Natural Product Reports

ROYAL SOCIETY OF CHEMISTRY

REVIEW

View Article Online
View Journal

Cite this: DOI: 10.1039/d5np00040h

Cultivar to chemotype: characterizing complex botanicals with mass spectrometry metabolomics

Joshua J. Kellogg, *[ab] R. Teal Jordan, [a] Madhusha M. Ranaweera, [b] Kelsey Custer, [a] Savannah G. Anez, [b] Julia Bendlin, [a] Francisco T. Chacon[b] and Xiaoling Chen [a]

Received 18th May 2025

DOI: 10.1039/d5np00040h

rsc.li/npr

Covering up to 2025

Plant products, including botanical dietary supplements, nutraceuticals, and herbal medicines, remain central to supporting human health and wellness. Their usage has been steadily increasing over the last few decades, which has also led to raised concerns about proper identification and characterization of plant materials. This information is crucial to evaluate the safety and efficacy of these botanical products and prevent misidentification or adulteration. While there are multiple analytical approaches to characterize botanicals, this review provides insight into how untargeted mass spectrometry metabolomics can profile these commonly complex mixtures and provide detailed datasets that are capable of taxonomically classifying samples, detecting adulteration, and providing insight into variation between plant materials and their nutritional, medicinal, or toxicological effects. We describe data analysis approaches for untargeted metabolomics, case studies on the various applications of this method for characterizing botanicals, and challenges that the growing field of mass spectrometry-based metabolomics is facing. The chosen topics reflect the current state of metabolomics analyses for complex systems with a look to the future of how to conceptualize botanical characterization.

1. Introduction

1.1. Importance of the characterization of botanicals

With a long therapeutic history, plants and other natural products remain a central element in promoting human health and preventing disease. Their relevance and use as dietary supplements (*i.e.*, "botanical dietary supplements") have expanded in the last several decades in the United States since the passage of the Dietary Supplement Health and Education Act. The number of Americans using dietary supplements has increased steadily; over 50% of the population reports using dietary supplements and up to 40% of those were reported to consume botanical supplements.[1–4] Between 2000 and 2022, the

[a]Department of Veterinary and Biomedical Sciences, Pennsylvania State University, University Park, PA 16802, USA. E-mail: jjk6146@psu.edu
[b]Intercollege Graduate Degree Program in Plant Biology, Pennsylvania State University, University Park, PA 16802, USA

sale of herbal supplements in the United States has increased from $4.25 to $12.12 billion.[5] This drive for botanical dietary supplements is rooted in attitudes toward 'natural' products as well as a growing body of evidence that is supportive of their positive effects on human health.[6]

The impact of botanicals on human health is perhaps not surprising; plants have diverse biosynthetic capabilities and produce a range of chemical structures that reaches complexities beyond many synthetic compound libraries.[7] In addition, their secondary metabolites are evolutionarily designed to interact with biological receptors and systems,[7] which has been reflected in their foundational role in human health and pharmaceutical development. Currently, there are *ca.* 150 active clinical trials on the US National Library of Medicine's clinical trial tracker (https://clinicaltrials.gov, accessed October 20, 2024) that incorporate "herbal" or "botanical" as part of their study focus,[8] and natural products have higher rates of clinical trial success.[9] Of the 1394 small molecule drugs approved by the United States Food and Drug Administration (FDA) between 1981 and 2019, 53.1% were from natural products or designed around natural product pharmacophores.[10]

Botanical dietary supplements are characteristically complex phytochemical mixtures, and this chemical composition can vary depending on abiotic and biotic factors during growth and processing, as well as the biosynthetic variation between species, genera, and families. Thus, the veracity of research, and the impact for consumers, is predicated on the authenticity of the botanical(s) under consideration and characterizing their chemical constituents; there is considerable evidence to suggest that misidentification and adulteration (either intentional or accidental) is rampant in the botanical dietary supplement arena. Two global assessments of 5957 commercial herbal products across 37 countries and

Joshua J. Kellogg

Dr Joshua J. Kellogg is an Assistant Professor of Metabolomics in the Department of Veterinary and Biomedical Sciences and the Huck Institutes of Life Sciences, Pennsylvania State University. Dr Kellogg received his B.S. in chemistry from the University of California Berkeley and PhD in natural product chemistry and ethnobotanical nutraceuticals from North Carolina State University. His lab at Penn State focuses on the characterization of biologically active molecules from plant and fungal sources, including new leads against chronic and infectious diseases as well as focusing on metabolomics approaches to characterize herbal preparations and enhance modeling of complex systems.

R. Teal Jordan

Teal Jordan works as the manager for the Kellogg Lab at The Pennsylvania State University where she conducts research and trains student scientists in the exploration of bioactive compounds and metabolomics in plants and fungi. She holds a B.S. in environmental science from the University of Georgia and an MS in forest resources from Penn State where her research examined the phytochemistry in ramps/wild leeks (Allium tricoccum Ait.) in Pennsylvania. She has industry experience in herbal product manufacturing and has studied medicinal plants for over a decade.

Madhusha M. Ranaweera

Madhusha Ranaweera is a PhD student in the Kellogg Lab at Penn State, originally from Sri Lanka. She received her B.S. (Honors) in Botany from the University of Peradeniya in 2021. Her current research is focused on investigating the bioactive metabolites of hemp (Cannabis sativa) and their therapeutic potentials for inflammatory bowel disease (IBD). She is excited to continue her passion for natural product discovery from plants to bridge the gap between traditional plant-based therapeutics and modern medicine, hoping to share her expertise gained at Penn State with her home country, Sri Lanka.

Kelsey Custer

Kelsey Custer is a recent undergraduate at The Pennsylvania State University, earning a B.S. in Pharmacology and Toxicology in spring 2025. She is currently a first-year graduate student at the University of North Carolina at Greensboro, pursuing a PhD in Chemistry and Biochemistry. She hopes to continue her studies in botanical natural products and drug discovery research.

six continents reported that 27% of tested products were determined to be adulterated in some way.[11,12] A study by Navarro *et al.* (2019) found that, of 272 products tested, 51% were mislabeled, where the labeled constituents did not match the mass spectrometry analysis.[13] During the Covid-19 pandemic, an analysis of elderberry (*Sambucus nigra* L. and *S. canadensis* L.) products revealed that 58 products out of 532 analyzed were adulterated, mostly with black rice.[14] Thus, the mis-identification, adulteration, or other alteration of botanical products is well-documented, and there is substantial risk for reduced efficacy and safety, increased toxicity and unforeseen adverse interactions if the identity of the botanical agent and its chemical composition are unknown or mischaracterized. Complicating this is the fact that botanical products can be obtained from multiple sources, potentially with multiple producers of the raw material and varied processing techniques. Therefore, careful identification and characterization of botanicals are crucial to ensuring that the products on the shelves are efficacious and safe.

1.2. Conventional authentication methods

To maintain safe and efficacious botanical products, multiple analytical approaches continue to serve as primary means of characterizing plant formulations; however, they possess distinct limitations in their accuracy, sensitivity, and/or prospects for development; this highlights a noticeable analytical gap in botanical profiling.

1.2.1. Classical taxonomic approaches: macroscopic and microscopic morphology. Botanical identification by morphology (macroscopic or microscopic characteristics) is one of the most accurate methods in confirming botanical identity and has a large role in phytochemical research.[15] Taxonomic treatments guided by species-specific combinations of morphological characteristics are used in botanical quality control and scientific investigations extensively.[16,17] As most taxonomic treatments are mostly based on the morphological characters of the reproductive organs (flowers, fruits), and the above-ground vegetative parts, identification of commercial

Francisco T. Chacon

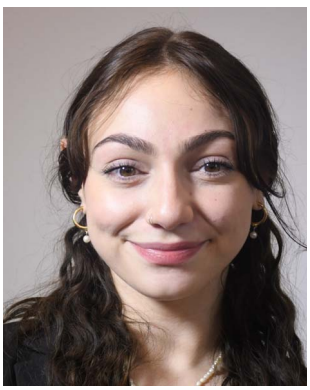
Francisco Chacon, PhD, earned his B.S. from New Mexico State University in Genetics and Biotechnology. He completed his PhD in Plant Biology from The Pennsylvania State University in October 2024, studying the synthesis and synergy of secondary metabolites from Cannabis sativa L. under Prof. Joshua Kellogg.

Xiaoling Chen

Xiaoling Chen is a PhD graduate student researcher in the Kellogg Lab at The Pennsylvania State University. She received her B.S. in Microbiology and BBA in Management from UMass Amherst and has conducted research on rare disease small molecule drug delivery and CRISPR-Cas9 gene therapy prior to starting her graduate studies. She is currently working on identifying novel AHR ligands derived from natural products, with a particular interest in Eastern medicinal plants that may help regulate abdominal pain and gut homeostasis.

Savannah G. Anez

Savannah Anez received her B.S. in Biochemistry from Notre Dame, and is currently a 4th-year PhD Candidate in the Plant Biology program at The Pennsylvania State University. She is using an ethno-directed approach to study the phytochemistry and therapeutic potential of the medicinal plant "ghost pipe" (Monotropa uniflora). After graduation, she hopes to continue her dissertation work while teaching and mentoring at a primarily undergraduate institution (PUI).



Julia Bendlin

Julia is a postgraduate student from Connecticut and graduated from The Pennsylvania State University in May 2024 with a Bachelor degree in Pharmacology and Toxicology as well as a minor in Biology. She is pursuing a PharmD at Northeastern University, with plans to explore the field of industrial pharmacy while maintaining a focus in natural products research and its intersection with modern therapeutics. She looks forward to connecting her passion for pharmacology, patient safety, and scientific discovery, and is excited to build on the experiences she gained at Penn State's Kellogg Lab.

botanical products through morphology can be challenging.[16,18] Although microscopic and organoleptic characteristics may still provide accurate means of identification in some cases where the botanicals are available fresh or dried,[16,18] their identification becomes impossible when the ingredients lose their diagnostic features during product development; for example, when they are substantially processed (*e.g.*, ground, extracted, or compounded).[19–23] The use of microscopic techniques such as scanning electron microscopy (SEM) and transmission electron microscopy (TEM) for anatomical studies leads to additional challenges due to the expensive equipment, extensive processing, and proper user experience required to minimize misleading artifacts.[24,25] While trained specialists have traditionally been required to provide accurate morphological identifications (a limitation of the method), technological advancements, including machine learning and artificial intelligence approaches, could bolster the use of morphological identification for plants,[26,27] though questions of interpretable accuracy remain.[28–30] Furthermore, the literature records for species level identification are sparse and sometimes contradictory,[31,32] and associated language-specific names can be varied as well.[33,34]

1.2.2. Genomic approaches. DNA barcoding is an approach applied in the identification and quality control of herbal products, which enables species identification using short standard DNA sequences (*e.g.*, DNA barcodes).[35–37] DNA barcoding coupled with high throughput sequencing (HTS), known as DNA metabarcoding, allows simultaneous high throughput multi-taxa identification from complex samples with DNA of different origins.[38,39] However, both the DNA barcoding and metabarcoding have limitations in producing positive authentication of plant ingredients from any amplifiable DNA and false negatives with degraded or lost DNA during processing or manufacturing, and their applicability is narrowed to only taxonomic authentication; barcoding is unable to provide any quantitative or qualitative information regarding the active metabolites in the plant samples in the context of quality control of herbal products.[35,40] Real-time quantitative polymerase chain reaction (qPCR) technology eliminates the need of sequencing by enabling real time detection of specific target sequences using gene specific primers and fluorescence.[41,42] The barcoding-high-resolution melting analysis (Bar-HRM) has emerged as a simple, highly specific, cost-effective, high throughput, and sensitive technique, having the potential of detecting and discriminating among closely related species in herbal products without the need of post PCR analysis with mini barcodes (<200 bp). However, its application to multicomponent samples might produce unreliable results.[19,43] Genomic approaches can extend to situations where the adulterant or mixture contains unknown species *via* the use of non-specific primers;[44,45] however, this is not a universal solution, as the genetic divergence of certain materials (*e.g.*, wild potatoes, *Euphrasia*) may not be large enough for species level separation.[46] In addition, genomic identification isn't applicable if the products are processed samples, which do not feature genomic material.[13,47]

1.2.3. Targeted chemical approaches. The use of targeted chemical analysis to identify and characterize botanical products can be extremely useful, as often there are distinct molecules present that are well known in the literature to represent the species or genus or there are specific monographs detailing their characteristic chemical profile (*e.g.*, the German Commission E,[48] US Pharmacopeia,[49] Tyler's Herbs of Choice,[50] and the American Herbal Pharmacopoeia[51]). Furthermore, previously identified marker compounds can be detected by a variety of chemical techniques, including charged aerosol detection (CAD) or ultraviolet-visible light spectroscopy (UV-vis), often coupled to separation methods like high performance liquid chromatography (HPLC), capillary electrophoresis (CE), or gas chromatography (GC). However, axiomatic to employing a set of biomarkers to characterize a botanical product is (a) *a priori* knowledge of the chemistry relevant to the plant and (b) availability (either commercially or *via* previous isolation efforts) of the requisite compounds in sufficient quantity and quality for the unambiguous characterization. This represents a major drawback to relying on defined subsets of the plant's chemistry, as not all botanicals have been exhaustively evaluated to have detailed literature or monographs that support targeted molecular approaches.

Additionally, targeted analyses rely on an oversimplification of the broader chemical landscape present in botanicals, and this leaves the analysis susceptible to duplicitous manipulation of the product, especially single-molecule analyses. Ginkgo (*Ginkgo biloba* L.) extracts were found to be adulterated with flavonoid-rich chemical mixtures to bypass authenticity markers that were built upon "flavone glycosides" as a broad category; 35% of those sampled were found to be a completely different species (*Styphnolobium japonicum* (L.) Schott).[52,53] For the elderberry case mentioned above, the samples had been spiked with black rice extract, which contains the elderberry marker cyanidin-3-*O*-glucoside.[14] Other cases have been well-documented where botanicals, dietary supplements, and nutraceuticals were spiked to bypass authentication or quality control efforts.[54] In one, weight loss dietary supplements were adulterated with various androgenic steroids, alkaloid derivatives, or even *Ephedra sinica* Stapf extracts.[55,56] A second weight loss supplement case found chemical analogs of the banned 1,3-dimethylamylamine (1,3-DMAA) present in commercial products.[57]

These inadequacies can be partially overcome by incorporating multiple chemical biomarkers into a single method, known as molecular "fingerprints." This has been employed to authenticate various botanical species, including *Coptis* species,[58] *Tinospora* species,[59] and even the Association of Official Agricultural Chemists' (AOAC) official method for authenticating ashwagandha (*Withania somnifera* (L.) Dunal), which utilizes 10 separate withanolide glycosides and aglycones.[60] However, multi-compound fingerprints are more labor- and time-intensive to establish and validate, and there are practical challenges in developing a method that can accommodate the potentially disparate physiochemical characteristics of a series of analytes.[61] This also does not address the issue of using predefined compounds to characterize a botanical product, as reliable sources of the analytes must be available for method development and further implementation.

