Sections:

- Metabolomics Instrumentation
- Metabolite Analyses
- Analysis Tips from Users

Metabolomics Instrumentation

Mass Spectrometry

JGI Metabolomics has 3 **Thermo Orbitrap** mass spectrometers (QE-HF, Exploris 120 and IQ-X) each coupled to an **Agilent 1290 UHPLC** for chromatographic separation. MS and MS/MS fragmentation spectra (as well as UV) are acquired during the same run from each sample extract, with each detected *m/z* having a mass resolution of 60k for MS (accuracy of ~3 ppm, typically observe less than 1 ppm) and 15k for MS/MS. With this information, metabolites are identified by comparison to standards using a “targeted” approach, and with an “untargeted” approach, putative chemical formulas are generated (from accurate mass) for unknowns and fragmentation spectra used for further insight into chemical structure and potential identification.

Cutting-edge mass spectrometry capabilities are in development using the Thermo Orbitrap Tribrid IQ-X, enabling up to 500k mass resolution for resolving molecular formula and fine isotope analysis and MSn (n=10) for deeper fragmentation and insight into molecular structure.

Documentation of methods, workflows and analyses developed and used by JGI Metabolomics is available at protocols.io in the workspace JGI/LBNL Metabolomics Repository.

Sample Preparation

Hamilton Vantage robotic liquid handler for automated sample extraction.

Additional equipment routinely used in sample extraction at the JGI includes a SpeedVac for drying solvent extracts, lyophilizer to freeze-dry aqueous samples as well as a bead-beater and probe sonicator for lysing cells and other biomass.

Metabolite Analyses

At the JGI we analyze metabolite profiles from a wide range of samples including microbes, plants and fungi, as well as the media in which they grow and environment in which they are found, including soil, bodies of water (e.g. ocean, lake), etc. Metabolites are extracted from each type of sample using a variety of methods and solvent systems optimized for specific metabolites of interest. Metabolites from extracts are then detected using our liquid-chromatography tandem mass spectrometry (LC-MS/MS) system, and analyzed using a **"targeted" approach**, in which standards run in-house are used for identification, and/or an **"untargeted" approach**, in which software and algorithms are used to detect features and putatively identify compounds, or both depending on the experiment hypotheses.

Currently, the following types of metabolite analyses can be requested in a CSP proposal, to which a "targeted" and/or "untargeted" analysis method is applied:

1. **Polar Metabolite Analysis** – small, hydrophilic polar metabolites such as amino acids, nucleic acids, sugars and small organic acids, often involved in primary metabolism. Uses normal phase HILICZ chromatography. Limit 200 samples for a typical CSP.
2. **Non-polar Metabolite Analysis** – generally non-polar metabolites not directly involved in primary metabolism, such as antibiotics, polyketides, phenolics. Uses reverse phase C18 chromatography. Limit 500 samples for a typical CSP.
3. **Lipid Metabolite Analysis** – based on detection of lipid, or lipid-like molecules, using a C18 chromatography optimized for lipid detection. Here, metabolites with "lipid-like" chemistry can be analyzed, including common lipid species (triacylglycerols, phospholipids, cholesterols, sterols/steroids, etc), ubiquinones, carotenoids and other pigments. Limit TBD for a typical CSP

The JGI also has capabilities to analyze isotopically-labeled compounds, so each polar, non-polar or lipid analysis can also include **stable isotope probing (SIP)**. Here, the relative amount of incorporated heavy isotope into a compound is measured in experimental samples. These analyses are limited to metabolites detected using a "targeted" approach.

Additional capabilities are always in development, and custom analyses can be considered (LC-MS method, extraction, analysis, etc). Please contact us to discuss available metabolomics opportunities complementary to your research.

Polar Metabolite Analysis

Polar metabolite analysis consists of small, polar metabolites such as amino acids, nucleic acids, sugars and small organic acids that are typically part of primary metabolism, often playing a role in normal growth and development and part of fundamental metabolic pathways essential for survival. Characterizing primary metabolites is important for examining interspecies interactions and cross-feeding, and can be used to determine what substrates are synthesized, taken up or released by different organisms under various environmental conditions.

