Physiologia Plantarum 132: 117–135. 2008

REVIEW

Metabolomic technologies and their application to the study of plants and plant–host interactions

J. William Allwood*, David I. Ellis and Royston Goodacre

School of Chemistry, Manchester Interdisciplinary Biocentre, The University of Manchester, 131 Princess Street, Manchester M1 7DN, UK

Correspondence
*Corresponding author,
e-mail: william.allwood@manchester.ac.uk

Received 31 August 2007; revised 10 October 2007

doi: 10.1111/j.1399-3054.2007.01001.x

Metabolomics is perhaps the ultimate level of post-genomic analysis as it can reveal changes in metabolite fluxes that are controlled by only minor changes within gene expression measured using transcriptomics and/or by analysing the proteome that elucidates post-translational control over enzyme activity. Metabolic change is a major feature of plant genetic modification and plant interactions with pathogens, pests, and their environment. In the assessment of genetically modified plant tissues, metabolomics has been used extensively to explore by-products resulting from transgene expression and scenarios of substantial equivalence. Many studies have concentrated on the physiological development of plant tissues as well as on the stress responses involved in heat shock or treatment with stress-eliciting molecules such as methyl jasmonic acid, yeast elicitor or bacterial lipopolysaccharide. Plant–host interactions represent one of the most biochemically complex and challenging scenarios that are currently being assessed by metabolomic approaches. For example, the mixtures of pathogen-colonised and non-challenged plant cells represent an extremely heterogeneous and biochemically rich sample; there is also the further complication of identifying which metabolites are derived from the plant host and which are from the interacting pathogen. This review will present an overview of the analytical instrumentation currently applied to plant metabolomic analysis, literature within the field will be reviewed paying particular regard to studies based on plant–host interactions and finally the future prospects on the metabolomic analysis of plants and plant–host interactions will be discussed.

Introduction

The constituents of the metabolome reflect the cellular processes that control the biochemical phenotype of the cell, tissue or whole organism. Measurements of intracellular metabolites, whether qualitative or quantitative, reveal the biochemical status of an organism and in turn can be used to monitor and assess gene function (Fiehn et al. 2000a). Within functional genomics, metabolomics

Abbreviations – APCI, atmospheric pressure chemical ionisation; ATR, attenuated total reflectance; CE, capillary electrophoresis; DAD, diode array detector; DFA, discriminant function analysis; DI, direct injection/infusion; EI, electron impact; EM, electromagnetic; ESI, electrospray ionisation; FI, flow injection/infusion; FT, Fourier transform; GA, genetic algorithm; GM, genetically modified; GP, genetic programming; ICR, ion cyclotron resonance; JA, jasmonic acid; LA, linolenic acid; LC, liquid chromatography; MAS, magic angle spinning; MeJA, methyl jasmonic acid; MS-MS, tandem MS; MW, molecular weight; PC, principal component; PCA, principal component analysis; PCD, programmed cell death; Q, quadrupole; Re, reflectron; SA, salicylic acid; SAR, systemic acquired resistance; TMV, tobacco mosaic virus; TOF, time of flight; UHPLC, ultra-high-pressure liquid chromatography.

may emerge as a more robust approach to predict gene activity than the currently more widely used transcriptomic and proteomic approaches. Studies in yeast using transcriptome and proteome data have not always predicted gene function accurately, with mRNA and protein levels often being poorly correlated (Gygi et al. 1999). Furthermore, once a protein has been synthesised it may or may not be enzymatically active, and as a result of allosteric control changes in either transcription or protein levels often do not correlate with those of the metabolome.

The ability to collect extremely large data sets with metabolomic approaches from carefully designed experiments permits the development of hypotheses by inductive reasoning (Kell and Oliver 2004). The quality of inductive approaches is improved by increasing the size of the data set. However, at present, it is impossible to quantify all the metabolites within a cell, tissue or organism of any kingdom in any system. The problem lies within the fact that not any single extraction technique or analytical instrument can isolate and detect every metabolite within a biological sample (Sumner et al. 2003). These problems are further compounded by issues such as human error in sample preparation and extraction, sample storage and instrument reproducibility (Dunn and Ellis 2005). To tackle this problem, the Metabolomics Standards Initiative (Castle et al. 2006, Fiehn et al. 2007, Sansone et al. 2007; and see http://msi-workgroups.sourceforge.net/) has started to discuss the standardisation of metabolomic experiments from growth of materials and their collection to chemical analysis and data processing.

Metabolomics is further complicated by the huge diversity of metabolites in any given species. This is especially the case in the plant kingdom that encompasses an estimated 100 000–200 000 metabolites (Oksman-Caldentey and Inzé 2004). This poses significant technical demands whether undertaking targeted metabolite analyses or comprehensive global metabolic profiling, covering both primary and secondary metabolism. Sumner et al. (2003) suggested that a comprehensive plant metabolic profile should include a minimum of carbohydrates, amino acids, organic acids, lipids/fatty acids, vitamins and various other compound classes such as phenylpropanoids, terpenoids, alkaloids and glucosinolates, with this list of secondary metabolites varying according to the species under study (Sumner et al. 2003).

The various metabolomic research areas have used differing terminology for the definition of metabolic approaches (Table 1). At the core of metabolomic research is a range of metabolite fingerprinting and profiling approaches. Fingerprinting involves the detection of all metabolites within a sample without regard for their identification. Such techniques [viz, NMR, Fourier transform (FT)-IR spectroscopy, Raman spectroscopy and electrospray ionisation (ESI)-MS] are commonly used to screen a biological system to see if differential metabolites are present between control and test material (Johnson et al. 2003), thus providing a cheap first-round approach prior to more costly metabolic profiling. Metabolic profiling involves the detection, quantification and, when possible, the identification of metabolites

Table 1. Terms and definitions used in metabolomics. Adapted from Dunn et al. (2005a) and Hall (2006). Original references Fiehn (2001), Harrigan and Goodacre (2003) and Goodacre et al. (2004).

Term	Definition
Metabolome	The entire biochemical complement present within an organism
Metabolomics	The non-targeted identification and quantification of all metabolites within an organism or system, under a given set of conditions. This is currently not possible with any single or combination of metabolomic approaches
Metabonomics	A term usually used in non-plant systems referring to the quantitative detection of metabolites that are dynamically altered by a living system in response to pathophysiological stimuli or genetic modification. Such metabolites are commonly monitored in biofluids
Metabolic profiling	The identification and quantification of metabolites related through their metabolic pathway(s) or similarities in their chemistry. Crude sample extracts are generally separated by chromatography prior to their detection by MS
Targeted metabolite analysis	Either based on existing knowledge or following broad-scope metabolomic analysis, in-depth biochemical profiling may be based on pre-defined groups of metabolites. Such an approach relies on optimised metabolite, extraction, separation and detection
Metabolite fingerprinting	Rapid and high-throughput methods where global metabolite profiles are obtained from crude samples or simple cellular extracts. In general, metabolites are neither quantified nor identified
Metabolite footprinting	The measurement of metabolites secreted from the intracellular complement of an organism into its extracellular growth medium. Sampling is rapid because metabolite quenching and extraction are not required, and this approach has commonly been employed in microbial metabolomics

within an extract, commonly by employing chromatographic separation [viz, GC or liquid chromatography (LC)] coupled with MS approaches.

In MS profiling, a sample's metabolites are detected prior to the selection and identification of differential metabolites between control and test material often by computer-aided inductive approaches (Goodacre 2005, Goodacre et al. 2004), followed by targeted metabolite identification commonly by MS-MS, MS^n or FT-MS-based instruments. The identified metabolites can then be quantified and correlated with one another and also with their potential metabolite pathways. The altered metabolite pathways are then related back to the systems biological significance in terms of phenotype or to assign gene function. In plants, this has commonly been undertaken by a correlation analysis method known as clique metabolite matrices, developed by researchers based at the Max Planck Institute (Kose et al. 2001). Although it is also important to recognise that just because altered metabolite levels may be correlated, there may not actually be a causal link between the changes (Camacho et al. 2005). However, flux analysis of specific pathways using heavy and stable isotope labels (so-called isotopomers analysis; Vo et al. 2006) has also been used to trace metabolite uptake and destiny (Hegeman et al. 2007, Huege et al. 2007, Kikuchi et al. 2004).

Instrumentation applied to metabolomic analysis

Central to metabolomics are a range of metabolite fingerprinting and profiling technologies that have a range of advantages and disadvantages over one another. However, it is generally accepted that to profile an entire extract without bias, several technologies that collect orthogonal data types should be employed if possible (Hall 2006). However, the extraction method applied to the sample has an equally significant influence on the metabolite range that is detected, as Gullberg et al. (2004) and Want et al. (2006) discovered when applying a range of extraction solvents to plant and serum samples respectively. The richest metabolite profiles will most easily be obtained by employing a range of extraction methods and analytical instruments due to the fact that none are without bias towards certain groups of compounds.

Solvent phase separation has been commonly applied to plant extraction methods (Fiehn et al. 2000b). Phase separation is usually dependent on the extraction solution containing methanol, water and chloroform. Metabolites with a high solubility such as carbohydrates are typically extracted in methanol–water, whereas those with a low solubility such as membrane lipids and fatty acids are typically extracted in chloroform. Further addition of water to the extract will separate the polar from the non-polar phase. These are typically then dried via speed vacuum concentration and stored at −80°C or directly analysed (Fiehn et al. 2000b). Dried-down samples are reconstituted in an appropriate organic solvent or are directly derivatised prior to analysis.

Spectroscopic approaches

Spectroscopic methods are based on the variable absorbance or redirection of electromagnetic (EM) radiation by chemical bonds. This can result in the radiation or transition of the samples atoms to a higher energy state. Spectroscopic methods hold a number of advantages, the most significant being that they offer a global metabolic fingerprint of a sample with rapid spectral acquisition (thus permitting high-throughput analysis) and also can permit spatial measurements of metabolites.

FT-IR spectroscopy

FT-IR spectroscopy is an established and constantly developing analytical technique, which enables the extremely rapid (seconds per sample), high-throughput (1000s of samples per day) and non-destructive analysis of an exceptionally wide range of sample types (Ellis and Goodacre 2006, Ellis et al. 2007). Central to this technology is the principle that when a sample is interrogated with an infrared beam, functional groups within the sample absorb infrared radiation and vibrate in one of the several predetermined ways, such as bending, stretching or deformation vibrations. These absorptions/vibrations can be correlated directly to (bio)chemical species and the resultant infrared spectrum can thus be described as an infrared, or indeed, metabolic, fingerprint characteristic of any (bio)chemical substance. Different classes of (bio)chemical compounds respond to specific IR wavelengths. The mid-IR region (4000–600 cm^{-1}) can be broken down into spectral windows relating to specific classes of compound (Fig. 1): fatty acids 2800–3050 cm^{-1}, amides 1600–1800 cm^{-1}, mixed region 1250–1450 cm^{-1} and polysaccharides 1000–1150 cm^{-1} (Harrigan and Goodacre 2003).

