# Hydrobiogeochemical Features and Function Across Basins

Mikayla Borton,[1,2*] Amy Goldman,[1] Angela Oliverio,[2] Rosalie Chu,[1] Robert Danczak,[1] Rory Flynn,[2] Sarah Fansler,[1] Vanessa Garayburu-Caruso,[1] Sophia McKever,[1] Lupita Renteria,[1] Mike Shaffer,[2] Rebecca Daly,[2] Hyun Song,[3] Josh Torgeson,[1] Jason Toyoda,[1] Kelly Wrighton,[2] and James Stegen[1]

[1]Pacific Northwest National Laboratory, Richland, WA; [2]Colorado State University, Fort Collins, CO; [3] University of Nebraska-Lincoln, Lincoln, NE

**Project Abstract:** *This element of the PNNL River Corridor SFA combines molecular data and numerical modeling to provide transferable principles that integrate organic matter (OM) chemistry and microbial gene expression.* We extend the RCSFA to the globe using crowdsourcing to further establish ESS as a global leader in open watershed science. We are leveraging WHONDRS data while shifting the sampling paradigm to efforts guided more directly by models, via deeper stakeholder engagement, and in partnership with other ESS investments and other agencies (e.g., USDA, NSF, DoD). For example, we are using AI models to guide new sampling locations (see ICON-ModEx presentation) and partnered with EXCHANGE and the University of Quebec to study OM chemistry from source to sea along the St. Lawrence River. We are also expanding to additional continents such as Africa and South America. From each of these sampling efforts, we also generate metagenomic and metatranscriptomic data with JGI. These data are being integrated into the Genome Resolved Open Watersheds database (GROWdb). GROWdb goes beyond the RCSFA and provides an open resource for global river corridor microbiomes. To date, 360 river microbiomes have been sampled across the globe and ongoing sampling aimed at filling in geographical gaps. Microbiomes from all sites were characterized using genome resolved metagenomics, enabling the reconstruction of thousands of unique microbial genomes and a corresponding catalog of over 2 million microbial genes, including sampling of 850 genes with known capacity to modulate carbon, nitrogen, sulfur, and hydrogen cycling in these watersheds. GROWdb has partnered with DOE KBase to make this resource publicly available and to enable integration of river-specific microbial processes into reactive transport models through genome-scale metabolic models, with 2,093 models built to date. Approximately a third of the samples have paired metatranscriptomes and metabolomes, distinguishing active from latent microbial processes and providing additional constraints for modeling efforts.