



Published in final edited form as:

Nat Prod Rep. 2017 January 04; 34(1): 6–24. doi:10.1039/c6np00048g.

Comparative mass spectrometry-based metabolomics strategies for the investigation of microbial secondary metabolites

Brett C. Covington^a, John A. McLean^{a,b}, and Brian O. Bachmann^a

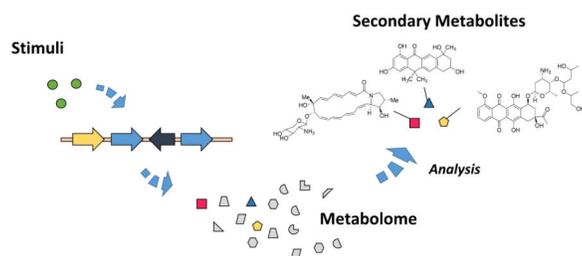
^a Department of Chemistry, Vanderbilt University, 7330 Stevenson Center, Nashville, TN 37235

^b Center for Innovative Technology, Vanderbilt University, 5401 Stevenson Center, Nashville, TN 37235

Abstract

The labor-intensive process of microbial natural product discovery is contingent upon identifying discrete secondary metabolites of interest within complex biological extracts, which contain inventories of all extractable small molecules produced by an organism or consortium. Historically, compound isolation prioritization has been driven by observed biological activity and/or relative metabolite abundance and followed by dereplication via accurate mass analysis. Decades of discovery using variants of these methods has generated the natural pharmacopeia but also contributes to recent high rediscovery rates. However, genomic sequencing reveals substantial untapped potential in previously mined organisms, and can provide useful prescience of potentially new secondary metabolites that ultimately enables isolation. Recently, advances in comparative metabolomics analyses have been coupled to secondary metabolic predictions to accelerate bioactivity and abundance-independent discovery work flows. In this review we will discuss the various analytical and computational techniques that enable MS-based metabolomic applications to natural product discovery and discuss the future prospects for comparative metabolomics in natural product discovery.

Graphical abstract



1 Introduction

Genomic sequencing of both cultivated microorganisms and uncultivated microbiomes has revealed that most of the biosynthetic potential of microorganisms remains inaccessible to date. Even genomes of extensively studied microorganisms contain a large fraction of secondary metabolite biosynthetic gene clusters for which natural products have not been identified.¹⁻³ It is now estimated that the products of greater than 90% of secondary metabolite gene clusters are either not expressed under standard laboratory growth

conditions^{4,5} and/or their products are difficult to identify within extracted metabolomes. Secondary metabolites play crucial roles in the chemical ecology of their producing organisms,⁶⁻⁸ and these roles often correlate translationally into applications in human medicine.⁹ Therefore, solving the linked problems of secondary metabolite gene expression and the identification of secondary metabolites within metabolomic inventories have become central efforts in the field of natural product discovery.

Natural product biosynthetic potential can be rapidly estimated from genomic sequence data via automated bioinformatics platforms capable of comparing sequenced biosynthetic gene clusters to previously sequenced microorganisms and inferring putative structures of natural products by biosynthetic inference.¹⁰⁻¹⁶ Recently, several reviews have described evolving computational tools for biosynthetic gene cluster analysis.^{17,18} Increasingly sophisticated methodologies have been developed to tackle the biosynthetic gene cluster expression problem, which may be subdivided into heterologous and native approaches. Heterologous strategies endeavour to recapitulate functional secondary metabolic biosynthetic gene clusters in surrogate producers. Gene clusters may be cloned, and/or synthesized and refactored into alternate organisms with the aim of detecting newly produced metabolites in comparison to a clean background.¹⁹⁻²⁴ The success of heterologous expression is dependent upon functional expression within the host organism, which is a function of successful transcription, translation, and precursor availability as discussed in other reviews.^{19,25,26} Given the phylogenetic diversity of microbial secondary metabolite producers, a number of hurdles must be addressed to successfully identify constructs for functional expression. In addition to optimizing genetic regulatory elements for heterologous expression, differences in protein stability, post-translational modification of biosynthetic enzymes, and precursor availability must be addressed. Alternatively, native expression methods endeavour to activate secondary metabolite production from within the native producer.^{4,27,28}

In addition to refactoring biosynthetic gene clusters via genome editing, many chemical and biological stimuli have been reported over the past few decades that activate secondary metabolite expression. The practice of exposing microorganisms to an array of growth conditions for the purpose of eliciting the production of multiple compounds is not a recent development.²⁹ However, contemporary studies of the impact of stimuli on microbial metabolism, in addition to fulfilling to goals of natural product discovery, now model the chemical ecology and environmental microbiology of microorganisms.^{4,27,30-32} For example, microbial secondary metabolite producers have shown responses to subinhibitory antibiotic exposure^{33,34} as well as vertically acquired antibiotic resistance mutations which engender mutations in transcription and translation machinery³⁵⁻⁴⁰ have been demonstrated to activate the production of a fraction of previously undiscovered metabolites. Similarly, rare earth metal exposure,⁴¹⁻⁴³ which may affect circulating levels of pleiotropic factors, has been demonstrated to modulate secondary metabolite production. A particularly successful strategy for activating secondary metabolite expression in microbes is via stimulation with competing organisms.⁴⁴⁻⁴⁸ Taken together these phenomenon suggest that secondary metabolites are indeed produced by microorganisms to respond to environmental stimuli⁴⁵ and this is supported by the apparent biosynthetic gene cluster activation selectivity of various stimulatory methods as well as the intrinsically complex nature of secondary metabolite gene cluster regulation.^{1,49}

Regardless, all categories of genome-prioritized natural product discovery require a means of measuring the modulation of secondary metabolite production within the extracted metabolome of native or heterologous producers, and metabolomics methods have been continually adapted to this task. Metabolomics is often defined as the comprehensive study of small molecules within a biological system and provides a direct measure of detectable secondary metabolite production within an organism of interest. Currently there are two analytical platforms to facilitate metabolome profiling for natural product discovery. Nuclear magnetic resonance (NMR) based metabolomic analyses, reviewed elsewhere,⁵⁰ are not biased by molecular class and provide enhanced structural information for metabolites but are limited by the inherently low sensitivity of NMR. In contrast, the metabolomic analyses through mass spectrometry (MS), which will be the focus of our review, are exceptionally sensitive but are exclusively biased towards ionisable metabolites. The structural diversity of secondary metabolites, which span a broad range of functionality, molecular weight, and ionization efficiency, renders comprehensive detection of all metabolites through MS a challenging endeavour, and there is no universal approach for bioanalytical detection. For this reason, the development of metabolomics methods with MS strategies necessitates a discussion of contemporary practices and advances in analytical instrumentation.

As the product of the central dogma, the metabolome also contains information regarding a wide variety of cellular processes unrelated, or indirectly related to secondary metabolism. Correspondingly, metabolomics information may encode insights into how secondary metabolite producing organisms respond to chemical and biological stimuli and may also provide a means of investigating the biological mechanisms of newly isolated natural products, antibiotics, and chemotherapeutics, from the metabolomic changes engendered within treated organisms. Secondary metabolites are generally biosynthetic end-products and unlike primary metabolites, they accumulate at higher levels than the fluxes observed in central metabolism.⁵¹ Hence, comparatively abundant secondary metabolites are well suited for comparative metabolomics work-flows. Other recent reviews have highlighted some applications of mass spectrometry for the discovery of natural products deriving from plant⁵² and microbial⁵³ sources. In this review we provide a foundational overview of the analytical techniques that underlie MS-based metabolomic applications to natural product discovery and describe how these various techniques provide differentiating molecular characteristics for detected metabolites. We discuss the computational methods used to process complex metabolomics data and bioinformatics methods that utilize the molecular characteristics of detected metabolites to prioritize and dereplicate leads for natural product discovery. Lastly, we describe how these metabolomic methods are being applied to investigate biological activities for natural products and discuss future prospects for the field.

2 Methods of generating inventories of microbial metabolites

A variety of MS techniques are available to acquire metabolomics data with corresponding advantages and challenges depending on the analytical descriptor(s) that is/are desired. In each method, the end result of the analysis is a set of metabolomic 'features', ions with a determined mass-to-charge ratio (m/z) and potentially additional descriptive information. This additional information may include descriptors such as mass accuracy, chromatographic retention time, isotopic envelope, size and shape information,

fragmentation data, and topological distribution, among others. A summary of key descriptors and the information they provide in secondary metabolite characterization is provided in Table 1. As the dimensionality of feature characterization has an impact upon subsequent effectiveness of comparative metabolomics analyses, we will briefly discuss in this section commonly utilized methods for MS acquisition and highlight several of these key molecular characteristics that can be obtained with MS.

2.1 Mass measurement accuracy

The mass-to-charge ratio (m/z) of detected metabolites is the most useful property used to initiate the process of dereplication. For more than a decade mass analyzers have been able to determine mass accuracy with an error of under 1 ppm.⁵⁴⁻⁵⁷ This level of mass accuracy allows for the determination of elemental composition boundaries for compounds under 600 Da^{58,59} when coupled with isotopic mass ratios.⁶⁰ While advances in Fourier Transform ion cyclotron resonance MS (FTICRMS) can now routinely perform at sub-ppm mass errors, typical instrumentation provides mass errors in the range of 1 to 10 ppm (e.g. time-of-flight MS). Unfortunately, this alone is insufficient to confidently dereplicate features, because of the extensive number of potential isomers for a given elemental composition.⁶¹ Early compound dereplication is thereby often dependent on obtaining additional distinguishing characteristics such as those listed in Table 1, or via additional characteristic such as UV/Vis spectrum and biological activity.^{62,63} It is also noteworthy that MS analysis is predicated on the ability to generate ions of the species of interest. Neutral or poorly ionizing species are transparent to MS, and because of this the number of detectable compounds from a metabolomic extract will vary depending on the analytical methods used during acquisition, in particular the specific ionization source and ionization conditions that are used.

2.2 Isotopic modeling

The isotopic envelope, comprised of both the major and minor isotopic contributions to the elemental formula, provides several opportunities for enhanced characterization information,⁶⁴ including: (i) the presence of heteroatoms,⁶⁵ and (ii) isotopic enrichment strategies for relative and absolute quantitation of the abundance of the secondary metabolite.⁶⁶ The MS analysis of most biological molecules is typically concerning elemental formula comprising C, H, O, and N. The shared characteristic of these elements is that the monoisotopic peak also corresponds to the lowest mass isotope and thus, the lowest mass peak in the envelope is also the highest abundance. However, the vast majority of the periodic table is characterized by isotopic abundances that are somewhat varied from lightest to heaviest mass isotope and their isotopic signatures are oftentimes used in MS-based atomic analyses for identification purposes.⁶⁷ The presence of heteroatoms, such as chlorine or bromine, are readily discernable in their contribution to the isotopic abundance observed for secondary metabolites and their stoichiometric contribution can be quickly verified through the use of isotopic calculator algorithms.⁶⁸ Furthermore, these approaches are equally well suited by the addition of non-natural isotopic enrichment or depletion for determining the relative or absolute abundance of the secondary metabolite that is expressed. One such approach termed stable isotope labeling in cell culture (SILAC) has been demonstrated as a facile tool for incorporating enrichment or depletion in experimental protocols.⁶⁹ Finally, genomic-based structural predictions, implying biosynthetic precursors,

can be combined with stable isotope studies to identify targeted metabolites within organisms.⁷⁰

2.3 Chromatographic retention time

Liquid chromatography (LC) is one of the most commonly used approaches to separate individual constituents of complex natural product extracts, and various LC methods and their applications have been previously reviewed.⁷¹⁻⁷³ For natural product separations, reversed phase LC, and hydrophobic interaction chromatography are most commonly employed with a water-acetonitrile, or water-methanol gradient.⁷⁴ This is typically performed on the basis of hydrophobicity, where reversed phase LC (RPLC) and hydrophobic interaction chromatography (HILIC) are most commonly utilized,⁷⁵ and column retention will be affected by the ionization of these groups. Mobile phase pH can thereby significantly affect the separation efficiency for natural product extracts. Due to the dependence of compound retention on pH, and to assist ionization, mobile phases are commonly buffered with either acetic acid, trifluoroacetic acid, or formic acid⁷⁶ to protonate acidic sites and facilitate retention. However, as low pH may suppress detection of negatively charged species in switched scanning modalities, neutral volatile buffers are often preferred.