Extraction is performed on a sample (e.g. media, microbial pellet) that has been lyophilized dry, then 100% methanol added to extract the primary metabolites while precipitating most proteins and salts. Each sample extract is run on the JGI LC-MS/MS system with HILIC chromatography, with acquisition of UV as well as MS and MS/MS fragmentation spectra. (~1 hr/sample)

Non-polar Metabolite Analysis

Non-polar metabolite analysis consists of metabolites that have non-polar or hydrophobic chemical properties, and often encompasses secondary metabolites such as polyketides and phenolics. Although not usually essential, secondary metabolites can provide evolutionary advantages important for survival. As more and more organisms are sequenced, more and more new, unique biosynthetic clusters are revealed each day, with the associated secondary metabolites yet to be discovered and functionally characterized.

Analyzing the non-polar or secondary metabolite profile of organisms under various conditions provides the opportunity to identify new compounds, and to gain insight and create linkages between sequence and function, especially in combination with transcriptomics performed on replicate samples. In synthetic biology, this information can be used to determine relative synthesis levels between constructs, discover new compounds as they relate to organism genomics, as well as identify pathway intermediate, shunt products and precursors to better understand pathway dynamics.

Since secondary metabolites are typically non-polar or hydrophobic in nature, either chloroform, methanol or ethyl acetate is used for extraction of a sample. Each sample extract is run on the JGI LC-MS/MS system with C18 chromatography, with acquisition of UV as well as MS and MS/MS fragmentation spectra. (~35 min/sample)

Lipid Analysis

A diversity of lipids are found across organisms and environments, each playing critical roles in membrane organization, structure, and signaling, responding to environmental

conditions such as desiccation or light, sporulation, as well as being important precursors for production of biofuels or involved in secondary metabolite biosynthesis. Often, species of lipids (e.g. phosphatidylcholine, or PC lipid) are distinguished by a common headgroup attached to fatty acid chains of various carbon length and degree of saturation. Although several hundred forms of each species may exist, fragmentation during mass spectrometry analysis often leads to a characteristic fragmentation ion (e.g. 184 for PC) or neutral loss (e.g. -179 for MGDG) allowing for identification.

Here, we are able to identify common lipid species that have characteristic fragmentation spectra, and are often able to purchase a standard of each species. Some lipid species we have recently annotated include TG, DG, PC, PE, DGTS, SQDG, SM, MGDG, DGDG, PG and respective lyso-lipids.

Extraction is performed on a sample (e.g. algae pellet) that has been lyophilized dry, then a chloroform-based extraction applied (alternatively, 100% can be used with varied results based on lipid solubility). Each sample extract is run on the JGI LC-MS/MS system with lipid-optimized C18 chromatography, with acquisition of UV as well as MS and MS/MS fragmentation spectra. (~1 hr/sample)

Stable isotope labeling

Stable isotope labeling analysis consists of profiling the relative amount of incorporation of a heavy isotope (e.g. 13C, 2H, 15N) into synthesized metabolites. Mass spectrometry is an ideal analysis tool for stable isotope labeling, since the isotope of each compound is detected in a typical mass spectrum. Here, by comparing ion intensities of isotopes for a metabolite in unlabeled vs. labeled (treated with 13C or deuterium source, for instance) samples, active metabolic pathways can be traced through an organism, fate of a carbon source can be identified, and newly synthesized compounds can be identified.

The heavy isotope can be introduced, for example, as a labeled metabolite serving as a carbon source (e.g. 13C-acetate) to a microbial culture, and then using mass spectrometry to track to what degree (and potentially where) 13C is incorporated into metabolites in various metabolic pathways. For these types of experiments, unlabeled control samples must also be prepared for analysis.

This type of analysis can be performed on any ion feature (unique m/z, RT pair) detected by mass spectrometry. Typical analysis is performed on "**targeted"** metabolites, although a more limited analysis is available for general ion features.

"Targeted" Approach

To identify metabolites using a **"targeted" approach**, we have already run a large library of metabolite standards (currently >1000 and growing) to which we are able to compare

retention time, m/z (detected mass/charge ratio) and fragmentation spectra to definitively identify metabolites in a sample. This approach can be used for any metabolite in which a standard is available, including secondary metabolites (e.g. violacein, prodigiosin).

