Metabolomics and the Microbiome: Characterizing Molecular Diversity in Complex Microbial Communities

Sara P Puckett*, Division of Medicinal Chemistry, Department of Pharmaceutical Sciences, University of Connecticut, Storrs, CT, United States
Robert M Samples*, Department of Chemistry, University of Connecticut, Storrs, CT, United States
Patrick D Schloss, Department of Microbiology and Immunology, University of Michigan, Ann Arbor, MI, United States
Marcy J Balunas, Division of Medicinal Chemistry, Department of Pharmaceutical Sciences, University of Connecticut, Storrs, CT, United States

1 Introduction

The search for new bioactive molecules from natural sources has undergone periods of intensive focus as well as times when these sources were abandoned for faster techniques. Various drug discovery strategies have fallen in and out of favor, such as an intensive approach using combinatorial chemistry that resulted in relatively few new drug leads, given the effort expended.[1–3] More recently, natural products drug discovery has returned to prominence as those faster techniques have resulted in compound libraries with low structural diversity and decreased drug-likeness. In addition, significant advances have occurred in our ability to explore genomic capacity and apply cutting-edge analytical chemistry tools to propel natural product drug discovery.

Historically, natural products research has focused on compounds from terrestrial plants, followed by the study of microbially-derived metabolites and investigations of molecules produced by marine organisms. More recently, natural products research has focused on complex microbial samples, such as microbial associates of eukaryotic hosts (e.g., humans[4–7], marine invertebrates[8–12]), as well as environmentally-derived microbiome samples (e.g., oil seeps,[13] glaciers and ice sheets,[14] and uranium-contaminated soil[15]). In part, this shift toward the microbiome derives from studies that have shown that many metabolites isolated from macroorganisms are in fact microbially-derived.[8,16] In addition, host-microbe symbioses are becoming an increasingly important focus in biomedical research as microbial symbionts are now recognized to be major contributors to human health and disease.[17–20]

Defining and contextualizing the term microbiome can pose both scientific and philosophical questions but there is typically agreement that to be considered a member of a microbiome, an organism must perform a function within that system.[21–23] However, function can be very challenging to define, especially when the experimental tools needed for functional assignment are still in development. In host-microbe interactions a distinction must be made between the essential components of the assemblage (i.e., holobiont) and the temporal microbial associates from the environment. For example, intracellular *Wolbachia* symbionts of aphids would clearly be considered members of the insect microbiome as neither host nor symbiont can survive alone,[24] but bacteria that are transiently associated with human skin raise questions as to their persistence and function in the system. Herein, we broadly define a microbiome as any microorganism or group thereof that colonizes or interacts with a particular environment or host organism in a manner which makes it useful to consider as a unit. Thus, for purposes of this review, we focus primarily on challenges associated with metabolomics of these complex microbial communities.

Traditional analysis of metabolites has focused on identification and comparison of single molecules across communities or on metabolite profiles of single bacterial strains grown in culture. However, metabolite production in complex microbial communities

---

*These authors contributed equally.

may be dependent upon spatial and temporal interactions within the system and/or with external factors (e.g., host, climate). Investigation of molecules from these complex microbial communities largely relies on metabolomics, or profiling of metabolites that are biosynthesized, degraded, or otherwise altered by the metabolic processes of an organism or group of organisms. Small molecule metabolomics is most often accomplished using liquid or gas chromatography (LC or GC) coupled to mass spectrometers (MS) collecting data for both MS parent ions and MS/MS fragmentation,[25–28] although there are continual efforts toward NMR-based metabolomics.[25,29] Small molecule interactions within microbiome communities play essential roles in the complex molecular conversations occurring between hosts and their microbiota and/or between organisms within environmental microbiomes. Production of extracellular metabolites by microbial systems is well-known[18,19,30,31] and may be the result of parasitic, mutualistic, or commensal interactions.[30] Furthermore, numerous studies have shown that host-associated microbial communities play important functional roles such as preventing disease, promoting immune response, and providing defense against pathogens.[20,31–38]

There are ongoing efforts to describe complex microbial systems via genetic techniques as well as via genome mining through biosynthetic gene cluster analysis.[4,6,39,40] Although they are powerful techniques, 16S rRNA gene and shotgun metagenomic sequencing do an inadequate job of characterizing the full complexity of microbial communities. These approaches detail which bacteria are present in a community and what they are functionally capable of, respectively.[41,42] Going beyond these DNA-based approaches, metatranscriptomics indicates which genes are being expressed, but can only approximate the proteome of the community.[43,44] Even if it were technically feasible to characterize the meta-proteome at sufficient resolution,[45,46] these data would provide limited information about the in situ activity of the microbiome and how it interacts with its host. Each of these approaches attempt to provide a surrogate to describe the function of the microbiome. In contrast, the metabolome represents the actual collection of chemical outputs and inputs that mediate the exchange of resources between the community and its host.[47–49]

Significant technical challenges limit the widespread use of metabolomic analysis in microbiome samples, including dependence on commercial analysis workflows that are designed for studying pure cultures, use of inadequate and/or proprietary databases, a wide range of analyte concentrations within each sample, and a lack of tools for merging metabolomic and sequence-based analyses. Perhaps the most significant challenge is that only a small fraction of the metabolites in complex microbial communities have been characterized, leading them to be excluded from many subsequent analyses that traditionally ignore unknown molecules.[50] In numerous ways, the state of metabolomics methods for complex microbial communities is similar to that of 16S rRNA gene sequence analysis 20 years ago. Through the development of open source software packages like mothur and QIIME,[51,52] the ability to analyze sequence data by microbiome researchers was standardized, democratized, and opened the field to the growing number of researchers interested in studying the microbiome. Both of these software packages are based on a framework that assumes that most sequence data have not been characterized. The ability of mothur and QIIME to overcome this limitation is the primary reason they have had such a significant impact.

To fully realize the potential of metabolomics and facilitate broader adoption of the technique, researchers need analytical pipelines that provide reliable data on both known and unknown compounds, especially given that a large fraction of the metabolites from a microbial community are unknown.[50] This is not surprising since many of the microbial sources of these metabolites have yet to be cultured and even when in culture, microbial metabolite production can vary substantially under different conditions (e.g., monoculture versus co-culture, changes in pH, nutrient availability, temperature, etc.).[53] Although efforts have recently been made to establish metabolomics databases analogous to GenBank (i.e., free and curated) (e.g., Refs. 54–56), these are still in their infancy and to date are maintained by disparate entities to varying levels of accuracy and completion or have remained proprietary and inaccessible to most users.

In response to the complexity of metabolomics analyses, many microbiology researchers outsource their metabolomic analysis to commercial companies that use proprietary methods and databases. The result is that researchers are restricted to the sample preparation/extraction methods and analyses the company will do and are subject to the substantial costs associated with those commercial analyses. Furthermore, companies rarely, if ever, provide raw data back to the researchers making it impossible to reproduce the analysis or properly interrogate the data. This is especially problematic since a large number of features represent unknown compounds that are typically ignored by commercial metabolomics companies. The inability to re-annotate unknown features/patterns when new reference compounds have been characterized is a significant limitation to advancing the collective understanding of microbiome metabolite diversity.

This article will cover the metabolomics pipeline as related to LC-MS/MS analyses of complex microbial samples (Fig. 1) including an overview of representative software for metabolomics of microbiome samples. This article is not meant to be a comprehensive detailing of all available metabolomics analyses but rather of parameters and their related freely-available software that are important to consider for complex microbial samples, ranging from sample collection, extraction, and sample preparation to data acquisition, file conversion, and pre-processing and delving into metabolite annotation, structural relationships, statistical analyses, and finally verification of metabolomics results.

2   Metabolomics Pipeline for Complex Microbial Samples

In some ways, existing pipelines for metabolomics of other sample types can be modified or adapted for use with complex microbial samples such as those from host-microbe systems or from environmental microbiomes. However, there are several areas of special concern with these substantially more complex samples that can comprise small samples with low abundance but
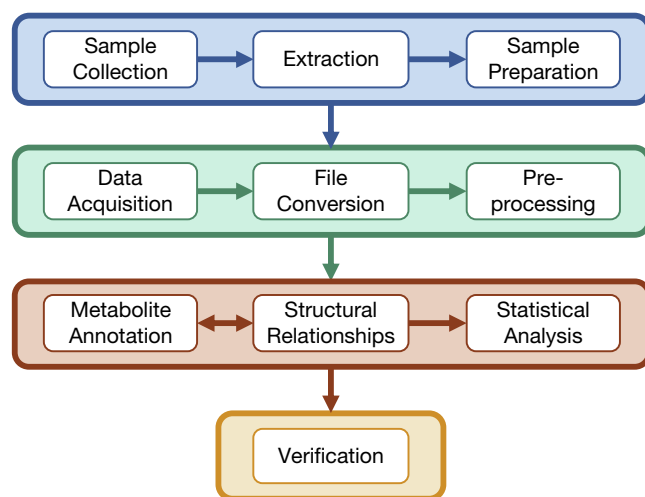
Fig. 1 Representative metabolomics pipeline for complex microbial samples. Methods for sample collection, extraction, and sample preparation occur prior to LC-MS/MS analyses. Parameters for LC-MS/MS data acquisition, file conversion, and pre-processing provide an important backdrop for further analyses such as metabolite annotation/identification, definition of structural relationships, and/or statistical analyses. Finally, given the complexity of both the microbiome samples as well as the immensity of data acquired and processed, all results need verification and possible reanalysis.

potentially significant metabolites. In addition to the generalizable metabolomics concerns related to sample preparation, data acquisition, file conversion, and pre-processing, there are additional variables to consider regarding sample collection and extraction techniques that can have substantial impacts on the metabolomics output, impacting metabolite annotation, analysis of structural relationships, and any statistical analyses that are performed. Because of these extra considerations for complex microbial community samples, verification of microbiome metabolomics represents a crucial step in the process of understanding the chemical communication in these samples and for utilizing this information for further natural products discovery.

2.1 Sample Handling

Preparation of microbiome samples for metabolomics requires careful consideration of several variables at multiple stages, including concerns for sample collection and storage, the extraction methods utilized, as well as how the samples are prepared for MS analyses. Because the full nature of metabolites produced by a complex microbial sample is frequently unknown to investigators, handling of these samples will involve compromises between spectrum of metabolites captured, background intensity, and concentration of analytes in the final sample, and these compromises must be made with the overall research question in mind (e.g., discovery of novel natural product metabolites versus understanding the chemical communication between a eukaryotic host and their specialized microbiome may alter sampling handling techniques). Because many of the microbiome metabolites of interest are likely present at low abundance, detection may require sample concentration, which can lead to compound degradation and/or loss of volatile compounds. In addition, microbiome metabolites may also be localized to host tissues and/or compartmentalized into cellular components,[16,57] and thus centrifugation or filtration may partition these molecules away from extracted material, depending on the protocols employed. An overview of metabolomics issues related to microbiome sample collection, extraction, and sample preparation is provided herein.

Sample collection: As with other aspects of microbiome research, specialized collection techniques are needed, dependent on the specific type of microbial samples and on the research questions to be addressed. Microbiome samples can be obtained from terrestrial, aquatic, or host-microbe environments. Maintaining sterility, locating and accurately sampling the desired environments, obtaining appropriate technical, biological and/or experimental replicates, and considering temporal variation such as time of year and/or time-course experiments all play a role in designing a robust collection protocol. Additional considerations include what sampling techniques will provide for accurate chemical representation of these complex microbial communities. For example, sampling of human skin done via swabbing will capture different metabolite profiles than samples collected via mechanical exfoliation with tape,[58] making it challenging to determine which more adequately represents the human skin microbiome. Furthermore, these techniques may be appropriate for analysis of microbiome communities and as a source of organisms for culture but may not provide enough material for direct chemical analysis of low abundance microbial metabolites and will not provide sufficient material for natural products isolation without employing a parallel culture-based pipeline.

Once samples are collected, sample storage can also affect metabolite composition. Processes such as air drying or lyophilization of samples can alter metabolite profiles (e.g., decreasing presence of volatile compounds).[59] Ambient conditions can also have an effect on the sample (e.g., oxidation) with one solution to store samples with heavier, labile gases such as argon to prevent these oxidative reactions. These considerations are typical of all natural products collections and not necessarily specific to microbiome collections but are particularly important to consider as samples are processed for metabolomics experiments.

Sample extraction: Extraction techniques for natural products have been extensively discussed elsewhere.[60] A variety of solvents including alcohols, esters, halocarbons, ketones, and hydrocarbons may be used depending on the research questions and sample source. As with other stages of sample handling, investigators should be aware that extraction artifacts can be imparted into samples, and given the complexity of microbiome samples, this can produce added complexity to determine the source of the detected metabolites. For example, acetal formation from alcohol and ketone solvents, ester formation due to use of alcohols, hydrolysis or solvolysis reactions, and nucleophilic addition of halocarbons to natural products are all well-known phenomena. Thus, with complex microbial community analyses, extraction solvent choices, extract conditions (e.g., minimizing exposure to heat, light, and air), and use of appropriate controls/blanks will all be essential to ensuring that the metabolomics results are inclusive of metabolites derived from the natural samples.[61]

Of notable concern, commonly utilized techniques such as rotary evaporation and lyophilization may cause compound degradation and/or loss of volatile compounds while extraction, by its nature, inevitably removes some components from a system.[28] In addition, solvent impurities and/or impurities from extracting with adsorbent beads (e.g., Diaion HP20, XAD resins) can be an issue during extraction but can be minimized by using high quality solvent, reducing contact between organic solvents and potentially leachable laboratory plasticware not rated for chemical exposure,[61] and by ensuring adsorbent beads are properly washed prior to and during extraction process.