1.3. Mass spectrometry/metabolomics methods of characterization

The metabolome is generally defined as the complete set of small molecules (<1200 Da) produced or metabolized by an organism/biological sample at a given point in time. This includes primary and secondary metabolites, the latter of which are designed for interaction with the external environment and are the best candidates for bioactivity.[7,62] Metabolomics is the universal, unbiased measurement of the metabolome (with the assertion that no single analytical platform can capture the entirety of small molecules in one experiment),[63] and using the relative intensity of the metabolite signals provides a broad dataset for comparing two samples chemically. This has been used for comparisons between samples' conditions (species, environment, geography, processing, storage, adulteration, *etc.*).[64–66] Crucially, untargeted metabolomics can be applied with no *a priori* knowledge of the chemistry of the system, representing a powerful agnostic tool for investigations into botanical products, dietary supplements, and nutraceuticals.[59,67]

While a variety of analytical techniques can be used to collect metabolome data – including Fourier-transformed infrared spectroscopy (FTIR), charged aerosol detection (CAD), and ultraviolet-visible (UV/VIS) spectrophotometry – the two primary approaches that have emerged for metabolomics studies are nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS). Mass spectrometry is a highly sensitive, highly accurate, high-resolution, robust approach to deciphering the chemical complexity of botanical products and nutraceuticals,[68] and has become a preeminent analytical method for collecting metabolomics data. The purpose of this review is not to discuss advancements in instrumentation related to metabolomics; that topic has been covered elsewhere in great detail.[69,70] Instead, this review seeks to summarize and highlight the analytical and data science applications of metabolomics methods for botanical characterization, with a focus on novel approaches developed in the last decade. This review will also feature current obstacles to advancing the science of botanical characterization and future directions for innovation.

2. Chemometric data analysis approaches for untargeted metabolomics

One of the greatest challenges with untargeted metabolomics data is analyzing the raw spectral dataset. Metabolomics data matrices often have more columns (independent variables, *e.g.*, *m*/*z*–retention time signals) than rows (samples) and are referred to as "landscape" matrices. There are two main approaches to analyzing metabolomics data: chemometrics and quantitative analysis. Chemometrics refers to the application of statistical methods to discover significant trends and patterns and maximize the information obtained from the chemical datasets, while quantitative analysis is traditionally employed when there is prior identification of relevant metabolites, which facilitates direct analysis of the subset of chemicals. Compared to quantitative analysis, chemometrics can be performed on unannotated data representing the entirety of the measurable metabolome, and is the primary statistical modeling performed on botanical metabolomic data (Fig. 1). To model and extract relevant information from these dense data matrices, chemometric pattern recognition algorithms are typically employed (those interested in the theory and derivations of chemometrics can find thorough discussions for these analyses elsewhere.[71,72])

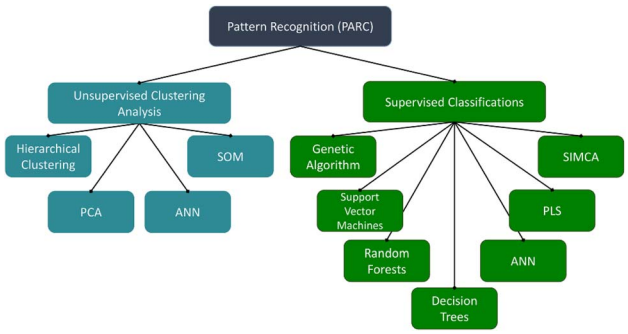


Fig. 1 Different unsupervised and supervised chemometric approaches for characterizing botanical materials. Reproduced with permission from Abraham and Kellogg.[71] Copyright 2021, Frontiers Media.

Regardless of the analysis, mass spectrometry data must be processed from their raw spectral format into a data matrix suitable for analysis; this is a multi-step process by which discreet features are obtained (unique *m*/*z*–retention time pairs) for input into the multivariate chemometric methods. While there are semi-automated software packages that function to process the data,[73–75] the multitudinous parameters needed to understand and optimize to produce a final dataset can be challenging for researchers. Furthermore, the centering, scaling, and normalization of data is crucial to control for heteroscedasticity and non-normal signal response but can also play a role in the shape of the final data matrix.[76,77]

2.1. Unsupervised linear multivariate statistics

Unsupervised methods are the foremost means for multivariate analysis of untargeted metabolomics data. These approaches are considered "unsupervised" as there are no associated data classifications or metadata that feed into the analysis; sample relationships are discerned from pattern recognition methods relying only on the chemical dataset. The most common unsupervised chemometric approach is Principal Component Analysis (PCA), a linear regression method in which the metabolomic data is projected into a smaller dimensional space comprised of orthogonal principal components that allows for characterization of the overall variation in the data. Frequently, the first two components are plotted in a pair-wise fashion (*e.g.*, "scores plot"), which allows for a spatial analysis of the overall chemical similarity of the objects/samples without any guiding principles and provides a corresponding look at potential contributions of variables to the PCA model (*e.g.*, "loadings plot").[78] This relatively simple multivariate analysis often represents a first step in analyzing botanical metabolomics data; however, the model does not inherently provide

quantitative measures of similarity or dissimilarity, and the choice of components to plot is often an *ad hoc* decision, which could complicate unbiased analysis.[79]

Beyond PCA, other unsupervised methods follow similar dimensional reduction approaches. Hierarchical Cluster Analysis (HCA) employs distance calculations between samples using and amalgamating the results to cluster based upon overall similarity. The clusters can be evaluated based upon different criteria (*e.g.*, how many groupings are desired) for describing sample similarity. Self-organizing maps (SOMs) are neural network-based algorithms that reduce the dimensionality to yield patterns of samples that are represented in a 2-dimensional map, with similar samples being mapped closer together.[80] All of these multivariate methods benefit from relative computational simplicity and an agnostic approach to classifying samples without any *a priori* decisions or metadata describing connections between the samples, and thus are powerful approaches for metabolomic analyses of botanical phytochemistry.

2.2. Supervised linear multivariate statistics

Supervised multivariate analyses have a different purpose compared to unsupervised approaches; supervised methods are used in identifying relevant biomarkers, classification of sample categories, and unknown prediction, which are beyond the scope of unsupervised methods. Supervised multivariate methods are generally linear regression models built to accommodate both independent and dependent variables to model correlative changes based on the independent variables. Indeed, a number of machine learning models are built upon supervised multivariate methods.[81] However, due to the imbalance of independent variables to samples in most metabolomic studies, supervised techniques are prone to overfitting the data;[82] even to the point of 'fitting' a model to completely random data.[83] Thus, steps must be taken to ensure that the model is robust and able to fit and predict the data without overfitting. One primary approach for supervised linear modeling is partial least squares (PLS), a dimension reduction method similar to PCA in which the model condenses the complex independent (*e.g.* chemical) data into a smaller set of latent variables but employs a dependent variable to supervise the overall construction of the model. The dependent variables can be nominal (*e.g.*, pre-defined classes or categories, known as PLS-Discriminant Analysis (PLS-DA))[84] or numerical in nature (*e.g.*, concentration, yield, or activity, known as PLS-Regression (PLS-R)).[85] PLS results in similar plots as PCA but can also be used to extend the analysis and highlight potential differentiating metabolites from the dataset.[86] Numerous variations of PLS have been developed,[87–89] and PLS is a principal linear regression algorithm for botanical characterization.[90–93] Soft Independent Modeling of Class Analogies (SIMCA) is a supervised extension of PCA; samples are grouped into *a priori* defined classes and then a PCA analysis is performed on each class. The two classes are compared, and unknown samples can be projected into this space to quantify the similarity against the classes' PCA space.[94] This has been used in multiple instances to distinguish similarities between reference sample sets and unknown samples.[79,95–97]

2.3. Non-linear machine learning approaches

While models such as PCA, PLS, and SIMCA are easily interpretable, their dependence on linear algorithms is a limitation, as they only model linear correlation and covariance. Chemical relationships in botanical systems are inherently non-linear in nature; thus, non-linear methods can be especially positioned to understand these relationships *via* metabolomics data. These machine learning approaches can be unsupervised or supervised in their approach. Non-linear models can use decision

Table 1 Comparison of reported chemometric methods and their advantages/limitations with respect to botanical identification and characterization

Chemometric method	Method type	Advantages	Limitations
Principal Component Analysis (PCA)	Unsupervised multivariate	Dimension reduction, straightforward calculation, noise reduction	Not quantitative, *ad hoc* component decisions, loss of information, sensitive to data scaling
Soft Independent Modeling of Class Analogies (SIMCA)	Supervised multivariate	Outlier detection, flexible classification, class-specific modeling, interpretable	Restricted to two class comparisons, sensitive to outliers, limited inference
Partial Least Squares (PLS)	Supervised multivariate	Robust, inference for feature importance, maximizes separation, handles high-dimensional data	Risk of overfitting, assumes linearity, sensitive to data scaling, model validation crucial
Random forest	Decision tree	Robust to overfitting, non-linear relationships, high accuracy	Computationally complex, struggles with sparse datasets
Support Vector Machines (SVMs)	Kernel separation	Robust to overfitting, kernel versatility in modeling, effective with high-dimensional data	Computationally complex for large datasets, sensitive to noise and outliers, difficult to choose starting conditions
Artificial Neural Networks (ANN)	Neural network	Adaptable, non-linear modeling	No interpretability, requires large datasets, computationally complex, prone to overfitting, *ad hoc* development

trees and ensemble learning (*e.g.*, random forest, gradient boosted tree model) to organize and classify data, take advantage of kernel tricks to plot data in higher dimensional space to separate data linearly by finding hyperplanes of covariance (*e.g.*, Kernel PCA (KPCA), Support Vector Machines (SVM)), or use deep-learning neural networks (*e.g.*, artificial neural network (ANN)). Random Forest (RF) is a method that builds an assemblage of decision trees, each tree of which is trained using the dependent variable(s).[98] Each tree is then employed to classify the unknown, and a consensus outcome is output as the model result. Using multiple, randomly generated decision trees allows for more accurate classifications, and is less prone to overfitting,[99] and can be readily applied for botanical classification.[100–102] More recent tree models such as gradient boosted tree model make use of iterative adaptations to better fit models to produce more accurate estimates of the outcome.[103]

Kernel approaches involve higher-dimensional latent spaces to find orthogonal planes that maximally distinguish the data points. Kernel PCA is a non-linear unsupervised method similar to PCA using non-linear models to understand similarities and differences between samples.[105] Support Vector Machines (SVMs) are a supervised machine learning approach that can be used for regression or classification, similar to PLS. While non-linear in nature, SVMs are prone to overfitting,[106] yet have found a space in botanical authentication and identification studies.[107–109] The third main type of non-linear chemometric approach are artificial neural networks (ANNs), deep-learning machine learning models which mimic the organization of the human brain, building 'neurons' to recognize complex patterns from large data sets. Using forward and backward progression through layers of these computational 'neurons' allows for a final output to be achieved.[110,111] These neural networks have the potential to be powerful classification tools for botanical products using metabolomic profiling as input.[112–114] However, their main limitation is the lack of interpretation allowed in neural networks; the presence of layers of 'neurons' with random weights and biases results in an inability to understand the chemical distinctions upon which the decisions are based (Table 1).

Complex botanical metabolomic datasets can provide rich information on their chemical similarities and differences. However, organizing this data can be a daunting task, and it is through these chemometric approaches that such relationships can be gleaned.

3. Applications of untargeted mass spectrometry metabolomics for botanical characterization

3.1. Taxonomic identification

Botanical products may consist of preparations of single or multiple plant parts, including roots and rhizomes, stems, leaves, flowers, fruits, seeds, and/or other organs of single or multiple species. The correct taxonomic identification of the plant species and botanical parts used is crucial to understanding the potential biological activity and/or adverse effects of a preparation. However, determining the genus, species, or cultivar of raw material by morphological characteristics can be difficult or impossible. Untargeted metabolomics approaches provide rigorous analytical methods to establish similarities and differences in phytochemistry against benchmarked taxa repositories (*e.g.*, herbaria, reference material libraries, botanical gardens) or to compare multiple botanical samples against each other.

For example, in seeds, there are limited distinguishing features to differentiate among species morphologically, but the chemical composition can be unique in comparison to other species, suggesting that untargeted metabolomics may represent a means of identification. Lesiak *et al.* employed direct analysis in real time mass spectrometry (DART-MS) to analyze the metabolomic profiles of seeds from various species of *Datura*, and unique chemical footprints allowed for the differentiation and identification of *Datura* species solely by seed chemistry.[116] Similar investigations have been performed with rice, lentil, and soybean cultivars, all demonstrating that unique seed chemical profiles can be used for future identification testing.[117–119]

The *Phyllanthus* genus consists of plants used to treat multiple conditions in traditional systems of Indian medicine, including, but not limited to, jaundice, dermatitis, and various respiratory diseases.[120] To examine differences among closely related species, researchers used an untargeted LC-MS approach to create unique and discernible profiles for nine *Phyllanthus* species for authentication and quality control. In a similar study, researchers used LC-MS metabolomics with PCA and OPLS-DA to characterize 10 cultivars of cranberry (*Vaccinium macrocarpon* Ait.), revealing distinct clustering amongst the cultivars (Fig. 2). These groupings were consistent with each cultivar's genetic background, with the most closely related cultivars clustering together,[104] offering evidence that sub-species discrimination is achievable *via* these methods.

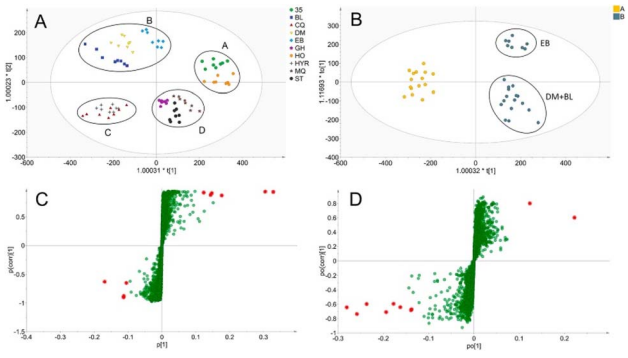


Fig. 2 An OPLS-DA model of untargeted LC-MS profiles of cranberry (*Vaccinium macrocarpon* Ait.) cultivars. (A) Scores plot for 10 cranberry cultivars; (B) scores plot for cultivar group A *vs.* B (including "Early Black" and "Demoranville" and "Ben Lear" cultivars); (C) *S*-plot for predictive component from model A *vs.* B; (D) *S*-plot for orthogonal component from model A *vs.* B (significant biomarkers highlighted in red). Reproduced with permission from Wang *et al.*[104] Copyright 2018, American Chemical Society.