"Untargeted" Approach

Often, a metabolite standard is not available and no database contains MS/MS fragmentation spectra, especially in the case of many secondary metabolites, or when little is known about the sample submitted for analysis and many unknown or novel compounds are expected. Here, metabolites are putatively identified using an **"untargeted approach."**

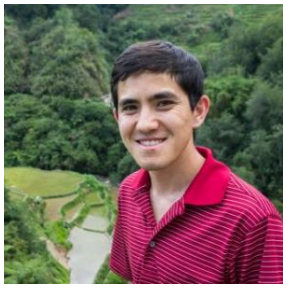
Our standard "untargeted" workflow is currently based on feature-finding using open-source software mzmine3, with putative identifications made using the web-based GNPS2 platform that provides interactive interfaces for further data analysis, molecular networking, and data sharing.

Another available tool is **Pactolus**, a new version of **MIDAS (Metabolite Identification via Database Searching)**, another solution for "untargeted" analysis developed by LBL researchers. This is essentially an open-source software tool that is able to predict MS/MS fragments from any chemical structure, real or theoretical, available in a database. A metabolite in a sample can then be putatively identified, with a percent probability, by comparing predicted fragments to actual measured fragments from a sample. In the absence of MS/MS fragmentation data from a standard, this is a powerful tool for metabolite identification and dereplication for discovery. Our current database has over 180,000 compounds for untargeted analysis. Additional software tools are in development for identification of unknowns and linkage to genomics.

Metabolomics Data Analysis – Tips From Users

Many of the standard procedures for processing ‘omics data sets for gene expression, protein abundance, ribosomal similarity, etc can be applied to metabolomics data as well. However, metabolites are unique in that they are the products of metabolism; where the other techniques lay the foundation for metabolism to occur. Example analysis approaches by JGI-metabolomics user’s are described below. These examples are not meant to provide in depth teaching, but a starting point for how one might approach their own analysis.

Daniel Caddell



Daniel Caddell is a Research Biologist at US Department of Agriculture (USDA) Agricultural Research Service (ARS). A useful first step in analyzing metabolomics data is to assess global trends in the data, beginning with assessing the robustness of sample replicates. For this, a scatterplot (log scale) can quickly be generated (in a spreadsheet program such as Microsoft Excel, or a programming language such as R) to compare trends in ion abundances between sample replicates. If the quality of the samples is high, very few significantly different ion abundances should be observed between replicates (Figure DC1A). In addition to determining the robustness of sample replicates, this method can be applied to probing relative peak heights of individual metabolites for outliers, whose ion abundances differ between sample type, location, or treatment, as seen in Figure DC1B-C. However, if the quantification of individual metabolites has not been performed, these relative ion abundances are not suitable for absolute metabolite level quantification (e.g. micrograms per gram of sample) or comparisons between different metabolites, due to differences in ionization efficiencies and the influence of the biological matrix.

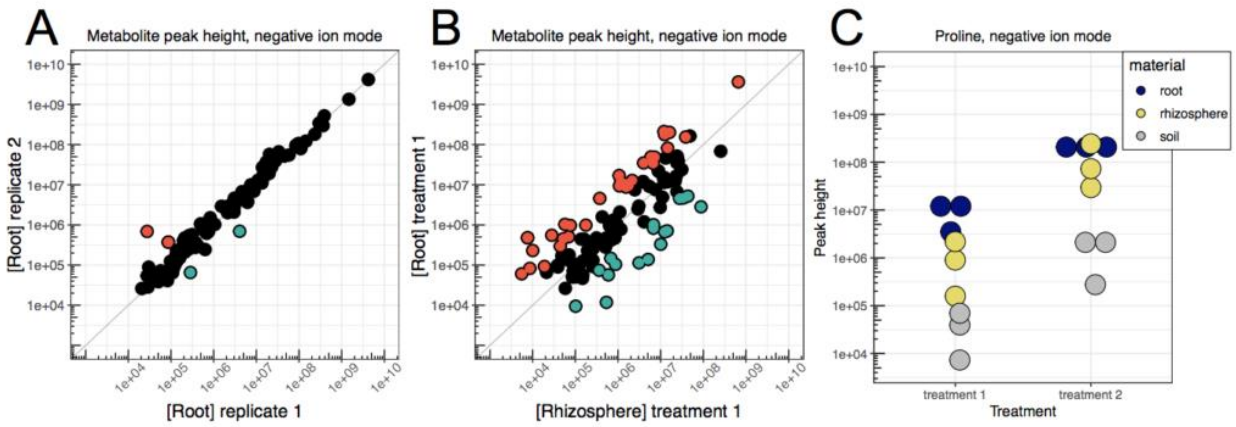


Figure 1. Comparison of ion abundances between (A) replicates and (B) sample types. Each dot represents an individual metabolite present in the dataset, with red or blue filled dots indicating the metabolites that were more abundant in one dataset or the other (fold change > 2). (C) Single metabolites can also be analyzed before or after normalization.