In comparison with mass spectrometric and other spectroscopic approaches, FT-IR is relatively inexpensive and readily lends itself as a rapid first-round screening method (Allwood et al. 2006). In addition, there are several sampling technologies such as high-throughput multiple sample (e.g. 96–384 samples per run) plate scanners, focal-plane array detector microscopes, where false colour chemical image maps of samples may be

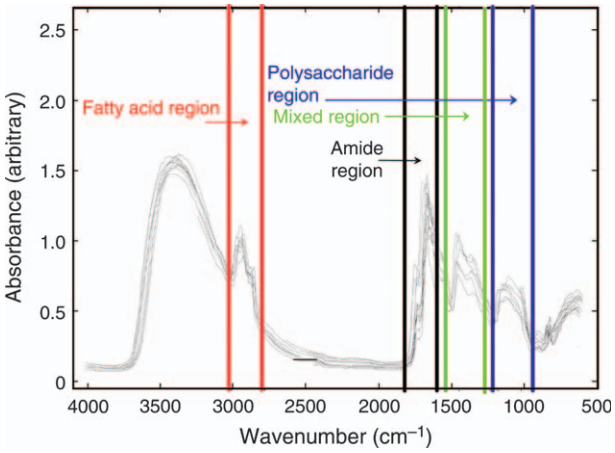

Fig. 1. Characteristic regions within FT-IR spectra. The mid-IR region of an FT-IR spectra can be broken down into subregions, which correspond to specific classes of compound: fatty acids 2800–3050 cm^{-1}, amides 1600–1800 cm^{-1}, mixed region 1250–1450 cm^{-1} and polysaccharides 1000–1150 cm^{-1}. The example spectrum was collected from *Arabidopsis thaliana* Col 0, which was ground in liquid N_2, scurried in water and dried on to an aluminium FT-IR plate. The plate was analysed with a Bruker IFS28 FT-IR (Bruker Optics Ltd, Banner Lane, Coventry, UK) in absorbance mode.

constructed, and attenuated total reflectance (ATR) methods (see below). Apart from the obvious advantages of FT-IR, such as its rapidity, robustness and reproducibility (which cannot be overstressed), sample preparation is also minimal as is the background training required for the instruments operation. Samples are either directly spotted onto a sample plate (made of an IR transparent material such as ZnSe, CaF or Si) or may be homogenised in water in advance, prior to dehydration before measurement takes place.

One potential disadvantage of FT-IR in the mid-IR is that the absorption of water is very intense, but this problem can be overcome in one of the several ways such as dehydration of samples (as already stated), subtraction of the water signal or application of ATR. ATR also allows for measurement of any samples that are compliant in nature (such as plant leaves, stem or root material and indeed whole plants), requires no sample preparation and could be said to minimise any spectral variation between sample operators (Ellis et al. 2002, 2005, Heberle and Zscherp 1996, Winder and Goodacre, 2004). In addition, and of particular relevance perhaps to the plant sciences, optical technology is rapidly developing, with instruments already available commercially as portable, hand-held and micro-devices. These would of course be ideally suited to collecting measurements not only in a laboratory environment but also in glasshouses and in situ in the field with growing specimens of either natural or agricultural origin.

NMR spectroscopy

NMR is also a non-destructive technique that may require minimal sample preparation and is also considered to be high throughput (hundreds of samples per day). The principle of NMR utilises nuclei with odd atomic or mass numbers, which act like magnets and can interact with an external magnetic field by a process termed nuclear spin (Hatada and Kitayama 2004).

In ^{1}H-NMR, an external magnetic field is employed to align the nuclear spin of responding proton nuclei. The nuclei are irradiated with EM radiation where at specific frequencies the different nuclei are said to be in resonance (where they are promoted from a low- to a high-energy spin state). The resulting emission of radiation during the nuclei relaxation process is detected as the absorption frequency. Protons will vary in specific absorption frequencies depending on the number of electrons that is orbiting the nucleus. The electrons establish magnetic fields in opposition to the external magnetic field of the instrument in a process termed as shielding, which results in the absorption of the EM radiation (Hatada and Kitayama 2004).

^{1}H-NMR can theoretically provide a unique signal for each chemically distinct hydrogen nucleus, thus allowing the operator to piece together a compound's structure. Because the vast number of biological metabolites contain hydrogen, ^{1}H-NMR can be considered as non-biased compared with other metabolomic approaches. NMR is therefore emerging as the standard metabolic profiling platform (Ward et al. 2003), despite its relative sensitivity being only half that of MS-based approaches (Sumner et al. 2003), although NMR sensitivity does vary depending on compound class with resolution and spectral crowding also being influential (Ardenkjaer-Larsen et al. 2003). Studies of intact tissues and cellular material are currently becoming more popular since the development of magic angle spinning (MAS). During MAS, the sample is spun at 54.7° (the magic angle) at high speeds while being heated, which greatly enhances sample homogeneity and spectral resolution (Griffin et al. 2003).

Mass spectrometric approaches

Metabolic profiling, especially in the plant sciences, often employs MS and combined chromatography–MS. Ions are generated and extracted into the analyser region of the MS where they are separated by their mass-to-charge (*m/z*) ratios. The separated ions are detected and the signal sent to a data system, where the *m/z* ratios are stored together with their relative abundance for presentation as a mass spectrum. The MS analyser, detector

and often the ionisation source too are maintained under pressure to give the ions a reasonable chance of travelling from one end of the instrument to the other without any hindrance from air molecules (Gross 2004, Niessen 1998).

To identify analytes, linear time of flight (TOF)-MS is often employed. Here, all ions are given the same kinetic energy and the mass of the ion directly controls its flight velocity along the flight tube, and therefore, the arrival time at the detector is proportional to the analytes mass (lighter analytes reach the detector first; Guilhaus et al. 1997). However, on occasion, mass ions of the same molecular weight (MW) but with differing flight velocities are detected, resulting in mass ion peak broadening (as a result of one mass ion reaching the detector in a shorter flight time than another). This problem can be avoided by employing a reflectron (Re)TOF-MS, which typically consists of a series of rings that function to create a retarding field. The mass ions penetrate and are held within the field (mass ions with greater flight velocities are retained for longer). Thus, the arrival time of the mass ions at the detector is equalised and peak broadening is prevented, resulting in greatly enhanced resolution (Guilhaus et al. 1997).

Mass ions are also analysed by single quadrupole MS, which works on the basis of electric fields generated between a set of four rods, which the mass ions pass through. Voltages are applied to the rods to create a quadrupolar field, which only permits the passage of mass ions within a certain range (mass ions outside that range hit the rods and are discharged). Positive mass ions are attracted to negative rods, while the converse is observed with negative mass ions. The quadrupole can therefore be regarded as a mass filter (Gross 2004, Niessen 1998).

Metabolite identity can be provided by tandem MS (MS-MS). MS-MS provides metabolite fragmentation via collision with an inert gas such as nitrogen or argon, which results in collision-induced dissociation (Wysocki et al. 2005). MS-MS approaches commonly use a single quadrupole mass filter linked to a TOF flight tube and detector (a Q-TOF; Chernushevich et al. 2000). However, there is also di-quadrupole MS-MS and now triple quadrupole MS-MS offering greater mass ion selectivity (Chernushevich et al. 2000, Gross 2004, Wysocki et al. 2005). Finally, it is also possible to perform several tandem MS analyses in series using an ion trap (Aharoni et al. 2002, Hirai et al. 2005, Hu et al. 2005) to select the ions to fragment, and this MS^n approach leads to increased structural analysis and hence metabolite identification.

The resolution of MS approaches is often improved with a prior sample fractionation stage using various chromatographic techniques. These various 'hyphenated'-MS approaches each have different advantages and disadvantages, but all represent sensitive and selective methods for qualitative or quantitative analyses (Gross 2004, Wysocki et al. 2005).

Gas chromatography–mass spectrometry

GC-MS is currently the most popular instrument for global metabolic profiling. It is long established, is relatively cheap compared with other MS methods and can resolve complex biological mixtures. GC-MS is naturally biased against non-volatile high-MW metabolites and favours metabolites that are volatile up to 250°C (e.g. alcohols, esters, and monoterpenes), and metabolites that are heat-labile are not detected. GC-MS is now applied to polar non-volatile analysis (e.g. amino acids, sugars and organic acids), which is made possible by chemical derivatisation. GC-MS analysis of plant extracts commonly employs a two-stage derivatisation process (Roessner et al. 2000) where first *O*-alkylhydroxylamine converts sample carbonyl groups to oximes for thermal stabilisation. A secondary treatment with a silylating compound [e.g. *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide] causes the formation of volatile trimethylsilyl esters.

GC-MS ion formation is commonly provided by electron impact (EI). EI involves the independent elimination of the sample's solvent prior to the vaporised sample being swept into the ionisation source where it is impacted by a steady flow of electrons with sufficient energy to ionise the vaporised molecule (Gross 2004). Mass detection is conducted by single quadrupole, TOF or ion-trap-based mass analysers. A single quadrupole mass analyser requires chromatographic times of over an hour to give good separation of a complex sample mixture. This may be reduced by as much as 75% by employing a high-speed TOF mass analyser coupled with deconvolution software (Dunn and Ellis 2005).

The sample metabolites are automatically identified by comparing and matching the retention time and/or retention index with those from mass spectra of chromatograms generated from pure chemical standards (Gross 2004). Mass spectral comparisons can be made with commercially available databases such as the National Institute of Standards and Technology library (www.nist.gov) or more commonly via in-house libraries. Recent advances have led to the development of instruments with greater sensitivity, selectivity and resolving power. These multidimensional separation techniques include GCxGC-TOF-MS where there is a partial separation of the extract through a non-polar GC column, followed by a full separation through a second shorter polar GC

column, thus allowing the visual separation of overlapping chromatographic peaks as well as increasing the instruments' dynamic range (Fig. 2, O'Hagan et al. 2007, Ong and Marriott 2002, Shellie et al. 2001).

Liquid chromatography–mass spectrometry

LC is often used prior to ESI or less frequently atmospheric pressure chemical ionisation (APCI), followed by MS. During ESI, the sample in a suitable solvent at atmospheric pressure is sprayed out of a small needle to which a high charge is applied. Small charged droplets are produced and following rapid solvent evaporation, the charged ions are swept into the MS where a TOF analyser detects the ions mass and the intensity (Gaskell 1997). ESI operates in both negative and positive ionisation modes where it detects metabolites that can be ionised by the loss of an H^+ (ESI-) or gain of an H^+, an NH_4^+, an Na^+ or a K^+ (ESI+). Although greater metabolite coverage is achieved when both ionisation modes are applied, the *m/z* identification is more difficult with ESI+ data because of potential adduct formation with any of four ions.

LC-MS has typically employed linear TOF-MS or single quadrupole MS, although Q-TOF has also been used. Quantification with LC-MS may be achieved with the use of chemical standards by either response or peak area ratios or external calibration (Gross 2004). The levels of resolution of LC-MS instruments are directly influenced by the chemistry and size of LC column employed, whereas the levels of sensitivity are controlled by both the LC column and the MS technique applied. Generally, LC columns do not have the resolving power necessary to separate complex biological mixtures (Tolstikov and Fiehn 2002). However, it has been found that LC-MS instruments with varying MS components (produced by a wide range of manufacturers) allow good mass library reproducibility, presenting a clear advantage (Huhman and Sumner 2002).

