

Sequencing and Annotation of 10 Loblolly Pine BACs

Anna Stambolia-Kovach¹, Jill L. Wegrzyn¹, Ian Korf², Andrew J. Eckert¹, Mark Yandell³, Carson Holt³, George E. Bruening¹, Douglas R. Cook¹, Charles H. Langley¹, David B. Neale¹

Loblolly pine has gained the interest of geneticists due to its importance as a commercial species and its emerging role as a dominant species in many threatened terrestrial ecosystems. While extensive mapping and association studies have shed light on locus-specific traits and processes, the enormous size of the pine genome has to date prohibited genome-wide sequencing and annotation. In order to gain insight into the content, structure, and organization of the relatively unexplored conifer genome, we sequenced ten loblolly pine bacterial artificial chromosomes (BACs), totaling nearly 1 Mb of genomic sequence. The BACs were selected by screening a BAC library for clones that contained genes associated with wood property traits. We used shotgun sequencing to acquire at least 6X coverage of each genomic insert through the standard Sanger sequencing method. The phred/phrap software package was implemented for automated basecalling and initial contig assembly, and the Arachne whole genome assembler was used to finish the scaffold assemblies. Automated annotation followed by manual curation revealed 19 putative genes in eight of the ten BACs sequenced. The gene set includes at least one lignin gene (*ccoamt*) that has previously been described in loblolly pine and other conifers. Perhaps more interesting, the gene set is largely comprised of putative homologues to previously described genes in pine. These potential genes include novel methyltransferases, kinases, LIM-domain transcription factors, chromatin remodelers, and S-adenosylmethionine synthetases. Since several of the known homologues to the novel genes are involved in wood property traits, these novel versions may prove to be important for the improvement of forest tree breeding. We offer the first structural and compositional analyses of large contiguous regions of the loblolly pine genome, the genetic patterns and properties contained therein, and a primer for future genomic exploration in conifer species.

¹ University of California at Davis

² University of California at Davis - Genome Center