

Annual Plant Reviews (2011) **43**, 1–24
doi: 10.1002/9781444339956.ch1

<http://onlinelibrary.wiley.com>



Chapter 1

PLANT METABOLOMICS IN A NUTSHELL: POTENTIAL AND FUTURE CHALLENGES

Robert D. Hall

*Plant Research International, Wageningen University and Research Centre (Wageningen-UR), PO Box 16, 6700 AA Wageningen, The Netherlands
Centre for BioSystems Genomics, PO Box 98, 6700 AB Wageningen, The Netherlands
Netherlands Metabolomics Centre, Einsteinweg 55, 2333 CC Leiden, The Netherlands*

Abstract: In just 10 years, plant metabolomics has been transformed from a purely theoretical concept into a highly valued and widely exploited technology. Moving on from the many and wide-ranging hopes, enthused upon in a multitude of early reviews, metabolomics for plant research has already proved itself despite the technology still experiencing certain limitations. We have a long road to travel before we reach our desired destination – a very happy place where large-scale, (semi-) automated, unbiased multiplex analyses of plant materials, leading to exhaustive lists of named metabolites become possible. This biochemical Holy Grail will, by definition, never be reached. Nevertheless, continued advances in hardware, software and biostatistics are enabling us to generate ever-advancing, detailed insights into the chemical composition of plants and how this is influenced by genetical and environmental perturbation. There continues to be a major driving force behind further developments, spurred on by myriad existing and potentially new applications in both applied and fundamental plant science. It is no longer so much a case of *hope springs eternal* but rather, *who dares wins!*

Keywords: applied metabolomics; fundamental research; crops; technology development; systems biology

1.1 The history and the goals of plant metabolomics

For thousands of years, man has understood the importance, value and even dangers of the chemicals that are to be found in plant organs, be they fresh,

Annual Plant Reviews Volume 43, *Biology of Plant Metabolomics*, First Edition. Edited by Robert Hall.
© 2011 Blackwell Publishing Ltd. Published 2011 by Blackwell Publishing Ltd.

2 ■ Biology of Plant Metabolomics

dried, fermented or processed in many other ways. Learning which plants are nutritious, poisonous, curative, fibrous, etc., has been of fundamental importance to the survival and evolution of humans and has shaped mankind as it is today. However, recognition is far from understanding and while even the ancient Greeks and Romans had a significant knowledge of the extraction and application of plant materials for a myriad of applications, it is only in modern times that we have had the technologies available to extract stably, separate and identify some of the individual components in plants that form the basis of their exploitable properties be they nutritional, medicinal or otherwise. Developments in analytical chemistry, related to aspects of both separation and detection, have been central to establishing our capacity to decipher the biochemical composition of usually, highly complex tissues and extracts. Recent developments have dramatically improved our analytical potential and the emergence of metabolomics as an analytical concept has initiated a paradigm shift in the way we think about and approach metabolic analyses and indeed, biological experiments in general.

Metabolomics, as a follow-on from transcriptomics and proteomics was a term coined at the end of the 1990s (Oliver *et al.*, 1998) (Table 1.1). The concept entailed the analyses of the metabolite composition of biological materials, aimed to be fully complementarily with the other, potentially unbiased or non-targeted 'omics' approaches. Restrictions were set, and primarily, metabolomics was specifically focused on the smaller metabolites, while the larger organic polymers were excluded. Concentrating on these smaller molecules enabled the focus to be directed to global sets of compounds involved both in primary metabolism as well as in so-called secondary metabolism, involving metabolites that may or may not (yet) have a proven role in the daily functioning of a plant cell or tissue but that might be of great importance in the long-term survival of the individual plant or whole species. Low molecular weight metabolites are the (end) products of a huge network of metabolic pathways and represent the activities of cell regulatory processes (Fiehn *et al.*, 2011). As such, they advertise the response of biological systems to a variety of genetical and environmental responses (Fiehn, 2002). Detecting and monitoring such global sets of metabolites quickly enables us to assess changes in the distribution and concentration of a broad range of potentially biochemically-unrelated compounds and such strategies therefore permit the detection of perturbations at multiple levels of organization: from cell to whole organ.

The primary goal of any plant metabolomics approach is therefore, to gain a helicopter view of metabolism at a specific point in time, in a chosen tissue, obtained either under control or experimental conditions. Extrapolating this to use time-resolved samples, taken at appropriate intervals, can then also introduce a degree of dynamics to the system. However, even when employing a number of extraction, separation and detection conditions (see below) the view gained will never be truly holistic as some element of bias will always be involved. This results from a failure to extract or detect certain compounds

Table 1.1 A short list of some of the most commonly encountered technological terms in the metabolomics literature

Plant metabolomics	An analytical approach focused on generating the least biased and most comprehensive qualitative and quantitative overview of the metabolites present in a tissue, organ or whole plant.
The plant metabolome	The complete complement of low molecular weight molecules present in a specific plant. Generally, 'low molecular weight' usually refers to those molecules smaller than 1500 Da.
Metabolic (metabolomic) fingerprinting	Screening of the metabolic composition of an organism, usually in a high-throughput approach involving large sample numbers. Quantification is usually relative to other (control/treated) samples and the initial goal is generally sample comparison and discrimination analysis. Identification of the metabolites present is not usually performed and differences are based upon contrasts in spectral pattern. Fingerprinting allows the most contrasting samples to be discriminated and selected further for more detailed analysis such as metabolic profiling.
Metabolic (metabolomic) profiling	In contrast to fingerprinting, metabolic profiling aims to identify and quantify at least some of the metabolites present, often focusing on those which have been identified through multivariate analyses as being discriminatory between samples/treatments/genotypes, etc. of the metabolites present in an organism. Such compounds may then form the basis of hypotheses linking them to genetic or phenotypic differences.
Targeted analysis	Generally refers to more traditional methods of chemical analysis where the extraction, separation and detection approach chosen has been focused on and optimized for a specific, chosen group of metabolites that have similar properties (e.g. amino acids) or that share a common biosynthetic pathway (e.g. flavonoids). Methods are usually fully quantitative and comprehensive.
Lipidomics	Specific metabolomic characterization of lipid species

that perhaps are unstable or that have chemical or chemico-physical properties unsuited to the methodologies chosen. Nevertheless, educated choices as to the best approaches to use and optimization of data collection and mining strategies have greatly enhanced our capacity to expand our biochemical knowledge of plant materials.

Initially, concepts such as 'targeted' and 'untargeted' analyses were used where the latter was essentially referring to a true metabolomics approach while the former was more equivalent to well-established analytical chemistry methods. These are still useful descriptors although the delineation is greying somewhat as the technologies develop. Lipidomics, for example, is a prominent and established form of metabolomics for 'untargeted' analyses

4 ■ Biology of Plant Metabolomics

but has been specifically optimized for the huge group of lipids and lipid-related compounds such as the sterols, phospholipids, glycerides, waxes etc (Lessire *et al.*, 2009). Other sub-themes for plants might also be expected to be developed in the future for other major groups of plant compounds such as the terpenoids, alkaloids or phenolics, each of which are already known to contain many thousands of different chemical structures, often with identical accurate masses and elemental formulae.

Other early terms regularly used were metabolic ‘fingerprinting’ and metabolic ‘profiling’. The former was generally taken to refer to the use of machine output as a potentially recognizable chemical pattern, specific to an individual sample. These unique fingerprints were usually the starting point for comparative metabolomics where the researcher wished to compare up to several hundred extracts in order to quickly assess the degree of variation and often, select the most divergent samples or genotypes for more detailed study (Hall, 2006; Saito *et al.*, 2006). There are multiple and diverse examples of such approaches being used in the plant field for a wide range of applications (Hall *et al.*, 2005, 2008; Saito *et al.*, 2006; Schijlen *et al.*, 2008). Specific software tools have also been developed to speed up and semi-automate this process and to optimize the output. In contrast, metabolic profiling is a term that has been employed to refer to a deeper form of analysis where one proceeds to complete metabolite structural elucidation. But again here also, as the technology is progressing, the boundaries are again becoming vague. For example, several labs have already developed extensive in-house databases for mainly primary plant metabolite analyses applicable with their own specific instrumentation. This enables them to identify unambiguously, up to 150 polar small molecules in a non-targeted approach (Ferne, 2007). In such cases, metabolic fingerprints are becoming more and more annotated and are moving towards extensive, true metabolite profiles. Metabolite identification, and particularly of so-called secondary plant metabolites remains however, a significant challenge, or indeed a major bottleneck. This aspect requires extensive, unified multi-disciplinary approaches crossing many different laboratory boundaries (see below).

