Part I
Central Metabolism

The Handbook of Plant Metabolomics, First Edition. Edited by Wolfram Weckwerth and Günter Kahl

1
Metabolic Profiling of Plants by GC–MS

Camilla B. Hill and Ute Roessner

1.1
Introduction

For numerous organisms, complete genomes have been sequenced [1–3] and transcriptome [4–6] and proteome studies [7–9] have been described, but only recently have metabolome analyses using mass spectrometry (MS)-based platforms attracted attention. Recent advances in analytical technologies have now allowed the analysis of complex metabolic structures in an organism.

Metabolomics is currently a very powerful tool for characterizing metabolites and metabolic pathways and aims to provide a "snapshot" of the biochemical state of a biological sample. The number of metabolites is expected to be significantly lower than the number of genes, proteins, or mRNAs, which reduces the complexity of the sample. However, the total number of metabolites in the plant kingdom is estimated to be between 100 000 and 200 000, which makes cataloging of all metabolites a challenging task [10,11]. The metabolic composition of plants is likely to be altered during different physiological and environmental conditions and can also reflect different genetic backgrounds. Metabolomics aims to provide a comprehensive and unbiased analysis of all metabolites with a low molecular weight present in a biological sample, such as an organism, a specific tissue, or a cell, under certain conditions [12].

Analytical strategies for plant metabolite analysis include metabolic profiling, metabolite target analysis, and metabolic fingerprinting and are chosen according to either the focus of the research or the research question [12–14]. Metabolite profiling aims to detect as many metabolites as possible within a structurally related predefined group, for example, organic acids, amino acids, and carbohydrates. Metabolic profiling does not necessarily aim to determine absolute concentrations of metabolites but rather their comparative levels. In contrast, the aim of targeted metabolite analysis is to determine pool sizes (e.g., absolute concentrations) of metabolites involved in a particular pathway by utilizing specialized extraction protocols and adapted separation and detection methods. A third conceptual approach in metabolome analysis is metabolic fingerprinting, which generally is not intended to identify individual metabolites, but rather provides a fingerprint of

The Handbook of Plant Metabolomics, First Edition. Edited by Wolfram Weckwerth and Günter Kahl

all chemicals measurable for sample comparison and discrimination analysis by nonspecific rapid analysis of crude metabolite mixtures. Depending on the analytical strategy, a number of different instrumental platforms with different configurations may need to be utilized to ensure optimal data acquisition [15].

Because of the diversity of structural classes of metabolites, ranging from primary metabolites such as carbohydrates, amino acids, and organic acids to very complex secondary metabolites such as phenolics, alkaloids, and terpenoids, there is no single methodology that can measure the complete metabolome in one step. It is necessary to combine different techniques to detect all metabolites in a complex mixture [13]. It is possible that two samples, although very different, may show the same metabolite profile using one strategy. Therefore, only by employing a combination of different instrument platforms and techniques can the suite of differences in the metabolite profiles be revealed.

Several extraction methods and instrument platforms have been established to analyze highly complex mixtures, and each has to be chosen according to particular interests. These include nuclear magnetic resonance (NMR), Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS or FT-MS), and mass spectrometry (MS) coupled with liquid chromatography (LC) [liquid chromatography–mass spectrometry (LC–MS)] or gas chromatography (GC) [gas chromatography–mass spectrometry (GC–MS)]. Section 1.2 focuses on the application of GC–MS to plant metabolomics studies; the advantages and disadvantages of other instrument platforms for metabolomics were discussed in Refs. [16–19].

The coupling of GC to electron impact ionization (EI) MS is possibly the oldest hybrid technique in analytical chemistry and is considered to be one of the most developed, robust, and highly sensitive instrument platforms for metabolite analysis [20–22]. GC–MS offers high chromatographic separation power, robust quantification methods, and the capability to identify metabolites with high fidelity, and is therefore often referred to as the "gold standard" in metabolomics [23]. GC–MS-based methodologies were among the first to be applied to metabolite profiling and target analysis, thus offering established protocols for machine setup, data mining, and interpretation. Compared with other instrument platforms, it offers the lowest acquisition, operating, and maintenance expenses [24]. Furthermore, both commercially and publicly available EI spectral libraries facilitate the use of GC–MS as a metabolomics platform [25].

Historically, the first chromatographic separation techniques were developed between 1940 and 1950 by Martin and Synge, who won the 1952 Nobel Prize for their invention of partition chromatography [26,27]. They further contributed substantially to the development of GC and high-performance liquid chromatography (HPLC). During the 1970s, the term "metabolite profiling" was coined and was first applied in studies of steroid and steroid derivatives, amino acids, and drug metabolites [28,29] in 1971. In the following years, metabolite research developed toward the utilization of metabolic profiling by GC–MS as a diagnostic technique in medicine to monitor metabolites present in urine [30]. But it was not until the 1990s that metabolomics found its way into plant research. In the late 1990s, Oliver *et al.*

were the first to introduce the terms metabolome and metabolomics [31]. About a decade ago, one of the first approaches for high-throughput, large-scale, and comprehensive plant metabolite analysis was conducted by Roessner *et al.* [21,32,33], who analyzed more than 150 compounds simultaneously within a single potato (*Solanum tuberosum*) tuber sample using GC–MS, and Fiehn *et al.* [20], who analyzed 326 distinct compounds from *Arabidopsis thaliana* leaf extracts of four genotypes by GC–MS, and identified ~50% of these compounds. Several studies have now implemented this approach, and it has been applied to various plant species and tissues, including *A. thaliana* leaf tissue [13], phloem exudates of buttercup squash (*Cucubita maxima*) [34], tomato leaves and fruit (*Solanum lycopersicum*) [35,36], and barley leaf and root tissue (*Hordeum vulgare*) [37]. GC–MS applications include studies that associate certain metabolites with biotic [38] and abiotic stress responses [39–42], define metabolic differences of genetically modified plants [32,33,35,43,44], or integrate genetic and metabolite data for plant functional genomics [45–49].

GC is the preferred technique for the separation of low molecular weight metabolites which are either volatile or can be converted into volatile and thermally stable compounds through chemical derivatization before analysis [15]. This includes especially primary metabolites, such as amino acids, amines, sugars, organic acids, fatty acids, long-chain alcohols, and sterols, whereas LC–MS analysis is favored for detecting a broader range of metabolites, including secondary metabolites such as alkaloids, terpenes, flavonoids, glucosinolates, and phenylpropanoids [50,51]. Derivatization is usually needed to increase volatility and to reduce the polarity of polar hydroxyl (—OH), amine ($—NH_2$), carboxyl (—COOH), and thiol (—SH) groups [25]. Exceptions include plant volatiles [52] and metabolites present in essential oils [53], which can be injected directly into the GC column.

The greatest challenge of any metabolomics project is to make sense of the wealth of data that has been produced during metabolite analysis. Targeted metabolite analysis employs optimized measurements of preselected metabolites, which are characterized by their mass spectrum and retention time/index, and allows the fast and easy construction of the data matrix [25,54]. It is a highly quantitative method with a very high detection rate for known metabolites, which must be available in purified form. To quantify metabolites, either external calibration (which requires preparation of standard solutions) or internal calibration (based on the relation between the peak area of the compound and that of an internal standard) can be employed [16].

