Invited Review

Recent applications of metabolomics in plant breeding

Nozomu Sakurai*

Bioinformation and DDBJ Center, National Institute of Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan

Metabolites play a central role in maintaining organismal life and in defining crop phenotypes, such as nutritional value, fragrance, color, and stress resistance. Among the 'omes' in biology, the metabolome is the closest to the phenotype. Consequently, metabolomics has been applied to crop improvement as method for monitoring changes in chemical compositions, clarifying the mechanisms underlying cellular functions, discovering markers and diagnostics, and phenotyping for mQTL, mGWAS, and metabolite-genome predictions. In this review, 359 reports of the most recent applications of metabolomics to plant breeding-related studies were examined. In addition to the major crops, more than 160 other crops including rare medicinal plants were considered. One bottleneck associated with using metabolomics is the wide array of instruments that are used to obtain data and the ambiguity associated with metabolite identification and quantification. To further the application of metabolomics to plant breeding, the features and perspectives of the technology are discussed.

Key Words: metabolomics, mass spectrometry, nuclear magnetic resonance, mQTL, mGWAS.

Introduction

The metabolites of agricultural and horticultural crops have a direct effect on their phenotypes through their influence on stress tolerance, inter-organismal interactions, color, taste, nutritional value, and shelf life. Metabolites can generally be divided into two groups; primary metabolites, which are essential for maintaining the fundamental life processes of the organism, and secondary (specialized) metabolites, which contribute to processes that are specialized to each organism. Together, the primary and secondary metabolites constitute the metabolome, which, among the 'omes' in biology (i.e., genome, transcriptome, and proteome) is considered to reflect the phenotype most closely (Guijas *et al.* 2018, Patti *et al.* 2012).

Metabolomics—technology for the comprehensive detection of small molecules in the samples (Hall 2006)—have been applied to numerous research fields, including plant science, zoology, clinical science, microbiology, environmental science, and plant breeding. Metabolomics has been used to clarify the chemical diversity between different crops, analyze stress response mechanisms, discover biomarkers for recognizing different genotypes and phenotypes, and assess material quality. Integrating metabolome data into genome and transcriptome data can be used to identify genes involved in the biosynthesis or degradation

of specific metabolites and quantitative trait loci (mQTL), and to perform metabolic genome-wide association studies (mGWAS). In recent years, the data obtained from these studies have been used for predicting phenotypes (metabolite genomic selection/prediction).

This review examined recent publications on the application of metabolomics to plant breeding. In most of these studies, a variety of analytical platforms were used, primarily because of a lack of a single technology that can detect and quantify the full spectrum of metabolites, i.e., the complete metabolome, produced by plants. Consequently, depending on the purpose of the study, researchers have to select the most optimal platforms by considering the types of chemicals to be analyzed, the required quantification accuracy, and the availability of the platforms. Therefore, the second part of this review provides and overview of the technical features of metabolome analysis with a brief introduction of the instruments and relevant technical points. Finally, future perspectives for the further application of metabolomics in this field are discussed. It is hoped that this information can be used as a guide to further the application of metabolomics in plant breeding.

Numerous excellent reviews on topics that were not covered in great depth in this review have been published recently; for example: sample preparation and data analysis for association mapping (Alseekh and Fernie 2021); mechanism analysis and phenotype predictions (Fernandez *et al.* 2021); integration with other omics data (Scossa *et al.* 2021); a comprehensive review of relevant publications (Razzaq *et al.* 2019, Sharma *et al.* 2021); safety assessment of genetically modified organisms (Bedair and Glenn 2020); elucidating and planning plant domestication (Alseekh

et al. 2021, Fernie and Yan 2019); and computational tools (Misra 2021). The reader is urged to refer to these studies in detail.

Application of metabolomics to plant breeding

A literature search of all studies published in the last two years (from January 2019 to July 2021) on the application of metabolomics to plant breeding (described later in detail) produced a total of 359 articles excluding reviews (**Supplemental Table 1**). The table also contains information on target plant, experimental factor, purpose of the study, instrument used, and links to the original paper to facilitate access to the article.

The literature search was conducted using PubMed, PubMed Central (Sayers *et al.* 2021) and European PubMed Central (Ferguson *et al.* 2021). All of the publications containing two metabolomics-related terms (i.e., metabolome or metabolomics) and ten plant breeding-related terms (i.e., breeding, crop improvement, quantitative trait loci, genome-wide association study, genomic selection, genomic prediction, metabolomic selection, phenotyping, marker, or biomarker) were searched for, with consideration given to inflection of the terms. Redundant results, and results containing animal- and clinical-specific terms, such as "cerebrospinal" were omitted, which gave 601 articles. The title, abstract, and original paper (available for 391 articles) were then checked manually, which left 359 original research articles in which metabolomics was applied to plant breeding. Articles indirectly related to plant breeding, such as studies for biological mechanisms using model plants and marker discovery for the geographical origin of traditional Chinese medicine, are included for use case purposes. Studies that did not use either a mass spectrometer (MS) or a nuclear magnetic resonance (NMR) spectrometer for chemical detection were omitted from the analysis.

Target plants

Metabolomics was applied to a wide variety of plants (**Table 1**). The 359 articles that were selected examined 160 plants or plant (sample) categories (hereafter referred to as "classes"); of these classes, only seven classes of major crops (including rice, tomato, maize, wheat, and tea tree) were reported in more than 10 publications. Eighteen plant classes were reported in more than 1% of articles (four publications), and 48% of all articles were related to these 18 plant classes. The remainder of the articles (52%) were related to 132 other classes, and 109 of those classes were reported in a single article. A rare medicinal plant *Ferula asafoetida* indigenous to Kashmir, Afghanistan, and Iran was included in the last criterion (Amini *et al.* 2019). As exemplified by coriander (Song *et al.* 2020) and *Medicago polymorpha* (Cui *et al.* 2021), metabolic profiling was conducted in reports on the genome sequences of these species.

Table 1. Plants or plant classes used in breeding-related metabolome analysis that were reviewed in this study (published from Jan 2019 to July 2021)

Plant (classes)	# Articles
Rice	24 (7%)
Tomato	22 (6%)
Maize	18 (5%)
Wheat	18 (5%)
Tea	13 (4%)
Brassicaceae except Arabidopsis	10 (3%)
Traditional Chinese medicine	10 (3%)
Poplar	8 (2%)
Soybean	8 (2%)
Grapevine	6 (2%)
Arabidopsis	5 (1%)
Barley	5 (1%)
Cassava	5 (1%)
Oat	5 (1%)
Sorghum	5 (1%)
Apple	4 (1%)
Medicago	4 (1%)
Olive	4 (1%)
Others	185 (52%)
Total	359

Purposes of the studies

Of the 359 articles, 166 (46%) analyzed metabolic mechanisms, 117 (33%) examined metabolic profiling, 58 (16%) examined marker discovery or discrimination, 18 (5%) examined mGWAS, 14 (4%) examined mQTL, 9 (3%) examined genomic/metabolic selection/prediction, and 25 (7%) examined other purposes. Many of the papers on mechanism analysis focused on the resistance mechanisms of biotic (insects and pathogens) and abiotic (salinity, temperature, and drought) stresses. The metabolic profiling studies included preliminary and descriptive findings for further detailed mechanism analysis and marker discovery. The genomic/metabolic prediction studies were conducted on major crops, such as rice, oat, barley, and wheat, as well as on less well known crops, such as blueberry (Ferrão *et al.* 2021). In addition, mGWAS was applied to hops (Feiner *et al.* 2021) and coffee (Gamboa-Becerra *et al.* 2019), and mQTL was applied to *Brachypodium distachyon* (Onda *et al.* 2019) and blackcurrant (Abreu *et al.* 2020).