1.2 The technologies

Perhaps one of the most formidable tasks facing any inexperienced researcher approaching metabolomics for the first time is to become familiar with the technologies available. These are not only extensive but also, are in a continual state of modification and improvement (Saito *et al.*, 2006; Weckwerth, 2007). User-friendliness is also in no way enhanced by the extensive use of abbreviations for the different approaches (see Table 1.2). Most metabolomics experiments involve combinations of separation and detection technologies that can also be used in serial or even in parallel combinations, often referred to as ‘hyphenated approaches’. Simpler forms are commonplace – such as LC-MS or GC-MS but extreme examples can be daunting – such as

Table 1.2 A non-exhaustive list of the most common technological abbreviations likely to be encountered in the metabolomics literature

APCI	Atmospheric pressure chemical ionization
APPI	Atmospheric pressure photo-ionization
CAS	Chemical abstracts service
CE	Capillary electrophoresis
CID	Collision-induced dissociation
DI-MS	Direct infusion mass spectrometry
ESI	Electrospray ionization
FIA/DIA	Flow injection analysis/direct infusion analysis (also FI-MS or DI-MS)
FT-ICR-MS	Fourier transform – ion cyclotron resonance – MS (or FTMS)
HCD	High-energy collision-induced dissociation
HILIC	Hydrophilic interaction chromatography
HPLC	High performance (pressure) liquid chromatography
HTP	High throughput
ICP-MS	Inductively coupled plasma MS
LC/GC	Liquid/gas chromatography
LTQ	Linear trap quadrupole
MALDI-MS	Matrix assisted laser desorption ionization – MS
MS	Mass spectrometry
MS/MS; MS ⁿ	Double (MS/MS) or multiple levels (MS ⁿ) of molecular fragmentation/re-fragmentation with MS detection (also Tandem MS)
<i>m/z</i>	Mass/charge ratio
netCDF	Network common data form
NIST	National Institute of Standards and Technology (metabolite database) (http://www.nist.gov/srd/analy.htm)
NMR	Nuclear magnetic resonance
PDA (DAD)	Photodiode array detection (diode array detection)
PI	Photo-ionization
SEC	Size exclusion chromatography
SPE/SPME	Solid phase extraction or solid phase micro-extraction
TOF	Time of flight (also ToF)
UPLC	Ultra performance liquid chromatography

Source: Modified from Hardy and Hall (2011).

HPLC-PDA-SPE-NMR-ESI-(ToF)MS (Moco & Vervoort, 2011). In this case, one separation technique (HPLC) has been combined subsequently with UV/Vis spectral detection (PDA) after which the sample is split and one fraction proceeds to an electrospray ionization (ESI) unit before entering an accurate mass Time of Flight MS ((ToF)MS). Simultaneously, the other fraction is adsorbed and concentrated using a Solid Phase Extraction (SPE) unit so that sufficient quantities of individual compound peaks can be collected for detection and identification using Nuclear Magnetic Resonance (NMR). Basically, the hyphenated code just refers to the analytical workflow used for that particular analysis. Many reviews have been written on the separation and detection technologies available for plant metabolomics so we include here only a nutshell-type introduction. For more detailed information the reader is referred to seminal volumes such as: Saito *et al.* (2006), Weckwerth (2007)

6 ■ Biology of Plant Metabolomics

and Hardy and Hall (2011). Compact summaries have also been provided by Browne *et al.* (2011) and Beale and Sussman (2011).

1.2.1 Extraction, separation and detection

In contrast to the other ‘-omics’ approaches, which are focused on DNA sequencing (genomics), gene expression analysis (transcriptomics) and proteins (proteomics), the major challenge for complementary, untargeted metabolomics approaches is related to the huge diversity of molecules present, their different physico-chemical properties and their dynamic range (Makkar *et al.*, 2007; Cevallos-Cevallos *et al.*, 2009; Fernie & Keurentjes, 2011). This chemical complexity within just the smaller molecules, involving both the basic structures as well as their different combinations of functional groups such as hydroxyls, carboxyls, amines, etc. has been well documented (Saito *et al.*, 2006; Fernie & Keurentjes, 2011). This complexity is also both the primary reason for our desire to develop metabolomics approaches in the first place, as well as being the primary reason why we shall unlikely ever succeed in obtaining a truly holistic overview of the complete metabolite profile of a plant (Hall, 2006). Dynamic range is a particular challenge as compounds with high biological relevance, relating for example, to their bioactivity or physiological importance, can be present in plant tissues at concentrations differing by a multitude of orders of magnitude – from molar to nanomolar or maybe even lower.

If a compound is not extracted from a sample it can, of course, never be subsequently detected. The choice and optimization of sample preparation and extraction procedures are therefore critical (Hall, 2006; Weckwerth, 2007; Hardy & Hall, 2011). Polar/semipolar lipophobic extraction procedures, often based on hot water or alcohol/water mixtures, as well as lipophilic extraction methods (e.g. often using chloroform) are widely used with the choice depending on the main groups of compounds of interest. Such protocols can also be used sequentially. Methanol-water-chloroform mixtures are also popular as they allow extraction of a range of both hydrophilic and hydrophobic compounds in a single method. For volatile components, organic solvents or Solid Phase Extraction approaches can be employed (Tikunov *et al.*, 2007; Verhoeven *et al.*, 2011). In all cases, the efficiency and balance of compounds moving from the biological sample into the extract determine the quality of the extract and thus how representative it is of the original sample. Inevitably, an element of bias is already introduced at this stage as few compounds will be extracted to 100%. This will later be reflected in the analytes ultimately detected and measured.

Chromatographic separation using either gas (GC) or liquid (LC) phases are very common and are widely applied for different groups of compounds. GC can be used for naturally volatile compounds at temperatures up to 250°C but is also used for heat stable molecules that can be structurally modified through a chemical derivatization process to make them so.

GC-MS of (semi)polar primary metabolite extracts is one of the most widely used metabolomics approaches currently employed (Fiehn *et al.*, 2011). LC methods often involving high pressure (HPLC) or Ultra Performance (pressure) (UPLC) are particularly popular with plant scientists (Verhoeven *et al.*, 2006). Protocols can be developed that are highly suited to many of the (semi)polar secondary metabolites in which plants can be particularly rich. A final separation option, which is not yet widely applied but which is becoming increasingly popular in certain labs, is Capillary Electrophoresis (CE) (Soga, 2007). CE-MS in certain circumstances can have particular advantages relating to sensitivity, rapidity and resolving power (Timischl *et al.*, 2008). For a full review of current CE developments, please see the special issue of *Electrophoresis*, 2009 volume 30, issue 10.

For metabolite detection after separation, there are basically two key players – NMR and MS. Each has its own advantages and disadvantages (see Table 1.3). NMR requires relatively minor sample preparation, is non-destructive and inherently quantitative. It is also not restricted to specific compound groups and has the potential to give unambiguous information for metabolite identification. The current greatest drawback relates to its lower sensitivity compared to MS-based approaches and its requirement for relatively large samples. However, recent improvements are making this less of an issue. MS has a wide dynamic range and high sensitivity but does require molecules to be ionized (charged) in order for them to enter the instrument. MS also requires extensive sample preparation and extraction. Furthermore, these highly complex extracts, typical of plants, despite good chromatography, can still be prone to significant ion suppression/matrix effects that can interfere with or mask molecule detection. This, together with variable ionization frequencies, makes MS-based quantification more difficult and totally dependent on available reference standards. Recent developments in improved mass accuracy machines such as the Orbitrap™ and FT-ICR-MS instruments, have created additional interest due to their greater applicability regarding empirical formula calculations and metabolite identification.

