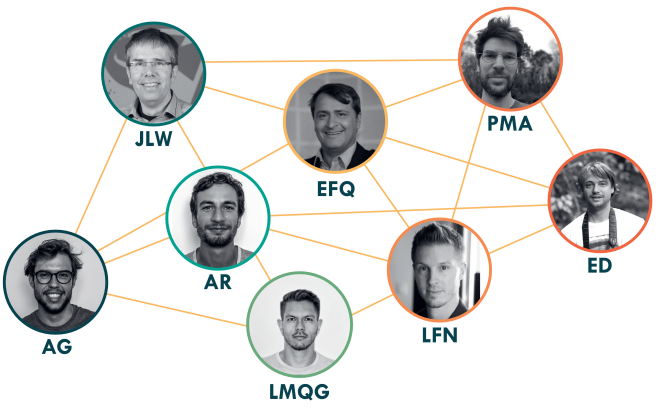
doi:10.2533/chimia.2022.954 Chimia 76 (2022) 954–963

Metabolomics in Ecology and Bioactive Natural Products Discovery: Challenges and Prospects for a Comprehensive Study of the Specialised Metabolome

Jean-Luc Wolfender[ab]*, Arnaud Gaudry[ab], Adriano Rutz[ab], Luis-Manuel Quiros-Guerrero[ab], Louis-Félix Nothias[ab], Emerson Ferreira Queiroz[ab], Emmanuel Defossez[cd], and Pierre-Marie Allard[abc]

Abstract: Metabolomics is playing an increasingly prominent role in chemical ecology and in the discovery of bioactive natural products (NPs). The identification of metabolites is a common/central objective in both research fields. NPs have significant biological properties and play roles in multiple chemical-ecological interactions. Classically, in pharmacognosy, NP chemical structure is determined after a complex process of isolating and interpreting spectroscopic data. With the advent of powerful analytical techniques such as liquid chromatography-mass spectrometry (LC-MS) the annotation process of the specialised metabolome of plants and microorganisms has improved considerably. In this article, we summarise the possibilities opened by these advances and illustrate how we harnessed them in our own research to automate annotations of NPs and target the isolation of key compounds. In addition, we are also discussing the analytical and computational challenges associated with these emerging approaches and their perspective.

Keywords: Bioactivity · Ecology · Metabolomics · Natural products · Structural elucidation



Jean-Luc Wolfender (JLW) is Professor at the University of Geneva where he leads the phytochemistry bioactive natural product (PBNP) group. In the 1990s he helped introduce LC-MS and LC-NMR for the profiling of natural extracts for dereplication for accelerating the discovery of novel bioactive compounds. He is currently developing innovative MS- and NMR-based metabolomics approaches in various aspects of natural products research. He is interested in NP-based drug discovery and evidence-based phytotherapy. His research also covers the search for inducible NPs in response to stimuli in microbial interactions and plant defense.

Arnaud Gaudry (AG) holds a MSc in pharmaceutical sciences from the University of Geneva (2018). He joined the PBNP group as a PhD candidate in 2018, under the supervision of Prof. Wolfender. As a pharmacist inspired about the potential of natural products, he is currently working on the development of computational methods for the efficient identification of bioactive compounds in natural extracts. The developed methods are applied to identify new antiparasitic compounds in a collaborative project involving the Swiss Tropical and Public Health Institute (STPH) and the Drug for Neglected Diseases Initiative (DNDi).

Adriano Rutz (AR) holds a PhD in pharmaceutical sciences from the University of Geneva (2022). He is now developing generic mass spectrometry-based metabolite profiling methods and tools to assess phytochemical and sensorial profiles of complex plant extracts. He is one of the founders of the LOTUS initiative for open knowledge management in natural products research.

Louis-Félix Nothias (LFN) is a researcher in the PBNP group since 2021. He received his PhD (2015) from the University of Corsica in co-direction with the Institut de Chimie des Substances Naturelles (CNRS, Université Paris-Saclay) where he studied metabolites from Mediterranean Euphorbia plants. In 2016, he joined the laboratory of Prof. Dorrestein (University of California San Diego) as a post-doctoral researcher to develop computational methods for mass spectrometry-based metabolomics analysis. He is also pioneering the use of integrative multi-omics approaches that can uncover microbially-related molecules in complex samples such as holobiont's.

Luis Quiros-Guerrero (LMQG) is a chemist with a MSc in Chemistry (2015) from the University of Costa Rica. His work was focused on the development of chemotaxonomic models with LC-MS. Since 2018 he has been part of the PBNP group, developing methodologies to discover chemical novelty in natural extracts libraries through the development of bioinformatic tools. He actively participates in several projects related to food and plant metabolomics, isolation of bio active natural products from diverse natural sources and chemical ecology.

*Correspondence: Prof. J.-L. Wolfender[ab], E-mail: jean-luc.wolfender@unige.ch, [a]Institute of Pharmaceutical Sciences of Western Switzerland, University of Geneva, CMU, CH-1211 Geneva, Switzerland; [b]School of Pharmaceutical Sciences, University of Geneva, CH-1211 Genève 4, Switzerland; [c]Department of Biology, University of Fribourg, CH-1700 Fribourg, Switzerland; [d]Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland

Emerson F. Queiroz (EFQ) holds a PhD in pharmaceutical sciences from the University of Paris-Sud. In 1999 he joined Prof. Hostettmann's group at the University of Lausanne for post-doctoral training. From 2006 to 2011 he worked as the head of the research and development department of Aché Laboratories in Brazil. Since 2011 he is a senior research associate in the PBNP group at the University of Geneva. His main research interests include the discovery of active molecules from diverse origins, biotransformation for generation of bioactive compounds, and the development of innovative approaches to the isolation and identification of bioactive NPs.

Emmanuel Defossez (ED) obtained his PhD in 2013 from the University of Grenoble (France). After a first post-doc at the Centre de Recherche en Écologie et Écologie Évolutive (CNRS-France, Montpellier), he continued his research as a postdoctoral fellow at the Laboratoire d'Écologie Fonctionnelle of the University of Neuchâtel (Switzerland). Since 2022, he leads the computational ecology research group of the University of Neuchâtel and he is the scientific manager of the Botanical Garden of Neuchatel. His current research focuses on using metabolomics to explore how evolutionary constraints and plant-environment interactions drive species coexistence and shape phytochemical diversity patterns.

Pierre-Marie Allard (PMA) is a researcher at the University of Fribourg where he launched the COMMONS Lab. This lab explores novel knowledge management strategies (from acquisition and organisation to dissemination of knowledge) for natural products research. The COMMONS Lab is strongly committed to the Open Science guidelines. We employ computational metabolomics approaches to explore the chemistry of living systems at large scale and build frameworks to organise and share the acquired information as Linked Open Data. Our main and long-term research objective is to inform, support and participate in biodiversity conservation.

1. Introduction

A precise estimate of the diversity of natural products (NPs) remains challenging to establish, and at present, at least 450,000 specialised metabolites have been fully characterised and documented.[1] It is worth mentioning at this stage that the full structural characterisation of any new NPs involves their isolation/purification and *de novo* structure determination by the interpretation of spectroscopic data (NMR, MS and when necessary, X-ray). NPs structures are reported in peer-reviewed articles which gives all the spectroscopic evidence for their identification. Such data are mainly compiled in the Dictionary Natural Product (DNP) (> 300,000 entries). The content of the DNP database and other chemical databases and the biological origin of all described NPs have been recently described by F. Ntie-Kang and D. Svozil.[1]

Although the biological functions of these NPs are not fully understood, many of these specialised metabolites play a key role in different types of interactions in chemical ecology.[2,3] They are also a historical source of medicines for humans, whether ingested as whole medicinal plants or preparations and or as purified substances in medication. Many of these NPs have been identified as a result of pharmacognosy studies aiming to describe, mainly in plants, the active principles responsible for a given biological activity.[4] Numerous NPs have also been identified in microorganisms during the antibiotics 'golden era'. To this end, investigations were classically carried out using so-called 'bioactivity-guided isolation' approaches (Fig. 1A). This research led to the discovery of a large number of NPs, several of which are at the origin of some of the most widely used drugs (*e.g.* morphine (Q81225), taxol (Q423762)). All structures can be found with their Wikidata Q identifiers. The effectiveness of bioactivity-guided drug discovery approaches was recently recognised when Prof. Tu Youyou was awarded the 2015 Nobel Prize in Medicine for her discovery of artemisinin (Q426921), an effective antimalarial compound from *Artemisia annua* (Q1308044).[5]

Artemisinin is an interesting example of an NP that is of medicinal interest but also plays a role in chemical ecology. This sesquiterpene lactone has been assumed to act both as a defence against insects (bitter principle) and as a phytotoxic allelochemical.[6] This interesting study illustrates the complexity associated with the elucidation of the metabolome's functions in ecological systems. Indeed, even a single compound for which the structure is fully determined can be linked to multiple and confounded effects depending on the numbers and identity of the connected actors. The complexity of this research endeavour (chemical ecology) increases, of course, when considering complete metabolomes for which structural, functional and relational indeterminations exist; the same reasoning applies to biotic interactions in ecological systems. Highlighting connections between these two deeply intricate networks (the chemical network on one side and the ecological network on the other) is a fascinating task but also a major challenge that can only be facilitated by a finer description at both levels.

