

Maize Diversity: Accessing Data behind Germplasm, QTL, and Breeding Studies

Terry Casstevens¹, Peter Bradbury^{2,3}, Payan Canaran⁴, Ken Youens-Clark⁴, John Doebley⁵, Jeff Glaubitz⁵, Dallas Kroon¹, Mike McMullen^{2,6}, Hector Sanchez-Villeda⁶, Steve Schroeder⁶, Qi Sun³, Doreen Ware^{2,4}, Zhiwu Zhang¹, Wei Zhao⁴, Edward S. Buckler^{1,2}

¹ Institute of Genomic Diversity, Cornell University, Ithaca, NY, USA

² USDA-ARS

³ Cornell Theory Center, Cornell University, Ithaca, NY, USA

⁴ Cold Spring Harbor Laboratory, NY, USA

⁵ Department of Genetics, University of Wisconsin, Madison, WI, USA

⁶ Division of Plant Sciences, University of Missouri, Columbia, MO, USA

Maize diversity is the basis of feeding a large proportion of the world. Unlocking the complex associations between the maize genome and observable traits require vast amounts of human and computational resources for data collection and analysis. A major challenge of this task involves storing large genotypic and phenotypic data sets and providing access to the data. We developed Panzea, a public website dedicated to making maize diversity data available to the community. The infrastructure includes a database schema (GDPDM; www.maizegenetics.net/gdpdm, sourceforge.net/projects/gdpdm), a Java middleware and advanced search (GDPC; www.maizegenetics.net/gdpc, sourceforge.net/projects/gdpc), and web-based tools (including: a sequence alignment-SNP viewer, Germplasm Search, Molecular Diversity Search, etc.) (www.panzea.org and www.gramene.org/db/diversity/diversity_view). Anyone can implement the diversity database schema (GDPDM) to organize their collections of genetic and phenotypic data. The middleware provides a highly-flexible advanced search tool to retrieve, integrate, and display customized data sets. A genetic association and diversity analysis tool (TASSEL; www.maizegenetics.net/tassel, sourceforge.net/projects/tassel) also uses the middleware to acquire data for analysis. We also provide web based tools for making data available on our website by utilizing existing tools when applicable and developing generic stand-alone tools. These include various interactive search pages and visualization tools. All these components are free open-source tools and can be used by other crop species (i.e. wheat, rice). We are also working (1) to incorporate more data from diversity projects, (2) to develop community upload tools, and (3) to create enhanced query, display and analysis tools. We encourage community input and collaboration on this effort, so that the largest possible community can access and productively use diversity data.