1.2.2 Data generation, storage, processing and mining

Metabolomics is a data-rich and also data-driven technology and the implementation of broad spectrum chemical analyses has generated considerable technical and intellectual challenges. Successful metabolomics research is as dependent upon the ‘dry’ side of science as on the ‘wet’ side. Effective data management tools are essential for storing both raw and processed data, and dedicated bioinformatics tools are crucial for subsequent data analysis, integration, visualization and ultimately therefore, their conversion into information and biological knowledge (Fiehn *et al.*, 2011; Redestig *et al.*, 2011). It is the size of the datasets that current instrumentation produces and the comprehensiveness of the biochemical analyses that can be performed that place all aspects of data management at the core of metabolomics technologies

8 ■ Biology of Plant Metabolomics

Table 1.3 The three main metabolomic approaches most commonly used with their contrasting advantages and disadvantages

Technology	Advantages	Disadvantages
NMR	<ul style="list-style-type: none"> Non-destructive Quantitative Robust, long-established technology Good software and metabolite database support Short measurement time Almost no sample preparation needed No separation required No derivatization required Detects all organic compound classes Structural identification of unknown compounds relatively straightforward Compatible with liquids and solids 	<ul style="list-style-type: none"> Relatively low sensitivity Expensive instrumentation Limited to protonated compounds Cannot detect salts and inorganic ions Relatively large (0.5 mL) samples required
GC-MS	<ul style="list-style-type: none"> Relatively inexpensive Quantitative (when standards available) Modest sample size required Good sensitivity Good software and metabolite database support Detects most organic and some inorganic molecules Excellent separation reproducibility 	<ul style="list-style-type: none"> Destructive Non-volatiles need derivatization Unsuitable for heat-labile components Extensive sample preparation procedures Separation step means longer measurement times (20–50 min) Identification of unknown compounds is difficult
LC-MS	<ul style="list-style-type: none"> Flexible technology Good sensitivity Detects most organic and some inorganic molecules Small sample size (mg) requirement Direct injection can be used for very rapid analysis (1–2 min) Has potential for detecting largest portion of metabolome 	<ul style="list-style-type: none"> Destructive Quantification limited Relatively expensive instrumentation Extensive sample preparation procedures Separation step means longer measurement times (20–50 min) Resolution and reproducibility poorer than GC Instrumentation less robust than NMR or GC-MS Software and metabolite database support poor Identification of unknowns is difficult

Source: Modified from Wishart (2008).

The choice of metabolomics approach is determined by a range of factors including experimental aims, compound groups of greatest interest, sample number and the instrumentation budget available.

(Sumner *et al.*, 2008). Indeed this is one of the key factors discriminating metabolomics from the more traditional, analytical chemistry approaches. Success of the technology will, however, only be achieved once our capacity to extract information from metabolomic data increases and allows proper, unlimited annotation of identified metabolites. This essentially requires tools

for rapid *de novo* metabolite identification, which we do not yet have, and for this reason, we also do not yet have any kind of reliable, even semi-accurate estimation of the true size of any plant species metabolome (Fiehn *et al.*, 2011).

Reliably and robustly extracting qualitative and quantitative information from metabolomics datasets has become a science in itself and requires a relatively high level of training. Many steps are involved once machine output has been collected and the specific steps chosen are also dependent upon the biological goals defined at the time of experimental design, the instrumentation used and the background biochemical information already available. Generally speaking, the process of data to knowledge conversion involves multiple steps including noise removal (filtering), peak alignment, peak picking and deconvolution, peak identification, etc. The final data matrix thus created usually then needs to be subjected to one or more statistical processing approaches appropriate both to the experimental design and the type and quality of the data generated. Commonly, multivariate statistical analysis tools are employed in order to perform effective comparative discriminative analyses between samples, genotypes, treatments, etc. Furthermore, as the technologies advance and research grows to involve multiple metabolomics approaches (e.g. LC-MS in positive/negative modes; GC-MS of both natural volatiles and non volatiles, NMR, etc.), which we may also later wish to combine with complementary genetic/genomic information (such as QTL or SNP data) or even macro-phenotypic information, the whole area of data integration or data fusion is developing rapidly (Fernie & Keurentjes, 2011; Moing *et al.*, 2011; Tikunov *et al.*, 2010). Through these data fusion approaches, we are gradually gaining a better ability to link differences at the molecular level to those at the level of the phenotype. There is also a growing requirement for useful visualization tools to bring this multitude of complex data, figuratively and literally, down to more user-friendly dimensions (Fiehn *et al.*, 2011; Redestig *et al.*, 2011).

The entire process of data generation and analysis is excellently reflected in the working examples published by Sumner *et al.* (2008), Kopka *et al.* (2011), and in various chapters in Saito *et al.* (2006) and Hall and Hardy (2011). Specific examples of the tools that have been developed for data (pre)processing have also been published by, for example, Lommen (De Vos *et al.*, 2007; Lommen, 2009) for metAlign; Kopka (Luedemann *et al.*, 2008) for TagFinder, and (<http://masspec.scripps.edu/xcms/xcms.php>) for XCMS. Tools for data analysis and biostatistics have also been broadly covered (see full reviews by Fiehn *et al.*, 2011; Redestig *et al.*, 2011). The scale of the task and the unique challenges that metabolomics throws at us mean that we are still distant from our ultimate objective, which is to perform (semi)automated comprehensive metabolic data analyses involving large sample numbers, to visualize this data in the form of pathways and gain a full overview of plant metabolic networks and how these are perturbed under a controlled and defined set of genetic and environmental conditions. As stated by Kopka *et al.* (2011), for several years, a major bottleneck has been the lack of options for multi-parallel

10 ■ Biology of Plant Metabolomics

chromatographic data processing. While this has now been at least partially overcome, current methodologies are not yet sufficiently robust to permit the deconstruction of MS data into quantified compound lists. Furthermore, the challenge will grow as there is a great desire for time-resolved metabolomics where aspects of the dynamics of metabolomic networks can be included and assessed. We still require many more bioinformatics and biostatistical tools for the metabolomics toolbox. Different analytical approaches, to tackle a wide range of biological questions, require different tailor-made solutions regarding data handling. Nevertheless, although we can still consider the data management aspect of metabolomics to be the biggest bottleneck – or rather, to be the main component in the current workflow, which is most sub-optimal – the progress that has been made is impressive, especially as it has involved multi-disciplinary collaboration between analytical chemists, computer scientists, statisticians, modellers and biologists.

1.3 The applications

Information on the biochemical composition of plants and plant-derived products including foodstuffs, medicinal preparations, etc. is very often of intrinsic biological interest. For the fundamental researcher, metabolites are the ultimate end points of gene activity but are by no means a direct true reflection thereof. Physical, temporal, or spatial differences in gene expression are often reflected in changes in biochemical pathway activity but additional levels of complexity are introduced by post-transcriptional and post-translational modifications as well as aspects of compound compartmentation, non-enzymatic (chemical) reactions, metabolite stability, vaporization, polymerization and complex-forming, etc. The link between metabolite profile and phenotype can however be immediate and consequently, scientists with a primary interest in the molecular organization of plants and how this is controlled often have a strong desire to identify biochemical changes as a potential basis of causality for phenomena of interest. For example, gene function analysis is an area of growing importance. The rapid increase in the number of whole genome sequences that are being generated through the recent revolution in hardware design of so-called Next Generation Sequencers has generated unimagined potential (Varshney *et al.*, 2009; Cyranoski, 2010). Mutations in many genes are phenotypically ‘silent’ whereas at the molecular/biochemical level, clear changes may take place. These changes can be of great importance but in a phenotypically, non-visible way. Random mutation approaches often involve genes of no predicted function and hence, a complementary, non-targeted metabolite screening approach is a very suitable starting point for pinning gene function to an enzymatic or structural role in the molecular organization of a cell (Hall, 2006; Fernie & Keurentjes, 2011).