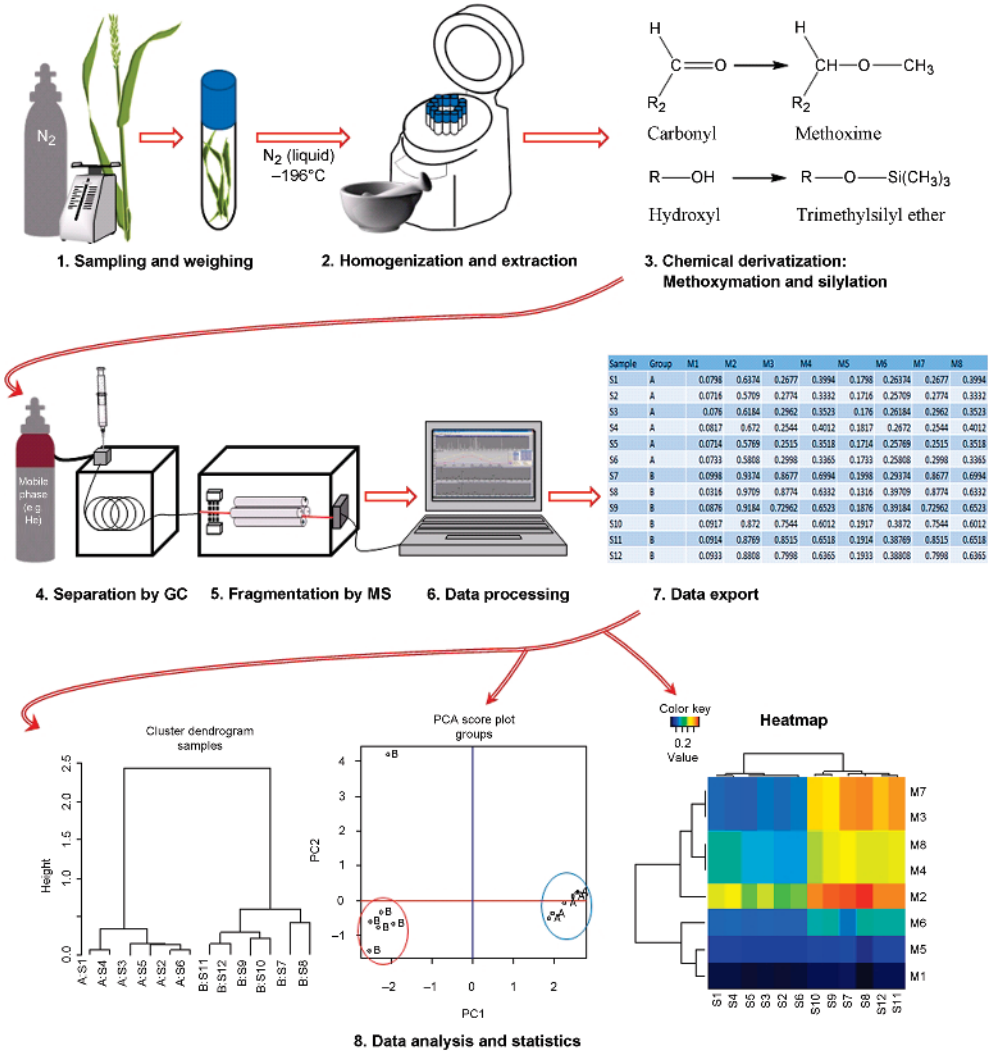
In contrast, untargeted analysis distinguishes all mass peaks above a certain threshold by their mass spectrum and retention time/index, with the majority of them not being identified, and can be used to detect novel metabolic markers. In this case, data mining is more complex than in targeted analysis and requires bioinformatics and statistical tools to avoid labor-intensive and time-consuming manual data handling.

In our laboratory, we routinely use GC–MS as a tool to investigate tolerance mechanisms of plants, particularly cereal crops such as wheat, rice, and barley,

under abiotic stress, including drought, cold, salinity, and nutritional deficiencies or mineral toxicities (www.acpfg.com.au; www.metabolomics.com.au).

Plant metabolite profiling using GC–MS involves the steps depicted in Figure 1.1.

The most relevant sections of this experimental workflow are detailed in Section 1.2. The chapter then turns to the implementation of GC–MS in plant metabolomics, portraying various examples of applications of this technology. The final section reports new developments in GC–MS technology.



Sample	Group	M1	M2	M3	M4	M5	M6	M7	M8
S1	A	0.0798	0.6374	0.2677	0.3994	0.1798	0.26374	0.2677	0.3994
S2	A	0.0716	0.5709	0.2774	0.3332	0.1716	0.25709	0.2774	0.3332
S3	A	0.076	0.6184	0.2962	0.3523	0.176	0.26184	0.2962	0.3523
S4	A	0.0817	0.672	0.2544	0.4012	0.1817	0.2672	0.2544	0.4012
S5	A	0.0714	0.5769	0.2515	0.3518	0.1714	0.25769	0.2515	0.3518
S6	A	0.0733	0.5808	0.2998	0.3365	0.1733	0.25808	0.2998	0.3365
S7	B	0.0998	0.9374	0.8677	0.6994	0.1998	0.29374	0.8677	0.6994
S8	B	0.0316	0.9709	0.8774	0.6332	0.1316	0.39709	0.8774	0.6332
S9	B	0.0876	0.9184	0.72962	0.6523	0.1876	0.39184	0.72962	0.6523
S10	B	0.0917	0.872	0.7544	0.6012	0.1917	0.3872	0.7544	0.6012
S11	B	0.0914	0.8769	0.8515	0.6518	0.1914	0.38769	0.8515	0.6518
S12	B	0.0933	0.8808	0.7998	0.6365	0.1933	0.38808	0.7998	0.6365

Figure 1.1 Workflow showing the general strategy and experimental steps of a GC–MS experiment.

1.2
Methods and Protocols

1.2.1
Sample Preparation

Sample preparation can be an important source of variations [24], and sampling and extraction methods vary according to the type of biological sample and the targeted class(es) of metabolites. This has to be considered for subsequent data analysis and interpretation of the experiment, since such biases may have an impact on the accuracy and precision of the information gained from the experiment. In particular, when studying plant samples, the influences of environmental factors such as harvesting time (day/night, season), light conditions, temperature, developmental stage of the plant or plant cells, the type of harvested tissue/plant organs, and genetic factors have to be considered [55].

1.2.1.1 **Sampling**

The first step in sample preparation for plant metabolite analysis is harvesting of plant tissue by rapid freezing in liquid nitrogen (−196 °C) and storing at −80 °C, or freeze-drying for longer storage until used. This will stop all enzymatic processes and avoid degradation and modification of metabolites in the sample. More uncommon ways to quench the metabolism involve the use of cold methanol, perchloric acid, or nitric acid [56]. Harvesting should be performed at the same time of day for all samples to minimize biological variations due to diurnal changes of metabolism. The number of replicates is dependent on the experimental sources of variation, but since the biological variation generally exceeds the analytical variation, a minimum of three to six biological replicates per line is recommended [57,58]. Technical replicates ensure that the effect of instrument variations during the analytical run are minimized.

1.2.1.2 **Homogenization and Extraction**

Before extraction of metabolites, the plant tissue has to be homogenized to a fine powder to allow the solvent to penetrate the tissue to extract metabolites effectively. This is typically done using one of the following methods: grinding with a mortar and pestle using liquid nitrogen [44,59], milling in a ball-mill with precooled holders [20], or using ULTRA-TURRAX tissue homogenizers [21,35,60].

The next step in sample preparation is the extraction of plant metabolites, which has to be optimized to ensure minimal losses of metabolites due to enzymatic conversion or chemical degradation. Blank samples containing only extraction solution and no metabolite extract should be derivatized along with other samples and analyzed in each analytical run to identify contaminants, which are then excluded from further analysis. Additionally, pooled samples are prepared by a combination of aliquots from each biological sample as suggested by Sangster *et al.* [61]. These are used to produce a set of replicates, which are analyzed together with the real samples at the beginning, at the end, and randomly throughout the

analytical run. Therefore, all metabolites of the real samples are present in the (pooled) reference samples, which can be used to normalize the metabolite levels in the real samples. Furthermore, using principal component analysis (PCA), the quality of the data set can be inferred from the clustering of the pooled quality control samples (see Section 1.2.4). Since the quality control samples are replicates of the same sample, they should have very similar values for their principal components, which ensure that instrument sensitivity and chromatography during the analytical run are not changed significantly.

Internal standards are compounds that are not present in the biological sample (e.g., stable isotope-labeled compounds) and are included before or during metabolite extraction. In the case of targeted analysis, stable isotope-labeled internal standards that have chemical properties identical with those of the target metabolites are often used.

