MetaboTandem and MetaboDirect: software pipelines for the analysis of high throughput metabolomics data for complex environmental samples

Christian Ayala-Ortiz[1]; Nathalia Graf-Grachet[2] ; Yuri Corilo[3], Malak M. Tfaily[1]

[1] Department of Environmental Science, University of Arizona, Tucson, Arizona, United States
[2] University of Arizona, Tucson, Arizona, United States
[3] Pacific Northwest National Laboratory, Richland, Washington, United States

INTRODUCTION

Metabolomics allows to characterize hundreds to thousands of small compounds within a biological system. These small compounds are usually the end-product resulting from complex biochemical cascades and can be considered the chemical currency used by microbes in the environment. Metabolomics is perfectly positioned linking phenotype to genotype to enable the discovery of the genetic basis of metabolic variation. Metabolomics studies generate large volumes of data making its analysis and integration a challenge, specially in the environmental field as they often require using and combining multiple data analysis pipelines. Our work aims to develop user-friendly, open-source metabolomics data analysis pipelines that permit optimizing the analysis and integration of complex and heterogeneous biological and environmental datasets.

METABODIRECT (Direct injection Mass Spectrometry)

IMPLEMENTATION

- Developed in Python 3.8 and R 4.2.
- Available through the Python Package Index (PyPI).
- Pipeline has 5 main steps and 2 optional steps.

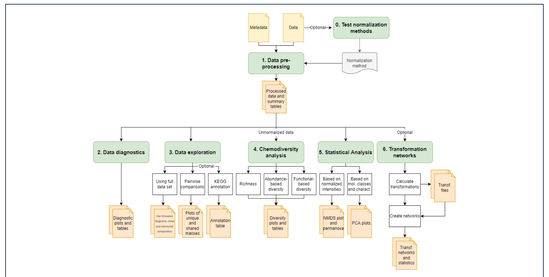

Figure 1. Main steps of the MetaboDirect pipeline

- MetaboDirect main steps can process ~200 samples in less than 3 minutes.

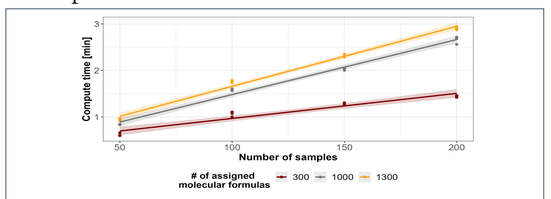

Figure 2. Compute times of MetaboDirect with different mock data sets

MetaboDirect paper is currently under revision, but you can check the preprint at bioRxiv [1].

APPLICATIONS

The influence of soil development on the depth distribution and structure of soil microbial communities

Mary-Cathrine Leewis [a, 1, 2], Corey R. Lawrence [b, 1], Marjorie S. Schulz [a], Malak M. Tfaily [c], Christian Orlando Ayala-Ortiz [c], Gilberto E. Flores [d], Rachel Mackelprang [d], Jack W. McFarland [a]

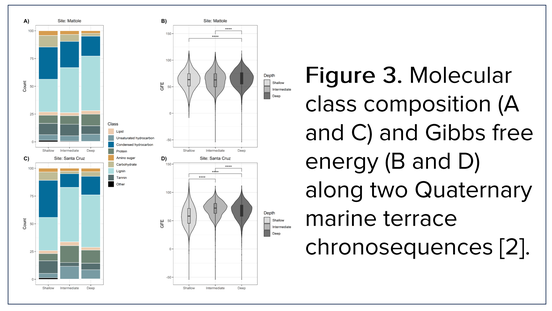

Figure 3. Molecular class composition (A and C) and Gibbs free energy (B and D) along two Quaternary marine terrace chronosequences [2].

Sequential Abiotic-Biotic Processes Drive Organic Carbon Transformation in Peat Bogs

Jane D. Fudyma, Jason G. Toyoda, Rosalie K. Chu, Karl K. Weitz, Heino M. Heyman, Elizabeth Eder, David W. Hoyt, Hans Gieschen, Nathalia Graf Grachet, Rachel M. Wilson, Malak M. Tfaily

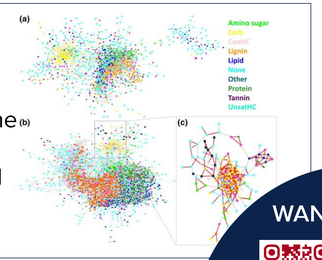

Figure 4. Transformation network analysis of FTICR–MS data in the control (a) and inoculum (b) colored by compound class [3].

