Genomic Resources for Sustaining and Improving American White Oak

Margaret Staton[1]*, Drew Larson[2], Beant Kapoor[1], Shenghua Fan[3], Jozsef Stork[4], Tetyana Zhebentyayeva[3,5], John Carlson[5], Albert Abbott[3], C. Dana Nelson[6], and Seth DeBolt[4]

[1]Department of Entomology and Plant Pathology, University of Tennessee, Knoxville, TN, USA; [2]Department of Biology, Indiana University Bloomington, Bloomington, IN, USA; [3]Forest Health Research and Education Center, University of Kentucky, Lexington, KY, USA; [4]Department of Horticulture, University of Kentucky, Lexington, KY, USA; [5]Department of Ecosystem Science and Management, Pennsylvania State University, State College, PA, USA; [6]USDA Forest Service, Southern Research Station, Forest Health Research and Education Center, Lexington, KY, USA;
*mstaton1@utk.edu

Quercus alba is a keystone species in eastern North American forests and provides valuable ecosystem services, including supporting immense biodiversity and sequestering carbon. Mature *Q. alba* trees are harvested for their lumber as well as specialty products including staves for bourbon barrels. Emerging concerns for white oak sustainability include reduced natural regeneration, increased harvest, new pests and pathogens, and climate change. To support the increasing genetic research and tree improvement programs in this species, we have produced a high-quality reference genome for *Q. alba*. It is haplotype resolved, with 796Mb from the primary haplotype and 793Mb from the secondary haplotype, and 97% of the genome anchored to 12 chromosomes for each. The genome was found to be highly heterozygous for both nucleotide and structural variation, reflecting the obligate outcrossing nature of the species. Resequencing *Q. alba* individuals from six different geographic locations and representatives from seven other common white oak species with sympatry to *Q. alba* has yielded a dataset of over 50 million SNPs to leverage for future genotype to phenotype studies. Comparison to the publicly available genomes from other oak species reveals a largely collinear structure at the chromosome scale but also significant expansion and contraction of specific gene families. We identified over 2,500 candidate R-genes, including 302 RLKs (receptor-like kinases), 308 RLPs (receptor-like proteins), 102 CC-NB-LRRs, and 19 TIR-NB-LRRs. As oaks have no recent whole genome duplication, we found that R-genes are primarily created through dispersed and proximal gene copying. This genome will provide the foundation for a white oak pangenome, an ambitious project that has the potential to accelerate oak research and provide new insights into the unique evolution of the *Quercus* species complex in North America.