THE TERRA PHENOTYPING REFERENCE PLATFORM: OPEN DATA AND SOFTWARE FOR PRECISION FIELD CROP MEASUREMENT AND ANALYSIS

Allocation: Director Discretionary/5 PB storage
PI: David LeBauer[1]
Co-PIs: Todd Mockler[2], Robert Pless[3], Roman Garnett[4], Vasit Sagan[5], Rob Kooper[6], Duke Pauli[1], Geoff Morris[7]

[1]University of Arizona
[2]Donald Danforth Plant Science Center
[3]George Washington University
[4]Washington University of St. Louis
[5]St. Louis University
[6]National Center for Supercomputing Applications
[7]Kansas State University

EXECUTIVE SUMMARY

The Transportation Energy Resources from Renewable Agriculture Phenotyping Reference Platform (TERRA–REF) aims to optimize breeding strategies for improving the yield and stress tolerance of sorghum. Sorghum (*Sorghum bicolor*) is a food and energy crop with high yield potential, efficient water use, and drought tolerance and has the capacity for genetic improvement using both traditional and genomic approaches. The water-use efficiency and drought tolerance of sorghum make it adaptable to future climates that are expected to be hotter and drier.

RESEARCH CHALLENGE

Improving crops is challenging, but humankind needs crops that are more productive, stress tolerant, and more efficient at using resources such as water and fertilizer. Intensive measurements of crop plants and their environment provide plant breeders with information that is required to select for these traits. Automated measurements can help breeders select the best plants more quickly and efficiently. Although new sensors and robotic platforms are becoming available, these relatively new tools are difficult for scientists to use to gather actionable information.

Figure 1: Field scanner system (with sensors in white box) operating on 200-meter steel rails in Maricopa, Arizona, U.S.A.

TERRA–REF is collecting automated high-throughput remote sensor data on diverse varieties of sorghum and wheat. Remote sensors mounted in a field scanner instrument box scan over one acre of plants, producing thousands of daily measurements with high spatiotemporal resolution. Sensor data types range from hyperspectral imaging to 3D reconstructions and thermal profiles, all at 1-mm resolution. The program is making the data available for researchers not only to study the plants but also the sensors themselves in order to learn what new information these sensors provide.

METHODS & CODES

Data from the sensors is streamed from the field scanner to the National Center for Supercomputing Applications, where it is calibrated and processed. Blue Waters' nearline tape storage is used both for backup and for staging of data for reprocessing. The data is then organized, annotated, and processed. This process uses RabbitMQ to recognize and handle each data set as it arrives or is created. Clowder extractors created using novel predictive models and algorithms process the raw data into high-resolution images, point clouds, and time series, and then extract features related to plant growth, chemistry, and physiological efficiency. Clowder stores metadata and tracks provenance for derived data sets.

Once reduced to plant and plot-level summaries, the means, trait data, and agronomic metadata are stored in the Biofuel Ecophysiological Traits and Yields database (BETYdb). Manually collected trait data and agronomic data are also imported into BETYdb for algorithm training, calibration, and validation.

RESULTS & IMPACT

All software used in the processing pipeline is available on GitHub with permissive open-source licenses. Similarly, curated data sets generated by this platform will be published and placed in the public domain to maximize the impact of these unprecedented data.

WHY BLUE WATERS

The Blue Waters nearline system provides TERRA–REF with a quickly accessible and reliable storage system for both data backup and staging data for reprocessing. Having reliable storage for backup allows the research team to focus on the data processing and delivery infrastructure.

PUBLICATIONS & DATA SETS

M. Burnette *et al.*, "TERRA–REF data processing infrastructure," in *Proc. PEARC 2018*, doi: 10.1145/3219104.3219152.

A Reference Phenotyping Platform for Crop Breeding and Improvement. [Online]. Available: https://github.com/terraref, Accessed on: Jul. 26, 2019.

Figure 2: An algorithm to count and measure sorghum panicles (seed heads) helps agricultural researchers understand genes that increase yields. It also helps plant breeders more efficiently develop and commercialize more productive and sustainable crops that will tolerate future climates.