1.2.1.3 Procedure for Polar Extraction of Metabolites

The procedure is outlined in Figure 1.2. Weigh 30 ± 3 mg (the amount depends on the origin of the sample and needs to be confirmed for each tissue type) of frozen sample plant tissue into a 2 ml soft tissue homogenizing tube with 1.4 mm ceramic beads (Bertin Technologies) (1), and add 0.5 ml of 100% methanol extraction solution to the plant sample (2). Record exact sample weights. Perform homogenization for 1×30 s at 6000 rpm using a high-throughput tissue homogenizer (Precellys 24, Bertin Technologies). Following incubation for 15 min at 70 °C in a thermomixer at 850 rpm (3), centrifuge the sample for 10 min at 14 000 rpm at room temperature (RT) (4). Transfer the supernatant into a new 1.5 ml reaction tube (5a) and add 0.5 ml of 50% aqueous methanol solution containing internal standards

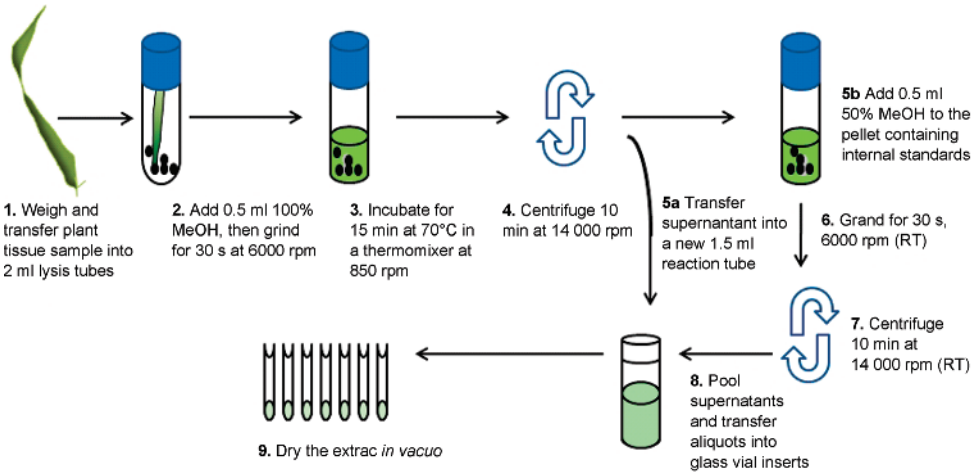


Figure 1.2 Experimental procedure for homogenization and polar extraction of plant metabolites for GC–MS profiling.

(20 µl per sample from a stock solution containing 0.2 mg/ml [^{13}C]sorbitol and 1 mg/ml [^{13}C]valine in 100% methanol) to the pellet (5b). After a second homogenization (6) and centrifugation step (7), pool the supernatants and transfer 50 µl aliquots (again, the amount needs to be confirmed for each tissue type for optimal analysis) into glass vial inserts suitable for GC–MS analysis (8). Dry all resulting aliquots *in vacuo* using a vacuum concentrator (9). For a subsequent GC–MS analysis, derivatize the sample immediately before analysis (see Section 1.2.3). *Note*: Prepare a sufficient amount of backup samples. Store the dried sample aliquots in plastic bags filled with silica gel beads at RT. For long-term storage, sample aliquots should be kept under argon to avoid oxidation and degradation of metabolites.

1.2.2 Chemical Derivatization: Methoxymation and Silylation

A variety of derivatizing agents with different properties have been developed, including alkylation, silylation, esterification, and acylation reagents [17,62]. Trimethylsilylation is a commonly used method to derivatize a broad range of metabolites, including sugars, sugar alcohols, amines, amino acids, and organic acids, in order for them to become volatile and thermally stable [21]. A two-step derivatization method involving oximation followed by silylation is commonly applied for GC–MS metabolite analysis: First, carbonyl groups are converted into the corresponding oximes using hydroxylamine or alkylhydroxylamine reagents (such as *O*-methylhydroxylamine hydrochloride, MeOx) to stabilize sugars in the open-ring conformation [16,17] (Figure 1.3 a). Oximes exist as two (*syn* and *anti*) stereoisomers, and therefore are often present as two peaks per compound in the chromatograms (denoted Mx1 and Mx2). This is followed by trimethylsilylation using silylating reagents such as *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide (MSTFA) + 1% trimethylchlorosilane (TMCS, a catalyst of the reaction), or alternatively *N*,*O*-bis(trimethylsilyl)trifluoroacetamide (BSTFA) + 1% TMCS, which replace active hydrogen in polar functional groups such as —OH, —COOH, —NH, and —SH with a TMS [—Si$(CH_3)_3$] group (Figure 1.3b). TMS derivatives are sensitive to moisture, which may cleave TMS derivatives. In contrast, tri-*tert*-butyldimethylsilyl (TBDMS) derivatives, which use *N*-methyl-*N*-(*tert*-butyldimethylsilyl)trifluoroacetamide (MTBSTFA) as a derivatization reagent, are more moisture resistant [63], but show a significant increase in molecular weight, which may lead to only partial derivatization due to steric hindrance [25]. *Note*: To ensure optimal sample stability, derivatization should be performed immediately before sample injection.

1.2.2.1 Procedure for the Chemical Derivatization of Plant Extracts

In our laboratory, a Gerstel MPS2XL GC–MS autosampler performs the derivatization procedure immediately before injection. Add the samples and the derivatization reagents (MeOx and BSTFA) to a glass vial and then place them in the autosampler tray. The autosampler mixes the sample with derivatization reagents automatically using the following program for derivatization using TMS. Plant extracts were derivatized for 120 min at 37 °C using 20 µl of MeOx solution (30 mg/ml MeOx

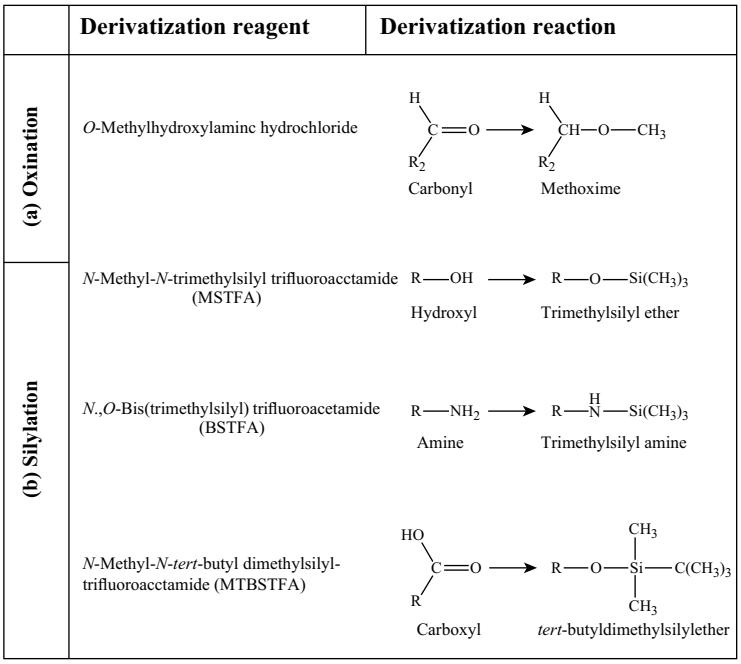


Figure 1.3 Chemical derivatization reactions commonly used for GC–MS-based plant metabolite analysis. (a) Methoxyamination of a carbonyl group. (b) Trimethylsilylation of a hydroxyl group and an amino group and tri-*tert*-butyldimethylsilylation of a carboxyl group.

dissolved in pyridine) per sample. This was followed by trimethylsilylation with 40 µl of BSTFA + 1% TMCS per sample for 30 min at 37 °C. Finally, 2 µl of retention time standard mixture [0.029% v/v *n*-dodecane, *n*-pentadecane, *n*-nonadecane, *n*-docosane, *n*-octacosane, *n*-dotriacontane, and *n*-hexatriacontane dissolved in pyridine; Sigma) per sample was added before injection into the GC column. *Note*: To prepare the MeOx solution, weigh 30 mg of MeOx in a reaction tube and after addition of 1 ml of pyridine heat the mixture for 5 min at 50 °C to dissolve the MeOx. Store the solution at RT for up to 1 month and avoid moisture. *Caution*: the derivatization reagents are extremely toxic and should be handled under a fume hood while wearing gloves.

1.2.3 GC–MS Analysis

Components are separated on the basis of differential partitioning between a mobile gas phase (typically helium) and a solid stationary phase (typically based on silicone polymers), which is bound to the inner surface of a fused-silica tube [18,64]. In the ion source, analytes are ionized by EI, creating distinct fragmentation patterns for each component. GC–MS traces of plant metabolites are commonly acquired using

a gas chromatograph coupled with either a single-quadrupole (QUAD), time-of-flight (TOF), or ion-trap (TRAP) mass analyzer, which separates the fragment ions according to their m/z values [20,21,50]. QUADs are comparably simple but versatile mass analyzers that consist of a set of four parallel metal rods that create an oscillating electric field when radiofrequency (RF) and DC voltages are applied to the rods [65]. Ions are separated depending on the stability of their trajectories through the electric field between the four rods. GC–QUAD-MS provides a large dynamic mass range of 2–4000 Da/e, but with a mass resolution around 1 : 1500 nominal mass accuracy and slow scan speeds compared with GC–TOF-MS [64,66]. Only recently have rapid-scanning QUADs been introduced, offering scan speeds of 10 000 amu/s [67].

