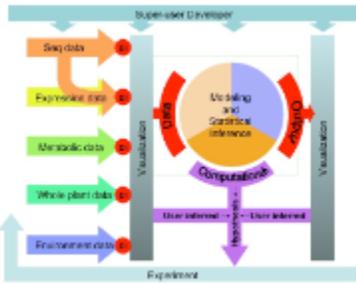


Home

Navigate space

Expand all Collapse all



iPG2P: Relating Genotypes to Phenotypes in Complex Environments

Elucidating the relationship between plant genotypes and the resultant phenotypes in complex (e.g., non-constant) environments is one of the foremost challenges in plant biology (NRC, 2008). Plant phenotypes are determined by often intricate interactions between genetic controls and environmental contingencies. In a world where the environment is undergoing rapid, anthropogenic change, predicting altered plant responses is central to studies of plant adaptation, ecological genomics, crop improvement activities (ranging from international agriculture to biofuels), physiology (photosynthesis, stress, etc.), plant development, and many many more.

A concerted attack on the G2P problem will require the combined and integrated efforts of specialists in functional-, quantitative-, and computational genetics/genomics, bioinformatics, modelers, physiologists, computer scientists (for computing to high performance computing to visualization), etc. Cyberinfrastructural innovations are required to facilitate collaborations this diverse. Planning efforts leading up to the iPG2P project have identified five, high priority areas where progress is needed:

1. Pipelining of NextGen sequence data into virtual genotype and molecular phenotype databases. Virtual databases are comprised of multiple, individual databases located at multiple sites, that are effectively integrated by middle ware that provides common access.
2. Data integration - the infrastructure necessary to combine/overlay data from such virtual databases to permit deeper insights, generation of hypotheses, evaluation of models, practical applications, etc.
3. Statistically-based tools for use in inferring relationships ranging from marker associations (primarily to, where practical, links in network structures. Many such tools exist, but the value-added aspect in the current context is to make them smoothly interoperable with the other features of the cyberinfrastructure.
4. Visual analysis tools. It is necessary to present integrated data to users in ways that are both concise and revealing. Such presentations must include capabilities for both static and (increasingly) dynamic/kinetic displays of plant biology information (e.g. .omics., ecophysiological data), be it multidimensional [2d (e.g. geographical, comparative genomics), 3d (e.g. PCA), 4d, or higher], and/or in the form of networks or pathways.
5. Modeling framework tools to support the construction, parameter estimation, sensitivity analysis, and utilization of models. Again, the value added is in interoperability. In the short term, operation within an integrated data environment will facilitate all forms of modeling (including statistical). Over the near-to-intermediate term, components of ecophysiological models will increasingly employ the results of gene-based network studies, thus enhancing their application in breeding and other contexts.

Working groups have been formulated in each of these areas ([NextGen Sequence Pipeline](#), [Data Integration](#), [Statistical Inference](#), [Visual Analytics](#), and [Modeling Tools](#)) with co-leaders to serve as points of contact.