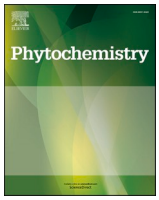
Phytochemistry 220 (2024) 114004

Contents lists available at ScienceDirect

Phytochemistry

journal homepage: www.elsevier.com/locate/phytochem



Essentials in the acquisition, interpretation, and reporting of plant metabolite profiles

Serhat S. Çiçek [a], Alfonso Mangoni [b], Franziska S. Hanschen [c], Niels Agerbirk [d], Christian Zidorn [e,*]

[a] Department of Biotechnology, Hamburg University of Applied Sciences, Ulmenliet 20, 21033, Hamburg, Germany
[b] Dipartimento di Farmacia, Università di Napoli Federico II, Via Domenico Montesano 49, 80131, Napoli, Italy
[c] Plant Quality and Food Security, Leibniz Institute of Vegetable and Ornamental Crops (IGZ) e. V., Theodor-Echtermeyer-Weg 1, 14979, Grossbeeren, Germany
[d] Department of Plant and Environmental Sciences, University of Copenhagen, Thorvaldsensvej 40, 1871, Frederiksberg, Denmark
[e] Pharmazeutisches Institut, Abteilung Pharmazeutische Biologie, Christian-Albrechts- Universität zu Kiel, Gutenbergstraße 76, 24118, Kiel, Germany

ARTICLE INFO

Keywords:
Plant metabolite profiling
Mass spectrometry
NMR
Misidentification
Transparency
Standards
References
Interpretation
Confidence levels
Chemophenetics
Chemosystematics
Chemotaxonomy

ABSTRACT

Plant metabolite profiling reveals the diversity of secondary or specialized metabolites in the plant kingdom with its hundreds of thousands of species. Specialized plant metabolites constitute a vast class of chemicals posing significant challenges in analytical chemistry. In order to be of maximum scientific relevance, reports dealing with these compounds and their source species must be transparent, make use of standards and reference materials, and be based on correctly and traceably identified plant material. Essential aspects in qualitative plant metabolite profiling include: (i) critical review of previous literature and a reasoned sampling strategy; (ii) transparent plant sampling with wild material documented by vouchers in public herbaria and, optimally, seed banks; (iii) if possible, inclusion of generally available reference plant material; (iv) transparent, documented state-of-the art chemical analysis, ideally including chemical reference standards; (v) testing for artefacts during preparative extraction and isolation, using gentle analytical methods; (vi) careful chemical data interpretation, avoiding over- and misinterpretation and taking into account phytochemical complexity when assigning identification confidence levels, and (vii) taking all previous scientific knowledge into account in reporting the scientific data. From the current stage of the phytochemical literature, selected comments and suggestions are given. In the past, proposed revisions of botanical taxonomy were sometimes based on metabolite profiles, but this approach ("chemosystematics" or "chemotaxonomy") is outdated due to the advent of DNA sequence-based phylogenies. In contrast, systematic comparisons of plant metabolite profiles in a known phylogenetic framework remain relevant. This approach, known as chemophenetics, allows characterizing species and clades based on their array of specialized metabolites, aids in deducing the evolution of biosynthetic pathways and coevolution, and can serve in identifying new sources of rare and economically interesting natural products.

1. Introduction

Plants, as other organisms, contain a relatively limited number of metabolites essential for the survival of cells in general, known as the primary or general metabolites. However, a significant aspect of phytochemistry concerns the vast number of additional metabolites, known as secondary or specialized metabolites (Wittstock and Gershenzon, 2002; Wink, 2003; Pichersky and Raguso, 2018; Alseekh et al., 2020; Negin and Jander, 2023). These metabolites are involved in a myriad of ecologically essential processes, such as signaling, resistance, and competition, all of which are relevant in plant breeding (Bednarek and Osbourn, 2009; Zhou and Jander 2021; Plaszkó et al., 2022). Furthermore, the secondary/specialized metabolites are of considerable interest for a healthy diet (Tang et al., 2017; Aune et al., 2017; Wink, 2022; Hill et al., 2023). The exact border between general and secondary/specialized metabolites is in many cases not clear-cut (Neilson et al., 2018; Erb and Kliebenstein, 2020). Therefore, in the following text the term 'metabolites' in a general sense will usually be used. As the number of metabolites in plants is much higher than the number of metabolites in mammals, reliable plant metabolite profiling and

* Corresponding author.
E-mail address: czidorn@pharmazie.uni-kiel.de (C. Zidorn).

https://doi.org/10.1016/j.phytochem.2024.114004
Received 11 November 2023; Received in revised form 22 January 2024; Accepted 23 January 2024
Available online 6 February 2024

metabolomics cannot simply be carried out using procedures and standards developed for medical metabolomics. Though the reliable documentation of plant metabolite profiles has always been challenging, growing access to sophisticated analytical equipment currently results in a higher number of papers published in the field. Unfortunately, also including a higher number of papers, in which the results are not correctly interpreted.

In particular, due to the progress in hyphenated analytical methods, the literature on the occurrence of plant metabolite profiles is currently growing at an exponential pace. However, due to lack of standards and reference materials and due to the incredible complexity of the analytical tasks, reliable plant metabolite profiling is still not trivial. Indeed, a considerable share of the current literature fails in one or more essential aspects, leading to a fragmented literature with many apparent false positives or overly interpreted reports.

Based on our joint experience in reviewing hundreds of manuscripts and published papers, selected essentials in reporting plant metabolite profiles will be highlighted and discussed. In support of our points, selected positive examples concerning the aspects at hand will be cited. Finally, some guidelines aiming at improving the scientific value of plant metabolite profiling reports will be proposed.

2. Plant sampling and identification

2.1. Choice of species and accessions

Before embarking on a dedicated study, at least a preliminary literature review should precede the choice of plant material, in order to secure relevance of the results and as far as possible eliminate the risk of omission of essential controls or reference groups. The species that first come to mind may be extremely well described already, while closely related species may represent essentially uncharted land. Adherence to the Nagoya protocol also needs consideration, even when the purpose is limited to metabolite profiling (David, 2018).

2.2. Crops and model species

Often, metabolite profiling concerns crops or model species. For a transparent report, the origin of the material should be indicated as precisely as possible for future reference (Kariya et al., 2023). Many crop species contain innumerable cultivars, often differing in metabolite profile. The exact commercial supplier of seeds of such varieties should always be indicated, in order to make access to the same seed material easier for researchers trying to repeat or build on the described experiments and to be able to judge the reliability of the indicated cultivar identification. For a report of a previously unknown metabolic feature, it is relevant information whether an *Arabidopsis thaliana* (L.) Heynh. accession or a specific crop cultivar was propagated for years in your own greenhouses, or obtained directly from a reliable supplier. Growth conditions should also be reported with as much detail as deemed necessary for a scientist in the field to reproduce the experiment.

2.3. Plant samples collected from the wild

For any publication in scientific journals, the full scientific name including the author citation of each plant species investigated is required. However, few authors consider the fact that the scientific name alone is not sufficient to characterize and delimit the investigated plant taxon (a species, a subspecies or a variety, in general terms a "taxon"). Although many biochemists may consider a species name definite, species and even subspecies and varieties are surprisingly heterogeneous entities, far from the apparently fixed combinations of characters listed in floras. Thus, the wild legume *Anthyllis vulneraria* L. is an easily recognizable taxon, which however can be split in at least 40 species, subspecies or varieties, depending on the personal judgement of the scientist working with this group (Cullen, 1968 versus Pignatti, 2017). In order to obtain specificity and transparency in the plant identification, a transparent documentation of the performed plant identification is needed. Transparency in this case includes specifying the actual flora used, the key characters observed and, if possible, one or more DNA barcode sequences. The term 'taxonym' was recently proposed for such a transparent link between scientific name and intended species definition/delimitation (Fischer, 2015). Another important aspect is the traceability of the plant identification; only with proper vouchers can future researchers ascertain that investigated plants were correctly identified. Numerous related critical aspects of plant identification and documentation, such as transparent sampling, precise indication of sampling coordinates and indication of various degrees of certainty in plant identification, have been covered before (Zidorn, 2017). Briefly, the following points were suggested: a) mandatory indication of a taxonym, i.e., the respective scientific name plus a reference clarifying its definition; b) the preparation of high quality voucher specimens, which also should be made available both in digital form and deposited in an Index Herbariorum listed herbarium; and c) a detailed and standardised geo-referencing of the collection site (including WGS84 coordinates and the altitude of the collection site above sea level).

2.4. Inclusion of reference species or cultivars

Often, scientific investigation is focused on field-collected material, such as sampling in supermarkets in order to assess the actual produce available to consumers, analysis of confidential breeding lines unavailable to other scientists, or sampling wild species in search of useful new germplasm. In those cases, inclusion of suitably related crops, model species or seed-bank material is highly advisable (Paguet et al., 2023; Grauso et al., 2023). Likewise, from any investigation of field-collected material, deposition of seeds from representative species in a publicly available seed bank should be considered, with due consideration of the Nagoya protocol (David, 2018). Inclusion of botanical references and deposition of seeds dramatically increases the value of the scientific report. Deposition in seed banks is particularly relevant in studies describing new enzymes or metabolites from plants (Blažević et al., 2020), because availability of reference materials is key to reliable metabolite identification (Section 3.3).

2.5. Precise description of sampling and replication

True biological replicates (as opposed to technical replicates and pseudo-replicates) are essential in plant science (Rogers et al., 2021). Additionally, a precise description of the sampling strategy is needed. Even though the rigorous sampling strategies of e.g., environmental analysis are rarely if ever realized in phytochemical studies, the theory of sampling (Petersen et al., 2005) is a useful basis for understanding just how much can go wrong during sampling of heterogeneous entities like entire plants or even natural vegetation. If a systematic strategy was used, e.g., for intended randomization or for minimizing variation (e.g., sampling only a specific leaf position), it should be indicated (Alseekh et al., 2021). A practice specifically prone to error is non-random subsampling, e.g., taking leaf discs in the middle of leaves or avoiding nerves or petioles, in that way not sampling the leaf margin, nerve or petiole with different metabolite profile (Shroff et al., 2015; Lorensen et al., 2023) or taking subsamples of roots without considering heterogeneity. Even if some degree of biased sampling practice is considered acceptable or inevitable (e.g., taking only lower branches of trees or taking roots accumulating in the bottom of pots), it should be reported transparently (Formato et al., 2022). In particular during field sampling campaigns of plants collected in their wild habitats, it should be considered that qualitative and quantitative variation of natural products can also depend on factors such as shading or sun exposition of the individual plant, age of the sampled plant, age of the sampled plant organ, seasonal variation of plant natural products, and other related

factors (Zidorn and Stuppner, 2001b; Zidorn, 2018; Li and Zidorn, 2022).

2.6. Metabolites from endophytes and microbiome

Microbes living in or closely attached to a plant part (endophytes, including fungi and prokaryotes) are ubiquitous in plants (Vandenkoornhuyse et al., 2015; Wang and Cernava, 2023), yet not visually apparent to the naked eye, and may produce or elicit a wide variety of metabolites (Tan and Zou, 2001; Ludwig-Müller, 2015; Liu et al., 2023). In traditional phytochemical analysis, usually based on extracts of apparently clean and healthy macroscopic plant parts including the invisible endophytes, distinction of true products of plant metabolism from endophyte metabolism is impossible, while specialized approaches allow distinction (Pang et al., 2021). From studies of endophytes in culture, numerous examples of proven endophyte-derived metabolites are known (Gao et al., 2018; Amirzakariya and Shakeri, 2022). However, endophytes in culture may not reveal their full metabolic potential unless exposed to specific elicitors (Toghueo et al., 2020; Su et al., 2023). Metabolites from endophytes can also be derivatives of plant metabolites, such as phytoalexins (Pedras and Thapa, 2020). There are almost certainly many yet undiscovered cases of a microbial origin of apparent plant metabolites. Ultimately, when the biochemistry and gene sequences behind a group of metabolites is known, a microbial origin will become apparent from e.g., the nature of the mRNA and gene sequences. In general plant metabolite profiling, the possibility of a microbial origin of the detected compounds should always be considered. Reporting details of the origin and nature of the investigated plant material will improve chances of later critical investigation of any irreproducible findings that may nevertheless be genuine findings due to variable effects of endophytes.

3. Analytical chemistry

3.1. Current plant metabolite libraries and their use in identification

Around 100 different databases containing natural products are currently accessible, of which half are open access (Sorokina and Steinbeck, 2020). Among them are the commercial databases SciFinder (Gabrielson, 2018), Reaxys (https://www.reaxys.com/#/search/quick), the National Institute of Standards and Technology (NIST) database (NIST, 2023) or the Dictionary of Natural Products (Dictionary of Natural Products 31.2., 2023), as well as the open-access collection of natural products (COCONUT) (Sorokina et al., 2021), covering between 200,000 and 400,000 natural products (Sorokina and Steinbeck, 2020). Out of these, the NIST database provides LC-MS/MS data for compound identification for 51,500 compounds (NIST, 2023). This is, however, far below the 843,000 compounds (NIST, 2023) available for GC-MS identification and thus highlights the (current) limitation of the LC-MS/MS approach. Likewise, natural product databases provided by some instrument manufacturers are not sufficient to achieve reliable confidence due to the limited number of included metabolites. However, there have been approaches to increase this number by linking the measured spectra to metadata-containing chemical libraries, e.g., ChemSpider (Hyland et al., 2019).

Apart from inadequacies of available LC-MS data, LC-MS/MS compound libraries also lack retention indices (RI), such as the Kovats-RI (Kováts, 1958), the Lee-RI (Lee et al., 1979), or the Fiehn-RI (Kind et al., 2009). These indices, which are based on *n*-alkanes, polycyclic aromatic hydrocarbons, or fatty acid methyl esters, play a key role in GC removing redundancies, e.g., from different temperature gradients. Different calibrants have been used to eventually establish a retention indexing system of LC or LC-MS, respectively, applying *n*-nitroalkanes (Hall et al., 2012), 2-dimethylaminoethylamine-labelled fatty acids (Zheng et al., 2018), or *N*-alkylpyridinium sulfonates (Quilliam, 2017). So far, none of these systems has been established due to different shortcomings of the calibrants, such as improper ionization with one of the two main ionization sources (ESI, APCI), their susceptibility to pH changes or their varying elution patterns in polar media, thus requiring logarithmic functions for calculation. Whether a recently proposed RI system applying cocamide diethanolamine homologues (Aalizadeh et al., 2022) is devoid of eventual shortcomings will be shown in the future.

3.2. Examples of problems related to attempted identification in GC-MS and LC-MS/MS

By using mass spectral databases together with measured retention indices, GC-MS experiments allow relatively reliable identification of known compounds. However, while the use of the Kovats-RI was long time essential – being the only identification tool in many earlier studies – the extended use of mass spectral libraries led to its neglect and compound identification is now often based on MS similarity indices only. Hence, instead of a relatively reliable identification of compounds a list of false positive hits may be generated. For instance, a recent investigation of essential oils from the legume tree *Vouacapoua americana* Aubl. by GC-MS revealed 14 sesquiterpenes after comparison of their RI values, MS similarity indices (SI) and reversed SI values [when the peaks in the unknown's spectrum that are not in the library's known reference spectrum are ignored (Çiçek et al., 2023; JordiLabs, 2023; NIST, 2023]. By using only SI and RSI values with acceptable scores, the list of identified compounds would give a total of 29 sesquiterpenes. Thus, twice as many compounds would have been reported, though half of them did not match in retention behavior and therefore would have been assigned erroneously.