In GC–TOF-MS instruments, bundles of ions are accelerated to high kinetic energy by an electric field and are separated along a flight tube as a result of their different velocities, depending on their m/z ratio [50,68]. GC–TOF-MS offers a higher m/z accuracy than conventional GC– QUAD-MS, which is important for the identification of unknowns [17]. Furthermore, GC–TOF-MS gives data acquisition rates with narrow high-resolution chromatographic peak widths (0.5–1 s), and therefore allows a higher sample throughput with shorter analysis times compared with QUAD- and TRAP-MS [50,66]. This is combined with a nominal mass resolution similar to that of a QUAD-MS [17].

TRAP instruments work by trapping and sequentially ejecting ions of successive masses [50]. Both QUAD and TRAP instruments are limited by low resolution; however, TRAPs are capable of reaction monitoring, which scans masses slowly over a predefined mass range to perform a second fragmentation step. This can facilitate compound identification and increases the mass resolution [50,65].

1.2.3.1 **Procedure to Acquire GC–MS Data**

In our laboratory, GC–MS traces are typically acquired using an Agilent 5975C gas chromatograph coupled with an Agilent Triple-Axis QUAD detector, operated by Chemstation software (Agilent). Samples are placed in random order on the sample tray and are analyzed along with several blank and pooled reference samples (see Section 1.2.1.2). Inject 1 µl of derivatized sample into the GC column using a hot needle technique with a 10 µl Hamilton syringe. Operate the injector in the splitless mode isothermally at 230 °C. Use helium as the carrier gas with a flow rate of 1 ml/min. Perform chromatographic separation on a 30 m VF-5MS column [with a 10 m Integra guard column of 0.25 mm i.d., 0.25 nm film thickness (Varian)]. Fix the MS transfer line to the quadrupole at 280 °C, the EI ion source at 250 °C, and the MS QUAD at 100 °C. Tune the mass spectrometer according to the manufacturer's protocols using tris(perfluorobutyl)amine (CF43).

Perform GC–MS analysis of plant tissue extracts using the following oven temperature program: set the injection temperature at 70 °C, followed by a 7 °C/min oven temperature gradient to a final 325 °C, and then hold for 3.6 min at 325 °C. The GC–MS system is then temperature equilibrated for 1 min at 70 °C before injecting the next sample. Ions are generated by a 70 eV electron beam at an ionization current of 2.0 mA and spectra are recorded at 2.91 scans per second with

an m/z scanning range of 50–550 amu. Retention time locking (RTL) of the chromatographic peak of mannitol before the sample run ensures repeatable retention times across the systems regardless of operator, detector type, and column maintenance. *Note*: For optimal sample analysis, GC–MS settings, including the injection temperature and the oven temperature gradient, have to be optimized and tailored for each type of plant sample and type of targeted metabolite class(es).

1.2.4 Data Preprocessing and Export

After the acquisition of mass spectra, the data sets have to be preprocessed, which includes the reduction of background noise, adjusting for baseline shifts and machine drift, peak alignment, peak detection, and mass spectral deconvolution, before they are subjected to searching against compound databases [69]. Software packages for effective *in silico* data preprocessing include the commercial software packages AnalyzerPro (SpectralWorks), Masshunter (Agilent), Xcalibur (Thermo-Fisher Scientific), and the freely available AMDIS (National Institute of Standards and Technology, Gaithersburg, MD, USA) (NIST) software (Table 1.1). The software detects component peaks in the chromatograms and calculates the relative amount by integration of the peak area below the peak, usually relative to the unique m/z of internal standards (standardization) [70]. To make the data suitable for statistical analysis (see Section 1.2.4.3), normalization has to adjust the data for experimental errors during sample preparation and changes in instrument sensitivity during the analytical run. Furthermore, retention time index (RI) systems based on either alkanes [71] or fatty acid methyl esters [72] are used for correct peak assignment, which depends on the relative elution of a compound between two RI standards. Compounds are identified by matching the RI and mass spectra of each compound, to minimize false peak assignment due to retention time shifts during the analytical run [73]. Automated calculation of the RI for all compounds and automated mass spectral deconvolution are implemented in most current software packages.

1.2.4.1 Procedure for Postacquisition Data Preprocessing

In this section, the data processing procedure using the commercially available AnalyzerPro software package (SpectralWorks, current version: 2.5.1.7) with the fully integrated NIST05 mass spectral search program (NIST) is described.

1) Import all data files into the AnalyzerPro (∗.swx) format.
2) Create a manual RI ladder by creating a ∗.csv file with alkane specifications (name/RI/RT).
3) Set up qualitative data processing of all data files of the pooled reference samples using a number of parameters in the "Processing Method" of AnalyzerPro. For targeted analysis, the use of the default settings is recommended: minimum masses = 4; area threshold = 500; height threshold = 1%; signal-to-noise ratio = 3; width threshold = 0.01 min; resolution = very low; scan windows = 3; smoothing = 3. The masses of m/z 73 (TMS), 147 (TMS-*O*-DMS), and 207

Table 1.1 Selected commercial and free software packages for data preprocessing.

Software	Publisher	Reference, URL
Chemstation	Agilent, USA	http://www.chem.agilent.com/en-US/products-services/software-informatics/openlab-cds-chemstation-edition/Pages/default.aspx
Xcalibur	Thermo Fisher Scientific, USA	http://www.thermoscientific.com/ecomm/servlet/productsdetail_11152___11961721_-1
AMDIS	National Institute of Standards and Technology (NIST), USA	[74], http://chemdata.nist.gov/mass-spc/amdis
AnalyzerPro	SpectralWorks, UK	www.spectralworks.com/analyzerpro.asp
Masshunter	Agilent, USA	http://www.chem.agilent.com/en-US/products-services/Software-Informatics/MassHunter-Workstation-Software/Pages/default.aspx
metAlign	Plant Research International (PRI), The Netherlands	[75], http://www.metalign.wur.nl
MZmine/MZmine2	Developed by Matej Orešič (VTT Technical Research Centre, Finland) and Mikko Katajamaa (Turku Centre for Biotechnology, Finland)	[76,77], http://mzmine.sourceforge.net
PyMS	Developed by group of V. Likić, Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Australia	http://code.google.com/p/pyms/
XCMS	Scripps Research Institute, USA	[78] http://metlin.scripps.edu/xcms/

(column bleed) appear in nearly every plant chromatogram after derivatization with MTSFA and therefore have to be excluded from further analysis. Leave the box for target component searching unchecked, and specify a library and the confidence threshold for targeted analysis. *Note*: It is important to adjust peak picking parameters according to the quality of the chromatogram (e.g., peak width, signal-to-noise ratio, and resolution) to be able to pick as many components as possible that are present in the biological sample; furthermore, deconvolution and peak picking parameters have to be optimized to avoid false positives (for a review on the quality of peak picking using different software programs, see [79]).

4) Generate a target component library (TCL) of the pooled reference sample file with the most deconvoluted components. The TCL contains a list of (identified) components of this representative chromatogram and has to be additionally specified by the target ion and ion ratios of the second and third most abundant ions (fill in manually). Additionally, perform background subtraction by choosing a blank sample data file to remove contaminants and components not present in the biological sample.
5) Enable and configure the Matrix Analyzer plug-in. Enable the box for target component searching. Match all components found in the other chromatograms against the TCL using the same initial parameter settings.
6) After processing of all data files, the Matrix Analyzer plug-in report can be accessed via the "reports" tab. Save the data matrix in one of the specified formats (∗.csv or ∗.xls) for further data mining (see Section 1.2.4.2).
7) Control the quality of the raw data. Ensure that peaks are accurately identified and peak areas are correctly integrated.
8) Normalize the data by dividing the integrated peak areas of all detected metabolites by the peak area of the internal standard and by the sample weight (in grams).

1.2.4.2 **Data Analysis and Statistics**

Following data preprocessing and normalization, the data are typically logarithm transformed to minimize possible effects of outliers [80,19]. Subsequently, effective statistical discriminant analysis is applied to the data set to extract biologically relevant information. This aims to find patterns or relationships within the data to extract the information needed to generate scientific hypotheses, which have to be further tested using Student's *t*-test and analysis of variance (ANOVA). Metabolite data can be mined using different pattern recognition methods to separate the data into classes, either knowing that classes exist, using supervised learning algorithms, or in the absence of any advanced knowledge, using unsupervised learning algorithms [81].

