UC Santa Cruz
UC Santa Cruz Previously Published Works

Title
Staring into the void: demystifying microbial metabolomics

Permalink
https://escholarship.org/uc/item/13t1m6z8

Journal
FEMS Microbiology Letters, 366(11)

ISSN
0378-1097

Authors
Grim, Cynthia M
Luu, Gordon T
Sanchez, Laura M

Publication Date
2019-06-01

DOI
10.1093/femsle/fnz135

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

OXFORD FEMS

FEMS Microbiology Letters, 366, 2019, fnz135

https://doi.org/10.1093/femsle/fnz135
Advance Access Publication Date: 18 June 2019
Minireview

MINIREVIEW – Incubator - Big Data Approaches for Industrial Microorganisms

Staring into the void: demystifying microbial metabolomics

Cynthia M. Grim[†,‡], Gordon T. Luu[†] and Laura M. Sanchez[*,§]

Department of Pharmaceutical Sciences, University of Ilinois at Chicago, 833 S Wood St, Chicago, IL 60612, USA

[*]**Corresponding author:** Department of Pharmaceutical Sciences, University of Ilinois at Chicago, 833 S Wood St, Chicago, IL 60612, USA. Tel: +312-996-0842; E-mail: sanchelm@uic.edu

One sentence summary: A short review on mass spectrometry based metabolomic approaches for microbes.

[†]Contributed equally, Listed in alphabetical order.

Editor: Marnix Medema

[‡]Cynthia M Grim, http://orcid.org/0000-0001-9638-6929
[§]Laura M Sanchez, http://orcid.org/0000-0001-9223-7977

ABSTRACT

Metabolites give us a window into the chemistry of microbes and are split into two subclasses: primary and secondary. Primary metabolites are required for life whereas secondary metabolites have historically been classified as those appearing after exponential growth and are not necessarily needed for survival. Many microbial species are estimated to produce hundreds of metabolites and can be affected by differing nutrients. Using various analytical techniques, metabolites can be directly detected in order to elucidate their biological significance. Currently, a single experiment can produce anywhere from megabytes to terabytes of data. This big data has motivated scientists to develop informatics tools to help target specific metabolites or sets of metabolites. Broadly, it is imperative to identify clear biological questions before embarking on a study of metabolites (metabolomics). For instance, studying the effect of a transposon insertion on phenazine biosynthesis in Pseudomonas is a very different from asking what molecules are present in a specific banana-derived strain of Pseudomonas. This review is meant to serve as a primer for a 'choose your own adventure' approach for microbiologists with limited mass spectrometry expertise, with a strong focus on liquid chromatography mass spectrometry based workflows developed or optimized within the past five years.

Keywords: mass spectrometry; metabolomics; microbiology

INTRODUCTION

Metabolomics

The metabolome comprises small molecules (metabolites) typically under 2000 Daltons (Da). These metabolites are typically involved in one or more metabolic pathways in any given organism and often underlie biological activity such as antimicrobial activity (Oliver et al. 1998; Raamsdonk et al. 2001). Metabolites can be further categorized into primary and secondary metabolites; primary metabolites have historically been defined as those that are directly involved in sustaining life, whereas secondary metabolites are not necessarily needed for growth and typically appear in cultures after the exponential phase of growth. The ecological role of secondary metabolites remains ambiguous in the majority of cases. Microorganisms can produce a wide array of metabolites with each having a number of possible roles for life, defense, and communication among others. Given the relative ease and affordability of genomic studies, there is a need to bridge the gap between the genome/transcriptome/proteome with phenotypic traits observed in microorganisms. We believe that metabolomics has gained significant interest in recent years due to its ability to help bridge the genotype–phenotype–chemotype gaps.

Received: 30 April 2019; **Accepted:** 14 June 2019

© The Author(s) 2019. Published by Oxford University Press on behalf of FEMS. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

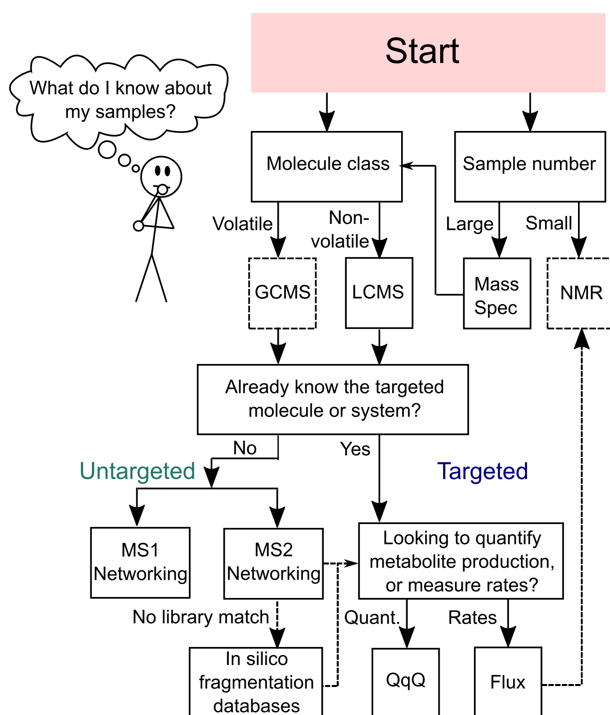

Figure 1. *Metabolomics Workflow Chart.* Start at the top box and ask yourself, what is the major molecule class I am interested in? How many samples do I have? Dashed boxes indicate sections not fully described in this text, and dashed arrows indicate optional steps. Those who are interested in the topics not covered such as GC-MS, NMR, or other mass spectrometry workflows should consult reviews by Beale et al., Halouska et al./Nagana Gowda & Raftery, and Luzzatto-Knaan et al., respectively (Halouska et al. 2013; Luzzatto-Knaan, Melnik and Dorrestein 2015; Nagana Gowda and Raftery 2017; Beale et al. 2018).

Analytical techniques

The use of analytical instrumentation is essential in the field of metabolomics. Nuclear magnetic resonance (NMR) and mass spectrometry (MS) can both be utilized in metabolomics, with MS typically being best suited for large sample sizes. Mass spectrometry based techniques detect ions, which can be either positively or negatively charged molecules that are denoted as mass-to-charge ratios (m/z) each with a corresponding intensity value. Two commonly used analytical techniques include gas chromatography mass spectrometry (GC-MS) and liquid chromatography mass spectrometry (LC-MS) (Fig. 1).

GC-MS is best suited for volatile metabolites or those that can be derivatized as such. Ions in GC-MS are created via electron ionization (EI), a form of hard ionization that produces molecular ions $[M]^{\cdot+}$. During ionization, metabolites fragment and are detected by the mass analyzer. This ability to fragment ions combined with high resolution, sensitivity, reproducibility, and large compound libraries makes GC-MS the preferred instrument for analysis of volatile metabolites under 1000 Da (Rowan 2011). However, the current limitations include the inability to analyze metabolites, especially unknowns, that don't fit those criteria, such as larger, non-volatile metabolites that are unable to be derivatized into volatile compounds. GC-MS is a robust and well understood form of mass spectrometry, but libraries and tools based on LC-MS are rapidly evolving to better understand the non-volatile chemical space.

