Effect of metagenome on stable-isotope-enabled metaproteomic analyses of soil microbial communities

Sneha P. Couvillion*, Samuel O. Purvine, Carrie D. Nicora, Anil K. Shukla, Evan P. Starr, Erin E. Nuccio, Kateryna Zhalnina, Eoin L. Brodie, Ulas Karaoz, Jillian Banfield, Jennifer Pett-Ridge, Mary K. Firestone and Mary S. Lipton

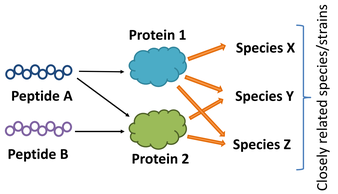
Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Overview

- Metaproteomic analysis of soil using isotopically labelled glucose as a marker for metabolic activity
- Three metagenomic databases were searched using MSGF+[1] and results compared
- Database selection can influence quality and quantity of peptide/protein identifications

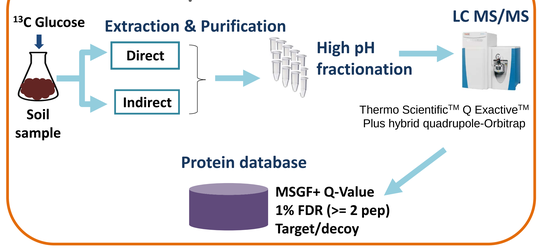
Introduction

- Microbial community composition and function can be important indicators of soil health, plant productivity and response to changing environmental conditions
- Field site: UC Hopland Research and Extension Station in northern California
- Isotopically labelled substrates as markers for metabolic activity is a particularly useful way to distinguish key metabolic players versus bystanders in complex communities
- Challenges of soil metaproteomics include
  - Protein extraction and purification
  - Database choice
  - Redundant protein identification
  - Identifying proteins based on unique peptide can be limiting



Methods

Soil metaproteomics workflow



Metagenomic databases for protein identification

# protein sequences 30,057,884	Comprehensive	→	Function
219,552	Binned	→	Taxonomy
202,248	Isolate		39 dominant bacterial isolates- fully sequenced genomes

Results

Comprehensive metagenome– broad function

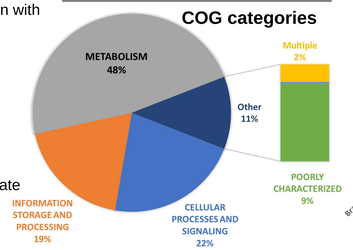
- Comprehensive metagenome represents the entire microbial community as a single entity providing functional information with no taxonomic identifications

Total proteins identified = 66,043
@ Q-Value<=0.0023

Proteins amongst top 20 most abundant include

- Glucose dehydrogenase
- Enolase
- Glyceraldehyde-3-phosphate dehydrogenase/erythrose-4-phosphate dehydrogenase
- ABC-type sugar transport system, periplasmic component

COG categories



Bin Metagenome: Initial taxonomic profiling

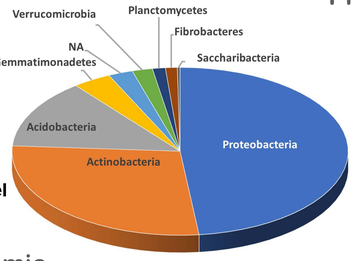
Total proteins identified = 7,118
@ Q-Value<=0.0103

- Binned metagenome contains only those gene models that could be sorted into bins based on tetranucleotide usage
- Representative of community and provides both taxonomic and functional information

Top 3 most abundant

Bradyrhizobium
Geodermatophilus
Streptomyces

Phylum level distribution



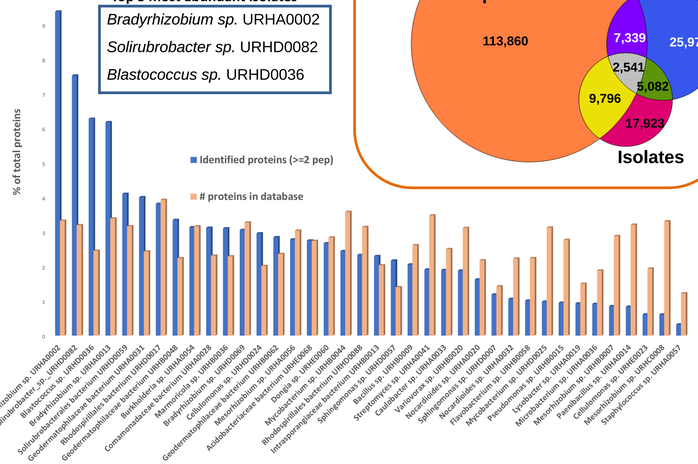
Isolate Metagenome : Precise taxonomic profiling

