Metagenome Analysis and Draft Genome Reconstruction of Produced Water Samples from Coalbed Methane Environments

Daniel E. Ross,[1,2]* Djuna Gulliver[1]
[1]National Energy Technology Laboratory, U.S. Department of Energy, [2]AECOM, (*Daniel.Ross@netl.doe.gov)

Research & Innovation Center

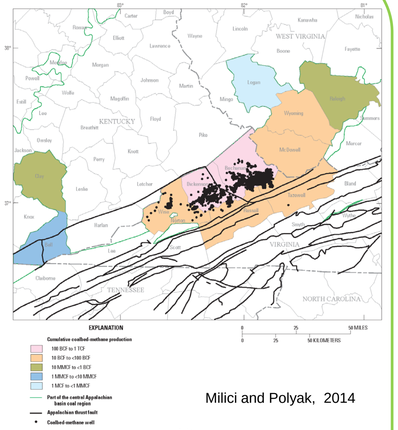
NATIONAL ENERGY TECHNOLOGY LABORATORY

Abstract

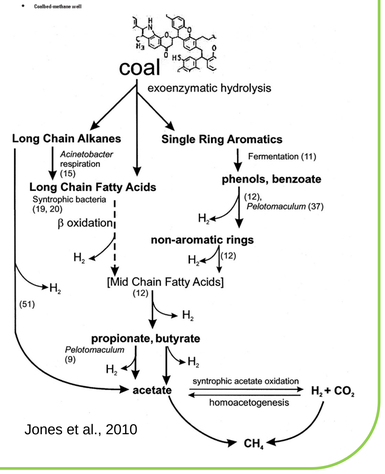
Biogasification is a process that utilizes the microbial community native to coalbeds to naturally convert currently unusable coal into readily available methane. One methodology involves injection of nutrients into the coal seams to stimulate biogenic coal degradation and methanogenesis. Identification of major functional pathways of biogenic coal degradation and subsequent methane production will lead to a better understanding of the coal-to-methane conversion, the microorganisms responsible for this conversion, and the nutrients required to bolster this conversion in situ. This study examines the metagenome of four produced water samples from the Central Appalachian Basin (Pocahontas 3 coal seam) to determine the composition (who's there) and the potential functional pathways (what can they do) of the resident microbial community. Nucleic acid was recovered from produced water samples using DNA isolation techniques and the quality and quantity of DNA was assessed. Illumina MiSeq next generation sequencing was employed, and the resultant nucleic acid sequence data was processed using a suite of bioinformatics software. Four metagenomes, named K34, K35, BB137, and L32A were obtained from produced water samples from a depth of 1704 ft, 1912 ft, 1980 ft, and 2578 ft, respectively. Methanogens were present in all samples, suggesting methanogenesis can occur. Furthermore, hydrocarbon degradation pathways were found, suggesting a route for biodegradation of coal. Importantly, a draft genome most closely related to *Pseudomonas stutzeri* CCUG was extracted from the K35 metagenome. Initial analysis of the draft genome revealed a complete nitrogen fixation pathway, and a naphthalene degradation pathway.

Goals:

- Investigate the microbial community in potential biogasification sites
- Characterize relevant functional pathways found in coal systems required for coal-to-methane bioconversion
- Construct draft genomes of abundant microorganisms in coal systems to complete a detailed characterization of prevalent functional potential

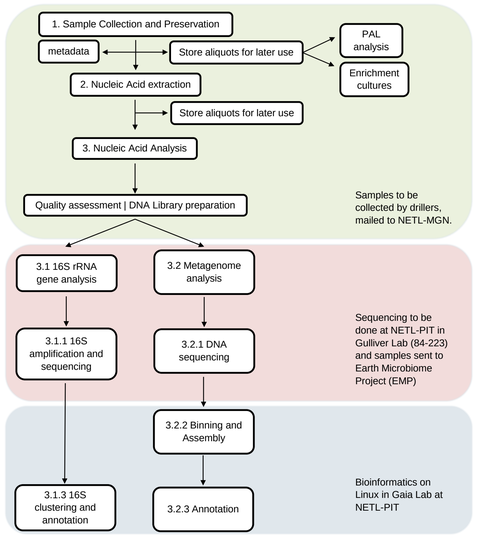


Milici and Polyak, 2014



Jones et al., 2010

Methods: Metagenome Analysis Pipeline (MAP)