LC-MS has become the standard technique for analysis of water-soluble metabolites and is one of the most utilized techniques according to a recent survey (Weber et al. 2017). Electrospray ionization (ESI) is the predominant form of ionization for LC-MS, and is classified as 'soft' ionization meaning the molecule is frequently identified as a protonated or deprotonated molecule, termed $[M + H]^+$ and $[M - H]^-$, although in-source fragmentation, multiply charged species ($[M + 2H]^{+2}$), and production of other adducts ($[M + Na]^+$, $[M + K]^+$, etc.) can occur. The biggest consideration for LC-based methods is the fundamental principle of 'like dissolves like' such that the solvent miscibility and solvents used can heavily dictate the type of molecules that are being ionized and analyzed. High-resolution mass spectrometry (HRMS), hyphenated by either chromatographic method, does not provide structural information but can provide accurate mass measurements allowing one to calculate a putative molecular formula. Resolution is the measure of separation of two mass spectral peaks, not to be confused with resolving power which is the ability of a mass spectrometer to separate ions of two different m/z values; these terms are often not used correctly (David Sparkman 2000). But when an instrument has high mass resolving capabilities (>20 K), it has the ability to resolve ions that have similar masses, such as the isobaric species choline (calculated exact mass of protonated molecule, 194.1070) and GABA (calculated exact mass of protonated molecule, 194.0706). Fragmentation via tandem mass spectrometry (MS/MS or MS^2) is used to glean information about an ion's structure. In MS/MS, precursor ions generated from first stage MS (MS^1) are selected to undergo collision-induced dissociation (CID) to fragment them into smaller fragment ions or product ions. It should be noted that fragmentation from $[M - H]^-$ molecules can be more difficult than from $[M + H]^+$ molecules. Being able to obtain fragmentation data in LC-MS/MS provides another layer of data to aid in the analysis of molecules of interest. For a thorough discussion of different types of mass analyzers including low resolving power (1000) and high resolving power (10K–1 Million) options, we recommend the reader to Henke and Kelleher's review of the topic which includes a discussion of how the mass analyzers and their mass resolving power can affect mass accuracy and ppm error in the measurement (Henke and Kelleher 2016).

Targeted versus untargeted metabolomics

Metabolomics experiments can be categorized into two approaches: targeted or untargeted (Fig. 2). Fig. 2b describes the principles behind targeted and untargeted metabolomics workflows. Targeted metabolomics is often hypothesis-driven, where specific metabolite(s) of interest are measured, quantified, and compared to known standards in an attempt to identify how corresponding metabolic pathways are modified in response to different stimuli. Phelan et al. previously examined the effects of a transposon insertion into the phenazine biosynthetic gene *phzF2*, and conducted fold change analyses of the known secondary metabolites. Interestingly, this insertion resulted in specific 48-fold increase in the siderophore pyochelin over WT, whereas the siderophore pyoverdine was not detected after insertion (Phelan et al. 2014). Another recent pilot study demonstrated the use of targeted metabolomics by measuring 221 metabolites to determine the changes in metabolism in methicillin-resistant *Staphylococcus aureus* (MRSA) and methicillin-susceptible *S. aureus* (MSSA) when treated with varying concentrations of glucose and sublethal doses of the antibiotic methicillin (Rutowski et al. 2019).

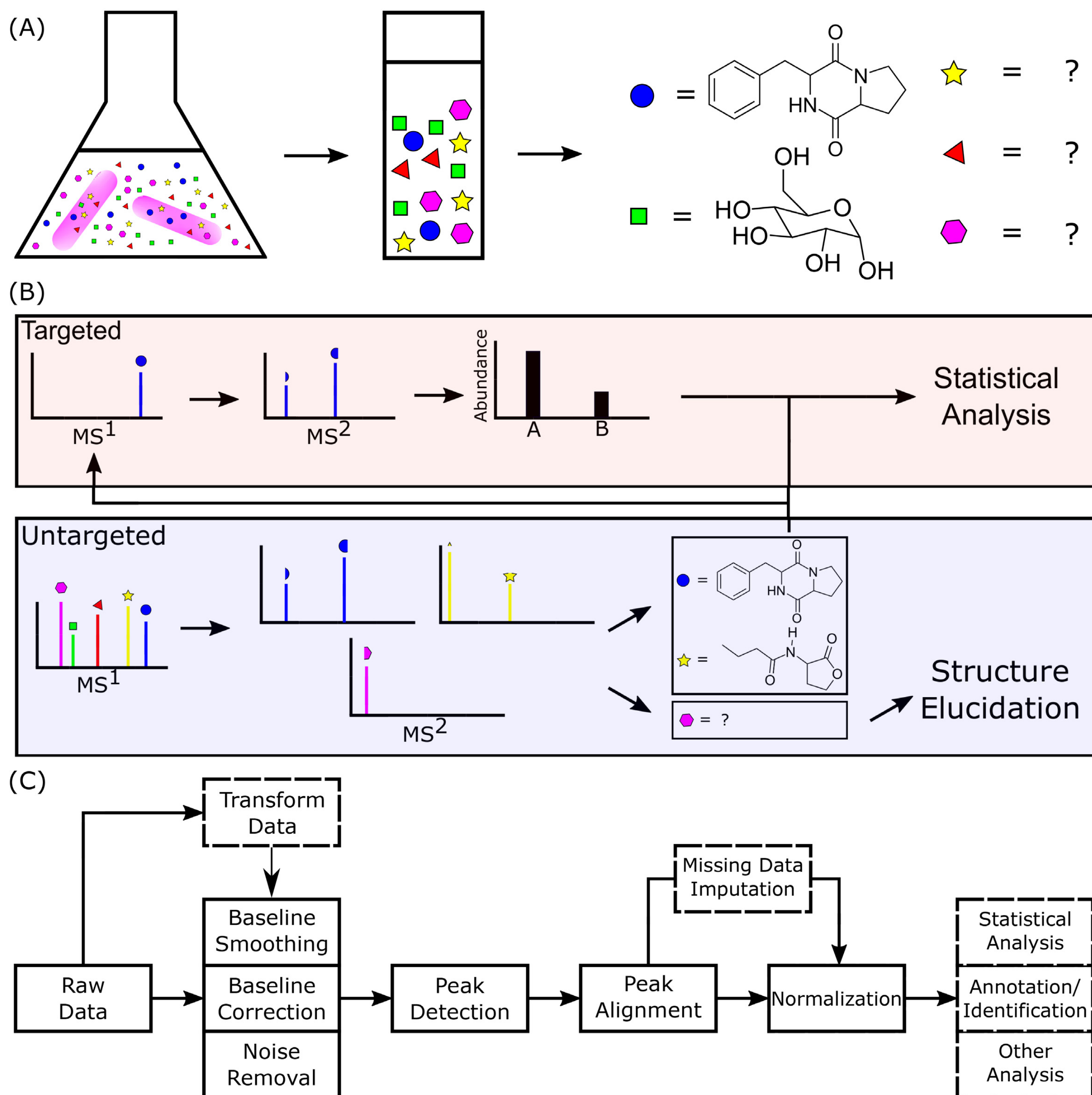

Figure 2. *Metabolomics Data Acquisition and Analysis*. (A) Samples may be grown in solid or liquid (pictured) based cultures and extracted using an extraction method of choice. Crude extracts may contain metabolites that are known or unknown, denoted by different colored shapes. Targeted or untargeted metabolomics experiments may then be performed on these extracts. (B) In targeted analysis (boxed in red, transition pairs consisting of precursor ions from MS^1 and product ions from MS/MS are acquired and compared across different conditions. Differences in metabolite expression can be used to infer biological significance. Untargeted analysis (boxed in blue) collects data on all metabolites using MS^1 and fragments each precursor ion in MS/MS into fragment ions (represented by partial shapes). MS/MS data can then be used for database searches to identify 'known unknowns'. If identified, further analysis or targeted experiments may be performed. If not identified, structure elucidation using MS and NMR can reveal the identity of the unknown metabolite(s). (C) Preprocessing workflow for metabolomics data. Solid boxes indicate steps that should be performed. Dotted boxes indicate optional steps. Stacked boxes indicate steps that may be performed in no particular order. It should be noted that statistical analysis and annotation/identification are not the only options for data analysis.

