

Rhizosphere metagenomics and metatranscriptomics of biofuel crops

Jiarong Guo^{1*}(guojiaro@gmail.com), Aaron Garoutte¹, Adina Howe², Qiong Wang¹, Titus Brown³, James Cole¹, **James Tiedje**¹

¹Michigan State University, East Lansing; ²Iowa State University, Ames; ³University of California, Davis, Davis

<https://www.glbrc.org/research/sustainability>

Project Goals: Our research supports a goal of the Great Lakes Bioenergy Research Center (GLBRC) which is to develop sustainable biofuel practices by optimizing soil, microbe, and plant interactions, since soil microbes form beneficial association with the crops' rhizospheres and also play a major role in ecosystem functions, such as N and C cycles. Our main goal is to find key eco-functional genes selected by biofuel crop rhizospheres, with a focus on nitrogen cycle genes, which is critical for sustainable biofuel production on marginal lands.

We applied deep sequencing to study rhizosphere microbiomes of biofuel crops. Through the past few years of GLBRC, our group has developed several tools (SSUsearch, khmer, and Xander) to deal with big metagenomic data that can not be analyzed by traditional bioinformatics methods. These tools enable us to explore the genetic and population components of rhizosphere communities, with special emphasis on ecological services. Here we showcase two deep sequencing enabled rhizosphere microbiome studies: 1) rhizosphere metagenomics of three biofuel crops; 2) core functional microbiome of switchgrass.

Rhizosphere metagenomics of three biofuel crops: Crop roots had strong influences on the soil microbial community. Thus large-scale plantation of biofuel crops will have significant impact on ecosystem functions regionally and beyond. We compared rhizosphere microbial communities of corn (annual) and switchgrass and *Miscanthus* (perennials). This is the first comparative study of these biofuel crops using shotgun metagenomics and one of the largest sequencing efforts to date (about 1 TB bp in total). We compared the rhizosphere metagenomes at three levels: overall community structure (SSU rRNA gene), overall function (annotation from global assembly), and N cycle genes (from Xander). All three levels showed corn had a significantly different community from *Miscanthus* and switchgrass (except for AOA). In terms of life history strategy, the corn rhizosphere was enriched with more copiotrophs while the perennials were enriched with oligotrophs, which is further supported by higher abundance of genes in "Carbohydrates" and higher fungi/bacteria ratios. In addition, corn also had a less rich and even community, so the perennials managed to maintain a more diverse community even though investing less C in the rhizosphere. Moreover, a larger dispersion of corn data in ordination plots and enriched *Penicillium* (non-beneficial fungi) also indicate corn may not be doing as well in controlling its community and selecting beneficial member. Furthermore, the nitrogen fixing community of corn was dominated by *Rhizobium* (perhaps a legacy from prior legume crops)

while the perennials had NifH sequences most related to *Coralimargarita*, *Novosphingobium* and *Azospirillum*, indicating that the perennials can better select beneficial members. Moreover, higher numbers of genes for nitrogen fixation and lower number of genes for nitrite reduction suggest better nitrogen sustainability of the perennials. Thus our study provides comprehensive evidence showing perennial bioenergy crops have advantages over corn in higher microbial species and functional diversity and in selecting members with beneficial traits, consistent with a higher level of sustainability of perennial biofuel crops.

Core functional microbiome of switchgrass: We used a “multi-omic” approach to establish a minimum core of actively transcribed functions in rhizosphere soils of switchgrass. We used metagenomics and metatranscriptomics to identify shared transcribed functions in samples. The minimum functional core accounts for 99% or 92% of functional annotations depending on annotation database (SEED Subsystems or RefSeq respectively). We then used metaproteomic data to further confirm our minimum functional core. The metaproteome data had 460 unique SEED Subsystem annotations with an abundance of 876,429. All but 12 of the SEED Subsystem annotations were found in the minimum functional core further validating the minimum functional core. The Subsystems with the greatest relative abundance in the metatranscriptome minimum functional core are the Clustering-based subsystem (defined as a cluster of genes known to work together but for which no function has been established), Protein metabolism, Carbohydrates (partly housekeeping functions and partly functions related to sugar utilization), Miscellaneous and RNA metabolism. Obviously the majority of these functions are either “known unknowns” or functions related to housekeeping. These results identify transcripts expressed at high levels and hence targets for further investigation of their function.

Our research is supported by GLBRC, a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.