Unfortunately, polar phase metabolites elute at the solvent front on LC columns, which reduces metabolite separation and resolution, thus reducing the amount of useful information that can be obtained. However, HydrophiLic Interaction Chromatography (HILIC) columns have been developed, which greatly aid the resolution of polar metabolites, as well as reverse-phase LC columns where the non-polar metabolites elute early. High-pressure columns are also being developed (i.e. ultra-high-pressure liquid chromatography (UHPLC)-MS; Plumb et al. 2006), although the extent of their success is still limited by ionisation suppression during the ESI process (Tolstikov and Fiehn 2002). However, ESI ionisation suppression will be significantly reduced because of the improved chromatographic separation offered by UHPLC. A further development has been that of LC-NMR-MS, which combines high-speed NMR screening along with the metabolite resolving powers and high sensitivity of LC-MS (Corcoran and Spraul 2003). Thus, LC-NMR-MS is capable of metabolite separation, followed by identification and quantification via comparison with internal standards.

Direct injection MS and flow injection MS

Direct injection mass spectrometry (DIMS) allows for the manual direct injection or infusion (DI) of a sample extract into an ESI-MS instrument. This provides a single mass spectrum representative of the samples metabolite composition. An alternative approach flow injection mass spectrometry (FIMS) (Vaidyanathan et al. 2002) provides sample delivery in a flowing solvent to the MS instrument via an automated sample stage. DIMS is more likely to produce information-rich spectra because the extract is not diluted in a flow solvent, although this may be at the expense of throughput (Allwood et al. 2006). However, in comparison with other MS approaches, either DIMS or FIMS represents high-throughput techniques (approximately one sample per minute), although they are both limited by ESI ionisation suppression effects (Dunn et al. 2005b). TOF-MS or high-resolution ReTOF-MS is commonly applied to ESI (although Q-TOF is also used).

Alternative technologies applied to MS

Capillary electrophoresis (CE)-MS applies high-resolution separations, sensitive mass determination and internal chemical standards to identify and quantify thousands of charged metabolites over both positive and negative ionisation modes (Sato et al. 2004). Currently, CE-MS is only really applicable to the analysis of polar metabolites that are highly water soluble. For example, Sato et al. (2004) revealed considerable problems during the analysis of non-water-soluble phospholipids and chlorophylls that coated the capillary walls, causing mass ion detection interference.

Problems associated with mass resolution and accuracy can be solved by FT ion cyclotron resonance (FT-ICR)-MS, an instrument with great potential despite its high purchase and running costs. All metabolite peaks except for identical mass stereoisomers can be resolved provided that structural isomers are separated with a prior chromatographic stage. Through highly accurate detection of mass, the calculation of molecular formulae is permitted. Fragmentation of mass ions is also possible via several methods but commonly by infrared multiphoton dissociation. High sample throughput is possible, as is the detection of lower metabolite concentrations than offered by alternative MS methods (Aharoni et al. 2002, Gross

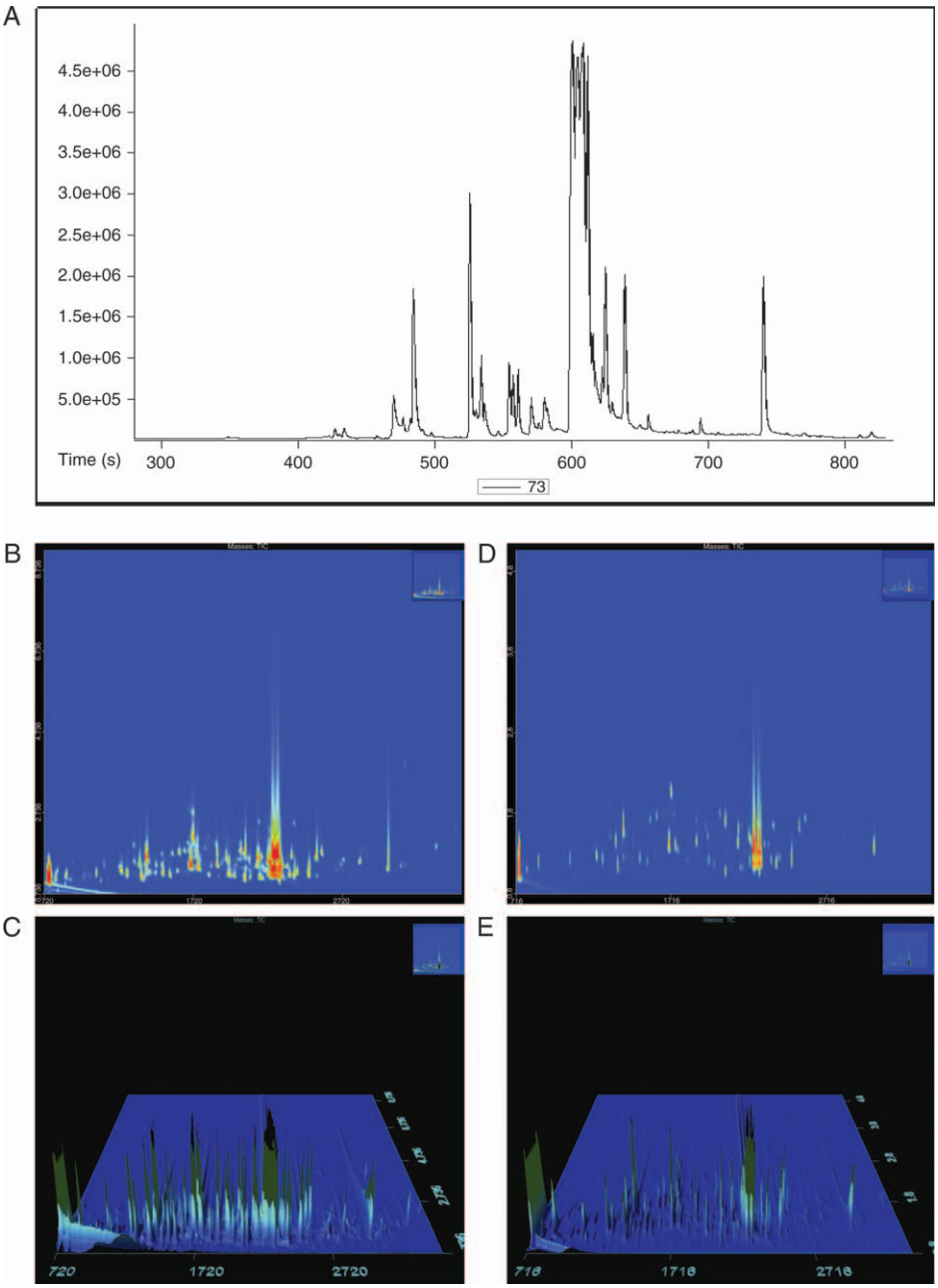

Fig. 2. A comparison of GC-TOF-MS and GCxGC-TOF-MS. (A) GC-TOF-MS single ion chromatogram for *m/z* 73 from a polar extract (277.5 mg ml^{-1} once derivatised; Fiehn et al. 2000a) of Broccoli (cv. Chevalier) generated following the yeast methods of O'Hagan et al. (2005) using a Leco Pegasus III GC-TOF-MS (Leco Ltd, Hazel Grove, Stockport, UK). The *m/z* 73 is an ion detected in mass spectra of all trimethylsilyl-derivatised metabolites and is therefore representative of this class of compounds. (B) GCxGC-TOF-MS chromatogram from the same polar extract (Fiehn et al. 2000a) of Broccoli (cv. Chevalier), derivatised following the yeast method of O'Hagan et al. (2005) and analysed via the instrument methods of O'Hagan et al. (2007). A high injection volume (3 μl) and a 6 s second-dimension separation time were used in the acquisition of this chromatogram to illustrate the greater range of metabolites detected by GCxGC-TOF-MS compared with conventional GC-TOF-MS. Analysis was undertaken with a Leco Pegasus III GCxGC-TOF-MS. (C) Three-dimensional representation of the chromatogram illustrated in B. (D) GCxGC-TOF-MS chromatogram from the same polar extract (Fiehn et al. 2000a) of Broccoli (cv. Chevalier), derivatised following the yeast method of O'Hagan et al. (2005) and analysed via the instrument methods of O'Hagan et al. (2007). A low injection volume (1 μl) and a 10 s second-dimension separation time were used in the acquisition of this chromatogram to illustrate the greater resolution offered by GCxGC-TOF-MS compared with conventional GC-TOF-MS. Analysis was undertaken with a Leco Pegasus III GCxGC-TOF-MS. (E) Three-dimensional representation of the chromatogram illustrated in D.

2004, Harrigan and Goodacre 2003, Hirai et al. 2005). A recent development known as the OrbiTrap FT-MS provides a more user-friendly interface than the FT-ICR-MS and is also thought to offer more rapid sample analysis times at a lower expense with minimal loss of mass accuracy (Hu et al. 2005).

From data to information and possibly knowledge

Raw data from any of the 'omic' fields are an eventual source of information and in turn a source of knowledge (Goodacre 2005). However, to make the leap from one to the next requires considerable data processing and statistical analysis as well as suitable data storage formats. The ability to mine data and to undertake reliable comparative analysis governs the success of metabolomic-based experiments (Goodacre et al. 2004). Metabolomic data analysis consists of four essential areas: raw data processing to allow comparison of different data sets, data mining to select key metabolite variables, data presentation in a simplistic format and effective data storage and database construction (Brown et al. 2005a, Hall 2006).

Raw data processing

All the previously discussed metabolomic technologies suffer from machine inaccuracies such as chromatogram shift and mass drift. Where a large multivariate data set is collected over several different MS runs, or between different instruments, deconvolution of overlapping chromatogram peaks and chromatogram alignment to overcome drift are both required. This may be achieved using a range of software tools including metabolomics spectral formatting and conversion tools (Duran et al. 2003); correlative optimised warping for chromatogram matching (Smedsgaard and Nielsen 2005) and METALIGN for baseline correction, noise reduction, chromatogram alignment and MS peak picking (Vorst et al. 2005); or by using multiple sample alignment algorithms for deconvolution (Jonsson et al. 2005). Baseline correction and noise reduction can be beneficial because they reduce data dimensionality, although this may be at the expense of losing some useful information.

Data mining

Data mining requires statistical tools that allow for high-throughput multivariate analyses of all the components within a data set, resulting in the selection of variables that show significant responses. Computer-based statistical applications must be capable of recognising which samples are similar or different as well as highlighting the key variables (Goodacre et al. 2004, Weckwerth and Morgenthal 2005). Metabolite profiles are information rich, a GC-MS profile typically contains 300–500 variables (Fiehn et al. 2000a), whereas an FT-IR spectra encompasses thousands of variables/wavenumbers, and with the potential of imaging FT-IR, this could include a total of 64×64 spectral measurements. This wealth of information results in significant challenges when it comes to data analysis, thus multivariate approaches are applied to reduce the dimensionality of data to overcome this.

Principal component analysis (PCA, Jolliffe 1986) and independent component analysis (Scholz et al. 2004) are unsupervised methods that perform dimensionality reduction. In this process, a large body of metabolite data is summarised by means of a few parameters with minimal loss of information. For example, PCA attempts to explain the variance within a multivariate data set according to an underlying set of variables known as the principal components (PCs). PCA involves the projection of an (X) matrix constructed from a given number of samples (N) with a given number of variables (K; e.g. peak area for MS peaks or intensity for FT-IR wavelengths) onto multidimensional space (Jolliffe 1986).