Even more problematic is the online detection by means of LC-MS/MS, with even the biggest databases only constituting a mere fraction of their GC-MS counterparts. This problem is increased by the fact that the number of non-volatile compounds is dramatically higher than the number of volatiles. In spite of this, the amounts of data from LC-MS/MS analysis have been growing exponentially in the last 10 years, and many computational tools for the automatized analysis of such large amounts of data have been proposed. These tools (e.g., DEREPLICATOR+, Mohimani et al., 2018) are aimed to automatize the analysis of LC-MS data with automatic identification of chromatographic peaks, accessing of public spectral databases and/or compound databases, and finally annotation of each peak with the putative relevant metabolite (or a ranked list of possible relevant metabolites). The most recent tools, such as Sirius (Dührkop et al., 2019), provide friendly user interfaces and are very easy to use, performing in minutes the entire workflow from raw LC-MS/MS data to annotated chromatographic peaks (often called *metabolic features* or simply *features* in this context), and are therefore now widely used all over the world.

These tools are very useful in guiding the analysis of LC-MS/MS data and can save researchers a great deal of time, but it should always be kept in mind that the annotations provided by these tools are by no means identifications and must be supported by additional information (see below).

In a provocative paper (Hoffmann et al., 2023), the Böcker research group, the creators of Sirius, presented a computational tool capable of identifying correctly 98% of metabolites based on LC-MS/MS data combined with searching in a compound database. Notably, the authors demonstrated that this astonishing performance was only the result of the most typical errors made in evaluating the performance of such tools. The authors concluded that they doubt that MS/MS data will ever allow reasonably certain automatic identification (>90% correctness), and this holds true only considering the planar structures because, in the authors' words, "establishing the stereochemistry from fragmentation spectra is highly challenging and beyond the power of *in silico* methods".

Still, in many recent papers the list of annotated compounds is uncritically considered as the list of metabolites actually present in plant extracts. This situation is particularly detrimental, because the

unsupported findings are often taken for granted in subsequent studies, thereby cementing putative results as scientific knowledge. At the same time, possible unreported metabolites remain uncovered, either by misassignments or by omitted isolation studies.

However, also the use of retention time plus MS data is prone to wrong assignments. For instance, the herbal medicine 'copaiba' contains three isomeric diterpene acids (**1**, **2**, and **3**) with similar retention times (Fig. 1). Especially compounds **2** and **3** were found to be co-eluting in all analytical LC and LC-MS studies and were therefore neither identified nor quantified, though both compounds are among the major diterpenoids in some of the medicinally used species of *Copaifera* (Santiago et al., 2015). It is questionable whether these compounds could be distinguished by tandem MS. At least for compound **1**, no meaningful fragmentation pattern was observed in a dedicated study (da Silva et al., 2017). In contrast, the three compounds can be easily distinguished by NMR (Çiçek et al., 2018), indicating a different degree of orthogonality.

Many more examples of the false positive identifications based on GC-MS and LC-MS/MS match could be mentioned, but the above should suffice to conclude: a GC-MS or LC-MS/MS match with an authentic standard is not generally to be considered a conclusive identification in phytochemistry, as discussed in detail below.

3.3. Identification confidence levels in phytochemistry

The historical foundations of phytochemistry were a set of rational synthetic and degradative approaches allowing conclusive structural elucidation or a specified number of alternative structures, exemplified by the Fisher elucidation of monosaccharide structures. This approach can be termed classical chemistry. However, conclusive identification of plant metabolites has for decades been accomplished by comprehensive spectroscopic data, more or less supplemented by classical chemistry.

Many metabolites are still identified in this comprehensive way, but many current suggested identifications are based on GC or LC hyphenated with MS/MS. Using this approach, identification is only conclusive when the molecular composition (sum formula) is determined and only one isomer matches LC-MS/MS or GC-MS data and all (!) other isomers can be excluded. Conclusive identification can in principle be carried out using chromatography hyphenated to high resolution MS/MS if all isomers are available as standards or can be distinguished by MS/MS. But the task is more challenging than often realized (Section 3.2). Considering an unusually simple plant metabolite with the formula $C_4H_{10}O$, eight isomers are possible, including all stereoisomers (Fig. 2). If each of these metabolites could be distinguished from the other ones using chiral chromatography combined with MS/MS, match of retention time (t_R) and MS/MS of the metabolite with one authentic standard would enable unambiguous identification. However, an increasing size of the molecule (increasing numbers of atoms), leads to a dramatic increase of the numbers of possible isomers.

Many authors report identification confidence levels with reference to the classification of the Metabolomics Standards Initiative (MSI) (Sumner et al., 2007) or the slightly modified confidence levels proposed by Schymanski et al. (2014). The MSI presents a four-level system with

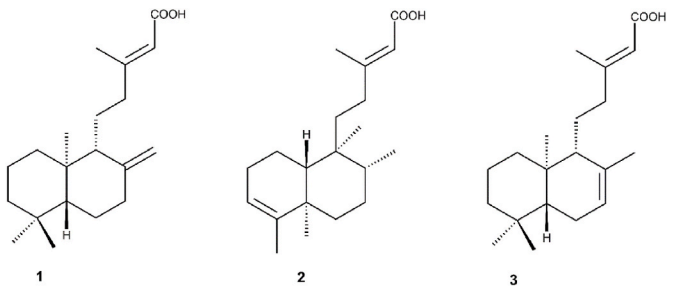


Fig. 1. Chemical structure of *ent*-copalic acid (**1**), kolavenic acid (**2**), and (13*E*)-*ent*-labda-7,13-dien-15-oic acid (**3**).

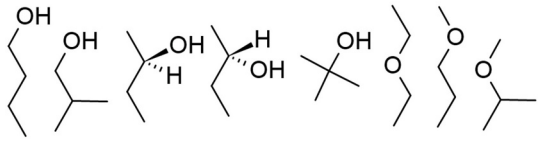


Fig. 2. Potential identities of a metabolite with a molecular formula of $C_4H_{10}O$.

the following confidence levels: identified compounds (1), putatively annotated compounds (2), putatively characterized compound classes (3), and unknown compounds (4) (Sumner et al., 2007). Identification in this system (level 1) requires at least two independent and orthogonal matches with an authentic standard (e.g., t_R and MS or accurate mass and tandem MS), while levels 2 and 3 are achieved without reference standards and rely on physicochemical properties and/or spectral similarities. Level 1 in the MSI system is defined remarkably broad (including e.g., full NMR comparison to an authentic standard!), and the accompanying discussion is still today worth reading (Sumner et al., 2007). However, the very broad level 1 definition in the MSI system renders the system weak in distinguishing between differing levels of relatively high certainty of identification. In a more recent discussion of human metabolomics, "confirmed structure" more specifically refers to a match in at least two independent properties with an authentic standard (e.g., t_R and MS/MS) (Schrimpe-Rutledge et al., 2016).

The number of metabolites in the human metabolome is estimated to be in the order 200,000 metabolites (Schrimpe-Rutledge et al., 2016). However, in the fields of environmental analysis and plant analysis, the potential complexity is much higher. For a broader variety of sample types, including environmental analysis, Schymanski et al. (2014) proposed five levels of identification confidence (including two sublevels) designated as confirmed structure (1), probable structure by library match (2a) or diagnostic evidence (2b), tentative structure (3), unequivocal molecular formula (4), and exact mass of interest (5). The confirmed structure (level 1) in that system is defined as t_R, MS and MS/MS match of the analyte with an authentic reference, which is slightly higher than the MSI minimum criteria, because MS/MS match is included. If "possible", the match should be tested with an orthogonal method (Schymanski et al., 2014), i.e., chromatography using another system. The Schymanski levels are generally accepted in environmental analysis (Alygizakis et al., 2023), which is the most complex of all analytical fields because any human made or natural compound, including any plant, animal or microbial metabolite, may be found in environmental samples, including a wide variety of plant metabolites (Nanusha et al., 2021; Liang et al., 2023).

The entire plant kingdom includes a staggering chemical complexity, but the metabolome of any individual plant species includes only a tiny subset of all plant chemicals, and the correct identification of this subset is essential for progress in plant biochemistry (Forman et al., 2022; Mancinotti et al., 2022; Zhao et al., 2023; Sun et al., 2023). Indeed, publishing false structures is devastating for later attempts to understanding the interplay between genome, proteome and metabolome in the investigated species as well as for any chemophenetic interpretation of literature data. This situation results in a requirement for higher identification standards in plant science than in some other fields. Extrapolating the number of alternatives in the very simple example above ($C_4H_{10}O$, Fig. 2) to complex metabolites, e.g., $C_{27}H_{30}O_{16}$ (Fig. 3) or $C_{62}H_{69}O_{34}$ (Tatsuzawa, 2019), it is evident that matching t_R, MS, and MS/MS with a standard isolated from another species is far from constituting a conclusive identification of a complex plant metabolite with multiple atoms and chiral centers. Phytochemical identification confidence levels must therefore match the molecular complexity of plant metabolites and the critical importance of the correctness for each structure (Table 1).

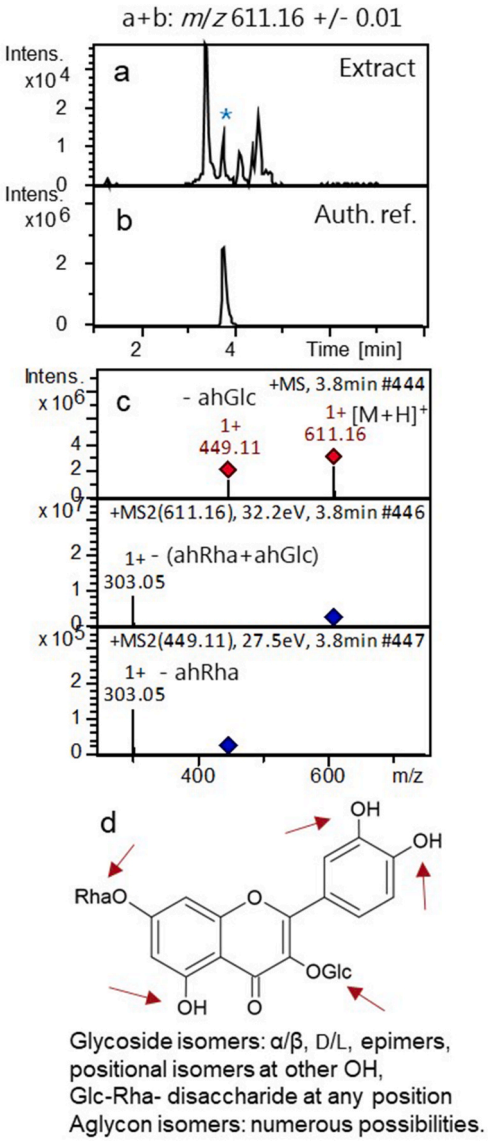


Fig. 3. Lack of conclusive identification despite comparison with an authentic standard. Panel a: extracted ion chromatogram of a plant crude extract chromatogram focused at $C_{27}H_{30}O_{16}$ alias *m/z* 611.16±0.01 Da, revealing a rather narrow retention time window almost saturated with peaks. A tentatively identified peak is labelled with an asterisk (Cárdenas et al., 2023). Panel b: extracted ion chromatogram focused at the same *m/z* value of an authentic standard of the flavonoid quercetin 3-*O*-*β*-D-glucopyranoside-7-*O*-*α*-L-rhamnopyranoside, showing similar retention time to the labelled minor peak in panel a. Panel c: The MS and MS^2 spectra of the standard in panel b, composed of the $[M+H]^+$ adduct as well as fragments due to loss of anhydro-sugars, i.e., either the glucose moiety, the rhamnose moiety or both. The simplicity of this MS^2 spectrum is typical for spectra of complex plant metabolites. Panel d: Structure of the analyzed standard, with arrows and cues indicating potential sites of isomerism. All indicated potential isomers would most probably show identical or very similar MS^2 (or MS^n) spectra.

3.4. Using identification confidence levels in plant metabolite profiling

In situations, where an authentic standard is not available, MS/MS data, combined with either X-ray crystallography (Pedras and To, 2016; Kariya et al., 2023) or NMR spectroscopic data (Tatsuzawa, 2019; Trabelcy et al., 2021), possibly supplemented with optical rotation or electronic circular dichroism spectroscopy results (Su et al., 2023; Liu et al., 2024), are needed to obtain truly confirmed structures of each metabolite reported. This then includes confirmed stereochemistry of all chiral centers. Even for apparently simple metabolites, NMR investigations can lead to revisions of previous assignments (Fechner et al., 2018). For complex metabolites like complex steroid glycosides or multiply glycosylated and/or acylated flavonoids (Vincken et al., 2007; Alseekh et al., 2020), the number of possible stereoisomers will be so high that additional classical chemistry degradation and comparison with authentic references, or x-ray data, will be needed in order to obtain identification confidence level A (Xu et al., 2023). For known metabolites, level A1 identification is essentially a "replication" of the original discovery.

From an attempt to avoid such a replication, or efficiently focusing on previously unknown metabolites, less conclusive protocols will typically be employed in cases where an authentic standard or comprehensive NMR spectral data is available (Li et al., 2019; Liang et al., 2021). For structurally simple metabolites (Fig. 2), exhaustive comparison with chiral chromatography and MS/MS may be realistic and equally conclusive, and then would constitute level A2 identification (Table 1). However, since all possible isomers will be available in exceptional cases only, level A2 will usually require protocols including NMR for conclusive comparison with the standard compound or NMR spectral data (Holscher and Schneider, 1999; Waridel et al., 2004; Aberham et al., 2010; Richard et al., 2013; Brkljaca and Urban, 2015; Liang et al., 2021; Hoffmann et al., 2022; Renz et al., 2024). Fortunately, modern NMR is highly sensitive and can provide conclusive ^{1}H NMR spectra (and indirectly ^{13}C chemical shifts from various 2D spectra) from a fraction of a milligram, so routinely automated (Staerk et al., 2009; Li et al., 2019; Liang et al., 2021) or manual (Agerbirk et al., 2014, 2015) interfacing the liquid chromatograph and the NMR spectroscopic instrument is feasible. Such an identification of a known structure in another plant species may consist of comparison to the authentic standard with respect to t_R, MS and less than comprehensive NMR data (e.g., simple one- or two-dimensional ^{1}H NMR). If the combined data can exclude all other isomers including enantiomers, e.g., by optical rotation data, the identification is to be considered conclusive (level A2). The classical method, testing for an unchanged melting point after mixing with an authentic standard, can also be used as part of level A2 identification, but is obsolete when analyte amounts are limited.

In many cases of partial structure elucidation, the entire structural connectivity of atoms and the relative configuration of the molecule are established, while its absolute configuration is not. In these cases, identification confidence level B is obtained. Level B evidence is still quite strong, mainly because enantiomers are rare or even undescribed from natural sources in many groups of metabolites. Level B identification is commonly reported (Agerbirk et al., 2014, 2015; de Graaf et al., 2015; Rasmussen et al., 2022; Mori et al., 2023; Krishnan et al., 2023). In cases where level B identification relies on very likely assumptions, such as glucose or glucose moieties being the D-isomer, standard amino acids or amino acid residues being the L-isomer (Dai et al., 2022), or where another biochemically conserved backbone can be assumed (Rasmussen et al., 2022), this is often considered as almost conclusive. Nonetheless, L-sugars and D-amino acids certainly do occur in nature. With modern high NMR sensitivity, obtaining reliable optical rotation measurements can be the limiting factor due to insufficient purity or limited analyte amounts, then resulting in level B identification (Hansen et al., 2022).

