

THE HARDWOOD GENOMICS PROJECT: AN ONLINE DATABASE FOR TREE GENETIC AND GENOMIC DATA

Matthew Huff¹, Abdullah Almasaeed¹, Bradford Condon¹, Ming Chen¹, Amanda Devine¹, Casey Richards¹, Raymond Senu¹, Patrick Sisler¹, Jill Wegrzyn², Doreen Main³, Stephen P. Ficklin³, Albert G. Abbott^{4,5}, John E. Carlson⁵, Margaret E. Staton¹

¹Department of Entomology and Plant Pathology, University of Tennessee, Knoxville, TN; ²Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT; ³Department of Horticulture, Washington State University, Pullman, WA; ⁴Forest Health Research and Education Center, University of Kentucky, Lexington, KY; ⁵Schatz Center for Tree Molecular Genetics, Pennsylvania State University, State College, PA

With the vast increase in genetic and genomic data for woody tree species, it is necessary for this data to be easily accessible to scientists. However, much of this data is not available publicly, or only raw or unannotated versions of the data are available. The Hardwood Genomics Website (HWG) is dedicated to housing and annotating tree genomic datasets, including ecologically relevant and threatened forest tree species. Our site houses genomes and transcriptomes of trees unavailable on any other platform and also provides searchable functional annotation for genes and transcripts, including plant protein databases, protein domains, KEGG pathways, and gene ontologies identified through BLAST and InterProScan; we further characterize gene sequences by housing gene expression data from high throughput RNASeq experiments. HWG houses simple sequence repeats (SSRs) and their primers from genome and gene sequences, either obtained from published data or developed by our team, where they are available for use as genetic markers. We have integrated a number of tools for researchers to easily access our data, including a powerful search engine, BLAST sequence similarity searching, and JBrowse for genome browsing. As a recent example, we have incorporated Galaxy software to run bioinformatics analysis workflows, allowing users to map DNA or RNA reads to a reference, perform differential expression analysis, call variants in reads, and others. Furthermore, this system allows users to use data available on HWG, as well as data uploaded from their computer, as input in a workflow. As the scope of forest tree breeding programs and genetic programs grows, we are also building support for high throughput genotyping and phenotyping data. As Hardwood Genomics grows, we welcome new data submissions, suggestions, and partnerships to continue development. HWG is supported by NSF Awards #1443040 and #1444573.