PCs are linear combinations of original variables (known as PC loadings), which are used in the projection of the X matrix. PCA models may be considered as being both linear and additive, with often a small set of 10 PCs explaining >90% of the total explained variance. PCA is an unsupervised method that does not require a priori knowledge of the data set structure and so models on the total variance within the data set. The data can be plotted with the plot area being defined according to the PCs, which offer the greatest explained variance. Thus, PCA provides a rapid means of visualising and comparing a data set, with the true variance between sample classes being indicated by differential clustering. By then, ranking the PC loading vectors from which the model is derived, the discriminant variables within the original data set can be selected (Goodacre et al. 2004).

For data sets showing a high degree of variability (which is especially the case with plant data) supervised statistical methods such as discriminant function analysis (DFA; Allwood et al. 2006, Goodacre et al. 2004, Manly 1994) may be applied where the algorithms are trained with an a priori knowledge of the data set structure (i.e. the number of experimental classes and replicates within them). DFA applies the knowledge of the experimental structure in order to produce measures of the between- and within-class differences. It then uses this knowledge to maximise the between-class difference while minimising the within-class difference (i.e. the Fisher ratio)

according to the trained experimental class structure (Allwood et al. 2006, Goodacre et al. 2004, Manly 1994).

The application of genetic algorithms (GAs) and genetic programming (GP) to metabolomics data was first pioneered in the late 1990s (Broadhurst et al. 1997, Goodacre 2005). GA represents a class of algorithms that imitate the processes involved in Darwinian natural selection and evolutionary genetics. Members of the population (samples) are chromosomes that are represented originally by bit strings of numbers or binary digits (each representing a sample variable). GA applies binary code, where variables represented by '1' are used for data modelling, whereas those represented by '0' are not. The GA cycle is fully explained in Goodacre (2005). A further development of GA has been GP where each individual or sample is represented as a tree (rather than strings of binary code) and input variables (metabolite signals or wavenumbers) form the leaves. Parent individuals are ranked according to fitness, which is indicated by the root mean square, and are mated in the same fashion as during GA (Goodacre 2005, Kell et al. 2001).

To validate PCA, PC-DFA and GA analyses independently, the data subsets (typically biological or experimental replicates) can be projected into the PCA or PC-DFA ordinate space (Allwood et al. 2006, Johnson et al. 2003). This involves each class within the data set being randomly split, thus forming a test sample set and a training set. PCA, PC-DFA or GA models are generated independently for the training set, and the test set is then subsequently projected into the same ordinate space. A close alignment between training and test clusters indicates the validity of the PCA, PC-DFA or GA model. This is particularly necessary when the number of variables is considerably larger than the number of samples and when one wants to ensure that the hypotheses generated are robust (Broadhurst and Kell 2006).

An emerging scientific field: plant metabolomics

Plant metabolomic research was not very common prior to this decade, although it has now become increasingly widespread (Fig. 3). Plant metabolomic analysis has been used to examine the effects of gene deletions and transgenes (Roessner et al. 2000, 2001, Weckwerth

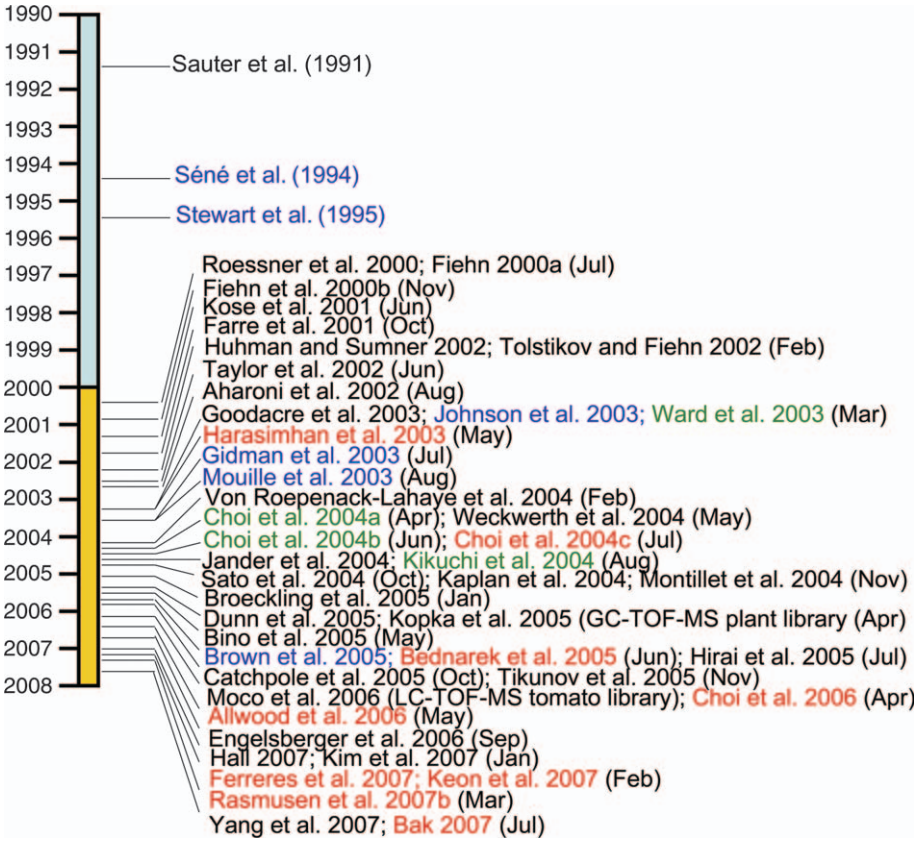

Fig. 3. A plant and plant–host metabolomics timeline. (——), General MS profiling (gene function analysis/physiological analysis/development/protocols); (——), plant–host studies; (——), FT-IR fingerprinting; (——), NMR fingerprinting/profiling.

et al. 2004), to elucidate the effects of physiological processes and physical stresses such as extremes of salinity or temperature (Johnson et al. 2003, Kaplan et al. 2004, Kim et al. 2007) and to study the genetics of metabolism (Keurentjes et al. 2006). Other applications include screening mutant plant populations, plant food sources (Hall 2007) and medicinal herbs or genetically manipulated (GM) crops to identify biomarkers relating to desirable or undesirable traits (Catchpole et al. 2005).

NMR applications have been limited for plant analysis because many metabolite concentrations are lower than the instrument detection limits. Rothamsted Research Institute assessed ^{1}H-NMR as a metabolite fingerprinting technique in *Arabidopsis thaliana* (Ward et al. 2003). The Verpoorte group applied similar methods to the analysis of wild-type and transgenic tobacco lines (Choi et al. 2004a) and to differentiate between various *Cannabis sativa* cultivars (Choi et al. 2004b). Kikuchi et al. (2004) revealed the potential of ^{1}H-NMR for stable isotope-labelling-based work. They also suggested that further developments could reduce plant tissue analysis times to as little as five minutes per sample, while permitting the collection of data from 500 to 1000 metabolites. For good reviews on the application of NMR to plant analysis, we direct the reader to Krishnan et al. (2005) and Ward et al. (2007).

FT-IR analysis has also been applied for metabolite fingerprinting of plant samples. FT-IR spectroscopy along with PC-DFA chemometric analysis was able to discriminate between control and salt-stressed tomato and suggested areas of metabolism that were important (Johnson et al. 2003). Interspecies competition between a monocotyledon *Brachypodium distachyon* and a dicotyledon *A. thaliana* in response to both abiotic and biotic perturbation was also revealed by FT-IR fingerprinting (Gidman et al. 2003). FT-IR has regularly been used for investigations of the plant cell wall make-up and cross-linking (Séné et al. 1994, Stewart et al. 1995), as well as for the classification and identification of various *A. thaliana* cell wall mutants (Brown et al. 2005b, Mouille et al. 2003). Allwood et al. (2006) applied FT-IR to differentiate between *Magnaporthe grisea*-challenged and non-challenged material from a resistant and a susceptible accession of *B. distachyon* over a time course. The initial success of the FT-IR analyses was used as an indication of the systems suitability for further targeted MS profiling.

GC-MS is currently proving to be the most popular profiling approach for plant metabolic profiling. Much of the development of GC-MS (and LC-MS) technologies for plant metabolomic analysis was originally conducted by the Max Planck Institute that adapted the methods first used by Sauter and colleagues (Sauter et al. 1991). Roessner et al. (2000) applied GC-MS-based metabolic profiling to transgenic potato tubers that overexpressed invertase compared with their natural wild-type. The analysis revealed that the transgenic tubers showed reduced starch accumulation, which resulted from the partitioning of carbon flux into the glycolysis cycle. In a subsequent study, chemometric analysis allowed the biochemical differentiation of potato tubers expressing invertase, glucose kinase and sucrose phosphorylase (Roessner et al. 2001). GC-MS-based approaches have also revealed subtle and yet significant metabolite changes in transgenic potato lines expressing sucrose synthase isoform II but which failed to exhibit a visible phenotype (Weckwerth et al. 2004).

Recent studies of potato have included a comparison of GM lines generated from the Désirée cultivar using a combination of FI-ESI-MS, LC-MS and GC-MS. Chemometric analysis revealed that when the metabolite changes associated with the transgene were removed, some differential clustering of the GM lines remained. This indicated wider effects of the transgene than originally indicated via the chemometric analysis (Catchpole et al. 2005). Yang and Bernards (2007) applied GC-MS to monitor the biosynthesis of the cell wall and membrane-associated suberin biopolymer in polar and non-polar extracts taken from wound-healing potato (*Solanum tuberosum* L.) over a time course. The different stages of the suberisation process were differentially clustered using PCA for both polar and non-polar profiles (Yang and Bernards 2007).

The model species *A. thaliana* is emerging as a popular target for metabolomic analysis because of its enormous genomic resources and the ease by which metabolites may be correlated with the gene transcripts of the relevant enzymes (Hirai et al. 2005). In pioneering work Fiehn and co-workers used GC-MS to quantify 326 distinct compounds to distinguish between wild-type and gene deletion mutants of *A. thaliana*, approximately 50% of the detected compounds were identified (Fiehn et al. 2000a, 2000b). Another *A. thaliana* GC-MS study allowed the parental lines and progeny of Columbia (Col0) × Columbia 24 crosses to be differentially clustered (Taylor et al. 2002).

Capillary LC-ESI-MS has been applied along with MS-MS for the selection and identification of a large range of *A. thaliana* secondary metabolites including nitriles, indoles, isothiocyanates and isoflavones (von Roepenack-Lahaye et al. 2004). LC-MS has also been applied to population screening where 10 000 *A. thaliana* mutants were analysed within half a year (Jander et al. 2004). Meyer et al. (2007) analysed metabolic signatures associated with high growth rates in *A. thaliana*. Kim et al. (2007) applied GC and LC-MS to

analyse salt stress in a temporal fashion within *A. thaliana* cell cultures. Results suggested that the methylation cycle, the phenylpropanoid pathway for lignin production and glycine betaine biosynthesis were synergetically induced as a short-term response against salt-stress treatment. The co-induction of glycolysis and sucrose metabolism as well as co-reduction of the methylation cycle was revealed as being a long-term response (Kim et al. 2007). Hirai et al. (2005) applied FT-ICR-MS along with HPLC to determine the effect of sulphur–nitrogen stress in *A. thaliana*. They revealed that many metabolite pathways showed significant changes, some of which had a high degree of coordination, an example being glucosinolate metabolism. This investigation elegantly illustrated the future potential for FT-ICR-MS-based analysis and its full resolving power as indicated by the large range of metabolites detected.