In addition, the use of organic, lipophilic, or aqueous extraction techniques will drastically impact the metabolite profile of the sample. Isolation and identification of natural products has traditionally focused on organic extracts while metabolites at either end of the polarity spectrum are minimally extracted.[60] However, lipophilic and/or aqueous metabolites may be of increased interest to fully understand the chemical communication within host-microbe systems and/or for environmental microbiomes. For a comprehensive representation of all metabolites in a complex microbial community, multiple types of extraction will need to be employed but these will be dependent on the amount of sample available and thus most studies use only a subset of possible extraction methods determined based on their overarching research questions.

While not often the focus of traditional natural product discovery, volatile compounds may be important factors mediating chemical signaling in microbiome communities. Analysis of volatile compounds requires specialized techniques such as volatile adsorbent traps or solid-phase microextraction (SPME).[62] These types of approaches can sample volatile organic compounds (VOCs) from fecal microbiome samples and pheromone-like products of insect microbiome samples.[63]

Sample preparation: Sample preparation techniques can also have a significant impact on metabolomic output from complex microbial communities. The manner in which a sample is dissolved can substantially alter the data output, often related to a mismatch in the polarity of the metabolites of interest with that of the LC-MS solvent system resulting in loss of poorly soluble metabolites. Samples of moderate polarity may dissolve well in like solvents (e.g., 50/50 methanol/water or 50/50 acetonitrile/water) but compounds at either edge of the polarity spectrum may require more lipophilic or more hydrophilic solvents for dissolution. In addition, these samples may require specialized columns for LC separation of metabolomics experiments. A variety of MS techniques exist to investigate molecules from minimally processed samples, such as desorption electrospray ionization (DESI) and matrix-assisted laser desorption/ionization (MALDI) which both allow direct sampling of tissue or microbiome samples, allowing for analyses not testable with traditional sample preparation methods. Considerations at numerous stages of sample handling can have significant impacts on the metabolomics of microbiome samples, especially important given the inherent complexity in these samples.

2.2 Data Acquisition

Given the broader research conclusions often derived from metabolomics workflows, data acquisition represents a crucial stage in this pipeline, especially when dealing with complex microbiome samples. LC-MS/MS data acquisition methods can pose particular challenges. Standardization of LC methods can make retention time data more generalizable but some sample sets, particularly from complex mixtures derived from microbiome samples, may need tailored methods that provide a more accurate representation of these metabolite profiles. Acquisition using high resolution MS (HRMS) instrumentation has become more commonplace and helps to ensure that annotations and conclusions drawn from the data are more accurate. Many high resolution acquisition approaches incorporate an internal standard or utilize lockspray correction where an error correction factor is applied to all masses based on the measured deviation of the known compound from its literature mass. In addition, the use of tandem MS (MS/MS or MS2) also provides additional data on the fragmentation of molecules that allows for further understanding of the composition of microbiome samples.

Acquisition methods: Tandem MS acquisition methods can broadly be grouped into data dependent and data independent acquisition methods (DDA and DIA, respectively[64]). In typical DDA methods, low energy MS1 scans are used to identify the most abundant molecular ions, typically the top three or top five, for further fragmentation using higher collision energy. Thus, MS2 spectra are obtained for each of the most abundant ions without interference from coeluting compounds, allowing for a concrete link between each selected MS1 parent ion and their MS2 fragment ions. However, conducting multiple MS2 scans between each MS1 scan will lengthen the overall sampling intervals for MS1 scans (e.g., Fig. 2), often resulting in poor chromatographic peaks and potentially missing data (e.g., Fig. 2C). Compounds may escape detection if they elute between MS1 scans, especially for UPLC-MS/MS DDA methods in which peaks are often very sharp.[65] Importantly, ions from complex samples may not be subjected to MS2 fragmentation because they are less abundant than the top three or five coeluting ions, of particular concern for microbiome studies due to the large number of compounds present at wide ranges in concentrations, biasing detection against compounds of interest in
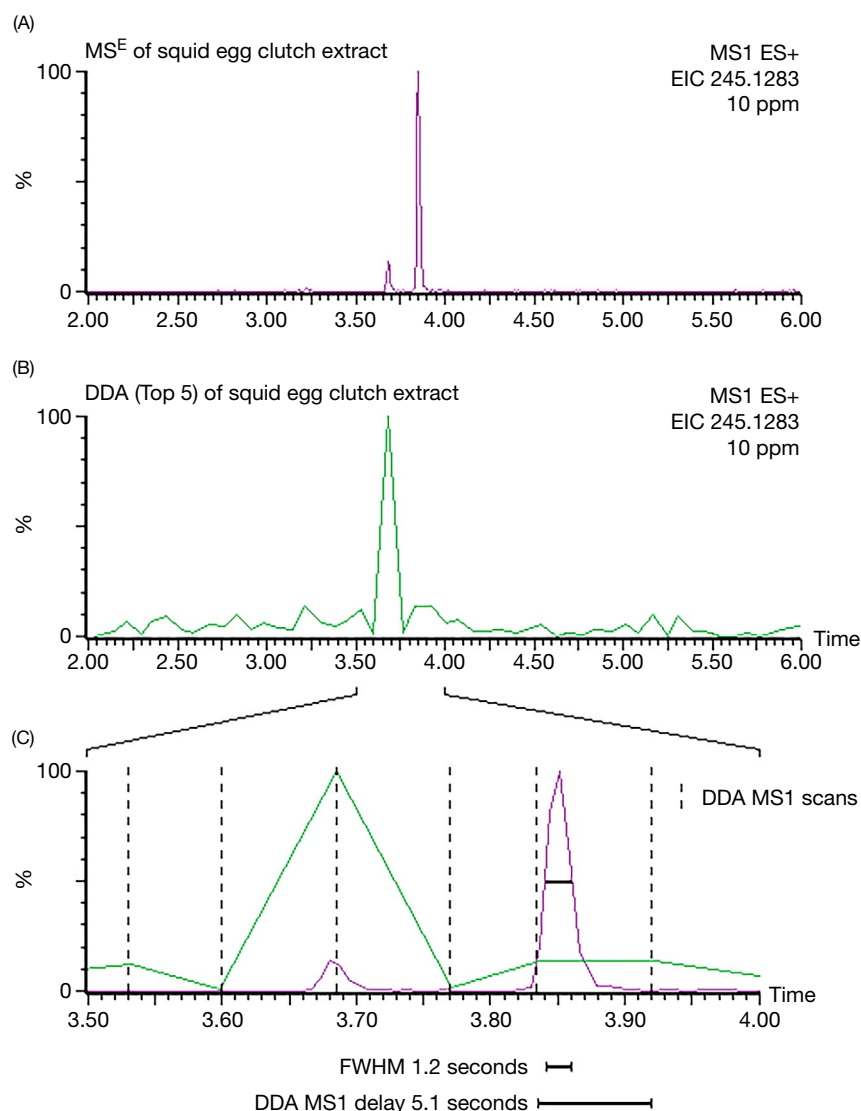
Fig. 2 Comparison of resolution and sampling from Data Dependent Acquisition (DDA) and Data Independent Acquisition (DIA). A representative microbiome sample (a squid egg clutch extract) was analyzed using both DDA (A) and DIA (B, specifically MS^E) methods using the same sample, column, and spectrometer. An expanded overlay of these traces (C) demonstrates the lengthy delay between DDA MS1 scan times (*dashed vertical lines*) that results in undersampling of some peaks (e.g., peak from MS^E at 3.85 min). With a substantially faster sampling rate, MS^E allows for better metabolomic coverage and more accurate quantification, although fragments are more challenging to match to parent ions.

favor of high abundance components.[65] In addition, because of the longer sampling time between MS1 scans, quantitation becomes increasingly inaccurate, eroding the value of metabolomics data in these studies. In some cases, DDA acquisition may also select in-source fragment ions leading to erroneous annotation of compounds. Various methods have been devised to counterbalance the sampling issues that arise from DDA acquisition (e.g., multiple runs and/or multiple iterative DDA scans, decoupling MS1 from MS2 to selectively target only ions of interest for MS2[66,67]). However, for complex microbial communities that often necessitate collection and analysis of large numbers of samples, multiple injections per sample are prohibitively expensive and come with other data consequences that can bias results.

In DIA methods, low energy MS1 scans establish the mass of molecular ions while concurrent higher energy MS2 scans capture fragment data. Faster data acquisition using DIA methods allows for substantial improvements in chromatographic peak shape, more accurate MS quantitation, higher coverage of narrow peaks, and fragmentation data for lower intensity compounds. However, all compounds eluting at a given point are fragmented simultaneously, making for significant challenges in peak deconvolution and accurate assignment of MS2 fragments to their corresponding MS1 parent ion. Thus, software strategies are employed to match ions with their fragments (e.g., matching peak shapes and elution times) but these can be especially cumbersome for samples with a large number of coeluting peaks. Orthogonal metabolite separations such as ion mobility or modified quadrupole scanning techniques

such as SCIEX SWATH® and Waters SONAR™ can also facilitate matching ions with corresponding fragments although this comes at a cost of reduced sensitivity. Ultimately, metabolite identification using DIA methods for data acquisition is significantly more challenging, especially given that many freely-available metabolomics programs are not designed to process DIA data and/or require extensive pre-processing prior to importing.

In addition to considering DDA versus DIA acquisition, MS data may be acquired in a raw full-spectrum form in which the mass spectrum at each timepoint is collected as a continuous signal, or can be acquired as collapsed point masses as in centroided data. Centroided files are dramatically smaller than full-spectrum files[68] however loss of information is intrinsic to centroiding, making troubleshooting and verification challenging. In addition, a single metabolomics sample, especially those from complex microbial communities, can contain analytes present at concentrations several orders of magnitude apart, causing detector oversaturation when attempting to analyze low abundance compounds. Detector oversaturation causes "ringing" in which erroneous small peaks will be present in between the main isotopic peaks. During centroiding, these ringing peaks can distort the accurate mass since this is no longer measured at the peak apex. However, processing of centroided data is easier computationally and analysis software typically requires centroided data as an input, which can be accomplished via file conversion tools as discussed below.

The type of MS instrumentation used can substantially affect data acquisition. Triple quadrupole instruments have a very fast sampling rate but have significantly lower mass accuracy than QToF instruments. Orbitrap instruments have higher scanning speed than QToF instruments and are available with a range of mass resolution capabilities. Because of their scanning speed orbitraps generally perform better with DDA acquisition than comparable QToF systems.

File conversion: Each of the many MS instrumentation vendors have their own proprietary acquisition software that provides complex data structures, often with each sample comprising multiple folders. These outputs typically cannot be read by typical text editors or most freely-available software, although many of these vendors do sell additional processing software that can read and/or convert data acquired on their instruments. To use most metabolomics software, these data must be converted to one of the open source formats (e.g., mzML, mzXML, mgf), often accomplished using the ProteoWizard Toolkit,[69] which will be discussed in detail in Section 3.1. Vendor-purchased data conversion programs include Bruker Compass DataAnalysis, Agilent MassHunter Qualitative Analysis, and Waters UNIFI, among others.

Many of the freely-available software for metabolomics discussed herein will accept mzML or mzXML formats with some support for vendor specific files (Table 1). These two file types are similar but differ in their overall structure. Historically, mzXML and mzDATA were the two preferred file types when processing MS data, but with minor differences and limitations.[70] As mzDATA was created to withstand advancements in software, mzXML was created using a more rigid structure. Both are widely available but difficulties were encountered when maneuvering between computer specialists and scientists or when a new acquisition file type was created. Thus, the best of both of these data formats were combined to mzML, including the ability to transform varying MS data formats, ease of adaptation as new instrumentation and techniques arise (e.g., conversion of Orbitrap data files). Several other open formats exist, including mgf (Mascot Generic Format),[71,72] a text-based file format widely utilized for proteomics and becoming more used in metabolomics processing. With microbiome samples, as with other sample types, it can be challenging to navigate file conversion, as much remains opaque with numerous options for user selection, some of which are likely used by only a subset of advanced users. However, some of the metabolomics software discussed herein will accept vendor files, alleviating the vagaries of file conversion.

Table 1   Acceptable file formats for representative metabolomic software.

XCMS Online	mzXML, mzData, NetCDF, wiff, d folders (Agilent, Bruker), RAW folders (Waters), RAW files (Thermo)
MetaboAnalyst	NetCDF, mzDATA, mzXML, txt, csv
MZmine	mzML, mzXML, mzData, NetCDF, RAW files (Thermo), RAW folders (Waters)
Global Natural Products Social (GNPS)	mzXML, mzML, mgf
SIRIUS 4	csv, ms, mgf

Pre-processing: Depending on instrumentation and acquisition software, some aspects of pre-processing may occur prior to file conversion. Pre-processing refers to a varied, multistep process of data manipulation wherein several functions are applied to the raw data, most often including peak picking, deconvolution, peak matching, and peak alignment,[73] with several steps along the way that are particularly important for pre-processing of microbiome samples. For example, within the peak picking step are noise reduction, background subtraction, and smoothing; for microbiome samples determination of noise cutoffs may be particularly important for any very low abundance metabolites whereas there may not be a straightforward background for subtraction other than instrument background. Deconvolution of multiple adducts, in-source fragments, and overlapping peaks is also especially important for complex microbial samples to ensure that metabolites are accurately represented within and between samples. Peak matching and peak alignment between samples are also of increased importance for microbiome samples as these will help address shifts in retention time, although some pipelines will also incorporate internal standards to measure relative retention times across wide sampling ranges. The importance of pre-processing cannot be overemphasized since the data manipulation steps utilized during this stage will propagate into all future analyses, with potential ramifications for metabolite annotation, assignment of structural relationships, and statistical analyses.

