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

Nathan Hughes

March 13, 2018

Outline

What is the project?

The project is aiming to use computational methods to answer biologically significant questions on wheat grain morphology and domestication using μCT images.

To do this, I will be using:

- Computer vision on 3D image sets
- Statistical analysis and data science
- Scientific theory to create reproducible results

Research Question:

Is it possible to use µCT imaging to answer questions about Wheat domestication?

I hope so!

Null-Hypothesis

- 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

**Hulling is a trait associated with domestication*

Outline

Materials (Plant)

I have a wide range of Wheat genotypes, these are:

- Ranged between diploid, tetraploid and hexiploid
- 12 total genotypes
- Divided between domestication status

Figure 1: A Primitive Wheat Spike

Methods (Image Generation)

Capture Methods

In order to generate this data a μCT100 scanner (ScancoMedical, Switzerland) was used Each spike was:

- ~1000 slices (51 slices per stack),
- 125 projections/180° were taken and a
- binning of 6 was used.
- Output images are a 0.2 megapixel (512 x 512) resolution (68.8 μ/pixel)

Outline

Data Gathering

MATLAB Software

Data has been extracted from the μCT images using a method I previously developed at IBERS [1]. A lot of additions and novel augmentations have been added in order to work with the range of Wheat genotypes.

- This is done via separating non-grain from grain ROIs in the image data
- Measurements are taken on a per grain level
- Measurements are in mm, mm^2 and mm^3 based on type
- Data is exported as CSV files

Improved Watershedding Algorithm

Algorithm

$$|x_1 - x_2| + (\sqrt{2} - 1), |x_1 - x_2| > |y_1 - y_2| \, (\sqrt{2} - 1) \, |x_1 - x_2| \, , \text{otherwise} \tag{1}$$

Visual Example

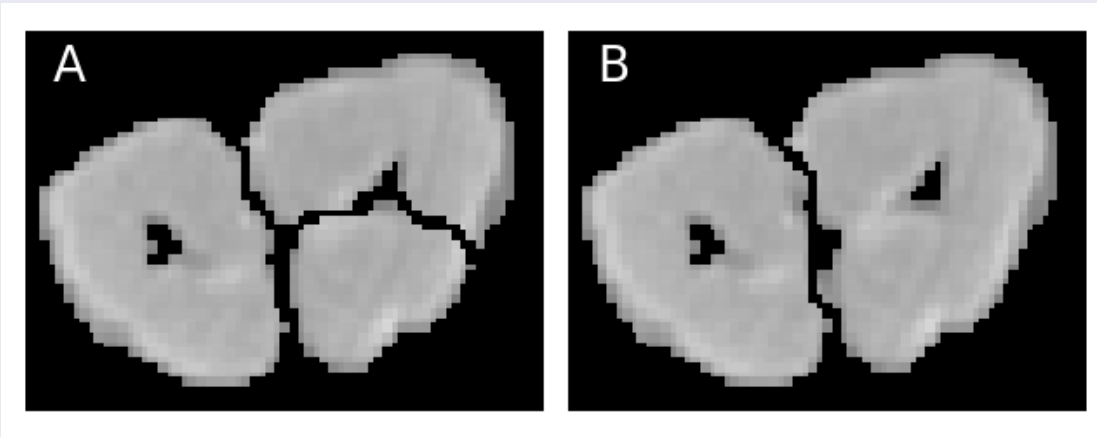

Figure 2: *A* showing the chessboard method, *B* improved quasi-euclidean method

Outline

Data Analysis *WIP*

Statistical Methods to be implemented

These will be implemented in Python using the library *SciPy* [2]

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

Graphing Methods to be implemented

All Graphing is being implemented using Matplotlib [3]

- Box
- Swarm
- Scatter / LM
- Violin
- Correlation Maps
- Frequency/Histograms

Reproducible Analysis GUI *WIP*

Figure 3: GUI Showing Dynamic Graphing

Outline

Additional Aims

Not yet started, but would like to implement by end of project

- Automate error finding and identification
- Visually Display Errors in Data
- Write testing suite for software

Ideas for future work

- Test software library on other datasets
- Grow more plants and create more replicates for more robust study

Outline

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, 2017.

[2] Eric Jones, Travis Oliphant, and Pearu Peterson.
{SciPy}: open source scientific tools for {Python}.
2014.

[3] J. D. Hunter.
Matplotlib: A 2d graphics environment.
Computing In Science & Engineering, 9(3):90–95, 2007.