From bulk organic matter profiling to specific metabolite identification: Improving metabolomics data analysis, annotation, interpretation, and integration.

Christian Ayala-Ortiz[1], Sumudu Rajakaruna[1], Jordan Rabus[2], Yuri E. Corilo[2], Dalal Alharthi[3], **Malak M. Tfaily**[1]

[1] Department of Environmental Science, University of Arizona, Tucson, Arizona, United States [2] Environmental Molecular Sciences Laboratory; Pacific Northwest National Laboratory, Richland, WA, 99354, USA [3] College of Applied Science & Technology, University of Arizona, Sierra Vista, AZ, 85635, USA

Overview

Metabolites are the chemical currency of the environment, they are readily used, transformed and exchanged by the microorganisms within a system [1]. Characterizing the metabolome is key to understanding mechanisms driving most ecosystem functions [2]. However, metabolite heterogeneity and complexity make the analysis of metabolomics data a challenging endeavor [3]. Even though metabolomics has long been overshadowed by other 'omics, metabolites are usually at the end of complex biochemical cascades allowing for a better understanding of phenotypic responses [1] by providing a functional snapshot of upstream biological processes that may have been missed by genomics and proteomics. As such, the objective of this work is to develop tools to **optimize**, **streamline**, and **improve** current metabolomics data analysis pipelines.

MetaboDirect

- Easy-to-use, command line-based pipeline for the bulk characterization of organic matter by FTICR-MS.
- Designed to facilitate data exploration, visualization, to perform chemodiversity and statistical analysis, and to build mass difference-based transformation networks [4].

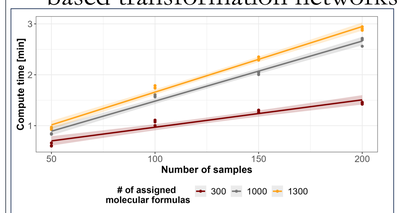
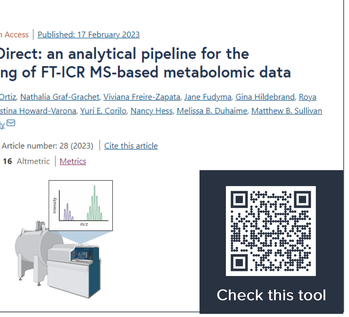


Figure 1. Compute times of MetaboDirect



Check this tool

What's next?

- MetaboDirect will be integrated with CoreMS (developed by EMSL at PNNL).
- It will exist as an optional module that can be run along the existing analysis pipelines provided by CoreMS.

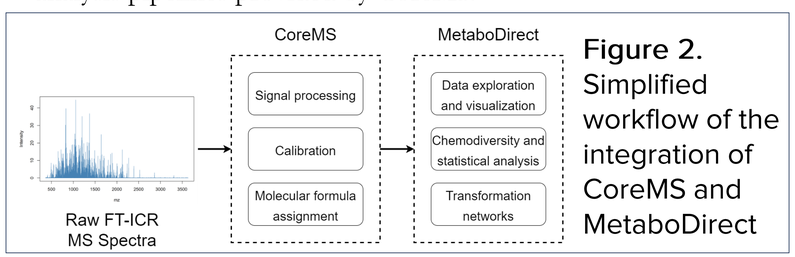


Figure 2. Simplified workflow of the integration of CoreMS and MetaboDirect

MetaboTandem

- Developed as both, an R package and a Shiny app for LC MS/MS.
- 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.

