

The Gramene Genetic Diversity Module: A Resource For Genotype-Phenotype Association Analysis In Grass Species

DeClerck, Genevieve¹, Casstevens, Terry⁴, Chen, Charles¹, Youens-Clark, Ken², Spooner, Will², Thomason, Jim², Yap, Immanuel¹, Tung, Chih-Wei¹, Avraham, Shulamit², Ren, Liya², Zhang, Jon⁴, Wei, Sharon², Bradbury, Peter³, Ware, Doreen^{2,3}, Jaiswal, Pankaj⁵, McCouch, Susan¹, Buckler, Edward^{3,4}

¹ Dept. of Plant Breeding and Genetics, 240 Emerson Hall, Cornell University, Ithaca, NY 14853

² Cold Spring Harbor Laboratory, 1 Bungtown Rd, Cold Spring Harbor, NY 11724

³ USDA-ARS, Robert W. Holley Center, Cornell University, Ithaca, NY 14853

⁴ Institute of Genomic Diversity, Cornell University, Ithaca, NY 14853

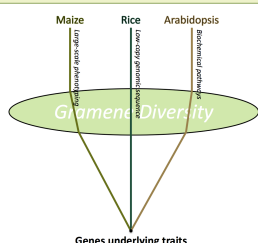
⁵ Dept. of Botany and Plant Pathology, 3082 Cordley Hall, Oregon State University, Corvallis, OR, 97331

⁶ European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK

⁷ Dept. of Biological Statistics and Computational Biology, Cornell University, 1198 Comstock Hall, Ithaca, NY, 14853, USA

Check out the other
Gramene PAG posters:
P817, P822, P835, P859

The Gramene Genetic Diversity module (gramene.org/diversity) focuses on the integration of diversity data from several plant species, concentrating on rice, maize and Arabidopsis. The goal is to enable users to elucidate genetic mechanisms underlying complex trait variation within and between these data sets by associating genotype with phenotype measurements. Recent additions to the diversity module include 160K rice SNP allele data points for 20 rice lines (McNally et al., PNAS 2009) and allele data from the 'Arabidopsis 2010' project 250K SNP-chip for 931 Arabidopsis ecotypes (Nordborg Lab website). A new tool called SNP Query (gramene.org/db/diversity/snp_query) has been developed for performing coordinate-based searches of these new large volume SNP data sets. Additional high-throughput SNP work from the Maize HapMap project as well as phenotype data for all three central Diversity species will be added as the experimental results are made public. For performing analyses, a Java Web Start enabled version of TASSEL can be launched with live Diversity database connectivity to streamline data loading. Filtered data sets can likewise be extracted for external use with the GDPC Browser. The Diversity module handles storage of genotype, phenotype and corresponding germplasm data through use of version 4.0 of the Genomic Diversity and Phenotype Data Model (GDPDM) database schema.



The Gramene Genetic Diversity Module: A lens for discovery of genes underlying traits in plants

Gramene Genetic Diversity stores genetic diversity information from rice, maize, and Arabidopsis, with each genomic dataset bringing complementary strengths. Pooling allelic data and phenotypic observations will facilitate association modeling within and between these three important plant systems and will help elucidate genes underlying traits in plants.



What's New in Gramene Genetic Diversity:

Two important new large-scale SNP variation studies

1. OryzaSNP.org rice SNP variation data released in McNally et al. (PNAS July 2009)
~160,000 rice SNP markers assayed against 20 diverse rice varieties.

2. Data from Arabidopsis '2010' Project's 250K SNP chip,
~216,000 Arabidopsis SNPs x 931 Arabidopsis ecotypes.

Summary of Data in the Gramene Diversity module

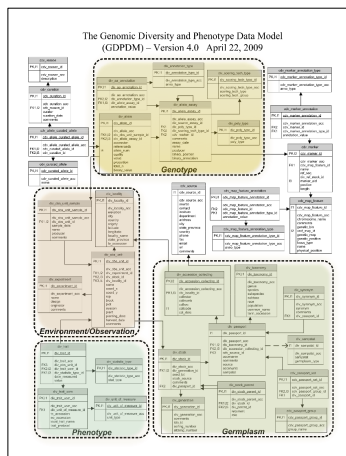
	Number of diversity experiments/studies	Number of germplasm	Number of traits for which there are phenotype measurements	Number of SSR, RFLP, small-scale SNP markers	Number of SNP markers from large-scale genotype array projects
Rice	16	847	28	1629	159,879
Maize	31	8037	150	13075	~coming soon~
Arabidopsis	3	931	-	-	216,508
Wheat	1	48	-	4097	-
Sorghum	6	1993	-	544	-

	Diversity data sources
Rice	QTL and SSR based diversity studies from 15 publications; Large scale SNP data from oryzaSNP.org
Maize	Molecular and Functional Diversity of the Maize Genome project (NSF DBI 0321467), Panzea.org
Arabidopsis	Large-scale SNP data from NSF '2010' Arabidopsis Project
Wheat	Haplotype Polymorphism in Polyploid Wheats and Their Diploid Ancestors project (NSF DBI 0321757)
Sorghum	Sorghum Diversity Project

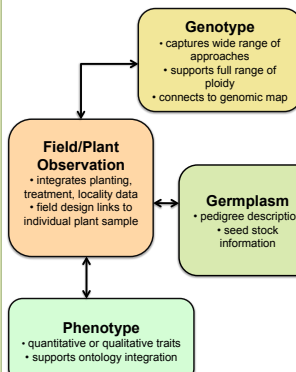
GDPDM Database Schema

Genomic Diversity and Phenotype Data Model

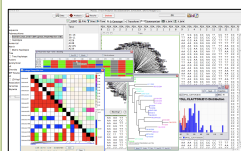
The Gramene Genetic Diversity database module implements the GDPDM schema. The focus of GDPDM is to hold quantitative phenotypic data and molecular genotypic data. This data may be the product of QTL mapping, association studies, large-scale SNP genotyping assay studies, field data from plant breeding programs, germplasm evaluation studies.



GDPDM: www.maizegenetics.net/gdpdm



Diversity Tools

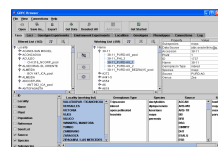


TASSEL (Trait Analysis by aSSociation, Evolution and Linkage) is a Java software package for evaluating trait associations, evolutionary patterns, and calculating and graphically visualizing linkage disequilibrium.

www.gramene.org/diversity/tools.html



The Gramene Diversity SNP Query tool will return subsets from the SNP-chip datasets based on position coordinates passed in by the user.



The Genomic Diversity and Phenotype Connection (GDPC) Browser is an advanced search and data filtering utility. The standalone Java-based GDPC Browser provides users with a platform for performing complex queries and data filtering steps directly with the Diversity database and/or with your own local datasets.

HOW TO: Search for allele profile of a specific germplasm accession using its accession name or accession number (IRIS or GRIN)

1. Go to: www.gramene.org/diversity and type a germplasm name in the quick search box...

2. Review germplasm and experiment information then click 'Show Allele Data' link

3. View and download allele profile of Basmati 1 (IRGC 27798) at 169 SSR loci

4. optionally, click on 'All Genotypes On Marker' link to also see allele profiles for SSR marker 'RM 1' for all 234 germplasm in this experiment

Download profile data in text format

