Domestication Draft

NH, CN, HO, JD

August 13, 2018

1 Introduction

<Waiting for Hugo>

2 **TODO**

☐ Report on standard deviations and errors, would be very interesting to highlight changes in this from prim to domesticated.

3 Results

Previous research demonstrated that certain species of wheat underwent a significant (a two-thirds increase) biomass shift during domestication [1]. Several more studies have confirmed the expected grain filling, volume and size change undergone during domestication [2, 3], but have been unable to provide detailed explanation of the effects leading to the observed changes in grain morphology.

Four key species, two diploid (Einkorn) and two tetraploid (Emmer) were selected to study the effect on grain morphology across domestication and polyploidisation. Domesticated *T. monococcum* and primitive *T. boeticum* were selected for being diploid; Domesticated *T. dicoccum* as primitive *T. dicoccoides* were chosen as representative species for tetraploids. (See Figure 1)

Key grain morphological traits were measured using micro-CT imaging, 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 ratio these are used to provide insight to potential trait interactions. Of these traits, length and width have already been highlighted as having vital importance in building genetic frameworks for wheat [4].

Traits were gathered using image analysis software [5], major modifications were made to grain identification algorithms in order to adapt to the wide variety in the species studied.

Analysis was performed using a Welch T-test for significance and confidence intervals. Kruchke's method was used for Bayesian estimation of difference [6] in order to quantify differences in population means. Estimation of domestication status was achieved through multiple regression (using ordinary least squares methods).
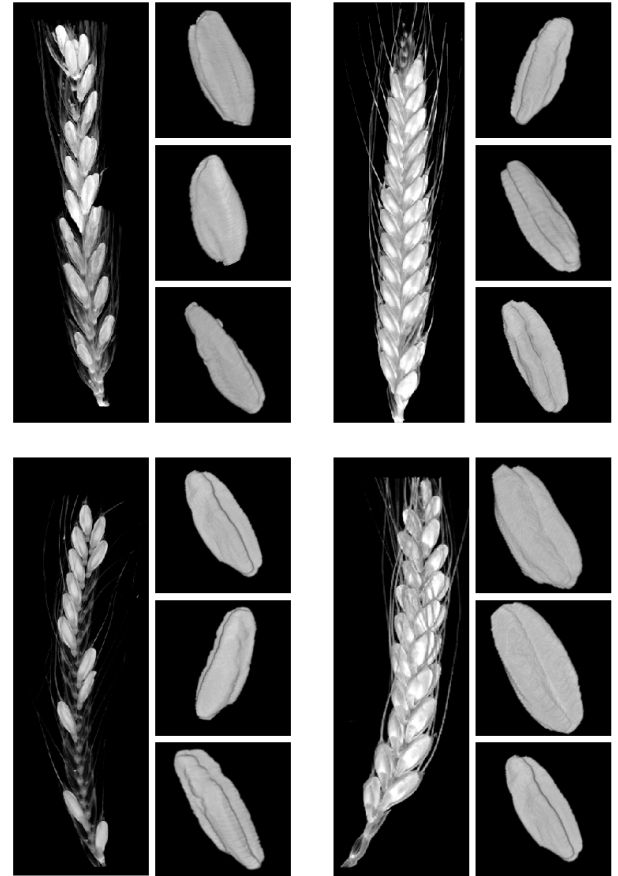

Figure 1: Einkorn (top) and Emmer (bottom) Wheat (Primitives left, Domesticated right)

Principle Component Analysis (PCA) was used to highlight separation of species based on grain measurements, in addition to the effect of specific traits in model though reported eigenvalues/coefficient loadings.

3.1 Einkorn

3.1.1 Grain Traits

No noticeable change has been observed in the length attribute between domesticated and wild einkorn ($p = 0.02$ and a predicted 35% probability of overlapping averages).
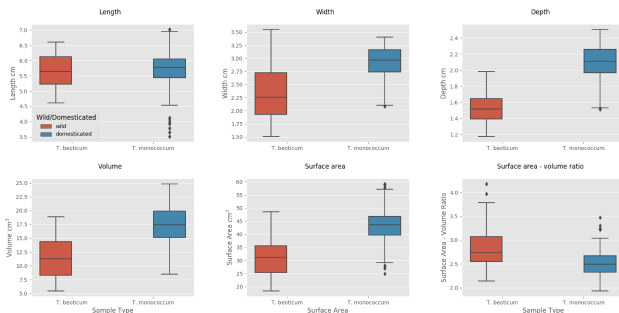

Figure 2: Einkorn Traits

However, other traits(figure:2): length, width, depth, volume and surface area all have a significant change between wild and domesticated ($p < 0.01$).