For the applied scientist, and particularly those in the plant breeding or food processing industries, metabolomics has also rapidly gained a strong

foothold. In both cases, key traits of importance to the industry are more often than not directly linked to the biochemical composition of the plant materials used. Endogenous features of food products such as nutritional value, sweetness, digestibility and bioavailability of nutrients as well as exogenous features such as general appearance, pigmentation, fragrance, etc. are fundamentally determined by the biochemical composition of the starting materials (Stewart *et al.*, 2011). For plant breeders, in progeny populations important features that are regularly selected for are disease resistance, pigmentation (appearance), shelf-life, etc. Once again, such characteristics are also to a large extent determined or mediated by the biochemical composition of the plants or plant organs (seeds, fruits, leaves, etc.) concerned. Many of these traits are complex and are not dependent on individual metabolites but rather, on a mix of metabolites that also generally have to be present in the correct balance. Taste and fragrance are two such complex traits that are determined by highly complex mixtures of metabolites from divergent chemical classes that act in harmony to produce the sensory sensation we experience (Hall *et al.*, 2008; Fitzgerald *et al.*, 2009; Bovy *et al.*, 2010). Involved here are not only components considered to give a positive sensory response but also many biochemicals that are linked to off-flavour or malodorous features. These are generally considered as being undesirable in food materials although they may of course be of importance in the growth and 'fitness' of the whole plant in nature.

1.3.1 Metabolomics and fundamental plant research

As is true for much fundamental plant research from the last two decades, extensive focus has been placed on the model plant *Arabidopsis thaliana*. The significant short life cycle and genetic advantages of this compact species have transformed a common weed into the most studied plant species on the planet. Inevitably, *Arabidopsis* has also rapidly become a favoured target for metabolomics studies (Beale & Sussman, 2011). Much functional genomics research has also been extended to include metabolic fingerprinting and profiling in the quest to understand better the function and molecular impact of specific gene mutations on the molecular patterning of plants. Databases for *Arabidopsis* metabolites have also been created and, as reviewed by Fiehn *et al.* (2011), the AraCyc pathway database has become one of the most extensive and most used in plant metabolomics, comprising more than 2500 metabolites, 5500 enzymes and 350 metabolic pathways. The intrinsic advantages of *Arabidopsis* as a model for other studies based, for example, on plant physiology, genetics or genomics have resulted in extensive metabolomics research taking place in these areas as well. An extended library of well-defined genetic mutants covering important features related to developmental, physiological and life cycle processes regularly provides materials that have been subjected to detailed metabolomics analyses. As might be anticipated, many of these mutations have been shown to incur extensive and broadly classed

12 ■ Biology of Plant Metabolomics

metabolic changes in comparison to the wild-type reference material (see references in Beale & Susman, 2011). Furthermore, this research emphasizes how wide the consequences of even point mutations in single genes can be in metabolic terms, even when the degree of visible phenotypic change may be limited. However, care needs to be taken regarding primary and secondary effects as fundamental changes in basic metabolism as a result of a mutation, has the potential to instigate a cascade of subsequent (secondary) metabolic effects, essentially only indirectly related to the initial mutation and thus lacking direct causality.

Metabolomics is also gaining ground in the area of general plant cell physiology research and also in the field of environmental research. Genotype × environment interaction is a topic of basic importance regarding how plants survive and function under what is normally an ever-changing environment (Ahuja *et al.*, 2010; Browne *et al.*, 2011; van Dam & van der Meijden, 2011). Furthermore, its importance in a crop production context enhances the value of this fundamental research (Browne *et al.*, 2011; Draper *et al.*, 2011). Metabolomics research has shown that changes in plant physiology, be they related to genetic or environmental perturbation, often involve widespread alteration in plant biochemistry not specifically restricted to particular pathways (Beale & Sussman, 2011). Abiotic stress for example, be it induced by temperature, drought or sub-optimal nutrient supply, has been shown to induce extensive reprogramming of the plant metabolome (see review by Browne *et al.*, 2011). Furthermore, in nature, the situation can become quite complex as abiotic stress factors also often occur in combination (e.g. heat + drought; salt + drought). Indeed, abiotically stressed plants may also have modified capacity to cope with additional biotic challenges such as insect or pathogen attack. Insight into the overall metabolic effects of these parallel stresses is highly desirable in our quest to understand better the molecular basis of plant stress responses in a natural, uncontrolled environment. This knowledge is needed in order to allow it to be exploited in the general context of crop yield research, which has a clear link to future food security issues, which are becoming very much hot topics considering the growing global population (Ahuja *et al.*, 2010).

The link between genotype and (metabolic) phenotype is also an area of growing interest and while *Arabidopsis* has here, once again a starring role (Keurentjes *et al.*, 2006; Keurentjes, 2009), other species such as tomato and melon are also proving interesting experimental subjects (Biais *et al.*, 2009; Fernie & Keurentjes, 2011; Moing *et al.*, 2011), particularly regarding fleshy fruit tissues that *Arabidopsis* cannot cover (de Vos *et al.*, 2011). Now that the genome of *Arabidopsis* has been drafted, widespread use of natural variants (e.g. ecotypes) and other artificial/genetic variants is being made to map the *Arabidopsis* genome and identify the genetic determinants of the biochemical composition of plants. This is leading to the development of the concept of genetical metabolomics as has been defined by Keurentjes (2009) as being an extension or refinement of the more general topic of genetical genomics (Jansen & Nap, 2001). Here, through such multi-disciplinary

(multi-omics) approaches, a more detailed functional genetic map of the Arabidopsis genome is being generated, which in turn is proving a highly valuable tool for the location and identification of genes playing central or structural roles in plant metabolism, which is ultimately translated from metabolite profiles into physical and chemical phenotypes.

1.3.2 Metabolomics and applied plant research

A logical and valuable extension of fundamental metabolomics approaches is to have these translated into a more applied context (Stewart *et al.*, 2011). Applied plant research is generally focused on crop plants, predominantly linked to the food industry but crops also have important roles to play in a variety of other industries related to fibre production for the textile industry, biomass production for use in energy generation or the synthesis of biofuels, wood for the building industry, etc. Furthermore, plants, being nature's most prolific biochemists (Hall, 2006), are a long-standing source of inspiration for industrial chemists in the search for novel bioactive compounds that can form the basis for the mass production of new synthetic drugs for the pharmaceutical industry, or biocides such as fungicides and insecticides for use in agriculture. For these latter examples, the link to metabolic profiles is immediately apparent. In addition, for the other applications, the value and quality of the crop products are generally directly linked to the chemical composition of the materials involved. For food products, the biochemical composition determines all aspects of nutritional value, taste and fragrance, appearance, etc. as well as potentially negative aspects such as toxicity, bitterness and astringency. It is therefore not surprising that exploitation of metabolomics approaches in applied plant science has proceeded, right from its inception, in parallel with technology development and its use in fundamental research as is reflected in many of the earliest publications at the start of the 1990s.

Metabolomics approaches involving food crops already include a wide variety of materials linked to both fresh and processed food products. Many of the main examples have been reviewed by Stewart (Stewart *et al.*, 2011). However, the list is considerably more extensive than has been covered there. Most of the world's major crops such as tomato (Bovy *et al.*, 2010), potato (Dobson *et al.*, 2008), rice (Fitzgerald *et al.*, 2009), legumes (Farag *et al.*, 2009), grape (Periera *et al.*, 2005), wheat (Graham *et al.*, 2009) and corn (Roehlig *et al.*, 2009) have been subjected to extensive metabolomics analyses, with the focus being placed on a variety of different biological questions. However, in addition, many more minor crops have been targeted, as widely diverse as raspberry (Beekwilder *et al.*, 2005), Brassica's (Hall *et al.*, 2010), *Citrus* and *Prunus* spp. (Arbona *et al.*, 2009) and strawberry (Hanhineva *et al.*, 2008). These analyses are regularly focused specifically on aspects of product quality but metabolomics approaches are also regularly applied to tackle problems related to crop fitness involving disease resistance and its molecular basis, plant–insect interactions, mycorrhizal interactions, stress effects, etc. (Ahuja *et al.*, 2010; Draper *et al.*, 2011). Through these approaches

14 ■ Biology of Plant Metabolomics

we are gradually gaining a much more detailed picture of crop biochemistry and its link to nutritional, sensory and other crop-related characteristics. Following on from this, we are developing new tools to improve those products offered to consumers in the local supermarket and also, to provide these consumers with additional information that they can use to make their own informed choices regarding the quality and health benefits of the food we buy. Included here are aspects such as tracing and tracking, where metabolomics is being applied to help identify biomarkers for proof of origin/authenticity as well as features related to product purity and contamination (Hall, 2006; Stewart *et al.*, 2011).