Liquid chromatography-mass spectrometry (LC-MS) acquisition can take minutes to hours per chromatographic separation, and environmental changes throughout the course of the sample set (column conditioning, instrumental sensitivity and accuracy drift, etc.) can affect the quality of the data. Consequently, for multiple extract samples analysed in a sequential fashion, conditional changes between the start and end of analysis could lead to significant artefactual differences in group metabolomes, which complicate interpretation of subsequent comparative analyses. While challenging, recent reviews have outlined metabolomic experimental design strategies to accommodate these technical problems.^{77,78}

2.4 Size and shape by ion mobility

Additional metabolomic feature information can be obtained by using gas-phase ion mobility-MS (IM-MS), without significantly increasing analysis time over MS-alone.⁷⁹ The mechanism and utility of IM-MS has been the topic of several recent reviews.⁷⁹⁻⁸¹ Briefly, in time-dispersive IM-MS, a uniform weak electric field is applied to a post-ionization ion drift tube containing an inert gas, where the ion velocity through the chamber is dependent upon thermal collisions with the background gas and its charge state.⁸² The number of collisions ions make as they traverse the drift cell are proportional to their collision cross-sectional area, providing distinguishing information regarding an ion's shape and/or conformation in the gas phase.⁸³ The separations in IM are very low energy in comparison with collisions used for fragmentation analysis, where in IM the ions experience approximately 10^4 to 10^6 collisions across a size separation versus 1 or several high energy collisions in collision induced dissociation (CID), respectively. Typical drift tube resolving power of IM-MS is sufficient such that conformationally restricted or extended metabolites, such as cyclic peptides, polycyclic polyketides, and polyenes often possess distinct ion mobility profiles that are obtained over the course of micro to milliseconds. IM-MS is often coupled with time-of-flight (TOF) MS that can rapidly acquire the m/z ratios for ions eluting from the IM-MS cell in a few microseconds. The frequency of data collection allows for

sufficient time sampling across chromatographic peaks, which occur over the course of minutes, to be coupled to IM-TOFMS.^{79,84} When applied to microbial metabolomics the enhanced separation and sensitivity provided by IM-MS has been beneficial for identifying known secondary metabolites, dereplication, and prioritizing features. Our laboratory has previously used IM-MS to help obtain high quality fragmentation data (IM-MS/MS) for all detected ions from crude extracts while comparing the differences between antibiotic resistant and wild-type *Nocardioopsis*.⁴⁰ This facilitated the putative identification of several metabolites over-produced in mutant strains. IM-MS has been applied to differentiate halogenated natural products in cyanobacteria⁸⁵ as well as peptide natural products from cave actinomycetes.⁸⁶ IM-MS has also been applied to investigate the 3-dimensional structures of lasso peptides, interlocked microbial peptides with a range of bioactivities,⁸⁷ and this will likely find other useful applications to natural product discovery as the technology becomes more widely available.

2.5 Ion fragmentation for structural information

Both time-dispersive (e.g. TOFMS) and scanning mass spectrometers (e.g. liner quadrupole MS and ion trap MS) can be used to acquire both precursor and fragment ion information (i.e. tandem MS),⁸⁸ which can provide a wealth of highly specific structural information that can be used to help identify and dereplicate metabolites.⁸⁹⁻⁹³ In metabolomics-driven natural product discovery workflows, fragmentation data is commonly collected via an automated data-dependent acquisition method in which the most abundant ions within a scanning cycle are automatically selected for fragmentation. Fragmentation data analysis facilitates natural product dereplication which, as will be discussed in more detail below, is a critical step in the process of natural product discovery.⁹⁴ There are a variety of methods applied to activate and induce dissociation of target ions, primarily categorized on the basis of how the ion is activated, collisionally, electron attachment, or through photon absorption, where the observed fragment ions will vary based on the method and parameters selected for fragmentation. For small molecules, collision induced dissociation (CID),⁹⁵ and surface induced dissociation (SID)⁹⁶ are commonly utilized. The degree of fragmentation observed using these methods depends on the number and degree of scissile bonds within a given molecule as well as the resulting internal ion energies used for analysis. In automated data-dependent tandem mass spectrometric fragmentation analysis, a given single set of dissociation parameters may not be appropriate for every feature of interest within a sample, requiring multiple experiments to determine optimal fragmentation parameters, and to effectively capture fragmentation data for a broad cross section of molecular classes. Ultimately, these methods provide characteristic fragmentation spectra that can be compared to established libraries of secondary metabolite fragmentation data to identify known secondary metabolites within the experimental sample.⁶³ Additionally, tandem mass spectrometric data are useful for elucidating the structures of peptide natural products and have been used in 'peptidogenomics' strategies to link ribosomal and non-ribosomal peptide natural products to their cognate biosynthetic gene clusters.⁹⁷⁻⁹⁹

2.6 Leveraging spatiotemporal metabolomics inventories to capture inter-organism interactions

Secondary metabolite producing microorganisms can be cultivated on agar medium¹⁰⁰⁻¹⁰² or in planktonic liquid¹⁰³⁻¹⁰⁹ culture medium, and several methods have been developed to extract and chromatographically separate resulting metabolomes.¹¹⁰⁻¹¹² However, microorganisms cultivated as monocultures or mixed cultures on agar may display planar metabolite distributions containing valuable information about chemical pleiotropism, nutrient dependence, and chemical ecology,¹¹³⁻¹¹⁵ and bulk liquid extractions discard the spatial metabolomic feature differentiation that could otherwise be observed.¹¹⁶ Correspondingly, imaging mass spectrometry (IMS) methods have been developed for agar cultivated¹¹⁷ and environmental¹¹⁸⁻¹²⁰ microbial samples to provide a second distinguishing ion characteristic, spatial localization. In IMS experiments, the area of a sample is divided into pixels which are individually analysed by the mass spectrometer. Matrix assisted laser desorption ionization (MALDI) is a commonly used ionization technique for IMS.^{121,122} MALDI requires the application of an ionization matrix to facilitate ionization of cell and agar embedded metabolites. This ablative technique has been applied to visualize spatial temporal distributions of secondary metabolite production in marine cyanobacteria^{51,120} and to elucidate microbial producers responsible for observed secondary metabolite biosynthesis.^{119,123} MALDI-IMS has also been used to visualize metabolic exchange between interacting organisms^{117,124-129} and identify novel antibiotic production in *Streptomyces*.¹³⁰ The efficiency of MALDI ionization is matrix dependent, and varies across metabolite classes. Correspondingly, Desorption Electrospray Ionization (DESI) and secondary ion mass spectrometry (SIMS), which do not require the addition of an ionizing matrix, have also been applied to visualize natural product distribution through IMS¹³¹⁻¹³⁴ among others.¹³⁵ Determining the spatial distribution of produced secondary metabolites can be useful in natural product research, and as these IMS technologies continue to develop they are expected to become an integral component of metabolomic investigations into microbial secondary metabolites.¹³⁶ Metabolomic features generated via IMS consist of m/z and its corresponding Cartesian coordinate in agar culture.

3 Preparation of high content mass spectral data for metabolomics studies

3.1 Strategies for formatting data for effective comparative analysis

Typical LC-MS analysis of extracted metabolites results in thousands of detectable features characterized by m/z , and retention time, as well as potentially ion mobility and fragmentation.¹³⁷ Unbiased manual comparison of features between samples is challenging, especially when analysing a large number of samples. Therefore, it is necessary to automate feature collection from the acquired data in several processing steps that will facilitate data analysis. There are a variety of non-compatible vendor-specific data file formats for mass spectrometric data which originally impeded the development of universal processing software. To address this issue the Protein Standards Initiative (PSI) group¹³⁸ developed a standardized format, mzData, to facilitate data exchange.¹³⁹ An additional format, mzXML, was developed by to serve as a standard format for MS and MS/MS data processing.^{140,141} While both of these formats were popular, the scientific community pushed for a unified standard format to simplify software development. A new format mzML was released to

replace both mzXML and mzData formats,¹⁴² however, all of these are still commonly used for metabolomics data. Correspondingly, one of the first steps in metabolomics data processing is to convert the data files from a vendor-specific format into one of the appropriate standard formats listed above for the processing software. One common utility for this is ProteoWizard's MSconvert,¹⁴³ which also has the ability to pre-filter the data with user defined parameters.

3.2 Methods and considerations for metabolite peak detection and alignment

After data format conversion, metabolite peaks must be identified and extracted from the data and aligned for all samples. A number of reviews have covered and compared the various processing packages and their algorithms.^{110,144-146} In this section we highlight a few of the common computational methods employed natural product discovery based metabolomics. The initial peak identification can be fairly challenging, as LC-MS ionization methods typically generate high levels of background chemical noise largely from mobile phases and buffers.¹⁴⁷ Therefore, the automated processing methods must be able to identify genuine sample features while omitting detected chemical background and instrumental noise, and there have been several algorithms developed to accomplish this task. Vectorized peak detection algorithms identify data points above a set intensity threshold in both the m/z and retention time dimensions.^{148,149} There are also a number of 1-dimensional LC-MS processing algorithms commonly used for peptide analysis which detect peaks by using the isotope patterns in the m/z dimension.¹⁵⁰⁻¹⁵² Another of the more common methods involves separating the LC-MS data into extracted ion chromatograms (EIC), each covering a very narrow m/z range. This process is called binning and, while fast and generally effective, this can lead to problems if the bin size is too large or too small. A matched filter¹⁵³ is commonly applied to EICs to select for m/z peak shapes in the chromatographic time domain, and if features are split between multiple bins due to inappropriate sizing, they can be excluded by the algorithm resulting in false negatives. The traditional XCMS peak detection algorithm, a widely used LC-MS processing software package, sections off 0.1 Dalton wide EICs and then applies a second derivative Gaussian filter that aids in the discrimination of authentic peaks from noise along with a 10:1 signal to noise intensity threshold.¹⁵⁴ An alternative to the binning approach for high resolution MS data is the centWave algorithm which identifies ion dense regions of interest in centroid data.¹⁵⁵ Peaks are detected along these regions using a continuous wavelet transform, which allows for a much more dynamic range of peak shapes.^{155,156} The quality and validation of peak detection from increasingly complex datasets remains an area of intense research efforts.^{157,158}

Another consideration in data processing is the tendency for retention times of features to vary between multiple injections due to changes in chromatographic conditions discussed previously. It is therefore necessary to match mass features between samples of an experiment, and align the retention times of matched peaks to generate a discrete feature list. Originally, internal reference standards were used to adjust retention times of each sample.^{159,160} However, retention time drifts throughout an acquisition are often not linear,^{154,161} and this also required additional sample preparation steps to incorporate the standards. A variety of algorithms have been developed to align features between sample

runs without the use of internal standards. The original XCMS alignment algorithm identified hundreds to thousands of peak groups that are present in a large number of samples. These “well behaved” groups are used as markers to align the remaining detected features. Typically, the number of these markers identified from metabolome extracts is sufficient to cover the chromatographic profile of samples and correctly align the nonlinear retention time drifts. Local regression, LOESS,¹⁶² is then used to approximate drifts for regions without sufficient peak markers. Several alignment algorithms have been developed to process LC-MS data,¹⁶³⁻¹⁶⁹ and in a comparative study of six freely available retention time alignment methods the XCMS algorithm was shown to be the best for processing metabolomics data.¹⁷⁰ However, it was noted that the appropriate selection of parameters used for the methods could have a large impact on the data output, such that the apparent success of any particular method is dependent upon the user’s experience. A software package, Isotopologue Parameter Optimization (IPO), was recently released to automatically optimize XCMS parameter settings using natural C13 isotopic peaks.¹⁷¹ This software applies to a variety of different sample types, chromatographic strategies, and instrument methods and aids to simplify and systematize method development while optimizing metabolomics processing for non-experts.

After peak alignment it is common for several mass/retention time features to possess few or even no matches between samples. This may be because some peaks are entirely unique to a subset of experimental samples but can also stem from errors in peak detection due to inappropriate parameter settings, noisy data, etc. Gap-filling is commonly used to ensure these are not false negatives and provide a non-zero value for subsequent statistical analyses. In the absence of a detectable peak, the values obtained through gap filling reflect noise within the region peaks that were detected in other samples. For low abundance features, the integrated noise level over the peak region may be similar to the value determined for the feature, and this can lead to the observation of a metabolite ion that statistically correlates with a single condition while its lower abundance isotopes show no correlations in subsequent statistical analyses.

4 Analysis of metabolomics data in the context of secondary metabolites

The next stage of metabolomics analysis consists of applying one or more methods to compare metabolomics datasets. Depending on the objectives of a given study, several complementary methods may be applied. In the following discussion we review extant methods for comparative metabolomics analysis. To illustrate the application of these methods, we apply them to the analysis of a metabolomics dataset focused on the cytotoxic macrolide producing organism, *Nocardioopsis sp FU40*, and its exposure to multiple competing organism in mixed culture. In selected mixed culture conditions, this organism increased production of the secondary metabolites called ciromicins, which we highlight throughout data analyses.