It was determined that the measured metabolic pathways such as valine/leucine/isoleucine degradation and phenylalanine/tyrosine/tryptophan biosynthesis in MRSA differed based on the concentration of glucose delivered in conjunction with methicillin. Untargeted metabolomics, on the other hand, involves detection of metabolite production between one or more microbial strains and the identification of metabolites. Identification can be done using fragmentation databases if the metabolites have been previously characterized (aka 'known unknowns') or structure elucidation if the metabolites have never been identified before (Fig. 2b). An untargeted metabolomics experiment was performed by Nguyen et al., where the bananamide compounds were identified from *Pseudomonas fluorescens* unique to the banana rhizoplane in the wetlands of Galagedara, Sri Lanka (Nguyen et al. 2016). Baptista et al. also used an untargeted approach to elucidate

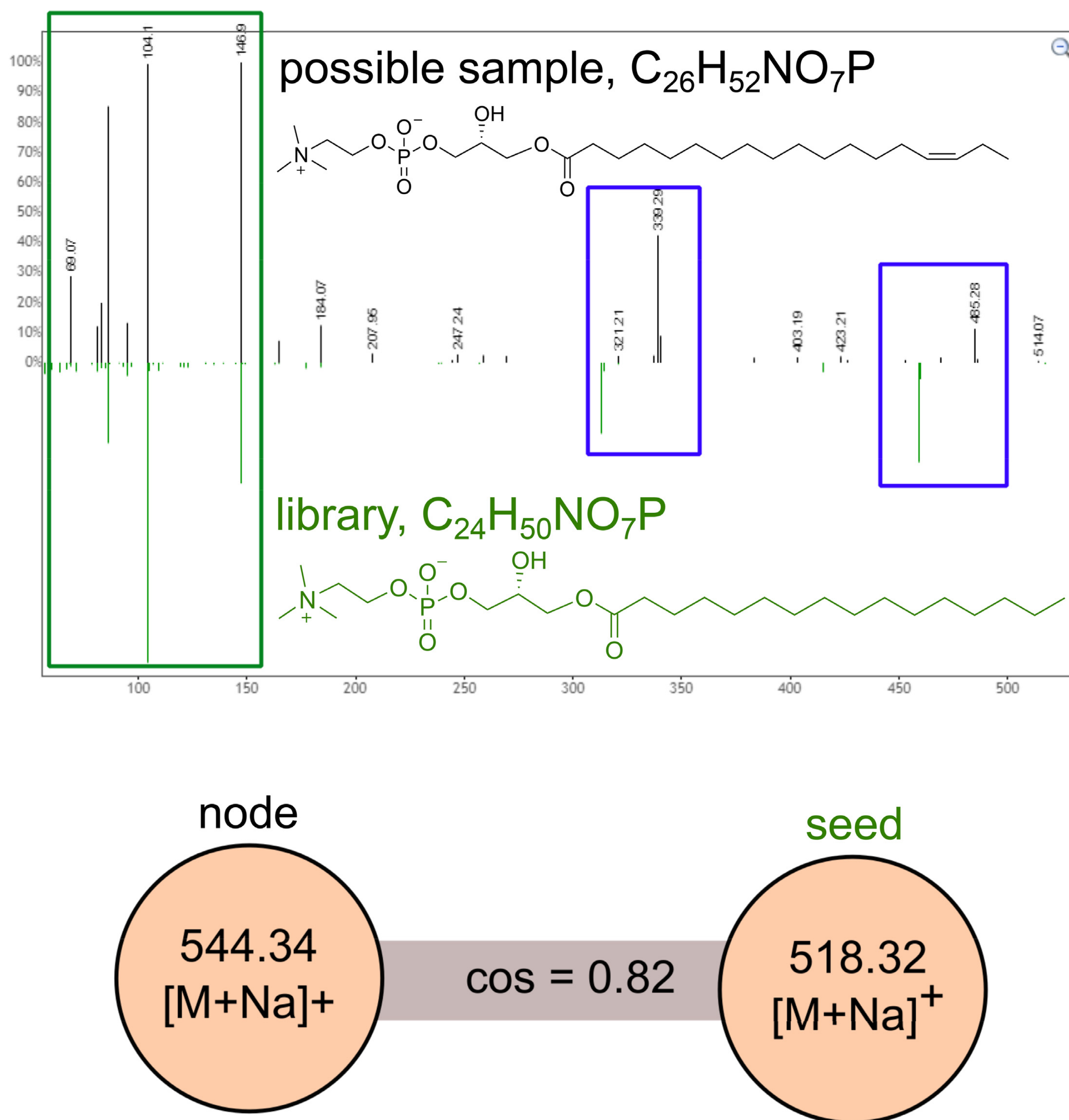

Figure 3. *GNPS Clustering.* MS/MS spectrum of a library match in a mirror plot with an experimental consensus MS/MS spectrum. The green box highlights ions that match the library spectra, while the blue boxes highlight ions that differ from the library. The power of inspecting the spectra can allow a user to infer which regions of the molecule might remain identical and pinpoint other regions that have changed. In this example with the mass difference of 26 Da (544–518) and coupled with the suggested molecular formulas acquired from an HRMS measurement would suggest that there is an addition of C_2H_2 to the library compound (also referred to as a seed, as it seeded this experimental node in the molecular network). Based on conserved product ions, which in this case correspond to the inherently charged polar head of this lipid, the best case for the addition of C_2H_2 to the experimental metabolite would be somewhere along the tail. More work would be needed to definitively ID the placement and the stereochemistry of the double bond (*E*/*Z*). It is worth noting that interpretation of the spectra can be difficult but we recommend readers to Demarque et al. for further explanation of rules that govern fragmentation patterns and mechanisms (Demarque *et al.* 2016).

the mechanism of action of pretomanid, an anti-tubercular drug currently in phase III clinical trials. Metabolome analysis of *Mycobacterium smegmatis* treated with a variety of antibiotics identified methylglyoxal as a unique toxic metabolite produced when treated with pretomanid (Baptista et al. 2018). Methylglyoxal is highly reactive and glycates to amino acids and nucleotides, causing damage to proteins and DNA that leads to cell death in *M. smegmatis* (Murata-Kamiya and Kamiya 2001; Thornalley *et al.* 2003, 2010).

Big data

The combination of ultra and high-performance LC systems and faster, high-accuracy mass spectrometers has helped to allow metabolomics to expand in experimental utility (Swartz 2005; Taleuzzaman et al. 2015; Sturm et al. 2016). Consequently, there are a wide variety of options when deciding on instruments in an experimental design. Multiple instruments/methodologies can

also be used based on their respective merits to perform orthogonal analyses that are necessary to confidently draw conclusions from experiments. This has resulted in an exponential increase in the amount of data that is generated. Haug et al. noted that as of May 2016, MetaboLights, one of the largest metabolomics data sharing repositories available, contained 4 TB of data and showed incredibly rapid growth since first being released. Currently (March 2019), MetaboLights contains 22.6 TB of submitted data. Metabolomics Workbench, its US-based sister site, contains 6 TB of data as well (Haug et al. 2013; Sud et al. 2016; Haug, Salek and Steinbeck 2017). It should be noted that this only includes data that has been shared publicly through these repositories and does not include data found on other repositories or data that has not been shared publicly. As the amount of data associated with experiments has grown, there has been proportional growth and development of tools used to process and analyze this data. An example of one such tool is the R package XCMS, which is the software of choice for many researchers for mass spectrometry data analysis (Smith et al. 2006; Weber et al. 2017). First released in 2006, XCMS was developed as a free, open source R package for mass spectrometry peak detection and has since released a web-based platform for automated data analysis (Tautenhahn et al. 2012).