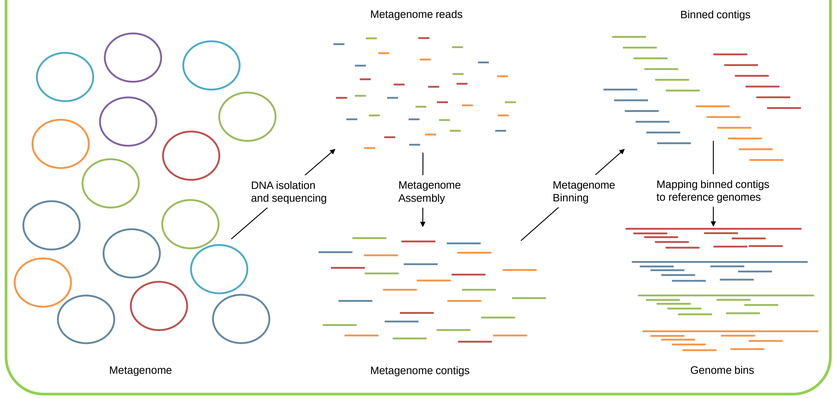


The first step in the metagenome analysis pipeline (MAP) involves careful and calculated sample collection. Samples are collected by drillers, or when possible on site by NETL researchers. Importantly, to preserve sample integrity and prevent nucleic acid degradation, samples are immediately aliquoted and frozen. After transport from the field, samples are thawed and processed for DNA extraction. The quality and quantity of DNA is assessed before preparing samples for sequencing.

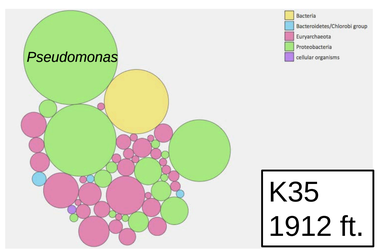
The second step involves processing DNA samples to generate a sequence library to be loaded into the sequencer (Illumina MiSeq). Processing involves cleaning, barcoding, and pooling DNA samples.

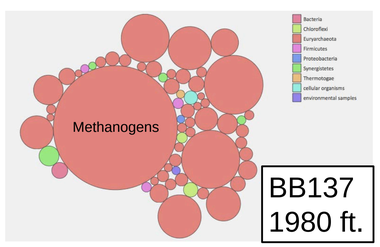
The third MAP step is the most time-consuming and computationally intensive. Here, data that is retrieved from the sequencer is processed and analyzed. Processing involves removing barcodes and trimming reads based upon the quality score (a measure of the confidence of each base call). Analysis involves metagenome assembly, binning, and annotation.

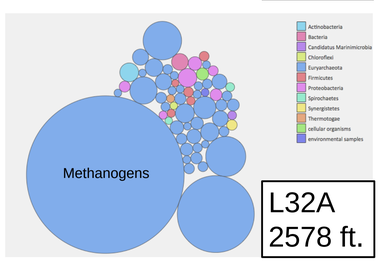
Metagenome Analysis



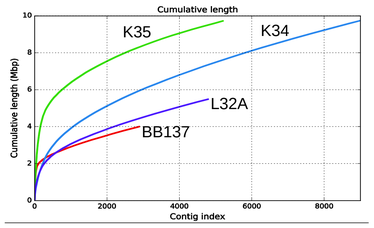
Taxonomic assessment of metagenome





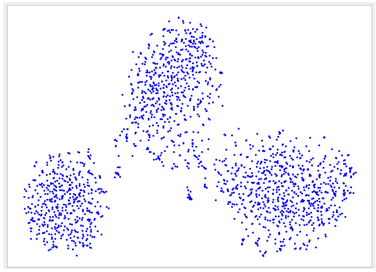


Metagenome Assembly



Sequence reads were assembled and contig length was plotted against contig number. A steeper slope represents a better assembly.

Metagenome Binning

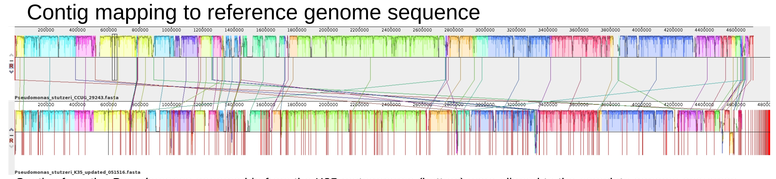


Contigs from the K35 metagenome assembly were binned. Each dot represents a contig and each cluster represents a potential genome.