Metabolomic analysis into tomato fruit has described volatile metabolite profiles (Tikunov et al. 2005) and differing flavour traits (Hall et al. 2005) as well as comparison of mutant and wild-type tomato fruit in order to assess the pleiotropic effect of a single gene deletion (Bino et al. 2005). Metabolic quantitative trait loci have also been analysed via metabolomic techniques in tomato (Schauer et al. 2006). Recently, an LC-MS tomato metabolite database has been published, which will greatly aid future analysis (Moco et al. 2006). Rice is also a popular metabolomic target. Sato et al. (2004) applied CE-MS to non-derivatised crude leaf extracts, successfully identifying and quantifying 88 of the main metabolites from the tricarboxylic acid cycle, the pentose phosphate pathway, photorespiration and the synthesis of amino acids. Tarpley et al. (2005) undertook comprehensive GC-MS metabolic profiling of rice during various stages of development and identified a range of specific developmental biomarkers. META-PHOR, a new European initiative, is currently analysing a range of natural and transgenic rice lines, along with melon and broccoli samples, in order to assess metabolomics as a tool for the screening of food quality and for identifying metabolites of health benefit (Hall 2007). Similar studies have previously been conducted looking at fruit development and ripening in strawberry via non-targeted FT-ICR-MS profiling (Aharoni et al. 2002).

The Sumner group (NOBLE Institute) has concentrated their efforts on the analysis of *Medicago truncatula* both as whole plants and as cell suspension cultures (Broeckling et al. 2005, Huhman and Sumner 2002). A recent investigation attempted to correlate suspension culture metabolite change resulting from treatment with certain elicitors including methyl jasmonate (MeJA), yeast elicitor and UV light. Although UV only had minimal effects, MeJA and yeast elicitor treatment mediated differential and common changes in polysaccharides, amino acids and organic acids (Broeckling et al. 2005). Jonsson et al. (2004) of the Umeå Plant Science Centre (Sweden) applied GC-MS to discriminate *Populus tremula* and *P. tremuloides* over a temporal period under both long and short day lighting conditions. Researchers at the University of Wales, Aberystwyth, defined a range of markers associated with flower initiation in *Pharbitis nil* sap. Plants were exposed to either flower-inducing long day or non-flower-inducing short day light regimens, and DIMS was used to acquire rapid fingerprint data. PC-DFA differentially clustered the metabolomic data according to day length (Goodacre et al. 2003). An increasingly important area of recent metabolomic research is analysing herbicide modes of action, which could potentially lead to new targets being discovered (reviewed by Grossmann 2005). For further reading on the application of MS to plant analysis, we direct the reader to the excellent articles of Fiehn et al. (2000a), Hall (2006) and Sumner et al. (2003).

Metabolomic analysis of plant–host interactions: an even greater challenge

Plants offer a nutrient-rich environment to microbial pathogens and symbionts as well as herbivorous insect grazers. As a result, plants must either reduce the damage caused by the inevitable infection of fungal, bacterial and viral pathogens, as well as members from other kingdoms such as nematodes and insects, or provide and receive beneficial metabolites with a symbiotic organism. Some pathogens are obligate biotrophs and will infect and colonise a plant but not kill it. Other pathogens are obligate necrotrophs where infection and colonisation result in plant host death. Some pathogens (hemibiotrophs) adopt an initial biotrophic phase prior to switching to a necrotrophic phase where plant death occurs (Table 2).

As a result of the huge variation between the interactions of plants with different interacting organisms (pathogenic or non-pathogenic), there is a huge variance in the biochemical responses of the plant (i.e. initial signalling cascades leading to either a mutually beneficial response with a symbiont or a targeted biochemical defence against a pathogen or grazer). Thus, plant–host interactions are not only extremely interesting in terms of metabolite richness but also extremely challenging with regard to the huge diversity of different chemical classes (requiring differing technologies for their comprehensive coverage) and also in terms of identifying the individual metabolite changes associated with the host and interacting organism within a dynamic system.

Considering the prominence of research into plant–host and plant–pathogen interactions in plant biology, it is

Table 2. Classes of organisms interacting with plants.

Interaction class	Definition	Example
Necrotrophic pathogen	A pathogen that kills cells prior to colonising and feeding upon them, by secreting toxins that diffuse ahead of the pathogens advance	*Septoria tritici* (wheat pathogen), *Erwinia carotovora* (soft rot)
Biotrophic pathogen	A pathogen that must establish an infection within living host cells in order to live and multiply	*Magnaporthe grisea* (rice blast disease; Allwood et al. 2006), *Pseudomonas syringae* (diseases of tomato and bean)
Hemibiotrophic pathogen	A pathogen that predominantly acts as a biotroph prior to switching to a necrotrophic lifestyle	*Mycosphaerella graminicola* (wheat pathogen; Keon et al. 2007), *Bipolaris sorokiniana* (wheat and barley pathogen)
Grazer	Predation upon plants by herbivorous insects, mammals or members of other kingdoms	*Phyllotreta nemorum* (flea beetle; Bak 2007), *Pieris brassicae* (cabbage white butterfly; Ferreres et al. 2007), cattle
Symbiont	A close relationship between organisms, usually over a prolonged period, where both can benefit	*Sinorhizobium meliloti* N_2 fixation in *Medicago sativa* (Barsch et al. 2006), *Mesorhizobium loti* N_2 fixation in *Lotus japonicus* (Desbrosses et al. 2005), *Epichloe festucae* (rye grass symbiont; Rasmussen et al. 2007b)
Parasite	Considered as a form of symbiosis, but one which usually results in the parasite benefiting more than the host	*Meloidogyne incognita* (parasitic nematode known to infect more than 700 hosts, including crops and ornamental plants)

perhaps surprising that to date few metabolomic studies have focused on the effects of two interacting organisms. However, a small number of groups have conducted research on systems of varying complexity. Narasimhan et al. (2003) used a combination of reverse-phase HPLC and ESI-MS to profile the 'rhizosphere metabolome' of root leaches from wild-type and mutant *A. thaliana*. A profile for *Catharanthus roseus* leaves infected with phytoplasma has also been described (Choi et al. 2004c). In this work, PCA of ^{1}H-NMR spectra from phytoplasma infections was associated with increases in phenylpropanoids and terpenoid indole alkaloids.

Montillet et al. (2004) profiled oxylipin products produced in *A. thaliana* as a result of oxidative stress as elicited by a harpin elicitor using HPLC and also ESI-MS. Comparisons with control plants revealed responses by a range of lipophilic polyunsaturated fatty acid hydroperoxides. Bednarek et al. (2005) used *A. thaliana* wild-type and mutant root cultures infected by the root-pathogenic oomycete *Pythium sylvaticum* to investigate aromatic metabolite profiles. Sixteen indolic, one heterocyclic and three phenylpropanoid compounds were screened and selected by ^{1}H-NMR prior to structural identification via ESI-MS and APCI-MS coupled with a Q-TOF-MS, thus providing MS-MS capabilities. The relative levels of most of the indolics greatly increased on infection, whereas the three phenylpropanoids decreased. These data indicated that roots differed greatly from leaves with regard to the nature and abundance of major soluble phenylpropanoid metabolites. However, for indolics, by contrast, the data indicated a close similarity between roots and leaves.

The previous plant–pathogen-based studies largely utilised ^{1}H-NMR as a fingerprinting technology, although some targeted metabolite identification has also been undertaken. These highlight the utility of metabolic fingerprinting approaches, which allow the rapid analysis and investigation of the sources of variation within a set of samples. For example, Allwood et al. (2006) employed FT-IR fingerprinting because it permitted much greater sample throughput at considerably lower costs and via a method that does not require skilled technical staff, when compared with ^{1}H-NMR (Dunn and Ellis 2005). The FT-IR data indicated that there were differential fatty acid metabolites present between resistant and susceptible cultivars of *B. distachyon* in response to *M. grisea*, thus DIMS was applied to profile non-polar extracts. DIMS revealed a number of differential phospholipids each containing linolenic acid (LA) constituents, which were identified by MS-MS on a Q-TOF. LA was further revealed to be driving the synthesis of the jasmonic acid (JA) plant defence signal (Allwood et al. 2006).

In the recent years, plant–host studies have grown in popularity; this is especially the case with a greater number of plant-pathology-based groups adapting their studies to encompass metabolomic analysis rather than just the more commonly applied transcriptomic and proteomic methods. Choi et al. (2006) employed the infection of tobacco with tobacco mosaic virus (TMV) as a model pathosystem for metabolomic analysis with ^{1}H-NMR. Initially, the host elicits a programmed cell death (PCD) of the infected cells to attempt to isolate the TMV, and this is followed by downstream signalling

cascades of compounds such as salicylic acid (SA), which results in the host gaining systemic acquired resistance (SAR), permitting a more powerful and rapid response against future TMV infection. On initial infection, the tobacco plants revealed significant alterations of phenylpropanoids involved in cell wall strengthening, caffeoylquinic acid was upregulated and is known to contribute to plant resistance, as was α-LA possibly contributing to JA synthesis. During SAR, tobacco was shown to downregulate the metabolites upregulated during the initial infection as well as several sugars. A number of sesquiterpenoids and diterpenoids were found to be upregulated specifically to the elicitation of SAR (Choi et al. 2006).

Keon et al. (2007) undertook a comprehensive analysis of the transcriptional adaptation of *Mycosphaerella graminicola* targeted against the PCD response of its wheat host. *M. graminicola* transcriptomic data suggested that as the wheat PCD occurred, the fungus switched to a more rapid growth rate and showed physiological adaptation with respect to membrane transport, chemical and oxidative stress mechanisms and metabolism. Using ^{1}H-NMR, they went on to reveal changes in the sugar levels within the wheat apoplastic fluid. Minimal changes were detected during the biotrophic phase of the fungal colonisation; however, as the host PCD occurred, the sugar levels showed a pronounced increase correlating with the indicated rise of the fungal growth rate (Keon et al. 2007).

Outside of plant–microbial pathogen-based studies, Kristensen et al. (2005) metabolically engineered the *Sorghum bicolor* gene cluster responsible for dhurrin synthesis into *A. thaliana*. Initially, the metabolome and transcriptome were monitored to assess the effect of expression of the dhurrin gene cluster. The presence of dhurrin in the transgenic *A. thaliana* plants was shown to confer resistance to the flea beetle *Phyllotreta nemorum*. In more recent analyses, the group have compared the metabolic profiles of wild-type and dhurrin-expressing *A. thaliana* as well as assessing the metabolic effect of grazing on the two plant lines in the flea beetle (Bak 2007). Ferreres et al. (2007) undertook a simultaneous metabolomic analysis of both plant and grazer based on a system utilising tronchuda cabbage leaves that had been fed upon by larvae of the cabbage white butterfly (*Pieris brassicae*). Using HPLC linked with a diode array detector (DAD) and an ESI-MS-MS (HPLC-DAD-ESI-MS/MS), 20 flavonoids of plant origin were identified, some of which were revealed to be selectively sequestered by the larval grazer. Further analyses of the larval samples revealed that the larvae were further metabolising the dietary flavonoids.

