A Morphometric analysis of Grain Size and Shape Variation during cereal domestication

Nathan Hughes, Hugo Oliveira, Fiona Corke, Terry Brown, Candida Nibau, and John Doonan

**Abstract**

**Introduction**

The collection of traits that emerged during the process of domestication and distinguish a crop from their wild progenitors is known as “domestication syndrome”. The term was coined by Hammer (1984) and has since been refined to refer to include morphological, biochemical, developmental, or physiological traits (Abbo et al. 2014). Some of these traits are unique to one particular “domesticated crop-wild progenitor” pair, for example, transformation from short and coarse fibers to long and fine ones in cotton domestication (Butterworth et al. 2009) or the increase in seed capsules in domesticated forms of poppy (Zohary et al. 2012).  Others are common to crops of a specific plant family, such as reduced grain shattering and loss of seed dormancy in grasses or indehiscent pods and reduction in seed toxicity in legumes (Zohary et al. 2012). Domestication syndrome traits are usually disadvantageous for wild populations and plants that possess them will only thrive under human management. Disarticulation of ears (“shattering”) is necessary in wild grasses to increase the area where spikelets might fall and germinate, but non non-shattering mutants were selected by humans during domestication of grass species such as wheat or rice as it makes it easier to harvest and store the edible grains (Doust et al. 2014).

Some  traits that differentiate crops from wild counterparts may not have emerged during domestication but afterwards, as particular varieties adapted to local environments or as crops were improved by selective selection and breeding. Examples include loss of day-length sensitivity in crops grown at northern latitudes and the emergence of free-threshing cereal varieties (Abbo et al. 2014). This makes it difficult to identify which traits resulted from the process of domestication.  Abbo et al (2014) propose that only traits having a clear dimorphic status that is always present in either domesticated or cultivated plants can be considered part of the domestication syndrome. Contrastingly, Milla et al (2015) extended the concept to include the physiological dynamics between different traits of the domesticated phenotype but also ecological aspects such as plant-soil feedbacks. Preece et al (2016) also emphasise the higher yields of crops when compared to wild progenitors as a key characteristic of domestication.  More subtle traits such as relative growth rate (the rate of dry matter production per unit of dry matter under optimal conditions) and seed mass have been postulated to differ substantially between some wild and domesticated species (Cunniff et al 2014).

In cereals such as wheat and barley, grain is the most used plant part and as such was affected by selection. Archaeobotanists distinguish the wild or domesticated status of cereals by observing the ear shape or the rachis fragments (tough rachis with a rough break scar and brittle rachis with smooth scar indicating domesticated and wild plants, respectively) (Brown et al 2009).  However, these are rarely found in the archaeological record, especially in the period of agriculture emergence in the Fertile Crescent, with charred grains being the more frequent type of plant remains. In the absence of diagnostic rachis, the possibility of distinguishing wild and domesticated cereals based on grain size and shape is a contentious issue that affects discussions on the place(s) and pace of domestication (Zohary et al., 2012).  From the analysis of archaeological grains, some authors agree that early domestication led to an increase in wheat and barley grain size and changes in its shape, independently from the evolution of rachis toughness (Brown et al. 2009; Nesbitt, 2002). Other authors suggest that the overlap in shape and sizes between archaeological and present-day wild and domesticated cereal grains indicate that these traits diverged after domestication and a secure diagnostic cannot be made (Abbo et al. 2014).  Though important, the phenotypic variation of cereal grain morphology is surprisingly understudied mainly due to the difficulty in quantifying this trait. Previous studies used a limited number of metrics that were analyzed discretely largely in single mapping populations (Giura and Saulescu, 1996; Campbell et al., 1999; Dholakia et al., 2003; Breseghello and Sorrels, 2007; Sun et al., 2009) and some preliminary morphometric studies suggest that grain shape does indeed differ between wild and domesticated wheat and barley (Bonnhomme and Wallace, 2016;  Gegas et al. 2010).

Wheat is one of the major food crops to have been domesticated in the fertile crescent around 10,000 years ago. The earliest cultivated forms were the diploid einkorn (*Triticium monococcum subsp. monococcum*, AA) and the tetraploid emmer *Triticum trugidum subsp. dicoccum,* AABB). These cultivated forms are considered landraces derived from the wild relatives *Triticum monococcum subsp. aegilopoides* (wild einkorn) and *Triticum turgidum subsp. dicoccoides* (wild emmer) respectively (Shrewry).  As they were cultivated and selected for by early farmers, the traits from the original progenitors were changed to better suit the agricultural and food needs of the early human population. Through further domestication the tetraploid wheat would then give rise to the modern tetraploid variety, *Triticum durum* (pasta wheat)  and through hybridisation with *Triticum tauschii* (DD) to the hexaploid *Triticum aestivum* (bread wheat) and *Triticum spelta* (spelt). Barley (*Hordeum vulgare*) was also domesticated in the fertile crescent from the wild progenitor *Hordeum spontaneum* and together with wheat are the founder crops on which western agriculture was built (Kilian, 2010). Although many studies look at the genetic underpinning of wheat and barley domestication, detailed evidence on how grain size and shape changed during the domestication process is still limited.

Originally developed as a medical diagnostic tool, X-ray micro computed tomography (μCT) is a non-invasive imaging technique based on differential X-ray attenuation by biological material that can be used to study complex plant traits. The capacity to provide and accurate 3D representation and quantification of internal structures in a non-invasive and non-destructive way, combined with the ability to automate the process, means that μCT is an attractive approach to study plant traits. High resolution μCT has been successfully used to various analyse plant traits and environmental responses (reviewed in Dhondt et al., 2010). We have recently shown that by using μCT scanning combined with an image analysis pipeline of dried wheat spikes we can accurately extract and measure grain and spike parameters (Hughes et al., 2017). Besides the traditional measurements of grain length and width, we can also measure depth, volume and surface area that combined provide a overall and accurate measure of grain size and shape. In addition, the ability to look at a large number of samples enables us to be analyse the influence of genotype and environmental factors on grain traits. The high throughput and automation provided by this technique combined with the ability to measure objects in three dimensions overcomes the limitations of more traditional approaches allowing for a better understanding of how grain shape and size changed during the domestication process.

Identifying the morphological and genetic basis of domestication syndrome traits is a rich field of research but it remains an open question if some there are traits that distinguish crops from their wild progenitors that have not yet been described (Stetter et al, 2017, Roucou et al. 2017).

In here, we describe the use of micro-CT analysis to dissect changes in cereal grain size and shape across three independent domestication events: einkorn and emmer wheat and barley. We demonstrate that grain shape changed during domestication and that change is driven to increased grain depth and to a less extent increased grain width with minimal changes in grain length. We provide evidence that grain shape is an important component of the domestication syndrome underscoring the need to use three dimensional measurements when looking at grain traits. Finally we developed a model able to, for a given species, classify grains based on their domestication status.

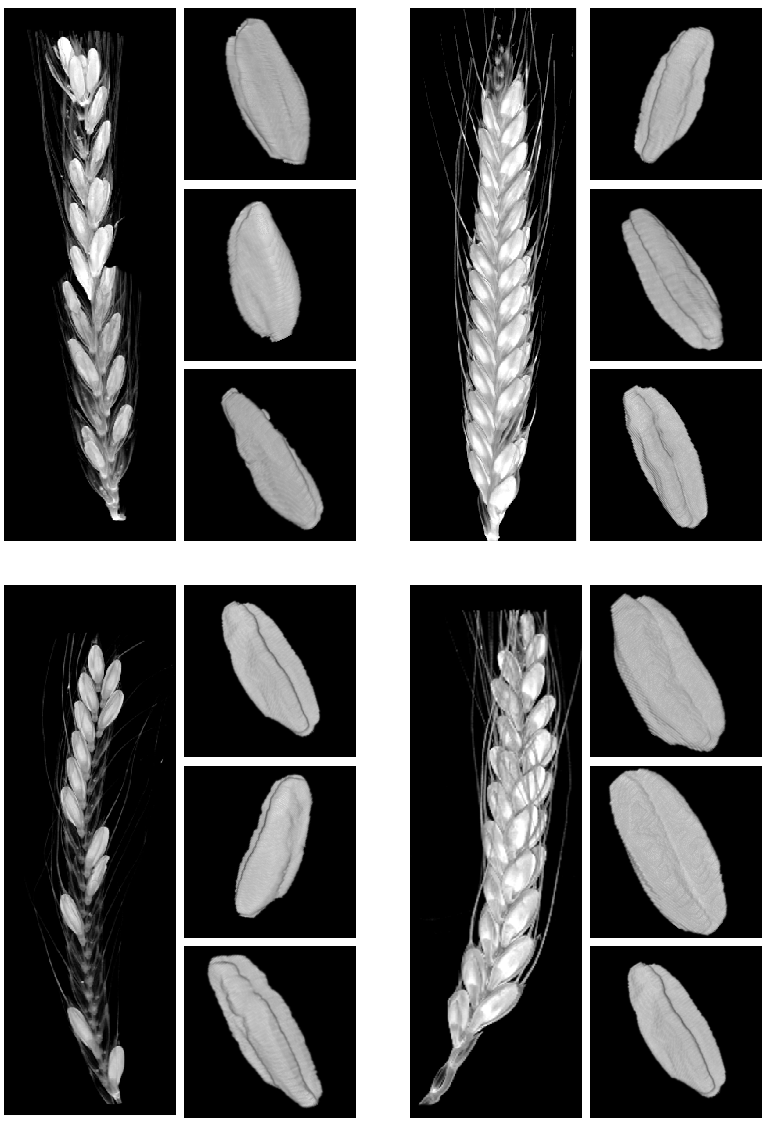
**Results**

**3D grain trait feature extraction and analysis**

It is well documented that during domestication there was a change in grain size and shape in wheat and other cereals. Several studies have confirmed the expected grain filling, volume and size change undergone during domestication (Gupta, Rustgi, and Kumar 2006; Piperno et al. 2004), but have been but a detailed analysis of 3D grain traits is lacking.