In an important study on the mechanism of stress response, Wang *et al.* (2021) identified the fluorescence in blue light (BrFLU) gene in Pak Choi; the gene was shown to be related to cold acclimation in a study that combined transcriptomics methods.

As an example of metabolic profiling, Qing *et al.* (2021) demonstrated the absence of genes and metabolites in the biosynthesis of colchicine, which is concerned about a factor for the limited market of *Hemerocallis citrina* Borani. Yu *et al.* (2020) elucidated metabolic profiles of 136 accessions of tea trees as part of a study on characterizing tea

genetic resources in China. In a study on marker discovery, Lemaitre-Guillier *et al.* (2021) demonstrated that the profiles of volatile organic compounds (VOCs) can be used as diagnostic markers for stress in grapevines.

Metabolomics was also used to depict the metabolic changes during the breeding and domestication processes of the crops. Schouten *et al.* (2019) conducted metabolic profiling of tomato cultivars in the Netherlands and concluded that increased diversity in the composition of aromatic volatiles from the 1990s onwards might reflect the efforts of breeders trying to improve fruit quality. Although not included in **Supplemental Table 1**, the use of metabolomics to clarify the domestication process has been reported in tomato (Zhu *et al.* 2018) and wheat (Batyrshina *et al.* 2020, Beleggia *et al.* 2016).

The use of metabolomics as an alternative to assessing the phenotype of a plant using organoleptic tests, which are difficult to standardize and perform at scale, was described by Ferrão *et al.* (2021). The relative scarcity of MS- and NMR-based metabolomics studies for large-scale phenotyping using other methods, such as GWAS and genomic predictions, is likely due to the lack of portable instruments. As a simple and portable phenotyping tool, near-infrared (NIR) spectrometry was used recently (Alves Filho *et al.* 2019, Lane *et al.* 2020, Rincent *et al.* 2018).

The stability of metabolome between crops grown under different conditions is often argued when the metabolomic data are subjected to statistical analysis. Chevalier *et al.* (2021) evaluated phenotypic plasticity in carrot varieties and concluded that a balance between constitutive content and the environmental sensitivity of key metabolites should be considered for quality improvement in carrot and other vegetables. For the assessments of the unintended effects of genetically modified crops, reports have been published on maize varieties (Liu *et al.* 2021) and tomato rootstock (Kodama *et al.* 2021).

Instruments and equipment used in metabolomics studies

In the 359 reviewed articles, liquid chromatography (LC)-mass spectrometry (MS) was the most commonly used instrument (239 articles, 67%), followed by gas chromatography (GC)-MS (104, 29%), nuclear magnetic resonance (NMR) spectrometry (22, 6%), capillary electrophoresis (CE)-MS (2, 1%), and others (19, 5%). LC-MS was often used using established methods to detect polyphenols such as flavonoids related to antioxidant activity and color production. GC-MS was used to obtain an overview of the metabolites produced by primary metabolism (sugars, amino acids, fatty acids, and organic acids), and also to detect flavor compounds. In numerous studies, multi-instrument platforms are used. Moing *et al.* (2020) used the widest variety of methods including LC-MS, GC-MS, NMR, and flow injection-MS (see section "Overview of metabolomics technology").
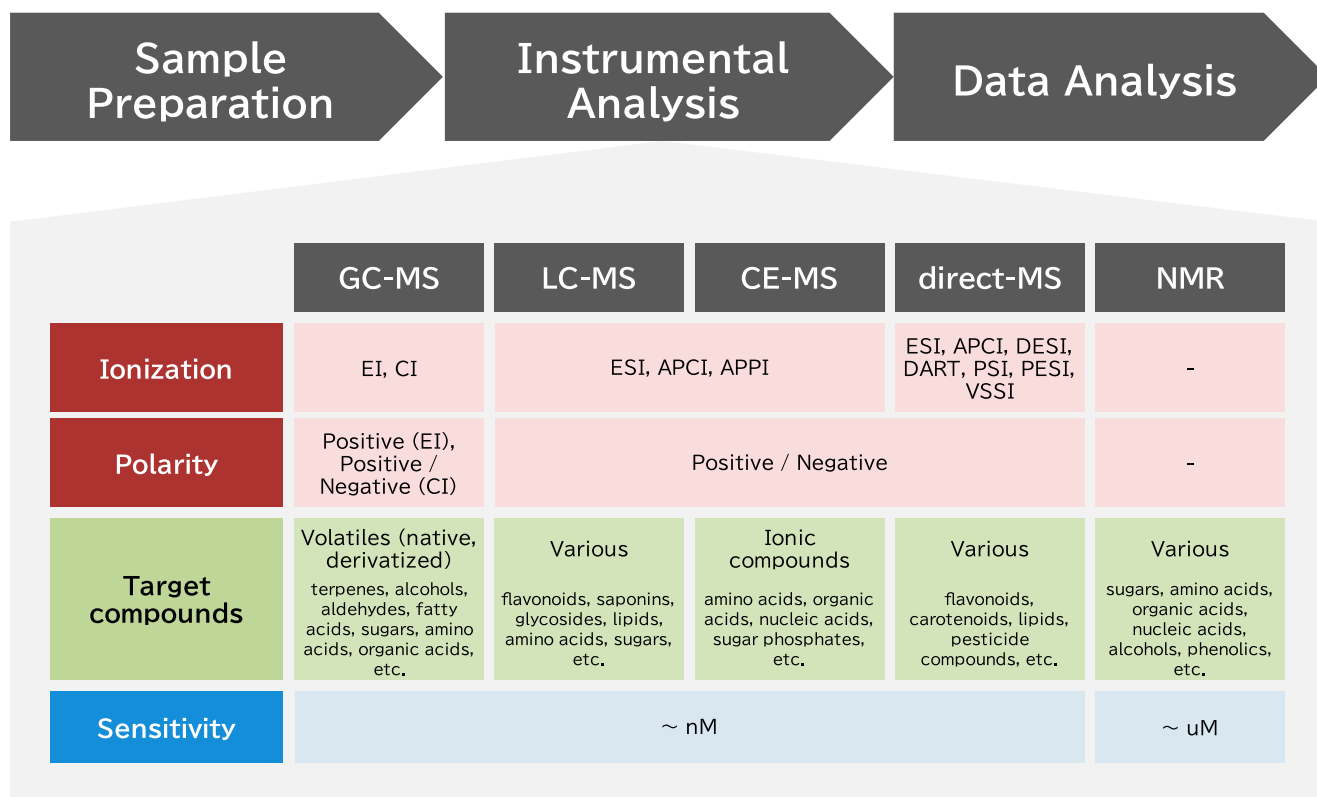
Overview of metabolomics technology