Notably, normalization of the data to account for background signals present in extraction blanks can be accomplished by two different methods. First, the background signal present in extraction blanks can simply be subtracted from the corresponding ion abundance in the experimental samples. Alternatively, a value representing the lower detection limit in the dataset (e.g. ~4,000 in Figure DC1) can replace any empty data points, either in extraction blanks or experimental samples, prior to normalization. The rationale for this substitution is that metabolites absent from a sample cannot be distinguished from metabolites present below the detection threshold. After normalization, metabolite peak height can be converted to percent relative abundance by setting the maximum peak height observed across all samples to 100%. While searching for the metabolites whose ion abundances have large fold changes is a useful heuristic for analysing metabolomic data, it can be beneficial to further subset metabolites by a combination of heuristic thresholds including significance (ie: $P<0.05$), fold change (ie: 2 or more), and minimum intensity (ie: 10x the background).

Ryan Lenz

Pathway analysis with MetaboAnalyst (Ryan Lenz). MetaboAnalyst is a useful online interface that allows a researcher to conduct many different types of analysis (Xia et al. 2015). This program is written in the R coding language allowing advanced users to change statistical and imaging parameters if desired. Generally, the online interface is sufficient for most analysis. To start, it is best to normalize the data before doing comparative statistics such as t-tests and fold change. MetaboAnalyst also has many choices for unsupervised and supervised modeling of metabolomic data and significant feature selection.

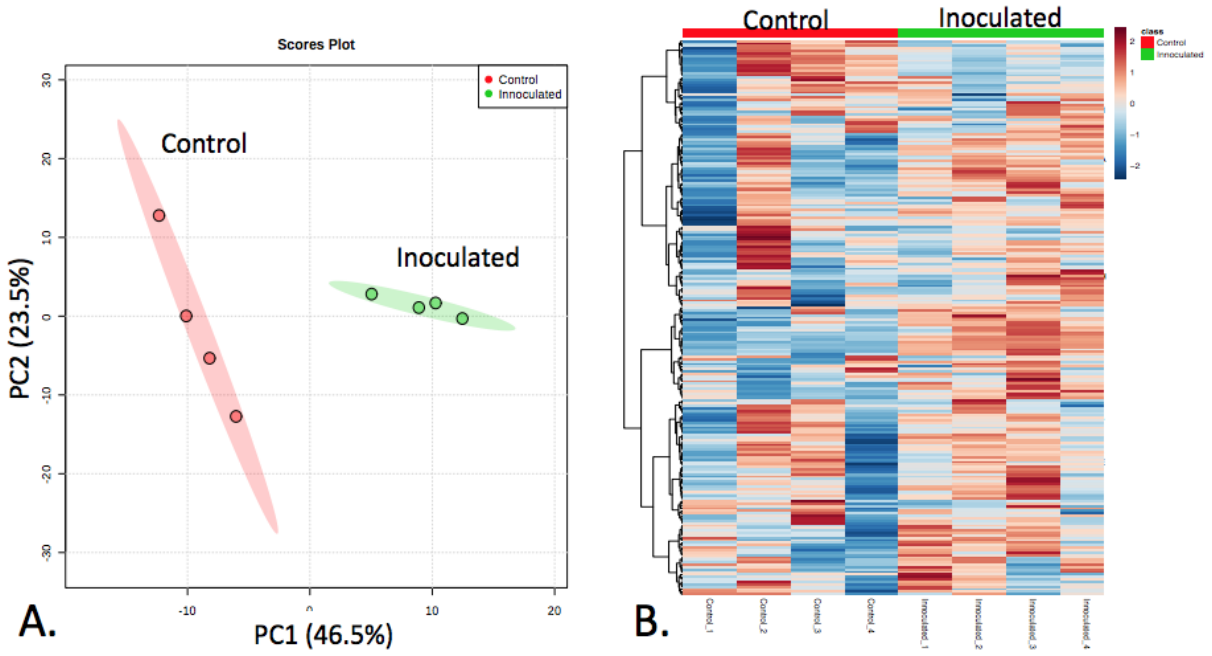


Figure 2. Overall data differentiation between mock-inoculated and inoculated stem tissue. (A) Principal component analysis (PCA) and (B) and Heatmap visualization of all (~250) metabolic features.

