Modelling the effects of domestication in Wheat through novel computer-vision techniques

Nathan Hughes (nah26@aber.ac.uk)

Introduction

This project aims to create a set of data, of novel phenotypic information, using a computer vision technique I developed at the National Plant Phenomics Centre for examining wheat grains using μ -Computed Tomography [2]. This technique allows for a detailed extraction of 3-dimensional morphometric data of seeds in wheat and other crops.

I have been able to gain access to a wide range or primitive, non-domesticated wheat varieties as well as their cultivated relatives. These genotypes also provide a nice gradient of ploidy which could also provide great insight into how genome structure effects grains. I aim to use the outlinned method to extract data from these crops and create statistical analysis software to answer various research questions.

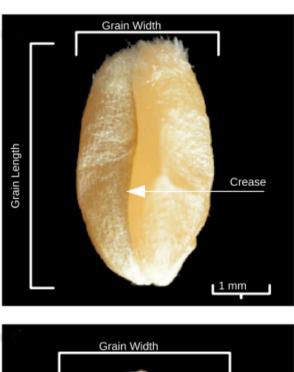
Research Aims

Primarily I have three main null-hypothesis which I hope this project will be able to reject:

- $ightharpoonup H_0$ = Domestication has no effect on the morphometric properties of wheat
- $ightharpoonup H_0$ = Ploidy has no effect on the morphometric properties of wheat
- $ightharpoonup H_0$ = There is no difference in hulled and non-hulled genotypes

Grain Parameters

The technique being used to produce these data provides a number of attributes for me to use in analysis. The primary ones can be seen in the figure 1



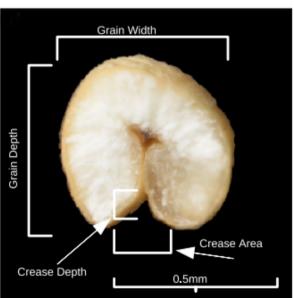


Figure 1: wheat grains labelled with data



Figure 2: wheat Spikes original (left) and scanned (right)

Software Development Aims

In order to make this work reproducible, as I believe it will justify many further studies I am creating a software library which acts as a specialist tool for handling the type of data which CT-Scanners output.

To do this I am using the PyQT library to create a GUI framework which will be useable by non-programmers in order to test hypothesis and build systematic models of the given input data.

Future Work

My main aims for the future is to fully implement my statistical library, once completed I can focus on testing my outlined hypothesis. I will also be aiming to make use of the *scipy* data models packages in Python in order to better model the data and to search for more vital underlying meanings in the data I've created.

Once this library is more complete, I shall be joining it into the GUI I have developed to allow for quick, easy and automatic analysis of future experiments. Should this project justify the need for more information, then it should be easily added through this additional application that I am developing with PyQt.

In order to address a broader picture, and to widen the scope of this research I am in the process of obtaining data used by Gegas et.al in 2011 to examine 2-Dimensional morphometric traits for a wide array of wheat that correlates nicely with what I have produced so far [1].

This would hopefully provide a lot of interesting information if the datasets were to compliment each other.

Current Progress

My work to date has been primarily focused on modifying and adapting my 3D analysis software to fit the required parameters to work with such a varied data source.

Additionally I have been building a data processing library whichs aims to:

- simplify data input
- perform data cleaning
- automatically generates plots
- handles data from multiple sources
- adding additional experiment parameters

Figure 3 shows one example of the auto analysis which my library can perform in its early stages of development.

