Untargeted metabolomics by high resolution LC-MS/MS revealed different metabolic profiles of oaks (*Quercus* spp.)

Nathalia Graf Grachet,[1*] (nathaliagg@arizona.edu), Megan N. Nickerson,[1] Jana M. U'Ren,[1] and **Malak M. Tfaily**[2]

[1]University of Arizona, Tucson

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.

Abstract

Untargeted tandem mass spectrometry by Fourier transform Ion Cyclotron Resonance preceded by liquid chromatography (LC-FTICR-MS/MS) is a high-throughput and sensitive metabolomics technique that yields high mass accuracy. Such experiments produce information-rich datasets that require the use of many software programs of different programming languages for data analysis. This problem is exacerbated by integrating multi-omics data with metabolomics because of the data heterogeneity. In addition, many of the existing state-of-art software were designed for model organisms, and have limited functionality for complex environmental data and non-model organisms, such as soil and plants. Therefore, our goals in this study were two-fold: i) demonstrate the capabilities of LC-FTICR-MS/MS for metabolic profiling, and ii) to develop a data analysis pipeline for a complex and non-model organism, oak (*Quercus* spp.).

North America is one of the centers of oak genetic diversity, having almost 300 species growing across Mexico, USA and Canada. Continental US has approximately 90 of those species growing from coast to coast (USDA, NRCS, 2021). Despite the incredible diversity of oak species and the social and ecological relevance of oak woodlands, the genetic resources of many species are limited (Plomion et al., 2016, 2018; Sork et al., 2016). Therefore, several aspects about oak species biology, and interactions with the environment still remains to be explored. Metabolomics is a fast-emerging technology in systems biology that remains relatively untapped in environmental and forest sciences. In this study, we used a LC-FTICR-MS/MS technique to perform a metabolic profiling of nine oak species, and to demonstrate its use as a tool for discovery of the genetic basis of phenotypic variation.

Metabolite extracts were prepared by Folch extraction (Folch et al., 1957) from healthy, living leaves. Tandem mass spectrometry was collected on a 21 Tesla (21T) Agilent FTICR-MS

equipped with a Waters ultra-performance liquid chromatography system as previously described in (Fudyma et al., 2019). Negative mode raw chromatography data were preprocessed using the R package XCMS (Benton et al., 2010; Smith et al., 2006). Fragmentation spectra (MS/MS or MS2) were searched against the GNPS High-Throughput Dereplication Comprehensive MS/MS Libraries (Wang et al., 2016). The remaining unannotated spectra were annotated *in silico* using Sirius command-line v.4.5 (Dührkop et al., 2019). Downstream multivariate statistical analyses were conducted with features with MS2 that received at least a molecular formula assignment. Genome-metabolome integration was done using MAGI (Erbilgin et al., 2019). The predicted amino acid sequences of the *Q. robur* reference genome (Plomion et al., 2018) were used in order to unveil potential reactions between the metabolites and genes.

One of the challenges of working with non-model organisms is the lack of useful annotation from publicly available MS2 databases. We only identified 19 compounds in GNPS MS/MS libraries. Using SIRIUS, we were able to assign *in silico* molecular formula and potential structure to 916 compounds, but many of these *in silico* predictions remained unknown because structural matches were not found in public databases. In addition, there is a lack of fragmentation spectra libraries generated by ultrahigh resolution mass spectrometers such as the 21T FTICR.

The main differences observed in this dataset were between deciduous and brevideciduous oak species. PCA was performed on the log-transformed MS2 feature intensities, and the groupings of brevideciduous versus deciduous species drove more the variance within the dataset (~35%). This was in agreement with PERMANOVA, which showed that leaf longevity was a significant factor ($R^2 = 0.28$, $p = 0.004$**). Deciduous oaks showed a higher number of organic acids and benzenoids compared to brevideciduous. Brevideciduous species showed a higher number of lipids and lipid-like molecules compared to deciduous species.

Using thermodynamics, Gibbs free energy (GFE) calculation can be used as an indication of the bioavailability of compounds in which lower GFE values indicate more bioavailable compounds such as carbohydrates and sugars (LaRowe & Van Cappellen, 2011). We observed a significant difference in the GFE values of brevideciduous and deciduous oak species (Kruskal-Wallis test, p = 3.5e-05 ***). Brevideciduous species had a higher median GFE value compared to deciduous, which suggests that brevideciduous have more non-polar compounds represented mainly by lipids and lipid-like molecules that could potentially play a role in plant cell wall formation.

