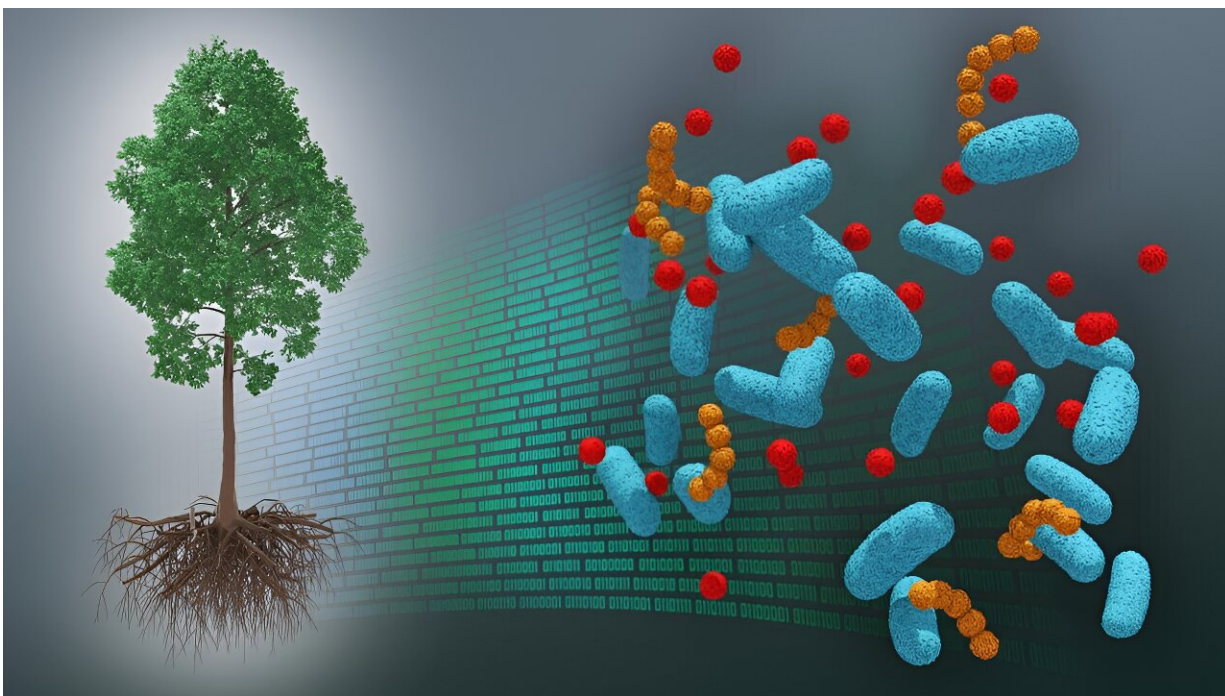


First-of-its-kind integrated dataset enables genes-to-ecosystems research

April 8 2024, by Stephanie G Seay



DOE national laboratory scientists led by Oak Ridge National Laboratory have developed the first tree dataset of its kind, bridging molecular information about the poplar tree microbiome to ecosystem-level processes. Credit: Andy Sproles/ORNL, U.S. Dept. of Energy.

A team of Department of Energy scientists led by Oak Ridge National Laboratory has released the first-ever dataset bridging molecular information about the poplar tree microbiome to ecosystem-level

processes. The project aims to inform research regarding how natural systems function, their vulnerability to a changing climate, and, ultimately, how plants might be engineered for better performance as sources of bioenergy and natural carbon storage.

The data, [described in *Scientific Data*](#), provides in-depth information on 27 genetically distinct variants, or genotypes, of *Populus trichocarpa*, a poplar tree of interest as a bioenergy crop.

The genotypes are among those that the ORNL-led Center for Bioenergy Innovation previously included in a [genome-wide association study](#) linking genetic variations to the trees' physical traits. ORNL researchers collected leaf, soil and root samples from poplar fields in two regions of Oregon—one in a wetter area subject to flooding and the other drier and susceptible to drought.

Details in the newly integrated dataset range from the trees' genetic makeup and [gene expression](#) to the chemistry of the soil environment, analysis of the microbes that live on and around the trees and compounds the plants and microbes produce.

The dataset "is unprecedented in its size and scope," said ORNL Corporate Fellow Mitchel Doktycz, section head for Bioimaging and Analytics and project co-lead. "It is of value in answering many different scientific questions." By mining the data with machine learning and statistical approaches, scientists can better understand how the genetic makeup, physical traits, and chemical diversity of *Populus* relate to processes such as the cycling of soil nitrogen and carbon, he said.

"The knowledge we generated from this one plant will be folded back into projects that produce biofuels from poplar," said Melanie Mayes, leader of ORNL's Ecosystem Processes group and a collaborator on the project. "The procedure we built here will be needed for bioengineering

of other plants and to help us build climate resilience—to advance soil carbon storage and reduce greenhouse gas emissions."

