Great Prairie Grand Challenge Soil Metagenome Sequencing Project

Janet K. Jansson[1,2]* (jrjansson@lbl.gov), Titus Brown[3], Patrick Chain[2], James Cole[3], Maude David[1], Jordan Fish[3], Adina Howe[3], Regina Lamendella[1], Kostas Mavromattis[2], Susannah Tringe[2], and James M. Tiedje[3].

[1]Lawrence Berkeley National Laboratory; [2]Joint Genome Institute; [3]Michigan State University;

Project Goals: Our overarching goal is to use deep metagenome sequencing to gain an understanding of the soil microbial community phylogenetic and functional gene repertoire in the U.S. great prairie and adjacent cultivated soils. This information will be used to determine the impact of land management (i.e. cultivation) on microbial community structure and function. In addition, this information will provide a database of microbial genes in native prairie soils that can serve as a baseline for understanding the impacts of perturbation and climate change on processes including soil carbon cycling. In order to achieve these goals the consortium is developing novel tools for metagenome sequencing, assembly, annotation and gene mining.

The United States Great Prairie contains approximately 35% of the total soil organic carbon stocks in the continental U.S. These carbon pools may be impacted in unknown ways by altered climate conditions and by land management. At the Joint Genome Institute, the U.S. Great Prairie was chosen as a "grand challenge" project for deep metagenome sequencing of a complex microbial community. This project is particularly challenging because of the high microbial diversity in soil that complicates metagenome assembly.

To date we have sequenced nearly 2 terabases of DNA from eight locations: 1) Wisconsin native prairie, 2) Wisconsin cultivated corn, 3) Wisconsin cultivated switchgrass, 4) Wisconsin restored prairie, 5) Iowa native prairie, 6) Iowa cultivated corn, 7) Kansas native prairie and 8) Kansas cultivated corn. Each site was sampled along a quadrant for a total of 8 samples per site. All of the samples were first profiled for their microbial community compositions by 454 pyrotag sequencing. In addition, the DNA from a central core from each location was extensively shotgun sequenced using the Illumina platform. A new algorithm was developed for assembly of the metagenome data. Different screening tools were developed to screen the assembled data and raw reads for key genes of interest in the carbon and nitrogen cycle.

454 pyrotag sequence data. We found that cultivation has a major impact on the composition of the soil microbial communities at all locations. There was similarity in the microbial community structures from the prairie soils in the three different locations. Some specific microbial species were more or less abundant in prairie soils, compared to cultivated soils, thus potential indicators of land use history. For example some Bradyrhizobia were more abundant in all of the native prairie soils. The restored prairie site had a microbial community composition that was intermediate between the cultivated corn and the native prairie soils, suggesting that this prairie community was progressing towards a restored (native) state.

Metagenome assembly. A novel approach was developed for metagenome assembly based on data reduction by filtering and normalization, followed by dividing the data into smaller disconnected sets to aid in assembly. This approach was tested on the Iowa corn and Iowa prairie metagenomes and is currently being applied to the remaining metagenomes. For more details about this approach, please see the poster presented by Adina Howe.

Metagenome mining. We first used a Hidden Markov Model (HMM) approach to screen for *nif*H genes in the Iowa corn and prairie assemblies. In addition, we developed a set of validated genes for different key pathways, including cycling of carbon and nitrogen. This gene database consists of "maudules" that represent individual pathways of interest. Currently we are using the maudules to mine the Iowa corn and prairie and the Kansas prairie metagenomes.

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231