The genome-metabolome integration was done using MAGI, and direct associations between amino acids and compounds were found based on known reactions and pathways found in the BioCyc database (Caspi et al., 2018). We leveraged MAGI for a compound-centric view of biochemical reactions between genes and compounds, and we identified seven compounds directly associated with oak genes. Beta-glucogallin (tannin), astragalin (flavonoid-3-o-glycosides), and quercetin (flavonoid, polyphenol) are phytochemical compounds with medicinal properties with potential uses as anti-inflammatory, antioxidant, and antimicrobial (Li et al., 2016; Puppala et al., 2012; Riaz et al., 2018). Uridine diphosphate (UDP) glucuronic acid and UDP galacturonate are both sugar-like molecules involved in the biosynthesis cascade of plant cell wall polymers such as hemicellulose (Mølhøj et al., 2004;

Reboul et al., 2011). Oxidized glutathione (peptide) and prostaglandin-like molecules (lipid fatty acid) are signaling molecules involved in pathways related to stress responses in plants (Mueller, 1998; Rahantaniaina et al., 2013).

Our study has demonstrated the application of LC-FTICR-MS/MS is metabolic profiling of nine different oak species. The profile of MS features were most distinct between brevideciduous and deciduous species rather than oak type, red vs white. Although many compounds remained unknown, we leveraged different software and strategies to structurally annotate unknown compounds, and were able to identify seven compounds with very important biological roles, including at least three phytochemical compounds well-known for their pharmacological applications.

References

Benton, H. P., Want, E. J., & Ebbels, T. M. D. (2010). Correction of mass calibration gaps in liquid chromatography–mass spectrometry metabolomics data. *Bioinformatics*, *26*(19), 2488–2489. https://doi.org/10.1093/bioinformatics/btq441

Caspi, R., Billington, R., Fulcher, C. A., Keseler, I. M., Kothari, A., Krummenacker, M., Latendresse, M., Midford, P. E., Ong, Q., Ong, W. K., Paley, S., Subhraveti, P., & Karp, P. D. (2018). The MetaCyc database of metabolic pathways and enzymes. *Nucleic Acids Research*, *46*(D1), D633–D639. https://doi.org/10.1093/nar/gkx935

Dührkop, K., Fleischauer, M., Ludwig, M., Aksenov, A. A., Melnik, A. V., Meusel, M., Dorrestein, P. C., Rousu, J., & Böcker, S. (2019). SIRIUS 4: A rapid tool for turning tandem mass spectra into metabolite structure information. *Nature Methods*, *16*(4), 299–302. https://doi.org/10.1038/s41592-019-0344-8

Erbilgin, O., Rübel, O., Louie, K. B., Trinh, M., Raad, M. de, Wildish, T., Udwary, D., Hoover, C., Deutsch, S., Northen, T. R., & Bowen, B. P. (2019). MAGI: A Method for Metabolite Annotation and Gene Integration. *ACS Chemical Biology*, *14*(4), 704–714. https://doi.org/10.1021/acschembio.8b01107

Folch, J., Lees, M., & Sloane Stanly, G. H. (1957). A simple method for the isolation and purification of total lipides from animal tissues. *The Journal of Biological Chemistry*, *226*(1), 497–509.

Fudyma, J. D., Lyon, J., AminiTabrizi, R., Gieschen, H., Chu, R. K., Hoyt, D. W., Kyle, J. E., Toyoda, J., Tolic, N., Heyman, H. M., Hess, N. J., Metz, T. O., & Tfaily, M. M. (2019). Untargeted metabolomic profiling of Sphagnum fallax reveals novel antimicrobial metabolites. *Plant Direct*, *3*(11), e00179. https://doi.org/10.1002/pld3.179

LaRowe, D. E., & Van Cappellen, P. (2011). Degradation of natural organic matter: A thermodynamic analysis. *Geochimica et Cosmochimica Acta*, *75*(8), 2030–2042. https://doi.org/10.1016/j.gca.2011.01.020

Li, Y., Yao, J., Han, C., Yang, J., Chaudhry, M., Wang, S., Liu, H., & Yin, Y. (2016). Quercetin, Inflammation and Immunity. *Nutrients*, *8*(3), 167. https://doi.org/10.3390/nu8030167

Mølhøj, M., Verma, R., & Reiter, W.-D. (2004). The Biosynthesis of d-Galacturonate in Plants. Functional Cloning and Characterization of a Membrane-Anchored UDP-d-Glucuronate 4-Epimerase from Arabidopsis. *Plant Physiology*, *135*(3), 1221–1230. https://doi.org/10.1104/pp.104.043745

Mueller, M. J. (1998). Radically novel prostaglandins in animals and plants: The isoprostanes. *Chemistry & Biology*, *5*(12), R323–R333. https://doi.org/10.1016/S1074-5521(98)90660-3