When an authentic standard is available, careful comparison using GC-MS or LC-MS/MS data is relatively informative, but less than assumed by Schymanski et al. (2014), leading only to identification confidence level C in phytochemistry (Aguiar et al., 2021; Clancy et al., 2023; Paguet et al., 2023). The underlying reasons of this problem are the high numbers of potential and actual isomers frequently encountered, the relatively short retention time windows in which they are to be expected, and the relatively non-informative MS/MS fragmentation patterns typically exhibited by plant metabolites when using common online compatible ionization techniques (Fig. 3). It is generally acknowledged that level C evidence in previously characterized

Table 1
Proposed identification confidence levels in phytochemical metabolite profiling, and comparison with two previously suggested systems aimed at medical metabolomics (MSI) (Sumner et al., 2007) and environmental analysis (Schymanski et al., 2014).

Phytochemical identification confidence levels proposed here	Corresponding general metabolomics confidence levels	
	Sumner et al. (2007)	Schymanski et al. (2014)
A. Confirmed structure including confirmed stereochemistry. Complete identification including isomer distinction by comprehensive data such as: **A1.** Comprehensive NMR and MS/MS or x-ray crystallography supplemented by chiroptical properties and classical chemistry as needed for stereochemical deductions. **A2.** Retention time and high res. MS/MS or partial NMR match with an authentic standard and confirmed lack of match with all possible isomers including all stereoisomers.	(Level 1)	(n/a)
B. Confirmed structure except for one or more stereochemical aspects. As level A, except for at least one remaining unresolved chiral center.	(Level 1)	(n/a)
C. Tentative identification matched with a standard compound. Match of at least t_R, MS and MS/MS with an actual authentic standard analyzed in parallel, preferably supported by other online data such as a UV–vis or ^{1}H NMR match.	Level 1	Level 1
D. Tentative identification based on libraries, model compounds etc. Match of t_R, MS and MS/MS with predictions from libraries, data from well-selected homologues or *in silico* predicted data. **D1.** Match with GC-MS library, including retention index or any future high-quality LC-MS/MS libraries validated for the particular class of metabolites → Relatively reliable evidence. **D2.** Match with current LC-MS/MS libraries not yet validated exhaustively or by detailed comparison with homologues or other model compounds → Relatively poor evidence.	Level 2	Level 2
E. Tentative candidate or tentative identification of metabolite class. Deductions from suitably informative data (t_R, high res. MS and suitably complex MS/MS data, poorly resolved NMR data) that might be compatible with expected (e.g., known) structures, but also with many other structures. Identification of metabolite class (MSI level 3) is also included here.	Level 3	Level 3
F. Data with no structural interpretation possible. Sum formulae, *m/z* values etc., with no available meaningful structural interpretation.	Level 4	Levels 4-5

genotypes (Ranner et al., 2023) is close to being conclusive and the only realistic level to be accomplished in current routine analysis. Using Fig. 3 as an example, since the authentic standard in panel b was isolated from the same genotype as analyzed in panel a, the assignment of the labelled peak in panel a was reasonable. However, for material never conclusively investigated before, the risk of mis-identification is evident. Understanding the phylogenetic distributions of various kinds of metabolites (Section 6.3.) may lead to a more qualified understanding of the situations where level C confidence can be considered satisfactory.

When lacking authentic standard compounds, a considerable number of common hydrocarbons and functionalized hydrocarbons, including terpenoids etc., can be identified with reasonable certainty using GC-MS data combined with high-quality libraries, based on retention time indices (level D1). This well-established method is generally considered reliable within the boundaries of the relatively simple pool of compounds covered by the libraries such as the often used "NIST" library (Aguiar et al., 2021; Clancy et al., 2023; Paguet et al., 2023) or the Adams library for essential oils (Adams, 2007).

When only LC-MS/MS analyses are possible, current libraries are far from exhaustive and the performance of various automated peak identification procedures are typically not yet validated, leading to level D2 identification, a relatively poor level of confidence. Satisfactory validation of a library and matching algorithm should be carried out for the specific compound class intended to be used with the library, e.g., alkaloids, flavonoids or triterpenoids. This would encompass testing the output of the algorithm with an extensive, suitably challenging set of common and rare known metabolites and isomers. The testing set must include synthetic isomers of natural metabolites or yet unidentified natural isomers not corresponding to any compound in the library, for judging the ability of the algorithm to recognize lack of match. Test procedures should ideally be designed by scientists at arms-length from developers of the library and algorithm, and successively subjected to peer-review. Returning to Fig. 3 as example, using a library only, the analyst would be clueless concerning which of the closely eluting peaks in panel a to assign as the single matching flavonoid (panel d) known for the species. As a literature example, glucosinolate profiles deduced at level D2 with the database "MWDB version 2" (Wang et al., 2022) of two varieties of canola (*Brassica napus*) were in conflict with glucosinolate profiles of 281 accessions of the same species identified at level C (Missinou et al., 2022). For statistical reasons and the higher certainty of C than D2, the conflicting report is hence considered wrong. For the individual analyst, preparing a tailor-made library of relevant reference compounds analyzed in parallel may be at least as strong circumstantial evidence as currently available LC-MS/MS data libraries (Ranner et al., 2023). Comparison with reference compounds is also very useful for proving lack of identity with all previously reported isomers (Agerbirk et al., 2022). In a well carried-out case, thorough analysis at identification confidence level D2 integrating UV–vis and MS data was used for distinction of three *Populus* species, and the authors took care not to over-interpret the MS/MS data (Alcalde-Eon et al., 2016). For dereplication, LC-MS/MS is an efficient way of guiding attention to previously unknown metabolites (Tang et al., 2019).

Although current LC-MS/MS libraries are by far not as powerful as GC-MS libraries, attempts to develop and validate LC-MS/MS analysis supported by exhaustive libraries is to be encouraged. Indeed, reliable methods validated for relevant metabolite classes would be of enormous relevance for phytochemical research. Match with such improved LC-MS/MS databases could qualify as level D1, if the database quality matched current GC-MS libraries. However, while databases for use in mammalian metabolomics are becoming increasingly reliable (Schrimpe-Rutledge et al., 2016), the authors are not aware of exhaustively validated plant metabolite libraries (Oberacher et al., 2020).

Confidence levels E and F are so faint that any structural proposals are not worth mentioning in abstracts or citing in later papers or critical reviews. However, discovery of correlation between a level E/F signal and a gene or another biological phenomenon represents an important stepping-stone for further research and can therefore still constitute a valuable scientific result (Nemesio-Gorriz et al., 2020).

In summary, use of authentic standards or reference materials is essential in studies based on LC-MS/MS data and allows level C identification. Except for very simple metabolites, inclusion of NMR data is needed to obtain truly conclusive identification at level A or B. The use

of standards and references should also be transparent, citing the origin and evidence for each of the standards used.

3.5. Detection limits and the need for reporting levels quantitatively

Detection limits are an important matter in phytochemical studies, even for qualitative aspects in focus in this review. In a technical sense, the detection limit refers to the analytical method used, e.g., a particular LC-MS or GC-MS set-up, describing the lowest concentration of analyte that can be detected, but not necessarily quantified as an exact value (EMA, 2018). The limits are usually estimated from signal-to-noise ratios or calculated from the standard deviation and slope of the calibration curve. Although a particular scientist may feel that the limit of detection in his or her work is incredibly low, history tells us that within just a few years, routine detection limits may be orders of magnitude lower. The developments in MS/MS detection are a good example of this trend. Realizing the need for stating detection limits automatically leads to a universal requirement for reporting approximate levels of detected metabolites. Even if a reported level is tentative, e.g., assuming similar detector response as a generally available external standard, specifying a level allows future authors to test the finding and critically compare apparently contrasting reports. Thus, reporting presence as well as absence of a metabolite is meaningless unless an approximate level and detection limit is specified. Though the main focus of the present paper is on qualitative aspects, i.e., presence or absence of a specific metabolite in a given source, quantitative aspects are of course also relevant and quantitation results are influenced by a multitude of factors, including, but not limited to, sample preparation, extraction procedure, completeness of extraction, solvents used for extraction, extraction temperature, number of extraction cycles, etc. Such differences will be more relevant for certain compound classes, e.g., larger molecules and polymers. In dedicated studies, where exact quantitation is paramount, internal standards to test recovery rates, and other parameters also used in quality control of pharmaceuticals have to be considered (Çiçek et al., 2018). Additionally, semiquantitative (Zidorn and Stuppner, 2001a) and quantitative differences (Zidorn et al., 2002) in concentrations of natural products in specified plant organs might also be used as similarity criteria in chemosystematics/chemophenetics.

4. Chemical reactivity before and during extraction and separation

In studies reporting specialized compounds from plants, the authors usually imply that the compounds isolated or detected are genuine plant natural products, i.e., compounds biosynthesized by the plants investigated. This assumption however is not necessarily true, because of the risk of chemical reactions during extraction and manipulation. Reporting metabolites after lengthy preparative procedures involving large-scale extraction, evaporation, preparative chromatography etc. should include demonstration of the metabolites in gently treated analytical scale GC or LC of crude extracts or minimally processed extracts.

At least three kinds of artefacts can be distinguished, based on the kind of control experiment used to reveal them. In some cases, presumed genuine plant natural products are in fact artefacts, involuntarily synthesized in the course of the isolation (or even structure elucidation) procedure (Section 4.1). In other cases, the original genuine plant natural products are prone to decompose, e.g., by elimination, oxidation or hydrolysis. This degradation step of the genuine plant natural product can occur at any time from sampling to detection (Section 4.2.). A third, physiologically interesting kind of “artefacts” are produced by involuntary activation of biochemical two component defense systems during extraction, typically because of incomplete inactivation of the activating glycosidase enzymes (Section 4.3.). Besides artefact formation, drying, prehandling, and storing of plant materials can lead to complete decomposition, and thus from an analytical point of view "disappearance" of previously present plant metabolites. It is thus paramount to critically check, both as an author and a reader, the time span between plant collection and analysis when evaluating analysis results (Di Lecce et al., 2022).

4.1. Artefact formation by reaction with solvents

One of the many possibilities of artefact formation is the reaction of genuine plant natural products with one of the organic solvents employed in extraction and column chromatography. An example from our own research is ring opening and ester formation starting from genuine sesquiterpene lactones after exposure to methanol during the extraction process (Fig. 4) (Grass et al., 2004). Designing control experiments for this group of artefacts is simple, yet efficient. If chemical reaction with a solvent is suspected, extraction and subsequent manipulations can be tested in a slightly different solvent. For example, if methyl esters are suspected to be artefacts of methanol used as solvent, repeating the procedure with ethanol will immediately clarify the situation.

4.2. Artefact formation due to inherent instability of the isolated metabolites

Discovering artefacts due to oxidation, dehydration or other inherent instability is less simple than the example above. Even though oxidation can be minimized by evaporation under e.g., N_2, the presence of oxygen throughout harvest, drying and extraction is difficult to avoid in preparative work. Deliberately increasing oxidation in a controlled way (e.g., repeated evaporation under an air-stream) is a simple way of testing whether an analyte might be an oxidation-artefact. However, the best control for such artefacts is analytical scale extraction of freshly harvested material at gentle conditions. Adding thiols or other antioxidants is also common practice in some fields, notably in enzyme work, but a control without the added agent is always advisable in metabolite studies.

4.3. Biochemical turnover during extraction

Some genuine natural products are stored within the plants in special cells or cell compartments and are only stable in undamaged living plants, while disruption of cell integrity leads to swift decomposition

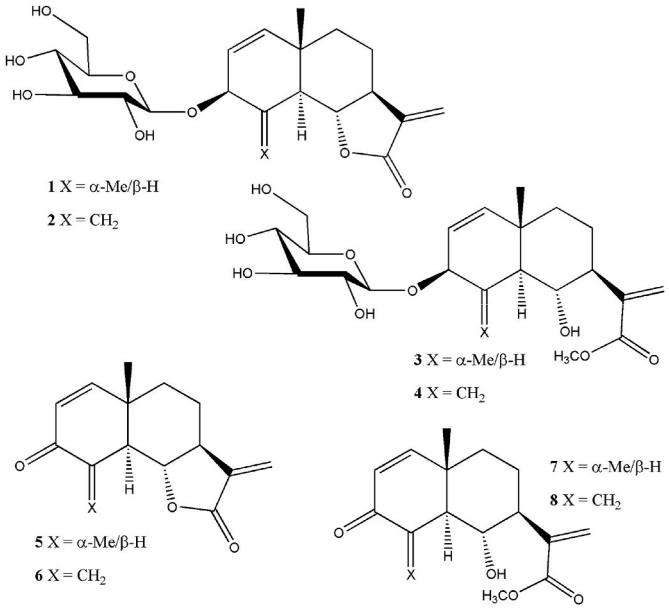


Fig. 4. Genuine natural products (**1–2**, **5–6**) and isolation artefacts (**3–4**, **7–8**) from *Hieracium intybaceum* All.

and successive formation of other often toxic/bioactive breakdown compounds, which in many cases act as defense compounds. Well-known examples are the compound classes of cyanogenic glycosides (Yulvianti and Zidorn, 2021) and glucosinolates (Blažević et al., 2020), where genuine natural products are activated in case of tissue damage to form toxic and feeding deterrent hydrogen cyanide and isothiocyanates, respectively. Numerous less well-known analogous systems exist (Morant et al., 2008; Nomura, 2017; Friedrich et al., 2022).

The suitable control experiment for artefacts caused by endogenous enzymes is extraction at conditions that immediately inhibit endogenous enzymes. Hot solvent is the traditional remedy (e.g., boiling 70% aq. MeOH) (at 70 °C) but simply keeping the organic component high can in some cases inhibit endogenous enzymes sufficiently (e.g., room temperature 80–100% MeOH). Conditions suitable for the specific plant material should be determined experimentally case by case. For example, the robust endogenous hydrolases in white mustard seed were active in 70% aqueous MeOH, almost inactivated by 1 min in boiling solvent and completely inactivated after three times 1 min in the boiling solvent (Agerbirk and Olsen, 2012).

In other instances, such as *Senna* and *Chelidonium*, (Lemli and Cuvelee, 1978; Paulsen et al., 2015), artefact formation already occurs at the stage of drying of the plant material. For example, the *post mortem* changes of the genuine main alkaloid of *Chelidonium majus* L. to the main alkaloid in dry plant material are depicted in Fig. 5.

5. Coverage of previous literature

5.1. Complete literature coverage

Before concluding presence or (in particular) absence of a metabolite based on literature data, all previously published results must be considered. It has to be taken into account that current literature databases are surprisingly inefficient in searching for and retrieving plant species names. In addition, species actually contained in multispecies screens are often not easily found using these data bases. In order to avoid such problems in the future, scientific names of the investigated species or genera should be included either in the title or the keywords.

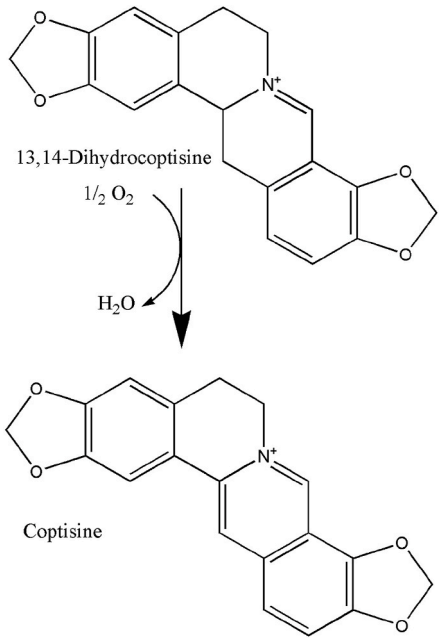


Fig. 5. Conversion of 13,14-dihydrocoptisine to coptisine during the drying process.

