

COMPUTATIONAL APPROACHES TO DECODE MEGAGENOMES AND DEVELOP DATABASE RESOURCES FOR THE FOREST TREE COMMUNITY

Jill L. Wegrzyn, Uzay Sezen, Nic Herndon, Emily Grau, Sean Beuhler, Sumaira Zaman, Alex Hart, Alex Trouern-Trend, and Madison Caballero

University of Connecticut, Storrs, CT, USA

Forest trees are represented by over 60,000 species and exist in a wide variety of climates. These perennial, long-lived and sessile species create an ideal system to study adaptation to biotic and abiotic stress. Among the numerous forest tree species, only a handful of reference genome sequences exist. Similar to many plants, the genomes of forest trees are both large and complex. The gymnosperms, primarily represented by conifers, are some of the largest genomes sequenced to date.

Research in our plant computational genomics lab seeks to develop tools and techniques to assemble and annotate these exceptional genomes. Recent efforts have focused on annotation of the latest release of the loblolly pine genome (v2.01). In addition to improving contiguity of the reference and accuracy of gene models, we develop software to aid in transcriptome assembly and annotation.

Generating these reference genomes is paired with developing resources to integrate, analyze, and visualize related genomic and genetic information. Recent development in the TreeGenes database (<http://treegenesdb.org>) focused on cross-database data access and advanced analytical pipelines through the Tripal framework (<http://tripal.info>). Development of web-based tools that integrate a variety of genomic, phenotypic, and environmental resources to facilitate association mapping analysis is the primary objective of the CartograTree application.

Contact Information: Jill Wegrzyn, University of Connecticut, 75 N. Eagleville Road, Storrs, CT 06269, Phone: 860-486-8742, Email: jill.wegrzyn@uconn.edu