Very little metabolomic research has been conducted on plant symbiotic relationships, although several groups have undertaken some analysis. Barsch et al. (2006) studied the interaction between *Medicago sativa* and *Sinorhizobium meliloti*, a relationship that requires a balanced physiological interaction to enable atmospheric nitrogen fixation by the microsymbiont. Via the comparative GC-TOF-MS analysis of *M. sativa* containing mutant non-nitrogen-fixing *S. meliloti*, wild-type *S. meliloti* and no *S. meliloti*, it was found that the relationship was based on alternating control by both partners during the various stages of root nodule development. Although distinct metabolic changes were detected during the different stages of nodule development, signals mediating the transition of each of the different stages could not be identified. Desbrosses et al. (2005) also applied GC-TOF-MS to the analysis of a nitrogen-fixing symbiont, *Mesorhizobium loti*, in *Lotus japonicus*. Comparative metabolic profiles were generated for nodules, roots, leaves and flowers from symbiotic plants. PCA permitted the metabolic differentiation of the different plant tissues and identified marker metabolites specific to each of the tissue types.

The symbiosis-based work of Barry Scott and colleagues at Massey University is also worth mentioning despite of it being based on genetic studies followed by targeted biochemical analysis of the effected pathways with thin layer chromatography and HPLC as opposed to 'true' metabolomics. Their work has focused on the relationship between rye grass and its fungal symbiont *Epichloe festucae*, which provides the plant with anti-insect and anti-mammalian toxins to reduce herbivore feeding upon the host (Rasmussen et al. 2007a, Tanaka et al. 2005, Young et al. 2006). Through creating gene deletion mutants in the fungal symbiont and combining genetic analyses with targeted MS biochemical analysis, the group have identified a number of the toxins produced by the fungus for the Rye grass host and their routes of biosynthesis (Rasmussen et al. 2007b).

Conclusions and future prospects

Historically, most plant metabolomic studies have focused on comparing situations where there are large chemical differences exhibited throughout the plant, for instance comparisons between mutant (or transgenic) and wild-type lines (e.g. Fiehn et al. 2000a, Roessner et al. 2001) or heating and chilling stress (Kaplan et al. 2004). Because of their dynamic, spatially varied and metabolite-rich qualities, plant interactions represent a very complex target for metabolomic analysis. Therefore, discerning key metabolic trends is a major challenge for the nascent field of plant metabolomics. Nevertheless, there have been precedents for metabolomic studies of such complex phenomena, perhaps most

notably human cancer. This has led to the current increase in popularity of metabolomics to assess the complex interactions between plants and other organisms.

In metabolomics, major technological problems also remain, with the most challenging being the dynamic range of MS-based instruments, limiting the range of metabolites that can be detected with a single technique and the deconvolution of overlapping chromatographic peaks. Although chromatogram deconvolution has been greatly aided for GC-TOF-MS data, significant efforts are still required for LC-MS deconvolution. Currently, the range of publicly available metabolite databases and libraries is also a major limitation. This problem arises from the limited availability of chemical standards, including many compounds that are relevant to the study of metabolic pathways. Some compounds are relatively unstable (being transitional within metabolic pathways) and therefore are never likely to become generally available. However, this may be viewed in a positive light, given the huge effort currently being expended by metabolomics groups in developing metabolite standards and libraries (http://msi-workgroups.sourceforge.net/).

The current lack of metabolite libraries perhaps impacts most greatly upon plant metabolomics, given that plants possess a huge range of secondary metabolites, many of which we do not currently have the ability to identify (Hall 2006). MS libraries and databases are not common because of difficulties in the comparison of data between instruments. This stems from technical difficulties that influence analytical reproducibility (Dunn and Ellis 2005). However, the creation of an open access GC-MS plant metabolite database has been initiated by the Max Planck Institute (Kopka et al. 2005), as has a tomato-based LC-MS database by Plant Research International (Moco et al. 2006).

Ultimately, databases are required that are based on a range of organisms from across different kingdoms, which are composed of data from their transcripts, proteins, metabolites, metabolic pathways and the enzymes involved within them (Hall,2006). Species-specific databases such as AraCyc (www.arabidopsis.org) would benefit greatly from the addition of information on rate constants between reactions within metabolic pathways and between metabolic networks. This would be very useful for systems biology where the relative accumulation of particular metabolites and possibly flux through a pathway could be modelled. Although this is an attractive prospect, one must bear in mind that rate constants are usually established using in vitro systems that may not reflect the in vivo situation.

Other metabolomic problems stem from sample heterogeneity. Tissues from whole plants are generally pooled potentially mixing leaves, petioles, shoots, flowers and roots. It should therefore be remembered that information regarding the compartmentation of metabolites to particular organelles is often lost. In the case of plant pathology, there are additional problems such as the sample containing plant, pathogen and infected and non-infected tissues. Differentiation of the metabolome from different tissue and cell types to the different cellular organelles will provide a more complex and yet surely improved appreciation of the plant metabolome (Sumner et al. 2003).

Some groups have already taken on the challenge of compartmentalised metabolomics. Farre et al. (2001) studied the compartmentalisation of glycolysis intermediates, sugars, nucleotides, sugar alcohols, amino acids and organic acids, in the potato tubers amyloplast, cytosol and vacuole. Recently, significant advances have been made for the compartmentalised analysis of metabolites by using stable isotope labelling with for example $^{15}NO_3$, ^{13}C or $^{15}K\ NO_3$ (Engelsberger et al. 2006, Kikuchi et al. 2004). Kikuchi et al. (2004) performed multidimensional heteronuclear-NMR-based analysis on the metabolic movement of carbon and nitrogen nuclei in *A. thaliana*, in which distinct ethanol-stress responses were investigated, as were nitrogen fluxes in ^{15}N-labeled seeds during germination.

To overcome the current limitations on metabolomic analysis discussed above successfully will require an interdisciplinary-based approach where biologists, chemists, statisticians and instrument manufacturers all have an input (Hall 2006, Hall et al. 2002, Sumner et al. 2003). Such an approach will maintain that the developments of both instrumentation and computational programs will benefit all research areas within metabolomics. This is especially the case within plant metabolomics with projects such as META-PHOR (Hall 2007) leading the way. With such coordinated efforts, further developments are inevitable, and so the future of metabolomics within the study of plant and plant–host-based systems appears to be rosy and will inevitably lead to a more comprehensive understanding of plant systems than the information provided by transcriptomics and proteomics alone.

Acknowledgements – J. W. A. and R. G. are both indebted and very grateful to the EU project META-PHOR (Food-CT-2006-03622) for research support and funding. R. G. is also indebted to the EBS committee of the UK BBSRC for financial support. The authors also wish to thank Drs Warwick B. Dunn and Paul Begley (School of Chemistry, Manchester Interdisciplinary Biocentre, University of Manchester) for their assistance with generating the GC-TOF-MS and GCxGC-TOF-MS profiles used in Fig. 2, and to Drs Annick Moing and Benoît

Biais as well as other colleagues at INRA (Bordeaux, France) for providing Broccoli material (also used in Fig. 2).

References

Aharoni A, de Vos CHR, Verhoeven HA, Maliepaard CA, Kruppa G, Bino R, Goodenowe DB (2002) Nontargeted metabolome analysis by use of Fourier transform ion cyclotron mass spectrometry. OMICS 6: 217–234

Allwood JW, Ellis DI, Heald JK, Goodacre R, Mur LAJ (2006) Metabolomic approaches reveal that phosphatidic and phosphatidyl glycerol phospholipids are major discriminatory non-polar metabolites in responses by *Brachypodium distachyon* to challenge by *Magnaporthe grisea*. Plant J 46: 351–368

Ardenkjaer-Larsen JH, Fridlund B, Gram A, Hansson G, Hansson L, Lerche MH, Servin R, Thaning M, Golman K (2003) Increase in signal-to-noise ratio of >10,000 times in liquid-state NMR. Proc Natl Acad Sci USA 100: 10158–10163

Bak S (2007) Lessons to be learned from metabolic engineering of cyanogenic glucosides. In: Bak S (ed) Metabolomics Society's 3rd Annual Conference. Manchester

Barsch A, Tellstrom V, Patschkowski T, Kuster H, Niehaus K (2006) Metabolite profiles of nodulated alfalfa plants indicate that distinct stages of nodule organogenesis are accompanied by global physiological adaptations. Mol Plant Microbe Interact 19: 998–1013

Bednarek P, Schneider B, Svatos A, Oldham NJ, Hahlbrock K (2005) Structural complexity, differential response to infection and tissue specificity of indolic and phenylpropanoid secondary metabolism in *Arabidopsis* roots. Plant Physiol 138: 1058–1070

Bino RJ, de Vos CHR, Lieberman M, Hall RD, Bovy A, Jonker HH, Tikunov Y, Lommen A, Moco S, Levin I (2005) The light-hyperresponsive high pigment-2(dg) mutation of tomato: alterations in the fruit metabolome. New Phytol 166: 427–438

Broadhurst D, Kell DB (2006) Statistical strategies for avoiding false discoveries in metabolomics and related experiments. Metabolomics 2: 171–196

Broadhurst D, Goodacre R, Jones A, Rowland JJ, Kell DB (1997) Genetic algorithms as a method for variable selection in PLS regression with applications to pyrolysis mass spectrometry. Anal Chim Acta 348: 323–326

Broeckling CD, Huhman DV, Farag MA, Smith JT, May GD, Mendes P, Dixon RA, Sumner LW (2005) Metabolic profiling of *Medicago trunculata* cell cultures reveals the effects of biotic and abiotic elicitors on metabolism. J Exp Bot 56: 323–326

Brown M, Dunn WB, Ellis DI, Goodacre R, Handl J, Knowles JD, O'Hagan S, Spasić I, Kell DB (2005a) A metabolome pipeline: from concept to data to knowledge. Metabolomics 1: 39–51

Brown DM, Zeef LAH, Ellis J, Goodacre R, Turner ST (2005b) Identification of novel genes in *Arabidopsis* involved in secondary cell wall formation using expression profiling and reverse genetics. Plant Cell 17: 2281–2295

Camacho D, de la Fuente A, Mendes P (2005) The origin of correlations in metabolomics data. Metabolomics 1: 53–63

Castle AL, Fiehn O, Kaddurah-Daouk R, Lindon JC (2006) Metabolomics Standards Workshop and the development of international standards for reporting metabolomics experimental results. Brief Bioinform 7: 159–165

Catchpole GS, Beckman M, Enot DP, Mondhe M, Zywicki B, Taylor J, Hardy N, Smith A, King RD, Kell DB, Fiehn O, Draper J (2005) Hierarchical metabolomics demonstrates compositional similarity between genetically modified and conventional potato crops. Proc Natl Acad Sci USA 102: 14458–14462