Choose your own adventure

As we have alluded to, we will delve deeper into specific aspects touched upon above in the following sections with a focus on LC-MS based workflows. In Fig. 1, we have created a flow chart which can also help direct a reader to the section they are most interested in applying to their microbial metabolomics questions.

Liquid chromatography

Before acquiring data in a metabolomics experiment, one must properly prepare the samples to be analyzed. A simplified workflow can be seen in Fig. 2a. Intracellular metabolite sample preparation involves the removal of media containing extracellular metabolites followed by quenching and sampling of cells and extraction of metabolites. Preparation of extracellular metabolite samples, on the other hand, requires only the removal of cells followed by extraction of metabolites from the supernatant. When performing global metabolite analyses, one should consider designing an extraction that can target both intracellular and extracellular metabolites. Some metabolites may be stored intracellularly or associated with the cell wall, so utilizing solvents that lyse cells walls or mechanical forms of extraction, such as sonication or flash freezing followed by grinding with a mortar and pestle, can greatly influence the ability of a researcher to capture a wide range of metabolites. Pinu et al. have previously discussed various techniques for preparation of both intracellular and extracellular metabolite samples (Pinu and Villas-Boas 2017; Pinu, Villas-Boas and Aggio 2017). Separation of metabolites preceding MS analysis is performed using various analytical chromatographic separations again using the same 'like dissolves like' principle (*vide supra*). There are three techniques employed, which vary based on research needs for the polarity of the metabolite(s): 1) hydrophilic interaction chromatography (HILIC) in which nonpolar metabolites will elute first and is typically carried out with an aqueous mobile phase; 2) normal phase LC in which the mobile phases are organic and can be used to separate both polar and non-polar metabolites and finally, 3) reverse phase (RP) LC in which polar compounds elute first and can be used with mixtures of aqueous and organic solvents, such as water and methanol or acetonitrile. Many 'drug-like' metabolites can be separated using RP which has increased its popularity in the last few decades. Therefore, samples containing more polar metabolites should be separated using HILIC. Conversely, samples containing more non-polar metabolites should be separated via RP LC. If analysis of both polar and non-polar metabolites are desired or the composition of the sample is completely unknown, HILIC and reverse-phase LC may be used together (Lei, Huhman and Sumner 2011; Yanes et al. 2011). The efficiency of the separation will depend on the choice of solvent and solvent gradient. Appropriate methods for sample preparation and sample separation should be chosen based on the goals of each experiment.

TARGETED METABOLOMICS AND FLUXOMICS

Targeted metabolomics instrumentation and data acquisition

The use of selected reaction monitoring (SRM) or multiple reaction monitoring (MRM) MS assays have been the gold standard for quantitation in targeted metabolomics workflows. These assays are generally performed on triple quadrupole (QqQ) MS and less commonly quadrupole ion trap (QIT) MS. Before beginning a targeted metabolomics experiment, one must determine 1) how many and 2) what specific metabolites of interest are to be analyzed. During reaction monitoring, the precursor ion is fragmented into product ion(s). This precursor ion and product ion pair, known as a transition, of a specific known metabolite(s) of interest are used to develop an assay that only acquires data for the *m/z* values that correspond to the precursor/product ion pair within a given tolerance window. These reaction monitoring experiments are unmatched in their sensitivity in measuring specific metabolites in complex biological mixtures. Götz et al. used a targeted metabolomics approach to quantify and confirm proline's role in the osmoregulatory mechanism of *Sulfurimonas denitrificans* (Götz et al. 2018).

Recently, parallel reaction monitoring (PRM), previously used in proteomics, has taken advantage of HRMS in targeted metabolomics experiments following a study performed by Ramanathan et al. in which the merits of HRMS for quantification were highlighted (Ramanathan et al. 2011). HRMS uses a data-dependent acquisition mode to collect MS/MS for abundant precursor ions in the MS^1 spectrum, as opposed to only specific transition pairs. Full-scan MS is nondiscriminatory and collects all data on ions that are passed through the instrument. PRM has been adapted to metabolomics applications to take advantage of this feature, allowing for collection of large amounts of data in exchange for acquisition speed and sensitivity (Gertsman, Gangoiti and Barshop 2014; Zhou et al. 2016); (Michalski et al. 2011; Zhou and Yin 2016).

Targeted metabolomics relies on rigorous assay development and validation to perform a successful experiment. Following optimization of parameters, the assay must be validated against a set of standards to ensure it is selective, sensitive, accurate, reproducible, and the sample can be recovered after detection and quantification. For example, Zhao et al. developed an MRM assay to quantify grape-derived polyphenol precursors and phenolic acid metabolites produced as a result of phase I/phase II metabolic enzymes by gut microbiota (Zhao *et al.* 2018). During development, parameters such as solvent used for sample dilution, LC solvent, and many technical parameters were all optimized. In addition, transitions for each of the two precursors and 16 phenolic acid metabolites of interest were determined.

Targeted metabolomics data analysis

Following acquisition of raw data for targeted experiments, several steps are involved in the analysis of collected data: identification, normalization, and quantification. Positive identification of metabolites requires a known standard analyzed on the same instrument with the same method and experimental metabolites to have matching transitions and retention times (Sumner et al. 2007). Sample intensities should be normalized based on internal standards. Metabolites are then quantified using one of several methods (peak height, full width at half-maximum peak area) based on their extracted ion chromatograms (EIC) and compared to make biological inferences. A variety of software, both commercial and open source, are available for targeted metabolomics analysis. For example, Xcalibur (ThermoFisher Scientific, Waltham, MA) and MultiQuant (SCIEX, Redwood City, CA) allow for quantitative analysis of targeted metabolomics data and export into open source formats. The included *chromatogram* class in the open source R package XCMS allows one to work with transition data generated during reaction monitoring experiments (Smith et al. 2006; Tautenhahn, Böttcher and Neumann 2008; Benton, Want and Ebbels 2010). Peak detection and alignment in XCMS also allows for comparison of samples, and packages in the R programming environment can be used for normalization, quantification, and subsequent univariate or multivariate statistical analysis using the stats, XCMS, or AssayR packages (Wills, Edwards-Hicks and Finch 2017; R Core Team 2019).

Fluxomics

Flux experiments involve using heavy labeled media to measure the rates of metabolic processes and exchange of metabolites between cells (Dai and Locasale 2017). While other metabolomics experiments focus on the presence, absence, or difference in metabolites, metabolite flux analysis (MFA) can determine the rate of metabolism, and focuses heavily on primary metabolism. In flux experiments, media with stable isotope-labeled reagents are fed to microbial cultures, which, in turn, ferment the reagents into amino acids and intracellular intermediates with the isotope atoms incorporated into their structure. For example, carbon-13 (^{13}C) is the most popular stable isotope for labeling. The location of the isotope on the metabolite's structure provides insight into what metabolic processes led to the formation of the metabolite. What becomes difficult is finding that specifically labeled atom in the structure, as a number of isotopologues can exist (Dai and Locasale 2017). It is worth noting that elements have isotopes, whereas molecules and metabolites have isotopologues for naming convention. Experiments have been designed to measure flux in colonies grown on solid agar (Wolfsberg, Long and Antoniewicz 2018), the effects of nutrition stress (Ji et al. 2018), and flux in co-culture (Gebreselassie and Antoniewicz 2015). Becker and Wittmann have thoroughly described a model MFA experiment using ^{13}C, GC-MS, and OpenFlux for *Corynebacterium glutamicum,* which provides step-by-step details on how a basic experiment is run (Becker and Wittmann 2014). For those looking for a diverse set of options when setting up a flux experiment with LC-MS/MS, we recommend the review written by Walvekar et al. (Walvekar et al. 2018). Table 2 contains a list of suggested software to look into for general flux experiments. The listed programs have been chosen for how recently they've been updated, how detailed their documentation is, and the breadth of experiments they can cover. However, this list is by no means exhaustive, and it may be necessary to search for programs better suited for the research question at hand (Dandekar et al. 2014).