Compared to transcriptome and proteome analysis, the characteristics of metabolome analysis are as follows: 1) a large variety of analytical instruments is used (**Fig. 1**), primarily because no single method is capable of covering the wide range of physicochemical properties and concentrations of metabolites typically found in samples (Saito and Matsuda 2010). 2) In terms of identifying metabolites, there is often some ambiguity in the obtained results (Viant *et al.* 2017). This is because a comparison of metabolite signals with those from the authentic standard chemical using the same instrumental conditions is required for metabolite identification. Therefore, the metabolomics data may contain known-, predicted- and unknown metabolites. A clear indication of the certainty of the annotation (level 1 to 4) in the results is strongly recommended by the Metabolomics Standards Initiative (MSI) (Sumner *et al.* 2007). 3) There is typically a difference in quantitative accuracy between the data. The quantitative accuracy of metabolite measurements depends on a variety of factors, such as the concentrations in the sample, pretreatment (extraction and concentration), analytical methods, and data processing methods. As a result of these intrinsic features of the technology, metabolome analysis can generally be divided into two categories: targeted analysis, in which specific metabolites are sought with high quantitative accuracy, and untargeted analysis, which tries to comprehensively detect undefined metabolites including unknowns (Gertsman and Barshop 2018). Combinations of targeted and untargeted metabolome analysis and other methods that are used to measure specific compounds can be used in a study. Therefore, selecting the most appropriate analytical method depending on the purpose of the study is necessary for robust metabolome analysis.

This section briefly describes the features of the instruments and some technical considerations for metabolomics experiments, such as sampling and data analysis. Although lipids are one of the largest chemical groups in organisms, this section omits the details of lipid-focused metabolomics (lipidomics). Lipids have multiple functions, and are involved in cellular compartmentation, energy storage, and signaling. The reader is urged to refer to the review by Zullig *et al.* (2020).

Analytical instruments

It is estimated that more than 1 million metabolites are produced in the plant kingdom (Afendi *et al.* 2012). Two of the most commonly used instruments in metabolomics, mainly because of the wide range of chemicals that can be assayed, are the mass spectrometer (MS), which measures the weight of the ionized molecule, and the nuclear magnetic resonance (NMR) spectrometer, which identifies features of the chemical structure of the molecule being assayed. As the mass measurement in MS is based on the

	GC-MS	LC-MS	CE-MS	direct-MS	NMR
Ionization	EI, CI	ESI, APCI, APPI		ESI, APCI, DESI, DART, PSI, PESI, VSSI	-
Polarity	Positive (EI), Positive / Negative (CI)	Positive / Negative			-
Target compounds	Volatiles (native, derivatized) terpenes, alcohols, aldehydes, fatty acids, sugars, amino acids, organic acids, etc.	Various flavonoids, saponins, glycosides, lipids, amino acids, sugars, etc.	Ionic compounds amino acids, organic acids, nucleic acids, sugar phosphates, etc.	Various flavonoids, carotenoids, lipids, pesticide compounds, etc.	Various sugars, amino acids, organic acids, nucleic acids, alcohols, phenolics, etc.
Sensitivity	~ nM				~ uM

Fig. 1. Instruments used for metabolome analysis.

motility of charged ions, an ionization unit (ion source) is placed in front of the MS. Several ionization methods are used. To separate the isomers and isobars (molecules sharing the same mass number) which cannot be distinguished based on mass values, a separation apparatus is coupled to the MS apparatus; for example, LC-MS, GC-MS, and CE-MS are often used.

Liquid chromatography-mass spectrometry (LC-MS)

A wide variety of metabolites can be separated and detected by selecting different column types and solvent conditions. Reversed-phase chromatography with a C_{18} column is often used to detect secondary metabolites with a low to moderate polarity, such as flavonoids, saponins, their glycosides, and polar lipids. Hydrophilic interaction chromatography (HILIC) is also used to better separate amino acids, sugars, and lipid classes. Electrospray ionization (ESI) is the most commonly used, as are atmospheric pressure chemical ionization (APCI) (Hao *et al.* 2020) and atmospheric pressure photoionization (APPI). By changing the voltage polarity, the molecules in a sample can be ionized and detected as their cation (in the positive mode) or anion (in the negative mode).

In addition to measuring the mass of the intact ionized molecule, the mass spectrum of the fragmented ion can be obtained using a modern LC-MS. Indeed, not only can 1) the ion mass be measured, but the mass spectrometer can also be used to 2) select ions with a specific mass value, and 3) fragment the molecular ions via collisions with an inactive gas (typically nitrogen) in a process referred to as collision-induced dissociation (CID). By combining these three functions of MS, the mass spectra of the fragmented ions of the selected precursor ion can be measured. The triple quadrupole-type MS (QqQ and Triple-Q) comprises three MS systems that are connected in tandem, where an ion is selected by the first MS, fragmented in the second MS, and the mass spectra are recorded in the third MS (MS/MS analysis). The Q-ToF-MS and Q-Exactive® have high-resolution MS, namely, time-of-flight (ToF)-MS and Orbitrap®-MS, respectively, as the third MS. The ion trap (IT)-type MS can repeat multiple cycles of the selection, fragmentation, and measurement processes (multiple-stage MS analysis). An apparatus combining IT-MS with high-resolution MS (IT-ToF-MS and Fusion Tribrid®) is also available. The mass spectrum provides information for the identification, prediction, and discrimination of the metabolites. Fragmentation is also used to quantify the metabolites with high selectivity and sensitivity by selecting the specific precursor ion and measuring the metabolite-specific fragment ion. These approaches are referred to as selected reaction monitoring (SRM), multiple reaction monitoring (MRM), and parallel reaction monitoring (PRM).

LC-MS can be used to measure a wide range of chemicals, as mentioned above, but the variability of the experimental conditions often makes it difficult to compare the results across studies. In addition, special procedures such as the addition of stable isotope-labeled internal controls is

required for highly accurate quantification of the metabolites in a crude sample, because ionization rate can be affected by co-existing molecules that are ionized preferentially. This phenomenon is called "ion suppression" or "matrix effect".

Gas chromatography-mass spectrometry (GC-MS)

GC-MS is widely used for the analysis of volatiles, such as terpenes, alcohols, aldehydes, and fatty acid esters. The metabolites that can be converted to volatile compounds by silylation and esterification (sugars, amino acids, organic acids, and fatty acids) can also be detected (Beale *et al.* 2018). An electron ionization (EI) setting of 70 eV is typically used. Under this condition, the molecule is fragmented during ionization and the mass spectrum is measured as the primary information of the metabolite signal. The fragmentation that occurs due to EI differs from that which occurs in CID (see section on LC-MS) and is more reproducible. Furthermore, the retention time of the metabolites can be standardized by indexing them based on the retention times of alkanes or fatty acid methyl esters. Taken together, these features make it easy to compare the data between studies. Numerous mass spectral libraries are available for compound identification and annotation, including the mass spectral library provided by National Institute of Standards and Technology (NIST, https://chemdata.nist.gov/), Wiley Registry of Mass Spectral Data (Wiley & Sons Inc.), the Golm Metabolome Database (Kopka *et al.* 2005), the library developed by FihenLib (Kind *et al.* 2009), FFNSC 3 for fragrances (Mondello 2015), as well as libraries for pesticides, pollutants, and forensic purposes. As an alternative to MS, a flame ionization detector (FID) is often used. Three studies that used an FID are given in **Supplemental Table 1** (Dar *et al.* 2020, Erzen *et al.* 2021, Klevorn *et al.* 2019).