3.1.2 Principle Component Analysis

A two component PCA reviled that no dominant trait appears to influence morphometric variation in einkorn species (figure:3).

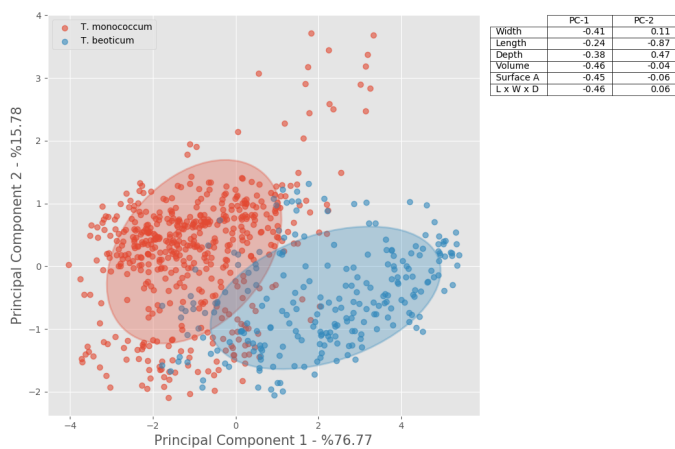|	PC-1	PC-2
Width	-0.41	0.11
Length	-0.24	-0.87
Depth	-0.38	0.47
Volume	-0.46	-0.04
Surface A	-0.45	-0.06
L x W x D	-0.46	0.06

Figure 3: Einkorn PCA

With the exception of grain length (-0.24), all other measurements (width, depth, volume, surface area and the interaction term LxWxD) have coefficients ranging -0.38 to -0.46 in the first principle component (76.77% of explained variation).

3.2 Emmer

3.2.1 Grain Traits

With length, Emmer displays no significant difference in grain length between wild and domesticated species ($p = 0.93$).
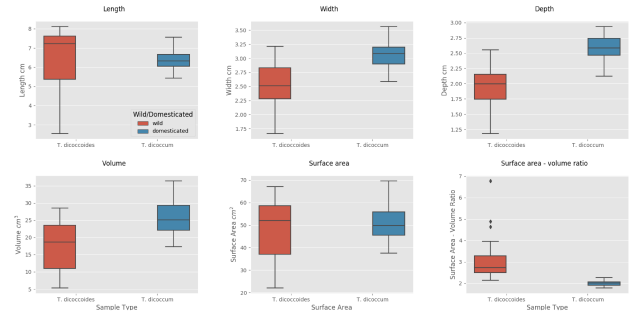
Figure 4: Emmer Traits

Additionally, surface area of grain is reported as non-significant ($p = 0.17$). Showing that grain compactness has significantly altered to preserve this trait (Figure: 4). The Bayesian model predicted a 20% probability of domesticated and wild types differing, showing that there is an indication of change. Though, with a high enough probability of this being down to chance.

The traits width, depth and volume have all shown significant change during Emmer domestication ($p < 0.01$).

3.2.2 Principle Component Analysis

A two component PCA, 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 moderate influence in grain morphology (Figure:5); length shows the lowest impact ($coef = -0.31$).

	PC-1	PC-2
Width	-0.42	-0.3
Length	-0.31	0.72
Depth	-0.39	-0.49
Volume	-0.46	-0.05
Surface A	-0.4	0.37
L x W x D	-0.46	-0.07

Figure 5: Emmer PCA

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.

3.3 Ploidy Effects

3.3.1 Wilds

In grain length, wild einkorn and wild emmer have shown to be just overlapping ($p = 0.025$) with a 9.4% probability of similar means (Figure: 6).