Univariate analysis is the simplest statistical method and is carried out with only one variable at a time. Basic univariate statistical measures are mean, variance, standard deviation, covariance, and correlation [82]. Multivariate statistics deal with the analysis of multiple variables simultaneously, and include unsupervised classification methods such as PCA, hierarchical cluster analysis (HCA), and self-

organization mapping (SOM) and supervised approaches such as partial least squares (PLS) to classify metabolites [15]. HCA and PCA are most widely used for comparison and visualization of similarities and differences between data sets. Additionally, tools displaying data sets on metabolic pathway maps are often used to visualize metabolic profiles, and can also be combined with gene expression profiles [83].

1.2.4.3 Procedure for Postacquisition Data Analysis

There is a huge amount of different commercially and freely available software packages to explore data sets statistically. Many statistical tests and classification methods, including PCA, PLS, and HCA, can be performed using The Unscrambler statistical software (CAMO) or using scripts integrated in the R software environment (www.r-project.org). Furthermore, normalization using internal standards and sample weight, log transformation, and statistical analyses can be performed using designed R scripts and Excel macros that are well documented and freely available at Metabolomics Australia (http://code.google.com/p/ma-bioinformatics/). Only recently have web-based metabolomic data tools such as MetaboAnalyst been made available; this combines several complex data analysis techniques including data processing, normalization, statistics, and pathway mapping and is freely available on a web server (www.metaboanalyst.ca; [84]). Further information is available in separate chapters on data analysis and multivariate statistics (Sun and Weckwerth, Chapter 16) and metabolite clustering and visualization (Kaever *et al.*, Chapter 14).

1.3 Applications of the Technology

Numerous applications have been reported in which GC–MS-based metabolomics has been used to investigate metabolites and pathways that are differentially regulated due to genetic or environmental perturbations. There have been extensive reports and reviews describing how GC–MS-based metabolite profiling has been employed to study plant metabolism in great detail [19,50,85]. Here, we mention just a few examples of research areas where metabolomics has already made a contribution.

Metabolite profiles generated by GC–MS can be used as biochemical readouts to classify organisms according to genetic and environmental stimuli and to identify the differences and similarities between the different conditions. As described in this chapter, GC–MS can generate hundreds of data points and, regardless of whether those data points can be referred to a known metabolite or not, the presence and relative abundance of those data points can be related to genetic background and environmental conditions similarly to a signature. DNA sequences are still the standard used for the identification of genetically different individuals. However, it is known that the biochemical readout of individuals even with similar genomes will be different with environmental changes. Therefore, metabolomics has already been successfully applied to classify genetically similar individuals grown at different

locations (provenance) or under different conditions. An example where metabolomics has been used to determine the geographical origin of samples was presented by Choi *et al.* [86], who used ^{1}H NMR-based metabolite profiling in combination with multivariate analysis to classify 12 *Cannabis sativa* cultivars based on the region they were grown (see also Chapter 3 by Choi *et al.*). Extensive studies on comparisons of metabolite profiles of plants grown in different conditions have been carried out. For example, the metabolomes of plants grown in unfavorable conditions such as abiotic and biotic stresses can increase our understanding of how plants respond and adapt to harsh environments. Researchers aim to understand how plants have evolved mechanisms to deal with stress and especially how some plants perform better than others. Abiotic stresses including cold, frost, heat, drought, and salinity cause massive losses in crop yields every year. An understanding of stress tolerance mechanisms and the transfer of those mechanisms to commercial crop varieties will reduce agricultural losses. Contributions made by metabolomics approaches to learning about the physiology and biochemistry of plants in different stress conditions have been reported, for example, for cold and heat stress in *Arabidopsis* [87], salinity in rice [88], *Lotus japonicus* [89], and barley [90], and water deficiency in *Arabidopsis* [91].

Metabolomics as a tool to characterize a plant chemically is becoming increasingly important for risk assessments of genetic modifications. Genetic alterations can have an impact not only on the visible phenotype but also on the biochemical composition of the cells, potentially leading to effects that are unexpected on the basis of current genetic or biochemical knowledge [92,93]. There have been a number of reports where the introduction or deletion of a gene has altered metabolism and therefore metabolite concentrations compared with wild-type controls [32,33,43]. It has also been demonstrated that the introduction of the same gene into different species could result in differential changes of the metabolomes [43]. A substantial equivalence concept is a framework for safety evaluations where existing crops and foods are taken as the baseline considered as being safe, and the properties of any new foods and crops are compared with the baseline. Therefore, it is important to monitor the metabolomes (and all other cell products) of genetically engineered plants and compare them with the natural variation of metabolomes of their wild-type counterparts [93,94].

The last example mentioned here is the application of metabolomics in breeding and quantitative trait loci (QTLs) analysis, which is recognized to have enormous potential. Often agronomic traits are controlled by many genes or QTLs potentially residing on different chromosomes but their expression works together as a network determining that particular phenotype or trait. Especially if the trait of interest is based on a metabolite of interest (e.g., vitamins or essential amino acids), the utilization of metabolomics as a strategy to link phenotypes with QTLs has already been demonstrated in a number of different species [46,47,95]. Now that metabolomics technologies have become faster and cheaper, it is possible to analyze huge numbers of compounds simultaneously in a large genetic mapping population. This new approach of combining conventional genetic methods such as QTL mapping with omics technologies such as transcriptomics, proteomics, and metabolomics,

also called genetic genomics, will allow the assessment of a large number of traits simultaneously and ultimately the identification of the function of underlying gene networks [58].

1.4 Perspectives

GC–MS-based metabolomics is still considered the workhorse for metabolite profiling of plants upon changing conditions, with numerous advantages over other analytical platforms such as robustness and high separation power and reproducibility. However, we need to be open to new and emerging developments utilizing GC–MS technology that will improve the current capabilities of GC–MS plant metabolomics. These improvements can be manifold, for example, increasing the number of compounds detectable as well as identifiable or increasing the speed of analysis for a higher throughput. Multidimensional or GC × GC–MS has already been successfully introduced to and applied in a number of metabolomics applications. The technology is well established for the analysis of volatiles in, for instance, wine [96], oil components of different origin [97,98], and fragrances [99]. So far, the analysis of semivolatiles or nonvolatiles in plant extracts using two-dimensional GC–MS has not been explored. However, first successes have been reported in the medical field [100].

To increase the speed of GC–MS analyses without reducing the separation power or deconvolution efficiency, fast-scanning mass spectrometers have been introduced in combination with fast heating and cooling GC ovens. Before utilizing fast GC or MS technology for any plant metabolomics applications, a careful investigation of the balance between the time of analysis and the number of sufficiently separated compounds needs to be carried out. Plant-derived extracts are extremely complex and also contain a huge numbers of sugars, including mono-, di-, and trisaccharides, which often produce the same or very similar mass spectra. This means that these compounds can only be separated chromatographically in order to identify and quantify them with confidence. However, increasing the rate of temperature change may reduce the separation power and therefore increase the coelution of compounds with similar mass spectra, so the trade-off between speed and number of detectable compounds needs to be established.

Additionally, to improve the GC–MS technology for better derived data, GC–MS only allows the detection of a few hundred metabolites and therefore limits the picture to be drawn in a biological and biochemical context. To increase our understanding of the biological system in question, as many metabolites as possible need to be analyzed. Therefore, GC–MS technology is a complementary analytical platform to others such as LC–MS, capillary electrophoresis (CE)–MS, and NMR spectroscopy [19]. Finally, to understand the system in question from a holistic viewpoint, it is important to interrogate metabolomics data with any other measurable traits such as genome sequence, transcript and protein expression, metabolic

fluxes, anatomic and physiological parameters, and also growth and performance upon any genetic or environmental stimuli.

Acknowledgments

The authors thank the Australian Centre for Plant Functional Genomics (ACPFG), which is funded by grants from the Australian Research Council (ARC) and the Grains Research and Development Corporation (GRDC), the South Australian Government, and the University of Adelaide, the University of Queensland, and The University of Melbourne. C.B.H. is a recipient of a Melbourne International Fee Remission Scholarship (MIFRS) and a Melbourne International Research Scholarship (MIRS). U.R. is also grateful to the Victorian Node of Metabolomics Australia, which is funded through Bioplatforms Australia Pty Ltd., a National Collaborative Research Infrastructure Strategy (NCRIS), Biomolecular Platforms and coinvestment from the Victorian State government and The University of Melbourne.