Capillary electrophoresis-mass spectrometry (CE-MS)

CE-MS, which was first applied to metabolomics by Soga *et al.* (2003), is well suited for the detection of hydrophilic ionic metabolites, such as amino acids, organic acids, nucleic acids, and sugar phosphates, i.e., most of the primary metabolites in cells (Buko 2017, Zhang and Ramautar 2021). While ESI is usually used for this purpose, APCI and APPI are also applicable (Hommerson *et al.* 2009a, 2009b).

Nuclear magnetic resonance (NMR) spectroscopy

NMR is advantageous for high throughput targeted analysis because the sample is measured in an enclosed glass vial which prevents contamination of the detector. In addition, sample preparation is straightforward as there is no need for extraction and pretreatment, and samples can be recovered after measurements and used for other experimental applications. Although ^{1}H-NMR is generally used for targeted analysis, NMR can also be used for untargeted analyses. The chemical structure of unknown metabolites can be directly elucidated or annotated by two-dimensional NMR, which is a very important advantage of NMR over MS (Emwas *et al.* 2019, Schripsema 2010). However, NMR requires a higher concentration of target metabolite

(~µM), which is disadvantageous when compared to MS (~nM). Nonetheless, quantitative accuracy is high (Pinu *et al.* 2019, Razzaq *et al.* 2019).

Other methods
Direct injection methods

Samples can be ionized directly and MS measurements can be performed without separating them by GC, LC, or CE. The "direct infusion" method typically uses a single syringe pump for injection of the sample. The "flow injection" method uses an autosampler and the pump systems of the LC without a column. The advantage of these direct methods is the short time required for analysis which makes this method well suited to high throughput analysis. The major disadvantages include the lower sensitivity, difficulties with elucidating the mass spectral data obtained from the mixture of metabolites in the samples, and the ion suppression unavoidable for the mixture. To overcome these issues, a spectral stitching method in which each mass scan is first separated into smaller mass ranges, which are then stitched together (Sarvin *et al.* 2020, Southam *et al.* 2016). Three examples of the use of direct injection-MS are provided in **Supplemental Table 1**: tomato fertilized with different nitrogen sources (Garcia-Casarrubias *et al.* 2019), metabolic profiling of melon (Moing *et al.* 2020), and polyphenol analysis of sorghum (Hodges *et al.* 2021)

Ambient ionization

Ambient ionization is a method that is used for direct ionization of the chemicals on the surface of an object under atmospheric pressure, such as chemicals on the leaves of a plant (Feider *et al.* 2019). Numerous methods based on a variety of principles have been reported to date, including solvent-based desorption electrospray ionization (DESI) and plasma-based direct analysis in real-time (DART) which is based on the same principle as APCI. These ionization techniques are advantageous for real-time monitoring of chemicals on a small area of an object's surface without pretreatment. These features of ambient ionization methods make them well suited to mass spectrometry imaging. The reader is urged to refer to other reviews on mass spectrometry imaging (Ferguson *et al.* 2019, Hu *et al.* 2021) as a description of these methods is beyond the scope of this review.

Ion mobility

This technique is well suited for high-throughput identification or discrimination of isomers or isobars which can be separated based on their molecular shape (collision cross section, CCS). Numerous MS devices are equipped with an ion mobility unit and CCS libraries of metabolites have recently been constructed (Dodds and Baker 2019, Zhou *et al.* 2020).

Supercritical fluid chromatography (SFC)

SFC is a separation technique that uses supercritical fluids with high diffusivity and low viscosity (Bamba *et al.* 2008, Gordillo 2021, van de Velde *et al.* 2020). These features enable a high-speed and high-resolution separation

using a long column. SFC is considered to be a prospective method for the comprehensive analysis of lipids.

Sample preparation
Sampling

Given the lack of portable instruments for MS- or NMR-based metabolomics analysis in the field, destructive sampling methods are usually necessary. In the case of NMR, non-destructive sampling can be performed if individuals can fit into sample vials. The amount of sample required for the analysis depends on the concentration of the target metabolites, pretreatment methods, and analytical methods, but it typically ranges from 10 to 100 mg fresh weight. Since the metabolome is sensitive to environmental factors, for large-scale sampling such as that required for mGWAS studies, controls for uniform sampling in a short time are required (Gemmer *et al.* 2020, 2021, Song *et al.* 2021, Yao *et al.* 2021). Immediately after sampling, a quenching procedure such as freezing in liquid nitrogen and extraction with organic solvents should be performed in order to avoid chemical changes by enzymatic and chemical reactions.

Metabolite extraction

Metabolite extraction should be performed appropriately for the metabolites targeted for the analysis. The major purpose of extraction is to improve the sensitivity of the analysis by concentrating the targeted metabolites and excluding contaminants, to protect instruments, to stop enzymatic reactions, and to remove proteins. Frequently used methods include the following: solid-phase microextraction (SPME) for volatiles (GC-MS); using a mixture of water, methanol, and chloroform for polar/lipid phase separation, including the Bligh and Dyer protocol and the Folch protocol (GC-MS, lipidome); and using 70–80% methanol for untargeted analyses (LC-MS). In sample preparation for ^{1}H-NMR, extraction can be omitted, but dilution with deuterated solvents is necessary in order to avoid absorption by the ^{1}H-containing solvents and to stabilize the magnetic field.

Derivatization

Based on the purpose or limitations of the study, derivatization of metabolites is conducted. The major reasons for performing derivatization are to improve the sensitivity of the analysis, to detect metabolites with specific functional groups, to protect instruments, to alter the preference of the detection instruments for specific moieties, and to improve the accuracy of quantification by stable isotope labeling. As mentioned in the section on GC-MS, less volatile compounds such as sugars, amino acids, organic acids, and fatty acids become detectable using GC-MS by converting them to volatile derivatives by silylation and esterification (Beale *et al.* 2018). GC-MS has been used for comprehensive detection of steroids. The derivatization of them to picolinyl esters facilitates the high sensitivity detection of steroids using standard LC-MS conditions with ESI (Honda *et al.* 2010).

Controls

Performing appropriate controls is essential for eliminat-ing false positives, especially for high-sensitivity detection of metabolites by MS. Blank samples, which are prepared excluding the addition of the samples and using the same procedures that are used to prepare the experimental samples, are used to subtract false-positive peaks which can be attributed to contaminants in the solvents and solvent-leaching compounds from plastic tubes and pipette tips. The internal standards (IS) added to the samples at fixed concentration are used to check for, and correct, the reproducibility of retention times and detector sensitivity. Precise checking and correction of the reproducibility and sensitivity can be performed by performing quality control (QC) of samples that contain a larger variety of metabolites. The QC samples are typically prepared by mixing an equal amount of all samples for testing.

Data analysis

Before analyzing and mining metabolomic data using computational methods such as statistical multivariate analyses, the raw data generated by the MS- and NMR-based instruments should be properly processed and compiled into a data matrix that contains the metabolite signal intensities (concentrations) of each sample to be compared. This preprocessing procedure is comprised of several steps: 1) peak detection and characterization; 2) peak alignment between the samples; and 3) peak identification or annotation.

