11. Metagenomic Insight into the Rhizospheres of Three Biofuel Crops

James R. Cole[1*] (colej@msu.edu), Aaron Garoutte[2], Jiarong Guo[1], Jordan Fish[1], Adina Howe[1,4], Bangzhou Zhang[1], Chao Xue[1], Qiong Wang[1], Caio Rachid[1], Elaine Gomes[1], C. Titus Brown[2,3] and **James M. Tiedje**[1,2]

[1]Center for Microbial Ecology, [2]Microbiology and Molecular Genetics, [3]Computer Science and Engineering, Michigan State University, East Lansing, MI 48824; [4]Argonne National Laboratory, Lemont, IL 60439

Project Goals: To improve the sustainable development of bioenergy, our project characterizes beneficial microbes in three important bioenergy crop systems and their impacts on critical biogeochemical processes, especially the nitrogen cycle, and ultimately explores ways to manage such microbiomes for improved biofuel sustainability.

An important issue in producing biofuel feedstock is long-term environmental sustainability. In previous years, we developed approaches to explore the high diversity present in complex communities of agricultural soils, with an emphasis on genes involved in nitrogen availability and cycling. In Phase II of this project, we apply approaches to identify critical biogeochemical cycling genes and populations in a rigorous sampling of multiple localized rhizosphere soil communities (n=7) from three major bioenergy crops: switchgrass, Miscanthus, and corn (continuous), from GLBRC's Kellogg Biological Station (KBS) intensive sites. This depth of replicated sampling significantly extends the statistical power of previous studies to identify relevant drivers of beneficial plant-microbe interactions and nitrogen cycling genes.

Each of all 21 samples was shotgun sequenced as one lane in Illumina HiSeq by JGI and produced high yield and good quality sequence reads. To mine this large volume of data for gene information, we developed a set of data mining processes/techniques, including our novel digital normalization (diginorm) and partitioning method for big data assembly, SSU rRNA gene fragment finding and analysis method, our Xander tool for gene targeted assembly from shotgun data, and N-cycle gene finding tools from shotgun data. Together, these tools are being used to reconstruct representative genetic references for bioenergy crop soil microbial communities in the rhizosphere.

Overall, the taxonomic composition of rhizosphere samples were consistent between using our SSU rRNA fragment finding method and our diginorm/partitioning assembly pipeline. They both implicated corn as the most different among the three microbiomes. Compared to the two perennial grasses, switchgrass and Miscanthus, corn-samples were more abundant in Proteobacteria but less in Acidobacteria. Proteobacteria are usually fast-growing and Acidobacteria are slow growers, consistent with corn growing new roots every year providing more new C for selection. The OTU based diversity analysis also suggests that corn-associated communities are less rich and diverse and significantly different from switchgrass and Miscanthus microbiomes.

Soil metagenomes were annotated against known nitrogen genes in the MG-RAST database. Novel nitrogen genes were also assembled using our own tool (Xander). Annotations of 10M read subsets of the short read replicates by MG-RAST showed that the relative abundances of N- cycle genes of the three crops are significantly different and, again, corn has the least abundance of the three. We found our Xander gene-targeted assembler provides higher sensitivity in determining contents of specific genes. We used Xander on all seven Miscanthus replicates (a total 1.6 billion reads or 100 billion

unique 30-mers after paired-end assembly and quality filtering) and found 60 unique contigs, and that the majority of contigs had a mean coverage of less than three, demonstrating that this technique can assemble genes with low coverage (and missed by any global assembly method). These contigs showed close relatedness(dissimilarity < 10%) to their closest *nifH* reference sequences, with the top two groups similar to family Rhodospirillales (69.6%), which includes Azospirillum, while another 15.7% were most similar to the Rhizobiales. Xander was also used to assemble 995 contigs of *nirK* gene from 50G sequences and examined the targeted sites for 12 *nirK* gene PCR primers. We found all literature primers would miss about 95% of soil *nirK* sequences (≤ 2 mismatches), while validated best MSU-developed primer provided a calculated > 90% of coverage (≤ 2 mismatches).

We also performed amplicon sequencing and analysis using bacterial 16S (V4), fungal (28S), and *nifH* genes from soil cores collected in 2013 at intensive experimental sites at KBS and Arlington from five different biofuel cropping systems: Continuous corn (G1), Continuous corn plus Cover crop (G2), Switchgrass (G5), Miscanthus (G6), and Native prairie (G10). We found clear differences between corn main plots (stover removal) and microplots (stover non-removal) both in G1 and G2 plots. Furthermore, the bacterial communities in the three biofuel crops were very different from native prairie, and the communities in continuous corn were more different from other two biofuel crops than the latter were to each other, consistent with the shotgun data analyses of the rhizosphere samples.

A benchmarking study was carried out to identify genes expressed in soil associated with the biofuel crop Miscanthus at the KBS site. Initial results indicated an increased abundance of transcripts derived from housekeeping genes and phage in the later sample, suggesting the possible effect of Miscanthus on gene expression of the soil microbial community.

This work was funded by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494).