## The Gramene Genetic Diversity module

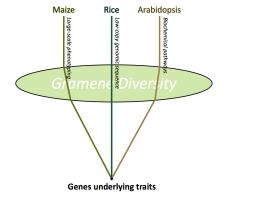
A resource for comparative genome analysis in plants

DeClerck, Genevieve<sup>1</sup>, Casstevens, Terry<sup>4</sup>, Yap, Immanuel<sup>1</sup>, Tung, Chih-Wei<sup>1</sup>, Avraham, Shulamit<sup>2</sup>, Ren, Liya<sup>2</sup>, Spooner, Will<sup>2</sup>, Thomason, Jim<sup>2</sup>, Wei, Sharon<sup>2</sup>, Youens-Clark, Ken<sup>2</sup>, Bradbury, Peter<sup>3</sup>, Ware, Doreen<sup>2,3</sup>, Jaiswal, Pankaj<sup>5</sup>, McCouch, Susan<sup>1</sup>, Buckler, Edward<sup>3,4</sup>

- <sup>1</sup> Dept. of Plant Breeding and Genetics, 240 Emerson Hall, Cornell University, Ithaca, NY 14853
- Cold Spring Harbor Laboratory, 1 Bungtown Rd, Cold Spring Harbor, NY 11724
  USDA-ARS, Robert W. Holley Center, Cornell University, Ithaca, NY 14853
- <sup>4</sup> Institute of Genomic Diversity, Cornell University, Ithaca, NY 14853
- <sup>5</sup> Dept. of Botany and Plant Pathology, 3082 Cordley Hall, Oregon State University, Corvallis, OR, 97331

Characterization of genetic diversity within and between species is a central activity in biology today. With the introduction of an increasing number of high-throughput sequencing and array-based genotyping technologies, identifying variation at the molecular level is becoming routine, while the ability to associate this type of data to variation at the phenotypic level continues to present significant challenges. The bottleneck encountered by studies aiming to elucidate genotypic-phenotypic associations is most acute in the areas of data post-processing, data storage, and in the computationallyintensive analyses of results. The Genetic Diversity module of the Gramene database (gramene.org/db/diversity/diversity\_view) 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) to store RFLP, SSR and SNP allele data, information about QTL, passport data for rice, maize, wheat, and Arabidopsis germplasm, and quantitative phenotypic data for some of these accessions. The module has direct connectivity with analysis packages such as TASSEL (maizegenetics.net/tassel), as well as other tools and potential external data sources through the Genomic Diversity and Phenotype Connection (GDPC; maizegenetics.net/gdpc). Of increasing importance for the Diversity module is the effort to develop associations between rice, maize, and Arabidopsis, which will allow users to integrate information from each of these model genomes, building on the advantages and compensating for the disadvantages of each system and dataset. In this way, Gramene's Genetic Diversity module aims to increase the efficacy of comparative genome analysis in the plant kingdom and to provide plant researchers with a platform for managing and interpreting

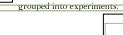
The database is updated 2 times a year, with the release (Build #29) on Search to 7,41(ble profile of a specific germplasm accession using its accession name or accession number (IRIS or GRIN) Go to: http://www.gramene.org/db/diversity/diversity\_view Allele profile of Basmati 1 (IRGC 2 8) at 169 SSR Show all germplasm records with Download all records in allele profiles for SSR marker, RM 1 tab-delimited text

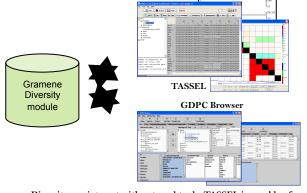


The Gramene Diversity module will increasingly become a lens through which information from rice, maize and Arabidopsis can be focused to elucidate genes underlying traits, with each genomic dataset bringing complementary strengths. Genes and genotype information can be linked to germplasm resources, traits or phenotypes, and environment through the GDPDM database schema (see below)

## GDPDM Database Schema

- · Germplasm: Supports basic passport information as well as complex genetic pedigrees.
- · Genotype: Captures wide range of approaches (SNP, SSR, sequence alignments, isozyme, etc). Supports full-range of ploidy and can capture background on methodology. Connects to genomic map data.
- Phenotype: Quantitative or qualitative traits. Supports ontology integration and a range of statistics can be recorded (mean, measure, median)
- Environment: Full field design can be linked to individual plant sample. Planting, treatment and locality data can be integrated. Observations can be





Gramene Diversity can interact with external tools. TASSEL is capable of quantitative trait analysis, diversity and linkage disequilibrium analyses, association analyses, SNP assay analyses, and a wide range of visualizations. The GDPC Browser can retrieve genomic and phenotypic data based on selected criteria. Data can be saved, exported, grouped.