Metabolomics of non food crops has also received considerable attention. Quite a number of these have been centred upon 'nature's chemical arsenal' where metabolomics is being used to investigate the biochemical composition of a range of species of importance to the pharmaceutical industry. Included here are key crops such as *Artemisia* (Ma *et al.*, 2009), *Echinacea* (Frederich *et al.*, 2010) and *Catharanthus* (El-Sayed & Verpoorte, 2004). The whole field of herbal extracts and their potential pharmaceutical application has also been a major target for metabolomics approaches (Xie *et al.*, 2008; Lan *et al.*, 2010). Tree species are also major world crops as sources of wood for both the building industry and for the paper industry. Here also the quality of the starting material is of paramount importance to the value of the product subsequently generated. Metabolomics has already been applied both to understand aspects of regeneration, tree growth and yield as well as aspects of the biochemical composition in species such as poplar (Morreel *et al.*, 2006), pine (Robinson *et al.*, 2009) and Douglas fir (Robinson *et al.*, 2007).

Taking one step back from crop production to crop varietal design, metabolomics is also attracting great interest in the plant breeding and seed production industries where once again, the technologies are being used to generate information of value in progeny selection to facilitate more targeted breeding approaches (Fernie & Schauer, 2008). Here also there is a focus on using metabolomics to identify biomarkers for both positive and negative crop traits or food quality characteristics, which can then be used to enhance and speed up the breeding process leading to the generation of new and improved varieties. However, whether these biomarkers remain, for the plant breeders, as chemical biomarkers or become linked to or translated into potentially more convenient DNA markers remains a topic of considerable debate. Identifying a DNA marker strongly linked to a biochemical component or profile that is associated with the trait of interest has potential advantages over the chemical marker itself as the breeder may include it with other already-existing, routinely used DNA-based markers (e.g. SNPs, SSRs) in one automated screen not reliant on additional and rather more complex, biochemical analyses. Nevertheless, current applications of metabolomics approaches purely as research tools to understand better the molecular basis of crop biochemistry will continue for many years to come. Our fundamental knowledge of key aspects of crop quality is still sadly lacking and it is

only now that we have such metabolomics approaches at our disposal that we can properly tackle biochemically based traits that are highly complex in the sense that they are dependent upon a qualitative and quantitative mix of chemically diverse components acting in harmony to form the molecular basis of key crop characteristics.

Taking the final step forward from crop production to food processing brings us again to a topic where metabolomics has industrial potential. Using metabolomics approaches to help us understand better those changes taking place between '*farm and fork*' will assist the food processing industry to improve further food processing strategies and optimize the individual steps involved. Consequently, the industry will gain a better position to serve the consumer by providing more nutritious, healthy and attractive products that may also have longer shelf lives and involve less waste. Relatively simple food processing procedures such as milling rice grains to remove the outer bran layers and expose the more attractive white endosperm (Fitzgerald *et al.*, 2009) clearly impact the biochemical composition of the final product. More complex procedures in food processing, typically involving a wider range of steps, from for example, washing and chopping to Pasteurization and cooking, also have significant step-by-step effects on the biochemical composition of the final product. This was for example, recently shown in detail for tomato paste by Capanoglu (Capanoglu *et al.*, 2008, 2010). Using an LCMS-based approach, changes in biochemical composition were clearly visible after each step. These could be often attributed to the manipulations involved and therefore facilitate the identification of those steps that might be targeted to help the processor improve the overall production process with the aim of generating a better quality end product with enhanced nutritive value or a higher, potentially healthier antioxidant content (de Vos *et al.*, 2011). Similar examples have also been recently reported for coffee, tea and wine (Stewart *et al.*, 2011).

The food processing industry itself has also readily embraced metabolomics, more often in relation to the fate and eventual (beneficial) effects of food products or individual food components upon ingestion and subsequent digestion and absorbance into the body (see Stewart *et al.*, 2011). Many nutritionists are using metabolomics to follow various aspects of the interaction between food and humans both related to general aspects of food metabolism as well as in the context of food-related diseases such as metabolic syndrome and diabetes. This whole field has been extensively reviewed, particularly with regard to plant/food–human health concepts (Gibney *et al.*, 2005; Wishart, 2008; Cevallos-Cevallos *et al.*, 2009).

1.4 The bottlenecks, the potential and future challenges

The progress made in the last (or rather, first) 10 years of plant metabolomics research is impressive. Publication numbers continue to increase

16 ■ Biology of Plant Metabolomics

exponentially (Hall & Hardy, 2011) and the field is gaining growing interest in a wide range of research areas. The diversity of plant species targeted, from algae to orchids, and the biological questions tackled, clearly emphasize the broad applicability of the technologies and demonstrate their added value and complementarity to other, already better established, ‘-omics’ approaches. Nevertheless, there remain some key areas of research that are of fundamental importance to all metabolomics experiments, which perhaps understandably lag behind other developments and represent the current limitations in this field. These are the areas that deserve and demand extra attention if we are to bring plant metabolomics to the next level.

1.4.1 Current limitations

Metabolomics is strongly data driven. Nevertheless, it is essential that the biologist maintains the lead in order to guarantee the development and proper implementation of those technologies specifically required to tackle unsolved, fundamental biological questions. However, with metabolomics being a multi-disciplinary technology, giving biologists the lead entails that these researchers need to have a broad understanding of a variety of scientific disciplines. This, however, does not require full individual technological proficiency. Indeed, most groups involve small teams of scientists with complementary scientific backgrounds to perform successful metabolomics experiments. Establishing such teams is a challenge in itself as the members may be required to adapt a significant change in mind set and the use of modified terminologies to achieve optimal communicative teamwork. Nevertheless, it is still essential that the lead biologist does have a basic understanding of all the disciplines involved. This entails a good understanding of the wet lab technologies as well as the statistics behind the whole data analysis procedure, starting with proper experimental design. Both the scale and the richness of metabolomics datasets carry with them an inherent risk of incorrect interpretation and the generation of false conclusions. This requires a new generation of biologists who are confident in multi-variate statistical approaches for complex, large-scale data management and manipulation. Such broad-minded and broadly disciplined individuals are not easy to find.

Converting machine output (data) into information (or knowledge) concerning the biochemical composition of plant material, represents perhaps the most significant single bottleneck to keeping us as biologists, from where we want to be. The challenge is significant. Unambiguous, preferably *de novo*, identification of (all of) the metabolites detected in an MS or NMR analysis of whatever type, is required if we are to take the next step to full scale analysis of metabolic flux, cross-talk between metabolic pathways and networks, and all time-resolved approaches. Tools for such systematic, metabolite annotation are being worked on but it will likely be several years before we are in a position to automatically reel-off lists of named metabolites with full confidence. Associated with this point are also the limitations incurred

through having a lack of authentic standards for the majority of compounds that can now be detected with the current hardware available. Such reference standards are essential for unambiguous formula assignment and the confirmation of metabolite identity and, without these, limitations to metabolic pathway elucidation and flux analysis are clear. Beale & Susman (2011) call for more community-based initiatives to tackle what is effectively a huge and potentially costly problem as the extraction and purification or alternatively, the biosynthesis of authentic standards is prohibitive.

While it is also stated above that we are unlikely ever to be able to perform fully holistic metabolite analyses for a range of different reasons, one major current limitation restricting a broader application of metabolomics approaches is the under-representation of key groups of low abundance molecules (Draper *et al.*, 2011). In particular, the general absence of data on key plant hormones such as the gibberellins, auxin and cytokinins restricts our ability to interpret observed phenotypic changes in the context of metabolite biosynthesis. Mechanistic interpretation of phenotypic and metabolomic modifications requires as broad an overview of metabolites as possible and the often complete absence of phytohormone data remains a significant limitation to biological interpretation.