4.1 Multivariate statistical analysis and data projections for identification of abundant covarying metabolites

Subsequent to pre-processing, metabolomics data can be analysed through multivariate statistical analyses (MVSA) which simplify and identify significant correlations within the data. Two common methods for metabolomics data analysis are partial least squares (PLS), or projections to latent structures, regression methods and principal component analyses (PCA) reviewed in more detail elsewhere.^{172,173} Briefly, PLS methods assume that changes within the data are largely driven by a subset of latent variables, which are not themselves measured within the data but are more abstract, such as experimental treatments/stimulants or conditions. With this assumption, a PLS analysis will identify latent vectors within the data that describe the maximal covariance between user defined groups.^{174,175} Alternatively, PCA makes no assumptions about the data and identifies the sources of the highest variance across the samples to distinguish the samples from one another.¹⁷⁶ The fundamental difference between these two analyses is that PLS are supervised with user defined groups while PCA are unsupervised variable reduction methods. Orthogonal signal corrections can be applied to PLS regressions to improve separation between predictive and non-predictive variation.¹⁷⁷ The product of these analyses are scoresplots, or projections of samples onto a hyperplane within the data describing sample covariance, from PCA and PLS analyses. Interpretation of scoresplots show the separation of samples based on feature variance to determine which samples are similar (nearby in Cartesian space) and dissimilar (far away) with regards to their most significantly varying features. Replicate analyses of the same sample should cluster within the scoresplot, and in this way scoresplots are a useful means of identifying errors in sample acquisition or data pre-processing. Additionally, a control comprised of pooled samples should locate close to the origin of a PCA plot. Another useful product of the PCA analyses are loadings plots, which show correlations between variables in the data and summarize these variables' impacts on the scoresplot. Nearby features are positively correlated, while distant features are negatively correlated, and features in the same region as samples in the scoresplot will be more abundantly or uniquely present in those samples.

4.1.1 Strain prioritization via principal component analysis—One approach to the discovery of new natural products has been to prioritize organisms distinguished as metabolically unique through a PCA analysis. There is often a great deal of redundancy in the compounds identified through microbial natural product screening endeavours, and this redundancy can be reduced through the selection of metabolically diverse microbial strains.¹⁷⁸ Under the hypothesis that organisms with similar secondary metabolic potential would cluster in PCA space, Hou et al. analysed 47 microbial strains to demonstrate how MVSA could prioritize strains with diverse secondary metabolic potential.¹⁷⁹ Similarly, PCA has been used to prioritize marine microbial symbionts¹⁸⁰ as well as phylogenetically similar *Streptomyces*¹⁸¹ for natural product isolation. Figure 2 demonstrates how metabolically unique organisms are distinguishable along the principal component vectors of the PCA scoresplot. This method may also be useful for identifying new classes of bioactive microbial compounds as has been done for plant extracts.¹⁸² However, caveats to this approach include (1) that the correlating features responsible for PCA prioritization of a subset of organisms from a library may not be secondary metabolites, which are generally

present in relatively low abundance within crude extracts, and (2) that low abundant secondary metabolites, will not be emphasized by these methods.

4.1.2 Secondary metabolite prioritization within metabolomics data via principal component and regression analyses

—One important application of PCA and PLS metabolomics for natural product discovery is to prioritize induced secondary metabolites in comparative analyses between chemically and/or biologically stimulated and control conditions. Secondary metabolite production can be activated in microorganisms through a variety of chemical and environmental stimulation,^{29,30,183,184} and PCA and PLS are commonly applied to identify abundantly produced features in these conditions.^{40,185-187} Binary comparisons using S-plots can be used to identify group specific features of a PLS model.¹⁸⁸ These graphs separate features by their covariance along the x-axis and their correlation to user defined groups on the y-axis. More simply, more abundant features are farther from the origin on the x-axis, and features with correlations closer to 1 or -1 are likely to be unique or specific to one group or the other. Volcano plots have also recently been used to identify significantly covarying metabolites in binary comparisons of natural product extracts.¹⁸⁹ Volcano plots show each features' statistical significance, p-value, on the y-axis and fold change along the x-axis.^{190,191} Similarly PCA loadings plots can be used to visualize significant feature differences between sample sets. Figure 3 demonstrates how S-plots, volcano plots, and loadings plots can distinguish induced metabolomic features in our *Nocardiosis* case study. The loadings plot tripartite comparison identifies features that correlate with either the *Nocardiosis* or *Rhodococcus* monocultures or a mixed culture where the two compete for nutrients. In this plot the induced cytotoxic macrolactam ciromicin is clearly distinguishable as positively correlated with the mixed culture extract. Similarly, ciromicin was clearly identified through the S-plot and volcano plot comparisons between the *Nocardiosis* monoculture and the mixed culture. These methods can be very powerful, and freely available online metabolomics packages, such as XCMS Online^{192,193} and Metaboanalyst^{194,195} can perform some routine MVSA data analyses in addition to data pre-processing. An alternative and fairly unique comparative analysis available through XCMS Online is the cloud plot¹⁹⁶. These plots convey feature fold changes, *m/z*, retention time, and statistical distribution in the same Figure, and can perform both binary and multigroup comparisons.¹⁹³

4.2 Discovering molecular inventories of microbial responses via self organizing map analytics

A strength in MVSA analysis of metabolomics datasets is the identification of the most unique and abundant features between small numbers of treatment conditions. However, these methods are limited to displaying data in two or three dimensions and are biased towards the largest differences within the entire dataset. Therefore, the utility of MVSA to represent an experiment diminishes as the number and diversity of samples increases. For instance, it is common to screen a target organism under dozens of stimulus conditions to optimize compound production, or to induce silent biosynthetic gene clusters, and in these cases we have previously demonstrated that an alternative method utilizing Kohonen self-organizing map (SOM) analytics can be more effective at representing multiplexed stimuli data than PCA.¹⁸⁴ As discussed above, metabolomic acquisition via LC-MS results in the

acquisition of thousands of detectable features. Through SOM analyses these features are organized using an artificial neural network into a 2-dimensional grid based on feature response patterns across all experimental conditions. Features that share similar trend patterns are grouped in nearby nodes of the map as shown in Figure 4. Through multiple iterations, typically several hundred, this organization is improved ultimately resulting in a feature map where features in this case correspond to clusters of similar response trends. Unlike MVSA, SOM analyses improve with increasing amount of data and response conditions (e.g. stimuli), as this leads to more varied response trends which in turn enhances feature organization. A metabolomics workflow – molecular expression dynamics inspector (MEDI), provides an open access methodology for SOM analysis from MS data and is readily applicable to microbial metabolomics.¹⁹⁷ In MEDI, each tile, or node, of the grid is coloured based on the centroid intensity of its features to generate heatmaps. Difference maps can be generated by subtractive analysis (e.g. control and stimulus conditions) to readily prioritize abundant and treatment-specific metabolomic features into regions of interest. We applied these SOM analytics to map stimuli-induced metabolomic responses from 23 distinct conditions in *Streptomyces coelicolor*.¹⁸⁴ In this study 16 detected secondary metabolites produced by *S. coelicolor* were induced in one of the 23 conditions and prioritized through the SOM analysis.¹⁸⁴ Application of SOM analyses to investigate metabolomic changes engendered through microbial competition in our *Nocardiosis* case study prioritized several metabolites unique to mixed culture conditions including the ciromicins. Additionally, the single SOM analysis recapitulates the results of multiple MVSA analyses. In Figure 5 three PCA loadings plots are compared with three SOM heatmaps from the *Nocardiosis* mixed culture example study. When the features held within regions of interest on the SOM maps are sorted by abundance they are highly consistent with the loadings plots from PCA analyses. Indeed, there is correspondence between PCA and the neural networks used for MEDI analysis. As the data and/or conditions become sparser, the SOM heatmap begins to decompose into a similar functional form as PCA.

4.3 Molecular networking to reveal structural uniqueness and relatedness in large datasets.

Microorganisms have been extensively mined for natural products throughout much of the past century in the search for new pharmaceuticals, and the rediscovery of known compounds or known families of compounds is quite common. Identifying and removing these rediscovered natural products, a process known as dereplication, is both critical and challenging.^{59,198} Typically accurate masses or determined molecular formula of extracted compounds are used to search databases of known natural products. However, the large number of isobaric compounds complicate dereplication. UV/Vis absorbance⁵⁹ spectra and chromatographic retention times¹⁹⁹ can be used to further match extracted features to database compounds, and as technologies and databases improve, it is likely that ion mobility will play a role in natural product dereplication as well.^{85,86,200} Fragmentation spectra acquired through tandem MS is another useful property for dereplication. Metabolite fragmentation patterns observed through MS/MS analysis can be matched to those in databases like PubChem, METLIN²⁰¹ and MassBank²⁰² to putatively identify MS features. Kernel based machine learning algorithms have recently been applied to dereplicate

metabolites using multiple levels of tandem MS,^{91,203} and while this works well for primary metabolites, public databases for microbial secondary metabolites with fragmentation spectra encompass only a small fraction of known natural products. For example, GNPS: Global Natural Products Social Molecular Networking, the largest natural product public database with MS/MS spectra, contains more than 140,000 natural products,¹⁹⁸ and there are an estimated 600,000 published natural compounds.²⁰⁴

Computational methods to generate theoretical fragmentation spectra have been employed to compensate for the lack of experimental data on natural products.²⁰⁵ These in-silico MS/MS spectral databases can further facilitate natural product dereplication when coupled with molecular networking,²⁰⁶ and as both experimental and in-silico database coverage improves, comparisons of fragmentation spectra may become the most useful method of natural product dereplication. In addition to matching fragmentation spectra with database compounds, fragmentation data can be used to cluster related classes of molecules by fragment similarity. Molecular networking analyses cluster families of molecules through vector correlations between fragment ions.²⁰⁷ Yang et al. demonstrated the utility of this approach for natural product discovery by dereplicating 58 natural products from marine and terrestrial microorganisms.²⁰⁸ Molecular networking in this study also identified a number of novel analogs to known compounds, which are more difficult to obtain through other dereplication methods. In Figure 6 we have applied molecular networking to our *Nocardiosis* example dataset. Using the network visualizer in the GNPS: Global Natural Products Social Molecular Networking website, fragmentation spectra for each object in the network can be easily viewed and compared to matched reference spectra from the GNPS library. The features in the network can also be coloured by the user-defined group or condition to which they are correlated. In Figure 6 we have highlighted features unique to mixed culture conditions in red. As shown, molecular networking identifies unique features which is unbiased by compound abundance. Several features of this dataset share no significant fragment similarity with the network and are isolated as “self-loops”. In fact, ciromicin A is among these uniquely fragmenting features, and this in itself may be another useful means to prioritize leads, as outlying features may be more structurally unique. Molecular networking analysis can be enhanced by combining additional metabolomics techniques. Klitgaard et al. used a combination of molecular networking and stable isotope labelling to identify novel analogs of nidulanin A and fungisporin in the well-studied fungus *Aspergillus nidulans*.⁷⁰ Fragment based clustering in this manner can also be used to identify modified natural products stemming from interactions between organisms. Moree et al. used molecular networking with imaging mass spectrometry to investigate the interkingdom interactions between *Pseudomonas* and *Aspergillus* and observed a variety of biotransformed metabolites arising from this microbial competition.²⁰⁹ Similarly, Briand et al. applied molecular networking to identify new compounds and analogs arising from intraspecific interactions between algae.²¹⁰ The application of molecular networking for the *Nocardiosis* mixed culture data shown in Figure 6 links a number features found in the *Nocardiosis* monoculture with similarly fragmenting features only detectable in mixed culture. These may represent compounds made by *Nocardiosis* that are stimulated or modified in some way by the competitor *Tsukamurella*.

Molecular networking can also prioritize features by linking observed natural products to their cognate biosynthetic gene clusters and gene cluster families^{99,211} when used in conjunction with genomic sequence analysis. This can be an advantageous means of prioritizing metabolite leads as demonstrated by the work from Kleigrewe et al. where molecular networking was combined with genomic sequence analysis to identify a novel group of acyl amides, termed columbamides, from marine cyanobacteria.²¹² The Crawford lab has recently employed 'pathway-targeted' molecular network analyses to identify metabolites from the colibactin gene cluster, which had been linked to increased virulence in *E. coli*.²¹³⁻²¹⁵ As previously discussed, heterologous hosts are often used for the production of microbial secondary metabolites,²¹⁶⁻²¹⁸ and molecular networking is a useful tool for comparative metabolomics to visualize the output of these heterologous hosts. Schorn et al. used molecular networking to identify novel eponemycin congeners produced through heterologous expression in *Streptomyces albus* J1046.²¹⁹ Molecular networking has even been applied to identify virulence factors in pathogenic organisms,^{220,221} and this method will become more beneficial for natural product discovery as databases and technologies improve.