'Bioactivity-guided isolation' approaches were initially carried out with no or limited prior information on the chemical constituents present in the extracts. From the 1980–90s onwards, analytical approaches have been introduced to rapidly highlight the presence of previously reported NPs, avoid the rediscovery of known active principles, and concentrate efforts on potentially new bioactive molecules. Such chemical profiling carried out upstream of activity-targeted isolation approaches is known as *'dereplication'*[7] and has mainly involved spectroscopic/spectrometric methods coupled to high-performance liquid chromatography (HPLC), a versatile technique for extract analysis.[8] These hyphenated techniques allow to obtain structural information through the interpretation of mass (LC-MS), UV (LC-PDA) and even sometimes NMR spectra (LC-NMR) obtained on-line (Fig. 1 A). In chemical ecology, similar approaches with bioactivity tests, *e.g.* on herbivores, have revealed NPs of fundamental ecological importance and have also driven the development of novel analytical chemistry methods.[9,10] With the development of these methods, more comprehensive approaches emerged in the early 2000s with the advent of so-called 'omics'. In this context, metabolomics, which aims at the most complete description (qualitative and quantitative) of the endogenous metabolites of a biological sample (metabolome), was defined by Fiehn.[11] Since then, this approach has become increasingly adopted in life sciences. Metabolomics allows the detailed metabolite profiling of a given natural extract and for this application, liquid chromatography hyphenated to mass spectrometry (LC-MS) is the primary technique used to generate the metabolite profiling of an extract, with currently unmatched sensitivity and dynamic range.[12] Since its development, metabolomics has thus become an essential tool in biomedical research for the analysis of biological fluids or for the search for biomarkers associated with pathologies.[13] This approach has also been widely adopted in NPs research and also more recently in several fields of ecology.[14–17]

As a result, the analytical approaches classically used for the dereplication of NPs in plant and microorganism extracts have evolved into high-throughput, high-resolution LC-MS metabolite profiling approaches that allow the annotation of NPs for dereplication but also to obtain a global view of their composition. These approaches allow researchers to have a better view of the wide range of metabolites produced by a given organism and offer a new reading of mechanisms in functional ecology.[18] Using metabolomic approaches, molecules of interest can be highlighted in a given extract prior to any preparative-scale separation process and then subjected to targeted isolation for *de novo* identification (Fig. 1B) and bioassay evaluation.

The applications of metabolomics are broad. Combined with the use of advanced multivariate data analysis (MVDA), they range from applied drug discovery objectives to more fundamental aspects such as chemical ecology studies, description of the me-

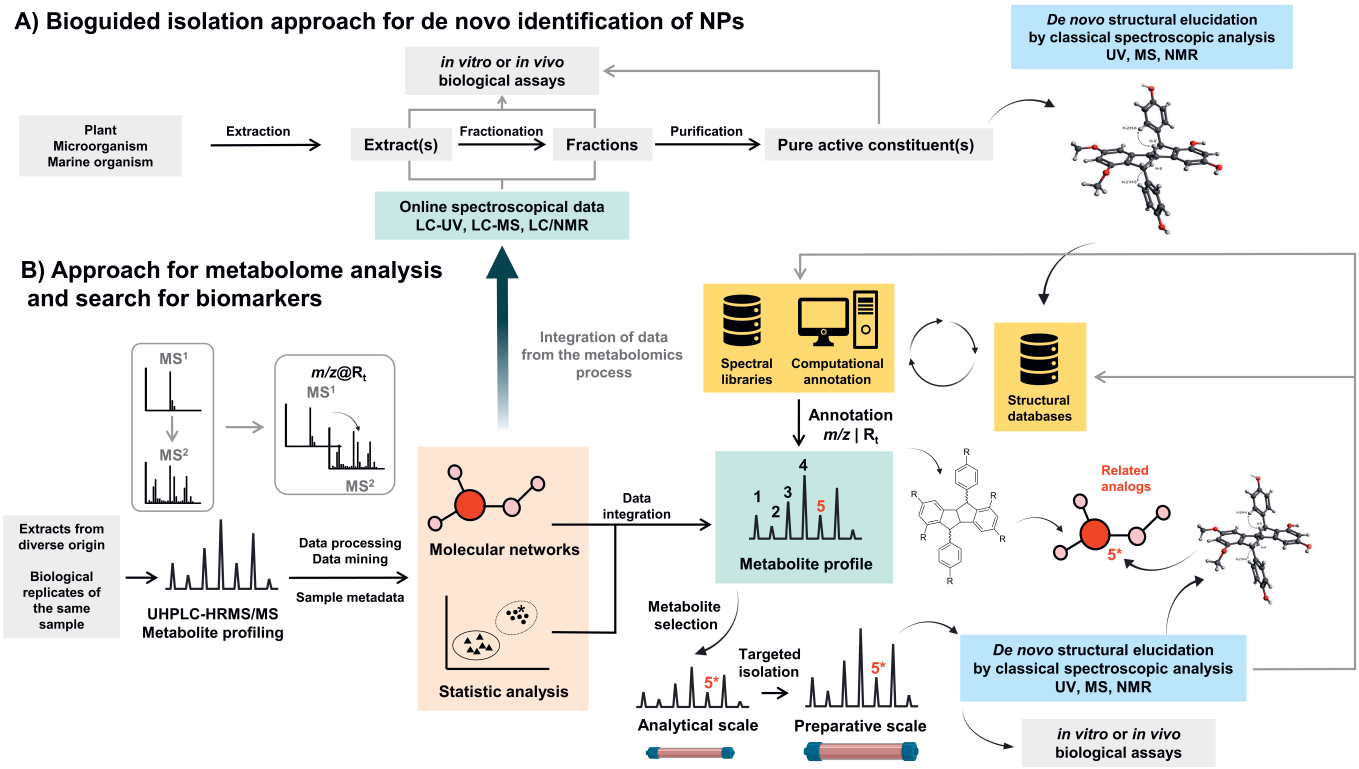


Fig. 1. Overview of the process for NPs identification/annotation and links between A, the bioguided isolation approach, and B, the approach for metabolome analysis and search for biomarkers. In A, after obtaining extracts of various polarities, their bioactivity is evaluated. The isolation is done by combining different preparative chromatography techniques. Fractions are screened for bioactivity until obtaining pure bioactive NPs. Dereplication is performed upfront by a combination of hyphenated methods. The isolated NPs are fully characterised by *de novo* structure elucidation, notably with 1D and 2D NMR and additional techniques. In B, metabolite profiles are acquired on many extracts (*e.g.* organisms obtained under diverse experimental conditions, biological replicates, or biodiverse sets) with untargeted UHPLC-HRMS/MS. HRMS (MS^1) and MS/MS (MS^2) data are used either to build MN for chemical exploration of the data set or for MVDA analysis for searching features that can be linked to chemical markers responsible for changes in the metabolome. Metabolites are annotated based on comparison with MS/MS experimental spectral libraries and completed by matching with *in silico* MS/MS libraries. If unambiguous structure identification or new NP determination is needed based on the annotation, targeted isolation based on metabolites profiles is conducted and led, as in A, to the *de novo* structure identification of **5**. Both A and B feed structural DBs of identified NPs. Similarly, efforts to populate the experimental libraries of unambiguously identified NPs following the FAIR principles (Findable, Accessible, Interoperable, Reusable) must be made to improve annotation tools and efficiency.

tabolome of a studied organism or chemotaxonomy studies.[19,20] Analytical platforms now routinely generate large volumes of spectral data in an untargeted manner for most detected molecules. Compared to more established omics sciences, such as genomics or proteomics, metabolomics deals with much more complex objects (non-polymeric molecules with no amplification step possible), which explains the additional challenges faced by researchers in this field. This is especially true for the annotation of metabolites and establishing clear links to other omics sciences, such as genetics, transcriptomics, and proteomics. Some of these challenges will be discussed below.

In this paper we aim to address the challenges in NP metabolomics from both a drug discovery and an ecology perspective. In particular, we will summarise the most recent developments of tools for metabolome annotation and specifically specialised metabolites identification. Different applications of metabolomics from our own research have been selected to illustrate some of the diverse topics that can be addressed with such holistic approaches.

2. Challenges in the Mining of Metabolite Profiling Data

The metabolite profiling of natural extracts is mainly done by LC-MS methods.[21] Modern spectrometers are capable of high-throughput acquisition of mass spectra with high accuracy, high resolution, high sensitivity, and a wide dynamic range. Together with these developments, HPLC evolved to ultra-high-performance liquid chromatography (UHPLC). UHPLC uses columns containing micrometric particles allowing high resolution analytical separations in a high throughput manner. The coupling of UHPLC with high-resolution mass spectrometry (UHPLC-HRMS) has quickly established itself as the gold standard for acquiring metabolomics data.[22] Additionally, mass spectrometers are working with increasingly higher acquisition frequencies, which allows for single MS (MS^1) analysis to be performed concomitantly with MS/MS (MS^2) spectral acquisitions in an untargeted manner.[23] The fragments observed in fragmentation spectra provide key structural information that can be used to improve molecular formula determination and structural annotation.[21]

Starting from limited amounts of crude extracts (typically in the range of 1–10 μg on column), such analysis generates several thousands of features (LC-MS^1 peaks characterised by their mass-over-charge ratio at a given retention time - *m/z* @ Rt) per sample. When grouping the LC-MS features produced by a given molecule (including isotopologues, adducts, in-source fragments, and multimers),[24] the number of compounds can decrease by an order of magnitude.[25]

To assist the reader, a glossary of the most commonly used terms and concepts in MS metabolomics is presented in Table 1.