Chernushevich IV, Loboda AV, Thomson BA (2000) An introduction to quadrupole-time-of-flight mass spectrometry. J Mass Spectrom 36: 849–865

Choi H-K, Choi YH, Verberne M, Lefeber AWM, Erkelens C, Verpoorte R (2004a) Metabolic fingerprinting of wild type and transgenic tobacco plants by 1H-NMR and multivariate analysis techniques. Phytochemistry 65: 857–864

Choi YH, Kim HK, Hazekamp A, Erkelens C, Lefeber AMW, Verpoorte R (2004b) Metabolomic differentiation of *Cannabis sativa* cultivars using 1H-NMR and principal component analysis. J Nat Prod 67: 953–957

Choi YH, Tapias EC, Kim HK, Lefeber AW, Erkelens C, Verhoeven JT, Brzin J, Zel J, Verpoorte R (2004c) Metabolic discrimination of *Catharanthus roseus* leaves infected by phytoplasma using 1H-NMR spectroscopy and multivariate data analysis. Plant Physiol 135: 2398–2410

Choi H-C, Kim HK, Linthorst HJM, Hollander JG, Lefeber AWM, Erkelens C, Nuizllard JM, Verpoorte R (2006) NMR metabolomics to revisit the tobacco mosaic virus infection in *Nicotiana tobacum* leaves. J Nat Prod 69: 742–748

Corcoran O, Spraul M (2003) LC-NMR-MS in drug discovery. Drug Discov Today 8: 624–631

Desbrosses GG, Kopka J, Udvardi MK (2005) *Lotus japonicus* metabolic profiling. Development of gas chromatography-mass spectrometry resources for the study of plant-microbe interactions. Plant Physiol 137: 1302–1318

Dunn WB, Ellis DI (2005) Metabolomics: current analytical platforms and methodologies. Trends Analyt Chem 24: 285–294

Dunn WB, Bailey NJC, Johnson HE (2005a) Measuring the metabolome: current analytical technologies. Analyst 130: 606–625

Dunn WB, Overy S, Quick WP (2005b) Evaluation of automated electrospray-TOF mass spectrometry for metabolic fingerprinting of the plant metabolome. Metabolomics 1: 137–148

Duran AL, Yang J, Wang L, Sumner LW (2003) Metabolomics spectral formatting, alignment and conversion tools (MSFACTs). Bioninformatics 19: 2283–2293

Ellis DI, Goodacre R (2006) Metabolic fingerprinting in disease diagnosis: biomedical applications of infrared and Raman spectroscopy. Analyst 131: 875–885

Ellis DI, Broadhurst D, Kell DB, Rowland JJ, Goodacre R (2002) Rapid and quantitative detection of the microbial spoilage of meat by Fourier transform infrared spectroscopy and machine learning. Appl Environ Microbiol 68: 2822–2828

Ellis DI, Broadhurst D, Clarke SJ, Goodacre R (2005) Rapid identification of closely related muscle foods by vibrational spectroscopy and machine learning. Analyst 130: 1648–1654

Ellis DI, Dunn WB, Griffin JL, Allwood JW, Goodacre R (2007) Metabolic fingerprinting as a diagnostic tool. Pharmacogenomics 8: 1243–1266

Engelsberger WR, Erban A, Kopka J, Schulze WX (2006) Metabolic labeling of plant cell cultures with $K^{15}NO_3$ as a tool for quantitative analysis of proteins and metabolites. Plant Methods 2: 14–25

Farre EM, Tiessen A, Roessner U, Geigenberger P, Tretheway RN, Willmitzer L (2001) Analysis of compartmentation of glycolytic intermediates, nucleotides, sugars, organic acids, amino acids, and sugar alcohols in potato tubers using a nonaqueous fractionation method. Plant Physiol 127: 685–700

Ferreres F, Sousa C, Valentao P, Pereira JA, Seabra RM, Andrade PB (2007) Tronchuda cabbage flavonoids uptake by *Pieres brassicae*. Phytochemistry 68: 361–367

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

Fiehn O, Kopka J, Dormann P, Altmann T, Trethewey RN, Willmitzer L (2000a) Metabolite profiling for plant functional genomics. Nat Biotechnol 18: 1157–1161

Fiehn O, Kopka J, Trethewey RN, Willmitzer L (2000b) Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry. Anal Chem 72: 3573–3580

Fiehn O, Robertson D, Griffin J, van der Werf M, Nikolau B, Morrison N, Sumner LW, Goodacre R, Hardy NW, Taylor C, Fostel J, Kristal B, Kaddurah-Daouk R, Mendes P, van Ommen B, Lindon JC, Sansone S-A (2007) The metabolomics standards initiative (MSI). Metabolomics 3: 175–178

Gaskell SJ (1997) Electrospray: principles and practice. J Mass Spectrom 32: 677–688

Gidman E, Goodacre R, Emmett B, Smith AR, Gwynn-Jones D (2003) Investigating plant-plant interference by metabolic fingerprinting. Phytochemistry 63: 705–710

Goodacre R (2005) Making sense of the metabolome using evolutionary computation: seeing the wood with the trees. J Exp Bot 56: 245–254

Goodacre R, York EV, Heald JK, Scott IM (2003) Chemometric discrimination of unfractionated plant extracts profiled by flow-injection electrospray mass spectrometry. Phytochemistry 62: 859–863

Goodacre R, Vaidyanathan S, Dunn WB, Harrigan GG, Kell DB (2004) Metabolomics by numbers – acquiring and understanding global metabolite data. Trends Biotechnol 22: 245–252

Griffin JL, Pole J, Nicholson JK, Carmichael P (2003) Cellular environment of metabolites and a metabonomic study of tamoxifen in endometrial cells using gradient high resolution magic angle spinning 1H NMR spectroscopy. Biochim Biophys Acta 1619: 151–158

Gross JH (2004) Mass Spectrometry. Springer, Berlin

Grossmann K (2005) What it takes to get a herbicide's mode of action: physionomics, a classical approach in a new complexion. Pest Manag Sci 61: 423–431

Guilhaus M, Mlynski V, Selby D (1997) Perfect timing: time-of-flight mass spectrometry. Rapid Commun Mass Spectrom 11: 951–962

Gullberg J, Jonsson P, Nordström A, Sjöström M, Moritz T (2004) Design of experiments: an efficient strategy to identify factors influencing extraction and derivatization of *Arabidopsis thaliana* samples in metabolomic studies with gas chromatography/mass spectrometry. Anal Biochem 331: 283–295

Gygi SP, Rochon Y, Franza BR, Aebersold R (1999) Correlation between protein and mRNA abundance in yeast. Mol Cell Biol 19: 1720–1730

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

Hall RD (2007) Food metabolomics: META-PHOR a new European research initiative. AgroFood industry hi-tech 18: 14–16

Hall R, Beale M, Fiehn O, Hardy N, Sumner L, Bino R (2002) Plant metabolomics: the missing link in functional genomic strategies. Plant Cell 14: 1437–1440

Hall R, de Vos CHR, Verhoeven HA, Bino RJ (2005) Metabolomics for the assessment of functional diversity and quality traits in plants. In: Vaidyanathan S, Harrigan GG, Goodacre R (eds) Metabolome Analyses: Strategies for Systems Biology. Springer, New York, pp 31–44

Harrigan GG, Goodacre R (eds) (2003) Metabolic Profiling: Its Role in Biomarker Discovery and Gene Function Analysis. Kluwer Academic Publishers, Boston

Hatada K, Kitayama T (2004) Basic principles of NMR. In: Hatada K, Kitayama T (eds) NMR Spectroscopy of Polymers. Springer, New York, pp 1–34

Heberle J, Zscherp C (1996) ATR/FT-IR difference spectroscopy of biological matter with microsecond time resolution. Appl Spectrosc 50: 588–596

Hegeman AD, Schulte CF, Cui Q, Lewis IA, Huttlin EL, Eghbalnia H, Harms AC, Ulrich EL, Markley JL, Sussman MR (2007) Stable isotope assisted assignment of elemental compositions for metabolomics. Anal Chem 79: 6912–6921

Hirai MY, Klein M, Fujikawa Y, Yano M, Goodenowe DB, Yamazaki Y, Kanaya S, Nakamura Y, Kitayama M, Suzuki H, Sakurai N, Shibata D, Tokuhisa J, Reichelt M, Gershenzon J, Papenbrock J, Saito K (2005) Elucidation of gene-to-gene networks in *Arabidopsis* by integration of metabolomics and transcriptomics. J Biol Chem 280: 25590–25595

Hu Q, Noll RJ, Li H, Makarov A, Hardman M, Cooks RG (2005) The Orbitrap: a new mass spectrometer. J Mass Spectrom 40: 430–443

Huege J, Sulpice R, Gibon Y, Lisec J, Koehl K, Kopka J (2007) GC-EI-TOF-MS of in vivo carbon-partitioning into soluble metabolite pools of higher plants by monitoring isotope dilution after $^{13}CO_2$ labelling. Phytochemistry 68: 2258–2272

Huhman DH, Sumner LW (2002) Metabolic profiling of saponin glycosides in *Medicago sativa* and *Medicago truncatula* using HPLC coupled to an electrospray ion-trap mass spectrometer. Phytochemistry 59: 347–360

Jander G, Norris SR, Joshi V, Fraga M, Rugg A, Yu S, Li L, Last RL (2004) Application of a high-throughput HPLC-MS/MS assay to *Arabidopsis* mutant screening; evidence that threonine aldolase plays a role in seed nutritional quality. Plant J 39: 465–475

Johnson HE, Broadhurst D, Goodacre R, Smith AR (2003) Metabolic fingerprinting of salt-stressed tomatoes. Phytochemistry 62: 919–928

Jolliffe IT (1986) Principal Component Analysis. Springer-Verlag, New York

Jonsson P, Gullberg J, Nordstrom A, Kusano M, Kowalczyk M, Sjöström M, Moritz T (2004) A strategy for identifying differences in large series of metabolomic samples analyzed by GC/MS. Anal Chem 76: 1738–1745

Jonsson P, Johansson AI, Gullberg J, Trygg J, Jiye A, Grung B, Marklund S, Sjöström M, Antti H, Moritz T (2005) High-throughput data analysis for detecting and identifying differences between samples in gc/ms-based metabolomic analyses. Anal Chem 77: 5635–5642

Kaplan F, Kopka J, Haskell DW, Zhao W, Schiller KC, Gatzke N, Sung DY, Guy CL (2004) Exploring the temperature-stress metabolome of *Arabidopsis*. Plant Physiol 136: 4159–4168

Kell DB, Oliver SG (2004) Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. Bioessays 26: 99–105

Kell DB, Darby RM, Draper J (2001) Genomic computing. Explanatory analysis of plant expression profiling data using machine learning. Plant Physiol 126: 943–951

Keon J, Antoniw J, Carzaniga R, Deller S, Ward JL, Baker JM, Beale MH, Hammond-Kosack K, Rudd JJ (2007) Transcriptional adaptation of *Mycosphaerella graminicola* to programmed cell death (PCD) of its susceptible wheat host. Mol Plant Microbe Interact 20: 178–193

Keurentjes JJB, Fu JY, de Vos CHR, Lommen A, Hall RD, Bino RJ, van der Plas LHW, Jansen RC, Vreugdenhil D, Koornneef M (2006) The genetics of plant metabolism. Nat Genet 38: 842–849