2.3   Data Analysis

Metabolomics data from microbiome samples can be subjected to multiple data analysis strategies, largely dependent on the overall research focus, with the vast majority of studies interested in metabolite annotation with increasing interest in structural relationships and statistical analyses. Many microbiome studies seek to employ a series of analyses to fully interrogate the chemical complexity of these samples. In practice, there is a considerable interplay between metabolite annotation and structural relationships, with exploration of structural relationships leading to better metabolite annotation and vice versa (e.g., Ref.[74]). In many ways, the same holds true for statistical analyses, especially as employed to address some of the challenges of data acquisition (e.g., utilization of statistical analyses on MS1 data to direct MS2 data acquisition circumventing some of the issues described above[66,67]). In all cases, accurate data analyses are dependent on the quality of data acquired and thus the importance of ensuring that microbiome samples are collected, extracted, and prepared in such ways that data acquisition and pre-processing provide the most representative data for analysis.

Metabolite annotation: Annotation of metabolites, or the assignment of compound names to MS features, remains a significant challenge especially for complex samples such as for metabolomics of microbiome samples. Detection of MS features in LC-MS/MS analyses has become increasingly straightforward, although determination as to whether these features comprise metabolites of interest remains challenging as pertaining to many of the sample handling and data acquisition issues discussed above. In addition, another confounding issue arises with in-source fragmentation during MS1 scans, which may be erroneously selected/annotated as parent ions, complicating determination that a particular MS feature should be considered a metabolite. An increasing body of literature, and accompanying software, aims to address these limitations with varying levels of success.[75]

In practice, annotation is only the first step in a lengthier process of accurate compound identification involving a suite of minimum data required to achieve the highest levels of confidence in these identifications.[76] High resolution LC-MS/MS data including molecular weight, mass defect, isotopic pattern, fragment weights, fragment abundance, fragmentation trees, and retention time allow for increased confidence in metabolite annotation.[26] Confidence in metabolite identification ranges from the lowest level of matching a high resolution exact mass (designated as a level 5 match), to confirming the molecular formula via MS isotope/adduct analysis (level 4), to having tentative candidates based on MS1 and MS2 (level 3), to having a probable structure via matching library spectrum (level 2), and finally to the highest confidence of a confirmed structure via matching to a reference standard (level 1). However, even at the highest levels, MS-based identifications can be subject to error and should also be verified via orthogonal techniques (discussed below in Section 2.4).

Metabolite annotations and identifications are only as comprehensive as the databases that are used for comparisons. Natural product databases have undergone considerable evolution over their history, beginning as compilations of physicochemical parameters such as molecular weights and melting points[77] and have now become computerized databases with terabytes of spectral, biological, and structural parameters. A subset of these natural products databases, metabolomics databases are rapidly growing and changing, with both in-house and publicly-available databases using either closed or community curation data platforms.[78] In addition to costly proprietary commercial databases, commonly utilized public databases include NIST (www.nist.gov/srd/nist1a.cfm), METLIN,[55] GNPS,[54] and KEGG.[79] Larger open databases such as PubChem (pubchem.ncbi.nlm.nih.gov) and Chemspider (www.chemspider.com[80]) are used by wide segments of industry and academia and offer a variety of experimental and computationally predicted physical and spectral data, but these may be challenging to use for natural products annotation due to the large number of synthetic or semisynthetic compounds that greatly outnumber the natural product hits. Additional specialized databases are tailored to metabolites from smaller taxonomic groups [e.g., PlantCyc (plants, www.plantcyc.org), StreptomeDB (*Streptomyces* sp.,[81]), and NPAtlas (microbes, www.npatlas.org[82])]. Some databases such as KEGG[83] and KNApSAcK[84] provide detailed pathway information related to annotated metabolites.

To date, selection of the most appropriate database can be challenging as most cannot currently provide comprehensive coverage for all types of metabolites and/or may not include ancillary data that can assist in accurate annotations (e.g., retention time, UV-Vis absorption, collisional cross section). Several databases include in silico predicted data (e.g., expected fragmentation pattern based on chemical structure) but this can prove problematic as these data may not be sufficiently reliable and as such some metabolomics databases have recently removed their in silico data in favor of expansion of their experimentally validated data (METLIN email communication June 2019). Additional obstacles can be encountered when working with community-curated databases since annotation accuracy can vary and thus orthogonal annotations and identification data are essential for accurate annotations based on these databases. Furthermore, many microbial metabolites remain uncharacterized and thus are not present in metabolomics databases. In these situations, such as for discovery of novel natural products compounds, metabolite annotation serves more for dereplication of the known molecules in a particular sample. In most cases, information from a variety of sources is necessary to bolster annotation confidence as well as to increase breadth of annotation coverage.[85]

Many software packages incorporate some automated annotation into their pipelines (Table 2). Automated software-based annotations are subjected to the same limitations as manual annotations with less human oversight and thus need to be interpreted cautiously. One objective metric for automated software annotations comes from the Critical Assessment of Small Molecule Identification (CASMI) contest, based on the Critical Assessment of protein Structure Prediction (CASP) challenge which first took place in 1994.[86] In both CASMI and CASP, a data set is provided (often comprising natural product molecules for CASMI) and requires participants to submit annotations. Although participants can manually annotation these data sets, there is a separate category for automated annotation and participants are scored in the categories of best molecular formula and best molecular structure,[87] providing an objective metric of each software.[88,89]

Table 2   Metabolomics capabilities of representative software for microbiome analyses.

Software	File conversion	Pre-processing	Annotation/identification	Structural relationship	Statistical analysis
ProteoWizard	X				
XCMS/XCMS Online		X	X		X
MetaboAnalyst		X	X		X
MZmine 2	X	X	X		X
GNPS			X	X	
SIRIUS 4			X		

Structural relationships: Analyzing and visualizing the relationships between related compounds produced by complex microbial communities allows for identification of compound families and can assist in targeting prospective novel compounds. There is a lack of structural similarity consideration in most existing open-source metabolomics software packages, with the exception of Global Natural Products Social Molecular Networking (GNPS, gnps.ucsd.edu,[54] described in detail in Section 3.5). However, understanding the structural relationships of metabolites within and between samples can provide significant advancement to describing the chemical composition of microbiome samples as well as to discovery of new molecules from these complex microbial samples. Using molecular networking or other clustering algorithms, computational comparisons are made between MS/MS fragmentation patterns from which structural similarity is then inferred, the details of which have been described elsewhere.[54] Metabolomic analyses of structural relationships can have significant impact on dereplication of known molecules from microbiome samples, to provide direction towards novel metabolites for natural product isolation and structure elucidation.[90]

Networking or clustering based on structural relationships can be especially useful for unknown compounds for which no annotation can be ascribed, since these unknowns may cluster with a known metabolite (e.g., Ref.[74]), pointing to the unknowns as possible analogs. Even when no known compound is present within a cluster, the importance of a series of metabolites may be inferred via comparisons between samples as in Fig. 3, in which a squid gland that houses a complex bacterial consortium (the accessory nidamental gland, ANG) was extracted and analyzed via LC-MS/MS and compared with extracts of bioactive bacterial
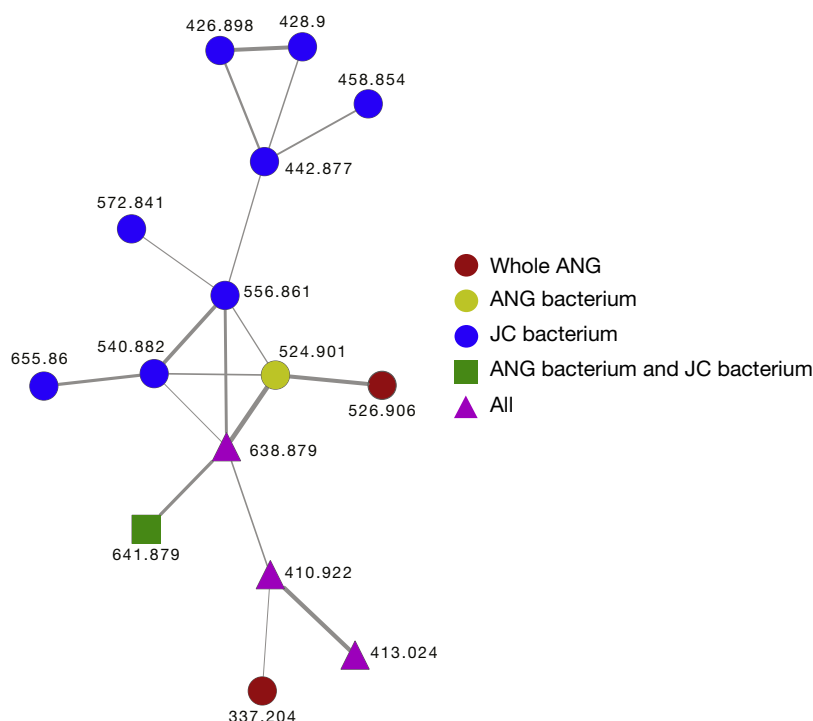

Fig. 3   Structural relationships between metabolites from microbiome samples using molecular networking. An expanded view of a portion of a network, focusing on a section with distinct but structurally related molecular ions from extracts of a squid gland that houses a complex bacterial consortium (the accessory nidamental gland, ANG, *red*), a bioactive bacterium isolated from the ANG (*yellow*), and a bioactive bacterium isolated from the jelly coat (JC, *blue*). Each node represents a consensus MS/MS spectrum from a single metabolite (labeled with parent molecular weights). Nodes are connected by lines, the thickness of which represents the degree of spectral/structural relatedness. Several nodes are found in all three extracts (magenta triangles), while two are only found in the whole ANG. Interestingly, a large number of these metabolites are found only in the JC isolate (*blue circles*), potentially indicating that additional secondary metabolite biosynthetic gene clusters are activated in the JC bacterium.

isolates from the ANG as well as from the squid egg jelly coat. Using the structural relationships that are inferred from these connected nodes, we were able to form testable hypotheses about the origin of these molecules and what these relationships might mean for their in situ function.

Despite the promise of structural relationship analyses in the field of microbiome metabolomics there remain significant technical obstacles. Metabolomic analysis of structural relationships began as an adaptation of similar analyses in proteomics.[91] Because of the biopolymeric nature of proteins and peptides and the extensive understanding of their fragmentation, analysis of their structural relationships using MS/MS data is now relatively straightforward. Because of their modularity, these proteomics-based metabolomics structural relationship analyses work well for NRPS and other biopolymeric metabolites from microbiome samples. However, for other types of microbiome metabolites, there are significant challenges due to structural diversity, complex scaffolds, smaller molecular weights, and/or greater variety of fragmentation possibilities. Small molecules can sometimes fragment in unpredictable ways and/or into uninformative fragments, and this fragmentation can be particularly dependent on the instrument and/or type of fragmentation technique employed.[27] Because of these issues structural relationship data should be carefully validated via cross comparisons with source data for confirmation.

Statistical analysis: Several types of statistical analyses can be applied to metabolomics data, dependent on the research questions to be addressed [e.g., univariate and multivariate analyses such as *t*-tests, analysis of variance (ANOVA) or multiple ANOVA (MANOVA), volcano plots, principal components analysis (PCA), partial least squares discriminant analysis (PLS-DA) or orthogonal PLS-DA (OPLS-DA), or hierarchical clustering analysis].[92] For microbiome samples, questions might range from determination of presence/absence of metabolites, to exploration of covarying metabolites from microbiome samples for patients with various disease states (e.g., carcinoma versus adenoma versus healthy), to comparison of disparate metabolite profiles from microbiome samples after experimental treatment (e.g., control versus treated), to impacts of drug treatment on microbiome metabolites, among many others.

Of primary importance for comparisons of metabolite production across samples is obtaining accurate abundance measurements which can pose considerable challenges for untargeted metabolomics experiments (reviewed elsewhere[93]). For complex microbial samples, employing untargeted metabolomics can provide an overall representation of sample metabolite profiles but can lead to challenges with measuring abundance.[94] Utilization of targeted metabolomics strategies can provide better metabolite quantitation but are restricted to a small number of metabolites[95] and thus not representative of the entirety of metabolites from microbiome samples. Variety in instrumentation can all affect quantitation as can types of data acquisition (e.g., DDA versus DIA as shown in Fig. 2).

Other statistical analysis issues arise for metabolomics data, including inadequate sample size, high false discovery rate, overfitting of models, inappropriate analyses, and bias due to confounding variables (as reviewed elsewhere[96]). Metabolomics allows researchers to investigate relationships between thousands of compounds and analysis of each metabolite must be considered an independent hypothesis, rapidly leading to testing of many thousand hypotheses in a single data set. Thus, methods for reducing the rate of false positives in multiple hypothesis testing should be employed to more accurately analyze metabolomics comparisons (e.g., Benjamini-Hochberg procedure,[97] Bonferroni correction[98]). In addition, as is true for nearly all scientific studies, sample replication is essential via biological, experimental, and technical replicates. Power analysis can be useful to determine the number of replicates required to detect the effect(s) in question for a particular metabolomics dataset. For metabolomics studies, injection replicates can provide an additional layer of controls, ensuring data integrity at the instrument level.

Combining multiple data analyses can also improve workflow efficiency and data reproducibility. For example, incorporation of structural relationship analysis with abundance data can indicate up or down regulation of a metabolite family produced by a microbiome under certain conditions such as pathogen presence or host stress. This approach can also be used to distinguish host-produced from microbially-produced metabolites by considering the presence or absence of related compounds in host or microbiome-only samples. Structural relationship studies may inform metabolite annotations and vice versa. Statistical analyses can be used to pinpoint which metabolites are most biologically relevant and thus focus annotation efforts on specific MS features from the thousands that typically arise using untargeted metabolomics.

2.4   Verification

Data verification is essential for conducting high quality metabolomics research as conclusions can be skewed by errors or bias in sample collection, extraction, data acquisition, peak detection, annotation, and/or statistics that are then propagated through the workflow making for challenges in determination of which MS features accurately represent sample data or comprise artifacts. Verification is essential at multiple levels beginning with confirmation of appropriate sample handling and data acquisition methods, checking consistency of processed data with raw data, evaluation of appropriate data analysis and statistical tools, and confirmation of metabolite annotation using orthogonal analytical techniques such as nuclear magnetic resonance (NMR) spectroscopy, electron cryomicroscopy (cryoEM), ultraviolet (UV), visible, and/or infrared (IR) spectral analysis, and physical properties.