Typically, a period of rapid technological proliferation is followed by a longer period of consolidation and refinement (Wishart, 2008). One potential limitation that is perhaps still underestimated or at least under-illuminated at present but that has long-term consequences for broader implementation is knowledge of the general robustness of methods and the interchangeability of data between labs. Some approaches and instrumentation are known to be more reliable than others (Browne *et al.*, 2011) but there are still concerns regarding inter-laboratory reproducibility. As a start, recommendations for the standardization of methodologies for data generation and collection – ‘minimum reporting standards’ – have been made in recent publications (Fiehn *et al.*, 2007; Sumner *et al.*, 2007). This should help to facilitate future desires for data exchange and long-term data validity. However, the ultimate desire to create common databases of experimental data and to have a system that enables direct comparison of data from different experiments and different laboratories is considerably more challenging. Contrasting experimental goals may result in different parameter choices for data collection as might also the precise instrumentation available in different laboratories. Furthermore, even after data generation, different labs have different experiences or preferences regarding data preprocessing. Regarding this issue, a number of inter-laboratory comparisons have recently been performed with the view to estimate both technical and non-technical (human) sources of variation in data generation. Data for both NMR (Viant *et al.*, 2009; Ward *et al.*, 2010) and GC-ESI-ToF-MS (Biais *et al.*, 2009) experiments, generated from up to five laboratories using identical, split samples and standard operating procedures, have been reported and early results are encouraging. Nevertheless, the strict requirement to follow precisely defined procedures

18 ■ Biology of Plant Metabolomics

at all steps in the data generation and data manipulation processes is evident. Many such comparisons are required in order to help define better fully standardized procedures that will enable robust and reliable cross-laboratory data exchange and the creation of common metabolite databases from various sources.

Lack of spatial resolution, which is inherently also linked to the sensitivity and resolving power of metabolomics instrumentation, also remains a fundamental limiting factor in the application of metabolomics approaches and the interpretation of the data produced. There is a great desire to reduce metabolomics analysis to the tissue or preferably, single cell level. Under normal circumstances, even adjacent cells in a plant organ can be highly different in their biochemical content. An epidermal cell in the leaf of a plant is often anthocyanin rich and chlorophyll free while the cell immediately below may have exactly the opposite characteristics. Glandular hair cells are also a clear case in point, being often filled with a rich mix of secondary metabolites that may hardly be detectable in adjacent leaf cells (Sumner *et al.*, 2011). Single cell metabolomics would also be highly desirable in investigating the responses of plants at the sub-tissue level at the earliest moments of pathogen attack (Draper *et al.*, 2011). Particularly, resistance mechanisms involving hypersensitive responses can be very rapid, involving just one or a small number of cells. Investigating such responses at the molecular level is particularly difficult as the number of cells responding is but a tiny fraction of the total number of cells present in a sample of the size needed to extract enough metabolites for analysis. This 'dilution' effect is also relevant in many other cases where our current lack of spatial resolution, determined by technical and handling limitations, masks highly relevant changes and makes them invisible. A start has been made to reduce analyses to the level of a small number of collected plant cells (e.g. 200 epidermal cells; Ebert *et al.*, 2010) or individual nerve cells (Lapainis *et al.*, 2009) but we still have some way to go before we can routinely analyze individual plant cells and success will only be possible once both sampling as well as instrument sensitivity issues have been solved.

While there are still clear limitations, challenges are still being tackled and overcome. Perhaps the ultimate goal of metabolomics as a technology should be to make the biological interpretation of the data the only real bottleneck remaining (Redestig *et al.*, 2011). Only then will we have realized the full potential of the technology.

1.4.2 Future potential

Plants clearly have huge natural biochemical diversity, perhaps in reality even greater than we currently imagine. Furthermore, the plant metabolome is not in a steady state but rather, is inherently highly amenable to variation. Under normal growing conditions, as experienced by plants in nature, basic processes such as growth, development, transition to flowering, etc. are accompanied or even determined by a global reprogramming of plant metabolism. Such reprogramming may be predominantly under genetic control in

relation to normal plant development or may be consequential to externally imposed environmental perturbation, which in turn is also predominantly under genetic control. Contemporary research approaches are progressing in their sophistication but nevertheless, we still have poor knowledge of the complexity of such changes and responses. There is a pivotal role here for advanced metabolomics approaches to play a part and help close this gap in our knowledge and help us elucidate the molecular mechanisms behind phenotypic traits.

Ultimately, we aim for a systems biology type of understanding (Coruzzi & Gutiérrez, 2009), where not only biochemical changes but also their inter-relationships and inter-dependencies can be visualized and interpreted at a cell, tissue or whole plant level. This entails effective, multi-disciplinary analysis approaches that also enable subsequent integration and cross-correlation between datasets. We are beginning to produce such heterologous datasets and these are offering tantalizing possibilities for teasing out causative relationships between metabolite levels, transcript levels, enzyme activities etc (Last *et al.*, 2007). Particularly, in relation to phenotypic analyses, metabolomics may end up forming the backbone of systems biology approaches (Redestig *et al.*, 2011) but for this to be realized we need more robust and broadly applicable approaches. Cutting-edge methods have been developed but must continue to be improved for more amenable data interpretation. Results already demonstrate the complexity and inter-dependence of metabolic pathways and their control. This inter-dependence is perhaps an in-built mechanism that has evolved to give the plant maximum biosynthetic flexibility. As a consequence, it does entail that deciphering metabolic processes and their controlling mechanisms is highly complex and will require the use of a wide range of approaches, including metabolomics and involving the exploitation of metabolic mutants and reverse metabolic engineering approaches to identify control elements.

1.4.3 Challenges for the future

Plants cover practically the entire land mass of our planet and (lower) marine plants are also very prominent in the sea. There are an estimated 300,000 plant species, highly adapted to, but also highly flexible in dealing with their environment's abiotic challenges. Furthermore, there is an estimated about 400,000 insect species that have plants as their daily diet (Schultz, 2002). Evolutionary pressure both to cover the planet and populate every suitable niche, however extreme, and to survive and reproduce there despite constant abiotic and biotic onslaught has resulted in a plant kingdom of exceptional diversity. Not more so than in the biochemical diversity, which has been developed to create the so-called biochemical arsenal (Fernie, 2007), that plants need to grow, survive and reproduce. It is this biochemical richness of plants that will always remain the fundamental challenge for metabolomics, but also its key driving force behind our desire for a better understanding of how plants live and function.

20 ■ Biology of Plant Metabolomics

Many of the points listed above represent the basis of most of the challenges plant metabolomics researchers face as we try to move to the next level in technology application. We have a long-standing, fundamental desire to take in the widest of biological horizons, to decipher the principles underlying plant metabolism and to unlock nature's biochemical secrets. Metabolomics is just one small step we can now take in addition to many other tools that have gone before. It is nevertheless, a very important one considering the role of metabolites in biological systems. The technologies will deliver the data needed to bring us considerably closer to linking genotype and phenotype that is for many fundamental and applied scientists of crucial importance in facilitating our understanding the richness that plants have to offer.

Acknowledgements

The author wishes to acknowledge financial support from the Centre for BioSystems Genomics and the Netherlands Metabolomics Centre, both initiatives under the auspices of the NWO-Netherlands Genomics Initiative. This work has been performed as part of the European Union Framework 6 Strategic Research Project META-PHOR (FOOD-CT-2006-036220) from which additional financial support has been received.

References

- Ahuja, I., de Vos, R.C.H., Bones, A.M. *et al.* (2010) Plant molecular stress programmes face climate change. *Trends in Plant Science* **15**, 664–674.
- Arbona, V., Iglesias, D.J. and Talon, M. (2009) Plant phenotype demarcation using non-targeted LC-MS and GC-MS metabolite profiling. *Journal of Agricultural and Food Chemistry* **57**, 7338–7347.
- Beale, M.H. and Sussman, M.R. (2011) Metabolomics of *Arabidopsis thaliana*. In: Hall, R.D. (ed.) *Annual Plant Reviews, Biology of Plant Metabolomics*. Blackwell Publishing Ltd., Oxford, UK. pp. 157–180.
- Beekwilder, J., Hall, R.D. and De Vos, C.H.R. (2005) Identification and dietary relevance of antioxidants in raspberry. *BioFactors* **23**, 197–205.
- Biais, B., Allwood, J.W., Deborde, C. *et al.* (2009) ¹H NMR, GC-EI-TOFMS, and data set correlation for fruit metabolomics: application to spatial metabolite analysis in melon. *Analytical Chemistry* **81**, 2884–2894.
- Bovy, A.G., Gomez-Roldan, V. and Hall, R.D. (2010) Strategies to optimize the flavonoid content of tomato fruit. In: *Recent Advances in Polyphenol Research* Santos-Buelga, C., Escribano-Bailon, M-T., Lattanzio, V. (Eds) *The Handbook of Polyphenols*, pp. 138–162.
- Browne, J., Bacic, A., Tester, M. and Roessner, U. (2011) Abiotic stress and metabolomics. In: Hall, R.D. (ed.) *Annual Plant Reviews, Biology of Plant Metabolomics*. Blackwell Publishing Ltd., Oxford, UK. pp.61–86.