GC-MS (Becker and Wittmann 2014) and MS/MS are all valid ways of conducting MFA analysis. MS/MS is still being developed and implemented in standard fluxomic workflows, and so while it may be useful in coming years, it is not currently a standard workflow (Choi and Antoniewicz 2019).

UNTARGETED METABOLOMICS

Untargeted metabolomics instrumentation

Unlike targeted metabolomics, untargeted metabolomics experiments do not have known transitions for metabolites in a complex mixture. Therefore, full-scan MS^1 and MS/MS can both be acquired during data acquisition. A wide variety of instruments are capable of acquiring untargeted metabolomics data and we will cover specifics below for MS^1 and MS/MS networking.

Untargeted metabolomics data analysis

Untargeted metabolomics experiments operate on the assumption that there is little to no prior knowledge of the metabolites in the sample being analyzed. Experiments rely on collection of as much data as possible. As a result, data analysis is much more complex in untargeted metabolomics and requires preprocessing which can directly influence how well one can interpret the acquired data. When planning an experiment, various experimental parameters and the goal of the experiment must be considered during its design. Likewise, no data analysis workflow is universal. Software used and steps taken to process and analyze data depend on the properties of the data and scientific question that has been posed. Therefore, it is crucial to understand the characteristics of data generated during an experiment and to choose appropriate software, algorithms, and parameters when analyzing data. Libiseller et al. and Manier et al. have demonstrated the impact of parameter settings in XCMS performance and have worked towards development of workflows for optimizing parameters based on provided datasets (Libiseller et al. 2015; Manier, Keller and Meyer 2019).

During data acquisition, most modern mass spectrometers will collect data in profile mode, where peaks are represented as a continuous waveform, as opposed to centroid mode, where peaks are represented as a bar. While profile mode data provides information related to peak shape, the relatively large size of the data can make processing and analysis much more computationally expensive and time-consuming. In some cases, the size of the data in experiments, large mass ranges, large number of samples, and a lack of available computational resources can make working with profile mode data impractical. Therefore, conversion of vendor-specific raw data formats to centroided universal file formats (.mzXML, .mzML, .cdf, .mgf) by using vendor-provided software or open-source solutions such as Proteowizard's msCovert can alleviate these issues (Chambers et al. 2012). After data conversion, preprocessing can be performed on the data. There are a number of options for processing data.The general workflow is outlined in Fig. 2c and the software can be found in Table 2. While certain steps of the data processing workflow may not always be necessary, it is not always apparent which steps are optional. Therefore, data should be processed at each step outlined in Fig. 2c and compared to the unprocessed data to determine the impact of processing. Data

Table 1. *GNPS Parameter Table*. Starting parameters for GNPS based on our own experiences in working with data collected from both LR and HR mass spectrometers as well as across different numbers of files.

Parameter	Description	HRMS Sugg.	LRMS Sugg.	Default
Precursor ion mass tolerance	Changes as a function of the resolving power of the instrument and number of data files. Lower values represent accurate mass data and a low number of files whereas a high number will identify more spectra as similar that may not be, but is needed for large datasets, represents the mass tolerance at the MS^1 level.	0.02	0.5	2 (for a high number of data files)
Fragment ion mass tolerance	Similar rationale exists for choosing values based on resolution and data files, represents the mass tolerance at the MS/MS level.	0.02	0.3	0.5
		Small dataset	Medium dataset	Large dataset
Min pairs cos	Sets the minimum cosine score to visualize an edge between nodes (identify relatedness across spectra from different metabolites). A smaller value leads to clustering less-related spectra and a higher value prevents additional edges. Consider adjusting this value if you notice strange connections in your clusters.	0.7	0.7	0.7
Minimum matched fragment ion (min-matched peaks)	The number of fragment ions that need to match between two different precursor ions in order to be connected by an edge. There are no hard rules to determine how many fragments will be produced from a precursor ion, so users are encouraged to look through their spectra to see how many major fragments tend to appear in their samples to determine this value.	6	4–6	4
Minimum cluster size	Consider setting a higher number for datasets that should have large amount of overlap or multiple replicates, or a lower number for the investigation of rare ions, or where there are no replicates. This parameter helps weed out ions that do not reproducibly fragment and instrument noise.	1	2–3	5+

can be plotted to visualize data before and after processing as well.

MS^1 networking

MS^1-based networking is a strategy that takes all of the ions from a given MS window (such as 200–2000 Da) and attempts to identify different metabolites of interest via molecular formula and/or retention time matching and visualizes related metabolites using a network. MetaMapR, MetaNetter2, and MetNet are examples of algorithms that have been developed for MS^1 networking (Grapov, Wanichthanarak and Fiehn 2015; Burgess *et al.* 2017; Naake and Fernie 2019). MetaMapR takes annotated mass spectra and corresponding metabolite structures from PubChem. It then creates a network based on structural similarity of the metabolites and annotates the molecular class of each metabolite. MetaNetter2 identifies metabolites and related ions such as $[M + Na]^+$ or $[M + K]^+$ based on differences in Dalton shifts within a specified window and identifies any ions that may correspond to biochemical reactions such as a loss of water from the metabolite structure (Δ18 Da). MetNet has been recently developed by Naake et al. in an attempt to bring MetaNetter2's functionality to the R programming environment, allowing for processing automation and integration with commonly used software such as XCMS and CAMERA (Kuhl et al. 2012; Naake and Fernie 2019). Understanding what biochemical reactions occur and how metabolites are transformed can provide insight on different biochemical pathways and microbial metabolism. Mangal et al. used MetaNetter2 in the analysis of algal-dissolved organic matter to determine biotransformations that could be correlated to microbial mercury uptake in normoxic environments, representing normal aquatic environments, and hypoxic environments, representing eutrophicated environments (Mangal et al. 2019). Using this approach, metabolites and certain biotransformations can be associated with eutrophication and used to determine the health of aquatic environments.

Global natural products social molecular networking (GNPS, MS/MS networking)

MS/MS networking is a visualization tool that can be used to query the relationships between ions (metabolites) from any number of input files. The relationships are based on both the MS^1 and the product ions from the MS/MS spectra. Relative and absolute mass differences and intensities between the products ions within a spectrum are compared across spectra to generate a cosine similarity score. The MS/MS spectra can be gathered in a number of ways from a microorganism including direct sampling of a developing microbial colony or by creating an extract from agar-based colonies or fermentation broth (Yang et al. 2013). This technique can, for example, identify molecules that may share a common core structure but have differing chemical modifications, such as additional amino acids, sugars, or different methyl patterns to name a few. This structural

Table 2. *Sample Software.* This is a list of some of the software mentioned within the text and where to find it. Some resources are also accessible to be downloaded for a personal, offline version and are indicated here.