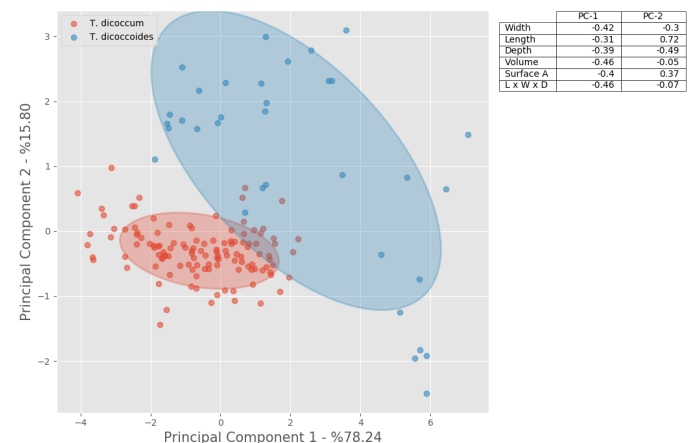Width follows a similar pattern with length, being non-significant ($p = 0.02$). A high probability exists that the averages of these wild types are similar (27% probability).

Grain depth, volume and surface area all report highly significant changes across ploidy ($p < 0.01$).
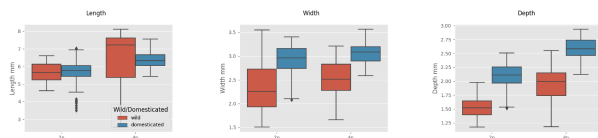

Figure 6: Ploidy Effects

3.3.2 Domesticates

Between 2N and 4N domesticated species all grain parameters (length, width, depth, volume and surface area) are shown to have a highly significant change ($p < 0.01$).

Whilst statistically significant, grain width has shown a high probability of 28%, that the mean between 2N and 4N overlaps. A wide standard deviation in einkorn provides this confusion.

All other measurements show a $< 5\%$ probability of overlapping values.

3.4 Modelling

A multiple regression model was used to obtain predictions of domestication using three key parameters (grain length, width and depth). A high variance was explained ($R^2 = 0.95$) using equation 1 and an ordinary leas squares method (OLS).

$$Y = \beta_0 + \times \beta_1 length \times \beta_2 depth \times \beta_3 width + \epsilon$$

Here, in addition to domestication status this model has differentiated across ploidy. Wild and primitives of einkorn and emmer are accurately predicted (figure:7).

The model produced an R2 value by using the sum of data variance over the explained sum of squares:
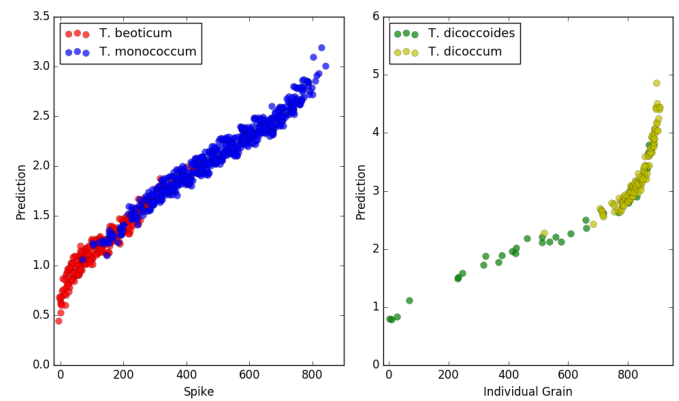
Figure 7: is showing a multiple regression of R=0.95 by using length, width and depth to correctly ID domestication status.

$$1 - \frac{\sum_{i=0}^{n} r^2}{\sum_{i=0}^{n} (y_i - \overline{y})^2}$$

4 Methods

4.1 Materials

< Plants which were used > . . .

4.1.1 How they were sourced

4.1.2 Any additional information

4.2 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 a CT100 scanner (Scanco Medical, Switzerland).

The imaging system was configured with an X-ray source ranging from 20 to 100 kVp, a detector of 3072 x 400 elements. A resolution of 68.8 micro-meters per pixel was used for all scans.

4.3 Computational Methods

Using software developed for previous wheat studies by the National Plant Phenomics Centre ([5]) . New and novel additions are implemented in the watershedding and segmentation processes, of the existing pipeline, in order to work with the more complex primitive species of wheat.

Due to the optimised resolution of the imaging technique (68.6μ meters per pixel) objects can appear connected which are not, particularly in primitive grain. A three dimensional watershedding algorithm is used to

correct any objects which appear connected when they should not be.

