Plant-Microbe Interfaces: High-Throughput Comparative Genomics for *Populus*-associated Microbes

Visanu Wanchai, Se-Ran Jun, Intawat Nookaew, Miriam Land, Preecha Patumcharoinpol, Michael Leuze, and **David W. Ussery*** (usserydw@ornl.gov)

Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

http://PMI.ornl.gov

Project Goals: The goal of the PMI SFA is to understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serves as the experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we focus on 1) characterizing host and environmental drivers for diversity and function in the *Populus* microbiome, 2) utilizing microbial model system studies to elucidate *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships and 3) develop metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

We have characterized 179 *Populus*-associated bacterial genomes. Based on full-length protein alignments, we find a core of only 11 protein families conserved among all 179 genomes, and a total of 294,668 unique protein families. Based on a larger set of more than 70,000 available genome sequences, we find that, within a given bacterial species, as much as 80% of the proteins in any one genome belong to the core. The core-genomes for *Populus*-associated bacterial genomes can be used to estimate basic metabolic pathways for a given species, which can be contrasted with strain-specific additional metabolic capabilities. To further explore functional space, we found a total of 7,343 unique protein functional domains in the 179 proteomes, with 47,075 architectures. In the larger set of 70,502 bacterial genomes, we find 12,496 total unique PfamA domains (12,496 out of 16,230 total domains in the database, or 77% of all domains). By storing all of the PfamA domain information in a graph database, we can quickly identify (within a few seconds) transcription factors unique to a specific taxonomic group, as well as from PMI genomes associated with a specific environment (e.g., endosphere, rhizosphere, tree). For example, we find that for the 179 PMI proteomes, 1,047 of the 18,225 architectures (5.7%) contain transcription factor domains, and a slightly larger number was found for all 70,502 genomes (14,111 out of 195,778, 7.2%). This large fraction of transcription factor architectures implies an evolutionary selection for diversity of regulatory proteins. Finally, we also examine sigma factors across all of the genomes.

The Plant Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S. Department of Energy, Office of Science, Biological and Environmental Research.