5. Investigations of secondary metabolite bioactivity

Natural products are intrinsically biologically active, however, the clinical relevance of this activity may not always be discernible. Typically, natural product structure and mode of action are determined fairly late in the natural product discovery pipeline, which contributes to high rediscovery rates. Therefore, prioritizing natural product leads by deep profiling of pharmacologically relevant biological activities would expedite natural product based drug discovery. Natural product extracts are commonly divided into multiple fractions which are then screened to identify the components underlying the desired biological activity.²²²⁻²²⁵ However, low abundance compounds can often be overlooked in complex extracts, and recently MVSA have been utilized to help link observed fraction bioactivity to detectable features from metabolomic analyses.^{39,226} Even after correlating metabolites with biological activity, determining the mode of action for active compounds can be difficult and expensive.²²⁷⁻²³⁰ One approach has been developed that uses the antibiotic spectrum of activities across different organisms, mode of action profiles (BioMAP), to group similar antibiotics.²³¹ This method was effectively able to cluster antibiotics of the same compound class and led to the identification of a novel naphthoquinone antibiotic, arromycin.²³¹ Gene expression profiling with either the entire transcriptome^{232,233} or a subset of reporter genes^{234,235} has also been used to predict modes of action for natural products. However, because these transcriptomic screens are still relatively costly, there is a great interest in applying metabolomics analyses to predict natural product modes of action using either natural product extracts²³⁶ or purified compounds.²³⁷⁻²⁴² Vincent et al. have recently shown untargeted metabolomics can effectively identify compound modes of action when specific metabolic pathways are the primary drug target.²⁴³ Metabolomic consequences of drug combinations may additionally be able to identify synergism, or antagonism between coadministered drug therapies.²⁴¹ In a study with *M. smegmatis*, Halouska et al. observed that antibiotics which share similar biological targets engender similar metabolomic changes and are grouped together through MVSA.²³⁹ The group additionally applied their

metabolomic methods to investigate antibiotics with unknown biological targets and found them to group with membrane disrupting antibiotics, ampicillin, D-cycloserine, and vancomycin.²³⁹ This methodology could prove very useful to prioritize compounds for isolation. Antimicrobial extracts which separate themselves metabolically through MVSA or other analyses may exert their activity through a novel biological target or mechanism. In this way pharmaceutically relevant natural products could be prioritized for isolation. These metabolomic analyses have even been applied to investigate the underlying methods by which known antibiotics kill pathogens.²⁴⁴ Another approach, cytological profiling, uses automated image and microscopy analyses to identify phenotypic changes induced from bioactive compounds,^{245,246} and this method has been used to classify biologically active compounds by their respective modes of action^{247,248} even within more complex marine derived bacterial extracts.²⁴⁹ A combined approach integrating these phenotypic screens with untargeted metabolomics has recently been developed to predict the modes of action for complex libraries of natural products and prioritize unique bioactive components.²⁵⁰ Applying this method, Compound Activity Mapping, on data from 234 natural product extracts led to the discovery of the quinocinnolinomycins, a new family of natural products implicated to induce endoplasmic reticulum stress based on further cytological profile clustering.²⁵⁰ Ultimately, these multi-omic combinatorial methods may become the preferred means of predicting molecular modes of action. Integrating the phenotypic data from cytological profiling and the transcriptomic functional signature ontologies²³⁵ with metabolomics data using one or combinations of the powerful analytical platforms discussed in this review, self-organizing maps, molecular networking, MVSA, etc., could provide new insights into the modes of action of bioactive compounds and greatly facilitate novel drug discovery.

6 Conclusions

Metabolomic analyses are powerful tools for natural product discovery. However, while metabolomics can provide a wealth of information regarding the activity and responses of microorganisms, with current technologies it is practically impossible to analyse the entire metabolome of an organism comprehensively due to variations in ionization efficiency and limitations in detection across wide concentration dynamic range. Instead, only detectable metabolites, which make up a fraction of the total metabolites present, are used to draw conclusions from current studies. While the full transcriptomic and proteomic potential of an organism can be determined through modern genome sequencing, there is no readily discernible limit to the number of metabolites present within organisms, so it is difficult to predict the number of metabolites omitted by current analyses. Due to these limitations, extra care must be taken when drawing conclusions from metabolomics datasets. Nonetheless, metabolomics analyses benefit microbial natural product discovery pipelines in a variety of ways as described in this report. These can be used to prioritize organisms, identify activated compounds from stimuli exposure, prioritize features through bioactivity spectrums or molecular class, and even dereplicate prioritized secondary metabolites. The metabolomics methods described herein may also facilitate investigations into the fundamental purpose behind secondary metabolite production within microbial communities. It is largely unclear how the production of secondary metabolites is regulated

in situ as well as which ecological stimuli trigger secondary metabolic production. Such studies would benefit natural product discovery endeavours by facilitating predictions of stimuli to induce secondary metabolite production within the endogenous producer and could additionally provide insight into human health and wellness. Microbial secondary metabolites have a significant impact on human health by means of both isolated pharmaceuticals and compounds produced *in situ* from within human microbiomes,^{13,251-255} and metabolomics may be able to offer insights into how these organisms modulate their secondary metabolism in response to diet, medicine, and endogenous host factors.

Comparative metabolomics methods are aiding in unleashing the repressed and/or hidden wealth of microbial secondary metabolism predicted by whole genome sequencing. The combination of complimentary methods (e.g. SOM and molecular networking) has the potential provide new tools to accelerate discovery by comparison. Ultimately, the purpose of these efforts is to identify biological roles for secondary metabolites, be they biochemical, chemical ecological, or translational in human medicine. We feel that the next era in secondary metabolite discovery and application will be facilitated by methods that combine high content biological activity data measurements for metabolites within metabolomes with corresponding multidimensional metabolomic data to illuminate effectors of natural small molecule interactions, and their roles in biological systems.

7 Acknowledgements

This work was supported by the National Institutes of Health (no. R01GM092218 awarded to B.O.B. and J.A.M., and T32 no. GM0650086 awarded to B.C.C.), the Vanderbilt Institute of Chemical Biology, the Vanderbilt Institute for Integrative Biosystems Research and Education, and the Vanderbilt University College of Arts and Sciences.

Biographies



Brett Covington received a B.S. in Chemistry from Austin Peay State University in 2012. He was awarded with an NIH sponsored Chemistry-Biology Interface training grant in 2013 and is currently a 4th year PhD candidate at Vanderbilt University under the supervision of Brian Bachmann. His research in the natural product discovery section of the Bachmann lab has been primarily focused on the application of untargeted metabolomic methods to identify microbial secondary metabolite responses to environmental stimuli.



Brian Bachmann graduated in 2000 with a PhD in Chemistry from The Johns Hopkins University, studying with Prof Craig Townsend. Following this, he joined Ecopia Biosciences in Montreal, where he was Director of Chemistry, prior to moving to Vanderbilt, where he is currently Professor of Chemistry and Associate Director of the Vanderbilt Institute for Chemical Biology. He has a long standing interest in studying and harnessing the biosynthetic capabilities of living systems in order to radically accelerate the discovery of new bioactive compounds, and revolutionize how molecules are synthesized.



John McLean graduated in 2001 with a PhD in Chemistry from George Washington University. Following postdoctoral research at Forschungszentrum Jülich in Germany and then at Texas A&M University with Prof. David H. Russell, he began at Vanderbilt University in 2006 where he is Stevenson Professor of Chemistry, Director of the Center for Innovative Technology, co-Director of the Automated Biosystems Core, and Deputy Director of the Institute for Integrative Biosystems Research and Education at Vanderbilt University. McLean and colleagues focus on the conceptualization, design, and construction of structural mass spectrometers, specifically targeting complex samples in systems, synthetic, and chemical biology.

8 References

1. Bentley SD, Chater KF, Cerdeño-Tárraga AMM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D, Bateman A, Brown S, Chandra G, Chen CW, Collins M, Cronin A, Fraser A, Goble A, Hidalgo J, Hornsby T, Howarth S, Huang CHH, Kieser T, Larke L, Murphy L, Oliver K, O'Neil S, Rabinowitsch E, Rajandream MAA, Rutherford K, Rutter S, Seeger K, Saunders D, Sharp S, Squares R, Squares S, Taylor K, Warren T, Wietzorrek A, Woodward J, Barrell BG, Parkhill J, Hopwood DA. *Nature*. 2002; 417:141–147. [PubMed: 12000953]
2. Ikeda H, Ishikawa J, Hanamoto A, Shinose M, Kikuchi H, Shiba T, Sakaki Y, Hattori M, Omura S. *Nat. Biotechnol.* 2003; 21:526–531. [PubMed: 12692562]
3. Zarins-Tutt JS, Barberi TT, Gao H, Mearns-Spragg A, Zhang L, Newman DJ, Goss RJM. *Nat. Prod. Rep.* 2016; 33:54–72. [PubMed: 26538321]

4. Scherlach K, Hertweck C. *Org. Biomol. Chem.* 2009; 7:1753–1760. [PubMed: 19590766]
5. Walsh CT, Fischbach MA. *J. Am. Chem. Soc.* 2010; 132:2469–2493. [PubMed: 20121095]
6. Shabuer G, Ishida K, Pidot SJ, Roth M, Dahse H-MM, Hertweck C. *Science (New York, N.Y.)*. 2015; 350:670–674.
7. Findlay BL. *ACS Chem. Biol.* 2016; 11:1502–1510. [PubMed: 27035738]
8. Vizcaino MI, Guo X, Crawford JM. *J Ind Microbiol Biotechnol.* 2014; 41:285–299. [PubMed: 24127069]
9. Newman DJ, Cragg GM. *J. Nat. Prod.* 2016; 79:629–661. [PubMed: 26852623]
10. Zakrzewski P, Fischbach MA, Weber T. *Nucl. Acids Res.* 2011; 39:339–346.
11. Li MHT, Ung PMU, Zajkowski J. *BMC Bioinformatics.* 2009;10. [PubMed: 19133123]
12. Ju K-SS, Gao J, Doroghazi JR, Wang K-KA, Thibodeaux CJ, Li S, Metzger E, Fudala J, Su J, Zhang JK, Lee J, Cioni JP, Evans BS, Hirota R, Labeda DP, van der Donk WA, Metcalf WW. *Proc. Natl. Acad. Sci. U. S. A.* 2015; 112:12175–12180. [PubMed: 26324907]
13. Donia MS, Cimermancic P, Schulze CJ, Wieland Brown LC, Martin J, Mitreva M, Clardy J, Linington RG, Fischbach MA. *Cell.* 2014; 158:1402–1414. [PubMed: 25215495]
14. Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, Lee SY, Fischbach MA, Müller R, Wohlleben W, Breitling R, Takano E, Medema MH. *Nucl. Acids Res.* 2015; 43:237–243. [PubMed: 25435544]
15. Jensen PR, Chavarria KL, Fenical W, Moore BS. *J. Ind. Microbiol. Biotechnol.* 2014; 41:203–209. [PubMed: 24104399]
16. Farnet, CM.; Zazopoulos, E. *Natural Products: Drug Discovery and Therapeutic Medicine.* Zhang, L.; Demain, AL., editors. Humana Press; Totowa, NJ: 2005. p. 95-106.
17. Ziemert N, Alanjary M, Weber T. *Nat. Prod. Rep.* 2016 DOI: 10.1039/C6NP00025H.
18. Medema MH, Fischbach MA. *Nat. Chem. Biol.* 2015; 11:639–648. [PubMed: 26284671]
19. Juan Pablo G-E, Mervyn JB. *J. Ind. Microbiol. Biotechnol.* 2013; 41:425–431. [PubMed: 24096958]
20. Komatsu M, Uchiyama T, mura S, Cane DE, Ikeda H. *Proc. Natl. Acad. Sci. U. S. A.* 2010; 107:2646–2651. [PubMed: 20133795]
21. Kim E, Moore BS, Yoon YJ. *Nat. Chem. Biol.* 2015; 11:649–659. [PubMed: 26284672]
22. Tang X, Li J, Millán-Aguiñaga N, Zhang JJ. *ACS Chem. Biol.* 2015; 10:2841–2849. [PubMed: 26458099]
23. Ross AC, Gulland LES, Dorrestein PC. *ACS Synth. Biol.* 2014; 4:414–420. [PubMed: 25140825]
24. Donia MS, Ruffner DE, Cao S, Schmidt EW. *ChemBioChem.* 2011; 12:1230–1236. [PubMed: 21542088]
25. Ongley SE, Bian X, Neilan BA, Müller R. *Nat. Prod. Rep.* 2013; 30:1121–1138. [PubMed: 23832108]
26. Luo Y, Enghiad B, Zhao H. *Nat. Prod. Rep.* 2016; 33:174–182. [PubMed: 26647833]
27. Bibb MJ. *Curr. Opin. Microbiol.* 2005; 8:208–215. [PubMed: 15802254]
28. Seyedsayamdost MR, Chandler JR, Blodgett JAV. *Org. Lett.* 2010; 12:716–719. [PubMed: 20095633]
29. Bode HB, Bethe B, Höfs R, Zeeck A. *ChemBioChem.* 2002; 3:619–627. [PubMed: 12324995]
30. Rutledge PJ, Challis GL. *Nat. Rev. Microbiol.* 2015; 13:509–523. [PubMed: 26119570]
31. Ochi K, Hosaka T. *Appl. Microbiol. Biot.* 2013; 97:87–98.
32. Reen FJ, Romano S, Dobson ADW, O'Gara F. *Mar. Drugs.* 2015; 13:4754–4783. [PubMed: 26264003]
33. Imai Y, Sato S, Tanaka Y, Ochi K, Hosaka T. *Appl. Environ. Microbiol.* 2015; 81:3869–3879. [PubMed: 25819962]
34. Wang W, Ji J, Li X, Wang J, Li S, Pan G, Fan K, Yang K. *Proc. Natl. Acad. Sci. U. S. A.* 2014; 111:5688–5693. [PubMed: 24706927]
35. Tokuyama S, Kaji A, Ikeda H, Ochi K. *Appl. Environ. Microbiol.* 2009; 75:4919–4922. [PubMed: 19447953]