In GC-MS analysis, the peak signal is represented as a mass spectrum (see the section on GC-MS). When fragment ions with the same mass values are shared among metabolites that are eluted at adjacent retention times, the signals of each fragment ion needs to be assigned appropriately to each metabolite. This process is called deconvolution and is included in step 1.

In LC-MS analysis, the mass values of the ions of intact molecules are detected in step 1. Here, discrimination of the type of the ion has to be performed to estimate the mass values of the original (usually uncharged) molecule. During the ESI process, the neutral molecules are ionized with ions in the peripheral environment (mainly from solvents) to produce adduct ions, such as protonated/deprotonated molecules ($[M + H]^+$/$[M - H]^-$) and sodium cationized molecules ($[M + Na]^+$). Discrimination of the adduct ions is based on the mass differences between the co-eluting mass peaks in the chromatogram. Using the same calculation, identification of the ^{13}C-derived stable isotopic peaks, multiply charged ion peaks, and peaks derived from in-source fragmentation in the MS needs to be performed. The proper assignment of MS/MS or a multi-stage MS spectrum to the precursor ion peak is performed in step 1.

Before or after constructing the data matrix by the peak alignment in step 2, normalization of the signals by QC or use of ISs, noise filtering based on the technical replicates and blank filtering are performed. Then the data calculated in step 1 are integrated into the data matrix and used for peak identification and annotation in step 3. The metabolite peaks are identified if the peak of the authentic standard is

present and aligned at the same metabolite on the matrix. If authentic standards are not available, the peaks are annotated by searching known metabolite databases and mass spectral libraries, and the structure of the underlying metabolites can then be predicted using the calculated deionized mass values and mass spectra; numerous databases, libraries, and predictive software tools have been developed for this purpose (Misra 2021, O'Shea and Misra 2020).

The same statistical analysis and data mining procedures that are used for the other omics analysis can then be applied using the resulting data matrix.

Future perspectives

Metabolomic analysis will be applied to the breeding and improvement of a much larger variety of crops in future. In addition to the major crops, minor crops are an important target for future food security (Tadele 2019). For the breeding of roots, tubers, and bananas (RTB), for which the underlying genetic mechanisms are generally poorly understood, large-scale metabolome data have been obtained and a chemotype core collection has been developed (Price et al. 2020). It has been reported that integrating the analysis of metabolome data with genomic data increases the predictability of crop traits, such as estimating lipid composition and yields (Campbell et al. 2021, Wang et al. 2019, Xu et al. 2021). However, it has also been suggested that the integration and standalone use of metabolome data does not increase the predictability of yield and plant height, respectively (Gemmer et al. 2020, Thorwarth et al. 2019). This discrepancy can partly be attributed to the fact that the plant metabolome is markedly affected by environmental factors. In order to more precisely predict complex and environmentally dependent traits, such as yield, further improvements in the selection of an appropriate set of metabolome data for each crop, development of species-specific bioinformatics techniques, and publication of an easy-to-use dataset for bioinformatics researchers are desired.

Improvements in accessibility to metabolomics technologies should also be considered to extend the range of applications of metabolomics to plant breeding because, as stated in this review, there are large variety of the instruments and their data quality. The establishment of networks between metabolomics laboratories and sharing information with plant breeders in each country should also be accelerated. Establishing a metabolomics center that can accommodate a variety of analytical requests, Wuhan MetWare Biotechnology Co., Ltd. in China as an example, is another solution.

Given the increasing demand for high-throughput analyses in mGWAS and prediction studies, the development of rapid and simple techniques is also required. The probe electrospray ionization (PESI) method—an ambient ionization method—is one such prospective technology. Using PESI-MS, the metabolites in a small amount of sample which is collected discontinuously using a probe can be measured rapidly. The advantage of the PESI method is the moderation of ion suppression by sequential and exhaustive ionization (Hiraoka et al. 2020, Mandal et al. 2011). In addition, vibrating sharp-edge spray ionization (VSSI), which does not require a high voltage, was recently proposed as a compact and inexpensive ionization method (Li et al. 2021). Beck et al. (2015) described a portable GC-MS for VOC detection in the field. Such technological developments for minimizing the size of detectors are desired. In addition to the methods for precisely detecting individual metabolites described above, improvements in the practical application of portable phenotyping devices is also necessary. For example, the application of near-infrared spectroscopy (NIR) for phenotyping has been recently reported (Alves Filho et al. 2019, Lane et al. 2020, Rincent et al. 2018). The use of chemical sensors, such as electronic noses, is also being investigated (Cui et al. 2018, Tholl et al. 2021). The development of portable, non-destructive, and inexpensive, but robust chemical detectors, combined with efficient methods reliably correlate device signals to accurate metabolite data will be a major breakthrough in this research area.

The identification of unknown peaks detected by MS- and NMR-based metabolomics is a necessary step for elucidating the underlying mechanisms of the phenotypes. In future breeding studies, crops harboring a wider range of specialized metabolites that characterize the crop phenotypes will be targeted. Among the large number of unknown peaks, such specialized metabolites should be prioritized for further detailed analysis, such as identification, annotation, and examination of functionality. In order to such prioritization, metabolome databases which facilitate comparisons among a wider range of species is needed. An example of a database is the Food Metabolome Repository (http://metabolites.in/foods/) (Sakurai and Shibata 2017), in which untargeted LC-MS data are compared among 222 foodstuffs, including more than 96 raw materials derived from plants. The expansion of such a concept is required in future studies.

Author Contribution Statement

NS prepared the manuscript.

Acknowledgments

Literature Cited

Abreu, I.N., R.M. Brennan, E.N. Kanichukattu, D. Stewart, R.D. Hancock, G.J. McDougall and C.A. Hackett (2020) Quantitative trait loci mapping of polyphenol metabolites in blackcurrant (*Ribes nigrum* L.). Metabolomics 16: 25.

Afendi, F.M., T. Okada, M. Yamazaki, A. Hirai-Morita, Y. Nakamura, K. Nakamura, S. Ikeda, H. Takahashi, M. Altaf-Ul-Amin, L.K. Darusman *et al.* (2012) KNApSAcK family databases: Integrated metabolite-plant species databases for multifaceted plant research. Plant Cell Physiol 53: e1.

Alseekh, S. and A.R. Fernie (2021) Using metabolomics to assist plant breeding. Methods Mol Biol 2264: 33–46.

Alseekh, S., F. Scossa, W. Wen, J. Luo, J. Yan, R. Beleggia, H.J. Klee, S. Huang, R. Papa and A.R. Fernie (2021) Domestication of crop metabolomes: Desired and unintended consequences. Trends Plant Sci 26: 650–661.

Alves Filho, E., L.M. Silva, Y. Lima, P. Ribeiro, E. Silva, G. Zocolo, K. Canuto, S. Morais, A.C. Castro and E. de Brito (2019) Metabolomic variability of different genotypes of cashew by LC-Ms and correlation with near-infrared spectroscopy as a tool for fast phenotyping. Metabolites 9: 121.

