

# Largest Populus SNP dataset holds promise for biofuels, materials, metabolites

January 17 2017

---



Scientists gleaned data from Populus trees in an ORNL greenhouse as part of the largest-ever single nucleotide polymorphism dataset of the species' genetic variations. The information can be useful in biofuels, materials science and secondary plant metabolism research. Credit: Genevieve Martin/Oak Ridge National Laboratory/US Dept. of Energy

Researchers at the Department of Energy's Oak Ridge National Laboratory (ORNL) have released the largest-ever single nucleotide polymorphism (SNP) dataset of genetic variations in poplar trees, information useful to plant scientists as well as researchers in the fields of biofuels, materials science, and secondary plant metabolism.

For nearly 10 years, researchers with DOE's BioEnergy Science Center (BESC), a DOE Bioenergy Research Center led by ORNL, have studied the genome of *Populus*—a fast-growing perennial tree recognized for its economic potential in biofuels production. Today, they released the Genome-Wide Association Study (GWAS) dataset that comprises more than 28 million [single nucleotide polymorphisms](#), or SNPs, derived from approximately 900 resequenced poplar genotypes. Each SNP represents a variation in a single DNA nucleotide, or building block, and can act as a biological marker, helping scientists locate genes associated with certain characteristics, conditions, or diseases.

The data "gives us unprecedented statistical power to link DNA changes to phenotypes [physical traits]," said Gerald Tuskan, a corporate fellow and leader of the Plant Systems Biology group in ORNL's Biosciences Division. Tuskan will present the GWAS data today at the Plant & Animal Genome Conference in San Diego. The results of this analysis have been used to seek genetic control of cell-wall recalcitrance—a natural characteristic of [plant cell walls](#) that prevents the release of sugars under microbial conversion and inhibits biofuels production.

BESC scientists are also using the dataset to identify the molecular mechanisms controlling deposition of lignin in plant structures. Lignin, the polymer that strengthens plant cell walls, acts as a barrier to accessing cellulose and thereby preventing cellulose breakdown into simple sugars for fermentation.

With the new poplar GWAS dataset, "we can identify the genes and

genetic variants [i.e. alleles] that move carbon through the lignin pathway, and then take that knowledge and, through genomic selection, develop plant materials that are tailored to work with microbes to yield the targeted product," Tuskan said. Such products include modified lignin customized for chemicals, polymers and materials. Although the dataset's most immediate applications are in plant science, ORNL researchers plan to use the GWAS data to inform bioscience work in areas such as cleaner, sustainable transportation fuels, carbon fiber for lightweight vehicles and alternatives to conventional plastics and building insulation materials.

Even the medical field could benefit from the work: ORNL researchers, for instance, have used the poplar GWAS to identify the genes that control callus formation, or cells covering a plant wound. The work has implications for cancer research.

"The genes related to callus formation are analogous to many genes involved in the formation of tumors in humans," Tuskan said. "This discovery, and the associated gene expression network surrounding such genes, could inform work related to the Cancer Moonshot," he added, referring to a federal initiative designed to speed progress in cancer research.

Tuskan, who holds a joint appointment at DOE's Joint Genome Institute in California, found inspiration for the work in the sequencing of the human genome about a decade ago. The researchers recognized how those types of studies could be used to address DOE challenges in carbon sequestration, bioprocessing and materials science.

Tuskan emphasized the importance of technological advances to the work. Sequencing capacity and computational abilities "made the work possible," he said. "We are working in the big data realm, and fortunately at the national lab we have the platforms and infrastructure

to do this type of analysis."

As part of their work, the researchers used the computational resources available at ORNL through its Compute and Data Environment for Science (CADES) program within ORNL's Computing and Computational Sciences Directorate, as well as the Titan supercomputer at the Oak Ridge Leadership Computing Facility, a DOE Office of Science User Facility.

The research also involves monitoring and cataloging phenotypes of [poplar trees](#) in regions from southern British Columbia to central California. "None of the sophisticated genomics and computational science would mean anything without the fieldwork. The genetics, the computational science, and measuring and cataloging phenotypes are the three legs of the platform we stand on at BESC," Tuskan said.

The researchers plan to expand the existing dataset and collaborate with other scientific groups to collect and analyze additional phenotypes.

Other ORNL scientists involved in the project include Wellington Muchero, Jay Chen, Daniel Jacobson, and Tim Tschaplinski. Contributing scientists at the Joint Genome Institute, which performed all the genetic sequencing, were Dan Rokhsar, Wendy Schackwitz, and Jeremy Schmutz. Steve DiFazio at West Virginia University's Department of Biology was also involved in the project. Mark Davis and others at DOE's National Renewable Energy Laboratory made contributions in characterizing the biochemistry of plant cell walls.

The dataset is available at: <http://bioenergycenter.org/besc/gwas>

Provided by Oak Ridge National Laboratory

Citation: Largest Populus SNP dataset holds promise for biofuels, materials, metabolites (2017, January 17) retrieved 26 April 2024 from <https://phys.org/news/2017-01-largest-populus-snp-dataset-biofuels.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.