References

1 Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **408**, 796–815.

2 Yu, J., Hu, S., Wang, J., Wong, G.K., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., Cao, M., Liu, J., Sun, J., Tang, J., Chen, Y., Huang, X., Lin, W., Ye, C., Tong, W., Cong, L., Geng, J., Han, Y., Li, L., Li, W., Hu, G., Huang, X., Li, W., Li, J., Liu, Z., Li, L., Liu, J., Qi, Q., Liu, J., Li, L., Li, T., Wang, X., Lu, H., Wu, T., Zhu, M., Ni, P., Han, H., Dong, W., Ren, X., Feng, X., Cui, P., Li, X., Wang, H., Xu, X., Zhai, W., Xu, Z., Zhang, J., He, S., Zhang, J., Xu, J., Zhang, K., Zheng, X., Dong, J., Zeng, W., Tao, L., Ye, J., Tan, J., Ren, X., Chen, X., He, J., Liu, D., Tian, W., Tian, C., Xia, H., Bao, Q., Li, G., Gao, H., Cao, T., Wang, J., Zhao, W., Li, P., Chen, W., Wang, X., Zhang, Y., Hu, J., Wang, J., Liu, S., Yang, J., Zhang, G., Xiong, Y., Li, Z., Mao, L., Zhou, C., Zhu, Z., Chen, R., Hao, B., Zheng, W., Chen, S., Guo, W., Li, G., Liu, S., Tao, M., Wang, J., Zhu, L., Yuan, L., and Yang, H. (2002) A draft sequence of the rice genome (*Oryza sativa* L. sp. *indica*). *Science*, **296**, 79–92.

3 The International Brachypodium Initiative (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature*, **463**, 763–768.

4 Fowler, S. and Thomashow, M.F. (2002) *Arabidopsis* transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. *Plant Cell*, **14**, 1675–1690.

5 The Rice Full-Length, cDNA, Consortium (2003) Collection, mapping, and annotation of over 28,000 cDNA clones from *japonica* rice. *Science*, **301**, 376–379.

6 Zhang, H., Sreenivasulu, N., Weschke, W., Stein, N., Rudd, S., Radchuk, V., Potokina, E., Scholz, U., Schweizer, P., Zierold, U., Langridge, P., Varshney, R., Wobus, U., and Graner, A. (2004) Large-scale analysis of the barley transcriptome based on expressed sequence tags. *Plant J.*, **40**, 276–290.

7 Froehlich, J.E., Wilkerson, C.G., Ray, W.K., McAndrew, R.S., Osteryoung, K.W., Gage, D.A., and Phinney, B.S. (2003) Proteomic study of the *Arabidopsis thaliana* chloroplastic envelope membrane utilizing alternatives to traditional two-dimensional electrophoresis. *J. Proteome Res.*, **2**, 413–425.

8 Santoni, V., Vingh, J., Pflieger, D., Sommerer, N., and Maurel, C. (2003) A proteomic study reveals novel insights into the diversity of aquaporin forms expressed in the plasma membrane of plant roots. *Biochem. J.*, **373**, 289–296.

9 Hajduch, M., Ganapathy, A., Stein, J.W., and Thelen, J.J. (2005) A systematic proteomic study of seed filling in soybean. Establishment of high-resolution two-dimensional reference maps, expression profiles, and an interactive proteome database. *Plant Physiol.*, **137**, 1397–1419.

10 Oksman-Caldentey, K.-M. and Inzé, D. (2004) Plant cell factories in the post-genomic era: new ways to produce designer secondary metabolites. *Trends Plant Sci.*, **9**, 433–440.

11 Last, R.L., Jones, A.D., and Shachar-Hill, Y. (2007) Towards the plant metabolome and beyond. *Nat. Rev.*, **8**, 167–174.

12 Fiehn, O. (2001) Combining genomics, metabolome analysis, and biochemical modeling to understand metabolic networks. *Comp. Funct. Genomics*, **2**, 155–168.

13 Fiehn, O. (2002) Metabolomics – the link between genotypes and phenotypes. *Plant Mol. Biol.*, **48**, 155–171.

14 Nielsen, J. and Oliver, S. (2005) The next wave in metabolome analysis. *Trends Biotechnol.*, **2** (11), 544–546.

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

16 Dunn, W.B. and Ellis, D.I. (2005) Metabolomics: current analytical platforms and methodologies. *Trends Anal. Chem.*, **24**, 285–294.

17 Dettmer, K., Aronov, P.A., and Hammock, B.D. (2007) Mass spectrometry-based metabolomics. *Mass Spectrom. Rev.*, **26**, 51–78.

18 Roessner, U. and Beckles, D.M. (2009) Metabolite measurements, in *Plant Metabolic Networks* (ed. J. Schwender), Springer Science+Business Media, Berlin.

19 Roessner, U. and Beckles, D.M. (2011) Plant metabolomics—applications and opportunities for agricultural biotechnology, in *Plant Biotechnology and Agriculture: Prospects for the 21st Century* (eds A. Altmann and P.M. Haegawa), Elsevier, Amsterdam.

20 Fiehn, O., Kopka, J., Doermann, P., Altmann, T., Trethewey, R.N., and Willmitzer, L. (2000) Metabolite profiling for plant functional genomics. *Nat. Biotechnol.*, **18**, 1157–1161.

21 Roessner, U., Wagner, C., Kopka, J., Trethewey, R.N., and Willmitzer, L. (2000) Simultaneous analysis of metabolites in potato tuber by gas chromatography–mass spectrometry. *Plant J.*, **23**, 131–142.

22 Sumner, L.W., Mendes, P., and Dixon, R.A. (2003) Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry*, **62**, 817–836.

23 Harrigan, G.G. and Goodacre, R. (2003) Metabolic profiling: pathways in drug discovery, in *Metabolic Profiling: Its Role in Biomarker Discovery and Gene Function Analysis* (eds G.G. Harrigan and R. Goodacre), Kluwer, Dordrechr.

24 Kanani, H., Chrysanthopoulos, P.K., and Klapa, M.I. (2008) Standardizing GC–MS metabolomics. *J. Chromatogr. B*, **871**, 191–201.

25 Halket, J.M., Waterman, D., Przyborowska, A.M., Patel, R.K.P., Fraser, P.D., and Bramley, P.M. (2004) Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. *J. Exp. Bot.*, **56**, 1–25.

26 Roda, A. (2010) ABC presents nobel prize winners in chemistry. *Anal. Bioanal. Chem.*, **396**, 1615–1617.

27 Martin, A.J.P. and Synge, R.L.M. (1941) Applications of partition chromatography. *Biochem. J.*, **35**, 1358–1366.

28 Devaux, P.G., Horning, M.G., and Horning, E.C. (1971) Benyzl-oxime derivatives of steroids; a new metabolic profile procedure for human urinary steroids. *Anal. Lett.*, **4**, 70–82.

29 Horning, E.C. and Horning, M.G. (1971) Human metabolic profiles obtained by GC and GC/MS. *J. Chromatogr. Sci.*, **9**, 129–140.

30 Pauling, L., Robinson, A.B., Teranishi, R., and Cary, P. (1971) Quantitative analysis of urine vapor and breath by gas–liquid partition chromatography. *Proc. Natl. Acad. Sci. U. S. A.*, **68**, 2374–2376.

31 Oliver, S.G., Winson, M.K., Kell, D.B., and Baganz, F. (1998) Systematic functional analysis of the yeast genome. *Trends Biotechnol.*, **16**, 373–378.

32 Roessner, U., Luedemann, A., Brust, D., Fiehn, O., Linke, T., Willmitzer, L., and Fernie, A.R. (2001) Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell*, **13**, 11–29.

33 Roessner, U., Willmitzer, L., and Fernie, A.R. (2001) High-resolution metabolic phenotyping of genetically and environmentally diverse potato tuber systems. Identification of phenocopies. *Plant Physiol.*, **127**, 749–764.

34 Tolstikov, V.V. and Fiehn, O. (2002) Analysis of highly polar compounds of plant origin: combination of hydrophilic interaction chromatography and electrospray ion trap mass spectrometry. *Anal. Biochem.*, **301**, 298–307.

35 Roessner-Tunali, U., Hegemann, B., Lytovchenko, A., Carrari, F., Bruedigam, C., Granot, D., and Fernie, A.R. (2003) Metabolic profiling of transgenic tomato plants overexpressing hexokinase reveals that the influence of hexose phosphorylation diminishes during fruit development. *Plant Physiol.*, **133**, 84–99.