- Capanoglu, E., Beekwilder, J., Boyacioglu, D. *et al.* (2008) Changes in antioxidant and metabolite profiles during production of tomato paste. *Journal of Agricultural and Food Chemistry* **56**, 964–973.
- Capanoglu, E., Beekwilder, J., Boyacioglu, D. *et al.* (2010) The effect of industrial food processing on potentially health-beneficial tomato antioxidants. *Critical Reviews in Food Science and Nutrition* **50**, 919–930.
- Cevallos-Cevallos, J.M., Reyes-de-Corcuera, J-I., Etxeberria, E. *et al.* (2009) Metabolomics analysis in food science: a review. *Trends in Food Science and Technology* **20**, 557–566.
- Coruzzi, G.M. and Guttierrez, R.A. (2009) *Plant Systems Biology*. Blackwell Publishing Ltd., Oxford, UK.
- Cyranoski, D. (2010) The sequence factory. *Nature* **464**, 22–24.
- De Vos, R.C.H., Moco, S., Lommen, A. *et al.* (2007) Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nature Protocols* **2**(4), 778–791.
- de Vos, C.H.R., Hall, R.D. and Moing, A. (2011) Metabolomics of a model fruit: tomato. In: Hall, R.D. (ed.) *Annual Plant Reviews, Biology of Plant Metabolomics*. Blackwell Publishing Ltd., Oxford, UK. pp. 109–156.
- Dobson, G., Shepherd, T., Verrall, S.R. *et al.* (2008) Phytochemical diversity in tubers of potato cultivars and landraces using a GC-MS metabolomics approach. *Journal of Agricultural and Food Chemistry* **56**, 10280–10291.
- Draper, J.H., Rasmussen, S. and Zubair, H. (2011) Metabolite analysis and metabolomics in the study of biotrophic interactions between plants and microbes. In: Hall, R.D. (ed.) *Annual Plant Reviews, Biology of Plant Metabolomics*. Blackwell Publishing Ltd., Oxford, UK. pp. 25–60.
- Ebert, B., Zoller, D., Erban, A. *et al.* (2010) Metabolic profiling of *Arabidopsis thaliana* epidermal cells. *Journal of Experimental Botany* **61**, 1321–1335.
- El-Sayed, M. and Verpoorte, R. (2004) Growth, metabolic profiling and enzyme activities of *Catharanthus roseus* seedlings treated with plant growth regulators. *Plant Growth Regulation* **44**, 53–58.
- Farag, M.A., Deavours, B.E., Fatima, A. *et al.* (2009) Integrated metabolite and transcript profiling identify a biosynthetic mechanism for hispidol in *Medicago truncatula* cell cultures. *Plant Physiology* **151**, 1096–1113.
- Fernie, A.R. (2007) The future of metabolic phytochemistry: larger numbers of metabolites, higher resolution, greater understanding. *Phytochemistry* **68**, 2861–2880.
- Fernie, A.R. and Keurentjes, J.J.B. (2011) Genetics, genomics and metabolomics. In: Hall, R.D. (ed.) *Annual Plant Reviews, Biology of Plant Metabolomics*. Blackwell Publishing Ltd., Oxford, UK. pp. 219–260.
- Fernie, A.R. and Schauer, N. (2008) Metabolomics-assisted breeding: a viable option for crop improvement? *Trends in Genetics* **25**, 39–48.
- Fiehn, O. (2002) Metabolomics – the link between genotypes and phenotypes. *Plant Molecular Biology* **48**, 155–171.
- Fiehn, O., Sumner, L.W., Rhee, S.Y. *et al.* (2007) Minimum reporting standards for plant biology context information in metabolomic studies. *Metabolomics* **3**, 195–201.
- Fiehn, O., Kind, T. and Kumar Barupal, D. (2011) Data processing, metabolomic databases and pathway analysis. In: Hall, R.D. (ed.) *Annual Plant Reviews, Biology of Plant Metabolomics*. Blackwell Publishing Ltd., Oxford, UK. pp. 367–406
- Fitzgerald, M.E., McCouch, S., Hall, R.D. (2009) Not just a grain of rice: the quest for quality. *Trends in Plant Science* **14**, 133–138.

22 ■ Biology of Plant Metabolomics

- Frederich, M., Jansen, C., de Tullio, P. *et al.* (2010) Metabolomic analysis of *Echinacea* spp. by H-1 NMR spectrometry and multivariate data analysis technique. *Phytochemical Analysis* **21**, 61–65.
- Gibney, M.J., Walsh, M., Brennan, L. *et al.* (2005) Metabolomics in human nutrition: opportunities and challenges. *American Journal of Clinical Nutrition* **82**, 497–503.
- Graham, S.F., Amigues, E., Migaud, M. *et al.* (2009) Application of NMR-based metabolomics for mapping metabolite variation in European wheat. *Metabolomics* **5**, 302–306.
- Hall, R.D. (2006) Plant metabolomics: from holistic hope, to hype, to hot topic. *New Phytologist* **169**, 453–468.
- Hall, R.D., de Vos, C.H.R., Verhoeven, H.A. *et al.* (2005) Metabolomics for the assessment of functional diversity and quality traits in plants. In: Vaidyanathan, S., Harrigan, G.G., Goodacre, R. (ed.) *Metabolic Profiling*. Kluwer Acad. Publishers, Dordrecht, NL, pp. 31–44.
- Hall, R.D., Brouwer, I.D., Fitzgerald, M.A. (2008) Plant metabolomics and its potential application for human nutrition. *Physiologia Plantarum* **132**, 162–175.
- Hall, R.D., de Vos, R.C.H. and Ward, J. (2010) Plant metabolomics applications in the Brassicaceae: added value for science and industry. *Acta Horticulturae* **867**, 191–205.
- Hanhineva, K., Rogachev, I., Kokko, H. *et al.* (2008) Non-targeted analysis of spatial metabolite composition in strawberry flowers. *Phytochemistry* **69**, 2462–2481.
- Hardy, N.W. and Hall, R.D. (2011) (eds.) *Plant Metabolomics Methods*. Humana Press, Ottawa, USA (in press).
- Jansen, R.C. and Nap, J.P. (2001) Genetical genomics: the added value from segregation. *Trends in Genetics* **17**, 388–391.
- Keurentjes, J.J.B. (2009) Genetical metabolomics: closing in on phenotypes. *Current Opinion in Plant Biology* **12**, 223–230.
- Keurentjes, J.J.B., Fu, J., de Vos, C.H. *et al.* (2006) The genetics of plant metabolism. *Nature Genetics* **38**, 842–849.
- Kopka, J., Walther, D., Allwood, J.W. and Goodacre, R. (2011) Progress in chemometrics and biostatistics for plant applications: or a good red wine is a bad white wine. In: Hall, R.D. (ed.) *Annual Plant Reviews, Biology of Plant Metabolomics*. Blackwell Publishing Ltd., Oxford, UK. pp. 317–342
- Lan, K., Zhang, Y., Yang, J. *et al.* (2010) Simple quality assessment approach for herbal extracts using high performance liquid chromatography –UV based metabolomics platform. *Journal of Chromatography A* **1217**, 1414–1418.
- Lapainis, T., Rubakhin, S.S. and Sweedler, J.V. (2009) Capillary electrophoresis with electrospray ionization mass spectrometric detection for single cell metabolomics. *Analytical Chemistry* **81**, 5858–5864.
- Last, R.L., Jones, A.D. and Shachar-Hill, Y. (2007) Towards the plant metabolome and beyond. *Nature Reviews: Molecular Cell Biology* **8**, 167–174.
- Lessire, R., Cahoon, E., Chapman, K. *et al.* (2009) Highlights of recent progress in plant lipid research. *Plant Physiology and Biochemistry* **47**, 443–447.
- Lommen, A. (2009) MetAlign: Interface-Driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. *Analytical Chemistry* **81**, 3079–3086.
- Luedemann, A., Strassburg, K., Erban, A. *et al.* (2008) TagFinder for the quantitative analysis of gas chromatography – mass spectrometry (GC-MS) – based metabolite profiling experiments. *Bioinformatics* **24**, 732–737.

