Topic 1- Exploring global river corridor OM chemistry through the ecological concept of core-satellite species. In a crowdsourced campaign in Summer 2019, the Worldwide Hydrobiogeochemistry Observation Network for Dynamic River Systems (WHONDRS) collected surface water and sediment data from 97 rivers corridors. Here, we present the dissolved organic matter (DOM) composition of a wide range of riverine settings measured at ultra-high resolution (FT-ICR-MS) and discuss how riverine DOM assemblages – fingerprints of substrates and metabolic products – are organized. To develop a transferable understanding of controls of DOM composition across sites and environments, we utilize the concept of core and satellite species, indicating coherences and similarities in substrates, dominant inputs or accumulated products. We discuss the patterns in terms of compound groups, lability, molecular size, and watershed characteristics (e.g. size, stream order, flow, soil characteristics, geology) for both surface water and sediments. We anticipate that this community-led effort will enhance our understanding of how organic matter is structured across global riverine systems and identify underlying factors that influence riverine organic matter composition.

Topic 2- Understanding the sources of organic molecules in river corridors. Understanding the origin of organic molecules across ecosystems remains challenging and requires data integration across various quantitative and qualitative measurements. To identify patterns of organic composition related to its source, we must analyze relatively large data sets (> n= 100s of samples) across vast spatial scales. Here we combine measurements of dissolved organic matter (DOM) composition in surface waters from 97 river corridors worldwide with hydrological and spatial catchment characteristics to answer "What molecular composition signatures are associated with different autochthonous and allochthonous sources of organic matter?," and "Under what ecological conditions are these sources active and/or transported?" With this approach, we intend to identify both co-occurring event-based processes (e.g., biological degradation, photochemical degradation) that structure the molecular composition of DOM across large spatial scales and how that information can be used to predict the fate of organic molecules in an environment.

Topic 3- Biogeochemical transformations of organic matter along the river continuum. Rivers are biogeochemical hotspots, but the environmental and biological factors that drive the transformation of organic matter flowing through rivers are poorly constrained. A synoptic collection of ultrahigh-resolution mass spectrometry (FT-ICR-MS) data and related meta-data across 97 watersheds from the WHONDRS 2019 campaign will allow us to investigate broad, global trends in organic matter (OM) composition and localized phenomena that cause biogeochemical transformations. These samples capture a range of settings, allowing us to examine how shared ("core") and distinct ("satellite") chemical assemblages emerge as a function of watershed properties such as stream order, river discharge, residence time, and dominant land use patterns. We will gain insight into both the geochemical characteristics of organic matter mobilized from particular types of watersheds into rivers and localized processes that transform this material load along the river continuum.

Topic 4- Examining the ecological processes influencing the assembly of molecules into OM assemblages. Dissolved organic matter (DOM) assemblages in freshwaters are formed from complex mixtures of compounds that are highly variable across time and space due to the environmental heterogeneity of river networks. We propose that ecologically significant DOM assemblages can be extracted from a continental-scale dataset of fluvial FTICR-MS organic matter molecular characterization, nutrient levels, stable water isotopes (2H and 18O), and other physicochemical parameters. To identify assemblages, we will apply artificial neural networks and machine learning on DOM samples collected across the continental US. Once DOM assemblages are identified we will use variation across space and time to investigate metabolomic processes in metacommunity ecology and community metabolomes. We will address questions about the origins of DOM assemblages in streams and rivers. Specifically, we will examine the relative importance of biotic versus abiotic processes, watershed characteristics, coupled nutrient cycles, and sediment metabolism on DOM assemblage formation. We argue that this data-driven approach will reveal common continental-scale DOM assemblages and the metabolomic processes that generate these groups of compounds.

Topic 5- A unified conceptual model of organic matter scaling in river corridors. Complex and heterogeneous biogeochemical processes across landscapes introduce a fundamental problem to macro ecological theory—how to scale measured properties to maximize predictive abilities of state-of-the art process models?. Scaling frameworks can include both allometric (*rarity* scales with *intensity*) and environmental (*rarity* scales with the spatial *area*) scaling properties. In attempting to characterize universal macroecological patterns associated with organic matter type and distribution, quantitative and qualitative scaling laws for prediction across biomes are required. Using a continental-scale compilation of dissolved organic matter molecular formulae data generated by the WHONDRS Project, we examine relationships of commonness (as measured by richness and evenness) and rarity (dominance and rarity). We look at how commonness/rarity scales with the intensity of molecules. We hypothesize that these rates are similar to environmental scaling properties found within each watershed and that universal patterns exist across allometric and environmental scaling parameters. Taking advantage of the natural hierarchical scales of aggregation found in watershed river corridors, we hypothesize that the rarity of molecules will be reduced as watersheds expand. We also plan to examine the effects within compound classes (lignin, tannin, etc.) to see if the scaling laws can predict the relative abundance of formulae across chemical groups, mapping roughly onto ecological function across spatial scales. We ask the fundamental question: Are scaling laws universal such that molecular scaling laws show similar behavior and co-variability with spatial scaling laws?

Topic 6- Model-driven identification of OM molecular signatures controlling biogeochemical transformation in river corridors. Organic matter (OM) assemblages in surface water and sediment are determined by complex biogeochemical processes, including those governed by microbes, and thus highly variable in spatiotemporal spaces. This continental-scale, ultrahigh-resolution river metabolomic dataset provides an opportunity to investigate global trends of riverine OM characteristics (e.g., source, mobility, bioavailability, reactivity, parent vs. product compounds) through both data-driven and physics-based modeling. In particular, models applying thermodynamic theory will allow better synthesis and identification of key OM signatures (e.g., thermodynamic favorability, reaction energy efficiency) that control their biogeochemical transformation (e.g., oxidative respiration rate), and can help improve the representation of biogeochemical mechanisms in more integrated models such as reactive transport models. Further, when coupled with existing meta-data of climate, hydrology, geology, and ecology, OM thermodynamics will allow better predictions of biogeochemical and ecosystem dynamics under changing climate. Finally, through data-driven approaches like Sparse Identification of Nonlinear Dynamics (SINDy), we will identify key compound sets that control OM thermodynamics and aerobic respiration rates. The results could be analyzed along with microbiome data to better understand microbe-metabolite interactions provided that microbiome data will be collected from the same locations in future

Topic 7- FTICR-MS data analysis, interpretation, and application. The complex and dynamic nature of organic matter across ecosystems makes it difficult to study, understand, and mechanistically link to emergent biogeochemical function. Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FTICR-MS) offers the potential of acquiring high resolution data for the untargeted characterization of organic matter in complex samples from natural systems, differentiating (by mass) many thousands of low concentration organic molecules. These data are particularly well-suited to ask diverse questions through an ecological lens. Here we provide numerous examples of how different science questions can be evaluated using FTICR-MS data. For each question we provide an example workflow for FTICR-MS data analysis and interpretation, including R scripts applied to a subset of the WHONDRS FTICR-MS global river corridor dataset. In accordance with a crowdsourced approach, the science questions and methods of analysis will be based on global community input and ideation. We anticipate this manuscript being a practical resource and a source of conceptual inspiration for researchers interested in using FTICR-MS to ask science questions within and beyond river corridors.