High-throughput Plant Phenotyping and Data Analysis

A report of a series of workshops and discussions held at:

The Plant and Animal Genome Conference XXIV
San Diego, CA USA
January 9-13, 2016

Phenotypic Prediction - Image Acquisition and Analysis
Ames, IA USA
February 23-25, 2016

Semantics for Harmonization and Integration of Phenotypic and Agronomic Data
Montpellier, France
May 9-13, 2016

Organizers
Ruth Bastow, *University of Warwick, UK*
Katherine Denby, *University of Warwick, UK*
Carolyn Lawrence-Dill, *Iowa State University, USA*
Nathan Miller, *The University of Wisconsin, USA*

Participants (cumulative)
Anne Francois Adam-Blondon, Adedotun Akintayo, Carson Andorf, Elizabeth Arnaud, Darshi Banan, Ivan Baxter, Ed Buckler, Darwin Campbell, Steven Cannon, April Agee Carroll, Jackson Cothren, Robert Davey, Abdalla Dao, John Doonan, Shawn Dorius, Kate Dreher, David Ertl, Noah Fahlgren, Richard Finkers, Maren Freisen, Baskar Ganapathysubramanian, Georgios Gkoutos, Fiona Goggin, Michael Gore, Sharon Gray, Mingze He, Pankaj Jaiswal, Toni Kazic, Paul Kersey, Pierre Larmande, Andrew Leakey, David LeBauer, Xianran Li, Kara Lind, Argelia Lorence, Helene Lucas, David Marshall, John McKay, Richard Michelmore, Seth Murray, Pascal Neveu, Henry Nguyen, Trevor Nolan, Jesse Poland, Cyril Pommier, Yumou Qiu, Vishnu Ramasubramani, Chris Rawlings, Jean-Jack Riethoven, Elena Rice, Sindhu Sankaran, Patrick Schnable, Oskar Siemianowski, A.K. Singh, Uli Schurr, Mark Sorrells, Josh Stein, David Stern, Yiwei Sun, Christopher Topp, Bjorn Usadel, Harkamal Walia, Ramona Walls, Xu Wang, Doreen Ware, Xiaoqing Yu, and Jianfeng Zhou

This report was prepared January through June of 2016 by the participants of the workshops for federal agencies in the US and UK. The workshops were sponsored by the National Science Foundation through Grant Number IOS 1613452 to Iowa State University as well as the Biotechnology and Biological Research Council through a grant to the University of Warwick. Any opinions, findings, conclusions, or recommendations expressed in this report are those of the participants and do not necessarily represent the official views, opinions, or policy of the funding agencies.

Executive Summary

Background

It is predicted that the world's population will exceed 9 billion by 2050. Ensuring a secure supply of crop plants for nutritious food, feed, fiber, and commercial uses is therefore a significant challenge for the years ahead. To meet this global need, crop production must be increased significantly, which requires a better understanding not only of the genetics that underpin agriculturally important traits but also how such traits are affected by growing conditions. Understanding how particular genotypes result in set phenotypes under specific environments is not only a goal important to implement sustainable production globally, but also a central question for modern biology as a discipline. Put another way, how are the observable physical or biochemical characteristics of an organism determined by the interaction of both genetic makeup and the environment? Linking complex traits with genetics and the environment requires skills in plant sciences, engineering, and data analytics and depends heavily on access to extensive, complex, computable datasets.

Activities

An interdisciplinary group of US and EU researchers met over the course of three workshops to discuss and identify key challenges in plant biology and applied breeding with emphasis on managing the datasets derived from the use of so called "high-throughput phenotyping" tools. The goals of these workshops were to identify obstacles to the development of a more data-enabled research environment for plant sciences and to develop a roadmap to support data management and data sharing. Invited speakers delivered presentations at these various venues on the broad topics of phenotyping, data collection, and information management to catalyze broad discussion. Panel discussions and brainstorming sessions engaged participants in open dialogue to articulate anticipated research trajectories and related challenges in the realms of data collection, management, and access for plant sciences research. Themes uncovered are described here. It is our hope that these findings and suggested strategies to enable concerted research efforts will serve to coordinate activities in support of data-driven research among funding entities, academic institutions, federal laboratories, independent research institutes, and beyond, thus contributing to the needed scale-change in agricultural gains required to support humanity over the coming years.

Findings

Across all venues, the group was unified in its perspective that **improved communication and collaboration opportunities that span national, international, and disciplinary boundaries are needed to share the current state-of-the-art and knowledge.** There are some national and international efforts to promote discussions and international collaborations, but these mainly occur through self-organization among scientists rather than via specifically funded strategic efforts. In the absence of increased, well supported community building activities, efforts to solve the same basic technical challenges will continue to develop in parallel and will negatively impact the rate of research. Such research coordination-building efforts are needed among scientists across diverse disciplines (e.g., plant sciences, engineering, and data sciences) as well as among the leaders of the agencies that fund plant sciences research to identify mechanisms to encourage collaborative activities across national and disciplinary boundaries. To achieve integrated research that involves the development and use of novel sensing instrumentation and data management and analysis principles for plant biological research, not only is collaboration with colleagues outside researchers' primary fields of expertise needed, **education and trans-disciplinary training opportunities at all career stages are critically needed**. We encourage a multi-pronged approach to training including, e.g., the development of explicit graduate opportunities, summer courses for students, postdocs, faculty, and staff, and immersion experiences for re-tooling (e.g., sabbatical work). As useful tools and data collection methods are developed and researchers acquire skills needed to collect large-scale, multifaceted datasets, **the development of standards, methods, and best practices to share the information gathered in machine-readable and computable formats will become key to maximizing impacts of research** through data discovery and re-use. Finally, the group recognized that the best of efforts among individual scientists and teams of researchers will ultimately fall short of their goals regarding 21st century science practice without commensurate institutional innovations. Reward structures in many academic institutions continue to give preeminent recognition to solo-authored work, published results over published data products, and scholarship in the home-discipline content domain at the expense of scholarship in the kinds of emergent, transdisciplinary domains favored by data scientists. To adapt to the changing research climate**, institutions of higher education and their peer organizations operating outside of the academy will need to make fundamental changes in reward structures** that incentivize: a) the sharing of data, b) more extensive involvement of individual faculty in transdisciplinary and trans-institutional research collaborations, c) equality of recognition for research products outside of a researcher's area of domain expertise, and d) enlargement of the definition of scientific scholarship to include the kinds of data rich activities at the center of data science. Failure to modernize reward structures in these and related areas will likely slow the pace of scientific discovery and mitigate current efforts aimed at building a more integrated, collaborative scientific scholarly community.