Plant-Microbe Interfaces: Sequencing of twenty-one *Pseudomonas* genomes and twenty-three genomes from diverse bacteria isolated from *Populus deltoides* rhizospheres and endospheres.

Sagar M. Utturkar[1,*] (sutturka@utk.edu), **Steven D. Brown**[1,2], Dawn M. Klingeman[2], Courtney M. Johnson[2], Stanton L. Martin[2], Miriam L. Land[2], Tse-Yuan S Lu[2], Christopher W. Schadt[1,2], **Dale A. Pelletier**[1,2], and **Mitchel J. Doktycz**[1,2]

[1]Genome Science and Technology Program, University of Tennessee, Knoxville TN; [2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

http://pmi.ornl.gov and http://genome.ornl.gov/microbial/PMI/

Project Goal: 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 an initial 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 will focus on 1) characterizing the natural variation in *Populus* microbial communities within complex environments, 2) elucidating *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships, and 3) performing metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

Plant-microbe interactions within the rhizosphere have been shown to have important roles in plant health and productivity (*1, 4*). As part of an ongoing effort to better understand the microbial communities associated with native *Populus deltoides* (Eastern cottonwood), we have under taken both cultivation independent and cultivation dependent assessments of microbial communities from *P. deltoides* rhizospheres and endospheres (*3*). Our goal is to understand the diversity of the *Populus* microbiome and to elucidate the metabolic and molecular mechanisms responsible for shaping the *Populus*-microbial interface.

We sampled *P. deltoides* at sites along the Caney Fork River in central Tennessee and at Yadkin River in North Carolina, USA. These sites represent ecotypes and soil conditions that are common to this region. Approximately 1,100 diverse bacterial strains were isolated from these sites over multiple seasons. The isolates comprise 7 classes and 85 genera of bacteria including, *Actinobacteria* (14%), *Bacilli* (17%), *Flavobacteria* (6%), *Sphingobacteria* (3%), and α- (22%) β- (16%) and γ- (22%) Proteobacteria. A number of our isolates were *Pseudomonas* species, which have considerable genetic and phenotypic variability and different members are pathogenic, biocontrol and plant growth promoting bacteria. Representatives were chosen for genome sequencing based on phenotypic traits and phylogeny. A total of 24 γ-Proteobacteria, 6- β-Proteobacteria, 10 α-Proteobacteria, 2-Bacteroidetes and 2-Bacilli were sequenced.

Paired-end DNA libraries with an average insert size of 500bp were created and draft genome data was generated using the Illumina (*2*) HiSeq2000 technology. CLC Genomics Workbench

(version 4.7.1) and FASTQC (*6*) were applied to trim reads for quality sequence data. We analyzed the effect of quality based trimming and filtering of reads on genome assemblies and compared assembly outputs generated by the Velvet assembler (Version 1.1.04) (*7*) and the CLC Genomics Workbench software. In each case found the selection of high quality reads was a key step for successful assemblies. Quality trimming dramatically improved the assembly i.e. reduced the number of contigs while maintaining the expected genome size. Velvet assemblies were further optimized by selecting different Kmer values and employing a scaffolding algorithm within the program. In the case of *Rhizobium* sp. PDO-076 a GS FLX shotgun dataset was also available, which was used to generate a hybrid assembly by combining 454 reads and shredded Velvet scaffolds using Newbler (Version 2.6). Initial assembly validation was performed based on parameters such as number of contigs, genome size, N50, contig length etc.

Draft genome sequences were annotated at Oak Ridge National Laboratory using the Microbial Genome Annotation Pipeline automated annotation pipeline at ORNL, which is based on the Prodigal gene prediction algorithm (*5*). Final validation was performed by annotating the assembled genome with ORNL genome annotation pipeline and comparing with known conserved sequences. We assembled diverse genomes that varied in size from 4.4MB to 10.7MB and had G+C contents in the range of 33% to 69%. The 4.9 Mb of *Phyllobacterium* genome was assembled in as few as 39 contigs with largest contig size of 1.4 Mb, while maximum number of contigs generated was the 778 for the 5.4 Mb *Caulobacter* genome. The N50 statistics varies from 14,172 to 506,356 bp with average value of 154,391 bp. These new genome sequences will allow more comprehensive comparisons for bacteria involved in plant-microbe interactions.

1. **Bais, H. P., T. L. Weir, L. G. Perry, S. Gilroy, and J. M. Vivanco.** 2006. The role of root exudates in rhizosphere interations with plants and other organisms. Annu. Rev. Plant Biol. **57:**233-266.
2. **Bennett, S.** 2004. Solexa Ltd. Pharmacogenomics **5:**433-438.
3. **Gottel, N. R., et al.** 2011. Distinct microbial communities within the endosphere and rhizosphere of *Populus deltoides* roots across contrasting soil types. Appl. Environ. Microbiol. **77:**5934-5944.
4. **Hodge, A., G. Berta, C. Doussan, F. Merchan, and M. Crespi.** 2009. Plant root growth, architecture and function. Plant Soil **321:**153-187.
5. **Hyatt, D., G. L. Chen, P. F. LoCascio, M. L. Land, F. W. Larimer, and L. J. Hauser.** 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11:**119.
6. **Kircher, M., P. Heyn, and J. Kelso.** 2011. Addressing challenges in the production and analysis of illumina sequencing data. BMC Genomics **12:**382.
7. **Zerbino, D. R., and E. Birney.** 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. **18:**821-829.

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