

Haplotype-resolved, chromosome-scale genome assembly of *Quercus rubra* L.

Beant Kapoor¹, Jerry Jenkins², Jeremy Schmutz³, Tatyana Zhebentyayeva⁴, Carsten Kuelheim⁵, Mark Coggeshall⁶, Chris Heim⁷, Jesse Lasky⁸, Laura Leites⁸, Nurul Islam-Faridi⁹, Jeanne Romero-Severson¹⁰, Victoria DeLeo⁸, Sarah Lucas⁸, Desanka Lazic¹¹, Oliver Gailing¹¹, John Carlson⁸, and Margaret Staton¹

¹University of Tennessee

²HudsonAlpha Institute for Biotechnology

³US Department of Energy

⁴University of Kentucky

⁵Michigan Technological University

⁶University of Missouri

⁷North Carolina State University at Raleigh

⁸Pennsylvania State University

⁹USDA Forest Service Southern Research Station

¹⁰University of Notre Dame Department of Biological Sciences

¹¹University of Göttingen

May 16, 2023

Abstract

Northern red oak (*Quercus rubra* L.) is an ecologically and economically important forest tree native to the northeastern United States. We present a chromosome-scale, haplotype-resolved genome of *Q. rubra*, a representative red oak species, generated by the combination of PacBio sequences and chromatin conformation capture (Hi-C) scaffolding. This is the first reference genome from the red oak clade (section Lobatae). The *Q. rubra* assembly spans 739 Megabases (Mb) with 95.27% of the genome sequences scaffolded into 12 chromosomes and 33,333 protein-coding genes. Comparisons to the genomes of *Q. lobata* and *Q. mongolica* reveal high collinearity, with intrachromosomal structural variants present. Orthologous gene family analysis with other oak and rosid tree species revealed that gene families associated with defense response were expanding and contracting simultaneously across the *Q. rubra* genome. *Quercus rubra* had the most CC-NBS-LRR and TIR-NBS-LRR resistance genes out of the nine species analyzed. Terpene synthase gene family comparisons further reveal tandem gene duplications in TPS-b subfamily, similar to *Q. robur*. Single major QTL regions were identified for vegetative bud break and marcescence which contain candidate genes for further research, including a putative ortholog of the circadian clock constituent cryptochrome (CRY2) and a family of eight tandemly duplicated genes for serine protease inhibitors, respectively. Genome-environment associations across natural populations identified candidate abiotic stress tolerance genes and predicted performance in a common garden. This high-quality red oak genome represents an essential resource to the oak genomics community which will further supplement the knowledge of *Quercus* genomics.

Hosted file

Red_Oak_Genome_MER_v2.docx available at <https://authorea.com/users/618769/articles/643714-haplotype-resolved-chromosome-scale-genome-assembly-of-quercus-rubra-1>