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

Nathan Hughes (nah26@aber.ac.uk)

Introduction

My Project's goal is to examine the effects of domestication in Wheat through the use of μComputed Tomography. This will be done by using a novel technique for a detailed extraction of 3-dimensional morphometric data of wheat grain [1].

I have a wide range or primitive, non-domesticated wheat varieties as well as their cultivated relatives. These genotypes also provide a nice gradient of genome shape and size (Ploidy) which could also provide great insight into how genome structure effects grains.

I aim to modify, improve and use the outlined method to extract data from these wheat varieties to create a unique data-set. I shall then create statistical analysis software which can reliably, produce answers to hypothesis aimed at these data.

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

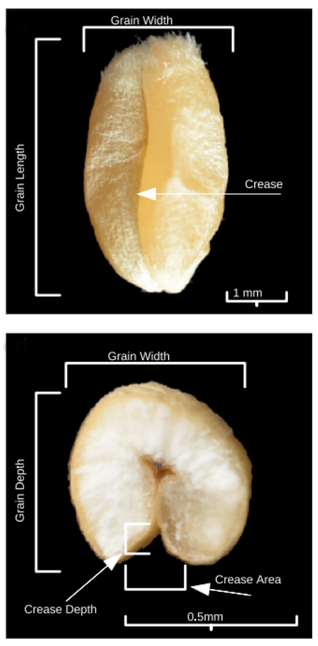

Figure 1: wheat grains labelled with data

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

Software Development and Pipeline

To make this research reproducible, for use in follow-up and future experiment analysis, I am creating a software library which acts as a specialist pipeline for handling CT-Scanner image data.

Figure:3 Illustrates a typical pipeline in which data will flow in this system. Primarily the process aims at taking raw uninteresting data and producing meaningful and useful output.

The pipeline will also allow a researcher to specify if they are interested in cleaning the data or if the outliers are of particular interest, additional options for joining in other external experiment information will also be possible in the final design, as flexibility is key.

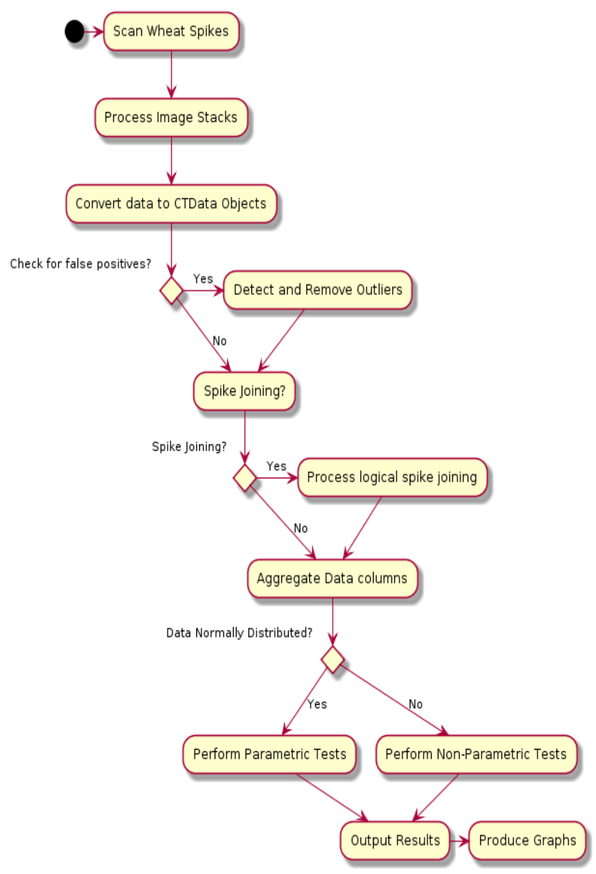

Figure 3: Data Processing Software Pipeline

Future Work

My main aim for the future is to fully implement the proposed 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.

Further 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 Also with Dr. Hugo Oliveira's work which can be found on Google scholar provides a great deal of insight into the primitive species and their phylogeny [2].

Research Aims

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

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

Current Progress

My work to date has been primarily focused on modifying and adapting 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 4 shows one example of the auto analysis which my library can perform in its early stages of development.

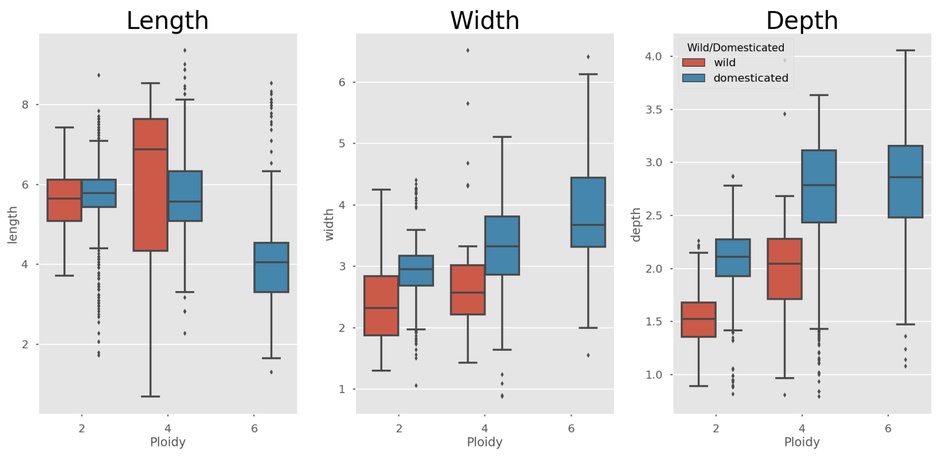

Figure 4: Initial Data analysis of Domestication and Ploidy

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, all of which will be incorporated into my analysing software library (Figure:3). I am planning to use:

- ANOVA
- MANOVA
- χ^2 Tests
- PCA
- GLMs
- T/F-Tests

From the data in Figure:4 you can see that the data is messy before processing. I have used QQ-Plots in Figure:5 to illustrate the non-parametric nature of the data. This is a problem I hope to solve in the future.

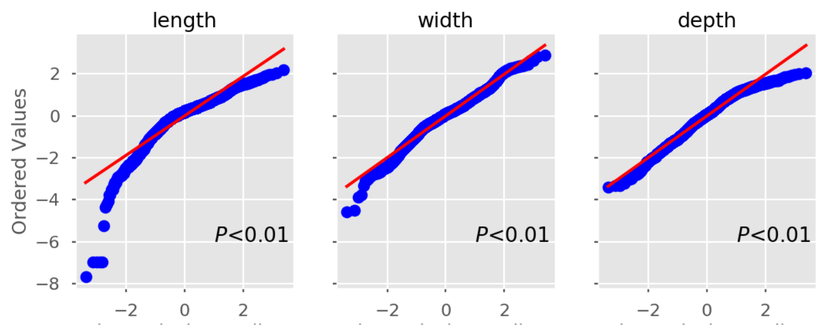

Figure 5: QQ-Plots showing Distribution

Acknowledgements and Thanks

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

- Dr. Wayne Aubrey
- Dr. Candida Nibau
- Prof. John Doonan
- Dr. Hugo Oliveira
- BBSRC
- Dr. Kevin Williams

References

[1] 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), nov 2017.

[2] Hugo R. Oliveira, Michael G. Campana, Huw Jones, Harriet V. Hunt, Fiona Leigh, David I. Redhouse, Diane L. Lister, and Martin K. Jones.
Tetraploid wheat landraces in the mediterranean basin: Taxonomy, evolution and genetic diversity.
PLoS ONE, 7(5):e37063, may 2012.