We selected four wheat species representing two parallel events in wheat domestication and two different ploidy levels. The diploids *Triticum monococcum subsp. aegilopoides* (wild einkorn) and *Trticium monococcum subsp. monococcum* (einkorn) and the tetraploids *Triticum turgidum subsp. dicoccoides* (wild emmer) and *Triticum trugidum subsp. dicoccum* (emmer) (Figure 1). In order to select traits that were a product of domestication and not of regional differences or growth conditions, we selected spikes from plants from different geographic origins and grown in different locations. Spikes were scanned in a micro-CT scanner as previously described (Hughes et al. 2017). Traits were gathered using the image analysis software described in Hughes et al. 2017.  Small Modifications were made to the grain identification algorithms in order to increase the robustness and applicability of the method to the species used. These modifications included an updated pixel distance algorithm, and improved grain segmentation. A schematic representation of the image analysis pipeline used is provided in Supplemental Figure 1.

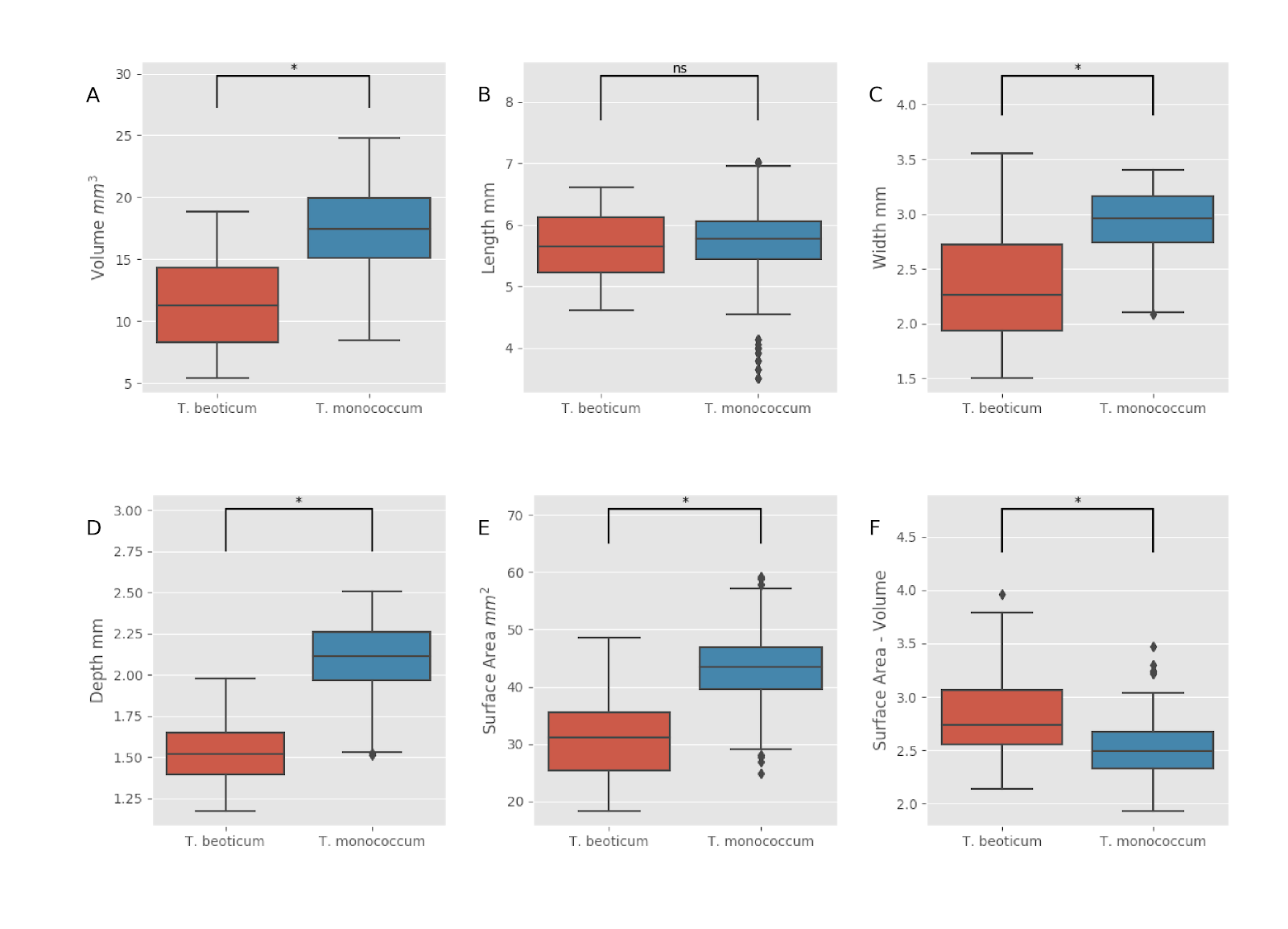
Key grain morphological traits were measured, these include length, width, depth, volume, and surface area. From these traits, two additional descriptors were created LWD (Length X Width X Depth) and Surface area / Volume to provide better insights into trait interactions.



*Figure 1-  3D visualization of reconstructed micro-CT images, whole spike (left) and representative isolated grains (right).  A- Wild Einkorn, B- Einkorn, C- Wild Emmer and D- Emmer.*

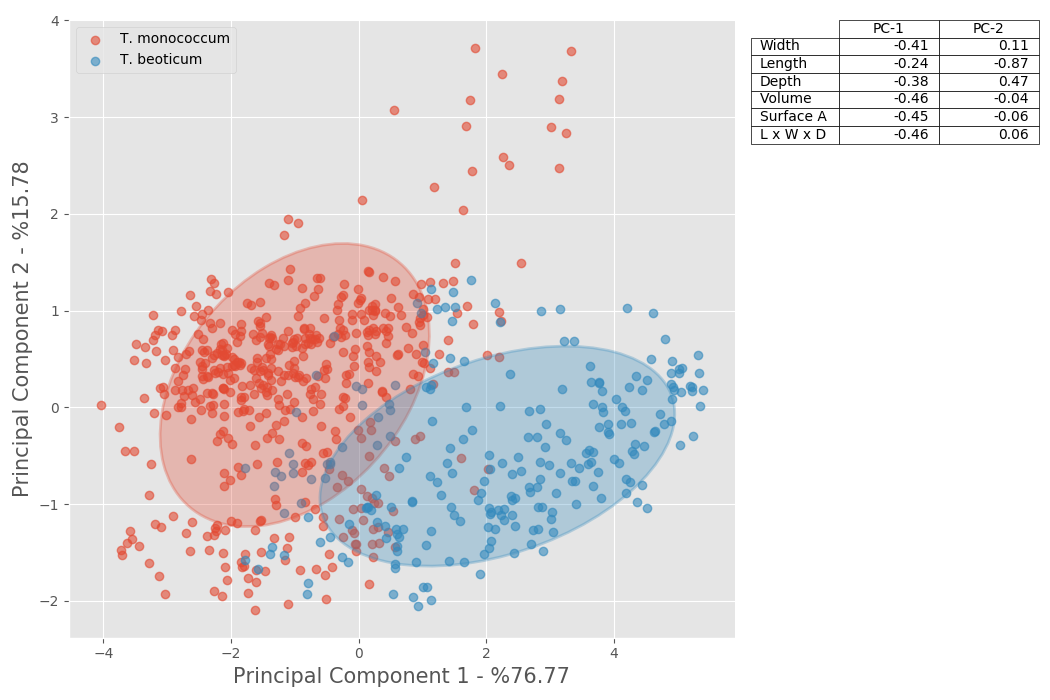
**Volume increase during Einkorn domestication is mainly due to changes in grain depth**

In order to identify domestication related grain traits in Einkorn wheat, 237 grains for the wild and 513 for domesticated were studied and morphological traits extracted.  As expected, we observed a significant increase in grain volume in the domesticated Einkorn when compared to wild (Fig. 2A). This is in agreement to the previously described increases in TGW (REF and Supplemental Figure 2). Surprisingly,  grain length was not significantly different between the wild and domesticated Einkorn (p=0.02 and a predicted 35% probability of overlapping averages by Kruche’s Bayesian estimation of difference) (Figure 2B and Supplemental Table 1). The major changes in domesticated grains were a small but significant change in width and a more pronounced change in depth (Fig. 2 C and D). The same trend were observed when the grains were analysed using MARVIN, a 2D imaging technique (Supplemental Fig. 2). As expected, we observed an increase in surface area (Fig. 2E) and a concomitant decrease in the surface area to volume ratio in domesticated Einkorn (Fig. 2F). This ratio is an important agronomic trait is the  which influences the milling properties of the grain.



*Figure 2- The effect of domestication on grain traits of wild (blue boxes) and domesticated (red boxes) Einkorn wheat. Boxplots represent median plus the interquartile range for grain Volume (A), Length (B), Width (C), Depth (D), Surface area (E) and Surface Area to Volume ratio (F). Asterisks indicate that the values are significantly different for p<0.01*

In order to identify the major sources of morphological variation, a two component PCA was performed (Fig. 3). Both wild and domesticated groups are separated and PC1 can explain 76.77% of the variation. Analysis of the loading for each component reveals that   no dominant trait appears to influence morphometric variation in einkorn species (figure:3) suggesting that the combination of traits, and thus shape is a major determinant.