The use of MS-based metabolomics has become more widespread in microbiome studies due to the high sensitivity, relatively low cost, and potential explanatory value of the data but there are several areas where verification is essential. Issues can occur during data acquisition and processing such as incorrect peak detection, inappropriate sample concentrations, and poor coverage of the full chemical space of a sample due to a poorly optimized separation method. Noise and detector saturation are particular issues due to the wide concentration range of individual analytes in microbiome samples so raw data must be inspected for ringing of high

concentration metabolites as well as for isotope patterns that may be lost in the noise for low concentration metabolites. Raw data must also be inspected for avoidance of poorly optimized separation methods that can result in unacceptable resolution with peak overlap and/or misrepresentation of metabolite composition due to solvent system and column selection. Following data acquisition and processing, processed data should be compared with raw data to confirm consensus masses with molecular masses, to validate structural relationships based on fragmentation patterns, and to ensure overall consistency between processed and raw data. Often these verification steps result in the need for new acquisition to reduce or increase injection concentrations, alter solvent systems to improve separation, and/or generally improve sample-specific metabolomics workflow.

Features of interest, based on annotation, fragmentation, and/or statistical analyses, must be verified, ideally via direct comparison to authenticated standards as discussed above. When authenticated standards are not available, it may be of interest to pursue isolation and identification using traditional natural products techniques (e.g., large-scale cultivation, chromatographic separation, 1D- and 2D-NMR structure elucidation), although not all microbiome samples will be amenable to bacterial isolation and cultivation. In these instances, new techniques such as microcrystal electron diffraction (microED), a cryoEM method that has gained recent recognition as a new way to elucidate compound structures,[99] may provide insights into metabolite composition of microbiome samples. Another possibility for verification of annotated features includes comparison of metabolomic data with metagenomic data, validating the occurrence of a MS feature with the parallel presence of the biosynthetic gene cluster in the metagenome. As these and other techniques mature, verification of metabolomics data for microbiome samples will become more robust, ensuring improved accuracy and increasing our ability to describe these complex microbial samples.

3   Representative Software for Metabolomics of Microbiome Samples

Once a researcher is able to obtain MS features and fragmentation patterns from their microbiome samples, there are commercial, vendor-specific software tools for analysis as well as a number of freely available tools for analysis (reviewed in ref.[73]). Many of these tools provide powerful methods for analyzing metabolomics data and several are suitable for microbial communities; however, they each have their own distinct limitations. In general, because many MS tools are web-based they are relatively easy for a researcher to interact with but the researcher is dependent on the availability and speed of the hardware on the server hosting the tools, which can be an issue for large microbiome datasets. This section will discuss several of the most widely-utilized, freely-available software tools for metabolomics analysis of complex microbial samples, including for file conversion, pre-processing, annotation/identification, structural relationships, and statistical analysis (Table 2), focused on ProteoWizard, XCMS/XCMS Online, MetaboAnalyst, MZmine2, GNPS, and SIRIUS 4.

3.1   ProteoWizard

ProteoWizard is a popular software package for processing, analyzing, and visualizing mass spectrometry data (proteowizard. sourceforge.net). It was created with the intention of being a framework for simplifying mass spectrometry file management and analysis and allowing more effective cross platform analysis through effective file conversion and processing.[100,101] The graphical user interface (GUI) support for ProteoWizard is limited relative to many other programs with msconvert, SeeMS, Skyline, and Topograph which handle file conversion, file viewing, MS/MS selective reaction monitoring, and metabolism processes, respectively. ProteoWizard finds significant utility in proteomics analysis but can also be used for small molecule applications. For metabolomics, the most notable tool in the ProteoWizard library is msconvert which allows users to select from a limited set of parameters for file conversion between mass spectrometry data from instrument-specific file formats to open data formats. ProteoWizard msconvert works across different vendor formats (e.g., AB SCIEX, Agilent, Bruker, Thermo, and Waters) to address the compatibility issues arising from widespread use of proprietary raw data formats among manufacturers of mass spectrometers. The msconvert GUI can convert many of these raw formats to mzML, mzXML, or mgf formats for further MS processing.

3.2   XCMS/XCMS Online

XCMS and XCMS Online are software packages for metabolite profiling that include retention time alignment, matched filtration, peak detection, and peak matching.[102,103] XCMS Online generally has similar capabilities as XCMS, but uses a GUI rather than a command-line interface, allowing for utilization by a greater audience. Some features within XCMS are lost in XCMS Online such as pre-processing abilities like peak picking.[102] XCMS Online was developed for use with untargeted metabolomics, as demonstrated by its robust ability to identify compounds using the associated METLIN database. Unlike most other programs discussed herein, XCMS Online is able to import some forms of raw data, allowing for a more streamlined workflow (Table 1; ref.[104]), although MZmine2 can also import a limited set of raw data as well. Several open file formats are accepted for upload into XCMS Online (Table 1) and additional criteria such as what type of LC-MS instrument (e.g., HPLC, UPLC, single quadrupole, Orbitrap, QToF) are also requested in order to best analyze samples. XCMS Online also allows for global modification of several parameters such as signal:noise ratio, annotation, and retention time correction, for the submitted runs as well as future runs. Once jobs are completed, all total ion chromatograms (TICs) can be visualized to correct shifts in retention time between samples. Extracted ion chromatograms (EICs) are also provided for each feature to allow for evaluation of presence/absence within the sample set(s). Within XCMS Online, there are three database searches (METLIN, isoMETLIN, and METLIN-MRM), dependent upon the analysis to be performed.

The METLIN database includes over 440,000 experimental MS/MS spectra (all in silico fragmentation data were recently removed as per email communication June 2019), while isoMETLIN focuses on analysis of isotopically labeled metabolites, and METLIN-MRM allows for identification and quantitation for multiple reaction monitoring experiments.[55,105]

Several statistical tools are available within XCMS Online, including common tools such as principal component analysis (PCA) and multidimensional scaling (MDS). XCMS Online will also create a "mirror plot" to monitor fold change between samples, in which compounds are represented by nodes starting at a threshold of zero indicative of no change between datasets. Nodes representing compounds found at higher concentrations are above the threshold while compounds found at lower concentrations are below and all compounds that are identified by METLIN are designated by a specific color. In addition, XCMS Online can create an exportable feature list for all molecular masses, including fold change, *p* value, *q* value, retention time, isotopic features, adduct formation such as common and neutral loss fragments, and whether the feature was identified by METLIN. XCMS Online can also be used concomitantly with external software such as genomics and proteomics platforms (e.g., BioCyc and UniProt, respectively) for integration into larger systems biology studies.[106,107]

In an example application of XCMS Online, toxin production by the fungus *Aspergillus carbonarius* was monitored in grape-vines.[108] This fungus produces the toxin, ochratoxin A, which is transmitted to humans upon consumption of grape products (e.g., juice and wine). GC-MS was used to monitor presence/absence of microbial volatile organic compounds (MVOCs), indicative of toxin production by *A. carbonarius*. Using XCMS Online statistical analysis tools to create an orthogonal projection to latent structures discriminant analysis (OPLS-DA) plot, moderately toxigenic strains were differentiated from highly toxigenic strains based on their metabolite profiles, with (*E*)-2-octen-1-ol, octanal, 1-octen-3-one, styrene, limonene, methyl-2-phenylacetate representative of moderately toxigenic strains whereas cuparene, (*Z*)-thujopsene, and methyl octanoate were reflective of highly toxigenic strains. These metabolites may serve as biomarkers for future analyses of *A. carbonarius* infection status of grape products.

In another study, XCMS was used to monitor interactions within a plant microbiome from *Cephalotaxus harringtonia*, a plant used for traditional medicine in Asia.[109] In preliminary spectroscopic studies, *Paraconiothyrium variabile*, a known endophytic fungus from *C. harringtonia*, was found to decrease abundance of the fungal pathogen, *Fusarium oxysporum*. Using XCMS, co-cultures of these two strains were analyzed, resulting in identification of the endophyte metabolite 13-oxo-9,11-octadecadienoic acid as important for *F. oxysporum* inhibition, namely by decreasing production of the mycotoxin beauvericin. Statistical analyses via XCMS pointed to the metabolite interactions within this microbiome, indicative of methods for monitoring relationships between endophytes and pathogenic fungi in other complex microbiome systems.

3.3 MetaboAnalyst

MetaboAnalyst is a freely accessible web-based server with many metabolomics applications (www.metaboanalyst.ca),[110] especially given the ability to read LC-MS, NMR, and GC-MS data, providing a much broader scope than other similar metabolomics software. Given that MetaboAnalyst will only accept file types of NetCDF, mzXML, mzDATA, and csv, pre-processing may be necessary (e.g., using ProteoWizard or vendor-specific export). MetaboAnalyst can then be used for data processing including peak detection and retention time correction for LC-MS or GC-MS data, as well as peak picking for LC-MS and NMR data. Data can then be normalized to decrease the number of systematic variances and to improve future statistical analyses. MetaboAnalyst will perform univariate statistical analyses such as measuring fold change differences, *t*-tests, and volcano plots as well as multivariate analyses such as PCA and PLS-DA. MetaboAnalyst can also perform hierarchical clustering to create dendrograms and/or heat maps, among various other types of analyses and visualizations. MetaboAnalyst will perform compound annotations for LC-MS or GC-MS (MS or MS/MS) or NMR (^{1}H, ^{13}C, or heteronuclear) data, although this may not be as robust as other software platforms. In the most recent version MetaboAnalyst 4.0,[111] improvements were made to integrate with metagenomics and transcriptomics data and to perform more robust identification of biomarker metabolites. In addition, the newest version of MetaboAnalyst will compare metabolite concentration with gene expression levels to suggest possible biosynthetic pathways.

MetaboAnalyst has been used to investigate metabolites from many different microbiome studies, including from both human, marine, and terrestrial environments. In one example, MetaboAnalyst was used to study the effects of histone deacetylase (HDAC) inhibitors on the fungus *Botryosphaeria mamane*, isolated from the roots of *Bixa orellana*.[112] Following treatment of the fungal isolate with the HDAC inhibitors suberoylanilide hydroxamic acid (SAHA) and sodium valproate, LC-MS was used to monitor secondary metabolite production. MetaboAnalyst was used to create volcano plots and to conduct partial least squares-discriminant analysis (PLS-DA) to pinpoint production of secondary metabolites resulting from addition of SAHA or sodium valproate to *B. mamane*. Although the study incorporated other metabolomics software platforms (MS-DIAL and GNPS), the statistical analyses provided by MetaboAnalyst were essential for identification of upregulated MS features.

MetaboAnalyst statistical tools have also been applied to gut microbiome research, including investigation of infants with neonatal hypoxic ischemic encephalopathy (HIE).[113] In an effort to offset some of the long-term negative consequences, a group of HIE infants were subjected to total body cooling and fecal samples were collected to compare their gut microbiome with those of healthy controls. Although significantly lower Bacterioides were found in the gut microbiota of HIE infants than in controls, no significant differences were seen in metabolomic samples as compared using MetaboAnalyst PCA and hierarchal clustering. Unlike many other microbiome metabolomics, this study utilized MetaboAnalyst to analyze ^{1}H NMR metabolite profiles from fecal samples, demonstrating the broad utility of MetaboAnalyst for incorporation of multiple types of spectroscopic data.

3.4   MZmine 2

MZmine 2 provides a metabolomics software framework to streamline importing, data processing, analysis, visualization, and export workflows (mzmine.github.io).[114] Such an approach requires a versatile, accessible, and modular software which allows users to construct data processing tools that meet their individual research needs. MZmine 2 improves upon the MZmine framework by improving modularity and flexibility, increasing ease of use and support for more varied, higher quality data. Notable changes include integrated annotation tools which use online databases such as PubChem and KEGG, as well as increased accuracy isotopic analysis, MS[n] data support, and new visualization techniques.[114]

Default MZmine 2 modules cover most aspects of general metabolomics processing, importing files formatted as mzXML, mzData, mzML, NetCDF, and thermo RAW. Imported data can then be filtered and smoothed before peak detection and deconvolution. Raw data is not removed at any stage of processing, allowing for review of processed and unprocessed data during verification stages. Unlike many proprietary software packages, intermediate stages of data processing are accessible to the user, making debugging and creation of new modules easier.[114] However, pre-processing with MZmine 2 is not without issues as highlighted by erroneous peak picking in some instances,[115] and thus users should carefully evaluate data to prevent propagation through their workflow. MZmine 2 can be used to align peak lists by mass to charge ratio and retention time using the RAndom SAmple Consensus (RANSAC) algorithm, allowing for more accurate peak alignment as well as supporting nonlinear alignment of chromatogram peaks where retention time shifts between injections vary. Annotation of peaks within MZmine 2 can be conducted with PubChem, KEGG, METLIN, HMDB,[116] or custom in-house databases. Mass data and individual spectra can be easily accessed within MZmine 2, as can various 2D and 3D visualization techniques, although only a limited suite of statistical analyses are available within the platform, (e.g., PCA, heatmaps, clustering).

As an example of MZmine 2 use for complex host-microbe samples, geographically distinct populations of the tunicate *Lissoclinum patella* were studied along with their symbiotic bacteria *Prochloron didemni* and *Candidatus* Endolissoclinum faulkneri.[117] MZmine 2 was used for processing (peak detection, noise cutoff, deconvolution, isotope peak grouping, and peak filtering) as well as to compare to a custom database of known tunicate-associated metabolites and obtain peak areas of these known metabolites. These data were then used to generate heatmaps and clustering of secondary metabolite production along this geographical range, following several NRPS-derived metabolites to determine that production of these metabolites related to variations in host phylogeny as well as to the two distinct symbionts.