The complete dataset comprises more than 25 terabytes. Links to the data are [available as part of the National Microbiome Data Collaborative](#), or NMDC, a DOE initiative supporting data-sharing on the association of microbiomes with environmental processes.

"The dataset represents the largest publicly available metagenomics repository on a tree endosphere," the plant tissue environment that is home to complex microbial communities, said Christopher Schadt, project co-lead and ORNL distinguished staff scientist.

Detailed analyses of the samples resulted in 318 metagenomes, revealing the diversity of microbes living in and around trees through genetic sequencing. Ninety-eight plant transcriptomes provide information on the full range of messenger RNA molecules expressed in the plant roots. The dataset includes 314 metabolomic profiles, supplying information on the small molecules produced by plants and microbes as they grow or in response to stress.

Data are also included on associated soil physical and biogeochemical characteristics, examining chemicals present and how they cycle through the environment.

Integrating this "multi-omics" data will provide essential information to scientists studying how plant-related molecular and cellular events are connected to ecosystem processes and behaviors.

Understanding plant, soil nitrogen cycling triggers

The Joint Genome Institute, a DOE Office of Science user facility at Lawrence Berkeley National Laboratory, was a close collaborator on the

project. JGI led the metabolomics profiling of the leaf, root, and soil environment, or rhizosphere, the plant root transcriptomics sequencing, and the soil rhizosphere and endosphere metagenomics work.

"The combination of metagenomics and metabolomics from leaf, root, and soils, along with *Populus* host transcriptomes, make this a truly unique dataset for the research community and could serve as a central data resource to explore plant-microbe interactions," said Emiley Eloe-Fadrosh, Metagenome Program head at JGI.

The project began as an ORNL pilot called Bio-Scales, supported by the Biological Systems Science Division in the DOE Office of Science's Biological and Environmental Research program. Bio-Scales pursues a better understanding of the plant-microbe relationship with a focus on nitrogen cycling. Nitrogen is an essential nutrient for life, but when overused in agriculture and other applications, it can harm water quality or be emitted as the potent greenhouse gas nitrous oxide, or N_2O .

"The project required the integration of a lot of diverse expertise," Doktycz said. "It started with a team who went out in the midst of COVID-19 to collect all these diverse materials and got them back to the lab, then prepared, analyzed and extracted data from them. We also had an incredible technical support team who processed hundreds of these samples in a tracked and coordinated way, interfacing with the Joint Genome Institute for the sequence analysis."

In addition to its size and scope, the dataset stands out as being heavily annotated with metadata—with precise details, for instance, on where and how the sampling took place, and a standard format for subsequent data reporting. Adding those elements to data makes information easier to find, understand, and reuse.

ORNL's Stanton Martin, who led data management for the project in

close coordination with the NMDC, noted that the data-first approach supports artificial intelligence and other analytical approaches to help resolve scientific questions.

"The data management we performed on this project is hugely valuable to data practices for other projects like the Plant-Microbe Interfaces Scientific Focus Area and the Center for Bioenergy Innovation at ORNL. It plays to ORNL's strengths in what I call [data management](#)'s three V's—data volume, variety, and velocity—and allowed us to take a first step in integrating very large 'omics data in a way that has not been done before."

The project started with Schadt and Mayes traveling to Oregon for sampling. "It normally would have been six scientists, but we had travel restrictions on groups traveling together due to the pandemic," Schadt said. They also had to work around encroaching wildfires, as Oregon experienced an active fire season that year. Schadt and Mayes worked with the assistance of Oregon State University volunteers to gather extensive geotagged samples at the two sites.

Beneficial bioengineering

Mayes said the project "gets at the role of genes in influencing not just the fate of the plant itself, but also the environment around it, such as the soil. For instance, we wanted to understand the potential of soil microbes to either make more nitrate or remove excess nitrate from the system. We wanted to learn more about how plant genomics influence what soil microbes are doing."

Knowing more about the plant and soil nitrogen cycle can affect emissions of N₂O, a gas that accounts for 6% of all [greenhouse gas emissions](#) in the United States.

"If you know which genes to target that result in the minimization of N₂O or nitrate production, then you have the potential to affect both greenhouse gas-related warming and water quality," Mayes said. "You could, for instance, select and further bioengineer plants with the best genetic profile for controlling these emissions."

"This project is unique because it gets at the connection between plant genomes and environmental outcomes like nitrous oxide emissions or nitrate production," Mayes said. "Building one of the first comprehensive datasets on the plant-microbe relationship also tells us how much we still can learn."

More information: Christopher Schadt et al, An integrated metagenomic, metabolomic and transcriptomic survey of *Populus* across genotypes and environments, *Scientific Data* (2024). [DOI: 10.1038/s41597-024-03069-7](https://doi.org/10.1038/s41597-024-03069-7)

Provided by Oak Ridge National Laboratory

Citation: First-of-its-kind integrated dataset enables genes-to-ecosystems research (2024, April 8) retrieved 2 May 2024 from <https://phys.org/news/2024-04-kind-dataset-enables-genes-ecosystems.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.