Kikuchi J, Shinozaki K, Hirayama T (2004) Stable isotope labeling of *Arabidopsis thaliana* for an NMR-based metabolomics approach. Plant Cell Physiol 45: 1099–1104

Kim JK, Bamba T, Harada K, Fukusaki E, Kobayashi A (2007) Time-course metabolic profiling in *Arabidopsis thaliana* cell cultures after salt stress. J Exp Bot 58: 415–424

Kopka J, Schauer N, Krueger S, Birkenmeyer C, Usadel B, Bergmüller E, Dörmann P, Gibon Y, Willmitzer L, Fernie AR, Steinhauser D (2005) GMDB@CSB.DB: the Golm Metabolome database. Bioinformatics 21: 1635–1638

Kose F, Weckwerth W, Linke T, Fiehn O (2001) Visualising plant metabolomic correlation networks using clique-metabolite matrices. Bioinformatics 17: 1198–1208

Krishnan P, Kruger NJ, Ratcliffe RG (2005) Metabolite fingerprinting and profiling in plants using NMR. J Exp Bot 56: 255–265

Kristensen C, Morant M, Olsen CO, Ekstrøm CT, Galbraith DW, Møller BL, Bak S (2005) Metabolic engineering of dhurrin in transgenic *Arabidopsis* plants with marginal inadvertent effects on the metabolome and transcriptome. Proc Natl Acad Sci USA. 102: 1779–1784

Manly BFJ (1994) Multivariate Statistical Methods: A Primer. Chapman & Hall, London

Meyer RC, Steinfath M, Lisec J, Becher M, Wituka-Wall H, Torjek O, Fiehn O, Eckardt A, Willmitzer L, Selbig J, Altmann T (2007) The metabolic signature related to high plant growth rate in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 104: 4759–4764

Moco S, Bino RJ, Vorst O, Verhoeven HA, de Groot J, van Beek TA, Vervoort J, deVos CHR (2006) A liquid chromatography-mass spectrometry-based metabolome database for tomato. Plant Physiol 141: 1205–1218

Montillet JL, Cacas JL, Garnier L, Montane MH, Douki T, Bessoule JJ, Polkowska-Kowalczyk L, Maciejewska U, Agnel JP, Vial A, Triantaphylides C (2004) The upstream oxylipin profile of *Arabidopsis thaliana*: a tool to scan for oxidative stresses. Plant J 40: 439–451

Mouille G, Robin S, Lecomte M, Pagant S, Höfte H (2003) Classification and identification of *Arabidopsis* cell wall mutants using Fourier-transform infrared (FT-IR) microspectroscopy. Plant J 35: 393–404

Narasimhan K, Basheer C, Bajic VB, Swarup S (2003) Enhancement of plant-microbe interactions using a rhizosphere metabolomics-driven approach and its application in the removal of polychlorinated biphenyls. Plant Physiol 132: 146–153

Niessen WM (1998) Advances in instrumentation in liquid chromatography-mass spectrometry and related liquid-introduction techniques. J Chromatogr A 794: 407–435

O'Hagan S, Dunn WB, Brown M, Knowles JD, Kell DB (2005) Closed-loop, multiobjective optimization of analytical instrumentation: gas chromatography/time-of-flight mass spectrometry of the metabolomes of human serum and of yeast fermentations. Anal Chem 77: 290–303

O'Hagan S, Dunn WB, Knowles JD, Broadhurst D, Williams R, Ashworth JJ, Cameron M, Kell DB (2007) Closed-loop, multiobjective optimization of two-dimensional gas chromatography/mass spectrometry for serum metabolomics. Anal Chem 79: 464–476

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

Ong RCY, Marriott PJ (2002) A review of basic concepts in comprehensive two dimensional gas chromatography. J Chromatogr Sci 40: 276–291

Plumb RS, Johnson KA, Rainville P, Smith BW, Wison ID, Castro-Perez JM, Nicholson JK (2006) UPLC/MS^E; a new approach for generating molecular fragment information for biomarker structure elucidation. Rapid Commun Mass Spectrom 20: 1989–1994

Rasmussen S, Parsons AJ, Bassett S, Christensen MJ, Hume DE, Johnson LJ, Johnson RD, Simpson WR, Stacke C, Voisey CR, Xue H, Newman JA (2007a) High nitrogen supply and carbohydrate content reduce fungal endophyte and alkaloid concentration in Lolium perenne. New Phytol 173: 787–797

Rasmussen S, Parsons AJ, Xue H, Newman JA (2007b) High nutrient supply and carbohydrate content reduce endophyte and alkaloid concentration. In: Popay AJ, Thom ER (eds) The 6th International Symposium on Fungal Endophytes of Grasses. Grassland Research and Practice Series No. 13. New Zealand Grassland Association, Christchurch, New Zealand, pp 141–146

von Roepenack-Lahaye E, Degenkolb T, Zerjeski M, Franz M, Roth U, Wessjohann L, Schmidt J, Scheel D, Clemens S (2004) Profiling of *Arabidopsis* secondary metabolites by capillary liquid chromatography coupled to electrospray ionization quadrupole time-of-flight mass spectrometry. Plant Physiol 134: 548–559

Roessner U, Wagner C, Kopka J, Tretheway RN, Willmitzer L (2000) Simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. Plant J 23: 131–142

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

Sansone S-A, Nikolau B, van Ommen B, Kristal BS, Taylor C, Robertson D, Lindon J, Griffin JL, Sumner LW, van der Werf M, Hardy NW, Morrison N, Mendes P, Kaddurah-Daouk R, Goodacre R, Fan T, Fiehn O (2007) The metabolomics standards initiative (MSI). Nat Biotechnol 25: 846–848

Sato S, Soga T, Nishioka T, Tomita M (2004) Simultaneous determination of the main metabolites in rice leaves using capillary electrophoresis mass spectrometry and capillary electrophoresis diode array detection. Plant J 40: 151–163

Sauter H, Lauer M, Fritsch H (1991) Metabolic profiling of plants a new diagnostic technique. In: Baker DR, Fenyes JG, Moberg WK (eds) American Chemical Society Symposium Series No. 443. American Chemical Society, Washington, DC, pp 288–299

Schauer N, Semel Y, Roessner U, Gur A, Balbo I, Carrari F, Pleban T, Perez-Melis A, Bruedigam C, Kopka J, Willmitzer L, Zamir D, Fernie AR (2006) Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. Nat Biotechnol 24: 447–454

Scholz M, Gatzek S, Sterling A, Fiehn O, Selbig J (2004) Metabolite fingerprinting: detecting biological features by independent component analysis. Bioinformatics 20: 2447–2454

Séné C, McCann MC, Wilson RH, Grinter R (1994) Fourier-transform Raman and Fourier-transform infrared spectroscopy (An investigation of five higher plant cell walls and their components). Plant Physiol 106: 1623–1631

Shellie R, Marriott P, Morrison P (2001) Concepts and preliminary observations on the triple-dimensional analysis of complex volatile samples by using GCxGC-TOFMS. Anal Chem 73: 1336–1344

Smedsgaard J, Nielsen J (2005) Metabolite profiling of fungi and yeast: from phenotype to metabolome by MS and informatics. J Exp Bot 56: 273–286

Stewart D, Wilson HM, Hendra PJ, Morrison I (1995) Fourier-transform infrared and Raman spectroscopic study of biochemical and chemical treatments of oak wood (*Quercus rubra*) and barley (*Hordeum vulgare*) straw. J Agric Food Chem 43: 2219–2225

Sumner LW, Mendes P, Dixon RA (2003) Plant metabolomics: large-scale phytochemistry in the functional genomics era. Phytochemistry 62: 817–836

Tanaka A, Taper BA, Popay A, Parker EJ, Scott B (2005) A symbiosis expressed non-ribosomal peptide synthetase from a mutualistic fungal endophyte of perennial ryegrass confers protection to the symbiotum from insect herbivory. Mol Microbiol 57: 1036–1050

Tarpley L, Duran AL, Kebrom TH, Sumner LW (2005) Biomarker metabolites capturing the metabolite variance present in a rice plant developmental period. BMC Plant Biol 5: Available at http://www.biomedcentral.com/1471-2229/5/8 (accessed December 2005)

Taylor J, King RD, Altmann T, Fiehn O (2002) Application of metabolomics to plant genotype discrimination using

statistics and machine learning. Bioinformatics 18: S241–S248
Tikunov Y, Lommen A, de Vos CHR, Verhoeven HA, Bino RJ, Hall RD, Bovy AG (2005) A novel approach for non-targeted data analysis for metabolomics: large-scale profiling of tomato fruit volatiles. Plant Physiol 139: 1125–1137
Tolstikov VV, 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
Vaidyanathan S, Kell DB, Goodacre R (2002) Flow-injection electrospray ionization mass spectrometry of crude cell extracts for high-throughput bacterial identification. J Am Soc Mass Spectrom 13: 118–128
Vo TD, Lim SK, Lee WNP, Palsson BO (2006) Isotopomer analysis of cellular metabolism in tissue culture: a comparative study between the pathway and network-based methods. Metabolomics 2: 243–256
Vorst OF, de Vos CHR, Lommen A, Staps RV, Visser RGF, Bino RJ, Hall RD (2005) A non-directed approach to the differential analysis of multiple LC/MS – derived metabolic profiles. Metabolomics 1: 169–180
Want EJ, Maille GO, Smith CA, Brandon TR, Uritboonthai W, Qin C, Trauger SA, Siudak G (2006) Solvent-dependent metabolite distribution, clustering, and protein extraction for serum profiling with mass spectrometry. Anal Chem 78: 743–752
Ward JL, Harris C, Lewis J, Beale MH (2003) Assessment of ^{1}H-NMR spectroscopy and multivariate analysis as a technique for metabolite fingerprinting of *Arabidopsis thaliana*. Phytochemistry 62: 949–957
Ward JL, Baker JM, Beale MH (2007) Recent applications of NMR spectroscopy in plant metabolomics. FEBS J 274: 1126–1131
Weckwerth W, Morgenthal K (2005) Metabolomics: from pattern recognition to biological interpretation. Drug discovery today 10: 1551–1558
Weckwerth W, Loureiro ME, Wenzel K, Fiehn O (2004) Differential metabolic networks unravel the effects of silent plant phenotypes. Proc Natl Acad Sci USA 101: 7809–7814
Winder CL, Goodacre R (2004) Comparison of diffuse-reflectance absorbance and attenuated total reflectance FT-IR for the discrimination of bacteria. Analyst 129: 1118–1122
Wysocki VH, Resing KA, Zhang Q, Cheng G (2005) Mass spectrometry of peptides and proteins. Methods 35: 211–212
Yang W-L, Bernards MA (2007) Metabolite profiling of potato (*Solanum tuberosum* L.) tubers during wound-induced suberization. Metabolomics 3: 147–159
Young CA, Felitti S, Shields K, Spangenberg G, Johnson RD, Bryan GT, Saikia S, Scott B (2006) A complex gene cluster for indole-diterpene biosynthesis in the grass endophyte Neotyphodium lolii. Fungal Genet Biol 43: 679–693

Edited by C. Guy