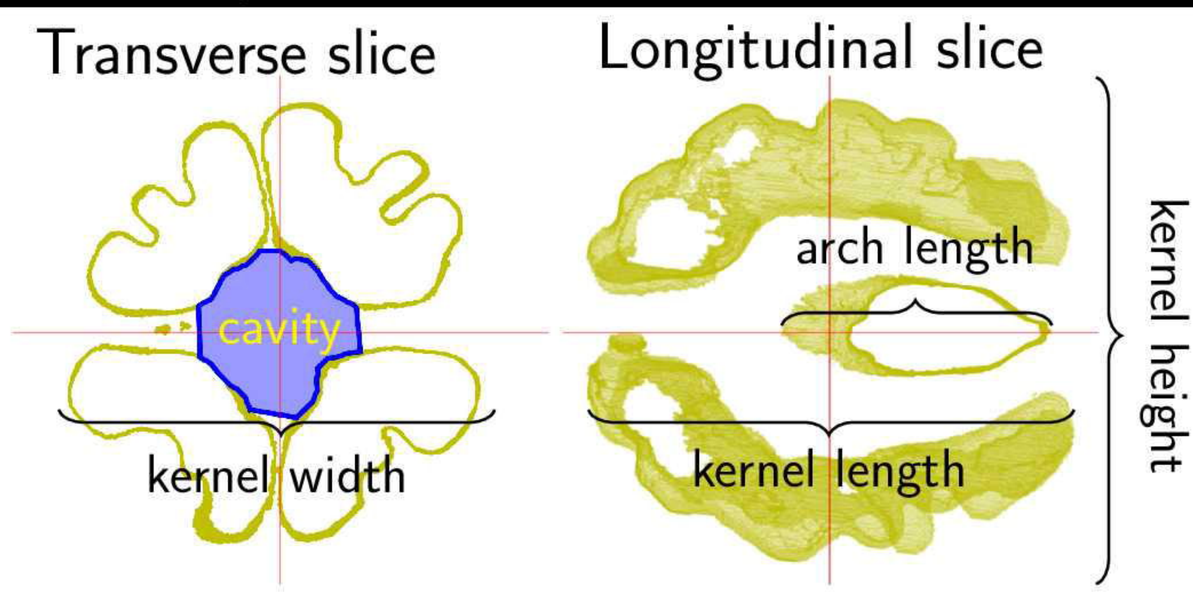
G

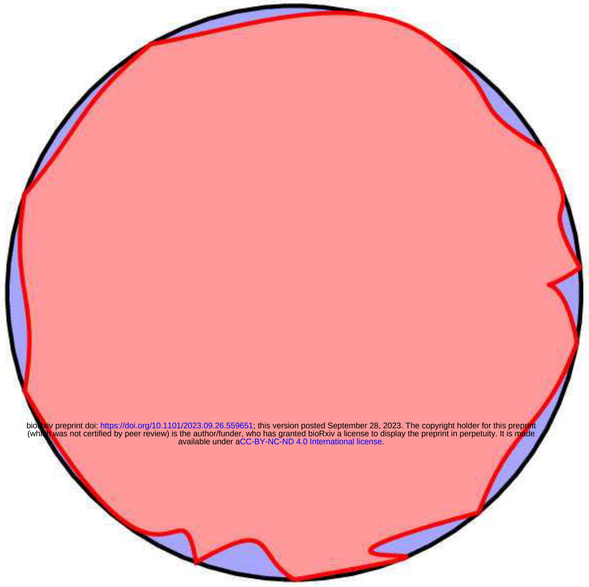
H

I

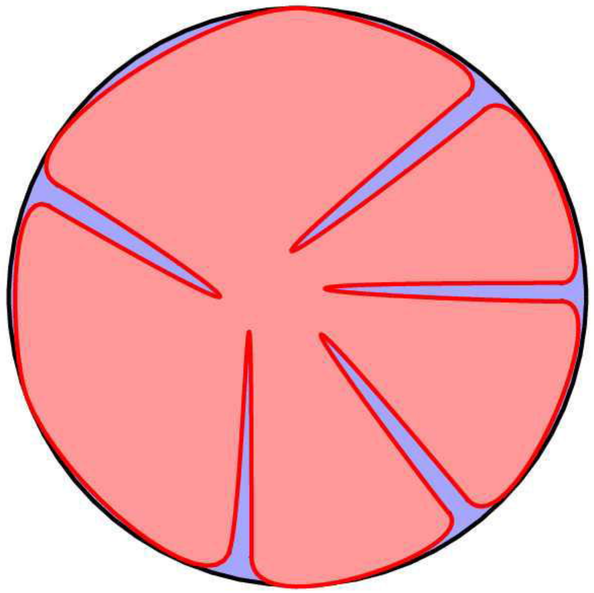


J

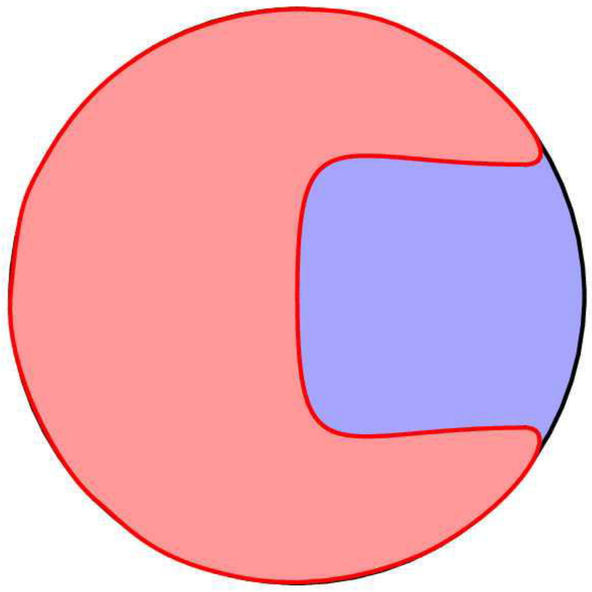




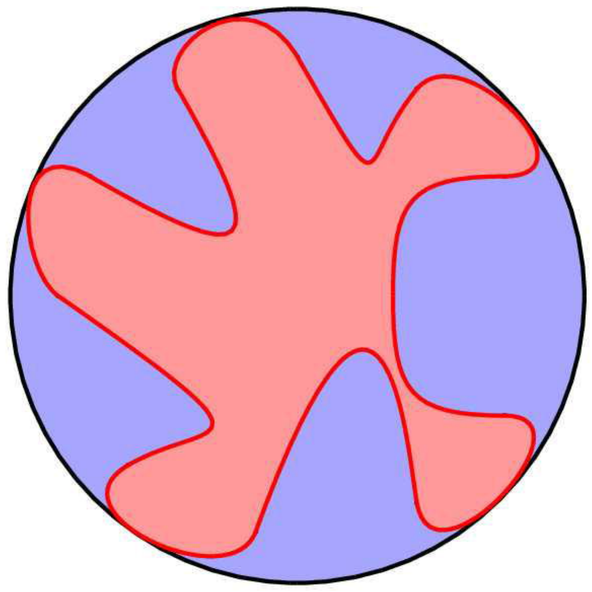
$$\text{Area} \approx \text{Convex.Area}$$
$$\text{Vol} \approx \text{Convex.Vol}$$
$$\frac{\text{Convex.Area}}{\text{Area}} \approx 1$$
$$\frac{\text{Volume}}{\text{Convex.Vol}} \approx 1$$



$$\text{Area} > \text{Convex.Area}$$
$$\text{Vol} \approx \text{Convex.Vol}$$
$$\frac{\text{Convex.Area}}{\text{Area}} < 1$$
$$\frac{\text{Volume}}{\text{Convex.Vol}} \approx 1$$



$$\text{Area} \approx \text{Convex.Area}$$
$$\text{Vol} < \text{Convex.Vol}$$
$$\frac{\text{Convex.Area}}{\text{Area}} \approx 1$$
$$\frac{\text{Volume}}{\text{Convex.Vol}} < 1$$



$$\text{Area} > \text{Convex.Area}$$
$$\text{Vol} < \text{Convex.Vol}$$
$$\frac{\text{Convex.Area}}{\text{Area}} < 1$$
$$\frac{\text{Volume}}{\text{Convex.Vol}} < 1$$