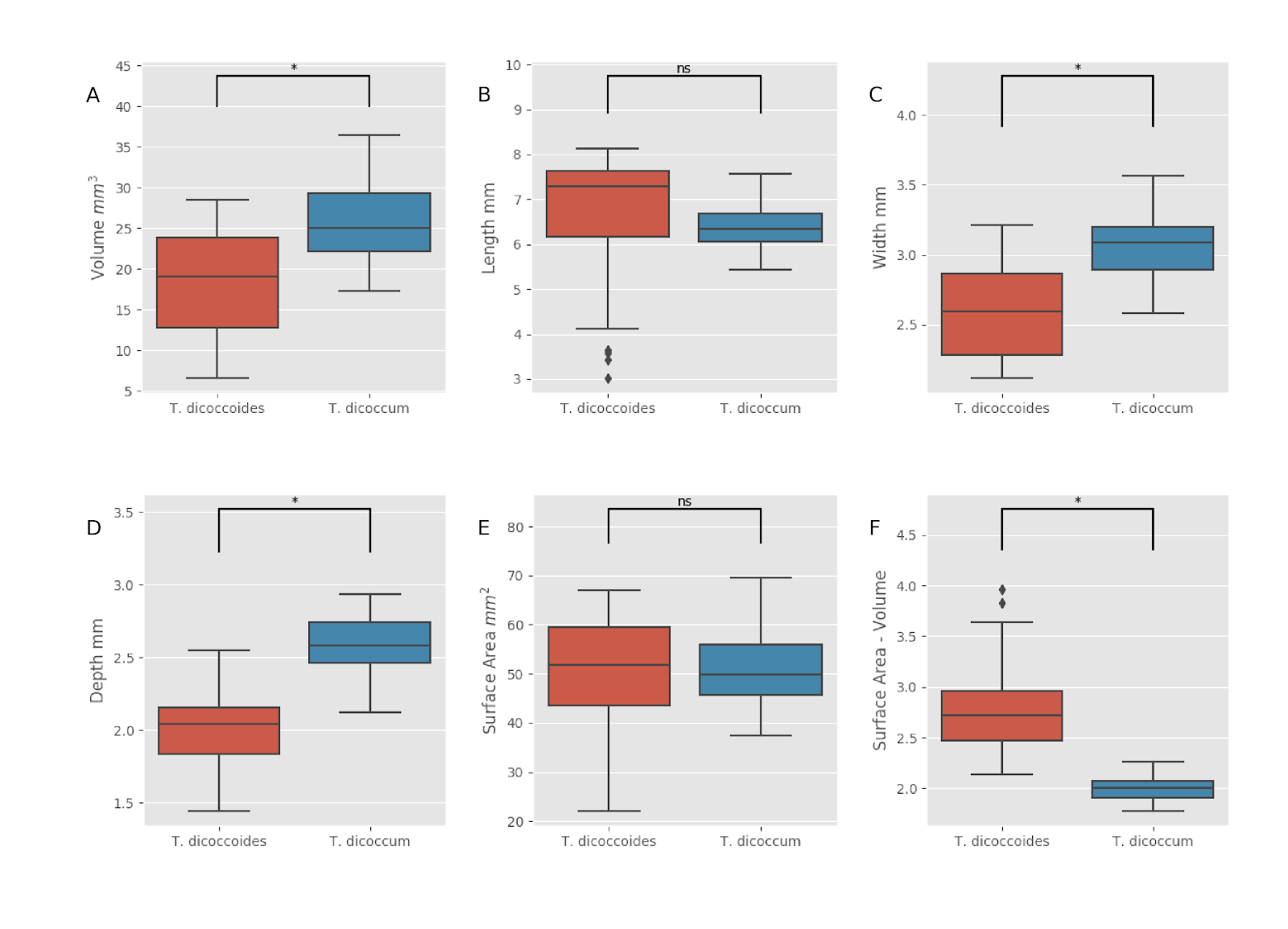


*Figure 3- Principle Component analysis of grain traits for wild (red circles) and domesticated (blue circles) Einkorn wheat, with two times standard deviation shaded. Table shows the loading values for PC1 and PC2.*

**Increased grain depth is also a major domestication trait in Emmer wheat**

As we observed for the diploid einkorn, in the tetraploid Emmer, domestication led to a significant increase in grain volume (Fig. 4A). This increase was due to increased width and depth of grain with grain length not changing significantly (Fig. 3B, C and D). Surprisingly we also did not observe a difference in grain surface area (Fig. 4E), showing that grain compactness has significantly altered to preserve this trait. The Bayesian model predicted a 20% probability of domesticated and wild types differing, showing that there is an indication of change (Supplemental Table 1). Though, with a high enough probability of this being down to chance. The observed decrease in the surface area to volume ratio is due to increases in grain volume (Fig. 4F).

These data suggest, as it was the case with Einkorn, that the changes in grain volume through domestication where the result of changes in size and shape with the depth component of the grain having a major effect in changing grain size and shape.



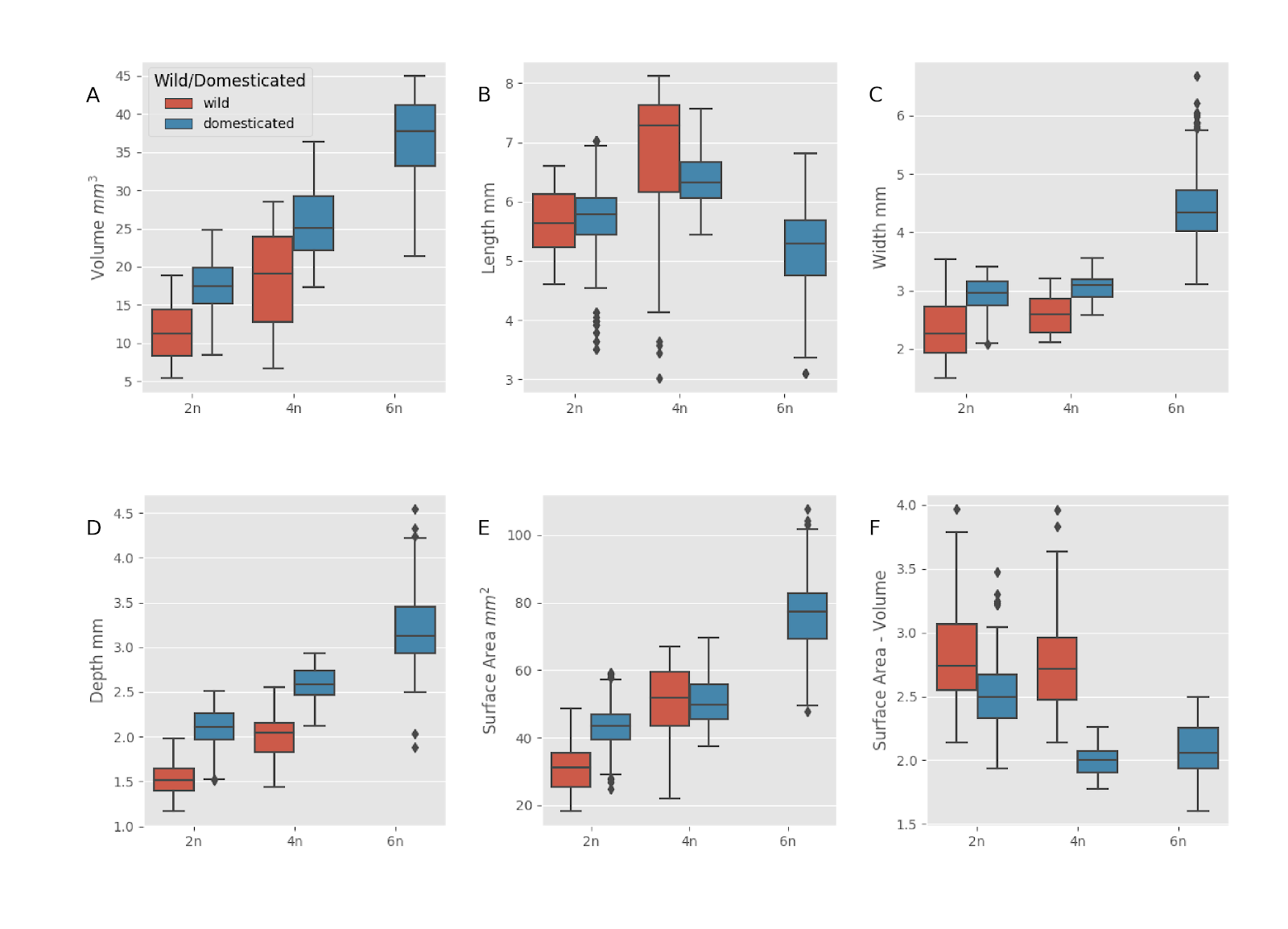
*Figure 4- Grain trait analysis of  wild (blue boxes) and domesticated (red boxes) Einkorn wheat. Boxplots represent median plus the interquartile range for grain Volume (A), Length (B), Width (C), Depth (D), Surface area (E) and Surface Area to Volume ratio (F). Asterisks  indicate that the values are significantly different for p<0.01. 31 grains were analysed for wild Emmer and 122 for the domesticated Emmer.*

A two component PCA for wild and domesticated Emmer grain traits, with PC1 and PC2 providing 78% and 15% respectively, shows that traits width, depth, volume, surface area as well as the interaction term LxWxD have the most influence in grain morphology (Supplemental Figure 3) while length shows the lowest impact as expected from the results above. The second principle component, with much less explanation, shows that the interaction between length, width and depth is much less important in explaining variance whilst length alone (coef=0.72) provides significant coverage (Supplemental Figure 3).

