Title: DNA-SIP enabled community genomics of cellulose degraders in an agricultural soil

Charles Pepe-Ranney[1*] (chuck.peperanney@cornell.edu), Ashley N Campbell,[2] and **Daniel H Buckley[1]**

[1]School of Integrative Plant Science, Cornell University, Ithaca; [2]Lawrence Livermore National Laboratory

Project Goals: Short statement of goals. (Limit to 1000 characters)
Project Goals: This research program will reveal fundamental aspects of soil C-cycling and provide ecological and metabolic insights on diverse non-cultivated soil microorganisms that play major roles in the global C-cycle. Specific goals include: 1) Map the C assimilation dynamics for thousands of non-cultivated microorganisms in soil by harnessing a full cycle microbial food web mapping approach that employs an array of ^{13}C-labeled molecules; 2) Map the C assimilation dynamics of soil microorganisms across soil systems as a function of soil characteristics; and 3) Evaluate ecological and seasonal patterns of activity and abundance for discrete microbial taxa across gradients of soil characteristics and as a function of their C-assimilation dynamics. These goals will be achieved by employing a newly developed microbial food web mapping approach, enabled by advances in ^{13}C-stable isotope probing of nucleic acids and next generation sequencing.

Microorganisms drive biogeochemical cycles and because soil is a large global carbon (C) reservoir, soil microorganisms are important players in the global C-cycle. Frustratingly, however, many soil microorganisms resist cultivation and soil communities are astoundingly complex and hence the dynamics of soil C metabolism remain incompletely described. Stable isotope probing (SIP) is a useful approach for establishing identity-function connections in microbial communities but has been challenging to employ in soil due to the inadequate resolution of microbial community fingerprinting techniques. High resolution DNA sequencing improves the resolving power of SIP transforming it into a powerful tool for studying the soil C cycle. We conducted a DNA-SIP experiment to track flow of cellulose-C, the most abundant global biopolymer, through a soil microbial community.

We found that uncultivated bacterial lineages among *Spartobacteria*, *Chloroflexi*, and *Planctomycetes* assimilated ^{13}C from ^{13}C-cellulose. These lineages are cosmopolitan in soil but little is known of their ecophysiology. In addition to cataloging the SSU rRNA genes of cellulose responsive microorganisms we used DNA-SIP as an approach for targeted community genomics. DNA-SIP enriches DNA of targeted microorganisms. For example, *Verrucomicrobia* cellulose degraders were enriched by nearly two orders of magnitude in the labeled DNA pool, and this "enriched" DNA can serve as template for community genomics. We employed a stripped down binning approach that coupled dimensional reduction of contig oligonucleotide signatures with density based clustering followed by complete linkage clustering of assembly contigs based on paired read links. This approach scales to large data, is easily interpreted and visualized, and

allows for flexible bin definitions such that contigs from strains highly similar in genomic content (that might be overlooked by conventional community genomics analyses) can be clustered into useful groups alongside strain specific genome bins. Using this approach, we produced draft genomes from soil cellulose degraders including microorganisms belonging to *Verrucomicrobia*, *Chloroflexi*, and *Planctomycetes*.

The metagenomic assembly of ^{13}C-enriched DNA yielded approximately 3,100 genes that are associated with the deconstruction of cellulose and 57 pangenome bins contained at least two of these cellulose-associated CAZymes. We also recovered 14 high quality single genome bins (>75% complete and < 10% contaminated as determined by assessment of single copy genes). For example, we recovered a nearly complete novel genome (96% complete) belonging to soil *Chloroflexi* which shared no more than 78% SSU rRNA gene identity to any cultured relative, and we also recovered a draft genome (75% complete) belonging to a lineage of *Verrucomicrobia* that possessed 14 cellulose-associated CAZymes. In addition, we identified a pangenome bin for *Cellvibrio* – a well-studied, model cellulose degrader – which consisted of several polymorphic strains all of which appear to have DNA labeled by 13C-cellulose. This study demonstrates how DNA-SIP can be used to study the ecophysiology of microbes important in terrestrial C cycling and to target guilds of microorganisms for characterization by community genomics.

Funding statement.
This material is based upon work supported by the Department of Energy Office of Science, Office of Biological & Environmental Research Genomic Science Program under Award Numbers DE-SC0004486 and DE-SC0010558.