The first step of metabolomic data analysis is to process the 'raw' LC-MS data and convert it into an *m*-by-*n* feature table, where *m* is the number of features and *n* the number of samples. This step is often done with open-source software like XCMS,[27]

Table 1. Glossary of terms used in MS-based metabolomics

Dereplication	Process of annotating previously known molecules from a complex sample such as natural extract without requiring their physical isolation at the preparative scale. Annotation is usually based on data obtained by a combination of hyphenated chromatographic methods (LC-UV-PDA, LC-MS, LC-NMR).
Annotation	The action of linking a putative chemical structure to an experimental feature (MS^1, MS^2 and/or MS/NMR).
Identification	Complete establishment of the three-dimensional structure of an analyte using a combination of spectroscopic methods (1D and 2D NMR, HRMS, when necessary, chiroptical and/or crystallographic methods).
HRMS	High resolution mass spectrometry/spectrum. The high resolution and accurate mass facilitates molecular formula calculation from the ions observed.
MS^1	Full scan spectrum. Such spectra show all the quasi-molecular ions features (adducts, isotopologues, multimers). The high mass accuracy in HRMS enables molecular formulae to be calculation.
MS/MS (MS^2)	MS/MS fragmentation spectra are obtained on tandem mass spectrometers by first filtering a given precursor ion mass that is then fragmented in the gas phase. The most common fragmentation technique used in metabolomics is collision induced dissociation (CID). In UHPLC-HRMS/MS metabolite profiling, the acquisition of MS/MS spectra on most detected features is typically automated by data dependent acquisition.
Feature	A feature in MS corresponds to a mass-to-charge ratio, *m/z*, at a given retention time (Rt). It can be observed in one sample and optionally aligned across multiple samples. Each sample contains typically hundreds of features that can be aligned across multiple samples. Features can also be characterised by their LC-MS peak shape, intensities and MS/MS.
Spectral library	A spectral library consists in a collection of experimental or *in silico* predicted MS/MS spectra. A structure is linked to each spectrum. MS/MS spectra from metabolite profiling can be compared and matched against such libraries with spectral similarity methods.
Structural database	A structural database consists of a set of chemical structures encoded in computer readable format such as their SMILES code or InChI that can be queried programmatically. InChIKey can be obtained from the InChI and allow for efficient indexing of molecular structures in databases. In the context of NP research, a structural database ideally links the structures of NPs with those of the organisms from which they originate.

Mzmine,[28] MS-DIAL[29] and produces quantitative and qualitative information[30] consisting of:

(1) *a **feature table*** documenting the intensity (peak height or area) of feature *m* in sample *n*.

(2) *a **spectral list*** that summarises MS/MS spectra and/or isotopic patterns of each features.

The ***feature table*** (1), can be interpreted through MVDA to assess variability within a given sample set. This will reveal statistically significant differences across the metabolic profiles of the samples and highlight the feature(s) responsible for these differences. In chemical ecology, this type of analysis can be employed for the non-targeted identification of compounds induced in a biological organism as a result of an interaction or a perturbation. It can also be used to compare changes in the metabolome of a crop plant at different times during its growth or at a given stage of growth. Supervised machine learning approaches such as random-forest methods can be alternatively employed to extract features of interest related to a specific experimental design.[31]

The ***spectral list*** (2), can be used to annotate most metabolites observed for the analysed organism. Spectral annotation provides a qualitative view of the putative chemical structures detected for a particular organism or can also be used to assess chemodiversity in a larger set of natural extracts. Such qualitative assessment is a crucial step upstream of drug discovery screening campaigns to anticipate the presence of active compounds[32,33] or to assess patterns of metabolic composition in relation to ecological aspects (eco-metabolomics).[14]

For either of the above options, the goal will be to identify either the biomarker differentially expressed in a significant manner across given experimental conditions or assess the overall composition of the metabolome of a target organism or a biodiverse set of organisms. The qualitative description of a metabolome, however, is by far the most complex task in metabolomics. Below we describe some of the latest approaches in this area.

3. Challenges in Natural Product Identification from a Metabolomics Perspective

Usually, the structural elucidation of NPs requires their isolation through chromatographic methods followed by NMR methods[34] complemented with MS information. For absolute configuration determination and thus complete metabolite identification in the case of possible stereoisomerism, complementary chiroptical methods such as electronic circular dichroism (ECD) and/or X-ray crystallography are required.

Since metabolomic data are mainly based on MS (HRMS, MS/MS), the annotation of structures in extracts will thus have limited accuracy, even when searching against reference MS/MS spectra in spectral libraries. Metabolite annotation *via* MS alone is far from a trivial task, because, as mentioned above, more than 450,000 NPs have been characterised and only a fraction have their experimental MS/MS fragmentation data available in public spectral libraries. Overall, these spectral databases are estimated to contain over 220,000 spectra (however the fraction of NP within is hard to estimate).[35] Moreover, during LC-MS analyses in electrospray ionisation, fragmentation spectra mode are generally obtained by collision-induced dissociation that is subject to variation in the fragmentation intensities depending on the instrument.[36] Unlike gas-chromatography hyphenated to mass spectrometry (GC-MS) analyses where the fragmentation energy is fixed by convention, fragmentation energies in LC-MS based metabolomics vary and are not readily standardisable due to instrumental constraints.

Therefore, in order to accurately annotate the structure of a NP, multiple approaches can be followed: i) after establishing the molecular formula for MS^1 features, this information can be searched

in NPs structural databases to provide clues about possible structures; ii) by matching the associated fragmentation spectra against the spectra of candidates molecules in public spectral libraries such as GNPS or MassBank.[37] In addition to experimental spectral library matches, *in silico* predicted fragmentation spectra can also be used to expand the searchable spectral space.[38,39]

To face this annotation challenge, innovative approaches such as the organisation of spectral data through molecular networks have been developed over the last ten years. This approach makes it possible, for a single or multiple extracts, to organise all the MS/MS fragmentation information in the form of a network that will link each of the features into clusters according to their spectral similarity.[37] Since fragmentation spectra reflect the analytes' chemical structures, the features can be organised and visualised as families of potentially structurally related compounds. If annotations are obtained for the MS/MS spectra for some of the nodes within a cluster, it is possible to propagate the annotation to other nodes in the same cluster by exploiting the existing spectral similarity links (Fig. 2B).

The use of such an approach combined with *in silico* fragmentation spectral libraries generated with CFM-ID,[40] such as the In-Silico DataBase (ISDB), which currently contains more than 270,000 compounds and their associated spectra,[38,39] allows to annotate a large number of detected features. This approach can be reinforced by the use of taxonomic considerations (see below). In addition, other computational methods have been established to rank possible structures for a given fragmentation spectrum. In particular SIRIUS[41] first identifies putative molecular formulas from both the MS^1 isotopic analysis and MS/MS fragmentation tree analysis. The generated fragmentation tree is then employed to predict candidate molecules with CSI:FingerID, a machine learning-based method employing Support Vector Machine (SVM) models that was trained to recognise the presence of structural features from fragmentation patterns.[42] This concept was extended with CANOPUS to predict chemical classes thanks to a deep neural network.[43] In addition, SIRIUS was recently upgraded by COSMIC,[44] the first method that can quantify the confidence level in the computational annotation generated by SIRIUS/CSI:FingerID. In summary, the tools available today can organise spectra by similarity, obtain candidate structures for a large number of features, or even propose reliable information on the chemical class of compounds for unknown spectra. Such capabilities can be led on hundreds to thousands of extracts but require access to significant computational resources.

Progress and limitations in computational annotation methods are periodically benchmarked during the Critical Assessment of Small Molecule Identification (CASMI) contest (http://casmi-contest.org/). Since its introduction ten years ago, the results of the CASMI editions are proving that automated approaches are continuously improving. Moreover, the CASMI consecrated the emergence of efficient computational methods for each task, including adducts and isotopologues recognition, molecular formula identification, and structure dereplication/annotation. For the CASMI 2022, our laboratory teamed up with the Boecker laboratory (Chair for Bioinformatics, Jena University) and the Dorrestein laboratory (University of California San Diego). In this edition, unknown exposomics (*lato sensu*) molecules were analysed by LC-HRMS/MS, and the data shared with the participants. Besides providing ion masses and retention times, no other information was available, neither was the biological source.