5.2. Consideration of synonyms of scientific plant names

More than a million scientific species names have been published for the approximately 400,000 currently known plant species (Dauncey et al., 2016). Synonyms are therefore common and many are still in use. Chemical databases do (currently) neither update published scientific names of source species, nor do they include synonyms in their search algorithms. Thus, e.g., older records for arbutin from *Arctostaphylos uva-ursi* (L.) Spreng.; need (in part) to be searched under the older synonym *Arbutus uva-ursi* L.

One out of hundreds of additional possible examples concerns a fern species named *Oreopteris limbosperma* (All.) H.P.Fuchs in the current Danish flora by Frederiksen et al. (2006). The same fern species was named in a previous flora from the same publishing house (Hansen, 2004) *Thelypteris limbosperma* (All.) H.P.Fuchs, indicating (only) *Dryopteris oreopteris* (Ehrh.) Maxon and *Thelypteris oreopteris* (Ehrh.) Slosson as possible synonyms. Also, the currently accepted name according to www.worldfloraonline.org is *Thelypteris limbosperma* (All.) H.P.Fuchs. Here, a total of 22 synonyms assigned to 10 different genera can be found (as of July 16, 2023).

Commercial providers of seeds and plants are notorious for conservatism in the use of outdated scientific names, which subsequently find their way to the scientific literature. Several online databases exist, and their status are still in some flux. A couple of authoritative databases should be consulted before literature search and publication. Examples of general databases are worldfloraonline (www.worldfloraonline.org) and plants of the world online (powo.science.kew.org). For some groups, additionally, dedicated databases should be consulted, such as the BrassiBase (brassibase.cos.uni-heidelberg.de) for the Brassicaceae, the Global Compositae Database (www.compositae.org) for the Asteraceae, the Legume Data Portal (www.legumedata.org) for the Fabaceae, or the Cichorieae Portal (www.cichorieae.e-taxonomy.net) for the Cichorieae tribe of the Asteraceae family.

6. Data interpretation

6.1. Pitfalls in interpretation of hyphenated MS data

Errors in qualitative MS interpretation are of at least two classes: misinterpretation of the individual spectrum (i) and lack of distinction of confidence levels (ii). A recent review discusses potential quantitative errors in depth (Alseekh et al., 2021).

An example of the first class is wrong identification of molecular ion or molecular adduct ion. If the true molecular ion fragments entirely, the proposed molecular mass and composition of the analyte will be wrong. Sometimes the proton adducts, but not the sodium or potassium adducts will fragment, in which case the true mass can be inferred (Fig. 6). Comparison with authentic standards will readily reveal these artefacts, as can *in silico* methods if sufficiently advanced. The ultimate error in MS interpretation is the claiming of previously unknown structures solely based on (often superficially interpreted) MS data. Unfortunately, even this error is frequent in modern plant metabolomics literature.

The second class, misinterpretation of the actual confidence level, is extremely frequent in the current literature, usually in metabolomics investigations carried out without authentic standards. Quite possibly, some authors may believe that results are only publishable if conclusiveness is claimed, or simply do not realize the number of alternative isomers that would also be in agreement with detected *m/z* values and fragmentations. It should however be stressed that the true quality of a report is not proportional to the confidence level claimed, but related to the correctness of data interpretation. Correctly interpreting the level of confidence is the hallmark of a good paper in metabolomics. The keys to obtaining quality reports are transparent distinction of identifications supported by authentic standard in contrast to suggestions not backed up by standards, as well as transparent reporting of the evidence behind tentative identifications. Some recent examples of systematic and

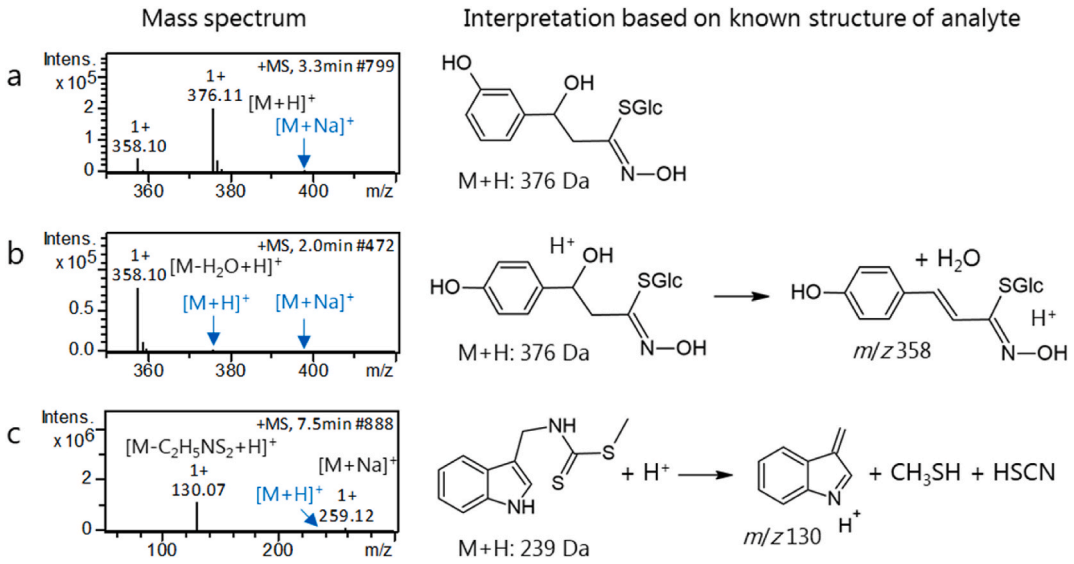


Fig. 6. Examples of diminutive proton adducts due to simple neutral losses in MS, resulting in resonance stabilized fragments that may be wrongly interpreted as $[M+H]^+$ ions. The true molecular formula could be concluded since authentic standards were used, but Na^+ and K^+ adducts also indicated the problem. Panel a, MS and structure of a desulfoglucosinolate with *meta*-substitution and hence moderately inclined for dehydration. Panel b, corresponding MS and structure of an isomer with *para*-substitution and thus highly inclined for dehydration, forming a resonance-stabilized product (Agerbirk et al., 2015). Panel c, MS and structure of the indole phytoalexin brassinin, subject to complex MS fragmentation, because a resonance-stabilized product is formed (Cárdenas et al., 2023).

precise reporting of the varying levels of confidence are listed here (Andini et al., 2019; Poveda et al., 2021; Missinou et al., 2022).

6.2. Pitfalls in the interpretation of NMR data

Major causes for erroneous compound identification or elucidation are the wrong interpretation of NMR spectra or an insufficient number of NMR experiments. Especially, interpretations of heteronuclear multiple bond correlation (HMBC) spectra are prone to misinterpretations. Firstly, the HMBC originates from two-bond and three-bond ^{13}C–^{1}H couplings (^{2}J(CH) and ^{3}J(CH)), which vary over wide and overlapping ranges of values (typically 0–8 Hz and 0–12 Hz, respectively). Therefore, an HMBC experiment is inherently unable to distinguish between two-bond and three-bond ^{13}C–^{1}H couplings, and mistaking a ^{2}J(CH) for a ^{3}J(CH), or vice versa is one of the most common reasons for HMBC misinterpretations (Mangoni, 2012). Several experiments have been proposed to detect selectively ^{2}J(CH) or ^{3}J(CH) couplings (Saurí et al., 2015), but none of them is exempt from problems and none has entered the standard repertoire of natural product chemists.

Secondly, the intensity of the observed HMBC cross peaks strongly depends on the chosen delay times in the pulse programs (Furrer, 2014). Routine HMBC experiments are recorded with a compromise delay time, suitable for a coupling constant of about 8 Hz (Furrer, 2012). This allows optimal correlation signals for coupling constants between 6 and 10 Hz with, however, weak or absent signals for smaller coupling constants. Intensity of HMBC couplings also depends on the fine structure of the relevant proton signal: evolution of the homonuclear couplings during the delay times in the pulse program may strongly affect the intensity of the signal, and in unfavorable cases may lead to the complete cancelation of a correlation peak, even if the relevant ^{13}C–^{1}H coupling is optimal. Though there have been approaches to overcome the problem of weakly observed cross correlations, such as the ACCORDION excitation or *J*-compensated HMBC experiments, these techniques have not been found superior to a set of two different HMBC experiments (e.g., one for a coupling constant of 3–4 Hz and one for a constant of about 8 Hz) (Furrer, 2014).

Another problem may come from particularly strong ^{4}J(CH) couplings (as may occur, for example, in unsaturated systems or in bridgehead polycyclic systems), which may be misinterpreted as weak ^{2}J(CH) or ^{3}J(CH) correlations. Any misinterpretation of an HMBC experiment may result in structures with, e.g., erroneous substitution patterns, bonds, or ring linkages. As long as the analyst is aware of these circumstances, however, the HMBC experiment is a powerful (and in most cases essential) tool for structure elucidation.

Also in other heteronuclear experiments, such as the often-used HSQC (heteronuclear single quantum coherence) experiment, intensities (or coupling constants, respectively) are different for different functional groups (Çiçek et al., 2019). As ^{1}J(CH) couplings typically range from 125 to 200 Hz, optimization of delay times for 145 Hz provides clear signals for all compounds. Though in this experiment only ^{1}J(CH) couplings are detected, eventual impurities might as well lead to misinterpretations, especially when the impurity-derived signals show experiment-favored coupling constants. A recent study on the mis-assignments in marine natural products found more than 200 reported wrong structures within the last decade (Shen et al., 2022), mostly because of misinterpretation of NMR spectra. As significantly more studies deal with (terrestrial) plant natural products, this number is likely to be surpassed for these.

As crucial as correct data interpretation is selecting (and analyzing) the necessary set of NMR experiments. A literature search for reported natural 3′,5′-disubstituted isoflavones (Fig. 7, left) yielded 13 compounds supposedly showing this substitution pattern (Çiçek et al., 2022). However, a closer look on the reported data revealed that the chemical shifts fit much better with a 3′,4′-substitution pattern (Fig. 7, right). The wrong assignments resulted from a higher order spin system and intricate proton signals, which at first glance suggested *meta*-substitution. However, in this case ^{13}C shift values did not support this pattern and clearly suggested a 3′,4′-substitution pattern. The fact that the missing substitution in position 4′ was implausible from a bio-synthetical point, could have hinted at a probably wrong structural feature. In addition, the application of NMR prediction tools could have been of help for those being less familiar with plant biosynthesis. Such

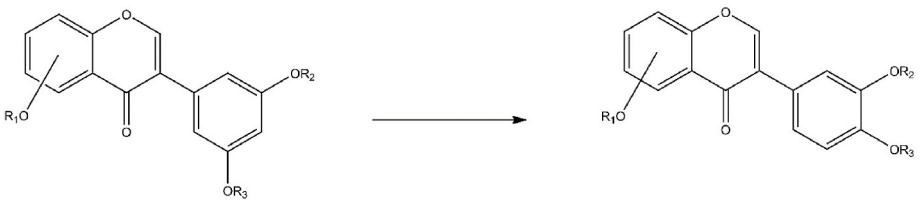


Fig. 7. Supposedly erroneous structure reports (left) and corrected structures (right) after thorough evaluation of shift values and coupling patterns. R_1=H, CH_3, or β-D-glucopyranoside; R_2/R_3=H or CH_3.

prediction tools are free of charge, e.g., the CSEARCH protocol (Robien, 2019), nmrshiftdb2 (Kuhn and Schlörer, 2015), and nmrdb.org (Binev et al., 2007), or commercially available (ACD/Labs). Regarding the mentioned isoflavones, nmrdb.org and nmrshiftdb2 predictions clearly highlight the strongly differing shift values. Interestingly, using the CSEARCH protocol, 3′,5′-dihydroxy and 3′,5′-dimethoxy substitutions are detected as erroneous, while no mistakes are found for the 3′-hydroxy-5′-methoxy variant (in the HOSE code prediction, but not in the neuronal network prediction).

However, even a full set of conducted experiments does not necessarily mean that all relevant conclusions were made from the data. For instance, the diterpenoid chagresnone and its deacetylated form were reported from *Myrospermum frutescens* Jacq. (Fabaceae) after isolation and subsequent hydrolyzation, respectively (Torres-Mendoza et al., 2004). After comprehensive NMR analysis the structure in Fig. 8 (left) was proposed, using NOESY experiments to determine the *cis* configuration of the cyclopropyl ring. Though elucidation of the structure and configuration of the cyclopropyl ring were accomplished in this study, key correlations for the orientation of the cyclopropyl group were missed in the NOESY spectrum and α-orientation instead of the correct β-orientation (Fig. 8, right) was suggested (Çiçek et al., 2023).

In particular diterpenoids are prone to inconsistent reports, due to the presence of "normal" and "wrong"-configurations, the latter being marked with an "*ent*"-prefix (Seaman et al., 1990). In particular, the use of the "*ent*"-prefix is neither mandatory nor generally applied, i.e., kolavenic acid being an *ent*-clerodane-type diterpenoid (Fig. 1). Apart from the use of the "*ent*"-prefix and the differentiation by the compounds' optical rotation, enantiomers may as well bear different trivial names. One example concerns the four naturally occurring stereoisomers of polyalthic acid, differentiated by their configurations in positions 4, 5, 9, and 10 (Fig. 9). Out of these four stereoisomers (-)-polyalthic acid and daniellic acid possess the "wrong" configuration, being epimers at position C-4, while the two "normally" configurated C-4 epimers are named (+)-polyalthic acid and lambertianic acid. Thus, one pair of enantiomers is denominated as such, whereas the other is not. However, in this case also the use of the "*ent*"-prefix is of no help, as the "*ent*"-configurated (-)-polyalthic acid was isolated first and named solely polyalthic acid. These ambiguities led to confusions in several publications as well as in some large databases, i.e., the name *ent*-polyalthic acid is leading to the structures of both, (-)-polyalthic acid and daniellic acid (Reaxys), or to daniellic acid only (SciFinder) (Çiçek et al., 2020). Alas, also less-complex structures are found erroneously in the databases, i.e., the structure of cirsimaritin leading to the name salvigenin or the search for 3′-*O*-methylrosmarinic acid giving both, the structure of the correct compound as well as the structure of 3-*O*-methylrosmarinic acid (clinopodic acid B) with equal priority (Reaxys).

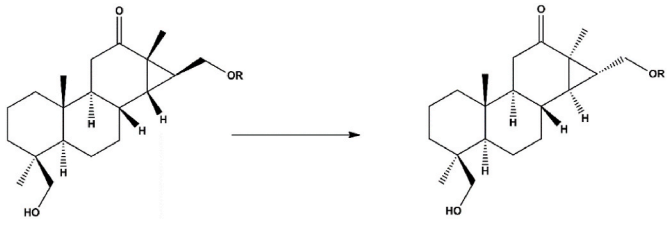


Fig. 8. Erroneous structure report (left) and corrected structures (right) of chagresnone (R=Ac) and deacetylchagresnone (R=H).

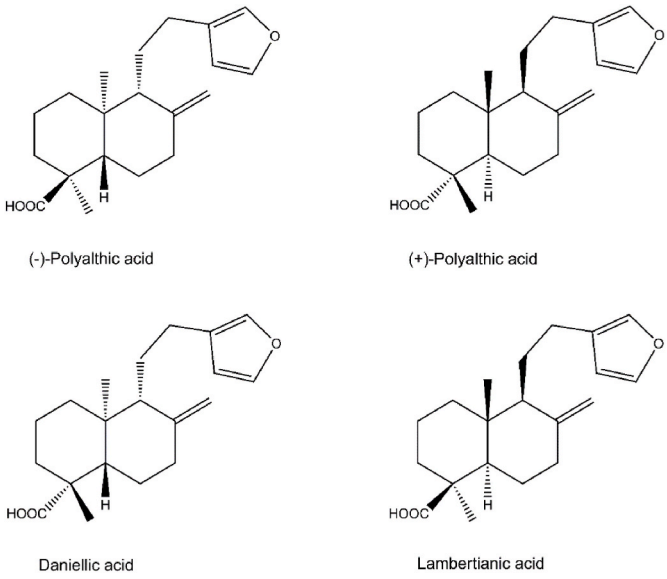


Fig. 9. Stereoisomers of polyalthic acid.