This has been expanded to interrogate other generally muddled herbal products; Xu *et al.* employed direct infusion-three-dimensional-mass spectrometry to disentangle the metabolomic profiles of herbal products that contain multiple Umbelliferae plants. This analysis bypassed traditional chromatographic input into the mass spectrometer, instead joining multiple ion monitoring (MIM) MS1 scan events, enhanced product ion (EPI) experiments for MS^2 data collection, and online energy-resolved MS to provide breakdown graphs of the MIM data to yield a multi-dimensional representation of the chemical profiles (Fig. 3) that provided rapid metabolomic analysis of botanical products.[115] Guo *et al.* (2023) employed an air flow-assisted desorption electrospray ionization-mass spectrometry imaging (AFADESI-MSI), to provide tissue analysis of xylem and phloem of two *Radix Puerariae* species, (*Puerariae lobata* (Willd.) and *P. thomsonii* (Benth.)).[121] The 3′-hydroxyl puerarin level was higher in the xylem of *P. thomsonii* and higher in the phloem of *P. lobata* (Fig. 4), suggesting qualitative and quantitative differences between these two species. Subsequent analyses of the metabolomes revealed 52 discriminating metabolites.

Metabolomic techniques were able to resolve plant organs and the originating biomes within the genus *Copaifera* (Fig. 5).[122] The leaves of *C. langsdorfii* (Desf.) contained higher levels of flavonoids that the reproductive organs and branches notably lack. It was also noted that plants stressed by temperature fluctuations produced higher levels of terpenes and

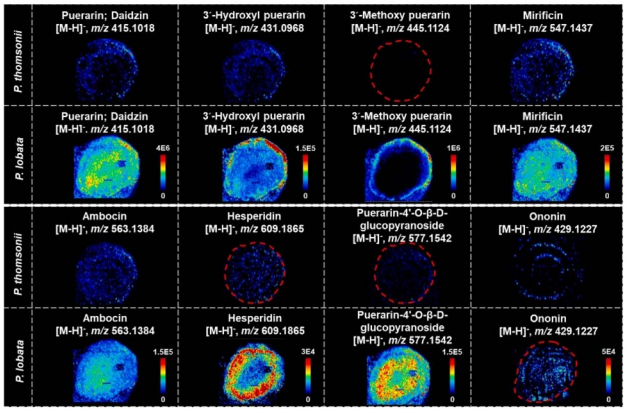


Fig. 4 Mass spec imaging (MSI) spectra of isoflavonoids in *P. thomsonii* and *P. lobata* xylem/phloem. Reproduced with permission from Guo *et al.*[121] Copyright 2023, Elsevier Ltd.

flavonoids, therefore creating a unique chemical footprint dependent on the biome in which it grew.[122] A study on American ginseng (*Panax quinquefolius* L.) roots revealed that morphological regions could be identified within the root system using ultrahigh-performance liquid chromatography with quadruple time-of-flight mass spectrometry (UHPLC-QTOF-MS). Examining the main, lateral, and fibrous root as well as the rhizome revealed 4–11 distinguishing chemical markers per morphological region.[66] Plant organs from *Piper*

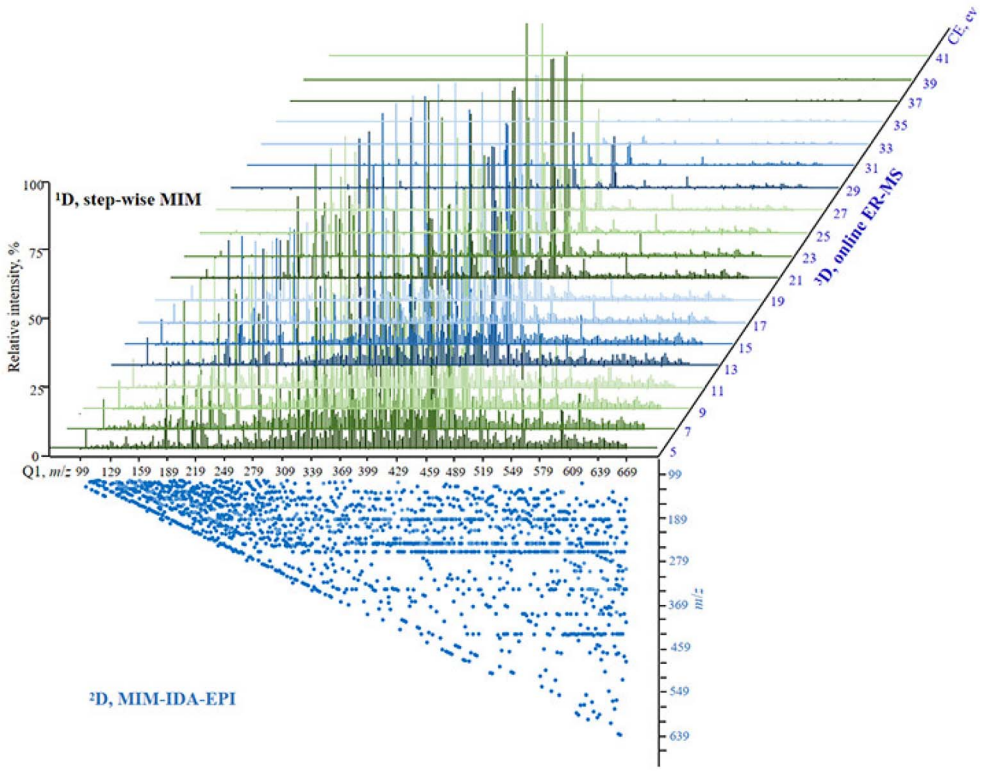


Fig. 3 Graphical representation of the direct infusion-3-dimensional mass spectrometry method for analyzing botanical products, combining a stepwise multiple ion monitoring (MIM) program, enhanced product ion (EPI) experiments for MS^2 data collection, and online energy-resolved MS (ER-MS) breakdown graphs for all MIM items. Reproduced with permission from Xu *et al.*[115] Copyright 2020, American Chemical Society.

spp.,[123] *Brachychiton acerifolius* (A. Cunn. ex G. Don) F. Muell.,[124] and *Salicornia perennis* Mill.[125] also revealed the sensitivity of metabolomics to differentiate botanical organs.

These techniques extend to discrimination within mixtures of multiple botanical species. Uncariae Ramulus Cum Uncis (Gou-Teng) is a traditional Chinese herbal medicine that can be derived from one of five different species of *Uncaria*.[126] To differentiate species comprising the medicine, untargeted LC-QTOF-MS analysis was combined with multivariate PLS-DA modeling, which was able to differentiate the various species and provide discriminating biomarkers. This resulted in a support vector machine (SVM) model which proved to be 100% accurate in identifying origin species from 20 commercial samples, identifying four samples which did not contain any of the five official species.[107]

The majority of metabolomics studies analyze samples that have all been procured and processed in a relatively short time from when they were first grown. Resende *et al.* (2020) pushed this boundary by investigating *Solanum* spp. herbarium vouchers, and found that, after 10 years of storage, samples from *S. argenteum* Dunal and *S. pseudoquina* A. St.-Hil. were still able to be differentiated by their chemical profiles.[127] Even older herbarium specimens, of the medicinal genus *Salvia* (some dating to 1862) were discriminated by metabolomics,[128] suggesting that the age of the specimen did not have a significant effect on the chemical composition for untargeted methods, but the specific compounds from a targeted approach were affected. Similar studies testing preserved samples from the genus *Nicotiana*[129] and the Gentianaceae family[130] have also demonstrated sufficient specificity and sensitivity to discriminate between botanical samples.

3.2. Adulteration

Adulteration, which includes the omission or inadvertent or purposeful exchange of botanical specimens in a formulation, is an increasingly concerning issue in the herbal and dietary supplements marketplace. Recent estimates of adulteration in five of the most popular and top-selling herbs in the United States, ginkgo, black cohosh (*Actaea racemosa* L.), elderberry, echinacea (*Echinacea* spp.), and turmeric (*Curcuma longa* L.) found 17–57% adulteration in tested products.[131] Adulteration includes many types of ingredient substitutions, dilutions, additions, and contaminants done intentionally for economic reasons or unintentionally due to misidentification or mishandling of materials along the supply chain.[41,131,132] Thus, it is imperative that strategies be deployed for characterizing botanicals to ensure that the proper material is being harvested, processed, and offered to consumers, and that fraudulent materials in otherwise known, claimed, or labeled ingredients are detectable, even in post-market analyses.

Detecting botanical adulteration includes using reference materials, often available through independent sources, that are ideally traced from harvest, authenticated taxonomically, and vouchered.[132] These validated samples have verified, unadulterated chemical profiles that can serve as standard fingerprints for comparison within metabolomic analyses,[132] as well as provide potential biomarker compounds that can then be utilized to differentiate species within a botanical mixture.[132] Sarker *et al.* used authenticated *Tinospora* spp. for LC-HRMS metabolomics analysis with OPLS-DA and PLS-DA chemometric models to determine the "normalized abundance" of seven biomarkers to distinguish *Tinospora cordifolia* Thunb., a popular Ayurvedic herb, from *T. crispa* L. and *T. sinensis* (Lour.) Merr., two similar species that can be mistakenly identified and substituted for *T. cordifolia*, but have potential adverse health effects.[59] Efforts to cover adulteration may also occur when a labeled ingredient is substituted or diluted with an unrelated species that has a similar chemical profile or a shared key biomarker known for having beneficial health effects, in attempt to bypass quality control systems that focus on individual or a small number of biomarker compounds.[131] Using LC-MS metabolomics, Wallace *et al.* were able to detect adulteration at the 10% m/m level of Chinese goldthread (*Coptis chinensis* L.) roots in goldenseal (*Hydrastis canadensis* L.) root preparations using a supervised method, soft independent modelling by class analogy (SIMCA).[133] Both plants contain berberine, a benzylisoquinoline alkaloid credited for much of the bioactivity of goldenseal, but only goldenseal contained the characteristic compounds hydrastine and canadine, while Chinese goldthread contained magnofluorine, coptisine, dihydrocoptisine, palmatine, and jatrorrhizine. As the relative intensity of unique compounds fluctuated, researchers were able to determine the level of adulteration, as well as quantify and view changes in PCA clustering which served as a visual representation of metabolomics data without having prior information on the variance of the dataset.[133]

Likewise, quality control screenings of botanicals that have historically been completed with UV detectors, such as assays using spectrophotometric quantification or chromatographic techniques with HPLC are often not sensitive enough to detect adulterants, and some producers may purposely use adulterants that evade these detection techniques. For example, cranberry fruit products are widely consumed for their flavor and

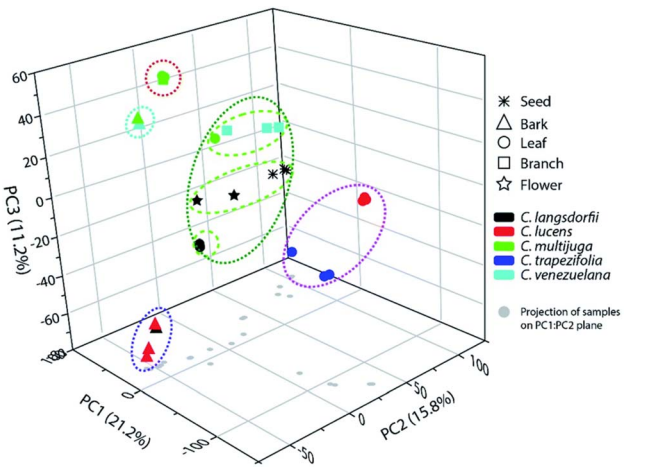


Fig. 5 PCA scores plot of *Copaifera* spp. showing clustering across species as well as plant physiological part. Reproduced with permission from da Silva *et al.*[122] Copyright 2021, Royal Society of Chemistry.

health benefits. The latter are thought to be at least in part due to bioactive A-type proanthocyanidins (PAC)s that have been shown to be uniquely effective for treating urinary tract infections. Quality screening methods used to detect PACs in cranberry products, such as the colorimetric 4-(dimethylamino) cinnamaldehyde (DMAC) reaction, cannot differentiate between A-type and B-type PACs, so frequent adulterations to cranberries have included less expensive fruit extracts such as apple, grape, or blueberries that have B-type PACs and would yield positive results in the DMAC assay, but do not possess the same bioactivity and efficacy in treating urinary infections. Peanut skins, which have A-type PACs, have also been used as an adulterant but may pose a serious health risk to those with peanut allergies. While colorimetric assays are susceptible to deception with non-A-type PACs, mass spectrometry metabolomics methods using UPLC-MS or MALDI-TOF are capable of selectively identifying individual PACs with differing bond structures.[134–136]

The herb saffron is obtained from the dried stigmas of *Crocus sativus* L., which give rise to its characteristic color and flavor. In comparing commercial samples, some of which were adulterated with other parts of the saffron flower (5–50% m/m stamens, tepals, or other constituents), an untargeted metabolomic approach was able to differentiate authentic *versus* adulterated samples at the 5% m/m level with high levels of model correlation ($R^2Y = 0.99$) and predictive capacity ($Q^2Y = 0.99$) (Fig. 6).[84] This provides an interesting case study on the differentiating potential of metabolomics characterization techniques; these were plant parts derived from the same species, and thus some other analytical approaches (*e.g.* genetic analyses) would have been unable to identify the impurity in the samples. Another study employed untargeted LC-MS metabolomics coupled with unsupervised dimensional reduction (PCA) to detect differences in the metabolomes between mass produced saffron and locally sourced saffron. Furthermore, supervised analysis (OPLS-DA) revealed 9(*S*),10(*S*),13(*S*)-trihydroxy-11(*E*)-octadecenoic acid as a significant biomarker able to distinguish mass produced saffron, and using oxidized crocins, the method was capable of differentiating quality levels of saffron regardless of geographic origin. Furthermore, this analysis was able to detect adulterations with paprika (*Capsicum annuum* L.) or turmeric at the 10% m/m level, representing a more sensitive approach compared to traditional methods of adulteration detection.[137,138]

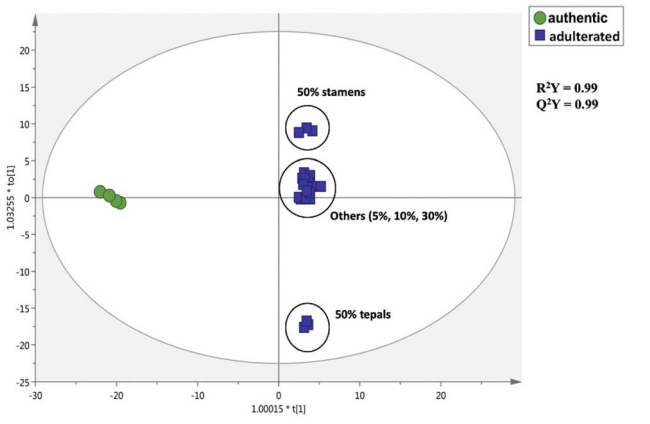


Fig. 6 Orthogonal Partial Least Squares-Discriminant Analysis (OPLS-DA) scores plot of authentic *vs.* adulterated saffron samples. Model metrics R^2Y and Q^2Y are provided. Reproduced with permission from Senizza *et al.*[84] Copyright 2019, Elsevier Ltd.