Pereira-Leal, J. B., Abreu, I. A., Alabaça, C. S., Almeida, M., Almeida, P., Almeida, T., Amorim, M., Araújo, S., Azevedo, H., Badia, A., Batista, D., Bohn, A., Capote, T., Carrasquinho, I., Chaves, I., Coelho, A., Costa, M., Costa, R., Cravador, A., … Ricardo, C. P. (2014). A comprehensive assessment of the transcriptome of cork oak (Quercus suber) through EST sequencing. *BMC Genomics*, *15*(1), 371. https://doi.org/10.1186/1471-2164-15-371

Plomion, C., Aury, J.-M., Amselem, J., Alaeitabar, T., Barbe, V., Belser, C., Bergès, H., Bodénès, C., Boudet, N., Boury, C., Canaguier, A., Couloux, A., Da Silva, C., Duplessis, S., Ehrenmann, F., Estrada-Mairey, B., Fouteau, S., Francillonne, N., Gaspin, C., … Kremer, A. (2016). Decoding the oak genome: Public release of sequence data, assembly, annotation and publication strategies. *Molecular Ecology Resources*, *16*(1), 254–265. https://doi.org/10.1111/1755-0998.12425

Plomion, C., Aury, J.-M., Amselem, J., Leroy, T., Murat, F., Duplessis, S., Faye, S., Francillonne, N., Labadie, K.,

Le Provost, G., Lesur, I., Bartholomé, J., Faivre-Rampant, P., Kohler, A., Leplé, J.-C., Chantret, N., Chen, J., Diévart, A., Alaeitabar, T., … Salse, J. (2018). Oak genome reveals facets of long lifespan. *Nature Plants*, *4*(7), 440–452. https://doi.org/10.1038/s41477-018-0172-3

Puppala, M., Ponder, J., Suryanarayana, P., Reddy, G. B., Petrash, J. M., & LaBarbera, D. V. (2012). The Isolation and Characterization of β-Glucogallin as a Novel Aldose Reductase Inhibitor from Emblica officinalis. *PLoS ONE*, *7*(4), e31399. https://doi.org/10.1371/journal.pone.0031399

Rahantaniaina, M.-S., Tuzet, A., Mhamdi, A., & Noctor, G. (2013). Missing links in understanding redox signaling via thiol/disulfide modulation: How is glutathione oxidized in plants? *Frontiers in Plant Science*, *4*. https://doi.org/10.3389/fpls.2013.00477

Reboul, R., Geserick, C., Pabst, M., Frey, B., Wittmann, D., Lütz-Meindl, U., Léonard, R., & Tenhaken, R. (2011). Down-regulation of UDP-glucuronic Acid Biosynthesis Leads to Swollen Plant Cell Walls and Severe Developmental Defects Associated with Changes in Pectic Polysaccharides*. *Journal of Biological Chemistry*, *286*(46), 39982–39992. https://doi.org/10.1074/jbc.M111.255695

Riaz, A., Rasul, A., Hussain, G., Zahoor, M. K., Jabeen, F., Subhani, Z., Younis, T., Ali, M., Sarfraz, I., & Selamoglu, Z. (2018). Astragalin: A Bioactive Phytochemical with Potential Therapeutic Activities. *Advances in Pharmacological Sciences*, *2018*, 1–15. https://doi.org/10.1155/2018/9794625

Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R., & Siuzdak, G. (2006). XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Analytical Chemistry*, *78*(3), 779–787. https://doi.org/10.1021/ac051437y

Sork, V. L., Fitz-Gibbon, S. T., Puiu, D., Crepeau, M., Gugger, P. F., Sherman, R., Stevens, K., Langley, C. H., Pellegrini, M., & Salzberg, S. L. (2016). First Draft Assembly and Annotation of the Genome of a California Endemic Oak Quercus lobata Née (Fagaceae). *G3 (Bethesda, Md.)*, *6*(11), 3485–3495. PubMed. https://doi.org/10.1534/g3.116.030411

USDA, NRCS. (2021). The PLANTS Database (http://plants.usda.gov, 11 January 2021). *National Plant Data Team, Greensboro, NC 27401-4901 USA*.

Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., Nguyen, D. D., Watrous, J., Kapono, C. A., Luzzatto-Knaan, T., Porto, C., Bouslimani, A., Melnik, A. V., Meehan, M. J., Liu, W.-T., Crüsemann, M., Boudreau, P. D., Esquenazi, E., Sandoval-Calderón, M., … Bandeira, N. (2016). Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature Biotechnology*, *34*(8), 828–837. https://doi.org/10.1038/nbt.3597

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.