36. Tsurumi Y, Kodani S, Yoshida M, Fujie A, Ochi K. *Nat. Biotechnol.* 2009; 27:462–464. [PubMed: 19396160]
37. Wang G, Hosaka T, Ochi K. *Appl. Environ. Microbiol.* 2008; 74:2834–2840. [PubMed: 18310410]
38. Ochi K, Okamoto S, Tozawa Y, Inaoka T, Hosaka T, Xu J, Kurosawa K. *Adv. Appl. Microbiol.* 2003; 56:155–184.
39. Wu C, Du C, Gubbens J, Choi YH. *J. Nat. Prod.* 2015; 78:2355–2363. [PubMed: 26438963]
40. Derewacz DK, Goodwin CR, McNeese RC, McLean JA, Bachmann BO. *Proc. Natl. Acad. Sci. U. S. A.* 2013; 110:2336–2341. [PubMed: 23341601]
41. Tanaka Y, Hosaka T, Ochi K. *J. Antibiot. (Tokyo)*. 2010; 63:477–481. [PubMed: 20551989]
42. Kawai K, Wang G, Okamoto S, Ochi K. *FEMS Microbiol. Lett.* 2007; 274:311–315. [PubMed: 17645525]
43. Hafnerburg G, Kothe E. *J. Basic Microbiol.* 2007; 47:453–467. [PubMed: 18072246]
44. Abdelmohsen UR, Grkovic T, Balasubramanian S, Kamel MS, Quinn RJ, Hentschel U. *Biotechnol. Adv.* 2015; 33:798–811. [PubMed: 26087412]
45. Derewacz DK, Covington BC, McLean JA, Bachmann BO. *ACS Chem. Biol.* 2015; 10:1998–2006. [PubMed: 26039241]
46. Angell S, Bench BJ, Williams H, Watanabe CMH. *Chem. Biol.* 2006; 13:1349–1359. [PubMed: 17185230]
47. Oh D-C, Kauffman CA, Jensen PR, Fenical W. *J. Nat. Prod.* 2007; 70:515–520. [PubMed: 17323993]
48. Cueto M, Jensen PR, Kauffman C, Fenical W, Lobkovsky E, Clardy J. *J. Nat. Prod.* 2001; 64:1444–1446. [PubMed: 11720529]
49. Brakhage AA. *Nat. Rev. Microbiol.* 2012; 11:21–32. [PubMed: 23178386]
50. Forseth RR, Schroeder FC. *Curr. Opin. Chem. Biol.* 2011; 15:38–47. [PubMed: 21071261]
51. Esquenazi E, Jones AC, Byrum T, Dorrestein PC, Gerwick WH. *Proc. Natl. Acad. Sci. U. S. A.* 2011; 108:5226–5231. [PubMed: 21393570]
52. Jarmusch AK, Cooks GR. *Nat. Prod. Rep.* 2014; 31:730–738. [PubMed: 24700087]
53. Krug D, Müller R. *Nat. Prod. Rep.* 2014; 31:768–783. [PubMed: 24763662]
54. Bereman MS, Lyndon MM, Dixon RB. *Rapid Commun. Mass Spectrom.* 2008; 22:1563–1566. [PubMed: 18421754]
55. Stroh JG, Petucci CJ, Brecker SJ, Huang N. *J. Am. Soc. Mass Spectrom.* 2007; 18:1612–1616. [PubMed: 17629494]
56. Sleno L, Volmer DA, Marshall AG. *J. Am. Soc. Mass Spectrom.* 2005; 16:183–198. [PubMed: 15694769]
57. Bristow AWT, Webb KS. *J. Am. Soc. Mass Spectrom.* 2003; 14:1086–1098. [PubMed: 14530089]
58. Kind T, Fiehn O. *Bioanal. Rev.* 2010; 2:23–60. [PubMed: 21289855]
59. Nielsen KF, Månsson M, Rank C, Frisvad JC, Larsen TO. *J. Nat. Prod.* 2011; 74:2338–2348. [PubMed: 22026385]
60. Kind T, Fiehn O. *BMC bioinformatics.* 2007;8. [PubMed: 17212835]
61. May JC, McLean JA. *Annu. Rev. Anal. Chem.* 2016; 9:387–409.
62. Kildgaard S, Mansson M, Dosen I, Klitgaard A. *Mar. Drugs.* 2014; 12:3681–3705. [PubMed: 24955556]
63. El-Elimat T, Figueroa M, Ehrmann BM, Cech NB, Pearce CJ, Oberlies NH. *J. Nat. Prod.* 2013; 76:1709–1716. [PubMed: 23947912]
64. Neumann S, Böcker S. *Anal. Bioanal. Chem.* 2010; 398:2779–2788. [PubMed: 20936272]
65. Hernandez MZ, Moreira DR, de Azevedo Junior WF, Leite AC. *Curr. Drug Targets.* 2010; 11:303–314. *C. S. M.* [PubMed: 20210755]
66. Kim JK, Harada K, Bamba T, Fukusaki E. *Biosci. Biotechnol. Biochem.* 2005; 69:1331–1340. [PubMed: 16041139]
67. Böcker S, Letzel MC, Lipták Z, Pervukhin A. *Bioinformatics.* 2009; 25:218–224. [PubMed: 19015140]
68. Haglund PS, Löfstrand K, Siek K, Asplund L. *Mass Spectrom.* 2013; 2:S0018.

69. Ong S-E, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M. *Mol. Cell Proteomics*. 2002; 1:376–386. [PubMed: 12118079]
70. Klitgaard A, Nielsen JB, Frandsen RJN, Andersen MR, Nielsen KF. *Anal. Chem*. 2015; 87:6520–6526. [PubMed: 26020678]
71. Bucar F, Wube A, Schmid M. *Nat. Prod. Rep*. 2013; 30:525–545. [PubMed: 23396532]
72. Ebada SS, Edrada RA, Lin W, Proksch P. *Nat. Protoc*. 2008; 3:1820–1831. [PubMed: 18989260]
73. Sticher O. *Nat. Prod. Rep*. 2008; 25:517–554. [PubMed: 18497897]
74. Wolfender JL, Marti G, Thomas A, Bertrand S. *J. Chromatogr. A*. 2015; 1382:136–164. [PubMed: 25464997]
75. Månsson M, Phipps RK, Gram L, Munro MH, Larsen TO, Nielsen KF. *J. Nat. Prod*. 2010; 73:1126–1132. [PubMed: 20509666]
76. Nielsen KF, Larsen TO. *Front. Microbiol*. 2015;6. [PubMed: 25713560]
77. Hounoum BM, Blasco H, Emond P, Mavel S. *Trac-Trend. Anal. Chem*. 2016; 75:118–128.
78. Beisken S, Eiden M, Salek RM. *Expert Rev. Mol. Diagn*. 2015; 15:97–109. [PubMed: 25354566]
79. May JC, McLean JA. *Anal. Chem*. 2015; 87:1422–1436. [PubMed: 25526595]
80. Lanucara F, Holman SW, Gray CJ, Eyers CE. *Nat. Chem*. 2014; 6:281–294. [PubMed: 24651194]
81. Cumeras R, Figueras E, Davis CE, Baumbach JI. *Analyst*. 2015; 140:1476–1490.
82. McLean JA, Ruotolo BT, Gillig KJ, Russell DH. *Int. J. Mass Spectrom*. 2005; 240:301–315.
83. Creaser CS, Griffiths JR, Bramwell CJ, Noreen S. *Analyst*. 2004; 129:984–994.
84. Kulchania M, Barnes CAS, Clemmer DE. *Int. J. Mass Spectrom*. 2001; 212:97–109.
85. Esquenazi E, Daly M, Bahrainwala T. *Bioorg. Med. Chem*. 2011; 19:6639–6644. [PubMed: 21784647]
86. Goodwin CR, Fenn LS, Derewacz DK. *J. Nat. Prod*. 2012; 75:48–53. [PubMed: 22216918]
87. Jeanne Dit Fouque K, Afonso C, Zirah S, Hegemann JD, Zimmermann M, Marahiel MA, Rebuffat S, Lavanant H. *Anal. Chem*. 2015; 87:1166–1172. [PubMed: 25495527]
88. Payne AH, Glish GL. *Methods Enzymol*. 2005; 402:109–148. [PubMed: 16401508]
89. Zhu Z-JJ, Schultz AW, Wang J, Johnson CH, Yannone SM, Patti GJ, Siuzdak G. *Nat. Protoc*. 2013; 8:451–460. [PubMed: 23391889]
90. Benton HP, Wong DM, Trauger SA, Siuzdak G. *Anal. Chem*. 2008; 80:6382–6389. [PubMed: 18627180]
91. Dührkop K, Shen H, Meusel M, Rousu J, Böcker S. *Proc. Natl. Acad. Sci. U. S. A*. 2015; 112:12580–12585. [PubMed: 26392543]
92. Hufsky F, Böcker S. *Mass spectrom. rev*. 2016; 9999:1–10.
93. Vaniya A, Fiehn O. *Trac-Trend Anal. Chem*. 2015; 69:52–61.
94. Tawfike AF, Viegelmann C, Edrada-Ebel R. *Methods Mol. Biol*. 2012; 1055:227–244.
95. Kertesz TM, Hall LH, Hill DW, Grant DF. *J. Am. Soc. Mass Spectrom*. 2009; 20:1759–1767. [PubMed: 19616966]
96. Wysocki VH, Joyce KE, Jones CM. *J. Am. Soc. Mass. Spectrom*. 2008; 19:190–208. [PubMed: 18191578]
97. Kersten RD, Yang Y-LL, Xu Y, Cimermanic P, Nam S-JJ, Fenical W, Fischbach MA, Moore BS, Dorrestein PC. *Nat. Chem. Biol*. 2011; 7:794–802. [PubMed: 21983601]
98. Kersten RD, Ziemert N, Gonzalez DJ, Duggan BM, Nizet V, Dorrestein PC, Moore BS. *Proc. Natl. Acad. Sci. U. S. A*. 2013; 110:4407–4416.
99. Nguyen DD, Wu C-HH, Moree WJ, Lamsa A, Medema MH, Zhao X, Gavilan RG, Aparicio M, Atencio L, Jackson C, Ballesteros J, Sanchez J, Watrous JD, Phelan VV, van de Wiel C, Kersten RD, Mehnaz S, De Mot R, Shank EA, Charusanti P, Nagarajan H, Duggan BM, Moore BS, Bandeira N, Palsson BØ, Pogliano K, Gutiérrez M, Dorrestein PC. *Proc. Natl. Acad. Sci. U. S. A*. 2013; 110:2611–2620. [PubMed: 23359698]
100. Reeta Rani S, Anil Kumar P, Carlos RS, Ashok P. *Biochem. Eng. J*. 2009; 44:13–18.
101. Ing-Lung S, Chia-Yu K, Feng-Chia H, Suey-Sheng K, Chienyan H. *J. Chin. Inst. Chem. Eng*. 2008; 39:635–643.

