Poster #21-71

Metabolic Network and Multi-omics Integration for Predictive Biogeochemical Modeling

Hyun-Seob Song[1*], Bill Nelson[1], Joon-Yong Lee[1], Chris Henry[2], Janaka Edirisinghe[2], Filipe Liu[2], James Stegen[1], Emily Graham[1], Jianqiu Zheng[1], David Moulton[3], Xingyuan Chen[1], and Tim Scheibe[1]

[1] Pacific Northwest National Laboratory, Richland, WA
[2] Argonne National Laboratory, Argonne, IL
[3] Los Alamos National Laboratory, Los Alamos, NM

Contact: hyunseob.song@pnnl.gov

BER Program: SBR
Project: PNNL SBR SFA
Project Website: https://sbrsfa.pnnl.gov/

This element of the PNNL SFA seeks to improve predictive biogeochemical process modeling through incorporation of the microbial control exerted on biogeochemical dynamics. Current models that take a reductionist approach provide a monolithic description of microbial processes, largely due to the difficulty in characterizing the complex communities. Rapid advancement of high-throughput profiling technologies has generated a large body of multi-omics data for environmental systems, addressing this limitation. To link this molecular data with biogeochemical and ecosystem models across scales, modeling frameworks that can fully take advantage of those unprecedented resources must be developed. Here, we present genome-scale metabolic networks as an ideal tool to fill this gap. While metabolic network modeling has been commonly used for studying single organisms growing in well-defined media, its application to complex environmental microbial systems poses several challenges to overcome as described below. By leveraging the DOE's KBase (https://kbase.us/) modeling and chemoinformatics tools, we propose new pipelines and workflows that enable network reconstruction and data integration for environmental systems to inform biogeochemical models.

First, we developed a new approach to construct metabolic networks from metagenomes. This is challenging in general due to the volume and complexity of metagenome data. As a key idea to address this issue, we reduced data handling by storing only identified activities and metadata such as their counts and estimated taxonomic origin, not all predicted genes. Subsequently, we applied this network reconstruction method to field metagenome data to model microbial communities sampled from two riverbank sediments with and without vegetation. Through a comparative analysis of the two constructed metabolic networks, we were able to identify metabolic pathways 1) that can uniquely characterize biochemical reactions pertaining to each site and 2) that are found in common at two sites. While this result demonstrates the capability of our approach that differentiates biogeochemical traits between two sites, metagenome-based networks could not capture all key differences as indicated by metabolite profiles obtained from FTICR-MS. Thus, we further developed a method to incorporate high- resolution metabolite profiles into metabolic networks. The resulting metabolite-informed networks provide a better representation of functional characteristics of the systems. This new development will significantly be expanded using metabolite data from WHONDRS, a global consortium led by the PNNL SFA to broadly study hydro-biogeochemical function of river corridors. As an important next step, we are undertaking efforts to integrate these new metabolic networks into reactive transport models through collaboration with the IDEAS project.