Summarizing, the elucidation or assignment of natural products by NMR spectroscopy is prone to errors and requires special attention. Thereby, a sufficient number of experiments and their comprehensive analysis is of utmost importance to not only correctly elucidate/assign the correct molecular structure, but also the correct configuration. In addition, the identified structures can be confirmed by comparison with NMR prediction tools. However, also biosynthetic pathways should be considered to evaluate the probability of the identified compound being a natural product. In order to confirm/exclude eventual structures the use of databases specialized on plant natural products, i.e., the COCONUT database (Sorokina et al., 2021) or the KNApSAcK family database (Afendi et al., 2012) could (at least theoretically) be of help. For instance, by searching for the name polyalthic acid in the KNApSAcK database, two hits for polyalthic acid are obtained. Though, they are not specified as the two enantiomers – both chemical structures show (-)-polyalthic acid and one of the two InChi Keys gives a wrong stereoisomer – at least the CAS IDs will lead to the two naturally occurring stereoisomers. Similarly, in the LOTUS database, the InChI or SMILES codes, but not the name of (-)-polyalthic acid (and also daniellic acid) will lead to the searched compound. The COCONUT database instead, presents a structure without defined stereochemistry for polyalthic acid along with the name illurinic acid, which is a synonym of daniellic acid and therefore not the correct compound. In addition, all "known" stereochemical variants are shown, which in fact are all theoretically

possible 16 isomers resulting from 4 stereocenters and not those that are naturally occurring. Hence, also natural product databases have to be regarded as what they are, a collection of correct and incorrect literature reports.

6.3. Chemophenetics rather than chemotaxonomy

In one field of plant metabolite profiling, variously named either chemosystematics or chemotaxonomy, taxonomic interpretation of metabolite profiles is still sometimes attempted, although this interpretation is outdated. This approach was based on a perception of secondary/specialized metabolites as ecologically neutral, but this perception has been completely abandoned by modern science (Ehrlich and Raven, 1964; Wink 2003; Wink et al., 2010; Züst et al., 2012; Negin and Jander, 2023). Modern plant phylogeny is based on macromolecule sequences (e.g., One Thousand Plant Trancriptomes Initiative, 2019), not metabolite profiles. In contrast, phytochemical profiles are still of significant interest to characterize the chemical/biochemical properties of taxa in a phylogeny, including similarities and differences within and between monophyletic groups. This approach has been termed chemophenetics to delimit it from former chemosystematic/chemotaxonomic studies. Notably, the experimental approach already used by chemotaxonomists, systematic chemical/biochemical comparison of multiple species, have remained unchanged in chemophenetics. Only the interpretation is radically different (Zidorn, 2019).

The basic publishable entity in this context is a chemophenetic report [formerly named chemosystematic report (Zidorn, 2008)], i.e., the reporting of one natural product from one source taxon. Compilations (reviews) of chemophenetic reports are the basis of any overview of the chemical diversity of plants. Many good classical chemotaxonomy papers were in reality such compilations, interpreted with due consideration of the possibility of convergent evolution or horizontal gene transfer (Griffin and Lin, 2000). The advantage of phylogenies created independently from metabolite profiles is enormous, as DNA based phylogenies can be matched with metabolite profiles to allow ecological, biosynthetic or evolutionary interpretation (Wink, 2003; Windsor et al., 2005; Alseekh et al., 2020; Czerniawski et al., 2021; Agerbirk et al., 2021; Okamura et al., 2023). Such comparison would of course be statistically meaningless if using phylogenies based on metabolite profiles. Data for such modern work can often be extracted from classic and recent high-quality papers published as chemotaxonomy (Maffei et al., 1996; Nørbæk et al., 2002; Braunberger et al., 2015).

7. Summary and concluding checklist authors and editors

In this viewpoint, essentials and potential pitfalls in plant metabolite profiling in the approximate sequence of a scientific investigation, from the planning and collection of material, over the analytical process, ending in discussion, interpretation and scientific publication have been discussed. We stress that the overall conclusiveness of even a qualitative phytochemical report relies on a combination of the conclusiveness of the compound identification and of the botanical identification. This point can be illustrated in a phytochemical quality matrix (Fig. 10). For quantitative work and physiological or ecological investigations, further dimensions should be considered (quantification, replication and sampling), with the truly conclusive paper still only occupying a single cell in the matrix. In conclusion, the following recommendations for all scientists working in the field are proposed.

7.1. Checklist for authors

Before the acquisition of plant metabolite profiles, the analyst should collect previous information of the plant in question and plan the investigation accordingly.

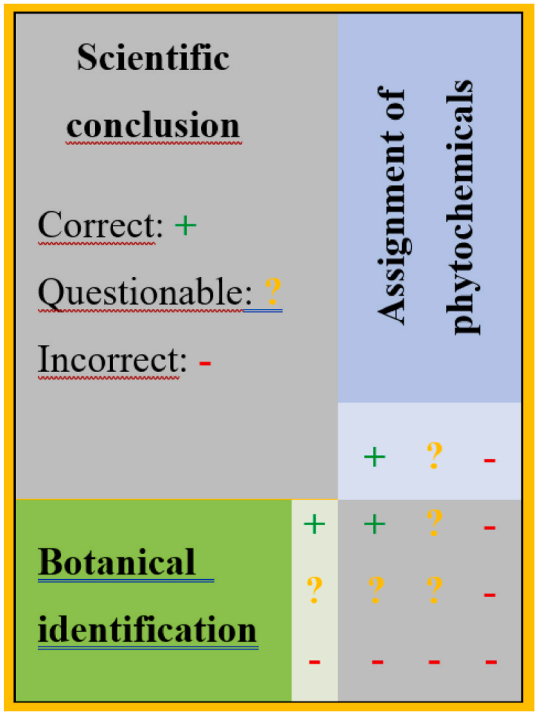


Fig. 10. Phytochemical data quality matrix illustrating that all aspects of identification must be conclusive in order to reach an overall conclusive link between compound and source species.

Plant collection should be reported in enough detail, to be reproducible, and species identification should be transparent (including a reference to a standard flora), and include a voucher. If at all possible, generally available material should be investigated in parallel and authentic standards or reference materials be acquired.

Attention should be paid to detection limits and quantification.

When evaluating data, identification levels suitable for phytochemistry should be considered. The so-called “level 1” in systems aimed at less metabolically complex organisms like mammals cannot be extrapolated to phytochemistry.

In studies including large scale preparative work, presence or absence of the isolated metabolites should be tested using gentle, small scale analytical procedures. Even in analytical studies, artefacts due to instability and incomplete inactivation of plant enzymes should be considered.

During spectral interpretation, common pitfalls in MS and NMR spectroscopy should be considered also using prediction tools for checking/verifying the elucidated/proposed structures.

When discussing and reporting data, the entire body of scientific information on the plant taxon investigated should be taken into account. Potential involvement of endophytic microorganisms should be kept in mind.

Phylogenies based on metabolite patterns should not be suggested anymore. In contrast, discussions on the evolution of metabolite patterns in the light of phylogenies created from independent data (DNA sequence data) are timely and warranted.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Serhat S. Çiçek: Writing – review & editing, Writing – original draft. **Alfonso Mangoni:** Writing – review & editing, Writing – original draft. **Franziska S. Hanschen:** Writing – review & editing, Writing – original draft. **Niels Agerbirk:** Writing – review & editing, Writing – original draft. **Christian Zidorn:** Writing – review & editing, Writing – original draft, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

Aalizadeh, R., Nikolopoulou, V., Thomaidis, N.S., 2022. Development of liquid chromatographic retention index based on cocamide diethanolamine homologous series (C(n)-DEA). Anal. Chem. 94, 15987–15996. https://doi.org/10.1021/acs.analchem.2c02893.

Aberham, A., Çiçek, S.S., Schneider, P., Stuppner, H., 2010. Analysis of sesquiterpene lactones, lignans, and flavonoids in wormwood (*Artemisia absinthium* L.) using high-performance liquid chromatography (HPLC)-mass spectrometry, reversed phase HPLC, and HPLC-solid phase extraction-nuclear magnetic resonance. J. Agric. Food Chem. 58, 10817–10823. https://doi.org/10.1021/jf1022059.

Adams, R.P., 2007. Identification of Essential Oil Components by Gas Chromatography/mass Spectrometry, fourth ed. Allured Publishing Corporation, Carol Stream, IL, USA.

Afendi, F.M., Okada, T., Yamazaki, M., Hirai-Morita, A., Nakamura, Y., Nakamura, K., Ikeda, S., Takahashi, H., Altaf-Ul-Amin, M., Darusman, L.K., Saito, K., Kanaya, S., 2012. KNApSAcK family databases: integrated metabolite-plant species databases for multifaceted plant research. Plant Cell Physiol. 53 (e1), 1–12. https://doi.org/10.1093/pcp/pcr165.

Agerbirk, N., Olsen, C.E., 2012. Glucosinolate structures in evolution. Phytochemistry 77, 16–45. https://doi.org/10.1016/j.phytochem.2012.02.005.

Agerbirk, N., Olsen, C.E., Cipollini, D., Ørgaard, M., Linde-Laursen, I., Chew, F.C., 2014. Specific glucosinolate analysis reveals variable levels of glucobarbarins, dietary precursors of 5-phenyloxazolidine-2-thiones, in watercress types with contrasting chromosome number. J. Agric. Food Chem. 62, 9586–9596. https://doi.org/10.1021/jf5032795.

Agerbirk, N., Olsen, C.E., Heimes, C., Christensen, S., Bak, S., Hauser, T., 2015. Multiple hydroxyphenethyl glucosinolate isomers and their tandem mass spectrometric distinction in a geographically structured polymorphism in the crucifer *Barbarea vulgaris*. Phytochemistry 115, 130–142. https://doi.org/10.1016/j.phytochem.2014.09.003.

Agerbirk, N., Hansen, C.C., Kiefer, C., Hauser, T.P., Ørgaard, M., Lange, C.B.A., Cipollini, D., Koch, M.A., 2021. Comparison of glucosinolate diversity in the crucifer tribe Cardamineae and the remaining order Brassicales highlights repetitive evolutionary loss and gain of biosynthetic steps. Phytochemistry 185, 112668. https://doi.org/10.1016/j.phytochem.2021.112668.

Agerbirk, N., Pattison, D.I., Mandáková, T., Lysak, M., Montaut, S., 2022. Ancient biosynthesis in an oil crop: glucosinolate profiles in *Limnanthes alba* and its relatives (Limnanthaceae, Brassicales). J. Agric. Food Chem. 70, 1134–1147. https://doi.org/10.1021/acs.jafc.1c07299.

Aguiar, J.M.R.B.V., Ferreira, G. de S., Sanches, P.A., Bento, J.M.S., Sazima, M., 2021. What pollinators see does not match what they smell: absence of color-fragrance association in the deceptive orchid *Ionopsis utricularioides*. Phytochemistry 182, 112591. https://doi.org/10.1016/j.phytochem.2020.112591.

Alcalde-Eon, C., García-Estévez, I., Rivas-Gonzalo, J.C., de la Cruz, D.R., Escribano-Bailón, M.T., 2016. Anthocyanins of the anthers as chemotaxonomic markers in the genus *Populus* L. Differentiation between *Populus nigra, Populus alba* and *Populus tremula*. Phytochemistry 128, 35–49. https://doi.org/10.1016/j.phytochem.2016.04.004.

Alseekh, S., de Souza, L.P., Benina, M., Fernie, A.R., 2020. The style and substance of plant flavonoid decoration: towards defining both structure and function. Phytochemistry 174, 112347. https://doi.org/10.1016/j.phytochem.2020.112347.

Alseekh, S., Apharoni, A., Brotman, Y., et al., 2021. Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. Nat. Methods 18, 747–756. https://doi.org/10.1038/s41592-021-01197-1.

Alygizakis, N., Lestremau, F., Gago-Ferrero, P., Gil-Solsona, R., Arturi, K., Hollender, J., Schymanski, E.L., Dulio, V., Slobodnik, J., Thomaidis, N.S., 2023. Towards a harmonized identification scoring system in LC-HRMS/MS based non-target screening (NTS) of emerging contaminants. Trends Anal. Chem. 159, 116944 https://doi.org/10.1016/j.trac.2023.116944.

Amirzakariya, B.Z., Shakeri, A., 2022. Bioactive terpenoids derived from plant endophytic fungi: an updated review (2011-2020). Phytochemistry 197, 113130. https://doi.org/10.1016/j.phytochem.2022.113130.

Andini, S., Dekker, P., Gruppen, H., Araya-Cloutier, C., Vincken, J.-P., 2019. Modulation of glucosinolate composition in Brassicaceae seeds by germination and fungal elicitation. J. Agric. Food Chem. 67, 12770–12779. https://doi.org/10.1021/acs.jafc.9b05771.

Aune, D., Giovannucci, E., Boffetta, P., Fadnes, L.T., Keum, N., Norat, T., Greenwood, D. C., Riboli, E., Vatten, L.J., Tonstad, S., 2017. Fruit and vegetable intake and the risk of cardiovascular disease, total cancer and all-cause mortality–a systematic review and dose-response meta-analysis of prospective studies. Int. J. Epidemiol. 1029–1056. https://doi.org/10.1093/ije/dyw319.

Bednarek, P., Osbourn, A., 2009. Plant-microbe interactions: chemical diversity in plant defense. Science 324, 746–748. https://doi.org/10.1126/science.1171661.

Binev, Y., Marques, M.M.B., Aires-de-Sousa, J., 2007. Prediction of ^{1}H NMR coupling constants with associative neural networks trained for chemical shifts. J. Chem. Inf. Model. 47, 2089–2097. https://doi.org/10.1021/ci700172n.

Blažević, I., Montaut, S., Burčul, F., Olsen, C.E., Burow, M., Rollin, P., Agerbirk, N., 2020. Glucosinolate structural diversity, identification, chemical synthesis and metabolism in plants. Phytochemistry 169, 112100. https://doi.org/10.1016/j.phytochem.2019.112100.

Braunberger, C., Zehl, M., Conrad, J., Wawrosch, C., Strohbach, J., Beifuss, U., Krenn, L., 2015. Flavonoids as chemotaxonomic markers in the genus *Drosera*. Phytochemistry 118, 74–82. https://doi.org/10.1016/j.phytochem.2015.08.017.

Brkljaca, R., Urban, S., 2015. HPLC-NMR and HPLC-MS investigation of antimicrobial constituents in *Cystophora monilifera* and *Cystophora subfarcinata*. Phytochemistry 117, 200–208. https://doi.org/10.1016/j.phytochem.2015.06.014.

Cárdenas, P.D., Landtved, J.P., Larsen, S.H., Lindegaard, N., Wøhlk, S., Jensen, K.R., Pattison, D.I., Burow, M., Bak, S., Crocoll, C., Agerbirk, N., 2023. Phytoalexins of the crucifer *Barbarea vulgaris*: structural profile and correlation with glucosinolate turnover. Phytochemistry 213, 113745. https://doi.org/10.1016/j.phytochem.2023.113742.