The software, developed in MATLAB ([7]), is freely available at <insert link to NPPC>.

4.3.0.1   Pipeline   The scanning and MATLAB routine pipeline:
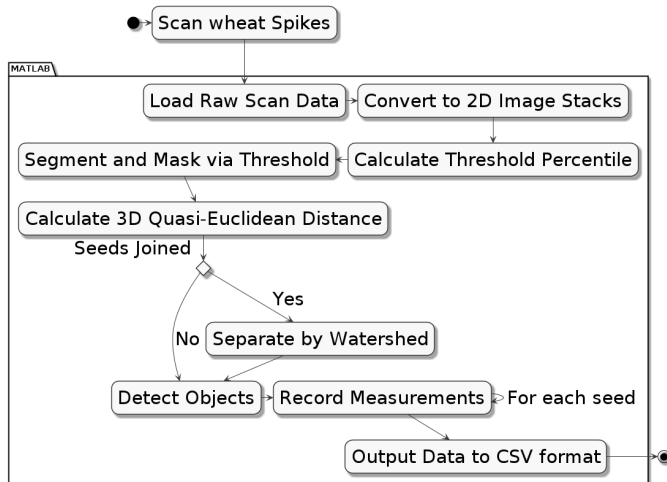
Figure 8: Pipeline

4.3.1   Morphometric Features

The features/phenotypes used are extracted during the imaging process.

- Length is calculated using the major axis of the whole grain.

- Width and depth are the major and minor axis of a cross section, found by selecting the grain's midpoint.

- Volume is a complete connected pixel count per grain

- Surface area is a single pixel perimeter calculation mapped in 3 dimensions

- Length   X   Depth   X   Width   is   a post-image-processing   value   calculated   by the   interaction   between   the   three   dimension descriptors.

Values used in statistical functions and measurements are presented as metric units, derived from μ-CT image pixel values. The equation:1 is presented here.

$$mm = \frac{pixel \times 68.8}{1000} \tag{1}$$

4.3.2   Error Removal

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 constraints are applied to the data based on findings from previous studies [5], this adds robustness.

4.3.3   LWD

An additional phenotype is created to describe the interaction between the geometric parameters; the interaction is described in equation:2.

$$\text{geometry interaction} = length \times depth \times width \tag{2}$$

4.3.4   Image Analysis Methods

I could provide a lot of info on this, but weary of going off-track, more can be added (or taken away) if needed, what's in comp methods could be enough I think.

4.4   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 $(P(A|B) \propto P(B|A) \times P(A))$ [6] along with Markov-Chain-Monte-Carlo (MCMC) to draw random samples from a normal population.

From Krusches' method a percentage likelihood of difference is produced.

4.5   Linear Modelling

A linear model allowed for an R^2 value of 0.91 in einkorn species when predicting domestication status.

4.5.1   Draft   Supplemental   figure   for   volume
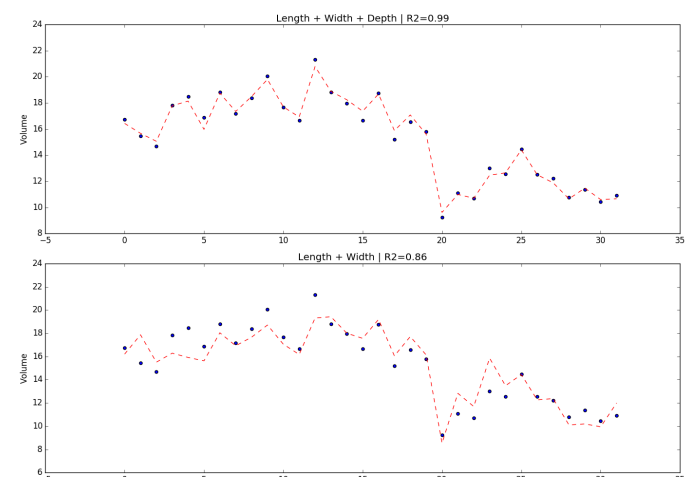
Figure 9: Showing the importance of 3rd dimension of depth

5 Misc Information to fold into discussion?

Fuller has evidenced that grain volume is used when initially identifying wheat grain which is recovered as being domesticated [8] . Therefore justification for our model is very useful.