Amini, H., M.R. Naghavi, T. Shen, Y. Wang, J. Nasiri, I.A. Khan, O. Fiehn, P. Zerbe and J.N. Maloof (2019) Tissue-Specific Transcriptome analysis reveals candidate genes for terpenoid and phenylpropanoid metabolism in the medicinal plant *Ferula assafoetida*. G3 (Bethesda) 9: 807–816.

Bamba, T., N. Shimonishi, A. Matsubara, K. Hirata, Y. Nakazawa, A. Kobayashi and E. Fukusaki (2008) High throughput and exhaustive analysis of diverse lipids by using supercritical fluid chromatography-mass spectrometry for metabolomics. J Biosci Bioeng 105: 460–469.

Batyrshina, Z.S., B. Yaakov, R. Shavit, A. Singh and V. Tzin (2020) Comparative transcriptomic and metabolic analysis of wild and domesticated wheat genotypes reveals differences in chemical and physical defense responses against aphids. BMC Plant Biol 20: 19.

Beale, D.J., F.R. Pinu, K.A. Kouremenos, M.M. Poojary, V.K. Narayana, B.A. Boughton, K. Kanojia, S. Dayalan, O.A.H. Jones and D.A. Dias (2018) Review of recent developments in GC-MS approaches to metabolomics-based research. Metabolomics 14: 152.

Beck, J.J., N. Porter, D. Cook, W.S. Gee, C.M. Griffith, A.D. Rands, T.V. Truong, L. Smith and I. San Roman (2015) In-field volatile analysis employing a hand-held portable GC-MS: Emission profiles differentiate damaged and undamaged yellow starthistle flower heads. Phytochem Anal 26: 395–403.

Bedair, M. and K.C. Glenn (2020) Evaluation of the use of untargeted metabolomics in the safety assessment of genetically modified crops. Metabolomics 16: 111.

Beleggia, R., D. Rau, G. Laido, C. Platani, F. Nigro, M. Fragasso, P. De Vita, F. Scossa, A.R. Fernie, Z. Nikoloski *et al.* (2016) Evolutionary metabolomics reveals domestication-associated changes in tetraploid wheat kernels. Mol Biol Evol 33: 1740–1753.

Buko, A.M. (2017) Capillary electrophoresis mass spectrometry based metabolomics. J Appl Bioanal 3: 5–20.

Campbell, M.T., H. Hu, T.H. Yeats, M. Caffe-Treml, L. Gutierrez, K.P. Smith, M.E. Sorrells, M.A. Gore and J.L. Jannink (2021) Translating insights from the seed metabolome into improved prediction for lipid-composition traits in oat (*Avena sativa* L.). Genetics 217: jyaa043.

Chevalier, W., S.A. Moussa, M. Medeiros Netto Ottoni, C. Dubois-Laurent, S. Huet, C. Aubert, E. Desnoues, B. Navez, V. Cottet, G. Chalot *et al.* (2021) Multisite evaluation of phenotypic plasticity for specialized metabolites, some involved in carrot quality and disease resistance. PloS One 16: e0249613.

Cui, J., Z. Lu, T. Wang, G. Chen, S. Mostafa, H. Ren, S. Liu, C. Fu, L. Wang, Y. Zhu *et al.* (2021) The genome of *Medicago polymorpha* provides insights into its edibility and nutritional value as a vegetable and forage legume. Hortic Res 8: 47.

Cui, S., P. Ling, H. Zhu and H.M. Keener (2018) Plant pest detection using an artificial nose system: A review. Sensors (Basel) 18: 378.

Dar, M.S., B.B. Dholakia, H. Shanmugam, V.S. Gupta, K.S. Subramanian, J. Subramanian and A.P. Giri (2020) Differential modulation in metabolites revealed with the improvement in the shelf-life of alphonso fruits. Mol Biotechnol 62: 508–520.

Dodds, J.N. and E.S. Baker (2019) Ion mobility spectrometry: Fundamental concepts, instrumentation, applications, and the road ahead. J Am Soc Mass Spectrom 30: 2185–2195.

Emwas, A.H., R. Roy, R.T. McKay, L. Tenori, E. Saccenti, G.A.N. Gowda, D. Raftery, F. Alahmari, L. Jaremko, M. Jaremko *et al.* (2019) NMR Spectroscopy for Metabolomics Research. Metabolites 9: 123.

Erzen, M., I.J. Kosir, M. Ocvirk, S. Kreft and A. Cerenak (2021) Metabolomic analysis of cannabinoid and essential oil profiles in different hemp (*Cannabis sativa* L.) phenotypes. Plants (Basel) 10: 966.

Feider, C.L., A. Krieger, R.J. DeHoog and L.S. Eberlin (2019) Ambient ionization mass spectrometry: Recent developments and applications. Anal Chem 91: 4266–4290.

Feiner, A., N. Pitra, P. Matthews, K. Pillen, L.A. Wessjohann and D. Riewe (2021) Downy mildew resistance is genetically mediated by prophylactic production of phenylpropanoids in hop. Plant Cell Environ 44: 323–338.

Ferguson, C., D. Araujo, L. Faulk, Y. Gou, A. Hamelers, Z. Huang, M. Ide-Smith, M. Levchenko, N. Marinos, R. Nambiar *et al.* (2021) Europe PMC in 2020. Nucleic Acids Res 49: D1507–D1514.

Ferguson, C.N., J.W.M. Fowler, J.F. Waxer, R.A. Gatti and J.A. Loo (2019) Mass spectrometry-based tissue imaging of small molecules. Adv Exp Med Biol 1140: 99–109.

Fernandez, O., E.J. Millet, R. Rincent, S. Prigent, P. Pétriacq and Y. Gibon (2021) Chapter Seven—Plant metabolomics and breeding. *In*: Pétriacq, P. and A. Bouchereau (eds.) Advances in Botanical Research, Academic Press.

Fernie, A.R. and J. Yan (2019) *De Novo* domestication: An alternative route toward new crops for the future. Mol Plant 12: 615–631.

Ferrão, L.F.V., R.R. Amadeu, J. Benevenuto, I. de Bem Oliveira and P.R. Munoz (2021) Genomic selection in an outcrossing autotetraploid fruit crop: Lessons from blueberry breeding. Front Plant Sci 12: 676326.

Gamboa-Becerra, R., M.C. Hernandez-Hernandez, O. Gonzalez-Rios, M.L. Suarez-Quiroz, E. Galvez-Ponce, J.J. Ordaz-Ortiz and R. Winkler (2019) Metabolomic markers for the early selection of *Coffea canephora* plants with desirable cup quality traits. Metabolites 9: 214.

Garcia-Casarrubias, A., R. Winkler and A. Tiessen (2019) Mass fingerprints of Tomatoes fertilized with different nitrogen sources reveal potential biomarkers of organic farming. Plant Foods Hum Nutr 74: 247–254.

Gemmer, M.R., C. Richter, Y. Jiang, T. Schmutzer, M.L. Raorane, B. Junker, K. Pillen and A. Maurer (2020) Can metabolic prediction

be an alternative to genomic prediction in barley? PLoS One 15: e0234052.