Çiçek, S.S., Galarza Pérez, M., Wenzel-Storjohann, A., Bezerra, R.M., Segovia, J.F.O., Girreser, U., Kanzaki, I., Tasdemir, D., 2022. Antimicrobial prenylated isoflavones from the leaves of the Amazonian medicinal lant *Vatairea guianensis* Aubl. J. Nat. Prod. 85, 927–935. https://doi.org/10.1021/acs.jnatprod.1c01035.

Çiçek, S.S., Pfeifer Barbosa, A.L., Girreser, U., 2018. Quantification of diterpene acids in *Copaiba* oleoresin by UHPLC-ELSD and heteronuclear two-dimensional qNMR. J. Pharm. Biomed. Anal. 160, 126–134. https://doi.org/10.1016/j.jpba.2018.07.034.

Çiçek, S.S., Pfeifer Barbosa, A.L., Wenzel-Storjohann, A., Segovia, J.F.O., Bezerra, R.M., Sönnichsen, F., Zidorn, C., Kanzaki, I., Tasdemir, D., 2023. Chemical and biological evaluation of Amazonian medicinal plant *Vouacapoua americana*. Aubl. Plants 12, 99. https://doi.org/10.3390/plants12010099.

Çiçek, S.S., Ugolini, T., Girreser, U., 2019. Two-dimensional qNMR of anthraquinones in *Frangula alnus* (*Rhamnus frangula*) using surrogate standards and delay time adaption. Anal. Chim. Acta 1081, 131–137. https://doi.org/10.1016/j.aca.2019.06.046.

Çiçek, S.S., Wenzel-Storjohann, A., Girreser, U., Tasdemir, D., 2020. Biological activities of two major copaiba diterpenoids and their semi-synthetic derivatives. Rev. Bras. Farmacogn. 30, 18–27. https://doi.org/10.1007/s43450-020-00002-y.

Clancy, M.V., Mamin, M., Flückiger, G., Quijano-Medina, T., Pérez-Niño, B., Abdala-Roberts, L., Turlings, T.C.J., Bustos-Segura, C., 2023. Terpene chemotypes in *Gossypium hirsutum*/wild cotton) from the Yucatan Peninsula, Mexico. Phytochemistry 205, 113454. https://doi.org/10.1016/j.phytochem.2022.113454.

Cullen, J., 1968. *Anthyllis* L. In: Tutin, T.G., Heywood, V.H., Burges, N.A., Moore, D.M., Valentine, D.H., Walters, S.M., Webb, D.A. (Eds.), Flora Europaea, vol. II. Cambridge University Press, Cambridge, pp. 177–182, 1968.

Czerniawski, P., Piasecka, A., Bednarek, P., 2021. Evolutionary changes in the glucosinolate biosynthetic capacity in species representing *Capsella, Camelina* and *Neslia* genera. Phytochemistry 181, 112571. https://doi.org/10.1016/j.phytochem.2020.112571.

Dai, D.-C., Xu, X.-F., Yan, H., Zhang, Y., 2022. Phenylpropanoid derivatives from *Ficus esquiroliana* and their chemotaxonomic significance. Biochem. Systemat. Ecol. 105, 104509 https://doi.org/10.1016/j.bse.2022.104509.

Dauncey, E., Irving, J., Allkin, R., Robinson, N., 2016. Common mistakes when using plant names and how to avoid them. Eur. J. Integr. Med. 8, 597–601. https://doi.org/10.1016/j.eujim.2016.09.005.

David, B., 2018. New regulations for accessing plant biodiversity samples, what is ABS? Phytochemistry Rev. 17, 1211–1223. https://doi.org/10.1007/s11101-018-9573-1.

Da Silva, J.J.M., Crevelin, E.J., Carneiro, L.J., Rogez, H., Veneziani, R.C.S., Ambrosio, S. R., Moraes, L.A.B., Bastos, J.K., 2017. Development of a validated ultra-high-performance liquid chromatography tandem mass spectrometry method for determination of acid diterpenes in *Copaifera* oleoresins. J. Chromatogr. A 1515, 81–90. https://doi.org/10.1016/j.chroma.2017.07.038.

de Graaf, R.M., Krosse, S., Swolfs, A.E.M., te Brinke, E., Prill, N., Leimu, R., van Galen, P. M., Wang, Y., Aarts, M.G.M., van Dam, N.M., 2015. Isolation and identification of 4-α-rhamnosyloxy benzyl glucosinolate in *Noccaea caerulescens* showing intraspecific

variation. Phytochemistry 110, 166–171. https://doi.org/10.1016/j.phytochem.2014.11.016.
Dictionary of Natural Products, 2023. CHEMnetBASE – Dictionary of Natural Products. https://dnp.chemnetbase.com/chemical/ChemicalSearch.xhtml?dswid=-2195. last accessed 10.03.2023.
Di Lecce, R., Galarza Pérez, M., Mangoni, A., Çiçek, S.S., Sönnichsen, F.D., Evidente, A., Zidorn, C., 2022. Sesquiterpenoids from *Leontodon tenuiflorus* (Asteraceae, cichorieae): first record of a hypocretenoid from *Leontodon* section Asterothrix. Biochem. Systemat. Ecol. 102, 104408 https://doi.org/10.1016/j.bse.2022.104408.
Dührkop, K., Fleischauer, M., Ludwig, M., Aksenov, A.A., Melnik, A.V., Meusel, M., Dorrestein, P.C., Rousu, J., Böcker, S., 2019. Sirius 4: a rapid tool for turning tandem mass spectra into metabolite structure information. Nat. Methods 16, 299–302. https://doi.org/10.1038/s41592-019-0344-8.
Ehrlich, P.R., Raven, P., 1964. Butterflies and plants: a study in coevolution. Evolution 18, 586–608, 2307/2406212.
EMA, 2018. ICH Q2(R2) Validation of Analytical Procedures - Scientific Guideline. European Medicines Agency. URL. https://www.ema.europa.eu/en/ich-q2r2-validation-analytical-procedures-scientific-guideline. (Accessed 10 March 2023). last accessed.
Erb, M., Kliebenstein, D.J., 2020. Plant secondary metabolites as defenses, regulators and primary metabolites: the blurred functional trichotomy. Plant Physiol. 184, 39–52. https://doi.org/10.1104/pp.20.00433.
Fechner, J., Kaufmann, M., Herz, C., Eisenschmidt, D., Lamy, E., Kroh, L.W., Hanschen, F. S., 2018. The major glucosinolate hydrolysis product in rocket (*Eruca sativa* L.), sativin, is 1,3-thiazepane-2-thione: elucidation of structure, bioactivity, and stability compared to other rocket isothiocyanates. Food Chem. 261, 57–65. https://doi.org/10.1016/j.foodchem.2018.04.023.
Fischer, M.A., 2015. Was ist ein Taxonym? Neilreichia 7, 195–229. URL: https://www.zobodat.at/publikation_volumes.php?id=46314.
Forman, V., Luo, D., Geu-Flores, F., Lemcke, R., Nelson, D.R., Kampranis, S.C., Staerk, D., Møller, B.L., Pateraki, I., 2022. A gene cluster in *Ginkgo biloba* encodes unique multifunctional cytochrome P450s that initiate ginkgolide biosynthesis. Nat. Commun. 13, 5143. https://doi.org/10.1038/s41467-022-32879-9.
Formato, M., Scharenberg, F., Pacifico, S., Zidorn, C., 2022. Seasonal variations in phenolic natural products in *Fagus sylvatica* (European beech) leaves. Phytochemistry, 113385. https://doi.org/10.1016/j.phytochem.2022.113385, 2003.
Frederiksen, S., Rasmussen, F.N., Seberg, O., 2006. Dansk Flora. Gyldendal, Københavnp. 701.
Friedrich, K., Wermter, N.S., Andernach, L., Witzel, K., Hanschen, F.S., 2022. Formation of sulfur volatile compounds and *S*-methyl-L-cysteine sulfoxide in *Brassica oleracea* vegetables. Food Chem. 383, 132544 https://doi.org/10.1016/j.foodchem.2022.132544.
Furrer, J., 2012. A comprehensive discussion of HMBC pulse sequences, part 1: the classical HMBC. Concepts Magn. Reson. 40A, 101–127. https://doi.org/10.1002/cmr.a.21232.
Furrer, J., 2014. A comprehensive discussion of HMBC pulse sequences. III. Solving the problem of missing and weakly observed long-range correlations. Concepts Magn. Reson. 43, 177–206. https://doi.org/10.1002/cmr.a.21317.
Gabrielson, S.W., 2018. SciFinder. J. Med. Libr. Assoc. 106, 588–590. https://doi.org/10.5195/jmla.2018.515.
Gao, H., Li, G., Lou, H.-X., 2018. Structural diversity and biological activities of novel secondary metabolites from endophytes. Molecules 23, 646. https://doi.org/10.3390/molecules23030646.
Grass, S., Zidorn, C., Ellmerer, E.P., Stuppner, H., 2004. Eudesmane derivatives from *Hieracium intybaceum*. Chem. Biodivers. 1, 353–360. https://doi.org/10.1002/cbdv.200490031.
Grauso, L., de Falco, B., Bochicchio, R., Scarpato, S., Addesso, R., Lanzotti, V., 2023. Leaf metabolomics and molecular networking of wild type and mutant genotypes of chia (*Salvia hispanica* L.). Phytochemistry 209, 113611. https://doi.org/10.1016/j.phytochem.2023.113611.
Griffin, W.J., Lin, G.D., 2000. Chemotaxonomy and geographical distribution of tropane alkaloids. Phytochemistry 53, 623–637. https://doi.org/10.1016/S0031-9422(99)00475-6.
Hall, L.M., Hall, L.H., Kertesz, T.M., Hill, D.W., Sharp, T.R., Oblak, E.Z., Dong, Y.W., Wishart, D.S., Chen, M.-H., Grant, D.F., 2012. Development of Ecom50 and retention index models for nontargeted metabolomics: identification of 1,3-dicyclohexylurea in human serum by HPLC/mass spectrometry. J. Chem. Inf. Model. 52, 1222–1237. https://doi.org/10.1021/ci300092s.
Hansen, K., 2004. Dansk Feltflora. 11. Oplag. Gyldendal, Copenhagen, p. 757.
Hansen, N.L., Kjaerulff, L., Heck, Q.K., Forman, V., Staerk, D., Møller, B.M., Andersen-Ranberg, J., 2022. *Tripterygium wilfordii* cytochrome P450s catalyze the methyl shift and epoxidations in the biosynthesis of triptonide. Nat. Commun. 13, 5011. https://doi.org/10.1038/s41467-022-32667-5.
Hill, C.R., Shafaei, A., Balmer, L., Lewis, J.R., Hodgson, J.M., Millar, A.H., Blekkenhorst, L.C., 2023. Sulfur compounds: from plants to humans and their role in chronic disease prevention. Crit. Rev. Food Sci. Nutr. 63, 8616–8638. https://doi.org/10.1080/10408398.2022.2057915.
Hoffmann, H., Andernach, L., Kanzler, C., Hanschen, F.S., 2022. Novel transformation products from glucosinolate-derived thioglucose and isothiocyanates formed during cooking. Food Res. Int. 157, 111237 https://doi.org/10.1016/j.foodres.2022.111237.
Hoffmann, M.A., Kretschmer, F., Ludwig, M., Böcker, S., 2023. MAD HATTER correctly annotates 98% of small molecule tandem mass spectra searching in PubChem. Metabolites 13, 314. https://doi.org/10.3390/metabo13030314.
Holscher, D., Schneider, B., 1999. HPLC-NMR analysis of phenylphenalenones and a stilbene from *Anigozanthos flavidus*. Phytochemistry 50, 155–161. https://doi.org/10.1016/S0031-9422(98)00495-6.
Hyland, K.C., Butt, C.M., Quinn-Paquet, A., 2019. Nontargeted Screening with SCIEX OS Software and ChemSpider. https://sciex.com/content/dam/SCIEX/tech-notes/technology/ruo-mkt-02-7777-a/Nontargeted-Screening_SCIEX-OS-ChemSpider_RUO-MKT-02-7777-A.pdf. (Accessed 25 October 2023). last accessed.
JordiLabs, 2023. NIST/EPA/NIH Mass Spectral Library Compound Scoring: Match Factor, Reverse Match Factor, and Probability. https://jordilabs.com/wp-content/uploads/2017/07/Whitepaper-NIST-EPA-NIH-Mass-Spectral-Library-Compound-Scoring.pdf. (Accessed 25 October 2023) (last accessed.
Kariya, K., Fujita, A., Ueno, M., Yoshikawa, T., Teraishi, M., Taniguchi, Y., Ueno, K., Ishihara, A., 2023. Natural variation of diterpenoid phytoalexins in rice: aromatic diterpenoid phytoalexins in specific cultivars. Phytochemistry 211, 112708. https://doi.org/10.1016/j.phytochem.2023.113708.
Kind, T., Wohlgemuth, G., Lee, D.Y., Lu, Y., Palazoglu, M., Shahbaz, S., Fiehn, O., 2009. FiehnLib: mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. Anal. Chem. 81, 10038–10048. https://doi.org/10.1021/ac9019522.
Kováts, E., 1958. Gas-chromatographische Charakterisierung organischer Verbindungen. Teil 1: Retentionsindices aliphatischer Halogenide, Alkohole, Aldehyde und Ketone. Helv. Chim. Acta 41, 1915–1932. https://doi.org/10.1002/hlca.19580410703.
Krishnan, S.R., Skiba, A., Luca, S.V., Marcourt, L., Wolfender, J.-L., Skalicka-Wozniak, K., Gertsch, J., 2023. Bioactivity-guided isolation of trypanocidal coumarins and dihydro-pyranocrhomones from selected Apiaceae plant species. Phytochemistry 213, 113770. https://doi.org/10.1016/j.phytochem.2023.113770.
Kuhn, S., Schlörer, N.E., 2015. Facilitating quality control for spectra assignments of small organic molecules: nmrshiftdb2 – a free in-house NMR database with integrated LIMS for academic service laboratories. Magn. Reson. Chem. 53, 582–589. https://doi.org/10.1002/mrc.4263.
Lee, M.L., Vassilaros, D.L., White, C.M., 1979. Retention indices for programmed-temperature capillary-column gas chromatography of polycyclic aromatic hydrocarbons. Anal. Chem. 51, 768–773. https://doi.org/10.1021/ac50042a043.
Lemli, J., Cuvelee, J., 1978. Les transformations des hétérosides anthroniques pendant le séchage des feuilles de *Cassia senna* et de *Rhamnus frangula*. Planta Med. 34, 311–318. https://doi.org/10.1055/s-0028-1097457.
Li, T., Kongstad, K.T., Staerk, D., 2019. Identification of α-glucosidase inhibitors in *Machilus litseifolia* by combined use of high-resolution α-glucosidase inhibition profiling and HPLC-PDA-HRMS-SPE-NMR. J. Nat. Prod. 82, 249–258. https://doi.org/10.1021/acs.jnatprod.8b00609, 2019.
Li, Y., Zidorn, C., 2022. Seasonal variation of natural products in European herbs. Phytochemistry Rev. 21, 1549–1575. https://doi.org/10.1007/s11101-021-09797-7.
Liang, C., Kjaerulff, L., Hansen, P.R., Kongstad, K.T., Staerk, D., 2021. Dual high-resolution α-glucosidase and PTP1B inhibition profiling combined with HPLC-PDA-HRMS-SPE-NMR analysis for the indentification of potentially antidiabetic chromene meroterpenoids from *Rhododendron capitatum*. J. Nat. Prod. 84, 2454–2467. https://doi.org/10.1021/acs.jnatprod.1c00454.
Liang, X., Christensen, J.H., Bucheli, T.D., Nielsen, N.J., 2023. Source-supported suspect screening (4S) of phytotoxins in terrestrial and aquatic environments: a field study of *Lupinus angustifolius* L. (blue lupin). Environ. Sci. Technol. 57, 2333–2340. https://doi.org/10.1021/acs.est.2c05387.
Liu, X., Matsumoto, H., Lv, T., Zhan, C., Fang, H., Pan, Q., Xu, H., Fan, X., Chu, T., Chen, S., Qiao, K., Ma, Y., Sun, L., Wang, Q., Wang, M., 2023. Phyllosphere microbiome induces host metabolic defence against rice false-smut disease. Nat. Microbiol. 8, 1419–1433. https://doi.org/10.1038/s41564-023-01379-x.
Liu, J.-J., Hao, J.-J., Tan, M., Liao, C.-C., Liu, D., Li, H.-M., Li, R.-T., 2024. Iridoids and other constituents from the leaves and stems of *Valeriana officinalis* var. *latifolia*. Phytochemistry 218, 113934. https://doi.org/10.1016/j.phytochem.2023.113934.
Lorensen, M.D.B.B., Bjarnholt, N., St-Pierre, B., Heinicke, S., Courdavault, V., O'Connor, S., Janfelt, C., 2023. Spatial localization of monoterpenoid indole alkaloids in *Rauvolfia tetraphylla* by high resolution mass spectrometry imaging. Phytochemistry 209, 113620. https://doi.org/10.1016/j.phytochem.2023.113620.
Ludwig-Müller, J., 2015. Plant and endophytes: equal partners in secondary metabolite production? Biotechnol. Lett. 37, 1325–1334. https://doi.org/10.1007/s10529-015-1814-4.
Maffei, M., 1996. Chemotaxonomic significance of leaf wax alkanes in the Gramineae. Biochem. Systemat. Ecol. 24, 53–64. https://doi.org/10.1016/0305-1978(95)00102-6.
Mancinotti, D., Frick, K.M., Geu-Flores, F., 2022. Biosynthesis of quinolizidine alkaloids in lupins: mechanistic considerations and prospects for pathway elucidation. Nat. Prod. Rep. 39, 1423. https://doi.org/10.1039/d1np00069a.
Mangoni, A., 2012. Strategies for structural assignment of marine natural products through advanced NMR-based techniques. In: Fattorusso, E., Gerwick, W.H., Taglialatela-Scafati, O. (Eds.), Handbook of Marine Natural Products. Springer, pp. 518–546. https://doi.org/10.1007/978-90-481-3834-0_8.
Missinou, A.A., de Carvalho, J.F., Marnet, N., Delhave, T., Hamzaoui, O., Sayed, D.A., Guitton, Y., Lebreton, L., Langrume, C., Laperche, A., Delourme, R., Manzanares-Dauleux, M.J., Bouchereau, A., Gravot, A., 2022. Identification and quantification of