**Grain size and shape change across ploidy**

As mentioned above, besides the wild and domesticated species, there was also variation in ploidy with Einkorn being a diploid species and Emmer tetraploid. We have also include grain traits for hexaploid wheat (from Hughes et al., 2017) so we have a complete ploidy series within the *Triticum*.  As can be seen in Fig. 5A, while grain volume increased with the domestication status, ploidy levels also had a significant effect with grains for both the wild and domesticated tetraploid Emmer being bigger than the corresponding diploid Einkorn types. Interestingly, increased ploidy in the hexaploid wheat did not result in increased volume compared with the domesticated tetraploid Emmer (Fig. 5A).  This observation implies that the increased yield in the hexaploid wheat (as determined by TGW REFs) must have been the result of grain shape and compactness change. To address this in more detail we looked at other grain traits. As described above, grain length did not significantly change with domestication or ploidy increase in Einkorn or Emmer but a significant decrease in length was observed for hexaploid wheat (Fig. 5B). Small changes in grain width happened with domestication and ploidy increases (Fig. 5C) but the more dramatic change is seen with grain depth (Fig. 5D). This observation underscores the importance that grain depth, a usually overlooked trait, has on grain size and shape changes during domestication.

As expected, the ratio of grain surface area to volume has changed with domestication and not with ploidy levels supporting the idea that this agronomically important trait was under strong selection during the early domestication events (Fig. 5E).

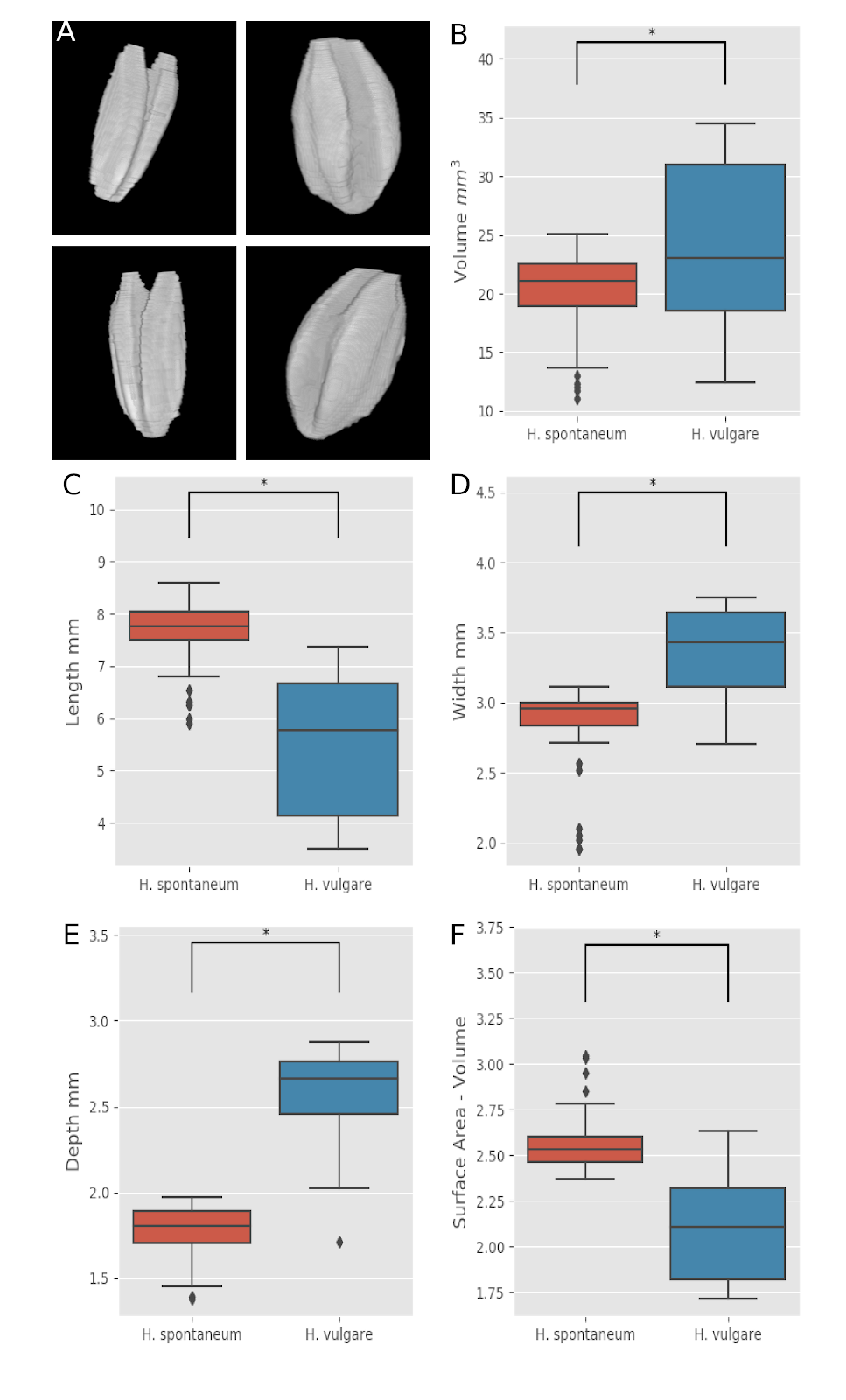


*Figure 5- The effect of ploidy (y axis) and domestication (wild - red boxes, domesticated - blue boxes) on wheat grain traits. Boxplots represent median plus the interquartile range for grain Volume (A), Length (B), Width (C), Depth (D), Surface area (E) and Surface Area to Volume ratio (F). Asterisks indicate that the values are significantly different for p<0.01.*

**Changes in barley grain shape during domestication**

Based on the results described in the sections above, we conclude that grain size changes in wheat were mostly due to changes in grain depth and to a smaller extent in grain width. To determine if this was specific for wheat domestication or also happened in other cereals we scanned spikes of wild (*Hordeum spontaneum*) and domesticated 2-row (*Hordeum vulgare*) barley. As can be seen in Fig. 6A, grain shape is different between wild and domesticated barley. Surprisingly, we only observed a small but significant change grain volume between wild and domesticated barley (Fig. 6B). In contrast to what was observed for wheat, there was a significant reduction in grain length accompanied by increases in grain width and most notably in depth (Fig. 6C-E). A similar trend was also observed using 2D MARVIN analysis (Supplemental Figure 2). The surface area to volume ratio also decreased as expected for a domestication event (Figure 6F).

Taken together, these data suggest that both in barley and wheat, grain volume increases happened in an allometric way with depth and width being responsible for the majority of the volume changes.

**

*Figure 6- Analysis of grain trait changes during barley domestication. A- 3D reconstruction of representative wild (left) and domesticated (right) barley grains. Grain traits measured were   Volume (B), Length (C), Width (D), Depth (E) and Surface Area to Volume ratio (F). Boxplots represent median plus the interquartile range for grain and asterisks indicate that the values are significantly different for p<0.01.*

**Developing a model to predict domestication in cereals**

The aforementioned results demonstrate that domestication affected grain size and shape in wheat and barley. We next asked if we could model those changes and develop a mathematical model capable of distinguishing wild and domesticated grains based on morphological parameters.  To this end, we used a multiple regression model to obtain predictions of domestication using three key parameters (grain length, width and depth). A high variance was explained using equation 1 and an ordinary least squares method (OLS).



*Eq1: OLS Regression model for separating grains based on domestication*

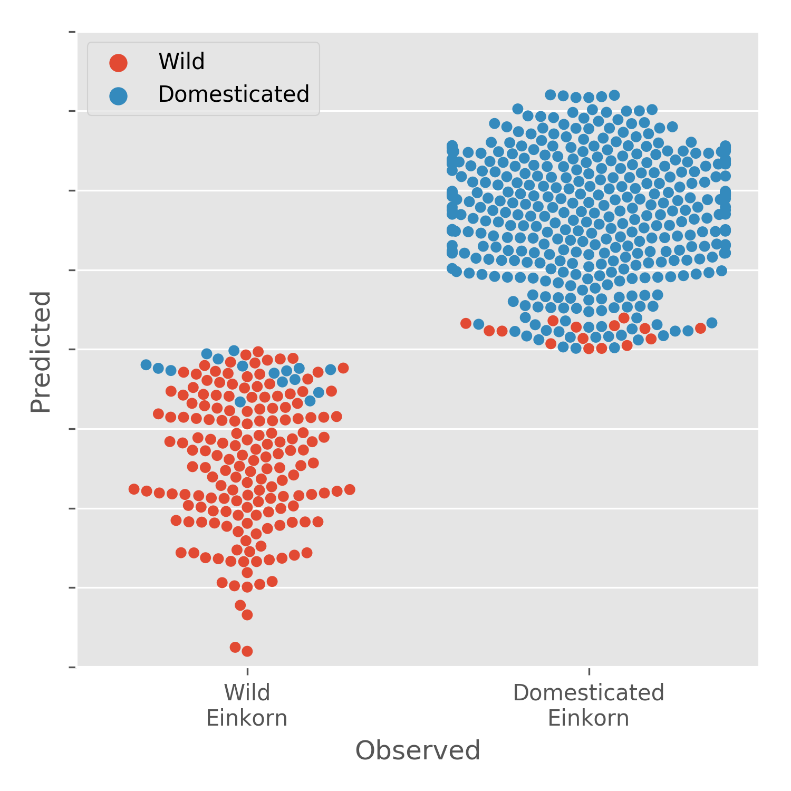
The model was built using the einkorn grain population (750 grains in total). A simple ordinary least squares model (see EQ above) was provided with 20% of the total grains (randomly sampled) and then tested with the remaining 80% of the data.

