79. Using GraftM to generate phylogenetically informed gene profiles in thawing permafrost metagenomes.

JA Boyd1* (j.boyd3@uq.edu.au), CM Singleton1, BJ Woodcroft1, and GW Tyson1

1 Australian Centre for Ecogenomics (ACE), University of Queensland, Australia; 2University of Arizona, Tucson http://ecogenomic.org/melting-permafrost

**Project Goals:** Microbes play a pivotal role in mediating the release and capture of carbon in the atmosphere and as a result are heavily involved in the regulation of the global carbon cycle. As Arctic and Antarctic frozen soil (permafrost) thaws in response to climate change, the extent of greenhouse gas (GHG) release from these large carbon sinks hinges on the microbial communities that inhabit them. This project focuses on establishing how microbes influence carbon release from these environments by identifying links between community dynamics, established by culture independent shallow metagenomics, and biogeochemical measurements. A model study site in Stordalen Mire, Sweden is used where a natural thaw gradient provides an ideal location for understanding how thaw progression impacts microbial communities. Here, a rapid and accurate community profiling tool (GraftM) is developed and applied to metagenomes extracted from samples taken at Stordalen Mire to recover the community composition of key GHG producing microbes.

Permafrost environments, once a collective carbon sink, are rapidly collapsing into GHG emitting wetlands as a result of climate change induced thaw. The potential for this GHG release to contribute to a positive feedback mechanism with global warming is worrying in light of the models that predict ~1700 Pg of carbon is cryosequestered within permafrost globally (1). The key producers of methane (methanogens), a GHG that is around 25 times more potent than carbon dioxide (2), are archaea that inhabit the expanding water saturated environments resulting from thaw. Previous work (3) has suggested that the abundance and population structure of methanogens is highly influential on the rate of methane release. However much of our current understanding of methanogen communities within these environments is limited to traditional microbiological approaches. These techniques, such as cultures or amplicon sequencing, have limitations such as culture bottlenecking and amplification bias which can negatively impact results. A view of methanogen community structure free from the limitations of traditional techniques would reveal how these populations respond to thaw, and clarify the implications of these responses for net CH4 release.

Here, culture independent shallow metagenomics is used to bypass traditional limitations in >200 environmental samples collected across the thaw gradient over multiple years and several depths at Stordalen Mire, Sweden. This massive dataset provides the unprecedented opportunity to explore and understand microbial communities in both spatial and temporal dimensions, however, rapidly recovering accurate community profiles from large metagenomic and metatranscriptomic data sets remains a non-trivial task. Moreover, the recent improvements in sequencing technology have pushed the yield of reads from sequencing into the billions, thus far outpacing the current bioinformatic tools available to analyze this data.

In this project, the methanogen population structure and abundance is elucidated by developing a novel metagenome/transcriptome analysis tool (“GraftM”), for use on the expansive metagenomic dataset at our disposal. GraftM uses Hidden Markov Models (HMMs), statistical models which use pattern recognition to rapidly identify fragments of key marker genes (such as 16S rRNA) within meta-omic data sets, and constructs community composition based on the placement of these fragments into phylogenetic trees. This approach sets GraftM apart from traditional BLAST based methods of identification and taxonomic assignment where the use of direct sequence-to-database comparisons are both slow and in the case of
complex environmental communities, potentially inaccurate. GraftM allows the exploration of microbial communities to remain unhindered by incomplete databases that often impede high resolution taxonomic assignment, and the size datasets being searched. These improvements make GraftM well suited to the task of analyzing the many complex metagenomes derived from Stordalen Mire in this project.

Here, the utility of GraftM as a profiling tool is demonstrated using in silico datasets where the community abundance was estimated consistently and accurately. Furthermore, GraftM is shown to perform more rapidly than the only published bioinformatic tool that utilizes tree insertion methods to infer phylogeny from marker gene fragments, PhyloSift. Using GraftM, the methanogen community composition was successfully recovered from metagenomes extracted from Stordalen Mire using three separate marker genes, two specific to methanogens (mer and mcrA) and one universal (16S rRNA). Highly significant correlations between the estimated abundance of methanogens as predicted by each of the three marker genes across metagenomes suggest that GraftM consistently assigns taxonomy using different marker gene sets. The population of methanogens was shown to diversify as thaw progresses, with a metabolically diverse community being found in the final stage of thaw hinting at an increase in potential methane production as permafrost thaws. Subsequent linking of key members within the methanogen community with biogeochemical data was found to be coherent with previously conducted amplicon studies on the site (3).

Future directions include the development of a marker gene database for use with GraftM, the integration of checks for copy number, and normalization of gene length. GraftM holds promise as a taxonomic gene profiler capable of rapid identification and classification of conserved marker and key functional genes in meta-omic datasets. This concept could further be expanded into a whole-pathway annotation tool for genome bins.

References

This work is supported by the U.S. Department of Energy under funding opportunity announcement number DE-FOA-0000866. Pathways to carbon liberation: a systems approach to understanding carbon transformations and losses from thawing permafrost.