Gemmer, M.R., C. Richter, T. Schmutzer, M.L. Raorane, B. Junker, K. Pillen and A. Maurer (2021) Genome-wide association study on metabolite accumulation in a wild barley NAM population reveals natural variation in sugar metabolism. PLoS One 16: e0246510.

Gertsman, I. and B.A. Barshop (2018) Promises and pitfalls of untargeted metabolomics. J Inherit Metab Dis 41: 355–366.

Gordillo, R. (2021) Supercritical fluid chromatography hyphenated to mass spectrometry for metabolomics applications. J Sep Sci 44: 448–463.

Guijas, C., J.R. Montenegro-Burke, B. Warth, M.E. Spilker and G. Siuzdak (2018) Metabolomics activity screening for identifying metabolites that modulate phenotype. Nat Biotechnol 36: 316–320.

Hall, R.D. (2006) Plant metabolomics: From holistic hope, to hype, to hot topic. New Phytol 169: 453–468.

Hao, Z., S. Liu, L. Hu, J. Shi and J. Chen (2020) Transcriptome analysis and metabolic profiling reveal the key role of carotenoids in the petal coloration of Liriodendron tulipifera. Hortic Res 7: 70.

Hiraoka, K., O. Ariyada, D.T. Usmanov, L.C. Chen, S. Ninomiya, K. Yoshimura, S. Takeda, Z. Yu, M.K. Mandal, H. Wada *et al.* (2020) Probe electrospray ionization (PESI) and its modified versions: Dipping PESI (dPESI), Sheath-Flow PESI (sfPESI) and Adjustable sfPESI (ad-sfPESI). Mass Spectrom (Tokyo) 9: A0092.

Hodges, H.E., H.J. Walker, A.J. Cowieson, R.J. Falconer and D.D. Cameron (2021) Latent anti-nutrients and unintentional breeding consequences in Australian *Sorghum bicolor* varieties. Front Plant Sci 12: 625260.

Hommerson, P., A.M. Khan, T. Bristow, M.W. Harrison, G.J. de Jong and G.W. Somsen (2009a) Drug impurity profiling by capillary electrophoresis/mass spectrometry using various ionization techniques. Rapid Commun Mass Spectrom 23: 2878–2884.

Hommerson, P., A.M. Khan, G.J. de Jong and G.W. Somsen (2009b) Capillary electrophoresis-atmospheric pressure chemical ionization-mass spectrometry using an orthogonal interface: Set-up and system parameters. J Am Soc Mass Spectrom 20: 1311–1318.

Honda, A., T. Miyazaki, T. Ikegami, J. Iwamoto, K. Yamashita, M. Numazawa and Y. Matsuzaki (2010) Highly sensitive and specific analysis of sterol profiles in biological samples by HPLC-ESI-MS/MS. J Steroid Biochem Mol Biol 121: 556–564.

Hu, W., Y. Han, Y. Sheng, Y. Wang, Q. Pan and H. Nie (2021) Mass spectrometry imaging for direct visualization of components in plants tissues. J Sep Sci 44: 3462–3476.

Kind, T., G. Wohlgemuth, D.Y. Lee, Y. Lu, M. Palazoglu, S. Shahbaz and O. Fiehn (2009) FiehnLib: Mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. Anal Chem 81: 10038–10048.

Klevorn, C.M., L.L. Dean and S.D. Johanningsmeier (2019) Metabolite profiles of raw peanut seeds reveal differences between market-types. J Food Sci 84: 397–405.

Kodama, H., T. Miyahara, T. Oguchi, T. Tsujimoto, Y. Ozeki, T. Ogawa, Y. Yamaguchi and D. Ohta (2021) Effect of transgenic rootstock grafting on the omics profiles in Tomato. Food Saf (Tokyo) 9: 32–47.

Kopka, J., N. Schauer, S. Krueger, C. Birkemeyer, B. Usadel, E. Bergmuller, P. Dormann, W. Weckwerth, Y. Gibon, M. Stitt *et al.* (2005) GMD@CSB.DB: the Golm Metabolome Database.

Bioinformatics 21: 1635–1638.

Lane, H.M., S.C. Murray, O.A. Montesinos-López, A. Montesinos-López, J. Crossa, D.K. Rooney, I.D. Barrero-Farfan, G.N. De La Fuente and C.L.S. Morgan (2020) Phenomic selection and prediction of maize grain yield from near-infrared reflectance spectroscopy of kernels. The Plant Phenome Journal 3: e20002.

Lemaitre-Guillier, C., C. Dufresne, A. Chartier, S. Cluzet, J. Valls, L. Jacquens, A. Douillet, N. Aveline, M. Adrian and X. Daire (2021) VOCs are relevant biomarkers of elicitor-induced defences in grapevine. Molecules 26: 4258.

Li, X., K. Attanayake, S.J. Valentine and P. Li (2021) Vibrating sharp-edge spray ionization (VSSI) for voltage-free direct analysis of samples using mass spectrometry. Rapid Commun Mass Spectrom 35 (Suppl 1): e8232.

Liu, W., H. Zhao, C. Miao and W. Jin (2021) Integrated proteomics and metabolomics analysis of transgenic and gene-stacked maize line seeds. GM Crops Food 12: 361–375.

Mandal, M.K., L.C. Chen and K. Hiraoka (2011) Sequential and exhaustive ionization of analytes with different surface activity by probe electrospray ionization. J Am Soc Mass Spectrom 22: 1493–1500.

Misra, B.B. (2021) New software tools, databases, and resources in metabolomics: Updates from 2020. Metabolomics 17: 49.

Moing, A., J.W. Allwood, A. Aharoni, J. Baker, M.H. Beale, S. Ben-Dor, B. Biais, F. Brigante, Y. Burger, C. Deborde *et al.* (2020) Comparative metabolomics and molecular phylogenetics of Melon (*Cucumis melo*, Cucurbitaceae) biodiversity. Metabolites 10: 121.

Mondello, L. (2015) Mass spectra of flavors and fragrances of natural and synthetic compounds. 3rd edn. Wiley & Sons.

Onda, Y., K. Inoue, Y. Sawada, M. Shimizu, K. Takahagi, Y. Uehara-Yamaguchi, M.Y. Hirai, D.F. Garvin and K. Mochida (2019) Genetic variation for seed metabolite levels in *Brachypodium distachyon*. Int J Mol Sci 20: 2348.

O'Shea, K. and B.B. Misra (2020) Software tools, databases and resources in metabolomics: Updates from 2018 to 2019. Metabolomics 16: 36.

Patti, G.J., O. Yanes and G. Siuzdak (2012) Innovation: Metabolomics: the apogee of the omics trilogy. Nat Rev Mol Cell Biol 13: 263–269.

Pinu, F.R., S.A. Goldansaz and J. Jaine (2019) Translational metabolomics: Current challenges and future opportunities. Metabolites 9: 108.

Price, E.J., M. Drapal, L. Perez-Fons, D. Amah, R. Bhattacharjee, B. Heider, M. Rouard, R. Swennen, L.A. Becerra Lopez-Lavalle and P.D. Fraser (2020) Metabolite database for root, tuber, and banana crops to facilitate modern breeding in understudied crops. Plant J 101: 1258–1268.