3.3. Intraspecies variation in the chemotype

The demonstrable differences in phytochemical makeup between taxa have been well established. Variations in plant biosynthetic pathways can underlie both the potential biological activity or toxicity and provide insight into the identification or misidentification of the botanical. However, it is well known that there are numerous other factors – both biotic and abiotic – that drive chemical variation within a given taxonomic group. Many of these effects are well-documented, and we will not attempt to reiterate those here.[139–141] As these have multifactorial impacts on the chemistry, metabolomics represents a powerful analytical tool for deciphering such alterations to the metabolome.

3.3.1. Geography and environment. The geographic location and environmental conditions surrounding a plant's growth have an unmistakable impact on the specialized metabolite production in the plant. This has been known for a long time in the wine and food industry, where a unique location's influences on secondary metabolite composition give a distinct aroma and flavor profile, or *terroir*.[142,143] The same principles can be applied to botanical supplements, medicinal plants, and nutraceuticals.[144]

In an examination of saffron commercial samples including some from protected designations of origin (*i.e.*, single authenticated origin samples, PDO), distinct metabolite profiles were made evident by LC-MS/MS, suggesting that geographic differences were reflected in the saffron metabolomes.[84] LC-MS/MS metabolomics was also used to differentiate different countries of origin for black pepper (*Piper nigrum* L.) between Brazil, Sri Lanka, and Vietnam and highlight putative biomarkers for discrimination of origin (Fig. 7A and B).[145,146] In two other studies, GC-MS metabolomics distinguished black peppers grown in different regions of Brazil[147] and China.[148] Untargeted analysis of lily (*Lilium* spp.) bulbs also yielded distinct clusters based upon 5 different geographic origin locations.[149] Untargeted gas chromatography-mass spectrometry (GC-MS) metabolomics highlighted different chemotypes of *Ferula assa-foetida* L. based upon geography and climate conditions,[150] and a combination of untargeted metabolomics and UV-vis differentiated geographic origins of *Cotinus coggygria* Scop.[151] Barbosa *et al.* (2020) employed non-targeted LC-MS profiling to discriminate between different geographic sources of paprika; the chemical fingerprints provided clear distinctions between paprika grown in Spain *versus* the Czech Republic.[92]

The geographic location of production of a botanical, whether a food product, nutraceutical, or medicinal plant, is a collective representation of multiple biotic and abiotic factors that exert influence over the biosynthetic regulation of small

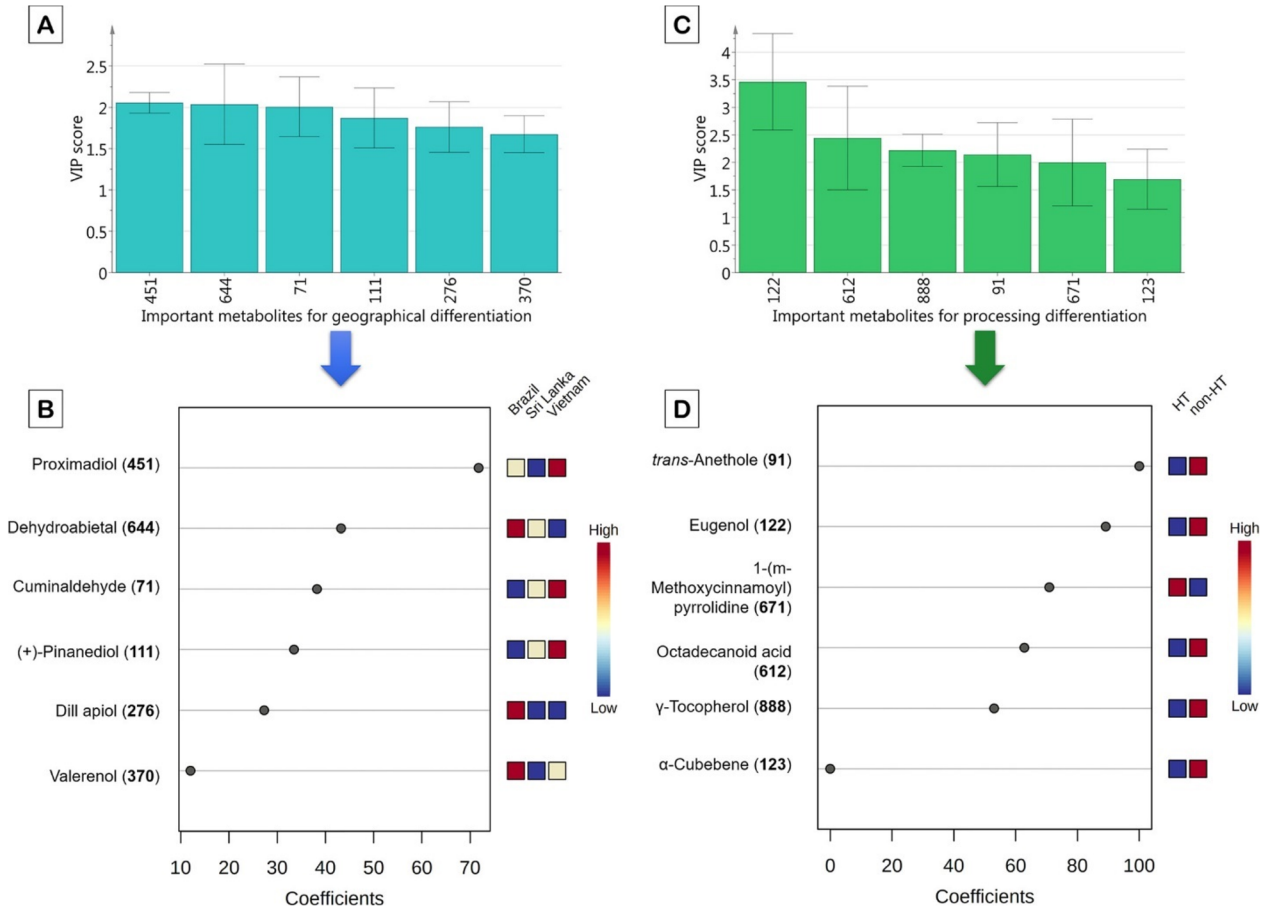


Fig. 7 Key biomarkers revealed by GC-MS analysis of black pepper (*Piper nigrum*) for (A) geographical authentication and (C) processing authentication. Relative contributions of the top six metabolites from the OPLS-DA model for both (B) geographical origin and (D) processing of black pepper. Relative concentrations of the corresponding metabolite are indicated by colored boxes on the right (relatively high levels are indicated in red, and the lower ones, in blue). Reproduced with permission from Rivera-Pérez *et al.*[146] Copyright 2021, American Chemical Society.

molecules in a plant and contribute to a unique small molecule profile, or chemotype. Indeed, the differences based upon geography are distinct and readily identified *via* metabolomics-based analyses; these have even been positioned as a potential marketing mechanism for botanical products to help 'verify' geographically indicated products for the marketplace.[152]

3.3.2. Production and processing effects. The production of medicinal and nutritional botanicals involves a number of variables that can impact the overall chemical profile of the plant sample. Furthermore, post-harvest processing plays a large role in the ultimate small molecule profile, and thus sensory and bioactive properties, of the final product.

Green tea (*Camellia sinensis* (L.) Kuntze) is a prime example of a botanical with nutritional and medicinal applications upon which production and processing can have large impacts. Green tea that was grown in the shade *versus* the sun demonstrated substantial differences in chemical composition; untargeted metabolomics revealed that sun-grown tea contained more galloylquinic acid, epigallocatechin, epicatechin, succinic acid, and fructosethan, whereas shade tea samples had more flavonoid derivatives, including gallocatechin, strictinin, apigenin glucosyl arabinoside, quercetin *p*-coumaroylglucosyl-rhamnosylgalactoside, kaempferol *p*-coumaroylglucosylrhamnosylgalactoside, malic acid, and pyroglutamic acid.[153] The tea plantation's elevation also proved a significant variable in determining green tea chemistry, with theogallin, citric acid, theanine, and sucrose increasing in concentration with increasing elevation.[154] Roasting of tea leaves initiates a number of biochemical reactions which fundamentally alter the chemical profile of the tea; untargeted metabolomics was employed to distinguish between white, green, black, and oolong teas,[155] while also showing the dynamic shifts in metabolome during the roasting process.[156] Metabolomics analysis was employed to understand the chemical differences produced by intercropping systems of green tea. Intercropping Chinese chestnut (*Castanea mollissima* Blume) with tea revealed substantial shifts to the phytochemical profile of the tea compared to monoculture tea plants, including allantoic acid, sugars, sugar alcohols, and oleic acid.[157]

Herbs that play a role both in culinary traditions as well as possessing potential nutritional and medicinal value can become heavily processed commodities through drying, grinding, cooking, and extracting processes. The application of sterilization or drying has the potential to impact a wide variety

of metabolites, and therefore metabolomics can serve as a tool for analyzing the changes brought on by processing in botanical products. In the case of black pepper, heat treatment resulted in a shift in key volatile markers that were captured by GC-MS analysis (Fig. 6C and D).[146] Similar approaches were utilized for thyme (*Thymus vulgaris* L.) to differentiate samples based upon production and processing parameters.[158,159]

3.4. Merging activity with metabolomics

Metabolomics datasets can be analyzed and interpreted either solely based upon their feature (metabolite) information (unsupervised approaches, see Section 2.1), or by adding dependent variables to guide the analysis (supervised approaches, Section 2.2). Most cases presented thus far use categorical data as the dependent variable (*e.g.*, country of origin, plant physiology, species), which facilitates the use of linear regression models incorporating non-numeric metadata. However, it is crucial to note that regression analyses with numeric dependent data can shed unique insight into the bioactive chemistry of botanical preparations. Often referred to as "biochemometrics,"[90] merging biological activity with chemical profiling represents a robust approach to supervise the downstream analysis and identify putative biomarkers for follow-up characterization, providing a crucial link between the chemical metabolome and overall phenotype of the sample.[160] The incorporation of sensory qualities, *in vitro* biological activities, or other phenotype data has powerful applications in characterizing botanical products. The exploration of biochemometrics for specific natural product discovery lies outside the scope of this particular review (others have accomplished this very well[160–163]), but these approaches still hold value for discriminating, characterizing, and identifying the botanical sources of biological activity.

One use for this approach is in grading and profiling the flavor chemistry of plant products. In analyzing monoculture tea and intercropped tea, Wu *et al.* noted that the intercropped tea was graded higher than a monoculture tea from the same plantation, suggesting there is a taste alteration associated with the intermingled system. The metabolome associated with green tea intercropped with Chinese chestnut had alterations in the amino acid and flavonoid composition, suggesting a shift to less bitter compounds, which could increase palatability.[157] Similarly, sensory and metabolomic analysis of different grades of Huangshan Maofeng green tea revealed procyanidins as the main quality markers differentiating the grades.[164] Analyses of ground coffee (*Coffea* spp.) used OPLS-DA to identify metabolites from GC-MS profiles associated with sensory quality; this led to the identification of three main metabolites (methyl pentanoate, 2-furfurylthiol, and L-homoserine) as significant markers of aroma quality from the ground coffee samples.[165] Beyond beverages, sensory analysis integrated with metabolomic profiling has been employed for fruits,[166,167] vegetables,[168] herbs,[169] and culinary vinegars.[170]

Incorporating bioactive data into chemometric analyses can serve to further differentiate botanical (sub)species and physiological parts. Bioactive molecule families (*e.g.*, phenolics, alkaloids, and terpenoids) were used to distinguish between three *Bryophyllum* species based on their differential expression.[171] The inhibitory activity of three different *Trigonella* species on digestive enzymes (α-amylase and α-glucosidase) was integrated with untargeted metabolomics to discriminate the species,[172] and in another study, antioxidant activity was used as a supervising variable to distinguish peach (*Prunus persica* L. Batsch) cultivars.[173] Additionally, anti-inflammatory data was instrumental in determining the responsible plant organs from Armenian cucumber (*Cucumis melo* var. *flexuosus* L.).[174] Thus, organoleptic and bioactive data can complement global metabolomics analyses to provide crucial clues to characterize botanical products.

4. Challenges

Metabolomics is a powerful analytical tool in the field of botanical characterization and identification, with numerous applications across plant biology, nutraceuticals, and natural products chemistry. However, metabolomics, and especially mass spectrometry-based metabolomics, is not without challenges and obstacles as a developing omics field. Limitations of the analytical instrument, the current shortcomings of metabolomics databases, and inherent variability in botanical samples represent three key challenges faced by researchers.

4.1. MS limitations

Mass spectrometers have grown more sensitive and more accurate in their measurement of molecular (and associated fragment) masses,[175] furthering the applications and robustness of MS as an analytical tool for metabolomics analyses. There are multiple instrument parameters that impact the quantity of masses detected, and influence the number and quality of fragmentation spectra obtained.