Figure 2 shows a principal component analysis and a heatmap to summarize the data. From here you can organize the fold-change table of all the metabolites and run it through enrichment and pathway analysis. This allows you to get a feel for the metabolomic

reactions most represented by the data. Once your data is uploaded, you can choose a pathway library from an assortment of model species including mammals, plants, and microbes. Figure 3 is an example of how MetaboAnalyst can organize the most impacted metabolic pathways from your data.

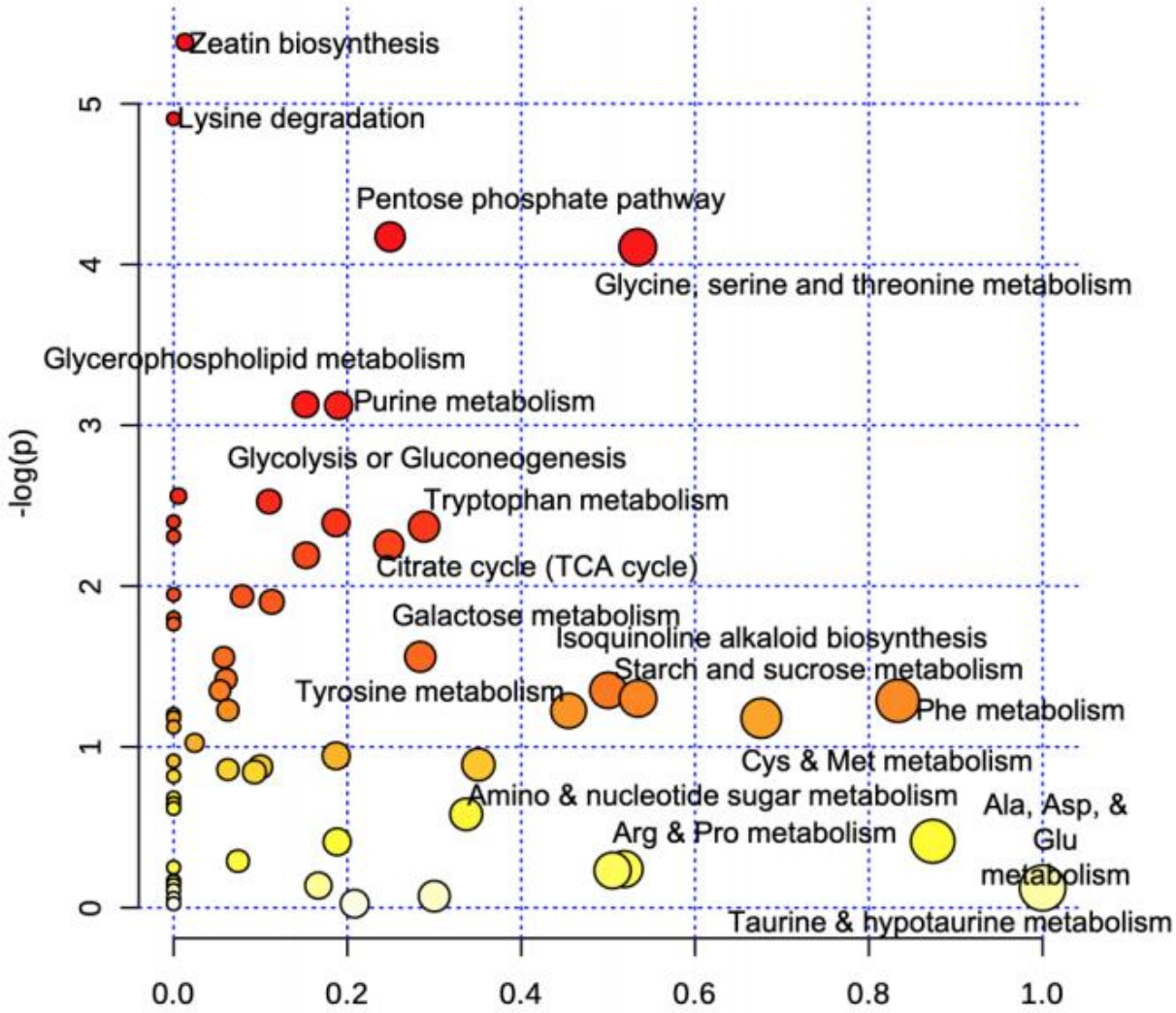


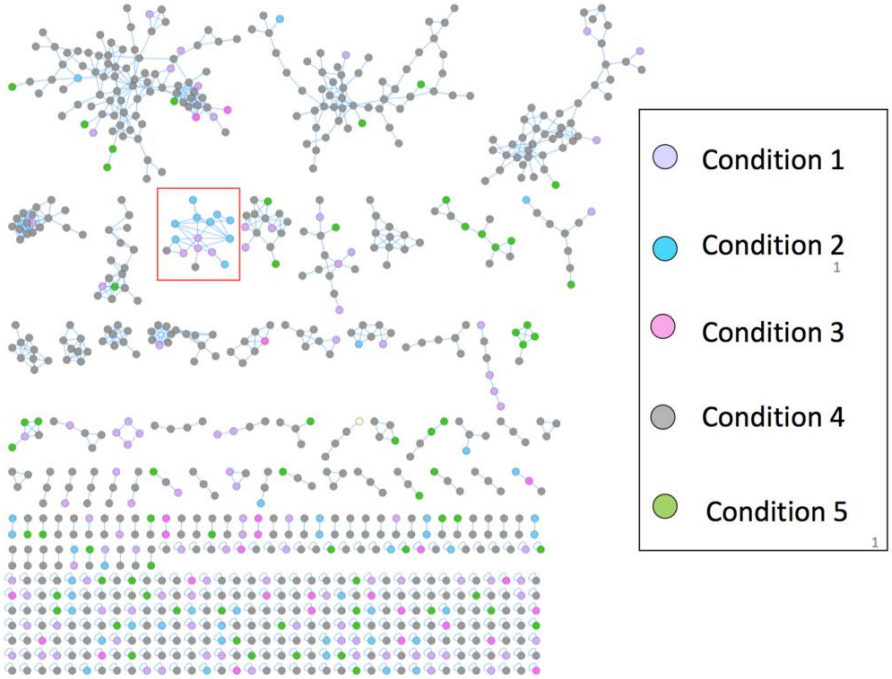
Figure 3. Metabolic pathways altered by inoculated stems organized by pathway enrichment analysis (p-values) and pathway topology analysis (pathway impact).

MAGI (https://magi.nersc.gov) is another tool that can add a layer of biological relevance to metabolomic data. Generally, MAGI allow users to screen an organism's genome for biochemical pathways involving a list of metabolites identified from metabolomics studies. In this way, users can confirm that significant features are produced/sourced from their treatments. It can also help decipher the origin of identified metabolites in treatments involving more than one organism. For example, a significant metabolite from an

experiment involving both a plant and a fungal pathogen was originally listed as putatively identified via LC-MS. This metabolite was screened with MAGI for both the plant and the fungus. The metabolite received a very low MAGI score for both organisms which indicates that it most likely is not produced in that context and is likely mis-identified. As a result, a user can reevaluate the m/z and retention times and select a biologically relevant metabolite for further analysis.