Qing, Z., J. Liu, X. Yi, X. Liu, G. Hu, J. Lao, W. He, Z. Yang, X. Zou, M. Sun *et al.* (2021) The chromosome-level Hemerocallis citrina Borani genome provides new insights into the rutin biosynthesis and the lack of colchicine. Hortic Res 8: 89.

Razzaq, A., B. Sadia, A. Raza, M. Khalid Hameed and F. Saleem (2019) Metabolomics: A way forward for crop improvement. Metabolites 9: 303.

Rincent, R., J.P. Charpentier, P. Faivre-Rampant, E. Paux, J. Le Gouis, C. Bastien and V. Segura (2018) Phenomic selection is a low-cost and high-throughput method based on indirect predictions: Proof of concept on wheat and poplar. G3 (Bethesda) 8: 3961–3972.

Saito, K. and F. Matsuda (2010) Metabolomics for functional genomics, systems biology, and biotechnology. Annu Rev Plant Biol 61: 463–489.

Sakurai, N. and D. Shibata (2017) Tools and databases for an integrated metabolite annotation environment for liquid chromatography-mass spectrometry-based untargeted metabolomics. Carotenoid Science 22: 16–22.

Sarvin, B., S. Lagziel, N. Sarvin, D. Mukha, P. Kumar, E. Aizenshtein and T. Shlomi (2020) Fast and sensitive flow-injection mass spectrometry metabolomics by analyzing sample-specific ion distributions. Nat Commun 11: 3186.

Sayers, E.W., J. Beck, E.E. Bolton, D. Bourexis, J.R. Brister, K. Canese, D.C. Comeau, K. Funk, S. Kim, W. Klimke *et al.* (2021) Database resources of the national center for biotechnology information. Nucleic Acids Res 49: D10–D17.

Schouten, H.J., Y. Tikunov, W. Verkerke, R. Finkers, A. Bovy, Y. Bai and R.G.F. Visser (2019) Breeding has increased the diversity of cultivated Tomato in the Netherlands. Front Plant Sci 10: 1606.

Schripsema, J. (2010) Application of NMR in plant metabolomics: Techniques, problems and prospects. Phytochem Anal 21: 14–21.

Scossa, F., S. Alseekh and A.R. Fernie (2021) Integrating multi-omics data for crop improvement. J Plant Physiol 257: 153352.

Sharma, V., P. Gupta, K. Priscilla, S. Kumar, B. Hangargi, A. Veershetty, D.P. Ramrao, S. Suresh, R. Narasanna, G.R. Naik *et al.* (2021) Metabolomics intervention towards better understanding of plant traits. Cells 10: 346.

Soga, T., Y. Ohashi, Y. Ueno, H. Naraoka, M. Tomita and T. Nishioka (2003) Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. J Proteome Res 2: 488–494.

Song, X., J. Wang, N. Li, J. Yu, F. Meng, C. Wei, C. Liu, W. Chen, F. Nie, Z. Zhang *et al.* (2020) Deciphering the high-quality genome sequence of coriander that causes controversial feelings. Plant Biotechnol J 18: 1444–1456.

Song, Y., P. Chen, A. Xuan, C. Bu, P. Liu, P.K. Ingvarsson, Y.A. El-Kassaby and D. Zhang (2021) Integration of genome wide association studies and co-expression networks reveal roles of *PtoWRKY 42-PtoUGT76C1-1* in *trans*-zeatin metabolism and cytokinin sensitivity in poplar. New Phytol 231: 1462–1477.

Southam, A.D., R.J.M. Weber, J. Engel, M.R. Jones and M.R. Viant (2016) A complete workflow for high-resolution spectral-stitching nanoelectrospray direct-infusion mass-spectrometry-based metabolomics and lipidomics. Nat Protoc 12: 310–328.

Sumner, L.W., A. Amberg, D. Barrett, M.H. Beale, R. Beger, C.A. Daykin, T.M. Fan, O. Fiehn, R. Goodacre, J.L. Griffin *et al.* (2007) Proposed minimum reporting standards for chemical analysis. Metabolomics 3: 211–221.

Tadele, Z. (2019) Orphan crops: Their importance and the urgency of improvement. Planta 250: 677–694.

Tholl, D., O. Hossain, A. Weinhold, U.S.R. Rose and Q. Wei (2021) Trends and applications in plant volatile sampling and analysis. Plant J 106: 314–325.

Thorwarth, P., G. Liu, E. Ebmeyer, J. Schacht, R. Schachschneider, E. Kazman, J.C. Reif, T. Wurschum and C.F.H. Longin (2019) Dissecting the genetics underlying the relationship between protein content and grain yield in a large hybrid wheat population. Theor Appl Genet 132: 489–500.

van de Velde, B., D. Guillarme and I. Kohler (2020) Supercritical fluid chromatography—Mass spectrometry in metabolomics: Past, present, and future perspectives. J Chromatogr B Analyt Technol Biomed Life Sci 1161: 122444.

Viant, M.R., I.J. Kurland, M.R. Jones and W.B. Dunn (2017) How close are we to complete annotation of metabolomes? Curr Opin Chem Biol 36: 64–69.

Wang, H., Z. Li, L. Yuan, H. Zhou, X. Hou and T. Liu (2021) Cold acclimation can specifically inhibit chlorophyll biosynthesis in young leaves of Pakchoi. BMC Plant Biol 21: 172.

Wang, S., J. Wei, R. Li, H. Qu, J.M. Chater, R. Ma, Y. Li, W. Xie and Z. Jia (2019) Identification of optimal prediction models using multi-omic data for selecting hybrid rice. Heredity (Edinb) 123: 395–406.

Xu, Y., Y. Zhao, X. Wang, Y. Ma, P. Li, Z. Yang, X. Zhang, C. Xu and S. Xu (2021) Incorporation of parental phenotypic data into multiomic models improves prediction of yield-related traits in hybrid rice. Plant Biotechnol J 19: 261–272.

Yao, L., P. Li, Q. Du, M. Quan, L. Li, L. Xiao, F. Song, W. Lu, Y. Fang and D. Zhang (2021) Genetic architecture underlying the metabolites of chlorogenic acid biosynthesis in populus tomentosa. Int J Mol Sci 22: 2386.

Yu, X., J. Xiao, S. Chen, Y. Yu, J. Ma, Y. Lin, R. Li, J. Lin, Z. Fu, Q. Zhou *et al.* (2020) Metabolite signatures of diverse *Camellia sinensis* tea populations. Nat Commun 11: 5586.

Zhang, W. and R. Ramautar (2021) CE-MS for metabolomics: Developments and applications in the period 2018–2020. Electrophoresis 42: 381–401.

Zhou, Z., M. Luo, X. Chen, Y. Yin, X. Xiong, R. Wang and Z.J. Zhu (2020) Ion mobility collision cross-section atlas for known and unknown metabolite annotation in untargeted metabolomics. Nat Commun 11: 4334.

Zhu, G., S. Wang, Z. Huang, S. Zhang, Q. Liao, C. Zhang, T. Lin, M. Qin, M. Peng, C. Yang *et al.* (2018) Rewiring of the fruit metabolome in Tomato breeding. Cell 172: 249–261.e12.

Zullig, T., M. Trotzmuller and H.C. Kofeler (2020) Lipidomics from sample preparation to data analysis: A primer. Anal Bioanal Chem 412: 2191–2209.