As can be seen in Figure 7, the model is able to predict the domestication status of the Einkorn grains with high accuracy and a minimal sum of squared errors (R^2 value of 0.95). In addition, the model also underscores the importance of grain depth when describing and classifying grains as the accuracy of the model is drastically affected when depth is not included (R^2 value of 0.31). A similar level of accuracy was obtained when the model was tested in the Emmer grain population (156 grains) with an R^2 value of 0.98 (Supplemental Figure 4).

In order to test applicability to other cereal grains, we ran the model on the barley grain dataset (746 grains) and as observed with wheat, the model as able to predict the domestication status of each grain with high accuracy (R^2 value of 0.98, Supplemental Figure 4).

Finally, and in order to further test the robustness of the model and exclude any geographical and environmental effects, we grew a population of wild and domesticated Einkorn wheat. Fully mature and dried primary spikes were scanned and grain parameters extracted, as described above and they showed similar trends as observed before (Supplemental Figure 5) When the model was applied to this new dataset we again found that it could accurately classify wild and domesticated grains with a R^2 value of 0.96 (Supplemental Figure 4). In addition, and as it was the case for the other datasets, the accuracy was decreased if depth wasn’t included (R^2 value of 0.39).

The data show that for each event, the model can accurately predict the domestication status via grain geometric parameters.

**

*Figure 7- Modelling grain domestication in Einkorn wheat. Multiple regression model shows the predicted (y axis) and actual (x axis) domestication status of Einkorn wheat with an R^2 value of 0.95 for 750 individual grains analysed.*

**Evaluating other domestication traits in Einkorn wheat**

As we were growing the Einkorn wheat as a test population for the domestication prediction model we also looked at other domestication related traits. As expected, we saw a decrease in the number of days the plants took to reach key developmental stages, namely GS39, GS55 and GS65 when the wild and domesticated Einkorn varieties were compared. This decrease was even more pronounced in the hexaploid wheat elite variety Paragon (Supplemental Figure 6). Total ear weigh and harvest index, two key indicators of domestication and crop yield, were also significantly higher in the domesticated Einkorn and the elite wheat (Supplemental Figure 6).

Taken together our data show that domestication has had a strong effect on grain shape and size in wheat and barley and the observed differences are due in part to changes in grain depth. In addition we have developed a model where we can predict if a grain of a certain species is wild or domesticated. The accuracy of the model is drastically increased by using grain depth in conjunction with width and length, again stressing the importance of looking at 3D parameters when comparing grain size and shape across ploidy and domestication events.

**Discussion**

One of the earlier and arguably more important domestication events in cereals was an increase in yield. This was brought about by increasing the number of seeds per spike and increasing grain weight. In open environments where agriculture initiated, larger seeds would have a significant advantage by being able to survive deeper burial and generally producing larger seedlings that germinated faster (Fuller et al., 2007) and would thus be selected for. It is possible that the need of seeds to absorb water and nutrients and seedlings to emerge from greater depths could have led to changes in grain length, width or crease depth (Kluyver et al. 2013).

Bigger grain, also meant increased levels of endosperm making them easier to mill and providing more flour. The increased yield by more and larger grains is proposed to be one of the earliest domestication traits to appear preceding non-shattering and naked types. Indeed, in rice, it only takes five generations to observe an increase in grain size from cultivating wild relatives (Oka and Morishima, 1971). Later on in the domestication process, more targeted breeding lead to further increases in size and a thin long grain was replaced by the shorter rounder grain found in modern varieties.

The two major factors determining weight are grain size and grain shape. Although several studies investigate the effect of domestication on grain size, there is little information on how wheat grain shape changed in early domestication events. One of the reasons being that accurate measurement of grain traits in three dimensions has been difficult and cumbersome. Here we took advantage of X-ray micro CT scanning and recently developed image analysis tools to overcome those limitations. We have previously shown that the scanning of dried intact spikes allows us to extract measurements of grain and spike traits with high accuracy and speed. Measurements like depth, volume and surface area are easily obtained with this method (Hughes et al., 2017). The majority of the studies so far in grain trait analysis in cereals have used 2D imaging techniques like MARVIN. The techniques provide measurements of length, and width, with the possibility of width and depth being measured interchangeably depending on how the grain was placed. Also measurements of surface area and the estimated TGW obtained with these methods only take in consideration two dimensions and not the general grain shape. As through domestication the grains tended to become more spherical this would increase the inaccuracy of these 2D methods. In our data, the inclusion of a third measurement (depth) drastically increased the prediction accuracy for a seed being from a wild sample or a domesticated one (see below).

Some evidence suggests that the major changes in grain size and shape happened within the earlier domestication events but in these studies only a small number of grain was used or only one domestication event (Gegas,Fuller, 2007, Gegas, Okamoto, 2012, Lin Qin, 2017).

In this paper we use uCT scanning coupled with a tailor-made software analysis package to further explore how early domestication events affected grain shape and size. We used three independent domestication events and by combining accessions from many geographical origins we hoped to minimize the environmental effect on grain size and shape and only look at traits that changed with domestication alone. In addition, by choosing einkorn and emmer we are also able to look at the effect of ploidy on grain traits. Finally the use of barley allowed us to test if the changes in domestication traits in wheat were also common to other independently domesticated cereals.

**Domestication had a major effect on grain shape through increases in grain depth**

Changes in seed shape and size associated with domestication have been reported for grape and date palm and serve as diagnostic features to distinguish true wild from cultivated or feral populations (Terral et al. 2010; Gros-Balthazard et al. 2016). These methods have barely been explored in cereals such as wheat and barley. Our work shows the potential of these methods to address archaeological or historical questions when grains of undetermined species are recovered.

Across the three domestication events analysed we observed the expected increases in grain volume. This increased volume has primarily driven by increases in width and more dramatically in grain depth. Changes in grain shape had previously been reported for einkorn and barley grains obtained from archeological sites in the near East (Fuller et al., 2007) and in populations of emmer (Okamoto) and our provide further support and establish that in wheat grain shape is also an important component of the domestication syndrome.

Some studies suggest that grain length was also affected by domestication (Okamoto - more?) but in our data, grain length was not affected by the domestication status of either einkorn or emmer and only the elite hexaploid wheat showed significantly shorter grain compared to the other varieties. This differences could be dependent on the type and size of the population used or the measuring methods, in our case the hull portion is removed by the segmentation algorithm in not measured.

In barley, we also observed changes in grain shape with increased width and depth in the domesticated varieties but this was accompanied by slight decrease in grain length. In general, there was a smaller volume increase in domesticated barley. This could be due to different agricultural and processing uses for barley or a domestication bottleneck. This small difference in volume is recognised and makes it difficult to distinguish wild and domesticated varieties in archaeological sites (Fuller 2007).

**Grain shape changes during domestication**

In the three domestication events we analysed, there was a clear change in grain shape towards more spherical grains. A good measure of this change is the ratio between grain surface area and grain volume and it has significant implications for the milling industry where a bigger endosperm to bran ratio is highly desirable. Using this 3D uCT technique we were able to get accurate measurements of both surface area and volume that were previously not possible.  For einkorn, emmer and barley we found a significant decrease in the SA/volume ratio with the domestication events. Surprisingly, in the hexaploid wheat the observed ratio was similar to the domesticated tetraploid emmer suggesting that there was a genetic bottleneck during the latest plolyplodisation event. (Shall we say more on this?)

As the three species were independently domesticated and each was probably independently domesticated several times (Fuller, 2017) our data indicate that grain shape was changing to respond to agricultural methods, like for example the use of sieves to remove smaller grains, tillage and human consumption habits as rounder grains would be easier to mill and provide more flour.   Several modelling studies point to the importance of grain shape in breeding varieties with better yield and milling properties, suggesting that large spherical grains would be preferable (Evers, 1997). Due to the difficulty in accessing grain shape it wasn’t traditionally been a target of breeding programs. Our data suggest that depth is a major driver in grain shape and showed greater changes during the domestication events and could still be an important target for increasing productivity. Indeed, in Paragon, an elite variety, we still see increases in width and depth and a small reduction in grain length.

The observed grain shape changes could have been brought about by changes in the floret structure. The structure and size of the different components of the floret is known to have an effect on final fruit size and shape. In rice, it has been shown that changes in grain width are controlled by GW2 that has an effect on the hull and by SW5 that regulates the size of the outer glume (Song, 2007 SW5 reference). It will be interesting to determine if a  similar set of genes controls grain shape in wheat and if modifying these genes will lead to changes in floret and consequently in changes in grain shape. Interestingly, a QTLs recently identified in tetraploid wheat that controls seed shape is proposed to be *tenatious glume* that controls the toughness of the glume and lemma, which might constrain grain expansion in with and depth (Nalam, 2007 Okamoto).

**Modelling grain domestication**

Taking in account the relationship between the three grain dimensions, length, width and depth, a multiple regression model was generated that allows the accurate classification of grains as wild or domesticated. The model also provided further evidence of the importance of grain depth when analysing changes in grain size and shape. For einkorn, when depth was included in the model, it predicted the domestication status with 95% accuracy while if only length and width were considered the accuracy decreased to 31%.

Grain identification (species and domestication status) is very important in archaeological sites to provide information on the customs and practices of the associated human population. They provide evidence not only of the agricultural practices but also on the evolution of nutrition and give information on migration routes and cultural aspects (Jacomet et al. 2006). While increased grain volume is often used in the identification of domesticated grain (Fuller et al, 2007) it is quite difficult to measure accurately and for some species like barley there was a very small change in volume with domestication (Willcox, 2004) making difficult to distinguish from its wild relatives. The model we have developed would be very useful in these situations where given the traits of the wild and domesticated grains for each species could help classify the grains found at the archeological sites.

