Data Management Plan   **Kevin Falk**
Course: BCB546X   PI: Asheesh K. Singh
Professors: Heath/Hufford/Lavrov   Department of Agronomy
Due Date: 12/01/2018   Root System Architecture in Soybean

Research Project

Studies of Root System Architecture in Soybean using Computer Vision. This project will create image files from 300 diverse soybean root accessions at different growth stages in both indoor and field experiments. The data management plan is devoted to ensure that the image files (as well as other data formats) are properly collected, handled and stored in addition to be made available publicly for future use. Due to the nature of image files, the management plan of this project is to share the software output files (pre-and post-analysis) in .CSV format. Genotype files of the 300 soybean accessions used in this study will also be made available.

Data Types

- image (.jpeg)

  - Capture images with a Canon T5i on maximum resolution, 18 megapixels or 5184 x 3456 pixels

  - Projected light images are collected using a DFK 33UX174 USB 3.0 which captures a 2.3 megapixel image (1920 x 1200 pixels)

- MATLAB (.m)

  - 3D reconstruction of projected light images are produced through MATLAB

- Video (.mp4)

  - Protocols and tutorials are created using videos from the Canon T5i and a GoPro camera

- Genotype data (.csv)

  - 50k SNP chip data for 300 soybean accessions

- Ratings, counts, staging, comments (.csv)

  - All field data captured within the growing season will be done using Excel
  - Additional comments are added to aid in analysis

Data Preservation

All raw data including images will be stored on the Singh Soybean Group server. Pre-analysis and post-analysis data will be shared on github. Links to software used in this analysis will also be made available at github. As a backup source, two Seagate 1.0 TB HDD will be used as redundant locations for all files and will be kept in G503 Agronomy Hall

- https://github.com/mighster/SoybeanRootProject

Data Sharing

Due to the raw data being in jpeg (image) format, it makes data sharing difficult as the total number of images taken in 2017 exceeds 100,000 (546 GB). Therefore, the management plan is to house the raw images in house at Iowa State University and share the software output (pre-analysis) online. The 50k SNP chip genotype data is available publicly through www.soybase.org. ARIA software output (image analysis) will be made available in .csv format at https://github.com/mighster/SoybeanRootProject. In addition, other .txt and .csv format data will also be stored at this location. Raw images available at request of corresponding author Asheesh K. Singh (singhak@iastate.edu)

Metadata

Metadata/documentation will be submitted alongside the data in order to make the data reusable. Scripts built into the software use barcodes within each image to rename each image file. Barcode and thus name of each file carries metadata of entry number, rep, growth chamber and day after planting. In addition to the name all images will have metadata (take the form of EXIF format within the .jpeg files) updated to include:

- Pixel density (pixels/cm) (essential for proper downstream measurement)

  - An image of a ruler is taken at each day of imaging. As well, the barcode tags have a known size, this information can also be used to capture pixels/cm.

- Timestamp, rep and growth chamber of capture

- Camera model, focal length information

- Other comments, errors and notes are also added in README.txt file for each replicate

  - Comments, errors and notes will be captured manually during the experiment and added to the file containing the images as a README

Quality Assurance

- Each day an image of a ruler will be captured to ascertain the pixels/cm

- After analysis, all images and analyzed images will return to Kevin for inspection.

- Problem images will be identified and modified

- Modified images will be reanalyzed

Future Proposals

As this particular project is one step within the program, these data will continue to be made available for future work including individuals within the Singh Soybean Group, its collaborators and individuals in the field.