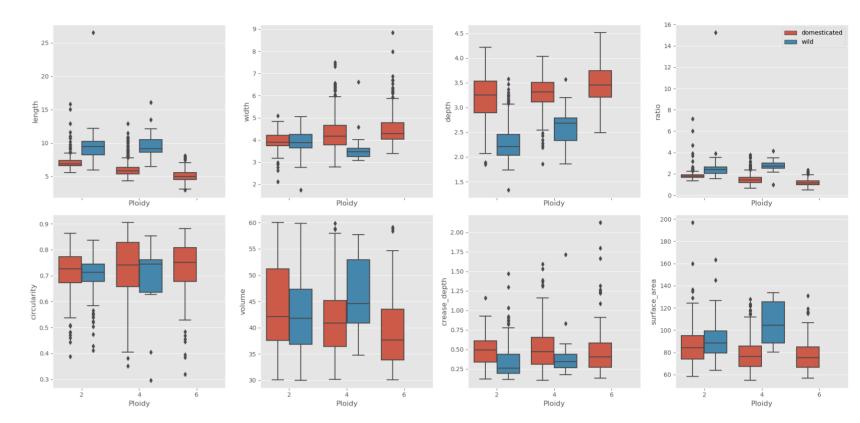


Figure 3: Initial Data analysis of Domestication and Ploidy

Statistical Methods and Testing

In order to answer these questions a number of statistical tests will be required, as will numerous elements of data science. I am planning to use:

- ► ANOVA
- MANOVA
- $\rightarrow \chi^2$ Tests

- ► PCA
- ► GLM
- ► T/F-Tests

From the data in Figure:3 you can see that the data is messy before processing. I have used QQ-Plots in Figure:4 to illustrate the non-parametric nature of the data. This is a problem I hope to solve in the future.

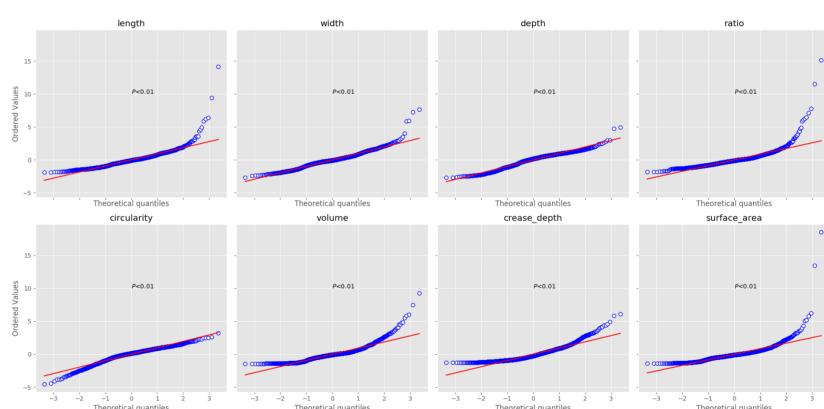


Figure 4: QQ-Plots showing Distribution

Futher Information

For more information on this project I recommend keeping up-to-date with my development blog which can be found at http://www.users.aber.ac.uk/nah26

Acknowledgements and Thanks

Thank you for the vital help and support from the following people and organisations.

- ► Dr. Wayne Aubrey
- ► Prof. John Doonan
- ► BBSRC

- ► Dr. Candida Nibau
- Dr. Hugo Oliveira
- ► Dr. Kevin Williams

References

- V. C. Gegas, A. Nazari, S. Griffiths, J. Simmonds, L. Fish, S. Orford, L. Sayers, J. H. Doonan, and J. W. Snape. A genetic framework for grain size and shape variation in wheat. The Plant Cell, 22(4):1046–1056, 2010.
- [2] N. Hughes, K. Askew, C. P. Scotson, K. Williams, C. Sauze, F. Corke, J. H. Doonan, and C. Nibau. Non-destructive, high-content analysis of wheat grain traits using x-ray micro computed tomography. *Plant Methods*, 13(1), nov 2017.
- [3] H. R. Oliveira, M. G. Campana, H. Jones, H. V. Hunt, F. Leigh, D. I. Redhouse, D. L. Lister, and M. K. Jones. Tetraploid wheat landraces in the mediterranean basin: Taxonomy, evolution and genetic diversity. *PLoS ONE*, 7(5):e37063, may 2012.