Our team proposed a novel computational integrative strategy that boosted SIRIUS[41] performances by leveraging public spectral libraries and repositories with GNPS[37] mass spectral search tools to retrieve meta-information pointing to source organism(s). The latter were then used to consider the most relevant structural database and candidate for each challenge. More specifically, we incorporated taxonomically informed scoring and the LOTUS initiative database (see next paragraph) to the SIRIUS graphical user interface. Our integrative strategy top-performed by accurately proposing the correct answer at the first rank for 228 challenges (94% correct for molecular formula, 26% correct for structure, and 69% correct for chemical class annotation) which demonstrated that computational annotation methods have undoubtedly matured into powerful tools that can guide researchers into exploiting the metabolomics data generated.

However, it is crucial to keep in mind that the structural annotations proposed by computational methods remain putative and thus orthogonal information is needed to increase confidence. In this context, taxonomic information related to the analysed samples can provide a valuable insight. Indeed, a central assumption in chemotaxonomy is that genetically close taxa will tend to produce similar metabolites.[45] Contextualisation of annotations from a taxonomic point of view can therefore greatly help to improve confidence in the annotation. To this end, we recently launched the LOTUS initiative with the aim of exploring and establishing open and collaborative NPs knowledge-sharing systems. The structures found in LOTUS linked to their source organisms can facilitate taxonomy-informed dereplication.[46] The users can directly interact with the data through Wikidata (and different other read-only access points are available (https://lotus.naturalproducts.net/).[47] Users can retrieve all structures found in a given organism (*e.g.* all compounds from the *Melochia* (Q837434) genus https://w.wiki/5LJf), or the other way round, all organisms where a given structure was found (*e.g.* all biological organisms where ergotamine (Q419186) was found in https://w.wiki/5ZqW). More examples can be found at https://www.wikidata.org/wiki/Wikidata:WikiProject_Chemistry/Natural_products#Queries. For structural annotation, this taxonomic information can be taken into account. In this context, we have developed a strategy that automatically reweighs the structural annotation and favours, in lists of candidate structures, those that were already reported from organisms taxonomically related to the one studied.[46] Implementing these different computational tools and their combination in data processing workflows significantly improves the interpretation of annotation obtained.

It is important to note that while taxonomic information is important for more reliable annotation, chemotaxonomy is less accurate than taxonomy in linking traits, for example in speciation processes. It is thus well known that the composition of specialised metabolites of a given botanical species can be strongly influenced by its environment. Therefore, if trends between taxonomy and chemical profile clearly exist, the exploitation of this type of relationship should not be overvalued.

Despite all the developments in the field, an inherent limitation of MS is that fragmentation spectra interpretation does not allow conclusions to be drawn regarding the stereochemistry of the annotated compounds. This information can only be obtained by techniques such as NMR (for relative configurations) and chiroptical or crystallographic approaches (for absolute configurations). However, both NMR and chiroptical or crystallographic approaches generally require physical isolation of the analyte. While the isolation process was often slow and complex a few years ago, it is now possible to target the isolation of chromatographic peaks of interest directly from metabolomic profiling data and to selectively isolate a metabolite at the scale of a few tens of micrograms based on its chromatographic retention time and *m/z* value using semi-preparative scale chromatographic gradient transfer calculations.[48]

To this end, we have developed protocols to perform such targeted isolations by keeping chromatographic resolution close to the one obtained at the analytical scale.[49–52] Thus, NMR data giving valuable additional MS information can be obtained con-

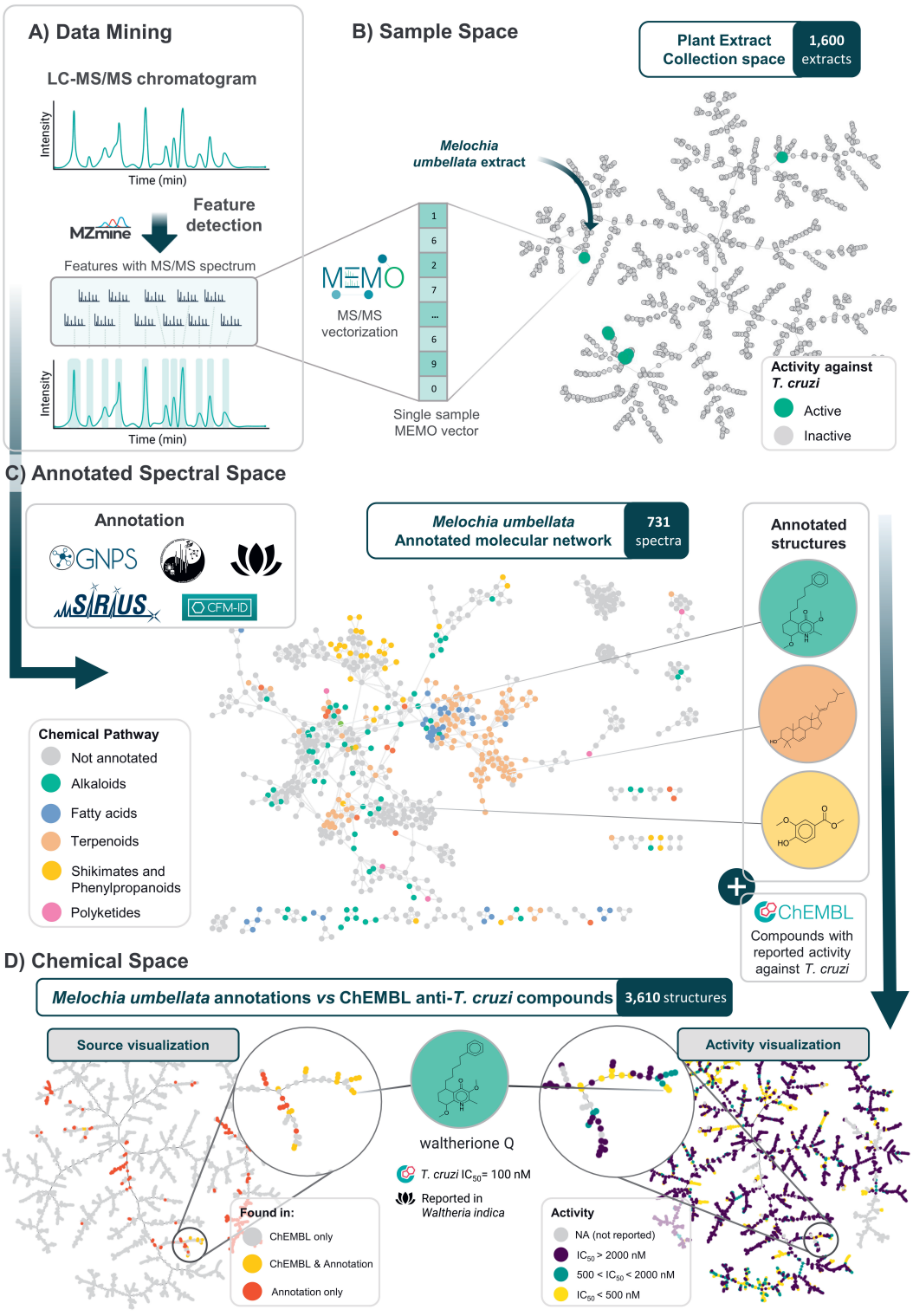


Fig. 2. A) After data processing, features along with their MS/MS spectra are detected in each sample. B) Based on the deconvoluted MS/MS spectra, samples can be compared using a vectorization process such as MEMO. To do so, MS/MS spectra of the detected features from each individual sample are vectorized with spec2vec[61] and aggregated to yield a MS/MS vector summarising the chemical diversity of the sample. In B), the same deconvoluted (here of the active *Melochia umbellata* (Q6813281) extract) spectra can be used to generate a molecular network (MN) to cluster features with similar fragmentation patterns. These spectra can also be annotated using different computational tools to provide putative structures for some spectra. Such a MN can be used to spot structural analogues and propagate the annotations from annotated to unannotated spectra. In C), the annotated compounds of the MN can be compared to different chemical DB (in this case ChEMBL) to further help the identification of biological or ecological properties. To help the exploration, visualisation methods such as TMAP can be used.[26] In this example, it is possible to spot the annotations of the waltherione Q (HDTZPKLRJBGXSO-OYKVQYDMSA-N, Q110090899), an alkaloid active against T. cruzi (Q150162) (IC_{50} = 100 nM) reported in *Waltheria indica* (Q7966688).[62]

veniently for selected metabolites of interest (Fig. 1B). Although NMR is inherently less sensitive than MS, it should be noted that today, high-field NMR platforms equipped with low-volume cryogenic probes allow the recording of information-rich NMR spectra with quantities down to the low μg range. This has considerably accelerated the complete identification of NPs and facilitated their rapid and efficient isolation to generate requisite sub-mg quantities. Propagation of such established structural information, mainly through molecular networks, then allows for a better annotation of a large number of structural analogues typically present in the metabolome of natural extracts.[53]

4. Selected Applications of Metabolomic Approaches from our own Research

Based on the previously described approaches, our group participated in various projects related to chemical and functional ecology as well as bioactive NPs discovery.