glucosinolates and phenolics in a large panel of *Brassica napus* highlight valuable genetic resources for chemical ecology and breeding. J. Agric. Food Chem. 70, 5245–5261. https://doi.org/10.1021/acs.jafc.1c08118.
Mohimani, H., Gurevich, A., Shlemov, A., Mikheenko, A., Korobeynikov, A., Cao, L., Shcherbin, E., Nothias, L.-F., Dorrestein, P.C., Pevzner, P.A., 2018. Dereplication of microbial metabolites through database search of mass spectra. Nat. Commun. 9, 4035. https://doi.org/10.1038/s41467-018-06082-8.
Morant, A.V., Jørgensen, K., Jørgensen, C., Paquette, S.M., Sánchez-Perez, R., Møller, B. L., Bak, S., 2008. β-Glucosidases as detonators of plant chemical defense. Phytochemistry 69, 1795–1813. https://doi.org/10.1016/j.phytochem.2008.03.006.
Mori, S., Hasegawa, Y., Moriguchi, Y., 2023. Color strategies of camellias recruiting different pollinators. Phytochemistry 207, 113559. https://doi.org/10.1016/j.phytochem.2022.113559.
Nanusha, M.Y., Krauss, M., Sørensen, B.G., Schulze, T., Strobel, B.W., Brack, W., 2021. Occurrence of plant secondary metabolite fingerprints in river waters from eastern Jutland, Denmark. Environ. Sci. Eur. 33, 25. https://doi.org/10.1186/s12302-021-00464-4.
Negin, B., Jander, G., 2023. Convergent and divergent evolution of plant chemical defenses. Curr. Opin. Plant Biol. 73, 102368 https://doi.org/10.1016/j.pbi.2023.102368.
Neilson, E.H., Goodger, J.Q.D., Woodrow, I.E., Møller, B.L., 2018. Plant chemical defense: at what cost? Trends Plant Sci. 18, 250–258. https://doi.org/10.1016/j.tplants.2013.01.001.
Nemesio-Gorriz, M., Menezes, R.C., Paetz, C., Hammerbacher, A., Steenackers, M., Schamp, K., Höfte, M., Svatos, A., Gershenzon, J., Douglas, G.C., 2020. Candidate metabolites for ash dieback tolerance in *Fraxinus excelsior*. J. Exp. Bot. 71, 6074–6083. https://doi.org/10.1093/jxb/eraa306.
NIST, 2023. NIST/EPA/NIH Mass Spectral Library 2023 Release. Wiley Science Solutions. URL: https://chemdata.nist.gov/dokuwiki/lib/exe/fetch.php?media=chemdata:asms2023:asms2023_nist23_features.pdf. (Accessed 25 October 2023) (last accessed.
Nomura, T., 2017. Function and application of a non-ester-hydrolyzing carboxylesterase discovered in tulip. Biosci. Biotechnol. Biochem. 81, 81–94. https://doi.org/10.1080/09168451.2016.1240608.
Nørbæk, R., Brandt, K., Nielsen, J.K., Ørgaard, M., Jacobsen, N., 2002. Flower pigment composition of *Crocus* species and cultivars used for a chemotaxonomic investigation. Biochem. Systemat. Ecol. 30, 763–791. https://doi.org/10.1016/S0305-1978(02)00020-0.
Oberacher, H., Sasse, M., Antignac, J.-P., Guitton, Y., Debrauwer, L., Jamin, E.L., Schulze, T., Krauss, M., Covaci, A., Caballero-Casero, N., Rousseau, K., Damont, A., Fenaille, F., Lamoree, M., Schymanski, E.L., 2020. A European proposal for quality control and quality assurance of tandem mass spectral libraries. Environ. Sci. Eur. 32, 43. https://doi.org/10.1186/s12302-020-00314-9.
Okamura, Y., Dort, H., Reichelt, M., Tunström, K., Wheat, C.W., Vogel, H., 2023. Testing hypotheses of a coevolutionary key innovation reveals a complex suite of traits involved in defusing the mustard oil bomb. Proc. Nat. Acad. Sci. USA 119 (51), e2208447119. https://doi.org/10.1073/pnas.2208447119.
One Thousand Plant Trancriptomes Initiative, 2019. One thousand plant transcriptomes and the phylogenomics of green plants. Nature 574, 679–685. https://doi.org/10.1038/s41586-019-1693-2.
Paguet, A.S., Siah, A., Lefevre, G., Moureu, S., Cadalen, T., Samaillie, J., Michels, F., Deracinois, B., Flahaut, C., Santos, H.A.D., Etienne-Debaecker, A., Rambaud, C., Chollet, S., Molinié, R., Fontaine, J.-X., Waterlot, C., Fauconnier, M.-L., Sahpaz, S., Rivere, C., 2023. Multivariate analysis of chemical and genetic diversity of wild *Humulus lupulus* L. (hop) collected in situ in northern France. Phytochemistry 205, 113508. https://doi.org/10.1016/j.phytochem.2022.113508.
Pang, Z., Chen, J., Wang, T., Gao, C., Li, Z., Guo, L., Xu, J., Cheng, Y., 2021. Linking plant secondary metabolites and plant microbiomes: a review. Front. Plant Sci. 12, 621276 https://doi.org/10.3389/fpls.2021.621276.
Paulsen, J., Yahyazadeh, M., Hänsel, S., Kleinwächter, M., Ibrom, K., Selmar, D., 2015. 13,14-Dihydrocoptisine – the genuine alkaloid from *Chelidonium majus*. Phytochemistry 111, 149–153. https://doi.org/10.1016/j.phytochem.2015.01.006.
Pedras, M.S.C., To, Q.H., 2016. Unveiling the first indole-fused thiazepine: structure, synthesis and biosynthesis of cyclonasturlexin, a remarkable cruciferous phytoalexin. Chem. Commun. 52, 5880–5883. https://doi.org/10.1039/c6cc02108e.
Pedras, M.S.C., Thapa, C., 2020. Unveiling fungal detoxification pathways of the cruciferous phytoalexin rapalexin A: sequential L-cysteine conjugation, acetylation and oxidative cyclization mediated by Colletotrichum spp. Phytochemistry 169, 112188. https://doi.org/10.1016/j.phytochem.2019.112188.
Petersen, L., Minkkinen, P., Esbensen, K.H., 2005. Representative sampling for reliable data analysis: theory of Sampling. Chemometr. Intell. Lab. Syst. 77, 261–277. https://doi.org/10.1016/j.chemolab.2004.09.013.
Pichersky, E., Raguso, R.A., 2018. Why do plants produce so many terpenoid compounds? New Phytol. 220, 692–702. https://doi.org/10.1111/nph.14178.
Pignatti, S., 2017. *Anthyllis* L in: Flora D'Italia, 2. 2. Aufl. – Edagricole, Milano, pp. 614–620.
Plaszkó, T., Szucs, Z., Vasas, G., Gonda, S., 2022. Interactions of fungi with non-isothiocyanate products of the plant glucosinolate pathway: a review on product formation, antifungal activity, mode of action and biotransformation. Phytochemistry 200, 113245. https://doi.org/10.1016/j.phytochem.2022.113245.
Poveda, J., Velasco, P., de Haro, A., Johansen, T.J., McAlvay, A.C., Möllers, C., Mølmann, J.A.B., Ordiales, E., Rodríguez, V.M., 2021. Agronomic and metabolomic side-effects of a divergent selection for indol-3-ylmethylglucosinolate content in kale (*Brassica oleracea* var. *acephala*). Metabolites 11, 384. https://doi.org/10.3390/metabo11060384.
Quilliam, M.A., 2017. Retention Index Standards for Liquid Chromatography. U.S. Patent. US2017/0102367A, 2017.
Ranner, J.L., Schalk, S., Martyniak, C., Parniske, M., Gutjahr, C., Stark, T.D., Dawid, C., 2023. Primary and secondary metabolites in *Lotus japonicus*. J. Agric. Food Chem. 71, 11277–11303. https://doi.org/10.1021/acs.jafc.3c02709.
Rasmussen, L.F., Anton, J., Kjaerulff, L., Zhao, Y., Semple, S.J., Chi, N., Buirchell, B., Møller, B.L., Staerk, D., 2022. Serrulatane diterpenoids with unusual side chain modifications from root bark of *Eremophila longifolia*. Phytochemistry 203, 113408. https://doi.org/10.1016/j.phytochem.2022.113408.
Reaxys. 2023 https://www.reaxys.com/#/search/quick (last accessed 10 Mar 23).
Renz, M., Andernach, L., Kaufmann, M., Rohn, S., Hanschen, F.S., 2024. Degradation of glucosinolates and formation of isothiocyanates, nitriles, amines, and N,N'-dialk(en)yl thioureas during domestic boiling of red cabbage. Food Chem. 435, 137550 https://doi.org/10.1016/j.foodchem.2023.137550.
Richard, T., Temsamani, H., Cantos-Villar, E., Monti, J.-P., 2013. Application of LC-MS and LC-NMR techniques for secondary metabolite identification. Adv. Bot. Res. 67, 67–98. https://doi.org/10.1016/B978-0-12-397922-3.00002-2.
Robien, W., 2019. Computer-assisted peer reviewing of spectral data: the CSEARCH protocol. Monatsh. Chem. 150, 927–932. https://doi.org/10.1007/s00706-019-02407-5.
Rogers, A., Dietz, K.-J., Gifford, M.L., Lunn, J.E., 2021. The importance of independent replication of treatments in plant science. J. Exp. Bot. 72, 5270–5274. https://doi.org/10.1093/jxb/erab268.
Santiago, K.B., Conti, B.J., Murbach Teles Andrade, B.F., Mangabeira da Silva, J.J., Rogez, H.L.G., Crevelin, E.J., Beraldo de Moraes, L.A., Veneziani, R., Ambrósio, S.R., Bastos, J.K., Sforcin, J.M., 2015. Immunomodulatory action of *Copaifera* spp oleoresins on cytokine production by human monocytes. Biomed. Pharmacother. 70, 12–18. https://doi.org/10.1016/j.biopha.2014.12.035.
Saurí, J., Martin, G.E., Furrer, J., 2015. A comprehensive discussion of HMBC pulse sequences: 4. Establishing two-bond correlations from HMBC and related experiments. Concepts Magn. Reson. 44A, 227–251. https://doi.org/10.1002/cmr.a.21362.
Schrimpe-Rutledge, A.C., Codreanu, S.G., Sherrod, S.D., McLean, J.A., 2016. Untargeted metabolomics strategies—challenges and emerging directions. J. Am. Soc. Mass Spectrom. 27, 1897–1905. https://doi.org/10.1007/s13361-016-1469-y.
Schymanski, E.L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H.P., Hollender, J., 2014. Identifying small molecules via high resolution mass spectrometry: communicating confidence. Environ. Sci. Techol. 48, 2097–2098. https://doi.org/10.1021/es5002105.
Seaman, F., Bohlmann, F., Zdero, C., Mabry, T.J., 1990. Diterpene analysis with emphasis on clerodanes. In: Seaman, F., Bohlmann, F., Zdero, C., Mabry, T.J. (Eds.), Diterpenes of Flowering Plants: Compositae (Asteraceae). Springer, New York, NY, pp. 493–550. https://doi.org/10.1007/978-1-4612-3274-2_6.
Shen, S.-M., Appendino, G., Guo, Y.-W., 2022. Pitfalls in the structural elucidation of small molecules. A critical analysis of a decade of structural misassignments of marine natural products. Nat. Prod. Rep. 39, 1803–1832. https://doi.org/10.1039/D2NP00023G.
Shroff, R., Schramm, K., Jeschke, V., Nemes, P., Vertes, A., Gershenzon, J., Svatoš, A., 2015. Quantification of plant surface metabolites by matrix-assisted laser desorption-ionization mass spectrometry imaging: glucosinolates on *Arabidopsis thaliana* leaves. Plant J. 81, 6. https://doi.org/10.1111/tpj.12760.
Sorokina, M., Merseburger, P., Rajan, K., Yirik, M.A., Steinbeck, C., 2021. COCONUT online: collection of open natural products database. J. Cheminf. 13, 2. https://doi.org/10.1186/s13321-020-00478-9.
Sorokina, M., Steinbeck, C., 2020. Review on natural products databases: where to find data in 2020. J. Cheminf. 12, 20. https://doi.org/10.1186/s13321-020-00424-9.
Staerk, D., Kesting, J.R., Sairafianpour, M., Witt, M., Asili, J., Emami, S.A., Jaroszewski, J.W., 2009. Accelerated dereplication of crude extracts using HPLC–PDA–MS–SPE–NMR: quinolinone alkaloids of *Haplophyllum acutifolium*. Phytochemistry 70, 1055–1061. https://doi.org/10.1016/j.phytochem.2009.05.004.
Su, S., Yang, X.Q., Yang, Y.-B., Ding, Z.-T., 2023. Antifungal and antifeedant terpenoids from *Paraphaeosphaeria* sp. cultured by extract of host *Ginkgo biloba*. Phytochemistry 210, 113651. https://doi.org/10.1016/j.phytochem.2023.113651.
Sun, R., Hong, B., Reichelt, M., Luck, K., Mai, D.T., Jiang, X., Gershenzon, J., Vassao, D. G., 2023. Metabolism of plant-derived toxins from its insect host increases the success of the entomopathogenic fungus *Beauveria bassiana*. ISME J. 17, 1693–1704. https://doi.org/10.1038/s41396-023-01480-3.
Sumner, L.W., Amberg, A., Barrett, D., Beale, M.H., Beger, R., Daykin, C.A., Fan, T.W.-M., Fiehn, O., Goodacre, R., Griffin, J.L., Hankemeier, T., Hardy, N., Harnly, J., Higashi, R., Kopka, J., Lane, A.N., Lindon, J.C., Marriott, P., Nicholls, A.W., Reily, M. D., Thaden, J.J., Viant, M.R., 2007. Proposed minimum reporting standards for chemical analysis chemical analysis working group (CAWG) metabolomics standards initiative (MSI). Metabolomics 3, 211–221. https://doi.org/10.1007/s11306-007-0082-2.
Tan, R.X., Zou, W.X., 2001. Endophytes: a rich source of functional metabolites. Nat. Prod. Rep. 18, 448–459, 10.1039/b100918o.