Metagenome Results. Four metagenomes were analyzed and classified taxonomically. Generally, all samples contained Bacteria and Archaea, mostly comprised of *Proteobacteria* and *Euryarchaeota.* With the exception of K35, all metagenomes were dominated by Methanogens from the order *Euryarchaeota.* Short DNA sequencing reads (250 bp) were assembled into longer contigs. The contigs were binned according to genomic signature. Each bin was individually analyzed for genome completeness by comparing contigs to a reference marker gene set. Based upon the presence or absence of these marker genes, the % genome completeness was estimated.

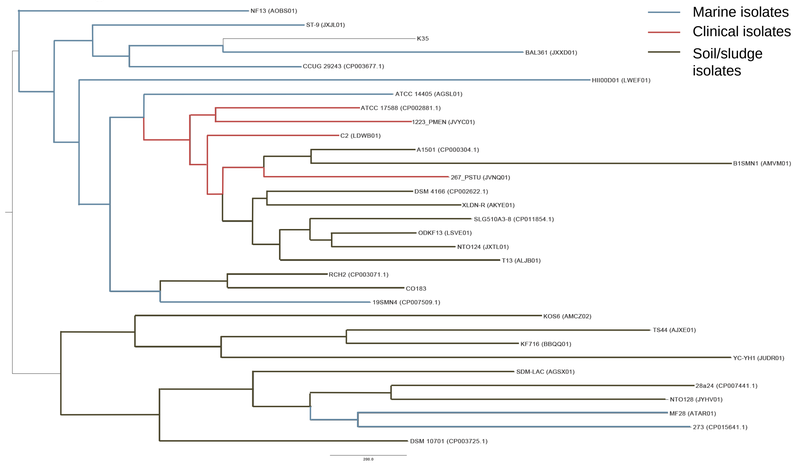
Genome Analysis

Contig mapping to reference genome sequence



Contigs from the *Pseudomonas* genome bin from the K35 metagenome (bottom) were aligned to the complete genome sequence of *Pseudomonas stutzeri* CCUG 29243 (top). Red lines demarcate contig boundary.

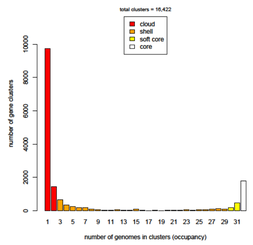
Whole genome comparisons



Pan-genome tree comparing the draft genome of *P. stutzeri* K35 to 31 complete and draft *Pseudomonas* genomes. The tree represents the degree of similarity between the predicted proteins encoded by each genome.

Genome Results. The K35 metagenome was estimated to be ~50% *Pseudomonas.* After careful contig binning and genome mapping, the *Pseudomonas* genome bin was 99.2% complete, estimated by the presence/absence of 833 marker genes. The pan-genome was determined and the core genome (genes common across all strains tested) was estimated. The draft genome encodes for a complete nitrification pathway as well as the upper and lower naphthalene degradation pathways. The work presented here represents an initial metagenomic/genomic approach to functional characterization of coalbed methane microbial communities.

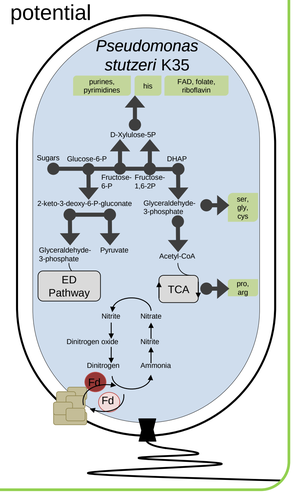
Pan-genome Analysis



Pan-genome of 32 *Pseudomonas stutzeri* strains. Each bar represents the number of gene clusters found in 1 to 32 strains examined.

Future work. Results presented here will provide a framework for tailoring nutrient amendments for microbial enhanced coalbed methane and provide a baseline for monitoring changes in the microbial community during amendments.

Overview of functional potential



REFERENCES:

1. Milici, R.C., and Polyak, D.E., 2014, Coalbed-methane production in the Appalachian basin, chap. G.2 of Ruppert, L.F., and Ryder, R.T., eds., Coal and petroleum resources in the Appalachian basin; Distribution, geologic framework, and geochemical character: U.S. Geological Survey Professional Paper 1708, 25 p., http://dx.doi.org/10.3133/pp1708G.2. (Chapter G.2 supersedes USGS Open-File Report 02–105.)
2. Jones EJP, Voytek MA, Corum MD, and Orem WH (2010). Stimulation of Methane Generation from Nonproductive Coal by Addition of Nutrients or a Microbial Consortium 76(21):7013-7022.

bioGasification

ACKNOWLEDGEMENT: *This technical effort was performed in support of the National Energy Technology Laboratory's ongoing research under the RES contract DE-FE0004000.*