Characterization of Structural Variation in a *Q. rubra* Three-Generation Pedigree

Beant Kapoor[1], Jerry Jenkins[2], Jeremy Shmutz[2], Mark Coggeshall[3], Jeanne Romero-Severson[4], John Carlson[5], and Margaret Staton[1]

[1]Department of Entomology and Plant Pathology, University of Tennessee, Knoxville, TN 37996, USA; [2]HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806, USA; [3]College of Agriculture, Food and Natural Resources, University of Missouri, Columbia, MO 65211, USA; [4]Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556, USA; [5]Department of Ecosystem Science and Management, Pennsylvania State University, University Park, PA 16802, USA

Northern red oak (*Quercus rubra* L.), an economically and ecologically important tree species of North America, is a member of the *Quercus* genus in the Fagaceae family. A chromosome-scale, haplotype-resolved genome of *Q. rubra* was generated by combining PacBio sequences, Illumina short reads and chromatin conformation capture (Hi-C) scaffolding. This resource represents the first reference genome from the red oak clade (section *Lobatae*). The *Q. rubra* assembly is 739 Mb, with 33,333 protein-coding genes and 12 chromosomes containing 95.27 percent of the genome's sequences. High overall collinearity along with intrachromosomal structural variation was observed when *Q. rubra* was compared to *Q. lobata* and *Q. mongolica* genomes. Based on orthologous gene family analysis with other oak and rosid tree species, the *Q. rubra* genome exhibited simultaneous expansion and contraction of gene families linked to the defense response. Out of the nine plant species examined, *Quercus rubra* had the highest number of CC-NBS-LRR and TIR-NBS-LRR resistance genes. The *Q. rubra* individual selected for the reference genome was a progeny of a two-generation pedigree. The genomes of the parents of the original cross and the two F_1 progeny parents of the reference tree were sequenced, scaffolded against the reference genome, and annotated. We characterized structural variants (SV) in this *Q. rubra* pedigree and produced a set of 117,854 high-quality SV events. We also found that most of these SVs (88.27%) called were less than 1 Kb in length which constitutes a novel genetic diversity resource in *Q. rubra* tree species. The high-quality red oak genomes generated are a crucial resource for the oak genomics community and will add to our understanding of *Quercus* genomics.