36 Schauer, N., Zamir, D., and Fernie, A.R. (2004) Metabolic profiling of leaves and fruit of wild species tomato: a survey of the *Solanum lycopersicum* complex. *J. Exp. Bot.*, **56**, 1–11.

37 Widodo, W., Patterson, J.H., Newbigin, E., Tester, M., Bacic, A., and Roessner, U. (2009) Metabolic responses to salt stress of barley (*Hordeum vulgare* L.) cultivars, Sahara and Clipper, which differ in salinity tolerance. *J. Exp. Bot.*, **60**, 1–15.

38 Broz, A.K., Broeckling, C.D., De-la-Peña, C., Lewis, M.R., Greene, E., Callaway, R. M., Sumner, L.W., and Vivanco, J.M. (2010) Plant neighbor identity influences plant biochemistry and physiology related to defense. *BMC Plant Biol.*, **10**, 115.

39 Roessner, U., Patterson, J.H., Forbes, M. G., Fincher, G.B., Langridge, P., and Bacic, A. (2006) An investigation of boron toxicity in barley using metabolomics. *Plant Physiol.*, **142**, 1087–1101.

40 Hernández, G., Ramőrez, M., Valdés-López, O., Tesfaye, M., Graham, M.A., Czechowski, T., Schlereth, A., Wandrey, M., Erban, A., Cheung, F., Wu, H.C., Lara, M., Town, C.D., Kopka, J., Udvardi, M.K., and Vance, C.P. (2007) Phosphorus stress in common bean: root transcript and metabolic responses. *Plant Physiol.*, **144**, 752–767.

41 Urano, K., Maruyama, K., Ogata, Y., Morishita, Y., Takeda, M., Sakurai, N., Suzuki, H., Saito, K., Shibata, D., Kobayashi, M., Yamaguchi-Shinozaki, K., and Shinozaki, K. (2009) Characterization of the ABA-regulated global responses to dehydration in *Arabidopsis* to metabolomics. *Plant J.*, **57**, 1065–1078.

42 Lugan, R., Niogret, M.-F., Leport, L., Guégan, J.-P., Larher, F.R., Savouré, A., Kopka, J., and Bouchereau, A. (2010) Metabolome and water homeostasis analysis of *Thellungiella salsuginea* suggests that dehydration tolerance is a key response to osmotic stress in this halophyte. *Plant J.*, **64**, 215–229.

43 Jacobs, A., Lunde, C., Bacic, A., Tester, M., and Roessner, U. (2007) The impact of constitutive heterologous expression of moss Na^+ transporter on the metabolomes of rice and barley. *Metabolomics*, **3**, 307–317.

44 Stamova, B.S., Roessner, U., Suren, S., Laudencia-Chingcuanco, D., Bacic, A., and Beckles, D.M. (2009) Metabolic profiling of transgenic wheat over-expressing the high-molecular-weight Dx5 glutenin subunit. *Metabolomics*, **5**, 239–252.

45 Kliebenstein, D.J., Gershenzon, J., and Mitchell-Olds, T. (2001) Comparative quantitative trait loci mapping of aliphatic, indolic and benzylic glucosinolate production in *Arabidopsis thaliana* leaves and seeds. *Genetics*, **159**, 359–370.

46 Schauer, N., Semel, Y., Roessner, U., Gur, A., Balbo, I., Carrari, F., Pleban, T., Perez-Melis, A., Breudigam, C., Kopka, J., Willmitzer, L., Zamir, D., and Fernie, A.R. (2006) Comprehensive metabolite profiling and phenotyping of interspecific introgression lines for tomato

improvement. *Nat. Biotechnol.*, **24**, 447–455.

47 Lisec, J., Meyer, R.C., Steinfath, M., Redestig, H., Becher, M., Witucka-Wall, H., Fiehn, O., Törjék, O., Selbig, J., Altmann, T., and Willmitzer, L. (2008) Identification of metabolic and biomass QTL in *Arabidopsis thaliana* in a parallel analysis of RIL and IL populations. *Plant J.*, **53**, 960–972.

48 Zanor, M.I., Rambla, J.-L., Chaõb, J., Steppa, A., Medina, A., Granell, A., Fernie, A.R., and Causse, M. (2009) Metabolic characterization of loci affecting sensory attributes in tomato allows an assessment of the influence of the levels of primary metabolites and volatile organic contents. *J. Exp. Bot.*, **60**, 2139–2154.

49 Spiller, M., Berger, R.G., and Debener, T. (2010) Genetic dissection of scent metabolic profiles in diploid rose populations. *Theor. Appl. Genet.*, **120**, 1461–1471.

50 Kopka, J., Fernie, A., Weckwerth, W., Gibon, Y., and Stitt, M. (2004) Metabolite profiling in plant biology: platforms and destinations. *Genome Biol.*, **5**, 109.

51 De Vos, R.C.H., Moco, S., Lommen, A., Keurentjes, J.J.B., Bino, R.J., and Hall, R.D. (2007) Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nat. Protoc.*, **2**, 778–791.

52 Verdonk, J.C., de Vos, R., Verhoeven, H.A., Haring, M.A., van Tunen, A.J., and Schuurink, R.C. (2003) Regulation of floral scent production in petunia revealed by targeted metabolomics. *Phytochemistry*, **62**, 997–1008.

53 Daferera, D.J., Ziogas, B.N., and Polissiou, M.G. (2000) Analysis of essential oils from some Greek aromatic plants and their fungitoxicity on *Penicillium digitatum*. *J. Agric. Food Chem.*, **48**, 2576–2581.

54 Shulaev, V. (2006) Metabolomics technology and bioinformatics. *Brief. Bioinform.*, **7**, 128–139.

55 Roessner, U. and Bacic, A. (2009) Metabolomics in plant research. *Aust. Biochem.*, **40**, 9–20.

56 Villas-Bôas, S.G. (2007) Sampling and sample preparation, in *Metabolome Analysis – An Introduction* (eds S.G. Villas-Bôas, U. Roessner, M.A.E. Hansen, J. Smedsgaard, and J. Nielsen), John Wiley & Sons, Inc., Hoboken, NJ.

57 Lisec, J., Schauer, N., Kopka, J., Willmitzer, L., and Fernie, A.R. (2006) Gas chromatography–mass spectrometry-based metabolite profiling in plants. *Nat. Protoc.*, **1**, 387–396.

58 Kliebenstein, D.J. (2007) Metabolomics and plant quantitative trait locus analysis – the optimal genetical genomics platform? in *Concepts in Plant Metabolomics* (eds B.J. Nikolau and E.S. Wurtele), Springer, Berlin, pp. 29–45.

59 Zoerb, C., Langenkämper, G., Betsche, T., Niehaus, K., and Barsch, A. (2006) Metabolite profiling of wheat grains (*Triticum aestivum* L.) from organic and conventional agriculture. *J. Agric. Food Chem.*, **54**, 8301–8306.

60 Orth, H.C.J., Rentel, C., and Schmidt, P.C. (1999) Isolation, purity analysis and stability of hyperforin as a standard material from *Hypericum perforatum* L. *J. Pharm. Pharmacol.*, **51**, 193–200.

61 Sangster, T., Major, H., Plumb, R., Wilson, A.J., and Wilson, I.D. (2006) A pragmatic and readily implemented quality control strategy for HPLC–MS and GC–MS-based metabonomic analysis. *Analyst*, **131**, 1075–1078.

62 Sobolevsky, T.G., Revelsky, A.I., Miller, B., Oriedo, V., Chernetsova, E.S., and Revelsky, I.A. (2003) Comparison of silylation and esterification/acylation procedures in GC–MS analysis of amino acids. *J. Sep. Sci.*, **26**, 1474–1478.

63 Birkemeyer, C., Kolasa, A., and Kopka, J. (2003) Comprehensive chemical derivatization for gas chromatography–mass spectrometry-based multi-targeted profiling of the major phytohormones. *J. Chromatogr. A*, **993**, 89–102.

64 Smedsgaard, J. (2007) Analytical tools, in *Metabolite Analysis: an Introduction* (eds S. Villas-Bôas, U. Roessner, M.A.E. Hansen, J. Smedsgaard, and J. Nielsen), John Wiley & Sons, Inc., Hoboken, NJ, p. 95.

65 De Hoffmann, E. (2005) *Mass Spectrometry. Kirk-Othmer Encyclopedia of*

Chemical Technology, John Wiley & Sons, Inc., Hoboken, NJ.