102. Robinson T, Singh D, Nigam P. *Appl. Microbiol. Biotechnol.* 2001; 55:284–289. [PubMed: 11341307]
103. Kunze B, Bohlendorf B, Reichenbach H. *J. Antibiot. (Tokyo)*. 2008; 61:18–26. [PubMed: 18305355]
104. Abdelfattah MS, Kharel MK, Hitron JA. *J. Nat. Prod.* 2008; 71:1569–1573. [PubMed: 18666798]
105. Surup F, Wagner O, von Frieling J. *J. Org. Chem.* 2007; 72:5085–5090. [PubMed: 17564461]
106. Ran i A, Sokovi M, Karioti A, Vukojevi J. *Environ. Toxicol. Pharmacol.* 2006; 22:80–84. [PubMed: 21783691]
107. Yoon TM, Kim JW, Kim JG, Kim WG, Suh JW. *J. Antibiot. (Tokyo)*. 2006; 59:640–645. [PubMed: 17191679]
108. Umezawa K, Ikeda Y, Kawase O. *J. Chem. Soc.* 2001; 1:1550–1553.
109. Murphy B, Anderson K, Borissow C, Caffrey P. *Org. Biomol. Chem.* 2010; 8:3758–3770. [PubMed: 20571619]
110. Zhou B, Xiao J, Tuli L, Resson HW. *Mol. BioSyst.* 2012; 8:470–481. [PubMed: 22041788]
111. Forcisi S, Moritz F, Kanawati B, Tziotis D. *J. Chromatogr. A.* 2013; 1292:51–65. [PubMed: 23631876]
112. Kuehnbaum NL, Britz-McKibbin P. *Chem. Rev.* 2013; 113:2437–2468. [PubMed: 23506082]
113. Straight PD, Kolter R. *Annu. Rev. Microbiol.* 2009; 63:99–118. [PubMed: 19566421]
114. Little AE, Robinson CJ, Peterson SB, Raffa KF, Handelsman J. *Annu. Rev. Microbiol.* 2008; 62:375–401. [PubMed: 18544040]
115. Ryan RP, Dow JM. *Microbiology.* 2008; 154:1845–1858. [PubMed: 18599814]
116. Watrous JD, Dorrestein PC. *Nat. Rev. Microbiol.* 2011; 9:683–694. [PubMed: 21822293]
117. Yang JY, Phelan VV, Simkovsky R, Watrous JD, Trial RM, Fleming TC, Wenter R, Moore BS, Golden SS, Pogliano K, Dorrestein PC. *J. Bacteriol.* 2012; 194:6023–6028. [PubMed: 22821974]
118. Lane AL, Nyadong L, Galhena AS. *Proc. Natl. Acad. Sci. U. S. A.* 2009; 106:7314–7319. [PubMed: 19366672]
119. Simmons TL, Coates RC, Clark BR, Engene N, Gonzalez D, Esquenazi E, Dorrestein PC, Gerwick WH. *Proc. Natl. Acad. Sci. U. S. A.* 2008; 105:4587–4594. [PubMed: 18250337]
120. Esquenazi E, Coates C, Simmons L, Gonzalez D, Gerwick WH, Dorrestein PC. *Mol. Biosyst.* 2008; 4:562–570. [PubMed: 18493654]
121. Eriksson C, Masaki N, Yao I, Hayasaka T. *Mass Spectrom.* 2013; 2:S0022.
122. Gessel MM, Norris JL, Caprioli RM. *J. Proteomics.* 2014; 107:71–82. [PubMed: 24686089]
123. Masatoshi T, Joshawna KN, Niclas E, Eduardo E, Tara B, Pieter CD, William HG. *J. Nat. Prod.* 2010; 73:393–398. [PubMed: 19839606]
124. Yu-Liang Y, Yuquan X, Paul S, Pieter CD. *Nat. Chem. Biol.* 2009; 5:885–887. [PubMed: 19915536]
125. Bleich R, Watrous JD, Dorrestein PC, Bowers AA, Shank EA. *Proc. Natl. Acad. Sci. U. S. A.* 2015; 112:3086–3091. [PubMed: 25713360]
126. Johannes K, Martin K, Bernd S, Maria-Gabriele S, Christian H, Ravi Kumar M, Erhard S, Aleš S. *Nat. Chem. Biol.* 2010; 6:261–263. [PubMed: 20190763]
127. Gonzalez DJ, Haste NM, Hollands A, Fleming TC, Hamby M, Pogliano K, Nizet V, Dorrestein PC. *Microbiology.* 2011; 157:2485–2492. [PubMed: 21719540]
128. Shih C-JJ, Chen P-YY, Liaw C-CC, Lai Y-MM, Yang Y-LL. *Nat. Prod. Rep.* 2014; 31:739–755. [PubMed: 24452118]
129. Harn YC, Powers MJ, Shank EA, Jojic V. *Bioinformatics.* 2015; 31:42–50.
130. Wei-Ting L, Roland DK, Yu-Liang Y, Bradley SM, Pieter CD. *J. Am. Chem. Soc.* 2011; 133:18010–18013. [PubMed: 21999343]
131. Traxler MF, Watrous JD, Alexandrov T, Dorrestein PC. *MBio.* 2013; 4:e00459–00413. [PubMed: 23963177]
132. Vaidyanathan S, Fletcher J, Goodacre R, Lockyer N, Micklefield J, Vickerman J. *Anal. Chem.* 2008; 80:1942–1951. [PubMed: 18290669]
133. Vaidyanathan S, Fletcher J, Lockyer N, Vickerman J. *Appl. Surf. Sci.* 2008; 255:922–925.

134. Esquenazi E, Dorrestein PC, Gerwick WH. *Proc. Natl. Acad. Sci. U. S. A.* 2009; 106:7269–7270. [PubMed: 19416917]
135. Bhardwaj C, Hanley L. *Nat. Prod. Rep.* 2014; 31:756–767. [PubMed: 24473154]
136. Bouslimani A, Sanchez LM, Garg N, Dorrestein PC. *Nat. Prod. Rep.* 2014; 31:718–729. [PubMed: 24801551]
137. Patti GJ, Yanes O, Siuzdak G. *Nat. Rev. Mol. Cell Biol.* 2012; 13:263–269. [PubMed: 22436749]
138. Orchard S, Hermjakob H, Apweiler R. *Proteomics.* 2003; 3:1374–1376. [PubMed: 12872238]
139. Hermjakob H. *Proteomics.* 2006; 6:34–38. [PubMed: 17031794]
140. Pedrioli PGA, Eng JK, Hubley R, Vogelzang M. *Nat. Biotechnol.* 2004; 22:1459–1466. [PubMed: 15529173]
141. Lin SM, Zhu L, Winter AQ, Sasinowski M. *Expert Rev. Proteomics.* 2005; 2:839–845. [PubMed: 16307524]
142. Martens L, Chambers M, Sturm M, Kessner D. *Mol. Cell Proteomics.* 2011:10.
143. Kessner D, Chambers M, Burke R, Agus D. *Bioinformatics.* 2008; 24:2534–2536. [PubMed: 18606607]
144. Hendriks M, Eeuwijk FA, Jellema RH, Westerhuis JA, Reijmers TH, Hoefsloot HCJ, Smilde AK. *Trac-Trend. Anal. Chem.* 2011; 30:1685–1698.
145. Mahieu NG, Genenbacher JL, Patti GJ. *Curr. Opin. Chem. Biol.* 2016; 30:87–93. [PubMed: 26673825]
146. Liland KH. *Trac-Trend. Anal. Chem.* 2011; 30:827–841.
147. Willem W, Phalp JM, Alan WP. *Anal. Chem.* 1996; 68:3602–3606.
148. Hastings CA, Norton SM, Roy S. *Rapid Commun. Mass Spectrom.* 2002; 16:462–467. [PubMed: 11857732]
149. Roy S, Shaler TA, Hill LR, Norton S, Kumar P. *Anal. Chem.* 2003; 75:4818–4826. [PubMed: 14674459]
150. Monroe ME, Toli N, Jaitly N, Shaw JL, Adkins JN. *Bioinformatics.* 2007; 23:2021–2023. [PubMed: 17545182]
151. Mueller LN, Rinner O, Schmidt A, Letarte S. *Proteomics.* 2007; 7:3470–3480. [PubMed: 17726677]
152. Sturm M, Bertsch A, Gröpl C, Hildebrandt A. *BMC Bioinformatics.* 2008:9. [PubMed: 18182098]
153. Rolf D, Dan B, Karin EM. *Anal. Chim. Acta.* 2002; 454:167–184.
154. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. *Anal. Chem.* 2006; 78:779–787. [PubMed: 16448051]
155. Tautenhahn R, Böttcher C, Neumann S. *BMC bioinformatics.* 2007:9. [PubMed: 17214880]
156. Du P, Kibbe WA, Lin SM. *Bioinformatics.* 2006; 22:2059–2065. [PubMed: 16820428]
157. French WR, Zimmerman LJ, Schilling B, Gibson BW, Miller CA, Townsend RR, Sherrod SD, Goodwin CR, McLean JA, Tabb DL. *J. Proteome Res.* 2014; 14:1299–1307. [PubMed: 25411686]
158. Aoshima K, Takahashi K, Ikawa M, Kimura T, Fukuda M, Tanaka S, Parry HE, Fujita Y, Yoshizawa AC, Utsunomiya S.-i. Kajihara S, Tanaka K, Oda Y. *BMC Bioinformatics.* 2014; 15:1–14. [PubMed: 24383880]
159. Dettmer K, Aronov PA, Hammock BD. *Mass Spectrom. Rev.* 2007; 26:51–78. [PubMed: 16921475]
160. Frenzel T, Miller A, Engel KH. *Eur. Food Res. Technol.* 2003; 216:335–342.
161. Podwojski K, Fritsch A, Chamrad DC, Paul W. *Bioinformatics.* 2009; 25:758–764. [PubMed: 19176558]
162. Cleveland WS, Grosse E. *Stat. Comput.* 1991; 1:47–62.
163. Perera V, Zabala MDT, Florance H, Smirnoff N, Grant M. *Metabolomics.* 2012; 8:175–185.
164. Lange E, Gröpl C, Schulz-Trieglaff O, Leinenbach A, Huber C, Reinert K. *Bioinformatics (Oxford, England).* 2007; 23:273–281.
165. Lommen A. *Anal. Chem.* 2009; 81:3079–3086. [PubMed: 19301908]

166. Bork C, Ng K, Liu Y, Yee A. *Biotechnol Prog.* 2013; 29:394–402. [PubMed: 23292764]
167. Voss B, Hanselmann M, Renard BY, Lindner MS, Köthe U, Kirchner M, Hamprecht FA. *Bioinformatics (Oxford, England).* 2011; 27:987–993.
168. Katajamaa M, Oresic M. *BMC bioinformatics.* 2004;6. [PubMed: 14736341]
169. Pluskal T, Castillo S, Villar-Briones A, Oresic M. *BMC bioinformatics.* 2010;11. [PubMed: 20053295]
170. Lange E, Tautenhahn R, Neumann S. *BMC Bioinformatics.* 2008;9. [PubMed: 18182098]
171. Libiseller G, Dvorzak M, Kleb U, Gander E, Eisenberg T, Madeo F, Neumann S, Trausinger G, Sinner F, Pieber T, Magnes C. *BMC bioinformatics.* 2014;16. [PubMed: 24428894]
172. Worley PR. *Curr. Metabolomics.* 2013; 1:92–107. B. [PubMed: 26078916]
173. Fonville JM, Richards SE, Barton RH. *J. Chemometr.* 2010; 24:636–649.
174. Wiley Interdiscip. Rev. Comput. Stat. 2010; 2:97–106. A. H.
175. Rosipal R, Krämer N. *Lect. Notes Comput. Sc.* 2006; 3940:34–51.
176. Bro R, Smilde AK. *Anal. Methods.* 2014; 6:2812–2831.
177. Bylesjö M, Rantalainen M, Cloarec O. *J. Chemometr.* 2006; 20:341–351.
178. Genilloud O, González I, Salazar O, Martín J, Tormo JRR, Vicente F. *J. Ind. Microbiol. Biotechnol.* 2011; 38:375–389. [PubMed: 20931260]
179. Hou Y, Braun DR, Michel CR, Klassen JL, Adnani N, Wyche TP, Bugni TS. *Anal. Chem.* 2012; 84:4277–4283. [PubMed: 22519562]
180. Macintyre L, Zhang T, Viegelmann C, Martinez I, Cheng C, Dowdells C, Abdelmohsen U, Gernert C, Hentschel U, Edrada-Ebel R. *Mar. Drugs.* 2014; 12:3416–3448. [PubMed: 24905482]
181. Forner D, Berrué F, Correa H, Duncan K, Kerr RG. *Anal. Chim. Acta.* 2013; 805:70–79. [PubMed: 24296145]
182. Norazwana S, Pei Jean T, Khozirah S, Faridah A, Hong Boon L. *Anal. Chem.* 2014; 86:1324–1331. [PubMed: 24405504]
183. Grond S, Papastavrou I, Zeeck A. *Eur. J. Org. Chem.* 2002; 2002:3237–3242.
184. Goodwin CR, Covington BC, Derewacz DK, McNees CR, Wikswo JP, McLean JA, Bachmann BO. *Chem. Biol.* 2015; 22:661–670. [PubMed: 25937311]
185. Krug D, Zurek G, Schneider B, Garcia R, Müller R. *Anal. Chim. Acta.* 2008; 624:97–106. [PubMed: 18706314]
186. Cortina NS, Krug D, Plaza A, Revermann O, Müller R. *Angew. Chem. Int. Ed. Engl.* 2012; 51:811–816. [PubMed: 22162209]
187. Krug D, Zurek G, Revermann O, Vos M, Velicer GJ, Muller R. *Appl. Environ. Microbiol.* 2008; 74:3058–3068. [PubMed: 18378661]
188. Wiklund S, Johansson E, Sjöström L. *Anal. Chem.* 2008; 80:115–122. [PubMed: 18027910]
189. González-Menéndez V, Pérez-Bonilla M, Pérez-Victoria I, Martín J, Muñoz F, Reyes F, Tormo J, Genilloud O. *Molecules.* 2016; 21:234.
190. Hur M, Campbell AA, Almeida-de-Macedo M, Li L, Ransom N, Jose A, Crispin M, Nikolau BJ, Wurtele ES. *Nat. Prod. Rep.* 2013; 30:565–583. [PubMed: 23447050]
191. Albright JC, Henke MT, Soukup AA, McClure RA, Thomson RJ, Keller NP, Kelleher NL. *ACS Chem. Biol.* 2015; 10:1535–1541. [PubMed: 25815712]
192. Tautenhahn R, Patti GJ, Rinehart D, Siuzdak G. *Anal. Chem.* 2012; 84:5035–5039. [PubMed: 22533540]
193. Gowda H, Ivanisevic J, Johnson CH, Kurczy ME, Benton HP, Rinehart D, Nguyen T, Ray J, Kuehl J, Arevalo B, Westenskow PD, Wang J, Arkin AP, Deutschbauer AM, Patti GJ, Siuzdak G. *Anal. Chem.* 2014; 86:6931–6939. [PubMed: 24934772]
194. Xia J, Wishart DS. *Curr. Protoc. Bioinformatics.* 2011 Chapter14: Unit14.
195. Xia J, Sinelnikov IV, Han B, Wishart DS. *Nucl. Acids Res.* 2015; 43:W251–257. [PubMed: 25897128]
196. Patti GJ, Tautenhahn R, Rinehart D, Cho K, Shriver L, Manchester M, Nikolskiy I, Johnson C, Mahieu N, Siuzdak G. *Anal. Chem.* 2013; 85:798–804. [PubMed: 23206250]

