PROJECT DESCRIPTION

1 Project Goals

In the post-genomics era, one of the major research goals for plant genomics community is to understand the functions of genes and their interaction networks. To achieve this goal, it is an urgent need for the community to access informatics tools and phenotyping protocols that provide quality of controlled data for keen study of phenotypes with both depth and breadth in diverse environment for numerous traits [Allerton Report]. However, there is minimal to no effort for the community to develop multiple-species standards, similar to the Plant Ontology project, for phenotype description and analysis, let alone more complex studies to associate traits with genes or results sharing without physical boundary barriers. A broad community involvement is expected to have high-throughput management and analysis of phenotypes, such as agronomic, morphological, cellular and molecular traits. A cohesive effort will require a high level of interaction within the plant community, dedicated cyberinfrastructure, and innovative informatics tools. A centralized information hub for repository, analysis, synthesis, and dissemination of phenomics media is proposed that can be maintained at any location.

To do this, based on our solid foundation in plant phenomics research sponsored by the National Science Foundation (DBI-0447794), we are proposing a Virtual Center for Plant Phenotypes (VCP^2) by closely working with individual plant genome databases. There are six research and development goals that will have the rich potential to broadly impact plant genome research community.

Aim 1: Develop Phenotype Collection Protocols to Allow Multi-Site Studies: We will develop protocols to collect phenotypic information with details about species, germplasm, anatomy, developmental stages, and environmental factors. We will introduce the concept of minimal data sets based on Plant Ontology (PO) for researchers to deposit commonly shared data for cross-institutional analysis and the concept of digital imaging and communications in plants (DICOP).

Aim 2: Develop a Data Warehouse for Community-based Deposit and Retrieval of Phenotypes: We will build a phenotype warehouse that provides a reliable space for community phenotype repository using fine-tuned relational databases, multimedia file systems, and customized database indexing mechanism to meet the unique need of large-scale phenotype data.

Aim 3: Develop Analysis Tools for High-Throughput Phenotype Data Process: We will develop automatic phenotype analysis tools to allow users to upload images, select algorithms from tool boxes, and process images in a batch mode with only a few clicks. Thousands of images are expected to be analyzed in minutes.

Aim 4: Develop Tools to Link Phenotype Data with Genotypes: We will develop tools to identify potential linkages between the phenotype data warehouse with genomic and gene function data in major plant genome databases, the protein data bank, pathways, and GO annotations.

Aim 5: Develop Advanced Retrieval Tools to Link Genotypes, Phenotypes, and Text Descriptions: We will develop unique phenotype retrieval tools to provide query by phenotype image example, query by semantics, query by text annotations, query by ontological information, query by genealogy information, query by map locations, and query by reverse genetic strategies.

Aim 6: Transfer Technology to Plant Databases by Outreach with Onsite Training: The ultimate goal of this project is to disseminate the entire framework to plant genome community by tech transfer to the major plant genome databases. We will organize biannual user's meetings to discuss the needs from the community. We will provide training for interested parties, at their site, both in use of the tools, and deployment in their computational environment.