Name	Function	Website	Documentation	Availability	Reference
Metabolomics Analysis					
XCMS	R package for preprocessing for targeted and untargeted metabolomics data from LC-MS	https://bioconductor.org/packages/release/bioc/html/xcms.html	https://bioconductor.org/packages/release/bioc/manuals/xcms/man/xcms.pdf	Offline	(Smith *et al.* 2006)
XCMS Online	Online implementation of XCMS	https://xcmsonline.scripps.edu/landing_page.php?pgcontent=mainPage	https://xcmsonline.scripps.edu/landing_page.php?pgcontent=documentation	Online	(Tautenhahn *et al.* 2012)
AssayrR	R package for quantification in targeted metabolomics and flux experiments	https://gitlab.com/jimiwills/assay.R	https://gitlab.com/jimiwills/assay.R	Offline	(Wills, Edwards-Hicks and Finch 2017)
CAMERA	R package for peak annotation	https://www.bioconductor.org/packages/release/bioc/html/CAMERA.html	https://www.bioconductor.org/packages/release/bioc/manuals/CAMERA/man/CAMERA.pdf	Offline	(Kuhl *et al.* 2012)
Thermo Xcalibur	Thermo Vendor Software			Offline	
AB SCIEX MultiQuant	AB SCIEX Vendor Software			Offline	
Data Conversion					
ProteoWizard's msConvert	Data conversion from vendor to open source formats	http://proteowizard.sourceforge.net/download.html	http://proteowizard.sourceforge.net/tools/msconvert.html	Offline	(Chambers *et al.* 2012)
Fluxomics					
SUMOFLUX	Free MATLAB software, contains workflows for *E. coli* and *B. subtilis* but can be expanded	https://gitlab.ethz.ch/z/sumoflux	https://gitlab.ethz.ch/z/sumoflux/blob/master/SUMOFLUXmanual.pdf	Offline	(Kogadeeva and Zamboni 2016)
WUFlux	Free MATLAB software, contains templates for chemoheterotrophic bacteria, cyanobacteria, and vanillin-degrading bacteria	http://13cmfa.org/	User manual included with download	Offline	(He *et al.* 2016)
MS^1 Networking					
MetaMapR	CytoScape plugin to group structurally similar annotated metabolites	https://github.com/dgrapov/MetaMapR	https://github.com/dgrapov/MetaMapR/blob/master/Manual/Metmapr%20v1.2.1%20tutorial%20v1.doc.pdf	Offline	(Grapov, Wanichthanarak and Fiehn 2015)
MetaNetter2	CytoScape plugin to group metabolites based on changes in mass and identify biochemical transformations	http://apps.cytoscape.org/apps/metanetter2		Offline	(Burgess *et al.* 2017)
MetNet	R package to provide MetaNetter2 functionality	https://bioconductor.org/packages/release/bioc/html/MetNet.html	https://bioconductor.org/packages/release/bioc/manuals/MetNet/man/MetNet.pdf	Offline	(Naake and Fernie 2019)
MS/MS Networking					
GNPS	MS/MS networking, spectra database	https://gnps.ucsd.edu	https://ccms-ucsd.github.io/GNPSDocumentation/	Online	(Wang *et al.* 2016)
In silico Tools and Databases					
Massbank of North America	Fragmentation database, experimental and *in silico*	http://mona.fiehnlab.ucdavis.edu/		Online	
European Massbank	Fragmentation database, experimental and *in silico*	https://massbank.eu/MassBank/		Online	
METLIN	Fragmentation database, experimental and *in silico*	https://metlin.scripps.edu/		Online	
DEREPLICATOR+	Peptide MS/MS annotation/dereplication, incorporated into GNPS	http://mohimanilab.cbd.cmu.edu/software/	https://bix-lab.ucsd.edu/display/Public/Insilico±Peptidic±Natural±Products±Dereplicator±Documentation	Online/Offline	(Mohimani *et al.* 2018)
SIRIUS4 + CSI:FingerID	Molecular formula calculator, MS/MS annotation/dereplication	https://bio.informatik.uni-jena.de/software/sirius/	https://bio.informatik.uni-jena.de/repository/dist-release-local/de/unijena/bioinf/ms/sirius/4.0.1/sirius-4.0.1-manual.pdf	Offline	(Dührkop *et al.* 2019)
LipidMaps	Lipid fragmentation database, experimental and *in silico*	https://www.lipidmaps.org/		Online	(The LIPID MAPS Lipidomics Gateway, http://www.lipidmaps.org/)
mzCloud	Fragmentation database, experimental and *in silico*	https://www.mzcloud.org/		Online	

information is captured in an MS/MS spectrum as differences in mass, which indicate losses of chemical modifications that are shared across related metabolites. This workflow is very powerful because the database of 'seed' metabolites is constantly growing allowing for deeper annotation of microbial metabolites from MS/MS datasets. A 'seed' is the spectrum of a known metabolite that can be found in the GNPS database, or the spectrum of a metabolite from an in-house database; the terminology seed is meant to represent a known spectrum that can ground the unknown data to a known entity. GNPS also acts as a repository itself, where members can upload their own reference spectra to grow the knowledge base (Wang et al. 2016).

Microbiologists have already started using GNPS to identify metabolite group differences across different extraction conditions or different media types (Crüsemann et al. 2016; Papazian et al. 2019). These experiments showed GNPS as a tool to identify major compound classes across many strains and conditions. Bauermeister et al. cultured six MAR4 *Streptomyces* strains and were able to compare metabolite production across strains as well as identify metabolite classes produced by the microbes. This led to a correlation of a cluster of novel ions with novel antibiofilm activity (Bauermeister et al. 2019).

It's worth noting that connectivity to a 'seed' in a molecular network still requires validation rather than considering this a definitive identification. This is an important tool that can help generate hypotheses but should not be considered a stand-alone resource for identifications. The molecular network is only as good as the data collected and is also heavily reliant on the use of appropriate parameters when setting up the network.

A few suggested starting parameters and their meanings are shown in Table 1, but users are encouraged to experiment with multiple sets of parameters and identify how that impacts their network. New users are encouraged to use Table 1, along with the network parameter presets when first learning GNPS. The presets are designed to handle a small (<5 files), medium (5–400 files), or large (>400 files) number of spectra and are good starting points for molecular networking.

In silico annotation, dereplication, and fragmentation databases

An important step in untargeted metabolomics experiments is the dereplication of 'known unknown' metabolites that are present among the hundreds/thousands in an extract. *In silico* database searches have proven to be an invaluable technique toward that end. Through the use of databases such as Massbank, METLIN, GNPS, and more, MS/MS peak data can be used to annotate metabolites and provide putative structural information that is absent in MS^1 data. Table 2 includes several other commonly used tools and databases. Despite its usefulness, *in silico* database searches are not absolute, and a compound's identity and structure must be experimentally verified using orthogonal methods such as MS/MS data, retention time, or NMR.

Although experimental data can provide the highest confidence in matching spectra to identify a metabolite, there is a bottleneck for how quickly as of yet uncharacterized metabolites can be isolated, fragmented, and uploaded to repositories. *In silico* databases, on the other hand, have a high number of searchable spectra that may still match to a number of identified structures. Therefore, *in silico* databases are able to assist in metabolite dereplication. Compared to GC-MS, the number of LC-MS/MS spectra that have been curated is much smaller. To fill this gap, *in silico* fragmentation has been used to generate a greater number mass spectra for comparison. METLIN has implemented *in silico* fragmentation simulations based on the spectra added to the database. As of 2016, METLIN contained 240 000 molecules where 13 000 had experimental MS/MS data and there were 160 000 *in silico* fragmentation structure predictions (Levin, Salek and Steinbeck 2016).