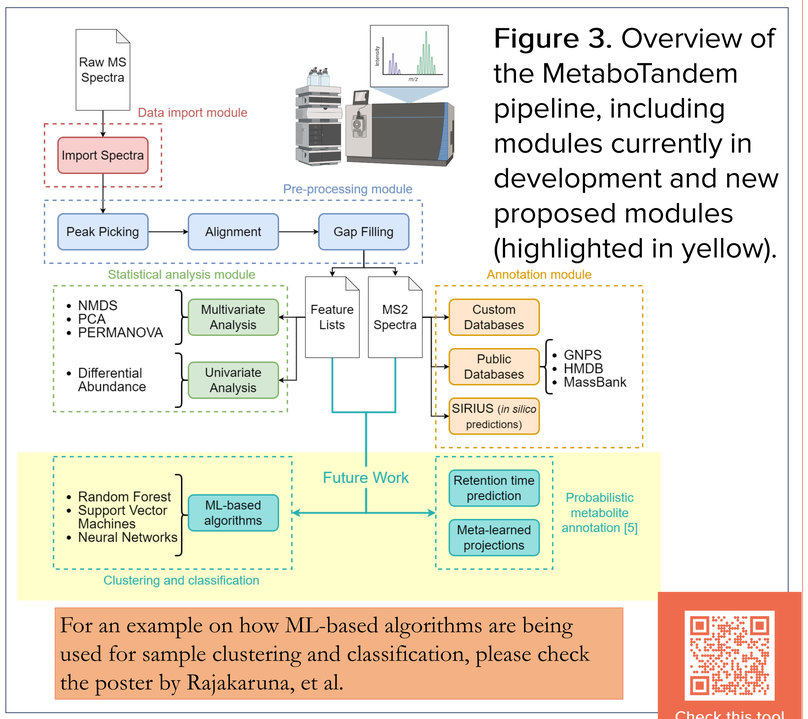


Figure 3. Overview of the MetaboTandem pipeline, including modules currently in development and new proposed modules (highlighted in yellow).

For an example on how ML-based algorithms are being used for sample clustering and classification, please check the poster by Rajakaruna, et al.

Check this tool

Future Work

- Future efforts intend to solve additional challenges in metabolomics: (A) improvement of metabolite annotation, and (B) dealing with larger and more complex data sets by leveraging computer science approaches (i.e., machine learning) to aid in the analysis of metabolomics data.

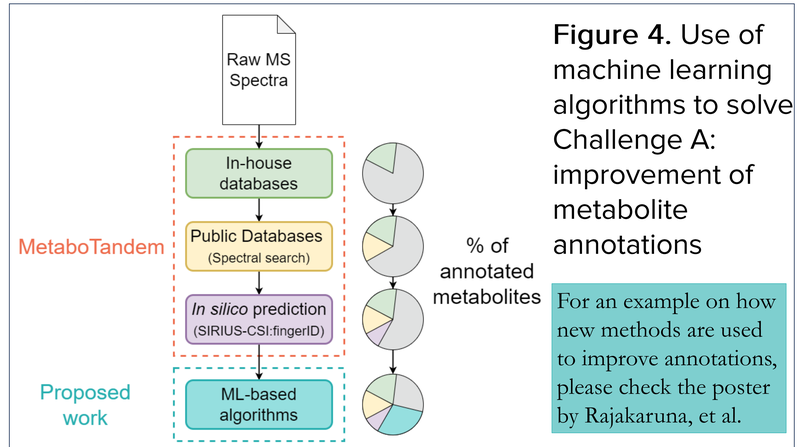


Figure 4. Use of machine learning algorithms to solve Challenge A: improvement of metabolite annotations

For an example on how new methods are used to improve annotations, please check the poster by Rajakaruna, et al.

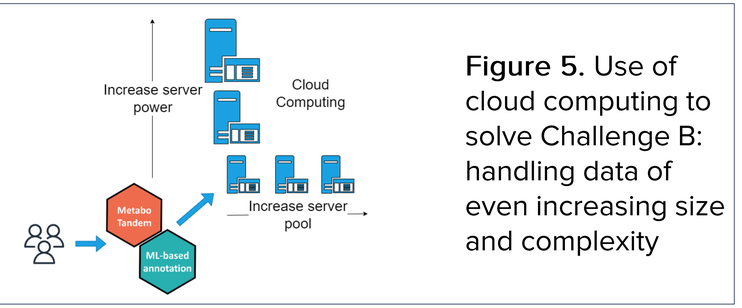


Figure 5. Use of cloud computing to solve Challenge B: handling data of even increasing size and complexity

REFERENCES

1. Fiehn, O., Metabolomics — the link between genotypes and phenotypes, in Functional Genomics, C. Town, Editor. 2002, Springer Netherlands: Dordrecht. p. 155-171.
2. Soule, M.C.K., et al., Environmental metabolomics: Analytical strategies. Marine Chemistry, 2015. 177: p. 374-387.
3. Cooper, W.T., et al., A History of Molecular Level Analysis of Natural Organic Matter by FTICR Mass Spectrometry and The Paradigm Shift in Organic Geochemistry. Mass Spectrometry Reviews, 2022. 41(2): p. 215-239.
4. Ayala-Ortiz, C., et al., MetaboDirect: an analytical pipeline for the processing of FT-ICR MS-based metabolomic data. Microbiome, 2023. 11(1): p. 28.
5. García, C.A., et al., Probabilistic metabolite annotation using retention time prediction and meta-learned projections. Journal of Cheminformatics, 2022. 14(1): p. 33.

Get this poster

ACKNOWLEDGEMENTS

- We thank members of the Tfaily Lab, and EMSL staff for feedback on design and implementation of MetaboDirect and MetaboTandem
- **Funding:** Department of Energy, Office of Science Biological and Environmental Research Grant (DE-SC0021349)