197. Goodwin CR, Sherrod SD, Marasco CC, Bachmann BO, Schramm-Sapyta N, Wikswo JP, McLean JA. *Analytical chemistry*. 2014; 86:6563–6571. [PubMed: 24856386]
198. Gaudêncio SP, Pereira F. *Nat. Prod. Rep.* 2015; 32:779–810. [PubMed: 25850681]
199. Eugster PJ, Boccard J, Debrus B, Bréant L, Wolfender J-L, Martel S, Carrupt P-A. *Phytochemistry*. 2014; 108:196–207. [PubMed: 25457501]
200. Stow SM, Goodwin CR, Kliman M. *J. Phys. Chem. B*. 2014; 118:13812–13820. [PubMed: 25360896]
201. Colin AS, Grace OM, Elizabeth JW, Chuan Q, Sunia AT, Theodore RB, Darlene EC, Ruben A, Gary S. *Ther. Drug Monit.* 2005; 27:747–751. [PubMed: 16404815]
202. 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 MY, 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. *J. Mass Spectrom.* 2010; 45:703–714. [PubMed: 20623627]
203. Shen H, Dührkop K, Böcker S, Rousu J. *Bioinformatics*. 2014:30.
204. Bérdy J. *J. Antibiot. (Tokyo)*. 2012; 65:385–395. [PubMed: 22511224]
205. Hufsky F, Scheubert K, Böcker S. *Nat. Prod. Rep.* 2014; 31:807–817. [PubMed: 24752343]
206. Allard P-MM, Péresse T, Bisson J, Gindro K, Marcourt L, Pham VC, Roussi F, Litaudon M, Wolfender J-LL. *Anal. Chem.* 2016; 88:3317–3323. [PubMed: 26882108]
207. Watrous J, Roach P, Alexandrov T, Heath BS, Yang JY, Kersten RD, van der Voort M, Pogliano K, Gross H, Raaijmakers JM, Moore BS, Laskin J, Bandeira N, Dorrestein PC. *Proc. Natl. Acad. Sci. U. S. A.* 2012; 109:1743–1752. [PubMed: 22232671]
208. Yang JY, Sanchez LM, Rath CM, Liu X, Boudreau PD, Bruns N, Glukhov E, Wodtke A, de Felicio R, Fenner A, Wong WR, Linington RG, Zhang L, Debonsi HM, Gerwick WH, Dorrestein PC. *J. Nat. Prod.* 2013; 76:1686–1699. [PubMed: 24025162]
209. Moree WJ, Phelan VV, Wu C-H, Bandeira N, Cornett DS, Duggan BM, Dorrestein PC. *Proc. Natl. Acad. Sci. U. S. A.* 2012; 109:13811–13816. [PubMed: 22869730]
210. Briand E, Bormans M, Gugger M, Dorrestein PC, Gerwick WH. *Environ Microbiol.* 2015; 18:384–400. [PubMed: 25980449]
211. Duncan KR, Crüsemann M, Lechner A, Sarkar A, Li J, Ziemert N, Wang M, Bandeira N, Moore BS, Dorrestein PC, Jensen PR. *Chem. Biol.* 2015; 22:460–471. [PubMed: 25865308]
212. Kleigrew K, Almaliti J, Tian IY, Kinnel RB. *J. Nat. Prod.* 2015; 78:1671–1682. [PubMed: 26149623]
213. Vizcaino MI, Crawford JM. *Nat. Chem.* 2015; 7:411–417. [PubMed: 25901819]
214. Trautman CJ. *Curr. Top. Med. Chem.* 2015; 16:1705–1716. EP.
215. Vizcaino MI, Engel P, Trautman E, Crawford JM. *J. Am. Chem. Soc.* 2014; 136:9244–9247. [PubMed: 24932672]
216. Luo Y, Li B-Z, Liu D, Zhang L, Chen Y, Jia B, Zeng B-X, Zhao H, Yuan Y-J. *Chem. Soc. Rev.* 2015; 44:5265–5290. [PubMed: 25960127]
217. Zhang H, Boghigian BA, Armando J. *Nat. Prod. Rep.* 2011; 28:125–151. [PubMed: 21060956]
218. Donia, Ruffner DE, Cao S, Schmidt EW. *Chembiochem.* 2011; 12:1230–1236. [PubMed: 21542088]
219. Schorn M, Zettler J, Noel JP, Dorrestein PC. *ACS Chem. Biol.* 2013; 9:301–309. [PubMed: 24168704]
220. Gonzalez DJ, Corriden R, Akong-Moore K, Olson J. *Chem. Biol.* 2014; 21:1457–1462. [PubMed: 25457179]
221. Gonzalez DJ, Vuong L, Gonzalez IS, Keller N, McGrosso D, Hwang JH, Hung J, Zinkernagel A, Dixon JE, Dorrestein PC. *Mol. Cell Proteomics.* 2014; 13:1262–1272. [PubMed: 24567418]
222. Macherla VR, Liu J, Sunga M, White DJ. *J. Nat. Prod.* 2007; 70:1454–1457. [PubMed: 17845000]
223. Desjardine K, Pereira A, Wright H. *J. Nat. Prod.* 2007; 70:1850–1853. [PubMed: 18044840]
224. McArthur KA, Mitchell SS, Tsueng G. *J. Nat. Prod.* 2008; 71:1732–1737. [PubMed: 18842058]

225. Teasdale ME, Shearer TL, Engel S. *J. Org. Chem.* 2012; 77:8000–8006. [PubMed: 22920243]
226. Kellogg JJ, Todd DA, Egan JM, Raja HA. *J. Nat. Prod.* 2016; 79:376–386. [PubMed: 26841051]
227. Salvador-Reyes LA, Luesch H. *Nat. Prod. Rep.* 2015; 32:478–503. [PubMed: 25571978]
228. Johnston CW, Skinnider MA, Dejong CA, Rees PN, Chen GM, Walker CG, French S, Brown ED, Bérdy J, Liu DY, Magarvey NA. *Nat. Chem. Biol.* 2016 DOI: 10.1038/nchembio.2018.
229. Rees MG, Seashore-Ludlow B, Cheah JH, Adams DJ, Price EV, Gill S, Javaid S, Coletti ME, Jones VL, Bodycombe NE, Soule CK, Alexander B, Li A, Montgomery P, Kotz JD, Hon CS, Munoz B, Liefeld T, Dan ík V, Haber DA, Clish CB, Bittker JA, Palmer M, Wagner BK, Clemons PA, Shamji AF, Schreiber SL. *Nat. Chem. Biol.* 2015; 12:109–116. [PubMed: 26656090]
230. Farha MA, Brown ED. *Nat. Prod. Rep.* 2016 DOI: 10.1039/C5NP00127G.
231. Wong WR, Oliver AG, Linington RG. *Chem. Biol.* 2012; 19:1483–1495. [PubMed: 23177202]
232. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC. *Science.* 2006; 313:1929–1935. [PubMed: 17008526]
233. Hutter B, Schaab C, Albrecht S, Borgmann M, Brunner NA, Freiberg C, Ziegelbauer K, Rock CO, Ivanov I, Loferer H. *Antimicrob. Agents Chemother.* 2004; 48:2838–2844. [PubMed: 15273089]
234. Hu Y, Potts MB, Colosimo D, Herrera-Herrera ML, Legako AG, Yousufuddin M, White MA, MacMillan JB. *J. Am. Chem. Soc.* 2013; 135:13387–13392. [PubMed: 23984625]
235. Potts MB, Kim HS, Fisher KW, Hu Y, Carrasco YP. *Sci. Signal.* 2013:6.
236. Yu Y, Yi Z, Liang YZ. *FEBS Lett.* 2007; 581:4179–4183. [PubMed: 17692848]
237. Liu Y, Wen J, Wang Y, Li Y, Xu W. *Chromatographia.* 2010; 71:253–528.
238. Halouska S, Chacon O, Fenton RJ, Zinniel DK, Barletta RG, Powers R. *J. Proteome Res.* 2007; 6:4608–4614. [PubMed: 17979227]
239. Halouska S, Fenton RJ, Barletta RG, Powers R. *ACS Chem. Biol.* 2012; 7:166–171. [PubMed: 22007661]
240. Vincent IM, Weidt S, Rivas L, Burgess K, Smith TK, Ouellette M. *Int. J. Parasitol. Drugs Drug Resist.* 2014; 4:20–27. [PubMed: 24596665]
241. Vincent IM, Creek DJ, Burgess K, Woods DJ. *PLoS Negl. Trop. Dis.* 2012; 6:e1618. [PubMed: 22563508]
242. Trochine A, Creek DJ, Faral-Tello P, Barrett MP. *PLoS Negl. Trop. Dis.* 2014; 8:e2844. [PubMed: 24853684]
243. Vincent IM, Ehmann DE, Mills S, Perros M. *Antimicrob. Agents Ch.* 2016; 4:20–27.
244. Lobritz MA, Belenky P, Porter CB, Gutierrez A, Yang JH, Schwarz EG, Dwyer DJ, Khalil AS, Collins JJ. *Proc. Natl. Acad. Sci. U. S. A.* 2015; 112:8173–8180. [PubMed: 26100898]
245. Perlman ZE, Slack MD, Feng Y, Mitchison TJ. *Science.* 2004; 306:1194–1198. [PubMed: 15539606]
246. Woehrmann MH, Bray WM, Durbin JK, Nisam SC. *Mol. Biosyst.* 2013; 9:2604–2617. [PubMed: 24056581]
247. Peach KC, Bray WM, Winslow D, Linington PF. *Mol. Biosyst.* 2013; 9:1837–1848. [PubMed: 23609915]
248. Navarro G, Cheng AT, Peach KC. *Antimicrob. Agents Chemother.* 2014; 58:1092–1099. [PubMed: 24295976]
249. Schulze CJ, Bray WM, Woehrmann MH, Stuart J. *Chem. Biol.* 2013; 20:285–295. [PubMed: 23438757]
250. Kurita KL, Glassey E, Linington RG. *Proc. Natl. Acad. Sci. U. S. A.* 2015; 112:11999–12004. [PubMed: 26371303]
251. Schmidt EW. *Nat. Chem.* 2015; 7:375–376. [PubMed: 25901811]
252. Sharon G, Garg N, Debelius J, Knight R, Dorrestein PC, Mazmanian SK. *Cell Metab.* 2014; 20:719–730. [PubMed: 25440054]
253. Hooper LV, Littman DR, Macpherson AJ. *Science.* 2012; 336:1268–1273. [PubMed: 22674334]

254. Clarke TB, Davis KM, Lysenko ES, Zhou AY, Yu Y. *Nat. Med.* 2010; 16:228–231. [PubMed: 20081863]
255. Koppel N, Balskus EP. *Cell Chem. Biol.* 2016; 23:18–30. [PubMed: 26933733]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

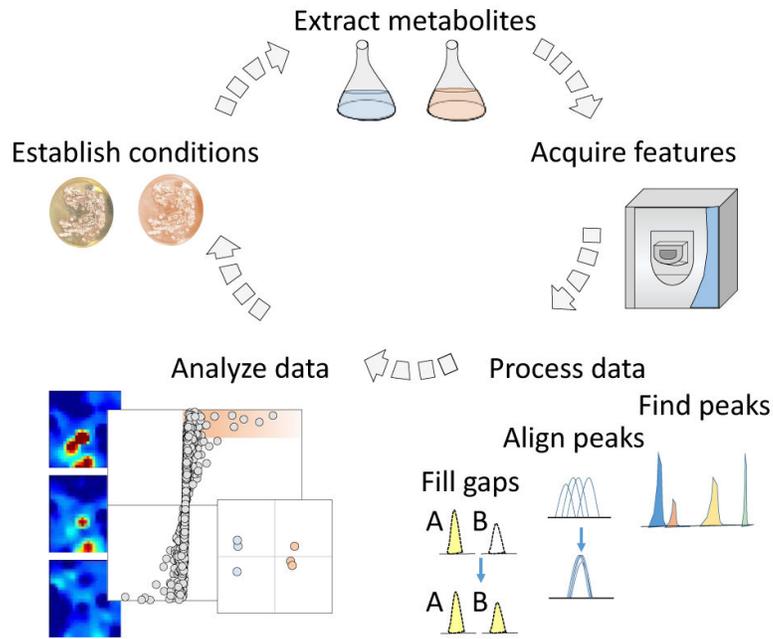


Figure 1.

General metabolomics workflow. metabolites are extracted from experimental conditions and detected through ms analysis. ms data is then formatted and processed before undergoing statistical analyses to determine important metabolomic changes between the sample groups. these results may then be used to direct new experiments to optimize secondary metabolite production or test biological hypotheses generated from the initial experiment.