Spearman coefficients of correlation
Volumes, thickness, and ratios
[all accessions]

	Air Volume	Kernel Volume	Shell Volume	Packing Volume	Air Vol Ratio	Kernel Vol Ratio	Shell Vol Ratio	Packing Vol Ratio	Kernel Vol Ratio wo Air	Shell Vol Ratio wo Air	Packing Vol Ratio wo Air	Nut Vol wo Air
Air Volume	1											
Kernel Volume	0.55 ***	1										
Shell Volume	0.33 ***	0.64 ***	1									
Packing Volume	0.57 ***	0.78 ***	0.63 ***	1								
Air Vol Ratio	0.75 ***	-0.05	-0.23 *	0.04	1							
Kernel Vol Ratio	-0.59 ***	0.21	-0.09	-0.14	-0.76 ***	1						
Shell Vol Ratio	-0.50 ***	-0.07	0.56 ***	-0.04	-0.72 ***	0.22 *	1					
Packing Vol Ratio	-0.39 ***	-0.07	-0.05	0.32 ***	-0.49 ***	0.24 *	0.21	1				
Kernel Vol Ratio wo Air	0.03	0.23 *	-0.46 ***	-0.21 *	0.11	0.51 ***	-0.60 ***	-0.31 ***	1			
Shell Vol Ratio wo Air	-0.16	-0.13	0.61 ***	-0.04	-0.30 **	-0.23 *	0.86 ***	-0.08	-0.77 ***	1		
Packing Vol Ratio wo Air	0.17	-0.14	-0.25 *	0.39 ***	0.27 **	-0.39 ***	-0.37 ***	0.64 ***	-0.30 **	-0.31 **	1	
Nut Vol wo Air	0.55 ***	0.94 ***	0.82 ***	0.87 ***	-0.08	0.05	0.12	0.02	-0.06	0.10	-0.05	1
Shell Thickness	-0.23 *	0.21	0.79 ***	0.20	-0.61 ***	0.15	0.93 ***	0.10	-0.58 ***	0.84 ***	-0.40 ***	0.40 ***
Nut Volume	0.85 ***	0.86 ***	0.67 ***	0.84 ***	0.32 ***	-0.27 **	-0.18	-0.18	-0.03	-0.02	0.05	0.90 ***
Nut Ratio Convex Area	-0.25 *	-0.16	-0.27 **	-0.17	-0.12	0.20	-0.08	0.13	0.13	-0.20	0.05	-0.21 *
Density Kernel vs Shell	0.25 *	-0.19	-0.65 ***	-0.13	0.58 ***	-0.23 *	-0.78 ***	-0.07	0.41 ***	-0.66 ***	0.40 ***	-0.33 ***
Ease Of Removal	-0.27 **	-0.01	0.25 *	0.12	-0.40 ***	0.13	0.43 ***	0.33 ***	-0.33 ***	0.31 **	0.04	0.10
Shell Strength	-0.31 **	0.06	0.52 ***	0.21 *	-0.60 ***	0.16	0.73 ***	0.39 ***	-0.56 ***	0.59 ***	-0.01	0.25 *
Kernel Weight Ratio	0.19	0.07	-0.60 ***	-0.10	0.38 ***	0.16	-0.82 ***	-0.13	0.78 ***	-0.88 ***	0.13	-0.17
Shell Texture	0.19	0.05	0.24 *	0.10	0.11	-0.23 *	0.11	-0.12	-0.21 *	0.23 *	-0.04	0.13

	Shell Thickness	Nut Volume	Nut Ratio Convex Area	Density Kernel vs Shell	Ease Of Removal	Shell Strength	Kernel Weight Ratio	Shell Texture
Shell Thickness	1							
Nut Volume	0.13	1						
Nut Ratio Convex Area	-0.13	-0.25 *	1					
Density Kernel vs Shell	-0.83 ***	-0.08	0.01	1				
Ease Of Removal	0.40 ***	-0.08	-0.01	-0.33 ***	1			
Shell Strength	0.71 ***	-0.00	-0.00	-0.58 ***	0.45 ***	1		
Kernel Weight Ratio	-0.81 ***	-0.02	0.18	0.65 ***	-0.41 ***	-0.74 ***	1	
Shell Texture	0.12	0.16	-0.66 ***	-0.07	0.02	0.04	-0.21 *	1

