

The Gramene Genetic Diversity module: A resource for comparative genome analysis in plants

Genevieve DeClerck¹, Terry Casstevens⁴, Immanuel Yap¹, Chih-Wei Tung¹, Shulamit Avraham², Liya Ren², Will Spooner², Jim Thomason², Sharon Wei², Ken Youens-Clark², Peter Bradbury³, Doreen Ware^{2,3}, Pankaj Jaiswal⁵, Susan McCouch¹, Edward Buckler^{3,4}

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

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

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

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

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

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 computationally-intensive 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 diversity data.