On the other hand [9] found that barley is much harder to identify from it's domesticated relatives. Volume is not significantly different! Which fits perfectly with our results!

Wondering outloud, with the wild 2n/4n an oberserved change in surface area. Would there be a reason for why grains would want to develop a larger surface area? Alternatively we can work off-of volume, as that changes also

6 Data Tables

6.1 einkorn

	pval	tval	diff_mean	0.25	0.975	Bayesian % likehood of similarity
length	0.0256	2.2389	0.094	0.0084	0.1788	0.365
width	0	15.2051	0.575	0.5154	0.634	0
depth	0	35.7882	0.562	0.5299	0.5939	0
volume	0	22.4279	6.005	5.501	6.5098	0
surface_area	0	22.3852	11.938	10.9794	12.8972	0
length_depth_width	0	25.9884	14.505	13.4528	15.5568	0
mean_length	0.1843	1.3591	0.128	-0.073	0.3286	0.49125
mean_width	0	6.9796	0.514	0.3809	0.6469	0.30915
mean_depth	0	7.4806	0.500	0.3662	0.6329	0.2112
mean_volume	0	9.1114	5.648	4.4402	6.8552	0.0657
mean_surface_area	0	12.885	11.363	9.6065	13.1195	0.17955
mean_length_depth_width	0	9.3852	13.289	10.5195	16.0577	0.02475
density	0	52.571	0.317	0.3018	0.3322	0

Einkorn

Figure 10: einkorn table

6.2 emmer

	pval	tval	diff_mean	0.25	0.975	Bayesian % likehood of similarity
length	0.9333	0.0843	0.024	-0.3037	0.3521	0.49165
width	0	7.7512	0.526	0.4244	0.6267	0.0638
depth	0	10.9044	0.640	0.5572	0.7229	0.00465
volume	0	6.4316	8.488	6.4431	10.5334	0
surface_area	0.1708	1.3959	3.391	0.0088	6.7725	0.22565
length_depth_width	0	6.6617	17.900	13.8293	21.9704	0
mean_length	0.9405	0.0797	0.022	-0.5466	0.5914	0.4972
mean_width	0.0437	3.4882	0.561	0.2832	0.8387	0.43615
mean_depth	0.0415	3.7346	0.689	0.3845	0.9926	0.3709
mean_volume	0.0201	3.3848	9.251	3.0649	15.4361	0.26695
mean_surface_area	0.2389	1.2933	4.131	-5.1452	13.4072	0.48195
mean_length_depth_width	0.0132	3.7335	19.299	7.4551	31.1422	0.25005
density			0.922	0.7566	1.0883	NaN

Emmer

Figure 11: emmer table

6.3 Barley

	pval	tval	diff_mean	0.25	0.975	Bayesian % likehood of similarity
length	0	21.4151	2.122	1.7956	2.4477	0
width	0	-7.1618	-0.329	-0.4419	-0.2157	0.14075
depth	0	-20.02	-0.501	-0.5896	-0.4127	0.0087
volume	0.1854	-1.3383	-0.782	-2.4715	0.9084	0.3003
surface_area	0.0001	4.3757	5.121	2.5149	7.7268	0.11745
length_depth_width	0.2952	-1.0552	-1.346	-5.084	2.3913	0.32525
mean_length	0	10.826	2.086	1.5037	2.6685	0.1957
mean_width	0.0039	-3.8732	-0.559	-0.7832	-0.3352	0.3283
mean_depth	0	-9.3088	-0.640	-0.8234	-0.4567	0.22105
mean_volume	0.0215	-2.5064	-3.594	-7.1728	-0.0157	0.29825
mean_surface_area	0.7265	0.3601	0.900	-3.3466	5.1467	0.4825
mean_length_depth_width	0.0249	-2.4229	-8.036	-16.5232	0.4512	0.2521
density	0	-26.6912	-0.529	-0.6338	-0.4251	0