METABOTANDEM (Liquid Chromatography Mass Spectrometry)

IMPLEMENTATION

- Developed as both, an R package and a Shiny app.
- Source code available directly from GitHub and as a Docker container.
- Focused on producing a more-in-depth annotation by combining the use of custom annotation databases, with public online databases and *in-silico* molecular structure predictions.

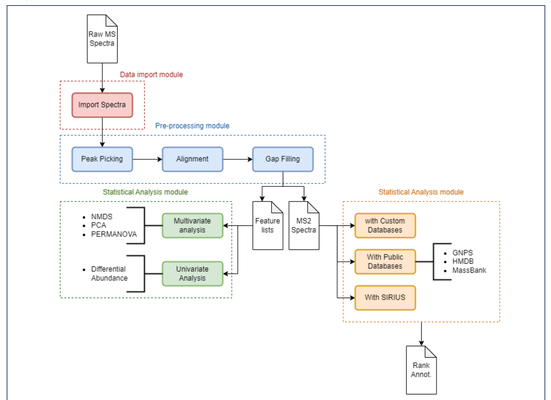

Figure 5. Modules that comprise the MetaboTandem pipeline

TESTING

- The MetaboTandem pipeline is currently under development, however it is currently being used to analyze LC-MS/MS data collected from the Saguaro National Park during the 2021 monsoon season.

If you want to know more about the arid ecosystem project, check the B25B-03 Oral presentation, happening on 12/13 at 2:25 pm

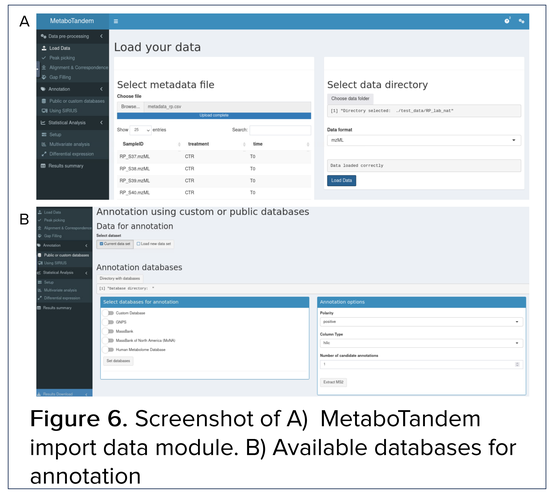

Figure 6. Screenshot of A) MetaboTandem import data module. B) Available databases for annotation

WANT TO TRY?

MetaboDirect

MetaboTandem

REFERENCES

1. Ayala-Ortiz, C. O., Graf-Grachet, N., Freire-Zapata, V., Fudyma, J., Hildebrand, G., AminiTabrizi, R., Howard-Varona, C., Corilo, Y. E., Hess, N., & Duhaime, M. B. (2022). *BioRxiv*
2. Fudyma, J. D., Toyoda, J. G., Chu, R. K., Weitz, K. K., Heyman, H. M., Eder, E., Hoyt, D. W., Gieschen, H., Graf Grachet, N., & Wilson, R. M. (2021). *Journal of Geophysical Research: Biogeosciences*, *126*(2), e2020JG006079.
3. Leewis, M.-C., Lawrence, C. R., Schulz, M. S., Tfaily, M. M., Ayala-Ortiz, C. O., Flores, G. E., Mackelprang, R., & McFarland, J. W. (2022). *Soil Biology and Biochemistry*, *174*, 108808.

ACKNOWLEDGEMENTS

- We thank members of the Tfaily Lab, and EMSL staff for feedback on design and implementation of MetaboDirect and MetaboTandem.
- Thanks to the executive committee of the EMERGE-BII and the Department of Environmental Science at the University of Arizona for providing additional funding to attend this conference.
- **Funding:** Department of Energy, Office of Science Biological and Environmental Research Grant (DE-SC0021349)

GET THIS POSTER