In an intriguing combination of platforms, MZmine 2 has been used to augment molecular network analysis using GNPS,[91] which does not consider retention time, can create multiple nodes for the same compound, and/or merges MS2 data for isomeric compounds. Implementing MZmine 2 pre-processing into the GNPS workflow was shown to address these issues by including feature detection (mass detection, chromatogram builder, deconvolution, and isotopic peak grouper), semi-quantitative data (via alignment, gap-filling, and area values), and identification (using molecular formulas and a suite of databases). Combination of these two metabolomics platforms may increase reproducibility by reducing the effects of different instrumentation on network structure.

3.5   Global Natural Products Social Molecular Networking (GNPS)

Global Natural Products Social Molecular Networking (GNPS) is a web-based community-curated metabolomics platform using MS/MS data to determine structural relationships between mass spectral features.[54] These structural relationships are then used to create molecular networks that group compounds based on similarities in MS/MS fragmentation patterns. Networks are typically visualized using Cytoscape[118] with distinct MS features delineated as nodes, grouped into clusters based on similarities in fragmentation (see Fig. 3 for an example cluster). Clustered nodes are connected by edges, represented by lines whose thickness represents the relatedness of their fragmentation patterns as calculated via cosine similarity scores.[54] GNPS can be used on individual samples to explore chemical composition or can be used to compare groups of samples based on research hypotheses of how their metabolite profiles might relate. Features and/or associated fragmentation patterns can also be compared with databases to identify known metabolites, although existing fragmentation databases are limited in scope. A unique aspect to GNPS is the concept of "living data" in which the ongoing annotation of mass spectral data results in continual analysis of previously created networks.[54] GNPS will notify users when new annotations have been found in their datasets, increasing the overall speed of the metabolomic pipeline. One issue is that without orthogonal software (e.g., combination of MZmine 2 described above), GNPS does not utilize retention time or abundance metrics to generate molecular networks, although retention time and spectral count data are retained for investigator inspection. In addition, as discussed in Section 2.3 above, GNPS fragmentation analysis is performed using algorithms based on proteomics and thus will bias modular compounds with predictable fragmentation patterns.

GNPS has seen multiple uses within the natural products community including for the dereplication of known metabolites from extracts,[90] in conjunction with imaging mass spectrometry,[119,120] for annotation of structurally related analogs,[74] to monitor metabolite profiles upon changes in bacterial growth conditions,[10,121] and in combination with genomic analyses,[122,123] among others. Dereplication of known compounds was one of the early and ongoing uses for GNPS. Mass spectral data can be queried against curated databases such as MassBank[124] and NIST (www.nist.gov/srd/nist1a.cfm) as well as against the publicly-curated GNPS database in in which users can upload and annotate samples. Publicly-curated samples are provided confidence scoring ranging from gold (highest) to bronze (lowest) so that users can judge the reliability of annotation results. As discussed in

Section 2.3 above, annotations are only the first step in metabolite identification and community-curated database matches should always be verified via comparisons with raw data and/or orthogonal identification techniques.

In addition to traditional dereplication, GNPS also features in silico fragmentation tools such as DEREPLICATOR which was developed for peptidic natural products,[125] and more recently DEREPLICATOR+ which was developed for non-peptidic natural products.[126] In an example use, DEREPLICATOR and GNPS were employed to study a bacterial biofilm community from the oral microbiome.[127] DEREPLICATOR was used to assist in metabolite annotations and to putatively identify the compound, mutano-bactin A, produced by *Streptococcus mutans* UA159, corroborated by comparison with an authenticated standard. Numerous other features of interest were found, highlighting the large number of peptidic small molecules from the oral microbiome for future research.

In another microbiome-related study, GNPS molecular networking was used to compare lung tissue from a cystic fibrosis patient and *Pseudomonas* spp. isolates from that same patient.[128] LC-MS/MS data were collected and networked to show that while some nodes and clusters were unique to lung tissue or to the *Pseudomonas* isolates, there were also metabolites in common. Database matches were used to annotate the network, including cholesterol and other sterols from human-only samples, as well as a series of fatty acids and lipids found in common between the lung tissues and the *Pseudomonas* isolates. Work on the lung microbiome was continued and expanded to construct three-dimensional metabolomic maps of a diseased human lung using GNPS to explore distribution of human, microbial, and medication-related metabolites throughout a cystic fibrosis lung,[129] allowing hypotheses to be generated as to location of bacterial biofilms and the penetration of pharmaceutical interventions.

3.6 SIRIUS 4

SIRIUS 4 software provides metabolite annotations and identifications, building upon previous releases of SIRIUS (Sum formula Identification by Ranking Isotope patterns Using mass Spectrometry) software.[130] SIRIUS 4 integrates structure identification and elucidation capabilities, using high-resolution isotopic patterns of molecular and fragment ions to predict molecular formula of compounds, including fragmentation trees to improve accuracy. This recent release adds CSI:FingerID to the pipeline, constructing a "fingerprint" of a molecule based on possible structural fragments, comparing these to databases and ranking prospective identifications by consistency of fingerprint, mass accuracy, isotope match, fragmentation tree quality, and percentage of MS/MS peaks explained in terms of number of peaks and weighted peak intensity (i.e., the fraction of the total ion intensity explained by a fragmentation tree). The use of weighted peak intensity sets SIRIUS 4 apart from other tools, in that while no peaks are excluded from analysis, low intensity peaks affect ranked scoring less, ensuring consideration of signal-to-noise ratio in ranking.

SIRIUS 4 is the first incarnation of SIRIUS software that can be directly incorporated into an LC-MS/MS metabolomics workflow as previous version of both SIRIUS and CSI:FingerID only supported analysis of single MS2 spectra.[130] By contrast, SIRIUS 4 will accept mgf files from LC-based experiments, automatically populating an ion list, characterizing each ion independently, with putative identifications able to be propagated through metabolomics workflows. SIRIUS 4 currently allows for only data dependent acquisition (DDA), which presents issues discussed above. In concert with GNPS, SIRIUS 4 can also provide more accurate annotations of network features than are possible with GNPS alone. Given these substantial improvements, SIRIUS 4 is likely to be important for compound annotation and identification in future microbiome metabolomics experiments.

4 Conclusions and Future Directions

Metabolomics is poised to make significant advances in our understanding of microbiome chemical communication, providing information on functional differences between complex microbial communities. However, the increasing interest and rapid developments in microbiome metabolomics require careful considerations for best practices, assessment of caveats based on current limitations, and intensive verification of results. Issues ranging from sample handling and storage, sample extraction, data acquisition and processing, to data analysis and verification, all have significant impact on results and conclusions drawn from these studies.

In addition, all of the software platforms discussed herein, as well as the proprietary commercial platforms, have significant limitations. None adequately deal with issues of metabolite abundance, in part due to limitations in mass spectrometric quantitation, although there are new avenues towards improvements in this area.[93] Software to better incorporate both known and unknown metabolites into downstream analyses will allow for more accurate representation of the metabolic profiles of these complex microbial systems.[50] Many current tools have closed source code, limiting the ability of the community to understand how the underlying algorithms function or contribute to improve the tools. Related to this, parameter choices are generally limited within these tools and the range of values that the parameters can take is restricted. Vulnerabilities in metabolomics workflows can be at least partially overcome by use of redundant processing with several programs for verification and improved results (e.g., combining MZmine 2 to improve data input into GNPS[91]). In addition, there are continually new techniques in development for improving metabolomics of complex samples. For example, the algorithm BLANKA was recently designed to distinguish media components from those produced by the microorganism of interest,[131] subtracting them before analysis with programs such as GNPS which cannot remove these components during processing. New software is needed to adequately address the limitations of existing platforms and to move the field towards more comprehensive community-based analyses.

Many microbiome metabolomics studies are conducted to better understand metabolite diversity and chemical communication between members of these complex communities. A subset of these studies may provide results appropriate for natural product drug discovery pipelines, in which case sufficient material will be needed to move forward with extraction, fractionation, and isolation of bioactive molecules. Although culture-independent technologies are in continuous development,[132–134] most of these studies will likely proceed via culture-dependent techniques, seeking to determine and culture the microbial producer of the bioactive phenotype. However, only a very small fraction of bacterial species in a given sample can be cultured with standard laboratory techniques. Methods such as iChip using soluble soil/microbiome components to maintain microenvironments of sample location may assist in culturing more members from microbiome communities.[135] Even if the organism(s) of interest are cultured, this type of reductive isolation can result in a loss of biological activity, due to a loss of chemical cues and interactions from within the original complex microbial system. Various methods have been devised to recapitulate these chemical interactions, including co-culture,[10] chemical elicitation,[53,136] as well as culturing organisms in bioreactors,[137] to more effectively emulate in situ microbiome environments. Overcoming these challenges with metabolomics-based microbiome research will allow for significant advances in natural products drug discovery from complex microbial communities.

As mentioned earlier, the state of metabolomics for microbiome research is similar to that of 16S rRNA gene sequence analysis 20 years ago. Through the development of open source software, the ability to analyze sequence data by microbiome researchers was standardized, democratized, and opened the field to the growing number researchers interested in studying the microbiome. Through the use of tools such as mothur and QIIME, researchers have been able to make associations between 16S rRNA gene sequence data and phenotypes in animal models, human populations, and from environmental microbiomes. The ability to compare samples with similar characteristics using an ecological framework has had a significant impact on the direction of the microbiome field. Missing from these studies has been an in-depth analysis of the metabolites found in these sites. In the cases where metabolomic analyses are performed, the data are presented independent from, rather than integrated with, the sequence data that the investigators collected and often do not contain analyses using unknown compounds. We can apply many of the approaches used with operational taxonomic unit (OTU) data to metabolomics data to better understand the function of microbial communities, and we can also integrate metabolomics and genomics datasets to reach deeper insights into the relationship between the structure and function of microbiome communities.

Acknowledgments

References

1. Cragg, G. M.; Newman, D. J. Natural Products: A Continuing Source of Novel Drug Leads. *Biochim. Biophys. Acta* **2013**, *1830*, 3670–3695.
2. Newman, D. J. Natural Products as Leads to Potential Drugs: An Old Process or the New Hope for Drug Discovery?*J. Med. Chem.* **2008**, *51*, 2589–2599.
3. Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs From 1981 to 2014. *J. Nat. Prod.* **2016**, *79*, 629–661.
4. Milshteyn, A.; Colosimo, D. A.; Brady, S. F. Accessing Bioactive Natural Products From the Human Microbiome. *Cell Host Microbe* **2018**, *23*, 725–736.
5. Wilson, M. R.; Zha, L.; Balskus, E. P. Natural Product Discovery From the Human Microbiome. *J. Biol. Chem.* **2017**, *292*, 8546–8552.
6. Donia, M. S.; Cimermancic, P.; Schulze, C. J.; Wieland Brown, L. C.; Martin, J.; Mitreva, M.; Clardy, J.; Linington, R. G.; Fischbach, M. A. A Systematic Analysis of Biosynthetic Gene Clusters in the Human Microbiome Reveals a Common Family of Antibiotics. *Cell* **2014**, *158*, 1402–1414.
7. Donia, M. S.; Fischbach, M. A. HUMAN MICROBIOTA. Small Molecules From the Human Microbiota. *Science* **2015**, *349*, 1254766.
8. Gulder, T. A.; Moore, B. S. Chasing the Treasures of the Sea—Bacterial Marine Natural Products. *Curr. Opin. Microbiol.* **2009**, *12*, 252–260.
9. Haygood, M. G.; Schmidt, E. W.; Davidson, S. K.; Faulkner, D. J. Microbial Symbionts of Marine Invertebrates: Opportunities for Microbial Biotechnology. *J. Mol. Microbiol. Biotechnol.* **1999**, *1*, 33–43.
10. Sung, A. A.; Gromek, S. M.; Balunas, M. J. Upregulation and Identification of Antibiotic Activity of a Marine-Derived *Streptomyces* sp. Via Co-Cultures With Human Pathogens. *Mar. Drugs* **2017**, *15*, 250.
11. Gromek, S. M.; Suria, A. M.; Fullmer, M. S.; Garcia, J. L.; Gogarten, J. P.; Nyholm, S. V.; Balunas, M. J. *Leisingera* sp. JC1, a Bacterial Isolate From Hawaiian Bobtail Squid Eggs, Produces Indigoidine and Differentially Inhibits Vibrios. *Front. Microbiol.* **2016**, *7*, 1342.
12. Morita, M.; Schmidt, E. W. Parallel Lives of Symbionts and Hosts: Chemical Mutualism in Marine Animals. *Nat. Prod. Rep.* **2018**, *35*, 357–378.
13. Ruff, S. E.; Biddle, J. F.; Teske, A. P.; Knittel, K.; Boetius, A.; Ramette, A. Global Dispersion and Local Diversification of the Methane Seep Microbiome. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 4015–4020.
14. Anesio, A. M.; Lutz, S.; Chrismas, N. A. M.; Benning, L. G. The Microbiome of Glaciers and Ice Sheets. *NPJ Biofilms Microbi.* **2017**, *3*, 10.
15. Akob, D. M.; Mills, H. J.; Kostka, J. E. Metabolically Active Microbial Communities in Uranium-Contaminated Subsurface Sediments. *FEMS Microbiol. Ecol.* **2007**, *59*, 95–107.
16. Zan, J.; Li, Z.; Tianero, M. D.; Davis, J.; Hill, R. T.; Donia, M. S. A Microbial Factory for Defensive Kahalalides in a Tripartite Marine Symbiosis. *Science* **2019**, *364*.
17. Bevins, C. L.; Salzman, N. H. The potter's Wheel: The Host's Role in Sculpting Its Microbiota. *Cell. Mol. Life Sci.* **2011**, *68*, 3675–3685.
18. Crawford, J. M.; Clardy, J. Bacterial Symbionts and Natural Products. *Chem. Commun.* **2011**, *47*, 7559–7566.
19. Piel, J. Metabolites From Symbiotic Bacteria. *Nat. Prod. Rep.* **2009**, *26*, 338–362.
20. Curtis, M. M.; Sperandio, V. A Complex Relationship: The Interaction among Symbiotic Microbes, Invading Pathogens, and Their Mammalian Host. *Mucosal Immunol.* **2011**, *4*, 133–138.
21. Klassen, J. L. Defining Microbiome Function. *Nat. Microbiol.* **2018**, *3*, 864–869.
22. Ursell, L. K.; Metcalf, J. L.; Parfrey, L. W.; Knight, R. Defining the Human Microbiome. *Nutr. Rev.* **2012**, *70* (Suppl 1), S38–S44.
23. Marchesi, J. R.; Ravel, J. The Vocabulary of Microbiome Research: A Proposal. *Microbiome* **2015**, *3*, 31.
24. Correa, C. C.; Ballard, J. W. O. *Wolbachia* Associations With Insects: Winning or Losing Against a Master Manipulator. *Front. Ecol. Evol.* **2016**, *3*, 153.