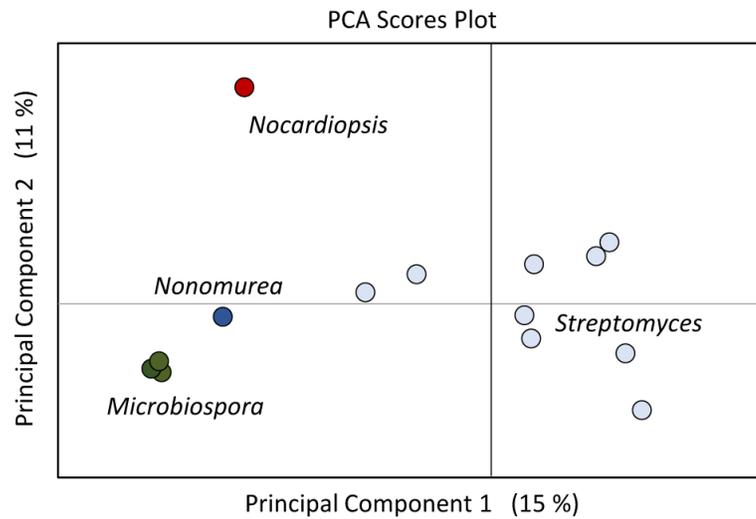


Figure 2.

Using a pca scores plot to prioritize microbial producers. a panel of actinomycetes including microbiospora, streptomyces, and nonomurea genera. in this analysis, 14 strains grown under identical conditions were compared and principal component analysis was used to display metabolomic feature variance between the strains. principal component 1 primarily groups streptomyces from other strains, and component 2 further distinguishes nocardioopsis sp. fu40 as metabolomically unique compared to other tested strains. percentages shown in parentheses correspond to the variance between the samples contained within the specific component.

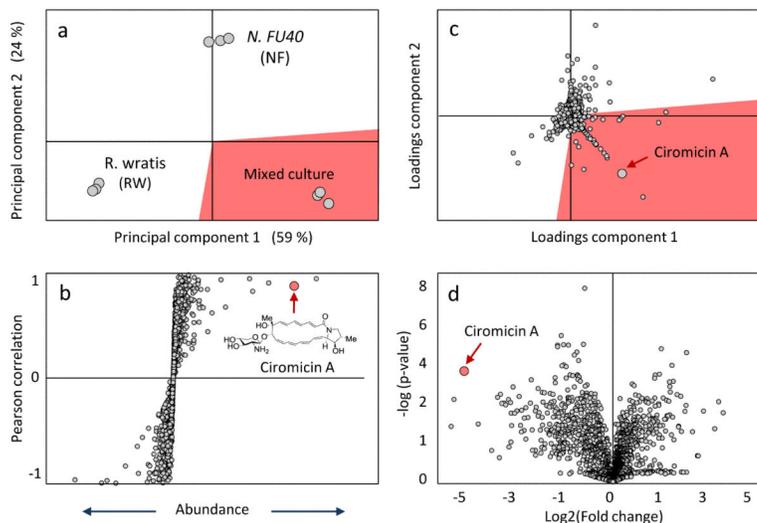


Figure 3. mvsa s-, loadings, and volcano plots to identify induced features. (a) the scoresplot reveals group separation between the nocardiosis monoculture (nf), the rhodococcus wratis competitor monoculture (rw), and the mixed culture (rw&nf). (b) s-plot shows ciromicin significantly correlates ($p < 0.1$) to the mixed culture group in a binary comparison vs the nocardiosis monoculture. (c) loadings plot of features shows ciromicin contributes significantly to group differentiation on the pca scoresplot. (d) volcano plot also prioritizes ciromicin ions which have a high correlation (low p-value) on the y-axis and high fold change on the x-axis. shading in panels a and c are used to highlight the data corresponding to the different sample subtypes.

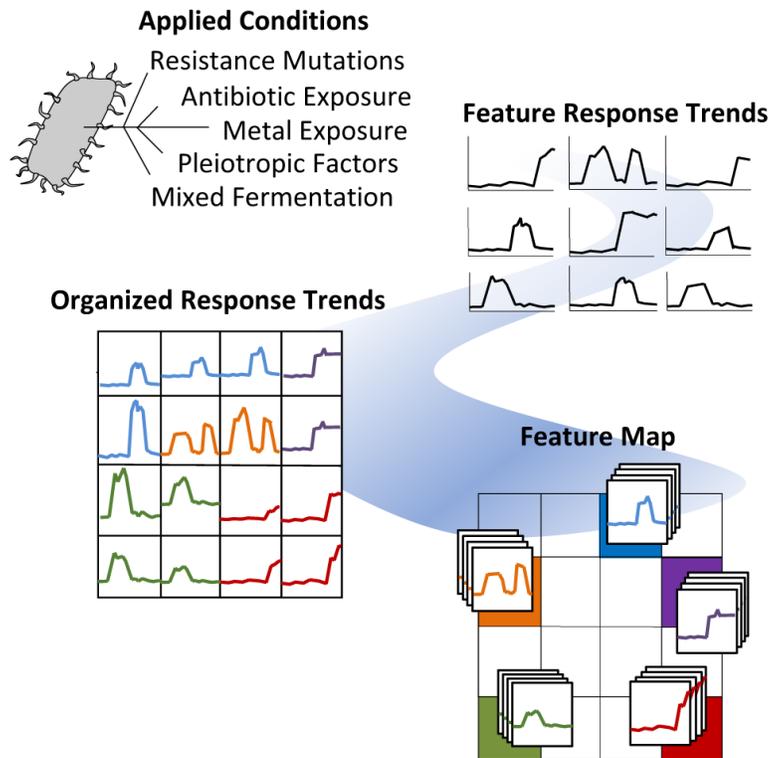


Figure 4. Feature organization within a self-organizing map analysis. feature abundance profiles are illustrated for each feature as a response trend across all experimental conditions shown in the upper right. these trends are organized for similarity as shown on the bottom right. these organized data serve as the basis for visual heatmap representations of the observed metabolomic content of experimental cultures.

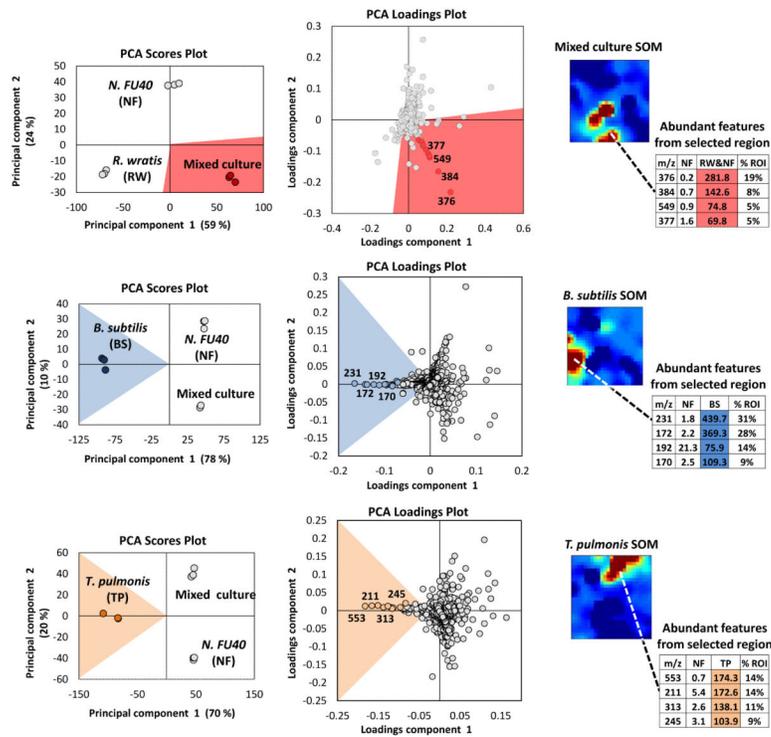


Figure 5. Three example comparisons of prioritized features through principal component and self-organizing map analyses on mixed cultures with nocardiosis fu40 (nf), rhodococcus wratis (rw), tsukamurella pulmonis (tp), and bacillus subtilis (bs). features prioritized within som regions of interest recapitulate pca tripartite analyses when sorted by abundance, or percentage of the region of interest (% roi). shading in scores and loadings plots used to highlight the data corresponding to the different sample subtypes.

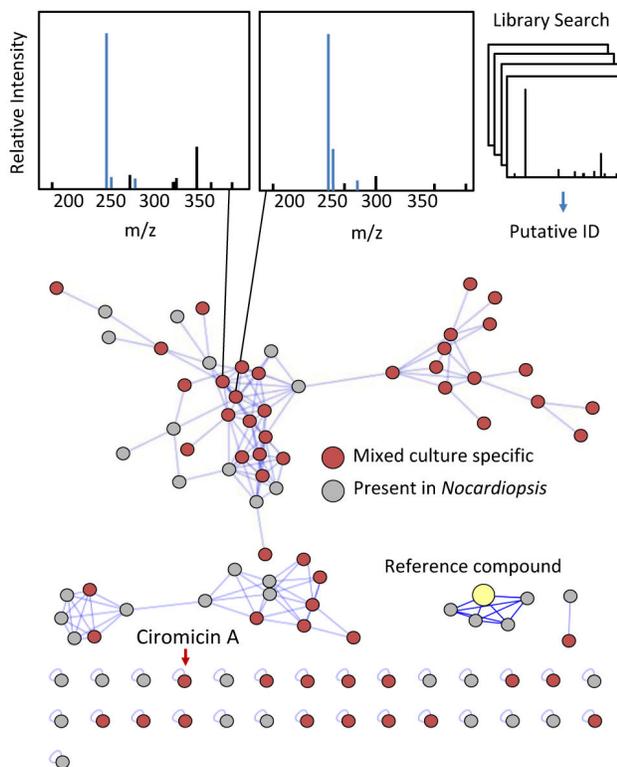


Figure 6. Applications of molecular networking to explore data. comparisons of acquired fragmentation spectra to established databases facilitates putative feature identification. connectivity between features shown with blue lines relates structural similarities. reference compounds seeded into the network can identify structural analogs. feature distributions between experimental conditions are indicated by node colouring, red for mixed culture specific, and grey for features detected within the monoculture.

Table 1

Overview of analytical descriptors relevant to metabolomics-based natural product discovery

Analytical Descriptor	Description		Analytical technique
Mass accuracy	Deviation of the experimentally determined m/z from the true m/z . Expressed as the mass error (<i>e.g.</i> ppm), with sufficiently small error an exact chemical formula can be determined.	Mass analyzer	
		•	Space-dispersive (<i>e.g.</i> ion trap, quadrupole)
		•	Time-dispersive (<i>e.g.</i> Time of flight)
Isotopic modeling	Comparison of the abundances of specific isotopes in the molecular isotopic envelope. Can provide rapid indication of amount and identity of heteroatoms.	Chemometrics	
		•	Theoretical isotope calculators
		•	Mass defect analysis
		•	Quantitation
Chromatographic retention time	Time required for fluid-solid phase partitioning across a column. Provides separation on the basis of a differentiating characteristic orthogonal to mass.	Chromatography	
		•	Hydrophathy (<i>e.g.</i> Liquid chromatography)
		•	Volatility (<i>e.g.</i> Gas Chromatography)
		•	Size and charge (<i>e.g.</i> Size exclusion, Charge exclusion, Ion capture)
Ion mobility drift time	Gas-phase electrophoretic separation based on size and shape of the metabolite as ions pass through a gas filled drift tube.	Ion drift tube	
		•	Time-dispersive (<i>e.g.</i> Drift time ion mobility, Traveling-wave ion mobility)
		•	Space-dispersive (<i>e.g.</i> Field-asymmetric ion mobility)
Fragmentation	Tandem MS using ion activation to provide characteristic fragment species. Provides metabolite structural information to prioritize which of multiple isomers are the likely identity for a given elemental formula.	Ion activation	
		•	Collisional (<i>e.g.</i> Collision induced dissociation, and Surface induced dissociation)
		•	Electron (<i>e.g.</i> Electron transfer dissociation and Electron capture dissociation)
		•	Photoactivation (<i>e.g.</i> Infrared multiphoton dissociation and Wavelength-Tunable Ultraviolet Photodissociation)

Table 2

Overview of methods for metabolomic data analysis

Method	Description	Applications	Disadvantages
Principle component analysis and Projections to latent structures	MVSA to identify significant covariance within data	<ul style="list-style-type: none"> • Identifying data outliers • Strain prioritization • Grouping samples • Compound prioritization 	<ul style="list-style-type: none"> • Less effective with large datasets
Self organizing Maps	Organizes features into a 2-dimensional map based on feature response trends across a variety of experimental conditions	<ul style="list-style-type: none"> • Grouping samples • Compound prioritization • Comparing large numbers of experimental conditions 	<ul style="list-style-type: none"> • Less effective with small numbers of conditions
Molecular networking	Organizes features into a connectivity network based on similarities in molecular fragmentation patterns	<ul style="list-style-type: none"> • Compound prioritization • Compound dereplication 	<ul style="list-style-type: none"> • Fragmentation can vary with instrument parameters