

The Gramene Genetic Diversity Module: a resource for genotype-phenotype association analysis in grass species

Abstract

With the deluge of data coming from high-throughput sequencing and array-based genotyping platforms, there is increasing interest in associating molecular genetic variation with physiological and phenotypic variation. The high volume of genotypic and phenotypic data being generated for diverse species poses significant challenges in terms of data post-processing, data storage and visualization, and the computationally intensive area of data analysis.

The Genetic Diversity module of the Gramene database (www.gramene.org/diversity) is specifically designed to handle these data and to facilitate data integration and analysis. It uses the Genomic Diversity and Phenotype Data Model (GDPDM; maizegenetics.net/gdpdm; GNU LGPL) to store RFLP, SSR and SNP allele data, information about QTL, passport data for wild and cultivated germplasm from rice, maize, wheat, and *Arabidopsis*, and quantitative phenotypic data for some of these accessions. With live database connectivity, large datasets of SNP variation can be queried by defining a genomic position of interest on a sequenced genome, and trait associations, patterns of linkage disequilibrium and diversity can be evaluated using a Java Web Start enabled version of TASSEL (gramene.org/diversity/tools.html; GNU GPL).

Of increasing importance for the Diversity module is the effort to leverage data generated in rice, maize, and *Arabidopsis* to allow users to integrate information from these model genomes to better understand the functional significance and evolutionary history of genes controlling plant morphology, development and response to the environment. Gramene scientists are developing tools and statistical techniques that allow users to build hypotheses about phenotype-genotype relationships in the plant kingdom based on the advantages and compensating for the disadvantages of each experimental system. The Gramene database is re-built and updated twice a year, with the most recent release (Build #31) completed in May 2010.