25. Amberg, A.; Riefke, B.; Schlotterbeck, G.; Ross, A.; Senn, H.; Dieterle, F.; Keck, M. NMR and MS Methods for Metabolomics. *Methods Mol. Biol.* **2017**, *1641*, 229–258.

26. Riekeberg, E.; Powers, R. New Frontiers in Metabolomics: From Measurement to Insight. *F1000Res* **2017**, *6*, 1148.

27. Schrimpe-Rutledge, A. C.; Codreanu, S. G.; Sherrod, S. D.; McLean, J. A. Untargeted Metabolomics Strategies-Challenges and Emerging Directions. *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 1897–1905.

28. Verpoorte, R.; Choi, Y. H.; Mustafa, N. R.; Kim, H. K. Metabolomics: Back to Basics. *Phytochem. Rev.* **2008**, *7*, 525–537.

29. Nagana Gowda, G. A.; Raftery, D. Can NMR Solve Some Significant Challenges in Metabolomics? *J. Magn. Reson.* **2015**, *260*, 144–160.

30. Phelan, V. V.; Liu, W. T.; Pogliano, K.; Dorrestein, P. C. Microbial Metabolic Exchange-the Chemotype-to-Phenotype Link. *Nat. Chem. Biol.* **2011**, *8*, 26–35.

31. Rath, C. M.; Dorrestein, P. C. The Bacterial Chemical Repertoire Mediates Metabolic Exchange Within Gut Microbiomes. *Curr. Opin. Microbiol.* **2011**, *15* (2), 147–154.

32. Maslowski, K. M.; Mackay, C. R. Diet, Gut Microbiota and Immune Responses. *Nat. Immunol.* **2011**, *12*, 5–9.

33. Maslowski, K. M.; Vieira, A. T.; Ng, A.; Kranich, J.; Sierro, F.; Yu, D.; Schilter, H. C.; Rolph, M. S.; Mackay, F.; Artis, D.; Xavier, R. J.; Teixeira, M. M.; Mackay, C. R. Regulation of Inflammatory Responses by Gut Microbiota and Chemoattractant Receptor GPR43. *Nature* **2009**, *461*, 1282–1286.

34. Hajishengallis, G.; Krauss, J. L.; Liang, S.; McIntosh, M. L.; Lambris, J. D. Pathogenic Microbes and Community Service Through Manipulation of Innate Immunity. *Adv. Exp. Med. Biol.* **2012**, *946*, 69–85.

35. Schaudinn, C.; Gorur, A.; Keller, D.; Sedghizadeh, P. P.; Costerton, J. W. Periodontitis: An Archetypical Biofilm Disease. *J. Am. Dent. Assoc.* **2009**, *140*, 978–986.

36. Dorer, M. S.; Talarico, S.; Salama, N. R. *Helicobacter pylori*'s Unconventional Role in Health and Disease. *PLoS Pathog.* **2009**, *5*, e1000544.

37. Florez, L. V.; Biedermann, P. H.; Engl, T.; Kaltenpoth, M. Defensive Symbioses of Animals with Prokaryotic and Eukaryotic Microorganisms. *Nat. Prod. Rep.* **2015**, *32*, 904–936.

38. Bellezza, I.; Peirce, M. J.; Minelli, A. Cyclic Dipeptides: From Bugs to Brain. *Trends Mol. Med.* **2014**, *20*, 551–558.

39. Schubert, A. M.; Rogers, M. A.; Ring, C.; Mogle, J.; Petrosino, J. P.; Young, V. B.; Aronoff, D. M.; Schloss, P. D. Microbiome Data Distinguish Patients With *Clostridium difficile* Infection and Non-*C. difficile*-Associated Diarrhea From Healthy Controls. *MBio* **2014**, *5*, e01021-01014.

40. Zackular, J. P.; Baxter, N. T.; Iverson, K. D.; Sadler, W. D.; Petrosino, J. F.; Chen, G. Y.; Schloss, P. D. The Gut Microbiome Modulates Colon Tumorigenesis. *MBio* **2013**, *4*, e00692-00613.

41. The Human Microbiome Consortium, Structure, Function and Diversity of the Healthy Human Microbiome. *Nature* **2012**, *486*, 207–214.

42. The Human Microbiome Consortium, A Framework for Human Microbiome Research. *Nature* **2012**, *486*, 215–221.

43. Mehta, R. S.; Abu-Ali, G. S.; Drew, D. A.; Lloyd-Price, J.; Subramanian, A.; Lochhead, P.; Joshi, A. D.; Ivey, K. L.; Khalili, H.; Brown, G. T.; DuLong, C.; Song, M.; Nguyen, L. H.; Mallick, H.; Rimm, E. B.; Izard, J.; Huttenhower, C.; Chan, A. T. Stability of the Human Faecal Microbiome in a Cohort of Adult Men. *Nat. Microbiol.* **2018**, *3*, 347–355.

44. McNulty, N. P.; Yatsunenko, T.; Hsiao, A.; Faith, J. J.; Muegge, B. D.; Goodman, A. L.; Henrissat, B.; Oozeer, R.; Cools-Portier, S.; Gobert, G.; Chervaux, C.; Knights, D.; Lozupone, C. A.; Knight, R.; Duncan, A. E.; Bain, J. R.; Muehlbauer, M. J.; Newgard, C. B.; Heath, A. C.; Gordon, J. I. The Impact of a Consortium of Fermented Milk Strains on the Gut Microbiome of Gnotobiotic Mice and Monozygotic Twins. *Sci. Transl. Med.* **2011**, *3*, 106ra106.

45. Ram, R. J.; Verberkmoes, N. C.; Thelen, M. P.; Tyson, G. W.; Baker, B. J.; Blake, R. C., 2nd; Shah, M.; Hettich, R. L.; Banfield, J. F. Community Proteomics of a Natural Microbial Biofilm. *Science* **2005**, *308*, 1915–1920.

46. Brown, C. T.; Xiong, W.; Olm, M. R.; Thomas, B. C.; Baker, R.; Firek, B.; Morowitz, M. J.; Hettich, R. L.; Banfield, J. F. Hospitalized Premature Infants Are Colonized by Related Bacterial Strains With Distinct Proteomic Profiles. *MBio* **2018**, *9*, e00441-18.

47. Bouslimani, A.; Porto, C.; Rath, C. M.; Wang, M.; Guo, Y.; Gonzalez, A.; Berg-Lyon, D.; Ackermann, G.; Moeller Christensen, G. J.; Nakatsuji, T.; Zhang, L.; Borkowski, A. W.; Meehan, M. J.; Dorrestein, K.; Gallo, R. L.; Bandeira, N.; Knight, R.; Alexandrov, T.; Dorrestein, P. C. Molecular Cartography of the Human Skin Surface in 3D. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, E2120–E2129.

48. Theriot, C. M.; Koenigsknecht, M. J.; Carlson, P. E., Jr.; Hatton, G. E.; Nelson, A. M.; Li, B.; Huffnagle, G. B.; J, Z. L.; Young, V. B. Antibiotic-Induced Shifts in the Mouse Gut Microbiome and Metabolome Increase Susceptibility to *Clostridium difficile* Infection. *Nat. Commun.* **2014**, *5*, 3114.

49. Zierer, J.; Jackson, M. A.; Kastenmuller, G.; Mangino, M.; Long, T.; Telenti, A.; Mohney, R. P.; Small, K. S.; Bell, J. T.; Steves, C. J.; Valdes, A. M.; Spector, T. D.; Menni, C. The Fecal Metabolome as a Functional Readout of the Gut Microbiome. *Nat. Genet.* **2018**, *50* (6), 790–795.

50. Peisl, B. Y. L.; Schymanski, E. L.; Wilmes, P. Dark Matter in Host-Microbiome Metabolomics: Tackling the Unknowns–A Review. *Anal. Chim. Acta* **2018**, *1037*, 13–27.

51. Schloss, P. D.; Westcott, S. L.; Ryabin, T.; Hall, J. R.; Hartmann, M.; Hollister, E. B.; Lesniewski, R. A.; Oakley, B. B.; Parks, D. H.; Robinson, C. J.; Sahl, J. W.; Stres, B.; Thallinger, G. G.; Van Horn, D. J.; Weber, C. F. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Appl. Environ. Microbiol.* **2009**, *75*, 7537–7541.

52. Caporaso, J. G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F. D.; Costello, E. K.; Fierer, N.; Pena, A. G.; Goodrich, J. K.; Gordon, J. I.; Huttley, G. A.; Kelley, S. T.; Knights, D.; Koenig, J. E.; Ley, R. E.; Lozupone, C. A.; McDonald, D.; Muegge, B. D.; Pirrung, M.; Reeder, J.; Sevinsky, J. R.; Turnbaugh, P. J.; Walters, W. A.; Widmann, J.; Yatsunenko, T.; Zaneveld, J.; Knight, R. QIIME Allows Analysis of High-Throughput Community Sequencing Data. *Nat. Methods* **2010**, *7*, 335–336.

53. Bode, H. B.; Bethe, B.; Hofs, R.; Zeeck, A. Big Effects From Small Changes: Possible Ways to Explore Nature's Chemical Diversity. *ChemBioChem* **2002**, *3*, 619–627.

54. 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.; Crusemann, M.; Boudreau, P. D.; Esquenazi, E.; Sandoval-Calderon, M.; Kersten, R. D.; Pace, L. A.; Quinn, R. A.; Duncan, K. R.; Hsu, C. C.; Floros, D. J.; Gavilan, R. G.; Kleigrewe, K.; Northen, T.; Dutton, R. J.; Parrot, D.; Carlson, E. E.; Aigle, B.; Michelsen, C. F.; Jelsbak, L.; Sohlenkamp, C.; Pevzner, P.; Edlund, A.; McLean, J.; Piel, J.; Murphy, B. T.; Gerwick, L.; Liaw, C. C.; Yang, Y. L.; Humpf, H. U.; Maansson, M.; Keyzers, R. A.; Sims, A. C.; Johnson, A. R.; Sidebottom, A. M.; Sedio, B. E.; Klitgaard, A.; Larson, C. B.; Boya, P. C.; Torres-Mendoza, D.; Gonzalez, D. J.; Silva, D. B.; Marques, L. M.; Demarque, D. P.; Pociute, E.; O'Neill, E. C.; Briand, E.; Helfrich, E. J.; Granatosky, E. A.; Glukhov, E.; Ryffel, F.; Houson, H.; Mohimani, H.; Kharbush, J. J.; Zeng, Y.; Vorholt, J. A.; Kurita, K. L.; Charusanti, P.; McPhail, K. L.; Nielsen, K. F.; Vuong, L.; Elfeki, M.; Traxler, M. F.; Engene, N.; Koyama, N.; Vining, O. B.; Baric, R.; Silva, R. R.; Mascuch, S. J.; Tomasi, S.; Jenkins, S.; Macherla, V.; Hoffman, T.; Agarwal, V.; Williams, P. G.; Dai, J.; Neupane, R.; Gurr, J.; Rodriguez, A. M.; Lamsa, A.; Zhang, C.; Dorrestein, K.; Duggan, B. M.; Almaliti, J.; Allard, P. M.; Phapale, P.; Nothias, L. F.; Alexandrov, T.; Litaudon, M.; Wolfender, J. L.; Kyle, J. E.; Metz, T. O.; Peryea, T.; Nguyen, D. T.; VanLeer, D.; Shinn, P.; Jadhav, A.; Muller, R.; Waters, K. M.; Shi, W.; Liu, X.; Zhang, L.; Knight, R.; Jensen, P. R.; Palsson, B. O.; Pogliano, K.; Linington, R. G.; Gutierrez, M.; Lopes, N. P.; Gerwick, W. H.; Moore, B. S.; Dorrestein, P. C.; Bandeira, N. Sharing and Community Curation of Mass Spectrometry Data With Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, *34*, 828–837.

55. Guijas, C.; Montenegro-Burke, J. R.; Domingo-Almenara, X.; Palermo, A.; Warth, B.; Hermann, G.; Koellensperger, G.; Huan, T.; Uritboonthai, W.; Aisporna, A. E.; Wolan, D. W.; Spilker, M. E.; Benton, H. P.; Siuzdak, G. METLIN: A Technology Platform for Identifying Knowns and Unknowns. *Anal. Chem.* **2018**, *90*, 3156–3164.

