TREEGENES: INTEGRATING GENOTYPIC, ENVIRONMENTAL, AND PHENOTYPIC DATA FOR FOREST HEALTH

Emily Grau[1], Nic Herndon[1], Sean Buehler[1], Taylor Falk[1], Peter Richter[1], Sumaira Zaman[1],
Stephen Ficklin[2], Margaret Staton[3], Doreen Main[2], and Jill Wegrzyn[1]

TreeGenes (https://treegenesdb.org/) is a web-based information resource designed to serve the diverse needs of the forest tree genomics research community by uniting information resources with visualization and analysis tools. TreeGenes has recently undergone a complete redesign using Tripal, a tool to create and manage genomic database websites. An open source project, Tripal allows developers the flexibility to create and share their own tools (modules) as well as open communication among Tripal supported repositories.

TreeGenes hosts a range of modules that have expanded functionality to deliver genomic and phenotypic data on >1,700 forest tree species. The new Tripal Galaxy module allows users to execute analytical workflows with next generation sequencing datasets on high performance computing (HPC) resources with the click of a button. The Tripal Elasticsearch module allows flexible searching and data retrieval between sites such that TreeGenes can share data with Hardwood Genomics Project and the Genome Database for Roseaceae.

TreeGenes is also developing new Tripal modules including the Tripal Plant PopGen Submit (TPPS) module, the TreeGenes TSeq module, TreeGenes OrthoQuery, and Cartogratree. TPPS is a pipeline for accepting direct submissions from researchers. This module collects relevant metadata/data for studies focused on the interactions between genotype, phenotype, and the environment while reducing the burden on the researcher for submission. The new Tripal Sequence Similarity (TSeq) Module module offers speed improvements over BLAST via DIAMOND and allows sequence similarity searches across numerous pre-indexed genomes and transcriptomes. TreeGenes OrthoQuery permits gene family and phylogenetic analysis for single sequences and/or entire proteomes with a robust and interactive visualization.

CartograTree (https://cartogratree.org) is a map-based open-source analytic module integrates across data that is curated through TPPS to facilitate association mapping and landscape genetics analysis. It uses genotypic and phenotypic data provided by two clade organism databases: TreeGenes via TPPS and Hardwood Genomics. Environmental data currently includes climate variables, land cover, canopy density, atmospheric metrics, and soil types. CartograTree leverages Galaxy workflows to support analysis in the web framework with High Performance Computing (HPC) resources.

---

[1] University of Connecticut, Storrs, CT 06269.
[2] Washington State University, Pullman, WA 99164.
[3] University of Tennessee, Knoxville, TN 37996, (jill.wegrzyn@uconn.edu).

Citation for proceedings: Nelson, C. Dana; Koch, Jennifer L.; Sniezko, Richard A., eds. 2020. Proceedings of the Sixth International Workshop on the Genetics of Host-Parasite Interactions in Forestry—Tree Resistance to Insects and Diseases: Putting Promise into Practice. e-Gen. Tech. Rep. SRS–252. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station. 170 p.