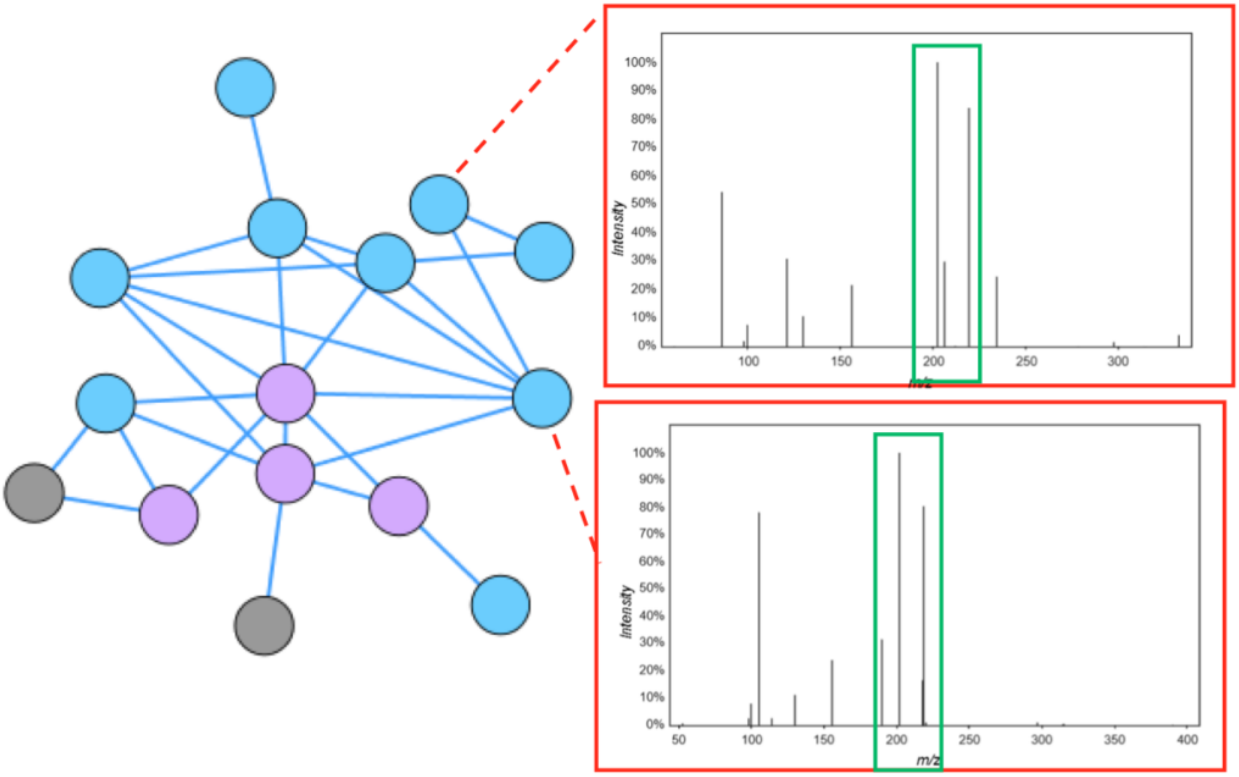
Candice Swift

Molecular Networking (Candice Swift, graduate student in the O'Malley lab at UC Santa Barbara.). Global Natural Products Social Molecular Networking (Ming et al. Nature Biotechnology 2016) GNPS) [Ming et al. Nature Biotechnology 2016] is a powerful technique for visualizing metabolomics datasets. In a molecular network, each node represents an MS/MS spectra for a particular m/z, retention time pair. Spectra are compared and given a cosine score between zero and one: a score of zero represents spectra without any similarity and a score of one represents a complete match. Similar nodes are connected by edges (the default threshold is 0.7), resulting in a network of clustered spectra. Mass differences between nodes can be used to gain structural insights into functional groups that may be present in the parent ions (for an example, see Fig. 2B of Watrous et al.PNAS 2012).



The GNPS data analysis pipeline used to create molecular networks has several useful features: 1) users can match unknown spectra to the GNPS compilation of spectral libraries, 2) it includes a built-in network visualization browser that allows visualization of clusters and

comparison of experimental spectra to the spectra of known compounds in the libraries, and 3) comparison of up to six different experimental conditions. This list is far from comprehensive, with more improvements and features constantly being added. Users are encouraged to explore GNPS for themselves. For more stringent library matching, be sure to adjust the mass difference tolerance, called Maximum Analog Search Mass Difference (default is 100 ppm).



Getting started is fairly straightforward, with video tutorials, in-depth documentation, and even regular office hours (see the main website here). Parameters to consider adjusting when creating a network include the following (default given in parenthesis): Min Pairs Cos (0.7), Maximum Connected Component Size (100), Minimum Cluster Size (2), and Maximum Analog Search Mass Difference (100.0)

When using GNPS, please cite Wang, Mingxun, et al. “Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking.” Nature Biotechnology 34.8 (2016): 828-837. PMID: 27504778

Marc Chevrette

Phylogeny and Metabolic similarity (Marc Chevrette). Metabolism is a complex trait shaped by ecological and evolutionary forces. As such, organismal metabolism can be explored in a phylogenetic framework to help explain underlying environmental (e.g. nutrient acquisition, flux) and species-species (e.g. host-microbe metabolic exchange, secondary metabolism) interactions. Gene-metabolite relationships (see MAGI section above) in the context of phylogenies offer insight into the evolutionary histories of pathways and allow for comparisons across gene topologies, population structure, and ecology.

Chevrette MG, Carlos-Shanley C, Louie KB, Bowen BP, Northen TR and Currie CR (2019) Taxonomic and Metabolic Incongruence in the Ancient Genus *Streptomyces*. *Front. Microbiol*. 10:2170. doi: 10.3389/fmicb.2019.02170