56. Wang, M.; Jarmusch, A. K.; Vargas, F.; Aksenov, A. A.; Gauglitz, J. M.; Weldon, K.; Petras, D.; Silva, R. D.; Quinn, R.; Melnik, A. V.; van der Hooft, J. J. J.; Caraballo Rodríguez, A. M.; Nothias, L. F.; Aceves, C. M.; Panitchpakdi, M.; Brown, E.; Di Ottavio, F.; Sikora, N.; Elijah, E. O.; Labarta-Bajo, L.; Gentry, E. C.; Shalapour, S.; Kyle, K. E.; Puckett, S. P.; Watrous, J. D.; Carpenter, C. S.; Bouslimani, A.; Ernst, M.; Swafford, A. D.; Zúñiga, E. I.; Balunas, M. J.; Klassen, J. L.; Loomba, R.; Knight, R.; Bandeira, N.; Dorrestein, P. C. MASST: A Web-Based Basic Mass Spectrometry Search Tool for Molecules to Search Public Data. *Nat. Biotechnol.* **2019**. in press.

57. Schmidt, E. W.; Donia, M. S. Life in Cellulose Houses: Symbiotic Bacterial Biosynthesis of Ascidian Drugs and Drug Leads. *Curr. Opin. Biotechnol.* **2010**, *21*, 827–833.

58. Ogai, K.; Nagase, S.; Mukai, K.; Iuchi, T.; Mori, Y.; Matsue, M.; Sugitani, K.; Sugama, J.; Okamoto, S. A Comparison of Techniques for Collecting Skin Microbiome Samples: Swabbing Versus Tape-Stripping. *Front. Microbiol.* **2018**, *9*, 2362.

59. Lin, L.; Zhuang, M.; Lei, F.; Yang, B.; Zhao, M. GC/MS Analysis of Volatiles Obtained by Headspace Solid-Phase Microextraction and Simultaneous-Distillation Extraction From *Rabdosia serra* (Maxim.) HARA Leaf and Stem. *Food Chem.* **2013**, *136*, 555–562.

60. Shimizu, Y.; Li, B. Purification of Water-Soluble Natural Products. In *Natural Products Isolation*; Sarker, S. D., Latif, Z., Gray, A. I., Eds.; Humana Press: Totowa, NJ, 2005;; pp 415–438.

61. Maltese, F.; van der Kooy, F.; Verpoorte, R. Solvent Derived Artifacts in Natural Products Chemistry. *Nat. Prod. Commun.* **2009**, *4*, 447–454.

62. Hough, R.; Archer, D.; Probert, C. A Comparison of Sample Preparation Methods for Extracting Volatile Organic Compounds (VOCs) From Equine Faeces Using HS-SPME. *Metabolomics* **2018**, *14*, 19.

63. Benda, N. D.; Boucias, D.; Torto, B.; Teal, P. Detection and Characterization of *Kodamaea ohmeri* Associated With Small Hive Beetle *Aethina tumida* Infesting Honey Bee Hives. *J. Apic. Res.* **2008**, *47*, 194–201.

64. Hu, A.; Noble, W.; Wolf-Yadlin, A. Technical Advances in Proteomics: New Developments in Data-Independent Acquisition. *F1000Res.* **2016**, *5*, 419.

65. Johnson, D.; Boyes, B.; Fields, T.; Kopkin, R.; Orlando, R. Optimization of Data-Dependent Acquisition Parameters for Coupling High-Speed Separations With LC-MS/MS for Protein Identifications. *J. Biomol. Tech.* **2013**, *24*, 62–72.

66. Hoffmann, T.; Krug, D.; Huttel, S.; Muller, R. Improving Natural Products Identification Through Targeted LC-MS/MS in an Untargeted Secondary Metabolomics Workflow. *Anal. Chem.* **2014**, *86*, 10780–10788.

67. Panter, F.; Krug, D.; Muller, R. Novel Methoxymethacrylate Natural Products Uncovered by Statistics-Based Mining of the *Myxococcus fulvus* Secondary Metabolome. *ACS Chem. Biol.* **2019**, *14*, 88–98.

68. De Vos, R. C.; Moco, S.; Lommen, A.; Keurentjes, J. J.; Bino, R. J.; Hall, R. D. Untargeted Large-Scale Plant Metabolomics Using Liquid Chromatography Coupled to Mass Spectrometry. *Nat. Protoc.* **2007**, *2*, 778–791.

69. Kessner, D.; Chambers, M.; Burke, R.; Agus, D.; Mallick, P. ProteoWizard: Open Source Software for Rapid Proteomics Tools Development. *Bioinformatics* **2008**, *24*, 2534–2536.

70. Martens, L.; Chambers, M.; Sturm, M.; Kessner, D.; Levander, F.; Shofstahl, J.; Tang, W. H.; Rompp, A.; Neumann, S.; Pizarro, A. D.; Montecchi-Palazzi, L.; Tasman, N.; Coleman, M.; Reisinger, F.; Souda, P.; Hermjakob, H.; Binz, P. A.; Deutsch, E. W. mzML—A Community Standard for Mass Spectrometry Data. *Mol. Cell. Proteomics* **2011**, *10*. R110.000133.

71. Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. Probability-Based Protein Identification by Searching Sequence Databases Using Mass Spectrometry Data. *Electrophoresis* **1999**, *20*, 3551–3567.

72. Kirchner, M.; Steen, J. A.; Hamprecht, F. A.; Steen, H. MGFp: An Open Mascot Generic Format Parser Library Implementation. *J. Proteome Res.* **2010**, *9*, 2762–2763.

73. Spicer, R.; Salek, R. M.; Moreno, P.; Canueto, D.; Steinbeck, C. Navigating Freely-Available Software Tools for Metabolomics Analysis. *Metabolomics* **2017**, *13*, 106.

74. da Silva, R. R.; Wang, M.; Nothias, L. F.; van der Hooft, J. J. J.; Caraballo-Rodriguez, A. M.; Fox, E.; Balunas, M. J.; Klassen, J. L.; Lopes, N. P.; Dorrestein, P. C. Propagating Annotations of Molecular Networks Using In Silico Fragmentation. *PLoS Comput. Biol.* **2018**, *14*, e1006089.

75. Godzien, J.; Gil de la Fuente, A.; Otero, A.; Barbas, C. Metabolite Annotation and Identification. In *Comprehensive Analytical Chemistry*; Jaumot, J., Bedia, C., Tauler, R., Eds.; 82; Elsevier, 2018; pp 415–445.

76. Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J. Identifying Small Molecules Via High Resolution Mass Spectrometry: Communicating Confidence. *Environ. Sci. Technol.* **2014**, *48*, 2097–2098.

77. Ōmura, S. *The Search for Bioactive Compounds From Microorganisms*; Springer-Verlag: New York, NY, 1992.

78. Johnson, S. R.; Lange, B. M. Open-Access Metabolomics Databases for Natural Product Research: Present Capabilities and Future Potential. *Front. Bioeng. Biotechnol.* **2015**, *3*, 00022.

79. Kanehisa, M.; Furumichi, M.; Tanabe, M.; Sato, Y.; Morishima, K. KEGG: New Perspectives on Genomes, Pathways, Diseases and Drugs. *Nucleic Acids Res.* **2017**, *45*, D353–d361.

80. Little, J. L.; Williams, A. J.; Pshenichnov, A.; Tkachenko, V. Identification of "Known Unknowns" Utilizing Accurate Mass Data and ChemSpider. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 179–185.

81. Klementz, D.; Doring, K.; Lucas, X.; Telukunta, K. K.; Erxleben, A.; Deubel, D.; Erber, A.; Santillana, I.; Thomas, O. S.; Bechthold, A.; Gunther, S. StreptomeDB 2.0—An Extended Resource of Natural Products Produced by *Streptomycetes. Nucleic Acids Res.* **2016**, *44*, D509–D514.

82. van Santen, J. A.; Jacob, G.; Singh, A. L.; Aniebok, V.; Balunas, M. J.; Bunsko, D.; Carnevale Neto, F.; Castaño-Espriu, L.; Chang, C.; Clark, T. N.; Clearly Little, J. L.; Delgadillo, D. A.; Dorrestein, P. C.; Duncan, K. R.; Egan, J. M.; Galey, M. M.; Haeckl, F. P. J.; Hua, A.; Hughes, A. H.; Iskakova, D.; Khadilkar, A.; Lee, J.-H.; Lee, S.; LeGrow, N.; Liu, D. Y.; Macho, J. M.; McCaughey, C. S.; Medema, M. H.; Neupane, R. P.; O'Donnell, T. J.; Paula, J. S.; Sanchez, L. M.; Shaikh, A. F.; Soldatou, S.; Terlouw, B. R.; Tran, T. A.; Valentine, M.; van der Hooft, J. J. J.; Vo, D. A.; Wang, M.; Wilson, D.; Zink, K. E.; Linington, R. G. The Natural Products Atlas: An Open Access Knowledge Base for Microbial Natural Products Discovery. *ACS Cent. Sci.* **2019**, *11*, 1824–1833.

83. Arakawa, K.; Kono, N.; Yamada, Y.; Mori, H.; Tomita, M. KEGG-Based Pathway Visualization Tool for Complex Omics Data. *In Silico Biol.* **2005**, *5*, 419–423.

84. Nakamura, Y.; Afendi, F. M.; Parvin, A. K.; Ono, N.; Tanaka, K.; Hirai Morita, A.; Sato, T.; Sugiura, T.; Altaf-Ul-Amin, M.; Kanaya, S. KNApSAcK Metabolite Activity Database for Retrieving the Relationships Between Metabolites and Biological Activities. *Plant Cell Physiol.* **2014**, *55*, e7.

85. Tsugawa, H. Advances in Computational Metabolomics and Databases Deepen the Understanding of Metabolisms. *Curr. Opin. Biotechnol.* **2018**, *54*, 10–17.

86. Moult, J.; Pedersen, J. T.; Judson, R.; Fidelis, K. A Large-Scale Experiment to Assess Protein Structure Prediction Methods. *Proteins* **1995**, *23*, ii-v.

87. Schymanski, E. L.; Neumann, S. CASMI: And the Winner Is. *Metabolites* **2013**, *3*, 412–439.

88. Allwood, J. W.; Weber, R. J.; Zhou, J.; He, S.; Viant, M. R.; Dunn, W. B. CASMI-the Small Molecule Identification Process From a Birmingham Perspective. *Metabolites* **2013**, *3*, 397–411.

89. Schymanski, E. L.; Ruttkies, C.; Krauss, M.; Brouard, C.; Kind, T.; Duhrkop, K.; Allen, F.; Vaniya, A.; Verdegem, D.; Bocker, S.; Rousu, J.; Shen, H.; Tsugawa, H.; Sajed, T.; Fiehn, O.; Ghesquiere, B.; Neumann, S. Critical Assessment of Small Molecule Identification 2016: Automated Methods. *Aust. J. Chem.* **2017**, *9*, 22.

90. Yang, J. Y.; Sanchez, L. M.; Rath, C. M.; Liu, X.; Boudreau, P. D.; Bruns, N.; Glukhov, E.; Wodtke, A.; de Felicio, R.; Fenner, A.; Wong, W. R.; Linington, R. G.; Zhang, L.; Debonsi, H. M.; Gerwick, W. H.; Dorrestein, P. C. Molecular Networking as a Dereplication Strategy. *J. Nat. Prod.* **2013**, *76*, 1686–1699.

91. Olivon, F.; Grelier, G.; Roussi, F.; Litaudon, M.; Touboul, D. MZmine 2 Data-Preprocessing to Enhance Molecular Networking Reliability. *Anal. Chem.* **2017**, *89*, 7836–7840.

92. Saccenti, E.; Hoefsloot, H. C. J.; Smilde, A. K.; Westerhuis, J. A.; Hendriks, M. M. W. B. Reflections on Univariate and Multivariate Analysis of Metabolomics Data. *Metabolomics* **2014**, *10*, 361–374.

93. Xiao, J. F.; Zhou, B.; Ressom, H. W. Metabolite Identification and Quantitation in LC-MS/MS-Based Metabolomics. *TrAC Trends Anal. Chem.* **2012**, *32*, 1–14.

94. Kumar, A.; Misra, B. B. Challenges and Opportunities in Cancer Metabolomics. *Proteomics* **2019**, *19* (21–22), e1900042.

95. Vidova, V.; Spacil, Z. A Review on Mass Spectrometry-Based Quantitative Proteomics: Targeted and Data Independent Acquisition. *Anal. Chim. Acta* **2017**, *964*, 7–23.

96. Broadhurst, D. I.; Kell, D. B. Statistical Strategies for Avoiding False Discoveries in Metabolomics and Related Experiments. *Metabolomics* **2006**, *2*, 171–196.

97. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Series B Stat. Methodology* **1995**, *57*, 289–300.

98. Simes, R. J. An Improved Bonferroni Procedure for Multiple Tests of Significance. *Biometrika* **1986**, *73*, 751–754.

99. Jones, C. G.; Martynowycz, M. W.; Hattne, J.; Fulton, T. J.; Stoltz, B. M.; Rodriguez, J. A.; Nelson, H. M.; Gonen, T. The CryoEM Method microED as a Powerful Tool for Small Molecule Structure Determination. *ACS Cent. Sci.* **2018**, *4*, 1587–1592.

100. Adusumilli, R.; Mallick, P. Data Conversion With ProteoWizard msConvert. *Methods Mol. Biol.* **2017**, *1550*, 339–368.

101. Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; Hoff, K.; Kessner, D.; Tasman, N.; Shulman, N.; Frewen, B.; Baker, T. A.; Brusniak, M. Y.; Paulse, C.; Creasy, D.; Flashner, L.; Kani, K.; Moulding, C.; Seymour, S. L.; Nuwaysir, L. M.; Lefebvre, B.; Kuhlmann, F.; Roark, J.; Rainer, P.; Detlev, S.; Hemenway, T.; Huhmer, A.; Langridge, J.; Connolly, B.; Chadick, T.; Holly, K.; Eckels, J.; Deutsch, E. W.; Moritz, R. L.; Katz, J. E.; Agus, D. B.; MacCoss, M.; Tabb, D. L.; Mallick, P. A Cross-Platform Toolkit for Mass Spectrometry and Proteomics. *Nat. Biotechnol.* **2012**, *30*, 918–920.

102. Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Anal. Chem.* **2006**, *78*, 779–787.