One parameter is the ionization method, which is essential for the detection of molecules in the mass analyzer. Multiple technical approaches can be employed for ionization, with (heated) electrospray ionization ((H)ESI) being the most widespread.[176] The polarity, voltage strength, and associated temperatures set during method development all create unique conditions for ionization, which are not universally adequate to ionize all potential metabolites in a sample and can also induce varying degrees of in-source fragmentation of the molecular ions.[177,178] Once ions are injected into the instrument, the mass resolution, cycle time, automated gain control, and ion injection time can all have an effect on measurement.[179,180] Furthermore, the collision energy, the level at which the precursor ions are fragmented to provide MS/MS tandem spectra, can vary between instruments, both in terms of method employed (*i.e.*, collision induced dissociation (CID) or higher-energy collisional dissociation (HCD)), and the energy level set to achieve fragmentation.[181] This will impact the size, abundance, and pattern of fragments, and impact the ability of the fragments to be matched against analytical standards or databases.[182,183]

Furthermore, phytochemical isomers are ubiquitously distributed in plants and frequently interact along different pathways and different physiological functions. Therefore, identifying and discriminating between isomers when annotating a metabolomics dataset can be critical in understanding the biological importance of metabolites. However, mass spectrometry, even MS/MS, is often "isomer-blind,"[185] as isomers possess identical MS^1 accurate masses, and often the same (or very similar) fragmentation patterns. Ion mobility spectrometry (IMS) is one means of separating isomeric metabolites in the gas phase post-LC elution and then passing them to the mass analyzer; this has been shown to improve isomeric detection and annotation in traditional LC-MS botanical sample analysis,[184] as well as mass spectrometry imaging experiments.[186] Li *et al.* (2021) identified 172 isoquinoline alkaloid analogues across four herbal samples (*Coptis chinensis*, *C. deltoidei* C. Y. Cheng & P. K. Hsiao, *C. teeta* Wall., and *Corydalis yanhusuo* W. T. Wang) using IMS/MS (Fig. 8).[184] Li *et al.* (2021) employed UPLC-IM-QTof-MS metabolomics to differentiate *Panax notoginseng* (Burkhill) F. H. Chen parts (root/rhizome *versus* leaf *versus* flower bud), characterizing 328 ginsenoside isomers in the process, and identifying five as potential discriminatory biomarkers.[187] For isomers that lack ion mobility, adduct formation increases the ionization efficiency of a sample, which in turn optimizes mass spectrometry signaling patterns. In a study comparing cyclic IMS with different adduct agents to distinguish catechin epimers, protonated adducts formed two stable adducts that IMS/MS could not directly detect, while sodium adducts gave clear separations of unmodified epimers.[188]

Compounding these instrumental intricacies is the immense complexity of botanical metabolomes. The variation in mass, structural connectivity, and potential for isomers, translate into a near impossibility of a single universal setting that is capable of analyzing every metabolite within a sample. Distinguishing between isomers alone remains a substantial challenge for mass spectrometry, especially given the different types of isomers capable of being present in a botanical sample (*e.g.*, structural, geometric, stereo).[188] And considering the breadth of structural motifs present in a botanical matrix, from non-polar (and poorly ionizable) lipids to the widely prevalent polar metabolites (*e.g.*, saccharides and amino acids), a single method would be unable to properly detect all of these compounds. Thus, untargeted metabolomics studies on botanical matrices are capable of measuring only a portion of the whole metabolome, even if efforts are taken to 'optimize' the MS conditions.[179] Some of these limitations are being addressed with advances in orthogonal techniques in ionization (*e.g.*, atmospheric pressure photo ionization, APPI, and atmospheric pressure chemical ionization, APCI),[189,190] additional separatory techniques (*e.g.*, ion mobility)[188] and data acquisition approaches.[191] However, there exists instrumental variation that cannot be recapitulated by simply mirroring ionization, detection, and fragmentation methods between different mass spectrometers, which further limits the interlaboratory translatability of mass spectrometric metabolomics analyses.[192–194] Furthermore, highlighted features that are discriminatory between different phenotypes, or as biomarkers/bioactive compounds, require additional authentication to confirm structure and activity. Mass spectrometry isn't capable of discerning some structural characteristics, and thus orthogonal approaches, such as NMR, are necessary to fully confirm the chemical structures.[195,196]

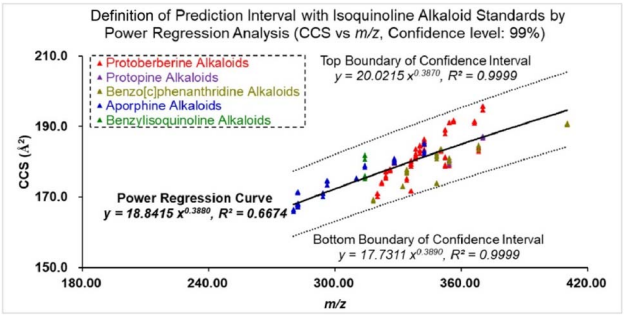


Fig. 8 Prediction interval for screening isoquinoline alkaloids from four botanicals, using collision cross-section and *m/z* with a power regression analysis and 99% confidence level. Reproduced with permission from Li *et al.*[184] Copyright 2021, Elsevier B.V.

4.2. Databases and annotation

Annotation of LC-MS data has traditionally matched experimentally-derived accurate mass and MS/MS spectra with those from authentic chemical standards.[197] As untargeted metabolomics has grown in the past two decades, the community has endeavored to create experimental MS/MS spectral databases to facilitate identification of unknown signals. A range of MS/MS databases such as NIST,[198] HMDB,[199] MassBank,[200] Global Natural Products Social Molecular Networking (GNPS),[201] LipidBlast,[202] and METLIN[203,204] have expanded the ease of access to reference standard peaks that may mirror an unknown compound of interest.[205] However, though valuable, matching unknown peaks from a library can be challenging, as spectral patterns from mass spectrometers suffer from the same instrument- and method-specific limitations described above (*e.g.*, instrument configuration, ionization, and collision energy). And most libraries are currently curated with lower collision energies, despite the potential advantages of higher energy CID/HCD spectra. Mismatched collision energies can impact the MS/MS 'fingerprint' of the precursor ion and lower the successful hit rates from databases; some have suggested that collecting MS/MS data over a collision energy ramp or multiple distinct energy levels (*e.g.*, 0, 10, 20, 30, 60 eV) would improve library searching.[197]

The ability of a database to annotate untargeted metabolomics data is also predicated on the spectra that are available in the library itself. While MS/MS accessions have been steadily increasing over the last decade, only 10–14% of metabolite signals are able to be annotated.[206,207] To circumvent the dearth of reliable MS^2 data, computational approaches such as GNPS[201] and SIRIUS[208] use fragmentation similarity scoring to construct molecular networks relating known and unknown

molecular structures. For both database and computational methods, the source of metabolites can be a source of inconsistency in the ability to annotate a metabolomics dataset; while multiple databases heavily feature human[199] or microbial[209] metabolites, less effort has been devoted to libraries focused on plant-based metabolites. It is possible to obtain spurious annotations when searching against libraries that are not specific to the taxonomic classification of the organism being analyzed. Thus, botanical-specific databases and libraries would be an essential tool for the advancement of annotations in mass spectrometry metabolomics; some efforts have been made to this effect, including the Sumner Spectral Library, Sam Sik Kang Legacy Library, and the newest Library Enabling Annotation of Botanical Natural Products (LEAFBot).[210] These libraries, to be well-populated and useful to researchers, will need to be constructed with sound structural data and expanded to more chemical families and compounds. This requires data from individual compounds, either isolated or purchased, that have been comprehensively elucidated using NMR[211] to ensure accurate annotation after being deposited into the library, and that have a robust set of MS fragmentation data across multiple energy levels and different mass analyzers.[197,198]

4.3. Chemical variation across botanical materials

Global metabolomics comparisons for botanical characterization often rely on comparisons against "botanical reference materials" (BRMs), which are samples whose identity has been confirmed by one or more analytical methods and is therefore widely accepted as representative of a species/cultivar.[212] These reference materials are invaluable for comparing unknown samples in an untargeted metabolomics experiment, but the use of BRMs is not without challenges. These botanical reference materials must be fully validated prior to acceptance and use, and there isn't adequate availability of reference materials for all potential botanicals under study. Reference materials may be available as just one or two different samples, many of which lack appropriate identifying information and accessible vouchers for validation.[16] The National Institute of Standards and Technology (NIST) maintains a robust catalog of reference materials, and efforts by the NIH Office of Dietary Supplements (ODS) have expanded reference material availability, but it is ultimately infeasible to have multiple (or even a single) BRMs for every botanical species and cultivar.[213]

However, as noted above, even when reference materials are available for comparison, there is not a guarantee that BRMs are sufficient to accurately characterize unknown samples; the final phytochemical metabolome depends on more than just genetics: climate,[214] nutrients and soil conditions,[215] herbivory and predation,[216] and pathogen exposure[217] all play a role in the development of a specific metabolome signature.

The impact of environment, climate, and post-harvest conditions results in potentially large variation in the phytochemical makeup of botanical samples, which could impact their biological activity, nutritional value, or potential toxic adverse effects. This could potentially render BRMs, which are procured from reputable sources or grown under controlled conditions, unable to adequately reflect variations in botanical metabolomes from plants grown in different geographic or climactic regions, or ones subjected to different post-harvest conditions. Abraham *et al.* (2025) employed reference samples of *Ocimum* species (basil) that were grown in a greenhouse and used as a training dataset to characterize commercial basil samples. However, the reference materials were unable to provide a sufficient basis for discriminating the commercial samples, which had unknown growing, harvesting, drying, or storage conditions.[218] Buț *et al.* (2023) reviewed the variation of phytosterol content in saw palmetto (*Serenoa repens* (W. Bartram) Small) dietary supplements.[219] The fluctuation in individual and total phytosterols was significant; this variability could be attributed to the botanical material being sourced, undeclared admixtures or adulterants, formulations including vegetable oils, or even differences in analytical methods used in analyzing the samples. Both represent enhanced variability in the final metabolome, and thus the authors concluded that this inconsistency, couple with the lack of reporting and standardization of the chemotype of the material, perhaps limits the overall clinical usefulness of saw palmetto. Fluctuating metabolome levels within green tea cultivars were also associated with different flavor profiles, which varied between years, seasons, and climate; these changes in the metabolome of the tea were posited to signify greater challenges for farmers to produce harvests with consistent quality and customer preference.[220] And metabolomics was used to correlate chemical profiles of *Sinocalycanthus chinensis* (W. C. Cheng & S. Y. Chang) with environmental conditions during growth to optimize harvest practices and improve biological activity.[221] These varying metabolomes stress the need for deeper investigations into how environmental stressors impact a plant's biochemical reactions and overall chemical profile, and to what degree this impacts desirable and adverse characteristics in the plant.

5. Conclusions

Botanicals are complex chemical systems, extremely dynamic and sensitive to genetics as well as biotic and abiotic influences. Accurate classification and characterization of botanicals is crucial for their continued adoption as medicines, dietary supplements, and nutraceuticals, and while multiple analytical techniques for such identification exist (*e.g.*, morphological, genetic, or single biomarker analyses), mass spectrometry-based metabolomics has emerged as a sensitive multi-faceted technique for discriminating botanical samples. The sensitivity and specificity of mass spectrometry enables large data collection that can be parsed using multivariate or machine learning statistical methods for discriminating patterns, evaluating similarities, and ultimately understanding the chemical underpinnings of phenotypic differences between samples. The field is growing rapidly, and while instrumental and logistic hurdles still limit the totality of compounds analyzed by mass spectrometry, advanced methods are endeavoring to distinguish between isomers, expand the range of compounds visualized by mass spectrometry, and enhance reproducibility

between instruments or labs. In addition, computational methods are expanding to facilitate annotation of these large datasets.

However, evidence is emerging that (sub)species identification might not be adequate to provide a full characterization of a botanical product, especially considering its potential nutritive/health/toxic effects. A plant's specialized metabolite composition is guided by the influence of abiotic and biotic factors, which can be unique to a particular geography, production characteristics, or even post-harvest handling. This has precedence in dietary products (*e.g.*, the *terroir* of wine or green tea) but has been expanding to other botanical products (*i.e.*, cannabis). The differences in metabolome could underlie differences seen at the clinical level, in which species identification is insufficient to encapsulate the chemical profile that achieves a desired effect (or avoids a deleterious consequence). Therefore, it is worth asking whether characterization and acceptability of botanical medicines, supplements, *etc.* needs to evolve beyond taxonomic identification as a gold standard and instead broaden to encompass the directly relevant molecular profiles that would be characteristic of a particular chemotype, which can be accurately measured by metabolomics approaches. It is our hope that greater consideration of the nuances of botanical secondary metabolite profiles will enable the design of rigorous and reproducible studies of complex botanical natural products and to increase our understanding of their biological effects.

6. Conflicts of interest

The authors declare no conflicts of interest.

7. Data availability

The current manuscript constitutes a review. It does not include any new, unpublished, primary research results, software, or code, and no new data has been generated in this review. All referenced works have been duly cited in the references section of the manuscript.

8. Acknowledgements

This work was supported by the USDA National Institute of Food and Agriculture and Hatch Appropriations under Project #PEN04956 and Accession #7006496 (JJK) and grant 2023-67017-39058 (JJK, RTJ, and XC) as well as the National Institutes of Health's National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK, T32DK120509, MMR) and the National Center for Complementary and Integrative Health (NCCIH, F31AT013158, SGA). Additional funding was provided by the Anne C. Chatham Fellowship in Medicinal Botany from the Garden Club of America (SGA), and through a research support agreement between Penn State College of Medicine and PA Options for Wellness (FTC).

