

MetaboDirect: An analytical pipeline for FT-ICR mass spectrometry data

Christian Ayala-Ortiz,¹ Nathalia Graf-Grachet,¹ Viviana Freire-Zapata,¹ Jane Fudyma,¹ Gina Hildebrand,¹ Roya AminiTabrizi,¹ Cristina Howard-Varona,² Yuri E. Corilo,³ Nancy Hess,³ Melissa Duhaime,⁴ Matthew Sullivan,^{2,5} **Malak Tfaily**,^{1,3*} (tfaily@arizona.edu)

¹ Department of Environmental Science; University of Arizona, Tucson, AZ, 85721, USA

² Department of Microbiology, and Center of Microbiome Science; Ohio State University, Columbus, OH, 43210, USA

³ Environmental Molecular Sciences Laboratory; Pacific Northwest National Laboratory, Richland, WA, 99354, USA

⁴ Department of Ecology and Evolutionary Biology; University of Michigan, Ann Arbor, MI, 48109, USA

<https://github.com/Coayala/MetaboDirect>

Project Goals: Advances in different -omics technologies have revolutionized biological research by enabling high-throughput monitoring of biological processes at the molecular level and their responses to environmental perturbation. Metabolomics is a fast-emerging technology in systems biology that aims to profile small compounds within a biological system that are often end products of complex biochemical cascades. Thus, metabolomics can enable discovery of the genetic basis of metabolic variation by linking the genotype to the phenotype. Despite increasing accessibility of multi-omics technologies, integration of multi-omics data in analysis pipelines remains a challenge especially in the environmental field. In addition, there are still many associated bottlenecks to overcome in metabolomics before measurements will be considered robust. The overarching goal of this proposal is to optimize the analysis of complex and heterogeneous biological and environmental datasets by developing a user-friendly, open-source metabolomics data analysis pipeline that is integrable with other multi-omics data sets.

Microbiomes are now recognized as driving ecosystem functions ranging from the oceans and soils to humans and bioreactors. However, a grand challenge in microbiome science is to characterize and quantify the chemical currencies of organic matter that microbes respond to, produce and alter. Critical to this has been the development of Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS), which has drastically increased molecular characterization of complex organic matter samples, but challenges users with hundreds of millions of data points where readily available, user-friendly, and customizable software tools are lacking. Here we build on years of analytical experience with diverse sample types to develop MetaboDirect, a comprehensive, open-source, command-line based pipeline for direct injection FT-ICR MS data analysis, as well as visualizing the resulting molecular formula assignments. MetaboDirect requires only to run a single line of code that launches a fully automated framework for the generation and visualization of organic matter transformation networks within few minutes. For more experienced users (python savvy), MetaboDirect allows users to customize plots, pipeline outputs, and analyses. Application of MetaboDirect to FT-ICR MS experimental data – from each a marine phage-bacterial infection and a *Sphagnum* leachate microbiome incubation – showcase the data discovery and exploration capabilities of the pipeline

that we hope will enable the FT-ICR MS research community to evaluate and interpret their data in greater depth and in less time.

Funding Statement: This work was funded by the DOE SC program in Biological and Environmental Research (BER) award number DE-SC0021349. A portion of this research was performed under the Facilities Integrating Collaborations for User Science (FICUS) initiative through an award to MT and used resources at the Environmental Molecular Sciences Laboratory, a DOE Office of Science User Facility.