

## **Integrative Network Modeling for Predicting Biochemical Signatures in Complex Soil Systems**

Hyun-Seob Song,<sup>1\*</sup> (hyunseob.song@pnnl.gov), Joon-Yong Lee,<sup>1</sup> Ryan McClure,<sup>1</sup> **Kirsten S. Hofmockel<sup>2</sup>**, and **Janet K. Jansson<sup>1</sup>**

<sup>1</sup>Biological Sciences Division; <sup>2</sup>Environmental Molecular Sciences Division, Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA

<https://www.pnnl.gov/biology/programs/>

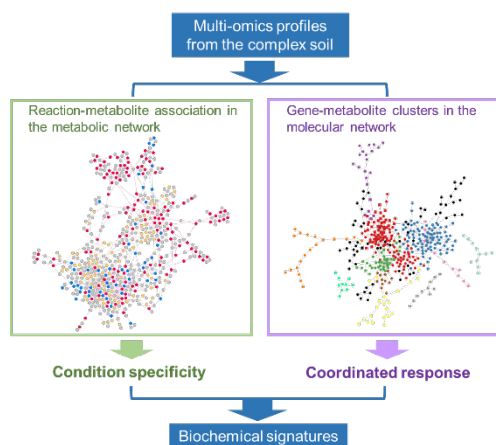
**Project Goals: PNNL's Soil Microbiome SFA aims to achieve a systems-level understanding of the soil microbiome's phenotypic response to changing moisture through spatially explicit examination of the molecular and ecological interactions occurring within and between members of microbial consortia. Integrated experiments were designed to confront both the scaling challenges and inter-kingdom interactions that regulate networks of biochemical reactions. Individual- and population-based models for predicting interspecies and inter-kingdom interactions were parameterized using experimental data, and predictions were tested in soil to reveal spatially explicit microbial interactions. Discoveries from controlled experiments are planned to be cross validated in the field, using moisture gradient experiments at a new local field site. Data was captured and shared through an optimized data management pipeline. Knowledge gained will provide fundamental understanding of how soil microbes interact to decompose organic carbon and enable prediction of how biochemical reaction networks shift in response to changing moisture regimes.**

**Abstract:** Soil microbiomes provide diverse ecosystem services that are vital for life on our planet. A greater understanding of how microbe-regulated soil biochemistry shifts in varying environments is critical for ensuring stable nutrient cycling and healthy ecosystem functioning. In an effort to achieve this goal, scientists have been keen to identify molecular signatures that may characterize soil biochemistry uniquely associated with specific environmental conditions. This requires a systems approach that integrates advanced experimental and modeling technologies<sup>1</sup>, but generalizable protocols have not been fully established yet. Central challenges include the lack of coordinated collection of multi-omics data and the difficulty in building predictive models that can successfully translate molecular-level data to soil metaphenomes<sup>2</sup>.

Leveraging PNNL's multi-omics profiling technology, here we present a general data-model integration approach that enables predicting biochemical signatures of complex soil systems. Our approach combines metabolic and molecular networks that are constructed from multi-omics data. Complementary predictions from these two networks make their integration synergistic: 1) the metabolic network predicts condition-specific reaction pathways and modules;

2) the molecular network identifies molecular clusters that show coordinated response to perturbations.

We demonstrated the proposed approach through a case study of native prairie soils sampled from multiple field locations in Kansas, which were subjected to extreme moisture perturbations. First, we predicted reaction pathways/modules specifically associated with dry and wet conditions by incorporating gene expression and metabolite profiles into a master metabolic network constructed by assembling all chemical reactions available in public databases such as KEGG and ModelSEED. Using the same omics data, we also constructed a molecular interaction network to identify clusters of strongly correlated genes and metabolites. Finally, we combined these outputs to determine molecular signatures, i.e., sets of condition-specific genes/metabolites that show similar response patterns to moisture perturbations (**Figure 1**). Those signatures were predicted particularly for dry soils.



**Figure 1.** Integration of multi-omics data (top), a metabolic network (left), and a molecular network (right) to define biochemical signatures

The general protocol of model building and integration proposed this work should be applicable to many other environmental systems. The metabolic network construction framework is currently being extended to incorporate high-resolution organic carbon profiles generated from high-accuracy mass spectrometry. This unprecedented development can significantly improve our ability to predict condition-specific biochemical footprints of complex environmental systems.

## References

1. Song H-S, Cannon W, Beliaev A, Konopka A. Mathematical modeling of microbial community dynamics: a methodological review. *Processes* **2**, 711-752 (2014).
2. Jansson JK, Hofmockel KS. The soil microbiome - from metagenomics to metaphenomics. *Curr Opin Microbiol* **43**, 162-168 (2018).

**Funding statement:** This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (OBER), as part of BER's Genomic Science Program (GSP), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Soil Microbiome Scientific Focus Area "Phenotypic Response of the Soil Microbiome to Environmental Perturbations." A portion of this work was performed in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by OBER and located at PNNL. PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RLO 1830.