9. References

1 E. L. Lee, N. Richards, J. Harrison and J. Barnes, *Drug Saf.*, 2022, **45**, 713–735.
2 C.-H. Wu, C.-C. Wang, M.-T. Tsai, W.-T. Huang and J. Kennedy, *J. Evidence-Based Complementary Altern. Med.*, 2014, **2014**, 872320.
3 K. M. Job, T. K. L. Kiang, J. E. Constance, C. M. T. Sherwin and E. Yu. Enioutina, *Expert Rev. Clin. Pharmacol.*, 2016, **9**, 1597–1609.
4 H. Vogtman, *Dietary supplement usage increases, says new survey*, https://www.crnusa.org/newsroom/dietary-supplement-usage-increases-says-new-survey, accessed 29 July 2021.
5 T. Smith, H. Bauman, H. Resetar and E. Craft, *HerbalGram*, 2024, 52–69.
6 M. Missenda, D. Morris and D. Nault, *J. Integr. Complementary Med.*, 2023, **29**, 584–591.
7 G. Porras, F. Chassagne, J. T. Lyles, L. Marquez, M. Dettweiler, A. M. Salam, T. Samarakoon, S. Shabih, D. R. Farrokhi and C. L. Quave, *Chem. Rev.*, 2021, **121**, 3495–3560.
8 US National Library of Medicine, *Clinical Trial Database*, https://clinicaltrials.gov/ct2/home, accessed 23 October 2024.
9 D. Domingo-Fernández, Y. Gadiya, A. J. Preto, C. A. Krettler, S. Mubeen, A. Allen, D. Healey and V. Colluru, *J. Nat. Prod.*, 2024, **87**, 1844–1851.
10 D. J. Newman and G. M. Cragg, *J. Nat. Prod.*, 2020, **83**, 770–803.
11 M. C. Ichim, *Front. Pharmacol*, 2019, **10**, 1227.
12 M. C. Ichim and A. Booker, *Front. Pharmacol*, 2021, **12**, 666850.
13 V. Navarro, B. Avula, I. Khan, M. Verma, L. Seeff, J. Serrano, A. Stolz, R. Fontana and J. Ahmad, *Hepatol. Commun.*, 2019, **3**, 792.
14 S. Gafner, T. Borchardt, M. Bush, S. Sudberg, N. G. Feuillère, Y. R. Tenon, J. H. Jolibois, P. J. N. Bellenger, H. You, R. E. Adams, J. Stewart, I. Dagan, T. Murray, D. L. Erickson and M. J. Monagas, *HerbalGram*, 2021, 24–32.
15 D. D. Soejarto, E. M. Addo and A. D. Kinghorn, *J. Ethnopharmacol.*, 2026, **354**, 120474.
16 W. L. Applequist and J. S. Miller, *Anal. Bioanal. Chem.*, 2013, **405**, 4419–4428.
17 N. Nazar, A. Saxena, A. Sebastian, A. Slater, V. Sundaresan and T. Sgamma, *Phytochem. Anal.*, 2025, **36**, 7–29.
18 A. Amin and S. Park, *Plants*, 2025, **14**, 2234.
19 L. Grazina, J. S. Amaral and I. Mafra, *Compr. Rev. Food Sci. Food Saf.*, 2020, **19**, 1080–1109.
20 K. M. Michetti, V. Pérez Cuadra and V. N. Cambi, *Rev. Bras. Farmacogn.*, 2019, **29**, 137–146.
21 R. S. Pawar, S. M. Handy, R. Cheng, N. Shyong and E. Grundel, *Planta Med.*, 2017, **83**, 921–936.
22 M. R. Joharchi and M. S. Amiri, *Avicenna J. Phytomed.*, 2012, **2**, 105.

23 H. Da-cheng, C. Shi-lin, X. Pei-gen and P. Yong, *Pharm. Biotechnol.*, 2009, **2**, 250–261.
24 M. C. Ichim, A. Häser and P. Nick, *Front. Pharmacol*, 2020, **11**, 876.
25 S. Sultana, M. A. Khan, M. Ahmad, A. Bano, M. Zafar and Z. K. Shinwari, *Pak. J. Bot.*, 2011, **43**, 141–150.
26 H. X. Kan, L. Jin and F. L. Zhou, *Pattern Recogn. Image Anal.*, 2017, **27**, 581–587.
27 J. Wäldchen and P. Mäder, *Methods Ecol. Evol.*, 2018, **9**, 2216–2225.
28 J. Pärtel, M. Pärtel and J. Wäldchen, *AoB Plants*, 2021, **13**, plab050.
29 A. G. Hart, H. Bosley, C. Hooper, J. Perry, J. Sellors-Moore, O. Moore and A. E. Goodenough, *People Nat.*, 2023, **5**, 929–937.
30 N. Lankasena, R. N. Nugara, D. Wisumperuma, B. Seneviratne, D. Chandranimal and K. Perera, *Comput. Biol. Med.*, 2024, **183**, 109349.
31 A. B. D. Selvam, *Asian J. Pharm. Res. Dev.*, 2018, **6**, 4.
32 A. Ouarghidi, G. J. Martin, B. Powell, G. Esser and A. Abbad, *J. Ethnobiol. Ethnomed.*, 2013, **9**, 59.
33 B. Rewald, C. Meinen, M. Trockenbrodt, J. E. Ephrath and S. Rachmilevitch, *Plant Soil*, 2012, **359**, 165–182.
34 E. A. Dauncey, J. T. W. Irving and R. Allkin, *J. Pharm. Pharmacol.*, 2018, **71**, 4–14.
35 H. J. de Boer, M. C. Ichim and S. G. Newmaster, *Drug Saf.*, 2015, **38**, 611–620.
36 A. Parveen, Y.-H. Wang, O. Fantoukh, M. Alhusban, V. Raman, Z. Ali and I. A. Khan, *J. Pharm. Biomed. Anal.*, 2020, **178**, 112894.
37 T. Sgamma, C. Lockie-Williams, M. Kreuzer, S. Williams, U. Scheyhing, E. Koch, A. Slater and C. Howard, *Planta Med.*, 2017, **83**, 1117–1129.
38 M. Staats, A. J. Arulandhu, B. Gravendeel, A. Holst-Jensen, I. Scholtens, T. Peelen, T. W. Prins and E. Kok, *Anal. Bioanal. Chem.*, 2016, **408**, 4615–4630.
39 P. Taberlet, E. Coissac, F. Pompanon, C. Brochmann and E. Willerslev, *Mol. Ecol.*, 2012, **21**, 2045–2050.
40 A. C. Raclariu, M. Heinrich, M. C. Ichim and H. de Boer, *Phytochem. Anal.*, 2018, **29**, 123–128.
41 M. Arya, I. S. Shergill, M. Williamson, L. Gommersall, N. Arya and H. R. Patel, *Expert Rev. Mol. Diagn.*, 2005, **5**, 209–219.
42 Y. Tao, S. Duan, K. Yu, X. Cheng, X. Li, W. Zhang, Y. Zhang and F. Wei, *Molecules*, 2025, **30**, 3763.
43 W. Sun, J. Li, C. Xiong, B. Zhao and S. Chen, *Front. Plant Sci.*, 2016, **7**, 367.
44 C. Wallinger, A. Juen, K. Staudacher, N. Schallhart, E. Mitterrutzner, E.-M. Steiner, B. Thalinger and M. Traugott, *PLoS One*, 2012, **7**, e29473.
45 S. Letsiou, P. Madesis, E. Vasdekis, C. Montemurro, M. E. Grigoriou, G. Skavdis, V. Moussis, A. E. Koutelidakis and A. G. Tzakos, *Appl. Sci.*, 2024, **14**, 1415.
46 D. M. Spooner, *Am. J. Bot.*, 2009, **96**, 1177–1189.
47 L. Grazina, J. S. Amaral and I. Mafra, *Compr. Rev. Food Sci. Food Saf.*, 2020, **19**, 1080–1109.
48 M. Blumenthal and W. R. Busse, *The complete German Commission E monographs*, American Botanical Council, Austin, TX, 1998.
49 U. S. Pharmacopeia, *Herbal Medicines Compendium*, http://hmc.usp.org/, accessed 13 August 2021.
50 V. E. Tyler, *Herbs of choice: the therapeutic use of phytomedicinals*, Pharmaceutical Products Press, 1994.
51 *Monographs*, https://herbal-ahp.com/collections/frontpage, accessed 25 October 2024.
52 A. Chandra, Y. Li, J. Rana, K. Persons, C. Hyun, S. Shen and T. Mulder, *J. Funct. Foods*, 2011, **3**, 107–114.
53 S. Gafner, *Botanical Adulterants Bulletin*, 2018, pp. 1–12.
54 S. Gafner, M. Blumenthal, S. Foster, J. H. Cardellina II, I. A. Khan and R. Upton, *Acta Hortic.*, 2020, 15–24.
55 N. Baume, N. Mahler, M. Kamber, P. Mangin and M. Saugy, *Scand. J. Med. Sci. Sports*, 2006, **16**, 41–48.
56 H. Geyer, M. K. Parr, K. Koehler, U. Mareck, W. Schänzer and M. Thevis, *J. Mass Spectrom.*, 2008, **43**, 892–902.
57 P. A. Cohen, J. C. Travis, P. H. J. Keizers, P. Deuster and B. J. Venhuis, *Clin. Toxicol.*, 2018, **56**, 421–426.
58 X. Lv, Y. Li, C. Tang, Y. Zhang, J. Zhang and G. Fan, *Pharm. Biol.*, 2016, **54**, 3264–3271.
59 R. Sarkar, N. Chatterjee, N. Shaikh, Z. Khan, B. Avula, I. Khan and K. Banerjee, *Ind. Crops Prod.*, 2023, **200**, 116835.
60 Official Methods of Analysis of AOAC International, *Withanolide Glycosides and Aglycones of Ashwagandha (Withania somnifera)*, AOAC International, Gaithersburg, MD, USA, 22nd edn, 2015.
61 T. J. Smillie and I. A. Khan, *Clin. Pharmacol. Ther.*, 2010, **87**, 175–186.
62 M. Commisso, P. Strazzer, K. Toffali, M. Stocchero and F. Guzzo, *Comput. Struct. Biotechnol. J.*, 2013, **4**, e201301007.
63 O. Fiehn, *Plant Mol. Biol.*, 2002, **48**, 155–171.
64 B. Avula, Y.-H. Wang, G. Isaac, J. Yuk, M. Wrona, K. Yu and I. A. Khan, *Planta Med.*, 2017, **83**, 1297–1308.
65 J. Yuk, K. L. McIntyre, C. Fischer, J. Hicks, K. L. Colson, E. Lui, D. Brown and J. T. Arnason, *Anal. Bioanal. Chem.*, 2013, **405**, 4499–4509.
66 Y. Jiao, Y. Si, L. Li, C. Wang, H. Lin, J. Liu, Y. Liu, J. Liu, P. Li and Z. Li, *J. Mass Spectrom.*, 2021, **56**, e4787.
67 E. D. Wallace, N. H. Oberlies, N. B. Cech and J. J. Kellogg, *Food Chem. Toxicol.*, 2018, **120**, 439–447.
68 L. Grazina, I. Mafra, L. Monaci and J. Amaral, *Compr. Rev. Food Sci. Food Saf.*, 2023, **22**, 3479–4185.
69 S. L. Collins, I. Koo, J. M. Peters, P. B. Smith and A. D. Patterson, *Annu. Rev. Anal. Chem.*, 2021, **14**, 467–487.
70 T. F. Jorge, A. T. Mata and C. António, *Philos. Trans. R. Soc., A*, 2016, **374**, 20150370.
71 E. J. Abraham and J. J. Kellogg, *Front. Nutr.*, 2021, **8**, 780228.
72 S. Ren, A. A. Hinzman, E. L. Kang, R. D. Szczesniak and L. J. Lu, *Metabolomics*, 2015, **11**, 1492–1513.
73 S. Hemmer, S. K. Manier, S. Fischmann, F. Westphal, L. Wagmann and M. R. Meyer, *Metabolites*, 2020, **10**, 378.
74 R. Schmid, S. Heuckeroth, A. Korf, A. Smirnov, O. Myers, T. S. Dyrlund, R. Bushuiev, K. J. Murray, N. Hoffmann, M. Lu, A. Sarvepalli, Z. Zhang, M. Fleischauer, K. Dührkop, M. Wesner, S. J. Hoogstra, E. Rudt,

O. Mokshyna, C. Brungs, K. Ponomarov, L. Mutabdžija, T. Damiani, C. J. Pudney, M. Earll, P. O. Helmer, T. R. Fallon, T. Schulze, A. Rivas-Ubach, A. Bilbao, H. Richter, L.-F. Nothias, M. Wang, M. Orešič, J.-K. Weng, S. Böcker, A. Jeibmann, H. Hayen, U. Karst, P. C. Dorrestein, D. Petras, X. Du and T. Pluskal, *Nat. Biotechnol.*, 2023, **41**, 447–449.

75 A. Klåvus, M. Kokla, S. Noerman, V. M. Koistinen, M. Tuomainen, I. Zarei, T. Meuronen, M. R. Häkkinen, S. Rummukainen, A. Farizah Babu, T. Sallinen, O. Kärkkäinen, J. Paananen, D. Broadhurst, C. Brunius and K. Hanhineva, *Metabolites*, 2020, **10**, 135.

76 R. Arneberg, T. Rajalahti, K. Flikka, F. S. Berven, A. C. Kroksveen, M. Berle, K. M. Myhr, C. A. Vedeler, R. J. Ulvik and O. M. Kvalheim, *Anal. Chem.*, 2007, **79**, 7014–7026.

77 S. Barnes, H. P. Benton, K. Casazza, S. J. Cooper, X. Cui, X. Du, J. Engler, J. H. Kabarowski, S. Li, W. Pathmasiri, J. K. Prasain, M. B. Renfrow and H. K. Tiwari, *J. Mass Spectrom.*, 2016, **51**, 535–548.

78 E. Jackson, in *A Use's Guide to Principal Components*, John Wiley & Sons, Ltd, 1991, pp. 26–62.

79 E. D. Wallace, D. A. Todd, J. M. Harnly, N. B. Cech and J. J. Kellogg, *Anal. Bioanal. Chem.*, 2020, **412**, 4273–4286.

80 N. P. Kalogiouri, R. Aalizadeh, M. E. Dasenaki and N. S. Thomaidis, *Anal. Chim. Acta*, 2020, **1134**, 150–173.

81 U. W. Liebal, A. N. T. Phan, M. Sudhakar, K. Raman and L. M. Blank, *Metabolites*, 2020, **10**, 243.

82 R. G. Brereton and G. R. Lloyd, *J. Chemom.*, 2014, **28**, 213–225.

83 K. Kjeldahl and R. Bro, *J. Chemom.*, 2010, **24**, 558–564.

84 B. Senizza, G. Rocchetti, S. Ghisoni, M. Busconi, M. De Los Mozos Pascual, J. A. Fernandez, L. Lucini and M. Trevisan, *Food Res. Int.*, 2019, **126**, 108584.

85 S. Medina, R. Perestrelo, P. Silva, J. A. M. Pereira and J. S. Câmara, *Trends Food Sci. Technol.*, 2019, **85**, 163–176.

86 B. D. Anderson, D. E. Sepulveda, R. Nachnani, A. Cortez-Resendiz, M. D. Coates, A. Beckett, J. E. Bisanz, J. J. Kellogg and W. M. Raup-Konsavage, *J. Pharmacol. Exp. Ther.*, 2024, **390**, 331–341.

87 K.-A. Lê Cao, S. Boitard and P. Besse, *BMC Bioinf.*, 2011, **12**, 253.

88 A. Foti, G. Musumarra, A. Trovato-salinaro, S. Scire, V. Barresi, C. G. Fortuna, G. Strazzulla, D. F. Condorelli, V. a Doria and S. Chimiche, *J. Chemom.*, 2007, 398–405.