As mentioned above, while morphological traits have somehow informed breeding programs, grain shape has been overlooked due to the difficulty of obtaining fast and accurate measurements. Our data strongly support the use of 3D grain shape to inform breeding programs aiming at developing wheat and barley varieties with optimised grain shape by increasing grain width and depth. The combined use of the 3D imaging of mapping populations and modelling as a phenotypic classification tool would provide the phenotypic information necessary for QTL identification in mapping populations.

Being one of the most important staple food crops in the world there is pressing demand to increase wheat production while maintaining current land use and reducing fertilizer inputs. In practice this means increasing seed number per plant and increase seed size.  While there have been several studies looking at the genetics underpinning of grain size and shape, the tools to look at grain morphology to better inform the genetic studies have been lacking. We provide morphological evidence that grain depth has been under domestication pressure and could potentially be a route to increase yield. The use of grain 3D imaging of mapping populations open up the exciting possibility of identifying QTLs responsible for the control of grain depth. Recently, Yu et al. , 2018 (biorxiv) showed that some of the genetic element controlling grain weight are dominant alleles which could facilitate their identification.

**Methods**

**Plant Materials**

Four different wheat species were used in this study.  The diploids *Triticum monococcum* subsp. *aegilopoides* (wild einkorn) and *Trticium monococcum* subsp. *monococcum* (einkorn) and the tetraploids *Triticum turgidum* subsp. *dicoccoides* (wild emmer) and *Triticum trugidum* subsp. dicoccum (emmer). In all cases, spikes for scanning were obtained from the John Innes Centre collection and were chosen to encompass a wide range of origin habitats. For the barley one accession of the wild species (*Hordeum spontaneum*) and 2 for the domesticated 2 -row (*Hordeum vulgare*) were grown at the NPPC, Aberystwyth and mature fully dried spikes were selected. This was also the case for the test Einkorn population. 2D grain measurements for length, width, area and estimated TGW were performed on XXX grains per genotype using a MARVIN grain analyser .

For the growth staging, plants were scored according to Zadoks scale (Zadoks et al., 1974) and the days taken to reach GS39 (flag leaf ligule just visible), GS55 (ear half emerged) and GS65 (anthesis half way) noted. After fully ripen, the total weight of all the ears of the plant and the harvest index (calculated by dividing the total ear weight by the total plant weight) was determined.

**3D Scanning of Spikes**

From the genotypes selected, fully dried, representative spikes were chosen for micro-CT scanning. Spikes were placed in plastic holders (34x70mm tubes) and imaged using a μCT100 scanner (Scanco Medical, Switzerland).

Spikes were scanned with the X-ray power set at 45 kVp, 200 µA and 9 W with an integration time of 200 ms. Each spike was ~ 1000 slices (51 slices per stack), 125 projections/180° were taken and a binning of 6 was used. Output images were produced with a 0.2 megapixel (512 × 512) resolution (68.8 µm/pixel) in a proprietary ISQ file type format (Scanco Medical, Switzerland).

**Morphometric Feature extraction**

Feature extraction was performed using previously developed MATLAB-based software (Hughes et al. 2017) and freely available at <insert link to NPPC>. Changes were made to the watershedding and segmentation processes, of the existing pipeline, in order to work with the more complex primitive species of wheat (Supplementary Fig. 1).The features extracted were length (calculated using the major axis of the whole grain), width and depth (the major and minor axis of a cross section respectively, found by selecting the grain’s midpoint), volume (a complete connected pixel count per grain) and surface area (single pixel perimeter calculation mapped in 3 dimensions). An additional phenotype was created to describe the interaction between the geometric parameters; the interaction is described in following equation: geometry interaction=lengthxdepthxwidth

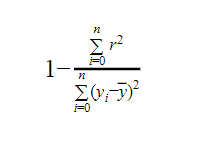
The data were checked for false positives, this is done by first removing outliers which are found by the 0.025 upper and lower percentiles of the data. Additionally for added robustness, constraints of expected grain limits were applied to the data based on findings from previous studies (Hughes et al. 2017).

**PCA analysis and Bayesian Modelling**

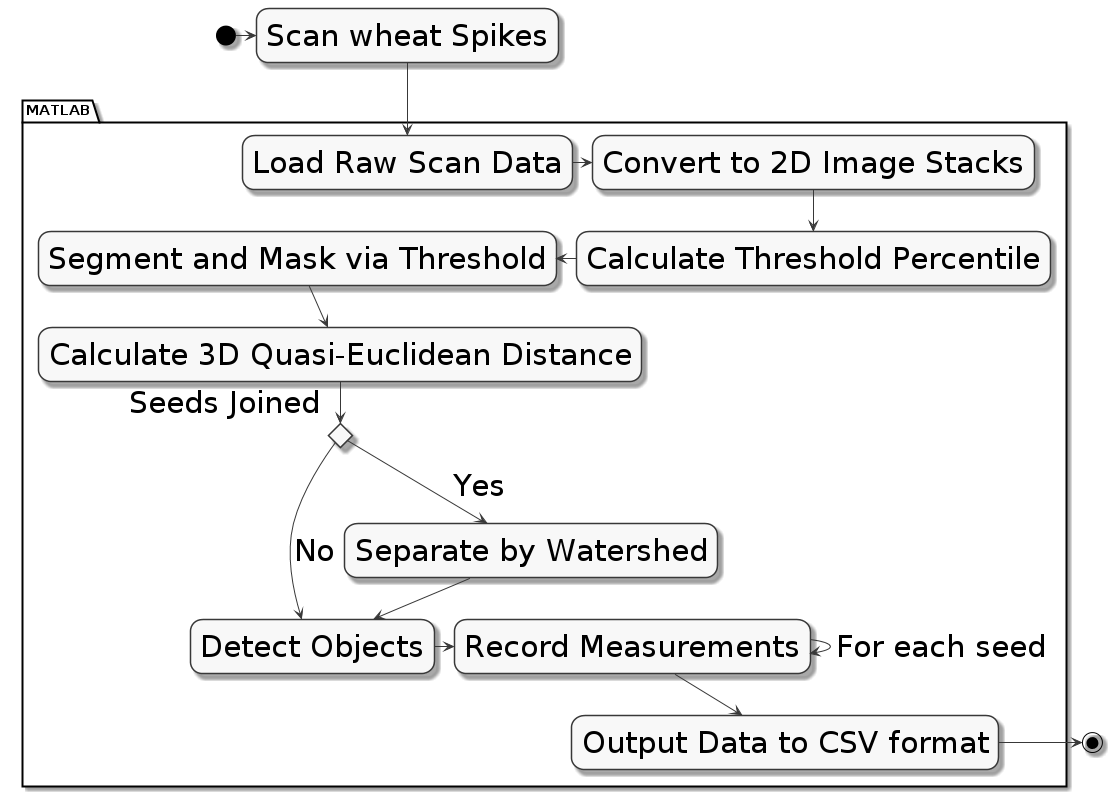
To provide deeper insight into the size of change or similarity in hypothesis testing, a Bayesian model is used. To estimate probability of two samples containing the same mean the method uses Bayes theorem (P(A|B)P(B|A)P(A)) (Kruschke 2013) along with Markov-Chain-Monte-Carlo (MCMC) to draw random samples from a normal population. In order to quantify differences in population means, a percentage likelihood of difference is produced using Krusches’ method . Significance analysis was performed using a Welch T-test for significance and confidence intervals. Principal Component Analysis (PCA) was performed using XXXX.

**Linear Modelling**

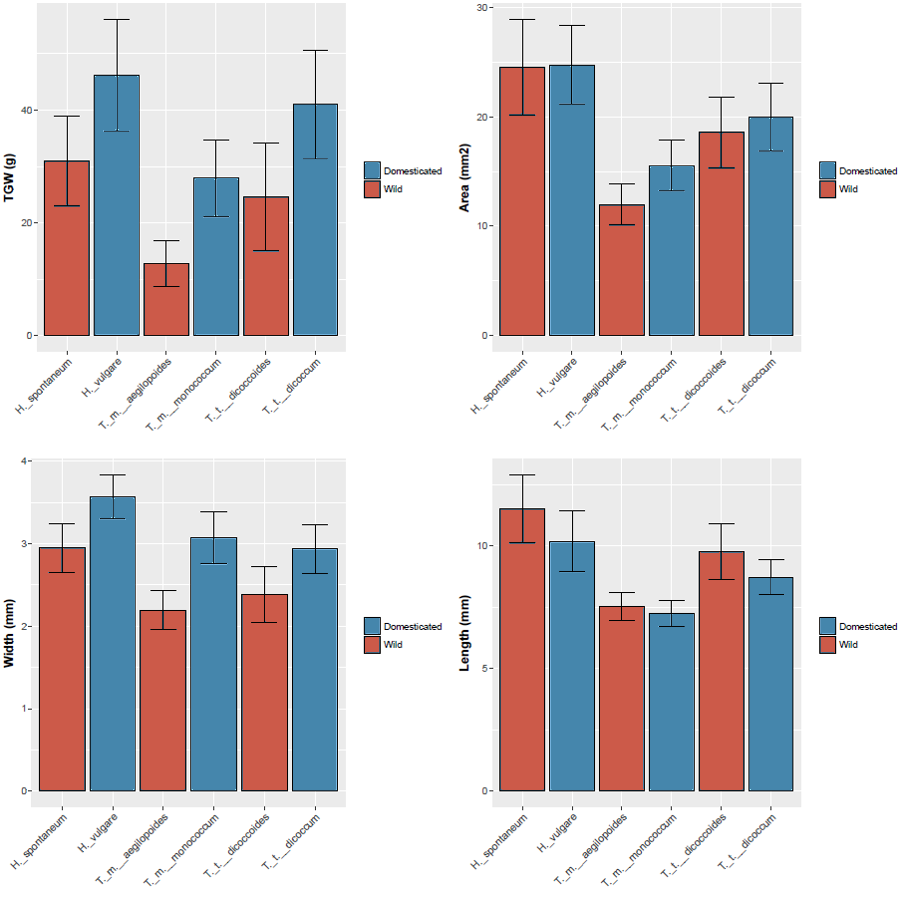
Each model was trained using 20% of the given data and 80% was used to test and calculate an r2 value. Estimation of domestication status was achieved through multiple regression (using ordinary least squares methods).The model produced an R2 value by using the sum of data variance over the explained sum of squares:



