

TASSEL Pipeline: Analyzing Diversity Data

Terry Casstevens¹, Peter Bradbury^{2,3}, Zhiwu Zhang¹, Edward S. Buckler^{1,2,4}

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

² USDA-ARS

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

⁴ Dept. of Plant Breeding and Genetics, Cornell University, Ithaca, NY, USA

TASSEL version 2.1 has been redesigned to allow command line batch analysis of diversity data. The various analysis functions are now built as individual modules (plugins). These plugins are used both in the graphical user interface and the new TASSEL pipeline. The plugins are used in the pipeline to create custom batch analysis jobs. Basically, data goes through the pipeline being acted upon by each plugin along the way. Also, more complicated logic can be designed into the pipeline if needed. Analysis plugins include functions, such as calculations of linkage disequilibrium, plots of linkage disequilibrium, association analysis using mixed linear model (MLM) and general linear model (GLM) algorithms, loading files, genotype transforms, and exporting results to files. Conditional plugins include functions, such as combining data sets, filtering, pass through, and setting analysis parameters. Once a pipeline has been designed and implemented, many data manipulations can be accomplished without user interaction. One pipeline that we developed can perform GLM and MLM analysis using flat file input (i.e. SNP data, Trait data, Population Structure, and Kinship Matrix) and producing tab delimited output of the results. TASSEL is an open source project (www.maizegenetics.net/tassel, sourceforge.net/projects/tassel).