89 J. Boccard and D. N. Rutledge, *Anal. Chim. Acta*, 2013, **769**, 30–39.

90 J. J. Kellogg, D. A. Todd, J. M. Egan, H. A. Raja, N. H. Oberlies, O. M. Kvalheim and N. B. Cech, *J. Nat. Prod.*, 2016, **79**, 376–386.

91 S. Ismail, M. Maulidiani, M. Akhtar, F. Abas, I. Ismail, A. Khatib, N. Ali and K. Shaari, *Molecules*, 2017, **22**, 1612.

92 S. Barbosa, J. Saurina, L. Puignou and O. Núñez, *Foods*, 2020, **9**, 486.

93 A. Walkowiak, Ł. Ledziński, M. Zapadka and B. Kupcewicz, *Spectrochim. Acta, Part A*, 2019, **208**, 222–228.

94 A. Rácz, A. Gere, D. Bajusz and K. Héberger, *RSC Adv.*, 2018, **8**, 10–21.

95 J. Harnly and R. Upton, *J. AOAC Int.*, 2023, qsad137.

96 G. Campmajó, J. Saurina, O. Núñez and S. Sentellas, *Food Chem.*, 2022, **390**, 133141.

97 J. Harnly, *J. AOAC Int.*, 2023, **106**, 1077–1086.

98 L. Breiman, *Mach. Learn.*, 2001, **45**, 5–32.

99 E. Vigneau, P. Courcoux, R. Symoneaux, L. Guérin and A. Villière, *Food Qual. Prefer.*, 2018, **68**, 135–145.

100 V. Deklerck, K. Finch, P. Gasson, J. V. den Bulcke, J. V. Acker, H. Beeckman and E. Espinoza, *Rapid Commun. Mass Spectrom.*, 2017, **31**, 1582–1588.

101 L. Hou, Y. Liu and A. Wei, *Ind. Crops Prod.*, 2019, **134**, 146–153.

102 X. Deng, Z. Liu, Y. Zhan, K. Ni, Y. Zhang, W. Ma, S. Shao, Y. Lv, Y. Yuan and K. M. Rogers, Predictive geographical authentication of green tea with protected designation of origin using a random forest model, *Food Control*, 2020, **107**, 106807.

103 A. Natekin and A. Knoll, *Front. Neurorob.*, 2013, **7**, 21.

104 Y. Wang, N. Vorsa, P. d. B. Harrington and P. Chen, *J. Agric. Food Chem.*, 2018, **66**, 12206–12216.

105 B. Schölkopf, A. Smola and K.-R. Müller, in *Artificial Neural Networks — ICANN'97*, ed. W. Gerstner, A. Germond, M. Hasler and J.-D. Nicoud, Springer, Berlin, Heidelberg, 1997, pp. 583–588.

106 Y. Li, L. Wang, L. Ju, H. Deng, Z. Zhang, Z. Hou, J. Xie, Y. Wang and Y. Zhang, *Toxicol. Sci.*, 2016, **150**, 390–399.

107 H. Pan, C. Yao, S. Yao, W. Yang, W. Wu and D. Guo, *J. Sep. Sci.*, 2020, **43**, 1043–1050.

108 J. Zhang, C. Wang, W. Wu, Q. Jin, J. Wu, L. Yang, Y. An, C. Yao, W. Wei, J. Song, W. Wu and D. Guo, *Arabian J. Chem.*, 2022, **15**, 104118.

109 S. Martín-Torres, A. M. Jiménez-Carvelo, A. González-Casado and L. Cuadros-Rodríguez, *J. Sci. Food Agric.*, 2019, **99**, 4932–4941.

110 Y. Pomyen, K. Wanichthanarak, P. Poungsombat, J. Fahrmann, D. Grapov and S. Khoomrung, *Comput. Struct. Biotechnol. J.*, 2020, **18**, 2818–2825.

111 K. M. Mendez, D. I. Broadhurst and S. N. Reinke, *Metabolomics*, 2019, **15**, 142.

112 C. Peng, Y. Ren, Z. Ye, H. Zhu, X. Liu, X. Chen, R. Hou, D. Granato and H. Cai, *Food Res. Int.*, 2022, **158**, 111512.

113 G. Binetti, L. Del Coco, R. Ragone, S. Zelasco, E. Perri, C. Montemurro, R. Valentini, D. Naso, F. P. Fanizzi and F. P. Schena, *Food Chem.*, 2017, **219**, 131–138.

114 A. J. Tušek, T. Jurina, M. Benković, D. Valinger, A. Belščak-Cvitanović and J. G. Kljusurić, *J. Appl. Res. Med. Aromat. Plants*, 2020, **16**, 100229.

115 X. Xu, W. Li, T. Li, K. Zhang, Q. Song, L. Liu, P. Tu, Y. Wang, Y. Song and J. Li, *Anal. Chem.*, 2020, **92**, 7646–7656.

116 A. Lesiak, R. Cody, A. J. Dane and R. Musah, *Anal. Chem.*, 2015, **87**, 8748–8757.

117 C. Hu, J. Shi, S. Quan, B. Cui, S. Kleessen, Z. Nikoloski, T. Tohge, D. Alexander, L. Guo, H. Lin, J. Wang, X. Cui, J. Rao, Q. Luo, X. Zhao, A. R. Fernie and D. Zhang, *Sci. Rep.*, 2014, **4**, 5067.

118 A. Tsopmo and A. D. Muir, *J. Agric. Food Chem.*, 2010, **58**, 8715–8721.
119 H. Lin, J. Rao, J. Shi, C. Hu, F. Cheng, Z. A. Wilson, D. Zhang and S. Quan, *J. Integr. Plant Biol.*, 2014, **56**, 826–836.
120 K. R. Kiran, P. S. Swathy, B. Paul, K. Shama Prasada, M. Radhakrishna Rao, M. B. Joshi, P. S. Rai, K. Satyamoorthy and A. Muthusamy, *J. Ethnopharmacol.*, 2021, **273**, 113928.
121 N. Guo, Z. Fang, Q. Zang, Y. Yang, T. Nan, Y. Zhao and L. Huang, *J. Ethnopharmacol.*, 2023, **313**, 116546.
122 A. da Silva Antonio, D. Santos Oliveira, G. R. C. dos Santos, H. M. Gualberto Pereira, L. S. Moreira Wiedemann and V. F. da Veiga-Junior, *RSC Adv.*, 2021, **11**, 25096–25103.
123 I. Ware, K. Franke, A. Frolov, K. Bureiko, E. Kysil, M. Yahayu, H. A. El Enshasy and L. A. Wessjohann, *Nat. Prod. Bioprospect.*, 2024, **14**, 30.
124 M. A. Farag, A. H. Abou Zeid, M. A. Hamed, Z. Kandeel, H. M. El-Rafie and R. H. El-Akad, *Nat. Prod. Res.*, 2015, **29**, 116–124.
125 N. N. Magni, A. C. S. Veríssimo, H. Silva and D. C. G. A. Pinto, *Metabolites*, 2023, **13**, 280.
126 C. P. Commission, *Pharmacopoeia of the People's Republic of China: (2015)*, China Medical Science Press, 2015.
127 J. V. Mendes Resende, N. M. D. de Sá, M. T. L. de Oliveira, R. C. Lopes, R. Garrett and R. Moreira Borges, *Phytochem. Lett.*, 2020, **36**, 99–105.
128 I. J. Foutami, T. Mariager, R. Rinnan, C. J. Barnes and N. Rønsted, *Front. Plant Sci.*, 2018, **9**, 1877.
129 D. Badillo-Sanchez, M. Serrano Ruber, A. M. Davies-Barrett, D. J. L. Jones and S. Inskip, *J. Archaeol. Sci.*, 2023, **153**, 105769.
130 A. Afzan, L. Bréant, D. U. Bellstedt, J. R. Grant, E. F. Queiroz, J.-L. Wolfender and J. Kissling, *Taxon*, 2019, **68**, 771–782.
131 N. Orhan, S. Gafner and M. Blumenthal, *Nat. Prod. Rep.*, 2024, **41**, 1604–1621.
132 C. Simmler, J. G. Graham, S.-N. Chen and G. F. Pauli, *Fitoterapia*, 2018, **129**, 401–414.
133 E. D. Wallace, D. A. Todd, J. M. Harnly, N. B. Cech and J. J. Kellogg, *Anal. Bioanal. Chem.*, 2020, **412**, 4273–4286.
134 S. Barbosa, N. Pardo-Mates, M. Hidalgo-Serrano, J. Saurina, L. Puignou and O. Núñez, *J. Agric. Food Chem.*, 2018, **66**, 9353–9365.
135 J. Kellogg, J. Wang, C. Flint, D. Ribnicky, P. Kuhn, E. G. De Mejia, I. Raskin and M. A. Lila, *J. Agric. Food Chem.*, 2010, **58**, 3884–3900.
136 D. Esquivel-Alvarado, E. Alfaro-Viquez, C. G. Krueger, M. M. Vestling and J. D. Reed, *J. AOAC Int.*, 2021, **104**, 223–231.
137 International Standard Organization (ISO), *ISO 3632-2*, https://www.iso.org/standard/44526.html, accessed 22 November, 2024.
138 L. Sabatino, M. Scordino, M. Gargano, A. Belligno, P. Traulo and G. Gagliano, *Nat. Prod. Commun.*, 2011, **6**, 1873–1876.
139 P. Pant, S. Pandey and S. Dall'Acqua, *Chem. Biodiversity*, 2021, **18**, e2100345.
140 C. V. Borges, I. O. Minatel, H. A. Gomez-Gomez and G. P. P. Lima, in *Medicinal Plants and Environmental Challenges*, ed. M. Ghorbanpour and A. Varma, Springer International Publishing, Cham, 2017, pp. 259–277.
141 D. P. Pavarini, S. P. Pavarini, M. Niehues and N. P. Lopes, *Anim. Feed Sci. Technol.*, 2012, **176**, 5–16.
142 L. Lucini, G. Rocchetti and M. Trevisan, *Curr. Opin. Food Sci.*, 2020, **31**, 88–95.
143 M. M. Artêncio, A. L. L. Cassago, R. K. da Silva, J. de Moura Engracia Giraldi and F. B. da Costa, *Rev. Bras. Farmacogn.*, 2023, **33**, 1251–1262.
144 E. M. Mudge, P. N. Brown and S. J. Murch, *Planta Med.*, 2019, **85**, 781–796.
145 A. Rivera-Pérez, R. Romero-González and A. Garrido Frenich, *Food Res. Int.*, 2021, **150**, 110722.
146 A. Rivera-Pérez, R. Romero-González and A. Garrido Frenich, *J. Agric. Food Chem.*, 2021, **69**, 5547–5558.
147 L. M. Barata, E. H. Andrade, A. R. Ramos, O. F. de Lemos, W. N. Setzer, K. G. Byler, J. G. S. Maia and J. K. R. da Silva, *Int. J. Mol. Sci.*, 2021, **22**, 890.
148 Y. Li, C. Zhang, S. Pan, L. Chen, M. Liu, K. Yang, X. Zeng and J. Tian, *LWT–Food Sci. Technol.*, 2020, **117**, 108644.
149 W. Long, S. Wang, C. Hai, H. Chen, H.-W. Gu, X.-L. Yin, J. Yang and H. Fu, *J. Food Compos. Anal.*, 2023, **118**, 105194.
150 A. Karimi, A. Krähmer, N. Herwig, J. Hadian, H. Schulz and T. Meiners, *J. Agric. Food Chem.*, 2020, **68**, 9940–9952.
151 A.-G. Ciocan, V. Tecuceanu, C. Enache-Preoteasa, E. M. Mitoi, F. E. Helepciuc, T. V. Dimov, A. Simon-Gruita and G. C. Cogălniceanu, *Plants*, 2023, **12**, 1762.
152 A. L. L. Cassago, M. M. Artêncio, J. de Moura Engracia Giraldi and F. B. Da Costa, *Eur. Food Res. Technol.*, 2021, **247**, 2143–2159.
153 K. M. Ku, J. N. Choi, J. Kim, J. K. Kim, L. G. Yoo, S. J. Lee, Y.-S. Hong and C. H. Lee, *J. Agric. Food Chem.*, 2010, **58**, 418–426.
154 H. Wang, X. Cao, Z. Yuan and G. Guo, *Food Chem.*, 2021, **352**, 129359.
155 X.-L. Li, X. Yu, J. Lin, X. Zhao, Y. Zhang, H. Lin, Z. Hao and X. Jin, *Food Sci.*, 2020, **41**, 197–203.
156 F. Liu, Z. Tu, L. Chen, J. Lin, H. Zhu and Y. Ye, *J. Sci. Food Agric.*, 2023, **103**, 213–220.
157 T. Wu, R. Zou, D. Pu, Z. Lan and B. Zhao, *BMC Plant Biol.*, 2021, **21**, 55.
158 A. Rivera-Pérez, R. Romero-González and A. Garrido Frenich, *Food Chem.*, 2022, **393**, 133377.
159 A. Rivera-Pérez, P. García-Pérez, R. Romero-González, A. Garrido Frenich and L. Lucini, *Food Res. Int.*, 2022, **162**, 112081.
160 G. Andrea Vitale, C. Geibel, V. Minda, M. Wang, A. T. Aron and D. Petras, *Nat. Prod. Rep.*, 2024, **41**, 885–904.
161 A. I. Calderon, *Comb. Chem. High Throughput Screening*, 2017, **20**(4), 278.
162 J. J. Kellogg, M. F. Paine, J. S. McCune, N. H. Oberlies and N. B. Cech, *Nat. Prod. Rep.*, 2019, **36**, 1196–1221.
163 J. J. Kellogg and N. B. Cech, in *Comprehensive Natural Products III*, Elsevier, 2020, pp. 271–279.
164 Z. Han, M. Wen, H. Zhang, L. Zhang, X. Wan and C.-T. Ho, *Food Chem.*, 2022, **374**, 131796.