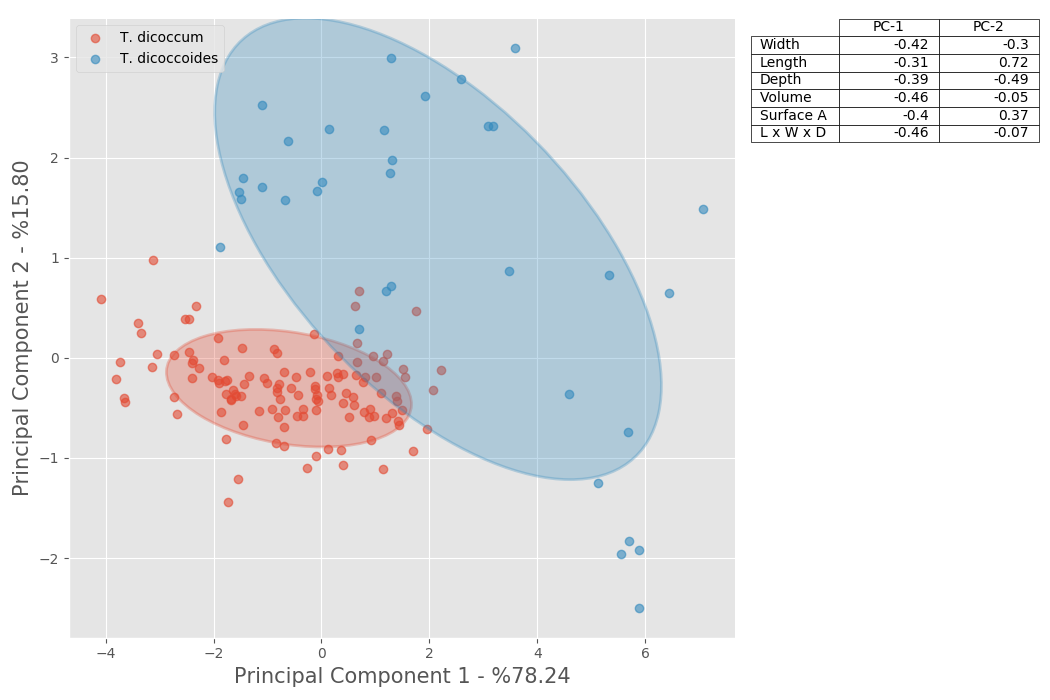
**Supplemental figures**



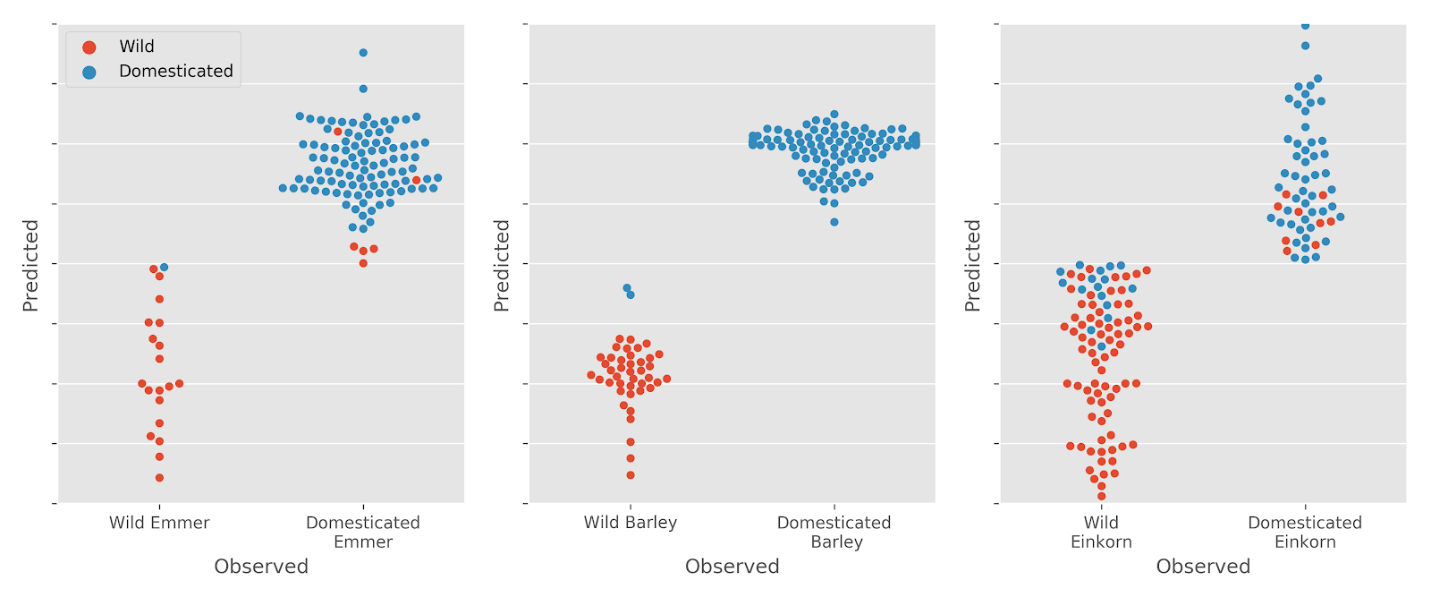
*Supplementary Figure 1- Schematics of the scanning, image analysis and feature extraction pipeline used in this study.*

**

*Supplemental Figure 2- 2D analysis of grain length, width, area and predicted thousand grain weight (TGW) of the indicated species using MARVIN imaging. Bars represent average ∓ SD for XXX grains.*

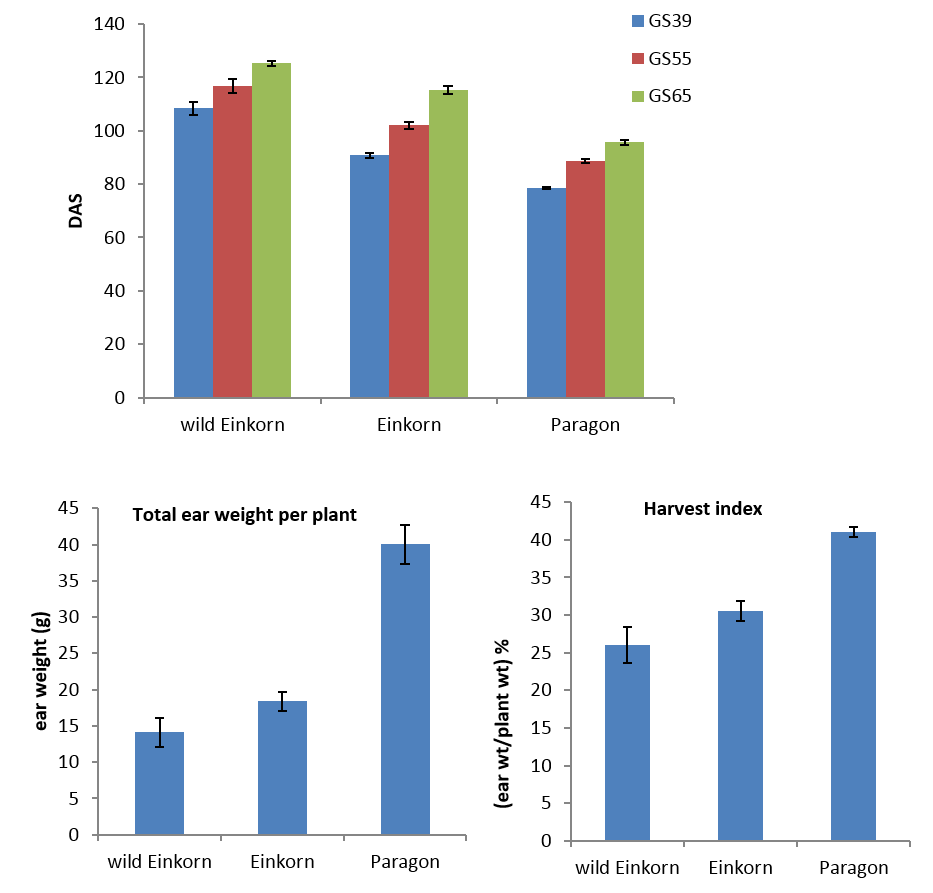


*Supplemental Figure 3-  Principle Component Analysis of grain traits for wild (red circles) and domesticated (blue circles) Emmer wheat. Two times standard deviation is represented by the shaded area.*

**

*Supplemental Figure 4- Domestication status modelling for emmer,  barley and einkorn grains. Multiple regression model shows the predicted (y axis) and actual (x axis) domestication status for each grain.  R^2 values were XXX for XXX grains for emmer, XXX for XXX grains for barley and XXX for XXX grains for the locally grown einkorn.*

*Supplemental Figure 5- Analysis of grain traits in wild (blue boxes) and domesticated (red boxes) einkorn wheat grown locally. Boxplots represent median plus the interquartile range for grain Volume (A), Length (B), Width (C), Depth (D). Asterisks indicate that the values are significantly different for p<0.01*