Barley

Figure 12: barley table

6.4 Domesticated 2N, 4N

	pval	tval	diff_mean	0.25	0.975	Bayesian % likehood of similarity
length	0	-12.3797	-0.600	-0.7084	-0.4909	0.04945
width	0	-5.3098	-0.127	-0.1801	-0.074	0.28765
depth	0	-27.1685	-0.504	-0.5452	-0.463	0.00025
volume	0	-18.5341	-8.357	-9.0506	-7.6633	0
surface_area	0	-11.4289	-7.875	-9.0581	-6.6919	0.0015
length_depth_width	0	-17.4632	-15.495	-16.9121	-14.0787	0
mean_length	0.0077	-3.6159	-0.615	-0.9231	-0.3068	0.4557
mean_width	0.1701	-1.5109	-0.111	-0.2495	0.0276	0.4957
mean_depth	0.0001	-6.3457	-0.510	-0.6833	-0.337	0.34185
mean_volume	0.0049	-4.5541	-8.250	-10.5045	-5.9951	0.20235
mean_surface_area	0.0325	-2.8588	-8.406	-11.9889	-4.8228	0.4036
mean_length_depth_width	0.0051	-4.4248	-15.579	-20.2621	-10.8958	0.2053
density	0	-12.6033	-0.562	-0.6091	-0.5155	0

Dom Einkorn / Dom Emmer

Figure 13: domesticated 2N, 4N table

6.5   Wild 2N, 4N

	pval	tval	diff_mean	0.25	0.975	Bayesian % likehood of similarity
length	0.0253	-2.3408	-0.669	-0.9386	-0.3996	0.09445
width	0.0205	-2.3854	-0.176	-0.3678	0.0154	0.2745
depth	0	-7.3629	-0.426	-0.5029	-0.3491	0.03255
volume	0	-4.6292	-5.874	-7.3684	-4.3797	0
surface_area	0	-6.8727	-16.423	-19.3996	-13.4457	0
length_depth_width	0	-4.6595	-12.100	-15.1929	-9.008	0
mean_length	0.084	-2.9689	-0.720	-1.0324	-0.4083	0.46945
mean_width	0.7187	-0.3976	-0.064	-0.3723	0.2444	0.4951
mean_depth	0.1949	-1.7954	-0.321	-0.5765	-0.0657	0.4351
mean_volume	0.1486	-2.1734	-4.647	-7.2984	-1.9955	0.3327
mean_surface_area	0.0019	-10.236	-15.638	-18.7554	-12.5203	0.32425
mean_length_depth_width	0.123	-2.3681	-9.569	-15.2499	-3.888	0.30845
density			0.043	0.0249	0.0614	NaN
		Wild Einkorn / Wild Emmer				

Figure 14: wild 2N, 4N table

References

[1] J. G. Waines and B. Ehdaie. Domestication and Crop Physiology: Roots of Green-Revolution Wheat. *Annals of Botany*, 100(5):991–998, jul 2007.

[2] Pushpendra K Gupta, Sachin Rustgi, and Neeraj Kumar. Genetic and molecular basis of grain size and grain number and its relevance to grain productivity in higher plants. 2006.

[3] Dolores R. Piperno, Ehud Weiss, Irene Holst, and Dani Nadel. Processing of wild cereal grains in the Upper Palaeolithic revealed by starch grain analysis. *Nature*, 430(7000):670–673, aug 2004.

[4] Vasilis C. Gegas, Aida Nazari, Simon Griffiths, James Simmonds, Lesley Fish, Simon Orford, Liz Sayers, John H. Doonan, and John W. Snape. A Genetic Framework for Grain Size and Shape Variation in Wheat. *The Plant Cell*, 22(4):1046–1056, 2010.

[5] Nathan Hughes, Karen Askew, Callum P. Scotson, Kevin Williams, Colin Sauze, Fiona Corke, John H. Doonan, and Candida Nibau. Non-Destructive, high-Content analysis of wheat grain traits using X-ray micro computed tomography. *Plant Methods*, 13(1), 2017.

[6] John Kruschke. Bayesian estimation Supersedes the t-test. *Journal of Experimental Psychology: General*, 142(2):573–603, 2013.

[7] The MathWorks Inc. MathWorks - Support - MATLAB & Simulink, 2018.

[8] D. Q. Fuller. Contrasting Patterns in Crop Domestication and Domestication Rates: Recent Archaeobotanical Insights from the Old World. *Annals of Botany*, 100(5):903–924, jul 2007.

[9] George Willcox. Measuring grain size and identifying Near Eastern cereal domestication: evidence from the Euphrates valley. *Journal of Archaeological Science*, 31(2):145–150, feb 2004.