165 G. Rocchetti, G. P. Braceschi, L. Odello, T. Bertuzzi, M. Trevisan and L. Lucini, *Metabolomics*, 2020, **16**, 127.
166 Z. Liu, H. Wang, J. Zhang, Q. Chen, W. He, Y. Zhang, Y. Luo, H. Tang, Y. Wang and X. Wang, *Food Chem.*, 2024, **439**, 138072.
167 Z. Sun, W. Zhao, Y. Li, C. Si, X. Sun, Q. Zhong and S. Yang, *Foods*, 2022, **11**, 3248.
168 M. H. Baky, S. N. Shamma, J. Xiao and M. A. Farag, *Food Chem.*, 2022, **383**, 132374.
169 V. Castro-Alves, I. Kalbina, A. Nilsen, M. Aronsson, E. Rosenqvist, M. A. K. Jansen, M. Qian, Å. Öström, T. Hyötyläinen and Å. Strid, *Food Chem.*, 2021, **344**, 128714.
170 R.-C. Liu, R. Li, Y. Wang and Z.-T. Jiang, *J. Food Compos. Anal.*, 2022, **112**, 104673.
171 P. García-Pérez, B. Miras-Moreno, L. Lucini and P. P. Gallego, *Ind. Crops Prod.*, 2021, **163**, 113322.
172 E. Shawky, A. A. Sobhy, D. A. Ghareeb, S. M. Shams Eldin and D. A. Selim, *Ind. Crops Prod.*, 2022, **182**, 114947.
173 X. Zhang, X. Li, M. Su, J. Du, H. Zhou, X. Li and Z. Ye, *Food Res. Int.*, 2020, **137**, 109531.
174 H. M. El-Sayed, D. M. Rasheed, E. A. Mahrous, B. M. Eltanany, Z. M. Goda, L. Pont, F. Benavente and E. Abdel-Sattar, *J. Pharm. Biomed. Anal.*, 2025, **252**, 116512.
175 C. Li, S. Chu, S. Tan, X. Yin, Y. Jiang, X. Dai, X. Gong, X. Fang and D. Tian, *Front. Chem.*, 2021, **9**, 813359.
176 E. Werner, J.-F. Heilier, C. Ducruix, E. Ezan, C. Junot and J.-C. Tabet, *J. Chromatogr. B*, 2008, **871**, 143–163.
177 A. Steckel and G. Schlosser, *Molecules*, 2019, **24**, 611.
178 K. Scheubert, F. Hufsky and S. Böcker, *J. Cheminf.*, 2013, **5**, 12.
179 H. A. Assress, M. G. Ferruzzi and R. S. Lan, *J. Am. Soc. Mass Spectrom.*, 2023, **34**, 1621–1631.
180 A. G. Marshall, G. T. Blakney, T. Chen, N. K. Kaiser, A. M. McKenna, R. P. Rodgers, B. M. Ruddy and F. Xian, *Mass Spectrom.*, 2013, **2**, S0009.
181 R. J. Cotter, *J. Am. Soc. Mass Spectrom.*, 2013, **24**, 657–674.
182 J. Lee, D. J. Tantillo, L.-P. Wang and O. Fiehn, *J. Chem. Inf. Model.*, 2024, **64**, 7470–7487.
183 W. M. A. Niessen, in *Encyclopedia of Spectroscopy and Spectrometry*, ed. J. C. Lindon, G. E. Tranter and D. W. Koppenaal, Academic Press, Oxford, 3rd edn, 2017, pp. 936–941.
184 M.-N. Li, B.-Q. Shen, X. Lu, W. Gao, S.-S. Wen, X. Zhang, H. Yang and P. Li, *J. Chromatogr. A*, 2021, **1657**, 462572.
185 Y. Song, Q. Song, W. Liu, J. Li and P. Tu, *TrAC, Trends Anal. Chem.*, 2023, **160**, 116982.
186 C. Zhang, K. Bieleszová, A. Žukauskaitė, P. Hladík, J. Grúz, O. Novák and K. Doležal, *Anal. Bioanal. Chem.*, 2024, **416**, 125–139.
187 W. Li, X. Yang, B. Chen, D. Zhao, H. Wang, M. Sun, X. Li, X. Xu, J. Liu, S. Wang, Y. Mi, H. Wang and W. Yang, *Arabian J. Chem.*, 2021, **14**, 103409.
188 C. R. de Bruin, M. Hennebelle, J.-P. Vincken and W. J. C. de Bruijn, *Anal. Chim. Acta*, 2023, **1244**, 340774.
189 Z. Liu, M. Zhang, P. Chen, J. M. Harnly and J. Sun, *J. Agric. Food Chem.*, 2022, **70**, 11138–11153.
190 Y. Niu, J. Liu, R. Yang, J. Zhang and B. Shao, *TrAC, Trends Anal. Chem.*, 2020, **132**, 116053.
191 Z. Zuo, L. Cao, L.-F. Nothia and H. Mohimani, *Bioinformatics*, 2021, **37**, i231–i236.
192 M. Ramos, V. Camel, E. Le Roux, S. Farah and M. Cladiere, *Anal. Bioanal. Chem.*, 2025, **417**, 311–321.
193 T. N. Clark, J. Houriet, W. S. Vidar, J. J. Kellogg, D. A. Todd, N. B. Cech and R. G. Linington, *J. Nat. Prod.*, 2021, **84**, 824–835.
194 C. Hoang, W. Uritboonthai, L. Hoang, E. M. Billings, A. Aisporna, F. A. Nia, R. J. E. Derks, J. R. Williamson, M. Giera and G. Siuzdak, *Anal. Chem.*, 2024, **96**, 5478–5488.
195 Á. López-López, Á. López-Gonzálvez, T. C. Barker-Tejeda and C. Barbas, *Expert Rev. Mol. Diagn.*, 2018, **18**, 557–575.
196 S. Li, N. Looby, V. Chandran and V. Kulasingam, *Metabolites*, 2024, **14**, 200.
197 T. Kind, H. Tsugawa, T. Cajka, Y. Ma, Z. Lai, S. S. Mehta, G. Wohlgemuth, D. K. Barupal, M. R. Showalter, M. Arita and O. Fiehn, *Mass Spectrom. Rev.*, 2018, **37**, 513–532.
198 D. F. McGlynn, L. D. Yee, H. M. Garraffo, L. Y. Geer, T. D. Mak, Y. A. Mirokhin, D. V. Tchekhovskoi, C. N. Jen, A. H. Goldstein, A. J. Kearsley and S. E. Stein, *J. Am. Soc. Mass Spectrom.*, 2025, **36**, 389–399.
199 D. S. Wishart, A. Guo, E. Oler, F. Wang, A. Anjum, H. Peters, R. Dizon, Z. Sayeeda, S. Tian, B. L. Lee, M. Berjanskii, R. Mah, M. Yamamoto, J. Jovel, C. Torres-Calzada, M. Hiebert-Giesbrecht, V. W. Lui, D. Varshavi, D. Varshavi, D. Allen, D. Arndt, N. Khetarpal, A. Sivakumaran, K. Harford, S. Sanford, K. Yee, X. Cao, Z. Budinski, J. Liigand, L. Zhang, J. Zheng, R. Mandal, N. Karu, M. Dambrova, H. B. Schiöth, R. Greiner and V. Gautam, *Nucleic Acids Res.*, 2021, **50**, D622–D631.
200 H. Horai, M. Arita, S. Kanaya, Y. Nihei, T. Ikeda, K. Suwa, Y. Ojima, K. Tanaka, S. Tanaka, K. Aoshima, Y. Oda, Y. Kakazu, M. Kusano, T. Tohge, F. Matsuda, Y. Sawada, M. Y. Hirai, H. Nakanishi, K. Ikeda, N. Akimoto, T. Maoka, H. Takahashi, T. Ara, N. Sakurai, H. Suzuki, D. Shibata, S. Neumann, T. Iida, K. Tanaka, K. Funatsu, F. Matsuura, T. Soga, R. Taguchi, K. Saito and T. Nishioka, *J. Mass Spectrom.*, 2010, **45**, 703–714.
201 M. Wang, J. J. Carver, V. V. Phelan, L. M. Sanchez, N. Garg, Y. Peng, D. D. Nguyen, J. Watrous, C. A. Kapono, T. Luzzatto-Knaan, C. Porto, A. Bouslimani, A. V. Melnik, M. J. Meehan, W. T. Liu, M. Crüsemann, P. D. Boudreau, E. Esquenazi, M. Sandoval-Calderón, R. D. Kersten, L. A. Pace, R. A. Quinn, K. R. Duncan, C. C. Hsu, D. J. Floros, R. G. Gavilan, K. Kleigrewe, T. Northen, R. J. Dutton, D. Parrot, E. E. Carlson, B. Aigle, C. F. Michelsen, L. Jelsbak, C. Sohlenkamp, P. Pevzner, A. Edlund, J. McLean, J. Piel, B. T. Murphy, L. Gerwick, C. C. Liaw, Y. L. Yang, H. U. Humpf, M. Maansson, R. A. Keyzers, A. C. Sims, A. R. Johnson, A. M. Sidebottom, B. E. Sedio, A. Klitgaard, C. B. Larson, C. A. P. Boya, D. Torres-Mendoza, D. J. Gonzalez, D. B. Silva, L. M. Marques, D. P. Demarque, E. Pociute, E. C. O'Neill, E. Briand, E. J. N. Helfrich, E. A. Granatosky, E. Glukhov, F. Ryffel, H. Houson, H. Mohimani,

J. J. Kharbush, Y. Zeng, J. A. Vorholt, K. L. Kurita, P. Charusanti, K. L. McPhail, K. F. Nielsen, L. Vuong, M. Elfeki, M. F. Traxler, N. Engene, N. Koyama, O. B. Vining, R. Baric, R. R. Silva, S. J. Mascuch, S. Tomasi, S. Jenkins, V. Macherla, T. Hoffman, V. Agarwal, P. G. Williams, J. Dai, R. Neupane, J. Gurr, A. M. C. Rodríguez, A. Lamsa, C. Zhang, K. Dorrestein, B. M. Duggan, J. Almaliti, P. M. Allard, P. Phapale, L. F. Nothias, T. Alexandrov, M. Litaudon, J. L. Wolfender, J. E. Kyle, T. O. Metz, T. Peryea, D. T. Nguyen, D. VanLeer, P. Shinn, A. Jadhav, R. Müller, K. M. Waters, W. Shi, X. Liu, L. Zhang, R. Knight, P. R. Jensen, B. Palsson, K. Pogliano, R. G. Linington, M. Gutiérrez, N. P. Lopes, W. H. Gerwick, B. S. Moore, P. C. Dorrestein and N. Bandeira, *Nat. Biotechnol.*, 2016, **34**, 828–837.

202 T. Cajka and O. Fiehn, in *Lipidomics: Methods and Protocols*, ed. S. K. Bhattacharya, Springer, New York, NY, 2017, pp. 149–170.

203 J. R. Montenegro-Burke, C. Guijas and G. Siuzdak, in *Computational Methods and Data Analysis for Metabolomics*, ed. S. Li, Springer US, New York, NY, 2020, pp. 149–163.

204 C. A. Smith, G. O. Maille, E. J. Want, C. Qin, S. A. Trauger, T. R. Brandon, D. E. Custodio, R. Abagyan and G. Siuzdak, *Ther. Drug Monit.*, 2005, **27**, 747–751.

205 G. Alvarez-Rivera, D. Ballesteros-Vivas, F. Parada-Alfonso, E. Ibañez and A. Cifuentes, *TrAC, Trends Anal. Chem.*, 2019, **112**, 87–101.

206 N. F. de Jonge, K. Mildau, D. Meijer, J. J. R. Louwen, C. Bueschl, F. Huber and J. J. J. van der Hooft, *Metabolomics*, 2022, **18**, 103.

207 L. Chen, W. Lu, L. Wang, X. Xing, Z. Chen, X. Teng, X. Zeng, A. D. Muscarella, Y. Shen, A. Cowan, M. R. McReynolds, B. J. Kennedy, A. M. Lato, S. R. Campagna, M. Singh and J. D. Rabinowitz, *Nat. Methods*, 2021, **18**, 1377–1385.

208 K. Dührkop, M. Fleischauer, M. Ludwig, A. A. Aksenov, A. V. Melnik, M. Meusel, P. C. Dorrestein, J. Rousu and S. Böcker, *Nat. Methods*, 2019, **16**, 299–302.

209 J. A. van Santen, E. F. Poynton, D. Iskakova, E. McMann, T. A. Alsup, T. N. Clark, C. H. Fergusson, D. P. Fewer, A. H. Hughes, C. A. McCadden, J. Parra, S. Soldatou, J. D. Rudolf, E. M.-L. Janssen, K. R. Duncan and R. G. Linington, *Nucleic Acids Res.*, 2022, **50**, D1317–D1323.

210 V. M. Anderson, M. M. Ranaweera, A. K. Jarmusch, A. E. Shay, D. A. Todd, N. B. Cech and J. J. Kellogg, *J. Am. Soc. Mass Spectrom.*, 2025, **36**, 926–929.

211 C. Wang, B. Zhang, I. Timári, Á. Somogyi, D. W. Li, H. E. Adcox, J. S. Gunn, L. Bruschweiler-Li and R. Brüschweiler, *Anal. Chem.*, 2019, **91**, 15686–15693.

212 R. Upton, B. David, S. Gafner and S. Glasl, *Phytochem. Rev.*, 2020, **19**, 1157–1177.

213 S. Hosbas Coskun, S. A. Wise and A. J. Kuszak, *Front. Nutr.*, 2021, **8**, 786261.

214 Y. Wang, Z.-T. Zuo, H.-Y. Huang and Y.-Z. Wang, *R. Soc. Open Sci.*, **6**, 190399.

215 S. F. Ullrich, A. Rothauer, H. Hagels and O. Kayser, *Planta Med.*, 2017, **83**, 937–945.

216 Y. Liu, B. Patra, S. K. Singh, P. Paul, Y. Zhou, Y. Li, Y. Wang, S. Pattanaik and L. Yuan, *Biotechnol. Lett.*, 2021, **43**, 2085–2103.

217 Y. Ding, D. M. Gardiner, J. J. Powell, M. L. Colgrave, R. F. Park and K. Kazan, *Plant, Cell Environ.*, 2021, **44**, 3756–3774.

218 E. J. Abraham, S. J. Chamberlain, W. H. Perera, R. T. Jordan and J. J. Kellogg, *Anal. Bioanal. Chem.*, 2025, **417**, 1479–1495.

219 M.-G. Buț, G. Jîtcă, S. Imre, C. E. Vari, B. E. Őcz, C.-M. Jîtcă and A. Tero-Vescan, *Plants*, 2023, **12**, 1722.

220 N. Kfoury, E. R. Scott, C. M. Orians, S. Ahmed, S. B. Cash, T. Griffin, C. Matyas, J. R. Stepp, W. Han, D. Xue, C. Long and A. Robbat, *Front. Plant Sci.*, 2019, **10**, 1518.

221 Y. Tong, X. Li, J. Wan, Q. Zhou, C. Jiang, N. Li, Z. Jin, J. Gu, F. Li and J. Li, *J. Sep. Sci.*, 2025, **48**, e70072.