Regarding the elucidation of chemical-ecological mechanisms, we have conducted several untargeted metabolomics studies looking for compounds induced by herbivore stress. In one of these studies, several other jasmonate derivatives (Q415713) having a characteristic dynamic induction behaviour were highlighted in an *Arabidopsis thaliana* (Q158695) model. This study allowed the identification of new jasmonate derivatives (Q415713) whose

structures were fully identified after targeted isolation.[54] The data obtained, both on local and distal leaves from the wound site (leaf wounding by forceps mimicking herbivory), also revealed jasmonate accumulations at fast post-injury induction times (180 seconds).[55] Such studies have shown that targeted approaches could both identify new phytohormone analogues and reveal novel patterns of induction. This also underlines the importance of adapted time-series experiments for highlighting defence biomarkers. This approach allowed to generate new hypotheses on the dynamics of induction during wounding. It was then followed by targeted analyses to further characterise the behaviour of the revealed biomarkers.

We also applied metabolomic approaches searching for defensive compounds induced in fungal microorganisms. In collaboration with the Agroscope (agroscope.ch/changins/en), the application of comparative metabolomics methods on fungal solid media co-culture allowed the identification of *de novo* induced compounds in the confrontation zone.[56] For this purpose, co-cultures of fungi known to interact in nature and random co-cultures of strains available at Agroscope mycotheca (https://www.mycoscope.ch/) were studied with the aim of highlighting dynamic metabolite induction mechanisms. As an example, two wood-decaying fungi involved in esca disease wood of the vines were investigated with this strategy. *Botryosphaeria obtusa* (Q4948840) and *Eutypa lata* (Q10647956) confront each other and form black interaction zones visible in the wood. Metabolomic analysis by GC- and LC-MS of the co-culture of the corresponding strains on solid media revealed the induction of volatile and non-volatile substances. For this purpose, complex multiblock MVDA methods had to be used to consider the data structure (GC and LC-MS time series) and to highlight the relevant metabolomic changes. This approach was useful for the assessment of the specific impact of controlled experimental factors and evidenced the induction of *O*-methylmellein (Q77494423) in the solid media and 2-nonanone (Q15726063) in the volatile part suggesting a response implying both soluble and airborne compounds.[57] In other cases, random co-culturing reveal the biosynthetic potential of cryptic genes. To search specifically for induced metabolites in these cocultures, multivariate data processing approaches had to be adapted. This was necessary since the comparison of two monocultures of microorganisms with their cocultures should generate different groups in MVDA, and only compounds that differ from the mixture of the confronted metabolomes should be highlighted. A data mining approach, called POCHEMON, was specifically developed to highlight compounds produced following the interaction.[58] Here again, targeted isolations carried out on selected co-cultures allowed the identification of new dynamically induced bioactive NPs, notably with anti-microbial activities.

At the wider environmental scale, we also employed untargeted metabolomic approaches to explore variation in phytochemical diversity across multiple elevational transects in the Swiss alps.[15] The eco-phytochemical dataset obtained included samples from 450 alpine plant species, collected across 42 sites spread along 5 elevation gradients. We explored the extent to which environmental factors could influence phytochemical diversity heterogeneity of alpine ecosystems. This dataset contributed firstly to evaluate the effect of climate warming on plant-herbivore interaction and biodiversity in the alpine environment[15] and was exploited secondly to map phytochemical diversity at the landscape level based on molecular distribution models. In this last study, we calculated the climatic niche of >6000 phytochemical families, allowing us to investigate the spatial and evolutionary predictability of phytochemical diversity and introduce the innovative concept of 'chemodiversity hotspots'.[14]

Over the course of studies aimed at investigating the metabolome of plants or microorganisms, we have carried out high-throughput metabolite profiling (UHPLC-HRMS/MS) on sets of plants, lichens, as well as on model plants with specific ecological niche (tropical palm tree, seagrass) and their endophyte community. For all these cases, we applied molecular networks (MN) that involved the comparison of tens up to more than a thousand of extracts depending on the data sets. A study was, for example, aimed at the profiling of a unique collection of Euphorbiaceae (Q156584) in collaboration with the Institut de Chimie des Substances Naturelles (ICSN) at the Centre National de la Recherche Scientifique (CNRS) of Gif-sur-Yvette, France. Euphorbiaceae (Q156584) are known to produce diterpene esters with interesting anticancer and antiviral activities. These diterpene esters also present MS/MS fragmentation patterns rich in structural information, which makes them well suited for this type of MN analysis. Starting from 297 extracts, more than 1.8 million spectra were clustered as 88,687 nodes, themselves grouped in 7,840 spectral families. In parallel to this massive chemical screening, all extracts were also evaluated for their anticancer activity in Wnt-pathway inhibition tests. The results of the bioassays were used as metadata for the molecular networks, thus creating what was defined as a massive 'multi-informative' molecular network. MN also allows researchers to assess whether a compound is extract-specific or not in a set. The combination of both sources of information allowed to highlight some clusters of features putatively related to the anti-cancer activity found in the corresponding crude extracts. This approach allowed the targeted isolation and identification of new phorbol esters. Interestingly, one of them turned out to be a very strong inhibitor of the Wnt-pathway; it has a scaffold related to tigilanol tiglate (Q5322564), a PKC activator currently in phase 2 clinical trials.[59] The stacked layers of chemical and biological information allowed to efficiently prioritise putative bioactive compounds from this dataset. The isolation process was straightforward since it didn't require the classical iterative bio-guided fractionation procedure.[32] This study shows that it is possible to combine molecular network approaches and biological assays for targeted and efficient isolation of bioactive compounds of interest. In the same way, we have now initiated a project to investigate the metabolome of a set of highly biodiverse plants (Pierre Fabre collection[60]). For this purpose, a subset (1,600 extracts) representing around 10% of the whole collection (159 Families, 533 Genus, 767 species) was profiled and the data were processed in the form of a massive MN.

Here, the size and chemical diversity of the data set can lead to huge MN hindering the possibility to get a global view of the spectral relations. We explored alternative tools to summarise high-dimension data, for example, chemical structures of such a set in the form of minimum spanning trees (TMAP) resuming the overall information by pruning links of lower importance (Fig. 2B,D).[26] This type of processing is applicable not only to metabolite profiling data but also to all structural information reported in the literature on NPs. This can be used to compare a set of annotations with a set of reported active compounds to highlight annotated active compounds or analogues (Fig. 2D).

The wealth of information acquired through MS analysis at the level of a large set of extracts also raises problems in terms of data alignment. Indeed, if samples cannot be analysed concomitantly, retention-time and intensity shift across batches hinders the feature alignment step. To overcome this problem, we have recently developed a vectorisation approach (MEMO, for MS/MS-Based Sample Vectorization) to explore sets of chemically diverse natural extracts. The MS/MS fragmentation data (peaks and losses to the precursor) of all the features of a given sample are abstracted from their spectral context using spec2vec[61] and summarised in the form of a single vector that can be compared to other vectors, *i.e.* samples. This tool allows to compare samples without tak-

ing into account retention time information and thus avoiding the alignment process necessary for the usual MVDA. This process enables the comparison of sample sets analysed over long periods of time, or even on different MS platforms and with different chromatographic conditions. After MEMO vectorisation, each sample can be summarised as a single dot and the visualisation of all the samples in a set can be done using the previously presented TMAP visualisation (Fig. 2B,D). The application of MEMO has already demonstrated its effectiveness to establish relationships on sets of thousands of extracts in a dozen minutes using a classical laptop.[63]

5. Conclusion: Prospects for Natural Products Metabolomics

As described above, MS has played a central role in the development of metabolomics. Today, the analytical platforms allow researchers to quickly obtain detailed information on most MS features and their MS/MS spectra even in complex matrices such as natural extracts. To transform this spectral information into coherent chemical compositional information, many data mining tools have been developed and are continuously improved. Capacities for structural annotation of known NPs, as well as the structural anticipation of unknown compounds (absent in the databases), is constantly improving.[64] The NPs chemists now possess powerful tools to document, interrogate and study the metabolome of natural extracts. The exclusive use of MS-based metabolomics, however, suffers from inherent limitations of the technique, in particular, to differentiate isomers. Complementary approaches can improve the quality of the data collected and the annotations obtained. This could be done by better exploitation of the orthogonal information acquired during metabolite profiling. This is particularly the case for the chromatographic retention times of the detected compounds. Ideally, tools that would allow accurate prediction of these LC dimensions would be useful. Efforts have been made in this direction and interesting solutions have been proposed to predict the elution orders of analytes.[65,66] However, it is known that this kind of prediction remains difficult to compute as the factors that govern the retention of compounds in HPLC are multiple.