66 Bedair, M. and Sumner, L.W. (2008) Current and emerging mass-spectrometry technologies for metabolomics. *Trends Anal. Chem.*, **27**, 238–250.

67 Adahchour, M., Brandt, M., Baier, H.U., Vreuls, R.J., Batenburg, A.M., and Brinkman, U.A. (2005) Comprehensive two-dimensional gas chromatography coupled to a rapid-scanning quadrupole mass spectrometer: principles and applications. *J. Chromatogr. A*, **1067**, 245–254.

68 Aebersold, R. and Mann, M. (2009) Mass spectrometry-based proteomics. *Nature*, **422**, 198–207.

69 Brown, M., Dunn, W.B., Ellis, D.I., Handl, J., Knowls, J.D., O'Hagan, S., Spasić, I., and Kell, D.B. (2004) A metabolome pipeline: from concept to data to knowledge. *Metabolomics*, **1**, 39–51.

70 Sysi-Aho, M., Katajamaa, M., Yetukuri, L., and Orešič, M. (2007) Normalization method for metabolomics data using optimal selection of multiple internal standards. *BMC Bioinform.*, **8**, 93.

71 Shuman, J.L., Cortes, D.F., Armenta, J.M., Pokrzywa, R.M., Mendes, P., and Shuaev, V. (2010) Plant metabolomics by GC–MS and differential analysis, in *Plant Reverse Genetics: Methods and Protocols* (ed. A. Pereira), Methods in Molecular Biology, vol. **678**, Springer Science + Business Media, Berlin, pp. 232–244.

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

73 Wagner, C., Sefkow, M., and Kopka, J. (2003) Construction and application of a mass spectral and retention time index database generated from plant GC/EI-TOF-MS metabolite profiles. *Phytochemistry*, **62**, 887–900.

74 Stein, S.E. (1999) An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. *J. Am. Soc. Mass Spectrom.*, **10**, 770–781.

75 Lommen, A. (2009) MetAlign: interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. *Anal. Chem.*, **81**, 3079–3086.

76 Katajamaa, M., Miettinen, J., and Oresic, M. (2006) MZmine: toolbox for processing and visualization of mass spectrometry-based molecular profile data. *Bioinformatics*, **22**, 634–636.

77 Pluskal, T., Castillo, S., Villar-Briones, A., and Oresic, M. (2010) MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinform.*, **11**, 395.

78 Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R., and Siuzdak, G. (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.*, **78**, 779–787.

79 Lu, H., Dunn, W.B., Kell, D.B., and Liang, Y. (2008) Comparative evaluation of software for deconvolution of metabolomics data based on GC–TOF-MS. *Trends Anal. Chem.*, **27**, 215–227.

80 Fait, A. and Fernie, A.R. (2009) Data integration, in *Plant Metabolic Networks* (ed. J. Schwender), Springer Science+ Business Media, Berlin, pp. 151–173.

81 Nobeli, I. and Thornton, J.M. (2006) A bioinformatician's view of the metabolome. *BioEssays*, **28**, 534–545.

82 Esbensen, K.H. (2004) *Multivariate Data Analysis – In Practice*, Camo Software, Oslo.

83 Thimm, O., Blasing, O., Gibon, Y., Nagel, A., Meyer, S., Kruger, P., Selbig, J., Mueller, L.A., Rhee, S.Y., and Stitt, M. (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.*, **37**, 914–939.

84 Xia, J., Psychogios, N., Young, N., and Wishart, D.S. (2009) MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Res.*, 37, W352–W660.

85 Bowne, J., Bacic, A., Tester, M., and Roessner, U. (2011) Abiotic stress and metabolomics, in *Biology of Plant Metabolomics*, 1st edn (ed. R. Hall), Annual Plant Reviews, vol. **43**, Blackwell, Oxford, pp. 61–85.

86 Choi, Y.H., Kim, H.K., Hazekamp, A., Erkelens, C., Lefeber, A.W., and Verpoorte, R. (2004) Metabolomic differentiation of *Cannabis sativa* cultivars using ^{1}H NMR spectroscopy and principal component analysis. *J. Nat. Prod.*, **67**, 953–957.

87 Kaplan, F., Kopka, J., Haskell, D.W., Zhao, W., Schiller, K.C., Gatzke, N., Sung, D.Y., and Guy, C.L. (2004) Exploring the temperature-stress metabolome of *Arabidopsis*. *Plant Physiol.*, **136**, 4159–4168.

88 Zuther, E., Koehl, K., and Kopka, J. (2007) Comparative metabolome analysis of the salt response in breeding cultivars of rice, in *Advances in Molecular-Breeding Toward Drought and Salt Tolerance Crops* (eds M.A. Jenks, P.M. Hasegawa, and S.M. Jain), Springer, Berlin, pp. 285–315.

89 Sanchez, D.H., Lippold, F., Redestig, H., Hannah, M.A., Erban, A., Krämer, U., Kopka, J., and Udvardi, M.K. (2008) Integrative functional genomics of salt acclimatization in the model legume *Lotus japonicus*. *Plant J.*, **53**, 973–987.

90 Widodo, W., Patterson, J.H., Newbigin, E., Tester, M., Bacic, A., and Roessner, U. (2009) Metabolic responses to salt stress of barley (*Hordeum vulgare* L.) cultivars, Sahara and Clipper, which differ in salinity tolerance. *J. Exp. Bot.*, **60**, 4089–4103.

91 Urano, K., Maruyama, K., Ogata, Y., Morishita, Y., Takeda, M., Sakurai, N., Suzuki, H., Saito, K., Shibata, D., Kobayashi, M., Yamaguchi-Shinozaki, K., and Shinozaki, K. (2009) Characterization of the ABA-regulated global responses to dehydration in *Arabidopsis* by metabolomics. *Plant J.*, **57**, 1065–1078.

92 Cellini, F., Chesson, A., Colquhoun, I., Constable, A., Davies, H.V., Engel, K.H., Gatehouse, A.M., Karenlampi, S., Kok, E.J., Leguay, J.J., Lehesranta, S., Noteborn, H.P., Pedersen, J., and Smith, M. (2004) Unintended effects and their detection in genetically modified crops. *Food Chem. Toxicol.*, **42**, 1089–1125.

93 Kuiper, H.A., Kok, E.J., and Engel, K.H. (2003) Exploitation of molecular profiling techniques for GM food safety assessment. *Curr. Opin. Biotechnol.*, **14**, 238–243.

94 Beale, M.H., Ward, J.L., and Baker, J.M. (2009) Establishing substantial equivalence: metabolomics. *Methods Mol. Biol.*, **478**, 289–303.

95 Keurentjes, J.J.B., Fu, J., Ric de Vos, C.H., Lommen, A., Hall, R.D., Bino, R.J., van der Plas, L.H.W., Jansen, R.C., Vreugdenhil, D., and Koornneef, M. (2006) The genetics of plant metabolism. *Nat. Genet.*, **38**, 842–849.

96 Robinson, A.L., Boss, P.K., Heymann, H., Solomon, P.S., and Trengove, R.D. (2011) Development of a sensitive non-targeted method for characterizing the wine volatile profile using headspace solid-phase microextraction comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry. *J. Chromatogr. A*, **1218**, 504–517.

97 Di, X., Shellie, R.A., Marriott, P.J., and Huie, C.W. (2004) Application of headspace solid-phase microextraction (HS-SPME) and comprehensive two-dimensional gas chromatography (GC×GC) for the chemical profiling of volatile oils in complex herbal mixtures. *J. Sep. Sci.*, **27**, 451–458.

98 Tranchida, P.Q., Shellie, R.A., Purcaro, G., Conte, L.S., Dugo, P., Dugo, G., and Mondello, L. (2010) Analysis of fresh and aged tea tree essential oils by using GC×GC–qMS. *J. Chromatogr. Sci.*, **48**, 262–266.

99 Dunn, M.S., Vulic, N., Shellie, R.A., Whitehead, S., Morrison, P., and Marriott, P.J. (2006) Targeted multidimensional gas chromatography for the quantitative analysis of suspected allergens in fragrance products. *J. Chromatogr. A*, **1130**, 122–129.

100 Koek, M.M., van der Kloet, F.M., Kleemann, R., Kooistra, T., Verheij, E.R., and Hankemeier, T. (2011) Semi-automated non-target processing in GC×GC–MS metabolomics analysis: applicability for biomedical studies. *Metabolomics*, **7**, 1–14.