

## **Microarray Design for Linking Agroecosystem and Gene Function in Soil Microbial Communities**

David Duncan<sup>1,2\*</sup> (dsduncan@wisc.edu), Richard LeDuc<sup>1</sup>, James Tiedje<sup>2,3</sup>, **Randall Jackson**<sup>1,2</sup>

<sup>1</sup>University of Wisconsin–Madison; <sup>2</sup> DOE Great Lakes Bioenergy Research Center, Madison, WI; <sup>3</sup>Michigan State University, East Lansing

<http://www.glbrc.org/>

<http://agronomy.wisc.edu/jackson>

**Project Goals: (1) Identify microbial genetic annotations whose abundances contrast among metagenomes sequenced from functionally different agricultural and grassland soil environments; (2) Design a microarray platform to quantify the abundances of genes that differentiate among environments and known microbial functional markers linked to relevant biogeochemical cycling processes; (3) Evaluate the explanatory power of the relative abundances of these sequences over greenhouse gas and N-cycling fluxes in bioenergy cropping systems toward the end of identifying predictive and diagnostic relationships that can be used to inform land management decisions, identifying links between genes and environmental function, and exploring to what extent knowledge of soil microbial community composition improves our capacity to model biogeochemical cycles. This poster presents our progress on the first and second goals and the approach to be taken with the third.**

Among the expectations placed on bioenergy production systems is that they sustainably provide ecosystem services equal to or beyond those obtained from traditional agriculture. Research at the GLBRC and elsewhere provides evidence that these goals can be furthered by perennial, high diversity cropping systems whose low management intensity and spatial extent poses a challenge to efficient and effective monitoring. In order to provide land managers with timely information on the health and functioning of their agroecosystem, we need both an improved understanding of the fundamental drivers of key ecosystem functions and novel techniques for assessing agroecosystem health.

For biogeochemical cycling, the soil microbial community presents a promising target as both a major driver of observed agroecosystem processes and a sensitive integrator of climatic, edaphic, and agronomic effects. However, the size and complexity of the soil microbial community pose a serious technical challenge to meaningfully operationalizing composition metrics. Shotgun metagenomic sequencing has greatly advanced our understanding of the genetic composition microbial communities, particularly since the advent of Next-Gen sequencing and improvements in the computational approaches to handle such data. Despite these advances, metagenomic sequencing remains a nontrivial technical challenge, making it ill-suited for widespread diagnostic use or studies with high biological replication. Narrower approaches to community characterization, such as targeted sequencing and microarrays, provide more tractable datasets,

but are limited in their scope to a predefined universe of genes and sequences. In this project, we are combining the open-ended exploratory benefits of shotgun metagenomic sequencing with the throughput and turnaround time of a microarray.

While generalized microbial community functional gene platforms exist, for this project we are designing a custom array populated primarily by sequences from the bulk soil and rhizosphere metagenomes of potential bioenergy crops grown in the Upper Midwest generated by the GLBRC. This will prevent the dilution of relevant sequences with sequences drawn from exotic environments and unlikely to be found in soil systems. In addition, we employ these metagenomes in a novel approach to identify candidate genes for empirical functional association. We identify a set of gene annotations with contrasting abundances among different metagenome sources (e.g. between rhizosphere and the bulk soil), then assay the abundance of these gene groups across a range of systems, potentially linking the genes to a system-scale process of interest. This helps mitigate the dependence of microarrays on *a priori* target selection without need to resort to a blind search. By selecting genes with systematically contrasting abundance in soil environments that have different properties, we expect to identify a gene set enriched for sequences that either drive ecosystem-level functions, or are highly responsive to the drivers of such functions.

In the upcoming growing season, we will use these newly-defined arrays to characterize the microbial community composition associated with potential bioenergy crops grown in the GLBRC Bioenergy Cropping Systems Trial. We will combine this compositional information with trace gas and N-cycling measurements generated through the GLBRC's sustainability research into biogeochemical cycling to identify connections between aspects of community composition and observable ecosystem properties. Going forward, we will revise the contents of the array to reflect our changing understanding of the composition of these soil communities and of their operational units, working toward the development of a platform for diagnosing agroecosystem health and functioning and toward a better understanding of the role of microbial community composition in agroecosystem biogeochemistry.

***Funding source:*** DOE Great Lakes Bioenergy Research Center; DOE BER Office of Science DE-FC02-07ER64494; DOE OBP Office of Energy Efficiency and Renewable Energy DE-AC05-76RL01830