New generations of mass spectrometers now allow the acquisition of Ion Mobility Spectrometry (IMS) data for all detected analytes. IMS allows the measurement of metabolite drift time which depends on Collisional Cross Section (CCS), itself related to their stereochemical structure and conformations. However, IMS is not orthogonal to MS as CCS approximately correlates with the molecular mass. Yet, improvement in the performance of IMS resolution and the development of dedicated computational CCS prediction methods[67] will provide additional confidence in the annotation process with the increasing adoption of IMS. In addition to these improvements in the annotation process, more sensitive, more specific, and more reproducible MS/MS fragmentation methods are being developed in LC-MS which will help to boost both the metabolome coverage and its annotation.[68,69]

For unambiguous structural determination, the targeted isolation of metabolites of interest is still required for NMR, chiroptical, or crystallographic data acquisition. The sensitivity of NMR has increased and a few micrograms of NPs are sufficient to acquire high-quality data. Recently, with the advent of microcrystal electron diffraction (micro-ED) methods, it has become possible to perform X-ray analysis on micro-quantities of collected NPs. This opens up the possibility of obtaining structural information as well as absolute configurations ideally on any isolated NPs. Recently, a proof-of-concept publication has indicated the potential for coupling micro-ED methods with UHPLC.[70] The challenges in this area remain significant, however, as although the micro-ED technique is extremely sensitive, it requires micro-crystalline powder which is difficult to obtain in a generic manner.

Metabolomic data alone can provide important information for chemical ecology or for the discovery of bioactive NPs. However, they need to be integrated with other omics data to better understand the links.[71] The links between genomic and metabolomic data are thus increasingly being exploited to predict the biosynthetic potential of natural organisms. In bacteria, it is becoming increasingly feasible to identify biosynthetic gene clusters and the combination of genomic and metabolomic data in this context is very effective in selecting producers of compounds of interest or, more globally, for a better understanding of the biosynthetic capacities of this kingdom.[72] This link between biosynthetic gene clusters and specialised metabolites is however more complicated to establish in the case of organisms displaying complex genomes such as plants or fungi.

Metabolomics continues to be a rapidly growing area of research in NPs chemistry. There is no doubt that the advancement of analytical techniques in conjunction with computational methods on the one hand and the integration of multi-omics data on the other will give the scientists involved fantastic tools to interrogate the living world at the molecular level. This would lead to more efficient discovery of new drug candidates and a better understanding of the chemical interactions between organisms that are at the root of biodiversity.

Acknowledgements

J-LW, AR and P-MA are thankful to the Swiss National Science Foundation for the funding of the project (SNF N° CRSII5_189921/1).

Received: August 25, 2022

[1] F. Ntie-Kang, D. Svozil, *Phys. Sci. Rev*. **2020**, *5*, https://doi.org/10.1515/psr-2018-0121.
[2] R. N. Bennett, R. M. Wallsgrove, *New Phytol.* **1994**, *127*, 617, https://doi.org/10.1111/j.1469-8137.1994.tb02968.x.
[3] M. Erb, D. J. Kliebenstein, *Plant Physiol*. **2020**, *184*, 39, https://doi.org/10.1104/pp.20.00433.
[4] D. J. Newman, G. M. Cragg, *J. Nat. Prod*. **2020**, *83*, 770, https://doi.org/10.1021/acs.jnatprod.9b01285.
[5] W. C. Van Voorhis, R. H. van Huijsduijnen, T. N. C. Wells, *Proc. Natl. Acad. Sci. U. S. A*. **2015**, *112*, 15773, https://doi.org/10.1073/pnas.1520952112.
[6] K. Knudsmark Jessing, S. O. Duke, N. Cedergreeen, *J. Chem. Ecol*. **2014**, *40*, 100, https://doi.org/10.1007/s10886-014-0384-6.
[7] J. Hubert, J. M. Nuzillard, J. H. Renault, *Phytochem. Rev*. **2017**, *16*, 55, https://doi.org/10.1007/s11101-015-9448-7.
[8] J.-L. Wolfender, *Planta Med*. **2009**, *75*, 719, https://doi.org/10.1055/s-0028-1088393.
[9] R. A. Raguso, A. A. Agrawal, A. E. Douglas, G. Jander, A. Kessler, K. Poveda, J. S. Thaler, *Ecology* **2015**, *96*, 617, https://doi.org/10.1890/14-1474.1.
[10] L. A. Dyer, C. S. Philbin, K. M. Ochsenrider, L. A. Richards, T. J. Massad, A. M. Smilanich, M. L. Forister, T. L. Parchman, L. M. Galland, P. J. Hurtado, A. E. Espeset, A. E. Glassmire, J. G. Harrison, C. Mo, S. 'ad Yoon, N. A. Pardikes, N. D. Muchoney, J. P. Jahner, H. L. Slinn, O. Shelef, C. D. Dodson, M. J. Kato, L. F. Yamaguchi, C. S. Jeffrey, *Nat. Rev. Chem*. **2018**, *2*, 50, https://doi.org/10.1038/s41570-018-0009-7.
[11] O. Fiehn, *Plant Mol. Biol*. **2002**, *48*, 155, https://doi.org/10.1023/A:1013713905833.
[12] D. D. Marshall, R. Powers, *Prog. Nucl. Magn. Reson. Spectrosc*. **2017**, *100*, 1, https://doi.org/10.1016/j.pnmrs.2017.01.001.
[13] A. D. Kennedy, B. M. Wittmann, A. M. Evans, L. A. D. Miller, D. R. Toal, S. Lonergan, S. H. Elsea, K. L. Pappan, *J. Mass Spectrom*. **2018**, *53*, 1143, https://doi.org/10.1002/jms.4292.
[14] E. Defossez, C. Pitteloud, P. Descombes, G. Glauser, P.-M. Allard, T. W. N. Walker, P. Fernandez-Conradi, J.-L. Wolfender, L. Pellissier, S. Rasmann, *Proc. Natl. Acad. Sci. U. S. A*. **2021**, *118*, https://doi.org/10.1073/pnas.2013344118.
[15] P. Descombes, C. Pitteloud, G. Glauser, E. Defossez, A. Kergunteuil, P.-M. Allard, S. Rasmann, L. Pellissier, *Science* **2020**, *370*, 1469, https://doi.org/10.1126/science.abd7015.
[16] B. E. Sedio, *New Phytol*. **2017**, *214*, 952, https://doi.org/10.1111/nph.14438.
[17] Y. Bai, C. Yang, R. Halitschke, C. Paetz, D. Kessler, K. Burkard, E. Gaquerel, I. T. Baldwin, D. Li, *Science* **2022**, *375*, eabm2948, https://doi.org/10.1126/science.abm2948.