Tang, Y., Li, N., Ai, Y., Ma, G.-L., Osman, E.E.A., Xiong, J., Li, J., Jin, Z.-X., Hu, J.-F., 2019. LC-MS guided isolation and dereplication of *Lycopodium* alkaloids from *Lycopodium cernuum* var. *sikkimense* of different geographical origins. Phytochemistry 160, 25–30. https://doi.org/10.1016/j.phytochem.2019.01.006.
Tang, G.-Y., Meng, X., Li, Y., Zhao, C.-N., Liu, Q., Li, H.-B., 2017. Effects of vegetables on cardiovascular diseases and related mechanisms. Nutrients 9, 857. https://doi.org/10.3390/nu9080857.
Tatsuzawa, F., 2019. Acylated pelargonidin glycosides from the red-purple flowers of *Iberis umbellata* L. and the red flowers of *Erysimum* × *cheiri* (L.) Crantz (Brassicaceae). Phytochemistry 159, 109–118. https://doi.org/10.1016/j.phytochem.2018.12.010.
Toghueo, R.M.K., Sahal, D., Boyom, F.F., 2020. Recent advances in inducing endophytic fungal specialized metabolites using small molecule elicitors including epigenetic modifiers. Phytochemistry 174, 112338. https://doi.org/10.1016/j.phytochem.2020.112338.
Torres-Mendoza, D., Ureña González, L.D., Ortega-Barría, E., Coley, P.D., Kursar, T.A., Capson, T.L., McPhail, K., Cubilla-Rios, L., 2004. Novel cassane and cleistanthane diterpenes from *Myrospermum frutescens*: absolute stereochemistry of the cassane diterpene series. J. Nat. Prod. 67, 1711–1715. https://doi.org/10.1021/np049890c.
Trabelcy, B., Chinkov, N., Samuni-Blank, M., Merav, M., Izhaki, I., Carmeli, S., Gerchman, Y., 2021. Investigation of glucosinolates in the desert plant *Ochradenus baccatus* (Brassicales: resedaceae). Unveiling glucoochradenin, a new arabinosylated glucosinolate. Phytochemistry 187, 112760. https://doi.org/10.1016/j.phytochem.2021.112760.
Vandenkoornhuyse, P., Quaiser, A., Duhamel, M., Van, A.L., Dufresne, A., 2015. The importance of the microbiome of the plant holobiont. New Phytol. 206, 1196–1206. https://doi.org/10.1111/nph.13312.
Vincken, J.P., Heng, L., de Groot, A., Gruppen, H., 2007. Saponins, classification and occurrence in the plant kingdom. Phytochemistry 68, 275–297. https://doi.org/10.1016/j.phytochem.2006.10.008.
Wang, M., Cernava, T., 2023. Soterobionts: disease-preventing microorganisms and proposed strategies to facilitate their discovery. Curr. Opin. Microbiol. 75, 102349 https://doi.org/10.1016/j.mib.2023.102349.
Wang, C., Li, Z., Zhang, L., Gao, Y., Cai, X., Wu, W., 2022. Identifying key metabolites associated with glucosinolate biosynthesis in response to nitrogen management strategies in two rapeseed (*Brassica napus*) varieties. J. Agric. Food Chem. 70, 634–645. https://doi.org/10.1021/acs.jafc.1c06472.
Waridel, P., Wolfender, J.L., Lachavanne, J.B., Hostettmann, K., 2004. Identification of the polar constituents of *Potamogeton* species by HPLC-UV with post-column derivatization, HPLC-MS^n and HPLC-NMR, and isolation of a new *ent*-labdane diglycoside. Phytochemistry 65, 2401–2410. https://doi.org/10.1016/j.phytochem.2004.06.031.
Wink, M., 2003. Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. Phytochemistry 64, 3–19. https://doi.org/10.1016/S0031-9422(03)00300-5.
Wink, M., 2022. Current understanding of modes of action of multicomponent bioactive phytochemicals: potential for nutraceuticals and antimicrobials. Annu. Rev. Food Sci. Technol. 13, 337–359. https://doi.org/10.1146/annurev-food-052720-100326.
Wink, M., Botschen, F., Gosmann, C., Schäfer, H., Watermann, P.G., 2010. Chemotaxonomy seen from a phylogenetic perspective and evolution of secondary metabolism. In: Wink, M. (Ed.), Annual Plant Reviews 40, second ed. Wiley-Blackwell, Chichester, UK, pp. 364–433. https://doi.org/10.1002/9781444320503.ch7.
Windsor, A.J., Reichelt, M., Figuth, A., Svatos, A., Kroymann, J., Kliebenstein, D.J., Gershenzon, J., Mitchell-Olds, T., 2005. Geographic and evolutionary diversification of glucosinolates among near relatives of *Arabidopsis thaliana* (Brassicaceae). Phytochemistry 66, 1321–1333. https://doi.org/10.1016/j.phytochem.2005.04.016.
Wittstock, U., Gershenzon, J., 2002. Constitutive plant toxins and their role in defense against herbivores and pathogens. Curr. Opin. Plant Biol. 5, 300–307. https://doi.org/10.1016/s1369-5266(02)00264-9.
Xu, Y., Xu, Y., Huang, Z., Luo, Y., Gao, R., Xue, J., Lin, C., Pawlowski, K., Zhou, Z., Wei, X., 2023. 3-Pentanol glycosides from root nodules of the actinorhizal plant *Alnus cremastogyne*. Phytochemistry 207, 113582. https://doi.org/10.1016/j.phytochem.2022.113582.
Yulvianti, M., Zidorn, C., 2021. Chemical diversity of plant cyanogenic glycosides: an overview about reported natural products. Molecules 26, 719. https://doi.org/10.3390/molecules26030719.
Zhao, Y., Hansen, N.L., Duan, Y.-T., Prasad, M., Motawia, M.S., Møller, B.L., Staerk, D., Bak, S., Miettinen, K., Kampranis, S.C., 2023. Biosynthesis and biotechnological production of the anti-obesity agent celastrol. Nat. Chem. 15, 1236–1246. https://doi.org/10.1038/s41557-023-01245-7.
Zheng, S.-J., Liu, S.-J., Zhu, Q.-F., Guo, N., Wang, Y.-L., Yuan, B.-F., Feng, Y.-Q., 2018. Establishment of liquid chromatography retention index based on chemical labeling for metabolomic analysis. Anal. Chem. 90, 8412–8420. https://doi.org/10.1021/acs.analchem.8b00901.
Zhou, S., Jander, G., 2021. Engineering insect resistance using plant specialized metabolites. Curr. Opin. Biotechnol. 70, 115–121. https://doi.org/10.1016/j.copbio.2021.03.005.
Zidorn, C., 2008. Sesquiterpene lactones and their precursors as chemosystematic markers in the tribe Cichorieae of the Asteraceae. Phytochemistry 69, 2270–2296. https://doi.org/10.1016/j.phytochem.2008.06.013.
Zidorn, C., 2017. Guidelines for consistent characterisation and documentation of plant source materials for studies in phytochemistry and phytopharmacology. Phytochemistry 139, 56–59. https://doi.org/10.1016/j.phytochem.2017.04.004.
Zidorn, C., 2018. Seasonal variation of natural products in European trees. Phytochemistry Rev. 17, 923–935. https://doi.org/10.1007/s11101-018-9570-4.
Zidorn, C., 2019. Plant chemophenetics – a new term for plant chemosystematics/plant chemotaxonomy in the macro-molecular era. Phytochemistry 163, 147–148. https://doi.org/10.1016/j.phytochem.2019.02.013.
Zidorn, C., Gottschlich, G., Stuppner, H., 2002. Chemosystematic investigations on phenolics from flowerheads of Central European taxa of *Hieracium* (Asteraceae). Plant Systemat. Evol. 231, 39–58. https://doi.org/10.1007/s006060200010.
Zidorn, C., Stuppner, H., 2001a. Chemosystematics of taxa from the *leontodon* section oporinia. Biochem. Systemat. Ecol. 29, 827–837. https://doi.org/10.1016/S0305-1978(01)00019-9.
Zidorn, C., Stuppner, H., 2001b. Evaluation of chemosystematic characters in the genus *Leontodon*. Taxon 50, 115–133. https://doi.org/10.2307/1224515.
Züst, T., Heichinger, C., Grossniklaus, U., Harrington, R., Kliebenstein, D.J., Turnbull, L. A., 2012. Natural enemies drive geographic variation in plant defenses. Science 338, 116–119. https://doi.org/10.1126/science.1226397.

Serhat Sezai Çiçek studied Pharmacy at the University of Innsbruck/Austria from 2002 to 2007. From 2007 to 2011 he pursued his doctoral studies in Pharmacognosy in the group of Professor Hermann Stuppner and in parallel absolved the Certificate of Advanced Studies in "Ethnobotany and Ethnomedicine" at the University of Zurich, Switzerland. He became a registered Pharmacist in 2011 and worked in the Hospital Pharmacy Salzburg/Austria from 2011 to 2016, where he was heading the Quality Control department for the last two years. In 2016 he became a postdoctoral researcher at the Pharmaceutical Biology of the Institute of Pharmacy at Kiel University/Germany, where he founded the Junior Research Group for Analytical Natural Product Chemistry in 2021. In September 2023, he obtained a professorship for Instrumental Bioanalysis at Hamburg University of Applied Sciences. His research interests are the identification, characterisation, and quantitation of natural bio-actives with modern analytical techniques. These techniques comprise chromatographic and spectroscopic methods, such as UHPLC-DAD-ELSD-MS/MS, GC-MS, and quantitative NMR spectroscopy. Further research fields are the identification of novel lead structures against multi-resistant pathogens and haematologic cancers as well as ethnopharmacognosy and food chemistry. Serhat Sezai Çiçek has published more than 50 papers in national and international journals and discovered several new natural products of different compound classes, such as diterpenes, triterpenes, isoflavones, and phenolic acids.

Alfonso Mangoni has been professor of Organic Chemistry since 2002 at the University of Naples Federico II, Department of Pharmacy. He was visiting professor at Columbia University (New York, NY, USA) in 1995 and at UCSD (San Diego, CA, USA) in 2006. The main object of his research on marine natural products is the chemistry of marine macro-organisms (with emphasis on Porifera) and microorganisms, including isolation, structure elucidation and bioactivity. Starting from this topic, his research developed in several directions: development and use of new methods for dereplication of natural extract (molecular networking), development and use of spectroscopic and computational methods for structural elucidation of organic compounds, biosynthetic studies on natural products from microbial symbionts of marine organisms, and total synthesis of analogues of marine natural products. This research led to the publication of 137 papers on peer-reviewed international journals and 3 book chapters.

Franziska S. Hanschen obtained her PhD in Food Chemistry at the Technische Universität Berlin in 2012 and received her habilitation degree in Food Chemistry from the Faculty of Mathematics, Informatics and Natural Sciences at the University of Hamburg in 2020. After a Postdoc at Leibniz Institute of Vegetable and Ornamental Crops (IGZ) in Großbeeren, in 2018 she became leader of the Leibniz Junior Research Group Optimization of Glucosinolate Degradation Pathways and in 2021 her research group was tenured at IGZ. Her research interests focus on the mechanistic understanding of the formation, role and fate of secondary plant metabolites in plants, foods and humans as well as on identification and characterization of natural compounds. Franziska Hanschen has published 60 peer-reviewed papers, obtained more than 2.8 Mio. € of third-party funding and discovered several new natural organosulfur compounds. In 2020 she was awarded the Werner-Baltes-Preis of the young scientist by the Society of Food Chemistry, a section of the German Chemical Society.

Niels Agerbirk (1967) is an Associate Professor at the University of Copenhagen and teaches biochemistry and chemistry. He graduated in Biology from Aarhus University (1994) and received his Ph.D. (1997) in Biochemistry from the Royal Veterinary and Agricultural University, Denmark. Doctoral and postdoctoral training (1994–2000) in natural product chemistry, enzymology and chemical ecology was in the laboratories of H. Sørensen and J.K. Nielsen (Copenhagen), S. Palmieri (Bologna) and J.A.A. Renwick (Ithaca), with a year in industry (yeast group, B. Rønnow, Danisco Biotechnology). His research is mainly on glucosinolates and related metabolites, with a broad perspective ranging from chemistry and enzymology to ecology and evolution.

Christian Zidorn studied Pharmacy at the University of Düsseldorf, Germany; he was registered as a pharmacist in 1995. From 1995 to 1998, Christian pursued doctoral studies in Pharmacognosy at the University of Innsbruck, Austria. After receiving his Dr. rer. nat. degree in 1998, he pursued postdoctoral studies in the Plant Extracts Research Unit of Crop and Food in Dunedin, New Zealand from 2000 to 2001 in the group of Dr. Nigel B. Perry. Subsequently, Christian completed habilitations (state doctorates) in Pharmacognosy (2003) and Botany (2007). From 2010 to 2013, Dr. Zidorn worked in the group of Dr. Anna Lönnroth in Brussels, Belgium as a scientific officer in the field of infectious diseases with the European Commission; here he was responsible for FP7 collaborative projects in the fields of antimicrobial resistance and emerging epidemics. After again working as an associate professor at the University of Innsbruck and a step-in professorship at the University of Kiel, Germany, he studied marine natural compounds as a Marie Curie IEF grant holder at the CNR, Istituto di Chimica Biomolecolare in Pozzuoli, Italy in the group of Dr. Margherita Gavagnin. Since April 2016, Christian Zidorn is director and full professor at the Department of Pharmaceutical Biology of the Institute of Pharmacy at the Kiel University in the far North of Germany. Currently, he heads the Section of Pharmacy at Kiel University and is ERASMUS coordinator for a total of ten universities in Italy, Lithuania, Norway, and Poland. Christian Zidorn's research interests include phytochemistry, chemophenetics, botanical systematics, and chemical ecology. Current projects are focused on seasonal variation of plant natural products, chemophenetics of the Cichorieae (Asteraceae), bio-actives from vegetables, and on marine natural products from seagrasses, their bioactivities, and ecological functions. Professor Zidorn has published more than 120 original papers in international journals, more than 20 international reviews, and 25 papers in journals of national interest. His research resulted in the discovery of dozens of new natural compounds from various compound classes such as sesquiterpenoids, diterpenes, phenolic acids, iridoids, lignans, diarylheptanoids, and polyacetylenes. Christian Zidorn is one of the editors in chief of Biochemical Systematics and Ecology and one of the Associate Editors of Fitoterapia.