- Ma, C., Wang, H., Lu, X. *et al.* (2009) Terpenoid metabolic profiling analysis of transgenic *Artemisia annua* by comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry. *Metabolomics* **5**, 497–506.
- Makkar, H.P.S., Siddhuraju, P. and Becker, K. (eds.) (2007) Plant secondary metabolites. *Methods in Molecular Biology*, Vol 393. Humana Press, Totowa, USA.
- Moco, S. and Vervoort, J. (2011) Chemical identification strategies by liquid chromatography (LC)-photo diode array (PDA)-solid phase extraction (SPE)-nuclear magnetic resonance (NMR)/mass spectrometry (MS). In: Hardy, N.G. and Hall, R.D. (eds.) *Methods for Plant Metabolomics*. Humana Press (in press).
- Moing, A., Aharoni, A., Biais, B. *et al.* (2011) Spatial and temporal metabolic profiling using multiple analytical platforms highlights the crosstalk between primary and secondary metabolites and mineral elements in melon fruit. *New Phytologist* (in press).
- Morreel, K., Goeminne, G., Storme, V. *et al.* (2006) Genetical metabolomics of flavonoid biosynthesis in *Populus*: a case study. *Plant Journal* **47**, 224–237.
- Oliver, S.G., Winson, M.K., Kell, D.B. *et al.* (1998) Systematic functional analysis of the yeast genome. *Trends in Biotechnology* **16**, 373–378.
- Periera, G.E., Gaudillere, J-P., van Leeuwen, C. *et al.* (2005) ¹H NMR and chemometrics to characterise mature grape berries in four wine – growing areas in Bordeaux, France. *Journal of Agricultural and Food Chemistry* **53**, 6382–6389.
- Redestig, H., Szymanski, J., Hirai, M.Y. *et al.* (2011) Data integration, metabolic networks and systems biology. In: Hall, R.D. (ed.) *Annual Plant Reviews, Biology of Plant Metabolomics*. Blackwell Publishing Ltd., Oxford, UK. pp. 261–316.
- Robinson, A.R., Ukrainetz, N.K., Kang, K-Y. *et al.* (2007) Metabolite profiling of Douglas Fir (*Pseudotsuga menziesii*) field trials reveals strong environmental and weak genetic variation. *New Phytologist* **174**, 762–773.
- Robinson, A.R., Dauwe, R., Ukrainetz, N.K. *et al.* (2009) Predicting the regenerative capacity of conifer somatic embryogenic cultures by metabolomics. *Plant Biotechnology Journal* **7**, 952–963.
- Roehlig, R.M., Eder, J., Engel, K-H. *et al.* (2009) Metabolite profiling of maize grain: differentiation due to genetics and environment. *Metabolomics* **5**, 459–477.
- Saito, K., Dixon, R., Willmitzer, L. (eds.) (2006) *Plant Metabolomics*. Springer-Verlag, Berlin.
- Schijlen, E.G.W.M., Beekwilder, J., Hall, R.D. *et al.* (2008) Boosting beneficial phytochemicals in vegetable crop plants. *CAB Reviews* **3**(25), 1–21.
- Schultz, J.C. (2002) How plants fight dirty. *Nature* **416**, 267.
- Soga, T. (2007) Capillary electrophoresis – Mass spectrometry for metabolomics. In: Weckwerth, W. (ed.) *Metabolomics: Methods and Protocols. Methods in Molecular Biology*, Vol 358. Humana Press, Totowa, USA. pp. 129–137.
- Stewart, D., Shepherd, L.V.T., Hall, R.D. *et al.* (2011) Crops and tasty, nutritious food – how can metabolomics help? In: Hall, R.D. (ed.) *Annual Plant Reviews, Biology of Plant Metabolomics*. Blackwell Publishing Ltd., Oxford, UK. pp. 181–218
- Sumner, L.W., Amberg, A., Barrett, D. *et al.* (2007) Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group Metabolomics Standards Initiative (CWAG MSI). *Metabolomics* **3**, 211–221.
- Sumner, L.W., Urbanczyk-Wochniak, E. and Broeckling, C.D. (2008) Metabolomics data analysis, visualization and integration. In: Edwards, D. (ed.) *Methods in Molecular Biology, Vol 406; Plant Bioinformatics*, pp. 409–436. Humana Press, Totowa, USA.

24 ■ Biology of Plant Metabolomics

- Sumner, L.W., Yang, D.S., Bench, B.J. *et al.* (2011) Spatially – resolved metabolomics – challenges for the future. In: Hall, R.D. (ed.) *Annual Plant Reviews, Biology of Plant Metabolomics*. Blackwell Publishing Ltd., Oxford, UK.
- Tikunov, Y., Verstappen, F.W.A. and Hall, R.D. (2007) Metabolic profiling of natural volatiles: headspace trapping GC-MS. In: Weckwerth, W. (ed.) *Metabolomics: Methods and Protocols. Methods in Molecular Biology*, Vol 358, pp. 39–54. Humana Press, Totowa, USA.
- Tikunov, Y., de Vos, R.C.H., Gonzalez-Paramas, A.M. *et al.* (2010) A role for differential glycoconjugation in the emission of phenylpropanoid volatiles from tomato fruit discovered using a metabolic data fusion approach. *Plant Physiology* **152**, 55–70.
- Timischl, B., Dettmer, K., Kaspar, H. *et al.* (2008) Development of a quantitative, validated capillary electrophoresis-time of flight - mass spectrometry method with integrated high-confidence analyte identification for metabolomics. *Electrophoresis* **29**(10), 2203–2214.
- van Dam, N.M. and van der Meijden, E. (2011) A role for metabolomics in plant ecology. In: Hall, R.D. (ed.) *Annual Plant Reviews, Biology of Plant Metabolomics*. Blackwell Publishing Ltd., Oxford, UK. pp. 87–108.
- Varshney, R.K., Nayak, S.N., May, G.D. *et al.* (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends in Biotechnology* **27**, 522–530.
- Verhoeven, H.A., De Vos, C.H.R., Bino, R.J. *et al.* (2006) Plant metabolomics strategies based upon Quadrupole Time-Of-Flight Mass Spectrometry (QTOF-MS). In: Saito K., Dixon, R.A. and Willmitzer, L. (eds.) *Plant Metabolomics*, pp. 34–47. Springer-Verlag.
- Verhoeven, H.A., Jonker, H., de Vos, R.C.H. *et al.* (2011) Solid Phase Micro-Extraction (SPME) GC - MS analysis of natural volatile components in melon and rice. In: Hardy, N.G. and Hall, R.D. (eds.) *Methods for Plant Metabolomics*. Humana Press (in press).
- Viant, M.R., Bearden, D.W., Bundy, J.G. *et al.* (2009) International NMR-based Environmental Metabolomics Intercomparison Exercise. *Environmental Science and Technology* **43**, 219–225.
- Ward, J.L., Baker, J.M., Miller, S.J. *et al.* (2010) An inter-laboratory comparison demonstrates the [¹H] – NMR metabolite fingerprinting is a robust technique for collaborative plant metabolomic data collection. *Metabolomics* **6**, 263–273.
- Weckwerth, W. (ed.) (2007) *Metabolomics: methods and protocols. Methods in Molecular Biology*, Vol 358. Humana Press, Totowa, USA.
- Wishart, D.S. (2008) Metabolomics: Applications to food science and nutrition research. *Trends in Food Science and Technology* **19**, 482–493.
- Xie, G.X., Ni, Y., Su, M.M. *et al.* (2008) Application of ultra-performance LC-TOF MS metabolite profiling techniques to the analysis of medicinal Panax herbs. *Metabolomics* **4**, 248–260.