While METLIN relies on the strength of their size of accumulated data to develop predictions, GNPS looks to improve confidence in fragmentation predictions through molecular networking. The Network Annotation Propagation tool (NAP) uses the fragmentation of neighboring nodes for the node in question to improve confidence in the structure prediction (da Silva et al. 2018). This feature is available through GNPS, but we recommend users to familiarize themselves with the basic molecular network techniques before attempting the experimental features. Additionally, since both of these tools are bolstered by the community's data, we encourage researchers to upload their own MS/MS data. Additionally, *in silico* fragmentation databases are compatible with any workflow that acquires MS/MS data. *In silico* databases and fragmentation have been more extensively covered in several recent reviews (Gil de la Fuente et al. 2017; Blaženović *et al*. 2018).

Future perspective

Overall, mass spectrometry is a key tool in designing metabolomics experiments. Before starting, be careful to select a workflow that is appropriate for the research question. Some questions will require multiple tools, such as in Papazian *et al.* where principal component analysis, imaging mass spectrometry, METLIN, and GNPS were used to identify and compare metabolites across eelgrass leaf extracts, show their spatial distribution, and relate that to their bioactivity (Papazian *et al*. 2019). Combining tools allows for answering overarching biological questions and we hope that microbiologists feel more equipped to think critically about their metabolomics experiments to enhance their science. In doing so, it is imperative that raw data be uploaded to public repositories along with publication. Data sharing allows for further validation of published results and provides training datasets when developing informatics tools.

ACKNOWLEDGEMENTS

We would like to thank JL Wolfender and M Litaudon for the library match spectrum featured in Fig. 3 which was accessed through GNPS.

This publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R01GM125943 and by the National Science Foundation grant 1817955.

Conflict of interests. None declared.

REFERENCES

Baptista R, Fazakerley DM, Beckmann M *et al*. Untargeted metabolomics reveals a new mode of action of pretomanid (PA-824). *Sci Rep* 2018;**8**:5084.

Bauermeister A, Pereira F, Grilo IR *et al.* Intra-clade metabolomic profiling of MAR4 Streptomyces from the Macaronesia Atlantic region reveals a source of anti-biofilm metabolites. *Environ Microbiol* 2019;**21**:1099–112.

Beale DJ, Pinu FR, Kouremenos KA *et al*. Review of recent developments in GC–MS approaches to metabolomics-based research. *Metabolomics* 2018;**14**:152. DOI :10.1007/s11306-018-1449-2

Becker J, Wittmann C. GC-MS-Based 13C metabolic flux analysis. *Methods Mol Biol* 2014:165–74.

Benton HP, Want EJ, Ebbels TMD. Correction of mass calibration gaps in liquid chromatography–mass spectrometry metabolomics data. *Bioinformatics* 2010;**26**:2488–9.

Blaženović I, Kind T, Ji J *et al*. Software Tools and Approaches for Compound Identification of LC-MS/MS Data in Metabolomics. *Metabolites* 2018;**8**:31. DOI :10.3390/metabo8020031

Burgess KEV, Borutzki Y, Rankin N *et al*. MetaNetter 2: A Cytoscape plugin for ab initio network analysis and metabolite feature classification. *J Chromatogr B Analyt Technol Biomed Life Sci* 2017;**1071**:68–74.

Chambers MC, Maclean B, Burke R *et al*. A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol* 2012;**30**:918–20.

Choi J, Antoniewicz MR. Tandem mass spectrometry for 13C metabolic flux analysis: Methods and algorithms based on EMU framework. *Front Microbiol* 2019;**10**:31.

Crüsemann M, O'Neill EC, Larson CB *et al*. Prioritizing natural product diversity in a collection of 146 bacterial strains based on growth and extraction protocols. *J Nat Prod* 2016;**80**:588–97.

Dai Z, Locasale JW. Understanding metabolism with flux analysis: From theory to application. *Metab Eng* 2017;**43**:94–102.

Dandekar T, Fieselmann A, Majeed S *et al*. Software applications toward quantitative metabolic flux analysis and modeling. *Brief Bioinform* 2014;**15**:91–107.

David Sparkman O. Mass spectrometry desk reference 2. *J Am Soc Mass Spectrom* 2000;**12**:1144.

Demarque DP, Crotti AEM, Vessecchi R *et al*. Fragmentation reactions using electrospray ionization mass spectrometry: an important tool for the structural elucidation and characterization of synthetic and natural products. *Nat Prod Rep* 2016;**33**:432–55.

Dührkop K, Fleischauer M, Ludwig M *et al*. SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nat Methods* 2019;**16**:299–302.

Gebreselassie NA, Antoniewicz MR. 13 C-metabolic flux analysis of co-cultures: A novel approach. *Metab Eng* 2015;**31**:132–9.

Gertsman I, Gangoiti JA, Barshop BA. Validation of a dual LC-HRMS platform for clinical metabolic diagnosis in serum, bridging quantitative analysis and untargeted metabolomics. *Metabolomics* 2014;**10**:312–23.

Gil de la Fuente A, Grace Armitage E, Otero A *et al*. Differentiating signals to make biological sense - A guide through databases for MS-based non-targeted metabolomics: General. *Electrophoresis* 2017;**38**:2242–56.

Götz F, Longnecker K, Kido Soule MC *et al*. Targeted metabolomics reveals proline as a major osmolyte in the chemolithoautotroph Sulfurimonas denitrificans. *Microbiologyopen* 2018;**7**:e00586.

Grapov D, Wanichthanarak K, Fiehn O. MetaMapR: pathway independent metabolomic network analysis incorporating unknowns. *Bioinformatics* 2015;**31**:2757–60.

Halouska S, Zhang B, Gaupp R *et al*. Revisiting protocols for the NMR analysis of bacterial metabolomes. *J Integr OMICS* 2013;**3**:120–37.

Haug K, Salek RM, Conesa P *et al*. MetaboLights—an open-access general-purpose repository for metabolomics studies and associated meta-data. *Nucl Acids Res* 2013;**41**:D781–6.

Haug K, Salek RM, Steinbeck C. Global open data management in metabolomics. *Curr Opin Chem Biol* 2017;**36**:58–63.

He L, Wu SG, Zhang M *et al*. WUFlux: an open-source platform for 13C metabolic flux analysis of bacterial metabolism. *BMC Bioinform* 2016;**17**:444.

Henke MT, Kelleher NL. Modern mass spectrometry for synthetic biology and structure-based discovery of natural products. *Nat Prod Rep* 2016;**33**:942–50.

Ji F, Shen Y, Tang L *et al*. Determination of intracellular metabolites concentrations in Escherichia coli under nutrition stress using liquid chromatography-tandem mass spectrometry. *Talanta* 2018;**189**:1–7.

Kogadeeva M, Zamboni N. SUMOFLUX: A generalized method for targeted 13C metabolic flux ratio analysis. *PLoS Comput Biol* 2016;**12**:e1005109.

Kuhl C, Tautenhahn R, Böttcher C *et al*. CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Anal Chem* 2012;**84**:283–9.

Lei Z, Huhman DV, Sumner LW. Mass spectrometry strategies in metabolomics. *J Biol Chem* 2011;**286**:25435–42.

Levin N, Salek RM, Steinbeck C. Chapter 11 - From databases to big data. In: Holmes E, Nicholson JK, Darzi AW *et al*. (eds). *Metabolic Phenotyping in Personalized and Public Healthcare*. Boston: Academic Press, 2016, 317–31.