Total proteins identified = 6,214
@ Q-Value<=0.011

- Metagenome based on concatenated genomes of single isolates from Hopland soil
- Highest level of functional & taxonomic attribution
- Limited to subset of community that can be cultured

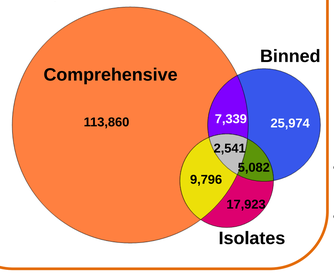
Top 3 most abundant isolates

Bradyrhizobium sp. URHA0002
Solirubrobacter sp. URHD0082
Blastococcus sp. URHD0036



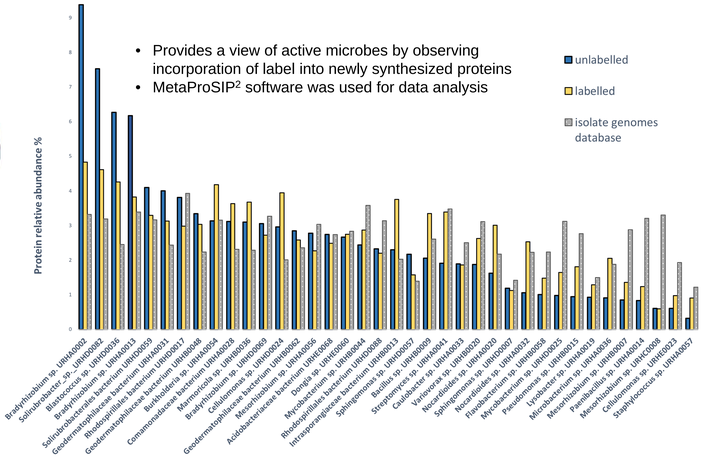
Peptide Overlap Analysis

Total unique=138,081



Protein-SIP: Taxonomic profiling of label incorporation

- Provides a view of active microbes by observing incorporation of label into newly synthesized proteins
- MetaProSIP[2] software was used for data analysis



Conclusions

- The three metagenomic databases provide complementary information and demonstrate how database selection can influence peptide identification
- Protein families that were identified from the glucose-amended samples using this metagenome included central metabolism, sugar transport, and sugar storage proteins - indicate a microbial community with an abundance of nutrients
- Stable isotope incorporation in proteins can be an indicator of microbial activity in soil microbial communities
- **Ongoing work**: Developing software for confident detection and visualization of stable isotope labelled peptides using improved fitting and optimization criteria

References

1. Kim, S. and P. A. Pevzner (2014). "MS-GF+ makes progress towards a universal database search tool for proteomics." *Nature communications* **5**: 5277.
2. Sachsenberg, T., et al. (2014). "MetaProSIP: automated inference of stable isotope incorporation rates in proteins for functional metaproteomics." *Journal of proteome research* **14**(2): 619-627.

Acknowledgements

This project was supported by the U.S. Department of Energy (DOE) Office of Biological and Environmental Research (OBER) Phage and Fauna program lead by Mary K. Firestone and portions of this work were performed in the Environmental Molecular Sciences Laboratory, a DOE OBER national scientific user facility on the PNNL campus. PNNL is a multiprogram national laboratory operated by Battelle for the DOE under contract DE-AC05-76RL01830. Work at LLNL was conducted under the auspices of the US Department of Energy by LLNL under Contract DE-AC52-07NA27344

Career Opportunities: For potential openings in Integrative Omics Group at PNNL please visit http://omics.pnl.gov/careers

CONTACT: Sneha (Polisetti) Couvillion, Ph.D.
Biological Sciences Division
Pacific Northwest National Laboratory
E-mail: sneha.couvillion@pnnl.gov

www.omics.pnl.gov