**

*Supplemental Figure 6- Analysis of domestication-related traits in wild and domesticated Einkorn wheat when compared with the modern hexaploid bread wheat variety Paragon. A- Number of days until growth stage GS39, GS55 and GS65 was reached for each species. B- Total ear weight at maturity. C- Harvest index for each species calculated as the total ear weight divided by the total plant weight.*

**Data Tables**

**einkorn**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **pval** | **tval** | **0.025** | **0.975** |
| **volume** | 0 | 22.3539 | 5.4784 | 6.4875 |
| **length** | 0.0281 | 2.2027 | 0.007 | 0.1777 |
| **width** | 0 | 15.1149 | 0.5128 | 0.6314 |
| **depth** | 0 | 35.6856 | 0.5291 | 0.5932 |
| **surface\_area** | 0 | 22.2893 | 10.9538 | 12.8747 |
| **length\_depth\_width** | 0 | 25.8976 | 13.416 | 15.5223 |
| **Surface Area - Volume** | 0 | -12.4141 | -0.3518 | -0.2675 |

*[fig:org62f1159] einkorn table*

**emmer**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **pval** | **tval** | **0.025** | **0.975** |
| **volume** | 0 | 5.8077 | 5.3877 | 9.4789 |
| **length** | 0.5387 | -0.6216 | -0.4893 | 0.1399 |
| **width** | 0 | 7.3497 | 0.3684 | 0.5672 |
| **depth** | 0 | 11.2744 | 0.5042 | 0.6613 |
| **surface\_area** | 0.3633 | 0.9213 | -1.1457 | 5.6405 |
| **length\_depth\_width** | 0 | 6.0425 | 11.6716 | 19.7904 |
| **Surface Area - Volume** | 0 | -9.0659 | -0.8766 | -0.6925 |

*[fig:org60d0038] emmer table*

**Barley**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **pval** | **tval** | **0.025** | **0.975** |
| **volume** | 0 | -5.6743 | -6.4365 | -2.4472 |
| **length** | 0 | 15.6396 | 1.8372 | 2.6041 |
| **width** | 0 | -10.841 | -0.6343 | -0.444 |
| **depth** | 0 | -28.0275 | -0.8857 | -0.7486 |
| **surface\_area** | 0.3145 | 1.0115 | -1.3518 | 4.0308 |
| **length\_depth\_width** | 0 | -5.3973 | -14.2941 | -4.9444 |
| **Surface Area - Volume** | 0 | 15.2648 | 0.398 | 0.5561 |

*[fig:org6fcf241] barley table*

**Domesticated 2N, 4N**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **pval** | **tval** | **0.025** | **0.975** |
| **volume** | 0 | -18.5341 | -9.0506 | -7.6633 |
| **length** | 0 | -12.3797 | -0.7084 | -0.4909 |
| **width** | 0 | -5.3098 | -0.1801 | -0.074 |
| **depth** | 0 | -27.1685 | -0.5452 | -0.463 |
| **surface\_area** | 0 | -11.4289 | -9.0581 | -6.6919 |
| **length\_depth\_width** | 0 | -17.4632 | -16.9121 | -14.0787 |
| **Surface Area - Volume** | 0 | 35.9876 | 0.4697 | 0.5535 |

*[fig:org778d7cd] domesticated 2N, 4N table*

**Wild 2N, 4N**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **pval** | **tval** | **0.25** | **0.975** |
| **volume** | 0 | -5.627 | -8.4026 | -5.4105 |
| **length** | 0.0042 | -3.095 | -1.1317 | -0.6016 |
| **width** | 0.0017 | -3.301 | -0.4302 | -0.0324 |
| **depth** | 0 | -9.508 | -0.5586 | -0.4063 |
| **surface\_area** | 0 | -7.3082 | -20.5777 | -14.506 |
| **length\_depth\_width** | 0 | -5.6696 | -17.3365 | -11.1307 |
| **Surface Area - Volume** | 0.6823 | 0.4128 | -0.1006 | 0.1741 |

*[fig:org11bf147] wild 2N, 4N table]*

Fuller, D. Q. 2007. “Contrasting Patterns in Crop Domestication and Domestication Rates: Recent Archaeobotanical Insights from the Old World.” *Annals of Botany* 100 (5). Oxford University Press: 903–24. doi:[10.1093/aob/mcm048](https://doi.org/10.1093/aob/mcm048).

Gegas, Vasilis C., Aida Nazari, Simon Griffiths, James Simmonds, Lesley Fish, Simon Orford, Liz Sayers, John H. Doonan, and John W. Snape. 2010. “A Genetic Framework for Grain Size and Shape Variation in Wheat.” *The Plant Cell* 22 (4): 1046–56. doi:[10.1105/tpc.110.074153](https://doi.org/10.1105/tpc.110.074153).

Gupta, Pushpendra K, Sachin Rustgi, and Neeraj Kumar. 2006. “Genetic and molecular basis of grain size and grain number and its relevance to grain productivity in higher plants.” doi:[10.1139/G06-063](https://doi.org/10.1139/G06-063).

Hughes, Nathan, Karen Askew, Callum P. Scotson, Kevin Williams, Colin Sauze, Fiona Corke, John H. Doonan, and Candida Nibau. 2017. “Non-Destructive, high-Content analysis of wheat grain traits using X-ray micro computed tomography.” *Plant Methods* 13 (1). doi:[10.1186/s13007-017-0229-8](https://doi.org/10.1186/s13007-017-0229-8).

Kruschke, John. 2013. “Bayesian estimation Supersedes the t-test.” *Journal of Experimental Psychology: General* 142 (2): 573–603. doi:[10.1037/a0029146](https://doi.org/10.1037/a0029146).

Piperno, Dolores R., Ehud Weiss, Irene Holst, and Dani Nadel. 2004. “Processing of wild cereal grains in the Upper Palaeolithic revealed by starch grain analysis.” *Nature* 430 (7000). Nature Publishing Group: 670–73. doi:[10.1038/nature02734](https://doi.org/10.1038/nature02734).

The MathWorks Inc. 2018. “MathWorks - Support - MATLAB & Simulink.” [https://es.mathworks.com/support.html?s{\\_}tid=gn{\\_}supp](https://es.mathworks.com/support.html?s%7B%5C_%7Dtid=gn%7B%5C_%7Dsupp).

Waines, J. G., and B. Ehdaie. 2007. “Domestication and Crop Physiology: Roots of Green-Revolution Wheat.” *Annals of Botany* 100 (5). Oxford University Press: 991–98. doi:[10.1093/aob/mcm180](https://doi.org/10.1093/aob/mcm180).

Willcox, George. 2004. “Measuring grain size and identifying Near Eastern cereal domestication: evidence from the Euphrates valley.” *Journal of Archaeological Science* 31 (2). Academic Press: 145–50. doi:[10.1016/J.JAS.2003.07.003](https://doi.org/10.1016/J.JAS.2003.07.003).

Abbo et al. 2014. Plant domestication versus crop evolution: a conceptual framework for cereals and grain legumes. Trends in Plant Science 19: 351-60.

Bonhomme V, Wallace M. 2016. The first shoots of a modern morphometrics approach to the origins of agriculture. Web Ecology 16(1-2).

Brown TA, Jones MK, Powell W, Allaby RG. 2009. The complex origins of domesticated crops in the Fertile Crescent. Trends in Ecology & Evolution 24(2): 103-109.

Butterworth KM et al. 2009. Initiation and early development of fiber in wild and cultivated cotton. Int. J. Plant Sci. 170(5):561–574.

Droust et al. 2014. Morphological diversity and genetic regulation of inflorescence abscission zones in grasses. American  Journal of Botany 101: 1759-1769.

Gros-Balthazard et al. 2016. The Domestication Syndrome in Phoenix dactylifera Seeds: Toward the Identification of Wild Date Palm Populations. PLoS ONE 11(3): e0152394.

Hammer K. 1984. Das domestikattionssyndrom. Kulturpflanze, 32 : 11-34.

Jones G, Valamoti S, Charles M. 2000. Early crop diversity: a “new” glume wheat from northern Greece. Vegetation History and Archaeobotany. 9(3):133-46.

Kislev ME. 1980. Triticum parvicoccum sp. Nov., the oldest naked wheat.   Israel Journal of Botany 28:95-107.

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(13):4101-8.

Meyer et al. 2012. Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. New Phytologist 196: 29–48.

Milla et al 2015. Plant domestication through an ecological lens. Trends in Ecology & Evolution 30: 463-469.

Mori N, Ohta S, Chiba H, Takagi T, Niimi Y, Shinde V, Kajale MD, Osada T. Rediscovery of Indian dwarf wheat (Triticum aestivum L. ssp. sphaerococcum (Perc.) MK.) an ancient crop of the Indian subcontinent. 2013. Genetic resources and crop evolution. 60(6):1771-5.

Nesbitt M (2002) When and where did domesticated cereals first occur in Southwet Asia? In: Cappers RTJ and Bottema S (eds.) *The Dawn of Farming in the Near East*. Berlin, ex oriente.

Preece et al. 2016. How did the domestication of Fertile Crescent grain crops increase their yields? Functional Ecology2017,31, 387-97.

Roucou A, Violle C, Fort F, Roumet P, Ecarnot M, Vile D (2018) Shifts in plant functional strategies over the course of wheat domestication. Journal of Applied Ecology. 55(1):25-37.

Stetter MG, Gates DJ, Mei W, Ross-Ibarra J. How to make a domesticate. 2017. Current Biology. 27(17):R896-900.

Terral et al. 2010. Evolution and history of grapevine (Vitis vinifera) under domestication: new morphometric perspectives to understand seed domestication syndrome and reveal origins of ancient European cultivars. Annals of Botany 105: 443–455.