Libiseller G, Dvorzak M, Kleb U *et al*. IPO: a tool for automated optimization of XCMS parameters. *BMC Bioinform* 2015;**16**:118.

Luzzatto-Knaan T, Melnik AV, Dorrestein PC. Mass spectrometry tools and workflows for revealing microbial chemistry. *Analyst* 2015;**140**:4949–66.

Mangal V, Stenzler BR, Poulain AJ *et al*. Aerobic and anaerobic bacterial mercury uptake is Driven by algal organic matter composition and molecular weight. *Environ Sci Technol* 2019;**53**:157–65.

Manier SK, Keller A, Meyer MR. Automated optimization of XCMS parameters for improved peak picking of liquid chromatography-mass spectrometry data using the coefficient of variation and parameter sweeping for untargeted metabolomics. *Drug Test Anal* 2019;**11**:752–61.

Michalski A, Damoc E, Hauschild J-P *et al*. Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol Cell Proteom* 2011;**10**:M111.011015.

Mohimani H, Gurevich A, Shlemov A *et al*. Dereplication of microbial metabolites through database search of mass spectra. *Nat Commun* 2018;**9**:4035.

Murata-Kamiya N, Kamiya H. Methylglyoxal, an endogenous aldehyde, crosslinks DNA polymerase and the substrate DNA. *Nucleic Acids Res* 2001;**29**:3433–8.

Naake T, Fernie AR. MetNet: Metabolite network prediction from High-Resolution mass spectrometry data in R aiding metabolite annotation. *Anal Chem* 2019;**91**:1768–72.

Nagana Gowda GA, Raftery D. Recent advances in NMR-Based metabolomics. *Anal Chem* 2017;**89**:490–510.

Nguyen DD, Melnik AV, Koyama N *et al*. Indexing the Pseudomonas specialized metabolome enabled the discovery of poaeamide B and the bananamides. *Nat Microbiol* 2016;**2**:16197.

Oliver SG, Winson MK, Kell DB *et al*. Systematic functional analysis of the yeast genome. *Trends Biotechnol* 1998;**16**:373–8.

Papazian S, Parrot D, Burýšková B *et al*. Surface chemical defence of the eelgrass Zostera marina against microbial foulers. *Sci Rep* 2019;**9**:3323.

Phelan VV, Moree WJ, Aguilar J *et al*. Impact of a transposon insertion in phzF2 on the specialized metabolite production and interkingdom interactions of Pseudomonas aeruginosa. *J Bacteriol* 2014;**196**:1683–93.

Pinu FR, Villas-Boas SG. Extracellular microbial metabolomics: The state of the art. *Metabolites* 2017;**7**:43. DOI:10.3390/metabo7030043

Pinu FR, Villas-Boas SG, Aggio R. Analysis of intracellular metabolites from microorganisms: Quenching and extraction protocols. *Metabolites* 2017;**7**. DOI :10.3390/metabo7040053

Raamsdonk LM, Teusink B, Broadhurst D *et al*. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat Biotechnol* 2001;**19**:45–50.

Ramanathan R, Jemal M, Ramagiri S *et al*. It is time for a paradigm shift in drug discovery bioanalysis: from SRM to HRMS. *J Mass Spectrom* 2011;**46**:595–601.

R Core Team. *R: A Language and Environment for Statistical Computing*. 2019.

Rowan DD. Volatile metabolites. *Metabolites* 2011;**1**:41–63.

Rutowski J, Zhong F, Xu M *et al*. Metabolic shift of *Staphylococcus aureus* under sublethal dose of methicillin in the presence of glucose. *J Pharm Biomed Anal* 2019;**167**:140–8.

da Silva RR, Wang M, Nothias L-F *et al*. Propagating annotations of molecular networks using in silico fragmentation. *PLoS Comput Biol* 2018;**14**:e1006089.

Smith CA, Want EJ, O'Maille G *et al*. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* 2006;**78**:779–87.

Sturm RM, Jones BR, Mulvana DE *et al*. HRMS using a Q-Exactive series mass spectrometer for regulated quantitative bioanalysis: how, when, and why to implement. *Bioanalysis* 2016;**8**:1709–21.

Sud M, Fahy E, Cotter D *et al*. Metabolomics Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucl Acids Res* 2016;**44**:D463–70.

Sumner LW, Amberg A, Barrett D *et al*. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* 2007;**3**:211–21.

Swartz ME. UPLC: An introduction and review. *J Liq Chromatogr Relat Technol* 2005;**28**:1253–63.

Taleuzzaman M, Ali S, Gilani SJ *et al*. Ultra performance liquid chromatography (UPLC)–a review. *Austin J Anal Pharm Chem* 2015;**2**:1056.

Tautenhahn R, Böttcher C, Neumann S. Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinform* 2008;**9**:504.

Tautenhahn R, Patti GJ, Rinehart D *et al*. XCMS Online: a web-based platform to process untargeted metabolomic data. *Anal Chem* 2012;**84**:5035–9.

The LIPID MAPS Lipidomics Gateway, http://www.lipidmaps.org/. *LIPID MAPS Lipidomics Gateway*.

Thornalley PJ, Battah S, Ahmed N *et al*. Quantitative screening of advanced glycation endproducts in cellular and extracellular proteins by tandem mass spectrometry. *Biochem J* 2003;**375**:581–92.

Thornalley PJ, Waris S, Fleming T *et al*. Imidazopurinones are markers of physiological genomic damage linked to DNA instability and glyoxalase 1-associated tumour multidrug resistance. *Nucl Acids Res* 2010;**38**:5432–42.

Walvekar A, Rashida Z, Maddali H *et al*. A versatile LC-MS/MS approach for comprehensive, quantitative analysis of central metabolic pathways. *Wellcome Open Res* 2018;**3**:122.

Wang M, Carver JJ, Phelan VV *et al*. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol* 2016;**34**: 828–37.

Weber RJM, Lawson TN, Salek RM *et al*. Computational tools and workflows in metabolomics: An international survey highlights the opportunity for harmonisation through Galaxy. *Metabolomics* 2017;**13**:12.

Wills J, Edwards-Hicks J, Finch AJ. AssayR: A simple mass spectrometry software tool for targeted metabolic and stable isotope tracer analyses. *Anal Chem* 2017;**89**:9616–9.

Wolfsberg E, Long CP, Antoniewicz MR. Metabolism in dense microbial colonies: 13C metabolic flux analysis of E. coli grown on agar identifies two distinct cell populations with acetate cross-feeding. *Metab Eng* 2018;**49**:242–7.

Yanes O, Tautenhahn R, Patti GJ *et al*. Expanding coverage of the metabolome for global metabolite profiling. *Anal Chem* 2011;**83**:2152–61.

Yang JY, Sanchez LM, Rath CM *et al*. Molecular networking as a dereplication strategy. *J Nat Prod* 2013;**76**:1686–99.

Zhao D, Yuan B, Carry E *et al*. Development and validation of an ultra-high performance liquid chromatography/triple quadrupole mass spectrometry method for analyzing microbial-derived grape polyphenol metabolites. *J Chromatogr B Analyt Technol Biomed Life Sci* 2018;**1099**:34–45.

Zhou J, Liu H, Liu Y *et al*. Development and evaluation of a parallel reaction monitoring strategy for Large-Scale targeted metabolomics quantification. *Anal Chem* 2016;**88**:4478–86.

Zhou J, Yin Y. Strategies for large-scale targeted metabolomics quantification by liquid chromatography-mass spectrometry. *Analyst* 2016;**141**:6362–73.