[18] T. W. N. Walker, J. M. Alexander, P.-M. Allard, O. Baines, V. Baldy, R. D. Bardgett, P. Capdevila, P. D. Coley, B. David, E. Defossez, M.-J. Endara, M. Ernst, C. Fernandez, D. Forrister, A. Gargallo-Garriga, V. E. J. Jassey, S. Marr, S. Neumann, L. Pellissier, J. Peñuelas, K. Peters, S. Rasmann, U. Roessner, J. Sardans, F. Schrodt, M. C. Schuman, A. Soule, H. Uthe, W. Weckwerth, J.-L. Wolfender, N. M. Dam, R. Salguero-Gómez, *J. Ecol*. **2022**, *110*, 4, https://doi.org/10.1111/1365-2745.13826.
[19] L. W. Sumner, Z. Lei, B. J. Nikolau, K. Saito, *Nat. Prod. Rep*. **2015**, *32*, 212, https://doi.org/10.1039/C4NP00072B.
[20] C. Kuhlisch, G. Pohnert, *Nat. Prod. Rep*. **2015**, *32*, 937, https://doi.org/10.1039/C5NP00003C.
[21] J. L. Wolfender, J. M. Nuzillard, J. J. J. van der Hooft, J. H. Renault, S. Bertrand, *Anal. Chem*. **2019**, *91*, 704, https://doi.org/10.1021/acs.analchem.8b05112.
[22] L. Perez de Souza, S. Alseekh, F. Scossa, A. R. Fernie, *Nat. Methods* **2021**, *18*, 733, https://doi.org/10.1038/s41592-021-01116-4.
[23] T. Kind, H. Tsugawa, T. Cajka, Y. Ma, Z. Lai, S. S. Mehta, G. Wohlgemuth, D. K. Barupal, M. R. Showalter, M. Arita, O. Fiehn, *Mass Spectrom. Rev*. **2018**, *37*, 513, https://doi.org/10.1002/mas.21535.
[24] C. Kuhl, R. Tautenhahn, C. Böttcher, T. R. Larson, S. Neumann, *Anal. Chem*. **2012**, *84*, 283, https://doi.org/10.1021/ac202450g.
[25] N. G. Mahieu, G. J. Patti, *Anal. Chem*. **2017**, *89*, 10397, https://doi.org/10.1021/acs.analchem.7b02380.
[26] D. Probst, J.-L. Reymond, *J. Cheminform*. **2020**, *12*, 12, https://doi.org/10.1186/s13321-020-0416-x.
[27] R. Tautenhahn, C. Böttcher, S. Neumann, *BMC Bioinformatics* **2008**, *9*, 504, https://doi.org/10.1186/1471-2105-9-504.
[28] T. Pluskal, S. Castillo, A. Villar-Briones, M. Oresic, *BMC Bioinformatics* **2010**, *11*, 395, https://doi.org/10.1186/1471-2105-11-395.
[29] H. Tsugawa, T. Cajka, T. Kind, Y. Ma, B. Higgins, K. Ikeda, M. Kanazawa, J. VanderGheynst, O. Fiehn, M. Arita, *Nat. Methods* **2015**, *12*, 523, https://doi.org/10.1038/nmeth.3393.
[30] L.-F. Nothias, D. Petras, R. Schmid, K. Dührkop, J. Rainer, A. Sarvepalli, I. Protsyuk, M. Ernst, H. Tsugawa, M. Fleischauer, F. Aicheler, A. A. Aksenov, O. Alka, P.-M. Allard, A. Barsch, X. Cachet, A. M. Caraballo-Rodriguez, R. R. Da Silva, T. Dang, N. Garg, J. M. Gauglitz, A. Gurevich, G. Isaac, A. K. Jarmusch, Z. Kameník, K. B. Kang, N. Kessler, I. Koester, A. Korf, A. Le Gouellec, M. Ludwig, C. Martin H, L.-I. McCall, J. McSayles, S. W. Meyer, H. Mohimani, M. Morsy, O. Moyne, S. Neumann, H. Neuweger, N. H. Nguyen, M. Nothias-Esposito, J. Paolini, V. V. Phelan, T. Pluskal, R. A. Quinn, S. Rogers, B. Shrestha, A. Tripathi, J. J. J. van der Hooft, F. Vargas, K. C. Weldon, M. Witting, H. Yang, Z. Zhang, F. Zubeil, O. Kohlbacher, S. Böcker, T. Alexandrov, N. Bandeira, M. Wang, P. C. Dorrestein, *Nat. Methods* **2020**, *17*, 905, https://doi.org/10.1038/s41592-020-0933-6.
[31] U. W. Liebal, A. N. T. Phan, M. Sudhakar, K. Raman, L. M. Blank, *Metabolites* **2020**, *10*, https://doi.org/10.3390/metabo10060243.
[32] F. Olivon, P.-M. Allard, A. Koval, D. Righi, G. Genta-Jouve, J. Neyts, C. Apel, C. Pannecouque, L.-F. Nothias, X. Cachet, L. Marcourt, F. Roussi, V. L. Katanaev, D. Touboul, J.-L. Wolfender, M. Litaudon, *ACS Chem. Biol*. **2017**, *12*, 2644, https://doi.org/10.1021/acschembio.7b00413.
[33] L.-F. Nothias, M. Nothias-Esposito, R. da Silva, M. Wang, I. Protsyuk, Z. Zhang, A. Sarvepalli, P. Leyssen, D. Touboul, J. Costa, J. Paolini, T. Alexandrov, M. Litaudon, P. C. Dorrestein, *J. Nat. Prod*. **2018**, *81*, 758, https://doi.org/10.1021/acs.jnatprod.7b00737.
[34] R. C. Breton, W. F. Reynolds, *Nat. Prowd. Rep*. **2013**, *30*, 501, https://doi.org/10.1039/c2np20104f.
[35] W. Bittremieux, D. H. May, J. Bilmes, W. S. Noble, *Nat. Methods* **2022**, *19*, 675, https://doi.org/10.1038/s41592-022-01496-1.
[36] D. P. Demarque, A. E. M. Crotti, R. Vessecchi, J. L. C. Lopes, N. P. Lopes, *Nat. Prod. Rep.* **2016**, *33*, 432, https://doi.org/10.1039/C5NP00073D.
[37] M. Wang, J. J. Carver, V. V. Phelan, L. M. Sanchez, N. Garg, Y. Peng, D. D. Nguyen, J. Watrous, C. A. Kapono, T. Luzzatto-Knaan, C. Porto, A. Bouslimani, A. V. Melnik, M. J. Meehan, W.-T. Liu, M. Crüsemann, P. D. Boudreau, E. Esquenazi, M. Sandoval-Calderón, R. D. Kersten, L. A. Pace, R. A. Quinn, K. R. Duncan, C.-C. Hsu, D. J. Floros, R. G. Gavilan, K. Kleigrewe, T. Northen, R. J. Dutton, D. Parrot, E. E. Carlson, B. Aigle, C. F. Michelsen, L. Jelsbak, C. Sohlenkamp, P. Pevzner, A. Edlund, J. McLean, J. Piel, B. T. Murphy, L. Gerwick, C.-C. Liaw, Y.-L. Yang, H.-U. Humpf, M. Maansson, R. A. Keyzers, A. C. Sims, A. R. Johnson, A. M. Sidebottom, B. E. Sedio, A. Klitgaard, C. B. Larson, C. A. B. P, D. Torres-Mendoza, D. J. Gonzalez, D. B. Silva, L. M. Marques, D. P. Demarque, E. Pociute, E. C. O'Neill, E. Briand, E. J. N. Helfrich, E. A. Granatosky, E. Glukhov, F. Ryffel, H. Houson, H. Mohimani, J. J. Kharbush, Y. Zeng, J. A. Vorholt, K. L. Kurita, P. Charusanti, K. L. McPhail, K. F. Nielsen, L. Vuong, M. Elfeki, M. F. Traxler, N. Engene, N. Koyama, O. B. Vining, R. Baric, R. R. Silva, S. J. Mascuch, S. Tomasi, S. Jenkins, V. Macherla, T. Hoffman, V. Agarwal, P. G. Williams, J. Dai, R. Neupane, J. Gurr, A. M. C. Rodríguez, A. Lamsa, C. Zhang, K. Dorrestein, B. M. Duggan, J. Almaliti, P.-M. Allard, P. Phapale, L.-F. Nothias, T. Alexandrov, M. Litaudon, J.-L. Wolfender, J. E. Kyle, T. O. Metz, T. Peryea, D.-T. Nguyen, D. VanLeer, P. Shinn, A. Jadhav, R. Müller, K. M. Waters, W. Shi, X. Liu, L. Zhang, R. Knight, P. R. Jensen, B. O. Palsson, K. Pogliano, R. G. Linington, M. Gutiérrez, N. P. Lopes, W. H. Gerwick, B. S. Moore, P. C. Dorrestein, N. Bandeira, *Nat. Biotechnol*. **2016**, *34*, 828, https://doi.org/10.1038/nbt.3597.
[38] P. M. Allard, T. Peresse, J. Bisson, K. Gindro, L. Marcourt, V. C. Pham, F. Roussi, M. Litaudon, J. L. Wolfender, *Anal. Chem*. **2016**, *88*, 3317, https://doi.org/10.1021/acs.analchem.5b04804.
[39] P.-M. Allard, J. Bisson, A. Rutz, **2022**, https://doi.org/10.5281/zenodo.6939173.
[40] F. Allen, R. Greiner, D. Wishart, *Metabolomics* **2015**, *11*, 98, https://doi.org/10.1007/s11306-014-0676-4.
[41] K. Duhrkop, M. Fleischauer, M. Ludwig, A. A. Aksenov, A. V. Melnik, M. Meusel, P. C. Dorrestein, J. Rousu, S. Bocker, *Nat. Methods* **2019**, *16*, 299, https://doi.org/10.1038/s41592-019-0344-8.
[42] K. Dührkop, H. Shen, M. Meusel, J. Rousu, S. Böcker, *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 12580, https://doi.org/10.1073/pnas.1509788112.
[43] K. Dührkop, L.-F. Nothias, M. Fleischauer, R. Reher, M. Ludwig, M. A. Hoffmann, D. Petras, W. H. Gerwick, J. Rousu, P. C. Dorrestein, S. Böcker, *Nat. Biotechnol*. **2020**, *39*, 462, https://doi.org/10.1038/s41587-020-0740-8.
[44] M. A. Hoffmann, L.-F. Nothias, M. Ludwig, M. Fleischauer, E. C. Gentry, M. Witting, P. C. Dorrestein, K. Dührkop, S. Böcker, *Nat. Biotechnol*. **2022**, *40*, 411, https://doi.org/10.1038/s41587-021-01045-9.
[45] O. R. Gottlieb, *Phytochemistry* **1990**, *29*, 1715, https://doi.org/10.1016/0031-9422(90)85002-W.
[46] A. Rutz, M. Dounoue-Kubo, S. Ollivier, J. Bisson, M. Bagheri, T. Saesong, S. N. Ebrahimi, K. Ingkaninan, J.-L. Wolfender, P.-M. Allard, *Front. Plant Sci*. **2019**, *10*, 1329, https://doi.org/10.3389/fpls.2019.01329.
[47] A. Rutz, M. Sorokina, J. Galgonek, D. Mietchen, E. Willighagen, A. Gaudry, J. G. Graham, R. Stephan, R. Page, J. Vondrášek, C. Steinbeck, G. F. Pauli, J.-L. Wolfender, J. Bisson, P.-M. Allard, *eLife* **2022**, *11*, https://doi.org/10.7554/eLife.70780.
[48] T. F. Molinski, *Curr. Opin. Drug Discov. Devel*. **2009**, 12, 197.
[49] L. Pellissier, A. Koval, L. Marcourt, E. Ferreira Queiroz, N. Lecoultre, S. Leoni, L.-M. Quiros-Guerrero, M. Barthélémy, B. L. Duivelshof, D. Guillarme, S. Tardy, V. Eparvier, K. Perron, J. Chave, D. Stien, K. Gindro, V. Katanaev, J.-L. Wolfender, *Front. Chem*. **2021**, *9*, 664489, https://doi.org/10.3389/fchem.2021.664489.
[50] A. Alfattani, L. Marcourt, V. Hofstetter, E. F. Queiroz, S. Leoni, P.-M. Allard, K. Gindro, D. Stien, K. Perron, J.-L. Wolfender, *Front. Mol. Biosci*. **2021**, *8*, 725691, https://doi.org/10.3389/fmolb.2021.725691.
[51] E. F. Queiroz, A. Alfattani, A. Afzan, L. Marcourt, D. Guillarme, J.-L. Wolfender, *J. Chromatogr. A* **2019**, *1598*, 85, https://doi.org/10.1016/j.chroma.2019.03.042.
[52] A. Azzollini, Q. Favre-Godal, J. Zhang, L. Marcourt, S. N. Ebrahimi, S. Wang, P. Fan, H. Lou, D. Guillarme, E. F. Queiroz, J.-L. Wolfender, *Planta Med*. **2016**, *82*, 1051, https://doi.org/10.1055/s-0042-108207.
[53] J. Y. Yang, L. M. Sanchez, C. M. Rath, X. Liu, P. D. Boudreau, N. Bruns, E. Glukhov, A. Wodtke, R. de Felicio, A. Fenner, W. R. Wong, R. G. Linington, L. Zhang, H. M. Debonsi, W. H. Gerwick, P. C. Dorrestein, *J. Nat. Prod*. **2013**, *76*, 1686, https://doi.org/10.1021/np400413s.
[54] G. Glauser, J. Boccard, S. Rudaz, J.-L. Wolfender, *Phytochem. Anal*. **2010**, *21*, 95, https://doi.org/10.1002/pca.1155.
[55] G. Glauser, E. Grata, L. Dubugnon, S. Rudaz, E. Farmer, J. L. Wolfender, *J. Biol. Chem*. **2008**, *283*, 16400, https://doi.org/10.1074/jbc.M801760200.
[56] S. Bertrand, N. Bohni, S. Schnee, O. Schumpp, K. Gindro, J. L. Wolfender, *Biotechnol. Adv*. **2014**, *32*, 1180, https://doi.org/10.1016/j.biotechadv.2014.03.001.
[57] A. Azzollini, L. Boggia, J. Boccard, B. Sgorbini, N. Lecoultre, P.-M. Allard, P. Rubiolo, S. Rudaz, K. Gindro, C. Bicchi, J.-L. Wolfender, *Front. Microbiol*. **2018**, *9*, https://doi.org/10.3389/fmicb.2018.00072.
[58] J. J. Jansen, L. Blanchet, L. M. C. Buydens, S. Bertrand, J.-L. Wolfender, *Metabolomics* **2015**, *11*, 908, https://doi.org/10.1007/s11306-014-0748-5.
[59] C. M. E. Barnett, N. Broit, P.-Y. Yap, J. K. Cullen, P. G. Parsons, B. J. Panizza, G. M. Boyle, *Invest. New Drugs* **2019**, *37*, 1, https://doi.org/10.1007/s10637-018-0604-y.
[60] Official European Commission Register of Collections, European Commission, **2020**.
[61] F. Huber, L. Ridder, S. Verhoeven, J. H. Spaaks, F. Diblen, S. Rogers, J. J. J. van der Hooft, *PLoS Comput. Biol*. **2021**, *17*, e1008724, https://doi.org/10.1371/journal.pcbi.1008724.
[62] S. Cretton, S. Dorsaz, A. Azzollini, Q. Favre-Godal, L. Marcourt, S. N. Ebrahimi, F. Voinesco, E. Michellod, D. Sanglard, K. Gindro, J.-L. Wolfender, M. Cuendet, P. Christen, *J. Nat. Prod*. **2016**, *79*, 300, https://doi.org/10.1021/acs.jnatprod.5b00896.
[63] A. Gaudry, F. Huber, L.-F. Nothias, S. Cretton, M. Kaiser, J.-L. Wolfender, P.-M. Allard, *Front. Bioinform*. **2022**, *2*, https://doi.org/10.3389/fbinf.2022.842964.
[64] A. E. Fox Ramos, C. Pavesi, M. Litaudon, V. Dumontet, E. Poupon, P. Champy, G. Genta-Jouve, M. A. Beniddir, *Anal. Chem*. **2019**, *91*, 11247, https://doi.org/10.1021/acs.analchem.9b02216.