A

Features used

Packing Vol Ratio
Shell Vol Ratio
Density Packing vs Kernel
Shell Volume
Protruding Shell Volume
Protruding Shell Ratio
Air Vol Ratio
Kernel Ratio Convex Area

1 2 3 4 5 6 7 8

B

EaseOfRemoval: OLS Linear Regression

R^2 score

0.6
0.5
0.4
0.3
0.2
0.1
0.0
−0.1

Train
Test

1 2 3 4 5 6 7 8

Number of morphological features used

C

Features used

Shell Thickness
Packing Vol Ratio
Shell Volume
Density Kernel vs Shell
Kernel Feret Ratio
Shell Vol Ratio
Kernel Vol Ratio wo Air

1 2 3 4 5 6 7

D

ShellStrength: OLS Linear Regression

R^2 score

0.8
0.6
0.4
0.2
0.0

Train
Test

1 2 3 4 5 6 7

Number of morphological features used

E

Features used

Shell Vol Ratio wo Air
Kernel Vol Ratio wo Air
Shell Thickness
Shell Vol Ratio

1 2 3 4

F

PercentKernel: OLS Linear Regression

R^2 score

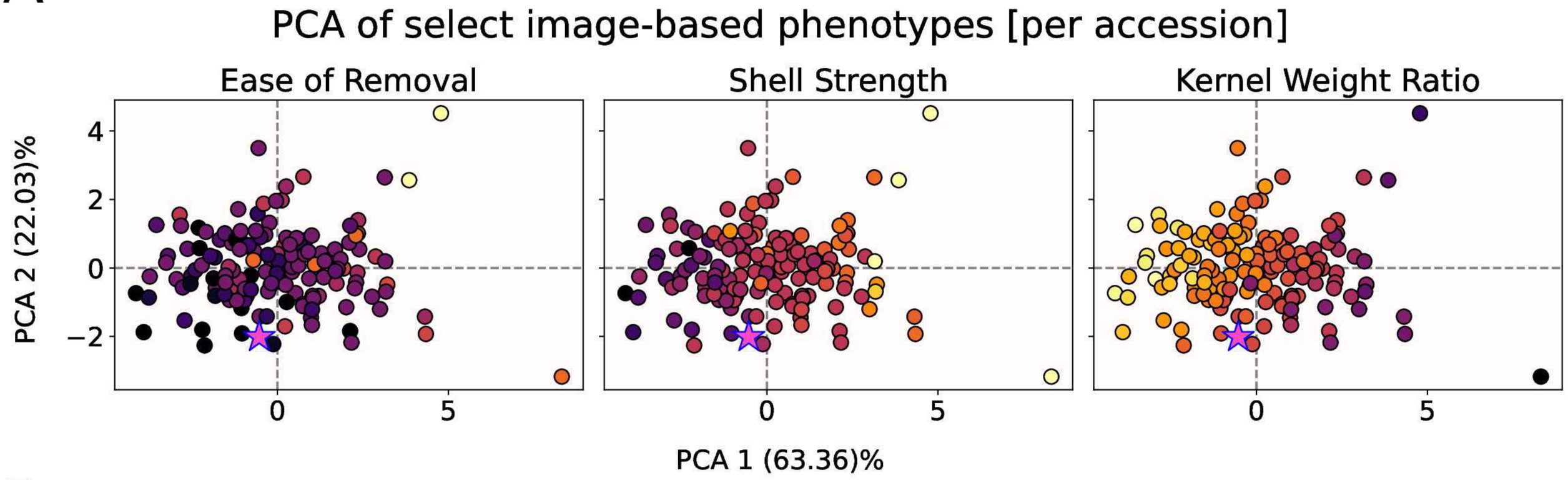
0.8
0.6
0.4
0.2
0.0

Train
Test

1 2 3 4

Number of morphological features used

A



B