103. Tautenhahn, R.; Patti, G. J.; Rinehart, D.; Siuzdak, G. XCMS Online: A Web-Based Platform to Process Untargeted Metabolomic Data. *Anal. Chem.* **2012**, *84*, 5035–5039.

104. Gowda, H.; Ivanisevic, J.; Johnson, C. H.; Kurczy, M. E.; Benton, H. P.; Rinehart, D.; Nguyen, T.; Ray, J.; Kuehl, J.; Arevalo, B.; Westenskow, P. D.; Wang, J.; Arkin, A. P.; Deutschbauer, A. M.; Patti, G. J.; Siuzdak, G. Interactive XCMS Online: Simplifying Advanced Metabolomic Data Processing and Subsequent Statistical Analyses. *Anal. Chem.* **2014**, *86*, 6931–6939.

105. Domingo-Almenara, X.; Montenegro-Burke, J. R.; Ivanisevic, J.; Thomas, A.; Sidibé, J.; Teav, T.; Guijas, C.; Aisporna, A. E.; Rinehart, D.; Hoang, L.; Nordström, A.; Gómez-Romero, M.; Whiley, L.; Lewis, M. R.; Nicholson, J. K.; Benton, H. P.; Siuzdak, G. XCMS-MRM and METLIN-MRM: A Cloud Library and Public Resource for Targeted Analysis of Small Molecules. *Nat. Methods* **2018**, *15*, 681–684.

106. Huan, T.; Forsberg, E. M.; Rinehart, D.; Johnson, C. H.; Ivanisevic, J.; Benton, H. P.; Fang, M.; Aisporna, A.; Hilmers, B.; Poole, F. L.; Thorgersen, M. P.; Adams, M. W. W.; Krantz, G.; Fields, M. W.; Robbins, P. D.; Niedernhofer, L. J.; Ideker, T.; Majumder, E. L.; Wall, J. D.; Rattray, N. J. W.; Goodacre, R.; Lairson, L. L.; Siuzdak, G. Systems Biology Guided by XCMS Online Metabolomics. *Nat. Methods* **2017**, *14*, 461–462.

107. Forsberg, E. M.; Huan, T.; Rinehart, D.; Benton, H. P.; Warth, B.; Hilmers, B.; Siuzdak, G. Data Processing, Multi-Omic Pathway Mapping, and Metabolite Activity Analysis Using XCMS Online. *Nat. Protoc.* **2018**, *13*, 633–651.

108. Cheng, Z.; Li, M.; Marriott, P. J.; Zhang, X.; Wang, S.; Li, J.; Ma, L. Chemometric Analysis of the Volatile Compounds Generated by *Aspergillus carbonarius* Strains Isolated From Grapes and Dried Vine Fruits. *Toxins (Basel)* **2018**, *10*, E71.

109. Combes, A.; Ndoye, I.; Bance, C.; Bruzaud, J.; Djediat, C.; Dupont, J.; Nay, B.; Prado, S. Chemical Communication Between the Endophytic Fungus *Paraconiothyrium variabile* and the Phytopathogen *Fusarium oxysporum*. *PLoS One* **2012**, *7*, e47313.

110. Xia, J.; Psychogios, N.; Young, N.; Wishart, D. S. MetaboAnalyst: A Web Server for Metabolomic Data Analysis and Interpretation. *Nucleic Acids Res.* **2009**, *37*, W652–W660.

111. Chong, J.; Soufan, O.; Li, C.; Caraus, I.; Li, S.; Bourque, G.; Wishart, D. S.; Xia, J. MetaboAnalyst 4.0: Towards More Transparent and Integrative Metabolomics Analysis. *Nucleic Acids Res.* **2018**, *46*, W486–w494.

112. Triastuti, A.; Vansteelandt, M.; Barakat, F.; Trinel, M.; Jargeat, P.; Fabre, N.; Amasifuen Guerra, C. A.; Mejia, K.; Valentin, A.; Haddad, M. How Histone Deacetylase Inhibitors Alter the Secondary Metabolites of *Botryosphaeria mamane*, an Endophytic Fungus Isolated From *Bixa orellana*. *Chem. Biodivers.* **2019**, *16*, e1800485.

113. Watkins, J.; Murphy, K.; Yen, S.; Carafa, I.; Dempsey, E. M.; O'Shea, C. A.; Vercoe, E. A.; Ross, R. P.; Stanton, C.; Ryan, C. A. Effects of Therapeutic Hypothermia on the Gut Microbiota and Metabolome of Infants Suffering Hypoxic-Ischemic Encephalopathy at Birth. *Int. J. Biochem. Cell Biol.* **2017**, *93*, 110–118.

114. Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresic, M. MZmine 2: Modular Framework for Processing, Visualizing, and Analyzing Mass Spectrometry-Based Molecular Profile Data. *BMC Bioinform.* **2010**, *11*, 395.

115. Myers, O. D.; Sumner, S. J.; Li, S.; Barnes, S.; Du, X. Detailed Investigation and Comparison of the XCMS and MZmine 2 Chromatogram Construction and Chromatographic Peak Detection Methods for Preprocessing Mass Spectrometry Metabolomics Data. *Anal. Chem.* **2017**, *89*, 8689–8695.

116. Wishart, D. S.; Feunang, Y. D.; Marcu, A.; Guo, A. C.; Liang, K.; Vazquez-Fresno, R.; Sajed, T.; Johnson, D.; Li, C.; Karu, N.; Sayeeda, Z.; Lo, E.; Assempour, N.; Berjanskii, M.; Singhal, S.; Arndt, D.; Liang, Y.; Badran, H.; Grant, J.; Serra-Cayuela, A.; Liu, Y.; Mandal, R.; Neveu, V.; Pon, A.; Knox, C.; Wilson, M.; Manach, C.; Scalbert, A. HMDB 4.0: The Human Metabolome Database for 2018. *Nucleic Acids Res.* **2018**, *46*, D608–d617.

117. Kwan, J. C.; Tianero, M. D.; Donia, M. S.; Wyche, T. P.; Bugni, T. S.; Schmidt, E. W. Host Control of Symbiont Natural Product Chemistry in Cryptic Populations of the Tunicate *Lissoclinum patella*. *PLoS One* **2014**, *9*, e95850.

118. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13*, 2498–2504.

119. Boya, P. C.; Fernandez-Marin, H.; Mejia, L. C.; Spadafora, C.; Dorrestein, P. C.; Gutierrez, M. Imaging Mass Spectrometry and MS/MS Molecular Networking Reveals Chemical Interactions Among Cuticular Bacteria and Pathogenic Fungi Associated With Fungus-Growing Ants. *Sci. Rep.* **2017**, *7*, 5604.

120. Watrous, J.; Roach, P.; Alexandrov, T.; Heath, B. S.; Yang, J. Y.; Kersten, R. D.; van der Voort, M.; Pogliano, K.; Gross, H.; Raaijmakers, J. M.; Moore, B. S.; Laskin, J.; Bandeira, N.; Dorrestein, P. C. Mass Spectral Molecular Networking of Living Microbial Colonies. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, E1743–E1752.

121. Crusemann, M.; O'Neill, E. C.; Larson, C. B.; Melnik, A. V.; Floros, D. J.; da Silva, R. R.; Jensen, P. R.; Dorrestein, P. C.; Moore, B. S. Prioritizing Natural Product Diversity in a Collection of 146 Bacterial Strains Based on Growth and Extraction Protocols. *J. Nat. Prod.* **2017**, *80*, 588–597.

122. Nguyen, D. D.; Wu, C. H.; Moree, W. J.; Lamsa, A.; Medema, M. H.; Zhao, X.; Gavilan, R. G.; Aparicio, M.; Atencio, L.; Jackson, C.; Ballesteros, J.; Sanchez, J.; Watrous, J. D.; Phelan, V. V.; van de Wiel, C.; Kersten, R. D.; Mehnaz, S.; De Mot, R.; Shank, E. A.; Charusanti, P.; Nagarajan, H.; Duggan, B. M.; Moore, B. S.; Bandeira, N.; Palsson, B. O.; Pogliano, K.; Gutierrez, M.; Dorrestein, P. C. MS/MS Networking Guided Analysis of Molecule and Gene Cluster Families. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, E2611–E2620.

123. Kleigrewe, K.; Almaliti, J.; Tian, I. Y.; Kinnel, R. B.; Korobeynikov, A.; Monroe, E. A.; Duggan, B. M.; Di Marzo, V.; Sherman, D. H.; Dorrestein, P. C.; Gerwick, L.; Gerwick, W. H. Combining Mass Spectrometric Metabolic Profiling with Genomic Analysis: A Powerful Approach for Discovering Natural Products From Cyanobacteria. *J. Nat. Prod.* **2015**, *78* (7), 1671–1682.

124. Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; Oda, Y.; Kakazu, Y.; Kusano, M.; Tohge, T.; Matsuda, F.; Sawada, Y.; Hirai, M. Y.; Nakanishi, H.; Ikeda, K.; Akimoto, N.; Maoka, T.; Takahashi, H.; Ara, T.; Sakurai, N.; Suzuki, H.; Shibata, D.; Neumann, S.; Iida, T.; Tanaka, K.; Funatsu, K.; Matsuura, F.; Soga, T.; Taguchi, R.; Saito, K.; Nishioka, T. MassBank: A Public Repository for Sharing Mass Spectral Data for Life Sciences. *J. Mass Spectrom.* **2010**, *45*, 703–714.

125. Mohimani, H.; Gurevich, A.; Mikheenko, A.; Garg, N.; Nothias, L. F.; Ninomiya, A.; Takada, K.; Dorrestein, P. C.; Pevzner, P. A. Dereplication of Peptidic Natural Products Through Database Search of Mass Spectra. *Nat. Chem. Biol.* **2017**, *13*, 30–37.

126. Mohimani, H.; Gurevich, A.; Shlemov, A.; Mikheenko, A.; Korobeynikov, A.; Cao, L.; Shcherbin, E.; Nothias, L. F.; Dorrestein, P. C.; Pevzner, P. A. Dereplication of Microbial Metabolites Through Database Search of Mass Spectra. *Nat. Commun.* **2018**, *9,* 4035.

127. Edlund, A.; Garg, N.; Mohimani, H.; Gurevich, A.; He, X.; Shi, W.; Dorrestein, P. C.; McLean, J. S. Metabolic Fingerprints From the Human Oral Microbiome Reveal a Vast Knowledge Gap of Secreted Small Peptidic Molecules. *mSystems* **2017**, *2,* e00058-17.

128. Garg, N.; Kapono, C.; Lim, Y. W.; Koyama, N.; Vermeij, M. J.; Conrad, D.; Rohwer, F.; Dorrestein, P. C. Mass Spectral Similarity for Untargeted Metabolomics Data Analysis of Complex Mixtures. *Int. J. Mass Spectrom.* **2015**, *377*, 717–719.

129. Garg, N.; Wang, M.; Hyde, E.; da Silva, R. R.; Melnik, A. V.; Protsyuk, I.; Bouslimani, A.; Lim, Y. W.; Wong, R.; Humphrey, G.; Ackermann, G.; Spivey, T.; Brouha, S. S.; Bandeira, N.; Lin, G. Y.; Rohwer, F.; Conrad, D. J.; Alexandrov, T.; Knight, R.; Dorrestein, P. C. Three-Dimensional Microbiome and Metabolome Cartography of a Diseased Human Lung. *Cell Host Microbe* **2017**, *22*, 705–716.e704.

130. Duhrkop, K.; Fleischauer, M.; Ludwig, M.; Aksenov, A. A.; Melnik, A. V.; Meusel, M.; Dorrestein, P. C.; Rousu, J.; Bocker, S. SIRIUS 4: A Rapid Tool for Turning Tandem Mass Spectra into Metabolite Structure Information. *Nat. Methods* **2019**, *16*, 299–302.

131. Cleary, J. L.; Luu, G. T.; Pierce, E. C.; Dutton, R. J.; Sanchez, L. M. BLANKA: An Algorithm for Blank Subtraction in Mass Spectrometry of Complex Biological Samples. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 1426–1434.

132. Hover, B. M.; Kim, S. H.; Katz, M.; Charlop-Powers, Z.; Owen, J. G.; Ternei, M. A.; Maniko, J.; Estrela, A. B.; Molina, H.; Park, S.; Perlin, D. S.; Brady, S. F. Culture-Independent Discovery of the Malacidins as Calcium-Dependent Antibiotics With Activity Against Multidrug-Resistant Gram-Positive Pathogens. *Nat. Microbiol.* **2018**, *3*, 415–422.

133. Katz, M.; Hover, B. M.; Brady, S. F. Culture-Independent Discovery of Natural Products From Soil Metagenomes. *J. Ind. Microbiol. Biotechnol.* **2016**, *43*, 129–141.

134. Piel, J. Bacterial Symbionts: Prospects for the Sustainable Production of Invertebrate-Derived Pharmaceuticals. *Curr. Med. Chem.* **2006**, *13*, 39–50.

135. Piddock, L. J. Teixobactin, the First of a New Class of Antibiotics Discovered by iChip Technology? *J. Antimicrob. Chemother.* **2015**, *70*, 2679–2680.

136. Tomm, H. A.; Ucciferri, L.; Ross, A. C. Advances in Microbial Culturing Conditions to Activate Silent Biosynthetic Gene Clusters for Novel Metabolite Production. *J. Ind. Microbiol. Biotechnol.* **2019**, *46* (9–10), 1381–1400.

137. Wissenbach, D. K.; Oliphant, K.; Rolle-Kampczyk, U.; Yen, S.; Hoke, H.; Baumann, S.; Haange, S. B.; Verdu, E. F.; Allen-Vercoe, E.; von Bergen, M. Optimization of Metabolomics of Defined In Vitro Gut Microbial Ecosystems. *Int. J. Med. Microbiol.* **2016**, *306*, 280–289.