[65] E. Bach, S. Szedmak, C. Brouard, S. Böcker, J. Rousu, *Bioinformatics* **2018**, *34*, i875, https://doi.org/10.1093/bioinformatics/bty590.
[66] P. Bonini, T. Kind, H. Tsugawa, D. K. Barupal, O. Fiehn, *Anal. Chem*. **2020**, *92*, 7515, https://doi.org/10.1021/acs.analchem.9b05765.
[67] S. M. Colby, D. G. Thomas, J. R. Nuñez, D. J. Baxter, K. R. Glaesemann, J. M. Brown, M. A. Pirrung, N. Govind, J. G. Teeguarden, T. O. Metz, R. S. Renslow, *Anal. Chem*. **2019**, *91*, 4346, https://doi.org/10.1021/acs.analchem.8b04567.
[68] Z. Zuo, L. Cao, L.-F. Nothia, H. Mohimani, *Bioinformatics* **2021**, *37*, i231, https://doi.org/10.1093/bioinformatics/btab279.
[69] V. Davies, J. Wandy, S. Weidt, J. J. J. van der Hooft, A. Miller, R. Daly, S. Rogers, *Anal. Chem*. **2021**, *93*, 5676, https://doi.org/10.1021/acs.analchem.0c03895.
[70] R. Ghosh, G. Bu, B. L. Nannenga, L. W. Sumner, *Front. Mol. Biosci*. **2021**, *8*, 720955, https://doi.org/10.3389/fmolb.2021.720955.
[71] M. A. Schorn, S. Verhoeven, L. Ridder, F. Huber, D. D. Acharya, A. A. Aksenov, G. Aleti, J. A. Moghaddam, A. T. Aron, S. Aziz, A. Bauermeister, K. D. Bauman, M. Baunach, C. Beemelmanns, J. M. Beman, M. V. Berlanga-Clavero, A. A. Blacutt, H. B. Bode, A. Boullie, A. Brejnrod, T. S. Bugni, A. Calteau, L. Cao, V. J. Carrión, R. Castelo-Branco, S. Chanana, A. B. Chase, M. G. Chevrette, L. V. Costa-Lotufo, J. M. Crawford, C. R. Currie, B. Cuypers, T. Dang, T. de Rond, A. M. Demko, E. Dittmann, C. Du, C. Drozd, J.-C. Dujardin, R. J. Dutton, A. Edlund, D. P. Fewer, N. Garg, J. M. Gauglitz, E. C. Gentry, L. Gerwick, E. Glukhov, H. Gross, M. Gugger, D. G. Guillén Matus, E. J. N. Helfrich, B.-F. Hempel, J.-S. Hur, M. Iorio, P. R. Jensen, K. B. Kang, L. Kaysser, N. L. Kelleher, C. S. Kim, K. H. Kim, I. Koester, G. M. König, T. Leao, S. R. Lee, Y.-Y. Lee, X. Li, J. C. Little, K. N. Maloney, D. Männle, C. Martin H, A. C. McAvoy, W. W. Metcalf, H. Mohimani, C. Molina-Santiago, B. S. Moore, M. W. Mullowney, M. Muskat, L.-F. Nothias, E. C. O'Neill, E. I. Parkinson, D. Petras, J. Piel, E. C. Pierce, K. Pires, R. Reher, D. Romero, M. C. Roper, M. Rust, H. Saad, C. Saenz, L. M. Sanchez, S. J. Sørensen, M. Sosio, R. D. Süssmuth, D. Sweeney, K. Tahlan, R. J. Thomson, N. J. Tobias, A. E. Trindade-Silva, G. P. van Wezel, M. Wang, K. C. Weldon, F. Zhang, N. Ziemert, K. R. Duncan, M. Crüsemann, S. Rogers, P. C. Dorrestein, M. H. Medema, J. J. J. van der Hooft, *Nat. Chem. Biol*. **2021**, *17*, 363, https://doi.org/10.1038/s41589-020-00724-z.
[72] S. A. Kautsar, K. Blin, S. Shaw, J. C. Navarro-Muñoz, B. R. Terlouw, J. J. J. van der Hooft, J. A. van Santen, V. Tracanna, H. G. Suarez Duran, V. Pascal Andreu, N. Selem-Mojica, M. Alanjary, S. L. Robinson, G. Lund, S. C. Epstein, A. C. Sisto, L. K. Charkoudian, J. Collemare, R. G. Linington, T. Weber, M. H. Medema, *Nucleic Acids Res*. **2020**, *48*, D454, https://doi.org/10.1093/nar/gkz882.

License and Terms

The definitive version of this article is the electronic one that can be found at https://doi.org/10.2533/chimia.2022.954